24/04/20

Project Contribution Sheet

Yash Kashiwala 19200059

Xianhong Cao 19200508

Daniel Haydock – 12424688

Mengxing Li - 19200093

All team members actively participated and contributed on this project with regular group meetings via skype from start to finish. A google drive was created in the beginning which enabled us to share our work and review each other's work in real time. Communication was excellent via WhatsApp and group members not afraid to ask questions or criticize each other's input. Work was evenly distributed, among the team and all team members came together for lengthy discussions to offer their own insights. Overall, we thoroughly enjoyed working with one another on this project and would have no hesitation working together again on future projects.

Signatures: *Daniel Haydock, Yash Kashiwala, Xianhong Cao Mengxing Li*

TELCO ABC

MARKETING

REPORT

Executive summary

Telecommunication companies around the world face escalating competition which is forcing them to aggressively market special pricing programs aimed at retaining existing customers and attracting new ones, (Rygielski, 2002). Therefore, customer churn prediction has become an essential issue in telecommunication business, (Keramati 2014). Telco ABC has commissioned the following report to develop a data-driven marketing communication strategy to improve customer retention and reduce customer churn.

This report details how Telco ABC customer training data was explored and cleaned. This cleaned data was then analysed using SAS Viya software where models were constructed and the customer data was segmented into chosen 'themes' via clustering.

Executive summary

These themes and models were then analysed further when compared to Telco ABC customer scoring data in order to draw insights and provide marketing communication strategy recommendations to help reduce customer churn.

Table of contents

Introduction

1.1Situation Overview

Telco ABC has been having issues with customer retention. The current churn rate is 49.7% and customer churn has been increasing steadily over time. Telco ABC now seeks assistance in deriving a communication strategy to reverse the current trend.

Using appropriate customer segmentation techniques and predictive analytics, a data driven marketing strategy has been devised to ascertain which of the existing customers are most likely to churn and how best to target customers at risk of churning. At the request of senior management, actionable recommendations have also been provided to assist the Telco ABC customer retention team.

1.2 Report Objectives

1. Assess the situation at Telco ABC and consider the various analytical approaches.

2. Provide a data quality report based on descriptive statistics.

3. Segment the existing customer base via clustering.

4. Conduct appropriate modelling techniques to predict those customers most likely to churn.

5. Provide recommendations to senior management on best to target these customers with a marketing communication strategy and reduce churn rates.

THE DATA LEVERAGED TO DELIVER BUSINESS OBJECTIVES WAS PRODUCED WITH PYTHON AND SAS VIYA SOFTWARE.

1.3 CRISP-DM

The CRISP-DM model for data-mining projects was utilised for this report. An overview of the strategy implemented is outlined below.

1. Business understanding

The initial phase focuses on Telco ABC's main objectives, namely reducing customer churn and requirements from a business perspective in order to ensure targets are met. At this stage an outline of the possible analytics solutions are ordered by importance. A list of the resources available for the project such as available data sources, system dependencies and project risks is made.

2. Data Understanding

In order to explore Telco ABC's existing data, a data quality report based on descriptive statistics for each of the variables is produced. **See appendix (2)** This is done to in order identify data quality problems, discover first insights and detect anything unusual or noteworthy such as outliers that may skew findings.

3. Data Preparation

To ensure model accuracy Telco ABC's data must be filtered first for any missing values and incorrect information. Input and target variables will be selected to reflect the project goal and unnecessary variables will be rejected.

4. Modelling

Modelling techniques are selected and applied to search for any underlying trends and identify insights. Two modelling techniques, decision trees and clustering were chosen for accuracy but also usability, i.e. models must be easily adoptable for Telco ABC’s retention team moving forward.

5. Business Validation

It is imperative that models are assessed at this stage using cumulative lift, and misclassification rate figures. Any business issues that may have been previously neglected may be identified, and a champion model selected.

6. Deployment

Generally projects do not end with just analytical solutions. Knowledge gained will have to be organised and presented in a way that is actionable. The deployment phase can be as simple as generating a report for senior management or as complex as implementing a repeatable process for the Telco ABC customer retention team.

1.4 Data Mining Techniques

Due to the nature of the Telco ABC business problem, the identification and retention of customers likely to return, the best data mining techniques for discovery we determined to be clustering and decision trees. Logistic regression was also explored but ultimately not used to form a marketing communication strategy.

1.4.1 Clustering

Clustering is the division of data into groups of similar objects. In clustering, some details are disregarded in exchange for data simplification. Clustering can be viewed as a data modeling technique that provides for concise summaries of the data, (Berkin,2006). What distinguishes clustering from classification is that clustering does not rely on a predefined classes, (Berry 2004).

1.4.2 Decision Trees

A decision tree is a supervised modelling technique in which an algorithm keeps splitting the data until it reaches pure sets. A decision tree is a tree-like structure which is drawn upside down having root at the top. The root splits into the branches based on the questions asked at each node in order to get a better-predicted value. The end at which the tree doesn’t split further is called a leaf.

The Decision tree is also called as Classification or Regression Trees (CART) as they are used for both classifications as well as regression problems. The most significant advantage of using a decision tree is that it is easy to interpret. The Decision tree clearly shows how the input variable predicts the churn rate of the customer.

1.4.3 Alternative Approaches—Logistic Regression

Logistic regression is better suited to predict the relationship between two variables, in which the target variable is categorical. In Telco ABC's case, the target variable, churn, is categorical, however, Logistic regression is also challenging to interpret from a business perspective and therefore a decision tree is preferred in this situation. Moreover, a decision tree provides greater transparency when predicting the target variable when compared to the logistic regression’s black box.

Data Quality Report

In order to understand the data and how it related to Telco ABC’s objectives, an initial data quality report was conducted in Excel**.** This report highlighted noteworthy results, as well detailed methods used to correct the raw data. **See appendix (2).**

2.1 Raw Data Observations

Using a maximum fetch size, a total of 8017 observations were recorded.

30 variables were observed, which included 8 categorical and 21 numeric and one target variable (churn).

8 categorical variables observed included: whether or not customers had children dependents, lines of credit, credit cards, customer marriage status occupation, and region type.

2.2 Raw Data Corrections

In categorical variables ‘region type’, 's' was converted to 'suburban', 't' was converted to 'town', 'r' was converted 'rural' and 'blank' and 'unknown' were converted to 'missing'.

For categorical variable 'occupation' over 70% of observations were 'blank' and converted to 'missing'.

For the 'marry' variable, 'unknown' observations were converted to 'missing'.

We compared custcareTotal (6 months) with custcareLast (1 month) and found that custcareTotal contained rows with lower values than in custcareLast. This should not be the case and signifies erroneous data.We filtered such rows in order to have clean and meaningful data. The same was done for the following variables (see Table 1).

All binary variables such as ‘creditcard’, ‘children’ and ‘churn’ were converted to ‘1’ or ’0’ for model readability.

(See appendix 2) for raw data analysis and (appendix 3) clean data visualisation

Last 6 months	**Last month (1 month)**
custcareTotal	custcareLast
directas	directasLast
dropvce	dropvceLast
mouTotal	mou
peakOffPeak	peakOffPeakLast
revenueTotal	revenue

Table1

Customer
Segmentation and

Profiling

Cluster 1 - Customer Satisfaction

Variables: 'custcareTotal', 'directas', 'dropvce', 'mouTotal', 'outcalls', 'overage', 'roam'
Cluster sizes k-means: [2573 865 435 428]

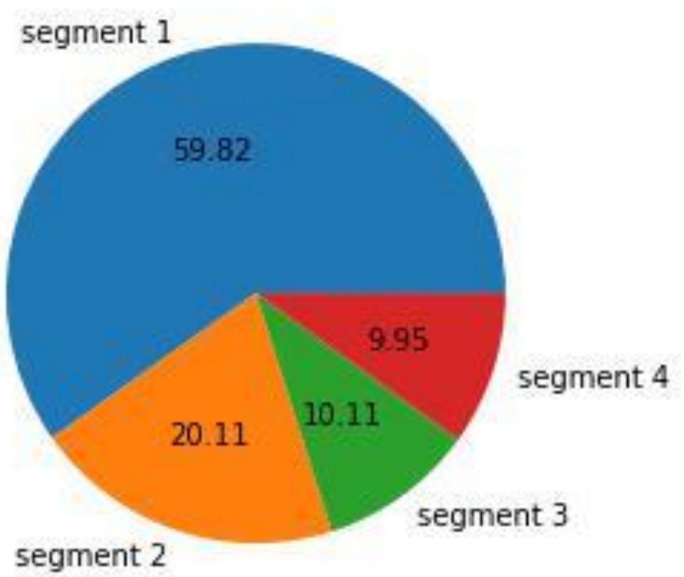


Fig.1 Cluster 1 Customer Satisfaction

Segment 1: 2573 (59.82%)

Observations

- Churn rate 60%, highest in cluster1.
- Highest proportion of customers with high credit scores a, aa: 64.59%.
- Highest average income (mean 4.58)
- Lowest activity in terms of the number of calls to customer calls, calls to directory assisted calls, disconnected calls and out-calls in the past six months.
- Lowest usage in terms of mean of the total number of minutes used in the last 6 months (40% of the total average).
- In the last month, the number of minutes over the customer's bundle used was also the lowest when compared to other segments (mean 6 minutes).

Recommended marketing activity

Credit and average income suggest that these members have well-paid jobs and a high level of disposable income. Despite this level of disposable income, customers in this segment, seldom use Telco ABC's services and maintain a high departure rate, which shows that our existing communication package does not attract them at all. In this case, it is suggested that Telco ABC provides these customers with exclusive new products after a return call to understand their needs.

Segment 2: 865 (20.11%)

Observations

- Churn rate: 46.59%, the only segment in cluster1 that has more people staying than leaving.
- In terms of user activity, this segment performed in stark contrast to segment1, providing the highest mean of the number of calls to customer calls and disconnected calls, which were both 3x the overall mean.
- Highest proportion of customers with low credit ratings (gy, z): 9.13% On average, this segment has the lowest mean of income.

Recommended marketing activity

High-frequency customer calls and disconnected calls show that customers in Segment 2 often encounter problems when using Telco ABCs services. Credit and income data tell us that this group may be very price-sensitive. To better capture these customers, affordable packages should be offered. The reasons for poor connection service should be determined. The location of these customers was examined assuming that location would be a factor for poor connectivity but there were too many missing values under regiontype. Customers should be surveyed to get a better idea of where Telco ABCs customers are worst affected.

Segment 3: 435 (10.11%)

Observations

- Churn rate: 56.32%
- Roaming rates are 6x the overall mean.
- Directory assisted calls at approximately 5.4x the overall mean.
- Lowest proportion of low-credit customers (gy, z), (64.69% of the overall mean).

Recommended marketing activity

It is fair to assume that these customers travel a lot due to the extremely high roaming rate of this group and the number of times these customers use directory assisted calls. Customers in this category should be targeted with tailored roaming and directory assisted call packages to reduce the churn rate.

Segment 4: 428 (9.95%)

Observations

- Churn rate: 59.6%
- Very high the number of minutes over the customer's bundle used this month (5.9x the overall mean).
- Average total call time in the past six months is the highest in cluster1, more than twice the overall average.
- Have a relatively small proportion of highly creditworthy customers (a, aa).

Recommended marketing activity

Customers in Segment 4 should be targeted with packages tailored for high call volumes. An 'unlimited minutes' package should be considered after taking into account rates of return. It is believed that such an offering will significantly reduce the churn rate of this group.

Cluster 2 - Customer contribution

Variables: 'revenueTotal', 'revenueChange', 'mouTotal', 'mouChange', 'outcalls', 'recchrge', 'peakOffPeak'
Cluster sizes k-means: [2085 495 459 1262]

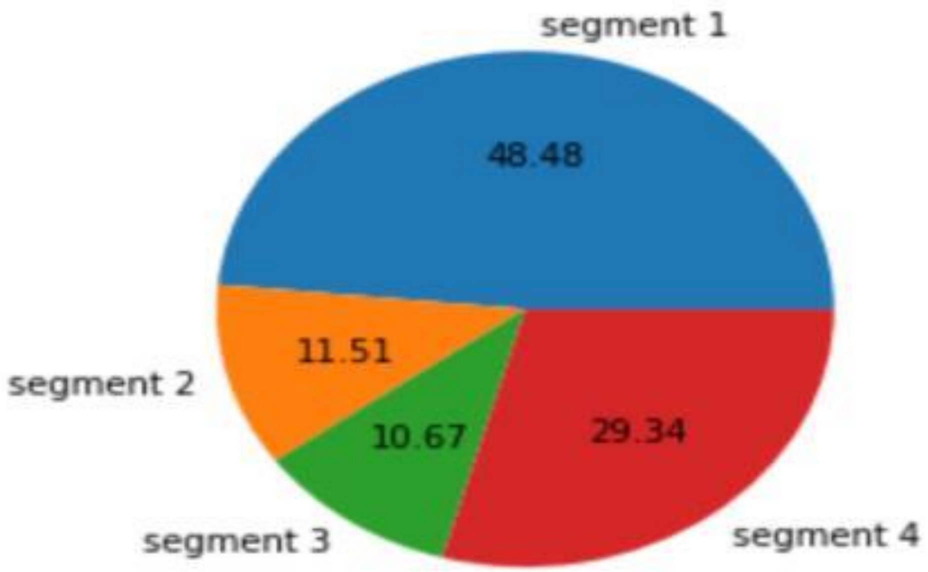


Fig.2 Cluster 2 Customer Contribution

Segment 1: 2085 (48.48%)

Observations

- Churn rate: 65% (this churn rate is about 10% higher than the overall rate).
- The lowest average call time last month among the four segments, (roughly 40% of the overall average).
- The lowest number of minutes of bundle usage this month (less than 1/3 of the overall average.)
- The lowest number of the average value of peak calls this month in cluster 2 (0.6x the population mean).
- Average revenue contributed by this segment last month is much lower than the overall average,(only about 60% of the overall average).

- Average income is higher than the overall level.
- 63.93% of customers have higher credit ratings (a,aa).

Recommended marketing activity
High churn rates, reduced bundle utilisation, and various call measures indicate that Telco ABC's existing products and services could be improved. This customer group accounts for nearly half of the total customer base, and the high churn rate leaves a big window of opportunity. As this group has high income and high credit ratings, the emphasis should be on service and not just price. A survey could be devised to ask this cohort what extra services they would like to see Telco ABC offer them. It is not recommended to sell existing products to them, as that would be a waste of our company's valuable marketing resources.

Segment 2: 495 (11.51%)

Observations

- Churn rate: 55.2%
- Only 3.4% of customers have a low credit rating (gy, z); (This figure is the lowest of the four segments).
- 70.1% of customers have a high credit rating (a, aa), (This figure is the highest in cluster2).
- The two previous points show that the customer reputation of this group is outstanding.
- On average, income is the highest of the four segments.
- Low usage of customer's bundles, this group provides the second-lowest mean, almost 1/2 of the overall average.
- The average value of peak calls is 3.5x the population mean this month, which is the highest one within the four segments.

Recommended marketing activity

Given the low usage of customer bundles but a high number of peak calls, it is recommended that these customers are better informed about Telco ABC's bundles that could save them money. Our marketing strategy is to recommend a telecommunications package with a high peak call quota for them. By the way, the quality of calls during peak hours should be ensured at the technical level.

Segment 3: 459 (10.67%)

Observations

- Churn rate: 100%
- Customer-service phone use averaged 0.61x the overall mean for the past six months, but last month's figures alone were twice the overall average. This suggests that the number of calls they have made to customer service has surged in the last month.
- Reported mean of drop calls in the last 6 months are 2.34x the overall average.
- Customer bundle usage this month is 2x the overall average.
- Contribution to revenue last month is 1.5x the overall average.
- The average value of roaming rates is 80% higher than average, (the highest of the four groups).

Recommended marketing activity

These customers are not satisfied with the current service, but due to their significant contribution to revenue should not be ignored. It should be explored as to why the customer service calls increased last month. Many dropped calls, excessive usage of bundles, and roaming demand indicate there are many travellers in this group. Competitor roaming packages should be reviewed and benchmarked against to help reduce customer churn in this cohort.

Segment 4: 1262 (29.34%)

Observations

- Churn rate: 26.78%,(roughly 60% of the overall average).
- The only segment in cluster2 that has more people staying than leaving.
- The mean of customer calls was high for the last 6 months, it is the only segment in cluster2 with a mean above the overall mean, 2.6x the overall mean.
- The average call time for last month was 1.85x the overall mean, highest among the four segments.
- Bundle usage for this month is 2x the overall mean.
- Revenue was high last month, 1.5x the overall mean.
- The average customer in this segment has a low income.

Recommended marketing activity

Although the churn rate for this group is low, there is still room for improvement. In the past month, there has been good demand for bundles and call time. This segment likes to make customer service calls. These customers should be encouraged to use less costly forms of communication, such as email or SMS. Since average income is low, Telco ABC should focus on pricing when targeting this segment as customers are more likely to be price sensitive.

Modeling

Predictive models provide measurable metrics for Telco ABC's retention team in their efforts to reduce customer churn. Models are used to identify patterns that if spotted, will alert the retention team to those customers most likely to churn. This data-driven approach means that Telco ABC will not have to rely on hunches when deploying their marketing communication strategy.

Decision Tree

In this report to predict the churn rate, different models for decision trees were built keeping different variables.

Prior to building the model the data was partitioned into Training, Validation and Test in order to ensure model accuracy with Training (60%), Validation (30%), Test (10%) partitions.

Firstly, the variables which were not contributing to the churn rate were removed in order to build an accurate model and with high precision.The main advantage of decision tree is that the algorithm is capable of handling categorical as well as continuous data as well as correctly handles missing values and outliers without any need to transform data. **(see appendix 4)**.

At First, A default decision tree was generated which gave importance to the columns such as ‘RevenueTotal’, ‘revenuechange’ and produced an accuracy of about 89%, but for the business perspective and keeping in mind the other departments, it is hard to deal with the information provided by the decision tree.

Keeping in mind the business objectives and departments, Another Decision tree was created using variables focusing on customer satisfaction and more inclined towards the business objectives such custcaretotal, dropvce moutotal, outcalls and overage

The Automatic decision tree generated using the above variables, gave a misclassification rate of 0.2126 with an accuracy of 0.78 **(see appendix 4)** even though the accuracy for this decision tree is lower than the previous one but it would be easy to interpret to the business people.

To follow up with these variables and to improve the accuracy and misclassification rate of the tree, following measures were taken.

The two Interactive Decision tree was created by changing the splitting criteria and fine-tuning some of the decisions made while splitting the node.

The same variables were used to form a Gradient Boosted tree. This model generated a misclassification rate of 0.2362 with an accuracy of 0.76.

Out of all the above decision trees, the automatic decision tree generated the best accuracy, even misclassification rate and error of prediction for churn rate is lower in the Automatic Decision tree and thus was selected as the champion model. **(see fig. 3)** for overall statistics and **(see appendix 5)** for detailed statistics**.**

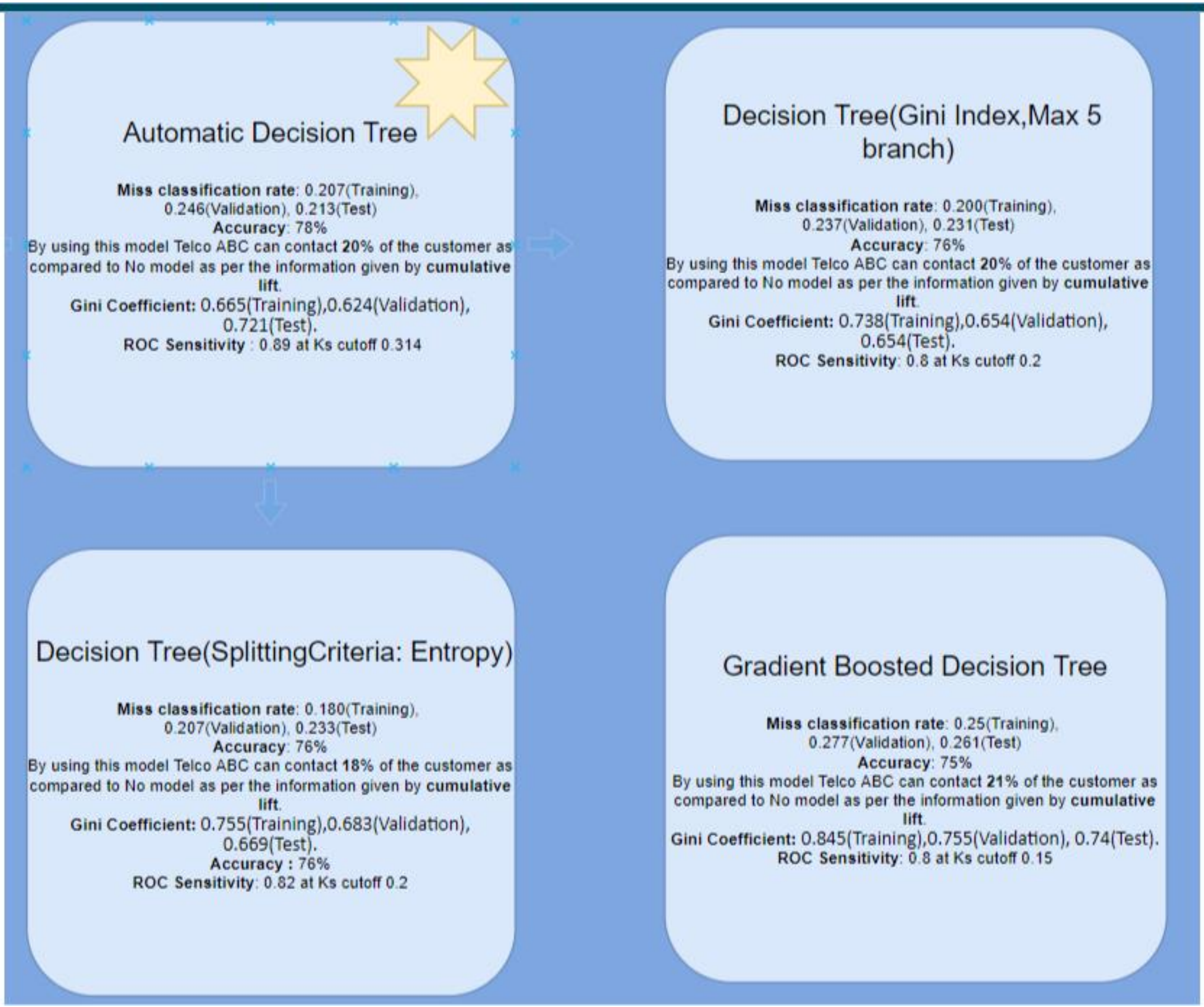


Fig. 3 Different Trees with Statistics

Model Development

The SAS VIYA Learner Scoring data was used to deploy customer data and ensure that Decision trees were correctly assembled for the purpose of the task.

The scoring data and its predicted values can be found in **Appendix 6**.

5.1 Scored Data Descriptive Feature:

Overall Churn rate in scored data Predicted.

I_churn means the predicted churn value by the model **(see figure 4).**

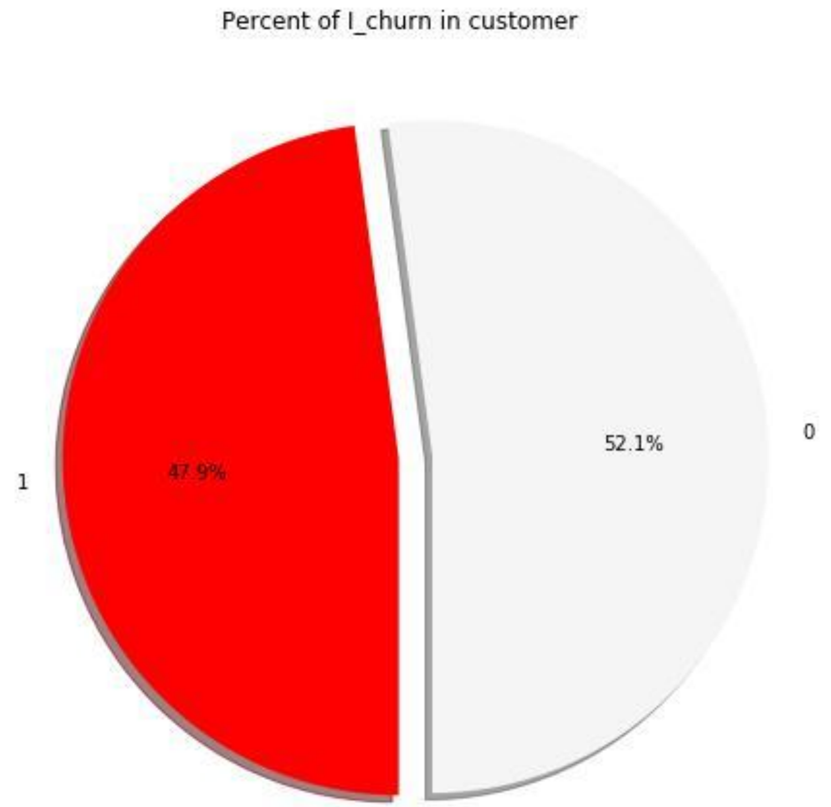


Fig.4 Churn rate overall statistics

The figure 5 shows that for the scored data, the number of customers who tend to make less calls to the customer service also shows a lower retention rate as compared to the customers making customer care calls.

