Data Science in the Business Environment: Customer Analytics Case Studies in SMEs

Jing Lu
Department of Digital Technologies, University of Winchester, Winchester, UK
Lisa Cairns
Winchester Business School, University of Winchester, Winchester, UK, and
Lucy Smith
Oakhaven Hospice Trust, Lymington, UK and Winchester Business School, University of Winchester, Winchester, UK

Abstract

Purpose

A vast amount of complex data is being generated in the business environment which enables support for decision making through information processing and insight generation. A process model for data-driven decision making is proposed to provide an overarching methodology covering key stages of the business analytics life cycle. The model is then applied in two small enterprises using real customer/donor data to assist the strategic management of sales and fundraising.

Design/methodology/approach

Data science is a multi-disciplinary subject that aims to discover knowledge and insight from data while providing a bridge to data-driven decision making across businesses. This paper starts with a review of established frameworks for data science and analytics before linking with process modelling and data-driven decision making. A consolidated methodology is then described covering the key stages of exploring data, discovering insights and making decisions.

Findings

Representative case studies from a small manufacturing organisation and an independent hospice charity have been used to illustrate the application of the process model. Visual analytics have informed customer sales strategy and donor fundraising strategy through recommendations to the respective senior management teams.

Research limitations/implications

The scope of this research has focused on customer analytics in small to medium-sized enterprise through two case studies. While the aims of these organisations are rather specific, they share a commonality of purpose for their strategic development which is addressed by this paper.

Originality/value

Data science is shown to be applicable in the business environment through the proposed process model, synthesising micro- and macro- solution methodologies and allowing organisations to follow a structured procedure. Two real-world case studies have been used to highlight the value of the data-driven model in management decision making.

Keywords:

Applied data science, Business analytics, Process modelling, Insight generation, Data-driven decision making, Customer analytics, Small enterprises, Strategic management

1. Introduction

Organisations have always been using data to inform their decisions. Some new business challenges and opportunities are now occurring due to the characteristics of big data, e.g. its volume (amount of data), variety (number of types of data), velocity (speed at which data is generated and moves around) and veracity (trustworthiness of the data). The availability of big data has not only enabled data-driven decision making but also created business research questions that cannot be addressed effectively with traditional analysis methods (Delen and Zolbanin, 2018). Data Science has been described as an emerging field that integrates systematic thinking, methodologies, approaches and technologies to develop intelligence with respect to real-world problems. It can involve the collection, management, processing, analysis, visualisation and interpretation of huge amounts of data (Donoho, 2017).

Insight has been defined as advanced understanding from knowledge which is actionable from data-driven findings in order to create business value (Chang *et al*., 2009). Moreover data science can be characterised as an umbrella of techniques used to generate insights from data – with the ultimate goal to improve decision making in the business context. A prototype conceptual architecture is introduced in Figure 1 to link the key components of data science graphically. The proximity of the 6 components is significant along with their 5 first-level intersections, which form related areas. The connection between data science and the particular organisational context comes through domain-based knowledge and expertise. This creates the substantive link from data science to (e.g.) the business environment, where business transformation and value creation through data are vitally important aspects.

Data analytics is one of the key components of Data Science – it has been interpreted with specialised meaning among different well-established disciplines such as mathematics, statistics, economics, operations research, computer science and industrial engineering (INFORMS, 2019). Analytics has become a new label for evidence-based management and data-driven decision making due to the business need, its availability and affordability, and culture change within organisations (Delen, 2020). Analytics is promising to provide managers with the insight they *need* to make better and faster decisions. Businesses now have a tremendous amount of data *available*, more than they can handle, and recent technological advances in software and hardware are *affordable* even for small to medium-sized enterprise (SME). And at the organisational level, there is a *culture change* from traditional intuition-driven decision making to new age data-driven decision making.

SMEs play an important role in the continuous growth and success of national economies (Liu *et al*., 2020). However, one of the main challenges related to the use of data science alongside big data is the skills requirement for analytics. A mechanism is needed to help SMEs get started with data science and analytics (Coleman *et al*., 2016).

The following research question is motivated for the study:

RQ: How can data science be adopted in the business environment to generate insights and make impacts for SMEs at strategic and operational levels?

With the rise in the use of quantitative methods to solve organisational problems, the business analytics community has adopted a paradigm that classifies analytics in terms of descriptive, diagnostic, predictive and prescriptive (INFORMS, 2019). This paper will address the research question through the design of a higher-level process model to guide data-driven decision making with this paradigm at its core. The applicability of the model will be demonstrated through customer analytics case studies in two small enterprises.

This research is a continuation of previous studies (Lu, 2018), where real-world examples in health informatics and marketing have been used to illustrate the application of a data-driven framework for business analytics – in particular using data mining and statistical techniques, machine learning algorithms and visual analytics. Furthermore data analytic-thinking can follow a set of fundamental principles (Provost and Fawcett, 2013), four of which are especially relevant here:

Principle 1: "*Data*, and the capacity to extract useful knowledge from data, should be regarded as key strategic assets"

Principle 2: "Extracting useful knowledge from data to solve business problems can be treated systematically by following a *Process* with reasonably well-defined stages"

Principle 3: "From a large mass of data, *Information Technology* can be used to find informative descriptive attributes of entities of interest"

Principle 4: "Formulating data analytical solutions and evaluating the results involves thinking carefully about the *Context* in which they will be used"

Based on the above principles, the goals of this research study can be characterised as follows:

A. Real-world data from a small manufacturing company and an independent hospice charity will be used to assist the strategic development of the two SMEs

B. A process model for data-driven decision making will be proposed to provide an overarching methodology covering key stages of the business analytics life cycle
C. Visual analytics techniques will then be applied to the data using the model and contemporary software for descriptive, diagnostic and predictive analytics
D. Evaluation of results from the two case studies will aim to generate important insights for the SMEs with potential for: innovation and change initiatives to target customers and markets more precisely; increased revenue through more targeted approaches and funding applications.

The remainder of this paper proceeds as follows: Section 2 provides an overview of key business analytics domains from the literature as well as evaluating solution methodologies for data science and analytics. A consolidated data-driven process model is then proposed in Section 3 covering the key stages of exploring data, discovering insights and making decisions. Two real-world case studies from UK-based SMEs are used to illustrate the application of the process model in Section 4: one is from a small manufacturing organisation, to help determine customer sales strategy; the other is from an independent hospice charity, to assist its donor fundraising strategy. A brief discussion of the business innovation and transformation potential of the case studies features in Section 5 before the paper draws to a close with some concluding remarks and a pointer to future work.

2. Literature Review

2.1 Key Business Analytics Overview

The generation of the right information and insights for decision makers is a major challenge for many organisations. The challenge lies in coping with a burgeoning amount of multifarious data, analysing the data and ensuring it reaches decision makers in a timely and meaningful manner. Business analytics refers to methods and practices of organisations to collect, manage and analyse data from a variety of sources in order to enhance the understanding of business processes, operations and systems as well as create value (Kraus *et al*., 2020). It has a clear role to generate competitive advantage in organisations (Hindle *et al*., 2020). Key analytics in business can help organisations apply tools to turn data into valuable insights that enable them to better understand their customers, optimise their internal processes, and identify cost savings and growth opportunities (Marr, 2016). Figure 2 shows some representative domain-based approaches within the business categories of Customer analytics, Marketing analytics, Financial analytics, Human Resource analytics and Operational analytics alongside Core analytics areas. Examples from each of the key business analytics domains are labelled in blue text corresponding to representative studies from the literature cited below.

Starting with marketing analytics, various quantitative approaches and modelling techniques have been employed in marketing research (Manrai, 2014). The business environment has been plagued with a high level of market uncertainty and rapid change in recent times. As an example from the competitor analytics area, a study was conducted to investigate factors contributing to supply chain agility in order to enhance the competitive capabilities of organisations (Ahmed *et al*., 2019). The model developed provides some insight into the context of dynamic capabilities for policymakers and decision makers. Moving onto financial analytics, a probabilistic model has been proposed to help direct marketing firms select or eliminate customers based on their lifetime profitability – so that efforts can be focused in a more targeted manner to increase total profits (Zhang and Seetharaman, 2018). In the human resource analytics domain, by applying factor analysis, a study has suggested that HR flexibility could exist across IT companies and that adopting a long-term focus throughout may promote higher employee performance (Sekhar *et al*., 2017). In terms of operational analytics, recent research has shown that certain behaviour factors may influence the decision making of supply chain members (Wang *et. al*., 2020). A linear utility function has been used to represent these factors and how they may affect the performance of supply chain transactions.

For the core analytics areas, visual analytics has become a prominent technology for organisations to discover insights. A content analysis approach has been applied to online evaluation reviews of visual analytics platforms to identify the determinants of visual analytics adoption in organisations from the practitioner's point of view (Daradkeh, 2019). Tableau software has been ranked positively according to interactive visual exploration, data source connectivity and ease of use for content authoring.

And finally customer analytics sets up the scope of study for this paper – it refers to the processes and technologies that give organisations insight into customers and help them make key business decisions. Customer analytics is especially important due to the impact on a company's long-term strategy and business growth. In particular, sales channel analytics looks at various ways to distribute products to identify the most profitable channels and reach the highest-value customers (Marr, 2016). As another example, customer engagement analytics is the process of assessing how well (or otherwise) organisations engage customers with products or services through various interactions (Holmlunda *et. al*., 2020). Using data to advance service has been highlighted as a critical and timely research topic. The rapid increase in customer-related data provides companies with numerous service opportunities to create customer value. In order to help organisations to develop specific customer processes and advanced services, a practical framework has been designed, applied, validated and further refined by analysing relevant cases (Lim *et al*., 2019). In terms of customer analytics for SMEs, there are different managerial and technical considerations. A study has suggested that a cloud-based approach may be the most appropriate way of gaining access to big data analytics technology (Liu *et al*., 2020).

2.2 Solution Methodologies for Data Science and Analytics

The Institute for Operations Research and the Management Sciences produced their Analytics Body of Knowledge recently (INFORMS, 2019), which represents perspectives on a wide variety of analytics-related topics. INFORMS introduces the notion of micro- and macro- solution methodologies for analytics projects. Micro-methodology applies particular techniques to solve very specific aspects of a problem including methods for exploration, discovery and understanding (e.g. statistics, hypothesis testing, linear regression), data-independent methods (e.g. probability, simulation, optimisation) and data-dependent methods (e.g. classification, clustering, time series analysis, neural networks). On the other hand, macro-methodology provides the more general project path and structure. It draws on scientific research, OR/MS (Operations Research / Management Science) and CRISP-DM (Cross-Industry Standard Process for Data Mining) – see Figure 3 for their relationship.

