BFS CAPSTONE PROJECT

Final - Submission

Ravi Shekhar Rai

Pavan ML

Subhanshu Rathi

Archana Mishra

Capstone Project Overview

Following is the bird eye view of this presentation

- Abstract
- Data Understanding
- Data Cleansing & Preparation
- Exploratory Data Analysis
- WOE & Information Value
- Model Building & Evaluation
- Conclusion

Abstract

- **BUSINESS UNDERSTANDING**

- CredX is a leading credit card provider receiving huge volume of applications every year.
- In recent times unfortunately CredX is experiencing an increase in credit loss due to not reaching the right customers during acquisition.

- **PROBLEM STATEMENT**

- For a provider like CredX it becomes very important to acquire right customers in order to increase their profitability by keeping their business costs in control.
- In this Project, We help CredX in exactly doing the same. We using our various predictive models help CredX acquire the right customers there by increasing their profits.
- In this process we also use our predictive models in determining the factors affecting the credit risk and creating the strategies to mitigate the acquisition risks.

Data Understanding

Data Source

Our data is mainly divided into two categories

- Demographic Data
- Credit Bureau Data

Demographic Data:

This is basically customer-level information on age, gender, income, marital status etc. given by customer himself via the credit card application

Credit Bureau Data:

Source of this data is via credit bureau and contains variables such as 'number of times 30 DPD or worse in last 3/6/12 months', 'outstanding balance', 'number of trades', etc.

Data Understanding

<u>Common Variable between the data sources</u>

"performance tag" variable is common in both the demographic and credit bureau data files which basically represents whether the applicant has gone 90 days past due or worse in the past 12-months after getting a credit card. In this case the customer will be treated as defaulted customer.

<u>Special Cases</u>

Case 1: All the fields in credit bureau data file are '0' and credit card utilization field is missing will be the case that says there is no hit in the credit bureau.

Case 2: Cases where the credit card utilization is simply missing says that the particular applicant doesn't hold any credit card.

Data Cleansing & Preparation

Data Observation

- Demographic Data – 71, 295 Records of 12 Variables

  Credit Bureau Data – 71, 295 Records of 19 Variables

- "Application ID" has been identified as a primary key in both the data sources
- No Duplicate Records found in Demographic data and Credit bureau data.
- 3 Discrepancy Records found for the same user in both data sources hence the discrepancy records are removed.
- 1577 Records has been observed with NA values in demographic data source.
- 3028 Records has been observed with NA values in credit bureau data source.
- 1425 Records from Demographic data and credit bureau data has been taken as rejected population as performance tag is NA (separate handling has been done for these rejected population)

NA value treatment of demographic data

Columns Name	Replaced column
Age	Age_Woe
Gender	Gender_Woe
Marital Status at the time of application	Marital Status at the time of application_Woe
No of dependents	No of dependents Woe
Income	Income_Woe
Education	Education_Woe
Profession	Profession_Woe
Type of residence	Type of residence_Woe
No of months in current residence	No of months in current residence_Woe
No of months in current company	No of months in current company_Woe

NA values have been replaced with WOE values

NA value treatment of credit bureau data

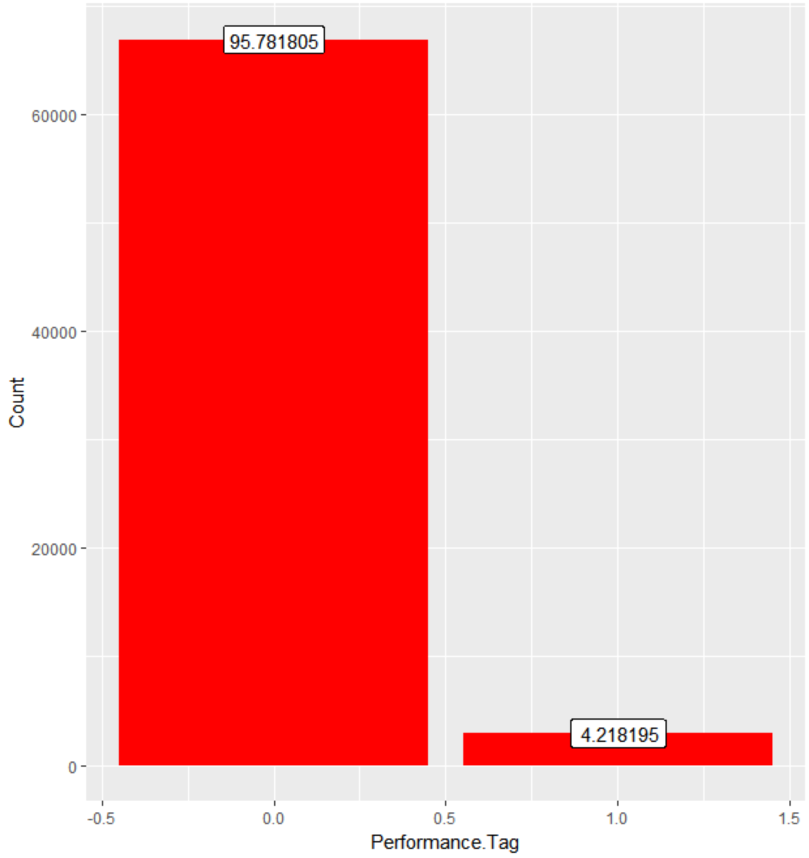
Columns Name	Replaced column
No.of.times.90.DPD.or.worse.in.last.6.months	No.of.times.90.DPD.or.worse.in.last.6.months_Woe
No.of.times.60.DPD.or.worse.in.last.6.months	No.of.times.60.DPD.or.worse.in.last.6.months_Woe
No.of.times.30.DPD.or.worse.in.last.6.months	No.of.times.30.DPD.or.worse.in.last.6.months_Woe
No.of.times.90.DPD.or.worse.in.last.12.months_Woe	No.of.times.90.DPD.or.worse.in.last.12.months_Woe
No.of.times.60.DPD.or.worse.in.last.12.months	No.of.times.60.DPD.or.worse.in.last.12.months_Woe
No.of.times.30.DPD.or.worse.in.last.12.months	No.of.times.30.DPD.or.worse.in.last.12.months_Woe
Avgas.CC.Utilization.in.last.12.months	Avgas.CC.Utilization.in.last.12.months_Woe
No.of.trades.opened.in.last.6.months	No.of.trades.opened.in.last.6.months_Woe
No.of.trades.opened.in.last.12.months	No.of.trades.opened.in.last.12.months_Woe

Columns Name	Replaced column
No.of.PL.trades.opened.in.last.6.months	No.of.PL.trades.opened.in.last.6.months_Woe
No.of.PL.trades.opened.in.last.12.months	No.of.PL.trades.opened.in.last.12.months_Woe
No.of.Inquiries.in.last.6.months..excluding.home...auto.loans	No.of.Inquiries.in.last.6.months..excluding.home...auto.loans._Woe
No.of.Inquiries.in.last.12.months..excluding.home...auto.loans	No.of.Inquiries.in.last.12.months..excluding.home...auto.loans._Woe
Presence.of.open.home.loan	Presence.of.open.home.loan_Woe
Outstanding.Balance	Outstanding.Balance_Woe
Presence.of.open.auto.loan	Presence.of.open.auto.loan_Woe
Total.No.of.Trades	Total.No.of.Trades_Woe

NA values have been replaced with WOE values

Exploratory Data Analysis – Univariate Analysis

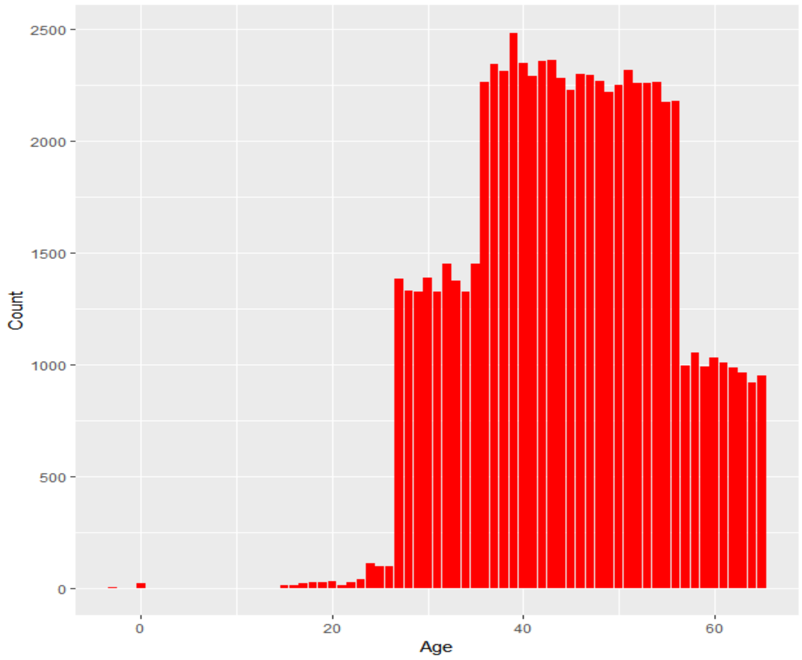
Demographic Data - Performance Tag

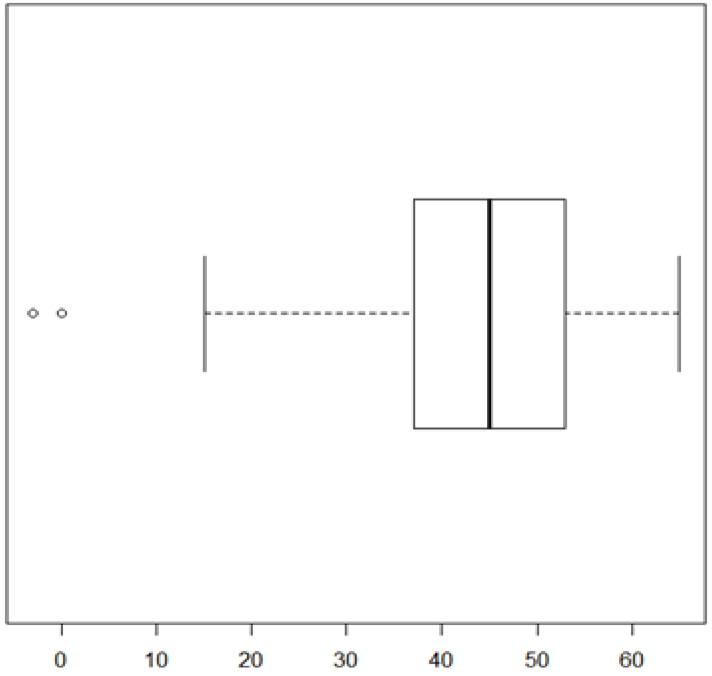


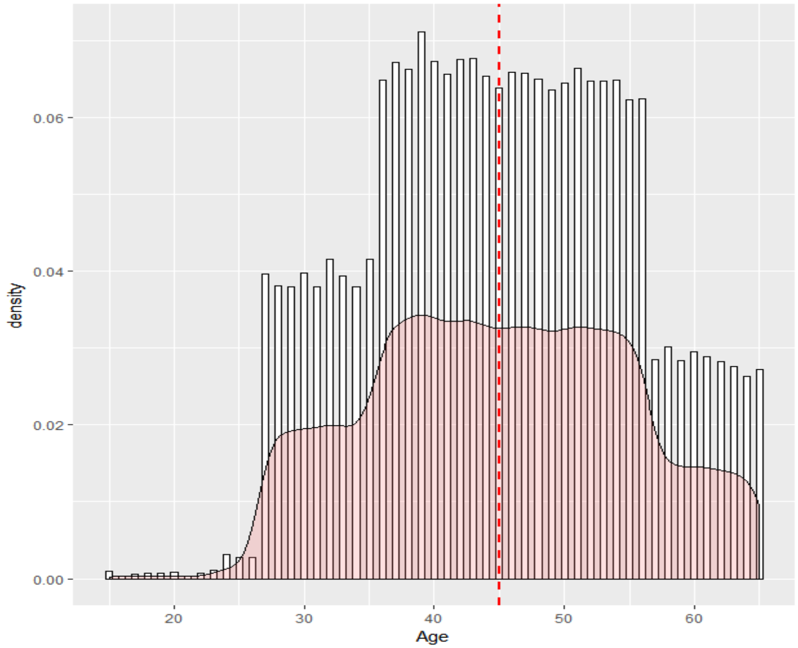
Observation :: Distribution of Performance Tag is Imbalanced where 1 represents Defaults and has share only 4.21 in the data.