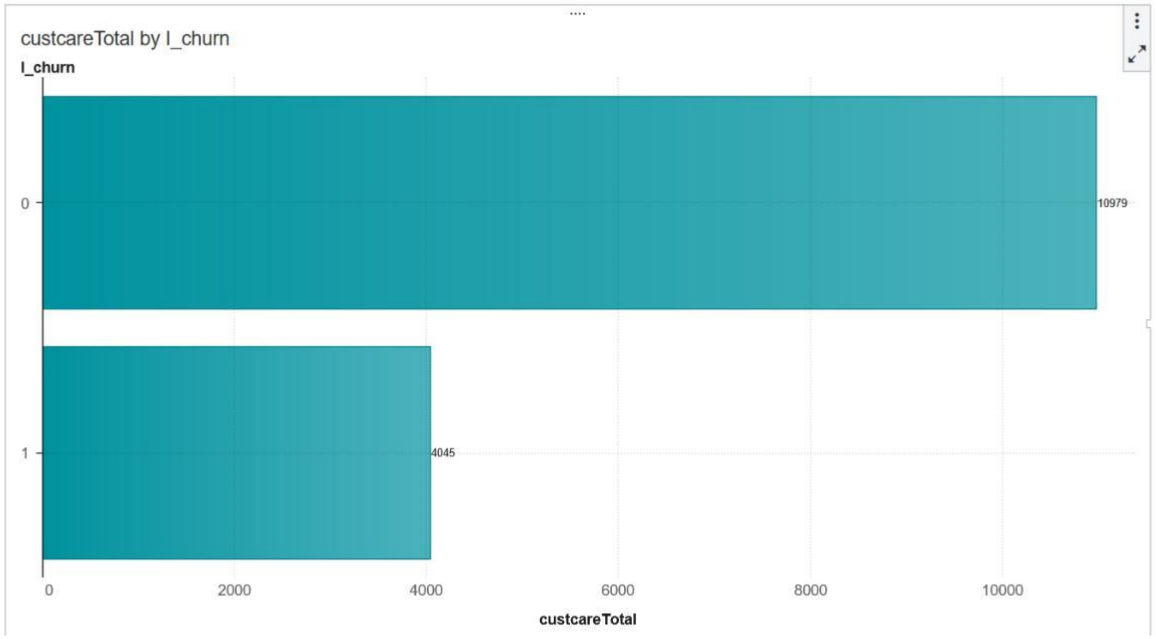


Fig. 5 Rate of Churn as compared to custcareTotal Calls

More customer calls lead to higher customer loyalty as relationships are built between customers and representatives(Koi-Akrofi,2013). This correlation indicates that unsatisfied customers tend to churn before looking for help. Insights gathered from customer interaction through calls can have a significant impact on customer services quality evaluation, (Gnroos, 2000).

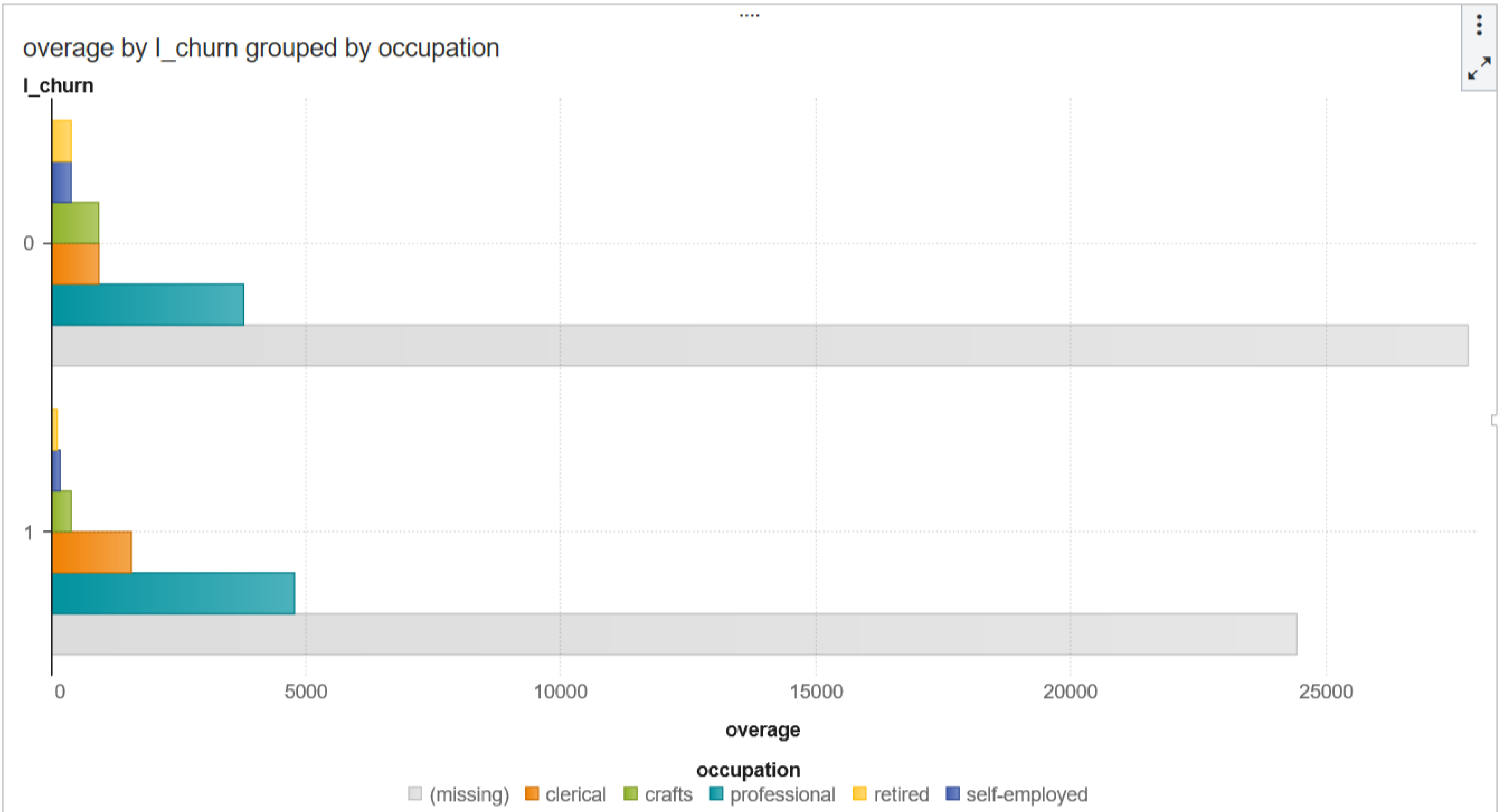


Fig. 6 Rate of Churn as compared to Overage within the group of occupation

With regards to the variables (see Fig. 6), 'occupation' the score data predicts that 'professionals' tend to have the highest churn rate, closely followed by people in 'clerical' positions.' 'Professionals' tend to have more ‘overage’, which means the number of minutes used by them exceeds the bundle pack offered by Telco ABC.

Business bundles should be reviewed to retain more professionals, and perhaps a higher usage 'premium business bundle' should be considered. Competitor business bundles should also be investigated as part of a benching marking exercise for Telco ABC.

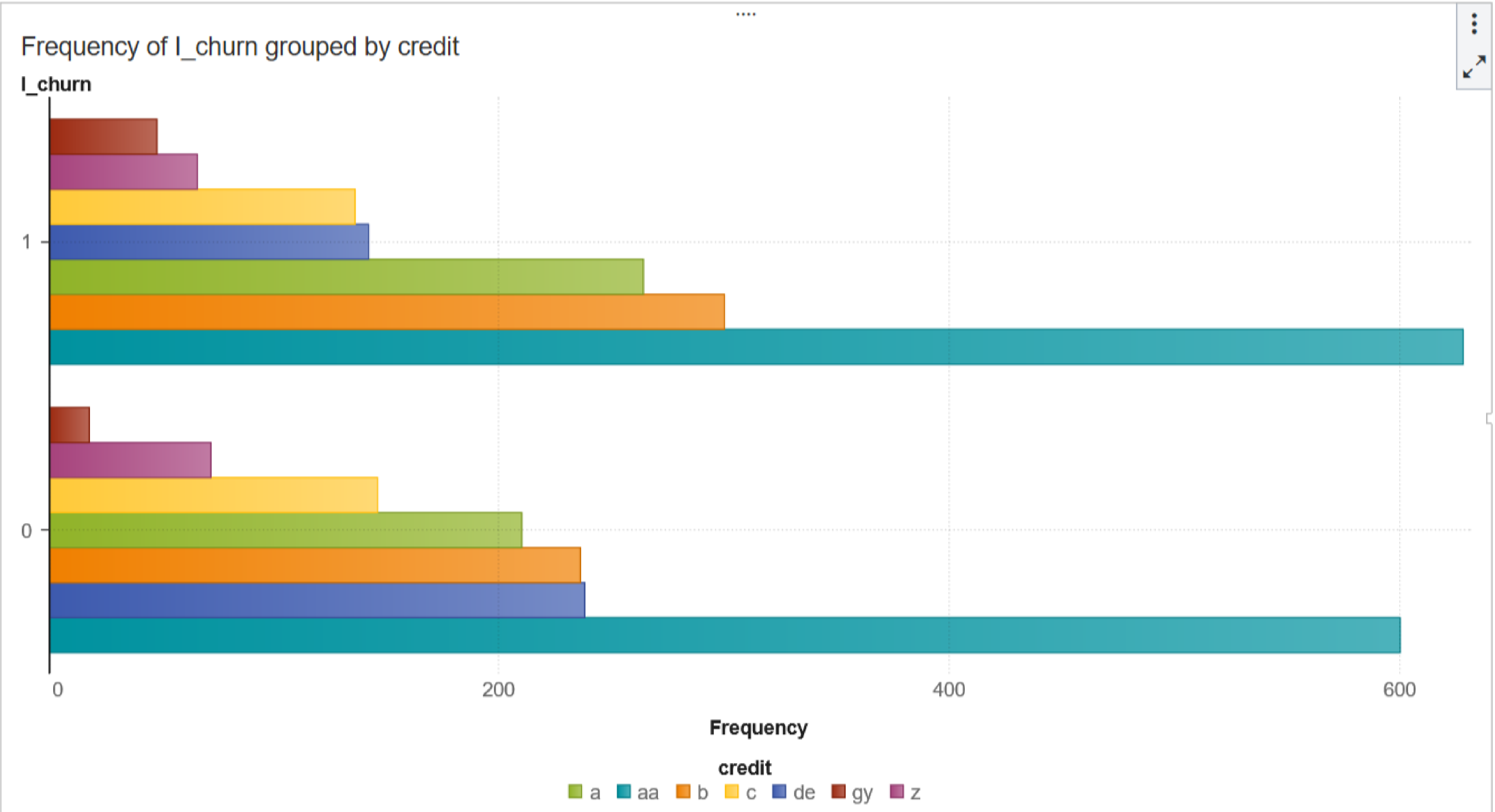


Fig.7 Churn rate vs. Credit

Based on the predictive churn rate (as seen in Fig. 7) there is a correlation between better credit ratings (aa, a) and a customer's likelihood to churn.'

Telco ABC should place particular emphasis on customers with high credit ratings. These customers are more than likely by incentivised by quality of service over price. Customers expect value for money, if a customer is expected to pay higher rates, then they expect an equivalent return in service quality, (see appendix 13). Telco ABC should look to communicate directly with these customers to see where their service can be improved to reduce churn.

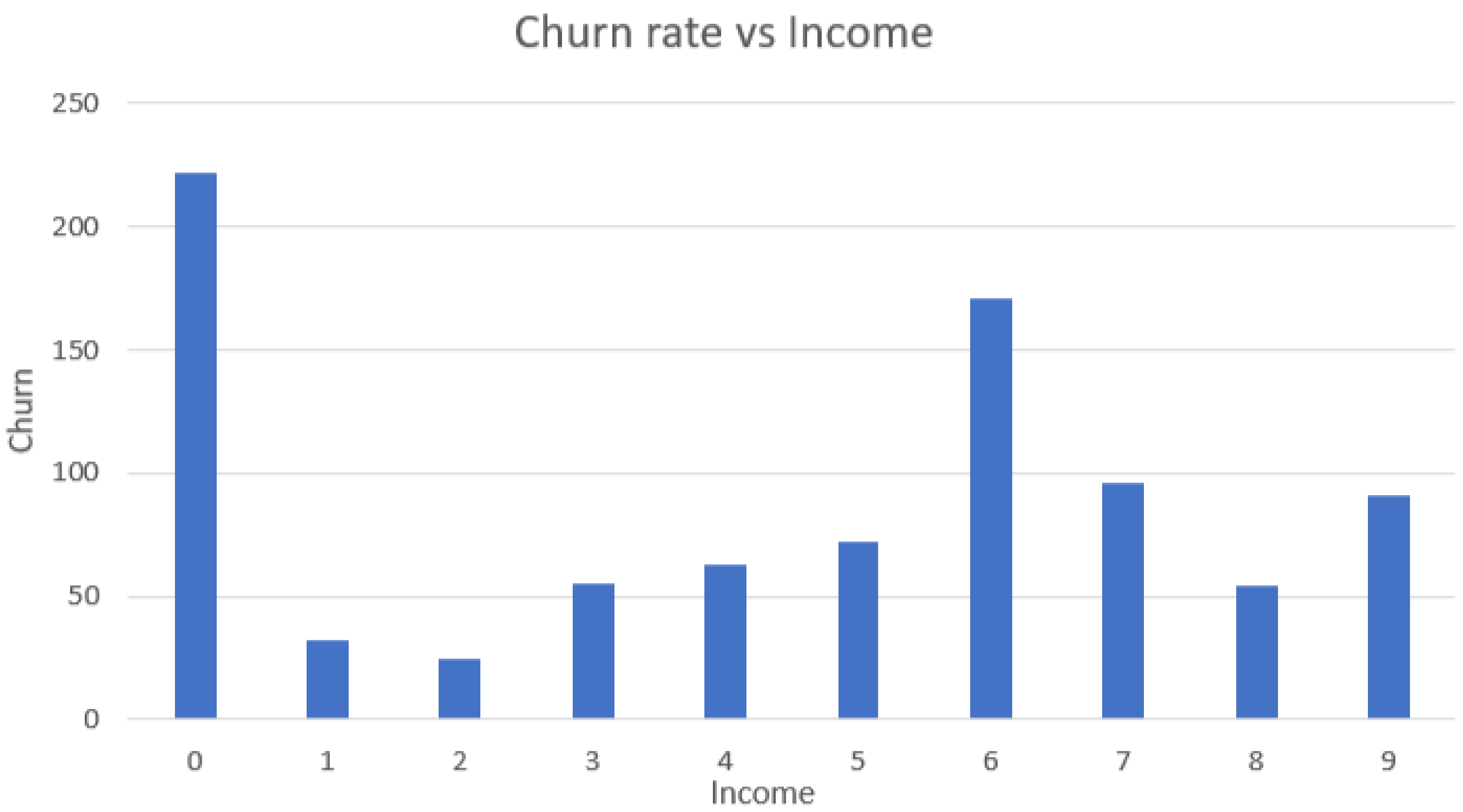


Fig.8. Predicted churn rate vs Income band

As seen in figure(8) customers in the lowest income band, '0' tend to churn more than customers, other income bands. These customers are more likely to be price-sensitive and may even be a risk of defaulting on their Telco ABC contracts. Customers should be provided with individual assistance to avoid unnecessary churn in this category. Allowances for late payments should be considered and a framework put in place for those most vulnerable to defaulting on their monthly payments. If it is the case that these customers are leaving for cheaper deals, it is essential to let these customers know that they are cared for as further price cuts may lead to a price war and subsequent race to the bottom, (Maille, 2008).

Decision making by using Scored data sets

Advantages

The main advantage of making decisions using a scored data set is that it helps Telco ABC to target a particular customer group based on the predicted churn rate. This helps Telco ABC take appropriate actions or prepare a marketing strategy ahead of customer churn.

Disadvantages

One cannot solely rely on the predictive value for identifying those most likely to churn. Domain experts should be consulted when devising a marketing strategy.

Biases may arise when one is presented with a predictive value and might ignore the customers who have been predicted as low risk of churning.

Use the scoring (unseen) data set and apply the Cluster models

Putting scoring data in Cluster1 model

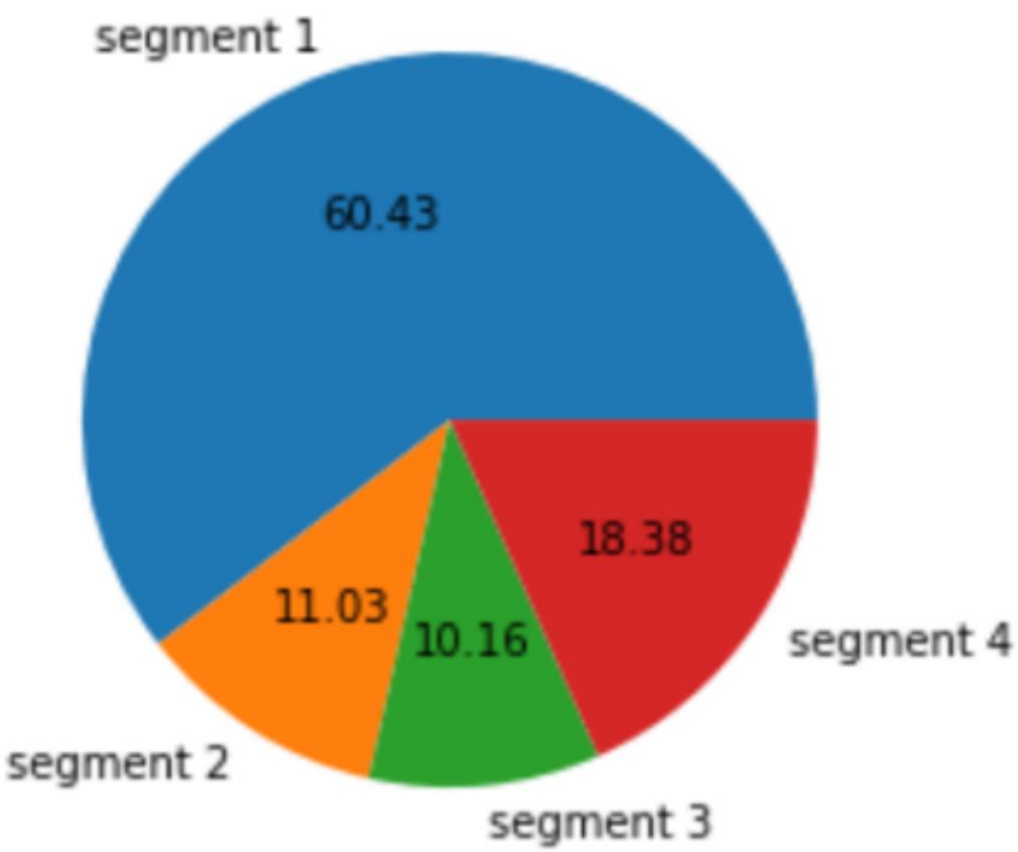


Fig. 9 Cluster 1 Visualisation

```
X_0.groupby('segment')['churn']
```

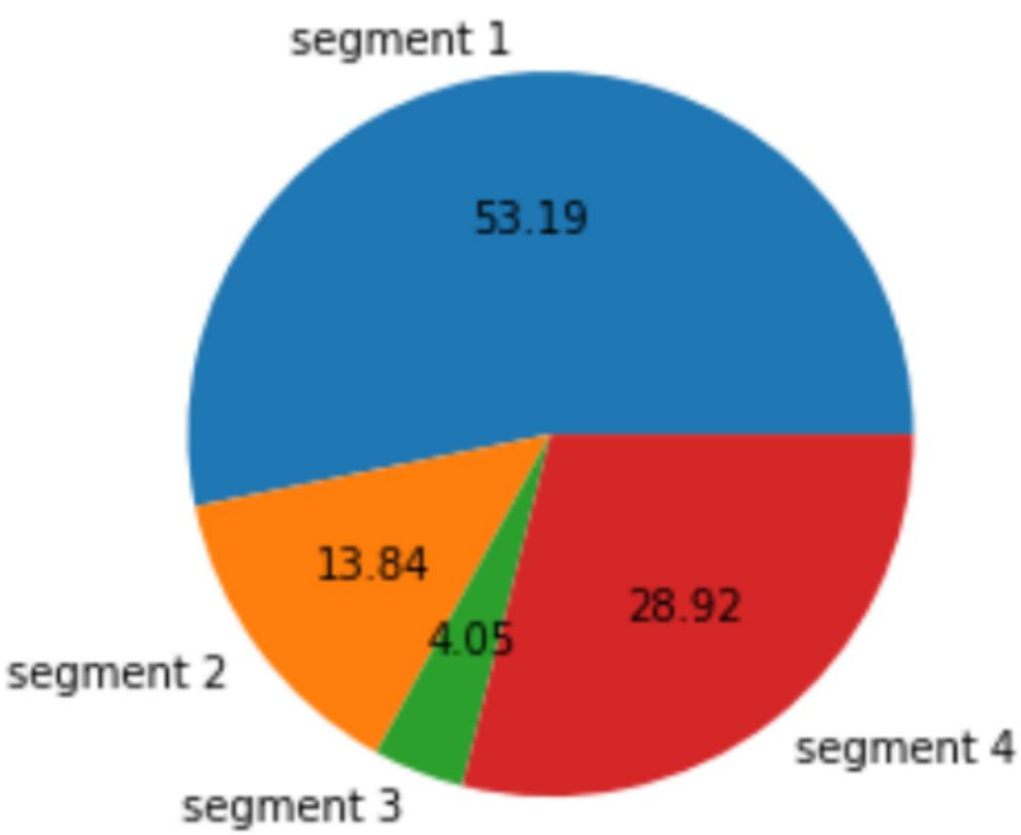
```
segment  churn
0        1        620
         0        498
1        1        111
         0         93
2        0        130
         1         58
3        0        190
         1        150
```

Fig. 10 Cluster 1 Statistics

Segment 1 (scoring) 1118: churn rate 55.47%
Segment 2 (scoring) 204: churn rate 54.41%
Segment 3 (scoring) 188: churn rate 30.85%
Segment 4 (scoring) 340: churn rate 44.12%

The Cluster 1 model is tested with scoring data, and the results about segment 1 and segment 2 are basically the same as those obtained with cleaned training data. There is a certain deviation for segment 3 and segment 4(see fig 9 and fig 10).

Putting scoring data in Cluster2 model




```
X_0.groupby('segment')['churn'].
```

```
segment  churn
0        1         589
         0         395
1        0         148
         1         108
2        1          75
3        0         368
         1         167
Name: churn, dtype: int64
```

Fig. 11 Cluster 2 Visualisation

Fig. 12 Cluster 2 Statistics

Segment1(scoring) 984: churn rate 59.86%
Segment2(scoring) 256: churn rate 42.18%
Segment3(scoring) 75: churn rate 100%
Segment4(scoring) 535: churn rate 31.21%

The Cluster 2 model is tested with scoring data, and the results about segment1, segment 3 and segment 4 are basically the same as those obtained with cleaned training data. There is a certain deviation for segment 2.

Conclusions

The cluster and decision tree models used in this report are accurate in identifying current Telco ABC customers most at risk of churning and recommendations on how to reduce churn rates through targeted marketing strategies. It is vital that Telco ABC reduces ongoing customer churn of 47.9%. It has been repeatedly indicated that customer retention in comparison to absorbing new customers is significantly more achievable and less expensive (Keramati, 2014). The key to retaining customers has been identified within each segment under Cluster 1 ‘Customer Satisfaction’ and Cluster 2 ‘Customer Contribution’ each segment supplies recommendations on how best to target that segment and why.

Modelling through the use of decision trees provides actionable recommendations for the Telco ABC customer retention team when compared to scoring data. If these recommendations are executed in a timely and systematic manner Telco ABC will successfully reduce customer churn rates.

Appendices

Appendix 1 Data Dictionary Colours signify pairs where there was data recorded for 1 month and 6 months.

