BI Norwegian Business School - campus Oslo

GRA 19703

Master Thesis

Thesis Master of Science

Churn Prediction: How Customers' Usage under Contract is Linked to Churn

Navn: Amanda Cecilie Limseth

Start: 15.01.2020 09.00

Finish: 01.09.2020 12.00

MSc. Business Analytics

Churn Prediction: How Customers' Usage under Contract is Linked to Churn

Supervisor: Rutger van Oest

Summary (i.)

Businesses will always have a certain amount of customers leaving for various reasons, whether it's for another competitor or because their perception of value isn’t met. The concept of churn prediction is based on investigating the pattern of behavior that has shown to lead to customers not renewing their membership in the past, and by identifying similar behavior in the current customer base to find the at-risk customers before they churn.

There exists large variability in the customer usage under contract, both in the total number of visits and in the proportional type of activity attended. The study will use two different statistical approaches to construct a binary model for how usage under contract is linked to churn. Both approaches will differentiate between the usage amount and type of activity under contract to ultimately explain and predict the likelihood of churn in the fitness industry. The results from the binary prediction model indicate that there is a significant difference in the estimated likelihood of churn in the fitness industry based on the customer usage. It found that members with monthly averages over the contract period at the mean or higher is generally less likely to churn and less affected by the change in explanatory variables like activity type, age, and proportional usage in the last months leading up to the renewal decision. The members that showed lower usage over the contract period had churn likelihoods that were considerably more affected by the change in usage variables.

The research will provide a statistical model that can be altered in order to work for any business that has contractual memberships where the usage amount and type of activity is recorded. The model will provide churn prediction probabilities before they occur, which can enable businesses to take targeted action ahead of time and attempt to lower their associated customer attrition by creating tailored promotions to entice the customer to renew their contracts.

Acknowledgment (ii.)

It has been a deep learning curve, but I've enjoyed working on this. I'd like to thank the Business Analytics department at BI Oslo, both Rutger van Oest and Auke Hunneman, for their valuable help and guidance.

I'd also like to thank SATS for allowing me to work with them on this topic.

Table of Content

1. Introduction **1**
1.1. Area of Study and Purpose 1

2. Positioning in the Literature **4**
2.1. Churn Management and Prediction 4
2.2. The shifted-Beta-Geometric Model 5
2.3. Ascarza-Hardie Model 6
2.4. Binary Regression and Classification Tree 7

3. Research Method **9**
3.1. GDPR Compliance 10
3.2. The Dataset 10
3.3. Creation of Dummy (Factor) Variables 11
3.3.1 Predicting Variable: Churn 11
3.4. Frequency of Usage and Type 12
3.5. Descriptive Statistics 12
3.6. Association Between Explanatory Variables and Churn 16
3.7. Outliers 20
3.8. Class Imbalance 20

4. Churn Prediction **21**
4.1. Evaluating the Alternative Models 22
4.1.1. P-values and Regression Coefficients 23
4.1.2. The Marginal Effects 24
4.1.3. AIC Score 24
4.1.4. Residual Plots 25
4.1.5. The Accuracy of the Model 25
4.1.6. Train and Test Set 26
4.2. Baseline Model 26
4.2.1. Confusion Matrix: Baseline 27
4.3. Logit Model 28
4.3.1. Logit Model: Estimated Results 29
4.3.2. Reduced Model: Logit 31
4.3.3. Average Marginal Effect 33
4.3.4. Confusion Matrix: Logit 38
4.4. The Probit Model 39
4.4.1. Probit Model: Estimated Results 40
4.4.2. Reduced Model: Probit 41
4.4.3. Average Marginal Effect 43
4.4.3. Confusion Matrix: Probit 47
4.5. Classification Tree Model 47
4.5.1. Confusion Matrix: Classification Tree 51
4.6. Model Comparison 51
4.6.1. Recommended Model 53

5. General Discussion **53**
5.1. Summary 53
5.2. Implication 55
5.3. Further Improvements 56
5.3.1. Length of Collection period 56
5.3.2. Switching between contracts 56
5.3.3. Cost variables 57

6. References **58**

Appendix I: R code for the Churn Prediction Models

1. Introduction

1.1. Area of Study and Purpose

Competition between businesses is increasingly being fueled by analytical decisions based on the processing of big data. Customer churn, which is when a customer is lost, is no exception. The concept of customer churn is a highly relevant topic as it explains how loyal or possibly dissatisfied the business' customers are. To some extent churn is unavoidable, as all companies experience that a portion of their customers move on, but understanding the type of usage behavior that is linked to churn will give businesses information that can be utilized for better informed decisions for what promotional activities would be relevant for the at-risk customers and ultimately maintain a higher number of customers than if they hadn't taken any action.

The way customer churn has been included in businesses over the years has evolved and the analysis can both be used to target specific promotional activities at the at-risk customers and used to offer a better service or product in the future. There is convincing evidence that data-driven decision-making and big data technologies substantially improve business performance (Provost & Fawcett, 2013, p 17). Analyzing the pattern in previously churned customers and investigating what usage pattern and demographic traits lies behind their decision can be an important factor in finding the right way to entice members to want to remain as members. The members could be leaving to go to another competitor because they're not happy with the services offered, or they could feel like they're not using the membership enough to account for the cost of the membership. Usage is an important factor to involve in the churn prediction as it so highly linked to the value the members get out of the contractual cost.

For most businesses customer churn is a problem because of the high expensive of acquiring new customers. For companies churn doesn't only mean the loss of revenue, but in most cases the cost associated with maintaining a customer relationship is cheaper than acquiring a new one. According to an industry report, it costs nine times as much to acquire a new member as it does to retain an existing one in the fitness industry (Curley, 2019). Once customer churn is a factor, the future value of the inactive customer is lost.

One of the defining characteristics of contractual and subscription-based business models and its related customer churn is that the departure of a given customer is easily observable as they actively have to contact the business in order to cancel their contract for the next period (Fader & Hardie, 2006). In non-contractual settings, i.e. stores, it is more complicated to estimate the number of churned customers as one can't know for sure if the absence of sales after the last purchase is related to something that is not observable in the sales numbers (e.g. moving, injury, vacation, or death). In businesses that operate with contractual memberships, one can take advantage of the easily observable churn and the recorded usage activity within contract to build the prediction model.

It's common among many industries to offer contractual memberships with a set monthly fee where the members are free to use the service as much as needed within the period and where this usage is observed and recorded. This membership method is common in gym memberships, magazine- and newspaper subscriptions, online gaming- and streaming services, and cell phone subscriptions to mention a few. The research aims to show a statistical approach that can easily be altered to be applicable to other industries with contractual memberships.

The fitness industry is a good industry to build the model on as the data available allows for flexibility of usage under contract, and where the renewal decision is easily observable as a binary decision. Many of the players in the fitness industry offer similar services at a comparable price point, and the switching cost is fairly low for the customers, which makes churn a major issue for this industry. In Norway, the players in the fitness industry are competing for the business of an estimated 800,000 members, with a yearly spending of over 5,100,000,000 NOK (Engmo, 2019). In the last 10 years the number of fitness centers in Norway has doubled, which means there is a big variety of prices and types of gyms that competes to provide value to the market. Most of the fitness players offer the customer variations of either monthly or yearly contract options where the customers can use the services as much as they want within the period with no additional cost.

Usage while under contract is an important factor to focus on in estimating the link to churn as it is both a likely crucial predictor, while it is practical in terms of it being data that is readily available in most business' databases already. Within a contract the customer can freely use the service as much as needed without any added costs, but if they don't use the service they still have to pay the monthly fee. The customer's renewal decision is likely strongly driven by the amount and type of usage within the period leading up to the renewal decision. Usage allows the model to take advantage of this time effect, where the observed usage can later lead to the churn decision. The research will investigate the link between the customer usage of the contractual service, i.e. the amount and type of gym activity within a gym contract and the related churn, in order to gain insight into what type of customer is more likely to churn when the usage goes down.

Several models exists specifically to predict churn in contractual settings. The established churn prediction models of statisticians Fader-Hardie (2007) and Ascarza-Hardie (2009) offer two different approaches for returning accurate churn predictions for contractual customers. The present study aims to offer an alternative statistical approach that differs from these already established models, and will utilize two different techniques in search of the best fitting prediction model for churn: binary regression and classification tree. Both techniques will allow for the inclusion of usage variables in the analysis and restricts the predicted outcome to two possible values, either the member churns or renews. The goal is that the binary model will provide an easy to implement option for churn prediction that will ultimately give a sufficient analysis of what effect the customer's usage has on the likelihood of churn in contractual memberships.

The present study will base the analysis on the summary of the aggregated behavior across the members, and predict the likelihood of churn on the previously collected behavioral activity patterns from 2017-2019.

2. Positioning in the Literature

2.1. Churn Management and Prediction

The churn rate for a certain time period is calculated by dividing the number of customers lost by the number of customer that the business had at the beginning of the period. The churn rate is simply quantifying the effect after it has happened and explains the extent of the issue in terms of both customers and revenues that have already been lost. With newer technology, predictive analytics has become increasingly important. Data mining and churn prediction is generally built and tested using events from the past (Provost & Fawcett, 2013, p 45). Predictive models aim to point out the members that are exhibiting a pattern of behavior that is likely to lead to churn before they actually take the action to cancel their contracts and engage them to want to stay. It goes beyond the descriptive aspect of the churn rate, and attempts to estimate the action the member will most likely take.

With the big data era's improvements in both the availability of data storage and technology, the level of customer relationship management and data analysis has skyrocketed. The increase in technology and decrease in cost of computing power has enabled businesses to save their customer data in their own data warehouses and conduct their own data analysis in-house. Data in vast amounts can offer invaluable insights and competitive advantage if the right technological and organizational resources support them (Morabito, 2015). The type of data that is now being collected and stored by the majority of businesses can be used in statistical software to better understand their customer base and to predict future behavior. With access to modern analysis tools, companies have moved from the straightforward calculation of churn rate, to actually predicting churn events ahead of time.

The topic of churn prediction has been visited by several statisticians that have created various models for both explaining and predicting churn in both contractual and non-contractual business settings, both using machine learning algorithms and traditional statistics. The different models require various types of data input as the basis for calculating the probability of customer churn. These range from simple models where they use few or no explanatory variables to accurately predict the expected churn rate, to more complex models where they allow for multiple factors to affect the predicted result that give each member a estimated probability of churn.

The models that will be introduced going forward and that have been used as inspiration in the research and in building the prediction models are based on different techniques for calculating the probability of future member churn. The present research will however go beyond just predicting the likelihood for churn occuring and aims to include additional activity variables in order to also explain the “why” behind the churn prediction.

2.2. The shifted-Beta-Geometric Model

Fader and Hardie (2007) present the shifted-Beta-Geometric (sBG) model as a simple but accurate probability model for predicting customer churn. The sBG model is purely a probability model that takes the past churn rates into account to forecast the estimated churn in future periods by using few measures.

The model requires only a few data measures in order to run a prediction that returns accurate quantifiable forecasts for a business’ overall churn rate, and all model parameter calculations are easily executed in Microsoft Excel using the Solver-tool. The sBG model allows for customer predictions to be affected by different probabilities depending on tenure to account for loyalty in order to compute the estimated customer churn, which means the proportion of members that churn will decrease for each time period. The predictions are based on maximizing the log-likelihood by calculating the best value for the alpha and beta, which are the measures that allow for the members to differ in terms of their probability of churn.

The model provides an easy to implement tool for accurately estimating the amount of overall members that will be active in the next period which is valuable for planning purposes and provides a good option for describing a business' churn situation. However, it doesn't allow for the inclusion of usage variables to find the correlating effect they have on churn or for the prediction at individual member level. The predictions are useful in terms of describing the expected churn environment, but it fails to go deeper into explaining what factors have affected the members from leaving or for identifying the members that are at-risk for churn so that they can be identified as at-risk and targeted.

2.3. Ascarza-Hardie Model

Ascarza and Hardie's (2009) model for churn and usage behavior in contractual settings is presented as a dynamic latent trait model where both the level of member usage, i.e. activity, and the renewal behavior is modeled simultaneously in order to predict churn. It's a complex model that goes beyond predicting the estimated number of churn cases like the sBG model, but it gives both the expected length of contract and predicts the expected number of activities attended for each individual member. The model takes length of contract into account in addition to usage in order to account for customer commitment and tries to accommodate the customer threshold for usage that would lead to churn.

The model parameters are estimated by the hierarchical Bayesian method and will identify changes in the members usage behavior and allow for early prediction of churn before the contract has expired. The Ascarza-Hardie model is a good alternative for predicting churn in contractual memberships as it both accommodates for the dependent variable being binary, and allows for counting the activity usage and renewal decisions at different time intervals. The model takes usage under contract into account and uses it to predict future usage and simultaneously the probability for churn and retention. It also has the option to include add-on expenses within contract, which could be relevant for more detailed analysis. In the gym industry, there are options for going beyond the monthly contract price, for personal trainer hours for example. This is further than what the regression model will consider, but worth mentioning.

