UNIVERSITY OF PORTSMOUTH
Systems Engineering Research Group

Department of Mechanical and Design Engineering

IMPROVING CUSTOMER GENERATION BY ANALYSING WEBSITE VISITOR BEHAVIOUR

By

Shalini Ramlall

BSc

This Thesis is submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy of the University of Portsmouth, following the research conducted by the author with the Systems Engineering Research Group, Department of Mechanical and Design Engineering, University of Portsmouth and in collaboration with MotionTouch Ltd.

April 2011

ABSTRACT

This dissertation describes the creation of a new integrated Information Technology (IT) system that assisted in the collection of data about the behaviour of website visitors as well as sales and marketing data for those visitors who turned into customers. A key contribution to knowledge was the creation of a method to predict the outcome of visits to a website from visitors' browsing behaviour.

A new Online Tracking Module (OTM) was created that monitored visitors' behaviour while they browsed websites. When a visitor converted into a customer, then customer and marketing data as well as sales activity was saved in a new Customer Relationship Management (CRM) system that was implemented in this research.

The research focused on service websites. The goal of these websites was to promote products and services online and turn enquiries into offline sales. The challenge faced by these websites was to convince as many visitors as possible to enquire.

Most websites relied on Search Engine Optimisation (SEO) and Pay Per Click (PPC) advertising for traffic generation. This research used PPC advertising to generate traffic. An important aspect of PPC advertising was landing page optimisation. The aim of landing page optimisation was to increase the number of visitors to a website who completed a specific action on the website. In the case of the websites investigated in this research the action consisted of completing and sending an enquiry form from the websites.

The research looked for meaningful commonalities in the data collected by MS CRM and the OTM and combined this with feedback from the collaborating company's sales team to create two personas for website visitors who had enquired. Techniques for improving landing pages were identified and these led to changes to landing pages. Some of these changes were targeted at a particular visitor persona. The effect of changes made to a landing page was measured by comparing its conversion rate and bounce rate before and after the changes.

Behavioural data collected by the OTM was then analysed using a data mining engine to find models that could predict whether a user would convert based on their browsing behaviour. Models were found that could predict the outcome of a visit to a service website.

CONTENTS

DECLARATION

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.

GLOSSARY AND ACRONYMS

ANN	Artificial Neural Network
ce	Classification efficiency
cerr	Classification error
cf	Classification failure
CI	Confidence interval
cp	Classification probability
CRM	Customer Relationship Management
CSS	Cascading Style Sheets
CSS module	Customer Satisfaction Survey module
CTR	Click Through Rate
DoF	Degree of Freedom
GUI	Graphical User Interface
KPI	Key Performance Indicator
LPO	Landing Page Optimisation
MS CRM	Microsoft Customer Relationship Management 3.0
OTM	Online Tracking Module
PPC	Pay Per Click
ROI	Return On Investment
RSq	R-Squared
SEO	Search Engine Optimisation

SERP	Search Engine Results Page
SQL	Structured Query Language
StdDev	Standard Deviation
StdErr	Standard error
URL	Uniform Resource Locator
VBA	Visual Basic for Application

ACKNOWLEDGEMENTS

I would like to express my gratitude to the following:

To Dr David Sanders, Dr Giles Tewkesbury and Dr David Ndzi for agreeing to supervise the research and for their help and assistance during the writing-up. Again to Dr David Sanders for providing me with the opportunity to conduct this research and for his leadership, commitment and encouragement throughout the research. For his support beyond academic matters and for being an inspirational mentor who, through his invaluable advice and guidance over the last 7 years has taught me a lot.

I would also like to thank MotionTouch Ltd for partly sponsoring my research, especially: Henry Powell (Company Director) for his encouragement and personal support; and Daniel Furmanek (junior web designer) for writing the code for the new dynamic website and OTM.

I would like to express my eternal gratitude:

To my family for their lifelong support, encouragement and love especially my parents for the selfless sacrifices that they have made to help me get to where I am now. And also to my sisters and my brother for always being there to lift my spirits and for their unwavering belief in me.

My sincere thanks go to:

My friends from Portsmouth for their support and camaraderie as well as the Friday night drinks which were the best remedy after a long stressful week.

PUBLICATIONS

Ramlall S., Sanders D.A., Tewkesbury G.E. & Ndzi D., (2011a). Can Keyword Length Indicate Web Users' Readiness to Purchase? Accepted by *7th International Conference on Web Information Systems and Technology (WEBIST).*

Ramlall S., Sanders D.A., Tewkesbury G.E. & Ndzi D., (2011b). Can Keyword Length Indicate Web Users' Readiness to Purchase? Accepted by *International Conference on Internet Computing (ICOMP).*

Ramlall S., Sanders D.A., Tewkesbury G.E., Powell H. & Ndzi D., (2009). Using Integrated IT Systems to Monitor Decay in Website Performance Following Individual Improvements. Journal of Computing in Systems and Engineering, 20, (ISSN: 1472-9083).

CHAPTER 1

INTRODUCTION

"*The Internet has brought about a fundamental change in the way users generate and obtain information, thereby facilitating a paradigm shift in consumer search and purchase patterns*" (Ghose and Yang, 2008b).

At the time of writing, the Internet had become a powerful advertising medium and an online presence was becoming essential for a successful sales and marketing strategy.

Online advertising had grown in popularity since its advent. Online advertising in the United Kingdom (UK) had a market share of 24.3% and was worth £1,968.6m in the first half of 2010. During that time, search advertising was the most popular online advertising format, accounting for 60% of the total online advertising spend (IAB UK, 2010).

The success of online advertising campaigns depended on the ability of websites to convert traffic generated by campaigns into customers. Implementing a website that had a high conversion rate involved designing web pages and browsing experiences that satisfied the needs of visitors. The design process was iterative and aimed to optimise a website by changing design elements in a controlled and measurable manner.

An important aspect of online advertising was to identify target customers, understand who they were and to use that knowledge to create advertising campaigns and design websites. It was also important to gather data to gain insight into how visitors behaved on websites so as to optimise their experience and encourage conversion.

1.1. Research Aims and Objectives

The aim of this research was to create new software systems to investigate whether conversion on a service website could be predicted based on visitors' behaviour on such a website. In addition, the research aimed to optimise Pay Per Click (PPC) advertising campaigns and landing pages so as to increase online conversion. Specific objectives were to:

- Investigate existing Customer Relationship Management (CRM) software.
- Implement a CRM strategy and software system.
- Customise and extend a CRM system by creating new user interfaces and new modules to support business processes.
- Investigate website design, navigation design and identify ways of measuring website performance.
- Create a new dynamic website.
- Create a new online tracking system that recorded detailed visitor activities and behaviour on a website.
- Investigate online advertising, Pay Per Click advertising and landing page design and optimisation.
- Create new PPC campaigns and landing pages.
- Use data collected in a CRM system and an online tracking system to understand the needs of website visitors and customers and to optimise PPC campaigns and landing pages so as to increase online conversion.
- Investigate factors that influenced visitors' behavior and their relation to conversion.

- Find ways to infer whether a visitor would convert based on their behavior on a website.

1.2. Methodology

CRM systems were investigated and the Microsoft Dynamics Customer Relationship Management 3.0 (MS CRM) system was chosen for implementation. The system was customised and extended in order to enable it to record customer data throughout a sales and product delivery cycle. A new Online Tracking Module (OTM) was implemented on existing websites so as to track visitors' activities. Data collected by the OTM was integrated with data collected by MS CRM so that a history of online behavioural activity and sales activity for website visitors who became customers was captured. PPC advertising campaigns were created to generate traffic to existing websites. Using the data collected about visitors and customers together with feedback from the sales department, basic visitor personas were created. These personas provided some understanding regarding who prospective customers were as well as their needs. Using these personas, landing pages were optimised in order to increase online conversion.

1.2.1. CRM

The collaborating company was using the GoldMine Business Contact Manager (GoldMine) system at the beginning of the research to manage some of its CRM activities. GoldMine was not a CRM system; it was a contact management system. It was difficult to customise and as a result collected little information about customers. The data that was collected provided limited insight into customers or their needs. Also the system was not suitable for managing customer relationship efficiently. Therefore, a new CRM system was needed for the research.

Requirements for a new CRM system were derived from the collaborating company's CRM strategy and business processes as well as feedback obtained from its staff. Several CRM systems were investigated, including MS CRM, SugarCRM and the latest version of GoldMine. MS CRM was chosen for implementation as the out-of-the-box system met most of the requirements that had been specified. The rest of the requirements could be met by developing custom add-ons for MS CRM.

Once MS CRM was deployed at the collaborating company, then custom modules were implemented to support business processes and meet requirements that the out-of-the-box product could not. Four new modules were implemented: a Customer Satisfaction Survey (CSS) module, Project Management module, Quality Control module and Opportunity Marker module.

Reports and dashboards were implemented. They used the data stored in MS CRM to display information about key business performance indicators (KPI's). The data collected by MS CRM was richer than had been available in GoldMine and was used to extract knowledge about customers. Integration with other IT systems was investigated but not carried out due to the costs involved.

1.2.2. Websites

At the beginning of the research, websites at the collaborating company were static and did not track visitor activity. As a result little was known about the type of traffic and visitors that online marketing campaigns generated.

A first version of a new Online Tracking Module (OTM) was implemented to capture visitors' marketing and behavioural data. This data was integrated with data collected by

MS CRM and was used to identify and optimise advertising campaigns that generated the types of customers that the collaborating company was targeting.

A new back-end was implemented for the collaborating company's main website so as to enable it to support dynamic content and navigation personalisation. The functionalities required to achieve this included the ability to infer customer interest from browsing behaviour, dynamic menu generation, dynamic content generation, dynamic inclusion of pictures and dynamic web page generation.

The first version of the OTM was built around the structure of existing websites and as a result some aspects of the implementation lacked flexibility. When the new backend of the main website was designed and implemented, it presented an opportunity to re-design some parts of the OTM as well as implement new functionality. The new version of the OTM could differentiate between new and returning visitors, used sessions to identify activities related to each returning visit and collected more details about visitors' behavioural activities.

After the implementation of the new version of the OTM, the front-end of the new dynamic main website was re-designed to give the website a professional and industry-appropriate look. During this re-design, the website's content was updated and its main menu was re-structured.

1.2.3. Pay Per Click advertising

Most Web users relied on search engines to find what they were looking for on the Internet. PPC advertising was a form of "*text based online advertising*" (Burns, 2005), where an Internet user entered keywords into a search engine and an advertisement was displayed on the top or the sides of a results page.

Google was one of the most popular search engines. In 2009, Google owned 82% of the global search engine market share (Netmarketshare, 2010). With such global reach, anyone advertising through Google's PPC service (called Google AdWords) was likely to reach a large number of potential customers.

The collaborating company provided two types of services: design and manufacturing. These services were advertised in the United Kingdom (UK) and United States (US). Each service had an advertising campaign and each campaign was duplicated and run in the US and the UK.

Each campaign had several of Ad Groups. The Ad Groups were created around the different sub-types of the design and manufacturing services for example plastic moulding, plastic manufacturing, mechanical design and electronic design.

The research created several Google AdWords campaigns to drive traffic to the collaborating company's websites. A campaign could have one or more advertisement groups. Organising campaigns into smaller and separate advertisement groups that were based around products, services or goals allowed for better targeting and customisation. Each advertisement group consisted of a set of keywords, text advertisements and landing pages that were aimed at a specific audience.

Google AdWords provided a number of performance measures for evaluating the effectiveness of campaigns, advertisement groups, keywords, advertisements and landing pages. Some of these measures were used during this research. They included: number of times an advertisement was displayed, click-through rate, cost per conversion, number of conversions and conversion rate.

In order to optimise the performance of advertisement groups, several experiments were conducted. First, text advertisements were tested to identify the best ones. The effectiveness of an advertisement was measured by its click-through rate (CTR). Then, changes were made to the landing pages of advertisement groups in order to increase conversion.

1.2.4. Landing page optimisation

In order to convert traffic generated by online advertising campaigns into customers, landing pages had to be optimised. Landing Page Optimisation (LPO) was achieved by improving the content and design of landing pages so that they appealed to targeted visitors and encouraged them to enquire about the services that were advertised.

Data collected by MS CRM and the OTM, together with feedback from the collaborating company's sales team were used to create two basic visitor personas for prospective customers. The personas attempted to identify who visitors were, what they were looking for, what problems they were trying to solve and where they were in the buying cycle.

Landing page design was investigated and elements that could influence visitor behaviour were identified. Some of the elements identified were used to implement new landing pages that could appeal to visitors who fitted the personas that had been created. Several experiments were conducted to test the performance of these new landing pages.

Design elements that were identified and implemented included:

- Using content to create targeted landing pages.

- Reinforcing and extending an advertisement by ensuring that a landing page provided what the advertisement promised and used keywords from the advertisement in its title.
- Improving visual design.
- Complying with usability guidelines when writing and structuring content.
- Visitor segmentation.

1.2.5. Experiments and Results

Data collected by the OTM about the way visitors interacted with the collaborating company's main website was analysed to find models that could predict whether a visitor would convert based on their browsing behaviour. Models were also sought to identify whether the length of search keywords could indicate readiness to convert based on a search-conversion model proposed in this research.

The data mining process was implemented in four stages:

- Population sampling – Research identified the criteria used to select the data that would be used for data mining.
- Data retrieval – Behavioural data was extracted from the OTM database.
- Data cleaning and transformation – Errors and inconsistencies in the data were identified and removed. A *Visual Basic for Applications* (VBA) macro was then used to transform the data into behavioural attributes that could be used by a data mining tool to find rules and patterns.
- Knowledge discovery – A data mining tool was used to analyse the data to find models that could predict conversions.

Several explorations of the data were carried out using three different prediction algorithms. A Neural Networks algorithm found the most accurate models for predicting conversion.

Search keyword length distribution was analysed. For the purposes of this research, search keyword length was defined as the number of words found in a visitor's search keyword. A search keyword could be a single word or a phrase that a visitor had typed into a search engine and that had led them to the websites used in this research.

Some evidence was found to support two hypotheses proposed by this research. The first hypothesis suggested that longer search keywords indicated that Web users were more ready to convert. The second hypothesis suggested that shorter search keywords indicated that Web users were less ready to convert.

Relevancy attributes for search keywords were derived using a scoring method, and then used to generate new models. However, the new attributes did not improve the prediction accuracy of these models.

The accuracy of the models found by Neural Networks was better than the accuracy of a naive prediction model. A naïve prediction is a model which given 2 outcomes 0 and 1 will predict all outcomes as either all 1 or all 0. A naïve prediction model is one that has not learnt from training data. It is the best guess that can be made without any other information. It was concluded that Neural Networks could be used to derive models from behavioural data to predict conversion. However, because the Neural Networks did not display the weights associated with input attributes, it was difficult to know how individual behavioural attributes influenced conversion. Nevertheless, a new method was created to extract attributes from behavioural data and use them to predict the outcome of visits on a service website.

1.3. Research Claims

Research into CRM systems, website design, online advertising, landing page optimisation and online customer behaviour was undertaken. The research work has resulted in the following achievements:

- A customised Customer Relationship Management (CRM) system was created that included new custom modules to:
    - Improve project management (Project Management module).
    - Improve quality control (Quality Control module).
    - Measure customer satisfaction (Customer Satisfaction Survey module).
    - Score sales opportunities (Opportunity Marker module).
- Dashboards and reports were created to provide real time, graphical representations of data stored in a CRM system.
- Data integration with a new Online Tracking Module was achieved to enable:
    - Web browsing data to be cross referenced with customer data stored in MS CRM.
    - Identification of the sources that customers originated from.
    - Identification of profitable online advertising campaigns.
    - Identification of the type of customers that each online campaign generated.
    - Calculation of Return On Investment for individual online advertising campaigns.
    - Collection of data about customers from the time they landed on a website and throughout a sales cycle.
    - Ability to associate customer’s online behavioural activities with sales profit.

- A new Online Tracking Module was created to enable:
  - The capture of detailed browsing activity.
  - Distinguishing between new and returning visitors.
  - Use of sessions to identify returning visits
  - Recording of lead quality scores both for email enquiries and telephone enquiries.
  - Associating lead quality score with individual online advertising campaigns.
  - Tracking of campaigns and advertisements that attracted Web users to a website.
  - Data integration with the MS CRM system.
- A dynamic website was created to provide the ability to:
  - Infer customer interest based on browsing activity.
  - Modify content to match the search keywords that a visitor used.
  - Match the pictures displayed on a web page to the search keywords that a visitor used.

New methods were created for:

- Collecting data about the behaviour of website visitors.
- Integrating CRM data with online behavioural data.
- Modelling search-conversion for inferring readiness to convert from keyword length.
- Optimisation of PPC campaigns and landing pages thorough the use of visitor personas. These personas were created using data collected by MS CRM and the OTM as well as feedback from the sales team at the collaborating company.
- Inferring experiential or goal-oriented behaviour from keyword length

- Deriving behavioural attributes from data recorded about the way users interacted with a website.
- Using Neural Networks to predict conversion.

New results were obtained:

- Changes in the conversion rate and bounce rate of landing pages were observed when modification based on the following concepts were made to the landing pages :
    - Using content to create targeted landing pages.
    - Reinforcing and extending an advertisement by ensuring that a landing page provided what the advertisement promised and used the wording of the advertisement in the title.
    - Improving visual design.
    - Complying with usability guidelines when writing and structuring content.
    - Visitor segmentation.
- The data collected by the OTM and MS CRM produced new metrics for measuring the impact of changes to landing pages:
    - quality of leads.
    - type of customers who enquire, for example, individuals, small companies or corporate.
- The distribution and occurrence of search keyword length in the data collected by the OTM was analysed. The analysis showed that
    - One and two word search keywords were associated with more non-conversions than conversions.
    - Longer search keywords (containing more than two words) were associated with more conversions than non-conversions.

- A data mining tool was used to analyse data collected by the OTM to find models that could predict conversion using visitors' behavioural attributes. New predictive models were found. These were more accurate than a naive prediction model.

1.4. Overview of Dissertation

Chapter 2 describes the literature research conducted into CRM systems, websites, website performance, online search behaviour, landing pages, data mining and knowledge discovery.

Chapter 3 describes the CRM system that existed at the beginning of the research and its limitations. It then describes the specifications for a new CRM system and its implementation as well as the custom modules that were developed to extend the functionality of the new CRM system. Finally, Chapter 3 describes the type of information that was extracted from the data collected by the new CRM system during initial analyses.

Chapter 4 describes the websites that existed at the beginning of the research and the implementation of the first version of the Online Tracking Module (OTM). It then goes on to describe the implementation of a new dynamic main website and a new and improved version of the OTM. Finally, it describes the changes made to the front-end of the new dynamic main website.

Chapter 5 describes the Google AdWords service and how it was used to set up PPC campaigns to drive traffic to the websites described in Chapter 4. It describes how advertising campaigns were structured, set up and configured and the metrics that were used to measure and monitor their performance.

Chapter 6 describes the visitor personas that were created by this research and the changes that were made to landing pages in order to achieve higher conversion rates.

Chapter 7 describes how data collected by the new OTM described in Chapter 4 was analysed and the results obtained.

Chapter 8 describes the research finding, discusses the successes and failures of the research and provides recommendations for future research.

CHAPTER 2

LITERATURE REVIEW

This chapter reviews the systems, technologies, methodologies and measures related to the creation of the new systems described in chapters 3, 4, 5 and 6 - including Customer Relationship Management (CRM) systems, website design and the Per Per Click (PPC) model.

2.1. Customer Relationship Management

Customer Relationship Management (CRM) has often been mistaken for CRM technology (Reinartz et al., 2005). CRM was "*a concept that comprises the establishment, development, maintenance and optimisation of long-term, mutually valuable relationships between customers and organisations*" (Payne and Ryals, 2001).

Shaw and Reed (1999) suggested that CRM entailed:

- Continuously collecting and updating knowledge on "*customer needs, motivations, and behaviour*" throughout the relationship.
- Using knowledge gained from customers in the form of success and failures to improve performance.
- Integrating various departments and business functions to achieve a common goal.
- Implementing systems to support:
    - acquisition and sharing of customer data and knowledge.

- Measurement of CRM effectiveness.

CRM technology or systems refer to the information technology (IT) implemented and deployed to support activities associated with customer relationship management, namely "*sales support, service support, analysis support, and data integration and access support*" (Chang et al., 2010).

CRM systems could gather large amounts of data which could be used to build knowledge about clients. Effective use of this knowledge could help companies understand "*customers' tastes and preferences*" (Lin et al., 2010) and build closer and more profitable relationships with clients.

2.1.1. Types of Customer Relationship Management Systems

Three types of CRM systems were identified by Marjanovic (2005):

a) Analytical customer relationship management.
b) Collaborative customer relationship management.
c) Operational customer relationship management.

a) Analytical Customer Relationship Management

The main aim of this type of CRM system was to store, analyse and report on data about customers and their interaction with a company. The data was used to understand customers' needs and assess customer satisfaction as well as predict future behaviour and help decision making regarding marketing, sales and customer support. By using data mining and data warehousing technologies analytical CRM systems could store and mine data.

b) Collaborative Customer Relationship Management

The main purpose of this type of system was to promote collaboration amongst different departments in a company, for example, sales, marketing and engineering, by enabling the sharing of customer data. This type of CRM system achieved collaboration through the use of emails, conferencing and chat tools to promote teamwork and information sharing within an organisation.

c) Operational Customer Relationship Management

The main purpose of this type of CRM system was to improve the efficiency of doing business with customers. It focused on automating business processes which were initiated by customers such as their interaction with sales and marketing, technical support and shipping.

2.1.2. Existing CRM systems

At the beginning of this research, the collaborating company used the GoldMine system to manage some of its CRM activities. GoldMine had a number of limitations which made it unsuitable for customer relationship management, for example, the system was difficult to customise and as a result, limited information about customers was collected.

A number of commercial CRM systems were investigated at the beginning of this research and the following systems were reviewed:

a) GoldMine 6.0
b) Microsoft CRM 3.0
c) SugarCRM

a) GoldMine 6.0

GoldMine 6.0 was "*a contact manager for individuals and teams, specifically designed for small-and mid-sized organisations to bridge the gap between traditional contact managers and complex Customer Relationship Management (CRM) solutions*" (FrontRange Solutions Inc, 2002).

The main features of GoldMine 6.0 were:

- Contact management.
- Activities and task management.
- Sales management.
- Reporting.

GoldMine 6.0 did not provide features that supported marketing campaigns, case management or customer service management. GoldMine 6.0 was not a complete CRM solution. It was more of a contact management product.

(FrontRange Solutions Inc, 2002)

b) Microsoft CRM 3.0

Microsoft CRM 3.0 (MS CRM) was a software platform that companies of all sizes could use to implement their CRM strategies. MS CRM was released in 2005 with modules for sales, marketing and customer service. MS CRM integrated directly into Microsoft Outlook and other Microsoft applications. This allowed users to work with applications with which they were already familiar, thus promoting adoption and use within an organisation.

(Snyder and Steger, 2006)

The main features of MS CRM included (Crm Connected, n.d.):

- Activity and task management.
- Account, contact, lead and opportunity management.
- Quotes, orders, invoices and case management.
- Workflow.
- Marketing automation.
- Microsoft CRM client for outlook (online and offline version).
- Reporting.

The main benefits of MS CRM were:

Working efficiently from Microsoft Outlook. MS CRM enabled users to work from Microsoft Outlook. The tight integration between MS CRM and Microsoft Outlook as well as other Microsoft products enabled users to work with MS CRM within their usual work environment.

Understanding customers. Microsoft SQL Server Reporting Services integrated with MS CRM, which enabled customer data to be analysed and reports to be created. This provided insight into customer preferences and behaviour.

Security of data. MS CRM allowed role-based and permission based access to customer data which helped ensure that users only had access to data that they were authorised to see.

Customisation. MS CRM had built-in customisation tools. These enabled quick customisation of application views, forms, data fields and relationships between entities. For more complex customisations, MS CRM could be extended using add-on software written in .NET.

Workflow. MS CRM workflow was a tool that helped set up and define business processes as a number of automated tasks. Workflow could be used to coordinate work between different teams or departments.

(SolutionMark, n.d.)

c) SugarCRM

SugarCRM was a popular open source CRM system that was also available as a fully licensed commercial product that offered a broad set of features. SugarCRM's open source architecture made customisation and integration of customer centric business processes relatively easy. SugarCRM was built on open-source technologies that included PHP, MySQL, and the Apache Web server.

The open source version of SugarCRM offered the following functionalities:

- Sales management which included lead, account and opportunity management.
- Marketing automation which included email marketing and management of marketing campaigns.
- Customer Support which included case management and tracking.
- Reporting.

The main advantages of the open source SugarCRM systems were:

Affordability. Since it was an open source product there were no financial costs involved in procuring SugarCRM thus making it the ideal solution for budget conscious organisations.

Customisation. The system was completely customisable as it was an open source solution. Source code and documentation were readily available and an active community also provided support in developing add-ons.

The main drawbacks of the open source SugarCRM system were:

Open source. Since SugarCRM was an open source product, organisations would require in-house programmers to customise it. Customisation and configuration efforts could be costly.

Small business solution. It was a small business only solution. The fully licensed commercial version of the product was the better solution for corporate organisations.

Poor reporting. Its built-in features for pipelines, forecasts and reporting were poor.

Administration and security. It had limited tools for system administration tasks such as merging accounts and mass modifications. Also, its security model was not flexible.

(Online-CRM.com, n.d.)

2.1.3. Customer Relationship Management Implementation

A CRM strategy could be expensive and difficult to implement. "*CRM is a multi-faceted, comprehensive phenomenon which includes strategic aspects, customer-oriented processes and organisational changes through projects as well as performance measurement. In addition, IS (information systems) implementation—which has mistakenly become a synonym for CRM—is an important element*" (Meyer and Kolbe, 2005).

Companies that believed that CRM facilitated the implementation of better customer-focused strategies invested heavily in CRM technology but it was found that only 30% of companies achieved improvement in performance (Chang et al., 2010). CRM solutions were not always successful and lost investments and absence of Return On Investment (ROI) have been cited by some academic authors (Bull, 2003, Kotorov, 2002). CRM implementation often focused on the deployment of a software package "*without an in-depth understanding of the issues of integrating culture, process, people, and technology within and across organisational context*" (Finnegan and Currie, 2010).

Factors affecting CRM Implementation

There were a number of factors that could affect the implementation and success of CRM (strategy and system). Some of these factors were:

a) Leadership.

b) Customer centric organisational culture.

c) Project Management.

d) Complexity.

e) Systems Integration.

a) Leadership

Effective leadership was important as the introduction of CRM technology in an organisation involved changes to business processes and introduction of new information technology (IT) or changes to existing IT structures. Leaders were also the ones who had the best understanding of an organisation's goals and how CRM could help achieve them (Bull, 2003).

b) Customer centric organisational culture

In order to deliver financial benefits, CRM activities needed to contribute to a company's performance and "*customer orientation and especially CRM are important preconditions for the realisation of profitability*" (Meyer and Kolbe, 2005).

CRM systems could not be regarded simply as contact management systems but rather as tools which, when used efficiently could help cultivate good relationship with customers (Ciborra and Failla, 2000).

Organisations needed to adopt a holistic approach to CRM whereby CRM systems were integrated with customer oriented business processes and customer service delivery. A holistic approach helped coordinate the various functions of CRM and ensured that operational CRM, analytical CRM and collaborative CRM complemented each other (Thompson et al., 2006).

c) Project management

Meyer and Kolbe (2005) found that bad project management and poor collaboration within organisations, with the latter mainly due to "*technical and organisational barriers*" were amongst the reasons why some CRM implementation failed.

d) Complexity

It appeared that a lack of alignment with an organisation's goals and underestimation of its complexity were the main reasons why CRM solutions failed (Bull, 2003, Piercy, 1998). This was because the implementation of a CRM strategy and system involved various stakeholders, with different skills and experiences. The coordination of the stakeholders throughout the implementation of the CRM solution could be challenging from a "*strategic, process and system perspectives*" (Mayer and Kolbe, 2005).

e) Information Systems (IS) Integration

The success of a CRM implementation depended on the coordination and collaboration that existed within an organisation. IS integration was crucial to a customer focused strategy (Bhatt and Troutt, 2005). Decision making for strategic input, relied on the use and analysis of data related to various business processes/activities that companies carry out. Gupta et al. (2006) suggested that organisations needed to exploit existing resources and knowledge to achieve efficiency in their operations.

There was a need for integration of business processes, software applications and data. Zahra and Nielsen (2002) stated that the '*integration of internal and external sources is positively associated with successful technology commercialisation*'. They suggested that both formal and informal integration contribute to this success. Other studies have supported that integration had a positive impact on systems performance, functional units and organisations (Kahn and Mentzer, 1998, Paashuis and Boer, 1997). Ultimately integration could lead to better relationship with customers and suppliers (Braganza, 2002).

It was assumed that information systems integration always resulted in net benefit to a company. However, Goodhue (1992) argued that the impact of data integration and benefits of information systems depended on organisational structure and that data integration could bring about "*gains in organisation-wide coordination and organisation-wide decision making, as well as losses in local autonomy and flexibility, and changes in system design and implementation costs*".

2.1.4. Measuring the impact of CRM

CRM could have an impact on a company's success and profitability (Wilson et al., 2002, Payne and Ryals, 2001, Dowling, 2002, Kotorov, 2002). CRM required considerable financial investments and changes in operational and organisational structures (Homburg et al., 2000, Wilson et al., 2002, Kotorov, 2002). It was important to be able to justify this investment with a ROI. ROI for CRM systems was usually achieved through customer retention and customer satisfaction. Reichheld and Sasser

(1990), showed that customer retention can have a big impact on profitability. The longer the customer relationship, the greater its profitability (Reichheld and Teal, 1996, Storbacka et al., 1994, Yeung and Ennew, 2000). Research has also shown that it was cheaper to retain existing customers than it was to acquire new customers (Reinartz et al., 2005, Phan and Vogel, 2010).

Customer satisfaction was important to the success of companies and CRM could help them achieve this. When CRM was used to increase customer satisfaction, the factors (service quality, perceived value, trust and commitment) that supported customer satisfaction became the antecedents to CRM success (Meyer and Kolbe, 2005). According to Liu and Zhu (2009), CRM not only improved customer satisfaction but also promoted loyalty. It followed from the research carried out into customer retention and customer satisfaction that if CRM was used effectively to support these two functions, it could deliver financial benefits.

2.1.5. Discussion

This Section described the CRM systems that were investigated at the beginning of this research as well as the factors that needed to be taken into consideration in order to successfully implement a CRM strategy.

MS CRM was the better choice both from a functionality point of view and a cost benefit point of view. It also combined analytical, collaborative and operational CRM into one product. Moreover, it integrated seamlessly with other Microsoft applications which meant that users worked within an environment that was already familiar to them. This could encourage adoption and use of the system. The main drawback of MS CRM was that as a relatively new product some features had limited capabilities and flexibility.

Some measures of CRM success were also identified, for example, increase in customer satisfaction, customer retention and profit. The existing literature suggested that an integrated CRM system could bring many benefits to an organisation (Braganza, 2002, Goodhue et al., 1992, Zahra and Nielsen, 2002). Integration did not necessarily need to be formal integration such as systems and data integration. Informal integration in the form of business process overlap within a CRM strategy could also bring positive results. Chapter 3 describes how MS CRM was implemented to support a CRM strategy in an SME.

2.2. Online Advertising

Online advertising could be regarded as the buying and selling of advertising space on the Internet. There were different types of online advertising:

- Search advertising or Pay Per Click (PPC) advertising where advertisements appeared on search engines' result page.
- Display advertising or banner advertising that appeared on websites.
- Classified listings which appeared on specialist websites.
- Internet email based advertising.

(Evans, 2008)

2.2.1. Brief history

Online advertising started when HotWire sold banner advertisements to a number of advertisers in 1994. Online advertising started gaining momentum until the dot-com bust reduced the demand for it. With the advent of the Web 2.0, online advertising regained momentum and by 2004 companies such as Overture, Advertising.com and Google were offering online advertising services (Evans, 2008).

Since then the online advertising market has grown steadily to become the second largest advertising medium; preceded only by television. It was worth £1,968.6m in the UK in the first half of 2010. This represented an increase of 10%, on a like for like basis, on the previous year. The breakdown of the market in terms of the different advertising formats was as follows:

- Search advertising – 60%
- Display advertising – 19%
- Classified – 19%

PPC advertising was by far the most popular online advertising format.

(IAB UK, 2010)

2.2.2. Pay Per Click Advertising

In the beginning, Internet advertising followed the offline advertising model. Advertising banners were the online equivalent of the offline print advertisements. Search engines that provided banner advertising faced a dilemma regarding whether to encourage visitors to stay on their website and view more advertising or to promptly send them to the websites that were listed in search results. In general, visitors found banner advertising irritating and distracting.

PPC advertising solved these issues both for search engines and visitors. PPC advertising effectively tied the revenue of search engines to the act of transferring a visitor to an advertiser's website. From the visitors' point of view, PPC advertising delivered relevance as advertisements were matched to their search terms. It was also unobtrusive since advertisements were delivered as and when a visitor carried out a search.

(Fain and Pedersen, 2006)

The PPC concept was first introduced in 1998 by Overture Services. Overture Services was acquired by Yahoo! in 2003 and re-branded as Yahoo! Search Marketing in 2005. Google launched its PPC service in 2002.

Search engines that provided PPC advertising not only displayed results on their own sites but also in space that they rented on other search engines' sites. For example when a Web user searched on AOL, the search query was passed to Google, which returned the results to AOL, which then displayed these on a search results page that it rendered to the Web user. Similarly Yahoo! rented space on MSN Search.

(Evans, 2008)

At the time of writing, Google was the most popular PPC advertising provider both in the United Kingdom and the United Stated of America (Charlton, 2010, Larsen, 2010).

Pay Per Click advertising was described as an advertising concept whereby advertisers pay a fee to Internet search engines to have their advertisements displayed alongside organic (non-sponsored) search results. The PPC mechanism worked as follows (Ghose and Yang, 2008b):

1. Advertisers identified keywords that described their products or services and submitted them to a PPC service provider.
2. Bid values were assigned to individual keywords. This determined how much the advertiser had to pay the provider when a Web user clicked on an advertisement that was triggered by the keywords. Bids also determined the position of the advertisement on the search engine result page (SERP).
3. When Web users searched for products or services online, they started their search by typing keywords (that can consist of multiple words) into a search engine (Rutz and Bucklin, 2007). Advertisements associated with their search

keywords were displayed on the SERP. Upon clicking on an advertisement, a Web user was taken to an advertiser's website. Search engines used propriety models to determine which advertisements were displayed and at what position on the SERP.

Companies used PPC advertising to drive customers to their websites. £1,180.1 million was spent on PPC advertising in the United Kingdom in the first half of 2010. This accounted for 59.9% of the online advertising revenue (IAB UK, 2010). In the United States of America, $4.4 billion was spent on PPC advertising during the first half of 2010 accounting for 36% of the total online advertising revenue (IAB USA, 2010).

PPC advertising allowed companies to address consumers directly during their electronic search for products or services. It was popular because it had "*enabled a shift in advertising from 'mass' advertising to more 'targeted' advertising*" (Ghose and Yang, 2008b). As such it delivered:

- More relevant search results to Web users hence improving their online experience.
- Higher quality visitors to websites.

Ghose and Yang (2008b) showed that "*on average the conversion rates, order values and profits from paid search advertisements were much higher than those from natural search*."

In order to run successful and optimised PPC campaigns that generated ROI, it was important to have a PPC strategy that considered the following (Burns, 2005):

a) Search Engine Optimisation (SEO).

b) Relevancy.

c) Campaign goals.

d) Targeted keywords.

e) Campaign structure.

f) Advertisement text.

g) Landing page.

h) Tracking results.

Burns (2005) described these factors as:

a) Search Engine Optimisation

Search engine optimisation (SEO) was "*a process that manipulated website characteristics and incoming links to improve a site's ranking in the search engines for particular search terms*" (Malaga, 2010). Although, not a so-called advertising technique, it played an important role in website exposure and traffic generation. SEO was thought of as a cheap way of generating traffic compared to PPC advertising. However, this was not always true as considerable investment both in terms of technical effort and SEO expertise was required in optimising and keeping a website optimised for first page listing.

SEO ensured that a website was well designed both for search engine crawlers and Web users. Crawlers were used by search engine to find and index sites. SEO reinforced the best practises in designing user-friendly websites with emphasis on content and structure.

b) Relevancy

Relevancy was important both in the advertisement and on the website. Advertisements that connected Web users with exactly what they were searching for were the most successful and profitable

c) Campaign goals

Specific goals needed to be established for advertising campaigns so that the right keywords, advertising messages and landing pages could be selected. Metrics for measuring these goals were also important.

d) Targeted keywords

Building relevant and targeted keyword lists was important to the success of PPC campaigns. A keyword list that included terms that generated clicks from Web users who were not looking for the advertised product or service could be costly.

There were free keyword generation tools that were available for example Google's Keyword Suggestion Tool. In general, it was better to use multi-word keywords rather than single keywords which were broad and vague. Using brand name and product words were also better for targeting.

e) Campaign structure

Organising keywords into separate advertisement groups that were based around products, services or goals allowed for generation of custom advertisements, web pages and content. Customisation increased relevancy which was important in turning visitors into customers. Figure 2.1 illustrates a PPC advertising campaign structure.

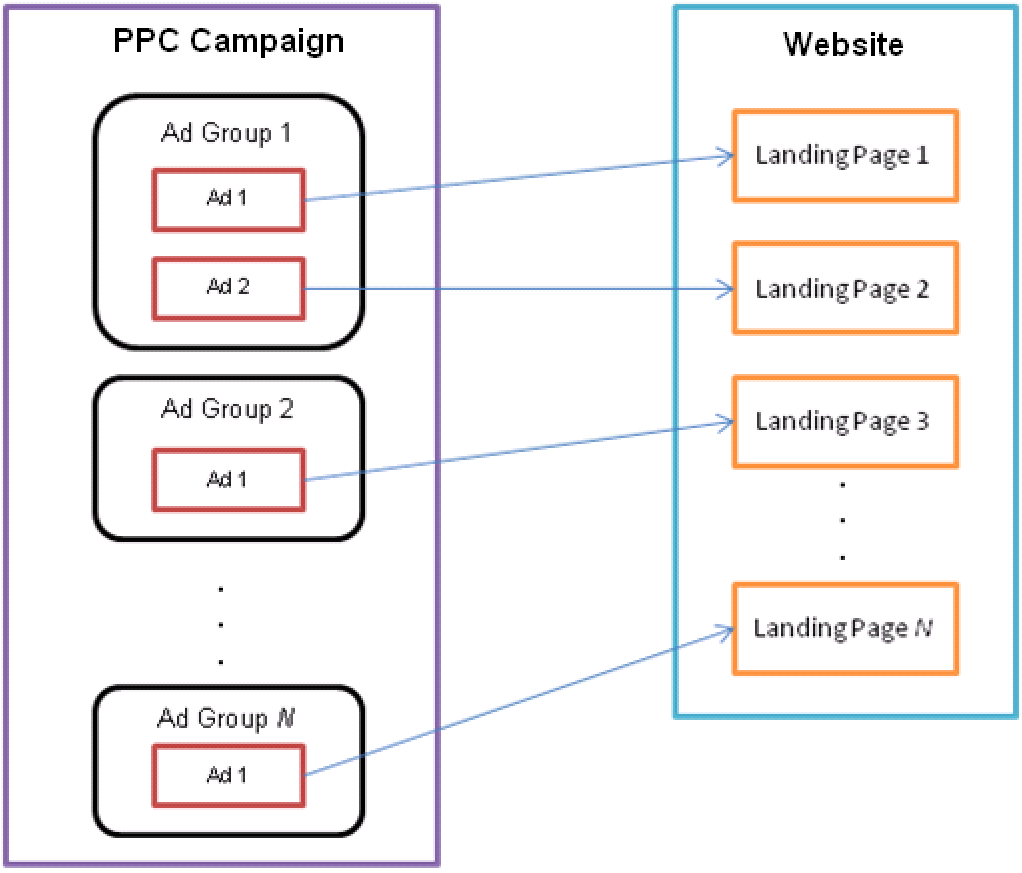


Figure 2.1: PPC advertising campaign structure.

f) *Advertisement text*

The quality of the advertisement's copy was crucial. It needed to be clear and compelling. Some of the techniques used generate good advertisement copy included:

- Writing and testing multiple advertisements for each ad group.
- Putting keywords into the copy to increase relevance.
- Using strong call to action.
- Putting forward a unique proposition or offer.

g) Landing page

Landing pages played a crucial role in the success of advertising campaigns. Success could be measured in term of conversion rate. Conversion rate was defined as the percentage of total visitors who completed a specific action on a website e.g. registered for a newsletter, sent an enquiry email or purchased a product. A website could consider one or more actions as a conversion.

Landing pages had a unique and distinct function from other pages in a website. As such there were unique challenges that had to be met when designing them. Loveday and Niehaus (2008) identified that landing pages faced particular challenges. They had to:

- singlehandedly take customers through the whole sales cycle. It had to "*create or reinforce interest, then instil desire, and finally guide visitors to take action*".
- perform quickly.
- deal with a high number of first time visitors, who were not "*familiar with the company and [had] no reason to trust it at first*".

Landing pages needed to be specific to the advertisement. The content and design of the landing pages had to relate to the advertisement; they needed to be an extension of the advertisement (Loveday and Neihaus, 2008).

Multiple landing pages could be used for individual advertisements as a way of finding the one that produced the best conversion rate. Over their lifetime, landing pages often went through an optimisation process in order to make them more effective.

h) Tracking and measuring results

In order to optimise ROI for PPC campaigns, it was important to try different keywords, advertisements, landing pages and content, and to measure how each performed.

PPC service providers reported on a number of metrics that could help gauge the performance of campaigns. These metrics included:

- Impressions, a record of the number of times an advertisement was displayed.
- Clickthrough rate (CTR) which was the number of clicks an advertisement received divided by the number of times the advertisement was displayed (impressions).
- Conversion was the completion of a unique goal or action on a website e.g. sending an email, buying a product, signing up for a service, downloading a product.
- Conversion rate was the number of conversions divided by the number of clicks.

Unique tracking Uniform Resource Locators (URLs) could be assigned to each advertisement or keyword to track visitors who clicked through to a website. This could indicate the advertisements and keywords that converted most clicks to sales (Burns, 2005).

2.2.3. Long Tail

Long tail was a phrase that became popular following the publication of a best-selling book by Anderson (2006). The long tail was a concept whereby demand for products shifted from popular products to niche products (Skiera et al., 2010). The idea of the long tail was adopted by online advertising industry and became a concept "*used to describe the hundreds to thousands of keywords and key phrases that a website is found for, yet rarely noticed or exploited by owners of the website*" (Bailey, n.d). Figure 2.2 illustrates the concept.

Long tail keywords

Long tail keywords were low-volume, obscure, infrequently searched-for keywords. When used as part of a PPC campaign they (Mitchell, 2009):

- Could provide significant search volume. While the number of long tail searches were individually insignificant compared to generic searches, together they could provide significant search volume.
- Had less competition than generic keywords and therefore were cheaper.
- Were more specific than generic keywords. Therefore advertisements and landing pages could be better customised to satisfy customer's needs.
- Could produce higher conversion rates as it was thought that Web users making long tail searches were likely to be further along the buying cycle and therefore more ready to buy than Web users who made generic searches.
- Could be more profitable then generic keywords as they were cheaper and more likely to produce a conversion.

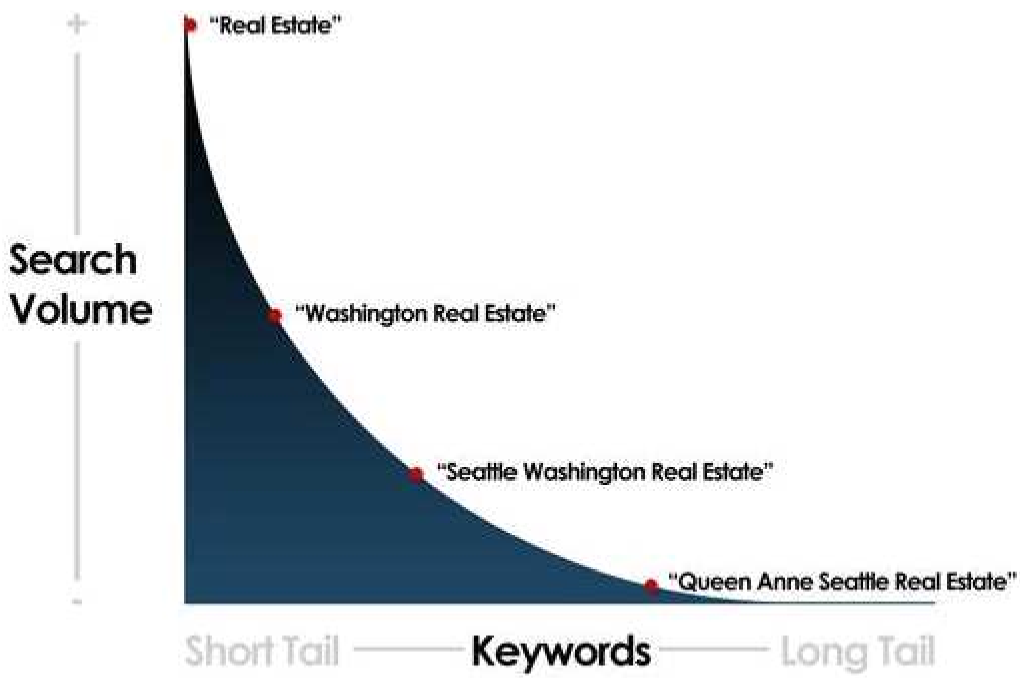


Figure 2.2: The short vs. long tail (Hoggart, n.d.).

Because long tail keywords tended to be specific they were usually longer than generic keywords. Since long tail keywords had higher conversion rates, it was assumed that longer keywords were more likely to convert than shorter keywords.

Online marketing agencies and experts believed that long tail keywords could boost conversion rates especially when keyword relevancy in the website content was good and elements of good landing page design were present (Mitchell, 2009, Search Engine Partner, n.d.).

However, Ghose and Yang (2008a), found that the "*length of a keyword negatively impacts the performance on all three metrics [conversion rate, order value and profit] for natural search listings but only affects the order value in paid search.*" Skiera et al (2010) found that "*the top 100 keywords - and not the very long tail composed of other keywords - generate the majority of searches, clicks and conversions*".

It was not clear whether the websites used in the research carried by Ghose and Yang (2008a) and Skiera (2010), were well designed and whether the content was relevant to long tail keywords. These could have affected their results.

2.2.4. Discussion

This Section described the concept of online advertising and the different advertising formats that it supported. The most popular format was Pay Per Click advertising.

The biggest advantage of PPC advertising was perhaps that it was relatively easy to set up campaigns and that it could generate traffic instantly. The biggest challenge was managing and optimising these campaigns so as to target customers, who were ready to buy, subscribe or complete an action that was profitable to advertisers. Landing pages played a crucial role in the success of PPC campaigns as they had to entice visitors to convert and become customers. Good design and optimisation was important in creating effective landing pages. Landing pages are discussed in more detail in Section 2.6.

Keyword selection was also important to ensure Return On Investment (ROI). Keywords had to be targeted and relevant. Part of the challenge in keyword selection was the identification of long tail keywords. Long tail keywords were thought to have higher conversion rates while being cheaper than generic short tail keywords. Although they were defined as being specific which, suggested longer keywords, there was not a number of words (keyword length) that characterised a long tail keyword.

Chapter 5 describes how PPC advertising was used to drive traffic to websites that were created in this research.

2.3. Websites

The Internet was an important marketing tool and a profitable selling channel. It was estimated that the Internet population would grow from 1.83 billion in 2010 to 2.10 billion in 2012 (Clickz, 2010).This represented a considerable rise in the number of potential customers for any business that used the Internet as a marketing or selling channel.

In order for businesses to take advantage of the growth in online advertising and population, they needed to design websites that could meet marketing objectives that included "*extended visit durations, repeat visits, positive consumer attitudes*" (Stanaland and Tan, 2010) and perhaps most importantly converting visitors into customers.

A website was "*an information resource on the World Wide Web (WWW) [that was] defined as a group of interface and functional attributes that are connected to each other to serve high levels of usability, performance, and beauty to users, to satisfy* *users' wants, and to obtain their satisfaction in a competitive market of online and offline sales and information-services*" (Lee and Koubek, 2010).

Lee and Koubek (2010) categorised websites into four types:

- Entertainment websites. These "*provided diversion and relaxation to users who wanted to escape from the stressful reality.*"
- Information websites. These "*made it possible for users to obtain useful information more quickly and more easily.*"
- Communication websites. These "*facilitated communicating with others with similar interests.*"
- Commerce websites. These "*provided an online market place where goods and services are purchased*."

2.3.1. Website design

"*In a practical sense, web[site] design is critical in building customer relationships, facilitating customer support, and converting visitors into customers in the online environment*" (Hausman and Siekpe, 2009). According to Lee and Koubek (2010), "*what constitutes a good website has been traditionally explained by relating it to user and usability. In other words, a successful and preferable website generally refers to one with high usability, which is user-friendly and user-centered in interface and functional aspects.*"

Studies by Levene (2006) and Widyantoro and Yen (2001) suggested that website visitors evaluated the relevance and usefulness of a website and formed an overall impression of the website in a short period of time. When a visitor landed on a website, "*a rapid and almost unconscious but complex thought process [was] activated*" that resulted in the visitor's first impression of the website and affected their subsequent decision regarding the website (Kim and Fesenmaier, 2008). Lindgaard et al (2006) stated that "*visual appeal [could] be assessed within 50ms, suggesting that web designers [had] about 50 ms to make a good first impression*."

First impressions created a long-lasting effect also called "*halo effect*", which led visitors to interpret information in such a way as to confirm their preconceptions (Lindgaard et al., 2006).

As a result Web users were consistent with their initial decision about a website. If the initial impression was good, they downplayed or ignored the negative aspects of a website. If the first impression was bad, most visitors left a website immediately. However, if they stayed on, they were less tolerant towards the negative aspects of the website and less impressed by the positive aspects (Kim and Fesenmaier, 2008).

2.3.2. Website design elements

There were a number of elements that made up the interface of a website and that affected visitors first and subsequent impression of a website The main elements included navigation, text, links, graphics and layout. Studies regarding website design focused on two levels of granularity, specific features, and categories that referred to a group of similar features (Zhang and von Dran, 2000). Some of these categories were:

a) Hygiene (dissatisfiers) and motivator (satisfiers) (Zhang and von Dran, 2000).
b) Human and computer (Hausman and Siekpe, 2009).
c) Information design, visual design and navigation design (Cyr, 2000).
d) Hygiene and potential (Kim and Fesenmaier, 2008).
e) Information, navigation, graphic and experience design (Ivory and Hearst, 2002).
f) Navigation.

a) Hygiene/Motivator factors

Hygiene factors were those factors that made a website functional and serviceable, and whose absence caused visitor dissatisfaction (hence dissatisfiers).

Motivator factors were those factors that added value to a website by contributing to user satisfaction (thus satisfiers).

(Zhang and von Dran, 2000)

b) Human/Computer factors

Hausman and Siekpe's (2009) computer factors and human factors were based on Zhang and von Dran's hygiene and motivator factors.

Human factors were website design elements that contributed to visitors' satisfaction, for example, (Hausman and Siekpe, 2009):

- Feedback features.
- Language options.
- Links to similar websites.
- Humour.
- Gift services.

Computer factors were those elements that provided functionality (Liang and Lai, 2002). According to Hausman and Siekpe (2009) computer factors included technical aspects, navigation, impartiality and information content for examples:

- Company logo.
- Clear displays of page contents.
- Indication of security/secure site.
- Presence of clear menu items on pages.

- Presence of shopping cart.
- Assurance of privacy.
- Company logo.
- Product images as thumbnails.

Hausman and Siekpe (2009) suggested that "*providing richer media with a more real environment (providing improved human factors) [had] a more positive influence on user involvement with the content over improved computer factors*." The importance of human factors was further supported by Fogg (2002) who found that about half of all consumers paid more attention to aspects linked to human factors than to the content of a website.

c) Information, visual and navigation design

Cyr (2000) described information, visual and navigation design as:

Information design referred to website elements that conveyed information (accurate or inaccurate) about products or services to visitors, for example text.

Visual design referred to website elements pertaining to the balance, emotional appeal, aesthetics, and uniformity of a website's graphical look, for example colours, photographs, shapes and fonts.

Navigation design referred to the navigational scheme used to enable visitors to access different sections of a website, for example menu, links and search facility.

d) Hygiene and potential factors.

Kim and Fesenmaier (2008) extended the model proposed by Zhang and von Dran (2000). They identified 2 types of website design categories:

- Hygiene factors which included elements pertaining to informativeness and usability.
- Potential factors which included elements pertaining credibility, inspiration, involvement, reciprocity.

Kim and Fesenmaier (2008) described these factors as:

Informativeness was the degree to which a website's content was accurate, relevant, useful, current and complete.

Usability was defined as the level of user-friendliness of websites that enabled visitors to easily navigate them with no (or a minimum level of) mental effort. Websites had to be designed so that visitors could easily understand who sponsored a website, the goals of a website and what they could achieve on it. Ease of navigation and perceived ease of use were important factors of usability. Perceived ease of use was considered an antecedent of positive behavioural intention, intention to purchase and satisfaction (Heshan and Zhang, 2006, Venkatesh and Morris, 2000).

Credibility could enhance visitors' perception of a website. Wang et al (2004) found that visitors inferred website credibility from credibility cues contained within a website. These cues were in the form of design elements such as awards from neutral sources, privacy and security components, the identity of the website operator, seals of approval, and references (Loveday and Neihaus, 2008, Fogg, 1999, Fogg et al., 2002, Fogg et al., 2001, Yang et al., 2003)

Inspiration could be evoked by truth, goodness, beauty, or superiority (Averill, 1975, Thrash and Elliot, 2003). Visual and auditory design elements could be used to create positive experiences that inspired visitors.

Involvement was the motivational force that determined behavioural outcomes. Interactivity promoted "*user engagement with content*" (Sundar and Kim, 2005). Early research had shown that visitors responded positively to interactive websites (Chung and Zhao, 2004, Jee and Lee, 2002, Stromer-Galley, 2004). Kim and Fesenmaier (2006) found that playful and enjoyable websites kept visitors entertained, encouraged them to browse and increased their depth of exploration. Interactivity positively affected visitors' attitude towards a website's brand as well as their intention to purchase (Ariely, 2000, Sundar and Kim, 2005). Interactivity also increased the amount of information processed by visitors and increased product and website likability (Sicilia et al., 2005). Wu (2006) found that control over website elements such as navigation, pace of interaction and content contributed to a website's perceived interactivity.

Reciprocity was defined as "*mutually gratifying patterns of exchanging goods and services*" (Kim and Fesenmaier, 2008). Kim and Fesenmaier (2008) suggested that a website that promoted reciprocity could have a favourable impression on visitors.

e) Information, navigation, graphic and experience design.

Information, navigation, graphic and experience design were identified by Ivory and Hearst (2002). Ivory and Megraw (2005) refined these factors into a conceptual model of website interface shown in Figure 2.3.

The model in Figure 2.3 shows that text, link, and graphic elements were the basic elements on which website interfaces were built. The next level up considered the formatting of basic elements while the subsequent level addressed the formatting/layout of a web page.

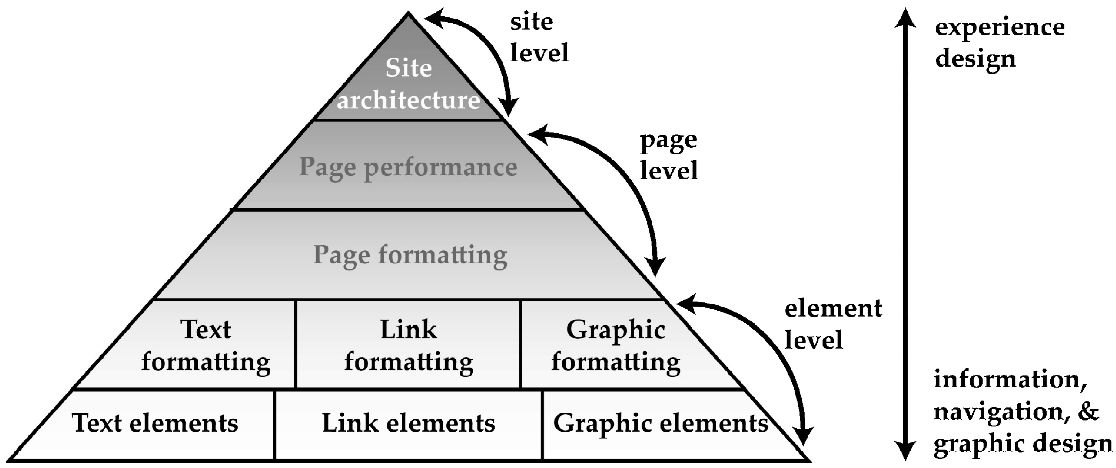


Figure 2.3: Conceptual model of website interface (Ivory and Megraw, 2005).

The top two levels considered page performance and site architecture including consistency, breadth, and depth of pages within websites. The top two levels were associated with design activities involved in creating an experience for visitors while the bottom three levels were associated with information, navigation, and graphic design activities. All levels influenced visitors' experience with a website (Ivory and Megraw, 2005).

f) Navigation.

Navigation structure acted as the roadmap of a website. According to Krug (2006), the purposes of navigation were "to help us find whatever it is we are looking for and tell us where we are" and also to tell visitors about the website's content by making the hierarchy visible. Ease and speed with which navigation allowed visitors to find what they were looking for were important factors that affected the usability of a website and the overall browsing experience of visitors. Disorientation was a frequently stated reason for negative experience with a website. Disorientation could lead to frustration which caused visitors to leave a website in search of a better website and left visitors with a negative impression which eliminated the possibility of recurring visits (Webster and Abuja, 2006).

2.3.3. Navigation Design

Loveday and Neihaus (2008) and Krug (2006) proposed the following guidelines for designing efficient navigation:

- Having a persistent navigation or global navigation which displayed a consistent navigation menu on all pages.
- Organising the content hierarchy into sections, sometimes called primary navigation.
- Having a search box. Some people prefer searching to browsing.
- Using breadcrumbs. Breadcrumbs were a series of links that showed the path that a visitor took while browsing a website. This made it easy for a visitor to navigate back to previous pages. Breadcrumbs were different to 'You are here' indicators which showed visitors where they were in the hierarchy of a website.

Websites with large amounts of content had deep hierarchies which could increase the time a visitor spent searching for a page. One of the issues with categorising large amounts of content into sectors was that sometimes it was not obvious to visitors which sector a particular piece of information was found in. This could be frustrating for visitors.

Design guidelines were sometimes not enough. There was a need for continuous optimisation of the navigation and linking structure over time as more visitors used a website, more content was added and also to accommodate the ever changing and evolving goals and search strategy of visitors.

2.3.4. Link Structure Optimisation

Designing a navigation structure was not straight-forward as designers did not always understand the browsing and behavioural pattern of visitors at the early stages of the design process. In most cases, the initial navigation structure had to be optimised to accommodate website changes as well as the behaviour and goals of visitors. Hollink et al. (2007) suggested two ways of improving link efficiency by optimising:

- Visual design of links, for example colour and placement on the page.
- Structure of links, that is the way pages were connected.

Methods existed that enabled the reduction of the time it took visitors to reach their goal page. These included:

a) Organising navigation based on page popularity.
b) Using recommender systems to predict pages that visitors were interested in.

a) Page Popularity

The model developed by Smyth and Cotter (2003) promoted menu items that visitors chose frequently to a higher position in the menu and reduced navigation time by almost 50% leading to increase in usage in excess of 30%.

Yen (2007) proposed an "*accessibility-popularity (A-P)*" model which modified the structure of a website by identifying popular pages and increasing their accessibility. The popularity of a page was measured as the number of times a page was viewed by visitors or the length of time visitors spent on the page. The A-P model followed two strategies:

- Push – where a website reacted to demand, that is the pages that were popular with visitors were made more accessible.

- Pull – where pages that a website wanted to promote were made more accessible.

b) Recommender Systems

"*Web page recommender systems predict the information needs of users and provide them with recommendations to facilitate their navigation. Given a user's current actions, the goal is to determine which Web pages will be accessed next. Many web sites on the Internet use web page recommender systems to increase their usability and user satisfaction*" (Göksedef and Gündüz-Ögüdücü, 2010).

Traditional recommender systems used web usage and web content mining techniques to make predictions. Better recommender systems in the form of hybrid recommender systems combine two or more prediction methods (Göksedef and Gündüz-Ögüdücü, 2010).

"*One of the most successful and widely used technologies for building recommendation systems is collaborative filtering (CF). A collaborative filtering system collects visitor opinions on a set of objects to form peer groups, using ratings which are provided by the users or are implicitly computed, then learns from these peer groups to predict a particular user's interest in an item. It is often based on matching, in real-time, the current user's profile against similar records (nearest neighbors) obtained by the system over time from other users*" (Demir et al., 2007).

2.3.5. Discussion

Websites were usually faced with the following questions:

- What elements were most important to its success?
- What elements influenced visitors' response?

This Section provided an overview of the main design elements that influenced visitors' response and contributed to good website design. However, because of the different types of websites that existed it was not clear how design elements contributed to the success of each website type. It was possible that depending on the purpose of a website, the importance of design elements varied.

The design elements identified in this Section were taken into consideration when implementing the front-end of the dynamic main website described in Chapter 4.

2.4. Behavioural Intention

The design of online advertisements could influence the way visitors responded to a website (Robinson et al., 2007). Similarly, visitors' perceptions and behaviours were governed by their motives for visiting a website (Rodgers et al., 2007).

Visitors' motives could affect the way they behaved on a website. Sanchez-Franco and Roldan (2005) identified two types of motives namely intrinsic and extrinsic. They described intrinsic motives as "*emphasising internal rewards such as pleasure and satisfaction from performing the behaviour*" and extrinsic motives as "*focusing on external rewards including, for instance, incentives and gratifications*". Sanchez-Franco and Roldan (2005) also suggested that "*research in the HCI (human-computer interaction) tradition has long asserted that the research of human factors is a key to the successful design and implementation of technological devices, and should include extrinsic and intrinsic motives*." Research in this area looked at applying the Technology Acceptance Model (TAM) to the online world in order to find out what factors affected website acceptance and usage.

Several models were found in literature regarding acceptance of technology. They included the Technology Acceptance Model (TAM) (Davis, 1986), Technology Acceptance Model 2 (TAM2) (Venkatesh & Davis, 2000), Innovation Diffusion Theory (Moore & Benbasat, 1991) and the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al, 2003).

Figure 2.4 shows the Technology Acceptance model presented by (Davis, 1986).

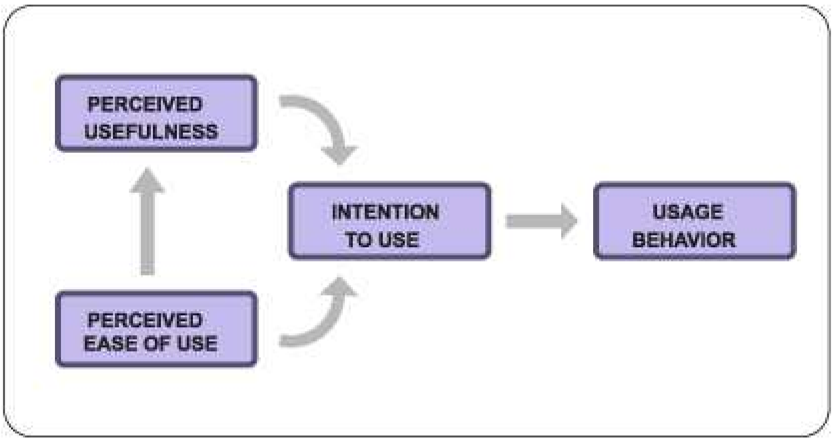


Figure 2.4: Technology Acceptance Model (TAM) (Davis, 1986).

The TAM suggested that acceptance of technology depended on:

- Perceived ease-of-use [PEOU].
- Perceived usefulness [PU].

Heshan and Zhang (2006) investigated the relationship between Perceived Enjoyment (PE) and Perceived Ease of Use (PEOU) and their effect on Behavioural Intention (BI) when applied to online tasks. They found that the "*PE→PEOU direction has an overall dominance over the PEOU→ PE direction in utilitarian system environments. PE does not have a direct impact on BI; instead, Perceived Usefulness (PU) and PEOU fully mediate its impacts.*" Their conclusions suggested that website design elements that contributed to PU and PEOU could influence visitor behaviour. Chen et al. (2002a) demonstrated that attitude toward websites was positively influenced by PU. Chen et al. (2002b) have also shown that Perceived Informativeness (PI) can influence attitude towards website in a positive way. None of these studies identified the design elements that contributed to PU, PEOU or PI explicitly.

Behavioural intention was also thought to be influenced by online flow. Sanchez-Franco and Roldan (2005) defined flow as "*a positive, highly-enjoyable state of consciousness that occurs when our perceived skills match the perceived challenges we are undertaking. When this occurs, an individual derives intrinsic enjoyment from the activity and tends to continue with it.*" They suggested that the level of perceived enjoyment of an activity could be a measure of flow.

