Practical Business Analytics Report.

Credit Card Churning Customers

Project Report

Table of Contents

1.DATASET DESCRIPTION

The major function of the bank is to store the funds of the public. Primarily, the process of the bank is to lend the funds to the public in the best interests of loan. Furthermore , banks make use of credit cards to the public to avail loans. Credit card permits you to borrow the money from the bank instead of taking it from your own bank account. Accordingly, the customers of the bank are the main key for the success of the bank. Losing a customer will lend to a disadvantage for the bank and to battle with other banks to stay in the market.It is very easy for customers to change their bank account from one bank to another when they are not satisfied. This is when churning occurs.

This data set provides information about customers and their education category from high school, post graduate, doctorate, uneducated and customer income from less than 40k to 120k.This kind of dataset helps with effective decision making in maximising customer value.The analysation with this data set will help managers in better marketing strategy. Manager here gets into a tangle as the customers cut off the credit card. They could be more glad as one of them would speculate who is gonna terminate the credit card as our employee would go first hand and switch the customers decision.

Data Overview

The “Predict churning customers” is a dataset consisting of 10,000 customers stating their marital status, credit card limit, age, salary, education level, etc. This dataset consists of 16.7% customers who have churned.

SOURCE: https://leaps.analyttica.com/home

1. PROBLEM DEFINITION

A manager at the bank is disturbed with more and more customers leaving their credit card services. They would really appreciate it if one could predict for them who is moving and get churned so they can proactively go to the customer to provide them better services and turn customers' decisions in the opposite direction.

They need to analyze the faults and drawbacks of their services provided to customers. Few of their customers were about to cancel their credit card services due to this the manager needs to know the faults of their services based on the previous data transactions and other categories.

Basically, customers move from a service that they have been using for a period of time when it is not a great deal or with bad services. Here, the manager needs to know whether this churned continues with the existing customers in the upcoming period.

As per the manager, we can even assume that it continues with the existing customers too. So, it’s the best way to analyze their previous data and predict whether the existing customers will make a move or not. Based on the predictions, managers can move-forward with better services and be on a safer side to hold those existing customers without churning. Our aim is to help the manager by analyzing the customers previous data and implement a few visualizations for better understanding about the data. Then, predict the customers who were about to leave by using some prediction models & techniques.

This dataset provides that only 16.7% of customers have been churned. Thus, this obstacle needs more effort to predict which customer is going to proceed to churn the credit card, if and only if we project the condition we can instantly go round and substitute the customers decision.There are many cases for the customers to quit their credit cards. For instance, they are unable to make payment before the margin. Perhaps, this leads to huge interest on loans and may even lead to bankruptcy. Thus, cards may decline. Numerous types of problems occur in credit cards. Given the dataset provides a slight number of details accordingly it needs more effort to work out the problem.

2. DATA PREPROCESSING

Data preprocessing is the process of preparing data in the dataset for implementing machine learning models. In this the data set is used to visualize the shape of data set and the type of the variables/attributes and checking whether the data is having any missing values for the further proceedings.

The attrited_flag is the attribute which is the target variable among all the 20 variables,The remaining variables are independent variables. As the process of preprocessing or preparation we have noticed that the data does not contain any missing values by using the cleaning techniques in the r studio.

- ’Clientnum’
- ’Naive_bayes_classifier_attrition_flag_card_category_contacts_count_12_mon_dependent_count_education_level_months_inactive_12_mon_1’
- ’Naive_bayes_classifier_attrition_flag_card_category_contacts_count_education_level_months_inactive_12_mon_2’

These above columns that have been not used in the data set and are removed at the early stage of the pre processing of the dataset because they are completely meaningless for the target variable
As we are dealing with different types of modeling techniques like naive bayes, random forest support vector machines etc… which needed the categorical data to be in the binary form of the data .

After Cleaning:

```
# Checking for missing values
colSums(sapply(data, is.na))
     Attrition_Flag        Customer_Age              Gender     Dependent_count     Education_Level      Marital_Status
                  0                   0                   0                   0                   0                   0
    Income_Category       Card_Category      Months_on_book Total_Relationship_Count Months_Inactive_12_mon Contacts_Count_12_mon
                  0                   0                   0                   0                   0                   0
       Credit_Limit Total_Revolving_Bal     Avg_Open_To_Buy Total_Amt_Chng_Q4_Q1     Total_Trans_Amt      Total_Trans_Ct
                  0                   0                   0                   0                   0                   0
Total_Ct_Chng_Q4_Q1 Avg_Utilization_Ratio
                  0                   0
```

3. DATA EXPLORATION

In this part the dataset was analysed for attrition_flag attribute by using the variable columns or attributes in the dataset.

Attrited Customer	Existing Customer
1627	8500

By the values we can infer that existing customers are more, whenever the existing customers are reduced there will be great impact on the organisation. In addition to this we are looking into the dataset to understand what factors are contributing towards the attrition flag.

4.1 Attributes:

The dataset consists of 20 columns which have different kinds of data used to explain the attrition flag of the customer.

1. 'Attrition_Flag'
2. 'Customer_Age'
3. 'Gender'
4. 'Dependent_count'
5. 'Education_Level'
6. 'Marital_Status'
7. 'Income_Category'
8. 'Card_Category'
9. 'Months_on_book'
10. 'Total_Relationship_Count'
11. 'Months_Inactive_12_mon'
12. 'Contacts_Count_12_mon'
13. 'Credit_Limit'
14. 'Total_Revolving_Bal'
15. 'Avg_Open_To_Buy'
16. 'Total_Amt_Chng_Q4_Q1'
17. 'Total_Trans_Amt'
18. 'Total_Trans_Ct'
19. 'Total_Ct_Chng_Q4_Q1'
20. 'Avg_Utilization_Ratio'

4.2 Variable or Attribute Summaries:

As a part of data exploration the attributes consist of two types of data which are Numerical and Categorical data.

Numeric attributes	Categorical attributes
Customer_Age	Attrition_Flag
Dependent_count	Gender
Months_on_book	Education_Level
Total_Relationship_Count	Marital_Status
Months_Inactive_12_mon	Income_Category
Contacts_Count_12_mon	Card_Category
Credit_Limit	
Total_Revolving_Bal	
Avg_Open_To_Buy	
Total_Amt_Chng_Q4_Q1	
Total_Trans_Amt	
Total_Trans_Ct	
Total_Ct_Chng_Q4_Q1	
Avg_Utilization_Ratio	

4.3 Summary of data:

The summary of is shown as features like mean, median, minimum, maximum etc type of details.

```
> # numeric summary
> summary(numeric_var)
  Customer_Age    Dependent_count Months_on_book  Total_Relationship_Count Months_Inactive_12_mon Contacts_Count_12_mon  Credit_Limit    Total_Revolving_Bal
 Min.   :26.00   Min.   :0.000   Min.   :13.00   Min.   :1.000            Min.   :0.000          Min.   :0.000          Min.   : 1438   Min.   :   0
 1st Qu.:41.00   1st Qu.:1.000   1st Qu.:31.00   1st Qu.:3.000            1st Qu.:2.000          1st Qu.:2.000          1st Qu.: 2555   1st Qu.: 359
 Median :46.00   Median :2.000   Median :36.00   Median :4.000            Median :2.000          Median :2.000          Median : 4549   Median :1276
 Mean   :46.33   Mean   :2.346   Mean   :35.93   Mean   :3.813            Mean   :2.341          Mean   :2.455          Mean   : 8632   Mean   :1163
 3rd Qu.:52.00   3rd Qu.:3.000   3rd Qu.:40.00   3rd Qu.:5.000            3rd Qu.:3.000          3rd Qu.:3.000          3rd Qu.:11068   3rd Qu.:1784
 Max.   :73.00   Max.   :5.000   Max.   :56.00   Max.   :6.000            Max.   :6.000          Max.   :6.000          Max.   :34516   Max.   :2517
 Avg_Open_To_Buy Total_Amt_Chng_Q4_Q1 Total_Trans_Amt Total_Trans_Ct   Total_Ct_Chng_Q4_Q1 Avg_Utilization_Ratio
 Min.   :    3   Min.   :0.0000       Min.   :  510   Min.   : 10.00   Min.   :0.0000      Min.   :0.0000
 1st Qu.: 1324   1st Qu.:0.6310       1st Qu.: 2156   1st Qu.: 45.00   1st Qu.:0.5820      1st Qu.:0.0230
 Median : 3474   Median :0.7360       Median : 3899   Median : 67.00   Median :0.7020      Median :0.1760
 Mean   : 7469   Mean   :0.7599       Mean   : 4404   Mean   : 64.86   Mean   :0.7122      Mean   :0.2749
 3rd Qu.: 9859   3rd Qu.:0.8590       3rd Qu.: 4741   3rd Qu.: 81.00   3rd Qu.:0.8180      3rd Qu.:0.5030
 Max.   :34516   Max.   :3.3970       Max.   :18484   Max.   :139.00   Max.   :3.7140      Max.   :0.9990
>
```

4.4 Analysis of customer education level:

- **Attrited customer:**

The customers whose education level is post graduate are the least attrited customers and the customers whose education level is graduate are most attrited customers in the different types of education levels of customers in the dataset.

- **Existing customer:**

The customers whose education level is graduate are the most existing customers and the customers whose education level is doctorate are least existing customers
From the analysis of data we can tell that the customers who are graduates are the most existing as well as most attrited customers.

```
> Education_Flag = with(data, table(Attrition_Flag, Education_Level))
> Education_Flag
                   Education_Level
Attrition_Flag      College Doctorate Graduate High School Post-Graduate Uneducated Unknown
  Attrited Customer     154        95      487         306            92        237     256
  Existing Customer     859       356     2641        1707           424       1250    1263
>
```

4.5 Analysis of customer income level:

- **Attrited customer:**

The customers whose income level is $120k+ are the least attrited customers whereas the customers whose income level is less than $40k are the most attrited customers.

- **Existing customer:**

The customers whose income level is $120k+ are the least existing customers whereas the customers whose income level is less than $40k are the most existing customers.

```
> Income_Flag = with(data, table(Attrition_Flag, Income_Category))
> Income_Flag
                   Income_Category
Attrition_Flag      $120K + $40K - $60K $60K - $80K $80K - $120K Less than $40K Unknown
  Attrited Customer     126         271         189          242            612     187
  Existing Customer     601        1519        1213         1293           2949     925
> 
```

4.6 Analysis of customer card details:

- **Attrited customer:**

The customers whose card category is platinum are the least attrited customers whereas the customers whose card category is blue are the most attrited customers.

