# A Conceptual Framework for the Direct Marketing Process using Business Intelligence

**A thesis submitted for the degree of Doctor of Philosophy**

**By**

**Adel Flici**

**Brunel Business School**
**Brunel University**

**July 2011**

# Abstract

Direct marketing is becoming a key strategy for organisations to develop and maintain strong customer relationships. This method targets specific customers with personalised advertising and promotional campaigns in order to help organisations increase campaign responses and to get a higher return on their investments. There are, however, many issues related to direct marketing, ranging from the highly technical to the more organisational and managerial aspects. This research focuses on the organisational and managerial issues of the direct marketing process and investigates the stages, activities and technologies required to effectively execute direct marketing.

The direct marketing process integrates a complex collection of marketing concepts and business analytics principles, which form an entirely ‘self-contained’ choice for organisations. This makes direct marketing a significantly difficult process to perform. As a result, many scholars have attempted to tackle the complexity of executing the direct marketing process. However, most of their research efforts did not consider an integrated information system platform capable of effectively supporting the direct marketing process. This research attempts to address the above issues by developing a conceptual framework for the Direct Marketing Process with Business Intelligence (DMP-BI). The conceptual framework is developed using the identified marketing concepts and business analytics principles for the direct marketing process. It also proposes Business Intelligence (BI) as an integrated information system platform to effectively execute the direct marketing process.

In order to evaluate and illustrate the practicality and impact of the DMP-BI framework, this thesis adopts a case study approach. Three case studies have been carried out in different industries including retailing, telecommunication and higher education. The aim of the case studies is also to demonstrate the usage of the DMP-BI framework within an organisational context. Based on the case studies’ findings, this thesis compares the DMP-BI framework with existing rival methodologies. The comparisons provide clear indications of the DMP-BI framework’s benefits over existing rival methodologies.

# Contents

# List of Tables

# List of Figures

# Dedication

This thesis is dedicated to my amazing Mother, Fatima-Zohra, my three beautiful Sisters (Nesrine, Nardjes and Sara), my fabulous Brother (Boualem), my wonderful Grandmother (Yamina) and my dearly departed Father, Grandfather and Uncle, Lâadi Flici, Boukhari Touil, and Mohamed Touil – with love and gratitude.

# Acknowledgement

I would like to express my gratitude to my supervisor Dr Kevin J. Lü for his guidance, advice and support during my PhD studies. I am also grateful to Dr Tariq Khan whose comments and feedbacks on my research are appreciated.

I would like also to thank all the members of the Business School at Brunel University, particularly, the IT team for their support throughout this journey. My thanks go to the staff that helped me collect the data to complete my case studies. A special thank you to all my friends, especially, Mohamed Kherachi, Atmane Benkara, Zoheir Sahel, Mouloud Ameni and Sharif Abbasi for their moral support. I also would like to thank the PhD students in the Business School and the Learning Technology Team at Brunel University. I extend my gratitude to Douja, Habib, Assia and their wonderful Children.

Last but not least, I would like to thank the lovely Asma Kahoul for her endless support throughout this journey.

# Publications Associated with this Thesis

Flici, A. & Lü, K. 2011, "A Framework for Direct Marketing using Business Intelligence", *The International Conference on Digital Information Processing and Communications,* Ostrava, Czech Republic.

Flici, A. 2009, "A Direct Marketing Framework to Facilitate Data Mining Usage for Marketers: A Case Study in Supermarket Promotions Strategy", *Brunel Business School Doctoral Symposium,* Brunel University, London, United Kingdom.

Flici, A. 2008, "A Systematic Evaluation of Business Intelligence Tools: An Investigation into Potential Tangible/Intangible Value", *Brunel Business School Doctoral Symposium,* Brunel University, London, United Kingdom.

Flici, A. 2008, "Business Intelligence Tools: Toward an Evaluation Framework", *Graduate School Poster Conference,* Brunel University, London, United Kingdom.

# Authors' Declaration

I, Adel Flici, declare that the ideas, research work, evaluation and conclusions reported in my PhD thesis *A Conceptual Framework for the Direct Marketing Process using Business Intelligence* are entirely my effort, except where otherwise acknowledged. I also certify that this thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work.

# CHAPTER 1: INTRODUCTION

## 1.1 Introduction

This chapter provides an overview of this research and outlines the scope of this thesis. Section 1.2 presents the research background. Section 1.3 provides the motivations and research questions behind this thesis. Section 1.4 explains the research aim and objectives. Section 1.5 describes the methodology used to evaluate the overall research. Section 1.6 outlines the expected research contributions. Finally, Section 1.7 describes the overall structure of the thesis.

## 1.2 Background

Organisations operate in highly competitive markets and a volatile environment, where customers' purchasing behaviour is constantly changing and difficult to predict. In such competitive markets, direct marketing has become a key method to enhance promotion campaigns as well as develop strong customer relationships (Chen, Chiu & Chang 2005); and (Martínez-López & Casillas 2008). Traditional one-size-fits-all general marketing techniques are no longer enough to develop effective marketing campaigns. Direct marketing methods are needed to increase marketing campaign responses and to lower costs. Unlike general marketing, which is a product-oriented strategy, direct marketing is a customer-oriented method. It uses customers' details, commonly held in databases, to understand their needs (Tapp 2008). This is achieved through a process which includes identifying customers' characteristics to recognise their market values and predict the likelihood that they will respond to marketing campaigns (Rao, Steckel Joel H. 1998); and (Bose, Chen 2009).

Direct marketing is becoming an increasingly popular strategy in many organisations, operating across a wide variety of industries. According to a report issued by the Direct Marketing Association (DMA 2008), direct marketing expenditure in the UK has reached £50.5 billion, with a 15.4% increase since 2006. The direct marketing budget allocation by organisations has had a 9.8% growth, with an increase in the total number of organisations. Due to its economic impact, direct marketing research is growing in popularity among both academics and practitioners. This popularity is reflected in a variety of places such as academic journals, industry reports, newspapers, magazines, and brochures.

## 1.3 Motivations and Research Questions

Despite widespread debates and various contributions to direct marketing, this thesis identifies several research gaps in the field. In fact, as the literature review demonstrates, existing contributions are mainly focused on the technical issues of direct marketing as derived from an information systems perspective. For example, studies such as (Ha, Cho & MacLachlan 2005), (Kaefer, Heilman & Ramenofsky 2005), (Cui, Wong & Lui 2006), (Tettamanzi et al. 2007), and (Kim 2009) have proposed improving data mining techniques and machine learning methods, for more effective direct marketing. The main contributions of these studies focus on merging or modifying algorithms in order to enhance the existing data mining methods for direct marketing purposes.

However, very little attention has been given to formalising the process of applying data mining and advanced analytics to the direct marketing process to allow non-analysts to use such tools to help them resolve direct marketing issues. This process involves incorporating a collection of marketing concepts and business analytics principles, which together form a ‘self-contained’ choice for marketers. This makes it a significantly challenging process to perform (Tapp 2008).

Moreover, the direct marketing process is commonly executed as several disconnected activities and operations in many organisations (Vesanen, Raulas 2006). It is also considered to be an ad hoc process as it is usually executed in different ways depending on the process objectives (Rao, Steckel Joel H. 1998). For example, the marketing strategy for a company focusing on a particular product is likely to differ from those of companies focusing on services. Meanwhile, the diversity of the parameters affecting the performance of a direct marketing process designed for a given situation makes it possible to have several different ways of executing that process. Consequently, marketers are facing many challenges in undertaking the direct marketing process effectively. In addition, the complexity of advanced analytics such as data mining makes marketers more reluctant to utilise the resulting models due to their difficulty, poor comprehensibility, and trust issues (Kim 2006); and (Cui, Wong & Lui 2006).

There are several direct marketing models and data mining methodologies that can be used to perform the direct marketing process. However, these models and methodologies do not provide or propose an integrated information system platform able to effectively support direct marketing’s various activities. For example, researchers such as (Vesanen, Raulas 2006) and (Peltier, Schibrowsky & Schultz 2003) did not suggest an integrated information system platform to support the direct marketing process (refer to Section 2.4.3). In addition, data mining methodologies such as CRISP-DM[1] can be used to execute the direct marketing process. However, these methodologies are not specifically tailored to direct marketing and therefore a high level of user judgement is required. This makes the process of extracting marketing intelligence in a direct marketing context difficult to achieve.

Given the above-mentioned issues, the direct marketing process clearly needs a sound foundation to be used effectively within organisations. Consequently, this thesis seeks to address the managerial and organisational issues of the direct marketing process. Owing to the socio-technical aspects related to direct marketing, this research does not exclusively focus on the marketing concepts and business analytics principles, but also on people, technologies, activities, operations, technical difficulties, collaboration, and strategy, all of which play an important part in direct marketing.

Based on the research motivations, the researcher set the following questions for this thesis:

- *How can the direct marketing process be executed more effectively?*
- *What are the stages, activities, and technologies needed to execute the direct marketing process effectively?*
- *Why do organisations need a framework to manage and execute the direct marketing process?*

Following the above questions, three other questions are formulated:

- *How can the developed framework be evaluated effectively?*
- *Can the developed framework be used to plan a direct marketing strategy in a real operating company?*
- *Does the developed framework provide users (e.g. marketers/analysts) with systematic and structured approach to develop a direct marketing strategy?*

[1] CRISP-DM is a process model aimed at facilitating the management of small to large data mining projects.

## 1.4 Research Aim and Objectives

This research aims to investigate the stages, activities, and technologies needed to effectively execute the direct marketing process. To this end, the thesis examines the fundamental concepts of direct marketing and various related business analytics practices. As will be discussed in depth later (refer to Section 2.4.2), direct marketing is executed by many organisations as an ad hoc process, which seems to affect direct marketing's usefulness and practicality within an organisational context. In order to avoid this, the thesis attempts to address and clarify the complex nature of the direct marketing process by drawing upon a variety of academic and industry literature including that for marketing, information systems, and the evaluation process.

Based on a thorough conceptual investigation of the direct marketing process, this thesis's primary objective is to develop a conceptual framework for the direct marketing process. Many scholars have been tackling various direct marketing issues in diverse ways, ranging from the highly technical to the more organisational and managerial. However, as will be shown later (refer to Section 2.4.3), those researchers into direct marketing process have failed to address some important issues, especially issues of the direct marketing process from an information systems perspective. Therefore, this thesis proposes Business Intelligence (BI) as an integrated information system platform on which the direct marketing process can be executed. As will be described later (refer to Section 3.3.3), BI tools offer organisations a complete set of functions that can effectively support the direct marketing process.

In summary, this thesis examines direct marketing concepts to identify the appropriate stages and activities needed to perform the direct marketing process. It also investigates direct marketing technologies to find a suitable information system platform to execute the direct marketing process. The aim is to develop a conceptual framework for the direct marketing process. The following is a summary of the main objectives of the research:

1. To comprehensively review the literature in the area of direct marketing, with particular emphasis on its process.
2. To investigate the main direct marketing concepts and business analytics practices.
3. To identify an integrated system platform to help execute the direct marketing process.

4. To develop a conceptual framework for the direct marketing process with business intelligence tools.
5. To illustrate the application of the conceptual framework in an organisational context.
6. To evaluate the usefulness and practicality of the conceptual framework using three case studies in different sectors.

## 1.5 Research Methodology

In order to achieve the research objectives, this thesis will employ a rigorous research methodology which will address five research principles: research philosophy, research method, research design, data collection, and data analysis. First, a positivist stance will be adopted as the research philosophy in this thesis. This research philosophy was adopted because of its suitability to address the research objectives. Similarly, a qualitative approach will be adopted to analyse the findings. This approach is appropriate where the research question is followed by "*How*" and "*What*", as will be discussed in the research methodology chapter. Second, a case study method will be used to perform the conceptual framework evaluation. Third, the research design will provide the overall structure of the thesis, with clear guidelines and procedures on the tasks needed to complete the research aim and objectives. Fourth, the data collection will be done from three organisations operating in the retailing, telecommunications, and higher education sectors. The data collected will be used to evaluate the conceptual framework performance in an organisational context. Finally, analytical strategies of the case study method will be used to assess each case study individually and collectively. Specifically, the "*relying on theoretical propositions*", "*using both qualitative and quantitative data*", and "*examining rival explanations*" analytical strategies will be used to examine the findings of the three case studies.

## 1.6 Contributions

This research aims to make a significant contribution to the direct marketing literature by introducing a new conceptual framework for the direct marketing process. The conceptual framework is intended to facilitate the execution of the direct marketing process by marketers and analysts within an organisational context.

Based on the theoretical and empirical discussions, this thesis is expected to contribute to the research on the direct marketing process both theoretically and practically. In theoretical terms, the thesis is expected to contribute to the following three main aspects:

- First, it will underpin the understanding of the main concepts of the direct marketing process. This provides a sound theoretical foundation for developing the conceptual framework as well as enriching the debates on the direct marketing process.
- Second, the thesis will contribute towards direct marketing theory by empirically confirming the appropriateness of various components of the direct marketing process and validating the conceptual framework in different industries.
- Third, the integration of business intelligence practices with the direct marketing process will contribute to advancing direct marketing studies. As mentioned earlier, whilst many researchers are tackling direct marketing issues, those efforts focus on technical issues and fail to properly address direct marketing process issues from a systems perspective. Therefore, through the development of a conceptual framework, this thesis will attempt to address direct marketing process issues, with particular emphasis on the systems perspective.

In practical terms, the thesis is expected to contribute to organisations in two ways:

- First, the conceptual framework intends to provide organisations with a clear and systematic process to execute direct marketing. This could be used as a standard procedure for marketers and analysts to execute the direct marketing process within an organisation.
- Second, the distinct integration of business intelligence with the direct marketing process can be highly valuable, for executing and implementing a direct marketing strategy within an organisational context.

In summary, the conceptual framework is expected to provide a new way to execute and manage the direct marketing process. There are many organisations that practise direct marketing in different forms without recognising that it is, in fact, a well-studied marketing process. This is due to the fact that most modern organisations keep data about their customers which could be used for direct marketing purposes. In these cases, direct marketing may not be used efficiently owing to the lack of a defined methodology or framework. Consequently, the conceptual framework implemented in this thesis is expected to make a significant contribution for contemporary organisations operating in various industries.

## 1.7 Structure of the Thesis

*Chapter 1* introduces the overall background of the thesis, with the initial motivations and the main research question. It outlines the aim and research objectives of this thesis. It also explains the research methodology used to achieve the aim and the objectives of the thesis. Finally, it provides an overview of the expected contributions of this research.

*Chapter 2* investigates the existing literature on issues in direct marketing. It begins with the growth and origins of the direct marketing discipline. It then discusses the terms used to refer to direct marketing, and how it differs from Customer Relationship Management. Next, the chapter explores the existing direct marketing concepts and practices. Specifically, it describes the direct marketing process and its applications as well as its relation to business analytics such as data mining. The chapter concludes by presenting state-of-the-art research on the direct marketing process. It categorises the current issues in direct marketing into two contrasting research schools: the technical and the social. This research focuses on the social school, where current literature that attempts to resolve direct marketing process issues is discussed. A summary of the limitations of existing direct marketing process models is given.

*Chapter 3* presents the conceptual framework for the Direct Marketing Process with Business Intelligence (DMP-BI). It starts with examining the most important concepts and practices to consider in the development of the DMP-BI framework. It determines that the marketing database and business analytics constitute the fundamental components of the direct marketing process. It then proposes Business Intelligence as an integrated information system platform to support the DMP-BI framework. It also provides an overview of the BI tools industry and the reasons they are selected for the DMP-BI framework. Next, it discusses the key literature used to develop the DMP-BI framework. It concludes by presenting a graphical illustration of the DMP-BI framework, along with a detailed explanation of how each stage is executed, and how the BI functions intend to support each stage.

*Chapter 4* addresses the research methodology. First, it discusses different research philosophies and identifies positivist as the most suitable one to adopt for this research. Second, it provides justifications for choosing a qualitative approach and the case study method to evaluate the DMP-BI framework performance. Third, it presents the research design for the entire thesis. This includes the unit of analysis used in a multiple-case design approach. It also provides the criteria that are used to evaluate the quality of a case study. Next, the chapter introduces the three organisations' datasets that are used in the case studies. It explains the reasons those organisations' datasets are used. It then provides the analytical techniques and reporting structures that are used to evaluate the case studies' findings. Finally, it concludes by covering the ethical issues related to the organisations under study.

*Chapter 5* presents the supermarket promotions case study, which is used to demonstrate and evaluate the usage of the DMP-BI framework in the retailing sector. This chapter is organised into two main parts: 1) Case Study Report, and 2) Case Study Evaluation. The first part is structured into three sections. The first section provides a brief introduction to the supermarket's promotions practices. The second section describes the dataset and the study proposition used to execute the direct marketing process. The third section is a step-by-step illustration of the usage of the DMP-BI framework in a supermarket promotions context.

The second part of *Chapter 5* evaluates the case study based on three themes: *the suitability of the stages of the DMP-BI framework, the applicability of the BI functions, and the organisation and structure*. This part is organised into three sections. The first section evaluates the stages of the DMP-BI framework. The second section examines the BI functions' impact on the direct marketing process. The third and last section assesses the links in the DMP-BI framework.

*Chapter 6* provides the second case study, which is used to illustrate and evaluate the DMP-BI framework in the telecommunication sector. The chapter is structured into two parts including the case study report and case study evaluation. The first part is organised into three sections. The first section provides an introduction to a major issue in the telecommunication sector, which is the high level of churning. The second section explains the dataset and study proposition used to carry out the direct marketing process using the DMP-BI framework. The third section is a step-by-step illustration of the application of the DMP-BI framework in the telecommunication sector.

The second part of *Chapter 6* presents the case study evaluation, which is based on three themes, namely: *the suitability of the stages of the DMP-BI framework, the applicability of the BI functions, and the organisation and structure*. This part is structured as follows: the first section evaluates the DMP-BI stages, the second one assesses the BI functions, and the third section examines the links in the DMP-BI framework.

*Chapter 7* presents the third case study, which is used to demonstrate the application of the DMP-BI framework in the higher education sector. Similarly to *Chapters 5 and 6*, this chapter is organised into two main parts: 1) Case Study Report, and 2) Case Study Evaluation. The first part is structured into three main sections. The first section briefly outlines the general marketing practices in the higher education sector. The second section presents the dataset and the study proposition used to execute the direct marketing process using the DMP-BI framework. The third section provides a step-by-step demonstration of the usage of the DMP-BI framework in a higher education context.

The second part of *Chapter 7* provides the case study evaluation, which is based on three themes including: *the suitability of the stages of the DMP-BI framework, the applicability of the BI functions, and the organisation and structure*. This part is organised into three sections. The first section evaluates the DMP-BI stages. The second section assesses the BI functions' impact on the direct marketing process. The third section examines the link between each stage within the DMP-BI framework.

*Chapter 8* evaluates and discusses the case studies' findings. The chapter begins by providing a cross-case evaluation table with a detailed explanation of the similarities and differences between the usages of the DMP-BI framework components for each case study. It then discusses the key benefits of the DMP-BI framework by comparing it with three rival methodologies. It finally concludes by providing a clear view of the main advantages that the DMP-BI framework has over the three rival methodologies.

*Chapter 9* concludes the thesis. It begins by discussing the theoretical and practical contributions of this research. It then provides this research's limitations including contextual, technological, and methodological limitations. Finally, it suggests future research directions in the area of direct marketing.

# CHAPTER 2: LITERATURE REVIEW

## 2.1 Introduction

This chapter provides a context to this study through an in-depth review of the published literature on Direct Marketing (DM) and its associated concepts, technologies, and applications. First, direct marketing development is presented and several misconceptions on direct marketing are addressed. Second, the most important concepts and applications in the direct marketing field are described. Analytical techniques and technologies that are used for direct marketing applications are also explained. Finally, an in-depth discussion on the state-of–the-art research and issues in direct marketing is provided.

## 2.2 The Development of the Direct Marketing Discipline

This section presents an overview of direct marketing history along with key factors that have contributed to its rapid development. It then provides a definition of direct marketing, and terms that are commonly used to refer to direct marketing. Finally, the section discusses the differences between direct marketing and Customer Relationship Management (CRM).

### 2.2.1 The Growth and Origins of Direct Marketing

Direct marketing originated as a mail-order discipline almost two centuries ago. This is in contrast to the claim of many journals and magazines that direct marketing is a new discipline (Tapp 2008). In fact, the USA experienced a major growth of mail-order sales during the 1800s. This was due to a significant rise in demand for goods in isolated communities. As a result, distribution and postal systems in those areas needed to be improved to provide better services. Similarly, some UK companies started to provide direct distribution services, such as lists of products and prices, in an attempt to reach customers who were geographically dispersed (Evans, O'Malley & Patterson 2004).

The expansion of direct marketing from mail order first started in the USA during the 1970s (Tapp 2008). American Express was one of the first main organisations to identify direct marketing’s potential. A decade later, the European market started adopting direct marketing when big UK companies such as British Telecom were evaluating direct marketing performance. In the late 1980s, the charity sector had made significant advancements in the usage of direct marketing, helping the emergence of highly accomplished practitioners in the 1990s (ibid).

In the last two decades, direct marketing has seen an enormous expansion, making it the fastest-growing marketing discipline worldwide (Tapp 2008). This is mainly because of the great benefits it offers to both buyers and sellers (Kolter, Armstrong 2008). For buyers, direct marketing offer customers access to a wealth of individually designed products anywhere in the world. For example, Dell provides customers with thousands of self-customised PC configurations, easily outnumbering its competitors in terms of PC choices sold in retail stores. For sellers, it offers a lower-cost, rapid and efficient alternative for reaching their markets. It is also very effective in terms of building strong and long-term customer relationships. Besides this, it eliminates boundaries to buyers by offering alternative channels to market products, e.g. internet mail (ibid). There are many other major factors that contributed to direct marketing's rapid development; the following are the most important ones:

- Database systems' rapid development: companies have access to highly sophisticated storage systems at a low cost. This allows them to keep a wide variety of data types such as customers' demographic information.
- Highly educated marketing managers: training has led to a new generation of well-educated marketing managers.
- Fragmentation of society: the massive growth in lifestyle options has made our society more complex, hence a more direct approach is more relevant than ever.
- Greater consumer power and sophistication: the dramatic increase in competition has forced companies to seek more effective methods to gain competitive advantage. Moreover, consumers' expectations of better services at a lower cost are ever increasing.
- Traditional marketing inefficiency: the poor impact and high cost of traditional marketing methods, such as mass marketing, have forced companies to look for alternative methods. Direct marketing can lower the cost of marketing campaigns as well as increase responses.
- Customer relationship management: the great benefits of customer retention and loyalty, especially cost wise, can only be improved by using direct marketing.
- The media: the development of satellite, cable, and digital has provided a whole new face of television broadcasting where the adoption of a mass-marketing approach comes at a higher cost.

- And finally, the internet: it has contributed to the economic growth of all sectors including direct marketing (Tapp 2008).

Direct marketing growth has mostly occurred in non-traditional business categories. In particular, heavy users of direct marketing include credit card companies, banks, investment companies, and insurance. Other direct marketing users comprise telecom, cable, and utility companies, airlines, associations, automobile manufacturers, retailers, and shopping centres (Stone, Jacobs 2008). Moreover, business-to-business direct marketing is also heavily used and is growing faster than in business-to-consumer direct marketing. Furthermore, it may well exceed consumer direct marketing in terms of total revenues in the near future. Finally, it is hard to find a company that does not use direct marketing in one form or another. The direct marketing discipline has grown to be the foundation of multichannel communications that provide measurable, targeted, one-to-one relationships between direct marketers and customers (ibid).

### 2.2.2 Direct Marketing Defined

Marketing is a process used by companies to create value for customers and build solid customer relationships in order to get a return from customers in the form of sales, profits, and long-term customer equity (Kolter, Armstrong 2008). There are two types of marketing methods: General Marketing and Direct Marketing. General marketing uses mass media such as television to target customers, regardless of their characteristics and preferences (Bose, Chen 2009). It is a product-oriented strategy, which aims to achieve market shares for specific products. In comparison, direct marketing is a customer-oriented method. It uses customers' details, commonly held in databases, to perform three main activities: 1) analyse customers' data, 2) formulate a marketing strategy, and 3) implement it to obtain direct responses from customers (Tapp 2008). The use of databases to analyse customers' purchasing behaviour is a fundamental building block of direct marketing (Stone, Jacobs 2008). Database analyses comprise a wide range of direct marketing applications including segmentation and targeting, which are discussed in more detail in the next section (i.e. Section 2.3). Strategy formulation and implementation involve the interactive use of advertising media to stimulate an immediate change in customers' purchasing behaviour (ibid). In other words, these activities aim to understand customers' preferences, which allow marketers to explicitly plan any subsequent interactive approach toward customers based solely on their needs (Tapp 2008).

There are many terms used to refer to direct marketing (e.g. personalised marketing and relationship marketing). The usage of such terms has made it more difficult to reach a consensus on a standard direct marketing definition (Stone, Jacobs 2008). However, the definition provided above comprises the fundamental elements of direct marketing. The following subsection seeks to clarify the confusion created from the usage of various terms to refer to direct marketing.

### 2.2.3 Direct Marketing Terms

Direct marketing has many terms used to refer to it. Some of the most widely used ones are: personalisation, customisation, database marketing, segmentation, one-to-one marketing, targeting marketing, profiling, loyalty marketing, and interactive marketing. Therefore, it is important to have a clear distinction between terms that are similar to direct marketing and terms that are misused to describe it.

There are various marketing books that view direct marketing as part of the marketing communication mix and do not mention database marketing at all (Tapp 2008). The marketing trade press also tends to use direct marketing and direct mail interchangeably. It even uses junk mail and cold-calling to refer to direct marketing. These terms are inaccurate for referring to direct marketing for various reasons. First, direct marketing originated from mail order, which was a distribution method rather than a means of communication. Second, direct marketing's core activity is analysis and communications such as direct mail are additional activities. Finally, direct marketing's key objective is to create a complete marketing strategy rather than being part of one (ibid).

The many terms used to refer to direct marketing have created confusion among both academics and practitioners. There are some terms that are correctly used and others that are inaccurate. Hence, it is critical to describe direct marketing and other terms carefully before using them as a reference to direct marketing. For example, personalisation marketing is described by (Vesanen, Raulas 2006) in a similar way to direct marketing. Therefore, the usage of this term to refer to direct marketing is accurate. In fact, personalisation marketing is used in several research papers and books such as (Adomavicius, Tuzhilin 2005) to refer to direct marketing. Another major confusion in the direct marketing discipline is the difference between direct marketing and CRM. The following subsection attempts to cover this issue.

### 2.2.4 Direct Marketing & Customer Relationship Management (CRM)

Modern marketing is based on two competing philosophies: the 'four Ps' and relationship marketing (Tapp 2008). First, the 'four Ps' is a transaction approach characterised by Product, Price, Place, and Promotion. Second, relationship marketing is more focused on building long-term customer relationships as well as striving for customer-led quality, service, and marketing within a company (ibid). CRM developed its roots from relationship marketing (Ryals, Knox 2001). Indeed, it is used to enhance relationship marketing by shifting it from the 'four Ps' transaction-based approach, with its focus on finding new customers, to customer retention, using its more innovative practices in the marketing field (Ryals, Knox 2001); and (Breur 2007).

CRM involves tailoring products and services based on customers' preferences rather than some general characteristics. It is certainly critical to have a good understanding of customers' needs and preferences in order to achieve an effective CRM application. In addition, the marketing function is very important for CRM effectiveness, as it is the way companies interact with their customers (Shaw et al. 2001). There are four categories of CRM dimensions: Customer Identification, Customer Attraction, Customer Retention, and Customer Development (Ngai, Xiu & Chau 2009). These four dimensions' main purpose is to provide a deeper understanding of customers in order to capitalise on their value for the organisation. First, customer identification involves target customer analysis and customer segmentation, which lead to the discovery of the population who are most likely to become customers. Second, customer attraction consists of looking into customer characteristics and carrying out direct marketing. Third, the customer retention dimension involves direct marketing, loyalty programmes, and complaints management. The main objective of this dimension is to keep customers satisfied. Finally, customer development entails lifetime value analysis, up/cross-selling, and market basket analysis. This dimension's primary aim is to intensify transaction value and individual customer profitability (ibid).

From the above CRM description, one can conclude that direct marketing is part of CRM. Specifically, CRM incorporates many activities and direct marketing is one of them. Unlike direct marketing, CRM needs to keep track of customer satisfaction and attempt to win back customers who are lost to the competition (Tapp 2008). Moreover, CRM uses many of the concepts and technologies used in direct marketing, but it uses them to integrate all areas of a business that affect customers, including marketing, sales, and customer service (Stone, Jacobs 2008). Indeed, CRM's ultimate purpose is to make information a driving force for the entire organisation and not only the marketing department. In short, CRM is a convergence of traditional direct marketing, database marketing decision-support tools, and digital marketing capabilities, making direct marketing an important part of CRM (ibid).

## 2.3 Direct Marketing Concepts and Practices

This section starts with a description of the direct marketing process, which is the foundation of direct marketing applications. This is then followed by an in-depth explanation of direct marketing major applications, along with key variables used in them. Finally, definitions of analytical techniques used to perform direct marketing applications are provided.

### 2.3.1 Direct Marketing Process

The basic aim of any direct marketing process is to obtain a measurable response which will generate an immediate or eventual profit (Stone, Jacobs 2008). A typical direct marketing process starts in two cases: when a new product or service is introduced, or when a channel is needed (Tapp 2008). It is followed by marketers gathering relevant data from databases or other sources. The next phase requires the analysis of data using statistical techniques or advanced analytical techniques such as data mining. The outcome of the analysis should provide marketers with enough information to sample a set of customers to be targeted (Bose, Chen 2009). Marketers are then required to develop customised products and communications for the target customers (Tapp 2008). This usually involves marketers choosing key components to produce a response, including lists/media, creative, and the offer (Stone, Jacobs 2008). These components can be divided into four parts, which are referred to as Elements of Promotion (see Figure 2.1) and are listed by importance: 1) media/lists (40%), 2) offer (20%), 3) copy (15%), 4) layout (15%), and 5) timing (ibid). Finally, those customers who respond are added to marketers' records, in order to evaluate the direct marketing campaign performance (Tapp 2008).

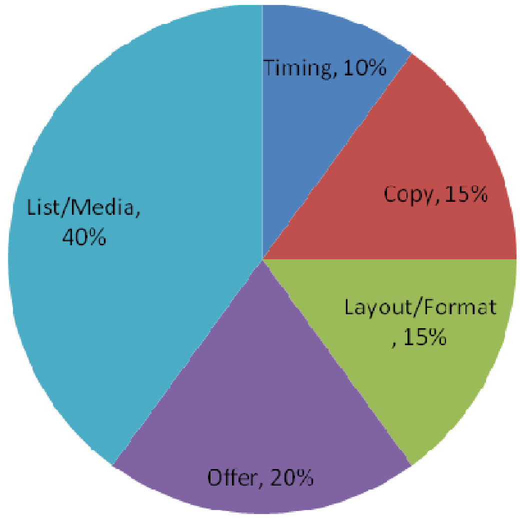



**Figure 2.1: Elements of Promotion (source: (Stone, Jacobs 2008, p.6))**

There are two fundamentals in the direct marketing process: understanding customers and interacting with customers. The latter consists of the 'Elements of Promotion' which are used to stimulate customers. On the one hand, the most creative or best offer may result in a low response if an inappropriate group of customers is targeted (Stone, Jacobs 2008). On the other hand, if a badly performed creative or poorly formulated offer is aimed at the right target group (i.e. a group that is interested in the product or service provided), it may depress consumer response, but not completely eliminate it. Thus, understanding customers' preferences and needs are more critical than an impressive creative or offer. This is because, regardless of the promotion's attractiveness, if the wrong group of customers is targeted, there will most probably be a low response. Additionally, the second fundamental part of the direct marketing process, i.e. "interacting with customers", is not the main subject of focus in this research. This is due to the fact that the "interacting with customers" area is very broad and could be a subject of research on its own right. In fact, there are many academics and practitioners studying this area. For example, Integrated Marketing Communications (IMC) is the management of all communications that strive to build strong relationships with customers and other shareholders (Stone, Jacobs 2008).

This research focuses on databases and analytical techniques to understand customers' needs and preferences. The success of the direct marketing outcome relies heavily on the effective application of these activities. The following subsection provides descriptions and best practices for understanding customers' needs and preferences.

### 2.3.2 Direct Marketing Applications

The technological developments in the collection, analysis, and use of customers' data have been a significant driver of many direct marketing applications (Evans, O'Malley & Patterson 2004). There are two main direct marketing applications: segmentation and targeting (Tapp 2008). These applications are primarily aimed to achieve customer acquisition and retention (Sargeant, Douglas 2001). Figure 2.2 demonstrates the usage of a firm's database to perform segmentation and targeting. The following subsections contain detailed explanations of these direct marketing applications.

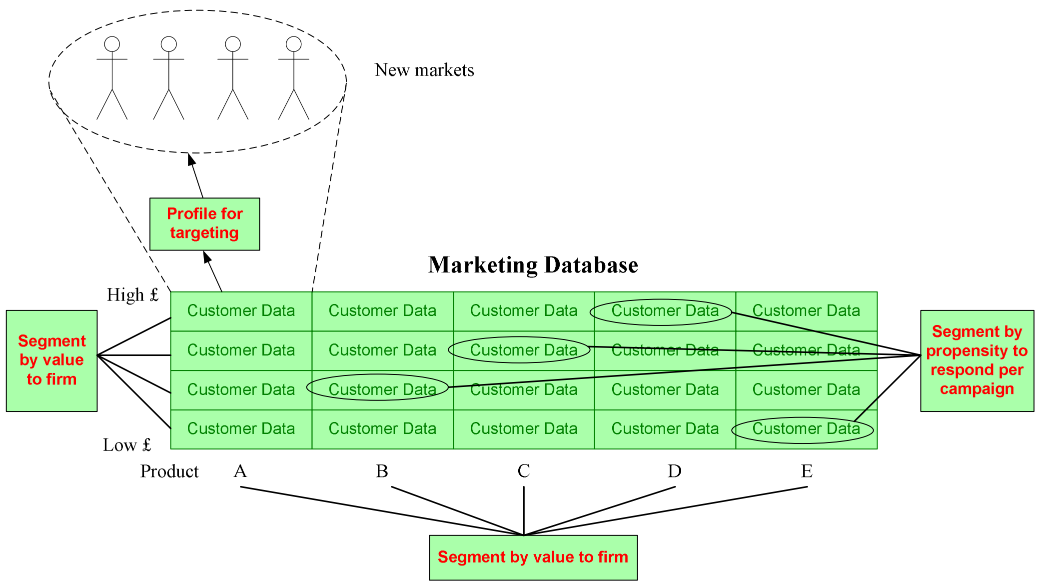



**Figure 2.2: Direct marketing applications using marketing database (source: (Tapp 2008, p.59))**

#### 2.3.2.1 Segmentation Application

Customer segmentation is the process of identifying homogenous groups of consumers which have similar tastes, needs, and preferences toward a company's products/services (Kumar, Werner 2006). The segmentation process is usually performed using two types of data: behaviour and profile data. On the one hand, behavioural data can help marketers to segment customers by value as well as forecast their lifetime value. On the other hand, profile data can be used to evaluate a campaign response and also to target new customers accurately (Tapp 2008).

There are two categories of segmentation: objective and subjective. First, an objective approach to segmentation may be measured unambiguously (e.g. by age and gender). Second, a subjective segmentation needs to be measured for the respondents and is usually psychographic, such as attitudes and intentions (Evans, O'Malley & Patterson 2004). Segmentation can be further classified into three levels. At the first level, general segmentation is based on permanent or long-term customers' characteristics (e.g. occupation, family composition, and lifestyle). At the second level, domain-specific segmentation takes into account specific product classes and consumption domains such as breakfast or commuting. At the final level, segmentation is based on specific levels, i.e. consumers are grouped into heavy and light users of a specific brand, for instance (ibid). Table 2.1 show the three levels of segmentation along with objective and subjective differences.

| **Level** | **Objective** | **Subjective** |
|---|---|---|
| General (Consumption) | Age, education level, and geographic area. | Lifestyle, general values, and personality. |
| Domain-specific (Product Classes) | Purchase/usage frequency, substitution, and complementarity. | Perception, attitude, preference, interests, opinions, and domain-specific values. |
| Specific level (Brand) | Brand loyalty (behaviour), and purchase/usage frequency. | Brand loyalty (attitude), brand preferences, and purchase intention. |

**Table 2.1: Segmentation Variable Classification (source: (Evans, O'Malley & Patterson 2004, p.160))**

The following subsections present the most frequently used variables in segmentation, including demographic, geographic and geo-demographic, behavioural, and psychographic/lifestyle variables.

***Demographic Variables***

The most common demographic variables include age, gender, family life cycle, and income/occupation (Sargeant, Douglas 2001). Of course, there are other variables that are not mentioned above, such as socio-economic group, religion, race, nationality, or education. Demographic variables are very effective in consumer markets as customers' needs and preferences are highly related to such variables. Moreover, demographic variables have been collected over many years, thus much research has been done on the consumer behaviour of each specific group (ibid).

Age is still a valid application of direct marketing segmentation as it holds important and, at times not too obvious, implications for marketers (Evans, O'Malley & Patterson 2004). For example, people in the 16–24s age group have become important spenders, but they tend to be especially individualistic and sceptical of marketing activities. Therefore, it has become harder to influence them, but they are still reachable. In fact, direct marketing provides additional tools that the 16–24s age group is looking for – greater interactivity as well as involvement in marketing communications (ibid).

Gender is also an important variable for customer segmentation. Specifically, gender provides a solid indication of a propensity to purchase a specific product or brand. For example, women are found to respond to information in a completely different way to men (Sargeant, Douglas 2001). In particular, the most significant characteristics to make a product/brand appeal to women are:

- intuition over reason,
- concern with appearance,
- persuasion rather than aggression,
- the need to nurture,
- quality rather than quantity.

Family life cycle tends to be very popular in direct marketing. Indeed, many companies such as financial organisations and supermarkets segment their customers on the basis of lifestage (Evans, O'Malley & Patterson 2004). This is achieved by keeping key variables such as age, marital status, presence and age of youngest child at home, gender, and labour force status. An example of lifestage could be ‘Full nest 1’: customer age between 18 and 35, young child 5 or under, home buying at peak, low expendable income, low savings, and high borrowings (Tapp 2008).

Income/occupation segmentation has proved to be very useful in providing marketers with a good indicator on a propensity to buy specific categories of products. However, it is usually difficult to determine someone’s income. Therefore, a popular method used by marketers is to categorise customers into social grades. Table 2.2 is an illustration of the social grade model (Sargeant, Douglas 2001).

| Social grade | Generic occupation |
|---|---|
| A | Senior professional/managerial |
| B | Middle professional/managerial |
| C1 | Supervisory management – clerical |
| C2 | Skilled manual labour – e.g. plumber |
| D | Unskilled manual – e.g. labourer |
| E | Unemployed, students, etc. |

**Table 2.2: Generic Social Grade Model (source: (Sargeant, Douglas 2001, p.129))**

***Geographic and Geo-demographic Variables***

Geography is very important for segmentation. In fact, customers' addresses are considered among customers' most informative characteristics (Linoff 2008). The first law of geography, which was attributed to Waldo Tobler: *"Everything is related to everything else, but near things are more related than distant things."* (Longley et al. 2005, p.65) supports the assumption that the closer the store is to customers, the more likely it is that they will shop at it (Longley et al. 2005). However, geographic segmentation is a very general approach and can provide little in the way of fine detail, specifically when analysing customer markets. As a result, geo-demographic segmentation is an attempt to address the limitations of the general geographic model (Sargeant, Douglas 2001).

Geo-demographics originated from work carried out by Webber[2] in 1973 (Sargeant, Douglas 2001). The work consisted of studying urban deprivation in Liverpool and classifying neighbourhoods into 25 different types using clustering techniques. The cluster analysis was based on population, housing, and socio-economic characteristics. The results showed that each neighbourhood had different mixes of problems and required a different type of social policy (ibid). The next significant development came in 1981, when some 40 census variables were analysed for the UK Census. This led to the creation of 39 neighbourhood types in the first geo-demographic system in the UK, known as ACORN[3] (Evans, O'Malley & Patterson 2004). There are a number of similar

[2] Melvin M. Webber (Hartford, May 6, 1920 - Berkeley, November 26, 2006) was an urban designer and theorist associated for most of his career with the University of California at Berkeley but whose work was internationally important (Wikipedia).
[3] A Classification of Residential Neighbourhoods (ACORN) developed by Consolidated Analysis Centres Incorporated (CACI).

geo-demographic systems including MOSAIC, PINPOINT, and FINPIN (Sargeant, Douglas 2001).

Geo-demographic systems commonly include information on: age, marital status, household composition, household size, employment type, travel to work, unemployment, car ownership, housing tenure, amenities, housing type, and socio-economic group. This information can have considerable potential for controlling the activities of the TGI (target group index) (Sargeant, Douglas 2001). The TGI is an annual report of 34 volumes of consumer profiles in most product markets, derived from samples of more than 20,000. From this, each geo-demographic group's interest in the product concerned can be identified (Evans, O'Malley & Patterson 2004). Hence, geo-demographic systems could help marketers perform direct marketing planning through market area analysis. For example, geographic expansion of a new entrant in retail is an obvious strategic application (Longley et al. 2005). This usually involves two distinct tasks: 1) identifying the geo-demographic profile of existing best customers, and 2) investigating the profile of the new area (using CACI or MOSAIC, for instance) and matching them with existing best customers, with the potential for targeting using customised communication such as direct mail and telemarketing (Evans, O'Malley & Patterson 2004).

***Behavioural Variables***

Behavioural variables can help marketers to segment customers based on their knowledge, attitude, use, or response to a product (Kotler 1991) cited by (Sargeant, Douglas 2001). They can also provide valuable information for marketers to better understand customers' value to a company. Understanding customers' value is a fundamental application for direct marketing success. In fact, practitioners have always related direct marketing's continuous growth to the key reason of its measurability. Accordingly, direct marketing has benefited from marketers' increasing requirements for more accountability from their advertising expenditures (Stone, Jacobs 2008). Each customer brings a certain value to the company. Generally, some customers are more important than others, depending on the value they add to the company (Tapp 2008). Hence, segmenting customers by value may be very useful as it is common that a small percentage of the market accounts for a large percentage of consumption. Indeed, Pareto's principle, known as the '80/20 rule', is based on the concept that 20 per cent of customers will frequently generate 80 per cent of an organisation's profit (Sargeant, Douglas 2001). Direct marketing purposely targets high-value customers differently from others to maximise return. For example, the Hilton Hotel provides special treatments for its most regular and valuable customers by keeping a record of their preferences, such as their favourite newspaper or preferred size of bed. Hence, when a valuable customer arrives at the hotel, the database flags them up as worthy of special attention (Tapp 2008).

Furthermore, the ever-growing competition generated by technologies such as long-distance and wireless telephone, internet service providers, and credit cards has made churn (the rate of consumer defection) very high (Stone, Jacobs 2008). These technologies have helped customers to be more aware of low, short-term promotional rates and make the most of them. For example, if a competitor offers similar attractive promotions, customers will commonly move and take advantage of the offerings. This is where behavioural variables can be used to achieve higher customer loyalty and could make the difference between retaining and losing a customer (ibid).

It is also widely accepted by both academics and industry professionals that conducting business with new customers will cost five times as much as doing it with existing customers (Sargeant, Douglas 2001). Indeed, an organisation's existing customers have proved their interest in the product/services provided. Therefore, they are more likely to make future purchases. For example, a supplier of garden seeds will have a higher response from mailing existing customers than using a cold list of prospects. This is due to the fact that some individuals from the prospects list may not be interested in gardening or are happy to make purchases from competitors (ibid). Consequently, understanding existing customers' individual needs is very important in marketing practices. Furthermore, customers want to be treated more as individuals (Evans, O'Malley & Patterson 2004) and direct marketing can help marketers to achieve that; in particular, using the behavioural variables held on each customer. For example, this type of data can help marketers to investigate the various benefits that the same product may provide to customers. This can enable direct marketing to be used for those customers with accuracy. For example, American Express (AE) segmented its customers based on similar purchasing behaviour using their credit card. Subsequently, they can identify the biggest spenders on petrol and look for partnerships with petrol providers to offer those spenders special discounts (Tapp 2008).

*Psychographic/Lifestyle Variables*

Psychographic segmentation is performed on the basis of three common attributes: personality, attitude, and values (Sargeant, Douglas 2001). First, the personality of customers can help marketers better understand their consumption behaviour. Indeed, a number of studies have proved that there is a correlation between personality variables and consumer purchasing behaviour of products such as mouthwash, alcoholic drinks, and headache remedies. An example of personality segmentation is First Direct Bank looking for customers who are risk takers. This segmentation was motivated by the fact that switching to a bank without physical branches (i.e. First Direct Bank) was perceived by many as a highly risky decision. Hence, the bank targeted risk takers who can be found in high-risk industries or even high-risk leisure activities such as paragliding and skiing (ibid). Second, attitude is commonly used in the not-for-profit sector where marketing purposes are to change societal attitudes to certain types of behaviour such as smoking and drinking. However, attitude can also be used for other business purposes. For example, direct marketing can be perceived by some customers as an invasion of their privacy. As a result, there are proposals to reduce marketing practices such as direct mail. However, by using consumer attitudes towards, for instance, direct mail, one can reduce privacy issues by targeting consumers who do not mind receiving direct mail promotions (Milne, Gordan 1993) cited by (Sargeant, Douglas 2001). Finally, values are customers' beliefs toward specific goals in life (i.e. terminal values) and modes of conduct (i.e. instrumental values) which are preferable to others. In other words, individual values can be used to guide customers' decision-making process. Values can be assessed using the following nine items, which are scaled as important/unimportant:

- sense of belonging,
- excitement,
- warm relationship with others,
- self-fulfilment,
- being well respected, fun and enjoyment of life,
- security,
- self-respect,
- sense of accomplishment (Sargeant, Douglas 2001).

