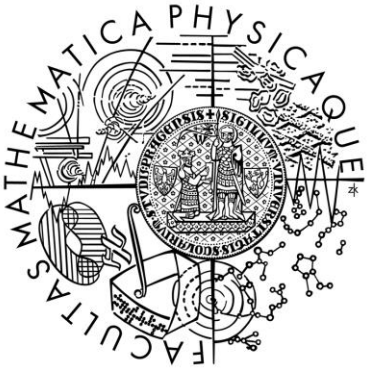
Charles University in Prague

Faculty of Mathematics and Physics

MASTER THESIS



Norbert Hanuska

Factors influencing customer profitability: an empirical examination in noncontractual settings

Department of Software Engineering

Supervisor of the master thesis: Ing. Vladimír Kyjonka

Study programme: Informatics

Specialization: Software systems

Prague 2014

I would like to thank the company Total Internet Group for providing the data and I also thank greatly supervisor Vladimír Kyjonka for his help to complete this thesis.

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague, 30.7.2014 signature

Název práce: Faktory ovlivňující profitabilitu zákazníka: empirický výzkum v nesmluvním prostředí

Autor: Norbert Hanuska

Katedra / Ústav: Katedra softwarového inženýrství

Vedoucí diplomové práce: Ing. Vladimír Kyjonka, Katedra softwarového inženýrství

Abstrakt: Porozumění, jak řídit vztah se zákazníky, se stalo důležitým tématem pro akademiky i praktiky v posledních letech. Efektivita a profitabilita obchodních vztahů může být do značné míry zlepšena identifikací hnacích faktorů nejvíc profitabilních zákazníků a jejich použití pro marketingové cílení na dané zákazníky. V této studii identifkujeme charakteristiky jako průměrná útrata během jedné obchodní transakce, délka vztahu zákazníka s firmou, úroveň nakupování různých produktových kategorií a také demografické charakteristiky jako věk a pohlaví jako důležité faktory nejlepších zákazníků. Výsledky této studie mají relevantní dopady jak pro akademiky zkoumající charakteristiky nejprofitabilnějších zákazníků v prostředí, kde zákazník není vázán kontraktem, tak pro praktiky, kterým výsledky můžou pomoct při vytváření efektivních marketingových strategií. Navíc výsledky procesu získávání vědomostí o zákaznících prostřednictvím různých data miningových technik přispívá k identifikaci vhodnosti použití těchto metod.

Klíčová slova: RFM analýza, data mining, profitabilita zákazníka, rozhodovací stromy

Title: Factors influencing customer profitability: an empirical examination in noncontractual settings

Author: Norbert Hanuska

Department: Department of software engineering

Supervisor: Ing. Vladimír Kyjonka, Department of Software Engineering

Abstract: Understanding of how to manage relationships with customers has become an important topic for both academic and practitioners in recent years. The effectiveness of business can be greatly improved by identifying the drivers of the most profitable customers and using them to target the right customers. In this study we identify exchange characteristics such as amount of money spent per purchase, customer relationship duration with firm, ratio of cross-buying and demographic characteristics such as age and gender as important drivers of the most profitable customers. The results of the study have important implications for academicians in understanding what drives the most profitable customers in noncotractual settings as well as practitioners to help design more effective marketing strategies. Moreover, the results of knowledge discovery about customers by different data mining techniques also contribute to help researchers identifying feasibility of these methods.

Keywords: RFM analysis, Data mining, Customer profitability, Decision trees

Contents

Introduction

An understanding of how to manage relationships with customers has become an important topic for both academic and practitioners in recent years (Reinartz, Krafft, Hoyer, 2004). As no business can exist without generating customer revenue, management thinking progresses and focuses more on customers. In fact, experiments conducted by Kumar, Venkatesan and Reinartz (2008) have shown that adopting a customer-focused sales campaign can significantly increase profits and return of investments. In addition, a customer-focused sales campaign can improve the quality of the relationship between the customer and the firm (Kumar, Venkatesan and Reinartz, 2008).

Companies have already recognized that they can increase profits by acknowledging that different groups of customers vary widely in their behavior, desires, and responsiveness to marketing (Zeithaml, 2001). Anderson and Narus (1991) note that every industry is characterized by its own transactional and relational changes, and moreover, customers with different characteristics may have different levels of repurchase behavior. It is very important for competitive industries to identify and retain customers with high value or profit potentials (Niraj, Gupta, Narashin, 2001; Chang et al. 2007, Chiliya et al. 2009, Mutandwa et al. 2009) and then to customize marketing strategies for those customers (Chang, 2010). Firms have discovered that customers are different and treating valuable or potentially valuable customers and limit investments to non-valuable is more effective and they need not serve all customers equally well (Zeithmal, 2001).

In order to manage, support and retain profitable customers and relationships, companies are turning more and more to Customer Relationship Management (CRM) techniques as a strategy that integrates the concepts of Knowledge Management, Data Mining and Data Warehousing (Cunningham et al., 2004). Yeh et al. (2009) stated that the concept of CRM is to acquire and retain the most profitable customers by understanding their values. By adopting a Recency, Frequency and Monetary (RFM) model, decision makers can identify these valuable customers effectively and then develop right marketing strategy (Wei, Lin, Wu, 2010). However, one of drawbacks of RFM is the focus on a company's current customers, and that it cannot be applied to prospecting for new customers as a marketer does not have transactions for prospects (McCarty and Hastak, 2007). That is why the understanding of what exactly drives customers in becoming profitable is essential in

many areas of decision making such as marketing decisions of customer acquisition and customer retention (Gupta, Lehman, 2001; Zeithaml 2000).

It is difficult to identify existing customers in terms of profitability, but it is even more difficult to identify potential customers who will ultimately be the best customers. According to Zeithmal (2000, p. 80),"developing a profile of the potentially profitable customer requires a combination of understanding existing profitable customers, delineating demographic and psychographic variables that predict profitability and creating and testing strategic approaches to obtain and qualify customers". To date, little empirical work has been reported on this topic (Zeithaml, 2000; Reinartz and Kumar, 2003). But because of increased managerial interest in how to manage the customer-firm relationship (Reinartz and Kumar, 2003), researchers have focused more on empirically measuring and modeling a customer's relationship with a firm (Bolton, 1998). Next, companies point out, that competitive situations in their industries lead to hesitation about validity of the research in different contexts and if it applies to them. Research how and if the relationship between customer characteristics and profitability vary by industry, country or category of business and remain to be conducted (Zeithaml, 2000). The situation is more difficult in noncontractual settings where the relationships between buyers and sellers are not governed by a contract, because a customer purchases completely at his or her discretion. Thus, our study would like to contribute by extending the research of antecedent variables that can characterize profitable customers in noncontractual settings. Database technology and data mining as the discovery of 'interesting,' non-obvious patterns hidden in a database that have a high potential for contributing (Peacock 1998) - are useful here (Zeithaml, 2000).

Research Objective

The main purpose of this research is to understand drivers of profitable customer relationships and empirically test the factors that affect becoming most profitable customers. Constructs from different scientific perspective will be presented and most profitable customers will be identified by adapting widely used RFM framework (Birant, 2011; Reinartz, Kumar, 2003). This aims to extend the current knowledge about factors that influence profitable customers. In addition, our goal is to demonstrate data mining techniques for predictive analytics by proposing predictive model and also to develop managerial implications for building and managing profitable customer relationships.

Problem statement

Presented motivation leads us to central research question:

What are the most influential drivers of profitable customers in noncontractual settings?

From the central research question, several sub-questions can be derived that will be addressed in the thesis:

- What drives profitability and how can we identify profitable customers?
- What are the customer characteristics and how they are related to profitability of a customer?
- How can we predict becoming the most profitable customer and which data mining techniques are useful for the purpose?

Scientific contribution

Studies contributed to customer characteristics and their relationship to customer's loyalty, satisfaction and profitability usually focus on certain industry - insurance, baking, retail or mobile commerce (Bolton, 1998; Reinartz and Kumar 2000, 2003). But there is very little research concerning how various factors of profitable customers differ across industries, especially in non-contractual settings (Reinartz and Kumar, 2003, Zeithaml, 2000).

RFM analysis is a method mostly used to identify high-response customers in marketing promotions, and to improve overall response rates, which is well known and is widely applied today (Wei, Lin, Wu, 2010). Less widely understood is the value of applying RFM scoring to a customer database and measuring customer profitability (Aggelis, 2004). In recent years, data mining applications based on RFM concepts have been proposed for different contexts such as the computer security, automobile industry and for the electronics industry (Birant, 2011). Our research extends application of traditional RFM analysis concept in new sectors – Webshop and Theater based in the Netherlands as these represent noncontractual product and service industry.

Investigation is supported by Data Warehousing and Data Mining techniques, which gives the opportunity to contribute in the field of Customer Relationship Management analysis by applying the general knowledge in specific environment (Cunningham, 2004).

Social contribution

Marketing managers can make use of the results as describing the most profitable customers and develop effective strategies to retain them, acquire more like them and improve overall profitability. It also helps to better understand the structure of profitable relationships and to deduce implications to better manage a customer's profitable tenure with the company.

Scope

The focus of this thesis will be on profitable customers and their characteristics in two noncontractual settings – Webshop selling pharmaceutical products and Dutch Theater. These contexts were chosen because they represent product and service orientation of business. Webshop selling goods represents product orientation as products are tangible items sold to customers who buy them off the shelf, from the lot or out of the catalog (Pine, Gilmore, 1999). On the other hand, services industry represents Theater as intangible activities performed for a particular client (Pine, Gilmore, 1999).

Research methodology

An extensive literature review is performed in order to gain insight and develop the hypotheses and to model factors that might influence becoming the most profitable customer. To answer the research question, the data from Webshop and Theater are used to empirically validate hypotheses. From the significantly large databases there are extracted relevant data and Data Warehouses developed. Results of the study are based on quantitative assessment of the data. We employ segmentation by RFM analysis and Partial Least Squares (PLS) modeling to estimate conceptual model. Next, data mining techniques for predictive modeling are utilized (regression analysis, decision trees) as well as descriptive analysis (association rules mining).

Structure of the thesis

The remainder of this study is structured as follows; first we introduce the basic concepts of Customer Relationship Management. Literature on customer profitability determination

will be discussed alongside with exploring existing theory and applications of RFM analysis. Next, use of RFM model for assessment of customer profitability will be explained. In addition, it will review data mining techniques for Customer Relationship Management analyses. Next, the conceptual research model will be presented alongside with hypotheses development. Then we will describe the research method used in this research, data gathering and pre-processing. Results of the study are presented and analyzed in the following section. This study will conclude with theoretical and practical implications of the research findings and will identify the limitations and possibilities for further research.

1. Theoretical Framework

This chapter will focus on the theoretical foundations of the research model of this study. First, the basics of the Customer Relationship Management will be presented. Then the concept of RFM analysis for definition and identification of profitable customers will be discussed, as well as other applications of the RFM model. Next, customer characteristics and factors that influence profitability of the customer will be elaborated on. We will also focus on concept of Data Mining techniques for analysis of customer and relationship to different types of businesses. For research purposes of this study two non-contractual settings were chosen and will be presented. Finally, several hypotheses will be developed and consequent research model depicted.

1.1 Customer Relationship Management

Customer Relationship Management has received an increasing amount of interest among academics and practitioners in recent years (Kannan, Rao, 2001; Kerstetter, 2001) and customer profitability becomes the main principle, core activity of CRM (Hawkes, 2000). Customers are being more demanding, sophisticated and selective and together with emergence of new technologies it creates new possibilities for companies to manage customer relations and how to interfere with their customers. Customer databases and the analysis of the data are essential parts of CRM (Shaw et al., 2001) and statistical techniques for providing those analyses and thus providing information for marketing decisions are essentials of database marketing and data mining (Petrison, Blattberg, Wang, 1993).

The purpose of CRM is to improve marketing productivity, which is achieved by increasing marketing efficiency and by enhancing marketing effectiveness (Parvatiyar & Sheth, 2001) and to forge closer and deeper relationship with customers. Simply put, focus of CRM is: “Attract new customers, know them well, give them outstanding service, and anticipate their wants and needs. When doing these things well, increased revenue and profits are likely to follow.” (Goodhue, Wixom & Watson, 2002, p.80)

Many companies gained significant benefits from their CRM initiatives. A few examples include American store chain, Lowe's Home Improvement Warehouse, which managed to achieve a 265 percent return on investment on its $11 million CRM investment (Stringefellow, Nie, & Bowen, 2004) and Harrah's Entertainment, which by mining the company's rich database to develop compelling customer incentives managed to turn its customers into the most devoted in the casino industry. However, not every CRM project is successful. As Gartner study has shown, 55 % of all CRM projects are not delivering results. Analysts of CRM failure have suggested that many CRM strategies fail because of implementation problems (Gartner, 2003).

CRM means different things to different people and it has been implemented in different ways. For the scope of this thesis, we contend that the strategic view provides the best foundation, since strategy can be defined as an "overall plan for deploying resources to establish a favorable position" (Grant, 1998, p.14), and CRM can be seen as a way to achieve competitive advantage (Reinartz & Kumar, 2003). Moreover, CRM is about identifying a company's best customers and maximizing the value from them by satisfying and retaining them and the tools and technologies of data warehousing and data mining enable new opportunities.

Understanding the most profitable customers is essential to retain customers (Hawkes, 2000).

Enterprises can shorten sales cycle and increase customer loyalty to build better close relationships with customers and further add revenues by good CRM. Thus, an excellent CRM can help enterprises keeping existing customers and attracting new ones (Cheng, Chen, 2009)

Moreover, enterprises also strengthen marketing and sales effectiveness in order to build good CRM. Kalakota and Robinson (1999) explained that the CRM is to integrate the function of the related fields with customer in the enterprise such as marketing, sales, services and technical support for customer needs, and it usually utilizes IT to help an enterprise managing relationships with customer in a systematic way, improving customer loyalty and increasing overall business profits (Cheng, Chen, 2009).

Data mining is a powerful new technique to help companies mining the patterns and trends in their customer's data, then to drive improved customer relationships, and it is one of

well-known tools given to customer relationship management - CRM (Cheng, Chen, 2009).

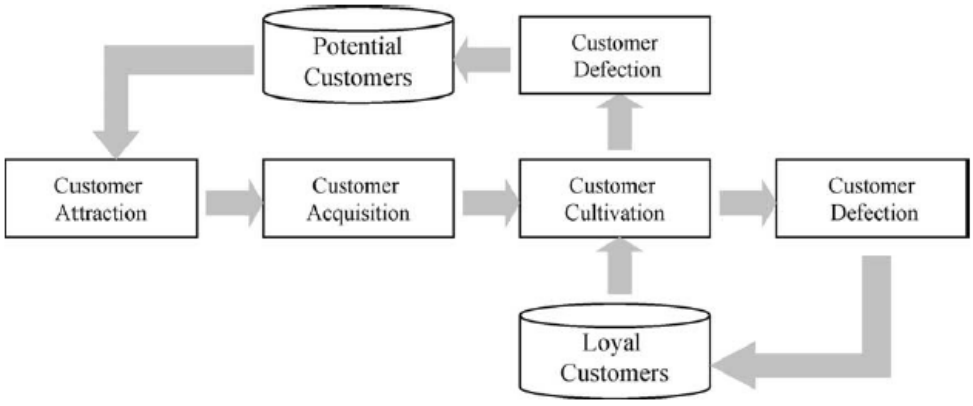


Figure 1: Scope of CRM according to Hwang, Jung, Suh (2004).

Because attracting new customers is expensive, retaining them is quite important and thus, the objective of CRM is to maximize customer lifetime value to the organization. Traditional CRM is about analyzing customer information for business decision with aim being to help organization understand customers' needs (Shen, Chaung, 2009). Differences between customers can be determined by market segmentations that imply that some customers are more profitable than others - it performs analysis of customer loyalty and profitability (Hughes, 1996). Customer relationship management (CRM) is one of the tools of business operation for acquiring and retaining customers, increasing customer value, loyalty and retention, and implementing customer-centric strategies. CRM, devoted to improve relationships with customer, focuses on a comprehensive picture on how to integrate customer value, requirements, expectations and behaviors via analyzing data from transaction of customer. And very effective method of customer segmentation is Recency, Frequency, Monetary (RFM) analysis (Newell, 1997).

1.2 RFM analysis, Customer Lifetime Value and Customer Profitability

Decision makers in managing customer relationship always wanted to project profitability of customers. RFM (Recency, Frequency and Monetary) model has been widely applied in many practical areas, particularly in direct marketing (Wei, Lin, Wu, 2010; McCarty and Hastak, 2007). Basically it is as an analytical technique allowing direct marketers to better segment their customer databases. RFM model is a behavior-based model used to analyze

behavior of a customer and then make predictions based on the behavior in the database (Hughes, 1996; Yeh et al., 2009). In spite of availability of more statistically sophisticated methods, a study by Verhoef et al. (2002) has shown that RFM is the second most common method used by direct marketers, after cross tabulation. There are several reasons for such popularity. In general, RFM analysis can be implemented very quickly and is easy to use (Kahan, 1998). Furthermore, it is a method that managers and decision makers can understand (Marcus, 1998). It is important for successful method to differentiate customers and do it in way that is easy to explain to decision makers.

1.2.1 Definition and scoring scheme

RFM model is the most frequently adopted segmentation technique that comprises three purchase related measures (recency, frequency and monetary) (Wei et al. 2010), which are combined into a three-digit RFM cell value, covering five equal quintiles (20%).

Recency refers to the interval between the last purchase and analyzed time. Usually it is number of days (weeks, months, years) from the latest consuming behavior of a customer to present. Recency is often regarded as the most important measure from those three, because the most recent purchasers are likely to purchase again, so they potentially create future value. However, according to prior findings this can be very firm-specific (Lumsden et al. 2008; Fader et al. 2005).

Frequency is the number of transactions / purchases made by the customer within a certain period of time. Higher frequency score indicates greater customer loyalty and greater demand for the products, because they are purchased repeatedly.

Monetary can represent the cumulative total of money spent by a customer within specific period of time or average dollar amount of all purchases (Hughes, 1996). According to Marcus (1998) it is better to use the average purchase amount as to reduce co-linearity of frequency and monetary. This measure gives us ability to identify customers, who are purchasing the most expensive goods.

The process to quantify customers via RFM is as follows. The segmentation starts with sorting database according to each dimension of RFM. It begins with sorting Recency - the customers are split into equal quintiles (20%) and given a score - the lower the recency, the higher the score. Top 20% of the customers are given a score of 5, the next 20% score 4 and so forth, until the last segment is coded 1. Example dataset is shown in Table 1 and the process of awarding score is illustrated in Table 2. Then the customers are sorted and

scored for frequency – from the most frequent customers to the least. Same process is undertaken also for monetary value. Finally, all customers are ranked by concatenating R, F and M values, which will give us (5x5x5) 125 segments – customers ranked 555, 554, 553, … , 111.

Customer ID	Recency (Days)	Frequency (Number of purchases)	Monetary (Average purchase value)
1	3	6	540
2	6	10	940
3	45	1	30
4	21	2	64
5	14	4	169
6	32	2	55
7	5	3	130
8	50	1	950
9	33	15	2460
10	10	5	190
11	5	8	840
12	1	9	1410
13	24	3	54
14	17	2	44
15	4	1	32

Table 1: An example dataset: customer transactions (Adapted from Birant, 2011)

In the Table below Customer quintiles and RFM values can be found.

ID	Rec.↓	R	ID	Freq.↑	F	ID	Mon.↑	M	ID	RFM
12	1	5	9	15	5	9	2430	5	1	544
1	3	5	2	10	5	12	1410	5	2	454
15	4	**5**	12	9	5	8	950	5	3	111
7	5	4	11	8	4	2	940	4	4	222
11	5	4	1	6	4	11	840	4	5	333
2	6	4	10	5	4	1	540	4	6	222

10	10	3	5	4	3	10	190	3	7	433
5	14	3	7	3	3	5	169	3	8	115
14	17	3	13	3	3	7	130	3	9	155
4	21	2	14	2	2	4	64	2	10	343
13	24	2	4	2	2	6	55	2	11	444
6	32	2	6	2	2	13	54	2	12	555
9	33	1	**15**	1	**1**	14	44	1	12	232
3	45	1	3	1	1	**15**	32	**1**	14	321
8	50	1	8	1	1	3	30	1	**15**	**511**

Table 2. Customer quintiles and RFM values

RFM provides a simple framework for quantifying customer behavior. For example, it is possible to infer from Table 2 that customer with id 9, which has RFM score 155, has made a high number of purchases with high monetary values but not for a long time. Something might have gone wrong with this customer, for example, he/she has most likely defected to a competitor's products and services or has found an alternate source and that is why his/her recency score is low. At this situation, marketers can contact with this customer and get feedbacks about how to do it better because he/she is one of the valuable customers according to his frequency and monetary values. Moreover, it is possible to plan a customer reactivation program and send him/her an extreme promotion in an effort to get his/her attention.

Researchers state that the RFM model is one of the well-known customer value analysis methods, where its advantage is to extract characteristics of customers by using fewer criterions (a three-dimension) as cluster attributes so that reduce the complexity of model of customer value analysis. Moreover, from view of the consuming behavior, RFM model is a long-familiar method to measure the strength of customer relationship. Retention cost is far less costly than acquisition cost to acquire a new customer therefore, enterprises utilize RFM analysis to mine databases for knowing about customers who spend the most money and create the biggest value for enterprises (Cheng, Chen, 2009).