Variable	Type	Description
customerID	Numeric	Customer ID
children	Categorical	There are children present in the customer's household {true, false}
credit	Categorical	The customers credit rating {a, aa, b, c, de, gy, z}
creditCard	Categorical	The customer owns a credit card {true, false}
custcare	Numeric	average number of calls to customer calls in the last 6 months
custcareTotal	Numeric	total calls to customer calls in the last 6 months
custcareLast	Numeric	calls to customer calls in the last month
directas	Numeric	The number of directory assisted calls made in the last 6 months
directasLast	Numeric	The number of directory assisted calls made last month
dropvce	Numeric	The number of calls dopped in the last 6 months
dropvceLast	Numeric	The number of calls dopped the last month
income	Numeric	The cutomer's income {0 - 9}
marry	Categorical	The customer's marital status {yes, no, unknown}
mou	Numeric	Number if minutes last month
mouTotal	Numeric	The total number of minutes used in the last 6 months
mouChange	Numeric	% change in minutes
occupation	Categorical	The occupation of the cutomer { clerical, crafts, homemaker, professional, retired, self-employed, student}
outcalls	Numeric	The number of calls made
overage	Numeric	The number of minutes over the customer's bundle used this month
overageMax	Numeric	Max overage
overageMin	Numeric	Min overage
peakOffPeak	Numeric	The total number of peak calls made the last 6 months
peakOffPeakLast	Numeric	The total number of peak calls made last month
recchrge	Numeric	The recurring bundle charge this month
regionType	Categorical	The type of region in which the customer lives {rural, suburban, town}
revenue	Numeric	Reveue from customer last month
revenueTotal	Numeric	total revenue in the last 6 months
revenueChange	Numeric	% change in revenue
roam	Numeric	The number of roaming events in the time period
churn	TARGET	Flag indicating if the customer has churned

Appendix 2 Raw Data Analysis

Row Labels	Count of children	% of Grand Total
FALSE	6086	75.91%
TRUE	1931	24.09%
Grand Total	**8017**	**100.00%**

Row Labels	Count of credit	% of Grand Total
aa	3059	38.16%
a	1372	17.11%
b	1353	16.88%
de	923	11.51%
c	789	9.84%
z	326	4.07%
gy	195	2.43%
Grand Total	**8017**	**100.00%**

Row Labels	Count of creditCard	% of Grand Total
TRUE	5339	66.60%
FALSE	2678	33.40%
Grand Total	**8017**	**100.00%**

Row Labels	Count of occupation	% of Grand Total
missing	5931	73.98%
professional	1361	16.98%
crafts	226	2.82%
clerical	168	2.10%
self-employed	155	1.93%
retired	105	1.31%
student	60	0.75%
homemaker	11	0.14%
Grand Total	**8017**	**100.00%**

Row Labels	Count of regionType	% of Grand Total
unknown	3914	48.82%
suburban	2473	30.85%
town	1196	14.92%
rural	388	4.84%
s	36	0.45%
t	6	0.07%
r	4	0.05%
Grand Total	**8017**	**100.00%**

Appendix 2 Raw Data Analysis

Row Labels	Count of churn	% of Grand Total
FALSE	4029	50.26%
TRUE	3988	49.74%
Grand Total	**8017**	**100.00%**

Appendix 2 Raw Data Analysis

variable name	Role	Number of levels (include missing)	Missing	Missing percentage	Mode	Mode percentage	Mode2	Mode percentage	table 1 : categorical features
children	input	2	0	0	0	75.91%	1	24.09%	False-0;True-1
credit	input	7	0	0	aa	38.16%	a	17.11%	
creditCard	input	2	0	0	0	33.40%	1	66.60%	False-0;True-1
marry	input	3	0	0	2	39.23%	1	35.94%	no-0;yes-1;unknown-2
occupation	input	8	5931	73.98%	professional	16.98%	crafts	2.82%	Missing means blankspace
regionType	input	7	3839	47.89%	suburban	30.85%	town	14.90%	Missing means blankspace
churn	output	2	0	0	0	50.26%	1	49.74%	False-0;True-1

Appendix 2 Raw Data Analysis

	mean	std	min	25%	50%	75%	max	missing	table 2 : continuous
custcare	1.7494075	6.2256	0	0	0	1	376	0	
custcareTotal	8.140701	33.829	0	0	0	6	2253	0	
custcareLast	1.7329425	6.2598	0	0	0	1	387	0	
directas	0.9017089	2.2346	0	0	0	1	55	0	
directasLast	0.8767619	2.2062	0	0	0	1	52	0	
dropvce	6.0253212	8.7648	0	1	3	8	114	0	
dropvceLast	5.9706873	8.7647	0	1	3	8	110	0	
income	4.2800299	3.1464	0	0	5	7	9	0	
mou	522.62784	545	0	149.45	363	711.69	6494.3	0	
mouTotal	2399.8305	2897.8	0	511.45	1447.8	3246.5	38966	0	
mouChange	0.6173898	15.919	-1	-0.0649	0	0.0659	967.42	0	
outcalls	159.34327	189.81	0	37	105	215	2716	0	
overage	42.037077	110.62	0	0	0	42.881	4313.1	0	
overageMax	44.668695	116.82	0	0	0	45.993	4565.2	0	
overageMin	39.390298	104.67	0	0	0	40.187	4098.3	0	
peakOffPeak	2.0850975	2.889	0	0.7602	1.3727	2.4566	79.333	0	
peakOffPeakLast	2.0875546	2.8926	0	0.759	1.3725	2.4615	79.333	0	
recchrge	46.263166	24.007	0	30	44.99	59.99	337.98	0	
revenue	61.740842	43.8	0	38.015	50.787	72.892	577.21	0	
revenueTotal	279.04486	243.16	0	121.12	239.94	359.94	3463.3	0	
revenueChange	0.0433287	0.5646	-1	-0.0168	0	0.0168	14.463	0	
roam	1.1554197	5.7708	0	0	0	0	149	0	

Appendix 3 Cleaned Data Analysis

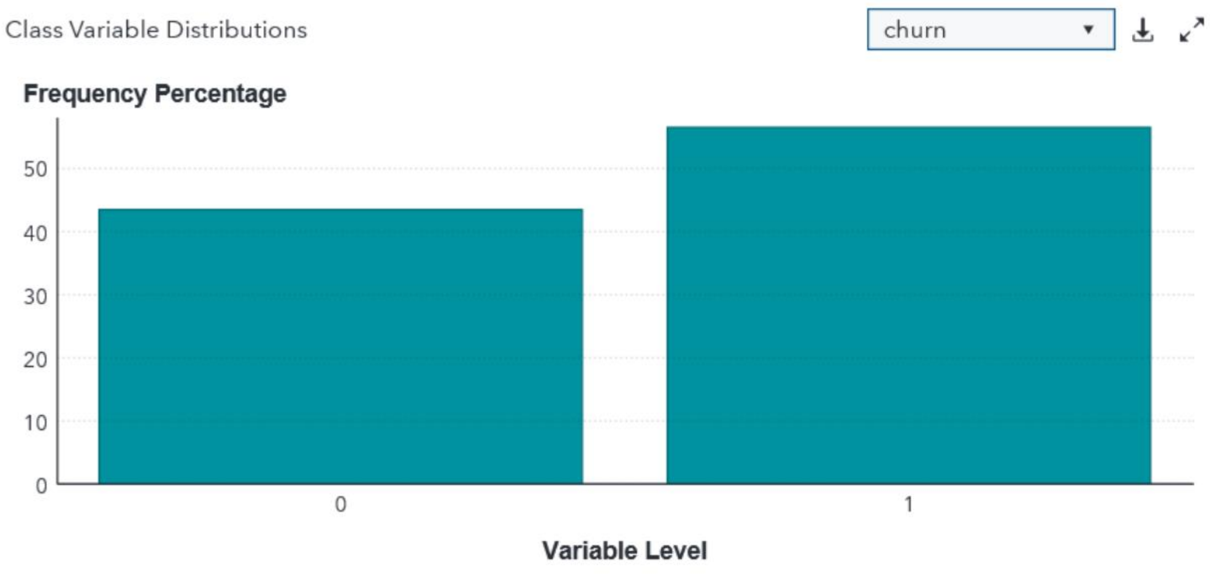
variable name	Role	Number of levels (include missing)	Missing	Missing percentage	Mode	Mode percentage	Mode2	Mode percentage
children	input	2	0	0	0	75.91%	1	24.09%
credit	input	7	0	0	aa	38.16%	a	17.11%
creditCard	input	2	0	0	0	33.40%	1	66.60%
marry	input	3	0	0	2	39.23%	1	35.94%
occupation	input	8	5931	73.98%	professional	16.98%	crafts	2.82%
regionType	input	4	3914	48.82%	suburban	31.30%	town	14.99%
churn	output	2	0	0	0	50.26%	1	49.74%

Appendix 3 Cleaned Data Analysis

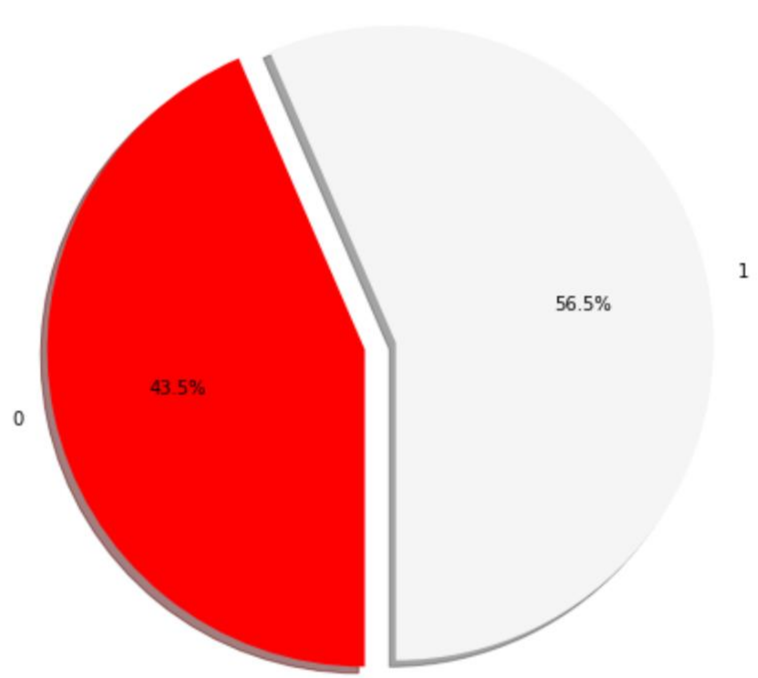
	count	mean	std	min	25%	50%	75%	max
children	4301	0.246454	0.430996	0	0	0	0	1
creditCard	4301	0.681934	0.465779	0	0	1	1	1
custcare	4301	1.352476	6.758898	0	0	0	1	376
custcareTotal	4301	5.733783	38.29992	0	0	0	3	2253
custcareLast	4301	1.340851	6.849823	0	0	0	1	387
directas	4301	0.693792	1.706467	0	0	0	1	28
directasLast	4301	0.658684	1.630448	0	0	0	1	25
dropvce	4301	4.496164	7.482743	0	0	2	6	109
dropvceLast	4301	4.26924	7.161269	0	0	2	5	109
income	4301	4.352011	3.141806	0	1	5	7	9
mou	4301	434.3412	482.6216	0	106.9492	294.2847	591.7508	5687.757
mouTotal	4301	1764.249	2368.36	0	304.4848	948.1822	2278.978	27430.24
mouChange	4301	1.146468	21.72042	-1	-0.0726	-0.00013	0.066003	967.4179
outcalls	4301	123.9893	158.4448	0	14	74	172	2235
overage	4301	33.58807	87.16411	0	0	0	29.73886	1512.281
overageMax	4301	35.32554	91.10486	0	0	0	31.86711	1598.581
overageMin	4301	31.78804	83.24252	0	0	0	28.02815	1472.172
peakOffPeak	4301	1.95587	3.140487	0	0.571429	1.269231	2.333333	79.33333
peakOffPeakLast	4301	1.928712	3.075827	0	0.565891	1.26087	2.310345	79.33333
recchrge	4301	43.69034	22.65921	0	29.99	41.98	54.98	309.99
revenue	4301	56.34652	38.58401	0	33.808	46.71	69.67	577.2117
revenueTotal	4301	226.8833	206.5928	0	87.41	183.66	299.94	3463.27
revenueChange	4301	0.080177	0.768395	-1	-0.01984	0	0.01899	14.463
roam	4301	1.046733	5.489559	0	0	0	0	149

