Aalto University
School of Business

Comparison of Owned, Earned and Paid Website Visitors: A Case Study

Master's Thesis
Jesse Ketonen
28 April 2017
Information and Service
Economy

Approved in the Department of Information and Service Economy

__ / __ / 20__ and awarded the grade

____

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Jesse Ketonen

Title of thesis Comparison of Owned, Earned and Paid Website Visitors: A Case Study

Degree Master of Science in Economics and Business Administration

Degree programme Information and Service Economy

Thesis advisor(s) Matti Rossi, Noora Pinjamaa

Year of approval 2017 **Number of pages** 103 **Language** English

Abstract

This thesis explores website traffic and visitors by analysing website customer behaviour. The thesis expands the current research on web analytics to consider the rising categorization of media into owned, earned and paid media types. The research is first of its kind to further explore if there is significant difference between owned, earned and paid website visitors measured by web metrics. In addition to academic contributions, it is desired that the research helps marketers and publishers to invest their resources between generating each type of traffic in order to reach their individual goals and maximize the return-on-investment.

In this paper, a framework for measuring owned, earned and paid website visitors is created. The research framework is tested in a case study where owned, earned and paid traffic is driven from Facebook to a fashion magazine's online articles. Data on visitor-level website behavior of 2739 visitors is collected from the case website using Piwik analytics. The data was analyzed using two quantitative methods: chi-square test of homogeneity and one-way analysis of variance. These methods were used in order to determine whether statistically significant differences in website between owned, earned and paid visitor groups exists. Further, the case study demonstrates how to use the framework and appropriate techniques to effectively collect, extract, and analyze website visitor's web behavior and the differences between owned, earned and paid website visitors.

The empirical research reveals that significant differences between different types of website visitors exists. The chi-square test of homogeneity indicated a statistical significant difference of binomial proportions of 'new / return user rate', 'bounce-rate' and 'mobile / desktop rate' variables. One-way ANOVA indicated a statistical significant difference between the means of owned, earned and paid visitors of "visit count" and "actions", but also a non-significant difference of "visit duration". Thus also the usability of the research framework is confirmed.

This thesis expands the research on clickstream data into social networking and earned media in media and journalism, and so contributes to the existing research on web analytics. This thesis also contributes to the existing literature on owned, earned and paid media and web analytics by adding owned and earned social media exposure to clickstream research and comparing them to paid social media exposure it in assessing user's behavioral response in a cross-site context. Thus the thesis also combines social marketing with web analytics and expands the use 'owned', 'paid' and 'earned' jointly in a digital environment. This study is also first one to apply 'heart rate monitoring' measurement, redefined visit duration and bounce-rate metrics. The thesis provides useful technical and methodological information about website visitor tracking and web metrics for both academics and businesses seeking benefits from web analytics and online channels.

Keywords Web analytics, clickstream, visitor statistics, owned media, earned media, paid media, social media

Tekijä Jesse Ketonen

Otsikko Omien, ansaittujen ja maksettujen verkkosivukävijöiden vertailu: tapaustutkimus

Tutkinto Master of Science in Economics and Business Administration

Tutkinto-ohjelma Tieto- ja palvelutalous

Ohjaajat Matti Rossi, Noora Pinjamaa

Hyväksymisvuosi 2017 **Sivut** 103 **Kieli** Englanti

Abstrakti

Tutkielma tarkastelee verkkosivuliikennettä ja verkkosivuvierailijoita analysoimalla verkkokäyttäytymistä. Tutkielma laajentaa nykyistä web-analytiikan tutkimusta käsittelemään omaa, ansaittua ja maksettua mediaa. Tutkielma on ainutlaatuinen tieteenalallaan tarkastellessaan omien, ansaittujen ja maksettujen verkkosivuvierailijoiden välisiä eroja web-metriikoilla mitattuna. Akateemisen kontribuution lisäksi tutkimuksen toivotaan auttavan markkinoijia ja julkaisijoita allokoimaan resurssejaan kullekin yksilöllisten tavoitteiden näkökulmasta paremmin eri tyyppisten liikennevirtojen generoimiseen, saavuttaen näin parhaan mahdollisen tuoton investoinneilleen.

Tutkielmassa esitellään uusi viitekehys omien, ansaittujen ja maksettujen verkkosivuvierailijoiden mittaamiseen. Tutkimuskehys testataan tapaustutkimuksessa, jossa omaa, ansaittua ja maksettua liikennettä ohjataan Facebookista valitun muotilehden verkkoartikkeleihin. Tapaustutkimuksen aineisto koostuu Piwik-analytiikkajärjestelmällä kerätyn 2789 verkkosivuvierailijan vierailija-tason verkkokäyttäytymiseen. Aineisto analysoidaan käyttäen kahta kvantitatiivista menetelmää: khii-neliön testiä ja yhdensuuntaista varianssianalyysia. Valituilla keinoilla selvitetään onko oman, ansaitun ja maksetun vierailijaluokkien välillä tilastollisesti merkitseviä eroja. Tapaustutkimus osoittaa, kuinka viitekehystä ja käyttötarkoitukseen soveltuvia tekniikoita käytetään omien, ansaittujen ja maksettujen verkkosivuvierailijoiden verkkokäyttäytymistietojen keräämiseen, analysointiin ja vertailuun. Tutkielman empiirinen tutkimus todistaa, että tilastollisesti merkitseviä eroja omien, ansaittujen ja maksettujen verkkosivuvierailijoiden välillä on olemassa. Khii-neliön testi osoittaa, että ero on olemassa 'uudet / palaavat käyttäjät', 'poistumissuhde" ja 'mobiili / työpöytä –käyttäjät' muuttujien tapauksessa. Yksisuuntainen varianssianalyysi osoittaa, että omien, ansaittujen ja maksettujen verkkosivuvierailijoiden keskiarvojen välillä on tilastollisesti merkitsevä ero 'vierailumäärän' ja 'toimintojen' tapauksissa, mutta myös ei-merkitsevä ero 'vierailun keston' tapauksessa. Näin ollen myös tutkimuskehyksen käytettävyys on todennettu tapaustutkimuksen kautta.

Tutkielma avartaa nykyistä klikkaustietoihin liittyvää tutkimusta sosiaalisiin verkostoihin sekä ansaittuun mediaan. Tämä tutkielma kontribuoi myös olemassa olevaan omaa, ansaittua ja maksettua mediaa sekä web-analytiikkaa käsittelevään kirjallisuuteen 1) tuomalla oman ja ansaitun sosiaalisen median näkyvyyden osaksi klikkaustietoihin liittyvää tutkimusta ja 2) vertaamalla näitä maksettuun sosiaalisen median näkyvyyteen verkkokäyttäjän käyttäytymissä ilmenevien reaktioiden arvioinnissa toisistaan erillisten verkkosivujen tapauksessa. Näin ollen tutkielma yhdistää sosiaalisen markkinoinnin web-analytiikkaan ja laajentaa oman, maksetun ja ansaitun median rinnakkaista seurantaa digitaalisessa ympäristössä. Tutkimus soveltaa 'sykemonitorointia' mittauksessa sekä uudelleen määriteltyjä vierailun kesto sekä poistumisaste –mittareita. Tutkielma tarjoaa hyödyllistä teknistä ja metodologista tietoa verkkosivuvierailijoiden seurannasta sekä web-metriikoista digitaalisten kanavien hyödyntämistä harkitseville akateemisille tutkijoille sekä yrityksille.

Avainsanat: Web-analytiikka, klikkaustieto, verkkosivukävijätilastot, oma media, ansaittu media, maksettu media, sosiaalinen media

Acknowledgements

I want to thank the thesis instructors Prof. Matti Rossi and Noora Pinjamaa for their extensive and lasting support thorough the research process. I also want to thank Prof. Merja Halme for her contribution and help with the selection and assessment of quantitative methods used in the study. I'm also obliged to the thesis opponents Aimo Järvinen and Ville Lehtonen from Aalto University. I want to give special recognition to Aller Media Oy and ELLE magazine that provided the opportunity and resources to profoundly explore the subject. I want especially thank to Aller's Insight & Analytics Director Kristina Hännikäinen, ELLE Editor-in-Chief Taru Marjamaa, Aller's former B2B Director Janne Pullinen and other people across the Aller IT-, digital- and Elle magazine departments who made the empirical research possible.

Table of Contents

List of figures

List of tables

1 Introduction

For the past several years, the media and entertainment industry has been in turmoil. Consumers are at an accelerated rate replacing their traditional media consumption with digital experiences (Berman, Battino, & Feldman, 2011). Online readership is hoped to save the struggling publishers from declining revenues (Vu, 2014). Digital experiences in digital channels grow their share of people's time. Companies aim to capitalize this by maximizing the attention towards themselves. Therefore, the role of digital marketing in a firm's marketing strategy is becoming central. This can be seen from increasing investments in digital marketing activities. An additional reason for the shift of budgets to digital marketing are the cost-effectiveness and easier measureability of its results (Järvinen & Karjaluoto, 2015).

Due to increased connectivity, growing interaction between customers and companies in digital channels, and development of technologies, companies need to serve and attract clients through digital channels irrespective of industry. According to Berman, Battino & Feldman (2011), especially established media and entertainment companies face challenges to monetize their digital channels successfully while they possess great opportunities that only a few relatively new industry entrants have successfully grasped. According to Berman, Battino, & Feldman (2011), media and entertainment companies must embrace new distribution platforms and strategically optimize between them capitalize on the opportunities. In order to successfully do this, the companies need to utilize analytics, have a consumer-centric approach, and embrace multi platform delivery (Berman et al., 2011). Unlike traditional media, Internet-driven digital media allows easy, fast and unobtrusive collection of information on individual activities on a detailed level (R. E. Bucklin & Sismeiro, 2009). This information can be harnessed for making better business decisions. For example, the retail giant Target has used its collected data to identify buying patterns, lift customer satisfaction, select promotions, predict trends, create strategies, and increase revenue (Corrigan, Craciun, & Powell, 2014).

For these reasons, companies need to utilize Web analytics (WA), defined as "the measurement, collection, analysis and reporting of Internet data for the purposes of understanding and optimizing Web usage" (Web Analytics Association, 2008, p. 3). Web analytics have become a popular subject ever since the rise of the Internet and company websites, as it helps companies to define the purpose and objectives of their web presence (Riihimäki, 2014). Nakatani and Chuang (2011) emphasize that these objectives are not limited to optimizing web sites but the ultimate objective is to drive the success of an organization's overall targets.

The digitalization of media has emphasized the importance of owned channels such as a company's website and mobile apps. Today, websites serve multiple purposes, which makes it also hard to evaluate their performance (Welling & White, 2006). Earlier research has investigated the value of websites from different perspectives. Welling & White (2006) propose that online sales is a key goal alongside customer support for a retail company. They continue saying brand awareness and education to be the purpose across different industries, whereas business-to-business companies commonly use it for recruiting. Benefits are gained by offering services and information to various stakeholders (Welling & White, 2006). Many firms use websites as the most important contact point between them and their potential or current clients (Riihimäki, 2014). However, the existence of online services and websites does not make sense without continuous and relevant high-quality traffic from external sources. Web traffic is a valid measure for performance for companies (Luo & Zhang 2013; Vaughan 2008; Vaughan & Yang 2013). Especially for online businesses, whose revenue relies purely on this traffic –with no traffic, there would be no revenue streams (Ghandour, Benwell, & Deans, 2010).

One way to study web traffic's relevance and quality is to look at a visitor's website behavior. Website behavior's importance has been widely acknowledged in previous studies across different academic disciplines. In the literature review, past research on website behavior from marketing, communications, and information and service economy perspectives is covered.

Based on the previous studies, it seems necessary to study web traffic more, and understand the different aspects of it. Given that we know that web traffic generally creates value, the existing research does not fully consider different traffic types that could have different effects on chosen performance metrics. Web traffic may be generated by various parties, channels, contents or mechanisms (Vaughan & Yang, 2013). Many of these have not yet been studied in detail. This research aims to narrow the gap of measuring and understanding different types of traffic on a visitor level. Taking into consideration the current state of the Internet, there are many different types of traffic, which can lead to very different results from a business perspective. Ghandour et al. (2010) state that online businesses can end up making no sales even with a high amount of traffic. This is because traffic may originate from different sources and there can be many different reasons why it has been generated. For example, a person might have seen a video on a social media platform, which has triggered him to click a link that leads to a company's website. The trigger behind the click might have been the content's headline or the video shown beside it or because it was shared by one of his best friends. The video might have come to him as a targeted ad, offering something new and interesting that took him to the website to explore more of its contents. Or maybe the ad was only generating a click but could not engage the visitor to further explore the website. This study proposes that website visitors who came through different kinds of media channels are different and behave differently on a website. Knowing these differences would bring marketing and media professionals an advantage in driving the most relevant and valuable traffic for their business goals instead using resources to generate traffic that does not provide return on investment. Companies and organizations should focus on bringing more relevant traffic to their own channels in order to maximize the positive business impact; in the context of our study – maximize the attention and reach the desired website behavior of their website visitors. In a case study by Järvinen & Karjaluoto (2015) one company saw the ability to measure website visits and traffic generated by different marketing actions as the greatest benefit of web analytics. Therefore it is necessary to more profoundly understand the existence of different types of traffic on the Internet. As the traffic is generated by different marketing communication activities, this raises the need to study how firms can measure the effects of different types of marketing communication and how they compare with each other (Trusov, Bucklin, & Pauwels, 2009).

The concepts of owned, earned and paid media

In today's business marketers, publishers and media professionals classify media into three types: owned, earned, and paid media. According to Stephen and Galak (2012), a typical situation is that a company has a combination of owned, earned and paid (OEP) media activity. This is true because most companies tend to advertise, have their own websites and engage in social media activities, which already covers all these types of media.

Businesses may intentionally increase the amount of any of these types of media reach for their company or brand using different strategies and tactics. For example, a company may focus on social networking through earned media. In this case, the company may target social influencers and aim to become a topic for networked social groups (Berman et al., 2011). Also, a company may aim to build large online communities to increase the baseline of their owned media exposure. The company may further utilize these strategies to drive traffic to their own channels and conduct campaigns utilizing their owned communities and social network to maximize the results. DiStaso & Brown (2015) found that many of the Fortune's "World's Most Admired Companies" and "Best Companies to Work For" listed companies utilize earned and owned media in their communications and that there is an increasing trend in the return on efforts regarding this activity. This study aims to further investigate the difference in the benefits that traffic generated by each type of media brings to marketers and publishers. With this kind of information, businesses can better plan their investments between different types of media according to their individual goals. Owned, earned and paid media is however rarely referred to in WA research. Article of Chaffey & Patron (2012) on WA driven digital marketing performance development is an exception, combining WA with the concept of owned, earned and paid. This thesis expands the analysis of clickstream data in social networking and owned, earned and paid media in media and journalism, and so contributes to the existing research on web analytics.

Today, businesses are seeing increasing amount of traffic coming from social media. The digital media analytics company Parse.ly has reported that social media is the lead traffic referrer for online publishers (Ingram, 2015; VanNest 2016). Especially Facebook has taken its place as a major traffic source (Ingram, 2015; VanNest 2016). In 2016 only, Facebook drove 42% of all traffic to Parse.ly's network of online publishers according to VanNest (2016). In addition to having a major role for publishers, owned, earned and paid media coexist in Facebook, which makes it excellent source material for study. This thesis introduces a way to divide owned, earned and paid traffic in a social media context and further explores the meaningful differences between visitors generated by each traffic type towards common web metrics. The case study focuses on a fashion media brand driving visitors from Facebook to its web articles during a measurement period.

The remainder of the article is organized as follows: First, the most important terms and concepts in web analytics, as well as owned, earned and paid media, traffic and visitors are defined. Next, the research problem and research question are introduced. After the introduction, the existing literature on web analytics and owned, earned and paid media is reviewed. Concluding the methodology section, the research framework and methodology are introduced. Next, our case study approach and data collection and analysis methods are explained in detail. This is followed by the actual data analysis and findings. In the final section, the findings are summarized, managerial implications and theoretical contributions are presented. Finally, the limitations of the study and future research opportunities are discussed.

1.1 Definitions

To have a clear picture of the research problem, various metrics should be understood. In August 2007, the Digital Analytics Association (DAA, former Web Analytics Association) has defined the most important web analytics metrics (Burby & Brown, 2007). DAA has also introduced a Definition Framework, which puts Web analytics metrics into three types: counts, ratios and KPIs. Count is the most basic unit of measurement and it is a single number (Burby & Brown, 2007). An example of count is 'time spent on page'. According to Burby and Brown (2007), ratio is usually a count divided by a count, but it may as well use either a count or a ratio in the numeral denominator. An example of a ratio is 'bounce rate'. Burby and Brown (2007) state that KPI (Key Performance Indicator) is used to distinguish the most important metrics to follow related to the business strategy. They amplify that KPI typically differs between site and process types. A KPI can be either a count or ratio (Burby & Brown, 2007).

DAA also adds a fourth type of definition for "terms that describe concepts instead of numbers." That is dimension, defined as "a general source of data that can be used to define various types of segments or counts and represents a fundamental dimension of visitor behavior or site dynamics." Traffic referral sources and website events are some examples. "They can be interpreted the same as counts above, but typically they must be further qualified or segmented to be of actual interest. Therefore these define a more general class of metrics and represent a dimension of data that can be associated with each individual visitor" (Burby & Brown, 2007). According to Burby and Brown (2007), "metrics are measured across the dimensions."

Next, we go through some definitions of common terms and concepts that are relevant for the topic and used in this thesis. We also discuss the means of each metric used in the case study.

1.1.1 Web analytics

Web analytics (WA) refers to a "tool that collects clickstream data regarding the source of website traffic (e.g. e-mail, search engines, display ads, social links), navigation paths, and the behavior of visitors during their website visits and that presents the data in a meaningful format" (Järvinen & Karjaluoto, 2015). Web analytics data are used to understand customer behavior online and to measure customer responses to digital marketing in online environment and to optimize elements and actions of digital marketing that have been identified to drive customer behavior that is beneficial for the business (Nakatani & Chuang, 2011).

1.1.2 Web Metrics

Web analytics enable website owners to monitor how website visitors behave online by providing web metrics (Tandoc, 2015). According to (Krall, 2009), web metrics is "any quantitative measure of passive viewing or consumption of content by internet users" According to (Burby & Brown, 2007), "a metric can apply to three different universes." They are, aggregate, segmented and individual. According to Burby and Brown (2007), "Aggregate is the total site traffic for a defined period of time". Segmented on is a "subset of the site traffic for a defined period of time, filtered in some way to gain greater analytical insight". Examples of filters are campaigns, referrers or visitor types such as new vs. returning visitors. Owned, earned and paid traffic are subsets of site traffic filtered by visitor type. The third universe is individual, defined as the "activity of a single web visitor for a defined period of time" (Burby & Brown, 2007).

1.1.3 Website Traffic

According to Marketingterms.com (2017), website traffic is "the amount of visitors and visits a Web site receives." In this thesis, we use web traffic as a synonym for website traffic.

1.1.4 Website Visitors

Website visitors represent "the number of unique users (identified by cookies) accessing a specific website" (Yang, Pan, & Song, 2013). Cookies are tiny text files saved in user's computer and they are sent to a web browser by a web server in order to identify the user. During the reporting period, each individual visitor is usually counted only once (Burby & Brown, 2007). However, a visitor does not always equal a unique person. A person may use different browsers or devices to access a website but cookies do not follow website visitors across them. An individual session initiated by a website visitor is called a "visit" (Yang et al., 2013). In this thesis we commonly refer to website visitors simply as "visitors". In practice, web traffic generates website visitors and their amount is equal on a single website.

1.1.5 Website Visits / Sessions

Website visits are "the number of individual sessions initiated by all visitors on a website" (Yang et al., 2013). One access by a unique IP address equals one visit. It is worth noting that most web analytics tools require a 30-min time interval between each access before it can be counted as an additional visit (Booth & Jansen, 2008). Website visits are also known as "sessions".

1.1.6 Bounce Visit

A bounce visit, also known as *Single Page View Visit,* is a visit that is formed on one page-view (Burby & Brown, 2007). A bounce visit is usually interpreted as a non-engaged visit, which might not be always true. If all the relevant content is placed on a single page, a bounced visitor may have actually spent a lot of time and consuming the content.

1.1.7 Bounce Rate

According to Plaza, Beatriz (2009), "Bounce rate is the percentage of visitors who enter a site (or a page) and then leave immediately without visiting any other pages." A similar definition is used by (Ghandour et al., 2010). According to Burby and Brown (2007), bounce rate is "single page view visits divided by entry pages". According to Plaza, Beatriz (2009), bounce rate could also be defined through visit duration regardless the number of page views. For example, visitors who spend 10 seconds or less on the site can be viewed as bounced visitors and therefore the percentage of visitors who spend under 10 seconds on site compared to total visitors counts as bounce-rate (Plaza, Beatriz, 2009).

According to Pakkala et al., (2012) bounce-rates indicate website relevancy to a visitor. Low bounce rate tells that a website is relevant to the visitor. Plaza, Beatriz (2009) state that high bounce rate visits are problematic and they are especially common amongst referring site visits.

1.1.8 Visit Duration

"Visit Duration is the length of time in a visit" (Burby & Brown, 2007). Burby and Brown (2007) clarify the measurement further: "calculation is typically the timestamp of the last activity in the session minus the timestamp of the first activity of the session". Visit duration is often mixed with "time spent on a page". The purpose of use is similar for both metrics, but there are technical differences that web analytics users should be aware of. As a default, most web analytics tools' calculation of visit duration record bounce visits as 0-second visits, where as time spent on a page excludes bounce visits. Therefore, due to bounce visits, many pages report lower visit duration than time spent on a page on average. This also means that as a default, visit duration is a skewed metric and does not report the true time of a visitor's visit. Further in this thesis, we explore this problem and try to find solutions to it.