Exploratory Data Analysis – Univariate Analysis

Demographic Data - Age



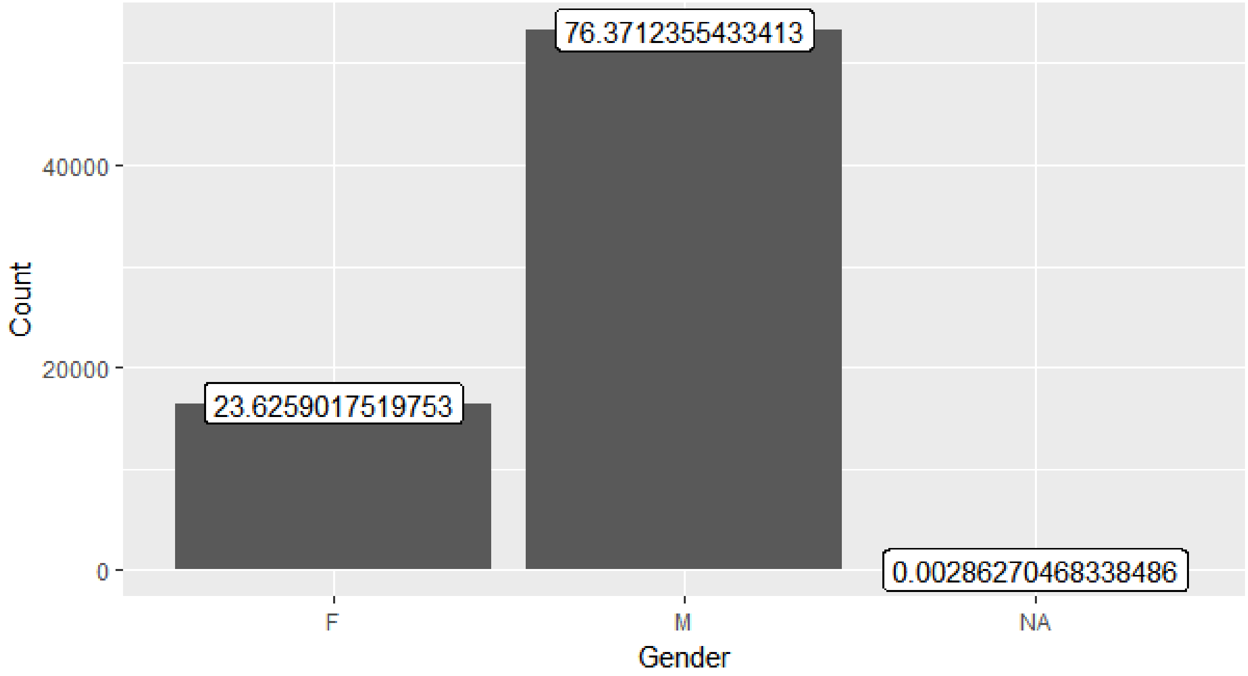


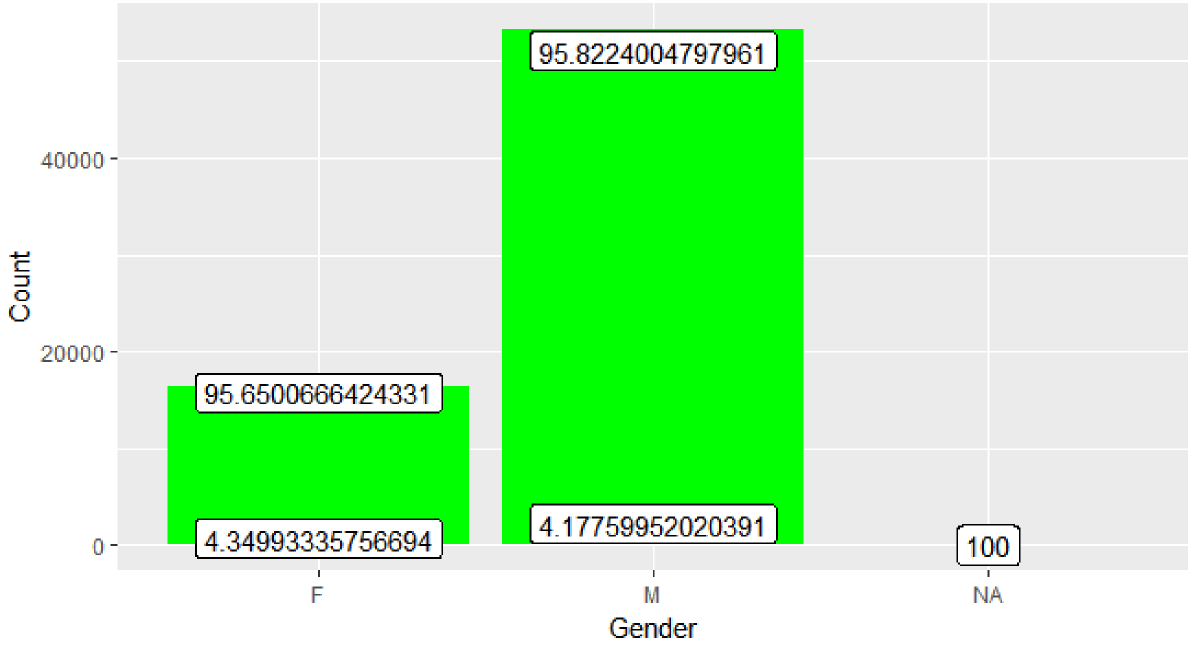


Replacing any ages less then 18 to 18 because 18 is the legal age to get credit card
Most of the users are in 30 to early 50 years age range
Some outliers (invalid age i.e. -3) values are present

Exploratory Data Analysis – Univariate Analysis

Demographic Data - Gender

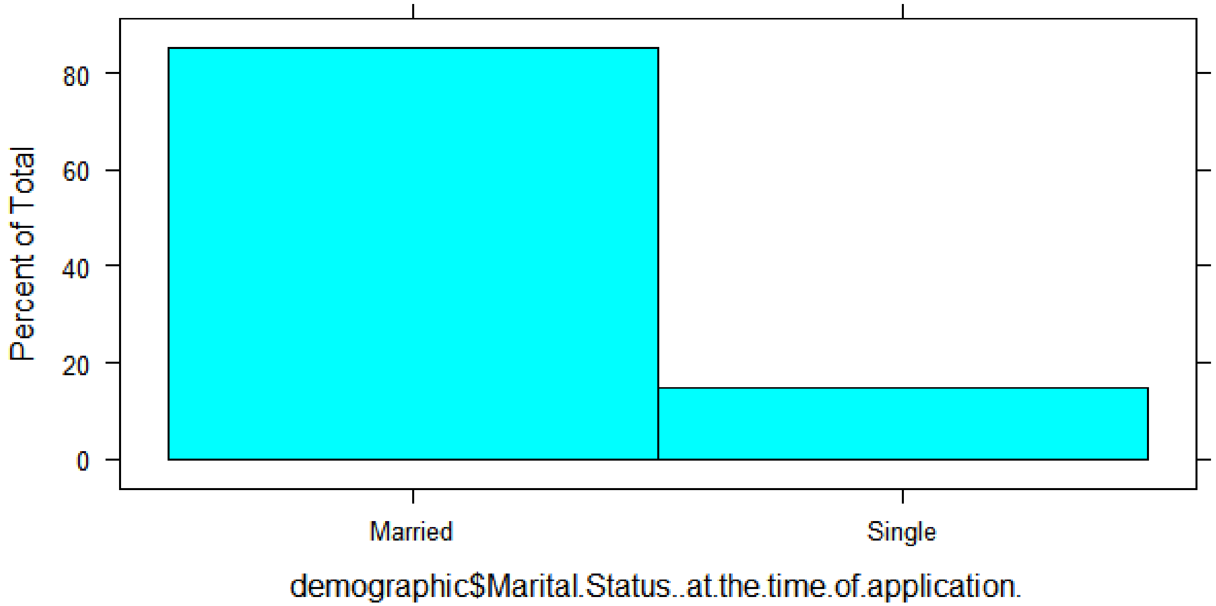


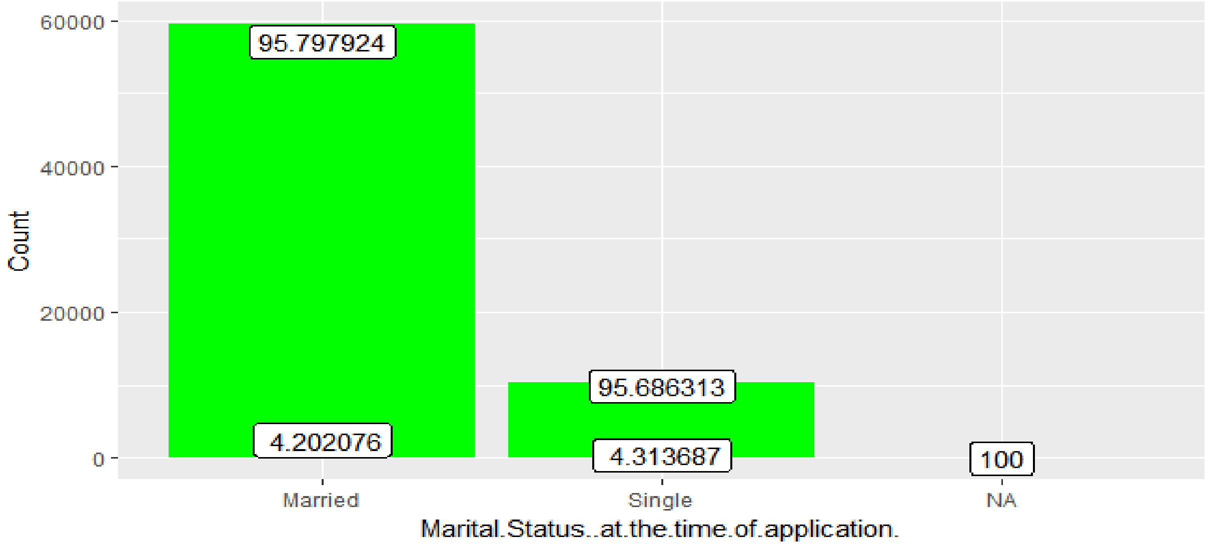


76.38 percentage of male constitutes the total population and remaining 23.63 percentage female constitutes the total population

Exploratory Data Analysis – Univariate Analysis

Demographic Data - Marital Status

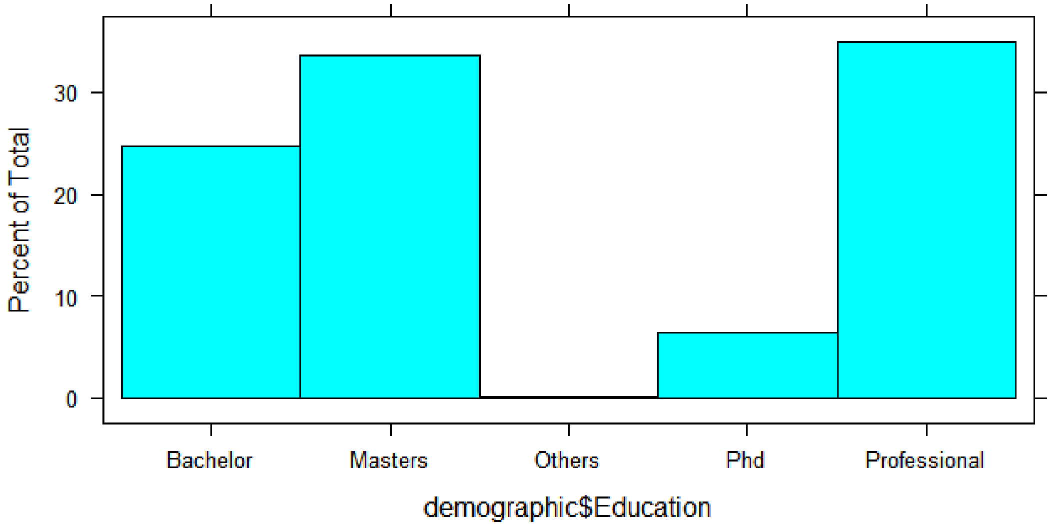


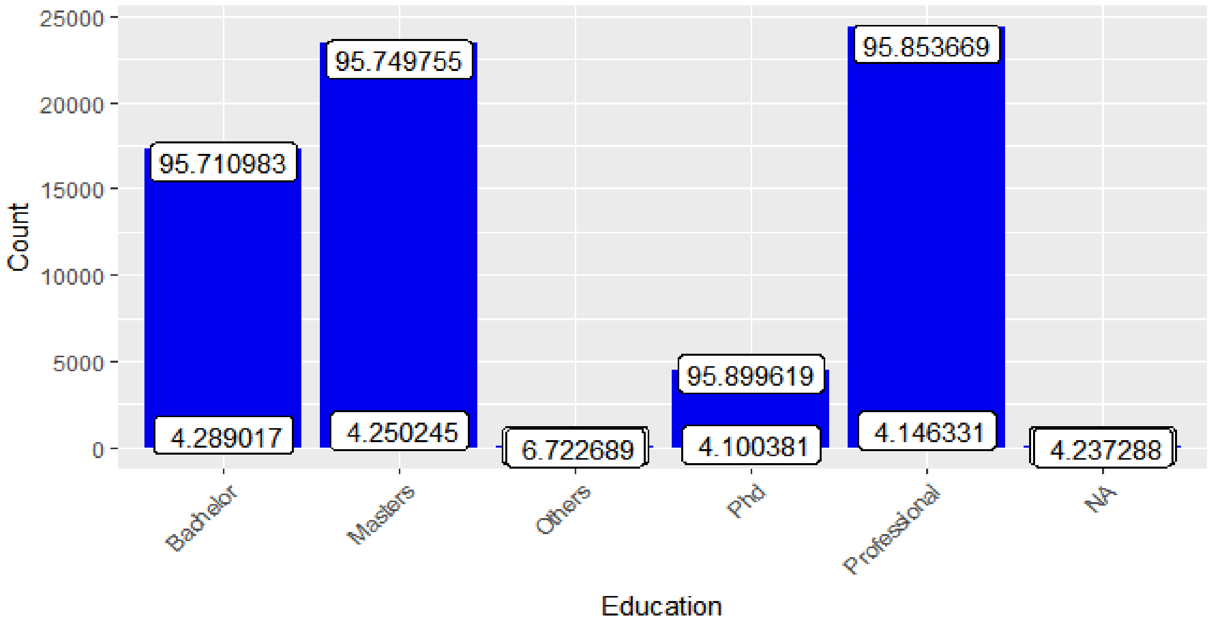


80 percentage of the total population constitutes of married people and remaining 20 percent single at time of applications