Several studies considered different variations of RFM analysis. The “quintile method” (Miglautsch, 2000) we described sorts customers in descending order from the best customer to worst. Its advantage is that the number of customer in each quintile is same

(20%); the segments have the same size. However, the final RFM cells differ in size from each other. Method known as hard-coding (McCarty and Hastak, 2007) scores customers according to their behavior so there can be different number of customers in the quintile. In practice, marketers decide thresholds for each segment in RFM dimension based on their knowledge and experience. For instance, coding intervals for recency might be based on marketer's preferences, history etc. as follows: 0 to 3 months, 3 to 6 months, 6 to 12 months, 12 to 24 months and over 24 months, coded 5,4,3,2,1 respectively. Also weighted approach to RFM scoring was proposed (Miglautsch, 2000). It consists of assigning weights to each of RFM measures and then adding the numbers to come up with final RFM score. The formulas of adding the weights to each RFM dimension can be quite different – Hughes (1994) suggests to use the same weights, so the cell (5, 3, 5) would end up after adding with composite score 13, Miglautsch (2000) indicated formula for summing the score as follows – (R x 3) + (F x 2) + (M x 1), but this summing formula can be product or industry specific (Stone, 1994). Assigning a weight to each of the RFM measure will become based on the past experience and the judgment of the particular database and marketer. That's why this approach is sometimes referred as judgment based RFM (McCarty and Hastak, 2007).

1.2.2 Applications of RFM

RFM is based on 80/20 pareto principle that 20% of the customers bring 80% of the revenue. Model is used to segment file (i.e. company's current customers) using information related to recency, frequency and monetary. It is not applicable to the prospecting for new customers, because a marketer would not have transactions information for prospects (McCarty and Hastak, 2004). Integration of RFM analysis and data mining techniques provides useful information for current and new customers. Rules discovered from customer demographic variables and RFM variables provide useful knowledge to predict future customer behavior. Because model measures when people buy, how often they buy and how much they buy and according to Swearingen (2009) the best predictor of future customer's behavior is past customer's behavior, RFM discovers customers that are profitable now, and also potentially profitable in the future.

Several studies incorporated RFM model to express Customer lifetime value (Liu and Shih, 2005; Sohrabi and Khanlari, 2007; Miglautsch 2000). Customer lifetime value

(CLV) is typically used to identify profitable customers and to develop strategies to target customers (Irvin, 1994) and is generally defined as the present value of all future profits obtained from a customer over his or hers relationship with a firm (Sohrabi, Khanlari, 2007). The previous researches contain several definitions of CLV but there is small difference between them (Hwang, Jung, Suh, 2004). Some of the definitions are presented below:

- Expected profits from customers, exclusive of costs related to customer management
- The total discounted net profit that a customer generates during her life on the house list
- The net present value of the stream of contributions to profit that result from customer transactions and contacts with the company
- The net present value of all future contributions to overhead and profit
- The net present value of all future contributions to profit and overhead expected from the customer

Thus, we can compute CLV in order to identify profitable customers because measuring RFM is an important method for assessing customer lifetime value (Liu, Shih, 2005). Alongside with its efficiency and other benefits presented before, we decided to utilize this method as a base to compute and express customer profitability.

Sohrabi and Khanlari (2007) in their study (Figure 2) used transactional data to segment customers based on the RFM model and then proposed strategies to retain them.

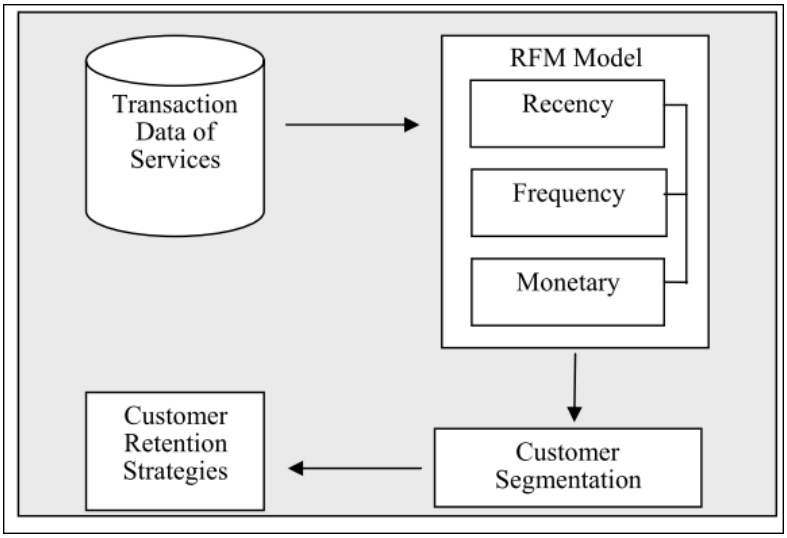


Figure 2: Conceptual framework of research by Sohrabi and Khanlari (2007)

In the process of computing CLV of a customer, form $x'=(x-x^S)/(x^L-x^S)$, was used to normalize the F (frequency) and M (monetary) values, since F and M positively influenced CLV. The cost form, $x'=(x^L-x)/(x^L-x^S)$, was used to normalize the R value, since it negatively impacted CLV. x' and x represented the normalized and original R (F, M) values, while x^L and x^S represented the largest and smallest R (F, M) value of all customers. This process was used also in other studies (Shen, Chuang, 2009; Liu et. al. 2011), where normalized RFM values of each customer were then multiplied by the relative importance of RFM variable, w_R, w_F and w_M (Liu, Shih 2005; Shen, Chuang, 2009).

RFM model for profitability evaluation

Similar approach but with z-score normalization instead of min-max normalization is also applicable. This model was used to analyze the relative profitability for each customer cluster. With models such as that, firms can quickly find the target clusters and adjust its marketing programs and business initiatives to provide right products, services and resources to the target clusters.

Based on the RFM scheme, the value of the customer, or profitability of a customer can be represented as $V(c_i) = W^R \times R(c_i) + W^F \times F(c_i) + W^M \times M(c_i)$ where $R(c_i)$, $F(c_i)$, and $M(c_i)$ represent the scores for customer c_i in terms of the R, F, and M criteria,

respectively. W^R, W^F, and W^M represent the importance weights for the RFM attributes. Then the scores retrieved from the original database are normalized by the z-score normalization before calculation profitability of a customer. Therefore the scores are redefined as:

$$R(c_i) = \frac{O_i^R - \mu^R}{\sigma^R}; F(c_i) = \frac{O_i^F - \mu^F}{\sigma^F};$$
$$M(c_i) = \frac{O_i^M - \mu^M}{\sigma^M}$$

Where where O_i^R,O_i^F, and O_i^M represent the original values for a customer c_i derived from database. μ^R,μ^F, and μ^M represent the averages for the O_i^R,O_i^F, and O_i^M values for all customers. σ^R,σ^F, and σ^M represent the standard deviations of the O_i^R,O_i^F, and O_i^M values for all customers. Then the profitability of customers can be defined to help company to offer customized products and services to specific customer cluster. The weights in this case were set as $W^R = 0.2$ $W^F = 0.4$ and $W^M = 0.4$.

This computation has strong benefit that can be easily implemented into ETL tool and the process automatized with other transformations before loading data extracted from original database to data warehouse. Our thesis adopts the RFM scoring approach of Miglautsch (2000). The main reasons are equal 5 customers groups and 25 segments by his valuation, which is very useful for data mining techniques we're planning to utilize (e.g. PLS path modeling, decision trees and association rules). Researches (Wu & Lin, 2005; Nowell 1997) showed that the bigger the value of Recency and Frequency is, the more likely customers are to produce a new trade with company and thus are more valuable in our definition of customer profitability, which is reflected in Miglautsch's formula. This approach to definition of best customer and customer profitability is supported also by Aggelis. Generally RFM segments with higher RFM scores are more profitable and loyal to the company. The best customers are those who are more likely to purchase again and those are customers with higher RFM values.

We use the terms "customer profitability" and "customer value (to the firm, e.g. best customers)" interchangeably in this research. Both expressions represent a measure of the economic value of a customer to the firm (Reinartz 2005). Conceptually, there may be reservations about using "customer lifetime value" because it implies complete knowledge

(past and future) about a customer's value to the firm. It is not within our scope to take such a viewpoint.

Customer Lifetime Value, Profitability and RFM

Customer lifetime value (CLV) is typically used to identify profitable customers and to develop strategies to target customers (Irvin, 1994; Shen, Chuang, 2009). Bitran and Mondschein (1996) define CLV as expected profits from customers, exclusive of costs related to customer management. There have been several studies considering RFM analysis to measure and calculate CLV (Miglautsch, 2000; Liu, Shih, 2004; Sohrabi, Khanlari, 2007). In addition, RFM model can be used to segment customers, observe customer behavior, estimate the response probability for each offer type and evaluate on-line reviewers (Wei, Lin, Wu, 2010). Essentially RFM analysis suggests that the customer exhibiting high RFM score should normally conduct more transactions and result in higher profit for the company (Aggelis, Christodoulakis, 2005).

RFM model has been widely applied in many practical areas, including nonprofits and financial organizations (banking and insurance industries) (Sohrabi and Khanlari, 2007), government agencies (King, 2007), on-line industries (Li et al., 2010), telecommunication industries (Li et al., 2008), travel industries (Ha and Park, 1998) and marketing industries.

Essentially RFM analysis suggests that the customer with high RFM score should conduct more transactions and result in higher profit (Aggelis, Christodoulakis, 2004). Moreover, customer with high RFM score is more beneficial to the business currently as well as in the future (Aggelis, 2004) it enables organizations to focus on the customers that provide the highest potential return.

Essential component in our study are the most profitable customers. Once the modeling process of a customer's value is established, an investigation into the factors that impact profitability can be performed. In our thesis we aim to find the most valuable customers for an outfitter and then definee customer value as "the economic value of the customer relationship to the firm" (Kumar and Reinartz 2006). By incorporating the concept in marketing strategy planning, a company can optimize its utility of marketing resources. Thus, the calculation of customer value is crucial to customer relationship management.

When calculating customer value, companies or academicians often use profitability indices such as contribution margins, net profits or customer life time value and RFM (recency, frequency, and monetary) metrics to understand customers' financial potential in direct marketing (Kumer and Reinartz, 2006).

1.3 Customer characteristics as antecedents of Customer Profitability

Measuring customer profitability encourages managers to focus on long term rather than short term (Gupta and Lehmann, 2003) and to focus on customer rather than products. The focus of our investigation is on demographics and variables that determine the nature of customer-firm exchange. Any exchange can be characterized by timing, scope and depth of buying (Reinartz and Kumar 2003). There is sufficient evidence in marketing literature that behavioral characteristics such as past purchases are strong predictors of future customer behavior (Dwyer 1997). In addition to the basic behavioral characteristics, we consider other relevant characteristics, for example signaling commitment (e.g. through participation in loyalty program). More specifically, customer behavioral characteristics that are recommended in the literature include relationship duration, loyalty program membership, total no. of services used, age, (Verhoef, Franses and Hoekstra, 2001) sex, education level, marital status, number of children in household, area of residence (Mittal and Kamakura, 2001), date of purchase, number of marketing initiated marketing campaigns in a specific time period, type of the campaigns, number of initiated contacts from the supplier, the level of cross-buying, the number of returns made by the customer and amount spent at each purchase occasion. The quality of the company's database, specifically the precision and depth of it influences the potential of database usage to several analyses (Clemens, 1994). But it's beyond our scope to investigate all of them; our limitation is the availability of the data in our data sets. Our decision in choosing variables is supported by Verhoef et al. (2002)'s study. This study was investigating which characteristics of customers and prospects are practitioners saving in databases. The results of availability of customer information were as showed in Table 3.

Characteristic	Percent	Characteristic	Percent
Name, address information	97.8	Channel of purchase	50.4
Date of first purchase	73.9	Response type	48.7
Source of customer	73.0	Offer characteristics	46.5
Type of product purchased	68.6	Interaction information	42.5
Amount of first purchase	61.9	Socio-demographic information	34.5
Number of offers	61.5	Lifestyle data	17.3
Date of all previous purchases	56.6	Satisfaction data	12.4
Amount of all previous purchases	56.6	Purchase data (other companies)	7.5

Table 3: Type of customer characteristics stored in the customer database (Verhoef et al., 2002)

We can see that most widely stored information was name and address of the customer followed by date of first purchase, type of products purchased or amount of all previous purchases. On the other hand, the least available customer information is usually socio-demographic information, lifestyle data or satisfaction data.

We are interested also in interdependencies between some of the variables, which is in line with findings by Reinartz and Kumar (2003). It has been empirically proven by those studies that purchase frequency, purchase amount, purchase composition and demographic characteristics of the customers affect the profitable lifetime duration of the customers but also that customer characteristics may explain profitability more than lifetime duration does.

Measuring Customer Lifetime Value

One of the most challenging issues in measuring individual-level CLV is data collection and management. Firms need to collect individual-level data about all their customers in order to compute CLV (Kumar et al., 2004). Key informational needs include demographic information, the amount spent on purchase, products purchased in each occasion, the number, time, and type of marketing contacts. Thus, firms should maintain a longitudinal database of all their customers (Kumar, George, 2007). Segmentation of customers based on their lifetime profits and relationship duration is depicted on the Figure 3.

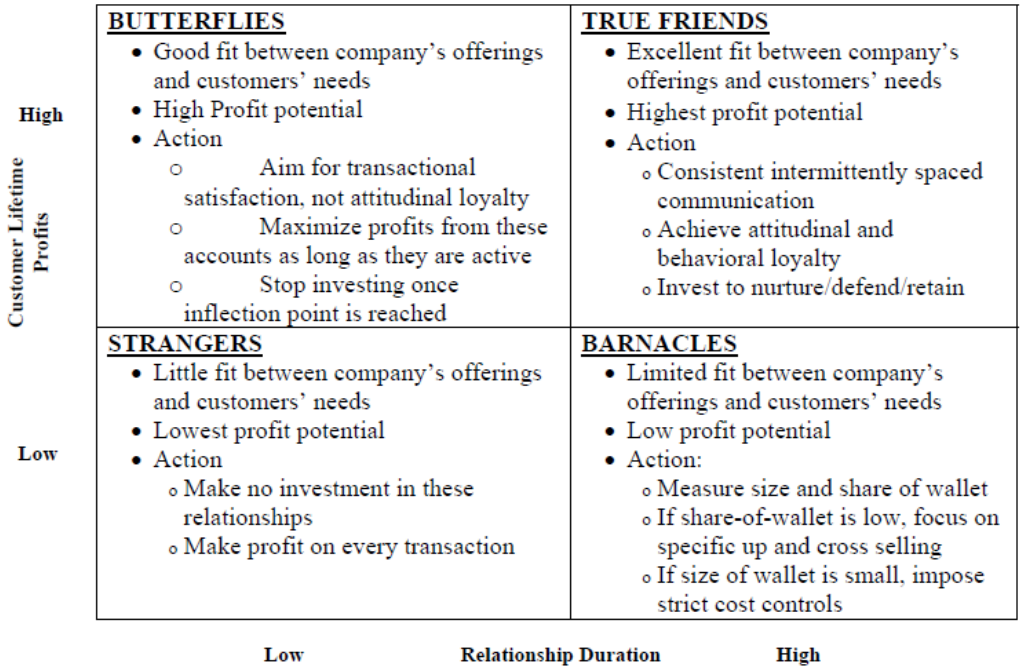


Figure 3: Segmentation of Customers Based on Customer Lifetime Profits and Relationship Duration (Reinhartz, Werner and Kumar, 2002).

1.4 Data mining process

Data mining techniques are a widely used information technology for extracting marketing knowledge and further supporting marketing decisions. During the past decade, there have been a variety of significant developments in data mining techniques and recently, the data mining techniques have been adopted to predict customer behavior (Song, Kim, & Kim, 2001). Data mining techniques search through a database to obtain implicit, previously unknown, and potentially useful information including knowledge rules, constraints and regularities (Chen, Han, & Yu, 1996). Data mining is a stage in Knowledge Discovery in Databases (KDD), involving the application of specific algorithms for pattern extraction (Mitra, Pal, & Mitra, 2002). Various successful applications have been reported in areas such as marketing, finance and banking. Applications in these domains generally involve the collection and storage of large amounts of data. Data mining brings various techniques together to discover patterns (rules) and to construct models from databases.

Retail market managers must not only provide high quality products and services, but also must react appropriately to changes in customer needs. Data mining can be applied to

identify useful customer behavior patterns from large amounts of customer and transaction data. As a result, the discovered information can be ascertained to support better decision-making. Data mining techniques have mostly been adopted to generate predictions and describe behaviors (Chen, Chiu, Chang, 2005). Data mining is also part of the broader Customer relationship management and Business Intelligence topics. It is a method or tool that can aid companies in their quest to become more customer-oriented (Rygielski et al., 2002). The goal of data mining in this context is to allow a business to improve its marketing, sales, and customer support operations through a better understanding of its customers. Hardly any of the data mining algorithms were first invented with commercial applications in mind. The commercial data miner employs a grab bag of techniques borrowed from statistics, computer science, and machine learning research.

Data mining is defined as a sophisticated data search capability that uses statistical algorithms to discover patterns and correlations in data. It is complementary to other data analysis techniques such as statistics, on-line analytical processing (OLAP), spreadsheets, and basic data access. In simple terms, data mining is another way to find meaning in data (Rygielski, 2002) and improve decision making (Olszak & Ziemba, 2007).

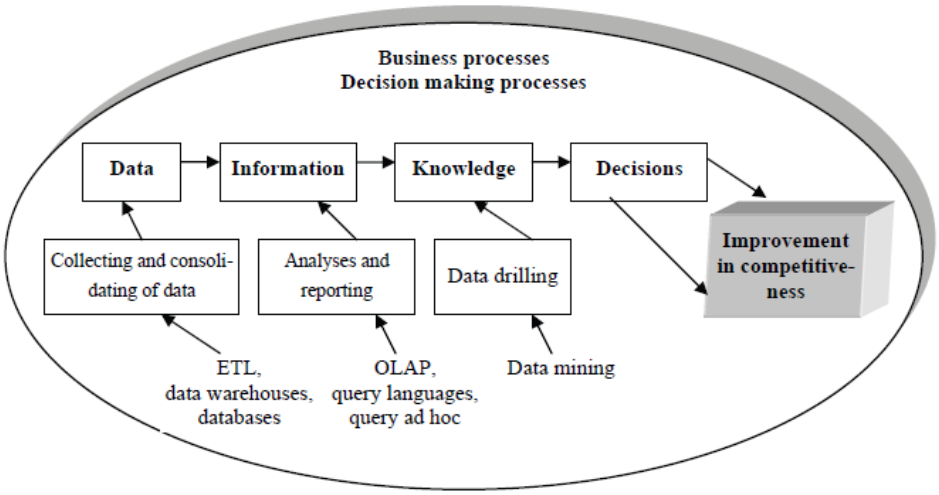


Figure 4: Decision making process (Olszak & Ziemba, 2007)

In our study we are following all the steps described by Olszak & Ziemba (2007), where we have two sets of data, which were pre-processed and transformed and knowledge is derived utilizing various data mining techniques. Then, competitiveness in business is improved through implementing companies' decisions based on our results.

1.4.1 Models for Data mining

In the data mining literature, various "general frameworks" have been proposed to serve as blueprints for how to organize the process of gathering data, analyzing data, disseminating results, implementing results, and monitoring improvements. One such model, CRISP (Cross-Industry Standard Process for data mining) was proposed in the mid-1990s by a European consortium of companies to serve as a non-proprietary standard process model for data mining. This general approach postulates the following general sequence of steps for data mining projects:

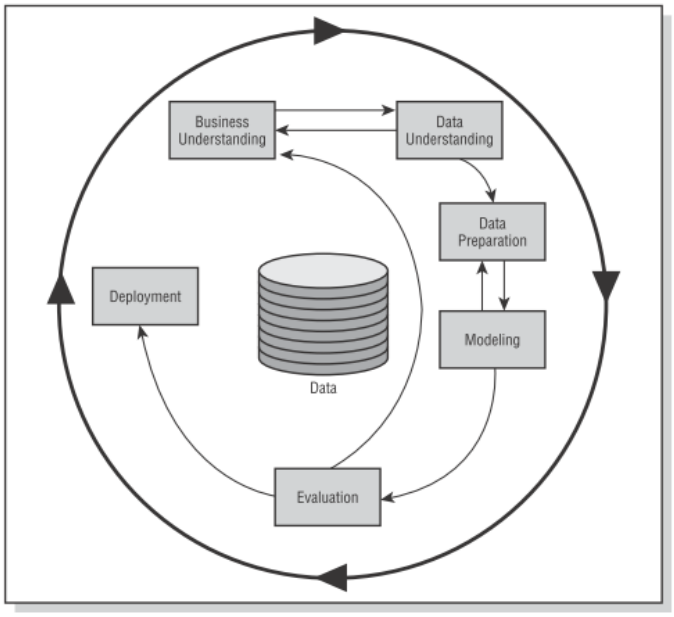


Figure 5: CRISP model

Next, following figures depict the whole process of knowledge discovery in more detail. Basically both approaches are essentially the same. From data stored in databases we are trying to gain valuable knowledge. At first, the data usually need to be cleaned and transformed and then loaded into data warehouse and if necessary, also transformed and reduced. This part of the knowledge discovery process is also called ETL (Extract, Transform, and Load). Next, on such data the data mining techniques are performed, the results are then evaluated and interpreted in comprehensible and understandable form. From this results the knowledge is gained, decisions are made and in the long run, the

business is more profitable and competitive. All of these steps are also performed and described in our study.