Visit duration is arguably one of the most used metric in Web Analytics, and its significance has been studied in prior literature. Panagiotelis et al., (2013) found a relation to sales, whereas quite a few studies have discovered a connection to purchase incidence (Moe & Fader, 2004; Montgomery, Li, Srinivasan, & Liechty, 2004; Panagiotelis et al., 2013). Danaher & Smith (2011) suggest that visit duration is linked to sales volume. For certain types of products, this might be especially true. Visit duration have been found to bring value through maximizing expected sales for books, travel service and digital media (Panagiotelis et al., 2013). Visit duration has been used as proxy indicator of web “stickiness”. This is due to its noted positive effect in visitor-to-buyer conversion rates and online loyalty (Xun, 2015). Long visit duration is not always necessarily an indicator of success: long visits may be caused by bad and difficult website design and unease use of a website leading to decreased sales (R. E. Bucklin & Sismeiro, 2009). Long visit durations without positive outcome are also a non-productive use of the web servers capacity (Wilson, 2010).

Pakkala et al. (2012) distinguishe two different groups of visitors: “experimenters” and “real visitors”. They define experimenters as visitors “who try the website and visit for less than 10 seconds and “real visitors” as visitors “who typically spend 1–10 minutes on the website”. An online media research company Compete has used visit duration to rank web sites based on the total time spent on a website compared to the total time Americans spend online (Zheng, Chyi, & Kaufhold, 2012). A similar metric, time spent on a page, may be used in assessing the time visitors spend consuming web site content and using its services. Pakkala et al., (2012) used the average time spent on a page to assess website importance as a source for different visitor groups. In our study, we focus on time spent on the whole website.

For businesses, it is essential to understand what affects visit duration. Danaher, Mullarkey and Essegaier (2006) concluded that variation in visit duration is driven mostly by the situation and less by qualities of the individual or website text, graphics advertising content and functionality. However, according to R. E. Bucklin & Sismeiro (2009), prior research suggests that “behavior follows regular patterns and that design issues might play an important role”. Therefore, we do not have clear viewpoint on whether visit duration will be significantly varied across visitor types.

1.1.9 Corrected Visit Duration

Corrected Visit Duration reports the real visit duration, where total time spent on a page is calculated using a different method that also counts the visit duration from single page view visits. We explore corrected visitor duration more precisely later in the thesis.

1.1.10 Adjusted Bounce Rate

This study proposes corrected visit duration as the determining factor of bounce rate. Therefore, adjusted bounce rate is defined as percentage of visitors who spend under 10 seconds on site, calculated from corrected visit duration.

Our definition of bounced visitors is that the visitors that truly spend under 10 seconds on a website will count “experimenters” as bounced visitors. Similarly as with bounce-rate, adjusted bounce-rate is proposed to indicate website relevancy to a visitor.

1.1.11 New Visitor

A unique visitor that records its “first-ever visit to a site during a reporting period” (Burby & Brown, 2007).

1.1.12 Return Visitor

A unique visitor that records a visit and has recorded a former visit to a site during a reporting period (Burby & Brown, 2007).

The rate of return visits has been identified a key web metrics that is connected with website effectiveness (Riihimäki, 2014). Both new visits and return visits should be followed because they correlate strongly with the total number of conversions (Riihimäki, 2014). Based on a regression analysis, Riihimäki (2014) find a that the “amount of new visits can be used as a predicator towards the conversion rate”. An insight from their regression and correlation analysis is there is a negative relationship between the rate of return and the conversion rate, meaning that new visitors are more likely to convert making them a more valuable than return visitors. Contradictory, Plaza, Beatriz (2009) claims that return visits foster longer visit duration, making it indirectly have a positive effect on conversion (Xun, 2015). Riihimäki (2014) found similarly the return visitors to spend a longer time on the site and have higher number of page views, but state that this does not turn into a higher conversion rate.

1.1.13 Mobile vs. Desktop users

The device used naturally affects how the website is experienced by the user. This study takes into account the differences in mobile vs. desktop traffic between the categories and investigates if device differences explain variations in other metrics.

1.1.14 Visitor Actions

According to Piwik (2017), Visitor Actions consist of page views, internal site searches, file downloads and clicks on external websites.

Montgomery et al., (2004) found that purchase conversion can be predicted from visitor page views on a site using a modeling approach. The more pages a visitor viewed, the more likely it led to a purchase.

1.1.15 Conversions

Conversion is a target action completed by the visitor (Burby & Brown, 2007). It can be very different depending on the objectives of the website and the business.

Conversions are commonly tracked in e-commerce and other transactional websites, where the end goal is to get the visitor to make a purchase (Croll & Power, 2009). In this case, the purchase is one kind of a conversion. In the way to purchase, visitor usually completes a set of different activities, which may also be seen as conversions, sometimes defined as “micro-conversions”. This path is from the beginning of visitor landing to a website to making a purchase is referred as conversion funnel (Croll & Power 2009). There are websites that serve different purposes so many different conversions are being used. For example, it is more difficult to determine a conversion on an information-oriented website. Downloads of a white paper or sign-ups on a newsletter are often monitored as conversions amongst them (Kaushik, 2010). In order to evaluate overall objectives, they should be broken into a more specific goals that visitors are wanted to complete on a website. A website can have multiple conversion goals. When at least one conversion goal is completed by a visitor, it is seen as converted visitor (Riihimäki, 2014). Conversions are often reported as rates of converted visitors. This conversion rate is the number of converted visitors compared to total unique visitors. According to Riihimäki (2014), “the total number of conversions can be used as an indicator for the overall performance of a site while the conversion rate tells about the quality of a single visit".

Two categories of conversion goals are introduced by Tonkin et al. (2010): "transaction goals" and "engagement goals". Transactional goals have direct monetary value whereas engagement goals "relate to a threshold or interaction without direct monetary value". An example of direct monetary value is purchase of a product or becoming a lead by registering an account. Indirect value can be spending a desired amount of time on web page consuming certain content. In our case a transactional conversion goal is set: sign-up in a competition through a form on the website.

1.1.16 Visit Count

Visit Count is the amount visits of a website visitor added together over time. Visit count is often referred as number of visits (Burby & Brown, 2007).

Visit count is commonly understood and referred to as visitor loyalty. According to Pakkala et al., (2012), visitor loyalty was used to determine whether visitors found website content and layout satisfactory enough so they would be willing to return to the website. Internet users have been found to have considerable website loyalty and high switching costs (R. E. Bucklin & Sismeiro, 2009). Moe and Fader (2004) found that changes in individual-level visit frequencies can indicate which visitors are more likely to buy online. On a retail site, they found that frequent visits translated into a higher likelihood of buying. The number of visits per visitor is widely accepted as being one of the key metrics to measure site performance whether the purpose of the site is to sell products or attract and retain regular readers (Moe & Fader, 2004). However, Sismeiro and Bucklin (2004) found that the number of site visits do not predict purchases. This may be because hedonic browsing and knowledge building may be a very different for different kind of product categories, such as cars and books. For example in the case of cars, many users might visit the site multiple times without making a purchase. We analyze visitor loyalty from the perspective of a publishing and content website that aims to maximize reader retention and the number of individual visits.

1.1.17 Owned, Earned and Paid Media

Xie and Lee (2015) suggest that owned media refers to “media activity generated by a company or its agents” in channels that they have control over. They suggest that earned media refers to “media activity that is not generated directly by the company”, but instead by other stakeholders such as consumers or journalists. According to Xie and Lee (2015) an alternative, widely used term for earned media is Word-Of-Mouth (WOM). More precisely, earned media in online channels have been named electronic Word-Of-Mouth (eWOM). A company’s marketing can support in generating earned media, but they do not generate the earned media activity directly (Xie & Lee, 2015). Paid media is often used as a synonym for advertising. According to Xie and Lee (2015), it refers to “media activity that a company or its agents generates” and pays for the distribution of the content. Today, these three types of media are widely recognized in the media industry.

1.1.18 Owned, Earned and Paid Traffic and Website Visitors

Web traffic can be divided to owned, earned and paid according to their source and redirect link distribution method. In this thesis it is proposed that owned, earned and paid traffic are defined after the media type that generates the traffic. For example, if the user comes via a link attached to owned media, we can say it belongs to owned traffic. If the user comes via a link that is not distributed by the brand or its agents, it is earned traffic and if the user comes through an advertised link, it would belong to paid traffic. Further, I propose that owned, earned and paid (web) traffic converts to owned, earned and paid (website) visitors when landing on a website by recording a visit on the website

The following table illustrates our conversion logic:

Owned media exposure	=>	Owned traffic	=>	Owned visitors
Earned media exposure	=>	Earned traffic	=>	Earned visitors
Paid media exposure	=>	Paid traffic	=>	Paid visitors

Table 1: OEP visitor conversion logic

To conclude definitions section, we present the following table to sum what implications can be made from each of the metric used in the case study:

Metric	Indication
Visit Duration	• Related to sales (Panagiotelis et al., 2013). • Connected to purchase incidence (Moe & Fader, 2004; Montgomery et al., 2004; Panagiotelis et al., 2013). • Linked to sales volume (P. J. Danaher & Smith, 2011). • Positive effect on visitor-to-buyer conversion rates and online loyalty (Xun, 2015). • Not always an indicator of success: may lead to decreased sales (Wilson, 2010).
Visitor Actions	• Page views can predict purchases. The more pages a visitor viewed, the more likely it leads to a purchase (Montgomery et al., 2004).
Bounce Rate	• Indicates website relevancy to a visitor. (Pakkala et al., 2012)
Adjusted Bounce Rate	• Indicates website relevancy to a visitor. (Pakkala et al., 2012) but fits better to websites with a lot of single page content and visits.
Conversion	• Conversion is a sign of quality of a single visit (Riihimäki, 2015).
New / Return visitors	• Return visits foster visit duration (Plaza, Beatriz, 2009). • Connected with website effectiveness (Riihimäki, 2014). • “Both new visits and return visits have strong correlation with the total number of conversions” (Riihimäki, 2014). • “Amount of new visits can be used as a predicator towards the conversion rate” (Riihimäki. 2014). • “Rate of return visits has a negative relationship with the conversion rate”, meaning that new visitors are more likely to convert making them more valuable than return visitors (Riihimäki, 2014). • Return visitors visit duration is higher and they have higher number of page views, but this may not translate into a higher conversion rates (Riihimäki, 2014).
Mobile / Desktop users	• Mobile users are hard to monetize for publishers (eMarketer, 2016). • Mobile users show better user engagement in video advertising (Heine, C., 2014). • Mobile users have undivided attention (Heine, C., 2014) • Mobile offers better qualifications for location based personalization (Heine, C., 2014) • Desktop users have better sales conversion is e-commerce (Heine, C., 2014)

	• Visual impact is greater for desktop users (Heine, C., 2014)
Visit Count	• More frequent retail site visitors have a greater propensity to buy. (Moe & Fader, 2004) • Visit frequency's evolution on individual level can explain which customers are more likely to buy online (Moe & Fader, 2004). • Visitor loyalty was used to determine whether visitors found website content and layout satisfactory enough so they would be willing to return to the website (Pakkala et al., 2012). • Change in individual-level visit frequency can indicate which visitors are more likely to buy online (Moe and Fader, 2004). On a retail site, they found visitors that visit that visit more often are more likely to buy. • Widely accepted to be one of the key metrics to measure site performance whether the purpose of the site is to sell products or attract and retain regular readers. (Moe & Fader, 2004). • Not predictive of purchase (Sismeiro and Bucklin, 2004)

Table 2: The chosen metrics and what they indicate

1.2 Research problem

Companies from various industries are seeing website traffic as increasingly important. Firms can design and execute online and offline marketing campaigns that aim to increase traffic to their digital channels (Järvinen & Karjaluoto, 2015). The campaign effectiveness is usually measured by its impact on website customer behavior (Järvinen & Karjaluoto, 2015). Website traffic is important especially for media companies, whose online businesses are tied to the amount and quality of their website traffic. This traffic is further monetized by selling advertisement solutions to advertisers or by converting the visitors into paying customers. Järvinen & Karjaluoto (2015) find that firms' ability utilize WA to improve marketing performance remains limited. They claim that the majority of marketers think that measuring digital marketing performance is important. However, they say that less than one third of marketers think they are doing it well.

As web traffic plays a key role in value generation for businesses in the digital era, firms should know about what kind of traffic to invest in and how they differ. The businesses should also understand better what makes the traffic relevant and of high quality. Plaza, Beatriz (2009) ask whether a strategic plan to increase website traffic should focus on a certain source of traffic. In order to answer this question and explain it, it is necessary to understand how to measure traffic source effectiveness. Analysis of different types of traffic, such as owned, earned and paid traffic can be similarly compared to the analysis of different sources of traffic. Research focusing on individual types of media and traffic has previously been done but the differences between owned, earned and paid traffic concurrently have been studied very little so far.

This study aims to find out if, and for what reasons firms should invest in generating owned and earned traffic in addition to paid traffic by investigating the website behavior of the visitors of an online fashion media site by analyzing its web metrics. Web metrics have been found to correlate with financial performance of an online business across different industries (Ghandour, Benwell & Deans 2010). Also, a categorization to owned, earned and paid media is being recognized. For these reasons, we want further to learn if there is significant difference between owned, earned and paid website traffic and visitors measured by web metrics. Online marketers need to decide how they invest their resources between generating each type of traffic in order to reach their goals and maximize the return-on-investment. eMarketer (2010) suggests that "some of today's greatest success stories in branding blend ingredients from the three kinds of marketing media: paid, owned and earned" but does not go further into the subject.

As mentioned earlier, traffic classification to owned, earned and paid has been studied very little. Terms 'owned', 'paid' and 'earned' are were used jointly in a digital environment by Srinivasan, Rutz, & Pauwels (2016). They explored how consumer activity metrics of owned, paid and earned influence differ in terms of brand performance and sales with fast moving consumer goods. They found that the difference is significant, and the effect on sales is positive. Their empirical study took place in a cross-channel context, whereas as the focus in this thesis is to study owned, paid and earned media that originates from social media. Instead of sales, this thesis focus solely on web metrics and website visitors. By exploring the subject from a new viewpoint, this study aims to complement the current research. It is a relevant subject as companies may affect the amount of owned, earned and paid traffic they are getting by investing in building online communities and directing people to their websites from them for example, or by creating shareable content that attracts people to spread links that direct to their website. Knowing the differences between different traffic types should direct a company's decision making. For example, if companies see the best returns from using paid traffic, it would be relevant for them to focus on establishing paid traffic generation strategies and invest in paid traffic over other types of traffic. This study will contribute to research on website behavior and metrics by exploring owned, paid and earned media in the context of the media industry.

Finally, it is also important to understand how the different types of traffic may be interconnected and how to actually define them before taking action to shape a company's marketing efforts. Companies have to make decisions on how to split their limited resources. There is a vast amount of options, so businesses need to have a good understanding of which drivers are the most critical for realizing their business goals. In online environments, investing in the right kind of media exposure and traffic is decisive. Therefore, this thesis investigates if investing based on the traffic type categorized as owned, earned and paid makes sense. This kind of categorization is already common practice amongst marketing and communications professionals, but the true value of it remains unclear. An example of the how companies invest in different types of media is given by the marketing communications manager of Lincoln Electric, Craig Coffey in based on their previous marketing campaign budget split in a CMO.com interview: "from a spend standpoint, it was 80% paid, 10% earned, and 10% owned, but from a conversion standpoint, it was almost the inverse: 81% earned, 12% owned, and 8% paid" (Schwarz, M., 2016). According to Statista (2017), marketing budgets in UK in 2015 were split 39% paid, 26% earned and 35% owned.

1.3 Research question

One of the key challenges for any organization's digital business is to learn how to identify the most impactful website visitors and sources to its website. The main research question for this thesis draws from the insight that website behavior measurement is critical for online businesses (R. E. Bucklin & Sismeiro, 2009) and web traffic analysis benefits firms performance (Luo & Zhang, 2013, Moe & Fader, 2004, Vaughan, 2008, Vaughan & Yang 2013). Website behavior analysis and web traffic analysis are equal to website visitor analysis. Today businesses typically have a combination of owned, earned and paid media activity (Stephen & Galak, 2012), yet their connection with website visitors is not well known. Therefore, the main research question is:

How does website behavior differ between owned, earned and paid website visitors?

The main research question is examined through a case study, where each type of traffic is being driven from Facebook to an online media website. The website behavior data is collected using the web analytics tool Piwik, and analyzed using Chi-Square analysis and One-Way ANOVA.

In order to find out answers to the main research question, there is a need to identify each type of website visitors and build a measurement model for owned, earned and paid traffic. Therefore, the supporting research questions is:

How to measure owned, earned and paid traffic and visitors?

A research framework is proposed for the measurement of traffic and visitors in third chapter. In the case study, we propose an application of the measurement model and test it in the social media context.

In the next chapter we will conduct a literature review about web analytics and then continue our review of owned, earned and paid media.

2 Literature Review

2.1 Web Analytics

Marketers have realized that interactions and performance of website visitors needs to be tracked as clients are interacting with companies through digital channels more and more (Chaffey & Patron, 2012). Traditionally, this has been measured by conducting market research and interviewing customers on their website experiences. According to Weischedel and Huizingh (2006), this method is expensive, time-consuming and requires a long time interval. However, it can answer the questions “how” and “why” (Weischedel & Huizingh, 2006). Web technology on the other hand can collect massive amounts of data on visitor traffic and activities on websites on a very detailed level (Ghandour et al., 2010). Data collection about people visiting the site can be automated and enable the aggregation of data over many visitors, granting managers the ability to assess their website performance more holistically (Ghandour et al., 2010; Schonberg, Cofino, & Hoch, 2000).

The modern method of utilizing web technology to assess digital channels is referred as web analytics. It is used to “understand online customers and their behaviors, design actions influential to them, and ultimately foster behaviors beneficial to the business and achieve the organization’s goal” (Nakatani & Chuang, 2011). An advantage of WA is its ability to collect objective data on actual “online customer behavior and subsequent business outcomes” (Järvinen & Karjaluoto, 2015). Järvinen & Karjaluoto (2015) refer to WA as a “tool that collects clickstream data regarding the source of website traffic (e.g., e-mail, search engines, display ads, social links), navigation paths, and the behavior of visitors during their website visits and that presents the data in a meaningful format.” Clickstream data are defined as the “electronic record of Internet usage collected by Web servers or third-party services” by (R. E. Bucklin & Sismeiro, 2009). Clickstream data is obtained by tracking website visitors’ mouse clicks on a website. Wilson (2010) conducted a study in b2b context that claims the data gathered can be used well to get insight on how visitors get and use online information, react to digital marketing and make purchases. Marketing executives may use performance measures obtained from clickstream data and web analytics software also as a competitive asset to increase their overall digital marketing effectiveness (Wilson, 2010). The clickstream data of WA help managers to answer the questions “when” and “what” (Weischedel & Huizingh, 2006). For example, when the visitors entered the website and when did they complete a conversion? What did the visitors do on the site before converting? A manager at a publishing company might ask what were the top performing articles of the month and so forth. Since clickstream datasets contain the “activities of online users and records the virtual trail each user leaves behind while surfing the Web” (R. E. Bucklin & Sismeiro, 2009), many possible questions may be answered by analyzing this data.

Even though web analytics enable its users with many capabilities and offers vast amount of potential use cases, the usual case is that WA remains underutilized. Academic research on WA is limited and majority of it shows that there is a great development potential for WA utilization (Järvinen & Karjaluoto, 2015). Hong (2007) and Welling & White (2006) report that WA is used poorly and mostly occasionally on an operational level. They point out that there is lack of WA usage on strategic level and benefits of the usage on long-term business benefits remain unclear. WA has been studied in the fields of marketing, sales, communications information systems, information management, computer science, informetrics, webometrics, customer relationship management, performance measurement, website design, e-commerce, statistics and journalism in both b2b and b2c context across different industries. The next section of this chapter explores these studies.

According to R. E. Bucklin & Sismeiro (2009), clickstream data analysis potential for both practical and academic marketing is clearly untapped. For example, the number of published academic papers that base on clickstream data is much higher in the fields of computer mediated interactions and computer science than marketing (R. E. Bucklin & Sismeiro, 2009). Use of WA and clickstream data is even less common in the fields of information systems and information technology. In the next section we will also review research on clickstream analysis.

2.2 Current research on Web Analytics

Research on web analytics has been covered across multiple different academic disciplines. In this section, we review the most relevant ones for our research objectives.

Chaffey & Patron (2012) study WA from a digital marketing performance perspective. They find that companies often fail to get the potential return from web analytics. In turn, other research show that digital marketing performance measurement with WA has positively affected the efficiency of marketing actions and later increased sales revenue (Phippen et al., 2004; Wilson, 2010). Chaffey & Patron (2012) propose that the scope of WA should be enlarged beyond website optimization to other forms of digital media in order to emphasize the role of marketing optimization activities in marketing performance improvement. When discussing about the scope, they bring up the model of owned, earned and paid media as a current model considering this. Compared to other WA research, this is a rare case where the model of OEP is mentioned in a web analytics focused research. Järvinen & Karjaluoto (2015) enlarge the scope to online customer behavior measurement, online customer response measurement, and optimization of digital marketing elements and actions (Nakatani & Chuang, 2011). Companies can demonstrate short-term results of marketing actions in digital environments, because actions and resulting outcomes are linked to each other directly (Järvinen, Töllinen, Karjaluoto, & Platzer, 2012). This can be further utilized to optimize results by adjusting marketing actions in real-time.