Appendix 3 Cleaned Data Analysis

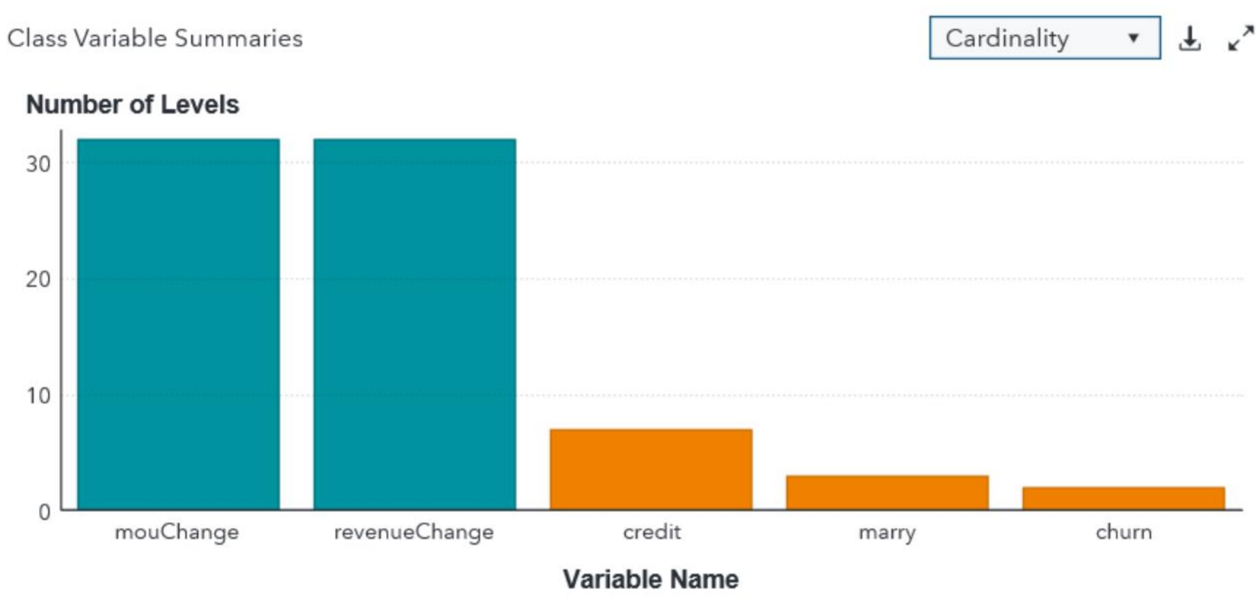
Class Variable Distributions



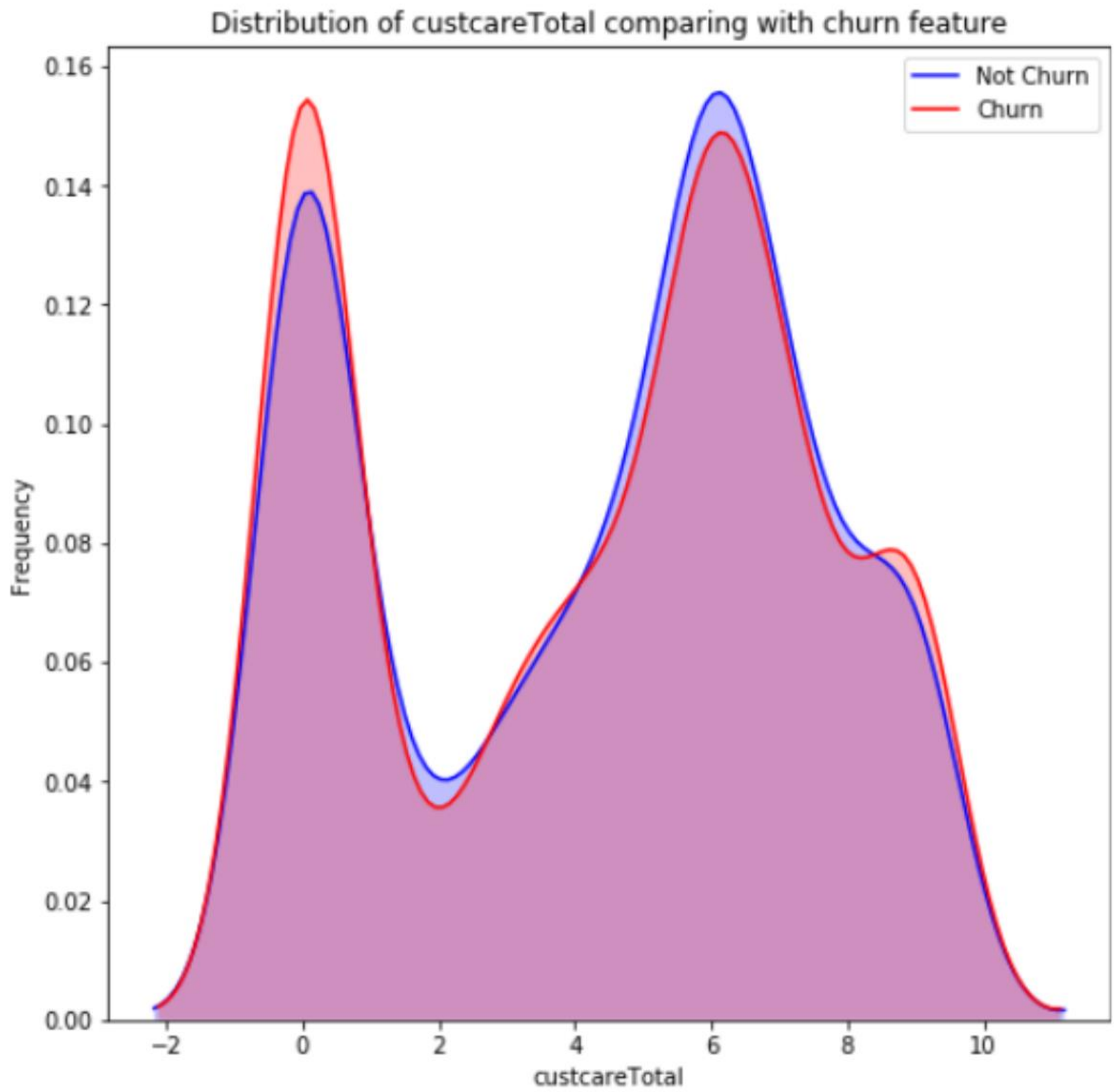
Percent of churn in customer

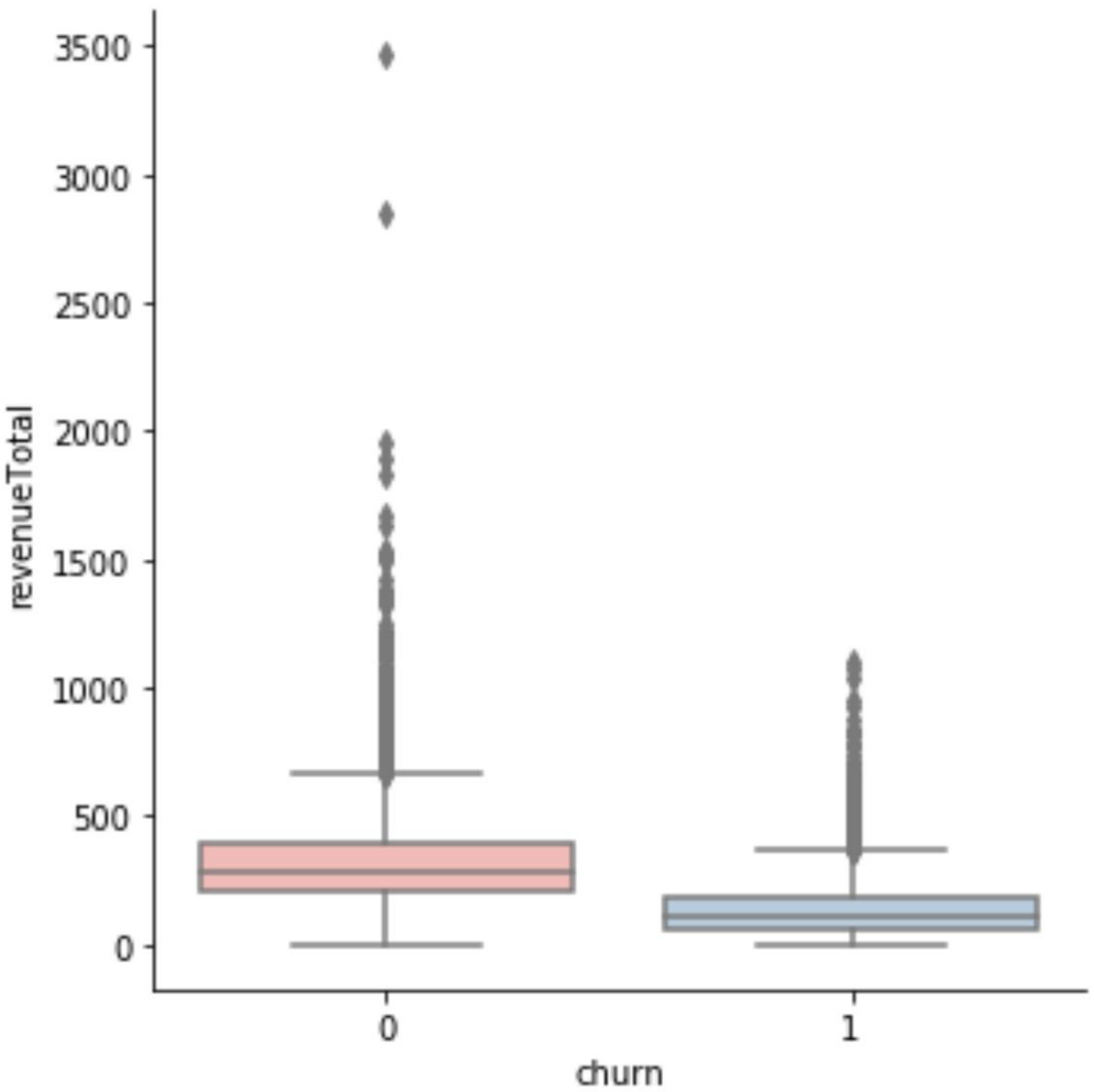


Class Variable Summaries

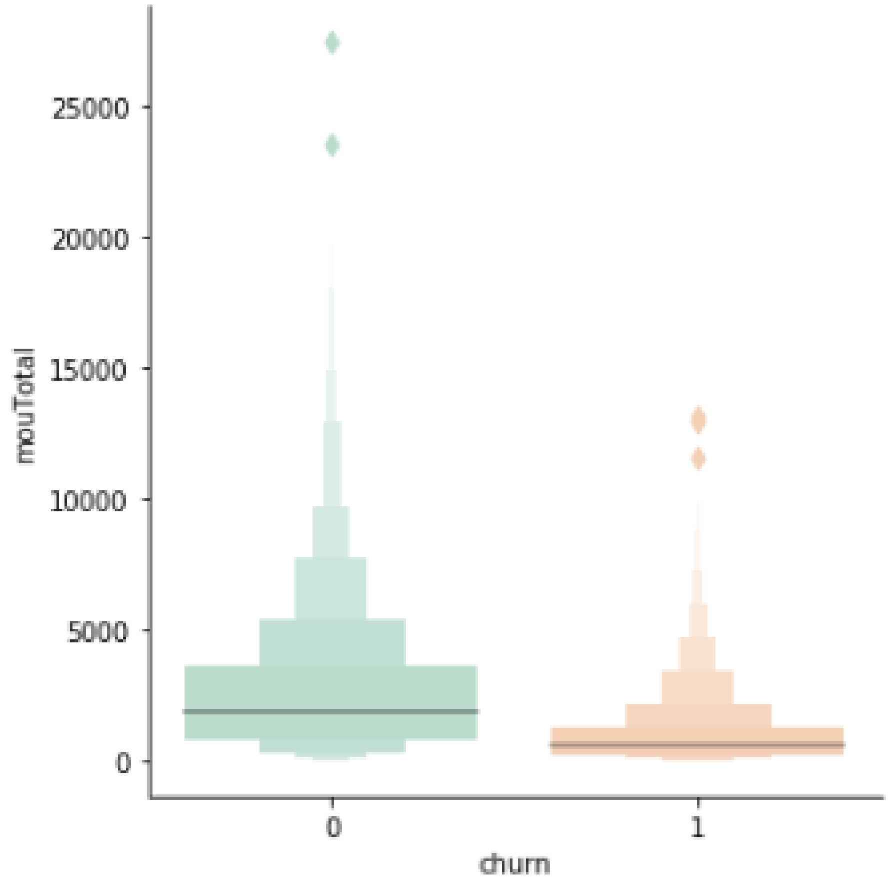


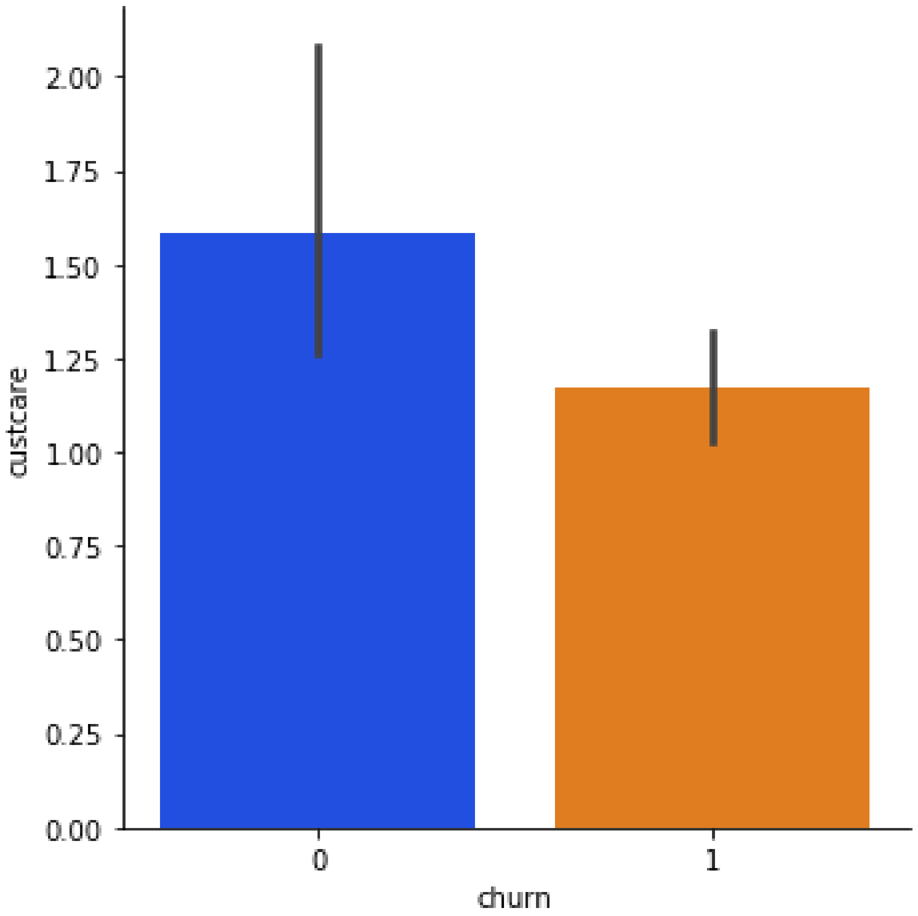
Appendix 3 Cleaned Data Analysis



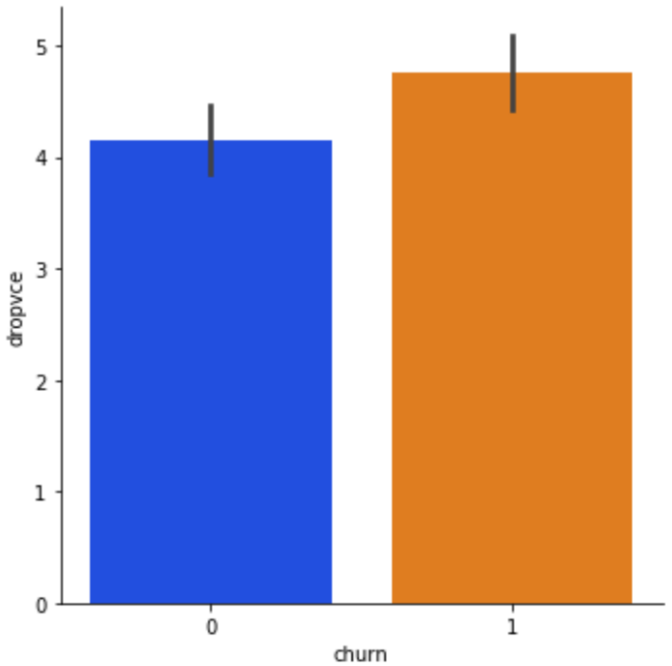


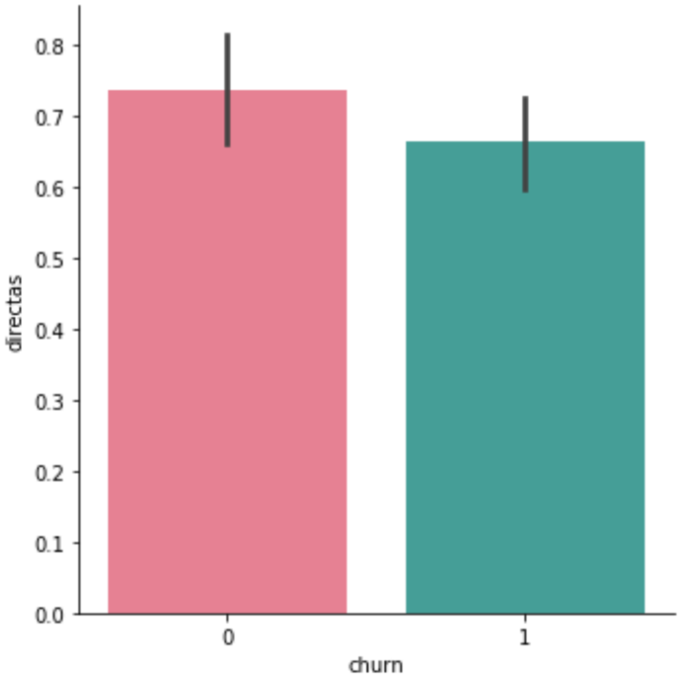
Appendix 3 Cleaned Data Analysis

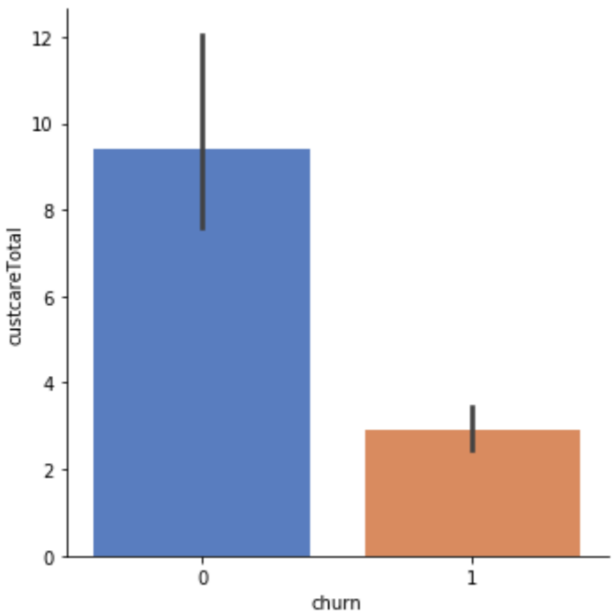


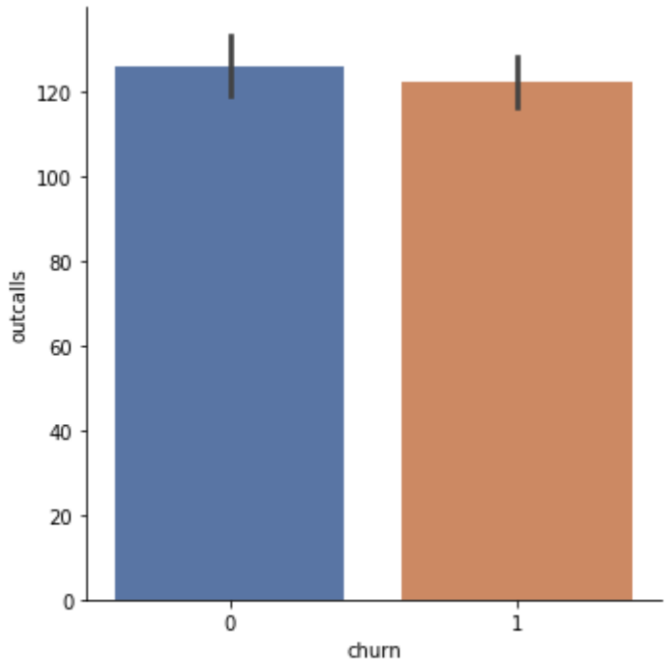


Appendix 3 Cleaned Data Analysis

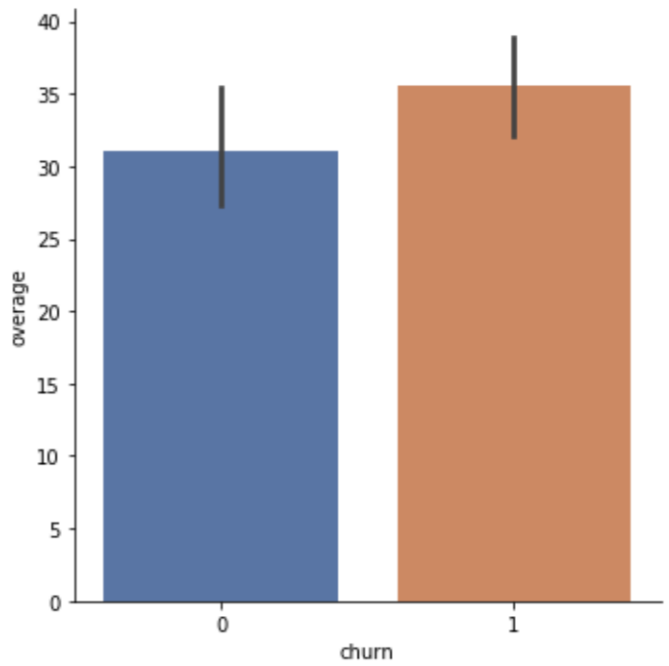


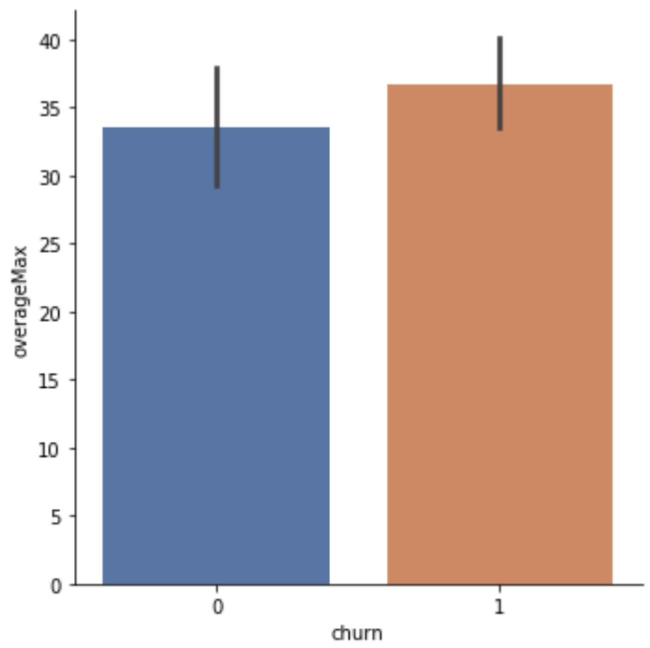


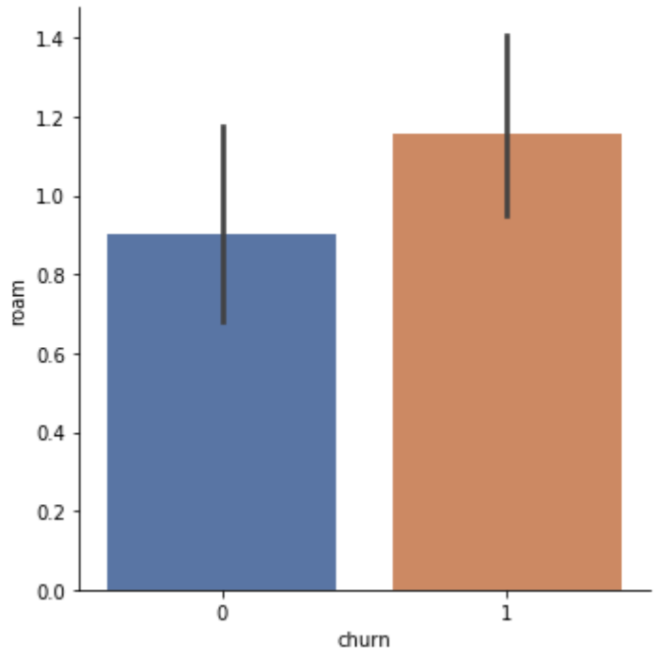




Appendix 3 Cleaned Data Analysis

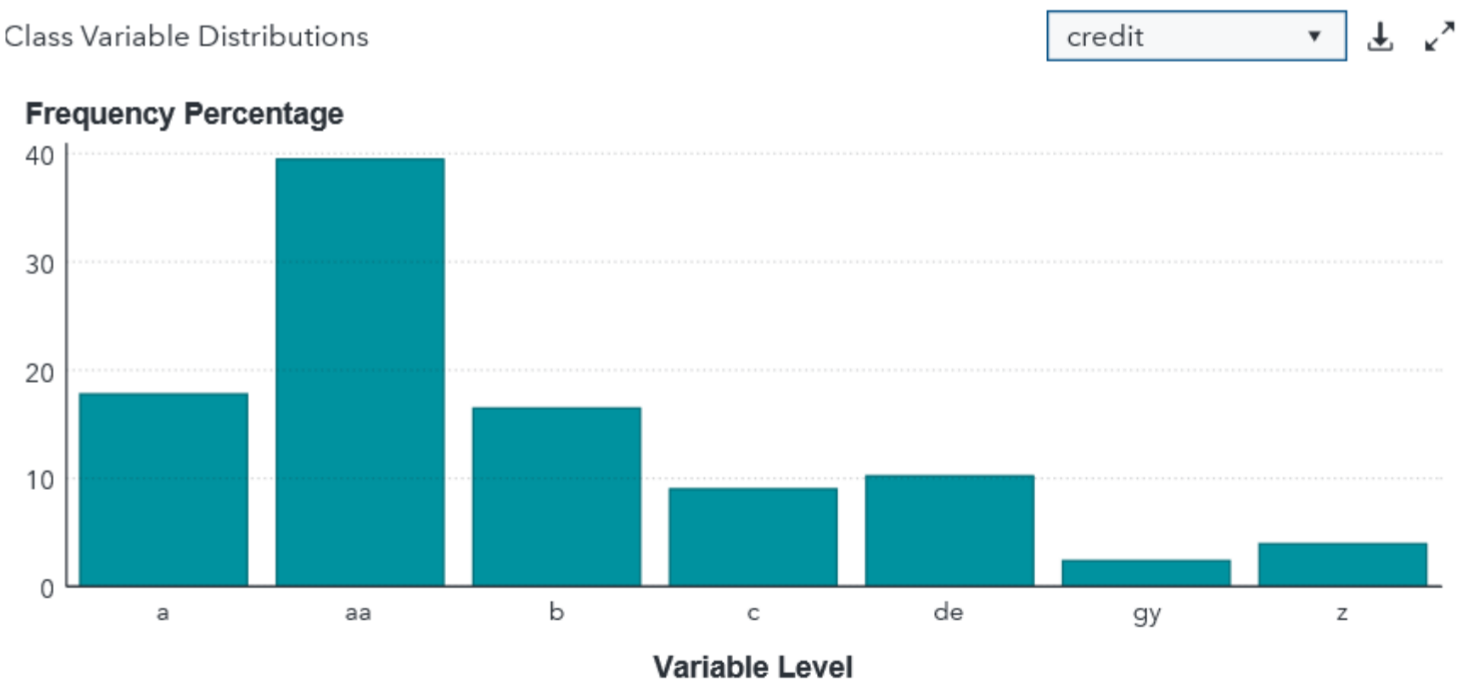




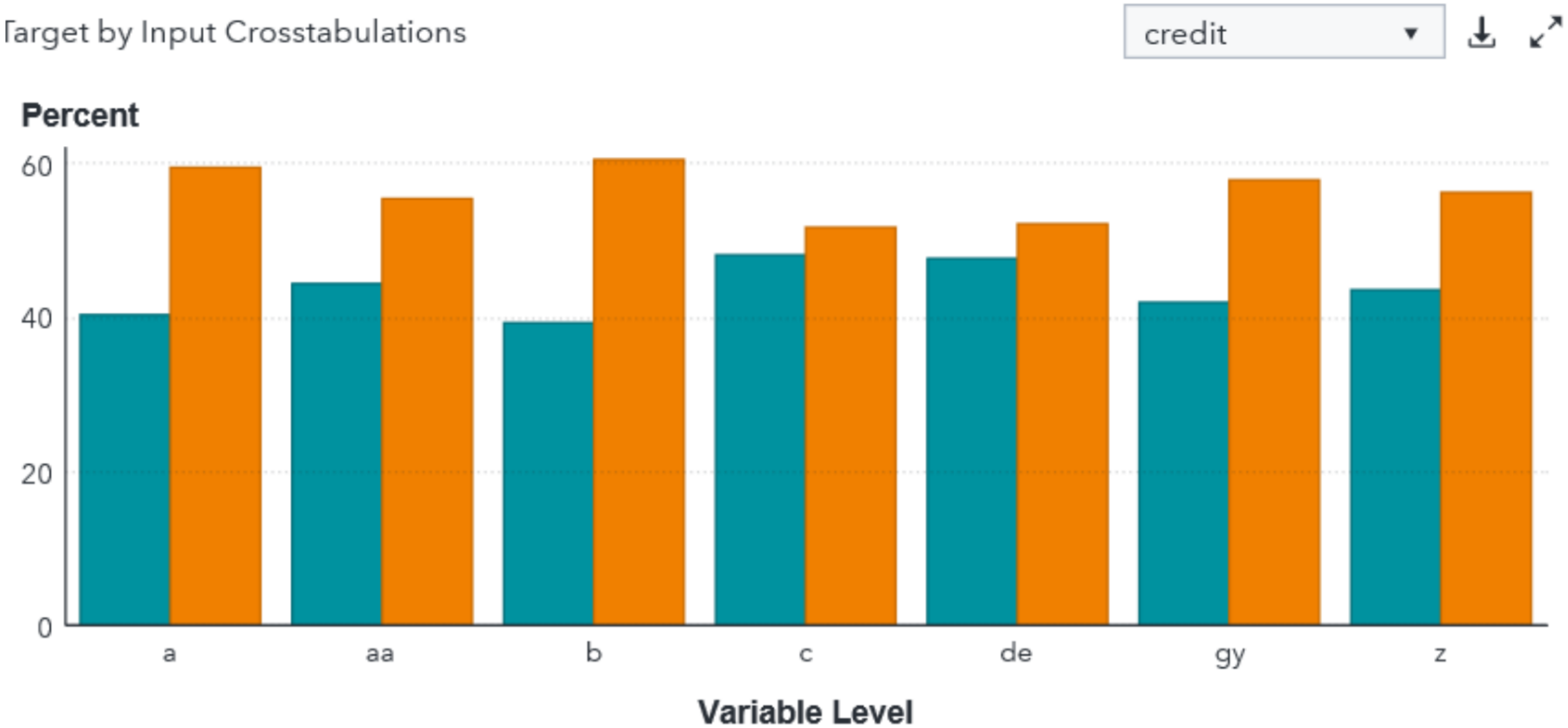


Appendix 3 Cleaned Data Analysis

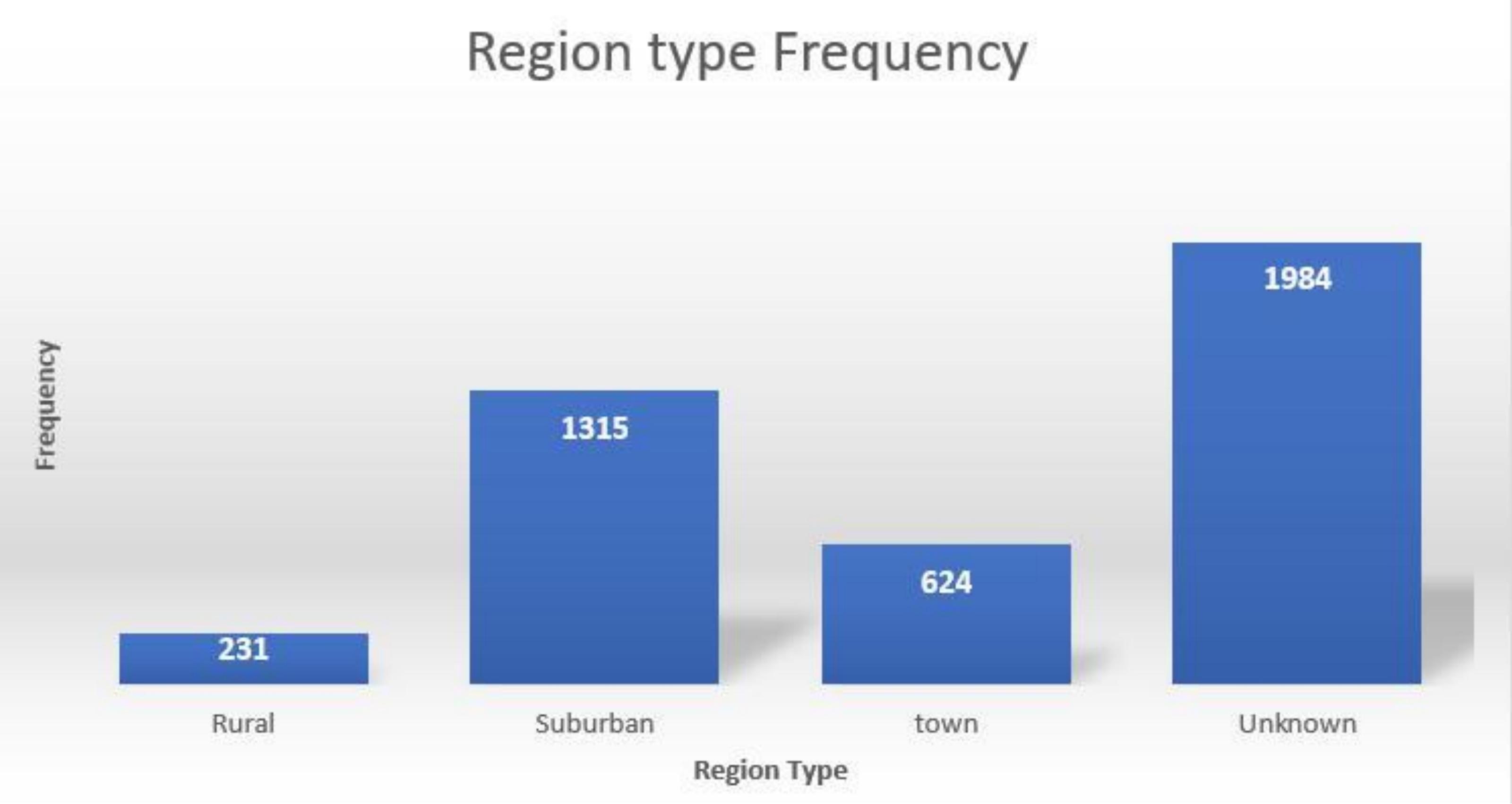
Class Variable Distributions



Target by Input Crosstabulations

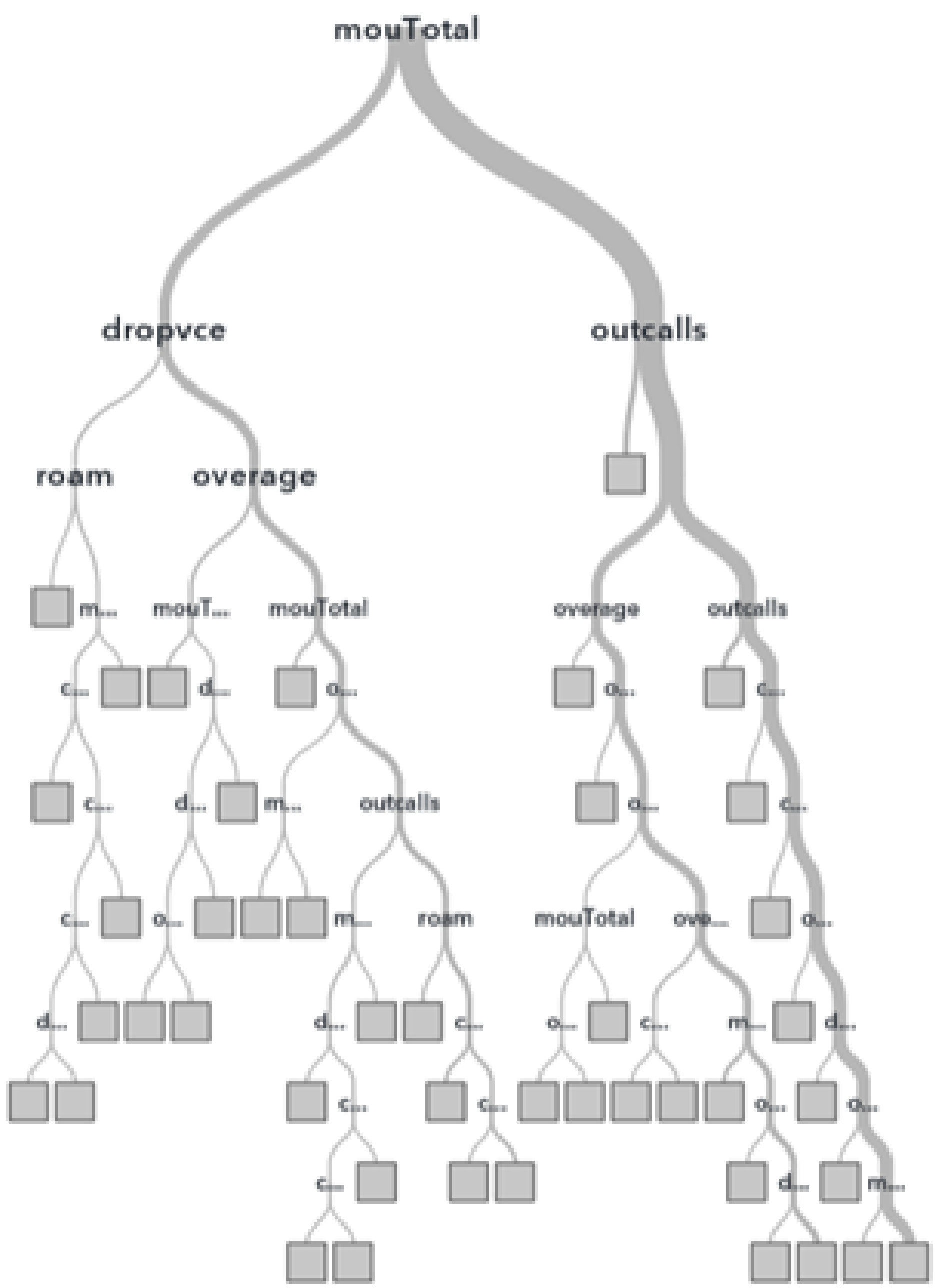


Appendix 3 Cleaned Data Analysis

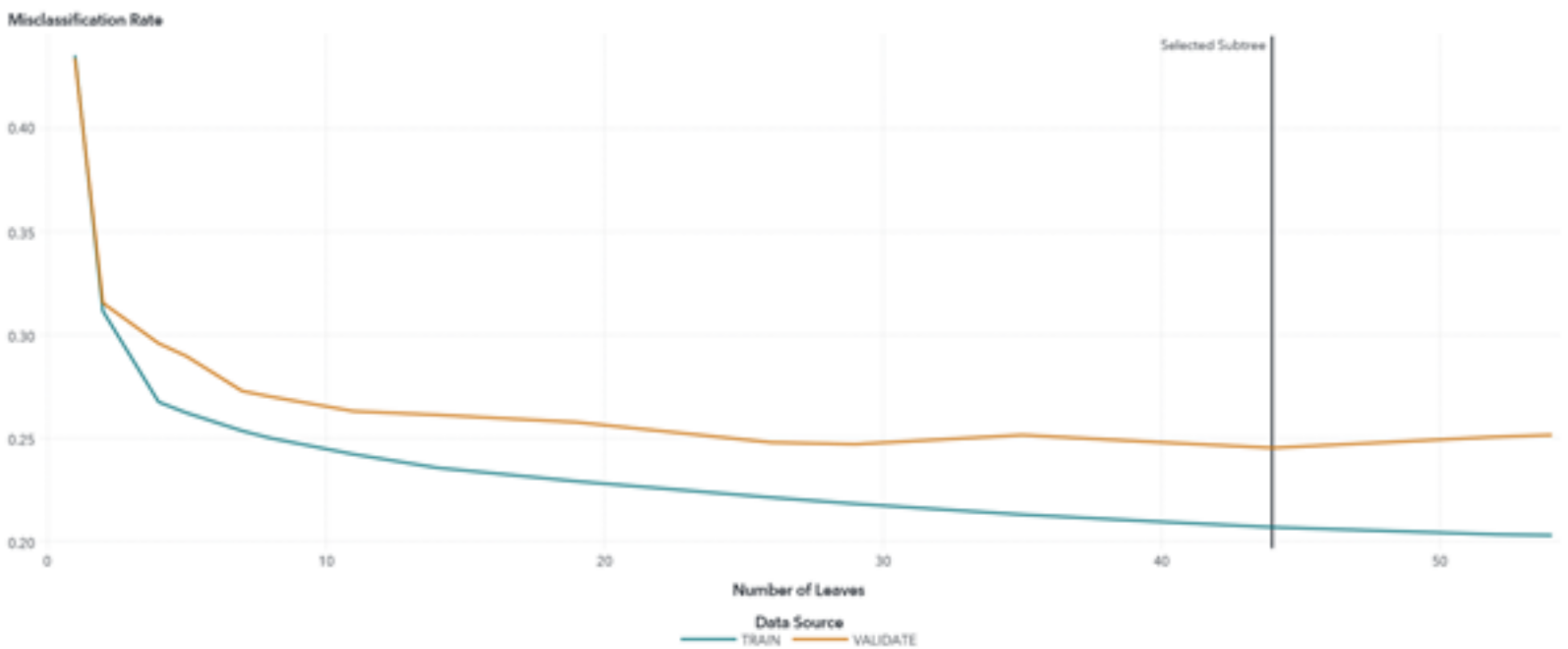


Appendix 4 Automatic Default Decision Tree Visualisation

Tree Diagram

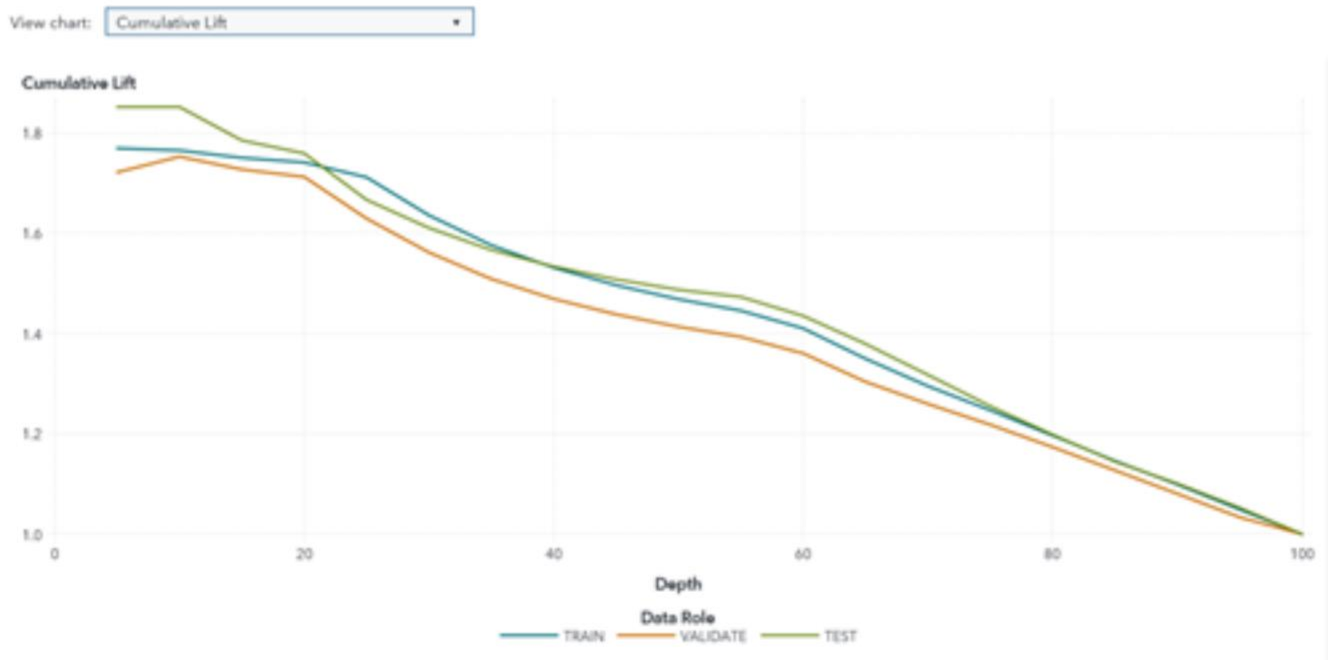


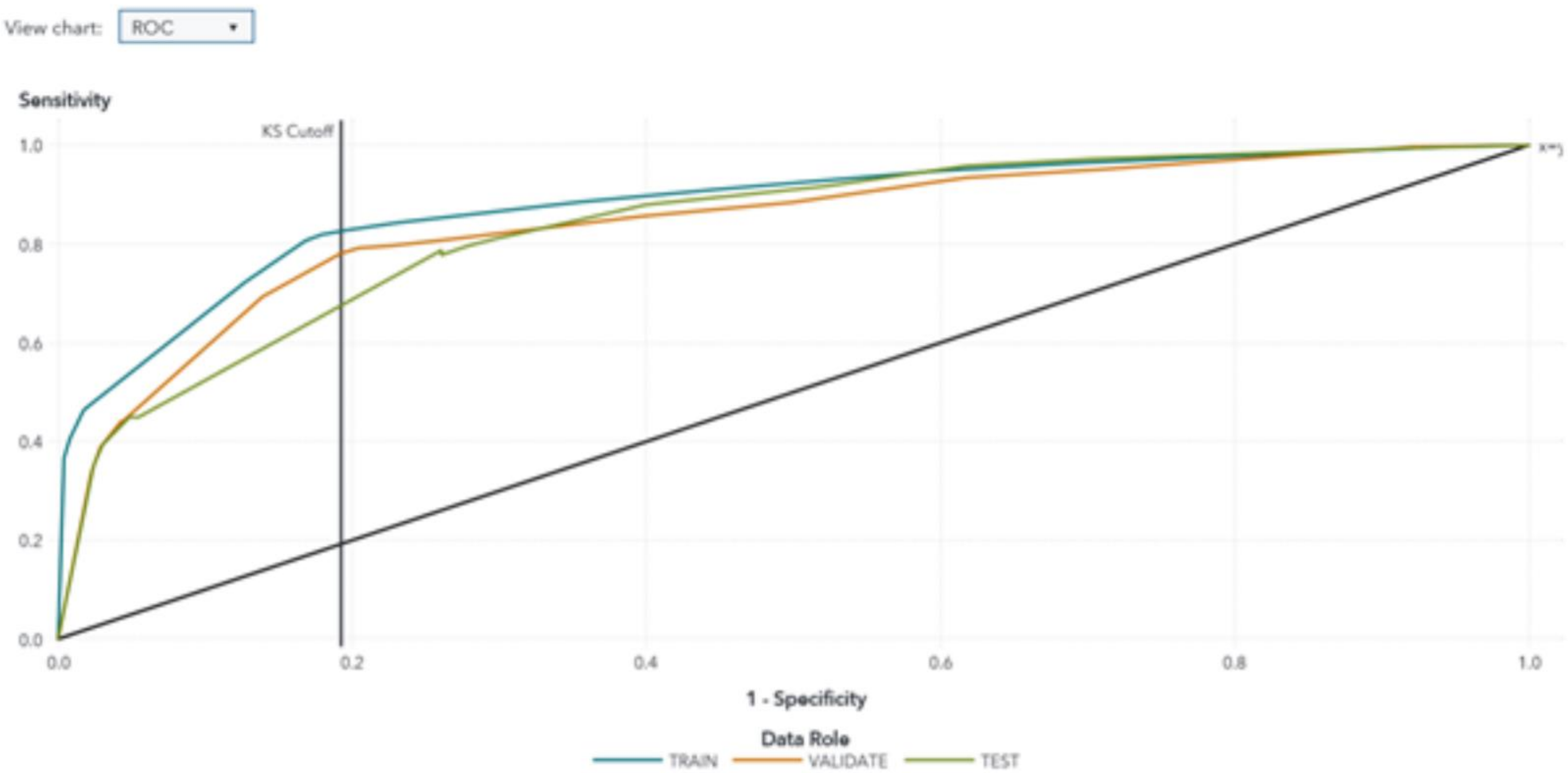
Appendix 4 Automatic Default Decision Tree Misclassification Rate



Fit Statistics for Selected Tree		
	Number of Leaves	Misclassification Rate
Training	44	0.2072
Validation	44	0.2456
Test	44	0.2126

Appendix 4 Automatic Default Decision Tree Cumulative lift and ROC curve





Appendix 4 Automatic Default Decision Tree Confusion Matrix

Event Classification

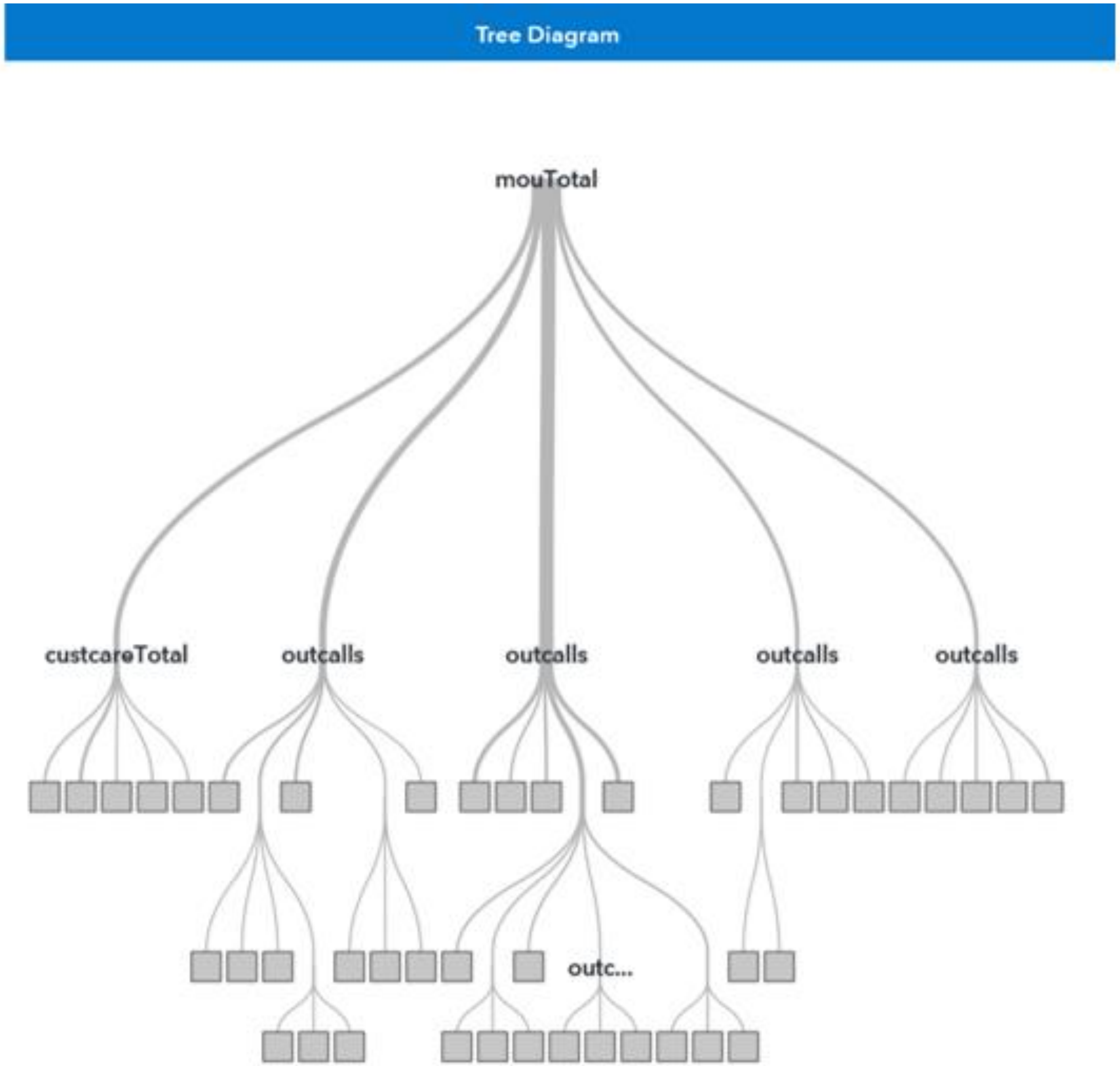
chart: Table

Cutoff	Cutoff S... ↑	Target Name	Response	Event	Value	Training Fr...	Validation Fr...	Test Frequency	Training P...	Validati...	Test Percentage
0.5000	Default	churn	Incorrect	1	False Positive	211	123	33	16.2558	19.2488	15.2778