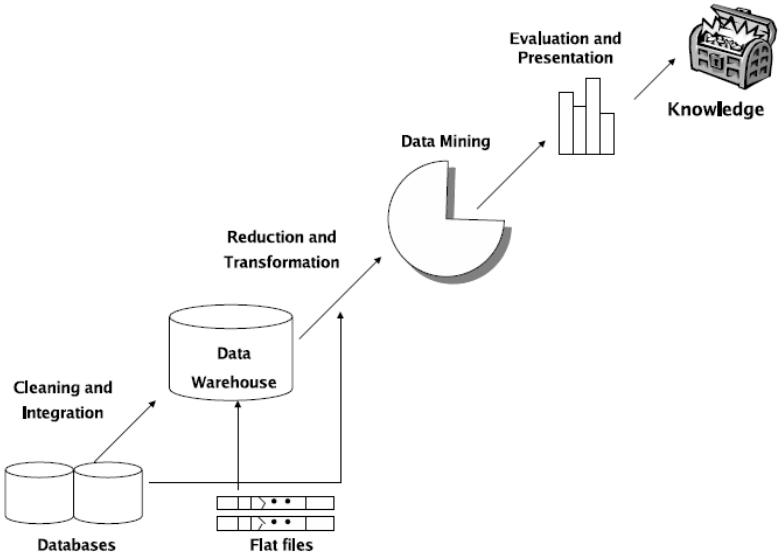


Figure 6: The process of Knowledge Discovery in databases (Maiom, Rokach, 2005)

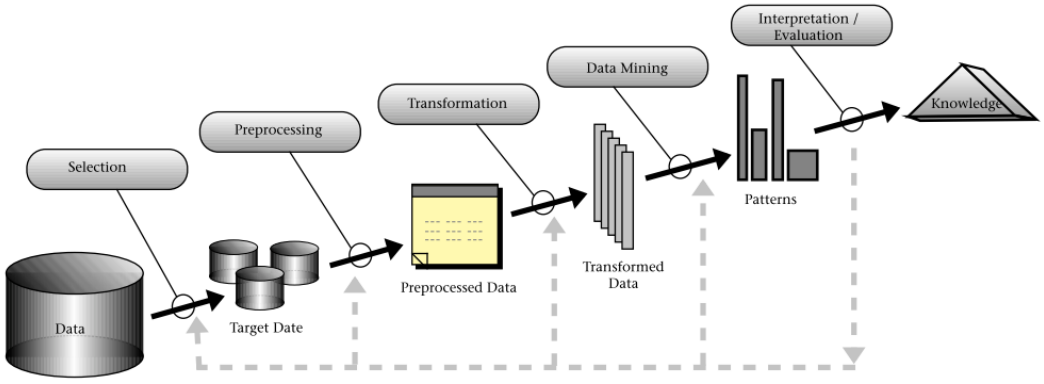


Figure 7: An overview of the steps that comprise knowledge discovery process (Fayyad et al. 1997)

1.4.2 Data mining techniques used in the study

Customers today are very diverse and to satisfy their needs the segmentation that divides markets into customer clusters with similar characteristics that are likely to exhibit similar purchasing behaviors. Segmentation is basically process to find classes in the data. With proper market segmentation, companies can arrange the right products, services and resources to target customer cluster and build a close relationship with them (Tsai, Chiu,

2004). As a consequence, market segmentation has been regarded as one of the most critical element in achieving successful modern marketing and customer relationship management. In our study we will perform segmentation by RFM analysis and after that we will perform classification, which tries to attach an element to an existing class or category. We will use decision trees to perform classification.

Basic overview of the data mining techniques is illustrated on the Figure below.

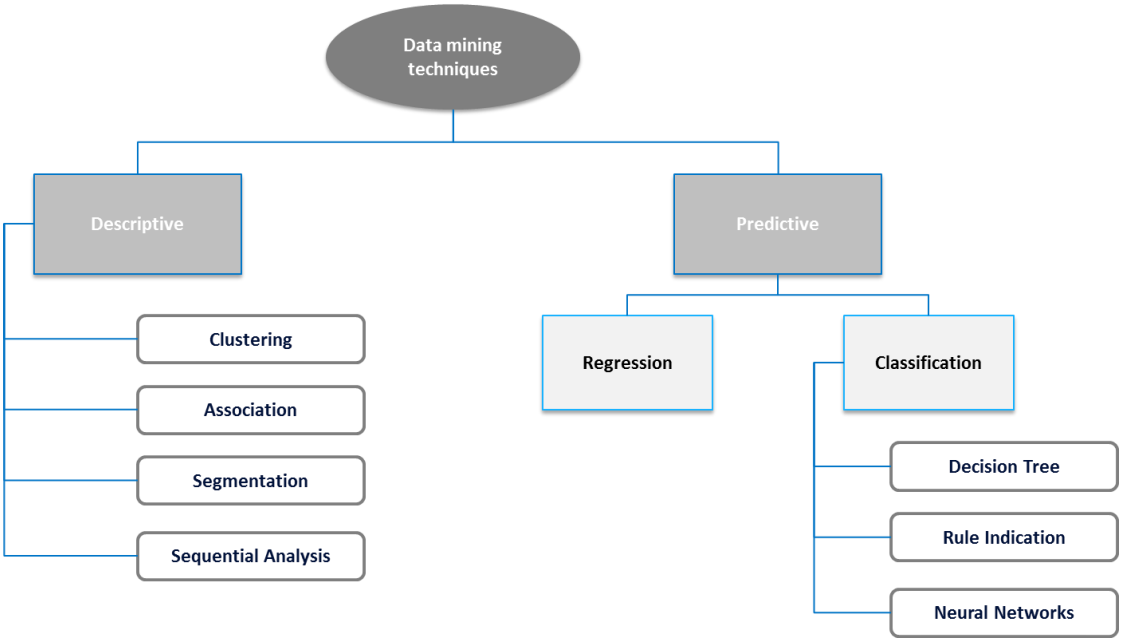


Figure 8: Overview of data mining techniques

1.4.1 Data mining tools utilized in the study

ETL tool Kettle is used for data pre-processing and segmentation of customers by RFM analysis, then Classification and Predictive modeling of WEKA data mining software is utilized and regression analysis (SPSS software utilized) are to perform data mining based on the data provided. Also SmartPLS tool is used to compute paths in the model by Partial Least Squares analysis algorithm. MySQL database was used to store and MySQL Workbench was used to model data warehouses and make basic data transformations.

1.4.2 Algorithms used in the study

Descriptive data mining techniques

Association analysis is the discovery of what are commonly called association rules. It studies the frequency of items occurring together and based on a threshold called support, identifies the frequent item sets. Another threshold, confidence, which is the conditional probability than an item appears in a transaction when another item appears, is used to pinpoint association rules.

Association analysis is commonly used for market basket analysis. For example, it could be useful for the video store manager to know what movies are often rented together or if there is a relationship between renting a certain type of movies and buying popcorn or pop. The discovered association rules are of the form:

X -> Y [s,c],

where X and Y are conjunctions of attribute value-pairs, and s (for support) is the probability that X and Y appear together in a transaction and c (for confidence) is the conditional probability that Y appears in a transaction when X is present. X is termed the left-hand-side (LHS), and is the conditional part of an association rule. Meanwhile, Y is called the right-hand-side (RHS), and is the consequent part. In Song et al. (2001), LHS of association rules chooses customer profile variables such as gender, age, yearly income, and so on; whereas RHS includes the product items bought. For example, the hypothetic association rule:

RentType(X, "game") AND Age(X, "13-19") -> Buys(X, "pop") [s=2%,c=55%]

would indicate that 2% of the transactions considered are of customers aged between 13 and 19 who are renting a game and buying a pop, and that there is a certainty of 55% that teenage customers who rent a game also buy pop.

Formal definitions of support and confidence are given below.

Frequency	Given an itemset pattern X, its *frequency fr(X)* is the number of cases in the data that satisfy X.
Support	*Support* is the frequency fr(X ^ Y).
Confidence	*Confidence* is the fraction of rows that satisfy Y among those rows that satisfy X, $c(X \rightarrow Y)=fr(X \wedge Y) / fr(X)$

In terms of conditional probability notation, the empirical accuracy of an association rule can be viewed as a maximum likelihood (frequency-based) estimate of the conditional probability that Y is true, given that X is true.

Customer behavioral data are generally the most effective predictive data in customer relationship management (Rud, 2001). Association rules can include any number of attributes on either side of the rule. Not all association rules are interesting to decision makers. Rule support and confidence are two measures of rule interestingness. An interesting rule must satisfy the minimum support and confidence determined by domain experts. In the algorithms for association rule mining, Apriori is one of the most widely used algorithms. It is typically used to find association rules by discovering frequent itemsets (sets of product items). An itemset is considered to be frequent if the support of that itemset exceeds a user-specified minimum support. Association rules that meet a user-specified minimum confidence can be generated from the frequent itemsets. In this study, Apriori algorithm is applied to discover customer behavior patterns. Association rules were initially applied to analyze the relationships of Customer Profitability.

Apriori algorithm is an influential algorithm for mining frequent itemsets for Boolean association rules. The algorithm contains a number of passes over the database. During pass *k*, the algorithm finds the set of frequent itemsets *Lk* of length *k* that satisfy the minimum support requirement. The algorithm terminates when *Lk* is empty. A pruning step eliminates any candidate, which has a smaller subset.

The pseudo code for Apriori Algorithm is following:

Ck: candidate itemset of size k
Lk: frequent itemset of size k

L1 = {frequent items};

For (k=1; Lk != null; k++) do begin

Ck+1 = candidates generated from Lk;

For each transaction t in database do

Increment the count of all candidates in

Ck+1 that are contained in t

Lk+1 = candidates in C k+1 with min_support

End

Return Lk ;

To illustrate the concept, we use a small example from the supermarket domain. The set of items is {milk, bread, butter, beer}. An example rule for the supermarket could be {butter, bread} -> {milk} meaning that if butter and bread is bought, customers also buy milk. This example is extremely small and in practical applications, a rule needs a support of several hundred transactions before it can be considered statistically significant, and datasets often contain thousands or millions of transactions.

Sometimes we will get rules that don't include the target class, which in our case, is customer profitability. This is why Apriori is often used for dataset exploration rather than for predictive modeling.

Predictive data mining techniques

Predictive analytics is data mining technology that uses customer data to build a predictive model specialized for certain business. This process learns from organization's collective experience by leveraging existing records of customer purchases, behavior and demographics. Trees and rules are primarily used for predictive modeling, both for classification (Apte and Hong 1996; Fayyad, Djorgovski, and Weir 1996) and regression, although they can also be applied to summary descriptive modeling (Agrawal et al. 1993).

Classification is learning a function that maps (classifies) a data item into one of several predefined classes. Examples of classification methods used as part of knowledge discovery applications include the classifying of trends in financial markets (Apte and

Hong 1996) and the automated identification of objects of interest in large image databases (Fayyad, Djorgovski, and Weir 1996). In another words, classification analysis is the organization of data in given classes. Also known as supervised classification, the classification uses given class labels to order the objects in the data collection. Classification approaches normally use a training set where all objects are already associated with known class labels. The classification algorithm learns from the training set and builds a model. The model is used to classify new objects. For example, after starting a credit policy, the video store managers could analyze the customers' behaviors via their credit, and label accordingly the customers who received credits with three possible labels "safe", "risky" and "very risky".

Regression is learning a function that maps a data item to a real-valued prediction variable. Regression applications are many, for example, predicting the amount of biomass present in a forest given remotely sensed microwave measurements, estimating the probability that a patient will survive given the results of a set of diagnostic tests, predicting consumer demand for a new product as a function of advertising expenditure, and predicting time series where the input variables can be time-lagged versions of the prediction variable.

A decision tree is a predictive machine-learning model that decides the target value (dependent variable) of a new sample based on various attribute values of the available data. The internal nodes of a decision tree denote the different attributes; the branches between the nodes tell us the possible values that these attributes can have in the observed samples, while the terminal nodes tell us the final value (classification) of the dependent variable.

By checking all the respective attributes and their values with those seen in the decision tree model, we can assign or predict the target value of this new instance. In our study in order to reduce the error and visualize the model we experimented with groups sizes. And since the model will make use to make predictions for data that we might receive in the future, we learned a model from all of the available training data.

It was seen that **J48 Decision trees** algorithm performed really well. In most cases, their performance was almost at par with that of Support Vector Machines. In fact in several

cases, it was seen that J48 Decision Trees had a higher accuracy than either Naïve Bayes, or Support Vector Machines. It turns out one can do even better by using more than one predictor at a time, combining them with a model. Creating this model is the very purpose of predictive analytics.

Since most real world applications of decision tree and classification learning involve continuous-valued attributes, or also known as numerical attributes, the problem of discretization these values is very important (Perner et al. 1996). A continuous-valued attribute is typically discretized during decision tree generation by partitioning its range into intervals. This method is called binary discretization. However, several studies have shown that multi-interval discretization (dividing values into more than two intervals) can outperform binary discretization in terms of accuracy and explanation capability (Perner, Trautzsch, 1998). In our study we would like to achieve reasonable level of accuracy of the resulting decision tree and the compactness of the tree.

Algorithm C4.5

Decision tree induction can be best understood as being similar to parameter estimation methods in econometric models. The goal of tree induction is to find the set of Boolean rules that best represents the empirical data. In this study, the trees were induced using the C4.5 method (Quinlan, 1993). Implementation of C4.5 algorithm in software tool Weka, which we utilize in this study is called J48.

In general, a tree consists of different layers of nodes. It starts from the root node in the first layer or first parent node. This parent node will split into daughter nodes on the second layer. In turn, each of these daughter nodes can become a new parent node in the next split, and this process may continue with further splits. A leaf node is a node, which has no offspring nodes. Nodes in deeper layers become increasingly more homogeneous. An internal node is split by considering all allowable splits for all variables and the best split is the one with the most homogeneous daughter nodes. The C4.5 algorithm recursively splits the sample space on X into increasingly homogeneous partitions in terms of Y, until the leaf nodes contain only cases from a single class. Increase in homogeneity achieved by a candidate split is measured in terms of an information gain ratio (Quinlan, 1993).

Partial Least Squares (PLS) variance analysis is estimation technique for structural equation modeling (Chin, 1998) used in this study. Interest in the PLS method has been increasing in recent years because of its ability to model constructs under conditions of non-normality and because of superior predictive power of its models, which were reasons to use PLS method also in this study. We utilized the software package Smart PLS of the University of Hamburg to compute factor loadings and Average Variance Extracted (AVE). In addition, SPSS (version 19) was used to compute minimum item-to-total correlations.

Partial Least Squares (PLS) path modeling is most often used in social sciences, econometrics, marketing and strategic management. PLS path modeling is a components based methodology that provides determinate construct scores for predictive purposes (Chin, 1998). PLS path modeling has encountered increasing popularity as an easy, yet powerful, estimation technique for structural equation models.

Market researchers and academics widely apply PLS path modeling to predict endogenous latent variables in their success driver studies. Moreover, PLS applications have been successfully used in the fields of strategic management, information technology management, media management and different sub-disciplines of marketing including international marketing, drivers and consequences of customer satisfaction and evolution of loyalty intentions, relationship marketing, and business-to-business marketing.

Partially Least Squares (PLS) modeling was used to assess measurement and to estimate the structural model. PLS was selected because our model is relatively complex with high amount of variables and paths between them, so it intends to develop theory rather than confirm it. Moreover it has advances in terms of minimal demands on data distribution and sample size (Chin, 1998). It is particularly useful when we need to predict a set of dependent variables from a (very) large set of independent variables (predictors). Although PLS can be used for theory confirmation, it can also be used to suggest where relationships might or might not exist and to suggest propositions for later testing.

2. Research Model and Hypotheses

To study the influence of customer characteristics on profitability, the research model presented in Figure 9 is presented. The first aim of this study is to test all the hypotheses included in the research model. When all hypotheses are tested by regression analysis, the relationships between customer demographic, situational characteristics and customer profitability will be further explored by PLS path modeling, decision trees and association rules mining.

The second aim of this study is to provide a framework for building up a metric for customer profitability variable included in the first aim of the study. This model will be built upon the surveyed literature, but availability of data and simplicity in implementation will be the primary practical limitations.

2.1 Hypotheses Development

The quality of the company's database, specifically the precision and depth of it influences the potential of database usage to several analyses (Clemens, 1994). In this section we will formulate expectation about the effect of the customer relationship characteristics and demographics on customer profitability as well the antecedents and consequences of some of them. We were able to formulate prepositions and construct research model from existing theoretical and empirical knowledge of marketing paradigm, literature recommended customer characteristics, antecedents of profitability related research and data availability in our datasets.

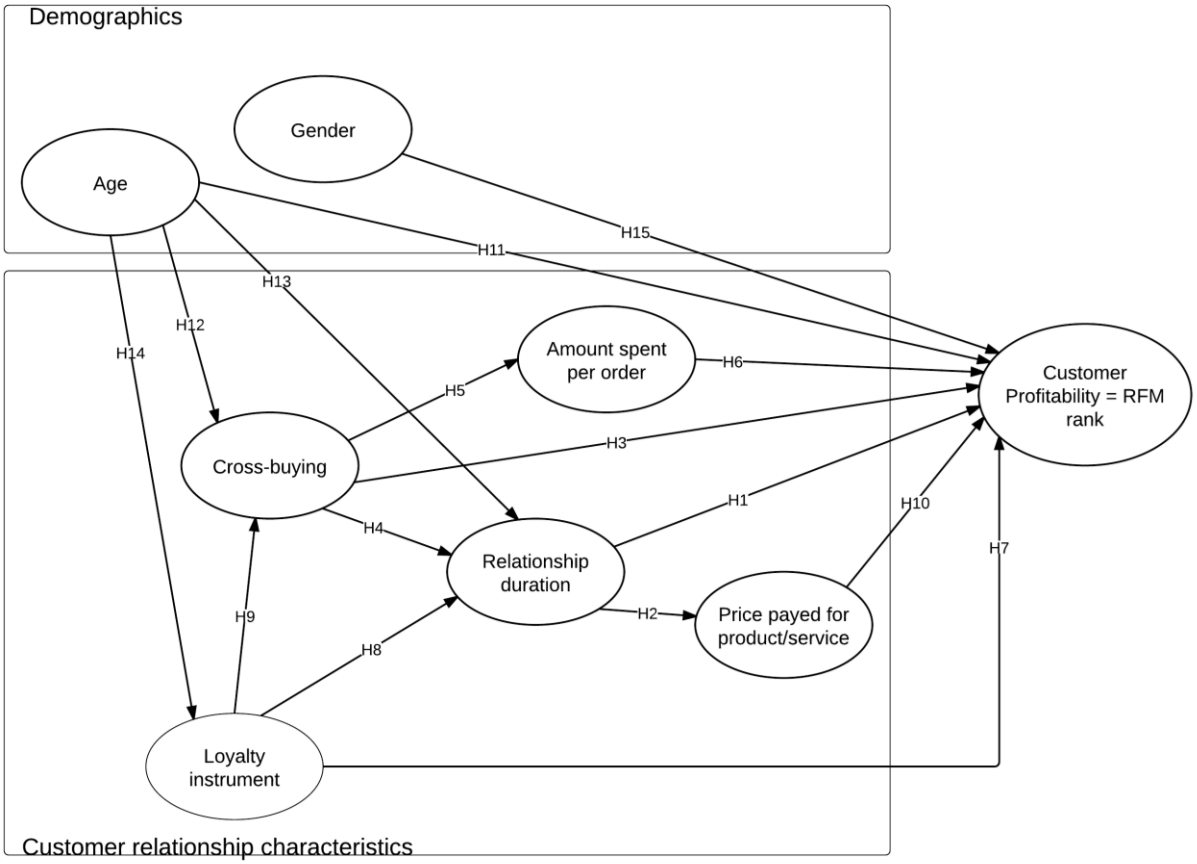


Figure 9: Research Model

2.1.1 Customer relationship characteristics

Relationship duration

Marketing theory and practice has become more and more customer centered and managers have increased their emphasis on long-term client relationships because the length of a customer's tenure is assumed to be related to long-run company revenues and profitability (Bolton et al., 2002 and Gupta et al., 2004). The building of strong customer relationship has been suggested as a means for competitive advantage (Berry, 1995). Because of lack of empirical evidence in the noncontractual setting and weak theoretical justification, we provide arguments for both for and against the positive relationship between customer tenure and profitability. Claim that the best customers are the loyal ones and firms benefit more from marinating long-term customer relationships as compared to short-term relationships was used in many studies (Reichheld and Sasser 1990; Bendapui and Berry 1997; Hallowell 1996). This assumption was advocated by the higher exchange efficiencies created by customer retention (Sheth and Parvatiyar 1995). Simply said,

longer relationship duration could indicate possibility for higher purchase frequency and volume and thus higher customer profitability. In addition, longer customer lifetime indicates greater satisfaction of a customer and satisfied customers are expected to by more loyal and repeat the purchase of the goods or services and thus be more profitable (Wang et al. 2004). Although this could be true in highly competitive markets where in order to retain customers it is important to satisfy them, in the non-competitive market it is less difficult to retain customer even if he's not that satisfied. Moreover, in the e-commerce industry people can consider a lot of choices before they actually place an order, which could indicate that long-life customers don't necessarily buy very frequently and therefore they don't have to be more profitable than short-time customers. This line of reasoning supports also recent study by Reinartz and Kumar (2000; 2003) with conclusion for non-contractual settings, that both long and short term customer – firm relationships can be profitable. In their empirical findings the link between customer lifetime and profits was weaker than expected. In addition, Dowling and Uncles (1997) claimed that saying that loyal customers are always more profitable is a gross oversimplification. However, because of our settings and more availability of evidence of positive lifetime-profitability relationship, we anticipate direct and positive relationship between lifetime duration and profitability.

We've already discussed arguments of Reichheld (1996a) and Reinartz, Thomson, Kumar (2005) that claim long-term customers buy more because as the length of the customer tenure rises, it allows more transactions which means higher volume and frequency. Another consequence of being long-life customer is connected with average price customers pay for product. As Reicheld and Teal (1996) argue, in most industries the existing customers pay higher prices than new ones, which would imply that the average price paid by customer and his lifetime duration could be positively related. Also Bolton, Lemon and Verhoef (2004) suggest loyal customers are sometimes assumed to be willing to purchase premium products instead of low-cost variants. We use this line of reasoning since long relationship length indicates strong intention for the continuity of the relationship and thus relationship duration can be considered as a measure of attitudinal loyalty as previously suggested by Pong and Yee (2001). However, Reinartz and Kumar (2000) argue that customers with lengthy relationships are higher value conscious. That is, if customers buy more product units for a given dollar amount, they exhibit a higher degree of value consciousness and they gain more "bang for he buck". If this observation

was true, it would contradict the evidence from Reichheld and Teal (1996a). A possible reason for higher value consciousness of long-term customers is that customers learn over time to trust lower-priced items or brands rather than established name brand products. Thus, there exists some reasonable evidence for both possibilities. Therefore, instead of proposing a directional effect, we test this proposition empirically.

