APPLICATION OF CUSTOMER RELATIONSHIP MANAGEMENT AND DATA MINING IN PREDICTING CUSTOMER PURCHASE BEHAVIOR IN MEDIUM AND LARGE SUPERMARKETS IN KENYA

CATHERINE WANJIRU NYOIKE

D65/ 80747/2012

A RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF THE DEGREE OF MASTER OF SCIENCE MARKETING, MARKETING RESEARCH, SCHOOL OF BUSINESS, UNIVERSITY OF NAIROBI

JULY 2015
DECLARATION

I declare that this project is my original work and has not been presented for any degree award in any other university.

Signature:……………………………………Date:………………………………………
Catherine Wanjiru Nyoike
D65/80747/2012

This research project has been submitted for examination with my approval as the university supervisor

Signature:……………………………………Date:………………………………………
Ms. Catherine Ngahu
Department of Business Administration
School of Business
University of Nairobi
ACKNOWLEDGEMENTS

I would like to acknowledge friends and family that have supported me through the MSc journey and everyone who has put in his/her effort in guiding and being there for me during preparation of my research project.

Special thanks too to the teaching staff of University of Nairobi, MSc Market Research Program and my supervisor Ms. Catherine Ngahu for her guidance throughout the project.

And finally to Mum for believing in me and supporting me both financially and emotionally throughout my academic journey.
DEDICATION

This research project is dedicated to all young people in pursuit of knowledge and a dream life. Growth is a personal journey only you can champion; pursue it in knowledge.

I also dedicate the project to my keen and future generations to pursue nothing but greatness in their lives no matter their circumstances.

Finally I wish to dedicate this to my mother who has tirelessly advocated for a good education throughout my life and her tireless contribution both financially and emotionally. God bless you immensely.
# TABLE OF CONTENTS

DECLARATION.......................................................................................................................... II

ACKNOWLEDGEMENTS ............................................................................................................. III

DEDICATION................................................................................................................................ IV

LIST OF TABLES ............................................................................................................................ VIII

LIST OF FIGURES .......................................................................................................................... IX

ABBREVIATIONS AND ACCRONMYS ...................................................................................... X

ABSTRACT .................................................................................................................................. XI

CHAPTER ONE: INTRODUCTION ............................................................................................... 1

1.1 Background to the Study ........................................................................................................ 1
1.1.1 Customer Relationship Management ............................................................................. 1
1.1.2 The Concept of Data Mining ......................................................................................... 2
1.1.3 Consumer Purchase Behavior ....................................................................................... 3
1.1.4 The Kenyan Retail Sector ............................................................................................... 4

1.2 Research Problem .................................................................................................................. 6

1.3 Objectives of the Study ......................................................................................................... 7

1.4 Value of Study ....................................................................................................................... 7

CHAPTER TWO: LITERATURE REVIEW ................................................................................... 8

2.1 Introduction ............................................................................................................................ 8

2.2 Theoretical Foundation of the Study .................................................................................... 8
2.2.1 Theory of Buyer Behavior ............................................................................................. 8
2.2.2 CRM Theory .................................................................................................................. 9

2.3 Customer Relationship Management ................................................................................... 10

2.4 The Relationship between CRM and Analytical Data Mining ........................................... 11

2.5 Data Mining Techniques ..................................................................................................... 12

2.6 Benefits of Data Mining ....................................................................................................... 13
2.7 Chapter Summary .............................................................................................................. 14

CHAPTER THREE: RESEARCH METHODOLOGY ......................................................... 15

3.1 Introduction .................................................................................................................. 15
3.2 Research Design ......................................................................................................... 15
3.3 Population .................................................................................................................... 16
3.4 Data Collection ........................................................................................................... 16
3.5 Data Analysis .............................................................................................................. 17
3.6 Chapter Summary ....................................................................................................... 17

CHAPTER FOUR: DATA ANALYSIS, RESULTS & DISCUSSION ......................... 18

4.1 Survey Findings .......................................................................................................... 18
4.2 Respondent Profile ..................................................................................................... 18
4.3 Data Mining Using Excel Software ........................................................................... 19
4.4 Data Types Captured in Retail .................................................................................. 20
4.5 Current State of Supermarkets on Data Mining ...................................................... 20
4.6 Performance of Data Capture Methods Used in Supermarkets ......................... 21
4.7 Data Mining Usage Across Departments .................................................................. 22
   4.7.1 CRM Usage ......................................................................................................... 25
4.8 Gap Analysis on Current Data Mining Capabilities ............................................... 25
4.9 Broadening Data Mining .......................................................................................... 26
4.10 Opportunities for Data Mining ................................................................................ 27
4.11 Discussion .................................................................................................................. 28
4.12 Chapter Summary ..................................................................................................... 29

CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS .......... 30

5.0 Introduction .................................................................................................................. 30
5.1 Summary of Findings ................................................................................................. 30
LIST OF TABLES

Table 1:  Respondent Profile................................................................. 18
Table 2:  Current Supermarket Status on Data Mining................................. 21
Table 3:  Performance of Data Capture Methods used ............................... 22
Table 4:  Departments that use Data Mining ........................................... 23
Table 5:  Departmental Application of Data Mining .................................... 24
Table 6:  CRM Rating .............................................................................. 25
LIST OF FIGURES

Figure 1: Uses of Excel Software ................................................................. 19
Figure 2: Variables Captured in the Data Mining System ........................................ 20
Figure 3: Gap Analysis on Application of Data Mining in Supermarkets......................... 26
Figure 4: Departments that Should Adopt Data Mining................................................. 26
Figure 5: Data that can be captured in Future............................................................ 27
ABBREVIATIONS AND ACCRONYMYS

POS    Point of Sale
SAS    Statistical Analysis System
CRM    Customer Relationship Management
RPS    Routine Problem Solving
SPSS   Statistical Package for Social Science
LPS    Limited Problem Solving
EPS    Extended Problem Solving
RPS    Routine Problem Solving
POS    Point of Sale
RMA    Return Merchandise Authorization
SQL    Structured Query Language
AX     Microsoft Axapta
SAS    Statistical Analysis System
GDP    Gross Domestic Product
ROI    Return on Investment
KPMG   Klynveld Peat Mackwick Goerdeler
The study topic was inspired by the growth in research dynamics for futurist organizations. Conventional research has concentrated on traditional methods that apply primary research and secondary research. One of the industries that has experienced rapid growth is the retail industry, so much that global brands like South Africa’s Shoprite have had a hard time penetrating the Kenyan market. Supermarkets are great custodians of big data which when mined can provide meaningful insights on basic organizational process as well as consumer behavior. The study thus sought to ascertain the extent of application of customer relationship management and data mining in predicting consumer behavior in medium and large supermarkets in Kenya. Random sampling was used for the study which was carried out amongst 30 supermarket staff in managerial cadre. Data was collected using self-administered questionnaires and face to face interviews which were also done to get optimal responses from the respondents and analyzed using descriptive statistical measures like frequency distributions and percentages. The study established that while CRM is applied in the retail industry to some extent, basic data mining techniques are still popular for mining insights and are generally not specialized for data mining functions. It thus emerged that Excel is the most widely used software at 92% with usage being mainly for measurement of sales and customer purchase frequency. On sources of customer data and the respective applications, it was established that loyalty cards would ideally provide the most comprehensive customer data. In terms of usage, it emerged that the current retail technologies are used mainly to capture sales at 100%, measure customer purchase frequency at 93% and measurement of stock movement at 90%. However, various variables were not captured due to use of manual systems. Based on the findings various recommendations were drawn, key amongst them the need to invest in advanced data mining systems and diversification of data types that can be captured by such technologies including and not limited to customer demographics, customer complaint handling mechanisms and tracking of lapsed customers. These shall be steered further by use of systems that capture data comprehensively, at best this can be captured by loyalty cards whose usage can be increased through campaign programs by the various retailers. There is still a long way to go for the industry to grow and maximize its data mining potential. Consequent research may look to establish reasons for the low uptake of data mining technologies in the industry given its known potential for a higher return on investment (ROI).
CHAPTER ONE: INTRODUCTION

1.1 Background to the Study

A competitive and volatile operating environment is a challenge every business is grappling with in the 21st century. Survival calls for day to day strategies to stay relevant; this requires a detailed understanding of consumers and as such their behavior patterns. A key source of consumer knowledge is past data otherwise known as big data which may have been captured for example from sales records, customer relationship management software and social media to name but a few of the many sources. Though relatively new, data mining pauses many opportunities in the study of consumer behavior and shaping strategy geared towards customer retention and continuous organizational growth. Traditionally, secondary data has been used to mine insights from customer data. Customer dynamics have however resulted in need for more structured systems where data retrieval is easy and efficient. One of the more common technologies in use currently is CRM (Customer Relationship Management) which enables customer profiling for better customer management.

