A MATHEMATICAL MODEL FOR CUSTOMER LIFETIME VALUE BASED OFFER MANAGEMENT

A Thesis

by

Zehra Can

Submitted to the
Graduate School of Sciences and Engineering
In Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in the
Department of Industrial Engineering

Özyeğin University
Jan 2018

A MATHEMATICAL MODEL FOR CUSTOMER LIFETIME VALUE BASED OFFER MANAGEMENT

Approved by:

Asst. Prof. Erinç Albey (Advisor)
Department of Industrial Engineering
Özyeğin University

Asst. Prof. Ihsan Yanıkoğlu
Department of Industrial Engineering
Özyeğin University

Assoc. Prof. Mehmet Güray Güler
Department of Industrial Engineering
Yıldız Technical University

Date Approved: 9 January 2018

To my parents, Hasan and Hanife Kendir, who have always loved me unconditionally and whose good examples have taught me to work hard for the things that I aspire to achieve.

This thesis work is also dedicated to my twins Esin Duru and Atahan Doruk and my husband, Selim, who has been a constant source of support and encouragement during the challenges of graduate school and life. I am truly thankful for having them in my life.

ABSTRACT

Customers with prepaid lines possess higher attrition risk compared to postpaid customers, since prepaid customers do not sign long-term obligatory contracts and may churn anytime. For this reason, mobile operators have to offer engaging benefits to keep prepaid subscribers with the company. Since all such offers incur additional cost, mobile operators face an optimization problem while selecting the most suitable offers for customers at risk. In this study, an offer management framework targeting prepaid customers of a telecommunication company is developed. Proposed framework choses the most suitable offer for each customer through a mathematical model, which utilizes customer lifetime value and churn risk. Lifetime values are estimated using logistic regression and Pareto/NBD models.

ÖZETÇE

Ön-ödemeli hat müşterileri uzun süreli taahütlü kontrat yapmadıkları ve herhangi bir zamanda operatörden ayrılabilecekleri için faturalı müşterilere göre daha yüksek gitme riskine sahiptirler. Bu sebepten mobil operatörler ön-ödemeli hat müşterilerini ellerinde tutmak için bağlayıcı teklifler sunmalıdır. Müşterilere sunulacak her bir teklifin maliyeti olacağından, mobil operatörler riskli müşterilerine en uygun teklifi seçme aşamasında optimizasyon problemi ile karşı karşıya kalmaktadırlar. Bu çalışmada, ön-ödemeli hat müşterilerine en uygun teklifin hesaplanabilmesi için özel bir teklif yönetim modeli geliştirilmiştir. Yaşam boyu değeri ve ayrılma riski bilgilerinden de faydalanarak önerilen yapı her bir müşteri için en uygun teklifi matematiksel model aracılığıyla seçer. Yaşamboyu değeri Pareto/NBD ve lojistik regresyon modelleri ile tahminlenmiştir.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my advisor, Dr. Erinç Albey for all his help and guidance that he has given me as I worked on my thesis. Secondly, I would also like to thank Ahmet Şahin for providing me support during this period.

Lastly, I would like to thank to my colleagues for encouraging me and inspiring me to always get things done better.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Subscriber Analysis In Mobile Industry

Under today' s challenging market conditions, competitions become more and more important for companies. The attention of a customer is disturbed by the competitors. Therefore the companies must be proactively analyze their customer behavior based on their CRM and behavioral data and offer the customer the best product or service to keep their attraction. Satisfaction with the product or service improves the loyalty of the customer with the brand. Customer loyalty encourages customer to spend more money with the company' s product and services, thus the revenue of the firm grows.

In mobile sector the churn rate of the customers are more dynamic than the other sectors. Especially predicting the behavior of the prepaid subscribers are more difficult than postpaid subscribers. Usually, it is accepted that prepaid subscribers individually generate less revenue than postpaid subscribers, as a result of those, operators mostly focus on the postpaid subscribers. However, in Turkey in last years, proportionally the volume of the prepaid subscribers' number converges to the postpaid subscribers' number so the prepaid revenue cannot be ignored. The trends of the postpaid and prepaid subscribers can be seen in Figure 1. This market data is published quarterly to report the market trends in mobile sector by BTK known as Bilgi Teknolojileri Kurumu which is the Governmental Organization of Information Technologies.

Telecommunication operators have to adapt to digital evolution of the customer services. Smart phones has changed the market conditions. Customers can experience

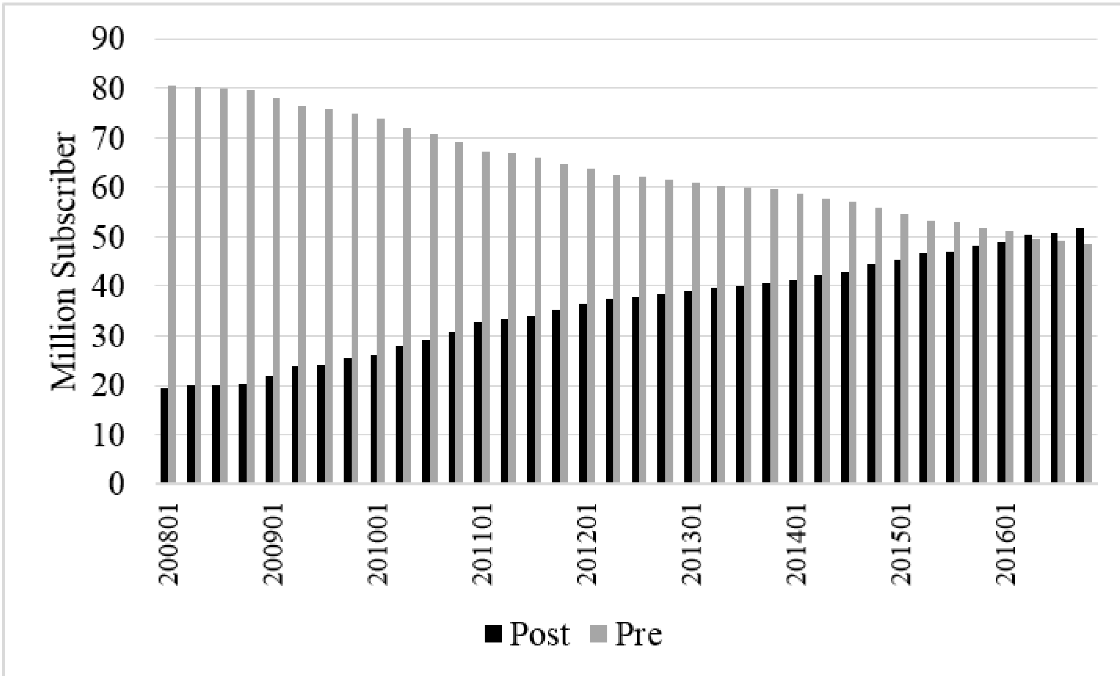


Figure 1: Postpaid Prepaid Subscriber Trends in Turkey Published By BTK

hundreds of digital services and applications, easily, with their smart phones. Therefore, telecommunication operators cannot be solely seen as communication providers anymore. Operators have to provide profitable services to survive in the business and to improve the customer experience to prevent customer churn.

The service and products of the operator should comply with the customer's needs. If this is not the case, it is highly probable that customers may leave and never come back or may never subscribe for a service of the operator. Customer experience management starts with potential customers perception about the company brand. Starting with first impression, customer experience management aims to understand the desires of customers and to provide them with easy, simple and seamless experience with the offered products and services. The higher the quality in customer experience, the more customer will be attracted to and stay with the company. Retaining customers has always been a costly challenge. Companies have developed different strategies to fight against customer attrition. Most of these efforts include proposing offers in varying forms, such as discounts, extra benefits etc. Almost in all of these efforts, increasing the amount of customer specific information pays off. In the digitalization era, the amount of customer specific data is more than ever, which yields numerous opportunities for developing granular and customized prediction models. When customer retention is considered, the challenge is turning this massive amount of data (and models built on the data) into business insights and eventually generate actions that decrease the retention cost, while increasing profit.

Churn analysis is one of the most fundamental customer behavior analysis in identifying silent, unhappy customers. The main output of churn analysis is a risk score, which quantifies the likelihood of customer churn. The output of churn analysis is best utilized when it is supported with extra customer specific features such as customer lifetime value (CLV). The relationship between churn and CLV is depicted in Figure 2. which is based on the idea in the [1]

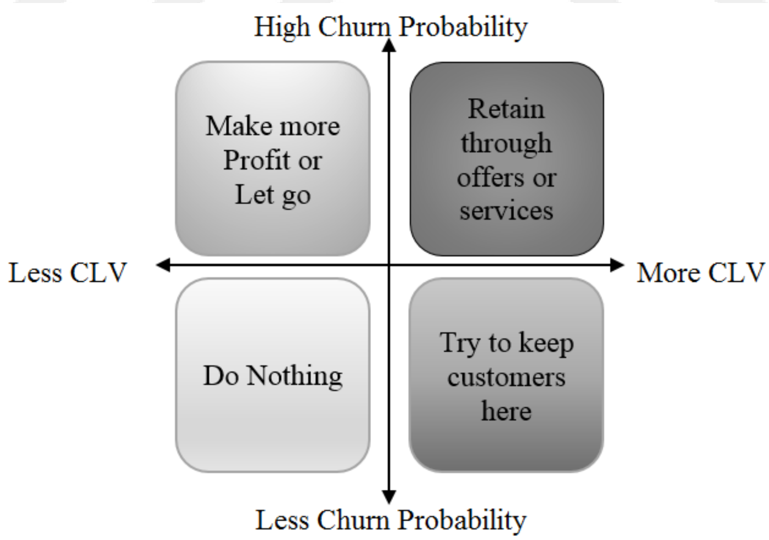


Figure 2: Churn and CLV Relationship

As can be seen in Figure 2, the ideal position of customers for a company is the bottom right, where customers produce the highest profit with lowest churn risk. It is recommended that companies should give special attention (provide offers or new exciting services) to the customers falling in the upper right region. Since, if retained in the company, these are the high profit promising customers. Bottom left and upper left regions are the regions, where less profitable customers reside. If possible, companies should try to increase profit made out of these customers. On the other hand, if the customer is not promising higher profit and possessing high churn risk, companies may simply let these customers go and try to acquire new customers, who fall into high CLV, low churn risk region.

Companies usually identify risky customers according to the lift of their favorite churn-prediction model. In a typical setting, they try to give the best offer to the customers possessing highest risk. In this scenario, managing the campaign is simple but its efficiency (in terms of cost) is questionable. Because, in such an execution, decision maker does not know whether offer receiving customers are worth the effort or not. On the other hand, integrating CLVs into the offer management process helps to increase return of the campaign by targeting, and possibly retaining, customers with higher CLV.

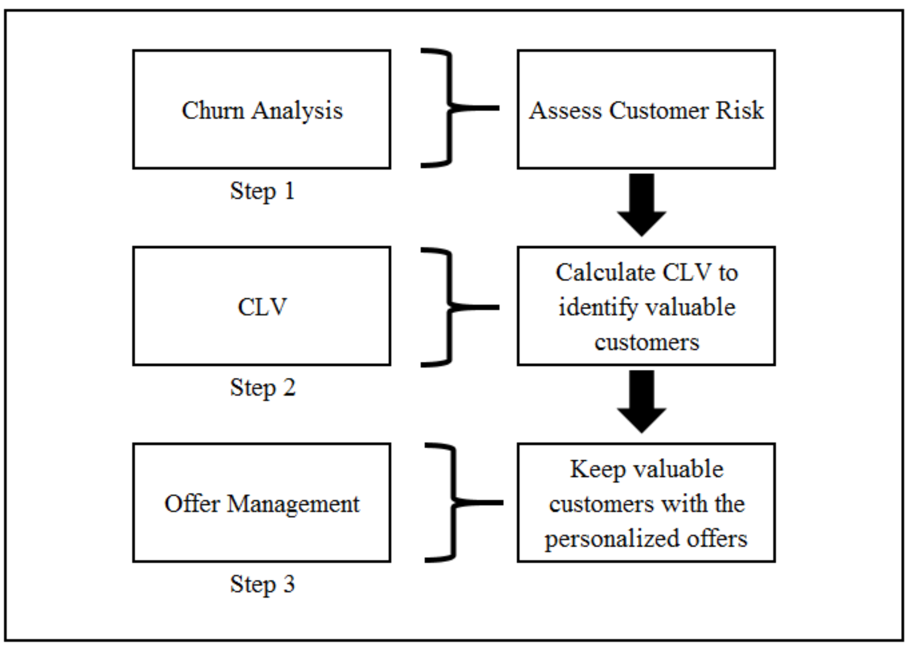


Figure 3: 3-Level Offer Management Process

Figure 3. shows our proposed 3-level offer management framework. The first step

calculates churn probability of customers, and the second step estimates CLVs. Without CLV calculation an operator churn management may cost higher than included with CLV calculation.

In study [2] the focus is on prepaid subscribers of a mobile operator. Mobile operators would prefer having postpaid subscribers more than prepaid subscribers. Because it is easier to manage the postpaid subscribers behavior and the revenue generation is more predictable than the prepaid subscribers. But in todays digital marketing ecosystem telecommunication operators have to act not only a call service provider (CSP), but also act like a full service provider (OTT Over the top provider) who offers many interactive social service products for their customers. Thus, keeping a customer is more valuable than before which means more users for the OTT services. Thus analyzing the churn of a customer must cover the behavioral effects of the people profile like their segment and tenure of the customer. Segment is the grouping of customers which gives indication about the behavior of the customer like his spending behavior or the age group of the customer. Beside segment customer tenure can also be used to classify customer behavior. Tenure can be defined as the age of the customer which gives the customer aliveness duration in time for the company.

The aim of the study is to analyze the churn behavior of the people according to their experience both with the operator and their social position. Therefore, we include both the segment and the tenure information of the customer in our data set.

1.2 Mobile Churn with Numbers in Turkey

Every year, millions of mobile subscribers churn from their mobile operator to the competitor operator. According to the 2017 2nd quarter report of the BTK (Information Technologies Organization), there is 76.6 million mobile subscribers in Turkey and mobile equipment penetration is 108% excluding the 0-9 ages population that can be seen in Figure 4 [3].