Lifestyle segmentation provides a useful insight into what makes consumers react positively. It is aimed at identifying consumers' activities, interests and opinions. This is achieved through presenting Likert-scaled statements to respondents, and asking them to give their degree of agreement with each. The resulting data can be very useful in determining the style and mood of promotional messages. Table 2.3 provides examples of the kinds of statement presented to respondents (Evans, O'Malley & Patterson 2004).

| **Statements** |
|---|
| I purchase clothes for comfort, not for style |
| Once I find a brand I like, I tend to stick to it |
| I always buy British whenever I can |
| I dress to please myself |
| My family rarely sits down to a meal together at home |
| I enjoy eating foreign food |
| I like to do a lot when I am on holiday |

**Table 2.3: Lifestyle statement examples (source: (Evans, O'Malley & Patterson 2004, p.176))**

The importance of lifestyle segmentation is such that a number of external databases are available commercially. These databases' original basis was the values and lifestyle (VALS) system (Tapp 2008). This system effectively merges psychographic data with product and service behaviour data, which provides the foundations of the lifestyle databases' analyses. For example, if an individual responds that he/she regularly goes on holiday, it is likely that he/she would be interested in sun cream products. Table 2.4 presents major lifestyle databases (ibid).

| **Supplier** | **Product** | **Data** | **Source** |
|---|---|---|---|
| Claritas (Acxiom) | Behaviour bank | 11 million consumers | Survey shopper |
| CACI | Lifestyle | 44 million people | Electoral roll; census data; investment data and lifestyle data from Experian |
| Experian | Canvasse | 44 million people 8 million survey responders | Electoral roll; census data |
| Equifax | Dimensions | 44 million people | Electoral roll; census data; financial data |

**Table 2.4: Major lifestyle databases (source: (Tapp 2008, p.108))**

#### 2.3.2.2 Targeting Application

The targeting application is very important for developing successful direct marketing programmes. It consists of identifying customers who are most likely to respond to a particular campaign. Targeting's key purpose is to maximise ROI per customer through determining new markets more accurately, thereby lowering the cost of acquisition (Tapp 2008). There are three main targeting application categories, namely customer profiling, setting up customer's budget, and maximising ROI per campaign.

***Profiling Existing Customers to Target New Prospects***

Profiling allow organisations to have a more thorough knowledge of their client base. This knowledge acquisition process will result in a better understanding of the marketing and communications required to more effectively target prospective customers (Stone, Jacobs 2008). In fact, profiling existing customers can help organisations to identify potential new customers. Pareto's principle can explain how profiling can be beneficial to marketers:

*"Your new customers will have a similar profile to your existing customers."*(Tapp 2008, p.63)

In other words, similar types of people will have common product preferences. Hence, accurate profiling of existing customers will allow more precision when targeting new customers (Tapp 2008). For example, an organisation has segmented its customers by their value. Normally, there would be fewer high-value customers than low-value ones, thus the segments are divided as shown in Figure 2.3. Profiling in this example would help marketers to target specific segments for acquisition purposes (Sargeant, Douglas 2001). They could, for instance, focus entirely on the profile of the organisation's high-value customers and target customers who match their profile with similar products purchased by them.

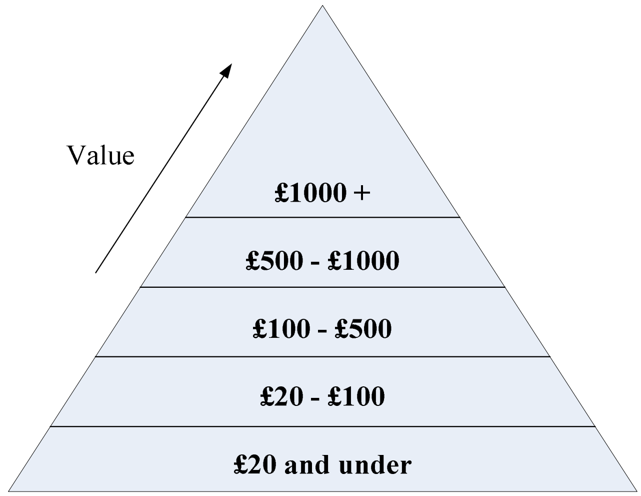



**Figure 2.3: Customer segments by value (source: (Sargeant, Douglas 2001, p.142))**

***Targeting to Maximise ROI per Campaign***

Marketers can use profile data to identify the most responsive customers. The primary aim is to allocate budgets on a per campaign basis such that return on investment (ROI) is maximised (Tapp 2008). For example, marketers are planning a direct promotion campaign and they have 100,000 customers' records stored on the database. However, they have a restricted marketing budget, able to cover only 20,000 customers. Consequently, the most efficient way to plan the direct marketing campaign is to investigate customers most likely to respond. This can ensure a better return on investment than selecting customers on a random basis (ibid). This illustrates the importance of customer targeting in maximising ROI per campaign.

***Setting up the Customer's Budget***

Setting up the customer's budget consists of predicting a customer's lifetime value. The aim is to construct a budget, on a per customer basis, which provides marketers with a clear figure of how much they can afford to spend. This is known as 'the allowable marketing spends per customer' and is very important in direct marketing practice (Tapp 2008). Setting up the customer's budget can be used to support five key management decisions:

- allocating acquisition allowances,
- selecting media for initial customer acquisition,
- identifying selection criteria for customer marketing,
- investing in the re-attraction of previous customers,
- allocating an asset value to the marketing database (Sargeant, Douglas 2001).

For example, a company that sells computers calculates the customer's average lifetime value at £100 gross profit per annum. This can help them set an allowable marketing spend, per customer, of say £30 per year. It is extremely important to calculate the allowable marketing spend, as it provides the basis for budget setting. Indeed, this is an essential direct marketing application which can be used as a starting point for both acquisition and retention budgets. In addition, it becomes critical when a restricted budget is provided for the marketing department. The calculation of a customer's lifetime value provides a sound basis to plan the budget spending (Tapp 2008).

#### 2.3.3 Direct Marketing & Analytical Techniques

Information technology's rapid development has strongly enhanced the way information is collected, stored, and utilised. This has significantly increased the usage of quantitative techniques to perform segmentation and targeting activities. There are two major categories of analytical techniques in direct marketing: 1) standard statistical techniques, and 2) data mining/machine leaning techniques (Bose, Chen 2009). The following subsections provide an in-depth insight into these two principal analytical techniques with a particular focus on data mining techniques. It is important to mention that statistical and data mining/machine learning techniques are very broad. These techniques are so vast that they can be the subject of a whole book. Therefore, this section is intended to cover the quantitative techniques used to perform direct marketing with the purpose of extracting marketing value and not to focus on each and every detail.

##### 2.3.3.1 Standard Statistical Techniques

There are two groups of standard statistical techniques: basic statistical techniques and advanced statistical techniques. First, basic statistical techniques comprise linear regression, logit/probit, tobit, beta/gamma, and discriminant analysis. Second, advanced statistical techniques include two-stage beta + gamma, two-stage logit + linear, two-stage probit + non-linear, latent class model logit, latent class model Poisson, and latent class model probit. Table 2.5 lists standard statistical techniques, including basic and advanced, along with each technique response and score type.

| **Basic/Advanced Statistical Techniques** | **Response Type** | **Score Type** |
|---|---|---|
| Linear regression | Response Interest Revenues | Continuous<br>Continuous |
| Logit/Probit | Binary Choice<br>Number of Products<br>Categorical Choice | Binary<br>Integer<br>Integer |
| Tobit | Revenues | Continuous |
| Beta/Gamma | Binary Choice | Binary |
| Discriminant analysis | Binary Choice | Binary |
| Two-stage Beta + Gamma | Binary Choice and Revenues | Continuous |
| Two-stage Logit + Linear | Binary Choice and Revenues | Continuous |
| Two-stage Probit + Non-linear | Binary Choice and Revenues | Continuous |
| Latent class model Logit | Binary Choice and Revenues | Continuous |
| Latent class model Poisson | Number of Products | Integer |
| Latent class mode Probit | Binary Choice | Binary |

**Table 2.5: Standard Statistical Techniques. Adapted from (Bose, Chen 2009, p.6))**

***Basic Statistical Techniques***

Linear regression is applied where variables are continuous and where the analysis objective is to identify trends. For example, this technique could be used to track sales and profitability over time. The dataset in Table 2.6 is used to illustrate linear regression practically. The aim is to investigate advertising expenditures (i.e. X) impact on sales revenue (i.e. Y) using linear regression (Sargeant, Douglas 2001). In order to calculate the best fitting straight line, the following equation is used:

$Y = A \oplus BX$ Equation 2.1

To calculate the slope (B), and the intercept (A), the following formulas were used:

- $B = \dfrac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2}$ Equation 2.2
- $A = \overline{Y} - B\overline{X}$ Equation 2.3 (Tabachnick, Fidell 2007).

The correlation (r) between the two variables (i.e. X and Y) is then calculated to identify whether the value of sales increases or decreases as the level of expenditure goes up (Sargeant, Douglas 2001). The following formula is used to achieve that:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{n s_x s_y} \quad \text{Equation 2.4}$$

The results of the equations are presented in Figure 2.4.

| Year | Advertising Expenditure | Sales Revenue |
|---|---|---|
| 1994 | £2000 | £60 000 |
| 1995 | £4000 | £90 000 |
| 1996 | £5000 | £100 000 |
| 1997 | £3000 | £70 000 |
| 1998 | £6000 | £110 000 |

**Table 2.6: Dataset for advertising expenditure and sales revenue (source: (Sargeant, Douglas 2001, p.249))**

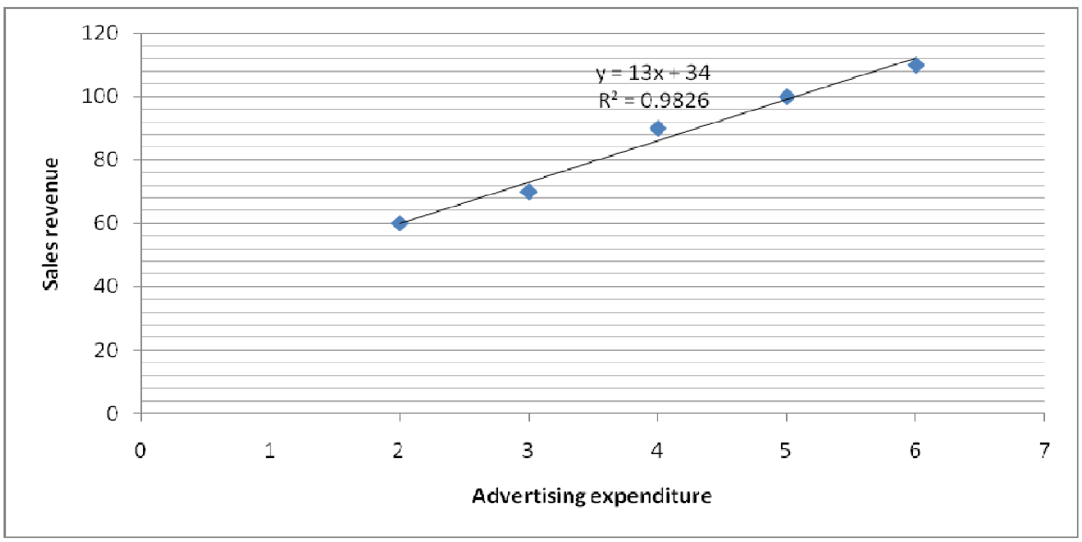



**Figure 2.4: Advertising impact on sales revenues (source: (Sargeant, Douglas 2001, p.250))**

Logit, probit, and tobit are also regression models (Bose, Chen 2009). They are similar to linear regression, but deal with discrete response directly by latent variables. Logit and probit use a direct approach in predicting the discretised response, which can yield a better predictive performance for large samples. Tobit models use latent variables to truncate values that are not within the threshold. For example, customers either generate revenues or do not; hence negative values in models do not have any meaning. The tobit technique transforms negative values to zero and holds only the positive value. Beta/gamma and discriminant analysis are also used for direct marketing and can produce more accurate results than regression analyses. However, the results' accuracy highly depends on the assumptions made prior to analyses. In other words, if assumptions are incorrect, it may lead to inaccurate estimation of parameters and hence overly inaccurate results (ibid).

***Advanced Statistical Techniques***

Advanced statistical models are commonly a combination of two basic statistical techniques. They typically involve a two-stage process. For example, the first stage is to model the probability of response and the second stage is to model the monetary value a customer could generate in response to the direct marketing activities (Bose, Chen 2009). Table 2.5 provides examples of such advanced techniques. For instance, logit techniques can be used to estimate customer response and linear regression to calculate monetary value. The other advanced techniques listed in Table 2.5 use similar approaches to identify customer response, monetary value, lifetime value, and so forth.

Lifetime Value (LTV) and Recency Frequency Monetary (RFM) are the most popular advanced statistical methods in direct marketing research and industry. They primarily cover the financial dimensions of direct marketing applications such as budget setting (Stone, Jacobs 2008). LTV aims to calculate the total contribution of a customer to an organisation over time. It is measured by estimating the costs and revenues related with managing a customer relationship (including communications, promotions, sales, costs, etc.) during his/her time with the organisation (Sargeant, Douglas 2001). In other words, lifetime values are the predicted net incomes expected from a customer at the present time (Tapp 2008). It should be emphasised that the money value can drop, i.e. £5 pound today is worth more than £5 next year. The principle is that money in the future is worth less than in the present day. For example, if £100 is invested, expectations are that it is going to be worth more next year. That is to say, it is more profitable to have £100 at present than in the future. This economic logic is known as the net present value (ibid). There are many different statistical techniques that can be used to calculate a customer's lifetime value. The basic formula to calculate individual customer value is as follows:

$$LTV = \sum_{i=1}^{n} C_i (1 \oplus d) \quad \text{Equation 2.5}$$

C = net contribution of each year's marketing activities,

d = discount/promotion rate,

i = the expected duration of the relationship in years (Sargeant, Douglas 2001).

A typical database would contain data on customers' date of last purchase (recency), number of purchases within a given period (frequency), and amount spent (monetary) (Stone, Jacobs 2008). RFM analysis involves the use of such records to identify customers with the highest monetary value and frequency of purchase (Evans, O'Malley & Patterson 2004). It consists of assigning each customer a score by assessing their purchase behaviour. This will allow marketers to find out which customers are most likely to respond to a marketing campaign. Moreover, it can also be used for segmentation, i.e. identifying customers by value and segmenting them as high, medium and low, for instance. In fact, the more a customer buys from you, the more likely he/she is going to spend with you in the future (Tapp 2008).

#### 2.3.3.2 Data Mining/Machine Learning Overview

Data mining and machine learning techniques are capable of analysing large amounts of customers' demographic, psychographic and behavioural data with the purpose of discovering hidden or interesting patterns, associations, and anomalies (Witten, Frank 2005); and (Kim 2006). Interesting patterns refer to a combination of validity, novelty, usefulness, and understandability. The use of data mining's resulting models can lead companies to new insights, and in a business context, to competitive advantage (Mitra, Pal & Mitra 2002). In a direct marketing context, data mining models can, for example, reveal a specific class of customers which is most likely to be interested in a particular product. This will subsequently allow the planning of a direct marketing campaign aimed toward that specific class of customers with the aim of achieving higher responses (Kim 2006).

Data mining and machine learning techniques can perform several tasks, which can be grouped into two main categories: classification and numerical prediction. First, classification is the process of examining the features of a new object and allocating it to a class. It is intended for building a model that can be applied to unclassified data in order to classify it. For example, credit applicants can be classified as low, medium, or high (Berry, Linoff 2000). Second, numerical prediction involves predicting continuous (or ordered) values for a particular input. For example, marketers/analysts can predict the income of college graduates with 15 years' experience, or the probability of how well a product may sell based on its price (Han, Kamber 2006). There are many techniques used to carry out classification and numerical prediction tasks. Table 2.7 presents the most common techniques used in the direct marketing field. The following subsections provide an overview of these techniques.

| Direct marketing models | Commonly used techniques | Direct marketing applications |
|---|---|---|
| Classification | Decision Trees,<br>Automatic Clustering Detection,<br>Neural Networks,<br>Association Rules,<br>*k*-nearest-neighbour. | To which segment does a customer belong? What characteristics define a segment? How can firms classify customers? |
| Numerical Prediction | Regression Analysis,<br>Neural Networks,<br>Naïve Bayesian. | How can companies predict customer response to targeted offers? What is the effect of prices and promotions on the probability of purchase? How likely is a customer to respond on a given advertisement? How can firms predict customers' relationship duration? |

**Table 2.7: Most common data mining and machine learning techniques for direct marketing applications. Adapted from (Murthi, Sarkar 2003, p.1354)**

***Neural Networks***

A neural network is a well-established data mining technique across a wide variety of industries. The neural network concept is derived from the human brain's connections, and if used correctly in a computer context can generate the ability to learn from data. However, unlike human brains, which can explain their given solution, training neural networks provide weights with no insights into why the solution is valid or how it has been achieved (Witten, Frank 2005). The neural network is the most popular and commonly used data mining technique for developing direct marketing response models and CRM (Ngai, Xiu & Chau 2009); and (Bose, Chen 2009).

A neural network's process is commonly activated by a linear or logistic mathematical function. While a linear function is best suited for numerical prediction (i.e. regression), a logistic function is more appropriate for classification (Nisbet, Elder & Miner 2009).

A neural network can be easily represented using standard linear regression models and numerous other functions such as $z = 3x \oplus 2y - 1$ (Equation 2.6) (Berry, Linoff 2000). In this instance, there are two variables input, x and y. These variables will return a value for z. Figure 2.5 represents a simple neural network, which consists of an input layer and output layer. The input layer is x and y and both have a network node. Most frequently, it is not the actual values of x and y that are given to the input layer but some transformation of them. Weights (i.e. 3 and 2 in Figure 2.5) combine the input layer using a *combination function* and then this is passed to a *transfer function*, which results in the output of the network. Both the *combination function* and the *transfer function* form the unit's *activation function.* Similarly to the input units, the resulting value of the output node's *activation function* is also generally transformed from the real output value (ibid).

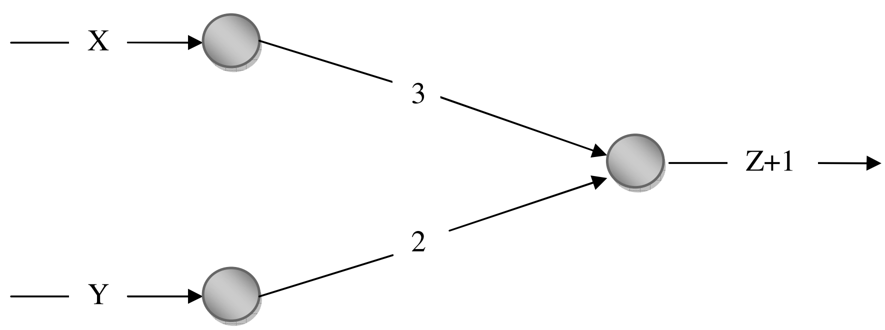



**Figure 2.5: Simple neural network using function $z = 3x \oplus 2y - 1$ (source: (Berry, Linoff 2000, p.122))**

In real-world cases, the neural network model is not as simple as the one shown in Figure 2.5. It is usually made of many more additional layers known as *hidden layers* and their units are *hidden units.* The function of the neural network becomes much more complicated and harder to represent as an equation (Berry, Linoff 2000). The hidden layer nodes provide the ability to model non-linear relationships between the input nodes and output nodes (the decision). This configuration makes a neural network a powerful classifier (Nisbet, Elder & Miner 2009). Backpropagation is a neural network algorithm that performs learning on a multilayer perceptron, and is also known as multilayer feedforward (see Figure 2.6). It is the most popular learning algorithm for neural networks and gained its reputation in the 1980s. It consists of repeatedly processing a dataset of tuples to predict each tuple and compare it with the current target value. The weights for each tuple are modified in order to minimise the mean squared error between the current and predicted value. The name backpropagation was given to this algorithm because the modifications of tuples are made backwards, starting from the output layer, through hidden layers to the first layer (Han, Kamber 2006).

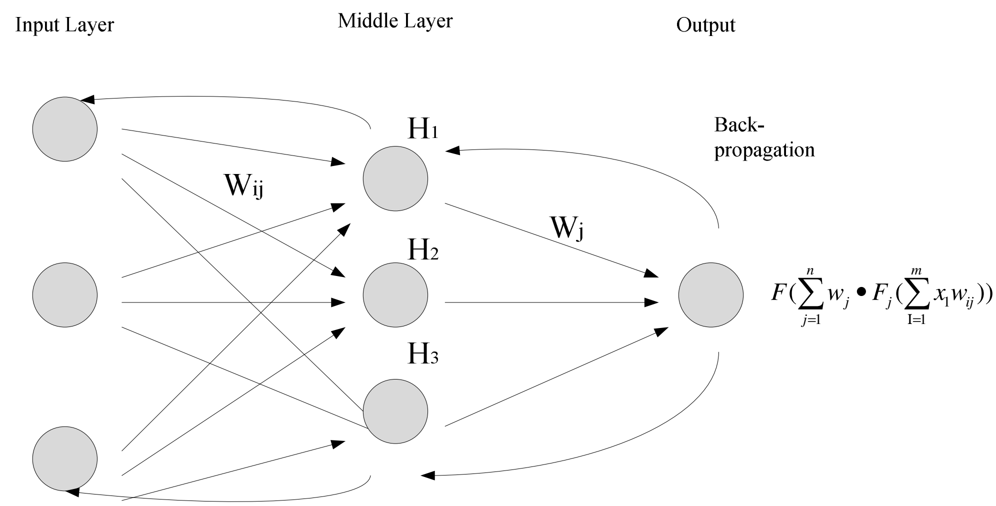



**Figure 2.6: A multilayer perceptron's neural network with hidden layer using backpropagation algorithm (source: (Nisbet, Elder & Miner 2009, p.131))**

***Decision Trees***

A decision tree is the application of simple decision rules to divide a large collection of records into smaller ones (Witten, Frank 2005). This technique uses a set of rules to divide large heterogeneous data into smaller, more homogenous classes based on a particular target variable (ibid). Decision trees can perform both classification and numerical prediction. First, classification trees assign label records to the appropriate classes. They can also give confidence that the classification is accurate. Second, a regression tree (numerical) estimates the value of a numeric variable. For example, it calculates the expected size of claims made by an insured customer (Berry, Linoff 2000). A basic decision trees algorithm uses an *attribute selection measure* to select the best attribute to partition the tuples into distinct classes. It then uses *tree pruning* to identify and remove branches' noise or outliers in the training data (Han, Kamber 2006). There are several decision tree algorithms, but the most popular ones are:

a) **CART:** Classification and Regression Tree,
b) **CHAID:** Chi-Square Automatic Interaction Detection.

There are three principle elements that define a decision tree algorithm and they are as follows:

- Data splitting is performed using rules at a node, according to the data value of one variable.

- A subtree is completed using a stopping rule.
- Each terminal leaf node is assigned to a class outcome (prediction).

As described above, the decision tree technique recursively partitions the data, at each step creating more homogeneous groups (Nisbet, Elder & Miner 2009). The difference between decision tree algorithms lies in the number of splits allowed at each level of the tree, how those splits are selected when the tree is created, and how the tree growth is limited to avoid over fitting (Berry, Linoff 2000). Most decision tree algorithms, such as ID3, C4.5, and CART, adopt an approach in which decision trees are built in a top-down recursive divide-and-conquer manner (Han, Kamber 2006). Figure 2.7 illustrates a typical decision tree model. As shown, a decision tree is a flowchart-like tree where each internal node is represented in a rectangle and leaf nodes are denoted by ovals. The objective of this example tree is to predict whether a customer is likely to purchase a computer (ibid).

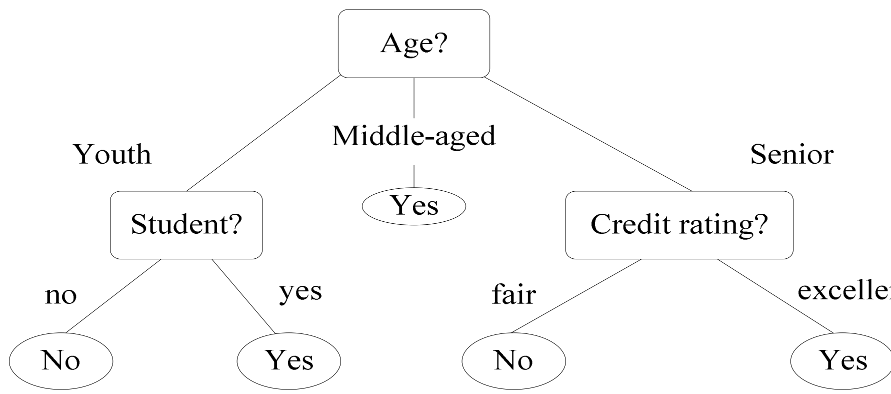



**Figure 2.7: Decision tree for the example where each internal (non-leaf) node represents a test on an attribute. Each leaf node represents a class (Yes or No) to indicate whether a customer is likely to purchase a computer (source: (Han, Kamber 2006, p.291))**

One of the greatest strengths of the decision tree technique is its ability to generate rules that can be translated into natural language or SQL. Indeed, complex decision trees can be dealt with by following the path through the tree to a particular leaf, which makes the explanation of any particular classification or prediction relatively easy (Berry, Linoff 2000).

***Automatic Cluster Detection***

A clustering algorithm aims to identify similar subgroups among a large collection of cases and allocate those subgroups to clusters. The identified clusters are given a sequential number to distinguish between them in the results report (Nisbet, Elder & Miner 2009). Clustering is different to classification and predication, which perform analysis on class-labelled data objects. The clustering technique performs analysis on data objects without referring to a known class label (Han, Kamber 2006). In fact, the class labels are commonly not present in the training data because they are unknown to begin with. A clustering algorithm can generate the class labels by grouping the objects based on the principle of *maximising the intraclass similarity and minimising the interclass similarity*. In other words, it detects clusters with objects that have high similarity in comparison to one another, but are very different to objects in other clusters. The resulting clusters can be viewed as a class of objects from which rules can be extracted. In addition, the clustering technique can facilitate **taxonomy formation**, i.e. the grouping of similar events through the organisation of observations into a hierarchy of classes. Figure 2.8 is an example of cluster analysis, where the aim was to identify homogeneous sub-populations of customers. The results of such analysis can be used to target specific groups for marketing campaigns (ibid).

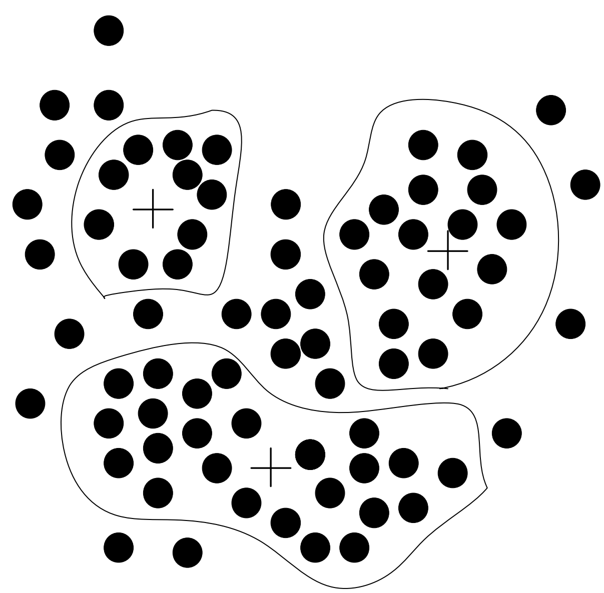


**Figure 2.8: Clusters of homogeneous sub-populations of customers with "+" representing the centre of the location (source: (Han, Kamber 2006, p.26))**

*K-means* is a well-known clustering technique and is widely used in many commercial data mining tools (Berry, Linoff 2000). It works by using a fixed number (*k*) of clusters, and assigning data objects to those clusters in which the means across clusters (all variables) are as dissimilar from each other as possible (Nisbet, Elder & Miner 2009). The calculation of the data objects' difference is performed based on one of several distance measures, which commonly comprise *Euclidean, Squared Euclidean, City-Block, and Chebychev* (ibid). This algorithm application is most suitable and effective when the input data is primarily numeric. For example, supermarket loyalty cards include customer-purchased products such as meat and cereal over the course of some period of time. Transaction data in this case will be numeric. Hence, the application of *k-means* will work quite well. The *k-means* algorithm will determine clusters of customers with similar purchasing behaviour (Berry, Linoff 2000).

***Bayesian Classification***

A Bayesian classifier can predict class membership probabilities using statistical classifiers (Han, Kamber 2006). Its basic operation is to identify the probability a given tuple belongs to a specific class. Naïve Bayesian is a Bayesian classifier algorithm which assumes that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is known as *class conditional independence*. Naïve Bayesian computations are simple and as such are referred to as "naïve" (ibid). Figures 2.9 and 2.10 demonstrate more clearly the concept of Naïve Bayesian (Nisbet, Elder & Miner 2009). First consider the past classification of the objects Green and Red, as shown in Figure 2.9. Naïve Bayesian aims to categorise new cases as they occur. Specifically, it will decide to which class label new cases belong. To achieve this, Naïve Bayesian uses *prior probabilities*, which are based on evidence from previous classifications. In this case, the percentage of green objects is almost twice the size of red. Based on this fact, it is reasonable to believe that any new case (e.g. X) is more likely to be green than red. However, the *likelihood* is also measured by considering a region around X (depicted by the larger circle), which includes a number (to be chosen a priori) of points irrespective of their class labels. Figure 2.10 clearly shows that the circle comprises three red objects and one green object. Hence, it is more likely that X is red than green. Naïve Bayesian combines both *prior probability and likelihood* in order to classify a new object "X" as green or red. This rule is known as *joint posterior probability* and in this example classifies X as red (ibid).

Although Naïve Bayesian is simple, its performance is comparable to sophisticated classification techniques such as decision trees. Furthermore, it has proven to be highly accurate and fast when applied to large datasets. A *Bayesian belief network* is also a Bayesian classifier, which differs from Naïve Bayesian in the way that it allows the representation of dependencies among subsets of attributes (Han, Kamber 2006).

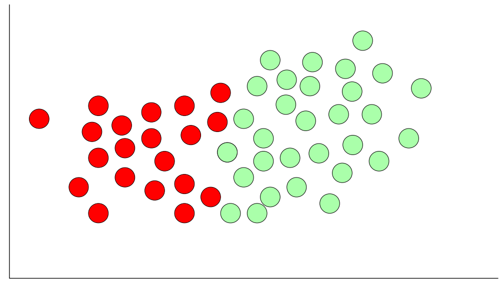


**Figure 2.9: Two groups of objects classified as red and green, plotted in an analysis space by two axes of similarity (two metrics) (source: (Nisbet, Elder & Miner 2009, p.254))**

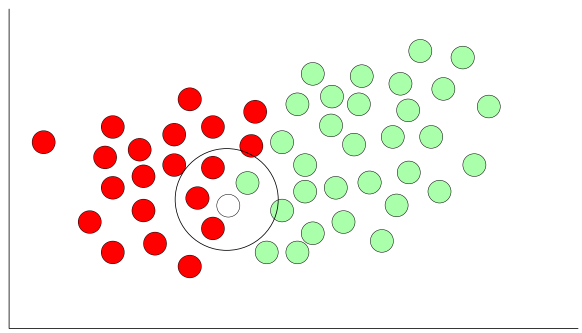


**Figure 2.10: A new object (white ball) position in the analysis space (source: (Nisbet, Elder & Miner 2009, p.255))**

***Other Major Data Mining Techniques***

**Association Rules** are used to detect relationships and associations between values in large datasets. For example, Amazon uses association rules through a *recommender engine* to identify which books are usually bought with a given book. This will provide them with the ability to recommend books to customers who have already purchased a specific book (Nisbet, Elder & Miner 2009). The following is the mined rule used by Amazon's transactional database:

*Buys(X, "book1")⇒ buys(X, "book2")[support 1%, confidence = 50%]*

Variable X represents a customer and 50% confidence means that if the customer purchases "book1", he/she will have a 50% probability of purchasing "book2". The 1% support suggests that 1% of the overall transactions which were analysed showed that "book1" and "book2" were purchased together. In this example, there is a single attribute or predicate (i.e. buys) that repeats (Han, Kamber 2006).

***k-nearest-neighbour*** is a very simple algorithm classifier based on the nearest neighbour approach (Nisbet, Elder & Miner 2009). This method's purpose is to find in the *N*-dimensional feature space the closest object from the training set to an object being classified. Specifically, the *k*-parameter specifies how many nearest neighbours to consider (an odd number is commonly selected to prevent ties). The closeness between objects is defined by the distance along the scale of each variable, which is converted to a similarity measure. The distance is known as the Euclidian distance (ibid).

The **Support Vector Machine** is a new classification method for both linear and non-linear data (Han, Kamber 2006). The tasks involved in SVM are as follows:

1. SVM uses non-linear mapping to transform the training data into a higher dimension.
2. It then searches for the linear optimal separating hyperplane (i.e. decision boundary separating the tuples of one class from another) within the new dimension.
3. Finally, using the support vector (essential training tuples) and *margins* (defined by the support vector), data from two classes can be separated by a hyperplane (ibid).

Figure 2.11 illustrates the basic concept behind support vector machines (Nisbet, Elder & Miner 2009). The original objects (left side of the schematic) are mapped using a set of mathematical functions known as kernels to a new dimension called *feature space*. This process results in the mapped objects (right side of the schematic), which are linearly separable. In other words, SVM avoids constructing the complex curve (left schematic) and find the optimal line that can separate green from red objects (ibid).

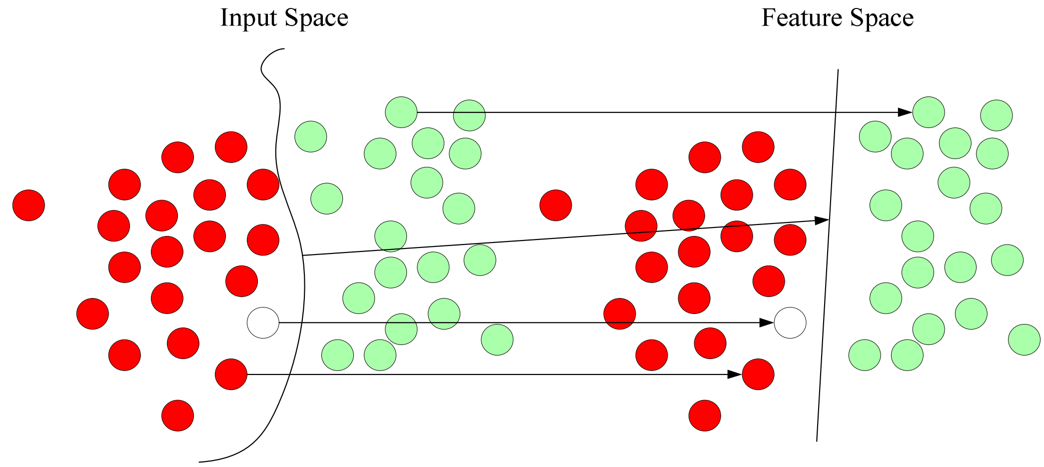



**Figure 2.11: Mapping of input data points to feature space, where linear separation is possible (source: (Nisbet, Elder & Miner 2009, p.164))**

Finally, **Genetic Algorithms (GA), Genetic Programming (GP), and Evolutional Programming (EP)** are growing in popularity in direct marketing and are fairly different from the previously introduced techniques. Indeed, these techniques consist of a search procedure that simulates natural selection and evolution including selection, crossover, and mutation steps. These techniques are appropriate for optimisation problems such as the selection of solicitation targets while meeting business requirements (Bose, Chen 2009).

## 2.4 Research on Direct Marketing

Owing to direct marketing's increasing importance, its issues and debates are being extensively researched by both academics and non-academics. Indeed, there are numerous magazines, newspapers, and academic journals dedicated to the discussion of direct marketing issues and trends. While its growing popularity intensifies and diversifies the debates, the issues of direct marketing mainly relate to two research schools, i.e. *the technical and the social schools.* In fact, the direct marketing community, including people from industry and academics, tends to focus on either the technical aspects of direct marketing or the social one. The technical school focuses mainly on the functions, usability, and performance of the technology, whereas the social school considers the impact of introducing new technology to an organisation's processes as well as on people (Grint, Woolgar 1997).

### 2.4.1 The Technical School

As presented in the previous section, analytical techniques are essential to direct marketing applications. As a result, issues related to statistical techniques, data mining, and machine learning algorithms' performance are among the most popular research topics in direct marketing. Statistical techniques can be powerful and have been used to build models of consumer responses. There are several research studies e.g. (Bodapati, Gupta 2004), (Van den Poel, Buckinx 2005), and (Baumgartner, Hrushka), which attempt to enhance statistical techniques in order to build more accurate direct marketing models. However, statistical techniques can only handle a limited number of variables and have limited explanatory ability. Data mining offers several distinctive benefits when dealing with large noisy datasets. Data mining overcomes statistical techniques' limitations by offering companies the capability to analyse large amounts of variables and providing better explanations (Cui, Wong & Lui 2006). Accordingly, data mining techniques have been the focus of most research conducted into direct marketing models' performance. For example, researchers such as (Ha, Cho & MacLachlan 2005), (Kaefer, Heilman & Ramenofsky 2005), (Cui, Wong & Lui 2006), (Tettamanzi et al. 2007), and (Kim 2009) have proposed improvements to data mining techniques and machine learning methods for more effective direct marketing models. These contributions generally involve merging or adding algorithm enhancements to existing data mining methods. Another area of interest is data preprocessing, where many researchers have provided tools to enhance data preparation for noisy datasets, e.g. (Heilman, Kaefer & Ramenofsky 2003), and (Crone, Lessmann & Stahlbock 2006).

The growing interest of researchers in the technical issues surrounding direct marketing, especially data mining, has resulted in the development of several open source data mining tools (e.g. Orange,[4] Rapid Miner,[5] WEKA, etc.). These tools are intended to provide researchers with an open source platform on which empirical studies can be performed. For example, Waikato Environment for Knowledge Analysis (WEKA) is a tool designed to help researchers focus on practical work rather than theoretical. In fact, (Witten, Frank 2005, p.365) describes WEKA as follows: *"WEKA is a collection of state-of-the-art machine learning algorithms and data preprocessing tools. It is designed so that you can quickly try out existing methods on new datasets in flexible ways. It provides extensive support for the whole process of experimental data mining,*

[4] Orange is an open source system for data visualisation and analysis using data mining (http://orange.biolab.si/ [Last Accessed: March 2011]).

[5] RapidMiner is an open-source system for data mining (http://rapid-i.com [Last Accessed: March 2011]).

*including preparing the input data, evaluating learning schemes, and visualising the input data and the result of learning."*

It also includes a user interface to facilitate navigation and provides users with a platform to compare different methods as well as identify those that are most suitable for the given problem. For instance, (Kim 2009) used WEKA to present the benefits and drawbacks of two ensemble methods compared to single classifiers. Results showed that ensemble models significantly improved prediction performance.

This research focuses on the issues related to the social school of the direct marketing field. The following subsection discusses these issues in detail.

### 2.4.2 The Social School

The social school of direct marketing research is mainly concerned with the variety of 'people', 'technologies', and 'activities', that are involved in the direct marketing process. There are many issues related to the direct marketing process found in the academic literature. These issues can be considered as mainly organisational and managerial. In fact, a direct marketing process is commonly executed as several disconnected activities and operations in many organisations (Vesanen, Raulas 2006). As a result, marketers are hindered by organisational and technical difficulties that complicate their efforts in capitalising on direct marketing benefits. Specifically, the direct marketing process incorporates a complex collection of marketing and business analytics principles, which form an entirely 'self-contained' choice for marketers (Tapp 2008). Moreover, direct marketing can be viewed as an ad hoc process, as it is usually executed in different ways depending on the process objectives (Rao, Steckel Joel H. 1998). For example, the marketing strategy for a company focusing on a particular product is more likely to differ from those focusing on services. The diversity of the parameters affecting the nature of the direct marketing process designed for a given situation makes it difficult to have a uniform way of executing the direct marketing process. In addition, the direct marketing literature does not sufficiently address the fact that the direct marketing process is iterative (Adomavicius, Tuzhilin 2005).

Previous research studies on direct marketing have predominantly related to data mining. Therefore, it is important to investigate the process issues of direct marketing in relation to data mining. First, the process of choosing mining objectives and methods for data mining in a direct marketing context is still unstructured and based mostly on judgement (Shaw et al. 2001). In fact, according to (Bose, Chen 2009) there are no research papers on data mining which provide detailed guidelines on how to extract marketing intelligence for direct marketing. Consequently, the process of extracting marketing intelligence can be difficult, time-consuming and highly uncertain. Second, data mining can be very complex to use and manipulate. For example, marketers with few data mining skills find it difficult to select from datasets if the features are numerous because of the extremely unbalanced class distribution (Ou et al. 2003). In addition, the complexity of advanced analytics such as data mining makes marketers more reluctant to utilise the resulting models due to their difficulty, poor comprehensibility, and trust issues (Kim 2006); and (Cui, Wong & Lui 2006).

Practitioners rarely seem to make use of the enhanced data mining models presented in academic papers to deal with their real-world problems (Martínez-López & Casillas 2008). This is not to question the relevance of those academic papers and their theoretic aspects, but rather their managerial application. Specifically, research endeavours should be directed toward understanding marketing managers' demands, and hence the framework application of data mining models (ibid). Accordingly, (Cui, Wong & Lui 2006) stressed the importance of strong collaboration between management researchers and data mining experts in order to take advantage of the technologically advanced features. The level of complexity of data mining often needs experts in data and statistical analysis. Besides, many organisations are becoming increasingly "Knowledge Centric", thus more employees need access to larger and more varied information sources in order to be more efficient (Cody et al. 2002). In addition, the high level of iteration between data analysts and business users causes the time needed for the overall cycle of collecting, analysing, and acting on enterprise data to be longer. This is due to the fact that business users, although experts in their own area, are still unlikely to be experts in data mining. For example, business users need to provide analysts with data, who in turn have to communicate its results, which may also raise further questions from either side. In brief, there is a significant gap between relevant analytics and users' strategic business needs (Kohavi, Rothleder & Simoudis 2002).

### 2.4.3 Existing Process Models in Direct Marketing

There are many research studies that attempt to overcome direct marketing process issues. Prior to starting the discussion on these research studies, it is important to mention that many of these studies refer to direct marketing as personalisation marketing. As described in Section 2.2.3, direct marketing can be referred to as personalisation marketing as long as the definition of personalisation marketing is similar to that of direct marketing. The following personalisation marketing studies all have a similar definition to direct marketing.

(Vesanen, Raulas 2006) proposed a process view of direct marketing to help marketers manage and execute it more effectively. The authors identified nine elements from the literature that are required to perform the direct marketing process: customer, dialogue with customer, customer data, analyses of customer data, customer profile, customisation, marketing output, delivery of marketing input, and Information Systems (IS). However, the authors excluded IS from their model. Figure 2.12 is a graphical representation of the model.

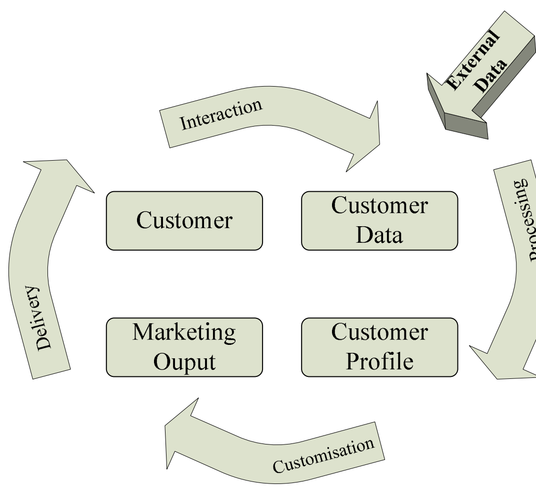



**Figure 2.12: The personalisation process (source: (Vesanen, Raulas 2006, p.10))**

Another direct marketing process model is proposed by (Adomavicius, Tuzhilin 2005). It is an iterative process that comprises three major stages, which constitute a cycle, as shown in Figure 2.13. The following are the main stages along with their sub-stages:

1) Understand: a) Data Collection, and b) Building Consumer Profile.
2) Deliver: c) Matchmaking, and d) Delivery and Presentation.
3) Measure: e) Measuring Personalisation Impact, f) Adjusting Personalisation Strategy.

Unlike (Vesanen, Raulas 2006), who did not consider information systems, (Adomavicius, Tuzhilin 2005) suggested recommender systems, statistics-based predictive approaches, and rule-based systems. The author focuses on recommender systems and argues that they are the most developed matchmaking technologies. These systems can be used, for instance, to recommend products or services to consumers that are similar to the ones they preferred in the past. The main drawback of recommender systems is that they can only support the completion of the first two stages of the direct marketing process, i.e. understand and deliver stages. The authors argue that there are no existing systems capable of supporting the whole process. Similarly, (Murthi, Sarkar 2003) have proposed a three-stage process which includes learning, matching and evaluation. They also suggested the use of recommendation systems.