Exploratory Data Analysis – Univariate Analysis

Demographic Data - Education

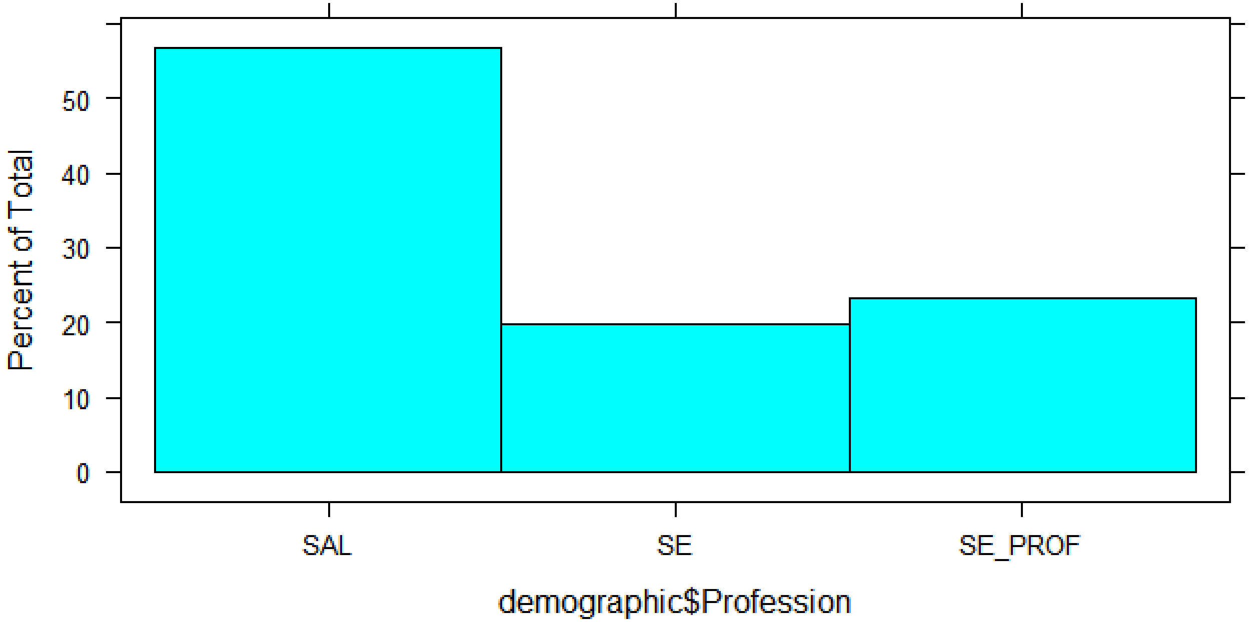


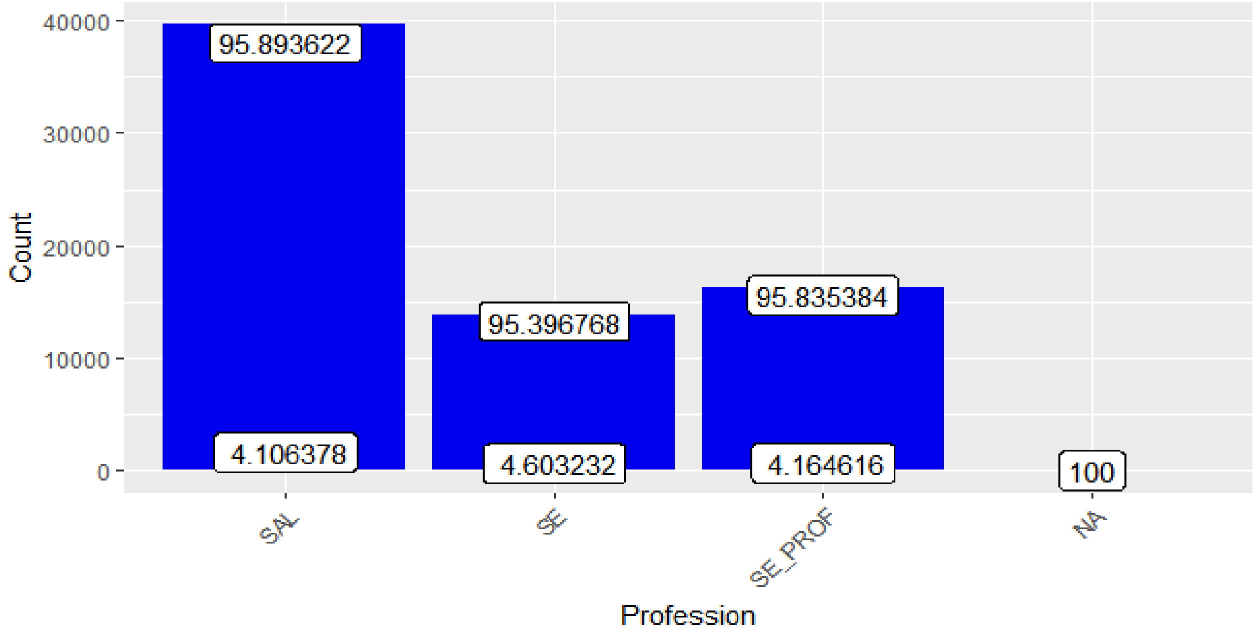


People who have done masters and belongs to professional background constitutes the majority of applications

Exploratory Data Analysis – Univariate Analysis

Demographic Data - Profession

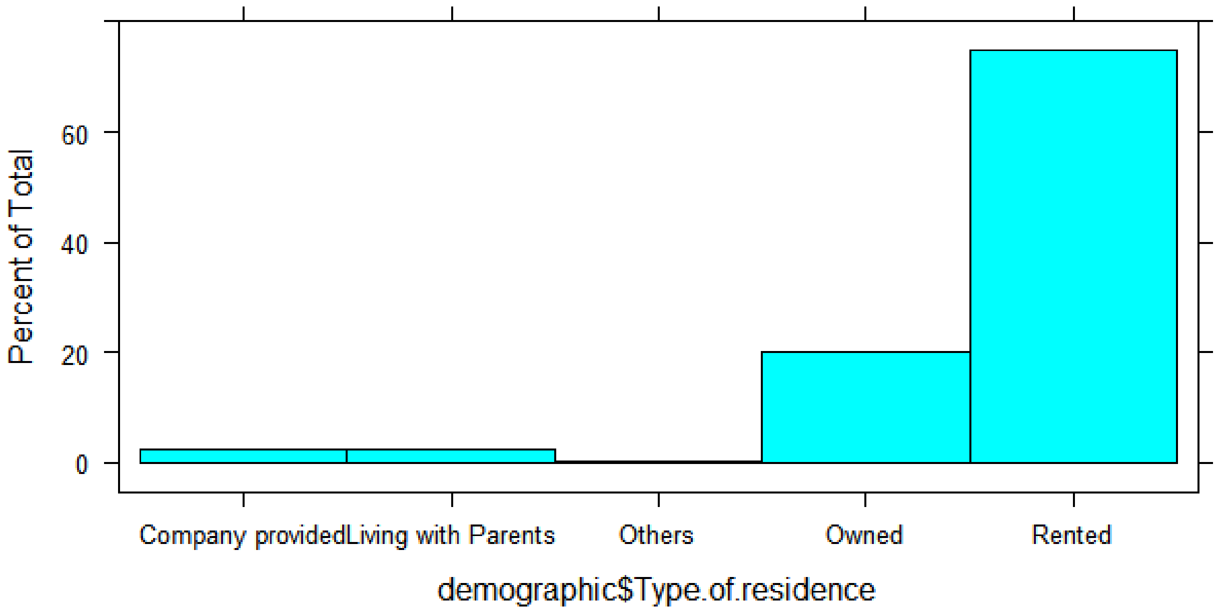


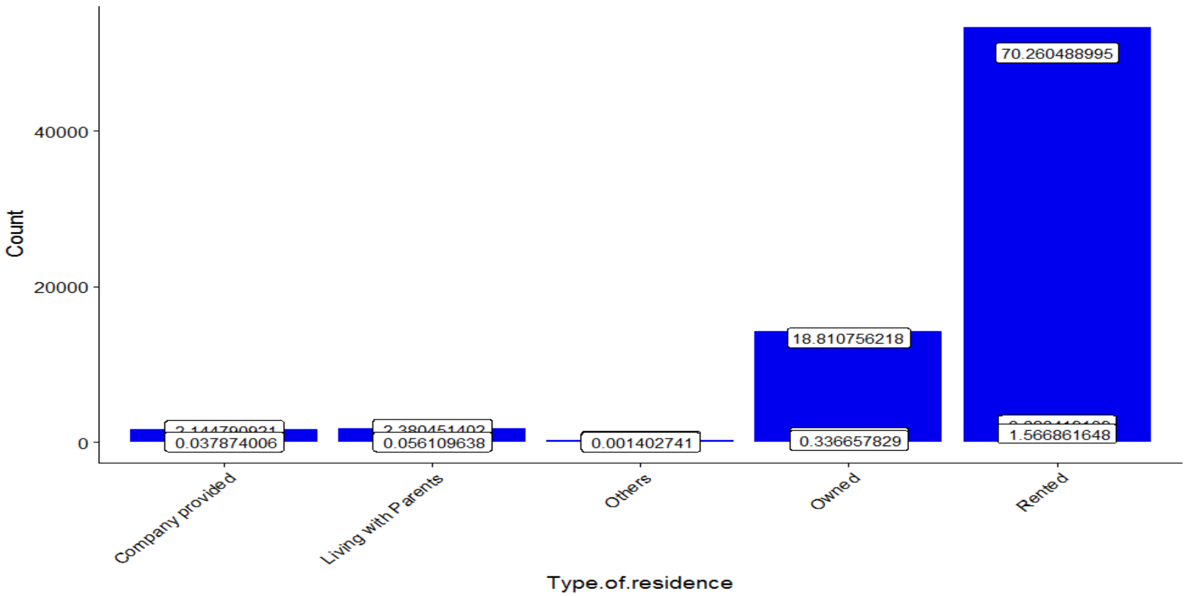


Salaried people constitutes the majority of applications

Exploratory Data Analysis – Univariate Analysis

Demographic Data - Type of Residence

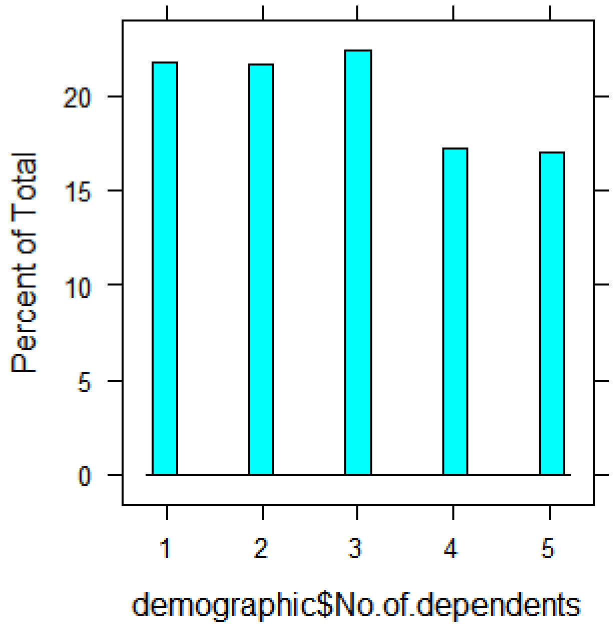


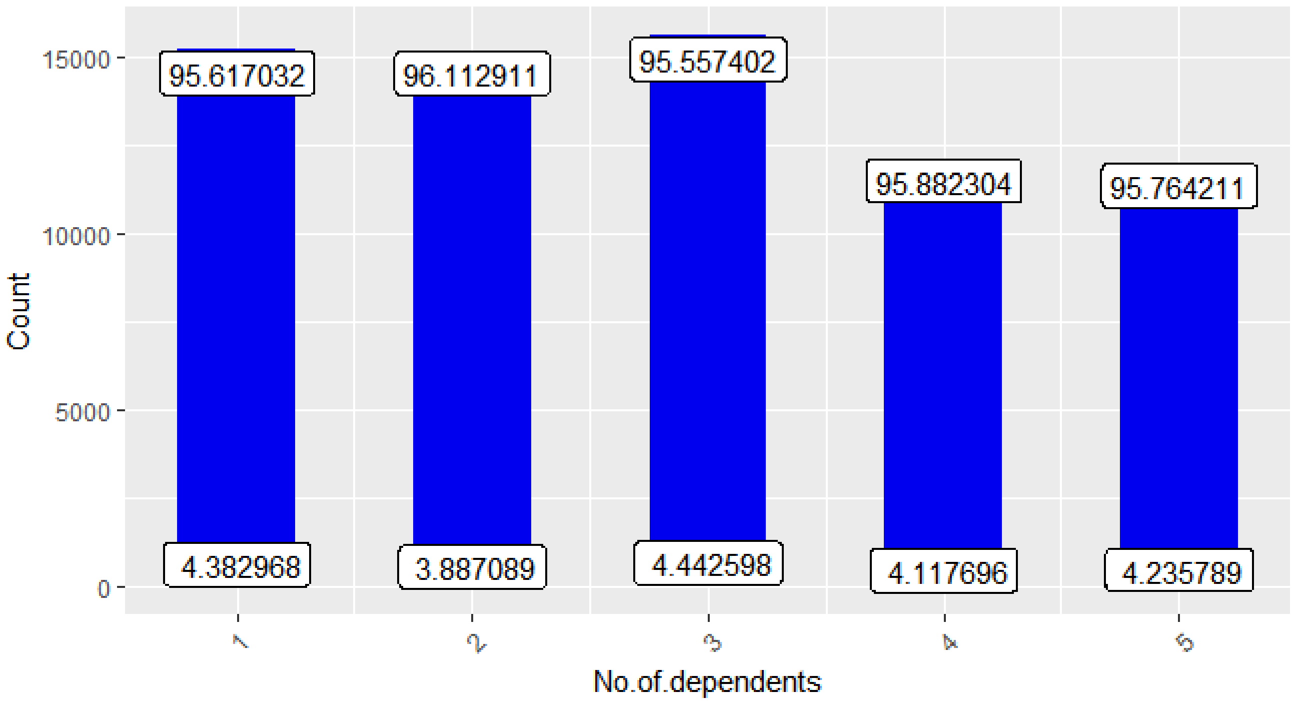


70.3 percentage of applicants stay in the rented home and around 18.9 stay in their own home