Hoffman and Novak (1996) suggested that online flow was a cognitive state experienced during navigation which could lead to more browsing, and ultimately purchase in an online shopping environment. According to their model, the factors that determined online flow were:

- high levels of skills and control.
- high levels of challenge and arousal.
- focused attention.
- interactivity and telepresence.

Hausman and Siekpe (2009) and Smith and Sivakumar (2004) suggested that flow facilitated online behaviours such as browsing, shopping, and repeat purchases. Novak at al. (2000) found that 47% of visitors experienced flow on the Internet at some point. However, little was known about how flow worked on websites and the design elements that promoted it (Hausman and Siekpe, 2009).

2.4.1. Types of website visitors

Visitors arrived at websites with different motives and goals in mind. These could affect behavioural intention. There were two generally recognised types of online visitors or visitor modes (Hoffman and Novak, 1996, Stanaland and Tan, 2010):

a) Goal-oriented or Seekers.

b) Experiential or Surfers.

a) Goal-oriented or Seekers

Goal-oriented consumers were motivated by external factors (extrinsic motivation), task oriented, and influenced by interests or concerns brought about by the particular situation or context that they were in. They used directed searches, characterised by work-like thoughts, to reach their goals (Novak et al., 2003).

According to Sanchez-Franco and Roldan (2005) "*using the web for its informational value and purchase utility – such as directly searching for information to complete a task or to reduce purchase uncertainty – are goal-directed behaviours.*" Goal-oriented visitors were usually constrained by time. They wanted to find what they were looking for as quickly as possible and access content and services according to "*instrumental decision-criteria*" and as a result perceived usefulness was "*likely to be weighed more strongly by goal-directed users*" (Sanchez-Franco and Roldan, 2005).

b) Experiential or Surfers

Experiential visitors were motivated by fun which led to browsing. Their searches were non-directed and they usually did not have an explicit goal in mind when carrying out searches. They also spent a lot of time browsing websites usually in an ad-hoc fashion.

"Relatively unstructured recreational use [of the web or websites] is experiential behaviour" Sanchez-Franco and Roldan (2005).

Furthermore, experiential visitors were thought to be less experienced while goal-oriented visitors were thought to be more experienced. As a result, ease of use was more important to experiential visitors compared to goal-oriented visitors. Since goal-oriented visitors were more skilled at using the Internet, they were less concerned by ease of use and considered usefulness as being more desirable. Perceived ease of use contributed to perceived usefulness and in the case of goal-oriented visitors the influence of perceived ease of use was moderated by perceived usefulness so that the former's influence on attitude was indirect in their case.

(Sanchez-Franco and Roldan, 2005)

The differences between goal-oriented and experiential visitors that were identified by Novak (2003) are summarised in Table 2.1.

2.4.2. Discussion

In order to design websites that catered to visitors' needs it was important to understand their motives. Visitor motives were determined by the goals that they had in mind when they browsed the Internet. Website visitors usually belonged in one of two categories: experiential visitors or goal-directed visitors. Each category was associated with characteristics that determined what visitors might expect from a website and how they might use it. It was therefore important for a website's success to identify who its visitors were and to adjust its design according to the preferences of those visitors. The literature did not explicitly identify website design elements that were important in providing an optimum experience to either type of visitors. Some landing pages created by this research (described in Chapter 6) were optimised for goal-oriented visitors.

Goal-oriented	Experiential
Extrinsic motivation	Intrinsic motivation
Instrumental orientation	Ritualised orientation
Situational involvement	Enduring involvement
Utilitarian benefits/value	Hedonic benefits/value
Directed (pre-purchase) search	Non-directed (ongoing) search; browsing
Goal-directed choice	Navigational choice
Cognitive	Affective
Work	Fun
Planned purchases; repurchasing	Compulsive shopping; impulse buys

Table 2.1: Comparison between goal-oriented and experiential visitors (Novak et al., 2003).

2.5. Websites Performance

There were a number of metrics that could be used to evaluate the performance of a website. Some of these metrics were implicit and could not be measured directly while others were explicit and easier to measure.

2.5.1. Implicit performance metrics

Chiou et al. (2010) identified that websites could be evaluated based on the following:

a) Website usability and design.

b) Content.

c) Quality.

d) User acceptance.

e) User satisfaction.

a) Website usability and design

"Website usability is a critical metric for assessing the quality of a firm's web presence. A measure of usability must not only provide a global rating for a specific website, ideally it can also illuminate specific strengths and weaknesses associated with site design" (Agarwal and Venkatesh, 2002).

b) Content

High quality content could "*effectively facilitate interpretation and understanding*" (Lin, 2010). Website content was different from design. "*The content component addresses the issue of what is included in the site and identifies the various types of information. The design component addresses presentation and navigational features*" (Robbins and Stylianou, 2003). Content could be evaluated in terms of content accuracy, content currency and content completeness (Nelson et al., 2005, Lin, 2010).

c) Quality

Website quality comprised of the following dimensions: interactivity, online completeness, ease of use, and entertainment (Kim and Niehm, 2009). Ahn et al. (2007) categorised website quality into system, information, and service quality. Ahn et al. (2007) also found that website quality had a "*significant impact on perceived ease of use, playfulness, and usefulness, and consequently, that it encouraged website use in the context of online retailing.*"

d) User acceptance

The Technology Acceptance Model (TAM) was originally developed by Davis (1986). Davis (1986) suggested that acceptance of technology depended on perceived ease-of-use and perceived usefulness. The TAM was applied to the online technology to predict the acceptance and use of the Internet and websites. Chung and Tan (2004) suggested that in order to access the effectiveness of a website it was important to understand user acceptance. For certain websites, the number of repeat visits could be an indication of user acceptance (Ahn et al., 2007).

e) User satisfaction

It was thought that effective website design, including navigation capability or visual appeal of a website, could potentially result in online trust (Gefen and Straub, 2003, Koufaris, 2002, Cyr, 2000) or satisfaction (Agarwal and Venkatesh, 2002, Anderson and Srinivasan, 2003). Cyr (2000) found that all causal relationships between design (information design, visual design, navigation design) and satisfaction were significant and that these design elements had the ability to elicit satisfaction in website visitors. An increase in satisfaction could lead to an increase in customer retention and profit. According to Reichheld and Schefter (2000), an increase in customer retention rates by only 5 percent could increase profits by 25% to 95%.

2.5.2. Explicit performance metrics

In order to stay competitive in the online arena, companies needed to understand and respond to their website visitors' needs. They could achieve this by analysing visitors' online activities and incorporating the results in their decision-making processes and strategy (Park et al., 2010). This gave birth to the concept of Web Analytics. The Web Analytics Association defined Web Analytics as "*the measurement, collection, analysis and reporting of Internet data for the purposes of understanding and optimising Web usage.*"

Web Analytics software was a key element of a company's Web Analytics strategy. A popular Web Analytics software was Google Analytics. Google Analytics was "*an enterprise-class web analytics solution*" that provided insights into a website's traffic and marketing effectiveness. It had powerful, flexible and easy-to-use features. It was also a free service (Google, n.d.-e).

Web Analytics software provided a number of metrics that enabled companies to measure the performance of their websites. Some of the metrics were (Burby et al., 2007) :

- Page views – "*The number of times a page (an analyst-definable unit of content) was viewed.*"
- Unique visitors – "*The number of inferred individual people (filtered for spiders and robots), within a designated reporting timeframe, with activity consisting of one or more visits to a site. Each individual is counted only once in the unique visitor measure for the reporting period*."
- New visitor – "*The number of unique visitors with activity including a first-ever visit to a site during a reporting period.*"
- Repeat visitor – "*The number of unique visitors with activity consisting of two or more visits to a site during a reporting period*."
- Visit duration – "*The length of time in a session. Calculation is typically the timestamp of the last activity in the session minus the timestamp of the first activity of the session*."
- Click-through – "*Number of times a link was clicked by a visitor*."
- Click-through rate/ratio – "*The number of click-throughs for a specific link divided by the number of times that link was viewed*."
- Page views per visit – "*The number of page views in a reporting period divided by number of visits in the same reporting period.*"

- Bounces – "*Visits that consist of one page-view*". A bounce occurred when a visitor left immediately without having viewed any page but the landing page (White, 2006).
- Bounce rate – This was calculated by dividing the number of bounces that a page generated by the number of visits that it had received.
- Conversion – "*A visitor completing a target action*."

2.5.3. Discussion

Small changes to a website's layout or content could have a large effect on visitors' response and the website's efficiency. This Section identified metrics that could be used to measure website performance. Website performance metrics have evolved over the years from the number of hits a website received to usability, customer satisfaction, conversion rate and bounce rate. Some of the metrics found were implicit and therefore more difficult to measure for example usability and customer satisfaction. The process of evaluating implicit metrics involved surveys and other interactive methods. Other metrics found were explicit and were relatively easier to measure. These metrics were readily available as part of sophisticated reports that were generated by Web Analytics software. Some of the explicit performance measures described in this Section were used to measure the performance of landing pages during landing page optimisation described in Chapter 6.

2.6. Landing Pages

Loveday and Neihaus (2008) described the goal of a landing page as "*moving the visitor to the primary desired action*." They stated that the selection of the action was driven by an online strategy for example if the business goal was to generate leads then the goal of the landing page was to motivate visitors to make contact.

The role of landing pages were different to other pages on a website and as such their design guidelines emphasised (Loveday and Neihaus, 2008):

- Trust and credibility - Since landing pages were usually the first point of contact with a company, they needed to make a good first impression.
- Professional and industry appropriate design – The appearance of a landing page impacted on visitors' first impression.
- References – References such as customer testimonials, industry awards and press quotes could be included on a landing page.
- Reducing or eliminating navigation – Since visitors were looking for something specific, navigation options could be minimised so as not to lead them away from a conversion path.
- Extension of an advertisement – Landing pages needed to be an extension of the advertisement that they serviced. This provided a consistent experience to the visitor.
- Providing what the corresponding advertisement promised – The landing page needed to do this, otherwise visitors would leave without converting.
- Segmentation – By providing segmentation options, landing pages were able to cater for the needs of a broader audience.
- Personalisation – "*Personalising the look and wording of a landing to a particular visitor or audience is a powerful way to capture and keep attention*."
- Consideration for reading patterns – Different visitors had different reading patterns. Therefore, the content needed to be formatted to appeal to all visitors.

2.6.1. Segmentation and personalisation

"*The goal of web personalisation is to deliver the right content to the right person at the right time and to maximise immediate and future business opportunities*" (Tam and Ho, 2006). Web personalisation is akin to the mirroring principle where a product should be built around the "*unique and particular needs*" of customers (Hammer, 1995).

In the online world, it was difficult to predict what group a website visitor belonged to. It was therefore difficult to cater to the specific needs of these segments. One way to solve this problem was to offer segmentation options which allowed visitors to qualify themselves. A web page with segmentation options offered "*a clear, audience-specific path for each of the most important audiences*" (Loveday and Neihaus, 2008). Further segmentation could be carried out on subsequent pages to pin down the visitor's interest. Providing segmentation options on a landing page meant that one page could capture the attention of a wide variety of visitors and offer information relevant to each group.

Segmentation could also take place at the advertisement level which meant that, the landing page could cater to a tightly defined set of visitors. In this case, the landing page could be personalised for these visitors. However, it was not always possible to run advertisement campaigns which were that well targeted and therefore, the level at which personalisation occurred within a website could vary.

Types of personalisation

There were different types of personalisation, ranging from "*user-driven personalisation to transaction and context-driven personalisation strategies*" (Tam and Ho, 2006). Tam and Ho (2006) presented the following definitions:

In user driven personalisation, visitors specified in advance the layout and content that matched their interests and preferences.

In transaction driven personalisation, a website was customised based on the previous transactions of visitors. The visitor's interests and preferences were inferred from their browsing history.

Context driven personalisation relied on understanding the context of individual visitors through click stream analysis or web mining. By identifying the context of visitors, their objectives could be determined and a website could be personalised to help them meet their goals.

User driven and transaction driven personalisation were not useful for landing pages as these pages attracted mostly new visitors for whom there was no historical data available. However, a landing page could offer visitors some options for personalising the website or page so that the next time they visited, the content and layout was tailored their preferences. Context driven personalisation could be attempted on landing pages. If an advertising campaign was well targeted, then the context of visitors who landed on a website through that campaign could be assumed.

Calculating how much of the content to customise on a landing page could be complex. Too much customisation could distract from the main aim of a landing page, for example offering too many customised links or options could take a visitor away from the optimal conversion path. It could be argued that landing page customisation had a positive impact as long as it re-enforced the primary aim of the landing page.

2.6.2. Trust and credibility

Some of the design elements that impacted trust according to Loveday and Neihaus (2008) were:

- Visual design - A landing page had to have a professional and industry-appropriate look.
- Quality and relevancy of content - A landing page had to provide well written, concise and accurate information.
- References - A landing page could include positive references from third parties.
- Security - It was important to make it obvious to visitors that a website was secure and that any credit card or personal details provided would be used in compliance with credit card company requirements and government regulation concerning storage of personal data.

2.6.3. Content relevancy/structure

The content of a landing page played an important role in its efficiency as it needed to catch visitors' attention and persuade them to explore a website. Information was also considered an important factor in influencing trust and satisfaction (Cyr, 2000). Small changes and adjustment in the content could have significant effects on visitors' behaviour. One of the few well documented facts about the Internet is that people do not read web pages, instead they scan them. Krug (2006) states that "*we tend to focus on words and phrases that seem to match (a) the task at hand or (b) our current or ongoing personal interests and (c) the trigger words that are hardwired into our nervous system like 'Free' and 'Sale'*". In his research Neilsen (1997) found that 79% of visitors always scanned new pages that they came across while only 16% read every word. In order to accommodate different reading patterns, Neilsen (1997) suggested that websites should use scannable text by:

- Highlighting keywords.
- Having meaningful sub-heading, not “clever” ones.
- Using bullet lists.
- Having one idea per paragraph.
- Using the inverted pyramid style.

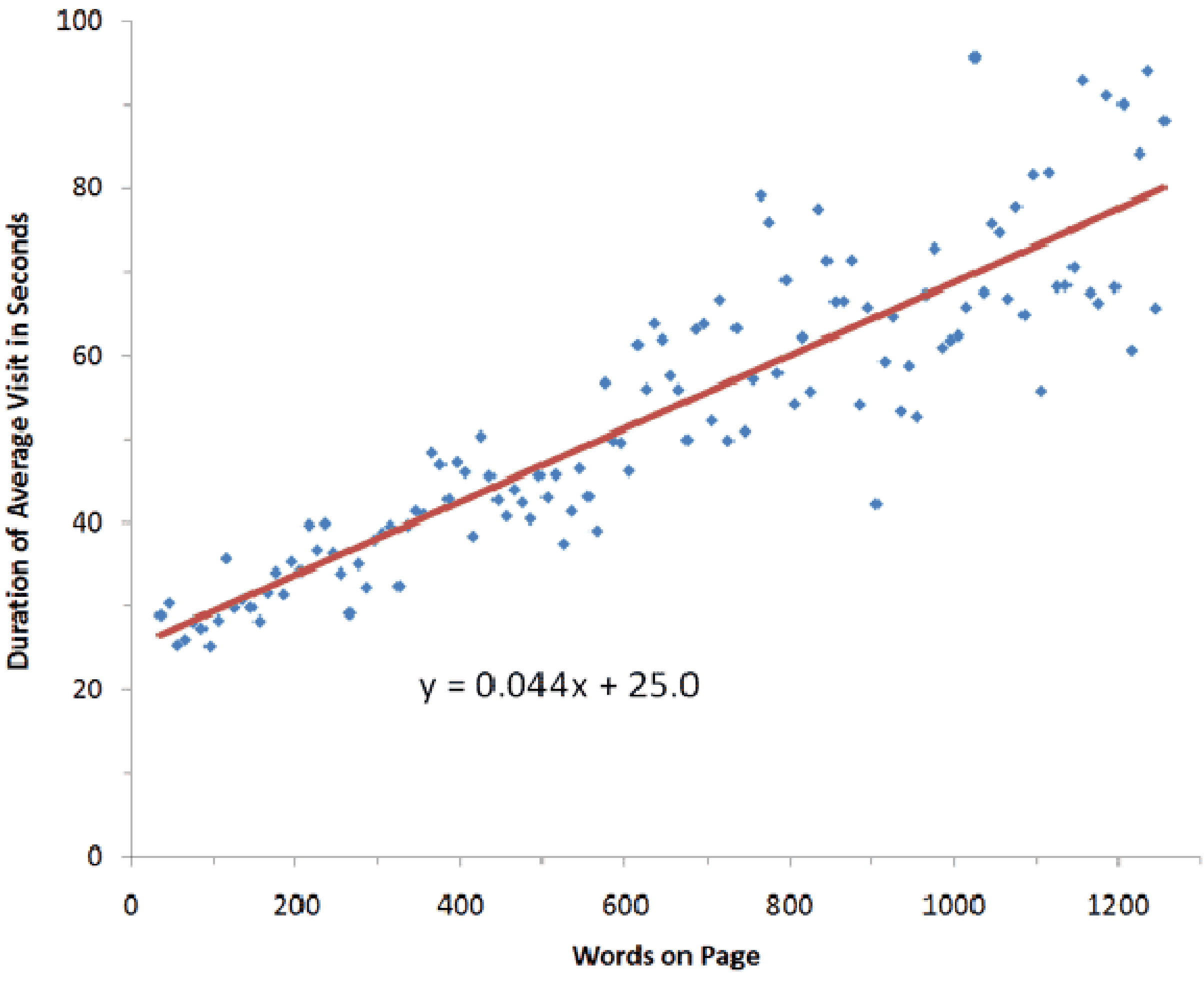
When measuring usability of a web page that followed his content formatting guidelines, Neilsen (1997) observed a 124% increase in usability.

Apart from content formatting and structure, there were other factors to consider when writing content such as (Loveday and Neihaus, 2008):

- Tone and language – This needed to match the language of the target audience.
- Benefits – The content needed to engage visitors with benefits and scenarios that they could relate to.
- The amount of copy – Depending on the complexity of the offer being made on a landing page the copy could vary in length. When there was too much copy a rule of thumb was to have the essential points above the fold. This allowed visitors who did not like to scroll to still be able to scan the offer.

Nielsen (2008) investigated how much content visitors actually read and found that “*on the average web page, users have time to read at most 28% of the words during an average visit; 20% is more likely.”* He was able to model visitors’ reading behaviour for pages containing between “*30 and 1,250 words”* using a linear equation shown in Graph 2.1.

As expected visitors spent more time on pages with more content. However, the best-fit formula showed that visitors spent only "*4.4 seconds more for each additional 100 words.*" There was also a fixed time of 25 second during which visitors were thought to be familiarising themselves with layout and navigation or looking at images (Nielsen, 2008). These results suggested that content had to be kept short and to the point as visitors read little content.



Graph 2.1: Average time visitors spend on pages with different word counts (Nielsen, 2008).

2.6.4. Discussion

This Section considered some of the best design practices regarding landing pages. These were taken into consideration when designing and optimising the landing pages that are described in Chapter 6.

2.7. Landing Page Optimisation

Landing page optimisation (LPO) was a popular approach for assessing and then improving website design. During LPO several prototypes were created and then tested with website visitors. LPO could be either target-based (pages were customised based on behavioural or "*self-profiled*" information of a visitor, for example purchase history) or experiment-based (pages were optimised for visitors' preferences that were inferred though experimentation). Landing page optimisation could be limited in time or ongoing.

(Gofman and Moskowitz, 2009)

The goal of LPO was to increase the number of website visitors who completed an action that represented a conversion. A conversion could be sending an enquiry, buying a product or downloading software. The definition of a conversion varied from website to website.

2.7.1. Landing page optimisation through targeting

Three main types of targeting could be used for landing page optimisation. They were (SEO Moves, 2010) :

- Associative content targeting (also called passive targeting). The design of the landing page was based on information obtained about the visitor, for example location, search criteria and other criteria that can be used to categorise a visitor.
- Predictive content targeting (also called active targeting). The content of the landing page was adjusted based on historical information collected about the visitor, for example behaviour, browsing pattern or demographics etc.
- Consumer directed targeting (also called social targeting). "*The page content is created using the relevance of publicly available information through a mechanism based on reviews, ratings, tagging, referrals, etc.*"

2.7.2. Landing page optimisation experimentation types

There were two main types of experimentations that could be conducted for LPO (Gofman and Moskowitz, 2009):

- Closed-ended experimentation. Several variations of landing pages were tried and visitor behaviour was observed. At the end of the experiment the better performing page was selected.
- Open-ended experimentation. Similar to closed-ended experimentation except that the experiment was ongoing or ran for a prolonged period of time. The design of the landing page was changed dynamically as the results changed.

2.7.3. Experimentation methodologies for LPO

There were three methodologies that could be used in experimentation based LPO, to find the best combination of design elements for an optimal landing page (Ash, 2008, Kaushik, 2006):

a) A/B testing.

b) Multivariate landing page optimisation.

c) Experience testing.

a) A/B testing

A/B testing was also known as A/B split testing. It was a method for testing two versions of a web page by randomly assigning (usually in equal proportions) new visitors to each page. A/B testing could be carried out sequentially or in parallel.

In sequential testing, pages were tested one at a time for a given period of time. Once all pages were tested, then their performance was compared. In parallel testing, both versions of a landing page were tested at the same time with traffic divided (in equal proportions) between the two pages.

It was possible to have more than two versions of a landing page in a split test. For example, if there was one original and two alternative versions. However, split tests were rarely used to test more than ten pages.

A/B testing was comparatively cheap as it did not require additional tools or resources before it could be implemented. Results were easy to interpret and did not require statistical analysis. However, in the case of sequential testing, it was difficult to control external factors such as traffic, seasonal trends and economic trends.

b) Multivariate landing page optimisation

Multivariate landing page optimisation (MVLPO) was a method whereby multiple combinations of design elements of a page could be tested. For example, a page may have had X choices for its title, Y choices for its featured image, Z choices for its content. This would give X x Y x Z design and layout options for the page. By running MVLPO it was possible to identify elements that tended to produce an increase in conversion. MVLPO supported open-ended experimentation and took a scientific approach towards understanding visitor’s preference. One of the drawbacks of MVLPO was that it focused on optimising one page at a time. Since website experiences that resulted in conversion relied on multiple pages, optimisation of a conversion path could take a long time.

c) Experience testing

Experience testing provided a method for changing the entire website experience for a visitor. With experience testing, everything about the experience of a website could be changed, not just the look of one page, or the navigation or a piece of text. Experience testing made it possible to create two or three persistent experiences on a website, instead of creating two or three individual websites (Kaushik, 2006). The advantage of experience testing was that the results represented the visitors experience with the whole website and not just one page. Two main disadvantages of experience testing were:

- It needed a platform that supported experience testing.
- It took longer to obtain results than with A/B testing or MVLPO.

2.7.4. Discussion

This section described landing page optimisation; the types of targeting that could be used to achieve LPO as well as the experimentation methodologies for LPO. These were taken into consideration during LPO described in Chapter 6.

2.8. Online Search Behaviour

As discussed in 2.4, visitors' motivation had an impact on their behaviour on websites. One of the characteristics differentiating goal-oriented visitors from experiential visitors was the type of search that they carried out. Goal-oriented visitors carried out "*directed (pre-purchase)"* searches while experiential visitors carried out "*non-directed (ongoing)"* searches (Novak et al., 2003). Therefore, it was possible that visitor motivation and purchase intent could be inferred from the keyword that they used to search for information.

Usually, Web users started their search with a generic search query and gradually refined it until they found a search query that led them to what they were looking for. Web users could adopt two types of strategies when searching: browsing and analytical strategies (Marchionini, 1995). These strategies were described as follows (Zhang and von Dran, 2000):

"*Browsing is an informal and natural information seeking approach that depends heavily on the information environment and the user's recognition of relevant information. Analytical strategies, in contrast, depend on careful planning, recall of query terms, iterative query reformulation, and examination of results*".

Pavlou and Fygenson (2006) found that the intention of buying a product occurred before the intention of acquiring information on a product. Following this reasoning, a Web user who had decided to buy a product would go online and try to express their decision in term of a search query. Pirolli (2007) made a similar distinction between task and need. He referred to a query as an external representation of need.

Jansen et al. (2008) suggested that the search query was not the only expression of intent and that other "*aspects of the interaction including number of query reformulations, selection of vertical, use of system feedback, and result page viewed*" were also expressions of intent. Broder (2002) identified three types of searches: informational, navigational, and transactional which were described by Jansen et al (2008) as:

- Informational searching: "*The intent of informational searching is to locate content concerning a particular topic in order to address an information need of the searcher. The content can be in a variety of forms, including data, text, documents, and multimedia. The need can be along a spectrum from very precise to very vague*".
- Navigational searching: "*The intent of navigational searching is to locate a particular website. The website can be that of a person or organisation. It can be*

a particular web page, site or a hub site. The searcher may have a particular website in mind, or the searcher may just 'think' a particular website exists."

- Transactional searching: "*The intent of transactional searching is to locate a website with the goal to obtain some other product, which may require executing some web service on that website. Examples include purchase of a product, execution of an online application, or downloading multimedia*."

The characteristics of these search types were (Jansen et al., 2008) :

- Navigational searching
  - queries containing company/business/organization/people names.
  - queries containing domains suffixes.
  - queries with 'Web' as the source.
  - queries length (i.e. number of terms in query) less than 3.
  - searcher viewing the first search engine results page.
- Transactional searching
  - queries containing terms related to movies, songs, lyrics, recipes, images, humor, and pornography.
  - queries with 'obtaining' terms (e.g. lyrics, recipes, etc.).
  - queries with 'download' terms (e.g. download, software, etc.).
  - queries relating to image, audio, or video collections.
  - queries with 'audio', 'images', or 'video' as the source.
  - queries with 'entertainment' terms (pictures, games, etc.).
  - queries with 'interact' terms (e.g. buy, chat, etc.).
  - queries with movies, songs, lyrics, images, and multimedia or compression file extensions (jpeg, zip, etc.).
- Informational searching

- uses question words (i.e. ‘ways to’, ‘how to’, ‘what is’, etc.);
- queries with natural language terms;
- queries containing informational terms (e.g. list, playlist, etc.);
- queries that were beyond the first query submitted;
- queries where the searcher viewed multiple results pages;
- queries length (i.e., number of terms in a query) greater than 2; and
- queries that do not meet criteria for navigational or transactional.

2.8.1. Discussion

Jansen et al (2008) have suggested that “by identifying the user intent of Web queries in real time, Web search engines can provide more relevant results to searchers and more precisely targeted sponsored links”. Search engine behaviour studies could be applied to the design of websites and landing pages. By using visitors’ search query to infer intent, their experience on a website could be customised to better cater to their needs. For example, a transactional query could indicate that a visitor was ready to buy and therefore, should be directed to a website’s online shop. If a visitor landed on a website with an informational query, they could be directed to pages that contained information that matched their query.

Search query could be used in visitor profiling as a way of determining what they were looking for when they landed on a website. This could be a powerful way of customising websites in real-time and minimising the effort required by visitors’ to reach their goals. This could lead to more satisfying interaction and experience on websites through increased customer satisfaction, Perceived Ease of Use and Perceived Informativeness.

The findings described in this Section led to an investigation into whether a relationship existed between search query length, search query relevancy and the likelihood that a visitor would convert. Chapter 7 describes the findings of the investigation.

2.9. Data Mining and Knowledge Discovery

Data mining was *"the exploration and analysis of large quantities of data in order to discover meaningful patterns and rules. The data mining process is sometimes referred to as knowledge discovery or Knowledge Discovery in Databases KDD*" (Berry and Linoff, 2004).

The data mining process consisted of five iterative stages:

- Data selection.
- Data cleaning – duplicate data and inconsistencies were found and removed.
- Data transformation – data was transformed into attributes that were used as input to data mining algorithm.
- Data mining and model creation – data discovery algorithms were used to find patterns and create models.
- Assess models – models were tested with a set of data separate and different to the one used to create the model. Accuracy measures were used to determine the reliability of models.

It was possible to go back a step or more at each stage in order to correct errors or optimise the process.

2.9.1. Web usage mining

Web mining was described as the extraction of useful information from web documents and web services using data mining methods (Etzioni, 1996). There were three types of web mining: content mining, web structure mining, and web usage mining (Kosala and Blockeel, 2000).

Web usage mining extracted knowledge and identified patterns from data collected in web server access logs, proxy server logs, referrer logs, browser logs, error logs, visitor profiles, registration data, visitor sessions or transactions, cookies, visitor queries, and bookmark data, mouse clicks and scrolls and any other data generated by the interaction between users and the web or a website (Das and Turkoglu, 2009).

The goal of web usage mining was to capture, analyse and model the behavioural patterns and profiles of visitors who interacted with a website (Liu, 2007, Wang and Liu, 2003, Das and Turkoglu, 2009). Figure 2.5 shows the web usage mining process.

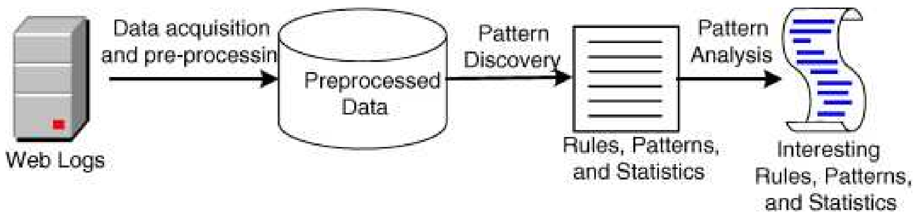


Figure 2.5: Web usage mining process(Das and Turkoglu, 2009).

Data mining algorithms that have been used for web usage mining included association rules, temporal sequences, clusters and path expressions (Cooley et al., 1997).

2.9.2. Predicting online behaviour

There has been growing interest in studying how people behave online and whether their behaviour can be predicted or influenced. Bucklin and Sismeiro (Bucklin and Sismeiro, 2008) reviewed the use of click stream data in understanding browsing and site usage, the efficacy of the Internet as an advertising medium and shopping behaviour on the Internet. Their review showed that research regarding online behaviour prediction had focused on shopping behaviour and purchase on e-commerce websites, not service websites.

Several approaches have been used to predict online purchase. Moe and Fader (2004a, 2004b) developed stochastic models of visitors' behaviour on ecommerce websites. They found a relationship between the frequency with which a visitor came to a retail website and their likelihood to purchase; the higher the frequency of visits the greater their propensity to buy. They used this relationship to predict purchase based on prior visit behaviour. A limitation of their method was that it did not consider browsing behaviour and the possibility that certain browsing actions could influence or be associated with online purchase

Montgomery et al (2004) modelled visitors' online browsing using path analysis and clickstream data. They proposed a multinomial probit model that could make probabilistic assessments about future paths including whether a visitor would make a purchase. Their model provided a dynamic forecast of a visitor's likelihood to purchase as they viewed pages on a website. This had some similarity to some of the early work described in this dissertation but the work in this dissertation did not use a multinomial probit model, and considered more attributes to predict behaviour.

Van den Poel and Buckinx (2005) investigated how different types of predictors affected purchasing at an online store and used logit modelling to predict whether a visitor would make a purchase during their next visit. They grouped predictors into four categories:

1. General clickstream behaviour.
2. More detailed clickstream behaviour.
3. Customer demographics.
4. Historical purchase behaviour.

Their results showed that variables from all four categories could be used to predict online purchase and that clickstream data (especially detailed clickstream behaviour) was important for predicting online purchase. They used previous browsing behaviour to predict future behaviour rather than real-time data.

2.9.3. Artificial Neural Networks

An Artificial Neural Network (ANN) was a network of interconnected elements which tried to mimic the way biological nervous systems worked (Picton, 2000). "*ANN can recognise patterns and learn from their interactions with the environment. The multilayer feed-forward network is widely used. ANN is adaptive and can handle complex systems. The architecture of ANN includes a number of nodes (neurons) or units organised in input and output layers as well as a number of hidden layers*" (Jafar et al., 2010).

"The two most significant properties of neural networks are their ability to learn and generalise" (Picton, 2000). ANNs learned by adjusting their weight and biases through an iterative process (Najjar et al., 1997).

There were two types of machine learning:

a) Unsupervised learning.

b) Supervised learning.

a) Unsupervised learning

In unsupervised learning an ANN was given a set of data as input but was not provided with corresponding target outputs (Ghahramani, 2004). Ghahramani (2004) suggested that it was "*possible to develop of formal framework for unsupervised learning based on the notion that the machine's goal is to build representations of the input that can be used for decision making, predicting future inputs, efficiently communicating the inputs* *to another machine, etc. In a sense, unsupervised learning can be thought of as finding patterns in the data above and beyond what would be considered pure unstructured noise. Two very simple classic examples of unsupervised learning are clustering and dimensionality reduction.*"

b) Supervised learning

"*The most common learning [method is] supervised learning, which provides a response value for every set of input values and requires a known (input) target value that the response is trying to guess"* (Jafar et al., 2010). Supervised training involved the following steps:

- Input values which had corresponding known output values were gathered. This set of data was usually called training data.
- An ANN was executed (trained) with training data. "*The goal of the machine is to learn to produce the correct output given a new input. This output could be a class label (in classification) or a real number (in regression)"* (Ghahramani, 2004). During execution the ANN weights were adjusted iteratively in order to minimize the error (Jafar et al., 2010).
- The model that was developed using the training data was tested with a separate set of data that contained inputs and known output values. This data set was usually called test data.

Najjar et al. (1997) identified some factors that needed to be considered during the development of ANNs:

a) Size of dataset.
b) Activation function.
c) Criteria for termination of training.

d) Number of training cycles (iteration).

a) Size of dataset

Najjar et al (1997) suggested "*that data to be used for training should be large enough to cover all possible variability within the application domain*" and that "*utilisation of 20 – 25% of the data for testing, from a relatively large mother database, is practically sufficient for examining the prediction accuracy of the network while maintaining a wide variety of patterns for the network to learn*."

b) Activation function

"*The activation (transfer) function is necessary to transform the weighted sum of all signals impinging onto a neuron into a state which determines the firing intensity of that neuron*" (Najjar et al., 1997).

c) Criteria for termination of training

"The convergence criteria is usually based on the error representing the difference between the target output(s) and the predicted output(s). Training is allowed to proceed until the predicted output(s) for any pattern agrees with the target output(s) within a pre-specified tolerance" (Najjar et al., 1997).

d) Number of training cycles (iteration)

It was important to determine the number of training cycles required to develop a good model. Using too many cycles could result in overtraining which was "*detrimental to the capacity of the network for generalising from unseen data (a network that can accurately predict the output of the testing patterns is said to have generalised)*" (Jafar et al., 2010).

2.9.4. Discussion

This Section provided an overview of data mining, web mining, online behaviour prediction and Artificial Neural Networks as well as some factors that needed to be considered during the development of ANN models. ANN algorithms were used during the data mining stage of the research, which is described in Chapter 7.

2.10. Chapter Discussion

This Chapter reviewed the background research that was carried out into several areas that were relevant to the research work described in this dissertation. The literature search identified a gap in knowledge in some key areas.

Existing research literature had proposed a number of website design and landing page design techniques that could be used to improve the performance of websites and landing pages. However, the literature did not present results to show how the application of these design techniques affected the conversion rate or bounce rate of a landing page or a website.

Existing research about predicting online conversion had focused on retail websites and the prediction of purchase. There is no research published regarding visitor behaviour on a service website and the prediction of conversion on such websites. Also, the methods used for predicting purchase conversion have included the use of Bayesian models, logit models and other statistical methods. It appears that linear regression models, Find Laws and Neural Networks had not been used as a method for predicting online conversion. Search keyword length and relevance had also not been previously used as predictors of conversion.

Section 2.1 described the different types of CRM systems, existing commercial CRM systems as well as important considerations when implementing CRM and measures to evaluate the impact of the introduction of CRM in an organisation. These findings were taken into consideration during the implementing a CRM system that is described in Chapter 3.

Section 2.2 gave an overview of online advertising, Pay Per Click advertising and the concept of long tail keywords. Chapter 5 describes how PPC advertising was used to drive traffic to websites that were created in this research.

Section 2.3 considered the importance of website design and navigation design. It identified a number of design elements that were considered during the design of websites that are described in Chapter 4.

Section 2.4 described behavioural intention and how it could affect visitors' behaviour on a website. It also identified two main types of website visitors. Section 2.5 described measures of website performance. Section 2.6 described design guidelines that were important for landing pages. Section 2.7 described LPO and LPO experimentation methodologies. These findings were considered during the optimisation of landing pages described in Chapter 6.

Section 2.8 described how Web users search for information and how a search query could be regarded an expression of a Web user's intent. Section 2.9 gave an overview of data mining, web mining and Artificial Neural Networks. The finding from Section 2.8 and 2.9 were used during data mining described in Chapter 7.

CHAPTER 3

CUSTOMER RELATIONSHIP MANAGEMENT SYSTEMS

A first step in improving customer generation was to understand who customers were as well as their needs. In order to achieve this, the research needed to collect data about customers from an initial visit to a website, through to product delivery.

This Chapter describes the Information Technology (IT) systems and architecture that existed at the beginning of the research. It describes the limitations of these systems and the implementation of the Microsoft Dynamics Customer Relationship Management 3.0 (MS CRM) system that replaced the existing GoldMine Business Contact Manager 5.7 (GoldMine) system. It also describes how the author customised MS CRM to capture customer data throughout the collaborating company's sales cycle and extended the system to collect additional customer data after product delivery. Finally, it describes how the data captured by MS CRM was used to extract knowledge about customers.

3.1. Existing IT architecture

The existing IT architecture is shown in Figure 3.1. The diagram shows interaction between users as well as interaction with software systems. The IT architecture shown in Figure 3.1 consisted of four systems:

1. GoldMine that was used to manage customers and sales activities. GoldMine was a popular solution for small businesses that needed to manage contacts,

track sales and marketing activities. GoldMine was particularly suited to environments where teams had to work together, share customer information and collaborate (FrontRange Solutions Inc, n.d.). GoldSync was part of the GoldMine system and enabled remote users to synchronise data with the GoldMine database.

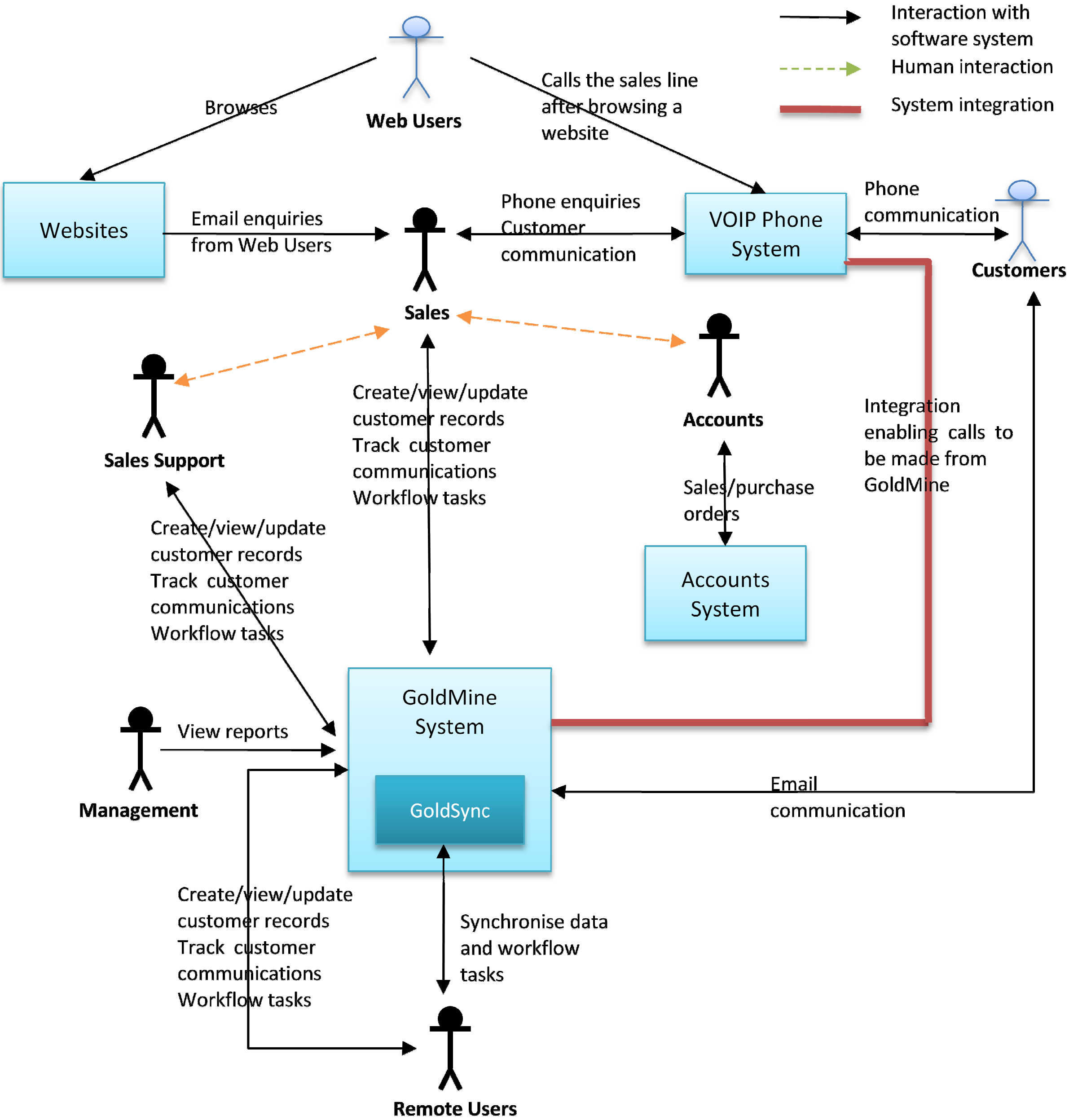


Figure 3.1: Existing IT infrastructure.

2. Voice Over Internet Protocol (VOIP) phone system which was accessible to staff in the form of a client application that ran on their computer. The GoldMine system interfaced with the VOIP phone system through a plug-in that allowed users to telephone customers from GoldMine.
3. Accounts system for processing sales and purchase orders.
4. Websites. There was a main website and a number of micro sites. The main website promoted all the services that the collaborating company offered whereas each micro site focused on a specific service. The websites were static with no back-end technology for tracking and logging the activities of visitors. As a result, there was a lack of knowledge regarding who website visitors were, where they came from, what they were looking for and how they behaved on the websites.

3.1.1. Limitations of existing IT architecture

Some limitations of the existing IT architecture were:

- Lack of automation.
- Limited information capture and reporting on KPIs.
- Little information flow.

3.1.2. GoldMine Business Contact Manager

GoldMine was used to capture and manage customer details and communication throughout the collaborating company's sales cycle. The main features of GoldMine included:

- An email system that allowed users to send and receive emails within GoldMine. GoldMine automatically linked emails to customer records thus keeping an accurate history of all communications with customers.

- A basic workflow system that could be customised to send tasks to company staff automatically.
- A basic interface for interrogating data stored in the GoldMine database using Structured Query Language (SQL) queries and for generating reports. Generating reports in GoldMine required technical skills. Staff found this to be a significant drawback of GoldMine.

It can be seen from Figure 3.1 that there was limited integration between the various systems. GoldMine interfaced with the phone system through a plug-in that allowed users to dial a customer’s phone number by clicking a button within the customer’s record. Sales staffs acted as an interface between GoldMine and the websites. The websites generated enquiries in the form of emails and sales staffs were responsible for creating a lead record in GoldMine for each enquiry and then assigning it to a member of the sales staff. Leads created from phone calls were assigned to the sales staff who answered the call.

The collaborating company’s sales process specified that details such as telephone number and email address had to be captured and stored for all customers. However, this rule could not be enforced in GoldMine as it did not allow for data fields to be made mandatory. As a result, customer information that was stored in GoldMine was sometimes incomplete.

Sales staff used GoldMine to manage customers from the lead stage to the order stage. When an order was placed, customer and order details were compiled into a sales order by a sales person. The sales order was then handed over to staff in the Finance department who entered it into the accounts system.

GoldMine ran workflow processes that assigned tasks to users. However, the workflow features were basic and did not allow for complex business processes to be completely automated and coordinated. It was difficult to implement business processes that had non-sequential or iterative stages. Human intervention and interaction was crucial in executing business processes.

GoldMine could be difficult to customise. Therefore, minimal customer data was captured. The data was not detailed enough for analysis and decision making. GoldMine did not classify records in terms of the common stages that were present in most sales cycles, for example leads, opportunities and orders. It was therefore difficult to:

- build and monitor a sales pipeline.
- measure the performance of the sales team at various stages of the sales cycle.
- identify bottlenecks in the sales cycle.

These limitations of GoldMine meant that a significant amount of work and effort was required to manage customers, convert leads to orders and perform day-to-day tasks. This restricted the number of enquiries and orders that the collaborating company could deal with at any given time. This made it difficult for the business to grow and expand. Overall, GoldMine could not store the data required for the research and did not provide the necessary tools to support the sales and management team.

In order to overcome these limitations of GoldMine, an alternative CRM system was required to provide:

- Better information capture.
- Reports that did not require technical skills to be created and could be generated by any member of staff.

- Workflow features that enabled easy automation of business processes and information flow.
- Flexibility through customisation and the ability to extend the system with either third party software or custom built software modules.
- Integration capabilities so that it could interface with other systems.

3.2. Specifications for an alternative CRM system

The specification of an alternative CRM system was defined after researching the different types of CRM systems described in Chapter 2, and after gathering information from sales and marketing staff at the collaborating company.

3.2.1. Users

Four different types of users were identified for the CRM system. These were:

a) Sales staff.
b) Sales support staff, for example, engineers, project managers and product sourcing staff.
c) Management.
d) Administrators.

a) Sales staff

Sales staff would use the CRM system on a daily basis to carry out tasks relating to sales. These tasks included communicating with customers via email, using calendar tools to schedule meetings and creating and modifying customer records in the CRM system. Sales staff typically had limited experience of using word processors, spreadsheet and email software.

b) Sales support staff

Sales support staff included project managers, product designers, electronic engineers and product sourcing staff who worked with the sales team to deliver customer orders. They would use the CRM system to access and update customer data as well as for communications.

c) Management

Management would use the CRM system to generate and view reports about performance at various stages of the sales cycle.

d) Administrators

Administrators would carry out maintenance and customisation of the CRM system, including modifying the interface, implementing or editing workflows and creating reports.

Sales staff were expected to use the CRM system on a daily basis to carry out all sales related tasks. They would use it more than other members of staff. Therefore, the system needed to be customised to meet their needs. Sales support staff were expected to use the CRM system mainly as an information repository and as a tool to track communications with customers.

Management required custom reports and dashboards that would display Key Performance Indicators (KPIs) derived from the data stored in the system. They would also require training to learn how to use the CRM system's reporting tools. Administrators needed to understand the technical details of the CRM system in order to maintain and customise it when required.

3.2.2. System Goals

The goal of the new CRM system was to increase company performance by:

- Increasing sales staff performance and thus increasing sales.
- Generating repeat business by improving customer satisfaction through improved customer relationships.
- Improving decision making through the use of reports and dashboards.
- Improving information flow and promoting collaboration between departments.
- Collecting detailed customer data. This data would be used to extract knowledge about customers and improve marketing.

3.2.3. System Attributes

The alternative CRM system needed to have the following attributes:

a) Ease of use.

b) Ease of customisation.

c) Extendibility.

d) Support for data mining and analysis.

a) Ease of Use

The alternative CRM system would be used mainly by sales staff with limited experience of computer applications. Therefore, the system needed to be easy to use. "*A salesperson's belief regarding CRM ease-of-use and CRM usefulness have a catalytic influence on sales performance*" (Avlonitis and Panagopoulos, 2005).

b) Ease of customisation

The alternative CRM system needed to be easily customisable especially with regards to the Graphical User Interface (GUI), data capture and workflows. Similarly, it was important to be able to easily edit or create new workflows which could automate existing and new business processes.

c) Extendibility

The alternative CRM system had to support business processes that would evolve and change. Therefore, it had to be extendable so that new features could be created when required. Extendibility was also important in case the alternative CRM system needed to be integrated with existing systems.

d) Support for data mining and analysis

The CRM system's database would act as a main data repository for customer data. This would be achieved through customisation of the CRM system's database and integration with existing systems. It was important for the CRM system to have reporting functions that users with limited technical abilities could use to analyse data stored in the system's database. It was also important for the back-end database to be directly accessible to more experienced users who might want to carry out more advanced data mining and analysis.

3.2.4. System functions

There were three main functions that were required from the alternative CRM system in order to meet the goals stated in Section 3.2.2:

a) Customer management.
b) Workflow service.
c) Reporting features.

a) Customer management

The CRM system needed to have functions to create, edit and save customer data easily. It needed to be able to track customer communications and have tools and features that would support the day to day tasks of the sales team.

b) Workflow service

A workflow was an application or service that ran constantly. Workflows evaluated CRM data according to conditions that were pre-defined in business processes. When the workflow service determined that a condition had been fulfilled, it started the appropriate workflow rules to run workflow actions. Typical workflow actions included sending an e-mail message, creating a task or updating a data field. Workflows were needed to automate business processes and sales processes and to coordinate tasks across departments.

c) Reporting features

The CRM system needed to have reporting functions so that knowledge could be extracted from data that was collected. The CRM system also needed to support quick and easy creation of custom reports. Examples of reports included sales pipeline, number of enquiries received in a period of time and number of quotes sent every month.

3.2.5. System implementation

Three CRM software packages were investigated namely MS CRM, SugarCRM and the latest version of GoldMine. These software packages were reviewed in Chapter 2. MS CRM was selected as it met the requirements for the system attributes (described in section 3.2.3) and system functions (described in section 3.2.4) while still being affordable. The process for rolling out MS CRM involved:

- Agreeing the system requirements with all stakeholders at the collaborating company.
- Building a test environment in which
    - MS CRM was evaluated against detailed requirements and then customised (including user interface, workflows, reports and dashboards) according to requirements gathered through meetings with sales staff and management.
    - Data migration from GoldMine to MS CRM was tested.
    - Selected users were asked to use MS CRM and provide feedback. Modifications were made to the system based on user feedback.
- Deploying MS CRM and customisations to a live server.
- Migrating selected data from GoldMine to MS CRM.
- Training staff.
- Rolling out MS CRM.
- Phasing out the use of GoldMine.
- Getting feedback about MS CRM and making modifications accordingly.

Initially the plan was to migrate all data stored in GoldMine to MS CRM. However, the database structures of the two systems were different so this was more complex than originally anticipated and increased timescales. As a result only a selected set of data was migrated. Once data had been migrated and staff trained, MS CRM was rolled out in the collaborating company in May 2006. The IT architecture at this stage is shown in Figure 3.2.

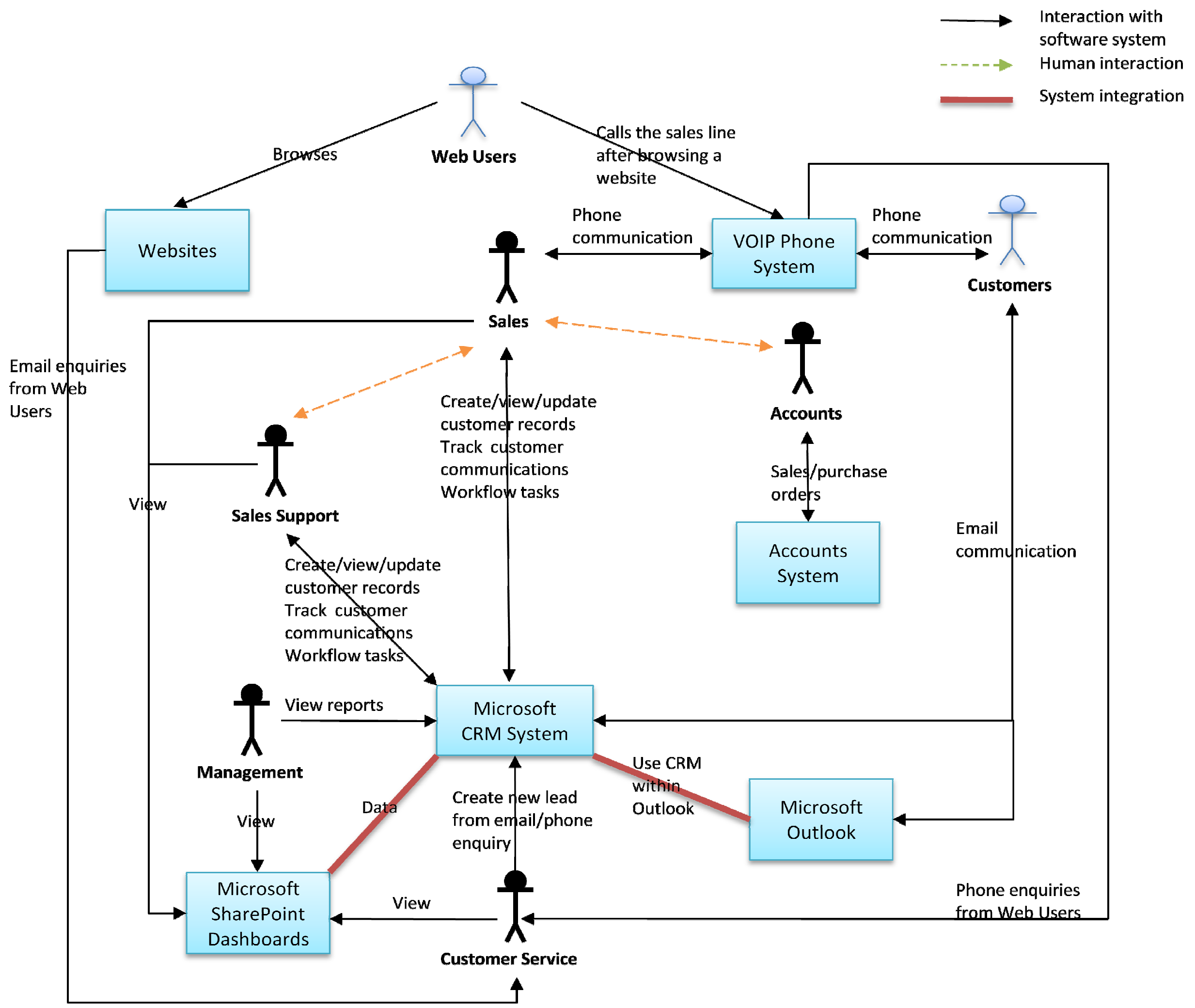


Figure 3.2: IT architecture after the introduction of MS CRM.

With the implementation of MS CRM, new business processes and information flow were introduced. Some of the changes included:

- A dedicated member of the Customer Service team was tasked with handling phone enquiries as well as email enquiries that were received from the websites. When a new customer made an enquiry, a lead record was created in MS CRM and populated with the customer's information. Selected fields were made mandatory in MS CRM to ensure that certain information was always captured for example email address and customer name. New data such as the source of an enquiry and the method of enquiry were also stored in customers' records.

Once a record was created, it was then assigned to a sales person. This triggered an automatic workflow which coordinated the efforts of the various departments that needed to work with the new customer so as to deliver an order.

- Dashboards were implemented and customised. They were available for both management and staff. MS CRM captured detailed data allowing reports to be generated and KPIs to be monitored. Dashboards displayed data stored in MS CRM graphically and provided easy access to statistics that were used in decision making.
- Remote staff could access MS CRM directly and did not need to synchronise with the system to see up-to-date data.

Over the months following its launch, MS CRM was extended to have custom modules that would provide project management functions, the ability to capture data from customer surveys, quality control functions and a scoring system for identifying high quality enquiries.

3.2.6. Custom Modules

The MS CRM system was not integrated with any of the other existing systems such as the websites, telephone system or accounting system. Human interaction was still required to carry out some tasks and achieve information flow especially for project management. In order to collect data beyond the sales cycle, the MS CRM system was extended with the following custom built modules:

a) Customer Satisfaction Survey (CSS) module.
b) Project management module.
c) Quality control module.

d) Opportunity marker.

a) Customer Satisfaction Survey module

At the beginning of the research, customer satisfaction surveys were not carried out. As part of the strategy to collect detailed customer data, a Customer Satisfaction Survey (CSS) was conducted by the marketing department. Marketing staff telephoned existing customers and asked them a series of questions. Answers to these questions were stored for analysis. This survey data was linked to customer data stored in MS CRM. In order to achieve this, the MS CRM was extended by creating a new CSS module. The CSS module worked within the MS CRM GUI, providing a seamless experience for users. Reports were created in MS CRM to view the data collected by the CSS module.

b) Project Management Module

At the beginning of the research, the collaborating company did not have a project management system. Sales staff managed the delivery of customers' orders without formal recordings of problems and with limited quality control. There were no formal procedures for managing and delivering customers' orders.

A project manager was recruited to take over project management responsibilities and to introduce procedures for managing and delivering customer projects. The project manager's initial responsibilities were to keep track of:

- The stages of the different projects.
- Payment dates and amount for customers and suppliers.
- Project delivery dates.
- Problems that occurred during production and delivery.

The project manager used a spreadsheet to record project data and used MS CRM to keep track of communications with customers. This method worked well in the beginning when there were few projects but became inefficient as the number of projects increased.

The main weaknesses associated with managing projects in this way were:

- It was difficult to update timescales. For example, if the start date of a project was delayed then this would affect subsequent dates. Updating the dates manually was time consuming and error prone.
- Payment information was updated by the project manager as well as by the product sourcing staff. This made version control problematic.
- It was difficult to generate reports from the data stored in a spreadsheet.

In order to address these weaknesses a project management module was created in MS CRM that had the same structure as the spreadsheet. Workflows were also created to help coordinate tasks that required collaboration between departments or teams.

The new module brought the following benefits:

- Staff across the company could access timescales and payment information for any project through MS CRM.
- It was possible to generate reports from the data captured by the module, for example, cash flow reports.

c) Quality control module

Following the introduction of quality control management processes at the collaborating company, a quality control module was created by extending MS CRM. This module collected quality control data that could be cross-referenced with existing customers' records in MS CRM.

d) Opportunity marker module

In order to allocate its resources efficiently, the collaborating company needed to identify enquiries that had a high potential of turning into sales. At the beginning of the research, enquiries were rated as ‘Hot’, ‘Warm’ or ‘Cold’ based on the opinion of individual sales staff. In an effort to improve the accuracy of these ratings, a scoring system was created. This system calculated the quality of enquiries based on a number of criteria. These criteria included:

- Size of business for example, small company, big multinational, individual, etc.
- Delivery timescales.
- The size of an order.
- Whether there would be repeat orders.

The opportunity marker module was created as a web form that displayed a series of questions. Each question had a set of pre-defined answers shown in a drop down menu. Answers were scored and the total was the quality score. Sales staff used the opportunity marker to score every enquiry. This score was stored automatically in MS CRM. Resources were allocated based on the score obtained by an enquiry.

The IT architecture at the time of writing is shown in Figure 3.3. The data stored in MS CRM was integrated with data stored by the main website via a unique customer ID that allowed data to be cross referenced between MS CRM and the main website’s Online Tracking Module (OTM). This is discussed in more detail in Chapter 4.

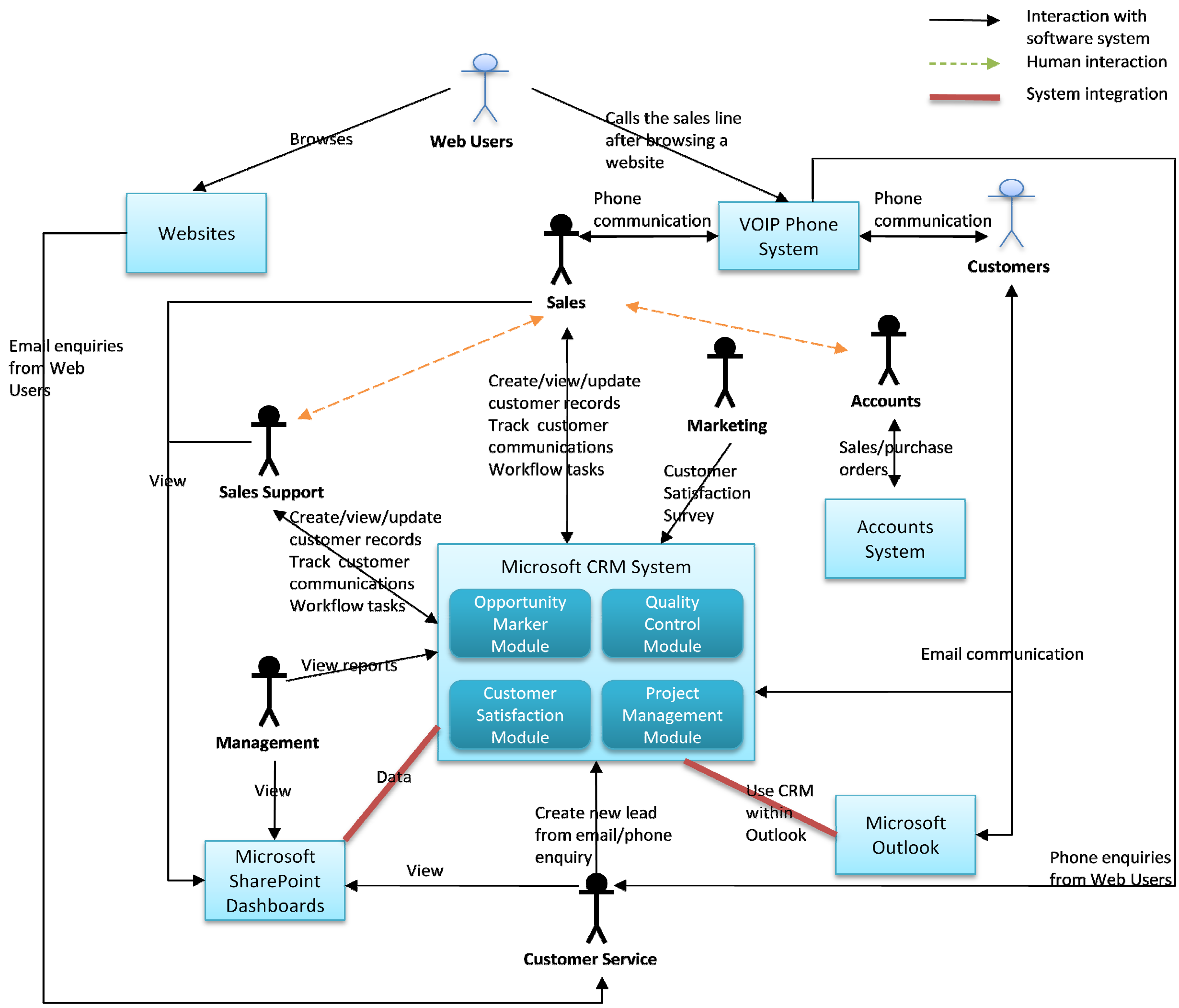


Figure 3.3: Current IT architecture.

3.2.7. Reporting

MS CRM enabled easier reporting on KPIs. A number of reports and dashboards were created to view the data stored in MS CRM. The dashboards were used by the collaborating company's management team to monitor the company's performance.

Other members of staff used the dashboards to keep track of their performance and ensure that they were meeting their targets. For example, sales staff was targeted on the number of opportunities that they opened every month. Dashboards were created to allow them as well as management to monitor this easily.

MS CRM reports enabled the extraction of knowledge about customers using the data that was stored in MS CRM. Data could also be manipulated and viewed by querying the MS CRM database directly.

3.2.8. Systems/Data Integration

MS CRM could be easily integrated with Microsoft products such as Microsoft Outlook and Microsoft SharePoint. Integration of MS CRM with non-Microsoft systems found in the IT architecture was investigated. It was found that because different technologies were used by the systems, integration would be expensive, complex and time-consuming. However, data integration between the main website and MS CRM was achieved by collecting behavioural data on the website for each visitor and then cross referencing this data with corresponding customer records in MS CRM. A unique ID was stored in both systems to enable cross referencing. Integration between MS CRM and the main website is explained in more detail in Chapter 4.

It was possible to integrate MS CRM with the main website so that online behavioural data for visitors who enquired by sending an email from the website, was automatically stored in MS CRM. However, this required another software component that was expensive. Therefore, this solution was not implemented.

3.2.9. Knowledge extraction

A few months after MS CRM was rolled out, enough data had been collected by the system for initial analyses. At this point in the research, the OTM was yet to be created. Therefore, there was no online behavioural data available for analysis. Initial analyses of the data captured by MS CRM generated information about:

- Enquiries, for example:

    - type of enquiry, that is whether an enquiry was about one of the following services: manufacturing , mechanical design or electronic design.
    - total number of enquiries received in a period of time.
    - total number of enquiries received for each type of enquiry.
    - the reason why an enquiry was rejected by the sales team.
    - types of enquiries commonly rejected by the sales team.
    - countries from which different types of enquiries originated.
- Customers, for example:
    - customer types, that is whether they were individuals, SMEs or corporate.
    - customers types who placed orders.
    - Whether a particular customer type was more likely to generate a particular type of enquiry.

Chapter 4 describes how online behavioural data was collected by an Online Tracking Module and how some of this data was stored in MS CRM.

3.3. Chapter Discussion

In order to improve customer generation it was important to first understand existing customers and their needs. This required the extraction of knowledge from customer data that had been collected and stored in various systems at the collaborating company.