Other more advanced technologies enable predictive analytics which is used to map trends and patterns within data upon which reliable and actionable forecasts can be made. Kenyan supermarkets are a perfect case of organizations with a wealth of consumer purchase data. According to the Economic Survey (2009), Kenya is the second most advanced country in terms of presence of supermarkets, after South Africa. Kenya has over 206 supermarkets and 18 hypermarkets. In a bid to retain customers and enable profiling, supermarkets have adopted smartcards whose points are redeemable every so often.

1.1.1 Customer Relationship Management

Customer Relationship Management (CRM) can be defined as the process of using information technology in implementing relationship marketing strategies, with particular emphasis on customer relationships (Ryals and Payne, 2001; Gummesson, 2008). CRM has its origin in the desire of combining the help desk, the customer support, the ERP and data mining (Peel, 2002)
and although most benefits of CRM are different in each business area, there are some benefits common to all businesses (Swift, 2000). These benefits are generally the following: lower cost of customers’ acquisition, improvement of customer services, customer retention and loyalty increase, higher customer’s profitability, easier identification of profitable customers and companies’ productivity increase (Alhaiou, 2011). CRM also promotes the increase of companies’ productivity since it enables the integration of all companies’ departments, such as information technology, finance and human resources (Romano and Fjermestad, 2003; Crosby, 2002; Kracklauer et al., 2001).

Following Swift (2000); Parvatiyar and Sheth (2002); Kracklauer et al. (2004), Ngai et al. (2009) categorizes analytical CRM on four dimensions: customer identification, customer attraction, customer development and customer retention. These four dimensions can be seen as a closed cycle of the customer’s management system (Au et al., 2003; Kracklauer et al., 2004; Ling and Yen, 2001). Only by analyzing customer’s data, companies can understand behaviors, identify buying patterns and trends and discover causal relationships. The insights obtained from the data help to model and predict future customer satisfaction and behavior more accurately, and may constitute a quantified foundation for strategic decision making (Reynolds, 2002).

1.1.2 The Concept of Data Mining

Data mining is the process of searching and analyzing data in order to find implicit, but potentially useful, information. It involves selecting, exploring and modeling large amounts of data to uncover previously unknown patterns, and ultimately comprehensible information, from large databases. Data mining uses a broad family of computations that include statistical analysis, decision trees methods, neural networks, rule induction and refinement, and graphic visualization (Shaw et al, 2001). Recently, organizations have adopted knowledge based marketing, strongly anchored on availability of large data in organizations. Among goals that can be achieved through data mining is customer profiling, designing marketing campaigns amongst other marketing functions.
The data mining process begins with a research question often a task for management and for which an answer is provided through data mining tools that analyze two key requirements; description and prescription. Data mining is also preceded by classification of data mining methods designed for description and prediction. Description is the process of discovering patterns, associations, and relationships among key customer characteristics such as demographic variables, frequency and purchase rate. Prediction uses this patterns and relationships to predict future trends and behavior. A range of data mining methods exist and are applied depending on the data types requirements of an organization and range from classification, regression, time series, clustering, association and sequence discovery. Classification, regression and time series are primarily used for description whereas clustering, association and sequence delivery are used to describe relationships that exist in the data (Hair et al, 2008).

Data mining in retail relies heavily on CRM systems that capture consumer information. Kotler (2006) defines CRM as the process of creating and retaining profitable customers with long term relations by delivering customer value and satisfaction. Hair et al (2008), have broken down the key organizational applications of CRM as improving efficiency of market segmentation, increasing profitability of repeat purchase behavior and enhancing sales and media effectiveness. A more relationship oriented application of CRM in modern marketing provides a wider array of CRM usage and is further refined to allow for exchange of information with customers, determining heavy users, determining lifetime customer value and building segment profiles.

1.1.3 Consumer Purchase Behavior

Engel, et al. (1986) defines consumer behavior as an act/s by individuals directly involved in obtaining, using, and disposing of economic goods and services. This includes the decision processes that precedes and determine these acts. Simple observation provides limited insight into the complex nature of consumer choice. Researchers have thus increasingly sought the more sophisticated concepts and methods of investigation provided by behavioral sciences in order to understand, predict, and possibly control consumer behavior more effectively.
Knowledge of consumer behavior helps the marketer to understand how consumers think, feel and select from alternatives like products, brands and the like and how the consumers are influenced by their environment, the reference groups, family, and salespersons amongst other influencers. A consumer’s buying behavior is influenced largely by cultural, social, personal and psychological factors. An understanding of consumer behavior can best be understood through consumer research which has been defined as the study of the processes involved when individuals or groups select, purchase, use, or dispose of products and services, ideas or experiences to satisfy needs and desires (Solomon, 1995). Consumer purchase behavior can be more easily predicted if the customers use loyalty or credit cards during their purchases as their demographics are readily available. Just to emphasize on benefits that can be sought from data mining, predictive analytics initiatives show a media ROI of 145%, in comparison to non-predictive business intelligence initiatives with a median ROI of 89% (Wayne, 2007).

1.1.4 The Kenyan Retail Sector

The Kenyan retail sector is seen as one of the most developed in sub-Saharan Africa, while Kenya’s Vision 2030; the government’s blueprint for the economy aims to increase efficiency in the retail sector, which it hopes will raise the market share of products sold through formal channels to 30% in the next few years (KPMG, Africa Consumer Theory, 2013). The Kenya economic report on the other hand highlighted that in 2012 some of the key sectors driving the economic sector was manufacturing at 3.2% and recently 4.4% in 2013. Consequently, the sector was seen as contributing to the country’s GDP at 9.5% in 2012 and 8.9% in 2013 (Kenya Economic Sector Report, 2014).

In a recent article by (Ngigi) 2014, Kenya was said to be the second most formalized African country in terms of formal retail penetration with retail penetration standing at around 30 per cent. Likewise according to research this is one of the fastest growing areas in the private sector and it is a good thing as it helps grow customer choices and provide employment. However there is need to introduce policies and procedures that will ensure all tax obligations are captured. Kenya's appeal on wholesale and retail trade was seen to have rebounded since post-election disruption and the global recession curbed sales growth experienced in the years
Retail sales expanded by 8% in 2010, 7.3% in 2011 and 6.4% in 2012 to account for 12.9% of factor-cost GDP, and continued growing by 6.5% in the first half of 2013, (The Economist, 2014). In Kenya, supermarkets started in up-market niches of the large cities of Nairobi and Mombasa, and then spread into middle-class and then poorer consumer markets, and from large cities to secondary towns within the country. Supermarkets have then spread from Kenya to poorer and less urbanized countries like Uganda, Rwanda and South Sudan (Mukuria, 2011). There has been a rapid growth and development of supermarkets in Kenya in the last two decades with Uchumi supermarket being the first supermarket to be developed in Kenya around 1975, Ebrahim's was the first supermarket in Nairobi. Over the years many supermarkets have come up such as Ukwala, Tuskys, Nakumatt, Naivas, Chandarana, Ebrahim's, Eastmatt and many other supermarkets now spread throughout Kenya.