Based on the reasoning above we anticipate:

H1*: Customer relationship duration has positive influence on the profitability of a customer*.
H2*: Customer relationship duration has influence on prices paid.*

Cross-buying

Cross-buying refers to the customer's practice of buying additional products and services from the existing service provider in addition to the ones s/he currently has (Ngobo 2004), in our case the degree to which customers purchase products or services from a different product categories.

Assuming that the firm offers products or services in different categories, consumers can purchase products in a focused manner or purchase across a variety of different categories. For example, in a general merchandise context, one customer might buy only women's shoes and formal wear, whereas another customer might purchase shoes, high-fashion, formal wear, sportswear, accessories and so on. In the latter case, the scope of interaction with the firm is rather broad, whereas in the former case, it is focused. The existing studies in the marketing domain have focused on the related, yet different, construct of cross selling (Chen et al. 1999; Drèze and Hoch 1998), but as compared with the cross-selling construct, far fewer studies deal with the cross-buying construct (Reinartz and Kumar, 2003). So what remains open in the literature is the effect of cross-buying on individual level outcomes such as customer lifetime duration (Reinartz and Kumar, 2003).

In contractual settings, cross buying positively influence a customer's incline to stay in a relationship (Morgan and Hunt, 1994; Venkatesan, 2002) because buying products or services across several categories is claimed to increase switching costs (Srivastava, Shervani and Fahey, 1998). In non-contractual settings, however, these switching costs don't exist. But customers still benefit from knowing the companies' product range and

quality level so it is believed that these customers are less likely to leave the relationship. Moreover, they might have recurrent needs and be more loyal, which could indicate positive influence on profitability (Bowman and Narayandas 2001; Reinartz and Kumar 2003). With respect to scope of purchases there is still little empirical evidence that links cross-buying with a customer's tenure.

In a cross-buying context buying from an additional product category might be considered as higher trust in company's product range which reduces the perceived risk (Kumar, George and Pancras, 2008). Because of this improved trust, customers may buy high-end products or simply more expensive products. In addition, increase in the choice of product categories and products (and thus higher degree of cross-buying) helps customer to shop for a wider range of products in a purchase occasion. This was already proved in study by Kumar et al. (2008) in retailing; our ambition is to test the relationship in regular Webshop and Theater setting. In addition, also Kamakura et al. (2003) claims that cross-buying as is an indicator of stronger relationships should have a potential impact on both relationship duration and customer profitability.

Based on the reasoning above we hypothesize:

H3*: The degree of customer's cross buying behavior has positive influence on customer profitability.*
H4*: The degree of customer's cross buying behavior has positive influence on customer relationship duration.*
H5*: The degree of customer's cross buying behavior has positive influence the customer's amount spent per order/purchase.*

Amount spent at each purchase occasion

Spending of a customer with firm over time can be decomposed into three components: purchase frequency, purchase composition and purchase amount per occasion (Reinartz and Kumar, 2003). This notion is also reflected by Kelley and Thibaut (1978) in more general context, they suggest that the interaction between two parties is demonstrated by frequency, scope and depth of interaction. So, when parties communicated more deeply, the relationship intensifies. Thus, when the customer buys more per each purchase

occasion, the relationship between him and vendor becomes more durable (Reinartz and Kumar, 2003). We expect that intensified relationship has a positive effect on customer purchase behavior. Sharp and Sharp (1997) stated that one desirable effect of the loyalty program would be to increase average amount bought on each purchase occasion. From this statement we could conclude they expect positive influence of average amount spent on each purchase occasion on profitability. Therefore, we offer proposition:

H6: *The customer's amount of money spent per each purchase occasion has positive influence on customer's profitability.*

Loyalty instrument

Another variable that should have an impact on exchange characteristics is whether a customer subscribes to the loyalty instrument of the company. In the context of Webshop customer can subscribe to the newsletter offering, in the context of Theater, loyalty instrument has a form of a discount card only applicable for cultural events. In case of a Webshop the marketing instrument rather than loyalty instrument might be used to address it, although signing up for the newsletter can suggest higher loyalty and intention to buy from company. Ownership of those loyalty instruments can be considered also as relationship intention, which is the intention of a customer to build a relationship with firm while buying a product or service of the firm, which Kumar, Bohling and Ladda (2003) employ as a measure of loyalty. Furthermore, Kumar, Bohling and Ladda (2003) claim that low serving cost, premium price, word-of-mouth promotion and company advertisement are consequences of relationship intention. The marketing literature provides a wide range of loyalty measures (Odin, & Valette-Florence, 2001), and their usefulness depends on the specific market and study objective. Repurchasing probability (Andreassen and Lindestad, 1998), likelihood of providing positive word-of-mouth, willingness to pay more (Srinivasan, Anderson and Ponnavolu, 2002) or customer longevity (Reinartz and Kumar, 2002) are all used as measures of customer loyalty. The main distinction in loyalty measures is between attitudinal loyalty and behavioral loyalty (Dick & Basu, 1994) and since we are interested in the effects of loyalty programs on actual purchase behavior, our key dependent variable captures behavioral loyalty.

As Reinartz an Kumar (2003) suggests we assess impact of loyalty instrument registration to the relationship duration, because ownership of a membership card or newsletter can suggest that customer has an interest in using the services of the provider more often and also can benefit from membership offers (discounts etc.). Moreover, ownership of a loyalty instrument can suggest better familiarity with current offers and thus higher probability for purchasing the product. Moreover, the results of Verhoef (2003) show that loyalty programs positively affect customer retention, although the effect of these variables is rather small. Subscription to a loyalty instrument can be considered as a participation in loyalty program, because loyalty programs are structured marketing efforts which reward, and therefore encourage loyal behavior (Sharp & Sharp, 1997).

If customers are affectively committed to a supplier, they are likely to buy additional services from the service organization (Morgan and Hunt 1994; Bolton, Lemon, 2004). This idea is also supported by Verhoef (2001), who reported a positive effect of commitment on cross-buying of financial services. This is in line with Ngobo (2004) who argues that members of the loyalty program are expected to be more likely to cross-buy in the retailing settings than non-member customers. We will empirically test this proposition in our settings. Signing up for a newsletter is also a mean for the company to communicate with the customer via direct mailing and prior research has shown that effective direct mailing promotions positively influence cross-buying (Verhoef et al. 2001).

***H7**: Subscription to loyalty instrument positively influences customer profitability.*
***H8**: Subscription to loyalty instrument positively influences customers' relationship duration.*
***H9**: Subscription to loyalty instrument positively influences degree of customers' cross-buying behavior.*

Price paid

Because of lack of the empirical evidence, we want to test whether more profitable customers pay, on average, higher or lower prices for their chosen products/services. We will compute for each transaction the ration of dollar spending over the number of items purchased and average this figure across purchase occasions for each customer. The line of reasoning is, if a customer purchases premium products or services, he generates higher revenue, which is positively reflected in our customer profitability model. Also Smirlock

(1985) argued that higher profits can result not only from lower costs but also from higher prices. Thus, we expect:

H10*: The height of the price paid by customer has positive influence on profitability.*

2.1.2 Demographic characteristics

Age and Gender

Mittal and Kamakura (2001) recommend gender, marital status, number of children in household, area of residence and education level as customer demographic characteristics that should be considered during customer profitability and relationship duration analysis. Gremler and Brown (1998b) and Verhoef, Franses and Hoekstra (2001) have employed age as a demographic characteristic for the same purpose. However, Reinartz and Kumar (2003) claimed there is lack of appropriate theory to posit directional hypothesis between age and profitable customer lifetime duration. We decided to use age and gender mainly because we were limited by the data available in our two databases.

According to Mittal and Kamakura (2001), old customers are thought to have stable preferences and to be more loyal than the younger ones. Older customer have already gained brand specific knowledge whereas younger customers in contrast might be in the stage of life cycle where they need to experiment, switch brands and search for information (Ratchford, 1999). Moreover, knowledge of older ones may also "divert" their attention from buying additional services (Ngobo, 2004). That could indicate greater profitability and longer relationship duration for older customers and higher degree of cross-buying for younger customers. In online buying behavior, younger customers tend to perform search for certain product or service more often than older customers, but older customers are more likely to buy product than younger customers (Sorce, 2005). In addition, other studies suggest older customers are more loyal (which is generally considered antecedent of profitability [Reichheld 1996a,b]) because of reducing their consideration set, aversion to change or socioemotional selectivity.

Considering gender as an antecedent of customer profitability, we could refer to Van Slyke et al. (2002) and Cyr and Bonanni (2005) that point out there are gender differences in on-line shopping characteristics such as compatibility, complexity or result demonstrability. The results of their research show significant gender differences in beliefs regarding

website design, website trust, website satisfaction and e-loyalty, with men having more favorable perceptions than women which could indicate inclination in greater purchase intention and thus higher profitability. However, there is still lack of appropriate theory in our context, so instead of proposing directional hypotheses, we will test empirically for the effect.

Due to the reasoning above, it is expected:

H11: *Age of the customer has positive influence on profitability.*

H12*: Age of the customer has negative influence on cross buying.*

H13*: Age of the customer has positive influence on relationship duration.*

H14*: Age of the customer has positive influence on registration for loyalty instrument.*

H15*: Gender of the customer influences his profitability.*

3. Research Methodology

In this chapter we will discuss the research method that was used during this study. The research design will be presented among with data preparation and collection. Then, the procedure will be explained and measures used in this study will be described. Finally, the data mining techniques used in the study will be explained.

First, we will apply RFM to measure profitability of the customers, and then identify factors that can explain variation in profitability. We will compare results for non-contractual industries – Webshop as a representative of product industry and theater as a representative of service industry. After RFM score is obtained for each customer, we are interested in knowing the factors or antecedents that drive the profitability and also antecedents and consequences of some of them. In order to do that, we will utilize data mining tools and techniques described later on.

3.1 Approach

Quantitative research was used to carry out this research. After the Proposal phase a thorough literature review was conducted. Based on the literature review, the appropriate assumptions were developed. Then the data were appropriately prepared – the whole process of cleansing, and transformation was performed. After that the chosen data mining methods (Partial Least Squares analysis, Decision Trees and Association Rules) were applied to the data using appropriated software tools. In the end, the results were analyzed, interpreted and conclusions were drawn. High-level research approach is depicted on the picture below.

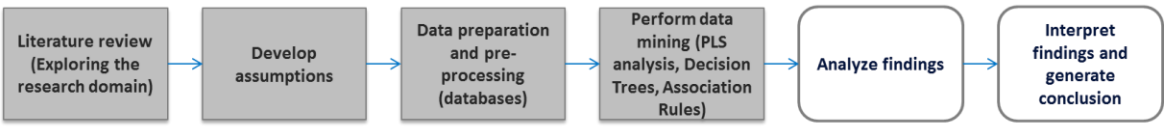


Figure 10: Research approach

3.1.1 Research context

As discussed previously the purpose of this study is to investigate the drivers of the profitability in non-contractual settings. Specifically, we want to identify the variables, obtainable from both databases, which act as drivers of cross-buy so that they can be used for the selection of customers for marketing efforts. We used Webshop and Theater as an example of non-contractual setting for the purpose of this study. In this environment customers can easily change their purchase behavior to competitors without being confronted with high switching costs and without informing the company about it.

To investigate profitability and the factors that influence it we study the customers of a Dutch Theater that has variety of shows on schedule (comedy, musical, cabaret…) and Webshop offering pharmaceutical and drugstore products. Both companies have database systems keeping track about their customers. But before we were able to use this data for the research, they needed to be appropriately prepared. The process of transforming the data, clearing them and discovering knowledge from them is illustrated on the picture below.

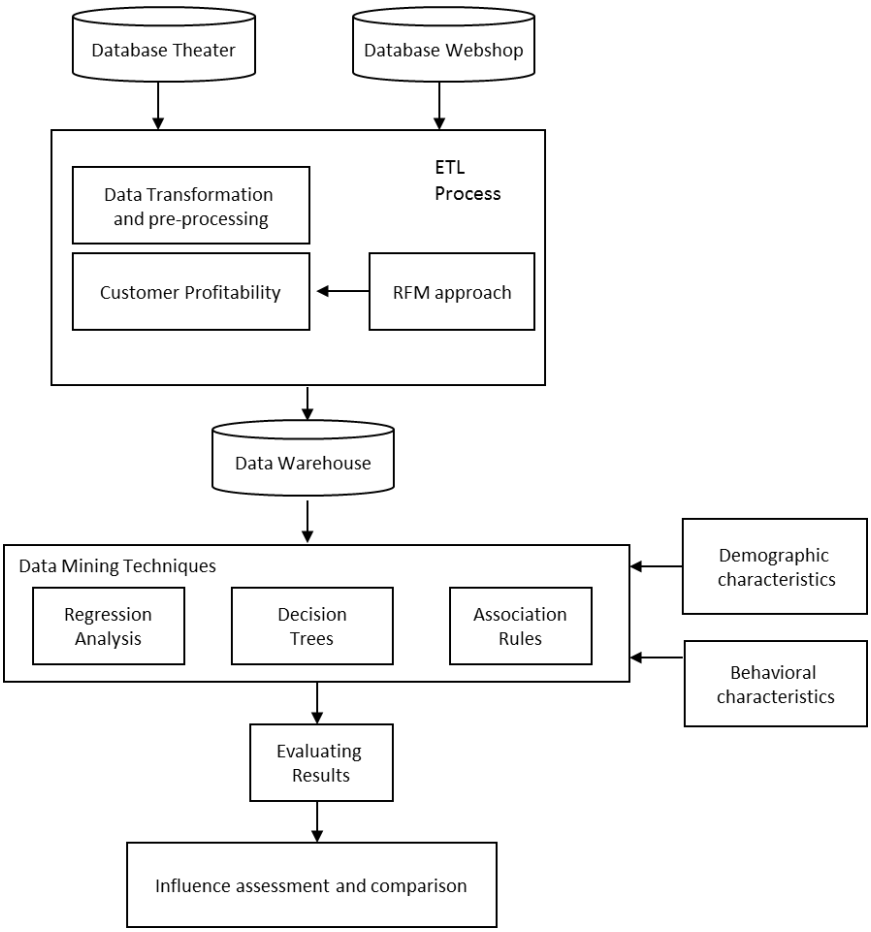


Figure 11: Data processing

To profile the customer, customer data is needed. The proposed data in Section 1.3 is not completely available. Information about lifestyles and income is missing. Usable data that will give knowledge about one customer is contained in the data we have about his transactions and history. But we are not trying to use every data available from the datasets – choices were made depending on the final purpose of the application, so we selected relevant data to understand the most profitable customers. But basically, the behavioral data is very important to look at, as well as demographic and financial information about the customers (Amat, 2002).

3.1.2 Proposed procedure

In this section we further explain the proposed procedure. It can be divided into following steps:

1. At first, select the dataset for empirical case study. Extracting the data and preprocessing the dataset is needed to make knowledge discovery easier and even possible. Thus, we firstly delete the records which include missing values or inaccurate values, transform the datum and other tasks explained in section above.

2. Segment customers by RFM model. The following step is to yield quantitative value of RFM attributes and compute the customer profitability based on formula by Miglautsch (2000). The detail process of this step is expressed:

(a) Define the scaling of three R–F–M attributes, which are 5, 4, 3, 2 and 1 that refer to the customer contributions to revenue for enterprises. The '5' refers to the most customer contribution to revenue and '1' refers to the least contribution to revenue.

(b) Sort the data of three R–F–M attributes by descendant order.

(c) Partition the three R–F–M attributes respectively into 5 equal parts and each part is equal to 20% of all. The five parts are assigned 5, 4, 3, 2 and 1 score by descendant order (see Table 4).

(d) Let *Rank* be the total RFM score of a customer, w_R, w_F and w_M relative importance of the RFM variables and R, F and M Recency, Frequency and Monetary quintiles as yielded in process (c) respectively. Then compute the profitability value with formula $Rank = w_R R + w_F F + w_M M$, where w_R was set to 3, w_F to 2 and w_M to 1. It expresses weight of each of the recency, frequency and monetary attribute, respectively. This leaves us with 25 customer segments, where the lowest value of customer's profitability is 6 and highest 30.

	R – Recency (%)	F – Frequency (%)	M – Monetary (%)
5 Score	0–20	0–20	0–20
4 Score	20–40	20–40	20–40
3 Score	40–60	40–60	40–60
2 Score	60–80	60–80	60–80
1 Score	80–100	80–100	80–100

Table 4: The scaling of RFM attributes, Cheng, Chen (2009)

3. Discretization

For the application of Decision trees and Apriori algorithm, continuous values that are present in data are discretized into bins. There are two methods used and compared for discretization, so one of those is in the end chosen for interpreting results and creating bins.

4. Generate association rules by Apriori algorithm and build decision tree as a predictive model for achieving an data insight and analytics

5. Evaluate the results of different data mining methods used. Finally, analyze the difference of these different methods, explore the reasons and propose the conclusions.

3.2 Operationalization of variables

The Table below summarizes operationalization of the depended variable and also independent variables. It also states how the variable is measured and summarizes hypothesized impact of the independent variables on dependent variables.

Dependent Variable	**Measured as**	
Customer Profitability	RFM score	
Independent Variables	**Measured as**	**Hypothesized Impact on Customer Profitability**
Relationship duration	Days	+
Cross-buying	Number of product categories shopped in / genres attended	+
Loyalty Instrument	Signing for newsletter / ownership of member card, 1 = yes, 0 = no	+
Focused buying	Ration of number of past purchases in the focal category (category with maximum purchases) to the total number of purchases.	No directional hypothesis
Age	Age of individuals in years	+
Gender	1= Male, 0 = Female	+

Dependent Variable	**Measured as**	
Relationship duration	Days	
Independent Variables	**Measured as**	**Hypothesized Impact on Relationship duration**
Cross-buying	Number of product categories shopped in / genres attended	+
Loyalty Instrument	Signing for newsletter / ownership of membership card, 1 = yes, 0 = no	+

Table 5: Operationalization of the variable

The description of dependent and independent variables is given in the table below.

Variable	Description
Customer relationship duration	This variable indicates how many days a customer already buys products or services from the firm. The use of this type of measurement is well-supported in several studies (Van den Poel, 2003). Relationship duration is measured as the number of months between the customer's first purchase and the last known purchase
Cross-buying	Measured as the total number of distinct product categories that the customer purchased in overall purchase occasions
Gender	Dummy variable 0/1. Women coded as 1, Men as 0.
Loyalty instrument	Loyalty instrument is operationalized as dummy variable 1/0, where 1 = signing for newsletter in Webshop / ownership of a membership card
Age	Variable indicates the age of the customer transformed from the birth date available in the database
Price Paid	Price paid represents the inclination of the customer to pay higher prices for the goods he's purchasing or buy premium services. Variable is operationalized as an average price of the product/service from all the purchases he made.
Amount spent per order	We found it interesting to look at the average amount of money the customer is spending for his order / purchase
Customer profitability	Variable computed for each customer according to their Recency, Frequency and Monetary

Table 6: Description of the variables

3.3 Data collection and preparation (pre-processing)

As said before, in this study data from two companies, Webshop and Theater are used. The customer purchase history for Theater is available for a period starting from 2005 to 2010 and for the Webshop starting from 2008 to 2010. All information was provided by Webshop and Theater residing in the Netherlands, whose databases were stored on the servers of the Total Internet Group company. It all concerned unprocessed transactional data at the individual customer so a lot of effort was put in preparing it for analysis.

The data we obtained were of reasonable quality, however with some missing values or ambiguities. We needed to transform and prepare or delete such records. After deleting cases with too many missing values on key variables, we obtained a sample size of *15 882* customers in sample 1 (Theather) and *19 441* in sample 2 (Webshop).

Before the data can be used for the actual data mining process, it needs to be cleaned and prepared in a required format. These data cleansing and correction tasks included:

- Discovering and repairing inconsistent data formats and inconsistent data encoding, spelling errors, abbreviations and punctuation.
- Deleting unwanted data fields.
- Interpreting codes into text or replacing text into meaningful numbers.
- Combining data, for instance the customer data, from multiple tables into one common variable.

For this task data transformations in ETL tool Kettle was used as well as MS Excel. More specifically, the following data preparations and pre-processing were performed during this research:

- Adding computed fields as inputs or targets.
- Discretization of the data - mapping continuous values into bins.
- Normalization of the variables.
- Checking abnormal, out of bounds or ambiguous values. Some of these outliers may be correct but this is highly unusual and thus hard to explain.
- Checking missing data fields or fields that have been replaced by a default value.
- Converting nominal data (for example yes/no in subscription for loyalty instrument) to metric scales.
- Converting from textual numeric data (e.g. Gender).

New fields important for our analyses had to be generated through combinations of frequencies, summations, averages and minimum/maximum values. The purpose of this approach is to reduce the number of variables and most of all gain the variables we want. Where there is large amount of data it is also useful to apply data reduction techniques

(data cube aggregation, dimension reduction, discretization). Especially discretization is very important to us, in order to gain good results of the Decision tree and Apriori algorithm, so we will examine it more detail in the next section.

Snapshots of the raw data and data transformation process are shown on the figures below.