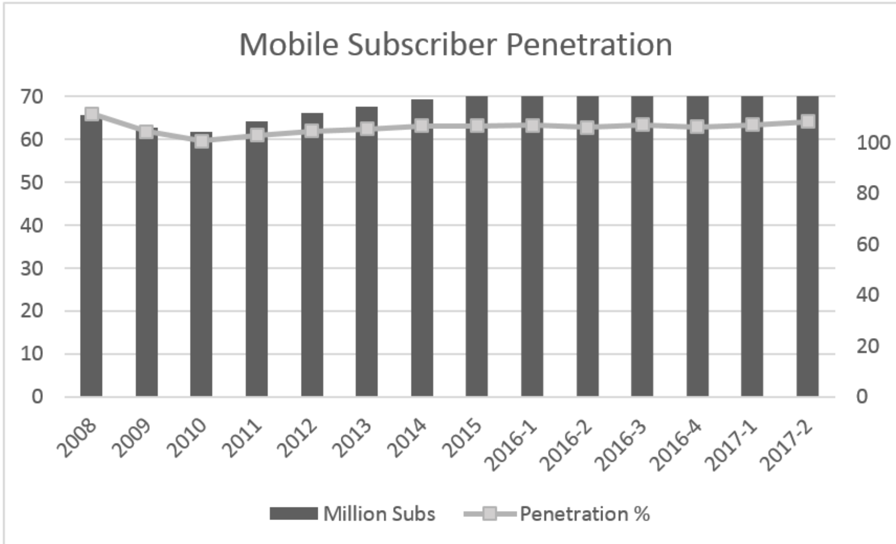


Figure 4: Mobile Subscriber Penetration in Turkey

The numbers in the same report also represented that there were 2.270.425 number portability transactions with 2.28% deccrease comparing to the previous 3 months period. The number of the transactions were 106.540.208 at total in the period from the beginning of the number portability in November of 2008 to May of 2015.

In the 2nd quarter of the 2017, there was an increase in the subscriber numbers of Vodafone, Avea and Turkcell by 0.36%, 0.27% and 2.27% respectively.

The ARPU amount for Turkcell, Vodafone and Avea is 28.5, 28.2 and 27.9 respectively. Simply, if the operators lose their subscribers, for example, 100.000 subscriber would change their operator in a month, the revenue of Turkcell, Vodafone and Avea would decrease 285.000, 282.000 and 279.000 in amount.

Moreover, SAC (Subscriber Acquisition Cost) is 35.5 TL according to the Turkcell 2015, 1st Quarter IFRS Interim Report. Keeping the existing customers, will prevent the mobile operators to lose revenue [4].

CHAPTER II

BACKGROUND

Prepaid subscribers have to first make top-up before making calls. Usage behavior of prepaid subscribers differs from each other. The credit is purchased at any time whenever the subscriber decides to. In Turkish mobile market if the mobile subscribers does not make any top-up, the line contract is terminated by the mobile operator. The period for the cancellation of the subscriber line contract is 270-days after the last purchase. However, the operator never knows when the subscriber would do the last credit purchase. Hence, the subscriber behavior can be easily defined with only recency/frequency data and Pareto/NBD model can give the probability of a customer being alive and the expected number of transactions for a customer. Based on these valuable data a mobile operator can calculate the Customer Life Time Value and monetary value of a customer [5].

The main goal of this study is proposing to the subscribers the most appropriate offers by using the results of a defined mathematical model which uses customer life time value of each customer and their probability of churn as the result of the Pareto/NBD model.

The Pareto/NBD model is mostly prefarable for sales transactions which have discrete purchase cycle among customers and non-contractual relationship with the company. The Pareto/NBD model and RFM data are studied in detail in the literature. [5], [6]. However, in this study, a benchmark with logistic regression is also included to show the power of the model and data.

The flow for this thesis is simply illustrated in the following figure 5.

Figure 5: The Flow For This Study

2.1 CLV

2.1.1 CLV Basics

CLV can be simply described as net the present value of the future cash flows associated with a customer [7]. The money has a time value which is represented by discount (interest) rate. In other words, future value of a money has a present value with a discount (interest) rate. This can be explained in the proceeding simple example. How much does one invest to get 100$ amount money at the end of each year for 3 years? It can be easily seen that the value of the money is not 100$. There has to be a rate between present and the future value. The present value of the money can be calculated if we accept the rate 5%;

$$TodayPrice = \frac{100}{(1+0.05)} + \frac{100}{(1+0.05)^2} + \frac{100}{(1+0.05)^3} \tag{1}$$

This definition can be generalized by the following equation

$$PV = \sum_{t=0}^{n} \frac{FV_t}{(1+r)^t} \tag{2}$$

- PV = Present Value
- FV = Future Value
- r = Discount Rate
- t = Time

This formula is the basis for the CLV calculation. The basic CLV calculation can be given in the basic formula below:

$$CLV = CM_i \left(\frac{R_r}{1 + \delta - R_r} \right) - AC_i \tag{3}$$

Where

- CM_i = Customer Contribution Margin; $CM_i = Rv_i - VC_i$ Where Rv_i = Customer Revenue and VC_i = Customer Variable Cost
- R_i = Retention Rate for Customers
- (δ) = Discount Rate (Cost of capital)
- AC_i = Customer Acquisition Cost

In detailed, retention rate is the proportion of customers who stay with the company. It can be simply calculated by the following formula;

$$CustomerRetentionsRate = ((E - N)/S) * 100$$

Where

- E : Number of customer at the end of a period.
- N : Number of new customers acquired during that period.
- S : Number of customers at the start of that period.

On the other hand, customer acquisition cost in mobile industry which is also called subscriber acquisition cost (SAC) simply means the price of acquiring a new customer for the first time. In mobile industry it includes typically the cost of sales and marketing, and handset subsidies and simcard costs.

CLV can be used to calculate the value of the customers of the company and allows company to manage their total costs for their customer relationships.

2.1.2 CLV In This Study

If a mobile operator wants a profitable customer management they have to calculate the CLV of a subscriber [1]. Loyal customers can mean more profit for the company. They tend to use more product and service of the company and can spread positive word of mouth about the brand if they feel satisfied with the product and/or services. The studies on customer life time value in literature can be classified in four main categories as in Table 1 from [8];

Table 1: CLV Models

Focus	Category	Model
Structural Model	Customer Unit	Individual model
		Segment (Customer base) model
	Prediction data	Retrospective model
		Prospective model
	Transaction	Contractual model
		Non-Contractual model
	Purchase cycle	Discrete model
		Continuous model
Strategic Model	Strategic use of CLV in management	
Normative Model	Relationship between Duration and cost	
Analytic Model	Resource allocation (Budget allocation) / Pricing	

The structural model is mostly focused on calculation the CLV based on the customer and customer groups. It is divided into 8 sub models. The CLV can be calculated based on each individual customer or a customer group can be selected to calculate the CLV of this segment. For the data set, historic transaction data or predicted future data can be used for the calculation. In mobile indusrty prepaid subscribers signe a subscription contract with the operator, but they do not promise a regular payment to the mobile operator. They can do top-up action whenever they want and the mobile operator does not know when the customer will defect. This model is defined as non-contractual model in CL calculation. The top-up action of the subscriber is also discrete action. There is maximum time limit for this top-up action, but within this duration subscriber can do anytime the purchase action. In this study we focused on both Non-Contractual model and Discrete model for purchase cycle which is based on Pareto/NBD model. In [5], the assumption is that monetary value is independent of the underlying transaction process. Two CLV calculation is used in the offer management model. One of them is CLV base which is calculated by the multiplication of the expected transaction count and the last refill amount of the subscriber :

$$CLVbase = (R)X(T) \tag{4}$$

where;

- R : Revenue of Refill amount
- T : Expected transaction count from the Pareto/NBD model

After accepting the offer, modified CLV is calculated which is;

$$CLVModifed = (R)X(D)X(P) \tag{5}$$

where;

- R : Revenue of Refill amount
- D : Discount rate for offer
- P : Probability of aliveness

2.2 Recency Frequency Monetary (RFM) Analysis

RFM data includes the transactional data of the customers. Recency can be calculated by the date of the customers' last transaction data, frequency can be calculated by the count of the customers' transaction which fall into between the first transaction date and last transaction date of the subscribers' lifecycle and the monetary value gives the transaction amount of the subscriber. This is the simplest way which can be used to define the non-contracted customers' behavior.

The RFM values can be also used to segment customers to identify the customers who have the highest probability to respond to campaigns. By using the RFM calculation, each subscriber would have a value score assigned to them.

The score is calculated first by dividing the customers in quintiles then the recency, the frequency and the monetary value of the customers are scored beginning from 5 to 1 in descending order. The customer who has the most recent value is scored with 5 then the less recent customers are scored with the following numbers 4,3,2,1. The same method than applied to the frequency and the monetary values of each customer. At the end, each customer has a score from 555 to 111. Totally, there could be 125 buckets in which the customers are segmented. The customers in the RFM bucket which is 555, could have the highest probability to respond the campaigns [9].

2.3 Pareto/NBD Model Assumptions

The Pareto/NBD model is defined by SMC (Schmittlein, Morrison, and Colombo) to model the repeat purchase for a non-contractual customers [6]. SMC states the model has several assumptions regarding customers: Individual customer;

- Poisson Purchases: While alive, each customer makes purchases according to a Poisson process with rate λ.
- Exponential Lifetime: Each customer remains alive for a lifetime which has an exponentially distributed duration with death rate μ.

Heterogeneity across customers;

- Individuals' purchasing rates are distributed following Gamma distribution with rate λ.

- Customers is distributed according to a gamma distribution across the population of customers (NBD distribution).

- Death rates also follow a Gamma distribution with rate μ, and customers have different gamma distribution across (Pareto distribution).

- Rates λ and μ are independent: The purchasing rates λ and the death rates μ are distributed independently of each other.

By using these distributions on the basic RFM data, SMC derived expressions for [10]

- The probability of a customer is still alive,

- The expected number of future transaction for a customer.

RFM data as a base data for Pareto/NBD model can be used by the mobile operator to offer their products or services after calculating the CLV and using this value as an input to a mathematical offer model. The Pareto/NBD model is defined by SMC (Schmittlein, Morrison, and Colombo) to define the repeat purchase for a non-contractual customers [6]. SMC states the model assumptions as below:

Individual customer;

1. Poisson Purchases: While alive, each customer makes purchases according to a Poisson process with rate λ.

$$P(X(t) = x|\lambda) = \frac{(\lambda t))^x e^{-\lambda t}}{x!}, x = 0, 1, 2, ... \tag{6}$$

   This is equivalent to assuming that the time between transactions is distributed exponential with transaction rate λ,

$$f(t_j - t_{j-1}|\lambda) = \lambda e^{-\lambda(t_j - t_{j-1})}, t_j > t_{j-1} > 0 \tag{7}$$

2. Exponential Lifetime: Each customer remains alive for a lifetime which has an exponentially distributed duration with death rate μ

$$f(\tau|\mu) = \mu e^{-\mu\tau} \tag{8}$$

Heterogeneity across customers;

3. Heterogeneity in transaction rates across customers follows a gamma distribution with shape parameter r and scale parameter α.

$$g(\lambda|r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\lambda\alpha}}{\Gamma(r)} \tag{9}$$

4. Heterogeneity in dropout rates across customers follows a gamma distribution with shape parameter s and scale parameter β.

$$g(\mu|s, \beta) = \frac{\beta^s \mu^{s-1} e^{-\mu\beta}}{\Gamma(s)} \tag{10}$$

5. The transaction rate λ and the dropout rate μ vary independently across customers

6. Death Rates Distributed Gamma: The customer's death rates μ are distributed according to a different gamma distribution across customers. (Pareto distribution).

7. Rates λ and μ are Independent: The purchasing rates λ and the death rates μ are distributed independently of each other.

Assumptions 1 and 3 gives the NBD model;

$$P(X(t) = x|r, \alpha) = \frac{\Gamma(r+x)}{\Gamma(x)x!} \left(\frac{\alpha}{\alpha+t}\right)^r \left(\frac{t}{\alpha+t}\right)^x \tag{11}$$

2 and 4 gives Pareto model;

$$f(\tau|s, \beta) = \frac{s}{\beta} \left(\frac{\beta}{\beta+\tau}\right)^{s+1} \tag{12}$$

$$F(\tau|s,\beta) = 1 - \left(\frac{\beta}{\beta+\tau}\right)^s \tag{13}$$

Likelihood function for a randomly-chosen individual with purchase history

$$(X, t_x, T)$$

:

$$L(r,\alpha,s,\beta|x,t_x,T) = \frac{\Gamma(r+x)\alpha^r\beta^s}{\Gamma(r)}\left\{\frac{1}{(\alpha+T)^{r+x}(\beta+T)^s} + \left(\frac{s}{r+s+x}\right)A_0\right\} \tag{14}$$

where $\alpha \geq \beta$

$$A_0 = \frac{F_1(r+s+x, s+1; r+s+x+1; \frac{\alpha-\beta}{\alpha+t_x}}{(\alpha+t_x)^{r+s+x}} - \frac{r+s+x, s+1; r+s+x+1; \frac{\alpha-\beta}{\alpha+T}}{(\alpha+T)^{r+s+x}} \tag{15}$$

where $\alpha \leq \beta$

$$A_0 = \frac{F_1(r+s+x, s+1; r+s+x+1; \frac{\beta-\alpha}{\beta+t_x}}{(\beta+t_x)^{r+s+x}} - \frac{r+s+x, s+1; r+s+x+1; \frac{\beta-\alpha}{\beta+T}}{(\beta+T)^{r+s+x}} \tag{16}$$

The four Pareto/NBD model parameters (r, α, s, β) can be estimated via the method of maximum likelihood. Suppose we have a sample of N customers, where customer i had x_i transactions in the period $(0, T_i]$, with the last transaction occurring at t_{x_i}. The sample log-likelihood function is given by;

$$LL(r,\alpha,s,\beta) = \sum_{i=1}^{N} ln[L(r,\alpha,s,\beta|x_i,t_{x_i},T_i)] \tag{17}$$

This can be maximized using standard numerical optimization routines [11]. With the derivations given above the aliveness probability of a customer can be obtained from the model in the following equation;

$$P(alive|r,\alpha,s,\beta,x,t_x,T) = \left\{1 + \left(\frac{s}{r+s+x}\right)(\alpha+T)^{r+x}(\beta+T)^s A_0\right\} \tag{18}$$

A_0 is given in the previous equations (15) and (16).

2.4 Logistic Regression

Logistic regression measures the relationship between categorical dependent variable and the independent variables. The independent variables can be one or more. The function which is used to calculate the probability of the relation between the dependent and independent variables is logistic function. Logistic regression is one of the most powerful methods to calculate the probability of an event.

In this paper logistic regression is used to benchmark the Pareto/NBD model. First logistic regression is applied to the based RFM data and then applied to other calculated variables which will be explained in the following section.

Logistic regression model is a prediction between a predictor variable (independent variable) and response variables (dependent variables). The difference from the linear regression is the predictor is transformed using a-non-linear function called the logistic function. It is simply non-linear transformation of linear regression. It is illustrated by the following simple graph Figure 6.