Even though the Ascarza-Hardie model would include all elements needed to predict churn and explain the most statistically significant moderators for contractual usage and renewal, the level of statistics needed to perform the analysis is so high that it could not practically be completed by any person in a short time frame without significant background and insight into the model. The Ascarza-Hardie model requires a high technical level of statistical knowledge, and can't easily be calculated in Microsoft Excel like the sBG model can, or in statistical software like the binary regression and classification tree model will.

2.4. Binary Regression and Classification Tree

The increased access to statistical software with user-friendly interfaces allow for implementing complex statistical models without much technological background. Instead of basing the churn predictions on statistical calculations like the Acarza-Hardie model, the binary regression and classification tree models use the benefits of statistical software which allows for easy implementation and the inclusion of more variables to account for the usage variables differing effect on churn for the contractual customers.

Binary regression modeling is a statistical process that studies the relationship between the dependent variable (response variable = churn) and a set of independent variables (explanatory variables), where the response is bound to return one of two possible response variable values. The binary regression technique returns values that can be used to evaluate the direct effect or the correlation that each individual independent variable has on the change in the response variable (Ciaburro, 2018, p. 21). Each change in one of the values can be expected to change the probability of the response variable by the listed effect as long as all other factors are kept constant.

Classification tree models take a different approach for predicting churn cases than the binary regression model. Classification trees use supervised learning in order to predict the dependent variable by creating a step-by-step graph instead of a equation with coefficients, where each level of the graph depicts a binary decision to split the customers using the value of the customers explanatory variables. Tree classification models provide a visual model that are easier to interpret visually than what binary regression equations are.

Both the binary regression and classification tree models will be given a list of several explanatory variables, which will allow them to find the correlating effect that the usage under contract has on the renewal decision. Only the variables that are deemed statistically significant for predicting churn across the members in the gym industry will be kept in the final model.

Table 1. Comparison of previous and present study approach

	Fader-Hardie model	**Ascarza-Hardie model**	**Binary Regression and Classification Tree models**
Inclusion of explanatory variables		**X**	**X**
Churn detection at member-level		**X**	**X**
Ease of implementation	**X**		**X**

Table 1 show the areas where the binary regression and classification tree models are set apart from the Fader-Hardie and the Ascarza-Hardie model. The binary regression and classification model are the only model options that crosses off every point with opening up for explanatory variables, early detection of churn at member-level, and easy implementation. While the Fader-Hardie model can accurately predict the churn rate using Excel, it doesn't allow for the inclusion of explanatory variables or prediction at the member level. The Ascarza-Hardie model allows for the inclusion of more variables and early prediction of churn at member level, but the downside with the model is that it requires a high level of statistics which makes it time consuming and affects the ease of implementation.

Typically statistical models aspire to include as few variables as possible in order to return their results. The general consensus is that including too many variables in a model will make the results less precise and more difficult to interpret. However, the regression model is attempting to go further than merely predicting the estimated value of the churn rate, which is already done very effectively in the sBG model with no explanatory variables. The aim of the present study is to build a model that can predict the likelihood of churn for each member in the sample before the renewal decision and allow for specifying how the usage under contract affects this likelihood.

3. Research Method

The aim is to build a churn prediction model by testing the best fit between the two statistical approaches, binary regression and classification tree model. The technique should provide the business with a well informed estimation of future churn while allowing for the investigation of the usage variables and their differentiating effect on customer churn. This will give businesses with contractual memberships access to a model that can provide insights into what drives their customer churning behavior, in order to target marketing campaign at the at-risk members to prevent some customers from leaving in the future.

From a business perspective, there is more value in a predictive model than merely a descriptive one as it lets you take action **before** the customer cancels their contract. The present study models will be based on input from the customer activity while under contract as it is easily observable and collected, as the customers have to check in at the entrance of the gym and type of activity is recorded. The pricing model is set up so that you pay a monthly cost, and with that you can use the service as much as you want, so the activity is not affected by cost.

3.1. GDPR Compliance

Since the research question is based on the collection and analysing of customer data and behavior for customer relationship management purposes, there are certain guidelines and regulations that come into place in order to protect the individuals' personal information. The new EU privacy regulation, the General Data Protection Regulation (GDPR, 2018) was put in place to protect individual persons from the processing of personal data (Art.1, GDPR). Any business that processes personal data will have to adhere to GDPR, which places strict requirements on how businesses collect, store, and use the information on their customers.

In order to process the customer data for the purpose of predicting churn, the identifiable portions of the data would have to be altered so that the data is anonymous and can't be used to trace back to the natural person behind it. Instead of each customer being identified by their name, they are listed by member ID in the data, that only the business should have the code for. Any unique identifiers in the data should be altered or taken out before the analysis is conducted.

The model aims to find and analyse the aggregated churn effect that exists in the large groups of customer segments in the gym industry, and doesn't contain any identifiable data. The model is not aimed at investigating individual behavior, but rather the large common patterns that exist in the data. This is all done in order to stay compliant with the GDPR regulations.

3.2. The Dataset

There are 768,058 rows in the raw data, where each row is an individual recorded visit at the gym for 10,000 unique customers. After aggregating the data to monthly intervals the data set is shaved down to 4196 rows, where each unique member has one row.

The data used for the models is detailed visit data from the fitness industry. The unit of analysis is the customer contract, where the dependent variable that the model is set to estimate is whether the contract is renewed at the end of the contract period. The various explanatory variables that are used to estimate the prediction are activity proportions, gender, customer type, and usage, which includes monthly average, and the three last month's proportional activity frequency.

There is only sufficient data to account for the prediction for members with yearly contracts.

3.3. Creation of Dummy (Factor) Variables

When a data set contains categorical variables, i.e. where the possible data input for a variable is limited to a certain amount of values, the creation of dummy variables is advisable. Dummy variables are numerical variables used to represent subgroups of the sample population by splitting up existing variables into another data structure. Instead of having all of the categorical variables listed in the same column, the dummy variables create a whole new column with binary inputs where the value is either set to be a 1 or 0. When the value of the dummy is set to 1 it generally means that the categorical variable is true for the member.

One of the benefits of including dummy variables in regression models is that they help separate and distinguish the different effects from the treatment groups present in the data, which helps isolate the most statistically significant variables.

3.3.1 Predicting Variable: Churn

Having a clear predicting variable is essential for a regression model to return any value, as it needs something specifically present to search for and present a prediction on. As the goal of the model is to predict customer churn, the predicting variable has to reflect this specific action of decision renewal in the members.

Churn = 1, if the member chooses to leave the company at the end of the contract period
Churn = 0, if the member chooses to renew their contract at the end of the contract period

3.4. Frequency of Usage and Type

In order to account for the difference in the members usage frequency leading up to the contract renewal decision, alternative explanatory variables will be included in the prediction models. Both the last months of usage leading up to the decision and the monthly average over the contract period are relevant variables to include in the study as it shows both the frequency and recency of the members usage of membership.

Usage is highly relevant to include as it allows for the inclusion of the customers activity level in the year leading up to the contract renewal and the last months before the decision is made. The usage will be added as the member's monthly average over the entire year, and the last months (12th, 11th, and 10th) will be included as their monthly proportion in relation to their overall yearly average. If the relative usage proportion in month 12 = 100 (%), the activity is at exactly the same level as the entire year, while 50 (%) means that the activity in that month is half of the yearly average.

Additionally, to account for the difference in the members type of activity usage each type of activity will be added as alternative explanatory variables, in the form of calculated proportions. The activity proportions will be a calculated percentage of the total activity frequency for each type of activity, divided by the total number of recorded activities for each of the members. There are three different activity types in the data: gym activity, group class activity, and personal trainer activity. Each member in the data will have a total activity proportion of 100 %, divided between the three activity types.

3.5. Descriptive Statistics

Binary regression and classification tree models are built on the basis of there being several explanatory variables that will be tested for their effect on the response variable. The summary shows the spread and distribution of the 12 explanatory variables that is used in the prediction models.

The binary regression models can accommodate for the inclusion of interaction terms, which is when one explanatory variable interacts with another variable to produce different coefficients for different values of the interacting variable. Interaction terms can greatly change the interpretation of variables, but is unfortunately not possible to include in a classification tree model.

Table 2. Summary of Explanatory Variables

Numerical Variables	**Min**	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max**
MonthyAverage	0.250	4.92	8.42	10.32	13.58	56.83
Age	13.00	25.00	28.00	31.70	35.00	90.00
GroupProp	0.00	0.00	3.99	25.52	50.00	100.00
PTProp	0.00	0.00	0.00	3.50	1.43	92.00
GymProp	0.00	44.02	89.44	70.98	100.00	100.00
Month12Prop	0.00	26.03	53.82	61.10	84.21	600.00
Month11Prop	0.00	13.77	46.15	53.73	78.58	600.00
Month10Prop	0.00	17.14	48.28	55.88	81.39	436.36
Factor Variables	**0**	**1**		**Factor Variables**	**0**	**1**
Churn	93.83 %	6.17 %		Student	71.45 %	28.55 %
GenderF	44.66 %	55.34 %		Private	66.66 %	33.34 %
				Corporate	66.97 %	33.03 %

Table 2 shows that out of the 4196 unique members that are active long enough to get to the renewal stage of their 1-year contract, 6 % end up churning at the end.

The gender distribution is somewhat even, but there are slightly more females than males, and there is a slight imbalance in the age of the members. 75 % of the members are placed below 35 years (below the 3rd quartile). The activity variables, monthly average and the monthly proportions show obvious heterogeneity, as the members have very different averages and vastly different total visits to the gym. The majority of the members are placed within a monthly average ranging from 5 to 13.5 (between the 1st and 3rd quartile), while some few members go in more extreme values in both directions with minimum and maximum averages from 0 and up to 56. In the last three months leading up to the renewal decision, there seems to be a trend that the majority of the members use their membership less than the average values from the year, as 75% of the members have proportional activity levels below 85 % of their average level for the year.

The distribution between the type of activities and their proportions vary greatly between the members. Gym has the highest values for both the mean and median, which suggests this is the most common type of activity. Group classes see more of a spread, where many members have low values that drag down the median, but the mean is fairly high at 25.5 %. PT is far more uncommon with the median placed at 0% and the 3rd quartile at 1.43% which means that 75 % of the customer are placed below this value of personal training proportional to their total activity.

There is about an equal distribution between the three different member types in the data, where the student proportion is slightly smaller than the private and corporate member types.

Figure 1. Age Distribution with Churn Proportion

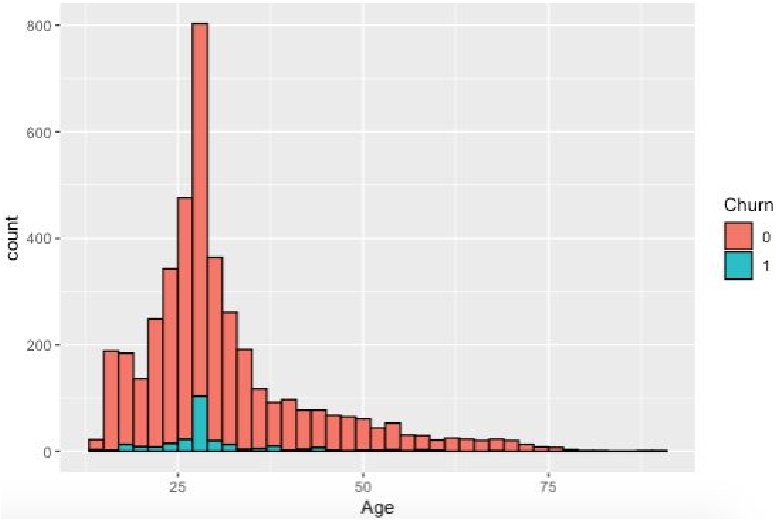
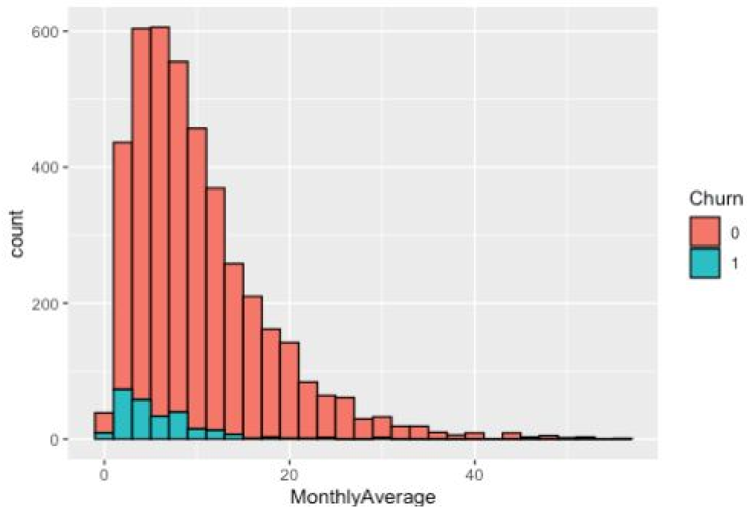


Figure 1 shows that the majority of the members' age in the data are located around 20 to 30 years, with a long skewed tail to the right representing the older proportion of members. The spread is representable for the average age distribution that is found in most gyms, and gives an indication that the data set is correctly balanced for the sample population in terms of the distribution of age groups. The lighter portion of the graph shows what proportion of churning customers belong to the age groups, and it looks like there are few customers in the higher age groups that end up churning.