In many former studies, the benefits of WA are discussed and proven regarding businesses whose transactions can be done online. Although WA is limited to digital environments, it drives the important development of more measurable marketing in a larger context (Järvinen & Karjaluoto, 2015). As a matter of fact, many offline marketing elements can already be tracked with WA (Järvinen & Karjaluoto, 2015). Also, Järvinen & Karjaluoto (2015) found that benefits from WA are extended also to business sectors in which transactions cannot be processed online. For B2B businesses it is still very typical that transactions are not processed online. Nonetheless, clickstream data analysis offers B2B marketers a more comprehensive picture of how Internet is affecting is B2B buying decisions according to previous studies (eg. Deeter-Schmelz & Kennedy, 2002; Wilson, 2010). Wilson (2010) support their claim by saying that clickstream analysis offers the ability to gain knowledge about website visitors and their website behavior, such as how much time they spend on a website and what products and services they are interested about. Connecting these web metrics to a database through cookie data or log-on information will give B2B a more complete picture of individual website visitor's behavior over time (Wilson, 2010).

Analysis of clickstream data, that is relevant for marketing, has developed through the last decades. R. E. Bucklin & Sismeiro (2009) group these advances in clickstream-based research into three broad research themes:

Theme 1. Peoples' navigation in the new medium including research papers of website choice, browsing behavior and the extent and nature of search across websites.

Theme 2. Online advertising methods including research papers of banner advertising, paid search, and email.

Theme 3. Online shopping and online purchases prediction including research papers on purchase conversion, completion of activities prior to purchase, online auctions and consideration sets.

2.2.1 Theme 1: Peoples' navigation in the new medium

Website browsing behavior was studied on the early days of clickstream analysis by Huberman, Pirolle, Pitkow, & Lukose (1998). They discuss the subject of understanding and predicting online behavior of an individual and suggest that cost-benefit perspective could be useful for that. However, they did not consider user navigation behavior changes over time and possible learning effects. Johnson, Bellman & Lohse (2003) explored the idea of visitors learning to use a website more effectively as they navigate and become more familiar with the website's content. They found evidence that website visitors' time per session decreases the more they visit the same website. Also R. Bucklin & Sismeiro (2003) explored learning effects and found that repeat visits by the same visitor had no effect on page view duration but led to fewer page views. Similar finding were made by Johnson, Bellman and Lohse (2003) with the addition that having fewer page views leads to decreased session duration but not less time spent viewing each page. When trying to determine whether having less or more page views or shorter or longer visit duration is better, it clearly depends on the business and its objectives. For an ecommerce business it might be good to get visitors to spend less time and browse fewer pages in order to speed up the purchase process. However, in media and entertainment it might be beneficial to maximize the number of page views in order to increase the amount of served ads on the site (R. E. Bucklin & Sismeiro, 2009), leading to increased revenue. If this is done by spreading content across different pages, the number of page views and page view duration may not be good metrics for site quality and users' engagement. In this case, the visit duration metric might be more appropriate in assessing site usage. (R. E. Bucklin & Sismeiro, 2009). According to R. E. Bucklin & Sismeiro (2009), detailed browsing records can be used to identify different site usage patterns across visitors. Moe (2003) did a cluster analysis based on visitor website behavior, where she identified four shopping strategies: search/deliberation, direct buying, hedonic browsers and knowledge-building. Each cluster was identified to have different navigation patterns and propensities to purchase from the site. Also Montgomery, Li, Srinivasan & Liechty (2004) studied browsing patterns by modeling website within-site transition choices. They explain browsing behavior through two states: deliberation and browsing. They suggested that visitor may switch from a state to other during a website visit and found evidence that this occasionally occurs. R. E. Bucklin & Sismeiro (2009) explored also other studies that investigate browsing behavior and search across multiple websites from visitor-level clickstream data. They brought up that doing this might help in predicting behavior on other websites based on the behavior in previous ones. Later, Park & Fader (2004) developed a stochastic timing model of cross-site visit behavior exploring the correlations in visit behavior between competing websites. They found that data on visit behavior on another site increases predictability of future visit behavior on another site. Park and Fader (2004**)** showed how this finding can be applied in forecasting when a visitor might make their first visit to a website based on their visits on competing websites.

R. E. Bucklin & Sismeiro (2009) remark that "click-stream based studies have found Internet users to have substantial site loyalty and high switching costs." Johnson, Moe, Fader, Bellman & Lohse (2004) report that shoppers and air travel site visitors stay loyal to a very few websites on a monthly basis. Smith and Brynjolfsson (2001) support these findings on their study of an Internet price-comparison service. They found that the biggest online book retailers could have a considerable price advantage over others that helps to get clicks from the comparison service. Visitor-level data on households was used by Goldfarb (2006) when discovering that users do have switching costs for online portals, which leads to loyalty that drives a large portion of the website traffic to the sites and generate substantial revenue.

Evidence on switching costs was also discovered by Chen and Hitt (2002) when they studied online brokerage firms' clickstream data. The variation of switching costs varied significantly across visitors, which was explained mostly system usage measures and firm characteristics that were associated with reduced switching costs.

The "lock-in" situation of having learning effects and switching costs, discovered from cross-site clickstream research is an important factor in daily Internet usage and recommends firms to pay attention to techniques in driving visitor retention (R. E. Bucklin & Sismeiro, 2009).

2.2.2 Theme 2: Online advertising methods

As reported by Bucklin and Sismeiro (2009), Clickstream data can be used to measure Internet advertising exposure and users' reactions to it, such as clickthroughs or purchases. This allows us to connect advertising exposure to a user's behavioral response (R. E. Bucklin & Sismeiro, 2009). The main categories of online advertising according to Bucklin and Sismeiro (2009) are display advertising, also known as banner advertising and paid search advertising, also known as search engine marketing, which usually occurs in services such as Google and Yahoo. Bucklin and Sismeiro (2009) found that clickstream research primarily focuses on banner advertising, but more research is happening around paid search. Also, e-mail has been studied as an advertising medium in clickstream-based research. Bucklin & Sismeiro (2009) point out that online word-of-mouth and recommendation systems and medium have become important topics but click-stream based studies have not been conducted in these areas in significant amounts. Social media, such as Facebook can be thought-of as an online word-of-mouth and recommendation system. Facebook also serves paid display advertising and therefore is an interesting subject for us to conduct clickstream related research.

Banner type advertising success can be measured with clickthroughs, but it is also argued that part of the value advertisers see in the format may come from other results that are generated by the exposure to paid banner advertising. All of this might not be measurable in clickstream data. (R. E. Bucklin & Sismeiro, 2009). Drèze and Hussherr (2003) suggest that brand awareness and recall would be more suitable measures for display advertising performance than clickthrough. Ilfeld and Winer (2002) found in their click-stream based research that digital advertising works well in attracting website visitors and it also has an effect on site awareness and brand. Banner advertising effects on visitor-level have been studied within websites. Manchanda, Dubé, Goh, and Chintagunta (2006) found that ad exposure on a given website had a positive impact on the repeat purchasing of existing customers. According to them, banner advertising may also have an impact on the website behavior within website. Rutz and Bucklin (2009) found that within-site banner advertising have segmented response effects on website behavior. Using individual-level site-centric data, they found the effect to be positive in one segment, negative in a second segment and in third segment banner ads had no effect.

However, there is little evidence regarding the effects of banner exposure on website behavior in a cross-site context. Our study will explore this further by analyzing banner advertising in a comparison of owned, earned and paid media exposure effects on cross-site web behavior.

2.2.3 Theme 3: Online shopping and online purchase prediction

Bucklin and Sismeiro (2009) report that understanding and modeling online purchase behavior has been one of the most active areas of clickstream research. Several different approaches have emerged with the focus on the factors that might predict online transactions, each utilizing clickstream data in different ways (R. E. Bucklin & Sismeiro, 2009). Moe and Fader (2004) used stochastic modeling to discover that visitors who visit a retail site more frequently will more be likely to make a purchase. The stochastic model used by them is limited, as it does not count user actions while browsing a site and their possible impact on purchases. Montgomery et al. (2004) found that purchase conversion can be predicted from visitor page views on a site using a modeling approach. The more pages a visitor viewed, the more likely it led to a purchase. A similar study has been conducted also in the field of statistics, where (Panagiotelis et al., 2013) propose a multivariate stochastic model to analyze website browsing behavior’s influence on purchase incidence and the sales for online retailers.

2.2.4 Research on Web Traffic and Behavior

Research on WA has also been conducted around web traffic. According to Zheng et al., (2012), web traffic data can be collected either offsite or onsite. From offsite data, one may compare a website's performance with others on the Internet (Clifton, 2010; Zheng et al., 2012). Onsite approaches track website visitors' interactions and engagements within a particular website (Clifton, 2010). Value of website traffic for website operating businesses has also been recognized in earlier research. For example, Luo & Zhang (2013) found a direct relationship between website traffic and firm performance. They argue that "web traffic also affects brand awareness and customer acquisition, and is a predictor of the performance of a firm's stock in the market." According to Vaughan (2008), web traffic to company websites can be used as an indicator of the firm's business performance. Vaughan & Yang (2013) support this claim, stating that traffic indicates business performance through net income and total sales but they better indicate performance of an online business. They also note that web traffic indicates academic quality. In their study, Vaughan & Yang (2013) used Spearman correlation tests that indicated a significant correlation between academic quality rankings and web traffic data. According to Moe & Fader (2004) retail and content provider online sites assess their overall success by routinely monitoring visitor traffic. Rajgopal, Kotha, and Venkatachalam (2000) say that web traffic can aid in establishing customer relationships and gather valuable purchase and website behavior data from the website visitors, which could potentially increase future revenues.

Usage of web traffic data has been studied also in the field on media and journalism. Tandoc (2015) found that online editors use WA primarily for traffic monitoring. This directs them to utilize audience information in their decision-making processes. Technology has enabled organizations to collect quantifiable information of an audience's news consumption, such as viewing time, number of shares, engagement and clicks (Napoli, 2010, Vu, 2014). Today, journalists can pull out figures and numbers in real time regarding their audience's behavior. When exploring visitor website behavior from web metrics, they are able to see how many readers there are at a given time, what content visitors prefer and where and how they discover the stories they consume (Vu, 2014). Therefore, web analytics can provide journalists information on its online audience, which can lead to making more relevant content or better content placement decisions (Tandoc, 2015). This naturally means new opportunities to base both business and editorial decisions on. It is reported that Washington

Post has made resource allocations solely based on web metrics. Some publishers, such as Bloomberg and Gawker media have started using web metrics as a basis of remuneration. (Nguyen, 2013). Also for other newsrooms, web metrics have become important performance indicators (Vu, 2014). In general, journalists are seeing an even higher pressure to follow web metrics and work towards goals that are based on the selected metrics (Nguyen, 2013). Online publishers are facing a paradigm shift when thinking about their content strategies. One of the challenges lies balancing between the old ways of making decisions and letting analytics be the judge of content publishing decisions.

In the travel industry, web traffic data from local destination marketing organizations provides value through its usability in hotel room demand prediction in a travel destination. Further, web traffic data is valuable for any business pursuing to observe future activities due to its universal availability (Yang et al., 2013). Bucklin and Sismeiro (2009) state that understanding online behavior is needed for the success of onliness businesses and websites as they compete in a complex environment. Ghandour et al. (2010) found a correlation between perceived success in website usage and perceived success of financial performance. Regarding this, they suggest to monitor website traffic in order to prevent downturn situations prior to their occurrence. According to Järvinen & Karjaluoto (2015), one of the advantages of web analytics is that it "offers a variety of objective, standardized, and quantitative metrics that are relatively easy to communicate to senior management". We can also quite comfortably say that web metrics are a good source of web behavior analysis.

According to Vaughan (2008), research based on analysis of web usage data is would be a natural extension to previous informetric research and also is related to webometric research, which has traditionally focused on analyzing *Web hyperlink data.* WA has also been identified as one of the main emerging research areas for almost the past two decades in the field of analytics and business intelligence (H. Chen & Storey, 2012). According to them, current research on web analytics on this field includes social media analytics, social search and mining, reputation systems and web visualization. They also see a potential growth in research that concerns social marketing, online auctions, internet monetization and web privacy/security, which many rely on development in various fields, such as social network analysis, text analytics and economic modeling.

2.3 Popularity of Web Analytics amongst practitioners

Over 60% of the top ten million globally most popular websites use WA (Web Technology Surveys, 2017). High adoption rate is driven by both the value of WA produced data and availability of free WA tools, such as Google Analytics (Järvinen & Karjaluoto, 2015). Google Analytics has an astounding 83,4% share of the web analytics market. In comparison, Piwik analytics is the 7th most popular tool with a 2% market share. (Web Technology Surveys, 2017). There are two main methods for web information gathering: page tagging and using web server log files (Pakkala et al., 2012). Using WA, the gathering of data can be standardized and automated (Russell, 2010). Therefore, data gathering is not considered to be a barrier of WA data usage (Järvinen & Karjaluoto, 2015). Google Analytics is a tracking application developed by Google. It records website traffic by inserting a small piece of HTML code in every page of the website. The application tells the user how visitors behave on the website and provides detailed statistics on the visitors (Plaza, Beatriz, 2009) such as their browsing and purchasing patterns (H. Chen & Storey, 2012). The statistics range from simple metrics like visitor counts, to monitoring multiple aspects of a visitor's session (Pakkala et al., 2012). Marketers commonly use Google Analytics to improve their website performance (Plaza, Beatriz, 2009) in order to increase the returns for their business. Plaza, Beatriz, (2009) studied website visitors visits internal performance depending on their traffic source using Google Analytics and clickstream data. Three types of traffic were analyzed: direct traffic, referring site traffic and search engine traffic. This gives us something to build on as our study aims also to analyze different types of traffic from clickstream data using WA tools.

According to Bucklin and Sismeiro (2009), the clickstream data analysis provides major development opportunities in handful new fast-growing areas of the Internet. Among these are social networking, word-of-mouth, multiple channel management and recommendation systems. Clickstream research is prone to lead methodological development in marketing as researchers and analysts try to solve challenges by using data to understand the impact of advertising, e-commerce, web usage and interactivity. This thesis expands the analysis of clickstream data in social networking and word-of-mouth (referred as earned media) in media and journalism, and so contributes to the existing research on web analytics.

2.4 Choosing the relevant metrics

A large variety of metrics makes the use of WA complicated (Järvinen & Karjaluoto, 2015). First of all, it is difficult to decide the most important metrics to be used and learn to use them (Phippen et al., 2004; Weischedel & Huizingh, 2006; Welling & White, 2006). Secondly, according to Zheng et al. (2012), one metric at a time does not give a comprehensive picture of website performance. Analyzing one measure at a time does not provide full understanding of a website's usage patterns (Zheng et al., 2012). For example, according to Zheng et al. (2012) analyzing visit duration alone without considering other metrics such as actions or conversions gives a very unsophisticated image of the website usage. Third, they state that the metrics can be measured on different levels. For example, data about the number of unique visitors tells about the popularity of the website compared to others but does not give information on how often a visitor visits the website (Zheng et al., 2012). Also, Zheng et al. (2012) find that reporting only some of the metrics sometimes gives a rosy picture of a given situation, which leads people to use metrics to whitewash their own situation. This is related to the fact that media organizations are not capable of proving their competitiveness through WA (Lacy, 2006). At the same time advertisers find little usable information that would guide their buying decisions (The Economist, 2007). However, both advertisers and media organizations depend of web metrics in assessing the market value of a website. Therefore, web metrics should be clearly defined, understood and chosen to disclose the value of a website in a given context.

Organizations should use web metrics to measure how well their websites support their business objectives. Therefore, web metrics should differ by website categories. According to (Chaffey & Patron, 2012), companies should begin the selection by identifying key performance indicators (KPIs), in other words metrics that are related to the business strategy or metrics that measure the primary goals, and differentiate them from other, less meaningful metrics. However, there is not much information about how organizations compile manageable and comprehensive set of WA metrics (Järvinen & Karjaluoto, 2015). For an online media site, it is important that the site and its individual pieces of content get traction and media attention. To assess the KPIs of a media company, one should understand the main revenue streams. Most of the media companies nowadays get the biggest chunk of their revenues from advertising. According to "State of News Media 2016" report, 72% of local publishers reported advertising generating half or more of their total revenues. Also other revenue streams such as subscription, sponsorship, grants, donations and events do exist (Mitchell & Holmcomb, 2016). Complete industry data of these revenue streams is unfortunately unavailable.

There are several common forms of advertising including banners, native advertising and video advertising. All of these are dependent on visitor exposure and in order to provide actual value for the advertisers, visitor attention. Another revenue model, which many media organizations have not successfully managed to employ in their digital channels, is paid content or paid reader subscription. In this case the user pays for the content they want to access. However, subscription still provides opportunities for digital publishers and companies such as New York Times have recently seen significant growth in their digital subscription revenues (New York Times, 2017). In this study, we concentrate on the advertising funded-model but acknowledge that studying online metrics may help editors in future content placement (Vu, 2014) that brings desirable results regardless of the business model.

For this study, it is essential to find out what core metrics should be and can be measured. In particular, here the interest is on the relevant metrics for an online publisher, whose content is mainly articles. The articles are mostly a combination of written text and pictures, but may sometimes include videos or fillable forms. The selected metrics are intended to help online marketers also from other industries to perform better.

First of all, the interest of this study is in visitors, not visits. As a standard, Google Analytics and many other web analytics tools give information about visits. Therefore, there is a need for tools that enable us to look at visitor-level data from websites. In order to delimit the study to website visitor behavior, all relevant metrics of an online publisher are not covered. A focus is also set to metrics that can be easily exported from the chosen web analytics tool. Many other metrics related to advertising, website performance and content itself are not covered in this thesis, but might be important for many publishers.

In this thesis, the chosen metrics for our case study were corrected visit duration, visitor actions, adjusted bounce-rate, conversions, new/return users ratio, mobile/desktop users ratio and visit count. Due to the explorative nature of the case study, learning what these metrics can reveal is a key point of interest. Also earlier research has found evidence on the importance and purpose of these metrics. However, the metrics that could be exported from Piwik analytics set a clear a limitation. In the next section, challenges related to the use of web analytics are presented.

2.5 Challenges with Web Analytics

Web Analytics does not come without challenges for both academics and practitioners. As a relatively new area of research, many of these challenges have not been covered in academic literature. In this section, challenges that are especially relevant for this study are discussed.

Firstly, web analytics contain certain technical constraints that make the data sets imperfect. As already previously stated, many web analytics tools use page tagging that relies on the use of JavaScript and cookies. However, some web users turn off both, and therefore become invisible for the measurement tool. Due to this, visible users do not currently represent reality perfectly. Due to the existence of robots, spiders and web crawlers that scan or download content from websites, some of the visitors that web analytics are reporting are non-human. Web analytics providers are using various techniques to identify and filter this kind of traffic. (Burby & Brown, 2007). Cookies may also be deleted by users, which may lead to tracking multiple visits of a same user as unique visits. However, according to Kaushik (2010), this is not considered a major handicap for WA.

One key challenge related to web analytics tools is the format of data you can export from them. Pakkala et al., (2012) introduced a challenge specific for Google Analytics. From the free version of GA, it is impossible to download “raw”, non-aggregated data, making it impossible for studying detailed distributions. The available data is mostly aggregated, preventing visitor-level behavior to analysis. In the next section, we discuss further why individual, visitor-level data is needed for the purposes of this study.

According to Järvinen & Karjaluoto (2015), web analytics data are based on the historical behavior of visitors, which makes it less helpful to analyze the future intentions of customers. They say WA data are only “quantitative and cannot be used to measure qualitative objectives” such as customer satisfaction or brand image. These measures might be finally more important for a company for their specific purposes (Järvinen & Karjaluoto, 2015).

Some of the challenges of web analytics are related to the definition and measurement techniques of web metrics. Often, practitioners do not have a full understanding of web metrics origins, which leads to false interpretations. One of them relevant for our study is the *visit duration challenge,* which is also connected to *bounce-rate challenge.* The visit duration metric is usually skewed by default and it does not report the true time of a visitor's visit. This is because bounce-visits are defined as single page visits and reported as 0-second visits by default. Web analytics tools do not usually count these visits into the sample whereof they derive the average time spent on a page metric. However, on a visitor-level data set, every visitor can be taken into account. Most of the visitors should be counted to give a more realistic picture of differences between visitor types. In an article-based online media, it is very common to have a high amount of single page visits as many visitors are arriving to the website through a link on social media and after consuming the content tend to return back. In today's website design it is also common to fit all the relevant content on a single web page. As a result, default web analytics settings would miss the measurement of visit duration of a large portion of the visitors. Bounce-rates are supposed to help in assessing the quality of the visitors, but it often fails to do so if the definition and measurement logic is not adjusted accordingly. Thus bounce-rate is here re-defined to measure all visits over 10 seconds utilizing corrected visit duration. This is referred as adjusted-bounce rate. The Piwik web analytics tool allowed applying a 'heart rate monitoring' tracking method on our case website to track the real visit duration of each visitor, even from single-page visits. Further, adjusted bounce-rate were measured because of their improved ability to indicate the actual quality of a website visitor how engaged the visitor is with the content. Staying more than 10 seconds on the web page is assumed to indicate real consumption of the website content. In the case study, the use of these re-defined metrics and their derived tracking methods are tested.