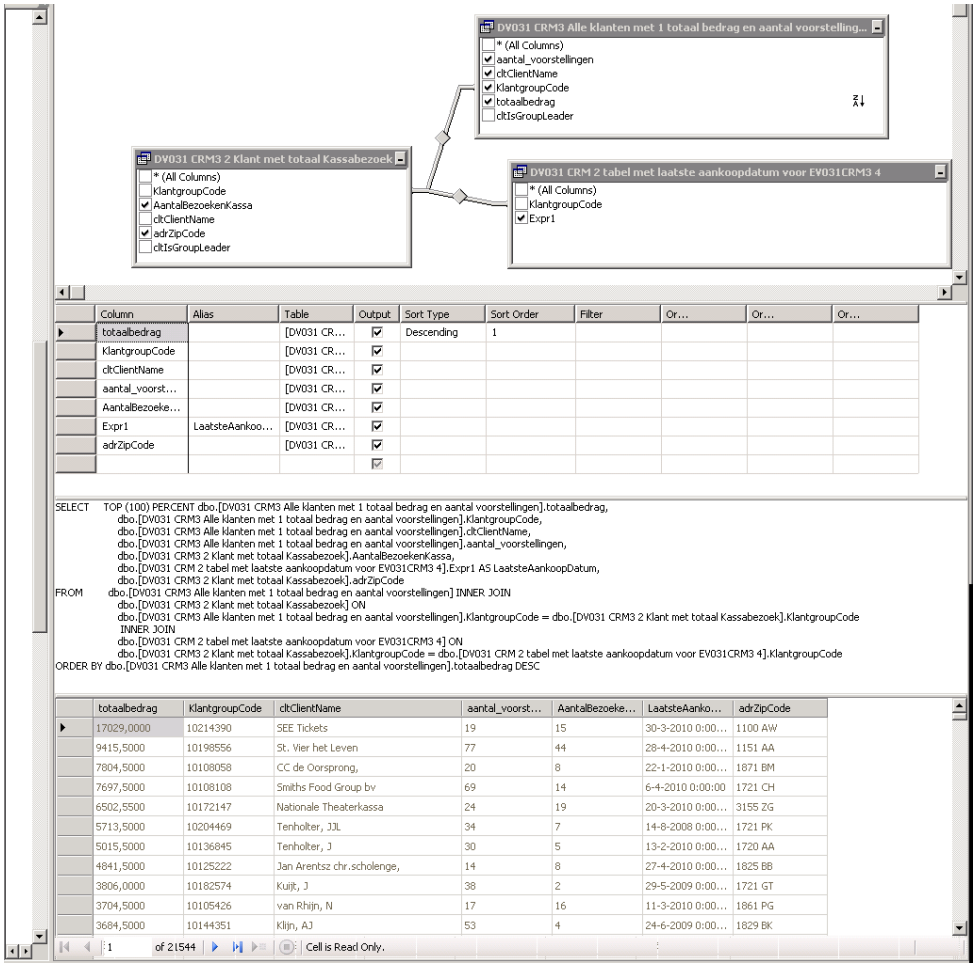
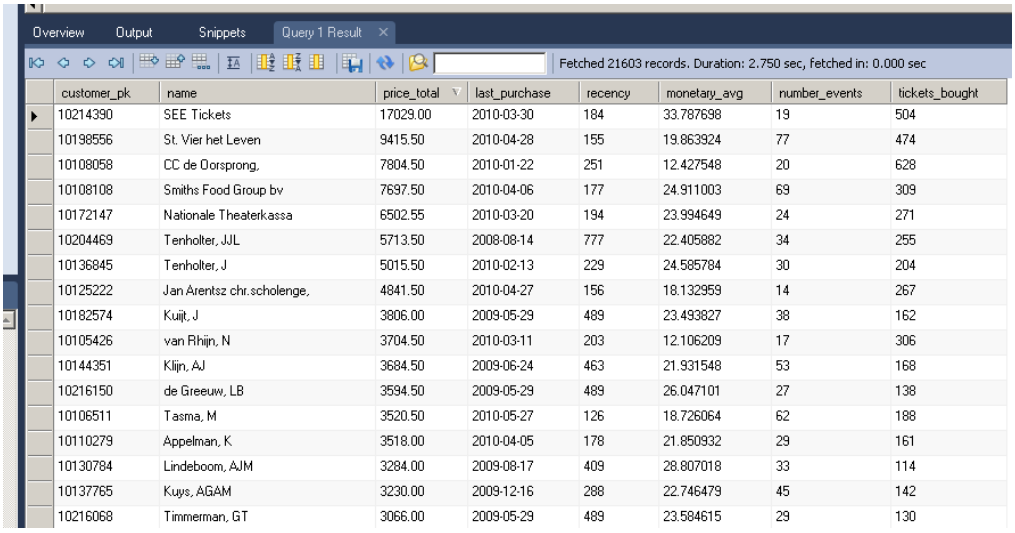


Figure 12: Raw data during transformation process

Data during transformation process and aggregations were loaded into data warehouse. Data as stored in database after several transformation steps are illustrated on the Figure 13.



Overview | Output | Snippets | Query 1 Result

Fetched 21603 records. Duration: 2.750 sec, fetched in: 0.000 sec

customer_pk	name	price_total	last_purchase	recency	monetary_avg	number_events	tickets_bought
10214390	SEE Tickets	17029.00	2010-03-30	184	33.787698	19	504
10198556	St. Vier het Leven	9415.50	2010-04-28	155	19.863924	77	474
10108058	CC de Oorsprong,	7804.50	2010-01-22	251	12.427548	20	628
10108108	Smiths Food Group bv	7697.50	2010-04-06	177	24.911003	69	309
10172147	Nationale Theaterkassa	6502.55	2010-03-20	194	23.994649	24	271
10204469	Tenholter, JJL	5713.50	2008-08-14	777	22.405882	34	255
10136845	Tenholter, J	5015.50	2010-02-13	229	24.585784	30	204
10125222	Jan Arentsz chr.scholenge,	4841.50	2010-04-27	156	18.132959	14	267
10182574	Kuijt, J	3806.00	2009-05-29	489	23.493827	38	162
10105426	van Rhijn, N	3704.50	2010-03-11	203	12.106209	17	306
10144351	Klijn, AJ	3684.50	2009-06-24	463	21.931548	53	168
10216150	de Greeuw, LB	3594.50	2009-05-29	489	26.047101	27	138
10106511	Tasma, M	3520.50	2010-05-27	126	18.726064	62	188
10110279	Appelman, K	3518.00	2010-04-05	178	21.850932	29	161
10130784	Lindeboom, AJM	3284.00	2009-08-17	409	28.807018	33	114
10137765	Kuys, AGAM	3230.00	2009-12-16	288	22.746479	45	142
10216068	Timmerman, GT	3066.00	2009-05-29	489	23.584615	29	130

Figure 13: Data as stored in the database

The data discretization and mining techniques were performed on the data from the data warehouse; the model of the data warehouse is shown on the picture below.

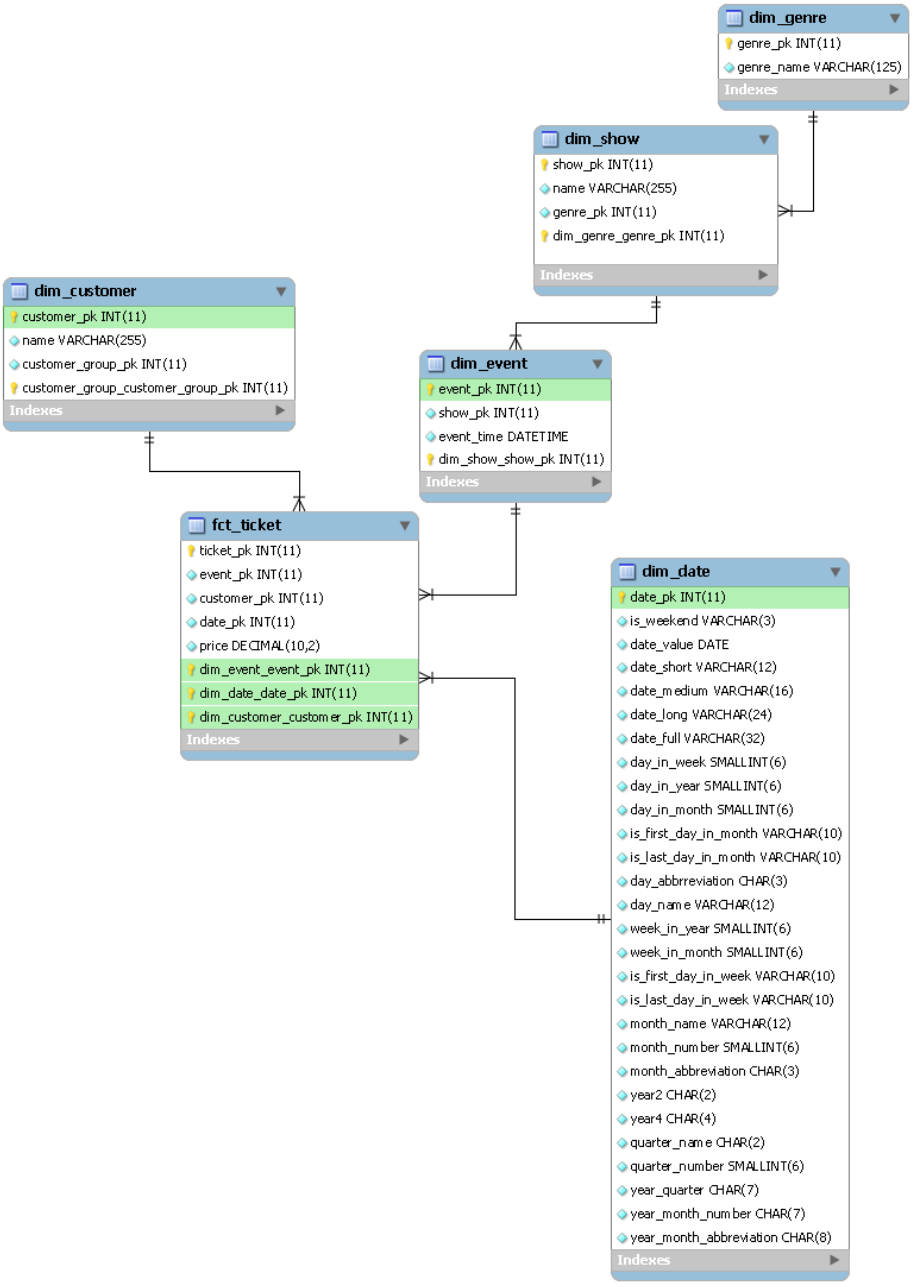


Figure 14: Data warehouse data model

3.3.1 Data discretization

Discretization is a technique for converting a possibly large number of values into a smaller set of values. Kohavi and Sahami (1995) compared different global discretization methods that used error-based and entropy-based criteria. They stated “naïve discretization of the data can be potentially disastrous for data mining as critical information may be lost due to the formation of inappropriate bin boundaries”.

It is said there is no "best" number of bins, and different bin sizes can reveal different features of the data. However, research has attempted to determine an optimal number of bins, but these methods of optimal binning generally make strong assumptions about the shape of the distribution. Depending on the actual data distribution and the goals of the analysis, different bin widths may be appropriate, so experimentation is usually needed to determine an appropriate width.

That's why we conducted number of experiments. We tried optimal binning algorithm but the result correctness was not sufficient, but with discretization to 3 groups with equal frequency (chosen experimentally, trial-error basis) we reached 68% correctness of the decision tree with even 73% correctly classified most profitable customers (profitable tier three).

In our study, for the experiments how to choose the optimal bin sizes, we used histogram-based method and compared the results. In the end, we used entropy-based multi-interval discretization based on Minimum Description Length Principle (MDLP) introduced by Fayyad and Irani (1993) because of the good results it has shown in study by Perner and Trausch (1998) and also it conveniently implemented in SPSS statistical software, which was also utilized for this task. Then the results were compared with the equal frequency method and based on that the final binning method was selected. We will discuss it in more detail further in the thesis. MDLP chooses cut-points to minimize entropy in the resulting bins. Entropy is a measure of the diversity of the bin on the nominal variable. Entropy is minimized (0) when there is only one nominal variable category represented by the cases in the bin. Entropy for a bin will be larger when there are multiple categories of nearly equal sizes.

Prior to analysis, data accuracy and consistency must be ensured to obtain truthful results. Generally, some useful variables can be hidden in a large quantity of raw data, such as customer behavioral variables (RFM), which are hidden in customer and transaction databases, and can be extracted from data integration and transformation. Since the data required for analyzing *association rules* as one of our utilized technique must be discrete, continuous variables are transformed to discrete variables, and a simple 3-4-5 rule is applied (Han & Kamber, 2001).

The screenshots from the experiments with the discretization are shown in the pictures below. They show specific distribution of the variables as an output from tool Weka.

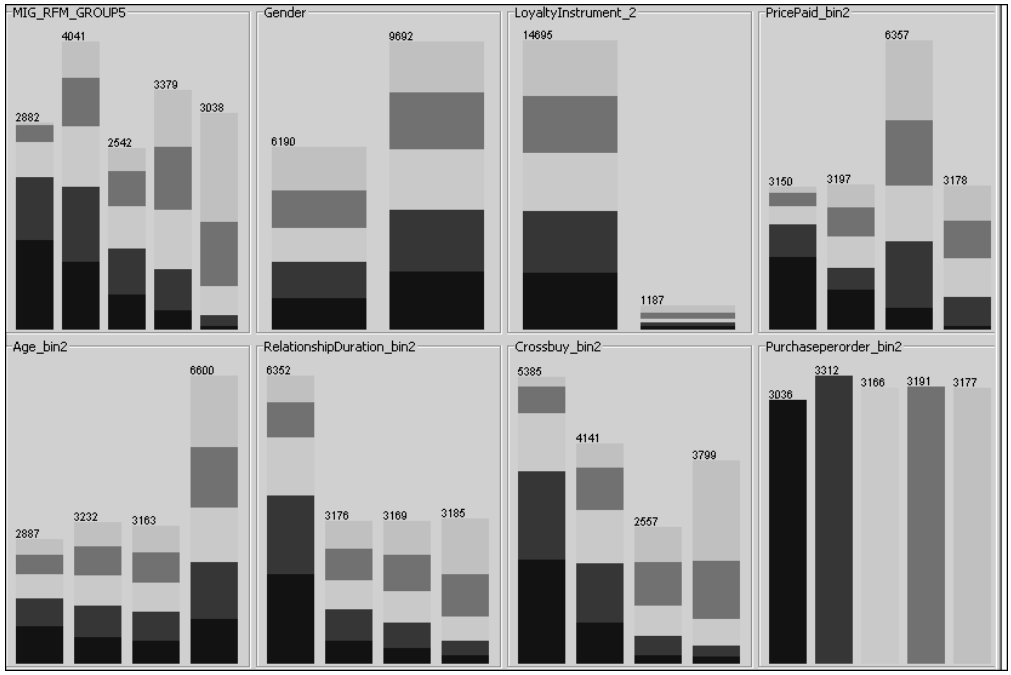


Figure 15: binning experiment Theater dataset

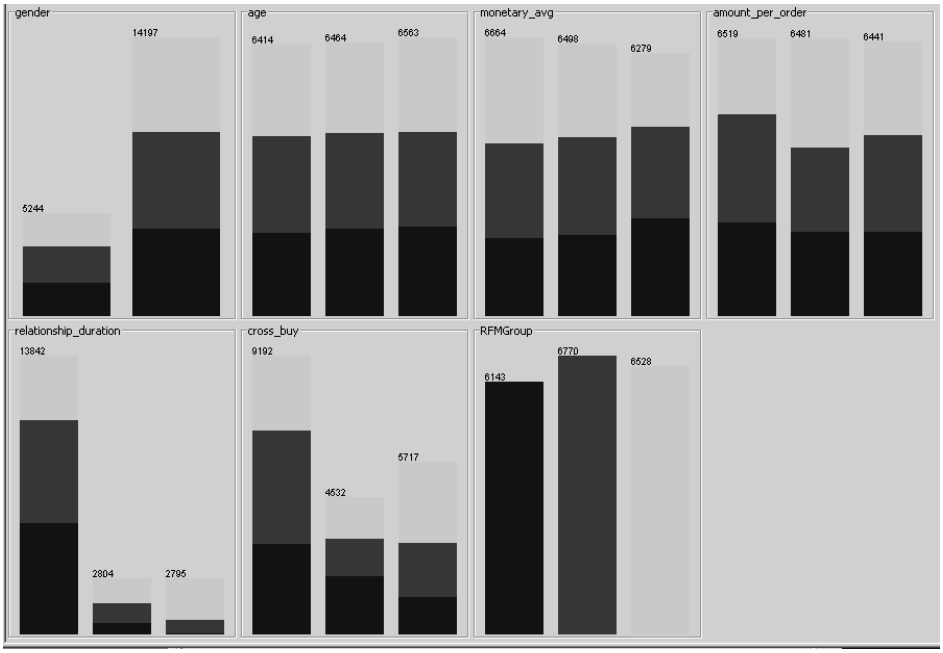


Figure 16: Binning experiment Webshop dataset

3.3.2 Customer valuation and segmentation

According to the RFM variables mentioned in previous section, the total customer value scores for individual customers are calculated to analyze market segmentation and target marketing. Recency of purchase is a more important indicator than purchase frequency and average spending per visit; hence, recency is assigned a greater weight than frequency and monetary amount (Scale by Miglautsch, 2000). Consequently, the maximum score of individual customer value is 30 (i.e. 3*5 + 2*5 + 1*5), whereas the minimum score is 6.

After data pre-processing, customers are segmented into various target markets in terms of customer value obtained by scoring RFM. Scale from 6 – 30 is used for PLS path modeling, however for purposes of decision trees and association rules mining, we will segment customers further into 3 groups according to tiers by Zeithmal (2000).

4. Empirical study – analysis and discussions of research findings

In this chapter we are going to present, analyze and discuss results of the different data mining methods for both datasets - the Theater and Webshop. To begin with, we are summarizing descriptive statistics about the data.

Theater (N = 15 882)

Construct	Min	Max	Mean	St. dev
Profitability	6	30	18.125	6.992
Cross-buying	1	18	3.644	3.004
Relationship duration	0	1953	697.833	695.99
Price paid	0	59.5	20.947	6.269
Amount spent per purchase	0	1262.5	109.145	94.232
Age	1	104	48.125	14.408
Loyalty Instrument*				
Gender*				

* = Dummy variables

Webshop (N = 19 441)

Construct	Min	Max	Mean	St. dev
Profitability	6	30	18.176	6.074
Cross-buying	1	26	2.306	2.044
Relationship duration	0	1026	74.509	171.509
Price paid	0	159	13.636	10.257
Amount spent per purchase	0	2897	37.999	37.312
Age	0	96	37.668	12.075
Loyalty Instrument*				
Gender*				

* = Dummy variables

Table 7: Descriptive Statistics

In order to test the hypothesis that are part of our research, 15 hypothesis were tested in 2 data sets. The dependent variables for PLS method - Profitability, Price paid, Relationship duration, Cross-buy and Loyalty instrument - were tested. We estimated the path coefficients (Beta) and R square values of structural model. Complete results can be found in Table 8 and Table 9 and Figure 18 and Figure 17. At the end also comparison of the two industries is conducted as shown in Table 10.

4.1 Theater dataset PLS analysis results

The results of the analysis are reported in Table 8. For Theater dataset, hypothesis H1, H3, H4, H5, H6, H9 and H11 were supported. The PLS-graph path coefficients were, respectively 0.640, 0.193, 0.741, 0.511, 0.145, 0.256, 0.294, which are statistically significant (at p-value 0.05). This means that study finds positive and strongest effect of relationship duration on profitability and the strongest influence of cross-buy to relationship duration. However, no significant evidence of the influence of other customer characteristics on profitability was found.

The model with the Beta values using PLS method is depicted on the picture below.

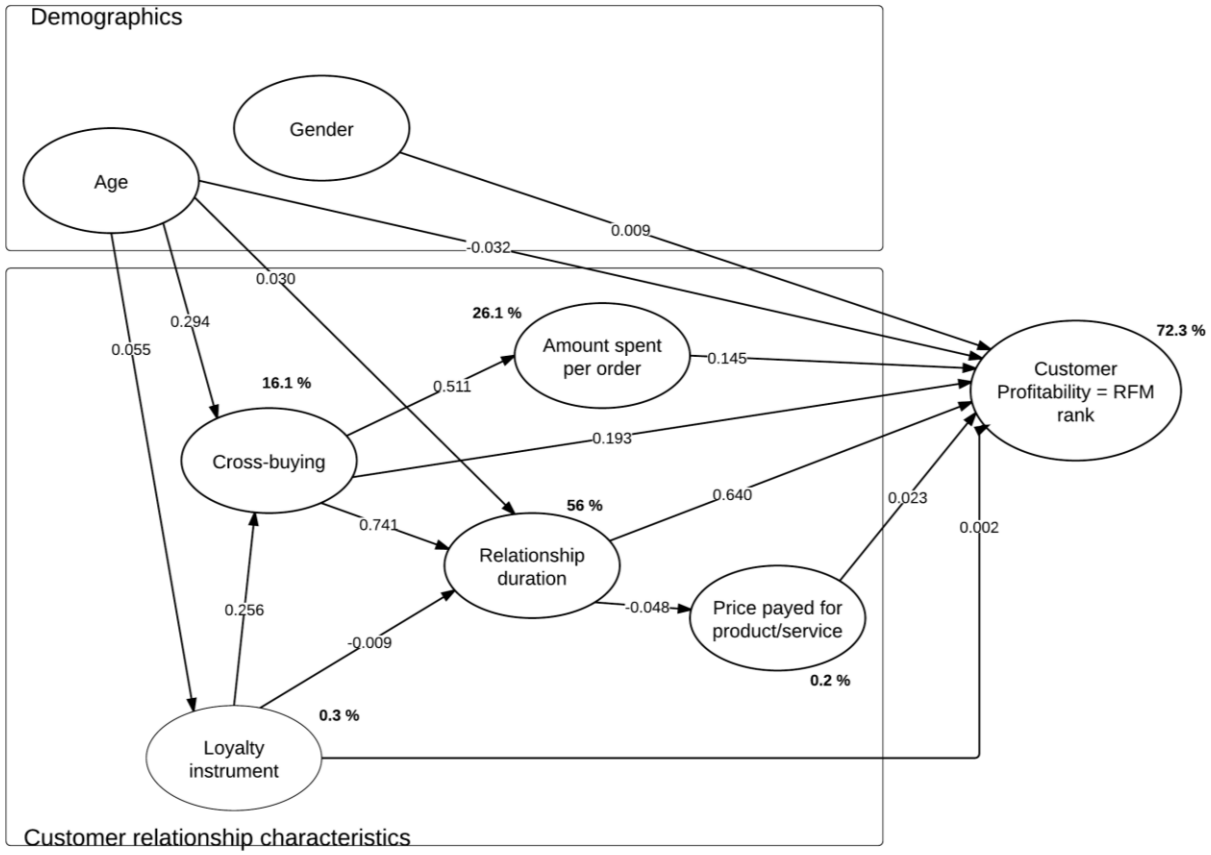


Figure 17: Model fit for Theater dataset using PLS method

From the following table you can find how well the model fits the data.