This Chapter described the IT structure, CRM system and business processes that existed at the beginning of the research. The existing GoldMine system had a number of limitations that made data collection and analysis challenging and also limited the performance and growth of the collaboration company. The GoldMine system was replaced with MS CRM to overcome these limitations.

The MS CRM system was customised and extended to collect customer data throughout the sales cycle and post product delivery. The MS CRM system was also integrated with the main website using a common ID (described in Chapter 4).

Detailed customer data was captured using the MS CRM system. This data was used to carry out initial analyses to extract knowledge about customers. Online behavioural data needed to be captured to enrich existing customer data so that more knowledge about customers could be extracted. Chapter 4 describes the implementation of an Online Tracking Module which captured data about website visitors' browsing behaviour.

CHAPTER 4

WEBSITES USED IN THIS RESEARCH

In order to gain insight into customers' needs and motivation, and build a knowledge base, data is often required. To achieve a similar goal in the research described in this dissertation a procedure was implemented to acquire data. A first step in achieving this was to implement and customise MS CRM to act as a central customer data repository. MS CRM captured data throughout a sales cycle and was extended to enable the collection of additional data after product delivery (described in Chapter 3). The next step involved collecting data about website visitors so that the research could gather data about customers from their initial visit to a website, through to product delivery.

In order to do this, an Online Tracking Module (OTM) was created to collect marketing and behavioural data for all visitors who browsed the collaborating company's websites. A solution was then implemented to allow data stored in the OTM to be cross referenced with data stored in MS CRM for visitors who had enquired.

At the beginning of the research, the collaborating company had static websites that had no visitor tracking or personalisation capabilities. This Chapter describes these websites and how a first version of the OTM was created. It goes on to describe how a dynamic main website that had personalisation capabilities was created to replace the existing main company website at the collaborating company and how the implementation of this dynamic main website led to the development of a second version the OTM with improved data capture functionalities.

The author identified the requirements and wrote the specifications for the new dynamic website and OTM. The author also designed these systems together with a junior web designer who worked for the collaborating company. The code for the new dynamic website and OTM were written by the junior web designer under the supervision of the author.

4.1. Existing websites

At the beginning of the research, the collaborating company had simple websites with static pages. These websites had little functionality and did not collect marketing and behavioural data about visitors. Figure 4.1 shows the set up of the websites and their traffic sources at the beginning of the research.

It can be seen from Figure 4.1 that the collaborating company had two types of websites:

- A static main website.
- Several static micro sites.

The collaborating company used its websites to promote its services. They offered two main types of services, product design and product manufacturing. Each type of service was divided into a number of specialist areas for example, product design included electronic design and mechanical design while product manufacturing included plastic production and tooling production. Some of these specialist areas were further broken down into specialist services for example, plastic production included different types of plastic moulding and plastic extrusion.

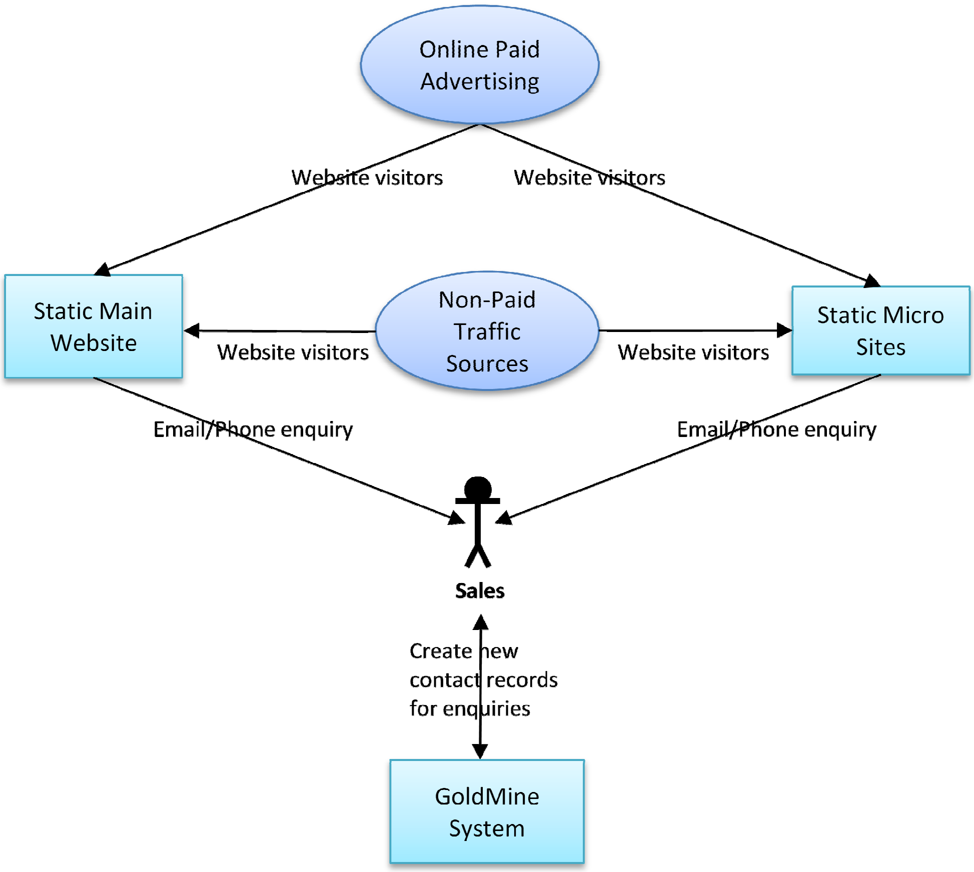


Figure 4.1: Existing websites.

The static main website promoted all the services that the company offered. It contained a large amount of content, giving the website a deep structure. The micro sites on the other hand were smaller websites. Each micro site focused on a specialist service area for example plastic production, electronic design etc.

Pay Per Click (PPC) advertising (discussed in Chapter 5) and non-paid online traffic sources such as organic search engine listings and web directory listings generated traffic to the websites.

Email and phone enquiries that were generated by the websites were handled by a member of the sales team. Customer details obtained from phone and emails enquiries were used to create unique contact records in GoldMine.

Several changes were made to the websites during this research. Table 4.1 summarises these changes which are discussed in the following Sections of this Chapter.

Stages	Changes
Stage 1	A first version of a new OTM was created and used to monitor visitors on all websites
Stage 2a	A new dynamic main website replaced the existing static main website. The back-end of the dynamic main website was new but the front-end was the same as the existing static main website. The research stopped generating traffic to the micro sites. They were not used from this point on; only the dynamic main website was in operation.
Stage 2b	A new improved version of the OTM was launched on the new dynamic main website only.
Stage 3	The front-end which included the visual design, layout and content of the new dynamic main website was changed.

Table 4.1: Stages of major changes to the websites.

4.2. Stage 1 – New Online Tracking Module

The first OTM was created in order to determine how online advertising campaigns were performing. By cross referencing customer data stored in MS CRM with advertising data collected by the OTM, it would be possible to identify the strengths and weaknesses of the online advertising campaigns.

Specifications of the OTM were defined after collating marketing information.

4.2.1. System goals

The goal of the OTM was to capture data about website visitors so as to enrich the data stored in MS CRM. Knowledge extraction from more detailed data could help improve understanding and performance of online advertising campaigns. The OTM captured data regarding:

- Online advertising campaigns that sent visitors to the websites. This data could be combined with data stored in MS CRM to identify the online advertising campaigns that generated enquiries.
- Search engines that sent visitors to the websites. This data could be used to understand which search engine visitors originated from and the type of visitors that search engines attracted to the websites.
- The search keywords that visitors had used to find the websites. This data could be used to identify search keywords that generated high quality enquiries or high value sales. This information could also be used to identify new search keywords for PPC campaigns.
- Browsing behaviour so as to understand how visitors browsed the websites.

4.2.2. System functions

There were two main functions that were required from the OTM:

- *Data collection:* The system needed functions to track and record data relating to online advertising, for example which search engines and advertising campaigns attracted visitors to the websites. The OTM also needed functions to track and record visitors' behavioural data that is, data relating to their browsing pattern on the websites. It was important for the OTM to be able to collect information in the background without interfering with visitors' browsing, for example by prompting them to fill in forms or provide information explicitly.

- *Reporting:* The system needed a simple interface to allow staff at the collaborating company with average computer skills to run reports on the data collected.

4.2.3. System implementation

The OTM's functions were implemented using Active Server Pages (ASP). A Microsoft Structured Query Language (MS SQL) database was used for storing data. Some of the core methods of the OTM included:

a) Generation of a unique identification code (*[mainID])* for each visitor.

b) Referer filtering.

c) Capturing visitors' browsing activity.

d) Identifying visitors' country.

e) Allowing a user to score the quality of enquiries and retrieve related marketing data.

f) Generating reports from data that was collected.

a) Generating a visitor's unique identification code ([mainID])

Each visitor was given a unique identification code called *[mainID]* so that they could be tracked. The *[mainID]* was generated using a counter (*[ID]*) that was stored in the OTM's database. When a visitor landed on a website that was monitored by the OTM, the value of *[ID]* was fetched from the database and assigned as *[mainID]* to the visitor. The value of the counter was then incremented and stored back into the database, overwriting the old counter value, and ready to be assigned to the next visitor.

Data stored in the OTM database about a particular visitor could be accessed using their *[mainID]*. The *[mainID]* was included in email enquiries whenever possible. If the *[mainID]* could be obtained from an email enquiry or from a person who enquired by telephone, it was stored in MS CRM and used to cross reference data between the MS CRM and OTM databases.

Website visitors enquired using three different methods:

1. An enquiry form found on the websites.
2. Their email client to send an email enquiry.
3. Telephoning the sales number advertised on the websites.

Method 1

When a visitor sent an enquiry form from a website, their *[mainID]* was automatically included in the body of the form. When the details contained in the enquiry form were used to create and populate a customer record in MS CRM, the *[mainID]* was also stored.

Method 2

Some visitors sent email enquiries by clicking on a contact email address link. Clicking on a contact email address link automatically started their email client. In such instances, tracking data that was included in the body of an email could be seen and deleted by a visitor. Therefore, it was not always possible to obtain a *[mainID]* from an email enquiry.

Website visitors could also manually open up their email client, enter the contact email address that was displayed on a website's **Contact Us** page, compose a message and then send the email. In such cases, it was not possible to include tracking data into the email and such enquiries could not be cross referenced with the OTM.

Method 3

Visitor's *[mainID]* was displayed as a promotional code in the footer of the websites and each visitor could only see their *[mainID].* When a website visitor made an enquiry by telephone, they were asked for a promotional code and if they were able to provide one, it was stored in their MS CRM record. It was not always possible to obtain a promotional code from website visitors who enquired by telephone.

Due to the issues associated with Methods 2 and 3, there were some instances where the data stored in MS CRM and the OTM could not be cross referenced.

b) Referer filtering

The HTTP protocol was a "*request/response protocol. A client sends a request to the server in the form of a request method, Uniform Resource Identifier, and protocol version, followed by a MIME-like message containing request modifiers, client information, and possible body content over a connection with a server. The server responds with a status line, including the message's protocol version and a success or error code, followed by a MIME-like message containing server information, entity meta information, and possible entity-body content*".

Request-header fields "*allow the client to pass additional information about the request, and about the client itself, to the server. These fields act as request modifiers, with semantics equivalent to the parameters on a programming language method invocation*."

The Referer was a "*request-header field [that] allows the client to specify, for the server's benefit, the address (URI) of the resource from which the Request-URI was obtained (the "referrer", although the header field is misspelled.)"* (Fielding et al., 1999)

The Referer was not sent if the Request-URI was obtained from a source that did not have its own URI, for example using a bookmarked URI or input from a user's keyboard. Therefore, the value of the Referer could either be a URI address or empty.

The Request Object and ServerVariables collection provided by ASP could be used to obtain predetermined environment variables and request header information (MSDN Library, n.d.). These were used to obtain the URI from the Referer field in the request header each time a website received a request. The URI was then parsed to extract a search engine name and a search keyword. These values were then stored in the OTM database along with the URI obtained from the Referer.

An issue with the third party code used to parse the Referer URI was that it checked the URI against a static pre-defined list of search engine names in order to identify the Referer website and extract a search keyword. Although, the list included the most popular search engines, it was not a complete list and did not include directories. As a result not all Referers could be matched successfully.

If a Referer was not matched against the list, NULL values were stored in the OTM database for the fields corresponding to search engine name and search keyword. If the Referer string request header was empty, then NULL values were stored in the OTM database for the fields corresponding to the Referer, search engine name and search keywords.

c) Visitor browsing activity

Data relating to user activity that needed to be recorded are shown in Table 4.2.

Data Field	Description
referer	URI from which the visitor originated.
keywords	The search keywords that the visitor had used.
engine	Name of search engine from which the visitor originated.
countryCode	Country from which visitor accessed the website.
land	Landing page, the first page that a visitor accessed.
decIP	The visitor's IP address in decimal format.
sent	Binary value indicating whether the visitor has sent an email enquiry from a website.
tStamp	Date and time at which a visitor landed on a website.
gc	Online campaign code. Each online advertising campaign was given a unique code.
mainID	Unique ID that identified a visitor
pID	Unique ID for each page visited.
dTime	Time in seconds spent on a web page.
vID	Video ID of a video viewed by a visitor.

Table 4.2: Data collected by OTM.

The structure of the OTM database as well as detailed table definitions can be found in Appendix A (Section A.1).

The OTM started tracking visitors as soon as they landed on a website and collected data relating to the activities described in Table 4.2. Each time a visitor accessed a page the tracking function was executed and the visitor's history was updated in the database. In order to calculate the time spent on a page (*dTime*), the time at which a visitor landed on a page was recorded and subtracted from the time at which the visitor landed on the next page.

If the visitor did not visit another page, then the time spent on the current page could not be calculated and was set to NULL. The time spent on a page (*dTime*), did not necessarily represent the actual amount of time that a visitor was active for on that page. The OTM could not detect when a visitor was inactive, for example if a visitor left a page open while working on something else or moved away from their computer.

d) Location

The Request Object and ServerVariables collection was used to obtain the Internet Protocol (IP) address of the remote host who made a request. The IP address was cross-referenced against a third party database to find the corresponding country which was then stored in the OTM database.

e) Enquiry quality score and marketing data retrieval

The OTM database was used to track the quality and type of enquiries that were generated by the website. A simple web page (see Appendix A, Figure A.3 and Figure A.4) was implemented so that the person in charge of handling email enquiries could store a value for lead type (*[leadType]*) and lead quality (*[leadClass]*) in the OTM database. This interface was also used to retrieve marketing data about visitors who had enquired. This data was then manually stored in visitors' MS CRM record. This was done for all enquiries that could be cross reference with the OTM database.

f) Reporting

A reporting interface was created to generate basic reports using the data collected by the OTM. The reporting interface was a web form that allowed a user to select a number of parameters and then click on a button to generate a report. Some of the parameters that were selected were used to construct SQL queries to retrieve data from the OTM database. Other parameters were used to format and display the data that had been retrieved and processed.

Reports generated using the OTM were mainly used to identify the type and quality of enquiries that online advertising campaigns were generating. Knowledge gained from these reports was used to identify online advertising campaigns that were performing poorly so that they could be improved.

4.2.4. Discussion

The OTM was successfully implemented on all websites. The data that it collected was used to generate performance reports about online advertising campaigns. The OTM had some weaknesses, for example it could not identify returning visitors. These were fixed in a new improved version of the OTM that was implemented later in the research (see Section 4.4).

4.3. Stage 2a – New dynamic main website

At the beginning of the research, the main company website was static. It was not flexible or customisable and could not support personalisation. *"The goal of web personalisation is to deliver the right content to the right person at the right time and to maximise immediate and future business opportunities"* (Tam & Ho, 2006). Web personalisation was akin to the mirroring principle where a product was built around the *"unique and particular needs"* of a customer (Hammer, 1995). This research created a new dynamic main website that could offer a better and more personalised browsing experience than the static main website.

4.3.1. System goals

The goal was to create a new dynamic main website that:

- Provided an improved browsing experience through content and navigation personalisation and generated high volume and high quality leads.
- Supported the separation of web page content from web page presentation, so as to provide more flexibility and control over presentation characteristics while maintaining consistency across the website.

4.3.2. System functions

The new dynamic main website consisted of a database back-end which stored the content for all pages on the website. The pages were generated by scripts that controlled various aspects of a page such as content, picture, menus and layout. The main features of the website that needed to be dynamic were the Left Hand Side (LHS) menu (Figure 4.2) and the content, which included both text and pictures. Functions required to achieve this included:

a) Customer interest evaluation. In order to achieve personalisation, it was important to understand what content a visitor was interested in. Visitors' search keyword and information about the type of pages that they browsed were used to infer interest.
b) Dynamic web page generation. Functions were required to dynamically generate pages using content and pictures that were stored in a database and layout specifications defined in a Cascading Style Sheet (CSS) file.
c) Dynamic LHS menu. Functions were required to display links in the LHS menu depending on the content of a page or a visitor's interest.

d) Dynamic content. Functions were required to change the textual content of pages based on a visitor’s search keywords or advertising campaign that they had originated from.

e) Dynamic inclusion of pictures. Functions were required to display images on pages based on a visitor’s search keywords or advertising campaign that they had originated from.

Figure 4.2: Screenshot showing the LHS menu that needed to be dynamic.

4.3.3. System implementation

The functions identified in Section 4.3.2 were implemented in ASP. A single database was created to store the content of the new dynamic main website and the data that would be collected by the new version of the OTM (described in Section 4.4). The database was referred to as the Content and Tracking (CAT) database. Its structure and detailed table definitions can be found Appendix A (Section A.2).

a) Customer interest evaluation

A list of the most popular search keywords used by visitors who browsed the static main website was identified using the data captured by the OTM.

Each search keyword was then given a design score and a manufacturing score between 0 and 1. These scores represented a customer's interest in either design or manufacturing content. When a customer landed on the dynamic main website, the customer's search keyword would be matched against the existing list of scored keywords to determine the customer's interest.

Each page on the dynamic main website was given a design score and a manufacturing score. Pages were given a score of 1 to indicate the bias of the content of the page, for example if a page's content was biased towards design services then it was given a score of 1 for *[ScoreDesign]* and 0 for *[ScoreManufacture]* and vice versa. Pages such as ***Contact Us*** pages and enquiry forms, which were not content pages, were given a *[ScoreDesign]* and *[ScoreManufacture]* of 0.

As visitors browsed the website, the page score and keyword score were used to calculate a score representing interest. This score (called *interest*) was stored in a session variable and was used to customise links in the LHS menu (see Figure 4.3). A visitor was shown more links to a certain type of content depending on their *interest* score.

b) Dynamic web page generation

Each page on the website would consist of a top horizontal menu, a LHS menu, a contents section and a footer. Figure 4.3 shows the structure of a web page and its components.

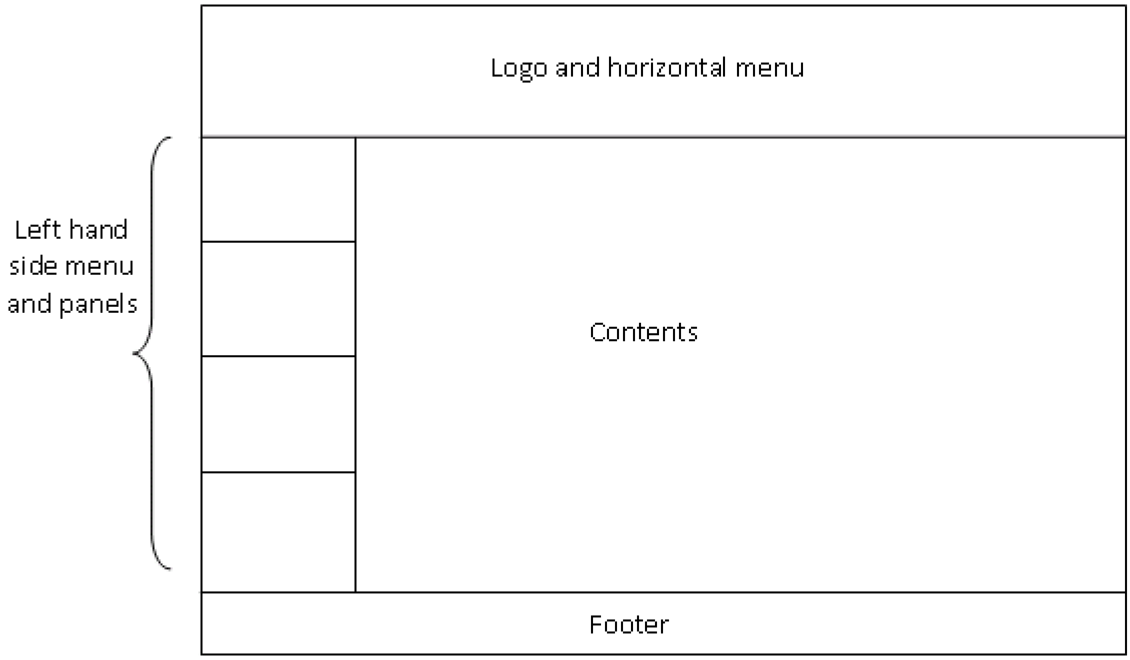


Figure 4.3: Structure of dynamic web page.

When a visitor accessed a web page, a script was executed which gathered the various components for the page from the database, formatted the page using a Cascading Style Sheet (CSS) file and presented it to the visitor.

c) Dynamic LHS menu generation

The left hand side menu was made up of a number of panels. The contents of the panels were defined in the database and each page was assigned a set of panels that made up its LHS menu. This allowed the LHS menu to be customised based on the content of each page, thus providing relevant links that could help visitors find information quickly. The LHS menu could also be customised based on a visitor's *interest* score.

d) Dynamic Content

The contents section (see Figure 4.3) contained both text and pictures. A new feature which dynamically altered the textual content of a web page was implemented. Particular words, sentences or whole paragraphs could be changed on-the-fly. A trigger or condition was required in order to activate this feature, for example the title of a page could be changed based on an advertisement that a visitor had clicked on, or particular words in the text could be replaced with the search keywords used by a visitor.

e) Dynamic inclusion of pictures

Pictures could be included on web pages dynamically if a predefined condition was satisfied, for example, a picture could be included based on an advertisement that a visitor had clicked on or based on the search keyword used by a visitor.

4.3.4. Discussion

Figure 4.4 shows the set up of the new dynamic main website and its traffic sources. After the launch of the new dynamic main website, the micro sites were taken offline. All traffic generated by PPC campaigns were directed to the new dynamic main website which had become the primary generator of enquiries. The look of the new dynamic main website was the same as that of the old static main website. However, the features and functionality of the new dynamic main website enabled content personalisation. This played an important role in designing landing pages that were effective at converting visitors into customers. This is discussed in more detail in Chapter 6.

The tracking module shown in Figure 4.4 was an improved version of the original OTM described in Section 4.2. Section 4.4 describes the new OTM in more detail.

4.4. Stage 2b – New and improved Online Tracking Module

The first version of the OTM was built around an existing website and its structure. As a result, some aspects of the implementation lacked flexibility. When the back-end of the dynamic site was created, it presented an opportunity to redesign some parts of the OTM.

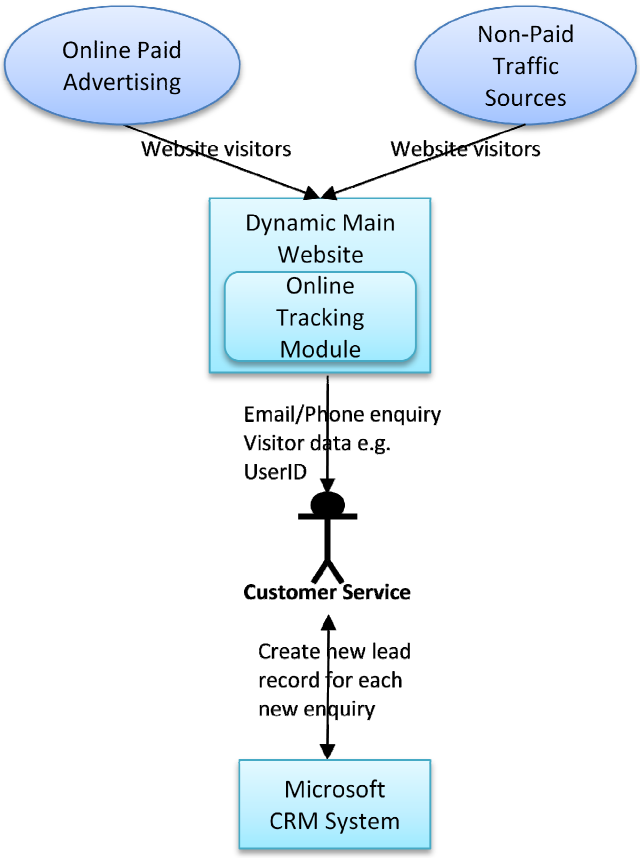


Figure 4.4: Current website systems.

4.4.1. System goals

The goals of for the new OTM included those describe in section 4.2. However, the new module also needed to:

- Capture complete browsing history for all visitors.
- Identify returning visitors and assign an ID for each visit so as to record browsing behaviour more accurately.
- Record which pages visitors sent email enquiries from.
- Have an interface that allowed
    - o easy access and sorting through the data that it had collected

    - reports to be generated using the new data that had been captured.
- Improve the ease with which
    - marketing data about visitors who had enquired was retrieved from the OTM and entered in MS CRM.
    - email enquiries were scored.

4.4.2. System functions

In order to meet the goals specified in section 4.4.1, the following functions were required:

- Returning visitors. The system needed to identify returning visitors and record their browsing behaviour under the same unique visitor ID (*UserID*) that had been assigned to them the first time they visited the website. Therefore, the system needed to generate and assign another ID (*microID*) to a visitor each time they returned to the website so as to differentiate between behaviour that took place during different visits.
- Identify where email enquiries were sent from. If a visitor sent an email enquiry from a website, the system needed to record which page it had been sent from.
- Quality score. All email enquiries received from the website were given a quality score. A better solution than the one implemented in the first version of the OTM was required to optimise the scoring of email enquiries and the retrieval of visitors' marketing data for subsequent input in MS CRM.
- Interface. An interface was required for staff at the collaborating company to generate reports from the data gathered by the OTM. This interface would be used primarily by the collaborating company's sales and marketing manager. More advanced reporting and data mining would be carried out using SQL to query the database.

4.4.3. System implementation

The OTM stored data that it collected in the CAT database (see Appendix A Section A.2). Functions were implemented to:

a) Identify returning visitors.

b) Record which page a visitor sent an enquiry from.

c) Simplify the scoring of the quality of email enquiries and retrieval of marketing data from the OTM.

d) Generate marketing reports from the data stored in the CAT database.

a) Returning visitors

Figure 4.5 shows how the OTM determined whether a visitor was a returning visitor or whether they were a first time visitor. The OTM relied on sessions and cookies to keep track of visitors.

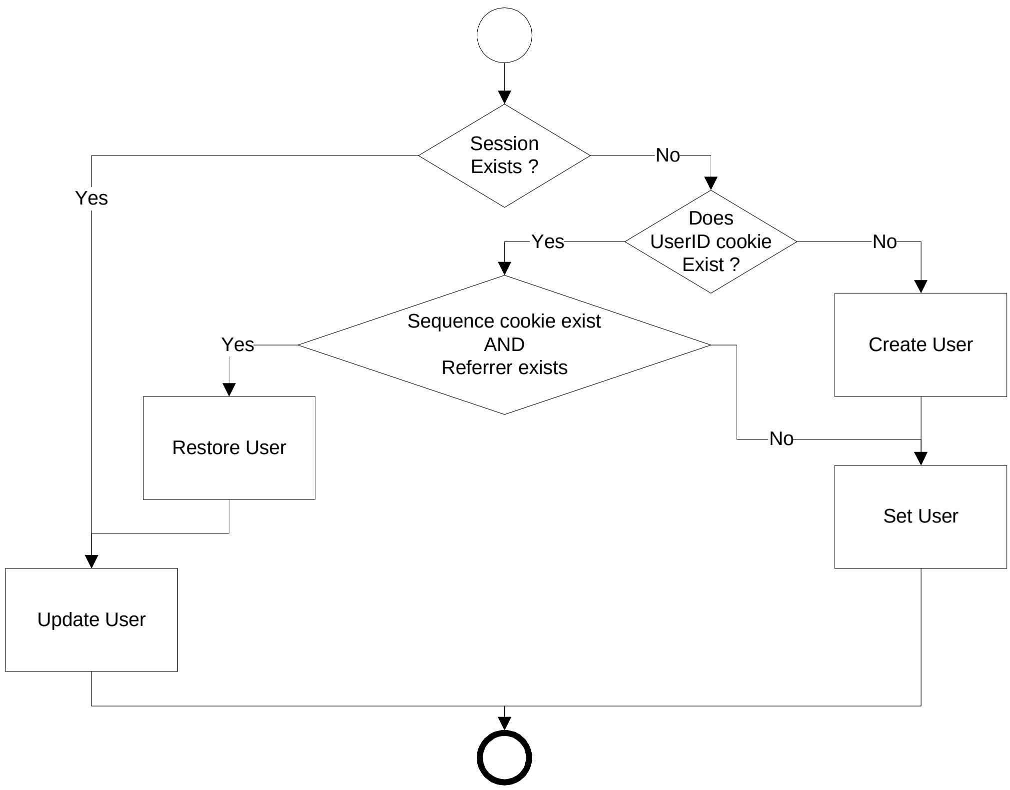


Figure 4.5: Flow diagram to identify returning visitors.

A session was defined as “*a sequence of actions undertaken by a user within a period of time. In order to determine when a* session *ends and the next one begins, a* session *timeout threshold (STT) is often used*” (Huynh and Miller, 2009). The OTM used a session timeout threshold that was set to 20 minutes.

“*HTTP cookies also known as Web* cookies *or just* cookies *are small parcels of text sent by a server to a web browser and then sent back unchanged by the browser if it accesses that server again. Cookies [were] originally designed to carry information between servers and browsers so that a stateful session could be maintained within the stateless HTTP protocol*” (Yue et al.).

The cookies created by the OTM were set to expire after a year except for a *[Sequence]* cookie which expired after 3,600 seconds. When a visitor landed on a website, the OTM checked whether there was an active session for that visitor. If there was, then the visitor was not regarded as a returning visitor. If a session did not exist, the OTM checked whether a cookie containing *[UserID]* existed on the visitor’s computer. If there was no cookie, then a *[UserID]* was created for the visitor. A cookie containing that *[UserID]* was stored on the visitor’s computer and a corresponding session was created.

If there was a cookie containing *[UserID]* on the visitor’s computer, then the *[Sequence]* cookie and the *[Referer]* cookie were checked. The *[Sequence]* cookie had a lifespan of an hour which was the period of time that the OTM used to determine whether a visitor had returned to the website or not. The *[Referer]* cookie stored the address of the website that the visitor had originated from. If the *[Referer]* value was NULL then it was assumed that the visitor has accessed the website directly by either typing the URL of the website into a browser or by using a bookmark. If a visitor did not have a *[Sequence]* value that was greater than 0 or was associated with a *[Referer]* whose value was NULL or contained the collaborating company's URI, then they were classified as returning visitors.

However, a visitor who had a valid *[UserID]* cookie was not considered a returning visitor if, they came back to the site within an hour of their last visit and did not access the site directly. Each returning visit was identified by a unique ID called a *[MicroID]*.

b) Website forms and emails

The website had a **Contact Us** page and various enquiry forms that visitors could use to enquire. Some pages had contact email address links. Upon clicking such links, visitors were taken to an enquiry form. There were various pages from which visitors could complete and send an enquiry form. In order to understand the point at which visitors decided to enquire, the *[PageID]* of the page that led them to an enquiry form was recorded. The *[PageID]* was only recorded in cases where an enquiry form was completed and sent.

c) Enquiry quality score and marketing data retrieval

All email enquiries that could be cross referenced with the OTM database were given a quality score and a type. Marketing data was also retrieved for all enquiries recorded by the OTM and stored in MS CRM.

When an email enquiry was sent from the website using an enquiry form, links similar to the ones shown in Figure 4.6 were appended to it. A user could click on the appropriate link to score the quality of an enquiry as well as specify the type of an enquiry. Upon clicking on a link, the user was taken to a web page that displayed marketing data related to the enquiry that was being scored (see Appendix A Section A.3).

```
Design
GOOD: http://www.motiontouch.com/tools/score.asp?LeadType=1&LeadClass=1&sentID=7125
MEDIUM: http://www.motiontouch.com/tools/score.asp?LeadType=1&LeadClass=0&sentID=7125
BAD: http://www.motiontouch.com/tools/score.asp?LeadType=1&LeadClass=-1&sentID=7125

Manufacturing
GOOD: http://www.motiontouch.com/tools/score.asp?LeadType=2&LeadClass=1&sentID=7125
MEDIUM: http://www.motiontouch.com/tools/score.asp?LeadType=2&LeadClass=0&sentID=7125
BAD: http://www.motiontouch.com/tools/score.asp?LeadType=2&LeadClass=-1&sentID=7125
```

Figure 4.6: Quality scoring links.

d) Reporting module

Figure 4.7 shows the interface for a simple reporting module that was implemented to allow staff at the collaborating company to generate reports from the data collected by the OTM. It was implemented as an online form. SQL queries were generated based on the selection a user made using the form. The SQL queries were then issued to the CAT database and results were formatted and displayed.

Figure 4.8 shows an example of a report that was generated. The reports generated using the reporting module were mainly used to understand and improve the online advertising campaigns that generated traffic to the website.

4.4.4. Discussion

The new improved OTM provided better tracking and data recording than the first version. In particular, it was able to identify returning visitors and assign a unique *[MicroID]* to each returning visit. This provided new insight in the search and browsing behaviour of visitors, which was used during the data selection stage of the data mining process described in Chapter 7.

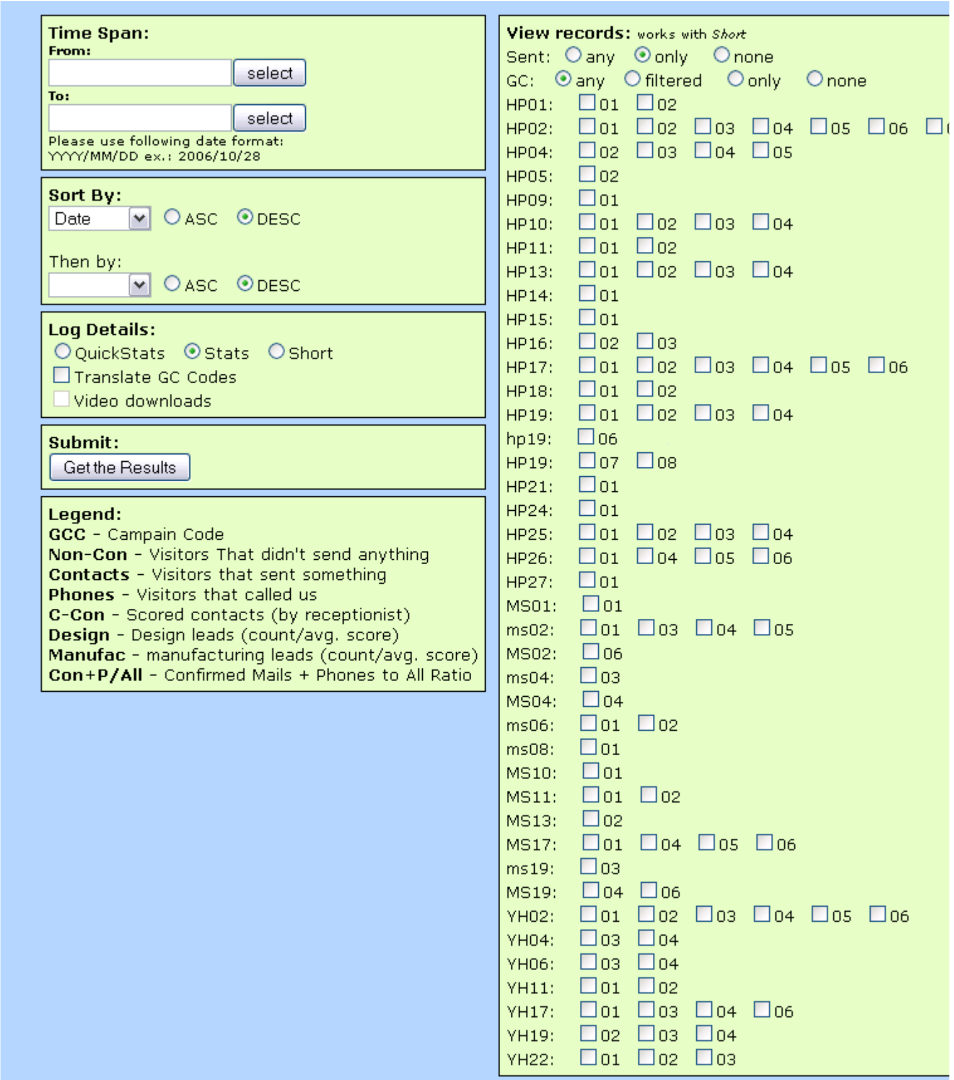
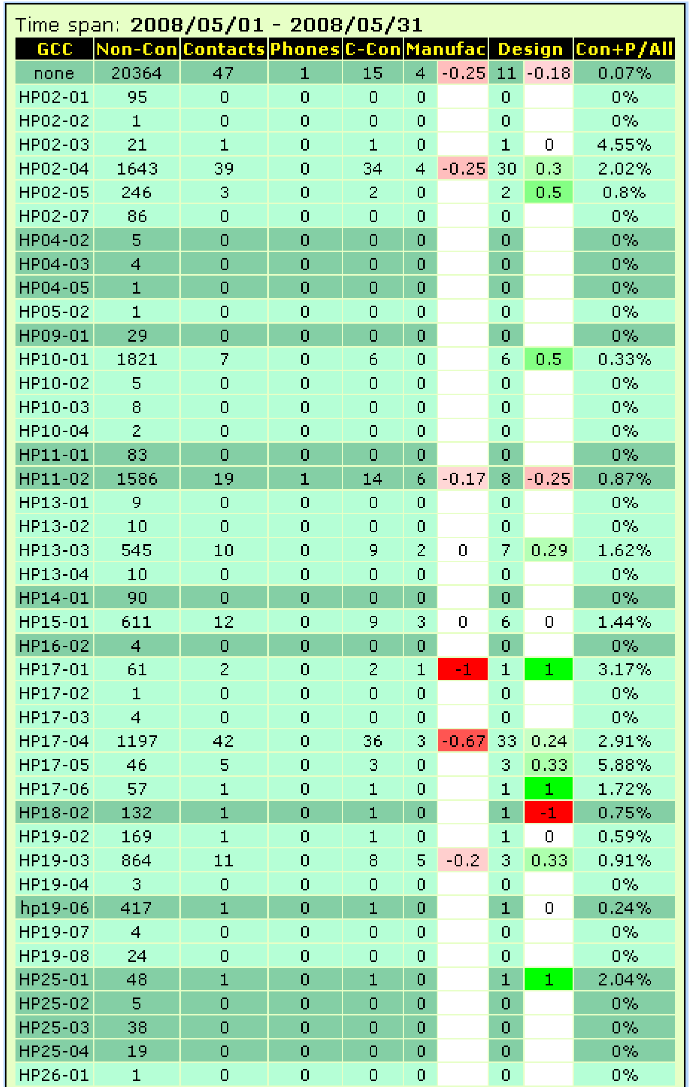


Figure 4.7: Reporting interface.



Time span: 2008/05/01 - 2008/05/31

GCC	Non-Con	Contacts	Phones	C-Con	Manufac		Design		Con+P/All
none	20364	47	1	15	4	-0.25	11	-0.18	0.07%
HP02-01	95	0	0	0	0		0		0%
HP02-02	1	0	0	0	0		0		0%
HP02-03	21	1	0	1	0		1	0	4.55%
HP02-04	1643	39	0	34	4	-0.25	30	0.3	2.02%
HP02-05	246	3	0	2	0		2	0.5	0.8%
HP02-07	86	0	0	0	0		0		0%
HP04-02	5	0	0	0	0		0		0%
HP04-03	4	0	0	0	0		0		0%
HP04-05	1	0	0	0	0		0		0%
HP05-02	1	0	0	0	0		0		0%
HP09-01	29	0	0	0	0		0		0%
HP10-01	1821	7	0	6	0		6	0.5	0.33%
HP10-02	5	0	0	0	0		0		0%
HP10-03	8	0	0	0	0		0		0%
HP10-04	2	0	0	0	0		0		0%
HP11-01	83	0	0	0	0		0		0%
HP11-02	1586	19	1	14	6	-0.17	8	-0.25	0.87%
HP13-01	9	0	0	0	0		0		0%
HP13-02	10	0	0	0	0		0		0%
HP13-03	545	10	0	9	2	0	7	0.29	1.62%
HP13-04	10	0	0	0	0		0		0%
HP14-01	90	0	0	0	0		0		0%
HP15-01	611	12	0	9	3	0	6	0	1.44%
HP16-02	4	0	0	0	0		0		0%
HP17-01	61	2	0	2	1	-1	1	1	3.17%
HP17-02	1	0	0	0	0		0		0%
HP17-03	4	0	0	0	0		0		0%
HP17-04	1197	42	0	36	3	-0.67	33	0.24	2.91%
HP17-05	46	5	0	3	0		3	0.33	5.88%
HP17-06	57	1	0	1	0		1	1	1.72%
HP18-02	132	1	0	1	0		1	-1	0.75%
HP19-02	169	1	0	1	0		1	0	0.59%
HP19-03	864	11	0	8	5	-0.2	3	0.33	0.91%
HP19-04	3	0	0	0	0		0		0%
hp19-06	417	1	0	1	0		1	0	0.24%
HP19-07	4	0	0	0	0		0		0%
HP19-08	24	0	0	0	0		0		0%
HP25-01	48	1	0	1	0		1	1	2.04%
HP25-02	5	0	0	0	0		0		0%
HP25-03	38	0	0	0	0		0		0%
HP25-04	19	0	0	0	0		0		0%
HP26-01	1	0	0	0	0		0		0%

Figure 4.8: Example report.

4.5. Stage 3 – Redesign of front-end of new dynamic main website

At the end of Stage 2, the front-end of the new dynamic main website was the same as that of the old static website. The design looked outdated. Therefore, the front-end was redesigned to have a professional and industry-appropriate look.

4.5.1. Redesign goals

The goals for the redesign of the front-end were to:

a) have a professional and industry-appropriate design.

b) improve the website's main menu.

c) improve and update the content of the website.

d) Use better quality pictures and graphics

4.5.2. Implementation

a) Website's look

The collaborating company's decided on a new visual design for the website. This design was then implemented using Cascading Style Sheets.

b) Main menu

In order to improve the navigation of the website, the main menu of the dynamic main website was changed. Its content was re-organised and categorised under new and more meaningful headings.

c) Content

New content for the dynamic main website was generated by the marketing department. When writing the textual content, the marketing staff followed some of the guidelines proposed by Neilsen (1997) to accommodate the different reading pattern of visitors.

d) Better graphics

The graphics and pictures used on the static main website were poor and sparse. Better graphics and pictures were included on every page on the dynamic main website. Good descriptive pictures were especially important as "*some visitors don't read at all; another reason to include at least one strong image*" (Loveday and Neihaus, 2008).

4.5.3. Discussion

The redesign of the front-end was successfully implemented. Landing pages would be further improved to increase the number of visitors who enquired. This is described in Chapter 6.

4.6. Chapter Discussion

The challenge for any online business is to turn website visitors into customers. A first step in achieving this is perhaps to gain knowledge about the type of visitors who come to a website as well as their needs.

One way to understand website visitors and their needs is to study their behaviour on a website and then infer motivation and need. This was what the research aimed to achieve through the implementation of the OTM. The research went a step further by cross referencing data between the OTM and MS CRM, so as to extract knowledge about visitors who become customers. The knowledge gained was used to improve online advertising campaigns and landing pages (described in Chapter 6).

A new dynamic main website was created to replace the old static main website. The new dynamic main website supported personalisation features. Chapter 6 describes how these features were used to design landing pages and improve the generation of enquiries.

The collaborating company relied on Pay Per Click (PPC) campaigns to generate traffic to its websites. The success of PPC campaigns relied not only on the ability of the campaigns to generate traffic but also on the ability of websites to convert the traffic into enquiries.

Chapter 5 describes the PPC model and explains how it was used to set up advertising campaigns that generated traffic to the websites described in this Chapter.

CHAPTER 5

PAY PER CLICK ADVERTISING

In order to attract customers, an online business needed a way of driving people to a website. This was an important part of any online business model without which it could not generate enquiries or online sales. There were several methods that could be used for gaining exposure online. One of the most popular methods was Pay Per Click (PPC) advertising. The Internet Advertising Bureau (IAB), reported that "*paid-for search continues to lead the way, growing by 28% year-on-year and was worth £981 million in the first half of 2008, with its market share marginally up to 58.3% of total online advertising*" (IAB UK, 2008).

This Chapter describes the PPC model and goes on to describe the Google AdWords model which was used during the research described in this dissertation to drive traffic to the websites described in Chapter 4.

5.1. Web search

Web search was key to the navigation and usage of the Internet (Fain and Pedersen, 2006). Most Web users relied on search engines to find what they were looking for on the Internet. The delivery of relevant results was an important part of the search experience. Sponsored search results which delivered highly targeted text advertisements could satisfy Web users' need for relevant search results. This also provided advertisers with high quality traffic to their websites. This win-win solution was perhaps why PPC advertising was so popular.

5.2. Pay Per Click (PPC) advertising

PPC advertising was also known as keyword search advertising (KSA), keyword advertising, or sponsored search advertising. PPC advertising was a form of "*text based online advertising*" (Burns, 2005) that used search engines to display text advertisements in response to the keywords that a Web user had typed into a search engine. The text advertisements were displayed alongside organic (non-paid) search results and were usually positioned at the top or on the right hand side of a search engine's results page. Advertisers could bid on keywords in response to which text advertisements were displayed. Advertisers were charged only when a Web user clicked on their advertisement. Upon clicking on an advertisement, Web users were taken to the advertiser's website. Search engines that offered PPC advertising included Yahoo, MSN and Google.

The history of PPC advertising as well as basic elements of the PPC advertising model were described in Chapter 2.

5.3. Google advertising

Google launched its first version of a PPC advertising service called Google AdWords in October 2000 with a total of 350 customers. Google AdWords was a self service advertising program where advertisers could set up and manage their advertising campaigns and that also provided performance feedback (Google, n.d.-f). Since its launch, Google AdWords has grown rapidly and has become one of the biggest PPC advertising service. The popularity of the Google search engine may have contributed to this achievement. In 2009, Google owned 82% of the global search engine market share (Netmarketshare, 2010). With such global reach, anyone advertising through Google AdWords was likely to reach a large number of potential customers.

The research described in this dissertation relied on Google AdWords to target customers and generate traffic to the websites described in Chapter 4. The campaigns and advertisements that were set up in Google AdWords to achieve this are described in Section 5.10.

The Google AdWords PPC model was based on the sponsored search model described in Chapter 2. Google AdWords had an online interface which allowed advertisers to create and manage advertising campaigns. The structure of a Google campaign is shown in Figure 5.1.

5.4. Google AdWords campaign

A Google AdWords campaign, "*uses keyword targeting or placement targeting to put advertisements on search results and content network placements across websites and other online content*" (Google, n.d.-d). A Google AdWords campaign had a number of settings that gave advertisers control over the campaign. Some of these setting included:

- Location targeting. This allowed an advertiser to specify the countries in which campaigns ran, that is countries in which text advertisements were displayed.
- Language targeting. Advertisers specified whether campaign advertisements were only displayed to users who communicated in a particular language.

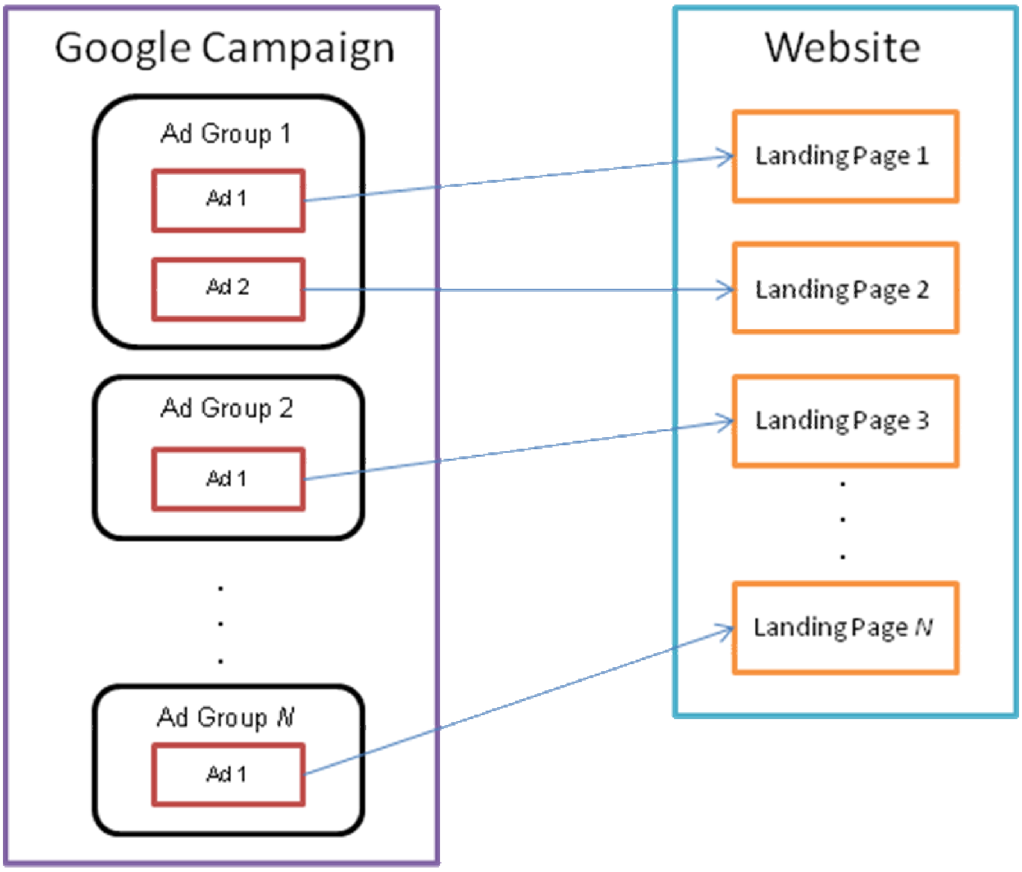


Figure 5.1: Google campaign structure.

- Networks. Where advertisements appeared. Google had "*a large group of websites and other products, such as email programs and blogs, who [had] partnered with Google to display AdWords ads*" (Google, n.d.-h). This was called the Google Network and was composed of a Search Network and a Display Network. Google AdWords advertisements could be placed based either on searches (Search Network) or website content (Display Network).
- Devices. The type of device advertisements appeared on, for example desktop and laptop computers or mobile devices such as iphones.
- Daily budget. The maximum spend for a campaign per day. Advertisements were not displayed once a daily budget had been spent.

- Advertisement scheduling. The advertiser could specify the hours or days on which advertisements were displayed.
- Advertisement delivery. Whether to display advertisements more evenly by rotating them or to optimise display by showing the better performing advertisements more often.

The settings of a campaign were inherited by all its Ad Groups. Ad Groups could not override these settings. Within a Google AdWords account, separate campaigns could be created to organise product lines and services. A Google AdWords account could have a maximum of 25 campaigns.

5.5. Google Ad Group

A Google Ad Group contained keywords and text advertisements associated with those keywords. Usually, an Ad Group focused on advertising a specific product or service. A campaign could have a maximum of 2000 Ad Groups. Each Ad Group could have up to 2000 keywords and 50 advertisements.

5.6. Keywords

Keywords were the terms or phrases that triggered an advertisement to appear in search results. Keywords had a maximum cost per click (CPC), which was defined by Google (n.d.-d) as "*the highest amount that [an advertiser] is willing to pay for a click on [their] ad*." The maximum CPC was set at keyword level. Google AdWords calculated an average position for each keyword. This was the position at which a text advertisement was displayed for a particular keyword. The average position depended on the maximum CPC of the keyword.

Keywords also had matching options that helped determine when advertisements were displayed in response to a search. Table 5.1 shows how the different matching options worked. Matching options were set on individual keywords using the punctuation shown below. The matching options were (Google, n.d.-d):

- Broad match: keyword *(no punctuation)*

  Showed advertisements for searches on similar phrases and relevant variations.

- Phrase match: "keyword"

  Showed advertisements for searches that matched the exact phrase.

- Exact match: [keyword]

  Showed advertisements for searches that matched the exact phrase exclusively.

- Negative match: -keyword

  Did not show advertisements for any search that included that term.

If a matching option was not specified, a keyword was set as a broad match by default. Some matching options provided greater advertisement exposure and broader targeting, for example broad match. Other matching options reached a smaller but more targeted audience, for example exact match. Upper-case and lower-case letters were ignored when keywords were matched.

The Ad Groups that were set up for this research used all four matching types. In order to work out which combination of keyword and matching option performed better, all keywords were configured with broad, phrase and exact match when a campaign was created. After a period of time, the performance of the keywords was reviewed and only combinations that performed well remained active.

Matching option	Advertisements may show on searches for	Advertisements will not show on searches for
Broad match e.g. Plastic manufacturer	plastic manufacturer plastic manufacturer in uk find plastic manufacturers plastic factory	
Phrase match e.g. “plastic manufacturer”	plastic manufacturer uk find plastic manufacturer	plastic toy manufacturer plastic manufacturers
Exact match e.g. [plastic manufacturer]	plastic manufacturer	plastic manufacturer uk find plastic manufacturer plastic manufacturers
Negative match e.g. plastic manufacturer -acrylic	plastic manufacturer uk find plastic manufacturer plastic manufacturers	acrylic plastic manufacturer manufacturer of acrylic plastic

Table 5.1: How matching options work.

5.7. Advertisements

Google AdWords offered 4 types of advertisement. They were:

- Text advertisements.
- Image advertisements.
- Mobile advertisements.
- Video advertisements.

The research described in this dissertation used text advertisements only to drive traffic to websites. Figure 5.2 shows an example text advertisement.

Figure 5.2: Example text advert (Google, n.d.-b).

The main components of a Google AdWords text advertisement were (Google, n.d.-b):

- Headline. The first line of a text advertisement was a link to a website. It was good practice to insert a keyword in the headline that related to keywords being searched.
- Lines of text. There were two lines of text that were available to convey a sales message. Each line of text was set to allow for a maximum number of characters.
- Display Uniform Resource Locator (URL). The last line in a text advertisement displayed the URL of the website that was being promoted. This URL was not the full URL of the destination page within the website.
- Destination URL. The destination URL was the exact URL of the page that Web users would be taken to when they clicked on an advertisement. The destination URL was not displayed in advertisements but had to be specified when creating an advertisement.

Figure 5.3 shows where advertisements were displayed on the Google search results page.

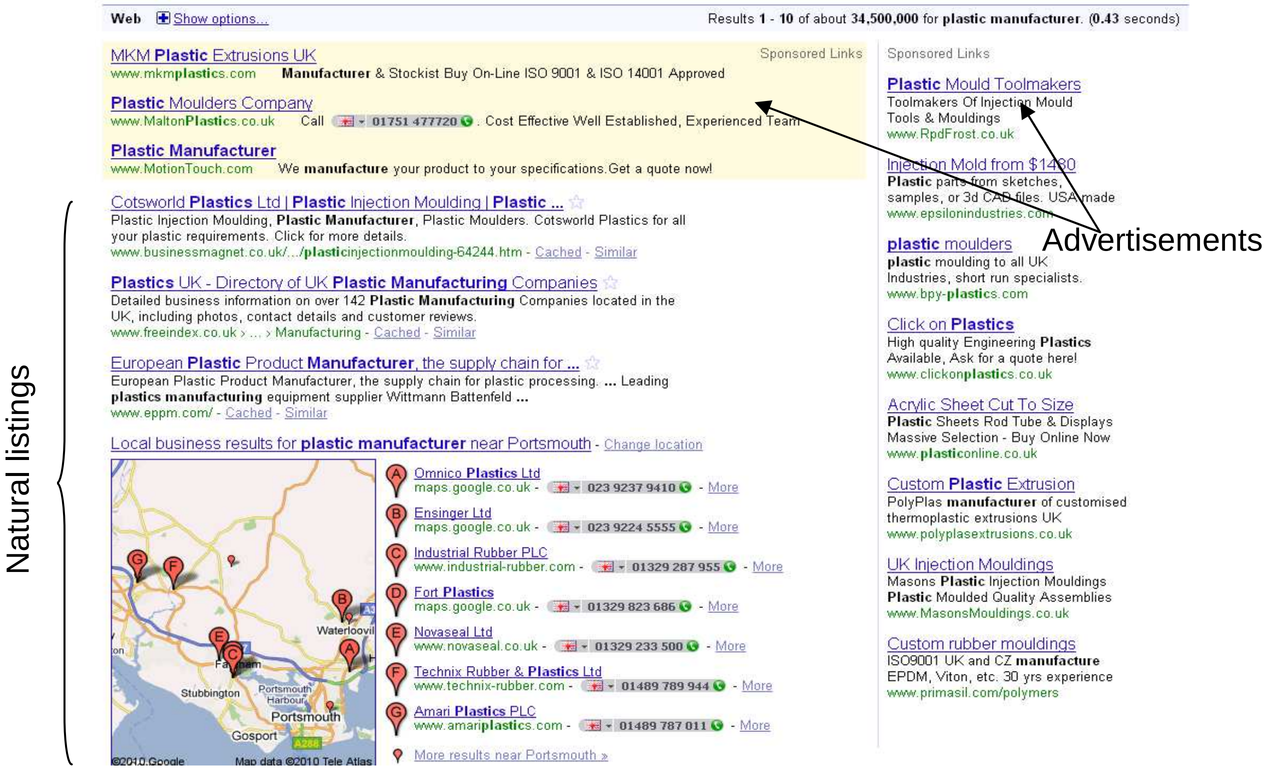


Figure 5.3: Google search results page.

The position of an advertisement depended on its AdRank. The advertisement with the highest AdRank appeared first. Google AdWords calculated AdRank using individual keyword's Quality Score and cost-per-click (CPC).

Quality Score was a measure that Google AdWords used to assess the relevance of advertisements, keywords and landing pages compared to a Web user's search keywords. A Quality Score was assigned to each advertisement and keyword found in an AdGroup. It was calculated using a variety of factors. Some of these factors were (Google, n.d.-c):

- The historical click-through rate (CTR) of a keyword.
- The historical CTR of all the advertisements and keywords in a Google account.
- The quality of a landing page that is, the page on the destination website that an advertisement was associated with.

Quality Score could have a big impact on the performance of Google AdWords campaigns. It affected:

- Costs. A high keyword Quality Score meant a lower cost-per-click.
- Whether advertisements were eligible for display. Keywords with a higher Quality Score performed better in the PPC auction that determined whether an advertisement was displayed.
- Advertisement position. An advertisement's position on a page depended the Quality Score and cost-per-click of the keywords that were associated with it.

5.7.1. Advertisement targeting

Advertisement targeting was an important factor in running successful Google AdWords campaigns. Advertisements were targeted using:

a) Keyword targeting.
b) Location and language.
c) Placement options on the Google Network.

a) Keyword targeting

Keyword targeting was the primary method for reaching potential customers. By creating a keyword list that was highly relevant to the type of audience that this research wanted to reach, advertisements could be targeted to the chosen audience. Building an effective keyword list for the campaigns that were created in this research was a process that took time. The process started with the compilation of a list of keywords that included as many relevant keywords as possible. These were then reviewed after a few weeks of being active and the list of keywords was refined by eliminating keywords that did not generate enquiries. The review process also enabled the identification of keywords that generated a high number of enquiries. By reviewing and refining a keyword list regularly, a targeted keyword list could be generated.

b) Targeting by location and language

For each Google AdWords campaign the language of the targeted audience could be configured. For example, advertisements could be displayed in Europe but to an English speaking audience only. Advertisers could also specify the location of the targeted audience. For example a campaign created during this research was targeted at the East coast region of the United States of America (US) while another was targeted at the whole of the US.

AdWords used the location and language settings to determine who saw advertisements. AdWords used the following factors to determine this:

- Web users' Google domain for example www.google.fr or www.google.co.uk.
- Web users' search keyword.
- Web users' IP address was used to determine geographical location.
- Language preference that a Web user had set for Google.

c) The Google Network

Each Google AdWords campaign enabled an advertiser to choose a network and device setting for the advertisements in that campaign. This was described in Section 5.4.

This research used a combination of the above targeting methods to create advertisements that could reach targeted audiences.

5.8. Landing Page

A landing page was a web page that a Web user was taken to after they clicked on an advertisement. Google AdWords assessed the page quality of all landing pages specified in advertisements. The landing page quality was determined by the "*usefulness and relevance of information provided on the page, ease of navigation for the user, page loading times, how many links are on the page, how links are used on the page and more*" (Google, n.d.-a). Google used the landing page quality to calculate the keyword quality score for all keywords that were associated with a landing page. Good landing page design was therefore important to achieve high keyword quality scores that would enable advertisements to reach high positions on search engine results pages.

Landing pages usually had a unique and distinct function from other pages in a website. They played an important role in converting website visitors into customers. Landing pages had to be specifically designed to support a marketing campaign. As such there were unique challenges that needed to be met when designing them.

Loveday and Neihaus (2008) found that landing pages faced the following issues:

- They had to singlehandedly take visitors through the whole sales cycle. They had to "*create or reinforce interest, then instill desire, and finally guide visitors to take action*".
- They had to perform quickly.
- They had to deal with a high number of first time visitors, who were not "*familiar with the company and [had] no reason to trust it at first*".

The landing pages used in this research went through an optimisation process which is described in Chapter 6.

5.9. AdWords performance metrics

The Google AdWords system provided a number of performance measures that could be used to evaluate the effectiveness of advertising campaigns, keywords and advertisements. Some of these measures were used during this research to measure the effect of changes made to advertising campaigns and to landing pages. Google AdWords metrics that were used during this research included:

- Impression - This was the number of times an advertisement was displayed on Google's search results page or the Google content network.
- Click - When a Web user saw an advertisement and clicked on it Google AdWords recorded this as a click. The number of clicks roughly represented the number of visitors to a website. The number of clicks recorded by Google AdWords was not unique, for example it sometimes recorded multiple clicks from the same Web user. However, Google AdWords monitored all clicks to ensure that there was no abuse. This included analysing clicks to see if they fitted a pattern of fraudulent use. Google's proprietary technology could distinguish between clicks from normal Web users and clicks generated by spammers and automated robots. Google claimed that it could filter out such clicks and that they did not show up on reports.
- Click through rate (CTR) - This was the number of clicks an advertisement received divided by the number of times the advertisement was displayed (impressions)

- Conversion - A conversion was the completion of a unique goal or action on a website for example sending an email, buying a product, signing up for a service or downloading a product. In the context of this research, the action of sending an email enquiry from the collaborating company's website was a conversion.
- Cost per conversion - This was the total cost of a campaign divided by the number of conversions that it had generated.
- Conversion rate – The conversion rate was the number of conversions divided by the number of clicks that a campaign or Ad Group or advertisement had received.

These metrics were available for campaigns, Ad Groups, keywords and advertisements. This enabled performance to be measured at a various levels.

5.9.1. Google AdWords Conversion Tracking

Google's AdWords Conversion Tracking was a tool that allowed advertisers to track and measure conversions. For tracking to take place advertisers had to place the AdWords Conversion Tracking code on their website. Depending on what represented a conversion, advertisers chose different conversion confirmation pages. Conversion confirmation pages were pages that users were sent to after completing a unique action for example a **Thank You** page following purchase or sign-up.

Once the code was placed on the appropriate pages, the AdWords Conversion Tracking worked by placing a cookie onto a visitor's computer or mobile phone when they clicked on an advertisement. If the visitor reached a conversion confirmation page, the cookie was connected to that page and AdWords recorded a conversion. The cookie that Google AdWords placed on a visitor's computer or mobile device expired after approximately 30 days.

AdWords Conversion Tracking could record 1-per-click and many-per-click conversions. 1-per-click conversion meant that if "*one click led to multiple conversions, they were counted only once*" (Mutlu, 2009). In the case of many-per-click conversion, the tracking system counted each conversion that occurred after a Web user clicked on an advertisement.

For the purposes of this research, only 1-per-click conversion was used since unique enquiries were of interest. Many-per-click conversion was better suited for measuring online purchases where a Web user clicked on an advertisement once but purchased more than one product.

For the websites used in this research the AdWords Conversion Tracking code was placed on **Thank You** pages. Visitors were automatically directed to this page after they completed and sent an enquiry form.

5.10. Google AdWords configuration

The research described in this dissertation created a number of AdWords campaigns in order to drive traffic to websites. Some of these campaigns are described in the following sections.

5.10.1. Campaign types

The collaborating company provided two types of services: design and manufacturing. These services were advertised in the United Kingdom (UK) and United States (US). Each service had an advertising campaign and each campaign was duplicated and run in the US and the UK.

Each campaign had several of Ad Groups. The Ad Groups were created around the different sub-types of the design and manufacturing services for example plastic moulding, plastic manufacturing, mechanical design and electronic design.

