Identifying the most valuable customers of a supermarket chain in Cyprus

MSc Marketing 2011-2012

ANDREAS PAPANTONIOU

Identifying the most valuable customers of a supermarket chain in Cyprus:
A Customer Database Analysis Using RFM

By

ANDREAS PAPANTONIOU

2012

A Dissertation presented in part consideration for the degree of MSc Marketing.

TABLE OF CONTENTS

ABSTRACT

The current study aims to illustrate the usefulness of the utilization of customer data into the retail industry and in particular to the supermarket sector. This piece of work tries to fill the gap in the literature regarding the application of database marketing techniques to real-life examples and also to be the starting point of more research related to the Cypriot retail context. The main objective of this study is to analyse the customer database of a supermarket chain based in Cyprus in order to segment their customers into homogeneous groups and then proceed to the identification of the most valuable customers. Apart from the theoretical contribution, this study aims to offer to the company's executives an opportunity to realise the potential of customer data. The RFM analysis was employed in order to segment the customer database and score each customer group according their Recency, Frequency and Monetary values. The findings suggest that the most valuable group of customers was consisted from 3657 customers belonging to the 555 and 455 segments. These customers represent more than the 34% of the total gross sales while they comprise more than 10% of the total cardholders. Also, a number of other interesting findings were also discussed, such as other valuable segments, middle-ranked segments and the least valuable customers.

ACKNOWLEDGEMENTS

The successful completion of my dissertation would have been impossible without the invaluable help of certain people. I take this opportunity to extend my sincere gratitude to those people who have helped me throughout the project.

First of all, I would like to thank my supervisor Dr. Andrew Smith for his understanding, invaluable guidance, advice and continuous support throughout the course of this project. He provided me with useful suggestions on how to approach my study and he was always willing to discuss my ideas, edit my work and ask the right questions to guide me on how to think about my problems.

Moreover, I would like to thank a number of people outside the academic field. First, I need to thank Papantoniou Supermarkets and in particular the Marketing and the IT Managers. They were very willing to help me acquire the customer data while they were always keen to

provide me with further information when it was needed. Second, I wish to thank RAI Consultants and their staff for their instant response to my inquiries. They were very helpful and without their contribution I would not be able to collect adequate information about the Cypriot retail context.

Finally, I would like to thank my family and my friends who have given me the encouragement and unfailing support in my pursuit of this master's degree.

1. INTRODUCTION

> "A *supermarket's most loyal customers are around 1,000 times more profitable than it's least loyal*" (Clark, 1997, p.146).

Clark (1997) signifies the imponderableness value of the most loyal customers and reveals the need to identify this group of customers so as to manage them more effectively. It could be said that the loyal customers are the most precious assets of a company since they are strongly related with its financial prosperity and profitability (Helgesen, 2006). Apart from their significant contribution to an organisation's financial performance, loyal customers benefit their favourite companies in many other, direct and indirect, ways. Gremler and Brown (1999) argue that the impact of loyal customers is "*analogous to the ripple caused by a pebble tossed into a still pond*" (p.1). In particular the authors support that loyal customers can benefit a company by encouraging new customer patronage or by engaging other actions and behaviours that create value for the organization, such as positive word-of-mouth communication. Hence, it is concluded that every company should identify, target and nurture its most valuable customers and an effective instrument to achieve this is a loyalty scheme.

Retailers have introduced the concept of loyalty schemes a long time ago but very few have managed to utilize it at its full potential. According to Pauler and Dick (2006) there are two perspectives from which one can perceive a loyalty scheme. The first one is the discount-orientated paradigm which supports that the main objective of a loyalty programme is to generate loyalty giving out rewards in mass promotions (Hughes, 1999; Curtis, 1999; Partch, 1999). What is often the case is that the members of a loyalty programme exchange their accumulated points for a shopping voucher or an item from the rewards catalogue. The second perspective is derived from the database marketing field. This paradigm utilizes the potential of data and goes beyond the mass promotion campaigns. In particular, database marketers argue that a loyalty card scheme cannot promote customer loyalty with simple mass marketing campaigns. Their view is that loyalty schemes' task is to collect useful transaction and socio-demographic data from cardholders and aim to the design of profitable, tailor-made special offers (Lee, 1996; Beenstock, 1999).

Supermarkets are among the first retailers which adopted such programmes. Often operating in a fiercely competitive environment, supermarkets need to develop a viable customer

retention strategy. Since the key to the development of such a strategy comes along with the implementation of a successful customer relationship management programme, supermarkets should be able to identify the most profitable ways to nurture and keep a loyal customer relationship. Database marketing offers to the retailers the opportunity to understand their customers' shopping and patronage behaviour. By understanding the purchasing behaviour of their customers, supermarkets would be able to divide their customer into a number of segments and as a result they will be able to develop efficient, tailor-made marketing campaigns for each customer group. But, most importantly, supermarkets would be able to identify their most valuable customers. This customer group is the main source of retailers' income so it is very important to be identified and treated accordingly.

As already said, few retailers have utilized the real capabilities of their loyalty programme. An example is the case of a supermarket chain in Cyprus, named Papantoniou Supermarkets. The particular retailer, although they have developed and successfully employing a loyalty card scheme, they never used it at any part of the development process of their marketing strategy. This study sets as its main objective to cluster the retailer's customer database and identify the most valuable segments. When completing this task, the current research would be the initial step towards a deeper and more comprehensive customer database analysis from the company's marketing department.

Conventional instruments of market research such as surveys, questionnaires and other similar techniques have faced severe criticism over the years. The main issue regarding these techniques is that they can only show what people say they do and not what they actually do. The solution to that problem is the utilization of the customer data. Customers' purchase behaviour can be tracked and recorded when customer data are analysed. A customer data set may include transaction histories, geo-demographic information and other valuable insights about consumer habits. This knowledge is then used for market segmentation; customers are grouped into different groups which are homogeneous within and heterogeneous in-between (Bessen 1993; Schultz and Wang, 1993; Lewington et al, 1996). Customer segmentation proved to be very efficient for retailers and especially for supermarkets. The most prominent example is the case of Tesco, who managed to become the largest retailer in the UK and one of the largest in the world after they exploited the power of the customer data (Humby et al, 2011).

Some key scholars conducted a number of studies which illustrate the importance and usefulness of customer data to the retail industry. For example, Min (2006) managed to develop the profiles of supermarket customers in the south-eastern United States by using data mining techniques. In particular, he classified the customer databases into distinct segments and then predicted a specific type of customer behaviour about selecting a specific supermarket outlet, the frequency of their visits and their basket value. Also, the patronage behaviour of the customers was examined by considering their demographic profile, shopping frequency, volume of purchase and a number of services offered by the supermarket. Among the most important findings was the conclusion that customer's loyalty to a particular supermarket and shopping frequency are important factors that influence his or her volume of grocery purchases. Also, the customer's age and whether he or she shops alone seem to influence the volume of shopping.

In addition, Ma et al (2009) and Shih and Liu (2003) used novel alternations of the RFM analysis in order to identify the most valuable customers of particular retailers. Both studies acquired customer data from large retailers; Ma et al (2009) used customer data from a shopping mall in Beijing while Shih and Liu (2003) acquired their data sets from a hardware retailing company. Both studies found that the most valuable segments of the particular retailers were consisted from a relatively small amount of customers. These findings are in line with a general rule in marketing which supports that a small number of customers are responsible for the largest proportion of the company's profits. Another important study within this field was conducted by Pauler and Dick (2006). The particular research developed a House of Profit Model which aims to maximize the profit of a food retailing chain by targeting the most valuable customers. In order to achieve this, the scholars used the loyalty card and scanner data of a food retailing chain. What these studies illustrate is the importance of identifying the most valuable customers of a company. As already said, this group of customers are very important since they are those who keep the company profitable and competitive.

Although there are a number of similar works, the current study aims to illustrate the usefulness of the utilization of customer data into the retail industry and in particular to the supermarket sector. This piece of work tries to fill the gap in the literature regarding the application of database marketing techniques to real-life examples and also to be the starting point of more research related to the Cypriot retail context. The main objective of this study is to analyse the customer database of a supermarket chain based in Cyprus in order to

segment their customers into homogeneous groups and then proceed to the identification of the most valuable customers. Apart from the theoretical contribution, this study aims to offer to the company's executives an opportunity to realise the potential of customer data. By the completion of this analysis the company would be able to use the work done as a foundation for a more extensive and sophisticate development of the concept. Moreover, an additional objective of this study is to identify and discuss any other important segments that may result from the analysis.

This piece of work neither follows a qualitative nor a quantitative style. There is a single exploratory question; who are the most valuable customers? Hence, this study is more exploratory than conclusive. It is an original, data-driven work which aims to draw conclusions based on what the data can tell rather than to expect a "shaped" outcome. The RFM analysis was employed in order to segment the customer database and score each customer group according their Recency, Frequency and Monetary values. Despite being one of the most recognised models for this kind of analysis and among the best regarding the identification of the most valuable customers, it was decided to introduce a minor novelty for more accurate results. Frequency and Monetary values were transformed from aggregate to average values so as to not underestimate the buying potential of recently joined cardholders. Also, it is important to mention that the term "most valuable customers" in this study refers to the customers who show patronage behaviour towards the specific retailer while they differ significantly from the rest customers in terms of Recency, Frequency and Monetary values. More details on the methodology and an extensive description of the data set will be discussed in the methodology section.

The current study is divided into five chapters. The first one is the introduction and discusses the importance of loyal customers and the value of customer data collection. Also, a brief overview of some recent related studies is undertaken. In addition, the statement of the exploratory question and the description of the study's purpose are discussed. The second chapter is divided into two sub-chapters. The first sub-chapter is a discussion of the relevant theories and a review of the related studies found in the literature. The second sub-chapter provides background information about the Cyprus retail context. Next, the third chapter is also divided to two sub-sections. The first offers an explanation of the method and the procedures used in order to analyse the database while the second describes the data set. The fourth chapter covers the outcomes of the analysis and their interpretations. The biggest part of the discussion is focused on the most valuable segments but other important segments

were also described. The fifth chapter concludes the study by discussing the conclusions of the research, its contribution and by listing a number of recommendations for future studies.

2. LITERATURE REVIEW

2.1 RELATED THEORIES

2.1.1 Customer Relationship Management (CRM)

It is widely accepted that developing and managing effective relationships with customers is one of the most important elements of a sustainable competitive advantage. Customer Relationship Management is recognised as the most effective marketing technique for this purpose. It is a customer centric strategy which involves human and technical dimensions and its main aim is to promote the commitment of the entire organization, aligning its people, process and technology in order to serve customers. Fox and Stead (2001) defined CRM as the development, establishment, maintenance and optimization of long-term valuable relationships between companies and customers. The scholars insist that CRM's success is based on understanding the needs and wants of the customers. These needs should be placed at the heart of the business by integrating them with the organization's strategy and practices.

The importance of CRM is illustrated by studies such as the Srinivasan and Moorman's (2005) work who managed to link CRM with greater customer satisfaction. Moreover, the studies from Bolton (1998), Forenell (1992) and Mittal et al (2005) argue that customer satisfaction is also related with greater customer loyalty, less customer complaints and increased shareholder value. In this highly competitive economy companies are focusing on customer retention and how to increase customer loyalty. It is well accepted that retaining a customer cost less than trying to acquire a new one. Rosenberg and Czepiel's (1984) study confirms that statement by supporting that marketers tend to focus on customer retention rather than to compete for new customer acquisitions.

One of the most important tasks of CRM strategies is to gain customer loyalty. Loyal customers are the most valuable assets for a company or a brand. Regarding the supermarket industry Clark (1997) supports that the loyal customers of a supermarket are a thousand times more valuable than the least loyal customers. He also argues that 65% of UK IT and

telecommunications marketers support that customer loyalty is the most important issue in determining a company's future. CRM can boost customer loyalty by collecting and then utilizing customer data. The companies that know their customers' needs are likely to cultivate loyalty more effectively. Loyal customers are likely to remain customers for longer periods, buy more and bring new customers by spreading positive word of mouth. For instance, it is argued that "*long-time customers tend to be less price-sensitive and provide free word-of-mouth advertising*" (Jain & Bagdare, 2011: pp.31-32). This argument is supported strongly within the marketing literature and practice communities, even if a number of scholars disagree (Reinartz and Kumar, 2002).

2.1.2 Loyalty

Defining customer loyalty has been a very difficult task for many scholars so far. In the marketing literature there are three distinct approaches that measure loyalty. First, there is the behavioural dimension, which considers the constant and repetitious purchase behaviour as an indicator of loyalty. This is not always the case, since, as Tepeci (1999) argues, repeat purchases are not always the result of a psychological commitment toward the brand. The second approach suggests that the attitudinal measurements mirror the emotional and psychological attachments inherent in loyalty; hence the attitudinal measurements are related with the degree of loyalty. A third approach support that loyalty can be measured by aggregating the behavioural and the attitudinal dimensions. According to Pritchard and Howard (1997), the use of both attitude and behaviour in a definition for loyalty increases the chances to predict loyal customers.

Behavioural loyalty

Traditionally, customer loyalty was seen as a behavioural measure. Under this definition, future loyalty was defined according customers' past behaviour. There are various techniques which define loyalty based on behavioural measures; some of these are the probability of purchase (Massey et al, 1970), probability of product repurchase (Lipstein, 1959; Kuehn, 1962), purchase frequency (Brody & Cunningham, 1968), repeat purchase behaviour (Brown, 1952) and purchase sequence (Kahn et al, 1986). Moreover, Magi (2003) argues that in the retailing context practitioners apply customer behaviour measures such as the share of purchase (SOP) and share of visits (SOV). SOP refers to the measurement of the relative share of a customer's purchase, compared to the total number of purchases. SOV compares

how many times a customer has visited a store with the total number of store visits. Moreover, Berger and Nasr (1998) suggest the Share of Wallet (SOW) measurement which compares the expenditure at a specific outlet against the total category expenditures. Under the same logic, Hughes (1996) proposes the RFM measurement, which refers to a customer's relationship with a business in Recency, Frequency and Monetary terms. The term Recency refers to the customer's latest purchase from the store, Frequency refers to how often a customer shops and Monetary refers to the dollar amount that the customer spent with a company. This model is the one which is going to be used in this study and is going to be discussed in detail in the methodology section.

The majority of loyalty programmes follow the above techniques to reward behavioural loyalty. This means that the more you spend with the company, the more benefits and rewards you get. According to Dowling and Uncles (1997), awarding behavioural loyalty lays the danger that customers may sometimes end up associating their loyalty towards a specific reward scheme rather than the brand. Moreover, Reinartz and Kumar (2002) raise their voice of concern by insisting that current loyalty schemes are problematic since the relationship between behavioural loyalty and profitability is weak. In particular, Reinartz and Kumar (2000; 2002) found empirical evidence that rejects the four commonly believed benefits of customer loyalty. According to Reicheld (1996) the benefits of customer loyalty is that loyal customers cost less to serve, are less price sensitive, spend more time with the company and spread positive word of mouth for their favourite brands. Therefore, it is concluded that behavioural loyalty by itself cannot be a measure of "actual" customer loyalty while it can be unreliable when used to predict customer profitability (Kumar and Shah, 2004). Another problem that was found in the marketing literature about the current loyalty schemes is the fact that most of these programmes are not forward looking since customers are rewarded for their instant or past purchases (Reinartz & Kumar 2003; Yi & Jeon, 2003). Reinartz and Kumar (2003) and Yi and Jeon (2003) insist that marketers who follow these approaches fail to consider the future potential of their customers.