2.6 Owned, Earned and Paid Media and Traffic

Web traffic is generated by owned, earned and paid media channels, which can be separated from each other with measurement techniques enabled by web analytics. The question remains, what is the reasoning for tracking these different types of media and traffic separately and how could this tracking be done? What possibilities does it hold for practitioners? Answers to these questions are suggested by exploring past literature and proposing measurement methods for tracking each type of traffic to be used in empirical research about the differences between owned, paid and earned website visitors.

Knowing how to measure online traffic helps organizations in various ways. For example, news organizations can better understand their impact in the online environment and make improvements in their practices based on the audience news consumption. Advertiser, on the other hand, may learn to better allocate their ad budgets to gain attention. Controlling the types of visitors a site has may help to achieve their specific goals better. According to the multidimensional model of Zheng et al. (2012), websites that get a lot of return visitors can drive the goal of getting repeated ad impressions. A site with high time spent per page may be suitable for ads that demand higher levels of concentration (Zheng et al., 2012). If owned, earned and paid traffic have different impact on metrics that can be connected to different objectives, they may be used strategically to pursue different goals.

Some previous studies have investigated the different purposes of owned, earned and paid media. Luo & Zhang (2013) suggest that firms should invest in paid online advertising to improve web traffic and conversion. A recent study by Ghose & Todri-Adamopoulos (2016) claims that paid display advertising has “statistically and economically significant effects on increasing consumers’ propensity to buy”. Paid display media also increases users’ tendency to look for the brand and product advertised. Hu, Du and Damangir, (2014) found that paid advertising can drive sales in a contest where information search is the norm prior to a purchase. This means that paid media may convert information seeking consumers into buyers. Vu (2014) discusses that earned media may help in attracting new customers.

From a publisher perspective, technological development has made it possible for online audiences to participate in news distribution. In the best case, the audiences may help spread the news and attract more website visitors who may also count as readers (Vu, 2014). Xie and Lee (2015) studied owned and earned media in social media for fast-moving consumer good brands using a two-stage decision model. They found that "exposures to earned and owned social media activities have significant and positive impacts on consumers' likelihood to purchase the brand". However, they did not find that similar exposure would increase the amount purchased offline when accompanying in-store promotions. Their study was based on a dataset of 12-month home scanned purchase records and Facebook brand page messages related to brands.

Stephen & Galak (2012) ran a multivariate time-series analysis on a data from an online microlending marketplace in order to analyze both traditional and social earned media effects on sales. They found that both traditional and social earned media have impact on sales. Social earned media is also significant in boosting traditional earned media. Harrison (2013) found that paid media have serves better in reaching non-buyers, whereas for earned media it is more difficult. Amongst the different types of media, Harrison (2013) reports that owned and earned media have the greatest traction amongst existing buyers. In addition to direct benefits and effects different types of media have for business performance, they also offer other possibilities for companies to transform their business. Berman et al. (2011) suggest that media and entertainment companies can innovate new sponsored revenue models utilizing owned, paid and earned media as components. Instead of traditional advertising, they could innovate new revenue models in order to appeal to the consumers (Berman et al., 2011).

In our study we are exploring the owned, earned and paid traffic in a social media context. In earlier studies, traffic generated by social media has been slightly discussed. The case companies in Järvinen & Karjaluoto (2015) study said that they were satisfied to identify the amount of traffic social media was driving to their website and the results of it.

Being the number one traffic generating social media for digital publishers (VanNest, 2016), Facebook was chosen to be examined in this study as the source of traffic. Also, all different types of media appear in Facebook. Specifically social media and Facebook have been connected to owned, earned and paid media in a previous study: "owned social media is social media activity that is generated by the brand owner (or his/her agents) in social networking services (e.g., Facebook) that it can control" (Xie & Lee, 2015). One example of this is posts that are published on company-owned pages/accounts on Facebook. Stephen and Galak (2012) define earned social media as "social media activity related to a brand that is not directly generated by a brand owner or its agents", but is related to the brand. "Earned media is online community posts in consumer-generated social media such as status updates in Facebook" (Stephen & Galak, 2012). A comparable example of paid media in similar social media context can be Facebook Ads. (Dehghani & Tumer, 2015; Duffett, 2015) Facebook ads are interactive in their essence: they offer users opportunity to "like" and "share" them (Dehghani & Tumer, 2015), thus enabling the paid media content to transform into earned media. Similarly, owned media posts on Facebook may be liked and shared and therefore turned into earned media. It is acknowledged that the boundaries are increasingly blurred between owned and earned media, especially in social media. Xie et al. (2015) say that most literature analyzes earned social media effectiveness without accounting for the effect of firm-generated owned social media. They further state that it is not well known how earned and owned social media interact with each other especially in current social media platforms where they coexist.

The following table presents the definitions of Owned, Earned and Paid Media as understood in the context of this thesis.

	Owned Media	Earned Media	Paid Media
Social Media Applied definitions	“Social media activity generated by the brand owner or its agents in social networking services it can control” (Xie & Lee, 2015). Equivalent in Facebook are organic page posts.	“Online community posts in consumer-generated social media” (Xie & Lee, 2015). Equivalent in Facebook are user generated posts or posts shared by users and other third parties.	Social media activity or posts that are generated by the company or the brand or its agents and has paid distribution. Equivalent in Facebook are paid ads, such as sponsored page post ads.
General Definitions	“Media activity related to a company or brand that is generated by the company or its agents in channels it controls” (Stephen & Galak, 2012).	“Media activity related to a company or brand that is not directly generated by the company or its agents but rather by other entities such as customers or journalists” (Stephen & Galak, 2012).	“Media activity related to a company or brand that is generated by the company or its agents” (Stephen & Galak, 2012). In return, a payment is done.

Table 3: Owned, Earned and Paid Media definitions

In our case study we refer to owned social media simply as owned media, earned social media as earned media and paid social media to paid media. We want to further explore three different types of traffic that social media generates and how they differ from website behavior perspective.

2.7 Measuring Owned, Earned and Paid Traffic

Some previous studies have used web traffic data (Chang & Huang, 2009; Wolk & Theysohn, 2007) in their research, yet it is still fairly little studied and utilized. Vaughan and Yang (2013) note that the growth of web-based activities and services on the Internet makes web traffic increasingly valuable. They find that several main sources of web traffic data exist, each of them with varying data collection methods and traffic metrics. The relative advantages of different measurement and reliability of different data sources raises questions amongst academics and business managers (Vaughan & Yang, 2013).

Vaughan & Yang (2013) have studied the differences between data sources in the field of webometrics. According to Vaughan & Yang (2013), previous studies of web traffic have used data gathered from one particular website or traffic data from multiple sites. In their study, they focus on data from multiple sites. In our study we focus on measuring traffic to a single website (www.elle.fi) from a single source (www.facebook.com).

In order to measure and study the differences between owned, earned and paid website visitors, we need to define what we mean by these visitors and then find a way to distinguish them from each other in web analytics. By “owned website visitors” we mean users, who come to the website through a link within owned media content. By earned website visitors we mean users, who come to the website through a link within earned media content. By paid website visitors we mean users, who come to the website through a link within paid media content. Therefore, we see that our challenge is to measure the traffic through a set of different links distributed by each type of media separately. Later in the case study we describe how we do this by applying a web analytics campaign tracking method. As a result, we get a data set of website visitors that originates from one of the different types of web traffic. A similar method can be applied in tracking traffic from different media and channels (Bałazińska, 2017).

Getting links tracked in owned and paid media is fairly straightforward, because these media types are controlled by the publisher (or its agents). Tracking earned media using clickstream data is significantly more challenging (R. E. Bucklin & Sismeiro, 2009).

Trusov, Bucklin, and Pauwels (2009) used clickstream data from an Internet social network site to track user generated email invitations aiming to get more users to join it. Invitations were paired with data on new member sign-ups. They were not able to link the data at the individual-level, but through aggregate time series analysis they found that earned media “significantly affects the number of people who subsequently sign up to join the network.” R. E. Bucklin & Sismeiro (2009) state that if researchers could track earned media and connect it to business critical metrics or other outcomes, clickstream research would be able to produce new insights into the performance of earned media in marketing communications. This study aims to track traffic from user generated Facebook posts in a controlled way. In addition, page posts and advertised posts made by the page owner are tracked to make the comparison.

In the next section the frameworks behind the analysis of owned, earned and paid website visitor are explained. Additionally, the research methodology used in the case study is introduced.

3 Research Framework and Methodology

This study combines 'Major components of an online marketing system' (Tonkin et al., 2010), behavior analysis of 'Trinity approach' (Kaushik, 2007) and visitor-level analysis from 'Five-dimensional model of web attention' (Zheng et al., 2012) to create a model to measure owned, earned and paid media traffic on a website on visitor-level and analyze the differences in the behavior of website visitors generated by each type of media. In the next section, each of the key WA frameworks is described.

3.1 Web Analytics frameworks

3.1.1 Major components of an online marketing system

The "general framework of the major components of an online marketing system" introduced by Tonkin et al., (2010) connects web analytics with online marketing. The framework highlights that a company website is part of an interconnecting system that is engineered to drive a company's business goals. In the framework, web analytics' aim is to "measure and analyze all key components of e-commerce" in order to facilitate reaching business goals. As part of the framework, eight inbound marketing channels are introduced. According to Tonkin et al., (2010), web analytics can be used to quantify the value of different channels and analyze the relationship between a channel and a website.

The eight inbound marketing channels presented in the framework may create different types of traffic. Whereas SEM creates paid traffic, SEO is considered to create earned and owned traffic. Display and rich media advertising, according to its name, generates paid traffic. Affiliate and partner marketing may generate earned or paid traffic. E-mail, directory marketing and social media and offline channels may generate paid, owned and earned traffic.

The framework of Tonkin et al. (2010) connects marketing channels to the customer ecosystem, which represents the journey and conversion process of a website visitor into a customer. Channel effectiveness can be assessed towards conversion effectiveness in the e-commerce context using this framework (Tonkin et al., 2010), but our intention is to use it to evaluate website visitor behavior and engagement in the media context. With web analytics, it is also possible to analyze the relationships between owned, paid and earned traffic and a website.

3.1.2 Trinity approach

The Trinity approach by Kaushik (2007) is used to gain actionable insights and metrics. It helps in understanding "visitors that drive strategic differentiation and competitive advantage". The framework has three components:

1) Behavior Analysis
2) Outcome Analysis
3) Experience Analysis

"Behavior analysis is about what the visitors do, experience analysis is about why the visitors do the things they do" and outcome analysis, is about "how well the website is achieving its goals" (Riihimäki, 2014). Web analytics enables us to analyze the behavior of owned, paid and earned visitors and explore the differences between them. Our study refers to the framework's behavior analysis when we analyze website behavior as a part of our study.

3.1.3 Levels of analysis & five-dimensional model of web attention

As the Internet has changed how audiences consume media, Zheng et al., (2012) stresses the importance of attention in the media, advertising and audience measurement industries. Attention is defined as "focused mental engagement on a particular item of information" (Davenport & Beck, 2002, pp. 20–21). Zheng et al., (2012) also underline that attention should be separated from exposure, which means unintentional or intentional viewing of media content (Price & Zaller, 1993). Traditional criteria in audience measurement and advertising have relied on exposure but nowadays, the Internet requires active and attentive audience, because web users need to take action in order to view content (Webster, Phalen, & Lichty, 2006). Zheng et al., (2012) argue that "money cannot reliably buy attention to Web sites." However, they continue by saying that websites that can raise attention towards them, can effectively monetize attention. Therefore, attention is a top priority for businesses on the Internet.

The conceptual model for measuring web attention proposed by Zheng et al., (2012) measures attention on Internet in 5 dimensions at 5 different levels of analysis: visibility (share per market), popularity (unique audience per site), loyalty (visits per person), depth (pages per visit), and stickiness (time per page). The hierarchical structure, presented in Figure 1 demonstrates these levels as market, web site, visitor, visit and page, where per visitor is higher than per visit because a visitor can pay multiple visitors on a site and a web site can have multiple visitors and so forth.

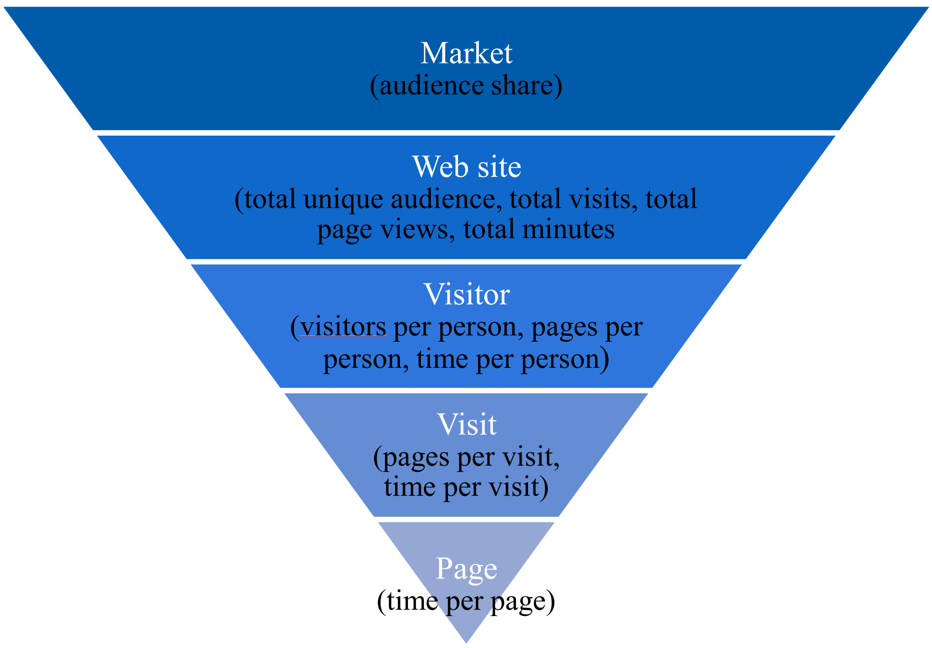


Figure 1: Levels of analysis and corresponding metrics (Zheng et al., 2012)

Measures within the same level of analysis have high correlations where as measures across levels have low. Therefore, each level represents a different scale by which performance of a website can be measured. These levels are not interchangeable with one another. Our study focuses on the Visitor –level. The relevance of Visitor –level analysis is also discussed in previous research: “Measuring visitor statistics is a core activity for any website provider” (Pakkala et al., 2012).

The metrics for our research should be in line with visitor-level metrics, generally approved website behavior metrics and they should tell us something about the visitor's attention towards the website. In the next chapter, introduce the rationale for our metrics chosen in our study.

3.2 Research Framework

Figure 2 presents the research framework used in this thesis. It combines frameworks from web analytics (Kaushik 2007; Tonkin et al., 2010) and media management (Zheng et al. 2012) and adds owned, earned and paid media known in management information systems (Xie et al., 2015), marketing (R. E. Bucklin & Sismeiro, 2009; Stephen & Galak, 2012; Trusov et al., 2009) and informetrics (Vaughan & Yang, 2013) literature to the framework in order to study differences between owned, earned and paid website visitors. The framework serves as a tool to measure the website behavior of different types of website visitors and uses the insight gained from the analysis of the measured data in marketing, content and media development. The framework provides a method to perform a comparison of owned, earned and paid visitors based on web metrics.

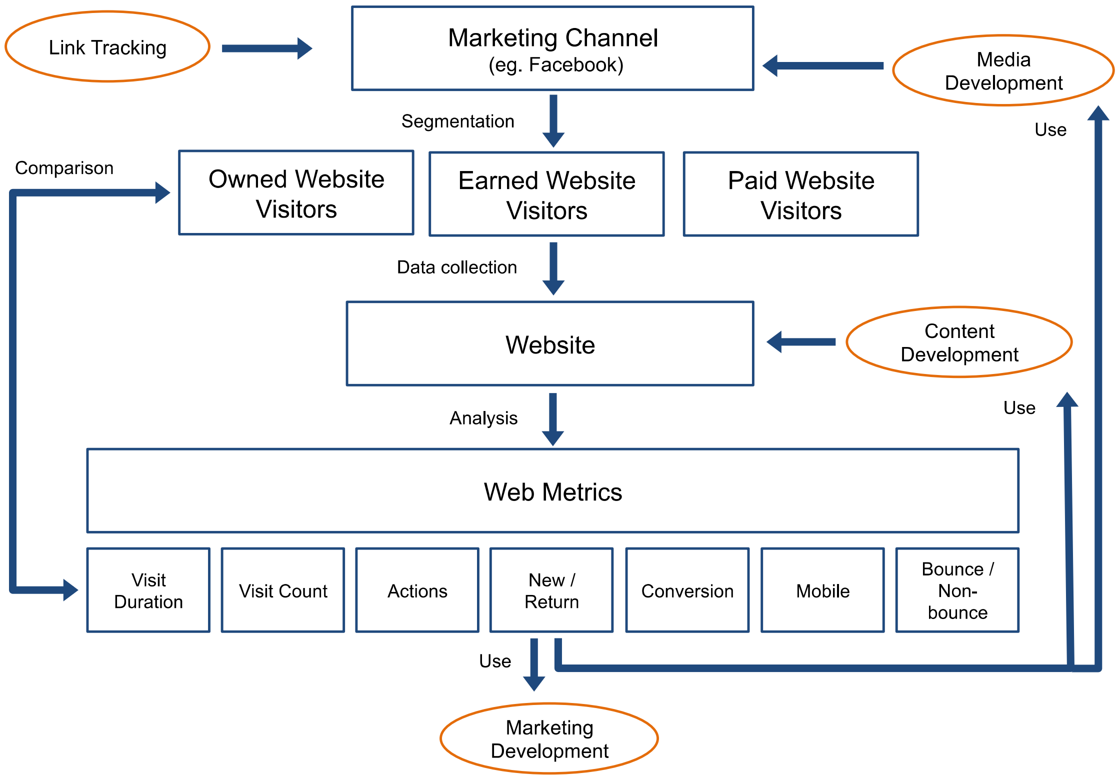


Figure 2: Analysis of website behavior of owned, earned and paid website visitors

The framework we propose brings an additional dimension to website visitor source (Tonkin et al., 2010): the "type of visitors". A link tracking method is placed on a chosen marketing channel separately in different types of media to distinguish different categories of website traffic from each other. These different types of traffic are tracked all the way in web analytics of the website as different types of visitors. Website behavior is analyzed by using the available set of web metrics and the differences are compared between different types of visitors. The web metrics shown in the framework may be used according to the need. Here we use the metrics chosen for our case study for illustrative purposes. The findings of the analysis are applied in developing marketing activities, website contents and media investments.

The objective of the case study is to examine if the proposed framework can be used in measuring owned, earned and paid traffic in a social media context so that differences between the visitor types may be quantifiably analyzed. The thesis aims also to find useful information for managers in their marketing, content and media development.

3.3 Research Methodology

This research methodology can be characterized as empirical exploratory research that combines theory-building and case study of a single case company. A literature review of existing research of web analytics and owned, earned and paid media was conducted. Prior research combining these two areas is limited due to their contemporary nature, and so a theoretical framework was built to connect key concepts from both areas of study. In order to test this framework and study a new subject, a case study approach was chosen. The case study approach is preferred when a contemporary phenomenon is investigated in its real-life context and boundaries between the context and phenomenon are not absolute (Yin, 1981). Both Web Analytics and Owned, Earned, Paid social media are contemporary phenomena as both of the fields have evolved recently due to the developments in technology and the Internet. Academic research and practical implications of the phenomena are still in their early stages. The case study was conducted on a single web page using a single source of traffic in order to avoid an overly complex setting and to have a better ability to control and assess influencing background variables. The research will serve as a complement to existing research but also as a novel approach to online media measurement and web analytics.

A quantitative approach was chosen as web analytics data are exclusively quantitative (Järvinen & Karjaluoto, 2015) and the research objectives support quantitative analysis on the data. As the aim was to determine whether there are any statistically significant differences in website behavior between the means of owned, earned and paid visitor groups, a one-way analysis of variance (one-way ANOVA) was chosen. According to Mahotra & Birks (2007), ANOVA is ”straightforward way to examine the differences between groups of responses that are measure on interval or ratio scales”, therefore it fits well for our reseach purposes. As there are also dichotomous variables amongst the chosen web metrics, chi-square test of homogeneity is used. Chi-square statistic is commonly used to examine whether a systematic association exists between two variables (Mahotra & Birks, 2007). Piwik web analytics is used to gather the data from the case website during a measurement period and campaign tracking is used to assort the website visitors into three categories. In the last section of this chapter the tools used for data gathering and analysis are introduced more precisely.

3.3.1 Website Visitor Grouping

The first step in gathering the data was to distinguish owned, earned and paid visitors from each other into separate groups. These groups form a basis of quantitative methods of analyzing differences between the visitor types. The solution to this was to use Piwik campaign tracking to assign each different type of post that directs traffic to the website an individual tracking tag. For example owned media links, which were distributed through owned media posts in a Facebook community, were marked as “owned”. Earned media posts were marked as “earned” and paid media posts were marked as “paid”. This way, a visitor who clicks a tracked link is assigned a category accordingly which we can later identify in the visitor data that is exported from the analytics software. Campaign tracking adds a column to complete exported dataset that enabled us to assign each individual visitor to one of the based on the category “owned”, “earned” or “paid”. Therefore, each individual visitors’ web behavior could be connected to their visitor type, allowing us to further investigate them as groups.

3.3.2 One-Way ANOVA

One-way analysis of variance (ANOVA) can be used to determine whether any statistically significant differences exist between two or more independent groups (Mahotra & Birks, 2007). In this study the independent variable is the traffic/visitor type, meaning that we have three independent groups (owned, earned and paid).