Hyp.	Path	Beta	t-value	P-value	Supported
1	Relationship duration -> Profitability	0.640	100.469	0.000	YES
2	Relationship duration -> Price paid	-0.048	-6.096	0.000	
3	Cross-buy -> Profitability	0.193	26.019	0.000	YES
4	Cross-buy -> Relationship duration	0.741	129.023	0.000	YES
5	Cross-buy -> Amount per order	0.511	74.831	0.000	YES
6	Amount per order -> Profitability	0.145	28.060	0.000	YES
7	Loyalty instrument ->	0.002	0.456	0.648	n.s.

	Profitability				
8	Loyalty instrument -> Relationship duration	-0.009	-1.642	0.101	n.s.
9	Loyalty instrument -> Cross-buy	0.256	35.169	0.000	YES
10	Price paid -> Profitability	0.023	5.161	0.000	
11	Age -> Profitability	-0.032	-7.127	0.000	
12	Age -> Cross-buy	0.294	40.440	0.000	YES
13	Age -> Relationship duration	0.030	5.377	0.000	
14	Age -> Loyalty instrument	0.055	6.898	0.000	
15	Gender -> Profitability	0.009	2.023	0.043	

Table 8: PLS Path coefficients and t-tests (n=15882)

4.2 Webshop dataset PLS method results

The results of the analysis of the Webshop dataset are reported in Table 9 and illustrated in Figure 18. For Webshop dataset, hypothesis H1, H3, H4, H5, H6 were supported. The PLS-graph path coefficients were, respectively 0.438, 0.178, 0.530, 0.260, 0.230, which are statistically significant (at p-value 0.05). Again, we find a strong influence of relationship duration on profitability and cross-buying on relationship duration, which makes it interesting finding in both datasets.

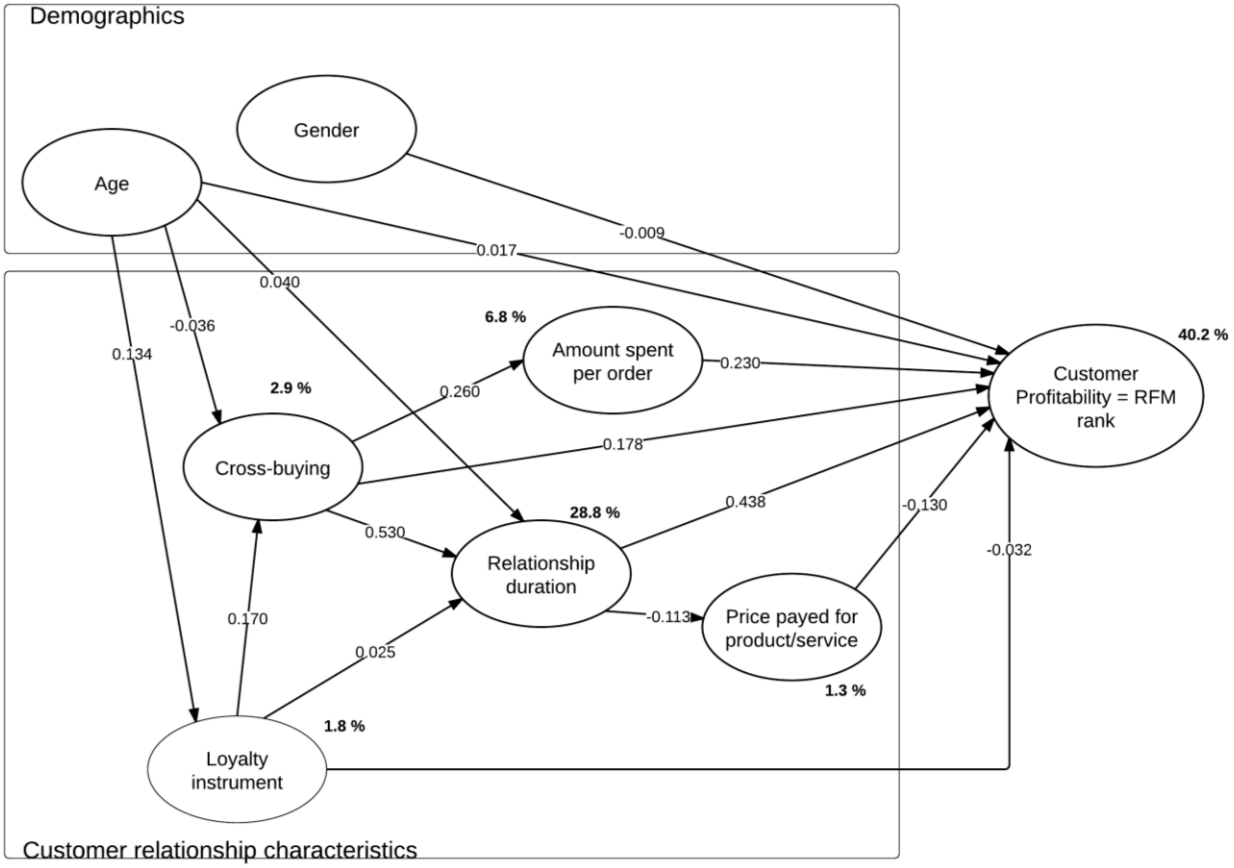


Figure 18: Model fit for Webshop dataset using PLS method

From the following table you can find how well the model fits the data. Overall, the results strongly confirm the predictive power of the model. The amount of variance explained is high, implying a good fit to the data. Four significant paths are of substantial strength (>0.19; Chin, 1998b), while significant but smaller effects are reported for ten paths.

Hyp.	Path	Beta	t-value	Significance	Supported
1	Relationship duration -> Profitability	0.438	65.522	0.000	YES
2	Relationship duration -> Price paid	-0.113	-15.870	0.000	
3	Cross-buy -> Profitability	0.178	24.687	0.000	YES
4	Cross-buy -> Relationship duration	0.530	86.358	0.000	YES
5	Cross-buy -> Amount per order	0.260	37.567	0.000	YES

6	Amount per order -> Profitability	0.230	37.404	0.000	YES
7	Loyalty instrument -> Profitability	-0.032	-5.705	0.000	
8	Loyalty instrument -> Relationship duration	0.025	3.987	0.000	
9	Loyalty instrument -> Cross-buy	0.170	23.892	0.000	
10	Price paid -> Profitability	-0.130	-21.786	0.000	
11	Age -> Profitability	0.017	3.091	0.002	
12	Age -> Cross-buy	-0.036	-4.981	0.000	
13	Age -> Relationship duration	0.040	6.531	0.000	
14	Age -> Loyalty instrument	0.134	18.851	0.000	
15	Gender -> Profitability	-0.009	-1.656	0.098	n.s.

Table 9: PLS Path coefficients and t-tests (n=19 440)

The comparison of the results from both data sets can be seen from the table below. The results of this analysis confirmed that customer value can be understood in terms of Relationship duration, Amount spent per order and Cross-buying.

Comparison Table

Hyp.	**Path**	**Webshop**		**Theather**	
		Beta	**Supported**	**Beta**	**Supported**
1	Relationship duration -> Profitability	0.438	YES	0.640	YES
2	Relationship duration -> Price paid	-0.113		-0.048	
3	Cross-buy -> Profitability	0.178	YES	0.193	YES
4	Cross-buy -> Relationship duration	0.530	YES	0.741	YES
5	Cross-buy -> Amount per order	0.260	YES	0.511	YES
6	Amount per order ->	0.230	YES	0.145	YES

	Profitability				
7	Loyalty instrument -> Profitability	-0.032		0.002	
8	Loyalty instrument -> Relationship duration	0.025		-0.009	
9	Loyalty instrument -> Cross-buy	0.170		0.256	YES
10	Price paid -> Profitability	-0.130		0.023	
11	Age -> Profitability	0.017		-0.032	
12	Age -> Cross-buy	-0.036		0.294	YES
13	Age -> Relationship duration	0.040		0.030	
14	Age -> Loyalty instrument	0.134	YES	0.055	
15	Gender -> Profitability	-0.009		0.009	

Table 10: Comparison of the PLS results

4.3 Discretization of the variables

In order to use Weka software and apply J48 decision trees algorithm, the data discretization was performed as described earlier. Discretization was performed on both datasets; the binning by optimal bin and equal frequency methods was performed in order to the able to select the best discretization of the variables. The binning results are summarized in the tables below.

Method	**Bins**	**Accuracy**	**Tree size**	**Precision Most Profitable Group**
Optimal Bin	Max 1000	71.14 %	679	0.86
Optimal Bin	Max 5	69.56 %	147	0.845
Optimal Bin	Max 4	69.52 %	99	0.859
Optimal Bin	Max 3	66.38 %	47	0.832
Equal Frequency	**3**	**68.70 %**	**51**	**0.83**

Table 11: Binning results Theater

For optimal binning in Webshop dataset we were unable to create bins for the Age variable because of weak association with the guided variable (Profitability).

Webshop

Method	**Bins**	**Accuracy**	**Tree size**	**Precision Most Profitable Group**
Optimal Bin	Max 1000	61.19 %	291	0.732
Optimal Bin	Max 5	56.67 %	105	0.749
Optimal Bin	Max 4	58.6 %	73	0.734
Optimal Bin	Max 3	50.06 %	35	0.772
Equal Frequency	**3**	**58.36 %**	**66**	**0.724**

Table 12: Binning results Webshop

From the tables we can see that MLDP method with no limit for bin amount proved to outperform limited bin amounts and also equal frequency but only by 0.5-3 % with significant increase in decision tree complexity. Binning attributes to 3 groups compared to optimal binning by MDLP has shown approximately the same results.

Based on research of Zeithaml (2000) on profitable customers groups, we decided to divide customers into three groups from least profitable customers to the most.

4.4 Theater dataset decision tree results

After the discretization of the variables and division of the attributes into bins of three (of course regardless of dummy variables), specific distribution of the values of each variable is shown in Table 13.

Construct	**Low**	**Medium**	**High**
Profitability	6 – 14	15 - 21	22 - 30
Cross-buying	0 - 1	2 - 4	>4

Relationship duration	0	1 - 1086	>1086
Price paid	<18	18 - 22	>22
Amount spent per purchase	<56	56 - 184	>184
Age	<41	41 - 54	>55
Loyalty Instrument	Yes (1)	No (0)	
Gender	Female (1)	Male (0)	

Table 13: Distribution of the variables - Webshop

Then the decision trees were built using Weka tool and results of this analysis are described as follows:

- If the relationship duration is the lowest (specifically 0) then the cross-buy can predict quite well the difference between most profitable and least profitable customers.
- If the cross-buy is lowest or even moderate, the customers are least profitable. Only if the cross-buy is the highest and the customers are buying the most expensive goods (tickets), the customers are most profitable, but it still depends on age – less than 40 or more than 55.
- However, if they pay the lowest prices for the goods, they are still in the lowest profitable group.

Another interesting observation is that in the moderate relationship duration sub-tree there is none most profitable customer whatsoever. The decision tree created by Weka can be seen below.

RelationshipDuration = '(-inf-0.5]'
| *Crossbuy = '(-inf-1.5]': 1 (4600.0/1269.0)*
| *Crossbuy = '(1.5-4.5]': 1 (755.0/301.0)*
| *Crossbuy = '(4.5-inf)'*
| | *PricePaid = '(-inf-18.305]': 1 (14.0/5.0)*
| | *PricePaid = '(18.305-22.505]': 2 (24.0/12.0)*
| | *PricePaid = '(22.505-inf)'*
| | | *Age = '(-inf-40.5]': 3 (2.0)*
| | | *Age = '(40.5-54.5]': 2 (10.0/3.0)*

| | | Age = '(54.5-inf)': 3 (10.0/5.0)

RelationshipDuration = '(0.5-1086.5]'

| Crossbuy = '(-inf-1.5]'

| | Purchaseperorder = '(-inf-56.845]'

| | | PricePaid = '(-inf-18.305]': 1 (242.0/102.0)

| | | PricePaid = '(18.305-22.505]': 2 (108.0/49.0)

| | | PricePaid = '(22.505-inf)'

| | | | Age = '(-inf-40.5]': 1 (11.0/2.0)

| | | | Age = '(40.5-54.5]'

| | | | | Gender = 0: 1 (7.0/1.0)

| | | | | Gender = 1: 2 (9.0/4.0)

| | | | Age = '(54.5-inf)': 2 (20.0/9.0)

| | Purchaseperorder = '(56.845-114.98]': 2 (214.0/107.0)

| | Purchaseperorder = '(114.98-inf)': 2 (69.0/22.0)

| Crossbuy = '(1.5-4.5]': 2 (3325.0/1464.0)

| Crossbuy = '(4.5-inf)': 2 (1231.0/473.0)

RelationshipDuration = '(1086.5-inf)'

| Crossbuy = '(-inf-1.5]'

| | Purchaseperorder = '(-inf-56.845]': 2 (44.0/13.0)

| | Purchaseperorder = '(56.845-114.98]': 2 (45.0/14.0)

| | Purchaseperorder = '(114.98-inf)': 3 (16.0/4.0)

| Crossbuy = '(1.5-4.5]'

| | Purchaseperorder = '(-inf-56.845]'

| | | LoyaltyInstrument_2 = 0: 2 (407.0/134.0)

| | | LoyaltyInstrument_2 = 1

| | | | PricePaid = '(-inf-18.305]': 2 (14.0/6.0)

| | | | PricePaid = '(18.305-22.505]': 3 (7.0/1.0)

| | | | PricePaid = '(22.505-inf)': 2 (3.0/1.0)

| | Purchaseperorder = '(56.845-114.98]'

| | | PricePaid = '(-inf-18.305]': 2 (157.0/76.0)

| | | PricePaid = '(18.305-22.505]'

| | | | Age = '(-inf-40.5]': 3 (84.0/39.0)

| | | | Age = '(40.5-54.5]': 3 (90.0/43.0)

| | | | Age = '(54.5-inf)': 2 (70.0/32.0)

| | | PricePaid = '(22.505-inf)'

| | | | LoyaltyInstrument_2 = 0

| | | | | Gender = 0: 2 (107.0/48.0)

| | | | | Gender = 1: 3 (132.0/59.0)

| | | | LoyaltyInstrument_2 = 1: 3 (23.0/7.0)

| | Purchaseperorder = '(114.98-inf)': 3 (342.0/87.0)

| Crossbuy = '(4.5-inf)': 3 (3690.0/457.0)

Number of Leaves : 33

Size of the tree: 51

The prediction model claims that the most profitable customers would be the ones that are in a relationship with company for the longest time (>1086) and attend the most genres (>4). This is in line with the results of PLS analysis. Also if the relationship duration is the longest but the cross-buy is the lowest, the most profitable customers would be those who are purchasing per order the most (>114 euro). If the relationship duration is the longest and cross-buy is moderate (1-3 genres), then the most profitable customers are the ones that have the highest purchase per order (> 114 euro).

Summary table for the decision tree is shown in the table below.

Correctly Classified Instances	10912	68.7067 %
Incorrectly Classified Instances	4970	31.2933 %
Kappa statistic	0.5295	
Mean absolute error	0.2826	
Root mean squared error	0.3771	
Relative absolute error	63.6819 %	
Root relative squared error	80.057 %	
Total Number of Instances	15882	

4.5 Theater dataset Apriori algorithm results

Then the Apriori algorithm was performed on the Theater dataset. Summary table of the algorithm results is shown below.

Minimum support: 0.15 (2382 instances)
Minimum metric <confidence>: 0.7
Number of cycles performed: 17
Generated sets of large itemsets:
Size of set of large itemsets L(1): 21
Size of set of large itemsets L(2): 56
Size of set of large itemsets L(3): 40
Size of set of large itemsets L(4): 5

Table 14: Summary table, Apriori algorithm, Theater dataset

From the best rules found, we can conclude that the least profitable customers and the ones that are the customers for the least time and they don't cross-buy (attend different genres) and don't own loyalty instrument with 98% confidence. Also female least profitable customers with lowest relationship duration don't own loyalty instrument with the same confidence.

Following other insight was gained from the generated rules:

- The least profitable customers who don't cross-buy (and also don't own loyalty instrument -membership card) are with 92% confidence customers for the shortest time.
- If the relationship duration is the highest and purchase per order also the highest we can conclude with 90 % confidence that those customers are in the most profitable group. Another interesting fact which was revealed is that most profitable customers who cross buy a lot (5 and more visited genres) are with 88% confidence customers for the longest time (>1086 days). Also another implication is valid with 88 % confidence – The customers who are with company for the longest time and cross-buy the most are the most profitable.
- Cross-buy and purchase per order, both the least groups imply the shortest relationship duration.

- Another interesting descriptive is that female customers who are customers for the shortest time cross-buy the least. That's not that surprising in consideration that they probably didn't really have time to attend many different shows, more interesting is, that this is with 86 % confidence true for women.

Best rules found:

1. MIG_RFM_GROUP=1 RelationshipDuration='(-inf-0.5]' Crossbuy='(-inf-1.5]' 3331 ==> LoyaltyInstrument_2=0 3255 conf:(0.98)

2. RelationshipDuration='(-inf-0.5]' Crossbuy='(-inf-1.5]' 4600 ==> LoyaltyInstrument_2=0 4494 conf:(0.98)

3. RelationshipDuration='(-inf-0.5]' LoyaltyInstrument_2=0 Purchaseperorder='(-inf-56.845]' 2693 ==> Crossbuy='(-inf-1.5]' 2629 conf:(0.98)

4. MIG_RFM_GROUP=1 Crossbuy='(-inf-1.5]' 3639 ==> LoyaltyInstrument_2=0 3552 conf:(0.98)

5. Gender=1 RelationshipDuration='(-inf-0.5]' Crossbuy='(-inf-1.5]' 3008 ==> LoyaltyInstrument_2=0 2936 conf:(0.98)

6. RelationshipDuration='(-inf-0.5]' Purchaseperorder='(-inf-56.845]' 2767 ==> Crossbuy='(-inf-1.5]' 2699 conf:(0.98)

7. MIG_RFM_GROUP=1 Gender=1 RelationshipDuration='(-inf-0.5]' 2451 ==> LoyaltyInstrument_2=0 2390 conf:(0.98)

8. MIG_RFM_GROUP=1 RelationshipDuration='(-inf-0.5]' 3804 ==> LoyaltyInstrument_2=0 3708 conf:(0.97)

9. RelationshipDuration='(-inf-0.5]' 5415 ==> LoyaltyInstrument_2=0 5278 conf:(0.97)

10. Gender=1 RelationshipDuration='(-inf-0.5]' 3507 ==> LoyaltyInstrument_2=0 3417 conf:(0.97)

11. RelationshipDuration='(-inf-0.5]' Crossbuy='(-inf-1.5]' Purchaseperorder='(-inf-56.845]' 2699 ==> LoyaltyInstrument_2=0 2629 conf:(0.97)

12. Crossbuy='(-inf-1.5]' 5385 ==> LoyaltyInstrument_2=0 5241 conf:(0.97)

13. RelationshipDuration='(-inf-0.5]' Purchaseperorder='(-inf-56.845]' 2767 ==> LoyaltyInstrument_2=0 2693 conf:(0.97)

14. Gender=1 Crossbuy='(-inf-1.5]' 3462 ==> LoyaltyInstrument_2=0 3366 conf:(0.97)

15. Crossbuy='(-inf-1.5]' Purchaseperorder='(-inf-56.845]' 3140 ==> LoyaltyInstrument_2=0 3052 conf:(0.97)

16. MIG_RFM_GROUP=1 5204 ==> LoyaltyInstrument_2=0 5049 conf:(0.97)

17. MIG_RFM_GROUP=1 Gender=1 3334 ==> LoyaltyInstrument_2=0 3229 conf:(0.97)

18. MIG_RFM_GROUP=1 Purchaseperorder='(-inf-56.845]' 2887 ==> LoyaltyInstrument_2=0 2795 conf:(0.97)

19. Age='(-inf-40.5]' 5243 ==> LoyaltyInstrument_2=0 5033 conf:(0.96)

20. Gender=1 Age='(-inf-40.5]' 3690 ==> LoyaltyInstrument_2=0 3537 conf:(0.96)

21. RelationshipDuration='(-inf-0.5]' Purchaseperorder='(-inf-56.845]' 2767 ==> Crossbuy='(-inf-1.5]' LoyaltyInstrument_2=0 2629 conf:(0.95)