Henceforth, the question that arises is if it is possible to develop a loyalty programme that promotes "true loyalty" and can reward customers "today" for their future purchases. According to Smith (1998) a loyal customer is the one who "*feels so strongly that you can best meet his or her relevant needs that your competition is virtually excluded from the consideration set and the customer buys almost exclusively from you*" (cited from Shoemaker and Lewis, 1999; p. 349). Shoemaker and Lewis (1999) argue that in order to develop a

"true" customer loyalty, practitioners should not only include customers' behavioural aspects but also the attitudinal side of their customers' purchasing behaviour drive.

Attitudinal loyalty

According to Shankar et al (2003) attitudinal loyalty refers to a long-term commitment of a customer to the company that cannot be measured based only on customer repeat purchase behaviour. Attitudinal loyalty is important because is likely to trigger future and repeat purchases (Liddy, 2000) and positive word of mouth spread (Reichheld, 2003). A group of marketing scholars (Buttle, 1996; Jones and Sasser, 1995; Oliver, 1999; Sirohi et al, 1998) define loyalty as being largely an attitude. Their main argument is that loyalty is a psychological attachment to a brand or firm, related to commitment. This assumes a favourable disposition towards a firm based in trust, familiarity, confidence, a perception of shared values and past experiences.

According to the grocery industry, Flavian et al (2001) argue that there is a change in customers' attitudinal behaviour in terms of store loyalty. In particular, the authors insist that during the 1970s and 1980s the loyal customer was described as *"a person who was not interested in discounts or advertising, who showed a certain aversion to purchasing and was not adventuresome*" (Flavian et al, 2001; p. 86). Moreover, Goldman (1978) argues that such customers made relatively little market search, visited a small number of stores and tended to buy only from firms that knew well. Conversely, the new loyal customer's profile is completely different. For instance, an evidence of this change is illustrated by studies supporting that the proximity from a customer's house to a grocery store plays a very important role to customer loyalty (McGoldrick and Andre, 1997). Similarly, East et al (1997) found a negative relationship between the degree of loyalty and the time needed by the customer to travel to an outlet. In other words, the closer the outlet is, the more loyal the customer is. In order to overcome this, loyalty programmes are used as a mean to "*reduce the importance of distance on consumer patronage by asking cardholders to drive by more proximate competitors to shop at the store where their patronage is rewarded*" (Allaway et al, 2006; p.1320).

Regarding the Cyprus context, the shopping behaviour of Cypriot consumers has changed dramatically during the last three years. The reason for this change is the global financial crisis and its results like unemployment, additional taxes and reductions in wages. The retail sector is suffering from this change as the consumers are becoming more price-sensitive and

less loyal. According to RAI consultants (2011) there was a 10% drop in retail sales' volume of 2011 compared to 2010. In particular, consumers are postponing large expenses while they are decreasing unnecessary expenditures like entertaining and dining out. Regarding the supermarket industry, the consumers now seem to consider brands less important than before and they turn their attention to private labels. In addition, they are waiting for good special offers in order to buy their groceries. Furthermore, a relatively large number of consumers use discounters as their main shopping channel.

A study for the impact of economic crisis on consumer behaviour in Cyprus revealed that, among others, shoppers prefer to shop cheaper products and brands when visiting supermarkets (26%), they avoid ordering food from out (22%), they limited their entertainment at home rather going out (20%) and they minimized the purchases of expensive alcohol drinks (15%) (RAI Consultants, 2011). Furthermore, the prominent characteristic in the "new era" of consumer behaviour is the decrease in average basket and the increase in the number of visits. In short, people prefer to make smaller and more frequent purchases than before. The findings of this study suggest that the Cypriot consumer has become more price-sensitive. As already said, this has a negative impact on brand and retailer loyalty. For instance, the study revealed that a customer is likely to visit at least three supermarkets in three months. When asked the consumers the reasons for choosing a hyper/supermarket, researchers found out that low pricing was by far the most important factor (33%) compared to product quality (15%) and convenient location (13%).

The change in consumers' behaviour is likely to influence the customer database clustering since it will be conducted regarding each customer's purchasing frequency, monetary value and date of last purchase. The data cover customer behaviour in period between 2009 and 2011 so the consequences of financial crisis should influence the findings.

Loyalty card programmes are mostly behaviour-centred since they do not award customers in order to change their attitudinal loyalty but instead they reward their past behaviour (Allaway et al, 2006; Sharp & Sharp, 1997). Loyalty schemes are data-rich environments and are suitable for exploring relational outcomes. Data acquired from loyalty cards such as consumer information, time, day, type of products bought, prices and amount spent, provide, among others, invaluable insights into consumer purchasing processes, patterns of long-term purchasing behaviour and the ability to measure the success of marketing campaigns. Allaway et al (2006) support that loyalty programmes are by nature, less appropriate for

generating attitudinal research. The authors insist that even if researchers try to study relational inputs they will come up with conclusions which may have serious statistical problems. The main reason of these problems would be that the researchers would need a random sample of the huge loyalty programme database which does not guarantee to cover the range of different behaviour groups within the scheme. This can only be applied after behaviour-centred research identifies distinct customer segments.

Considering the above assumptions, the aim of this study is to focus on behavioural loyalty by studying customers' shopping behaviour based on recency of last purchase, frequency of visits and monetary amount spent to a specific retailer.

2.1.3 Retail Reward Programs

> "*A loyalty programme helps segment and reshape the profile of the customer base, an essential task if marketing spend is to be directed mainly at the best customers. Choosing the most profitable customers and accurately targeting them and nurturing them, while virtually deselecting the least profitable customers, is one way of vastly improving bottom-line profits*" (Clark, 1997; p.146)

In the early 1990s the big players in the food and drug retail industry introduced the first customer reward programs as part of their efforts to increase customer loyalty to a particular store or company by maximizing the potential of the customer base (Passingham, 1998). The main objective of the concept was to reward customers for their loyalty to a specific company. In the retail industry, the most typical loyalty development strategies have taken the form of card-based reward schemes. Most common schemes involve the issuance of coded, scanner-readable cards which are "swipped" or scanned at checkout. Depending on card usage, customers receive various rewards and benefits. Rewards can take the form of immediate cost savings, members-only offers, gifts, special promotions or points which can be exchanged with a range of products. Allaway et al (2006) argue that loyalty card programs aim to increase customer propensity to visit one retailer over a competitor in spatially configured markets. In other words, it may be interpreted as a type of a "switching barrier" which makes customers more reluctant to move to a competitor. But in an environment that all competitors employ similar strategies, the switching costs are relatively low or even non-existent hence loyalty schemes do not act as switching barriers anymore.

Joining a loyalty programme does not mean that you become a loyal customer. Members of loyalty programmes tend to express their loyalty differently, while some others show no loyalty at all (Jones & Sasser, 1995). There are various categories of patronage behaviour within a loyalty scheme. According to Allaway et al (2006) a loyalty scheme may include shoppers who increase their patronage of the firm, cardholders who are not satisfied and quit, customers who belong to competitive loyalty programmes as well, those who shop only when there are offers and those who their serving costs exceed the benefits towards the company and must be abandoned. Furthermore, a group of customers is likely to be loyal without any reward schemes. For this group loyalty programmes are a waste of money. The existence of such diversity within a retailer's customer base illustrates the need of segmentation. By clustering their databases into a number of categories, retailers are able to target more effectively each group of customers with tailor-made selling propositions.

Firms should be benefited from such schemes in two ways. The first and the most obvious way is that, at least theoretically, members of such schemes should be encouraged to make repeat purchases from the particular firms; hence the firms will increase their sales. The second major benefit of a loyalty card scheme is the collection of customer information. Such information includes customers' name, age, marital status, area of living and other important details. This knowledge, coupled with customers' historical transactions is essential for the firms in order to set up effective marketing strategies.

The importance and usefulness of loyalty schemes is illustrated by the fact that some of the largest retailers base a big part of their strategic decision making on the customer knowledge acquired by those schemes. The most prominent example is the case of Tesco's Clubcard. Tesco's Clubcard was the main reason behind the giant retailer's success in the past decade. Before the introduction of Clubcard Tesco was UK's second supermarket. Nowadays Tesco is the UK's number one grocer, the world's most successful internet supermarket, one of the fastest-growing financial services companies in Europe and one of the most successful exponents of CRM. It is not a secret that Tesco's management praise Clubcard for these excellent results. According to Humby et al (2011):

"*The events of 13 February 1995 changed the way Tesco makes decisions, develops products, manages its stores and, most important, the way it serves its customers. On that day, Tesco launched Clubcard, its customer loyalty programme*" (p.2).

The information acquired from loyalty card schemes, has direct marketing implications on all retail mix variables while can significantly impact on store location decisions. For instance, Tesco, because of the customer knowledge acquired from the Clubcard, was able to cluster their customers to six life-stage segments (Rayner, 1996). The six segments receive a different, tailor-made of the Clubcard magazine. Furthermore, each segment is divided into micro-segments based on their shopping behaviour. These sub-segments receive personalised offers and invitations to events relevant to their needs.

2.1.4 Database marketing

Database Marketing and CRM

Wehmeyer (2005) makes a distinction between database marketing and CRM. According to the author, "*database marketing is understood to be IT-enhanced direct marketing"* while CRM is seen as "*transactional marketing-mix marketing by direct means*" (p.244). Database marketing uses the unlimited power of databases to broaden and support the entire mix. Its main capabilities which are applied to marketing are segmentation, value analysis, controlling and reporting (Wehmeyer, 2005). CRM is also strongly related with IT usage. In fact, Ryals and Payne (2001) argue that CRM acts as a "strategic bridge between IT and marketing strategies". CRM is the part of relationship marketing which focus on customer retention, on long-term and profitable relationships development and on the maximization of customer value for the company. CRM approaches differ significantly from those of transactional marketing and often need sophisticated IT solutions. Although direct and database marketing strategies can be used to achieve CRM goals, Wehmeyer (2005) insists that CRM is not always an original sub-task for direct and database marketers. In contrast, he argues, CRM is mostly discussed at a strategic level, where customer retention and relationships development are emphasized.

Definition

Schoenbachler et al (1997) define database marketing as the "*collection of data, such as customers' names, addresses and purchases, which provides marketers with information that enables them to make better decisions in working toward accomplishing the company's objectives*" (p. 5). In an alternative definition, Jutkins (1994) emphasizes the benefits of database marketing both to marketers and consumers. In particular, he defines database

marketing as "*gathering, saving and using the maximum amount of useful knowledge about your customers and prospects … to their benefit and your profit*" (Jutkins, 1994; p.40).

Lewington et al (1996) suggests that "*harnessing the power of database marketing can be the basis of competitive advantage*" (p.329). The evolution of relationship marketing and the realization from the companies that in order to be competitive they should develop strong relationships with their customers made database marketing an integral element in strategic marketing. It is now well known that it is easier and much cheaper to retain an existing customer than it is to acquire a new one. The main advantage of database marketing for the companies is that it enables them to build a profitable individual relationship with each customer. For example, a marketer, using the customer database can "read" the preferences of each customer, or group of customers, and target them with specialized promotions. The aim of this relationship is to make the customer feel unique, that she or he is recognized and receives personal service and attention. As Jutkins (1994) stated, database marketing is in favour of both companies and consumers. Companies are benefited because through database marketing are able to collect customer information and make better decisions. This leads to increased profitability because of more efficient promotional efforts. On the other side, consumers profit by enjoying more opportunities to make purchases they are likely to be interested in, often at advantageous prices.

It is well accepted in the marketing literature that retailers which adopt a database-oriented relationship marketing approach gain important competitive advantages compared to the firms which do not (Davis, 1997). Some key benefits of the various database marketing approaches are found in the literature. For instance Derks (1994), Berry (1995), Jackson and Wang (1994) and Jutkins (1994) argue that companies which adopt a database-driven relationship marketing approach have increased knowledge about customers and as a result more informed decisions, they are able to track customer buying patterns and understand their motivations. Retailers are also able to target promotional efforts only to customers who are most likely to respond as well as to offer varied messages to different customers. Moreover, according to the scholars, a database-oriented marketing approach gives to the retailers the opportunity to customize promotions, prices and services to individual customers, and as a consequence, to create long-term customer relationships. In addition, organisations may achieve greater results in their customer retention strategies as it is likely to achieve business growth by maintaining existing customers. Furthermore, when knowing who to target and how do to it effectively can reduce the company's marketing costs and increase profits in the

long run. Finally, database marketing techniques enable the retailers to augment main offerings with valued incentives, personalize dialogue encounters as appropriate and increase customer awareness and sales (Jackson & Wang, 1994; Derks, 1994; Berry, 1995;).

Responding to the increased diversity in consumers' preferences, modern marketing is becoming more and more customer-centric. In particular, traditional marketing methods such as distribution of marketing messages through mass media are becoming less dominant. Marketers nowadays are seeking ways to target consumers offering tailor-made propositions to meet each consumer's unique needs. Bessen (1993) argues that mass advertising campaigns have become less and less effective since diversity increases within consumers. A communication strategy for reaching consumers effectively demands one-to-one promotions and targeted advertising campaigns. This exact need makes the acquisition of extensive customer information systems such as customer database analysis a must. Moreover, companies with customer information systems are able to target small groups or niches. These micromarketing techniques, according to Bessen (1993), "*are critical to the survival of large players, since they allow big companies to own niches in the manner of smaller, more flexible competitors*".

Loyalty cards are the tools that retailers use in order to acquire customer knowledge that would help them to strengthen their store loyalty and build stronger consumer relationships. In particular, Mauri (2011) supports that the exploitation of loyalty cards as a knowledge tool follows a dynamic process that is continuously redesigned according to new knowledge additions. This process, according to the author, combines the application of basic statistical techniques and geo-marketing tools to customer database with the outlets' scanner data and demographic household data. Mauri (2011) argues that this process has three specific goals. First, retailers want to identify and describe their customers (macro-segmentation). Second, they need to set up a geographical delimitation of their attraction area. Third, and most important, loyalty cards are used to identify the best customers which comprise the 20% who are responsible for the 80% of revenue and an even greater percentage of profit. It is known in the marketing literature that the highest profitability derives from the heaviest spending customers, who are considered to be the most loyal ones.

2.2 RELATED STUDIES

Table 1. Studies using customer segmentation techniques.