Exploratory Data Analysis – Univariate Analysis

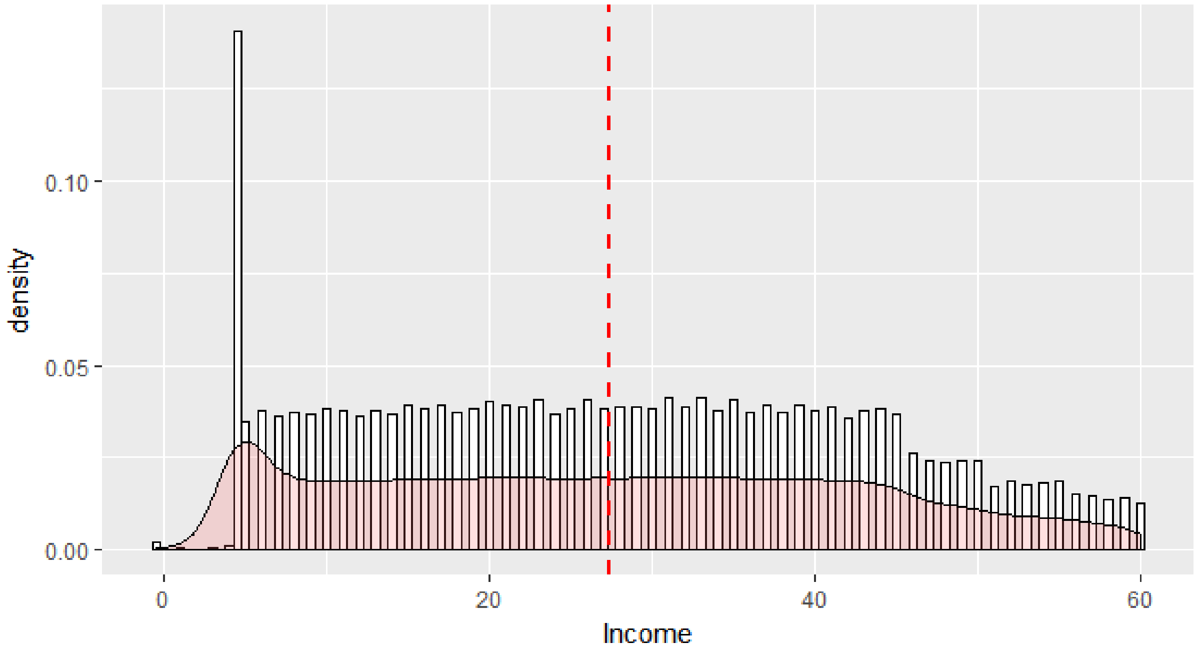
Demographic Data - Dependents



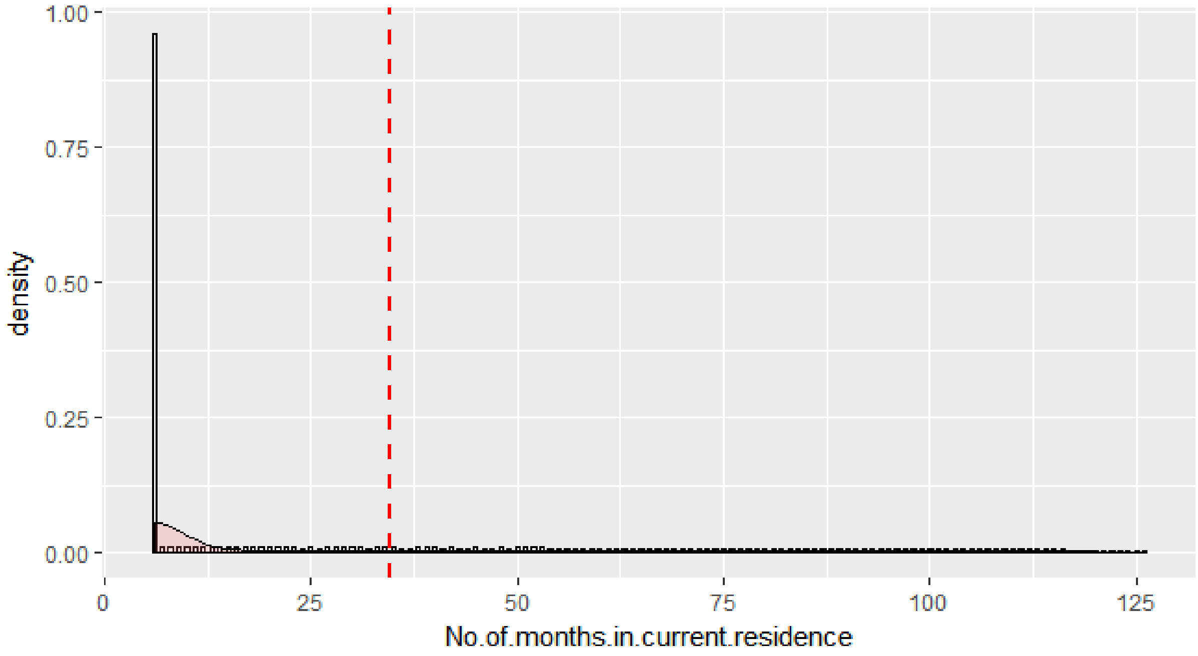


Exploratory Data Analysis – Univariate Analysis

Demographic Data - Income

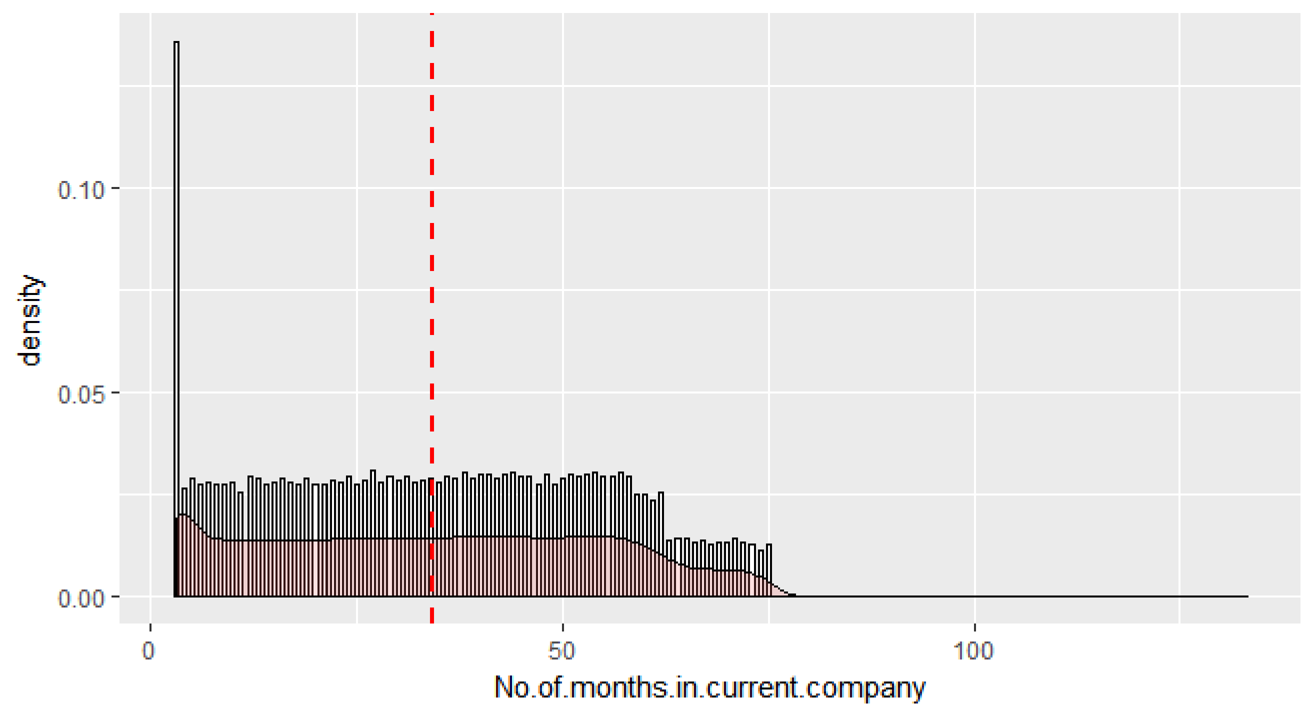


Demographic Data - Residence Duration



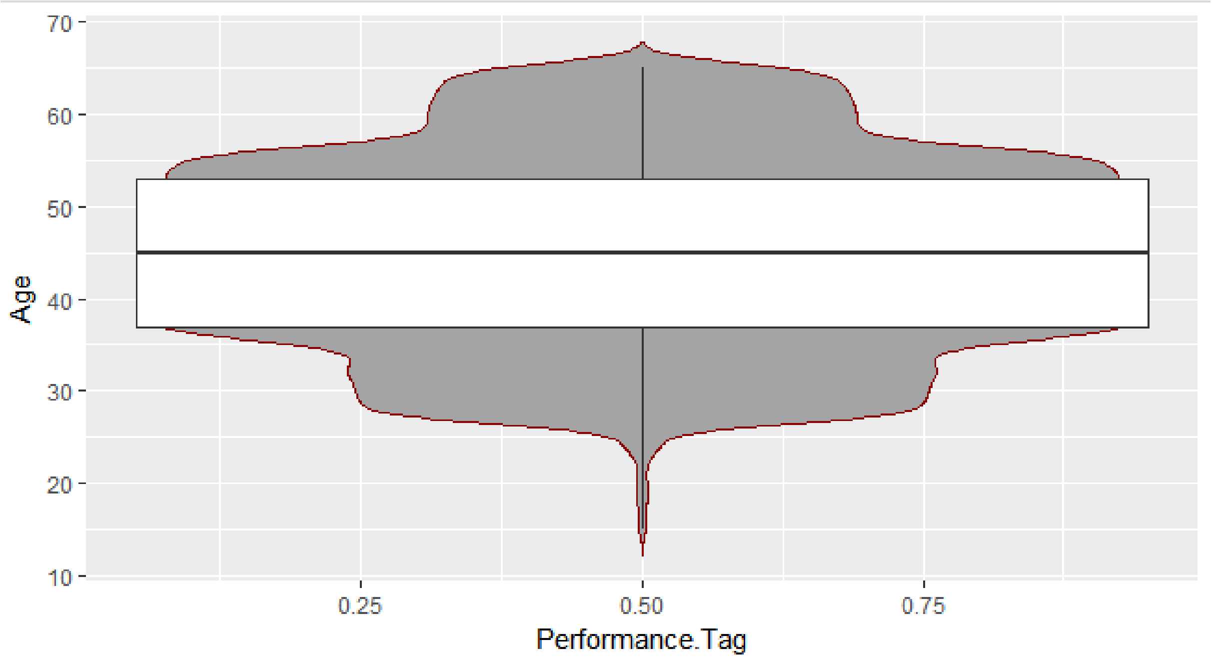
Exploratory Data Analysis – Univariate Analysis

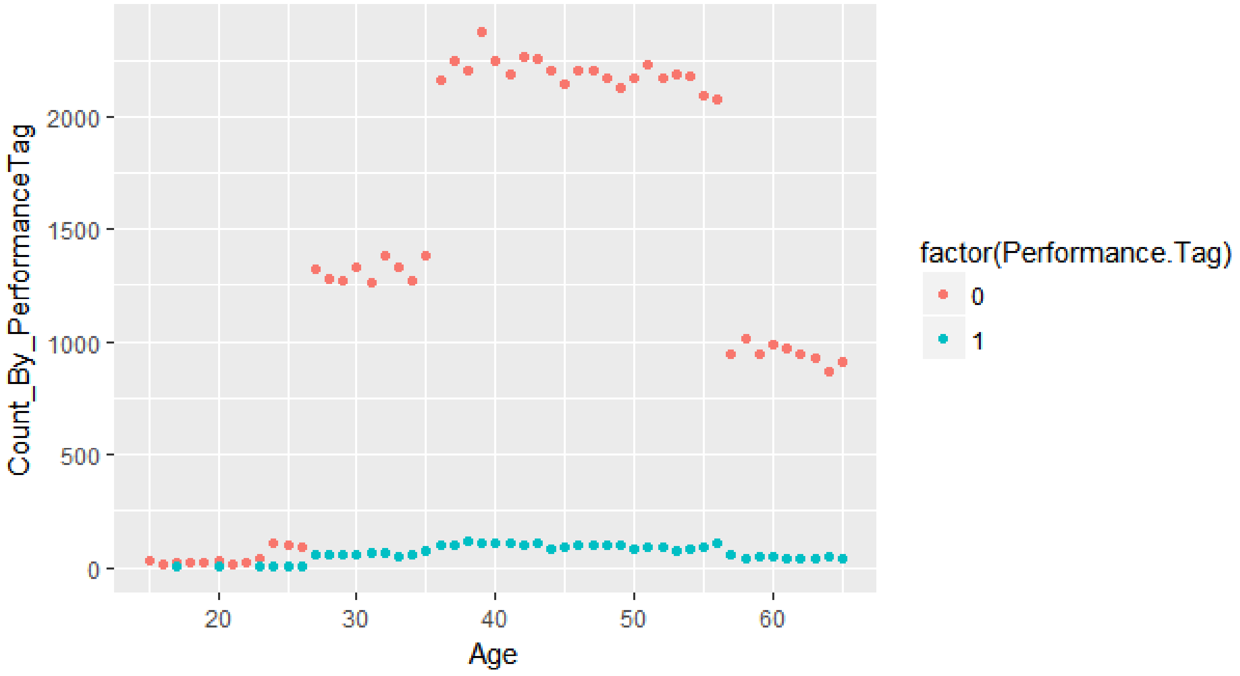
Demographic Data - Current Company



Exploratory Data Analysis – Bivariate Analysis

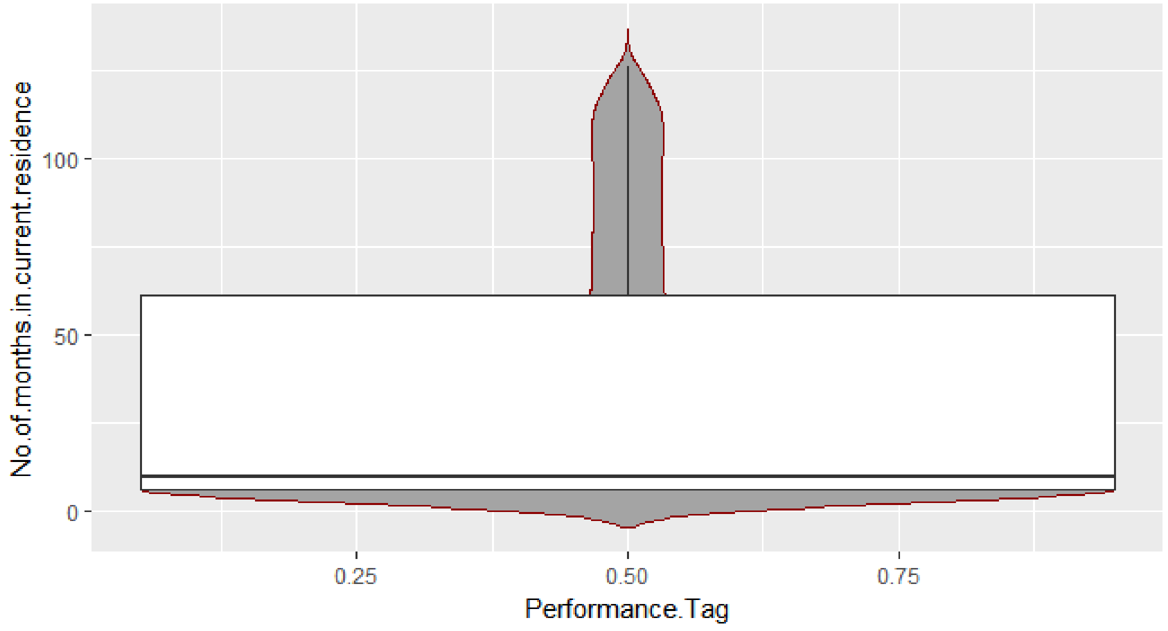
Demographic Data - Performance Tag Vs Age