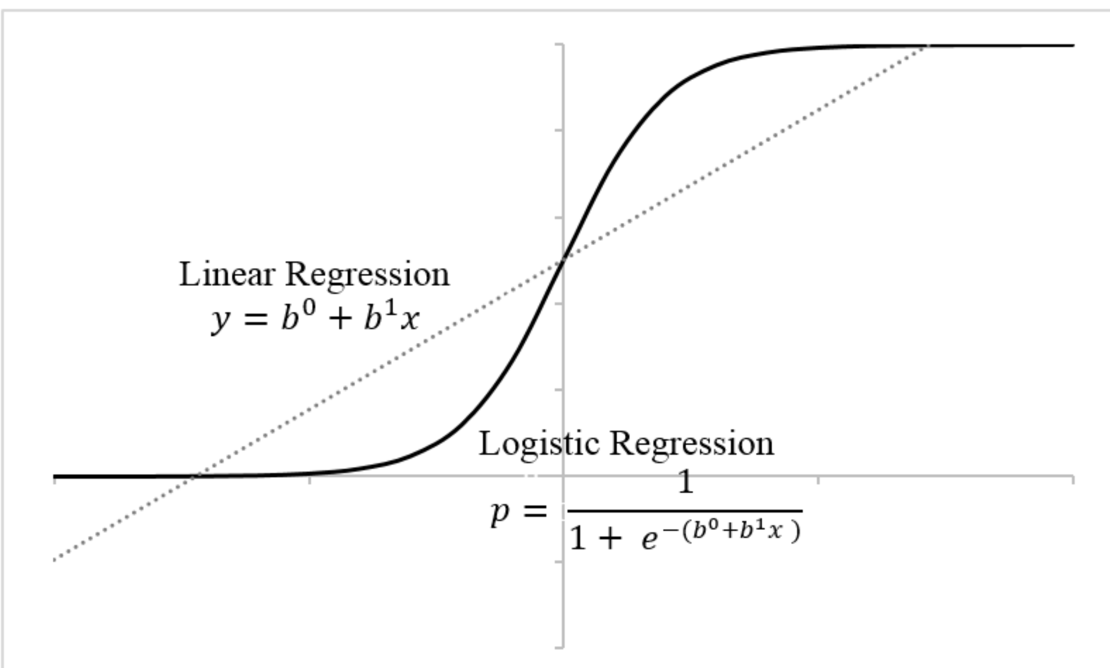


Figure 6: Linear Regression & Logistic Regression Comparison

There are three types of logistic regression.

- Binary Logistic Regression : Used when the response is binary like in our model if the subscriber has churned or is still active.
- Nominal Logistic Regression : Used when there are nominal categories and there is no order between these categories.
- Ordinal Logistic Regression : Used when there are ordinal categories and the categories can be sorted.

Binary logistic regression is mostly used with two outcomes scenarios like in predicting churn of subscribers. Unlike lineer regression a link function is used to predict the related outcomes.

$$t = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \epsilon \tag{19}$$

The link function for the logistic regression;

$$t = g(\theta) = log\left(\frac{\theta}{1-\theta}\right) \tag{20}$$

Inverse of link function;

$$\theta = \frac{e^t}{(1+e^t)} \tag{21}$$

The link function transforms the probabilities (between 0 and 1) into regression scores (between $-\infty$ and $+\infty$)

2.5 Offer Management

In the first years of mobile communication, there was very limited number customer services that was only voice and short message services(SMS) were available. As mobile phone technology evolved, GSM service providers could not only provide voice or SMS services. They have to adapt to the rapidly changing technology more than other industries. Because the subscribers can easily leave the mobile operator, if the mobile operator does not propose a value to the customer. Therefore during the years they generated many product and services for their subscribers to keep them loyal with their brand. The mobile operators created their own product lines in a very broad perspective. One can easily describe each of the products as offer.

Offers includes marketing activities to have the attraction of the customers over any kind of channel. Campaign management is to contact with the right customers at the right time over the right channel and with the right offer Figure 7.

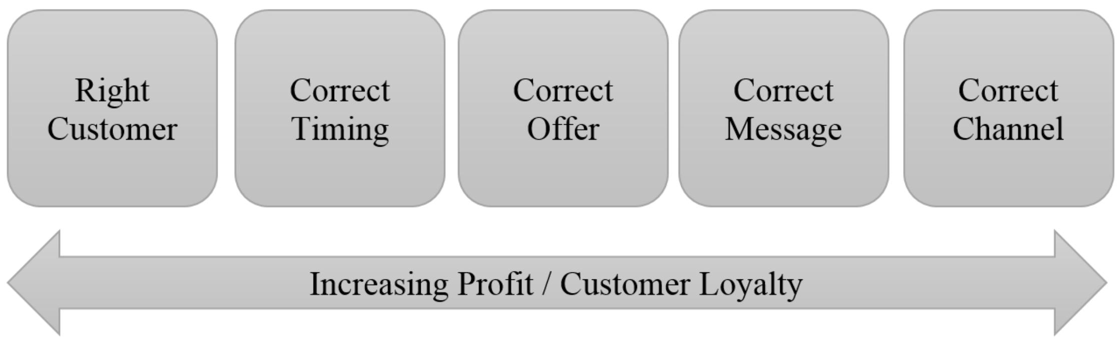


Figure 7: Offer Management Cycle

If a mobile operator can achieve a good campaign management, then has the probability to increase their revenue and also decrease their churn rate. We said that a mobile operator could not do anymore only churn analysis. It has to be adapted in a campaign management system. We can define an offer management framework for a mobile operator like below Figure 8;

The Average Revenue per User in Turkish Lira (TL) (ARPU-TL) and subscriber trend is shown in Figure 5 and Figure 6, respectively. This market data is published quarterly to report the market trends in mobile sector in Turkey by Bilgi Teknolojileri Kurumu (BTK) which is the Governmental Organization of Information Technologies. It can be seen that ARPU of the postpaid subscribers is higher than the prepaid subscribers and the time shows that the number of the postpaid and prepaid subscribers converge to each other. It is also known that keeping a subscriber is cheaper than acquiring a new customer. If a mobile operator calculate the CLV of the prepaid

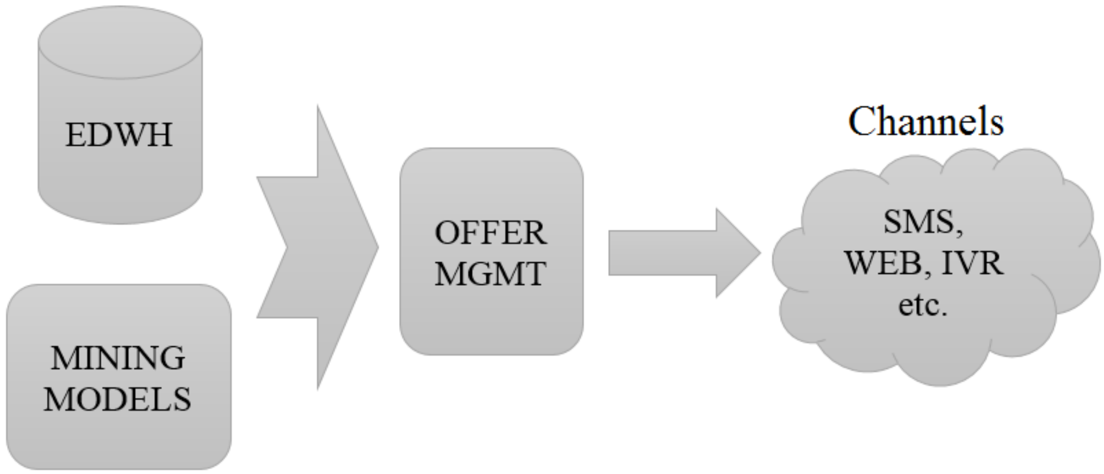


Figure 8: Offer Management Framework

subscriber and propose them the right offer than it is obvious that the operator can increase their revenue Figure 9.

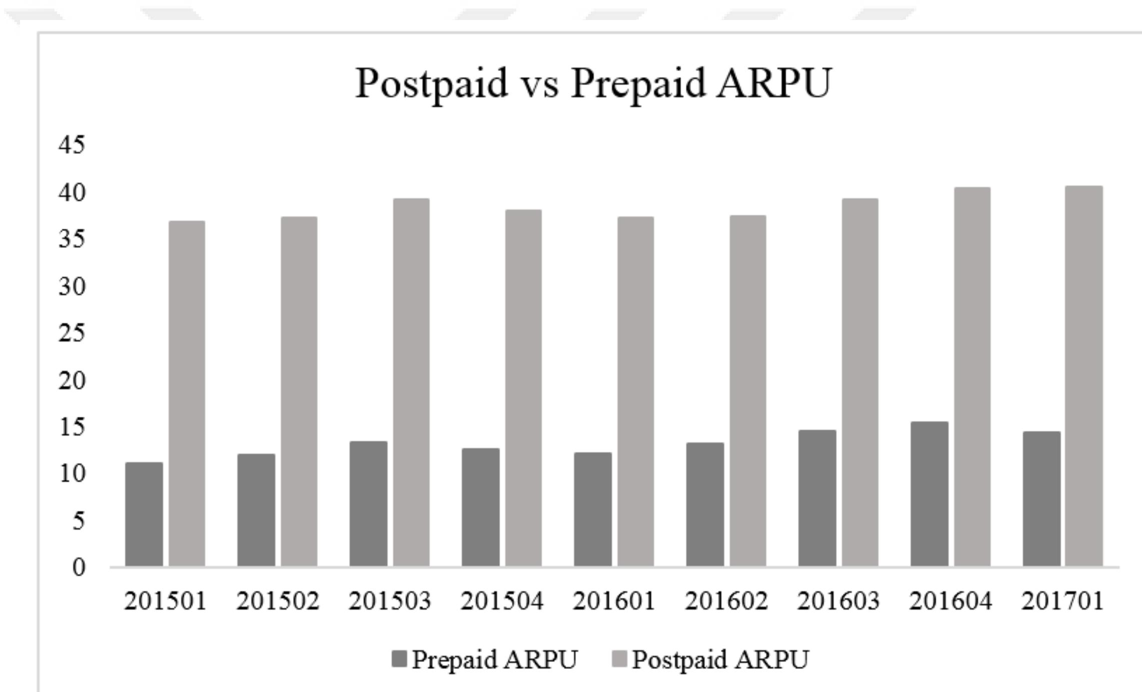


Figure 9: ARPU trend for Postpaid & Prepaid Subscribers

CHAPTER III

RELATED WORK

The activation of the prepaid subscribers can be defined as non-contractual process. They stay alive while continuing to purchase from the mobile operator. There are some purchasing actions that resets their 270-day period to zero day and the subscriber continues to generate revenue. Otherwise, they become inactive at the end of the 270-day period.

The churn rate of the prepaid subscribers is three times higher than the postpaid subscribers. In section 1.2, the penetration rate in Turkey Mobile Market has been given. It can be seen that the penetration rate is higher than the population which %108 at the end of the 2017-Q2. Therefore, increasing the customer loyalty or offering new product services to the subscribers will give the control to the mobile operator rather than the customer. As a result, the mobile operator can increase their revenue.

There are a lot of modelling works with postpaid subscribers in mobile sector. However, defining and managing prepaid subscribers' future value is harder than postpaid ones because of their unobserved behavioral format.

There is a publication about this behavior of prepaid subscribers [12].They use decision tree algorithm to segment the prepaid subscribers to define who are going to churn based on CDR and SIM data.

There is a similar work as mentioned above. The data set in this study [13] includes prepaid subscribers model variables. The churn prediction is done by using the logistic regression, linear regression and Fisher linear discriminant analysis and decision trees. There are other studies for mobile operator churn prediction. However, most of them does not directly focus on prepaid subscribers' behavior [14]. In this study, specifically churn prediction is studied based on CDR data without specifying any customer base. They work on feature selection methods and supervised learning algorithms to predict the churn score for the subscribers. [15], [16], [17] are again the studies that focus on postpaid subscribers churn propensities. Another paper [18] again analysis the churn prediction but does not define a customer base.

As stated at the beginning, prepaid subscribers can be accepted non-contractual subscribers. In marketing non-contractual based data can be widely analyzed with RFM data [9], [19]

In this paper [20], RFM analysis, logistic regression and decision trees are used to compare with each other based on the accuracy of the data.

Some other publications also focused on the performance of the churn prediction models. [21] applies many data mining methods to a mobile operator data. But the main focus is to improve the model accuracy. [22] used hybrid churn prediction model for prepaid subscribers and compared the accuracy of the data mining models. [23] proposed a hybrid model for churn prediction.

Although there are many churn pediction publications, there are a few with mobile prepaid subscribers data. Most of them focuses on the performance improvement of the data mining methods. In this work, the power of RFM data is used to predict prepaid subscribers behavior.

This study uses refill history of prepaid subscribers which can be accepted as a purchase order thus can be defined easily with RFM data. RFM data is a simple but a strong data to analyze valuable customers. RFM data is reviewed in the study [24]. In this paper it is emphasized that there are a lot of advantage especially to use this data which can boost company revenue in a short term despite there are also disadvantages some of which are;

- It only focuses on best customers. If customer do not buy often or spend little or do not generate any transaction lately, it provides little information about

RFM scoring.

- It only focuses on limited variables. It is better to take into account other customer relation variables when using RFM data.

Although some of the disadvantage of the RFM data, the simplicity of the data variables and the preparation of the data makes it still a powerful and useful data. Specifically for mobile company. The RFM data can find for itself many kinds of applicable areas [25]. Furthermore RFM data can be used for calculating CLV of customers. The calculation of customer based CLV is not a common study area for a mobile operator company. There are a lot of churn analysis in the literature but mostly, they do not directly focus on customer value. Only doing churn analysis does not give the right customers all the times. Because there are millions of subscribers, a single subscriber based small cost can produce a high cost for the company if you apply your analysis results in a thousands of subscribers. Therefore a mobile operator has to improve their churn analysis with the power of the CLV calculation. There are some studies which covers CLV calculation for telco industry. But the customer base is not prepaid mobile subscribers. The study [8] deals with wireless telecommunication subscribers and calculate the CLV based on a new approach of Markov chain model. The study [26] is another study on land-line churn prediction. The study [27] is focused on contracted telecommunication services (land-line or mobile phone or internet line). Some hypothesis are defined in terms of validation and calculation of the CLV and customer equity values. CLV calculation cannot be ignored in todays' challenging marketing. In the big data era, making descriptive analysis on historical data is not enough for improving customer loyalty. The competition conditions in any market compelled the companies to analyze the future behavior of their customer to improve the customer experience with their product. Thus, the CLV calculation is a must in todays' market. Especially, like mobile market where the prepaid subscribers are dominated, the mobile companies have to take care on their prepaid subscribers which has more complicated behavior than the postpaid subscribers.

In this study we also use the CLV for the best offer calculation for the subscriber to retain them with the company. One can base their best offer calculation just on the churn probabilities of the customers. In this case, the model false positives rates becomes very important because if you offered wrong customer the churn offer, you would probably wasting your budget on wrong customers which would not be leave your company. The study [28] is about cost-sensitive approach of offer management for churners.

The RFM data was extended by [5] for the CLV calculation. As stated in the paper, the study focused on using only the RFM data for the CLV calculation in customer base. The customer base data is the transactions for the non-contractual setting. This CLV calculation based on RFM data can be easily adopted to prepaid subscribers in mobile sector. However, there is no application of this method for mobile prepaid subscribers, we can give some other applications of CLV calculation based on RFM data. This method is used widely in many industries, there are many case studies like retail in these papers [29], [30].

In [5], the Pareto/NBD model is used. Pareto/NBD model is used to define the non-contractual purchase transactions for customers who you will never know when they leave your brand. There are many other models that can be used instead of Pareto/NBD model. In [30] a comparison for different models are made.