Reference	Purpose of research	Industry Sector	Model(s) / techniques	Inputs / Variables
Fader, Hardie & Lee (2005)	To link RFM paradigm with Customer Lifetime Value (CLV) & estimate the total CLV for a cohort of new customers of the online music site CDNOW	e-retailing	RFM and CLV. Iso-value curves used to vusialize the interactions and trade-offs among the RFM measures and CLV	Customers' Recency, Frequency & Monetary values.
Pauler & Dick (2006)	To set up a House of Profit model, an approach to maximize the profit of a food retailing chain by targeting and promoting valuable customers.	Food retail	A refined version of Niraj's (2001) segmentation scheme & k-mean clustering.	Total sales & profit per household, sales gap coefficient.
Bult & Wansbeek (1995)	To introduce a comprehensive methodology for the selection of targets from a mailing list for direct mail. The proposed model is called the PM-approach (profit maximization). The model was validated using data from a marketing company selling books, periodicals and music in the Netherlands.	Retail	Based on the gains chart principles (Banslaben, 1992) but adds some new aspects and overcomes some limitations.	No. of books ordered in the last year & in the year before last year, no. of books ordered since date of entry, no. of mailings received, no. of no-envelopes returned and no. of quarters since last book order.
Shih & Liu (2003)	To propose a method for customer lifetime value ranking	Hardware Retail	Weighted RFM, k-means clustering, Analytic Hierarchy Process (AHP).	Customers' Recency, Frequency & Monetary values.
Ma, Li, Wang & Ran (2009)	To target valuable customers within a retail reward program database	Retail	Weighted RFMG	Customers' Recency, Frequency, Monetary & Gap coefficient values.
Chang & Tsai (2011)	To propose GRFM (group RFM) to identify high loyal and contribution	Retail	GRFM (Group RFM), PICC algorithm to	Customers' Recency, Frequency, Monetary

	customers. Also it discovers potential customers for products promotion, based on their purchasing patterns in certain product categories.		dynamically cluster customers according to a specific demand in terms of constraints associated with a product category.	values and their transactions history.
Colombo & Jiang (1999)	To choose the customers who are most likely to respond to an offer.	Market Research	RFM for the customer segmentation. They use Frequency and Recency to predict response probability and then they combine it with monetary value to predict an expected contribution.	Customers' Recency, Frequency, Monetary values, response information, transaction history, cost of goods, cost of contacting the customer.
Dan (2008)	To build RFM-based customer segmentation model to assist students' loan subsidy valuation by analysing consumption transactional histories in a university's canteen.	Finance	Weighted RFM, AHP, K-means cluster algorithm.	Customers' Recency, Frequency, Monetary values.
Spring et al (1999)	To introduce, perform, and evaluate a methodology for determining which direct mail offer should be sent to which target segment		RFM, logit regression	Customers' Recency, Frequency, Monetary values, questionnaire response variables
Allaway et al (2006)	To investigate the potential for deriving meaningful, managerially relevant customer segments within a retail loyalty-type program	Retail	Cluster analysis, scree testing, discriminant analysis, OLS regression.	Recency, Monetary, Frequency, number of items purchased each time, change of activity between year's halfs.
Mauri (2011)	To identify the heaviest users of the loyalty card and to compare the differences between their shopping behaviour and the behaviour of the light users.	Retail	Answer Tree	Card code, name of cardholder, street address, household size, data of subscription, frequency, monetary

As already said, the aim of this study is to find the most valuable (loyal) customers of a supermarket chain by clustering the loyalty card's customer database. Customers will be analysed using the RFM (Recency, Frequency, Monetary Value) method. Several studies in the marketing literature had similar objectives. RFM and its variations was the most prominent approach for database clustering. However, even if this method is seen as the most popular, an important number of researchers employed some other methodologies as well. The main disadvantage of the traditional RFM clustering approach is the assumption that its outcomes are not accurate enough in some industries as it does not considers the element of profitability. Several scholars argue that even if a customer group has the best RFM scores, it may include customers which are not profitable or exclude some others which are very profitable to the company. In this section, a review of related studies will be conducted in order to provide an overall picture of the database analysis field.

According to Chen et al (1996) clustering is a data mining tool which is used to discover knowledge procedures. Specifically, "*clustering aims to maximise variance among groups while minimising variance within groups*" (Shih & Liu, 2003: p.161). Marketing practitioners, in particular database marketers, tend to use statistical techniques in order to analyse customer data and to provide information for marketing decisions. Verhoef et al (2002) argue that customer segmentation in database marketing is used to group customers into clusters. Those clusters are homogeneous internally and heterogeneous between them. In marketing terms, this means that the members of a segment respond similarly to marketing initiatives but their reaction differs from the reaction of other clusters' members.

A research by Allaway et al (2006) analysed the loyalty card customer database of a US major retailer. They managed to identify and analyse distinct patronage segments. Their research indicates that only a small percentage of loyalty card members demonstrate purchasing behaviours that can be interpreted as truly loyal. The research was conducted by collecting one-year's data from the retailer's loyalty card programme. Then, a cluster analysis was used in order to generate a variety of potential market structures on a number of managerially relevant variables. After that, scree testing and discriminant analysis was used so as to select the most appropriate market structure. Lastly, they used multinomial logistic regression to regress a set of variables representing drives of patronage on to the clusters. The patronage-related variables were six: the distance of each cardholder to the store, to the nearest competing store, to the nearest billboard advertising the loyalty program and the distance to the nearest Very Early Adopter (a person who adopted the card program during its

first two days). Moreover, they considered the number of competing stores with a two mile radius of each cardholder, and the number of Very Early Adopters living very close to cardholder (Allaway et al, 2006).

Despite the contribution of this study, there are certain limitations. For instance, the data used were acquired from a single loyalty card program. As a result, the outcomes of this study cannot be applied in other regions without further confirmation. Also, the researchers were not able to contact a sample of the cardholder population to conduct an attitudinally-oriented research in order to verify the findings. Furthermore, the size of the data set precluded the researchers to get information like cardholder shopping habits at the particular stores before the launch of the loyalty card. Also, a better analysis should be conducted so as to find possible limitations in the spatial coding of the distance data.

Venkatesan and Kumar (2004) analysed the usefulness of Customer Lifetime Value (CLV) as a metric for customer selection and resource allocation strategy. According to the customer selection, the researchers compared their proposed method (CLV) against three other metrics: previous-period customer revenue (PCR), past customer value (PCV) and customer lifetime duration (CLD). Furthermore, they compared the customer selection capabilities of CLV with other customer-based metrics such as share of wallet and RFM. Their data was from a large multinational computer hardware and software manufacturer, which was selling mainly to business customers. Their data set was comprised from two groups of observations. The first had 1316 observations with the first purchase taking place at 1997 while the second had 873 observations and 1998 as the first purchase year. In order to conduct the metrics comparison, Vankatesan and Kumar rank-ordered customers from best to worst according to each metric and then compared the costs, sales and profits from the top 15%, 10% and 5% of customers. Their findings suggest that CLV metric better identifies profitable customer than the other metrics do. In particular, Vanketsan and Kumar (2004) argue that the difference in total profit from use of PCR, PCV, CLD and CLV across the top 5% to 15% of the entire customer base may yield more than $1 million. The findings of this study may be useful for practitioners in high-technology industry but its main limitation according to the customer selection method is that it represents only a small sample on a specific industry. In order to overcome this limitation, the study should be replicated using customer data sets from other industries and settings.

A novel approach of customer database clustering is proposed by Chang and Tsai (2011). Chang and Tsai (2011) used the classic RFM method as their basic metric with the addition of one extra dimension. This dimension was included so as to consider the characteristics of the purchased items of each customer. The authors argue that in this way the clustering method is strongly related to customer purchases and can correctly reflect their actual consumption behaviour. In addition, the GRFM (Group RFM) employs a constrained clustering method named PICC (Purchased Items-Constrained Clustering) which, based on a sophisticated purchase pattern (OPRA) table, could adjust original purchase records to satisfy various clustering constrains and decrease re-clustering time. The main advantage of GRFM is the fact that allows an individual to belong to more than one cluster. As a result, a customer can be associated with different loyalties and contributions according to the distinct characteristics of the purchased items. This difference allows GRFM to discover with accuracy the sales trend for the purchased items. This approach can benefit marketing managers to decide when to launch a specific sales promotion since the clustering results of PICC contains extra information about the distribution status inside each cluster. According Chang and Tsai (2011) previous studies that used the RFM approach (Cheng & Chen, 2009; Yeh et al, 2008; Miglautsch, 2000) failed to provide effective information for promotion of specific items. Because of that, they propose their GRFM metric as a solution to that problem.

An evolution of RFM model is also proposed by Shih and Liu (2003). The authors' proposition is to apply the analytic hierarchy process (AHP) to determine the relative importance of RFM variables. They did so because, as they insist, "*applying AHP to determine the relative importance of RFM variables is important, since the RFM weights may vary with product and industry characteristics*" (Shih & Liu, 2003: p.170). In order to apply the AHP, decision makers were asked to make intuitive judgements about the ranking order of RFM variables. There are a number of steps when conducting an AHP analysis. First, the researchers asked the decision makers to make pairwise comparisons of the relative importance of RFM variables. Then, after assessing the inconsistency index to be less than 0.1 so as to be acceptable, they proceed on computing the relative weights. The authors employed eigenvalue computation in order to derive the weights of RFM. Then, they use the weighted RFM metric so as to evaluate customer lifetime value (CLV). When the customers' RFM value is defined, the authors employ clustering techniques in order to derive CLV ranking. In particular, K-means clustering is employed to group customers with similar

lifetime value or loyalty, based on weighted RFM. In this way they can identify and compare the customer segments clearly.

In order to demonstrate the usefulness of the proposed approach, the researchers applied it to a hardware retailer. Specifically, the researchers collected and analysed a two-year data set on consumer transactions. Their findings suggest that their proposed method can produce more reasonable CLV rankings than methods which do not consider the relative importance of RFM variables. The weighted RFM method can be applied both to general and high consumption industries. Regarding the limitations of this study, it should be replicated in a different context. For example, a dataset from a completely different industry should be used in order to verify the applicability of this method to all industries.

Another study which proposes RFM as a method to segment customers and find the most valuable ones within a loyalty programme is the one by Ma et al (2009). This study, similarly to previous works, introduces an additional dimension to the traditional RFM metric. Specifically, the researchers created the gap coefficient variable in order to consider each customer's purchases made at competitors. The model, named RFMG, is tested for statistical validity using the customer database of a loyalty card scheme from a mall located in Beijing. Their findings support the view that the proposed model performs better than traditional RFM in terms of identifying and targeting the most valuable customer group in a loyalty program database. In particular, this model has two important alternations compared to the traditional RFM. First, in traditional RFM Frequency and Monetary Value are whole-time aggregates. As a result, the value of newly joined customers who have the potential to become valuable to the company is not reflected so these customers are underestimated. In order to overcome this problem, the researchers calculate RFM values as average variables. Second, Ma et al (2009) introduce the gap coefficient as the fourth variable of the model. Traditional RFM does not consider purchases made at competitors so it does not give an accurate view about the potential spending of customers. The solution to this limitation is the creation of the Gap Coefficient. Gap Coefficient represents the purchases made at competitors and it is represented as the quotient of the standard deviation of money spent for each customer's purchase divided by the average money spent of the customer.

Sales Gap Coefficient was also utilised by Pauler and Dick (2006). Pauler and Dick (2006) proposed the House of Profit Model as an approach "*to maximize the profit of a food retailing chain by targeting and promoting their most valuable customers*" (p.1263). The model is

based on four components; segmentation of households using loyalty card and scanner data, price and promotion elasticity analysis, simulation of effects of pricing and promotion and price and promotion optimization to maximize profit. According to the customers' segmentation, the House of Profit Model uses three segmentation measures. In particular, each household is ranked by its total sales, total profit and its sales gap coefficient. The gap coefficient value is calculated exactly as it was in Ma et al (2009) study. One significant difference though is the fact that the researchers in this study are interested for household values and not individual customers. In order to aggregate to the household level, they compute "Durations of households" (Pauler & Dick, 2006; p.1264). This is measured by the duration of the oldest card in the household. In particular, loyalty cards within a household. Totals of households with life duration between three months and one year are scaled by dividing them by duration while households with duration less than three months are discarded as insufficient data. When every household was ranked, hey were clustered into eight distinct groups by running k-mean clustering on their three segmentation attributes.

In short, the House of Profit Model is suggested as a framework for managers, offering differential promotions to customers by distinguishing consumer groups according their loyalty and profitability. This segmentation scheme provides better estimates for price and promotion elasticity, resulting higher profit maximization. According to the authors, this study has a number of limitations. First, they argue that the length of the time series constrains the maximum number of independent variables that can be analysed in OLS regressions. A remedy to this would be the analysis of aggregated product categories but this may result poor model fit. A second limitation of this approach is the fact that the researchers did not include the pricing and promotion of the retailer's competitors. The researchers insist that aggregating unit prices of products at competitors into product categories which are comparable without knowing their sales data is a hard task. Third, a dynamic approach of price elasticity analysis along with a rolling time window and ARMA techniques should be used so as to flatten time series and control autocorrelation.

In addition, Chan (2008) proposed an approach which combines customer targeting and customer segmentation for promotion strategies using RFM to identify customer behaviour and a CLV model to evaluate the proposed segmented customers. The applicability of this approach was tested on 4000 customers acquiring their information from the customer database of a Nissan retailer. In particular this method intends to find the appropriate customers that will be more likely to engage with a marketing campaign. The findings proved

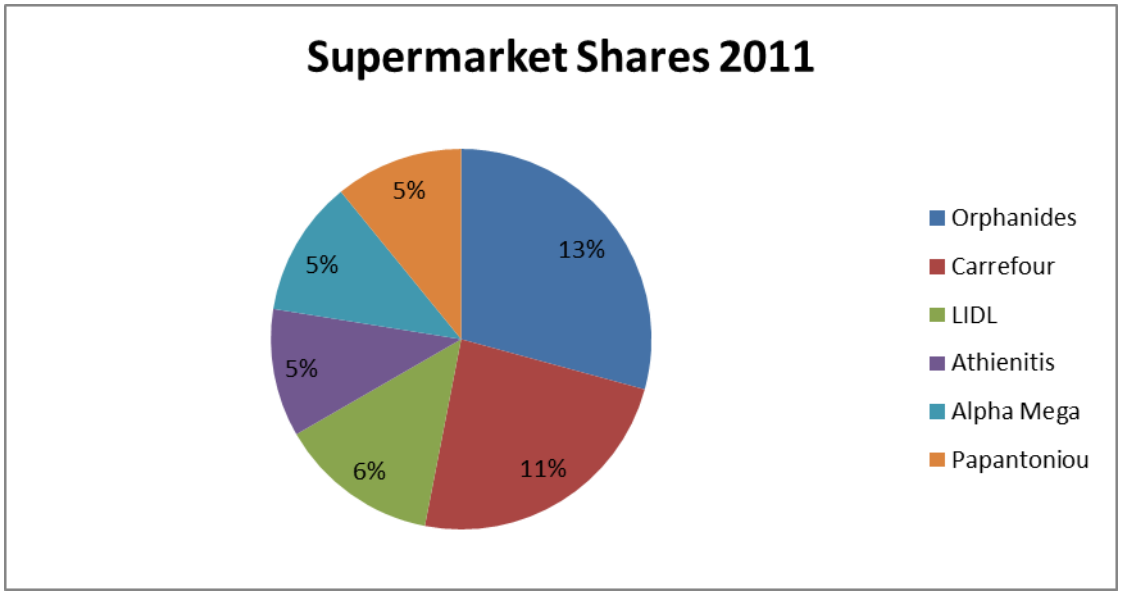
that Chan's method produces better results in targeting valuable customers than random selection while it has the potential to increase customer loyalty and customer lifetime value. A similar study by Spring et al (1999) developed a combination strategy of target selection and the selection of the strongest offer. The strategy utilises a response model which makes target selection specific to a proposed offer. This is achieved by deploying a logit model with standard RFM variables to predict the response probability for each offer type. The conclusions of this study support that combination strategy produces greater profits compared to typical two-stage strategy.