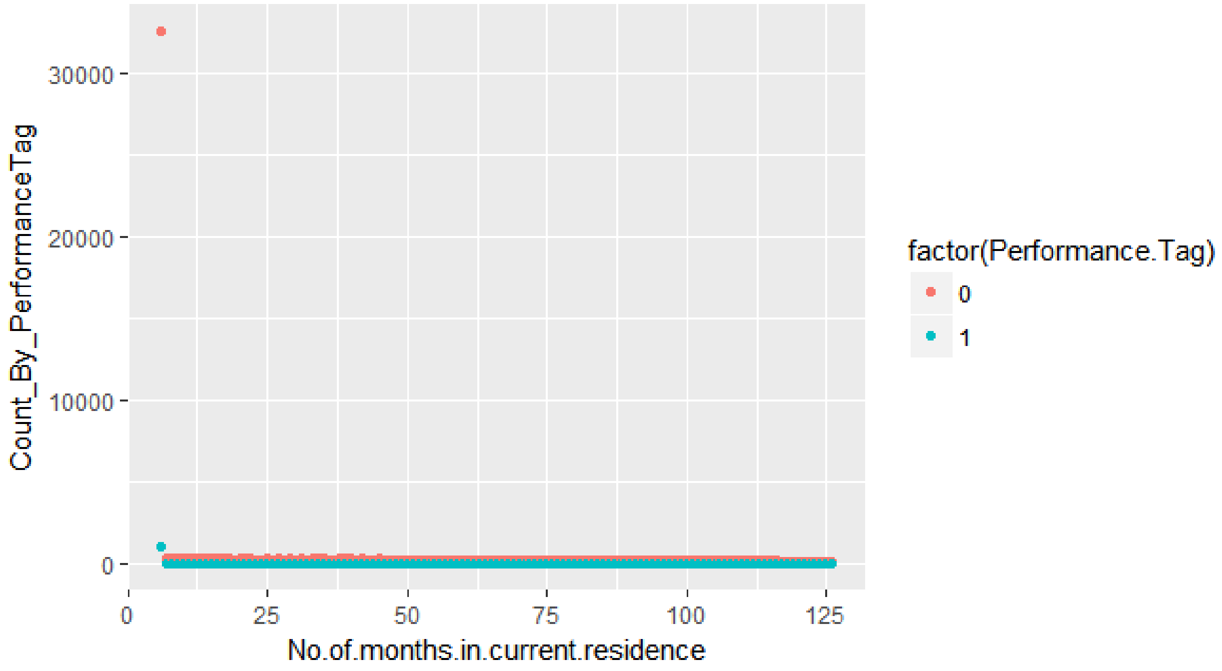




Exploratory Data Analysis – Bivariate Analysis

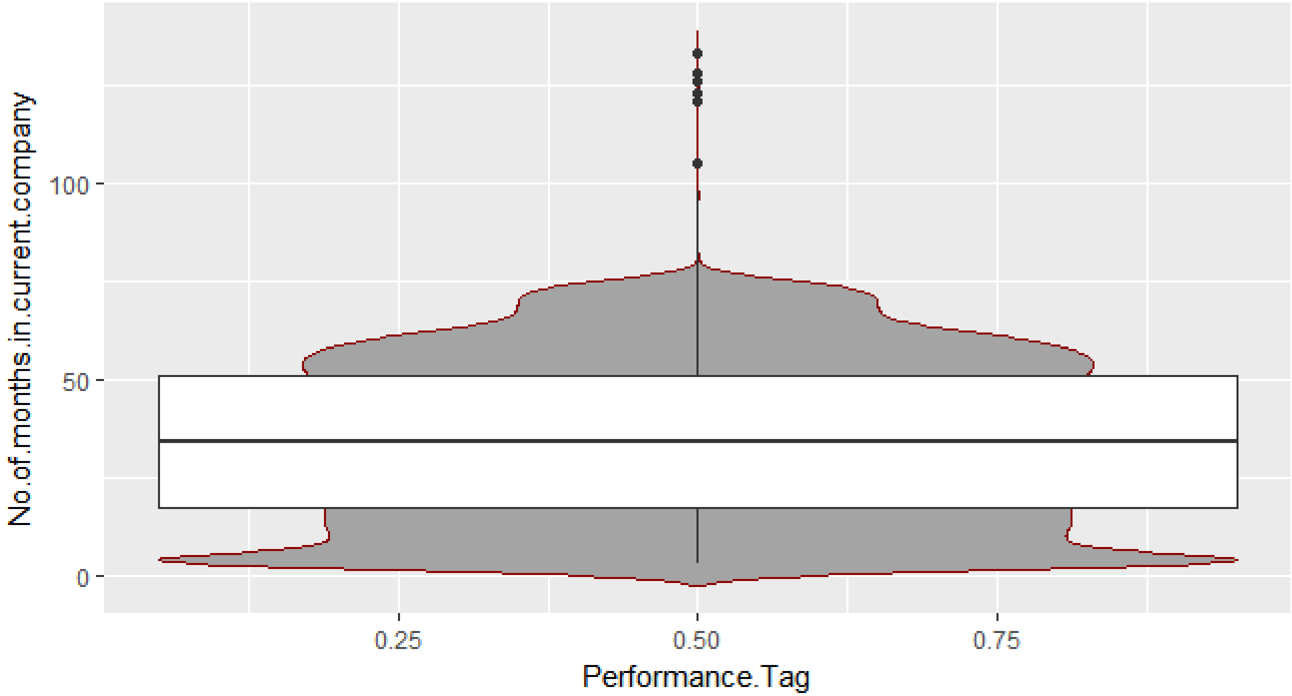
Demographic Data - Performance Tag Vs Current Residence

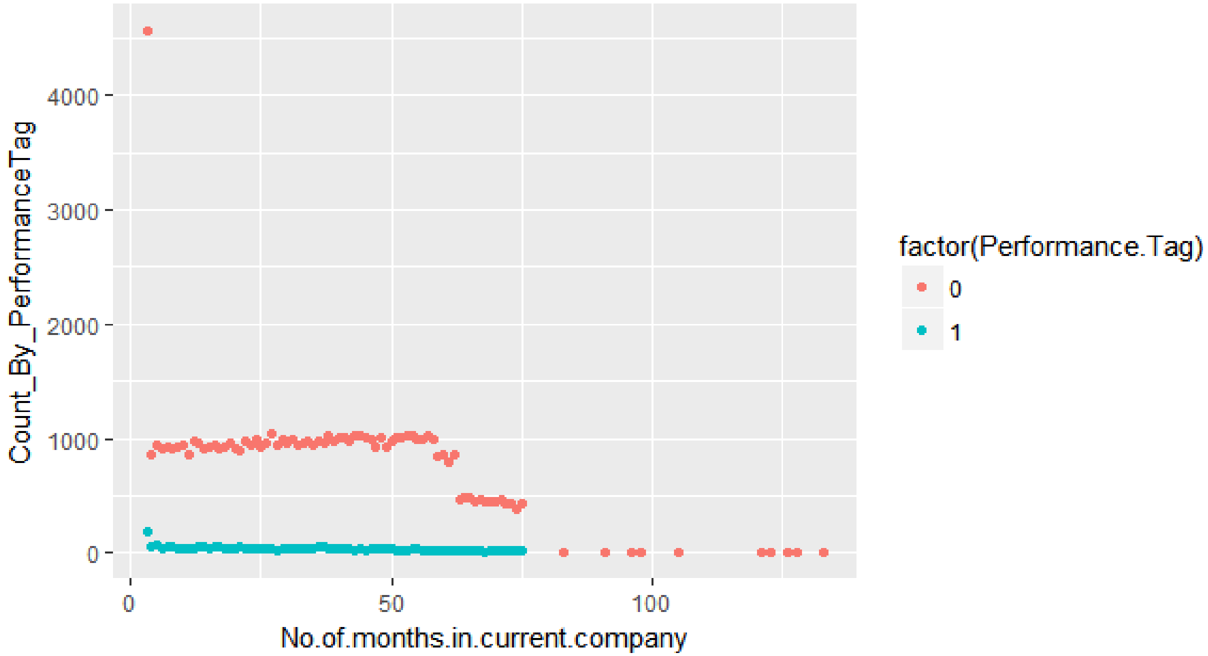




Exploratory Data Analysis – Bivariate Analysis

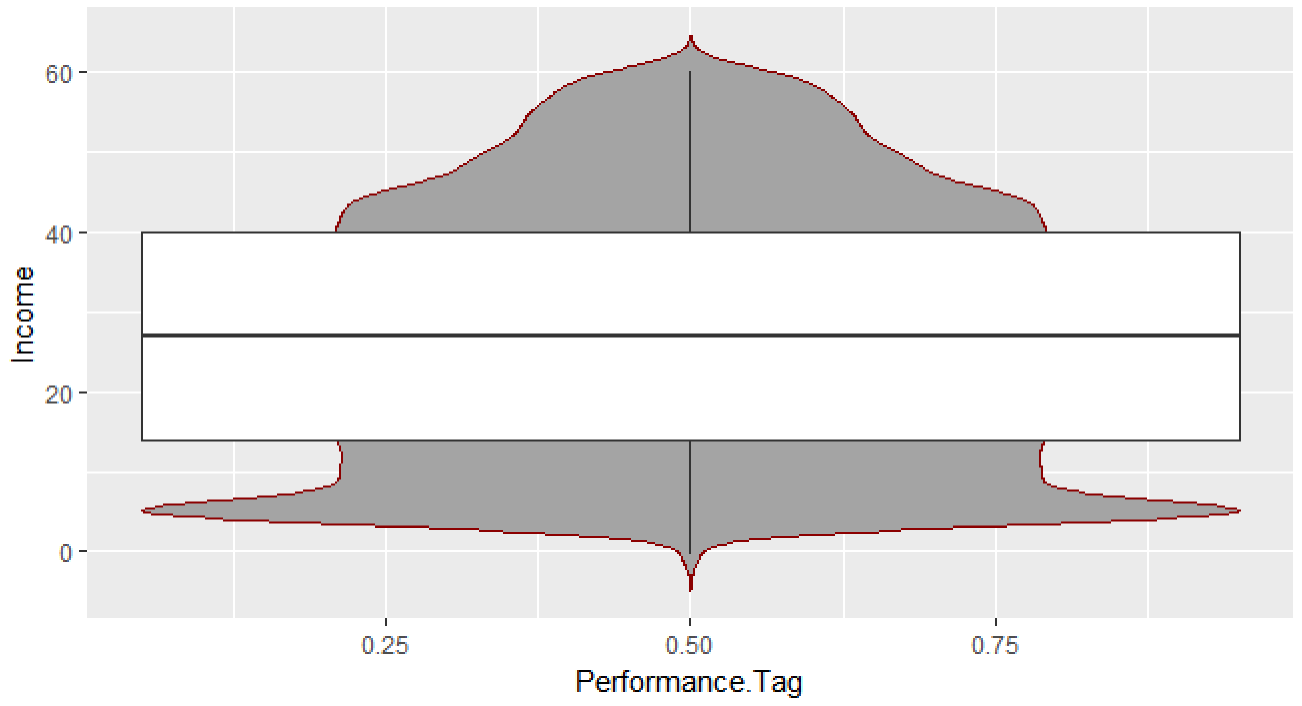
Demographic Data - Performance Tag Vs Current Company