One-Way ANOVA may be used to understand if there are differences in a dependent variable or variables based on the independent variable as it is the most commonly used method. If the result is statistically significant, the result will be followed by a post hoc tests. This procedure is recommended in many textbooks (eg. Keppel & Wickens, 2004; Malhotra & Birks, 2007). In our case, dependent variables are the chosen ratio web metrics that represent visitor website behavior: visit duration, number of actions, number of conversions, visit count and dichotomous web metrics: new/returning user, mobile/desktop user and bounced/non-bounced user. In the case of these dichotomous variables, chi-square test of homogeneity is used. The Chi-square test is introduced in the next subsection.

In order to run One-Way ANOVA, there are six assumptions that need to be considered (Laerd Statistics 2017; Mahotra & Birks, 2007):

Assumption 1)

At least one dependent variable that is measured at the continuous level exists.

Assumption 2)

One independent variable that consists of two or more categorical, independent group exists.

Assumption 3)

There is independence of observations. This means that there is not relationship between the observations in each group of the independent variable or between the groups themselves.

Assumption 4)

There should be no significant outliers in the groups of independent variable, meaning that their value is very small or large compared to the other values.

Assumption 5)

The dependent variable should be approximately normally distributed for each group of the independent variable.

Assumption 6)

There should be homogeneity of variances.

After testing how our data relates to the six assumptions, the procedure of conducting one-way ANOVA introduced by Mahotra & Birks (2007) is followed. First they suggest that the dependent and independent variables are identified, then the total variation is decomposed After that, they instruct that the strength of the effects are measured and significances are tested. Finally, the results are interpreted. (Mahotra & Birks, 2007).

The one-way ANOVA calculates an F ratio based on the variability between groups versus the variability within groups (Laerd Statistics, 2017). The probability (p-value) of finding F ratio as large as the one calculated by the one-way ANOVA is used to either reject or not reject the null hypothesis. If the probability is less than 0,05, then there is less than 5% chance of the F ratio being as large as calculated, given that the null hypothesis is true (Laerd Statistics, 2017). According to Laerd Statistics (2017), this is interpreted as a real difference between group means in the population. When this occurs, the result is statistically significant (Laerd Statistics, 2017).

According to Laerd Statistics (2017), One-way ANOVA is an omnibus test statistic so it only tells that at least two groups were different. However, it cannot tell which groups were significantly different from each other amongst all (Laerd Statistics, 2017). Since we have three groups, it is important to know which of these differ from each other. This can be done with follow-up tests, which can be customer contrasts or post hoc tests. We choose the post hoc test. Because we do not have prior hypotheses about which groups might differ we run a post hoc test that tests all possible group comparisons. The Tukey post hoc test is said to be decent (Westfall et al., 2011) and recommended (Kirk, 2013), but our study contains a design imbalance (inequal number of cases in each group) and therefore a modified version of the test, Tukey-Kramer post hoc test is used in our study as it gives the same result as Tukey post hoc test (Laerd Statistics, 2017).

3.3.3 Chi-square test of homogeneity

The Chi-square test of homogeneity can be used to determine whether there is a difference in three or more proportions (Mahotra & Birks, 2007), which is a common way to think about dichotomous variables. Chi-square test of homogeneity can be used also in cases where the independent variable has two groups (Laerd Statistics, 2017). In our case we use the chi-square test of homogeneity to determine whether the proportions of owned, earned and paid website visitors are new versus returning users, mobile versus desktop users and bounced versus non-bounced users are equal.

According to Laerd Statistics (2017), to run chi-square test of homogeneity, five basic assumptions must be considered:

Assumption 1): There is one dependent variable that is measured at the dichotomous level. Dichotomous variables can be nominal or ordinal, meaning that the independent groups are either unordered or ordered.

Assumption 2): One independent variable that has three or more categorical, independent groups.

Assumption 3): Independence of observations is required, which means that there is no relationship between the observations in each group of the independent variable or between the groups themselves.

Assumptions 4): Only certain types of sampling/study design can be used with the chi-square test of homogeneity: a) single sample is taken and from this single sample, the participants are randomly assigned to groups or (b) when prospective or retrospective types of purposive sampling have been conducted. If purposive sampling is used, a specific number of people for each group is pre-specified: each group being based on having one particular characteristic that is different to the other groups.

Assumptions 5): Sufficiently large sample size is required so that the approximation to the chi-squared distribution is valid.

After the chi-square test of homogeneity is done, we want to know where the differences are. To determine this between the groups a post hoc test must be conducted (Laerd Statistics, 2017). In this case, pairwise comparisons are used. Using SPSS, a z-test of two proportions is run, which tests every possible combination of differences between the proportions of our groups (Laerd Statistics, 2017).

3.4 Research tools

Web analytics provide capabilities to collect and process high volumes of data (Nakatani & Chuang, 2011). According to (Nakatani & Chuang, 2011) web analytics tools "collect click-stream data, track users navigation paths, process and present the data as meaningful information." At first Google Analytics was considered as our main data collection tool. According to Pakkala et al. (2012), Google Analytics is easy to use and gives useful and versatile information about website visitors. Also, they recommended using similar tools available. Google Analytics uses page tagging and is also used by over 80% of the websites that use traffic analysis tools and are commonly visited (Pakkala et al., 2012).

Ultimately, it was decided to use Piwik analytics. Piwik is very similar to Google analytics, but there were few critical differences that made it ideal for this study: the ability to export visitor data on user-level and possibility to reconfigure the way the visit duration is measured. Pakkala et al. (2012) confirm that with the free version of Google Analytics it is impossible to download non-aggregated raw data of the visits, while – based on tests – Piwik allows this. Piwik was used to gather the clickstream data from the case website, and its campaign tracking was used to separate owned, paid and earned traffic. Facebook and its advertising tool were used to drive traffic to the case website. After the data collection, the exported data was refined in Excel. Data analysis, One-way ANOVA and Chi-square test of homogeneity were conducted by using IBM SPSS Statistics software.

3.5 Getting visitor-level data

Most web analytics tools provide aggregated data, which further can be segmented according to user need. In order to conduct proper quantified analysis, such as Chi Square test of homogeneity or One-Way ANOVA in our case, individual-level data is needed. R. Bucklin & Sismeiro, (2003) also say that aggregate-level tracking statistics should be interpreted with care as aggregate-level tracking can be misleading. The free version of Google Analytics does not give individual-level data to us, so other tools had to be evaluated in order to do the analysis. Therefore, PIWIK analytics was chosen, as it offers a possibility to export individual data from the web visitors for our chosen period of time. In our case we were able to export excel sheets where each row represented an individual visitor, showing the different metrics we are analyzing in our study per visitor.

According to Moe and Fader (2004), aggregate measures may seem good for site managers on the surface, but may in reality be misleading. They claim that any attempts to make conclusions of visit times directly from observed data might not lead to making "accurate estimates of true rates of repeat visiting" and learn true visit dynamics on a website. According to them, individual visitors' behavior often changes as they continually adapt to a new environment. If many new visitors flow in, this may change the aggregate reports significantly (Moe & Fader, 2004).

For the reason above, it is essential in our study to conduct the analysis on a visitor level that was introduced by Zheng et al., (2012). Our research also focuses especially on understanding the differences between owned, earned and paid visitors, which can be analyzed from visitor level statistics.

4 Case study and data

4.1 Case Introduction

Elle is a global lifestyle magazine that focuses on fashion, beauty, health, and entertainment. At the moment, Elle is the world's largest fashion magazine with 43 international editions in over 60 countries. In Finland, Aller Media Oy publishes Elle as a monthly appearing print version, but also frequently produces content, such as articles, blogs and competitions on its web page, www.elle.fi.

Supporting Elle's content and publishing strategy, Elle had a Facebook page consisting 12 000 fans during the time research was made. Recent Elle's online articles are shared through their Facebook page in addition to other original content. Also, Facebook advertising is used to generate traffic to Elle's website during campaigns. Some of the articles are shared in social networks by the audiences, which also creates a remarkable amount of traffic to the website. Therefore, the three types of traffic investigating in this thesis (owned, earned, paid) are constantly directed to Elle's website directly from Facebook.

In order to track owned, earned and paid Facebook traffic to Elle.fi in a controlled manner, certain preparations were needed. First, in all of the cases, using Piwik campaign tracking every post of an article was separately distinguished, assigning them a category (owned, earned, paid) based on the way of being published in Facebook. This way, Piwik would detect and tell the category of each individual visitor who came through a tracked link. For example, during the measurement period, owned posts in Facebook were assigned a tracked link that indicate its type, source and article name as follows:

http://www.elle.fi/ArticleURL/?pk_campaign=Owned-Facebook-ArticleName

Example of owned media tracking for a single article:

http://www.elle.fi/kilpailut/adidas-elle-juoksukoulu/?pk_campaign=Owned-Facebook-Adidas

Picture 1: Elle Owned Facebook Post example

Paid posts in Facebook were assigned a tracked link that indicates its type as follows:

http://www.elle.fi/ArticleURL/?pk_campaign=Paid-Facebook-ArticleName

Example of paid media tracking for a single article:

http://www.elle.fi/kilpailut/adidas-elle-juoksukoulu/?pk_campaign=Paid-Facebook-Adidas

Picture 2: Elle Paid Facebook Post example

Paid traffic was tracked across desktop and mobile devices. Content was made as identical as possible with the owned page posts. In targeting, Facebook look-a-like custom audiences from Elle's Facebook page fans were used. In practice, this mean that Facebook creates an audience to match Elle's current fan-base, so people who received these ads would be similar to the ones who follow Elle on Facebook and therefore receive the owned posts. With this targeting, we wanted the influence or targeting method to be as minimal as possible when comparing paid media with owned media.

Earned posts in Facebook were assigned a tracked link by configuring the Facebook share button on the article web page to generate a link that includes a tracked link that indicates its type as follows:

http://www.elle.fi/kilpailut/adidas-elle-juoksukoulu/?pk_campaign=Earned-Facebook-adidas-elle-juoksukoulu

Picture 3: Elle Earned Facebook Post example

Therefore everyone who shares the article through the Facebook share button found on the website after scrolling down to the end of the article, would share a tracked link so we can see which visitors from Facebook came through the earned posts. It is good to remember that there are also two other ways of getting earned traffic. First, when the article is shared by copying the link from the address bar and shared then on Facebook. Second, earned traffic is generated when the owned or paid post is shared inside Facebook. Unfortunately, we were unable to track these two kinds of earned traffic at this time. When owned and paid Facebook posts are being shared, there is also a risk of the traffic falling into wrong category, as the tracked link remains the same after sharing. However, during the measurement period no detected shares directly from our Facebook posts were made and we believe that amount of this kind of traffic remains small.

In the next section the data gathered in the case study during the measurement period is introduced. This data is later used in analyzing the web behavior of owned, earned and paid Elle website visitors. Further, the findings, reflecting prior research is assessed. The case study also serves as a first use case for our proposed research framework and provides valuable insight on the usability and usefulness of it.

4.2 Case Data

In this study, data was collected from one particular website, Elle.fi. The Piwik measurement tag was placed on each web page of the website in order to collect the clickstream data. The data set was gathered during a 25-day measurement period between April 24th and May 18th of 2016. The tracked traffic from Facebook drove posts to three different articles on Elle.fi. Each Facebook link post type (owned, earned and paid) was tracked separately for each of the published articles using the measurement methodology discussed. Based on their study of different traffic data collection methods, (Vaughan & Yang, 2013) claim that no particular method must be used as traffic data can be similar despite the method. They also state that any potential biases could be corrected with proper data normalization methods. Thus, our chosen data collection method should be appropriate for our purpose.

Visitor-data sets were downloaded as excel spreadsheets from Piwik analytics. During the measurement period, a total of 2739 website visitors from Facebook were tracked. 877 of the visitors were owned, 1079 were earned and 783 were paid. In its original form the data conflicts with the assumption 3) of One-Way ANOVA (independence of observations). In the data there are some duplicate visitors who either have visited many times, falling into the same or some other visitor category during the measurement period. Therefore, the duplicate visitors were removed from the data and only the first visit of a visitor during the measurement period is considered and included in the data. This also means that less return visitors will be present in the data. In the second data set 314 duplicate visitors were found and removed. After removing the duplicates, 2425 visitors remain. According to Plaza, Beatriz, (2009), over 1000 entries need to be collected in order to make statistically robust conclusions from Google Analytics. As Piwik is a similar tool, we may assume that our data set is sufficient for statistical analysis. The table below sums the final data used in the analysis and how it is assigned to different categories.

	Owned Visitors	Earned Visitors	Paid Visitors
Article 1	148	30	159
Article 2	444	910	441
Article 3	162	61	70
Total	754	1001	670

Table 4: Breakdown of case study data set

4.2.1 Comments on metrics

Prior to our actual case data set gathering, test data was gathered during a 7-day test period between March 9th and March 16th of 2016. After the test data collection, changes were made on the measurement of visit duration. As a standard, most web analytics tools, such as Google Analytics or Piwik do not measure the time a user spends on the last page of their visit to a site. This happens because they use the time of the next page view to tell the time spent staying at the current page. This means that visit duration is counted only from the visits that have entered more than one page on the site. The visitors that have left the site from the same page as entered will be counted as a zero second visit. Also, when the visitor views more than one page, the last page view of the visit will have a "time spent on page" of 0 second. Zero second visits are also considered as bounce visits, which causes the bounce-rate to rise. Generally, high bounce rate is considered to indicate that site entrance pages are not relevant to your visitors since they seem to leave the site immediately. This is problematic in the case of social traffic, especially for publishers, as visitors often tend to only visit the landing article and then return back to the social media to continue what they were doing. To test and to tackle this problem, we set up measurement method called "heart rate" to our Piwik settings "so that it accurately measures the time spent on the last page of a visit" (Piwik, 2017). With the new configuration, Piwik then sent "requests to count the actual time spent on the page, when the user is actively viewing the page" (Piwik, 2017). According to Piwik (2017), "these heartbeat requests will not track additional actions or pageviews". By default, Piwik sends a heartbeat request every 15 seconds but we changed the default interval to send a request every 10 seconds to allow a more accurate adjusted bounce rate (Piwik, 2017). Heart rate therefore allows us also to calculate close to the true visit duration of single page visits. We defined the visit duration measured by this technique "corrected visit duration". Also, bounce-rate was redefined as "adjusted bounce-rate" to be calculated from all visits under 10 seconds.

By looking at the data as whole, an overall adjusted bounce rate of 8,6% was observed, which is significantly lower than the bounce rate in the test-data set (91,6%). Therefore, we can conclude that actually 91,2% of the visitors were "real", meaning that they spent over 10s on the site. The number of mobile users (mobile phones + tablet) was as high as 77,4%. It is significantly higher compared to the figures reported in other prior research. For example, Pakkala et al (2012) reported just 4 years earlier that only 1,6-2,2% of visits were made by either mobile phone or tablet in their study. In our case, 76,9% of the visitors were new to the site and the average number of earlier visits was 2. The conversion rate was 2,1%. In our case, conversion was counted when the visitor completed a form submission in one of our articles. The average number of actions was 1,2. Average visit duration on the Elle site was 2 minutes 8 seconds.

This chapter introduced the case study and summarized the data that was gathered for the empirical, statistical analysis part. In the next chapter the analysis and findings made from this data are presented and analyzed.

5 Data analysis and findings

Statistical analysis involves hypothesis testing (Malhotra & Birks, 2007). In our case data analysis, the interest was in finding out whether a statistical significance exists between the groups of owned, earned and paid visitors. Accordingly, a **null hypothesis** was formed:

H0: all visitor type group population means are equal (i.e., µowned = µearned = µpaid)

With the sample data, we are trying to find evidence against this null hypothesis and accept the **alternative hypothesis**, which states that there are differences between the group population means (Malhotra & Birks, 2007):

H1: at least one group population mean is different (i.e., they are not all equal*)*

The null hypothesis is tested using chi-square test of homogeneity and one-way ANOVA in the next sections. First the results of the chi-square test of homogeneity are interpreted for our dichotomous variables. This is done separately for each of the variables. Second, the results of one-way ANOVA are interpreted and the results from it are summed up. Finally, the results from both statistical tests are discussed.

5.1 Chi-Square test of Homogeneity

First it can be concluded that the sample data complies with all the basic assumptions of Chi-Square test of Homogeneity. Analysis for each dichotomous variable in our data was run using IBM SPSS software. In this section, the results of the analysis are presented for each variable separately.

5.1.1 Bounce vs. Non-bounce visitors

			Owned	Earned	Paid	Total
Bounce	No	Count	680_a	938_b	598_a	2216
		%	90,2%	93,7%	89,3%	91,4%
	Yes	Count	74_a	63_b	72_a	209
		%	9,8%	6,3%	10,7%	8,6%
Total		**Count**	**754**	**1001**	**670**	**2425**

Table 5: Bounce / Non-bounce contingency table

First it must be stated that the sample size meets the requirements of chi-square test. Minimum expected frequency of the cells is 57,74, which is greater than 5.

In total, 2425 website visitors were analyzed. 754 of them being owned visitors, 1001 earned visitors and 670 paid visitors. To check if the proportion of bounced visitors was the same across the visitor categories the chi-squared test was run and produced $\chi^2(2) = 12{,}1$, $p = 0{,}002$ – thus the visitor type and bounce/non-bounce were dependent variables. A total of 74 (9,8%) of the owned visitors bounced compared to 63 (6,3%) of the earned visitors and 72 (10,7%) of paid visitors. Post hoc analysis involved pairwise comparisons using the test of two proportions with Bonferroni correction. The proportion of bounced visitors amongst earned visitors was significantly lower than amongst owned visitors or paid visitors, $\chi^2(2) = 12{,}1$, $p = 0{,}002$.

5.1.2 Mobile vs. Desktop visitors

			Owned	Earned	Paid	Total
Mobile	No	Count	155_a	277_b	106_c	549
		%	22,0%	27,7%	15,8%	22,6%
	Yes	Count	588_a	724_b	564_c	1876
		%	78,0%	72,3%	84,2%	77,4%
Total		**Count**	**754**	**1001**	**670**	**2425**

Table 6: Mobile / Desktop contingency table

First it must be stated that the sample size meets the requirements of chi-square test. Minimum expected frequency of the cells is 139, which is greater than 5.

In total, 2425 website visitors were analyzed. 754 of them being owned visitors, 1001 earned visitors and 670 paid visitors. To check if the proportion of mobile visitors was the same across the visitor categories the chi-squared test was run and produced $\chi^2(2) = 34{,}4$, $p < 0{,}001$ – thus the visitor type and mobile/desktop were dependent variables. A total of 558 (78,0%) of the owned visitors were on mobile device compared to 724 (72,3%) of the earned visitors and 564 (84,2%) of paid visitors. Post hoc analysis involved pairwise comparisons using the test of two proportions with Bonferroni correction. All pairwise comparisons were statistically significant, $\chi^2(2) = 34{,}4$, $p < 0{,}001$.

5.1.3 New vs. Returning visitors

		Owned	Earned	Paid	Total
Return	Count	268$_a$	138$_b$	155$_c$	561
	%	35,5%	13,8%	23,1%	23,1%
New	Count	486$_a$	863$_b$	515$_c$	1864
	%	64,5%	86,2%	76,9%	76,9%
Total	**Count**	**754**	**1001**	**670**	**2425**

Table 7: New / Return contingency table

First it must be stated that the sample size meets the requirements of chi-square test. Minimum expected frequency of the cells is 142,04, which is greater than 5.

In total, 2425 website visitors were analyzed. 754 of them being owned visitors, 1001 earned visitors and 670 paid visitors. To check if the proportion of new users was the same across the visitor categories the chi-squared test was run and produced $\chi^2(2) = 114{,}5$, $p < 0{,}001$ – thus the visitor type and new/returning were dependent variables. A total of 486 (64,5%) of the owned visitors were new users compared to 863 (86,2%) of the earned visitors and 515 (76,9%) of paid visitors. Post hoc analysis involved pairwise comparisons using the test of two proportions with Bonferroni correction. All pairwise comparisons were statistically significant, $\chi^2(2) = 114{,}5$, $p < 0{,}001$.

5.1.4 Conversions

			Owned	Earned	Paid	Total
Conversion	No	Count	120_a	29_a	139_a	288
		%	81,1%	96,7%	87,4%	85,5%
	Yes	Count	28_a	1_a	20_a	39
		%	18,9%	3,3%	12,6%	14,5%
Total		**Count**	**148**	**30**	**159**	**337**

Table 8: Conversion contingency table

Minimum expected frequency is of the cells is 4,4, which is less than 5. Therefore, we do not have an adequate sample size to run the chi-square test of homogeneity.

5.1.5 Summary of findings

	Owned		**Earned**		**Paid**	
	Count	Percentage	Count	Percentage	Count	Percentage
New / Return**	486	**64,5%**	863	**86,2%**	515	**76,9%**
Bounce-rate**	74	**9,8%**	63	**6,3%**	72	**10,7%**
Mobile-rate**	588	**78,0%**	724	**72,3%**	564	**84,2%**
Conversion-rate	Inadequate sample size					

Table 9: Chi-square tests of homogeneity summary

**statistically significant difference between groups ($p < 0,01$)

From the summary of chi-square test of homogeneity table it can be seen that new / returning users rate, bounce-rate and mobile-rate were statistically different ($p < 0,01$) between groups. Therefore we can reject the null hypothesis and accept the alternative hypothesis. Only conversion-rate did not have an adequate sample size to conduct the analysis and therefore no conclusions should be made about the differences between groups for that variable.