Furthermore, a study by Colombo and Jiang (1999) presents a stochastic RFM model. The proposed model is used to target customers in the in a firm's customer database by considering only recency and frequency in order to predict response probability. It is also used to predict the expected contribution of customers combining the response probability and the monetary value of each customer group. The concept of identifying the most profitable customers was also considered by Hsieh (2004). He proposed an integrated data mining and behavioural model to analyse the credit card customers of a bank. The method used a self-organizing maps neural network so as to predict profitable customer segments based on their repayment behaviour and RFM scoring. The findings of this study suggest that the values of RFM and repayment behaviour can be employed as behavioural scoring predictors affecting customer segmentation. Consequently, the newly created customer segments were profiled through customers' feature attributes and credit card usage while different marketing strategies were developed for different groups of customers. Another recent study on customer loyalty has been made by Mauri (2011). This study analyses the loyalty card database of an Italian supermarket in order to identify the heaviest users of the card, their shopping behaviour compared with light users' behaviour and the identity of the key cardholders. In order to conduct this study Mauri (2011) used the Answer Tree (SPSS Package) technique. Some of the most important findings of this research support that the more consumers buy on promotion the higher their card loyalty and that the customers buy promotions which are on absolute amounts rather than percentages. In addition, Mauri (2011) argues that more frequent consumers have a higher level on their card loyalty. Furthermore, this study supports that, for the specific retailer the key customers are those who show the highest per capita spending together with the highest purchases of special offers.

2.3 CYPRUS RETAIL CONTEXT.

Cyprus is a small yet important island in the east side of the Mediterranean Sea. Being the eastern part of Europe, coupled with its strategic geographical location, had always been the centre of attention for various political and financial reasons. Cyprus has five major cities and a population of 862,000, most of them (67.4%) concentrated in urban areas (Cyprus Statistical Service, 2012). The capital of Cyprus is Nicosia with 336,000 population and the rest cities are Limassol with 241,300 residents, Larnaca with 146,300, Paphos with 90,800 and Famagusta with 47,600. Regarding the retail industry, Cyprus in 2010 had a retail market worth of €5.562 billion, excluding motor vehicles and motor cycles market (Cyprus Statistical Service, 2012b).

The retail market structure in Cyprus follows the western European standards in general with a major exception. The market is characterised by the existence of a large number of kiosks. A kiosk is a small store, similar to convenient shops but smaller in size. For instance, there were 1172 kiosks in 2001 while ten years later the number decreased slightly to 993 (Toumazou, 2012). Cyprus retail market is divided in five main types of retailers. In particular, according to the retail Census made by RAI Consultants in 2011, there are 201 large supermarkets and hypermarkets, 608 grocery stores, 993 kiosks, 892 convenience stores and minimarkets and 346 bakery outlets (RAI Consultants, 2011). Two of these categories increased significantly from 2007, since supermarkets and hypermarkets increased by 101% and convenience stores by 62.8%. The main reasons behind these increases are the opening of new large supermarkets from the big retailers along with their expansion with smaller convenience stores. In addition, some big retailers acquired smaller groceries upgrading them to supermarkets. To be more specific, the term "supermarket" refers to stores with size between 500 m² and 4000 m² and "hypermarket" refers to stores bigger than 4000 m². Mini markets and convenience stores are between 100 m² and 500 m² and kiosks and groceries' size varies from 200 m² to 50 m².



Respect to the supermarket/hypermarket category, there are six main competitors who have a combined market share of 45%. According to RAI Consultants (2012) the market leader with 13% market share is Orphanides, a local supermarket chain having presence in every major city, owning 11 supermarkets and 18 convenience stores. The second biggest retailer in terms of market share is the multinational giant Carrefour with 11%. Carrefour is also established in all major cities, operating 7 hypermarkets and 5 smaller supermarkets. Lidl, another multinational retail chain penetrated Cyprus market recently and managed to acquire 6% of the total market share. Local retail chains Alpha Mega, Papantoniou and Athienitis are the ones who follow with a market share of 5% each.

Consumer DNA

The population of Cyprus consist of 79% Cypriots and 21% foreigners who live permanently at the island. Among the foreigners, the larger groups are Greeks (17.3%), British (14.8%) and those from Bulgaria and Romania (24.3%). One third of the population has a higher education degree while the ageing population (65+) consist the 13% of the total population (Toumazou, 2012). Furthermore, there are 300,000 households in Cyprus with an average size of 2.76 persons per household.

3. METHODOLOGY

The design of this research is more exploratory than conclusive. The main goal of this study is to explore the customer database of a supermarket chain in Cyprus in order to identify the most valuable customers. A specific hypothesis is not applicable in this kind of research since the objective for exploratory studies is to provide insights and in-depth understanding about the subject. In addition, although the outcomes of this study could be used as a kind of advice for future decision making, further validation and confirmation of the results is needed. Even if the data set used in this study provides useful behavioural characteristics for the customers of the retailer, a more detailed analysis would be more appropriate in order to reach valid outcomes. Unfortunately the retailer was not able to provide further data such as customer transaction history or demographic details about their customers. Moreover, since this is a dissertation for an MSc programme, there were significant time constraints so a extended analysis was not feasible.

3.1 RFM

> *"Never assume a CHAID program or even a regression model will outperform an old-fashioned RFM analysis if the RFM has been refining the model for more than 20 years"*
>
> (David Shepard, cited from Hughes, 2012; p.101)

The purpose of this study is to identify the most valuable customers of the company. In order to achieve this, the customer database of the company's loyalty scheme is going to be analysed. The database analysis will be undertaken using the RFM method. According to the particular methodology, the most valuable customers are those who purchase very often, have small purchase intervals and spend a large amount of money with the company. Being able to identify those customers is very useful for the company since they are the core source of income and should be treated accordingly.

The RFM model is a common segmentation technique that combines three measures (recency, frequency and monetary values) into a three-digit RFM score, covering five equal quintiles (20% group). In the traditional RFM model, Recency is regarded as the most

important measure. However, Lumsden et al (2008) argue that the weight of each measure tends to vary from industry to industry.

In order to conduct a database segmentation based on the RFM the following procedure is followed. First, the database is sorted by each element of RFM and then divided into five equal clusters. For recency, the customers are sorted by the purchase dates. Recency is measured by the number of periods since the last purchase, which measures the interval between the most recent transaction date and last date of the observation (months or days). This means the higher the score of recency, the lower the number of months or days. According the RFM literature, a customer with a high recency score is more likely to make a repeat purchase (Hughes, 1996; Kahan, 1998; Tsai and Chiu, 2004; Wei et al, 2010). The top 20% cluster is coded as 5, the next 20% cluster is coded as 4 and so forth. Eventually, each customer should have a number from 5 to 1 denoting his or her recency score.

For frequency, the customer database is sorted by the number of purchases made in a pre-set time period. The definition of frequency considers two states; single and repeated purchases. As with recency, the top quintile is coded as 5, the next 4 and so forth. High frequency scores indicate that there is a greater possibility for an individual to be a repeated customer. For monetary value, the database is sorted according the total or average amount of money spent with the company during a specified period of time. Marcus (1998) argues that average purchase amount is better to use compared to the total accumulated purchase amount. Using the average purchase amount helps to reduce co-linearity of frequency and monetary.

When the customer database is sorted and segmented to RFM variables, the "top-of-the-top" customer segment is coded as 555, while the least valuable segment is coded as 111. It is important to mention here that although the 555 segment seems to be the most valuable one, other segments with relatively high scores tend to have same or very similar importance with the top segment. In particular this is true for customers with high Recency and Frequency scores (4 and 5) and monetary scores at or above the median (3, 4, 5).

Advantages & Disadvantages

The RFM model has been one of the most commonly used database clustering method and there are several reasons for that. First, RFM is cost-effective in acquiring basic customer behaviour analysis. Also it is easy to quantify customer behaviour where customer and transactional data are stored in electronic form (Kahan, 1998, Miglautsch, 2000). As a result,

managers and other decision makers can easily understand the application of RFM model without any additional training. Secondly, RFM is a very effective way of modelling since it summarizes consumer behaviour in only three variables. Thirdly, with RFM is possible to predict customer response and boost company's profits in a relatively short term (Baecke and Van den Poel, 2009). A fourth advantage of the RFM model is that its variables are acquired from the company's internal database and they contain customer-specific information regarding the transaction history and are not gathered through the aggregate level information in the demographic datasets (Wei et al, 2010). For this reason, according to Kaymak (2001), RFM is more meaningful and accurate for targeting particular customers. Lastly, as Wang (2010) argues, RFM can effectively identify valuable customers so it is an ideal method of measuring the strength of customer relationships.

As with every business tool, so RFM does not come without any disadvantages. First of all, the major goal of RFM is to identify the most valuable customers. As a result, it focuses only on the best customers. It provides less meaningful results on recency, frequency and monetary when most of the customers do not purchased lately, did not show up regularly, or spent little. Consequently, RFM ignores the analysis on new firms operating for a short time and customers who only purchased one time making small orders. This type of customers is recognised as the 111 segment, and according to Miglautsch (2002) they may have the greatest untapped potential. Second, the limited number of selection variables that can be used by the proposed model it has been an issue of criticising. Most household characteristics have important effect on the probability of customer response so not considering them may produce inaccurate results. McCarty and Hastak (2007) argue that it is preferable to consider relational information when using RFM models. The third recognised disadvantage of RFM is the fact that it focuses only on current customers and cannot be applied to potential ones because of the lack of their behavioural history. Fourth, Wei et al (2010) argue that RFM estimates only one response model for all customers in the database, assuming that the database is homogeneous but this is contrary with the real situation as it is likely that the database may have a sizeable heterogeneity. Last, Yeh et al (2009) support that the importance of each RFM variable varies among industries, so every research, in order to be accurate, should calculate the appropriate weights of each variable before running RFM analysis.

RFM vs other models

Although RFM is a useful technique for database analysis, it seems that, at least for certain situations, some other techniques may be more appropriate. For instance, McCarthy and Hastak (2007) examined different methods for database segmentation such as RFM, Chisquare Automatic Interaction Detection (CHAID) and logistic regression. They concluded that CHAID performs better than RFM under certain circumstances. Moreover, Rust and Verhoef (2005) proposed a fully personalized model for optimizing multiple marketing interventions in intermediate-term (CRM) by comparing their model to other traditional models such as demographic models, RFM and finite mixture models. The results prove that their model outperforms traditional segmentation models in predicting the effectiveness of the intermediate-term (CRM). Furthermore, Wang (2010) used a hybrid method which incorporates kernel induced fuzzy clustering techniques in order to find outliers efficiently and to segment customers in a more effective way. This method included the robust "possibilistic" clustering method and robust fuzzy clustering method. The findings of this research support that the proposed method proved to be more effective than classic models.

On the other hand, RFM is recognised as a better method for database analysis compared to demographic models. For example, when Hughes (1996) compared RFM against demographic modelling, he found that RFM provides accurate results which are seldom offered by any demographic model. The reason for this is the fact that demographics provide information on what people are, like their income, age, marital status and home address while RFM measures what people do and how they do what they do. The goal of a marketer is to predict the future behaviour of the customer so systems based on customer behaviour such as RFM are more likely to be accurate than any combination of demographic information.

3.2 EMPIRICAL CASE STUDY

In order to conduct this study a supermarket chain in Cyprus was contacted in order to acquire their loyalty scheme customer database. Although the initial aim was to collect customers' transaction history and basic demographic information, eventually only certain customer data were acquired. There were multiple reasons for not acquiring the full data. First, the retailer does not collect demographic data; although address and postal code is requested in the signing form of the loyalty programme, only a limited number of customers give out their information. Second, acquiring and most important analysing the transaction history of every customer in the particular data set could put the completion of this study into

question since a lot of time was required. The dataset acquired for this study were anonymous and only the retailer would be able to link the findings with the actual customers.

Apart from the acquisition of the customer data set further steps have be taken in order to obtain information about the Cyprus retail context. Because there was not adequate information about the retailing background available in the marketing literature or in any internet sources, a marketing research firm was contacted. Their contribution was important as they provided their insight about the structure of the retail market as well as valuable information according the behaviour of the Cypriot consumer. This knowledge was necessary in order to describe the situation in Cyprus as well as to interpret better any related findings.

3.2.1 Data description

This section will describe the process of identifying the most valuable customer segment by using a loyalty scheme dataset from the supermarket chain under consideration. This dataset included details of 40233 customers owning a loyalty card. The loyalty program of the particular supermarket includes almost all of its customers. This assumption arises from the fact that in order for a customer to be eligible for the regular special offers that the supermarket has, he or she must be a loyalty card holder. In addition, cardholders collect points based on their total spending which can later exchange them with money coupons which can be used in every department of the supermarket. In order for an individual to become a loyalty card holder he or she should provide his or her name, phone, city and street address. Moreover, it is important to mention that the loyalty program is divided into five categories; corporate customers – such as restaurants and hotels, students, large families (four children or more), staff and the rest customers. The largest category is the general customers and the smallest is the students.

In addition, the dataset includes behavioural information about the purchase activity of these cardholders. More specifically, the dataset provides information such as the total monetary amount spent by each customer, the total visits as well as the loyalty card creation date and the date of last time shopped. The dataset covers a 36 months period; starting at the 3rd of January of 2009 until the 31st of December of 2011. Data capture for the reward card program was made via checkout scanner. Every transaction in which the card was used was recorded in the customer database and linked with the customer account.

The initial 40233 customer dataset included some cardholders who should be eliminated. For instance, every cardholder should be member of the loyalty scheme for at least 1 month.

Consequently, cardholders who joined after the 1st of December 2011 were neglected. Furthermore, cardholders under consideration should have visited one of the outlets of the supermarket chain at least 3 times; hence customers who visited the company less than that were eliminated. Moreover, the company preserves some special loyalty card accounts for non-profit organizations and charities so as the shoppers to "donate" their loyalty points to them. These accounts were identified and removed from the dataset so as to avoid any inaccurate conclusions. Outliers, frauds, and employees' data were also eliminated. When the unwanted cardholders were eliminated the database was left with 36216 records.

The traditional RFM metric uses the aggregate amount of purchases for every customer as well as his or her total visits to the shop. As already pointed out, this particular methodology was the reason for a lot of criticism, because in this way the recently joined customers are underestimated and their future shopping potential is not considered. In order to solve this problem, the aggregate values of Frequency and Monetary are replaced with average values. Specifically, for Monetary cardholders would be scored according their average monthly spending and for Frequency according their average visits per month. Using this method, recently joined customers who spend much with the company and make frequent visits should be ranked with a high score as well. Regarding the Recency metric, it estimates how many days passed from the last purchase date to the observation date. The observation date was set to the 1st of January 2012.