According to a research conducted by Neven and Readorn (2005), it was found that between 1990 and 2003, supermarkets in Kenya grew by a rate of 18% per year. They also found that supermarkets spread from the capital to intermediate and small towns, with 44% of supermarket sales and 58% of supermarket stores located outside Nairobi in 2003. The pattern of expansion in Kenya is similar to that of South Africa (Njenga, 2006). This pattern of first penetrating upper class urban market and then moving into lower income and rural-town markets shows that there will be a steady and rapid increase in supermarkets in East Africa.

The survey was thus necessitated by the increasing growth of supermarkets in Kenya and diverse market changes including buyouts and possibility of new international market entrants. According to Euromonitor International, Kenya is set for a major retail boom with the entry of four global retail chains such as Wal-Mart planning to be present in Kenya through its South African subsidiary, Massmart Game Stores, Jet and Edgars who plan to open retail outlets in Kenya, further intensifying competition in the Kenyan retail sector, while local formal retailers like Nakumatt, Tuskys, Uchumi and Naivas continue to expand (Retailing in Kenya, June 2014, Euromonitor International). There is also a growing middle class in Kenya with increasing demands and with need for more specialized products and services. Forward looking organizations can only manage to tap into this segment if they invest in customer relationship management systems that enable better understanding of their customer base.
1.2 Research Problem

Today’s customers have such varied tastes and preferences that it is not possible to group them into large homogenous populations to develop marketing strategies. In fact, each customer wants to be served according to her individual and unique needs. Database marketing characterized by marketing strategies based on the great deal of information available from the transaction databases and customer databases has become popular and most organizations have built up massive databases about their customers and their purchase transactions. However, due to lack of appropriate tools and techniques to analyze these huge databases, a wealth of customer information and buying patterns is permanently hidden and unutilized in such databases. Knowledge-based marketing, which uses appropriate data mining tools and knowledge management framework, addresses this need and helps leverage knowledge hidden in databases (Shaw et al, 2001).

In 2012, the leading sub-sectors were wholesale and retail trade, transport and communication, financial intermediation, and education sector services. These four sub-sectors accounted for about 38.2 per cent of overall growth in GDP. The wholesale and retail sub-sector was the second single largest source of growth after agriculture. While the sector is a major source of employment and growth, it is characterized by a large number of small and micro-enterprises, and informality. However, supermarkets have grown spectacularly by about 32 between 2000 and 2010 (Kenya Economic Report, 2013). In line with the rapid growth in the Kenyan retail sector, various studies have been done on different aspects of supermarkets; Karemu (1993) studied the state of strategic management practices in retailing sector with a particular focus on supermarkets in Nairobi.

Lagat (1995) studied the state of marketing using online activities in Kenya retailing sector while Munyoki (1997) did an analysis of the factors affecting pricing strategies of selected consumer goods in the retail market with a specific focus on supermarkets in Nairobi Kenya. Kiumbura (2003) studied retailer brands & channel conflict of supermarkets in Nairobi while Njenga (2006) investigated the attitudes of selected stakeholders towards growth strategies that were followed by large scale retailers in Kenya. The research gap indicates that none of the studies has
been known to focus on application of customer relationship management and data mining in predicting consumer behavior in supermarkets in Kenya. The study thus sought to answer the research question: What is the influence of customer relationship management and data mining on consumer purchase behavior in medium and large supermarkets in Kenya?

1.3 Objectives of the Study

The objective of the study was to determine the extent of application of customer relationship management and data mining in predicting consumer purchase behavior in medium and large supermarkets in Kenya.

1.4 Value of Study

The survey sought to make an understanding of how data mining is perceived, data capture techniques used and data types captured in medium and large supermarkets in Kenya. Little is known of retailers using data mining extensively except in banking and insurance. The financial sector utilizes data mining especially in measurement of risk i.e. based on market trends or credit worthiness of those borrowing. Further, the players have been known to engage in advanced analytics like predictive analytics which forecasts future happenings. This is mostly applied by actuarists in insurance companies and credit controllers and financial engineers in banks.

The survey also sought to unearth usage of POS (Point of Sale) data in Kenya by supermarkets. Particular areas of interest were on the key custodians of the data in the supermarkets, data capture and mining technologies used currently, effectiveness of the data mining programs, how the data mining process can be improved and the future outlook. By researching on this area with little research, the study sought to contribute to theory by giving scholars, academics and students more reference material on the concept of data mining in a retail context.
CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

This chapter details the theoretical background of the study and empirical research; it further provides an overview of the application of customer relationship management and data mining in predicting consumer purchase behavior in retail.

2.2 Theoretical Foundation of the Study

Consumer decision making has long been of interest to researchers. Beginning about 300 years ago early economists, led by Nicholas Bernoulli, John von Neumann and Oskar Morgenstern, started to examine the basis of consumer decision making (Richarme, 2007). This early work approached the topic from an economic perspective, and focused solely on the act of purchase (Loudon & Della Bitta 1993). The most prevalent model from this perspective is ‘Utility Theory’ which proposes that consumers make choices based on the expected outcomes of their decisions. Consumers are viewed as rational decision makers who are only concerned with self-interest (Schiffman & Kanuk 2007, Zinkhan 1992). The theories highlighted in the study are based on buyer behavior and customer relationship management and lay the foundation for data mining and enabling consumer behavior prediction in a retail context.

2.2.1 Theory of Buyer Behavior

Howard developed the first consumer decision-model in 1963 (Du Plessis, Rousseau et al. 1991). This model was developed further in 1969 by Howard and Sheth to become the Theory of Buyer Behavior also known as the Howard and Sheth Model (Howard and Sheth, 1969). The model provides for a sophisticated integration of the various social, psychological and marketing influences on consumer choice into a coherent sequence of information processing, (Foxall, 1990). The author’s interest was in constructing a comprehensive model that could be used to analyze a wide range of purchasing scenarios, and as such the term ‘buyer’ was preferred over ‘consumer’ so as to not exclude commercial purchases (Loudon & Della Bitta 1993).
Howard and Sheth suggested that consumer decision making differs according to the strength of the attitude toward the available brands; this being largely governed by the consumer’s knowledge and familiarity with the product class. In situations where the consumer does not have strong attitudes they are said to engage in Extended Problem Solving (EPS), and actively seek information in order to reduce brand ambiguity. In such situations the consumer will also undertake prolonged deliberation before deciding which product to purchase or whether to make any purchase. As the product group becomes more familiar, the processes will be undertaken less conscientiously as the consumer undertakes Limited Problem Solving (LPS) and eventually Routine Problem Solving (RPS) (Foxall, 1990).

2.2.2 CRM Theory

CRM can be defined as the process of using information technology in implementing relationship marketing strategies, with particular emphasis on customer relationships (Ryals and Payne, 2001; Gummesson, 2008). The first CRM initiatives were launched in the early 1990s and were mainly focused on call center activities (Rahimi, 2007). The promising emergence of CRM was influenced by the advances in information technologies, data management systems, improved analytics, enhanced communications, systems integration and internet adoption (Greenberg, 2001). Currently, in information technology terms, CRM means the integration of technologies such as: data warehouse, website, intranet/ extranet, help desk, sales, accounting, ERP and data mining. Indeed, all information technology able to gather data is integrated in order to provide the information required to create a more personal interaction with customers (Bose, 2002).

In retail, CRM systems are best placed to capture meaningful information from consumers including demographics like age, number of members in family, geographical location, gender and career among other things. The analytical bit of CRM makes use of secondary data by allowing for data mining. According to Reynolds (2002), the most critical CRM component is the analysis. Analytical CRM solutions allow managing effectively the relationship with customers. Only by analyzing customer’s data, companies can understand behaviors, identify buying patterns and trends and discover causal relationships.
The insights obtained from the data help to model and predict future customer satisfaction and behavior more accurately, and may constitute a quantified foundation for strategic decision making. According to Reynolds (2002), each CRM component is dependent on the others. For instance, analytics drives the decision making in operational CRM, e.g. sales arrangement, marketing actions and customer service processes. On the other hand, without the data collected via operational CRM processes, analytical CRM would not be possible. Moreover, the data processed by analytical CRM tools could not be used effectively and strategic decision-making would not occur without collaborative CRM; summing up, operational CRM, collaborative CRM and analytical CRM work together to create business value.