Offer management is also another aspect of predicting churn of subscribers. A mobile operator have to manage their costs for their customer loyalty. In [1], it is emphasized that companies cannot just only focus on profitable customer but also finding the loyal customers is more important than profitable customers. It is also important that companies has to solve the optimization problems for the promotions and calculate the maximum profit from their offer campaigns [31].

CHAPTER IV

DATA PREPARATION

In offer management model, the results are used from the base model which is Pareto/NBD. As stated in the thesis flow in Section 2, a benchmark is done for the RFM data and the extra subscriber features data. For this reason two data set is prepared for ;

1. RFM data
2. Features for the logistic regression

4.1 RFM Data

Pareto-NBD model uses RFM data. In this thesis, we analyze the active mobile prepaid subscribers. The data is provided by one of the mobile operators in Turkey. If prepaid mobile subscribes do not make any top-up, their contract will end in a defined period of time, 270 days in Turkish market. But you never know when they will stop making top-up so when they will end their contract. In this thesis the RFM data set includes prepaid mobile subscribers who make their first activation with the prepaid charging method and do not change their charging method in the selected time interval. The selected time interval is 2 year period from 1st of February, 2015 to 31th of January, 2017.

The number of distinct subscribers is 386K. The refill amount and refill count is calculated based on the refill transactions. The number of transactions is 2.7M. The subscribers with no transactions are eliminated from the base subscriber set after elimination the observer subscribers with refill transactions are 327K.The refill

Table 2: A small sample of the RFM Data from Refill Transactions

Subscriber Id	Refill Date	Refill Amount (TL)
132047392	20160812	15
132047392	20160901	20
...	...	...
132054290	20150328	25
132054290	20150416	12

transactions for the same subscriber on the same date are merged into one record so the 2.700.349 record has become 2.578.681 distinct transactions.

For the simplicity of the model, not all the transactions are fed into the PARETO/NBD model, the data is split into 50 buckets with ORA-HASH function. The size of the one of the bucket has approximately 55K refill transaction and the distinct subscriber is 6500. The model is run with 6 sets of this data. Each transaction contains a Subscriber Id which uniquely defines the customers, refill date the date of the transaction, refill amount the amount of refill transaction in TL in other words the monetary value of the refill transaction. The data is order by subscriber id and refill date in ascending. A small sample is presented in Table 2.

Prepaid subscribers do refill action when they decide. The lifecycle period for the prepaid subscriber is 270-days. If they do not do any refill action, after 270-days the subscription will be ended by the mobile operator. But this period will be reset by refill transaction and another 270-days new period starts. The total subscriber used in this study is 6480 which were activated two year period starting from 1.2.2015 to 31.1.2017. Refill transactions are also for the same period for the activated subscribers. The subscribers are chosen among prepaid subscribers who did not do any charging method change and remains prepaid subscriber for their lifetime.

In this thesis we also add segment information of the subscribers to the subscriber set. We want to group the subscribers in to the same segment which usually has the same behavior for creating the transactions. Our segment information includes

Table 3: The Number of Distribution

Segment Information	OTHER	YOUTH	MASS	Total
# of Subscriber	1,063	1,352	4,065	6,480
# of Refill Transaction	6,290	12,620	34,338	53,248

YOUTH, MASS, and OTHER respectively which mean the subscribers under age 26, all existed subscribers and not segmented new subscribers that the mobile operator do not have any enough information about them. The number of distribution can be seen in Table 3 below. For the RFM data, the refill transactions belong to again the same two year period with the subscribers. For recency parameter we have taken the refill date, for the monetary value we have taken the refill amount in TL. The number of rows is 53.249. The number of distribution can be seen in Table 3.

4.2 Features Selection for the Model

For model benchmarking with logistic regression the defined variables were prepared. Usage (Data, Voice, SMS) behaviors,

- Usage statistics which includes if the subscriber has any usage in the last 3 months.
- Refill behaviors which holds the sum of the last 12 months refill transaction amounts. The refill transactions include both voice and data separately.
- Package usage properties which hold if the subscriber make any package refill in the last 12 months.
- ARPU (average revenue per user) properties.
- Call center and online interaction transaction variables were also included.

The subscriber set for this variable is the same with RFM data set. The variables are calculated in monthly bases. The selected month is the last month before the calibration date which is January of 2016.

Table 4: Basic properties of the selected data sets.

Transaction Count	~50K
Distinct Subscriber Count	~6.500
Minimum Refill Date	01.03.2015
Maximum Refill Date	31.01.2017
Days Between Dates	708

4.3 Exploratory Data Analysis

Six buckets of the transaction data is used. 15% of the subscribers, which is 59.810 subscribers, never made a top-up during the selected period. They are removed from the used data set. The basic properties of the data sets are shown in Table 4.

The refill subscriber' s behavior cannot be estimated beforehand like postpaid subscribers. The demographic information for the prepaid subscribers usually differs from the postpaid ones. Mostly young people prefer to use prepaid lines which is shown in the Figure 10.

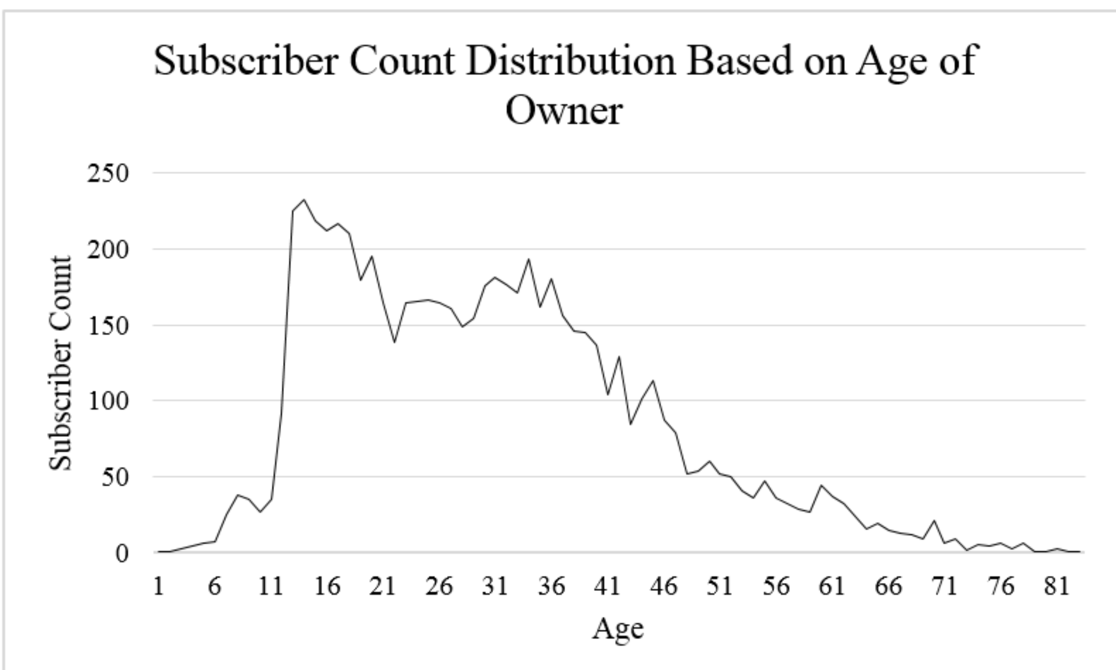


Figure 10: Subscriber Distribution Based on Age

The basic statistics of the sets are given below in Table5. The statistics are

Table 5: Basic statistics of the selected data sets.

Set	Trans	Min	Max	Mean	Var	Stddev
Set 1	53,320	2	180	24	122	11
Set 2	52,683	2	360	24	123	11
Set 3	53,091	1	180	24	123	11
Set 4	51,920	1	180	24	124	11
Set 5	53,900	1	360	25	130	11
Set 6	53,248	2	360	24	147	12

Table 6: The distributions of the data sets.

Set	Q5	Q25	Median	Q75	Q95
Set 1	10	19	25	30	40
Set 2	10	19	25	30	40
Set 3	10	19	25	30	40
Set 4	10	19	25	30	45
Set 5	10	19	25	30	45
Set 6	10	19	25	30	40
Main	10	19	25	30	40

generated based on the Refill Amount (TL).

The statistics for the subsets are nearly the same as the main set which has 2.7M transactions. This ensures that we can use one of the subsets to create the model. The 5th, 25th, 50th, 75th and 95th percentiles of the sets are listed in Table 6.

According to Table 7 it can be easily seen that the sets are shown similar distributions. To see the distribution of the subscriber behavior for the days between transactions, we can see there is positive skewness for the distribution of the days.

Table 7: Values of the days between transactions.

Set	Min	1st.Qu.	Median	Mean	3rd.Qu.	Max
Set 1	1	9	24	31.72	34	593
Set 2	1	9	24	31.93	34	515
Set 3	1	10	26	32.74	35	590
Set 4	1	10	25	32.45	35	545
Set 5	1	9	24	31.45	34	462
Set 6	1	9	25	32.21	35	536

4.4 *Benchmarking Pareto/NBD Model with Logistic Regression*

There are 395 features selected for the logistic regression benchmark. The set includes different type of features which are;

- Call or usage behavior, includes incoming or outgoing calls and the types of communication like voice, data or short message service (sms).
- Network metrics, includes call drop rates.
- Refill metrics, includes the behavior of the refill transactions of the subscriber like refill amount and also usage type based refill utilization like if the subscriber uses mostly sms, data or voice.
- ARPU, includes the average revenue of per user in TL amount.
- # of lines, includes if the subscriber has other lines, or has one or more line.
- Age, gives the age of the subscriber.
- VAS usage / subscription, includes the subscriber usage behavior of value added service or if the subscriber has any subscription for the value added services.
- Pre-calculated SNA metrics, the features also include social network analysis metrics which are calculated in another product and taken as input for the logistic regression
- Network types (2g, 3g, 4g), includes the network type of the subscriber
- Equipment types, give if the subscriber has a smartphone or other basic types.

Features set elimination is done as following;

- 395 features is analyzed with statistical methods.

- First the features whose minimum and maximum values are equal are eliminated. 383 features are left.
- Correlation analysis is made and 308 features are left.
- With domain knowledge only 50 variables are selected.
- Stepwise regression with backward elimination is run over the 50 variables and 10 features are selected.

Logistic Regression is applied as in the following order and also shown in Figure 11 ;

- Result 1: Logistic Regression with RFM
- Result 2: Logistic Regression with 10 features
- Result 3: Logistic Regression with 10 features and RFM
- Result 4: Logistic Regression with 50 features
- Result 5: Logistic Regression with 50 features and RFM

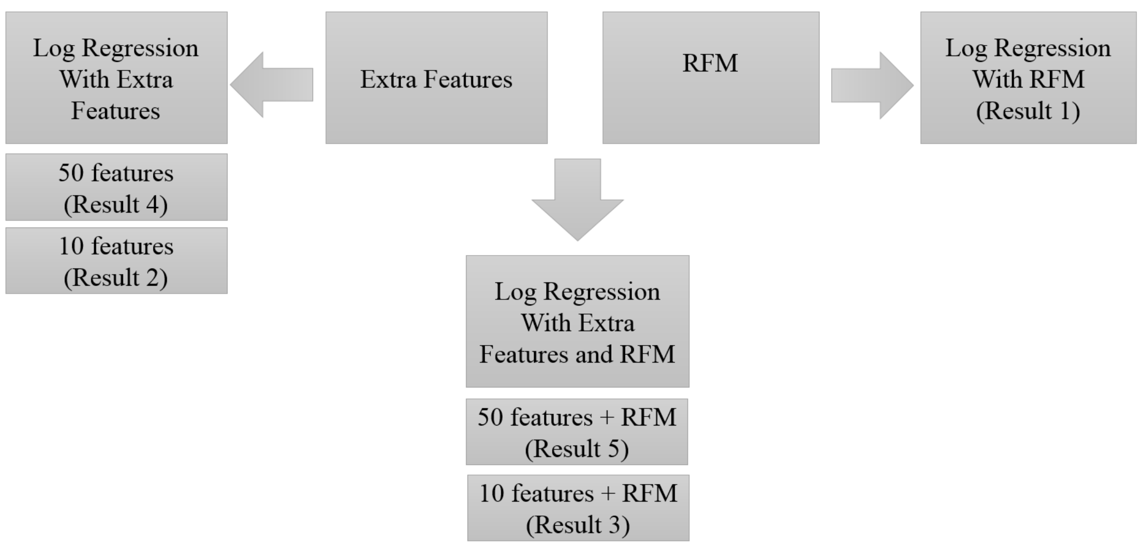


Figure 11: Flow of Benchmarking

The logistic regression benchmark for RFM and the calculated subscriber features results are shared in Table 8.

Table 8: Benchmark of Data with Logistic Regression

Cutoff	Result	Set Name	Accuracy	TP Rate	FP Rate	Precision
0.5	Result 1	RFM	95%	98%	24%	96%
	Result 2	Var 10	91%	100%	64%	90%
	Result 3	Var 10 + RFM	95%	100%	31%	95%
	Result 4	Var 50	92%	100%	56%	91%
	Result 5	Var 50 + RFM	96%	100%	28%	95%
0.6	Result 1	RFM	94%	97%	22%	96%
	Result 2	Var 10	92%	100%	50%	92%
	Result 3	Var 10 + RFM	96%	100%	24%	96%
	Result 4	Var 50	93%	99%	41%	93%
	Result 5	Var 50 + RFM	96%	100%	23%	96%
0.7	Result 1	RFM	93%	95%	20%	96%
	Result 2	Var 10	94%	99%	34%	94%
	Result 3	Var 10 + RFM	97%	100%	20%	97%
	Result 4	Var 50	95%	99%	30%	95%
	Result 5	Var 50 + RFM	97%	100%	18%	97%
0.8	Result 1	RFM	92%	93%	17%	97%
	Result 2	Var 10	94%	98%	29%	95%
	Result 3	Var 10 + RFM	97%	99%	15%	97%
	Result 4	Var 50	95%	98%	24%	96%
	Result 5	Var 50 + RFM	97%	99%	14%	98%
0.9	Result 1	RFM	86%	86%	13%	97%
	Result 2	Var 10	95%	98%	24%	96%
	Result 3	Var 10 + RFM	98%	99%	12%	98%
	Result 4	Var 50	95%	98%	20%	97%
	Result 5	Var 50 + RFM	98%	99%	11%	98%

4.5 Pareto/NBD Model & CLV

Firstly, we only focused on subscriber refill transactions and did not include any tenure or segment information of the subscriber in the transaction data. We found out that without any tenure or segment information the model failed in the test period while showing a good performance in the training period.

Our segment information includes YOUTH, MASS, and NOT SEGMENTED respectively which mean the subscribers under age 26, all existed subscribers and new subscribers that the mobile operator do not have any enough information about them. First we generated the results based on the segment information then included 8 tenure groups which were based on the 3-months sub period information. The tenure was calculated according to the first refill action date. This is illustrated in Figure 12.