After the data conversion, three new variables were created.

$$\mathbf{R(c_i)} = \text{Observation date} - \text{Last purchase date}$$

$$\mathbf{F(c_i)} = F_i / D_i$$

$$\mathbf{M(c_i)} = M_i / D_i$$

Where $R(c_i)$ refers to the number of days between the observation date and the last purchase date for customer c_i. Where $F(c_i)$ and $M(c_i)$ represent the average monthly visits and average monthly spending of customer c_i respectively, F_i and M_i represent the aggregate visits and purchases respectively of customer c_i while D_i stands for the membership duration (in months) of the customer c_i. After the creation of these variables, an RFM score should be assigned to each customer. In order to achieve this, an RFM analysis was conducted using IBM's SPSS Statistics.

3.3 APPLYING RFM

RFM analysis was performed using the Arthur Hughes method (Hughes, 2012). This method bins each of the RFM attributes independently into five equal frequency bins. As a result 125 cells are created (5x5x5). More specifically, each customer should be assigned an independent score for each variable of RFM. For example, the customers with very recent purchases were assigned a "5" score for recency while those who had not purchased for a very long period of time were scored with a "1". A common problem with this method is that customers who are on the threshold point of two groups can be on either side. For instance, the recency dividing line between 5 and 4 may occur on the 28th of December 2011. Due to the sorting process that SPSS RFM analysis use some customers with a most recent purchase date of 28th of December 2011 may show up as "5s" while others with the same date may be appeared as "4s". This is a frequent problem of RFM but because this analysis should be done regularly, any arbitrary number assigned this month will be corrected next month.

Similarly, the same applied for the Frequency and Monetary variables. After this procedure was finished, each customer had an independent scoring for Recency, Frequency and Monetary. The real power of the RFM technique though, comes when the three variables (Recency, Frequency and Monetary) are combined into one three digit "RFM score". The purpose of the RFM scoring is to predict the customer behaviour, so it is important to translate the customer behaviour into numbers which then will be used to produce an accurate segmentation. The RFM score is calculated as follows:

RFM score = Recency score x 100 + Frequency score x 10 + Monetary score

When this process is finished, all customers will end up with a three digit cell assigned to their database records. This means that, for example, a customer with a 5 score for Recency, Frequency and Monetary has a "555" combined RFM score. Similarly, a customer with 4 for Recency, 2 for Frequency and 1 for Monetary has a combined RFM score of "421" and so on.

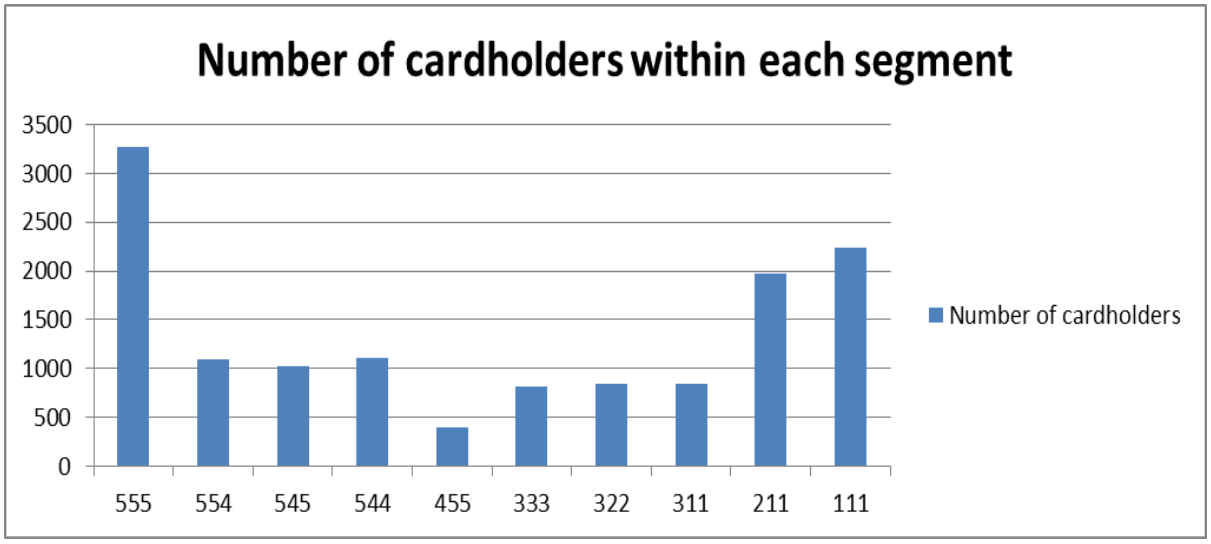
As already said, the database was clustered into 125 distinctive groups based on RFM scores. The next chapter will discuss the results of the analysis focusing on the identification and the discussion of the most valuable customer segment, the peculiarities of the analysis and the various interpretations of the findings.

4. FINDINGS AND DISCUSSION

A general rule in marketing argues that a small percentage of a company's customers contribute to a very large percentage of its revenue. It is well accepted in the marketing literature that, in general, the existence of a business is highly depended on a relatively small number of customers. These customers, often characterised as loyalists, apostles (Jones and Sasser, 1995) or true believers (Reinartz & Kumar, 2002), are the most precious assets and the lifeblood of every company. Thus, it is necessary for every company which wants to be viable and competitive to identify and retain this group of customers. In addition, companies should put much effort to increase the loyalty level of their customers. Apart from the fact that customer retention is much easier, most importantly, it is significantly cheaper than new-customer hunting (Rosenberg & Czepiel, 1984).

This chapter will demonstrate the most important findings of this study, by discussing some of the most significant segments that RFM analysis produced (Appendix, Table 1 and Figure 1). Because of the large number of the segments (125), analysing every single one would be very time consuming and unproductive, so it was decided to focus only on the most important ones (Figure 1). Customer segments such as the most valuable shoppers, the least valuable shoppers, the middle-ranked customers and a couple of other segments that present some peculiarities are going to be discussed more analytically.

Figure 1 The most important segments



4.1 General Findings

Table 2. data set Descriptive statistics

Descriptive Statistics							
	N	Minimum	Maximum	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
Recency	36216	1	991	104.04	1.072	204.080	41648.546
Frequency	36216	.0865	96.6512	6.006490	.0390966	7.4402879	55.358
Monetary	36216	.1561	4337.0938	191.710580	1.1544267	219.6931928	48265.099
Valid N (listwise)	36216						

Looking at the descriptive statistics (Table 1) a general overview of the data set is illustrated. First of all, Table 1 shows that the dataset consists of 36216 cardholders. In addition, it is demonstrated that the database has been analysed in terms of the three RFM variables. Analysing the database in terms of Recency, Frequency and Monetary a better understanding of the retailer's customers is possible. For instance, regarding the recency dimension, the minimum value is 1 day since last purchase and the maximum is 991 days. The mean for the days since last purchase is approximately 104. For Frequency, a considerable variance is observed. For instance, the minimum visits per month are 0.865 while the maximum is 96.65. Furthermore, the biggest variance (standard deviation = 219.69) is detected in monthly spending. In particular, there are customers who spend from €0.16 to €4337 per month for their shopping at the specific retailer. The considerable large variation observed within this segment was expected due to the nature of the purchases made in grocery retailing. That is, there are customers who spent much with the specific retailer while others may have shopped only a limited number of times.

As already said, the aim of this study is to identify the most valuable customers of a supermarket chain. In order to achieve this, the company's customer database had to be analysed and clustered into several groups. This analysis was undertaken using the RFM model. In particular, 36216 customers were analysed and clustered into 125 distinct groups. Each customer was assigned a three-digit RFM score; the most valuable customers are those with a "555" score while the least valuable ones are those with a "111". The results of the database analysis support that from the 36217 customers under consideration, 3267 of them belong to the 555 segment. In other words, these 3267 individuals form the most valuable

group of the company's customers. For the sake of the discussion, this group will be referred to as the "highly loyal shoppers". Moreover, a second group displays very similar behavioural characteristics so it was considered part of the most valuable customers group. It is the 455 segment and is consisted of 390 customers. In addition, the second largest segment in terms of the number of its members is the 111 cluster, which represents the least valuable customers (Figure 1). As seen from the figures for the RFM scores (Figures 1 and 2, Appendix) the particular data set shows a dipole concentration on the two ends, since the largest clusters are the first one (111) and the last one (555). Also, a considerable number of customers are concentrated in the 211 group as well as within some middle groups (311, 322, 333). Another interesting trend represented in the histogram (Figure 1, Appendix) is the relatively large segments near the 555 cluster. A more detailed, segment by segment analysis and interpretation will follow.

4.2 THE MOST VALUABLE SEGMENTS

TABLE 3 THE HIGHLY LOYAL CUSTOMERS

	The average "555" customer		
Total Spending	€ 18,993.42		
Monthly Spending	€ 582.59		
Monthly Visits	20.51		
Average spending per visit	€ 28.41		
Days since last purchase	1 day		
Duration of Membership	32.65 months		
RFM score	555		
	Segment characteristics		
Number of cardholders	3267		
Size compared to total cardholders	9.02%		
Value	€ 62,051,498.33		
Contribution to total sales	31.03%		
	Descriptive statistics		
	Minimum	Maximum	Std. Deviation
Recency	1 day	1 day	0.0
Frequency	9.08 Vpm[1]	96.65 Vpm	11.86
Monetary	€ 300 p/m	€ 3813 p/m	346.70

Analysing the highly loyal shoppers more deeply, clearly confirms the initial assumption that these customers are considered the "retailer's patrons". For instance, the average customer in

[1] Vpm = visits per month

this group has been member in the loyalty scheme for a long period, spent a large amount of money with the company each month and visited one of the supermarkets of the chain more than two times every three days. The average monetary amount per visit was approximately €28.41, a relatively low amount compared to other segments. This was expected though, since this set of customers shop very often so they do not need to stock up (Allaway et al, 2006). Also, the monthly spending ranges from €300 to €3813; this fact, confirmed with the large standard deviation (346.70) denotes that there is considerable variation between the monthly spending of customers within this segment.

The observed variation is likely to occur because corporate customers were not excluded from the data set. The term "corporate customers" refers to restaurants, hotels and other customers who tend to make large purchases from the retailer. Although it is common to exclude such customers from this kind of database analysis it was decided to keep them because the retailer itself treats them as regular customers with the exception that they are offered an extra discount on certain product categories. They are not considered as business customers though because they do not purchase very large quantities so as to be handled by the company's B2B operations. Larger companies which are customers at the Business-to-business level were not included in the loyalty scheme at all. Although corporate customers are included, when considering the large number of customers within the highly loyal shoppers segment and the relatively small average purchase amount per visit, it is concluded that even if there are some corporate customers they are not so many so as to affect the validity of the outcome.

In addition, a considerable variation is observed for the monthly visits as well. The minimum visits per month are 9 while the maximum is 96. While the minimum number of visits looks fine, the maximum value raises some questions. One possible explanation for the maximum trend is the fact that some customers use to visit one of the retailer's shops multiple times in a day. Also, since basically each loyalty card represents a household and not an individual customer it is likely that more than one person within the family uses the card. For example, one member of the family may shop for his or her needs in the morning (e.g before going to school or work) and another member will shop for the home needs later during the day.

In addition, such frequent visits can be justified if it is assumed that these customers live or work close to one of the retailer's outlet. Marketing literature suggests that the most loyal customers of a retailer might be the customers that live or work nearby the retailers' outlets

(Pan & Zinkhan, 2006; Min, 2006). Furthermore, it could be said that shoppers tend to make their everyday shopping on their way back from work. This assumption could be confirmed if the retailer had the customers' addresses or post codes stored in their database records. Unfortunately, the company does not hold a complete record with street addresses linked to cardholders so a further analysis was not feasible.

This group, apart from the fact that is the most important it is also the biggest. More specifically, the group consists of 3267 customers who comprise the 9% of the total cardholders. This finding is contradictory with the results of some similar studies in the marketing literature (Allaway, 2006; Ma et al, 2009; Min, 2006). One possible explanation for this outcome is the peculiarities of the market that the specific retailer is operating in. For instance, a large number of the retailer's customers come from a single city since four of their six supermarkets are located there. This fact, coupled with the strong relationships that the company developed with the local population gives them a competitive advantage compared to other big retailers which operating in the same city but they have fewer outlets or weaker relationships with the local residents. In fact, Macintosh and Lockshin (1997) argue that an important element that helps to develop the customer's loyalty is the interpersonal relationships between retail salespeople and customers. Furthermore, the total spending of the group represents the 31.03% of the company's revenue from all of the cardholders during the observed period. The fact that this group makes almost one third of the company's revenue confirms their patronage towards the organisation.