2.3 Customer Relationship Management

In recent years, the advent of information technology has transformed the way marketing is done and how companies manage information about their customers. The availability of large volume of data on customers, made possible by new information technology tools, has created opportunities as well as challenges for businesses to leverage the data and gain competitive advantage. Wal-Mart, the largest retailer in the U.S., for example, has a customer database that contains around 43 tera-bytes of data, which is larger than the database used by the Internal Revenue Services for collecting income taxes, (Graen, 1999).

The first CRM initiatives were launched in the early 1990s and were mainly focused on call center activities (Rahimi, 2007). The promising emergence of CRM was influenced by the advances in information technologies, data management systems, improved analytics, enhanced communications, systems integration and internet adoption (Greenberg, 2001). Currently, in information technology terms, CRM means the integration of technologies such as: data warehouse, website, intranet/extranet, help desk, sales, accounting, ERP and data mining. Indeed, all information technology able to gather data is integrated in order to provide the information required to create a more personal interaction with customers (Bose, 2002).
In the past, researchers generally applied statistical surveys to study customer behavior. Recently, however, data mining techniques have been adopted to predict customer behavior (Giudici & Passerone, 2002; Song et al, 2001). As such, Giudici & Passerone (2002) cite that retail market managers must not only provide high-quality products and services, but must also react appropriately to changes in customer needs. Data mining can be applied to identify useful customer behavior patterns from large amounts of customer and transaction data. Kenya boasts of a growing retail environment with supermarkets chains like Nakumatt going regional. Such growth necessitates thorough understanding of customers for better positioning in the various markets.

2.4 The Relationship between CRM and Analytical Data Mining

Berry and Linoff (2000) defines data mining as the process of exploring and analyzing huge datasets, in order to find patterns and rules which can be important to solve a problem. Berson et al. (1999); Lejeune (2001); Berry and Linoff (2004) define data mining as the process of extracting or detecting hidden patterns or information from large databases. Data mining is motivated by the need for techniques to support the decision maker in analyzing, understanding and visualizing the huge amounts of data that have been gathered from business and are stored in data warehouses or other information repositories. Data mining is an interdisciplinary domain that gets together artificial intelligence, database management, machine learning, data visualization, mathematic algorithms, and statistics.

According to Ngai et al. (2009), association, classification, clustering, forecasting, regression, sequence discovery and visualization cover the main data mining techniques. For customer identification purposes, classification and clustering techniques are the most used. If the objective is to attract customers, classification techniques are the most frequently used, while if the objective is to retain customers, association and classification are the most frequently used. Concerning customers’ development, association techniques are the most frequent. Despite this, it is known that a combination of data mining techniques is often required to support each CRM analytical dimension (Ngai et al, 2009).
2.5 Data Mining Techniques

According to Ngai et al. (2009), association, classification, clustering, forecasting, regression, sequence discovery and visualization cover the main data mining techniques. Classification aims to map a data item into one of several predefined categorical classes (Berson et al., 1999; Mitra et al., 2002; Chen et al., 2003; Ahmed, 2004). For example, a classification model can be used to identify loan applicants as low, medium, or high credit risks. Clustering, similarly to classification models, aims to map a data item into one of several categorical classes (or clusters). Unlike classification in which the classes are predefined, in clustering the classes are determined from the data. Clusters are defined by finding natural groups of data items, based on similarity metrics or probability density models (Berry and Linoff, 2004; Mitra et al., 2002; Giraud-Carrier and Povel, 2003; Ahmed, 2004). For example, a clustering model can be used to group customers who usually buy the same group of products.

Forecasting estimates the future value of a certain attribute, based on records’ patterns. It deals with outcomes measured as continuous variables (Ahmed, 2004; Berry and Linoff, 2004). The central elements of forecasting analytics are the predictors, i.e. the attributes measured for each item in order to predict future behavior. Demand forecast is a typical example of a forecasting model whose predictors could be for example price and advertisement.

Sequence discovery intends to identify relationships among items over time (Berson et al., 1999; Mitra et al., 2002; Giraud-Carrier and Povel, 2003). For example, sequence analysis can be developed to determine, if customers had enrolled for plan A, then what is the next plan that customer is likely to take-up and in what time-frame. Visualization is used to present the data such that users can notice complex patterns (Shaw, 2001). Usually it is used jointly with other data mining models to provide a clearer understanding of the discovered patterns or relationships (Turban et al., 2010). Examples of visualization applications include the mind maps.
2.6 Benefits of Data Mining

The use of data mining techniques to extract meaningful information from data is very promising with many companies having collected and stored data resulting from the interactions with customers, suppliers and business partners. However, according to Berson et al. (1999), the inability to find valuable information in the data has prevented companies from converting these data into valuable and useful knowledge. This has seen a growing need for data scientists who can barely meet global demand owing to the negligible number.

Data scientists are trained to retrieve and make sense of the most complex data. Within the analytical CRM dimension, data mining techniques are becoming popular ways of analyzing customer data. In fact, the employment of data mining to support CRM analytical dimension is seen as an emerging tendency (Ngai et al., 2009). Data mining techniques can be used to support competitive marketing strategies by analyzing and understanding customer behaviors and characteristics, so as to acquire and maintain customers and maximize customer value. In the past, companies focused on operational and collaborative tools, but this tendency seems to be changing (Reynolds, 2002). Decision-makers have realized that analytical tools are necessary to drive strategy and tactical decisions, related to customer identification, attraction, development and retention.

Analytical CRM is mainly focused on analyzing the data collected and stored, in order to create more meaningful and profitable interactions with customers. To achieve this purpose the data is processed, interpreted and reported using several tools (Greenberg, 2004). The data analyzed is part of a large reservoir of information, i.e. a data warehouse, which contains data from both external and internal sources, obtained using operational tools. The data gains value when the knowledge extracted becomes actionable.
Data mining techniques can be used to support competitive marketing strategies by analyzing and understanding customer behaviors and characteristics, so as to acquire and maintain customers and maximize customer value. The selection of appropriate data mining techniques which can extract useful knowledge from large customer databases is of utmost importance, (Ngai et al, 2009).

The survey interlopes various variables that cannot be separated; data mining systems, secondary data and application of the mined customer data. Data mining cannot be without secondary data and in retail, it is best captured via CRM systems. The study entails understanding the current data mining systems in place amongst them CRM, their efficiency and uses in day to day organizational running.

2.7 Chapter Summary

The chapter gives the theoretical background of the study which is founded on consumer behavior and CRM theories. CRM provides the basis for data mining as it is the main and most comprehensive way to capture data in retail. Further, the chapter enlightens the user on the various data categories that exist and the technological capacity required to mine the various data types. Lastly and most importantly, the chapter highlights the benefits of data mining as experienced by organizations that have embraced the technique.
CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Introduction

This section covers the research methodology that was be used for the study. Further the chapter highlights the research design applied, a description of the population, sampling design, data collection instruments used and analysis.

3.2 Research Design

The study employed descriptive cross sectional survey design whereby access to the widest possible amount of data from the entire medium and large supermarkets in Kenya was sought. Nachmias and Nachmias (2007) pointed that a researcher can employ survey design to measure variables or test hypothesis using survey data collection instruments to get a feel of a population or industry. The methodology applied was chosen as it provided an added benefit of overcoming the deficiencies that can result from employing constricted design like case study which has generalization challenges.

Survey method was used since the study required data from different supermarkets to get an unbiased view of how data mining is applied and to what extent. This method has been applied before e.g. by Bartram et al (2003) and Freeman, Cox and Wright (2006). Descriptive statistics was presented in form of frequencies, percentages and cross tabulation. Cross tabulation was used to compare and enable conclusions on the extent of application of data mining in supermarkets in Kenya.
3.3 Population

The target population of the study was all the medium and large supermarkets in Nairobi and its satellite towns. According to Busha and Harter (1980), the concept of the population surveyed is fundamental to research and refers to the group of persons or objects from which the research plans to draw inferences.