5.2 One-Way ANOVA

The first step of one-way ANOVA is to test the assumptions. As the original data does not comply with assumption #3, duplicate visitors were removed from the data sets. In addition, our data conflicts with the assumption #4, which requires the existence of no significant outliers. It is expected that some of the website visitors are being idle and focusing on something else in the meanwhile, therefore distorting the results. Our solution to this is to run One-way ANOVA analysis with three different data sets: first with the original data with all outliers included, second with data with 1% outliers removed and third with data with 5% outliers removed. The removal of outliers differs from the usual because outliers are not removed from the lower end. The findings suggest that with 1% outliers removed, statistically significant results can be produced for two of the variables and no significant changes in results can be noticed with data with 5% outliers removed.

The Shipiro-Wilk test was conducted to test the normality of the sample data. Our results show that engagement scores were not normally distributed, as assessed by Shapiro-Wilk's test ($p>0,05$). Therefore our data also fails to meet assumption #5. This is, however said to not be an obstacle if the sample size is not small and groups are similarly skewed (Sawilowsky & Blair, 1992), which is true in our case. Therefore one-way ANOVA can still provide robust results.

After dealing with the assumptions, one-way ANOVA was run using SPSS statistics.

5.2.1 Results of One-way ANOVA

One-way ANOVA with no outliers removed

Web Metric	Visitor type: Owned		Earned		Paid	
	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Visit Duration	2m 27s	281,40	2m 17s	287,56	1m 58s	266,30
Visit Count**	**3,57**	**8,81**	**1,38**	**4,12**	**1,30**	**1,52**
Actions	1,19	0,74	1,16	0,83	1,12	0,41

Table 10: One-way ANOVAs summary (no outliers removed)

**statistically significant difference between groups (p < 0,01)

Owned visitors have the highest average number of earlier visits (n = 754, M = 3,57, SD = 8,81). Earned and paid visitors had almost equally less earlier visits on average, earned visitors having slightly more (n = 1001, M = 1,38, SD = 4,12) than paid visitors (n = 670, M = 1,30, SD = 1,52).

Earned visitors also have the highest average visit duration (n = 1001, M = 137,1, SD = 287,6). Owned visitors have the second highest average (n = 754, M = 126,8, SD = 281,4), whereas paid visitors have the shortest average visit duration (n = 670, M = 116,8, SD = 266,3).

Average number of actions was the highest amongst owned visitors (n = 754, M = 1,19, SD = 0,74). Earned visitors have the second highest average (n = 1001, M = 1,16, SD = 0,83), where as paid visitors have the least actions on average (n = 670, M = 1,12, SD = 0,41).

To run the ANOVA for the variable means the homogeneity of variances was tested. Levene's test showed that variables in the visitor groups for "number of visits" and "actions" had heterogeneous variances with p < .001.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
visitDuration	Between Groups	168796,431	2	84398,215	1,077	,341
	Within Groups	189758796,214	2422	78347,975		
	Total	189927592,645	2424			

Table 11: One-way ANOVA output (no outliers removed)

Running standard ANOVA for visit duration showed no differences across visitor types. As homogeneity of variances was violated for variables visit count and number of actions, Welch's ANOVA is used.

The mean of the number of visits is different across types of visitors (F(2; 1327,419) = 24,312, p < .001). The same holds for the mean of the number of actions (F(2,;1550,980) = 2,908, p = 0,055),

One-way ANOVA after removing 1% of outliers (ZVALUE >-3.29)

From this data set we removed rows that were included in the top 1% (ZVALUE >-3.29.) of standard deviation for each dependent variable separately. After this, One-way ANOVA was conducted normally for each of the variables separately.

Visitor type

		Owned		Earned		Paid	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Web Metric	Visit Duration	1m 29s	154,47	1m 35s	160,89	1m 22s	141,06
	Visit Count**	**2,44**	**3,31**	**1,19**	**0,92**	**1,26**	**1,08**
	Actions**	**1,13**	**0,41**	**1,08**	**0,31**	**1,11**	**0,37**

Table 12: One-way ANOVAs summary (1% outliers removed)

**statistically significant difference between groups (p < 0,01)

Owned visitors have the highest average number of earlier visits (n = 728, M = 2,44, SD = 3,31). Earned and paid visitors had almost equally less earlier visits on average, earned visitors having slightly less (n = 998, M = 1,19, SD = 1,26) than paid visitors (n = 669, M = 1,26, SD = 1,08).

Earned visitors have the highest average visit duration (n = 970, M = 94,9, SD = 160,9). Owned visitors have the second highest average (n = 734, M = 88,5, SD = 154,5), whereas paid visitors have the shortest average visit duration (n = 654, M = 81,5, SD = 141,1).

Average number of actions was the highest amongst owned visitors (n = 744, M = 1,13, SD = 0,41). Paid visitors have the second highest average (n = 667, M = 1,11, SD = 0,37), whereas earned visitors have the least actions on average (n = 988, M = 1,08, SD = 0,31).

To run the ANOVA for the variable means the homogeneity of variances was tested. Levene's test showed that variables in the visitor groups for "number of visits" and "actions" had heterogeneous variances with $p < .001$.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
visitDuration	Between Groups	70560,748	2	35280,374	1,495	,224
	Within Groups	55567783,695	2355	23595,662		
	Total	55638344,443	2357			

Table 13: One-way ANOVA output (1% outliers removed)

Running standard ANOVA for visit duration showed no differences across visitor types. As homogeneity of variances was violated for variables visit count and number of actions, Welch's ANOVA is used.

The mean of the number of visits is different across types of visitors ($F(2; 1265,175) = 49,469$, $p < .001$). The same holds for the mean of the number of actions ($F(2; 1509,424) = 4,061$, $p = 0,017$)

Post Hoc Tests

Tukey post hoc –tests showed that earned visitors had a lower mean for the number of visits (M = 1,19 SD = 0,92) than the owned visitors (M = 2,44, SD = 3,31), a difference between means of 1,25, 95% CI [0,96, 1,55]), (p < 0,01). Also, paid visitors (M = 1,26, SD = 1,08) had a lower mean compared with the owned visitors (M = 2,44, SD = 3,31), a difference between means of 1,18, 95% CI [0,88, 1,49], (p < 0,01).

In number of actions means, the earned visitors group (M = 1,08, SD = 0,31) had a lower mean than the owned visitors (M = 1,13, SD = 0,41), a difference between means of 0,05, 95% CI [0,01, 0,09], (p = 0,014).

One-way ANOVA was also conducted after removing 5% of outliers (ZVALUE >-1.96). This test provided no differences to the significance of results.

5.2.2 Summary of findings

One-way ANOVA analysis was conducted with three datasets in order to find the impact of removing outliers. The dataset with 1% of outliers removed provided already statistically significant results from One-way ANOVA, so the results will be interpreted using data without 1% of the outliers.

Number of earlier visits

A one-way Welch ANOVA was conducted to determine if the number of visits was different for different types of visitors. Visitors were classified into three groups: Owned (n = 728), Earned (n = 998), and Paid (n = 669). The data was not normally distributed for each group, as assessed through a graphical view on the data. Homogeneity of variances was violated, as assessed by Levene's Test of Homogeneity of Variance (p < .001). Number of earlier visits was statistically significantly different between different types of visitors, Welch's F(2, 1265.175) = 46.469, p < .001. The Earned visitors group (M = 1,19, SD = 0,92) had a lower mean compared with the Owned visitors group (M = 2,44, SD = 3,31), a difference between means of 1,25, 95% CI [0,96, 1,55], which was statistically significant ($p < 0,01$). Also, the Paid visitors group ($M = 1,26$, $SD = 1,08$) had a lower mean compared with the Owned visitors group ($M = 2,44$, $SD = 3,31$), a difference between means of 1,18, 95% CI [0,88, 1,49], which was statistically significant ($p < 0,01$). The Earned visitors group ($M = 1,19$, $SD = 0,92$) had a lower mean compared with the Paid visitors group ($M = 1,26$, $SD = 1,08$), a difference between means of 0,07, 95% CI [-0,05, 0,19], which was not statistically significant ($p = 0,347$).

The group means were statistically significantly different ($p < 0.05$) and, therefore, we can reject the null hypothesis and accept the alternative hypothesis.

Visit duration

Earned visitors have the highest average visit duration ($n = 970$, $M = 94,9$, $SD = 160,9$). Owned visitors have the second highest average ($n = 734$, $M = 88,5$, $SD = 154,5$), whereas paid visitors have the shortest average visit duration ($n = 654$, $M = 81,5$, $SD = 141,1$). The data was not normally distributed for each group, as assessed through a graphical view on the data. There was homogeneity of variances, as assessed by Levene's test for equality of variances ($p = .130$). However, there was no statistically significant differences in visit duration for different types of visitors, $F(2, 2335) = 1,495$, $p = .224$.

The group means were not statistically significant different ($p > .05$) and, therefore, we cannot reject the null hypothesis and we cannot accept the alternative hypothesis.

Actions

A one-way Welch ANOVA was conducted to determine if the number of actions was different for different types of visitors. Visitors were classified into three groups: Owned ($n = 744$), Earned ($n = 988$), and Paid ($n = 667$). The data was not normally distributed for each group, as assessed through a graphical view on the data. Homogeneity of variances was violated, as assessed by Levene's Test of Homogeneity of Variance ($p < .001$). Number of earlier visits was statistically significantly different between different types of visitors, Welch's $F(2, 1509,424) = 4,061$, $p = 0,017$. In number of actions, the Earned visitors group ($M = 1,08$, $SD = 0,92$) had a lower mean compared with the Owned visitors group ($M = 1,13$, $SD = 0,41$), a difference between means of 0,05, 95% CI [0,01, 0,09], which was statistically significant ($p = 0,014$). The Paid visitors group ($M = 1,11$, $SD = 0,37$) had a lower mean compared with the Owned visitors group (M = 1,13, SD = 0,41), a difference between means of 0,02, 95% CI [0,01, 0,09], but which was not statistically significant (p = 0,398). The Earned visitors group (M = 1,08, SD = 0,31) had a lower mean compared with the Paid visitors group (M = 1,11, SD = 0,37), a difference between means of 0,03, 95% CI [-0,02, 0,06], which was not statistically significant (p = 0,364).

The group means were statistically significantly different (p < .05) and, therefore, we can reject the null hypothesis and accept the alternative hypothesis.

5.3 Discussion

Previous research by R.E. Bucklin & Sismeiro (2009) explored the influence of online advertising exposure to user behavior response and pointed out that clickstream research primarily focuses on paid banner advertising and paid search advertising. However, they suggested that online earned media have become important but lack profound research. This thesis contributes to the existing literature by adding owned and earned social media exposure to clickstream research and comparing them to paid social media exposure it in assessing user's behavioral response in a cross-site context. The thesis also combines social marketing with web analytics, which was seen as a potential growth area in the field of analytics and business intelligence by Chen & Storey (2012). Xie and Lee (2015) mention that most current research only minimally takes earned social media into consideration together with owned media and thus this thesis expands the current literature on owned, paid and earned media. The efforts of this study to track earned media and connect it to the other outcomes will provide clickstream an ability to clickstream research to provide new insights into the effectiveness of earned media in marketing communications (R. E. Bucklin & Sismeiro, 2009). It is acknowledged that empirical results of this thesis focus on a single case study on a specific industry and context. However, the framework is general in its applicability and may be used for future clickstream research in a different context and for different research objectives.

The framework created and used in the thesis worked for the purpose of the study and provided a statistical comparison between owned, earned and paid website visitors. Using the framework, we were able to analyze the relationship between the different media types on Facebook and the website, which supports the theory of Tonkin et al. (2010) framework and adds media types and visitor level web metrics to it (Zheng et al, 2012). The framework proposed in this thesis also uses behavior analysis introduced by Kaushik (2007) and connects it to the theory of Tonkin et al. (2010). The statistical methods used in this thesis were successfully applied using the thesis research framework and web metrics that were included in it. The Chi-square test of homogeneity indicated a statistical significant difference of binomial proportions of 'new / return user rate', 'bounce-rate' and 'mobile / desktop rate' variables. One-way ANOVA indicated a statistical significant difference between owned, earned and paid visitors of "visit count" and "actions", but also a non-significant difference of "visit duration".

The findings show that the model of owned, earned and paid is relevant extension to WA and marketing performance research as Chaffey & Patron (2012) suggested. There are significant differences between the media types on the online customer behavior, which was considered by Järvinen & Karjaluoto (2015). Based on their claims, we can assume that the insight about differences of owned, earned and paid visitors can help managers to improve their digital marketing performance measurement practices even further as well as demonstrate the business impact of their marketing actions. Next, we will discuss the results of the statistical analysis on each chosen metric and what assumptions can be made reflecting earlier research on WA.

Owned visitors tend to visit more often (μ=2,44), which is natural also from the Facebook perspective: these visitors are active followers of Elle's Facebook page and therefore they are being exposed to Elle's owned media content more often. E. J. Johnson et al. (2003) studied the learning effects of visitors finding evidence of changes in web behavior over time when visitors visit a site more often. According to them, visitors that visit more often tend to spend less time per session, but the results in this thesis do not directly support this claim as there is not a significant difference in visit duration compared to earned and paid visitors with significantly lower amount of visits. The findings of this research is supported other academics as well, seeing that repeat visits have no effect on page view duration, but led to fewer page views (R. Bucklin & Sismeiro, 2003; E. J. Johnson et al., 2003). Fewer page views can however be harmful to a publisher whose business model relies on banner advertising since more page views generally leads to more served ad impressions site (R. E.

Bucklin & Sismeiro, 2009; Zheng et al., 2012), generating more revenue for the publisher. However, due to the common "lock-in" of Internet users discovered by many researchers (eg. R. E. Bucklin & Sismeiro, 2009; P. S. Chen & Hitt, 2002; Goldfarb, 2006; E. J. Johnson et al., 2003; Smith & Brynjolfsson, 2001), efforts to increase the amount of owned visitors might serve a good technique to drive desirable visitor retention (R. E. Bucklin & Sismeiro, 2009). Both paid visitors (μ=1,26) and earned visitors (μ=1,19) have fairly low average number of earlier visits, which can indicate more page views (R. Bucklin & Sismeiro, 2003; E. J. Johnson et al., 2003) and therefore more ad revenue (R. E. Bucklin & Sismeiro, 2009). Earned visitors have the least average number of earlier visits and therefore earned media exposure may bring mostly new users to a website. The rate of new users supports this claim as 86,2% of earned visitors are new and therefore earned media appears to be a good way to attract new users. This study thus supports the proposal of Vu (2014) that earned media can help in attracting new customers. The rate of new users amongst owned visitors is 64,5%, which is the lowest compared to other types of visitors and may explain the difference in visit count. Paid visitors also are often fairly new, 72,3% of them being first time visitors with. Therefore it seems that also paid media may be considered in attracting new users when not precisely targeted to known existing visitors.

Owned visitors tend to take also more actions on the site (μ=1,13). This might be due to that these visitors are already engaged with the brand both in Facebook and have visited the website more in the past. Earned visitors tend to take least actions of the different types of visitors (μ=1,08) and for paid visitors, the number of actions on average (μ=1,12) is between owned and earned visitors, but the differences are relatively small compared to other types of visitors. More actions might also mean more page views, and therefore have positive effect on ad revenue for a publisher. Thus, owned visitors perform better on one of the website attention dimensions, *depth* (Zheng et al., 2012). High number of actions can also be a positive signal of engagement. However, 9,8% of owned visitors bounce from the page (adjusted bounce rate = spending less than 10 seconds on the page). The bounce-rate of earned visitors is clearly the lowest amongst the types of visitors. With only 6,3% bounce-rate, it seems that earned visitors are very engaged with the website content. Generally, adjusted bounce-rate is often seen as indicator of quality because visitors who spend less than 10 seconds on a website can be said to be fairly unengaged with its content. So, from that perspective, earned visitors seem to be of a higher quality than owned and paid visitors. For paid visitors, bounce-rate is the highest, 10,7%. Therefore, it seems paid visitors are the lowest quality, relatively speaking. According to (Pakkala et al., 2012) this means that the website is most relevant for earned visitors, second most relevant for owned visitors and least relevant for paid visitors. Thus a website and its content might be more relevant for a person when it is recommended by a friend or other third party.

78% of owned visitors access the website with a mobile device whereas 72,3% of the earned visitors did so, which is the lowest amount compared to others. Paid visitors are also the most mobile-penetrated as 84,2% of visitors visited with a mobile device. From current research there are no clear explanations for such differences, but generally the mobile penetration rates are nowadays higher compared to those in the past research (Pakkala et al., 2012). eMarketer (2016) claim that mobile users to be are hard to monetize for publishers, in this case would mean that paid visitors are the hardest to monetize. However, according to Heine, C. (2014), mobile users have undivided attention and therefore paid visitors and owned visitors would have relatively high attention towards the publisher. They also note that mobile offers better qualifications for location based personalization, so paid visitors could be served better with location-based personalization. Desktop users, in turn have better sales conversion in e-commerce and visual impact is greater for them (Heine, C, 2014), which would make earned visitors the most ideal group for e-commerce retailers in terms of business outcomes.

From the multiple pairwise comparison table (Appendix Table A1) it can be seen that not all of the pairwise comparisons are significant. For visit count, the comparison of earned and paid visitors is not significant. Regarding actions, the comparison between owned and earned visitors is the only significant one. This should be considered, when making more precise assumptions based on results of this case study.

Unfortunately, not enough conversions occurred in order to statistically assess difference between the groups. Visit duration is the only statistically non-significant metric we discovered in our case study. This might be due to the large variance between individual visitors. Visit duration is also a controversial metric as its purpose has been claimed to vary considerably. There is no clear and common understanding regarding how this metric should be interpreted (Panagiotelis et al., 2013; Wilson, 2010), so it is not that surprising to not detect significant differences of it.

The results of the case study expand the use ‘owned’, ‘paid’ and ‘earned’ jointly in a digital environment. While the previous studies have explored how consumer activity metrics of owned, paid and earned influence differ in terms of brand performance and sales

(Srinivasan et al., 2016), this thesis focus solely on web metrics and website visitors and thus contributes to the current research.

The case study also expands the use of alternative web analytics tools in academic research. Piwik analytics was used to gather the data and test the framework successfully, whereas Google Analytics has been the primary tool in the most WA research (Bekavac & Praničević, 2015; Järvinen & Karjaluoto, 2015; Pakkala et al., 2012; Plaza, Beatriz, 2009; Turner, 2010). To the best of my knowledge, this study is also first one to apply 'heart rate monitoring' measurement, redefined visit duration and bounce-rate metrics and one to one comparable measurement of visitor level owned, paid and earned website statistics. Therefore, the thesis contributes also to methodological development of clickstream research in marketing (R. E. Bucklin & Sismeiro, 2009) and provides useful insight about website measurement for those who operate or study article-based websites such as online media and journalism organizations and researchers. In addition, the thesis aims to helps both managers and academics in filling the information gap (Järvinen & Karjaluoto, 2015), that is defining and finding the most relevant WA metrics for their individual purposes, as suggested by Chaffey & Patron (2012).

6 Conclusions

Consumers are at accelerated rate replacing their traditional media consumption with digital experiences (Berman, Battino, & Feldman, 2011). Online readership is hoped to save the struggling publishers from declining revenues (Vu, 2014). Web analytics have become a popular subject with the rise of the Internet and company websites, as it helps companies to define the purpose and objectives of their websites (Riihimäki, 2014). The digitalization of media has emphasized the importance of owned channels such as a company's website and mobile apps. However, the existence of online services or websites does not make sense without continuous and relevant high-quality traffic from external sources. Website traffic has critical value for any business that operates a website. Understanding of online behavior is needed for the success of online businesses and websites as they compete in a complex environment (Bucklin and Sismeiro, 2009). Knowing online traffic measurement helps news organizations understand the effect of their work and improve their practices in the online environment (Graves, Kelly & Gluck 2010). Advertisers benefit from this knowledge as it helps them allocate their limited resources better when trying to maximize their share of consumer attention (Zheng et al., 2012).

Existing research on owned, earned and paid media is limited. Today, traffic can be divided to owned, earned and paid according to their source and redirect link distribution method. However, the differences between these types of traffic and website visitors have not been studied. Our aim was to fill the research gap by examining "owned", "earned" and "paid" media exposure in a cross-site context from the perspective of website visitors and their website behavior. Companies and organizations should focus on having more relevant traffic and visitors to their own channels in order to maximize the positive business impact of their website visitors. Therefore is it necessary to understand deeper the existence of different types of traffic and visitors on Internet.

6.1 Thesis Summary

The main research questions for the thesis was:

How does website behavior differ between owned, earned and paid website visitors?

Through the analysis of chosen web metrics in the case study, we were able to notice a statistically significant difference between owned, earned and paid website visitors as assessed by chi-square test of homogeneity and one-way analysis of variance. Each of the metrics have unique extent in their mean difference. Especially owned visitors tend to have more and earlier visits than other types and have the highest portion of return visitors. Earned visitors also seem to have notably lower bounce-rates than other types of visitors. Earlier literature on web analytics and marketing show the metrics can indicate different things in different settings. Assessing the impact, decision-making and directing the organization needs to be based on relevant metrics for the chosen business. Therefore choosing the right metrics to follow is very important for any company dealing with web analytics. Sometimes finding relevant metrics requires reconfiguring the given standard metrics. During the research process we defined two new modified web metrics in order to make them more relevant for our research purposes: corrected visit duration and adjusted bounce rate. Both managers and academics should carefully consider the interpretation of metrics and how to use them as a basis of managerial decisions.

In order to find out answers to the main research question, there was also a need to build a measurement model for owned, earned and paid traffic. Therefore the supporting research questions was:

How to measure owned, earned and paid traffic and visitors?