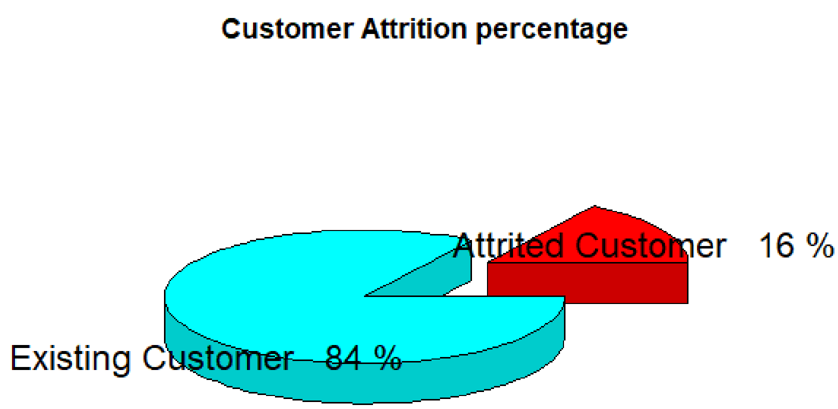
- **Existing customer:**

The customers whose card category is platinum are the least existing customers whereas the customers whose icard category is blue are the most existing customers.

```
> Card_Flag = with(data, table(Attrition_Flag, Card_Category))
> Card_Flag
                   Card_Category
Attrition_Flag      Blue Gold Platinum Silver
  Attrited Customer 1519   21        5     82
  Existing Customer 7917   95       15    473
> 
```

4. EXPLORATORY DATA ANALYSIS

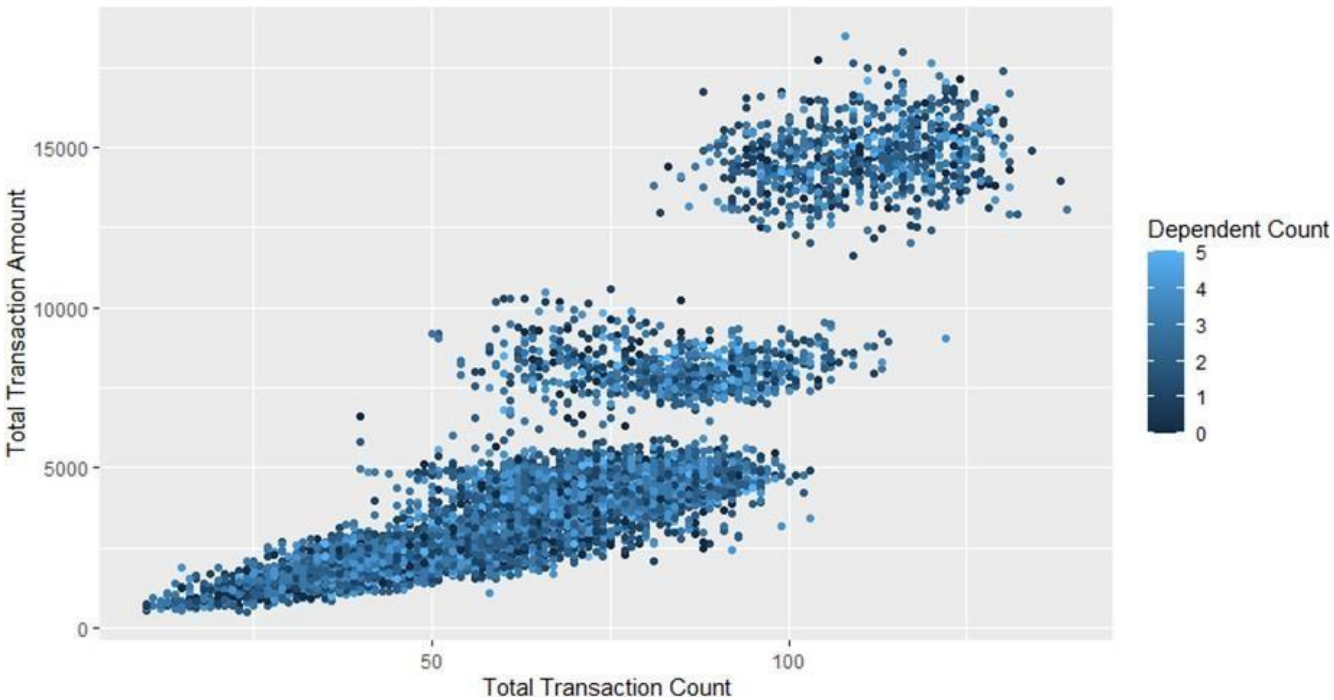
5.1 Attrition Flag



We can conclude from the Pie chart that current customers account for roughly85% of the total, whereas attired customers account for around 15%.

We're attempting to anticipate whether existing customers will remain or departbased on various key factors.

5.2 Transaction Count



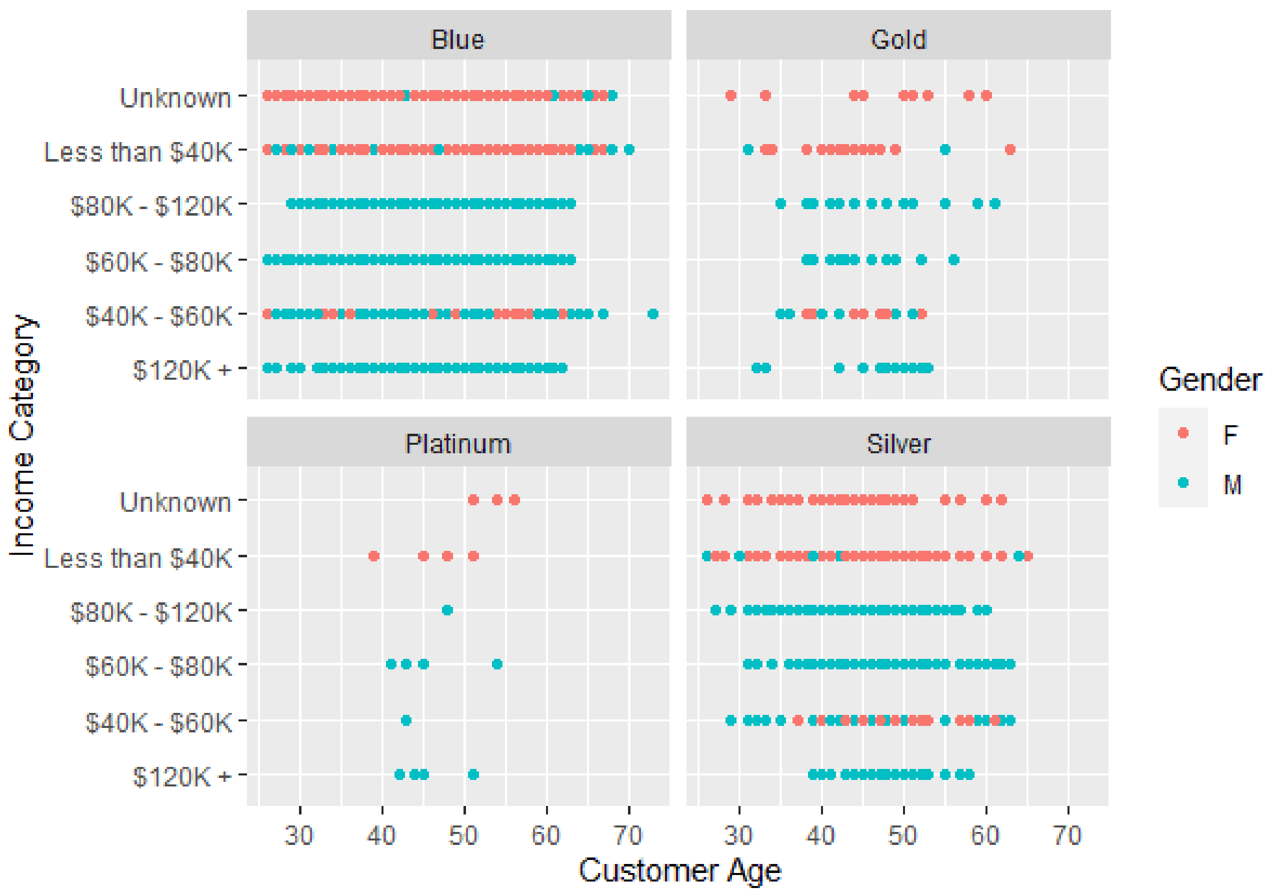
For a specific time period, the plot shows the combinations of "Total Transaction Count" and "Total Transaction Amount" with the Dependent Count (the number ofpeople dependent on the person who did these transactions).

As can be seen in the graph, three clusters have developed, with the majority of the initial clients falling below the $5K mark.

The minimum and maximum transactions in the Second Cluster are roughly 7.5K and 10K, respectively.

The third cluster is pretty obvious; all of them have completed at least 7.5K transaction amounts with a minimum of 80 transactions.

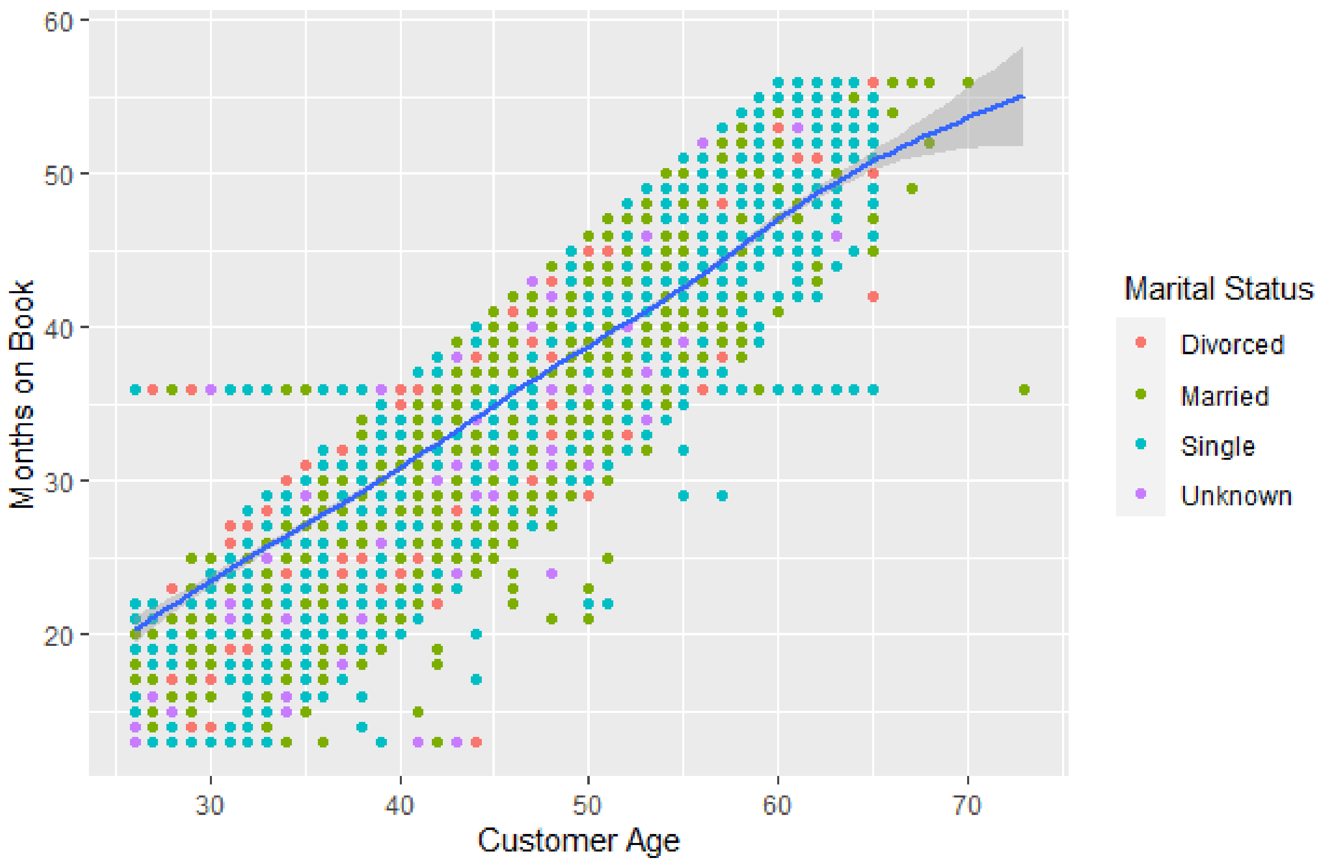
5.3 Customer Age, Income Category and Genders



We can learn a lot from the above graph. For instance, all of the female clients are in the 40k-60k range and below, and if we look closely, nearly everyone in the Unknown income group is a woman.