For many years the most widely used approach for data analytics has been CRISP-DM (Provost and Fawcett, 2013). This methodology breaks up the overall task of finding patterns from data into a set of well-defined sub-tasks: business understanding, data understanding, data preparation, modelling, evaluation and deployment. Analytics professionals in many fields have found it useful for almost all types of projects and have considered CRISP-DM as the closest thing to a "standard" over the years (INFORMS, 2019). According to a KDnuggets survey CRISP-DM was the most popular methodology for analytics, data mining and data science projects – but a replacement for the now unmaintained CRISP-DM is overdue (Piatetsky, 2014).

IBM Analytics has proposed a 10-stage Foundational Methodology for Data Science (IBM, 2016) which extends CRISP-DM. It spans various technologies and approaches to form an iterative process for using data to uncover insights. The methodology has been adapted in Figure 4 to integrate the corresponding questions to consider at each stage.

The Team Data Science Process (TDSP) from Microsoft Azure is another data science methodology, which aims to deliver predictive analytics solutions and intelligent applications (Microsoft Azure). TDSP has been redrawn here in Figure 5 (left-hand side) in a linear layout for the purpose of comparing it from a business analytics perspective with other frameworks, such as SMART (Marr, 2015) and the Data Analytics Lifecycle (EMC, 2015).

The SMART model suggests a (big) data analytics project starts with strategic objectives instead of data. This will point to questions that need to be answered and reduce data requirements into manageable scopes (Marr, 2015). The Measure Metrics & Data stage includes tasks such as understanding various types and forms of data; gaining access to relevant data; clarifying metrics in order to best address previously set questions. Apply Analytics then aims to extract insights using appropriate tools and techniques that can help answer the strategic questions. The insights alone are not useful unless Results are Reported through visualisation – making sure the right people get the right information in the right format to make the right decisions. Finally, SMART uses insights gained to Transform Business through digital innovation.

The Data Analytics Lifecycle in Figure 5 (right-hand side) is designed for tackling big data problems through data science team projects (EMC, 2015). The Discovery phase assesses available resources and frames business problems as analytical challenges. Following Data Preparation, the team needs to plan methods, techniques and workflow before building and executing models. The final two phases include tasks such as conveying findings, quantifying business values and implementation in a production environment.

The frameworks reviewed here share some common elements: (1) address business problems or strategic objectives; (2) data acquisition and preparation; (3) modelling and analytics; (4) evaluation and deployment. These set up reference points for the proposed process model for data-driven decision making in the next section, regardless of the explicit starting point.

3. Process Modelling

3.1 Framework

As one of the key components for data science, data analytics can be broadly considered as a process by which a team of people helps an organisation make better decisions (the objective) through the analysis of data (the activity). Data-driven decision making refers to the practice of making decisions from data analysis, rather than just intuition. It normally starts by pulling together as much relevant data as possible, analysing that data to identify patterns that lead to insight and (hopefully) facilitating better informed decisions (INFORMS, 2019).

As noted in Section 1, Principle 2, extracting useful knowledge from data to solve business problems can be treated systematically by following a process with reasonably well-defined stages. Figure 6 represents the proposed process model for data-driven decision making which consolidates several features from the frameworks described in Section 2.2. The model synthesises micro- and macro-solution methodologies and consists of three distinctive stages – Explore Data, Discover Insights and Make Decisions – before culminating in Business Innovation and Transformation. Moreover it provides a project path and structure which organisations can adopt in a procedural manner.

3.2 Explore Data

This first stage starts with exploring available data – internal or external to the organisation – outlining each type of data needed for analysis and determining whether it exists within the organisation (or how to get it). Data may have already been brought into a common location – e.g. a data warehouse – then it is just a matter of extracting it for analysis. However, it is often the case that not all of the data needed is easy to access or ready to retrieve. It is worth planning what steps are necessary to get the data and checking whether it is feasible to do so.

The shaded area in Figure 6 indicates the other issues that should be considered during the Explore Data stage, such as data complexity (diversity and variety of data sources and types), data security (legal, ethical and policy standards for data sharing), data quality (completeness, accuracy and validity) and data transformation (converting data into a suitable format).

Descriptive analytics provides summarised information, helps understand the current state of the problem and answers the question *what* has happened in the past. Computing descriptive statistics and corresponding visualisation are the final tasks when exploring data. The dotted line in Figure 6 indicates that, after descriptive analytics, it is possible to go back to identify/define/refine the business issue and/or management problem. It may also be necessary to state hypotheses and their impact on the business outcomes, executive decisions and/or employee actions.

3.3 Discover Insights

Based on the complexity of the problem and the time available, different analytical techniques can be used during the Discover Insights stage: diagnostic analytics, predictive analytics and prescriptive analytics. Diagnostic analytics helps understand *why* an event has happened or the underlying causes for an observation. It typically tries to go deeper into a specific reason or hypothesis following the descriptive analytics outcomes. Predictive models can then be built based on the understanding gained from descriptive and diagnostic analytics. Predictive analytics is more forward looking and envisions what could happen in the future. Finally, by looking at what happened in the past, the present state and all the future possibilities, prescriptive analytics goes beyond providing recommendations to actually taking the decisions or executing the actions that are right for a particular situation. The boundaries between them are not always precise – however, when moving from descriptive to predictive and then prescriptive models, the analysis tends to become more advanced.

The Discover Insights stage also provides examples in Figure 6 of different types of techniques for analytics. For example, cluster analysis divides a set of data in a way that objects in the same group (or cluster) are more similar than those in other clusters. The technique is often used in market segmentation in order to understand differences among customers. Logistic regression is a traditional, powerful and flexible statistical process for investigating (causal) relationships that can be used in predictive analytics. Time series analysis is a forecasting technique to identify patterns (trends and/or seasonality) that can then be extrapolated into the future (Marr, 2016). Finally simulation is the imitation of the operation of a real-world process or system over time. It requires a model that represents the key characteristics or behaviours of a selective system or process.

3.4 Make Decisions

After insight generation, it is necessary to communicate the results of the analysis, making assessments that drive a decision and planning actions. Accordingly it is the Make Decisions stage where insights will be linked with actionable recommendations and an execution strategy.

Analytics produces outputs such as descriptive statistics, regression models and classification trees – these need to be translated into language that decision makers understand. This is also the time to combine the *art* of executives' instinct and experience with the *science* of data and analytics. The conclusion of the process is about reviewing the outcomes that can transform the business in terms of long-term objectives and solutions.

The process model for data-driven decision making (Figure 6) allows organisations to perform data-driven decision making in a structured way and provides a procedure to follow when working with business to solve problems. It will be further illustrated in the next section through real-world SME case studies in the customer analytics domain.

4. Case Studies

4.1 Datasets Overview

Two case studies are used to showcase the application of the data-driven process model: one is from a small manufacturing organisation, for customer sales strategy analysis; the other is from an independent hospice charity, regarding analysis of its donor fundraising strategy. Table I gives a summary of data types and sources including data volume, quality and objectives for each case study.

4.2 Small Manufacturing Organisation – Customer Sales strategy

As a fast-growing enterprise which manufactures cutting-edge systems using the latest technology, its identity and product range will remain anonymous for this case study. The company designs systems to support researchers and clinicians in uncovering new insights and making breakthrough discoveries. Its mission is to contribute to science by developing systems that enable research and provide the technology to improve people's lives. Five types of product are manufactured and distributed – although there is some crossover between product types, they are generally targeted at separate applications and customer needs.

The aim of this case study is to analyse the company's sales data for the three years from 2016 to 2018 to determine the sales approach which will most effectively increase revenue compared to previous years.

EXPLORE DATA

Primary data includes sales orders, invoices, customer records, surveys and customer feedback. The information is already in an easy-to-use format – internal sales reports and the customer database can be used to extract individual records into a dataset which can be analysed subsequently.

Firstly, the data was anonymised by removing the customer name and leaving only the customer number. The customer segmentation information was then merged into the sales reports to create one set of data. This data was cleansed to remove repairs, warranty replacements, system accessories (as a customer cannot use an accessory unless they have purchased a system) and in-house demonstration orders. The data was then split into two worksheets containing sales and loans orders before using the Tableau software package (https://www.tableau.com/).

Customer Types

Figure 7 shows that Resellers provided the highest revenue (£3.39m in 2018). The second highest revenue was from Microscope Manufacturers (£2.43m in 2018). Both groups have shown a growth rate of 109% from 2016 to 2018. The lowest revenue came from zSystem accounts which had no sales in 2018. OEM customers have had a 33% decline from £340k in 2016 to £227k in 2018.

Product Types

Product 3, as shown in Figure 8, provided the highest sales revenue (£1.7m in 2016 to £4m in 2018) and a growth rate across the 3-year period of 136%. Product 4 has had a 47% growth rate (£825k to £1.2m). The largest growth rate was for Product 5 which has increased from £85k in 2017 to £383k in 2018 (350% growth). This product was introduced in 2017 so there is no sales data for 2016. Product 2 has declined over the period showing a decrease from £337k to £268k (-20%). Sales for Product 1 have remained stable over the period.

Territory

Figure 9 shows that the highest sales revenue was provided by NEE (North East Europe, £1.3m in 2018). This territory has had a growth rate of 33% from 2016 to 2018. In 2018, the second highest revenue came from USA (£961k), but this territory has had a significantly higher growth rate since 2016 (418%). SAMCAN (South America and Canada) and UK&I (United Kingdom & Ireland) are the only territories which have decreased in sales revenue (SAMCAN £104k to £102k, UK&I £925k to £880k).

DISCOVER INSIGHTS

Three questions have been specified for this case study, as follows.

1) *Which customer type should be targeted?*

Figure 10 shows that Resellers and Microscope Manufacturers have purchased the highest numbers of products (total 2119 and 1506 respectively). They are the top customers for each product apart from Product 2, where OEM customers are the second largest purchaser (56 units). Further analysis has shown that Resellers and Microscope Manufacturers provide the highest revenue and the largest sales volume, although they did not place the largest average order size or pay the highest price. The dashboard in Figure 11 summarises further insight generated, e.g. the bottom left diagram shows the forecast trend of customer types for the next 5 years (2019-2023) in terms of their purchases.