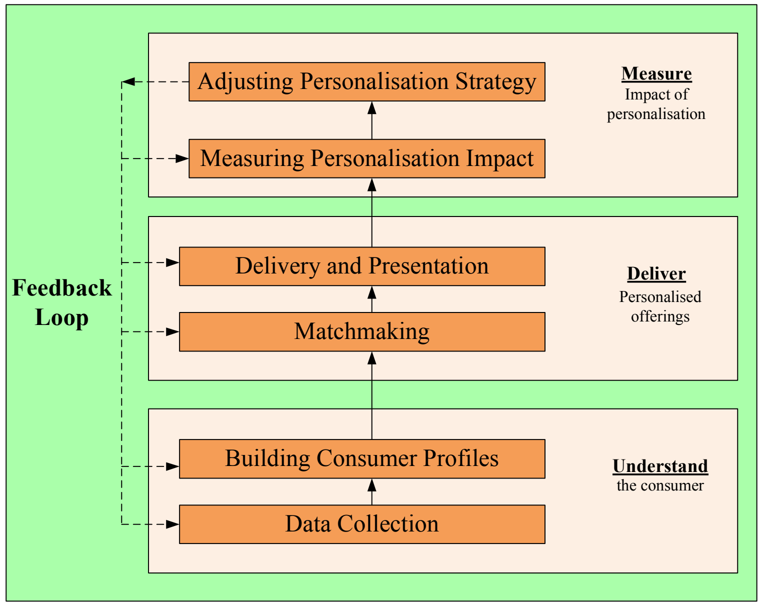



**Figure 2.13: The Iterative Personalisation Process (source: (Adomavicius, Tuzhilin 2005, p.85))**

Finally, (Peltier, Schibrowsky & Schultz 2003) provided a conceptual model for direct marketing, which links the use of a database to the creation of customised and electronic media in order to build interactive and integrated marketing communication. Again, the model does not suggest an information system platform on which the process model can run.

### 2.4.4 Data Mining Methodologies & Knowledge Discovery

There are many direct marketing studies, e.g. (Gersten, Wirth & Arndt 2000), (Shaw et al. 2001), and (Chen, Chiu & Chang 2005), that use data mining methodologies to execute the direct marketing process. However, these methodologies were designed to suit any data mining project within any industry. Therefore, they provide abstract guidelines and marketers/analysts may face many difficulties trying to execute the process. Furthermore, the general nature of these methodologies increases the risk of marketers/analysts having to deal with a lot of uncertainties.

There are three major data mining methodologies that are widely used in both academia and industry: 1) CRISP-DM, 2) Knowledge Discovery in Databases (KDD), and 3) Simple, Explore, Modify, Model, and Assess (SEMMA). These methodologies are process models aimed at facilitating the identification of interesting patterns in large data repositories. They all provide three major steps to find useful patterns including data preprocessing, data modelling, and model assessment. The following is a more detailed explanation of each methodology.

First, CRISP-DM is an industry standard for data mining and predictive analytics (Chapman et al. 2000). It is a process model aimed at facilitating the management of small to large data mining projects. It is applicable to a wide variety of industries. Figure 2.14 shows the steps of the CRISP-DM data mining methodology. In CRISP-DM, the phases take place sequentially, but it is also possible to move back and forth between different phases. The outer circle in the methodology shows the iterative process of data mining projects, which is not always finished once a solution is deployed. The first step is business understanding. Since data is the major factor in data mining projects, data understanding and preparation are the next important steps in data mining projects. The modelling phase consists of selecting one or several data mining techniques to apply for the given problem. Before deployment, it is necessary to more thoroughly evaluate the model and review the methodology steps in order to ensure the achievement of the business objectives. The deployment step is the creation of the model and the follow-up depends on the business requirements.

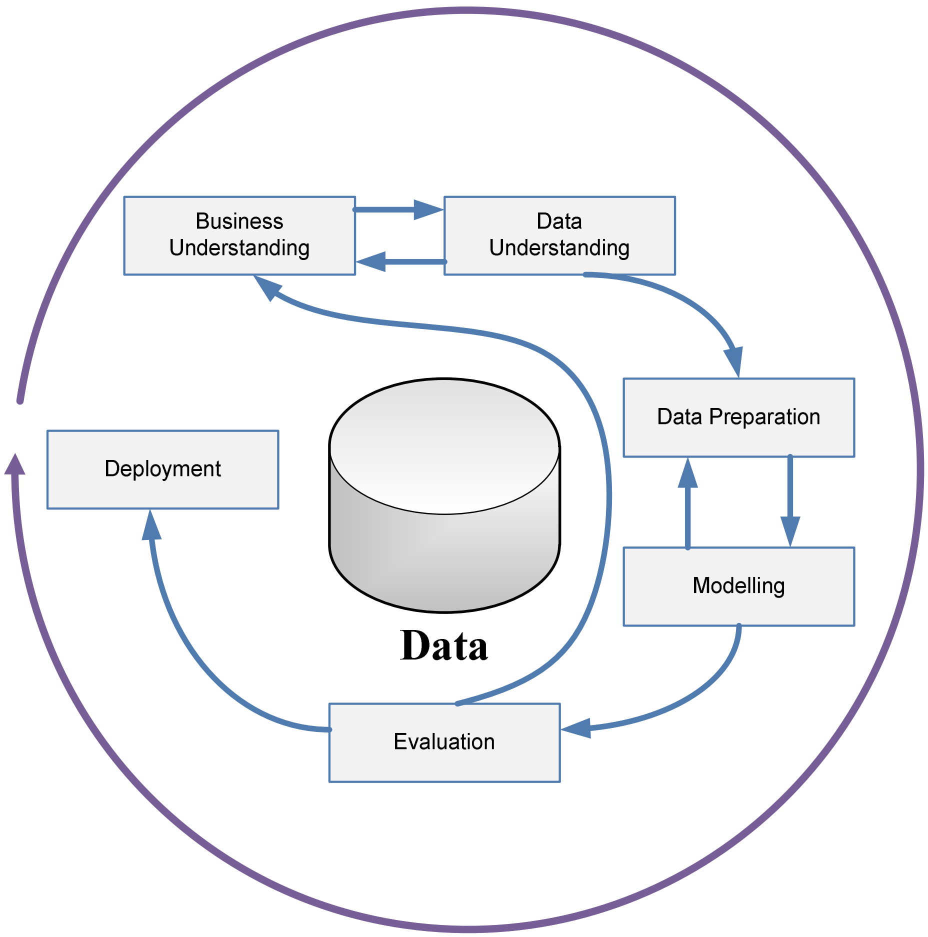



**Figure 2.14: The CRISP-DM data mining methodology (source: (Chapman et al. 2000, p.13))**

Second, KDD can be described as a non-trivial process that aims to identify valid, novel, useful, and understandable patterns in data. It is an interactive and iterative process that goes through several steps, with many decisions made by users. It focuses on the overall process of knowledge discovery from large amounts of data, starting with the storage and accessing of such data, scaling algorithms to huge datasets, interpretation and visualisation of model results, and finally the modelling and support of the overall human machine interaction (Fayyad, Piatetsky-shapiro & Smyth 1996); and (Mitra, Pal & Mitra 2002). Figure 2.15 represents the KDD process with data mining being an essential step in the overall process.

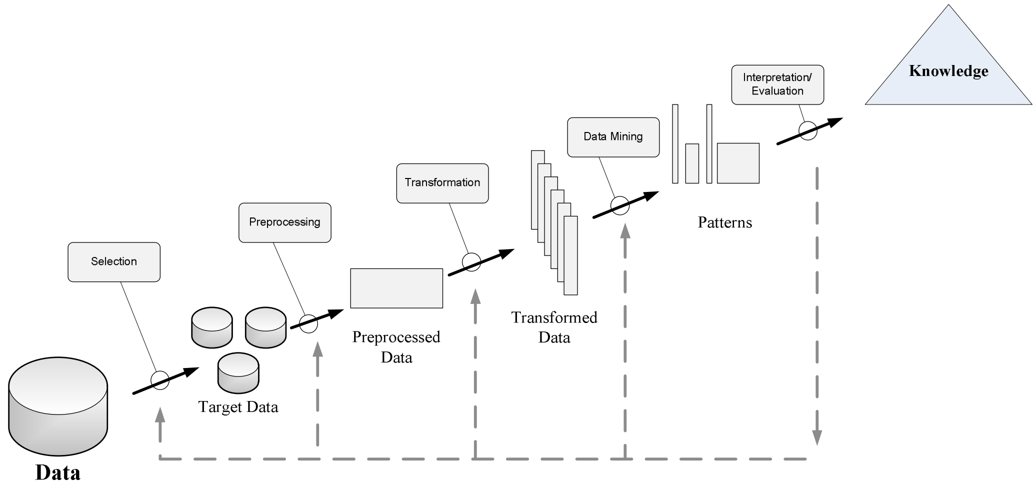



**Figure 2.15: Knowledge Discovery in Databases Process (source: (Fayyad, Piatetsky-shapiro & Smyth 1996, p.41))**

Finally, SEMMA is another popular data mining methodology. It was developed by the SAS[6] Institute to help data analysts perform statistical, data mining, and visualisation tasks. Unlike the other two methodologies, SEMMA is not solely a data mining methodology but a logical organisation of SAS Enterprise Miner functional tools including data mining (SAS Institute 1998).

### 2.4.5 Limitations of Existing Direct Marketing Process Models

There is no correct way in which direct marketing process should be carried out. The style and format of the direct marketing process will differ significantly based on the organisation (Sargeant, Douglas 2001). Several scholars have devoted significant contributions to direct marketing models' performance. However, there has been a lack of research on issues related to the direct marketing process.

The application of data mining methodologies to the direct marketing process is possible. But these methodologies are not specifically tailored to direct marketing; hence a high level of user judgement is required. This makes the process of extracting marketing intelligence using these methodologies difficult to achieve. In fact, current data mining methodologies are general guidelines where marketers can find difficulty achieving focus. Focus is a really important factor for processes that involve the use of data mining. This is because they commonly include a large amount of attributes and values to be considered, which can lead to endless combinations (Mitra, Pal & Mitra 2002). Consequently, the direct marketing process can take longer to finish, leading to

---

[6] SAS is the leader in business analytics software and services, and the largest independent vendor in the business intelligence market (http://www.sas.com/ (Last Accessed: Feb 2011).

an ineffective usage of time and making marketers more uncertain about the overall process outcome.

Most existing direct marketing process models do not consider an information system capable of supporting the activities involved in the process. In the case where an information system is suggested, the author(s) does/do not illustrate its usage in the direct marketing process (e.g. (Adomavicius, Tuzhilin 2005); and (Murthi, Sarkar 2003)). Therefore, users such as marketers and analysts have to identify the appropriate functions within a given information system to execute a specific task.

The following is a summary of the main limitations and problems surrounding the direct marketing process:

1. The direct marketing process incorporates a great variety of marketing concepts and business analytics principles, making it a rather challenging process to perform.
2. Existing direct marketing process models do not provide an integrated information system platform capable of supporting the execution of all direct marketing activities.
3. Data mining methodologies are not specifically tailored to the direct marketing process. Therefore, the use of these methodologies can be difficult, time-consuming, and highly uncertain.
4. Marketers are reluctant to use direct marketing models that are developed using data mining because of their difficulty, poor comprehensibility, and trust issues related to data mining technology.
5. The high level of iteration between data analyst and marketer causes the time needed for the overall cycle of collecting, analysing, and acting on enterprise data to be longer.

## 2.5 Chapter Conclusion

The purpose of this chapter was to provide an overview of the emerging direct marketing studies. It investigated the concepts, practices, and models used in the direct marketing discipline. The chapter also studied the various activities that are commonly used to execute the direct marketing process. Furthermore, it discussed the key issues of the direct marketing process and the need to develop a conceptual framework to tackle these issues.

This chapter was structured into three different parts in order to present a comprehensive background on direct marketing. The first part presented the main factors that contributed to direct marketing's fast development. It also discussed the confusion created by the many terms used to refer to direct marketing, and clarified several misconceptions relating to the direct marketing discipline. The second part provided a description of direct marketing concepts and practices. This included an explanation of the activities that are commonly involved in the direct marketing process. This part also outlined the most common business analytics used to perform direct marketing applications. The final part presented contemporary issues in the direct marketing field. It primarily focused on the social issues of the direct marketing process with particular emphasis on issues in its management and execution. This part also examined existing methods for the direct marketing process and discussed their main drawbacks.

This chapter has clearly illustrated the need to develop a conceptual framework, which integrates the appropriate concepts, practices and technologies in order to execute the direct marketing process more effectively. In fact, it has been demonstrated that organisations are facing many issues to effectively execute the direct marketing process, which can have a negative effect in direct marketing campaign(s) performance.

# CHAPTER 3: DEVELOPMENT OF THE PROPOSED CONCEPTUAL FRAMEWORK

## 3.1 Introduction

The previous chapter provided a review of major direct marketing concepts and practices. It also discussed the main issues surrounding direct marketing in general and its process in particular. The purpose of this chapter is to develop a conceptual framework for the direct marketing process. The development of the conceptual framework will not only attempt to address the issues of the direct marketing process, but also to overcome the limitations of previous direct marketing process models. To achieve this, the conceptual framework will be developed through three steps: define the scope of the direct marketing process, identify the most important components of the direct marketing process, and propose an information system platform.

This chapter is organised into three main sections. The first section investigates the marketing concepts and business analytics to consider when developing the conceptual framework. While addressing these, the section also defines the scope of the direct marketing process in this research.

The second section investigates an information system platform to effectively support the direct marketing process. This results in identifying Business Intelligence (BI) as an appropriate information system platform. This section also provides the reasons BI is selected, and a background on BI concepts and technologies.

The third section starts by introducing the academic and industry literature used to develop the conceptual framework. It then presents a graphical representation of the developed framework; namely the Direct Marketing Process with Business Intelligence (DMP-BI). This is followed by a detailed description of the developed framework. This chapter concludes with a brief summary of the main points covered in the chapter.

## 3.2 Marketing Concepts and Business Analytics

Prior to the development of the conceptual framework, it is important to investigate the marketing concepts and business analytics that need to be considered in the direct marketing process. Indeed, (Tapp 2008, p.9) defined direct marketing as "*a rather* *complex collection of principles and practices which together make up an entirely 'self-contained' choice for marketers*".

This section starts by providing the scope of this study in terms of developing the conceptual framework. It then describes database marketing and business analytics, which represent the foundation of the direct marketing process.

### 3.2.1 Scope of this Study

As mentioned in Section 2.3.1, the direct marketing process involves two fundamental concepts including "understanding customers" and "interacting with customers". Direct marketing always involves analysing customers' data. Most organisations seek to understand customers in order to attract them more efficiently. Interacting with customers is also an essential element for attracting and stimulating customers' responses. However, if both concepts are compared, understanding customers is considered more vital than interacting with customers, as already discussed in Section 2.3.1. This research scope is limited to the study of understanding customers in the direct marketing process.

Understanding customers is a fundamental part of the direct marketing process and comprises two key components; namely a marketing database and business analytics. The following subsections provide a detailed explanation of these two components and their importance in a direct marketing context.

### 3.2.2 Marketing Database

Marketing databases are usually intended to keep customers' data. This data can be the driving factor in companies' direct marketing success or failure. A marketing database is described by (Tapp 2008, p.32) as *"a list of customers' and prospects' records that enable strategic analysis, and individual selections for communication and customer service support. The data is organised around the customer."*. In other words, marketing databases are central to a direct marketing strategy. This is because organisations can no longer know their customers individually. Consequently, marketing databases have become an essential technology to allow organisations to serve customers as individuals, ensuring ongoing dialogues to customise the relationships between marketers and customers (Stone, Jacobs 2008).

Figure 3.1 illustrates a typical marketing database system along with key resources where data can be collected about customers. The growing variety of sources where data can be collected (e.g. Internet) make it more complicated for organisations to create an efficient marketing database (Stone, Jacobs 2008). (Tapp 2008) suggested a set of minimum requirements that a marketing database should have in order to provide a solid foundation for executing the direct marketing process. The following are those minimum requirements:

- customer data,
- purchase (transaction) data: what the customer has bought,
- communication data: campaign history and responses.

The marketing database system presented in Figure 3.1 captures historical and behavioural data from an organisation's most common marketing activities. This can enable marketers to build strong customer relationships. Information is an ever-growing strategic resource, which can be used to drive product, channel, and marketing communications programmes (Stone, Jacobs 2008).

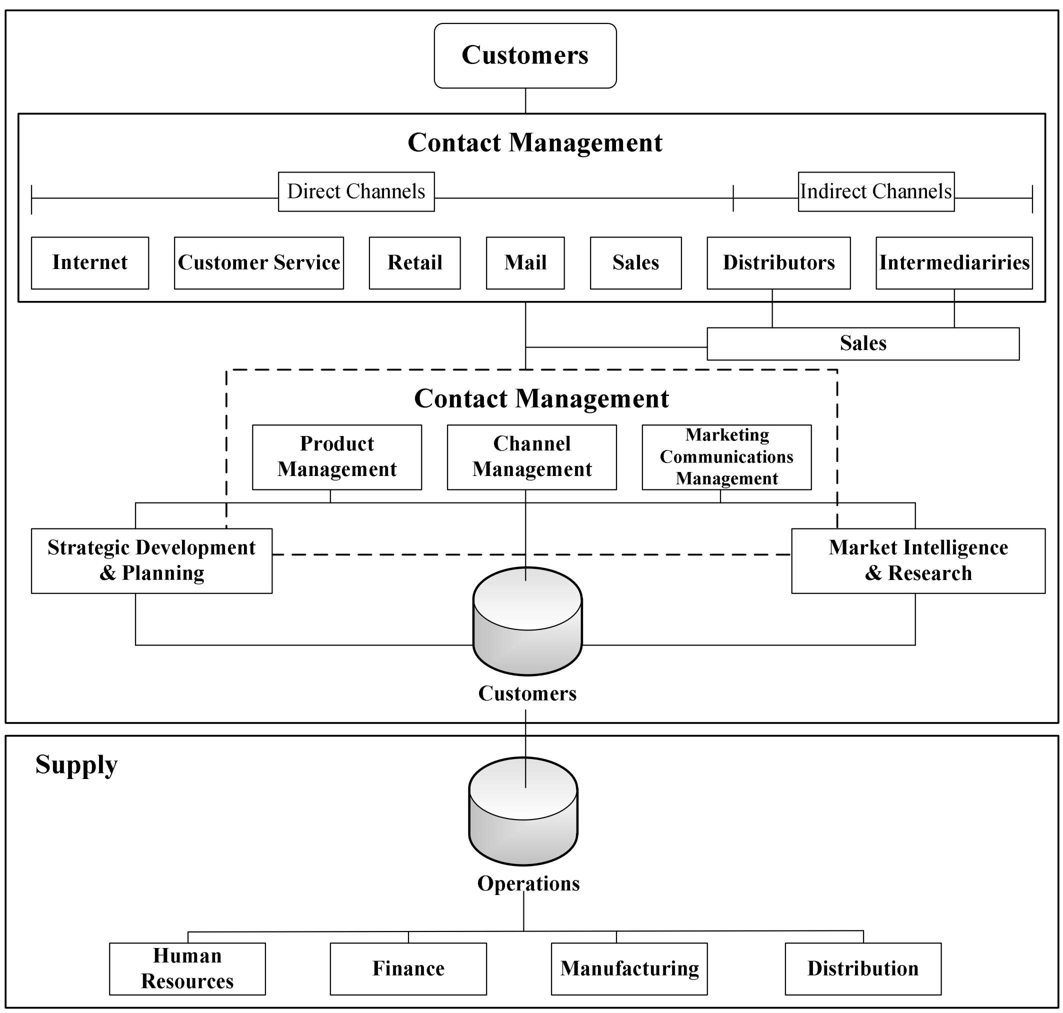



**Figure 3.1: A Marketing Database System Adapted (Stone, Jacobs 2008, p.45)**

### 3.2.3 Business Analytics

As indicated in Section 2.4.1, data mining is one of the most researched business analytics in the direct marketing field. Most data mining techniques' descriptions have been covered in Section 2.3.3. Therefore, this section will focus on discussing the important aspects of data mining to consider when developing the conceptual framework. As mentioned in Section 2.4.2, data mining issues in direct marketing mainly relate to complexity, usability, and interaction. In fact, data mining commonly requires highly skilled analysts to perform the analysis part of the direct marketing process. This affects the process in three aspects: 1) time, 2) management, and 3) trust issues. First, the high level of iteration that may arise between the marketer and the analyst can only delay the process. Second, management of the process is in essence more complicated when multiple users are involved. Third, marketers can have trust issues towards the model results because of the difficulty and poor comprehensibility of the technology, i.e. data mining.

From the above discussion, one can conclude that "ease of use" is a critical aspect of data mining success. In fact, if data mining is accessible to a wider user audience, it can reduce the direct marketing issues that were discussed above. Specifically, if data mining models are more self-explanatory, i.e. easier to interpret through, for instance, better visualisation techniques, marketers can undertake the process without the need for expert knowledge. This will eliminate the high level of iteration between the marketer and the analyst as well as involve fewer users in the process. As a result, issues that are related to time and management (as discussed above) will be significantly reduced. Furthermore, marketers will have greater trust in model results that they understand and have deployed. For example, databases emerged some 40 years ago and were not easy to use for non-expert users. But nowadays, database applications are accessible to a greater variety of users with both expert and non-expert knowledge, and that is one of the main reasons that databases are very successful. This is supported by (Berg, Breur 2007) when stating that for data mining to become effective, ease of model deployment is one of the decisive factors.

## 3.3 Information Systems & the Direct Marketing Process

This section investigates an integrated information system platform to use for developing the conceptual framework. There are numerous IS tools which can be used to support the direct marketing process. However, they are commonly specialised for specific activities within the direct marketing process. In other words, they only partially support the various activities involved in the direct marketing process. For example, there are specialised IS tools for clustering such as BayesiabLab,[7] perSimplex,[8] and CLUTO[9]. These tools could only be used to perform clustering activities, whereas direct marketing may involve other forms of analyses. In addition, clustering techniques may not be appropriate for a given direct marketing process. There are several other tools that are specialised for other forms of analyses such as Social Network Analysis, Text Mining, or Recommendations services (a list of most available open source and commercial software can be found in KDnuggets[10] website).

Organisations cannot afford to purchase a specialised tool for each direct marketing activity. This research proposes the use of Business Intelligence (BI) tools as an integrated information system platform to effectively execute the direct marketing process. The following are the three main reasons for choosing BI tools.

Firstly, BI tools offer organisations software applications, technologies, and analytical methodologies that help them produce accurate and timely marketing knowledge. They also provide them with the functionality, scalability, and reliability of modern database management systems (Cody et al. 2002). This can allow organisations to efficiently store and manage data in relational databases. As a result, organisations' employees can have access to an organised collection of data rather than unstructured ones, which are commonly stored in different spreadsheets.

Secondly, BI tools are well integrated with transactional systems, allowing a close link between operations and analysis activities. This provides efficient access to data which can be analysed, and results can have a direct impact on marketing activities (Kohavi, Rothleder & Simoudis 2002).

---

[7] BayesiaLab is commercial clustering software which uses Bayesian classification algorithms for data segmentation and uses Bayesian networks to perform automatic cluster detection.
[8] perSemplex is also commercial clustering software that performs clustering based on fuzzy logic.
[9] CLUTO is a free and open source clustering software which offers a set of partitional clustering algorithms that treat a clustering problem as an optimisation problem.
[10] http://www.kdnuggets.com/software/index.html: Accessed October 2010. KDnuggets provides top resources on data mining and analytics news, **software**, data and more.

Thirdly, BI tools' advanced analytics offer organisations a better understanding of their market dynamics and customers' future behaviour. Indeed, advanced analytics can analyse large amounts of data and create models that can be used to perform activities such as prediction (Tettamanzi et al. 2007). In fact, most BI tools have a wide range of business analytics techniques which can be used to cover all major forms of direct marketing analysis.

BI tools offer a wide variety of functionalities which support all major activities involved in the direct marketing process. The following subsections provide a comprehensive insight into BI tools' development, description, and capabilities.

### 3.3.1 Business Intelligence Development

The Information Systems market has identified that BI evolved in 15-year cycles. The first cycle started in 1975 and lasted until 1990 and was characterised by production reporting on mainframes. BI was known as Decision Support Systems (DSS) and its level of complexity and high cost restricted its use to only highly trained users and large organisations such as SAS,[11] IBI,[12] and IBM.[13] The second cycle, from 1990 to 2005, saw the beginning of the "modern era" of BI, characterised by end-user-friendlier client/server-based BI tools from vendors such as Business Objects, Cognos, and Hyperion. The term BI originated from Howard Dresner at the Gartner Group in 1989 (Ou, Peng 2006). Eventually, query, reporting, and OLAP technology migrated from client/server to web-based architecture with the development of broad suites of BI. It is also worth mentioning that it was in the early 1990s that BI emerged within the industrial world. Its main purpose was to provide managers with effective tools to better understand their business environment, which could facilitate their decision-making process. Academic interest did not come until later in the mid-1990s and has evolved greatly ever since (Golfarelli, Rizzi & Cella 2004). Figure 3.2 shows the BI development life cycle as well as the level of complexity of each BI function.

When we look back in a few years, we will see that 2005 was another turning point in the BI market and the beginning of a new wave of investment in BI by organisations across all industries. The current market cycle is expected to last until 2020 and will be focused on expanding the reach of BI to more users both inside and outside the

---

[11] SAS is a provider of enterprise intelligence software and services.
[12] IB (Information Builders) also provides business intelligence for business users.
[13] IBM is an IT and consulting services company.

organisation and a move to automate more decision processes by combining BI functionalities.

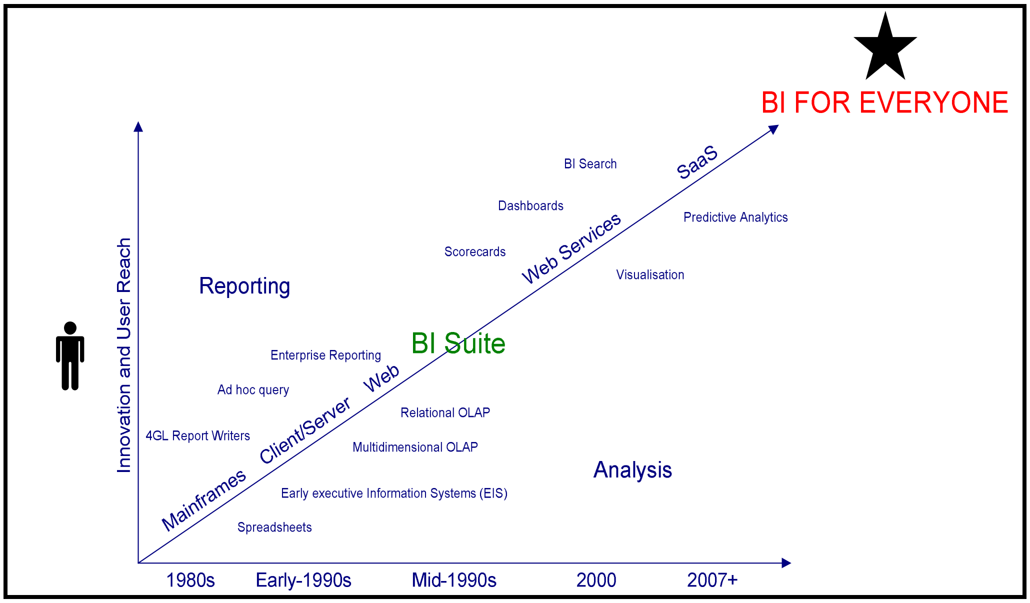



**Figure 3.2: Evolution of BI tools (source: (Howson 2008, p.10))**

### 3.3.2 BI Overview

The development of new technologies and the increasing level of sophistication in computational and analytical capabilities, as well as computer hardware and software, resulted in enterprises generating a huge variety of information inputs (Negash 2004). Figure 3.3 summarises the most common information formats in modern organisations. As illustrated in Figure 3.3, BI tools can use organisational data to support decision making.

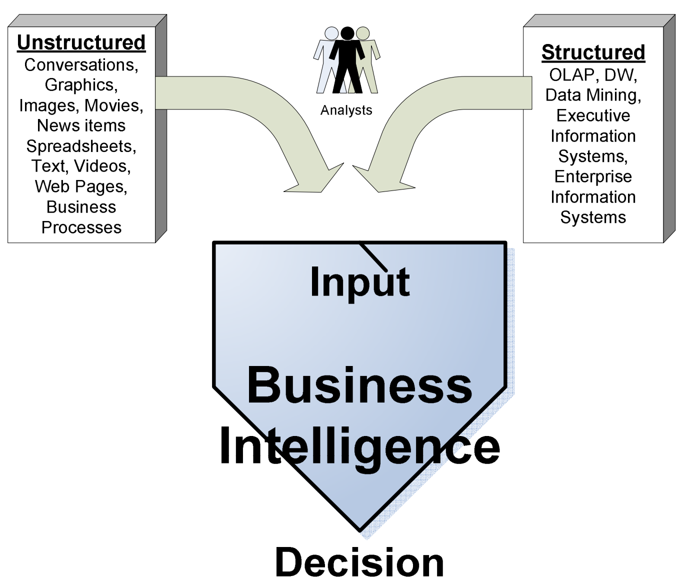



**Figure 3.3: Inputs to BI tools (source: (Negash 2004, p.178))**

(Howson 2008, p.2) describes BI as follows: *"Business Intelligence allows people at all levels of an organisation to access, interact with, and analyse data to manage the business, improve performance, discover opportunities, and operate efficiently."*. In order for BI tools to be effective, BI has to be characterised by the following features: real-time data warehousing, data mining, automated anomaly and exception detection, proactive alerting with automatic recipient determination, seamless follow-through workflow, automatic learning and refinement, and data visualisation (Negash 2004). Figure 3.4 represents a BI system that queries a data source, and uses well-founded approaches such as Online Analytical Processing (OLAP) and data mining to analyse information in the source and report results (Ortiz 2002).

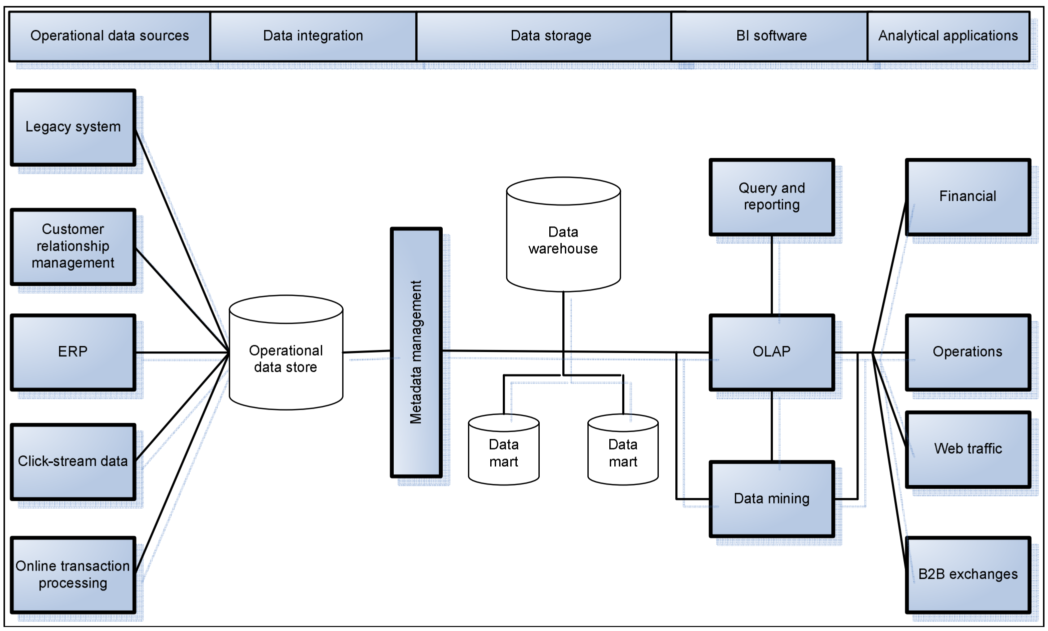



**Figure 3.4: BI process diagram (source: (Ortiz 2002, p.12))**

### 3.3.3 BI Capabilities

There are two main analytical methodologies used in the BI tools market. First, Query, Reporting, and Analysis (QRA) consist of analysis tools, i.e. dashboards that support ad hoc data access and report building. Second, Advanced Analytics employs data mining, statistical software, and knowledge discovery that includes different techniques, i.e. classification and clustering in data mining to obtain valuable information and knowledge from large amounts of data often thought of as useless. These techniques can be used in applications such as market basket analysis and loan applications (Vesset, McDonough July 2006).

BI tools' leading vendors include SAP, IBM, SAS, Oracle, and Microsoft. BI vendors' market share is greater than $10 million in worldwide BI tools revenue (Vesset 2010).

Most BI tools' vendors provide six main functions: 1) reporting, 2) data integration, 3) visualisation techniques, 4) database management, 5) analytical techniques, and 6) dashboards & scorecards (Butler Group 2006); (Harinath, Quinn 2006); and (Howson 2008). In other words, even if BI tools differ, they will only differ in small features rather than in their key capabilities.

## 3.4 The Conceptual Framework for the Direct Marketing Process with Business Intelligence (DMP-BI)

This section introduces the Direct Marketing Process with Business Intelligence (DMP-BI) framework. The conceptual framework is developed after an extensive review of the direct marketing and business intelligence literature. The main focus of the review is to explore and identify appropriate direct marketing concepts and business intelligence practices in order to develop an effective conceptual framework. Specifically, the review aims to identify: 1) the stages of the direct marketing process, 2) the main activities in each stage, 3) the expected outcome of each stage, and 4) the BI functions to support the process in general, and each stage and activity in particular. The following subsections present key direct marketing and business intelligence literature used to develop the conceptual framework. They also provide detailed descriptions of the stages and functions that constitute the conceptual framework.

### 3.4.1 Process Models used to Develop the DMP-BI Framework

This subsection describes the identified literature used to develop the structure and organisation of the DMP-BI framework. Figure 3.5 describes the DMP-BI framework graphically. The overall structure of the DMP-BI framework is a synthesis of five process models. The first model is the CRISP-DM methodology, which is aimed at facilitating the management of small to large data mining and predictive analytics projects in various industries (Chapman et al. 2000). This model was deemed appropriate because a typical direct marketing process involves data analysis tasks (Tapp 2008); and (Bose, Chen 2009). It is also due to the fact that data analysis in direct marketing is commonly performed using data mining and predictive analytics (e.g. (Ou et al. 2003); (Changchien, Lee & Hsu 2004); (Chen, Chiu & Chang 2005); (Reutterer et al. 2006); (Wang, Hong 2006); and (Bose, Chen 2009)) and CRISP-DM methodology is the business analytics industry standard (Chapman et al. 2000).

Four other direct marketing process models are used to develop the structure and organisation of the DMP-BI framework including the *Personalisation Process* (Adomavicius, Tuzhilin 2005, p.85), *The Process of Personalisation* (Vesanen, Raulas

2006, p.5), *A Typical Direct Marketing Process* (Tapp 2008, p.13), and *A Systems Perspective of Direct Marketing Models* (Bose, Chen 2009, p.2). These models were selected because they provide the common activities of a typical direct marketing process as well as the order in which these activities are executed.

Based on the above process models, the overall structure of the DMP-BI framework is composed of four stages: 1) Direct Marketing Objectives, 2) Data Preparation, 3) Data Modelling, and 4) Direct Marketing Planning. The first stage was derived from (Kolter, Armstrong 2008); and (Tapp 2008), which set 'direct marketing objectives' as the starting point of the process. The second and third stages are based on data mining methodologies, which were adapted for direct marketing by the following researchers: (Shaw et al. 2001); (Heilman, Kaefer & Ramenofsky 2003); (Harinath, Quinn 2006); (Kim 2006); and (Ngai, Xiu & Chau 2009). The final stage was adapted from (Rao, Steckel Joel H. 1998); and (Tapp 2008). Table 3.1 provides a summary of the literature used to develop the DMP-BI framework.

### 3.4.2 BI Functions Used

BI tools are used to support the DMP-BI framework from a system perspective. They provide a complete set of functions capable of supporting each stage involved in the DMP-BI framework. The following BI functions have been selected to support each stage of the framework (Butler Group 2006); (Harinath, Quinn 2006); and (Howson 2008).

1) Reporting: Report Building, Enterprise Reporting, Report Management, and Report Publishing,
2) Data Integration,
3) Visualisation Techniques,
4) Database Management,
5) Analytical Techniques,
6) Dashboards & Scorecards,
7) BI Search.

These BI functions support the execution of the four stages involved in the DMP-BI framework, as shown in Figure 3.5.

This study does not focus on the BI tools' vendors, but on the fundamental functions provided by them. Indeed, most BI tools' vendors provide the seven main functions mentioned above. In other words, even if BI tools differ, they will only differ in small features rather than in their key capabilities. The following subsections provide detailed explanations of how each stage and supporting BI functions of the DMP-BI framework are used to execute a direct marketing process.

| **Direct Marketing Process Activities** | **Sources** |
|---|---|
| Overall Structure & Organisation of the framework. | (Chapman et al. 2000); (Adomavicius, Tuzhilin 2005); (Vesanen, Raulas 2006); (Tapp 2008); and (Bose, Chen 2009). |
| 1. Direct Marketing (DM) Objectives<br>- A new product or service is introduced, or new channel is needed<br>- List of Objectives | (Tapp 2008); and (Kolter, Armstrong 2008). |
| Data Preparation<br>- Attribute Selection<br>- Data Sampling<br>- Data Key Facts Summary | (Fayyad, Piatetsky-shapiro & Smyth 1996); (Shaw et al. 2001); (Mitra, Pal & Mitra 2002); (Ou et al. 2003); (Chen, Chiu & Chang 2005); (Kim 2006); (Tapp 2008); (Bose, Chen 2009); and (Case Study I). |
| Data Modelling:<br>-Model Deployment:<br>o Classification<br>o Estimation<br>o Prediction<br>o Association Rules<br>o Clustering<br>o Description<br>-Model Assessment | (Fayyad, Piatetsky-shapiro & Smyth 1996); (Shaw et al. 2001); (Mitra, Pal & Mitra 2002); (Ou et al. 2003); (Changchien, Lee & Hsu 2004); (Chen, Chiu & Chang 2005); (Kim 2006); (Wang, Hong 2006); and (Bose, Chen 2009). |
| Direct Marketing (DM) Planning:<br>-Plan Direct Marketing Campaign<br>-Evaluate Direct Marketing Performance | (Ou et al. 2003); (Changchien, Lee & Hsu 2004); (Chen, Chiu & Chang 2005); (Wang, Hong 2006); (Bose, Chen 2009); and (Lin, Hong 2009). |
| BI Functions: Reporting, Data Integration, Visualisation Techniques, Database Management, Analytical Techniques, Dashboards & Scorecards, and BI Search. | (Butler Group 2006); (Harinath, Quinn 2006); and (Howson 2008). |

**Table 3.1: DMP-BI Conceptual Framework Literature Summary (source: Author)**

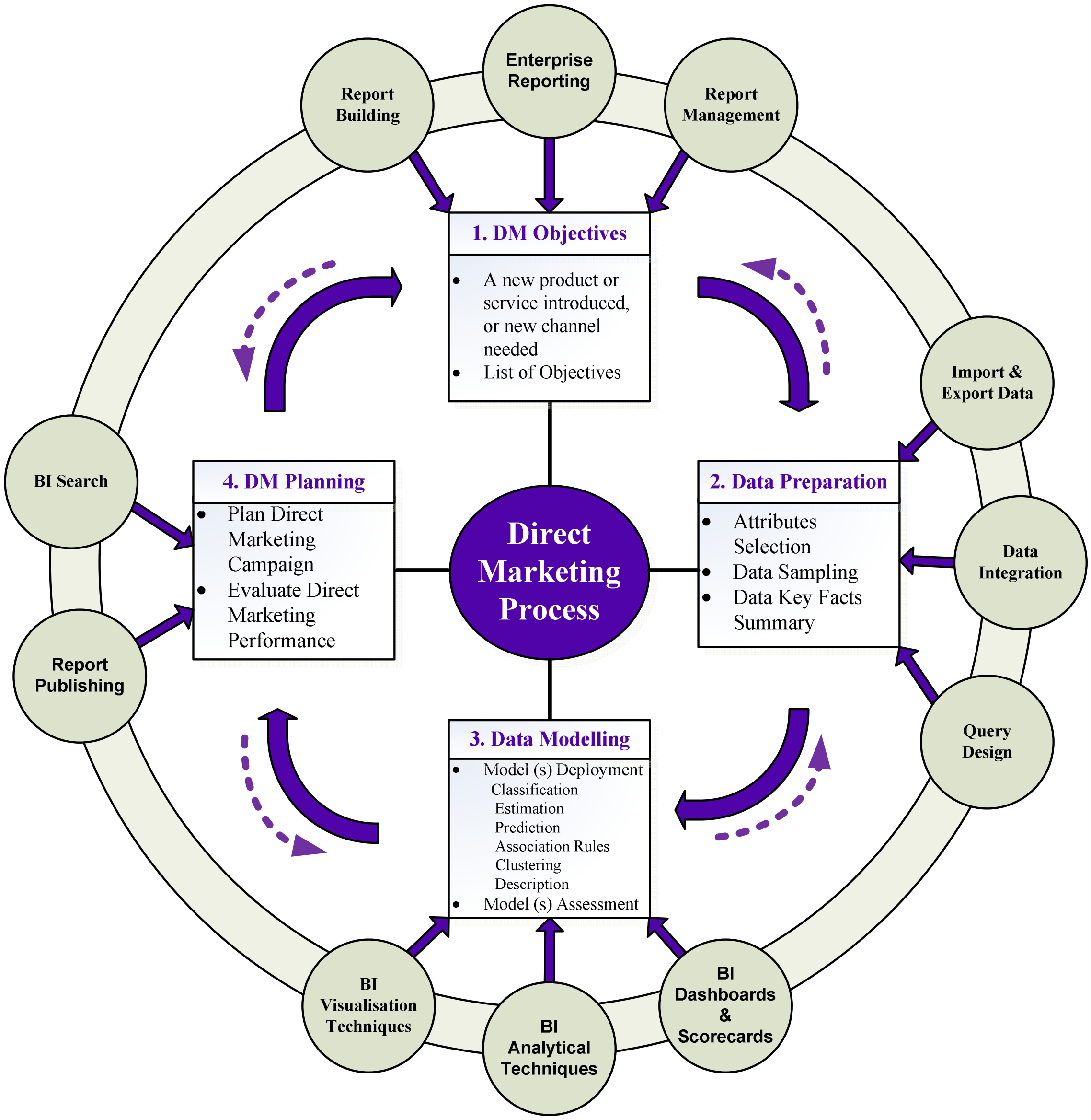



**Figure 3.5: Direct Marketing Process with Business Intelligence (DMP-BI) Framework (source: Author)**

### 3.4.3 Direct Marketing Objectives

This stage is commonly initiated by an internal or external entity, i.e. new product, service, or channel, which needs to be marketed. An internal entity is usually the outcome of a direct marketing process. Indeed, the DMP-BI framework is iterative. Hence, the outcome of a direct marketing process can become the starting point for a new process. An external entity is commonly provided by, for instance, the product development department. For example, when a new product or service is introduced, or when a channel is needed, the department concerned sends a request for a direct marketing campaign to be planned (Tapp 2008).

This stage aims to define the objectives of the direct marketing process. While addressing the direct marketing objectives, marketers/analysts need to consider key questions: what product/service will be offered, how they will be positioned/channelled, and what customers will be targeted (Rao, Steckel Joel H. 1998). Table 3.2 provides a useful insight on ways to define the direct marketing scope/objectives.

| **Market Scope** | **Decision** | **Consumers' behaviour questions** |
|---|---|---|
| New Products | Product Design | What benefits do customers want in new products? |
| | Positioning | How are existing brands perceived in the marketplace? |
| | Price points | How price-sensitive are customers? |
| | Customer targeting | What classes of customer want certain benefits and have certain perceptions? |
| Existing Products | Product Modification | What benefits do customers want from specific products? |
| | Positioning | How are existing brands (including ours) perceived in the marketplace? |
| | Promotions | How price-sensitive are customers? |
| | Customer selection | What customers would be most receptive to our offerings? |

**Table 3.2: Direct marketing scope identification. Adapted from (Rao, Steckel Joel H. 1998, p.24)**

After defining the process objectives, marketers/analysts are required to identify a suitable data source to achieve these objectives. This commonly entails the search of organisations' databases or other data sources to select an appropriate dataset for analysis. There are numerous variables which can help marketers/analysts achieve direct marketing objectives. Transaction variables are commonly the most useful because they can reveal customers' behaviour (Bose, Chen 2009). This allows a better understanding of customers' buying habits, thus helping maximise direct marketing campaigns toward various customer segments. Table 3.3 provides a summary of the most common variables used in the direct marketing context.

| Category | Variables | Importance | Accessibility | Variability |
|---|---|---|---|---|
| Customer | Demographic, lifestyle, socio-graphic | Low | External | High |
| Transaction | Transaction records, feedbacks, web browsing, log files | High | Internal, accumulating | High |
| Product | Size, colour, price, design style | High | Internal | Low |

**Table 3.3: Direct marketing data categories and variables. Adapted (Bose, Chen 2009, p.5)**

The outcome of this stage is a list of objectives and a dataset to execute the direct marketing process. In this stage, BI reporting is used to report the objectives of the direct marketing process to share with departments that are concerned or for future record.

### 3.4.4 Data Preparation

The second stage of the framework requires marketers/analysts to prepare data for the data modelling stage. The direct marketing process often exposes marketers/analysts to a large amount of attributes and values within a dataset. This may result in a high number of possibilities to formulate from the attributes and values. Therefore, it is important to achieve attribute focus by selecting the most important ones (Mitra, Pal & Mitra 2002).

This stage typically involves three key activities: attribute selection, data sampling, and data key facts summary. To begin with, attribute selection is choosing a set of attributes by removing attributes that are redundant or not relevant for the given objectives (Kim 2006). This aims to improve the focus, comprehensibility, and the quality of the resulting model. Next, data sampling entails choosing a small number of examples to deploy in a direct marketing model. There are several advantages related to data sampling including increasing model accuracy and robustness. Marketers/analysts need to perform data sampling very carefully as biased sampling can affect the true patterns in the original data and lead to incorrect solutions (ibid). (Heilman, Kaefer & Ramenofsky 2003) stressed the importance of conducting a sensitivity analysis to select the optimal amount of data to use for the data modelling phase. Direct marketing objectives are the best references for selecting the appropriate amount of data, and hence achieve efficient data sampling. Finally, the data key facts summary consists of the mean for each attribute selected for analysis. This data summary can provide useful information to perform cross-analysis activities with the deployed models. It is also recommended that the minimum, maximum and standard deviation are calculated for each selected attribute. Again, these could be used for cross-analysis purposes. This activity was included from conducting the first case study (for more details refer to Section 5.3.1).