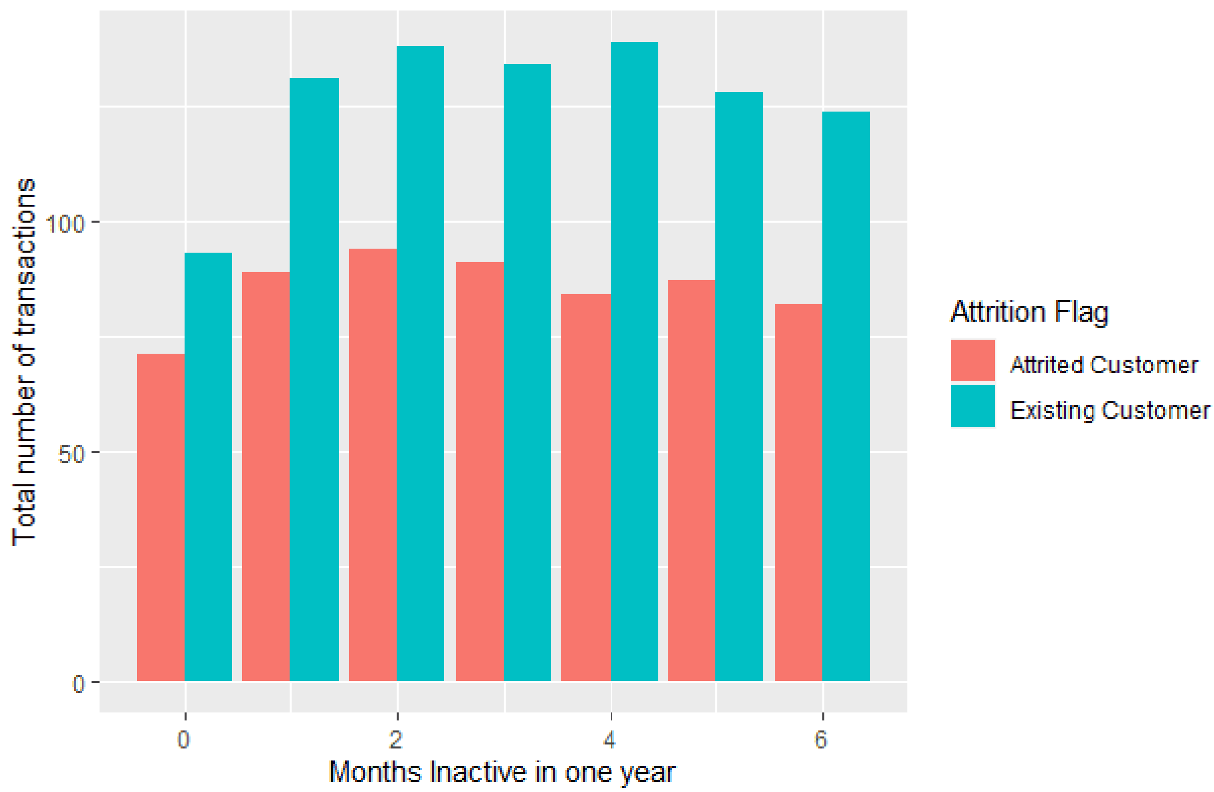
Furthermore, the majority of customers pick Blue Cards regardless of their financial range. In addition, the Platinum Card is the least popular of the four.

5.4 Customer Age, Months on book and Marital Status



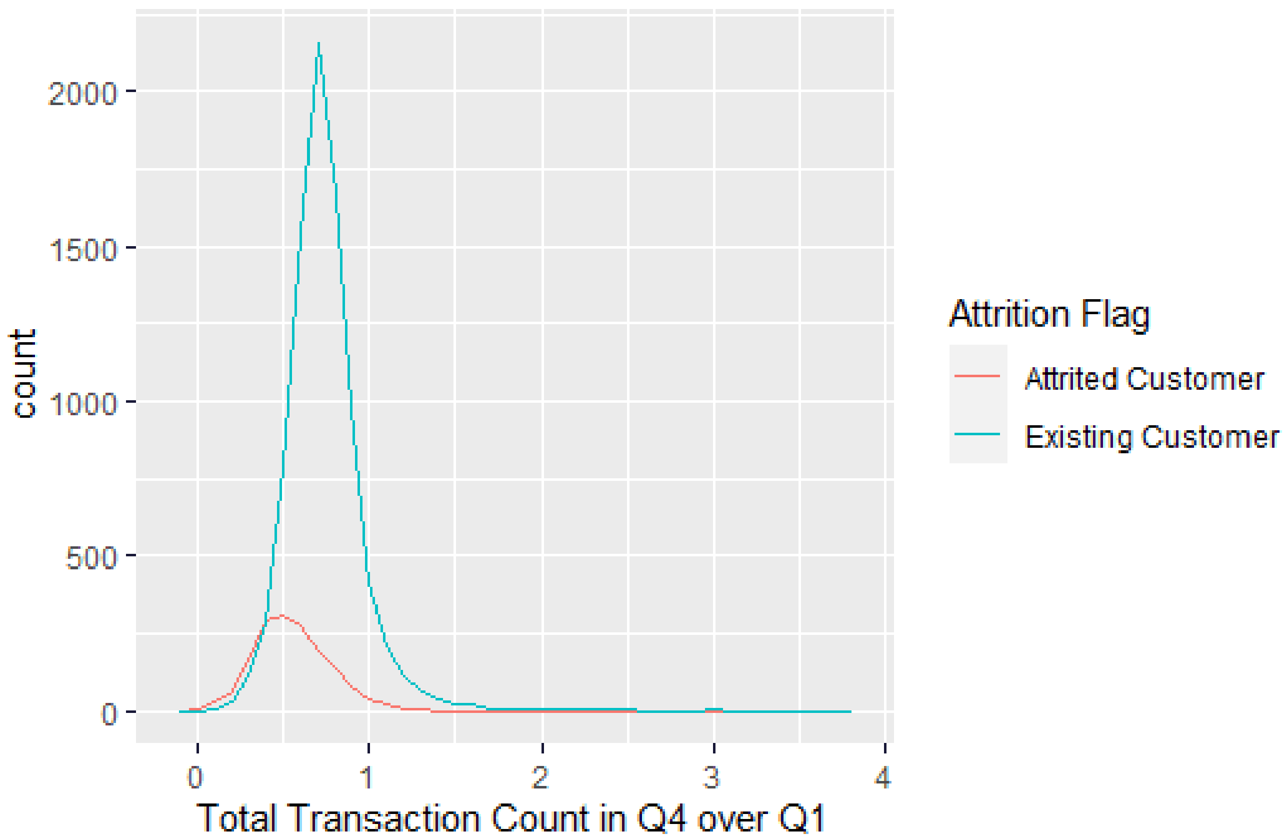
From the above graph, we can learn that there is a linear relation, established between the Customer's age and the period of relationship with the bank. Customers that are older are more likely to remain longer. Additionally, we can see there are few clusters for marital Status.

5.5 Months inactive in one year, Total number of transactions & Attrition Flag



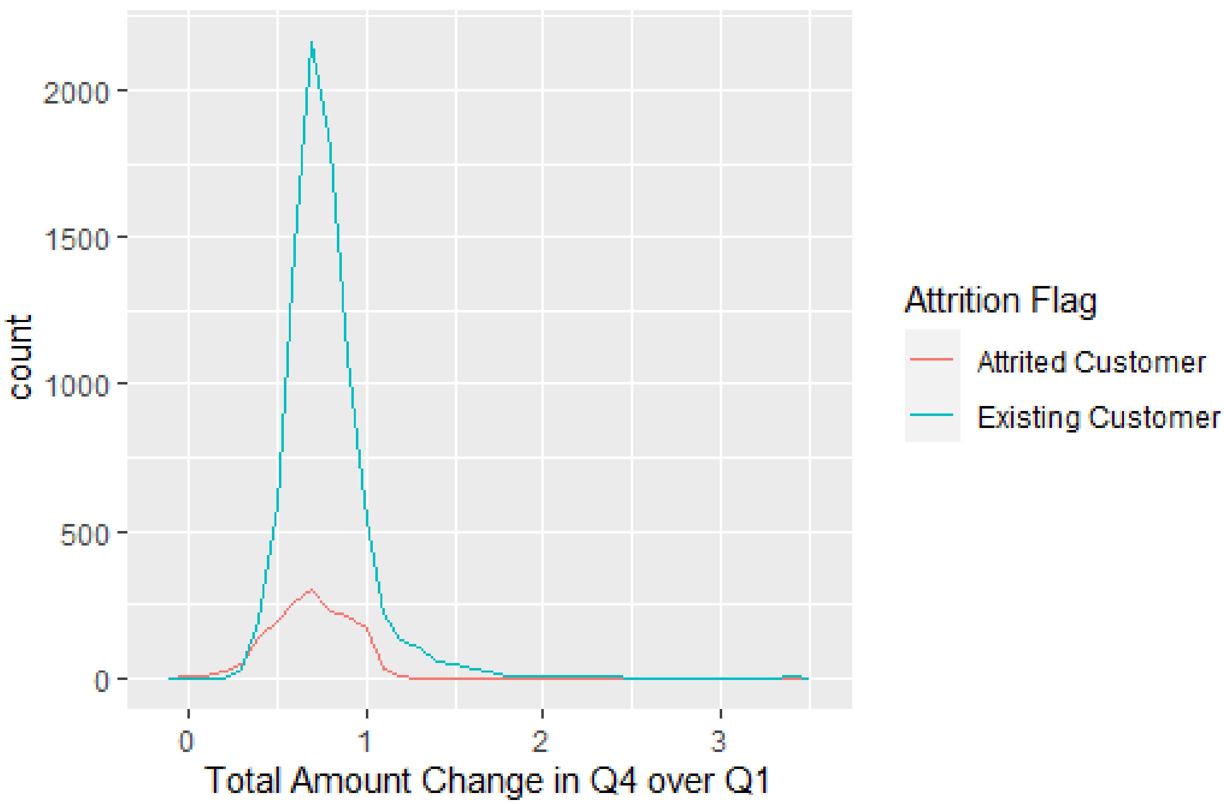
The attrited customers never completed any transactions over 100 on the bar graph, indicating that if a customer crosses the 100 transaction barrier, he is less likely to quit the bank.

5.6 Total Transaction Count in Q4 over Q1



When compared to Q1, the majority of customers did not complete the same number of transactions in Q4. In Q4, attrited consumers had an average of roughly 0.5, while existing customers had an average of around 0.7. Finally, the majority of consumers have a credit score of less than 1, which is a disadvantage for the bank because it expects a typical transaction flow.

5.7 Total Amount change in Q4 over Q1



The Total amount Change in quarters is reflecting the same from the Transaction Graph from the previous graph. The change in the amount is Below 1.

5.8 Number of Contacts in one year

"Contact count in one year" (number of times the consumer contacted the bank in one year) and "Attrition flag" is used to visualize this plot. The goal of this graph is to discover the relationship between these two combinators.