Figure 2. Monthly Average with Churn Proportion



The histogram in figure 2 shows the monthly activity average with the proportion of the members in the final data who end up churning after the 12 contractual months with a binwidth of 2. The lighter part of the graph represent the proportion of members within the activity level that end up ending their contract.

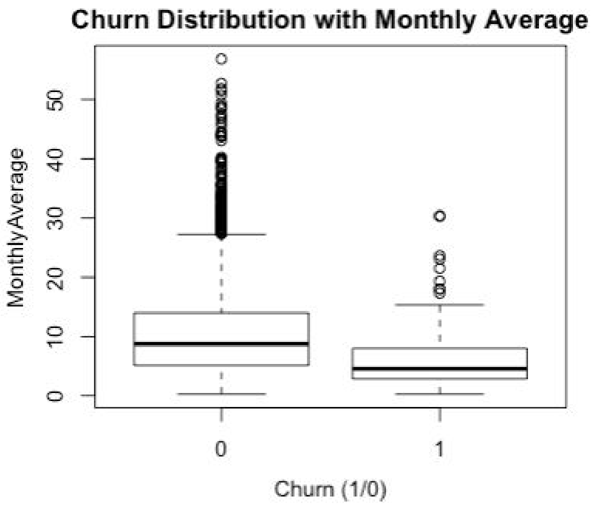
The distribution shows that it's not just members with low or no monthly activity that leave the company, but that the majority actually uses their membership 2-8 times per month over the entire 1 year period, and even some as high as 14. Once the monthly activity average goes above 16 there is little churn present in the members.

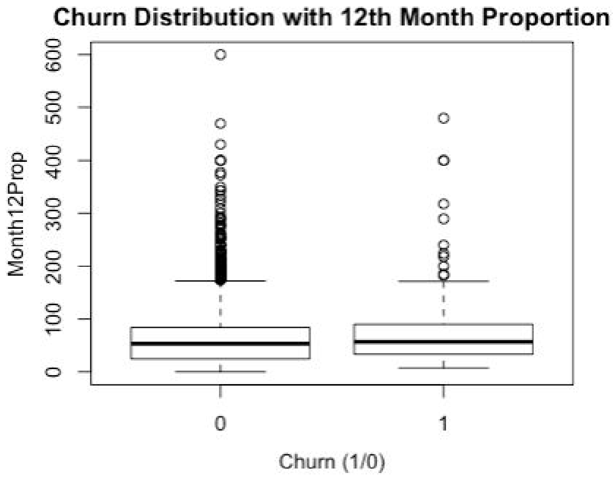
3.6. Association Between Explanatory Variables and Churn

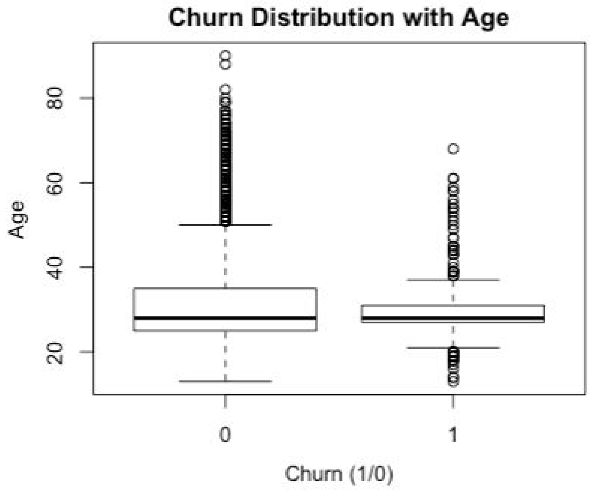
Investigating the spread of the boxplots of the explanatory variables against the dependent variable of churn can help show the association between the variables and their effect on churn. The boxplots can show early patterns of what can be expected as being the significant variables at the modeling stage.

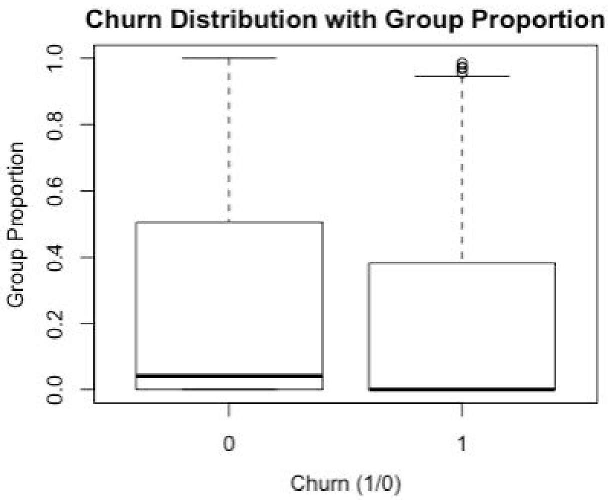
The boxplots will show a comparison of the combined members' average values between the Churn variable at value 0 (no churn) and at value 1 (churn). The variables that show significant differences between the two stages of churn, are more likely to be found as significant variables, as they show a difference between the types of customers who end up leaving the company at the end of the contract period.

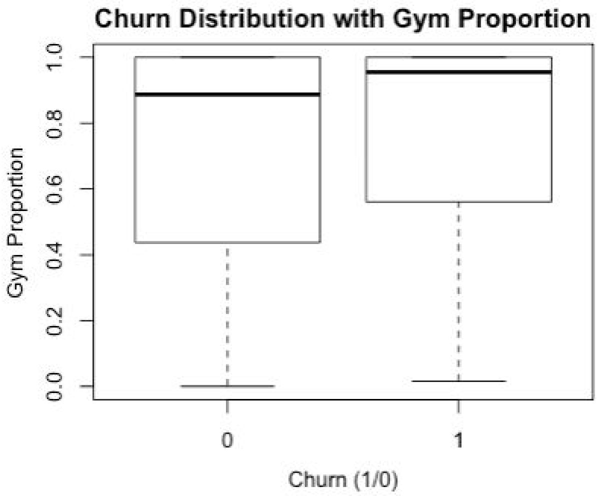
Figures 3-8. Boxplots of Variables and Churn Distribution

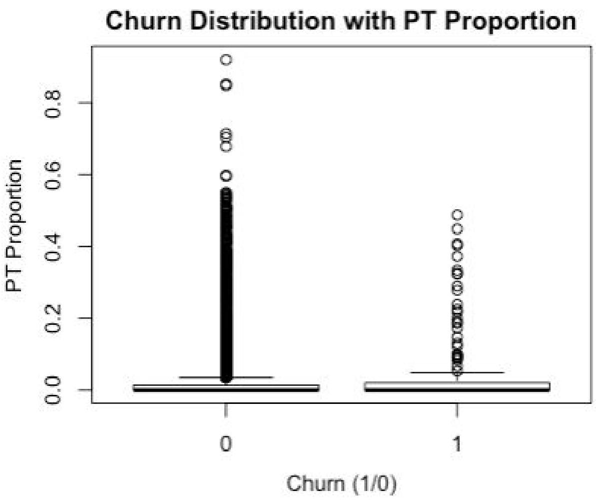



Churn Distribution with 12th Month Proportion
Month12Prop
0
100
200
300
400
500
600
0
1
Churn (1/0)


Churn Distribution with Age
Age
20
40
60
80
0
1
Churn (1/0)


Churn Distribution with Group Proportion
Group Proportion
0.0
0.2
0.4
0.6
0.8
1.0
0
1
Churn (1/0)





None of the boxplots above seems completely identical at first glance. The median line in the boxplots are different for the two stages of churn for several of the variables, which does suggests that the variables have some sort of effect on the occurrence of churn in the sample population. From the boxplots, one can see how members that end up churning at the end of the contract tend to have significantly lower monthly averages and group proportions. While the gym proportions are higher for the churning members. The apparent differences in the averages between the two states of churn in these variables suggest that there is a correlation between the usage of customers and churn.

The “Month12Prop”, “Age”, and “PTProp” boxplots show a similar median between both stages of churn, but the spread is narrower for the churning population, which is likely linked to there being less data points in the churning population and not enough to lead to a conclusion for there being a correlation between these variables and churn.

Figure 9. Spread of Monthly Tenure over the Collection Period

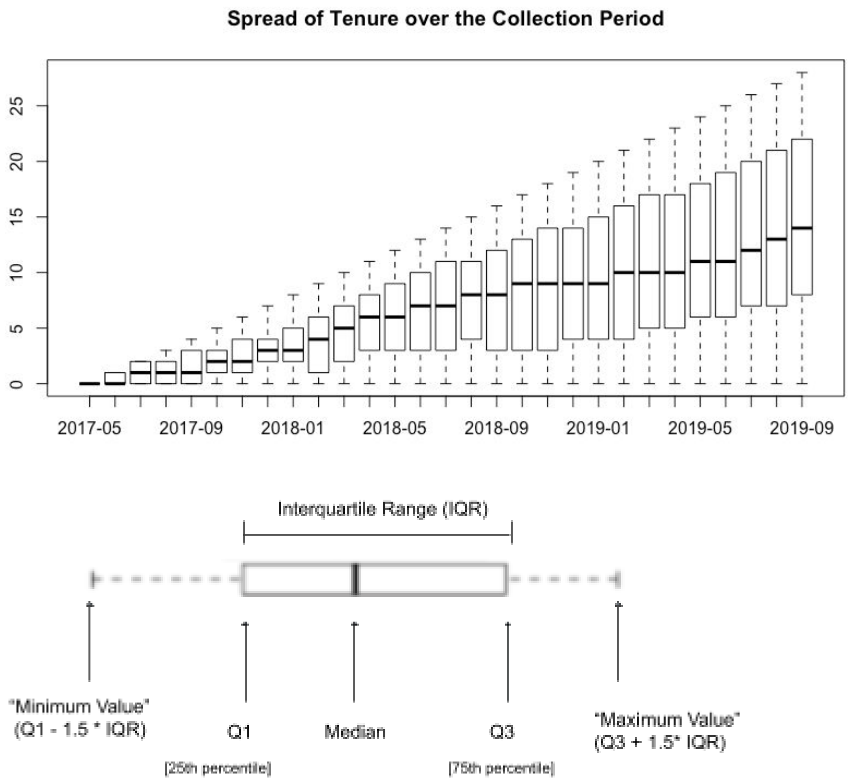


Figure 10 shows the variability between the members tenure for each observed month, and shows how the data is spread out over time. The values between Q1 and Q3 constitutes the interquartile range, which means that members from the 25th to 75th percentile lie between these values. Between the lower and upper whisker of each box you find the minimum and maximum data point excluding outliers which would have been shown as single points.

Most of the members in the sample data have different start dates for their contracts, which is reflected in the spread of tenure. It shows that for each observed month there are new members added, as the whiskers always go down to 0, which represents members with a tenure of 0 months. Over the collection period of 29 months, there is only a small portion of members who are in the system long enough in order to reach a tenure of over 26 months. At the end, the average tenure for all of the 10,000 members is at around 12 months, but the variability increases the further out we move.

3.7. Outliers

Outliers are explained as data points that differ significantly from the other data points in a sample. Sometimes outliers represent mistakes in the recorded data, or in other cases just represent members that are way more active than the average member. Outliers will greatly affect a regression model if they are left untouched, as they tend to change the direction of the regression line as they are given more weight than they are worth (Ciaburro, 2018). Boxplots and histograms are great visual aids for identifying outliers as they map out the point where the majority of data points are grouped. Leaving the outliers in the data can alter the effect that is given to the regression coefficients.

There are several outliers in the data that are statistically different from the mean, for example, the mean average of the monthly activity is placed at 8, while the highest observed value for the same variable is 56. These outlier values are however not large enough to signal that there is a mistake in the observed values. The values will be significant to leave in the data as they show a segment of customer that is highly active, and could be very relevant for the prediction.

3.8. Class Imbalance

The presence of churning members in the data set is only at 6 %, while the proportion of non-churning members are the remaining 94 %. The unbalanced distribution between the churning and non-churning members is known as the class imbalance problem, which can affect the quality of the regression model as the algorithm isn't given enough information on the minority class, where churn = 1, in order to make accurate predictions.

Ideally there should be an equal balance between the distribution of the binary dependent variable class. However, in many cases where one is interested in predicting the occurrence of a specific state, this is rare and does naturally occur less frequently, as it is with cancer detection in scans, detecting fraud in credit card customers, and similarly in customer churn prediction. In these cases the dependent class is often only present in a small percentage of the data, but these are the most important percentages for a model to be able to detect.

In binary regression models, predicting the outcome is found through an estimation of the probability of survival, which is used as the basis for deciding whether the members are placed in one of the binary outcomes, Churn = yes, or Churn = no. A specific percentage decides where the cut-off point of probabilities differentiate between the two options. In a balanced data set where there is an equal distribution between the two outcomes, the threshold value for probability is placed at 50 %, where an estimated probability above the threshold would result in a probability of the prediction to be placed at true (no churn), and opposite for the probabilities under the given threshold (churn). Because of the imbalance of churn cases present in the data set, the models threshold value is placed at the value of churn proportion at 6%. An estimated probability that is calculated to be lower than this threshold value, will result in a prediction of churn, while a probability above the threshold value will predict that the member will remain active in the next period.

