

**BC2406 Analytics I: Visual & Predict Technology**

**Semester 1 AY 2019/20**

**Project Report**

**Seminar Group: {Redacted}**

**Team Number: {Redacted}**

**Elliot {Redacted}**

**Kathleen {Redacted}**

**Shao Yu {Redacted}**

**Ayush {Redacted}**

**Wanxuan {Redacted}**

**Supervising Tutor: {Redacted}**

**Submission Date: 1 November 2020**

**Table of Contents**

[1. Executive Summary 3](#_Toc55162145)

[2. Business Understanding 4](#_Toc55162146)

[2.1 Background 4](#_Toc55162147)

[2.2 Business Problem 4](#_Toc55162148)

[2.3 Performance Measure 5](#_Toc55162149)

[3. Analytics Approach 5](#_Toc55162150)

[3.1 Analytics Problem 5](#_Toc55162151)

[3.2 Analytics Solution 5](#_Toc55162152)

[3.3 Introducing our dataset 5](#_Toc55162153)

[4.1 Description of Data Characteristics 6](#_Toc55162154)

[4.2 Understanding the Data 6](#_Toc55162155)

[4.3 Data Cleaning 7](#_Toc55162156)

[5. Modelling & Evaluation 9](#_Toc55162157)

[5.1 Identifying significant factors influencing customer satisfaction 9](#_Toc55162158)

[**5.1.1 Analysing Customer Satisfaction Variables with Logistic Regression Analysis** 9](#_Toc55162159)

[**5.1.2 Analysing effect of Demographic Factors on customer satisfaction with CART** 12](#_Toc55162160)

[5.2 Predicting Customer Churn 14](#_Toc55162161)

[**5.2.1 Logistic Regression Analysis** 14](#_Toc55162162)

[**5.2.2 Classification and Regression Tree (CART)** 14](#_Toc55162163)

[6. Proposed Business Recommendations 16](#_Toc55162164)

[7. Further Research 20](#_Toc55162165)

[8. Bibliography 22](#_Toc55162166)

[9. Appendices 28](#_Toc55162167)

[Appendix 0 - CorrPlot for customer churn dataset 28](#_Toc55162168)

[Appendix 1.1: Data Dictionary for Dataset 1 28](#_Toc55162169)

[Appendix 1.2: Data Dictionary for Dataset 2 30](#_Toc55162170)

[Appendix 2.1: Logistic Regression for By-phone service satisfaction 31](#_Toc55162171)

[Appendix 2.2: Logistic Regression for By-phone outcome satisfaction 32](#_Toc55162172)

[Appendix 2.3: Logistic Regression for Writing service satisfaction 32](#_Toc55162173)

[Appendix 2.4: Logistic Regression for Writing outcome satisfaction 33](#_Toc55162174)

[Appendix 2.5: Logistic Regression for In-Person service satisfaction 33](#_Toc55162175)

[Appendix 2.6: Logistic Regression for In-Person outcome satisfaction 34](#_Toc55162176)

[Appendix 2.7: Logistic Regression for Email service satisfaction 34](#_Toc55162177)

[Appendix 2.8: Logistic Regression for Email outcome satisfaction 35](#_Toc55162178)

[Appendix 2.9: Odds Ratio Confidence Interval for By-phone service satisfaction 35](#_Toc55162179)

[Appendix 2.10: Odds Ratio Confidence Interval for By-phone outcome satisfaction 35](#_Toc55162180)

[Appendix 2.11: Odds Ratio Confidence Interval for Writing service satisfaction 36](#_Toc55162181)

[Appendix 2.12: Odds Ratio Confidence Interval for Writing outcome satisfaction 36](#_Toc55162182)

[Appendix 2.13: Odds Ratio Confidence Interval for In-Person service satisfaction 36](#_Toc55162183)

[Appendix 2.14: Odds Ratio Confidence Interval for In-Person outcome satisfaction 37](#_Toc55162184)

[Appendix 2.15: Odds Ratio Confidence Interval for Email outcome satisfaction 37](#_Toc55162185)

[Appendix 2.16: Odds Ratio Confidence Interval for Email service satisfaction 37](#_Toc55162186)

[Appendix 2.17: Maximal Tree root node error and CP values 38](#_Toc55162187)

[Appendix 2.18 Pruned Tree CP and Variable Importance 38](#_Toc55162188)

[Appendix 3.1 - Statistical Significance and Odds Ratio of Data Set 2 39](#_Toc55162189)

[Appendix 3.2 - Confusion Matrix 40](#_Toc55162190)

[Appendix 3.3 - Maximum Tree 40](#_Toc55162191)

[Appendix 3.4 - Pruning Sequence of the Max Tree 41](#_Toc55162192)

[Appendix 3.5 - Final Pruned Tree 42](#_Toc55162193)

[Appendix 3.6 - printcp of the Pruned Tree 43](#_Toc55162194)

[Appendix 3.7 - Plotcp of the pruned tree 43](#_Toc55162195)

[Appendix 3.8: Implications of the confusion matrix statistics for our customer churn CART model 44](#_Toc55162196)

[Appendix 3.9 Conditional probability of customer churning 45](#_Toc55162197)

# 

# 

# 1. Executive Summary

As the asset management landscape gets increasingly competitive, clients now have a wide array of companies to choose from and firms are increasingly struggling to retain their clients. The COVID-19 pandemic has exacerbated the situation as clients start to withdraw from asset management plans amidst an impending market downturn and possible global recession.

This report aims to address the business problem of **what strategies should White Rock adopt to reduce its customer churn.** Framing it into an analytics problem, our focus will be to **identify** significant factors affecting customer satisfaction scores and customer churn, **analyse** the impact of these factors and derive ways to **predict** customers who are likely to churn. To answer the analytics problem, we made use of the Logistic Regression and Classification Tree Model. 2 datasets were used *(Irish Civil Service Customer Satisfaction Survey 2017 and Kaggle Customer Churn)* to illustrate the analytics modeling, and insights were derived based on them.

With regards to customer satisfaction scores, we identified the amount of holding time, the staff’s demeanour, the security/privacy features offered by the firm, and the quality of their external reports to be significant factors that impacted customer satisfaction scores. We also noticed 3 demographic factors that played a crucial role in profiling customers who were likely to churn; these included the tech-savvy customers, the ones holding a bachelor’s degree or higher, and the ones who communicated with their manager only on a bimonthly/quarterly basis. Lastly, from the customer churn dataset, we realised that clients who use more than 2 of the firm’s services are left dissatisfied, and are almost certain of churning. A “non-active” customer was also at a much higher risk of churning than an active one, thus emphasizing the importance of having active communication efforts. A lower percentage (11.7%) is seen to have churned amongst the younger customers (Age <43), compared to 41.6% for the complement group.

With these findings, several business recommendations are proposed --- (1) the phasing out of auxiliary service offerings to focus efforts on specialising in two services; (2) the development of digital capabilities to retain their tech-savvy clientele: done via the use of an omni-channel strategy, by investing in a Natural Language Processing chatbot, or through the introducing of a digital payment portal; (3) increasing communication frequency to strengthen client relationships and improve overall client experience.

The report ends with suggestions for further research and the limitations of our dataset that have inhibited our analysis.

# 

# 

# 

# 

# 

# 2. Business Understanding

## 2.1 Background

The pressure on asset managers is intensifying as the wealth management landscape gets increasingly competitive. Companies like White Rock face rising costs from having to comply with regulations, achieve greater transparency and comparability and develop their digital capabilities (PricewaterhouseCoopers, 2020). They also have to deal with new and tougher competitors such as internet giants, Amazon and Google, who are looking to partner with J.P. Morgan (Fonseca & Mukerjee, 2018).

## 2.2 Business Problem

In an industry where competition is tough and regulatory requirements are increasing, clients can easily select another firm if they are not satisfied. While gaining new clients seems tough enough, retaining them is even more difficult (Fonseca & Mukerjee, 2018). One-third of clients have switched wealth management providers or moved assets in the past three years and another third plan to do so in the next three years (Birkin et al., 2019). Not only does customer churn cost the firm dearly, it also has the potential to disrupt market trends significantly among institutional investors as well (Saketh & Sridhar, 2018).

Furthermore, in a time of a major, global black-swan event that has led to an economic downturn, such as the COVID-19 pandemic, the effect on customer churn rates has been significant. In fact, its impact is seen to mirror that of the 2008 financial crisis. Private Banker International’s surveying during and after the Global Financial Crisis of 2008 highlighted the negative effect it had on retention rates, as clients (at least in part) blamed their advisers for lost portfolio value (Private Banker International, 2020). Loyalty ranking fell from 70 before the crisis to 46.7 in 2009. 6 in 10 wealth managers also believe that volatile market conditions have negatively affected client retention rates.

This report will thus focus on addressing the business problem of what strategies White Rock should adopt in order to reduce customer churn

## 2.3 Performance Measure

The performance metric will be the customer churn rate across a period of time. A decreasing rate will suggest that customers are overall satisfied with the firm and willing to partner with it long-term. An increasing rate will suggest otherwise.

# 

# 

# 

# 

# 

# 3. Analytics Approach

## 3.1 Analytics Problem

This report will focus on:

1. Identifying significant factors affecting
   1. Customer satisfaction scores
   2. Customer churn
2. Analysing the impact of these significant factors
3. Predicting active customers who are likely to churn

Understanding the reasons behind customers feeling dissatisfied and customers leaving will reveal deficiencies in their operations. Additionally, being able to predict customers who are likely to churn allows asset managers to intervene early and implement relevant measures to retain clients.

## 3.2 Analytics Solution

This report aims to analyse the degree of impact different variables have on customer satisfaction scores and customer churn rates through:

1. Logistic Regression Model
2. Classification Tree Model

## 3.3 Introducing our dataset

Our first dataset is retrieved from the Irish Civil Service Customer Satisfaction Survey 2017, published by Ireland’s Department of Public Reform and Expenditure (DPER). The purpose of the survey was to ascertain satisfaction levels with services received, as well as more general perceptions of, and attitudes to, the Civil Service. Additionally, the survey was published in order to allow the government to understand possible reasons for dissatisfaction with the Irish Civil Service. We had selected this dataset for our analysis as it provides individual customers’ scoring for various elements throughout the service process. These customer service variables are applicable to most businesses and especially that of the asset management industry, which shares many similarities with the civil service. Analysing this dataset provides us with insights into which factors customers generally prioritise and have the greatest impact on overall customer satisfaction.

The second dataset we have procured is one relating to Customer Churn. While the dataset was initially meant to be from a bank’s perspective, we have modified our interpretation of the variables to suit the present context (refer to [Appendix 1.2](#A1p2)). The variables or instances themselves have not been touched. Global banks have begun prioritising asset management divisions in recent years given their growing contributions to group profits and revenues (Marriage, 2015). In 2020, Goldman Sachs’s revenue climbed 30% to 10.78 billion, driven by the trading and asset management divisions (Son, 2020). With banks increasingly providing asset management services, similarities can be drawn between the clients of White Rock and those of the banks. Insights gathered from this dataset will thus be transferable and applicable to White Rock as well.

**4. Data Preparation**

## 4.1 Description of Data Characteristics

Dataset 1 - Irish Civil Service Customer Satisfaction Survey 2017

|  |  |
| --- | --- |
| Data Source | Department of Public Expenditure and Reform |
| Collection Period | 5 April 2017 to 1 April 2019 |
| Sample Size | 2,027 rows, 44 columns |

Dataset 2 - Customer Churn

|  |  |
| --- | --- |
| Data Source | Kaggle |
| Collection Period | Not Available |
| Sample Size | 10,000 rows, 14 columns |

## 4.2 Understanding the Data

Dataset 1 - Irish Civil Service Customer Satisfaction Survey 2017

First off, to understand the data, we investigate the significance and meaning behind the variables. (Refer to [Appendix 1.1](#A1p1) for explanation of variables)

Our group was able to retrieve the full csv dataset from the Irish government’s website; however, the link for the accompanying list of interview questions was invalid, making it difficult to decipher the dataset which contained question numbers as the headers. Fortunately, we were able to rely on the Report on Findings which pointed us towards identifying key columns in the dataset. Following which, we manually transposed the questions into the headers. Out of the 233 columns in the original dataset, 36 columns were retained with transposed headers, across 2,027 rows.

There are multiple methods of communication through which customers can interact with government departments and offices, with phones being the most popular communication method (73%), followed by in person (51%), writing (36%) and email (32%). As customers may use one or more methods for contact, for each entry, columns pertaining to unutilised methods were left blank, resulting in a high number of null values in the dataset.