2) *Which product should be promoted?*

To evaluate future product performance, the top right diagram from the dashboard (Figure 11) extrapolates current year-on-year sales trends into sales through to 2023. From this it can be seen that, if sales trends continue as they have, Product 3 will be by far the highest selling product and will reach approximately £10m. Product 4 and Product 5 will have similar sales of approximately £2m. However, as Product 5 was only introduced in 2017, the growth rate and future predicted sales may be falsely inflated due to part year sales in 2017. It is unknown from this 3-year data what the usual sales pattern will be once the product is established. Taking into account the results in Figures 8 and 11, Product 3 and Product 4 provided the highest sales and best predicted future performance. Although other elements such as competitor activity, market demands and new internal product releases may affect future sales, this data gives an indication of what sales potential exists for each product. Product 3 also provided the highest volume of units sold (Figure 10).

3) *Which territory should be focused on?*

To evaluate the future performance of each territory, the bottom right diagram from the dashboard (Figure 11) shows an extension of the current trends for the next 5 years. If trends continue as they are, in 2023 the USA will be the highest revenue territory (approximately £3m), NEE will be the second highest (>£2m) and APAC (Asia Pacific) will be third (>£1.5m).

As an additional piece of analysis for the sales strategy, an evaluation of the provision of loan systems to try before sale has been carried out. It shows a positive correlation between the loans and sales. When a customer has loaned more products, they are also likely to purchase more products.

MAKE DECISIONS/RECOMMENDATIONS

Following the insights discovered above, the recommended sales strategy would be to focus on promoting Product 3 and Product 4 to Resellers and Microscope Manufacturers in NEE, USA, China and APAC. This will ensure that time and resources are allocated to promoting the best-selling products to the highest performing customer type in the territories where the product is already established. Approaching the strategy is this way reduces wasted resources in trying to sell less popular products and reduces travel in between territories. Using Reseller and Microscope Manufacturers as a sales channel rather than targeting End-users directly also makes most efficient use of the sales resource.

As sales for Product 2 are in decline, this product should be made obsolete after converting the remaining customers to an alternative product. The impact of this would be small as the contribution of this product to the overall sales portfolio is minimal. Product 1 should be maintained as an available product but not heavily promoted at this stage as sales are declining in favour of Product 3 and Product 4. Facilitating loans to those customers would also be recommended.

Since the findings have been shared with the management team, there have been a number of changes within the organisation. An additional member of the Sales Team has been employed for USA so that the USAW (USA West) territory can be targeted more effectively. Enquiries have begun into opening an office in China to provide local sales and support functions. Product 2 is being phased out and customers have been advised that there is a limited availability time. A project is also underway to build closer relationships with Microscope Manufacturers, increasing loyalty while remaining aware of competitors.

4.3 Independent Hospice Charity – Donor Fundraising strategy

As an independent charity-run hospice, generating income through fundraising is essential in order to contribute to the annual running costs of the organisation. While regular donors provide the focus here, customer analytics principles will be applied in this case study. The Oakhaven Trust is an independent hospice serving a population of 150,000 people in the south of England. With only 10% of total running costs received from statutory funding, fundraising and income generation are essential to ensure the continued and effective operation of hospice services.

Regular donors is an area that could provide an increased revenue – however, defining how this is achieved and gaining insight into the current activity of regular donors needs to be further explored. To this end, the aim of the case study is to analyse data from regular financial donations in order to contribute towards future fundraising strategy. Corresponding objectives have been developed as follows:

- Establish key lines of enquiry in order to shape data analysis
- Analysis of internal structured data of regular donations using suitable software
- Effective reporting of findings using appropriate data visualisations
- Development of recommendations and a communication plan for dissemination of findings.

EXPLORE DATA

The hospice has collected data on regular donors over many years – however, this growing resource has not been used to inform future actions. This is addressed in the case study by using the Tableau software for data analysis. Insights gained will be used by the Senior Management Team to inform future fundraising strategy.

The data displayed in Figure 12 clearly shows that significant amounts of money can be generated on an annual basis through regular donors, at the very outset establishing regular donors as a worthy source of income.

DISCOVER INSIGHTS

In order to identify what the key issues are, the following two questions have been pursued within this case study.

1) *What does the data show in relation to the package ID/amount of money made per time period?*

Packages or campaigns are designed by the hospice fundraising team to encourage regular giving donations. Each package is themed in a different way – for example, DIRNPL is direct giving 'No Place Like Home', a campaign designed to generate regular donations to support the Hospice at Home service. DIREPC is direct giving 'Every Penny Counts', aimed at generating funds for the general running of the hospice in its entirety. The intention is to explore if certain packages are more popular and, if so, which characteristics they have.

Six key appeals in Figure 13 have generated the most records, i.e. the most donations. These key appeals are predominantly linked to significant clinical activity – for example nursing care, no place like home and carers support.

With regards to gift amount, again, six key appeals have generated the most value in their lifetime, therefore overall money. These are the same six appeals as above with one exception: DIR TEA (Direct giving Tea at 3). DIR TEA had fewer records but raised more money than DIR EPC. This is likely to be linked to the nature of the appeal and the amount of individual donations. One thought might be the name of the appeal could influence the amount donated – for example, every penny counts might link to more but lower-level donations. DIR TEA may be an appeal that targeted fewer people but those who could donate greater amounts.

The other appeals that have lower record numbers tend to be related to appeals not directly linked to clinical care – for example the Hospice 21st birthday appeal and the newsletter appeal. The most popular appeal is the Regular giving appeal, DIR REG, which is pitched as 'keep our hospice going'. This again could be perceived as linking directly to the clinical aspects of the hospice's work, as the public will accurately view a hospice as a clinical environment.

In relation to Figure 14, patterns of distribution can also be seen. The three packages have been selected as each demonstrates a different pattern. DIR NPL shows that people make a consistent level of donations over the years with a smaller average gift against the higher number of donors. DIR REG shows that some people donated much bigger amounts (especially in 1999) while the rest fall within a certain money range. Finally, with a relatively high average gift amount, DIR TEA indicates the distribution of donations is less evenly spread with some lower and some higher across the 13-year period.

Considerations for the hospice regarding future appeals could ask which, if any, is better? Is it better to have shorter campaigns aiming for higher donations, or longer ones with lower-level donations. It may be considering the overall activity picture of fundraising, combined with the notion ‘little acorns can grow to mighty oak sized gifts’ (More Partnership), that the aim of regular donors is little and often.

2) *What does the data show in relation to postcode – which provide the most regular donations?*

This may provide insight into areas that could benefit from more concentrated fundraising attention, which is also a key question for future strategy as the hospice has recently extended its geographical catchment area.

This information in Table II clearly shows that the three key postcode areas relate to the three areas in hospice catchment that are most densely populated: BH25, SO41 and SO45. It is perhaps surprising that SO40 has a similar number of constituents supporting with regular donations as SO43. This is because a large part of SO40 was not even part of the hospice catchment area until April 2018. Also, SO43 could be considered an area of wealth that accesses the hospice facilities, therefore could be a significant target for regular donation appeals in the future. It would of course be pertinent to look at these results in the broader context of all fundraising activity however. It may be that this wealthy area contributes more significantly in other fundraising ways.

The dashboard in Figure 15 is represented for several purposes. The first is to show that significant donations from individual constituents occur in two of the three main postcode areas (BH25 and SO41). This insight may well open potential doors for further fundraising conversations with these individuals. For example, it is clear to see that tens of thousands of pounds are donated annually – therefore this funding may be considered to support specific developments – and also small acorns can grow to larger donations, so conversations about larger donations and the benefit of these may be appropriate to consider in line with a major donor mindset.

Secondly, it is pertinent to note that Figure 15 shows donations over £5,000 per year from individual donors in BH25 and SO41. The same approach has been applied to postcode area SO45 which shows donations over £500 from individual donors which rise up to £1,200 across the years. This is arguably a reflection of the postcode: BH25 and SO41 could be considered areas with pockets of extreme wealth, whereas SO45 has a different socio-economic demographic – the amounts of money donated on a regular basis therefore differ significantly. Such an insight could help the appropriate targeting of fundraising and regular donation appeals in the future.

Finally, the box plots take the three main areas and identify within them the specific postcodes where the largest donations have originated. The postcode BH25 5RW shows a positive skew, indicating that most money donated was around the £8,000 - £9,000 mark. Some larger donations are apparent however, hence the positive skew and the evident outlier at £10,000. SO41 6AS suggests a bell curve distribution, indicating that donations are more evenly spread. The levels of donations are also evident over a much wider range, starting at just £200 through to just under £10,000.

Key Insights Gained

- Direct giving appeals linked to obvious clinical activity appear to be most popular
- Three key postcodes, the most densely populated areas within the hospice catchment, generate the most regular giving donations (BH25, SO41, SO45)
- Out of hospice catchment postcodes also generate significant regular donations
- SO40 and SO43 postcodes are currently similar in constituent numbers
- Certain constituents regularly donate significant amounts of money
- BH25 and SO41 postcodes show higher amounts of money donated regularly – in SO45 the amount of money regularly donated is significantly less.

MAKE DECISIONS/RECOMMENDATIONS

In order to inform future regular donor fundraising strategy, it is reasonable to make the following predictions:

- Regular donations will increase significantly in the SO40 postcode with increased awareness of the hospice and current campaigns
- SO40 has a similar demographic to SO45, therefore regular donations will be of lower sums than other postcode areas, as noted with SO45
- SO43 is an untapped resource for significant regular donations – appropriate targeting and marketing will increase numbers and amounts raised through regular donations
- Clearly linking campaigns to familiar clinical activity will enhance the number of regular donations.

Now insights have been gained and predictions made, the final step is to communicate the findings effectively. The plan allows for an initial discussion with the Senior Management Team at the hospice and then, with agreement, discussion at a Board of Trustees meeting subsequently – the purpose being to raise awareness of the benefits of data insight in advantageously shaping the actions and progression of the hospice.

In conclusion, five key recommendations include:

- To link regular giving appeals directly to clinical care within the hospice – within this to consider branding such that the clinical element is explicitly clear to the public
- To target regular giving appeals to different postcode areas within catchment to appeal to different demographics
- To target certain postcodes where constituents who regularly donate are low in number
- To maximise regular donations outside the hospice catchment area – through alumni type opportunities for example
- To use 'SMART Board' thinking to identify data needs from an organisational perspective.