22. RelationshipDuration='(0.5-1086.5]' Crossbuy='(1.5-4.5]' 3325 ==> LoyaltyInstrument_2=0 3156 conf:(0.95)

23. MIG_RFM_GROUP=2 Crossbuy='(1.5-4.5]' 2791 ==> LoyaltyInstrument_2=0 2647 conf:(0.95)

24. Crossbuy='(1.5-4.5]' 5516 ==> LoyaltyInstrument_2=0 5228 conf:(0.95)

25. Purchaseperorder='(-inf-56.845]' 5295 ==> LoyaltyInstrument_2=0 5015 conf:(0.95)

26. Gender=1 Purchaseperorder='(-inf-56.845]' 3383 ==> LoyaltyInstrument_2=0 3203 conf:(0.95)

27. PricePaid='(22.505-inf)' 5167 ==> LoyaltyInstrument_2=0 4878 conf:(0.94)

28. Gender=1 Crossbuy='(1.5-4.5]' 3393 ==> LoyaltyInstrument_2=0 3200 conf:(0.94)

29. MIG_RFM_GROUP=2 5670 ==> LoyaltyInstrument_2=0 5343 conf:(0.94)

30. PricePaid='(22.505-inf)' Gender=1 3105 ==> LoyaltyInstrument_2=0 2919 conf:(0.94)

31. MIG_RFM_GROUP=2 Gender=1 3487 ==> LoyaltyInstrument_2=0 3274 conf:(0.94)

32. RelationshipDuration='(0.5-1086.5]' 5236 ==> LoyaltyInstrument_2=0 4914 conf:(0.94)

33. Gender=1 RelationshipDuration='(0.5-1086.5]' 3216 ==> LoyaltyInstrument_2=0 3002 conf:(0.93)

34. Purchaseperorder='(56.845-114.98]' 5291 ==> LoyaltyInstrument_2=0 4937 conf:(0.93)

35. MIG_RFM_GROUP=2 RelationshipDuration='(0.5-1086.5]' 2948 ==> LoyaltyInstrument_2=0 2749 conf:(0.93)

36. Gender=0 6190 ==> LoyaltyInstrument_2=0 5759 conf:(0.93)

37. Gender=1 Purchaseperorder='(56.845-114.98]' 3341 ==> LoyaltyInstrument_2=0 3105 conf:(0.93)

38. PricePaid='(-inf-18.305]' Purchaseperorder='(-inf-56.845]' 2971 ==> LoyaltyInstrument_2=0 2761 conf:(0.93)

39. PricePaid='(18.305-22.505]' 5423 ==> LoyaltyInstrument_2=0 5027 conf:(0.93)

40. Gender=0 Age='(54.5-inf)' 2783 ==> LoyaltyInstrument_2=0 2577 conf:(0.93)

41. Gender=1 9692 ==> LoyaltyInstrument_2=0 8936 conf:(0.92)

42. PricePaid='(18.305-22.505]' Gender=1 3155 ==> LoyaltyInstrument_2=0 2907 conf:(0.92)

43. Age='(54.5-inf)' 5225 ==> LoyaltyInstrument_2=0 4789 conf:(0.92)

44. MIG_RFM_GROUP=1 Crossbuy='(-inf-1.5]' LoyaltyInstrument_2=0 3552 ==> RelationshipDuration='(-inf-0.5]' 3255 conf:(0.92)

45. MIG_RFM_GROUP=1 Crossbuy='(-inf-1.5]' 3639 ==> RelationshipDuration='(-inf-0.5]' 3331 conf:(0.92)

46. PricePaid='(-inf-18.305]' Gender=1 3432 ==> LoyaltyInstrument_2=0 3110 conf:(0.91)

47. PricePaid='(-inf-18.305]' 5292 ==> LoyaltyInstrument_2=0 4790 conf:(0.91)

48. Age='(40.5-54.5]' 5414 ==> LoyaltyInstrument_2=0 4873 conf:(0.9)

49. RelationshipDuration='(1086.5-inf)' Purchaseperorder='(114.98-inf)' 2765 ==> MIG_RFM_GROUP=3 2485 conf:(0.9)

50. Purchaseperorder='(114.98-inf)' 5296 ==> LoyaltyInstrument_2=0 4743 conf:(0.9)

51. Gender=1 Age='(40.5-54.5]' 3560 ==> LoyaltyInstrument_2=0 3187 conf:(0.9)

52. MIG_RFM_GROUP=1 Crossbuy='(-inf-1.5]' 3639 ==> RelationshipDuration='(-inf-0.5]' LoyaltyInstrument_2=0 3255 conf:(0.89)

53. Gender=1 Purchaseperorder='(114.98-inf)' 2968 ==> LoyaltyInstrument_2=0 2628 conf:(0.89)

54. MIG_RFM_GROUP=3 Crossbuy='(4.5-inf)' 3656 ==> RelationshipDuration='(1086.5-inf)' 3233 conf:(0.88)

55. MIG_RFM_GROUP=1 RelationshipDuration='(-inf-0.5]' LoyaltyInstrument_2=0 3708 ==> Crossbuy='(-inf-1.5]' 3255 conf:(0.88)

56. RelationshipDuration='(1086.5-inf)' Crossbuy='(4.5-inf)' 3690 ==> MIG_RFM_GROUP=3 3233 conf:(0.88)

57. MIG_RFM_GROUP=1 RelationshipDuration='(-inf-0.5]' 3804 ==> Crossbuy='(-inf-1.5]' 3331 conf:(0.88)

58. MIG_RFM_GROUP=3 Crossbuy='(4.5-inf)' LoyaltyInstrument_2=0 3042 ==> RelationshipDuration='(1086.5-inf)' 2654 conf:(0.87)

59. Gender=1 Crossbuy='(-inf-1.5]' LoyaltyInstrument_2=0 3366 ==> RelationshipDuration='(-inf-0.5]' 2936 conf:(0.87)

60. RelationshipDuration='(1086.5-inf)' Purchaseperorder='(114.98-inf)' 2765 ==> Crossbuy='(4.5-inf)' 2407 conf:(0.87)

61. Gender=1 Crossbuy='(-inf-1.5]' 3462 ==> RelationshipDuration='(-inf-0.5]' 3008 conf:(0.87)

62. RelationshipDuration='(1086.5-inf)' Crossbuy='(4.5-inf)' LoyaltyInstrument_2=0 3061 ==> MIG_RFM_GROUP=3 2654 conf:(0.87)

63. MIG_RFM_GROUP=3 Purchaseperorder='(114.98-inf)' 3042 ==> LoyaltyInstrument_2=0 2623 conf:(0.86)

64. Crossbuy='(-inf-1.5]' LoyaltyInstrument_2=0 Purchaseperorder='(-inf-56.845]' 3052 ==> RelationshipDuration='(-inf-0.5]' 2629 conf:(0.86)

65. RelationshipDuration='(1086.5-inf)' 5231 ==> LoyaltyInstrument_2=0 4503 conf:(0.86)

66. Crossbuy='(-inf-1.5]' Purchaseperorder='(-inf-56.845]' 3140 ==> RelationshipDuration='(-inf-0.5]' 2699 conf:(0.86)

67. Gender=1 RelationshipDuration='(-inf-0.5]' LoyaltyInstrument_2=0 3417 ==> Crossbuy='(-inf-1.5]' 2936 conf:(0.86)

68. MIG_RFM_GROUP=3 5008 ==> LoyaltyInstrument_2=0 4303 conf:(0.86)

69. Crossbuy='(4.5-inf)' Purchaseperorder='(114.98-inf)' 3302 ==> LoyaltyInstrument_2=0 2833 conf:(0.86)

70. Gender=1 RelationshipDuration='(-inf-0.5]' 3507 ==> Crossbuy='(-inf-1.5]' 3008 conf:(0.86)

During the experiments quite interesting results were obtained when the Apriori algorithm was performed with the Loyalty instrument variable left out from the dataset. Summary table of the algorithm is shown below, as well as overview of the best rules found.

Minimum support: 0.1 (1588 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 18
Generated sets of large itemsets:
Size of set of large itemsets L(1): 20
Size of set of large itemsets L(2): 125
Size of set of large itemsets L(3): 40
Size of set of large itemsets L(4): 6

Table 15: Summary of the Apriori algorithm, Theater dataset, and Loyalty instrument left out

Rule	**Confidence**
(Profit Group = Low) & (Rel. duration = Low) & (Purchase per order = Low) => (Cross buy = Low)	0.98
(Profit Group = High) & (Age = High) & (Cross buy = High) => (Rel. duration = High)	0.92
(Profit Group = Low) & (Cross buy = Low) => (Rel. Duration = Low)	0.92
(Rel. Duration = High) & (Purchase per order = High) => (Profit Group = High)	0.90
(Profit Group = High) & (Rel. Duration = High) & (Purchase per order = High) => (Cross buy = High)	0.89
(Age = High) & (Rel. Duration = High) & (Cross buy = High) => (Profit Group = High)	0.88

Table 16: Best rules using Apriori algorithm, Theater dataset, Loyalty instrument left out

4.6 Webshop dataset decision tree results

We start our analysis of the Webshop dataset decision tree by explaining the most interesting predictions for the most profitable customer group. Relationship duration has the highest information gain, so it's the first node we will use for prediction and we left out age because of the very low information gain.

The results showed that:

- The most profitable customers are most likely customers who are the customers for the longest time (3rd relationship duration group).

- If the relationship duration is 0 (-inf – 0.5) and amount per order is 0 - 19.5 and monetary average < 8.5 (least) and cross-buy is none (1) and newsletter is 0 the these customers are within the least profitable group.
- If the relationship duration is 0 (-inf – 0.5) and amount per order is (0 - 19.5) and monetary average > 14.5 (highest) and cross-buy is either none (1) or highest (2) then the customers fall into least profitable group; if cross-buy is in the middle, than moderate profitable group (2). This result shows us the reasonable influence of the cross-buy to profitability of the customer.

The only occasion where least profitable customer is predicted if the relationship duration is more than zero (group > 1) is the case:

- Relationship duration is middle tern and amount per order is the least and cross buy is the least and customer is signed to newsletter and monetary average is the highest (>14.5).

This confirms our theory for the relationship duration as a strong predictor of customer profitability.

Construct	**Low**	**Medium**	**High**
Profitability	6 - 14	15 - 20	21 - 30
Cross-buying	1	2	>2
Relationship duration	0	1 - 193	>193
Price paid	<9	9 - 14	>14
Amount spent per purchase	<20	20 - 43	>43
Age	<41	41 - 54	>55
Loyalty Instrument	Yes (1)	No (0)	
Gender	Female (1)	Male (0)	

Table 17: Distribution of the variables for Webshop dataset

The full decision tree looks like follows:

relationship_duration = '(-inf-0.5]'
| *amount_per_order = '(-inf-19.5]'*

```
|   |   monetary_avg = '(-inf-8.5]'
|   |   |   cross_buy = '(-inf-1.5]'
|   |   |   |   newsletter <= 0: 2 (1293.0/640.0)
|   |   |   |   newsletter > 0: 1 (228.0/116.0)
|   |   |   cross_buy = '(1.5-2.5]': 1 (512.0/264.0)
|   |   |   cross_buy = '(2.5-inf)': 2 (97.0/55.0)
|   |   monetary_avg = '(8.5-14.5]': 1 (1966.0/881.0)
|   |   monetary_avg = '(14.5-inf)'
|   |   |   cross_buy = '(-inf-1.5]': 1 (1103.0/341.0)
|   |   |   cross_buy = '(1.5-2.5]': 2 (2.0)
|   |   |   cross_buy = '(2.5-inf)': 1 (0.0)
|   amount_per_order = '(19.5-43.5]'
|   |   monetary_avg = '(-inf-8.5]'
|   |   |   newsletter <= 0: 2 (895.0/529.0)
|   |   |   newsletter > 0
|   |   |   |   gender <= 0
|   |   |   |   |   cross_buy = '(-inf-1.5]': 3 (18.0/9.0)
|   |   |   |   |   cross_buy = '(1.5-2.5]': 2 (16.0/8.0)
|   |   |   |   |   cross_buy = '(2.5-inf)': 1 (8.0/4.0)
|   |   |   |   gender > 0: 1 (219.0/123.0)
|   |   monetary_avg = '(8.5-14.5]'
|   |   |   cross_buy = '(-inf-1.5]': 2 (610.0/321.0)
|   |   |   cross_buy = '(1.5-2.5]': 1 (536.0/274.0)
|   |   |   cross_buy = '(2.5-inf)'
|   |   |   |   newsletter <= 0: 2 (118.0/62.0)
|   |   |   |   newsletter > 0: 1 (37.0/20.0)
|   |   monetary_avg = '(14.5-inf)': 1 (1573.0/704.0)
|   amount_per_order = '(43.5-inf)'
|   |   monetary_avg = '(-inf-8.5]'
|   |   |   cross_buy = '(-inf-1.5]': 3 (234.0/96.0)
|   |   |   cross_buy = '(1.5-2.5]': 2 (275.0/122.0)
|   |   |   cross_buy = '(2.5-inf)': 2 (684.0/344.0)
|   |   monetary_avg = '(8.5-14.5]'
|   |   |   newsletter <= 0: 3 (857.0/450.0)
```

| | | newsletter > 0: 2 (296.0/146.0)

| | monetary_avg = '(14.5-inf)'

| | | cross_buy = '(-inf-1.5]': 2 (974.0/484.0)

| | | cross_buy = '(1.5-2.5]'

| | | | gender <= 0: 2 (242.0/136.0)

| | | | gender > 0: 1 (623.0/351.0)

| | | cross_buy = '(2.5-inf)': 2 (426.0/185.0)

relationship_duration = '(0.5-193.5]'

| amount_per_order = '(-inf-19.5]'

| | cross_buy = '(-inf-1.5]'

| | | newsletter <= 0: 2 (269.0/143.0)

| | | newsletter > 0

| | | | monetary_avg = '(-inf-8.5]': 2 (24.0/13.0)

| | | | monetary_avg = '(8.5-14.5]'

| | | | | gender <= 0: 3 (7.0/3.0)

| | | | | gender > 0: 2 (13.0/5.0)

| | | | monetary_avg = '(14.5-inf)': 1 (11.0/4.0)

| | cross_buy = '(1.5-2.5]'

| | | gender <= 0: 2 (65.0/33.0)

| | | gender > 0

| | | | monetary_avg = '(-inf-8.5]': 3 (100.0/50.0)

| | | | monetary_avg = '(8.5-14.5]': 3 (60.0/32.0)

| | | | monetary_avg = '(14.5-inf)': 2 (16.0/7.0)

| | cross_buy = '(2.5-inf)'

| | | newsletter <= 0: 3 (135.0/57.0)

| | | newsletter > 0: 2 (46.0/20.0)

| amount_per_order = '(19.5-43.5]': 3 (1125.0/446.0)

| amount_per_order = '(43.5-inf)': 3 (933.0/237.0)

relationship_duration = '(193.5-inf)': 3 (2795.0/308.0)

Number of Leaves: 41

Size of the tree: 66

From this decision tree we can make following predictions for the most profitable customers:

- If the relationship duration of the customer is moderate (1 - 193.5 days) and amount per order is more than 19.5 euro (2^{nd} and 3^{rd} group) then we can predict the customer will be the most profitable.
- If the female customer is in moderate relationship duration group (1 – 193.5 days) and amount per order is less than19.5 euro and cross-buying is also moderate (1.5 – 2.5) and moreover the customer is not buying the most expensive products (group 1 and 2) then she will eventually become the most profitable customer.
- And one more prediction about the most profitable customers and moderate relationship duration group. If the average money spent per order is the lowest (< 19.5 euro) and cross-buy is the highest and the customer doesn't want to receive newsletter, then the customer is likely to become the most profitable.

4.7 Webshop dataset Apriori algorithm results

Then the Apriori algorithm was performed on the Webshop dataset. The summary table of the algorithm is shown below.

Minimum support: 0.2 (3888 instances)
Minimum metric <confidence>: 0.7
Number of cycles performed: 16
Generated sets of large itemsets:
Size of set of large itemsets L(1): 20
Size of set of large itemsets L(2): 44
Size of set of large itemsets L(3): 15
Size of set of large itemsets L(4): 1

Table 18: Summary of the Apriori algorithm, Webshop dataset

From the results of the algorithm, we could draw following conclusions:

- if a customer signed up for the loyalty instrument (newsletter) and he's the least profitable one, or just the least profitable one, than the relationship duration is shortest (0 in this case, which means the customer made purchase just once) with 97% confidence. Female least profitable customer is also usually in the shortest relationship duration group (96% conf.).
- Cross buying also revealed some interesting rules. Especially with combination with amount spent per order, gender and newsletter. The customers who are not cross-buying (buying only from 1 product category) and female or without newsletter or making small purchases are usually members of the shortest duration group, with confidence 89%, 89%, 90%, respectively.
- Concerning most profitable customer group, Apriori algorithm revealed only two rules with more than 70% confidence. That is, the most profitable customer is female with 74 % confidence and the most profitable customer doesn't want to receive newsletters (73 % conf.).

These conclusions are summarized in the best rules found table below.

Rule	**Confidence**
(Newsletter = 0) & (Profit Group = Low) => (Rel. duration = Low)	0.97
(Profit Group = Low) => (Rel. duration = Low)	0.97
(Purchase per order = Low) & (Cross buy = Low) => (Rel. duration = Low)	0.90
(Gender = Female) & (Cross buy = Low) => (Rel. duration = Low)	0.89
(Profit Group = Low) => (Newsletter = 0)	0.78
(Cross buy = High) => (Gender = Female)	0.77
(Cross buy = Low) => Rel. Duration = Low & (Newsletter = 0)	0.74
(Profit Group = High) => (Gender = Female)	0.74
(Profit Group = High) => (Newsletter = 0)	0.73
(Price paid = High) => (Gender = Female)	0.71

Table 19: Best rules found for Webshop dataset

Best rules found – Weka output:

1. newsletter=0 MIG_GROUP3=1 4403 ==> relationship_duration='(-inf-0.5]' 4256 conf:(0.97)

2. MIG_GROUP3=1 5632 ==> relationship_duration='(-inf-0.5]' 5438 conf:(0.97)

3. gender=1 MIG_GROUP3=1 4089 ==> relationship_duration='(-inf-0.5]' 3943 conf:(0.96)

4. amount_per_order='(-inf-19.5]' cross_buy='(-inf-1.5]' 4950 ==> relationship_duration='(-inf-0.5]' 4456 conf:(0.9)

5. gender=1 cross_buy='(-inf-1.5]' newsletter=0 5216 ==> relationship_duration='(-inf-0.5]' 4684 conf:(0.9)

6. gender=1 cross_buy='(-inf-1.5]' 6415 ==> relationship_duration='(-inf-0.5]' 5741 conf:(0.89)

7. cross_buy='(-inf-1.5]' newsletter=0 7602 ==> relationship_duration='(-inf-0.5]' 6763 conf:(0.89)

8. cross_buy='(-inf-1.5]' 9192 ==> relationship_duration='(-inf-0.5]' 8140 conf:(0.89)

9. amount_per_order='(-inf-19.5]' relationship_duration='(-inf-0.5]' 5201 ==> cross_buy='(-inf-1.5]' 4456 conf:(0.86)

10. age='(-inf-30.5]' relationship_duration='(-inf-0.5]' 4679 ==> newsletter=0 3997 conf:(0.85)

11. amount_per_order='(-inf-19.5]' cross_buy='(-inf-1.5]' 4950 ==> newsletter=0 4186 conf:(0.85)

12. newsletter=0 MIG_GROUP3=2 5658 ==> relationship_duration='(-inf-0.5]' 4781 conf:(0.84)

13. amount_per_order='(-inf-19.5]' relationship_duration='(-inf-0.5]' 5201 ==> newsletter=0 4375 conf:(0.84)

14. age='(-inf-30.5]' 6414 ==> newsletter=0 5352 conf:(0.83)

15. relationship_duration='(-inf-0.5]' cross_buy='(-inf-1.5]' 8140 ==> newsletter=0 6763 conf:(0.83)

16. amount_per_order='(-inf-19.5]' 6519 ==> newsletter=0 5405 conf:(0.83)

17. MIG_GROUP3=2 7122 ==> relationship_duration='(-inf-0.5]' 5894 conf:(0.83)

18. cross_buy='(-inf-1.5]' 9192 ==> newsletter=0 7602 conf:(0.83)

19. gender=1 MIG_GROUP3=2 5134 ==> relationship_duration='(-inf-0.5]' 4245 conf:(0.83)

20. gender=1 age='(-inf-30.5]' 4847 ==> newsletter=0 3980 conf:(0.82)

21. gender=1 relationship_duration='(-inf-0.5]' cross_buy='(-inf-1.5]' 5741 ==> newsletter=0 4684 conf:(0.82)

22. gender=1 cross_buy='(-inf-1.5]' 6415 ==> newsletter=0 5216 conf:(0.81)

23. relationship_duration='(-inf-0.5]' MIG_GROUP3=2 5894 ==> newsletter=0 4781 conf:(0.81)

24. amount_per_order='(-inf-19.5]' newsletter=0 5405 ==> relationship_duration='(-inf-0.5]' 4375 conf:(0.81)

25. gender=0 5244 ==> newsletter=0 4225 conf:(0.81)

26. monetary_avg='(14.5-inf)' newsletter=0 4880 ==> relationship_duration='(-inf-0.5]' 3909 conf:(0.8)

27. amount_per_order='(-inf-19.5]' 6519 ==> relationship_duration='(-inf-0.5]' 5201 conf:(0.8)

28. relationship_duration='(-inf-0.5]' 13842 ==> newsletter=0 11027 conf:(0.8)

29. MIG_GROUP3=2 7122 ==> newsletter=0 5658 conf:(0.79)

30. monetary_avg='(14.5-inf)' relationship_duration='(-inf-0.5]' 4943 ==> newsletter=0 3909 conf:(0.79)

31. monetary_avg='(14.5-inf)' 6279 ==> relationship_duration='(-inf-0.5]' 4943 conf:(0.79)

32. gender=1 relationship_duration='(-inf-0.5]' 10002 ==> newsletter=0 7836 conf:(0.78)

33. relationship_duration='(-inf-0.5]' MIG_GROUP3=1 5438 ==> newsletter=0 4256 conf:(0.78)

34. gender=1 MIG_GROUP3=2 5134 ==> newsletter=0 4015 conf:(0.78)

35. MIG_GROUP3=1 5632 ==> newsletter=0 4403 conf:(0.78)

36. monetary_avg='(14.5-inf)' 6279 ==> newsletter=0 4880 conf:(0.78)

37. monetary_avg='(8.5-14.5]' 6498 ==> newsletter=0 5040 conf:(0.78)

38. amount_per_order='(-inf-19.5]' newsletter=0 5405 ==> cross_buy='(-inf-1.5]' 4186 conf:(0.77)

39. cross_buy='(2.5-inf)' 5717 ==> gender=1 4415 conf:(0.77)

40. age='(30.5-41.5]' 6464 ==> newsletter=0 4989 conf:(0.77)

41. amount_per_order='(-inf-19.5]' 6519 ==> cross_buy='(-inf-1.5]' 4950 conf:(0.76)