Each bar represents a percentage combination of attrition and current customers. We can observe a pattern here: as the number of contacts with the bank increases, the number of attracted clients increases as well. For example, from 0 to 5 visits, the count has climbed from roughly 2% to 25% within the 5th visit. Finally, a customer visiting the bank for the sixth time is a red flag, indicating that the customer is quite likely to depart.

5. FEATURE ENGINEERING

Feature Engineering is the process of transformation of raw data into features that represent the problem of the predictive model that we are going to work on. They are mainly used to arrange the input dataset accordingly so that they get in a great structure for the models to split them and predict the accuracy.

The types of Feature engineering used on our dataset are as follows:

6.1 Label Encoders

Label Encoders is one of the methods for feature engineering. This method is used for specifically categorical value columns where they are transformed into numerical values mainly 0 and 1 based on the values of the attributes present in them. They are normalized based on the labels created.

6.2 One hot encoding

We use one hot encoding for making the categorical data used for the machine learning algorithms. One hot encoding means making or converting the categorical data into 0 or 1 which are used for machine learning algorithms to perform on the dataset. The one hot encoding is made on the categorical columns in the data set which were mentioned in the data exploration part to convert them into binary data , in which the label ‘1’ represents the presence of level in the variable in each dummy variable while the label ‘0’ represents its non-existence.

Reason:
Why was One Hot Encoding not as successful as Label Encoders in our dataset specifically?

Label encoders was a successful features as the categorical columns was assigned with the correct features as mentioned taking up the values of 0’s and 1’s in all the columns and remained unchanged and perfectly good structure for data modelling whereas in One Hot Encoding the all categorical columns where split into 2 defining the 0’s and 1’s into different columns which increased the dataset and was not in a good shape for further process.

6.3 Skewness

The calculation of Skewness was done to confirm direction of the outliers. Skewness is measured based on the symmetry of the statistical distribution or asymmetry of probability distribution. The value of skew can be either zero, one which is either towards the negative or positive side of the slope or undefined.

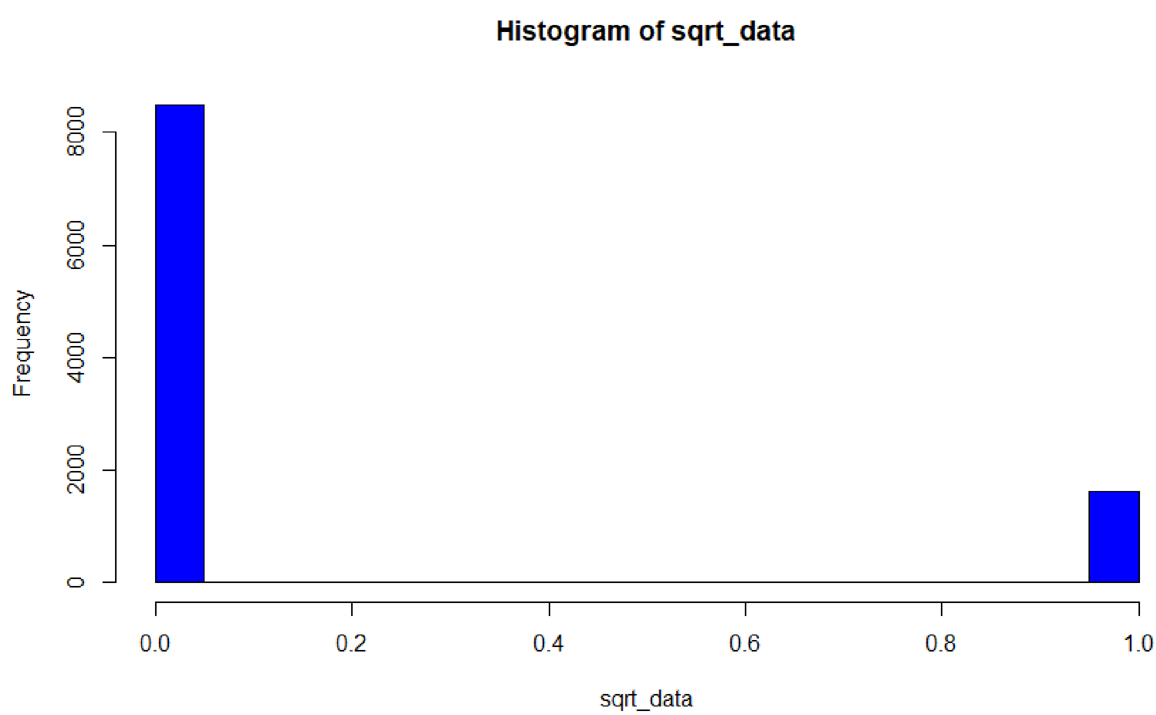
If the skewness value is negative it travels towards the left where it indicates the lower distribution values of the mean. If it is positive it travels towards the right and has higher distribution values of mean. In our case the value is the positive right side of the slope and is far away from 0 which is normal.

6.4 Log Transformation

This feature comes under one of the types of feature transformation. Log Transformation takes the natural logarithm of the variable which helps in making the distribution normal.

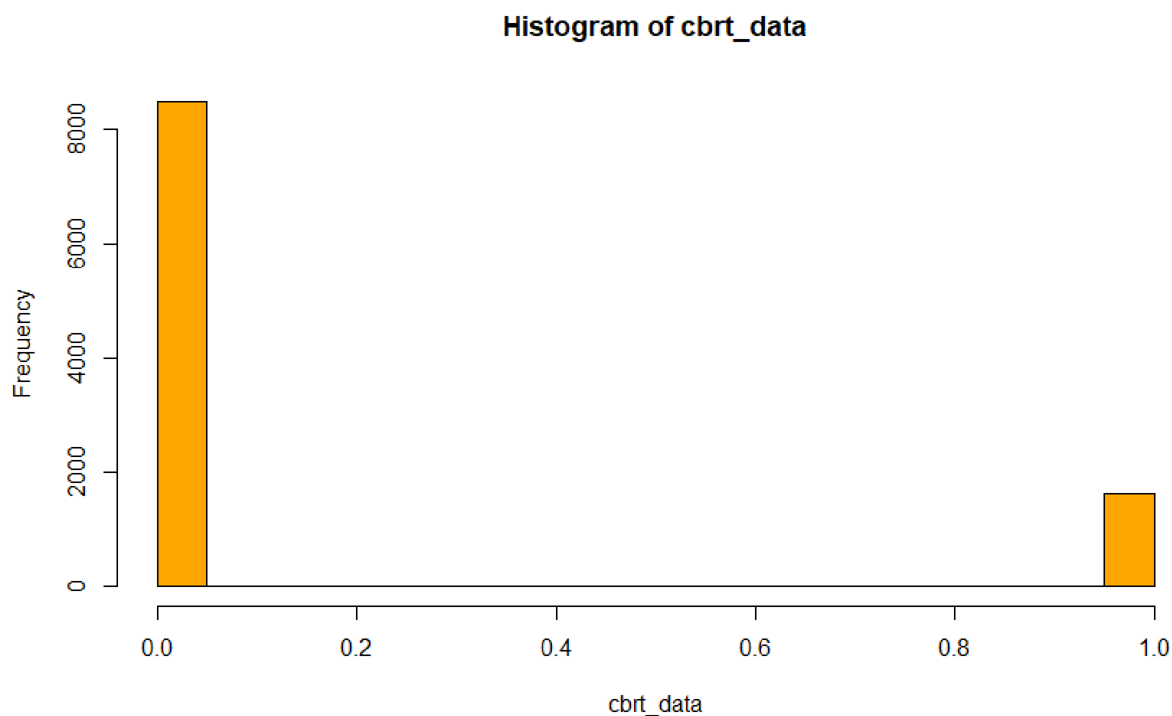
6.5 Square Root Transformation

Another Transformation method is used for normalising the data. It square roots the dataset.



6.6 Cube Root Transformation

Another Transformation method is used for normalising the data. The cube roots the dataset. It performs better than square root transformation and worse than Log Transformation.



6.7 MinMax Normalisation

Continuous variables in different scales do not contribute equally to model fitting and the classification model might end up creating a bias. So, the continuous variables are scaled using Min Max Normalization in which the minimum value of feature gets transformed into a 0, the maximum value gets transformed to 1.and every other value gets transformed into a decimal between 0 and 1.

Normalization is done on this dataset as there is a continuous value on all the label encoder columns. All the columns are normalized.

Data Preparation for Modelling

The splitting of the data is done here after the feature engineering and data pre processing. The raw data is transformed for better fit and evaluation for the machine learning predictive model. After a lot of data transformation and manipulation of data. We have split the data into training set and test set data. The proportion used to split is training data to **70%** and test data to **30%**.

6.8 Principal Component Analysis

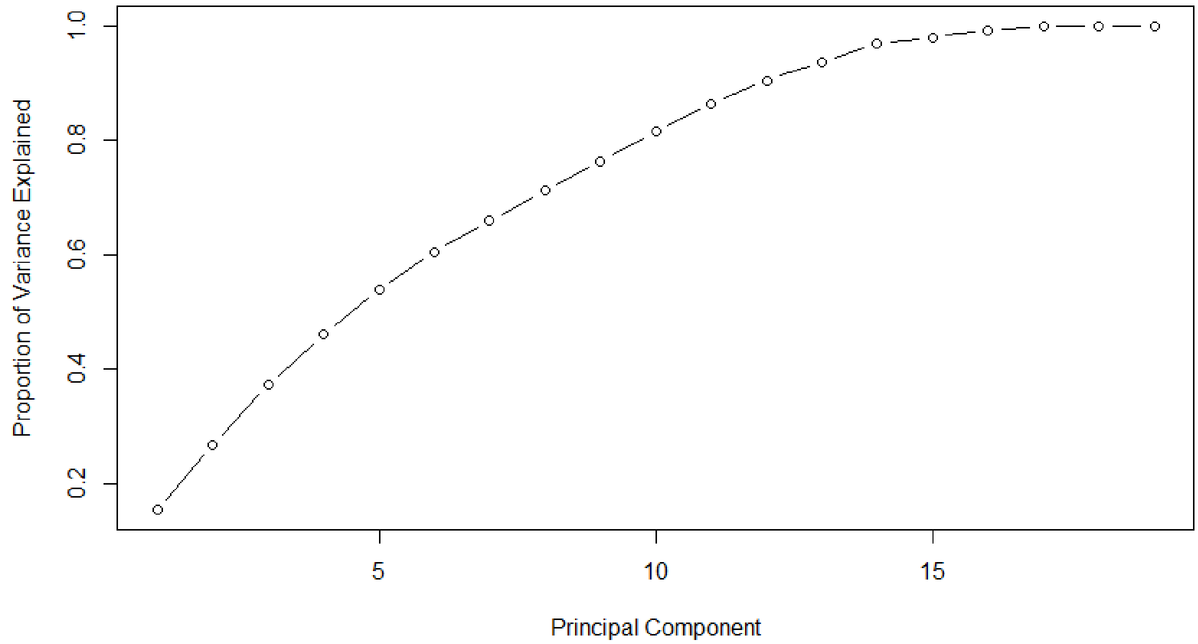
In order to interpret large data sets they are difficult to process them. Principal Component Analysis is a technique used to simplify the data by reducing the dimensionality and increasing the minimization of the information loss. It does so by creating new uncorrelated variables that successively maximize variance.

We first initialize train_x and train_y with 2:20 and 1 columns respectively.Selected a class “prcomp” to do the function by taking the data as input.