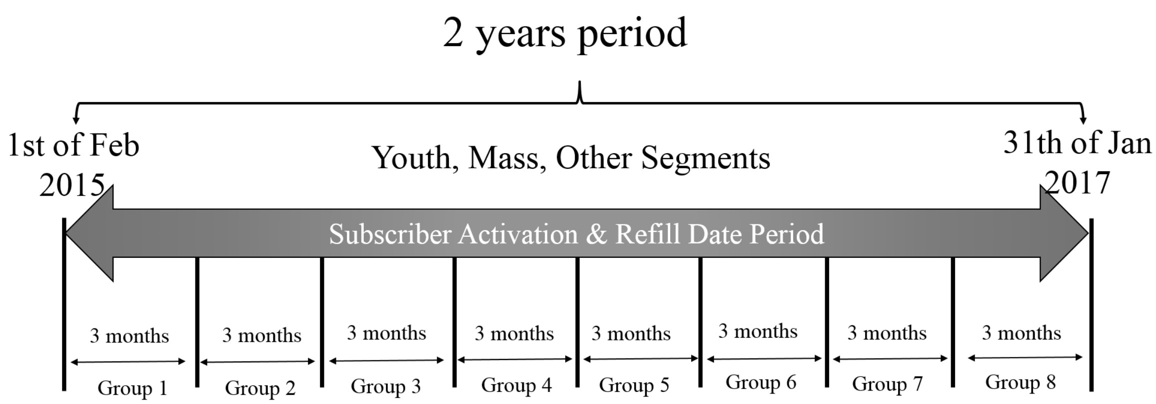


Figure 12: Tenure Groups

The Pareto/NBD model was run based on these data sets which is described above. The results are shared in the Table 9. The Pareto/NBD model results is defined as;

- Result 6: Pareto/NBD model is run with RFM without segment and tenure information

- Result 7: Pareto/NBD model is run with RFM with segment information

Table 9: Pareto/NBD Model Results

Cutoff	Result	Set Name	Accuracy	TP Rate	FP Rate	Precision
0.5	Result 6	w/o Segment + Tenure	88%	98%	69%	89%
	Result 7	Segment	95%	99%	29%	95%
	Result 8	Segment + Tenure	94%	98%	29%	95%
0.6	Result 6	w/o Segment + Tenure	88%	97%	64%	90%
	Result 7	Segment	96%	99%	18%	97%
	Result 8	Segment + Tenure	96%	98%	18%	97%
0.7	Result 6	w/o Segment + Tenure	89%	97%	58%	91%
	Result 7	Segment	96%	98%	13%	98%
	Result 8	Segment + Tenure	96%	97%	13%	98%
0.8	Result 6	w/o Segment + Tenure	90%	96%	46%	92%
	Result 7	Segment	97%	98%	10%	98%
	Result 8	Segment + Tenure	96%	97%	10%	98%
0.9	Result 6	w/o Segment + Tenure	91%	93%	18%	97%
	Result 7	Segment	97%	97%	6%	99%
	Result 8	Segment + Tenure	96%	96%	6%	99%

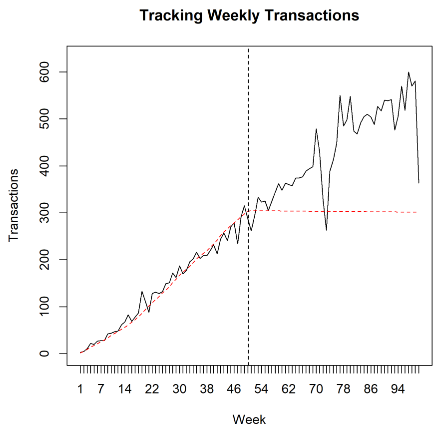
- Result 8: Pareto/NBD model is run with RFM with segment and tenure information

The model performance is given in the figures below (Figure 13) for the all set, mass, youth and not segmented sets. The segment & tenure based data sets have the best performance for the Pareto/NBD model.

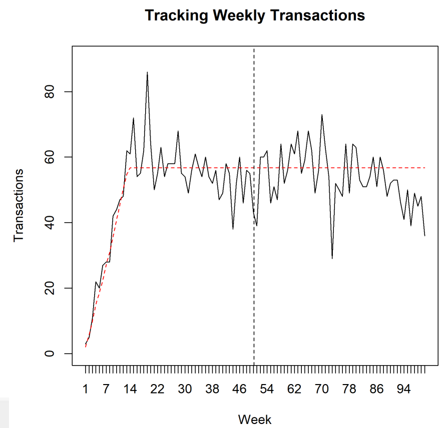
In preceding study since we did not get any good performance in the test period, we could not calculate the CLV value of the subscribers. In this study by adding the tenure and segment information we have got a good performance for the model in test period for that the expected number of transactions and probability of alive of each subscriber are calculated to be used in the offer model.

4.6 Proposed Mathematical Model for Offer Management

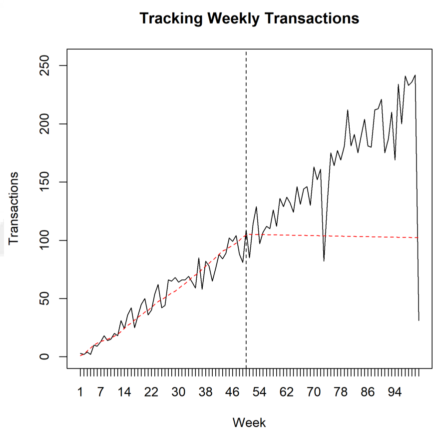
After identifying valuable customers to prevent them from churn, a best offer selection mathematical model is proposed. The model calculates output for each customer an eligible offer. Furthermore, the model maximizes the total CLV by proposing the



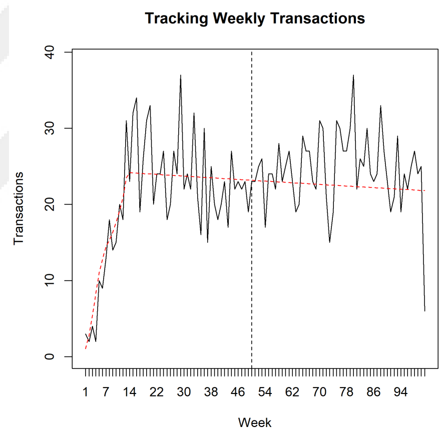
(a) Mass Segment



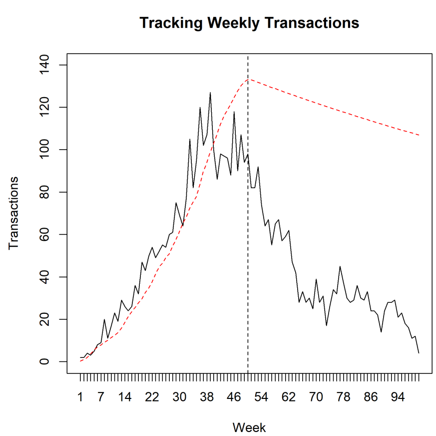
(b) Mass Segment with Tenure



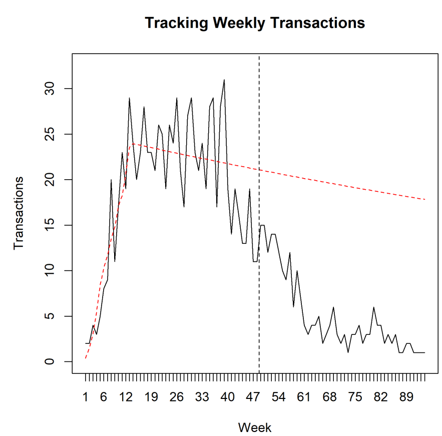
(c) Youth Segment



(d) Youth Segment with Tenure



(e) Others



(f) Others with Tenure

Figure 13: Pareto/NBD Model Results

customers with the eligible offers. Base CLV is calculated by the Pareto/NBD model which is based on the RFM data.

The model is defined as;

Sets;

- i= Customer index, $i \epsilon I$
- j= Offer index, $j \epsilon J$

Parameters;

- $clvbase_i$: Base Customer Lifetime Value of customer i. This amount should be interpreted as the expected present net worth of customer i for the following cases: *i)* customer receives no offer or *ii)* customer refuses the proposed offer.
- p_i : Churn probability of customer i to company before receiving the offer,
- o_i : Upper bound of discount for customer i,
- w_i : Upper bound of offer j,
- c_j : Discount rate of offer j,
- f_i : Monthly payment of customer i to company before receiving the offer,
- f_{avg} : Monthly average payment of customers,
- β_{ij} : Probability of accepting offer j by customer i $= p_i^{c_j \frac{f_{avg}}{f_i}}$
- p_{ij} : Churn probability of customer i after taking offer $= p_i(1 - p_i^{c_j \frac{f_{avg}}{f_i}})$,
- clv_{ij} : Modified Customer Lifetime Value of customer i,
- t : Monthly budget for discount,
- M : Large positive number.

Decision Variables;

- $x_{ij} = \begin{cases} 1, & \text{if customer i get offer j} \\ 0, & \text{"o/w"} \end{cases}$

$$maxz = \sum_i [\sum_j x_{ij}[\beta_{ij} clv_{ij} + (1-\beta_{ij}) clvbase_i] + (1 - \sum_i x_{ij}) clvbase_i] \tag{22}$$

$$\sum_i \sum_j f_i c_j x_{ij} \leqslant t \tag{23}$$

$$\sum_i f_i c_j x_{ij} \leqslant o_i, \forall i \tag{24}$$

$$\sum_i x_{ij} \leqslant w_j, \forall j \tag{25}$$

$$x_{ij} \epsilon 0, 1, \forall i, j \tag{26}$$

The objective function aims to maximize expected customer lifetime value while increasing retention of customers by giving the right offer. In other words, the expected CLV depends on whether the customer is received an offer or not, and if the customer is received an offer, it also uses the probability of the accepting the offered or rejecting the offer by the customer. The model is simply illustrated in Figure 14.

Constraint 23 states that discounts made cannot exceed the total budget. Constraint 24 indicates the given offer cannot be predetermined discount limit for each customer. However, it occurs some infeasibilities for some customers. Constraint 25 ensures that number of offers given cannot exceed the specified limit.

We use Pareto/NBD results for CLV and churn probabilities (Result 7). We calculate probability of accepting each offer by customers (β_{ij}) depending on the customer's package refill payment (f_i), churn probability (p_i) and average package refill payment

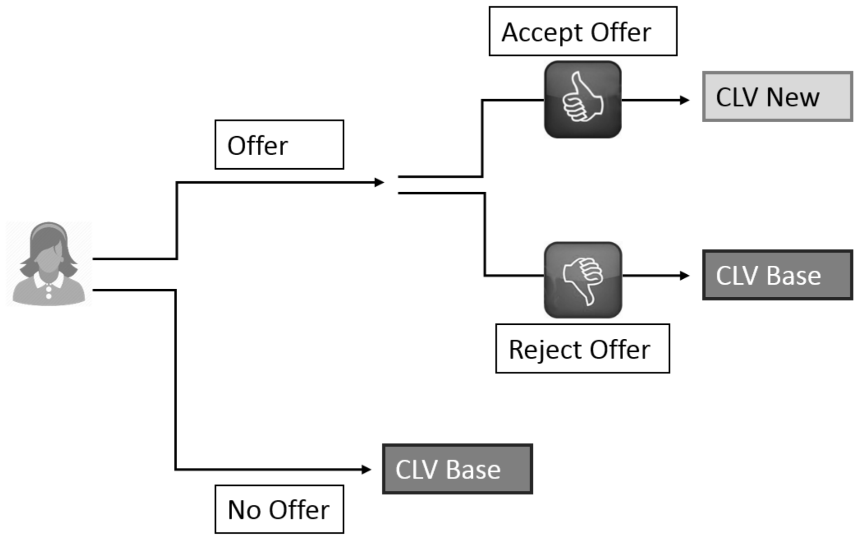


Figure 14: Offer Model

of all customers (f_{avg}). Having lower churn probability, higher package refill payment from average and higher discount rate give higher probability of accepting an offer:

$$\beta_{ij} = p_i^{c_j \frac{f_{avg}}{f_i}}, \forall i, j \tag{27}$$

In the same way, we calculate new churn probabilities (ρ_{ij}) after receiving each offer depending probabilities of accepting offer (β_{ij}) and churn probabilities (p_i). Having higher churn probability (ρ_i) and higher probability of accepting an offer (β_{ij}) gives lower new churn probability after receiving an offer:

$$\rho_{ij} = p_i(1 - p_i^{c_j \frac{f_{avg}}{f_i}}), \forall i, j \tag{28}$$

Then, we use the new churn probabilities (ρ_{ij}) to calculate modified CLVs (clv_{ij}). Modified CLV is the expected present net worth of customer i assuming customer accepts offer j. A mixed integer based mathematical model for offer management is developed, and solved by using GAMS software with CPLEX Solver. We solved problem with 6480 customers and 5 offers in Table 10.

Different scenarios were run with the 6480 customers and the following results are

Table 10: Offers

Offer	**Discount %**
Offer 1	10
Offer 2	20
Offer 3	30
Offer 4	40
Offer 5	50

observed Table 11. We did not limit the customer set and used all the customers as input for the model.

- Scenario1, the budget was set to 10.000 and offer count was set to 300. There is no cut off value for the customer probability.

- Scenario2, the budget was set to 10.000 and no offer limit was applied. There is no cut off value for the customer probability.

- Scenario3, the budget was set to 5.000 and offer count was set to 300. There is no cut off value for the customer probability.

- Scenario4, the budget was set to 5.000 and no offer limit was applied. There is no cut off value for the customer probability.

- Scenario5, the budget was set to 1.000 and offer count was set to 300. There is no cut off value for the customer probability.

- Scenario6, the budget was set to 1.000 and no offer limit was applied. There is no cut off value for the customer probability.

- Scenario7, the budget was set to 10.000 and no offer limit was applied with 0.5 cut-off for the customer probability.

In table 14, the results show that when there is no limit for the offers, more customers accept the offer. The high-performance model is the "Scenario4". Because

Table 11: Offer Model Scenarios

	Budget	**Offer Limit**	**Cutoff**	**No of Customers**	**No of Received**
Scenario1	10,000	300	0	6,480	741
Scenario2	10,000	inf	0	6,480	3,398
Scenario3	5,000	300	0	6,480	532
Scenario4	5,000	inf	0	6,480	1,498
Scenario5	1,000	300	0	6,480	168
Scenario6	1,000	inf	0	6,480	168
Scenario7	10,000	inf	0.5	781	562

Table 12: Churn Rate Breakdown

	Churn Rates		
	p<0.5	p>=0.5	Total Subs
Scenario1	294	447	741
Scenario2	3,331	67	3,398
Scenario3	294	238	532
Scenario4	1,485	13	1,498
Scenario5	168		168
Scenario6	168		168
Scenario7		562	562

Table 13: Accepted Offer Distribution

	Discount Offers					
	Offer1 %10	**Offer2 %20**	**Offer3 %30**	**Offer4 %40**	**Offer5 %50**	**Total Subs**
Scenario1	300	57	39	60	285	741
Scenario2	3352	10	2	11	23	3398
Scenario3	300	22	13	35	162	532
Scenario4	1495	3				1498
Scenario5	168					168
Scenario6	168					168
Scenario7	121	57	39	60	285	562

Table 14: Calculated CLV

		CLV Calculated	CLV Base	CLV Diff	Churn Prob Base Avg = 12%
Offer Limit = 300	**Scenario1**	1,239,305	1,203,300	36,005	55%
	Scenario3	1,239,043	1,203,300	35,743	42%
	Scenario3	1,239,043	1,203,300	35,743	2%
No Offer Limit	**Scenario2**	1,267,980	1,203,300	64,680	3%
	Scenario4	1,257,194	1,203,300	53,894	2%
	Scenario6	1,228,630	1,203,300	25,330	2%

mostly Offer1 is accepted and the higher CLV is got from the model. Limiting total budget does not have much affect over the CLV maximization. In Scenario2, more customers accept the offer and higher CLV is got. However, comparing to Scenario4, with lowest churn probability and limmited budget, Scenario4 has the higher performance.