The dataset contains two dependent variables, overall satisfaction with service and overall satisfaction with outcome. Excellent service during contact with the customer may not result in a satisfactory outcome; likewise, customers may be satisfied with the outcome of their interaction but dissatisfied with the quality of service throughout the interaction. Since satisfaction levels of both the process and outcome can be key to customers’ decisions to churn, we decided to keep both dependent variables and carry out our analysis separately, before coming to a conclusion on whether both types of customer satisfaction were different enough.

All dependent variables in the dataset are categorical variables while all independent variables are continuous variables.

Dataset 2 - Customer Churn

Our customer churn dataset contained 10,000 records, spread across 14 different columns. We went over the dataset to find the nature and the distribution of each of the variables.

The dataset consisted of a lot more instances of customers not churning (7,963) than customers who churned (2,037). Given this proportion, we realised that if one was to randomly pick a row and “predict” that the particular customer did not churn, he would have a 79.6% chance of being correct in his guess, with no other considerations made. Hence, we settled on the target accuracy of our model to exceed that, and ideally better it by at least 5 percentage points (84.6%).

Secondly, the *Balance* column had a lot of ‘0’ values, which mandated cleaning. Lastly, the *Age* distribution heavily sided towards the younger end of the spectrum. With a median age of 37, and standard deviation and IQR of 10.4 and 12 years respectively, our insights will thus be most applicable to White Rock’s younger clientele.

No multicollinearity issues arose from correlated variables as the maximum magnitude of correlation being found was merely -0.26 between *Age* and *NumofProducts* ([Appendix 0](#A0)).

## 

## 4.3 Data Cleaning

|  |  |  |
| --- | --- | --- |
| **First dataset:** From the Irish Civil Service Customer Satisfaction Survey 2017 | | |
| **Variable /Column** | **Action** | **Rationale** |
| *service\_satisfaction*  *outcome\_satisfaction* | Removed values *>5* | 2 dependent variables (*service\_satisfaction and outcome\_satisfaction)* exist in our dataset.  There was a single value of 6 and a value of 7 in the dependent variable *service\_satisfaction*. There were also 11 sixes and sevens for the dependent variable *outcome\_satisfaction*, all of which represented results categorised as ‘Don’t know / No answer’.  For more accurate analysis, we removed these anomalies to restrict the range of values between 1 to 5. |
| Communication Mediums | Removed the medium  “Online via a PC/Laptop” and “Online via a mobile device/tablet” | These 2 mediums involved overlapping methods of communication. They were removed due to their small sample size, with a usage of only 25% and 6% respectively. The report on findings also did not reveal any further investigation done on the satisfaction levels for both mediums, thus there were no variables available for analysis |
| *phone\_*  *writing\_*  *inperson\_*  *email\_* | Segregated the dataset by the four main communication mediums | We grouped the variables in the dataset according to 4 types of communication medium (*phone, writing, in-person, email)*. Under each method of communication, the values of the categorical variables range from 1 to 7. Segregating the dataset also ensures that null values for unanswered columns would not affect the integrity of the dataset after trimming. |

|  |  |  |
| --- | --- | --- |
| **Second dataset:** Kaggle Customer Churn | | |
| **Variable /Column** | **Action** | **Rationale** |
| *RowNumber* | Variable was removed | We removed this row, as it served no purpose. The in-build indexing in R helped us with the index of each row |
| *CustomerId* | Variable was removed | While CustomerId serves an important role in database management systems, it served no purpose other than that of a unique identifier here, which wasn’t required for our purpose. |
| *Surname* | Variable was removed | Surname served no real purpose for our modelling requirements. |
| *Balance* | Replaced 0 values with the mean balance of non-zero entries of the dataset | There were a very large number of 0 values, to the point that the 1st quartile value was also 0. Given our interpretation of ‘Balance’ being the investment balance, it made no sense to have a nil amount invested. |

# 

# 5. Modelling & Evaluation

## 5.1 Identifying significant factors influencing customer satisfaction

This section details the methods we used to analyse the relationship between various variables and customer satisfaction scores.

### **5.1.1 Analysing Customer Satisfaction Variables with Logistic Regression Analysis**

The aim of the logistic regression analysis is to determine which variables have the greatest significance on both outcome satisfaction and service satisfaction rates, based on the 4 categories in the communication medium - *phone, email, in-person* and *writing*.

Two new columns, b\_service\_satisfaction and b\_outcome\_satisfaction were generated whereby the service\_satisfaction and outcome\_satisfaction were discretized into binary categories with scores of 1 to 3 as Not Satisfied and 4 to 5 as Satisfied. The minimum threshold was elected to be a score of 4 rather than 3 as we believe that a neutral rating is unlikely to prevent customer churn especially in a competitive sector such as asset management.

Based on the logistics regression model, the variables with p-value < 0.05 **(**[**Appendix 2.1**](#A2p1) **-** [**2.8**](#A2p8)**)** would be recognised as having a statistically significant impact on customer satisfaction rate. We also checked to ensure that Odds Ratio Confidence Intervals **(**[**Appendix 2.9**](#A2p9) **-** [**2.16**](#A2p16)**)** of significant variables do not include 1.

The following variables were identified as having a significant impact on both service satisfaction and outcome satisfaction: *phone*\_*easeoffindingphonenumber, phone\_manner, phone\_holdingtime, writing\_easeoffindingaddress, inperson\_knowledge, inperson\_system, inperson\_privacy, inperson\_formdesign.*

The low p-values of the variables identified indicate that these variables have a statistically significant impact on customer satisfaction rate, and therefore, a change in these variables would correspond to an increase or decrease in the customer satisfaction rate.

The first observation from looking at the p-values of all the variables is that no variables under the email category has a significant impact on both service satisfaction and outcome satisfaction. One possible reason is that due to the comparatively lower adoption rate of using email as a method of communication, the regression model was unable to identify significant variables from the smaller sample size.

Another observation was that the significant variables for predicting service satisfaction as well as outcome satisfaction were largely similar, thus we can conclude that targeting the identified significant variables will allow us to raise satisfaction levels for both service rendered and outcome of the service. This was further confirmed by the high correlation between service satisfaction and outcome satisfaction (0.858).

Significant Variables & Their Relationship With Overall Customer Satisfaction

As we are concerned with helping WhiteRock with lowering customer churn, the likelihood to churn can be most easily reduced by raising customer satisfaction in both service processes and outcome of the service. Different communication mediums have different elements through which a customer’s satisfaction may be affected. From the identified significant variables, we selected the following variables to be discussed in more detail: *phone\_holdingtime, phone\_manner, inperson\_privacy, inperson\_knowledge, writing\_quality.* These variables were chosen for their relevance to an asset management firm, as well as their positive correlation to overall customer satisfaction scores.

Implications and Recommendations

**By phone: Amount of time left holding - *phone\_holdingtime***

The satisfaction score given to the category of holding time had a positive correlation of 0.49186 with service satisfaction and 0.65271 with outcome satisfaction, with a higher satisfaction score assumed to be a result of shorter holding times.. Holding time has long been studied as a significant factor on customer satisfaction. According to a Forrester survey, 73% of customers listed valuing their time as the most important aspect of good customer service (Morgan, 2019). In another survey from Velaro, 60% of customers feel that a single minute on hold during a call was too long a wait, with 32.3% also having zero tolerance for any holding time during a call (Cseley, 2020).

White Rock can firstly adopt methods to optimise service department processes in order to reduce hold time. A possible solution would be expanding the knowledge base of service representatives and improving their information accessibility so as to allow them to bypass the higher chain of command. Given the exclusivity and limited size of an asset management firm’s customer base, having dedicated service agents or hotlines for selected customers may also be a feasible method of increasing capacity and increasing customer satisfaction. However, such solutions may require sacrificing privacy for the sake of convenience, as service agents get hold of clients’ details.

Additionally, WhiteRock can find ways to reduce the perceived length of holding time, by devising an optimal holding programme or adopting Music on hold (MOH). When on hold in a phone conversation, the use of music (subject to an ideal tempo and volume) or delivery of information while holding can improve waiting time tolerance (Guégen & Jacob, 2002).

**By phone: Manner in which staff provided information - *phone\_manner***

This variable is significant and has a high positive correlation of 0.67109 with service satisfaction and 0.70319 with outcome satisfaction in the logistic regression model, thus implying that greater satisfaction with the manner through which staff communicated with customers will likely result in a higher customer satisfaction rate.

Characteristics of satisfactory manners could be the projection of an amicable or professional image. Therefore, WhiteRock can have an employee training program to train asset managers’ interpersonal skills. Strong client relationships depend on strong communication. It is important that asset managers are able to resolve any issues that clients are facing with clarity and competence. This can greatly enhance client experience, and hence satisfaction levels.

**In person: Privacy - *inperson\_privacy***

This variable has a positive coefficient of 0.44993 with overall outcome satisfaction, indicating that the variable moderately influences customer satisfaction. Our hypothesis is that stringent privacy measures put in place will allow customers to feel assured that their private information is secure throughout the transaction process, thus increasing overall satisfaction of the firm’s services.

In recent years, customers’ rights to privacy have been thrown into the spotlight with stronger legislation and regulations being put in place, such as Europe’s General Data Protection Regulation (GDPR) and Singapore’s Personal Data Protection Act (PDPA). Customers may be more vigilant of the absence of sound data security and privacy practices set in place, as well as their exposure to significant vulnerabilities. Hence, it is critical to protect the confidentiality, integrity and availability of customer data in order to keep customers satisfied with the firm they have engaged to manage their assets.

To reduce the risk of cyber attacks and data leaks, WhiteRock can consider having a dedicated cyber-security department that is reliable, knowledgeable and able to combat cyber threats that they could have encountered. One data security strategy that the firm can adopt is having system logs keep track of instances of access to data, with access records to the client database being reviewed regularly by the management. Moreover, to prevent loss of customer data or unavailability of customer data due to malicious attacks on the database, the firm can make use of state of the art antivirus software and firewalls that are regularly updated and conduct regular backups of customer data. At the firm’s offices, staff must also ensure that collection of personal information is carried out with clear consent obtained from the customers, with regulatory procedures followed stringently.

**Writing: Quality of information received - *writing\_quality***

We hypothesize that quality of information (regarding asset management) received by the client will have a significant impact on customer churn rate, as there is a positive correlation of 0.429 with overall outcome satisfaction. When deciding on an asset management company, the client would consider the financial products being proposed by the company. An effective asset management plan means that the firm is able to develop, operate, maintain, upgrade and dispose of asset cost effectively. Additionally, asset managers must use reasonable judgment to perform analysis and make decisions that are advantageous to the clients.

According to Clearwater Analytics benchmarking survey (Clearwater Analytics, 2015), 87% of survey respondents ranked reporting quality as an ‘Important’ to ‘Extremely Important’ criteria when evaluating asset managers. As such, in order to meet clients’ expectations, asset managers should opt to devise high-quality reports that would also create high-performing returns which appeal to clients.

### **5.1.2 Analysing effect of Demographic Factors on customer satisfaction with CART**

Next, we look at the effect of demographic factors on customer satisfaction. By analysing the profile of clients, we may identify commonalities between their profile and the satisfaction scores they give. We can then preempt customers’ satisfaction scores, and gain better visibility into customer sentiments. Basing off their predicted scores, White Rock will be able to identify dissatisfied clients and take proactive steps such as routing dissatisfied clients to more experienced staff, prioritising their needs etc.

Variables we looked at include: *age, gender, monthly\_income, highest\_education\_level, occupation, marital\_status, number\_of\_dependents, tech\_savviness, communication*

A Classification Tree was used to model the target variable - Customer Satisfaction Score (cust\_sat). Scores took on integer values between 1 and 5. After fully growing the tree (inclusive of all variables, minsplit=2, cp=0), it was pruned at the optimal CP value (refer to [Appendix 2.17](#A2p17)). The top 3 significant variables identified based on variable importance (in order of *decreasing* importance) were: *tech\_savviness, communication, highest\_education\_level* (refer to [Appendix 2.18](#A2p18))

Significant Variables & Their relationship With Customer Satisfaction Scores

**1. Digital Savviness of clients - *tech\_savviness***

The more tech-savvy clients are (*tech\_savviness scores of 4 and 5)*, the more dissatisfied they appear to be. As people become more adept with technology, they are increasingly expecting for their asset managers to be technologically proficient as well. Asset management firms thus need to display technological sophistication in order to keep customers satisfied.

With major future clients becoming increasingly tech-savvy, firms that succeed in satisfying customers will be those that are able to effectively integrate digital solutions in a client-centric model (Additiv, 2020). White Rock will thus need to develop and enhance their digital capabilities.