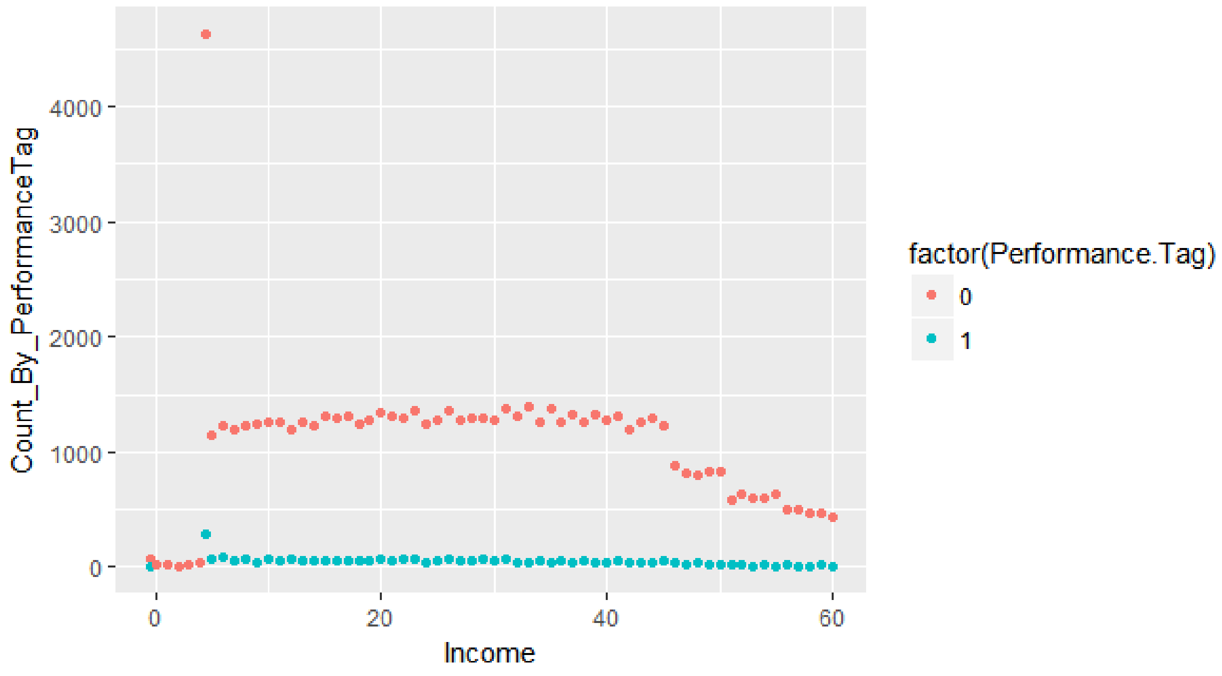




Exploratory Data Analysis – Bivariate Analysis

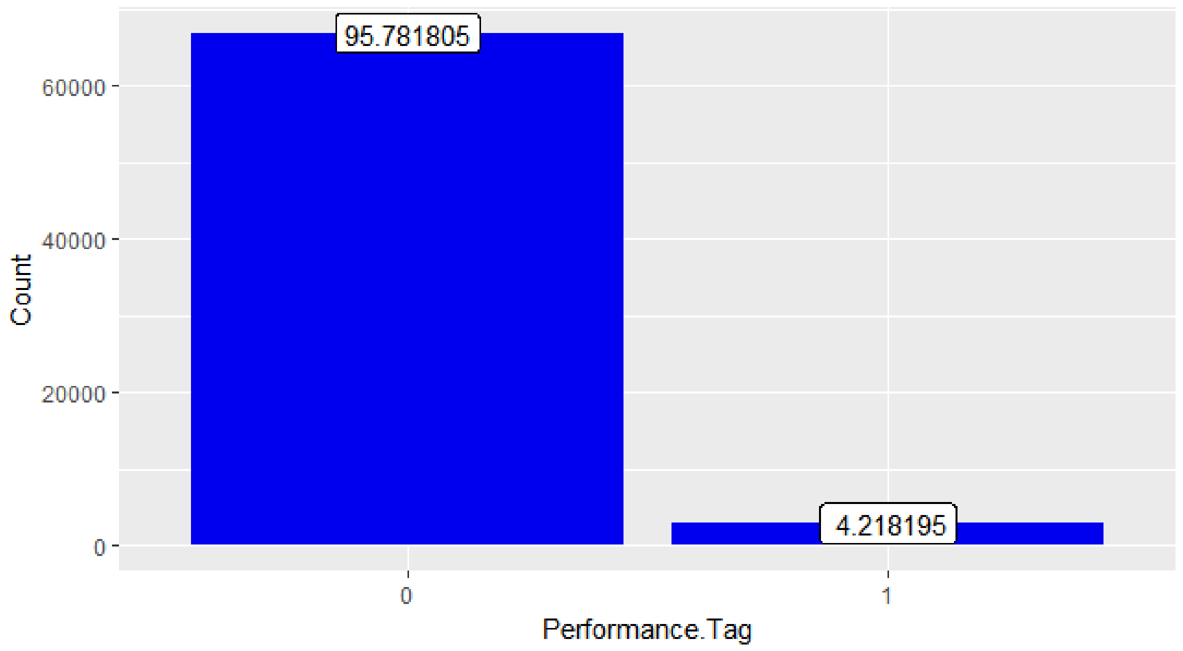
Demographic Data - Performance Tag Vs Income



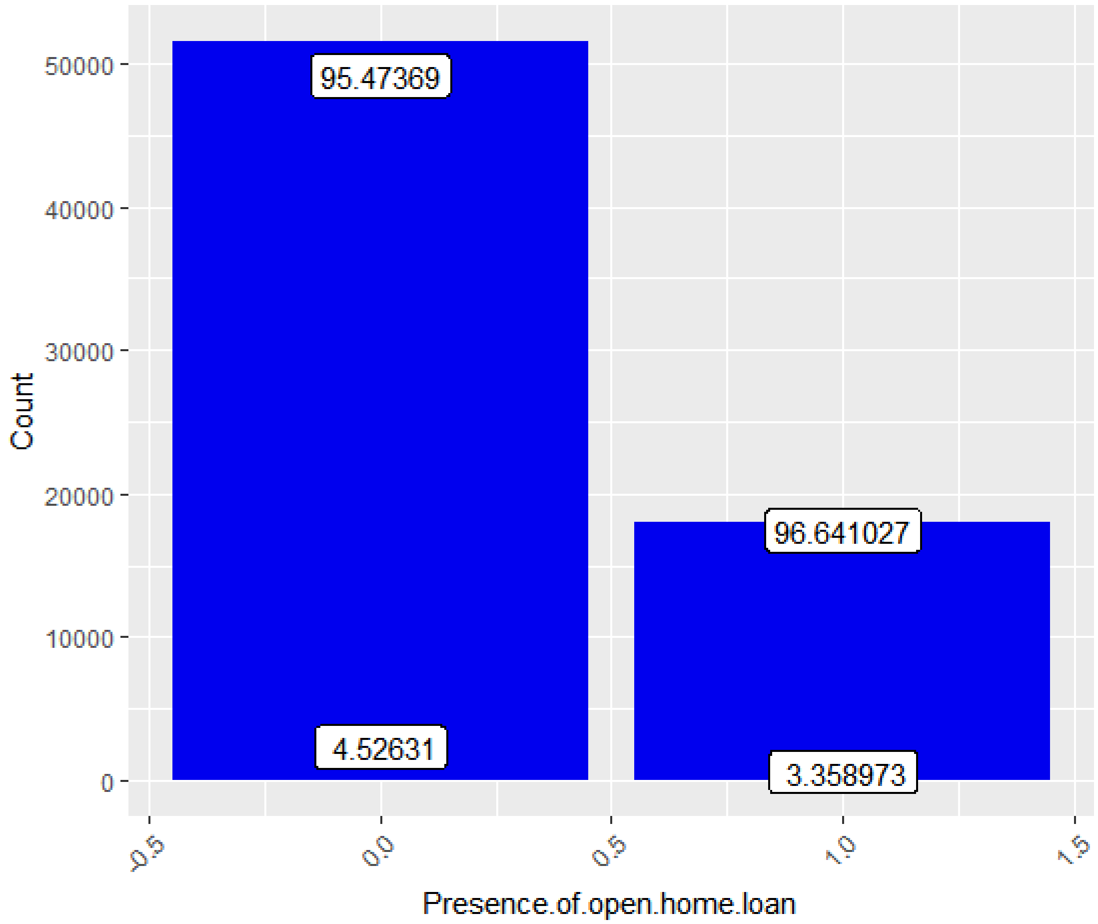


Exploratory Data Analysis – Univariate Analysis

Credit Bureau Data – Performance Tag

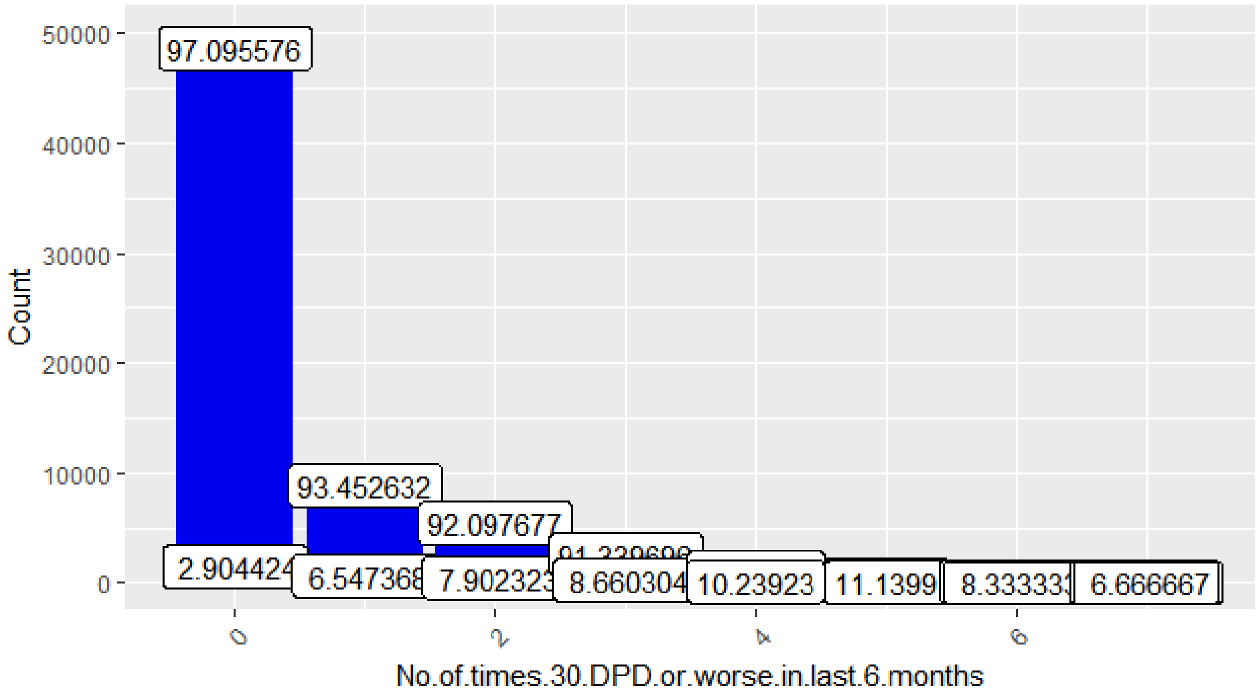


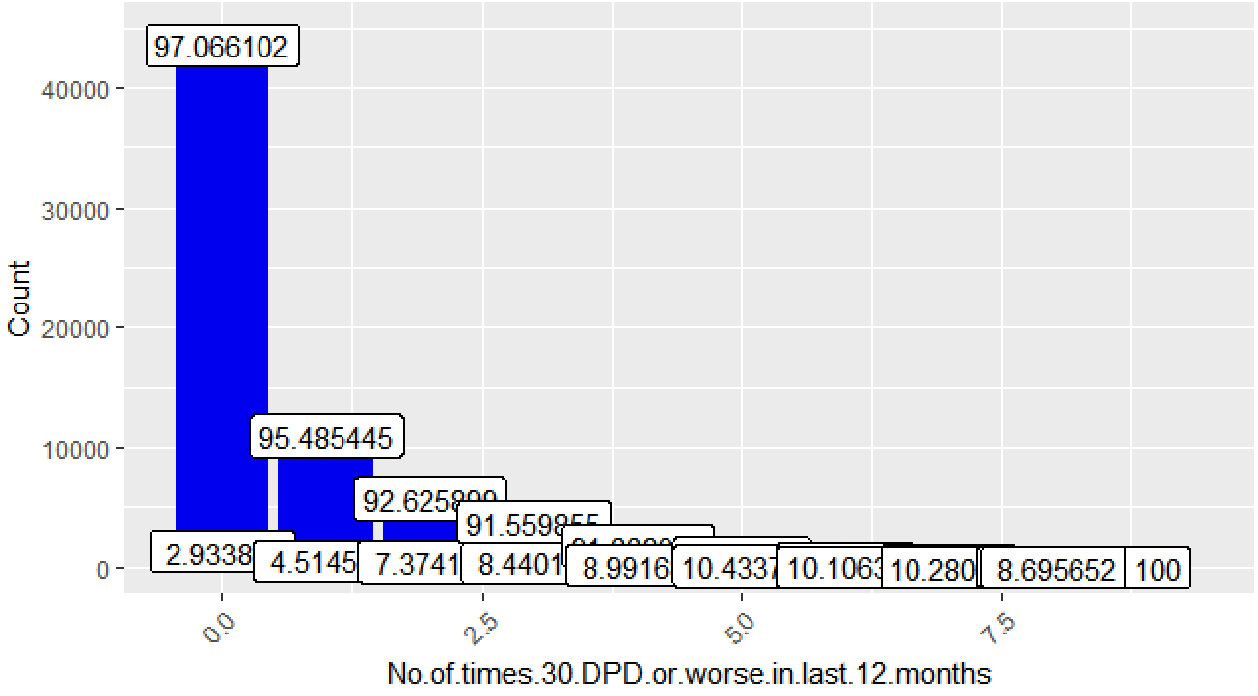
Credit Bureau Data – Presence of Open Home Loan



