LUT
Lappeenranta
University of Technology

Lappeenranta University of Technology
School of Business and Management
Strategic Finance and Business Analytics

Mikko Forsström

Analysis of loyalty program data and profitability – Case of a company

Master's thesis
2020

1st Supervisor: Professor, D.Sc. (Econ. & BA), Mikael Collan

2nd Supervisor: Post-Doc. Researcher, D.Sc. (Tech.), Jan Stoklasa

ABSTRACT

Author: Forsström, Mikko
Title: Analysis of loyalty program data and profitability – Case of a company
Faculty: LUT, School of Business and Management
Major: Finance
Year: 2020
Master's Thesis: Lappeenranta University of Technology
46 pages, 3 charts, 8 tables, 12 figures and 6 equations
Examiners: Professor, D.Sc. (Econ. & BA), Mikael Collan
Post-Doc. Researcher, D.Sc. (Tech.), Jan Stoklasa
Keywords: customer value, customer profitability, customer loyalty, loyalty program, customer value metrics, customer lifetime value

Customer loyalty program (LP) can be an important part of the retail strategy and the data collected from the loyalty program can give insight and knowledge for the management when it is analyzed properly. A part of the benefits of the LP are monetary, which influences the profitability of the loyalty program. Also, the members of the LP are differently profitable as some do purchases with the retail price whereas others might only purchase discounted products. This brings up two aspects to be analyzed on the LP data. The first aspect is the current reward structure of the LP and if the structure rewards the right customers and motivates the members to act more profitably. The profitability of the LP members now and in the future, is the second aspect. This is important to be analyzed to ensure that the future investments within the LP can be allocated profitably, and to the right customers. In this thesis, the profitability of the LP and its structure was analyzed with descriptive methods: the reward structure is looked at by checking if it generates points-pressure to earn bigger discount vouchers. And how different level vouchers generate spending that surpass the vouchers value. These measures and results could possibly be used for optimizing the reward structure or as indicators on how the LP is performing. The future profitability of the customers is approached by implementing forward-looking models for predicting the future spending of the customers to get a better view on the customer lifetime value. The CLV prediction models, linear regression and Pareto/NBD, were applied to the newly signed and active Finnish members of the LP as for this cohort the needed data was fully available. The accuracy of the model's predictions is compared to see which model performs the best. The CLV prediction could be an important metric for marketing function's and LPs resource allocation as it would make sure that the invested amount to serve and reward a certain customer would not be bigger than the future value the customer brings to the company. The data used is from the loyalty program's database and represents the second half of 2016 and the first half of 2017. The conclusion of the research is that the LPs database is a good source of information, but to get more reliable results on the analyses of the reward structure and on the performance of the CLV prediction models, the LPs cost side and the individual margins of the customers should be collected and included in the analyses.

TIIVISTELMÄ

Tekijä: Forsström, Mikko
Tutkielman nimi: Kanta-asiakasjärjestelmän datan ja kannattavuuden analyysi
Tiedekunta: Kauppatieteellinen tiedekunta
Pääaine: Rahoitus
Vuosi: 2019
Pro gradu tutkielma: Lappeenrannan teknillinen yliopisto
46 sivua, 3 kaaviota, 8 taulukkoa, 12 kuvaa ja 6 kaavaa
Tarkastajat: Professori, KTT, Mikael Collan, Tutkijatohtori,TkT, Jan Stoklasa
Avainsanat: asiakkaanarvo, asiakaskannattavuus, asiakasuskollisuus, kanta-asiakasohjelma, asiakkaanarvomittarit, asiakkaan elinkaariarvo

Kanta-asiakasohjelma voi olla tärkeä osa vähittäiskaupan strategiaa, ja kanta-asiakasohjelmasta kerätyt tiedot voivat antaa, oikein analysoituna, ymmärrystä ja tietoa johdolle. Osa kanta-asiakasohjelman eduista ovat rahallisia, mikä vaikuttaa ohjelman kannattavuuteen. Lisäksi ohjelmaan kuuluvat jäsenet eroavat kannattavuuksiltaan, sillä jotkut ostavat tuotteita vähittäismyyntihinnalla, kun taas toiset ostavat vain alennustuotteita. Tämä tuo esille kaksi näkökulmaa, joita analysoidaan ohjelman tuottaman datan pohjalta. Ensimmäinen näkökulma on kanta-asiakasohjelman nykyinen palkkiorakenne ja palkitseeko ohjelman nykyinen rakenne oikeat asiakkaat ja motivoiko rakenne toimimaan kannattavammin yhtiön näkökulmasta. Ohjelman jäsenten kannattavuus nyt ja jatkossa on toinen näkökulma. Tätä on tärkeä tarkastella sen varmistamiseksi, että tulevat sijoitukset asiakkaisiin ohjelman puitteissa voidaan kohdistaa kannattavasti ja oikeille asiakkaille. Tässä pro gradussa kanta-asiakasohjelmaa ja sen rakenteen kannattavuutta analysoitiin kuvailevilla menetelmillä: palkkiorakennetta tarkastellaan tutkimalla kannustaako palkkioiden ansaintarakenne asiakkaita kuluttamaan lisää päästäkseen seuraavalle etu-tasolla ja kuinka eri tason etusetelit lisäävät kulutusta. Tuloksia voidaan hyödyntää palkkiorakenteen optimoinnissa tai mittarina siitä, kuinka hyvin ohjelma toimii. Asiakkaiden kannattavuutta jatkossa lähestytään toteuttamalla ennustemalleja asiakkaiden tulevasta kulutuksesta, jotta asiakkaan elinkaariarvosta saadaan parempi käsitys. Asiakkaan elinkaariarvo- ennustemalleja (lineaarinen regressio- ja Pareto/NBD- malli) sovellettiin kanta-asiakasohjelmaan juuri liittyneiden ja aktiivisten asiakkaiden elinkaariarvon ennustamiseen, koska tälle segmentille data oli kokonaan saatavilla. Mallien ennusteiden tarkkuutta verrattaan keskenään, jotta selviää mikä malleista toimi parhaiten. Mallin tuottama ennuste asiakkaan elinkaariarvosta voi olla tärkeä mittari markkinoinnin ja etenkin kanta-asiakasohjelman resurssien allokoinnissa, koska se varmistaisi, että sijoitus tiettyyn asiakkaaseen ei ole suurempi kuin tuleva elinkaariarvo, jonka asiakas tuo yritykselle. Käytetyt tiedot ovat kanta-asiakasohjelman tietokannasta ja edustavat vuoden 2016 jälkipuoliskoa ja vuoden 2017 ensimmäistä puoliskoa. Johtopäätöksenä voidaan todeta, että kanta-asiakasohjelman tietokanta on hyvä tietolähde, mutta luotettavampia tuloksia varten palkkiorakenteen analyyseihin ja elinkaariarvojen-ennustemalleihin on tarpeellista lisätä dataa kanta-asiakasohjelman kulupuolesta ja yksilötason asiakaskatteista.

Table of contents

1. Introduction

Customer loyalty programs are popular in a variety of industries and they are among the most popular marketing tools for companies to collect data. With loyalty programs companies can collect data of customer behavior and use this information to guide decision making and resource allocation in business and marketing management. The customer loyalty programs often offer monetary benefits to members, which influences the profitability of the program.

The customer loyalty program collects vast amounts of data of its members. The reward system is based on rewarding good customers bi-annually with a voucher with a value that is based on the previous purchases. This research sets out to analyze the loyalty program data to analyze the profitability and usability of the loyalty program and current reward system for good customers. The implemented research is planned to give managerial contributions and practical implications to the principal organizations marketing department.

1.1. Motivation of the study

A customer loyalty program can be a fundamental part of the retail strategy and the data collected via the loyalty program can offer a lot of knowledge and insights for the management, when analyzed properly. It should be ensured that a company gets all the possible benefits from the loyalty program data. This is done by analyzing the data and finding out useful analyses and models that give insight for guiding the business.

In the case company, the marketing management of a brand is interested in the profitability of the loyalty program and new ways of utilizing the data gathered through the program. This study sets out to accomplish this by analyzing the data provided and trying to find out proper analyses and models that give insight on profitability of the current reward system and structure of the loyalty program. So, the motivation of this study is to ensure and better the profitability of the loyalty program and gain information through the LPs data that can help form the retail marketing strategy of the brand further. Aim is to introduce and implement analyses that can be used to gain information on the data and the programs profitability.

1.2. Theoretical background and motivation

In their book Customer Relationship Management: Concept, Strategy, and Tools (2012 p. 184) Kumar & Reinartz define customer loyalty programs as marketing processes that reward customers based on their purchases (size and repetition). And as a CRM tool that collects data that firms can use to identify, award and retain profitable customers. They write that loyalty programs (LP) aim to keep customers by offering rewards to them for repeat purchases. The success of a LP depends on the profitability gained from the customers. They also emphasize that LP is a method to gather data to improve the efficiency and effectiveness of the marketing function. (Kumar, V. & Reinartz, W.2012p.201-204)

LPs objectives are building behavioral and attitudinal loyalty, efficiency profits, effectiveness profits and value alignment. Behavioral loyalty is the observed actions by customers; the most used measure of this is share of wallet. Attitudinal loyalty entails the attitudes and perceptions of the customer. Efficiency profits are the profits that result from change in customers buying behavior due to LP net of LPs costs. Effectiveness profits are the long-term consequences realized by learning about customer preferences and can generate competitive advantages and higher profits in the long run. Value alignment aims to match the cost of serving a particular customer with the value they provide to the firm. (Kumar, V. & Reinartz, W. 2012 p. 201-204)

Kumar and Reinartz also describe the importance and benefits of data mining and give an overview of the process. They state that nowadays the availability of vast amounts of data, computing power, mass storage and data analysis methods give companies access to a powerful asset: information. To gain value from the data, companies must implement a successful data mining procedure that transforms data into new knowledge that can be transformed into business action and value. They present a five-step data-mining process that consists of defining the business objectives, getting raw data, identifying relevant variables, gaining customer insight, and acting. (Kumar, V. & Reinartz, W. 2012 p. 143-144)

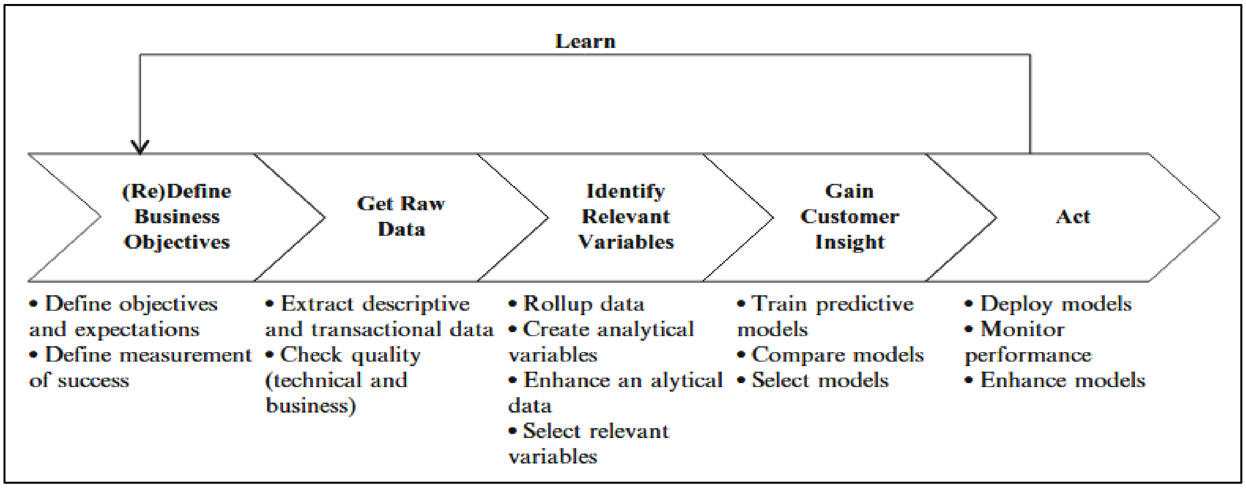

Figure 1. Overview of the data-mining process (Kumar, V. & Reinartz, W. 2012 p. 145)

Before the steps are opened, data manipulation is introduced to data-mining as dimensionality of data might change dramatically through the process. The manipulations on columns are listed as: transformation (for example birth date to age), derivation (creating new variables based on existing ones), elimination (excluding a whole variable). The row manipulations are listed as: Aggregation (for example counts, means and standard deviations of specific type), Change detection (to detect if certain variables change their value), missing value detection and outlier detection. The data is common to be split into sets for different purposes when it is prepared for modeling. These sets are: train set (used to build the models), test set (for out-of-sample test of the model and selecting final model candidate) and scoring data (for model-based prediction) (Kumar, V.& Reinartz, W.2012p.146-147).

In the first step of data-mining project defining business objectives consists of specifying goals, target variables, methods and rules, identifying cost and revenue drivers, creating a project plan and establishing criteria for evaluating success. The second step is getting the raw data which consists of looking for, loading and checking the quality of the data. The third step of identifying relevant variables can consist creating analytical customer view/ variables and selecting predictable variables. The fourth step of gaining insight consists of preparing data samples, (predictive) modeling and selecting a model. The fifth step of Acting consists of delivering and archiving results and learning from the results. (Kumar, V. & Reinartz, W. 2012 p. 148-158)

1.3. Scope and focus of the study

In this thesis literature from four different branches of science are combined: marketing, business analytics, finance and accounting. Focus of this study is on the customer loyalty program as a system and on profitability of its structure and effects, and usability of the data it collects.

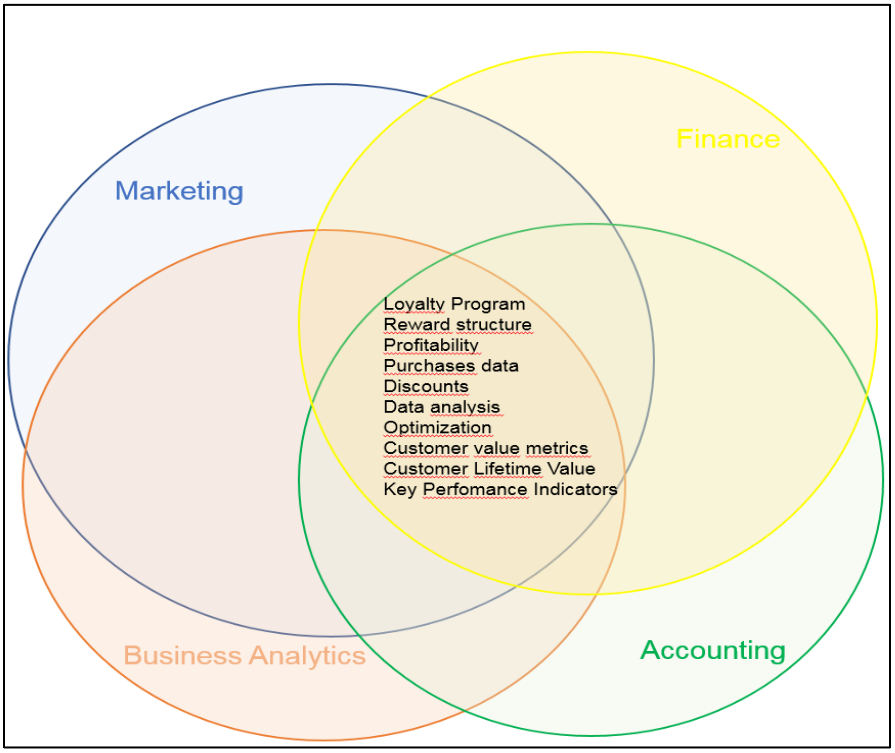

Figure 2. Combination of different branches of science literature in the study

The scope of this study doesn't include business to business markets as the data used in this study is on business to consumers markets. The scope is also limited to the short-term effects of a loyalty program as the data used is one year's purchases. This means that the long-term effects of a continuous loyalty program are excluded. The value of the used vouchers is the main cost to be researched as the data gives some limitations on the cost side of profitability calculations.

1.4. Outline and research questions

In her thesis, Kaisa Kuokka stated that there are two pressing issues about loyalty programs. The first issue is if they really work effectively to enhance loyalty or are the already loyal customers selecting to become members to benefit from the program. The second issue is the profitability of the loyalty program as companies might use resources for marketing activities targeted to members of the loyalty program such as special offers, price discounts and vouchers. The members might act differently on these activities: some may buy only products that are on sale and others buy products with the normal retail price. This makes the customer margins of differently acting members different. (Kuokka, K. 2014. p. 7). Therefore, the profitability of the members and the program is something to be analyzed more closely.

The research questions of the thesis are:

1. What general insights can be gained from data-based analysis of a loyalty program by way of (business) analytics methods?
2. How can the loyalty program of the case company be enhanced?
   i. Is the current reward system optimal/profitable?
   ii. Should there be more and bigger classes in the reward system?
   iii. Is it possible to develop new and usable measures from the collected data (KPI)?

So, this thesis tries to find answers to these questions by reviewing current literature and by conducting a case study on a data set provided by the principal organization.

1.5. The structure of this thesis

The thesis is structured as follows. Next the literature and researches are reviewed on the following topics: customer loyalty programs and their profitability, reward systems, customer metrics on observable measures and loyalty program data. Then in section three the literature and scientific side of research methods are described. In fourth chapter, the case is presented: the data is introduced; conducted research is described and the results presented. In the fifth section, the research questions are answered in the light of the analyses and results, implications for real world and academic results are presented, limitations are discussed and topics for future research are suggested.

2. Literature review (state of the art)

The literature review displays the search for the current knowledge, in the international environment, usable for this thesis: previous findings, theoretical and methodological contributions of previous research on the topics of loyalty programs, data, profitability, reward structure and customer value metrics. The quest to find relevant research articles started from defining terms that can be used as search words in international library databases. The following search-string was developed: customer loyalty program (data or profitability or reward structure or metrics). Elsevier's Science Direct database was used as a source. The search was limited to publications of the period 2000 to present day. This query provided a stock of 437 Journals. The query was refined by choosing relevant fields of science: Business, Management and Accounting, Computer Science and Economics, Econometrics and Finance. Also, search was limited to subscribed publications and open access articles.

Next the results were filtered by excluding irrelevant publications (Tourism management and Human resource management). This resulted in 252 journals. These 252 journals were sorted by relevance, scanned and irrelevant articles were excluded. First scan was done by reading the titles and excluding clearly irrelevant journals. This resulted in 60 journals remaining. Next the abstracts of remaining journals were read and relevant articles chosen and downloaded. This resulted in a stock of 30 journals for more thorough examination.

The last step of the literary selection process was backward tracking. It is an effort to spot major contributions not captured by the search-string. The 30 articles were reviewed to find major contributions not included in the initial stock. This was done by going through the articles introductions and literary reviews. Also a few articles, that were found in the planning phase of this thesis and deemed relevant by the author, were added in this phase.