0.5000	Default	churn	Incorrect	0	False Negative	265	154	48	26.5265	31.4928	29.0909
0.5000	Default	churn	Correct	0	True Negative	734	335	117	73.4735	68.5072	70.9091
0.5000	Default	churn	Correct	1	True Positive	1,087	516	183	83.7442	80.7512	84.7222

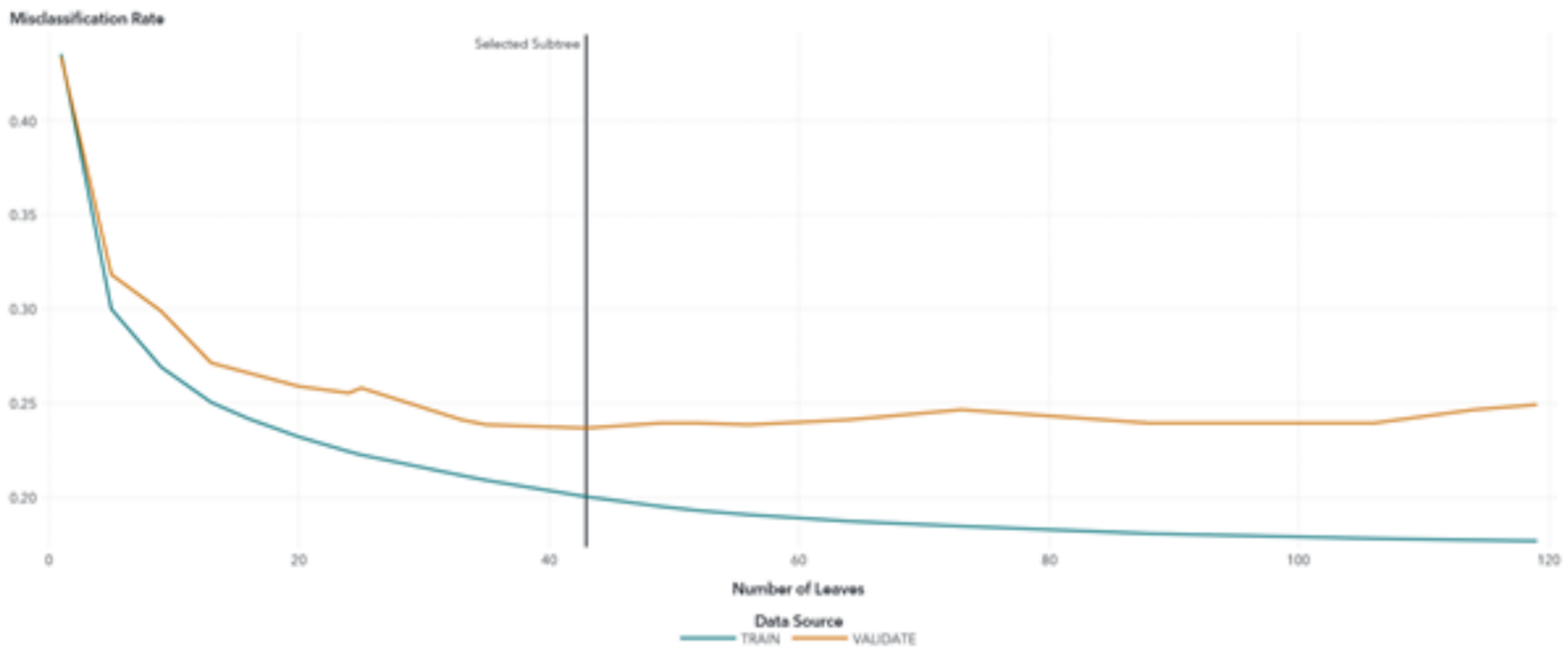
Fit Statistics

Target ...	Data Role	Partitio...	Formatt...	Sum of ...	Averag...	Divisor f...	Root Av...	Misclas...	Multi-Cl...	KS (You...	Area Un...	Gini Co...	Gamma
churn	TEST	2	2	381	0.1590	381	0.3987	0.2126	0.4865	0.5563	0.8327	0.6654	0.7494
churn	TRAIN	1	1	2,297	0.1467	2,297	0.3830	0.2072	0.4429	0.5722	0.8601	0.7202	0.8019
churn	VALIDATE	0	0	1,128	0.1744	1,128	0.4176	0.2456	0.6815	0.4926	0.8119	0.6238	0.7018

Appendix 4 Automatic Default Decision Tree Decision Tree (GINI) Visualisation

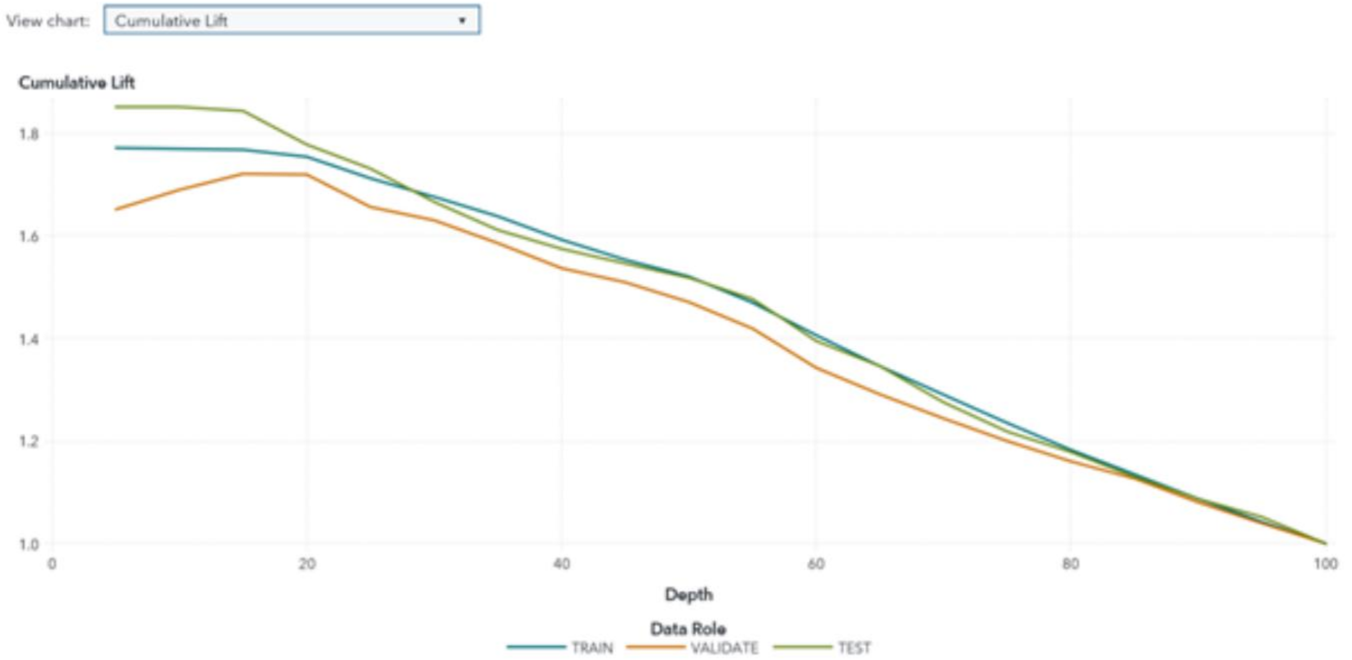


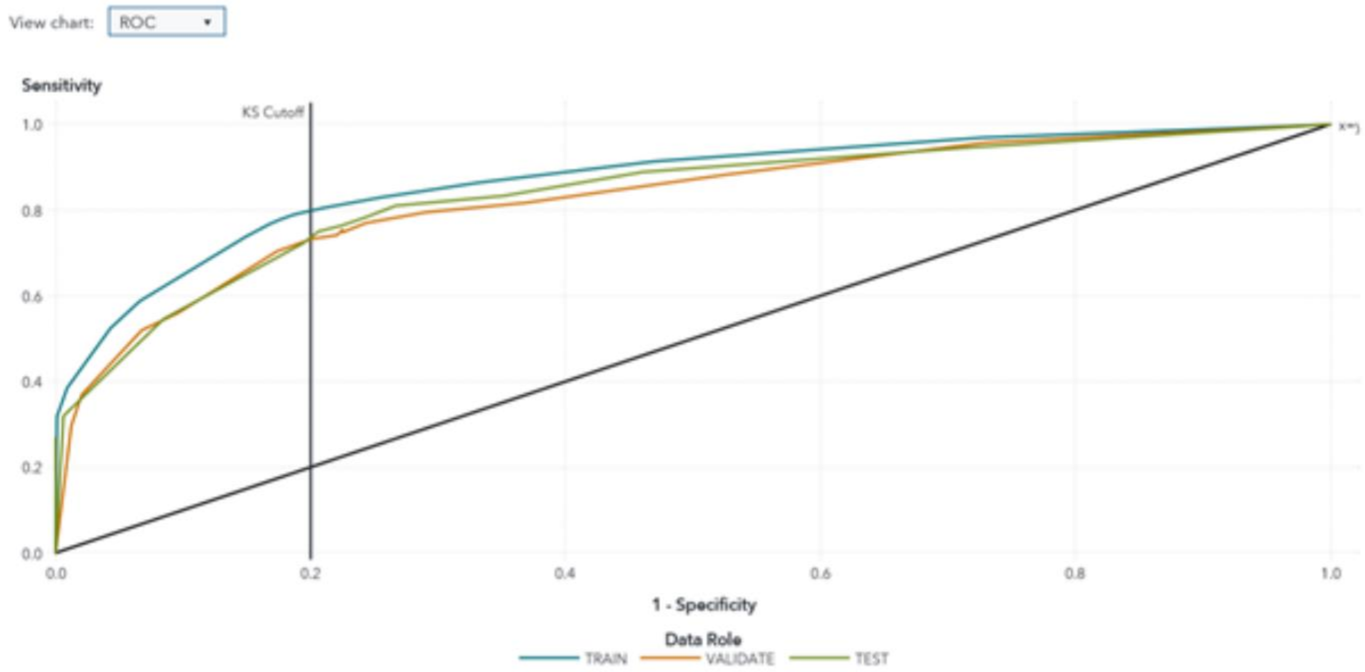
Appendix 4 Automatic Default Decision Tree Decision Tree (GINI) Misclassification Rate



Fit Statistics for Selected Tree		
	Number of Leaves	Misclassification Rate
Training	43	0.2003
Validation	43	0.2367
Test	43	0.2310

Appendix 4 Automatic Default Decision Tree Decision Tree (GINI) Cumulative lift and ROC Curve





Appendix 4 Automatic Default Decision Tree Decision Tree (GINI) Confusion Matrix and Fit Statistics

Event Classification

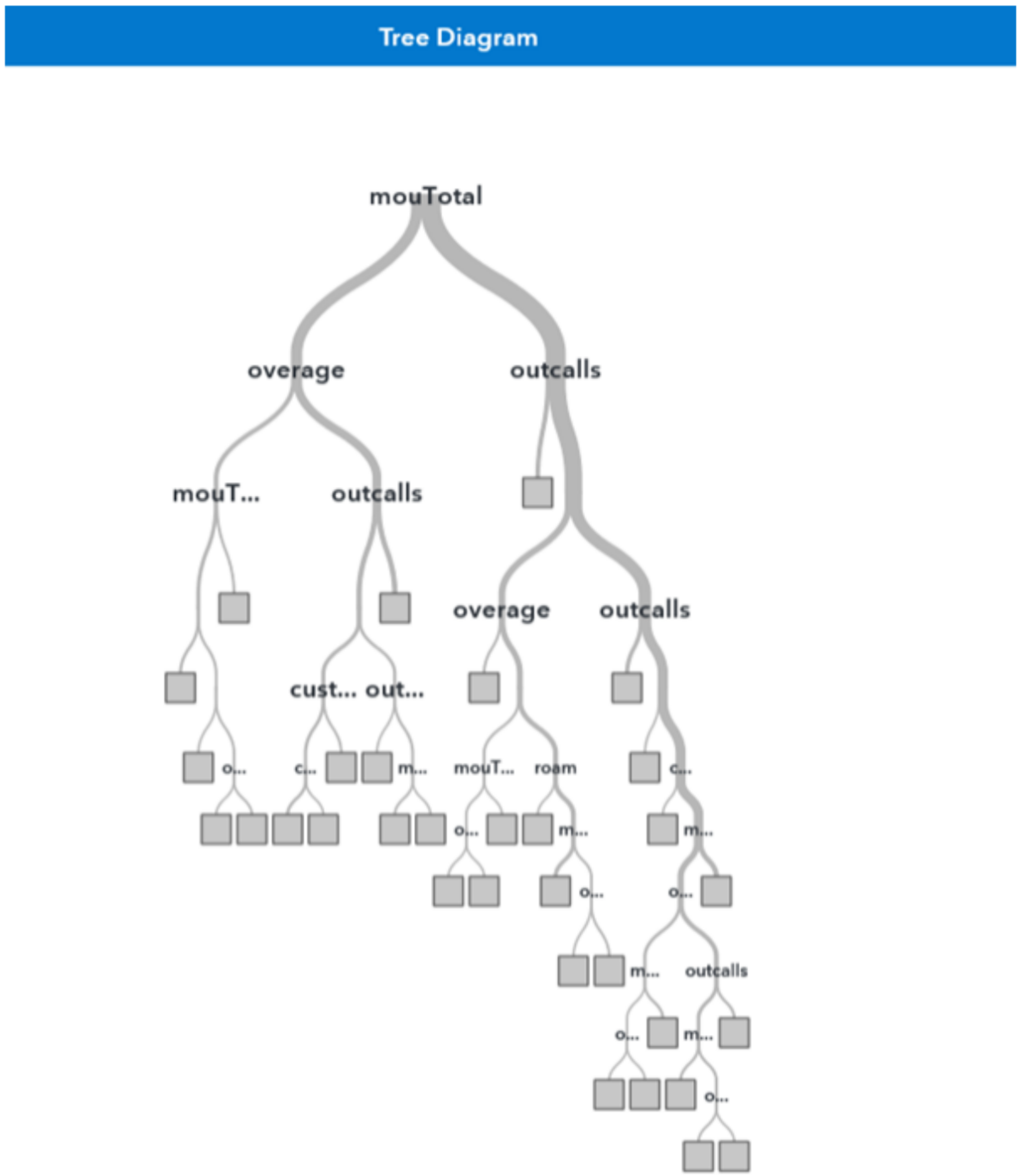
chart: Table

Cutoff	Cutoff Sou...	Target Name	Response	Event	Value	Training Fr...	Validation ...	Test Frequ...	Training P...	Validation ...	Test Perce...
0.5000	Default	churn	Correct	1	True Positive	1,027	482	165	79.1217	75.4304	76.3889
0.5000	Default	churn	Incorrect	1	False Positive	271	157	51	20.8783	24.5696	23.6111
0.5000	Default	churn	Correct	0	True Negative	810	379	128	81.0811	77.5051	77.5758
0.5000	Default	churn	Incorrect	0	False Negative	189	110	37	18.9189	22.4949	22.4242