**2. Frequency of communication - *communication***

Clients that communicate only bimonthly or quarterly with their asset managers tend to give lower satisfaction scores than those that communicate at least monthly with them. This is in line with a study by YCharts which found the current frequency of advisors’ interactions with clients unsatisfactory and insufficient. 66% of clients also responded that more frequent communication would increase their confidence in their advisor (YCharts, 2019). Regular communication is thus vital to build trust, strengthen relationships and retain clients.

In times of market volatility, such as the COVID-19 pandemic, improving and increasing client engagement is even more vital in boosting client satisfaction and loyalty. White Rock needs to adopt a communication strategy that aids advisors in offering the type of immediacy clients are increasingly demanding (Deyo, 2017). Keeping communication personalised and proactive, which are characteristics highly valued by clients (YCharts, 2019) will be advantageous in the long run.

**3. Highest Education Level of clients - *highest\_education\_level***

Clients that are more educated (*bachelor’s, master’s and PhDs)* appear to be more dissatisfied. It is likely that being more knowledgeable, they grow to have higher expectations of their asset managers. More informed clients also have better understanding of the nuances of pricing and are more aware of cheaper alternatives, such as FinTech (Birkin et al., 2019).

Hence, there is growing dissatisfaction with fees charged based on assets under management (AUM). Satisfaction is seen to be the lowest among the youngest and more knowledgeable clients who believe that they have been charged unfairly (Birkin et al., 2019). Firms thus need to prove that their services are worth the fees charged. White Rock needs to increase the transparency and predictability of their fee structures, and communicate effectively with clients on the value of their services.

## 5.2 Predicting Customer Churn

This section details the methods we used to analyse the relationship between various variables and customer churn rates.

### **5.2.1 Logistic Regression Analysis**

As part of our preliminary analysis of the customer churn dataset, we used a Logistic Regression Model to analyse the various factors to see which are the main determinants of whether a customer churns or not.

Logistic Regression was done by using the “Exited” factor as a categorical Y, with “1” for having churned and “0” for not churning. We used a train/test split of 70/30 ratio in order to evaluate the accuracy of the model, as well as the statistical significance of the factors.

Running a summary of the model gives us a greater understanding of which factors were significant. Based on p-value and confidence intervals , the following variables with low p-value (<0.05) and odds ratio confidence intervals not containing 1 were identified to have a significant impact on whether a customer churns: Geography, Gender, Age, Number of Products and whether or not they are an active member ([Appendix 3.1](#A3p1)).

A confusion matrix for the logistic model ([Appendix 3.2](#A3p2)) is developed to further analyse our

model. At first glance, the accuracy of 80.57% seems to be a satisfactory model. The model also has a fairly high sensitivity of 0.8286, which implies that the model can fairly accurately determine that a customer will not churn. However, a lower sensitivity of 0.5556 implies that the model is unable to accurately determine that a customer will churn. A Kappa value of 0.2332 further suggests that while the logistic model certainly works better than simply guessing by chance, its accuracy still leaves much to be desired.

In order to generate further insights into this matter, we decided to employ the use of a Classification and Regression Tree (CART) as well.

### **5.2.2 Classification and Regression Tree (CART)**

We further employed a Classification Tree model to determine some of the important factors that determine customer churn. In order to gauge the true accuracy of our model, we performed a train/test split at the very start. The chosen ratio was 70/30.

The trainset was used to develop the mode which involved fully growing the tree (setting minsplit = 2 and cp = 0) and then pruning at its optimal CP (refer to [Appendix 3.3](#A3p3) - [3.7](#A3p7)). To test the accuracy of the model, we used it to predict the churn outcomes in our testset. The overall accuracy of the model was 0.857 (down from 0.8658 for the trainset) (refer to [Appendix 3.8](#A3p8)).

Other important metrics to be considered are the sensitivity, specificity and the positive and negative predicted values of the test:

|  |  |
| --- | --- |
| **Metric** | **Score** |
| Sensitivity | 87.04% |
| Specificity | 75.71% |
| Positive Predicted Value | 96.40% |
| Negative Predicted Value | 43.86% |

The detailed interpretations and implications of these metrics are captured in [Appendix 3.8](#A3p8).

One of the likely reasons for the model to have such a skewed accuracy towards being able to accurately predict the non-risk customers is due to the nature of the dataset itself. With 10,000 instances of customer data, only 2,037 of the instances are of customers actually churning. With the sheer volume (>80%) of the customers staying, the model is able to understand the pattern of the data that follows this distribution much better than the latter. It also implies that as the model gathers more data, especially on customers who left White Rock, it should be able to firm up the weaker side of its predicting power as well.

The top 5 most significant variables are as follows:

Age = 45 | NumProd = 31 | IsActiveMember = 16 | Geography = 5 | Balance 2

Significant Variables & Their relationship with Customer Churn

**Age**

In a dataset with clients aged between 18 and 92, and 75% of them being below the age of 44, the only cluster of clients who are aged below 43 and who churned are the ones who availed of 3 or greater services, which comprise merely 2% of the entire population in the dataset.

Inference: Younger customers may be less likely to churn due to the fact that millennials are, on average, more [loyal](https://articles.bplans.com/the-4-keys-to-building-brand-loyalty-with-millennials/#:~:text=Millennials%20are%20the%20most%20brand,authentic%2C%20people%20will%20take%20notice.) to their brands than Gen-X or baby boomers. In fact, a CrowdTwist report called them “the most loyal” brand generation (Oracle CrowdTwist, 2015). Other facts such as this generation already having long-term relationships with brands at a 60% rate further substantiate why White Rock needs to keep them a priority.

**NumOfProducts**

Number of products used is an extremely key determinant of whether or not a customer churns, as all those customers who are not based in Germany have churned when they have availed 3 or greater of the firm’s services.

Inference: If we were to assume White Rock’s services are being broken down to ‘core’ (2) and ‘auxiliary’ (3), they are doing well as far as their core services are concerned. However, their auxiliary services are not up to scratch in terms of the quality or delivery, and are actually causing customers to churn.

**IsActiveMember**

Another extremely important factor in determining White Rock’s customer churn is whether or not the customer was “an active member”. We defined this as requesting frequent communications or check-ins with their asset manager.

IsActiveMember appeared in our CART model once and fell under the cluster of customers aged greater than 43. Within this sub-class (Active and aged >43 - comprising 16% of all customers), only 2% of all customers churn. However, if the client is above 43 and inactive (comprising 13% of all cases), 9% of them are eventually classified as “churning”, constituting an extremely high volume of the sub-class.

Inference: If the client is an active member, they have a much lower churn rate and a much higher retention rate. Furthermore, given our interpretation of what ‘active’ means, it would seem that regular communication between clients and asset managers play a vital role in convincing existing clients to stick around.

# 

# 6. Proposed Business Recommendations

**6.1 White Rock should phase out their auxiliary service offerings**

We know that asset management is no longer restricted to mere “managing investments”. Today, asset managers themselves offer services including financial planning, financial advisory via RIAs (Registered Investment Advisory), wealth and risk management etc (The Balance, 2020). Further, we know from the problem statement that White Rock “is primarily” an asset manager, suggesting it offers services in other domains as well.

Given the business implication mentioned above about a large majority of churned customers being those who used more than 2 products, it may be in White Rock’s best interests to play to its strengths - its core product offerings, and phase out ‘secondary’ services altogether.

From a data standpoint, if we consider the conditional probability of a customer exiting, given that they used 1 or 2 products, it is 18.16%. However, this shoots up astronomically to 85.89% if the number of products used exceeds 2 ([Appendix 3.9](#A3p9)).

As such, we believe that it would be in White Rock’s best interests to keep it simple, and focus on the products which customers perceive to be the most satisfactory. We believe such a ‘specialist’ approach would help them build, but more importantly retain a happy and loyal customer base, as opposed to an approach aimed at offering a basket of mediocre services.

While this may at first seem to deteriorate White Rock’s customer base, and hence their RoI and/or profitability, we believe it is the best approach with the long-term health of the business in mind. White Rock would be able to boost its retention metrics from focusing on their core competencies, which in turn would boost their lifetime value (LTV) metric. If they continue to excel in providing these ‘core’ services, they may consider changing their fee structure to perhaps increasing the % of AUM fee they charge, or increasing the minimum threshold of funds to be managed to ensure profits are not dampened in the long run.

Ideally, this would translate to a higher LTV: CAC ratio via being able to retain existing clients and focusing efforts on enhancing current offerings instead.

**6.2 White Rock should improve digital efficiency and capabilities.**

6.2.1 - Omni-Channel Strategies

In recent years, new digital capabilities have been shown to improve customer retention. For example, omni-channel strategies have also been achieving 91% higher yearly customer retention rates compared to those that don’t (Benbria, 2018).

As customers become increasingly tech savvy, they would naturally grow to expect great technologically integrated customer service, and conversely, would be dissatisfied with old fashioned and inefficient customer service. This would inevitably lead to customer churning.

In order to remain competitive on this front, our group proposes that White Rock opens up its communication channels and to move towards an omnichannel approach. By building digitally enabled interaction channels (i.e client portals) for advisors to engage more broadly with clients and compliment face to face consultations.

Omni-channel approaches appeal to customers by increasing their points of contact with the brand. They increase customer conversion and retention while minimizing their acquisition costs, ultimately retaining almost three times more customers than non-omni-channel businesses (Commbox, 2019)

In addition, the omnichannel approach helps to combine what the customer needs with the same customer experience across all channels. This allows the customer to decide for himself via what channel he engages with the brand, while he can rest assured that the customer experience will live up to his expectations. A key ingredient in boosting customer loyalty.

Furthermore, this integration of processes and communications helps to create greater synergy across channels and departments. For example, a customer’s query over live chat can be easily accessed by other relevant departments, greatly speeding up the request as well as minimizing miscommunications across departments (Stirista, 2020). Over 42% of customers cite receiving inconsistent information from customer service agents as prohibitive to their decisions. Further, 72% of customers consider it poor customer service if they have to explain their problem multiple times (Commbox, 2019). The usage of an omnichannel approach with real-time information sharing can therefore help minimize these issues by improving the customer experience, and therefore reducing churn.

6.2.2 - Invest in an NLP chatbot to better answer customer queries

As part of the omni-channel approach, facilitated by information sharing, we recommend White Rock to invest in a Natural Language Processing (NLP) chatbot, to better answer customer queries, especially ones of a standardised nature, such as requesting updates or scheduling of check-in sessions/meetings.

The chatbot can recognise and answer multiple forms of the same question and give instant, standardised responses to customers. This will be a huge asset for a company like White Rock, whose international operations would likely cut across customers with different proficiencies in different languages, and non-uniform speech/typing patterns, when it comes across sentence structure, grammar and so on.

For the customer, this means faster, and more consistent support with uniform answers, and for the service agents, fewer repetitive and transaction questions. Further, given the fact that the chatbot will not have ‘hours of operation’, it can be used to provide round the clock service at no waiting/holding time. All of these will overcome major obstacles of phone service mentioned in the report previously.

With regards to scheduling meetings, the asset managers can have their calendars synchronised into the chatbot’s algorithm for ease of slotting in clients for their update/check-in sessions. This will ensure that there does not need to be a constant ‘back and forth’ just for the client to meet up with their manager.

Further, given that the chatbot will be a one-time (fixed) investment for the company, as opposed to paying all service representatives on a recurring basis, it may prove to improve profitability in the long run.

6.2.3 - Digital payment portal

One point of tension between clients and asset management firms arises from dissatisfaction surrounding management fee structure and the lack of transparency surrounding such fees. As mentioned in Section 5.2.1, clients nowadays have a wider knowledge of financial products available on the market and are constantly lured by up and coming alternatives in the form of New Financial Products (NFP), such as FinTech. In a YouGov survey commissioned by Netwealth, 72% of investors ranked Fee Transparency as the most important factor when it comes to choosing a wealth manager (Netwealth, 2018). However, only 56% of asset management clients on average have full awareness of the fees paid (Birkin et al., 2019), with fee awareness the lowest for higher-aged and less financially-savvy clients. Hence, by curbing growing dissatisfaction with fees charged for services rendered and strengthening the value-for-money proposition, White Rock stands a chance to prevent loss of customers to competitors with more attractive fees.

Therefore, as part of its digital transformation, our group proposes for White Rock to introduce a digital payment portal to enable payment of management fees, so as to increase the transparency and predictability of their fee structures. A digital payment portal provides features such as on-to-go and round-the-clock fee payment, detailed breakdown of fee payment invoices, straightforward monitoring of investment assets’ performance, as well as timely notifications.