According to economic abstract (2010), there were 23 mainstream supermarkets in Nairobi. These were the primary target of the study. A census was carried out given the small number of the population of interest.

3.4 Data Collection

Primary data was used in the survey and has been described by Kothari (1985) as original information collected for the first time. The questionnaire method was selected as the research method for the survey as it makes the quantification of information possible, opinions of respondents can also be obtained in a structured manner. A semi-structured questionnaire was used to gather primary data with the survey targeting marketing customer relationship and operations managers in 23 supermarkets. The interview questions comprised both closed and open-ended questions. The open-ended questions were important for respondents to express their views and experiences as freely as possible on the Kenyan retail environment and daily organizational running. The questionnaire was divided into two sections; A, and B. Section A addressed the general information about the supermarkets while Section B addressed the data mining tools used in daily running of the organization and the extent of application of the same in predicting consumer purchase behavior.

The researcher then briefed the research assistants thoroughly on how to conduct the interviews after which field assistants distributed the final questionnaire to marketing managers in the selected supermarkets for conducting face to face interviews. The choice of respondents was because marketing managers are most involved in customer related functions in an organization.
An introductory letter signed at the university accompanied the questionnaire to give authenticity to the research and explain the purpose of the survey.

3.5 Data Analysis

The data analysis process entailed coding, entry and analysis. Analysis was preceded by coding and data entry. A code list was provided by the researcher for open ended questions to guide the research assistants during coding and was developed based on the themes that were prominent for more accuracy.

Questionnaires were collected, serialized and cleaned before they were coded. Data entry was then done before analysis. Data captured was analyzed based on descriptive statistics that included mean scores, frequencies and percentages. The results from the analysis were then presented using tables, pie charts and bar graphs for easier interpretation.

Content analysis described as a method of analyzing qualitative data by Saunders and Pinhey (1983) was used in analyzing the few open-ended questions relating to the perspectives and opinions of the managerial cadre interviewed. Open ended questions were however limited as closed questions provide for more accuracy than the latter.

3.6 Chapter Summary

These chapter highlights the population of interest and the sampling method applied. As discussed in the chapter, the population of interest was managers and supervisors in supermarkets within the Nairobi metropolis. The chapter also details the data collection method used; which is in this case quantitative as well as the data processing process and the data presentation formats used for easier reader interpretation and comprehension of the survey findings.
CHAPTER FOUR: DATA ANALYSIS, RESULTS & DISCUSSION

4.1 Survey Findings

The chapter presents the research findings based on data collected and analyzed. The findings were presented in form of charts and frequency tables and were classified as per research objectives of this survey. The study targeted 30 respondents in managerial cadre in medium and large supermarkets in Nairobi and their usage of data mining in customer relationship management and predictive analytics. The survey established that 93% of the respondents were aware of and applied data mining in daily organizational running.

4.2 Respondent Profile

Table 1: Respondent Profile

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accountant</td>
<td>6</td>
<td>20%</td>
</tr>
<tr>
<td>Data clerk</td>
<td>4</td>
<td>13%</td>
</tr>
<tr>
<td>Customer service manager/Attendant</td>
<td>10</td>
<td>33%</td>
</tr>
<tr>
<td>Branch supervisor</td>
<td>2</td>
<td>7%</td>
</tr>
<tr>
<td>No response</td>
<td>8</td>
<td>27%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>30</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

Source: Researcher (2015)

Usually data handing in supermarkets is a preserve of the managerial cadre, this is due to the sensitivity accorded such data and the high competition in the retail industry. The survey thus targeted 30 middle and senior level managers in medium and large supermarket in Nairobi. Of this, the highest representation was by customer service managers at 33% followed by accountants at 20%. Data clerks accounted for 13% of the respondents whereas branch supervisors were least at 7%.
4.3 Data Mining Using Excel Software

Awareness is what determines usage of a product at any one time. The interviewer sought to establish awareness levels for data mining amongst the respondents. Awareness of data mining amongst the respondents was high at 93%. Of those that were aware of data mining, those that use data mining used simple software like Excel which was mentioned prominently at 92%. Other software used though in a limited manner included SQL, SPSS, Lawson Portal System, ZIZI, AX and RMA. Further, the surveyor correlated various variables to establish exact usage applied via Excel software. The uses mentioned are as below.

![Excel Software Uses](image)

**Figure 1: Uses of Excel Software**
Source: Researcher (2015)

The most applied use of Excel was determining sales at 92% followed by measurement of customer purchase frequency at 83%. The least captured data types using the software were customer demographics like age and gender at 25%, as well as measurement of lapsed customers also at 25%. Other uses applied by Excel software included monitoring stock movement (79%), measurement of the average cost of purchases per customer (71%) and the average value of goods replaced (63%). It is important to note that while measurement of customer purchase frequency was relatively high at 71%, measurement of repeat purchases was only utilized 33% of the time which could imply there is limited customer database that is complete for any accurate prediction of consumer behavior.
4.4 Data Types Captured in Retail

The survey sought to ascertain the types of data captured in the existing organizational systems. The data captured was on:-

![Figure 2: Aspects Captured in the Data Mining System](image)

Sales forecasting was basically captured 100% of the time which makes sense given it is one of the key variables used to measure organizational profitability. The current technologies in use are also able to capture customer purchase frequency 93% of the time, together with stock movement at 90%. Variables currently not captured as ought by current systems includes customer demographics (20%), lapsed customers (27%), number of customer repeat purchases at (33%) and customer complaints at (47%).

Current systems are also able to capture the average cost of purchases by customers and average value of goods replaced. Measurement of customer value is important to enable profiling and segmentation. This also enables more targeted promotions and campaigns largely improving the success of such campaigns and minimizing on costs arising from poorly targeted campaigns.

4.5 Current State of Supermarkets on Data Mining

The interviewer asked various questions to ascertain the general state of data mining in supermarkets. The statements were ticked as true or false based on the respective supermarkets.
Statements that were said to be true included availability of adequate data capture systems within the organization, use of customer data by managerial cadre for decision making and managerial recognition for need to mine data at 93% scoring. Management can be seen to be very supportive of data mining in supermarkets and could be an appreciation of the benefits sought from the procedure. Table 4 presents the general state of data mining.

**Table 2: Current Supermarket Status on Data Mining**

<table>
<thead>
<tr>
<th>Variable</th>
<th>TRUE</th>
<th>FALSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>The organization has adequate data capture systems</td>
<td>93%</td>
<td>7%</td>
</tr>
<tr>
<td>Managerial cadre uses customer data for decision making and organizational strategy</td>
<td>93%</td>
<td>7%</td>
</tr>
<tr>
<td>Management recognizes the need for mining customer purchase data</td>
<td>93%</td>
<td>7%</td>
</tr>
<tr>
<td>The organization is investing in technologies that will enable better capture of customer details</td>
<td>87%</td>
<td>13%</td>
</tr>
<tr>
<td>Management of customer data is the role of a particular department in the organization</td>
<td>77%</td>
<td>23%</td>
</tr>
<tr>
<td>There are not sufficient technologies to enable the capture of consumer data effectively</td>
<td>13%</td>
<td>87%</td>
</tr>
<tr>
<td>The organization lacks skilled people that can retrieve and make sense of customer data</td>
<td>7%</td>
<td>93%</td>
</tr>
<tr>
<td>It is difficult mining data from the customer database</td>
<td>0%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Source: Researcher (2015)

### 4.6 Performance of Data Capture Methods Used in Supermarkets

The survey sought to establish the data capture methods applied in the supermarkets, consequently, the survey also sought to establish performance of the data capture methods on efficiency, ease of data capture and retrieval and ease of data integration amongst other metrics. The best rated data capture method was POS at 89% with efficiency and easy interpretation of data being rated highest at 92%. Customer Relationship Management Systems (CRM) was the
least used of the data capture methods at 72%. Ability to integrate customer data from other sources was rated relatively low for Loyalty cards at 75% in as much as the system rated highest on adequacy of customer data at 92%.