Data preparation activities are commonly performed using BI database management functions such as data integration and query design. In fact, the BI database management platform offers all major capabilities for adding, changing, and manipulating dataset attributes. The primary purpose of data preparation is to provide a solid foundation on which data models can be deployed. The direct marketing model's accuracy and reliability relies heavily on the integrity of the prepared data.

### 3.4.5 Data Modelling

Data modelling for direct marketing is commonly performed using data mining techniques (see Section 2.4.1). At this stage, advanced modelling is required in order to ensure marketers/analysts have a strong platform to search for interesting patterns and extract marketing intelligence. According to (Berry, Linoff 2000), direct marketing modelling using data mining can be classified into two main approaches: 1) directed data mining approach, and 2) undirected data mining approach. Firstly, directed data mining involves the tasks of classification, estimation, prediction, and profiling. The objective of directed data mining methods is to search and find patterns that explain a specific result. Secondly, undirected data mining entails the tasks of clustering, finding association rules, and description. The objective of undirected data mining is to determine whether or not the patterns identified are relevant. The following is a brief insight[14] into directed and undirected data mining tasks in direct marketing:

- Classification is the process of examining the features of a new object and allocating it to a class. It is intended for building a model that can be applied to unclassified data in order to classify it. For example, customers can be assigned to predefined segments.

[14] Note: refer to Section 2.2.3.2 for a more detailed description of data mining techniques in direct marketing.

- Estimation performs similar tasks to classification. While classification uses discrete outcomes such as yes or no for credit cards, mortgages, or car loans, estimation assigns a value, such as a number between 0 and 1, to estimate the probability of customer responding positively to an offer, for instance.
- Prediction is similar to classification or estimation, except that the data is classified based on future customer behaviour. For example, marketers/analysts can predict which customers are mostly likely to leave in the next six months.
- Association Rules is the process of identifying which things may go together. For example, supermarkets might want to know which products are sold together in a single shopping basket. This is also known as basket analysis and can help supermarkets plan their shelving of products more effectively.
- Clustering is the process of segmenting a diverse group into more similar subgroups or clusters. The difference between classification and clustering lies in the fact that clustering does not rely on predefined classes. Data is grouped together based on self-similarity. For example, a cluster of symptoms can represent a particular disease.
- Description and Visualisation techniques are simply used to describe and visualise a dataset with the aim of providing a better understanding of the data. They also provide a comprehensible visualisation of direct marketing models for interpreting results.

There are no universally best data mining techniques and selecting a specific one or combination of methods needs subjective judgements on the suitability of an approach (Mitra, Pal & Mitra 2002). This stage proposes two approaches to facilitate marketers/analysts selection of data mining technique(s). The first approach is based on the direct marketing objectives. Marketers/analysts need to map the direct marketing objectives with one of the two categories provided in Table 2.7. After that, the data mining and machine learning methods that are suggested on the table are used to deploy the direct marketing models (Murthi, Sarkar 2003). The second approach is mass modelling and entails using multiple, or a combination of, data mining techniques. According to (Gersten, Wirth & Arndt 2000), mass modelling is a pragmatic method yielding good results in real-world marketing projects. In fact, the direct marketing process often requires a combination of data mining models to achieve the process objectives. For example, clustering can be used to perform customer segmentation for the initial classes of the data modelling. In this case, clustering only supports preliminary direct marketing analyses, and prediction of customers' behaviour is the main purpose of the analyses (Ngai, Xiu & Chau 2009). The following direct marketing activities are typical examples of analyses involved in the direct marketing process (Shaw et al. 2001):

a) Customer profiling: it uses customer attributes, such as demographic and purchase transactions, in order to perform dependency analysis, class identification, and concept description.
b) Deviation analysis: it allows marketers/analysts to detect anomalies and changes. For example, deviation analysis can reveal results that occurred on recent price changes and promotions, and can evaluate the impact of the changes.
c) Trend analysis: it identifies patterns that occur continuously over a period of time. There are short-term trends, such as an increase in sales during a promotion campaign, or long-term trends, such as a slow drop in sales over a few years. Trends are usually used to perform forecasting of future sales.

After the model(s) deployment(s), marketers/analysts have to evaluate the accuracy of the results. Lift chart models can be used to estimate the accuracy of data mining models. This technique is a graphical model, where the x-axis represents the percentage of data used for prediction, and the y-axis represents the percentage of prediction correctness. A model is considered reliable if its accuracy is over 50% (Harinath, Quinn 2006).

The final activity of this stage is to select the models to perform the analysis. It is rational to first use the most accurate model to perform the analysis. It is also suggested that if the most accurate model provides enough information to achieve the objectives, the other model(s) should not be considered for further analysis. Similarly, if the most accurate model does not fulfil the process objectives, the following most accurate model should be subject to further analysis and so on until the process objectives are achieved.

BI analytical techniques are used to complete this stage. BI tools provide a variety of analytical techniques ranging from standard statistical techniques to more advanced data mining techniques. Model assessment is commonly performed using lift chart methods and most BI tools provide this capability. This stage can be quite tedious in terms of interpreting the model(s) results. However, it is a critical stage in the direct marketing process, because it is where patterns are identified. The following stage involves the planning of a direct marketing campaign using the model(s) results.

### 3.4.6 Direct Marketing Planning

Direct marketing model(s) analysis provides marketers/analysts with information that can help them develop customised products, services, channels, and communications for customers. For example, particular products can be marketed to customers with different colour preferences, discount coupons, and customised leaflets. Customers are then contacted through personal media (e.g. direct mail or post), in order to maximise the direct marketing impact. The outcome of the campaign responses is added to the marketing database (Kolter, Armstrong 2008). It is common for companies to have an end date for their promotions. Therefore, updating the marketing database should be performed after the deadline of the promotions has passed. Finally, the responses data can be used to evaluate the campaign impact. This will show whether the direct marketing campaign (e.g. targeting specific customer segment) was effective (Tapp 2008). Typically, it is done by reviewing the number of responses achieved.

There are cases where the direct marketing process leads to the discovery of more patterns than originally intended. This can result in formulating a new direct marketing process in order to further investigate the patterns discovered. This is the reason the direct marketing process is iterative. BI search and reporting can be used in this stage to search for previous direct marketing or for documentation purposes.

## 3.5 Chapter Conclusion

This chapter identified the stages, activities, and technologies needed to execute the direct marketing process. It described the development process of the DMP-BI framework to tackle the issues of the direct marketing process.

The chapter started by discussing the appropriate concepts and technologies to consider when developing the conceptual framework. These concepts and technologies include understanding customers, the marketing database, and business analytics. The chapter then proposed business intelligence tools as an integrated system platform. It also presented the main reasons why BI tools were selected. After that, an overview of BI tools was provided. The final section presented the academic and industry literature that was explored and adopted to develop the conceptual framework. It concluded by explaining how marketers/analysts can use the DMP-BI framework to execute a direct marketing process.

This chapter investigated the main direct marketing concepts and business analytics practices and identified an integrated system platform to help execute the direct marketing process. The chapter also presented the DMP-BI framework, which was developed to address the key issues facing organisations to effectively execute the direct marketing process.

# CHAPTER 4: RESEARCH METHODOLOGY

## 4.1 Introduction

This chapter explains the research methodology and the research design of this thesis. To achieve the research objectives and evaluate the conceptual framework, this chapter is organised as follows. Section 4.2 describes the research philosophies and approaches used in the IS discipline. It discusses the main characteristics of the three research philosophies in IS, namely positivist, interpretivist, and critical. It then provides the main reasons for adopting positivist for this research. Next, it gives an overview of quantitative and qualitative research and the justifications for adopting the latter approach. Section 4.3 describes the case study research method and the justifications for selecting it to perform this research. Section 4.4 presents the research design used in this thesis. Section 4.5 explains the data collection process and why organisations' datasets were used to evaluate the conceptual framework. Section 4.6 introduces the analytical strategies used to evaluate the case studies. Section 4.7 addresses the different reporting structures available for case studies. It also presents the two types of compositional structure that are used in this research. Section 4.8 outlines the ethical considerations related to this thesis. Finally, a summary of the chapter is provided.

## 4.2 Research Philosophy

A research philosophy is a set of basic beliefs about ontology (what is the form and nature of reality?), epistemology (what is the relationship between the inquirer and the known?), and methodology (how can the inquirer gain knowledge of the world?) (Guba, Lincoln 1994). These philosophical assumptions are known as paradigms. All research is based on four major paradigms of choice in informing and guiding the inquiry: positivism, postpositivism, critical theory, and constructivism (ibid). In an information systems context, ontology refers to the fundamental units which are assumed to exist in the object system (Hirschheim, Klein & Lyytinen 1995). Epistemology is concerned with how researchers inquire into the object system and see phenomena in it. This influences the format in which the knowledge or perception of the object system is represented (ibid). Methodology is the process by which knowledge is to be generated (Mingers 2001).

All research methods in IS are based on three broad-brush philosophical paradigms: 1) positivist, 2) interpretivist, and 3) critical research. First, positivist can be subdivided into positivist and post-positivist approaches. Second, interpretivist comprises approaches based on the philosophy of hermeneutics, phenomenology or constructivism. Third, critical research encompasses a whole range of different approaches including Marxism, feminism and queer research (Oates 2006). The reason behind having so many paradigms lies in the fact that philosophical assumptions in IS draws upon a very wide range of disciplines including technology, psychology, economics, sociology, mathematics, linguistics, and semiotics. These disciplines comprise very broad and different research traditions (Mingers 2001). Therefore, debates, discussions, and questions on which research methods are more suitable for the information systems discipline have been a focus of concern for some time (ibid). In other words, philosophical paradigms research is an academic discipline in its own right, with extensive literature. Hence, this chapter will only discuss in detail the paradigm used to perform this research. The following is a brief description of the three philosophical paradigms in IS research.

First, a positivist approach in IS research is usually suitable if there is evidence of formal propositions, quantifiable measures of variables, hypothesis testing, and the drawing of inferences about phenomena from a representative sample to a stated population (Orlikowski, Baroudi 1991) cited by (Klein, Myers 1999). One of the most popular examples of the positivist approach is the work of (Yin 1994) and (Benbasat, Goldstein & Mead 1987) on case study research (Klein, Myers 1999).

Second, IS research can be categorised as interpretivist when researchers' knowledge of reality is gained only by social constructions such as documents, tools, and shared meaning. Interpretivist research focuses on understanding the complexity of human sense-making as the situation emerges. In other words, it does not predefine dependent and independent variables, but attempts to understand phenomena through the meaning that people assign to them. Interpretivist methods are intended for constructing an understanding of the context of the information system, and also the process on which the IS impacts, and is impacted by, the context (Walsham 1993) cited by (Klein, Myers 1999). Good examples of interpretivist methods in qualitative research include the work of (Boland 1991) and (Walsham 1993).

Third, critical research consists of social critique, whereby restrictive alienating conditions of the status quo are brought to light (Klein, Myers 1999). In other words, critical research aims to support the removal of the causes of unwarranted alienation and dominations, thus improving the opportunities for fulfilling human potential. Examples of critical research in IS include the work of (Forester 1992), and (Ngwenyama 1992).

Finally, Table 4.1 provides a comparison of the three major IS paradigms: positivist, interpretivist, and critical. This research is performed based on the positivist paradigm. The following section provides more details on the positivist paradigm and presents the justification for choosing this paradigm.

| Assumption | Positivist | Interpretivist | Critical |
|---|---|---|---|
| Ontology | "Naïve Realism", in which an understandable reality is assumed to exist, driven by immutable natural laws. True nature of reality can only be obtained by testing theories about actual objects, processes or structures in the real world | Relativist: the social world is produced and reinforced by humans through their actions and interactions | Historical realist: social reality is historically constituted; human beings, organisations, and societies are not confined to existing in a particular state |
| Epistemology | • Verification of hypothesis through rigorous empirical testing<br>• Search for universal laws or principles<br>• Tight coupling among explanation, prediction, and control | • Understanding of the social world from the participants' perspective, through interpretation of their meanings and actions<br>• Researchers' prior assumptions, beliefs, values, and interest always intervene to shape their investigations | • Knowledge is grounded in social and historical practices<br>• Knowledge is generated and justified by a critical evaluation of social systems in the context of researchers' theoretical framework adopted to conduct research |

| | | | |
|---|---|---|---|
| Relationship between Theory and Practice | It is possible to discover universal laws that govern external world | Generative mechanisms identified for phenomena in the social sciences should be viewed as "tendencies", which are valuable in explanations of past data but not wholly predictive for future situations | • Generalisations point to regularities of process rather than cross-sectional differences<br>• Generalisation in critical research focuses on the "totality" of relationships<br>• There can be no theory-independent collection and interpretation of evidence to conclusively prove or disprove a theory |
| Role of the Researcher | Objective, impartial observer, passive, value-neutral | Interactive: the researcher interacts with the human subjects of the enquiry, changing the perceptions of both parties | Transformative: initiating change in social relations and practices, helping to eliminate the bases of alienation and domination |

**Table 4.1: A comparison of the key rhetoric of major IS research paradigms (source: (Khazanchi, Munkvold 2003, p.5))**

### 4.2.1 Research Philosophy Adopted

From the above descriptions, one can conclude that the research philosophy is the strategy that defines all fundamentals which guide the researcher's path to address and explore the research questions. In this study, the positivist paradigm appears to be the most relevant compared to the other two paradigms. Indeed, the positivist paradigm characteristics are more appropriate for this research's questions, objectives, and study proposition (i.e. the conceptual framework). The positivist paradigm is characterised by four major research evaluation criteria: good research should make controlled observation, should be able to be replicated, should be generalisable, and use formal logic. These evaluation criteria are suitable for this research (Cavaye 1996). Therefore, it is necessary to further explain these evaluation criteria and why they are appropriate for this research.

The positivist paradigm assumes an objective physical and social world that is independent of humans' involvement, and which can also be easily apprehended, characterised, and measured (Orlikowski, Baroudi 1991). For example, organisations have structure and reality, regardless of their employees' actions. Hence, a researcher can investigate the objective physical and social reality by making accurate measures, in order to identify and gauge those dimensions of reality that interest the researcher. Moreover, researchers adopting the positivist paradigm assume that human action is intentional and rational. Specifically, it assumes that humans interact in a fairly stable and orderly way, and that conflict is not endemic to organisation and society (ibid). This research focuses on the theories, concepts, and practices involved in the direct marketing process, and not the actors involved in the process. In other words, the aim is to evaluate the components and characteristics of the conceptual framework and not how marketers, data analysts, or any other users execute it.

The positivist paradigm utilises deduction reasoning, which consists of starting with theory development and verifying its validity through relevant data collection. The interpretivist paradigm uses induction reasoning, which consists of finding a case and observing relationships to construct a general theory (Cavana, Delahye & Sekaran 2001). In other words, positivist starts with theory and collects data, whereas interpretivist begins with collecting data and then develops theory.

Therefore, the positivist paradigm matches the objectives and the nature of this research. Specifically, the conceptual framework was developed through theory and data is collected for evaluation purposes. Furthermore, the positivist paradigm is most relevant, when there is a study proposition which needs to be verified. In line with that, this research has a formal proposition (i.e. the DMP-BI framework).

The positivist paradigm is the scientific explanation of phenomena, and the finding of objective cause-effect relationships between IS and human beings, business processes, and organisations. It can also be the finding of objective universal laws for the effect of IS on human beings, business processes, and organisations (Howcroft, Trauth 2005). This research evaluates the direct marketing process and the BI functions' cause-effect relationships through the developed conceptual framework. In fact, it is not possible to achieve this evaluation using interpretivist and critical paradigms, as they believe there is no single objective reality, but multiple constructed realities (Oates 2006). In other words, there is no benchmark against which to verify any findings. It is also intended to verify the DMP-BI framework against replicated observation to support or refute the framework's causal relations. This may result in 'falsification', in which the study propositions are further refined into more accurate ones (Howcroft, Trauth 2005).

This research aims to verify the DMP-BI framework through objective observation and rigorous empirical testing in the real world context. Interpretivist and critical paradigms reject the notion of objective observation, which excludes human experience in the process (Khazanchi, Munkvold 2003). Specifically, the interpretivist paradigm seeks not only to understand information systems in their social context (e.g. organisational, political and cultural contexts), but also aims to understand people's (e.g. IS users or developers) feelings, values, norms, interests, motivations, and actions (Howcroft, Trauth 2005). The critical paradigm criticises intepretivist as being too relativist and passive; for attempting to understand social reality instead of acting on it. Indeed, the critical paradigm goes a step further than interpretivist and seeks to change the world – actors, information systems, organisations, and society, including their dynamic, complex, and emergent relationships (ibid). Finally, unlike the positivist paradigm, which aims to collect value-free and unbiased facts, interpretivist and critical paradigms believe that there will always be biased facts and consider all values to be equally important.

Repeatability is an important aspect of this research endeavour. Indeed, the conceptual framework is mainly intended for marketers and data analysts to perform the direct marketing process using business intelligence tools. Unlike the positivist paradigm, which is characterised by repeatability, in interpretivist and critical research, repeatability is improbable (Oates 2006). This is because the researcher influences the study and its findings, thus another researcher is unlikely to obtain similar results.

Finally, generalisation can be achieved using the positivist paradigm, which looks for general patterns or laws, and not findings that can be related only to one case. According to (Yin 2009), if two or more cases are shown to support the same study proposition, replication may be claimed. However, interpretivist and critical research accepts the uniqueness of contexts and individuals, making the matching of results in other contexts less likely (Oates 2006). This research attempts to generalise the usage of the conceptual framework across all industries. Hence, the positivist paradigm is more appropriate to achieve generalisation.

#### 4.2.2 Justification of Qualitative Approach

There are two major research approaches: qualitative and quantitative. They are both standardised research methods that have developed in parallel as two independent spheres of empirical social research (Flick 2009). A qualitative approach is used to carry out this research. (Creswell 1998, p.15) describes the qualitative approach as "*an inquiry process of understanding based on distinct methodological traditions of inquiry that explore a social or human problem. The researcher builds a complex, holistic picture, analyses words, reports detailed views of informants, and conducts the study in a natural setting*". The qualitative approach has a variety of methods including Action Research, Case Study Research, Ethnography, Grounded Theory, Semiotics, Discourse Analysis, Hermeneutics, and Narrative and Metaphor (Myers 2009). Examples of qualitative approach data sources comprise observation, documents and texts, and the researcher's impressions and reactions. The primary usage of the qualitative approach in this context is to help us understand people, processes, data, models, and technology within specific organisational contexts.

The quantitative approach comprises different methods including surveys, laboratory experiments, simulation, mathematical modelling, structured equation modelling, statistical analysis, and econometrics (Myers 2009). After comparing qualitative and quantitative approach characteristics, it was deemed more appropriate to use a qualitative approach to carry out this research. Indeed, based on the research questions and objectives, a quantitative approach is not suitable for three main reasons. First, a quantitative approach, such as experiments, manipulates instances, whereas the qualitative approach studies instances in a real-world context (Dul, Hak 2008). Furthermore, even when the quantitative approach studies instances in a real-life context, such as survey methods, it needs a large sample of a population and the scores obtained are statistically analysed (ibid). The nature of this research makes the use of survey methods very complex, because the target audience is not expected to be large. In fact, there are not many organisations with employees who practise both direct marketing and business intelligence concepts. Therefore, qualitative methods, such as interviews and corporate records, are more feasible and suitable methods to adopt. Second, quantification of organisational processes or phenomena is not commonly an effective method of research. This is because it is usually very difficult to assign meaning to a phenomenon without considering the context, and without understanding the role of people who affect or are affected by the phenomenon (Cavaye 1996). Third, it is common in quantitative research that the researcher trades context (e.g. social, cultural, and organisational aspects), or treats it in a superficial manner, for the ability to generalise across many people or many organisations (ibid). In this research, organisational context is an important component of the overall research. Therefore, quantitative analyses are not suitable for performing this research.

There are many reasons the qualitative approach is suitable for this research. First, the main research questions are followed by "*How*" and "*What*", where the qualitative approach is most appropriate (Yin 2003b). Second, the qualitative approach is particularly relevant when the research needs to be explored (Creswell 1998). Specifically, it is relevant for research that studies a particular subject in depth (e.g. in one or a few organisations), and when the particular topic is quite new and there are not many research publications on that topic. In fact, the direct marketing process variables were complex to identify and a collection of marketing concepts and business intelligence practices needed to be combined to explain the direct marketing process. Third, there is a need to have a detailed view of the topic as a distant panoramic shot is not enough to tackle the problem. Fourth, the qualitative approach allows the researcher to observe and understand the context within which decisions and actions take place (Myers 2009). In order to evaluate the conceptual framework effectively, it is essential to observe and understand how the DMP-BI framework is applied within a real-world context. Finally, a qualitative approach is more suitable for this research because the researcher role is that of an active learner, who can tell the story from the participants' view rather than that of an expert who passes judgement on participants (ibid).

## 4.3 Proposed Research Method: Case Study

During the last four decades, case studies have been constantly used in an evaluative context. They have been applied to specific programmes, projects, initiatives, or sites and have become an integral part of evaluation research. That is one of the main reasons that case studies have been associated with process evaluations. However, case study research can be used for other reasons, such as public- and private-funded programmes, to document and analyse the outcome of the sponsored programmes (Yin 2003a). As mentioned earlier, there are several research methods that can be used under the qualitative approach; however, case study was selected as the most suitable method to investigate and answer this research question.

The most important factor when selecting a research methodology is to identify the type of question being asked (Yin 2003b). In this case, the primary research questions cover mainly the how and the why. Therefore case study is a very suitable method to evaluate the conceptual framework. Indeed, a defining characteristic of the case study method is its focus on asking "how" and "why" questions (Myers 2009). For example, the researcher in this research seeks to understand how and why a specific direct marketing activity is needed, and how and why the direct marketing process is executed in a particular way. There are two other important factors when selecting case study research: the extent of control of a researcher on actual behavioural events, and the degree of focus on contemporary rather than historical events (Yin 2009). This research is characterised by these two factors. Indeed, the researcher has little control over the results of the case studies' scenario, and also studies current issues in an organisational context. In addition, this research is characterised by the five main applications in which case study research is more relevant. First, one of the most important applications of case studies is to explain the supposed causal links in real-life events that are too complicated for other research methodologies such as survey or experiment (Yin 2003b). Second, it describes the events and real-life context in which things happen. Third, the illustration of a certain topic within an evaluation context is performed in descriptive mode. Fourth, case studies can explore events with no clear, single set of outcomes. Fifth, a case study can be a meta-evaluation, where a study is an evaluation of a study (ibid).

A case study is commonly performed to understand complex social phenomena (Yin 2003a). It is an empirical investigation of real-life events, where the boundaries between phenomenon and context are not clearly visible (Yin 2003b). The case study method can deal with technically distinctive situations, where there are many more variables to consider than data points. It usually relies on multiple sources of evidence needing triangulation techniques (Yin 2009). In a business context, case study research uses empirical evidence from one or more organisations, where the research attempts to study the subject matter in context (Myers 2009). Positivist case study research is considered as a method for testing and refining hypotheses or propositions in the real world. It is conducted to provide empirical evidence to convince other researchers of the applicability of the developed theory or proposition (ibid). The case study investigation benefits from prior formulation of hypotheses or propositions to guide the data collection and analysis (Yin 2009).

There are three elementary types of case study: exploratory, descriptive, and explanatory (Yin 2009). Case study research can be used for exploratory research, which usually involves the search for relevant features, factors, or issues of a research topic that might apply in other similar situations (Myers 2009). A descriptive study typically requires rich, detailed analysis of a specific phenomenon and its context. The analysis involves telling a story, which includes the discussion of what happened and how different people perceive what happened (Oates 2006). It can also be used for explanatory research, where there is already a lot of literature on the subject. In this case, a case study approach is used to test a theory, to develop a causal explanation, or even to compare theories (Myers 2009). In essence, case study is an all-encompassing method, covering the logic of design, data collection techniques, and data analysis approaches (Yin 2009).

According to (Myers 2009), the following criteria can be used as general guidelines to evaluate case study research in business-related disciplines:

1. The case study has to be interesting.
2. The case study has to display sufficient evidence.
3. The case study should be "complete".
4. The case study has to consider alternative perspectives.
5. The case study should be written in an engaging manner.
6. The case study should contribute to knowledge.

First, the case study has to be interesting in a sense that it should provide something new. Second, it has to display sufficient evidence such that the study arguments make sense and are plausible. Third, the case study should be complete by including all the relevant evidence to prove or disprove a particular hypothesis or proposition. Fourth, it has to consider alternative perspectives, meaning it has to consider different theories, alternative cultural views, or disagreements among the subjects. Fifth, the write-up of the case study should be done in an engaging manner, where creativity has to captivate the readers. Finally, the research case should contribute to knowledge (Myers 2009).

Case study is the most popular method in qualitative research used in business-related disciplines. This is because case studies are usually contemporary stories, in which the case documents one or more organisations' attempts to deal with issues of current importance to other organisations, which are highly likely to be facing similar challenges (Myers 2009). Moreover, case study research allows researchers to explore the theories or propositions within a real-world context. This allows the researcher to get 'close to the action', which make it easier for the researcher to identify the complexities faced by organisations (ibid). Furthermore, it allows direct observation of the events being studied. Another major strength of a case study is its ability to deal with various sources of evidence such as documents, artefacts, and observations (Yin 2009). Case study research has several disadvantages, most notably a lack of scientific generalisation, gaining access to organisations, the researcher's difficulty in focusing only on important issues, and finally being a time-consuming method (Myers 2009); and (Yin 2009).

The case study approach can be summarised into five major dimensions. First, the "focus" dimension refers to the development of an in-depth analysis of a single case or multiple cases. Second, the case study approach originated from these disciplines: a) political science, b) sociology, c) evaluation, d) urban studies, and e) other social sciences. Third, the "data collection" dimension involves collecting data from single or multiple sources including documents, archival records, interviews, observations, and physical artefacts. Fourth, the "data analysis" dimension may include descriptions, themes, or assertions. Finally, the "narrative form" dimension requires in-depth study of the case or cases (Creswell 1998).

The following sections cover the research design and the above dimensions with clear indications of the case studies' focus, data collection, data analysis, and reporting structure for case studies.

## 4.4 Research Design

Every research has an implicit, if not explicit, research design. In the most basic sense, the research design is the rational sequence that connects the empirical data to the research questions and, ultimately, to its findings and conclusions (Yin 2009). In other words, research design is the road map of the research project, with clear guidelines and procedures on the tasks required to undertake the project. In the research design, the research study needs to decide on all the various components of the research project: philosophical assumptions, research method, data collection techniques, approach to qualitative data analysis, and a written record of the findings (Myers 2009). It is important to differentiate between a work plan and research design. In fact, research design is much more than a work plan. The research design is aimed to help the researcher avoid the situation where the empirical evidence does not address the initial study questions. In particular, the research design copes with a logical problem and not a logistical problem (Yin 2009).

### 4.4.1 The Research Design Components

There are five important components of research design for case studies: 1) *A study's questions*, 2) *Its propositions*, if any, 3) *Its unit(s) of analysis*, 4) *the logic linking the data to the propositions*, 5) *the criteria for interpreting the findings* (Yin 2009). Firstly, the study's "research questions" are an important factor in selecting a research method, which has been already discussed in the previous sections. This research's main questions start with "how" and "why", which are ideal for the case study research method.

Secondly, the *study propositions* guide the focus to the important aspects, components, or constructs that should be examined within the scope of the research. In this case, the direct marketing process investigation starts with the following questions: How can the direct marketing process be executed more effectively? Why is there a need for a framework to manage and execute the direct marketing process? These questions led to the development of the study proposition, which is the DMP-BI framework. The study proposition does not only help the researcher reflect on important theoretical issues, but also starts to tell the researcher where to look for relevant evidence. For example, it can help the researcher to define and ascertain the extent of the conceptual framework benefits.

Thirdly, the *unit of analysis* component is related to the fundamental problem of describing what the "case" is – individual, event or entity. This can be identified using the study propositions. If the case being studied is an individual person, s/he is the unit of analysis. Hence, the primary aim is to collect relevant information about individual(s) or "case(s)" which might be included in a multiple-case study. The study proposition will also help the researcher to identify the relevant information to be collected about the individual. Indeed, it is common for researchers to find difficulty in narrowing the case within feasible limits. The "case" being studied can also be some event or entity. Examples of case studies that have been done on some event or entity include decisions, programmes, the implementation process, and organisational change. In this situation, the researcher has to be aware of the difficulty in defining the beginning or the end points of the "case(s)". For example, the DMP-BI framework may be executed differently depending on the perspective of the actors. In other words, if a marketer executes the direct marketing process using the DMP-BI framework, it will most certainly differ from an analyst executing it. However, this research focuses more on the entities and events that occur during the execution of the direct marketing process. Specifically, the researcher is looking to evaluate each stage of the conceptual framework in terms of its input and expected output. This involves asking the following question: does the input result in the expected output described in the DMP-BI framework? Furthermore, the researcher is testing the BI functions' applicability in each stage of the conceptual framework. In particular, whether the functions have supported the framework stages in the way defined in the DMP-BI stages. Evidently, this will also involve the examination of the causal relationship between each stage within the conceptual framework. Therefore, the units of analysis in this research are the entities and events that happen when executing the direct marketing process using the DMP-BI framework (this is discussed in detail in Section 4.6). Again, the actors' perspective is not included in the evaluation, as already discussed in Section 4.2.1. Finally, this research intends to compare the findings of the case studies with previous research. This is done using frameworks or models used previously to execute the direct marketing process. As a result, the frameworks and models' entities and events that occur when executing the direct marketing process are also considered as units of analysis. The aim of such a comparison is to clearly illustrate the DMP-BI framework's characteristics and features that make it more effective than existing direct marketing models and methodologies.

Fourthly, the *linking data to propositions* component can be performed using the following analytical techniques: pattern matching, explanation building, time-series analysis, logic models, and cross-case synthesis. Selecting an analytical technique will depend upon the case study data and the initial study propositions. Fifthly, the *criteria for interpreting a study's findings* are the final component of the research design for case studies. A significant strategy to interpret findings is to identify and address rival explanations for the findings. The fourth and fifth components are addressed in detail in Section 4.6.

### 4.4.2 Research Design Adopted

The overall research design (Figure 4.1) is composed of three major phases: 1) Define & Design, 2) Prepare, Collect, & Analyse, and 3) Evaluate & Conclude. The first phase begins with an in-depth review of the literature in the direct marketing field. The aim was to identify contemporary issues related to direct marketing, and to formulate the study scope and research questions. This was followed by the development of a conceptual framework, which attempts to address the identified issues and answer the research questions. The first phase ends with the identification of a research method to evaluate the validity of the conceptual framework. This involved selecting a research philosophy (i.e. positivist paradigm), a research approach (i.e. qualitative approach), and a research method (i.e. case study). These were covered in the last three sections.

The second phase includes selecting cases, collecting data, and conducting the case studies. The primary objective is to have a strong platform on which the conceptual framework can be effectively evaluated. Cases were selected from different industries to diversify the testing context of the conceptual framework. This is aimed at increasing the generalisability factor of the framework. The data collection and the case study design are discussed in detail in the next sections.

The final phase involves the evaluation of each individual case study, and then the performance of a cross-case evaluation. This will allow the researcher to evaluate the benefits of the conceptual framework. Finally, the research draws to a conclusion, covering the research contributions, limitations and future work.

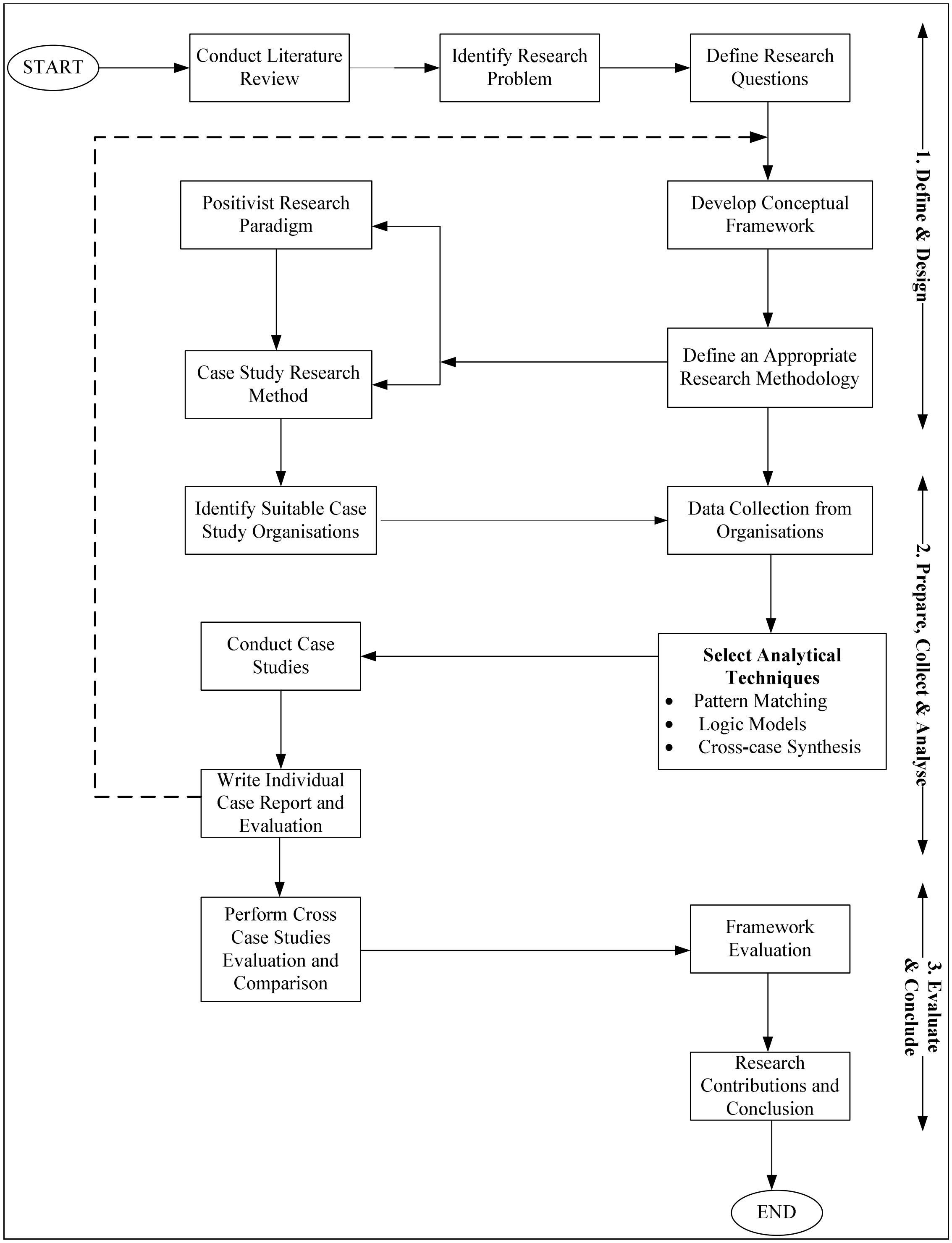



**Figure 4.1: Empirical Research Design (source: Author)**

### 4.4.3 Case Study Design Adopted

A multiple-case design approach is considered more compelling, allowing the overall study to be regarded as more robust (Yin 2009). This is because it often reflects the logic of replication. There are two types of replication logic underlying the use of case studies: a) predicts similar results (literal replication), and b) predicts contrasting findings but for anticipatable reasons (a theoretical replication). The latter type of multiple-case design is more relevant to this research. Specifically, the conceptual framework is evaluated in three case studies in order to achieve replication. It is also the intention to use rival direct marketing methodologies to perform comparison. The ultimate aim is to illustrate how the DMP-BI framework can be used to effectively execute the direct marketing process as well as to show how the rival methodologies may lack important characteristics. The findings can be the foundation for generalising the applicability of the DMP-BI framework across multiple industries.

Prior to undertaking the case studies, it is important to determine the cases and measures that will be used in collecting evidence (Yin 2009). Indeed, each case is regarded as a "whole" study in which evidence can be collected. The aim is to seek the facts and conclusions of each case in order to identify the information needing replication by other cases. A summary report is then produced, including both the individual cases and the multiple-case findings. The individual case report should indicate the reasons a specific proposition was illustrated or not. The multiple-case report should indicate the extent of the replication logic, and also the reasons that certain cases were predicted to have certain results or to have contrasting results. The dashed-line feedback loop (see Figure 4.1) represents the event where important findings happen while undertaking the case studies. This may lead to one or more reconsiderations of the study's original framework. The study should address this change(s) to avoid being accused of accommodating the original design (ibid).

An important part of Figure 4.1 is the choice of an embedded single case study. The purpose is to provide a strong platform to compare the case studies. Therefore, the research can evaluate the conceptual framework with the logical replication, i.e. theoretical replication as already discussed. Figure 4.2 illustrates the type of case study design adopted in the research design.

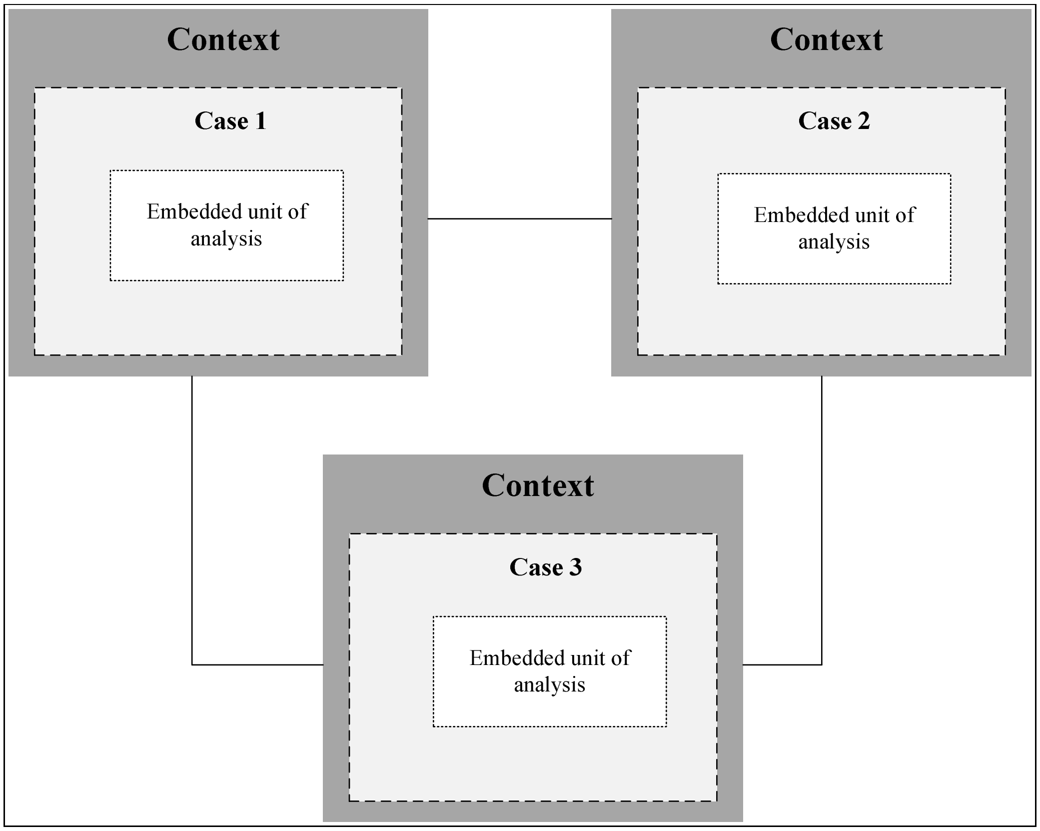


**Figure 4.2: Case Study Design (source: (Yin 2009, p.46))**

### 4.4.4 Evaluation of Case Study Quality

There are four tests that are used to establish the quality of case study research: *construct validity, internal validity, external validity, and reliability* (Yin 2009). Firstly, construct validity refers to the identification of appropriate operational measures for the concepts under study. Researchers can increase the construct validity by using multiple sources of evidence in a way that encourages convergent lines of inquiry, and this can be done during data collection. Secondly, internal validity is relevant only for explanatory case studies, where the study seeks to establish causal relationships (e.g. explain how and why event x led to event y). Internal validity is also concerned with the broader problem of making inferences. An inference from the researcher might occur when an event cannot be directly observed. This inference is commonly based on interview and documentary evidence collected as part of the case study. In order to deal with the overall problem of making inferences and therefore the particular problem of internal validity, asking the following questions can help the researcher to deal with this problem (Yin 2009):

- Is the inference correct?
- Have all the rival explanations and possibilities been considered?

- Does it appear to be airtight?

Thirdly, the external validity test involves defining the domain to which the research findings are generalisable, beyond the immediate case study. For example, if the conceptual framework is applied to other organisations, will it have the same impact? Indeed, one of the major concerns that researchers have with case studies is their external validity. However, external validity, which can also be referred to as generalisability, can be achieved through replicating the findings in a second or third case study. The two or three case studies should show that the study proposition has the same impact on different contexts, hence supporting the study proposition. Finally, the reliability test deals with the problem of demonstrating that the operations of a study can be repeated, and result in the same findings. This is achieved by documenting the procedures followed in the case study. The objective is to allow a later investigator to conduct the case study in the same way, and for them to arrive at the same findings and conclusions. Table 4.2 is a summary of the criteria described above to evaluate case study research.

| **Test** | **Case Study Tactic** | **Phase of research in which tactic occurs** |
| --- | --- | --- |
| Construct validity | • Use multiple sources of evidence<br>• Establish chain of evidence<br>• Have key informants review draft case study report | Data collection<br>Data collection<br>Composition |
| Internal validity | • Do pattern matching<br>• Do explanation building<br>• Address rival explanation<br>• Use logic models | Data analysis<br>Data analysis<br>Data analysis<br>Data analysis |
| External validity | • Use theory in single case study<br>• Use replication logic in multiple case studies | Research design<br>Research design |
| Reliability | • Use case study protocol<br>• Develop case study database | Data collection<br>Data collection |

**Table 4.2: Case Study Tactics for Four Evaluation Criteria (source: (Yin 2009, p.41))**

## 4.5 Data Collection

The selection of a data collection method depends on the research topic, research method, and availability of data. The data collection method should allow the researcher to gather all the information to answer the research questions and achieve the objectives (Myers 2009).

The data collection method was by far the biggest challenge of the entire research project. The primary aim of data collection was to get as close to the organisations under study as possible. However, the nature of the research requires access to organisations that work in a very dynamic and competitive market. As a result, it was very difficult to persuade managers to take part in the research project, given their unavailability due to many other commitments. Indeed, one of the main disadvantages of case study research is the difficulty in gaining access to the particular organisation(s) that the researcher intends to study (Myers 2009). This is because organisations are usually sceptical, in terms of the value of the research for themselves. Specifically, organisations may worry that the researcher will take too much of their time, and also that the findings might have a negative effect on the company image (ibid). Moreover, the conceptual framework integrates direct marketing with business intelligence; hence it was difficult to identify organisations that use both. In particular, the study proposes the framework as the solution to a major problem related to the direct marketing process. Therefore, it was not expected that many organisations would make use of this combination.

According to (Myers 2009), organisations' records, such as datasets, can provide some evidence that may allow the research study to build a richer picture than could be obtained by interviews or fieldwork. In fact, such records can be so valuable that they can become the object of extensive analysis (Yin 2009). Also, records are usually accurate and reliable sources of evidence. Moreover, they are stable and can be reviewed repeatedly. Furthermore, they are unobtrusive (i.e. not created as a result of the case study) and have broad coverage including a long time span, many events, and many settings. As mentioned earlier, the research attempts to gain access to the resources of the organisations under study which are relevant to the study propositions and questions. Accordingly, it was decided that the most effective way to achieve that was to obtain the dataset records which were used to perform direct marketing campaigns in these organisations. These provided sufficient access to companies' records to perform case study research. The purpose was to use these datasets to execute a direct marketing process using the DMP-BI framework. This could give a clear observation of the usefulness and practicality of the conceptual framework using real-world data, which is used in real-world projects.

Therefore, the data used to evaluate the conceptual framework was acquired from three organisations. The first dataset was acquired from a major supermarket chain in the United Kingdom (due to confidentiality, the identification of the supermarket chain is kept anonymous and referred as "Supermarket_1"). The second case study is of a major telecommunication company in the UK. The final case study dataset was acquired from Brunel Business School marketing services. In addition, the set-up of the work environment was also taken into consideration. In particular, the researcher installed two Business Intelligence tools, which provided similar work environments to those one would expect to find in any organisation.

The following subsection provides an overview of the case study protocol used to deal effectively with data collected from the organisations.

## 4.6 Data Analysis

Analysing case study evidence can be a very challenging process. This is related to the fact that there are few standard procedures or rules to guide case study analysis. (Yin 2009) proposes a strategy to guide the analysis of case study evidence. It consists of three analytical strategies: a) *relying on theoretical propositions*, b) *using both qualitative and quantitative data*, and c) *examining rival explanations*.

### 4.6.1 Relying on Theoretical Propositions Strategy

The *relying on theoretical propositions* strategy is to follow the conceptual framework that led to the case study research. Indeed, the DMP-BI framework components can help focus attention on certain data. Based on the DMP-BI framework, four interrelated themes were selected to help evaluate the case studies.

The first theme is the suitability of the stages in the DMP-BI framework. This includes the suitability of the activities, guidance, and recommendations provided by the DMP-BI framework. For example, the data modelling stage includes inputs and expected outputs, along with specific BI functions for support. These components can be the focus of attention when conducting the case studies, where evidence supporting (or not supporting) their validity is one of the key subjects of analysis.