A digital payment portal firstly enables clients to view the breakdown of the annual fees paid, differentiating between the key constituents such as the investment management charges, administration charges, performance fees and advice fees. The improved transparency will allow clients to re-evaluate whether the price-to-entry to an exclusive financial service is noteworthy, while enabling the firm to build up trust with their clients.

The same digital payment portal can also disclose the evolution of clients’ investment assets over time with the help of visualisations, in order to demonstrate the efficacy of asset management plans in helping clients achieve their financial goals. The closer gross returns are to the intended target, the easier it is for White Rock to justify fees charged, especially for charges linked to performance and returns.

**6.3 Increasing Communication Frequency**

As previously mentioned, frequent communication with their asset managers strengthens clients’ relationship with the firm and thus increases the likelihood of firms retaining them.

White Rock should commit to a regular cadence of communication with their clients since clients want frequent, proactive communications. In a YCharts survey, when asked if it is important for advisors to reach out proactively, 35 percent responded with "Strongly Agree," and 40 percent stated "Somewhat Agree" (YCharts, 2019). A four-pronged communication strategy that White Rock can adopt would be having:

1. Scheduled broadcast check-ins
2. Scheduled personalized check-ins
3. Ad Hoc broadcast updates
4. Ad Hoc personalized updates

Scheduled check-ins will be conducted on a monthly basis, which involves:

1. Broadcasting information such as changes in economic and market conditions, changes in firm strategies, economic outlook and commentaries
2. Personalized updates about important portfolio milestones or statistics and visuals related to the assets of individual clients

Ad Hoc updates will be broadcasted in response to unprecedented events affecting all such as the Covid-19 crisis and when needed. Personalized Ad Hoc updates relate to reaching out to individual clients during special circumstances such as when planning an asset allocation shift prior to retirement.

The use of technology also complements this strategy whereby White Rock can automate parts of it such as scheduled check-ins. Automation will improve efficiency and ensure the timely delivery of these high volume reportings. Touching base regularly strengthens client relationships and improves the client experience which ultimately helps White Rock retain them (Slemmer, 2020).

# 7. Further Research

**Limitation of Dataset 1**

Our group had selected the 2017 edition of the Irish government’s Civil Service Customer Satisfaction survey due to the comprehensive coverage of variables affecting overall customer satisfaction, with different variables relating to the main communication mediums used for interaction. However, it is observed that under the primary method of most recent contact with customers, the use of communication through email has a very low rate of 9% with only 9 responses as compared to communication through in-person and by-phone, which have a user rate of 29% and 41% respectively. The lack of data collected for communication through email affects the significance level of the variables under email, namely *ease\_of\_finding\_address, clarity, quality\_of\_advice* and *speed\_of\_response*.

A 2019 edition of the same survey was published on 3rd September 2019, but unfortunately, the dataset was not made available to the public. From the report of findings of the 2019 survey, we have noticed a few key differences between the results from both years. In 2017, the most popular interaction method was by phone, with 42% of customers surveyed using phone calls as their primary mode of interaction. E-mail was the fourth most popular method then, with only 9% using it as their primary mode. By 2019, within the span of two years, email has risen in popularity to become the most popular interaction method (29%), with interaction by phone sliding to a close second at 28%.

As WhiteRock is an asset management firm catering to high net-worth individuals and businesses, we believe that a significant portion of WhiteRock’s customers will be likely to utilise email messaging as their main mode of contact. While we were unable to identify significant variables affecting service satisfaction specific to service via email from the 2017 dataset, the 2019 dataset may allow us to do so given the larger sample size of surveyees using email as a communication method.

**Limitation of Dataset 2**

Our group had selected a dataset from Kaggle about a bank’s customer churning data, as a substitute for an asset management firm’s customer churning data. We acknowledge that there are inherent differences between an asset management firm’s customer base and a bank’s customer base in terms of their spending power, demographics, substitute options etc.

Further, given that the period of data collection was not specified, it is difficult to ascertain over what time period the data has been collected. If the period was short, there are likely to be little issues, if at all, with the implications and the suggestions put forth by our group. However, if the collection period was large, so as to span multiple years, the implications may have been rendered moot by other actions taken by White Rock.

In order to create a more comprehensive and accurate analysis and model of churn patterns in asset management firms, more accurate and specific data is needed in order to get more relevant results. To this end, one possible alternative is to acquire specific data sets from data vendors such as Acxiom in order to get more accurate analysis and models.

# 

# 

# 

# 

# 

# 8. Bibliography

Additiv. (2020). Becoming a Digital Wealth Leader. Thought Leadership, No. 1. Retrieved November 1, 2020, from https://additiv.com/assets/uploads/2020/02/Thought-Leadership-Becoming-a-Digital-Wealth-Leader.pdf

Benbria. (2018, May 08). The Key Omni-Channel Statistics You Need To Know. Retrieved November 01, 2020, from https://www.benbria.com/the-key-omni-channel-statistics-you-need-to-know/

Bielen, Frédéric & Demoulin, Nathalie. (2007). Waiting time influence on the satisfaction-loyalty relationship in services. Managing Service Quality. 17.174-193. 10.1108/09604520710735182.

Birkin, A., Wightman, M., & Nanayakkara, N. (2019, April 16). When wealth management clients want to switch, will you be their first choice? Retrieved November 01, 2020, from https://www.ey.com/en\_gl/wealth-asset-management/when-wealth-management-clients-want-to-switch-will-you-be-their-first-choice

Birkin, A., Wightman, M., & Nanayakkara, N. (2019, April 26). How to align pricing with value through transparency and variety. Retrieved November 01, 2020, from

https://www.ey.com/en\_gl/wealth-asset-management/how-to-align-pricing-with-value-through-transparency-and-variety

Brown, S. (2020, March 30). Client Communications Matrix To Strengthen Relationships. Retrieved November 01, 2020, from https://www.kitces.com/blog/client-communication-matrix-report-personalization-ycharts-sean-brown-frequency-style-channel/

Celent. (n.d.). Designing the Digital Wealth Management Client Experience. Retrieved November 01, 2020, from https://visit.dowjones.com/newswires/content/celent-digital-wealth-management/

Civil Service Public Customer Surveys. (n.d.). Retrieved November 01, 2020, from https://www.gov.ie/en/policy-information/a38d80-civil-service-general-public-customers/?referrer=http%3A%2F%2Fwww.per.gov.ie%2Fwp-content%2Fuploads%2FCustomer-Survey-2017-Full-Reprt-final-version-March-23-ilovepdf-compressed.pdf

Clearwater Analytics. (2015, August). How Clients Choose and Evaluate Asset Managers. Retrieved November 1, 2020, from http://d1pvbs8relied5.cloudfront.net/resources/white-papers/How-Clients-Choose-Evaluate-Asset-Managers-White-Paper.pdf

Commbox. (2019, November 12). Why Your Organization Should Adopt An Omnichannel Approach To Customer Communication. Retrieved November 01, 2020, from https://www.commbox.io/omnichannel-approach-for-customer-service/

Cseley, E. (2020, July 06). Waiting On Hold A Minute Too Long. Retrieved November 01, 2020, from https://www.jacada.com/blog/waiting-on-hold-a-minute-too-long/

Deyo, K. (2017, September 28). Digitizing the Wealth Management Experience: Omnichannel as the new Business Imperative. Retrieved November 01, 2020, from https://www.orchestracms.com/Digitizing-the-Wealth-Management-Experience-Omnichannel-as-the-new-Business-Imperative

Fonseca, I., & Mukerjee, M. (2019, August 28). Client Retention in Wealth Management. Retrieved November 01, 2020, from https://digitalreasoning.com/blog/analyzing-communications-the-next-step-in-client-retention-for-wealth-managers/

Guéguen, Nicolas & Jacob, Céline. (2002). The Influence of Music on Temporal Perceptions in an on-hold Waiting Situation. Psychology of Music - PSYCHOL MUSIC. 30. 210-214. 10.1177/0305735602302007.

Harari, O. (2017, June). Wealth Management Digitalization changes client advisory more than ever before. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/de/Documents/financial-services/Wealth%20Management%20Digitalization.pdf

Iacurci, G. (2020). Young investors are going digital. Financial advisors need to adapt with them. CNBC. Retrieved November 1, 2020, from https://www.cnbc.com/2020/10/14/millennials-gen-z-want-robo-advisors-and-digital-financial-advice.html

IE Staff. (2018, November 15). IG Wealth Management introducing unbundled pricing for all clients. Retrieved November 01, 2020, from https://www.investmentexecutive.com/news/products/ig-wealth-management-introducing-unbundled-pricing-for-all-clients/

Investopedia. (2020, January). Investment Strategies To Learn Before Trading. Retrieved from https://www.investopedia.com/investing/investing-strategies/

Jaacob, Zulnaidi. (2014). The Effect of Employee Training and Participation on Customer Satisfaction. 10.12776/QALI.V1.1.

Langan, M. (2012, October 02). How Long Will You Wait On Hold for Customer Service? Retrieved November 01, 2020, from http://www.prweb.com/releases/2012/10/prweb9964730.htm

Lu, Yina and Musalem, Andres and Olivares, Marcelo and Schilkrut, Ariel, Measuring the Effect of Queues on Customer Purchases (November 6, 2012). Columbia Business School Research Paper No. 12/23, Available at SSRN: https://ssrn.com/abstract=1851643 or http://dx.doi.org/10.2139/ssrn.1851643

Marriage, M. (2015, November 30). Banks embrace ‘the age of asset management’. Financial Times. Retrieved from https://www.ft.com/content/40c39d3e-9440-11e5-b190-291e94b77c8f

Morgan, B. (2019, September 24). 50 Stats That Prove The Value Of Customer Experience. Retrieved November 01, 2020, from https://www.forbes.com/sites/blakemorgan/2019/09/24/50-stats-that-prove-the-value-of-customer-experience/?sh=1ae1ebd4ef22

Nagarajan, S. (2020, July 9). The world's richest people are getting more and more annoyed at how much it costs them to look after their money | Markets Insider. Retrieved November 01, 2020, from https://markets.businessinsider.com/news/stocks/world-rich-want-to-switch-asset-managers-high-fees-2020-7-1029379900

Netwealth Investments Ltd. (2018, July 12). 72% of Investors say Fee Transparency Matters Most when choosing a Wealth Manager. Retrieved November 01, 2020, from https://www.netwealth.com/ourviews/investors-say-fee-transparency-matters-most-when-choosing-a-wealth-manager/

PricewaterhouseCoopers. (n.d.). Asset Management 2020: A Brave New World. Retrieved November 01, 2020, from https://www.pwc.com/gx/en/industries/financial-services/asset-management/publications/asset-management-2020-a-brave-new-world.html

Private Banker International. (2020, March) Retrieved from: https://www.privatebankerinternational.com/comments/2008-crisis-investor-behaviour-coronavirus/

Saketh, S., & Sridhar, G. (2020, March 27). Analytics and Machine Learning for Asset Management. Retrieved November 01, 2020, from https://www.latentview.com/blog/uses-of-analytics-and-machine-learning-in-asset-management-organizations/

Slemmer, M. (2020, October 25). Client communication best practice for asset management firms. Retrieved November 01, 2020, from https://www.bobsguide.com/guide/news/2017/Nov/16/client-communication-best-practice-for-asset-management-firms/

Son, H. (2020, October 14). Goldman Sachs wallops profit estimates on stronger-than-expected bond trading. Retrieved November 01, 2020, from https://www.cnbc.com/2020/10/14/goldman-sachs-gs-earnings-3q-2020.html

Stirista. (2020, February 26). Omnichannel Marketing: Six Benefits Your Brand Can Achieve. Retrieved November 01, 2020, from<https://www.stirista.com/omnichannel-marketing-six-benefits-brand-can-achieve>

The Balance. (2020, September). What Is Asset Management? Retrieved from: https://www.thebalance.com/asset-management-companies-fo r-beginners-4048203