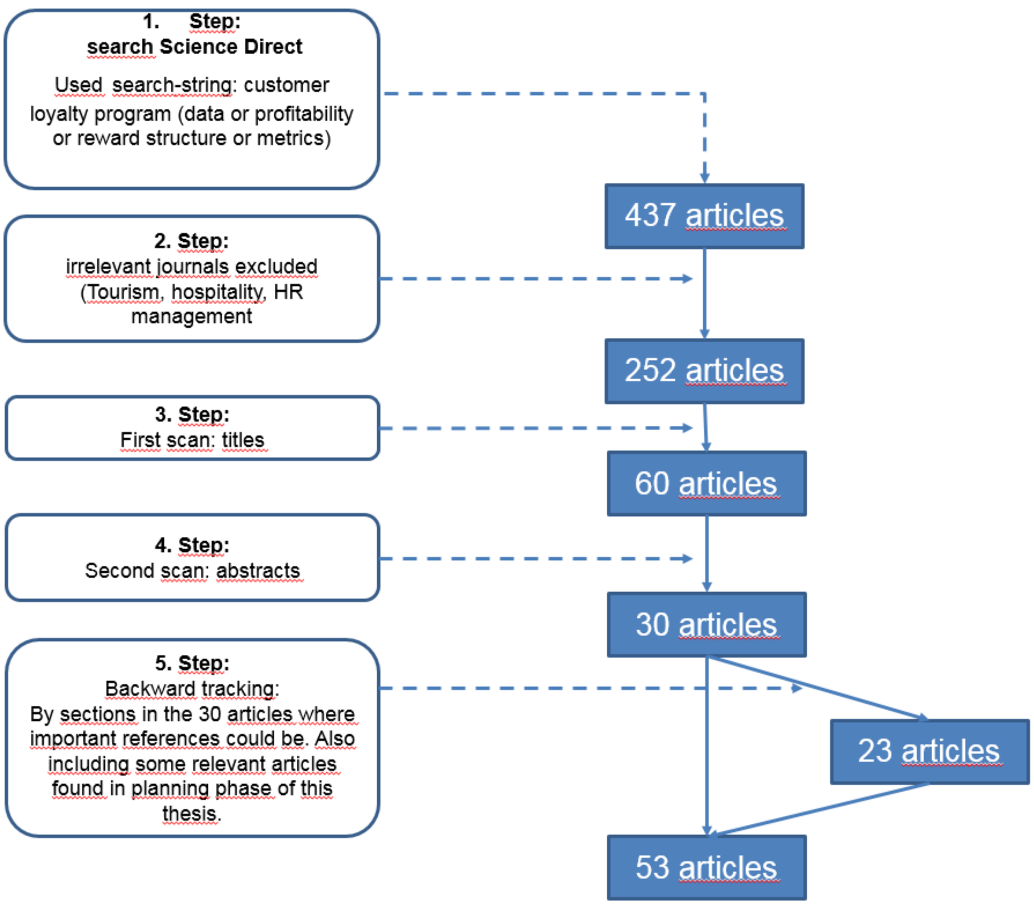

Figure 3. The literature selection process

First the literature on customer loyalty program data is reviewed, in chapter 2.1., to gain knowledge on current literature and results on data diversity and quality and how the data can be used to gain insight. A main point that is gained from this chapter is that, even though the raise in data quality drives improved outcomes, the profit-maximizing level of data quality will be realized with not-perfect data as marginal benefits of improved quality will decrease with increased cost. Next, in 2.2., the literature on profitability of customer loyalty programs is reviewed. Here the metric of Customer Lifetime Value is already mentioned as a forward-looking metric to evaluate the profitability. Also, the link between LP's profitability and reward structure is brought up in the papers.

In chapter 2.3. the articles on reward structure are reviewed. The rewards structures effects on loyalty (to LP vs company) and profitability are viewed. Also, complexity's amount in reward structure and the point's pressure -effect are presented. A 2- tier LP framework is presented, as a more optimal option of a reward structure, and analyzed.

Chapter 2.4. goes through literature on customer value metrics with comparisons between backward- and forward-looking metrics and an emphasis on CLV as it is a working forward-looking metric.

2.1. Data analytics in customer loyalty program context

Verhoer et al. (2010) provide an overview of retailing's CRM in data-rich environments. They discuss how retailers can gather customer data using, for example loyalty programs, and how these data can be analyzed to gain useful customer insights. They define CRM as the practice of analyzing and utilizing marketing databases to determine corporate practices and methods that will maximize the lifetime value of each customer. They build a conceptual model where diverse data (from customer, Point-of-sale (POS) and supply chain) is collected and integrated, integrated data is utilized by customer analytics and managerial insights are derived to make better decisions and improve performance. They give an overview of studies in their conceptual model and the most important findings of these topics. The paper's focus is on individual customer data. They state that data quality and data integration can be extremely difficult for firms that collect data across multiple channels. The data quality is positively related to firm performance. But that, marginal benefits of improved quality will decrease with increased cost, which implies that the profit-maximizing level of data quality will be realized with not-perfect data. (Verhoef et al. p. 121-125)

In their article “The Role of Big Data and Predictive Analytics in Retailing” (2017) Bradlow et al. examine the opportunities in and possibilities coming from big data in retailing, along major data dimensions - customers, products, time, location and channel. They ran a field experiment with a large national chain in U.S. and 42 of its stores and used a predictive econometric model to optimize profit in the framework of price and promotion elasticity and accounting for other factors such as seasonality. They conclude that model-based elasticity price optimization improves gross margin dollars significantly, both managerially and statistically. They state that it is data quality rather than only raise in data volumes that drive improved outcomes and that much of the increase in quality comes from new data sources, smart application of statistical tools and domain knowledge with theoretical insights. (Bradlow et al 2017. p.90-93)

Chang et al. investigate the influence of the completeness of CRM relational information processes on customer based relational performance and profit performance in their study “The effects of customer relationship management relational information processes on customer-based performance” published in Decision Support Systems journal in 2014. They developed a framework that resulted in 5 hypothesis, three of which were supported based on the analyses conducted: CRM relational information processes positively affect customer-based relational performance (Hypothesis 1), Customer-based relational performance positively affects customer-based profit performance (Hypothesis 3) and the positive effect of CRM relational information processes on customer-based relational performance is higher for firms with higher interaction orientation (Hypothesis 4a). The paper defines CRM relational information processes to include reciprocity, capture, integration, access, and use of information to facilitate CRM performance. Customer-based relational performance is indicated by three measures: customer satisfaction, customer ownership, and positive word-of-mouth (WOM). Customer-based profit performance is measured with three indicators: identification of profitable customers, acquisition and retention of profitable customers, and conversion of unprofitable customers to profitable ones. (Chang et al.2014. p. 146-149)

The study used a qualitative case study on three firms to ensure rationality and reliability of the framework with individual in-depth interviews. Then a questionnaire survey with 7-point Likert-scale was done to obtain quantitative data. The hypotheses were tested with a confirmatory factor analysis(CFA) and structural equation modeling(SEM). The results from the analyses showed that completeness of CRM information processes affect positively customer-based relational performance, which in turn enhances profit performance. (Chang et al.2014. p. 152, 157)

Peltier et al. study the relationship between organizational processes, customer data quality and firm performance by developing and testing an organizational framework and conducting a questionnaire for executives in financial services industry, in their article published in the Journal of Interactive Marketing in 2013. (Peltier et al. 2013 p. 1)

In their theoretical framework, they list that the quality of customer data is dependent of the following components quality: customer touchpoints (i.e. Internet contacts, email), transaction data (i.e. Purchase history), loyalty/satisfaction data (LPs, satisfaction surveys) and customer lifetime value data (i.e. retention, share of wallet). They define that "*Customer data are of high quality when the information collected across multiple transactions, touchpoints, and channels accurately reflects the behaviour and sentiments of customers, both collectively and individually."* (Peltier et al. 2013 p. 2)

The study was conducted by testing the relationships of the framework using a structural equation model to analyse the collected data. The testing supported many of the relationships developed in the hypothesized framework and based on this a final model for a framework on organizational learning and data quality was presented:

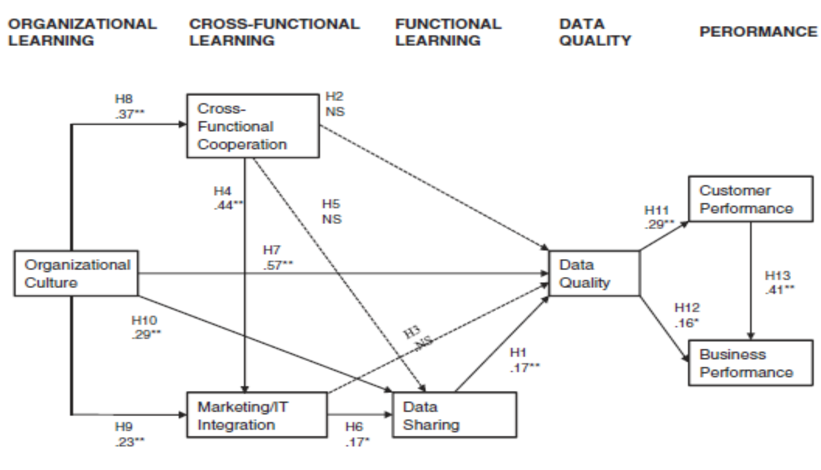

Figure 4. Organizational learning and data quality framework (Peltier et al. 2013 p. 8)

2.2. Profitability of customer loyalty program

In their paper A stochastic model on the profitability of loyalty programs (Computers & Industrial Engineering, 2011, Vol.61(3), 482-488) Gandomi and Zolfaghari develop an analytical model on the profitability of loyalty programs. Their model consists of a revenue-maximizing firm selling a product through two periods. The customers earn a loyalty reward in the form of a discount on the product in the second period if they buy in both periods. The objective of the model is to maximize the revenue in terms of three decision variables: prices of the product in first and second period and the loyalty reward amount. Their paper gives insights into profitability of LPs: it shows that the optimal price in period 1 generally increases with loyalty reward, the increased price offsets the costs of the LP and that the price in period 1 increases as customers intention to repurchase rises.

V. Kumar et al. examine the questions: "What is the right metric to manage customer programs, for example, customer loyalty programs? Can CLV outperform traditional metrics?" in their article Managing retailer profitability—one customer at a time! (2006). They write that retailers usually use historic measures and backward-looking metrics (past spending or frequency of purchase) as objective measures to manage customer loyalty. They conducted a study where they modeled the correlation of the backward-looking measures of years 2001-2003 to customer transactions of 2004 to see how well historic measures help on determining future customer profitability. Their study found out that often these prove to have a poor correlation with the future profitability of the customer. Hence, they suggest that retailers should have a forward-looking metric such as CLV (customer lifetime value) implemented before investing valuable resources into loyalty management programs as the CLV metric could ensure simultaneous management of loyalty and profitability of the customer. (Kumar et al. 2006 p. 281 & 291)

So et al. examine the types of values customers derive from LP membership and the relationship between program value, loyalty to the program, and loyalty to the firm in their article published in 2015 in Australasian Marketing Journal. They further examine program and brand loyalty's effect on behavioral responses such as share of wallet, share of purchase, word of mouth and willingness to pay more. (So et al. 2015 p.196) The research was done with a survey to members of two Australian stand-alone loyalty programs and structural equation modeling with maximum likelihood estimation was used to analyze the answers and test the hypothesis. The article's findings suggest that reward attractiveness, knowledge benefit, experiential benefit, disclosure comfort, group belongingness and required effort are aspects of value that customer derives from joining a LP. The study resulted further in that reward attractiveness, knowledge benefit and required effort influence the perception of experiential benefits, which affects program loyalty. And that brand loyalty is driven by group belongingness, disclosure comfort and program loyalty. Program loyalty is found to enhance LP members share of wallet, share of purchase and word of mouth (WOM), but has a reverse impact on willingness to pay more. Brand loyal customers are found to have larger share of purchases, be enthusiastic advocates for the firm(WOM) and dedicated in obtaining the firm's offerings (willing to pay more). The article gives a few key managerial implications in that by fostering program loyalty a firm can cultivate brand loyalty especially through customization of the LP. (So et al. 2015 p.201-204)

Kang et al. investigate how customers in a loyalty program perceive the benefits from the LP that enhance their loyalty to the program and to the company indirectly through program loyalty in their article Customer–company identification and the effectiveness of loyalty programs published in the Journal of Business Research in 2015. The article conceptualizes a model with eight hypotheses: the (1) financial and (2) social benefits a LP offers relate positively to program loyalty, (3) program loyalty relates positively to customer-company identification(CCID), (4) program loyalty relates positively to company loyalty, (5) CCID relates positively to company loyalty, (6) program loyalty relates positively to share of wallet, (7) program loyalty can increase and (8) company loyalty decrease firm's latent financial risk. The data was collected through an online panel of LP members and five-point Likert type scales was used as measures. The hypotheses were tested with a partial least squares structural equation modeling(PLS-SEM). (Kang et al. 2015 p. 464-467)

The research found out that loyalty programs can contribute to company loyalty and program loyalty and that specifically customers perceptions of the values and benefits available through participating in a LP grow loyalty toward the program, which can boost loyalty to the company. About the latent financial risks of a firm the study found out that program loyalty can increase it and company loyalty reduce it. So, transforming program loyal customers to company loyal can reduce the negative effects of program loyalty. The study highlights the importance of accounting customer-company identification aspects in LPs as forming CCID can make the customers more loyal to the company. For this the balancing of social and financial benefits is mentioned as an important measure of LPs' effectiveness for building company loyalty. (Kang et al. 2015 p. 468-

469)

Bridson et al. asses the relationship between loyalty programs hard and soft attributes, store satisfaction and store loyalty in their article published in 2008 in the journal of retailing and consumer services. They define LPs hard attributes to be tangible elements such as discounts or gifts that have economic value whereas soft attributes such as special communications and treatment are more emotionally oriented and don't carry intrinsic dollar worth. The article defines a conceptual framework where LP has a positive effect on store loyalty, but where store satisfaction acts as a mediator between these two. The data was collected by a survey on LP members of a health and beauty retailer and regression analysis and SEM was used to analyze the data. (Bridson et al. 2008 p. 365-369) The study found that LP is a significant predictor of store loyalty and that both hard and soft attributes are significant predictors of loyalty, hard on WOM and behavioral and soft on all including commitment. The attributes and LP were also found to be significant predictors of store satisfaction and that store satisfaction is positively related to store loyalty. The knowledge from this study can be applied to the evaluation of existing LPs by assessing the worth of the blend of attributes currently in use. (Bridson et al. 2008 p. 371-372)

2.3. Reward structure design in loyalty programs

Kreis and Mafael study "The influence of customer loyalty program design on the relationship between customer motives and value perception" in their article published in 2014 in the Journal of Retailing and Consumer Services. They present a framework that shows loyalty programs with differing designs as a moderating tool in a means-end relationship between customer motives and value. Their framework divides LPs into two classes based on the LPs design: monetary or treatment (status or contact) based. The monetary based LPs provide economic value to the customers and are distinguished based on the way they deliver the value: reward redemption (bonus points used as currency) or money saving (offering discounts). The research supports the argument that a LP can be an effective tool that creates value by itself, and not only something that adds to value of a product or a service. The paper notes that this can be the case only for programs that target prevailing customer motives and this way provide a higher level of perceived value. (Kreis & Mafael. 2014, p. 590-591)

In their article “The effect of loyalty program fees on program perceptions and engagement” published in 2016 in the Journal of Business Research, Ashley et al. study how loyalty program fees change consumer perceptions and how the loyalty program structure interacts with loyalty program fees to affect customer engagement. They conducted three different studies to answer these research questions. (Ashley et al. 2016, p. 964) The data was collected by online survey, student survey and consumer panel and the data collected was analyzed with regression, GLM and anova-models. The studies found out that fee-based programs can help separate more desirable customers through screening or because fee-paying generates favorable behavior. And that the structure of the LP has an interaction with the fees to customer engagement: when LP fees are nominal or free simple reward structure (for example dollar to 1 point accrual) seem to be the best, but when the cost of a membership to a LP rises a more complex denominations of accrual system is likely to increase membership intentions as the focus of processing is taken from the fee of the program to rewards (cost/benefit analysis). (Ashley et al. 2016, p. 971)

Blattberg et al discuss the issue of how earned rewards relate to customer lifetime value in their article published in 2009 in the journal of interactive marketing. They start by listing the important differences between earned rewards and standard price discounts. The main thing is that rather than giving a certain discount at the time of the purchase, the reward is generally earned conditional on the customers desirable purchase behavior in a certain period. This means that reward programs may serve to increase relationship duration and increase revenues as they create “points pressure” for the customer to accumulate enough points to qualify for a reward. They bring up the question of whether the receipt of a reward will positively affect future behavior of a customer. They refer to another article from Blattberg et al. that summarizes the concepts of points pressure and rewarded behavior and show empirical evidence regarding them, but end in that more evidence is needed to quantify and verify their impact on CLV. (Blattberg et al. 2009, p. 161)

Kumar and Shah propose a conceptual framework for building and sustaining profitable customer loyalty, by linking loyalty to profitability and a forward-looking metric in the form of CLV by reviewing related research in their article published in the journal of retailing in 2004. They present a two-tiered rewards structure for marketers to operationalize the framework. (Kumar & Shah. 2004 p.317)

In their review of concept of customer loyalty three problems are brought up. First is that by rewarding only behavioral loyalty (the more you spend the more rewards you earn) customers may associate their loyalty towards the rewards program and not the company. Second is that LPs that reward behavior without considering profitability run the risk of imminent failure. Third is that LPs that reward customers by actions committed today (instant reward) or in the past (delayed reward) fail to consider future potential of the customer and that research indicates that previously well performing customers need not to perform similarly in the future in terms of spending or profitability to the firm. After this they bring up the question of is it possible to develop a LP that can pro-actively reward customers 'today' for their 'future' spending? (Kumar & Shah. 2004 p.319)

The following conceptual framework is presented where the two-tiered reward structure is also included:

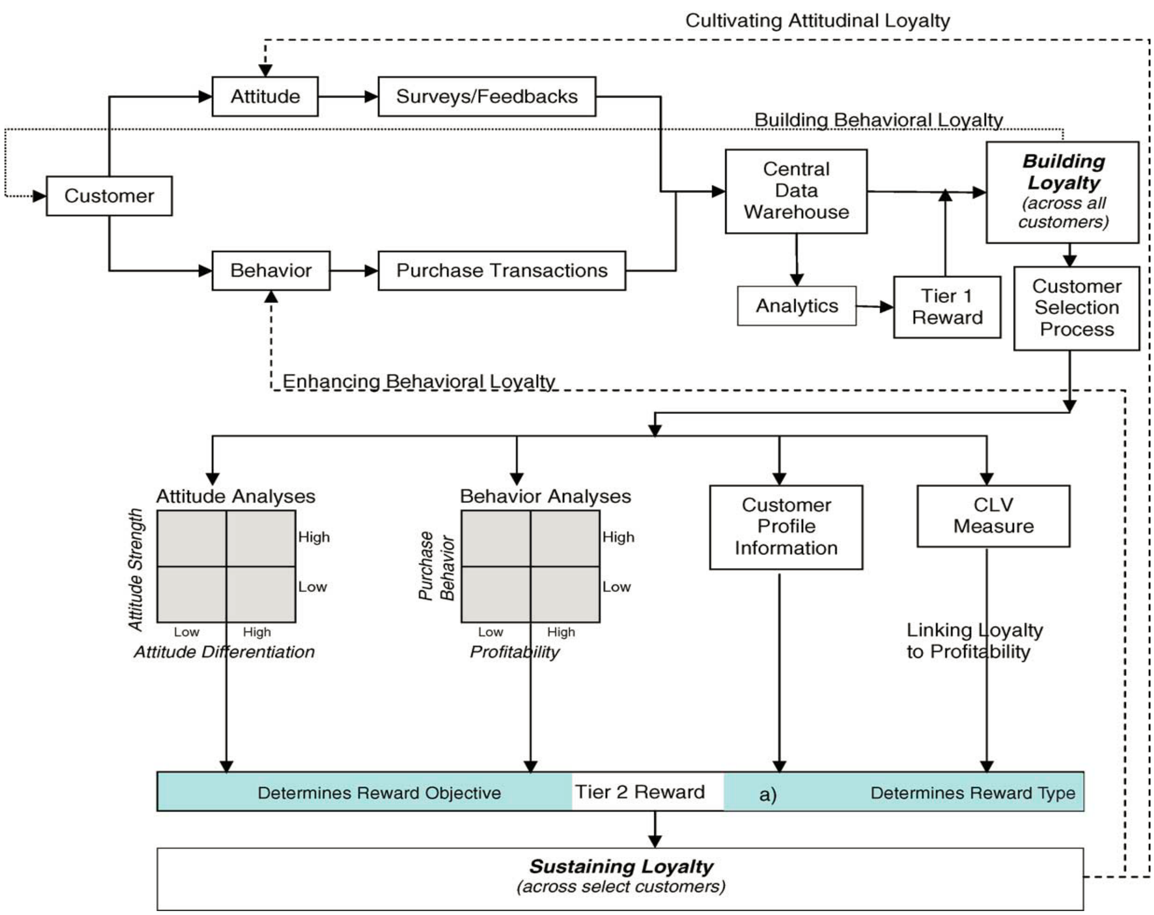

Figure 5. Conceptual framework for building and sustaining profitable customer loyalty. (Kumar & Shah. 2004 p.320)

The components of the framework are explained by discussing three fundamental objectives fulfilled by the framework: Building and enhancing behavioral loyalty, (cultivating attitudal loyalty) and linking loyalty to profitability. The main point in their framework for building and enhancing behavioral loyalty is to move from aggregate level of purchase behavior (everybody who spends 100, gain the same amount of points) to customer level by mapping customers by purchase behavior measures and profitability. This analysis would serve as decision support for marketing function to choose right actions for right group of customers. In the framework, this is presented by a 2-by-2 matrix between these two dimensions. (Kumar & Shah. 2004 p.321) For linking loyalty to profitability, the computation of CLV is discussed as it is a forward-looking metric that incorporates all elements of revenue, expense and customer behavior that drive profitability into one and is consistent with the customer centric paradigm of marketing. In the framework, the CLV is a decision support tool that sets the maximum monetary value limit for marketing investment on a loyal customer and nulls the risk of over-spending (investing more than what the CLV of the customer is). CLV can ensure profitability without compromising loyalty and, being a forward-looking metric, it can be used for pro-active marketing interventions (vs traditional reactive). (Kumar & Shah. 2004 p.322)

A two-tiered rewards strategy is proposed as a possible means to operationalize the suggested framework. The tiers are classified by their end-objectives and the level of differentiation. Tier 1 rewards are said to have the following three objectives. Providing a simple, fair and explicit reward system to reward all customers for their present and past behavior, without including attitude and purchase pattern factors, to ensure all customers know of the reward program. Provide means to capture customer transaction data. And to ensure scalability of the LP by rewarding customers in proportion to their spending (ideally based on profitability, practically more feasible to be based on spending). Hence Tier 1 would be a standard unidimensional reward strategy that gives rewards or points based on customers total spending and serves as means for instant gratification. Tier 1 can be administered at an aggregate level for building loyalty and would be stated as general policy.