The second theme is the applicability of the BI functions in the DMP-BI framework. This theme seeks to verify the usability of the BI functions within the DMP-BI framework. The three case studies will require the usage of various BI functions to complete the direct marketing process. Therefore, the case studies will provide empirical information to evaluate and discuss whether the recommended BI functions in the DMP-BI framework were useful.

The third theme is the structure and organisation of the stages within the DMP-BI framework. This theme is concerned with evaluating whether the DMP-BI framework provided a systematic approach using the three case study findings. For example, the case studies can help analyse the appropriate causal links between each stage in the conceptual framework.

The fourth theme focuses on the BI tools used to perform the direct marketing process. This is to assess whether the current BI tools provide the necessary functions to undertake any direct marketing process.

In other words, the conceptual framework should help organise the focus points of analysing case study evidence. In addition, it will help in identifying suitable rival methodologies (see research design section) which provide alternative explanations for more rigorous comparison evidence.

### 4.6.2 Qualitative & Quantitative Strategy

The use of both a qualitative and quantitative data strategy can yield considerable benefits for the case study analyses. Indeed, the three case studies used for evaluating the conceptual framework include both quantitative and qualitative data. However, the qualitative data remains central to each entire case study. In this research, quantitative data is part of the events in the case studies. Specifically, it is an essential part of the direct marketing process execution. For example, quantitative models are deployed in the data modelling stage to provide support for the direct marketing planning stage.

### 4.6.3 Examining Rival Explanations

Third, *examining rival explanations* is very powerful strategy, which can help yield strong evidence supporting (or not supporting) the validity of the conceptual framework. Indeed, identifying rival explanations and performing comparisons with the study propositions can add more confidence to the overall research findings. This strategy consists of defining rival explanations to the DMP-BI framework. In this context, the research identified three rival methodologies used to execute the direct marketing process. The rival methodologies were selected from the literature review, where major direct marketing methodologies were discussed (see Sections 2.4.3 and 2.4.4). Table 4.3 provides a summary of the major types of rival explanation. This study adopted current direct marketing methodologies, which can be categorised as both direct and rival theory.

After clarifying the overall strategy used for analysing the case study evidence, the following subsection presents the analytical techniques used to undertake the analysis strategies.

### 4.6.4 Analytical Technique Used

There are five major analytical techniques that can be used as part of and along with the analysis strategies: 1) pattern matching, 2) explanation building, 3) time-series analysis, 4) logic models, and 5) cross-case synthesis (Yin 2009).

The logic models technique is particularly useful in evaluating case studies. This technique is most relevant when the study proposition is a complex chain of events which occur over an extended period of time. The events occur in a cause-effect-cause-effect pattern, where the dependent variable (i.e. event) at a previous stage becomes the independent variable (i.e. causal event) for the next stage. This technique consists of matching empirically observed events to theoretically predicted events. The case studies intend to evaluate the conceptual framework by observing whether the designed stages produce the predicted outcomes.

There are four types of logic model: i) *individual-level logic model*, ii) *firm or organisational logic model*, iii) *an alternative configuration for an organisational-level logic model*, and iv) *program-level logic model*. The choice of logic model type is mainly related to the unit of analysis. In this case, the conceptual framework represents the unit of analysis. Hence, the *firm or organisational logic model* type has been selected, because it primarily deals with events occurring in an individual organisation. Similarly, each case study in this research is tracing direct marketing process events occurring in an organisation. The aim is to evaluate whether the conceptual framework improved the execution of the direct marketing process.

| Type of Rival | Description or Examples |
| --- | --- |
| **Craft Rivals:** | |
| 1. The Null Hypotheses | The observation is the result of chance circumstances only |
| 2. Threats to Validity | E.g. history, maturation, instability, testing, instrumentation, regression, selection, experimental, and selection-maturation interaction |
| 3. Investigator bias | E.g. “experimenter effect”; reactivity in field research |
| **Real-Life Rivals:** | |
| 1. Direct Rival (Practice or Policy) | An intervention (“suspect 2”) other than the target intervention (“suspect 1”) accounts for the results (“the butler did it”) |
| 2. Commingled Rival (Practice or Policy) | Other interventions and the target intervention both contributed to the results (“it wasn’t only me”) |
| 3. Implementation Rival | The implementation process, not the substantive intervention, accounts for the results (“did we do it right?”) |
| 4. Rival Theory | A theory different from the original theory explains the result better (“it’s elementary, my dear Watson”) |
| 5. Super Rival | A force larger than but including the intervention accounts for the results (“it’s bigger than both of us”) |
| 6. Societal Rival | Social trends, not any particular force or intervention, account for the results (“the times they are changing”) |

**Table 4.3: Rival Explanation Types with Brief Descriptions (source: (Yin 2009, p.135))**

This research is based on three case studies, thus *cross-case synthesis* can be a valuable technique for analysing case study evidence. The analysis of multiple case studies is likely to be easier and produce more robust findings. This technique treats each case study individually. It consists of designing a table which displays the data from individual cases based on some uniform framework. The aim is to facilitate the identification of cross-case patterns. In other words, it provides a solid foundation on which the research can build strong, plausible, and fair arguments to support the validity of the conceptual framework with empirical data.

Finally, (Yin 2009) suggested four main criteria by which high-quality analysis can be achieved:

- The analysis should consider all the evidence.
- The analysis should consider all major rival explanations.
- The analysis should consider the significant aspect of each case study.
- The researcher should use prior, expert knowledge in each case study.

## 4.7 Reporting Case Studies

There is no stereotype form for reporting case studies. However, it is important to identify a structure for the case studies to follow. This is to ensure coherence as well as clarity for the audience. In this case, the main target audience is composed of academics and practitioners, especially data analysts and marketers. The importance of identifying the target audience is to design the overall case study report so that it serves the needs and wants of the audience (Yin 2009). Indeed, academics usually tend to focus on the importance of the relationships among the case studies, their findings, and previous theory or research (ibid). Whereas, data analysts and marketers will probably focus on the way the direct marketing process is executed, and whether the findings are relevant. Therefore, the case studies' report structure, including emphasis, details, compositional forms and even length, should aim to serve both audiences. In addition, the case study reports should attempt to be as descriptive as possible in order to target the non-specialist audience as well.

Case study research's usefulness goes far beyond the normal research report, which is typically targeted at researchers rather than practitioners or a non-specialist audience. In fact, case study reports can communicate empirical information about a phenomenon (e.g. direct marketing process) to a variety of audiences such as non-specialists and practitioners (Yin 2009).

The format for writing the case study reports is based on the inputs and outputs of the conceptual framework stages. This is aimed at facilitating the process for cross-case comparisons for both the researcher and the reader. (Yin 2009) proposes six illustrative structures: linear-analytic, comparative, chronological, theory-building, "suspense", and unsequenced. Table 4.4 provides a summary of the main structures and their applications to the different purposes of case studies.

This research adopts a linear-analytic structure, which begins by presenting the problem, and reviews the relevant literature. It then demonstrates how the conceptual framework stages are executed, and reports the findings and conclusions. Such a structure is most advantageous when researchers are the main audience for the case study.

| **Type of Compositional Structure** | **Explanatory Case Study** | **Descriptive Case Study** | **Exploratory Case Study** |
|---|---|---|---|
| 1. **Linear-analytic** | X | X | X |
| 2. **Comparative** | X | X | X |
| 3. **Chronological** | X | X | X |
| 4. **Theory-building** | X | | X |
| 5. **"Suspense"** | X | | |
| 6. **Unsequenced** | | X | |

**Table 4.4: Six Illustrative Structures for Case Study Compositions (source: (Yin 2009, p.176))**

## 4.8 Ethical Considerations Related to Research

Research ethics are an important part of the research methodology process. The Oxford English dictionary describes ethics as "*the moral principles governing and influencing conduct*". In a qualitative research context, ethics are described by (Myers 2009) as a moral stance that involves "*respect and protection for the people actively consenting to be studied*".

This research work follows the Brunel Ethics Code, which is defined as follows: "*Any research that involves human participation, the collection or study of their data, organs, and/or tissue, and that is carried out on Brunel University premises and/or by Brunel University staff, or students under the supervision of Brunel University staff requires ethical approval*" (Brunel University 2010, p.3). This research includes identifiable data relating to humans. In other words, the three case studies contain different levels of detail of human data. The first and second case studies include transaction data for customers without any personal details. Therefore, these two case studies are not concerned with ethical approval. However, the third case study of this research comprises students' personal details, hence ethical approval needed to be sought.

There are some basic issues that need to be addressed by the researcher when conducting research using human personal information (Brunel University 2010). First, it is important to protect the privacy of human research participants. Second, the information on human participants needs to conform to generally accepted scientific principles, and be based on thorough knowledge of the scientific sources of information. Third, the organisation providing the data for the participants should be informed of the aims, methods, and anticipated benefits and risks of the study. Also, the organisation should be made aware that the data will be kept strictly confidential. Finally, the organisation should be informed if any publication of the results is planned.

## 4.9 Chapter Conclusion

This chapter presented the overall research design, with an overview of the most important components of the research design for case studies. An effective research design is important as it ensures the research generates empirical evidence to address the initial research questions. The data collection, analysis, and reporting structures for case studies are provided with a clear indication of the procedures taken for this research. By following this research design, a systematic approach for collecting evidence for the conceptual framework validity can be carried out effectively. The research design also ensures a strong platform on which the conceptual framework's usefulness and practicality can be validated. Specifically, three case studies were selected to ensure the validity of the conceptual framework. Moreover, three rival methodologies will be used as comparison benchmarks, upon which the performance of the conceptual framework can be scrutinised.

The chapter also explained the methodological considerations adopted in this research. It provided a description of the philosophical assumption (positivist), research approach (qualitative), and method (case study) that will underpin and validate the conceptual framework.

This chapter provided a solid foundation on which the DMP-BI framework can be evaluated effectively. Three case studies in different sectors will be conducted in order to generalise and ensure the validity of the DMP-BI Framework.

# CHAPTER 5: CASE STUDY I SUPERMARKET PROMOTIONS

## 5.1 Introduction

This chapter presents the first case study used to evaluate the DMP-BI framework. The chapter is organised into two main parts: 1) Case Study Report and 2) Case Study Evaluation. The first part presents the supermarket promotions case study, which aims to illustrate the usage of the DMP-BI framework in a retailing context. This case study adopts linear-analytic reporting, as already discussed in Section 4.7. The purpose of using linear-analytic as a standard reporting structure is to ensure having an effective platform to perform cross-case analysis. The first part of this chapter is organised into three main sections. Section 5.2.1 briefly introduces supermarket sales promotions' practices. Section 5.2.2 provides an overview of the transaction dataset used to illustrate the utility of the DMP-BI framework. It also describes the study proposition used to carry out the direct marketing process. Section 5.2.3 demonstrates the usage of the DMP-BI framework to execute a direct marketing process.

The second part evaluates the case study findings based on the three themes which were explained in Section 4.6.1. These themes include *the suitability of the stages of the DMP-BI framework, the applicability of the BI functions, and the organisation and structure*. This part is structured into three sections: 1) the DMP-BI stages, 2) BI functions, and 3) the DMP-BI structure. Section 5.3.1 evaluates the four stages of the DMP-BI framework. Section 5.3.2 assesses the BI functions' impact on the direct marketing process. Section 5.3.3 examines the links in the DMP-BI framework. The chapter concludes with a summary.

## 5.2 Case Study Report

This part of the chapter demonstrates the usage of the DMP-BI framework in a retailing context. Specifically, it investigates supermarket sales promotions. It starts by providing an overview of the supermarket's sales promotions' practices. It then introduces the transaction dataset used to demonstrate the practicality of the DMP-BI framework. It also presents the study proposition used to execute the direct marketing process. Finally, it illustrates the application of the DMP-BI framework to execute the direct marketing process for supermarket sales promotions.

### 5.2.1 Introduction

Supermarkets are always looking for new strategies to achieve higher growth and increase their market share. Sales promotions are a short-term strategic tool used to influence consumer shopping behaviour. There are four types of sale promotion widely used in UK supermarkets: coupons, price discounts, samples, and buy one get one free (Gilbert, Jackaria 2002). A study conducted by (Gilbert, Jackaria 2002) found that buy one get one free and price discount were the most popular promotions among consumers. Moreover, (Lim, Currim & Andrews 2005) also found that the impact of price discounts on sales volume is constantly found to be high. In addition, there is evidence that a high volume of sales are made on promotions. That is why these types of promotion share a relatively high portion of the marketing budget (Lim, Currim & Andrews 2005). Supermarkets are continuously increasing sale promotions to achieve higher growth as well as competitive advantage (Tolson 2007). However, according to the Professional Assignments Group report, many retailers might be over-discounting products, which leads to low profit margins and a struggle to sustain aggressive price discounting (Kitts, Hetherington 2005). This is the main reason that analysing previous promotional impact on a given product is becoming more critical, in order to avoid poor promotional planning and increase promotional impact.

### 5.2.2 An Overview of the Transaction Dataset

The aim of this case study is to illustrate the conceptual framework's practicality and usefulness in executing a direct marketing process. This is achieved using a real transactional dataset provided by a major supermarket chain in the United Kingdom. The supermarket does not want to be identified, as promotion strategies and transactions are confidential. For convenience, the supermarket is referred to "Supermarket_1".

Supermarket_1 is committed to direct marketing practice, which makes it an ideal case for this study. There are 4,300 transactions in the dataset, and 80 potential predictors. These predictors can be categorised into five main attributes: a) product type contains three different products, b) there are two promotion types, c) sales transactions are from May 2006 to January 2008, d) region attribute is grouped into ten major regions, and e) there are five customer segments. Figure 5.1 provides a graphical illustration of the dataset attributes. Appendix 5.1 provides each product code and description. For example, "B1" is a beef product code which corresponds to "Premium Roasting Beef".

The dataset contains a broad range of information which could be used to investigate several consumer behaviours. However, a study proposition is adopted to execute the direct marketing process in a specific context. The following is the study proposition:

- Each region in the UK has different consumer purchasing behaviours; therefore it requires a customised promotion campaign.

The DMP-BI framework is applied to the dataset to provide evidence for the study proposition. The aim is to prove that each region should have a specific promotion strategy. This study uses only *beef products* because the case study only needs one range of products to illustrate the DMP-BI framework in practice.

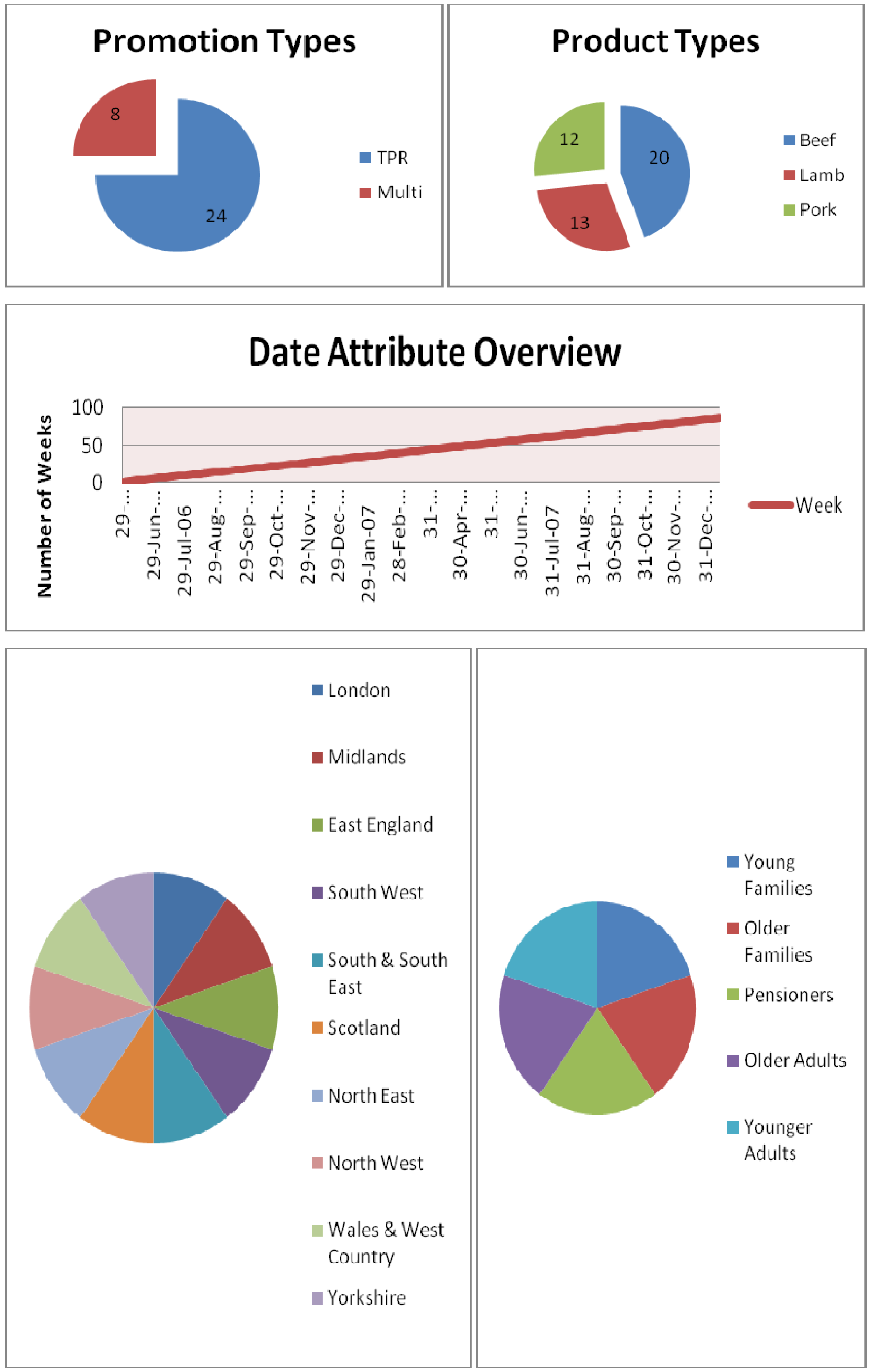



**Figure 5.1: Dataset Main Attributes**

### 5.2.3 DMP-BI Application in Retailing

This section is organised into four subsections: 1) Direct Marketing Objectives, 2) Data Preparation, 3) Data Modelling, and 4) Direct Marketing Planning. Each subsection represents one stage of the DMP-BI framework (see Figure 3.5). The aim is to provide a step-by-step illustration of the DMP-BI framework application to the dataset[15]. However, there is a need to install a business intelligence tool to execute the direct marketing process with the DMP-BI framework. Microsoft SQL Server 2008 is a Business Intelligence tool which provides all the major business intelligence functions including data integration, reporting, and analysis. SQL Server 2008 was selected because it is one of the market leaders in the business intelligence industry (Vesset 2010).

#### 5.2.3.1 Direct Marketing Objectives

This stage involves the formulation of direct marketing process objectives. The study proposition is used as a basis to formulate the objectives. There are four main objectives for this direct marketing process:

1. Investigate beef products sales in each region.
2. Examine promotions' impact on beef sales in each region.
3. Identify lifestage purchasing behaviour in each region.
4. Determine differences between regions' sales, promotions, and lifestage.

#### 5.2.3.2 Data Preparation

This stage aims to prepare the dataset for deploying direct marketing models. It is important to select the appropriate set of data because the quality of the deployed models (in the following stage) relies heavily on the data preparation stage. In this process, there are four main tasks involved in the data preparation: 1) data import, 2) attribute selection, 3) data key facts summary, and 4) data integration.

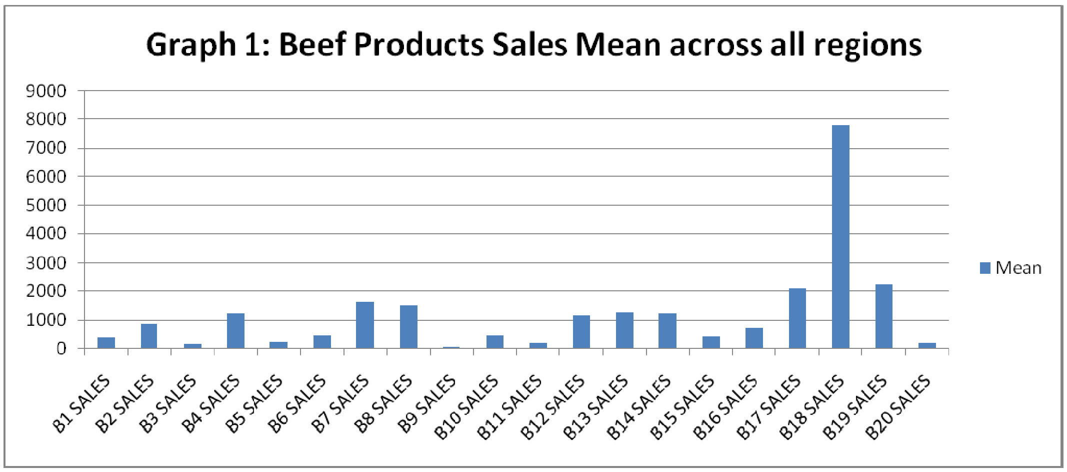
Firstly, the dataset was provided in a Microsoft Excel format, hence it was necessary to import the data into the database management platform. This was done using SQL Server's "Import and Export Data" function.

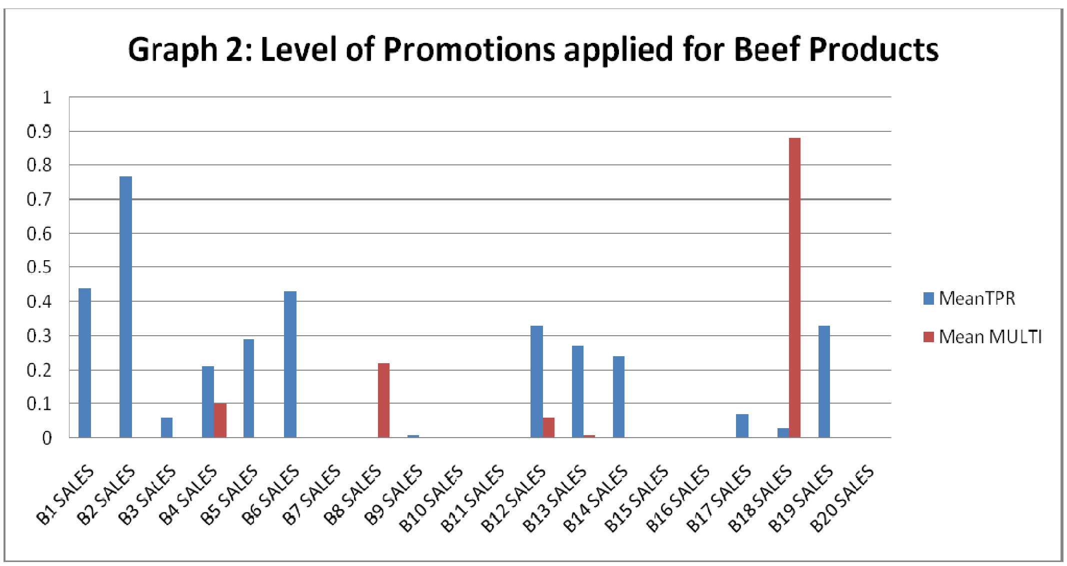
Secondly, the attribute selection is performed to select the most appropriate attributes for analysis. The attributes are selected based on the direct marketing process objectives. In fact, it is clear from the objectives that the *beef products, region, lifestage,*

[15] Extensive reporting of various configurations results to complete the case study can be provided by the author.

*and promotions* attributes are the most appropriate for analysis. SQL Management Studio "Query" functions are used to perform the attribute selection.

Thirdly, prior to model development, a summary of the dataset key facts is helpful for cross-analysis tasks with deployed models (see Section 3.4.4). These key facts are the mean of the beef sales and promotions level applied to each beef product. Graph 1 and 2 were derived from the table included in Appendix 5.2.




Finally, the dataset did not contain an attribute that has unique values. In fact, the original dataset was ordered by transaction date starting from 29/05/2006 to 14/01/2008. Therefore, it was necessary to add a transaction ID attribute in order to uniquely identify each transaction in the dataset. In other words, a primary key has been added to the dataset. This was done using the BI "Data Integration" function.

#### 5.2.3.3 Data Modelling

This stage involves five activities: 1) select data mining approach, 2) model(s) deployment, 3) model(s) accuracy, 4) model(s) selection, and 5) model(s) analyses (refer to Section 3.4.5).

Firstly, it is necessary to select which data mining approach to use. There are two data mining approaches: directed and undirected. Based on the direct marketing objectives, a directed data mining approach is more suitable for this process.

Secondly, the mass-modelling approach is selected to deploy the direct marketing models. There are nine data mining techniques available in SQL Server 2008 for this process including *association rules, clustering, decision trees, logistic regression, linear regression, Naïve Bayes, neural networks, sequence clustering, and time series*. For each technique, a direct marketing model has been deployed.

Thirdly, Table 5.1 presents the models' accuracy from the highest to the lowest. As shown in Table 5.1, there are no models deployed for *linear regression, sequence clustering and time series*. This is because these techniques support numerical data only, whereas the data used to build the models contains textual data (i.e. regions and lifestage attributes). Therefore, data models have been successfully deployed for *neural networks, logistic regression, Naïve Bayes, decision trees, association rules, and clustering*. As explained in Section 3.4.5, a data model is considered reliable only if the accuracy exceeds 50%. In this case, the association rules and clustering models' accuracy is below 50%, hence these models will not be considered.

| Data Mining Techniques | Accuracy/not relevant |
|---|---|
| ***Neural Networks*** | 73% accurate |
| ***Logistic Regression*** | 69% accurate |
| ***Decision Tree*** | 68% accurate |
| ***Naïve Bayes*** | 58% accurate |
| ***Clustering*** | 29% accurate |
| ***Association Rules*** | 12% accurate |
| ***Linear Regression*** | Relevant for numerical dataset only |
| ***Sequence Clustering*** | Relevant for numerical dataset only |
| ***Time Series*** | Relevant for numerical dataset only |

**Table 5.1: Models' Accuracy and Relevancy**

Fourthly, the *neural network* model is the most accurate model, and hence is selected for the analysis. If the *neural network* model does not fulfil the direct marketing objectives, then the following most accurate model (i.e. *logistic regression*) will be used for further analysis. Again, if the *logistic regression* model does not fulfil the objectives, the next most accurate model is subjected to further analysis (refer to Section 3.4.5).

The following subsections are the fifth activity, which is analysing the selected models.

***Neural Network Model***

The neural network model provides a comprehensive and extensive set of information on beef sales, promotions impact, and lifestage purchasing behaviour. Figure 5.2 illustrates an example of the model's results, which shows a comparison between the regions of London and the Midlands. The neural network model displays the results in a table which is made up of four columns:

- **Attribute:** this column contains the analysed attributes, i.e. *beef product, promotion type, or lifestage*. For example, "B2 Sales" is a beef product attribute as shown in Figure 5.2.

- **Value:** this column displays the value of the attribute. For instance, "B2 sales" is "1,231.519-2,563.246" units as indicated in Figure 5.2's table in row number 2.

- **Output Attribute:** this represents the last two columns, which in this example are called "Favors London" and "Favors Midlands". These columns' values can change to any of the other eight regions (see Figure 5.2). The marketers/analysts can select which two regions they want to compare in terms of sales, promotions impact, and lifestage purchasing behaviour. The blue thick line in each row, which is displayed in either column, signifies that the value of the attribute occurs more in the region with the blue thick line than in the other region with no line. For example, the “B2 sales” attribute with a value “1,231.519-2,563.246” occurs more in the London region than in the Midlands. The score of the blue thick line will vary between 0 and 100; the higher it is, the more “B2 sales” occur in the London region, for instance.

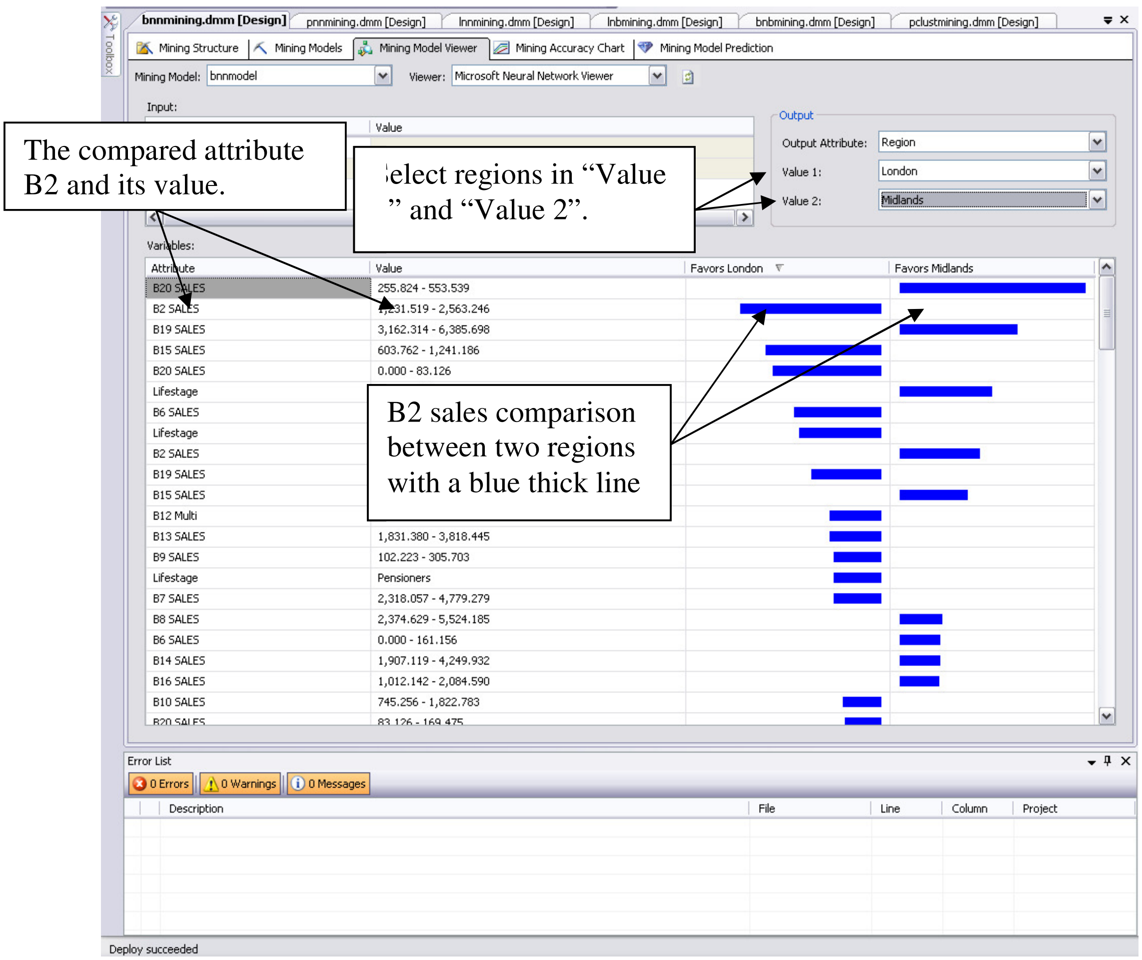



**Figure 5.2: Example of neural network model results for London and the Midlands**

As shown in Figure 5.1, the dataset contains ten regions. The following formula is used to deduce the number of possible comparisons between regions:

$C = \frac{(N(N-1))}{2}$ where C is the number of comparison possibilities and N is the number of regions, i.e. 10 (Tabachnick, Fidell 2007).

Forty-five comparison possibilities are performed between the regions. Table 5.2 is a summary of key patterns identified from the analyses performed in these comparisons. This table presents three categories of results: 1) products sales that occur more in one region than the others, 2) promotions that had more impact in one region than the others, and 3) lifestage that makes more purchases in one region than the others. Based on the information provided by the neural network model, one can conclude that beef sales are different in each region. Also, lifestage purchasing behaviour is different between regions. For example, the "East England" region favours sales of "B4" and "B11" products more than any other region, and the "Old Adults" lifestage makes the most purchases in that region. However, there is no evidence that promotions had more impact in one region than the others. Therefore, the neural network model did not help identify all the patterns to fulfil the direct marketing objectives (refer to Section 5.2.3.1). In fact, the model's results have achieved objectives 1, 3, and partially 4. Specifically, it identified beef sales in each region (objective 1), it found lifestage purchasing behaviour in each region (objective 3), and it determined the differences between regions' sales and lifestage, but not promotions. Objectives 2 and 4 still need to be achieved. In this case, the next most accurate model is subject to further analysis in order to try and achieve objectives 2 and 4 (refer to Section 3.4.5).

The next most accurate model is logistic regression. However, due to the similarity of results between the logistic regression and the neural network models, it was deemed more appropriate to use the decision tree model. In fact, a comparison was performed between the two models' results and many similarities were found. For example, the "B2 sales" and "B20 sales" comparison in the regions of London and the Midlands revealed the exact same patterns for both neural network and logistic regression. The difference was in the value range for both products, which was only a slight difference. Indeed, the "B2 sales" range for the neural network was "1,231.519-2,563.246" and "1,237.080-2,559.608" for logistic regression. Similarly, the "B20 Sales" range for the neural network was "255.824-553.539" and "254.722-548.120" for logistic regression. In addition, the decision tree model is only 1% less accurate than logistic regression.

| A Region vs Other Regions | Products Sales | Promotions | Lifestage Active Buyers |
|---|---|---|---|
| London | B2, B13, and B15 | / | Pensioners |
| Scotland | B12, B17, and B18 | B18 TPR | / |
| North East | / | / | / |
| North West | B2 and B7 | / | Older Families |
| Wales & West Country | B7 and B3 | / | Old Adults |
| South West | / | / | Older Families, Young Families, and Older Adults |
| Yorkshire | B8 and B20 | / | Older Families, Young Families, and Young Adults |
| East England | B4 and B11 | / | Old Adults |
| South & South East | B5, B6, and B7 | / | Pensioners |
| Midlands | B14, B17, B19, and B20 | B4 Multi | Pensioners (apart from London and South & South East), Young Adults |

**Table 5.2: A Region's sales, promotions and lifestage favoured more than other regions**

***Decision Tree Model***

Figure 5.3 illustrates the decision tree model's results. The decision tree technique split the data into several nodes. Each node represents a specific attribute, i.e. products, promotions, or lifestage with a value assigned to it, e.g. "B7 sales >=834" as shown in Figure 5.3. Specifically, the nodes provide the number of cases that particular value most occurred. For example, "B7 Sales" with a value that is greater than or equal to 834 units has occurred in 123 cases within the dataset and 35 cases were found in the "North West" region. In other words, the node reveals that "B7 Sales" recorded its highest sales in the "North West" region. The other layers of nodes provide more information on specific products' sales. The most important patterns are summarised below:

- After an in-depth cross-analysis of the *beef* tree model, the following patterns are found to be useful. The **B2** product records its highest sales in the *region* of *London* with *TPR* promotions having a relatively positive impact. Although the *TPR* promotions level on the **B2** product is as high as 77%, **B2** promotions have a low impact on sales in most other *regions*, especially in the *North East* and *South West*, where sales are at their lowest.
- **B3** sales are very low across all regions, with only 6% *TPR* promotions applied to it. The decision tree model reveals that *East England* records the lowest **B3** sales. It also shows that the *North West* and *Wales & West Country* record the highest sales.
- **B7** sales are fairly high with a mean close to 1,601 units. The decision tree shows that the highest-selling region for the **B7** product is the *North West*. It also reveals that the **B7** lowest-selling region is *Scotland*. It is important to mention that there is no promotion strategy applied to this product.
- The **B13** product has 33% *TPR* and 6% *Multi* promotions. The product sales are relatively high across most regions. The decision tree reveals that Yorkshire is the region with relatively low sales.
- **B14** sales are relatively high with 24% *TPR* promotions. The decision tree model shows that sales are the highest in the *Midlands, South & South East*, *East England*, and *Wales & West Country*, while the lowest by far are in *Scotland*.
- The **B15** product has very low sales in most regions. The decision model shows that *Yorkshire* has the highest sales for **B15** and *South West* the lowest. There are no promotions applied to this product.
- **B16** records its highest sales in the *Midlands* with sales exceeding 621 units. *Scotland* has the lowest sales for this product. There are no promotions used for this product.
- **B17** has 7% *TPR* promotions applied to it. The tree model shows that *Wales & West Country* has very low sales figures.
- The **B18** product has the highest sales as well as the highest number of promotions of all beef products. The *South West* has the highest sales units of **B18**. The tree model does not show many cases that have low sales of **B18**.

Indeed, **B18** *Multi* promotions have a significant impact on **B18** sales across all regions.

- **B19** has high sales in *Wales & West Country* and the *South & South East*, and low sales in the *North West*. This product has 33% *TPR* promotions.

Although the above analyses seek to relate promotions to sales performance, the sales figures may be high or low due to factors other than promotions. For example, the **B2** product's high sales in London could be related to the large population of the city rather than the 77% *TPR* promotion applied to it. From the above analyses, it is clear that the decision tree did not find new patterns which had not already been found in the neural network model's results. In fact, objectives 2 and 4 still need to be completed. This is the reason the Naïve Bayes model will be analysed in attempt to achieve those remaining objectives.

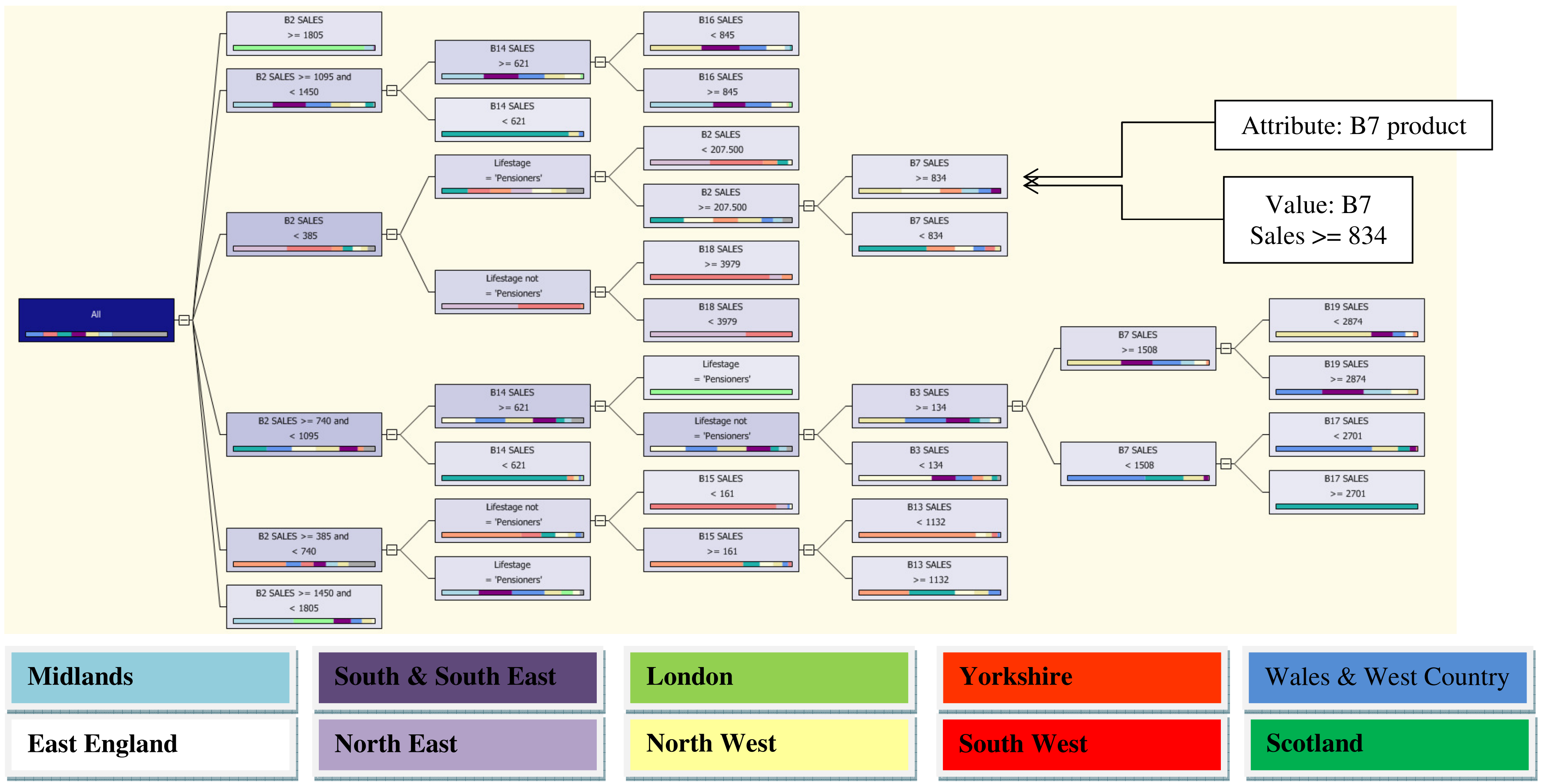



**Figure 5.3: Beef Products Sales in Different Regions using Decision Tree Technique**

***Naïve Bayes Model***

Figure 5.4 shows the Naïve Bayes model, which is represented in a table containing the following four columns:

- **Attributes**: beef products and promotions,
- **States**: sales figures,
- **Population**: the data sample analysed,
- **Regions**: each region along with the sample size analysed.

The Naïve Bayes model has divided sales figures into the following four categories:

a) High-selling products with sales exceeding 3,000 units are highlighted in violet.

b) Average-selling products with sales between 500 and 3,000 units are illustrated in light green.

c) Low-selling products with average sales between 200 and 500 units are shown in light red.

d) Very low-selling products with sales below 200 units are presented in light blue.

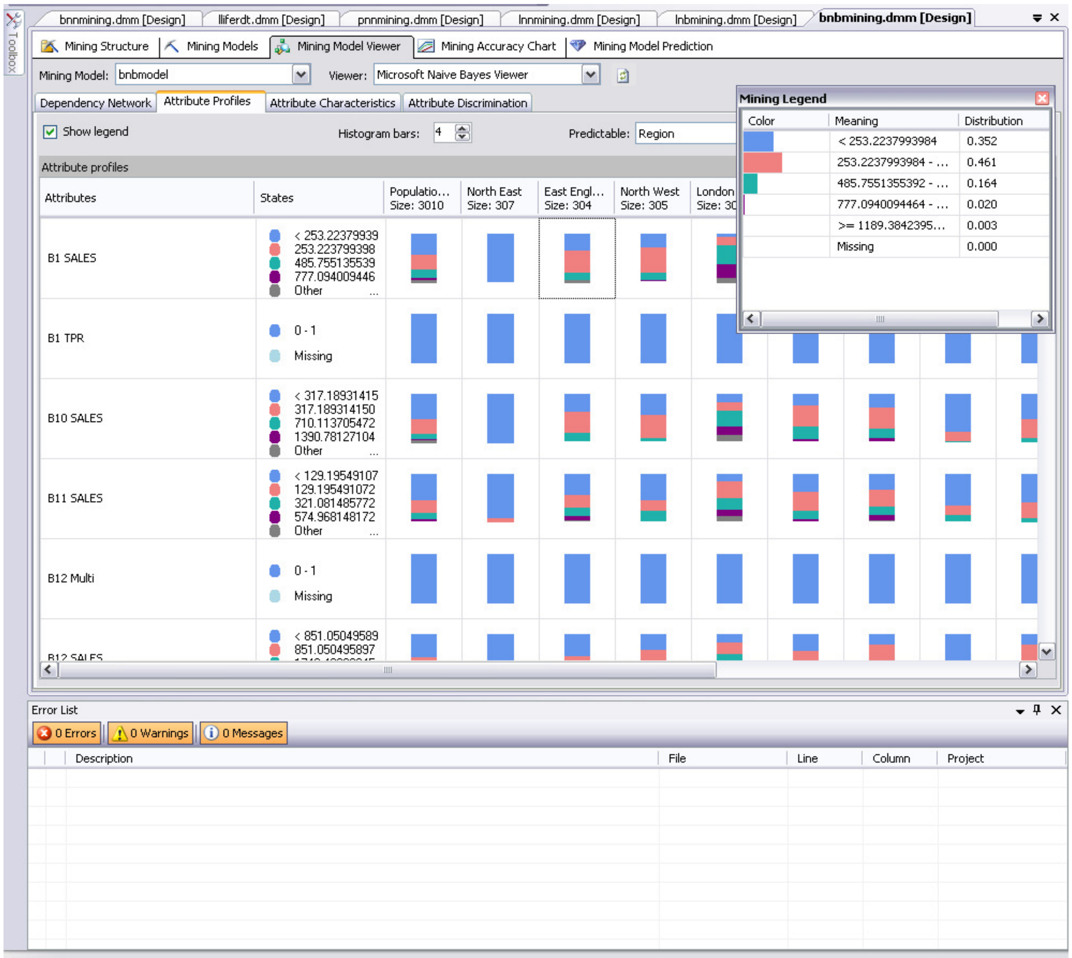



**Figure 5.4: Naïve Bayes Model for Beef products Sales in each Region**

Table 5.3 is a summary of the information revealed by the Naïve Bayes model. London and the Midlands are found to be the highest-selling regions closely followed by East England. Regions with average sales comprise the North West, Scotland, South & South East, and Wales & West Country. The lowest sales regions were by far the North East, South West and Yorkshire. The Naïve Bayes model also identified the highest-selling products, *i.e.* B18, B17, B19, B7, and B2; and the lowest-selling products, *i.e.* B5, B3, B9, B11, and B20. These patterns provide evidence that most regions have different product preferences in which sales are higher for one product in one region than the others. However, the model did not reveal any interesting information for promotions' impact. Therefore, objectives 2 and 4 are still not fulfilled. There are no more models available for analysis, as clustering is the following most accurate model with an accuracy of 29%, making it unreliable for analysis (refer to Section 3.4.5).

| Regions | High-Selling Products (Sales > 3,000 units) | Good-Selling Products (Sales between 1,000 & 3,000 units) | Average-Selling Products (Sales between 500 & 1,000 units) | Low-Selling Products (Sales between 200 & 500 units) | Very Low-Selling Products (Sales < 200 units) |
|---|---|---|---|---|---|
| **East England** | B18 | B17, B14, B18, B19, B2, B4, B6, B8, B7 | B1, B12, B13, B16 | B10, B15 | B5, B3, B9, B11, B20 |
| **North West** | B18 | B2, B7, B17 | B1, B12, B13, B14, B19, B4, B6, B8, B16 | B5, B10, B15 | B3, B9, B11, B20 |
| **London** | B18, B17 | B1, B12, B8, B13, B14, B19, B2, B4, B6, B7, B16 | B5, B10, B15 | B20 | B3, B9, B11, |
| **South & South East** | B18 | B2, B4, B14, B6, B19, B8, B13, B7, B17 | B1, B5, B12, B16 | B10, B15 | B3, B9, B11, B20 |
| **Midlands** | B 19, B18 | B2, B4, B8, B13, B12, B14, B7, B16, B17 | B1, B6, B15 | B5, B10, B20 | B3, B9 |
| **Yorkshire** | B18 | B19, B7, B17 | B2, B4, B14, B13, B8, B12 | B1, B6, B10, B15, B16 | B5, B3, B9, B11, B20 |
| **Scotland** | B18 | B19, B8, B17 | B2, B4, B13, B12, B7 | B1, B14, B6, B10, B15, B16 | B3, B5, B9, B11, B20 |
| **Wales & West Country** | B18 | B19, B8, B7, B17 | B4, B14, B12, B2, B13, B16 | B1, B6, B10, B15 | B5, B3, B9, B11, B20 |
| **North East** | / | B18 | B12, B13, B14, B17, B19, B4, B7, B8 | B1, B10, B15, B16, B2, B6 | B11, B20, B3, B5, B9 |
| **South West** | / | B18 | B12, B13, B14, B17, B19, B4, B7, B8 | B1, B10, B15, B16, B2, B6 | B11, B20, B3, B5, B9 |

**Table 5.3: Summary of information revealed by Naïve Bayes model**

### 5.2.3.4 Direct Marketing Planning

The neural network, decision tree, and Naïve Bayes models have revealed very useful patterns in different ways. These patterns have been used to fulfil the direct marketing objectives. Firstly, the neural network model provided information on beef products sales and lifestage purchasing behaviour. The results showed that there were significant differences in beef sales and lifestage purchasing behaviour in each region. Secondly, the decision tree and Naïve Bayes models confirmed the patterns discovered using the neural network model. Both models provided evidence that different regions in the UK have different beef products' purchasing behaviour. Therefore, these models have helped achieve objectives 1, 3, and partially 4, as already discussed earlier. However, none of the models provided tangible information on promotions making an impact on sales or purchasing behaviour. As a result, objective 2 and partially 4 have not been achieved using the above models. This could be interpreted in two different ways: 1) the models failed to identify pattern(s) on promotions' impact in each region, or 2) there was no promotions' impact, hence the current promotions strategy is not effective.