Table 3: Performance of Data Capture Methods used

<table>
<thead>
<tr>
<th>Variable</th>
<th>Loyalty cards</th>
<th>POS</th>
<th>Credit Cards</th>
<th>CRM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency</td>
<td>86%</td>
<td>92%</td>
<td>79%</td>
<td>72%</td>
</tr>
<tr>
<td>Ease of data retrieval</td>
<td>90%</td>
<td>90%</td>
<td>80%</td>
<td>70%</td>
</tr>
<tr>
<td>Adequacy of customer data</td>
<td>92%</td>
<td>85%</td>
<td>86%</td>
<td>73%</td>
</tr>
<tr>
<td>Ability to integrate with other data</td>
<td>75%</td>
<td></td>
<td>87%</td>
<td>83%</td>
</tr>
<tr>
<td>sources</td>
<td></td>
<td></td>
<td></td>
<td>71%</td>
</tr>
<tr>
<td>Easy interpretation of customer data</td>
<td>93%</td>
<td>92%</td>
<td>77%</td>
<td>75%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>87%</strong></td>
<td><strong>89%</strong></td>
<td><strong>81%</strong></td>
<td><strong>72%</strong></td>
</tr>
</tbody>
</table>

Source: Researcher (2015)

Satisfaction with overall performance of supermarkets was assigned a percentage score of 86% by the managers, of this 21(70%) were satisfied with the other 9(30%) being fully satisfied with extent of application of the data mining component in the organization.

4.7 Data Mining Usage across Departments

The surveyor also sought to establish departments that make the most use of customer purchase data within the supermarkets. The customer care department was said to be the biggest user of the data at 57% followed by marketing at 43%.
Table 4: Departments that use Data Mining

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer care</td>
<td>17</td>
<td>57%</td>
</tr>
<tr>
<td>Marketing</td>
<td>13</td>
<td>43%</td>
</tr>
<tr>
<td>Finance</td>
<td>8</td>
<td>27%</td>
</tr>
<tr>
<td>Logistics &amp; operations</td>
<td>5</td>
<td>17%</td>
</tr>
<tr>
<td>IT</td>
<td>2</td>
<td>7%</td>
</tr>
</tbody>
</table>

Source: Researcher (2015)

It is interesting to note the minimal application of the data by the operations department which was said to be applied at 17%. The IT department also applies the data mining function minimally at 7%. Usage of customer data by the various departments was as shown in table 3.

The highest usage of the data was by the customer care department for customer management purposes at 23% followed by tracking customer spending at 20%. It thus stands that the customer care department is the biggest consumer/user of customer data in supermarkets. The least applied uses of the data were mentioned as being measurement of customer tastes and preferences, storage of information, interaction with customers, sales forecasting, tracking customer points and measurement of customer basket size all at 7%.
Table 5: Departmental Application of Data Mining

<table>
<thead>
<tr>
<th>Department</th>
<th>Nature of Use</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistics &amp; Operations</td>
<td>Customer management</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td>Store information</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>Asses tastes and preferences by customers</td>
<td>7%</td>
</tr>
<tr>
<td>Customer Care</td>
<td>Customer management</td>
<td>23%</td>
</tr>
<tr>
<td></td>
<td>Customer usage/spending</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>Store information</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td>Interact with customers</td>
<td>13%</td>
</tr>
<tr>
<td></td>
<td>Basket size</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>Asses the taste of customers &amp; preferences</td>
<td>7%</td>
</tr>
<tr>
<td>Marketing</td>
<td>Improve sales/sales forecasting</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td>Making orders</td>
<td>13%</td>
</tr>
<tr>
<td></td>
<td>Research &amp; development</td>
<td>13%</td>
</tr>
<tr>
<td></td>
<td>Customer usage/spending</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>Tracking customer points</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>Interact with customers</td>
<td>7%</td>
</tr>
<tr>
<td>Finance</td>
<td>Making orders</td>
<td>13%</td>
</tr>
<tr>
<td></td>
<td>Tracking customer points</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>Research &amp; development</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>Improve sales/sales forecasting</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>Prepare the budget</td>
<td>7%</td>
</tr>
<tr>
<td>IT</td>
<td>Interact with customers</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>Store information</td>
<td>7%</td>
</tr>
</tbody>
</table>

Source: Researcher (2015)
4.7.1 CRM Usage

The CRM software was rated 72% on efficiency. It was the least used of the data sources showing need to integrate consumer data. The software is seen to perform well easy interpretation of data at 75% as well as capturing adequate data at 73%.

Table 6: CRM Rating

<table>
<thead>
<tr>
<th>Feature</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy interpretation of customer data</td>
<td>75%</td>
</tr>
<tr>
<td>Adequacy of customer data</td>
<td>73%</td>
</tr>
<tr>
<td>Efficiency</td>
<td>72%</td>
</tr>
<tr>
<td>Ability to integrate with other data sources</td>
<td>71%</td>
</tr>
<tr>
<td>Ease of data retrieval</td>
<td>70%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>72%</strong></td>
</tr>
</tbody>
</table>

Source: Researcher (2015)

4.8 Gap Analysis on Current Data Mining Capabilities

Various functions can be achieved through the application of data mining in the organization. The respondents were asked to rate relevance of various functions vs the extent of application within the organizations. This was on a 5 point scale with 1 meaning poor and 5 excellent.

Of the variables rated on importance, customer profiling (4.79) and basket analysis (4.82) were rated highest. In-store promotion was one of the variables assigned lesser importance at a mean of (3.75). Satisfaction with achievement of the various metrics was rated highest for stock management at a mean of (4.17) followed by basket analysis at mean of (4.03). The greatest variation between importance and satisfaction was seen with customer profiling (-0.86) and sales forecasting (-0.79). This could mean current systems in use are not able to capture this particular detail adequately and thus the need for improvement.
4.9 Broadening Data Mining

As a way of comparing current users and prospective users of the data, the surveyor further enquired on which departments would be in most need of using customer data for planning and decision making purposes. Usage by the research and development department was recommended at 33%. It is interesting to note that usage of customer data by the research and development department is still low in the retail sector despite its high importance in forecasting and ascertaining the organizational direction. 13% are also of the opinion that data mining should be applied by the logistics and operations department mostly tasked with sourcing for stock supplies.
Figure 4: Departments that Should Adopt Data Mining
Source: Researcher (2015)

4.10 Opportunities for Data Mining

As an indicator to what could benefit the retail industry if adopted, the surveyor sought to establish what other variables the respondents would be interested to measure in the future. This was mainly based on their daily experiences and thus the data needs cited were those that would be ideal for adopting in the future. The question was open ended to allow for a more varied response. Figure 5 below highlights the responses given in detail.

Figure 5: Data that can be captured in Future
Source: Researcher (2015)

Capturing of customer complaints and an efficient customer feedback mechanism were the leading concerns at 50% followed by lapsed customers at 25%. Incidentally, these were the least captured variables by the current data capture mechanisms. As was established, customer
complaints are largely captured manually rather than electronically making it difficult to ascertain and trend past customer patterns.

4.11 Discussion

Data mining in the retail sector is still in its early stages of development with Excel being largely applied i.e. 92% of the time. Data mining applications vary from the most basic forms i.e. classification to the more advanced applications like predictive analytics. Application of the said uses is based on the researchers needs and especially in ascertaining the degree of various relationships of data. Given the current basic usage as indicated by the popularity of the Excel software, data mining at this point can be said to at the classification level. Classification aims to map a data item into one of several predefined categorical classes (Berson et al., 1999; Mitra et al., 2002; Chen et al., 2003; Ahmed, 2004). Other advanced analytics that can be achieved through data mining include clustering, regression and predicting amongst other uses.