Tier 2 rewards are described as forward-looking rewards that are aimed for influencing customers behavior and/or attitude in the future based on the performance of the customer observed on attitudinal and behavioral dimensions of the framework. This makes the tier 2 to be special rewards to select customers for cultivating attitudinal loyalty and/or enhancing behavioral loyalty. The tier 2 rewards are internally controlled by the firm to decide ‘who’ should receive them?; ‘what’ the reward should be?; and ‘how’ much it should be worth? For tier 2 rewards customers are chosen by customer selection process (measuring CLV for each customer). High and medium CLV customers are extracted, as they represent high value customers, and queried on four parameters: attitude and behavioral analyses to specify the objectives to be fulfilled by tier 2 reward for a specific customer, and customer profile information and CLV to determine the type and value of the tier 2 reward. Thus tier 2 rewards are highly differentiated and awarded selectively at individual level to customers that the firm is interested in sustaining loyalty. The rewards are also more invisible to the competition as they are administered at the discretion of the firm on a customer-by-customer basis. For designing the right reward for the right customer, research and systematic data mining of the frameworks components can enable the development of an algorithm to find the most optimal and relevant reward.

The operating on these two tiers concurrently can give considerable flexibility to any LP and most importantly can help achieve attitudinal and behavioral loyalty and profitability simultaneously and give marketers the power to invest in their best customers today based on their future potential, not just history of transactions. (Kumar & Shah. 2004 p.322-324)

2.4. Customer value and segmentation: current and future metrics

Petersen et al. assess marketing literature regarding marketing metrics in their article “Choosing the Right Metrics to Maximize Profitability and Shareholder Value” published in the journal of retailing in 2009. They develop a framework identifying key metrics that can give firms a better picture of how they got were they are and insights on growing in the future. Then they bring up challenges that need to be addressed to being able to build capabilities of collecting right data, choosing right metrics, and linking the metrics to customer value and firm performance. The article lists and address five questions that managers need to answer to choose the appropriate metrics: 1. What metrics are now in place in different firms? 2. What metrics should be in place? 3. How strategic actions can be linked to these metrics? 4. How these metrics relate to customer value to the firm and firm performance? 5. What are the challenges faced when migrating to these new metrics? (Petersen et al 2009 p. 95-96)

The articles review of marketing metrics divide the metrics into seven categories: (1)Brand, (2)Customer and (3)Word of mouth/referral value metrics, (4)Retention and acquisition, (5)Cross-buying and up-buying, (6)Multi-channel shopping and (7)Product return metrics. (Petersen et al 2009 p. 97) The paper discusses all of the categories and provides tables summarizing articles and their findings on each individual category. The findings on categories 2, 4, 5 and 6 are presented more in detail as they are the most relevant for this thesis.

For customer value metrics, the paper brings up two from the vast amount of metrics developed:

customer lifetime value (CLV) for individual level and customer equity for the aggregate level. The authors say that to this point the purpose of these measures has been for optimal customer selection in marketing campaigns and to measure effectiveness after. They refer to a research that found a link between CLV and shareholder value, where marketing campaigns that increase customer value will increase shareholder value, but say that this is only the start of improving CLV measures and linking of CLV to financial outcomes. (Petersen et al 2009 p. 100)

For retention and acquisition metrics the paper mentions RFM (Recency, Frequency, and Monetary Value) and CLV as metrics that should be reviewed and linkages studied by firms. The authors state that the ideal way is to maximize profits from CLV by simultaneously managing acquisition and retention of customers. Once links between RFM and CLV are established the next step is marketing strategies that use metrics like RFM to increase CLV. They mention a research that found out that series of direct marketing campaigns or loyalty programs that build affective commitment potentially lead to increases in future customer spending.

Cross-buying and up-buying give firms a chance to increase revenue and profits from current customers as it has been shown that customers that cross-buy are more profitable than customers that do not. For implementation of strategies to increase cross- and up-buying it is important to determine: 1. Which customers are likely to cross-buy, 2. Which products they are likely to buy, 3. What marketing message to send, 4. When are those customers likely to purchase. To address these issues two metrics are mentioned: average interpurchase time to identify ideal customers within firm's databases to be responsive for cross- and up-selling campaigns, and Next-Product-to-Buy model to assist with identifying customers likely to buy a specific product. (Petersen et al 2009 p. 101)

Multi-channel shopping metrics answer to the challenge of understanding how each channel impact customer purchase behavior and customer profitability. Research has found that customers who shop in multiple channels are more profitable than customers who shop in a single channel. Multi-channel users are also found to score highly on metrics like revenue, likelihood to stay active and future profit potential. (Petersen et al 2009 p. 101) The paper organizes the metrics into four groups (at the customer and store levels for the current and the future) and visualizes it with a following figure of Metrics we need to know.

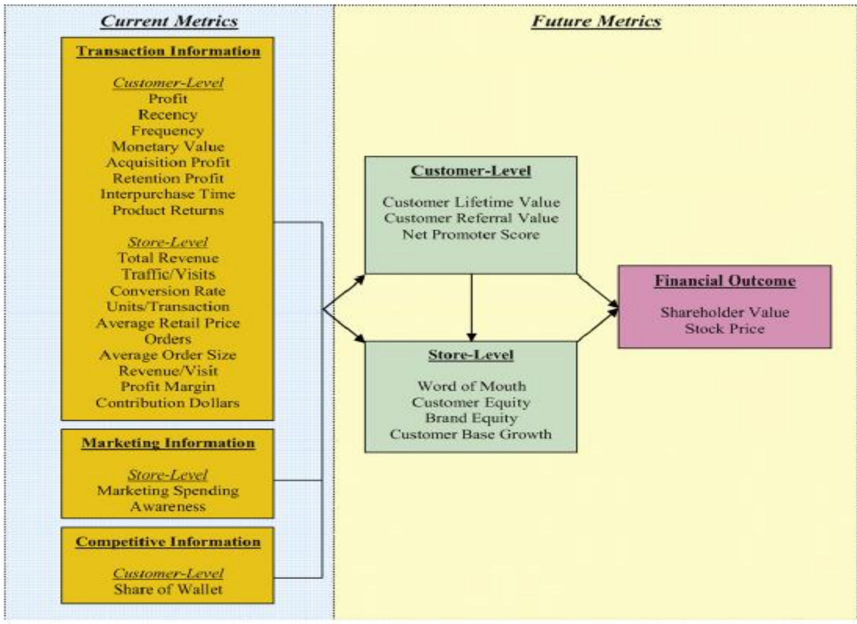

Figure 6. (Petersen et al 2009 Metrics we need to know. p. 103)

Thakur and Workman use Customer Portfolio Management (CPM) approach to examine how companies can define the value of customers and segment customers into portfolios in their article published in the Journal of business research in 2016. The segmentation of customers can help a company to better understand the importance of each customer to the company's profit and help with retaining valuable customers and creating additional value by relationship development. The article introduces a conceptual framework of the CPM matrix focusing on two issues: 1. cost to serve and 2. value of a customer to the firm. The framework segments the customer base of a company into four portfolios, platinum, gold, silver, and bronze. The article defines the portfolios in a following way:*" Platinum customers are very loyal, highly profitable, and less demanding, whereas gold customers are loyal and profitable but more demanding than platinum customers because they require higher costs. On the other hand, silver and bronze customers are defined as low value customers. Bronze customers are least profitable and highly demanding in nature".* (Thakur & Workman 2016 p. 4095-4096) The CPM matrix's focus is on allocating resources to the customers who are deemed to offer more value to the firm according to the segmentation to the four categories based on the customers' lifetime value. This would result in that the firm is able to target more marketing resources toward customers of greater value. The paper builds on the traditional CPM matrix by linking cost and value, strategy and external environment to the CPM process. This is demonstrated by three matrices:

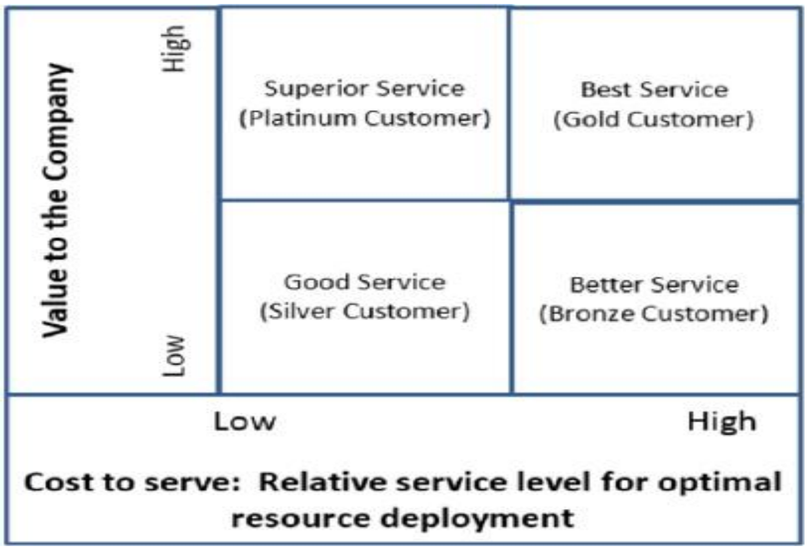

Figure 7. Basic CPM matrix (Thakur & Workman 2016 p. 4096)

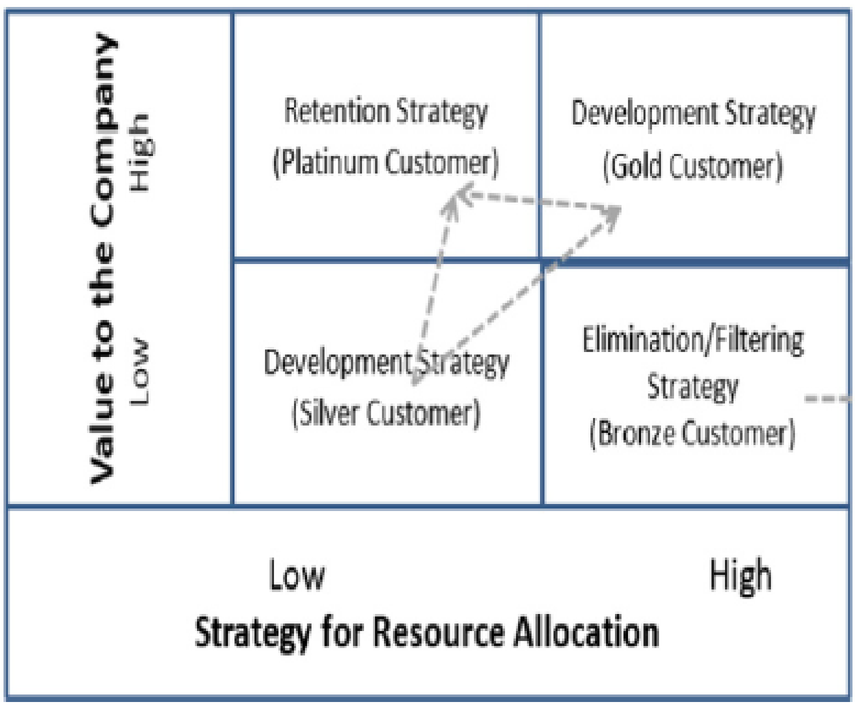

Figure 8. Strategic resource Allocation based on Customer Value (Thakur & Workman 2016 p. 4098)

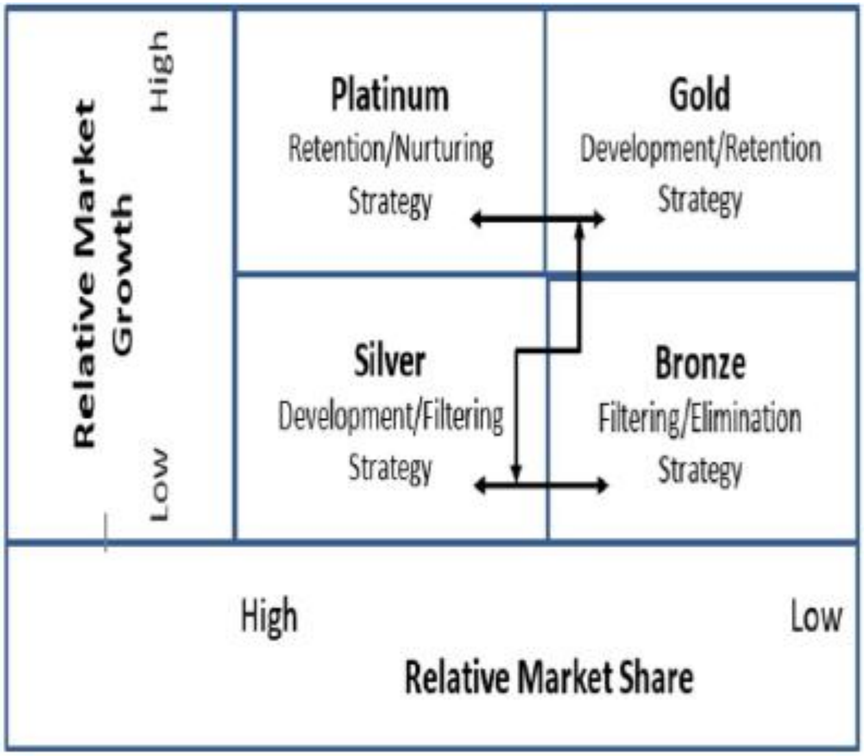

Figure 9. CPM and the External Environment (Thakur & Workman 2016 p. 4098)

The study provides theoretical implications: segmentation of customer base into portfolios, attempt to answer a key question of which customer relationships to develop, maintain or discard, gives strategies that can be applied to portfolios to enhance market value and for adjusting firms resource allocation, and applies the Boston Consulting Groups (BCG) growth/share matrix to customers. And managerial implications: identifying, nurturing and developing valuable customers to grow customer share, shifting of focus to customer value from financial metrics, gaining ability to convert customers to higher level relationships (for example from silver to gold) and gaining informed understanding of key customers that have large effect on the profitability of the company. *(Thakur & Workman 2016 p. 4100)*

In their article published in 2004 in the journal of interactive marketing Kumar et al. present two CLV computation approaches and best practice applications and address the challenges firms can face in implementing the CLV approach to marketing. First, they define CLV as the sum of cumulated cash flows (discounted with the weighted average cost of capital) of a customer over his or her entire lifetime with a certain firm. The two CLV approaches presented are firm-level average CLV and customer-level individual CLV. (Kumar et al. 2004 p. 61)

They present the following formula for average CLV that starts to resemble a real business situation:

$$\text{CLV} = \frac{1}{\sum_{k=0}^{\infty} n_k}\left[\sum_{k=0}^{\infty}\frac{n_k}{(1+d)^k}\sum_{t=k}^{\infty}\left\{\frac{(GC_{t-k}-M_{t-k})}{(1+d)^{t-k}}r^{t-k}\right\} - \sum_{k=0}^{\infty}\frac{n_k A_k}{(1+d)^k}\right]$$

where,

n_k= number of customers in the kth cohort

GC= the average gross contribution

A= the average acquisition cost per customer

M= Marketing Costs per customer

r=rate of retention

d= cost of capital

Equation 1. Average Customer Lifetime Value formula (Kumar et al. 2004 p. 61)

For the individual level CLV approach the following formulas are presented for probability of activity of a customer (P(Active)) and CLV calculation of Net present value(NPV) of expected gross contribution (EGC) from customer in particular period:

$$P(Active) = (T/N)^n$$

where,

n= the number of purchases in the observation period

T= time elapsed between when the customer was acquired and the most recent purchase

N= time elapsed between when the customer was acquired and the period for which P(active) needs to be determined.

Equation 2. individual level CLV approach probability of activity of a customer (Kumar et al. 2004 p. 63-64)

$$\text{NPV of EGC}_{it} = \sum_{n=t+1}^{t+x} P(Active)_{in} * \text{AMGC}_{it}\left(\frac{1}{1+d}\right)^{n}$$

where,

$AMGC_{it}$ = average gross contribution margin in period t based on all prior purchases

d = discount rate

i = customer

t = the period for which NPV is being estimated

x = the future time period

n = the number of periods beyond t (acting as a counter)

$P(Active)_{in}$ = the probability that customer i is active in period n

Equation 3. CLV calculation of Net present value of expected gross contribution from customer (Kumar et al. 2004 p. 63-64)

3. Methods: CLV Prediction

This section describes the approaches applied on the case data for CLV prediction to see if a usable model is possible to be made based on the LP members purchasing data. Two different approaches were chosen for this: the Pareto/NBD-model with Gamma Gamma Monetary-model and a multivariate linear regression model.