TABLE 4 THE 455S

	The average "455" customer
Total Spending	€ 16,657.99
Monthly Spending	€ 525.01
Monthly Visits	15.16
Average spending per visit	€ 34.63
Days since last purchase	2.61 days
Duration of Membership	32.06 months
RFM score	455
	Segment characteristics
Number of Cardholders	390
Size compared to total cardholders	1.08%

Value	€ 6,633,779.18		
Contribution to total sales	3.32%		
	Descriptive statistics		
	Minimum	Maximum	Std. Deviation
Recency	2 days	4 days	0.756
Frequency	9.13 Vpm	71.10 Vpm	6.849
Monetary	€ 300 p/m	€ 4337 p/m	342.623

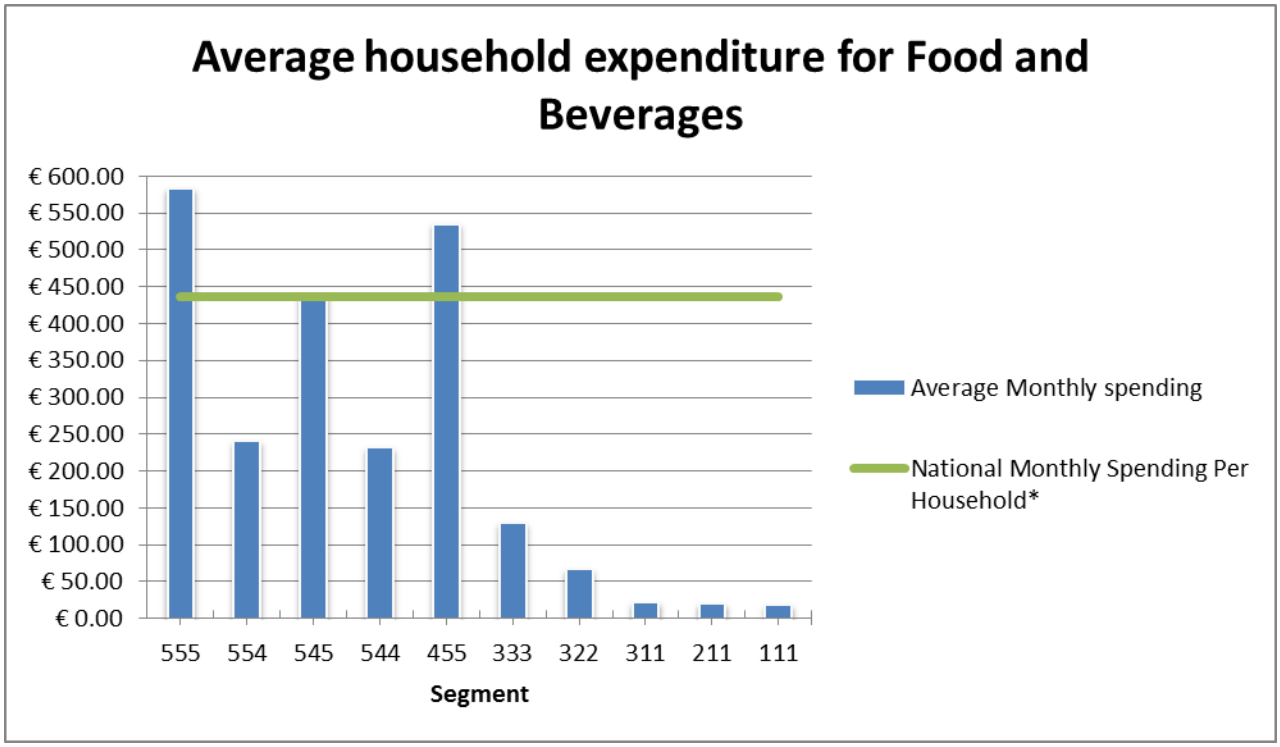
A very similar group to the highly loyal shoppers is the "455" segment. This segment is slightly different from the previous one but, practically, it could be seen as of equal value with the highly loyal shoppers group. At a first glance, the "4" in the assigned RFM score denotes that the customers within this group did not shop recently. Analysing this dimension deeper though, it was found that the average period since last purchase was 2.61 days. Bearing in mind that this time period is considered very good within the grocery retailing and that this segment has shown patronage behaviour since they are frequent shoppers (5 for Frequency) and heavy buyers (5 for Monetary), it was concluded that the particular customer group may not be seen as less valuable to the company. In addition, it could be said that this result is, at a certain degree, arbitrary since RFM assigned a recency score of "5" to customers with the most recent purchase date but a score of "4" to customers who bought 2-4 days before that date. This distinction may be important in other industries but in the retail industry, and especially for grocery retailing, this is a very good result for customers' purchase interval metric. Also, the fact that the customers belonging to this group have been members of the loyalty programme almost for the same period as the highly loyal shoppers should be considered.

Similarly with the highly loyal shoppers, this segment presents a considerable variation in the frequency of visits and in the value of monthly purchases. In particular, the outliers of this group are 9 and 71 visits per month while the minimum spending per month is €300 and the maximum €4337. As already discussed, this group has very similar behavioural characteristics as the previous one so this variance was expected and justified with the reasons discussed for the 555 segment.

On the other hand, this segment consists of a significantly smaller number of customers compared to the previous one. It has also a greater average spending per visit, which is

justified by the lower monthly visits. In particular, the average customer of this segment visits one of the supermarket's outlets approximately one time every two days while he or she spends almost €35 each time. The average total spending of the segment is € 16,657.99, which is marginally above the 3% of the total sales of the chain. Observing this trend one can argue that these customers were correctly separated from the previous segment but, considering the fact that these metrics are seen as excellent for the grocery retail market it is suggested that this group should be treated equally with the highly loyal shoppers.

FIGURE 2 Average household expenditure for Food and Beverages



An additional indicator of the loyalty levels of these two segments is the fact that their average monthly spending exceeds the national average monthly spending per household in Cyprus. According to the Cyprus Statistical Service (2011), the average monthly consumption expenditure for food and beverages per household in 2009 was € 436.87[2]. The average customer in segment 555 spends €582.59 per month while the average monthly

[2] Cyprus Statistical Service conducted a research to discover the average annual consumption expenditure per household for 2009. The findings showed that the average Cypriot spends €38,547 per annum for his expenditures. Of the total amount the 26.6% is spent on housing, 13.6% for food and beverages, 13.9% in transportation, 8.5% in hotels and restaurants, 6.8% for clothing and footwear, 5.4% for recreation and culture, 5.7% for furniture and household equipment, 5.3% for health, 3.4% in education, 3.5% in communication and 7.2% on other goods and services. The average monthly spending for groceries was calculated by multiplying €38,547 by 13.6% (Food and Beverages) and then dividing the quotient by 12 in order to find the monthly rate (€436.87).

spending for the customer group 455 is €534.79. This trend shows that both groups have customers who are very loyal to the retailer since their monthly groceries purchases are significantly larger than the average consumer in Cyprus. Also, observing this trend it could be assumed that the specific customers are retailer loyal since they are likely to spend a large proportion of their grocery shopping budget with the retailer. Such assumption though needs further confirmation since the available data for this research were not so extensive so as to reach valid conclusions about this issue.

An additional interpretation for this finding is that this metric denotes that the retailer attracts many "big spenders". The term "big spenders" refers to shoppers who tend to spend more than the average for their grocery shopping. Knowing these insights is very valuable for the retailers since they could have a better understanding of their customer and develop more personalised promotions and marketing campaigns in order to improve their revenue and profitability. As already said in the previous chapters, database marketing can facilitate the retailers in collecting and analysing customer information so as to improve their strategies, and this study could be an initiative towards that direction.

Consequently, if both groups are "merged" and considered as one, a single segment is resulting with highly loyal customers that worth approximately € 68,685,277. This amount represents more than the 34% of the total gross sales while it comes from the 10.1% of the total cardholders. The fact that more than one third of the company's total retail sales come from this group of customers signifies the segment's importance.

Regarding the least valuable segments 111 and 211, the difference on their monthly shopping compared to the national average is tremendous. More specifically, they have an average monthly spending of €18.14 and €19.86 respectively. A possible interpretation for this trend is that these shoppers spend the rest of their grocery budget to other competitors, hence there might not be valuable for the company. A more detailed analysis of these two segments though will follow in the next sections.

4.3 OTHER VALUABLE SEGMENTS

TABLE 5 OTHER VALUABLE SEGMENTS.

	Segment name		
	554	**545**	**544**
Total Spending (mean)	€ 7,365.60	€ 14,270.80	€ 7,424.86
Monthly Spending	€ 239.99	€ 439.04	€ 230.88
Monthly Visits	14.72	7.05	6.71
Average spending per visit	€ 16.30	€ 62.27	€ 34.41
Days since last purchase	1 day	1 day	1 day
Duration of Membership	30.54 months	32.6 months	32.09 months
RFM score	554	545	544
	Segment characteristics		
Number of cardholders	1089	1027	1115
Size compared to total cardholders	3.01%	2.84%	3.08%
Value	€ 8,021,136.90	€14,656,109.53	€8,278,721.94
Contribution to total sales	4.01%	7.33%	4.14%

There are a number of other segments that show similar behavioural characteristics with the most valuable customers group and need to be discussed. Also, they comprise a relatively large proportion of the customer base since they include more than 1000 customers each. The customers within these groups may not have shown the patron behavioural characteristic at a certain degree so as to be considered as the most valuable customer groups but their contribution to the company's revenue is significant. For instance, segment 554 is consisted from 1089 customers who shop at the retailer almost once every two days, spending on average more than €16 per visit. In addition, customers in segment 545 tend to show some very interesting behavioural characteristics. In particular, the average 545 customer spends € 439 per month, a number which is very close to the national average spending per household for food and beverages (Figure 2). This particular group consists of 1027 customers who contribute more than €14,5M to the company's total gross sales. This contribution is reasonably large since it is more than the 7% of total sales. Additionally, customer group 544

consists of 1115 customers. These customers, like the customers of the previous two segments, have been members of the retailer's loyalty scheme for a long period, spend more than €230 per month and shop from the retailer's outlets nearly 7 times per month or almost twice per week (Table 4). The average customer has a total spending of €7,424.86 during the specific time period while the segment has contributed more than €8M in the total gross sales.

4.4 THE LEAST VALUABLE CUSTOMERS

TABLE 6 THE LEAST VALUABLE CUSTOMERS – 111S

	The average "111" customer		
Total Spending	€ 343.00		
Monthly Spending	€ 18.14		
Monthly Visits	0.53		
Average spending per visit	€ 9.63		
Days since last purchase	337.71 days		
Duration of Membership	19.01 months		
	Segment characteristics		
Number of Cardholders	2235		
Size compared to total cardholders	6.17%		
Value	€ 766,597		
Contribution to total sales	0.38%		
	Descriptive statistics		
	Minimum	Maximum	Std. Deviation
Recency	131 days	991 days	198.059
Frequency	0.10 Vpm	1.15 Vpm	0.268
Monetary	€ 0.37 p/m	€ 42.57 p/m	10.57

Although the major task of this study is to identify the most valuable customers, the findings "compelled" the discussion of a number of additional segments; one of them is the "111s". It is the second biggest customer segment and it is named as the "least valuable customers" segment. This cluster consists of customers who have the lowest total spending with the company, visited the chain's outlets very rarely and made their last purchase a long time ago. The group's average membership duration hardly exceeds the 19 months.

The variation for the last date of purchase within this group is relatively large. For instance, there is at least one customer who shopped 131 days ago while another one made his or her last purchase 991 days ago. As recency is considered an important variable for the determination of the customer's loyalty, it is assumed that these customers are not loyal to the retailer at all. Furthermore, the minimum spending for customers within this group is only €0.37 per month and the maximum €42.57 per month. As said before, these customers are likely to be loyal to the competitors. Another explanation based on the customers' Recency, Frequency and Monetary values is that they do not consider the retailer as one of their grocery shops at all, and they just happen to visited one of the outlets a couple of times. In fact, observing the mean of the group's monthly visits (0.53) and monthly spending (€ 18.14) confirms this interpretation.

Considering these low metrics, this group of customers should normally be isolated by any future marketing campaigns because it looks like they do not provide any value to the company. They seem to be customers who have been with the company for a limited time period and then switched (or returned) to a competitor. This change could have been occurred due to several reasons; bad shopping experience, a competitor opened near their residence or work and so on.

Although some would argue that the supermarket chain should not bother to retain these customers it is believed they might worth a try. After all, it is the second largest segment, and even if half of the customers within this group move higher in the loyalty rankings then the company would be benefited dramatically. In fact, Miglautsch (2002) argues that the 111 customers may have the greatest untapped potential since they are a relatively large proportion of the total customers. Because the aim of this study was to discover the most valuable customers of the supermarket chain the RFM analysis was chosen since it is one of the best methods in marketing literature and practice for that purpose. On the other hand though, there are certain limitations according the least valuable customers. Even if average values for Frequency and Monetary variables were used, the possibility that there are viable customers buried in the 111 segment could not be eliminated. In order to understand better this segment, the company needs to conduct further analysis.

Miglautsch (2002) argues that "*sub-segmentation is the key to 1-1-1 viability and profitability*" (p.7). In order to achieve this, a number of new variables should be introduced since RFM is not designed to break up the 111 segments. First, internal purchase information

from the customers' transaction history should be added to the existing RFM variables. Second, an attribute with the geo-demographic information of each customer should be included. For instance, a postal code would be useful for the developing of certain assumptions for the current customers as well as for the future prospects. Third, according to Miglautsch (2002), "custom variables" which refer to a combination of inside and outside data should be included. For example, information from list counts could be linked with postal code data in order to produce valuable combinations. Unfortunately, the company could not provide this kind of information so a deeper analysis of the 111 segment was not possible.

4.5 Other Segments

Table 7 The 211s

	The average "211" customer		
Total Spending	€ 525.60		
Monthly Spending	€ 19.86		
Monthly Visits	0.58		
Average spending per visit	€ 11.52		
Days since last purchase	73.16 days		
Duration of Membership	26.35 months		
RFM score	211		
	Segment characteristics		
Number of cardholders	1974		
Size compared to total cardholders	5.45%		
Value	€ 1,037,539		
Contribution to total sales	0.52%		
	Descriptive statistics		
	Minimum	Maximum	Std. Deviation
Recency	27 days	130 days	29.929
Frequency	0.865 Vpm	1.156 Vpm	0.274
Monetary	€ 0.16 p/m	€ 42.07 p/m	10.84

Another large segment, the third largest compared to the highly loyal shoppers and the least loyal customers is the one with the "211" customers. Similar to the 111 segment this group

consists of customers who didn't show up recently, tend to visit the supermarket chain rarely and spend a relatively low amount in every visit. In particular, the average customer of the 211 segment has not visited any store of the chain for 73 days while he or she makes one purchase, approximately, every two months on average. The mean of the segment's monthly spending is almost €20 and the average spending per visit is about €11.50. The recency values range from 27 to 130 days while the minimum monthly spending for a customer within this group is €0.16 and the maximum €42.07.

Observing these findings a number of conclusions could be drawn. The first assumption for this group is that they are individuals who may not belong to the frequent customers of the supermarket chain but are active at a certain level. More specifically, this segment may consist of customers who shop at the company's stores only on certain occasions. For example, some may shop only when a good offer is out or only on a special occasion. If this is the case, these kinds of customers tend to do the same with every competitor. They are characterised as price-sensitive and they shop only whenever and wherever there are good promotions and special offers. Also, their low averages for visit's spending and monthly spending show that these customers purchase only a limited number of items. This adds to the assumption that this segment consists of customers who shop only when promotions and discounts are running. The second interpretation for the findings of this group is, like the previous group, that these customers are more likely to be loyal customers of other competitors and happen to buy from one of the retailer's outlets on random occasions. This is supported by their metrics since they clearly show that the retailer under consideration is not their first choice for their grocery shopping.

According to whether the company should make an effort to retain them, if the first assumption is true then the marketing literature suggests that it is very unlikely to change shopping behaviour. Despite the fact that they comprise a big segment, it would be more efficient to target the 111 customers rather than this group because they developed a certain type of consumer behaviour which would be either unchangeable or too costly to alter it. In addition, if they are loyal customers of other retailers then, as Pauler and Dick (2006) argue, promoting to them would most likely create a poor return since this segment is unlikely to increase its purchases.

TABLE 8 THE MIDDLE-RANKED

	Segment name		
	333	**322**	**311**
Total Spending	€ 3,921.20	€ 1,976.36	€ 626.36
Monthly Spending	€ 129.27	€ 66.53	€ 21.68
Monthly Visits	3.56	1.80	0.61
Average spending per visit	€ 36.31	€ 36.96	€ 13.22
Days since last purchase	12.79 days	13.65 days	14.43 days
Duration of Membership	30.24 months	29.77 months	28.91 months
RFM score	333	322	311
	Segment characteristics		
Number of cardholders	811	848	837
Size compared to total cardholders	2.24%	2.34%	2.31%
Value	€ 3,180,092.66	€ 1,675,950.54	€ 530,526.54
Contribution to total sales	1.59%	0.84%	0.27%

As already pointed out, there is a significant concentration of customers in the middle-ranked segments. For instance, three particular segments seem to distinguish from the rest in the middle ranking area. More specifically, segments 333, 322 and 311 consist of 811, 848 and 837 customers respectively. Each of these three segments is between 2.20-2.35% of the total customer base. As seen from the graph (Figure 1 in Appendix) their size is relatively large compared to other segments. The 333 segment consist of customers who spend nearly €130 per month. They visit the retailer's stores 3.5 times per month or nearly once a week and they spend approximately € 36 every time. Furthermore, their average recency is nearly 13 days while they have been members of the loyalty programme for a large period (more than 30 months). Their total value exceeds the €3M, a number which is 1.59% of the total gross sales.