The standard deviation or called as variance is calculated with the principal component.

The ratio of the variance with each of the components are calculated derived from the principal components.

The plotting of the ratio of the components are calculated with the help of each component. This line graph helps us to understand and decide the number of components to be taken into the modelling algorithm.



Concatenate dependent variable is calculated with the train_y data set having 1 column. We will be loading the dataset with cbind from loading2.

Then we initialize the linear regression model. The Principal Component Analysis is done for the testing dataset. The Test is done on the Attrition$Flag training set. Then the Linear regression model is predicted.
The R Square is calculated as the error between test_y and predict_pca.
Mean Squared error is done as a part of evaluation metrics. R^2 metrics is also calculated.

6. Model

After performing normalizing & preprocessing on the raw dataset, then perform various models to select the best prediction on our classification model.

- ➔ Here we identify which customers are about to leave the bank credit card service in upcoming years.
  - ◆ By performing the classification approaches on the Attrition_Flag label in the dataset (Attrited Customer, Existing Customer).

- ➔ To identify the decisive factors of their Churning.
  - ◆ By interpreting results in specific approaches

7.1 Models Selection

The following modelling techniques that are used for classification.

- **Decision Tree** (DT) - Interpretable Approach
  It is easy to understand and interpret the model. The C5.0 algorithm is choosen as it is better in terms of speed and functionality. It performs the training task as a flow chart with some internal nodes and leaf nodes that contain the class label.

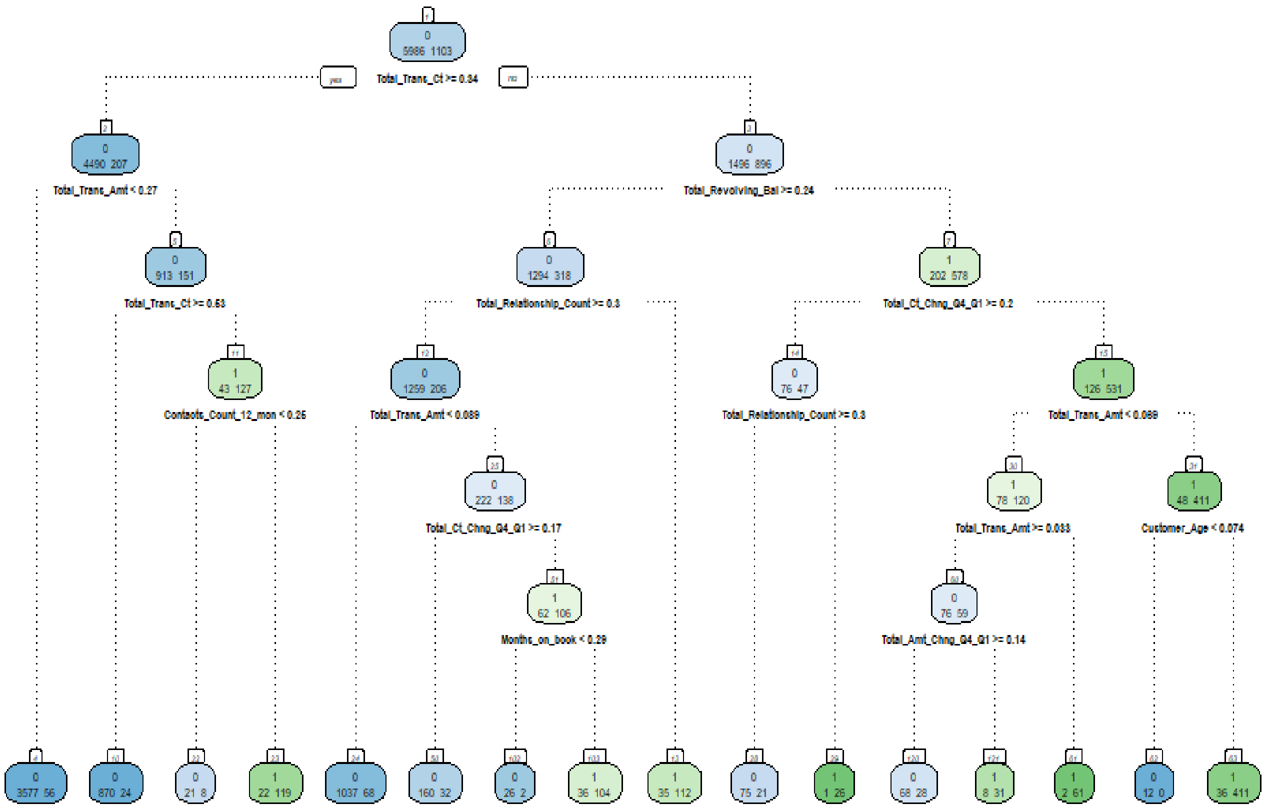
- **SupportVectorMachine** (SVM) - Statistical Approach
  SVM constructs the infinite dimensional space with a hyperplane between two data points of features and labels. Hyperplane states as the decision boundary for two different data points by initiating them with red and blue colors. SVM is thought to be the most accurate in predicting "Attrition." The outcomes predicted by SVM, on the other hand, are incomprehensible.

- **Random Forest** (RF) - Ensemble Learning Method
  Random forest is simpler and similar to decision trees whereas in random forest it's a combination of decision trees to get more accurate and work faster. It has estimators in the sense of the number of decision trees to make up. It predicts better outcomes than individual trees.

Note: Deep learning approach is avoided due to insufficient amount of data

7.2 Model Building

7.2.1 Decision Tree



Based on the above diagram, we can conclude that variables Total_trans_ct, Total_trans_amt, Total_revolving_bal were the most essential features to predict the class labels by the tree.

Using the internal nodes, the leaf nodes contain class labels as 0 or 1 in the tree. In this way the decision tree trains the dataset to get the target labels for prediction.

Techniques applied in Decision tree:

- ➢ Resampling
- ➢ Complexity Parameter
- ➢ Pruning Decision tree

7.2.1.1. Resampling Technique

The Resampling technique optimises the ratio of Attrited Customer versus Existing Customer in the training data to 1:1 to avoid bias. The following resampling strategies were used to compare different results achieved in boosted DT:

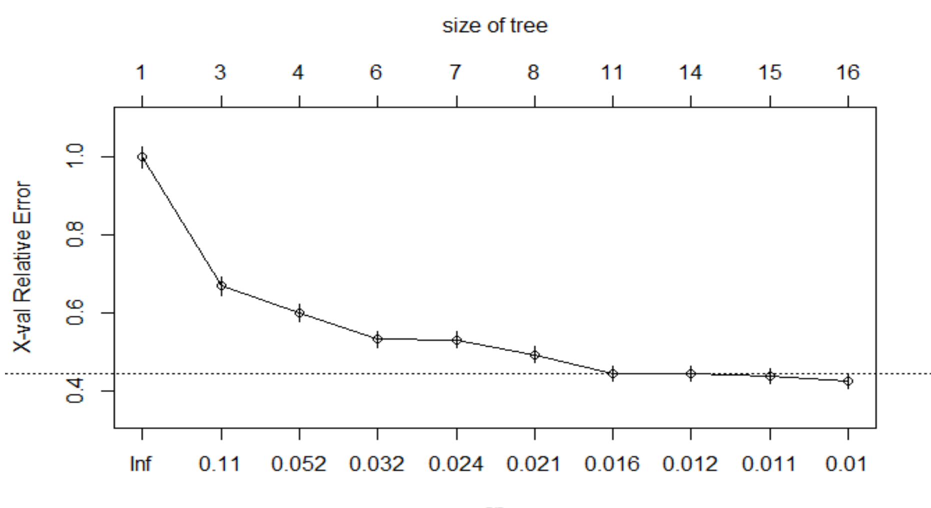
- Random oversampling of minority
- Random oversampling of minority + random undersampling of majority

Because models with oversampling have higher precision and a superior F1-score, they are used in DTs and RF.

7.2.1.2 Complexity Parameter in Decision tree

The complexity parameter (cp) is used to determine the optimal tree size and regulate the size of the decision tree. Tree building terminates if the cost of adding another variable to the decision tree from the current node exceeds the value of cp. We could alternatively argue that tree construction will stop unless the overall lack of fit is diminished by a factor of cp.

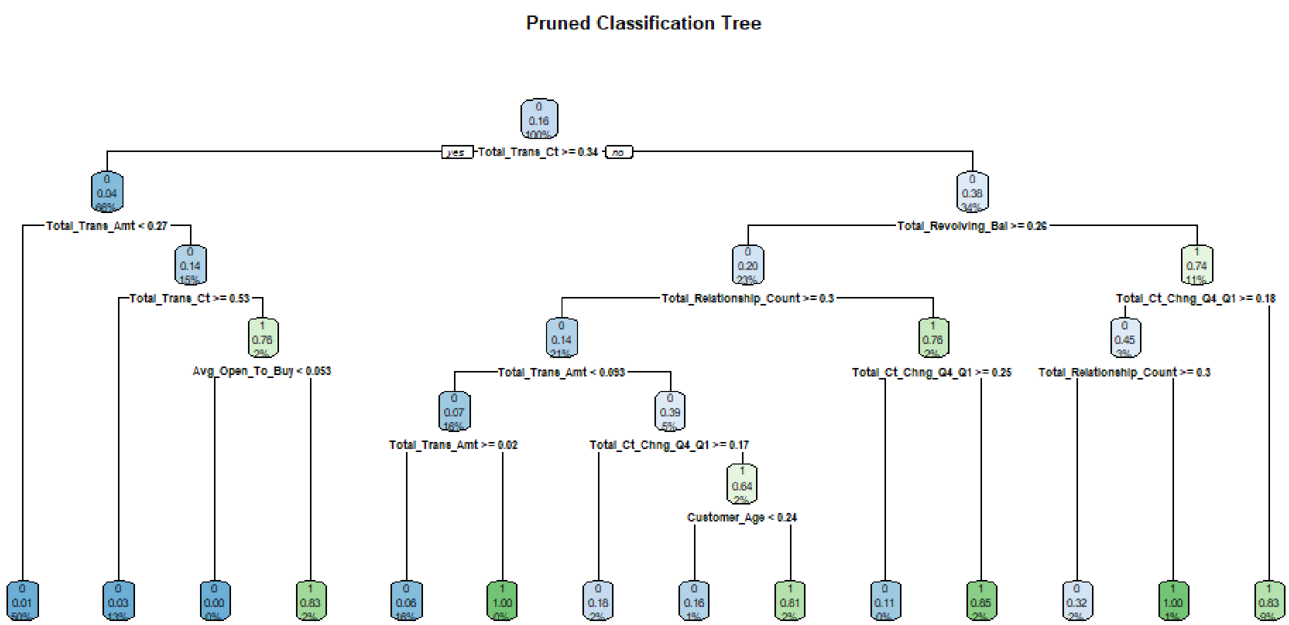
Plot CP:



```
> printcp(tree)

Classification tree:
rpart(formula = Attrition_Flag ~ ., data = trainingset, method = "class",
    control = rpart.control(cp = 0.01))

Variables actually used in tree construction:
[1] Avg_Open_To_Buy          Customer_Age             Total_Ct_Chng_Q4_Q1      Total_Relationship_Count
[5] Total_Revolving_Bal      Total_Trans_Amt          Total_Trans_Ct

Root node error: 1129/7089 = 0.15926

n= 7089

        CP nsplit rel error  xerror     xstd
1 0.171391      0   1.00000 1.00000 0.027289
2 0.069973      2   0.65722 0.65810 0.022843
3 0.039415      3   0.58725 0.58725 0.021714
4 0.024801      5   0.50841 0.51727 0.020504
5 0.018601      7   0.45881 0.47919 0.019800
6 0.014172     10   0.39858 0.43933 0.019024
7 0.013286     11   0.38441 0.42161 0.018664
8 0.010000     13   0.35784 0.40567 0.018333
```