3.1. Pareto/NBD Model and Gamma Gamma Monetary Model for CLV

The Pareto/NBD-model is a probabilistic model for CLV modeling in a non-contractual context. It requires three past purchasing behavior measures of every customer: recency (when the last transaction happened), the length of time of observed purchasing behavior and frequency (how many transactions happened within the observed time). Based on these three measures the Pareto/NBD model's parameters can be estimated by fitting a probability distribution to observed values, and a forecast of future customer activity can be made (Glady et al. 2009). As the models forecast gives the estimate of the number of future transactions, the Gamma

Gamma Monetary Model is applied to generate the CLV forecast. This is done by multiplying the number of future transactions in a future period by the Pareto/NBD model with the expected average purchase of the customer given by the estimated population mean of the Gamma/Gamma Monetary Model.

$$\widehat{\mathrm{CLV}}_{i,h} = \sum_{k=1}^{h} \frac{(\hat{x}_{i,T_i+k} - \hat{x}_{i,T_i+k-1})\hat{m}_i}{(1+d)^k}$$

where,

$\hat{x}_{i,T_i+k} - \hat{x}_{i,T_i+k-1}$ = the expected number of transactions during k^{th} future time period

m_i = the monetary value

d = discount rate

Equation 3. CLV estimation with Pareto/NBD model (Glady et al. 2009 p. 7)

3.1.1. Pareto/NBD Model and assumptions

For the Pareto/NBD model (initially proposed by Schmittlein et al. in 1987) to forecast future activity, based on parameters estimated on past observations, there are five assumptions (Glady et al. 2009) to be made about the buying event process and time a customer stays active:

- While active a customer makes purchases according to a Poisson process with rate n_i
- Every customer stays active during a time being exponentially distributed with death rate u_i
- The purchase rate n_i for different customers is distributed by a Gamma distribution across the population of customers
- The death rates u_i are distributed according to a different Gamma distribution across customers
- The purchase rate n_i and death rate u_i are considered as distributed independently of each other.

3.1.2. Gamma Gamma Monetary Model assumptions

Fader and Hardie present the derivation of the gamma gamma model of monetary value and the details of how to estimate the model parameters for predicting likely spend per transaction at the customer level in the future in their article from 2013. The method assumes that monetary value follows a gamma-gamma distribution.

The model is based on three assumptions (Glady et al. 2009):

- The monetary value of a transaction varies randomly around their average transaction value.
- Average transaction values vary across customers but do not vary over time for any given individual.
- The distribution of average transaction values across customers is independent of the transaction process.

3.2. Linear Regression Baseline Model CLV

A multivariate linear regression model can be built on CLV for existing customers. This is done by taking a certain time periods purchases as independent variables and the total revenue in the observation period as the dependent variable and building the regression model on this data (DataCamp, Customer Lifetime Value, 2018). This is formulated as:

$CLV = \beta_0 + \beta_1 X_1 \ldots \beta_{ith} X_{ith}$

where,

CLV	= total purchases of a customer in the observation period
β_0	= intercept
$\beta_1 X_1 - \beta_{ith} X_{ith}$	= monthly purchases made by the customer in the first six months of the observation period

Or:

$CLV = \beta_0 + \beta_1 X_1 \ldots \beta_{ith} X_{ith} + \beta_7 X_7$

where, $\beta_7 X_7$ is added to the equation as the value of the voucher awarded to be spend in the second half of the observation period.

Equation 4. CLV estimation with linear regression models

4. Case of a company’s loyalty program

The case of this study is to analyze the data of a loyalty programs members. To answer the research questions the data is analyzed from different perspectives: firstly, finding out if the LP is profitable in the time horizon of one year, if the reward structure is optimal and whether there should be bigger reward classes. Secondly predictive models are built for CLV estimations of newly signed LP members, where models are built with and without the LP’s vouchers awarded to the customers to see if the voucher is a feature that enhances the predictions accuracy on future spending.

4.1. Data

The data for this master ‘s thesis research is collected from a Finish company’s loyalty program databases. The data is at the customer level and it is detailed. It also contains the marketing information on the customer level. There are two limiting factors in the data: the customer level margin that weren’t available from the database at the time and the time span of only one year. The data will be described more in detail in the following subsections.

4.1.1. Data description

The data to be used in this thesis consists of one year’s purchases under the customer loyalty program between July 2016 and June 2017, email-marketing results (10/2016-6/2017)) and data on the usage of the reward vouchers. The reason for using only this data is that the older data was in an older system and not easily compatible with the data from the new software and considering the scale of this thesis the decision to include only the data from new system was made.

The original purchases data set had 649 927 observations with the following information:

- Customer integration Id
- Mailing Country
- Sign date to the loyalty program
- Basket
- Channel Name
- Channel
- Items purchased
- Receipt Id
- Purchase value
- Purchase date

The usage of vouchers data set had 76 957 observations and had the following information:

- Customer integration Id
- Name (number code)
- Currency
- Value of the voucher
- Balance on the voucher
- Voucher starting period
- Status of the voucher
- Receipt Id
- Voucher amount used
- Purchase date
- Total of the purchase
- Voucher ending period

4.1.2 Background from the data

In the time period July 2016 to June 2017 the loyalty program had over 300 000 active members that made an observable transaction. Over half of the active members came from Finland and the rest are mostly from other European countries. For the remainder of this case research the data will be consisting of only Finnish members. It is not possible to use the whole member base as one entity because of the size of the data and available resources and because different markets might act very differently. The aim is to use the sample of Finnish members as an example for analysis to be conducted throughout the loyalty programs member base.

4.1.3 Sampling/Limitation for analyses

Due to the size of the data sets, the data is first filtered by demographic area and the focus from here on out is on Finnish customers of the LP. The purchase data set was filtered based on Mailing Country to include only the Finnish customers. This resulted in a data table with 368 840 rows/observations.

The Data was further filtered by excluding based on the sales channel: The channels outside of Finland were filtered out of the data with one exception: one channel close to the Finnish boarder was kept as it has a significant amount of purchases (>6000) in the data of Finnish customers. These observations were transformed to euro to get the data comparable to other observations. This resulted as data table of 368194 rows/observations. Which tells that only 646 observations were removed by removing channels that are outside of Finland or otherwise barely used by Finnish LP members.

The remaining 28 channels are in general outside of Finland or generated together only 626 purchases with the Finnish customers of the loyalty program.

VoucherPeriodStartDate_VouchIssuedFinn	VoucherPeriodEndDate_VouchIssuedFinn	GroupCount
01-Jul-2016	31-Dec-2016	12285
01-Jul-2016	09-Jan-2017	1
01-Jul-2016	31-Jan-2017	17
01-Jan-2017	30-Jun-2017	22756
01-Feb-2017	30-Jun-2017	3

Table 1. Voucher periods observed in the data

The check of the issued vouchers showed that there were 21 vouchers with differing issue/expired dates from the standard dates of the vouchers. These were filtered out of the data as outliers. After this the data included only vouchers with the correct issue/ expiration dates of 1.7.2016 – 31.12.2016 and 1.1.2017 – 30.6.2017.

4.1.4 Voucher system: reward structure and limits

The vouchers that are sent out twice a year are the biggest benefit the members get. They are created according to the purchases an individual LP member makes: shopping at any of the company's store, outlet or webstore can increase the amount of the voucher.

Currently the vouchers have four limits with the following structure:

Six month purchases at least	Voucher	Max discount
150.0 €	10.0 €	6.67%
250.0 €	20.0 €	8.00%
500.0 €	50.0 €	10.00%
1 000.0 €	100.0 €	10.00%

Table 2. Voucher limits, awarded voucher, maximal possible discount

The vouchers are written in February and August and they are usable in all stores. In general, the vouchers are valid from the start of January or July and expire in the end of June or December. In this research, the data includes vouchers from two periods: 1.7.2016 – 31.12.2016 and 1.1.2017 – 30.6.2017.

Regarding to the vouchers it has not been regulated for what they are earned and what they are spent on. The worst scenario from the LP's managements point of view would be a member, who buys only highly discounted products, gets a voucher according to the purchases, and then uses it on highly discounted products. (Kuokka, K. 2014. p. 41-42)

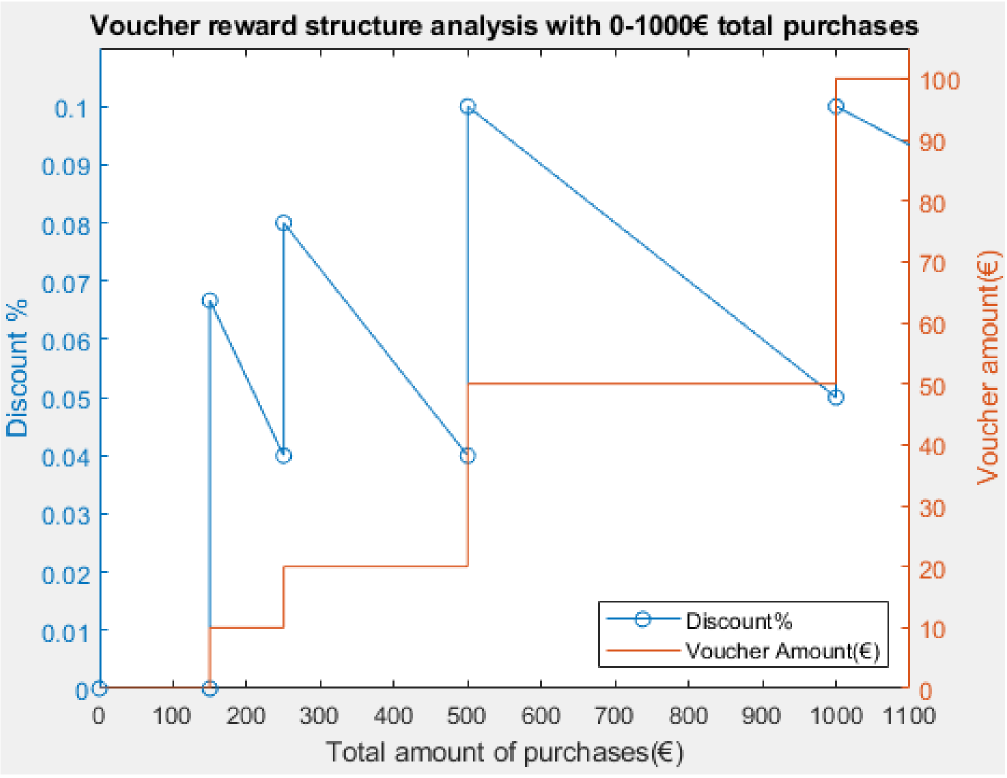

Chart 1. Visualization of the relation of awarded Discount % to total spending and voucher value in the loyalty programs reward system

From the graph, we see how the different valued vouchers (10, 20, 50, 100) discount percentage corresponds to the amount spent to earn a voucher with a certain value. The blue line shows that when between two voucher limits the discount percentage gets lower until reaching the next limit. Also notable is that after spending 1000€ to earn the biggest voucher (100€), further spending is not rewarded by the system as the 100€ voucher will correspond to smaller discount % the more you buy (for example spending 2000 € will be rewarded with 100€ voucher which would correspond to 5% discount %).

4.1.5 Descriptive statistics on the voucher issue and usage data of Finnish customers

29772 unique Finnish LP customers, who shopped on the Finnish channels (incl. Webs stores), were awarded 34 299 vouchers (11 889 for the 1.7.2016 – 31.12.2016 and 22 365 for the 1.1.2017 – 30.6.2017 period). The issue and usage of vouchers by vouchers € value is following in the two time periods:

Value of Vouch Issued	Part of vouchers with certain value	% of issued € value	Percentage of issued vouchers getting used
10.00	55,5%	28,94%	60,90%
20.00	33,1%	34,59%	66,30%
50.00	8,8%	22,95%	76,80%
100.00	2,6%	13,52%	83,80%

Table 3. Voucher issue and usage for 1.7.2016 – 31.12.2016

Value of Vouch Issued	Part of vouchers with certain value	% of issued € value	Percentage of issued vouchers getting used
10.00	55,5%	29,24%	31,74%
20.00	33,0%	34,78%	38,52%
50.00	9,2%	24,12%	52,39%
100.00	2,3%	11,86%	63,89%

Table 4. Voucher issue and usage for 1.1.2017 – 30.6.2017

From the tables, we see that the more valuable the voucher is the more likely it is to be used for purchases (has one or more receipt id included in data). The Issued € voucher value is already calculated from the sole voucher data but is discussed more in detail in the chapter 3.2. where it is compared with the analyzed purchase data in relation to purchases made and vouchers value used.

Next the voucher usage is looked from time perspective: when the different value vouchers are used and how much before expiry they are used.

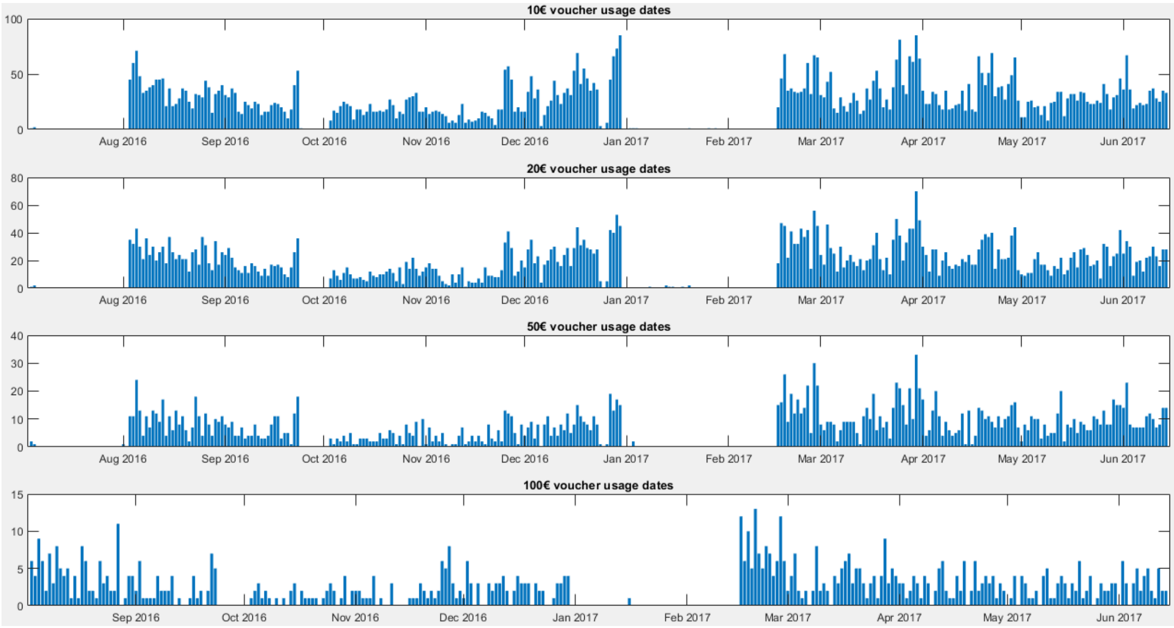

Chart 2. Usage of the vouchers in the usage period

The vouchers seem to be used rather evenly on the issue periods (1.7.2016 – 31.12.2016 and 1.1.2017 – 30.6.2017). The data show two date period gaps coming from the fact that the vouchers are written in February and August and so aren't usable before that.

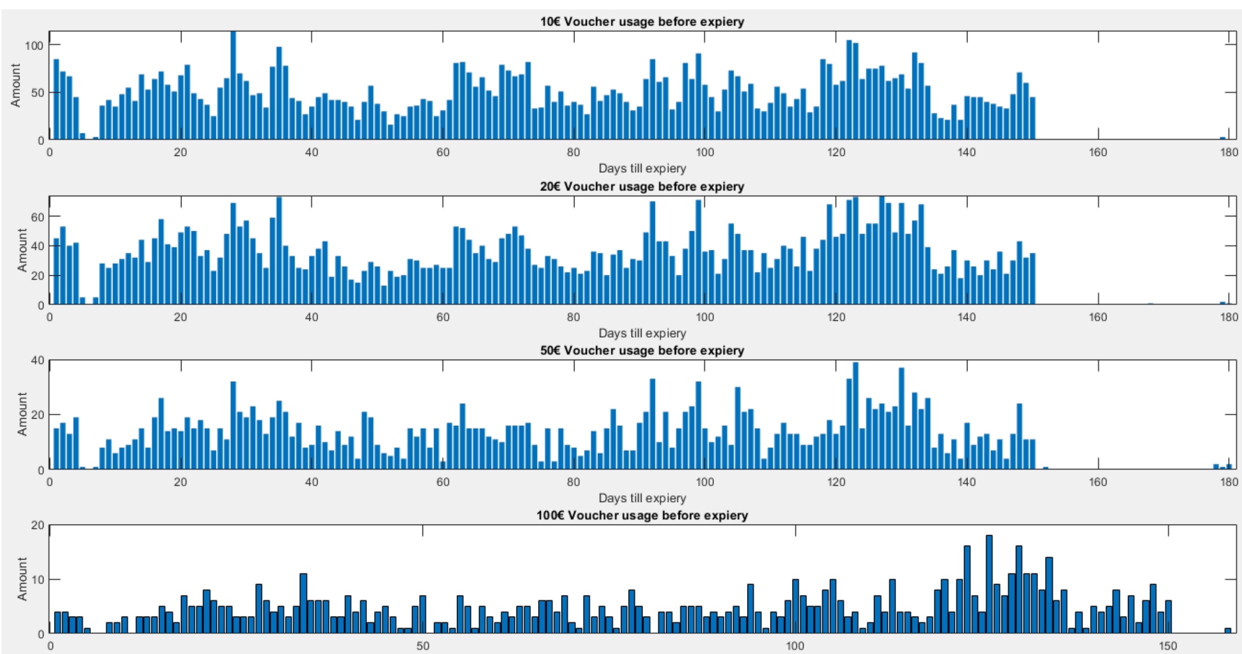

Chart 3. Usage of the vouchers before expiry

This figure shows also that the vouchers are use evenly on period of being valid. There seems to be a small last-minute usage boost on the last days of being valid (last 5 days).

4.1.6 Descriptive statistics on the data of Finnish customers

In the time period of 2.7.2016 to 30.6.2017, +150 000 Finnish members of the LP made more than 350 000 purchases, that were registered in the database. The smallest purchase value on single receipt is 0.09 € and the largest is 6009,92 €. The mean of purchase values is 57.98 €, the median is 36.12 € and standard deviation 78.17 €.

There are 61 channels where Finnish LP members did purchases in the time period. The purchases and voucher usage are:

Sales Channel	Purchases	Monetary spending	Voucher Usage
B2C WEB STORE FINLAND	3,97%	7,54%	10,73%
OUTLET HELSINKI	5,88%	7,47%	7,01%
OUTLET TURKU	5,85%	6,17%	3,81%
OUTLET ESPOO, LOMMILA	4,82%	5,43%	6,41%
OUTLET VANTAA, TAMMISTO	4,65%	5,01%	6,64%
OUTLET BRAND OUTLET	4,11%	4,17%	3,67%
OUTLET LAHTI, RENKOMÄKI	5,08%	4,04%	3,66%
STORE ESPLANADI	2,55%	3,93%	4,22%
OUTLET RIIHIMÄKI	4,22%	3,91%	3,19%
OUTLET HUMPPILA	4,71%	3,81%	3,57%
OUTLET VAASA	4,10%	3,46%	3,54%
OUTLET OULU	3,41%	3,31%	2,26%
OUTLET KOTKA, JUMALNIEMI	3,49%	3,27%	3,67%
OUTLET PALOKKA	3,89%	2,95%	3,48%
OUTLET PIRKKALA	3,56%	2,88%	5,19%
OUTLET JÄRVENPÄÄ	2,90%	2,80%	2,44%
OUTLET KYYJÄRVI	3,11%	2,76%	2,59%
OUTLET KUOPIO, MATKUS	3,17%	2,49%	1,95%
OUTLET JYVÄSKYLÄ, VIHERLAND	2,91%	2,47%	2,44%
OUTLET LEPPÄVIRTA, UNNUKKA	2,64%	2,26%	1,80%
OUTLET NAPAPIIRI	2,45%	2,18%	1,90%
STORE SELLO	1,90%	1,92%	1,53%
MAIN BRAND SHOP	2,32%	1,80%	1,40%
STORE JUMBO	1,79%	1,78%	1,76%
STORE ISO OMENA	1,70%	1,78%	1,74%
OUTLET LAPPEENRANTA	2,37%	1,76%	1,52%
OUTLET HAPARANDA	1,69%	1,71%	1,18%
LIELAHTI	1,41%	1,62%	0,61%
STORE MYLLY	1,30%	1,42%	1,01%
STORE ITÄKESKUS	1,46%	1,42%	1,62%
OUTLET NUUTAJÄRVI	1,11%	1,19%	1,80%
STORE KAMPPI	1,31%	1,11%	1,44%
STORE AIRPORT HKI	0,15%	0,17%	0,20%

Table 5. Finnish LP members purchases (count and monetary) and voucher usage by sales channel

The 33 sales channels based on amount of purchases, monetary value and voucher usage are listed above. The remaining 28 channels are in general outside of Finland or generated together only 626 purchases with the Finnish customers of the loyalty program.