The perception by the respondents interviewed however is that the capacities are well of age and fully supported by management. Of the uses mostly applied currently, sales measurement is the key use mentioned by 100% of the respondents, followed by measurement of customer purchase frequency at 93%. Stock movement was also ascertained 90% of the time. The largest user department is the customer care department at 57% followed by the marketing department at 43%. On proposed usage by other departments, Research and Development was mentioned 33% of the time. It can be seen that data handling is largely the preserve of the customer care department who may lack the requisite skills that may be available in IT for in-depth data mining. The data capture methods used currently include POS, Loyalty Cards, Credit Cards and CRM. Loyalty Cards and POS are the most used at 89%. They are preferred for efficiency and easy interpretation of data captured at 92%. It is however important to note that it is difficult to merge data from the various sources and thus the need for a specific data mining software that would easily integrate with other data sources. The interviewer further ascertained that the customer data usage perceived as most important includes customer profiling and basket analysis at 96% followed by sales forecasting at 90%. Variables assigned lesser importance included customer segmentation at 79% and planning in store promotions at 75%. It is a point of concern
that customer segmentation is not very highly regarded as it informs customer profiling as well as strategic marketing. Consumer purchase behavior can be more easily predicted if the customers use loyalty or credit cards during their purchases as their demographics are readily available. Just to emphasize on benefits that can be sought from data mining, predictive analytics initiatives show a media ROI of 145%, in comparison to non-predictive business intelligence initiatives with a median ROI of 89% (Wayne, 2007).

Moving forward, it pertinent that retailers tap the wealth of their warehouse data to guide organizational strategy, this especially with the fast growth of the retail sector in Kenya and prospecting entry by multinationals. As has been mentioned, data mining techniques can be used to support competitive marketing strategies by analyzing and understanding customer behaviors and characteristics, so as to acquire and maintain customers and maximize customer value (Reynolds, 2002). Finally, it will also be important to use systems that allow capture of more customer details and more specifically one that provides for an efficient and effective customer feedback mechanism as well as an automated complaint handling mechanism proposed for capture at 50%. The system should also capture details of lapsed customers to ease follow through and conversion rates. Further opportunities in data mining are in amongst others establishing an automated customer complaint solving mechanism and a system that identifies lapsed customers for follow up purposes.

4.12 Chapter Summary

The chapter details study findings on the state of data mining in the retail industry in Kenya and includes findings on the software used to mine data in organizations currently, types of data captured by the various software, user departments as well as specific usage for each of the data types captured. Further, the chapter highlights challenges encountered in mining customer data and current information gaps while highlighting future organizational data needs.
CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.0 Introduction

This chapter presents the summary of the findings, conclusions and recommendations of the research study.

5.1 Summary of Findings

The purpose of this research study was to ascertain the extent of application of data mining practices within supermarkets in managing customer relationships. The research utilized descriptive research where the survey was conducted on medium and large supermarkets in Nairobi. Out of the 30 surveyed supermarkets, 93% were aware of data mining practices and likewise applied them in the organization. Structured questionnaires were used to collect data. The collected data was analyzed using descriptive tools and was presented using frequency distribution tables and charts.

From the study, it was established that application of data mining was high in the surveyed supermarkets, systems applied however are not modern and do not allow for detailed data mining and thus basic analysis. Excel is widely used at 92%. The research established that whereas the systems can capture various customer details, they are still largely unable to detect customer complaints as well as lapsed customers. This can be attributed to manual processes and lack of adequate penetration of loyalty cards amongst customers. The most applied use emerged as measurement of sales at 92%, measurement of customer purchase frequency at 83% and monitoring of stock movement at 79%. It also emerged that the marketing and customer care departments were the main users with IT and finance departments applying the least usage.
5.2 Conclusion

The purpose of this study was to ascertain the extent of application of data mining practices in managing customer relationships in retail and predicting consumer behavior. Little progress has been in adopting data mining for the retail sector. In the same breadth it was established that technologies in use currently do not allow for in-depth data mining. Excel is still widely used at 92% with usage being mainly for measurement of sales and customer purchase frequency. On sources of customer data and the respective applications, it was established that loyalty cards would ideally provide the most comprehensive customer data. Though gaining on popularity amongst Kenyan consumers, many are yet to own loyalty cards making customer profiling difficult. In terms of usage, it emerged that the current retail technologies are used mainly to capture sales at 100%, measure customer purchase frequency at 93% and measurement of stock movement at 90%. However, various variables were not captured due to use of manual records. This applies for complaint handling. About 50% of respondents would like the capture of complaint automated for better client management.

Variables currently not captured adequately by current systems include customer demographics (20%), lapsed customers (27%), number of customer repeat purchases at (33%) and customer complaints at (47%). Customer demographics are very important for segmenting and profiling customers for easier targeting. In retail, CRM systems are best placed to capture meaningful information from consumers including demographics like age, number of members in family, geographical location, gender and career among other things (Bose, 2002). This thus means current systems in use are not exploiting customer usage data to capacity or are not technologically up to date. Despite data mining methods being seen to be largely traditional, respondents interviewed perceive themselves to be advanced in terms of data mining capacity and capabilities with managerial support being one of the drivers for usage. This could be emanating from a lack of awareness of existence of advanced systems which provide for detailed data mining or a lack of understanding of benefits that can be sought from adopting a highly efficient data mining system.
5.3 Recommendations

Data Mining remains key in fueling growth of companies especially through an increased return on investment. Supermarkets are a great custodian of big customer data that can be mined to understand customer habits, make meaningful decisions and guide organizational strategy. The study arrived at the following recommendations.

5.3.1 Technological Growth

Technologies in use currently do not allow for in-depth data mining with Excel still widely applied at 92%. There is need to invest in advanced technologies like SAS which allow for detailed data mining. These would have to emanate from awareness of capabilities that can be achieved through advanced software. Consequently, current data mining systems are not able to capture data from all sources losing the customer journey in between. This also provides for loopholes in the customer management system. The organizations should invest in CRM (Customer Relationship Management) software to enable proper management of the ever dynamic clientele and from which patterns can be easily established for predictive analytics.

5.3.2 Create Consumer Awareness on Adoption of Loyalty Cards

It was established that the most used data capture methods were loyalty cards and the POS system. Loyalty cards provide for a more comprehensive data base but are not very largely used by consumers as ought. Retailers should thus invest in customer awareness programs on adoption of loyalty cards as a way to enable customer relationship management. Credit cards also proved to be minimally used as opposed to POS and Loyalty cards which emerged as the post popular data capture methods. This can be explained by the low adoption rates for credit cards amongst users and the growing preference for mobile money. Organizations should device ways to diversify the capture of customer details including those who use the mobile money platform.
5.3.3 Diversify Data Types that can be captured

The current systems in use are able to capture various consumer variables including average cost of goods purchased, basket goods combination and frequency of customer purchases. One of the key service determinants; the complaint handling mechanism however remains undeveloped and provides for possibilities for poor customer relations. This is largely because customer complaints are captured manually, automation of this factor may be one of the most crucial ways to retain customers which is largely needing in all the surveyed supermarkets. Variables currently not captured by the data capture systems include customer demographics (20%), lapsed customers (27%), number of customer repeat purchases at (33%) and customer complaints at (47%). Systems developed henceforth should integrate this functions for a better defined customer journey. As such, integration of current customer data sources emerged as a challenge with POS being easiest to use. It is important that the data is integrated and shared with all the necessary departments to improve the value chain as well as service times.