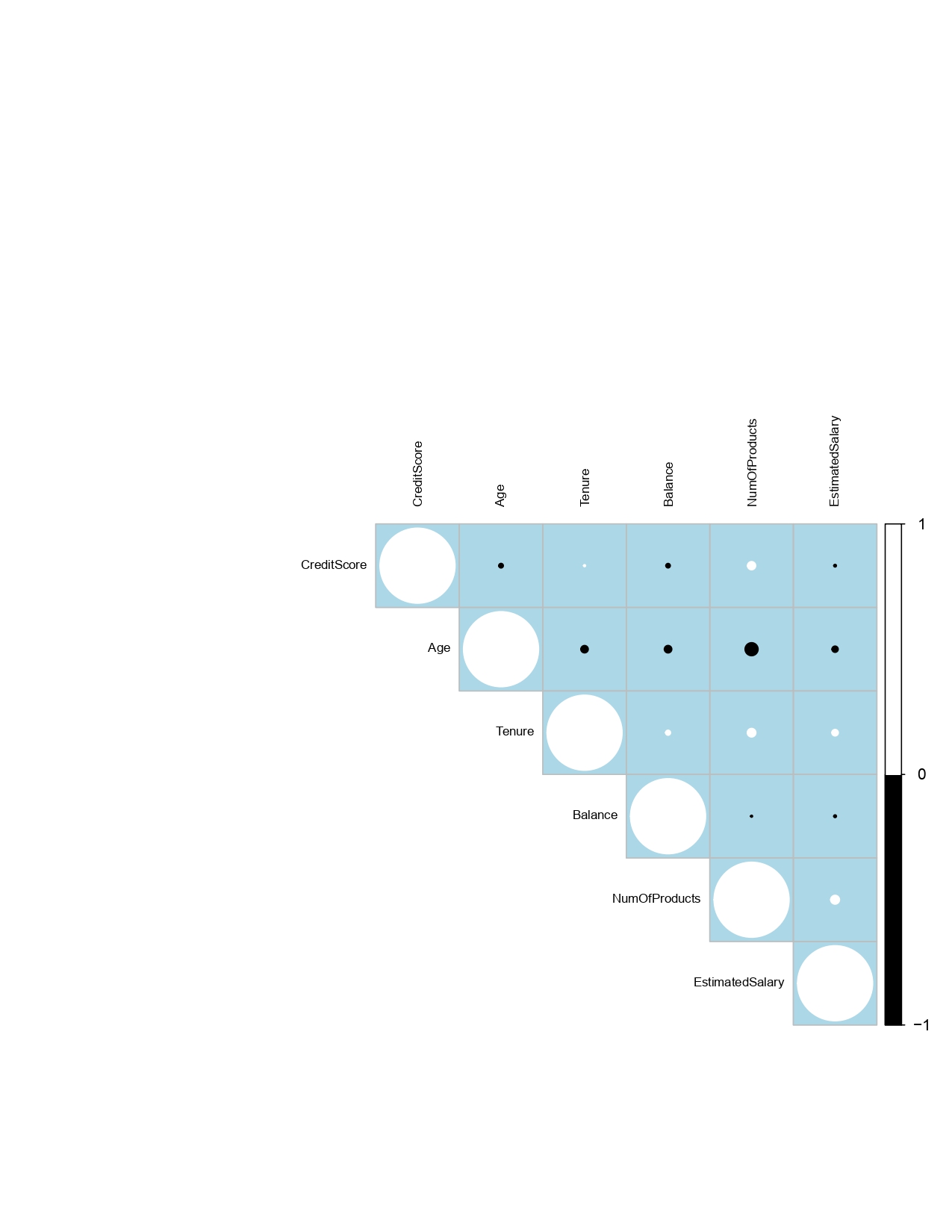
Wolla, S., & Sullivan, J. (2017, January). Education, Income, and Wealth. Retrieved November 01, 2020, from https://research.stlouisfed.org/publications/page1-econ/2017/01/03/education-income-and-wealth/

YCharts. (2019, December). How Can Advisors Better Communicate With Clients? YCharts Client Communications Survey. Retrieved from https://go.ycharts.com/hubfs/YCharts\_Client\_Communications\_Survey.pdf\

Vanguard. Retrieved from: https://investor.vanguard.com/financial-advisor/financial-advice/

# 9. Appendices

## Appendix 0 - CorrPlot for customer churn dataset



## Appendix 1.1: Data Dictionary for Dataset 1

|  |  |
| --- | --- |
| Column Name | Context |
| s.n | Serial Number assigned to survey response |
| age | Age of survey respondent |
| service\_satisfaction | Overall satisfaction with service received |
| outcome\_satisfaction | Overall satisfaction with outcome of the contact had |
| phone\_easeoffindingnumber | Ease of finding the telephone number |
| phone\_helpfulness | Helpfulness of staff |
| phone\_knowledge | Knowledge of staff |
| phone\_quality | Quality of advice / information received |
| phone\_manner | Manner in which staff explained issues / provided information |
| phone\_queryspeed | Speed / efficiency with which query was dealt with |
| phone\_answeringspeed | Speed with which phone was answered |
| phone\_holdingtime | Amount of time left holding |
| phone\_menu | Telephone menu / automated telephone services |
| phone\_voicemail | Voicemail service |
| writing\_easeoffindingaddress | Ease of finding correct address / contact person |
| writing\_clarity | Clarity of language used in written communication |
| writing\_design | Design and layout format |
| writing\_quality | Quality of advice / information given |
| writing\_speedresponse | Speed and efficiency of response to query |
| inperson\_knowledge | Knowledge of staff |
| inperson\_helpfulness | Helpfulness of staff |
| inperson\_qualityadvice | Quality of advice / information received |
| inperson\_location | Location of dept/office |
| inperson\_manner | Manner in which staff explained issues / provided information |
| inperson\_welcome | Welcome / reception received |
| inperson\_bizhours | Hours of business |
| inperson\_facilities | Public service area facilities |
| inperson\_queuing | Queuing system |
| inperson\_formdesign | Design and layout of forms |
| inperson\_speedquery | Speed / efficiency with which query was addressed |
| inperson\_helpwithforms | Help received in filling out forms |
| inperson\_privacy | Privacy of conversation / transaction |
| email\_clarity | Clarity of language used in e-mail |
| email\_easeoffindingaddress | Ease of finding correct email address / contact |
| email\_qualityadvice | Quality of advice / information received |
| email\_speedresponse | Speed / efficiency of response to query |

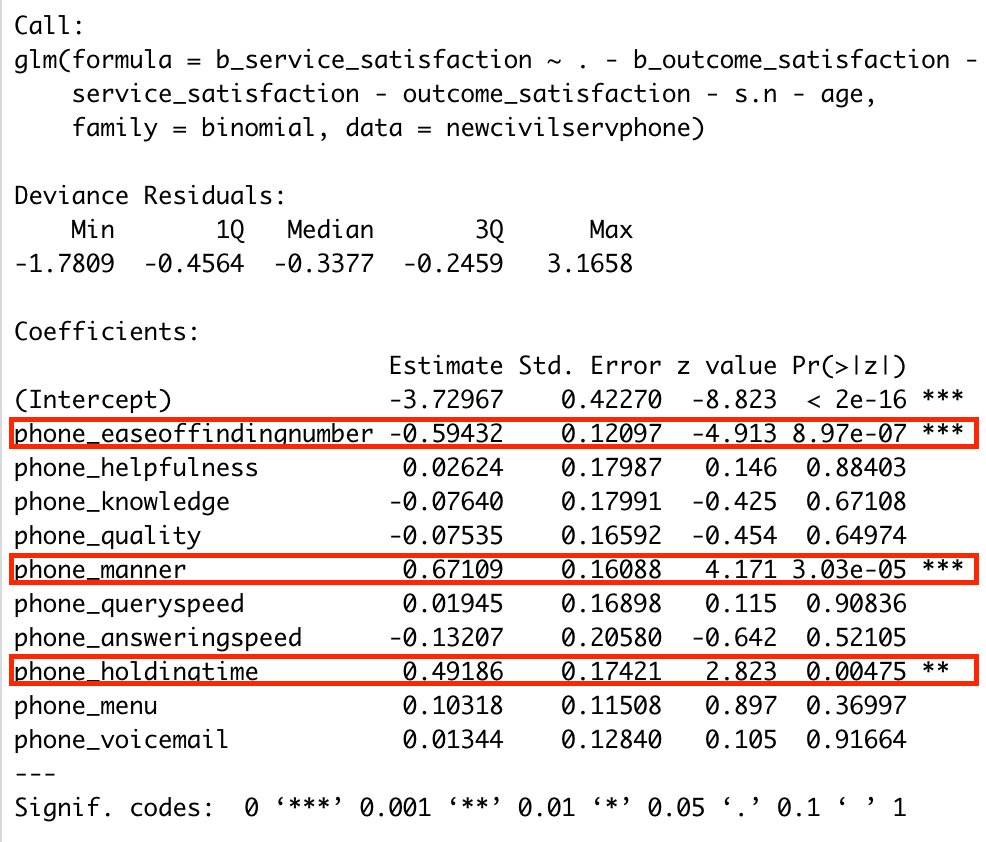
## 

## Appendix 1.2: Data Dictionary for Dataset 2

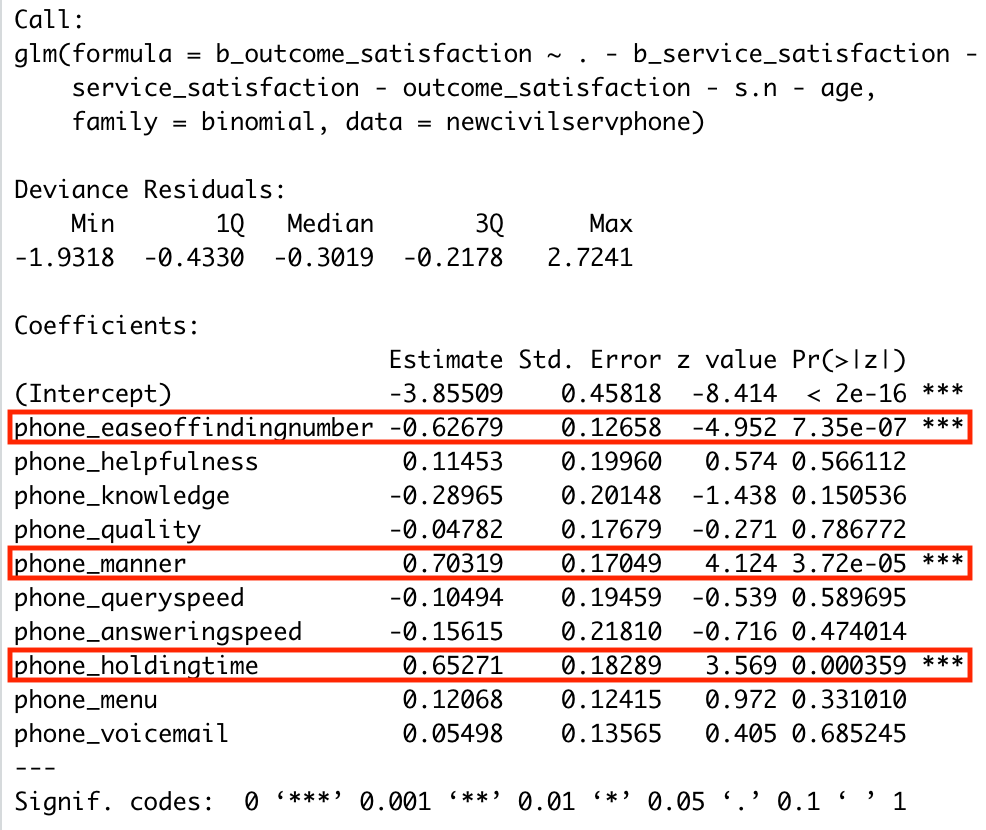
Below depicts the treatment of each variable in our analysis:

|  |  |
| --- | --- |
| **Column Name** | **Context** |
| RowNumber | The particular row number |
| CustomerId | A unique identifier |
| Surname | The particular customer’s surname |
| Geography | The customer’s nation of residence |
| Gender | The customer’s gender |
| Age | The customer’s age |
| Tenure | The period of time for which the customer has been with the firm |
| BankBalance | The estimated bank balance of the particular customer with a particular bank. |
| NumOfProducts | The number of services White Rock is providing the customer |
| HasCrCard | Whether or not the customer has a credit card |
| IsActiveMember | Whether or not the customer is an active member. Our interpretation of this is whether or not they request frequent check-ins or consultations with their asset manager |
| EstimatedSalary | The estimated salary of the particular customer |
| Exited | Whether or not the customer terminated their relationship as a client of White Rock |