Next the purchases without voucher usage, all purchases with voucher usage, and purchases separately with different voucher values were compared on the aggregate level by calculating interesting comparable statistics for them.

OriginalVariableNames	NoVoucherPurchases	AllVoucherPurchases	Vouch100Purchases	Vouch50Purchases	Vouch20Purchases	Vouch10Purchases
PurchValueMean	57,30	72,88	172,29	104,84	69,44	56,97
PurchValueMedian	35,90	47,68	109,91	70,04	47,80	37,80
PurchValueStdeviation	77,04	98,47	246,02	126,75	74,87	67,41
Vouch € used / Issued		50,44%	68,63%	58,88%	46,81%	40,13%
VouchAmountUsedCount	352033,00	16155,00	734,00	2161,00	5530,00	7721,00
VouchAmountUsedMean	0,00	20,38	75,93	42,18	19,18	9,88
VouchAmountUsedMedian	0,00	10,02	99,99	50,00	20,00	9,99
VouchAmountUsedStdeviation	0,00	18,40	33,00	14,43	3,21	0,98
MoneySpentOverVouchAmount UsedCount		16155,00	734,00	2161,00	5530,00	7721,00
MoneySpentOverVouch AmountUsedMean		52,50	96,36	62,66	50,26	47,09
MoneySpentOverVouch AmountUsedMedian		27,33	19,58	25,00	28,00	27,80
MoneySpentOverVouch AmountUsedStdeviation		93,79	238,59	123,25	74,45	67,35
PercentageOfOnlyVouchUsed		10,69	28,61	21,19	10,36	6,29
VouchAmounts VS MoneySpentOver		2,58	1,27	1,49	2,62	4,76
MoneySpentOverVouchAmount UsedSumTotal / Issued € Voucher Value		1,30	0,87	0,87	1,23	1,91

Table 6. Finnish LP members purchases divided by voucher values, spending over the voucher value and descriptive statistics for comparison

From this table, many useful insights can be derived:

- average purchase with no voucher is 57.50 €, whereas average purchase with vouch used is 72.88 € with 52.50 € spent over the vouchers value.
- The more valuable the voucher the more likely it is to be used (Vouch100 69% of issued value used vs Vouch10 40%) and more likely the voucher is used on its own (Vouch100 28% used on its own for a purchase vs Vouch10 6%)
- Issuing smaller value vouchers generate more spending over the vouchers value
-

4.1.7 Points pressure: getting over the closest voucher level

For looking at the points pressure of the current reward system the purchases data aggregated by months was utilized by dividing the months based on the earning periods (07/2016-12/2016, 01/2017-06/2017) and calculating the cumulative sums month by month. The possibility of points pressure is looked from the transformed data set by observing the cases where LP members are below a certain voucher award limit before the last month of reward earning period and looking at how many make a purchase and get over the limit and how many make a purchase but don’t by enough to get over the limit.

Voucher earning period	Month before earning periods end. Cumulative purchases between	LP members who's purchases under limit and who did purchases in the last month	Part of customers whose purchases got them over the limit in last month
Winter 16	50 - 150	6174	44,90%
Summer 17	50 - 150	4497	41,43%
Winter 16	150 - 250	1972	55,43%
Summer 17	150 - 250	1455	50,86%
Winter 16	250 - 500	1285	30,89%
Summer 17	250 - 500	949	25,40%
Winter 16	500 - 1000	367	19,62%
Summer 17	500 - 1000	300	20,33%

Table 7. Points pressure by voucher earning period and voucher limit with the Finnish LP members who made purchases in the last month of the voucher earning period.

There seems to be some points pressure with the LP members for the two lower voucher earning limits (150 and 250), as the percentage of customers that purchase over the limit is rather high. With the two larger voucher limits the percentages are smaller so probably the points pressure is not so evident. This could also be because the needed spending gets bigger with the two biggest voucher limits.

4.2. Research method

To answer the last research question of “Is it possible to develop new and usable measures from the collected data (KPI)?” predictive CLV models were built and compared to predict the future spending in the second voucher period (CLV) based on the purchase behavior in the first voucher period (07/2016 – 12/2016) after signing up as a member to LP.

Two CLV based approaches, for non-contractual setting, were chosen to analyze the, newly signed up, Finnish LP customers data. The idea is to build a multivariate linear regression model and probabilistic pareto/NBD model to predict the CLV of the whole timespan and compare which perform better.

For this predictive model part of the research, the scope is further limited to the new Finnish members of the LP as for this cohort of customers we have the transactional data from the beginning of the membership. So, the cohort includes the LP members who signed up to the LP and made a purchase during the first half of the observation period (4th of Jul 2016 – 25th of December). This filtering, to LP members that we have the whole purchase behavior history, was done to assure better comparability of the models and the results as the data is more consistent compared to a case where part of the customers might have had from one up to nine years of purchase history as a LP member that is not included in the data provided. The filtering resulted in a cohort of 18268 Finnish LP member customers.

The linear regression model utilizes the purchases of the first six months (07/2016 – 12/2016) purchase behavior and voucher amount awarded (to be spent in the next six months) to predict the customers CLV for the whole time period (07/2016 – 06/2017). The pareto/NBD model uses the RFM values of the first six months to predict the repeat purchases for the second six-month period and derive the CLV based on the predicted repeat purchases and average monetary value per transaction. The use of vouchers is taken into count in the CLV calculations by using the actual money spent on a purchase for the calculations (purchase values minus the voucher amount used for the purchase).

A notably difference between the models is that for the regression model the data needs to be split into training and test sets (70/30) for being able to estimate the performance of the model, this is not the case for the Pareto/NBD model, as the CLV can be predicted without having a single value observed for it.

The analyses were done using mainly MATLAB, Excel was used for some data pre-processing phases (also for the Gamma Gamma monetary model's average transaction distribution estimation, the Solver analysis tool was used). For linear regression model and purchase point analyses the long format purchase data was transformed into wide format aggregated to month level: month and year information was extracted to separate column from the purchase date column and the data was unstacked to get a table where the columns are the months from 07/2016 to 06/2017, the rows are customer ids that were active during the period and the values are the monthly purchase amounts in euros. For the Pareto/NBD -model the data was transformed into wide format aggregated to weekly level (52 weeks, starting from 4th July 2016 to be the first day of week 1 and ending into 30th June 2017 as the last day for week 52). For the pareto model the week-level data was divided into two periods: first 25 weeks for model building and weeks 26 to 52 for evaluation of the model (matching the six-month time periods used in linear regression model). Both the count and monetary values of weekly transactions were collected to gather the necessary data in right format to generate predictions for the clv of individual customers.

4.2.1 Linear regression model for CLV prediction

The CLV was calculated for every customer in the filtered cohort of newly signed LP members by summing every month's purchases together (the voucher usage was considered by deducting the used vouchers amount from the purchase amount where it was used based on receipt data). For the multivariate linear regression model building the purchases of the first six months (07/2016-12/2016) were chosen as feature variables (alternative model was also built where the voucher amount rewarded for the next six months was added as 7th independent variable) and the "CLV" was chosen as the target variable:

$CLV = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$

where,

CLV	= total purchases of a customer in the observation period
β_0	= intercept
$\beta_1 X_1$ to $\beta_6 X_6$	= monthly purchases made by the customer in the first six months of the observation period

Or:

$CLV = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 V_7$

where,

$\beta_7 V_7$ is added to the equation as the value of the voucher awarded to be spend in the second half of the observation period.

Equation 5. CLV estimation with linear regression baseline models with and without awarded voucher value

First the train-test split was set to 70/30: 70% of the rows for building the model and 30% to test the model. To get more reliable results a 10-fold cross validation method was used: the data was split into 10 folds and every fold was used as test set when other folds were used as training set, the mean values were calculated from the different model's performance measures. These mean measures of accuracy are presented in the model comparison table. The data of all the newly signed and active members within the first six months was used to train and test the model. Other grouping/filtering ways were also tested: the different voucher award groups and web store users were filtered and tested if the built model was better compared to the initial one that had all the active LP members included. But these results aren't presented as they are not comparable with the Pareto/NBD models results.

4.2.2 Pareto/NBD model for clv prediction

For the Pareto/NBD model based clv prediction a new data set was generated from the weekly level purchases data described in the end of last paragraph of 4.2. The data set includes the customer level RFM and monetary measures to generate predictions on CLV: Customer ID(ID), time for repeat purchases in weeks (T, the time between the first purchase and 25th Dec 2015, last day of period the period used for model estimation), time of last purchase in weeks (t_x, time between last purchase and 25th Dec 2015, last day of the period used for model estimation), number of repeat purchases in model estimation period (p1x, weeks 1-25, 4th Jul 2016 – 25th Dec 2016), Average monetary value of repurchase in model estimation period (zbar). And measures needed to evaluate the model's predictions: the actual number of repeat purchases in model prediction period (p2x, weeks 26 - 52, 26th Dec 2016 – 30th Jun 2013), the actual monetary value of customers repeat purchases in model prediction period (M2x, weeks 26 - 52, 26th Dec 2016 – 30th Jun 2013).

The measures T, t_x and p1x are used to estimate the Pareto/NBD model's parameters and measures p1x and zbar to estimate the Gamma Gamma Monetary Models parameters. After the two models' parameters are estimated, the models are used to get the estimated count of repetitive purchases for the forecast period (in this case 26 weeks) from the Pareto/NBD model and the weighted average monetary value of a purchase from Gamma Gamma Monetary Model. These are used to generate the CLV estimations which are then compared to the actual monetary values (M2x) of customers in the second period to estimate the predictions performance.

For the Pareto/NBD model implementation the MATLAB scripts from Fader's et al. paper *"A Note on Implementing the Pareto/NBD Model in MATLAB"* article, published in 2005, were utilized (Fader et al.). The code was used partly and slightly modified to give also the predictions of individual level repeat purchases for the prediction period instead of only providing the cumulative repeat purchases of the whole cohort. For the Gamma Gamma Monetary model implementation, the paper from Fader & Hardie published in 2013 *"The Gamma-Gamma Model of Monetary Value"* was followed and the example excel and MATLAB solutions were utilized to generate the weighted average monetary values for repeat transactions.

4.2.4 Comparison of models, measures of accuracy

For comparing the performance of the three models (linear regression with 6-months purchases, linear regression with 6-months + voucher value provided for use in the next period and Pareto/NBD model) presented previously, the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) are calculated between the predicted CLV and the actual value. The formulas of the used measures of accuracy:

RMSE: $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$ MAE: $\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$

where,

y_i = actual CLV value of individual customer

$\hat{y}_i$ = predicted CLV value of the same individual customer.

Equation 6. RMSE and MAE measures for accuracy

With these measures of accuracy, it is possible to compare how well the models worked and check if the added voucher value variable increases the accuracy of forecasting future spending.

4.3. Research results

For the LP's active, newly signed Finnish members, the CLV was predicted for the next six-month (Jan-2017 - Jun-2017) period based on the first six-month period's (Jul-2016 – Dec-2016) purchases. Then the predicted values were compared to the actual value of the CLV. The reader is good to keep in mind the difference between model's estimation approaches: for the linear regression models the data was split to train and test sets (90/10, with 10 – fold cross validation) and the evaluation models' accuracy is done on the test set whereas the Pareto/NBD model is evaluated on the whole data set as split is not necessary because the CLV can be forecasted without having any observations of the actual CLV.

Model	RMSE	MAE
Pareto/NBD model	88.26	57.41
Linear Regression 6-month purchases	79.05	35.18
Linear Regression 6-month purchases + Voucher	**78.86**	**35.01**

Table 8. Model comparison for the CLV prediction using the new Finnish LP members purchase data (10 -fold cross validation done for linear regression models)

From the table 7 we can see that the linear regression model with the voucher value added performs the best by both RMSE and MAE. From this results comparison, we can observe that by adding the voucher value to a prediction model we can get more accurate prediction of the actual future spending as the voucher usage was controlled for in the CLV calculations. Even though the probabilistic Pareto/NBD model performed worst, it can be a useful approach to forecast CLV for new LP members as it only needs three measures for generating the prediction and no train/test split is needed. The model can be a useful tool in the situations when new customers have the three needed measures (recency, frequency, time had for repeat purchases) observed but don't yet have specific enough behavior/characteristics for being placed to a specific customer segment/cohort where a more specified model could be used.

4.3.1 Multivariate linear regression prediction models for CLV

Here the models with the initial individual 70/30 train test split are presented to get a more detailed view on the multivariate linear regression models and the voucher value as a predictive variable.

Linear regression model with new Finnish LP member customer cohort:

y ~ 1 + x1 + x2 + x3 + x4 + x5 + x6

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	7.77	0.80	9.66	0.00
x1	1.14	0.01	107.65	0.00
x2	1.13	0.01	99.07	0.00
x3	1.12	0.01	85.28	0.00
x4	1.20	0.01	86.09	0.00
x5	1.14	0.01	104.08	0.00
x6	1.14	0.01	141.46	0.00

Number of observations: 12757, Error degrees of freedom: 12750

Root Mean Squared Error: 72

R-squared: 0.814, Adjusted R-Squared 0.814

F-statistic vs. constant model: 9.3e+03, p-value = 0

Model’s prediction performance with test data:

RMSE = 92.60 MAE = 33.52

Linear regression model with awarded voucher value added as predictive variable:

y ~ 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7

Estimated Coefficients:

	Estimate	SE	tStat	pValue	
(Intercept)	9.13	1.01	9.02	0.00	
x1	1.09		0.01	76.18	0.00
x2	1.05	0.02	65.35	0.00	
x3	1.05	0.02	55.20	0.00	
x4	1.11	0.02	59.12	0.00	
x5	1.10	0.02	72.88	0.00	
x6	1.16	0.01	77.94	0.00	
x7	0.94	0.16	5.75	0.00	

Number of observations: 12757, Error degrees of freedom: 12749
Root Mean Squared Error: 77
R-squared: 0.799, Adjusted R-Squared 0.799
F-statistic vs. constant model: 7.26e+03, p-value = 0
Model’s prediction performance with test data:
RMSE = 82.23 MAE = 33.28

4.3.2 Pareto/NBD model for CLV

The Pareto/NBD model parameters:

Using starting values of r = 1, α = 1, s = 1, β = 1, MATLAB’s optimization toolbox’s fmincon routine terminates at the following point when it finds the values of r, α, s, β that maximize the log-likelihood function(LL): r = 0.6917, α = 5.3249, s = 0.2695, β = 0.06, LL =27248.57

With these model parameters, the expected repeat transactions within the forecast period of 26 weeks are estimated for every customer and by the whole cohort, to assess the model's performance on the cohort level.

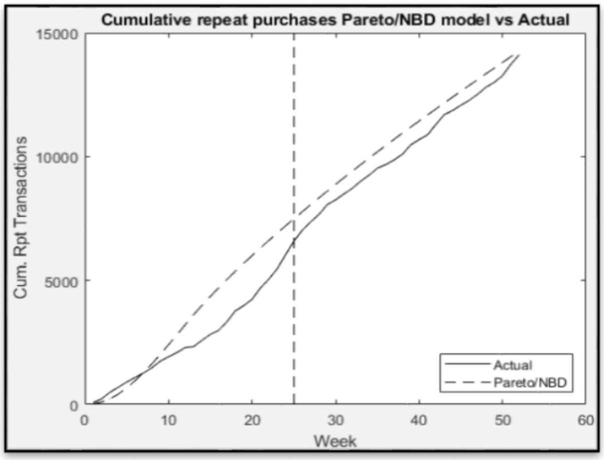

Figure 10. Comparison of the cohort's cumulative repeat purchases in estimation and forecast period weeks 26-52(Jan-2017-Jun-2017), Pareto/NBD model vs actual

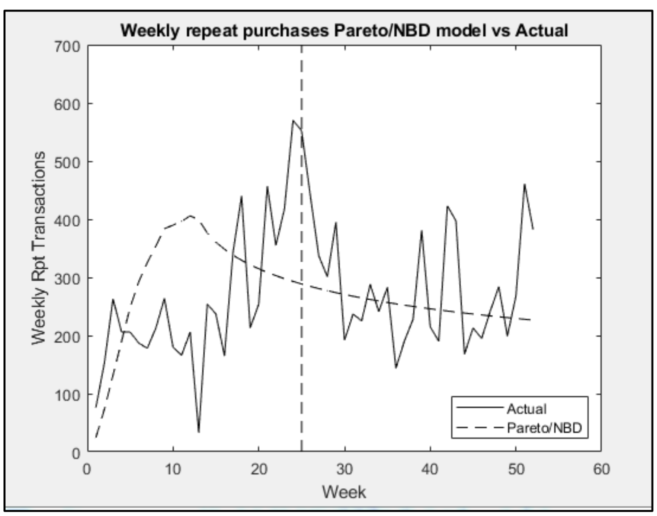

Figure 11. Comparison of the cohort's weekly repeat purchases in estimation and forecast period, Pareto/NBD model vs actual

On the cohort level, the model seems to perform rather well for predicting cumulative repeat purchases. For the weekly level of repeat purchases the model shows indication of the general trend, but the data span could be longer to get a better comparison as now some seasonality (week 25 Christmas time and week 52 last time to use voucher) might be the reason of ill-fitting parts between the actual and the model. The cumulative repeat purchases of the model still indicate that the model could be a useful way to estimate the CLV as the repeat purchases are forecasted rather well for weeks 26-52.

Next the Gamma Gamma Monetary model's average transaction values for repurchases is compared to the actual average repurchase value to see if the model's estimation values can reasonably be used for calculating the CLV forecast of the Pareto/NBD approach.