Definitions of owned, earned and paid website traffic and visitors were made and a literature review was conducted on prior research of web analytics and owned, earned and paid media. Combining features from prior web analytics frameworks, the thesis proposes a framework for analysis of website behavior of owned, earned and paid website visitors. The framework was tested successfully in a case study of an online publisher, resulting in quantitative assessment of the characteristics and website behavior of owned, earned and paid website visitors. Therefore, the framework answers the supporting research question by providing a method of measuring owned, earned and paid traffic and visitors, which is further exemplified through the case study. The framework may be used by both academics when conducting further research and managers when developing their online marketing and measurement.

In the next section, the usability of the results of our case study is analyzed from a managerial perspective.

6.2 Managerial implications

As proposed previously, the application of different metrics should be considered differently in different contexts. The managerial implications described here focus on applications mainly for media companies and online publishers, but could be applied across different types of websites within appropriate circumstances.

The first suggestion for managers is to consider the configurations of “visit duration” and “bounce-rate” web metrics. The redefined metric “corrected visit duration” tells the actual duration of single page visits instead of counting it as zero, and therefore provides a better picture for managers regarding content consumption and visitor engagement. “Corrected visit duration” may be used to all of the individual visitors better on site and helps making better decisions regarding the website content and its distribution. “Adjusted bounce-rate” on the other hand is valuable for assessing both the quality of the visitors and quality of the landing page especially in a single page context. As most of the online media and publisher pages are single page articles, it is valuable to measure them accordingly. “Corrected bounce-rate” gives a more realistic picture, as well as a much higher quality picture of these articles, as seen on tests run during our research. The differences between the old “visit duration” and “corrected visit duration” were dramatic. Reporting the new metric makes it therefore possible to justify the real value of web articles better.

From the significant differences between the web behavior of owned, earned and paid website visitors – as assessed by our chosen web metrics –cautious suggestions can be made for businesses. The multidimensional model by Zheng et al. (2012) helps set expectations for different sites on different web attention dimensions. According to them, frequent visits to a site or sites that manage to get users to view more pages could be more effective if we aim to create repeated ad impressions. Based on this, it can be deduced that visitor types that have more page views and frequent visits are better for the same goals. Owned visitors tend to have a bigger average count in both number of earlier visits and number of actions, so from a managerial perspective building and actively managing own social media communities would be beneficial in order to increase the amount of repeated ad impressions. Increasing owned media also seems a good way to attract and retain regular readers (Moe & Fader, 2004).

On retail sites, owned visitors may be considered when aiming to increase sales as more frequent retail site visitors have a greater propensity to buy (Moe & Fader, 2004). However, Sismeiro and Bucklin (2004) found that number of site visits was not predictive of purchase. According to Pakkala et al., (2012), visitor loyalty can be used to determine whether the website content and layout is satisfactory enough for the visitors to be willing to return to the website. Therefore, more loyal owned visitors seem more satisfied with the layout and content of the website. We make this assumption with cause, as our research time frame was relatively short. Loyalty is also favorable goal amongst Internet users since they have found to have high switching costs (R. E. Bucklin & Sismeiro, 2009).

As the rate of new visitors is the highest among earned visitors, second highest among paid visitors and lowest amongst owned visitors. This confirms the finding of Vu (2014), who believed that earned media may help in attracting new customers. As new visitors are claimed to convert more likely, it can be suggested that earned visitors are most likely to convert, making them more valuable than paid and owned visitors. Managers could aim to maximize the number of earned visitors for maximized conversions and prefer paid visitors more than owned visitors for the same purpose. This conclusion however should be considered carefully, as it may conflict with the earlier assumption of owned visitors being more likely to buy. Also, Plaza (2009) finds that return visitors may be more valuable due to their longer visit duration. Return visitors tend to spend a longer time on the site and have higher number of page views, but this does not necessarily turn into a higher conversion rate (Riihimäki, 2014).

Earned visitors have clearly the lowest adjusted bounce-rate of all users. As bounce-rate is an indicator of quality and website relevancy to a visitor, it seems recommendable that companies focus on increasing earned traffic in order to get more high quality and relevant traffic on a website.

For some reason, paid visitors access the website more through a mobile device. This might be because Facebook advertising algorithms that prefer to serve mobile ads over desktop ads. Earned visitors score lowest on mobile penetration rate. For a publisher, mobile users bring their own challenges: monetization (eMarketer, 2016), lower sales conversion, smaller visual impact but also benefits through better user engagement, undivided attention and location based personalization possibilities (Heine, C., 2014).

Finally, the thesis provides a framework for analyzing owned, earned and paid website visitors for managers. Businesses could utilize this framework as a part of their marketing, website content and media selection development in order to gain better returns for their online businesses. Managers are encouraged to test the framework in different contexts and with different metrics that fits their individual purpose.

6.3 Theoretical contributions

This thesis contributes mainly to research on web analytics, clickstream analysis, owned, earned and paid media and social marketing. The thesis also complements the academic disciplines of information systems science, management science, marketing and journalism in various ways. First of all, during the research process, new definitions were made. Owned, earned and paid traffic were defined alongside owned, earned and paid visitors. Two new metrics were defined: corrected visit duration and adjusted bounce-rate. These definitions may be applied in future research, where they may better meet specific research purposes and help other academics to reach more accurate results in research that utilize clickstream data or web analytics.

Secondly, the research compiles existing frameworks on web analytics and marketing in a new theoretical framework for website behavior measurement between different types of visitors. Further, the framework is tested it in an empirical setting. The framework provides a method to perform a comparison of owned, earned and paid visitors based on web metrics and may be used in future research in different contexts and with different set of metrics. The research also may help academic in choosing the relevant web metrics for their research purposes.

Thesis contributes also to methodological development of clickstream research in marketing (R. E. Bucklin & Sismeiro, 2009) by applying web analytics tools and configurations and introduces a way to distinguish owned, earned and paid traffic from each other in social media. In this thesis, a way to track earned media and connect it to business critical metrics or other outcomes is tested and therefore researchers conducting clickstream research will able to bring fresh insights into the performance evaluation of earned media in marketing communications (R. E. Bucklin & Sismeiro, 2009).

Exploration of owned, earned and paid traffic in social media context itself contributes to previous marketing research as previous research has focused on different channels. The research also adds variables of marketing mix to visit behavior across sites, as OEP model can be considered a part of the marketing mix. The impact of additional variables, such as OEP exposure can be used to complement research on individual visitor behavior and its prediction across sites as suggested by Park & Fader (2004). The research contributes to the existing literature web analytics and marketing by adding owned and earned social media exposure to clickstream research and comparing them to paid social media exposure it in assessing user's behavioral response in a cross-site context. The thesis also pairs social marketing with web analytics, contributing to current research in the field of analytics and business intelligence. According to Xie and Lee (2015) considering earned social media jointly with owned media expands the current literature on owned, paid and earned media.

Finally, significant differences were found between owned, earned and paid visitors which builds foundation for future research.

6.4 Limitations

The thesis has clear limitations that should be considered. First, data was collected from a single website and over a short, single period of time. Only three articles were measured with a relatively limited set of web metrics. Traffic was also driven solely from Facebook, which clearly does not represent all possible sources of owned, earned and paid traffic and visitors. Inside Facebook, only page posts were considered. Each page post represented only a single version of the content. Regarding paid media, we used a single targeting method, which contained a look-a-like audience from the case company's Facebook site fans. Also, every Facebook community is unique, which affects especially the composition of owned and paid visitors. Some of the earned visitors may have been missed due to the difficult measurement of earned media. Some visitors may also have ended up in wrong groups due to the nature of web analytics campaign tracking method.

The data was modified slightly when duplicate visitors and 1% of the outliers were removed.

6.5 Suggestions for future research

Further research could be done to extend the approach of the thesis and use of research framework in other kind of websites or landing pages of other types of businesses. Also, the media types could be published in different media. A longer measurement period or multiple measurement periods in different moments of time could be considered. Additional web metrics, such as number of pages visited and more types of conversions could be considered. Difference of owned, earned and paid website visitors could be considered also in relation to different sets of data. The study was limited solely on clickstream data, but for other types of quantitative and qualitative measures could be considered. For example, a questionnaire for the different kind of visitors might reveal some enlightening findings and deepen our understanding of them. Finally, the thesis relies on certain kind of research methods, but others might provide interesting results for the exploration of owned, earned and paid website visitors. Now that owned, earned and paid website traffic and visitors have been defined and a framework for their measurement is provided, future studies have an opportunity to explore them from multiple different aspects and deepen the understanding around the subject.

References

Bałazińska, Ewa. "A Complete Guide to Campaign Tracking in Your Web Analytics Platform." Piwik PRO - Cloud and Enterprise Analytics. N.p., 08 Mar. 2017. Web. 27 Apr. 2017

Bekavac, I., & Praničević, D. G. (2015). Web analytics tools and web metrics tools: An overview and comparative analysis. *Croation Operational Research Review*, *6*(6), 373–386. http://doi.org/10.17535/crorr.2015.0029

Berman, S. J., Battino, B., & Feldman, K. (2011). New business models for emerging media and entertainment revenue opportunities. *Strategy & Leadership*, *39*(3), 44–53. http://doi.org/10.1108/10878571111128810

Bucklin, R. E., & Sismeiro, C. (2009). Click Here for Internet Insight: Advances in Clickstream Data Analysis in Marketing. *Journal of Interactive Marketing*, *23*(1), 35–48. http://doi.org/10.1016/j.intmar.2008.10.004

Bucklin, R., & Sismeiro, C. (2003). A Model of Web Site Browsing Behavior Estimated on Clickstream Data. *Journal of Marketing Research*, *40*(3), 249–267. http://doi.org/10.1509/jmkr.40.3.249.19241

Booth, D., & Jansen, B. (2008). The methodology of search log analysis. In B. J. Jansen,A.Spink,&I.Taksa (Eds.), Handbook of research on Web log analysis (pp. 143–164). Hershey, PA: Idea Group Inc.

Burby, J., & Brown, A. (2007). *Web Analytics Definitions*. Retrieved from http://www.digitalanalyticsassociation.org/Files/PDF_standards/WebAnalyticsDefinitionsVol1.pdf

Chaffey, D., & Patron, M. (2012). From web analytics to digital marketing optimization: Increasing the commercial value of digital analytics. *Journal of Direct, Data and Digital Marketing Practice*, *14*(1), 30–45. http://doi.org/10.1057/dddmp.2012.20

Chang, S.-S., & Huang, C.-Y. (2009). Commonality ofWeb Site Visiting Among Countries. *Journal of the American Society for Information Science and Technology*, *60*(6), 1168–1177. http://doi.org/10.1002/asi

Chen, H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data To Big Impact. *Mis Quarterly*, *36*(4), 1165–1188. http://doi.org/10.1145/2463676.2463712

Chen, P. S., & Hitt, L. M. (2002). Measuring switching cost and the determinants in Internet-enabled business: A Study Industry of the online brokerage industry. *Information Systems Research*, *13*(3), 255–274. http://doi.org/10.1287/isre.13.3.255.78

Clifton, B. (2010). Advanced Web metrics with Google Analytics. Indianapolis, IN: Wiley

Corrigan, H. B., Craciun, G., & Powell, A. M. (2014). How Does Target Know So Much About Its Customers ? Utilizing Customer Analytics to Make Marketing Decisions. *Marketing Education Review*, *24:2*, 159–166. http://doi.org/10.2753/MER1052-8008240206

Croll, A., & Power, C. (2009). Complete web monitoring. Sebastopol: O'Reilly Media Inc

Danaher, P. J., Mullarkey, G. W., & Essegaier, S. (2006). Factors Affecting Web Site Visit Duration: A Cross-Domain Analysis. *Journal of Marketing Research*, *43*(September), 182–194. http://doi.org/10.1509/jmkr.43.2.182

Danaher, P. J., & Smith, M. S. (2011). Modeling Multivariate Distributions Using Copulas: Applications in Marketing. *Marketing Science*, *30*(1), 4–21. http://doi.org/10.1287/mksc.1090.0491

Davenport, T. H., & Beck, J. C. (2002). The attention economy. Boston, MA: Harvard Business Press.

Deeter-Schmelz, D. R., & Kennedy, K. N. (2002). An exploratory study of the Internet as an industrial communication tool. *Industrial Marketing Management*, *31*(2), 145–154. http://doi.org/10.1016/S0019-8501(01)00178-X

Dehghani, M., & Tumer, M. (2015). A research on effectiveness of Facebook advertising on enhancing purchase intention of consumers. *Computers in Human Behavior*, *49*, 597–600. http://doi.org/10.1016/j.chb.2015.03.051

DiStaso, M. W., & Brown, B. N. (2015). From Owned to Earned Media: An Analysis of Corporate Efforts About Being on Fortune Lists. *Communication Research Reports*, *32*(3), 191–198. http://doi.org/10.1080/08824096.2015.1016149

Drèze, X., & Hussherr, F. X. (2003). Internet advertising: Is anybody watching? *Journal of*

Interactive Marketing, *17*(4), 8–23. http://doi.org/10.1002/dir.10063

Duffett, R. G. (2015). Facebook advertising's influence on intention-to-purchase and purchase amongst Millennials Rodney. *Internet Research*, *25*(4), 498–526. http://doi.org/http://dx.doi.org/10.1108/MRR-09-2015-0216

The Economist. "Many ways to skin a cat." The Economist. The Economist Newspaper, 29 Nov. 2007. Web. 27 Apr. 2017.

eMarketer (2010, September 21). Blending Paid, Owned and Earned Media for Branding. Retrieved March 29, 2017, from http://www.emarketer.com/Article/Blending-Paid-Owned-Earned-MediaBranding/1007936#sthash.tiPQzvwq.dpuf

eMarketer (2016, January 27). Marketers Face Mobile Advertising Challenges. Retrieved March 30, 2017, from https://www.emarketer.com/Article/Marketers-Face-Mobile-Advertising-Challenges/1013514

Ghandour, A., Benwell, G., & Deans, K. (2010). The relationship between website metrics and the financial performance of online businesses. *Icis*.

Ghose, A., & Todri-Adamopoulos, V. (2016). Toward a Digital Attribution Model: Measuring the Impact of Display Advertising on Online Consumer Behavior. *MIS Quarterly*, *40*(4), 889–910.

Goldfarb, A. (2006). State dependence at internet portals. *Journal of Economics and Management Strategy*, *15*(2), 317–352. http://doi.org/10.1111/j.1530-9134.2006.00102.x

Graves, L., Kelly, J., & Gluck, M. (2010). Confusion online: Faulty metrics and the future of digital journalism. New York: Tow Center for Digital Journalism, Columbia University Graduate School of Journalism. Retrieved April 27, 2017, from http://www. towcenter.org/research/confusion-online-faulty-metrics-and-the-future-of-digital-journalism/

Harrison, F. (2013). Digging deeper down into the empirical generalization of brand recall: Adding owned and earned media to paid-media touchpoints. *Journal of Advertising Research*, *53*(2), 2–7. http://doi.org/10.2501/JAR-53-2-181-185

Heine, C. (2014, October 2). Mobile vs. Desktop: See Which Medium Wins in 5 Key

Comparisons. Retrieved March 30, 2017, from http://www.adweek.com/digital/mobile-vs-desktop-see-which-medium-wins-5-key-comparisons-160458

Hong, I. B. (2007). A survey of web site success metrics used by Internet-dependent organizations in Korea. *Internet Research*, *17*(3), 272–290. http://doi.org/10.1108/10662240710758920

Hu, Y. E., Du, R. E. X. Y., & Damangir, S. (2014). Decomposing the Impact of Advertising: Augmenting Sales with Online Search Data. *Journal of Marketing Research (JMR)*, *51*(3), 300–319. http://doi.org/10.1509/jmr.12.0215

Huberman, B. A., Pirolle, P. L. T., Pitkow, J. E., & Lukose, R. M. (1998). Strong Regularities in World Wide Web Surfing. *Science*, *280*(5360), 95–97. http://doi.org/10.1126/science.280.5360.95

Ilfeld, J. S., & Winer, R. S. (2002). Generating Web Site Traffic: An Empirical Analysis of Web Site Visitation Behavior. *Journal of Advertising Research*, (May 2001), 1–34.

Ingram, M. (2015, August 19). Facebook now drives more traffic to media sites than Google. Retrieved February 15, 2017, from http://fortune.com/2015/08/18/facebook-google/

Järvinen, J., & Karjaluoto, H. (2015). The use of Web analytics for digital marketing performance measurement. *Industrial Marketing Management*, *50*, 117–127. http://doi.org/10.1016/j.indmarman.2015.04.009

Järvinen, J., Töllinen, A., Karjaluoto, H., & Platzer, E. (2012). Web analytics and social media monitoring in industrial marketing - Tools for improving marketing communications measurement. In *Proceedings of the 2012 Academy of Marketing Science (AMS) Annual Conference* (pp. 477–486).

Johnson, E. J., Bellman, S., & Lohse, G. L. (2003). Cognitive Lock-In and the Power Law of Practice. *Journal of Marketing*, *67*(2), 62–75. http://doi.org/10.1509/jmkg.67.2.62.18615

Johnson, E., Moe, W., Fader, P., Bellman, S., & Lohse, G. (2004). On the Depth and Dynamics of Online Search Behavior. *Management Science*, *50*(3), 299–308. http://doi.org/10.1287/mnsc.1040.0194

Kaushik, A. (2010). Web Analytics 2.0. Indianapolis: Wiley Publishing, Inc.

Kirk, R. E. (2013). Experimental design: Procedures for the behavioural sciences (4th ed.).

Thousand Oaks, CA: Sage.

Krall, J. (2009). Using social metrics to evaluate the impact of online healthcare NECESSARY FOR GAUGING. *Journal of Communication*, *2*(4), 387–394.

Lacy, Sarah. "Web Numbers: What's Real?" Bloomberg.com. Bloomberg, 23 Oct. 2006. Web. 27 Apr. 2017.

Laerd Statistics (2017). Statistical tutorials and software guides. Retrieved from https://statistics.laerd.com/

Luo, X., & Zhang, J. (2013). How Do Consumer Buzz and Traffic in Social Media Marketing Predict the Value of the Firm? _EBSCOHost. *Journal of Management Information Systems*, *30*(2), 213–238. http://doi.org/10.2753/MIS0742-1222300208

Malhotra, N. K., & Birks, D. F. (2007). Marketing research: an applied approach (3rd European Edition). Harlow: Prentice Hall, 502-555.

Manchanda, P., Dubé, J.-P., Goh, K. Y., & Chintagunta, P. K. (2006). The Effect of Banner on Internet Advertising Purchasing. *Journal of Marketing Research*, *43*(1), 98–108. http://doi.org/10.1509/jmkr.43.1.98

Marketingterms.com (n.d.). What is Web Site Traffic? - Definition & Explanation. Retrieved March 29, 2017, from http://www.marketingterms.com/dictionary/web_site_traffic/

Mitchell, A., & Holmcomb, J. (2016). State of the News Media 2016. *Pew Research Center, State of News Media 2016 Report*, (June). http://doi.org/10.1002/ejoc.201200111

Moe, W. W. (2003). Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream. *Journal of Consumer Psychology*, *13*(1–2), 29–39. http://doi.org/10.1207/S15327663JCP13-1&2_03

Moe, W. W., & Fader, P. S. (2004). Capturing evolving visit behavior in clickstream data. *Journal of Interactive Marketing*, *18*(1), 5–19. http://doi.org/10.1002/dir.10074

Montgomery, A. L., Li, S., Srinivasan, K., & Liechty, J. C. (2004). Modeling Online Browsing and Path Analysis Using Clickstream Data. *Marketing Science*, *23*(4), 579–595. http://doi.org/10.2307/30036690

Nakatani, K., & Chuang, T. (2011). A web analytics tool selection method: an analytical hierarchy process approach. *Internet Research*, *21*(2), 171–186.

http://doi.org/10.1108/10662241111123757

Napoli P. M. (2010). Audience Evolution: New Technologies and the Transformation of Media Audiences. New York: Columbia University Press.

New York Times (2017, January 17). Journalism That Stands Apart. Retrieved March 30, 2017, from https://www.nytimes.com/projects/2020-report/

Nguyen, A. (2013). *Online news audiences: the challenges of web metrics. In: Allan, S. and Fowler-Watt, K., eds. Journalism: New Challenges*. (K. Fowler-Watt & S. Allan, Eds.)*Journalism: New Challanges*. Centre for Journalism & Communication Research, Bournemouth University. http://doi.org/10.1007/s13398-014-0173-7.2

Pakkala, H., Presser, K., & Christensen, T. (2012). Using Google Analytics to measure visitor statistics: The case of food composition websites. *International Journal of Information Management*, *32*(6), 504–512. http://doi.org/10.1016/j.ijinfomgt.2012.04.008

Panagiotelis, A., Smith, M. S., & Danaher, P. J. (2013). From Amazon to Apple: Modeling Online Retail Sales, Purchase Incidence and Visit Behavior. *Journal of Business & Economic Statistics*, *15*(November), null-null. http://doi.org/10.1080/07350015.2013.835729

Park, Y.-H., & Fader, P. S. (2004). Modeling Browsing Behavior at Multiple Websites. *Marketing Science*, *23*(3), 280–303. http://doi.org/10.1287/mksc.1040.0050

Phippen, A., Sheppard, L., Furnell, S., Phippen, a., Sheppard, L., & Furnell, S. (2004). A practical evaluation of Web analytics. *Internet Research*, *14*(4), 284–293. http://doi.org/10.1108/10662240410555306

Piwik (n.d.). Glossary. Retrieved March 29, 2017, from https://glossary.piwik.org/

Piwik (2017). *JavaScript Tracking Client: Integrate - Piwik Analytics - Developer Docs - v3*. N.p., 2017. Web. 28 Apr. 2017, from https://developer.piwik.org/guides/tracking-javascript-guide#accurately-measure-the-time-spent-on-each-page

Plaza, Beatriz, U. of the B. C. (2009). Monitoring web traffic source effectiveness with Google Analytics. *Aslib Proceedings*, *61*(5), 474–482.