At the beginning of this research, there were two Google AdWords accounts. One ran campaigns for the main site and the other ran campaigns for the micro sites. The campaigns were set up in this way because they were bidding on the same keywords but the advertisements led to different websites. Only one advertisement from a Google AdWords account could be shown on the Google search results page at any given time. By having separate accounts, it was possible to have two advertisements appear for the same keyword but leading to different sites. This set up ran from 2005 to 2007.

This research also used MSN's and Yahoo!'s search marketing service. However, these campaigns were not cost effective and generated few, poor quality enquiries. As a result these campaigns were suspended in 2008.

In order to track which advertisements, Ad Groups and campaigns generated traffic, each of them was given unique identifier called a campaign code. The destination URL of all advertisements was appended with a campaign code corresponding to their Ad Group. The Online Tracking Module described in Chapter 4 used this code to track which Google AdWords campaign visitors came from. This code was also used to dynamically configure landing pages (described in Chapter 6) and menus in order to provide better content personalisation (described in Chapter 4). The campaign code was recorded in MS CRM for those visitors who enquired using enquiry forms found on the websites.

5.10.2. Core campaigns

Over the course of this research, some of the campaigns that were created became core campaigns as they generated the majority of enquiries. There were also several small campaigns that were targeted at smaller audiences that generated fewer leads. The core campaigns were the:

a) Inventor campaign.

b) Manufacturing campaign.

a) Inventor Campaign

The Inventor campaign advertised the collaborating company's design services. It was targeted at individuals and inventors who had new product ideas that they wanted to develop. The matching options used for the keywords in the two main Google AdWords campaigns that ran in the United Kingdom can be found in Appendix B (and at the beginning of this Chapter). Keywords included:

- Invention
- Inventor
- product idea
- product development

The advertisements used for this campaign are shown in Figure 5.4. Both advertisements had the same destination URL. Over the course of the research described in this dissertation, the landing page for the advertisements was changed several times. Figures 5.5, 5.6 and 5.7 show the last three landing pages for the UK campaign.

Figure 5.4: Inventor campaign advertisements.

The screenshot in Figure 5.5 shows the landing page used for the campaign from May 2007 to June 2008. The landing page was changed to the one shown in Figure 5.6 in July 2008. This landing page was active until the end of October 2008 when it was replaced by the one shown in Figure 5.7.

Figure 5.5: Landing page for the inventor campaign from May 2007 to June 2008.

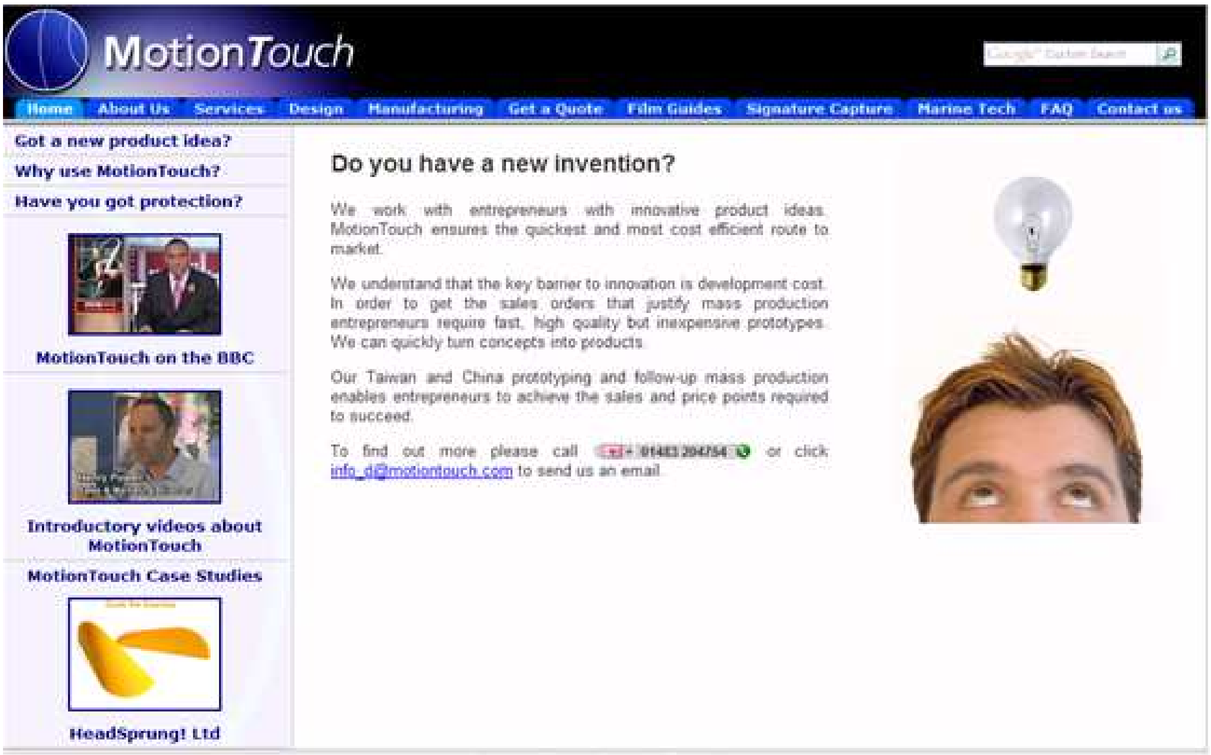


Figure 5.6: Landing page between July 2008 and October 2008.

Figure 5.7: Landing page from November 2008 to present.

The US version of this campaign was paused in February 2009 due to cuts in advertising budget. This campaign used the same landing pages as the UK campaign. The landing pages followed a similar progression as those of the UK campaign. However, the last landing page of the US campaign (active from October 2008 to February 2009) was different to that of the UK campaign. This landing page is shown in Figure 5.8.

Figure 5.8: Landing page for US campaign from October 2008 to February 2009.

b) Manufacturing Campaign

The Manufacturing campaign advertised the collaborating company's manufacturing services. It was targeted at individuals and SMEs who were looking for plastic manufacturing services. Keywords included:

- Plastic manufacturer

- Plastic manufacturing
- Plastic product

This campaign ran both in the UK and in the US. A number of different advertisements and landing page combinations were used to achieve high conversion rates. Some of these landing pages are described in Chapter 6.

5.10.3. Running advertisements in parallel

In order to work out the most effective combination of landing page and advertisement, several advertisements were run at the same time. The effectiveness of an advertisement was measured by its Click-Through Rate (CTR). To find the best advertisement, several advertisements were created, each with a different message. The campaign containing the advertisements was then configured to serve the advertisements evenly by rotating them. These advertisements were run for a minimum of two weeks before comparing their CTR to find the one that had performed best.

Once the advertisement with the highest CTR was found, then landing pages were tested. The advertisement was duplicated and each copy was given a destination URL corresponding to each landing page that needed to be tested. Once again the campaign containing the advertisements was configured to serve them evenly by rotating them. The advertisements were run for a minimum of two weeks. At the end of the experiment, the conversion rates of the advertisements were compared to find the best landing page.

There were a number of external factors that affected the performance of campaigns and landing pages, for example seasonal trends and the recession that started in 2008 (BBC News, 2009). It was difficult to attribute the changes in performance to the changes made to the campaigns or landing pages unless there were control experiments also running. By comparing performance against that of the control group, the influence of external factors could be minimised.

It was not possible to run control groups as the collaborating company only ran high performing advertisements and landing pages. Control groups could be expensive to run and could take some of the traffic away from the better performing advertisements and landing pages.

5.11. Chapter Discussion

This Chapter described the PPC model, in particular the Google AdWords model used during this research to generate traffic to the websites described in Chapter 4. The various configurations used to target advertisements at specific audiences via language, location and keywords were described in section 5.7. The metrics provided by Google AdWords to measure the performance of campaigns, Ad Groups, keywords and advertisements were described in section 5.9. These performance measures were used to monitor the performance of the advertising campaigns that were created during this research.

Google AdWords campaigns were successfully setup to drive targeted traffic to the websites described in Chapter 4. The next step in this research was to optimise landing pages so as to encourage website visitors that PPC campaigns had attracted, to convert into customer by enquiring. Chapter 6 describes how landing page optimisation was carried out.

CHAPTER 6

LANDING PAGE OPTIMISATION

Chapter 5 described how Google AdWords was used to generate traffic to the websites described in Chapter 4. In order to convert traffic generated by online advertising campaigns into enquiries, these websites needed to be improved to have higher conversion rates. Landing page optimisation (LPO) was especially important as the landing page was the page that visitors landed on after clicking on an advertisement.

This Chapter describes some of the techniques for improving landing pages that were found in literature. These techniques led to changes to the landing pages of the two core advertising campaigns (described in Chapter 5) that ran in the United Kingdom (UK). The changes to the landing pages that are described in this Chapter took place on the collaborating company's main website.

This Chapter also describes the experiments that were set up to determine how the changes that were made affected the conversion rate and bounce rate of landing pages. In order to determine whether the changes made to a landing page had led to an improvement, the conversion rate and the bounce rate (in some cases) of the landing page before and after the changes was compared. The statistical significance of the differences observed in conversion rates or bounce rates was calculated using statistical techniques such as chi-square test, confidence interval and t-test.

The results of the experiments that were carried out on the landing pages are presented and discussed in this Chapter. The literature search carried out during this research identified a gap in knowledge whereby existing research literature did not present results to show how the application of certain landing page design techniques affected the conversion rate or bounce rate of a landing page. This Chapter presents data that showed the effect that some landing page design techniques had on conversion rate and bounce rate. Such results have not been previously available in research literature.

6.1. Understanding a website's audience

Ash (2008) claimed that "*before you can even look at the specific issues and problems with your landing page, you must try to see it through the eyes of your audience*". Therefore, it was important to understand who website visitors were, where they came from and what their intent was.

This research looked for meaningful commonalities in the data collected by MS CRM and the OTM and combined this with feedback from the sales team to create two personas for website visitors who had enquired. They were:

- The Inventor. This was an individual who had an idea for an invention that they wanted to develop into a product. Usually, they did not have previous experience in product development and were looking for a company that could guide them through the process. An inventor had limited or sometimes no budget. Inventors used keywords such as "invention", "product idea", "invention idea".
- The Manufacturer. This was usually a person from a small to medium company that wanted to manufacture a product in large quantities. Their main goal was to get a quote. They were goal-oriented, time conscious and had a budget. They used keywords that were variations of the word "plastic manufacturer"

The keywords that were frequently used by each persona were identified from data that was stored in MS CRM and the OTM. The other characteristics of the personas were identified by the sales team and were based on their interaction with customers.

6.2. Landing page optimisation

6.2.1. Changes to landing pages

This research focused on optimising the landing pages for the core UK advertising campaigns described in Chapter 5 and for the two visitor personas that had been identified. Table 6.1 provides a summary of the changes that were made to the landing pages of the core campaigns during the landing page optimisation process.

Change	Description	Applied to
Change 1	Changed landing page to provide more targeted content.	Inventor campaign
Change 2	Included keywords from advertisement in title of landing page.	Inventor campaign
Change 3	Improved visual design.	All pages of the new dynamic main website
Change 4	Changed style in which content was written and laid out.	Manufacturing campaign
Change 5a	Implemented user directed segmentation using pictures.	Manufacturing campaign
Change 5b	Implemented user directed segmentation using a questionnaire.	Manufacturing campaign

Table 6.1: Changes to landing pages.

These changes are described in the following sections. Performance measures that were used to assess the impact of the changes made to landing pages are also presented. Detailed performance measures can be found in Appendix E.

6.2.2. Performance measures

The performance of landing pages was measured using conversion rate and bounce rate.

The conversion rate of a landing page was calculated by dividing the number of conversions that it had generated by the number of views that it had received. Data collected by Google AdWords was used to calculate conversion rate.

The bounce rate of a landing page was calculated by dividing the number of bounces that it generated by the number of visits that it received. Data collected by Google Analytics was used to calculate bounce rate.

The number of clicks (or views) that was recorded by Google AdWords was different from the number of visits recorded by Google Analytics. This was because Google AdWords tracked clicks, while Google Analytics tracked visits. Clicks represented the number of times that an advertisement was clicked by visitors, while visits indicated the number of unique sessions initiated by visitors. For example, if a Web user clicked on an advertisement twice within thirty minutes without closing their browser, this was registered by Google Analytics as one visit while Google AdWords recorded two clicks (Google, n.d.-g).

6.2.3. Statistical indicators

In order to determine the statistical significance of the results that were obtained during LPO, the following statistical indicators were used:

a) Pearson chi-square test.

b) confidence interval.

c) t-test.

a) Pearson chi-square test

The Pearson chi-square test (commonly referred to as chi-square test) was used when the data being analysed was “*nominal in nature i.e. when it is recorded as frequencies in discrete categories”* (Jones, 2002). Since the data collected from the experiments described in this Chapter was nominal the chi-square test was used to determine whether there was a correlation between the frequencies associated with the rows and those recorded in the columns that is, it was used to test the null hypothesis and to determine the statistical significance of results.

The chi-square test compared the tallies or counts of categorical responses between two (or more) independent groups. Equation 6.1 shows how the chi-square statistic for a 2 x 2 contingency table (Table 6.2) was calculated (The Mathbeans Project, n.d.).

$$X^2 = \frac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(b+d)(a+c)}$$

Equation 6.1: Chi-square.

	Data Type 1	**Data Type 2**	**Total**
Category 1	a	b	a + b
Category 2	c	d	c + d
Total	a + c	b + d	a + b + c + d = N

Table 6.2: Example 2 x 2 contingency table.

The chi-square statistic (X^2) was used to calculate a p-value by comparing the value of the statistic to a chi-squared distribution for a number of degrees of freedom (DF). Equation 6.2 shows how the number of degrees of freedom was calculated. r was the number of rows and c was the number of columns in a contingency table.

$$DoF = (r-1)(c-1)$$

Equation 6.2: Degrees of freedom.

The p-value indicated the statistical significance of data presented in a contingency table. This research used a significance level of 0.05. Therefore, if a p-value was smaller than 0.05, then test results were deemed to be statistically significant.

This research used two free online tool, one provided by the Mathbeans Project (The Mathbeans Project, n.d.) and another by GraphPad (GraphPad, n.d.-a) to carry out chi-square tests and calculate p-value for data presented in this Chapter.

b) Confidence interval

The research used confidence interval to confirm the conclusions drawn from chi-square tests. If the p-value calculated by a chi-square test was greater than 0.05 then, it was concluded that the results obtained were not statistically significant. This meant that the null hypothesis could not be rejected but did not mean that the null hypothesis was true. Motulsky (1995) suggested that when a high p-value indicated results that were not significant, confidence interval (CI) could be used to evaluate the study.

A confidence interval (CI) was "*a range of values for a variable of interest constructed so that this range has a specified probability of including the true value of the variable. The specified probability is called the confidence level, and the end points of the confidence interval are called the confidence limits. It is conventional to create confidence intervals at the 95% level*" (Davies and Crombie, 2009).

When evaluating CI on the difference between proportions, if the confidence interval included the value reflecting no-effect (which was 1), then the difference that was observed was statistically non-significant (for a 95% confidence interval, non-significance was at the 5% level). If the confidence interval did not enclose the value reflecting no-effect, then the difference that was observed was statistically significant. Equation 6.3 shows how the CI on the difference between proportions was evaluated.

$$CI = RR(e^{\mp 1.96\sqrt{v}})$$

Equation 6.3: Confidence interval (Winner, 2011).

Equation 6.4 and Equation 6.5 show how RR and v were calculated (using Table 6.2 as example):

$$RR = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

Equation 6.4: Relative Risk (Winner, 2011).

$$v = \frac{(1 - \frac{a}{a+b})}{a} + \frac{(1 - \frac{c}{c+d})}{c}$$

Equation 6.5: v (Winner, 2011)

c) t-test

The t-test was a commonly used method for evaluating the differences in means between two groups. A t-test produced a p-value that represented the probability of error regarding the existence of a difference. In other words the p-value indicated whether the difference in means was statistically significant. If the difference was in a predicted direction, then only one half (one "tail") of the probability distribution was considered. In this case, the standard p-value reported by a t-test (a "two-tailed" probability) was divided by two (StatSoft, 2010).

This research used GraphPad (GraphPad, n.d.-b), which was a free online tool to carry out t-tests.

6.3. Change 1 – Targeted landing page

6.3.1. Concept

Understanding the audience of a website was an important part of LPO (Ash, 2008, Loveday and Neihaus, 2008). When designing landing pages *“it is important to keep in mind who your audience is and [to] make sure that the information you provide is relevant to them”* (Mason, 2007). Thompson (2009) suggested that using content to create targeted landing pages for different audience segments could help keep visitors on a website longer and lead to more conversion. He claimed that the more targeted a landing page’s content was towards a niche market, the more likely it was to convert visitors.

6.3.2. Concept application and landing page testing

At the beginning of this research, the Inventor campaign described in Chapter 5 had the **Design Overview** page as landing page. The Inventor campaign ran two text advertisements that had different headlines but the same text, display URL and landing page (the **Design Overview** page). Figure 6.1 shows the **Design Overview** page and the advertisements that used it as landing page. The Inventor campaign mostly attracted visitors who fitted the Inventor persona.

Advertisement delivery for the Inventor campaign was configured so that the two advertisements were displayed evenly by rotating them. Google AdWords determined which advertisement to display in response to a search by a Web user. Each advertisement received approximately the same number of impressions but the number of clicks that each received depended on Web users’ preference.

The content of the **Design Overview** page was focused on the design services that the collaborating company offered and was not targeted at the Inventor persona. Therefore, a new landing page called **Product Idea** (see Figure 6.2) was tested against the **Design Overview** page. The advertisements shown in Figure 6.1 were duplicated and configured to have the **Product Idea** page as landing page. The duplicated advertisements co-existed with the original advertisements in the Inventor campaign.

Figure 6.1: Design Overview page and advertisements that were run by the Inventor campaign.

The content of the **Product Idea** page was relevant to the Inventor persona because:

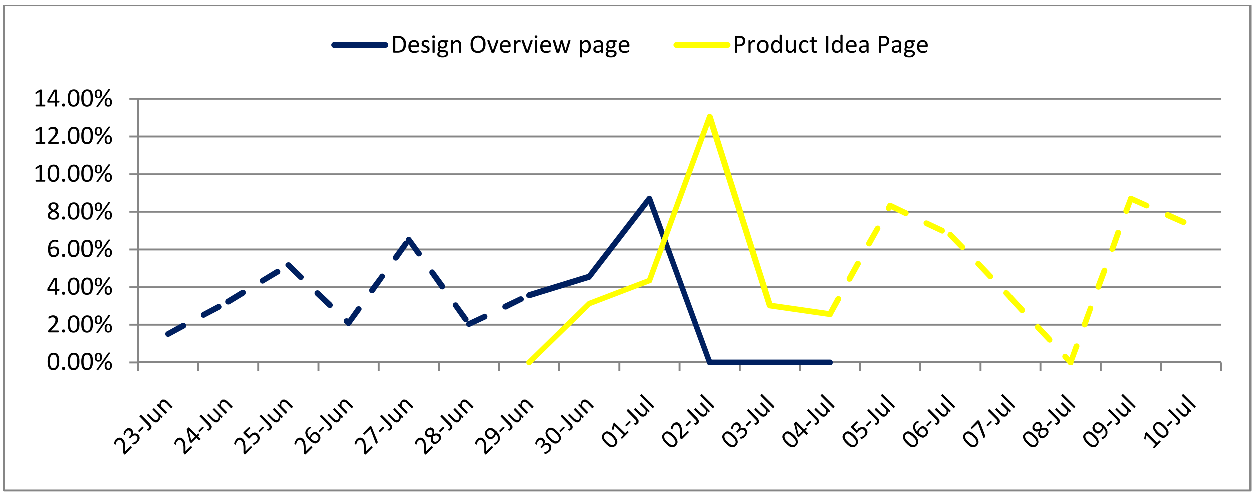
- of the heading of the page.
- the first paragraph of the text specified that the company worked with individuals and could turn their idea into a product.

- the picture on the page showed a person having a light bulb moment which was evocative of an inventor coming up with an idea.

The **Product Idea** page had clear calls of action at the end of the page to encourage visitors to make an enquiry by either calling the sales line or by email. The **Design Overview** page did not have any calls of actions. The content of the **Product Idea** page was written in short easy to read paragraphs while the content of the **Design Overview** page was written in longer paragraphs which could have required more effort to read and looked less appealing to a visitor. The **Design Overview** page displayed a list of links at the bottom of the page while the **Product Idea** page did not.

Figure 6.2: Duplicated advertisements used the Product Idea page as landing page.

The two landing pages ran in parallel for 6 days. Conversion rates were recorded by Google AdWords and used to compare the performance of the two landing pages.



Graph 6.1: Conversion rate for Design Overview page and Product Idea page before and after testing.

Graph 6.1 shows how the conversion rate of the two landing pages varied during the test. The solid lines show the conversion rate during the testing period. The dotted lines show the performance of the pages before and after the test. It can be seen that at the beginning of the test the **Design Overview** page performed better that the **Product Idea** page. The conversion rate of the **Product Idea** page gradually increased and peaked at 13.04%.The conversion rates for the landing pages at the end of the test are shown in Table 6.3.

Landing page	Conversions	Non-Conversions	Views	Conversion Rate
Product Idea	7	145	152	4.61%
Design Overview	5	139	144	3.47%

Table 6.3: Conversion rates for Design Overview page and Product Idea page.

A chi-square test was used to determine the statistical significance of the results shown in Table 6.3. The chi-square test yielded a p-value of 0.62. A 95% Confidence Interval

(CI) ran from 0.43 to 4.08. The CI was wide and included 1 which confirmed that the results were not statistically significant.

6.3.3. Conclusion

It was concluded that the increase in conversion rate observed for the **Product Idea** page was not statistically significant. Because the test was carried out over a short period of time and collected a relatively small amount of data, it might not have detected the effect of the **Product Idea** page. The **Product Idea** page remained active and was used to carry out new changes and tests that are described in the next section.

6.4. Change 2 - Include keywords from advertisement in title of landing page.

6.4.1. Concept

Loveday and Neihaus (2008) suggested that in order to provide a seamless and consistent experience to visitors, landing pages had to be an extension of their advertisements. They proposed the following methods to achieve this:

- Ensuring that a landing page provided what the corresponding advertisement promised. This had to be explicit and obvious. If an advertisement offered design services then the content of the landing page had to be about design services. Otherwise visitors could think that they were victim of the "*bait and switch*" trick and leave with a bad impression of the website.

- Matching the wording of an advertisement on its landing page. Since visitors decided whether a page was relevant to their search within seconds (Lindgaard et al., 2006), including the wording found in an advertisement on its landing page could help to convince visitors that they had found a suitable website, thus encouraging them to stay and browse.

- Maintaining the language and tone of the advertisement on the landing page, for example if the advertisement used intellectual language then the landing page had to do the same.

Maintaining consistency between an advertisement and its landing page drew on the idea of mirroring. The mirroring principle suggested "*that the efficacy of tools or messages can be maximised by ensuring that they contain features that mirror the preferences of the target market*" (Moss et al., 2008).

The mirroring principle had been used in various fields and translated into different views. Two views that were relevant to website design were the communication field and social psychology. In the communication field, mirroring represented the notion that persuasiveness is enhanced by the similarity between source and receiver (Moss et al., 2008, Brock, 1965). In social psychology, it translates to "*the 'matching hypothesis' or 'similarity-attraction' which implied that increased similarity leads to increased attention and attraction*" (Moss et al., 2008).

6.4.2. Concept application and landing page testing

This research assumed that when a Web user clicked on a particular advertisement out of a list of advertisements, they were expressing a preference towards the advertisement's message. Therefore, maintaining features of the advertisement on its landing page could have a positive impact on conversion rate. This idea was tested on the **Product Idea** page that was introduced during Change 1 (described in Section

6.3.2). Figure 6.3 shows the advertisements (referred to as Advertisement 1 and Advertisement 2) that used the **Product Idea** page as landing page.

Advertisement 1 **Advertisement 2**

Figure 6.3: Product Idea page as landing page for Advertisement 1 and Advertisement 2

The title of the **Product Idea** page included the keyword ("product idea") from Advertisement 1 only. The two advertisements ran in parallel over a period of 3 weeks. The conversion rate of the **Product Idea** page resulting from views generated by each advertisement are shown in Table 6.4.

A chi-square test yielded a p-value of 0.02. A 95% CI ran from 1.12 to 3.63. It was concluded that the higher conversion rate associated with Advertisement 1 was statistically significant. Table 6.5 shows the bounce rate (recorded by Google Analytics) associated with visits to the **Product Idea** page generated by each advertisement.

Landing page	Conversions	Non-Conversions	Views	Conversion Rate
Advertisement 1	26	352	378	6.88%
Advertisement 2	18	510	528	3.41%

Table 6.4: Conversion rates of the Product Idea page resulting from views generated by Advertisement 1 and Advertisement 2 over 3 weeks.

The bounce rates of the **Product Idea** page resulting from visits generated by each advertisement are shown in Table 6.5. A chi-square test on the data presented in Table 6.5 yielded a p-value of 0.03. A 95% CI ran from 0.69 to 0.98. It was concluded that the lower bounce rate associated with Advertisement 1 was statistically significant.

Landing page	Bounces	Non-Bounces	Visits	Bounce Rate
Advertisement 1	120	201	321	37.38%
Advertisement 2	205	246	451	45.45%

Table 6.5: Bounce rate for landing pages associated with Advertisement 1 and 2.

Based on the conversion rates shown in Table 6.4, and the bounce rates shown in Table 6.5, it was concluded that the **Product Idea** page performed better when used in combination with Advertisement 1. It was assumed that this was the case because the keywords used in Advertisement 1 had been included in the title of the **Product Idea** page. In order to determine whether, by using this principle, the conversion rate of the **Product Idea** page could be increased when used with Advertisement 2, the **Product Idea** page was modified so that its title changed dynamically depending on the advertisement that a visitor had clicked on. Figure 6.4 shows how the dynamic **Product Idea** page looked when a visitor landed on it after clicking Advertisement 2. The look of the dynamic **Product Idea** page for Advertisement 1 stayed the same as the non-dynamic **Product Idea** page shown in Figure 6.3. Advertisement 1 and Advertisement 2 ran in parallel for a period of 12 weeks.

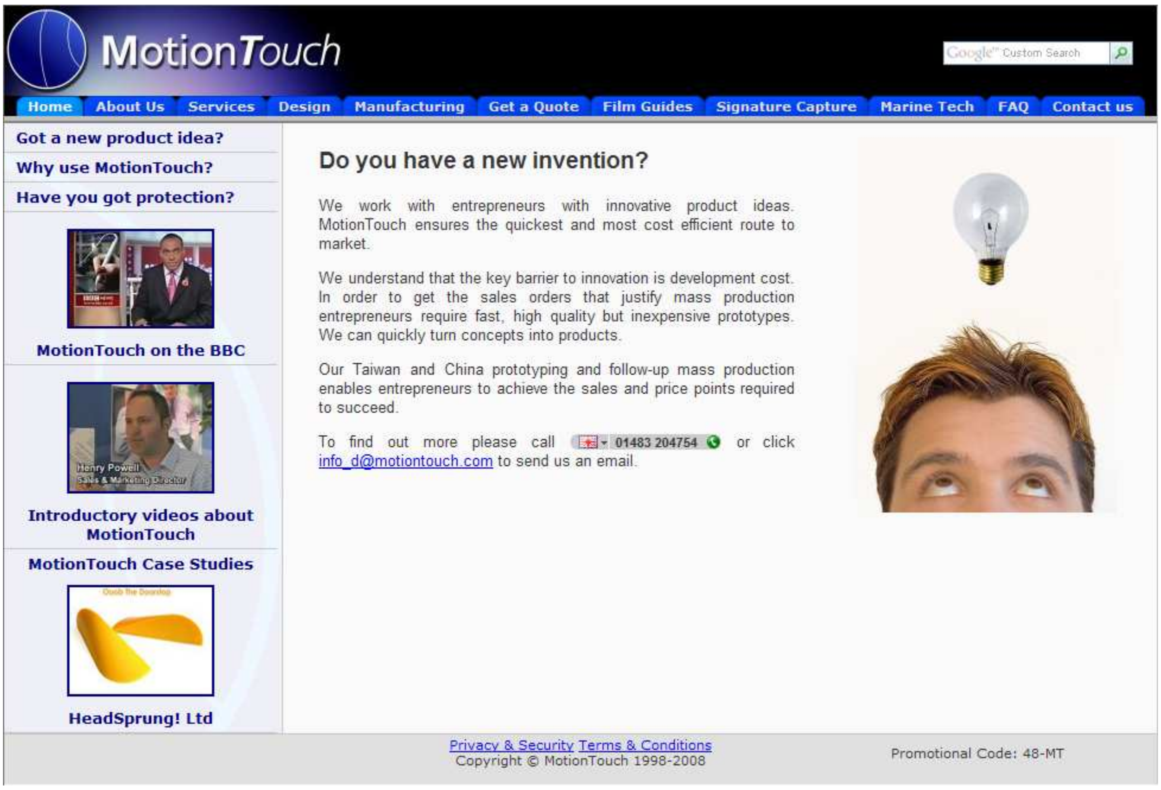


Figure 6.4: The dynamic Product Idea page for Advertisement 2.

The conversion rate of the dynamic **Product Idea** page resulting from views generated by Advertisement 1 and Advertisement 2 is summarised in Table 6.6. Because the setup shown in Figure 6.3 where Advertisement 1 and Advertisement 2 used the non-dynamic **Product Idea** page as landing page, ran for a 3 week period only, the results shown in Table 6.6 (obtained over 12 weeks) were averaged over a 3 week period before the performance of the dynamic and non-dynamic **Product Idea** page could be compared.

Table 6.7 shows the performance (over a 3 week period) of the **Product Idea** page for Advertisement 2 before and after the change to the page’s title was made.

Landing page	Conversions	Non-Conversions	Views	Conversion Rate
Advertisement 1	64	1691	1755	3.65%
Advertisement 2	121	2404	2525	4.79%

Table 6.6: Conversion rate of the dynamic Product Idea page resulting from views generated by Advertisement 1 and Advertisement 2 over 12 weeks.

Landing page	Conversions	Non-Conversions	Views	Conversion Rate
Dynamic **Product Idea** page	30	601	631	4.75%
Non-dynamic **Product Idea** page	18	510	528	3.41%

Table 6.7: Conversion rate of the dynamic and non-dynamic Product Idea page resulting from views generated by Advertisement 2 over a 3 week period.

A chi-square test on the data presented in Table 6.7 yielded a p-value of 0.25. A 95% CI ran from 0.79 to 2.47. It was concluded that the results were not statistically significant.

Data collected about the **Design Overview** page and Advertisement 2 over a 12 week period prior to Change 1 (describe in Section 6.2.2) was compared with data collected about the dynamic **Product Idea** page and Advertisment 2. The conversion rates and bounce rates are shown in Table 6.8 and Table 6.9 respectively.

Landing page	Conversions	Non-Conversions	Views	Conversion Rate
Dynamic **Product Idea** page	121	2404	2525	4.79%
Design Overview page	96	3495	3591	2.67%

Table 6.8: Conversion rate of the dynamic Product Idea page and Design Overview page resulting from views generated by Advertisement 2 over a 12 week period.

A chi-square test on the data presented in Table 6.8 yielded a p-value of 0.00. A 95% CI ran from 1.38 to 2.33. The p-value and the CI confirmed that the increase in conversion rate that was associated with the dynamic **Product Idea** page was statistically significant when compared to the **Design Overview** page. There was a relative increase in conversion rate of 79.40%.

A chi-square test was carried out on the data presented in Table 6.9. This produced a p-value of 0.47. A 95% CI ran from 0.96 to 1.09. The CI was narrow and included 1. Therefore, the increase in bounce rate that was observed for the dynamic **Product Idea** page was not statistically significant.

Landing page	Bounces	Non-Bounces	Visits	Bounce Rate
Dynamic **Product Idea** page	964	1128	2092	46.08%
Design Overview page	1355	1653	3008	45.05%

Table 6.9: Bounce rate of the dynamic Product Idea page and Design Overview page resulting from visits generated by Advertisement 2 over a 12 week period.

Data collected about the **Design Overview** page and Advertisement 1 over a 12 week period prior to Change 1 (described in Section 6.2.2) was compared with data collected about the dynamic **Product Idea** page and Advertisment 1. This is shown in Table 6.10. A chi-square test yielded a p-value of 0.34 and a 95% CI ran from 0.54 to 1.24. The decrease in conversion rate associated with the dynamic **Product Idea** page was not statistically significant.

Landing page	Conversions	Non-conversions	Views	Conversion Rate
Dynamic **Product Idea** page	64	1691	1755	3.65%
Design Overview page	33	709	742	4.45%

Table 6.10: Conversion rate of the dynamic Product Idea page and Design Overview page resulting from views generated by Advertisement 1 over a 12 week period.

6.4.3. Conclusion

The results obtained were inconclusive in determining whether the inclusion of keywords used in an advertisement into its landing page's title could increase conversion rate. However, for Advertisement 2, the dynamic **Product Idea** page had a higher conversion rate than the **Design Overview** page that was statistically significant. For Advertisement 1, the dynamic **Product Idea** page had a lower conversion rate than the **Design Overview** page but this was not statistically significant. Therefore, it could be concluded that the landing page for Advertisement 2 had been improved and that a relative increase in conversion rate of 79.40% (compared to the **Design Overview** page) that was statistically significant had been achieved.

These conclusions had limitations. External factors such as search and seasonal trends could have affected the results obtained when comparing the performance of the dynamic **Product Idea** page and the **Design Overview** page. The effect of these external factors could have been reduced if the performance of the two landing pages had been compared over a similar time frame.

Another limitation was that the advertisement message was reinforced on the landing page through the title only. The content of the page was not changed to match the message of the advertisement. It could be argued that changing the title alone was not enough to match the message of the advertisement and that if the content had been matched to the advertisement message; higher conversion rates could have been observed.

6.5. Change 3 - Visual design

6.5.1. Concept

A study by Lindgaard et al (2006) showed that visitors formed an impression of a web page in the first 50ms based on visual appeal, for example, design layout, colour, etc. This implied that the main features and general appearance of a landing page had a greater initial impact on visitors than its content (Gofman and Moskowitz, 2009).

6.5.2. Concept application and landing page testing

The visual design of landing pages was changed when the front-end of the dynamic main website was redesigned to have a professional and industry-appropriate design (described Chapter 4). The performance of the dynamic **Product Idea** page before and after the change was compared to measure its impact on conversion rate. Table 6.11 shows the conversion rate of the dynamic **Product Idea** page resulting from views generated by Advertisement 1 over a 3 month period before and after the change in visual design.

Landing page	Conversions	Non-Conversions	Views	Conversion Rate
Product Idea page (after change)	62	1286	1348	4.60%
Product Idea page (before change)	64	1691	1755	3.65%

Table 6.11: Conversion rate of the dynamic Product Idea page resulting from views generated by Advertisement 1 over a 3 month period before and after the change in visual design.

It can be seen that there was an increase in conversion rate after the change in design. A chi-square test yielded a p-value of 0.18. A 95% CI ran from 0.90 to 1.78. The increase in conversion rate was deemed not to be statistically significant.

Table 6.12 shows the conversion rate of the dynamic **Product Idea** page resulting from views generated by Advertisement 2 over a 3 month period before and after the change in visual design.

Advertisement	**Conversions**	**Non-Conversions**	**Views**	**Conversion Rate**
Product Idea page (after change)	101	3111	3212	3.14%
Product Idea page (before change)	121	2404	2525	4.79%

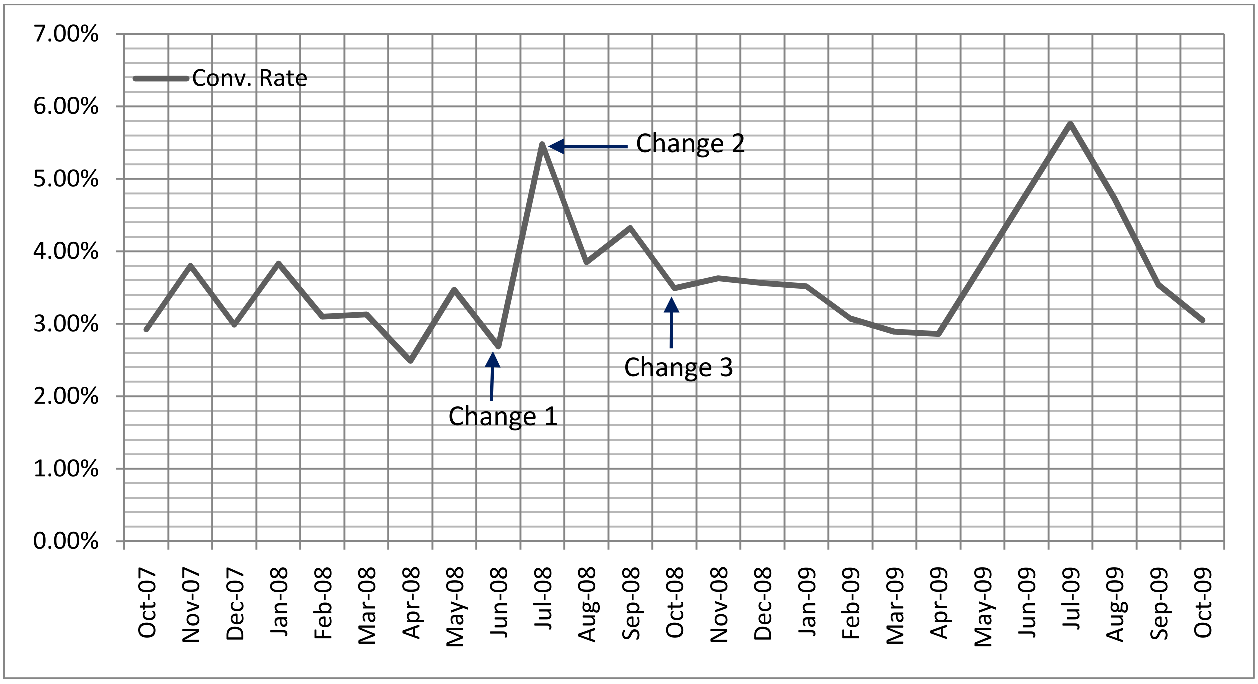
Table 6.12: Conversion rate of the dynamic Product Idea page resulting from views generated by Advertisement 2 over a 3 month period before and after the change in visual design.

For Advertisement 2, a decrease in conversion rate was observed following the changes to the website's design. A chi-square test yielded a p-value of 0.00. The CI for a relative conversion rate of 0.66 in ran from 0.51 to 0.85. The decrease was statistically significant.

6.5.3. Conclusion

When used as landing page by Advertisement 1, the dynamic **Product Idea** page achieved an increase in conversion rate but this was not statistically significant. When used as landing page by Advertisement 2, the dynamic **Product Idea** page achieved a decrease in conversion rate that was statistically significant. The observations made had limitations as the data used for comparison had been collected over different time periods. Therefore, external factors such as search and seasonal trends could have affected the results.The effect of these external factors could have been reduced if the performance of the landing pages had been compared over a similar time frame. Also the effects of the recession that had started in the second quarter of 2008 (BBC News, 2009) had become more severe in the months following the changes made to the dynamic main website's visual design.

Graph 6.2 shows how the conversion rate of the Inventor campaign varied over a period of 2 years. The average conversion rate associated with the Inventor campaign before Change 1 was 3.22%. The average conversion rate after Change 1 (and including Change 2 and 3) was 3.84%. It appeared that there had been an increase in average conversion rate after Change 1. This could have been due to seasonal trends and some results associated with Changes 1, 2 and 3 were not statistically significant.



Graph 6.2: Conversion rate of Inventor campaign over time.

A t-test produced a value of 1.7864 at a DF of 20. The corresponding one-tailed probability (p-value) was 0.04. Therefore, the relative increase in conversion of 19.25% was statistically significant. It appeared that the combined effect of Changes 1, 2 and 3 had a positive influence on the overall conversion rate of the landing pages. This increase in average conversion rate was observed despite external factors such as the economic recession that started in the second quarter of 2008 and a cut in advertising budget (of 50%) at the beginning of February 2009.

It was not possible to conclusively determine whether individual changes to the landing pages of the Inventor campaign had impacted conversion rate. However, it appeared that a combination of the following techniques: using targeted content (Change 1), including keywords from advertisement into landing page title (Change 2) and improving visual design (Change 3) had resulted in an increase in average conversion rate that was statistically significant despite a recession and reduction in advertising budget.

6.6. Change 4 - Content Structure

6.6.1. Concept

In his research Nielsen (1997) found that 79% of users always scanned new pages that they came across while only 16% read every word. In order to accommodate the different reading patterns, Nielsen (1997) suggested that websites should use scannable text by:

- Highlighting keywords.
- Having meaningful sub-headings, not “clever” ones.
- Using bullet lists.
- Having one idea per paragraph.
- Using the inverted pyramid style.

When measuring the usability of a web page against all the guidelines outlined above, Nielsen observed a 124% increase in usability.

Loveday and Neihaus (2008) suggested that the following should be considered when writing content landing page:

- Tone and language. This needed to match the language of the target audience, for example if the audience was technical, then technical jargon could be used.
- Engaging visitors with benefits and scenarios they can relate to.
- The amount of text. Depending on the complexity of the offer being made on a landing page, the text could vary in length. If there was too much text, a rule of thumb was to have essential points above the fold. This allowed visitors who did not like to scroll to still see the offer.

6.6.2. Concept application and landing page testing

Figure 6.5 shows the **Plastic Manufacturing** page which was the original landing page of the Manufacturing campaign.

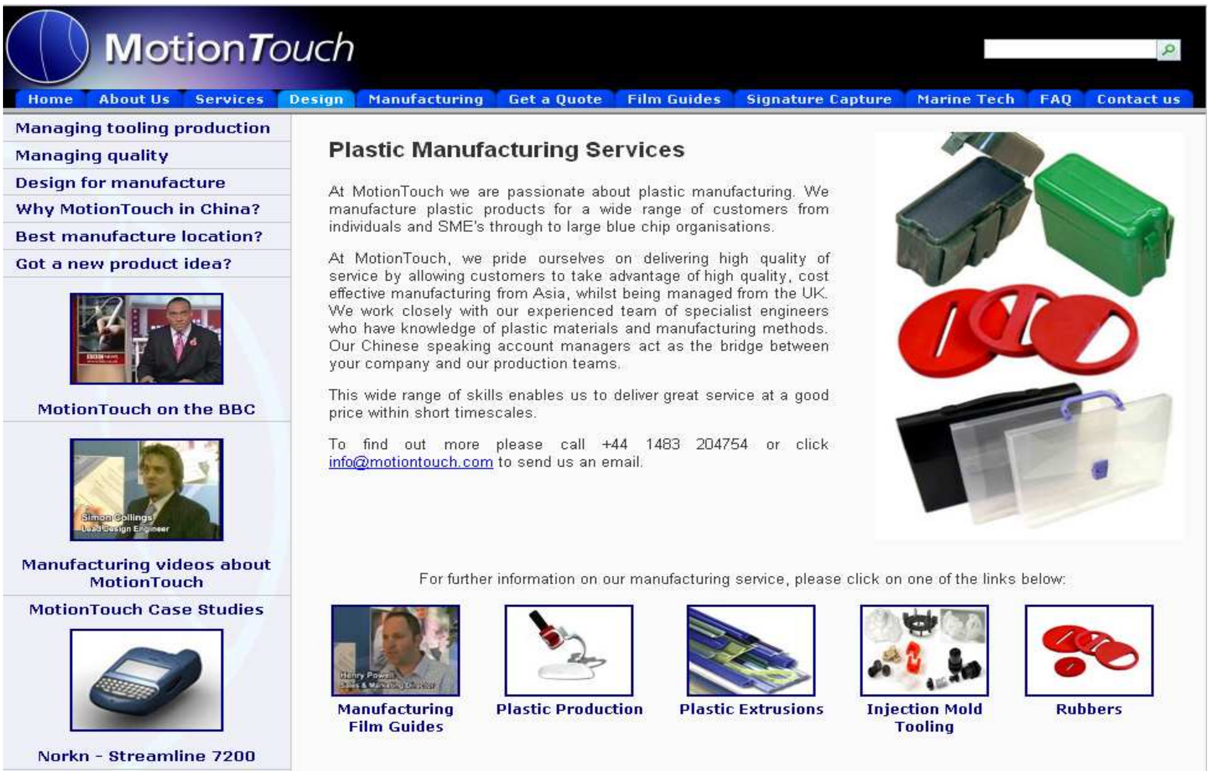


Figure 6.5: Plastic Manufacturing page. Original landing page for Manufacturing campaign.

The **Plastic Manufacturing** page ignored the guidelines suggested by Nielsen (1997) and Loveday and Neihaus (2008). A new landing page called **Product Manufacture** (see Figure 6.6) was implemented based on some of the concepts described in Section 6.6.1. This landing page was similar to the **Plastic Manufacturing** page except for the structure of its content. The content structure followed some of Nielsen's (1997) guidelines (bullet points, one idea per paragraph) and suggestions by Loveday and Neihaus (2008).

Figure 6.6: Product Manufacture page.

In order to test the landing pages, the advertisement that led visitors to the **Plastic Manufacturing** page was duplicated. The duplicate advertisement was then set so that it led visitors to the **Product Manufacture** page. The advertisements and landing pages ran in parallel over a period of 2 months and the results obtained are shown in Table 6.13.

Landing page	Conversions	Non-Conversions	Views	Conversion Rate
Product Manufacture	21	947	968	2.17%
Plastic Manufacturing	31	1603	1634	1.90%

Table 6.13: Conversion rates for landing pages shown in Figure 6.5 and Figure 6.6.

The results showed that the **Product Manufacture** page that followed the guidelines for structuring and writing content had a higher conversion rate. A chi-square test produced a p-value of 0.63. A 95% CI ran from 0.66 to 1.98. The increase in conversion was not statistically significant. The bounce rates of the landing pages are shown in Table 6.14.

Landing page	Bounces	Non-Bounces	Visits	Bounce Rate
Product Manufacture	311	395	706	44.05%
Plastic Manufacturing	1115	817	1894	58.87%

Table 6.14: Bounce rates for landing pages shown in Figure 6.5 and Figure 6.6.

The results showed that the **Product Manufacture** page had a lower bounce rate. A chi-square test produced a p-value of 0.00. A 95% CI ran from 0.66 to 0.82. The relative decrease of 25.17% in bounce rate was therefore statistically significant.

6.6.3. Conclusion

Although the **Product Manufacture** page had a higher conversion rate, this was not statistically significant. Therefore, it could not be demonstrated that good content structuring and writing had an effect on conversion rate. However, the bounce rate of the **Product Manufacture** page that followed guidelines for good content structuring and writing was lower than that of its counterpart. This suggested that good content structuring and writing could encourage visitors to stay on a website and browse rather than leave without browsing.

6.7. Change 5 – Visitor segmentation

6.7.1. Concept

According to Talerico (2010) "*online traffic is like a mixed fruit basket, similar in many ways, but also very different. Marketing to a mixed basket is difficult—how can you possibly know the right image, message, tone and offer that will appeal to anyone & everyone? This task becomes much more manageable when you group similar people together into segments.*" Leap (2010) suggested that for online marketing, segmentation could be carried out pre-click and post-click. Pre-click segmentation involved keyword-based segmentation, so as to maximise the relevancy of paid clicks. Post-click segmentation involved segmenting visitors when they reached a landing page. Such a landing page presented visitors with simple choices to make about who they are or what they are looking for. The subsequent pages that a visitor was directed to were more relevant and tailored to the visitor's audience segment. Leap (2010) called this type of segmentation "*user-directed"* segmentation as visitors directed how they were segmented, and Leap claimed that this could increase conversion rate and the quality of enquiries.

6.7.2. Concept application and landing page testing

This research used two types of landing page design to implement user-directed segmentation:

a) Segmentation using pictures.

b) Segmentation using a questionnaire.

6.7.3. Change 5a - Segmentation using pictures

It was assumed that a segmented landing page that could help visitors find relevant content quickly would be better suited to visitors who fitted the Manufacturer persona, that is, visitors who were goal-oriented and time conscious. Therefore, the original landing page (**Plastic Manufacturing** shown in Figure 6.7) of the Manufacturing campaign was replaced with a landing page that implemented segmentation using pictures (**Picture-segmented** page shown in Figure 6.8).

Figure 6.7: Plastic Manufacturing page. Original landing page for Manufacturing campaign.

Due to a limited advertising budget, the two landing pages did not run in parallel. In order to compare performance, data from two different years (2008 and 2009) had to be used. To minimise the effect of seasonal trends, data corresponding to the same months (January to August) of the different years was used. Table 6.15 summarises the performance of the landing pages.

Figure 6.8: New Picture-segmented landing page with segmentation options.

Landing page	Conversions	Non-Conversions	Views	Conversion Rate
Picture-segmented	237	9167	9404	2.52%
Plastic Manufacturing	111	9051	9162	1.21%

Table 6.15: Performance of Plastic Manufacturing page between January 2008 and August 2008 and Picture-segmented page between January 2009 and August 2009.

A chi-square test on the data shown in Table 6.15 yielded a p-value of 0.00. A 95% CI ran from 1.66 to 2.60. Therefore, the relative increase in conversion rate of 108.26% for the new **Picture-segmented** landing page was statistically significant. The bounce rate of the landing pages were compared using data collected by Google Analytics. This is shown in Table 6.16. A relative improvement of 22.96% in bounce rate was observed for the new **Picture-segmented** landing page.

Landing page	Bounces	Non-bounces	Visits	Bounce Rate
Picture-segmented	3583	3770	7353	48.73%
Plastic Manufacturing	10005	5813	15818	63.25%

Table 6.16: Bounce rates for Picture-segmented and Plastic Manufacturing landing pages.

A chi-square test for the data shown in Table 6.16 yielded a p-value of 0.00. A 95% CI ran from 0.75 to 0.79. Therefore, it was concluded that relative improvement of in bounce rate that was observed for the new **Picture-segmented** landing page was statistically significant.

6.7.4. Conclusion

The statistically significant increase in conversion rate and decrease in bounce rate demonstrated that the new **Picture-segmented** landing page performed better than the **Plastic Manufacturing** page. There were limitations associated with the data used to draw this conclusion. The two landing pages that were compared were different in terms of look and navigation. Also, data used to compare performance was not from the same time period. Therefore, it was not possible to conclusively attribute the increase in conversion rate solely to the use of a picture-segmented landing page as external factors could have also contributed to this.

6.7.5. Change 5b - Segmentation using questionnaire

A landing page that used a questionnaire (see Figure 6.9) to segment visitors was tested. This landing page (**Questionnaire**) ran in parallel with the **Picture-segmented** landing page shown in Figure 6.8. A summary of the performance of these landing pages over an eight months period is shown in Table 6.17.

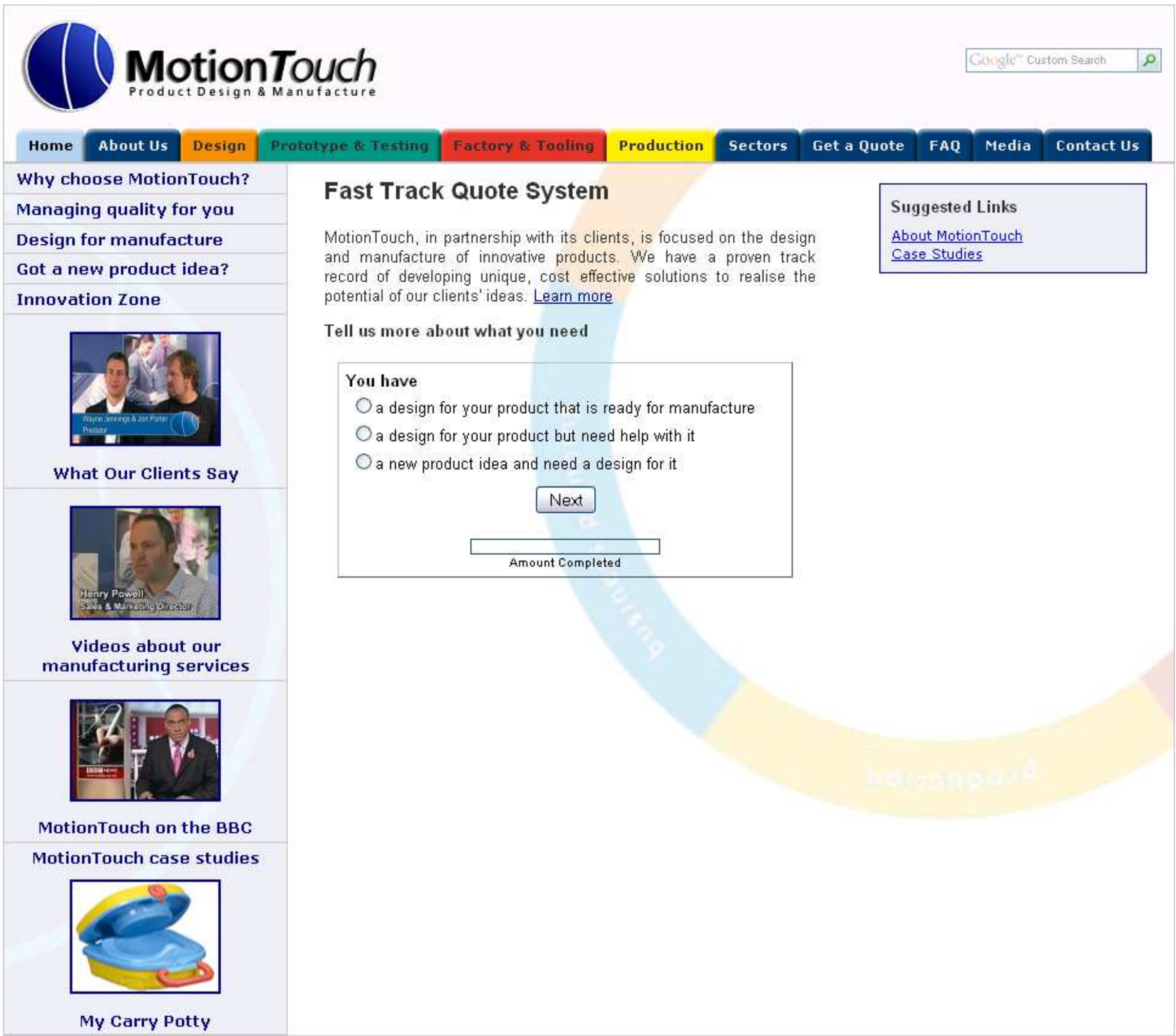


Figure 6.9: Questionnaire landing page for Manufacturing campaign.

A chi-square test for the data shown in Table 6.17 yielded a p-value of 0.22. A 95% CI ran from 0.92 to 1.43. The CI was narrow and included 1 which suggested that the increase in conversion rate associated with the **Questionnaire** landing page was not statistically significant. The bounce rates (shown in Table 6.18) of the two landing pages were compared.

Landing page	Conversions	Non-Conversions	Views	Conversion Rate
Questionnaire	156	3963	4119	3.79%
Picture-segmented	150	4395	4545	3.30%

Table 6.17: Performance data for Questionnaire and Picture-segmented landing pages.

The questionnaire landing page had a higher bounce rate than the picture-segmented landing page. A chi-square test yielded a p-value of 0.00. A 95% CI ran from 1.22 to 1.35. The higher bounce rate associated with the questionnaire landing page appeared be statistically significant.

Landing page	Bounces	Non-Bounces	Visits	Bounce Rate
Questionnaire	1934	1390	3324	58.18%
Picture-segmented	1391	1673	3064	45.40%

Table 6.18: Bounce rates for Questionnaire and Picture-segmented landing pages

The comparison between the performance of the **Questionnaire** landing page and the **Picture-segmented** landing page were inconclusive. The **Questionnaire** landing page was then compared with the original landing page (**Plastic Manufacturing**) shown in Figure 6.7. Data over the same months (May to August) but from two different years (2008 and 2009) were used for comparison since the landing pages did not run in parallel. Table 6.19 shows the conversion rates of the two landing pages.

Landing page	Conversions	Non-Conversions	Views	Conversion Rate
Questionnaire	89	1962	2051	4.34%
Plastic Manufacturing	61	5170	5231	1.17%

Table 6.19: Performance data for Questionnaire and Plastic Manufacturing landing pages.

A chi-square test on the data presented in Table 6.19 yielded a p-value of 0.00. A 95% CI ran from 2.70 to 5.13. Therefore, the relative increase in conversion rate of 270.94% for the **Questionnaire** landing page was statistically significant.

Landing page	Bounces	Non-Bounces	Visits	Bounce Rate
Questionnaire	1010	651	1661	60.81%
Plastic Manufacturing	3953	2461	6414	61.63%

Table 6.20: Bounce rate for original and questionnaire landing pages.

Table 6.20 shows the bounce rate of the **Plastic Manufacturing** page and the **Questionnaire** landing page. A chi-square test yielded a p-value of 0.54 and a 95% CI ran from 0.94 to 1.03. Therefore, the relative decrease in bounce rate associated with the **Questionnaire** landing page was not statistically significant.

6.7.6. Conclusion

It was not possible to ascertain whether the **Questionnaire** landing page was better than the **Picture-segmented** landing page. However, the **Questionnaire** landing page had a higher conversion rate (that was statistically significant) than the original landing page of the Manufacturing campaign (**Plastic Manufacturing** page shown in Figure 6.7). There were limitations associated with this conclusion. The **Plastic Manufacturing** page and the **Questionnaire** landing page were different in terms of look and navigation. Also, data used to compare performance was not from the same time period. Therefore, it was not possible to conclusively attribute the increase in conversion rate to the use of a landing page that provided user-directed segmentation through a questionnaire.

6.8. Chapter Discussion

This Chapter described how landing page optimisation was carried out during the research. Different landing pages were tested and their performance was then evaluated using conversion and bounce rate. Not all tests yielded results that were statistically significant. Therefore, it was not possible to conclusively determine the impact that individual changes to the landing pages had on their performance.

However, it was observed that for the Inventor campaign the average conversion rate following all the changes made to its landing pages was higher than the average conversion rate of its original landing page. This increase in average conversion rate was statistically significant.

Three new landing pages were tested for the Manufacturing campaign. The first landing page was created by changing the content structure and writing of the original landing page. The changes did not have a significant effect on conversion rate. However, the bounce rate of the new landing page was lower than that of the original landing page and the difference was statistically significant.

Two landing pages that supported user-directed segmentation were created. Both had the same visual design but used different types of segmentation. The visual design of these landing pages was different to that of the original landing page of the Manufacturing campaign. Both segmentation landing pages achieved statistically significant increases in conversion rate compared to the original landing page of the Manufacturing campaign. Overall, it appeared that the new landing pages that were implemented during the research had higher conversion rates than the original landing page of the Manufacturing campaign.

The conclusions drawn from the results obtained had some limitations. In some cases, multiple changes were made to a landing page. This affected comparison to an original landing page. When landing pages did not run in parallel, performance data from different time frames were used for comparison. As a result, the effect of external factors could not be controlled. However, to minimise the effect of seasonal trends, comparisons were carried out with data from the same months (but different years) whenever possible.

External factors such as the economic recession that started in the second quarter of 2008 and a cut in advertising budget (of 50%) at the beginning of February 2009 also affected the results presented in this Chapter.

The aim of the landing page optimisation process had been to encourage as many visitors as possible to browse the dynamic main website and to enquire (convert). The OTM described in Chapter 4 collected data about these visitors' browsing behaviour. This research analysed the data captured by the OTM to determine whether visitors' browsing behaviour could predict their intention to convert. Chapter 7 describes the analysis that was carried out and the results that were obtained.

CHAPTER 7

EXPERIMENTS AND RESULTS

This Chapter describes the data mining and analysis that was carried out on the data collected by the Online Tracking Module (OTM) described in Chapter 4. The aim was to analyse data collected by the OTM to find rules or models that enabled the prediction of conversions from recorded visitor activity on a website.

Models or rules that could predict whether visitors were likely to convert could help identify attributes and website design elements that affected conversion positively. This could be used to improve website design and increase conversion, which could ultimately lead to increased sales and profit. Also, the ability to predict conversion could help create systems that could monitor visitor’s browsing and then guide them along an optimised or personalised path.

This Chapter first describes the different stages of the data mining and analysis process namely population sampling, data retrieval, data cleaning and transformation and knowledge discovery. It then describes the data mining tool and algorithms that were used to find models. The results of the various explorations of the data are then described. Finally, the results as well as their limitations are discussed and conclusions are drawn.

7.1. Experiment Design

Experiments were designed to retrieve and analyse data about visitors' browsing behaviour on the collaborating company's dynamic main website. Collected data was analysed using the PolyAnalyst data mining engine to find models to predict online conversion.

Data captured by the Online Tracking Module (OTM) described in Chapter 4 was used. The OTM saved data in a Microsoft Structured Query Language (SQL) database. Data selection and cleaning was performed in the SQL database and was then exported to a Microsoft Excel file. A Visual Basic for Applications (VBA) macro was then used to transform the data so as to derive and format values for predictor attributes (described in Section 7.3.3). These were then imported into a data mining engine called PolyAnalyst to find models.

7.2. The data mining process

The data mining process used in this research is shown in Figure 7.1. Data extraction encompassed Stages 1 and 2 of the experiments. It consisted of identifying the criteria for the population sample required and extracting data from the OTM database for that sample. The data was then saved in Microsoft Excel files.

Data cleaning consisted of removing inconsistent and erroneous data. Data transformation consisted of deriving and formatting the values of predictor attributes from the data saved in Microsoft Excel files. These predictor attributes were then imported into PolyAnalyst to find models.

PolyAnalyst's data mining algorithms were run during the data mining phase. This provided a set of models. PolyAnalyst generated accuracy measures for each model

model analysis was repeated (based on the findings of previous iterations) in order to improve the accuracy of models.

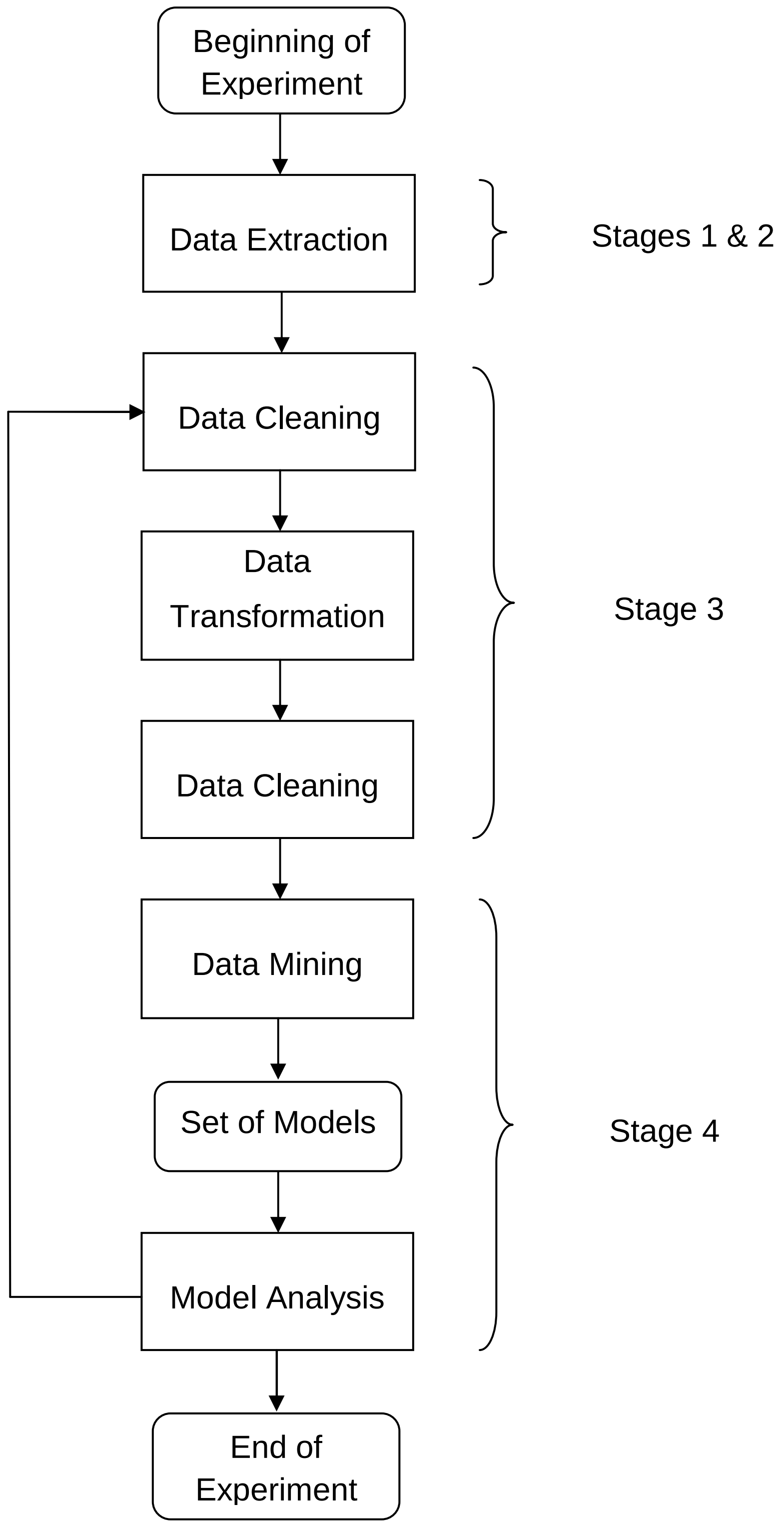


Figure 7.1: The data mining process.