The main patterns that were discovered by these models can be summarised as follows: firstly, the *North East, South West and Yorkshire* regions are the most poorly performing regions. Secondly, **B5, B3, B9, B11, and B20** were the lowest-selling beef products across all regions. It is interesting to mention that these products had very low or no *TPR* promotions applied to them. Thirdly, the highest sales regions were London and the Midlands. This might be due to the fact that these regions have the highest populations compared to other regions. Finally, **B18** was the highest-selling product with an 88% *Multi-buy* promotion applied to it.

Based on the models' results, a series of actions can be suggested and they are as follows:

1. It has been demonstrated that regions have different purchasing behaviours, and therefore a region-specific promotions strategy should be implemented.
2. A new direct marketing process should be performed, with the objective of investigating the reasons for low sales in the *North East, South West and Yorkshire* regions.
3. Similarly, a new direct marketing process should be executed to examine ways to increase sales for **B5, B3, B9, B11, and B20** products.

4. Investigate ways to introduce more Multi-buy promotions to a greater variety of regions and products, due to its positive impact in making **B18** the highest-selling product.

## 5.3 Supermarket Case Study Evaluation

The second part of this chapter evaluates the impact of the DMP-BI framework in the supermarket promotions case study. It is organised into three sections: 1) the DMP-BI stages, 2) BI functions, and 3) the DMP-BI structure. The first section examines the four stages involved in the DMP-BI framework. The second one evaluates the impact of the BI functions on the direct marketing process. The third section assesses the links in the DMP-BI framework.

### 5.3.1 The DMP-BI Stages

This section investigates the impact of the activities provided in each DMP-BI stage. The aim of these activities is to facilitate the process of executing the direct marketing process more effectively. Therefore, it is critical to verify whether these activities have provided the predicted benefits in executing the direct marketing process. The following evaluates the impact of each stage on the direct marketing process execution.

Firstly, the "Direct Marketing Objectives" stage aims to provide a basis to execute the direct marketing process. It recommends three main activities for marketers/analysts to follow including *identify process initiator(s), consider key questions, and identify a suitable data source*. In this case, the latter activity was not required because the dataset was already provided by Supermarket_1. The other two activities have been used. The first one requires marketers/analysts to identify the process initiator(s) (*Is the process initiated internally or externally*?) In this case, it was initiated by the study proposition, which is categorised as an external entity. This is because the study proposition is not related to a previous direct marketing process (see Section 3.4.3). The second activity recommends a set of key questions that can assist marketers/analysts to formulate the direct marketing objectives. In this case, the process used two key questions: "*How price-sensitive are customers?*" and "*What customers would be most receptive to our offerings?*" These questions are derived from Table 3.2 and are selected because the study proposition falls into the "Promotions" and "Customer Selection" categories (refer to Table 3.2). The outcome of this stage was, as expected, a list of objectives and a dataset to execute the process.

Secondly, the "Data Preparation" stage provides marketers/analysts with three key activities: 1) attribute selection, and 2) data sampling, and 3) data key facts summary. The "Attribute Selection" activity aims to select the appropriate attributes to use for the data modelling stage. As recommended by the DMP-BI framework stage, the selection process was based on the "Direct Marketing Objectives" stage. This ensures data integrity in terms of redundancy and relevance towards the process objectives. The "Data Sampling" activity was not necessary in this stage as all the data is needed for analysis. The "Data Key Facts Summary" activity was used, and the mean, minimum, maximum and standard deviations of the selected attributes were calculated as recommended. It is important to mention that this activity was identified from conducting the case study as useful for performing cross-analysis between the deployed models and the selected attributes. This activity was added to the DMP-BI framework as explained in Section 4.4.2, where the research design adopted allows reconsideration of the original framework design. In addition, the "Data Preparation" stage involved two more activities, namely *data import* and *data integration*. These activities are part of the BI supporting functions and will be discussed later in Section 5.3.2.

Thirdly, the "Data Modelling" stage involves five activities: 1) select a data mining approach, 2) model(s) deployment, 3) model(s) accuracy, 4) model(s) selection, and 5) model(s) analysis. In this case, all of the activities have been used to complete the stage. The stage began with selecting an appropriate data mining approach. Given the process objectives, directed data mining was selected because the process was searching for specific results (see Section 3.4.5). Next, mass modelling was performed to complete the models' deployment activity. It is important to clarify that marketers/analysts have the other option of using Table 2.7 to select the data mining methods. However, in this case, it was more suitable to use the mass-modelling approach. Selecting data mining techniques is highly subjective and requires marketers/analysts to choose the best approach based on the specific process objectives. The "Model(s) Accuracy" activity was then calculated using the lift chart method. This allowed the identification of the most accurate model to use for analysis. The neural network model was the most accurate, and hence selected for analysis. However, the neural network model's analysis did not reveal sufficient patterns to achieve the process objectives. In this case, the DMP-BI framework recommends marketers/analysts use the next most accurate model and so on. Therefore, the decision tree and Naïve Bayes models were used for more in-depth analysis in order to find patterns to fulfil the process objectives.

Finally, the "Direct Marketing Planning" stage recommends marketers/analysts to perform the following:

- Activity 1: Check whether the process objectives are achieved.
- Activity 2: Revise key patterns.
- Outcome: Provide a direct marketing campaign or a list of suggestions.

In this case, all the objectives were achieved apart from identifying the promotions' impact in each region. The "revise key patterns" activity was performed and a series of suggestions were provided for further analysis. Planning a direct marketing campaign was not feasible because of the lack of access to the organisation's information, such as budget allowance.

### 5.3.2 BI Functions

This section examines the impact of BI functions in each stage of the DMP-BI framework. The aim of BI functions is to provide an integrated information system platform on which the process can be executed. Therefore, it is important to investigate whether Microsoft's SQL Server 2008 business intelligence suite has provided all the required functions to execute the direct marketing process.

In the first stage, BI enterprise reporting is recommended for use in order to facilitate documentation sharing in organisations. However, this study was performed outside the organisation environment. Therefore, there was no requirement to share the direct marketing process documentation with other departments concerned.

The second stage used three database management functions, namely *data import, data integration,* and *attribute selection*. These functions have all been recommended by the DMP-BI framework and were available in the SQL Server 2008 BI suite.

The third stage involved deploying the data models. The SQL Server 2008 BI tool provides the necessary data mining techniques to deploy the models. In this case, mass modelling was performed and the nine techniques available in SQL Server 2008 were used to deploy the models. After that, the lift chart accuracy technique was used to evaluate the performance of each model. This BI function allows the identification of the most accurate model.

The final stage included two BI functions: BI search and report publishing. These two functions were not used in this process because there was no requirement to share results across the organisation, and no requirement to search for a new direct marketing process to execute.

### 5.3.3 The DMP-BI Structure

This section evaluates the structure and organisation of the DMP-BI framework. Specifically, it investigates the links between each stage to verify their suitability. The link between each stage should be a causal link. This means each stage that is executed will cause the start of the next stage. However, the link between the stages is not limited to a simple causal link. It is also an iterative process, where marketers/analysts can move back to the previous stage and make changes (see Figure 5.5).

In this case study, the stages have occurred in a systematic way, where each stage has caused the next stage to start. Also, each stage provided the relevant information for the next stage to be executed, except for the "Data Preparation" stage. In fact, the user had to navigate back from the "Data Modelling" stage to the "Data Preparation" stage. This is because of the incomplete models which were deployed, using the dataset without a field that uniquely identified each field. In addition, SQL Server 2008 did not allow the use of a combination key as an alternative way to uniquely identify each transaction. This has led the user to navigate to the "Data Preparation" stage and perform the data integration activity. Overall, the sales promotion case study has illustrated that the DMP-BI structure is suitable for executing a direct marketing process.

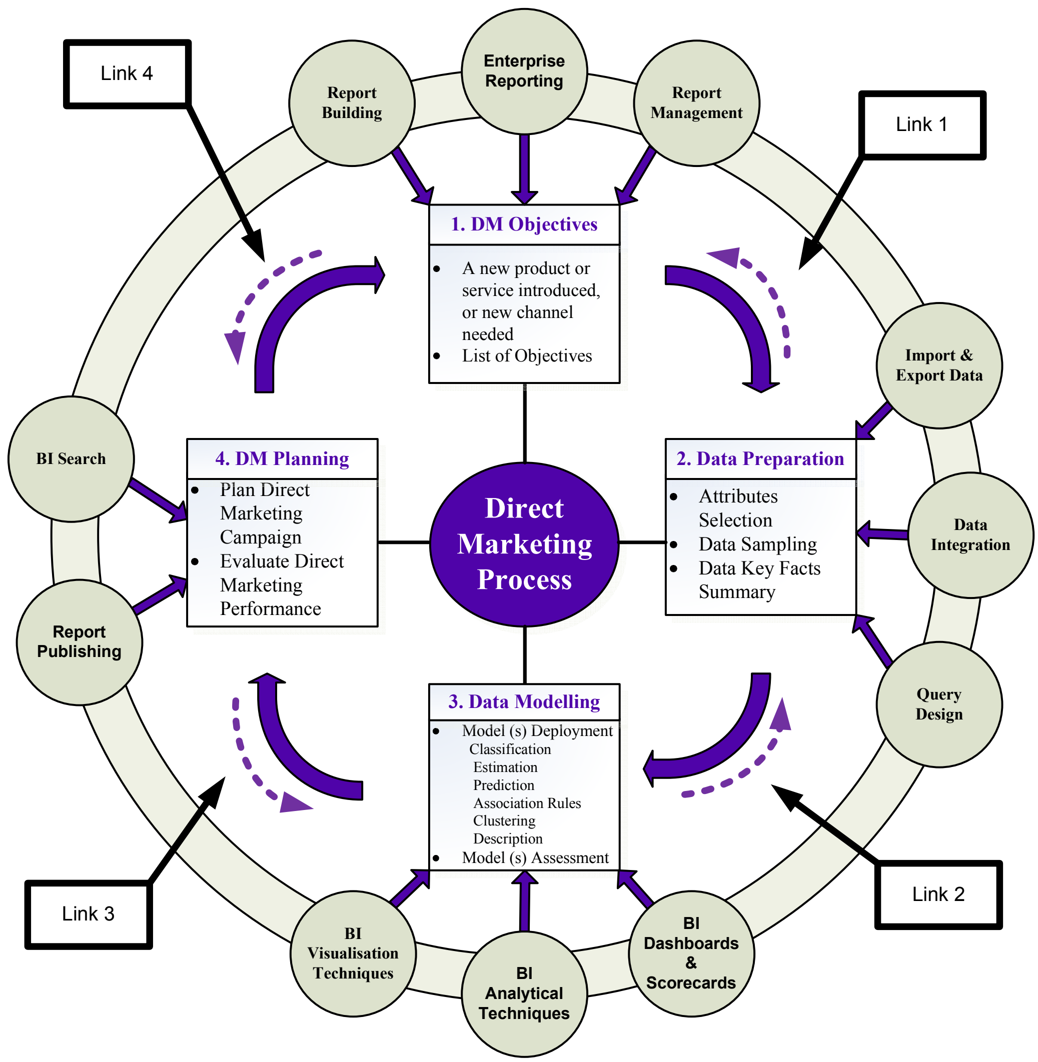



**Figure 5.5: The DMP-BI Framework Structural Links (source: Author)**

## 5.4 Chapter Conclusion

This chapter demonstrated and evaluated the usage of the DMP-BI framework in a retailing context. It presented the supermarket promotions case study, which aimed to show the application of the DMP-BI framework within a real-world organisational context. This was achieved through a step-by-step illustration of the DMP-BI framework using the supermarket's transactional dataset.

The chapter concluded by providing an in-depth evaluation of the impact of the DMP-BI framework in executing a direct marketing process. This included assessing the practicality of the four stages, the BI functions, and the linkage of the DMP-BI framework.

This chapter illustrated the application of the developed conceptual framework in retailing context. It has also demonstrated the practicality and usefulness of the developed framework in executing direct marketing process in retailing context.

# CHAPTER 6: CASE STUDY II TELECOMMUNICATION SECTOR

## 6.1 Introduction

This chapter presents the second case study used to evaluate the DMP-BI framework. The chapter is structured into two main parts including the case study report and case study evaluation. The first part of the chapter presents the case study report, which aims to demonstrate the usage of the DMP-BI framework in the telecommunication sector. Similarly to the first case study, this case study adopts linear-analytic reporting (refer to Section 4.7 for more details). This part of the chapter is organised into three sections. Section 6.2.1 introduces a major issue within the telecommunication sector; that is, the high level of churning. Section 6.2.2 presents the dataset and study proposition used to demonstrate the usage of the DMP-BI framework. Section 6.2.3 demonstrates the usage of the DMP-BI framework in the telecommunication sector.

The second part of the chapter provides an evaluation of the case study based on the three themes that were discussed in Section 4.6.1. These themes include *the suitability of the stages of the DMP-BI framework, the applicability of the BI functions, and the organisation and structure*. This part is organised into three sections: 1) the DMP-BI stages, 2) BI functions, and 3) the DMP-BI structure. Section 6.3.1 assesses the four stages of the DMP-BI framework. Section 6.3.2 evaluates the BI functions' impact on the direct marketing process. Section 6.3.3 examines the links in the DMP-BI framework. Finally, a summary of the chapter is provided.

## 6.2 Case Study Report

The aim of the case study report is to demonstrate the usage of the DMP-BI framework in the telecommunication sector. This part is organised in three sections. It starts by introducing the high churn issue within the telecommunication industry. Next, it provides details of the dataset and study proposition used to execute the direct marketing process with the DMP-BI framework. It then demonstrates the application of the DMP-BI framework to execute the direct marketing process.

### 6.2.1 Introduction

The telecom industry is considered as one of the most competitive markets (Feireira, 2004). It is extremely dynamic, with new technologies, services and competitors constantly changing the market environment. In fact, wireless telecom companies are always offering new rates and incentives, with the aim of attracting new customers and also luring customers away from rival companies (Mozer et al. 2000). In such a dynamic market, companies are always looking for new methods to identify customers who are most susceptible to churn. This will help companies to act upon those customers before they switch away.

Churn has become one of the most important business issues faced by telecom companies (Ferreira et al. 2004). The great financial benefits of preventing customer churn versus the costly process of acquiring new customers can only be seen as a competitive advantage. Therefore, there is a significant endeavour in research to find methods to enhance customer churn prediction. Indeed, companies need an accurate and timely prediction of future churners to act on. Also, they need accurate classifications to identify the reasons that cause customers to churn (Ferreira et al. 2004).

### 6.2.2 An Overview of the Dataset

The dataset used in this case study was acquired from “TERADATA Centre for Customer Relationship Management at Duke University”. It is a scaled-down version of a real database provided by “Cell2Cell”, an anonymous mobile phone company. The dataset has 71,047 customers and 78 variables (The Fuqua School of Business 2002). Appendix 6.1 provides a detailed overview of the variables names and descriptions. It also presents key statistical facts for each variable including the minimum, maximum, mean, and standard deviation (ibid).

The dataset provides a wide range of information which could be used to investigate many aspects of customers’ behaviour. However, the “Churn Game” study material[16] is used to provide the direct marketing process with a specific context. In other words, a study proposition is presented based on the “Churn Game” study material. The following is the study proposition:

- Customer churn is motivated by specific variables within the dataset. The identification of those variables can help plan a direct marketing campaign for enticing customers to remain with Cell2Cell.

[16] The Churn Game case study materials can be accessed through this web link: http://www.fuqua.duke.edu/centers/ccrm/datasets/cell/ (Last accessed January/2011).

The DMP-BI framework is applied to the dataset to predict customers churning, and identify the variables which could be a significant cause for customers churning. The purpose is to identify a list of high potential churners, and find variables which can be used to entice these churners to remain with Cell2Cell. Moreover, the aim is to illustrate the process of planning a direct marketing campaign using the DMP-BI framework.

#### 6.2.3 DMP-BI Framework Application in Telecommunication

This section presents the application of the DMP-BI framework to the Cell2Cell dataset. It is organised into four subsections, which also correspond to the number of stages involved in the DMP-BI framework. Indeed, each subsection represents one stage of the DMP-BI framework and they are as follows; 1) Direct Marketing Objectives, 2) Data Preparation, 3) Data Modelling, and 4) Direct Marketing Planning. The purpose of this structure is to provide a step-by-step demonstration of the DMP-BI framework executing the direct marketing process[17] (refer to Figure 3.5). Unlike the previous case study, where the SQL Server 2008 Business Intelligence tool was used to execute the DMP-BI framework, this case study uses SPSS Clementine Business Intelligence Version 12. The purpose is to show that the DMP-BI framework is not restricted to one tool.

##### 6.2.3.1 Direct Marketing Objectives

The study proposition is used as a basis to formulate the direct marketing objectives. The following is the set of objectives used to execute this direct marketing process:

1. Develop a direct marketing model(s) for predicting customer churn.
2. Identify the most significant drivers of churn using the deployed model(s).
3. Plan a direct marketing campaign to try and retain customers.

##### 6.2.3.2 Data Preparation

In this stage, the dataset is subjected to a number of manipulations. The aim is to identify the best possible set of data to deploy the model(s). The purpose of this/these model(s) is to predict customer churn and identify the features that lead them to churn. Therefore, it is important to perform rigorous data preparation to ensure maximum accuracy.

This stage involves two main activities: data type selection, attribute selection, and data sampling. First, the BI tool's data manipulation capabilities are used to select the

[17] Extensive reporting of various configurations and results to complete the case study can be provided by the author.

appropriate data type for each variable within the dataset. The following activities are performed:

1. Add the “SPSS File” node to a stream (a stream is the explorer pane in SPSS Clementine) and import the data file.
2. In the Filter tab, delete “Customer”, “Churn”, and “Calibrat” from further processing.
3. Add a “Type” node to display all the fields of the dataset. This node is used to define the correct data type for each field. For example, the “Occupation” fields are set as “Range Type”, but they are “Flag Type” because they are based on two values, either 1 or 0.
4. The type node is also used to define the field that is targeted for prediction. In this case, “CHURNDEPT” is the target value for prediction and the rest of the fields are considered as input values.

Second, attribute selection is performed by adding the “Feature Selection” node. This node is used to remove attributes that do not provide any useful information related to the target prediction field, i.e. “CHURNDEP”. The following are the activities performed to complete attribute selection:

1. Add and execute the feature selection node.
2. Open the feature selection results, which will show the fields within the dataset that will provide the most useful information for predicting existing customer churn. The other fields are simply categorised as marginal or unimportant.
3. Click on “Generate” and then “Filter…” to select only the features that will provide useful information. Marginal and unimportant fields are not included in the generated filter.
4. The filter with the important fields is generated. It is also attached to the type node.

5. Add a data audit node to search for fields with a high number of missing data. Any field that is less than 50% complete needs some amendments to ensure the integrity of the deployed churn model. Open and execute the data audit node.
6. The data audit node provides a table with the name, sample graph, type, minimum value, maximum value, mean and standard deviation of each field. It also includes the quality tab where the user can check the percentage of missing data in each field and, if required, make any relevant amendments. In this case,

there is no significant amount of data missing; hence no further changes are needed.

After performing both data type selection and attribute selection activities, the data preparation stage is completed. The following is the data modelling stage, where customer churn predictive models are deployed.

Data sampling was performed during the next stage (i.e. Data Modelling). This is because a neural network model needs a training dataset and a testing dataset. However, this activity is part of the "Data Preparation" stage, thus requiring the user to return to this stage and complete this activity.

#### 6.2.3.3 Data Modelling

This stage involves five main activities: 1) select data mining approach, 2) deploy the model(s) using data mining techniques, 3) identify model(s) accuracy, 4) select the most accurate model(s) for analysis, and 5) perform the analyses.

Based on the process objectives, a directed data mining approach is most appropriate for this stage. This is because the process is looking for specific patterns (refer to Section 3.4.5). After selecting the directed data mining approach, the second activity is to deploy data models using data mining techniques. The BI tool used to carry out this process encompasses a wide variety of data mining techniques. It is commonly a good practice to perform mass modelling. In this case, the main purpose of the process is to perform numerical prediction of customer churn. There are three data mining techniques that are commonly used for numerical prediction including regression analysis, neural networks, and Naïve Bayes (refer to Table 2.7). Therefore, these techniques are used to deploy data models for predicting customer churn.

The next activity is to measure the model's accuracy. Table 6.1 presents the model's accuracy for each data mining technique. As shown in Table 6.1, the neural network model is by far the most accurate. As a result, it is selected to be used for analysis.

| **Data Mining Techniques** | **Accuracy** |
|---|---|
| **Neural Networks** | 71.475% |
| **Logistic Regression** | 58.2% |
| **Naïve Bayes** | 32.335% |

**Table 6.1: Predicting customer churn models accuracy**

The final activity is to analyse the neural network model to achieve the process objectives. The following section presents the neural network model with an in-depth analysis of its results.

***Neural Network Model***

Prior to the deployment of the neural network model, it was necessary to perform some further changes to the dataset. This is because a neural network needs an almost equal number of 1s and 0s for the target variable (i.e. CHURNDEP) to train properly. Figure 6.1 is an illustration of the overall process used to build the neural network model and evaluate its predictive power.

Figure 6.2 presents the neural network model, but only shows the first 18 variables (predictors) in descending order and according to their importance. The top four variables correspond to the "Number of days of the current equipment", "% Change in minutes of use", "Mean total recurring charge", and "Months in service". These are the variables that have the most impact in causing customers to churn. Figure 6.3 is a summary of the model including its accuracy and all the variables used to build the model.

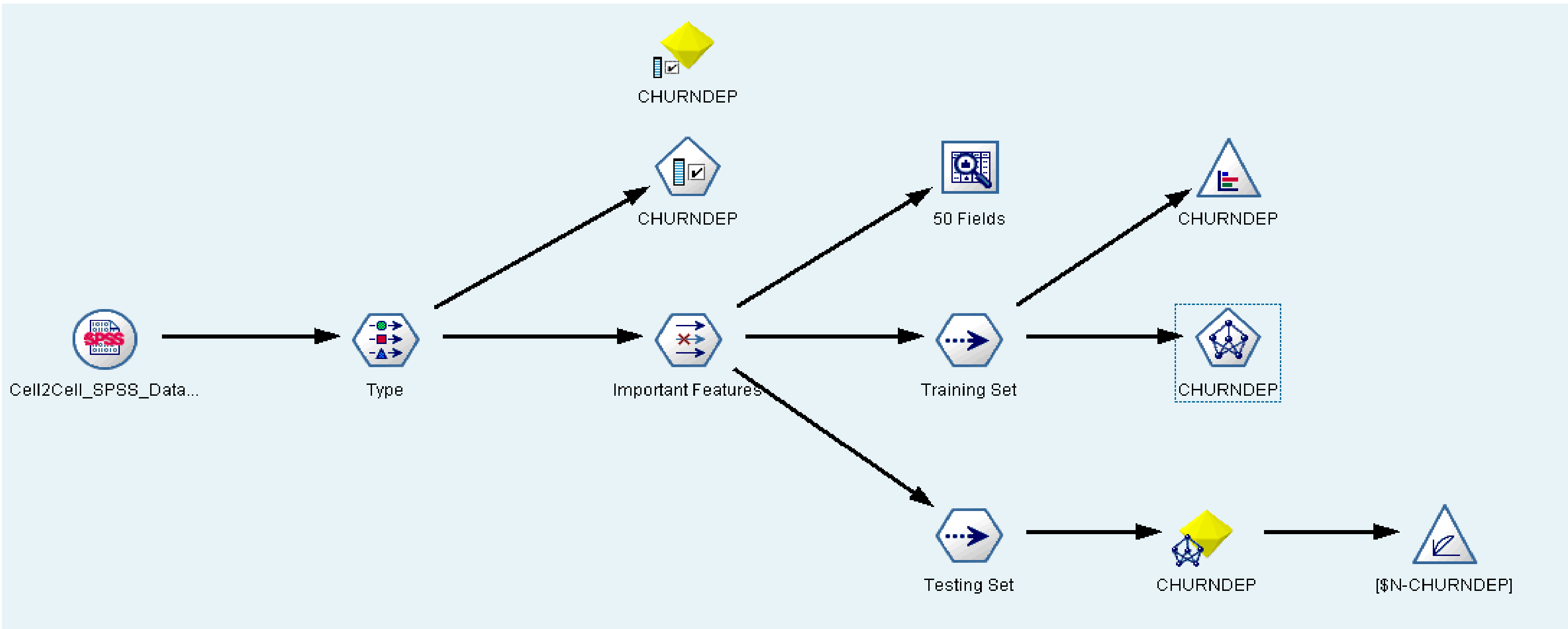



**Figure 6.1: Overall process used to build the neural network model**

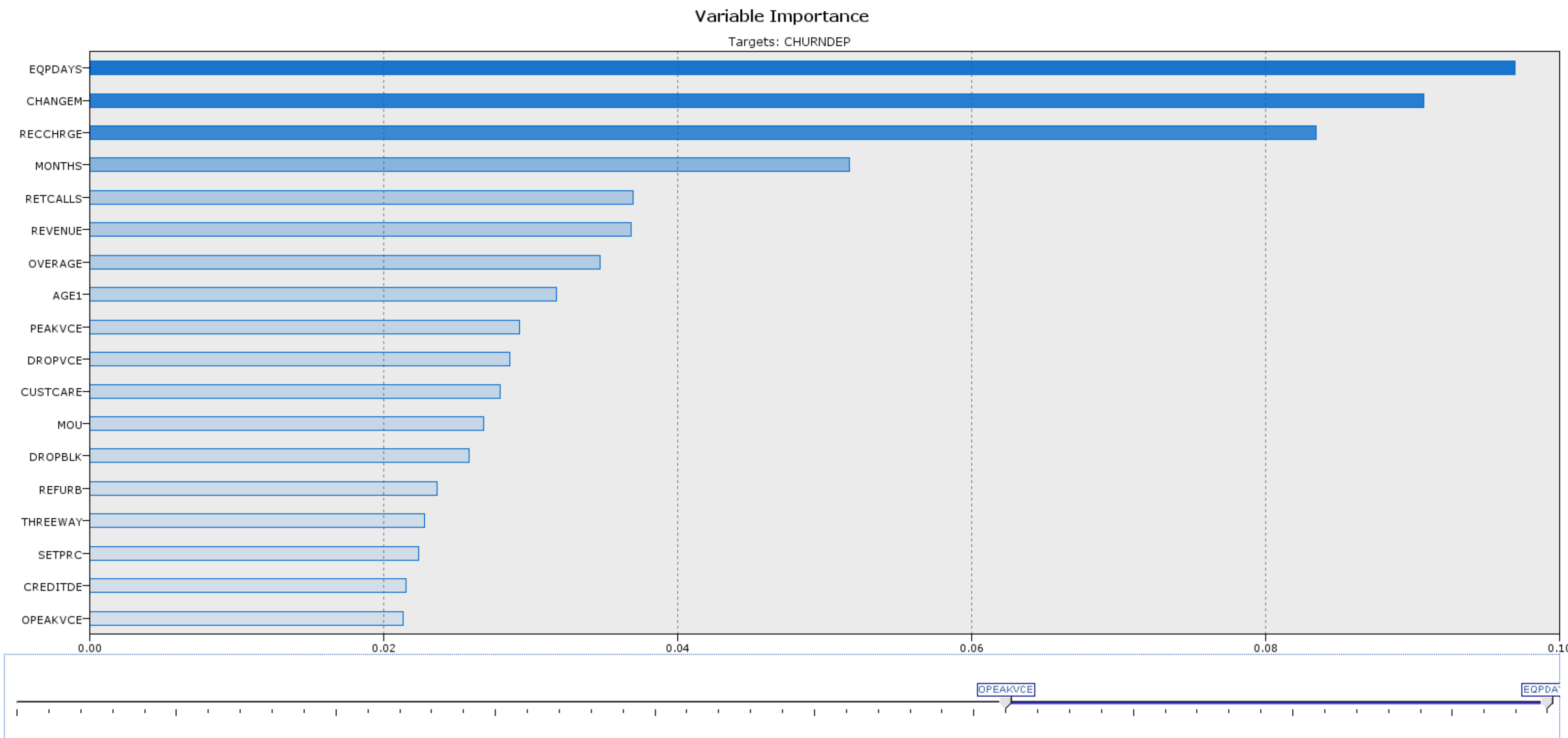



**Figure 6.2: Neural Network Model with variable importance**

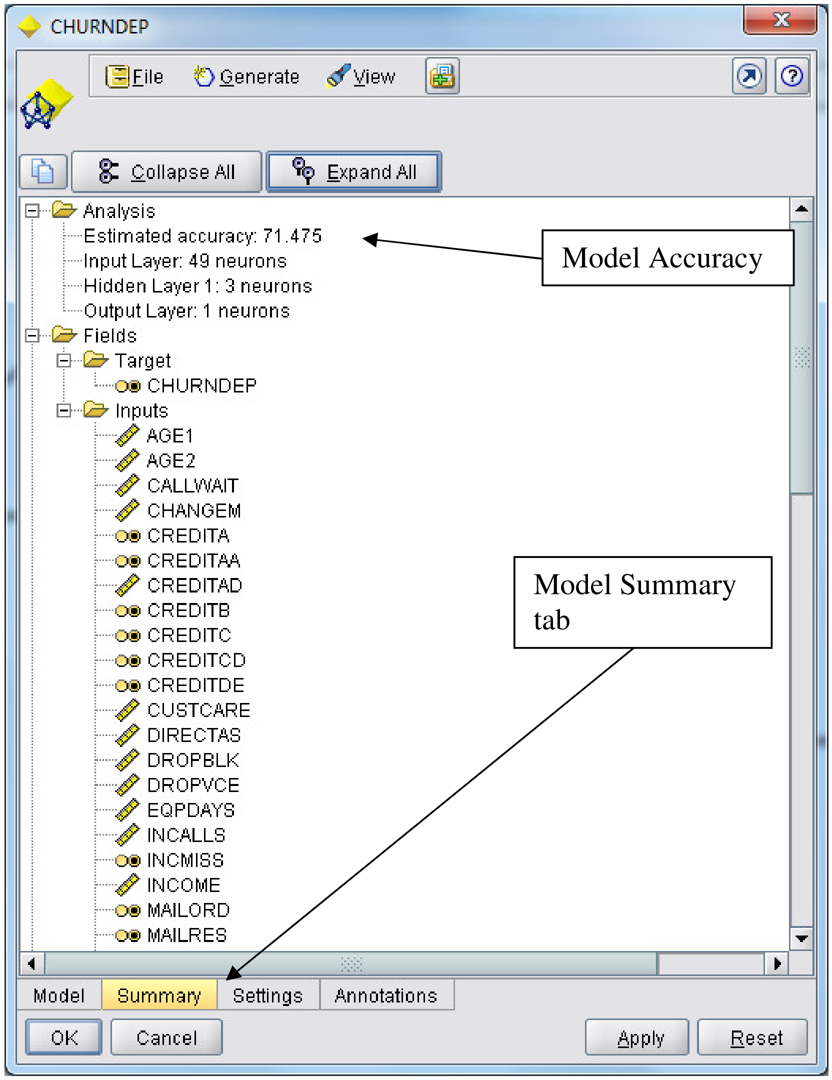



**Figure 6.3: Model Summary including its accuracy**

There are several methods for evaluating the predictive power of the neural network model. However, this stage uses cumulative and incremental lift chart methods, which are sufficient to evaluate the neural network model. First, the cumulative lift chart shows the lift index on the vertical axis, which demonstrates how much better than the random prediction rate was the prediction built by the model (see Figure 6.4). For example, in the 20$^{th}$ percentile, the neural net model performed over 1.3 times better than random prediction (red line).

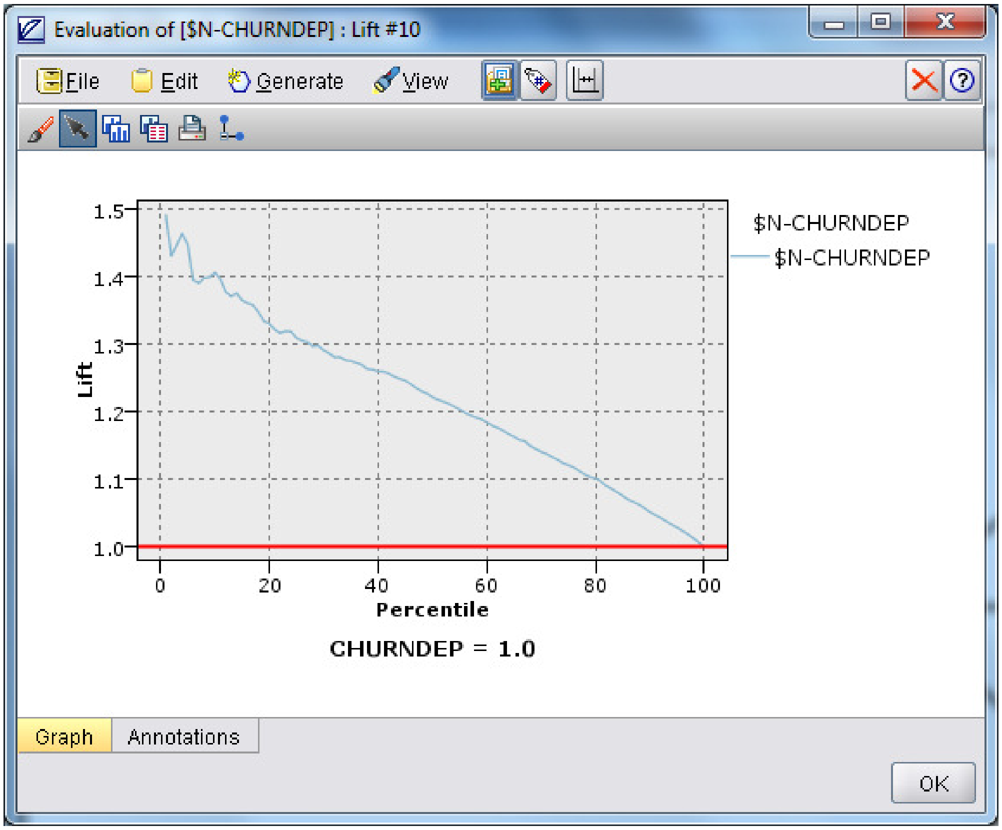



**Figure 6.4: A Cumulative lift chart for the neural network model**

Figure 6.5 presents the incremental lift chart. It shows the lift in each percentile without any accumulation. The lift chart's most significant point is when the lift line goes below the random (red) line at nearly 54%. This shows that all of the benefit of the model (compared to random expectation) is realised in the first 54% of the records. It is important to mention here that the lift chart is built from a list, which is sorted on the predicted probability in descending order. This model can be useful, in a sense, since it can guide the marketers/analysts to provide an incentive to only the top 54% (or whatever percentage the marketing budget permits) in this sorted scored list, and expect to target 1.3 times the number of high probability churners than normal. The response rate can be enhanced by direct marketing campaigns toward high probability churners. Similarly, the effectiveness of the direct marketing campaign can be measured using the response rate as a benchmark.

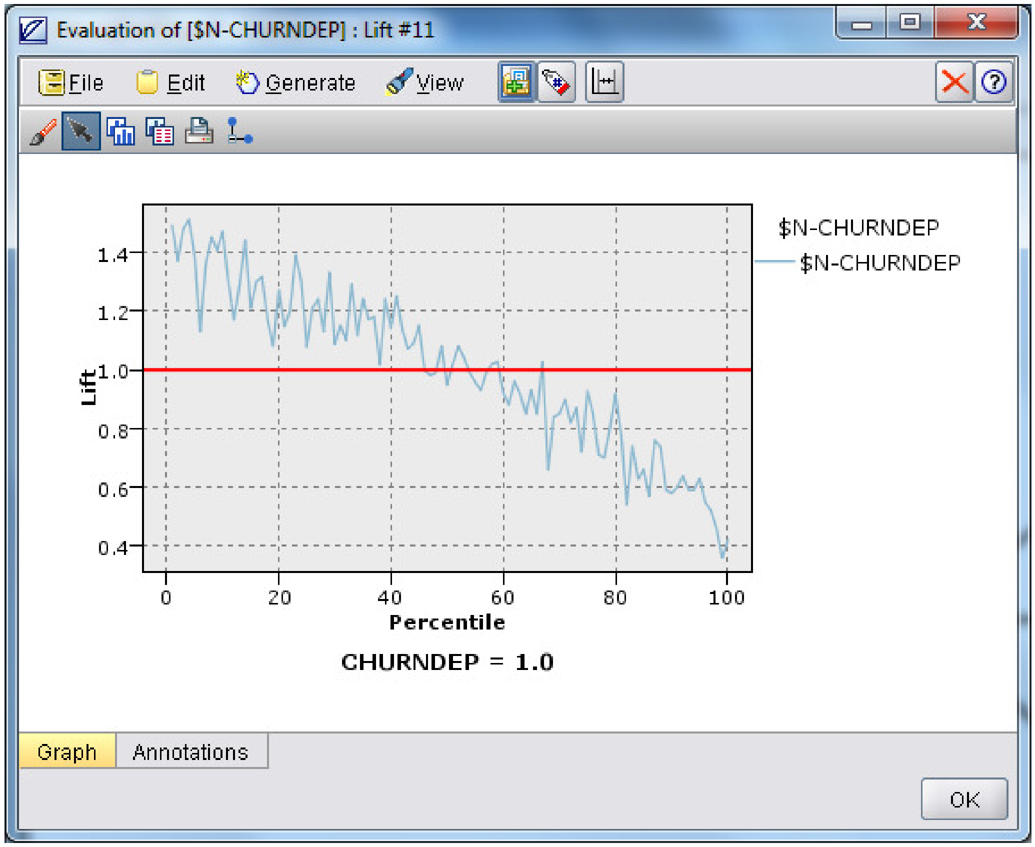



**Figure 6.5: The incremental lift chart for the neural network model**

### 6.2.3.4 Direct Marketing Planning

The neural network model has provided a rich set of useful information for planning a direct marketing campaign. The model revealed the most significant variables that can cause customers to churn. These variables should be used as an incentive to prevent customers from changing their wireless service provider. The following are the four most significant variables that cause customer churn:

1. EQPDAYS: Number of days of the current equipment,
2. CHANGEM: % Change in minutes of use,
3. RECCHRGE: Mean total recurring charge,
4. MONTHS: Months in service.

It is clear that the equipment (i.e. the mobile handset) is a significant driver of a customer's decision to remain with Cell2Cell. Therefore, if customers are targeted with a new mobile handset as an incentive, it would significantly decrease their probability of churning. Similarly, the CHANGEM and RECCHRGE variables can be used as incentives to retain customers. For example, the retention team can target customers with more attractive offers in terms of minutes of use and monthly cost. Indeed, if the retention team can identify an attractive offer, which combines an increase in the minutes of use with keeping the customer's bill at an adequate level, it could have a major impact on customers' decision to remain with the company. The MONTHS variable is useful for informing the retention team of the best period to contact customers with new offers.

After identifying the appropriate offers with which to target customers, it is essential to have a list of customers to directly contact. This is where the incremental lift chart can provide important information. As described earlier, the incremental lift chart provided a list of high probability churners in descending order. Hence, depending on the marketing budget, the highest churners in descending order (maximum 54%) can be contacted with a tailored incentive in an attempt to retain them.

The model has provided a list of high probability churners. This list can be subject to another direct marketing process. For example, the new direct marketing process can analyse the preferences of the customers included in the list of high probability churners. The results can provide further information for using a more direct approach in targeting customers. In fact, customers can be classified into different groups based on their preferences. After that, each group is targeted with offers that aim to satisfy their specific preferences.

## 6.3 Telecommunication Case Study Evaluation

This section evaluates the findings of the Telecommunication case study. It begins by investigating the impact of the activities, which were provided in the stages of the DMP-BI framework. It then examines the impact of BI functions in executing the direct marketing process. Finally, it provides an evaluation of the links between the stages in the DMP-BI framework.

### 6.3.1 The DMP-BI Stages

This subsection evaluates the impact of the four stages involved in the DMP-BI framework. First, the activities provided by the "Direct Marketing Objectives" stage were successfully applied. The "Identify Process Initiator(s)" activity was executed, and the process initiator was classified as an external entity. For the "Consider Key Questions" activity, the study proposition (the study proposition was used to initiate the process, refer to Section 6.3.2) attempts to retain existing customers through "*Customer Selection*" and "*Promotions*" methods. This involved using four questions from Table 3.2 to facilitate the formulation of the process objectives. The following are those questions:

- What benefits do customers want in specific product? (Product Modification)
- How are existing brands (including ours) perceived in the marketplace? (Positioning)
- How price-sensitive are customers? (Promotions)
- What customers would be most receptive to our offerings? (Customer Selection)

The "Identify Suitable Data Source" is the last activity, and was not applicable in this case as the dataset had already been provided.

Second, the "Data Preparation" stage involves three activities including *attribute selection, data sampling,* and *data key facts summary*. In this case, only the first two activities were required. The last activity was not executed as the "Churn Game" case study already provided a "*data key fact summary*" table. Although the "*Data Type Selection*" activity was not included in the DMP-BI framework, it was necessary to execute it in order to ensure data integrity. This activity involves selecting the right data format (refer to Section 6.2.3.2 for more details). As discussed in Section 2.4.2, the direct marketing process is characterised by uncertainty. Therefore, it was expected that in some cases, such as this one, more activities would be required to complete the stage.

Third, the "Data Modelling" stage successfully used the five recommended activities in the DMP-BI framework. It is important to place an emphasis on the "Model(s) deployment" activity. In fact, this activity can be executed using two methods: mass modelling, or selective modelling (select data mining techniques from Table 2.7). In this case, selective modelling was performed because the process objectives are related to "Numerical Prediction". Therefore, the most common data mining techniques in "Numerical Prediction" were used.

Finally, the "Direct Marketing Planning" stage used all the recommended activities in the DMP-BI framework (refer to Section 6.2.3.4).

### 6.3.2 BI Functions

This section evaluates the BI functions used to execute the direct marketing process. It is important to mention that the second case study used the SPSS Clementine business intelligence suite, and not SQL Server 2008. The following are the BI functions used in each stage of the DMP-BI framework.

Similarly to the first case study, the BI reporting functions were not used in the "Direct Marketing Objectives" stage. This is because there was no need to perform document sharing or reporting in this process.

In the "Data Preparation" stage, there were three BI functions used. First, the "Feature Selection" function was used to perform the "*Attribute Selection*" activity. Second, the "Type" function was utilised to perform the "*Data Type Selection*" activity. It was interesting to find that the "*Data Type Selection*" activity was required when using the SPSS Clementine tool. This is because SQL Server 2008, which was used to perform the first case study, performs this activity automatically. This clearly shows that the use of different BI tools to support the DMP-BI framework could have various effects. Finally, the "Data Audit" function was performed to verify data integrity. This activity is not recommended in the DMP-BI framework, but was found to be useful in this particular process.