CHAPTER V

CONCLUSION

The mass amount of the mobile subscribers are still prepaid subscribers Figure 1. In years both the number of postpaid and prepaid subscriber has converged to each other as a result of data usage increasing due to 4g network upgrade by the mobile operators. In this study we have stated that life time value is important part of the churn prediction. Therefore a mobile operator should use the results in their offer management lifecycle. They have to optimize their cost for churn retention program and know the subscribers who will really value to propose an offer. Otherwise without calculating the CLV, the mobile operator has to depend on the results of the churn prediction model and accept the burden of the false positive subscriber sets effect over their cost.

In this study, the segment and tenure information is also included in the data set. It can be seen from the (Figure 13) that the model gave good results for the youth and mass segments in the test period. If the customer has other segment, the model failed in Figure 13e and Figure 13f.

For CLV calculation we have used the subscriber sets result with segment information which is Result 7 in 9. The selected set has the highest Accuracy among the other sets. Instead of proposing an offer to every churner, the mathematical model gives us the flexibility of the offering to the most valuable subscribers who has the higher potential to accept the offers.

Different scenarios were tested. The maximized CLV results are shared 14. As stated in the previous section 4.6, the highest performance is got from the Scenario4. This means that if no offer limit is applied most customers accept the proposed offer and it also increases the retention rate which means more loyal customers.

On the other hand, parameter estimation has to be improved for the model. There are some different studies about the parameter estimation [32], [33] and [34] However we decided to focus on the data set to get the results for the CLV and offer management.

Lastly, we may try to test the framework and measure the real performance of the system by collecting also the response rates of the subscribers.

Appendices

APPENDIX A

PARETO/NBD R CODE

```
############################################
# INSTALL AND LOAD NEEDED PACKAGES
############################################
install.packages(ggplot2)
install.packages(BTYD)
install.packages(reshape2)
install.packages(plyr)
install.packages(lubridate)
install.packages(xlsx)
install.packages(gsl)
install.packages(stringr)
library(ggplot2)
library(BTYD)
library(reshape2)
library(plyr)
library(lubridate)
library(xlsx)
library(gsl)
library(stringr)


############################################
# USED BTYD PACKAGE IS FIXED ONE : https://github.com/theofilos/BTYD
############################################
h2f1 <- function(a,b,c,z){
  lenz <- length(z)
  j = 0
  uj <- 1:lenz
  uj <- uj/uj
  y <- uj
  lteps <- 0
  while (lteps<lenz){
    lasty <- y
    j <- j+1
```

```
    uj <- uj*(a+j-1)*(b+j-1)/(c+j-1)*z/j
    y <- y + uj
    lteps <- sum(y==lasty)
  }
  y
}

zc.pnbd.cbs.LL <- function(params, cal.cbs) {

  dc.check.model.params(c("r", "alpha", "s", "beta"), params, "pnbd.cbs.LL")

  tryCatch(x <- cal.cbs[, "x"], error = function(e) stop("Error in pnbd.cbs.LL: cal.
      cbs must have a frequency column labelled \"x\""))
  tryCatch(t.x <- cal.cbs[, "t.x"], error = function(e) stop("Error in pnbd.cbs.LL:
      cal.cbs must have a recency column labelled \"t.x\""))
  tryCatch(T.cal <- cal.cbs[, "T.cal"], error = function(e) stop("Error in pnbd.cbs.
      LL: cal.cbs must have a column for length of time observed labelled \"T.cal\"")
      )

  if ("custs" %in% colnames(cal.cbs)) {
    custs <- cal.cbs[, "custs"]
  } else {
    custs <- rep(1, length(x))
  }
  return(sum(custs * zehra.pnbd.LL(params, x, t.x, T.cal)))
}


zc.pnbd.LL= function (params, x, t.x, T.cal)
{
  max.length <- max(length(x), length(t.x), length(T.cal))
  if (max.length%%length(x))
    warning("Maximum vector length not a multiple of the length of x")
  if (max.length%%length(t.x))
    warning("Maximum vector length not a multiple of the length of t.x")
  if (max.length%%length(T.cal))
    warning("Maximum vector length not a multiple of the length of T.cal")
  dc.check.model.params(c("r", "alpha", "s", "beta"), params,
                        "pnbd.LL")
  if (any(x < 0) || !is.numeric(x))
    stop("x must be numeric and may not contain negative numbers.")
```

```
  if (any(t.x < 0) || !is.numeric(t.x))
    stop("t.x must be numeric and may not contain negative numbers.")
  if (any(T.cal < 0) || !is.numeric(T.cal))
    stop("T.cal must be numeric and may not contain negative numbers.")
  x <- rep(x, length.out = max.length)
  t.x <- rep(t.x, length.out = max.length)
  T.cal <- rep(T.cal, length.out = max.length)
  r <- params[1]
  alpha <- params[2]
  s <- params[3]
  beta <- params[4]
  maxab <- max(alpha, beta)
  absab <- abs(alpha - beta)
  param2 <- s + 1
  if (alpha < beta) {
    param2 <- r + x
  }
  part1 <- r * log(alpha) + s * log(beta) - lgamma(r) + lgamma(r + x)
  part2 <- -(r + x) * log(alpha + T.cal) - s * log(beta + T.cal)
  if (absab == 0) {
    part2_times_F1_min_F2 <- ( (alpha+T.cal)/(maxab+t.x) )^(r+x) * (beta+T.cal)^s /
      ((maxab+t.x)^s) -
      ( (alpha+T.cal)/(maxab+T.cal) )^(r+x) * (beta+T.cal)^s /
      ((maxab+t.x)^s)
  }
  else {
    part2_times_F1_min_F2 = h2f1(r+s+x, param2, r+s+x+1, absab / (maxab+t.x)) *
      ( (alpha+T.cal)/(maxab+t.x) )^(r+x) * (beta+T.cal)^s /
      ((maxab+t.x)^s) -
      h2f1(r+s+x, param2, r+s+x+1, absab / (maxab+T.cal)) *
      ( (alpha+T.cal)/(maxab+T.cal) )^(r+x) * (beta+T.cal)^s /
      ((maxab+t.x)^s)
  }
  return(part1 + part2 + log(1+(s/(r+s+x))*part2_times_F1_min_F2) )
}

zc_pnbd.EstimateParameters <- function(cal.cbs, par.start = c(1, 1, 1, 1),
                                      max.param.value = 10000) {

  dc.check.model.params(c("r", "alpha", "s", "beta"), par.start, "zehra.
      EstimateParameters")
```

```
  ## helper function to be optimized
  pnbd.eLL <- function(params, cal.cbs, max.param.value) {
    params <- exp(params)
    #params[params > max.param.value] <- max.param.value
    return(-1 * zc.pnbd.cbs.LL(params, cal.cbs))
  }
  logparams <- log(par.start)

  # results <- optim(logparams, pnbd.eLL, cal.cbs = cal.cbs, max.param.value = max.
      param.value, method = "L-BFGS-B")
  results <- optim(logparams, pnbd.eLL, cal.cbs = cal.cbs, max.param.value = max.
      param.value, method = "L-BFGS-B", control=list(trace = TRUE))

  estimated.params <- exp(results$par)
  #estimated.params[estimated.params > max.param.value] <- max.param.value
  return(estimated.params)

}

zc_ga.EstimateParameters <- function(cal.cbs, par.start = c(1, 1, 1, 1),
                                       max.param.value = 10000) {

  dc.check.model.params(c("r", "alpha", "s", "beta"), par.start, "zehra.
      EstimateParameters")

  ## helper function to be optimized
  pnbd.eLL <- function(params, cal.cbs) {
    params <- exp(params)
    #params[params > max.param.value] <- max.param.value
    return(-1 * pnbd.cbs.LL(params, cal.cbs))
  }
  logparams <- log(par.start)

  lst_CV_data <- cal.cbs
  theta_min <- c(r = 0.1, alpha = 0.1, s = 0.1, beta = 0.1)
  theta_max <- c(r = 9000, alpha = 9000, s = 9000, beta = 9000)

  theta_min <-log(theta_min)
  theta_max <-log(theta_max)

```

```
  results <- ga(type = "real-valued", fitness = pnbd.eLL, lst_CV_data,
                names = c("r", "alpha", "s", "beta"),
                min = theta_min, max = theta_max,
                popSize = 20, maxiter = 1000)

  estimated.params <- exp(results@solution)

  #estimated.params[estimated.params > max.param.value] <- max.param.value
  return(estimated.params)

}

###########################################
# LOAD DATA
###########################################
refill_file <- file.choose()
refill_file_name <- substr(basename(refill_file), 1, str_locate(basename(refill_file), ".CSV")[1]-1)

L <- readLines(refill_file, n = 1)
if (grepl(",", L))
{
  elog  <- read.csv(refill_file)
} else
{
  print("Choose correct file format")
}

elog<-elog[,c(1,2,4)] # we need these columns ; customer id, transaction date, refill count
names(elog) <- c("cust","date","sales") # model functions expect these names

# format date
elog$date <- as.Date(as.character(elog$date), format="%Y%m%d")

###########################################
# DIVIDE DATA
###########################################
# into a calibration phase and a holdout phase , determine middle point for splitting
(end.of.cal.period <-min(elog$date)+as.numeric((max(elog$date)-min(elog$date))/2))

```

```r
# split data into train(calibration) and test (holdout) and make matrices
data <- dc.ElogToCbsCbt(elog, per="week", T.cal=end.of.cal.period, merge.same.date=TRUE, statistic = "freq")   # which CBT to return

# extract calibration matrix
cal2.cbs <- as.matrix(data[[1]][[1]])

##########################################
# ESTIMATE PARAMETERS FOR MODEL
##########################################
#(params2 <- pnbd.EstimateParameters(cal2.cbs))
(params2 <- zc_pnbd.EstimateParameters(cal2.cbs))

# look at log likelihood
(LL <- pnbd.cbs.LL(params2, cal2.cbs))

##########################################
# HOW WELL DOES MODEL DO IN HOLDOUT PERIOD?
##########################################

# get holdout transactions from dataframe data, add in as x.star
x.star   <- data[[2]][[2]][,1]
cal2.cbs <- cbind(cal2.cbs, x.star)

holdoutdates <- attributes(data[[2]][[1]])[[2]][[2]]
holdoutlength <- round(as.numeric(max(as.Date(holdoutdates))-min(as.Date(holdoutdates)))/7)

# plot predicted vs actual by week

# get data without first transaction, this removes those who buy 1x
removedFirst.elog <- dc.SplitUpElogForRepeatTrans(elog)$repeat.trans.elog
removedFirst.cbt <- dc.CreateFreqCBT(removedFirst.elog)

# get all data, so we have customers who buy 1x
allCust.cbt <- dc.CreateFreqCBT(elog)

# add 1x customers into matrix
tot.cbt <- dc.MergeCustomers(data.correct=allCust.cbt, data.to.correct=removedFirst.cbt)

```

```
lengthInDays <- as.numeric(max(as.Date(colnames(tot.cbt)))-min(as.Date(colnames(tot.cbt))))
origin <- min(as.Date(colnames(tot.cbt)))

tot.cbt.df <- melt(tot.cbt,varnames=c("cust","date"),value.name="Freq")
tot.cbt.df$date<-as.Date(tot.cbt.df$date)
tot.cbt.df$week<-as.numeric(1+floor((tot.cbt.df$date-origin+1)/7))

transactByWeek <- ddply(tot.cbt.df,.(week),summarize,sum(Freq))
names(transactByWeek) <- c("week","Transactions")

T.cal <- cal2.cbs[,"T.cal"]
T.tot <- round(as.numeric(max(elog$date)-min(elog$date))/7)

comparisonByWeek <- pnbd.PlotTrackingInc(params2, T.cal,T.tot, actual.inc.tracking.data=transactByWeek$Transactions)

###########################################
#GET PALIVES AND CONFUSSION MATRIX
###########################################
p.alives <- pnbd.PAlive(params2, cal2.cbs[,"x"], cal2.cbs[,"t.x"], cal2.cbs[,"T.cal"])
ggplot(as.data.frame(p.alives),aes(x=p.alives))+ geom_histogram(colour="grey",fill="gray")+ylab("Number of Customers")+xlab("Probability Customer is 'Live'")+theme_minimal()
df.prob.alives <- data.frame(names(p.alives), as.vector(p.alives))
names(df.prob.alives) <-c("subscriberid","probability")
write.xlsx(df.prob.alives, paste(refill_file_name, sep = "", "_PROB_OF_ALIVES.xlsx"))

###########################################
#SPEND - MONETARY
###########################################
elog_cal <- elog[which(elog[,"date"] < end.of.cal.period),]
elog_cal$Week <- as.Date(cut(elog_cal$date,breaks = "week",start.on.monday = TRUE))
cu.tracking <- aggregate(elog_cal$cust, by=list(elog_cal$Week), FUN=length) #ddply(elog, .(elog$Week), transform, Cumulative.Sum = cumsum(sales))
est.params <- params2
m.x <- aggregate(elog_cal$sales, list(elog_cal$cust), mean)
cdnowSummary <- list(cal2.cbs, as.vector(cu.tracking[][2]), as.vector(est.params), as.vector(m.x[][2]))
m.x.vector <-as.vector(m.x[][2])
```

```
names(m.x.vector) <- c("m.x")
calCBS <- cbind(m.x.vector,cal2.cbs)
names(cdnowSummary) <- c("cbs","cu.tracking","est.params","m.x")
#str(cdnowSummary)
ave.spend <- cdnowSummary$m.x
tot.trans <- cdnowSummary$cbs[,"x"]

# Now we can estimate model parameters. spend.LL, which is used
# by spend.EstimateParameters, will give you warning if you pass
# it a customer with zero transactions. To avoid this, we can
# remove customers with zero transactions
ave.spend <- ave.spend[which(tot.trans > 0),]
tot.trans <- tot.trans[which(tot.trans > 0)]
params_spend <- spend.EstimateParameters(ave.spend, tot.trans)
spend.cal.expected <- spend.expected.value(params_spend, calCBS[,"m.x"], calCBS[,"x"]) #-> abone bazinda matrix ya da dataframe olustur
names(spend.cal.expected) <- c("expected_value")
calCBS.expected.value <- cbind(calCBS, expected.value=as.vector(spend.cal.expected))
df.expected.value <- data.frame(calCBS.expected.value)
write.xlsx(df.expected.value, paste0(refill_file_name, "_EXOECTEDVALUE.xlsx"))
spend.plot.average.transaction.value(params_spend, ave.spend, tot.trans)

# Uses Pareto/NBD model parameters and a customer's past transaction behavior to return the number
# of transactions they are expected to make in a given time period.
# pnbd.ConditionalExpectedTransactions(params, T.star, x, t.x, T.cal)
# T.star <- holdoutlength
# predict<-pnbd.ConditionalExpectedTransactions(params2, T.star = holdoutlength, x = cal2.cbs[,"x"], t.x   = cal2.cbs[,"t.x"], T.cal = cal2.cbs[,"T.cal"])

# pnbd.DERT Pareto/NBD Discounted Expected Residual Transactions
# Calculates the discounted expected residual transactions of a customer, given their behavior during the calibration period.
# pnbd.DERT(params, x, t.x, T.cal, d)

############################################
# DATA EXPLORATION
############################################
# plot predicted vs seen conditional freqs and get matrix 'comp' w values
T.star <- holdoutlength
censor <- 10 # Bin all order numbers here and above
```

```
comp <- pnbd.PlotFreqVsConditionalExpectedFrequency(params2, T.star,cal2.cbs, x.star,
     censor)
rownames(comp) <- c("act", "exp", "bin")
comp


##########################################
# PLOT GROUP DISTRIBUTION OF PROPENSITY TO PURCHASE, DROPOUT
##########################################
# par to make two plots side by side
par(mfrow=c(1,2))
# Plot the estimated distribution of lambda
# (customers' propensities to purchase)
pnbd.PlotTransactionRateHeterogeneity(params2, lim = NULL)

# lim is upper xlim
# Plot estimated distribution of gamma
# (customers' propensities to drop out).
pnbd.PlotDropoutRateHeterogeneity(params2)
# set par to normal
par(mfrow=c(1,1))


# look at days between orders
# model describes rates via a gamma distribution across customers
purchaseFreq <- ddply(elog, .(cust), summarize, daysBetween = as.numeric(diff(date)))
ggplot(purchaseFreq,aes(x=daysBetween))+geom_histogram(fill="darkgrey")+xlab("Time
    between purchases (days)")+theme_minimal()
```