5. Discussion

The data-driven process model proposed in this paper is geared to strategic development in SMEs. The model has been applied in two small enterprises to help determine their respective strategies for customer sales and donor fundraising. From the perspective of business innovation and transformation, the potential impact on the organisations in each case study is summarised below.

Small Manufacturing Organisation – Customer Sales strategy

Using a data-driven model, innovation and change initiatives can be developed to target customers and markets with greater precision and, as a result, with greater success. The insight provided by the analysis in Section 4.2 has given the ability to assess the potential impact of changes to existing products, such as reducing product ranges or product enhancement innovations. The data has been used in conjunction with market research and customer interviews to develop a new product pipeline to further meet the needs of existing customers and target new customers. The forecasted sales have then been fed into the analysis to view the potential impact.

From the perspective of change management, a data-driven model can significantly improve buy-in to proposed changes and increase the likelihood of project success. As buy-in to a project is a significant factor influencing change management, the ability to use data in this way is important. The management team and the company directors have limited resources to work with and must decide where these are used most effectively. The use of data insight to demonstrate the predicted post-change sales revenue is an extremely effective way to gain buy-in at the start of a project as well as demonstrate success at the post-launch review of a product.

Independent Hospice Charity – Donor Fundraising strategy

The first of several transformative opportunities identified is to work in a more targeted, streamlined or 'Lean' way. This would focus fundraising activity based on knowledge and fact derived from the type of data analysis in Section 4.3, rather than broader more general approaches which are currently the norm. The outcome of an approach such as this could be increased income for the hospice without increasing the workload of the current fundraising team. A more targeted approach may lead to an increase in fundraising team capacity, which in turn could be used in alternative ways.

There is also the potential to increase revenue into the hospice through targeted funding applications. Such applications are either aimed at trusts with a philanthropic approach or for monies associated with specific projects of relevance. It is a current weakness of the hospice that, when compiling applications of this nature, the lack of robust supporting data and evidence is a significant disadvantage. With the aid of a data-driven model, specific supporting data and information may well enhance the quality of such applications, in turn enhancing this stream of revenue generation.

Finally the development of specific roles in relation to data analysis represents another opportunity, the benefit of which would be to further inform all areas of the hospice. For example, through data analytics, clinical services may be shaped and refined in line with robustly identified local need, enhancing care and care experience for patients at end of life and their families. As a fundamental principle of the modern hospice movement, any robust and reliable information that helps shape clinical services in the most skill, time and cost appropriate way is arguably essential in the best care provision.

6. Conclusion

Data science projects and indeed research can be conducted via problem, goal and data-driven processes (Cao, 2017), where the ultimate objective is to improve decision making. Data-driven approaches require understanding of the data characteristics, complexity and challenges – in the context of the nature of the business issues and data availability. A data-driven organisation can gain competitive advantage through investment in high-quality data and sophisticated analysis techniques, improving its decision making capabilities. Although there are different perceptions about the nature and scope of data analytics, there is general agreement that it involves data-driven decision making (Delen and Zolbanin, 2018). Additionally, data analytics has been viewed as a *process* that takes various forms of data as input and generates value for the organisation as output. Below is a summary of the contribution and novelty of this research study into data science in the business environment:

Scientific Value – Process Model

Based on a review of established frameworks for data science and analytics, a consolidated process model for data-driven decision making is proposed in this paper to provide an overarching methodology covering key stages of the business analytics life cycle. The model synthesises micro- and macro-solution methodologies – with the descriptive, diagnostic, predictive and prescriptive analytical paradigm at its core – allowing organisations to follow a structured data-driven procedure.

Applicability – SMEs and Data Visualisation

The process model has been applied in two small enterprises using real customer/donor data to assist the strategic management of sales and fundraising. These case studies have demonstrated the feasibility of the proposed model and its applicability in the business environment, in particular the customer analytics domain.

Tableau is one of the leading visual analytics software packages with customers ranging from small to large organisations across all industries (Ryan, 2018; Loth, 2019). Using Tableau software with the real-world SME data, some interesting visualisations and valuable results have been discovered through descriptive, diagnostic and predictive analytics techniques.

Organisational Context – Business Intelligence

The connection between data science and the particular organisational context comes through domain-based knowledge and expertise. Departmental managers from the manufacturing company and the independent charity have been involved with this research. Their respective expertise has enabled interpretation of the insights from the analytical process. Moreover their domain knowledge has facilitated the assessment of business value across the organisations, improving strategic planning and enhancing the potential for business innovation.

Future work could deepen the findings here by exploring the application of prescriptive analytics, which is often considered as the ultimate step towards increasing data analytics maturity, leading to optimised decision making ahead of time for business performance improvement (Lepeniotia *et al*., 2020). It is also worth adding that this research could be widened to other key business analytics areas, subject to the availability of suitable real-world data.

References

Ahmed, W., Najmi, A., Mustafa, Y. and Khan, A. (2019), "Developing model to analyze factors affecting firms' agility and competitive capability: A case of a volatile market", *Journal of Modelling in Management*, Vol. 14 No. 2, pp. 476-491.

Cao, L.B. (2017), "Data science: A comprehensive overview", *ACM Computing Surveys*, Vol. 50 No. 3, pp. 43:1-43:42.

Chang, R., Ziemkiewicz, C., Green, T.M. and Ribarsky, W. (2009), "Defining insights for visual analytics", *Visualization Viewpoints*, IEEE Computer Society, Vol. 29, pp. 14-17.

Coleman, S., Göb, R., Manco, G., Pievatolo, A., Tort-Martorell, X. and Reis, M.S. (2016), "How can SMEs benefit from big data? Challenges and a path forward", *Quality and Reliability Engineering International*, Vol. 32, pp. 2151-2164.

Daradkeh, M. (2019), "Determinants of visual analytics adoption in organizations: Knowledge discovery through content analysis of online evaluation reviews", *Information Technology & People*, Vol. 32 No. 3, pp. 668-695.

Delen, D. (2020), *Prescriptive Analytics: The final frontier for evidence-based management and optimal decision making*, Pearson.

Delen, D. and Zolbanin, H.M. (2018), "The analytics paradigm in business research", *Journal of Business Research*, Vol. 90, pp. 186-195.

Donoho, D. (2017), "50 Years of data science", *Journal of Computational and Graphical Statistics*, Vol. 26 No. 4, pp. 745-766.

EMC Education Services (2015), *Data Science and Big Data Analytics: Discovering, analyzing, visualizing and presenting data*, Wiley.

Hindle, G., Kunc, M., Mortensen, M., Oztekin, A. and Vidgen, R. (2020), "Business analytics: Defining the field and identifying a research agenda", *European Journal of Operational Research*, Vol. 281 No. 3, pp. 483-490.

Holmlunda, M., Vaerenbergh, Y.V., Ciuchita, R., Ravald, A., Sarantopoulos, P., Ordenese, F.V. and Zaki, M. (2020), "Customer experience management in the age of big data analytics: A strategic framework", *Journal of Business Research*, available online 6 February 2020, in Press.

IBM Analytics White Paper (2016), "Foundational methodology for data science", https://tdwi.org/~/media/64511A895D86457E964174EDC5C4C7B1.PDF (accessed 26 February 2020).

INFORMS (2019), *Analytics Body of Knowledge*, Hoboken, NJ: John Wiley & Sons.

Kraus, M., Feuerriegel, S. and Oztekin, A. (2020), "Deep learning in business analytics and operations research: Models, applications and managerial implications", *European Journal of Operational Research*, Vol. 281 No. 3, pp. 628-641.

Lepeniotia, K., Bousdekis, A., Apostolou, D. and Mentzas, G. (2020), "Prescriptive analytics: Literature review and research challenges*", International Journal of Information Management*, Vol. 50, pp. 57-70.

Lim, C., Kim, M.J., Kim, K.H., Kim, K.J. and Maglio, P. (2019), "Customer process management: A framework for using customer-related data to create customer value", *Journal of Service Management*, Vol. 30 No. 1, pp. 105-131.

Liu, Y., Soroka, A., Han, L.X., Jian, J. and Tang, M. (2020), "Cloud-based big data analytics for customer insight-driven design innovation in SMEs", *International Journal of Information Management*, Vol. 51.

Loth, A. (2019), *Visual Analytics with Tableau*, John Wiley & Sons, Inc.

Lu, J. (2018), “A data-driven framework for business analytics in the context of big data”, In: A. Benczr et al. (Eds.) *New Trends in Databases and Information Systems. Communications in Computer and Information Science*, Vol. 909, Springer Cham, pp. 339-351.

Manrai, A. (2014), “Quantitative approaches and modeling in marketing research”, *Journal of Modelling in Management*, Vol. 9 No. 3.

Marr, B. (2015), *Big Data: Using SMART big data, analytics and metrics to make better decisions and improve performance*, Chichester: John Wiley & Sons.

Marr, B. (2016), *Key Business Analytics: The 60+ business analysis tools every manager needs to know*, Harlow, Pearson.

Microsoft Azure, Team Data Science Process Documentation. https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/ (accessed 21 August 2019).

More Partnership, The truth about regular giving: 10 fundraising proverbs compiled from 250 million lines of data. https://www.morepartnership.com/library/The%20truth%20about%20regular%20giving.pdf. Dundee: More Partnership Ltd (accessed 22 August 2019).

Piatetsky, G. (2014), “CRISP-DM, still the top methodology for analytics, data mining, or data science projects”, KDnuggets News 28 October 2014. https://www.kdnuggets. com/2014/10/crisp-dm-top-methodology-analytics-data-mining-datascience-projects.html.

Provost, F. and Fawcett, T. (2013), *Data Science for Business: What you need to know about data mining and data-analytic thinking*, O'Reilly Media, California.

Ryan, L. (2018), *Visual Data Storytelling with Tableau*, Addison-Wesley Professional.

Sekhar, C., Patwardhan, M. and Vyas, V. (2017), “Causal modelling of HR flexibility and firm performance in Indian IT industries”, *Journal of Modelling in Management*, Vol. 12 No. 4, pp. 631-651.

Wang, N., Gu, J.B., Gou, Q.L. and Yue, J.F. (2020), *Supply Chain Contracting with Linear Utility Function*, IGI Global.

Zhang, Q. and Seetharaman, P. (2018), “Assessing lifetime profitability of customers with purchasing cycles”, *Marketing Intelligence & Planning*, Vol. 36 No. 2, pp. 276-289.