The "Data Modelling" stage used five BI functions including *neural networks, logistic regression, Naïve Bayes, cumulative lift chart, and incremental lift chart methods*. These functions were all suggested by the DMP-BI framework.

Finally, the "Direct Marketing Planning" stage did not require the usage of BI functions to be completed.

### 6.3.3 The DMP-BI Structure

This section investigates the links between the DMP-BI stages. Specifically, it evaluates the causal link between each stage, and assesses whether the link between each stage is effective. Overall, there are four links to evaluate (refer to Figure 5.5):

- First, the link between the "Direct Marketing Objectives" stage and the "Data Preparation" stage was suitable for the process.
- Second, the link between the "Data Preparation" stage and the "Data Modelling" stage was appropriate for the process. In addition, the iterative link between the two stages was used. Specifically, during the "Data Modelling" stage, additional data sampling was required. Therefore, it was necessary to move back to the "Data Preparation" stage, and perform the "Data Sampling" activity.
- Third, the link between the "Data Modelling" stage and the "Direct Marketing Planning" stage was suitable for the process.

- Fourth, the link between the "Direct Marketing Planning" and the "Direct Marketing Objectives" stages was not applicable in this process.

In this case study, the DMP-BI framework structure, organisation, and iterative feature were suitable for executing the process effectively.

## 6.4 Chapter Conclusion

This chapter illustrated and assessed the usage of the DMP-BI framework in the telecommunication sector. It presented the case study, which aimed to show the application of the DMP-BI framework within a real-world organisational context. This was performed through a step-by-step demonstration of the DMP-BI framework using a telecom company's dataset.

The chapter also provided an evaluation of the impact of the DMP-BI framework in executing the direct marketing process. The evaluation process was based on three themes including *the suitability of the stages of the DMP-BI framework, the applicability of the BI functions, and the organisation and structure*.

This Chapter presented the application of the DMP-BI framework in the telecommunication sector. It has been established that the conceptual framework stages and functions are useful and practical to execute the direct marketing process.

# CHAPTER 7: CASE STUDY III HIGHER EDUCATION SECTOR

## 7.1 Introduction

This chapter presents the third and final case study used to evaluate the DMP-BI framework. The chapter is organised in a similar way to the previous two chapters. It is divided into two main parts, namely case study report and case study evaluation. The first part provides the case study report, which aims to illustrate the usage of the DMP-BI framework in the higher education sector. This case study adopts linear-analytic reporting, which was explained in Section 4.7. This part is structured into three sections. Section 7.2.1 provides a brief overview of general marketing practices in the higher education sector. Section 7.2.2 describes the dataset and the study proposition used for the DMP-BI framework. Section 7.2.3 illustrates the usage of the DMP-BI framework in a higher education context.

The second part evaluates the case study based on the three themes that were described in Section 4.6.1. These themes include *the suitability of the stages of the DMP-BI framework, the applicability of the BI functions, and the organisation and structure*. This part is structured into three main sections. Section 7.3.1 examines the four stages of the DMP-BI framework. Section 7.3.2 assesses the BI functions' impact on supporting the DMP-BI framework to execute the direct marketing process. Section 7.3.3 evaluates the appropriateness of the links between each stage in the DMP-BI framework.

## 7.2 Case Study Report

This case study evaluates the DMP-BI framework in a higher education context. Unlike the other case studies, which both investigated consumer behaviour, this case study investigates the potential benefits of implementing a geographic direct marketing strategy for student attraction and retention. The DMP-BI framework is used to demonstrate the process of developing this strategy, using students' addresses.

This part of the chapter starts by introducing the importance of marketing in the higher education sector. It then presents the dataset and study proposition used to carry out the direct marketing process. Finally, it demonstrates the usage of the DMP-BI framework within a higher education context.

### 7.2.1 Introduction

In the last eight years, there has been a constant increase in student numbers in UK universities. In fact, the number of students between 2002 and 2009 has increased by around 15% (UCAS 2010). However, recent government tax increases and spending cuts will result in a significant reduction in graduates. Consequently, while the latest increase in student numbers is expected to slow, the competition to attract students will probably intensify. As a result, universities' marketing departments need to consider more effective marketing methods to attract students. Indeed, the production of a quality prospectus and recruitment open days are no longer enough to gain a competitive edge (Tapp, Hicks & Stone 2004).

In a marketing context, higher education institutions have to think of students as 'customers' and education as a 'product' (Berger, Wallingford 1997). The availability of large databases provides marketers/analysts with rich data. This data can be used to perform advanced geographic analyses to identify useful patterns, which can enhance the advertising and promotion strategy for higher education (ibid). For example, geographic analyses can assist marketers/analysts in planning direct marketing campaigns through market area analysis.

### 7.2.2 An Overview of the Student Database

This case study aims to illustrate the usage of the DMP-BI framework within a higher education context. This is achieved using data held by Brunel University's[18] student administration system. The system comprises three main components: 1) MAS (Marketing and Admissions System), 2) SRS (Student Registration System), and 3) CAMS (Credit Accumulation System). Table 7.1 presents a summary of the main features of the three parts of the system.

[18] Brunel University is a British higher education institution situated in Uxbridge, West London (http://en.wikipedia.org/wiki/Brunel_University: (Last Accessed 2010).

| MAS | SRS | CAMS |
|---|---|---|
| • Recording enquiries and dispatch of marketing materials<br>• Applications management<br>• Direct application processing<br>• UCAS application processing<br>• GTTR application processing<br>• NMAS application processing<br>• Marvin/Hercules offer codes<br>• Interviews and Open days<br>• Administration of the UCAS link<br>• Statistics and reports<br>• Transfer of applicants to SRS and CAMS | • Enrolment and re-enrolment<br>• Fees processing<br>• Production of HESA and HESES returns<br>• Student statistics and reports<br>• Student progression<br>• Research degree administration<br>• Student loans administration | • Scheme definition and management<br>• Teaching resource structure<br>• Student programme planning<br>• Examination scheduling<br>• Student module scheduling<br>• Student assessment<br>• Student awards |

**Table 7.1: System's main features. Adapted from (SITS Support 2005, p.3-5)**

As shown in the above table, the administration system contains a wide variety of data on past, current, and future students. This offers marketers/analysts useful data to plan and execute many different direct marketing campaigns. However, a study proposition is suggested for putting the direct marketing process within a specific context. The following is the study proposition:

- Geographic analyses of past and present students' location can lead to finding regions with a high student population. Therefore, a direct marketing strategy can be planned for those regions with a high student population or with low student population.

The DMP-BI framework is used to perform the geographic analyses of students' location. The aim is to assess whether geographic direct marketing can be planned using the DMP-BI framework.

### 7.2.3 DMP-BI Application in Higher Education

This section illustrates the DMP-BI framework application in a higher education context[19]. It is organised in four subsections, which are consistent with the framework's four stages (refer to Figure 3.5). These subsections are as follows; 1) Direct Marketing Objectives, 2) Data Preparation, 3) Data Modelling, 4) Direct Marketing Planning. The SQL Server 2008 Business Intelligence tool is used to support the DMP-BI framework execution.

#### 7.2.3.1 Direct Marketing Objectives

This stage entails two main activities: 1) formulate the process objectives, and 2) identify a suitable data source (refer to Section 3.4.3). First, the study proposition is used to formulate the direct marketing objectives, which are as follows:

1. Perform geographic analyses on Brunel students' location, and create clusters which can identify students that are located within 50km of any given location.
2. Look for interesting patterns on Brunel students' home location.
3. Suggest a geographic direct marketing strategy to attract potential students for Brunel.

Second, a suitable data source is required. In this case, the process objectives are to search for patterns of students' locations, thus it was deemed that student addresses are the most appropriate data to use. A dataset containing the names and addresses of students was derived from Brunel's student administration system.

#### 7.2.3.2 Data Preparation

In order to fulfil the process objectives, the Brunel student administration system is used to extract and prepare data. This system contains a wide variety of information on Brunel students, ranging from students' exam results to their home addresses. Therefore, the first activity of this stage is to select the appropriate attributes from this system. The attributes were selected based on the process objectives, and they are as follows:

- Students' Addresses: Address 1, Address 2, Address 3, Postcode.
- Students' Course Mode: Full-time and Part-time.
- Students' Course Group: Undergraduate and Postgraduate.

[19] Some descriptions and techniques were adapted from (Aitchison 2009). Extensive reporting of various configurations to complete the case study can be provided by the author.

This direct marketing process investigates home students only, i.e. students whose addresses are based in the United Kingdom. The resulting dataset has 11,770 students' home addresses.

After completing the attribute selection, the next activity is to import the dataset to the SQL Server Management Studio. Owing to the nature of the dataset, a number of configurations are required to make the SQL Management Studio capable of supporting the geographic data type, i.e. students' addresses. To this end, there are two important activities to complete: 1) data transformation, and 2) configuring the analyses environment. The first activity involves adding a "Geocoding" function to the BI tool in order to identify a precise location for each student in the dataset. The second activity creates a Virtual Earth Map to display students' addresses data.

***Data Transformation***

Prior to completing the data transformation activity, the "Geocoding" function needs to be configured. This is because SQL Server 2008 does not provide a built-in function to perform geocoding. MapPoint Web Service is used to integrate the "FindAddress()" method to perform geocoding.

After configuring the "Geocoding" function, the data transformation activity is performed. This involves transforming students' addresses into a structured spatial representation, which is the coordinates of a single point geometry for each student address. For example, Brunel University's address has been geocoded, and the result obtained is the coordinates of 51.5327 degrees latitude, and -0.4728 degrees longitude. These coordinates correspond to a single point located at Brunel University's address. Figure 7.1 shows the geocoding process of Brunel University's address.

| Brunel University,<br>Uxbridge, Middlesex,<br>UB8 3PH<br>UK | **geocoding** → | **(51.5327, -0.4728)** |
|---|---|---|

**Figure 7.1: The process of geocoding Brunel University's address to a precise location (source: Author)**

After adding the "Geocoding" function to the BI tool, the geocoding of Brunel students' addresses was performed successfully. This was achieved through populating the existing student addresses table[20] with a geography column called "Location" using the following query:

```
ALTER TABLE L1std_UGBBS
ADD Location geography
GO

Using the existing table columns for addresses and the Geocoder
function, we add the geographic information to existing table:

UPDATE L1std_UGBBS
SET Location =
geography::STPointFromText(dbo.Geocoder('','','',Postcode,'UK'), 4326)
GO
```

***Displaying Data using Web Mapping Services***

This task involves the development of a new website, where a Virtual Earth Map application programming interface (API) is implemented. The purpose of this website is to display geographic data from the BI tool directly onto the Virtual Earth Map. Figure 7.3 is the web page with the Virtual Earth Map displaying the region of London and its surrounding cities and towns.

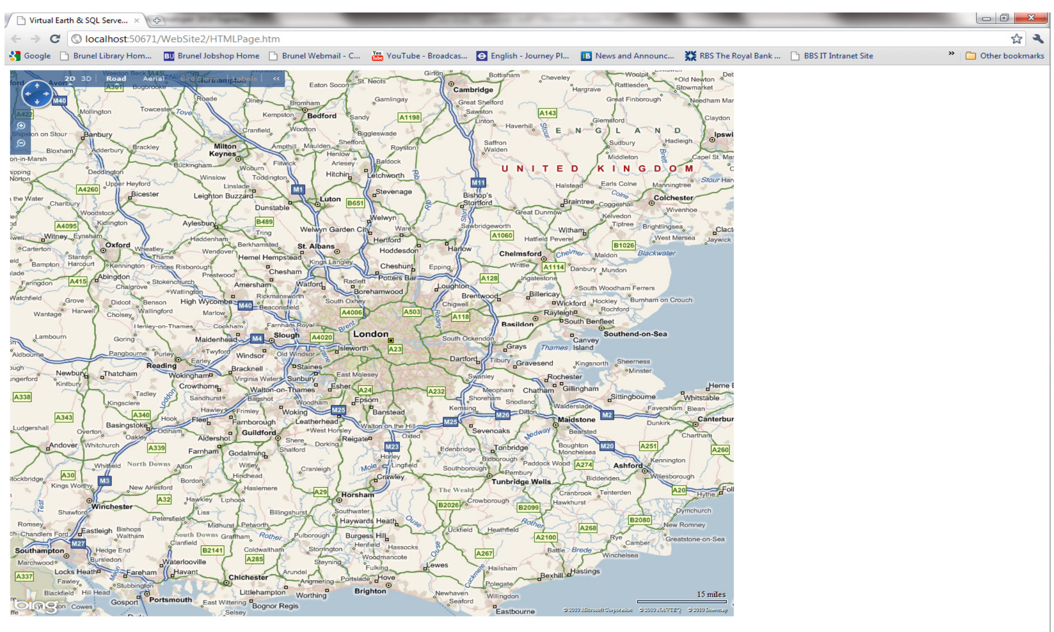



**Figure 7.2: The new webpage with Virtual Earth Map**

### 7.2.3.3 Data Modelling

This stage aims to build data models to look for patterns by analysing Brunel students' locations. There are usually five activities involved in this stage: 1) select data mining

[20] The name of the table that contains the student addresses is L1std_UGBBS.

approach, 2) model(s) deployment, 3) model(s) accuracy, 4) model(s) selection, and 5) model(s) analyses

First, an undirected data mining approach is appropriate for this stage, based on the process objectives. An undirected data mining approach entails clustering, finding association rules, and description (refer to Section 3.4.5). In this case, clustering seems to be more appropriate, based on the process objectives (refer to Section 7.2.3.1). Therefore, a clustering technique can be used to perform the second activity, which is model(s) deployment. However, owing to the unavailability of a geographic clustering technique in SQL Server 2008, an alternative technique is proposed. It consists of developing an advanced query to display Brunel's past and present students' locations within a 50km radius of any given location. The reason for displaying data on a 50km basis is to facilitate the analysis process.

Model accuracy and selection activities are not relevant in this process because of the models deployed. Specifically, the model(s) deployed consist of data in maps, where accuracy is not relevant. Also, the model(s) selection activity cannot be performed without model(s) accuracy. The following subsection is the final activity; that is, model(s) analyses.

***Analyses of Models***

The .NET handler executes the advanced query and displays the latitude and longitude parameters of the point that was clicked on the map. In this case, Brunel University was clicked and Figure 7.4 illustrates the results[21]. It displays students whose home addresses are within 50km of Brunel University. The advantage of the created procedure is that it can identify Brunel students' locations within 50km of any point on the map that is clicked on. For example, Brunel students who live within 50km of East London or Kent can also be displayed. Owing to confidentiality, models that were used for analyses cannot be included in this study.

[21] To preserve anonymity, some records in the dataset have been recoded.

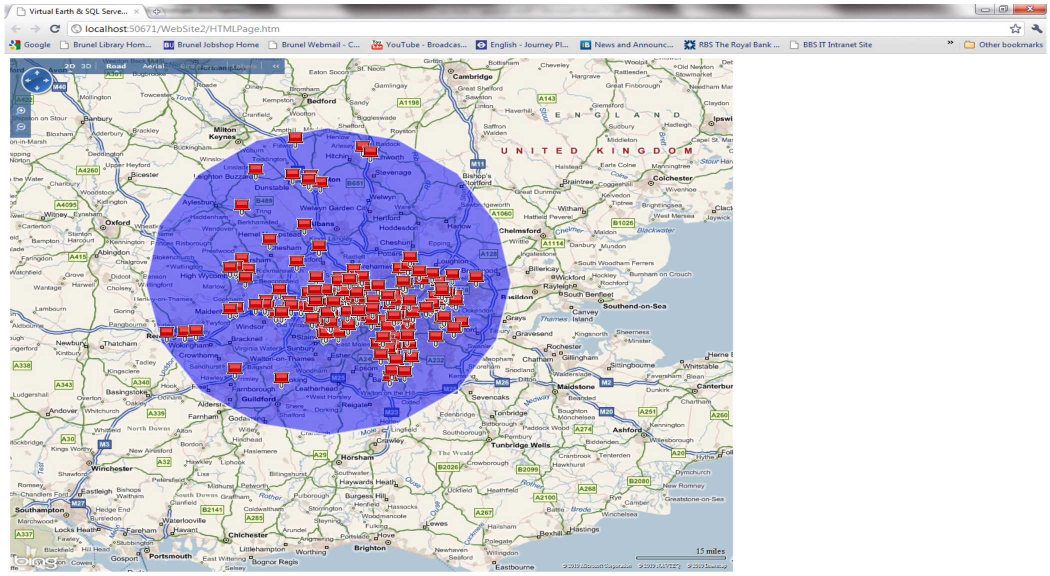



**Figure 7.3: Students' home addresses located within 50km of Brunel University**

Using the .NET handler technique, a number of models are deployed. These models are divided into five different parts.

The first model is built from the overall population of the dataset. This includes all past and current Brunel students from all the different pathways (i.e. undergraduate, postgraduate, and research), and course modes (i.e. full-time and part-time). The model's results demonstrate that a significant percentage of Brunel students are from within a 50km radius of Brunel University. The results also show that the following regions have the highest Brunel University past and present student populations:

- London, especially west London and its surroundings,
- Birmingham and its surroundings,
- Winchester and its surroundings, including Bournemouth, Portsmouth and Southampton,
- Manchester and its surrounding, including Blackburn, Bradford, and Bolton,
- Norwich and its surroundings, including Ipswich,
- Bristol and its surroundings.

The second model is built from all past and current Brunel students on a full-time course mode. The model's results reveal that a high number of students are from Birmingham and its surroundings. Also, the results show that Brunel is a local university, based on the large number of students found within a 50km radius of Brunel

University. In addition, the model identified six other regions with high numbers of students, and they are as follows:

- Manchester and its surroundings,
- Ipswich and its surroundings,
- Bristol and its surroundings,
- Plymouth and its surroundings,
- Bournemouth and its surroundings,
- Canterbury and its surroundings.

The third model is built from all past and current Brunel students on a part-time course mode. The model's results reveal four main cities with high numbers of Brunel students. The following are the identified cities:

- Brighton and its surroundings,
- Birmingham and its surroundings,
- Bristol and its surroundings,
- London and its surroundings,
- Manchester and its surroundings.

The fourth model is built from all past and current Brunel students who are enrolled on an Undergraduate course, whether full-time or part-time. The model's results show the following cities as being the most populated with Brunel students:

- London and its surroundings,
- Bournemouth and its surroundings,
- Bristol and its surroundings,
- Ipswich and its surroundings,
- Birmingham and its surroundings.

Finally, the model is built from full-time and part-time postgraduate students. It reveals that most students on postgraduate courses are located in London, Birmingham, Manchester, and Southampton and their surroundings.

#### 7.2.3.4 Direct Marketing Planning

This stage is used to interpret the analyses performed in the data modelling stage. It is also intended to design a direct marketing plan for use towards specific targets. Based on the analyses' findings, one can identify many interesting patterns that can be used to plan a direct marketing strategy. The following is a summary of key findings and suggestions for possible actions that could be taken by Brunel's marketing department in order to improve marketing campaigns.

Brunel University is primarily a local university. This is illustrated by the significant percentage of Brunel students whose home addresses are located within 50km of Brunel University. Birmingham is the second city from which many students are attracted to Brunel. Indeed, the five models deployed for the analyses revealed that Birmingham was a major source of Brunel students. The proximity of Brunel University to Birmingham could be the reason for it. Next, Manchester and Bristol are also found to have high number of Brunel students. This could be related to the big population of both cities. The regions of Bournemouth, Southampton, and Winchester come after Manchester and Bristol in terms of student numbers. They are followed by Ipswich, Norwich, Plymouth, Canterbury, and Brighton. In terms of course mode, it is interesting to find that Canterbury and Plymouth have a high percentage of Brunel students enrolled on full-time basis. For part-time courses, it is equally interesting to find Brighton is a major source of Brunel part-time students.

Based on these findings, a direct marketing strategy could be planned to target more extensively the local colleges in the cities which are found to have a high percentage of Brunel students. This might be achieved through organising special events in the targeted colleges. These events could offer an overview of current courses available at Brunel University and help potential students to apply for Brunel. Canterbury, Plymouth and Brighton could be targeted with specific direct marketing campaigns. These campaigns should be tailored to meet the specific requirements of the course mode characteristics. For example, part-time students should be made aware of Brunel's Virtual Learning tool, which provides course materials and central services online.

## 7.3 Higher Education Case Study Evaluation

This section provides a comprehensive evaluation of the third case study. It starts by investigating the usage of the activities included in each stage of the DMP-BI framework. Next, it evaluates the impact of BI tools' functions in the direct marketing process. Finally, it examines the link between each stage and assesses whether the structure and organisation of the DMP-BI framework are effective.

### 7.3.1 DMP-BI Stages

In this subsection, an in-depth evaluation of the DMP-BI stages is provided. Specifically, it investigates the impact of the activities provided by the DMP-BI framework on the direct marketing process. There follows an evaluation of each stage.

The "Direct Marketing Objectives" stage used all the three recommended activities of the DMP-BI framework. The first activity was to categorise the process into an external entity (refer to Section 7.2.3.1). The second activity used the following questions from Table 3.2 to formulate the process objectives:

- How are existing brands (including ours) perceived in the marketplace? (Positioning)
- How price-sensitive are customers? (Promotions)
- What customers would be most receptive to our offerings? (Customer Selection)

The last activity of this stage was to identify a suitable data source to fulfil the direct marketing objectives.

The "Data Preparation" stage used only the "Attribute Selection" activity from the recommended activities in the DMP-BI framework. There were additional activities performed, but these activities are not included in the DMP-BI framework. This was due to the data type used in this process, i.e. students' addresses. The additional activities were "Data Transformation" and "Analyses Environment Configuration". As mentioned earlier and in Section 2.4.2, the direct marketing process is highly uncertain, and this process is another example of it.

In the "Data Modelling" stage, three activities were executed, namely *select data mining approach, model(s) deployment,* and *model(s) analyses*. The "*select data mining approach*" activity was completed using the process objectives. The "*model(s) deployment*" activity was executed using an advanced query that creates 50km clusters for students' locations. This query was developed because SQL Server 2008 does not provide a data mining technique which supports spatial analyses. Therefore, neither the "*model(s) accuracy*", nor the "*model(s) selection*" activities were performed.

The "Direct Marketing Planning" stage used all the activities that are recommended by the DMP-BI framework. In fact, this stage was successfully completed and provided a series of suggestions to plan a geographic direct marketing campaign.

### 7.3.2 BI Functions

In this subsection, the BI functions' impact is assessed. The section is organised into four parts, where each part describes the impact of BI functions in each stage of the DMP-BI framework.

First, the "Direct Marketing Objectives" stage did not use any BI functions. This is similar to the other two case studies, where BI functions were not applicable. The reason is that the direct marketing process does not need to be shared across the organisation under study.

The "Data Preparation" stage performed four activities using BI functions: *Attribute Selection*, *Data Import*, *Data Transformation*, and *Environment Configuration*. "*Attribute Selection*" is the only activity that is suggested by the DMP-BI framework. The other activities are additional, but they are equally important for the "Data Modelling" stage. "*Attribute Selection*" and "*Data Import*" were performed using SQL Server database management functions. The other two activities are used to configure an analysis platform in the SQL Server tool. This is because SQL Server 2008 does not by default support spatial data. Therefore, it was necessary to configure two BI functions, namely *geocoding, and Virtual Earth Map*, to perform "*Data Transformation*" and "*Environment Configuration*". This involved a user with technical skills configuring these functions. In fact, the user skills level is a common issue in the direct marketing process, as already discussed in Section 2.4.2. This stage illustrates the complexity and the different level of skills that are required to execute a direct marketing process.

The "Data Modelling" stage used an advanced query function with a .NET handler technique to perform "*Model(s) Deployment*" and "*Model(s) Analyses*" activities. It would have been more appropriate to use a data mining clustering technique to perform these activities. However, neither SQL Server 2008 nor SPSS Clementine tools provide a data mining clustering technique for spatial data analysis. This is considered as a major drawback.

The "Direct Marketing Planning" stage was completed without the use of BI functions. This is because the proposed BI functions for this stage are not applicable in this particular process. Specifically, there is no need for either document sharing or reporting activities in this process.

### 7.3.3 The DMP-BI Structure

This subsection investigates the DMP-BI framework structure and organisation in terms of providing the correct linkage between different stages to execute the direct marketing process. In this case study, the process was hindered by many uncertainties, mostly related to the unavailability of important BI functions. However, the stages were executed in a systematic way, where each stage provided the relevant information and outcomes to perform the next stage.

## 7.4 Chapter Conclusion

This chapter demonstrated and evaluated the usage of the DMP-BI framework within a higher education context. It provided a brief overview of the importance of marketing in the higher education sector a step-by-step illustration of the DMP-BI application within an organisational context. It also assessed the usefulness of each stage and activity within the DMP-BI framework. Moreover, the chapter examined the impact of the BI functions in supporting the DMP-BI framework to execute the direct marketing process. Furthermore, it also discussed the relevancy of the links between each stage of the DMP-BI framework.

The chapter showed the application of the DMP-BI framework in the higher education sector. Although, the BI tool used to execute the direct marketing process did not provide key functions, it has been demonstrated that the guidance of the developed framework is still a valuable support to complete the process.

# CHAPTER 8: CROSS-CASE STUDY EVALUATION AND COMPARISON

## 8.1 Introduction

This chapter provides a cross-case evaluation of the three case studies and compares the DMP-BI framework with three rival methodologies. The chapter is organised into two parts. The first part evaluates the three case studies using the "Relying on theoretical propositions" strategy, which was described in Section 4.6.1. This technique is used to investigate the impact of each component of the DMP-BI framework on the direct marketing process. The DMP-BI framework is composed of six components: a) the "Direct Marketing Objectives" stage, b) the "Data Preparation" stage, c) the "Data Modelling" stage, d) the "Direct Marketing Planning" stage, e) the BI functions, and f) the "Iterative Process" feature. These components are used to facilitate the evaluation process of the DMP-BI framework. The purpose of the cross-case evaluation section is to compare the impact of similar components between each case study. This aims to assess the overall impact of each component in the three case studies.

The second part provides a discussion on the practicality and usefulness of the DMP-BI framework by comparing it with three rival methodologies, which are selected from the literature review chapter (see Sections 2.4.3 and 2.4.4). Indeed, the comparison is based on the "Examining rival explanations" strategy (see Section 4.6.3). This strategy involves comparing the DMP-BI components with similar components in the rival methodologies, which will allow identification of the key benefits that the DMP-BI framework has over these methodologies. Finally, a summary of the chapter is provided.

## 8.2 Cross-Case Studies Evaluation

This part of the chapter provides a cross-case evaluation of the three case studies. It uses the "Cross-case Synthesis" technique that was described in Section 4.6.4. This technique requires the researcher to design a table that displays the patterns which were identified in each case study. The aim is to investigate cross-case patterns, and enable the building of strong, plausible, and fair arguments in order to empirically support the DMP-BI framework's effectiveness. Table 8.1 illustrates the key patterns identified in each case study. The table includes the six components that constitute the DMP-BI framework: 1) Direct Marketing Objectives, 2) Data Preparation, 3) Data Modelling, 4)

Direct Marketing Planning, 5) DMP-BI System Platform, and 6) Iterative Process. It also contains the activities involved in each stage. The aim is to identify and compare the usage of these components and activities between the three case studies. The following is a comparison of the impact of each component, activity, and function between the three case studies.

### 8.2.1 The DMP-BI Stages

This subsection compares the usage of the DMP-BI stages between the case studies. The following is the comparison.

First, the “Direct Marketing Objectives” stage includes three activities. The first two activities have been used to identify the process objectives in all three case studies. This shows the essentiality of these activities in completing this stage in particular, and the overall process in general. The third activity is only used in the third case study, because the datasets for the first two case studies were already provided by the organisations. This also demonstrates that the activities in each component are recommendations. Their usage is dependent on the process objectives, and there are some cases where they are not necessary. The outcome of the first component is a list of objectives that will guide the rest of the process.

The second component is the “Data Preparation” stage, which comprises three activities. The first activity is “*Attribute Selection*” and it was used in all the three case studies. The second and third activities are only used in the second and third case studies. As mentioned earlier, the usage of these activities depends on the process objectives. The outcome of the second component is a prepared dataset, ready for model(s) deployment.

The “Data Modelling” stage is the third component, and involves five activities. The first and second case studies used the five activities to complete the stage. It is interesting to mention that both case studies aimed to investigate consumers’ behaviour. On the other hand, the third case study did not involve the usage of the “*Model(s) accuracy*” and “*Model(s) Selection*” activities because they are not applicable to the models that were deployed. Specifically, owing to the unavailability of a data mining clustering technique for geographic data in the BI tool used, it was necessary to use an alternative method (i.e. an advanced SQL query) to perform the analyses.

| DMP-BI Components and Activities | Case Study 1 | Case Study 2 | Case Study 3 |
|---|---|---|---|
| **DMP-BI Stage 1:** | **Direct Marketing Objectives** | | |
| **Activity 1:** Identify & Classify the Entity that Initiates the Process as Internal or External | X | X | X |
| **Activity 2:** Use Table 3.2 Questions | X | X | X |
| **Activity 3:** Identify a Suitable Data Source | | | X |
| **Outcome:** List of Objectives | X | X | X |
| **DMP-BI Stage 2:** | **Data Preparation** | | |
| **Activity 1:** Attribute Selection | X | X | X |
| **Activity 2:** Data Sampling | | X | |
| **Activity 3:** Data Key Facts Summary | X | | |
| **Outcome:** Prepared Data | X | X | X |
| **DMP-BI Stage 3:** | **Data Modelling** | | |
| **Activity 1:** Select a Data Mining Approach | X | X | X |
| **Activity 2:** Model(s) Deployment | X | X | X |
| **Activity 3:** Model(s) Accuracy | X | X | |
| **Activity 4:** Model(s) Selection | X | X | |
| **Activity 5:** Model(s) Analyses | X | X | X |
| **Outcome:** Model(s) Analyses | X | X | X |
| **DMP-BI Stage 4:** | **Direct Marketing Planning** | | |
| **Activity 1:** Objectives Achieved? | X | X | X |

| | | | |
|---|---|---|---|
| **Activity 2:** Review Key Patterns | **X** | **X** | **X** |
| **Outcome 1:** Suggestions for more Analyses | **X** | **X** | **X** |
| **Outcome 2:** A Direct Marketing Campaign | **N/A** | **N/A** | **N/A** |
| **DMP-BI System Platform** | **BI Functions** | | |
| Reporting | **N/A** | **N/A** | **N/A** |
| Data Integration | **X** | | **X** |
| Visualisation Techniques | **X** | **X** | **X** |
| Database Management | **X** | **X** | **X** |
| Analytical Techniques | **X** | **X** | |
| Dashboards & Scorecards | | | |
| BI Search | **N/A** | **N/A** | **N/A** |
| **Iterative Process** | **X** | **X** | |

**Table 8.1: Key Patterns in Each Case Study (source: Author)**

It is interesting to observe that this case study was not related to investigating consumers' behaviour. In fact, the aim of the third case study was to analyse students' locations in an attempt to identify the regions that provide most students. In other words, it was not required to analyse the characteristics of the students, but only their location. In short, this direct marketing process context (geographic analyses) differs from the first two case studies (consumers' behaviour).

The fourth component is the "Direct Marketing Planning" stage, which entails two activities. These two activities have been successfully employed in the three case studies. Indeed, these activities facilitated the process of identifying a series of suggestions to plan a direct marketing campaign or perform a new direct marketing process. However, the three case studies' outcomes did not provide an explicit direct marketing campaign. This is because of the lack of access to the organisations' information such as budget allowance. Therefore, "Outcome 2" was set to Not Applicable (N/A) in Table 8.1.

### 8.2.2 The DMP-BI System Platform

This section compares the usage of the BI functions between the case studies. There are seven functions that are suggested for supporting the DMP-BI stages.

First, the "*BI Reporting"* and *"BI Search*" functions were not applicable in the three case studies. These functions are used to support the first and fourth stage of the DMP-BI framework. Since the research has limited access to organisational settings, it was not necessary to use these functions, as no document sharing or reporting are required.

Second, the "*Database Management*" set of functions have been used in the three case studies. These include attribute selection, data manipulation, data audit, query design, and import/export.

Third, the "*Data Integration*" function is used in the first and third case studies. Similarly to the activities of the DMP-BI stages, BI functions are required based on the process objectives. There are examples, such as the second case study, where the "*Data Integration*" function is not required.

Fourth, "*Dashboards & Scorecards*" are not used in the three case studies. This is because the process objectives do not require these functions.

Fifth, the "*Analytical Techniques*" functions were used in the first and second case studies, and were critical in completing the "Data Modelling" stage. However, the BI tools that were used for analyses did not provide analytical techniques that support geographic data analyses. Therefore, analytical techniques were not used in the third case study. Finally, visualisation techniques are also used in all case studies. They are used by other BI functions to facilitate the user process of finding interesting patterns. For example, a decision tree model can be visualised in different ways.

### 8.2.3 The Iterative Process

The fifth component represents the iterative nature of the DMP-BI framework. This component was used in the first and second case studies, where further data preparation was needed to effectively deploy the models. This shows the validity of the framework's iterative feature, where the DMP-BI framework offers the ability to move back to previous stages.

### 8.2.4 Summary of Key Findings

This section presents the key findings concluded from the comparison of the components of the DMP-BI framework between the case studies. The cross-case evaluation of the DMP-BI framework has revealed interesting findings. Most notably, it shows the importance of the process context in using the components' activities and functions. In particular, the first and second case studies both investigated consumers' behaviour, whereas the third case study performed geographic analyses. The difference has been reflected in the usage of the components' activities and functions when comparing the first and second case studies with the third case study.

Moreover, although a different BI tool was used for the first and second case studies, they mostly used the same activities and functions apart from the "*Data sampling*" activity and the "*Data integration*" function. Furthermore, the first and third case studies used the same BI tool, but this has not resulted in similarity in terms of the usage of the DMP-BI components.

Based on the above findings, the DMP-BI components are effective where the direct marketing process is related to investigating consumers' behaviour. However, if the context changes, like in the third case study, then the components of the DMP-BI framework may be less effective than when the process is related to consumers' behaviour.

## 8.3 Comparison & Discussion

The discussion section compares the DMP-BI framework with three rival methodologies, namely CRISP-DM (Chapman et al. 2000), the Personalisation Process (referred to as the PPV model) (Vesanen, Raulas 2006), and the Iterative Personalisation Process (referred to as the IPA model) (Adomavicius, Tuzhilin 2005). These models were described and discussed in Sections 2.4.3 and 2.4.4. This section also aims to clearly define the benefits of the DMP-BI framework over these rival methodologies. In order to achieve this, Table 8.2 includes the components of the DMP-BI framework along with the three rival methodologies. This provides a platform on which an effective comparison of the DMP-BI framework and rival methodologies can be performed. The following is a discussion and comparison of the six components that constitute the DMP-BI framework with the selected three rival methodologies.

### 8.3.1 Direct Marketing Objectives

The first component of the DMP-BI framework consists of three main activities and one major outcome. This component is compared with the first stage(s) of the three rival methodologies. First, the CRISP-DM methodology provides two stages prior to the data preparation stage, namely "*Business Understanding*" and "*Data Understanding*" (Chapman et al. 2000). The "*Identify a Suitable Data Source*" activity and "*List of Objectives*" outcome are the only common features between the DMP-BI framework and the CRISP-DM methodology. However, the "*Identify & Classify the Entity that Initiates the Process*" activity can be related to the "*Determine Business Objective*" activity in the CRISP-DM methodology. Indeed, both activities aim to find a common entity, which describes the entire process purpose and which can be used to identify the list of objectives. In addition, CRISP-DM suggests the usage of reporting in the "*Business Understanding*" and "*Data Understanding*" stages. This is also the case in the DMP-BI framework, where the BI reporting function is proposed for supporting the first stage.

The other two models (i.e. PPV and IPA) both include the "*Identify a Suitable Data Source*" activity. But, they did not include the other activities, or the outcome involved in the DMP-BI framework. Instead, they both included the "*Building Consumer Profiles*" activity in the early stages of their process models.

| DMP-BI Components and Activities | CRISP-DM | PPV Model | IPA Model |
|---|---|---|---|
| **DMP-BI Stage 1:** | **Direct Marketing Objectives** | | |
| **Activity 1:** Identify & Classify the Entity that Initiates the Process as Internal or External | X | X | X |
| **Activity 2:** Use Table 3.2 Questions | X | X | X |
| **Activity 3:** Identify a Suitable Data Source | √ | √ | √ |
| **Outcome:** List of Objectives | √ | X | X |
| **DMP-BI Stage 2:** | **Data Preparation** | | |
| **Activity 1:** Attribute Selection | √ | X | X |
| **Activity 2:** Data Sampling | √ | X | X |
| **Activity 3:** Data Key Facts Summary | X | X | X |
| **Outcome:** Prepared Data | √ | X | X |
| **DMP-BI Stage 3:** | **Data Modelling** | | |
| **Activity 1:** Select a Data Mining Approach | X | X | X |
| **Activity 2:** Model(s) Deployment | √ | √ | √ |
| **Activity 3:** Model(s) Accuracy | √ | X | X |
| **Activity 4:** Model(s) Selection | √ | X | X |
| **Activity 5:** Model(s) Analyses | √ | √ | X |
| **Outcome:** Model(s) Analyses | √ | X | X |
| **DMP-BI Stage 4:** | **Direct Marketing Planning** | | |

| | | | |
|---|---|---|---|
| **Activity 1:** Objectives Achieved? | **X** | **X** | **X** |
| **Activity 2:** Review Key Patterns | **X** | **X** | **X** |
| **Outcome 1:** Suggestions for more Analyses | √ | √ | √ |
| **Outcome 2:** A Direct Marketing Campaign | **X** | √ | √ |
| **DMP-BI System Platform** | **BI Functions** | | |
| Reporting | **X** | **X** | **X** |
| Data Integration | **X** | **X** | **X** |
| Visualisation techniques | **X** | **X** | **X** |
| Database Management | **X** | **X** | **X** |
| Analytical Techniques | **X** | **X** | **X** |
| Dashboards & Scorecards | **X** | **X** | **X** |
| BI Search | **X** | **X** | **X** |
| **Iterative Process** | √ | √ | √ |

**Table 8.2: The DMP-BI Framework vs Three Rival Methodologies (source: Author)**

### 8.3.2 Data Preparation

The second component includes three main activities and one outcome. On the one hand, the PPV and IPA models do not include any of these activities and do not mention anything related to data preparation. On the other hand, the CRISP-DM methodology comprises a data preparation stage, which includes similar activities to the DMP-BI data preparation stage. The only difference between the two stages is the "data key facts summary" activity. In fact, while CRISP-DM includes this activity in the prior stage (data understanding), the DMP-BI framework includes it at the end of the data preparation stage. This is because the selected dataset, in the prior stage(s) to data preparation, may be large and complicated to summarise. Moreover, the data preparation stage will only include data that is relevant to the process, whereas if the data key facts summary is done in the early stages, it might includes attributes or data that is not appropriate for the process objectives.

### 8.3.3 Data Modelling

The data modelling stage includes five activities and three BI supporting functions. As expected, the "*Model(s) Deployment*" activity is present in all three rival methodologies, given its essentiality. On the other hand, the "*Select a Data Mining Approach*" is not included in any of the rival methodologies. The CRISP-DM methodology involves model selection, building, description, and assessment. These activities are included in the DMP-BI framework, but with more recommendations compared to CRISP-DM. For example, while the DMP-BI framework recommends that marketers/analysts use either mass modelling or selective modelling, CRISP-DM does not recommend any of these activities. In addition, CRISP-DM involves the model analyses in its evaluation stage, whereas the DMP-BI includes them in the data modelling stage.

The PPV model includes a "Processing" stage which only involves two activities, namely "*Profiling*" and "*Segmentation*" of customers. Similarly, the IPV model only suggests constructing consumers' profiles, and then performing matchmaking to develop personalisation applications. It also includes technologies that are used for matchmaking, such as the recommender system and statistically based predictive approaches. In comparison, the DMP-BI framework not only suggests profiling and segmentation of customers, but it also recommends other analyses activities such as deviation and trend analysis. This clearly shows that the DMP-BI provides more comprehensive reporting of the activities involved in investigating consumers' behaviour.

### 8.3.4 Direct Marketing Planning

The "Direct Marketing Planning" stage in the DMP-BI framework corresponds to the "Deployment" stage in the CRISP-DM methodology. The activities in both DMP-BI and CRISP-DM differ in context, but the overall purpose is similar. In other words, both stages aim to plan a strategy of actions for the process objectives. However, unlike the CRISP-DM methodology, which provides general recommendations, the DMP-BI framework provides suggestions in a specific context (i.e. direct marketing).

The PPV model does not include the activities suggested in the "Direct Marketing Planning" stage, but it does provide marketing-related recommendations similar to the ones in the DMP-BI framework. For example, it recommends that marketers/analysts should personalise the product/service and communication method to interact with the targeted customers.

The IPA model suggests the usage of visualisation, lists ordered by relevance, and unordered lists of alternatives to deliver personalised information to consumers. Similar activities are suggested in the DMP-BI framework. In addition, the IPA model requires the measuring of the personalisation impact, which is also recommended in the DMP-BI framework.

It is also important to consider the expected outcomes in the DMP-BI framework compared with the three rival methodologies. Indeed, the first outcome is reflected in the description of all three rival methodologies, while the second outcome is only reflected in the PPV and IPA models. This could be related to the context of direct marketing that both models possess, whereas CRISP-DM attempts to provide a methodology to cover a large scope of contexts. This makes it abstract when compared with models tailored to the direct marketing context.

### 8.3.5 DMP-BI System Platform

The CRISP-DM and PPV models did not provide a specific system platform on which the process can be executed. The IPA model did suggest the usage of recommender systems, statistically based predictive approaches, and rule-based systems. However, these are abstract suggestions, without a description of the functions and capabilities of these systems. Therefore, none of the rival methodologies provided an integrated information system platform with specific functions to support the stages and activities involved in each methodology. As mentioned in Section 2.4.5, an integrated information system platform is an important aspect to consider when executing a direct marketing process.

It is important to mention that BI tools were effectively used for the first two case studies. However, the BI tool used in the third case study was less effective, where an alternative analytical method was developed due to the lack of analytical techniques able to support geographic data. This prompted the researcher to investigate other tools able to support more effectively the third case study's direct marketing process. Microsoft MapPoint 2010,[22] along with the MPCluster,[23] add-in was found to be a better specialised tool to deal with geographic analysis. In fact, unlike MS SQL Server 2008, where the user has to make many configurations prior to deploying models, the MapPoint 2010 tool supports geographic data and does not need any particular configurations. Furthermore, the availability of the MPCluster add-in provided the ideal analytical technique to achieve the third case study's process objectives. Therefore, there are specialised tools that can, in some cases, provide more appropriate functions for the direct marketing process than BI functions.

### 8.3.6 Iterative Process

The iterative process feature is provided by all the rival methodologies. This is because direct marketing and data mining processes are highly likely to involve more analyses, hence requiring a new direct marketing process or data mining project. Also, the possibility of moving back and forward between stages is available in the CRISP-DM methodology, but the PPV and IPA models did not mention the possibility of moving between stages. In fact, they just mentioned that a direct marketing process is commonly a loop which triggers more analyses to be performed.

## 8.4 Chapter Conclusion

This chapter provided a comprehensive evaluation and discussion of the DMP-BI framework. It provided a cross-case evaluation of the impact of the DMP-BI framework in the three case studies. Indeed, an in-depth cross-case evaluation of the usage of each component of the DMP-BI framework between the three case studies was performed. This aimed to identify cross-case patterns and provide empirical findings to support the

[22] MapPoint 2010 is a Microsoft product that empowers organisations with visualisation techniques to display business data, communicate insights with instant impact and integrate maps into the work they do in Microsoft Office. It also includes important features for map settings such as display details and expanded pushpins (MapPoint Homepage: http://www.microsoft.com/uk/mappoint/default.aspx).

[23] MPCluster is a Microsoft MapPoint add-in which identifies groups or clusters in an organisation's MapPoint data. This clustering technique draws a boundary shape around each cluster and/or marks each cluster's centre with a pushpin (MPCluster Homepage: http://www.mpcluster.com/index.shtml).

usefulness and practicality of the DMP-BI framework. Table 8.1 presented the key patterns identified in each case study. In addition, the cross-case evaluation has confirmed the validity of the activities and functions of the DMP-BI framework.

The chapter also provided a comparison between the DMP-BI framework and three rival methodologies including the CRISP-DM, IPA, and PPV models. This comparison resulted in a detailed discussion that evaluated the benefits which the DMP-BI framework has over the rival methodologies. This resulted in a further validation of the practicality and usefulness of the DMP-BI framework over existing methodologies in executing the direct marketing process more effectively.

This chapter clearly demonstrated the usefulness and practicality of the conceptual framework in three sectors namely retail, telecommunication, and higher education. The chapter also validated the concepts, practices and technologies used in the conceptual framework. This was further verified through a comparison between the developed framework and well-established direct marketing models within both the industry and academic literature.

# CHAPTER 9: CONCLUSIONS

## 9.1 Introduction

This thesis has explored the issues related to the direct marketing process and proposed a conceptual framework to address those issues. Owing to the direct marketing process's diverse concepts and practices, the study stretched across a wide variety of research fields such as business intelligence research, organisational process, and IS evaluation studies. However, the thesis's main focus was directed towards the effective execution of the direct marketing process in contemporary business and organisational contexts.

In this final chapter, the prospective theoretical and practical contributions are outlined, the research limitations are discussed, future research directions are proposed, and concluding remarks are presented.

## 9.2 Summary of Contributions

Having discussed a variety of issues related to the direct marketing process, this research can significantly contribute to the debates surrounding those issues, both theoretically and practically. The following subsections present the theoretical and practical contributions of this research.