For class imbalance problems it is more appropriate to evaluate the quality of models by using confusion matrices, as it will show the spread of the correct and missed prediction cases instead of the overall accuracy as it could easily be misinterpreted as a good fitting model if the accuracy is falsely high (Clemente, Giner-Bosch, & Matías, 2012). A model that can predict some occurrence of churn cases will be more valuable to a business than one that can only predict the majority class of non churning cases.

4. Churn Prediction

In regular regression models, the response variables is typically returned as a numerical value. However, since the dependent variable “Churn” is binary, the regression model is restricted to classification of the two possible alternatives, i.e. Churn: yes/no. A classification prediction performed through regression will explain the probability of an event happening (Ciaburro, 2018, pg 121). The classification model for churn will return the differing effects that the explanatory variables have on the probability of the response variable.

The renewal point for the members, which for yearly contracts is at month 12, is extracted as a subset to highlight the decisions the members take after staying for the contractual 12 months. The model will be aimed at finding the statistical significant factors that can help explain why the members choose to leave the company.

Figure 10. Graphical Description of Churn

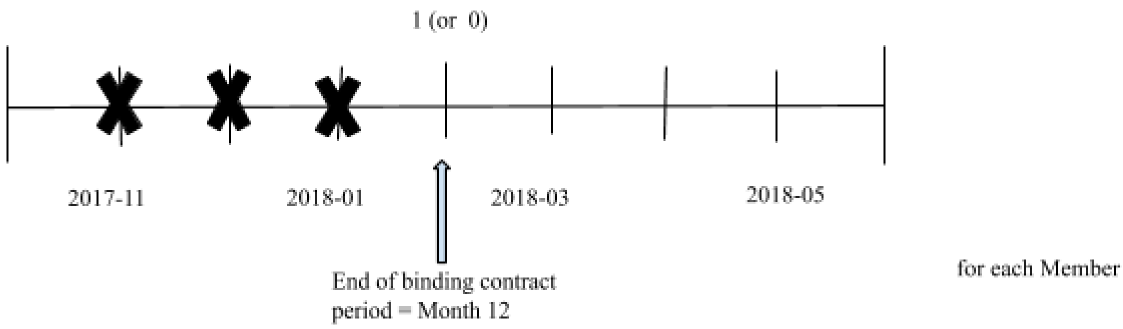


Figure 10 shows how each individual member is graphically represented in the non-aggregated data set. Each row in the Churn-column will return a 0 up until the point where the member is no longer an active member. At the point where the member decides to cancel their membership, the row will return a 1 to signal member churn. For every member that decides to remain as an active member at the contract renewal stage, the row will show a 0. In the data set that is used for running the binary models, only the 12th month will be included, as this is where the renewal decision takes place. The 11 months prior or not of interest as these are bound to contract to continue on to the next period.

4.1. Evaluating the Alternative Models

Since the dependent variable of in this case is binary, with only two possible outcomes for the prediction of the member decision, the model options are limited to binary regression and classification models. There exists a variety of different statistical classification models that will be trialed in order to construct the most accurate churn prediction model for contractual memberships. Once the data is structured and cleaned enough to start the modeling phase, several options will be tested and evaluated against each other:

- Binary Logit
- Binary Probit
- Classification Tree Model

The three types of churn prediction models will be tested going forward, with an analysis of their estimated coefficients and prediction distributions, and the decision for what explanatory variables are significant and should be included in the final model. In the end, the best fitting model will be concluded for the churn prediction of contractual models with the inclusion of member activity by evaluating the distribution of the predictions in their confusion matrices.

There are several parameters and values that are returned in the summary of the models that can be considered when one compares the accuracy and statistical significance of the alternative. Below are some of the values that will be used to evaluate the best fitting model for the churn prediction.

4.1.1. P-values and Regression Coefficients

The p-values and regression coefficients found in the summary statistics for each regression model will be used in order to determine what explanatory variables are found to have an effect that is statistically significant for the dependent variable of churn, and will give the estimated values for how they differ between the members.

The p-value will be an important indicator for what variables should be taken out of the model. The general rule is that the p-values should be smaller than the significance level of 0.05 in order for the explanatory variable to be deemed statistically significant. Once the p-value has shown that the variable is likely to be significant, the value of the associated regression coefficient will give the estimate for the effect the variable has on the dependent variable. The regression coefficient gives the estimated change in the dependent variable for one unit change in the explanatory variable while holding all other explanatory variables constant.

The sign of the estimated coefficients will determine if the explanatory variables has a positive or negative effect on the prediction of churn. A negative coefficient indicates the variable will decrease the likelihood of churn, while a positive coefficient will increase the predicted likelihood.

4.1.2. The Marginal Effects

Marginal effect is an important measure to investigate in order to understand what the change in explanatory variables does with the probability of churn occurrence. The marginal effect is used to help explain the variables effect on the likelihood of churn which is used to predict whether the member is placed as 0 or 1. In a linear model, the coefficients has a linear effect which means the marginal effect is simply β_i (Fernihough, 2011). In a binary model the probability distribution is s-shaped in order to restrict the outcome between 0 and 1, as it is interested in whether the member will take one of the two outcomes, not the probability itself. In binary models increasing x will increase or decrease the likelihood of y being equal to 1.

The marginal effect on the churn likelihood in a logit model is found by exponentially transforming the coefficients in the model, while the probit model coefficient transformation is based on the cumulative normal distribution.

4.1.3. AIC Score

In order to separate out the best prediction model from the tested alternatives, the Akaike information criterion can be helpful in model selection. The AIC score compares the quality of model fit by finding the trade off between 2 areas: the value of the models log-likelihood (how well the models fits the data) and number of explanatory variables (the models complexity).

$$AIC = -2(\text{log-likelihood}) + 2K, \quad (1)$$

where k is the number of explanatory variables including the intercept.

The values from several models can in comparison choose the model and explanatory variables that help best explain the variability in the data. The model with the lowest AIC score should be selected as the best quality one.

4.1.4. Residual Plots

Each binary regression model will have an associated residual plot that shows the models predicted performance against the actual observed values of churn (0/1), where the residual is the computed difference between the two. The residual plot visualizes each models difference between the observed and predicted values, and is a good plot to investigate in order to check the quality of the churn predictions.

$$\text{Residual} = \text{Observed value (0 or 1)} - \text{Predicted value (probability)} \quad (2)$$

In linear models, the residual plot should ideally show a random dispersion of points when the model is a good fit. For binary regression models binned residual plots are the most appropriate. The binned residual plots divides the data into a series of bins based on their fitted values, and plots the average residual versus the average fitted value for each of the bins. The collection of bins compromise bands, that constitute plus and minus two standard-error bounds for each bin, that ideally the data points should fall within for it to be considered a good fitting model.

4.1.5. The Accuracy of the Model

Determining the accuracy of the model will be based on both the sensitivity and specificity of the model, as the overall accuracy is not a good evaluation method when the prediction class is imbalanced (Ferri et al, 2002). The sensitivity will return the percentage of the actual non-churn cases that it is able to predict correctly in relation to the predictions of no churn that are incorrect.

$$\text{Sensitivity} = \text{true negative} / \text{true negative} + \text{false positive} \quad (3)$$

The specificity is the opposite, and returns the percentage of the actual churn cases that is correctly predicted by the mode in relation to the ones it gets wrong. Confusion matrices for each model will show the dispersion of the spread of missed and correct classification of predictions.

$$\textit{Specificity} = \textit{true positive} \,/\, \textit{true positive} + \textit{false negative} \tag{4}$$

Table 3. Confusion Matrix

	Actual		
Predicted		0	1
	0	True Negative	False Negative
	1	False Positive	True Positive

4.1.6. Train and Test Set

The first step before moving forward with the baseline model, is to split up the final transformed data set in a test and train set. The train set will be used as the basis for building the model, while the test set will be a part of the data that the model doesn’t have access to until the model is finished. The test set will used to see if the finished model can accurately predict cases outside of the data it was built on. This is an essential step in assessing the accuracy of the various models and by avoiding overfitting, as the test set will return the percentage of cases where the models can correctly predict churn.

The training set will contain 70% of the data, which equals 2937 rows, while the test set will be the remaining 30% and 1259 rows.

4.2. Baseline Model

We know from the distribution in the sample data that member churn is less common than members staying with the company past the end of the contract. The baseline model is a simple prediction where one predicts all members to remain active past the renewal decision at month 12, i.e. no members are predicted to churn.

This takes advantage of the imbalance of the churn class in order to get return a high accuracy score. In this case, out of the test data of 1259 members, the model will correctly predict Churn = 0 in 1181 of the cases in the data set, and 78 will be predicted wrong where one predict 0 (no churn), but they are actually 1's (churn). This leads to a baseline model accuracy of 94 %, and the associated confusion matrix is shown below.

4.2.1. Confusion Matrix: Baseline

Table 4. Confusion Matrix Baseline Model

n = 1259	**Actual: No Churn**	**Actual: Churn**
Predicted: No Churn	1181 (100 %)	78 (100 %)
Predicted: Churn	0 (0 %)	0 (0 %)

As a model, the baseline prediction seems to do a pretty good job from the high accuracy score, but as the confusion matrix highlights, in practice it isn't predicting any churn cases at all. The specificity of the model, which calculates how many of the positive values has been correctly predicted, where churn = 1, is at 0 %. The baseline model isn't actually providing any value as it's not picking out any member churn which is what it is aiming to do.

The goal for the regression models going forward is to compete with the baseline model, where the accuracy is likely to be lower, but where they can actually identify some of the churn cases from the data. This requires attempting to move some of the churn cases down to the lower right part of the confusion matrix without moving too many of the non-churning cases down to the lower left part of the matrix.

4.3. Logit Model

The equations behind the binary regression models stem from extensions of the simple linear model equation, but with added alterations to account for different distributions of the data and multiple explanatory variables.

As the linear function is not appropriate for churn prediction, one would have to alter the linear equation to restrict the estimate of the dependent variable for it to fit with a binary regression model. Applying an exponential function to the linear equation restricts the values that the dependent variable of churn can return, to two outcomes, i.e. 0 or 1, and including the possibility of multiple explanatory variables allows for the model to distinguish between the variables' partial effect on the estimation of churn.

Which leads to the binary logit equation:

$$P(y = 1) = exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n) /$$
$$1 + exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n), \qquad (5)$$

where P(y=1) is the probability for Churn = 1,
x is the *n* independent variables, and
the coefficients β_i measures the x_i variables effect on P(y=1), for each unit change in x_i, when all other independent variables are kept constant.

The coefficients that are returned when running the binary logistic regression are converted to the odds ratio for a more straightforward interpretation of their effect on churn.

Reframing the equation to the odds ratio gives us:

$$P(y = 1) / 1 - P(y = 1) = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n} \qquad (6)$$

In order to reach a regression equation that returns coefficients' effect on the probability of churn that can be easier to interpret their effect on the prediction, the logarithm is taken on both sides of the odds ratio equation. Taking the logarithm on both sides returns a final equation that looks similar to the linear regression equation. Which leads to:

$$log(P(y = 1) / 1 - P(y = 1)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \quad (7)$$

This final multivariate logistic regression equation is the basis for the binary logit regression model. It accounts both for the presence of several independent variables, and limits the response variable to binary outputs of churn. The model predicts the probability of the binary variable being equal to 1, which in this case equals churn, and the coefficients are given values that affect the probability.

In binary logistic regression models, the value of the coefficients that result from the regression equation is not an expected quantifiable value, but the logarithmic odds that the members belong to the specific group, which is the dependent variable of churn. The coefficients (under "Estimate") for each variable given in the summary will show the change in log-odds of the outcome of churn that each variable will have for each one unit increase in a independent variable, or for the factor variables, the change in log odds whether the case is true or not for the member.

Converting the log odds to the odds ratio will give the effect the variable has on the dependent variable of churn, but only in terms of the coefficients increasing or decreasing effect on the likelihood of churn. The coefficients will have to be converted to give the average marginal effect in order to investigate the actual effect the variables have on the estimation of probability of churn.

4.3.1. Logit Model: Estimated Results

The first logit model is a full logistic binary regression model on the training data, where all explanatory variables are included, in addition to interaction terms between the continuous activity variables.

Table 5. Estimated coefficients: Full Logit Model

	Estimate	Std.Error	z-value	p-value
(Intercept)	- 3.038	0.849	- 3.580	0.0003 ***
GenderF1	0.126	0.182	0.692	0.489
MonthlyAverage	- 0.108	0.022	- 4.901	9.54e-07 ***
Age	- 0.016	0.009	- 1.747	0.081
GroupProp	0.009	0.005	2.054	0.040 *
PTProp	0.017	0.017	1.009	0.313
Student1	1.289	0.753	1.713	0.087
Private1	1.526	0.736	2.075	0.038 *
Corporate1	0.992	0.741	1.339	0.181
Month12Prop	0.002	0.001	1.259	0.207
Month11Prop	0.002	0.001	1.953	0.051
Month10Prop	0.004	0.001	3.032	0.002 **
GroupProp:MonthlyAverage	- 0.002	0.008	- 2.769	0.005 **
PTProp:MonthlyAverage	- 0.004	0.287	- 1.263	0.206
Null deviance	1359.4			
AIC	1240.9			

From the high p-values, it is apparent that there are many insignificant variables in the full model that do not contribute valuable information to the model that helps it make prediction estimates. In this model, the most significant variables with the lowest p-values are “MonthlyAverage”, “Month10Prop”, and the interaction term for “GroupProp:MonthlyAverage”.