42. amount_per_order='(19.5-43.5]' 6481 ==> newsletter=0 4918 conf:(0.76)

43. age='(-inf-30.5]' 6414 ==> gender=1 4847 conf:(0.76)

44. MIG_GROUP3=1 5632 ==> relationship_duration='(-inf-0.5]' newsletter=0 4256 conf:(0.76)

45. gender=1 14197 ==> newsletter=0 10710 conf:(0.75)

46. monetary_avg='(-inf-8.5]' 6664 ==> newsletter=0 5015 conf:(0.75)

47. amount_per_order='(19.5-43.5]' 6481 ==> gender=1 4859 conf:(0.75)

48. age='(-inf-30.5]' newsletter=0 5352 ==> relationship_duration='(-inf-0.5]' 3997 conf:(0.75)

49. MIG_GROUP3=3 6687 ==> gender=1 4974 conf:(0.74)

50. age='(-inf-30.5]' newsletter=0 5352 ==> gender=1 3980 conf:(0.74)

51. monetary_avg='(-inf-8.5]' 6664 ==> gender=1 4949 conf:(0.74)

52. age='(30.5-41.5]' 6464 ==> gender=1 4789 conf:(0.74)

53. newsletter=0 14935 ==> relationship_duration='(-inf-0.5]' 11027 conf:(0.74)

54. monetary_avg='(8.5-14.5]' 6498 ==> gender=1 4786 conf:(0.74)

55. cross_buy='(-inf-1.5]' 9192 ==> relationship_duration='(-inf-0.5]' newsletter=0 6763 conf:(0.74)

56. gender=1 newsletter=0 10710 ==> relationship_duration='(-inf-0.5]' 7836 conf:(0.73)

57. gender=1 cross_buy='(-inf-1.5]' 6415 ==> relationship_duration='(-inf-0.5]' newsletter=0 4684 conf:(0.73)

58. amount_per_order='(-inf-19.5]' 6519 ==> gender=1 4756 conf:(0.73)

59. age='(-inf-30.5]' 6414 ==> relationship_duration='(-inf-0.5]' 4679 conf:(0.73)

60. MIG_GROUP3=3 6687 ==> newsletter=0 4874 conf:(0.73)

61. MIG_GROUP3=1 5632 ==> gender=1 4089 conf:(0.73)

62. relationship_duration='(-inf-0.5]' MIG_GROUP3=1 5438 ==> gender=1 3943 conf:(0.73)

63. relationship_duration='(-inf-0.5]' 13842 ==> gender=1 10002 conf:(0.72)

64. MIG_GROUP3=2 7122 ==> gender=1 5134 conf:(0.72)

65. relationship_duration='(-inf-0.5]' MIG_GROUP3=2 5894 ==> gender=1 4245 conf:(0.72)

66. newsletter=0 14935 ==> gender=1 10710 conf:(0.72)

67. amount_per_order='(43.5-inf)' 6441 ==> newsletter=0 4612 conf:(0.72)

68. amount_per_order='(43.5-inf)' 6441 ==> relationship_duration='(-inf-0.5]' 4611 conf:(0.72)

69. amount_per_order='(43.5-inf)' 6441 ==> gender=1 4582 conf:(0.71)

70. monetary_avg='(14.5-inf)' 6279 ==> gender=1 4462 conf:(0.71)

4.8 Discussion of Apriori algorithm results

The results of Apriori algorithm applied to the data were interesting in both cases. Age surprisingly was not significant at all. However, we could witness the relationship duration phenomenon – both long term and short term customers fell into the most profitable category, which confirms theory that both long term and short term customers in non-contractual settings could be profitable.

4.9 Discussion of different output classes

In this study, a discussion about the issue of different classes on input and output is conducted. Using Webshop dataset, the classification accuracy is 58.03 %, 49.42 % and 42.45 % in 3, 4 and 5 classes on output, respectively. We can clearly see that the less the output classes are, the higher is accuracy. The lower the output classes, the lower is the selection of target customers, which means there has to be a trade-off selection for managers if they prefer higher accuracy of the results or better ability to target customers (Cheng, Chen, 2009).

During the experiments we compared following discretization methods:

- Optimal binning method vs equal frequency
- Output bins – 3 vs 4 vs 5 – with the results that the less bins the better accuracy
- Regular classes bins – 3 vs 4 vs 5 – with the result the more bins the better accuracy

These results confirm findings of Cheng, Chen (2009).

4.10 Discussion of the computing process

Computing process using both datasets can be summarized once again as follows:

1) Data preprocessing

At first, selected dataset was extracted from transactional database using ETL tool Kettle with all the necessary computations and stored in MySQL database. We ended up with several fields, which data are characterized by: (i) Customer_ID, (ii) Recency, (iii) Frequency, (iv) Monetary, (v) Age, (vi) Gender, (vii) Loyalty Instrument, (viii) Cross buy, (ix) Amount per Purchase, (x) Price Paid and (xi) Relationship duration. Then delete records which include missing values and inaccurate values. Next, change the data in

appropriate formats. Finally this leaves us with 15,884 instances in Theater and 19,445 in Webshop.

2) Segment customer and determine profitability value

This step is based on the literature we previously discussed, mainly Hughes (1994) and Miglautsch (2000). It can be divided into parts:

(a) Define the scaling of three R-F-M attributes, which are 5, 4, 3, 2 and 1.

(b) Sort the data of three R-F-M attributes by descendant order.

(c) Partition the real data of R-F-M attributes into 5 scaling.

(d) Yield final RFM value by formula (R x 3) + (F x 2) + (M x 1), which leaves us with 25 profitability segments.

3) PLS path modeling was performed using SmartPLS tool. The model fit was evaluated and results for both datasets compared.

4) Data discretization

In this step we performed two discretization methods and compare the results. One is MDLP based optimal binning utilizing the SPSS software and second one is histogram base / equal frequency discretization into 3 clusters for easier explanation ability of final predictions.

5) Decision trees and Apriori rules were generated using Weka tool

6) The results of the decision trees and Apriori rules were interpreted

Conclusion

The statement "retaining customers is more profitable than building new relationships" seems to be true especially in changing Internet market. A worldwide survey conducted by the Conference Board (Bell, 2002) found that customer retention was the most important challenge that CEO's believed they faced. According to Massey et al. (2001), acquiring new customers can cost five times more than it costs to retain existing ones. It has been estimated that it costs five times as much to attract a new customer as it does to retain an existing one, according to research by the American management Association (Massey 2001, Peppers & Rogers, 1996). Niraj, Gupta, and Narasimhan (2001) also argue that estimating profitability at the individual level is important to distinguish the more profitable customers from the less profitable ones.

Specifically, the characteristics of a non-contractual setting, as explained previously, result in both long and short lifetime customers being profitable to the firm. Thus, there exists a need for conceptual and empirical exploration of the antecedent variables that can characterize profitable customers and not just the longer lifetime customers - thereby advancing our relationship management understanding (Reinartz and Kumar, 2003). In our study we tried to contribute, address this need and extended research conducted in this field.

Key Findings

We have empirically validated conceptual model by investigating two non-contractual industries. Key finding are:

- Customer behavioral data are generally the most effective predictive data in customer relationship management. This supports the research by Rud (2001).
- Best customer is long-life customer, who is spending a lot of money at each purchase occasion, or buying wide range of products (cross buying) but buying low end products. Demographics turned out not to be a good predictor.

Benefits

The direct benefit of analysis presented in this study lies in the insight it provides about the uneven distribution of revenues and profits over customers. The results extend the literature by demonstrating that that customers place high importance on product variety and conducting cross-buying. In addition, customers may buy from competitors that provide a higher degree of product variety. Thus, product variety may be a strong competitive tool in a competitive market.

Conclusion

As marketing strives to become more accountable, we need metrics and models that help us assess the return on marketing investment. Many CRM researches pertain to develop a comprehensive model of customer profitability since the question 'Who are profitable customers?' is a starting point of CRM. The easy availability of transaction data and increasing sophistication in modeling has made customer lifetime value an increasingly important concept in both academia and practice. In this study, we suggested a customer lifetime value model considering the recency, frequency and monetary at the same time. It clusters customers into segments according to their lifetime value expressed in terms of RFM and describes their profitability according to it.

Discussion

Based on the results from various statistical and data mining techniques the findings that behavioral variables are better predictors of a profitable customers than demographics were confirmed. This conclusion was also supported by Schmittlein and Peterson (1994), who advocate the use of demographic information but they point out that past purchase behavior generally outpredicts geodemographic information.

Scientific and social implications

Although the importance of an analysis of the dynamic customer–firm relationship is hardly disputed, empirical evidence is rather scarce (Reinartz and Kumar, 2003). Our study helps to widen this field of research as well as gives an answer to call from Zeithaml (2000) "What demographic and psychographic variables are most effective in characterizing profitability tiers?" Because once firms identify profitability segments, they need to be able to characterize these segments into identifiable, measurable, and accessible groups of customers. This is particularly true with companies who deal with end consumers, for these companies must look out at a vast array of potential customers and decide whom to market to and who to focus on once they have a set of customers. Among industries, telecommunications and banking are ahead in these efforts, and more generalizable approaches are needed for other industries.

Data mining techniques are widely used information technology for extracting marketing knowledge and further supporting marketing decisions (Shen, Chuang, 2009). In this study, we also employ data mining techniques to gain insight into our best customers and this knowledge about the customers can support marketing decisions and customer relationship management.

Given the managers are always interested in chasing customers, is it possible to develop some early warning indicators to distinguish longer-life and shorter-life or more profitable and less profitable customers. Our study contributes exactly to this line of research. However, also our research has some limitations and there are ways to extend it.

Limitation and directions for further research

Further research in this topic can focus on:

- In-depth analysis of the industries - examine customers in more industries (contractual / non-contractual)
- Further understanding of the customer lifetime values - developing more complex models for measuring customer profitability (life time value)
- Understanding competitive effects - use of additional variables, to gain deeper insight about the most profitable customers such as satisfaction or attitudinal loyalty. We had to leave this test to future inquiries because of the unavailability of appropriate data

- Applying different Data Mining techniques and see if there are some more precise or interesting results

Bibliography

Aggelis V. (2004). Data Mining for Decision Support in e-banking area. Proc. 1st International Conf. on Knowledge Engineering and Decision Support.

Aggelis V., Christodoulakis D., (2004). Customer clustering using RFM analysis. Proceedings of the 9th WSEAS International Conference on Computers, World Scentific and Engineering Academy and Society, Steven Point, Wisconsin, USA

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases, ACM SIGMOD Conference Washington, DC, USA , 254–259

Anderson J.C., Narus (1991). Partnering as a focused market strategy. California Management Review.

Apte, C., and Hong, S. J. 1996. Predicting Equity Returns from Securities Data with Minimal Rule Generation. In Advances in Knowledge Discovery and Data Mining, eds. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, 514–560. Menlo Park, Calif.: AAAI Press.

Bell, D. (2002). The CEO Challenge: Top Marketplace and Management Issues–2002. Research Report No. R-1322-02-RR, The Conference Board, New York.

Birant D., (2011). Data Mining Using RFM Analysis. Knowledge-Oriented Applications in Data Mining, InTech Publishing.

Bolton, R N. (1998). A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: the Role of Satisfaction., Marketing Science, 17 (1), 45-63.

Bouman R., Dongen J. v., (2009). Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL. Wiley Publishing.

Chang, E.-C., Huang, S.-C., Wu, H.-H (2010). Using K-means method and spectral clustering technique in an outfitter's value analysis. Qual. Quant., 44(4): 807-815

Chang, E.-C., Huang, S.-C., Wu, H.-H., & Lo, C.-F (2007). A Case Study of Applying Spectral Clustering Technique in the Value Analysis of An Outfitter's Customer Database, 2007-2010

Chen Injaz J., Popovich Karen, "Customer relationship management (CRM): People, process and technology", Business Process Management Journal

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

Cheng CH., Chen YS (2009). Classifying the segmentation of customer value via RFM model and RS theory. Expert Syst. Appl., 36:4176-4184

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In: G. A. Marcoulides (Ed.), Modern Methods for Business Research (pp. 295–358). Mahwah, NJ: Lawrence Erlbaum Associates

Chiliya N, Herbst G, Roberts-Lombard M. 2009. The impact of marketing strategies on profitability of small grocery shops in South African townships. Afr. J. Bus. Manage., 3 (3): 70-79.

Clemons, E.K., Weber, B.W. (1994). Segmentation, differentiation and flexible pricing: experiences with information technology and segment-tailored strategies, Journal of Management Information Systems 11 (2) 9 – 36.

Cunningham C., Song Il-Yeol. Chen, Peter P. (2004). Data warehouse design to support customer relationship management analysis. Proceedings of the 7th ACM international workshop on Data warehousing and OLAP, 14-22.

Fayyad U. M. and Irani K. B. (1993). Multi-interval discretization of continuous-valued attributes for classification learning. In Ruzena Bajcsy, editor, Proceedings of the 13th International Joint Conference on Artificial Intelligence., volume 2, pages 1022–1029. Morgan Kaufmann, 1993.

Fayyad U., Piatetsky-Shapiro G., Smyth P., (1997). From Data Mining to Knowledge Discovery in Databases. AI Magazine Volume 17 Number 3, pp. 37- 54

Fayyad, U. M.; Djorgovski, S. G.; and Weir, N. (1996). From Digitized Images to On-Line Catalogs: Data Mining a Sky Survey. AI Magazine 17(2): 51–66.

Feldman, R. and Dagan, I., (2005). Knowledge discovery in textual databases (KDT). In Proc. 1st Int. Conf. Knowledge Discovery and Data Mining, pp. 112-117.

Gartner (2003, March 3). CRM success is in strategy and implementation, not software. http://www.gartner.com

Goodhue, D.L., B.H. Wixom, and H.J. Watson. 2002. Realizing business benefits through CRM: Hitting the right target in the right way. MIS Quarterly Executive 1(2): 79-94.

Gupta S, Lehmann DR (2004). Valuing customers. Journal of Marketing.

Hallowell, R., (1996). The Relationships of Customer Satisfaction, Customer Loyalty, and Profitability: An Empirical Study, International Journal of Service Industry Management, Vol. 7 No. 4, pp. 27-42

Hawkes, V. A. (2000). The heart of the matter: the challenge of customer lifetime value. CRM Forum Resources.

Hosseini SMS, Maleki A, Gholarmian MR (2010). Cluster analysis using data mining approach to develop CRM methodology to asses the customer loyalty. Expert Syst. Appl., 37: 5259-5264.

Hughes, A. M. (1994). Strategic database marketing. Chicago: Probus Publishing Company.

Hughes, A.M. (1996). Boosting Response with RFM. Marketing Tools, 3(3), 4–5.

Irvin, S., Using lifetime value analysis for selecting new customers. Credit World 82 3 (1994), pp. 37–40.

Kalakota, R., & Robinson, M. (1999). e-Business roadmap for success (1st ed.). New York, USA: Addison Wesley Longman Inc.. pp. 109–134.

Kannan, P.K., Rao, H.R., (2001), Introduction to the special issue: decision support issues in customer relationship management, Decision Support Systems 32 (2) (2001) 83–84;

Kerstetter, J. (2001), Software highfliers, Business Week, June 18, 2001, pp. 62– 63.

Kohavi, R., Sahami M., Dougherty J., (1993): "Supervised and Unsupervised Discretization of Continuous Features", Morgan Kaufmann Publishers, San Francisco, CA

Krakhmal V., (2006). Customer profitability analysis in service industries. In: BAA Annual Conference, 11-13 Apr 2006, Portsmouth, UK.

Kumar V., George M., (2007), Measuring and maximizing customer equity: a critical analysis, Journal of the Academy of Marketing Science June 2007, Volume 35, Issue 2, pp 157-171

Kumar, V. and Reinartz, W.J. (2006). Customer Relationship Management: A Databased Approach, John Wiley, New York.

Kumar, V., Venkatesan, R., Reinartz W. (2008). Performance Implications of Adopting a Customer-Focused Sales Campaign. Journal of Marketing, Vol. 72, 50-68

Li, Y.K. and Fu, K.S. (1983): „Automatic Classification of Cervical Cell using a Binary Tree Classifier", Pattern Recognition, vol. 16, pp. 69-80

Liu DR, Shih YY (2005). Integrating AHP and data mining for product recommendation based on customer lifetime value. Inform. Manage., 42: 387-400.

Liu, B., Hsu, W., Han, H. S., & Xia, Y. (2000). Mining changes for real-life applications. Second International Conference on Data Warehousing and Knowledge Discovery , 337–346.

Liu, B., Hsu, W., Mun, L. F., & Lee, H. Y. (1999). Finding interesting patterns using user expectations. IEEE Transactions on Knowledge and Data Engineering, 11(6), 817–832.

Maiom O., Rokach L., (2005). Data mining and knowledge discovery handbook. ISBN: 978-0-387-09822-7

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

Massey Anne P., Montoya-Weiss Mitzi M., Holcom Kent (2001).Re-engineering the customer relationship: leveraging knowledge assets at IBM. Decision Support Systems, 3 (2) (2001), pp. 155 – 170.

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

Miglautsch, J. R. (2000). Thoughts on RFM Scoring. International Society for Strategic Marketing, The Journal of Database Marketing, 8(1), 67–72.

Mittal, Vikas and Wagner Kamakura (2001), "Satisfaction, Repurchase Intent, and Repurchase Behavior: Investigating the Moderating Effect of Customer Characteristics," Journal of Marketing Research, 38 (February), 131–42.

Mutandwa E, Kanuma NT, Rusatira E, Kwiringirimana T, Mugenzi P, Govere I, Foti R (2009). Analysis of coffee export marketing in Rwanda: Application of the Boston consulting group matrix. Afr. J. Bus. Manage., 2(4): 210-219.

Newell, F. (1997). The new rules of marketing: How to use one-to-one relationship marketing to be the leader in your industry. New York: McGraw-Hill

Niraj, R., Gupta M., Narasimhan Ch. (2001). Customer profitability in a Supply Chain. Journal of Marketing, 65 (July), 1-16.

Peppers, D., & Rogers, M. (1996). The one to one future: Building relationships one customer at a time. NY: Doubleday

Perner, P.; Belikova, T. and Yashunskaja, I. (1996): „Knowledge Acquisition by Symbolic Decision Tree Induction for Interpretation of Digital Images in Radiology", In Proc. Advances in Structural and Syntactical Pattern Recognition, Springer Verlag, pp. 208-219

Petrison L.A., Blattberg R.C., Wang P., (1993). Database marketing past, present and future, Journal of Direct Marketing 7 (3), 27– 43

Pine B.J, Gilmore, J.H. (1999). The Experience Economy: Work is Theater and Every Business is a Stage.

Quinlan, J, (1993).: C4.5: programs for machine learning, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993, 55860-238-0

Reichheld, Fredrick F. , (1996b). The Loyalty Effect. Cambridge, MA: Harvard Business School Press.

Reichheld, Fredrick F. (1996a). Learning From Customer Defections. Harvard Business Review 73 (2): 56-69.

Reinartz W., Kumar, V. (2000). On the profitability of Long-Life customers in a noncontractual settings: An empirical investigation and implications for marketing, Journal of Marketing, 64 (October), 17-35

Reinartz W., Kumar, V. (2003). The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration. Journal of Marketing, 67(1), 77–99

Reinartz, W., (2002), The Mismanagement of Customer Loyalty, Harvard Business Review, 80 (July), 86–94.

Reinartz, W., Krafft, M. & Hoyer, W.D. (2004) "The Customer Relationship Management Process: Its measurement and impact on performance", Journal of Marketing Research (JMR), 41(3), pp.293-305

Reinartz, W., Kumar V. (2000), "On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing," Journal of Marketing, 64 (October), 17–35.

Rodgers, S., Harris, M. A. (2003). Gender and E-Commerce: An Exploratory Study. Journal of Advertising Research, 43, pp 322-329

Rud, O. P. (2001). Data Mining Cookbook. New York: Wiley

Schijns Jos M. C. andGaby J. Schröder (1999). Segment selection by relationship strength. Journal of Direct Marketing, Volume 10, Issue 3, pages 69–79, Summer 1996

Schmittlein D. and Peterson Robert A. (1994), Customer Base Analysis: An Industrial Purchase Process Application, Marketing Science, 13 (Winter), 41-67.

Shaw, M.J., Subramaniam C., Tan G.W., Welge M.E., (2001). Knowledge management and data mining for marketing, Decision Support Systems 31 (1), 127– 138

Sohrabi B, Khanlari A (2007). Customer lifetime value (CLV) measurement based on RFM model. Iranian Acc. Aud. Rev., 14(47): 7-20.

Song IY, Levan-schultz, K, (1999). Data Warehouse Desgin for E-Commerce Environment. Proceedings of ER Workshops.

Song, H. S., Kim, J. K., & Kim, S. H. (2001). Mining the change of customer behavior in an internet shopping mall. Expert System with Applications, 21(3), 157–168.

Tsai C.-Y., Chiu C.-C. (2004). A purchase-based market segmentation methodology. Expert Systems with Applications 27 (2004) 265–276

Verhoef, Peter, C., Franses P. H., Hoekstra, Janny C. (2002). The Effect of Relational Constructs on Customer Referrals and Number of Services Purchased From a Multiservice Provider: Does Age of Relationship Matter?, Journal of the Academy of Marketing Science 30 (3): 202-212

Wei Jo-Ting, Lin Shih-Yen, Wu Hsin-Hung, "A review of the application of RFM model", African Journal of Business Management, Vol 4(19), 2010

Yeh IC, Yang KJ, Ting TM (2009). Knowledge discovery on RFM model using Bernoulli sequence. Expert Syst. Appl., 36: 5866-5871.

Zeithaml V. A. (2000). Service quality, profitability, and the economic worth of customers: what we know and what we need to learn. Journal of the Academy of Marketing Science.

Zeithaml V. A., Roland T. Rust & Katharine N. Lemon (2001). The Customer Pyramid: Creating and Serving Profitable Customers. California Management Review 43, no. 4, Summer 2001

List of figures

List of Tables