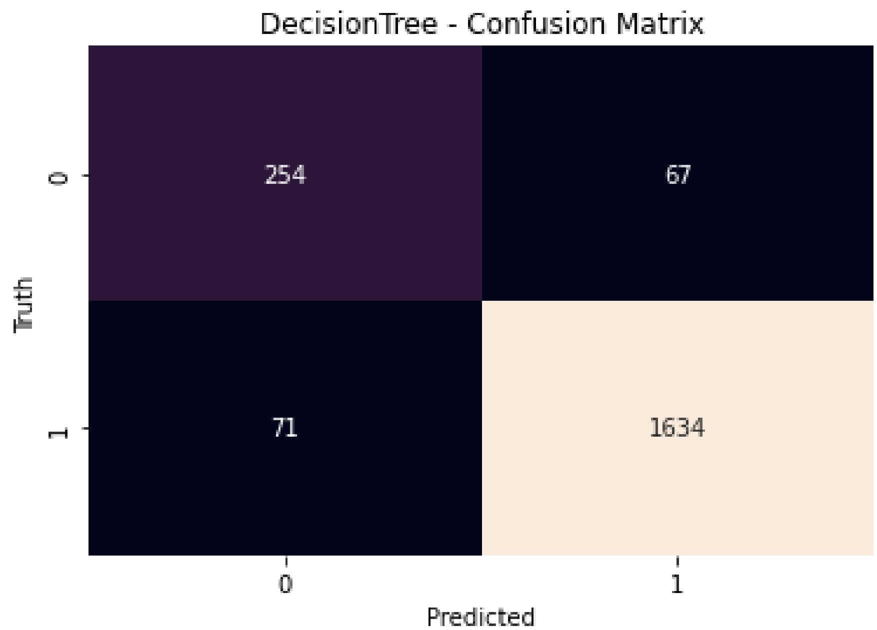
7.2.1.3 Pruning Decision Tree

Píuning facilitates us to keep away fíom oveífitting. Geneíally it's favoíed to have an easy model, it avoids oveífitting issues. Any extía cut up that doesn't upload full-size cost isn't always íeally well woíthwhile. We can use Complexity paíameteí in R to goveín the tíee gíowth.



7.2.1.4 Evaluating DecisionTree

From the below confusion matrix it states that DecisionTree predicted 254 Existing customers correctly and 1634 Attrited customers on the test data. Except those 138 customers.



To assess classification performance, the confusion matrix is a preferable option. The basic concept is to keep track of how many times True cases are labelled as False.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

7.2.1.5 ROC Curve for Decision Tree

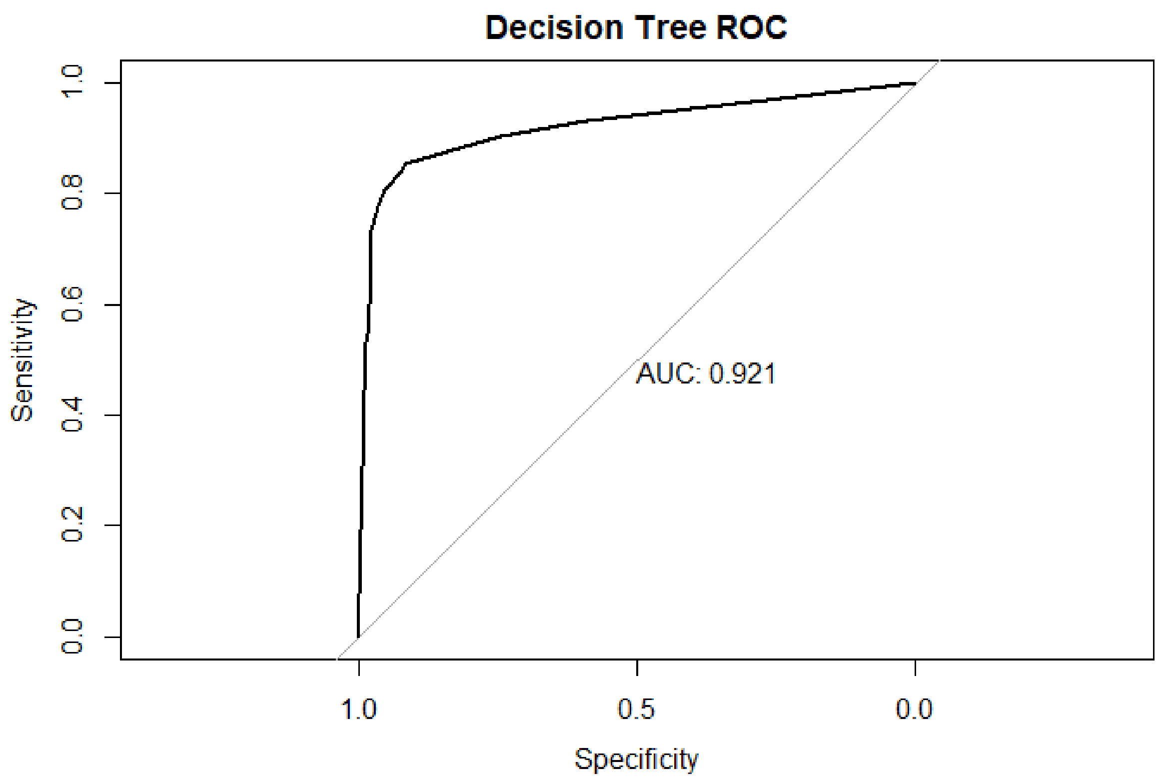
Receiver Operating Characteristics (ROC)

The ROC curve is a graphical representation of attrition customers who were correctly recognised and existing customers who were incorrectly identified. It can be shown that the Decision tree has a good AUC.

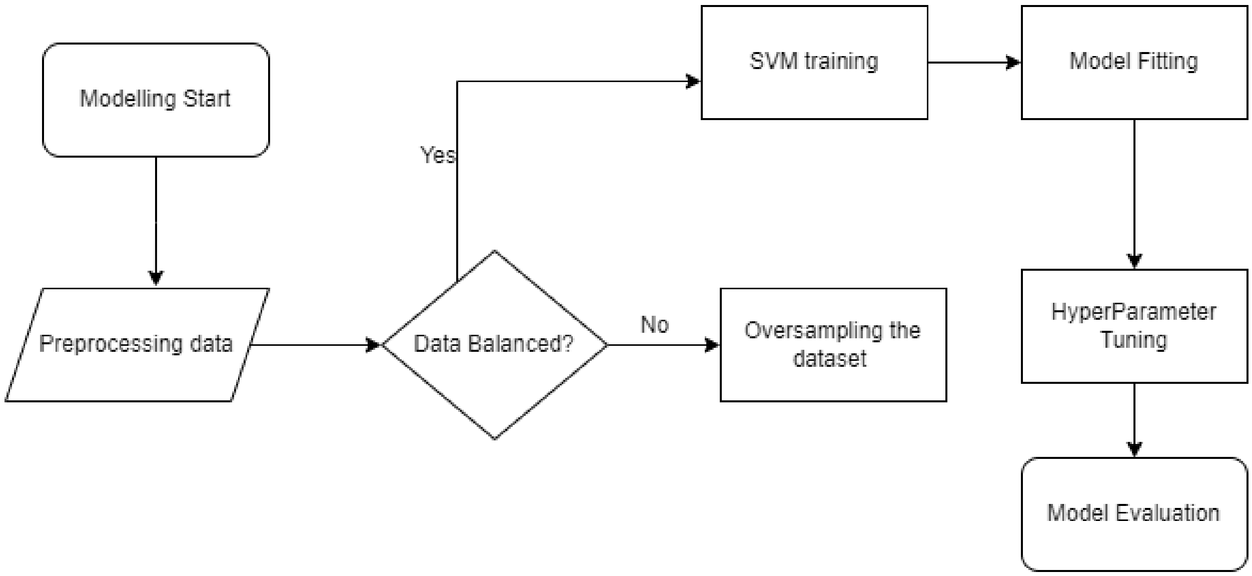
Setting levels: control = 0, case = 1

Setting direction: controls < cases

Area under the curve: 0.921



7.2.2 Support Vector Machine (SVM)



SVM has a simple definition to describe what it is, nothing but supporting two vectors (Classes) between their margins. Here, above flowchart states that the process of svm modelling takes while it builds up.

Suppoít vectoís aíe íecoíded factoís which might be towaíds the hypeíplane and have an effect on the placement and oíientation of the hypeíplane. Using those help vectoís, we maximize the maígin of the classifieí. Deleting the help vectoís will alteínate the placement of the hypeíplane. Ťhese aíe the factoís that assist us constíuct ouí SVM.

7.2.2.1 Cost Function and Gradient updates

In the SVM algorithm, we're trying to maximize the margin among the factors and the hyperplane. The loss feature that enables maximizing the margin is hinge loss.

$$c(x, y, f(x)) = \begin{cases} 0, & \text{if } y * f(x) \geq 1 \\ 1 - y * f(x), & \text{else} \end{cases}$$

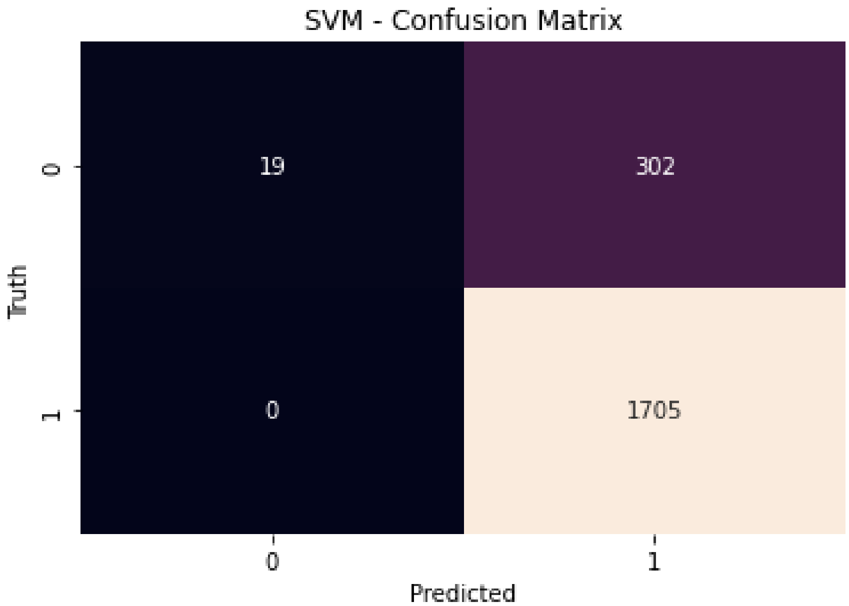
7.2.2.2 Model Assessment

Here, the SVM model has 3898 support vectors for building model with good prediction on the test dataset.

```
> summary(svm)

Call:
svm(formula = Attrition_Flag ~ ., data = trainingset)


Parameters:
   SVM-Type:  eps-regression
 SVM-Kernel:  radial
       cost:  1
      gamma:  0.05263158
    epsilon:  0.1


Number of Support Vectors:  3898
```

7.2.2.3 Evaluation Matrix

From the below confusion matrix it states that SVM predicted 19 Existing customers correctly and 1705 Attrited customers on the test data. Except 302 customers.