7.3. Implementing the data mining process

The experiments described in this Chapter were implemented in four stages:

1. Population sampling.
2. Data retrieval.
3. Data cleaning and transformation.
4. Knowledge discovery.

7.3.1. Stage 1: Population sampling

The OTM collected data for all visitors who browsed the collaborating company's dynamic main website. Google Analytics (described in Chapter 2) was also used to capture data about the dynamic main website's traffic. Reports from Google Analytics showed that:

- the dynamic main website received an average of 220 visits a day.
- traffic to the dynamic main website was generated by the following sources:
  - Search Engines (86%),
  - Direct Traffic (9%) and
  - Referring Sites (5%).
- 82% of visits were unique visits while the remaining 18% were returning visits.
- On average a visitor viewed 3 pages per visit and spent an average of 159 seconds per visit.

Data collected by MS CRM showed that visitors who enquired varied from individuals to corporate businesses. Data collected by the OTM showed that visitors were mainly from the United States or Europe (predominantly the United Kingdom and Ireland). For the purposes of this research only a subset of the available data was selected. The research described in this dissertation considered a population sample that satisfied the following criteria:

- First time visitors.
- Visitors who landed on the dynamic main website after a keyword search. This excluded search keywords that contained the collaborating company's name or variations of it, for example "MotionTouch", "Motion touch design", etc.
- Visitors who did not bounce off the dynamic main website. A bounce occurred when a visitor left immediately without having viewed any page but the landing page (White, 2006).

This research did not consider returning visits or visits that were not associated with a search keyword.

7.3.2. Stage 2: Data retrieval

SQL commands were used to retrieve data from the OTM database for a sample population that satisfied the criteria defined in Stage 1. This data was then saved to Microsoft Excel files which were used in Stage 3. Data retrieval from the OTM database took place on the collaborating company's web server. The data retrieval involved a series of steps. They were:

1. Extracting and combining elements of browsing behaviour and web page structure for each visitor from various tables in the OTM database. This data was then saved in a table called Results. This table contained records for visits that had converted, had not convert and visits that had bounced.
2. Records that were not associated with a conversion were retrieved. For each record, the *[Total Time on site]* was calculated and the results were stored in a

table called NonConvAll. Bounces were identified from table NonConvAll by selecting records that had a *[Total Time on site]* that was equal to NULL.

3. Data associated with conversions and non-conversions were extracted from the Results database and saved in Microsoft Excel files ready for the transformation stage. Visits that originated from:
    a. China were excluded from the data set as the majority of email enquiries originating from China were spam emails.
    b. Poland were excluded as these represented visits by staff who worked in the collaborating company's Polish office.
    c. Netherlands were excluded as an individual with malicious intent was accessing the dynamic main website from the Netherlands.

The SQL commands used in Steps 1, 2 and 3 are listed in Appendix C. Figure 7.2 is a screenshot showing the structure of the Microsoft Excel files created in Step 3.

The Microsoft Excel files had the following data columns:

- *[UserId]* – Unique id generated and assigned by the OTM to a visitor.
- *[Title]* – Title of a web page.
- *[PageId]* – Unique id for each web page.
- *[PageType]* – Each web page was given a type that described the function of that page. *[PageType]* could have one of the following values: *Video, Download, Company Info, Other, Services, Contact Us, Form, Media Access, FT Quote, Case Study, PR.* These are described at the end of this section.

	1	2	3	4	5	6	7	8
1	UserId	Title	PageId	PageType	MicroId	Sequence	TimeSpent	Keywords
24	625543	Product Ideas - MotionTouch	501	Services	1	1	36	invention sales
25	625543	Have you got protection? - MotionTouch	68	Services	1	2	14	invention sales
26	625543	Manufacture your product with us - MotionTouch	76	Services	1	3	49	invention sales
27	625543	Prototype & Testing - MotionTouch	36	Services	1	4	12	invention sales
28	625543	Product Testing & Certification - MotionTouch	38	Services	1	5	50	invention sales
29	625543	The Design Process - MotionTouch	34	Services	1	6	44	invention sales
30	625543	Product Ideas - MotionTouch	501	Services	1	7	29	invention sales
31	625543	Fast Track Quote - MotionTouch	169	FTQuote	1	8	21	invention sales
32	625543	Product Design - Fast Track Quote - MotionTouch	81	Form	1	9	38	invention sales
33	625543	Product Replication - Fast Track Quote - MotionTouch	82	Form	1	10	466	invention sales
34	625543	Fast Track Quote - MotionTouch	169	FTQuote	1	11	28	invention sales
35	625543	Product Design - Fast Track Quote - MotionTouch	81	Form	1	12	345	invention sales
36	625543	Send	0	Other	1	13	20	invention sales
37	625543	Message Sent	13	Other	1	14	12	invention sales
38	625543	Product Ideas - MotionTouch	501	Services	1	15	1	invention sales
39	625543	The Design Process - MotionTouch	34	Services	1	16	1	invention sales
40	625543	Prototype & Testing - MotionTouch	36	Services	1	17	1	invention sales
41	625543	Manufacture your product with us - MotionTouch	76	Services	1	18	1	invention sales
42	625543	Have you got protection? - MotionTouch	68	Services	1	19	0	invention sales

Figure 7.2: Structure of Microsoft Excel file that stored visitors' browsing data.

- *[MicroId]* – This represented a session number and was used to identify returning visits. If *[MicroId]* was 1, it indicated a first visit. If it was greater than 1 then it indicated returning visits, in which case the value represented the number of returning visits. Chapter 4 described how the OTM assigned a *[MicroId]* to returning visitors.
- *[Sequence]* – This indicated the order in which web pages were accessed by a visitor.
- *[TimeSpent]* – The number of seconds that had elapsed between a visitor landing on a page and accessing the next page. *[TimeSpent]* was not available for the last page visited because of the method used by the OTM to calculate this metric. *[TimeSpent]* was regarded as the time a visitor spent on a web page but there was no way of knowing how the visitor actually spent this time. The visitor could have been browsing a page or could have been away from their computer.
- *[Keywords]* – The search keyword used by a visitor.

Definition of *[Page Type]*

Case Study – Pages that displayed case studies about projects that had been completed by the collaborating company. Each case study had its own page.

Company Info – Pages that provided information about the collaborating company, for example the **Home** and **About Us** pages.

Contact Us – This was the **Contact Us** page. The **Contact Us** page displayed contact details, for example office addresses and email addresses.

Download – Pop up pages that allowed visitors to download files.

Form – Pages that displayed enquiry forms.

FT Quote – These pages offered segmentation options that allowed visitors to guide themselves to an enquiry form best suited for the type of enquiry that they wanted to make.

Media Access – Pages that listed files such as videos and case studies. Visitors accessed individual *Video and Case Study* pages via a *Media Access* page.

PR – Pages that displayed press releases.

Services – Content pages that described the services that the collaborating company offered, for example the **Design Overview** page, the **Product Sector Overview** page and the **Quality Services** page.

Video –These were pop up pages that contained embedded videos and no content or navigation.

Other – Pages that did not fit in any category, for example a **Thank you** page.

7.3.3. Stage 3: Data cleaning and transformation

Data saved during Stage 2, was cleaned and then transformed into attributes that could be imported into PolyAnalyst. A VBA macro was implemented to transform the data. The data sets obtained from Step 3 of Stage 2 were cleaned and transformed separately but using the same process. The transformed data was then used to create two new data sets called ***Training Data*** and ***Test Data***.

Data cleaning - *[Keywords]* column

Records were removed when the string stored in the *[Keywords]* column contained:

- URL fragments.
- only alphanumeric string.
- the collaborating company's name or variations of it, for example, “MotionTouch”, “Motion touch design”, etc.

The *[Keywords]* column was checked a second time for invalid values after the data had been transformed. This was done manually.

Data cleaning - *[PageType]* column

All records associated with a *[UserId]* that had a *[PageType]* entry that was equal to NULL were removed. A *[PageType]* that was equal to NULL indicated that a visitor had browsed the collaborating company's online shop. This website was not considered in this research.

Data transformation

Once the data had been cleaned, then it was transformed. In order to do this a VBA macro was implemented to transform the data saved in Microsoft Excel files. The VBA macro saved the transformed data within the same file but in a different spreadsheet called ***Clean Data***. The spreadsheet with data saved from Stage 2 was called ***Raw Data***. The transformation process derived the following attributes for each visitor record found in ***Raw Data***:

- *[Browsing Time]* – Time in seconds spent browsing a website before enquiring. For visitors who did not enquire this represented the time spent browsing a website before leaving. *[Browsing Time]* excluded time spent filling in enquiry forms. This also excluded time spent on a **Contact Us** page if a visitor accessed and then sent an enquiry form via the **Contact Us** page.
- *[Browsed ContactUs]* **–** Number of times a **Contact Us** page was visited without the visitor subsequently accessing and sending an enquiry form.
- *[Browsed ContactUsTS]* **–** Time spent browsing a **Contact Us** page without subsequently accessing and sending an enquiry form.
- *[Case Studies]* – Number of pages of type *Case Study* accessed.

- *[Case StudiesTS]* – Total time spent on pages of type *Case Study*.
- *[Company Info]* – Number of pages of type *Company Info* accessed.
- *[Company InfoTS]* – Total time spent on pages of type *Company Info*.
- *[ContactUs Sent]* – This value was set to 1 if an enquiry ensued from a visit to a **Contact Us** page. Otherwise, the value was set to 0.
- *[ContactUs SentTS]* – Time spent on a **Contact Us** page that generated an enquiry.
- *[Conversion]* – 1 or Yes for a conversion and 0 or No for a non-conversion.
- *[Download]* – Number of files downloaded.
- *[DownloadTS]* – Time spent downloading files.
- *[FTQuote]* – Number of pages of type *FTQuote* accessed.
- *[FTQuote TS]* – Total time spent on pages of type *FTQuote*.
- *[Keyword Length]* – The number of words in a search keyword.
- *[Media Access]* – Number of pages of type *Media Access* that were accessed.
- *[Media AccessTS]* – Total time spent on pages of type *Media Access*.
- *[Other]* – Number of pages of type *Other* accessed. A page called **Sent** was loaded by the script that sent emails from a website. If an email was successfully sent, then the script redirected the user to a **Thank You** page. The **Sent** page and **Thank You** page were of type *Other.*
- *[OtherTS]* – Total time spent on pages of type *Other*.
- *[PR]* – Number of pages of type *PR* accessed.
- *[PRTS]* – Total time spent on pages of type *PR*.
- *[Sent Form]* – Number of pages of type *Form* that were accessed and sent. The value could be either 1 or 0.

- *[Sent FormTS]* – Total time spent on pages of type *Form* that were accessed and then sent.
- *[Services]* – Number of pages of type *Services* accessed.
- *[ServicesTS]* – Total time spent on pages of type *Services*.
- *[Total Time on site]* – Total time in seconds that a visitor spent on a website excluding time spent on the last page they visited. Since the OTM had no method of knowing when a visitor left the website, it recorded time spent on the last page visited as NULL.
- *[Unsent Form]* – Number of pages of type *Form* that were accessed but not sent.
- *[Unsent FormTS]* – Total time spent on pages of type *Form* that were accessed but not sent.
- *[UserId]* – Unique ID for each visitor.
- *[Video]* – Number of videos that a visitor accessed.
- *[VideoTS]* – Time spent watching videos.

The structure of the ***Clean Data*** spreadsheet is shown in Figure 7.3 and Figure 7.4. Figure 7.3 shows a screenshot of the first fifteen columns of the spreadsheet and Figure 7.4 shows a screenshot of the next seventeen columns. The VBA macro transformed the browsing history shown in Figure 7.2 into the values stored in row five of the spreadsheet shown in Figure 7.3 and Figure 7.4.

In order to calculate *[Browsing Time]*, each time the VBA macro came across a page of type *Form*, it checked whether the next page visited had a *[PageId]* of 0. This sequence of action indicated that an enquiry form had been sent and that the VBA macro needed to adjust its calculation of *[Browsing Time]* to exclude *[TimeSpent]* on the page of type *Form* and any pages accessed after it. For example, when the macro parsed the data

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Userid	Keyword Length	Keyword	Video	VideoTS	Download	DownloadTS	Company Info	Company InfoTS	Services	ServicesTS	Unsent Form	UnsentFormTS	Case Studies	Case StudiesTS
2	626271	4	how to invent something	0	0	0	0	0	0	1	33	0	0	0	0
3	625838	2	Invention developers	0	0	0	0	0	0	1	26	0	0	0	0
4	625728	5	toilets made from plastic uk	0	0	0	0	0	0	0	0	1	16	0	0
5	625543	2	invention sales	0	0	0	0	0	0	7	234	2	504	0	0
6	624603	5	productionline of plastic jersey barriers	0	0	0	0	0	0	1	168	2	1149	0	0
7	624128	2	product development	0	0	4	103	1	103	15	770	0	0	1	19

Raw Data | Clean Data | Sheet1 | Sheet3 | Sheet2

Figure 7.3: First fifteen attribute columns containing values derived from Raw Data spreadsheet.

	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1	Media Access	Media AccessTS	PR	PRTS	Other	OtherTS	Browsing Time	Total Time on site	Conversion	Browsed ContactUs	Browsed ContactUsTS	Sent Form	Sent FormTS	Contactus Sent	Contactus SentTS	FTQuote	FTQuote TS
2	0	0	0	0	2	23	33	428	1	0	0	1	119	0	0	0	0
3	0	0	0	0	2	1	86	87	1	0	0	1	0	0	0	1	60
4	0	0	0	0	2	14	34	435	1	0	0	1	387	0	0	2	18
5	0	0	0	0	2	32	787	1168	1	0	0	1	345	0	0	2	49
6	0	0	0	0	2	0	1497	2071	1	0	0	0	0	1	34	2	180
7	1	29	0	0	3	10	1049	3628	1	0	0	1	108	0	0	1	25

Raw Data | Clean Data | Sheet1 | Sheet3 | Sheet2

Figure 7.4: Next seventeen attribute columns containing values derived from Raw Data spreadsheet.

shown in Figure 7.2, it did not include *[TimeSpent]* on pages beyond row 34 when calculating *[Browsing Time]*. This was because row 35 showed that a page of type *Form* had been accessed and then sent (rows 36 and 37). If the page preceding a page of type *Form* (that was subsequently sent) was the **Contact Us** page then *[Browsing Time]*, did not include *[TimeSpent]* on the **Contact Us** page.

[Total Time on site] included *[TimeSpent]* on all pages browsed by a visitor without exception. In the example shown in Figure 7.2, the visitor continued to browse the site after sending an enquiry. Therefore, rows 24 to 42 were used to calculate *[Total Time on site]* for that visitor.

Once data had been transformed, it was cleaned a second time to remove errors that had not been detected in the first attempt. Using the filter feature available in Microsoft Excel, records with particular criteria were identified and removed from the data. In the data set for conversions the following criteria were identified and corresponding records were deleted:

- *[Total Time on site]* was equal to 0.
- *[Other]* was equal to 0. This highlighted unusual browsing patterns or an error in the tracking.
- Both *[Sent Form]* and *[ContactUs Sent]* were equal to 0. This highlighted unusual browsing patterns or an error in the tracking.

The following criteria were identified and corresponding records were deleted in the data set for non-conversions:

- *[Total Time on site]* was equal to 0.

- *[Other]* was greater than 0. This highlighted unusual browsing patterns or an error in the tracking where a **Sent** page had been accessed but no enquiry had been received or recorded by the OTM.

Once cleaned, the data contained 1431 conversion records and 23,348 non-conversion records. This represented a 6%-94% split between conversion and non-conversion. The data was used to create two new data sets, ***Training Data*** and ***Test Data*** that were used in Stage 4 to find and test models. Conversion records were split between ***Training Data*** and ***Test Data***. Each data set had a 50%-50% split between conversion and non-conversion records.

It was assumed that a training sample with a 50%-50% split between conversion and non-conversion records could produce more accurate models since prediction algorithms could develop equal knowledge about conversion and non-conversion with such a training sample.

It was also assumed that if a test sample that had a 6%-94% split was used, then it could be difficult to assess the accuracy of the models at predicting conversions since accuracy measures could be influenced by the models' prediction accuracy of a high number of non-conversions. To avoid this, a training sample with a 50%-50% split between conversion and non-conversion was used.

Training Data contained:

- A total of 1428 records.
- 714 records of visits that had converted, that is, had a value of 1 for *[Conversion]*.
- 714 records of visits that had not converted, that is had a value of 0 for *[Conversion]*.

Test Data contained:

- A total of 1434 records.
- 717 records of visits that had converted, that is had a value of 1 for *[Conversion]*.
- 717 records of visits that had not converted, that is had a value of 0 for *[Conversion]*.

7.3.4. Stage 4: Knowledge discovery

Data from Stage 3 was imported into PolyAnalyst to seek models to automatically determine whether a visitor would enquire based on the way that they interacted with the dynamic main website.

The knowledge discovery stage took place at the University of Portsmouth. ***Training Data*** from Stage 3 was imported into a data mining engine called PolyAnalyst. Some models were successfully found by the data mining engine. These models were then tested using ***Test Data*** from Stage 3. Section 7.5 describes the knowledge discovery stage in more detail.

7.4. Data Mining

Once data was retrieved, cleaned and transformed, it was imported into PolyAnalyst in order to generate models to predict whether a visit would generate a conversion.

PolyAnalyst 6 was a data mining tool that had a selection of algorithms that could perform automated learning and knowledge discovery operations. It was selected to perform data mining because it provided tools for importing, cleaning and manipulating data as well as pattern discovery, prediction and reporting (PolyAnalyst, n.d.). PolyAnalyst 6 provided algorithms for general statistical analysis; categorisation, predictive analysis and text analysis.

7.4.1. Prediction algorithms

Three algorithms were selected: “Linear Regression”, “Find Laws” and “Neural Networks”. “Linear Regression” and “Find Laws” produced human-readable models that could be used to predict target variables from a set of input attributes. Unlike the “Linear Regression” and “Find Laws” algorithms, which clearly displayed the involved terms and each term's impact (each term's coefficient), the output of “Neural Networks” could not be observed. The model that it created was essentially the weight of the connections between neurons (nodes) within that model.

The “Linear Regression”, “Find Laws” and “Neural Networks” algorithms are described as:

Linear regression

“Linear Regression” was one of the oldest and most frequently used methods of statistical prediction. “Linear Regression” analysed the relationship between variables and produced a linear regression model that could be used to understand dependency between variables and predict numerical values.

The output of “Linear Regression” was a linear solution. If the relationship was expected to be linear, then the algorithm provided high quality output. In cases where the relationship was not linear, “Linear Regression” was useful in identifying attributes that influenced the target attribute. These attributes could then be included in subsequent analyses with “Find Laws” and “Neural Networks” to find better models. PolyAnalyst’s multi-parametric stepwise linear regression worked with any number of attributes.

PolyAnalyst performed rigorous significance testing on linear regression models. It used two different methods to perform this. The first method was based on checking the value of F-ratio (see description in Section 7.4.2) of every term that was included in a model. If a term had a value of F-ratio less than a specified threshold (set to 3 for the research described in this dissertation), then this term was discarded from the model. The second method used randomised testing. While PolyAnalyst searched for the best regression model for the explored data set, it solved the same problem for several randomised data sets. Randomised data sets were created from the original data set using random permutation of the target attribute values for various records. PolyAnalyst considered the created regression model as significant only if accuracy obtained for the real data was much higher than for any randomised data. Otherwise, PolyAnalyst concluded that it did not have enough data to create a significant model.

Find Laws

PolyAnalyst's "Find Laws" was a proprietary, nonlinear regression algorithm that could be used for predictive analysis. "Find Laws" searched for multi-dimensional non-linear relations in data and presented discovered relations as explicit mathematical formulae. "Find Laws" algorithm generated miniature equations and then combined these together to form larger equations so as to produce a final equation that effectively modelled the dependence of a target attribute on a set of one or more independent attributes. PolyAnalyst performed rigorous significance testing on "Find Laws" models using randomised testing and created significant models only.

Neural Networks

"Neural Networks" are a well known and researched approach for modelling data and identifying patterns. In comparison to linear regression, a neural network is non-linear. Data was inputted into a neural network to train a model. The trained model was then applied to new records to assess the accuracy of the model. "Neural Networks" was computationally intensive and could take extended periods of time to train a model. The number of attributes used in “Neural Networks” could be minimised by using “Linear Regression” and “Find Laws” beforehand and then running “Neural Networks” with the attributes found to be of high relevance.

7.4.2. Accuracy measures

PolyAnalyst provided statistical indicators to assess the accuracy of models, such as standard error, r-squared and standard deviation. When “Neural Networks” was utilised, the statistical indicators for the accuracy of the models developed included classification probability, classification efficiency, classification error and classification failure.

Standard Deviation (σ, s or stdev) was a measure of the degree of variation of data from its mean value. A large standard deviation indicated that data points were far from the mean and spread over a large range of values. A small standard deviation indicated that they were close to the mean (Jordan and Smith, 2002). Equation 7.1 shows how the standard deviation for a given population was calculated. Equation 7.2 shows how the standard deviation for a sample of values from a larger population was calculated.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2}{N}}$$

Equation 7.1:Standard deviation.

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n-1}}$$

Equation 7.2: Standard deviation for a sample from a larger population.

Standard Error (stderr) was the standard deviation of the predicted values of a target variable with respect to the real values of the target variable. Equation 7.3 shows how the standard error was calculated. N was the number of data samples. P_i and p_i were the real and predicted values for the target variables in each sample.

$$stderr = \sqrt{\frac{\sum_{i=1}^{N}(p_i - P_i)^2}{(N-1)}}$$

Equation 7.3: Standard Error.

PolyAnalyst calculated standard error by dividing standard deviation by dispersion. Dispersion was the square of standard deviation. Values for standard error lay between 0 and 1. A value of 0 corresponded to an absolutely accurate prediction model.

R-Squared (RSq) was a measure of how well future values were likely to be predicted by a model. R-Squared was calculated as $1 - r^2$, where r was the standard error. Its values lay between 0 and 1, with a value of 1 indicating an absolutely accurate model.

F-ratio was used to determine how much predictive power independent attributes had in a linear relationship. F-ratio was calculated as the square of the ratio of a term's value to the term's standard deviation. If an independent attribute present in a linear regression model had a high F-ratio (greater than 3), it meant that the attribute was important in predicting the value of the model's target variable. The threshold value of F-ratio was set to 3 for the research work discussed in this dissertation.

Classification Probability (cp) was the number of times (expressed as a percentage) that a prediction algorithm made a correct classification. Equation 7.4 shows how the classification probability of an algorithm was calculated. $N_{correctA}$ and $N_{correctB}$ were the number of correctly classified records for classes A and B. N_A and N_B were the total number of records of class A and B.

$$cp = 100 \times \frac{N_{correctA} + N_{correctB}}{N_A + N_B}$$

Equation 7.4: Classification probability (Bergasa-Suso, 2005).

The percentage of visitors in each class (conversion and non-conversion) in ***Test Data*** determined the minimum accuracy required by the models that were found. For example, a naïve prediction rule that considered every visit as a non-conversion would have an accuracy of 50% if 50% of visitors in the dataset did not convert. Any algorithm that claimed to learn from a set of input data had to predict better than a naïve algorithm (Adriaans and Zantinge, 1996).

Classification Efficiency (ce) provided a measure of the accuracy of a prediction algorithm with respect to a naïve prediction algorithm . Equation 7.5 shows how the classification efficiency of a prediction algorithm was calculated. $N_{correctA}$ and $N_{correctB}$ were the number of correctly classified records for classes A and B. N_A and N_B were the total number of records of class A and B.

$$ce = 100 \times \frac{N_{correctA} + N_{correctB} - \max(N_A, N_B)}{\min(N_A, N_B)}$$

Equation 7.5: Classification Efficiency (Bergasa-Suso, 2005).

If 80% of records in a data set were from class A and 20% were from class B, then a naïve algorithm would classify all records as class A and would have a classification probability of 80%, but a classification efficiency of 0% (see Equation 7.6, ce_{naive}). A prediction algorithm that classified all records correctly would have a classification efficiency of 100% (see Equation 7.6, $ce_{accurate}$) (Bergasa-Suso, 2005).

$$ce_{naive} = 100 \times \frac{80 + 0 - 80}{20} = 0\% \quad ce_{accurate} = 100 \times \frac{80 + 20 - 80}{20} = 100\%$$

Equation 7.6: Classification efficiency for naïve and accurate classification rules (Bergasa-Suso, 2005).

Classification Error (cerr) was the percentage of times that a model made an incorrect prediction.

Classification Failure (cf) was the percentage of times that the prediction algorithm failed to make a prediction.

7.5. Experiment Results

7.5.1. Initial explorations

Initial explorations were carried out to predict whether a visit would result in a conversion. These explorations were carried out using PolyAnalyst's "Linear Regression", "Find Laws" and "Neural Networks". During these explorations, errors were found in the VBA macro that was used to transform the data retrieved from the database into the data that was imported into the PolyAnalyst algorithms. This introduced errors in the data as *[Total Time on site]* was calculated erroneously in instances where an enquiry was sent from a **Contact Us** page. Also, the VBA parser macro did not filter out visits to the collaborating company's online shop. The results that were obtained from these initial explorations can be seen in Appendix D.

7.5.2. First exploration - Using all attributes

The VBA parser macro was modified to filter out visits to the collaborating company's online shopping website and to include *[TimeSpent]* on a **Contact Us** page when calculating *[Total Time on site]* in instances where an enquiry was sent from a **Contact**

Us page. The first exploration of the data was carried out using “Linear Regression”, “Find Laws” and “Neural Networks”. The attributes used in the first exploration are shown in Table 7.1.

The target attribute was *[Conversion].* Since “Linear Regression” and “Find Laws” accepted numerical values only, *[Conversion]* was imported into these algorithms as integer values; conversion was represented by a value of 1 and non-conversion by a value of 0.

Attributes used in first exploration	
[Browsed ContactUs]	*[Media Access]*
[Browsed ContactUsTS]	*[Media AccessTS]*
[Browsing Time]	*[PR]*
[Case Studies]	*[PRTS]*
[Case StudiesTS]	*[Services]*
[Company Info]	*[ServicesTS]*
[Company InfoTS]	*[Total Time on site]*
[Download]	*[Unsent Form]*
[DownloadTS]	*[UnsentFormTS]*
[FTQuote]	*[Video]*
[FTQuote TS]	*[VideoTS]*
[Keyword Length]	

Table 7.1: Attributes used in the first exploration.

Linear Regression

“Linear regression” found prediction model LR1 using *Training Data*. Equation 7.7 shows model LR1.

$$Conversion = +0.387559 - 0.0335942 * [Company\ Info] - 0.0164385 * Services - 0.106221 * [Unsent\ Form\ ] - 0.000319465 * [Media\ AccessTS] - 0.000693281 * [Browsing\ Time] + 0.000784759 * [Total\ Time\ on\ site] - 0.104908 * [Browsed\ ContactUs] + 0.0388765 * FTQuote$$

Equation 7.7: Prediction model LR1.

The accuracy measures for the prediction model LR1 are shown Table 7.2.

StdErr	RSq	StdDev
0.80	0.36	0.40

Table 7.2: Accuracy measures for prediction model LR1 derived from *Training Data*.

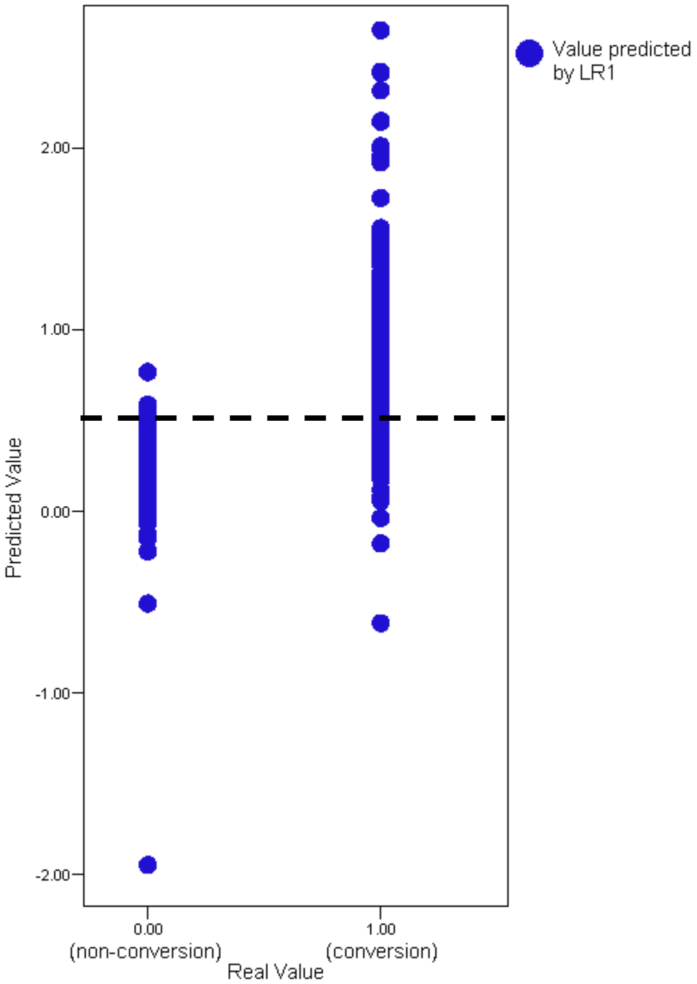
It can be seen from Table 7.2 that LR1 had a high standard error and low RSq value. This suggested that the model was not accurate. The model was tested with ***Test Data***. The accuracy measures are shown in Table 7.3. These confirmed the low accuracy of model LR1.

StdErr	RSq	StdDev
0.82	0.33	0.41

Table 7.3: Accuracy measures for LR1 after testing model with *Test Data*.

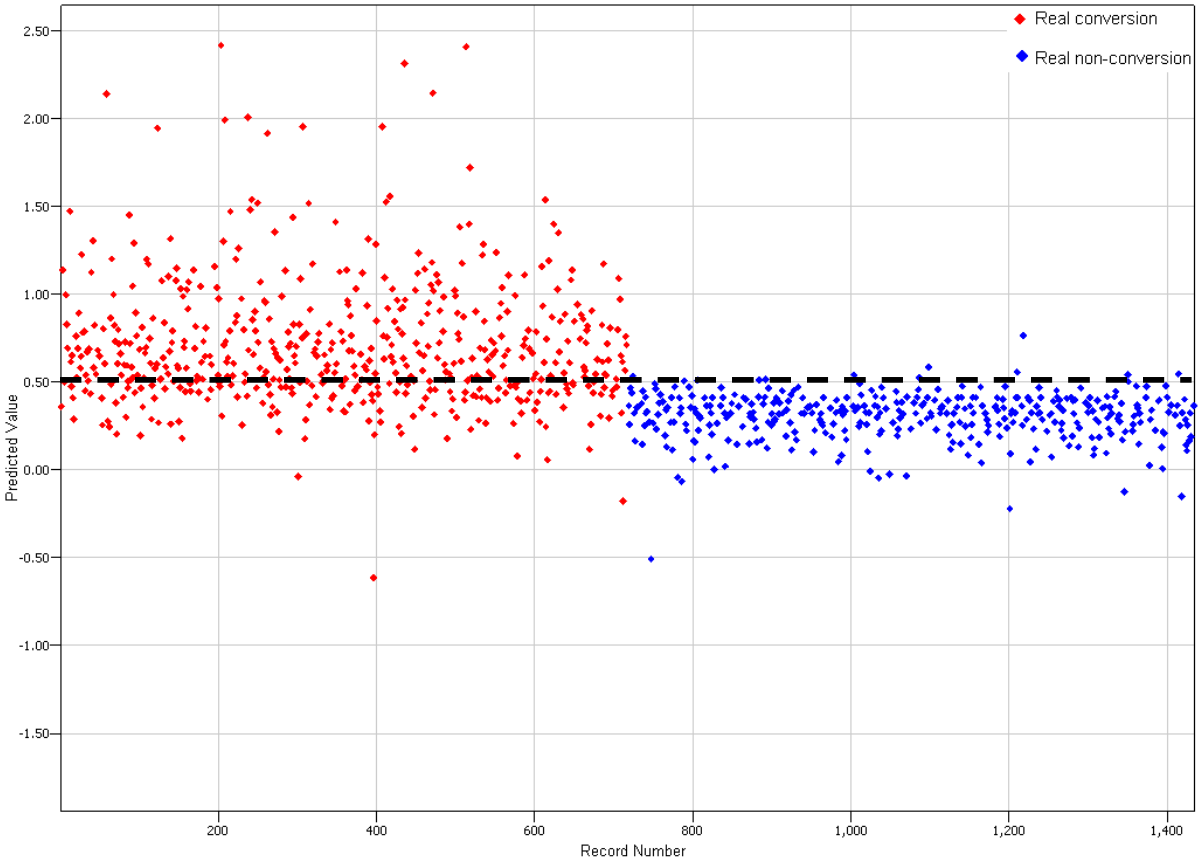
Graph 7.1 shows *[Conversion]* values predicted by LR1 plotted against real *[Conversion]* values in ***Test Data.*** The x-axis of Graph 7.1 shows the real value of *[Conversion]* for the records in ***Test Data***, that is 0 for non-conversion and 1 for conversion. The y-axis represents the value that LR1 predicted for *[Conversion]* for each record found in ***Test Data***. When the real *[Conversion]* value was 0 (non-conversion), the values predicted by LR1 ranged from -1.92 to 0.76. When the real *[Conversion]* value was 1 (conversion) the values predicted by LR1 range from -0.62 to 2.64. It was observed from Graph 7.1 that for real *[Conversion]* values of 1 most of the predicted *[Conversion]* values were greater than 0.50 and that most of the predicted

[Conversion] values for real non-conversions was less than 0.50. Therefore, a boundary value of 0.50 (shown as a dotted line on Graph 7.1) was considered.



Graph 7.1: *[Conversion]* value predicted using LR1 vs real *[Conversion]* value (using *Test Data*).

Graph 7.2 shows the value predicted by LR1 for each record in ***Test Data***. The x-axis of the graph shows the record number for each record found in ***Test Data***. The y-axis shows the *[Conversion]* values predicted by LR1. The red dots on the graph represent records whose real *[Conversion]* value was 1(conversion) while the blue dots represent records whose real *[Conversion]* value was 0 (non-conversion).



Graph 7.2: *[Conversion]* values predicted by LR1 for individual records found in *Test Data*.

Graph 7.2 confirmed the observations made from Graph 7.1, whereby most of the *[Conversion]* values predicted by LR1 for records that had a real *[Conversion]* value of 0 were less than 0.50 and most of the *[Conversion]* values predicted by LR1 for records that had a real *[Conversion]* value of 1 were greater than 0.50. Therefore, a boundary value of 0.50 (shown as a dotted line on Graph 7.1 and Graph 7.2) was applied.

A confusion matrix for LR1 (using ***Test Data***) with a boundary value for identifying conversions set at 0.50 is shown in Table 7.4. The classification probability and efficiency of LR1 were derived from the values shown in Table 7.4. They were 84.59% and 69.18% respectively. The model predicted 70.99% of the target 1s correctly, which suggested good accuracy. Also Table 7.4 shows that LR1 (with a boundary value set at

0.50) predicted that 522 records were conversions. Out of these predictions, 97.51% (509) were correct.

Actual \ Predicted	> 0.50 (conversion)	≤ 0.50 (non-conversion)	Total
1 (conversion)	509	208	717
0 (non-conversion)	13	704	717
Total	522	912	1434

Table 7.4: Confusion matrix for LR1 with boundary set to 0.50 (using *Test Data*).

Find Laws

"Find Laws" found prediction model FL1 (see Equation 7.8) using ***Training Data***.

$$\begin{aligned} Conversion \;=\; & (-3.64356e-009 * [Total\ Time\ on\ site] * [Total\ Time\ on\ site] \\ & + 1.00094 * [Total\ Time\ on\ site] - 1.00093 * [Browsing\ Time] \\ & - 3.81618e-012 * [Browsed\ ContactUsTS] * [Browsed\ ContactUsTS] \\ & * [Total\ Time\ on\ site])/([Total\ Time\ on\ site] - 0.999726 \\ & * [Browsing\ Time] + 0.0248078) \end{aligned}$$

Equation 7.8: Prediction model FL1.

The accuracy measures for the prediction model FL1 are shown in Table 7.5. The model was tested with ***Test Data***. The accuracy measures are shown in Table 7.6. FL1 had a high RSq value and low standard error. This suggested that the model had high accuracy.

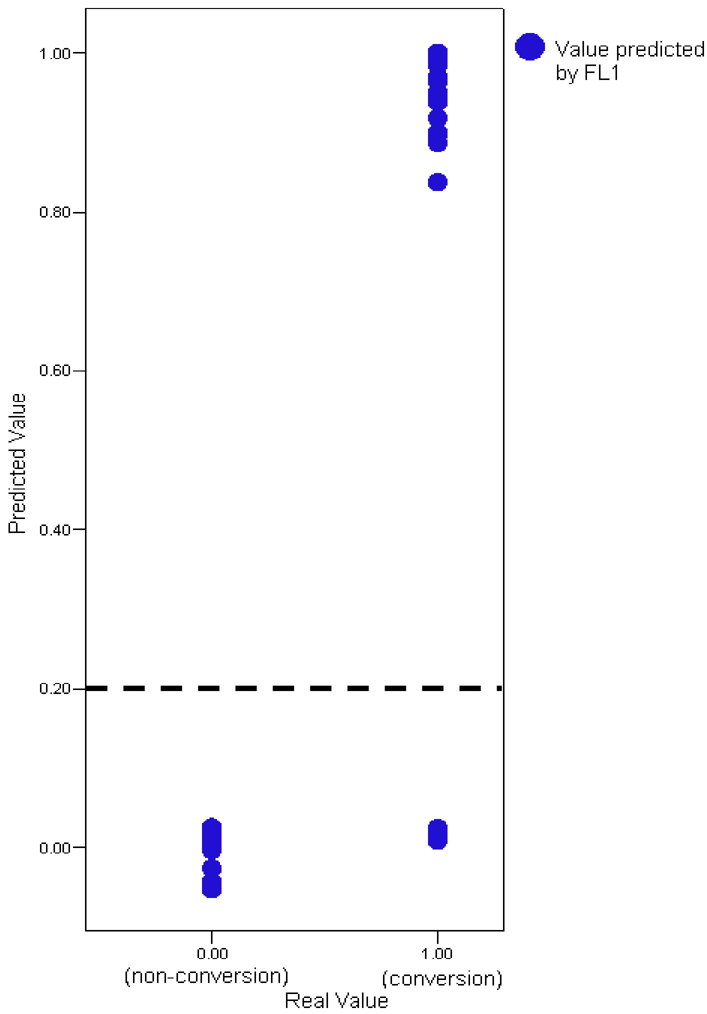
StdErr	RSq	StdDev
0.13	0.98	0.06

Table 7.5: Accuracy measures for prediction model FL1 derived from *Training Data*.

StdErr	RSq	StdDev
0.16	0.98	0.08

Table 7.6: Accuracy measures for FL1 after testing model with *Test Data*.

Graph 7.3 shows *[Conversion]* values predicted by FL1 plotted against real *[Conversion]* values found in ***Test Data.***

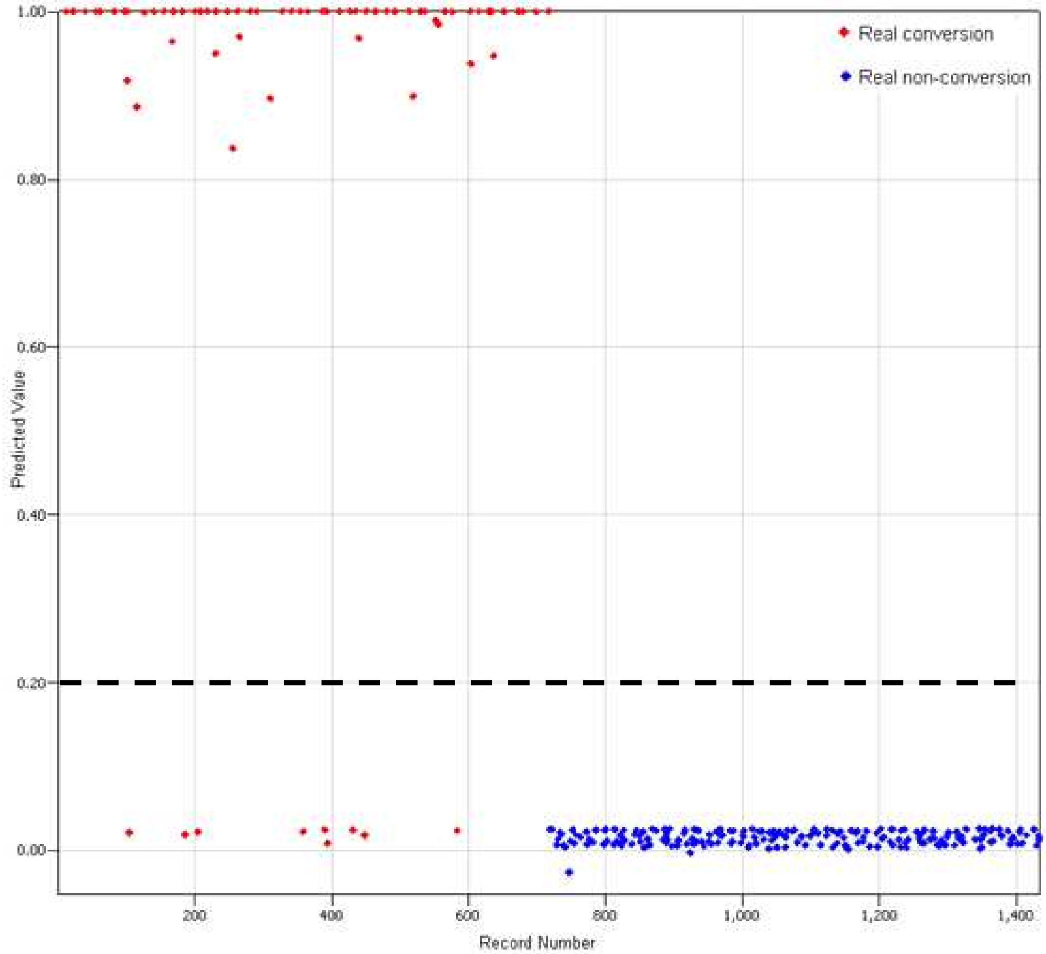


Graph 7.3: *[Conversion]* value predicted using FL1 vs A real *[Conversion]* value (using *Test Data*).

The x-axis of Graph 7.3 shows the real value of *[Conversion]* for the records in ***Test Data****,* that is 0 for non-conversion and 1 for conversion. The y-axis represents the value that FL1 calculated for *[Conversion].* When the real *[Conversion]* value was 0 (non-conversion), the values predicted by FL1 ranged from -0.05 to 0.04. When the real

[Conversion] value was 1 (conversion), the values predicted by FL1 ranged from 0.01 to 1. A boundary value could be set anywhere between 0.05 to 0.80. An arbitrary value of 0.20 (shown as a dotted line on Graph 7.3) was considered as boundary value.

Graph 7.4 shows the *[Conversion]* value predicted by FL1 for each record found in ***Test Data***.



Graph 7.4: *[Conversion]* values predicted by FL1 for individual records found in *Test Data*.

The x-axis of the graph shows the record number for each record found in ***Test Data***. The y-axis shows the *[Conversion]* values predicted by FL1. The red dots on the graph represent records whose real *[Conversion]* value was 1 (conversion) while the blue dots represent records whose real *[Conversion]* value was 0 (non-conversion). Graph 7.4

confirmed the observations made from Graph 7.3, whereby most of the *[Conversion]* values predicted by FL1 for records that had a real *[Conversion]* value of 0 were less than 0.20 and most of the *[Conversion]* values predicted by FL1 for records that had a real *[Conversion]* value of 1 were greater than 0.20. Therefore, a boundary value of 0.20 (shown as a dotted line on Graph 7.3 and Graph 7.4) was set.

A confusion matrix for FL1 is shown in Table 7.7 when the boundary value for identifying 1s was set to 0.20.

Predicted / Actual	**> 0.20 (conversion)**	**≤ 0.20 (non-conversion)**	**Total**
1 (conversion)	708	9	717
0 (non-conversion)	0	717	717
Total	708	726	1434

Table 7.7: Confusion matrix for FL1 with boundary was set to 0.20.

The classification probability and efficiency of FL1 were derived from the values shown in Table 7.7. They were 99.37% and 98.74% respectively. The model predicted 98.74% of the target 1s correctly which suggested good accuracy. Also Table 7.7 shows that FL1 (with a boundary value set at 0.20) predicted that 708 records were conversions. This represented 100% prediction accuracy in the prediction set.

FL1 appeared to be a much better model than LR1 with higher classification probability, efficiency and percentage of correct predictions for conversion in both the target and predicted sets. Table 7.8 shows the accuracy measures for FL1 and LR1.

Model	cp %	ce %	Accuracy of conversion prediction in target set %	Accuracy of conversion prediction in predicted set %
LR1	84.59	69.18	70.99	97.51
FL1	99.37	98.74	98.74	100.00

Table 7.8: Accuracy measures of LR1 and FL1.

Neural Network

Unlike “Linear Regression” and “Find Laws”, “Neural Networks” did not provide symbolic rules as output. The model was the weight of the connections between neurons within the model. PolyAnalyst calculated classification probability, classification efficiency, classification error and classification failure for “Neural Networks” models. These accuracy measures were used to assess the accuracy of the models developed. These accuracy measures were described in Section 7.4.

PolyAnalyst’s “Neural Networks” algorithm accepted either Boolean or binary values as target values. PolyAnalyst generated accuracy measures such as classification probability (cp), classification efficiency (ce), classification error (cerr), classification failure (cf) and confusion matrices only if the target value was Boolean. These measures were considered to be better for assessing the quality of “Neural Networks” models than the measures (standard error, r-squared and standard deviation) that would have been derived if binary target values had been used. Therefore, this research used Boolean target values instead of binary values when utilising “Neural Networks”. The Boolean values were set as Yes for conversion and No for non-conversion.

“Neural Networks” was given the same input parameters as LR1 and FL1. ***Training Data*** was used to derive model NN1, which was then tested with ***Test Data***. The accuracy measures for the tested model NN1 is shown in Table 7.9.

cerr	cp	cf	ce
3.56%	96.44%	0%	92.89%

Table 7.9: Accuracy measures for NN1 when tested with *Test Data*.

NN1 had a classification probability that was higher than a naïve model (cp =50%) as well as a high classification efficiency. Table 7.10 provides a breakdown of the model's predictions. It can be seen that NN1 predicted 94.14% of Yes correctly in the target set. The confusion matrix for NN1 is shown in Table 7.11.

Target	No of records	Error %	Correct%	Undefined%
Yes (conversion)	717	5.86	94.14	0.00
No (non-conversion)	717	1.26	98.74	0.00
Total	1434	3.56	96.44	0.00

Table 7.10: Breakdown of predictions for NN1 when tested with *Test Data*.

From Table 7.11, it can be seen that model NN1 predicted a total of 684 records as conversions. Out of these predictions, 98.68% (675) were correctly predicted.

Actual \ Predicted	Yes (conversion)	No (non-conversion)	Total
Yes (conversion)	675	42	717
No (non-conversion)	9	708	717
Total	684	750	1434

Table 7.11: Confusion matrix for NN1 when tested with *Test Data*.

Summary of results obtained from first exploration

The first exploration used all available predictor attributes as input data to find models using "Linear Regression", "Find Laws" and "Neural Networks". "Linear Regression" and

“Find Laws” found models LR1 and FL1 that are represented by Equation 7.7 and Equation 7.8 respectively. “Neural Networks” found model NN1. Table 7.12 summarises the classification accuracy for LR1, FL1 and NN1.

Model	cp %	ce %	Accuracy of conversion prediction in target set %	Accuracy of conversion prediction in predicted set %
LR1	84.59	69.18	70.99	97.51
FL1	99.37	98.74	98.74	100.00
NN1	96.44	92.89	94.14	98.68

Table 7.12: Accuracy measures for LR1, FL1 and NN1.

From the measures shown in Table 7.12, it was concluded that FL1 was the best model as it had the highest classification efficiency, classification probability and prediction accuracy in both the target and predicted sets. NN1 was the second most accurate model while LR1 was the least accurate.

7.5.3. Second exploration – Removing *[Total Time on site]* as an attribute

In the second exploration, *[Total Time on site]* was removed as an attribute from the data. *[Total Time on site]* made it easy for models LR1, FL1 and NN1 to predict conversions as on average *[Total Time on site]* for visits that generated conversions tended to be higher when compared to visits that did not generate conversions. Also, this research was trying to find a model that could predict conversions based on data about the behaviour of website visitors prior to an action that represented a conversion. Since visitors could continue browsing a website after they had converted, *[Total Time on site]* was not a suitable attribute. The attributes used in the second exploration are shown in Table 7.13.

Attributes used in second exploration	
[Browsed ContactUs]	*[Keyword Length]*
[Browsed ContactUsTS]	*[Media Access]*
[Browsing Time]	*[Media AccessTS]*
[Case Studies]	*[PR]*
[Case StudiesTS]	*[PRTS]*
[Company Info]	*[Services]*
[Company InfoTS]	*[ServicesTS]*
[Download]	*[Unsent Form]*
[DownloadTS]	*[UnsentFormTS]*
[FTQuote]	*[Video]*
[FTQuote TS]	*[VideoTS]*

Table 7.13: Attributes used in second exploration.

Linear Regression

"Linear regression" found prediction model LR2 (see Equation 7.9) using ***Training Data***.

$$\begin{aligned} Conversion \ = \ &+0.529064 \ - 0.0142550 * Services \ - 0.137899 * [Unsent\ Form\ ] \\ &- 0.0920242 * [Media\ Access] + 0.000167695 * [Browsing\ Time] \\ &- 0.118354 * [Browsed\ ContactUs] + 0.0515357 * FTQuote \end{aligned}$$

Equation 7.9: Prediction model LR2.

The accuracy measures for the prediction model LR2 are shown in Table 7.14.

StdErr	RSq	StdDev
0.97	0.06	0.49

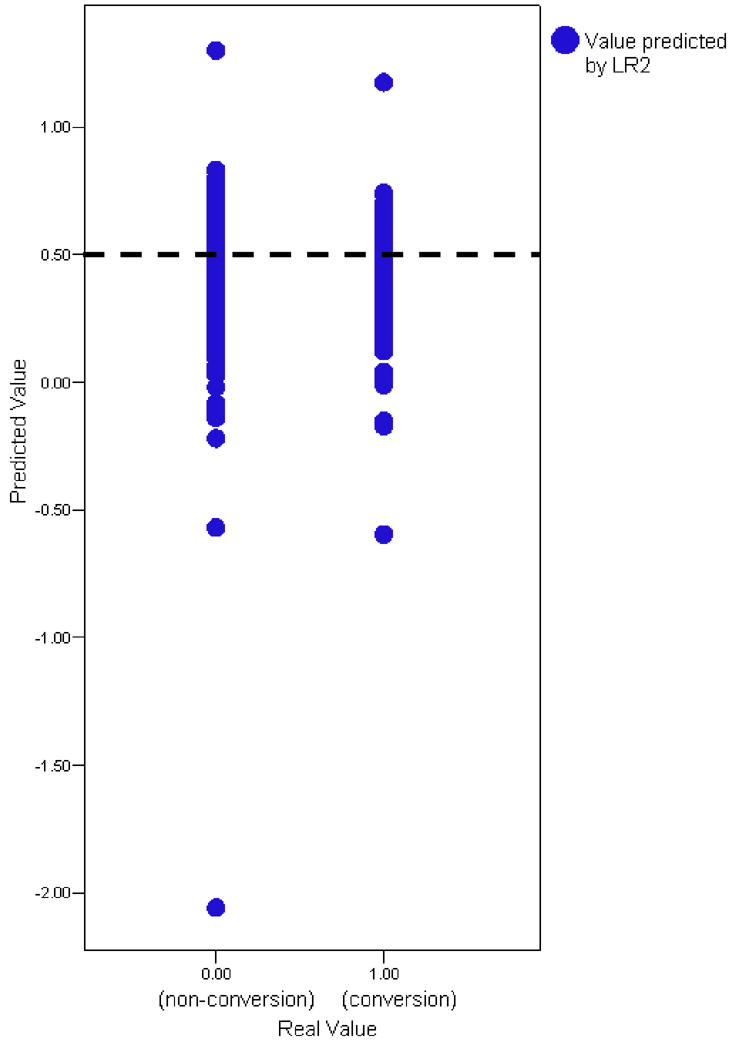
Table 7.14: Accuracy measures for prediction model LR2 using *Training Data*.

It can be seen from Table 7.14 that LR2 had a high standard error and low RSq value. This suggested that the rule was not accurate at predicting *[Conversion]*. The accuracy measures obtained when LR2 was tested with ***Test Data*** are shown in Table 7.15

StdErr	RSq	StdDev
1.00	-0.01	0.50

Table 7.15: Accuracy measures for prediction model LR2 using *Test Data*.

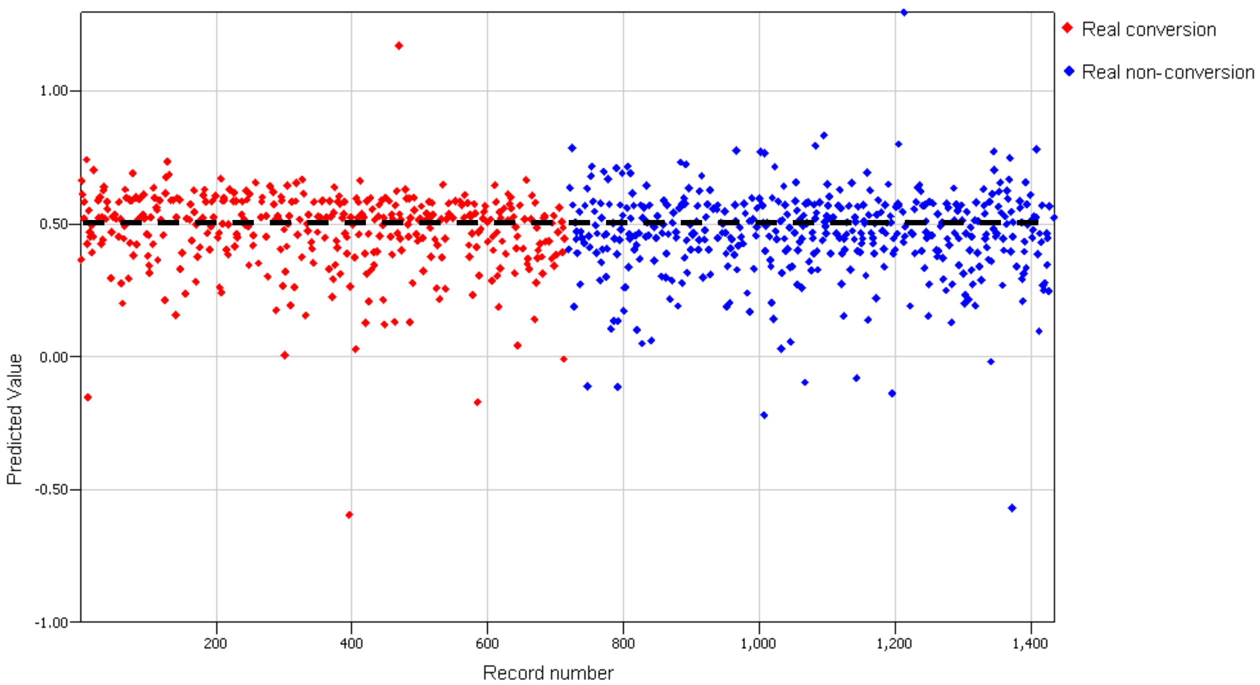
Standard error was 1 and the Rsq was -0.01. This suggested that model LR2 was poor at predicting *[Conversion]*. Graph 7.5 shows *[Conversion]* values predicted by LR2 plotted against actual *[Conversion]* values, using ***Test Data.***



Graph 7.5: *[Conversion]* value predicted using LR2 vs real *[Conversion]* value (using *Test Data*).

The x-axis of Graph 7.5 shows the real value of *[Conversion]* for the records in ***Test Data***, that is 0 for non-conversion and 1 for conversion. The y-axis represents the values that LR2 calculated for *[Conversion]* using Equation 7.9 and ***Test Data***. When the real value was 0 (non-conversion), the values predicted by LR2 ranged from -2.06 to 1.3. When the real value was 1 (conversion), the values predicted by LR2 ranged from -0.60 to 1.22. It was observed from Graph 7.5 that the range of predicted values overlapped for most of the data. A boundary value of 0.50 was considered. A lower value could cause LR2 to predict the value of a high number of real non-conversions as conversions and a value higher than 0.50 could predict fewer real conversions as conversions.

Graph 7.6 shows the *[Conversion]* values predicted by LR2 for individual records found in ***Test Data***. The x-axis of the graph shows the record number for each record found in ***Test Data***. The y-axis shows the *[Conversion]* values predicted by LR2.



Graph 7.6: *[Conversion]* values predicted by LR2 for individual records found in *Test Data*.

The red dots on the graph represent records whose real *[Conversion]* value was 1 (conversion) while the blue dots represent records whose real *[Conversion]* value was 0 (non-conversion). Graph 7.6 confirmed the observations made from Graph 7.5 whereby most of the predicted values were within the same range of values, thus making it hard to find an obvious boundary value. A boundary value of 0.50 was chosen (shown as a dotted line on Graph 7.5 and Graph 7.6).

A confusion matrix for LR2 (using ***Test Data***) when a boundary value for predicting *[Conversion]* was set to 0.50 is shown in Table 7.16. The classification probability and efficiency of LR2 were derived from the values shown in Table 7.16. They were 61.85% and 23.71% respectively. The model predicted 70.15% of the target 1s correctly. Also Table 7.7 shows that LR2 (with a boundary set at 0.50) predicted that 836 records were conversions. Out of these predictions, 60.17% (503) were correct.

Predicted / Actual	> 0.50 (conversion)	≤ 0.50 (non-conversion)	Total
1 (conversion)	503	214	717
0 (non-conversion)	333	384	717
Total	836	598	1434

Table 7.16: Confusion matrix for LR2 with boundary value set to 0.5 (using *Test Data*)

Find Laws

The "Find Laws" algorithm was unable to find a model.

Neural Networks

The accuracy measures for model NN2 generated by "Neural Networks" and tested using ***Test Data*** is shown in Table 7.17.

cerr	cp	cf	ce
30.68%	69.32%	0%	38.63%

Table 7.17: Accuracy measures for model NN2.

NN2 had a classification probability that was lower than NN1 but still higher than a naïve prediction model (cp =50%) as well as a high classification efficiency. Table 7.18 provides a breakdown of the model's predictions. It can be seen that NN2 predicted 65.69% of conversions correctly. The confusion matrix for NN2 is shown in Table 7.19.

Target	No of records	Error %	Correct%	Undefined%
Yes (conversion)	717	34.31	65.69	0.00
No (non-conversion)	717	27.06	72.94	0.00
Total	1434	30.68	69.32	0.00

Table 7.18: Breakdown of predictions for model NN2.

From Table 7.19, it can be seen that model NN2 predicted a total of 665 records as conversions. Out of these predictions, 70.83% (471) were correct.

Predicted / Actual	Yes (conversion)	No (non-conversion)	Total
Yes (conversion)	471	246	717
No (non-conversion)	194	523	717
Total	665	769	1434

Table 7.19: Confusion matrix for NN2.

Summary of results obtained from second exploration

The second exploration produced a "Linear Regression" model LR2 and a "Neural Network" model NN2. The "Find Laws" algorithm was unable to find a model to predict *[Conversion]*. LR2 had poor prediction accuracy as indicated by an RSq of -0.01. "Neural Networks" generated a model which had a classification probability of 69.32%.

This was lower than the classification probability of NN1 but higher than that of a naïve prediction model (50%). Table 7.20 summarises the accuracy measures of the models found in the first and second explorations.

Model	cp %	ce %	Accuracy of conversion prediction in target set (%)	Accuracy of conversion prediction in predicted set (%)
LR1	84.59	69.18	70.99	97.51
FL1	99.37	98.74	98.74	100.00
NN1	96.44	92.89	94.14	98.68
LR2	61.85	23.71	70.15	60.17
FL2	-	-	-	-
NN2	69.32	38.63	65.69	70.83

Table 7.20: Summary of accuracy measures of models found in the first and second explorations.

7.5.4. Third exploration – Using attributes with high F-Ratio

Based on the results of the second exploration, it was assumed that "Neural Networks" was better than "Linear Regression" and "Find laws" at finding predictive models and non-linear dependencies in the data used in this research. However, "Linear regression" provided a method of determining how much predictive power each attribute had in a linear relationship based on F-ratio. PolyAnalyst used F-ratio (described in Section7.4) to determine how much predictive power independent attributes had in a linear relationship. F-ratio was calculated as the square of the ratio of a term's value to the term's standard deviation.

In order to determine whether the attributes identified by "Linear Regression" could be used to generate a model that was more accurate than NN2, "Neural Networks" was run with the attributes of LR2 (see Equation 7.9). Table 7.21 shows the F-ratio for the attributes of LR2.

Attribute	F-ratio
[Services]	4.82
[FTQuote]	7.20
[Media Access]	14.57
[Browsing Time]	12.29
[Browsed ContactUs]	19.52
[Unsent Form]	29.31

Table 7.21: F-Ratio for attributes of LR2.

Model NN3 was generated by "Neural Networks" using the attributes shown in Table 7.21 as input. NN3 was tested with ***Test Data*** and its accuracy measures are shown in Table 7.22.

cerr	cp	cf	ce
30.61%	69.39%	0%	38.77%

Table 7.22: Accuracy measures for model NN3.

When comparing the accuracy measures of NN2 (see Table 7.17) and NN3 (see Table 7.22), it was observed that there was no significant improvement in classification probability or classification efficiency. A breakdown of the predictions made by NN3 is shown in Table 7.23.

Target	No of records	Error %	Correct%	Undefined%
Yes (conversion)	717	39.19	60.81	0.00
No (non-conversion)	717	22.04	77.96	0.00
Total	1502	30.61	69.39	0.00

Table 7.23: Breakdown of predictions for model NN3.

From Table 7.18 and Table 7.23, it can be seen that NN2 was better at predicting conversions (Yes) in the target set than NN3. Table 7.24 shows the confusion matrix for NN3.

Actual \ Predicted	**Yes (conversion)**	**No (non-conversion)**	**Total**
Yes (conversion)	436	281	717
No (non-conversion)	158	559	717
Total	594	840	1434

Table 7.24: Confusion matrix for model NN3.

Table 7.24, shows that model NN3 predicted a total of 594 records as conversion. Out of these predictions, 73.40% (436) were actually correct. NN3 was slightly better than model NN2 at predicting conversions in the predicted set.

Summary of results obtained from third exploration

The third exploration did not produce a model that was significantly more accurate than NN2. It appeared that using “Linear Regression” as a method for selecting attributes for “Neural Networks” could not produce more accurate NN models. The patterns that existed in the data may have been too complex for an algorithm such as “Linear regression” to identify.

7.5.5. Fourth exploration – Search keyword length

Web users’ perceptions and behaviours were governed by their motives for visiting a website (Rodgers et al., 2007). Pavlou and Fygenson (2006) found that the intention of buying a product occurred before the intention of acquiring information on a product.

Jansen, Booth and Spink (2008) suggested that a search query was one way in which Web users expressed intent.

Hölscher and Strube (2000) suggested that when Web users searched for information online they usually went through a process of typing a search term into a search engine and browsing websites returned by the search engine. If they did not find what they were looking for, Web users reformulated their query by changing one word for another, or adding or subtracting words (Jansen et al., 2000) and went through the search process again.

This research theorised that during the search process, Web users went through an experiential stage where they looked for information and tried to refine their search terms so as to find relevant information. During this stage, Web users were focused on research and were not ready to convert. Once Web users felt that they had enough or the right information, they would enter a goal-oriented stage. At this stage Web users were ready to convert and browsed websites with the intention to convert.

This research proposed the following hypotheses:

H1: Longer search keywords (containing 3 words or more) indicated that Web users were more ready to convert.

H2: Shorter search keywords (containing less than 3 words) indicated that Web users were less ready to convert.

H3: Visitors' search terms could indicate intent.

H4: Goal-oriented visitors were more likely to convert than experiential visitors.

Based on these hypotheses the research proposed a search-conversion model shown in Figure 7.5.