The null deviance is a fit index that is used to compare the value that a model with only the intercept (β_0) can do in relation to the model with explanatory variables included. The lower the null deviance model, the better it is. In this case the null deviance of 1359.4 is considerably higher than the AIC of the full logit model at 1240.9, which tells us that the full model is a better quality model and performs better predictions than the null deviance model even when the logit model has several variables that are not significant for the predicting variable.

Before analysing the meaning and effect of the logistic model coefficients on churn, the number of variables should be decreased so that the statistically insignificant variables are removed. Only keeping the most significant variables will leave the model easier to interpret and more accurate as there are fewer variables to account for in the binary logistic regression equation.

4.3.2. Reduced Model: Logit

The explanatory variables that are kept in the reduced logit model is "MonthlyAverage", "Age", "GroupProp", "Private", "Month11Prop", "Monthly10Prop", and the interaction terms for "MonthlyAverage:GroupProp". In the reduced logit model, all variables now have a p-value smaller than 0.05, which makes them all statistically significant for the estimation of churn.

Several usage variables are found to be significant, which leads to the assumption that both usage type and amounts are good indicators for what decision a member will take at the end of their contract. The proportional activity in month 12 was not found to be statistically significant, which could suggest that the renewal decision has already been decided before the final month of the contract, and the usage months before the renewal decision can be an important indicator for churn.

Table 6. Estimated coefficients: Reduced Logit Model

	Estimate	Std.Error	z-value	p-value
(Intercept)	- 1.459	0.311	- 4.697	2.64e-06 ***
MonthlyAverage	- 0.121	0.021	- 5.663	1.49e-08 ***
Age	- 0.025	0.008	- 3.053	0.002 **
GroupProp	0.010	0.005	2.215	0.027 *
Private1	0.442	0.162	2.715	0.007 **
Month11Prop	0.003	0.001	2.274	0.023 *
Month10Prop	0.004	0.001	3.082	0.002 **
MonthlyAverage:GroupProp	- 0.212	0.078	- 2.731	0.006 **
Null deviance	1359.4			
AIC	1237.7			

The estimated parameter coefficients lead to the logit regression equation of:

*log(Pr(churn)/1-Pr(Churn) = - 1.459 - 0.121 * MonthlyAverage - 0.0253 * Age + 0.010 * GroupProp + 0.442 * Private1 + 0.003 * Month11Prop + 0.004 * Month10Prop - 0.212 * MonthlyAverage:GroupProp (8)*

The logit regression equation will return a different log odds for each member dependent on their characteristics and activity levels over the recorded period. The lower the value of the members log odds, the less likely the probability of churn, while the higher the log odds value, the more likely the members are to churn at the end of their contract.

The model consists of several continuous variables, that will affect the log odds by the coefficient value multiplied by the variable value for each member. The coefficient estimate of -0.121 for "MonthlyAverage" can be interpreted as the estimated change in log odds for a one unit increase in the monthly activity in the year leading up to the contract decision. The member values for monthly average can go as high as 56, which in that case would reduce the value of the log odds by

-0.121 * 56, which equals - 6.78. The “Month11Prop” can be interpreted as a 1 % increase in the proportion will lead to a 0.2 % increase in the log odds.

The factor variable, “Private1”, will either affect the log odds of the model with the value of the estimated coefficient of 0.442 if the member type is Private, or will have no effect on the log odds when the member type is not Private. Even though the estimated coefficient of Private is the highest value, the effect of Private can be comparably small to the other continuous variables.

4.3.3. Average Marginal Effect

The log odds can be complicated to interpret on their own, but by transforming the log odds using the inverse logit function, it will return the marginal effect each explanatory variable has on the probability of churn occurring. By running the inverse logit function on different variable values of log odds, one can easily see the explanatory variables marginal effect on the probability of churn when all other variables are kept constant.

The lower the combined value of log odds, the lower the associated probability of churn will be. This means that the higher the estimated coefficient for a variable is, the more significant it is for affecting the probability of churn for each member.

Table 7. Logit Model Average Marginal Effects

Explanatory Variable	**Average Marginal Effect**
(Intercept)	- 0.080
MonthlyAverage	- 0.007
Age	- 0.001
GroupProp	0.0006
Private1	0.024
Month11Prop	0.0001
Month10Prop	0.0002
MonthlyAverage:GroupProp	- 0.012

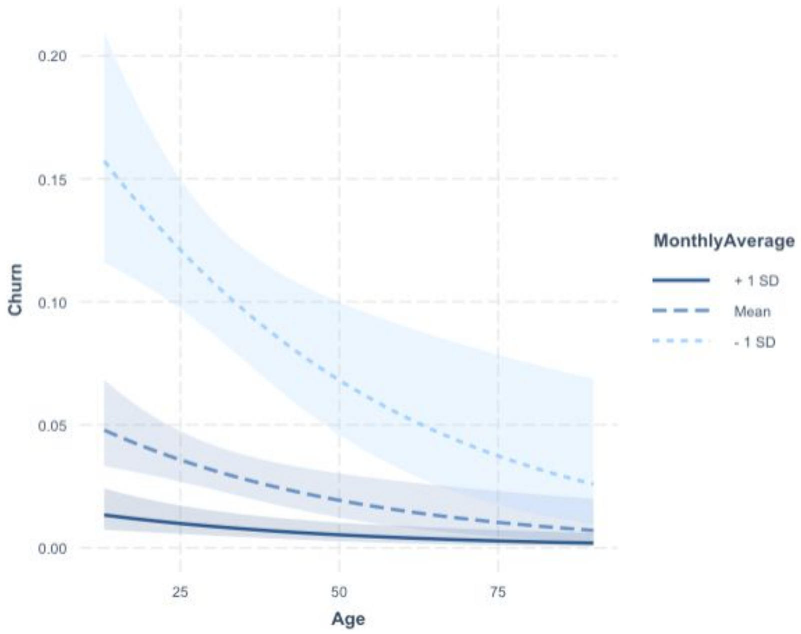
When the average marginal effect is positive, the likelihood of churn will increase by the marginal effect for each unit increase.

There are three variables that decrease the likelihood of churn according to the logit model are “MonthlyAverage”, “Age”, and the interaction between monthly averages and group proportion. The age can consequently be said to lower the likelihood of churn by 0.1 % for each yearly unit increase, which means the older members are not as subjected to the risk of churn when the activity levels go down. Each unit increase in the monthly average attendance will lower the likelihood of churn by 0.7 %, and the combination of a high monthly average and group proportion will decrease the likelihood of churn further. The importance of a high monthly average suggests that habitual usage over the entire year is the best way to ensure that the members remain active after the contract end.

The remaining variables in the model are found to increase the likelihood of churn when their value increases. Surprisingly enough, having higher activity levels in Month 10 and 11 is found to increase the likelihood of churn, while the 12th month is not significant. For each percentage increase in the “Month11Prop”, the likelihood of churn will increase by 0.01 %, and each percentage increase in “GroupProp” will increase the likelihood by 0.06 %. For members that have private memberships, their likelihood of churn increases by 2.4 %, which suggests that students and corporate memberships are not as affected by the change in usage as the private members. This could be related to Private memberships being placed in a higher price range and more susceptible for competing brands, than for the members that are offered student or corporate discounts.

Showing the marginal and interaction effect in a plot together will visually show how the different values for the explanatory variables will affect the churn prediction.

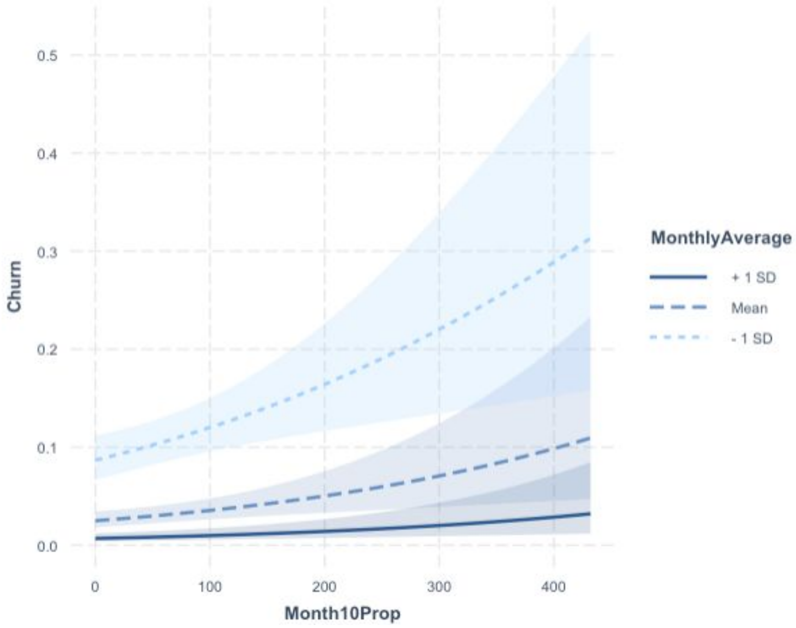
Figure 11. Age and Monthly Average Effect on Churn



The likelihood of churn decreases with each unit increase in age, and figure 11 shows the effect on three different lines representing the mean and + - 1 standard deviation of the monthly average. It is apparent that lower monthly averages are found dramatically more likely to churn than the higher averages. The slope for the lowest monthly averages has the widest confidence bands, but it is also the group of activity that is lowered the most percentage points when age increases. From this plot the customers with monthly averages 1 standard deviation above the mean look to be unlikely to churn in general and isn't very affected by the change in age.

The majority of the members are placed below 35 years, which means that this plot is highly relevant to be aware of. The younger members are at a higher risk to churn when their monthly average drops below the mean. This suggests that the marketing promotion should attempt to increase the members activity level in order to decrease the churn rate.

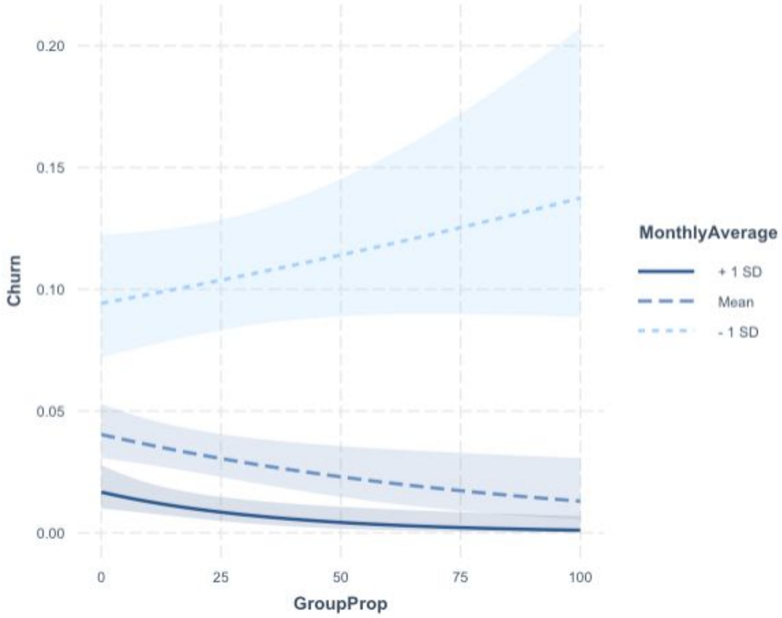
Figure 12. Month 10's Proportion and Monthly Average Effect on Churn



The plot in Figure 12 show the increasing effect that the relative proportions in the 10th month have on the estimation of churn. The relative member proportions that are up to 100 % or below the monthly average show a moderate increase in the likelihood of churn, while the higher the proportional levels gets, the stronger the correlation with churn gets. The members that in addition have had monthly averages 1 standard deviation below the mean, are much more subject to an increase in likelihood of churn with higher relative activity proportions in Month 10. This shows how it's mainly members with vastly different activity levels in the last couple months before the end of the contract that have shown churning behavior, while member above the monthly average mean is affected in a small degree.

However, 75 % of the members in the data had relative proportions in Month 10 at 78 % or below their monthly average, which means that the members that are represented in the higher proportions in the graph is the minority in terms of usage. In general, one can say that the group proportion doesn't have as big an effect on the estimation of churn that the plot indicates, and shouldn't necessarily be used as the basis for marketing promotions. Again, high monthly averages has a more sturdy position and the change in monthly proportion has little effect on the estimation of churn. The promotional activities would therefore still be recommended at increasing the member attendance over the entire contractual period.