APPENDIX B

GAMS CODE FOR MATHEMATICAL MODEL

```
option optcr = 0.00001;
Sets
i customer
j offer

$call GDXXRW offer.xlsx dset=i rng=main!a2 rdim=1 dset=j rng=main!c2 rdim=1"
$gdxin offer.gdx
$load i j
$gdxin

display j;

Parameters
clv_base(i)  The customer lifetime value of the customer i
p(i) The churn probability of the customer i
f(i) Monthly payment of customer i to company before taking the offer
o(i)  The upper bound of offer monetary value for customer i
clv(i,j)  The customer lifetime value of the customer i
bb(i,j)
pp(i,j)
c(j)  The monetary value of offer j
;

execute "GDXXRW offer.xlsx par=c rng=main!as2 Rdim=1  par=clv_base rng=main!K2 Rdim=1 par=p rng=main!E2 Rdim=1 par=f rng=main!H2 Rdim=1 par=o rng=main!ap2 Rdim=1 par=b rng=main!aj2 Rdim=1 par=n rng=main!am2 Rdim=1"
execute_load 'offer.gdx' c  clv_base p f o
display c  ,clv_base, p, f, o;

execute "GDXXRW offer.xlsx par=bb rng=bb!a1 Rdim=1 Cdim=1 "
execute_load 'offer.gdx'  bb

execute "GDXXRW offer.xlsx par=pp rng=pp!a1 Rdim=1 Cdim=1 "
execute_load 'offer.gdx'  pp
```

```
execute "GDXXRW offer.xlsx par=clv rng=clv!a1 Rdim=1 Cdim=1 "
execute_load 'offer.gdx'  clv

display bb,pp,clv;

Scalar
t Monthly budget for offers/
10000/;

Variable z;

Binary Variable
x(i,j) If customer i get offer j

Equations
obj
c3
c4(i)
c5(j);

obj.. z =e= sum(i, sum(j,x(i,j)*(bb(i,j)*clv(i,j)+(1-bb(i,j))*clv_base(i)))+(1-sum(j,
    x(i,j)))*clv_base(i));
c3.. sum ((i,j), f(i)*x(i,j)*c(j)) =l= t;
c4(i).. sum (j, f(i)*x(i,j)*c(j)) =l= o(i) ;
c5(j).. sum (i, x(i,j)) =l= 300;

model offer_management /all/;
Solve offer_management using mip maximizing z;

file out2 /offer_ham-ek2.txt/;
put out2;
loop((i,j), put$(x.l(i,j)) i.tl j.tl/);
```

APPENDIX C

FEATURES FOR LOGISTIC REGRESSION

1. Subscriber Activation Reason
2. Counter Slope in the last six months (Usage + Fee counters)
3. Trend of usage + fee credits used in last 6 months
4. Variance of credits used in last 6 months
5. End date of subscription
6. End reason of subscribtion
7. Average number of dropped calls for last six months
8. Average number of dropped calls for last month
9. Last package date
10. Last tariff group change year month
11. Minimum packet fill date
12. Minimum packet fill year month
13. Minimum sms fill date
14. Minimum sms fill year month
15. Last MNP port in year month
16. Last Portout Year Month
17. End Date of latest acquisition given in the last month
18. Start Date of latest acquisition given in the last month
19. Last activation date
20. Last Activation year month of subscriber
21. Contract Signed Year Month
22. First activation date
23. VAS Package purchase year month
24. Last month line count less than 2
25. OMO CC Incoming call flag
26. OMO CC outgoing call flag
27. ONNET CC call flag
28. Flag for ONNET prefix
29. 4g flag
30. Gnctrkcll flag
31. Last equipment 4g support flag
32. Last Equipment Mobil TV Flag
33. Last Equipment screen dimensions
34. Last equipment 3G flag

35 Youth segment flag
36 Subscriber Id
37 Customer Id
38 Average call revenue (usage + fee) for last six months
39 Average number of incoming different B number(voice) for last six months
40 Average number of incoming different onnet B number(voice) for last six months
41 Average number of incoming voice calls for the last six months
42 Average Incm Voice MOU for last six months for last six months
43 Average Night MOU for last six months
44 Average Offnet Night MOU for last six months
45 Average Onnet Night MOU for last six months
46 Average Not Roaming Incm Not Vas Mms OnnetCALL for last six months
47 Not roaming incoming not vas voice OMO MoU in the last month
48 Not roaming incoming not vas voice OnnetMoU in the last month
49 Last month incoming TT mou
50 Last month Not roaming Incm Not Vas SMS OnnetCall
51 Average Not roaming Incm Not Vas SMS OnnetCall for last six month
52 Average Not roaming Incm Not Vas Voice OnnetCall for last six month
53 Average Not Roaming Incm Sms other operator 1 CALL for last six months
54 Average Not Roaming Incm SMS CALL for last six months
55 Average Not Roaming Incm Sms Int CALL for last six months
56 Average Not Roaming Incm Not Vas Sms Nat CALL for last six months
57 Average Not Roaming Incm Sms Omo CALL for last six months
58 Average Not Roaming Incm Sms other operator 2 CALL for last six months
59 Average Not Roaming Incm Voice other operator 1 CALL for last six months
60 Average Not Roaming Incm Voice other operator 1 MOU for last six months
61 Average Not Roaming Incm Voice CALL for last six months
62 Average Not Roaming Incm Voice International CALL for last six months
63 Average Not Roaming Incm Voice National CALL for last six months
64 Average Not Roaming Incm Voice Omo CALL for last six months
65 Average Not Roaming Incm Voice Onnet CALL for last six months
66 Average Not Roaming Incm Voice other operator 2 CALL for last six months
67 Average Not Roaming Incm Voice other operator 2 MOU for last six months
68 Average Not Roaming Incm Voice CALL for last six months
69 Average Not Roaming Incm Voice Int MOU for last six months
70 Average Not Roaming Incm Voice MOU for last six months
71 Average Not Roaming Incm Not Vas Voice Nat MOU for last six months
72 Average Not Roaming Incm Voice Omo MOU for last six months
73 Average Not Roaming Incm Voice TT MOU for last six months
74 Average Not Roaming Incm VAS SMS Call for last six months
75 Last month Not Roaming Not VAS Incm OMO SMS Call

76 Total not roaming incoming not vas SMSOMO calls in the last month

77 Last month outgoing OMO mou

78 Not Roaming Outgoing Not Vas Last month outgoing Onnetmou

79 Last month outgoing TT mou

80 Average Not roaming Outg Not Vas SMS Yurtici Call for last six months

81 Last month Not roaming Outg Not Vas SMS OnnetCall

82 Average Not roaming Outg Not Vas SMS OnnetCall for last six months

83 Average Not roaming Outg Not Vas Voice Yurtici Call for last six months

84 Average Not roaming Outg Not Vas Voice Yurtici MOU for last six months

85 Average Not roaming Outg Not Vas Voice OnnetCall for last six months

86 Average Not roaming Outg Not Vas Voice OnnetMOU for last six months

87 Average Not Roaming Outg Not Vas Voice CALL for last six months

88 Average Not Roaming Outg Not Vas Voice MOU for last six months

89 Average not roaming outgoing SMS other operator 1 calls in the last six months

90 Average Not Roaming Outg SMS Call for last six months

91 Average Not Roaming Outg SMS Int Call for last six months

92 Average Not Roaming Outg SMS Omo Call for last six months

93 Average not roaming outgoing SMS VDF calls in the last six months

94 Average Not Roaming Outg Voice other operator 1 Call for last six months

95 Average Not Roaming Outg Voice other operator 1 MOU for last six months

96 Average Not Roaming Outg Voice Int Call for last six months

97 Average Not Roaming Outg Voice Int MOU for last six months

98 Average Not Roaming Outg Voice OMO Call for last six months

99 Average Not Roaming Outg Voice Onnet Call for last six months

100 Average Not Roaming Outg Voice Onnet MOU for last six months

101 Average Not Roaming Outg Voice other operator 2 CALL for last six months

102 Average Not Roaming Outg Voice other operator 2 MOU for last six months

103 Average Not Roaming Outg Voice Omo MOU for last six months

104 Average Not Roaming Outg Voice TT CALL for last six months

105 Average Not Roaming Outg Voice TT MOU for last six months

106 Average not roaming outgoing VAS SMS calls in the last six months

107 Last 6 months sent and received sms average

108 Last month Number of call forwarding

109 Number of call forwarding for last 3 months

110 Monthly Distinct Cell City Count for last three months

111 Monthly Distinct Cell Site Count for last three months

112 Total number of distinct top3 call for the last six months

113 Average number of Night Sms Calls for last six months

114 Average number of Offnet Night Sms Calls for last six months

115 Average number of Onnet Night Sms Calls for last six months

116 Number of tariff change in last 6 months

117 Sum of calls made to top1 OMO number for last six moths
118 Last month Call Revenue excluding Package Fee for all call types
119 Average number of outgoing different B number(voice) for last six months
120 Average number of outgoing different OnnetB number(voice) for last six months
121 Average total top1 mou for last six months
122 Average total top1 calls for last six months
123 Average total top(1+2+3)mou for last six months
124 Average total top(1+2+3) call count for last six months
125 Total Amount of Discounted MOU for the last six months
126 Total number of discounted sms usage for the last six months
127 Average Week End MOU for last six months
128 Days since last CHURN campaign ACCEPTED
129 Days since last offer campaign
130 GPRS package count in last month
131 Number of ACCEPTED churn campaigns for last 3 months
132 Number of ACCEPTED campaigns for last 3 months
133 Utilization of data package in last 3 months
134 Utilization of sms package in last 3 months
135 Utilization of voice package in last 3 months
136 Contracted Packeg Fee in last month
137 Voice package count in last month
138 Number of MOC Status Changes for last twelve months
139 Number Of Postpaid Lines (excluding deactive lines)
140 Number Of Prepaid Lines (excluding deactivated lines)
141 Total number of status change in the last twelve months
142 Number of lines for the customer (excluding deactivated lines)
143 Incoming TT count in last month
144 Outgoing TT count in last month
145 Incoming TT vo ce count in last month
146 Outgoing TT voice count in last month
147 Data cell count in last 3 month
148 Change in postpaid line count
149 Change in prepaid line count
150 Number of SRs for the last month
151 Number of SRs for six month
152 Discounted Mou Average for last six months
153 Last 6 month Average Amount of dedicated free counter earned
154 Average amount of package sms calls for the last six months
155 Total Number of SMS PACKAGE Calls for last 6 months
156 Total number of refills excluding refills(event type refilland konbara) for last six moths

```
157 Total Number of Voice PAKET Mou for last 6 months
158 Web usage count in last 3 month
159 Web last usage in last 3 month
160 Total number of port-ins
161 Total number of port-outs
162 Total OnnetTenure for port-in customers
163 Top 3 incoming call subscriber port count in last month
164 Top 3 outgoing call subscriber port count in last month
165 Tenure since the last portin
166 Total amount of refills for last six months
167 Average of Daily Balance for the last six months
168 Standard Deviation of SDP counter balance for last six months
169 Total count of negative SDP balance days for last six months
170 Advanced credit usage in last 6 months
171 Average number of days between refills
172 Maximum number of days between refills
173 Minimum number of days between refills
174 Standart deviation of day count between refills
175 Number of days since last refill
176 Total number of refills for last six months
177 Sum of prepaid package purchases made for the last six months
178 Count of NAR package in last 6 months
179 Count of other package in last 6 months
180 Average data utilization
181 Maximum data utilization
182 Minimum data utilization
183 Average of other package utilization in last 6 months
184 Average voice utilization
185 Maximum voice utilization
186 Minimum voice utilization
187 Average SMS utilization
188 Maximum SMS utilization
189 Minimum SMS utilization
190 Refill count from Shop in last 6 months
191 Refill count other than POS/Shop/Bank channel in last 6 months
192 Average utilization
193 SNA Average offnet subscriber count in last six months
194 SNA Change in offnet subscriber count in last six months
195 SNA Standart deviation of offnet subscriber count in last six months
196 SNA Trend of offnet subscriber in last six months
197 SNA Average onnet subscriber count in last sisx month
```