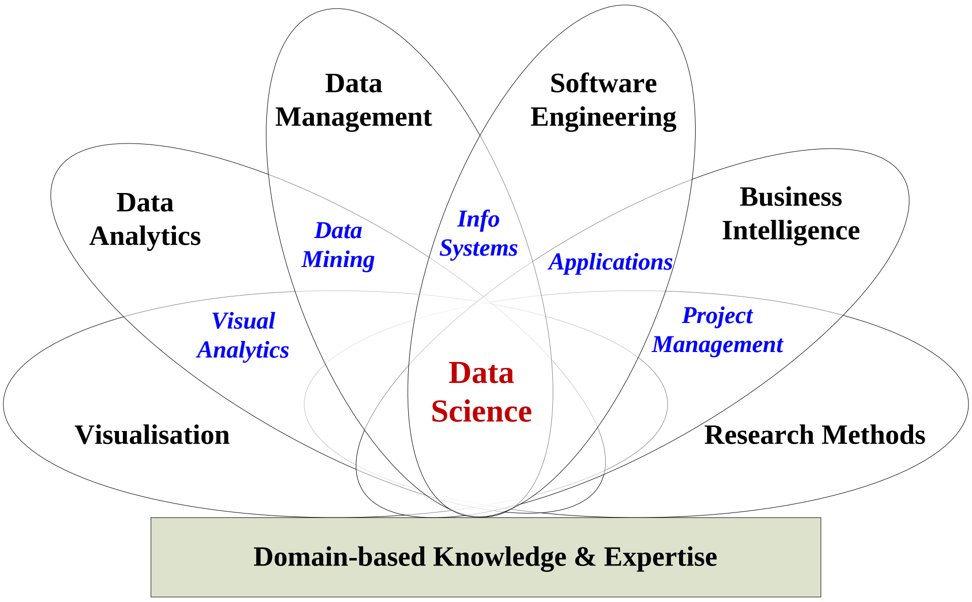

Figure 1. A prototype conceptual architecture for data science

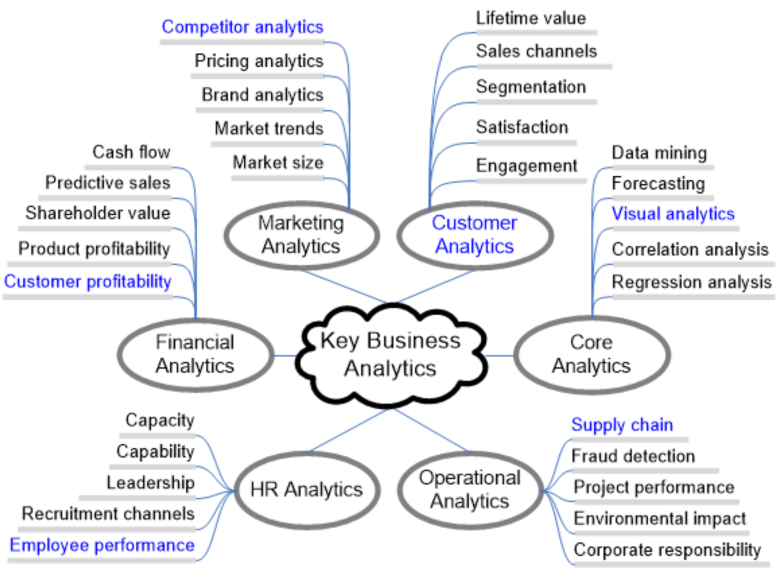

Figure 2. Domain-based business analytics

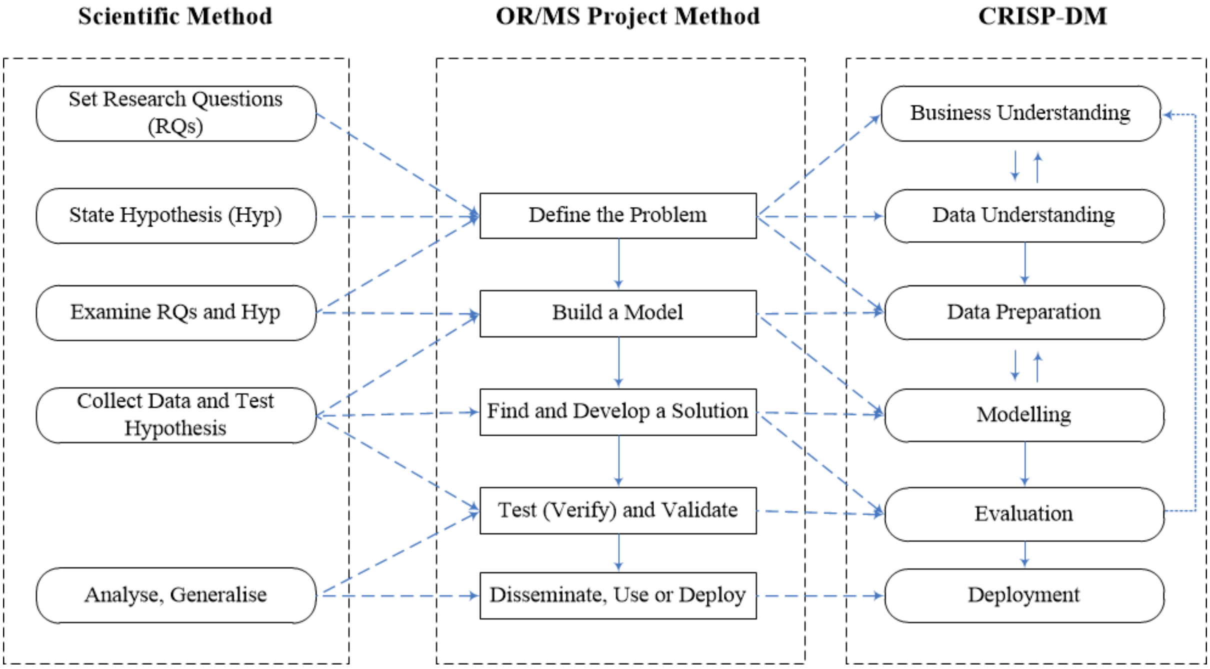

Figure 3. Relationship among macro-methodologies

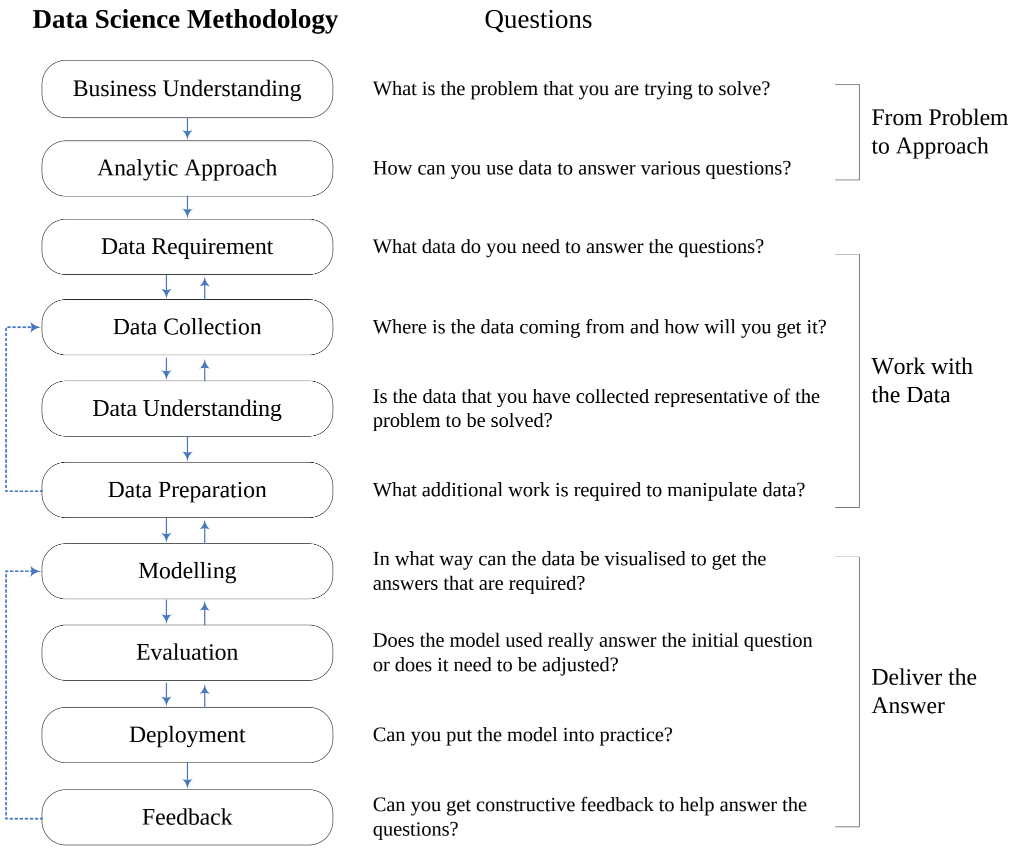

Figure 4. Adapted foundational methodology for data science

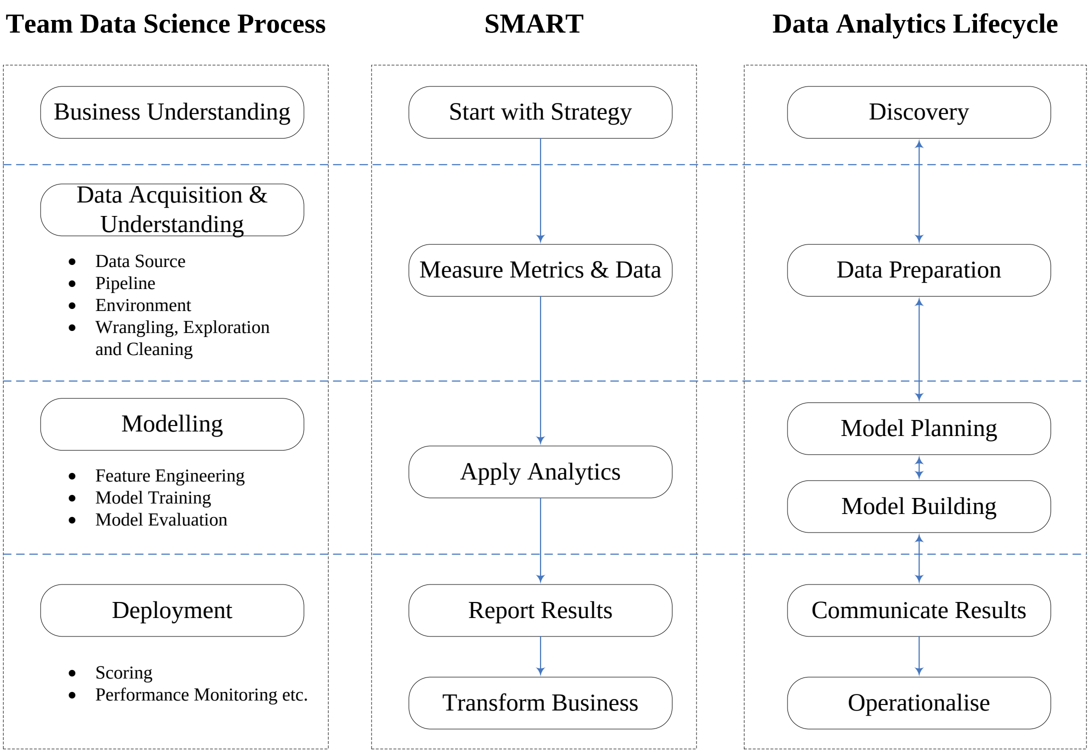

Figure 5. Comparison of three data science/analytics process models

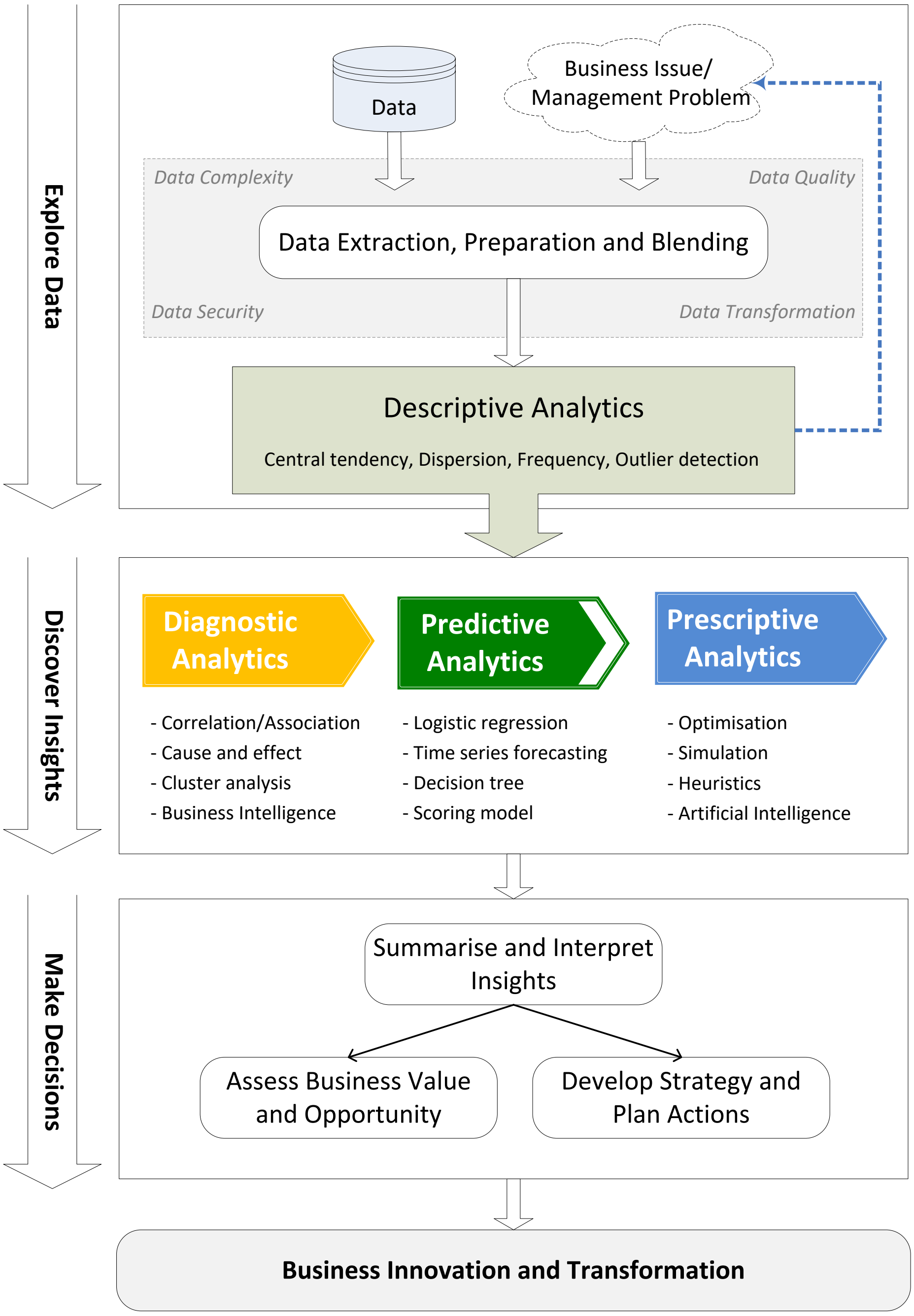

Figure 6. A process model for data-driven decision making

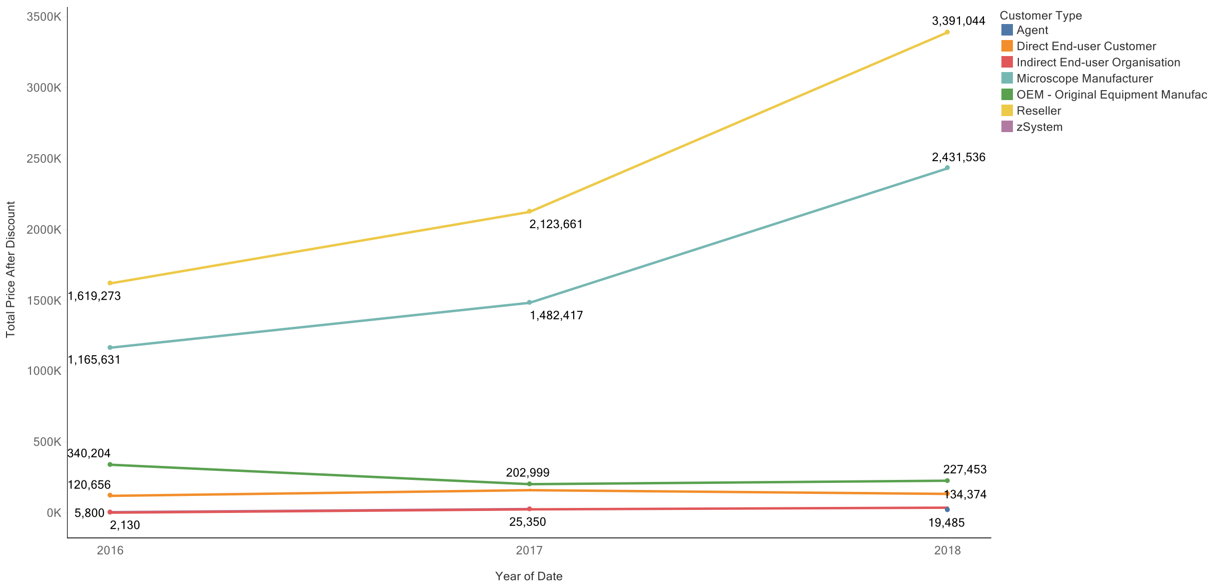