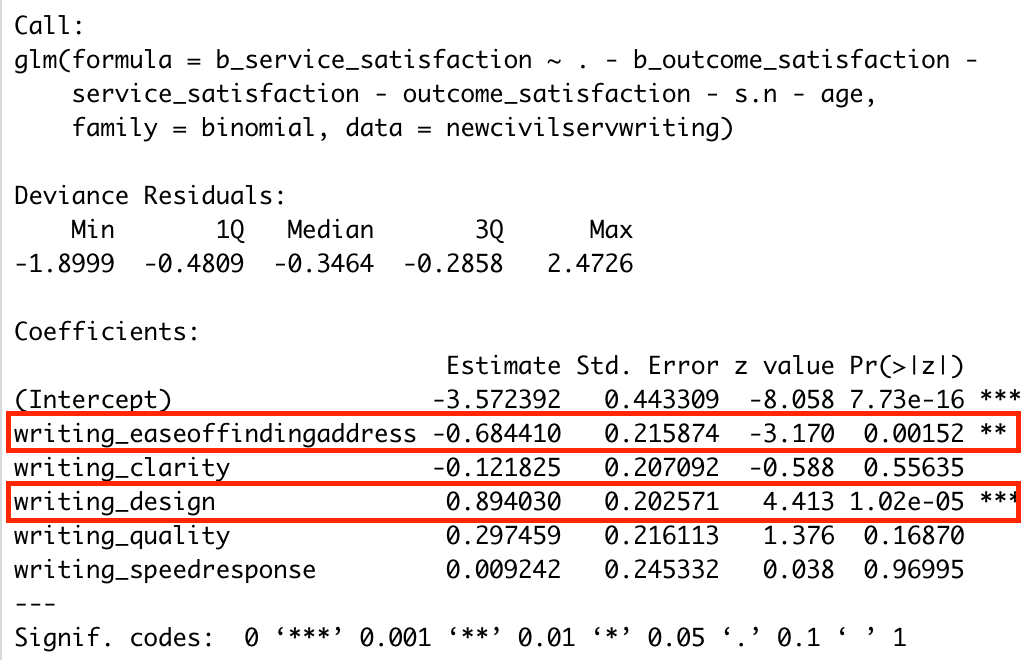
## Appendix 2.1: Logistic Regression for By-phone service satisfaction



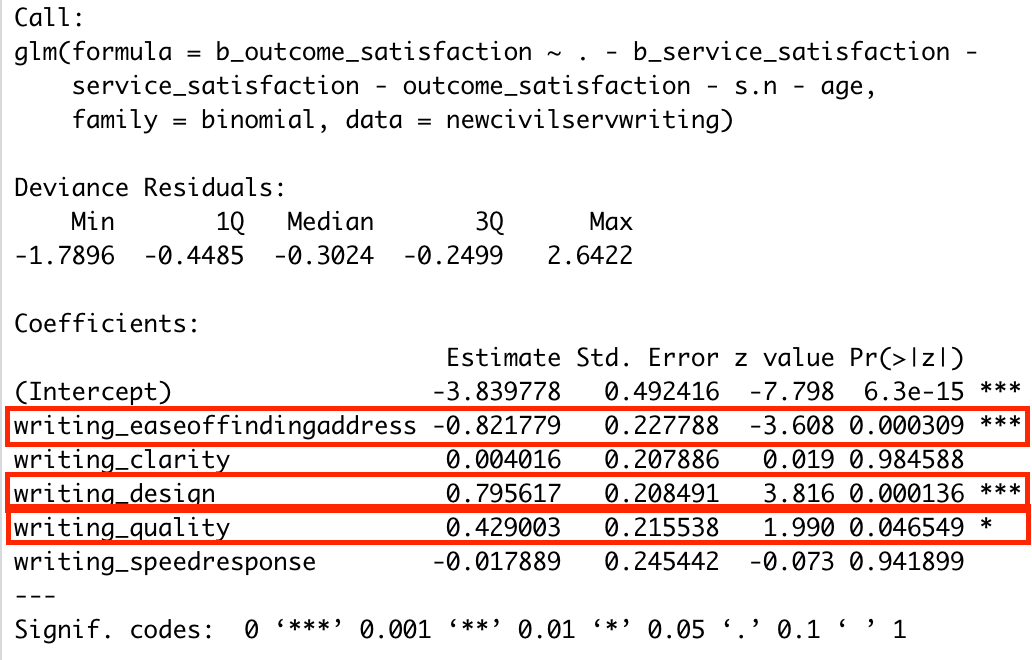
## Appendix 2.2: Logistic Regression for By-phone outcome satisfaction



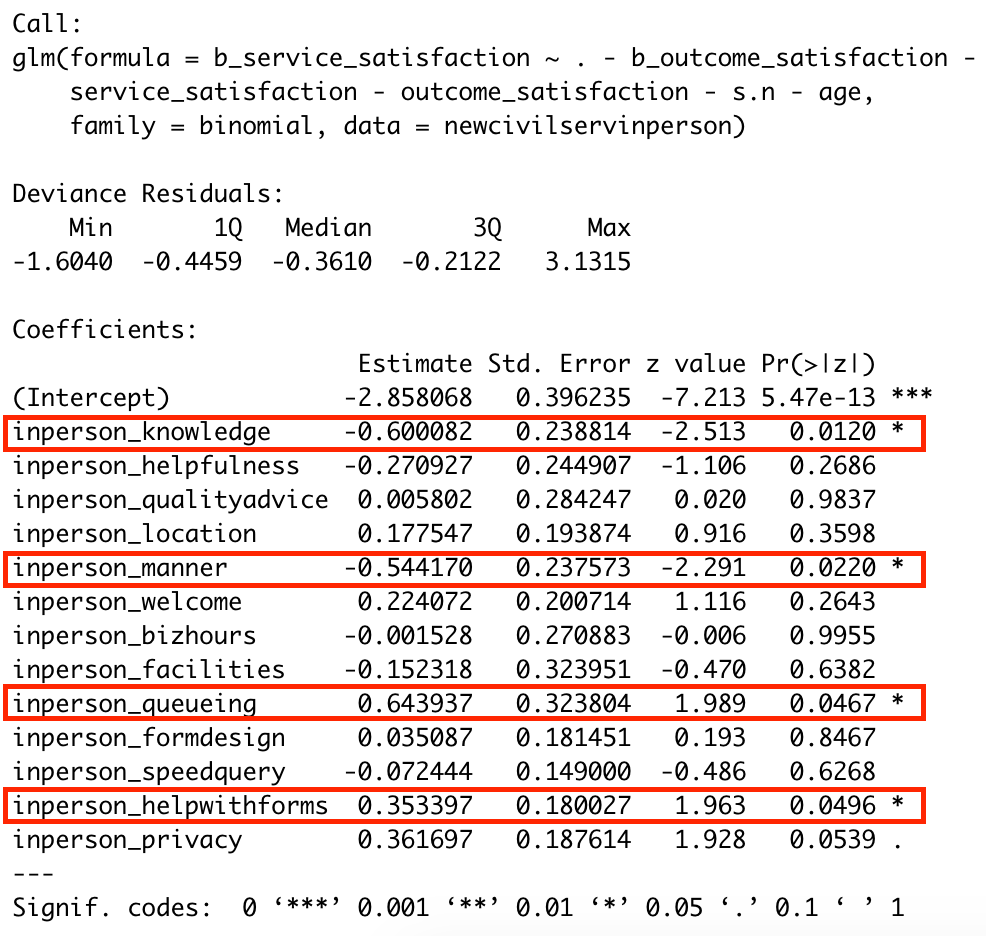
## Appendix 2.3: Logistic Regression for Writing service satisfaction

****

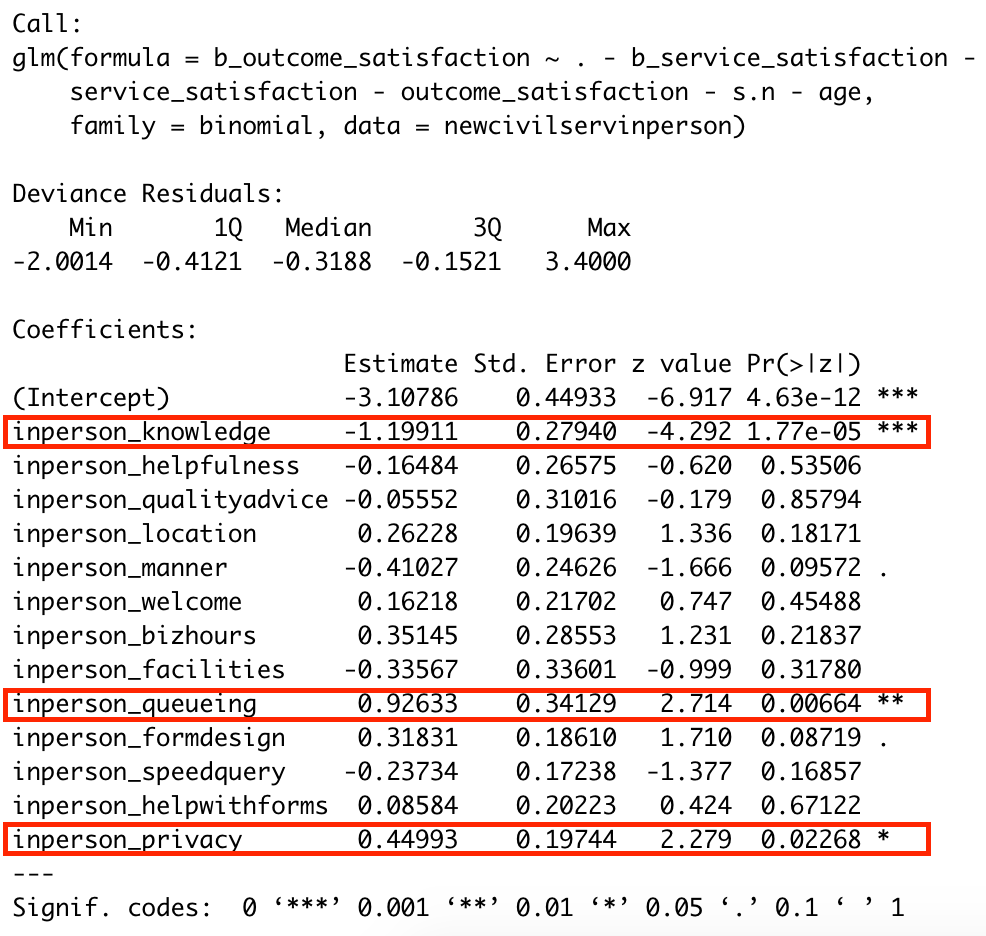
## Appendix 2.4: Logistic Regression for Writing outcome satisfaction

****

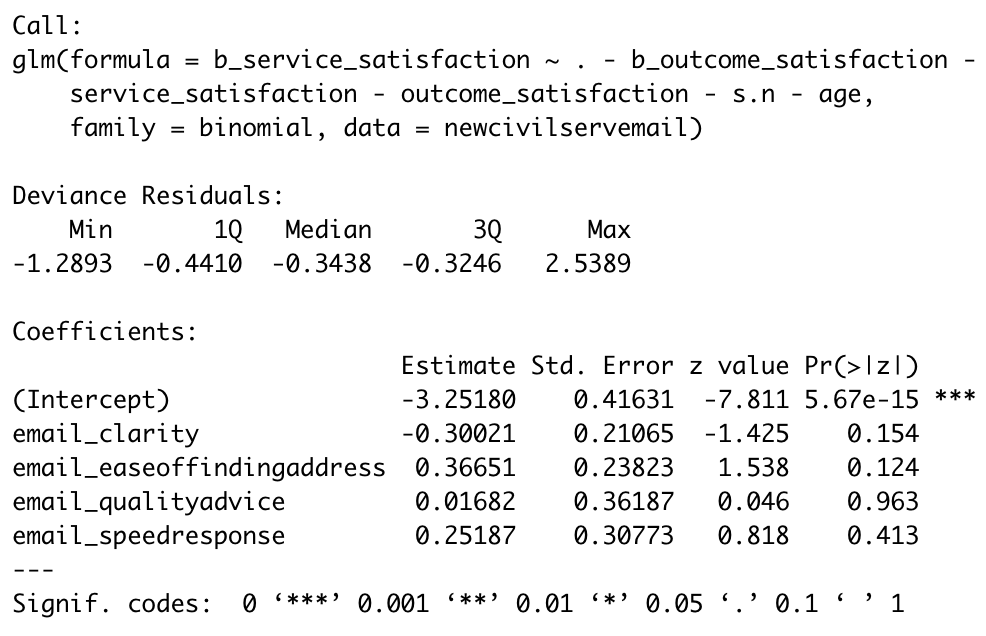
## Appendix 2.5: Logistic Regression for In-Person service satisfaction