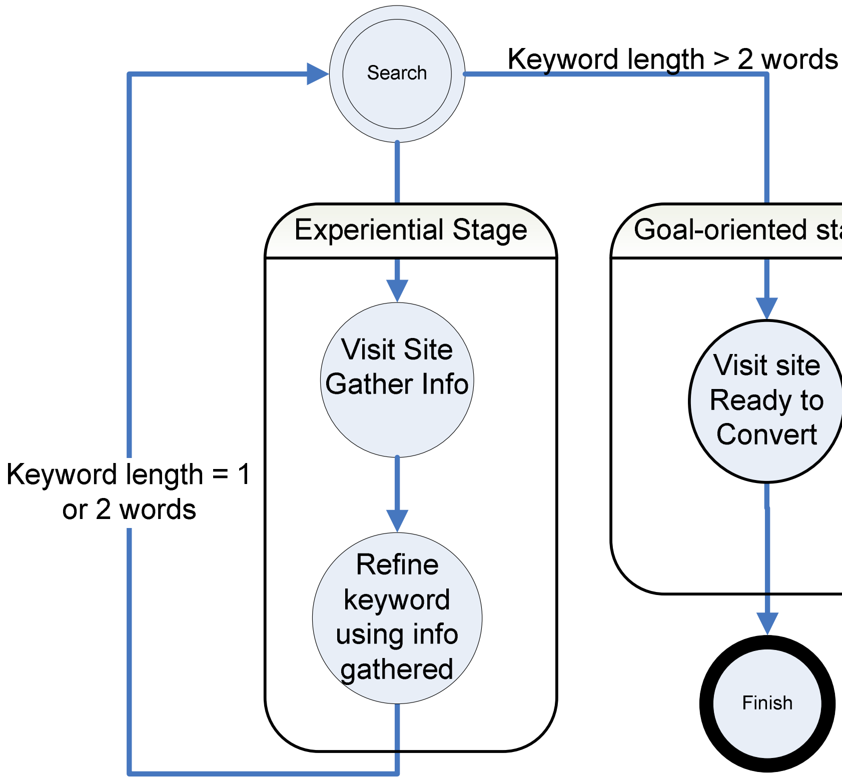
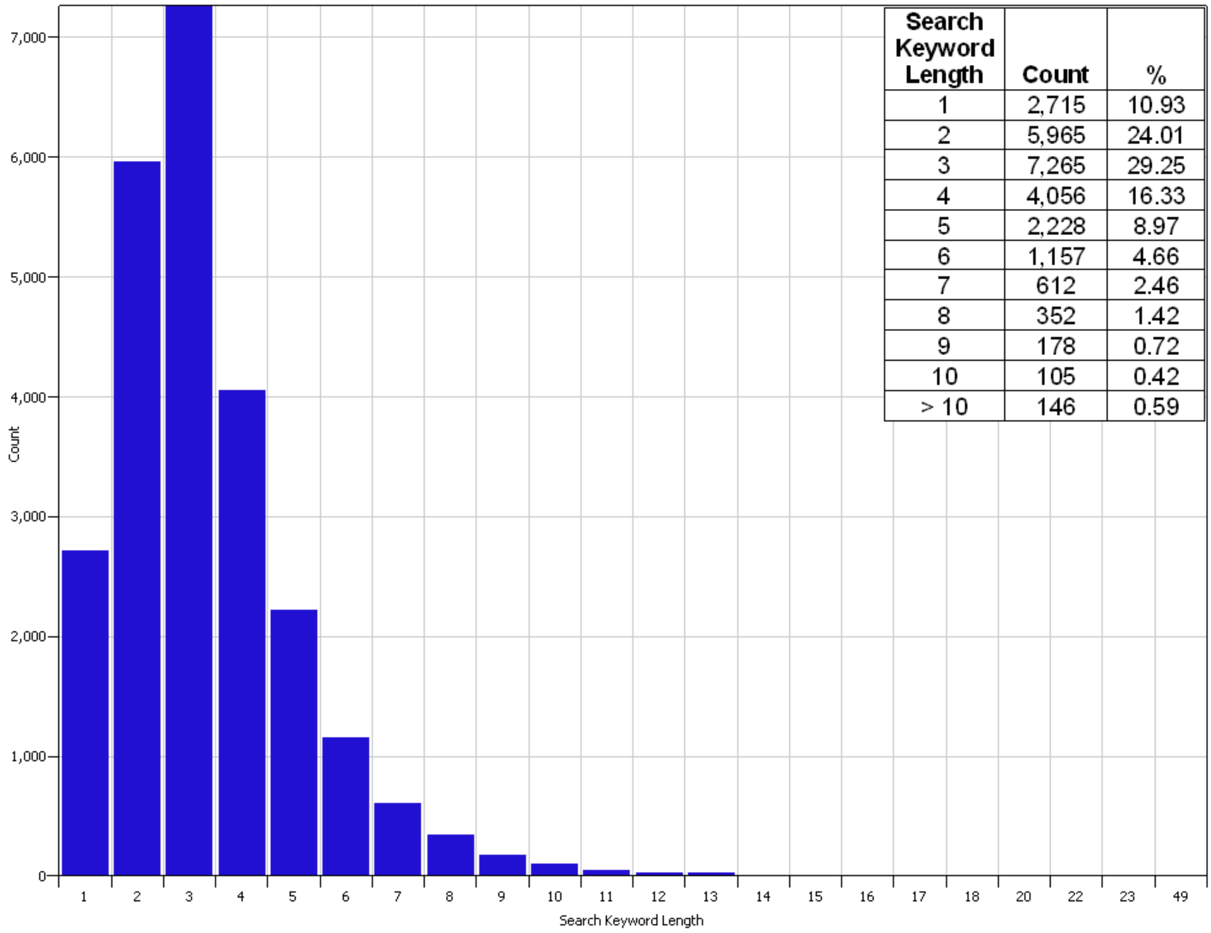


Figure 7.5: Search-conversion model.

It was assumed that Web users who were at the experiential stage were at the beginning of the search process and would use one or two word search keywords. Web users who had reached the goal-oriented stage were likely to have more specific and therefore longer search keywords, that is, search keywords containing more than 2 words. Therefore, it was hypothesised that the length of a search keyword could indicate whether a visitor was ready to convert.

The search keyword length distribution for the data collected by the Online Tracking Module (OTM) described in Chapter 4 was analysed. The analysis used the conversion and non-conversion data sets generated in Step 3 of the data mining process (described in Section 7.3.3). Graph 7.7 shows the distribution of search keyword length in this data.



Search Keyword Length	Count	%
1	2,715	10.93
2	5,965	24.01
3	7,265	29.25
4	4,056	16.33
5	2,228	8.97
6	1,157	4.66
7	612	2.46
8	352	1.42
9	178	0.72
10	105	0.42
> 10	146	0.59

Graph 7.7: Search keyword length vs frequency of occurrence (count).

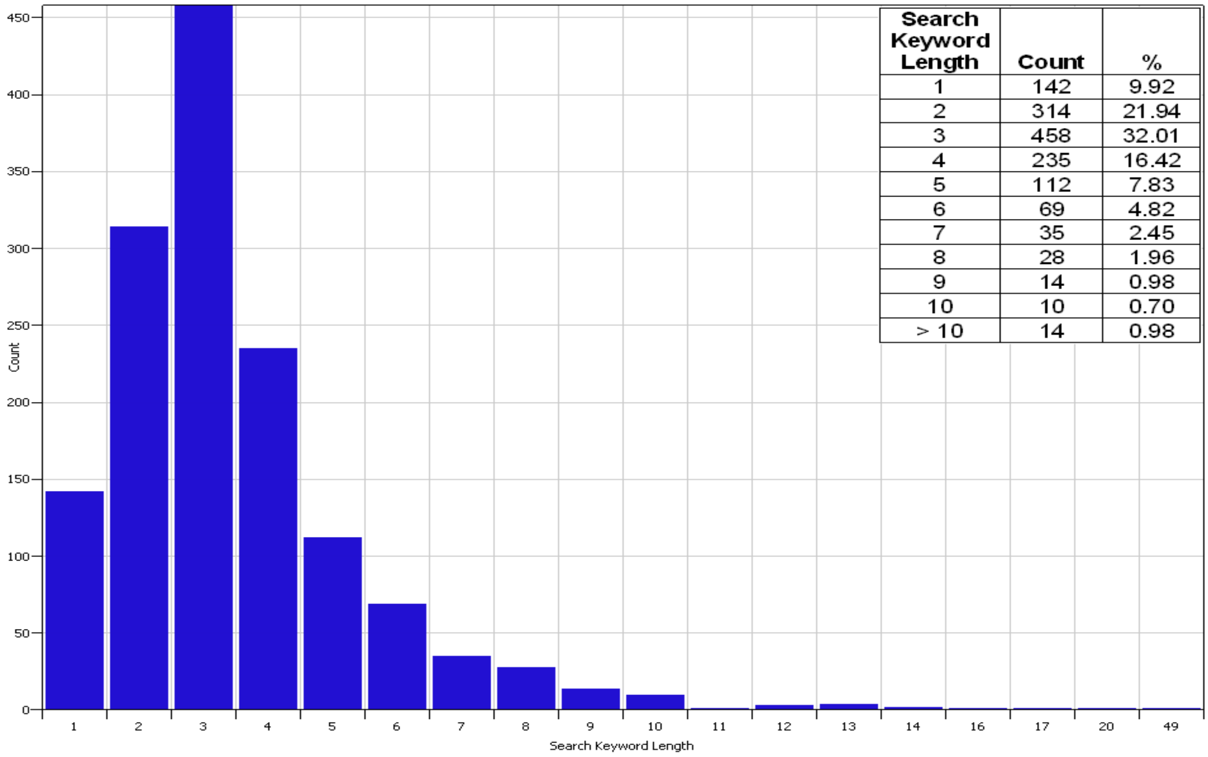
It can be seen from the graph that search keywords that contained two or three words had the highest occurrence followed by search keywords that contained four words. Search keyword length appeared to decay exponentially from search keyword length three onwards. That part of the graph appeared to follow the long tail distribution. The long tail is a "*colloquial name given to a product distribution curve at the long-tail end because the demand for the products is low. This technique has helped Amazon and Netflix satisfy customers' demands for obscure products that traditional stores would not stock*" (Phan and Vogel, 2010). The long tail concept has been applied in retail marketing to describe niche markets, where large number of unique or obscure products is sold in small amounts. The concept has been also applied to search keywords. Search keywords could be categorised as:

- Short-tail search keywords: One or two word keyword phrases which were very popular and generated high volume search and traffic.

- Long-tail keywords: "*Long-tail keywords are those low-volume, obscure, infrequently searched-for keywords*" (Mitchell, 2009). Mahaney (2006) describes long-tail keywords as "*three and four keyword phrases which are very, very specific to whatever you are selling*."

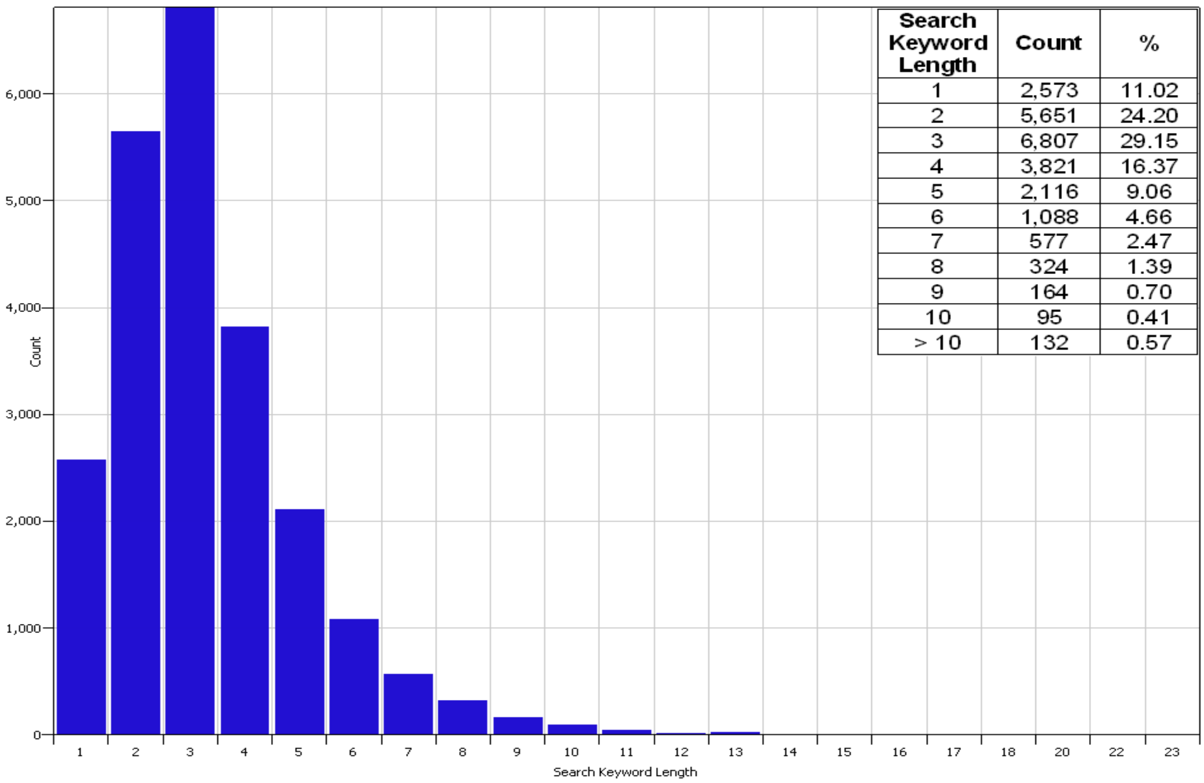
There was no formal definition of the length of a short-tail or long-tail search keyword. Search keywords used in Pay Per Click (PPC) advertising campaigns tended to have a natural bias towards a certain length that depended on the products or services advertised. Therefore, the length of short-tail and long-tail search keywords could vary. The PPC campaigns (described in Chapter 5) created during this research to drive traffic to the websites described in Chapter 4 were biased towards search keywords that contained one and two words.

Graph 7.8 shows the search keyword length distribution for visits that had generated an email enquiry (equivalent to a conversion in this research). Graph 7.8 follows the same kind of distribution as Graph 7.7. Most conversions (32.01%) were generated from three word search keywords. The distribution peaked at search keyword length three then decayed exponentially.



Search Keyword Length	Count	%
1	142	9.92
2	314	21.94
3	458	32.01
4	235	16.42
5	112	7.83
6	69	4.82
7	35	2.45
8	28	1.96
9	14	0.98
10	10	0.70
> 10	14	0.98

Graph 7.8: Search keyword length distribution for visits that had generated a conversion.



Search Keyword Length	Count	%
1	2,573	11.02
2	5,651	24.20
3	6,807	29.15
4	3,821	16.37
5	2,116	9.06
6	1,088	4.66
7	577	2.47
8	324	1.39
9	164	0.70
10	95	0.41
> 10	132	0.57

Graph 7.9: Histogram of keyword length vs frequency of occurrence (count) for visits that did not convert.

There were more conversions originating from search keywords containing four words (16.42%) than one word. When the distribution was examined on either side of the highest point on the graph (search keyword length = 3), it was observed that 31.86% of conversions were associated with search keywords that contained less than three words, while 68.14% of conversions were associated with search keywords that contained three words or more. This appeared to support Hypothesis 1 (H1) which suggested that visitors who used search keywords containing 3 words or more were more ready to convert.

The distribution of search keywords that did not produce conversions was examined. This is shown in Graph 7.9.

It can be seen that the shape of the distribution shown in Graph 7.9 is similar to that of Graph 7.7 and Graph 7.8. It was observed that three word search keywords accounted for 29.15% of visits that did not convert. The distribution peaked at search keyword length three then decayed exponentially. There were more non-conversions originating from search keywords containing four words (16.37%) than from one word search keywords (11.02%).

When the distribution was examined on either side of the highest point on the graph (search keyword length = 3), it was observed that 35.22% of non-conversions were associated with search keywords containing less than three words, while 35.63% of non-conversions were associated with search keywords containing more than three words. 64.78% of non-conversions were associated with search keywords containing three or more words.

Table 7.25 compares the occurrence of search keyword length in the conversion and non-conversion data sets. From Table 7.25, it was observed that one and two word search keywords were associated more non-conversions (35.22%) than conversions (31.86%). This appeared to support Hypothesis 2 (H2), which suggested that visitors who used search keywords containing less than 3 words were less ready to convert.

Search Keyword length	% in Conversion data	% in Non-conversion data	Difference
1	9.92	11.02	-1.10
2	21.94	24.20	-2.26
3	32.01	29.15	2.86
4	16.42	16.37	0.05
5	7.83	9.06	-1.12
6	4.82	4.66	0.16
7	2.45	2.47	-0.02
8	1.96	1.39	0.57
9	0.98	0.70	0.28
10	0.70	0.41	0.29
>10	0.98	0.57	0.41

Table 7.25: Occurrence of search keyword length in conversion and non conversion data sets.

Search keywords containing three words generated more conversions than non-conversions as did search keywords containing four words. However, four word search keywords generated only 0.05% more conversions than non-conversions. Five and seven word search keywords generated more non-conversions than conversions, while six, eight, nine, ten and longer than ten word keywords generated more conversions than non-conversions. Overall, longer search keywords (containing more than 3 words) were associated with more conversions (68.14%) than non-conversions (64.78%). This appeared to support Hypothesis 1 (H1), which suggested that longer search terms indicated that visitors were more ready to convert.

Summary of results obtained from fourth exploration

Data collected by the OTM was analysed to investigate H1 and H2. The results that were obtained by comparing the occurrence of search keyword length in the conversion and non-conversion data sets provided some evidence to support H1 and H2. This also supported part of the search-conversion model that was proposed. H3 and H4 were not tested as the research ran out of time. This is an area where future work is required. There were some limitations to the conclusions drawn regarding H1 and H2. Not all keywords found in the data were relevant to the content of the dynamic main website. Some PPC advertisements were triggered by keywords that had low relevancy. For example, search keywords such as “james dyson inventions” or “largest manufacturer of plastic parasol bases in the uk” were not relevant to the dynamic main website’s content and were not expected to convert. Irrelevant search keywords could have affected the results obtained during the analysis of search keywords associated with non-conversions. Another limitation was that the search keyword data that was analysed had been collected from one website only. Search keywords from other websites should be analysed and the results compared so as to further validate H1 and H2.

7.5.6. Fifth exploration - Search keyword Type

The relevancy of a visitor’s search keyword to the content of a website could affect the likeliness that the visitor would browse or convert. The relevancy of a search keyword did not depend on its length but rather on the words that made up the search keyword. Increased relevancy was thought to lead to increased browsing which could in turn lead to conversions. This research evaluated the relevancy of a search keyword by determining its type. Search keyword types were identified by reviewing the search keywords in the data and finding common attributes between them. The search keyword types identified were:

1. *[Product]* – search keywords that indicated a need for a product, for example "plastic container designer".
2. *[Service]* – search keywords that indicated a need for a service, for example "make a prototype", "rubber manufacture" and "plastic manufacturers ".
3. *[Location]* – search keywords that contained a location, for example "magnet manufacturers uk toys" and "plastic product manufacturing uk".
4. *[Other]* – search keywords that did not fit under any type or which were not relevant to the main website e.g. "buy plastic coffin", "membership for inventors" and "invention grant".

While assigning types to the search keywords found in the data, it was observed that most search keywords belonged to more than one type for example "plastic tray manufacturer uk" was of type *[Product]*, *[Service]* and *[Location]*. Therefore, rather than using keyword types as categories, they were used as individual predictor attributes. *[Product]*, *[Service]* and *[Location]* were given a value of 1 or 0 depending on whether the search keyword satisfied the requirements for that type. In the case of *[Other]*, the values ranged from -1 for search keywords that were irrelevant to the dynamic main website to 1 for search keywords that were relevant and 0 for search keywords that were not of type *[Other]*. Table 7.26 shows examples of how keywords were scored against keyword types.

Key phrase	Product	Service	Location	Other
plastic feed buckets	1	0	0	0
custom plastic molding	0	1	0	0
sell a toy idea	0	0	0	-1
plastic manufacturers in bristol	0	1	1	0

Table 7.26: Attribute for search keyword relevance.

The attributes were scored manually as it was difficult to automate the process due to the uniqueness of keywords in the data set. As a result, the new attributes were derived for small samples taken from ***Training Data*** and ***Test Data***.

The new data sets ***Keyword Training Data*** and ***Keyword Test Data*** used 26% of the data from ***Training Data*** and ***Test Data*** respectively***. Keyword Training Data*** contained:

- A total of 376 records.
- 188 records of visits that had converted, that is had a value of 1 for *[Conversion]*.
- 188 records of visits that had not converted, that is had a value of 0 for *[Conversion]*.

Keyword Test Data contained:

- A total of 378 records.
- 189 records of visits that had converted, that is had a value of 1 for *[Conversion]*.
- 189 records of visits that had not converted, that is had a value of 0 for *[Conversion]*.

Keyword Training Data and ***Keyword Test Data*** were explored using the "Linear Regression", "Find Laws" and "Neural Network" algorithms. Table 7.27 shows the attributes that were used in the fifth exploration.

Attributes used in fifth exploration	
[Browsed ContactUs]	*[Media Access]*
[Browsed ContactUsTS]	*[Media AccessTS]*
[Browsing Time]	*[Other]*
[Case Studies]	*[PR]*
[Case StudiesTS]	*[PRTS]*
[Company Info]	*[Product]*
[Company InfoTS]	*[Service]*
[Download]	*[Services]*
[DownloadTS]	*[ServicesTS]*
[FTQuote]	*[Unsent Form]*
[FTQuote TS]	*[UnsentFormTS]*
[Keyword Length]	*[Video]*
[Location]	*[VideoTS]*

Table 7.27: Attributes used in fifth exploration.

Linear Regression

“Linear Regression” produced model LR5 (see Equation 7.10). Out of the four search keyword attributes used as input for the “Linear Regression” algorithm, *[Service]* and *[Location]* were identified by LR5 as being useful for predicting *[Conversion]*.

$$\begin{aligned} Conversion = {} & +0.345505 + 0.000808075 * VideoTS - 0.0958917 \\ & * [Company\ Info] - 0.179565 * [Unsent\ Form\ ] - 0.0800860 \\ & * [Media\ Access] + 0.000238593 * [Browsing\ Time] - 0.117527 \\ & * [Browsed\ ContactUs] + 0.162107 * FTQuote + 0.167554 \\ & * Service - 0.192616 * Location \end{aligned}$$

Equation 7.10: Prediction model LR5.

The accuracy measures for LR5 derived using ***Keyword Training Data*** are shown in Table 7.28.

StdErr	RSq	StdDev
0.92	0.16	0.46

Table 7.28: Accuracy measures for LR5 derived using *Keyword Training Data*.

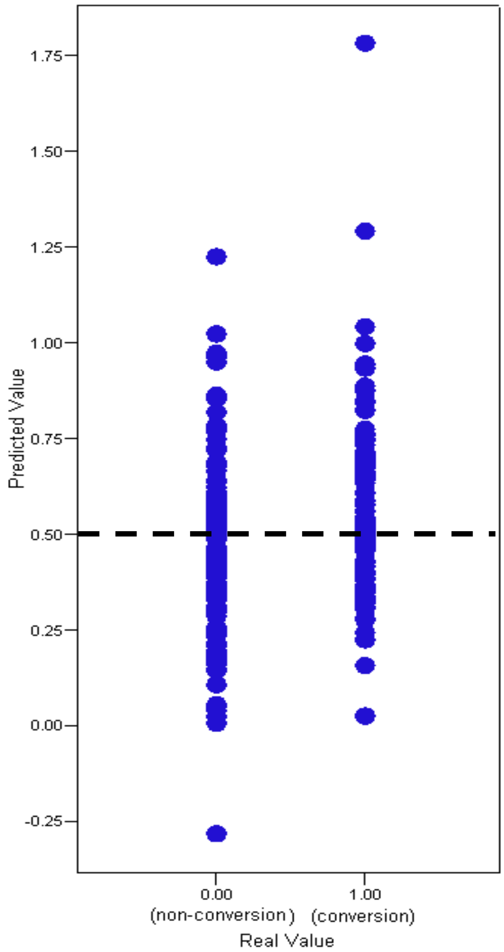
It can be seen from Table 7.28 that LR5 had a high standard error and low RSq value. This suggested that LR5 was not accurate at predicting *[Conversion]*. The accuracy measures for LR5 when tested with ***Keyword Test Data*** are shown in Table 7.29. The high standard error and Rsq confirmed that LR5 was poor at predicting *[Conversion]*.

StdErr	RSq	StdDev
0.98	0.04	0.49

Table 7.29: Accuracy measures for LR5 using (*Keyword Test Data*).

Graph 7.10 shows how the *[Conversion]* values predicted by LR5 varied compared to the real *[Conversion]* value in ***Keyword Test Data.*** The x-axis shows the real value of *[Conversion]* for the records in ***Keyword Test Data***, that is 0 for non-conversion and 1 for conversion.

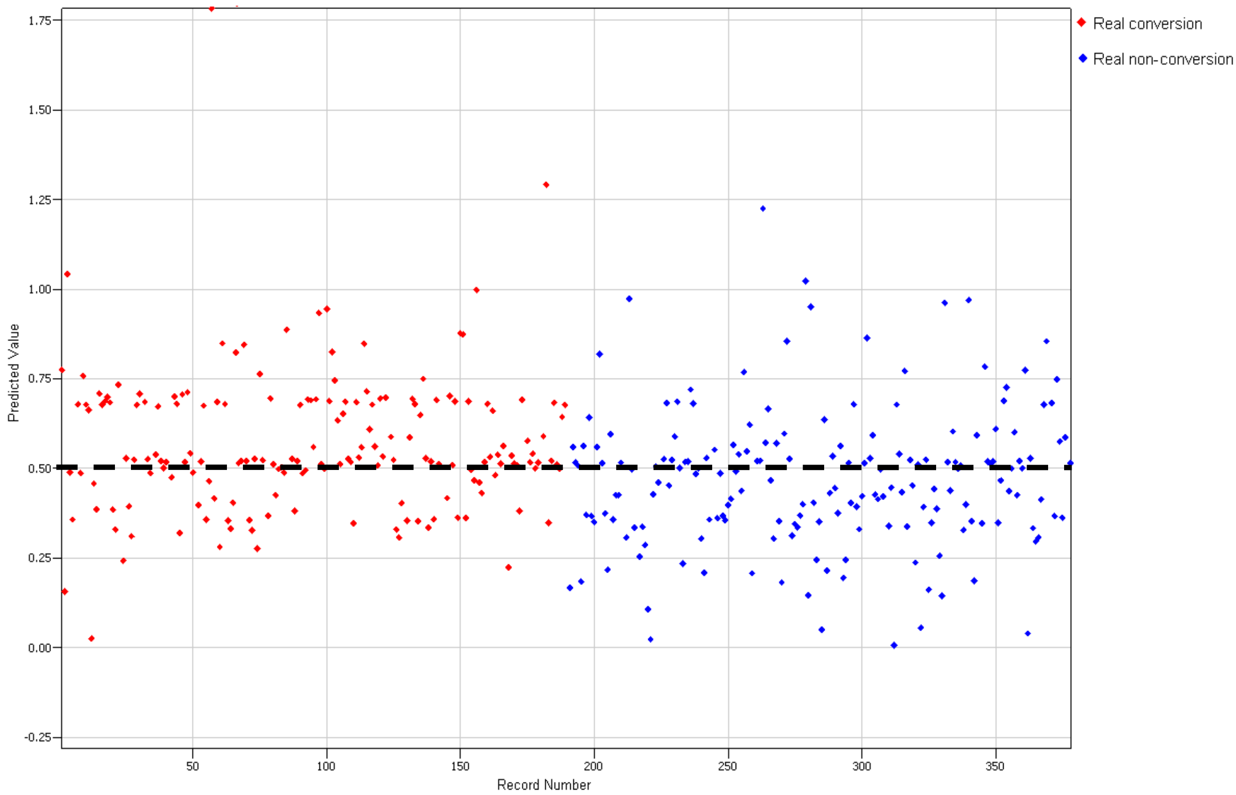
The y-axis represents the value that LR5 calculated for *[Conversion]* using Equation 7.10 and ***Keyword Test Data***. When the real value was 0 (non-conversion), the values predicted by LR5 ranged from -0.28 to 1.22. When the real value was 1 (conversion), the values predicted by LR5 ranged from 0.03 to 1.78. It was observed from Graph 7.10 that the range of predicted values overlapped for most of the data. A boundary value of 0.50 was considered.



Graph 7.10: *[Conversion]* value predicted using LR5 vs real *[Conversion]* value (using *Keyword Test Data*).

Graph 7.11 shows the *[Conversion]* values predicted by LR5 for individual records found in ***Keyword Test Data***. The x-axis of the graph shows the record number for each record found in ***Keyword Test Data***. The y-axis shows the *[Conversion]* values predicted by LR5. The red dots on the graph represent records whose real *[Conversion]* value was 1 (conversion) while the blue dots represent records whose real *[Conversion]* value was 0 (non-conversion).

Graph 7.11 confirmed the observations made from Graph 7.10 whereby most of the predicted values for conversion and non-conversion overlapped, thus making it difficult to find an obvious boundary value. A boundary value of 0.50 was decided upon (shown as a dotted line on Graph 7.10 and Graph 7.11).



Graph 7.11: *[Conversion]* values predicted by LR5 for individual records found in *Keyword Test Data*.

A confusion matrix for LR5 (using ***Keyword Test Data***) with a boundary value set to 0.50 is shown in Table 7.30. The classification probability and efficiency of LR5 were derived from the values shown in Table 7.30. They were 62.43% and 24.87% respectively. The model predicted 69.84% of target 1s correctly, which suggested good accuracy. Also Table 7.30 shows that LR5 (with a boundary value set at 0.50) predicted that 217 records were conversions. Out of these predictions 60.83% (132) were correct.

Predicted / Actual	**> 0.50 (conversion)**	**≤ 0.50 (non-conversion)**	**Total**
1 (conversion)	132	57	189
0 (non-conversion)	85	104	189
Total	217	161	378

Table 7.30: Confusion matrix for LR5 with boundary set to 0.50 (using *Keyword Test Data*)

Model LR2 which was found during the second exploration was tested with ***Keyword Test Data*** and the results were compared with LR5. Table 7.31 shows the accuracy measures for LR2 when tested with ***Keyword Test Data.***

StdErr	RSq	StdDev
0.97	0.07	0.48

Table 7.31: Accuracy measure for LR2 when tested with *Keyword Test Data.*

A confusion matrix for LR2 (using ***Keyword Test Data***) with a boundary value set to 0.50 is shown in Table 7.32.

Predicted / Actual	**> 0.50 (conversion)**	**≤ 0.50 (non-conversion)**	**Total**
1 (conversion)	140	49	189
0 (non-conversion)	88	101	189
Total	228	150	378

Table 7.32: Confusion matrix for LR2 with boundary set to 0.50 (using *Keyword Test Data*)

The classification probability and efficiency of LR2 when tested with ***Keyword Test Data*** were derived from the values shown in Table 7.32. They were 63.76% and

27.51% respectively. The model predicted 74.07% of target 1s correctly. Also Table 7.32 shows that LR2 (with a boundary value set at 0.50) predicted that 228 records were conversions. Out of these predictions 61.40% (140) were correct.

Table 7.33 summarises the classification accuracy of LR5 and LR2 when tested with ***Keyword Test Data.*** It was concluded that LR2 was a better prediction model than LR5.

Model	cp %	ce %	Accuracy of conversion prediction in target set %	Accuracy of conversion prediction in predicted set %
LR5	62.43	24.87	69.84	60.83
LR2	63.76	27.51	74.07	61.40

Table 7.33: Classification accuracy of LR5 and LR2 when tested with *Keyword Test Data.*

Find Laws

The “Find Laws” algorithm was unable to find a model.

Neural Networks

The predictor attributes shown in Table 7.27 were used as input for “Neural Networks”. The accuracy measures for model NN5 which was tested with ***Keyword Test Data*** are shown in Table 7.34.

cerr	cp	cf	ce
33.86%	66.14%	0%	32.28%

Table 7.34: Accuracy measures for model NN5.

The classification probability and efficiency of NN5 were less than that of NN2. However, the classification probability was still higher than that of a naïve model. Table 7.35 shows the breakdown of the predictions made by model NN5.

Target	No of records	Error %	Correct%	Undefined%
Yes (conversion)	189	31.75	68.25	0.00
No (non-conversion)	189	35.98	64.02	0.00
Total	378	33.86	66.14	0.00

Table 7.35: Breakdown of NN5.

Table 7.36 shows the confusion matrix for NN5. From Table 7.36 it can be seen that model NN5 predicted a total of 197 records as conversions. Out of these predictions, 129 were correct (65.48%).

Actual \ Predicted	Yes (conversion)	No (non-conversion)	Total
Yes (conversion)	129	60	189
No (non-conversion)	68	121	189
Total	197	181	378

Table 7.36: Confusion matrix for NN5.

Model NN2, which was found during the second exploration was tested with ***Keyword Test Data***. The accuracy measures for the tested model are shown in Table 7.37.

cerr	cp	cf	ce
28.84%	71.16%	0%	42.33%

Table 7.37: Accuracy measures when NN2 was tested with *Keyword Test Data*.

Table 7.38 shows the breakdown of the predictions made by model NN2 when tested with ***Keyword Test Data***. It can be seen that the percentage error in the target set is less than that of NN5 suggesting that NN2 was a more accurate prediction model.

Target	No of records	Error %	Correct%	Undefined%
Yes (conversion)	189	29.63	70.37	0.00
No (non-conversion)	189	28.04	71.96	0.00
Total	378	28.84	71.16	0.00

Table 7.38: Breakdown of NN2 when tested with *Keyword Test Data*.

Table 7.39 shows the confusion matrix for NN2 when it was tested with ***Keyword Test Data***. From Table 7.39, it can be seen that model NN2 predicted a total of 186 records as conversions. Out of these predictions, 133 were correct (71.51%). The accuracy measures for NN5 and NN2 when tested with ***Keyword Test Data*** are summarised in Table 7.40.

Predicted / Actual	Yes (conversion)	No (non-conversion)	Total
Yes (conversion)	133	56	189
No (non-conversion)	53	136	189
Total	186	192	378

Table 7.39: Confusion matrix for NN2 (*using Keyword Test Data*).

From the data presented in Table 7.40, it was concluded that NN2 was a better and more accurate prediction model that NN5.

Model	cp %	ce %	Accuracy of conversion prediction in target set %	Accuracy of conversion prediction in predicted set %
NN5	66.14	32.28	68.25	65.48
NN2	71.16	42.33	70.37	71.51

Table 7.40: Classification accuracy of NN5 and NN2 when tested with *Keyword Test Data.*

Summary of results obtained from fifth exploration

The fifth exploration produced a "Linear Regression" model LR5 and a "Neural Network" model NN5. The "Find Laws" algorithm was unable to find a model to predict *[Conversion]*. In order to compare the accuracy of LR5 and NN5 with models LR2 and NN2 obtained during the second exploration, LR2 and NN2 were tested with ***Keyword Test Data*** and then the accuracy measures were compared to those of LR5 and NN5. It was found that the models produced during the second exploration were more accurate than those produced during this exploration. The keyword relevancy attributes introduced during this exploration had not produced more accurate prediction models. There were two reasons why this could have been the case:

1. Keyword relevancy was not an important factor in predicting conversions.
2. The simple classification model used to determine keyword relevancy was not accurate.

7.5.7. Sixth exploration – new attributes [Keyword Score] and [Ratio of Score to Keyword Length]

Table 7.41 shows the predictor attributes that were used as input for the algorithms used in this exploration.

In order to find a more accurate model than LR5 and NN5, two new attributes *[Keyword Score]* and *[Ratio of Score to Keyword Length]* were derived. *[Keyword Score]* was calculated by adding the scores of the individual relevancy scores. *[Ratio of Score to Keyword Length]* was the ratio of *[Keyword Score]* to *[Keyword Length]*. These two new attributes replaced the keyword relevance attributes in ***Keyword Training Data*** and ***Keyword Test Data***.

Attributes used in sixth exploration	
[Browsed ContactUs]	*[Keyword Score]*
[Browsed ContactUsTS]	*[Media Access]*
[Browsing Time]	*[Media AccessTS]*
[Case Studies]	*[PR]*
[Case StudiesTS]	*[PRTS]*
[Company Info]	*[Ratio of Score to Keyword Length]*
[Company InfoTS]	*[Services]*
[Download]	*[ServicesTS]*
[DownloadTS]	*[Unsent Form]*
[FTQuote]	*[UnsentFormTS]*
[FTQuote TS]	*[Video]*
[Keyword Length]	*[VideoTS]*

Table 7.41: Attributes used in sixth exploration.

Linear Regression

The "Linear Regression" algorithm found prediction model LR6 (see Equation 7.11) using ***Keyword Training Data***.

$$\begin{aligned} Conversion = {} & +0.463718 + 0.000818116 * VideoTS - 0.0907351 \\ & * [Company\ Info] - 0.175348 * [Unsent\ Form\ ] - 0.0826452 \\ & * [Media\ Access] + 0.000277556 * [Browsing\ Time] - 0.150006 \\ & * [Browsed\ ContactUs] + 0.136613 * FTQuote \end{aligned}$$

Equation 7.11: Prediction model LR6.

The prediction model LR6 did not use either of the new attributes. The prediction accuracy of LR6 derived from ***Keyword Training Data*** is shown in Table 7.42.

StdErr	RSq	StdDev
0.93	0.13	0.47

Table 7.42: Measures for prediction model LR6 derived from *Keyword Training Data*.

It can be seen from Table 7.42 that prediction model LR6 had a high standard error and low RSq value. This suggested that the rule had low prediction accuracy. The prediction accuracy measures for LR6 when tested with ***Keyword Test Data*** are shown in Table 7.43. These confirmed the low accuracy of model LR6.

StdErr	RSq	StdDev
0.97	0.06	0.48

Table 7.43: Accuracy measures for prediction model LR6 when tested with *Keyword Test Data*.

Although "Linear Regression" was able to find a prediction rule, it did not use the new attributes *[Keyword Score]* and *[Ratio of Score to Keyword Length]*. "Linear Regression" did not find a relationship between these attributes and *[Conversion]*.

Find Laws

"Find Laws" was unable to find a prediction model suggesting that the relationship that existed in the data might have been too complex for this algorithm.

Neural Networks

Accuracy measures for model NN6 that was generated by "Neural Networks" and tested with ***Keyword Test Data*** is shown in Table 7.44.

cerr	cp	cf	ce
31.22%	68.78%	0%	37.57%

Table 7.44: Accuracy measures for model NN6.

The classification probability and efficiency of NN6 were higher than those of NN5 but still lower than those of NN2 when tested with ***Keyword Test Data*** (shown in Table 7.37). Table 7.45 shows the breakdown of the predictions made by model NN6 and Table 7.46 shows its confusion matrix.

Target	No of records	Error %	Correct%	Undefined%
Yes	189	35.45	64.55	0.00
No	189	26.98	73.02	0.00
Total	378	31.22	68.78	0.00

Table 7.45: Breakdown of predictions for model NN6.

Actual \ Predicted	Yes (conversion)	No (non-conversion)	Total
Yes (conversion)	122	67	189
No (non-conversion)	51	138	189
Total	173	205	378

Table 7.46: Confusion matrix for NN6.

It can be seen from Table 7.45 that NN6 predicted 64.55% of the target Yes correctly. It was observed from Table 7.46 that NN6 predicted a total of 173 records as conversions. Out of these predictions, 122 were correct (70.52%).

Summary of results obtained from sixth exploration

The sixth exploration produced a "Linear Regression" model LR6 and a "Neural Network" model NN6. The "Find Laws" algorithm was unable to find a model to predict *[Conversion]*. "Linear Regression" produced a model which did not use either of the new attributes introduced in this exploration. "Neural Networks" produced model NN6 whose overall classification accuracy was better than NN5. Table 7.47 shows the classification accuracy of NN6, NN5 and NN2 when tested with ***Keyword Test Data***. The difference in accuracy between the three models was small. NN2 had the highest classification accuracy.

Model	cp %	ce %	Accuracy of conversion prediction in target set %	Accuracy of conversion prediction in predicted set %
NN6	68.78	37.57	64.55	70.52
NN5	66.14	32.28	68.25	65.48
NN2	71.16	42.33	70.37	71.51

Table 7.47: Classification accuracy of NN6, NN5 and NN2 when tested with *Keyword Test Data.*

Even though NN5 and NN6 were less accurate than NN2, they were still better than a naïve prediction model.

The scoring system used to score the keyword relevancy attribute and calculate *[Keyword Score]* and *[Ratio of Score to Keyword Length]* had some limitations which could have accounted for the fact that they did not produce models which had better accuracy than NN2.

It was observed that most keywords that had not converted were still relevant to the websites. It was therefore difficult to evaluate keyword relevancy using an attribute based scoring system, since keywords that did not convert could still satisfy the relevancy criteria and score high marks. The keywords used to search for design services were shorter than those used to search for manufacturing services. As a result the shorter design keywords scored consistently less than the longer manufacturing keywords. This made the scoring system inaccurate. By using the *[Ratio of Score to Keyword Length]*, the inaccuracy was reduced.

A better way of calculating keyword relevancy would have been to weigh each word in the keyword phrase against the frequency with which those words appeared on websites. However, due to the time required to implement such a scoring system this method was not used. This is an area where future work is required.

7.6. Data Interpretation

The way data was collected and transformed might have impacted on the experiments and results described in this Chapter.

7.6.1. Contact Us page Vs FT Quote pages

When calculating *[Browsing Time]*, the time spent on the **Contact Us** page was not included if the user accessed the **Contact Us** form immediately afterwards. If the user visited the **Contact Us** page but did not access the **Contact Us** form next, the time spent on the **Contact Us** form was saved as [*Browsed ContactUs TS]* and included in *[Browsing Time]*.

This method was used as it was assumed that if a user enquired after going to the **Contact Us** page, then they had already made a decision to enquire when they accessed the **Contact Us** page. Therefore, they were not in browsing-mode when they reached the **Contact Us** page

The same reasoning could have been applied to the **Fast Track Quote** (**FT Quote**) pages but it was not. Time spent on the **FT Quote** pages was always included in *[Browsing Time]*. **FT Quote** pages were treated differently because they were used as landing pages by the Manufacturing advertising campaign described in Chapter 5. As such, it was difficult to determine whether a visitor had made a conscious decision to visit an **FT Quote** page or whether they had been sent to one after clicking on an advertisement.

7.6.2. [Browsing Time]

The Online Tracking Module (OTM) described in Chapter 4 was unable to determine *[TimeSpent]* on the last page that a visitor browsed. The system worked out

[*TimeSpent]* on a page by recording and then subtracting the time at which a visitor landed on the page from the time at which the visitor landed on the next page. This did not affect *[Browsing Time]* for visitors who converted as there was always another page following the point from which *[Browsing Time]* stopped being calculated. For visitors who did not enquire, *[Browsing Time]* did not include *[TimeSpent]* on the last page visited as this was the page from which visitors left the website.

It could be argued that this did not affect the results as the issue is consistent across all non-converted visits. If it had randomly affected the records used in the experiments then the data would have been inconsistent and the results unreliable.

The calculation of the number of pages visited for each page type did not suffer from this issue. Pages were counted even if the corresponding value for *[TimeSpent]* was NULL.

7.6.3. Training and test samples relative sizes

The distribution of converted and unconverted visits in the data collected by the OTM was 6% conversion and 94% non-conversion. If this data was split equally into a training sample and a test sample, they would each retain the 6% - 94% split. If a neural network was trained with such a sample, it would learn from less examples of conversion but more examples of non conversion. This could mean that the “Neural Networks” algorithm understood and therefore predicted non-conversions better.

However, if a training sample that had a 50%-50% split was used then the “Neural Networks” algorithm could develop equal knowledge about converted and unconverted visits and could make more accurate predictions. Based on this reasoning a training sample with a 50%-50% split was used to train the “Neural Networks” algorithm used in the experiments described in this Chapter.

The test sample used to test the models that were found in the various explorations also had a 50%-50% split. If the test sample had a 6%-94% then it could have been difficult to assess the accuracy of the models at predicting conversions since the classification probability and efficiency could have been influenced by the prediction accuracy of the high number of non-conversions. Since the research was only interested in predicting conversions, it was looking for a model that was better at predicting conversions than non-conversions. Therefore, a test sample with a 50%-50% split was found to be better for demonstrating the efficiency of a model that was trying to predict conversion.

7.7. Chapter Discussion

Experiments were designed using data collected by the Online Tracking Module (OTM) described in Chapter 4. The aim of the experiments were to analyse data collected by the OTM to find rules or models that enabled the prediction of conversions from recorded user activity on a website.

Data used for the experiments was collected on the collaborating company’s dynamic main website. Initial analysis yielded Neural Network models that had good classification probability and efficiency. However, it was found that errors were introduced in the data due to bugs in the VBA parser macro used to transform the data obtained from the OTM’s database into input data for the “Neural Networks” algorithm. *[Total Time on site]* was calculated with errors in instances where an enquiry was sent from the **Contact Us** page. Also, the VBA parser macro did not filter out visits to the collaborating company’s online shop. Since the research and experiments were focused on service websites, data associated with an online shop introduced errors. The VBA parser macro was modified to filter out visits to the collaborating company's online shopping website and to ignore *[TimeSpent]* on a **Contact Us** page when calculating *[Total Time on site]* in instances where an enquiry was sent from a **Contact Us** page.

The first exploration of the data used all available attributes. "Linear Regression", "Find Laws" and "Neural Networks" were used to find rules. "Linear Regression" and "Find Laws" found rules to calculate a value for *[Conversion]*. Boundary values were identified for each rule and used to determine whether predicted values were considered to be a conversion (indicated by a value of 1). "Neural Networks" also found a prediction model. The "Find Laws" rule was found to be the most accurate with a classification probability that was higher than the "Linear Regression" rule and the "Neural Networks" model.

For visitors who did not convert, *[Total Time on site] was* always the same as *[Browsing Time]*, while for those who converted, *[Total Time on site]* was always greater than *[Browsing Time]*. This relationship made it easy for a model to predict *[Conversion]*. The "Find Laws" rule FL1 identified and exploited this relationship which explained its high classification probability and efficiency. Moreover, when trying to predict *[Conversion]*, only activities that took place before a conversion should have been considered since events that occurred after a conversion would not have affected the outcome of a visit.

In order to find better rules, a second exploration was carried out in which *[Total Time on site]* was not used as input. "Linear Regression" found a rule which had low accuracy while "Find Laws" was unable to find a rule. "Neural Networks" found a model which had a classification probability of 69.32%. The model was better than a naïve model.

There was a disparity between *[Browsing Time]* for visitors who converted and those who did not. The OTM did not detect when a visitor left the website and as a result *[TimeSpent]* on the last page that a visitor browsed was not recorded. In general, this did not affect the calculation of *[Browsing Time]* for visitors who converted. However, for visitors who did not convert, *[Browsing Time]* excluded the time that they spent on the last page that they browsed.

A third exploration was carried out to determine whether the “Neural Networks” model found in the second exploration could be improved by using attributes identified by the “Linear Regression” algorithm. Attributes with high F-Ratios were selected and used as input to the “Neural Networks” algorithm. A more accurate model was not found.

The fourth exploration tried to identify whether a relationship existed between *[Keyword Length]* and *[Conversion]*. The research proposed a search-conversion model based on four hypotheses. Hypothesis 1 (H1) and Hypothesis 2 (H2) were tested by analysing how search keyword length was associated with conversion and non-conversion using data sets derived from data collected by the OTM. It was found that one and two word search keywords were associated with more non-conversions than conversions. It was also found that search keywords that contained more than two words were usually associated with more conversions than non-conversions. Some evidence supporting H1, H2 and thus part of the search-conversion model was found. Hypothesis 3 (H3) and Hypothesis 4 (H4) could not be tested as the research ran out of time. Future work is required to test H3 and H4 and to further investigate the search-conversion model proposed.

It was thought that there could be a relationship between keyword relevancy and *[Conversion]*. In order to investigate this, keywords were scored against a matrix of keyword attributes in the fifth exploration. Only a subset of the data used in previous explorations was utilised for this. The accuracy measures of the rules found in this exploration were still lower than those of the second exploration.

Two new attributes *[Keyword Score]* and *[Ratio of Score to Keyword Length]* were derived from the data and introduced in the sixth exploration. However, these did not produce models that were more accurate than NN2. These results did not necessarily indicate that keyword relevance was not a predictor of *[Conversion]*. The method used to evaluate keyword relevancy had limitations and may not have assessed keyword relevancy consistently. This is an area where future work is required.

PolyAnalyst did not provide information regarding the weights it assigned to the different inputs attributes of “Neural Networks”. As a result, it was not possible to determine how individual attributes contributed in predicting *[Conversion]* or which attributes were more important and which ones had less impact. This is an area where future work is required.

The experiments described in this Chapter showed that the relationships between the various attributes that represented browsing activity were complex and were best modelled by Neural Networks. The fact that models were found that were consistently better than a naive model demonstrated that visitors’ browsing activity on a service website could be used to predict whether visitors were likely to convert. This could help identify attributes and website design elements that affected conversion positively, thus providing a method of improving website design and increasing conversion which could ultimately result in increased sales and profit. The ability to predict conversion could also help create systems that could monitor visitor’s browsing and guide them along an optimised route determined from a conversion model.

CHAPTER 8

DISCUSSION AND CONCLUSION

This Chapter describes the conclusions and recommendations resulting from the research described in this dissertation.

A challenge at the beginning of the research was to improve the data capturing ability of an existing Customer Relationship Management (CRM) system, in order to collect customer data from an initial visit to a website, through to product delivery. To achieve this, the existing CRM system was replaced by Microsoft Dynamic CRM 3.0 (MS CRM). MS CRM was customised and new modules were created to provide additional functionality and data capture.

The research then focused on improving websites. Existing websites at the collaborating company were static and did not record data about visitors' browsing behaviour. A new front-end was implemented to improve the usability and aesthetics of the collaborating company's main website. A new back-end was also created to allow this website to support personalisation features. The back-end incorporated an Online Tracking Module (OTM) which recorded data about visitor's browsing activities.

Data collected by the OTM was integrated with data collected by MS CRM so that a complete history of online behavioural activity and sales activity for website visitors who became customers was captured. PPC advertising campaigns were created and optimised to generate traffic to the collaborating company's main website.

Knowledge was extracted from data collected by MS CRM and the OTM, and was used together with feedback from the collaborating company's sale team to create personas for visitors who could be prospective customers. These personas were used to create and optimise the landing pages of the main online advertising campaigns.

Data collected by the OTM was then analysed to find rules that could predict whether a visitor would enquire based on their browsing behaviour. Rules and models were also sought to identify whether the length of search keywords could indicate readiness to convert based on the stage that a visitor had reached in a search-conversion model proposed in this research.

A Neural Network algorithm was used to build and test models. The models that were found were consistently better than a naive model and demonstrated that visitors' browsing activity on a service website could be used to predict whether visitors were likely to convert.

8.1. Research summary

A challenge at the beginning of the research was to improve data capture. Therefore, Microsoft Dynamics Customer Relationship Management 3.0 (MS CRM) was customised and extended in order to enable it to record customer data throughout a sales and product delivery cycle.

A new Online Tracking Module (OTM) was implemented on existing websites so as to track visitors' activities. Data collected by the OTM was integrated with data collected by MS CRM so that a history of online behavioural activity and sales activity for website visitors who became customers was captured. A new dynamic website was also implemented to replace an existing static website. PPC advertising campaigns were then created to generate traffic to existing websites.

Using the data collected about visitors and customers together with feedback from the sales department, basic visitor personas were created. These personas together with some landing page design techniques were used to improve the conversion rate of landing pages. The effect of the changes to the landing pages was measure using conversion rate and bounce rates.

Data collected by the Online Tracking module about the way visitors interacted with the main website were analysed to find rules/models that could predict whether a visitor would convert based on their browsing behaviour. Rules and models were also sought to identify whether the length of search phrases could indicate readiness to convert based on the stage that a visitor had reached in a search-conversion model proposed by the research.

8.2. Resolution of Research Aims and Objectives

a) *Implement a CRM strategy and software system.*

A new CRM system was implemented, customised and extended by creating new user interfaces and new modules to support business processes. This was described in Chapter 3.

b) *Investigate website design, navigation design and identify ways of measuring website performance.*

Important elements of website and navigation design were identified and discussed in Chapter 2. Some of the elements that were identified were used in the design of a new website described in Chapter 4 as well as landing pages described in Chapter 6. Ways of measuring website performance were identified and discussed in Chapter 2

c) *Create a new dynamic website.*

Chapter 4 described the implementation of new websites that were dynamic and could provide customised and personalised content to website visitors.

d) *Create a new online tracking system that recorded detailed visitor activities and behaviour on a website*.

Chapter 4 described the implementation of a first version of the OTM which recorded visitors' browsing activity. A second version of the OTM was then implemented to overcome some limitations.

e) *Investigate online advertising, Pay Per Click advertising and landing page design and optimisation.*

Online advertising, Pay Per Click advertising and landing page design were investigated in Chapter 2. The findings were taken into account when implementing PPC campaigns described in Chapter 5 and when designing landing pages described in Chapter 6.

f) *Create and optimise new PPC campaigns and landing pages*

PPC campaigns and landing pages described in Chapters 5 and 6 were created and optimised to drive targeted traffic to websites and to increase conversion rate.

g) *Investigate factors that influenced users' behavior online and their relation to conversion.*

Chapter 2 investigated and identified factors regarding website and landing page design as well as online search behavior that influenced online behavior and likeliness to convert.

h) *Find ways to infer whether a website visitor would convert based on their behavior on a website.*

Data mining algorithms were applied to data recorded by the new OTM to find rules to infer conversion. Results were analysed and compared to find the most accurate rules and to determine whether search keywords length could indicate increased likelihood to convert.

8.3. Key research successes and contribution

Research into CRM systems, website design, online advertising, landing page optimisation and online customer behaviour was undertaken. The research work delivered the following achievements:

New systems created:

- Custom modules to extend and improve an existing Customer Relationship Management (CRM) software package:
    - Project Management module.
    - Quality Control module.
    - Customer Satisfaction Survey module.
    - Opportunity Marker module
- New dynamic website that supported content personalisation.
- New Online Tracking Module that collected data pertaining to visitors' browsing behaviour on a website.
- New Lead quality scoring and reporting interface for the Online Tracking Module.

New Methods:

- Capture of customer and sales data from inquiry to order delivery using a customised CRM system.
- Capture of browsing behaviour of website visitors.
- Integrating behavioural data captured on a website with customer and sales data stored in a CRM system
- Search-conversion model for inferring readiness to convert from keyword length.
- Optimisation of PPC campaigns through the creation of user stereotypes from CRM data.
- Optimisation of landing page through segmentation.
- Inferring experiential or goal-oriented behaviour from keyword length
- Creation of metrics to measure the impact of changes to web design, for example:
    - quality of leads.
    - type of customers who enquired, for example, individuals, small companies or corporate.
- Design elements that had significant impact on user behaviour (for example: content, segmentation, navigation, structure, text, pictures etc).
- Method for deriving behavioural attributes from data recorded about the way visitors interacted with a website.
- Using Neural Networks to predict conversion.

8.4. Improvements to this research

There were several areas in this research where limitations restricted the achievement of some objectives and that require improvement.

8.4.1. Online Tracking Module

The Online Tracking Module (OTM) described in Chapter 4 was unable to track *[TimeSpent]* on the last page that a visitor browsed. The module calculated *[TimeSpent]* on a page by recording and then subtracting the time at which a visitor landed on the page from the time at which the visitor landed on the next page. This did not affect *[Browsing Time]* for visitors who converted as there was always another page following the point from which *[Browsing Time]* stopped being calculated. For visitors who did not enquire, *[Browsing Time]* did not include *[TimeSpent]* on the last page visited as this was the page that visitors left the site from.

It could be argued that this did not affect the results as the issue was consistent across all non-converted visits. If it had randomly affected the records used in the experiments described in Chapter 7 then the data would have been inconsistent and the results unreliable.

The calculation of the number of pages visited (for the each page type) did not suffer from this issue. Pages were counted even if the corresponding value for *[TimeSpent]* was NULL.

The time spent on the last page could have been inferred by calculating the average time a visitor spent on other pages in that session

8.4.2. Search-conversion model

This research proposed a search-conversion which was based on hypotheses that were presented in Chapter 7. Hypothesis 1 and 2 (H1 and H2) were tested by analysing the occurrence of search keyword length in the conversion and non-conversion data sets derived from data collected by the OTM.

Some evidence was found to support H1 and H2. However, these had limitations. Not all keywords found in the data were relevant to the content of the dynamic main website. Therefore, irrelevant search keywords could have affected the results obtained during the analysis of search keywords associated with non-conversions. Irrelevant search keywords should have been identified and removed from the data set before analysis.

Another limitation was that the search keyword data that was analysed had been collected one website only. Search keywords from other similar websites should have been analysed and the results compared so as to further validate H1 and H2.

Hypothesis 3 and 4 (H3 and H4) were not tested as this research ran out of time. As a result the search-conversion model was not fully tested and validated.

8.4.3. Keyword relevancy

The scoring system used to score the keyword relevancy attribute and calculate *[Keyword Score]* and *[Ratio of Keyword Score to Keyword Length]* in Chapter 7 had some limitations. It was observed that most keywords that had not converted were still relevant to the websites. It was therefore difficult to evaluate keyword relevancy using an attribute based scoring system, since keywords that did not convert could still satisfy the relevancy criteria and obtain a high score.

Also the average keyword length varied depending on the service that visitor's searched for. For example, keywords used to search for design services tended to be shorter than those used to search for manufacturing services. As a result the shorter keywords scored consistently less than the longer keywords. The scoring system was therefore biased.

A better way of calculating keyword relevancy would have been to weigh each word in the keyword phrase against the frequency with which those words appeared on websites.

The sample data that included keyword score was relatively small due to difficulty in scoring keywords automatically. A bigger sample could have yielded more accurate rules.

8.4.4. Neural Networks

The Neural Network algorithm that was used did not provide detailed breakdowns of the models that it found. It only calculated accuracy measures for the models that it had found. By using a Neural Network algorithm that showed the weights assigned to each attribute in a model, it would have been possible to identify the type of pages that had the greatest influence on visitors' behaviour.

8.5. Suggestions for future work

Following from the discussion presented in Section 7.4 the following are areas where further work could be undertaken:

- Testing and refining the search-conversion model.
- Identifying more accurate methods for calculating the relevancy of search keywords used by website visitors.
- Developing a methodology to infer whether visitors are experiential or goal-oriented based on their browsing behaviour.
- Writing or using new Neural Network algorithms to determine the importance of individual attributes used in the models.

- Identifying methods for calculating the boundary value of Linear Regression and Find Laws automatically or using statistical methods.

8.6. Thesis conclusion

Research was successfully undertaken in the area of Information Technology to create an integrated system that could assist in the collection of data about the browsing behaviour of website visitors as well as sales and marketing data for those website visitors who turned into customers.

The research resulted in the customisation and extension of a CRM software package that was used to capture data from the enquiry stage through to the product stage and beyond in the form of customer satisfaction surveys. The research also resulted in the creation of a dynamic website that had content personalisation features and an Online Tracking Module that recorded the browsing behaviour of website visitors. Data integration was achieved between the OTM and the CRM system, enabling knowledge to be extracted about website visitors and customers. This knowledge was used to improve online marketing and the website's design.

A key contribution to knowledge was the creation of a method to predict the outcome of visits to a website from visitors browsing behaviour. The objectives of the research were broadly achieved. Problems and limitations were encountered. These have been explained and possible solutions have been recommended.

The research demonstrated that visitors' browsing activity on a service website could be used to predict whether visitors were likely to convert. Such conversion prediction models could help identify attributes and web design elements that affect conversion positively, thus providing a method for improving website design and increasing conversion that could ultimately result in increased sales and profit. The ability to predict conversion could be used to create systems that monitor visitor's browsing and guide them along an optimised route determined from a conversion model.

REFERENCES

Adriaans, P. & Zantinge, D. (1996). *Data Mining*, Addison-Wesley Longman.

Agarwal, R. & Venkatesh, V. (2002). Assessing a firm's Web presence: A heuristic evaluation procedure for measurement of usability. *Information Management Research,* 13**,** 168-186.

Ahn, T., Ryu, S. & Han, I. (2007). The impact of Web quality and playfulness on user acceptance of online retailing. *Information & Management,* 44**,** 263-275.

Anderson, C. (2006). *The Long Tail: how endless choice is creating unlimited demand,* London, Random House Business.

Anderson, R. E. & Srinivasan, S. S. (2003). E-satisfaction and e-loyalty: A contingency framework. *Psychology & Marketing,* 20**,** 128-138.

Ariely, D. (2000). Controlling the information flow: effects on consumers' decision making and preferences. *Journal of Consumer Research,* 27**,** 233–248.

Ash, T. (2008). *Landing Page Optimisation: The definitive guide to testing and tuning for conversions*, Wiley Publishing, Inc.

Averill, J. R. (1975). A Semantic Atlas of Emotional Concepts. *JSAS Catalogue of Selected Documents in Psychology,* 52**,** 330.

Avlonitis, G. J. & Panagopoulos, N. G. (2005). Antecedents and consequences of CRM technology acceptance in the sales force. *Industrial Marketing Management,* 34**,** 355-368.

Bailey, M. (n.d). Keyword Strategies - The Long Tail. Retrieved 31 October 2010 from http://www.searchengineguide.com/matt-bailey/keyword-strategies-the-long-tail.php

BBC News (2009). UK in recession as economy slides. Retrieved 01 February 2011 from http://news.bbc.co.uk/1/hi/business/7846266.stm

Bergasa-Suso, J. (2005). Intelligent software systems to assist in using the internet. *Mechanical Design and Engineering.* University of Portsmouth.

Berry, M. J. A. & Linoff, G. S. (2004). *Data Mining Techniques For Marketing, Sales, and Customer Relationship Management*, Wiley.

Bhatt, G. D. & Troutt, M. D. (2005). Examining the relationship between business process improvement initiatives, information systems integration and customer focus: an empirical study. *Business Process Management Journal* 11**,** 532-558.

Braganza, A. (2002). Enterprise integration: Creating competitive capabilities. *Integrated Manufacturing Systems,* 13**,** 562–72.

Brock, T. C. (1965). Communicator-recipient similarity and decision change. *Journal of Personality and Social Psychology***,** 650-654.

Broder, A. (2002). A taxonomy of Web search. *ACM SIGIR Forum,* 36**,** 3–10.

Bucklin, R. E. & Sismeiro, C. (2008). Click Here for Internet Insight: Advances in Clickstream Data Analysis in Marketing. *Journal of Interactive Marketing, Forthcoming*.

Bull, C. (2003). Strategic issues in customer relationship management (CRM) implementation. *Business Process Management Journal* 9**,** 592–602.

Burby, J., Brown, A. & Committee, W. S. (2007). Web Analytics Definitions. WAA Standards Committee.

Burns, K. (2005). New Technology Briefing: Ten golden rules to search advertising. *Interact Market,* 6**,** 248-252.

Chang, W., Park, J. E. & Chaiy, S. (2010). How does CRM technology transform into organizational performance? A mediating role of marketing capability. *Journal of Business Research,* 63**,** 849-855.

Charlton, G. (2010). Search market report reveals increase in PPC and SEO spending. Retrieved 30 October 2010 from http://econsultancy.com/uk/blog/6031-search-market-report-reveals-increase-in-ppc-and-seo-spending

Chen, L. D., Gillenson, M. L. & Sherrell, D. L. (2002a). Enticing online consumers: an extended technology acceptance perspective. *Information & Management,* 39**,** 705-719.

Chen, Q., Clifford, S. & Wells, W. (2002b). Attitude toward the site II: new information. *Journal of Advertising Research,* 42**,** 33-45.

Chiou, W.-C., Lin, C.-C. & Perng, C. (2010). A strategic framework for website evaluation based on a review of the literature from 1995-2006. *Information & Management,* 47**,** 282-290.

Chung, H. & Zhao, X. (2004). Effects of Perceived Interactivity on Web Site Preference and Memory: Role of Personal Motivation. *Journal of Computer-Mediated Communication,* 10.

Chung, J. & Tan, F. B. (2004). Antecedents of perceived playfulness: an exploratory study on user acceptance of general information-searching websites. *Information & Management,* 41**,** 869-881.

Ciborra, C. & Failla, A. (2000). Infrastructure as a process: the case of CRM in IBM. *From Control to Drift: they Dynamics of Corporate Information Infrastructures.* Oxford University Press.

Clickz (2010). Stats – Web Worldwide. Retrieved 15 March 2010 from http://www.clickz.com/stats/web_worldwide

Cooley, R., Mobasher, B. & Srivastava, J. (1997). Web Mining: Information and Pattern Discovery on the World Wide Web. *Ninth IEEE International Conference on Tools with Artificial Intelligence***,** 558 - 567

Crm Connected (n.d.). Features. Retrieved 27 October 2010 from http://www.crmconnected.com/features.html

Cyr, D. (2000). Modeling Web Site Design Across Cultures: Relationships to Trust, Satisfaction, and E-Loyalty. *Journal of Management Information Systems,* 24**,** 47-72.

Das, R. & Turkoglu, I. (2009). Creating meaningful data from web logs for improving the impressiveness of a website by using path analysis method. *Expert Systems with Applications,* 36**,** 6635-6644.

Davies, H. T. & Crombie, I. K. (2009). What are confidence interval and p-values? Retrieved 03 February 2011 from http://www.whatisseries.co.uk/whatis/

Davis, F. D. (1986). A Technology Acceptance Model for Testing New End-User Information Systems: Theory and Results. *School of Management.* Cambridge, MIT Sloan

Demir, G. N., Goksedef, M. & Etaner-Uyar, A. S. (2007). Effects of Session Representation Models on the Performance of Web Recommender Systems. *Data Engineering Workshop, 2007 IEEE 23rd International Conference on***,** 931-936.