Figure 13. Group Proportions Effect on Churn

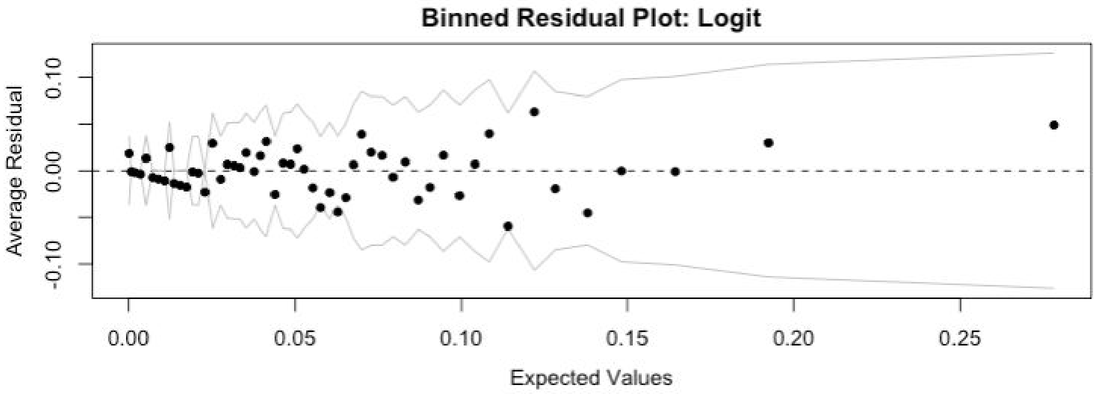


The different levels of group activity proportions and the differing effects for the various usage levels of the monthly average are shown in Figure 13. From the positive value of the marginal effect that "GroupProp" has, one would expect to see an increase in the likelihood of churn when the relative proportions increases, but the plot shows that there are split effects between the different levels of monthly averages. The higher the proportion of group activity attendance, the lower the estimated likelihood of churn is for the members with monthly averages at the mean or 1 standard deviation above the mean. The members with monthly averages below the mean show the expected effect, where the increase in group proportion actually leads to an increase in the likelihood of churn.

This is an important effect to be aware of, as the marginal effect alone doesn't show how the largest portion of members actually will be more likely to remain active if they have high attendance in group classes. Once again, the more active members are not affected as strongly by the change in group proportions as the less active members, which suggest finding a way to increase attendance is the best way to lower the churn rate in the industry.

The AIC score has reduced from 1240.9 from the full logit model, to 1237.7 for the reduced model, which suggest that the reduced logit model is a slightly better fitting model for churn prediction.

Figure 14. Binned Residual Plot: Logit



The binned residual plot for the logit model looks like it provides a pretty good fit for the higher values, but that there are some problem areas where the prediction misses in the lower expected values. The model seems to be able to capture most of the points that are placed inside of the grey bands above 0.03, but there are some areas of the plot where there are outliers which means the model could be improved in order to get a more accurate prediction.

4.3.4. Confusion Matrix: Logit

In order to test whether the model also predicts well outside of sample, it should be tested on the test set that has been separated from the model building. The confusion matrix below shows the distribution of the churn prediction of the binary logit model on the test set.

The threshold value is set to 0.06, so when the probability of survival is placed below the threshold value of 6 %, the members is predicted to churn.

Table 8. Logit: Confusion Matrix

n = 1259	**Actual: No Churn**	**Actual: Churn**
Predicted: No Churn	707 (59.86 %)	21 (26.92 %)
Predicted: Churn	474 (40.14 %)	57 (73.08 %)

The sensitivity of the model explains what percentage of the non-churning cases the model can correctly predict, which is placed at 59.86%, which means that in order for the model to correctly capture churn it is at the cost of misclassifying a considerably large group of non-churning members.

The specificity however, which looks at the percentage of churn cases it correctly predicts is higher, at 73.08 %, which means it is getting about 27% of the actual churn cases wrong. Compared to the baseline model where the specificity of predicting churn cases was 0 %, the binary logit model offers a big improvement of quality and prediction value.

4.4. The Probit Model

The binary probit regression model is similar to the binary logit model, but the probit model uses a different cumulative distribution function. Both the binary probit and binary logit models can capture the nonlinear relationships between the explanatory variables in a model.

The equation for the probit model incorporates the cumulative standard normal distribution function (Φ) with the inclusion of multiple explanatory variables, and restrict the predicted value for churn to one of the two outcomes, 0 or 1.

Which leads to the binary probit equation:

$$Pr(y = 1) = \Phi(\beta_0 + \beta_1 X_1 + ... + \beta_n x_n), \tag{9}$$

where Pr(y=1) is the probability for Churn = 1,

x is the independent variables,

n is the number of independent variables, and

the *βi* can be translated as the effect a one unit change in the independent variable *Xj* has on Φ, if all other independent variables are kept constant.

The estimated coefficients returned from the probit model will be a little more complicated to interpret than the coefficients from the logit model. The coefficients in a probit model affects the z-value of the model, so that for each unit increase in the continuous variables, the z-value will increase by the value of the coefficient. The higher the z-score, the more likely the probability of churn is.

4.4.1. Probit Model: Estimated Results

The first probit model is run as a full binary probit regression model on the training data, where all explanatory variables and interaction terms for usage are included.

Table 9. Estimated coefficients: Full Probit Model

	Estimate	Std.Error	z-value	p-value
(Intercept)	- 1.678	0.375	- 4.477	7.59e-06 ***
GenderF1	0.054	0.089	0.608	0.543
MonthlyAverage	- 0.050	0.010	- 5.022	5.11e-07 ***
Age	- 0.009	0.005	- 1.801	0.072
GroupProp	0.003	0.002	1.177	0.239
PTProp	0.010	0.009	1.211	0.226
Student1	0.591	0.321	1.838	0.066
Private1	0.697	0.312	2.237	0.025 *
Corporate1	0.436	0.314	1.389	0.165
Month12Prop	0.001	0.001	1.335	0.182
Month11Prop	0.001	0.001	1.940	0.052
Month10Prop	0.002	0.001	3.073	0.002 **
GroupProp:MonthlyAverage	- 0.001	0.001	- 1.959	0.050
PTProp:MonthlyAverage	- 0.002	0.001	- 1.435	0.151
Null Deviance	1359.4			
AIC	1247.1			

The p-values show that there are multiple values with high p-values that are not significant for the prediction of churn. According to the full probit model, the most statistically significant explanatory variables are “MonthlyAverage”, “Month10Prop”, and “Private”.

4.4.2. Reduced Model: Probit

In an attempt to improve the full probit prediction model, explanatory variables will be removed from the model one by one, until all variables are statistically significant before analysing the variables effect on churn. The model removes all variables but “MonthlyAverage”, “Age”, “Private”, “Month11Prop”, and “Month10Prop”.

Unlike the binary logit model, the binary probit doesn't find any of the activity types to be significant, but both the overall average monthly usage frequency and the second and third to last activity proportions are determined to be statistically important for the estimation of the churn prediction. Based on the selection of variables in the probit model, observing usage amount under contract is the best way to estimate churn for members. Similarly like the logit model, the last month under contract is not considered significant.

Table 10. Estimated coefficients: Reduced Probit Model

	Estimate	**Std.Error**	**z-value**	**p-value**
(Intercept)	- 0.875	0.147	- 5.965	2.45e-09 ***
MonthlyAverage	- 0.067	0.009	- 7.886	3.12e-15 ***
Age	- 0.013	0.004	- 3.294	0.001 ***
Private1	0.207	0.081	2.575	0.010 *
Month11Prop	0.001	0.001	2.214	0.027 *
Month10Prop	0.002	0.001	3.119	0.002 **
Null deviance	1359.4			
AIC	1246.3			

The estimated parameter coefficients lead to the final probit equation of:

$$Pr(churn=1) = \Phi(- 0.875 - 0.067 * MonthlyAverage - 0.013 * Age + 0.207 * Private1 + 0.001 * Month11Prop + 0.002 * Month11Prop) \quad (10)$$

Similarly to the logit model, the lower the value of the probit equation is for each member, the less likely they are to churn at the end of the contract period.

Simply looking at the coefficients signs, the conclusion can be drawn that the variables "Private", "Month11Prop", and "Month10Prop" increase the z-value and ultimately the likelihood of churn because the coefficients are positive. While increasing the value for "Age" and "MonthlyAverage" decreases the z-value, which in turn affects the estimated likelihood of churn, as these have negative estimated coefficients.

There are multiple continuous variable in the model that have the potential to be the most significant coefficient. "MonthlyAverage" has the highest coefficient value and the activity level can go as high 56, where the coefficient value would be - 0.067 * 56 = - 3.75.

The AIC score of the reduced probit model is placed at 1246.4, which slightly reduced from the AIC of 1247.1 from the full probit model. Removing the insignificant variables has made the reduced probit model a slightly better fitting model than the full probit model.

4.4.3. Average Marginal Effect

Transforming the coefficients using the cumulative normal distribution will return the average marginal effect for the explanatory variables in the binary probit model. Instead of interpreting the model coefficients in terms of their effect on the z-value, the transformation to the average marginal effect allows for investigating the coefficients in terms of their direct effect on the likelihood of churn.

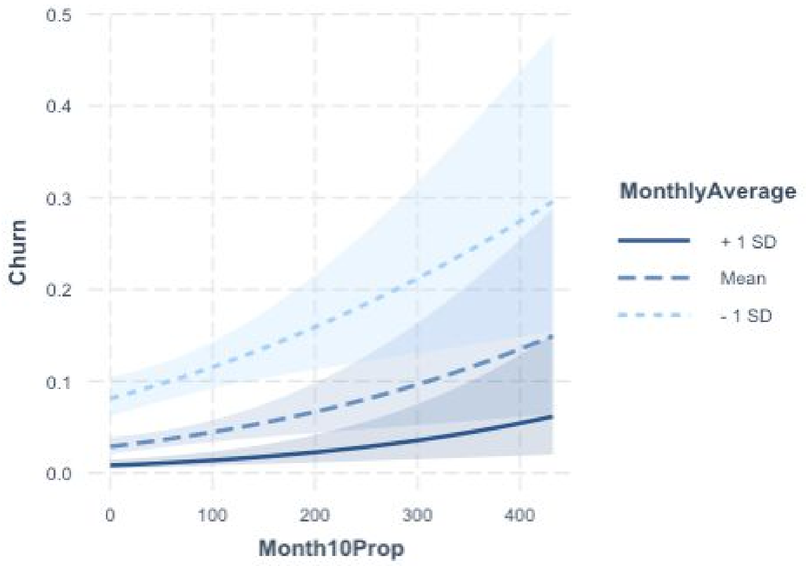
Table 11. Probit: Marginal Effect

Explanatory Variable	**Average Marginal Effect**
(Intercept)	- 0.098
MonthlyAverage	- 0.008
Age	- 0.001
Private1	0.023
Month11Prop	0.0002
Month10Prop	0.0002

The value of the character of the coefficient can be directly applied to the marginal effect, so if a coefficient is negative, it will reduce the likelihood for churn, while a positive coefficient will increase the likelihood of churn. "MonthlyAverage" is again the continuous variable with the highest value of marginal effect that will decrease the likelihood of churn by 0.8 % for each unit increase. The age can consequently be said to lower the likelihood of churn by 0.1% for each yearly unit increase. The marginal effects show similar effects like the binary logit model, where high monthly averages and high age is found to be the best predictors for the estimation of likelihood of churn.

Showing the marginal and interaction effect in a plot will show how the different values of usage affects the estimation of churn.

Figure 15. Monthly Usage Effect on Churn



There is a similar increasing slope of churn for the differing monthly averages depending on the usage proportion in month 10, where the highest estimated churn is for the members with lower averages, which is shown in Figure 15. For the relative proportions placed at 100 % of the monthly average or below, the effect is slight, but increasing. For the members with proportions that are placed several times higher their monthly average, the likelihood of churn has a steeper increasing slope.

As also mentioned in the logit model, the majority of the members have relative proportions placed at 78 % or below of their average, which means it’s just the lower part of the graph that is representable for the majority of the members. This portion of the graph shows a slight increase in estimated churn when the group proportion increases. This highlights the importance of higher monthly averages in order to reduce the likelihood of member churn.