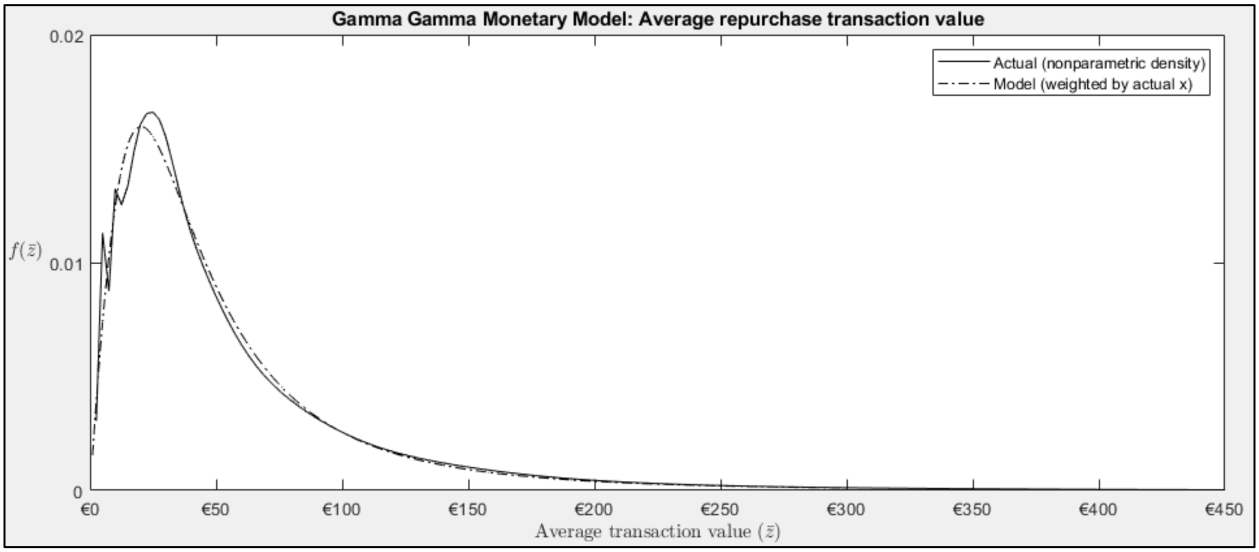

Figure 12. Comparison of the cohort's observed vs theoretical distribution of average transaction value across customers, Gamma Gamma Monetary model vs actual

The Gamma Gamma Monetary model weighted average transaction value distribution seems to fit reasonably well with the actual distribution of average repurchase values.

Then the two models individual level measures (repeat purchases and weighted average monetary value of repurchase) are multiplied together to get the forecasts of individual CLV. Next the comparison of the predicted CLV to the actual CLV was made with following results: RMSE = 88.26, MAE = 57.41

5. Conclusions

Here the research questions are answered based on reviewed literature and research done on the case data: exploratory analyses on all the active Finnish LP members purchase and voucher usage and predictive CLV model for the new active members cohort research.

1. **What general insights can be gained from data-based analysis of a loyalty program by way of (business) analytics methods?**

The loyalty program and its database can be a great source for gaining insights and driving business by analytical methods. The profitability of LPs members and from that the profitability of the LP as a whole can be analyzed and with forward looking metrics, such as CLV, the future profitability of the program can be managed and the structure of the LP could be optimized further. To be able to get a more precise view on the profitability of the LP members, the LP as whole and CLV of the members, the cost side of the loyalty program should be included in the analyses with individual level customer margins. With these added, the actual values of the LPs positive effects that came up in the analyses (added spending with voucher usage, points pressure) could be calculated more precise. This would also make the predictive models CLV predictions more usable as instead of looking at predicted future spending the prediction could give better insight on the profitability of the customers.

2. **How can the loyalty program of the case company be enhanced?**

To answer the question of how to enhance the LP, the current reward system and data generated from the LP was analyzed. Answers on ways to enhance the LP are based on the research results and the literature review and are presented next:

i. **Is the current reward system optimal/profitable?**

Based on the analyses and literary research the LPs reward system can be viewed profitable, but it could be optimized. In the one year’s timespan and using only the used rewarded vouchers as the cost of reward system it can be said that the reward system is profitable as the system is observed to generate extra spending in two ways: spending to earn a certain voucher (points pressure) and extra spending over the value of the voucher when the awarded voucher gets used. But to get a clearer picture of the current profitability the cost side of the LP and individual profit margins would be needed. Also, to verify the effectiveness of current reward system points pressure the LP members purchase data should be compared to non-LP members purchases.

Three approaches to optimize the LPs reward structure came up based on the analysis and literature: the current logic could be tweaked to enhance the effects and the profitability of the LPs reward system by modifying the amount of reward classes and values of the rewarded vouchers. Second approach would be to change the logic of the reward system from rewarding past spending to rewarding LP members based on their individual future value (CLV) to the company. The third approach would be a two-tier reward structure, suggested in the literature, where the first tier is the same kind of reward structure as is now in place: rewarding customers in proportion to their spending (ideally based on profitability, practically more feasible to be based on spending). The second-tier would be based on forward looking metrics like individual CLV of the LP members. The second-tier rewards could be differentiated and awarded selectively at individual level to customers that the firm is interested in sustaining loyalty, and the invested amount for the reward would be limited by the CLV calculation to null the risk of over-spending (investing more than what the CLV of the customer is).

ii. Should there be more and bigger classes in the reward system?

Smaller general class and bigger personalized class could be useful to optimize the reward system. So, a voucher class below 150/10 € could be an improvement, for example 100/5 €, as the data indicates that the smaller valued vouchers generate more spending over the vouchers value in the purchase occasion when the voucher is used. And, the points pressure to get to a certain reward level was more evident in the smaller reward classes. This would also suggest having more classes between 250 and 1000 € limits as it would make the gap to get to the next reward level smaller, which could motivate LP members to spend more to get to the next level, thus improving the points pressure effect in the bigger classes and improve profitability of the system.

It could also be fitting to reward the big spenders with a bigger class in reward system, for example fixed 10% discount voucher after 1000€ purchases. So, individual voucher amount that is based on the previous purchases: 1100 € spent -> 110€ vouch, 1500 € spent -> 150€ vouch, this way the discount awarded would not dilute indefinitely when spending goes beyond the largest voucher limit.

iii. Is it possible to develop new and usable measures from the collected data (KPI)?

Two kinds of measures were developed using the collected data, that could be useful and meaningful for managing and improving the loyalty program: descriptive measures of the reward structure and predictive models that give insight to members future spending.

Aggregate level measures were calculated on the different vouchers values usage: How the use of different vouchers generates spending over the number of vouchers. For example, the 10 € vouchers were used on 7721 purchases for 76309 €, but on these purchases 363545.46 € was spent over the voucher amount whereas with 100 € vouchers were used for 55730.92 € and only 70730.90 € was spent over on these purchases (28.6% of the 100 € voucher usage were cases where only the vouchers value was used). For these indicators, the margins of customers and products would be beneficial to include to get a true picture of actual profitability instead of only spending.

For more individual level indicators two models were built to predict CLV (on spending level) for newly signed active customers to have a forward-looking indicator for helping with marketing effort decisions. The linear regression models seemed to be more accurate compared to the Pareto/NBD model, but the Pareto-model can still be useful with newly signed LP members as the model only needs three measures (recency, frequency, time had for repeat purchases) for forecasting. So, the model could be used to get initial CLV indicator when these measures are observed but the customer doesn't yet have specific enough behavior/characteristics for being placed to a more specific customer segment/cohort where a more specified model could be used.

From the relevant literature review came up the limiting factor of improving data quality: the data quality and the performance of the firm are positively related, but that the profit-maximizing level of data quality is reached with non-perfect data as marginal benefits of improved data quality will decrease with increased cost. This needs to be addressed when investments are considered to improve data quality for more specific measures to get more detailed data or to better the accuracy of predictive models.

5.1. Limitations of research

The main limitations of this work are the fact that only one year's data was available and that the costs of the loyalty program and customer level margins were not available for analysis. This means that the results are not as precise as they could be for the profitability and CLV calculations and the predictive CLV models used were mainly focused on the future spending of the newly signed members.

Another limiting factor for the results on effects and profitability of the LP is the lack of comparable data on non-LP members. Including comparable data on non-LP members to the analyses would validate which effects are evident because of the LP and the significance of the effects.

The sampling to only include Finnish members of LP in the analyses, limits the results as only one market area was used in the analysis. To get more generalizable results the same analyses could be done on all the LP members data or separately on the different market areas.

5.2. Future research

Many future research ideas came up during this master's thesis project. Firstly, a more longitudinal approach could be added to the analyses as the program has been in place for more than 10 years. This could be done by utilizing data from multiple years instead of just one year's data. This would give a better view on: the long-term effects of the LP, the LP members true CLV and how the predictive models would perform with more data or a longer prediction period.

Secondly, more sophisticated models for the CLV predictions could be researched and developed (for example deep neural network models (Google. 2019 Predicting Customer Lifetime Value with AI Platform), where additional features can be incorporated) and compared to the models, utilized in this thesis, which are based only on the RFM features. These features could include specified customers segment, reactions to E-mail marketing, previously bought product categories and amount of returns. It would be interesting to see how much better predictive performance would be possible with more complex models and further feature engineering.

Third research idea comes from a limitation of this thesis: lack of comparable non-LP member data. Research between LP members versus non-members or on LP members before and after signing up to the loyalty program could provide more solid and generalizable results on the effects and profitability of the LP.

Fourth research idea that came up, when writing this thesis, is looking at the possibilities of gamification of the loyalty program to enhance effects and profitability of the LP and how it could be best implemented for a specific loyalty program to suit the profitable LP members and add value to both the customers and the company.

References

Allaway A.W., D’Souza G., Berkowitz D., Northington W.M. (2017) Trajectory-Based Segmentation of Loyalty Program-Type Data. In: Campbell C.L. (eds) The Customer is NOT Always Right? Marketing Orientationsin a Dynamic Business World. Developments in Marketing Science: Proceedings of the Academy of Marketing Science. Springer, Cham

Gupta, S. & Zeithaml, V. 2006, "Customer Metrics and Their Impact on Financial Performance", *Marketing Science*, vol. 25, no. 6, pp. 718-739.

Pauwels, K. 2015, "Truly Accountable Marketing: The Right Metrics for the Right Results", *GfK Marketing Intelligence Review*, vol. 7, no. 1, pp. 8-15.

Uncles, M.D., Dowling, G.R. & Hammond, K. 2003, "Customer loyalty and customer loyalty programs", *The Journal of Consumer Marketing*, vol. 20, no. 4, pp. 294-314.

Ashley, C., Gillespie, E.A. & Noble, S.M. 2016, "The effect of loyalty program fees on program perceptions and engagement", *Journal of Business Research*, vol. 69, no. 2, pp. 964-973.

Blattberg, R.C., Malthouse, E.C. & Neslin, S.A. 2009, "Customer Lifetime Value: Empirical Generalizations and Some Conceptual Questions", *Journal of Interactive Marketing*, vol. 23, no. 2, pp. 157-168.

Bradlow, E.T., Gangwar, M., Kopalle, P. & Voleti, S. 2017, "The Role of Big Data and Predictive Analytics in Retailing", *Journal of Retailing*, vol. 93, no. 1, pp. 79-95.

Bridson, K., Evans, J. & Hickman, M. 2008, "Assessing the relationship between loyalty program attributes, store satisfaction and store loyalty", *Journal of Retailing and Consumer Services*, vol. 15, no. 5, pp. 364-374.

Chang, H.H., Wong, K.H. & Fang, P.W. 2014, "The effects of customer relationship management relational information processes on customer-based performance", *Decision Support Systems*, vol. 66, pp. 146-159.

Fader, P.S., Hardie, B.G.S. & Lee, K.L. 2005, “A Note on Implementing the Pareto/NBD Model in MATLAB”, http://brucehardie.com/notes/008/

Fader, P.S. & Hardie, B.G.S. 2013, “The Gamma-Gamma Model of Monetary Value“, http://brucehardie.com/notes/025/

Gandomi, A. & Zolfaghari, S. 2011, "A stochastic model on the profitability of loyalty programs", *Computers & Industrial Engineering*, vol. Vol.61(3), pp. 482-488.

Glady, N., Baesens, B. & Croux, C. 2009, “A modified Pareto/NBD approach for predicting customer lifetime value”, Expert Systems with Applications, Volume 36, Issue 2, Part 1, pp. 2062-2071

Kang, J., Alejandro, T.B. & Groza, M.D. 2015, "Customer–company identification and the effectiveness of loyalty programs", *Journal of Business Research,* vol. 68, no. 2, pp. 464-471.

Kreis, H. & Mafael, A. 2014, "The influence of customer loyalty program design on the relationship between customer motives and value perception", *Journal of Retailing and Consumer Services,* vol. 21, no. 4, pp. 590-600.

Kumar, V., Ramani, G. & Bohling, T. 2004, "Customer lifetime value approaches and best practice applications", *Journal of Interactive Marketing,* vol. 18, no. 3, pp. 60-72.

Kumar, V. & Reinartz, W. 2012, *Customer Relationship Management : Concept, Strategy, and Tools,* 2nd ed. 2012. edn, Springer Berlin Heidelberg : Imprint: Springer, Berlin, Heidelberg.

Kumar, V. & Shah, D. 2004, "Building and sustaining profitable customer loyalty for the 21st century", *Journal of Retailing,* vol. 80, no. 4, pp. 317-329.

Kumar, V., Shah, D. & Venkatesan, R. 2006, "Managing retailer profitability—one customer at a time!", *Journal of Retailing,* vol. 82, no. 4, pp. 277-294.

Kuokka, K. 2014, “APPLYING CUSTOMER VALUE MEASUREMENT TO A CUSTOMER LOYALTY PROGRAM: Case of implementing a key performance indicator based on customer value.”, Master’s Thesis, Lappeenranta University of Technology, Accounting

Liu, Y. 2007, "The Long-Term Impact of Loyalty Programs on Consumer Purchase Behavior and Loyalty", *Journal of Marketing,* vol. 71, no. 4, pp. 19-35.

Peltier, J.W., Zahay, D. & Lehmann, D.R. 2013, "Organizational Learning and CRM Success: A Model for Linking Organizational Practices, Customer Data Quality, and Performance", *Journal of Interactive Marketing,* vol. 27, no. 1, pp. 1-13.

Petersen, J.A., McAlister, L., Reibstein, D.J., Winer, R.S., Kumar, V. & Atkinson, G. 2009, "Choosing the Right Metrics to Maximize Profitability and Shareholder Value", *Journal of Retailing,* vol. 85, no. 1, pp. 95-111.

Rackley, J. 2015, *Marketing Analytics Roadmap,* Apress.

So, J.T., Danaher, T. & Gupta, S. 2015, "What do customers get and give in return for loyalty program membership?", *Australasian Marketing Journal (AMJ),* vol. 23, no. 3, pp. 196-206.

Szczepańska, K. & Gawron, P. 2012, "Loyalty Programs Effectiveness", vol. Foundations of Management, no. 3(2), pp. 89.

Thakur, R. & Workman, L. 2016, "Customer portfolio management (CPM) for improved customer relationship management (CRM): Are your customers platinum, gold, silver, or bronze?", *Journal of Business Research*, vol. 69, no. 10, pp. 4095-4102.

Verhoef, P.C., Venkatesan, R., McAlister, L., Malthouse, E.C., Krafft, M. & Ganesan, S. 2010, "CRM in Data-Rich Multichannel Retailing Environments: A Review and Future Research Directions", *Journal of Interactive Marketing*, vol. 24, no. 2, pp. 121-137.

Online documents:

Cloud, Google. (2019) Predicting Customer Lifetime Value with AI Platform.

[Online document] [Referred to 20.11.2019]
Available: https://cloud.google.com/solutions/machine-learning/clv-prediction-with-offline-training-train

DataCamp. (2018) Customer Lifetime Value tutorial, learn how to calculate Customer Lifetime Value in Python. [Online document] [Referred to 20.11.2019]
Available: https://www.datacamp.com/community/tutorials/customer-life-time-value

Appendix

Appendix 1. VoucherStructureAnalysis.m

```
clc
clear all
close all

%Load voucher structure
Voucherstruct=readtable('VoucherStructure.xlsx')

Voucherstruct =

  13×3 table

    SixMonthPurchasesAtLeast    Voucher    Discount
    ________________________    _______    ________

              0.00               0.00       0.00
            149.99               0.00       0.00
            150.00              10.00       0.07
            249.99              10.00       0.04
            250.00              20.00       0.08
            499.99              20.00       0.04
            500.00              50.00       0.10
            999.99              50.00       0.05
           1000.00             100.00       0.10
           1500.00             100.00       0.07
           2000.00             100.00       0.05
           5000.00             100.00       0.02
          10000.00             100.00       0.01

%plot voucher structure large scale
figure(1)
yyaxis left
plot(Voucherstruct.SixMonthPurchasesAtLeast,Voucherstruct.Discount,'-o')
axis([0 2500 0 0.11])
yyaxis right
plot(Voucherstruct.SixMonthPurchasesAtLeast,Voucherstruct.Voucher)
```

```
axis([0 2500 0 105])
title('Voucher reward structure analysis(0-2500€ purchases)')
xlabel('Total amount of purchases(€)')
legend('Discount%','Voucher Amount','Location','southeast')

%plot only to 1000
figure(2)
yyaxis left
plot(Voucherstruct.SixMonthPurchasesAtLeast,Voucherstruct.Discount,'-o')
axis([0 1100 0 0.11])
ylabel('Discount %')
yyaxis right
plot(Voucherstruct.SixMonthPurchasesAtLeast,Voucherstruct.Voucher)
axis([0 1100 0 105])
ylabel('Voucher amount(€)')
title('Voucher reward structure analysis with 0-1000€ total purchases')
xlabel('Total amount of purchases(€)')
legend('Discount%','Voucher Amount(€)','Location','southeast')
```