5.4 Suggestions for further Research

Research on data mining in Kenya is still very low, this by large can be associated with the low uptake amongst the business community. The area is fairly new in Africa although developed markets have long adopted the technology. Further, western countries have gone beyond basic data mining to concentration on big data which enable establishment of complex relationships between data over time. It has been noted that data mining is largely a preserve of the financial sector, who have applied the technique for ages. Despite the low industry uptake, the importance of data mining cannot be underestimated as it has been known to help businesses cut costs, minimize risks and even increase the overall return on investment. Further research could be carried out to determine the data mining systems that would be most appropriate for the Kenyan market given application is currently very basic creating a vague picture of benefits that could be achieved from applying the technique. Uptake is also slow in Kenya vis a vis other modernized countries it would thus be necessary to conduct further research on factors influencing the low adoption of the data mining systems in the retail sector.
5.5 Chapter Summary

The chapter summarizes conclusions from the survey as well as recommendations based on survey findings. Further, the chapter highlights further recommendations for research based on identified research gaps from the survey.
REFERENCES


APPENDICES

Appendix I: Introductory Letter

C/o University of Nairobi,
P.O Box 30197-00100,
NAIROBI
KENYA

TO WHOM IT MAY CONCERN

Dear Respondent,

REF: MSc RESEARCH STUDY

I am a student pursuing a Masters of Science, Marketing Research at the University of Nairobi. In partial fulfillment of the requirements to the award of the Masters degree, I am required to carry out a study on “Application of Customer Relationship Management & Data Mining for Predicting Consumer Purchase Behavior in Medium & Large Supermarkets in Kenya”. The choice is based on your strategic importance in the achievement of organizational goals hence improved performance of the company in terms of return on investment and sustainable business. I kindly request your assistance by availing time to respond to the questionnaire. A copy of the final report will be made available to you at your request. The information given will be treated with utmost confidentiality for the purpose of this study only. Your assistance will be highly appreciated.
Appendix II: Questionnaire

SECTION A: GENERAL INFORMATION

(Kindly put your response on the spaces provided below)

<table>
<thead>
<tr>
<th>Date</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Name of supermarket</td>
<td></td>
</tr>
<tr>
<td>Respondent’s Name</td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
</tr>
<tr>
<td>Respondent’s Number OR Email</td>
<td></td>
</tr>
<tr>
<td>Interviewer’s Name</td>
<td></td>
</tr>
</tbody>
</table>

Q1. When did the supermarket start operations? .................................

Q2. How many branches does the supermarket have? ............................

Q3. How many employees does the supermarket currently have? ............

Q4. What is the yearly marketing spending of the supermarket in Ksh? ....
SECTION B: APPLICATION OF DATA MINING

Q5. Data mining is a technique that enables decision making through retrieving and making sense of customer data stored within organizational systems e.g. from loyalty cards and POS (Point of Sale Systems) in supermarkets.

a) Are you aware of the meaning of the term data mining?

Yes [ ] If yes go to (b) No [ ] (explain meaning then Proceed if aware to b, close if unaware)

b) Does your organization in any way apply this technique?

Yes [ ] No [ ]

Q6. The capture of Customer data is a very important aspect of any business and requires application of specialized software to capture customer details which if any of the following software is applied in your organization?

<table>
<thead>
<tr>
<th>Software</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a) SAS</td>
<td>1</td>
</tr>
<tr>
<td>b) SQL</td>
<td>2</td>
</tr>
<tr>
<td>c) SPSS</td>
<td>3</td>
</tr>
<tr>
<td>d) EXCEL</td>
<td>4</td>
</tr>
<tr>
<td>e) Other</td>
<td></td>
</tr>
</tbody>
</table>
Q7. Below is a list of data capture methods used in various organizations. How would you rate the methods applied in your organization on the following aspects on a 5 point scale (1, 2, 3, 4, and 5) where 1 is the lowest score and 5 the highest score (enter rating in the boxes below)

<table>
<thead>
<tr>
<th>Loyalty Cards</th>
<th>POS Systems</th>
<th>Credit Cards</th>
<th>CRM Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Efficiency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) Ease of data retrieval</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) Adequacy of customer data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) Ability to integrate with other data sources</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e) Easy interpretation of customer data</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q8a. Based on your observation, which department/s makes most use of customer purchase data and for what purposes?

<table>
<thead>
<tr>
<th>Department</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q8b. Which department do you think is in most need of using past customer data for planning and decision making?

<table>
<thead>
<tr>
<th>Department</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistics &amp; operations</td>
<td>1</td>
</tr>
<tr>
<td>Customer care</td>
<td>2</td>
</tr>
<tr>
<td>Marketing</td>
<td>3</td>
</tr>
<tr>
<td>Finance</td>
<td>4</td>
</tr>
<tr>
<td>Research and Development</td>
<td>5</td>
</tr>
</tbody>
</table>
Q9. Various functions can be achieved through the use of data mining in an organization. Please rate the relevance of the following functions in your organization VS the extent of application of the same for decision making purposes on a scale of 1 to 5 where 1 means Poor and 5 means Excellent for implementation, and a scale of 1 to 5 where 1 means not important at all and 5 very important on relevance?

<table>
<thead>
<tr>
<th>Level of Importance/Relevance</th>
<th>Level of Implementation 1= Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1= Not Important at All</td>
<td>5= Excellent</td>
</tr>
<tr>
<td>5= Very Important</td>
<td>(Rate Accordingly)</td>
</tr>
<tr>
<td></td>
<td>(Rate Accordingly)</td>
</tr>
<tr>
<td>1 2 3 4 5 Customer profiling</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>1 2 3 4 5 Sales forecasting</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>1 2 3 4 5 In-store promotions</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>1 2 3 4 5 Stock management</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>1 2 3 4 5 In-store planning and layout</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>1 2 3 4 5 Customer segmentation</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>1 2 3 4 5 Basket analysis (e.g. analyzing items bought per customer)</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>1 2 3 4 5 Other(Please Specify)</td>
<td>1 2 3 4 5</td>
</tr>
</tbody>
</table>
Q10. Basing your answer on how your organization is run on a daily basis, how would you rate them on application of customer data for decision making purposes?

<table>
<thead>
<tr>
<th></th>
<th>1(Poor)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5(Excellent)</th>
</tr>
</thead>
</table>

Q11. From your observation and experience working in this organization, the following statements are true or false (Tick appropriately)

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>FALSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) The organization has adequate data capture systems</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2) Managerial cadre uses customer data for decision making and organizational strategy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) It is difficult mining data from the customer database</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) There are not sufficient technologies to enable the capture of consumer data effectively</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5) The organization lacks skilled people that can retrieve and make sense of customer data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6) The organization is investing in technologies that will enable better capture of customer details</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7) Management recognizes the need for mining customer purchase data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8) Management of customer data is the role of a particular department in the organization</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Q12. Data capture systems can enable capturing of various important data. Which if any of the following ASPECTS is your data mining system able to capture? **Code Appropriately**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer purchase frequency</td>
<td>1</td>
</tr>
<tr>
<td>Average cost of purchases per customer</td>
<td>2</td>
</tr>
<tr>
<td>Customer demographics e.g. Age and Gender</td>
<td>3</td>
</tr>
<tr>
<td>Sales</td>
<td>4</td>
</tr>
<tr>
<td>Nu. of customer repeat purchases</td>
<td>5</td>
</tr>
<tr>
<td>Customer complaints</td>
<td>6</td>
</tr>
<tr>
<td>Average value of goods replaced</td>
<td>7</td>
</tr>
<tr>
<td>Stock movement</td>
<td>8</td>
</tr>
<tr>
<td>Lapsed customers/customers that no longer make purchases</td>
<td>9</td>
</tr>
</tbody>
</table>

Q13  (a) In your opinion what other data do you feel would be important to capture in future?

Thank you for taking part in the survey
Appendix III: List of Major Supermarkets in Kenya

1. Chandarana Supermarkets
2. Cleanshelf Supermarkets
3. Eastmatt Supermarkets
4. Jaharis Supermarkets
5. Kassmart Supermarkets
6. Maguna-Andu Supermarkets
7. Metro Cash & Carry
8. Naivas
9. Maathai Supermakets
10. Nakumatt
11. Quickmart Supermarkets
12. Rikana Supermarkets
13. Selfridges Supermarkets
14. Tumaini Supermarkets
15. Tuskys
16. Uchumi Supermarkets
17. Ukwala Supermarkets
18. Karrymatt Supermarkets