Figure 16. Monthly Average and Month 10 Proportion Effect on Churn

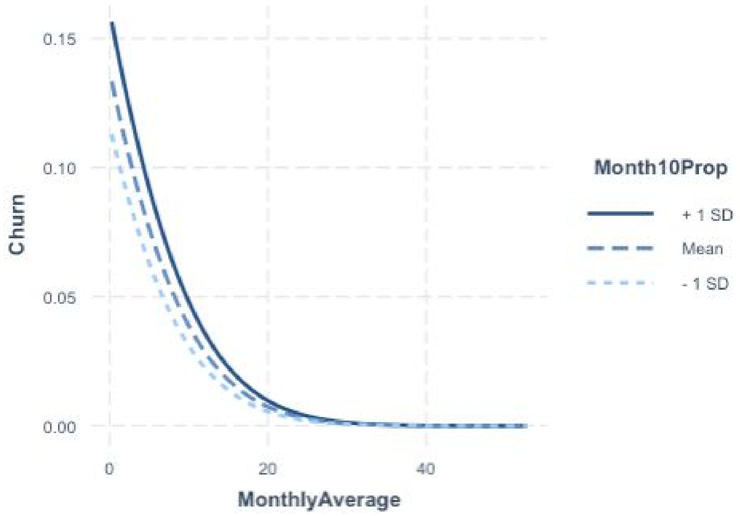
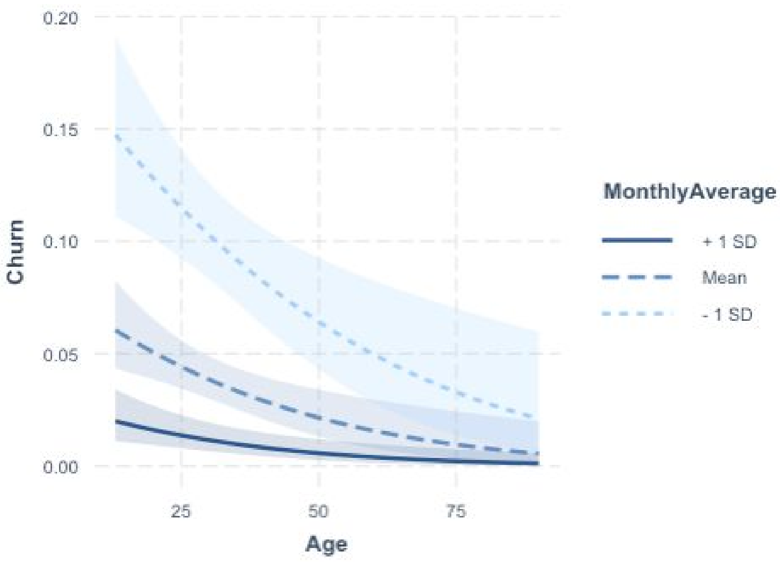


Figure 16 show how the values for monthly average is linked to the estimated likelihood of churn with lines that differentiate between the mean and + - 1 standard deviation of the relative usage proportion in month 10. For the lower values of monthly averages, the members with the highest levels of activity in month 10 is estimated as more likely to churn. As the monthly average increases, the estimated churn has a steep decrease up until around an average of 20. Above this value, it’s hard to distinguish between the three groups as all three are considered unlikely to churn.

The smoothing out of the curve around monthly averages over 20 helps to show more clearly that the proportional effect from Figure 15 isn’t necessarily true for the majority of the members. Members with low monthly averages are actually found to be less affected by the proportion in month 10, as the estimated churn is lower. Essentially, the increase in monthly average is the best way to decrease the churn rate. The members with monthly averages above 20 are generally unaffected by the change in proportion.

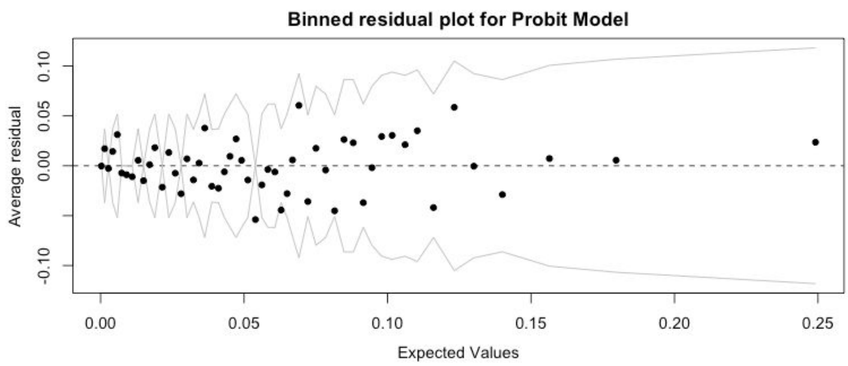
Figure 17. Age and Monthly Average Effect on Churn



The difference in monthly averages and how age plays a role in the effect on churn shown in Figure 17. The younger the customer is, the more likely they are to churn, but higher monthly averages lower the estimated likelihood by quite a significant amount. The members with monthly averages 1 standard deviation above the mean, are as shown in the logit model, not very affected by the increase in age as the likelihood is generally low for all levels.

As mentioned earlier, the majority of the members in the data is 35 years or below. In this age group, there are significant differences in the churn rate depending on their monthly average. From the members 1 standard deviation below and those above the monthly mean, there are about a 8 % difference in the likelihood of churn. In order to reduce the likelihood of churn the promotional activities should be focused on getting the members to increase their monthly average for the contractual year.

Figure 18. Binned Residual Plot: Probit



The binned residual plot for the probit model seems to provide a similar fit like the logit models, where there are still several points that are placed outside of the grey bands in the values below the 0.05 area of expected values. There are a good amount of points that are placed within the grey bin, which suggest that the model provides a pretty good fit, but there will be some degree of misclassification.

4.4.3. Confusion Matrix: Probit

The confusion matrix is conducted on the test set with a threshold value of 0.06.

Table 12. Probit Model: Confusion Matrix

n = 1259	**Actual : No Churn**	**Actual: Churn**
Predicted: No Churn	659 (55.80 %)	20 (25.64 %)
Predicted: Churn	522 (44.20 %)	58 (74.36 %)

The sensitivity of the model is 55.80 %, and the specificity is 74.36 %. Compared to the logit model, the sensitivity of the probit model is slightly lower by about 3 %, but the specificity is slightly higher by 1 %.

The probit model has a slight improvement in the prediction of true positives cases in relation to the logit model, where it is able to predict 1 additional actual churn case correctly than what the logit model could (58 vs 57 out of 78). This makes the probit model better in regard to actual churn prediction than the logit model. The probit model has more false positives than the logit model, which makes it worse in regard to the logit model, as the probit model uses 48 more attempts to get the 1 additional churn cases that are correctly predicted compared to the logit model.

4.5. Classification Tree Model

The classification tree builds a step-by-step graph where each level depicts a binary decision to split the members using the value of the explanatory variables. For creating the classification tree, the model is given access to all explanatory variables, and decides on which ones are statistically significant and should be included in the model in order to create the predictions of churn.

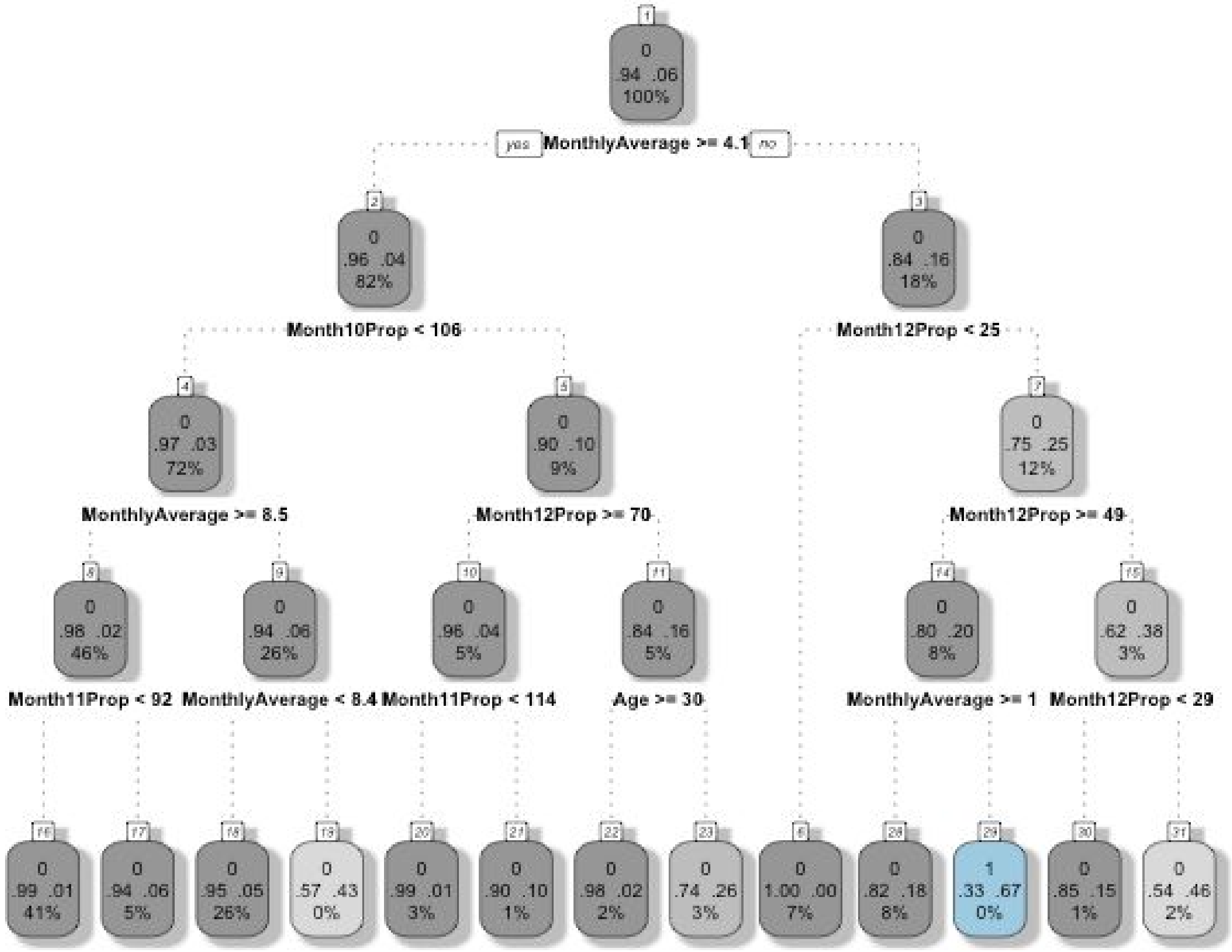
The algorithm behind the tree creates partitioned brackets, by using the most statistically significant variables as “leaves”. For each step of the tree it will pick a threshold value for the chosen significant variable that will split the members with different behavior or characteristics to a specific direction down the tree. At the end of the final step the tree will identify through the value of the final node whether the customer behavior up to that point would lead to a prediction of churn or no churn.

The classification tree model tries to construct a model that has the lowest number of misclassification errors. The misclassification error is based on the Gini-index, which is an error metric that bases each split on the improvement in the purity of the dependent variable class. If a region in the split contains points from mostly one class, the purity is considered high and the Gini index will be low, while a split that contains half of each class will have the highest Gini Index at ½, and be considered highly impure.

The variables that are deemed statistically significant and included in the model is “Age”, “Month10Prop”, “Month11Prop”, “Month12Prop”, and “MonthlyAverage”. These are different from the selection of variables kept in the other models as it doesn’t allow for taking interaction terms into the selection criteria.

The complexity parameter (CP) will stop the growing of the tree if a split doesn’t reduce the overall complexity by a set amount. In general, a smaller tree can better avoid overfitting, but when the data is imbalanced smaller trees result in all members predicted as no-churn, similar to the baseline model. The tree model should include enough information in order to split the data and identify some of the churn cases. Forcing the model to fully grown even though the splits will be impure, will provide more information.

Figure. 19: Classification Tree



For predicting a response using the classification tree model, one starts at the top and move down for each decision statement. If the the binary threshold statement is in true for the member you go left down the tree, or right if it is false. This continues for each node until you reach the bottom where each final node will contain either a 0 if the prediction is "no churn" and a 1 for "churn". Each step of the tree also shows the relative proportion of the total members that are left in each of the leaves at the bottom, and middle values shows the actual distribution of churn and no-churn within that node, which explains the purity of each node along each step.

The top node shows how when all data is included in one node, the proportion between churn and non-churn cases is as seen earlier, 6 % churn, while the remaining 94 % remain as active members. After the first step the model manages to split out 82 % of the data to the left where there are only 4 % of churners present. The right side of the tree has 18 % of the total data where 16 % are churners. Within the second step the right side of the model can with 100 % purity remove another 7 % of the total members that are non-churning to the left. The further down the tree, the leaves become more impure and the model has a harder time accurately predicting the churn occurrences.

In 4 out of the 12 end-nodes that predict no-churn, the purity of non churning members is placed above 98 %, which suggest they are more pure than the initial data and lead to more accurate predictions of the non-churning members than the baseline model. The remaining 8 nodes that predict no churn, have more impure proportions of actual churn cases that are placed in the non-churning node, which suggest that the model could be making mistakes in the predictions for the members with the characteristics that lead to these end nodes. In the node that predicts churn, 67 % of the members are predicted correctly, but the overall percentage of data is placed at 0 %, which suggests that following the graph would misclassify a good amount of members.

Similarly to the results in the binary regression models, the classification tree also favors customers with low activity levels in the months leading up to the renewal decisions. 13 % of the members with less than 50 % relative proportional attendance in the 12th month are all predicted to remain active with 100 % accuracy.

Running the tree models algorithm by lowering the cut off threshold to 0.06 leads to the confusion matrix of the churn prediction below.

4.5.1. Confusion Matrix: Classification Tree

Table 13. Classification Tree: Confusion Matrix

n = 1259	**Actual : No Churn**	**Actual: Churn**
Predicted: No Churn	999 (84.6 %)	39 (50 %)
Predicted: Churn	182 (15.4 %)	39 (50 %)

The classification tree has the lowest number of correctly predicted churn cases between the three models, and significantly lower number of misclassified non-churn cases, as it only misclassified 182 no-churn cases, while both the logit and probit were closer to 500.