Dowling, G. (2002). Customer relationship management: in B2C markets, often less is more. *California Management Review* 44**,** 87–104.
Etzioni, O. (1996). The world wide web Quagmire or gold mine. *Communications of the ACM* 39**,** 65–68.
Evans, D. S. (2008). The Economics of the Online Advertising Industry. *Review of Network Economics,* 7.
Fain, D. C. & Pedersen, J. O. (2006). Sponsored Search: A Brief History. *Bulletin of the American Society for Information Science and Technology,* 32**,** 12-13.
Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P. & Berners-Lee, T. (1999). Hypertext Transfer Protocol -- HTTP/1.1. Retrieved 11 January 2011 from http://www.ietf.org/rfc/rfc2616.txt
Finnegan, D. J. & Currie, W. L. (2010). A multi-layered approach to CRM implementation: An integration perspective. *European Management Journal,* 28**,** 153-167.
Fogg, B., Kameda, T., Boyd, J., Marshall, J., Sethi, R., Sockol, M. & Trowbridge, T. (2002). Stanford–Makovsy Web credibility study 2002: Investigating what makes websites credible today. . Retrieved 2 November 2010 from http://captology.stanford.edu/pdf/Stanford-MakovskyWebCredStudy2002-prelim.pdf
Fogg B, K. T., Boyd J, Marshall J, Sethi R, Sockol M, et al. (2002). Stanford–Makovsy Web credibility study 2002: investigating what makes websites credible today. A research report by the Stanford Persuasive Technology Lab
and Makovsky & Co. . *Stanford University Press*.
Fogg, B. J. (1999). Persuasive Technologies. *Persuasive Technologies,* 42**,** 27-29.
Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Jyoti, P., Rangnekar, A., Shon, J., Swani, P. & Treinen, M. (2001). What Makes A Web Site Credible? A Report on a Large Quantitative Study. *Proceedings of ACM CHI 2004 Conference on Human*
Factors in Computing Systems.
FrontRange Solutions Inc (2002). Using GoldMine 6.0. Retrieved 27 Oct 2010 from http://www.ticomix.com/manuals/Goldmine/Ticomix-GM60userguide.pdf
FrontRange Solutions Inc (n.d.). Latest GoldMine Achieves Perfect Score in PC Magazine Review Retrieved 13 December 2010 from http://www.frontrange.com/company/pressreleases.aspx?id=750
Gefen, D. & Straub, D. (2003). Managing user trust in B2C e-services. *e-Service Journal,* 2**,** 7-24.
Ghahramani, Z. (2004). Zoubin Ghahramani. *Advanced Lectures on Machine Learning*.
Ghose, A. & Yang, S. (2008a). Comparing Performance Metrics in Organic Search with Sponsored Search Advertising. *International Conference on Knowledge Discovery and Data Mining***,** 18-26.
Ghose, A. & Yang, S. (2008b). An empirical analysis of sponsored search performance in search engine advertising. *International conference on Web search and web data mining*.
Gofman, A. & Moskowitz, H. R. (2009). Integrating science into web design:consumer-driven web site optimization. *Journal of Consumer Marketing,* 26**,** 286–298.
Göksedef, M. & Gündüz-Ögüdücü, S. (2010). Combination of Web page recommender systems. *Expert Systems with Applications,* 37**,** 2911-2922.
Goodhue, D. L., Wybo, M. D. & Kirsch, L. J. (1992). The Impact of Data Integration on the Costs and Benefits of Information Systems. *Management Information Systems Quarterly,* 16**,** 293-311.

Google (n.d.-a). Ad and Site Quality. Retrieved 26 January 2011 from http://adwords.google.com/support/aw/bin/static.py?hl=en&page=guide.cs&guide=23686&topic=23689
Google (n.d.-b). AdWords Beginner's Guide Retrieved 15 May 2010 from http://adwords.google.com/support/aw/bin/static.py?hl=en-uk&topic=21903&guide=21899&page=guide.cs
Google (n.d.-c). AdWords Beginner's Guide : Importance of relevancy Retrieved 15 May 2010 from http://adwords.google.com/support/aw/bin/static.py?hl=en-uk&topic=21900&guide=21899&page=guide.cs&answer=146307
Google (n.d.-d). AdWords Help. Retrieved 12 May 2010 from http://adwords.google.com/support/aw/bin/answer.py?hl=en-uk&answer=86879
Google (n.d.-e). Enterprise-class web analytics made smarter, friendlier and free. Retrieved 3 November 2010 from http://www.google.com/intl/en_uk/analytics/#utm_source=en_gb-ha-uk-bk_analytics&utm_medium=ha&utm_campaign=en_gb&utm_term=google%20analytics
Google (n.d.-f). Google Milestones. Retrieved 12 May 2010 from http://www.google.com/corporate/history.html
Google (n.d.-g). What's the difference between clicks, visits, visitors, pageviews, and unique pageviews? Retrieved 16 February 2011 from
Google (n.d.-h). What is the Google Network? Retrieved 15 January 2011 from http://adwords.google.com/support/aw/bin/answer.py?hl=en&answer=6104
GraphPad (n.d.-a). Analyze a 2x2 contingency table. Retrieved from http://www.graphpad.com/quickcalcs/contingency2.cfm
GraphPad (n.d.-b). t-test calculator. Retrieved from http://www.graphpad.com/quickcalcs/ttest1.cfm
Gupta, A., Smith, K. & Shalley, C. (2006). The interplay between exploration and exploitation. *Academy of Management Journal,* 49**,** 693-706.
Hammer, M. (1995). *Reengineering the Corporation*, Nicholas Brealey Corporation, London.
Hausman, A. V. & Siekpe, J. S. (2009). The effect of web interface features on consumer online purchase intentions. *Journal of Business Research,* 62**,** 5-13.
Heshan, S. & Zhang, P. (2006). Causal Relationships between Perceived Enjoyment and Perceived Ease of Use: An Alternative Approach. *Journal of the Association for Information Systems,* 7**,** 618-645.
Hoffman, D. L. & Novak, T. P. (1996). Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of Marketing,* 60**,** 50-68.
Hoggart, G. (n.d.). Short vs. Long-Tail Keywords. Retrieved 01 November 2010 from http://www.gabehoggarth.com/2009/10/short-vs-long-tail-keywords/
Hollink, V., Van Someren, M. & Wielinga, B. J. (2007). Navigation behavior models for link structure Optimization. *User Modeling and User-Adapted Interaction,* 17**,** 339-377.
Hölscher, C. & Strube, G. (2000). Web search behavior of Internet experts and newbies. *Computer Networks,* 33**,** 337-346.
Homburg, C., Workman, J. P. & Jensen, O. (2000). Fundamental changes in marketing organization: the movement toward a customer-focused organizational structure. *Journal of the Academy of Marketing Science,* 28**,** 459–78.
Huynh, T. & Miller, J. (2009). Empirical observations on the session timeout threshold. *Information Processing & Management,* 45**,** 513-528.

IAB UK (2008). Internet advertising spend up 21% despite economic downturn. Retrieved 29 March 2010 from http://www.iabuk.net/en/1/internetadvertisingspendup21071008.mxs
IAB UK (2010). IAB Online Adspend Factsheet - H1 2010. Retrieved 28 October 2010 from http://www.iabuk.net/media/images/iabresearch_adspend_adspendfctshthh12010_7139.pdf
IAB USA (2010). IAB Internet Advertising Revenue Report. Retrieved 28 October 2010 from http://www.iab.net/media/file/IAB_report_1H_2010_Final.pdf
Ivory, M. & Hearst, M. (2002). Improving Web site design. *Ieee Internet Computing,* 6**,** 56-63.
Ivory, M. Y. & Megraw, R. (2005). Evolution of Web Site Design Patterns. *ACM Transactions on Information Systems,* 23**,** 463-497.
Jafar, R., Shahrour, I. & Juran, I. (2010). Application of Artificial Neural Networks (ANN) to model the failure of urban water mains. *Mathematical and Computer Modelling,* 51**,** 1170-1180.
Jansen, B. J., Booth, D. L. & Spink, A. (2008). Determining the informational, navigational, and transactional intent of Web queries. *Information Processing & Management,* 44**,** 1251-1266.
Jansen, B. J., Spink, A. & Saracevic, T. (2000). Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing & Management,* 36**,** 207-227.
Jee, J. & Lee, W. (2002). Antecedents and Consequences of Perceived Interactivity: An Exploratory Study. *Journal of Interactive Advertising,* 3.
Jones, D. (2002). *Pharmaceutical Statistics*, Pharmaceutical.
Jordan, D. W. & Smith, P. (2002). *Mathematical Techniques - An introduction for the engineering, physical and mathematical sciences.*, Oxford University Press.
Kahn, K. B. & Mentzer, J. T. (1998). Marketing's integration with other departments. *Journal of Business Research,* 42**,** 53–62.
Kaushik, A. (2006). Occam's Razor. Experimentation and Testing: A Primer. Retrieved 25 2010 from http://www.kaushik.net/avinash/2006/05/experimentation-and-testing-a-primer.html
Kim, H. & Fesenmaier, D. R. (2006). First Impression and Persuasive Design in Destination Web sites. *Proceedings of the Annual Conference of the Travel and Tourism Research Association, Travel and Tourism Research Association*.
Kim, H. & Fesenmaier, D. R. (2008). Persuasive Design of Destination Web Sites: An Analysis of First Impression. *Journal of Travel Research,* 47**,** 3-13.
Kim, H. & Niehm, L. S. (2009). The Impact of Website Quality on Information Quality, Value, and Loyalty Intentions in Apparel Retailing. *Journal of Interactive Marketing,* 23**,** 221-233.
Kosala, R. & Blockeel, H. (2000). Web mining research: A survey. *SIGKDD: SIGKDD explorations: Newsletter of the special interest group (SIG) on knowledge discovery and data mining ACM,* 2**,** 1–15.
Kotorov, R. P. (2002). Ubiquitous organization: organizational design for e-CRM. *Business Process Management Journal,* 8.
Koufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behavior. *Information Systems Research,* 13**,** 205-223.
Krug, S. (2006). *Don't make me think, A common sense approach to web usability*, Berkeley California New Riders.

Larsen, K. (2010). Latest Search Market Share: Google, Bing and Yahoo Comparisons. Retrieved 30 October 2010 from http://www.ppcsummit.com/newsletter/google-adwords/latest-search-market-share-google-bing-and-yahoo-comparisons/

Leap, M. (2010). 7 Ways to Segment Your Landing Page Visitors. Retrieved 09 February 2011 from http://www.wordstream.com/blog/ws/2010/03/08/segment-landing-pages

Lee, S. & Koubek, R. J. (2010). The effects of usability and web design attributes on user preference for e-commerce web sites. *Computers in Industry,* 61**,** 329-341.

Levene, M. (2006). *An Introduction to Search Engines and Web Navigation*, Reading, MA: Addison-Wesley.

Liang, T. P. & Lai, H. J. (2002). Effect of store design on consumer purchases: van empirical study of on-line bookstores. *Information & Management,* 39**,** 431-444.

Lin, H.-F. (2010). An application of fuzzy AHP for evaluating course website quality. *Computers & Education,* 54**,** 877-888.

Lin, R. J., Chen, R. H. & Chiu, K. K. S. (2010). Customer relationship management and innovation capability: an empirical study. *Industrial Management & Data Systems,* 110**,** 111-133.

Lindgaard, G., Fernandes, G., Dudek, C. & Brown, J. (2006). Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology* 25**,** 115-126.

Liu, B. (2007). *Web data mining:Exploring hyperlinks, contents and usage data*, Springer.

Liu, C. N. & Zhu, X. W. (2009). A Study on CRM Technology Implementation and Application Practices. *Computational Intelligence and Natural Computing, 2009. CINC '09. International Conference on* 2**,** 367 - 370

Loveday, L. & Neihaus, S. (2008). *Web design for ROI, Turning Browsers into buyers & prospects into leads*, New Riders.

Mahaney, S. (2006). Three good reasons to target long tail keywords! Retrieved 15 June 2010 from http://www.wordtracker.com/academy/three-good-reasons-to-target-long-tail-keywords

Malaga, R. A. (2010). Search Engine Optimization - Black and White Hat Approaches. *Advances in Computers, Vol 78.* San Diego, Elsevier Academic Press Inc.

Marchionini, G. (1995). *Information seeking in electronic environments*, Cambridge: Cambridge University Press.

Marjanovic, O. (2005). Process-Oriented CRM Enabled by Component-Based Workflow Technology. *18th Bled eConference eIntegration in Action*.

Mark, S. (n.d.). Top 10 Benefits to Microsoft CRM 3.0. Retrieved 27 October 2010 from http://www.solutionsmark.com/solutions/products/crm/Pages/Benefits.aspx

Mason, J. (2007). Think Beyond The Click: How To Build Landing Pages That Convert. Retrieved 02 February 2011 from http://searchengineland.com/think-beyond-the-click-how-to-build-landing-pages-that-convert-12939

Meyer, M. & Kolbe, L. M. (2005). Integration of customer relationship management: status quo and implications for research and practice. *Journal of Strategic Marketing,* 13**,** 175-198.

Mitchell, A. (2009). The 5 Benefits of Long-Tail Keywords. Retrieved 15 June 2010 from http://www.alanmitchell.com.au/techniques/benefits-of-long-tail-keywords/

Montgomery, A. L., Li, S., Srinivasan, K. & Liechty, J. C. (2004). Modeling Online Browsing and Path Analysis Using Clickstream Data. *Marketing Science,* 23**,** 579-595.

Moss, G. A., Gunn, R. W. & Kubacki, K. (2008). Gender and Web Design: The Implications of the Mirroring Principle for the Services Branding Model. *Journal of Marketing Communications,* 14**,** 37-57.

Motulsky, H. (1995). Intuitive Biostatistics: Interpreting Nonsignificant P values. *Intuitive Biostatistics.* Oxford University Press.

MSDN Library (n.d.). Request.ServerVariables Collection. Retrieved 13 March 2011 from http://msdn.microsoft.com/en-us/library/ms525396%28v=vs.90%29.aspx

Mutlu, E. (2009). Google Blog: Updates to AdWords conversion metrics. Retrieved 17 May 2010 from http://adwords.blogspot.com/2009/04/updates-to-adwords-conversion-metrics.html

Najjar, Y., Basheer, I. A. & Hajmeer, M. N. (1997). Computational neural networks for predictive microbiology : Methodology. *International Journal of Food Microbiology,* 34**,** 27–49.

Nelson, R. R., Todd, P. A. & Wixom, B. H. (2005). Antecedents of information and system quality:An empirical examination within the context of data warehousing. *Journal of Management Information Systems,* 21**,** 199–235.

Netmarketshare (2010). Search Engine Market Share. Retrieved 12 May 2010 from http://marketshare.hitslink.com/report.aspx?qprid=4&qptimeframe=M&qpsp=120&qpnp=12

Nielsen, J. (1997). How Users Read on the Web. Retrieved 30 March 2010 from http://www.useit.com/alertbox/9710a.html

Nielsen, J. (2008). How Little Do Users Read? Retrieved 30 March 2010 from http://www.useit.com/alertbox/percent-text-read.html

Novak, T. P., Hoffman, D. L. & Duhachek, A. (2003). The influence of goal-directed and experiential activities on online flow experiences. *Journal of Consumer Psychology,* 13**,** 3-16.

Novak, T. P., Hoffman, D. L. & Yung, Y. F. (2000). Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science,* 19**,** 22-42.

Online-CRM.com (n.d.). Sugar CRM Retrieved 27 October 2010 from http://www.online-crm.com/sugarcrm.htm

Paashuis, V. & Boer, H. (1997). Organizing for concurrent engineering: an integration mechanism. *Integrated Manufacturing Systems* 8**,** 79–89.

Park, J., Kim, J. & Koh, J. (2010). Determinants of continuous usage intention in web analytics services. *Electronic Commerce Research and Applications,* 9**,** 61-72.

Pavlou, P. A. & Fygenson, M. (2006). Understanding and predicting electronic commerce adoption: An extension of the theory of planned behavior. *MIS Quarterly,* 30**,** 115-143.

Payne, A. & Ryals, L. (2001). Customer relationship management in financial services: towards information-enabled relationship marketing. *Journal of Strategic Marketing,* 9**,** 3-27.

Phan, D. D. & Vogel, D. R. (2010). A model of customer relationship management and business intelligence systems for catalogue and online retailers. *Information & Management,* 47**,** 69-77.

Picton, P. (2000). *Neural Networks*, Palgrave.

Piercy, N. F. (1998). Barriers to implementing relationship marketing: analysing the internal market-place. *Journal of Strategic Marketing,* 6**,** 209–22.

Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information*, Oxford University Press.

PolyAnalyst (n.d.). PolyAnalyst. Retrieved 11 October 2010 from http://www.megaputer.com/polyanalyst.php
Reichheld, F. F. & Sasser, W. E. (1990). Zero defections: quality comes to services. *Harvard Business Review,* 68**,** 105–111.
Reichheld, F. F. & Schefter, P. (2000). E-loyalty: Your secret weapon on the Web. *Harvard Business Review,* 78**,** 105-114.
Reichheld, F. F. & Teal, T. (1996). *The Loyalty Effect: The Hidden Force Behind Growth, Profits and Lasting Value*, Harvard Business School Press.
Reinartz, W. J., Thomas, J. S. & Kumar, V. (2005). Balancing acquisition and retention resources to maximize customer profitability. *Journal of Marketing,* 69**,** 63–79.
Robbins, S. S. & Stylianou, A. C. (2003). Global corporate web sites: an empirical investigation of content and design. *Information & Management,* 40**,** 205-212.
Robinson, H., Wysocka, A. & Hand, C. (2007). Internet advertising effectiveness: The effect of design on click-through rates for banner ads. *International journal of advertising,* 26**,** 527-541.
Rodgers, S., Wang, Y., Rettie, R. & Alpert, F. (2007). The web motivation inventory. *International Journal of Advertising,* 26**,** 447–476.
Rutz, O. & Bucklin, R. (2007). A model of individual keyword performance in paid search advertising. *Working Paper*.
Sanchez-Franco, M. J. & Roldan, J. L. (2005). Web acceptance and usage model: A comparison between goal-directed and experiential web users. *Internet Research,* 15**,** 21-48.
Search Engine Partner (n.d.). Long Tail Keywords. Retrieved 01 November 2010 from http://www.searchenginepartner.com/Latest-SEO-News/seo-trends-utilysing-lsi-and-the-long-tail.html
SEO Moves (2010). Landing Page Optimization. Retrieved 07 October 2010 from http://www.seomoves.org/blog/search-engine-optimization/landing-page-optimization-1849/
Shaw, R. & Reed, D. (1999). *Measuring and Valuing Customer Relationships: How to Develop the Measures that Drive Profitable CRM Strategies*, London: Business Intelligence Ltd.
Sicilia, M., Ruiz, S. & Munuera, J. L. (2005). Effects of interactivity in a web site. *Journal of Advertising Research,* 34**,** 31–46.
Skiera, B., Eckert, J. & Hinz, O. (2010). An analysis of the importance of the long tail in search engine marketing. *Electronic Commerce Research and Applications,* 9**,** 488-494.
Smith, D. N. & Sivakumar, K. (2004). Flow and Internet shopping behavior - A conceptual model and research propositions. *Journal of Business Research,* 57**,** 1199-1208.
Smyth, B. & Cotter, P. (2003). Intelligent Navigation for Mobile Internet Portals. *IJCAI'03 Workshop on AI Moves to IA: Workshop on Artificial Intelligence, Information Access, and Mobile Computing.* .
Snyder, M. & Steger, J. (2006). *Working with Microsoft Dynamics CRM 3.0*, Microsoft Press.
Stanaland, A. J. S. & Tan, J. (2010). The impact of surfer/seeker mode on the effectiveness of website characteristics. *International Journal of Advertising,* 29**,** 569-595.
StatSoft (Ed.) (2010) *Electronic Statistics Textbook,* Tulsa, OK: StatSoft. web: http://www.statsoft.com/textbook/

Storbacka, K., Strandvik, T. & Gronroos, C. (1994). Managing customer relationships for profit: the dynamics of relationship quality. *International Journal of Service Industry Management,* 5**,** 21–38.
Stromer-Galley, J. (2004). Interactivity as Process and Interactivity as Product. *The Information Society,,* 20**,** 391-94.
Sundar, S. S. & Kim, J. (2005). Interactivity and persuasion: influencing attitudes with information and involvement. *Interactivity and persuasion: influencing attitudes with information and involvement,* 5**,** 5-18.
Talerico, A. (2010). Landing Page Segmentation For Higher Conversions. Retrieved 09 February 2011 from http://online-behavior.com/targeting/landing-page-segmentation-1044
Tam, K. Y. & Ho, S. Y. (2006). Understanding the impact of web personalization on user information processing and decision outcomes. *MIS Quarterly,* 30**,** 865-890.
The Mathbeans Project (n.d.). The Chi Square Statistic. Retrieved 16 February 2011 from http://math.hws.edu/javamath/ryan/ChiSquare.html
Thompson, M. (2009). Website & Landing Page Design Elements for Usability and SEO. Retrieved 30 January 2011 from http://www.stayonsearch.com/webpage-landing-page-design-elements-for-usability-and-seo
Thompson, S. H., Devadoss Teoa Paul & L., P. S. (2006). Towards a holistic perspective of customer relationship management (CRM) implementation: A case study of the Housing and Development Board, Singapore *Decision Support Systems,* 42**,** 1613-1627
Thrash, T. M. & Elliot, A. J. (2003). Inspiration as a Psychological Construct. *Journal of Personality & Social Psychology,* 84**,** 871-89.
Van den Poel, D. & Buckinx, W. (2005). Predicting online-purchasing behaviour. *European Journal of Operational Research,* 166**,** 557-575.
Venkatesh, V. & Morris, M. G. (2000). Why Don't Men Every Stop to Ask for Directions? Gender, Social Influence, and Their Role in Technology Acceptance and Usage. *MIS Quarterly,* 24**,** 115-39.
Wang, B. & Liu, Z. (2003). Web mining research. *Fifth international conference on computational intelligence and multimedia applications (ICCIMA'03)***,** 84.
Wang, S., Beatty, S. E. & Foxx, W. (2004). Signalling the Trustworthiness of Small Online Retailers. *Journal of Interactive Marketing Management,* 18**,** 53-69.
Webster, J. & Abuja, J. S. (2006). Enhancing the Design of Web Navigation Systems: The Influence of User Disorientation on Engagement And Performance. *MIS Quarterly,* 30**,** 661-678.
White, K. (2006). How high is your bounce rate? Retrieved 18 January 2011 from http://newsletter.blizzardinternet.com/how-high-is-your-bounce-rate/2006/02/09/
Widyantoro, D. H. & Yen, J. (2001). A Fuzzy Ontology-Based Abstract Search Engine and Its User Studies. *Proceedings of the International Conference on Fuzzy Systems, Melbourne, Australia, December.*
Wilson, H., Daniel, E. & McDonald, M. (2002). Factors for success in customer relationship management (CRM) systems. *Journal of Marketing Management,* 18**,** 193-219.
Winner, L. (2011). Categorical Data Analysis
Wu, G. (2006). Conceptualizing and measuring the perceived interactivity of websites. *Journal of Current Issues and Research in Advertising,* 28**,** 87–104.
Yang, X., Zafar, A., Ghingold, M., Boon, G., Mei, T. & Hwa, L. (2003). Consumer Preferences for Commercial Web Site Design: An Asia-Pacific Perspective. *Journal of Consumer Marketing,* 20**,** 10-27.

Yen, B. (2007). The design and evaluation of accessibility on web navigation. *Decision Support Systems,* 42**,** 2219-2235.
Yeung, M. C. H. & Ennew, C. T. (2000). From customer satisfaction to profitability. *Journal of Strategic Marketing,* 8**,** 313–26.
Yue, C., Xie, M. & Wang, H. An automatic HTTP cookie management system. *Computer Networks,* 54**,** 2182-2198.
Zahra, S. A. & Nielsen, A. P. (2002). Sources of capabilities, integration and technology commercialization. *Strategic Management Journal* 23**,** 377–98.
Zhang, P. & von Dran, G. M. (2000). Satisfiers and dissatisfiers: A two-factor model for Website design and evaluation. *Journal of the American Society for Information Science,* 51**,** 1253-1268.

APPENDIX A

A.1. *First Online Tracking Module*

A.1.1. Table relationship in OTM database

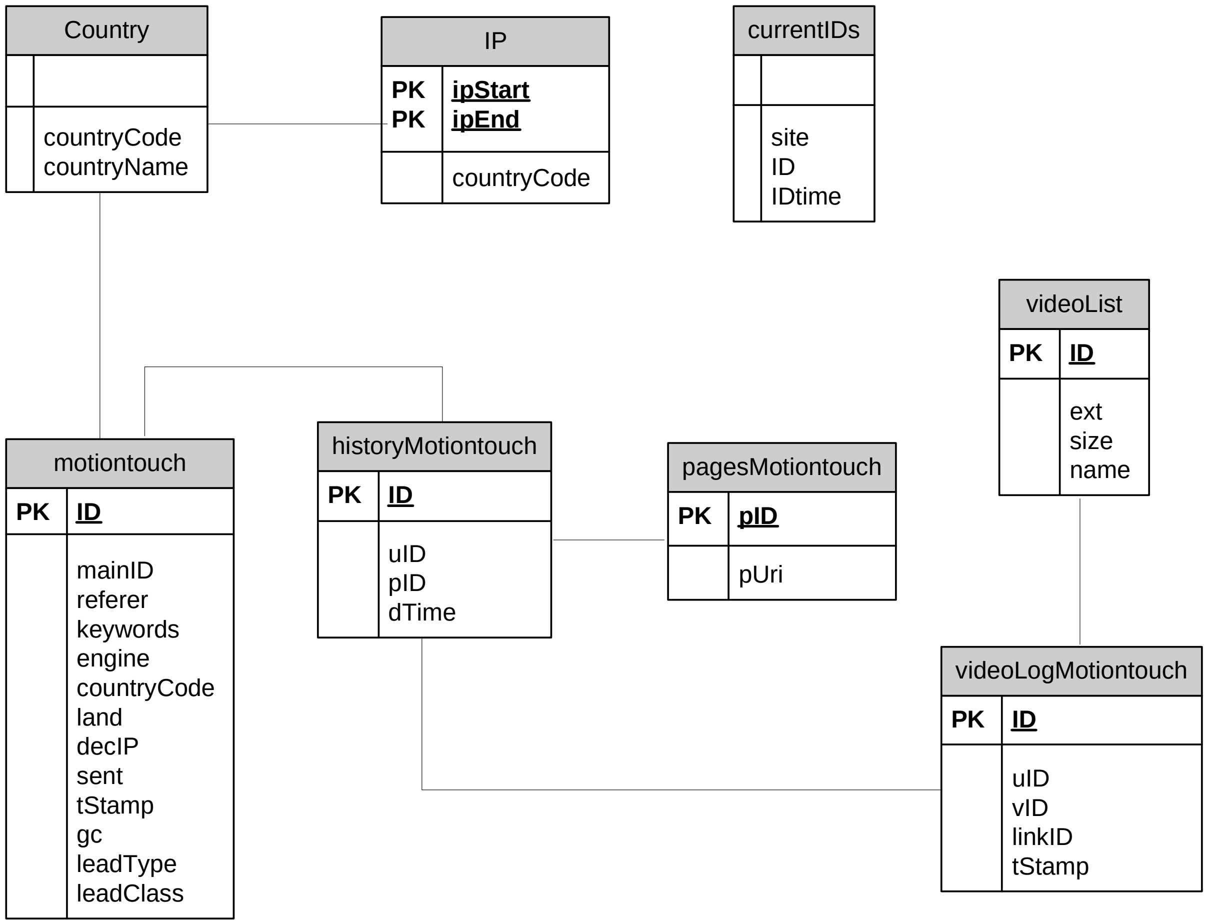


Figure A.1: Table relationship in OTM database.

A.1.2. Definition of tables found in OTM database

	Table	Description
1	CurrentID	Stores next free userID
2	IP	GeoIP DB
3	Country	GeoIP DB
4	motiontouch	Stored details of visitor origin
5	historyMotiontouch	Stores pages visited by users and time spent on each page
6	pagesMotiontouch	Stores the URI for all page on the website
7	videoList	Stores details of all videos on the website
8	videoLogMotiontouch	Stores view history of videos

Table A.1: Definition of tables found in OTM database.

A.1.3. Definition of columns for tables in OTM database

Column Name	*▪ ample Data*	*Description*
Table: country		
countryCode	GB	Primary Key: Unique ISO country code, + Ax for anonymous proxies
countryName	United Kingdom	Iso Country names + Anonymous Proxy
Table: IP		
countryCode	GB	Primary Key: Unique ISO country code, + Ax for anonymous proxies
ipStart	-2113487304	First IP in netblock
ipEnd	-2113487297	Last IP in netblock
Table: currentIDs		
site	MT	Identifies the website that the ID is for
ID	468701	The ID that will be assigned to the next visitor
IDtime	23/06/2009 08:45:15	Time and data at which ID was updated
Table: historyMotiontouch		
ID	1	Primary Key: unique ID
uID	21827	unique ID assigned to each new visitor
pID	23	Foreign Key : unique ID of the page that a visitor browsed
dTime	124	Time in seconds that a visitor spent on a page
Table: pagesMotiontouch		
pID	324	Primary Key: unique page ID
pUri	/manufacturing/plastics_production.asp	URL of page
Table: videoList		
ID	1	Primary Key: Unique video ID
name	Electronic-Design	Name of the video

ext	wmv, avi	Extension indicating format of video
size	2.2	Sizeof video in MB
Table:videoLogMotiontouch		
ID	12	Primary Key: unique ID
uID	21827	Foreign Key: unique ID assigned to each new visitor
vID	1	Foreign Key: unique video ID stored as ID in table videoList
linkID	/manufacturing/video/Default.asp	URL of page where video is embedded
tStamp	20/06/2007 13:39:34	Time and data at which a visitor accessed a video
Table: motiontouch		
ID		Primary Key: unique ID
referer	http://www.google.com/search?q=job+agencys+that+employ+injection+moulders&hl=en&rlz=1B2GGFB_enGB204&start=20&sa=N	URL of the website that from which a visitor originated
keywords	plastic cans	Keyword that a visitor used to carry out a search in a search engine
engine	google	Search engine name
countryCode	GB	Visitor's country of origin
land	www.motiontouch.com/manufacturing/plastics_production.asp	First page on which a visitor landed
decIP	-721618680	Visitor's IP address
sent	1 or 0	Flag indicating whether an email enquiry was sent from the website
tStamp	10/09/2008 04:35:53	Data and time at which a visitor arrived on the website
gc	MT01-01	Unique code given to online marketing campaigns
mainID	21266	FK: Unique visitor ID stored as uID in other tables
leadType	2	Data provided by person creating leads. Type represents lead type i.e. design (1) or manufacturing(2)
leadClass	1	Class represents the quality of leads, good(1) medium(0) or bad (-1)

Table A.2: Definition of columns for tables in OTM database.

A.2. *Table relationship and definition for CAT database*

The relationships between tables found in the CAT database are shown in Figure A.2. The descriptions of the tables found in the CAT database are summarised in Table A.3.

	Table	Description	
1	URL	Contains PageID and corresponding user frendly URI	O
2	Page	Contains content of the page	C
3	Header	Contains right header for the page	C
4	IP	GeoIP DB	G
5	Country	GeoIP DB	G
6	Panel	Stores information that binds pages and panel sections together	C
7	PanelSection	Contains elements for the left panel	C
8	Misc	Contains predefined code to insert into pages	C
9	History	Tracking – main table with user IDs	T
10	MicroHistory	Tracking – detailed log of pages visited	T
11	Sent	Tracking – Contains log of sent events	T
12	VideoLog	Tracking - Log of videos viewed by visitors	T
13	Video	List of available videos	V
14	VideoFile	List of video files for each video, needed for streaming/download	V
15	VideoCategory	Which video belongs to which category, for auto populate video pages	V
16	VideoCategoryList	List of categories that videos belong to	V
17	SentName	Stores names of pages from where enquiries can be sent	T
18	EngineType	Defines if it's a search engine, or directory listing	T
19	CS	Case studies list, with links and pics and short desc, for list of case studies	C
20	CSCategory	List of Case Study categories mappings to Case Studies	C
21	CSCategoryList	List of Case Study Categories	C
22	Keywords	List of search phrases and corresponding interest scores	S
23	CurrentID	Stores next free userID	T
24	SiteSearch	Contains phrases used in page search	T
25	EngineType	Defines name of the Engine	T
26	User	Stores registered visitor details – for authentication	T
27	File	Lists files available for download	C
28	FileCategory	Defines file – category	C
29	FileCategoryList	Defines list of categories for files	C
30	FileHistory	Saves download events	T
31	Questionnaire	Stores questionnaire details	C
32	QLink	Stores link recommendation for questionnaire answer	C
33	QHistory	Stores details of how visitors answered questionnaire	T

Table A.3: Descriptions of tables found in the CAT database.

Database tables are grouped as follows:

Tracking – this group is used to track user behaviour

Scoring – This group is used to calculate visitor interest score

Content – Contains different parts of page content

GeoIP – GeoIP tables

Videos – Video content

Other – things that don’t belong to any other group

#MBD – must be defined while creating record

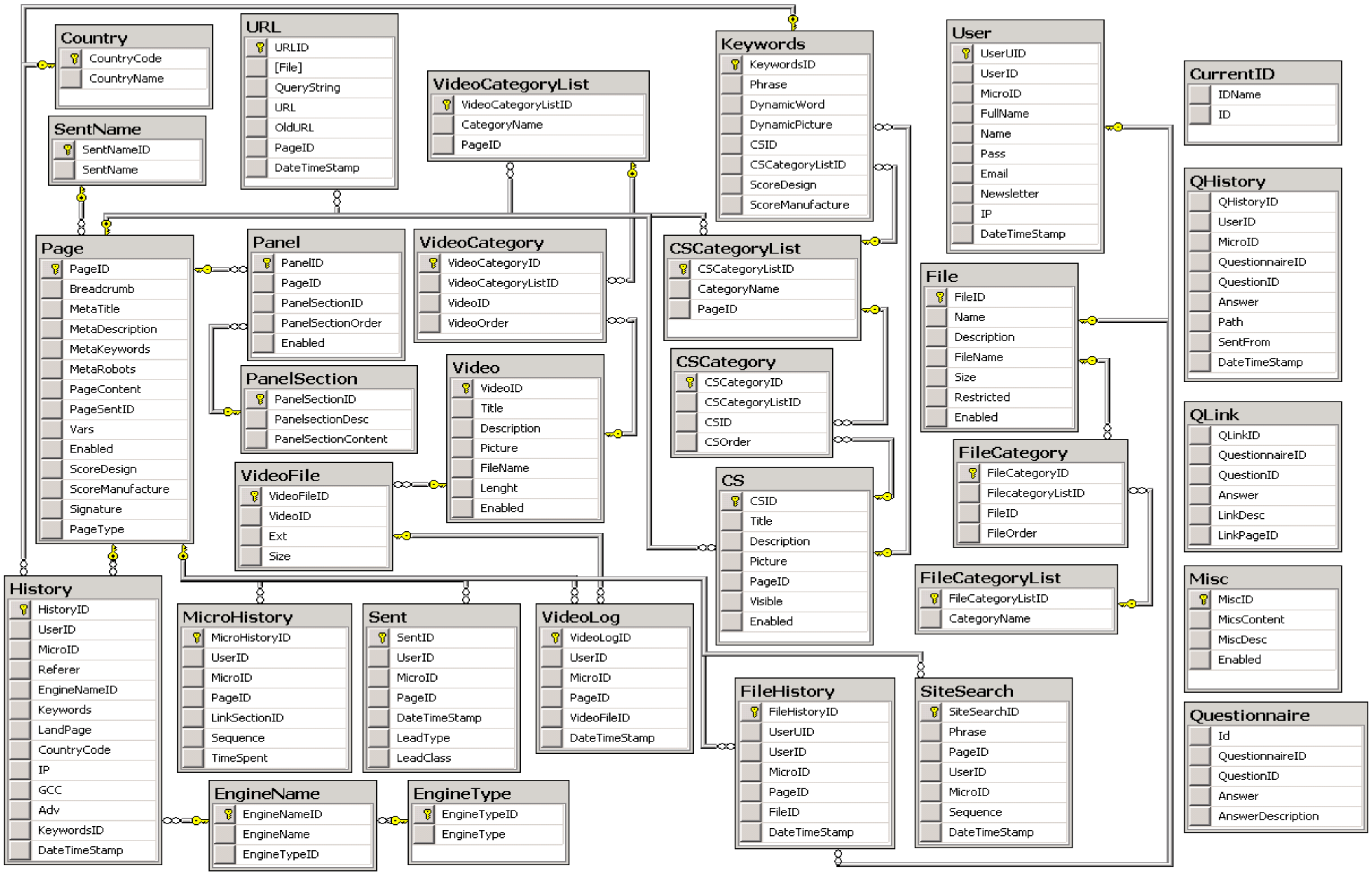


Figure A.2: Relationships between tables found in the CAT database.

A.2.1. Definition of columns for tables in CAT database

Column Name	***Sample Data***	***Default Value***	***Groups***	***Description***
URL			**Other**	
URLID	324	UID		PK - URL unique ID
URL	/design/plastic.asp	#MBD		URL of the page to display to the user, instead of the Page.PageID
PageID	324	UID		FK: Page.PageID – to which page given url refers to
File	View.asp	<empty>		Which file will handle that url – works with IIRF
QueryString	page	<empty>		How to pass page id in the url – works with IIRF
OldURL	/news/today.asp	NULL		If the page was in the old system under differen url this wil generate 301 redirect from old url to new
DateTimeStamp	2008/06/13 1:12:05	Now()		When the row was edited
Page			**Content**	
PageID	324	UID		PK - Page unique ID
PageContent	<h1>Design Dep…	#MBD		HTML content to display as a main text for the page
PageSentID	2	0		FK: Sent. SentID – Describes type of sent event
Vars	var1 = 22	<empty>		Additional Predefined Variables
ScoreDesign	0.23	0.5		Design Score of that page
ScoreManufacturing	0.23	0.5		Manufacturing Score of that page
Enabled	1	1		Is the page enabled to view
MetaTitle	Design - MotionTouch	<empty>		Used to fill in meta title tag
MetaKeywords	Plastic, production	<empty>		Used to fill in meta keywords tag
MetaDescription	Mt is a leading....	<empty>		Used to fill in meta description tag
metaRobots	all	<empty>		Used to fill in meta robots tag
Signature	1/0	<empty>		Flag to indicate whether page belongs to signature pad microsite
PageType	Services	<empty>		Each web page was classified under a given type. The different types were: *Video, Download, Company Info, Other, Services, Contact Us, Form, Media Access, FT Quote, Case Study, PR*
Country			**Geo**	
CountryCode	UK	#MBD		PK - Unique ISO country code, + Ax for anonymous proxies
CountryName	United Kingdom	#MBD		Iso Country names + Anonymous Proxy
Keywords			**Scoring**	
KeywordsID	324	UID		PK - Keywords unique ID
Phrase	Plastic design	#MBD		Phrase to match search engine keywords against
DynamicWord	Plastic	#MBD		DynamicWord that dynamic wording feature uses

DynamicPicture	/images/pl01.jpg	#MBD		DynamicPicture that dynamic images feature uses
ScoreDesign	0.23	#MBD		Initial Score that will be assigned to user based on Keywords he used in SE
ScoreManufacturing	0.23	#MBD		
CSID	22	NULL		Shows related CS ID to given keyword
CSCategoryListID	2	NULL		Shows related cs category id
Panel			**Content**	
PanelID	324	UID		PK - Left panel unique id
PageID	22	#MBD		On which page display the panel section
PanelSectionID	3	#MBD		Which panel section to display on that page
PanelSectionOrder	50	#MBD		What's the order of the panels ASC
Enabled	1	#MBD		Is the rule enabled
PanelSection			**Content**	
PanelSectionID	324	UID		PK - PanelContent unique ID
PanelSectionContent	<tr><td>Favorite…	#MBD		Content that will be feed into Panel.Content
PanelSectionDesc	3 links	NULL		Contains basic description for the given section
Misc			**Content**	
MiscID	324	UID		PK - Misc unique ID
MiscContent	<tr><td>Featured...	#MBD		HTML code that can be attached on any page in main text area (featured Video, etc)
MiscDesc	Mentis video	NULL		Basic info on what's this misc do
Enabled	1	#MBD		Is the section enabled
SentName			**Tracking**	
SentNameID	324	UID		PK - Video unique ID
SentName	FastTrack Design	#MBD		Name of the sent event
Video			**Videos**	
VideoID	324	UID		PK - Video unique ID
Title	MT introduction	#MBD		The real user friendly title of the video
Description	Why Mot..	#MBD		HTML code to put under the video, the usual questions
Picture	\videos\324.jpg	#MBD		Path to video thumbnail
FileName	Intro to MT	#MBD		Video file name, without ext
Length	4:21	#MBD		Video Length (time)
Enabled	1	1		Is the video enable to access
VideoFile			**Videos**	
VideoFileID	324	UID		PK - VideoFile unique ID
VideoID	234	UID		FK: Video.VideoID
Ext	swf	#MBD		File ext that decides how to play the file
Size	1.5	#MBD		Size in MB

VideoCategory			**Videos**	
VideoCategoryID	324	UID		PK - VideoCategoryID
VideoCategoryListID	3214	UID		FK: VideoCategoryList.VideoCategoryListID
VideoID	324	UID		FK: Video.VideoID – which video falls in which category
VideoOrder	55	50		The order of the video I current category
VideoCategoryList			**Videos**	
VideoCategoryListID	324	UID		PK - unique identifier
CategoryName	Manufacturing	#MBD		User friendly way to name categories
PageID	2001	#MBD		Which page displays videos in that cat.
VideoLog			**Tracking**	
VideoLogID	324	UID		PK - Video Log unique ID
UserID	324	UID		FK: Motiontouch.UserID
MicroID	324	UID		FK: Motiontouch.MicroID
PageID	324	UID		FK: Page.PageID
VideoFileID	324	UID		FK: Video.VideoID
DateTimeStamp	12:00 01/12/2008	Now()		Time Stamp
History			**Tracking**	
HistoryID	324	UID		PK - Motiontouch unique ID
UserID	324	UID		User main ID aka UserID that identifies each single user
MicroID	324	UID		User micro ID aka MicroID that identifies each single visit of the user
Referer	http://www.googl..	0		Referrer requested from user browser
EngineNameID	1	0		FK: Describes Which search engine it was
KeywordsID	34	NULL		FK: Did it match any of the defined keywords
Keywords	Plastic design	0		Keywords extracted from referrer - if possible
LandPage	Page.PageID	#MBD		FK: Page.PageID Page that user used as a start page
CountryCode	UK	0		FK: Country.CoutryCode Country code of the user
IP	5416873	0		Dec representation of user IP address
GCC	HP11-03	0		Google Campain Code – if present
DateTimeStamp	12:00 01/12/2008	Now()		Time Stamp
MicroHistory			**Tracking**	
MicroHistoryID	324	UID		PK - History Motiontouch unique ID
UserID	324	UID		FK: Motiontouch.UserID
MicroID	324	UID		FK: Motiontouch.MicroID
PageID	324	UID		FK: Page.PageID
Sequence	3	#MBD		Sequence number for any given MicroID of visited pages – for paths
TimeSpent	21	0		Time in seconds that user spent on the page

LinkSectionID	2	NULL		Where the link was clicked, left menu, main menu, the site itself, features
Sent			**Tracking**	
SentID	324	UID		PK - SentMotiontouch unique ID
UserID	324	UID		FK: Motiontouch.UserID
MicroID	324	UID		FK: Motiontouch.MicroID
PageID	324	UID		FK: Page.PageID
DateTimeStamp	12:00 01/12/2008	Now()		Time Stamp
LeadType	2	<NULL>		Data provided by person creating leads. Type represents lead type i.e. design(1) or manufacturing(2)
LeadClass	1	<NULL>		Class represents lead quality, good(1) medium(0) or bad(-1)
EngineType			**Tracking**	
EngineTypeID	22	UID		PK – unique ID
EngineType	Search Engine	#MBD		User friendly name of engine type: search engine, directory listing
CS			**Content**	
CSID	324	UID		PK – unique ID
Title	Netronome	#MBD		Title of the CS
Description	Service type: des..	#MBD		Short desc for cs listings
Picture	/images/cs/324.jpg	#MBD		Path to img thumb for cs
PageID	324	#MBD		Where the actual CS is located
Visible	1	1		Is the CS visible in listings
Enabled	1	1		Is the CS accessible to users
CurrentID			**Tracking**	
IDName	motiontouch	#MBD		Which Site the id is for
ID	25658	#MBD		Current id
CScategoryList			**Content**	
CScategorylistID	234	UID		PK – unique Table id
CategoryName	design	#MBD		Category name
PageID	234	#MBD		Which page displays given category
CSCategory			**Content**	
CSCategoryID	2435	#MBD		PK – unique ID
CScategoryListID	45	#MBD		FK
CSID	234	#MBD		FK
CSOrder	34	#MBD		CS order in given category
User			**Content**	
UserUID	234	UID		PK – unique ID
UserID	234	#MBD		FK: user id
MicroID	234	#MBD		FK: user micro id

FullName	Dave Watson	#MBD		Full name
Name	Watt	#MBD		Desired login name
Pass	Dw343	#MBD		Password
Email	user@mail.com	#MBD		Email address
Newsletter	0	#MBD		Does the user wants to receive newsletter
IP	98463987496	#MBD		User ip (dec)
DateTimeStamp	2008-06-18	Now()		Time of record creation
File			**Content**	
FileID	3	UID		PK – unique ID
Name	Mt T&Cs	#MBD		Name of the file
Description	Terms and cond.	#MBD		A short desc. For downloads page
FileName	T&c.pdf	#MBD		An on-disc filename
Size	2.5	#MBD		Size in MB
Restricted	0	#MBD		Is the file available only to registered users
Enabled	1	#MBD		Is the file enabled for download
FileCategory			**Content**	
FileCategoryID	23	UID		PK – unique ID
FileCategoryListID	2	#MBD		FK- category list id
FileID	43	#MBD		Fk – FileID
FileOrder	50	#MBD		Order in category
FileCategoryList			**Content**	
FilecategorylistID	23	UID		PK – unique ID
CategoryName	Legal docs	#MBD		Name of the category
EngineName			**Tracking**	
EngineNameID	2	UID		PK – unique ID
EngineName	Google	#MBD		SE name
EngineTypeID	1	#MBD		Type of the SE
SiteSearch			**Tracking**	
SiteSearchID	23	UID		PK – unique ID
Phrase	Toys	#MBD		Phrase that user typed in to search box
PageID	24	#MBD		FK – Page id
UserID	23	#MBD		FK – user id
MicroID	54	#MBD		FK – micro id
Sequence	3	#MBD		Sequence on the site
DateTimeStamp	2008-06-18	Now()		Current date time stamp
FileHistory			**Tracking**	
FileHistoryID	45	UID		PK – unique ID

UserUID	34	NULL		FK – user – registration id
UserID	345	#MBD		FK –user id
MicroID	34	#MBD		FK –user micro id
PageID	34	#MBD		FK –page id
FileID	534	#MBD		FK –file id
DateTimeStamp	2008-06-18	Now()		Date Time stamp
Questionnaire			**Tracking**	
ID	23	#MBD		PK – Unique ID
QuestionnaireID	1	#MBD		Unique ID of questionnaire
QuestionID	24	#MBD		Unique ID of questions
Answer	3	#MBD		Unique ID for answers to questionnaire's question
AnswerDescription	Medium Business	#MBD		Description of answer
QHistory			**Tracking**	
QHistoryID	4	#MBD		PK – Unique ID
UserID	34	#MBD		FK –user id
MicroID	34	#MBD		FK –user micro id
QuestionnaireID	1	#MBD		FK – unique ID of questionnaire
QuestionID	3	#MBD		FK – unique ID of questions
Answer	3	#MBD		FK - unique ID for answers to questionnaire's question
Path	2.1	#MBD		Path indicating the options visitor chose in the questionnaire
SentFrom	<empty>	NULL		Point in the questionnaire at which it was sent
DateTimeStamp	2008-06-18	Now()		Date Time stamp
Link			**Tracking**	
QLinkID	1	#MBD		PK – Unique ID
QuestionnaireID	1	#MBD		FK - unique ID of questionnaire
QuestionID	24	#MBD		FK - unique ID of questions
Answer	3	#MBD		FK - unique ID for answers to questionnaire's question
LinkDesc	Design Overview	#MBD		Title to be displayed as a link
LinkPageID	39	#MBD		PageID of page to link to

PK: Primary Key

FK: Foreign Key

UID: Unique ID

A.3. *Enquiry quality score and marketing data retrieval*

A.3.1. First version of OTM

Figure A.3 shows a screenshot of a simple web page that was implemented for the first version of the OTM. This web page was used to access quality score and marketing data for visitors who had enquired. On this page, the user entered the *[mainID]* corresponding to the visitor whose enquiry needed to be scored.

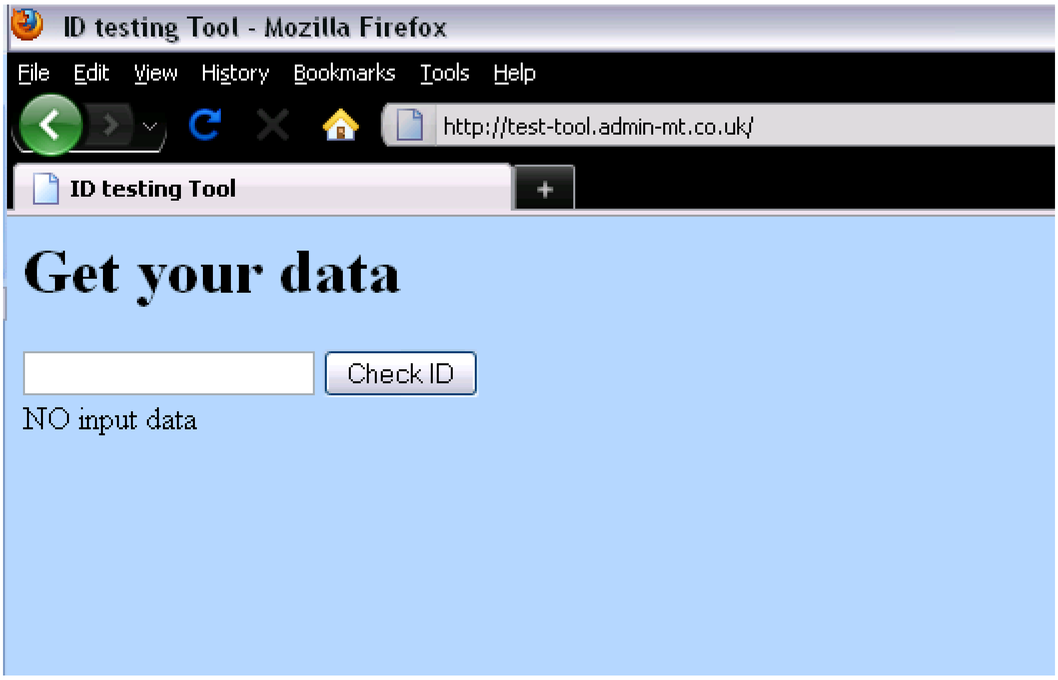


Figure A.3: Screenshot of web page used to view visitor's data

Figure A.4 shows the page displayed when the user had entered a valid *[mainID]*. The user could then score the enquiry as well as obtain marketing data that needed to be manually entered and stored in the visitor's MS CRM record.

Figure A.4: Screenshot of web page used to score enquiries

A.3.2. Improved OTM (implemented in Stage 2b)

Figure A.5 shows a screenshot of the links that were appended to email enquiries by the improved OTM. A user could click on the appropriate link to score the quality of an enquiry as well as specify the type of an enquiry. Upon clicking on a link, the user was taken to a web page (Figure A.6) that displayed marketing data related to the enquiry that was being scored.

```
Design
GOOD: http://www.motiontouch.com/tools/score.asp?LeadType=1&LeadClass=1&sentID=7125
MEDIUM: http://www.motiontouch.com/tools/score.asp?LeadType=1&LeadClass=0&sentID=7125
BAD: http://www.motiontouch.com/tools/score.asp?LeadType=1&LeadClass=-1&sentID=7125

Manufacturing
GOOD: http://www.motiontouch.com/tools/score.asp?LeadType=2&LeadClass=1&sentID=7125
MEDIUM: http://www.motiontouch.com/tools/score.asp?LeadType=2&LeadClass=0&sentID=7125
BAD: http://www.motiontouch.com/tools/score.asp?LeadType=2&LeadClass=-1&sentID=7125
```

Figure A.5: Screenshot of links that were appended to email enquiries.

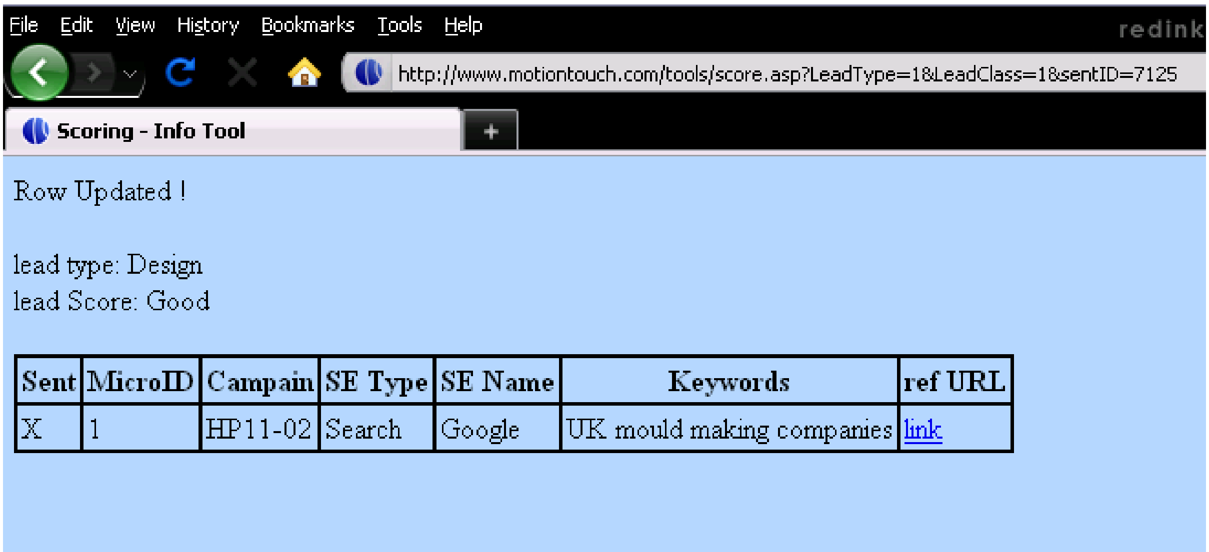


Figure A.6: Screenshot of web page used to score enquiries

APPENDIX B

The matching options used for the keywords in the two main Google AdWords campaigns that ran in the United Kingdom were defined as:

- **Broad match:** keyword *(no punctuation)*

  Showed advertisements for searches on similar phrases and relevant variations.

- **Phrase match:** "keyword"

  Showed advertisements for searches that match the exact phrase.

- **Exact match:** [keyword]

  Showed advertisements for searches that match the exact phrase exclusively.

- **Negative match:** -keyword

  Did not show advertisements for any search that includes that term.

The list of keywords is commercial in confidence at the time of writing. The list may be requested from the author or the collaborating company after 2 years (that is, from June 2013).

APPENDIX C

C.1. Stage 2: Data retrieval

C.1.1. Step 1: Extract and combine elements of browsing history

```
INSERT INTO motiontouch.Results
(userid, metatitle,
pageid,pagetype,pagelength,textlinkcount,imglinkcount,listcount,listwordcount
, microid,sequence, linksectionid,timespent, keywords)
select motiontouch.microhistory.userid, metatitle,
motiontouch.page.pageid,pagetype,pagelength,textlinkcount,imglinkcount,listco
unt,listwordcount,motiontouch.microhistory.microid,sequence,linksectionid,tim
espent,keywords from motiontouch.page, motiontouch.microhistory,
motiontouch.history
where motiontouch.page.pageid = motiontouch.microhistory.pageid
and motiontouch.microhistory.userid = motiontouch.history.userid
and motiontouch.microhistory.microid = motiontouch.history.microid
and motiontouch.history.microid = '1'
and keywords not like '%motiontouch%'
and keywords not like '%motion touch%'
and keywords not like 'MONTION TOUCH'
and keywords not like 'www.%'
and keywords not like 'ttp:%'
and keywords not like 'http%'
and keywords not like '%.com'
and keywords not like '%.co.uk'
and keywords not like '%.org'
and keywords not like '%.net'
and keywords != ''
and countrycode != 'NL'
and countrycode != 'CN'
and countrycode != 'PL'
and motiontouch.microhistory.userid != 4209 -- Malicious user
and enginenameid not in ('2','4','30','32','33','34')
order by motiontouch.microhistory.userid,microid,sequence
```

C.1.2. Step 2: Retrieve all non-converted records

```
insert into motiontouch.NonConvAll
(userid,pages,totaltimespent)
select userid, count(*)as Pages, sum (TimeSpent) as TotalTimeSpent  from
motiontouch.microhistory
where userid not in
(select userid from motiontouch.sent where microid = '1')
and microid = '1'
group by userid
order by pages, userid asc
```

C.1.3. Step 3: Retrieve data associated with conversions and non-conversions

Data associated with conversions

```
select userid,metatitle, pageid,pagetype,microid,sequence,timespent,keywords
from motiontouch.results
where userid in
(select userid from motiontouch.sent
where leadtype !=''         -- legitimate lead as indicated by receptionist
and datetimestamp <= '2010-06-01'
and microid = '1'
)
order by userid desc,microid,sequence
```

Data associated with non-conversions (excluding bounces)

```
select userid,metatitle, pageid,pagetype,microid,sequence,timespent,keywords
from motiontouch.results
where userid in
(select  motiontouch.NonConvAll.userid from
motiontouch.NonConvAll,motiontouch.history
where motiontouch.NonConvAll.userid = motiontouch.history.userid
and totaltimespent is not NULL -- not a bounce
and datetimestamp <= '2010-06-01'
)
order by userid desc,microid,sequence
```

APPENDIX D

INITIAL DATA EXPLORATIONS

D.1. First exploration

Linear Regression

“Linear regression” found the following prediction model (LR0.1) using ***Training Data***:

Conversion = +0.395528 -0.000281744*VideoTS -0.0238632*Services -0.0588671*[Unsent Form] -0.000545676*[Media AccessTS] -0.000553811*[Time on site (s) b4 enquiring] +0.000789077*[Total Time on site (s)] -0.107448*[Browsed ContactUs]

Equation D.1: Prediction model LR0.1.

The accuracy measures for the prediction model LR0.1 are shown in Table D.1.

StdErr	RSq	StdDev
0.81	0.34	0.41

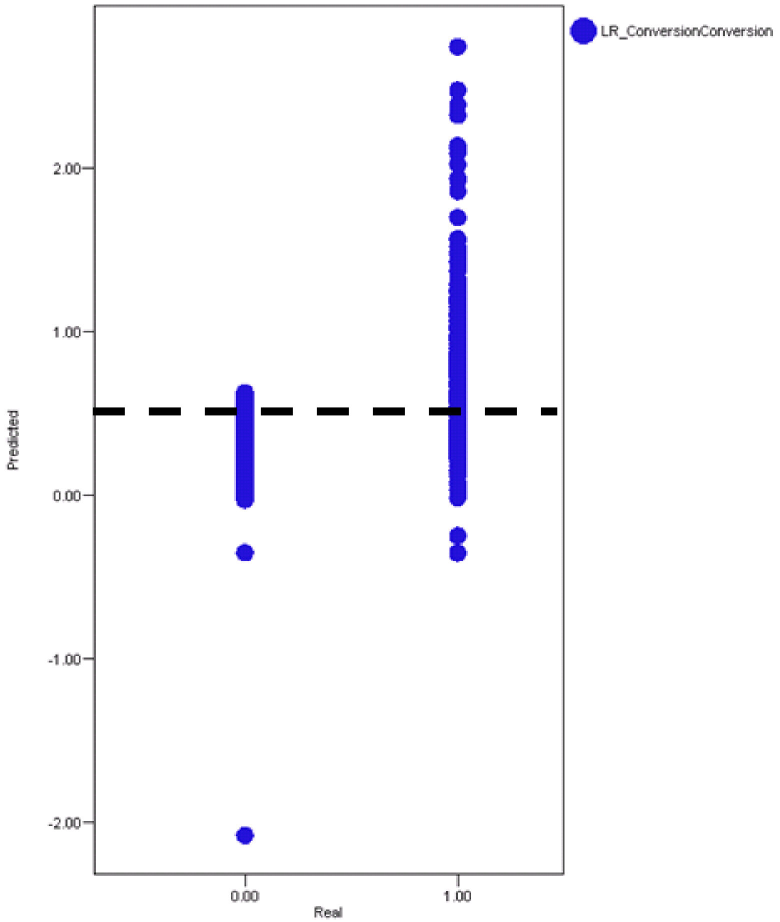
Table D.1: Accuracy measures for prediction model LR0.1 derived from Training Dataset.

It can be seen from Table D.1 that prediction model LR0.1 had a high standard error and low RSq value. This suggested that the rule was not accurate in predicting conversions. The model was tested with ***Test Data***. The accuracy measures are shown in Table D.2.

StdErr	RSq	StdDev
0.86	0.27	0.43

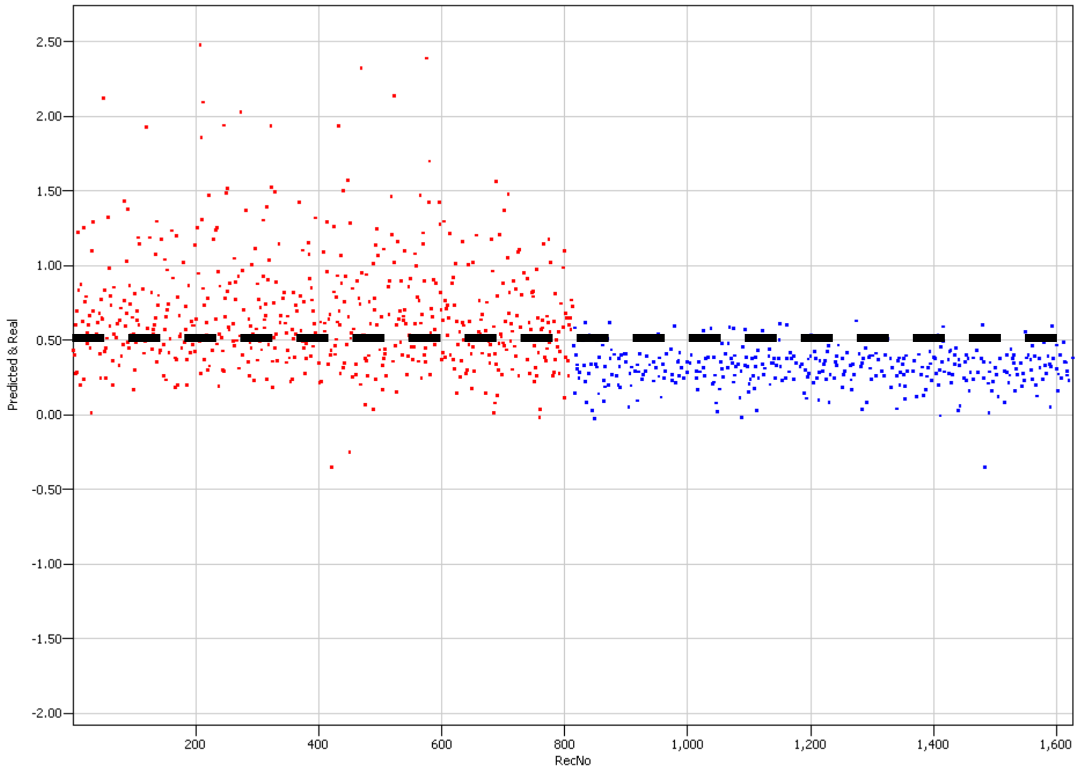
Table D.2: Accuracy measures for LR0.1 after testing the model with *Test Data*.

Graph D.1 shows how the predicted values varied compared to the real value in ***Test Data***.



Graph D.1: Predicted vs Real Value for LR0.1 (using Test Dataset).

Graph D.2 shows how the predicted values and real varied for individual records in ***Test Data***. From Graph D.1 and Graph D.2, it was observed that most target values of 1 had a predicted value that was greater than 0.50.



Graph D.2: Predicted and real value for individual records (using Test Dataset).

A confusion matrix for LR0.1 (using ***Test Data***) is shown in Table D.3 when the boundary value for identifying 1s was set to 0.50.

Actual/Predicted	**1**	**0**	**Total**
1	529	284	813
0	31	782	813
Total	560	1066	1626

Table D.3: Confusion matrix for LR0.1.

The classification probability and efficiency of LR0.1 were derived from the values shown in Table D.3. They were 80.62% and 61.25% respectively. The model predicted 65.07% of the target 1s correctly which suggested good accuracy. Also Table D.3

shows that LR0.1 (with a boundary for 1s set at 0.50) predicted that 560 instances of the data had a conversion value of 1. Out of these instances 94.46% (529) were predicted correctly.