****

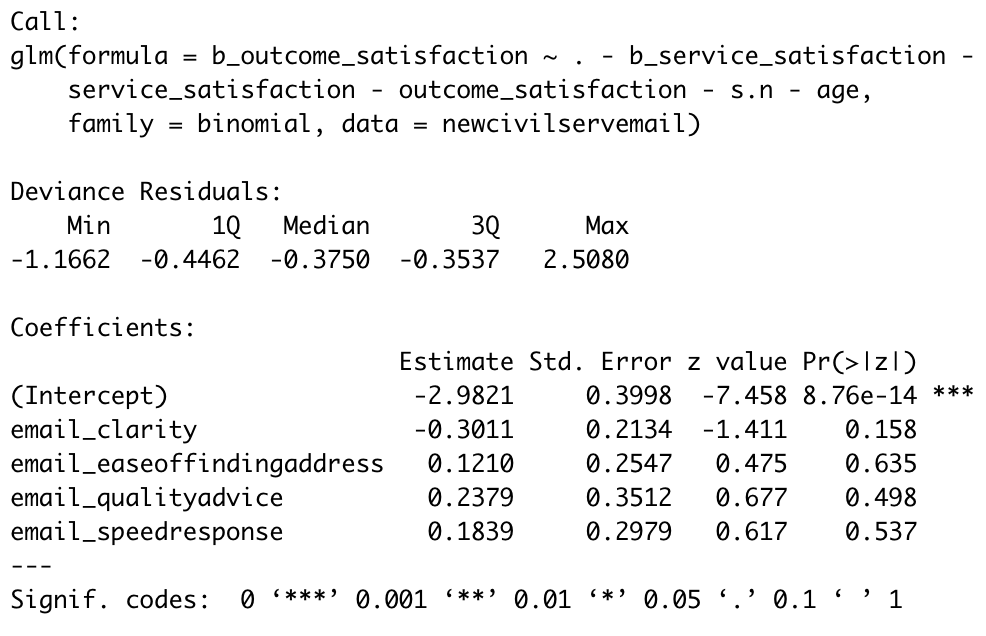
## Appendix 2.6: Logistic Regression for In-Person outcome satisfaction

****

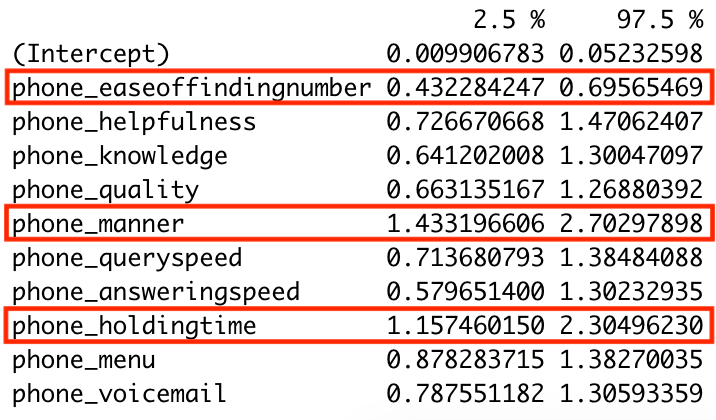
## Appendix 2.7: Logistic Regression for Email service satisfaction

****

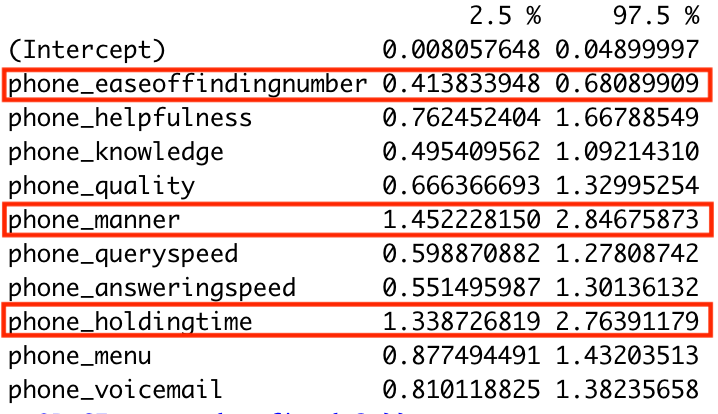
## Appendix 2.8: Logistic Regression for Email outcome satisfaction

****

## Appendix 2.9: Odds Ratio Confidence Interval for By-phone service satisfaction

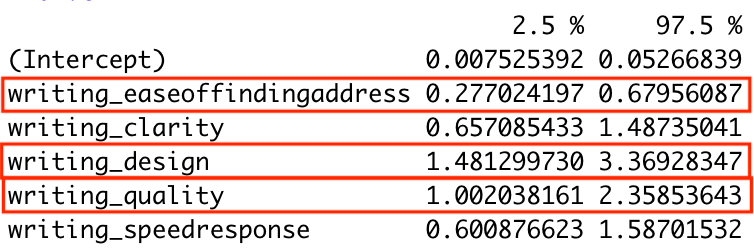


## Appendix 2.10: Odds Ratio Confidence Interval for By-phone outcome satisfaction

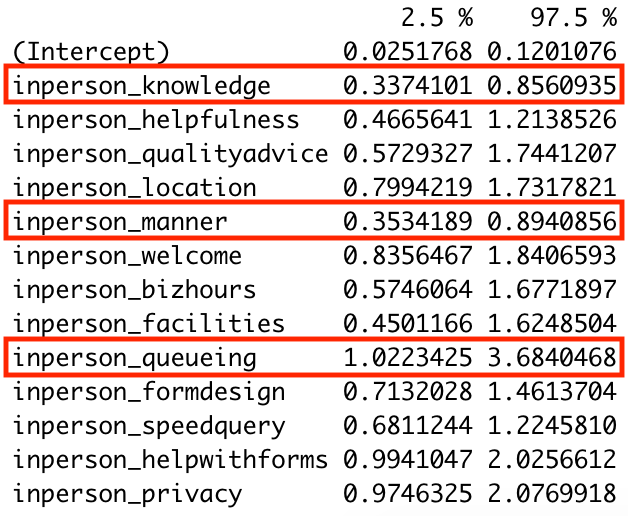


## Appendix 2.11: Odds Ratio Confidence Interval for Writing service satisfaction

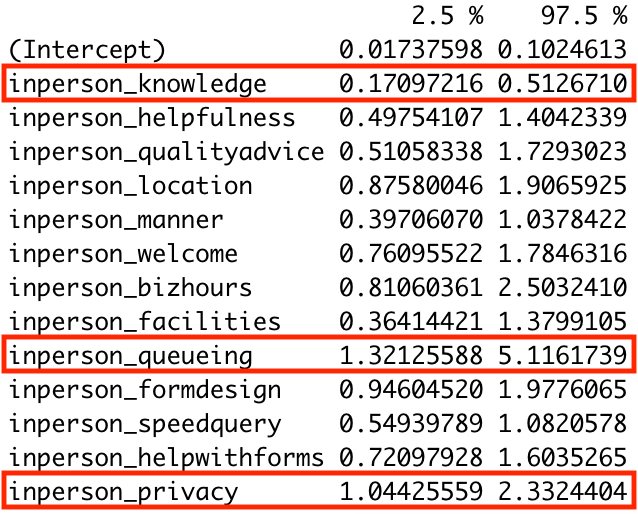
## Appendix 2.12: Odds Ratio Confidence Interval for Writing outcome satisfaction



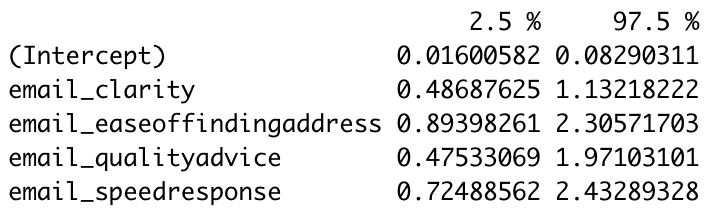
## Appendix 2.13: Odds Ratio Confidence Interval for In-Person service satisfaction



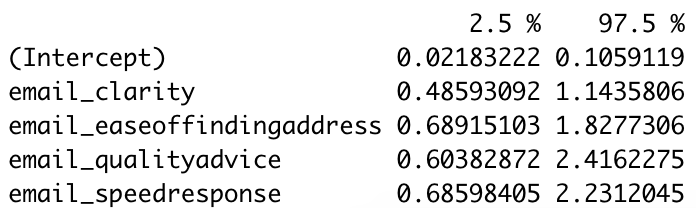
## Appendix 2.14: Odds Ratio Confidence Interval for In-Person outcome satisfaction



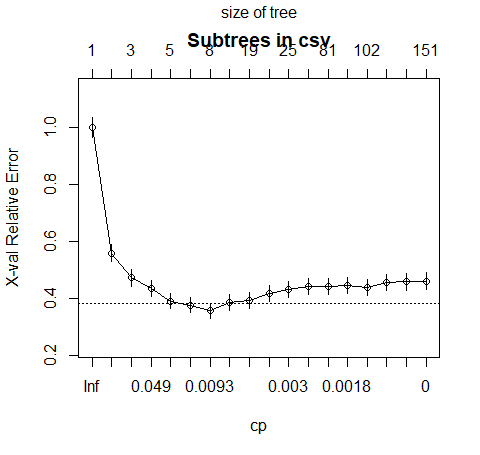
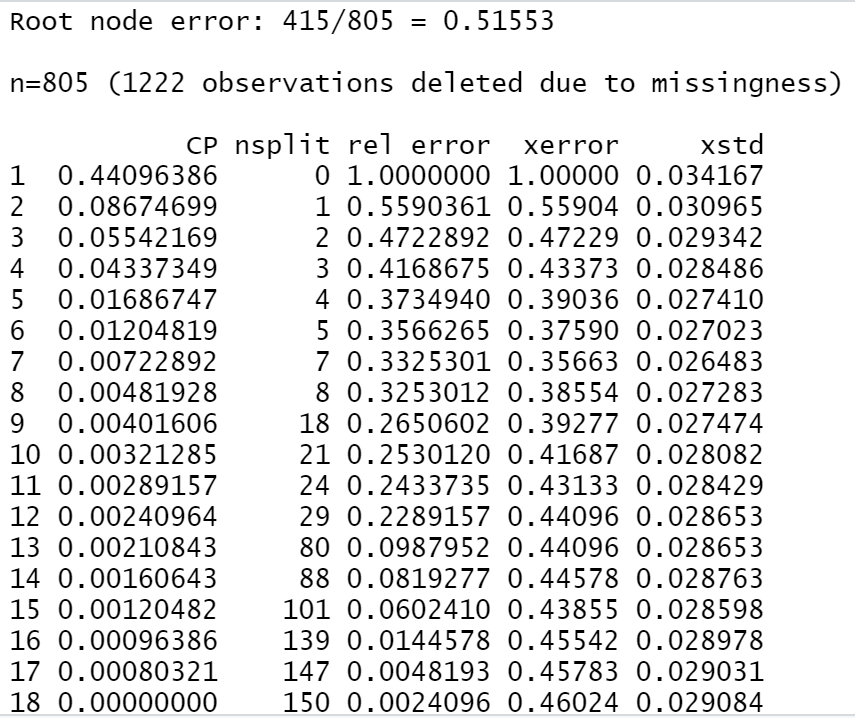
## Appendix 2.15: Odds Ratio Confidence Interval for Email outcome satisfaction



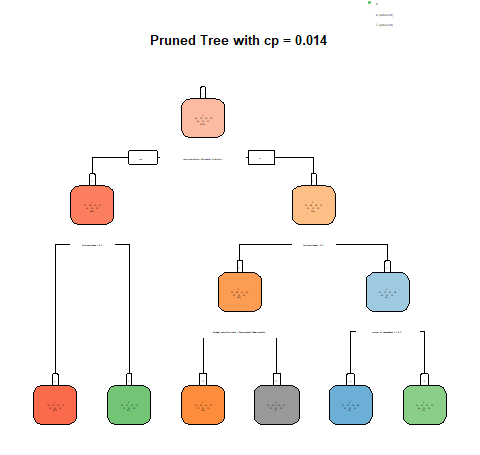
## Appendix 2.16: Odds Ratio Confidence Interval for Email service satisfaction