Figure 7. Year-on-year sales by customer type

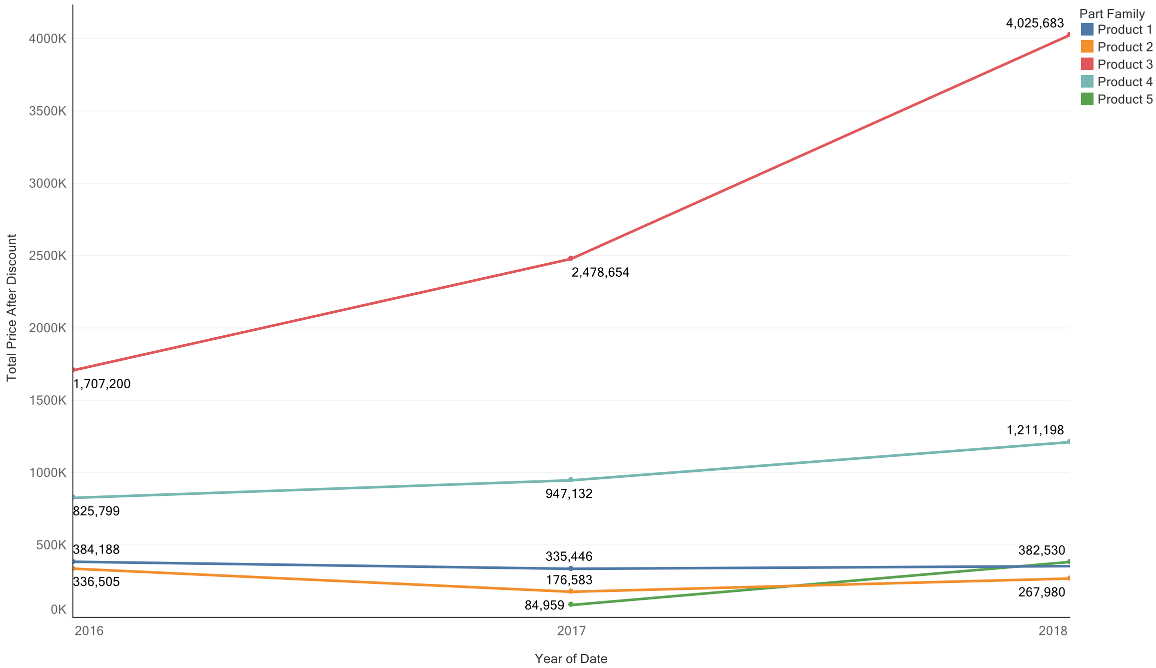

Figure 8. Year-on-year sales by part

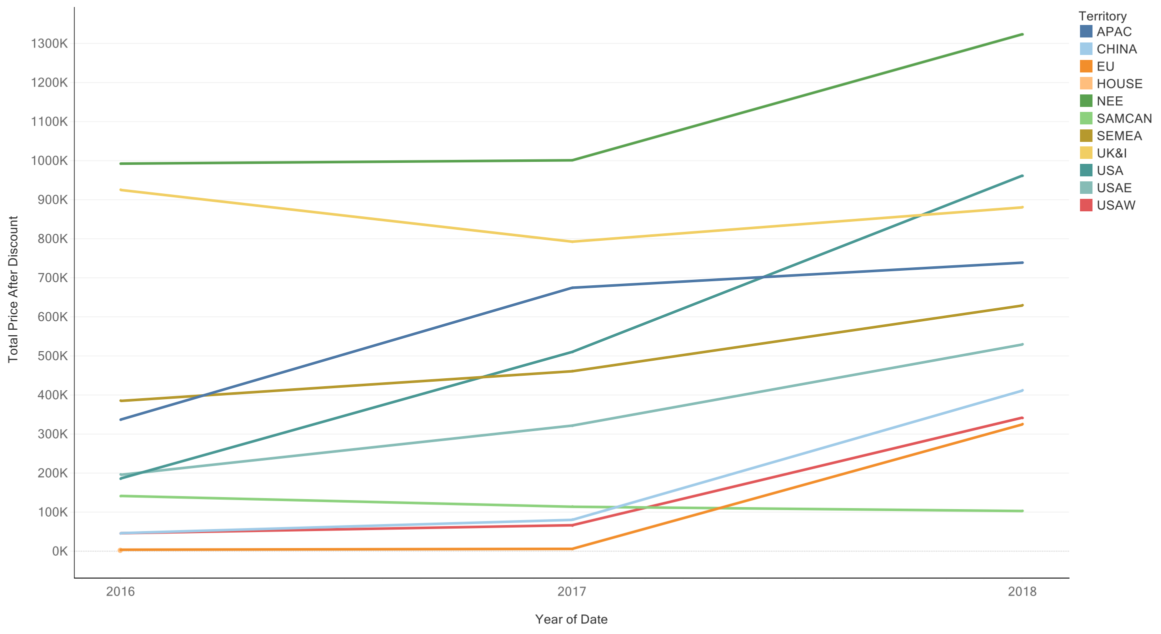

Figure 9. Year-on-year sales by territory

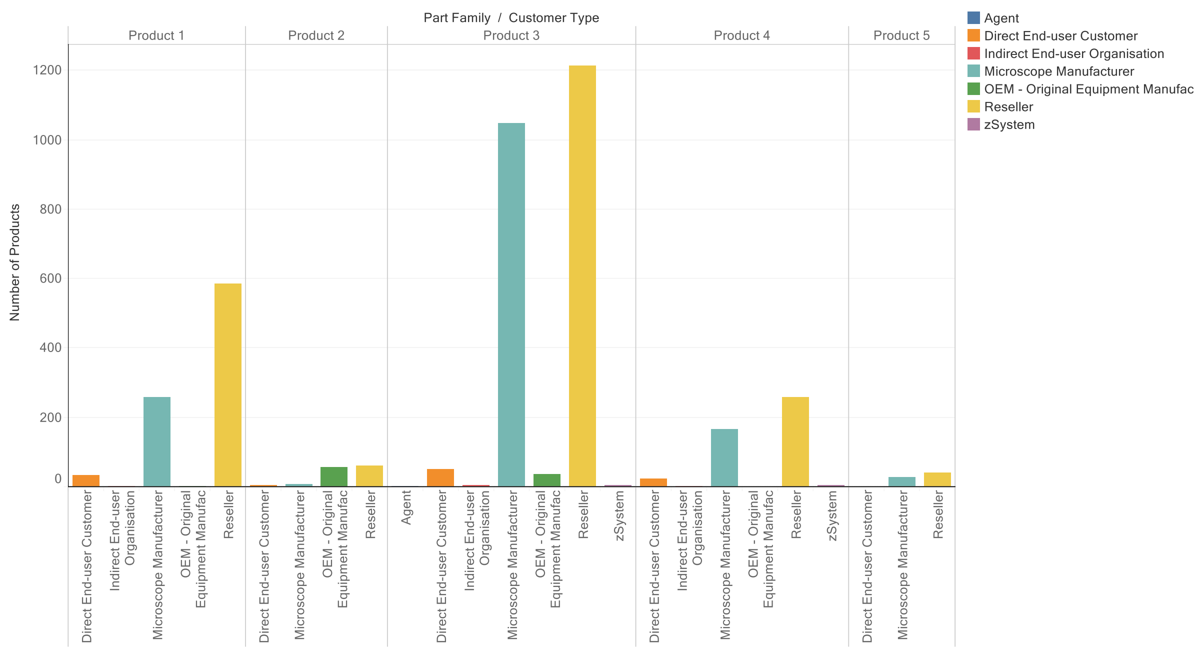

Figure 10. Number of parts by customer type

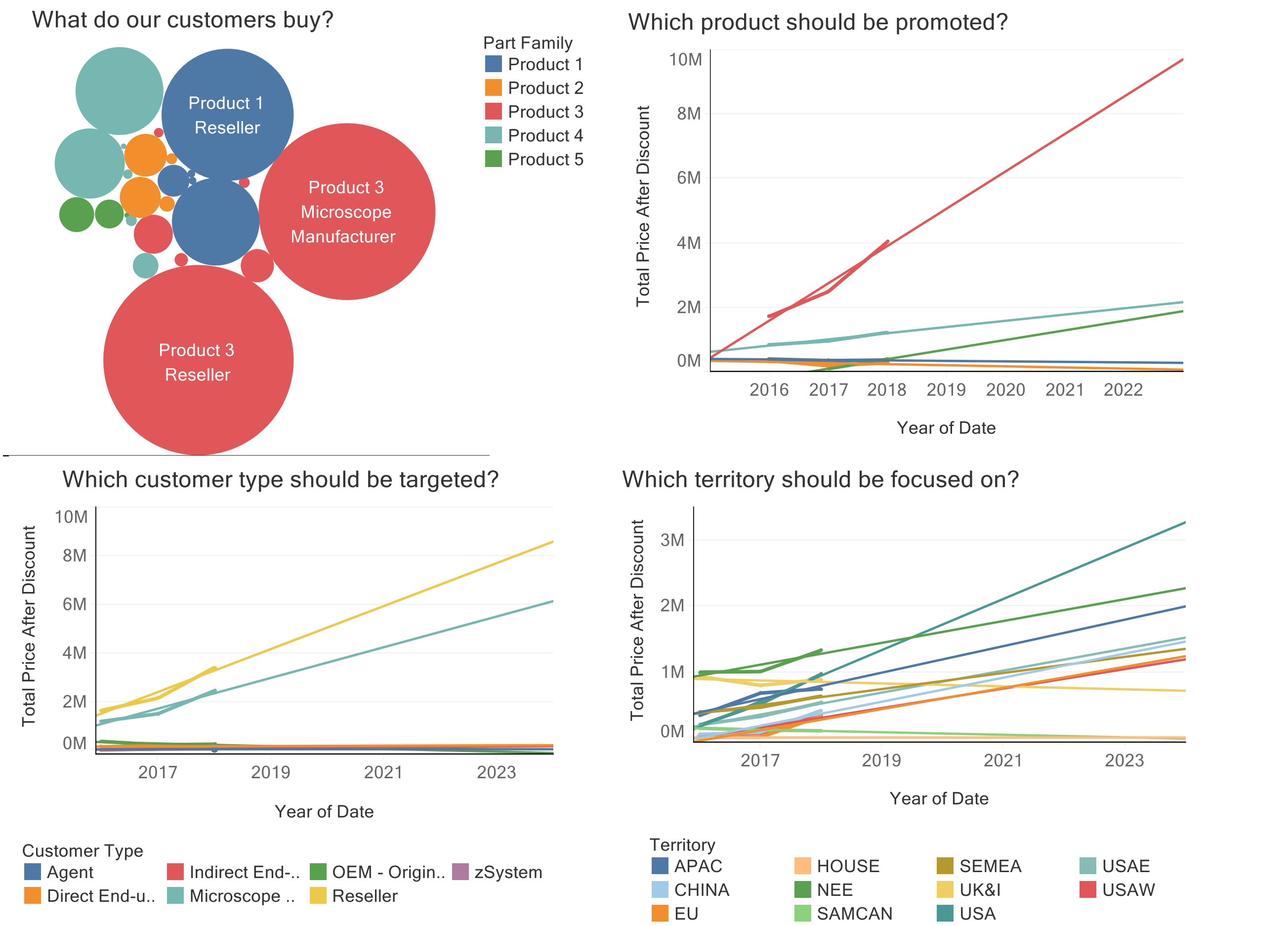

Figure 11. Tableau Dashboard for insight generation

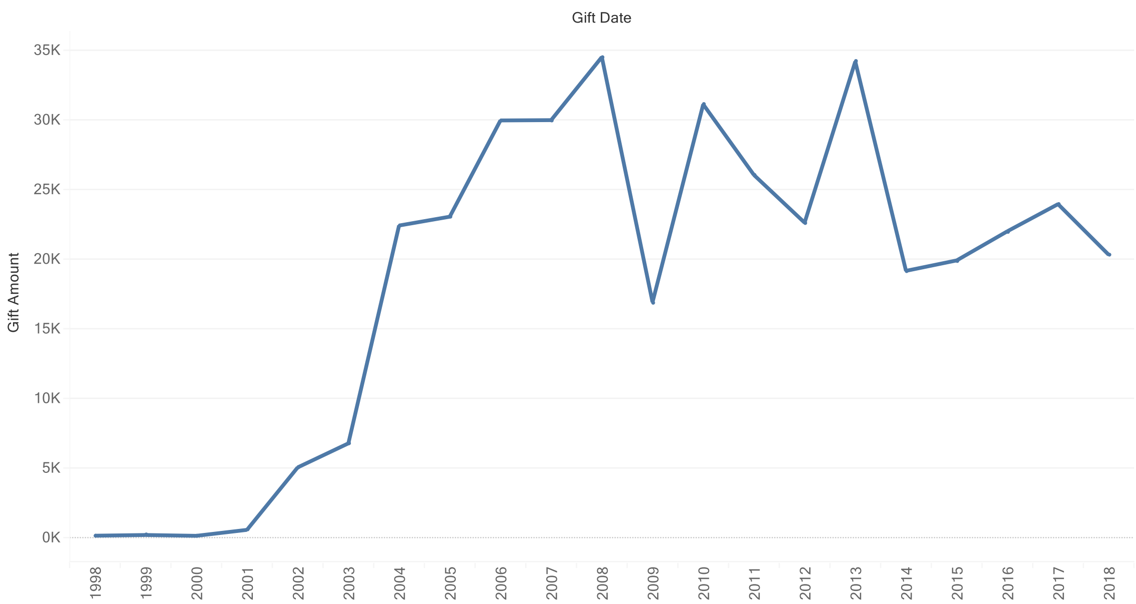

Figure 12. Amount of money raised through regular donations per year

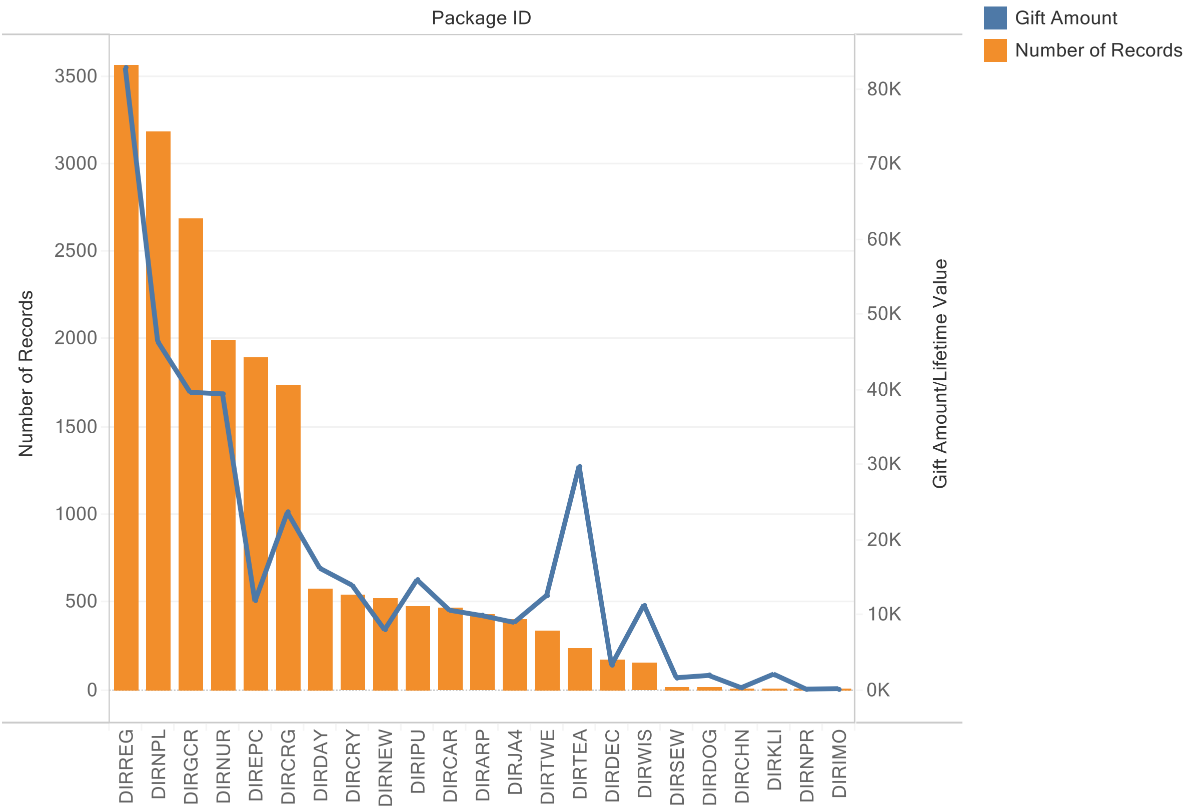

Figure 13. Number of records and gift amount per package

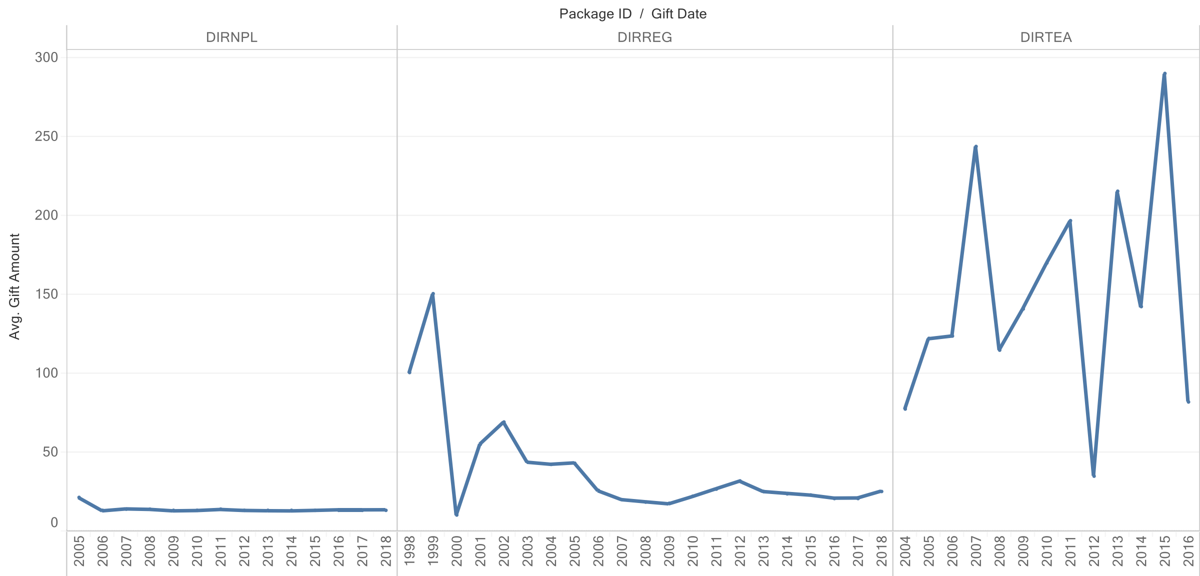