The sensitivity of the tree model is 84.59 %, as it is able to predict the majority of the non-churning cases, and doesn't mistake too many of the churning cases for no churn. The specificity of the model is 50 %, which means it is missing half of the actual churn cases, which is significantly worse than the regression models. The classification tree only predict 39 of the churn cases correct, compared to 57 and 58 in the logit and probit model.

The tree model is intuitive to follow along in order to explain what variables are significant in dividing between the members that are likely to churn and those who are likely to stay. It is a model that can be great in provides quick insight for departments that are not sufficiently interested in numbers, and who understand better by visually investigating the tree graph, but it can not compete at the prediction value and quality in this case.

4.6. Model Comparison

The three different tested statistical models for churn prediction: binary logit, binary probit and tree classification have all resulted in churn predictions that provided dramatically more value than what the initial baseline model could. The overall accuracy is lower for all three models, but with significantly higher proportions of specificity, where portions of the small class of churners are identified.

The table below shows all the models side-by-side in order to see the difference between their churn predictions in order to base a recommendation for what model provides the best fit and should be used going forward to get the best result.

Table 14. Comparison of Prediction Models

	Baseline Model	**Logit Model**	**Probit Model**	**Classification Tree Model**
Sensitivity	100 %	59.86 %	55.80 %	84.59 %
Specificity	0 %	73.08 %	74.36 %	50.00 %
AIC	-	1237.7	1246.3	-
# of variables	0	7	5	5

In comparison to the predictions from the simple baseline model, all models performance is of higher quality as they are all able to some extent predict the occurrence of churn cases. Even though the baseline gets 100% of the non-churning cases, this is just a fallacy that relates to the issue of having an imbalanced data set. If the model was to be used on a balanced data set, the baseline model would perform at the same level as random guessing, with about 50 % accuracy.

The logit and probit functions have both shown to be somewhat similar in the estimation of the significant model coefficients and their marginal effect on churn, both in terms of sensitivity and specificity, but the logit model found more variables significant than the probit model. Between the two it’s generally hard to conclude a better fitting model, but based on the logit having a lower AIC score, it is chosen as a better fit for this data. The lowered value can be contributed to the fact that the logit model included group type and interaction terms as statistically significant variables, which the probit model didn’t. The classification tree model partition the most significant variables in a step-by-step figure that is intuitively easy to understand and implement, but has shown to provide the lowest amount of correctly predicted churn cases.

4.6.1. Recommended Model

The best fitting model for predicting the likelihood of churn for contractual memberships with including the customer usage levels, is generally the model that can provide the highest number of actual churn predictions with the lowest misplaced non-churners. Between the logit and probit model, there is 1 churn case difference in correct predictions, but the logit misplaces considerably less non-churners. Based on the model parameters that assess the quality of fit, the logit model would be the recommended model. The logit model offers a lower AIC score which means it's a better quality model. The binary logit model offers a detailed explanation of the link between the usage variables and churn.

5. General Discussion

5.1. Summary

The present study aimed to produce a model that could be used to predict customer churn in any industry that use contractual memberships by using the customer usage under contract. From the lowest AIC value, and the related sensitivity and specificity percentages, the binary logit regression model was found to be the model that would offer the best quality for predicting the link between usage under contract and the associated churn prediction. The value gained from running and analysing the model can be used for both the planning and efficient implementation of future marketing campaigns that aim to reduce the associated churn rate.

Ultimately having a high monthly average over the entire contract period is shown as the most significant for whether a member decides to remain active past the contract renewal. The members with monthly averages 1 standard deviation above the mean were found to be generally unaffected by the change in the usage variables, while the members on the opposite side with 1 standard deviation below the monthly mean were significantly more volatile to the change in usage.

The marginal effects and interaction terms from the binary logit model showed that it wasn't merely low usage that would indicate the most at risk members, as low age would in some cases also be found to correlate with the at-risk churning behavior. The type of activity also had a certain correlating effect, as high levels of group class proportion helped lower the estimated likelihood of churn for the members with monthly averages at the mean or 1 standard deviation above the mean. For members with monthly averages below the mean it did show that higher group class proportions would actually increase the likelihood of churn.

The model found that the last month under contract was not statistically significant for the estimation of the likelihood of churn, which suggests that the renewal choice is made before the 12th month. The activity levels in month 10 and 11 are considered better variables to estimate future churn, which allows for the prediction to be run months ahead before the renewal point and gives the business time to affect the members' renewal decision. The promotional activity should therefore be targeted in advance of the 12th contractual month in order to see the best result in the reduction of churn rate.

The members who don't use their membership evenly over the contractual year are more likely to leave because they're not getting enough value out of their level of usage to account for the associated cost. The ideal promotional activity directed at the low-usage members should be aimed at finding ways to entice members to want to use their membership more frequently, maybe by getting them to attend more group classes, as this has a decreasing effect on their likelihood of churn when the activity is placed at the mean or above (which is about 10 attendances per month).

The predictive quality of the model where it has shown to correctly place 60 % of the non-churners, will reduce the risk of trying to entice members that are already active to become more active. Instead of having to promote higher activity levels for all members, the model will identify and limit the group of members to include the most at risk based on their usage activity.

5.2. Implication

In conclusion, the model offers a way to estimate the effects that usage behavior has on the industry's churn rate. There is value in being able to identify specific segments that are more likely to leave at the end of the contract period, in order to let the marketing and promotional activities be directed with more precision, which will give them a higher chance of returning the wanted effect.

The churn prediction model can easily be converted to other businesses in various industries that have contractual memberships and a way to observe the amount of member usage under contract. As long as the data is transformed to include the same information, which includes a dependent variable of churn and a way to differentiate between the activity proportions, the model should replicate quite easily. It would only require some level of understanding of the statistical software R. The code that was used for the research will be available in the appendix, and would have require some updating to fit new data.

Lowering the threshold cut off value in the models have resulted in a good amount of churners being identified, but at the expense of non-churners being misclassified. Had the cut off value been placed at a higher threshold value it would lower the misclassified proportion of non-churners, but at the expense of losing actual predictions of churn. The sensitivity would increase, but the specificity which is the main goal of the model would decrease with a higher threshold for predicting churn. The consequence of the amount of high misclassified non-churners depends on the marketing approach the business decides to take in order to entice the at-risk members for staying.

5.3. Further Improvements

Because of both the time constraints and the extent of data collected, there were certain limitations that affected the research. There are therefore several areas that could be improved upon if one wanted to get better insight into this type of contractual memberships and its related churning behavior.

5.3.1. Length of Collection period

The collection period of 29 months is relatively short, so there is essential information missing that could have an impact on the churn prediction. There are no members in the sample that have memberships that date back later than May 2017, which consequently tell us that the sample is not a reliable representation of the actual population of members in a company that has been in business for 20+ years.

For a better representation and analysis, the data set would have to be larger in order to include a larger variability in the members tenure. It is likely to assume that members who have been with the company for 5+ years is more loyal and less likely to leave the company even if they would show some months with lower activity.

5.3.2. Switching between contracts

The data set had a column that should include information on whether a customer has downgraded, upgraded, or kept the same type of contract as before. However, this column didn’t have any data in it. It is likely that the change of contract type is statistically significant as an early warning for churn, and would be interesting to include in a model.

If there had been data on switching between contract, the multinomial logit model could have been an interesting option. Would be interesting as there would be several options for the customers at the end of the contract.

choice = 1, churn

choice = 2, take expensive contract (month)

choice = 3, take cheap contract (year)

5.3.3. Cost variables

There are no variables in the data that account for the cost of membership for the different members. If this had been included, one could look at the differing effect between the more expensive and less expensive memberships. This could be in combination with the recording of switching between the contracts. One would suspect that downgrading to a less expensive membership is an early indicator for churn in the future, as this could represent that the member's perception of value is not met with the cost.

The cost variable could also be included as a variable in the confusion matrix where the different marketing efforts related to retaining a member that is predicted to churn would have a different cost or penalize the model for misclassifying certain types of errors. This would be a factor that would be used to further improve the model based on the types of errors that are preferred based on cost.

6. References

Ascarza, E., & Hardie, B. (2009) *Modeling Churn and Usage Behavior in Contractual Settings*. Retrieved from: https://pdfs.semanticscholar.org/3b39/e182a17f3139823b8a482155340173f5bcde.pdf

Baecke, P., & Van den Poel, D. (2011). *Data Augmentation by Predicting Spending Pleasure Using Commercially Available External Data*. J. Intell. Inf. Syst.. 36. 367-383. 10.1007/s10844-009-0111-x.

Beck, C. (2012) *Handling date-times in R.* Vanderbilt Edu. Retrieved from http://biostat.mc.vanderbilt.edu/wiki/pub/Main/ColeBeck/datestimes.pdf

Blattberg, R., Kim, B., & Neslin, S. (2008) *Database Marketing: Analyzing and Managing Customers.* Springer.

Breen, R., Karlson, K.B., & Holm, A. (2018) *Interpreting and Understanding Logits, Probits, and other Nonlinear Probability Models.* Annual Review Of Sociology. Retrieved from: http://repec.sowi.unibe.ch/varia/Odense/Logit-Interpretation-Problems/Breen-etal-2018.pdf

Ciaburro, G. (2018) Regression Analysis with R. Packt Publishing.

Clemente, M. & Giner-Bosch, V., & Matías, S. (2012). *Assessing classification methods for churn prediction by composite indicators*. Retrieved from https://www.researchgate.net/publication/269631632_Assessing_classification_methods_for_churn_prediction_by_composite_indicators

Cook, D. P. Dixon, W. M. Duckworth, M. S. Kaiser, K. Koehler, W. Q. Meeker and W. R. Stephenson. (2001). *Binary Response and and Logistic Regression Analysis.* Chapter 3. Iowa State University. Retrieved from http://pages.stat.wisc.edu/~mchung/teaching/MIA/reading/GLM.logistic.Rpackage.pdf

Curley, E. (2019) *The Gym Membership Retention Statistics Worth Retaining in 2019*. glofox. Retrieved from: https://www.glofox.com/blog/the-gym-membership-retention-statistics-worth-retaining-in-2019/

Date, S. (2019) *The Akaike Information Criterion.* Towards Data Science. Retrieved from: https://towardsdatascience.com/the-akaike-information-criterion-c20c8fd832f2

Engmo, D., & Lysvold, S.S., & Kristoffersen, J.K. NRK. (2019) *Trener for 5 milliarder - Folk jakter paa den perfekte kroppen.* Retrieved from: https://www.nrk.no/nordland/trener-for-fem-milliarder_-_-folk-jakter-pa-den-perfekte-kroppen-1.14612059

Fader, P. & Hardie, B. & Liu, Y., Davin, J., & Steenburgh, T. (2019). *Erratum to '"How to Project Customer Retention" Revisited: The Role of Duration Dependence'* [Journal of Interactive Marketing 43 (2018) 1–16]. Journal of Interactive Marketing. 47. 198. 10.1016/j.intmar.2019.07.001.

Fader, P. & Hardie, B. (2007) *How to Project Customer Retention.* Journal of Interactive Marketing, 76–90. Retrieved from: https://faculty.wharton.upenn.edu/wp-content/uploads/2012/04/Fader_hardie_jim_07.pdf

Fader, P. & Hardie, B. (2009) *Probability Models for Customer-Base Analysis.* Journal of Interactive Marketing, 61–69. Retrieved from https://faculty.wharton.upenn.edu/wp-content/uploads/2012/04/Fader_hardie_jim_09.pdf

Fawcett, T. (2005) *An Introduction to ROC Analysis.* Science Direct. Retrieved from: https://people.inf.elte.hu/kiss/13dwhdm/roc.pdf

Ferri, C., & Flach, P., & Hernández-Orallo, J. (2002) *Learning Decision Trees Using The Area Under the ROC Curve*. Retrieved from: http://dmip.webs.upv.es/papers/ICML2002presentation.pdf

Fernihough, A. (2011) *Simple Logit and Probit Marginal Effects in R.* University College Dublin. Retrieved from: https://www.econstor.eu/bitstream/10419/72194/1/742493032.pdf

Galarnyk, M. (2018) Towards Data Science. *Understanding Boxplots.* Retrieved from https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51

General Data Protection Regulation. (2018) European Union. Retrieved from https://gdpr.eu

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017) *An Introduction to Statistical Learning with Application in R.* Springer.

Menard, S. (2010) *Logistic Regression: From Introductory to Advanced Concepts and Applications.* SAGE Publishing.

Panel Data Models. *Principles of Econometrics with R.* Retrieved from https://bookdown.org/ccolonescu/RPoE4/panel-data-models.html

Provost, F. & Fawcett, T. (2013) *Data Science for Business: What You Need to Know About Data Mining and Data-Analytical Thinking.* O'Reilly Media. ISBN: 978-1-449-36132-7

Statistics Norway, *"Trening og Mosjon"*. Retrieved from https://www.ssb.no/a/publikasjoner/pdf/sa38/Kap7.pdf

Wickham, H., & Grolemund, G. (2016) *R for Data Science. Import, Tidy, Transform, Visualize, and Model Data.* O'Reilly Media, Inc.