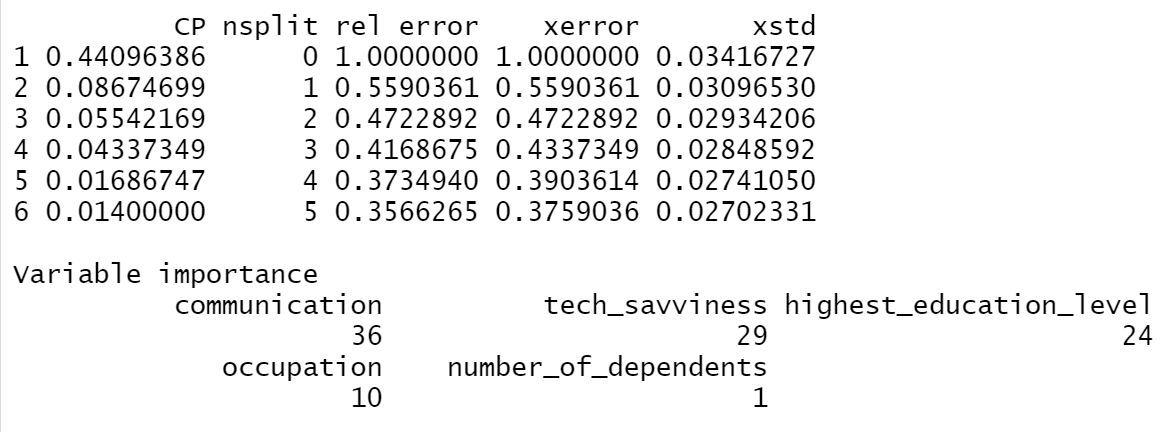


## Appendix 2.17: Maximal Tree root node error and CP values

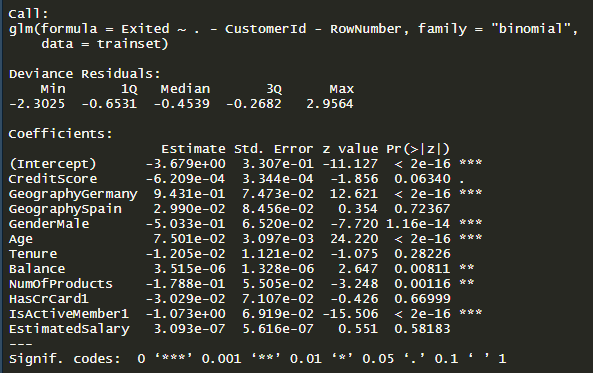
****Pruned at CP of sqrt(0.01686747 \* 0.01204819) = 0.014

## Appendix 2.18 Pruned Tree CP and Variable Importance

****

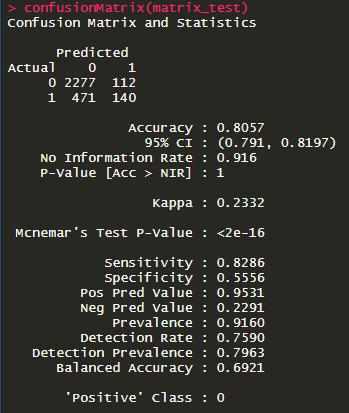
****

## Appendix 3.1 - Statistical Significance and Odds Ratio of Data Set 2

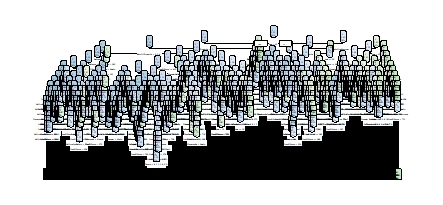
****

****

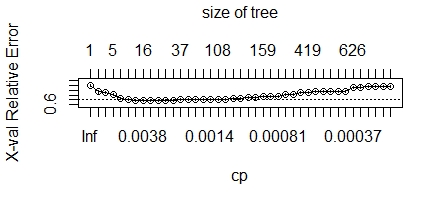
## Appendix 3.2 - Confusion Matrix

****

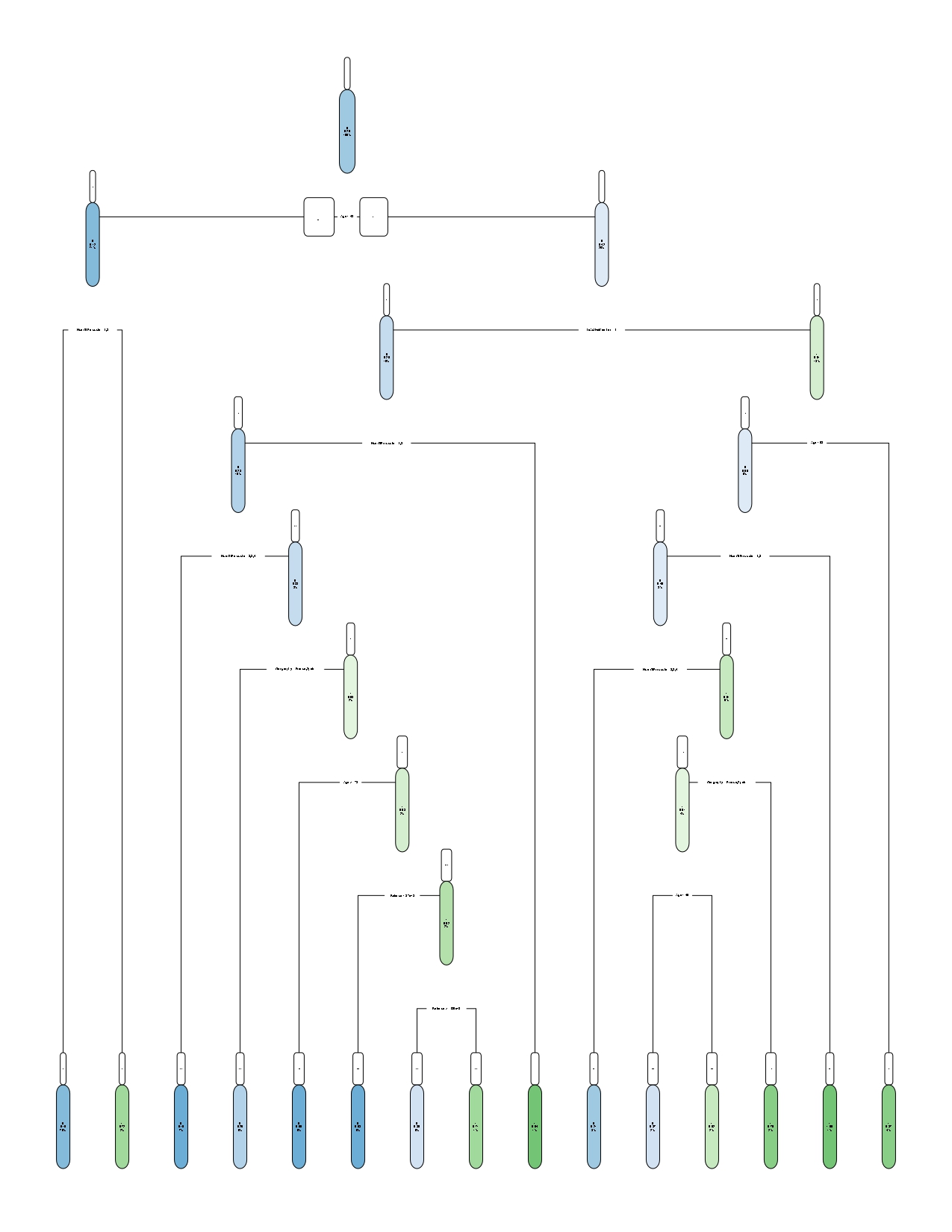
## Appendix 3.3 - Maximum Tree

****

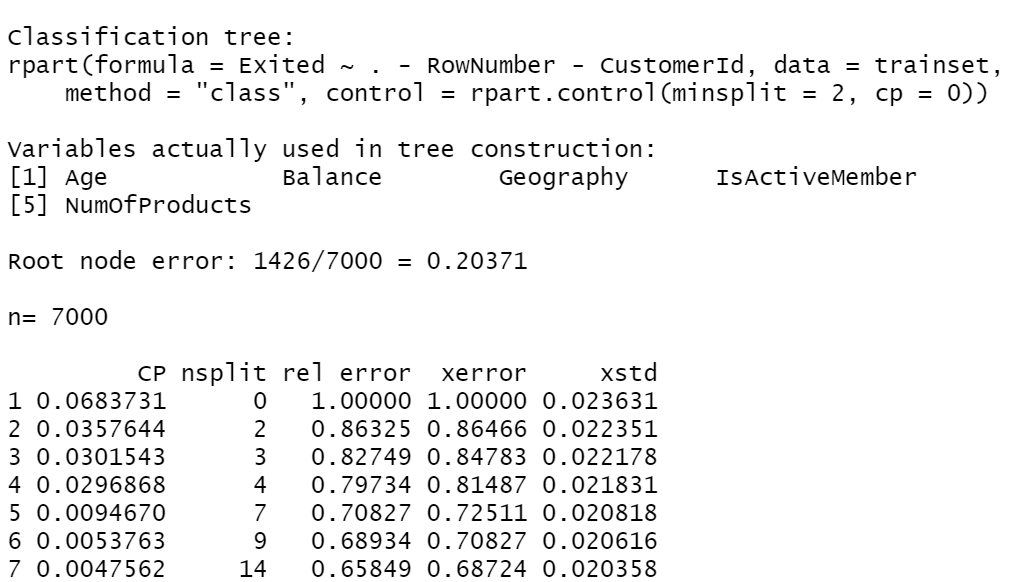
## Appendix 3.4 - Pruning Sequence of the Max Tree

****

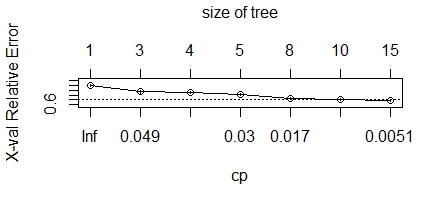
## Appendix 3.5 - Final Pruned Tree

****

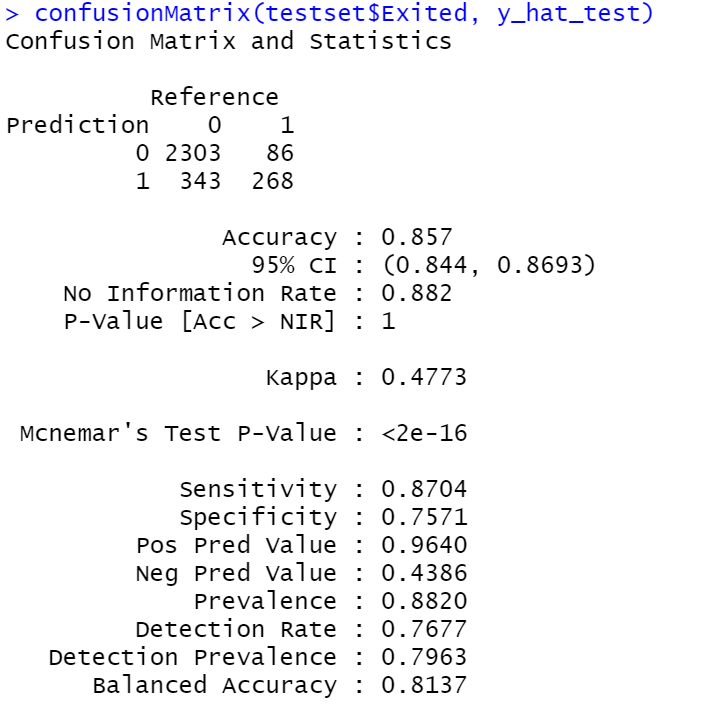
## Appendix 3.6 - printcp of the Pruned Tree

****

## Appendix 3.7 - Plotcp of the pruned tree

****

## Appendix 3.8: Implications of the confusion matrix statistics for our customer churn CART model



These results show that our model is fairly accurate at identifying both, actually positive as well as actually negative cases. Further, the positive predictive value test (Or precision) is very high, which implies that 96.40% of those who were identified as positive (Or not churning), were actually positive, or did not actually churn. Hence, there is a very good possibility that those customers who our model marks as will not churn, will actually stay with White Rock.

On the other hand, the negative predicted value leaves something to be desired. This test statistic tells us that of those identified negative (Or “churned”), only 43.86% were actually negative, i.e., did actually discontinue their relationship with White Rock.

What this means practically is that if the model were to tell us that a particular customer is not at risk of churning, we can rely on the model and mark that person ‘safe’ or ‘covered’ with a reasonably high degree of accuracy. However, if the model were to tell us that a person is indeed at risk of churning, we would need to follow up with perhaps another test to truly gauge if the client will churn.

What this does is that instead of directly giving us the at-high-risk customers, it tells us who will not churn for sure, which leaves the remainder of customers as the ones with potentially higher risk. In a way, these are merely 2 sides of the same coin, just with the additional step involved of taking the complement of the set of positive values.

## Appendix 3.9 Conditional probability of customer churning