Appendix 2. Code for currency transformations, filtering on Ids and channels, descriptive analyses on purchases and visualsations Filter_Descript_Code.m

```
clc
clear all
close all

load 'PurchDataFin'

load 'VoucherDataEUR'

PurchData=PurchDataFin;
length(unique(PurchData.IntegrationId))
VoucherData=VocherDataEUR;

%% Join Datatables together by ReceiptId
JoinedDataAllOrig=outerjoin(PurchData,VoucherData,'Type','left');
JoinedDataVouchersOnlyOrig=innerjoin(PurchData,VoucherData);

%% Fix Haaparanda purchase amounts and voucher amount used from SEK to EUR with
ConvRateEURtoSEK=9.491500
ConvRateEURtoSEK=9.491500
ConvRateSEKtoEUR=1/ConvRateEURtoSEK

JoinedDataAllOrig_H = JoinedDataAllOrig((JoinedDataAllOrig.ChannelName=='OUTLET HAPARANDA'...
    & JoinedDataAllOrig.CurrencyIsoCode~='SEK'),:);
JoinedDataAllOrig_H.PurchaseValue = JoinedDataAllOrig_H.PurchaseValue*ConvRateSEKtoEUR;
JoinedDataAllOrig_H.VoucherAmountUsed = JoinedDataAllOrig_H.VoucherAmountUsed*ConvRateSEKtoEUR;
JoinedDataAllOrig_Oth = JoinedDataAllOrig((JoinedDataAllOrig.ChannelName~='OUTLET HAPARANDA'),:);

JoinedDataAll = [JoinedDataAllOrig_Oth;JoinedDataAllOrig_H];

JoinedDataVouchersOnlyOrig_H =
JoinedDataVouchersOnlyOrig((JoinedDataVouchersOnlyOrig.ChannelName==...
    'OUTLET HAPARANDA' & JoinedDataVouchersOnlyOrig.CurrencyIsoCode~='SEK'),:);
JoinedDataVouchersOnlyOrig_H.PurchaseValue =
JoinedDataVouchersOnlyOrig_H.PurchaseValue*ConvRateSEKtoEUR;
JoinedDataVouchersOnlyOrig_H.VoucherAmountUsed =
JoinedDataVouchersOnlyOrig_H.VoucherAmountUsed*ConvRateSEKtoEUR;
JoinedDataVouchersOnlyOrig_Oth =
JoinedDataVouchersOnlyOrig((JoinedDataVouchersOnlyOrig.ChannelName~=...
    'OUTLET HAPARANDA'),:);

JoinedDataVouchersOnly = [JoinedDataVouchersOnlyOrig_Oth;JoinedDataVouchersOnlyOrig_H];

%% Add days between voucher sign date to usage and expiery date - usage
```

```matlab
JoinedDataAll.DaysBetweenVoucherStartAndUsageDate =
days(JoinedDataAll.PurchaseDate_VoucherData...
    - JoinedDataAll.VoucherPeriodStartDate);
JoinedDataAll.DaysBetweenVoucherExpieryAndUsageDate = days(JoinedDataAll.VoucherPeriodEndDate...
    - JoinedDataAll.PurchaseDate_VoucherData);

JoinedDataVouchersOnly.DaysBetweenVoucherStartAndUsageDate =
days(JoinedDataVouchersOnly.PurchaseDate...
    - JoinedDataVouchersOnly.VoucherPeriodStartDate);
JoinedDataVouchersOnly.DaysBetweenVoucherExpieryAndUsageDate =
days(JoinedDataVouchersOnly.VoucherPeriodEndDate...
    - JoinedDataVouchersOnly.PurchaseDate);

%% Purchases and voucher usage by ChannelName
Channel_PurchValue = JoinedDataAll(:,{'ChannelName','PurchaseValue','VoucherAmountUsed'});
Channel_PurchValue.VoucherAmountUsed(isnan(Channel_PurchValue.VoucherAmountUsed)) = 0;
ChannelgroupSums = varfun(@sum,Channel_PurchValue,'GroupingVariables','ChannelName');
ChannelgroupSumsSort = sortrows(ChannelgroupSums,{'sum_PurchaseValue'},'descend')
ChannelsWithMoreThan200Purch = ChannelgroupSumsSort.ChannelName(ChannelgroupSumsSort.GroupCount >
200.00)
%writetable(ChannelgroupSumsSort,'FinnPurchVouchByChannel.xlsx')

% Filter outlands and random channels out from the data
JoinedDataAllFinChan =
JoinedDataAll(ismember(JoinedDataAll.ChannelName,ChannelsWithMoreThan200Purch),:);
JoinedDataAllFinChan.VoucherAmountUsed(isnan(JoinedDataAllFinChan.VoucherAmountUsed)) = 0;
JoinedDataVouchersOnlyFinChan =
JoinedDataVouchersOnly(ismember(JoinedDataVouchersOnly.ChannelName,...
    ChannelsWithMoreThan200Purch),:);

% Pick finnish LP members IDs who did purchases in finnish channels,
% webstore and Haaparanta
IDsForAnalysis = unique(JoinedDataAllFinChan.IntegrationId_PurchData);

%% How the different sized vouchers are spent? Does the voucher amount spent affect the money
spent over it
JoinedDataVouchersOnlyFinChan.MoneySpentOverVouchAmountUsed =...
    (JoinedDataVouchersOnlyFinChan.PurchaseValue-
JoinedDataVouchersOnlyFinChan.VoucherAmountUsed);
Voucher10data = JoinedDataVouchersOnlyFinChan((JoinedDataVouchersOnlyFinChan.Value==10),:);
Voucher20data = JoinedDataVouchersOnlyFinChan((JoinedDataVouchersOnlyFinChan.Value==20),:);
Voucher50data = JoinedDataVouchersOnlyFinChan((JoinedDataVouchersOnlyFinChan.Value==50),:);
Voucher100data = JoinedDataVouchersOnlyFinChan((JoinedDataVouchersOnlyFinChan.Value==100),:);

%% How purchases without voucher usages compare?
FinPurchasesWithoutVouchers =
JoinedDataAllFinChan(not(ismember(JoinedDataAllFinChan.ReceiptId_PurchData...
    ,JoinedDataVouchersOnlyFinChan.ReceiptId)),:);
sum(FinPurchasesWithoutVouchers.PurchaseValue)
length(FinPurchasesWithoutVouchers.PurchaseValue)
mean(FinPurchasesWithoutVouchers.PurchaseValue)
median(FinPurchasesWithoutVouchers.PurchaseValue)
std(FinPurchasesWithoutVouchers.PurchaseValue)

% Result table
ResultRows =
{'NoVoucherPurchases';'AllVoucherPurchases';'Vouch100Purchases';'Vouch50Purchases';...
    'Vouch20Purchases';'Vouch10Purchases'}
PurchValueSumTotal = [sum(FinPurchasesWithoutVouchers.PurchaseValue);
    sum(JoinedDataVouchersOnlyFinChan.PurchaseValue);
    sum(Voucher100data.PurchaseValue);
    sum(Voucher50data.PurchaseValue);
    sum(Voucher20data.PurchaseValue);
    sum(Voucher10data.PurchaseValue)]

PurchValueCount = [length(FinPurchasesWithoutVouchers.PurchaseValue);
    length(JoinedDataVouchersOnlyFinChan.PurchaseValue);
    length(Voucher100data.PurchaseValue);
    length(Voucher50data.PurchaseValue);
```

```
    length(Voucher20data.PurchaseValue);
    length(Voucher10data.PurchaseValue)]

PurchValueMean = [mean(FinPurchasesWithoutVouchers.PurchaseValue);
    mean(JoinedDataVouchersOnlyFinChan.PurchaseValue);
    mean(Voucher100data.PurchaseValue);
    mean(Voucher50data.PurchaseValue);
    mean(Voucher20data.PurchaseValue);
    mean(Voucher10data.PurchaseValue)]

PurchValueMedian = [median(FinPurchasesWithoutVouchers.PurchaseValue);
    median(JoinedDataVouchersOnlyFinChan.PurchaseValue);
    median(Voucher100data.PurchaseValue);
    median(Voucher50data.PurchaseValue);
    median(Voucher20data.PurchaseValue);
    median(Voucher10data.PurchaseValue)]

PurchValueStdeviation = [std(FinPurchasesWithoutVouchers.PurchaseValue);
    std(JoinedDataVouchersOnlyFinChan.PurchaseValue);
    std(Voucher100data.PurchaseValue);
    std(Voucher50data.PurchaseValue);
    std(Voucher20data.PurchaseValue);
    std(Voucher10data.PurchaseValue)]

VouchAmountUsedSumTotal = [sum(FinPurchasesWithoutVouchers.VoucherAmountUsed);
    sum(JoinedDataVouchersOnlyFinChan.VoucherAmountUsed);
    sum(Voucher100data.VoucherAmountUsed);
    sum(Voucher50data.VoucherAmountUsed);
    sum(Voucher20data.VoucherAmountUsed);
    sum(Voucher10data.VoucherAmountUsed)]

VouchAmountUsedCount = [length(FinPurchasesWithoutVouchers.VoucherAmountUsed);
    length(JoinedDataVouchersOnlyFinChan.VoucherAmountUsed);
    length(Voucher100data.VoucherAmountUsed);
    length(Voucher50data.VoucherAmountUsed);
    length(Voucher20data.VoucherAmountUsed);
    length(Voucher10data.VoucherAmountUsed)]

VouchAmountUsedMean = [mean(FinPurchasesWithoutVouchers.VoucherAmountUsed);
    mean(JoinedDataVouchersOnlyFinChan.VoucherAmountUsed);
    mean(Voucher100data.VoucherAmountUsed);
    mean(Voucher50data.VoucherAmountUsed);
    mean(Voucher20data.VoucherAmountUsed);
    mean(Voucher10data.VoucherAmountUsed)]

VouchAmountUsedMedian = [median(FinPurchasesWithoutVouchers.VoucherAmountUsed);
    median(JoinedDataVouchersOnlyFinChan.VoucherAmountUsed);
    median(Voucher100data.VoucherAmountUsed);
    median(Voucher50data.VoucherAmountUsed);
    median(Voucher20data.VoucherAmountUsed);
    median(Voucher10data.VoucherAmountUsed)]

VouchAmountUsedStdeviation = [std(FinPurchasesWithoutVouchers.VoucherAmountUsed);
    std(JoinedDataVouchersOnlyFinChan.VoucherAmountUsed);
    std(Voucher100data.VoucherAmountUsed);
    std(Voucher50data.VoucherAmountUsed);
    std(Voucher20data.VoucherAmountUsed);
    std(Voucher10data.VoucherAmountUsed)]

MoneySpentOverVouchAmountUsedSumTotal = [NaN;
    sum(JoinedDataVouchersOnlyFinChan.MoneySpentOverVouchAmountUsed);
    sum(Voucher100data.MoneySpentOverVouchAmountUsed);
    sum(Voucher50data.MoneySpentOverVouchAmountUsed);
    sum(Voucher20data.MoneySpentOverVouchAmountUsed);
    sum(Voucher10data.MoneySpentOverVouchAmountUsed)]

MoneySpentOverVouchAmountUsedCount = [NaN;
    length(JoinedDataVouchersOnlyFinChan.MoneySpentOverVouchAmountUsed);
    length(Voucher100data.MoneySpentOverVouchAmountUsed);
```

```
    length(Voucher50data.MoneySpentOverVouchAmountUsed);
    length(Voucher20data.MoneySpentOverVouchAmountUsed);
    length(Voucher10data.MoneySpentOverVouchAmountUsed)]

MoneySpentOverVouchAmountUsedMean = [NaN;
    mean(JoinedDataVouchersOnlyFinChan.MoneySpentOverVouchAmountUsed);
    mean(Voucher100data.MoneySpentOverVouchAmountUsed);
    mean(Voucher50data.MoneySpentOverVouchAmountUsed);
    mean(Voucher20data.MoneySpentOverVouchAmountUsed);
    mean(Voucher10data.MoneySpentOverVouchAmountUsed)]

MoneySpentOverVouchAmountUsedMedian = [NaN;
    median(JoinedDataVouchersOnlyFinChan.MoneySpentOverVouchAmountUsed);
    median(Voucher100data.MoneySpentOverVouchAmountUsed);
    median(Voucher50data.MoneySpentOverVouchAmountUsed);
    median(Voucher20data.MoneySpentOverVouchAmountUsed);
    median(Voucher10data.MoneySpentOverVouchAmountUsed)]

MoneySpentOverVouchAmountUsedStdeviation = [NaN;
    std(JoinedDataVouchersOnlyFinChan.MoneySpentOverVouchAmountUsed);
    std(Voucher100data.MoneySpentOverVouchAmountUsed);
    std(Voucher50data.MoneySpentOverVouchAmountUsed);
    std(Voucher20data.MoneySpentOverVouchAmountUsed);
    std(Voucher10data.MoneySpentOverVouchAmountUsed)]

PercentageOfOnlyVouchUsed = [NaN;
    (sum(JoinedDataVouchersOnlyFinChan.MoneySpentOverVouchAmountUsed <= 0)...
    /length(JoinedDataVouchersOnlyFinChan.MoneySpentOverVouchAmountUsed)*100);
    (sum(Voucher100data.MoneySpentOverVouchAmountUsed <= 0)...
    /length(Voucher100data.MoneySpentOverVouchAmountUsed)*100);
    (sum(Voucher50data.MoneySpentOverVouchAmountUsed <= 0)...
    /length(Voucher50data.MoneySpentOverVouchAmountUsed)*100);
    (sum(Voucher20data.MoneySpentOverVouchAmountUsed <= 0)...
    /length(Voucher20data.MoneySpentOverVouchAmountUsed)*100);
    (sum(Voucher10data.MoneySpentOverVouchAmountUsed <= 0)...
    /length(Voucher10data.MoneySpentOverVouchAmountUsed)*100);]

Resmatrix = table(PurchValueSumTotal,PurchValueCount,PurchValueMean,...
    PurchValueMedian,PurchValueStdeviation,VouchAmountUsedSumTotal,...
    VouchAmountUsedCount,VouchAmountUsedMean,VouchAmountUsedMedian,...
    VouchAmountUsedStdeviation,MoneySpentOverVouchAmountUsedSumTotal,...
    MoneySpentOverVouchAmountUsedCount,MoneySpentOverVouchAmountUsedMean,...
    MoneySpentOverVouchAmountUsedMedian,MoneySpentOverVouchAmountUsedStdeviation,...
    PercentageOfOnlyVouchUsed,'RowNames',ResultRows)
Resmatrix.VouchAmountsVSMoneySpentOver = Resmatrix.MoneySpentOverVouchAmountUsedSumTotal...
    ./ Resmatrix.VouchAmountUsedSumTotal
rows2vars(Resmatrix)
%writetable(rows2vars(Resmatrix),'resmatrix_table_all.xlsx');

%% Voucher usage days before expiery only
ValueDaysBeforeExp =
JoinedDataVouchersOnlyFinChan(:,{'Value','DaysBetweenVoucherExpieryAndUsageDate'});

UsageOfVouchBasedOnExpiery = varfun(@sum, JoinedDataVouchersOnlyFinChan...
    (:,{'Value','DaysBetweenVoucherExpieryAndUsageDate'}),'GroupingVariables',...
    {'Value','DaysBetweenVoucherExpieryAndUsageDate'});

Vouch10use = UsageOfVouchBasedOnExpiery((UsageOfVouchBasedOnExpiery.Value == 10),:);
Vouch20use = UsageOfVouchBasedOnExpiery((UsageOfVouchBasedOnExpiery.Value == 20),:);
Vouch50use = UsageOfVouchBasedOnExpiery((UsageOfVouchBasedOnExpiery.Value == 50),:);
Vouch100use = UsageOfVouchBasedOnExpiery((UsageOfVouchBasedOnExpiery.Value == 100),:);

figure(1)
subplot(4,1,1)
bar(Vouch10use.DaysBetweenVoucherExpieryAndUsageDate, Vouch10use.GroupCount)
title('10€ Voucher usage before expiery')
ylabel('Amount')
xlabel('Days till expiery')
subplot(4,1,2)
```

```
bar(Vouch20use.DaysBetweenVoucherExpieryAndUsageDate, Vouch20use.GroupCount)
title('20€ Voucher usage before expiery')
ylabel('Amount')
xlabel('Days till expiery')
subplot(4,1,3)
bar(Vouch50use.DaysBetweenVoucherExpieryAndUsageDate, Vouch50use.GroupCount)
title('50€ Voucher usage before expiery')
ylabel('Amount')
xlabel('Days till expiery')
subplot(4,1,4)
bar(Vouch100use.DaysBetweenVoucherExpieryAndUsageDate, Vouch100use.GroupCount)
title('100€ Voucher usage before expiery')
ylabel('Amount')
xlabel('Days till expiery')

%%UsageDates
UsageOfVouchBasedOnExpiery = varfun(@sum,
JoinedDataVouchersOnlyFinChan(:,{'Value','PurchaseDate'})...
    ,'GroupingVariables',{'Value','PurchaseDate'});
Vouch10use = UsageOfVouchBasedOnExpiery((UsageOfVouchBasedOnExpiery.Value == 10),:);
Vouch20use = UsageOfVouchBasedOnExpiery((UsageOfVouchBasedOnExpiery.Value == 20),:);
Vouch50use = UsageOfVouchBasedOnExpiery((UsageOfVouchBasedOnExpiery.Value == 50),:);
Vouch100use = UsageOfVouchBasedOnExpiery((UsageOfVouchBasedOnExpiery.Value == 100),:);

figure(2)
subplot(4,1,1)
bar(Vouch10use.PurchaseDate, Vouch10use.GroupCount)
title('10€ voucher usage dates')
subplot(4,1,2)
bar(Vouch20use.PurchaseDate, Vouch20use.GroupCount)
title('20€ voucher usage dates')
subplot(4,1,3)
bar(Vouch50use.PurchaseDate, Vouch50use.GroupCount)
title('50€ voucher usage dates')
subplot(4,1,4)
bar(Vouch100use.PurchaseDate, Vouch100use.GroupCount)
title('100€ voucher usage dates')
```

Appendix 3. MultivariateLinRegCLVModelNewMembersCohort10FoldAnalysis.m

```
clc
clear all
close all

%Load Datasets
load 'FinnIdsPurchByMonth'
load 'IDsVouchValue2017'
load 'NewLPCustcohort'

IDsVouchValueIssuedFor2017.Properties.VariableNames{1} = 'IntegrationId_PurchData';

%Filter monthlty purchase data to only include customers of new to lp
%cohort
FinnIdsPurchByMonth = FinnIdsPurchByMonth(ismember...
    (FinnIdsPurchByMonth.IntegrationId_PurchData,NewLPCustcohort),:);
%Alter NaNs, order columns, calculate CLV (Total purch of timeperiod) for
%Ids
PurchByMonth = fillmissing(FinnIdsPurchByMonth,'constant',0,'DataVariables',...
    {'x1_2017','x2_2017','x3_2017','x4_2017','x5_2017','x6_2017','x7_2016','x8_2016'...
    ,'x9_2016','x10_2016','x11_2016','x12_2016'});
PurchByMonth = movevars(PurchByMonth,{'x7_2016','x8_2016','x9_2016','x10_2016',...
    'x11_2016','x12_2016'},'Before','x1_2017');

PurchByMonth.CLVJuly2016June2017 = sum(PurchByMonth{:,2:end},2);
```

```matlab
PurchByMonth = outerjoin(PurchByMonth,IDsVouchValueIssuedFor2017,'Type','Left');
PurchByMonth = fillmissing(PurchByMonth,'constant',0,'DataVariables','Value');

%How good of predictor is LinReg model for 1 Years total based on july-dec
%periods purchases?

%Independent variables
x = PurchByMonth{:,{'x7_2016','x8_2016','x9_2016','x10_2016','x11_2016','x12_2016','Value'}};

%Dependent variables
y = PurchByMonth{:,{'CLVJuly2016June2017'}};

%Split to train and test sets
seed = 3
rng(seed)

%Random 10-Fold cross
%Indices for 10-fold cross-validation
foldsize = round(max(size(y))/10)

indices = []
for i = 1:10
    parti = repmat(i,foldsize,1);
    indices = [indices;parti];
end
hist(indices)

Shufledataind = randperm(height(PurchByMonth))'

PurchByMonth.ShuffleIndex = Shufledataind;
PurchByMonth = sortrows(PurchByMonth,'ShuffleIndex');

%Independent variables
x = PurchByMonth{:,{'x7_2016','x8_2016','x9_2016','x10_2016','x11_2016','x12_2016','Value'}};

%Dependent variables
y = PurchByMonth{:,{'CLVJuly2016June2017'}};

Rmse10 = []
MAE10 = []

for ith = 1:10
    testset = (indices == ith);
    trainset = ~testset;
    x_train = x(trainset,:);
    x_test = x(testset,:);
    y_train = y(trainset,:);
    y_test = y(testset,:);

    %LinearModel All Ids as a whole
    mdlAll = fitlm(x_train, y_train);
    mdlAll;

    Ypredictions = predict(mdlAll,x_test);

    [r2 rmse] = rsquare(y_test,Ypredictions)
    Rmse10 = [Rmse10,rmse];
    %Mean absolute error
    err = y_test-Ypredictions;
    abserr = abs(err);
    meanabserr = mean(abserr);
    MAE10 = [MAE10,meanabserr];
end

mean(MAE10)
mean(Rmse10)
```

```
mdlAll
rmse
meanabserr
```

Appendix 4. Resulting model(10^{th}) from 10-fold linear regression model script with mae and rmse values

```
mdlAll =


Linear regression model:
    y ~ 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7

Estimated Coefficients:
                   Estimate       SE        tStat       pValue
                   ________    ________    ______    __________

    (Intercept)     9.2366      0.93094    9.9218    3.8908e-23
    x1              1.0779     0.013684    78.765             0
    x2              1.0571     0.014762     71.61             0
    x3              1.0343     0.016347    63.269             0
    x4              1.1592     0.017186    67.448             0
    x5              1.1279     0.015149    74.458             0
    x6              1.1303     0.013409    84.297             0
    x7              1.0086      0.15337    6.5765    4.9628e-11


Number of observations: 16398, Error degrees of freedom: 16390
Root Mean Squared Error: 79.6
R-squared: 0.79,  Adjusted R-Squared 0.79
F-statistic vs. constant model: 8.83e+03, p-value = 0

rmse =

   68.3641


meanabserr =

   31.8495
```