These customers could be seen as individuals who live alone and shop only for their own needs. This assumption comes from the relatively low monthly spending and the less frequent visits. On the other hand, they show a relatively large average spending per visit compared to other groups; they tend to spend more than the 555 and 455 groups. The most

likely explanation for this is that they like to stock up since they do not make frequent trips for shopping. Also, their long membership duration is an indicator that they are long-term and stable customers of the retailer.

The customers in the 322 segment share some similar behavioural characteristics with those in the 333 segment. In particular, the customers within this segment have, like those in the previous group, show relatively large monthly spending and long membership duration. On the other hand, they show limited monthly spending, a little more than €66 per month, and only 1.8 visits per month. On average, the customers within this segment did not shop recently (almost 14 days). The segment's value is € 1,675,950, a number which comprises the 0.84% of the total company's sales within the considered time period.

Segment 311 has significant differences from the other two groups. Even if the customers within this group have similar membership duration like those of the previous two segments (nearly 29 months), they have lower monthly spending (€21.68) and their visits to the retailer's outlets are limited only to 0.61 per month. Furthermore, they spend just over €13 per visit for their shopping. These customers, although they have been customers of the particular supermarket chain for a long time period and they are active since they made their last purchase nearly fifteen days before, they don't seem to prefer this retailer for their grocery shopping. The low frequency and monetary rates add to this assumption. It is very likely that these customers prefer alternative solutions for their shopping but they seem to shop occasionally to the retailer. This might be happening because they are interested for certain special offers or because they like some specific products that might be available only from this retailer.

All in all, the findings revealed the most valuable customers of the company. Moreover, a number of other important segments were also discovered. Clustering the company's customer database using RFM, and analysing each segment further using additional metrics, proved to be very illustrative. Discussing the ten most important customers segments, valuable conclusions were drawn and strong assumptions were made according their profiles and behaviour. As already pointed out in earlier sections, customer data have unlimited capabilities and when utilized appropriately can generate invaluable information, making the development and the application of marketing strategy more effective and efficient.

5. CONCLUSION AND RECOMMENDATIONS

5.1 SUMMARY

To sum up, the major goal of this study was to cluster the customer database of a supermarket chain based in Cyprus and identify their most valuable customers. First of all, the relevant literature was discussed. Related topics such as CRM, customer loyalty and loyalty schemes were discussed while the usefulness of customer data collection and applications of database marketing were illustrated. The scope of the literature review section was to position this study in the wider field of relationship marketing. It is well accepted in the marketing literature that the truly loyal customers are the most important assets of an organisation. Their contribution to a company's success is invaluable since they are responsible for the largest proportion of the company's sales. Considering this fact, it is a necessity for every organisation to identify those customers in order to analyse, understand and serve them in the appropriate way in order to be able to enjoy both the short and long-term benefits that these customers bring to the company. The retailer under consideration, although they had a fully functioned and developed loyalty scheme they did not proceed to an analysis of their customer database hence they did not had the opportunity to work their relationships with the customers better. This study plans to be the initial step towards this direction since it provides the foundations for a further exploitation of their customer database.

The next section's discussion was about two main issues. Initially, the method which was used in order to conduct the database analysis was described. Since the exploratory question of this study was to identify the most valuable customers of the supermarket chain, the RFM model was chosen as the appropriate method to analyse the customer database. RFM is widely recognised for its ability to identify the most valuable customers within a customer database. The particular model divided the database into 125 distinct customer clusters with similar characteristics and purchasing behaviours. After the creation of these 125 clusters, each one was ranked according its value. Each customer was scored according his Recency, Frequency and Monetary values. The most valuable one was the 555 segment while the least valuable was the 111. Then, the focus of the discussion moved into the dataset description. In particular, it was mentioned that the dataset had gone through a number of steps so as to be suitable for the analysis. First, the whole dataset run through a "clean up" procedure so as to disregard any inappropriate records. Second, Frequency and Monetary variables were

transformed from aggregate into average variables so as to not underestimate the full buying potential of recently joined customers. After that, the application of the RFM analysis was discussed.

After the clustering process the data set was segmented into 125 customer groups. The most valuable group of customers was the 555 segment which was consisted from 3267 customers and comprised approximately the 9% of the total customers of the retailer. Moreover, a very similar segment, named 455, was also considered as part of the most valuable customers since its difference from the 555 group was negligible. This group consists of 390 cardholders and together with the 555 customers they worth approximately € 68,685,277. This amount represents more than the 34% of the total gross sales while they comprise more than 10% of the total cardholders. These findings answer the research question of this study since they reveal the most valuable customers of the supermarket chain. Apart from these findings, a number of other interesting outcomes were also discussed, such as other valuable segments, middle-ranked segments and the least valuable customers.

Certainly, this work has a number of limitations and further research should be conducted in order to verify the findings and the interpretations discussed earlier. For instance, this study should be replicated using different context and data from different industries in order to confirm this study's findings. Also, additional studies should be able to acquire and handle large amount of data such as customers' transactions history and demographic data. This would clarify a number of blurred conclusions since this study lacks this kind of data. For example, if a researcher couples the demographic data with the customer data from the loyalty scheme, he or she will be able to identify the geographical position of each segment. Additionally, the RFM model can be revolutionised by building on the current changes that this study proposed in order to be able to provide even more accurate results. By adding a number of additional variables, the particular model would be appropriate for a wide range of purposes such as the analysis of the 111 customers. Lastly, a combination of RFM and other statistical models may provide more insight according the database segmentation and the identification of the most valuable customers.

5.2 CONTRIBUTION AND RECOMMENDATIONS

Managerial perspective

Although the retailer developed and used actively their loyalty scheme for several years, the customer data collected were not taken under consideration for the development of their marketing strategy. This study could be used as a starting point for the marketing department of the retailer in order to draw on some important conclusions and start using customer data more efficiently. In this study the customer information is anonymous but the retailer will be able to link the findings with their actual customers. Additional information such as customer address, region, email and loyalty points combined with the transaction history of every customer could be very useful in order to analyse the database more thoroughly. In particular, the outcomes of the analysis for the most valuable customers can give the retailer an important first insight of who their best customers are, how they behave and what is their actual contribution to their total gross sales.

Academic perspective

At a theoretical level, the current work provides some meaningful conclusions regarding the generic marketing literature. First of all, the findings of this study confirm the well accepted argument that a small percentage of a company's customers are responsible for the largest proportion of the company's earnings. Moreover, some of the conclusions of this research are aligned with the findings of some other studies. For instance, marketing literature supports that loyal customers are likely to live or work nearby the retailer's outlets (Pan & Zinkhan, 2006; Min, 2006) and this was also found in this study. Additionally, it was found that loyal shoppers usually make small and frequent purchases because they do not stock up, something which was also supported by Allaway et al (2006). Moreover, it provided support regarding Macintosh and Lockshin's (1997) argument that the interpersonal relationships between retail salespeople and customers are an important element for the developing of customer loyalty. Furthermore, even if it has been done before, this study tries to overcome one of the RFM limitations by using average values for Frequency and Monetary. This novelty tries to measure the buying potential of recent customers too, something which is neglected in the traditional RFM approach.

Furthermore, this study could, at a certain degree, provide useful information on the shopping behaviour of the consumers in Cyprus. In particular, the segments' information and the trends found here can be used as an initial guideline in order to construct a customer map for the clientele of the supermarket industry in Cyprus. Since there is not a similar study for the Cyprus retail context and considering the fact that in general there is not much literature for the Cyprus grocery retailing, this piece of work might be the starting point for future studies related with the subject.

REFERENCE LIST

Allaway, W. A., Gooner, M. R., Berkowitz, D. and Davis, L. (2006). Deriving and exploring behaviour segments within a retail loyalty card program. *European Journal of Marketing*, **40** (11), 1317-1339

Baecke P. and Van den Poel, D. (2009) Data Augmentation by Predicting Spending Pleasure Using Commercially Available External Data. *Journal of Intelligent Information System.* 36, 367–383.

Beenstock, S., (1999). Supermarkets entice ultra-customers. *Marketing, London,* April 15, 20

Berger D. P. and Nasr, I. N. (1998) Customer lifetime value: Marketing models and applications, *Journal of Interactive Marketing*, **12**, 17-30.

Berry, L. (1995) Relationship Marketing Of Services – Growing Interest. Emerging Perspectives, *Journal of the Academy of Marketing Science*, **24** (3), 236-245

Bessen, J. (1993) Riding the Marketing Information Wave, *Harvard Business Review*, **71** (5), 151-160

Blattberg, R. C. and Deighton, J. (1991) Interactive Marketing: Exploiting the Age of Addressability, *Sloan Management Review*, **32 (**5), 5-14.

Bolton, Ruth N. (1998), "A Dynamic Model of the Duration of Customer"s Relationship with a Continuous Service Providers: The Role of Satisfaction", *Marketing Science*, **17** (1), 45-65.

Brody, P. R. and Cunningham M. S. (1968) Personality variables and the consumer decision process, *Journal of Marketing Research*, **5** (1), 50-57

Brown, G. H. (1952) Brand loyalty – Fact or fiction? Advertising Age, **23** (9), 53-55.

Bult, J.R. and Wansbeek, T. (1995) Optimal selection for direct mail, *Marketing Science*, **14** (4), 378-394.

Buttle, F. (1993) Relationship Marketing, in Buttle, F.(Ed.) *Relationship Marketing: Theory and Practice*, Paul Chapman, New York, NY, 1-16.

Chan, C. C. (2008) Intelligent value-based customer segmentation method for campaign management: a case study of automobile retailer. *Expert Systems with Applications*, **34** (4), 2754–2762.

Chang, H. and Tsai, H. (2011) Group RFM analysis as a novel framework to discover better customer consumption behaviour, *Expert Systems with Applications*, **38**, 14499-14513.

Chen, M. S., Han, J. and Yu, P. S. (1996) Data mining: An overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering*, **8** (6), 866-883.

Cheng, C. H. and Chen, Y. S. (2009) Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications*, 4176-4184.

Clark, R. (1997) Looking after business: linking existing customers to profitability. *Managing Service Quality*, **7** (3), 146-149

Colombo, R. and Jiang, W. (1999) A stochastic RFM model, *Journal of Interactive Marketing*, **13** (3), 2-12

Curtis, J., (1999). Cards versus cuts in the loyalty war. *Marketing, London*, October 7, 37

Cyprus Statistical Service (2011) *Household Budget Survey 2009*, Cyprus Statistical Service, Cyprus.

Cyprus Statistical Service (2012) *Population By District, 1996-2011*, Cyprus Statistical Service, Cyprus, accessed at 20 July 2012 from < http://www.mof.gov.cy/mof/cystat/statistics.nsf/populationcondition_21main_en/populationcondition_21main_en?OpenForm&sub=1&sel=2>

Cyprus Statistical Service (2012b) *Wholesale And Retail Trade, 2005-2010*, Cyprus Statistical Service, Cyprus.

Dan, Z. (2008) Integrating RFM model and Cluster for Students Loan Subsidy Valuation, *International Seminar on Business and Information Management*, 461-464.

Davis, J. (1997), Maintaining Customer Relationships Through Effective Database Marketing: A Perspective For Small Retailers, *Journal Of Marketing Theory & Practice*, **5** (2), 31-44

Derks, P. R. (1994) Business Trends: The competitive advantages of database marketing, *Information Strategy*, **10** (3), 5-11

Dowling, R. G. and Uncles, M. (1997) Do customer loyalty programs really work? *Sloan Management Review*, **38** (4), 71-82.

East, R., Harris, P., Willson, G., Lomax, W. and Perkins, D., (1997) Firststore loyalty to US and British supermarkets. In: *Proceedings of the 26th European Marketing Academy Conference*, Warwick, **1** , pp. 401-415.

Fader, S. P., Bruce, G.S. H. and Lee, L. K. (2005) RFM and CLV: Using Iso-Value Curves for Customer Base Analysis, *Journal Of Marketing Research*, **42**, 415-430.

Flavian, C., Martinez, E. and Polo, Y. (2001) Loyalty to grocery stores in the Spanish market of the 1990s, *Journal of Retailing and Consumer Services*, **8** (2), 85-93

Forenell, C. (1992), "A Customer Satisfaction Barometer: The Swedish Experience", *Journal of Marketing*, **56** (2), 6-21.

Fox, T. and Stead, S. (2001) *Customer relationship Management: Delivering the benefits*, CRM (UK) and SECOR Consulting, New Malden.

Goldman, A., (1978). The shopping style explanation for store loyalty. *Journal of Retailing*, **53** (Winter), 33-46.

Gremler, D. D. and Brown, W. S. (1999),The loyalty ripple effect: Appreciating the full value of customers, *International Journal of Service Industry Management*, **10** (3), 271 – 293

Helgesen, O. (2006): Are Loyal Customers Profitable? Customer Satisfaction, Customer (Action) Loyalty and Customer Profitability at the Individual Level, *Journal of Marketing Management*, **22 (**3-4), 245-266.

Hsieh NC (2004). An integrated data mining and behavioral acoring model for analyzing bank customers. Expert Syt Appl., 27: 623-633.

Hughes, A. M. (1996) Boosting response with RFM, *Marketing Tools*, **5**, 4-10

Hughes, A. M. (2012) *Strategic Database Marketing*, 4th Edition, McGraw-Hill, New York.

Hughes, A.M., (1999). How safeway built loyalty. *Target Marketing*, **22** (3), 46–49.

Humby, C., Hunt, T. and Philips, T. (2011) *Scoring Points. How Tesco continues to win customer loyalty*, 2nd edition, Kogan Page, London.

Jackson, R. B. and Wang, P. (1994) *Strategic Database Marketing*, Lincolnwood, Ill: NTC Books.

Jain, R. and Bagdare, S. (2011) CRM in Retailing: A Behavioural Perspective. *Journal of Marketing & Communication*, **7** (2), 31-37

Jones, O., T. and Sasser, W. E. (1995) Why satisfied customers defect, *Harvard Business Review*, Novermber-December, 88-99.

Jutkins, R (1994) Just imagine! Database marketing targets the right customers – and keeps them coming back, *Direct Marketing*,**12** (56), 38-40

Kahan R (1998). Using database marketing techniques to enhance your one-to-one marketing initiatives. J. Consum. Mark., 15(5): 491-493.

Kahn, E. B., Kalwani, U. M. and Morrison, G. D. (1986) Measuring variety-seeking and reinforcement behaviours using panel data, *Journal of Marketing Research*, **23** (2).

Kaymak, U. (2001) Fuzzy target selection using RFM variables, IFSA World Congress 20th NAFIPS International Conference, 2, 1038-1043.