Find Laws

“Find Laws” found the following prediction model (FL0.1) using ***Training Data***:

Conversion = (3.70233e-007*[Time on site (s) b4 enquiring]*[Total Time on site (s)]*[Total Time on site (s)]+1.00196*[Time on site (s) b4 enquiring]*[Total Time on site (s)]-1.00198*[Time on site (s) b4 enquiring]*[Time on site (s) b4 enquiring]+0.000376032*[Total Time on site (s)]-0.000864941)/([Time on site (s) b4 enquiring]*[Total Time on site (s)]-0.99881*[Time on site (s) b4 enquiring]*[Time on site (s) b4 enquiring]+8.59465*[Media Access]*[Time on site (s) b4 enquiring])

Equation D.2: Prediction model FLD.1

The accuracy measures for the prediction model FLD.1 are shown in Table D.4.

StdErr	RSq	StdDev
0.30	0.91	0.15

Table D.4: Accuracy measures for prediction model FL0.1 derived from *Training Data*

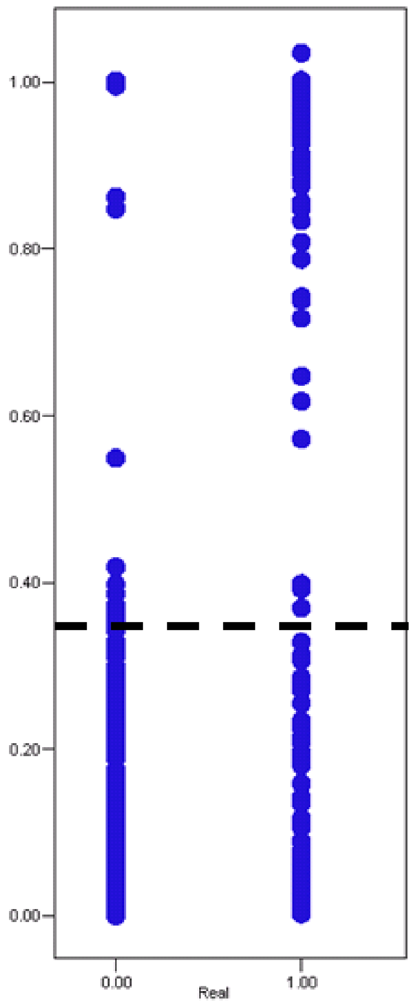
The model was tested with ***Test Data***. The accuracy measures are shown inTable D.5.

StdErr	RSq	StdDev
0.48	0.77	0.24

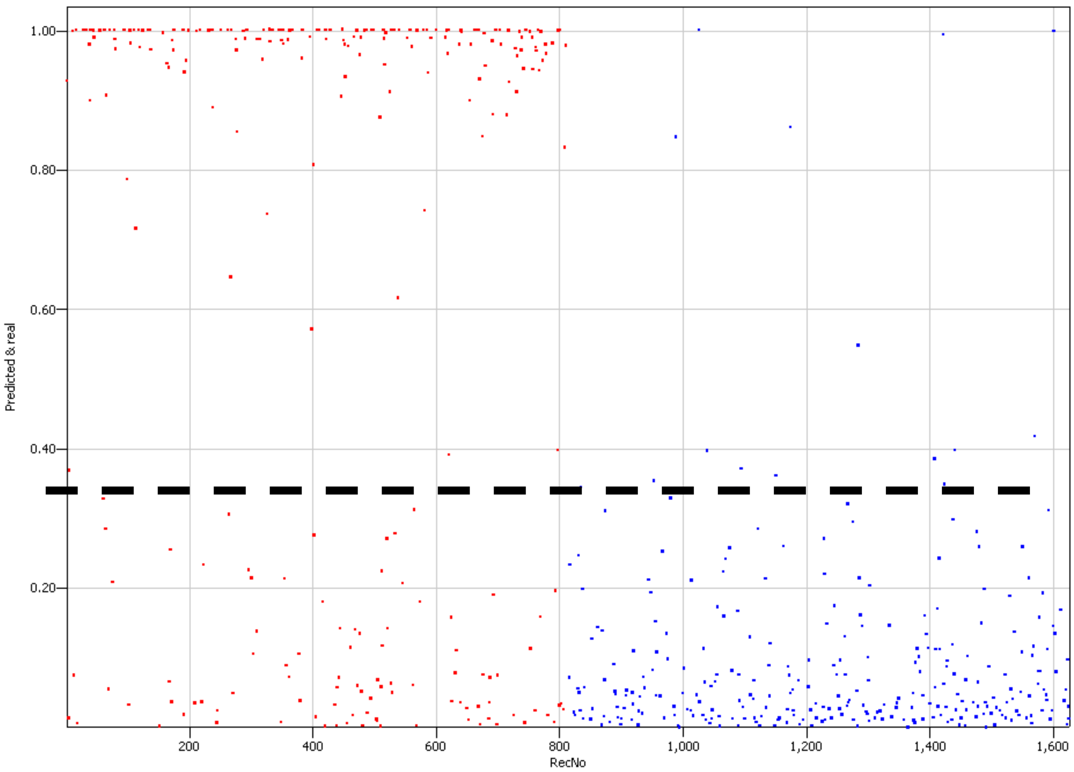
Table D.5: Accuracy measures for FL0.1 after testing FL0.1 with *Test Data*.

RSq for FL0.1 was high suggesting a model that was more accurate than LR0.1. From Graph D.3 and Graph D.4, it can be seen that predicted values that were greater than

0.35 were likely to correctly predict 1s. Table D.6 shows the confusion matrix for FL0.1 when the boundary value for identifying 1s was set to 0.35.



Graph D.3: Predicted vs Real Value for FL0.1.



Graph D.4: Predicted and Real value vs record number.

Actual/Predicted	1	0	Total
1	707	106	813
0	13	800	813
Total	720	806	1626

Table D.6: Confusion matrix for FL0.1 when boundary value for identifying 1s was set to 0.35

The classification probability and efficiency of FL0.1 were derived from the values shown in Table D.6. They were 92.68% and 85.36% respectively. The model predicted 86.96% of the target 1s correctly which suggested good accuracy.

Also Table D.6 shows that FL0.1 (with a boundary for 1s set at 0.35) predicted that 720 instances of the data had a conversion value of 1. Out of these instances 98.19% (707) were predicted correctly.

Table D.7 shows the confusion matrix for FL0.1 if the boundary value for identifying 1s was set to 0.5 (as with LR0.1).

Actual/Predicted	**1**	**0**	**Total**
1	704	109	813
0	6	807	813
Total	710	916	1626

Table D.7: Confusion matrix for FL0.1 when boundary value for identifying 1s was set to 0.5

The classification probability and efficiency of FL0.1 derived from the values shown in Table D.7 were 92.92% and 85.84% respectively. The model predicted 86.59% of the target 1s correctly.

Table D.7 also shows that FL0.1 (with a boundary for 1s set at 0.5) predicted that 710 instances of the data had a conversion value of 1. Out of these instances 99.15% (704) were predicted correctly.

Changing the boundary value for identifying 1s from 0.35 to 0.5 had little impact on FL0.1. It was however easier to compare FL0.1 with LR0.1 when they both had the same boundary value for identifying 1s. FL0.1 appeared to be a much better model than LR0.1 with higher classification probability efficiency and correct predictions for target 1s. Table D.8 compares the measures for FL0.1 and LR0.1.

Model	cp	ce	Accuracy in target set	Accuracy predicted set
LR0.1	80.62	61.25	65.07	94.46
FL0.1	92.92	85.84	86.59	99.15

Table D.8: Accuracy measures of LR0.1 compared to FL0.1.

Neural Networks

"Neural Networks" were given the same input parameters as LR0.1 and FL0.1. ***Training Data*** was used to derive mode NN0.1, which was then tested with ***Test Data***. The accuracy measures for the tested model NN0.1 is shown in Table D.9.

cerr	cp	cf	ce
10.15	89.85	0%	79.70%

Table D.9: Accuracy measures for NN0.1.

This model had a classification probability that was higher than a naïve prediction (cp =50%) as well as a high classification efficiency. Table D.10 provides a breakdown of the model's predictions.

Target	No of records	Error %	Correct%	Undefined%
Yes	751	17.22	82.78	0.00
No	751	3.08	96.92	0.00
Total	1502	10.15	89.85	0.00

Table D.10: Breakdown of predictions for NN0.1.

It can be seen that NN0.1 predicted 96.01% of conversions correctly in the target set. The confusion matrix for NN0.1 is shown in Table D.11.

Actual/Predicted	Yes	No	Total
Yes	673	140	813
No	25	788	813
Total	698	928	1626

Table D.11: Confusion matrix for NN0.1.

From Table D.11, it can be seen that model NN0.1 predicted a total of 698 records as having a value of Yes for conversion. Out of these predictions, 96.42% (673) were predicted correctly.

D.2. Second Exploration

In the second exploration, *[Time spent on site (s)]* was removed as an attribute from the data.

Linear Regression

Linear regression found the following prediction model (LR0.2) using ***Training Data***:

Conversion = +0.525756 -0.0185710*Services -0.107911*[Unsent Form] -0.106552*[Media Access] +0.00317758*PRTS +0.000245167*[Time on site (s) b4 enquiring] -0.111964*[Browsed ContactUs] +0.0371440*FTQuote

~~Equation D.3: Accuracy measures for prediction model LR0.2.~~

The accuracy measures for the prediction model LR0.2 are shown Table D.12.

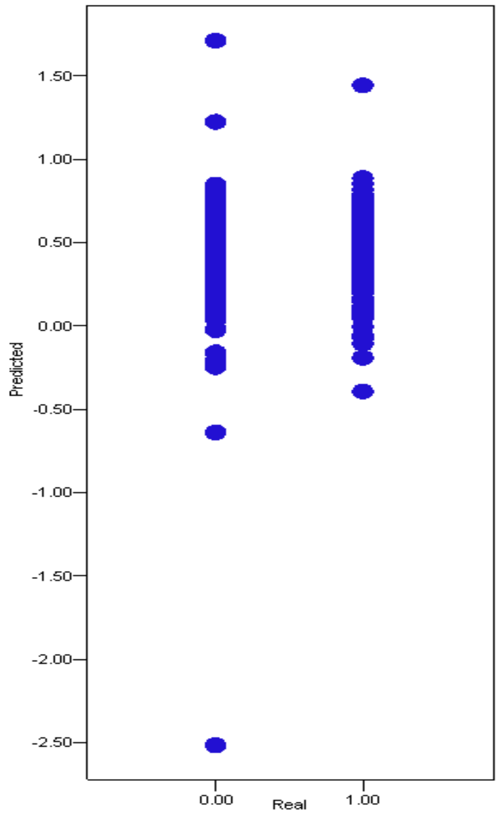
StdErr	RSq	StdDev
0.9719	0.0555	0.4861

Table D.12: Accuracy measures for prediction model LR0.2

It can be seen from Table D.12 that prediction model LR0.2 had a high standard error and low RSq value. This suggested that the rule was not accurate. The model was tested with ***Test Data***. The accuracy measures are shown in Table D.13. These confirmed that LR0.2 had low accuracy.

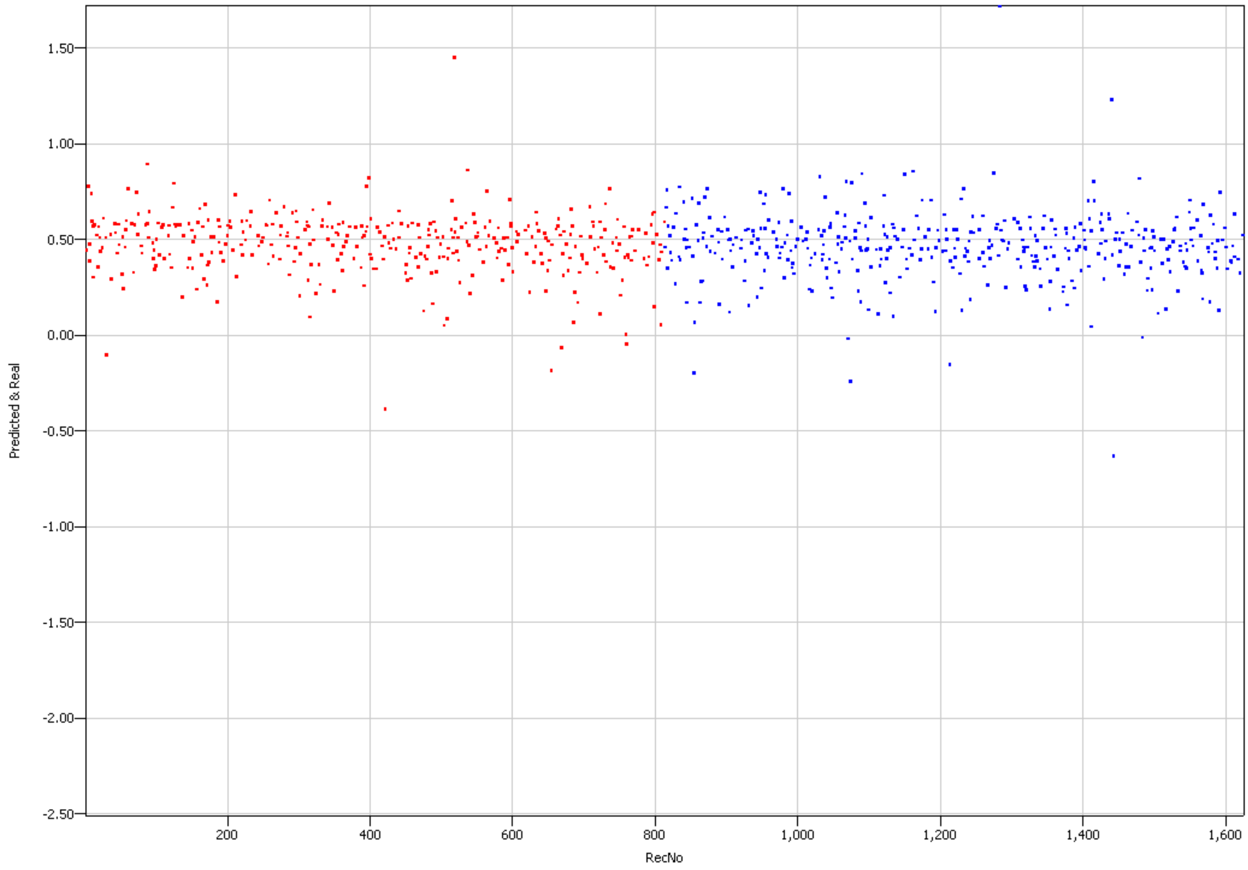
StdErr	RSq	StdDev
1.01	-0.03	0.51

Table D.13: Accuracy measures for LR0.2 after testing the model with *Test Data*.



Graph D.5: Predicted values against real value for LR0.2.

Graph D.5 shows the values predicted by LR0.2 plotted against the real values. It can be seen that the range of predicted values for 0 and 1 overlap for most of the data.



Graph D.6: Predicted, real vs record number.

Graph D.6 shows how the predicted values varied compared to the real values for individual records. It confirmed that prediction model LR0.2 had low accuracy.

Find Laws

"Find Laws" was unable to find a rule.

Neural Networks

The accuracy measures for model NN0.2 generated by "Neural Networks" is shown in Table D.14.

cerr	cp	cf	ce
23.64%	76.36%	0%	52.73%

Table D.14: Accuracy measures for model NN0.2.

This model had a classification probability that was higher than a naïve prediction (cp =50%) as well as a high classification efficiency.

Table D.15 provides a breakdown of the model's predictions. It can be seen that NN0.2 predicted 72.17% of Yes correctly and 80.56% of No correctly. This indicated that NN0.2 was better at predicting No than it was at predicting Yes. The confusion matrix for NN0.1 is shown in Table D.16.

Target	No of records	Error %	Correct%	Undefined%
Yes	751	27.83	72.17	0.00
No	751	19.44	80.56	0.00
Total	1502	23.64	76.36	0.00

Table D.15: Breakdown of predictions for model NN0.2.

From Table D.16, it can be seen that model NN0.2 predicted a total of 688 records as Yes. Out of these predictions, 542 were actually correct (79%) while 146 (21%) were incorrect. NN0.2 was more accurate than a naïve model.

Actual/Predicted	Yes	No	Total
Yes	542	209	751
No	146	605	751
Total	688	814	1502

Table D.16: Confusion matrix for NN0.2.

D.3. Third Exploration

"Neural Networks" was better than "Linear Regression" and "Find laws" at finding predictive models and non-linear dependencies. However, the "Linear regression" provided a method of determining how much predictive power each attribute had in a linear relationship using F-Ratio.

In order to determine whether the attributes identified by the “Linear Regression” could generate more accurate models, “Neural Networks” was run with attributes identified by the predictive model LR0.2. Table D.17 shows the F-Ratio for these attributes.

Attribute	**F-Ratio**
[Services]	9.28
[Unsent Form]	19.38
[Media Access]	21.77
[PRTS]	3.35
[Time on site (s) b4 enquiring]	30.58
[Browsed ContactUs]	17.82
[FTQuote]	4.06

Table D.17: F-Ratio for attributes identified by predictive model LR0.2

The accuracy measures of the new NN model (NN0.3) are shown in Table D.18.

cerr	**cp**	**cf**	**ce**
24.77%	75.23%	0%	50.47%

Table D.18: Accuracy measures for model NN0.3.

It can be seen from Table D.18 and Table D.19 that while there was no improvement in accuracy measures, the difference between the accuracy measures of NN0.2 and NN0.3 was actually small. The breakdown of NN0.3’s predictions is shown in Table D.19.

Target	No of records	Error %	Correct%	Undefined%
Yes	751	29.16	70.84	0.00
No	751	20.37	79.63	0.00
Total	1502	24.77	75.23	0.00

Table D.19: Breakdown of predictions for model NN0.3.

From Table D.20, it can be seen that model NN0.3 predicted a total of 685 records as Yes. Out of these predictions, 532 were actually correct (78%) while 153 (22%) were incorrect. This was better than a naïve prediction.

Actual/Predicted	Yes	No	Total
Yes	532	219	751
No	153	598	751
Total	685	817	1502

Table D.20: Confusion for NN0.3.

APPENDIX E

Data used to measure the performance of landing pages described in Chapter 6 are provided in this Appendix. Conversion rate were obtained from Google AdWords and bounce rate was obtained from Google Analytics.

E.1. Data used to measure Change 1

Day	Page	Clicks	Conv	Conv Rate
23-Jun	Design Overview	66	1	1.52%
24-Jun	Design Overview	62	2	3.23%
25-Jun	Design Overview	58	3	5.17%
26-Jun	Design Overview	48	1	2.08%
27-Jun	Design Overview	46	3	6.52%
28-Jun	Design Overview	49	1	2.04%
29-Jun	Design Overview	56	2	3.57%
30-Jun	Design Overview	22	1	4.55%
01-Jul	Design Overview	23	2	8.70%
02-Jul	Design Overview	21	0	0.00%
03-Jul	Design Overview	16	0	0.00%
04-Jul	Design Overview	6	0	0.00%
During Overlap		144	5	3.47%
29-Jun	Product Idea	2	0	0.00%
30-Jun	Product Idea	32	1	3.13%
01-Jul	Product Idea	23	1	4.35%
02-Jul	Product Idea	23	3	13.04%
03-Jul	Product Idea	33	1	3.03%
04-Jul	Product Idea	39	1	2.56%
05-Jul	Product Idea	36	3	8.33%
06-Jul	Product Idea	59	4	6.78%
07-Jul	Product Idea	58	2	3.45%
08-Jul	Product Idea	50	0	0.00%
09-Jul	Product Idea	46	4	8.70%
10-Jul	Product Idea	55	4	7.27%
During Overlap		152	7	4.61%

E.2. Data used to measure Change 2

Conversion Rate (over 3 weeks) of Product Idea page for views from Advertisement 1

Day	Advertisement 1	Views	Conversion	Conversion Rate
29/06/2008	Got a great product idea?	1	0	0
30/06/2008	Got a great product idea?	13	1	7.69%
01/07/2008	Got a great product idea?	9	0	0.00%
02/07/2008	Got a great product idea?	7	1	14.29%
03/07/2008	Got a great product idea?	10	1	10.00%
04/07/2008	Got a great product idea?	16	0	0.00%
05/07/2008	Got a great product idea?	15	1	6.67%
06/07/2008	Got a great product idea?	30	3	10.00%
07/07/2008	Got a great product idea?	27	1	3.70%
08/07/2008	Got a great product idea?	20	0	0.00%
09/07/2008	Got a great product idea?	17	2	11.76%
10/07/2008	Got a great product idea?	27	4	14.81%
11/07/2008	Got a great product idea?	13	1	7.69%
12/07/2008	Got a great product idea?	6	0	0.00%
13/07/2008	Got a great product idea?	10	0	0.00%
14/07/2008	Got a great product idea?	14	0	0.00%
15/07/2008	Got a great product idea?	18	2	11.11%
16/07/2008	Got a great product idea?	19	0	0.00%
17/07/2008	Got a great product idea?	17	1	5.88%
18/07/2008	Got a great product idea?	12	1	8.33%
19/07/2008	Got a great product idea?	13	1	7.69%
20/07/2008	Got a great product idea?	20	3	15.00%
21/07/2008	Got a great product idea?	33	3	9.09%
22/07/2008	Got a great product idea?	11	0	0.00%
		378	26	6.88%

Conversion Rate (over 3 weeks) of Product Idea page for views from Advertisement 2

Day	Advertisement 2	Views	Conversion	Conversion Rate
29/06/2008	Got a great invention?	1	0	0.00%
30/06/2008	Got a great invention?	19	0	0.00%
01/07/2008	Got a great invention?	14	1	7.14%
02/07/2008	Got a great invention?	16	2	12.50%
03/07/2008	Got a great invention?	23	0	0.00%
04/07/2008	Got a great invention?	23	1	4.35%
05/07/2008	Got a great invention?	21	2	9.52%
06/07/2008	Got a great invention?	29	1	3.45%
07/07/2008	Got a great invention?	31	1	3.23%
08/07/2008	Got a great invention?	30	0	0.00%
09/07/2008	Got a great invention?	29	2	6.90%
10/07/2008	Got a great invention?	28	0	0.00%
11/07/2008	Got a great invention?	30	1	3.33%
12/07/2008	Got a great invention?	10	1	10.00%
13/07/2008	Got a great invention?	9	0	0.00%
14/07/2008	Got a great invention?	23	2	8.70%
15/07/2008	Got a great invention?	29	0	0.00%
16/07/2008	Got a great invention?	20	0	0.00%
17/07/2008	Got a great invention?	36	2	5.56%
18/07/2008	Got a great invention?	24	0	0.00%
19/07/2008	Got a great invention?	22	0	0.00%
20/07/2008	Got a great invention?	28	1	3.57%
21/07/2008	Got a great invention?	30	1	3.33%
22/07/2008	Got a great invention?	3	0	0.00%
		528	18	3.41%

Bounce rate of Product Idea page for visits from Advertisement 1 and Advertisement 2

www.motiontouch.com				
Top Landing Pages				
June 29, 2008 - July 22, 2008				
# --				
Landing Page	**Ad Content**	**Entrances**	**Bounces**	**Bounce Rate**
Advertisement 2	Got a great invention?	451	205	45.45%
Advertisement 1	Got a great product idea?	321	120	37.38%

Conversion Rate (over 12 weeks) of dynamic Product Idea page for views from Advertisement 1

Day	Ad	Clicks	Conv.	Conv. rate
22/07/2008	Got a great product idea?	18	2	11.11%
23/07/2008	Got a great product idea?	20	2	10.00%
24/07/2008	Got a great product idea?	20	1	5.00%
25/07/2008	Got a great product idea?	26	3	11.54%
26/07/2008	Got a great product idea?	8	0	0.00%
27/07/2008	Got a great product idea?	16	2	12.50%
28/07/2008	Got a great product idea?	21	3	14.29%
29/07/2008	Got a great product idea?	11	0	0.00%
30/07/2008	Got a great product idea?	14	0	0.00%
31/07/2008	Got a great product idea?	8	0	0.00%
01/08/2008	Got a great product idea?	9	1	11.11%
02/08/2008	Got a great product idea?	9	0	0.00%
03/08/2008	Got a great product idea?	18	0	0.00%
04/08/2008	Got a great product idea?	26	0	0.00%
05/08/2008	Got a great product idea?	36	0	0.00%
06/08/2008	Got a great product idea?	23	1	4.35%
07/08/2008	Got a great product idea?	22	1	4.55%
08/08/2008	Got a great product idea?	23	2	8.70%
09/08/2008	Got a great product idea?	12	0	0.00%
10/08/2008	Got a great product idea?	14	0	0.00%
11/08/2008	Got a great product idea?	28	0	0.00%
12/08/2008	Got a great product idea?	31	1	3.23%
13/08/2008	Got a great product idea?	19	0	0.00%
14/08/2008	Got a great product idea?	22	3	13.64%
15/08/2008	Got a great product idea?	18	1	5.56%
16/08/2008	Got a great product idea?	13	0	0.00%
17/08/2008	Got a great product idea?	27	0	0.00%
18/08/2008	Got a great product idea?	22	1	4.55%
19/08/2008	Got a great product idea?	19	2	10.53%
20/08/2008	Got a great product idea?	14	1	7.14%
21/08/2008	Got a great product idea?	15	0	0.00%
22/08/2008	Got a great product idea?	10	1	10.00%
23/08/2008	Got a great product idea?	4	0	0.00%
24/08/2008	Got a great product idea?	10	0	0.00%
25/08/2008	Got a great product idea?	16	0	0.00%
26/08/2008	Got a great product idea?	22	0	0.00%
27/08/2008	Got a great product idea?	22	0	0.00%
28/08/2008	Got a great product idea?	17	0	0.00%
29/08/2008	Got a great product idea?	18	1	5.56%

30/08/2008	Got a great product idea?	11	0	0.00%
31/08/2008	Got a great product idea?	20	0	0.00%
01/09/2008	Got a great product idea?	13	2	15.38%
02/09/2008	Got a great product idea?	17	0	0.00%
03/09/2008	Got a great product idea?	18	0	0.00%
04/09/2008	Got a great product idea?	16	0	0.00%
05/09/2008	Got a great product idea?	21	0	0.00%
06/09/2008	Got a great product idea?	10	0	0.00%
07/09/2008	Got a great product idea?	15	0	0.00%
08/09/2008	Got a great product idea?	19	0	0.00%
09/09/2008	Got a great product idea?	17	1	5.88%
10/09/2008	Got a great product idea?	17	0	0.00%
11/09/2008	Got a great product idea?	32	1	3.12%
12/09/2008	Got a great product idea?	14	1	7.14%
13/09/2008	Got a great product idea?	12	0	0.00%
14/09/2008	Got a great product idea?	23	1	4.35%
15/09/2008	Got a great product idea?	33	3	9.09%
16/09/2008	Got a great product idea?	29	1	3.45%
17/09/2008	Got a great product idea?	30	0	0.00%
18/09/2008	Got a great product idea?	32	3	9.38%
19/09/2008	Got a great product idea?	19	0	0.00%
20/09/2008	Got a great product idea?	11	1	9.09%
21/09/2008	Got a great product idea?	19	1	5.26%
22/09/2008	Got a great product idea?	32	1	3.12%
23/09/2008	Got a great product idea?	25	1	4.00%
24/09/2008	Got a great product idea?	29	0	0.00%
25/09/2008	Got a great product idea?	23	2	8.70%
26/09/2008	Got a great product idea?	9	0	0.00%
27/09/2008	Got a great product idea?	20	2	10.00%
28/09/2008	Got a great product idea?	18	2	11.11%
29/09/2008	Got a great product idea?	29	1	3.45%
30/09/2008	Got a great product idea?	27	0	0.00%
01/10/2008	Got a great product idea?	29	0	0.00%
02/10/2008	Got a great product idea?	32	2	6.25%
03/10/2008	Got a great product idea?	13	0	0.00%
04/10/2008	Got a great product idea?	14	0	0.00%
05/10/2008	Got a great product idea?	19	0	0.00%
06/10/2008	Got a great product idea?	21	0	0.00%
07/10/2008	Got a great product idea?	24	0	0.00%
08/10/2008	Got a great product idea?	20	0	0.00%
09/10/2008	Got a great product idea?	13	0	0.00%
10/10/2008	Got a great product idea?	11	0	0.00%
11/10/2008	Got a great product idea?	7	0	0.00%
12/10/2008	Got a great product idea?	9	0	0.00%

13/10/2008	Got a great product idea?	27	1	3.70%
14/10/2008	Got a great product idea?	18	1	5.56%
15/10/2008	Got a great product idea?	13	0	0.00%
16/10/2008	Got a great product idea?	21	1	4.76%
17/10/2008	Got a great product idea?	11	1	9.09%
18/10/2008	Got a great product idea?	14	1	7.14%
19/10/2008	Got a great product idea?	11	0	0.00%
20/10/2008	Got a great product idea?	17	0	0.00%
21/10/2008	Got a great product idea?	25	2	8.00%
22/10/2008	Got a great product idea?	25	2	8.00%
		1755	64	3.65%

Conversion Rate (over 12 weeks) of dynamic Product Idea page for views from Advertisement 2

Day	Ad	Clicks	Conv.	Conv. rate
22/07/2008	Got a great invention?	15	2	13.33%
23/07/2008	Got a great invention?	20	1	5.00%
24/07/2008	Got a great invention?	18	2	11.11%
25/07/2008	Got a great invention?	28	0	0.00%
26/07/2008	Got a great invention?	14	0	0.00%
27/07/2008	Got a great invention?	25	2	8.00%
28/07/2008	Got a great invention?	26	3	11.54%
29/07/2008	Got a great invention?	14	0	0.00%
30/07/2008	Got a great invention?	16	1	6.25%
31/07/2008	Got a great invention?	19	2	10.53%
01/08/2008	Got a great invention?	19	0	0.00%
02/08/2008	Got a great invention?	15	0	0.00%
03/08/2008	Got a great invention?	34	1	2.94%
04/08/2008	Got a great invention?	26	0	0.00%
05/08/2008	Got a great invention?	33	3	9.09%
06/08/2008	Got a great invention?	28	2	7.14%
07/08/2008	Got a great invention?	22	2	9.09%
08/08/2008	Got a great invention?	22	0	0.00%
09/08/2008	Got a great invention?	22	1	4.55%
10/08/2008	Got a great invention?	19	1	5.26%
11/08/2008	Got a great invention?	43	4	9.30%
12/08/2008	Got a great invention?	29	1	3.45%
13/08/2008	Got a great invention?	43	5	11.63%
14/08/2008	Got a great invention?	23	0	0.00%
15/08/2008	Got a great invention?	16	0	0.00%

16/08/2008	Got a great invention?	12	1	8.33%
17/08/2008	Got a great invention?	27	2	7.41%
18/08/2008	Got a great invention?	31	2	6.45%
19/08/2008	Got a great invention?	32	0	0.00%
20/08/2008	Got a great invention?	12	0	0.00%
21/08/2008	Got a great invention?	18	0	0.00%
22/08/2008	Got a great invention?	11	0	0.00%
23/08/2008	Got a great invention?	5	2	40.00%
24/08/2008	Got a great invention?	8	1	12.50%
25/08/2008	Got a great invention?	18	0	0.00%
26/08/2008	Got a great invention?	28	3	10.71%
27/08/2008	Got a great invention?	19	1	5.26%
28/08/2008	Got a great invention?	23	0	0.00%
29/08/2008	Got a great invention?	21	1	4.76%
30/08/2008	Got a great invention?	25	0	0.00%
31/08/2008	Got a great invention?	30	2	6.67%
01/09/2008	Got a great invention?	22	5	22.73%
02/09/2008	Got a great invention?	34	0	0.00%
03/09/2008	Got a great invention?	33	0	0.00%
04/09/2008	Got a great invention?	29	0	0.00%
05/09/2008	Got a great invention?	33	2	6.06%
06/09/2008	Got a great invention?	20	0	0.00%
07/09/2008	Got a great invention?	34	3	8.82%
08/09/2008	Got a great invention?	38	2	5.26%
09/09/2008	Got a great invention?	58	5	8.62%
10/09/2008	Got a great invention?	41	2	4.88%
11/09/2008	Got a great invention?	48	2	4.17%
12/09/2008	Got a great invention?	27	1	3.70%
13/09/2008	Got a great invention?	31	1	3.23%
14/09/2008	Got a great invention?	31	1	3.23%
15/09/2008	Got a great invention?	45	4	8.89%
16/09/2008	Got a great invention?	56	4	7.14%
17/09/2008	Got a great invention?	46	0	0.00%
18/09/2008	Got a great invention?	46	3	6.52%
19/09/2008	Got a great invention?	24	0	0.00%
20/09/2008	Got a great invention?	20	0	0.00%
21/09/2008	Got a great invention?	27	0	0.00%
22/09/2008	Got a great invention?	46	2	4.35%
23/09/2008	Got a great invention?	38	1	2.63%
24/09/2008	Got a great invention?	39	2	5.13%
25/09/2008	Got a great invention?	39	3	7.69%
26/09/2008	Got a great invention?	23	1	4.35%
27/09/2008	Got a great invention?	25	0	0.00%
28/09/2008	Got a great invention?	25	3	12.00%

29/09/2008	Got a great invention?	29	1	3.45%
30/09/2008	Got a great invention?	29	0	0.00%
01/10/2008	Got a great invention?	31	2	6.45%
02/10/2008	Got a great invention?	27	3	11.11%
03/10/2008	Got a great invention?	23	1	4.35%
04/10/2008	Got a great invention?	17	0	0.00%
05/10/2008	Got a great invention?	23	0	0.00%
06/10/2008	Got a great invention?	29	0	0.00%
07/10/2008	Got a great invention?	34	0	0.00%
08/10/2008	Got a great invention?	24	2	8.33%
09/10/2008	Got a great invention?	28	2	7.14%
10/10/2008	Got a great invention?	25	1	4.00%
11/10/2008	Got a great invention?	20	0	0.00%
12/10/2008	Got a great invention?	30	2	6.67%
13/10/2008	Got a great invention?	27	3	11.11%
14/10/2008	Got a great invention?	43	3	6.98%
15/10/2008	Got a great invention?	23	0	0.00%
16/10/2008	Got a great invention?	30	2	6.67%
17/10/2008	Got a great invention?	35	0	0.00%
18/10/2008	Got a great invention?	13	1	7.69%
19/10/2008	Got a great invention?	21	1	4.76%
20/10/2008	Got a great invention?	25	2	8.00%
21/10/2008	Got a great invention?	24	0	0.00%
22/10/2008	Got a great invention?	28	0	0.00%
		2525	121	4.79%

Conversion Rate (over 12 weeks) of Design Overview page for views from Advertisement 1

Day	Advertisement 2	Clicks	Conv.	Conv. rate
28/03/2008	Got a great product idea?	2	1	50.00%
29/03/2008	Got a great product idea?	0	0	0.00%
30/03/2008	Got a great product idea?	4	0	0.00%
31/03/2008	Got a great product idea?	3	0	0.00%
01/04/2008	Got a great product idea?	6	1	16.67%
02/04/2008	Got a great product idea?	6	0	0.00%
03/04/2008	Got a great product idea?	0	0	0.00%
04/04/2008	Got a great product idea?	5	0	0.00%
05/04/2008	Got a great product idea?	2	0	0.00%
06/04/2008	Got a great product idea?	1	0	0.00%
07/04/2008	Got a great product idea?	3	0	0.00%
08/04/2008	Got a great product idea?	3	0	0.00%

09/04/2008	Got a great product idea?	4	0	0.00%
10/04/2008	Got a great product idea?	3	0	0.00%
11/04/2008	Got a great product idea?	0	0	0.00%
12/04/2008	Got a great product idea?	2	0	0.00%
13/04/2008	Got a great product idea?	4	0	0.00%
14/04/2008	Got a great product idea?	5	0	0.00%
15/04/2008	Got a great product idea?	2	0	0.00%
16/04/2008	Got a great product idea?	3	0	0.00%
17/04/2008	Got a great product idea?	0	0	0.00%
18/04/2008	Got a great product idea?	2	0	0.00%
19/04/2008	Got a great product idea?	2	0	0.00%
20/04/2008	Got a great product idea?	4	0	0.00%
21/04/2008	Got a great product idea?	6	0	0.00%
22/04/2008	Got a great product idea?	4	0	0.00%
23/04/2008	Got a great product idea?	6	0	0.00%
24/04/2008	Got a great product idea?	7	0	0.00%
25/04/2008	Got a great product idea?	4	0	0.00%
26/04/2008	Got a great product idea?	1	0	0.00%
27/04/2008	Got a great product idea?	5	0	0.00%
28/04/2008	Got a great product idea?	6	1	16.67%
29/04/2008	Got a great product idea?	6	0	0.00%
30/04/2008	Got a great product idea?	0	0	0.00%
01/05/2008	Got a great product idea?	1	1	100.00%
01/05/2008	Got a great product idea?	12	1	8.33%
02/05/2008	Got a great product idea?	21	1	4.76%
03/05/2008	Got a great product idea?	16	0	0.00%
04/05/2008	Got a great product idea?	21	1	4.76%
05/05/2008	Got a great product idea?	14	0	0.00%
06/05/2008	Got a great product idea?	18	1	5.56%
07/05/2008	Got a great product idea?	27	1	3.70%
08/05/2008	Got a great product idea?	18	0	0.00%
09/05/2008	Got a great product idea?	23	0	0.00%
10/05/2008	Got a great product idea?	13	1	7.69%
11/05/2008	Got a great product idea?	15	0	0.00%
12/05/2008	Got a great product idea?	21	0	0.00%
13/05/2008	Got a great product idea?	9	3	33.33%
14/05/2008	Got a great product idea?	9	1	11.11%
15/05/2008	Got a great product idea?	12	0	0.00%
16/05/2008	Got a great product idea?	10	1	10.00%
17/05/2008	Got a great product idea?	5	1	20.00%
18/05/2008	Got a great product idea?	2	0	0.00%
19/05/2008	Got a great product idea?	4	0	0.00%
20/05/2008	Got a great product idea?	5	0	0.00%
21/05/2008	Got a great product idea?	1	0	0.00%

22/05/2008	Got a great product idea?	6	0	0.00%
23/05/2008	Got a great product idea?	3	1	33.33%
24/05/2008	Got a great product idea?	2	0	0.00%
25/05/2008	Got a great product idea?	1	0	0.00%
26/05/2008	Got a great product idea?	1	0	0.00%
27/05/2008	Got a great product idea?	2	0	0.00%
28/05/2008	Got a great product idea?	0	0	0.00%
29/05/2008	Got a great product idea?	1	0	0.00%
30/05/2008	Got a great product idea?	0	0	0.00%
31/05/2008	Got a great product idea?	3	0	0.00%
01/06/2008	Got a great product idea?	0	0	0.00%
02/06/2008	Got a great product idea?	0	0	0.00%
03/06/2008	Got a great product idea?	0	0	0.00%
04/06/2008	Got a great product idea?	2	0	0.00%
05/06/2008	Got a great product idea?	0	0	0.00%
06/06/2008	Got a great product idea?	0	0	0.00%
07/06/2008	Got a great product idea?	0	0	0.00%
08/06/2008	Got a great product idea?	0	0	0.00%
09/06/2008	Got a great product idea?	0	0	0.00%
10/06/2008	Got a great product idea?	1	0	0.00%
11/06/2008	Got a great product idea?	0	0	0.00%
12/06/2008	Got a great product idea?	0	0	0.00%
13/06/2008	Got a great product idea?	11	0	0.00%
14/06/2008	Got a great product idea?	18	1	5.56%
15/06/2008	Got a great product idea?	24	0	0.00%
16/06/2008	Got a great product idea?	19	2	10.53%
17/06/2008	Got a great product idea?	29	1	3.45%
18/06/2008	Got a great product idea?	24	1	4.17%
19/06/2008	Got a great product idea?	16	0	0.00%
20/06/2008	Got a great product idea?	20	2	10.00%
21/06/2008	Got a great product idea?	22	1	4.55%
22/06/2008	Got a great product idea?	25	0	0.00%
23/06/2008	Got a great product idea?	22	1	4.55%
24/06/2008	Got a great product idea?	28	1	3.57%
25/06/2008	Got a great product idea?	24	2	8.33%
26/06/2008	Got a great product idea?	14	0	0.00%
27/06/2008	Got a great product idea?	20	3	15.00%
28/06/2008	Got a great product idea?	16	1	6.25%
		742	33	4.45%

Conversion Rate (over 12 weeks) of Design Overview page for views from Advertisement 2

Day	Advertisement 2	Clicks	Conv.	Conv. rate
28/03/2008	Got a great invention?	51	2	3.92%
29/03/2008	Got a great invention?	41	4	9.76%
30/03/2008	Got a great invention?	38	0	0.00%
31/03/2008	Got a great invention?	56	2	3.57%
01/04/2008	Got a great invention?	49	1	2.04%
02/04/2008	Got a great invention?	46	2	4.35%
03/04/2008	Got a great invention?	33	0	0.00%
04/04/2008	Got a great invention?	50	4	8.00%
05/04/2008	Got a great invention?	32	2	6.25%
06/04/2008	Got a great invention?	48	1	2.08%
07/04/2008	Got a great invention?	43	1	2.33%
08/04/2008	Got a great invention?	43	0	0.00%
09/04/2008	Got a great invention?	41	3	7.32%
10/04/2008	Got a great invention?	40	1	2.50%
11/04/2008	Got a great invention?	33	0	0.00%
12/04/2008	Got a great invention?	27	1	3.70%
13/04/2008	Got a great invention?	37	1	2.70%
14/04/2008	Got a great invention?	45	1	2.22%
15/04/2008	Got a great invention?	33	0	0.00%
16/04/2008	Got a great invention?	36	0	0.00%
17/04/2008	Got a great invention?	22	0	0.00%
18/04/2008	Got a great invention?	28	1	3.57%
19/04/2008	Got a great invention?	42	1	2.38%
20/04/2008	Got a great invention?	33	1	3.03%
21/04/2008	Got a great invention?	42	0	0.00%
22/04/2008	Got a great invention?	25	0	0.00%
23/04/2008	Got a great invention?	26	0	0.00%
24/04/2008	Got a great invention?	29	1	3.45%
25/04/2008	Got a great invention?	31	1	3.23%
26/04/2008	Got a great invention?	8	0	0.00%
27/04/2008	Got a great invention?	24	1	4.17%
28/04/2008	Got a great invention?	51	1	1.96%
29/04/2008	Got a great invention?	30	2	6.67%
30/04/2008	Got a great invention?	38	0	0.00%
01/05/2008	Got a great invention?	15	3	20.00%
01/05/2008	Got a great invention?	11	0	0.00%
02/05/2008	Got a great invention?	9	1	11.11%
03/05/2008	Got a great invention?	15	1	6.67%
04/05/2008	Got a great invention?	22	1	4.55%

05/05/2008	Got a great invention?	15	1	6.67%
06/05/2008	Got a great invention?	31	2	6.45%
07/05/2008	Got a great invention?	24	0	0.00%
08/05/2008	Got a great invention?	30	0	0.00%
09/05/2008	Got a great invention?	31	3	9.68%
10/05/2008	Got a great invention?	25	0	0.00%
11/05/2008	Got a great invention?	29	0	0.00%
12/05/2008	Got a great invention?	43	3	6.98%
13/05/2008	Got a great invention?	37	0	0.00%
14/05/2008	Got a great invention?	53	1	1.89%
15/05/2008	Got a great invention?	51	3	5.88%
16/05/2008	Got a great invention?	44	1	2.27%
17/05/2008	Got a great invention?	39	0	0.00%
18/05/2008	Got a great invention?	54	2	3.70%
19/05/2008	Got a great invention?	63	1	1.59%
20/05/2008	Got a great invention?	62	0	0.00%
21/05/2008	Got a great invention?	55	2	3.64%
22/05/2008	Got a great invention?	48	0	0.00%
23/05/2008	Got a great invention?	41	2	4.88%
24/05/2008	Got a great invention?	34	2	5.88%
25/05/2008	Got a great invention?	58	3	5.17%
26/05/2008	Got a great invention?	63	2	3.17%
27/05/2008	Got a great invention?	61	0	0.00%
28/05/2008	Got a great invention?	60	4	6.67%
29/05/2008	Got a great invention?	48	0	0.00%
30/05/2008	Got a great invention?	46	0	0.00%
31/05/2008	Got a great invention?	45	2	4.44%
01/06/2008	Got a great invention?	53	0	0.00%
02/06/2008	Got a great invention?	55	0	0.00%
03/06/2008	Got a great invention?	49	1	2.04%
04/06/2008	Got a great invention?	47	0	0.00%
05/06/2008	Got a great invention?	51	0	0.00%
06/06/2008	Got a great invention?	41	0	0.00%
07/06/2008	Got a great invention?	46	0	0.00%
08/06/2008	Got a great invention?	35	1	2.86%
09/06/2008	Got a great invention?	50	1	2.00%
10/06/2008	Got a great invention?	48	0	0.00%
11/06/2008	Got a great invention?	54	2	3.70%
12/06/2008	Got a great invention?	45	1	2.22%
13/06/2008	Got a great invention?	40	1	2.50%
14/06/2008	Got a great invention?	21	0	0.00%
15/06/2008	Got a great invention?	28	2	7.14%
16/06/2008	Got a great invention?	41	3	7.32%
17/06/2008	Got a great invention?	30	2	6.67%

18/06/2008	Got a great invention?	40	0	0.00%
19/06/2008	Got a great invention?	28	2	7.14%
20/06/2008	Got a great invention?	22	1	4.55%
21/06/2008	Got a great invention?	25	0	0.00%
22/06/2008	Got a great invention?	24	1	4.17%
23/06/2008	Got a great invention?	44	0	0.00%
24/06/2008	Got a great invention?	34	1	2.94%
25/06/2008	Got a great invention?	34	1	2.94%
26/06/2008	Got a great invention?	34	1	2.94%
27/06/2008	Got a great invention?	26	0	0.00%
28/06/2008	Got a great invention?	33	0	0.00%
		3591	96	2.67%

Bounce rate of Design Overview page for visits from Advertisement 1 and Advertisement 2 (March 28, 2008 - June 28, 2008)

# --				
www.motiontouch.com				
Top Landing Pages				
March 28, 2008 - June 28, 2008				
# --				
Page	Ad Content	Entrances	Bounces	Bounce Rate
Design Overview (Advertisment 2)	Got a great invention?	3008	1355	45.05%
Design Overview (Advertisment 1)	Got a great product idea?	636	252	39.62%

Bounce rate of dynamic Product Idea page for visits from Advertisement 1 and Advertisement 2 (July 22, 2008 - October 22, 2008)

# --				
www.motiontouch.com				
Top Landing Pages				
July 22, 2008 - October 22, 2008				
# --				
Page	Ad Content	Entrances	Bounces	Bounce Rate
Product Idea (Advertisement 2)	Got a great invention?	2092	964	46.08%
Product Idea (Advertisement 1)	Got a great product idea?	1441	618	42.89%

E.3. Data used to measure Change 3

Conversion Rate (over 3 months) of dynamic Product Idea page for views from Advertisement 1 after change in visual design

Day	Ad	Clicks	Conv.	Conv. rate
01/11/2008	Got a great product idea?	15	2	13.33%
02/11/2008	Got a great product idea?	17	0	0.00%
03/11/2008	Got a great product idea?	29	2	6.90%
04/11/2008	Got a great product idea?	36	2	5.56%
05/11/2008	Got a great product idea?	26	0	0.00%
06/11/2008	Got a great product idea?	27	1	3.70%
07/11/2008	Got a great product idea?	16	0	0.00%
08/11/2008	Got a great product idea?	16	2	12.50%
09/11/2008	Got a great product idea?	38	1	2.63%
10/11/2008	Got a great product idea?	29	1	3.45%
11/11/2008	Got a great product idea?	38	1	2.63%
12/11/2008	Got a great product idea?	29	2	6.90%
13/11/2008	Got a great product idea?	27	2	7.41%
14/11/2008	Got a great product idea?	17	0	0.00%
15/11/2008	Got a great product idea?	14	1	7.14%
16/11/2008	Got a great product idea?	30	0	0.00%
17/11/2008	Got a great product idea?	49	3	6.12%
18/11/2008	Got a great product idea?	27	3	11.11%
19/11/2008	Got a great product idea?	30	1	3.33%
20/11/2008	Got a great product idea?	33	2	6.06%
21/11/2008	Got a great product idea?	21	1	4.76%
22/11/2008	Got a great product idea?	13	0	0.00%
23/11/2008	Got a great product idea?	21	1	4.76%
24/11/2008	Got a great product idea?	27	2	7.41%
25/11/2008	Got a great product idea?	28	1	3.57%
26/11/2008	Got a great product idea?	26	0	0.00%
27/11/2008	Got a great product idea?	27	2	7.41%
28/11/2008	Got a great product idea?	18	2	11.11%
29/11/2008	Got a great product idea?	35	1	2.86%
30/11/2008	Got a great product idea?	12	0	0.00%
01/12/2008	Got a great product idea?	24	0	0.00%
02/12/2008	Got a great product idea?	20	0	0.00%
03/12/2008	Got a great product idea?	18	1	5.56%
04/12/2008	Got a great product idea?	16	2	12.50%
05/12/2008	Got a great product idea?	11	1	9.09%
06/12/2008	Got a great product idea?	16	0	0.00%

07/12/2008	Got a great product idea?	14	0	0.00%
08/12/2008	Got a great product idea?	18	2	11.11%
09/12/2008	Got a great product idea?	18	1	5.56%
10/12/2008	Got a great product idea?	15	1	6.67%
11/12/2008	Got a great product idea?	22	1	4.55%
12/12/2008	Got a great product idea?	16	0	0.00%
13/12/2008	Got a great product idea?	19	1	5.26%
14/12/2008	Got a great product idea?	17	1	5.88%
15/12/2008	Got a great product idea?	20	1	5.00%
16/12/2008	Got a great product idea?	21	0	0.00%
17/12/2008	Got a great product idea?	17	1	5.88%
18/12/2008	Got a great product idea?	8	0	0.00%
19/12/2008	Got a great product idea?	5	0	0.00%
20/12/2008	Got a great product idea?	9	1	11.11%
21/12/2008	Got a great product idea?	5	0	0.00%
22/12/2008	Got a great product idea?	7	1	14.29%
23/12/2008	Got a great product idea?	7	0	0.00%
24/12/2008	Got a great product idea?	2	0	0.00%
25/12/2008	Got a great product idea?	2	0	0.00%
26/12/2008	Got a great product idea?	7	0	0.00%
27/12/2008	Got a great product idea?	6	0	0.00%
28/12/2008	Got a great product idea?	9	1	11.11%
29/12/2008	Got a great product idea?	14	1	7.14%
30/12/2008	Got a great product idea?	12	1	8.33%
31/12/2008	Got a great product idea?	8	0	0.00%
01/01/2009	Got a great product idea?	11	3	27.27%
02/01/2009	Got a great product idea?	25	1	4.00%
03/01/2009	Got a great product idea?	16	0	0.00%
04/01/2009	Got a great product idea?	16	0	0.00%
05/01/2009	Got a great product idea?	17	0	0.00%
06/01/2009	Got a great product idea?	19	1	5.26%
07/01/2009	Got a great product idea?	7	0	0.00%
08/01/2009	Got a great product idea?	10	0	0.00%
09/01/2009	Got a great product idea?	3	0	0.00%
10/01/2009	Got a great product idea?	5	1	20.00%
11/01/2009	Got a great product idea?	9	0	0.00%
12/01/2009	Got a great product idea?	12	1	8.33%
13/01/2009	Got a great product idea?	8	1	12.50%
14/01/2009	Got a great product idea?	3	0	0.00%
15/01/2009	Got a great product idea?	4	0	0.00%
16/01/2009	Got a great product idea?	1	0	0.00%
17/01/2009	Got a great product idea?	2	0	0.00%
18/01/2009	Got a great product idea?	2	0	0.00%
19/01/2009	Got a great product idea?	1	0	0.00%

20/01/2009	Got a great product idea?	0	0	0.00%
27/01/2009	Got a great product idea?	0	0	0.00%
28/01/2009	Got a great product idea?	2	0	0.00%
29/01/2009	Got a great product idea?	0	0	0.00%
30/01/2009	Got a great product idea?	0	0	0.00%
31/01/2009	Got a great product idea?	1	0	0.00%
		1348	62	4.60%

Conversion Rate (over 3 months) of dynamic Product Idea page for views from Advertisement 2 after change in visual design

Day	Ad	Clicks	Conv.	Conv. rate
01/11/2008	Got a great invention?	16	0	0.00%
02/11/2008	Got a great invention?	22	2	9.09%
03/11/2008	Got a great invention?	43	1	2.33%
04/11/2008	Got a great invention?	43	0	0.00%
05/11/2008	Got a great invention?	37	0	0.00%
06/11/2008	Got a great invention?	31	1	3.23%
07/11/2008	Got a great invention?	26	0	0.00%
08/11/2008	Got a great invention?	22	1	4.55%
09/11/2008	Got a great invention?	30	3	10.00%
10/11/2008	Got a great invention?	50	2	4.00%
11/11/2008	Got a great invention?	41	0	0.00%
12/11/2008	Got a great invention?	38	1	2.63%
13/11/2008	Got a great invention?	40	2	5.00%
14/11/2008	Got a great invention?	25	0	0.00%
15/11/2008	Got a great invention?	17	1	5.88%
16/11/2008	Got a great invention?	50	1	2.00%
17/11/2008	Got a great invention?	50	0	0.00%
18/11/2008	Got a great invention?	41	1	2.44%
19/11/2008	Got a great invention?	47	1	2.13%
20/11/2008	Got a great invention?	48	3	6.25%
21/11/2008	Got a great invention?	33	1	3.03%
22/11/2008	Got a great invention?	23	0	0.00%
23/11/2008	Got a great invention?	28	1	3.57%
24/11/2008	Got a great invention?	49	2	4.08%
25/11/2008	Got a great invention?	47	2	4.26%
26/11/2008	Got a great invention?	50	0	0.00%
27/11/2008	Got a great invention?	40	1	2.50%
28/11/2008	Got a great invention?	27	2	7.41%
29/11/2008	Got a great invention?	28	0	0.00%

30/11/2008	Got a great invention?	33	2	6.06%
01/12/2008	Got a great invention?	46	1	2.17%
02/12/2008	Got a great invention?	49	2	4.08%
03/12/2008	Got a great invention?	30	0	0.00%
04/12/2008	Got a great invention?	31	2	6.45%
05/12/2008	Got a great invention?	37	3	8.11%
06/12/2008	Got a great invention?	32	1	3.12%
07/12/2008	Got a great invention?	44	0	0.00%
08/12/2008	Got a great invention?	46	2	4.35%
09/12/2008	Got a great invention?	60	1	1.67%
10/12/2008	Got a great invention?	41	1	2.44%
11/12/2008	Got a great invention?	42	2	4.76%
12/12/2008	Got a great invention?	38	1	2.63%
13/12/2008	Got a great invention?	26	2	7.69%
14/12/2008	Got a great invention?	26	0	0.00%
15/12/2008	Got a great invention?	32	1	3.12%
16/12/2008	Got a great invention?	33	0	0.00%
17/12/2008	Got a great invention?	26	1	3.85%
18/12/2008	Got a great invention?	25	0	0.00%
19/12/2008	Got a great invention?	16	2	12.50%
20/12/2008	Got a great invention?	10	1	10.00%
21/12/2008	Got a great invention?	16	0	0.00%
22/12/2008	Got a great invention?	19	0	0.00%
23/12/2008	Got a great invention?	15	0	0.00%
24/12/2008	Got a great invention?	4	0	0.00%
25/12/2008	Got a great invention?	6	0	0.00%
26/12/2008	Got a great invention?	16	0	0.00%
27/12/2008	Got a great invention?	10	0	0.00%
28/12/2008	Got a great invention?	12	0	0.00%
29/12/2008	Got a great invention?	23	0	0.00%
30/12/2008	Got a great invention?	37	3	8.11%
31/12/2008	Got a great invention?	14	1	7.14%
01/01/2009	Got a great invention?	25	1	4.00%
02/01/2009	Got a great invention?	35	2	5.71%
03/01/2009	Got a great invention?	44	0	0.00%
04/01/2009	Got a great invention?	48	3	6.25%
05/01/2009	Got a great invention?	55	2	3.64%
06/01/2009	Got a great invention?	55	3	5.45%
07/01/2009	Got a great invention?	61	3	4.92%
08/01/2009	Got a great invention?	48	0	0.00%
09/01/2009	Got a great invention?	45	0	0.00%
10/01/2009	Got a great invention?	40	2	5.00%
11/01/2009	Got a great invention?	47	1	2.13%
12/01/2009	Got a great invention?	72	4	5.56%

13/01/2009	Got a great invention?	78	0	0.00%
14/01/2009	Got a great invention?	71	0	0.00%
15/01/2009	Got a great invention?	67	2	2.99%
16/01/2009	Got a great invention?	39	2	5.13%
17/01/2009	Got a great invention?	49	1	2.04%
18/01/2009	Got a great invention?	66	4	6.06%
19/01/2009	Got a great invention?	76	1	1.32%
20/01/2009	Got a great invention?	8	0	0.00%
27/01/2009	Got a great invention?	10	1	10.00%
28/01/2009	Got a great invention?	75	1	1.33%
29/01/2009	Got a great invention?	52	3	5.77%
30/01/2009	Got a great invention?	51	4	7.84%
31/01/2009	Got a great invention?	58	3	5.17%
		3212	101	3.14%

Conversion Rate (over 3 months) of dynamic Product Idea page for views from Advertisement 1 before change in visual design is the same as the data for conversion rate (over 12 weeks) of dynamic Product Idea page for views from Advertisement 1 (found in Section E.2)

Conversion Rate (over 3 months) of dynamic Product Idea page for views from Advertisement 2 before change in visual design is the same as the data for conversion rate (over 12 weeks) of dynamic Product Idea page for views from Advertisement 2 (found in Section E.2)

Conversion rate of Inventor campaign over 2 years (Oct 2007 – Oct 2009)

Month	Views	Conv.	Conv. rate		
Oct-07	1027	30	2.92%		
Nov-07	1369	52	3.80%		
Dec-07	804	24	2.99%		
Jan-08	1541	59	3.83%		
Feb-08	1098	34	3.10%		
Mar-08	1595	50	3.13%		
Apr-08	1167	29	2.49%		
May-08	1558	54	3.47%	Avg Conv from Oct-May	3.22%

Jun-08	1525	41	2.69%	Change 1 made here	
Jul-08	1295	71	5.48%	Change 2 made here	
Aug-08	1376	53	3.85%		
Sep-08	1665	72	4.32%		
Oct-08	1406	49	3.49%	Change 3 made here	
Nov-08	1846	67	3.63%		
Dec-08	1265	45	3.56%		
Jan-09	1449	51	3.52%		
Feb-09	1271	39	3.07%		
Mar-09	968	28	2.89%		
01/04/2009	70	2	2.86%		
01/07/2009	243	14	5.76%		
01/08/2009	275	13	4.73%		
01/09/2009	395	14	3.54%		
01/10/2009	197	6	3.05%	Avg Conv from Jul-Oct	3.84%

E.4. Data used to measure Change 4

Conversion Rate (over 2 months) of Plastic Manufacturing page

Month	Landing Page	Views	Conv.	Conv. rate
01/07/2008	Plastic Manufacturing	1549	30	1.94%
01/08/2008	Plastic Manufacturing	85	1	1.18%
		1634	31	1.90%

Conversion Rate (over 2 months) of Product Manufacture page

Month	Landing Page	Views	Conv.	Conv. rate
01/07/2008	Product Manufacture	73	2	2.74%
01/07/2008	Product Manufacture	32	0	0.00%
01/07/2008	Product Manufacture	51	2	3.92%
01/08/2008	Product Manufacture	434	10	2.30%
01/08/2008	Product Manufacture	378	7	1.85%
		968	21	2.17%

Conversion Rate (over 2 months) for Plastic Manufacturing and Product Manufacture page

www.motiontouch.com				
Top Landing Pages				
July 1, 2008 - August 31, 2008				
Page	Ad Content	Entrances	Bounces	Bounce Rate
Plastic Manufacturing	Plastic Manufacturer	1894	1115	58.87%
Product Manufacture	Plastic Manufacturer	706	311	44.05%

E.5. Data used to measure Change 5a

Conversion Rate (over 9 months) for Plastic Manufacturing page

Month	Landing Page	Views	Conv.	Conv. rate
01/01/2008	Plastic Manufacturing	991	18	1.82%
01/02/2008	Plastic Manufacturing	927	10	1.08%
01/03/2008	Plastic Manufacturing	1201	14	1.17%
01/04/2008	Plastic Manufacturing	374	2	0.53%
01/04/2008	Plastic Manufacturing	438	6	1.37%
01/05/2008	Plastic Manufacturing	712	10	1.40%
01/06/2008	Plastic Manufacturing	2885	20	0.69%
01/07/2008	Plastic Manufacturing	1549	30	1.94%
01/08/2008	Plastic Manufacturing	85	1	1.18%
		9162	111	1.21%

Conversion Rate (over 9 months) for Picture-segmented page

Month	Landing Page	Views	Conv.	Conv. rate
01/01/2009	Picture-segmented	2060	31	1.50%
01/02/2009	Picture-segmented	1807	45	2.49%
01/03/2009	Picture-segmented	1606	40	2.49%
01/04/2009	Picture-segmented	1451	39	2.69%
01/05/2009	Picture-segmented	780	28	3.59%
01/05/2009	Picture-segmented	225	0	0.00%
01/06/2009	Picture-segmented	407	17	4.18%
01/07/2009	Picture-segmented	532	14	2.63%
01/08/2009	Picture-segmented	536	23	4.29%
		9404	237	2.52%

Bounce Rate (over 9 months) for Plastic Manufacturing page

www.motiontouch.com				
Top Landing Pages				
January 1, 2008 - August 31, 2008				
Page	**Ad Content**	**Entrances**	**Bounces**	**Bounce Rate**
Plastic Manufacturing	Plastic Manufacturer	15818	10005	63.25%

Bounce Rate (over 9 months) for Picture-segmented page

# --				
www.motiontouch.com				
Top Landing Pages				
January 1, 2009 - August 31, 2009				
# --				
Page	Ad Content	Entrances	Bounces	Bounce Rate
Picture-Segemented	Plastic Manufacturer	7353	3583	48.73%

E.6. Data used to measure Change 5b

Conversion Rate (over 8 months) for Picture-segmented page

Month	Page	Views	Conv.	Conv. rate
01/05/2009	Picture-segmented	1005	28	2.79%
01/06/2009	Picture-segmented	407	17	4.18%
01/07/2009	Picture-segmented	532	14	2.63%
01/08/2009	Picture-segmented	536	23	4.29%
01/09/2009	Picture-segmented	553	19	3.44%
01/10/2009	Picture-segmented	620	19	3.06%
01/11/2009	Picture-segmented	575	17	2.96%
01/12/2009	Picture-segmented	317	13	4.10%
		4545	150	3.30%

Conversion Rate (over 8 months) for Questionnaire page

Month	Ad	Views	Conv.	Conv. rate
01/05/2009	Questionnaire	360	15	4.17%
01/06/2009	Questionnaire	622	25	4.02%
01/07/2009	Questionnaire	573	30	5.24%
01/08/2009	Questionnaire	496	19	3.83%
01/09/2009	Questionnaire	589	19	3.23%
01/10/2009	Questionnaire	590	15	2.54%
01/11/2009	Questionnaire	566	20	3.53%
01/12/2009	Questionnaire	323	13	4.02%
		4119	156	3.79%

Bounce Rate for Questionnaire page and Picture-segmented page

# --				
www.motiontouch.com				
Top Landing Pages				
May 1, 2009 - December 31, 2009				
# --				
Page	Ad Content	Entrances	Bounces	Bounce Rate
Questionnaire	Plastic Manufacturer	3324	1934	58.18%
Picture-Segmented	Plastic Manufacturer	3064	1391	45.40%

Conversion Rate (over 4 months) for Plastic Manufacturing page

Month	Ad	Clicks	Conv.	Conv. rate
01/05/2008	Plastic Manufacturing	712	10	1.40%
01/06/2008	Plastic Manufacturing	2885	20	0.69%
01/07/2008	Plastic Manufacturing	1549	30	1.94%
01/08/2008	Plastic Manufacturing	85	1	1.18%
		5231	61	1.17%

Conversion Rate (over 4 months) for Questionnaire page

Month	Ad	Clicks	Conv.	Conv. rate
01/05/2009	Questionnaire	360	15	4.17%
01/06/2009	Questionnaire	622	25	4.02%
01/07/2009	Questionnaire	573	30	5.24%
01/08/2009	Questionnaire	496	19	3.83%
		2051	89	4.34%

Bounce Rate for Plastic Manufacturing page

# --				
www.motiontouch.com				
Top Landing Pages				
May 1, 2008 - August 31, 2008				
Page	**Ad Content**	**Entrances**	**Bounces**	**Bounce Rate**
Plastic Manufacturing	Plastic Manufacturer	6414	3953	61.63%

Bounce Rate for Questionnaire page

# --				
www.motiontouch.com				
Top Landing Pages				
May 1, 2009 - August 31, 2009				
# --				
Page	Ad Content	Entrances	Bounces	Bounce Rate
Questionnaire	Plastic Manufacturer	1661	1010	60.81%