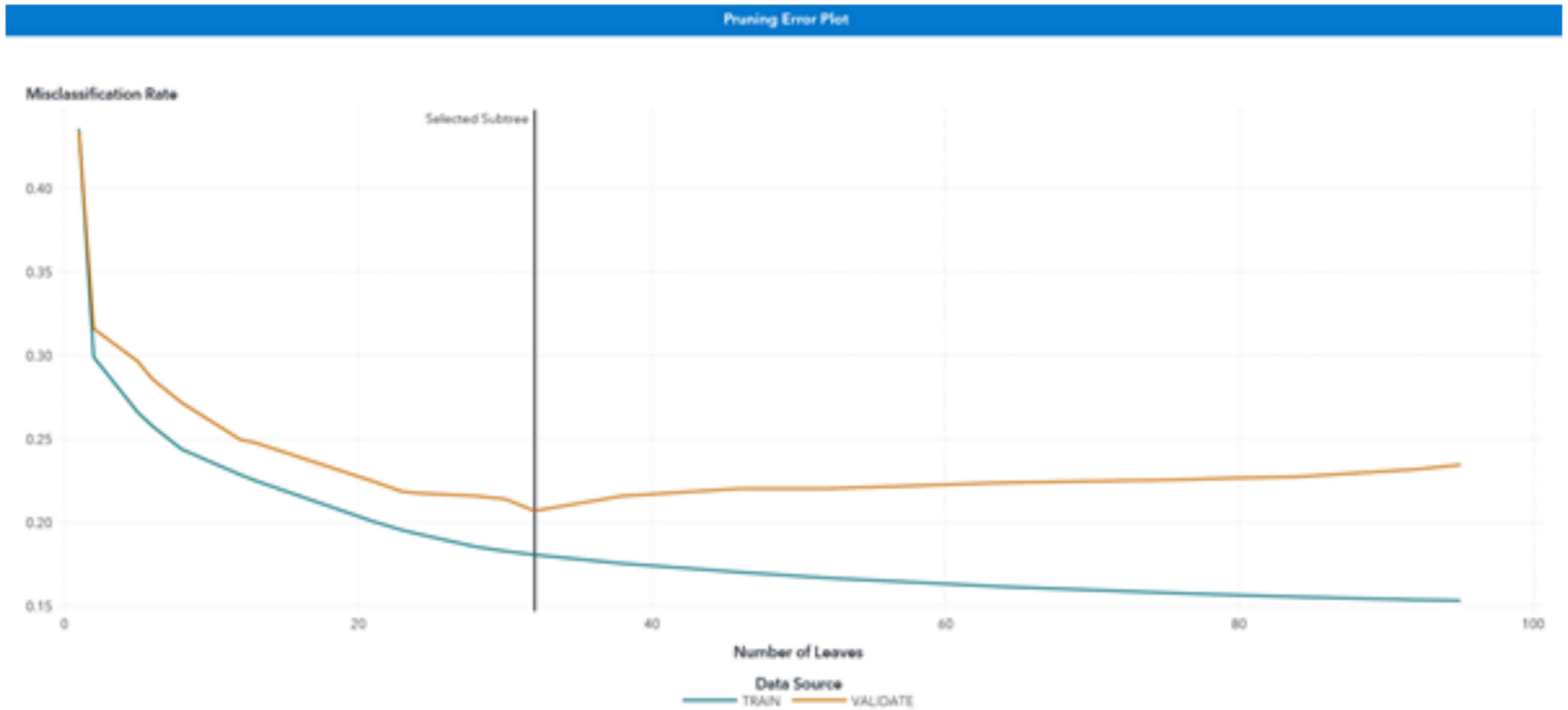
Fit Statistics

Target ...	Data Role	Partitio...	Formatt...	Sum of ...	Averag...	Divisor f...	Root Av...	Misclas...	Multi-Cl...	KS (You...	Area Un...	Gini Co...	Gamma
churn	TEST	2	2	381	0.1624	381	0.4030	0.2310	0.4846	0.5439	0.8354	0.6708	0.7207
churn	TRAIN	1	1	2,297	0.1435	2,297	0.3788	0.2003	0.4356	0.6027	0.8692	0.7383	0.7803
churn	VALIDATE	0	0	1,128	0.1676	1,128	0.4094	0.2367	0.6017	0.5320	0.8269	0.6539	0.6993

Appendix 4 Automatic Default Decision Tree Decision Tree (Entropy) Visualisation

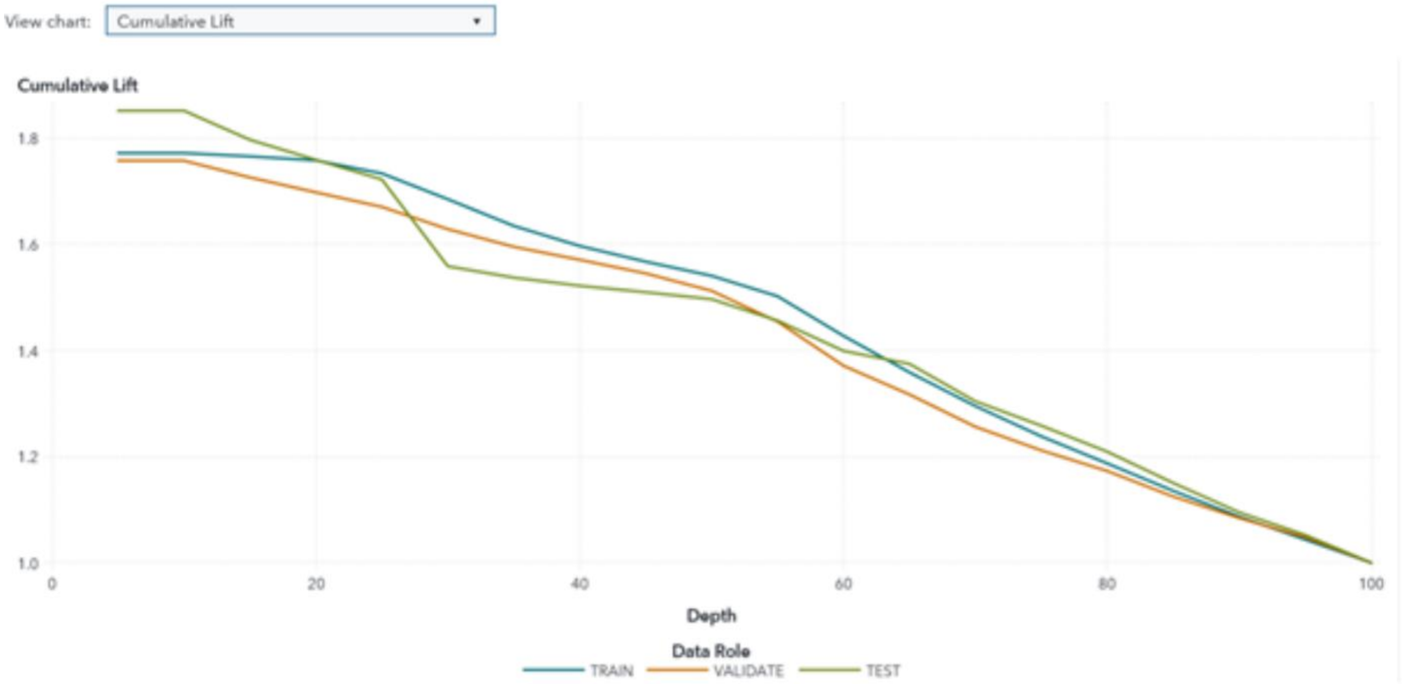


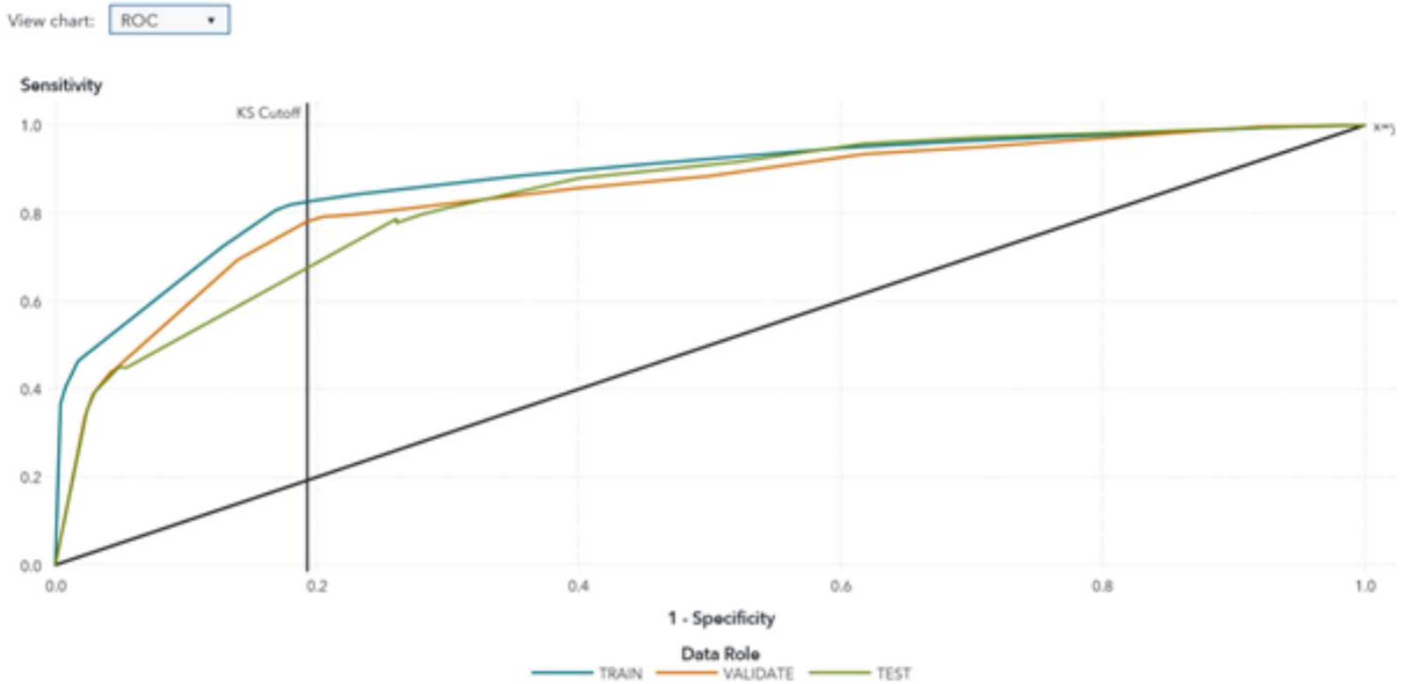
Appendix 4 Automatic Default Decision Tree Decision Tree (Entropy) Misclassification Rate



Fit Statistics for Selected Tree		
	Number of Leaves	Misclassification Rate
Training	32	0.1802
Validation	32	0.2066
Test	32	0.2336

Appendix 4 Automatic Default Decision Tree Decision Tree (Entropy) Cumulative lift and ROC Curve





Appendix 4 Gradient Boosted Decision Tree Decision Tree (Entropy) Confusion Matrix and Fit Statistics

Event Classification

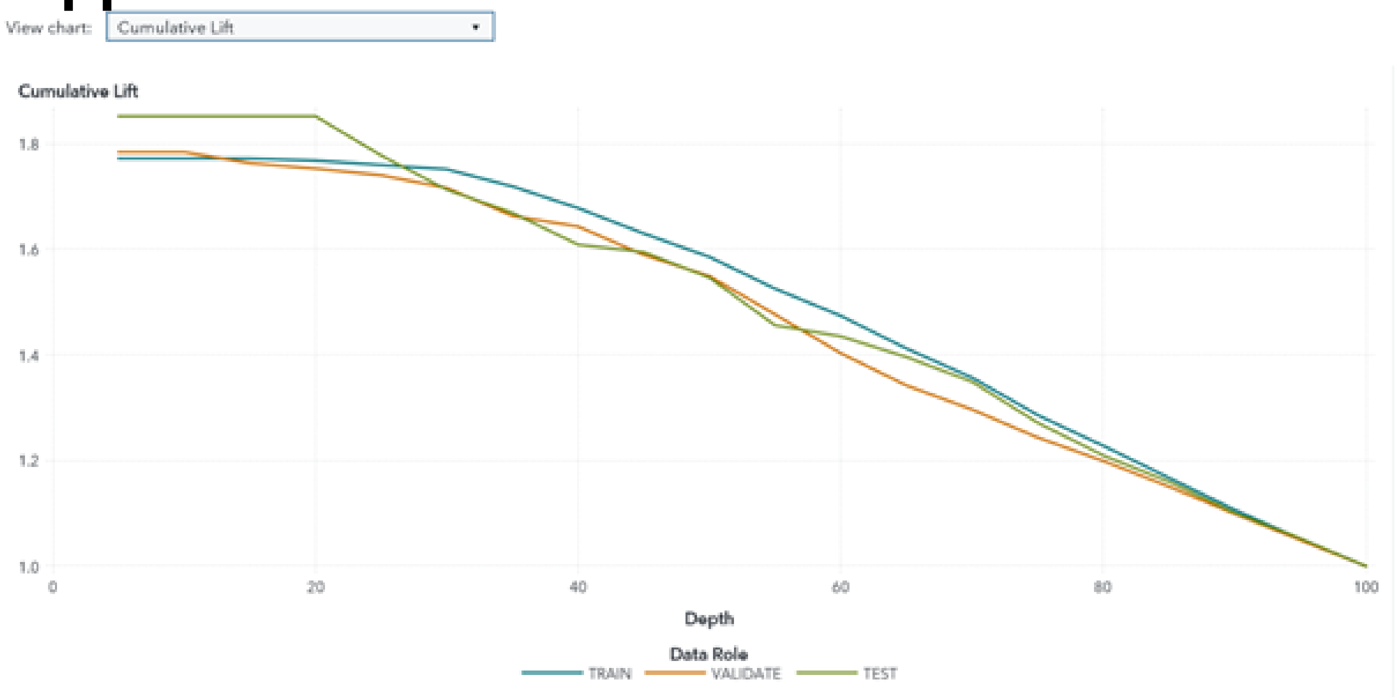
ew chart: Table

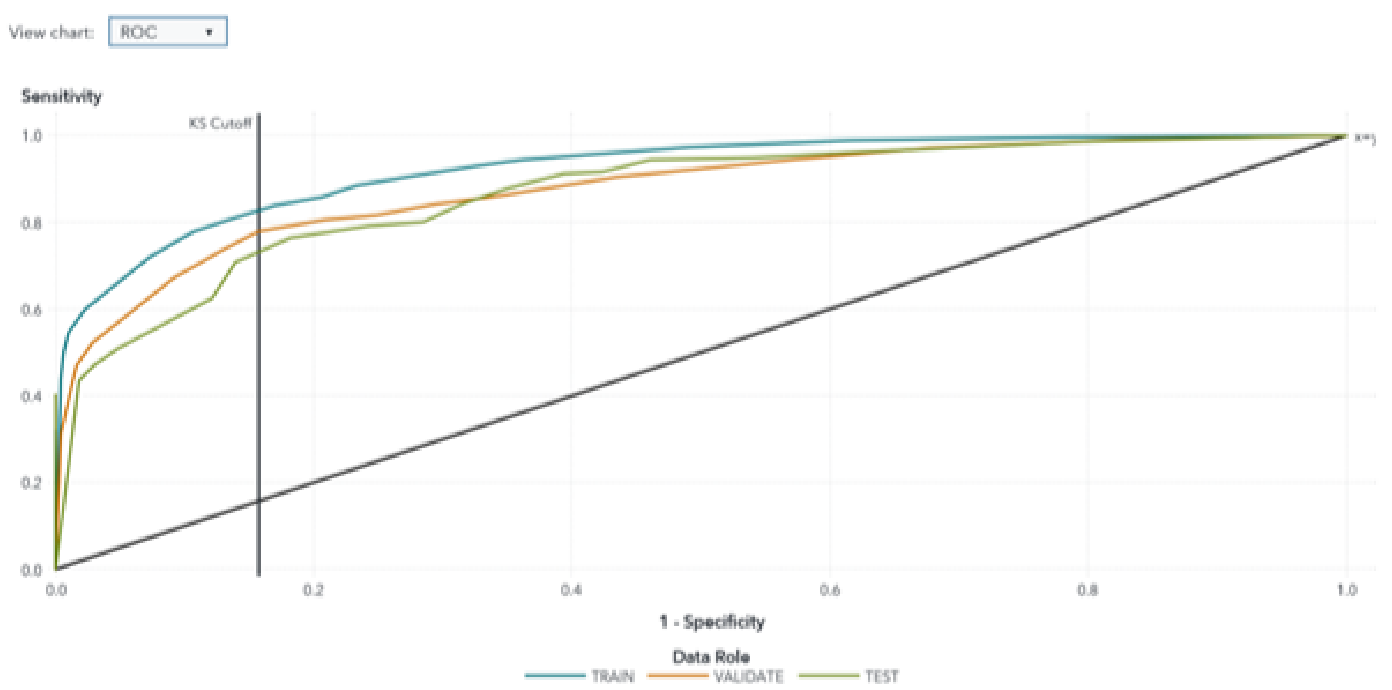
Cutoff	Cutoff S... ↑	Target Name	Response	Event	Value	Training Fr...	Validation ...	Test Frequ...	Training P...	Validation ...	Test Perce...
0.5000	Default	churn	Correct	1	True Positive	1,063	506	170	81.8952	79.1862	78.7037
0.5000	Default	churn	Incorrect	1	False Positive	235	133	46	18.1048	20.8138	21.2963
0.5000	Default	churn	Correct	0	True Negative	820	389	122	82.0821	79.5501	73.9394
0.5000	Default	churn	Incorrect	0	False Negative	179	100	43	17.9179	20.4499	26.0606

Fit Statistics

Target ...	Data Role	Partitio...	Formatt...	Sum of ...	Averag...	Divisor ...	Root Av...	Misclas...	Multi-Cl...	KS (Yo...	Area U...	Gini Co...	Gamma
churn	TEST	2	2	381	0.1670	381	0.4086	0.2336	0.5037	0.5264	0.8349	0.6698	0.7274
churn	TRAIN	1	1	2,297	0.1352	2,297	0.3677	0.1802	0.4165	0.6398	0.8777	0.7554	0.8019
churn	VALIDATE	0	0	1,128	0.1554	1,128	0.3943	0.2066	0.5119	0.5887	0.8417	0.6833	0.7302

Appendix 4 Gradient Boosted Decision Tree Cumulative lift and ROC Curve





Appendix 4 Gradient Boosted Decision Tree Confusion matrix and Fit Statistics

Event Classification

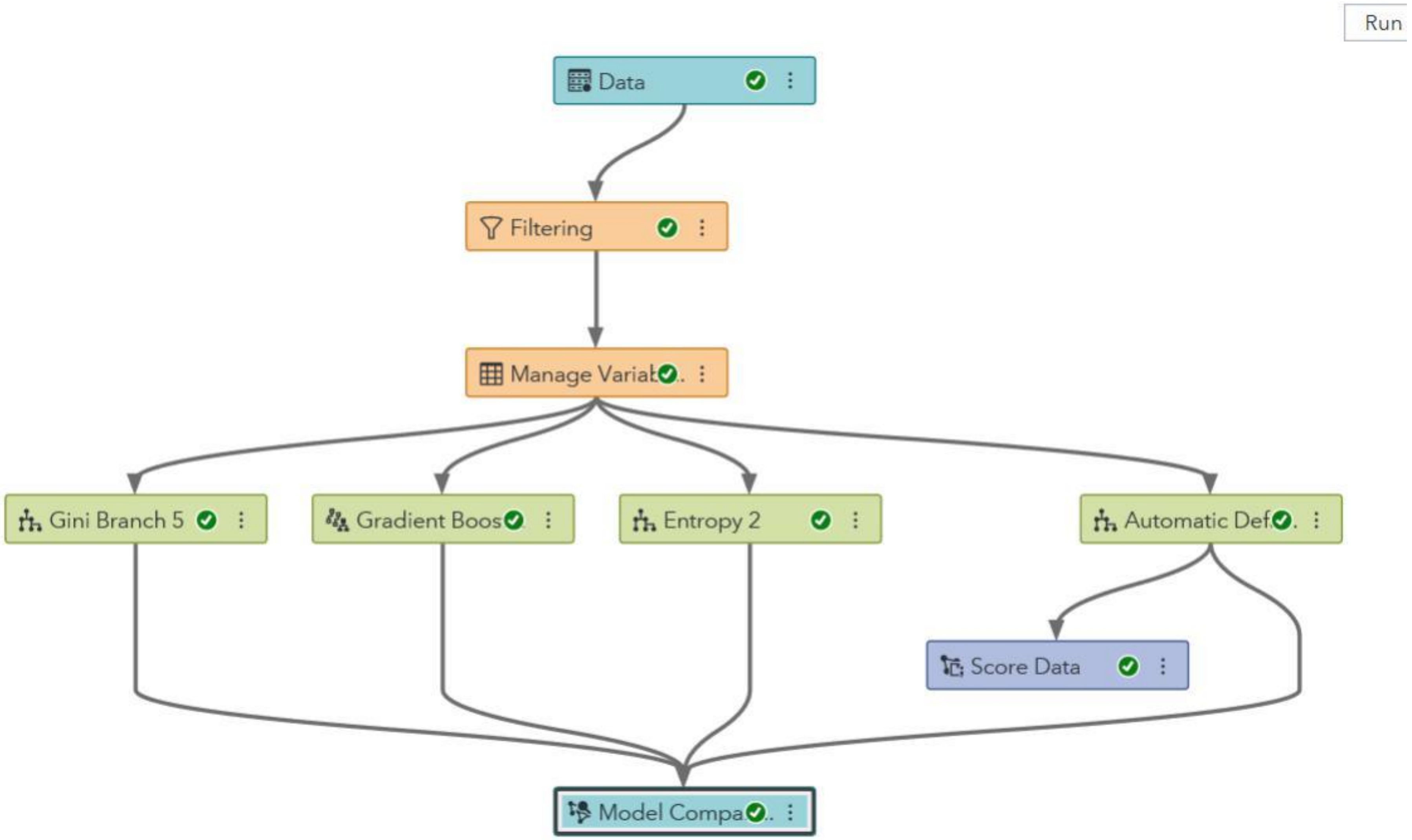
View chart: Table

Cutoff	Cutoff Sou...	Target Name	Response	Event	Value	Training Fr...	Validation ...	Test Frequ...	Training P...	Validation ...	Test Perce...
0.5000	Default	churn	Correct	1	True Positive	1,114	522	173	85.8243	81.6901	80.0926
0.5000	Default	churn	Incorrect	1	False Positive	184	117	43	14.1757	18.3099	19.9074
0.5000	Default	churn	Correct	0	True Negative	793	367	118	79.3794	75.0511	71.5152
0.5000	Default	churn	Incorrect	0	False Negative	206	122	47	20.6206	24.9489	28.4848

Fit Statistics

Target ...	Data Role	Partitio...	Formatt...	Sum of ...	Averag...	Divisor ...	Root Av...	Misclas...	Multi-Cl...	KS (Yo...	Area U...	Gini Co...	Gamma
churn	TEST	2	2	381	0.1453	381	0.3811	0.2362	0.4376	0.5821	0.8691	0.7382	0.7601
churn	TRAIN	1	1	2,297	0.1139	2,297	0.3375	0.1698	0.3584	0.6720	0.9226	0.8453	0.8637
churn	VALIDATE	0	0	1,128	0.1418	1,128	0.3765	0.2119	0.4311	0.6219	0.8773	0.7546	0.7761

Appendix 5 Tree Comparison Pipeline

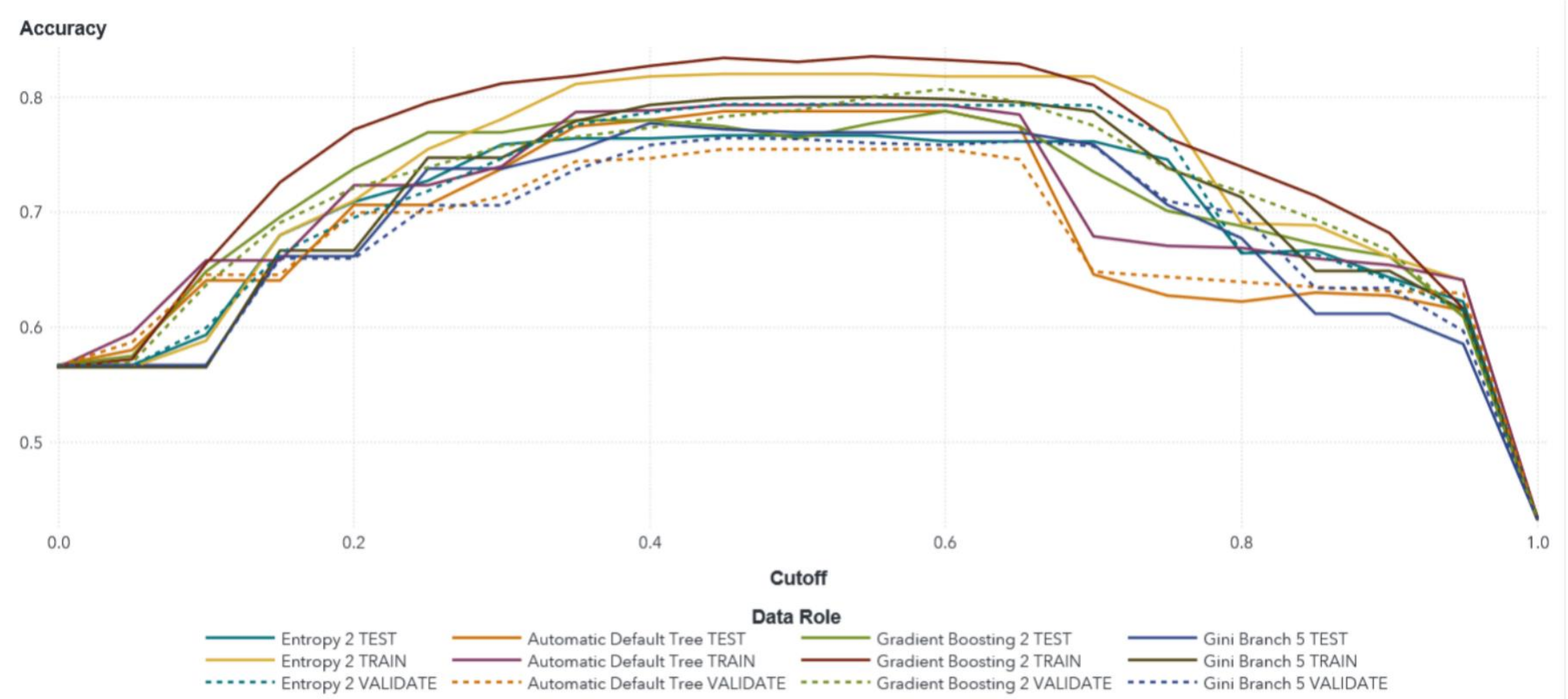


Model Comparison

Champion	Name	Algorithm Name	Accuracy	Misclassification Rate
★	Automatic Default Tree	Decision Tree	0.7874	0.2126
	Gini Branch 5	Decision Tree	0.7690	0.2310
	Entropy 2	Decision Tree	0.7664	0.2336
	Gradient Boosting 2	Gradient Boosting	0.7638	0.2362