Kuehn, A. (1962) Consumer brand choice as a learning process. *Journal of Advertising Research*, **2**, 10-17

Kumar, V. and Shah, D. (2004) Building and Sustaining Profitable Customer Loyalty for the 21st Century. *Journal of Retailing*, **80** (4), 317-330

Lee, J., (1996) Service with a smile. *Marketing, London*, November 28, 46

Lewington, J., De Chernatony, L. and Brown, A. (1996) Harnessing the Power of Database Marketing, *Journal of Marketing Management*, **12** (4), 329-346

Liddy, A. (2000) Relationship marketing, loyalty programmers and the measurement of loyalty, *Journal of Targeting, Measurement Analysis for Marketing*, **8** (4), 351-362.

Lipstein, B. (1959) The dynamics of brand loyalty and brand switching. In: Proceedings of the fifth annual conference on the advertising research foundation, New York: Advertising Research Foundation, 101-108.

Lumsden SA, Beldona S and Morison AM (2008). Customer value in an all-inclusive travel vacation club: An application of the RFM framework. *J. Hosp. Leisure Mark.*, **16** (3), 270-285

Ma, B., Li, F., Wang, G. and Ran, L. (2009) Targeting Valuable Customers within a Retail Reward Program Database by RFMG Model," *Management and Service Science, 2009. MASS '09. International Conference*, 1-4,

Macintosh, G. and Lockshin, S. L. (1997) Retail relationships and store loyalty: A multi-level perspective, *International Journal of Research in Marketing*, **14** (5),487-497

Magi, W. A. (2003) Share of Wallet in retailing: the effects of customer satisfaction loyalty cards and shopper characteristics, *Journal of Retailing*, **79** (2), 97-106

Marcus C (1998). A practical yet meaningful apprach to customer segmentation. J. Consum. Mark., 15(5): 494-504.

Massey, F. W., Montgomery, B. D. and Morrison, G. D. (1970) *Stochastic Models of Buyer Behaviour*, MIT Press, Cambridge

Mauri, C. (2011) Card Loyalty. A New Emerging Issue In Grocery Retailing. SDA Bocconi, Research Division, Working Paper No. 01-42. Available at SSRN: http://ssrn.com/abstract=271768 or http://dx.doi.org/10.2139/ssrn.271768

McCarty JA and Hastak M (2007). Segmentation approaches in data-mining: A comparision of RFM, CHAID, and logistic regression. *Journal of Business Research*, 60, 656-662.

McGoldrick, P.J., Andre, E., (1997). Consumer misbehaviour: promiscuity of loyalty in grocery shopping. *Journal of Retailing and Consumer Services*, **4** (2), 73-81

Miglautsch JR (2002). Application of RFM principles: What to do with 1-1-1 customers? J. Database Mark., 9(4):319–324.

Miglautsch, J. R. (2000) Thoughts on RFM scoring. *Journal Of Database Marketing*, **8** (1), 67-72.

Min, H. (2006) Developing the Profiles of Supermarket Customers through Data Mining, *The Service Industries Journal*, **26** (7), 747-763.

Mittal, V., Anderson,E., Sayrak, A. and Tadikamalla, P. (2005), Dual Emphasis and the Long-Term Financial Impact of Customer Satisfaction, *Marketing Science*, **24** (4), 544–55

Oliver, R. L. (1999) Whence consumer loyalty? *Journal of Marketing*,**63**, 33-44.

Pan,Y. and Zinkhan, G. M. (2006) Determinants of retail patronage: A meta-analytical perspective, *Journal of Retailing*, **82**(3), 229-243

Partch, K., (1999). For most, loyalty cards haven't changed a thing. *Supermarket Business*, **54** (7), 114

Passingham, J. (1998) "Grocery Retailing and the Loyalty Card: The Journal of the Market Research Society" *The Journal of the Market Research Society*, 40 (1), 55-63.

Pauler, G. and Dick, A. (2006) Maximizing profit of a food retailing chain by targeting and promoting valuable customers using Loyalty Card and Scanner Data. *European Journal of Operational Research*, **174**, 1260-1280.

Pritchard, M. P., & Howard, D. R. (1997). The loyal traveller: examining a typology of service patronage. Journal of Traveller Research, 35 (4), 83-95.

RAI Consultants (2011) *Retail Census 2011 and Impact of economic crisis on retail market*. Nicosia: RAI Consultants

RAI Consultants (2012) *Supermarket Shares 2010-2011*. Nicosia: RAI Consultants

Rayner, S. (1996) Customer Loyalty Schemes. Effective implementation and management, FT Management Report, FT Retail & Consumer Publishing Pearson Professional, London.

Reichheld, F. F. (1996) *The loyalty effect: The hidden force behind growth, profits and lasting value*, Boston: Harvard Business School press.

Reichheld, F. F. (2003) The one number you need to grow, *Harvard Business Review*, **81** (12), 46-54.

Reinartz, W. J. & Kumar, V. (2000) The Profitability of long-life customers in a noncontruactual setting: An empirical investigation and implications for marketing. *Journal of Marketing*, **64** (4), 17-35.

Reinartz, W. J. & Kumar, V. (2002) The mismanagement of customer loyalty. *Harvard Business Review*, **80** (7), 86-94

Reinartz, W. J. & Kumar, V. (2003) The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing*, **63** (1), 77-99

Rosenberg, L. J and Czepiel, J. A. (1984) A marketing approach for Customer Retention, *Journal of Consumer Marketing*, **1** (2), 45-51.

Rust, RT and Verhoef, PC (2005) Optimizing the marketing interventions mix in intermediate-term CRM, *Marketing Science*, **24** (3), 477-489.

Ryals, L. and Payne, A. (2001) 'Customer relationship management in financial services: towards information-enabled relationship marketing', *Journal of Strategic Marketing*, Vol. 9, No. 1, pp. 3–27.

Schoenbachler, D. D., Gordon, L. G., Foley, D. and Spellman L. (1997) Understanding consumer database marketing, *Journal of Consumer Marketing*, **14** (1), 5-19

Schultz, Don E. and Wang, Paul (1993) Database Marketing Applications: An industry Survey of Traditional Direct Marketers in the United States, *The Journal of Database Marketing*, **1** (2), 113-129.

Shankar, V., Smith, A. K., and Rangaswamy, A. (2003) Customer satisfaction and loyalty in online and offline environments, *International Journal of Research in Marketing*, **20** (2), 153-175

Sharp, B and Sharp, A. (1997) Loyalty Programs and their impact on repeat-purchase loyalty patterns, *International Journal of Research in Marketing*, **14** (5), 473-486.

Shih, Y. and Liu, C. (2003) A method for customer lifetime value ranking – Combing the analytic hierarchy process and clustering analysis, *Database Marketing & Customer Strategy Management*, **11** (2), 159-172.

Shoemaker S. and Lewis, R. (1999) Customer Loyalty: The future of hospitality marketing, *Hospitality Management*, 18 (4), 345-370

Sirohi, N., McLaughlin, E.W., and Wittink, D.R. (1998) Model of consumer perceptions and store loyalty intentions for a supermarket retailer, *Journal of Retailing*, **74** (2), 223-255.

Smith, R., (1998). Can you bribe your way to customer loyalty? Frequency marketing strategies. *Strategic Research Institute*, New York.

Spring, P., Leeflang, P. and Wansbeek (1999) The Combination Strategy to Optimal Target Selection and Offer Segmentation in Direct Mail, *Journal of Market Focused Management*, 4, 187–203

Srinivasan, R and Moorman, C. (2005) Strategic firm commitments and rewards for customer relationship management in online retailing, *Journal of Marketing*, **69** (4), 193-200

Stone, M. and Shaw, R. (1987), Database Marketing for Competitive Advantage, *Long Range Planning*, **20** (2), 12-20.

Tepeci, M. (1999) Increasing brand loyalty in the hospitality industry, *International Journal of Contemporary Hospitality Management*, **11** (5), 223-9

Toumazou, O. (2012) What mews in the rood? (Τι κάνει νιάου νιάου στα κεραμύδια; Written in Greek.), In: 9th Retail Forum, Nicosia, Cyprus, Hilton Park. Accessed at 25 July 2012 from < http://www.imh.com.cy/assets/files/CONFERENCES%202012/OLYMPIOS%20TOUMAZOU.pdf>

Venkatesan, R. and Kumar, V. (2004) A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing*, **68** (4), 106-125.

Verhoef, C. P., Spring, N. P., Hoekstra, C. J. and Leeflang, S.H. P. (2002) The commercial use of segmentation and predictive modelling techniques for database marketing in the Netherlands, *Decision Support Systems,* **34**, 471-481.

Wang, CH (2010) Apply robust segmentation to the service industry using kernel induced fuzzy clustering techniques. *Expert Systems with Applications*, 37, 8395-8400.

Wehmeyer, K. (2005) Aligning IT and marketing —The impact of database marketing and CRM, *Database Marketing & Customer Strategy Management*, **12** (3), 243-256

Wei, J., Lin, S. and Wu, H. (2010) A review of the application of RFM model. *African Journal of Business Management*, **4** (19), 4199-4206.

Yeh IC, Yang KJ and Ting TM (2008). Knowledge discovery on RFM model using Bernoulli sequence. *Expert Systems with Applications*, 36: 5866-5871.

Yi, Y. and Jeon, H. (2003) Effects of loyalty programs on value perception program loyalty, and brand loyalty, *Journal of the Academy of Marketing Science,* **31** (3), 229-240

APPENDIX

Table 9 Frequency score * Monetary score * Recency score Crosstabulation

Recency score			Monetary score 1	2	3	4	5	Total
1	Frequency score	1	2235	477	65	8	1	2786
		2	427	669	291	79	7	1473
		3	78	376	377	256	53	1140
		4	15	135	326	337	179	992
		5	2	18	164	291	399	874
	Total		2757	1675	1223	971	639	7265
2	Frequency score	1	1974	547	75	6	0	2602
		2	518	1118	515	112	13	2276
		3	70	408	505	270	58	1311
		4	6	99	225	242	121	693
		5	0	18	45	113	143	319
	Total		2568	2190	1365	743	335	7201
3	Frequency score	1	847	327	77	5	3	1259
		2	348	848	586	187	27	1996
		3	58	419	811	571	165	2024
		4	8	120	388	637	391	1544
		5	0	8	84	216	353	661
	Total		1261	1722	1946	1616	939	7484
4	Frequency score	1	155	69	10	1	1	236
		2	86	284	164	67	4	605
		3	28	231	365	310	93	1027
		4	6	71	288	448	337	1150
		5	0	7	62	238	390	697
	Total		275	662	889	1064	825	3715
5	Frequency score	1	229	116	19	2	0	366
		2	107	390	283	95	12	887
		3	40	331	626	548	199	1744
		4	6	126	589	1115	1027	2863
		5	0	31	304	1089	3267	4691
	Total		382	994	1821	2849	4505	10551
Total	Frequency score	1	5440	1536	246	22	5	7249
		2	1486	3309	1839	540	63	7237
		3	274	1765	2684	1955	568	7246
		4	41	551	1816	2779	2055	7242
		5	2	82	659	1947	4552	7242
	Total		7243	7243	7244	7243	7243	36216

Figure 3 – All customer segments (125)

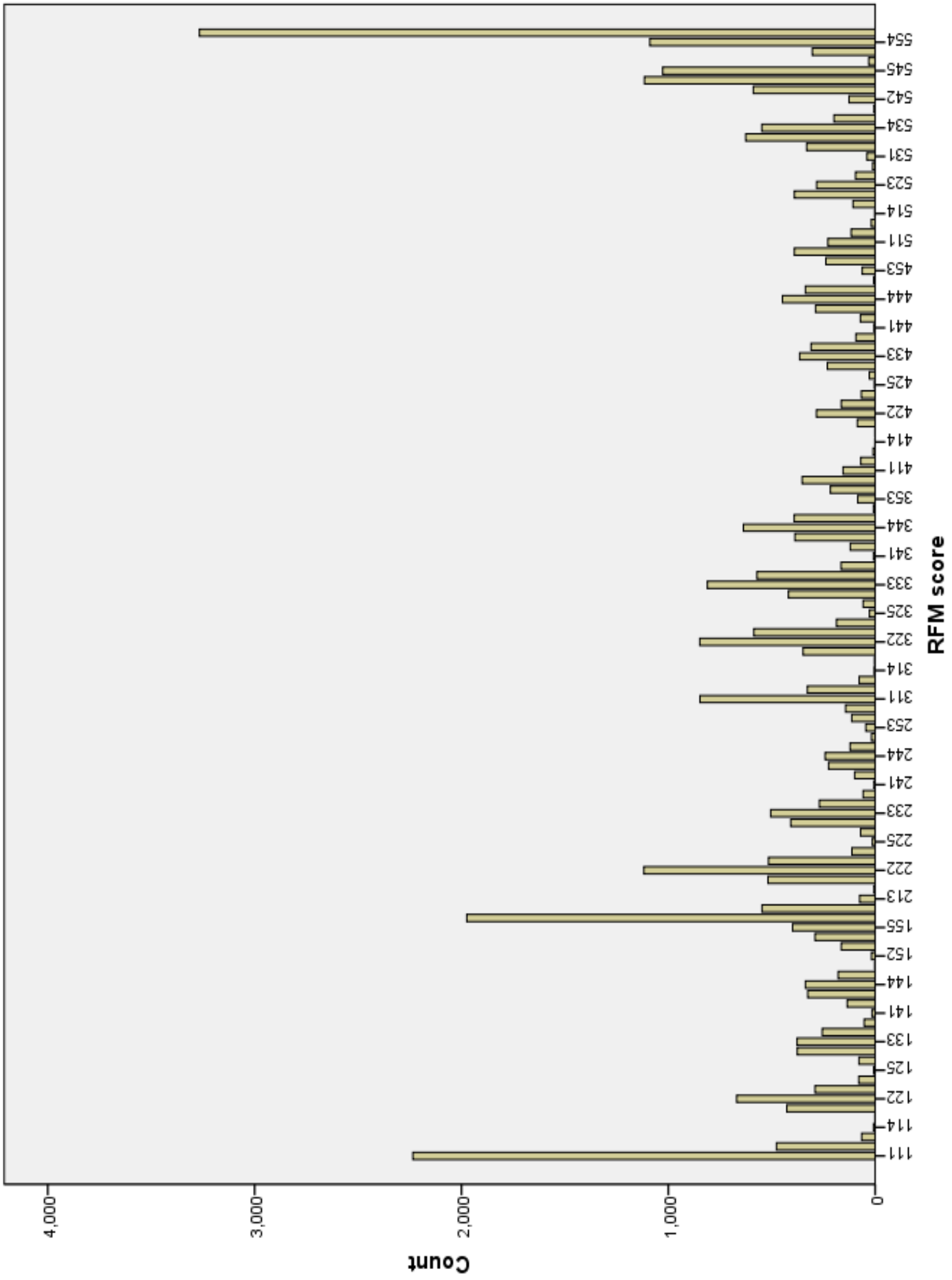


Figure 4. RFM Results for the most important segments