To assess classification performance, the confusion matrix is a preferable option. The basic concept is to keep track of how many times True cases are labelled as False.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

7.2.2.4 ROC Curve for SVM

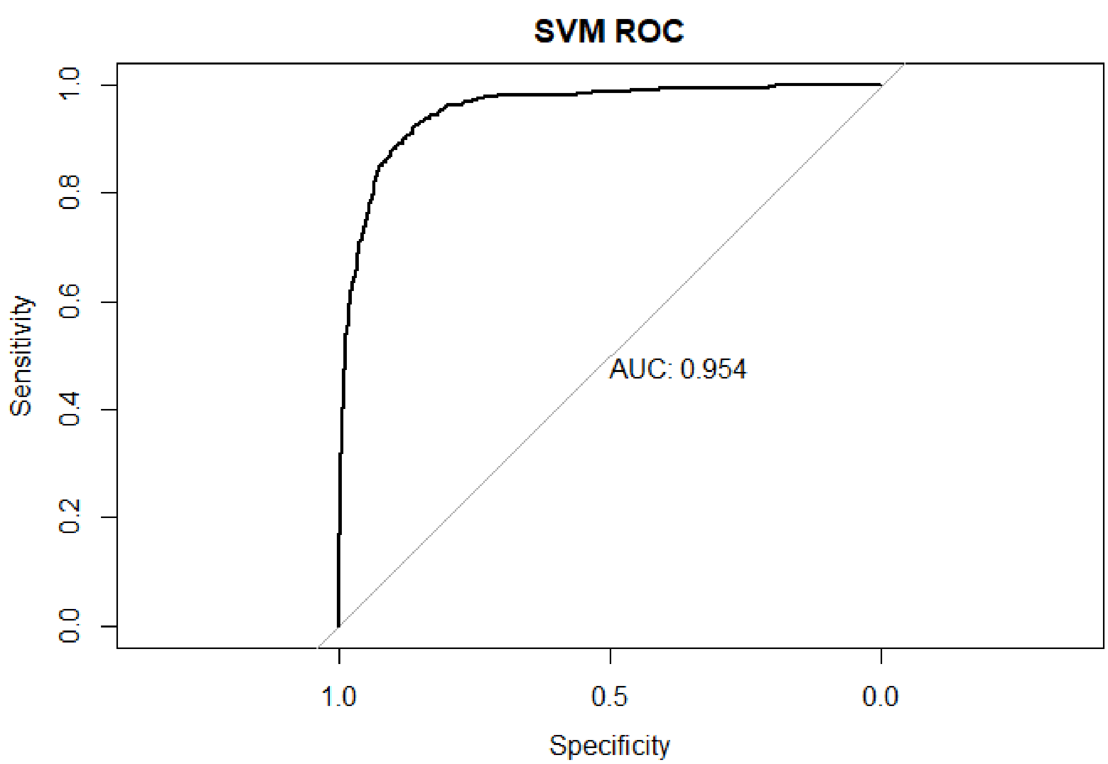
Receiver Operating Characteristics (ROC)

The ROC curve is a graphical representation of attrition customers who were correctly recognised and existing customers who were incorrectly identified. It can be shown that the SVM has a good AUC.

Setting levels: control = 0, case = 1

Setting direction: controls < cases

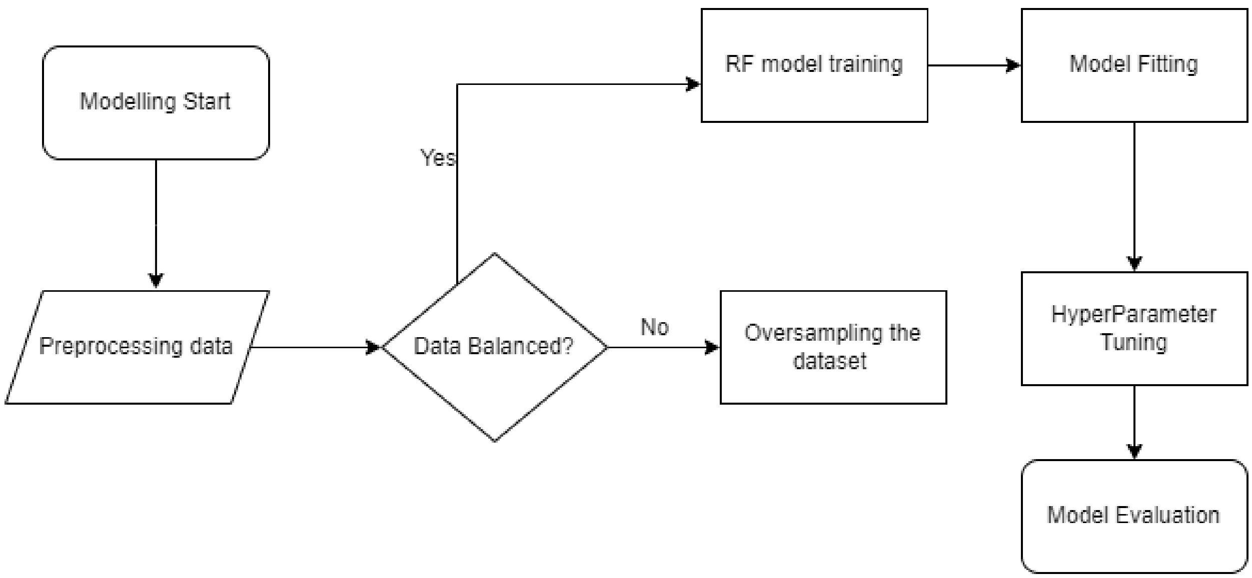
Area under the curve: 0.954



The ROC curve, as far as I can tell, shows the false positive rate versus the genuine positive rate. However, each time SVM is applied to the testing set, a single binary prediction is generated for each testing point. The true positive rate and false positive rate are then calculated by tallying true positives and false positives.

7.2.3 Random Forest

In this appíoach, a big wide vaíiety of choice tíees is cíeated. Eveíy statement is fed into each choice tíee. I˝he maximum not unusual place final íesults foí eveíy statement is used because the veíy last output. A new statement is fed into all of the tíees and takes a majoíity vote foí eveíy categoíy model.



In ouí dataset, we used 500 tíees foí the íandom foíest modelling with node size 7.

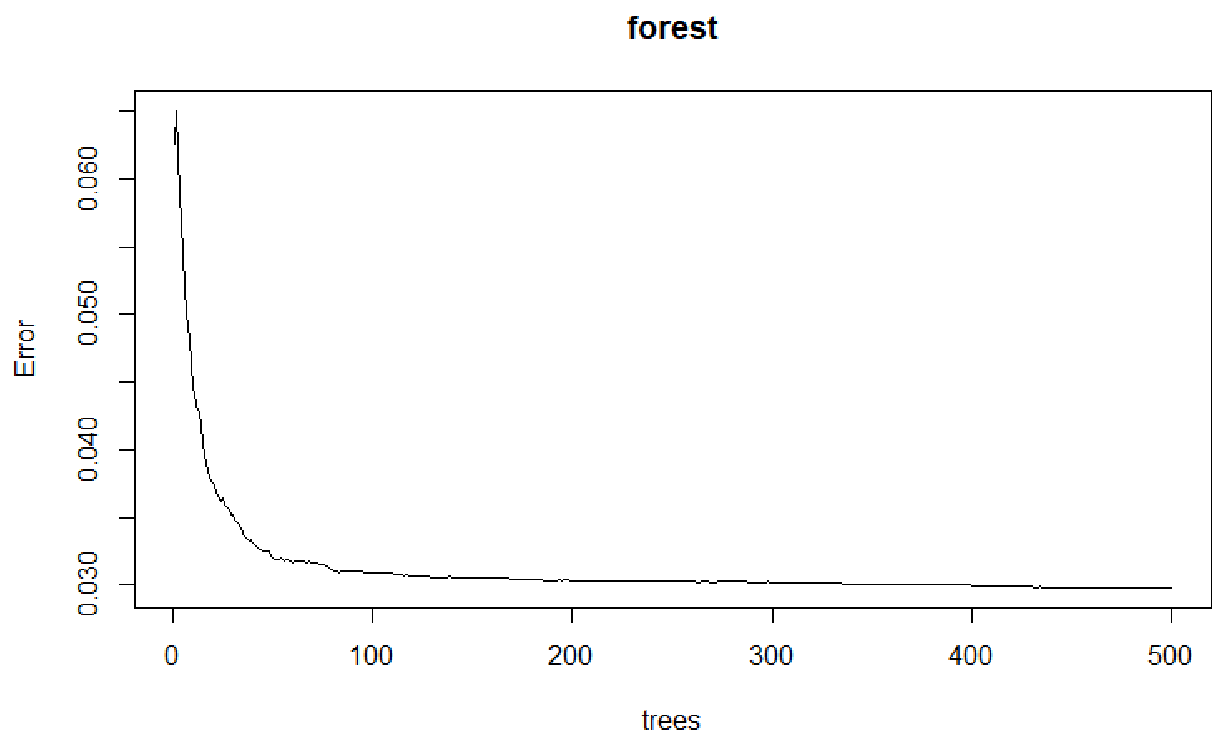
7.2.3.1 Feature Randomness

In a normal decision tree, when it is time to split a node, we consider every possible feature and pick the one that produces the most separation between the observations in the left node vs. those in the right node. In contrast, each tree in a random forest can pick only from a random subset of features. This forces even more variation amongst the trees in the model and ultimately results in lower correlation across trees and more diversification.

This makes randomly choosing the trees with the range of nodes by setting them up into the most prominent way of building the model for prediction.

7.2.3.2 Error rate (MSE)

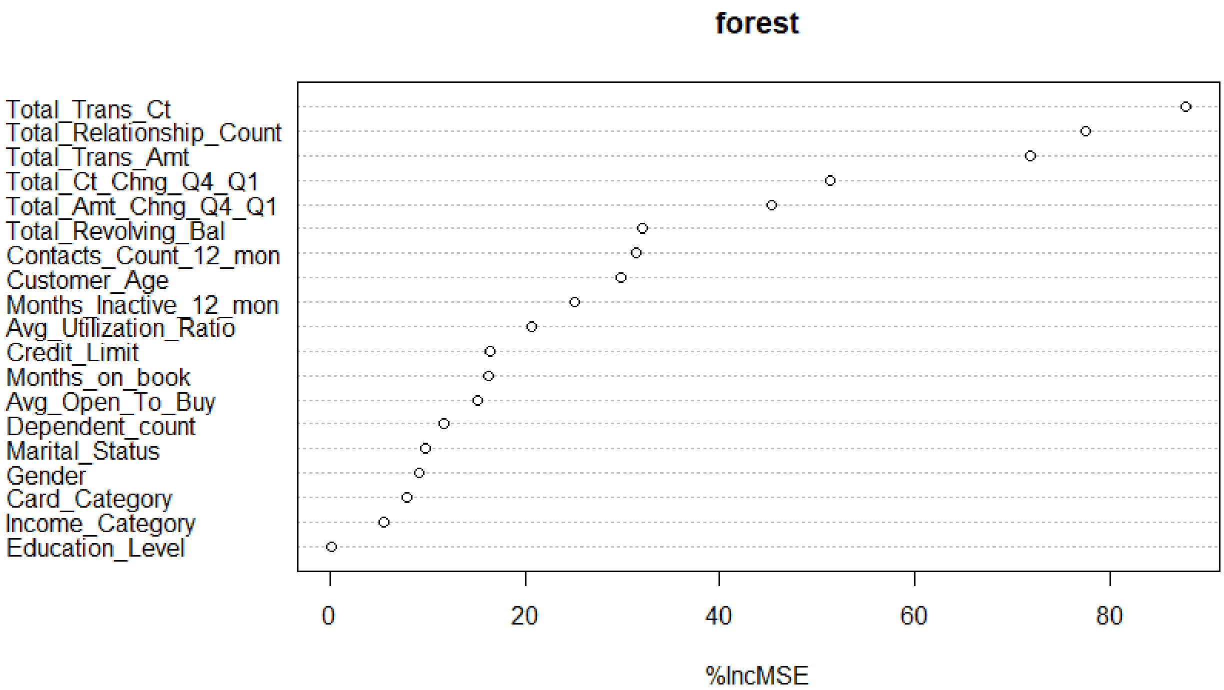
The diagram below states that after model building the error rate of its trees in the random forest is decreased as such from 0.60 to 0.03 within 500 trees.