Appendix 5. ParetModelDataAndPredictionResultsCompared.m

```
clc
clear all
close all

%Load Datasets
load 'PurchDataFin'

load 'VoucherDataEUR'

%load 'VoucherDataEURALL'

load FinnIDsThatHaveVouch
```

```
PurchData=PurchDataFin;
length(unique(PurchData.IntegrationId))
VoucherData=VocherDataEUR;

%% Join Datatables together by ReceiptId
JoinedDataAllOrig=outerjoin(PurchData,VoucherData,'Type','left');
JoinedDataVouchersOnlyOrig=innerjoin(PurchData,VoucherData);
JoinedDataAllOrig.VoucherAmountUsed(isnan(JoinedDataAllOrig.VoucherAmountUsed)) = 0;

%% Fix Haaparanda purchase amounts and voucher amount used from SEK to EUR with
ConvRateEURtoSEK=9.491500
ConvRateEURtoSEK=9.491500
ConvRateSEKtoEUR=1/ConvRateEURtoSEK

JoinedDataAllOrig_H = JoinedDataAllOrig((JoinedDataAllOrig.ChannelName...
    =='OUTLET HAPARANDA' & JoinedDataAllOrig.CurrencyIsoCode~='SEK'),:);
JoinedDataAllOrig_H.PurchaseValue = JoinedDataAllOrig_H.PurchaseValue*ConvRateSEKtoEUR;
JoinedDataAllOrig_H.VoucherAmountUsed = JoinedDataAllOrig_H.VoucherAmountUsed*ConvRateSEKtoEUR;
JoinedDataAllOrig_Oth = JoinedDataAllOrig((JoinedDataAllOrig.ChannelName~='OUTLET HAPARANDA'),:);

JoinedDataAll = [JoinedDataAllOrig_Oth;JoinedDataAllOrig_H];

JoinedDataVouchersOnlyOrig_H =
JoinedDataVouchersOnlyOrig((JoinedDataVouchersOnlyOrig.ChannelName...
    =='OUTLET HAPARANDA' & JoinedDataVouchersOnlyOrig.CurrencyIsoCode~='SEK'),:);
JoinedDataVouchersOnlyOrig_H.PurchaseValue = JoinedDataVouchersOnlyOrig_H.PurchaseValue...
    *ConvRateSEKtoEUR;
JoinedDataVouchersOnlyOrig_H.VoucherAmountUsed =
JoinedDataVouchersOnlyOrig_H.VoucherAmountUsed...
    *ConvRateSEKtoEUR;
JoinedDataVouchersOnlyOrig_Oth =
JoinedDataVouchersOnlyOrig((JoinedDataVouchersOnlyOrig.ChannelName...
    ~='OUTLET HAPARANDA'),:);

JoinedDataVouchersOnly = [JoinedDataVouchersOnlyOrig_Oth;JoinedDataVouchersOnlyOrig_H];

%% Add Column where Voucher Usage is deducted from purch amounts
JoinedDataAll.PurchMinusVoucher = JoinedDataAll.PurchaseValue - JoinedDataAll.VoucherAmountUsed;
JoinedDataVouchersOnly.PurchMinusVoucher = JoinedDataVouchersOnly.PurchaseValue...
    - JoinedDataVouchersOnly.VoucherAmountUsed;
%% Purchases and voucher usage by ChannelName
Channel_PurchValue = JoinedDataAll(:,{'ChannelName','PurchaseValue','VoucherAmountUsed'});
%Channel_PurchValue.VoucherAmountUsed(isnan(Channel_PurchValue.VoucherAmountUsed)) = 0;
ChannelgroupSums = varfun(@sum,Channel_PurchValue,'GroupingVariables','ChannelName');
ChannelgroupSumsSort = sortrows(ChannelgroupSums,{'sum_PurchaseValue'},'descend')
ChannelsWithMoreThan200Purch = ChannelgroupSumsSort.ChannelName...
    (ChannelgroupSumsSort.GroupCount > 200.00)
%writetable(ChannelgroupSumsSort,'FinnPurchVouchByChannel.xlsx')
histogram(JoinedDataAll.PurchaseValue,500)
% Filter outlands and random channels out from the data
JoinedDataAllFinChan = JoinedDataAll(ismember(JoinedDataAll.ChannelName...
    ,ChannelsWithMoreThan200Purch),:);
%JoinedDataAllFinChan.VoucherAmountUsed(isnan(JoinedDataAllFinChan.VoucherAmountUsed)) = 0;
JoinedDataVouchersOnlyFinChan =
JoinedDataVouchersOnly(ismember(JoinedDataVouchersOnly.ChannelName...
    ,ChannelsWithMoreThan200Purch),:);

NewLPMem0407_25_12_coh=JoinedDataAll((JoinedDataAll.SignedDate > '04-Jul-2016'...
    & JoinedDataAll.SignedDate < '25-Dec-2016' & JoinedDataAll.PurchaseDate_PurchData...
    <= '25-Dec-2016'),{'IntegrationId_PurchData','SignedDate','PurchaseDate_PurchData',...
    'PurchaseValue','VoucherAmountUsed'});
NewLPCustcohort = unique(NewLPMem0407_25_12_coh.IntegrationId_PurchData)
save('NewLPCustcohort.mat','NewLPCustcohort')

% Map of Days 4th jul 2016 to end and week order number for every day
DaysWeekNumMap = readtable('DaysWeeksMapForPareto.xlsx')
summary(DaysWeekNumMap)
DaysWeekNumMap.Date = regexprep(DaysWeekNumMap.Date,"'","")
```

```
DaysWeekNumMap.Date=datetime(DaysWeekNumMap.Date,'InputFormat','dd-MMM-yyyy');
DaysWeekNumMap.Properties.VariableNames = {'PurchaseDate_PurchData','Week'};

%Test with only new members cohort!
JoinedDataAllFinChan =
JoinedDataAllFinChan(ismember(JoinedDataAllFinChan.IntegrationId_PurchData,...
    NewLPCustcohort),:);

JoinedDataAllWeeks = outerjoin(JoinedDataAllFinChan,DaysWeekNumMap,'Type','Left');
% Monetary values of the two periods purchases
FinnIdsPurchMoneyUsedByWeek = unstack(JoinedDataAllWeeks, 'PurchMinusVoucher',...
    'Week', 'GroupingVariables', 'IntegrationId_PurchData', 'AggregationFunction', @sum);
FinnIdsPurchMoneyUsedByWeek(:,2:end) =
fillmissing(FinnIdsPurchMoneyUsedByWeek(:,2:end),'constant',0);
FinnIdsPurchMoneyUsedByWeek.Monetary1_25 = sum(FinnIdsPurchMoneyUsedByWeek{:,2:26},2);
FinnIdsPurchMoneyUsedByWeek.Monetary26_52 = sum(FinnIdsPurchMoneyUsedByWeek{:,27:53},2);
FinnIdsPurchMoneyUsedByWeek.Monetary1_52 = sum(FinnIdsPurchMoneyUsedByWeek{:,2:53},2);
FinnIdsPurchMoneyUsedByWeek(FinnIdsPurchMoneyUsedByWeek.Monetary1_52==0,:)

% Monetary value of rep purchase: pick only first purchase value and deduct
% it from the Monetary1_25 value
JoinedDataAllWeeksFirstPurch = sortrows(JoinedDataAllWeeks(:,{'IntegrationId_PurchData',...
    'PurchaseDate_PurchData_JoinedDataAllFinChan','Week','PurchMinusVoucher'}),...

{'IntegrationId_PurchData','PurchaseDate_PurchData_JoinedDataAllFinChan'},{'ascend','ascend'});
[~,idx] = unique(JoinedDataAllWeeksFirstPurch(:,1));
JoinedDataAllWeeksFirstPurch = JoinedDataAllWeeksFirstPurch(idx,:);
JoinedDataAllWeeksFirstPurch = JoinedDataAllWeeksFirstPurch((JoinedDataAllWeeksFirstPurch.Week <=
25),:);

FinnIdsPurchMoneyUsedByWeek = outerjoin(FinnIdsPurchMoneyUsedByWeek,...

JoinedDataAllWeeksFirstPurch(:,{'IntegrationId_PurchData','PurchMinusVoucher'}),'Type','Left');
FinnIdsPurchMoneyUsedByWeek.PurchMinusVoucher =
fillmissing(FinnIdsPurchMoneyUsedByWeek.PurchMinusVoucher,...
    'constant',0);
FinnIdsPurchMoneyUsedByWeek.RepMonetary1_25 = FinnIdsPurchMoneyUsedByWeek.Monetary1_25...
    -FinnIdsPurchMoneyUsedByWeek.PurchMinusVoucher;

%Weekly counts!!!
JoinedDataAllWeeks.purchase = double(any(JoinedDataAllWeeks.PurchaseValue,2));
FinnIdsPurchCountByWeek = unstack(JoinedDataAllWeeks, 'purchase', 'Week',...
    'GroupingVariables', 'IntegrationId_PurchData', 'AggregationFunction', @sum);
FinnIdsPurchCountByWeek(:,2:end) = fillmissing(FinnIdsPurchCountByWeek(:,2:end),'constant',0);

%Time to do repeat purch (T) first purhc up to 25th dec 2016 in weeks 1 to
%25
JoinedDataAllFinChanBefore25Dec =
JoinedDataAllFinChan(JoinedDataAllFinChan.PurchaseDate_PurchData...
    <= datetime('25-Dec-2016')& JoinedDataAllFinChan.PurchaseDate_PurchData > datetime('03-Jul-
2016'),:);
FirstPurchase =
sortrows(JoinedDataAllFinChanBefore25Dec(:,{'IntegrationId_PurchData','PurchaseDate_PurchData'}).
..
    ,{'IntegrationId_PurchData','PurchaseDate_PurchData'},{'ascend','ascend'});
[~,idx] = unique(FirstPurchase(:,1));
FirstPurchaseOnly = FirstPurchase(idx,:);
FirstPurchaseOnly = FirstPurchaseOnly(FirstPurchaseOnly.PurchaseDate_PurchData <= datetime('25-
Dec-2016')...
    &FirstPurchaseOnly.PurchaseDate_PurchData > datetime('03-Jul-2016'),:)
FirstPurchaseOnly.DateOfInterest = repmat(datetime('25-Dec-2016'),height(FirstPurchaseOnly),1);
summary(FirstPurchaseOnly)
FirstPurchaseOnly.TimeForRepeat =  days(FirstPurchaseOnly.DateOfInterest...
    - FirstPurchaseOnly.PurchaseDate_PurchData)/7;

IdsThatPurchInFirstPeriod = FirstPurchaseOnly.IntegrationId_PurchData

%Repeat purchases in 1 to 25 weeks and weeks 26 to 52 (p1x,p2x)
```

```
FinnIdsPurchCountByWeek1to25 =
FinnIdsPurchCountByWeek(ismember(FinnIdsPurchCountByWeek.IntegrationId_PurchData...
    ,IdsThatPurchInFirstPeriod),:);
FinnIdsPurchCountByWeek1to25.RepPurch1_25 = sum(FinnIdsPurchCountByWeek1to25{:,2:26},2);
FinnIdsPurchCountByWeek1to25.ActualRepPurch1_25 = FinnIdsPurchCountByWeek1to25.RepPurch1_25-1;
FinnIdsPurchCountByWeek1to25.RepPurch26_52 = sum(FinnIdsPurchCountByWeek1to25{:,27:53},2);
FinnIdsPurchCountByWeek1to25.RepPurchTotal = sum(FinnIdsPurchCountByWeek1to25{:,2:53},2);

CheckLogic = FinnIdsPurchCountByWeek1to25.RepPurchTotal ==
(FinnIdsPurchCountByWeek1to25.RepPurch1_25...
    +FinnIdsPurchCountByWeek1to25.RepPurch26_52);

%Time of last repeat purchase in weeks 1 to 25
JoinedDataAllFinChanBefore25Dec =
JoinedDataAllFinChan(JoinedDataAllFinChan.PurchaseDate_PurchData...
    <= datetime('25-Dec-2016')& JoinedDataAllFinChan.PurchaseDate_PurchData > datetime('03-Jul-
2016'),:);
LastPurchase = sortrows(JoinedDataAllFinChanBefore25Dec(:,{'IntegrationId_PurchData',...

'PurchaseDate_PurchData'}),{'IntegrationId_PurchData','PurchaseDate_PurchData'},{'ascend','descen
d'});
[~,idx] = unique(LastPurchase(:,1));
LastPurchaseOnly = LastPurchase(idx,:);
LastPurchaseOnly.DateOfInterest = repmat(datetime('25-Dec-2016'),height(LastPurchaseOnly),1);
summary(LastPurchaseOnly)
LastPurchaseOnly.TimeOfLastRepeat =  days(LastPurchaseOnly.DateOfInterest...
    - LastPurchaseOnly.PurchaseDate_PurchData)/7;

%Create table with ID,p1x(Repeat purch in first period),t_x(time of last
%repeat purch),T(time of first purch),p2x(purchases in second period)
ParetoModelDataSet = FirstPurchaseOnly(:,{'IntegrationId_PurchData','TimeForRepeat'});
ParetoModelDataSet = outerjoin(ParetoModelDataSet,LastPurchaseOnly(:,...
    {'IntegrationId_PurchData','TimeOfLastRepeat'}),'Type','Left');
ParetoModelDataSet = ParetoModelDataSet(:,{'IntegrationId_PurchData_ParetoModelDataSet',...
    'TimeForRepeat','TimeOfLastRepeat'});
ParetoModelDataSet.Properties.VariableNames = {'IntegrationId_PurchData',...
    'TimeForRepeat','TimeOfLastRepeat'};
ParetoModelDataSet = outerjoin(ParetoModelDataSet,FinnIdsPurchCountByWeek1to25(:,...
    {'IntegrationId_PurchData','ActualRepPurch1_25','RepPurch26_52'}),'Type','Left');
ParetoModelDataSet = ParetoModelDataSet(:,{'IntegrationId_PurchData_ParetoModelDataSet',...
    'TimeForRepeat','TimeOfLastRepeat','ActualRepPurch1_25','RepPurch26_52'});
ParetoModelDataSet.Properties.VariableNames =
{'IntegrationId_PurchData_FinnIdsPurchMoneyUsedByWeek',...
    'TimeForRepeat','TimeOfLastRepeat','ActualRepPurch1_25','RepPurch26_52'};
ParetoModelDataSet = outerjoin(ParetoModelDataSet,FinnIdsPurchMoneyUsedByWeek(:,...
    {'IntegrationId_PurchData_FinnIdsPurchMoneyUsedByWeek','Monetary1_25','Monetary26_52',...
    'RepMonetary1_25'}),'Type','Left');
ParetoModelDataSet =
ParetoModelDataSet(:,{'IntegrationId_PurchData_FinnIdsPurchMoneyUsedByWeek_ParetoModel'...
    ,'TimeForRepeat','TimeOfLastRepeat','ActualRepPurch1_25','RepPurch26_52','Monetary1_25',...
    'RepMonetary1_25','Monetary26_52'});
ParetoModelDataSet.Properties.VariableNames = {'ID','T','t_x','p1x','p2x',...
    'M1x','RepM1x','M2x'};
ParetoModelDataSet.zbar = fillmissing((ParetoModelDataSet.RepM1x./...
    ParetoModelDataSet.p1x),'constant',0);
ParetoModelDataSet.zbar(isinf(ParetoModelDataSet.zbar)) = 0;
ParetoModelDataSet.t_x(ParetoModelDataSet.t_x==ParetoModelDataSet.T)=0;


%writetable(ParetoModelDataSet,'ParetoModelDataNewCustOnly.xlsx')

%ParetoModelDataSet.p1x(ParetoModelDataSet.p1x<0)=0;

%Summary stats on averag rep purchase
TempGamma = ParetoModelDataSet((ParetoModelDataSet.p1x > 0 & ParetoModelDataSet.zbar > 0),:);
summary(TempGamma);
mode(TempGamma.zbar)
std(TempGamma.zbar)
mean(TempGamma.zbar)
```

```
quantile(TempGamma.zbar,[0.25 0.50 0.75])

%Cumulative repeat purchases
%Keep customers that purch in first 25 weeks, remove first purch, unstack
%to weeks and sum by week cumsum
%OR join week nums to FirstPurchTable sum by week, deduct from total cumsum
FirstPurchByWeek1_25 = outerjoin(FirstPurchaseOnly,DaysWeekNumMap,'Type','Left');
Week1_25FirstPurch = varfun(@sum,FirstPurchByWeek1_25(:,{'Week'}),'GroupingVariables','Week');

sumbyweek = sum(FinnIdsPurchCountByWeek1to25{:,2:53},1)
RepurchsumbyWeek = sumbyweek-[Week1_25FirstPurch.GroupCount' zeros(1,27)]
RepurchsumbyWeek = RepurchsumbyWeek'
CumulativeRepurchByweek = cumsum(RepurchsumbyWeek,1);

%Sales forecast for every individual Before running this part matlab scripts
%load_data.m, estimate_pareto_nbd.m, create_tracking_plot.m,
%compute_pactive.m, compute_ce.m and functions h2f1.m and pareto_nbd_ll.m
%from A Note on Implementing the Pareto/NBD Model in MATLAB- paper (Fader
%et al. 2005) <http://brucehardie.com/notes/008/> modified to context need to be ran and kept in
%working space and parameters(p=2.1343155298357,q=3.15620393708097,gam=60,9915811586191) from the
%Gamma-Gamma Model of Monetary Value(Fader et al. 2013) <http://brucehardie.com/notes/025/>
%brought also to working space

endwk = 26;
endday = endwk*7;

tmp1 = r*beta/(alpha*(s-1));
tmpcumsls1 = [];
i = endday
tmp2 = (beta/(beta+i/7))^(s-1);
tmpcumsls1(i) = tmp1*(1-tmp2);


%Temp Pareto
TempParetoModelDataSet = ParetoModelDataSet;
TempParetoModelDataSet.ParetoEstp2x = repmat(tmpcumsls1(end),height(TempParetoModelDataSet),1);
sum(TempParetoModelDataSet.p2x)
sum(TempParetoModelDataSet.ParetoEstp2x)

%CLV calc and stats compared to real
%p=2.00724623554296;
%q=3.26491945192269;
%gam=65.0380836834493;

%For new only cohort
p=2.1343155298357;
q=3.15620393708097;
gam=60,9915811586191;

Ez = p*gam/(q-1)

TempParetoModelDataSet.weight = (q-1)./(p*TempParetoModelDataSet.p1x+q-1);
TempParetoModelDataSet.excpectavrpurch = TempParetoModelDataSet.weight*Ez+...
    (1-TempParetoModelDataSet.weight).*TempParetoModelDataSet.zbar;
TempParetoModelDataSet.ParetoGammaCLV = TempParetoModelDataSet.ParetoEstp2x.*...
    TempParetoModelDataSet.excpectavrpurch;

sum(TempParetoModelDataSet.ParetoGammaCLV)
sum(TempParetoModelDataSet.M2x)

[r2 rmse] = rsquare(TempParetoModelDataSet.M2x,TempParetoModelDataSet.ParetoGammaCLV)

%Mean absolute error
err = TempParetoModelDataSet.M2x-TempParetoModelDataSet.ParetoGammaCLV;
abserr = abs(err);
meanabserr = mean(abserr)
```