Price, V., & Zaller, J. (1993). Who Gets the News? Alternative Measures of News Reception and Their Implications for Research. *The Public Opinion Quarterly*, *57*(2), 133–164.

Riihimäki, T. (2014). *Evaluating the Value of Web Metrics*. *Master's Thesis*. Aalto University School of Business.

Russell, M. G. (2010). A Call for Creativity in New Metrics for Liquid Media. *Journal of Interactive Advertising*, *9*(2), 44–61. http://doi.org/10.1080/15252019.2009.10722155

Rutz, O. J., & Bucklin, R. E. (2009). Does banner advertising affect browsing for brands? Clickstream choice model says yes, for some. *Quantitative Marketing and Economics*, *10*(2), 231–257. http://doi.org/10.1007/s11129-011-9114-3

Sawilowsky, S. S., & Blair, R. C. (1992). A more realistic look at the robustness and Type II error properties of the t test to departures from population normality. Psychological Bulletin, 111, 352-360.

Schonberg, E., Cofino, T., & Hoch, R. (2000). Measuring success. *Communications of the …*, *43*(8), 53–57. Retrieved from http://dl.acm.org/citation.cfm?id=345142

Schwarz, M. (2016, June 15). Marketers Search For Paid, Earned, Owned Balance. Retrieved April 21, 2017, from http://www.cmo.com/features/articles/2016/5/3/marketers-search-for-paid-earned-own-balance.html#gs.ZjgDvYY1

Smith, M. D., & Brynjolfsson, E. (2001). Consumer Decision-Making at an Internet Shopbot : Brand Still Matters. *The Journal of Industrial Economics*, *49*(4), 541–558. http://doi.org/10.1111/1467-6451.00162

Srinivasan, S., Rutz, O. J., & Pauwels, K. (2016). Paths to and off purchase: quantifying the impact of traditional marketing and online consumer activity. *Journal of the Academy of Marketing Science*, *44*(4), 440–453. http://doi.org/10.1007/s11747-015-0431-z

Statista (2017). Marketing budget split by paid, owned and earned media UK 2015. Retrieved April 21, 2017, from https://www.statista.com/statistics/467153/marketing-budget-split-by-paid-owned-and-earned-media-uk/

Stephen, A. T., & Galak, J. (2012). The Effects of Traditional and Social Earned Media on Sales : A Study of a Microlending Marketplace. *Journal of Marketing Research*, *XLIX*(October), 624–639. http://doi.org/10.1509/jmr.09.0401

Tandoc, E. C. (2015). Why Web Analytics Click. *Journalism Studies*, *16*(6), 782–799.

http://doi.org/10.1080/1461670X.2014.946309

Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). “Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site.” *Journal of Marketing*, *Vol. 73*(September), 90–102. http://doi.org/10.1509/jmkg.73.5.90

Turner, S. J. (2010). Website statistics 2.0: Using Google analytics to measure library website effectiveness. *Technical Services Quarterly*, *27*(3), 261–278. http://doi.org/10.1080/07317131003765910

VanNest, A. (2016, December 14). Where Is Your Site Traffic Coming From? Retrieved February 15, 2017, from https://blog.parsely.com/post/5194/referral-traffic/

Vaughan, L. (2008). A New Frontier of Informetric and Webometric Research: Mining Web Usage Data. *Collnet Journal of Scientometrics and Information Management*, *2*(2), 29–35. http://doi.org/10.1080/09737766.2008.10700851

Vaughan, L., & Yang, R. (2013). Web traffic and organization performance measures: Relationships and data sources examined. *Journal of Informetrics*, *7*(3), 699–711. http://doi.org/10.1016/j.joi.2013.04.005

Vu, H. T. (2014). The online audience as gatekeeper: The influence of reader metrics on news editorial selection. *Journalism*, *15*(8), 1094–1110. http://doi.org/10.1177/1464884913504259

Weischedel, B., & Huizingh, E. K. R. E. (2006). Website Optimization with Web Metrics: A Case Study. *Business*, *Fredericto*, 463–470. http://doi.org/10.1145/1151454.1151525

Welling, R., & White, L. (2006). Web site performance measurement: promise and reality. *Managing Service Quality*, *16*(6), 654–670. http://doi.org/10.1108/09604520610711954

Webster, J. G., Phalen, P. F., & Lichty, L. W. (2006). Ratings analysis: The theory and practice of audience research. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Web Technology Surveys (2017, January). Usage of traffic analysis tools for websites. Available at: http://w3techs.com/technologies/overview/traffic_analysis/all (Accessed 29 January 2017)

Westfall, P. H., Tobias, R. D., & Wolfinger, R. D. (2011). Multiple comparisons and multiple tests using SAS (2nd ed.). Cary, NC: SAS Publishing.

Wickens, T. D., & Keppel, G. (2004). Design and analysis: A researcher's handbook.

Wilson, R. D. (2010). Using clickstream data to enhance business-to-business web site performance. *Journal of Business & Industrial Marketing*, *25*(3), 177–187. http://doi.org/10.1109/INFOCOM.2015.7218617

Wolk, A., & Theysohn, S. (2007). Factors influencing website traffic in the paid content market. *Journal of Marketing Management*, *23*(7–8), 769–796. http://doi.org/10.1362/026725707X230036

Xie, K., & Lee, Y. (2015). Social Media and Brand Purchase : Quantifying the Effects of Exposures to Earned and Owned Social Media Activities in a Two-Stage Decision Making Model Social Media and Brand Purchase : Quantifying the Effects of Exposures to Earned and Owned Social Media. *Journal of Management Information Systems*, *1222*(October). http://doi.org/10.1080/07421222.2015.1063297

Xun, J. (2015). Return on web site visit duration: Applying web analytics data. *Journal of Direct, Data and Digital Marketing Practice*, *17*(1), 54–70. http://doi.org/10.1057/dddmp.2015.33

Yang, Y., Pan, B., & Song, H. (2013). Predicting Hotel Demand Using Destination Marketing Organization's Web Traffic Data. *Journal of Travel Research*, *53*(4), 433–447. http://doi.org/10.1177/0047287513500391

Yin, R. K. (1981). The Case Study Crisis: Some Answers. *Administrative Science Quarterly*, *26*(1), 58–65. http://doi.org/10.1177/107554708100300106

Zheng, N., Chyi, H. I., & Kaufhold, K. (2012). Capturing "Human Bandwidth": A Multidimensional Model for Measuring Attention on Web Sites. *JMM: The International Journal on Media Management*, *14*(2), 157–179. http://doi.org/10.1080/14241277.2011.619153

Appendices

Appendix 1: Chi-square test of homogeneity outputs:

Output 1: Bounce-rate

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Bounce * Owned 1, Earned 2, Paid 3	2425	100,0%	0	0,0%	2425	100,0%

Bounce * Owned 1, Earned 2, Paid 3 Crosstabulation

			Owned 1, Earned 2, Paid 3			Total
			1	2	3	
Bounce	0	Count	680_a	938_b	598_a	2216
		Expected Count	689,0	914,7	612,3	2216,0
		% within Owned 1, Earned 2, Paid 3	90,2%	93,7%	89,3%	91,4%
	1	Count	74_a	63_b	72_a	209
		Expected Count	65,0	86,3	57,7	209,0
		% within Owned 1, Earned 2, Paid 3	9,8%	6,3%	10,7%	8,6%
Total		Count	754	1001	670	2425
		Expected Count	754,0	1001,0	670,0	2425,0
		% within Owned 1, Earned 2, Paid 3	100,0%	100,0%	100,0%	100,0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	12,090[a]	2	,002
Likelihood Ratio	12,431	2	,002
Linear-by-Linear Association	,245	1	,620
N of Valid Cases	2425		

Each subscript letter denotes a subset of Owned 1, Earned 2, Paid 3 categories whose column proportions do not differ significantly from each other at the ,05 level.

Output 2: Mobile / Desktop

Mobile * Owned 1, Earned 2, Paid 3 Crosstabulation

			Owned 1, Earned 2, Paid 3			Total
			1	2	3	
Mobile	0	Count	166_a	277_b	106_c	549
		Expected Count	170,7	226,6	151,7	549,0
		% within Owned 1, Earned 2, Paid 3	22,0%	27,7%	15,8%	22,6%
	1	Count	588_a	724_b	564_c	1876
		Expected Count	583,3	774,4	518,3	1876,0
		% within Owned 1, Earned 2, Paid 3	78,0%	72,3%	84,2%	77,4%
Total		Count	754	1001	670	2425
		Expected Count	754,0	1001,0	670,0	2425,0
		% within Owned 1, Earned 2, Paid 3	100,0%	100,0%	100,0%	100,0%

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
New 1/Return 0 * Owned 1, Earned 2, Paid 3	2425	100,0%	0	0,0%	2425	100,0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	32,431[a]	2	,000
Likelihood Ratio	33,280	2	,000
Linear-by-Linear Association	6,746	1	,009
N of Valid Cases	2425		

Output 3: New / Returning user

New 1/Return 0 * Owned 1, Earned 2, Paid 3 Crosstabulation

			Owned 1, Earned 2, Paid 3			Total
			1	2	3	
New 1/Return 0	0	Count	268a	138b	155c	561
		Expected Count	174,4	231,6	155,0	561,0
		% within Owned 1, Earned 2, Paid 3	35,5%	13,8%	23,1%	23,1%
	1	Count	486a	863b	515c	1864
		Expected Count	579,6	769,4	515,0	1864,0
		% within Owned 1, Earned 2, Paid 3	64,5%	86,2%	76,9%	76,9%
Total		Count	754	1001	670	2425
		Expected Count	754,0	1001,0	670,0	2425,0
		% within Owned 1, Earned 2, Paid 3	100,0%	100,0%	100,0%	100,0%

Each subscript letter denotes a subset of Owned 1, Earned 2, Paid 3 categories whose column proportions do not differ significantly from each other at the ,05 level.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
goalConversions * Owned 1, Earned 2, Paid 3	337	100,0%	0	0,0%	337	100,0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	114,489[a]	2	,000
Likelihood Ratio	114,264	2	,000
Linear-by-Linear Association	34,631	1	,000
N of Valid Cases	2425		

Output 4: Conversions

goalConversions * Owned 1, Earned 2, Paid 3 Crosstabulation

			Owned 1, Earned 2, Paid 3			Total
			1	2	3	
goalConversions	0	Count	120_a	29_a	139_a	288
		Expected Count	126,5	25,6	135,9	288,0
		% within Owned 1, Earned 2, Paid 3	81,1%	96,7%	87,4%	85,5%
	1	Count	28_a	1_a	20_a	49
		Expected Count	21,5	4,4	23,1	49,0
		% within Owned 1, Earned 2, Paid 3	18,9%	3,3%	12,6%	14,5%
Total		Count	148	30	159	337
		Expected Count	148,0	30,0	159,0	337,0
		% within Owned 1, Earned 2, Paid 3	100,0%	100,0%	100,0%	100,0%

Appendix 2: One-way ANOVA outputs:

Output 1: No outliers removed

Descriptives

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
visitCount	Owned	754	3,57	8,807	,321	2,94	4,20	1	152
	Earned	1001	1,38	4,115	,130	1,12	1,63	1	110
	Paid	670	1,30	1,523	,059	1,18	1,41	1	29
	Total	2425	2,04	5,726	,116	1,81	2,27	1	152
visitDuration	Owned	754	126,79	281,402	10,248	106,67	146,91	0	2198
	Earned	1001	137,14	287,556	9,089	119,31	154,98	0	1889
	Paid	670	116,79	266,300	10,288	96,59	136,99	0	1798
	Total	2425	128,30	279,916	5,684	117,15	139,45	0	2198
actions	Owned	754	1,19	,739	,027	1,14	1,25	1	14
	Earned	1001	1,16	,825	,026	1,11	1,21	1	16
	Paid	670	1,12	,414	,016	1,09	1,15	1	4
	Total	2425	1,16	,706	,014	1,13	1,19	1	16

Test of Homogeneity of Variances

	Levene Statistic	df1	df2	Sig.
visitCount	100,541	2	2422	,000
visitDuration	1,816	2	2422	,163
actions	6,293	2	2422	,002

Robust Tests of Equality of Means

		Statistic[a]	df1	df2	Sig.
visitCount	Welch	24,312	2	1327,419	,000
actions	Welch	2,908	2	1550,980	,055

a. Asymptotically F distributed.

Output 2: 1% outlies removed

Descriptives

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
visitCount	Owned	728	2,44	3,311	,123	2,20	2,68	1	20
	Earned	998	1,19	,915	,029	1,13	1,24	1	15
	Paid	669	1,26	1,083	,042	1,17	1,34	1	20
	Total	2395	1,59	2,080	,042	1,50	1,67	1	20
visitDuration	Owned	734	88,52	154,472	5,702	77,33	99,72	0	1044
	Earned	970	94,93	160,891	5,166	84,79	105,07	0	1032
	Paid	654	81,54	141,062	5,516	70,71	92,37	0	1043
	Total	2358	89,22	153,641	3,164	83,02	95,43	0	1044
actions	Owned	744	1,13	,405	,015	1,10	1,16	1	3
	Earned	988	1,08	,311	,010	1,06	1,10	1	3
	Paid	667	1,11	,366	,014	1,08	1,13	1	3
	Total	2399	1,11	,358	,007	1,09	1,12	1	3

Test of Homogeneity of Variances

	Levene Statistic	df1	df2	Sig.
visitCount	305,232	2	2392	,000
visitDuration	2,040	2	2355	,130
actions	16,376	2	2396	,000

Robust Tests of Equality of Means

		Statistic[a]	df1	df2	Sig.
visitCount	Welch	49,469	2	1265,175	,000
actions	Welch	4,061	2	1422,716	,017

a. Asymptotically F distributed.

Appendix 3: One-Way ANOVA Post Hoc tests outputs

Table A 3.1: One-Way ANOVA Post Hoc tests (Multiple comparisons with 1% outliers removed)

Games-Howell

Dependent Variable	I	J	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Visit Count	Owned	Earned	**1,255***	,126	**,000**	,96	1,55
		Paid	**1,184***	,130	**,000**	,88	1,49
	Earned	Owned	-1,255*	,126	,000	-1,55	-,96
		Paid	-,071	,051	,347	-,19	,05
	Paid	Owned	-1,184*	,130	,000	-1,49	-,88
		Earned	,071	,051	,347	-,05	,19
Actions	Owned	Earned	**,050***	,018	**,014**	,01	,09
		Paid	,027	,021	,398	-,02	,07
	Earned	Owned	-,050*	,018	,014	-,09	-,01
		Paid	-,023	,017	,364	-,06	,02
	Paid	Owned	-,027	,021	,398	-,07	,02
		Earned	,023	,017	,364	-,02	,06

*. The mean difference is significant at the 0.05 level.

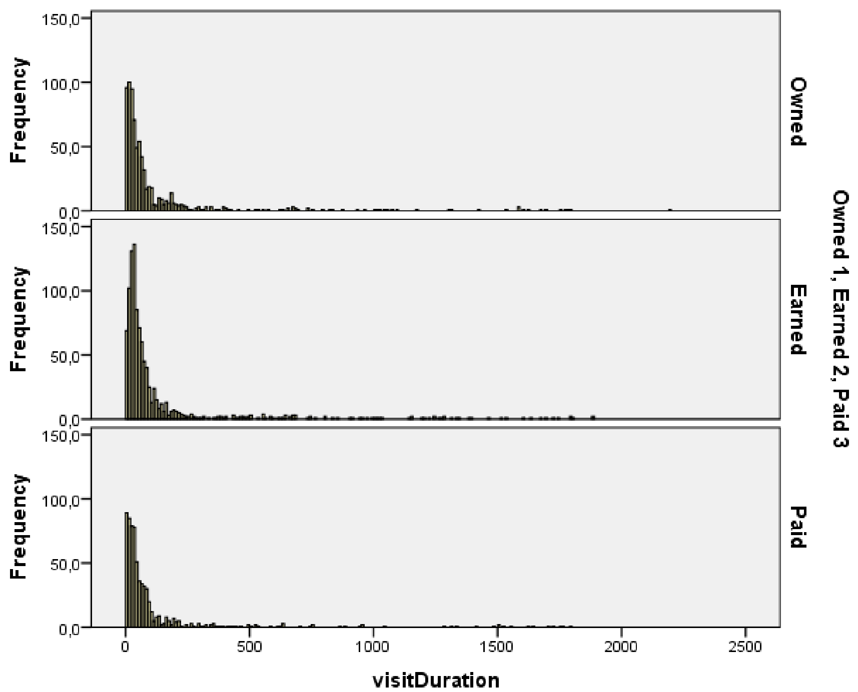
Table A3.2: One-Way ANOVA Post Hoc tests (Multiple comparisons without outliers removed)

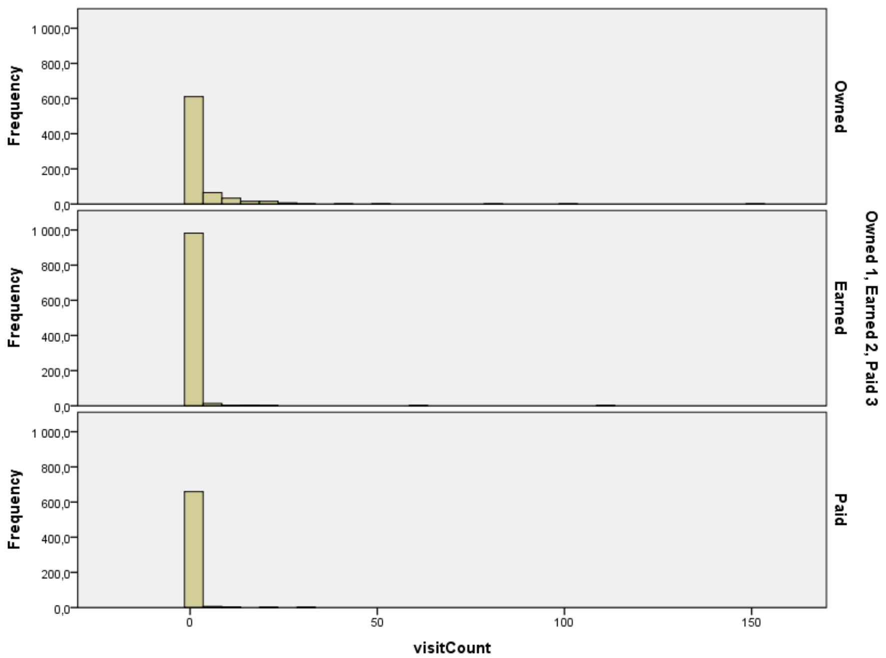
Games-Howell

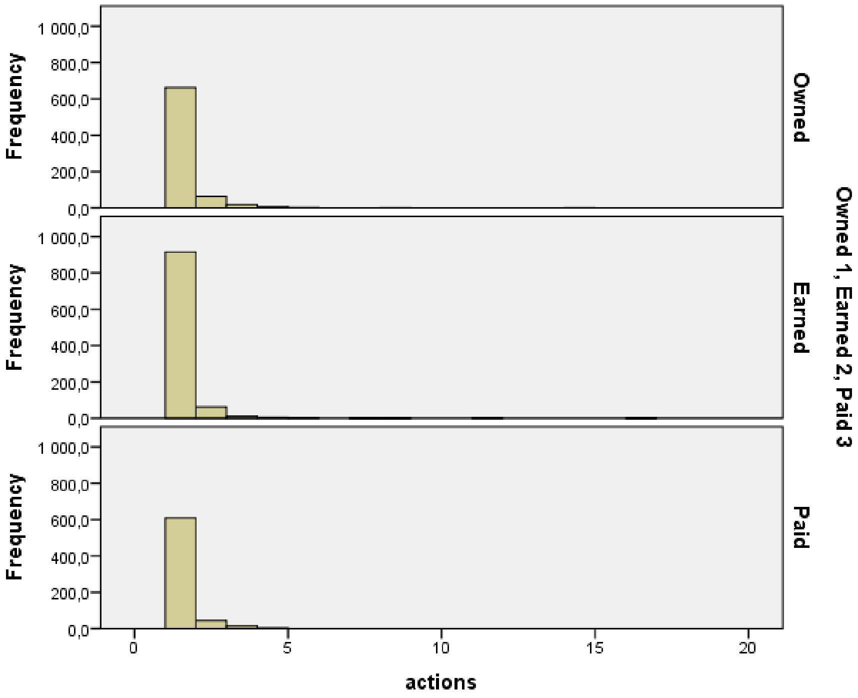
Dependent Variable	I	J	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Visit Duration	Owned	Earned	2,194*	,346	,000	1,38	3,01
		Paid	2,274*	,326	,000	1,51	3,04
	Earned	Owned	-2,194*	,346	,000	-3,01	-1,38
		Paid	,080	,143	,841	-,25	,42
	Paid	Owned	-2,274*	,326	,000	-3,04	-1,51
		Earned	-,080	,143	,841	-,42	,25

*. The mean difference is significant at the 0.05 level.

Appendix 4:Visualization of the data (Histograms)






Frequency
1 000,0
800,0
600,0
400,0
200,0
0,0
Owned
Earned
Paid
Owned 1, Earned 2, Paid 3
0
5
10
15
20
actions