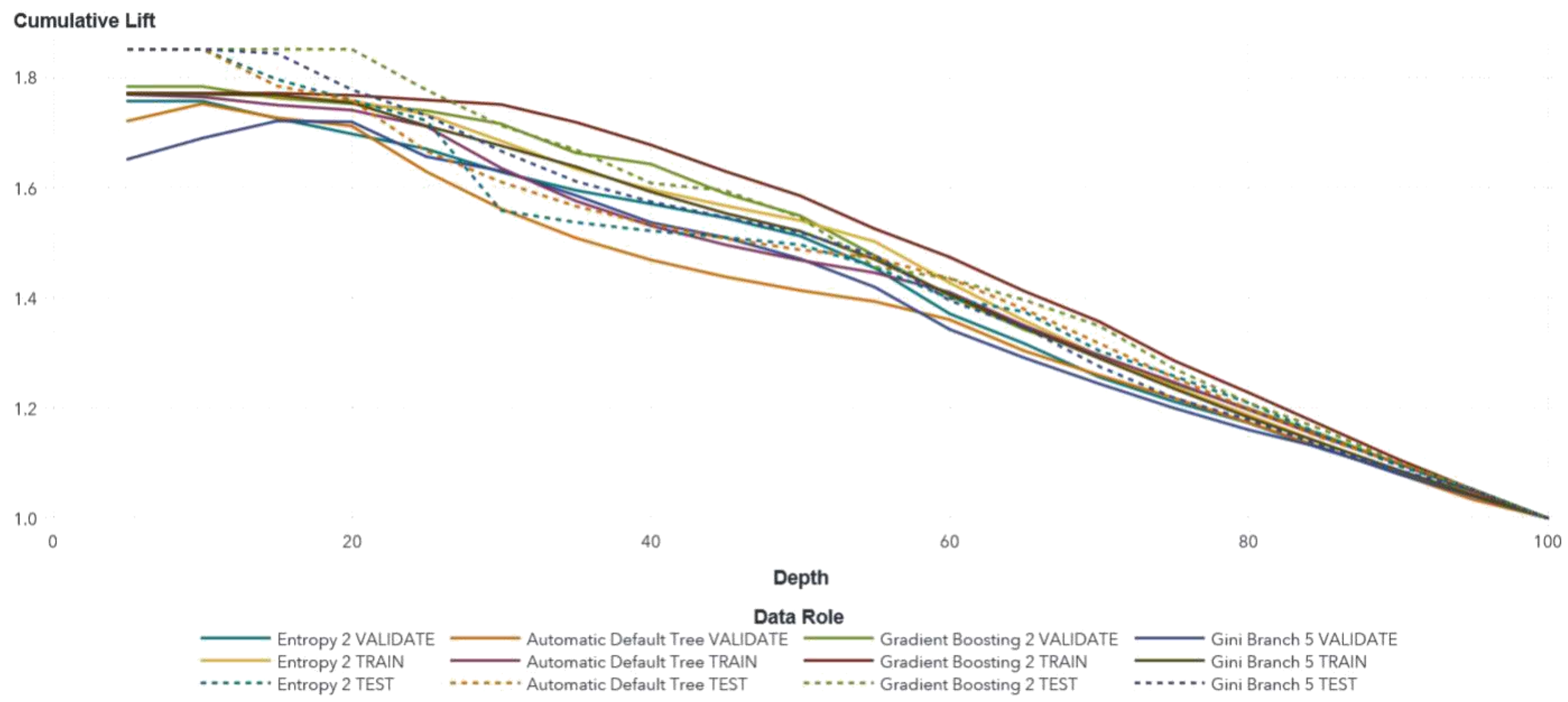
ROC Reports

View chart: Accuracy ▾



Lift Reports

View chart: Cumulative Lift



Fit Statistics

Statistic...	Train: Gi...	Validate:...	Test: Gin...	Train: Gr...	Validate:...	Test: Gr...	Train: Au...	Validate:...	Test: Aut...	Train: En...	Validate:...	Test: Ent...
Area Under ROC	0.8691	0.8269	0.8354	0.9226	0.8773	0.8690	0.8600	0.8118	0.8326	0.8777	0.8416	0.8348
Average Squared Error	0.1434	0.1676	0.1624	0.1139	0.1417	0.1452	0.1467	0.1743	0.1589	0.1352	0.1554	0.1669
Divisor for ASE	2,297	1,128	381	2,297	1,128	381	2,297	1,128	381	2,297	1,128	381
Formatted Partition	1	0	2	1	0	2	1	0	2	1	0	2
Gamma	0.7802	0.6992	0.7207	0.8637	0.7760	0.7601	0.8018	0.7017	0.7493	0.8019	0.7302	0.7273
Gini Coefficient	0.7383	0.6538	0.6708	0.8452	0.7546	0.7381	0.7201	0.6237	0.6653	0.7554	0.6833	0.6697
KS (Youden)	0.6027	0.5319	0.5439	0.6719	0.6218	0.5820	0.5721	0.4925	0.5563	0.6397	0.5886	0.5264
KS Cutoff	0.5500	0.6500	0.6500	0.6000	0.6000	0.6000	0.4500	0.4500	0.4500	0.4500	0.6000	0.4500
Misclassification at Cutoff	0.2002	0.2367	0.2309	0.1697	0.2118	0.2362	0.2072	0.2455	0.2125	0.1802	0.2065	0.2335
Misclassification Rate	0.2002	0.2367	0.2309	0.1697	0.2118	0.2362	0.2072	0.2455	0.2125	0.1802	0.2065	0.2335
Misclassification Rate (Event)	0.2002	0.2367	0.2309	0.1697	0.2118	0.2362	0.2072	0.2455	0.2125	0.1802	0.2065	0.2335
Multi-Class Log Loss	0.4355	0.6016	0.4845	0.3584	0.4311	0.4375	0.4429	0.6815	0.4865	0.4165	0.5119	0.5036
Partition Indicator	1	0	2	1	0	2	1	0	2	1	0	2
ROC Separation	0.6020	0.5293	0.5396	0.6520	0.5674	0.5160	0.5721	0.4925	0.5563	0.6397	0.5873	0.5264
Root Average Squared Error	0.3787	0.4093	0.4029	0.3375	0.3765	0.3811	0.3830	0.4175	0.3987	0.3677	0.3942	0.4086
Sum of Frequencies	2,297	1,128	381	2,297	1,128	381	2,297	1,128	381	2,297	1,128	381
Target Name	churn	churn	churn	churn	churn	churn	churn	churn	churn	churn	churn	churn
Tau	0.3630	0.3214	0.3302	0.4156	0.3709	0.3634	0.3541	0.3066	0.3275	0.3714	0.3359	0.3297

Appendix 6 Scored Data

First 65 rows whole sheet would be sent to Telco ABC or can be merged into CRM.

_leaf_id_	children	churn	credit	creditCard	custcare	custcareLast	custcareTotal	customerID	directas	directasLast	dropvce	dropvceLast	l_churn	income	marry	mou	mouChange	mouTotal	occupation	outcalls	overage	overageMax	overageMin	P
5	0	1	a	1	0	0	0	1000001	0	0	1	1	1	4	no	236	0	235.859	professiona	88	0	0	0	C
13	0	1	de	1	0	0	0	1000022	0	0	24	20	1	5	yes	###	-0.0503936	4331.89		643	0	0	0	
47	0	0	a	1	0	0	0	1000034	0	0	1	1	0	7	yes	45	-0.1348335	270.971	professiona	21	14.27	15.2106629	12.634987	
43	1	0	a	1	0	0	0	1000100	0	0	7	6	0	7	yes	540	0.01146045	3241.7	homemake	135	0	0	0	
20	0	0	b	0	5	4	27	1000143	44	40	10	11	0	5	unkno	###	0.00802144	19110		1108	0	0	0	C
15	1	0	a	0	0	0	0	1000164	0	0	2	2	0	4	yes	109	-0.0942105	655.791		46	0	0	0	
25	0	0	z	0	4	3	21	1000196	1	1	3	4	0	0	unkno	283	0.01743896	1700.16		117	64.73	68.4987788	59.71014	C
20	1	1	a	1	5	6	27	1000275	0	0	2	2	0	2	yes	###	0.02662476	5553.25		186	256.49	272.971664	225.99217	C
16	1	1	a	1	0	0	0	1000291	0	0	0	0	1	5	unkno	221	0.00531999	442.282		54	0	0	0	
20	0	1	a	1	0	0	0	1000334	1	1	8	9	0	4	no	###	-0.054444	5969.27		492	513.33	537.163576	489.16523	C
51	0	1	a	1	0	0	0	1000447	0	0	0	0	0	9	yes	3.4	0.15649475	13.6383		1	0	0	0	C
26	0	0	a	0	0	0	0	1000455	1	1	2	2	0	0	unkno	134	0.03101354	804.813		39	0	0	0	
16	0	1	c	0	1	1	2	1000482	1	1	6	6	1	0	unkno	448	0.03725323	896.655		134	41.862	42.1345605	41.588667	
16	0	1	a	1	0	0	0	1000579	1	1	4	4	1	9	yes	167	0.0909387	667.749		85	0	0	0	
20	1	0	b	1	1	1	6	1000636	1	1	30	24	0	9	no	991	0.08890956	5947.66		194	84.546	89.2380742	79.713669	C
5	0	1	c	0	0	0	0	1000660	0	0	0	0	1	0	unkno	26	-0.9957893	26.4034		2	0	0	0	C
21	0	1	a	1	1	1	2	1000715	3	3	46	45	1	8	unkno	###	0.11363457	3957.17	professiona	681	141.22	145.551324	136.89467	
27	1	1	a	0	0	0	0	1000743	0	0	7	7	1	4	yes	560	0.03459586	2238.47		259	59.279	62.2518048	53.64593	
16	0	1	a	0	0	0	0	1000865	0	0	0	0	1	0	unkno	28	0.03390279	85.3522		23	0	0	0	
9	0	0	a	0	0	0	0	1000895	2	2	1	1	0	0	unkno	132	0.01286107	792.995		38	0	0	0	
20	0	0	b	0	0	0	0	1000929	22	20	32	28	0	3	no	###	-0.0106419	8438.1		566	240.97	255.850269	222.93762	C
9	0	0	a	1	3	3	18	1000971	0	0	8	7	0	6	yes	904	-0.0420588	5421.94		201	0	0	0	
15	0	0	a	0	0	0	0	1000995	0	0	2	3	0	6	yes	119	0.10711229	712.018		37	38.85	40.7809484	36.129423	
31	1	0	b	1	0	0	0	1001022	0	0	0	0	0	7	yes	272	0.03040517	1629.55	professiona	97	0	0	0	
12	0	0	a	1	0	0	0	1001058	0	0	2	2	0	8	no	179	-0.0532062	1076.02		33	0	0	0	
16	0	1	a	1	0	0	0	1001068	0	0	0	0	1	6	yes	26	-0.0161111	78.7742	professiona	4	0	0	0	
27	0	1	a	0	0	0	0	1001121	0	0	0	0	1	0	unkno	23	0.06429097	91.1605		5	0	0	0	
5	0	1	b	0	0	0	0	1001211	0	0	7	7	1	0	unkno	457	-0.1429329	457.326		213	38.87	38.8700644	38.870064	C
9	0	0	a	1	0	0	0	1001219	0	0	0	0	0	7	unkno	326	0.0107794	1956.79		0	0	0	0	
20	0	0	a	0	0	0	0	1001288	2	1	8	8	0	0	unkno	###	-0.0974034	9557.37		680	0	0	0	C
23	0	0	a	1	0	0	0	1001331	0	0	4	5	0	6	no	430	0.09566554	2580.64		203	0	0	0	

overage	overageMax	overageMin	P_churn0	P_churn1	peakOffPeak	peakOffPeakLast	Predicted for churn	Probability for churn =1	Probability of Classification	recchrge	regionType	revenue	revenueChange	revenueTotal	roam
0	0	0	0.041344	0.958656	2.38461538	2.384615385	1	0.958656331	0.958656331	22.5	suburban	23.98	0	23.98	0
0	0	0	0	1	2.6346516	2.63030303	1	1	1	56.25	suburban	61.11	0.014299486	183.33	0
14.27	15.2106629	12.634987	1	0	0.5060241	0.461538462	0	0	1	16.99	town	21.13	0.081067721	126.78	0
0	0	0	1	0	5.42063492	5.380952381	0	0	1	75	rural	75.075	0	450.45	0
0	0	0	0.824427	0.175573	4.31070288	4.317073171	0	0.175572519	0.824427481	212.99	unknown	262.53	-0.061377025	1575.18	0
0	0	0	1	0	5.825	5.428571429	0	0	1	31.98		32.352	0.037869458	194.11	0
64.73	68.4987788	59.71014	0.914894	0.085106	4.99145299	5.1	0	0.085106383	0.914893617	30		46.912	0.019813001	281.47	0
256.49	272.971664	225.99217	0.824427	0.175573	3.50485437	3.487179487	0	0.175572519	0.824427481	60	s	128.03	-0.058280793	640.15	0
0	0	0	0	1	3.90909091	3.909090909	1	1	1	50		55.535	-0.036244696	111.07	0
513.33	537.163576	489.16523	0.824427	0.175573	2.32040472	2.319444444	0	0.175572519	0.824427481	59.99		206.17	0.026580495	824.69	0
0	0	0	0.571429	0.428571	0	0	0	0.428571429	0.571428571	16.99	suburban	17.225	-0.052426102	68.9	0
0	0	0	1	0	1.19626168	1.222222222	0	0	1	69.99		69.99	0	419.94	0
41.862	42.1345605	41.588667	0	1	5.84615385	5.8	1	1	1	50	suburban	55.715	0.054766734	111.43	0
0	0	0	0	1	2.62765957	2.64	1	1	1	30	suburban	30.575	0.02295082	122.3	0
84.546	89.2380742	79.713669	0.824427	0.175573	3.18705036	3.181818182	0	0.175572519	0.824427481	84.99	suburban	103.16	0.037228845	618.96	0
0	0	0	0.041344	0.958656	0	0	1	0.958656331	0.958656331	19.91	town	20.73	-0.962218415	20.73	0
141.22	145.551324	136.89467	0	1	1.73843058	1.736641221	1	1	1	79.99		126	-0.026857725	251.99	0
59.279	62.2518048	53.64593	0	1	8.96153846	9	1	1	1	50		63.22	-0.054503464	252.88	0
0	0	0	0	1	1.8	1.777777778	1	1	1	47.17		47.17	0	141.51	0
0	0	0	1	0	1.92207792	1.916666667	0	0	1	44.99		44.99	0	269.94	0
240.97	255.850269	222.93762	0.824427	0.175573	2.88888889	2.888111888	0	0.175572519	0.824427481	119.99		217.5	-0.03549882	1305.01	4
0	0	0	1	0	1.01505017	1.019607843	0	0	1	44.99	suburban	44.99	0	269.94	0
38.85	40.7809484	36.129423	1	0	1.96	2	0	0	1	16.99		31.243	0.034582133	187.46	0
0	0	0	1	0	10.4509804	9.888888889	0	0	1	35.99	town	35.99	0	215.94	0
0	0	0	1	0	1.3902439	1.384615385	0	0	1	14.99	suburban	28.355	0.014508524	170.13	3
0	0	0	0	1	0	0	1	1	1	29.99	suburban	30.39	-0.03847387	91.17	0
0	0	0	0	1	3.75	4	1	1	1	69.99	suburban	69.99	0	279.96	0
38.87	38.8700644	38.870064	0.041344	0.958656	0.90178571	0.901785714	1	0.958656331	0.958656331	35		46.5	0.033563014	46.5	0
0	0	0	1	0	0	0	0	0	1	44.99	suburban	44.99	0	269.94	0
0	0	0	0.824427	0.175573	4.54076087	4.52991453	0	0.175572519	0.824427481	154.99	suburban	160.28	0.005881609	961.69	0
0	0	0	1	0	3.92307692	3.931818182	0	0	1	50.01		51.4	0.023583994	308.4	2

34	5	0	1	c	1	0	0	0	1001488	0	0	0	0	1	8	unkno	21	-0.8883309	20.9135		8	0	0	0	0.0
35	9	0	0	a	0	0	0	0	1001551	0	0	1	1	0	0	unkno	386	0.01798841	2318.88		181	0	0	0	
36	31	0	0	gy	0	0	0	0	1001636	1	1	1	1	0	0	unkno	319	0.17540463	1911.94		124	12.296	13.4801951	11.493086	
37	20	0	0	a	0	12	11	74	1001717	0	0	45	41	0	4	no	###	0.00734902	18797.6		769	252.96	262.438013	236.81535	0.8
38	43	0	0	c	0	2	2	12	1001737	1	1	19	21	0	0	unkno	###	0.12252506	7494.21		632	0.244	0.52820975	0	
39	5	1	1	a	1	0	0	0	1001880	0	0	3	3	1	9	yes	99	-0.818027	99.4911		48	0	0	0	0.0
40	31	1	0	a	1	0	0	0	1001937	0	0	2	2	0	6	no	337	-0.155891	2024.04		146	5.3261	5.87123653	5.0889847	
41	33	0	1	a	1	0	0	0	1001943	0	0	0	0	0	7	unkno	0	0	0		0	0	0	0	0.7
42	26	0	0	b	0	3	2	17	1001987	1	1	9	8	0	0	unkno	508	-0.0349815	3045.48		148	3.6599	11.433884	0	
43	26	0	0	b	0	0	0	0	1002002	2	2	4	3	0	0	unkno	402	0.14471782	2409.13		184	90.358	94.9457186	86.310822	
44	20	0	0	aa	0	0	0	0	1002023	4	4	8	7	0	0	unkno	###	0.12884511	15300.7		589	82.43	86.6086132	78.676634	0.8
45	9	0	0	a	0	0	0	0	1002128	0	0	1	1	0	7	yes	136	-0.0272244	815.85		28	0	0	0	
46	43	0	0	a	1	0	0	0	1002167	0	0	9	9	0	6	no	###	-0.0145064	6513.56		300	22.354	23.7678662	19.793369	
47	13	0	1	a	1	1	1	5	1002483	0	0	0	0	1	9	yes	240	-0.0498663	1198.66	professiona	56	0	0	0	
48	9	0	0	a	0	0	0	0	1002494	0	0	5	5	0	0	unkno	116	0.0338289	694.52		59	0	0	0	
49	34	0	1	gy	1	0	0	0	1002544	0	0	0	0	1	9	no	0	0	0		0	0	0	0	
50	9	0	0	c	1	6	7	36	1002561	1	0	3	2	0	7	no	460	0.02847979	2760.78	clerical	96	13.785	15.4165267	13.183768	
51	5	1	1	a	1	0	0	0	1002669	0	0	4	4	1	8	unkno	233	0	232.524	professiona	75	0	0	0	0.0
52	5	0	1	a	1	0	0	0	1002701	0	0	1	1	1	7	unkno	149	-0.8715228	148.776		34	5.0792	5.07921872	5.0792187	0.0
53	16	0	1	de	0	4	4	7	1002723	0	0	2	2	1	0	unkno	123	-0.0252906	246.8		43	1.3168	1.320687	1.3129981	
54	31	0	0	a	1	1	1	6	1002775	0	0	1	1	0	6	no	60	0.01984567	359.32	professiona	40	31.41	33.251962	28.952289	
55	43	0	0	b	1	0	0	0	1007088	0	1	4	4	0	6	unkno	751	0.0023604	4508.63		259	0	0	0	
56	23	0	0	b	0	7	8	41	1007102	0	0	6	5	0	0	unkno	559	0.00713635	3354.05		284	32.884	34.4586991	31.296583	
57	11	1	1	a	1	0	0	0	1007168	2	2	5	5	1	6	yes	172	-0.0182783	859.414	professiona	84	4.8738	5.16121143	4.3788681	
58	5	0	1	a	1	0	0	0	1007216	0	0	5	5	1	4	yes	98	-0.0739427	98.2304	crafts	33	10.435	10.4352771	10.435277	0.0
59	5	1	1	gy	1	0	0	0	1007313	0	0	1	1	1	7	yes	242	0.14804301	241.501	professiona	36	0	0	0	0.0
60	16	0	1	a	1	0	0	0	1007552	0	0	1	1	1	5	no	65	-0.0100717	129.344		24	0	0	0	
61	12	0	0	c	1	0	0	0	1007623	0	0	0	0	0	6	no	57	0.06271472	344.798		24	0	0	0	
62	31	1	0	b	1	0	0	0	1007644	0	0	0	0	0	4	yes	139	-0.0123123	832.079		20	6.6454	7.37832694	6.1094068	
63	5	0	1	a	1	0	0	0	1007731	0	0	0	0	1	8	unkno	0	-1	0	professiona	0	0	0	0	0.0
64	13	1	1	b	1	0	0	0	1007750	0	0	0	0	1	9	yes	136	0.16578674	272.494		50	0	0	0	
65	5	1	1	b	1	0	0	0	1007950	1	1	0	0	1	6	yes	813	0.02378584	2438.13		0	0	0	0	0.0

33		38	0	0	0	0.041344	0.958656	2.16666667	2.166666667	1	0.958656331	0.958656331	53.29		53.29	-0.111685281	53.29	2
34		8	0	0	0	0.041344	0.958656	0.6	0.6	1	0.958656331	0.958656331	30		30	-0.51171875	30	0
35		181	0	0	0	1	0	0.39130435	0.392592593	0	0	1	44.99		44.99	0	269.94	0
36		124	12.296	13.4801951	11.493086	1	0	0.73770492	0.736842105	0	0	1	30	town	34.867	0.005528077	209.2	0
37		769	252.96	262.438013	236.81535	0.824427	0.175573	0.98195101	0.982543641	0	0.175572519	0.824427481	75		137.71	-0.001982233	826.27	0
38		632	0.244	0.52820975	0	1	0	2.32136602	2.317073171	0	0	1	77.96		79.702	-0.040137897	478.21	0
39		48	0	0	0	0.041344	0.958656	4.33333333	4.333333333	1	0.958656331	0.958656331	29.99		30.97	-0.380476095	30.97	0
40		146	5.3261	5.87123653	5.0889847	1	0	2.68776371	2.675675676	0	0	1	34.19		36.233	0.008871639	217.4	0
41		0	0	0	0	0.705882	0.294118	0	0	0	0.294117647	0.705882353	29.99	rural	29.99	0	149.95	0
42		148	3.6599	11.433884	0	1	0	2.60323887	2.615384615	0	0	1	67.7		67.895	0.007680945	407.37	0
43		184	90.358	94.9457186	86.310822	1	0	1.48758465	1.5	0	0	1	34.98		65.09	0.024889729	390.54	1
44		589	82.43	86.6086132	78.676634	0.824427	0.175573	1.26231606	1.260536399	0	0.175572519	0.824427481	79.99	suburban	137.72	-0.065431059	826.31	0
45		28	0	0	0	1	0	2.60869565	2.5	0	0	1	44.95		44.95	0	269.7	0
46		300	22.354	23.7678662	19.793369	1	0	5.92692308	5.906976744	0	0	1	74.99	town	80.027	0.069625096	480.16	0
47	professiona	56	0	0	0	0	1	1.98924731	1.944444444	1	1	1	44.99	suburban	44.99	0	224.95	0
48		59	0	0	0	1	0	4.04285714	4.166666667	0	0	1	44.99	suburban	44.99	0	269.94	0
49		0	0	0	0	0	1	0	0	1	1	1	28.63		28.85	0.015368495	57.7	0
50	clerical	96	13.785	15.4165267	13.183768	1	0	2.38235294	2.392857143	0	0	1	20.32		41.18	-0.064387731	247.08	6
51	professiona	75	0	0	0	0.041344	0.958656	2	2	1	0.958656331	0.958656331	30		37.38	1.131128848	37.38	0
52		34	5.0792	5.07921872	5.0792187	0.041344	0.958656	0.61904762	0.619047619	1	0.958656331	0.958656331	29.99	suburban	33.22	-0.855136927	33.22	0
53		43	1.3168	1.320687	1.3129981	0	1	0.65384615	0.653846154	1	1	1	29.75		38.17	-0.074653216	76.34	0
54	professiona	40	31.41	33.251962	28.952289	1	0	6.11764706	5.833333333	0	0	1	30.02		35.275	0.011235955	211.65	0
55		259	0	0	0	1	0	0.62722513	0.628205128	0	0	1	76.76		76.76	0	460.56	0
56		284	32.884	34.4586991	31.296583	1	0	0.66731898	0.666666667	0	0	1	43.38	suburban	49.44	-0.058742902	296.64	0
57	professiona	84	4.8738	5.16121143	4.3788681	0	1	5.29850746	5.214285714	1	1	1	30	town	34.178	0.008985507	170.89	0
58	crafts	33	10.435	10.4352771	10.435277	0.041344	0.958656	2	2	1	0.958656331	0.958656331	24.99		30.34	-0.069325153	30.34	0
59	professiona	36	0	0	0	0.041344	0.958656	0.56521739	0.565217391	1	0.958656331	0.958656331	34.99	rural	45.97	2.93916024	45.97	2
60		24	0	0	0	0	1	1.4	1.4	1	1	1	32.49	suburban	32.49	0	64.98	0
61		24	0	0	0	1	0	2.04166667	2.125	0	0	1	29.99	town	30.395	0	182.37	0
62		20	6.6454	7.37832694	6.1094068	1	0	0.53246753	0.538461539	0	0	1	32.49		34.582	0.04981774	207.49	0
63	professiona	0	0	0	0	0.041344	0.958656	0	0	1	0.958656331	0.958656331	24.99	suburban	24.99	-0.167	24.99	0