Exploratory Data Analysis – Univariate Analysis

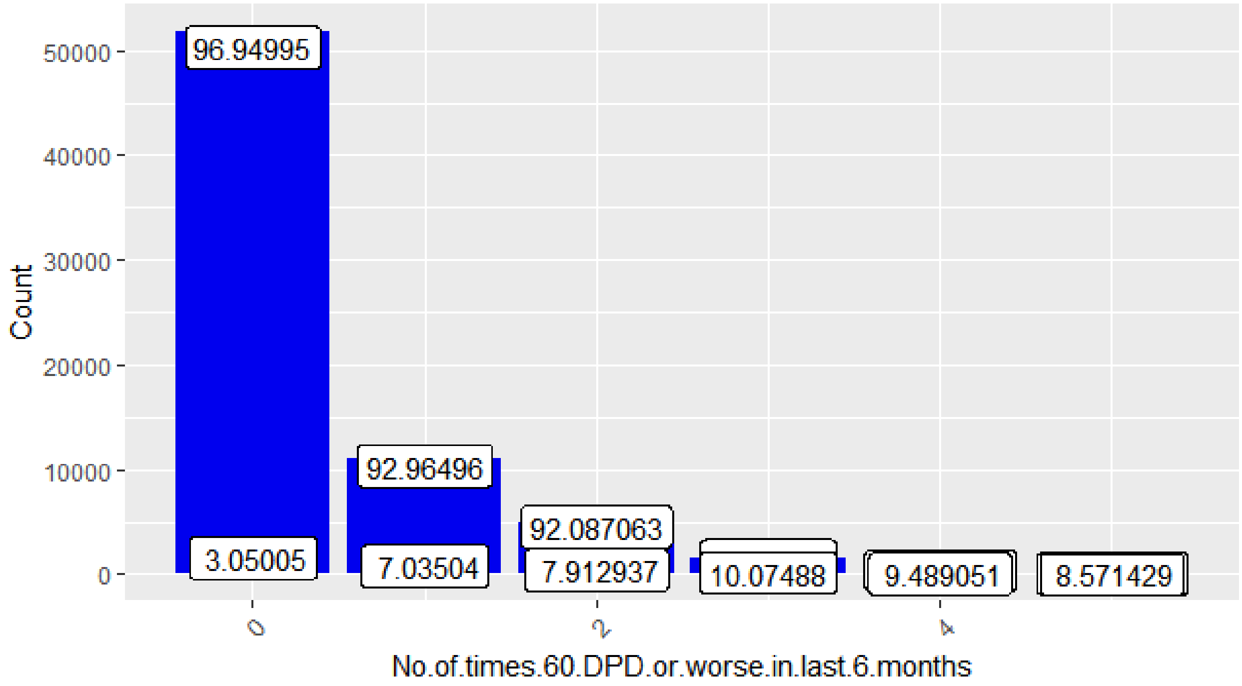
Credit Bureau Data – No of times 30 DPD or worse in last 6 Months & 12 Months

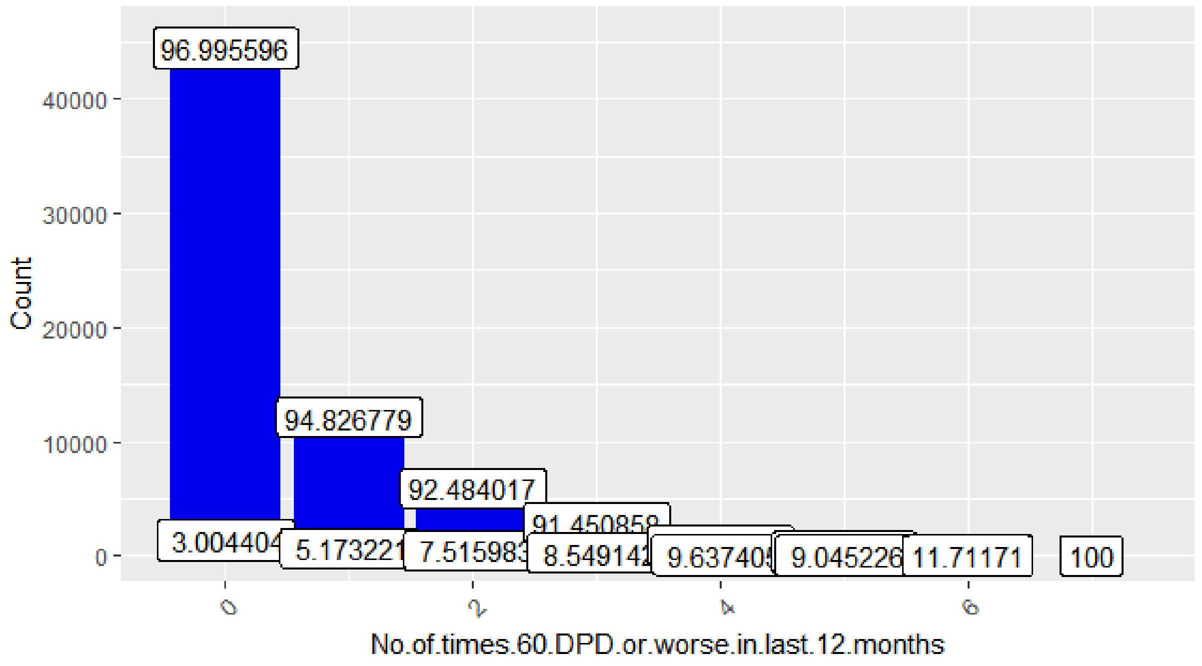




Exploratory Data Analysis – Univariate Analysis

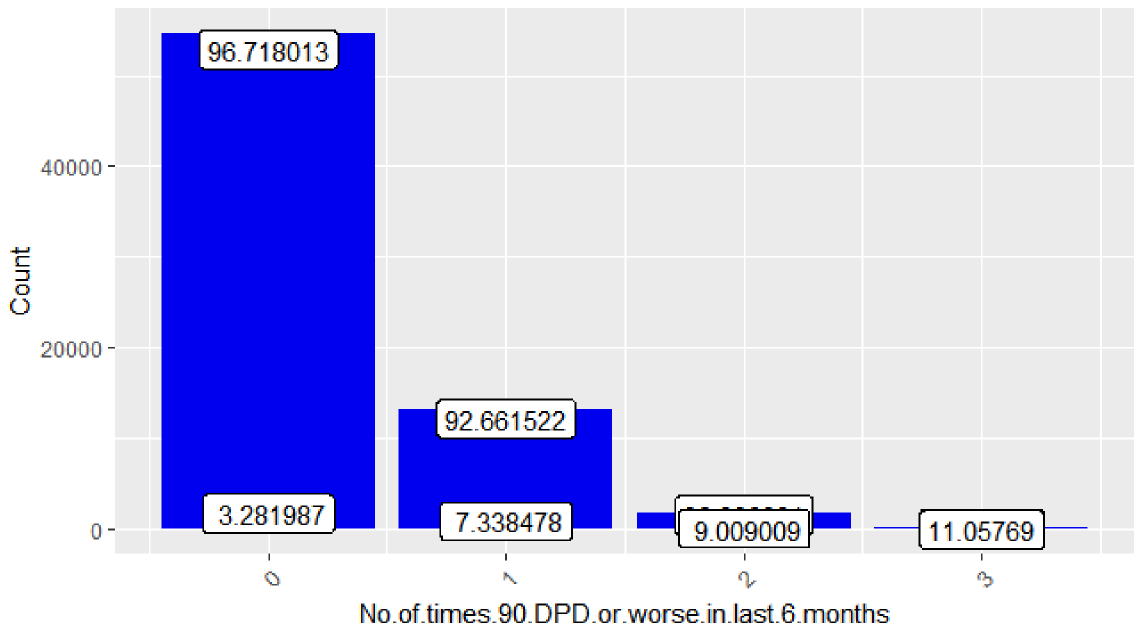
Credit Bureau Data – No of times 60 DPD or worse in last 6 Months & 12 Months

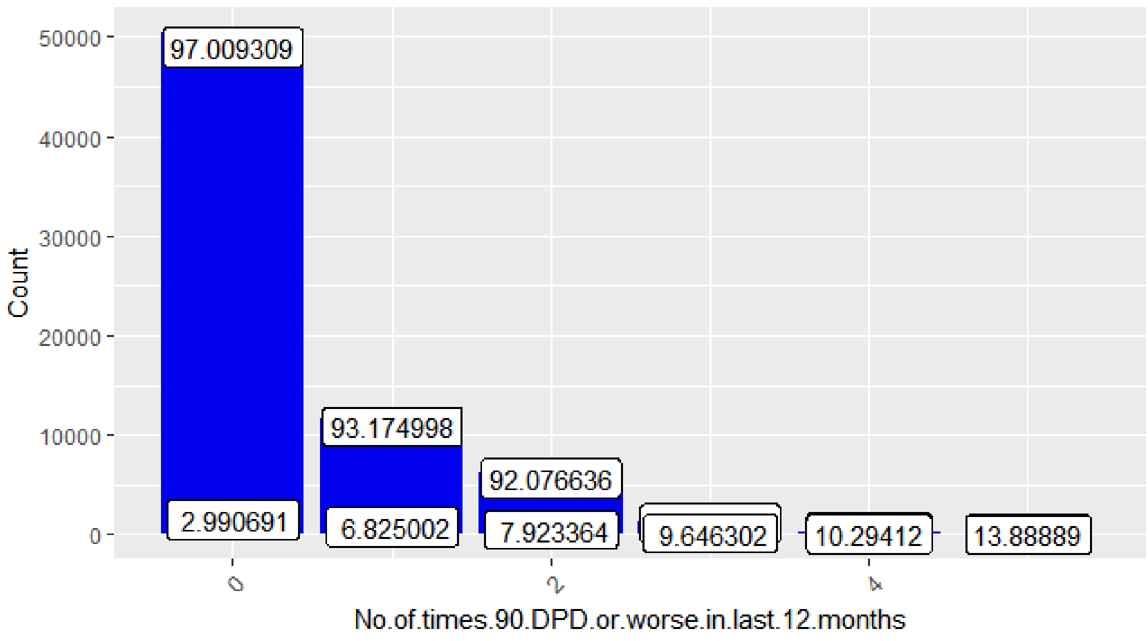




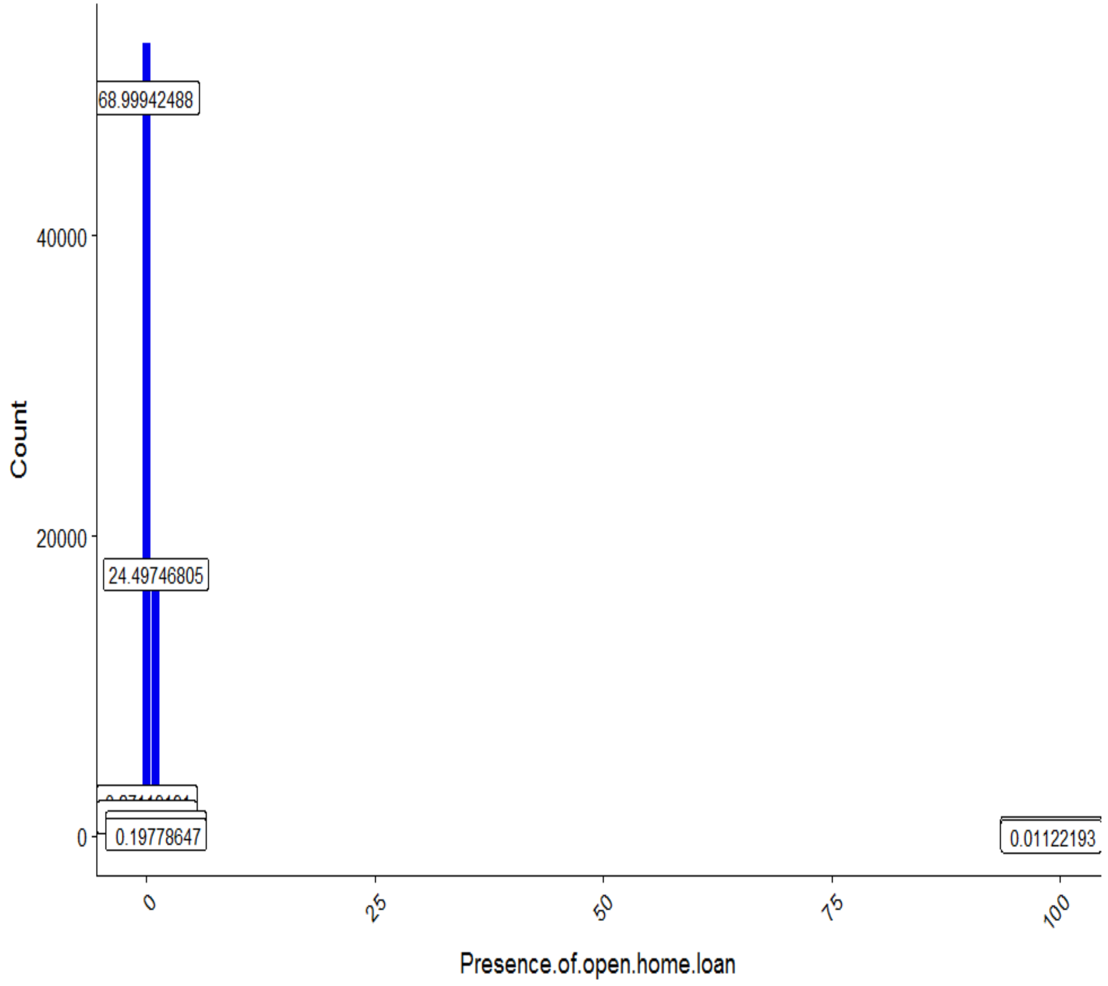
Exploratory Data Analysis – Univariate Analysis