Figure 14. Gift amount and gift date for three specific packages

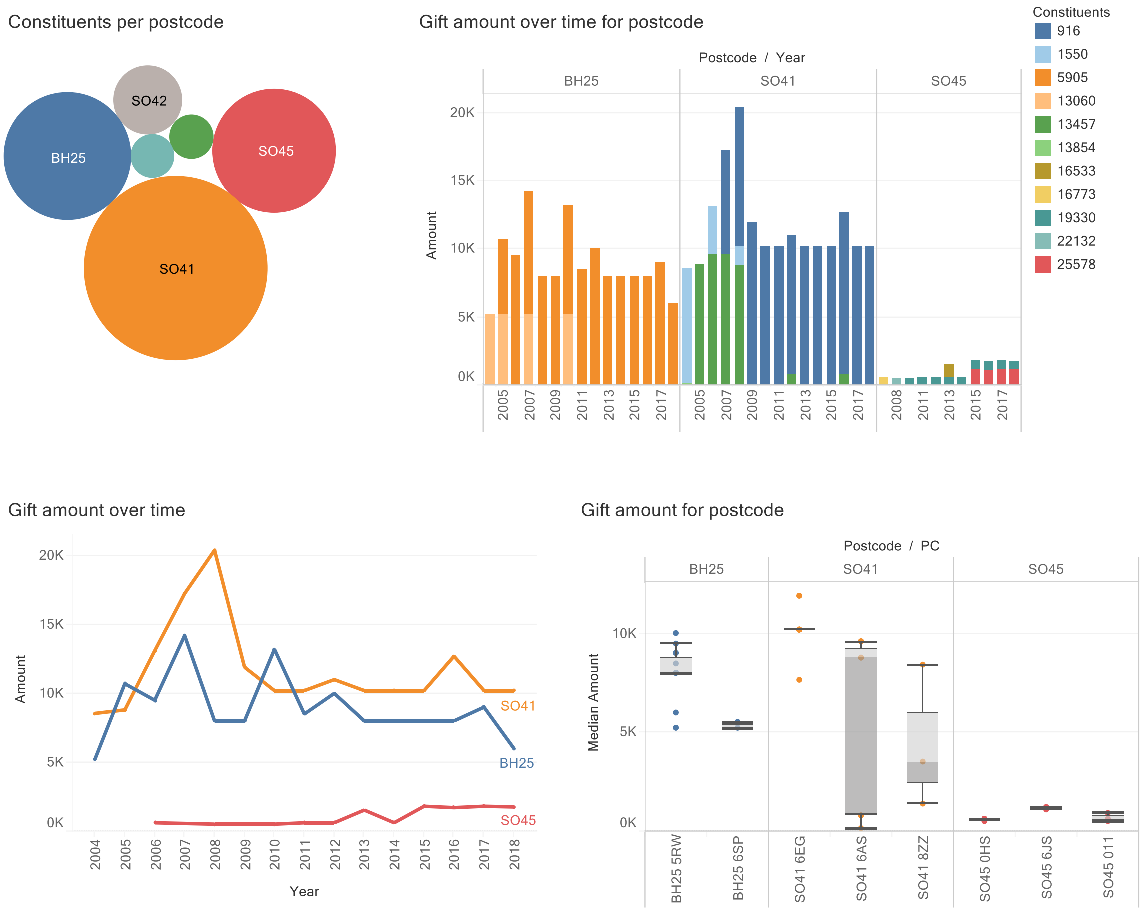

Figure 15. Tableau Dashboard based on postcode areas

Table I: Datasets for the two case studies

Nature of Business	Small Manufacturing Organisation – Customer Sales strategy	Independent Hospice Charity – Donor Fundraising strategy
Dataset Name	Company Sales Data 2016-2018	Regular Monetary Donations 1998-2018
Data Description	Sales of Technology (customers, products, territories)	Information regarding regular Donations (value, constituents, postcodes, frequency of donations, linkage with appeal/package)
Ownership	Internal Company data	Internal Hospice data
Data Volume	Sales (15 variables, 3896 rows) Loans (14 variables, 937 rows)	13 variables, 19573 rows
Data Quality	No data missing	Two columns are not filled in at all – minimal null entries in other data columns
Objectives	Target customers Promote products Sales channels Focus on territories	Popular packages Postcode segmentation Lifetime value Preferred form of donation

Table II: Summary of constituents and postcodes

Postcode area	No. of constituents within the postcode area	No. of individual postcodes within the postcode area
BH25	315	231
SO40	37	34
SO41	654	415
SO42	92	63
SO43	38	34
SO45	296	225
Other/outside of hospice catchment area	269	250