198 SNA Change in onnet subscriber count in last six months
199 SNA Standart deviation of onnet subscriber in last six months
200 SNA Trend of onnet subscriber in last six months
201 SNA Average postpaid subscriber count in last six months
202 SNA Change in postpaid subscriber count in last six months
203 SNA Standart deviation of postpaid subscriber count in last six months
204 SNA Trend of postpaid subscribers
205 SNA Average prepaid subscriber count in last six months
206 SNA Change in prepaid subscriber count in last six months
207 SNA Standart deviation of prepaid subscriber count in last six months
208 SNA Trend of prepaid subscribers
209 SNA upsell count in last month
210 Last SNA Role of the subscriber
211 Last acquisition sum in last month
212 Last acquisition usage in last month
213 Age
214 ARPU Average for last six months
215 ARPU Slope for last six months
216 ARPU trend in the last six months
217 Call Center Cost
218 Usage + Fee Revenue Slope for last six months
219 Usage + fee revenue trend in the last six months
220 Total number of CSI logins for last six months
221 Num of actions made on CSIMI in last 6 months
222 Num of actions made on CSITS in last 6 months
223 Num of actions made on CSIWS in last 6 months
224 Day since last equipment changed
225 Customer Small Business flag
226 Final Cost
227 Final Profit
228 Interconnect Cost
229 Network Cost
230 Number of different equipment used in 6 months
231 Outgoing MOU Slope for last six months
232 Outgoing MOU Trend for last six months
233 Outgoing different B number(voice)call Slope for last 6 months
234 Outgoing diffB number trend in the last six months
235 Subscriber General ARMU for previous month
236 Soho Club flag
237 Tenure in Months (Months after last portin for portin subscribers months after last activation for others)

```
238 Tax revenue
239 Treasure Cost
240 3g data usage ratio in last month
241 Data usage between 08:00 - 19:00 (MB)
242 Average Data Volume
243 Standart deviation of Data Volume
244 Data Volume Standard Deviation
245 Data Volume Trend
246 Data TL Average
247 Data TL Change - StandardDeviation/Mean
248 Data TL Standard Deviation
249 Data TL Trend
250 Data usage between 06:00 - 08:00 (MB)
251 Incoming MOU Average
252 Incoming MOU Change - StandardDeviation/Mean
253 Incoming MOU Standard Deviation
254 Incoming MOU Trend
255 Incoming OMO MOU Average
256 Incoming OMO MOU Change - StandardDeviation/Mean
257 Incoming OMO MOU Standart Deviation
258 Incoming OMO MOU Trend
259 Incoming Onnet MOU Average
260 Incoming Onnet MOU Change - StandardDeviation/Mean
261 Incoming Onnet MOU Standart Deviation
262 Incoming Onnet MOU Trend
263 Incoming SMS Count Change - StandardDeviation/Mean
264 Incoming SMS Trend
265 Incoming Voice Count Change - StandardDeviation/Mean
266 Incoming Voice Count - StandardDeviation
267 Incoming Voice Count Averge
268 Incoming Voice Count Trend
269 Last six months usage days
270 Data usage between 23:00 - 07:00 (MB)
271 Data usage between 19:00 - 23:00 (MB)
272 Charged Outgoing Voice TL Average
273 Charged Outgoing Voice TL Change - StandardDeviation/Mean
274 Charged Outgoing Voice TL - StandardDeviation
275 Charged Outgoing Voice TL Trend
276 Discounted Outgoing Voice TL Average
277 Discounted Outgoing Voice TL Change - StandardDeviation/Mean
278 Discounted Outgoing Voice TL - StandardDeviation/Mean
```

279 Discounted Outgoing Voice TL Trend
280 Last month last call date difference
281 Last month last data usage date difference
282 Outrgoing MOU Average
283 Outgoing MOU Change - StandardDeviation/Mean
284 Outrgoing MOU - StandardDeviation
285 Outgoing MOU Trend
286 Outgoing OMO MOU Average
287 Outgoing OMO MOU Change - StandardDeviation/Mean
288 Outgoing OMO MOU - StandardDeviation
289 Outgoing OMO MOU Trend
290 Outgoing OMO Voice Count Average
291 Outgoing OMO Voice Count Change - StandardDeviation/Mean
292 Outgoing OMO Voice Count - StandardDeviation
293 Outgoing OMO Voice Count Trend
294 Outgoing OMO Voice TL Average
295 Outgoing OMO Voice TL Change - StandardDeviation/Mean
296 Outgoing OMO Voice TL - StandardDeviation
297 Outgoing OMO Voice TL Trend
298 Outgoing Onnet MOU Average
299 Outgoing Onnet MOU Change - StandardDeviation/Mean
300 Outgoing Onnet MOU - StandardDeviation
301 Outgoing Onnet MOU Trend
302 Outgoing Onnet Voice Count Average
303 Outgoing Onnet Voice Count Change - StandardDeviation/Mean
304 Outgoing Onnet Voice Count Average
305 Outgoing Onnet Voice Count Trend
306 Outgoing Onnet Voice TL Average
307 Outgoing Onnet Voice TL Change - StandardDeviation/Mean
308 Outgoing Onnet Voice TL - StandardDeviation
309 Outgoing Onnet Voice TL Trend
310 Outgoing SMS Change - StandardDeviation/Mean
311 Outgoing SMS Trend
312 Outgoing Voice Count Average
313 Outgoing Voice Count Change - StandardDeviation/Mean
314 Outgoing Voice Count - StandardDeviation
315 Outgoing Voice Count Trend
316 Outgoing Voice TL Average
317 Outgoing Not Roaming Voice TL Change - StandardDeviation/Mean
318 Outgoing Voice TL - StandardDeviation
319 Outgoing Not Roaming Voice TL Trend

320 Ratio of incoming and outgoing call
321 Other operator 1 incoming call ratio in all calls
322 Onnet incoming call ratio in all calls
323 Other operator 2 incoming call ratio in all calls
324 Other operator 1 outgoing call ratio in all calls
325 Onnet outgoing call ratio in all calls
326 Other operator 2 outgoing call ratio in all calls
327 Total call in last month
328 Total outgoing call to operator 1 in last month
329 Total voice package excess in last month
330 Total outgoing call to operator 2 in last month
331 Vas 3g Tenure
332 Data usage KB in the last month
333 Average volume of data used for last six months
334 VAS entertainment package lus for last month
335 VAS entertainment package maximum for last month
336 VAS entertainment package rec for last month
337 VAS entertainment package slope for last month
338 VAS Total amount of incoming credit transfer for last three months
339 VAS Prepaid SMS package in last 12 months
340 VAS Social media uasge in last 12 months
341 VAS Subscription count
342 VAS mobile apps maximum usage in last 12 months
343 VAS mobile apps record usage in last 12 months
344 VAS mobile apps slope usage in last 12 months
345 VAS Total TL revenue from MMS usage for last 3 months
346 VAS Total TL revenue from SMS usage for last 3 months
347 VAS Usage count for last 3 months
348 Last economic package
349 Last shared package
350 Most frequent Cell City for last month
351 Current Tariff Group
352 Previous Tariff Group
353 Subscriber to benefit based upon the most recent company name for the six months
354 Operating System of the equipmnent
355 Contract Type
356 Most frequent Data Cell City
357 Last Port-IN Operator
358 Last portin operator
359 Last Port-In Reason
360 Last portout operator

```
361 Last portout reason
362 Average refill card unit
363 Last refill card unit
364 Maximum amount of refill card unit
365 Minimum amount of refill card unit
366 Most frequently refilled card unit
367 Standard Deviation of amounts of refill card unit
368 Activation Status Reason of the Subscriber
369 Activation Reason of the Subscriber
370 Bill City for postpaid
371 Activation dealer city
372 Gender
373 Subscriber Last Equipment Manufacturer Group
374 Last Equipment Marketing Model
375 Current Simcard Type
376 Subscriber Subscription system (Postpaid/Prepaid)
377 Last 3 monhts PSTN usage segmentation
378 Last month End Of Month Contract status
379 General Segment
380 General Sub Segment
381 Last equipment category
382 Number of incoming other operator 1 different B numbers for last month
383 Number of incoming other operator 1 different B numbers for previous month
384 Number of incoming Onnet different B numbers for last month
385 Number of incoming Onnet different B numbers for previous month
386 Number of outgoing other operator 1 different B numbers for last month
387 Number of outgoing other operator 1 different B numbers for previous month
388 Number of outgoing Onnet different B numbers for last month
389 Number of outgoing Onnet different B numbers for previous month
390 Number of incoming other operator 2 different B numbers for last month
391 Number of incoming other operator 2 different B numbers for previous month
392 Number of outgoing Telsim different B numbers for last month
393 Number of outgoing other operator 2 different B numbers for previous month
394 Last Tariff Duration
395 Year Month of the Variable pool data
```

Bibliography

[1] V. Kumar and B. Rajan, "Profitable customer management: Measuring and maximizing customer lifetime value," *Management Accounting Quarterly*, vol. 10, no. 3, pp. 1–18, 2009.

[2] Z. Can and E. Albey, "Churn prediction for mobile prepaid subscribers," in *Proceedings of the 6th International Conference on Data Science, Technology and Applications*, pp. 67–74, 2017.

[3] "Türkiye elektronik haberleşme sektörü, üç aylık pazar verileri raporu." https://www.btk.gov.tr/tr-TR/Sayfalar/Pazar-Verileri, 2017. 2017 Q2.

[4] "Turkcell Q1 2015 IFRS Interim Report." http://s1.turkcell.com.tr/hakkimizda/en/yatirimciiliskileri/InvestorReportLibrary/IFRS-Report-Q12015.pdf, 2015.

[5] P. S. Fader, B. G. Hardie, and K. L. Lee, "Rfm and clv: Using iso-value curves for customer base analysis," *Journal of Marketing Research*, vol. 42, no. 4, pp. 415–430, 2005.

[6] D. C. Schmittlein, D. G. Morrison, and R. Colombo, "Counting your customers: Who they are and what will they do next?," *Management Science*, vol. 33, no. 1, pp. 1–24, 1987.

[7] P. E. Pfeifer, M. E. Haskins, and R. M. Conroy, "Customer lifetime value, customer profitability, and the treatment of acquisition spending," *Journal of Managerial Issues*, vol. 17, no. 1, pp. 11–25, 2005.

[8] H.-S. Hwang, "A dynamic model for valuing customers: A case study," *Advanced Science and Technology Letters*, vol. 120, no. 10, pp. 56–61, 2015.

[9] D. Birant, "Data mining using rfm analysis," *Knowledge-Oriented Applications in Data Mining*, vol. 154, no. 1, pp. 91–108, 2011.

[10] P. S. Fader, B. G. Hardie, and K. L. Lee, "A note on deriving the pareto/nbd model and related expressions." http://brucehardie.com/notes/009/. 2005.

[11] P. S. Fader, B. G. Hardie, and K. L. Lee, "A note on implementing the pareto/nbd model in matlab." http://brucehardie.com/notes/008/. 2005a.

[12] A. O. Dairo and T. Akinwumi, "Dormancy prediction model in a prepaid predominant mobile market: A customer value management approach," *International Journal of Data Mining & Knowledge Management Process*, vol. 4, no. 1, pp. 33–39, 2014.

[13] M. Owczarczuk, "Churn models for prepaid customers in the cellular telecommunication industry using large data marts," *Expert Syst. Appl.*, vol. 37, pp. 4710–4712, 2010.

[14] M. R. Khan, J. Manoj, A. Singh, and J. Blumenstock, "Behavioral modeling for churn prediction: Early indicators and accurate predictors of custom defection and loyalty," in *Proceedings of the 2015 IEEE International Congress on Big Data*, BIGDATACONGRESS '15, (Washington, DC, USA), pp. 677–680, IEEE Computer Society, 2015.

[15] C. K. Kirui, L. Hong, W. K. Cheruiyot, and H. Kirui, "Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining," in *International Journal of Computer Science Issues (IJCSI)*, vol. 10, p. 165, 2013.

[16] J. Lu, "Predicting customer churn in the telecommunications industry - an application of survival analysis modeling using sas," *SUGI 27*, pp. 114–27, 2002.

[17] J.-H. Ahn, S. P. Han, and Y.-S. Lee, "Customer churn analysis: Churn determinants and mediation effects of partial defection in the korean mobile telecommunications service industry," *Telecommunications Policy*, vol. 30, pp. 552–568, 11 2006.

[18] K. Dahiya and S. Bhatia, "Customer churn analysis in telecom industry," in *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, pp. 1–6, 9 2015.

[19] S. A. Neslin, R. C. Blattberg, and B.-D. Kim, "Rfm analysis," *Database Marketing*, vol. 18, pp. 323–337, 2008.

[20] K. Coussement, F. Van den Bossche, and K. De Bock, "Data accuracy's impact on segmentation performance: Benchmarking rfm analysis, logistic regression, and decision trees," *Journal of Business Research*, vol. 67, p. 27512758, 01 2014.

[21] A. Keramati, R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari, and U. Abbasi, "Improved churn prediction in telecommunication industry using data mining techniques," *Applied Soft Computing*, vol. 24, no. Supplement C, pp. 994 – 1012, 2014.

[22] G. Olle, "A hybrid churn prediction model in mobile telecommunication industry," *International Journal of e-Education, e-Business, e-Management and e-Learning*, 01 2014.

[23] Y. Huang and T. Kechadi, "An effective hybrid learning system for telecommunication churn prediction," *Expert Systems with Applications*, vol. 40, no. 14, pp. 5635 – 5647, 2013.

[24] J.-T. Wei, S.-Y. Lin, and H.-H. Wu, "A review of the application of rfm model," *African Journal of Business Management*, vol. 4, pp. 4199–4206, 12 2010.

[25] T. S. Zabkowski, "Rfm approach for telecom insolvency modeling," *Kybernetes*, vol. 45, pp. 815–827, 05 2016.

[26] B. Huang, T. Kechadi, and B. Buckley, "Customer churn prediction in telecommunications," *Expert Systems with Applications: An International Journal*, vol. 39, pp. 1414–1425, 01 2012.

[27] J.-R. Segarra-Moliner and M. Moliner, "Customer equity and clv in spanish telecommunication services," *Journal of Business Research*, vol. 69, 04 2016.

[28] A. Correa Bahnsen, D. Aouada, and B. Ottersten, "A novel cost-sensitive framework for customer churn predictive modeling," *Decision Analytics*, vol. 5, pp. 1–15, 06 2015.

[29] M. Khajvand, K. Zolfaghar, S. Ashoori, and S. Alizadeh, "Estimating customer lifetime value based on rfm analysis of customer purchase behavior: Case study," *Procedia Computer Science*, vol. 3, 01 2010.

[30] M. Platzer, "Stochastic models of noncontractual consumer relationships," Master's thesis, Vienna University of Economics and Business Adminstration, 2008.

[31] C. Homburg, N. Koschate, and W. Hoyer, "Do satisfied customers really pay more? a study of the relationship between customer satisfaction and willingness to pay," *Journal of Marketing - J MARKETING*, vol. 69, pp. 84–96, 04 2005.

[32] S.-H. Ma and J.-L. Liu, "The mcmc approach for solving the pareto/nbd model and possible extensions," in *Proceedings - Third International Conference on Natural Computation, ICNC 2007*, vol. 2, pp. 505 – 512, 2007.

[33] M. Abe, ""counting your customers" one by one: A hierarchical bayes extension to the pareto/nbd model," *CIRJE, Faculty of Economics, University of Tokyo, CIRJE F-Series*, vol. 28, 01 2008.

[34] M. Platzer and T. Reutterer, "Ticking away the moments: Timing regularity helps to better predict customer activity," *Marketing Science*, vol. 35, 05 2016.