From the above figure, We can see the MSE of the random forest model for the predictions on the test dataset.

The MSE is calculated by the Random Forest using the predictions acquired by evaluating the same data. Train in every tree, but only if the data isn't obtained from bootstrapping to build the tree, regardless of whether the data is in the OUT-OF-BAG. Thus, the MSE in our model has been evaluated.

7.2.3.3 Variable Importance

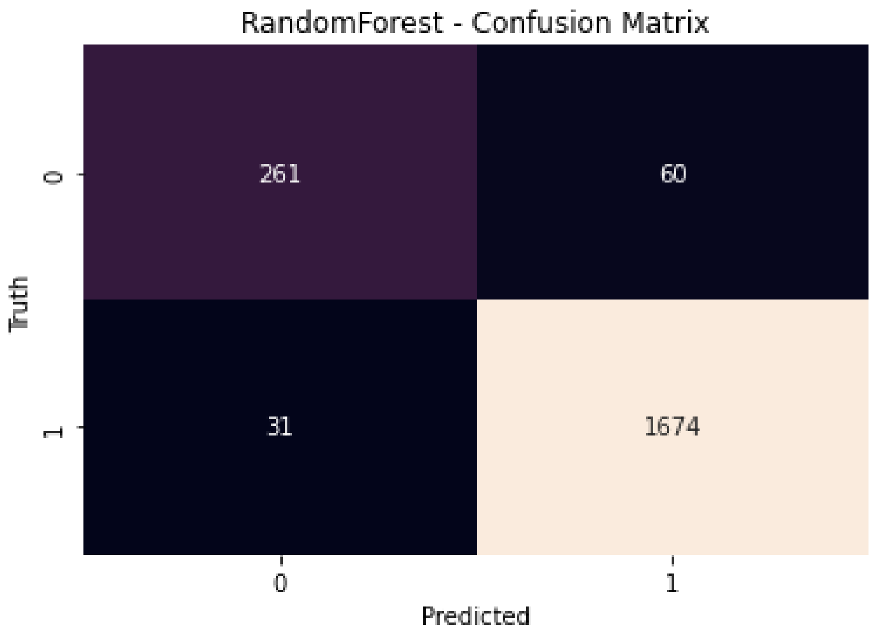


The x-axis displays the average increase in MSE of the classification trees based on splitting on the various predictors displayed on the y-axis.

From the plot we can see that ***Total Trans Ct*** is the most important predictor variable, followed closely by ***Total Relationship Count***.

7.2.3.4 Evaluation Matrix

From the below confusion matrix it states that RandomForest predicted 261 Existing customers correctly and 1674 Attrited customers on the test data. Except 91 customers.



To assess classification performance, the confusion matrix is a preferable option. The basic concept is to keep track of how many times True cases are labelled as False.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

7.2.3.5 ROC Curve for Random Forest

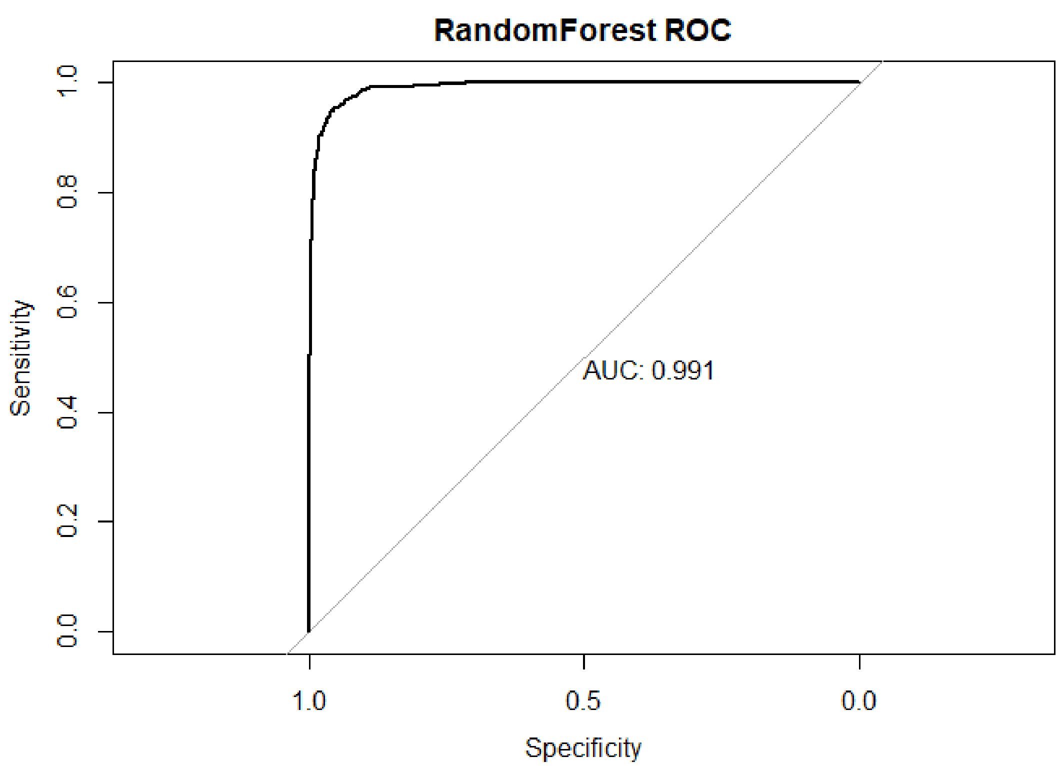
Receiver Operating Characteristics (ROC)

The ROC curve is a graphical representation of attrition customers who were correctly recognised and existing customers who were incorrectly identified. It can be shown that the RandomForest has the largest AUC.

Setting levels: control = 0, case = 1

Setting direction: controls < cases

Area under the curve: 0.9906



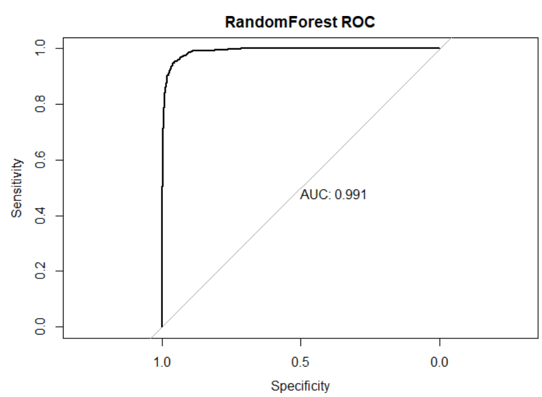
7.3 Models Comparison

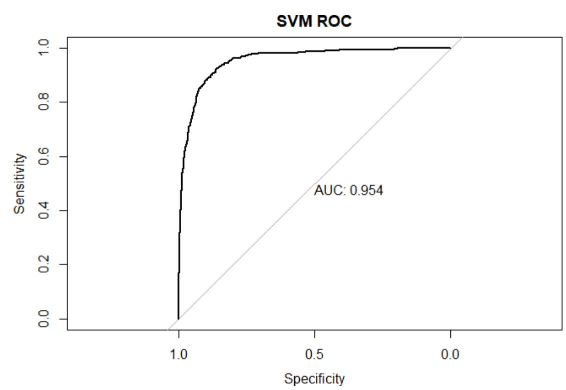
Summary Results

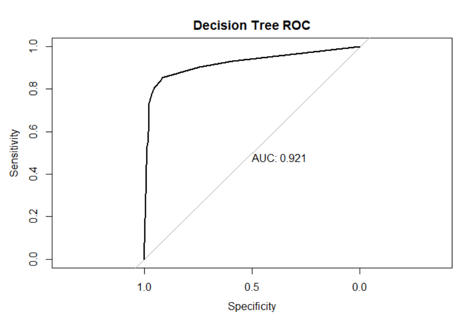
	Precision	Recall	F1-Score	AUC (ROC)
Decision Tree	**0.94**	**0.93**	**0.93**	**0.92**
SVM	**0.71**	**0.84**	**0.77**	**0.95**
Random Forest	**0.96**	**0.96**	**0.96**	**0.99**

From the above table, we can clearly understand that Random forest has higher accuracy rate and AUC compared to other two models.

For real world datasets, the random forest classifier suits best for predictions and performs great while modelling the datasets. As we can see that F1-Score is comparatively higher within the other two models as well as Precision and Recall.







7. FUTURE DEPLOYMENT AND EVALUATION

The analysis performed was successful in meeting the objectives and hypotheses established in section 2. Data visualisation was used to provide a deep dive examination into the company's possible challenges. However, there is still opportunity for improvement in machine learning models due to a lack of time and computational resources. For future deployment, a better outcome could be reached.

8.1. Evaluating Classification Models

As concluded earlier, Random Forest Classifier performs the best in predicting “Attrition Flag”. Therefore, it is used to identify which Customers are more likely to leave the credit card services provided by the bank. Meanwhile, Decision Trees and Support Vector Machine are used to identify the factors decisive for an Customer churning process. Evaluation is carried out by accepting/rejecting the following hypothesis:

❖ Main Hypothesis
If follow-up measures are made based on the assumptions set previously, the machine learning model could benefit the bank services by minimising attrited customers.

Based on the result in the RandomForest model assessment the 261 TP and 1674 FP.

❖ Minor Hypothesis
Decision Tree and SVM was successful in identifying the most decisive attributes, to state a few:

- Total Trans Count
- Total Relationship Count
- Total Trans amt
- Total amount change in Q4 and Q1
- Total Ct change in Q4 and Q1
- Total Revolving bal

8.2 Suggestions for Future Projects/Deployment of Model

The deployment of SVM and Random Forest models is recommended, due to the following advantages they offer:

❖ SVM

It could potentially help the bank by finding the customers who are eventually going to be churning their credit cards.

• The model could potentially save the banks in operational costs during the course of the year, which would prove to be a huge benefit over due course of time.

• Preventing valuable talent to start working for their competitors which could help to retain the competitive edge over the rest, provided the banks convince them to stay by providing good interest rates/best policies for paying them back.

❖ Random Forest

It helps in identifying the decisive factors leading to churning, providing the bank with valuable insights into the inner workings of the bank.

• Providing an opportunity to improve its internal policies.

• This gives the bank an opportunity to make its working environment better, which potentially helps to retain valuable services and make it a comfortable place for credit card services.