### 9.2.1 Theoretical Contributions

This research focused on direct marketing studies in general and its process in particular. There are many terms used to refer to direct marketing in both academia and industry. As discussed in *Chapter 2*, the diversity of terms used to refer to direct marketing has created confusion among both academics and practitioners. This confusion resulted in two main problems in direct marketing studies. First, there is a lack of distinctive common ground on which debates and research on the direct marketing process can be done. Second, there is a lack of attempts at theorisations. Having considered this, the thesis attempted to deal with a variety of direct marketing process issues from theoretical perspectives, and to identify the significance of the concept of the direct marketing process in contemporary business and organisational contexts. *Chapter 2*, in particular, addresses the various concepts and practices in direct marketing studies, and serves as a conceptual foundation on which the discussions concerning direct marketing process were built.

Based on the conceptual work on direct marketing, the research scope was narrowed down to direct marketing process issues. As discussed in *Chapter 2*, the direct marketing process, being one of the most important factors for an effective direct marketing strategy, needs to be addressed critically. The main reason for this is that the direct marketing process is fuelled by a wide range of marketing concepts and business analytics principles. It is also concerned with "people", "technologies", and "activities", making it a complex process to execute. This research tackles these issues by proposing a conceptual framework which incorporates the appropriate marketing concepts and business analytics included in the direct marketing process. Specifically, *Chapter 3* investigated and discussed the theoretical foundations of the direct marketing process including the marketing database, business analytics, and information systems.

Based on those theoretical discussions, the researcher selected the most important concepts and technologies related to the direct marketing process and developed a conceptual framework to tackle this thesis's main research question: *"How can the direct marketing process be executed more effectively, and what are the stages, activities, and technologies needed to achieve that?"*

The developed DMP-BI framework is a detailed process model for executing and managing the direct marketing process's various tasks and functions. This framework stems from two perspectives. First, it attempts to overcome the limitations of existing direct marketing process models and data mining methodologies. Second, it proposed BI tools as an integrated system platform to overcome the lack of an information system in previous process models. Therefore, the DMP-BI framework is believed to enhance the execution of the direct marketing process as it focuses on closely integrating BI functions with direct marketing process activities. Moreover, the framework structures and organises direct marketing's various activities and functions in a systematic way to ensure a well-defined path for marketers/analysts to execute the process.

The DMP-BI framework is the main contribution of this thesis. It was developed and presented in detail in *Chapter 3*, and empirically investigated in *Chapters 5, 6, and 7*. The findings of this investigation were the basis for evidence to support the practicality and usefulness of the developed framework. The exploratory evaluation and discussion in *Chapter 8* illuminated some of the key benefits of the DMP-BI framework over rival methodologies. Along with several key concepts and process models submitted in *Chapter 3*, the researcher particularly proposed business intelligence as being the fundamental concept for making direct marketing process execution more effective. Business intelligence tools use specific practices for supporting and facilitating the dynamic and heterogeneous direct marketing tasks of marketers/analysts. BI tools are not unique as an information system platform that can be used for the direct marketing process; it is clear that many different information systems can be used for direct marketing in different research contexts. However, this research argues that BI tools group together all the required functions to support any given direct marketing process, whereas other information systems would only partially support the direct marketing process.

### 9.2.2 Practical Contributions

Whilst this thesis mainly aimed to develop a conceptual framework to tackle direct marketing process issues, it also provided organisations with practical implications for primarily two issues: marketers/analysts' execution of the direct marketing process, and organisations' direct marketing strategies.

The DMP-BI framework can benefit marketers/analysts in an organisation in terms of providing them with a standard procedure for executing the direct marketing process effectively.

As discussed in *Chapter 2*, direct marketing has become an essential part of many organisations seeking to enhance marketing campaign responses and higher returns on their investments. However, most of the research on direct marketing focuses on the technical aspects of data mining to improve analytical models' accuracy. Yet this research fails to consider the organisational and managerial issues related to the direct marketing process from an information systems perspective. The specific integration of business intelligence with the direct marketing process in the DMP-BI framework, which has been provided in *Chapter 3*, illustrated in *Chapters 5, 6, and 7*, and evaluated and discussed in *Chapter 8*, can be highly valuable for executing and implementing a direct marketing strategy within an organisational context. Indeed, the DMP-BI framework provides organisations with a clear and systematic guide for executing the direct marketing process. Previous attempts to develop direct marketing process frameworks were either very broad in terms of giving general guidelines for executing the direct marketing process, or lacked an information system to support the process, which could result in a highly uncertain, difficult and time-consuming process for organisations.

The developed framework includes a list of tasks for marketers and analysts to execute the direct marketing process more effectively. As described in *Chapter 3*, these tasks are extracted from widely accepted literature and are linked with the most common direct marketing process models and data mining methodologies. Although a number of previous models are proposed in the direct marketing literature, the framework presented in this work has identified the issues in existing models and attempted to resolve them. By focusing on the process from an information systems perspective, the direct marketing process's effectiveness was improved. Moreover, the identification of relevant information system functions to support the process tasks, which were emphasised in this research, led to clear direct marketing process stages with specific BI functions to support each stage.

The three case studies are also an important practical contribution in this research. The case studies can be used as examples or references for organisations performing consumer behaviour analysis or geographic analysis for direct marketing purposes. Furthermore, the three case studies could be used as examples to observe the DMP-BI framework application in retailing, telecommunication, and the higher education sector.

As discussed above, this research holds several practical implications. However, the most significant finding from this research with regard to an organisation's direct marketing strategy is that business intelligence and the direct marketing process can be well integrated. Through case study illustrations of the DMP-BI framework, this research has demonstrated the complementary link between direct marketing tasks and business intelligence functions, which can greatly enhance organisations' direct marketing strategies.

The DMP-BI framework provides a new way to execute and manage direct marketing. Many organisations nowadays use the direct marketing process without even recognising it. In fact, data is today kept by almost any organisation and can be used for direct marketing purposes. This, in effect, can involve the usage of the DMP-BI framework. Although the DMP-BI framework was developed for the direct marketing process, the researcher believes that it also reflects the marketing process in many other situations. Therefore, the developed framework can be used in any marketing situations where customer data and information systems with the relevant functions are involved.

## 9.3 Research Limitations

This research has several limitations which can be improved in future work. These limitations are primarily related to the DMP-BI framework, the BI tools used, and the research methodology adopted. Therefore, this section is divided into three main subsections: 1) contextual constraints, 2) BI concept vs BI technology, and 3) methodological limitations.

### 9.3.1 Contextual Constraints

The DMP-BI framework's main limitations are concerned with the target users. Indeed, the framework needs a basic level of perquisite knowledge with regards to direct marketing concepts and BI technology practices. Throughout this thesis, the researcher used marketers/analysts to refer to the prospective users of the developed framework. The reason why marketers and analysts were chosen as target users is that both possess perquisite knowledge to execute the direct marketing process using the DMP-BI framework. Marketers should be familiar with direct marketing concepts; hence the framework can provide support in terms of BI practices. Analysts, on the other hand, should be familiar with analytical techniques and could require support to deal with direct marketing tasks. In summary, the DMP-BI framework is intended for users who are either familiar with marketing concepts, or have perquisite knowledge in business analytics.

Moreover, the scope of the development of the DMP-BI framework was limited to understanding customers. As described in *Chapter 3*, the direct marketing process is composed of two fundamental activities: 1) understanding customers, and 2) interacting with customers. The latter activity was not considered when developing the framework due to the limited access to the organisations selected for this study.

### 9.3.2 BI Concept vs BI Technology

This research was restricted to the usage of two BI tools only. These tools were selected based on their availability. In fact, the unavailability of many BI tools for evaluation purposes has significantly constrained this research. As a result, this research has covered business intelligence as a concept instead of focusing on its particular tools. In fact, the main concept used throughout the thesis to build the DMP-BI framework was based on the description of business intelligence by (Howson 2008, p.2): *"Business Intelligence allows people at all levels of an organisation to access, interact with, and* *analyse data to manage the business, improve performance, discover opportunities, and operate efficiently."*

The BI tools that are available on the market seek to offer their customers this whole concept. However, not all BI tools provide the exact same functions and principles. This can be confirmed by key findings derived from the third case study (i.e. higher education case study). In fact, it has been found that the BI tool used to perform the third case study does not support geographic data by default but requires additional configurations. Moreover, geographic analytics are not supported in either of the BI tools used in this research. For example, clustering techniques, which enable marketers/analysts to investigate various patterns related to customers' locations, are not provided in either BI tool. However, other BI tools may well be providing geographic clustering techniques, but this could not be confirmed due to time constraints and the unavailability of evaluation versions. In short, the BI concept in general does not exclude geographic analysis and therefore existing BI tools should provide a comprehensive set of geographic functions to support the direct marketing process.

It is important to mention that the versions of the BI tools used for this research were available during the time of conducting this study, which is between 2008 and 2010. Newer versions may well have included new functions such as geographic analytics.

### 9.3.3 Methodological Limitations

The DMP-BI framework needed a rigorous research methodology to evaluate its usefulness and practicality. A case study method was applied using customers' datasets acquired from three organisations as discussed in *Chapter 4*. A qualitative approach was also adopted as this study is associated with human and organisational issues. This allowed the researcher to find the rich contextual data that this research required. However, there were several weaknesses in the adopted research methodology. First, qualitative research is commonly criticised for the subjective influence the researcher's interpretation might have on the study findings. To address this issue, this research adopted rigorous analytical strategies and a common case study reporting structure to avoid bias where possible.

Second, the case studies were performed with a greater emphasis upon executing the process directly from data, as opposed to users. This is intended to prevent conflict when prospective users with different knowledge levels on direct marketing and business intelligence are involved or execute the process themselves. This is done using a common reporting structure for the three case studies. The case studies reporting structure followed the guidelines provided by the well-established work of (Yin 2009) on case study research.

Third, this research used existing organisational data to illustrate the practices of the DMP-BI framework within organisational settings. Therefore, there was no data collection process where users were involved. This was due to many constraints including time to collect data from three different organisations, access limitations to the organisations' human resources, and uncertainty of the availability of BI technology and the users' expertise level within the organisations.

Finally, generalisation is also considered as a drawback since the multiple-case studies method does not ensure automatic generalisation of the DMP-BI framework. This research attempted to generalise the usage of the DMP-BI framework by employing case studies from three different industries. Indeed, this research has adopted a multiple-case design approach which is characterised by the logic of replication and that can significantly enhance the robustness of research. However, it can still be argued that the framework validity is limited to supermarket promotions, telecommunication, and higher education industries.

## 9.4 Future Research

This section addresses future research directions for investigating the direct marketing process. The researcher suggests evaluating the proposed framework using action research as a research methodology. Action research is a process where the researcher enters a real-world situation in order to enhance it and acquire more knowledge about it (Checkland, Holwell 2007). This requires access to an organisation which has business intelligence and practises direct marketing. Action research is particularly relevant due to its suitability for IS-related disciplines. This is due to the fact that the IS discipline is by nature an applied field (Baskerville, Wood-Harper 2001).

There are three key elements that are involved in any piece of research work: Framework of ideas (F), Methodology (M), and Area of concern (A) (Checkland, Holwell 2007). In this case, future studies could use the proposed framework as the "Framework of ideas", action research as the "Methodology", and the DMP-BI performance as the "Area of concern". The objective is not only to learn lessons about DMP-BI practicality using action research but also to evaluate the adequacy of the latter. The learning process in action research can lead to changes and modifications that improve the DMP-BI framework application within live organisational settings.

Greater access to organisations would also permit the integration of the second fundamental activity of the direct marketing process, i.e. interacting with customers. In fact, scholars could incorporates the “interacting with customers” activity within the DMP-BI framework and be able to evaluate it in a real-world context.

The DMP-BI framework has been evaluated within three industries. Greater generalisation of the framework could significantly increase its validity. Quantitative methods, such as survey, can enable a larger-scale study and increase the framework generalisation. In fact, the usage of BI technology and direct marketing practices are expected to grow in the future, hence more organisations can be targeted for larger-scale studies which could improve, verify, and add relevant components to the DMP-BI framework. More case studies could also further validate the usage of the DMP-BI framework. However, due to time constraints, it was not possible to perform further case studies in this research.

Although this research focused on providing a direct marketing process framework for marketers and analysts, the proposed framework could also be used to manage and execute direct marketing in teams of analysts, marketers, and managers, for instance. This implies that there is a need to expand the discussion at an individual level towards more collective levels. It is important to realise and explain the collective features of individual users and their direct marketing use, as real marketing processes incorporate dynamic and complex collaborations amongst diverse users. As discussed in *Chapter 2*, the high iteration between marketers and analysts can cause the time required for the overall cycle of collecting, analysing, and acting on enterprise data to be longer. Therefore, this has to be taken into consideration while investigating ways to expand the usage of the DMP-BI framework application.

# References

Adomavicius, G. & Tuzhilin, A. 2005, "Personalization technologies: a process-oriented perspective", *Commun.ACM,* vol. 48, no. 10, pp. 83-90.

Aitchison, A. 2009, *Beginning Spatial with SQL Server 2008,* Apress, United States of America.

Baskerville, R.L. & Wood-Harper, A.T. 2001, "A critical perspective on action research as a method for information systems research", *Journal of Information Technology,* vol. 11, no. 3, pp. 235.

Baumgartner, B. & Hrushka, H. 2003, "Allocation of catalogs to collective customers based on semiparametric response models", *European Journal of Operational Research,* vol. 3, no. 162, pp. 839-849.

Benbasat, I., Goldstein, D.K. & Mead, M. 1987, "The Case Research Strategy in Studies of Information Systems", *MIS Quarterly,* vol. 11, no. 3, pp. 369.

Berg, B.v.d. & Breur, T. 2007, "Merits of interactive decision tree building: Part 1", *Journal of Targeting, Measurement & Analysis for Marketing,* vol. 15, no. 3, pp. 137-145.

Berger, K.A. & Wallingford, H.P. 1997, "Developing Advertising and Promotion Strategies for Higher Education", *Journal of Marketing for Higher Education,* vol. 7, no. 4, pp. 61.

Berry, M.J.A. & Linoff, G.S. 2000, *Data Mining techniques: For Marketing, Sales, and Customer Relationship Management*, Wiley Publishing, Inc, United States of America.

Bodapati, A. & Gupta, S. 2004, "A direct approach to predicting discretized response in target marketing", *Journal of Direct Marketing,* vol. 20, no. 1, pp. 73-85.

Boland, R.J.J. 1991, "Information System Use as a Hermeneutic Process" in *Information Systems Research: Contemporary Approaches and Emergent Traditions* North-Holland, Amsterdam, pp. 439-464.

Bose, I. & Chen, X. 2009, "Quantitative models for direct marketing: A review from systems perspective", *European Journal of Operational Research,* vol. 195, no. 1, pp. 1-16.

Breur, T. 2007, "How to evaluate campaign response — The relative contribution of data mining models and marketing execution", *Journal of Targeting, Measurement & Analysis for Marketing,* vol. 15, no. 2, pp. 103-112.

Brunel University 2010, *, Code of Research Ethics* [Homepage of Brunel University], [Online]. Available: http://intranet.brunel.ac.uk/registry/minutes/researchethics/CoEv7.pdf.

Butler Group 2006, *Business intelligence: a strategic approach to extending and standardising the use of BI,* Butler Group, Hull.

Cavana, R., Delahye, B.L. & Sekaran, U. 2001, *Applied Business Research: Qualitative and Quantitative Methods,* Wiley, Milton, Australia.

Cavaye, A.L.M. 1996, *Case study research: a multi-faceted research approach for IS*, Blackwell Publishing Ltd.

Changchien, S.W., Lee, C. & Hsu, Y. 2004, "On-line personalized sales promotion in electronic commerce", *Expert Systems with Applications,* vol. 27, no. 1, pp. 35-52.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. 2000, "
CRISP-DM 1.0: Step-by-step data mining guide", CRISP-DM Consortium.

Checkland, P. & Holwell, S. 2007, "Action Research", *Information Systems Action Research*, vol. 13, pp. 3-17.

Chen, M., Chiu, A. & Chang, H. 2005, "Mining changes in customer behavior in retail marketing", *Expert Systems with Applications,* vol. 28, no. 4, pp. 773-781.

Cody, W.F., Kreulen, J.T., Krishna, V. & Spangler, W.S. 2002, "The integration of business intelligence and knowledge management", *IBM Systems Journal,* vol. 41, no. 4, pp. 697-713.

Creswell, J.W. 1998, *Qualitative inquiry and research design : choosing among five traditions,* Sage Publication.

Crone, S.F., Lessmann, S. & Stahlbock, R. 2006, "The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing", *European Journal of Operational Research,* vol. 173, no. 3, pp. 781-800.

Cui, G., Wong, M.L. & Lui, H. 2006, "Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming", *Management Science,* vol. 52, no. 4, pp. 597-612.

DMA 2008, *Economic Impact Analysis 2008: The Direct Marketing Industry*, Direct Marketing Association Research.

Dul, J. & Hak, T. 2008, *Case Study Methodology in Business Research,* First edn, Elsevier, London.

Evans, M., O'Malley, L. & Patterson, M. 2004, *Exploring Direct and Relationship Marketing,* Second edn, Thomson, Surrey.

Fayyad, U., Piatetsky-shapiro, G. & Smyth, P. 1996, "From data mining to knowledge discovery in databases", *AI Magazine,* vol. 17, pp. 37-54.

Ferreira, J.B., Vellasco, M., Pacheco, M.A. & Barbosa, C.H. 2004, "Data mining techniques on the evaluation of wireless churn", *Scientific Literature Digital Library and Search Engine.*

Flick, U. 2009, *An introduction to qualitative research,* 4th edn, Sage, London.

Forester, J. 1992, "Critical Ethnography: On Field Work in an Habermasian Way" in *Critical Management Studies* Sage Publications, London.

Gersten, W., Wirth, R. & Arndt, D. 2000, "Predictive modeling in automotive direct marketing: tools, experiences and open issues", *KDD '00: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining,* ACM, New York, NY, USA, pp. 398.

Gilbert, D.C. & Jackaria, N. 2002, "The efficacy of sales promotions in UK supermarkets: a consumer view", *International Journal of Retail & Distribution Management,* vol. 30, no. 6.

Golfarelli, M., Rizzi, S. & Cella, I. 2004, "Beyond data warehousing: what's next in business intelligence?", *DOLAP '04: Proceedings of the 7th ACM international workshop on Data warehousing and OLAP*ACM, New York, NY, USA, pp. 1.

Grint, K. & Woolgar, S. 1997, *The Machine at Work,* Polity Press and Blackwell Publishers, USA.

Guba, E.G. & Lincoln, Y.S. 1994, "Competing paradigms in qualitative research" in *Handbook of Qualitative Research* SAGE, , pp. 105-117.

Ha, K., Cho, S. & MacLachlan, D. 2005, "Response models based on bagging neural networks", *Journal of Interactive Marketing,* vol. 19, no. 1, pp. 17-30.

Han, J. & Kamber, M. 2006, *Data Mining: Concepts and Techniques,* 2nd edition edn, Morgan Kaufmann, San Francisco.

Harinath, S. & Quinn, S.R. 2006, *Analysis Services 2005 with MDX,* Wiley Publishing, Inc, Indianapolis, Indiana.

Heilman, C.M., Kaefer, F. & Ramenofsky, S.D. 2003, "Determining the appropriate amount of data for classifying consumers for direct marketing purposes", *Journal of Interactive Marketing,* vol. 17, no. 3, pp. 5-28.

Hirschheim, R., Klein, H.K. & Lyytinen, K. 1995, *Information Systems Development and Data Modeling,* Cambridge University Press.

Howcroft, D. & Trauth, E.M. 2005, *Handbook of Critical Information Systems Research: Theory and Application,* Edward Elgar Publishing, Cheltenham.

Howson, C. 2008, *Successful Business Intelligence: Secrets to Making BI a Killer App,* McGraw-Hill, London.

Kaefer, F., Heilman, C.M. & Ramenofsky, S.D. 2005, "A neural network application to consumer classification to improve the timing of direct marketing activities", *Computers & Operations Research,* vol. 32, no. 10, pp. 2595-2615.

Khazanchi, D. & Munkvold, B.E. 2003, "On the Rhetoric and Relevance of IS Research Paradigms: A Conceptual Framework and Some Propositions", *Hawaii International Conference on System Sciences,* vol. 8, pp. 252b.

Kim, Y. 2009, "Boosting and measuring the performance of ensembles for a successful database marketing", *Expert Systems with Applications,* vol. 36, no. 2, Part 1, pp. 2161-2176.

Kim, Y. 2006, "Toward a successful CRM: variable selection, sampling, and ensemble", *Decision Support Systems,* vol. 41, no. 2, pp. 542-553.

Kitts, B. & Hetherington, K. 2005, "Price Optimization in Grocery Stores with Cannibalistic Product Interactions", *Proceedings of the First International Workshop on Data Mining Case Studies*2005 IEEE International Conference on Data Mining, , pp. 74.

Klein, H.K. & Myers, M.D. 1999, "A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems", *MIS Quarterly,* vol. 23, no. 1, pp. 67-93.

Kohavi, R., Rothleder, N.J. & Simoudis, E. 2002, "Emerging trends in business analytics", *Communications of the ACM,* vol. 45, no. 8, pp. 45-48.

Kolter, P. & Armstrong, G. 2008, *Principles of Marketing,* 12th edn, Pearson Prentice Hall, New Jersey.

Kotler, P. 1991, *Marketing Management,* Prentice Hall, USA.

Kumar, V. & Werner, J.R. 2006, *Customer Relationship Management: A Database Approach,* John Wiley & Sons, Inc, USA.

Lim, J., Currim, I.S. & Andrews, R.L. 2005, "Consumer heterogeneity in the longer-term effects of price promotions", *International Journal of Research in Marketing,* vol. 22, no. 4, pp. 441-457.

Lin, C. & Hong, C. 2009, "Development of a marketing information system for supporting sales in a Tea-beverage market", *Expert Systems with Applications,* vol. 36, no. 3, Part 1, pp. 5393-5401.

Linoff, G.S. 2008, *Data Analysis Using SQL and Excel,* Wiley Publishing, Inc., Indianapolis, Indiana.

Longley, P.A., Goodchild, M.F., Maguire, D.J. & Rhind, D.W. 2005, *Geographical Information Systems and Science,* Second edn, John Wiley, Chichester.

Martínez-López, F.J. & Casillas, J. 2008, "Marketing Intelligent Systems for consumer behaviour modelling by a descriptive induction approach based on Genetic Fuzzy Systems", *Industrial Marketing Management,* vol. In Press, Corrected Proof.

Milne, G.R. & Gordan, M.E. 1993, "Direct Mail Privacy Efficiency Trade-offs Within an Implied Social Contract Framework", *Journal of Public Policy and Marketing*, no. 12, pp. 206-215.

Mingers, J. 2001, "Combining IS Research Methods: Towards a Pluralist Methodology", *Information Systems Research,* vol. 12, no. 3, pp. 240-259.

Mitra, S., Pal, S.K. & Mitra, P. 2002, "Data mining in soft computing framework: a survey", *Neural Networks, IEEE Transactions on,* vol. 13, no. 1, pp. 3-14.

Mozer, M.C., Wolniewicz, R., Grimes, D.B., Johnson, E. & Kaushansky, H. 2000, "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry", *Neural Networks, IEEE Transactions on,* vol. 11, no. 3, pp. 690-696.

Murthi, B.P.S. & Sarkar, S. 2003, "The Role of the Management Sciences in Research on Personalisation", *Management Science,* vol. 49, no. 10, pp. 1344-1362.

Myers, M.D. 2009, *Qualitative Research in Business & Management,* SAGE Publications, London.

Negash, S. 2004, "Business Intelligence", *Communications of the Association for Information Systems*, vol. 13, pp. 177-195.

Ngai, E.W.T., Xiu, L. & Chau, D.C.K. 2009, "Application of data mining techniques in customer relationship management: A literature review and classification", *Expert Systems with Applications,* vol. 36, no. 2, Part 2, pp. 2592-2602.

Ngwenyama, O.K. 1992, "The Critical Social Theory Approach to Information Systems: Problems and Challenges" in *Information Systems Research: Contemporary Approaches and Emergent Traditions* North-Holland, Amsterdam, pp. 267-280.

Nisbet, R., Elder, J. & Miner, G. 2009, *Handbook of Statistical Analysis and Data Mining Applications,* Elsevier Inc, USA.

Oates, B.J. 2006, *Researching Information Systems and Computing,* Sage Publications, London.

Orlikowski, W.J. & Baroudi, J.J. 1991, "Studying Information Technology in Organizations: Research Approaches and Assumptions", *Information Systems Research,* vol. 2, no. 1, pp. 1-28.

Ortiz, S.,Jr. 2002, "Is business intelligence a smart move?", *Computer,* vol. 35, no. 7, pp. 11-14.

Ou, C., Liu, C., Huang, J. & Zhong, N. 2003, "On Data Mining for Direct Marketing", *Lecture Notes in Computer Science,* vol. 2639, pp. 575.

Ou, L. & Peng, H. 2006, "Knowledge and Process Based Decision Support in Business Intelligence System", *Computer and Computational Sciences, 2006. IMSCCS '06. First International Multi-Symposiums on*, pp. 780.

Peltier, J.W., Schibrowsky, J.A. & Schultz, D.E. 2003, "Interactive integrated marketing communication: combining the power of IMC, the new media and database marketing", *International Journal of Advertising*, vol. 22, pp. 93-115.

Rao, V.R. & Steckel Joel H. 1998, *Analysis for Strategic Marketing,* Addison-Wesley, Reading, MA.

Reutterer, T., Mild, A., Natter, M. & Taudes, A. 2006, "A dynamic segmentation approach for targeting and customizing direct marketing campaigns", *Journal of Interactive Marketing,* vol. 20, no. 3-4, pp. 43-57.

Ryals, L. & Knox, S. 2001, "Cross-functional issues in the implementation of relationship marketing through customer relationship management", *European Management Journal,* vol. 19, no. 5, pp. 534-542.

Sargeant, A. & Douglas, C.W. 2001, *Direct and Interactive Marketing,* Oxford University Press.

SAS Institute 1998, *Data mining and the case for sampling*, SAS Inc.

Shaw, M.J., Subramaniam, C., Tan, G.W. & Welge, M.E. 2001, "Knowledge management and data mining for marketing", *Decision Support Systems,* vol. 31, no. 1, pp. 127-137.

SITS Support 2005, *An Introduction to SITS,* Brunel University Intranet Website.

Stone, B. & Jacobs, R. 2008, *Successful Direct Marketing Methods,* Eighth Edition edn, McGraw-Hill, United State of America.

Tabachnick, B.G. & Fidell, L.S. 2007, *Using Multivariate Statistics,* Fifth edn, Pearson Education, Boston.

Tapp, A. 2008, *Principles of Direct and Database Marketing,* Fourth edn, Pearson Education Limited, England.

Tapp, A., Hicks, K. & Stone, M. 2004, "Direct and database marketing and customer relationship management in recruiting students for higher education", *International Journal of Nonprofit and Voluntary Sector Marketing,* vol. 9, no. 4, pp. 335-345.

Tettamanzi, A., Carlesi, M., Pannese, L. & Santalmasi, M. 2007, "Business Intelligence for Strategic Marketing: Predictive Modelling of Customer Behaviour Using Fuzzy Logic and Evolutionary Algorithms", *Applications of Evolutionary Computing*, pp. 233-240.

The Fuqua School of Business 2002, *CELL2CELL: The Churn Game*, Available: http://www.fuqua.duke.edu/centers/ccrm/datasets/cell/ [Last Accessed: Oct/2010]

Tolson, D. 2007, "Promotions at Top 4 increase to help drive business – Up to a third of all sales now going through on promotion", *Nielson Company*.

UCAS 2010, *Annual Datasets,* UCAS Website [Last Accessed: Sept/2010].

Van den Poel, D. & Buckinx, W. 2005, "Predicting online-purchasing behaviour", *European Journal of Operational Research*, vol. 2, no. 16, pp. 557-575.

Vesanen, J. & Raulas, M. 2006, "Building bridges for personalization: A process model for marketing", *Journal of Interactive Marketing,* vol. 20, no. 1, pp. 5-20.

Vesset, D. 2010, *Worldwide Business Intelligence Tools 2009 Vendor Shares*, IDC.

Vesset, D. & McDonough, B. July 2006, "Worldwide Business Intelligence Tools 2005 Vendor Shares", vol. 1, IDC.

Walsham, G. 1993, *Interpreting Information Systems in Organizations,* 1st edn, John Wiley \& Sons, Inc, New York, NY, USA.

Wang, H. & Hong, W. 2006, "Managing customer profitability in a competitive market by continuous data mining", *Industrial Marketing Management,* vol. 35, no. 6, pp. 715-723.

Witten, I. & Frank, E. 2005, *Data Mining: Practical Machine Learning Tools and Techniques,* 2nd edn, Morgan Kaufmann, San Francisco.

Yin, R.K. 2009, *Case Study Research: Design and Methods,* Fourth edn, Sage Publications, London.

Yin, R.K. 1994, *Case Study Research, Design and Methods,* 2nd edn, SAGE Publications, California.

Yin, R.K. 2003a, *Applications of Case Study Research,* 2nd edn, Sage Publications, Inc.

Yin, R.K. 2003b, *Case Study Research,* 3rd edn, Sage Publications, inc.

# Appendices

## Appendix 5: Supermarket Dataset Variables

### 5.1 Appendix: Meat Products and their Description

| Variables | Meaning | Variables | Meaning |
|---|---|---|---|
| **B Sales** | **Beef Products** | | |
| B1 SALES | Premium Roasting Beef | L3 SALES | Healthy Fry/Grilling Lamb |
| B2 SALES | Premium Fry/Grilling Beef | L4 SALES | Organic Mince Lamb |
| B3 SALES | Premium Diced Beef | L5 SALES | Healthy Mince Lamb |
| B4 SALES | Premium Mince Beef | L6 SALES | Organic Diced Lamb |
| B5 SALES | SR/Organic Roasting Beef | L7 SALES | Healthy Diced Lamb |
| B6 SALES | Specially Reared/Organic Fry/Grill Beef | L8 SALES | Premium Roasting Lamb |
| B7 SALES | Healthy Diced Beef | L9 SALES | Std Roasting Lamb |
| B8 SALES | Healthy Beef Mince | L10 SALES | Std Mince Lamb |
| B9 SALES | Healthy Fry/Grill Beef | L11 SALES | Premium Mince Lamb |
| B10 SALES | Organic Beef Mince | L12 SALES | Std Fry/Grilling Lamb |
| B11 SALES | Specially reared/ Organic Diced Beef | L13 SALES | Premium Fry/Grilling Lamb |
| B12 SALES | Std Other Fry/Grill Beef | **P Sales** | **Pork Products** |
| B13 SALES | Std Sirloin Steak Beef | P1 SALES | Premium Roasting Pork |
| B14 SALES | Std Rump Steak Beef | P2 SALES | Premium Fry/Grill Pork |
| B15 SALES | Std Fillet Steak Beef | P3 SALES | Std Roasting Pork |
| B16 SALES | Value Fry/Grill Beef | P4 SALES | Specially Reared/ Organic Roasting Pork |
| B17 SALES | Std Diced Beef | P5 SALES | Value Roasting Pork |
| B18 SALES | Std Beef Mince | P6 SALES | Std Fry/Grill Pork |
| B19 SALES | Std Roasting Beef | P7 SALES | Specially Reared/Organic Fry/Grill Pork |
| B20 SALES | Value Roasting Beef | P8 SALES | Value Fry/Grill Pork |
| **L Sales** | **Lamb Products** | P9 SALES | Healthy Fry/Grilling Pork |
| L1 SALES | SR/Organic Roasting Lamb | P10 SALES | Std Mince Pork |
| L2 SALES | SR/Organic Fry/Grilling Lamb | P11 SALES | SR/Organic Mince Pork |
| | | P12 SALES | Healthy Diced Pork |

## 5.2 Appendix: Data Key Fact Summary for the Supermarket Dataset

| Variables | Meaning | Mean.Sales | Mean.TPR | Mean.MULTI |
|---|---|---|---|---|
| B1 SALES | Premium Roasting Beef | 361.99 | 0.44 | 0 |
| B2 SALES | Premium Fry/Grilling Beef | 849.68 | 0.77 | 0 |
| B3 SALES | Premium Diced Beef | 119.87 | 0.06 | 0 |
| B4 SALES | Premium Mince Beef | 1237.01 | 0.21 | 0.1 |
| B5 SALES | SR/Organic Roasting Beef | 210.48 | 0.29 | 0 |
| B6 SALES | Specially Reared/Organic Fry/Grill Beef | 445.24 | 0.43 | 0 |
| B7 SALES | Healthy Diced Beef | 1600.9 | 0 | 0 |
| B8 SALES | Healthy Beef Mince | 1460.29 | 0 | 0.22 |
| B9 SALES | Healthy Fry/Grill Beef | 43.07 | 0.01 | 0 |
| B10 SALES | Organic Beef Mince | 429.97 | 0 | 0 |
| B11 SALES | Specially reared/ Organic Diced Beef | 185.61 | 0 | 0 |
| B12 SALES | Std Other Fry/Grill Beef | 1174.42 | 0.33 | 0.06 |
| B13 SALES | Std Sirloin Steak Beef | 1251.62 | 0.27 | 0.01 |
| B14 SALES | Std Rump Steak Beef | 1239.88 | 0.24 | 0 |
| B15 SALES | Std Fillet Steak Beef | 420.76 | 0 | 0 |
| B16 SALES | Value Fry/Grill Beef | 702.82 | 0 | 0 |
| B17 SALES | Std Diced Beef | 2098.82 | 0.07 | 0 |
| B18 SALES | Std Beef Mince | 7818.21 | 0.03 | 0.88 |
| B19 SALES | Std Roasting Beef | 2224.54 | 0.33 | 0 |
| B20 SALES | Value Roasting Beef | 169.17 | 0 | 0 |
| L1 SALES | SR/Organic Roasting Lamb | 125.29 | 0.21 | 0 |
| L2 SALES | SR/Organic Fry/Grilling Lamb | 324.62 | 0.34 | 0 |
| L3 SALES | Healthy Fry/Grilling Lamb | 322.98 | 0 | 0 |
| L4 SALES | Organic Mince Lamb | 158.17 | 0.02 | 0 |
| L5 SALES | Healthy Mince Lamb | 85.42 | 0 | 0 |
| L6 SALES | Organic Diced Lamb | 28.31 | 0 | 0 |
| L7 SALES | Healthy Diced Lamb | 1626.24 | 0 | 0 |
| L8 SALES | Premium Roasting Lamb | 137.3 | 0 | 0 |
| L9 SALES | Std Roasting Lamb | 561.85 | 0.74 | 0 |
| L10 SALES | Std Mince Lamb | 712.05 | 0 | 0.21 |
| L11 SALES | Premium Mince Lamb | 9.53 | 0 | 0 |
| L12 SALES | Std Fry/Grilling Lamb | 1809.54 | 0 | 0 |
| L13 SALES | Premium Fry/Grilling Lamb | 14.74 | 0 | 0 |
| P1 SALES | Premium Roasting Pork | 879.97 | 0.26 | 0 |
| P2 SALES | Premium Fry/Grill Pork | 74.84 | 0.33 | 0 |
| P3 SALES | Std Roasting Pork | 732.65 | 0.57 | 0.1 |
| P4 SALES | Specially Reared/ Organic Roasting Pork | 276.52 | 0.43 | 0 |
| P5 SALES | Value Roasting Pork | 113.22 | 0 | 0 |
| P6 SALES | Std Fry/Grill Pork | 1883.72 | 0.42 | 0.26 |
| P7 SALES | Specially Reared/Organic Fry/Grill Pork | 144.8 | 0.19 | 0 |
| P8 SALES | Value Fry/Grill Pork | 229.17 | 0 | 0 |
| P9 SALES | Healthy Fry/Grilling Pork | 201.23 | 0 | 0 |
| P10 SALES | Std Mince Pork | 671.61 | 0.08 | 0 |
| P11 SALES | SR/Organic Mince Pork | 18.27 | 0 | 0 |
| P12 SALES | Healthy Diced Pork | 1449.75 | 0 | 0 |
| **B SALES** | **Beef products** | **1202.218** | **0.174** | **0.0635** |
| **L SALES** | **Lam Products** | **455.08** | **0.100769231** | **0.016153846** |
| **P SALES** | **Pork Sales** | **526.8891** | **0.19** | **0.03** |

# Appendix 6: Variables Summary

| Position | Variable Name | Variable Descriptiion | N | Minimum | Maximum | Mean | Standard Deviati |
|---|---|---|---|---|---|---|---|
| 1 | revenue | Mean monthly revenue | 70831 | -6.1675 | 1223.38 | 58.852803 | 44.24358324 |
| 2 | mou | Mean monthly minutes of use | 70831 | 0 | 7667.75 | 525.72839 | 530.1342588 |
| 3 | recchrge | Mean total recurring charge | 70831 | -11.29 | 399.99 | 46.876304 | 23.91509477 |
| 4 | directas | Mean number of director assisted calls | 70831 | 0 | 159.39 | 0.8940274 | 2.19770883 |
| 5 | overage | Mean overage minutes of use | 70831 | 0 | 4320.75 | 40.095361 | 96.34710282 |
| 6 | roam | Mean number of roaming calls | 70831 | 0 | 1112.4475 | 1.2210712 | 9.08108861 |
| 7 | changem | % Change in minutes of use | 70545 | -3875 | 5192.25 | -10.84646 | 255.3143148 |
| 8 | changer | % Change in revenues | 70545 | -1107.74 | 2483.4825 | -1.205634 | 38.77029226 |
| 9 | dropvce | Mean number of dropped voice calls | 71047 | 0 | 221.66667 | 6.0099676 | 9.006124595 |
| 10 | blckvce | Mean number of blocked voice calls | 71047 | 0 | 384.33333 | 4.067833 | 10.67078555 |
| 11 | unansvce | Mean number of unanswered voice calls | 71047 | 0 | 848.66667 | 28.355903 | 38.90424823 |
| 12 | custcare | Mean number of customer care calls | 71047 | 0 | 365.66667 | 1.8659174 | 5.160761938 |
| 13 | threeway | Mean number of threeway calls | 71047 | 0 | 66 | 0.3001675 | 1.161560303 |
| 14 | mourec | Mean unrounded mou received voice calls | 71047 | 0 | 3287.25 | 114.93533 | 166.3057292 |
| 15 | outcalls | Mean number of outbound voice calls | 71047 | 0 | 644.33333 | 25.396526 | 35.14751212 |
| 16 | incalls | Mean number of inbound voice calls | 71047 | 0 | 519.33333 | 8.1767985 | 16.51905923 |
| 17 | peakvce | Mean number of in and out peak voice calls | 71047 | 0 | 2090.6667 | 90.580946 | 104.9148661 |
| 18 | opeakvce | Mean number of in and out off-peak voice calls | 71047 | 0 | 1572.6667 | 67.818418 | 93.32899044 |
| 19 | dropblk | Mean number of dropped or blocked calls | 71047 | 0 | 489.66667 | 10.149744 | 15.46058152 |
| 20 | callfwdv | Mean number of call forwarding calls | 71047 | 0 | 81.333333 | 0.0118372 | 0.562186873 |
| 21 | callwait | Mean number of call waiting calls | 71047 | 0 | 212.66667 | 1.8530339 | 5.556201766 |
| 22 | churn | Churn between 31-60 days after obs_date | 71047 | 0 | 1 | 0.2900756 | 0.453800218 |
| 23 | months | Months in Service | 71047 | 6 | 61 | 18.750827 | 9.787568466 |
| 24 | uniqsubs | Number of Uniq Subs | 71047 | 1 | 196 | 1.5295509 | 1.131774041 |
| 25 | actvsubs | Number of Active Subs | 71047 | 0 | 53 | 1.3516545 | 0.66004932 |
| 26 | csa | Communications Service Area | 71047 | Character string variable | | | |
| 27 | phones | # Handsets Issued | 71046 | 1 | 28 | 1.808617 | 1.33612025 |
| 28 | models | # Models Issued | 71046 | 1 | 16 | 1.561791 | 0.908280483 |
| 29 | eqpdays | Number of days of the current equipment | 71046 | -5 | 1823 | 380.26563 | 254.2946923 |
| 30 | customer | Customer ID | 71047 | 1000001 | 1099999 | 1050487.5 | 29199.11481 |
| 31 | age1 | Age of first HH member | 69803 | 0 | 99 | 31.375113 | 22.08219498 |
| 32 | age2 | Age of second HH member | 69803 | 0 | 99 | 21.157715 | 23.91758552 |
| 33 | children | Presence of children in HH | 71047 | 0 | 1 | 0.2423888 | 0.428531301 |
| 34 | credita | Highest credit rating - a | 71047 | 0 | 1 | 0.1676637 | 0.373569968 |
| 35 | creditaa | High credit rating - aa | 71047 | 0 | 1 | 0.3708812 | 0.48304413 |
| 36 | creditb | Good credit rating - b | 71047 | 0 | 1 | 0.1645249 | 0.370753809 |
| 37 | creditc | Medium credit rating - c | 71047 | 0 | 1 | 0.1044379 | 0.305829932 |
| 38 | creditde | Low credit rating - de | 71047 | 0 | 1 | 0.1284783 | 0.334624006 |
| 39 | creditgy | Very low credit rating - gy | 71047 | 0 | 1 | 0.022647 | 0.148776362 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 40 | creditz | Lowest credit rating - z | 71047 | 0 | 1 | 0.041367 | 0.19913893 |
| 41 | prizmrur | Prizm code is rural | 71047 | 0 | 1 | 0.047743 | 0.213223557 |
| 42 | prizmub | Prizm code is suburban | 71047 | 0 | 1 | 0.3211114 | 0.466906767 |
| 43 | prizmtwn | Prizm code is town | 71047 | 0 | 1 | 0.1484229 | 0.355521173 |
| 44 | refurb | Handset is refurbished | 71047 | 0 | 1 | 0.1396118 | 0.346586267 |
| 45 | webcap | Hanset is web capable | 71047 | 0 | 1 | 0.9028108 | 0.296217289 |
| 46 | truck | Subscriber owns a truck | 71047 | 0 | 1 | 0.1872141 | 0.390086043 |
| 47 | rv | Subscriber owns a recreational vehicle | 71047 | 0 | 1 | 0.0811998 | 0.273143583 |
| 48 | occprof | Occupation - professional | 71047 | 0 | 1 | 0.173899 | 0.379025249 |
| 49 | occcler | Occupation - clerical | 71047 | 0 | 1 | 0.0200571 | 0.140196765 |
| 50 | occcrft | Occupation - crafts | 71047 | 0 | 1 | 0.0296423 | 0.16959978 |
| 51 | occstud | Occupation - student | 71047 | 0 | 1 | 0.0075725 | 0.086690344 |
| 52 | occhmkr | Occupation - homemaker | 71047 | 0 | 1 | 0.0031528 | 0.056061986 |
| 53 | occret | Occupation - retired | 71047 | 0 | 1 | 0.0145115 | 0.119587364 |
| 54 | occself | Occupation - self-employed | 71047 | 0 | 1 | 0.0178333 | 0.132346086 |
| 55 | ownrent | Home ownership is missing | 71047 | 0 | 1 | 0.3319211 | 0.470906161 |
| 56 | marryun | Marital status unknown | 71047 | 0 | 1 | 0.3848157 | 0.486555142 |
| 57 | marryyes | Married | 71047 | 0 | 1 | 0.3653778 | 0.481539344 |
| 58 | marryno | Not Married | 71047 | 0 | 1 | 0.2498065 | 0.432903954 |
| 59 | mailord | Buys via mail order | 71047 | 0 | 1 | 0.3619717 | 0.480574059 |
| 60 | mailres | Responds to mail offers | 71047 | 0 | 1 | 0.377201 | 0.484689296 |
| 61 | mailflag | Has chosen not to be solicited by mail | 71047 | 0 | 1 | 0.014413 | 0.119186659 |
| 62 | travel | Has traveled to non-US country | 71047 | 0 | 1 | 0.0574831 | 0.232764974 |
| 63 | pcown | Owns a personal computer | 71047 | 0 | 1 | 0.1854125 | 0.388634551 |
| 64 | creditcd | Possesses a credit card | 71047 | 0 | 1 | 0.6764255 | 0.467843068 |
| 65 | retcalls | Number of calls previously made to retention team | 71047 | 0 | 4 | 0.0370037 | 0.20582259 |
| 66 | retaccpt | Number of previous retention offers accepted | 71047 | 0 | 4 | 0.0179177 | 0.14148458 |
| 67 | newcelly | Known to be a new cell phone user | 71047 | 0 | 1 | 0.1929427 | 0.394611209 |
| 68 | newcelln | Known not to be a new cell phone user | 71047 | 0 | 1 | 0.1387814 | 0.345720673 |
| 69 | refer | Number of referrals made by subscriber | 71047 | 0 | 35 | 0.0508537 | 0.290443694 |
| 70 | incmiss | Income data is missing | 71047 | 0 | 1 | 0.2498346 | 0.432920222 |
| 71 | income | Income (0=>missing) | 71047 | 0 | 9 | 4.3342295 | 3.137063069 |
| 72 | mcycle | Owns a motorcycle | 71047 | 0 | 1 | 0.0134559 | 0.115217218 |
| 73 | creditad | Number of adjustments made to customer credit rating (up or down) | 71047 | 0 | 25 | 0.053162 | 0.374988053 |
| 74 | setprcm | Missing data on handset price | 71047 | 0 | 1 | 0.5665123 | 0.495559854 |
| 75 | setprc | Handset price (0=>missing) | 71047 | 0 | 499.98999 | 35.79858 | 57.04096076 |
| 76 | retcall | Customer has made made call to retention team | 71047 | 0 | 1 | 0.0340338 | 0.181317322 |
| 77 | calibrat | Calibration sample = 1; Validation sample = 0; | 71047 | 0 | 1 | 0.5630076 | 0.496017648 |
| 78 | churndep | Churn (=missing for validation sample) | 40000 | 0 | 1 | 0.5 | 0.50000625 |