Appendix 7 Cluster 1 Customer Satisfaction

	custcareTotal							
	count	mean	std	min	25%	50%	75%	max
segment								
1	2573	1.649048	6.571597	0	0	0	0	120
2	865	17.82659	82.5978	0	0	5	17	2253
3	435	5.657471	12.30376	0	0	0	6	93
4	428	5.92757	12.51068	0	0	0	6	107

	directas							
	count	mean	std	min	25%	50%	75%	max
segment								
1	2573	0.155849	0.391643	0	0	0	0	2
2	865	0.627746	1.029728	0	0	0	1	9
3	435	3.698851	3.500345	0	2	3	5	28
4	428	1.007009	1.658827	0	0	1	1	12

	dropvce							
	count	mean	std	min	25%	50%	75%	max
segment								
1	2573	1.388651	1.942941	0	0	1	2	16
2	865	12.20694	11.58593	0	5	9	15	109
3	435	5.468966	5.569488	0	2	4	7	39
4	428	6.60514	7.342773	0	2	5	8	66

	mouTotal							
	count	mean	std	min	25%	50%	75%	max
segment								
1	2573	711.9492	806.6208	0	118.1793	434.1814	1059.174	6315.982
2	865	3602.37	2863.081	176.1932	1721.268	3013.064	4640.099	20848.97
3	435	2446.943	2276.325	77.61443	854.0334	1705.05	3216.755	14983.14
4	428	3681.592	3610.637	275.8614	1257.704	2526.472	4884.146	27430.24

	outcalls							
	count	mean	std	min	25%	50%	75%	max
segment								
1	2573	42.48037	45.46817	0	2	30	70	278
2	865	287.3156	192.7457	0	166	235	346	2235
3	435	178.0644	148.0943	0	89.5	147	231.5	970
4	428	228.9486	186.5549	0	126.75	183	288.25	1263

	overage							
	count	mean	std	min	25%	50%	75%	max
segment								
1	2573	6.022448	15.69348	0	0	0	0.213621	130.4379
2	865	30.39666	59.95388	0	0	2.211363	44.33262	1039.54
3	435	40.2587	62.60565	0	0	7.01082	58.91996	436.4495
4	428	198.974	178.0674	3.684986	89.45344	149.4711	242.7927	1512.281

	roam							
	count	mean	std	min	25%	50%	75%	max
segment								
1	2573	0.347454	1.289595	0	0	0	0	17
2	865	0.667052	2.246425	0	0	0	0	28
3	435	6.108046	15.60825	0	0	0	4	149
4	428	0.873832	2.503013	0	0	0	1	26

```
X_0.groupby('segment')['churn'].value_counts()
```

```
segment  churn
0        1        1527
         0        1046
1        0         462
         1         403
2        1         245
         0         190
3        1         255
         0         173
Name: churn, dtype: int64
```

```
X_0.groupby('segment')['credit'].value_counts()
```

```
segment  credit
0        aa        1122
         a          540
         b          389
         c          189
         de         186
         z           83
         gy          64
1        aa         278
         b          152
         de         133
         c          113
         a          110
         z           56
         gy          23
2        aa         161
         b           77
         a           73
         de          66
         c           40
         z           11
         gy           7
3        aa         141
         b           95
         de          59
         c           50
         a           46
         z           24
         gy          13
Name: credit, dtype: int64
```

```
X_0.groupby('segment')['occupation'].value_counts()
```

```
segment  occupation
0        missing          1788
         professional      516
         crafts             80
         retired            59
         clerical           58
         self-employed      46
         student            22
         homemaker           4
1        missing           666
         professional      116
         crafts             29
         clerical           21
         self-employed      18
         student             9
         retired             6
2        missing           346
         professional       68
         self-employed       8
         crafts              5
         clerical            4
         retired             2
         student             2
3        missing           348
         professional       56
         crafts             11
         clerical            8
         self-employed       3
         student             2
Name: occupation, dtype: int64
```

```
X_0.groupby('segment')['regionType'].value_counts()
```

```
segment  regionType
0        missing       1144
         suburban       866
         town           411
         rural          152
1        missing        482
         suburban       242
         town           107
         rural           34
2        missing        210
         suburban       132
         town            67
         rural           26
3        missing        214
         suburban       138
         town            57
         rural           19
Name: regionType, dtype: int64
```

	Churn	Credit	Occupation	Income	Region type
segment1 2573 (59.82%)	About 60% of the customers choose churn	64.59% of customers were high-rated (a,aa). 5.7% of the customers were low-rated customers (gy,z).	The largest proportion is missing data about 70%.Next is professional, at about 20% It's the only segment in cluster2 that has all occupation types.	On average, this group had the highest income at 4.58.	The largest proportion is missing data about 45%. Next is suburban, at about 33.7%. The smallest number of people live in rural areas, only about 6 %
segment2 865 (20.11%)	53.4% of customers will not choose to churn. It's the only segment in cluster1 that has more people staying than leaving	44.86% of customers were high-rated (a,aa). 9.13% of the customers were low-rated customers (gy,z).	The largest proportion is missing data about 78%. Next is professional, at about 13% No homemaker	On average, this group had the lowest income at 3.85.	The largest proportion is missing data about 56%. Next is suburban, at about 28%
segment3 435 (10.11%)	With 56.32% choosing churn	53.79% of customers were high-rated (a,aa). 4.14% of the customers were low-rated customers (gy,z).	The largest proportion is missing data about 80%. Next is professional, at about 15.6%. No homemaker	On average, this group had the second highest income at 4.34.	The largest proportion is missing data about 48%. Next is suburban, at about 30%
Segment4 428 (9.95%)	With 59.6% choosing churn	43.69% of customers were high-rated (a,aa). 8.64% of the customers were low-rated customers (gy,z).	The largest proportion is missing data about 81.3%. Next is professional, at about 13% No homemaker, retired	On average, this group had the second lowest income at 3.95.	The largest proportion is missing data about 50%. Next is suburban, at about 32%

Appendix 8 Cluster 2 Customer Contribution

```
X_0.groupby('segment')['revenueChange'].describe()    # numerical
```

segment	count	mean	std	min	25%	50%	75%	max
0	2085.0	-0.059793	0.216868	-1.000000	-0.031786	0.000000	0.008165	1.286898
1	495.0	-0.021749	0.153048	-0.788919	-0.020739	0.000000	0.013598	0.888420
2	459.0	1.047224	2.060536	-0.591422	0.000000	0.222272	1.132611	14.463000
3	1262.0	-0.000316	0.031806	-0.086298	-0.013700	0.000000	0.011175	0.097727

```
X_0.groupby('segment')['revenueTotal'].describe()    # numerical
```

3]:

segment	count	mean	std	min	25%	50%	75%	max
0	2085.0	130.193914	84.967082	0.00	59.020	117.160	193.130	463.60
1	495.0	207.228909	112.085339	11.26	119.325	201.330	273.105	634.30
2	459.0	111.895948	54.501255	31.56	71.060	98.620	143.240	458.05
3	1262.0	436.158764	251.705370	64.50	295.440	360.565	497.585	3463.27

```
X_0.groupby('segment')['mouTotal'].describe()    # numerical
```

]:

segment	count	mean	std	min	25%	50%	75%	max
0	2085.0	606.173201	661.846483	0.000000	101.015750	373.403221	927.738818	4209.902519
1	495.0	947.152771	780.675964	12.985968	377.145076	788.936568	1323.124782	5589.956496
2	459.0	1062.264276	682.779159	0.000000	568.710814	922.899033	1397.810962	5687.757309
3	1262.0	4253.364178	3022.500106	0.000000	2370.342419	3452.209677	5211.606006	27430.238983

```
X_0.groupby('segment')['mouChange'].describe()    # numerical
```

:

segment	count	mean	std	min	25%	50%	75%	max
0	2085.0	0.053029	1.588887	-1.000000	-0.090197	-0.005555	0.051139	42.452077
1	495.0	0.499457	12.042880	-0.980959	-0.088731	-0.023057	0.046143	267.712621
2	459.0	9.955965	64.607512	-1.000000	-0.038396	0.146720	2.881694	967.417868
3	1262.0	0.002672	0.079992	-0.185868	-0.056414	0.000598	0.058489	0.238160

```
X_0.groupby('segment')['outcalls'].describe()    # numerical
```

segment	count	mean	std	min	25%	50%	75%	max
0	2085.0	43.575060	52.005017	0.0	0.00	25.0	70.0	315.0
1	495.0	79.525253	61.659727	4.0	34.00	62.0	110.0	344.0
2	459.0	263.969499	196.270220	0.0	149.00	218.0	333.5	2235.0
3	1262.0	223.373217	192.798488	0.0	105.25	178.0	293.0	1287.0

```
X_0.groupby('segment')['recchrge'].describe()    # numerical
```

segment	count	mean	std	min	25%	50%	75%	max
0	2085.0	30.950139	13.889321	0.00	20.00	30.00	39.99	79.99
1	495.0	42.828929	16.759423	9.02	30.00	39.99	50.00	149.99
2	459.0	57.752331	20.172979	0.00	44.99	54.99	69.99	169.99
3	1262.0	59.962345	23.843965	5.16	44.99	57.45	69.99	309.99

```
X_0.groupby('segment')['peakOffPeak'].describe()    # numerical
```

segment	count	mean	std	min	25%	50%	75%	max
0	2085.0	1.153415	1.113529	0.000000	0.000000	1.000000	1.818182	5.500000
1	495.0	6.922992	6.816003	1.596154	3.700575	4.971429	7.120192	79.333333
2	459.0	1.521390	1.116076	0.000000	0.759455	1.251497	2.000000	7.954545
3	1262.0	1.491385	1.376814	0.000000	0.716375	1.164935	1.923216	14.025510

```
segment  churn
0        1        1360
         0         725
1        1         273
         0         222
2        1         459
3        0         924
         1         338
Name: churn, dtype: int64
```

```
X_0.groupby('segment')['credit'].
```

```
]: segment  credit
   0        aa        893
            a         440
            b         315
            de        160
            c         152
            z          74
            gy         51
   1        aa        226
            a         121
            b          62
            c          38
            de         31
            gy         13
            z           4
   2        aa        153
            b          91
            de         64
            a          56
            c          52
            z          31
            gy         12
   3        aa        430
            b         245
            de        189
            a         152
            c         150
            z          65
            gy         31
   Name: credit, dtype: int64
```

```python
X_0.groupby('segment')['custcareTotal'].describe()   # numerical
```

segment	count	mean	std	min	25%	50%	75%	max
0	2085.0	1.530456	6.373429	0.0	0.0	0.0	0.0	120.0
1	495.0	1.949495	8.315604	0.0	0.0	0.0	0.0	142.0
2	459.0	3.529412	7.493571	0.0	0.0	1.0	4.0	81.0
3	1262.0	14.964342	69.033117	0.0	0.0	3.0	15.0	2253.0

```python
X_0.groupby('segment')['custcareLast'].describe()   # numerical
```

]:

segment	count	mean	std	min	25%	50%	75%	max
0	2085.0	0.436930	1.498713	0.0	0.0	0.0	0.0	17.0
1	495.0	0.442424	1.559942	0.0	0.0	0.0	0.0	22.0
2	459.0	2.586057	4.939227	0.0	0.0	1.0	3.0	31.0
3	1262.0	2.733756	11.928831	0.0	0.0	1.0	3.0	387.0

```python
X_0.groupby('segment')['dropvce'].describe()   # numerical
```

segment	count	mean	std	min	25%	50%	75%	max
0	2085.0	1.820624	3.213524	0.0	0.0	1.0	2.0	60.0
1	495.0	2.606061	3.272441	0.0	1.0	2.0	3.0	34.0
2	459.0	10.520697	12.579056	0.0	3.0	7.0	13.0	109.0
3	1262.0	7.466719	8.770223	0.0	2.0	5.0	10.0	97.0

```python
X_0.groupby('segment')['income'].describe()   # numerical
```

]:

segment	count	mean	std	min	25%	50%	75%	max
0	2085.0	4.566427	3.094175	0.0	1.0	5.0	7.0	9.0
1	495.0	4.969697	3.001533	0.0	3.0	6.0	7.0	9.0
2	459.0	4.087146	3.153467	0.0	0.0	5.0	7.0	9.0
3	1262.0	3.851823	3.192886	0.0	0.0	4.0	6.0	9.0

```python
X_0.groupby('segment')['mou'].describe()    # numerical
```

segment	count	mean	std	min	25%	50%	75%	max
0	2085.0	179.884895	183.389939	0.000000	36.003345	131.619609	267.185991	1343.108323
1	495.0	231.576517	172.333958	9.222124	108.507794	197.234142	311.367896	931.659416
2	459.0	795.977941	548.639300	0.000000	459.859773	675.658834	1005.734364	5687.757309
3	1262.0	802.739392	557.763081	0.000000	438.874909	648.356996	1018.060091	4571.706497

```
X_0.groupby('segment')['overage'].describe()   # numerical
```

segment	count	mean	std	min	25%	50%	75%	max
0	2085.0	10.215347	27.930318	0.0	0.0	0.00000	4.077513	358.910218
1	495.0	17.897293	39.431948	0.0	0.0	0.00000	15.137035	275.855594
2	459.0	70.798227	121.094315	0.0	0.0	29.29401	96.978508	1121.700546
3	1262.0	64.823879	127.291121	0.0	0.0	6.78374	77.803362	1512.281420

```
X_0.groupby('segment')['peakOffPeakLast'].describe()   # numerical
```

]:

segment	count	mean	std	min	25%	50%	75%	max
0	2085.0	1.138658	1.098157	0.000000	0.000000	1.000000	1.800000	5.500000
1	495.0	6.771593	6.674503	1.571429	3.628289	4.882353	7.000000	79.333333
2	459.0	1.519905	1.115503	0.000000	0.758882	1.250000	2.000000	7.954545
3	1262.0	1.483134	1.365883	0.000000	0.714471	1.162585	1.909965	13.800000

```python
X_0.groupby('segment')['revenue'].describe()   # numerical
```

segment	count	mean	std	min	25%	50%	75%	max
0	2085.0	36.252388	17.336045	0.000000	29.528333	36.350000	44.990000	194.780000
1	495.0	50.127970	20.069233	10.000000	35.651000	46.236667	59.990000	155.830000
2	459.0	83.218872	42.250658	22.266667	59.990000	73.310000	94.295000	458.050000
3	1262.0	82.210255	46.460662	10.750000	54.439167	69.990000	93.478958	577.211667

```python
X_0.groupby('segment')['roam'].describe()   # numerical
```

]:

segment	count	mean	std	min	25%	50%	75%	max
0	2085.0	0.567386	2.619301	0.0	0.0	0.0	0.0	48.0
1	495.0	0.814141	2.853686	0.0	0.0	0.0	0.0	29.0
2	459.0	1.884532	8.393031	0.0	0.0	0.0	1.0	111.0
3	1262.0	1.625198	7.852108	0.0	0.0	0.0	0.0	149.0

```
X_0.groupby('segment')['credit'].value_counts()
```

```
segment  credit
0        aa        1122
         a          540
         b          389
         c          189
         de         186
         z           83
         gy          64
1        aa         278
         b          152
         de         133
         c          113
         a          110
         z           56
         gy          23
2        aa         161
         b           77
         a           73
         de          66
         c           40
         z           11
         gy           7
3        aa         141
         b           95
         de          59
         c           50
         a           46
         z           24
         gy          13
Name: credit, dtype: int64
```

```
X_0.groupby('segment')['occupation'].value_counts()
```

```
segment  occupation
0        missing          1788
         professional      516
         crafts             80
         retired            59
         clerical           58
         self-employed      46
         student            22
         homemaker           4
1        missing           666
         professional      116
         crafts             29
         clerical           21
         self-employed      18
         student             9
         retired             6
2        missing           346
         professional       68
         self-employed       8
         crafts              5
         clerical            4
         retired             2
         student             2
3        missing           348
         professional       56
         crafts             11
         clerical            8
         self-employed       3
         student             2
Name: occupation, dtype: int64
```

	Churn	Credit	Income	CustcareTotal	CustcareLast	Dropvce	Mou
segment1 2085 (48.48%)	About 65% of the customers chose churn	63.93% of customers were high-rated (a,aa). 6% of the customers were low-rated customers (gy,z).	On average, this group had the second highest income at 4.57.	The mean of this group is only 0.26 times the mean of the population	The mean of this group is only 0.32 times the mean of the population	The mean of this group is only 0.4 times the mean of the population	The average call time last month was the least among the four groups, only about 40% of the overall average.
segment2 495 (11.51%)	55.2% of customers chose to churn.	70.1% of customers were high-rated (a,aa). 3.4% of the customers were low-rated customers (gy,z).	On average, this group had the highest income at 4.99.	The mean of this group is only 0.34 times the mean of the population	The mean of this group is only 0.33 times the mean of the population	The mean of this group is only 0.58 times the mean of the population	The average call time last month was the second lowest of the four groups, only 53% of the overall average.
segment3 459 (10.67%)	All the members of this group chose to leave, that is, the churn is 100%	45.53% of customers were high-rated (a,aa). 9.37% of the customers were low-rated customers (gy,z).	On average, this group had the second lowest income at 4.08.	The mean of this group is only 0.61 times the mean of the population	The mean of this group is 1.93 times the mean of the population	This is the group with the highest average, which is 2.34 times the overall average.	The average value of this segment is 1.83 times of the whole mean.
Segment4 1262 (29.34%)	With 73.22% choosing stay. It's the only segment in cluster2 that has more people staying than leaving	46.12% of customers were high-rated (a,aa). 7.6% of the customers were low-rated customers (gy,z).	On average, this group had the lowest income at 3.85.	This is the only segment in cluster2 with a mean above the population mean, 2.6 times the population mean.	The mean of this group is 2.04 times the mean of the population	This segment mean is 1.66 times the overall average.	The average call time last month was the highest among the four groups, 1.85 times the overall average.

	Overage	peakOffPeakLast	Revenue	Roam
segment1 2085 (48.48%)	the lowest mean of this figure, less than 1/3 of the overall average.	The mean of this group is 0.6 times the population mean. It is the lowest one within four segments of cluster2.	The average revenue contributed by this group last month is much lower than the overall average, only about 60% of the overall average.	Having the lowest mean within the four segments, which is about half of the population mean.
segment2 495 (11.51%)	the second lowest mean, almost 1/2 of the overall average.	The mean of this group is 3.5 times the population mean. It is the top one in this cluster.	The mean of the segments is approximately the same as the population mean	Having the second lowest mean, which is about 77% of the population mean.
segment3 459 (10.67%)	the highest mean, more than twice the overall average.	The mean is 78% of the population mean	The mean of this group was about 1.5 times the mean of the population	Having the highest mean within the four segments, 80 percent above the population mean.
Segment4 1262 (29.34%)	Mean of this segment is about twice the overall average	The mean is 77% of the population mean	The mean of this group was about 1.5 times the mean of the population.	The segment mean is more than 50 percent higher than the population mean.

Appendix 9 Strategy Recommendation


CustomerThermometer
OVER 100
Customer Service, Satisfaction and Experience Statistics
Customer Service,Satisfaction and Experience
Here are some interesting customer service, satisfaction and experience stats which demonstrate how times are changing.
1 54%
54% of customers have higher expectations for customer service today compared to one year ago.
2 This percentage jumps to 66% for consumers aged from 18 to 34 years old.
3 Gartner predicts that 89% of businesses are expected to compete mainly on customer experience.
4 50%
Gartner have also predicted that by 2019 more than 50% of organizations will redirect their investments to customer experience innovations.
5 Worldwide, 67% of people believe that customer service as a whole is improving.
6 64% In 2017, 64% of Americans contacted some form of customer service.
7 52% 52% of people around the globe believe that companies need to take action on feedback provided by their customers.
0 × 880


TELECOMMUNICATIONS
Are you paying attention to your customers?
Your customers are speaking up—are you paying attention? IBM's most recent global telecommunications consumer survey can be summed up in two words: social disruption. Here are a few primary customer concerns from the 2014 IBM Global Telecommunications Consumer Survey.
I need better service from you
Only 1 out of 6 customers are real advocates for their provider
The telecommunications industry ranked near the bottom in multiple customer satisfaction surveys
I am vocal about a good customer experience—or a bad one
53% of subscribers will always or often tell others about a good experience
40% of subscribers will always or often speak up about a bad one
I don't often contact customer care when I have issues
49% of respondents don't call up their service provider's call center because they have to wait too long in the queue to speak to a call center agent
I rely heavily on my peers to choose communications products and services
60% of subscribers rely on recommendations from friends or family to choose communications products and services
I need you to be more proactive about improving the customer experience
68% of respondents rate their service provider "Average to Poor" about the following aspect: "Proactively tries to improve the user experience"
Listen to me and collect the right information to meet my communication needs
Only 12% of respondents strongly agreed that their provider listens to them and collects the right amount of information to meet their communication needs
(Additional source: 2013 IBM Global C-Suite Survey)
Learn more about IBM's Industry Analytics Solutions:
ibm.co/industrylaunch
IBM.

8 Aspects of a Customer Retention Strategy That Can't be Overlooked

Get these fundamentals right and you'll keep your customers loyal straight through to 2017 and beyond!

1. Define what you mean by churn

Before you begin working on proactive retention practices, it is vital to have a clear objective definition of what churn means for your company. Will you be focusing on a specific customer segment, or will your efforts be directed on those who are nearing the end of their contact?

2. Segment your customer base to build a more accurate profile of your users

Customer segmentation isn't new, but it is an evolving and ongoing pursuit. Accurate segmentation can ensure you're using your retention budget in the right way and it can aid you in developing products and services that can drive brand loyalty for specific clients.

Moreover, segmentation can in turn feed into the offers you design for your customers and the channels you choose to deliver them on.

3. Innovate and shout about your differentiating qualities

According to Ernst & Young, failure to adopt new routes to innovation was listed as one of the top ten risks in telecommunications in the last year.

Considering the lack of trust in some telecoms providers, or the idea that networks bundle all their customers together – this is a call to arms to be truly innovative in the way you approach customer management. Being innovative with your loyalty schemes is your route to standing out.

Demonstrating that you're a thought leader who has just what a customer expects at just the right time is vital. As is

Accurate segmentation can ensure you're using your retention budget in the right way and it can aid you in developing products and services that can drive brand loyalty for specific clients.

4. Focus on what your customers really care about

Taking customers for granted drives churn, that's a fact. By visibly addressing the aspects of your product or service offering that really matter to your customers you'll keep them happy and therefore loyal for longer.

If you don't know what those things are, there's no better time than now to find out and feed this into your loyalty and retention programmes for the year ahead.

5. Re-vamp your loyalty programmes

Always look for new ways to up-sell and cross-sell services that break traditional models. The crucial thing – which many operators get wrong – is not knowing what to offer their customers in order for them to stay engaged. The use of genuine customer insight in this area is vital as once you know what your customers want you'll be able to tailor your offerings to match their needs.

6. Ensure proper front-line training and materials

We can all think back to an occasion where we have had a terrible experience talking to someone at a call centre, the frustration that builds up and the negativity you feel towards the brand.

Without the correct training or a poor script, it's impossible for front-line representatives to handle customer calls in the right way. Overlooking this simple step will undermine all the hard work put into strategic planning, so invest in the right training tools and technologies.

approach and instead offer products and services that address the needs of different segments," notes Michelle Nowak, vice president of product management at CSG International.

2. Invest in first-contact resolution: Customers want their problems addressed and rectified immediately. However, as Burton notes, many telecom providers rate poorly when it comes to first-contact resolution, reflecting negatively on their customer satisfaction scores. There are a number of steps that companies can take to resolve issues immediately, including investing in training for contact center agents and providing agents with access to the information they need to give customers the answers they need. Ryan Pellet, senior vice president for strategic consulting services at Nexidia, also recommends using insights from contact center calls to determine what is making customers unhappy and making them want to churn. Information from individual calls, he notes, needs to be put in context of the relationship between a client and the company to better understand whether this is the first time that particular customer had a problem or whether this is a recurring complaint.

3. Focus on an omnichannel experience: With customers communicating with brands over multiple channels, organizations need to make sure that they can connect these different touchpoints and provide continuous experience across the board. While brands are trying to provide new ways for their customers to get in touch with them, the different channels are often siloed. Burton uses the example of a customer who starts an inquiry with a company by engaging with chat support when he realizes that the issue is complicated and he would rather get on the phone with an agent. "But many times there is no direct way to connect to the agent, and when the customer finally navigates the provider's IVR maze, he's often connected with someone new who doesn't have visibility into the chat conversation." In order to improve the experience, telecom companies need to consider investing in technologies like unified communications platforms and the cloud, as well as unifying the organization internally to connect the disparate channels and provide seamless transitions for customers.

4. Provide proactive warnings and information: In today's connected world, customers expect to be able to communicate anytime and anywhere. Loss of a wireless connection or Internet access is, for many people, an experience they will go to extremes to avoid. Nowak says one service that telecom companies should invest in is warnings and alerts when customers are about to reach their data limits. This, she notes, is especially important for those on shared plans who might not know how much data others in the group have used. "Use real-time data to notify customers about what they're buying and what they're consuming," Nowak recommends. Ulla Koivukoski, senior vice president for Comptel's analytics business unit, agrees. "We need to proactively reach out to a customer before he calls us," she notes.

Organizations should go a step further and use such opportunities to inform customers of other plans that might fit their needs better. "Understand their needs, quantify them, and offer them the right services." Koivukoski uses the example of Bangladesh-based pre-paid mobile provider Robi Axiata, which is lever-

References

Berkhin, P., 2006. A survey of clustering data mining techniques. In *Grouping multidimensional data* (pp. 25-71). Springer, Berlin, Heidelberg.

Berry, M.J. and Linoff, G.S., 2004. *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.

Grönroos, C., 2000. Service management and marketing: A customer relationship management approach. 2nd Edn., John Wiley and Sons Ltd.

Kelleher, J., Namee, B., D'Arcy, A., 2005. Fundamentals of Machine Learning for predictive Data Analytics.London, England:MIT Press.

Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M. and Abbasi, U., 2014. Improved churn prediction in the telecommunication industry using data mining techniques. *Applied Soft Computing*, *24*, pp.994-1012.

Koi-Akrofi, G.Y., Koi-Akrofi, J. and Welbeck, J.N., 2013. Relationship marketing tactics and customer loyalty-a case of the mobile telecommunication industry in Ghana. *Asian Journal of Business Management*, *5*(1), pp.77-92.

Maillé, P., Naldi, M. and Tuffin, B., 2008, September. Competition for migrating customers: a game-theoretic analysis in a regulated regime. In *IEEE GLOBECOM 2008-2008 IEEE Global Telecommunications Conference* (pp. 1-5). IEEE.

Rygielski, C., Wang, J.C. and Yen, D.C., 2002. Data mining techniques for customer relationship management. *Technology in society*, *24*(4), pp.483-502.