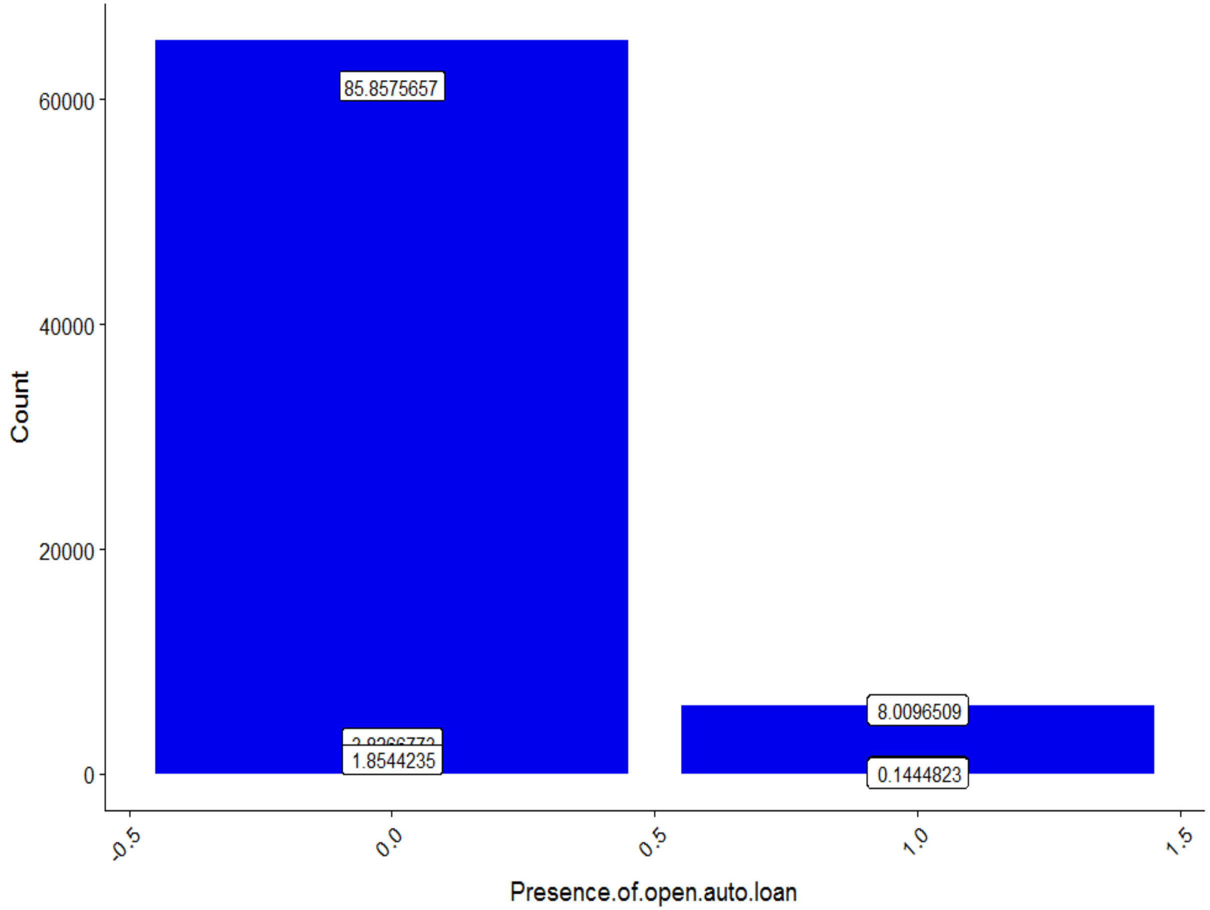
Credit Bureau Data – No of times 90 DPD or worse in last 6 Months & 12 Months





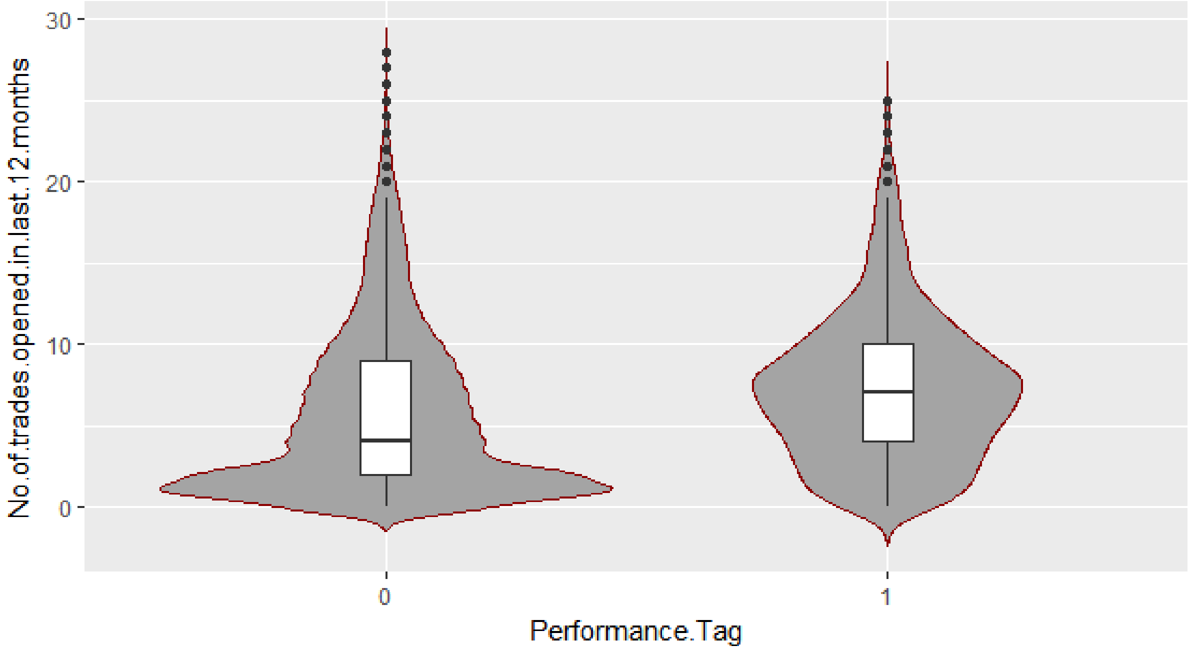
Exploratory Data Analysis – Univariate Analysis





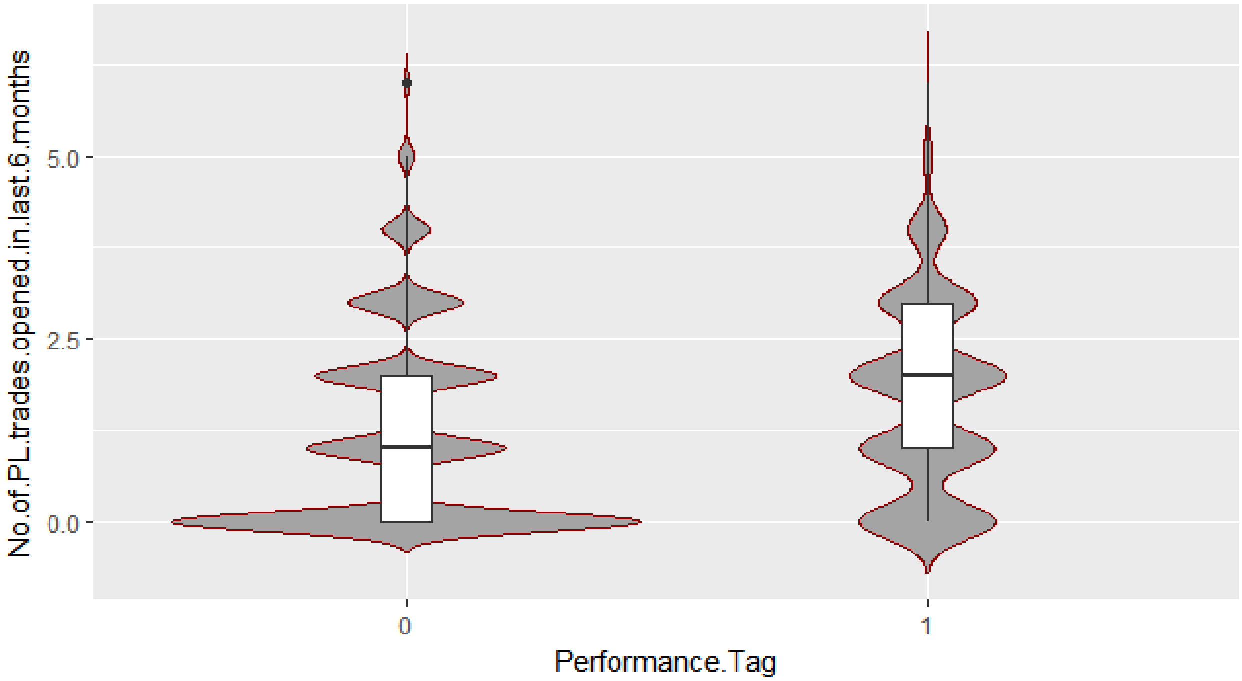
Exploratory Data Analysis – Bivariate Analysis

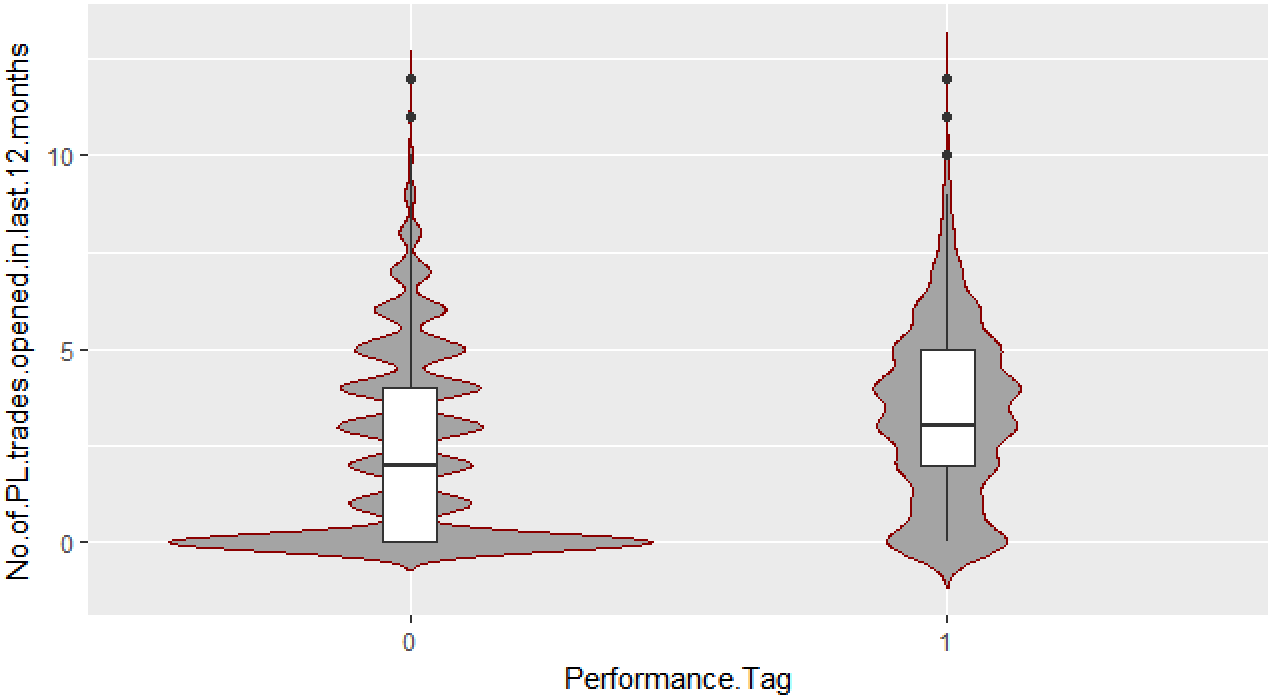
Credit Bureau Data – Performance Tag Vs No of trade opened in last 12 Months



Exploratory Data Analysis – Bivariate Analysis

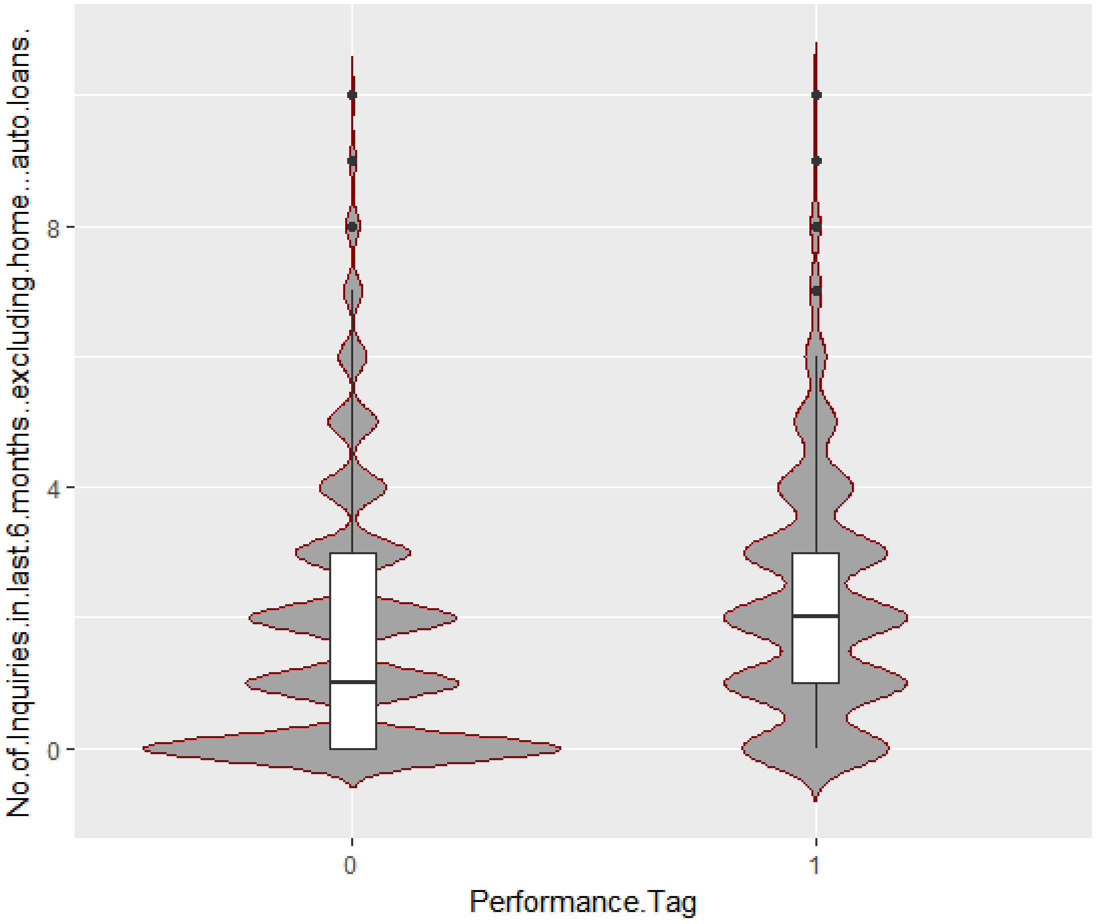
Credit Bureau Data – Performance Tag Vs No of PL trades opened in last 6 & 12 Months

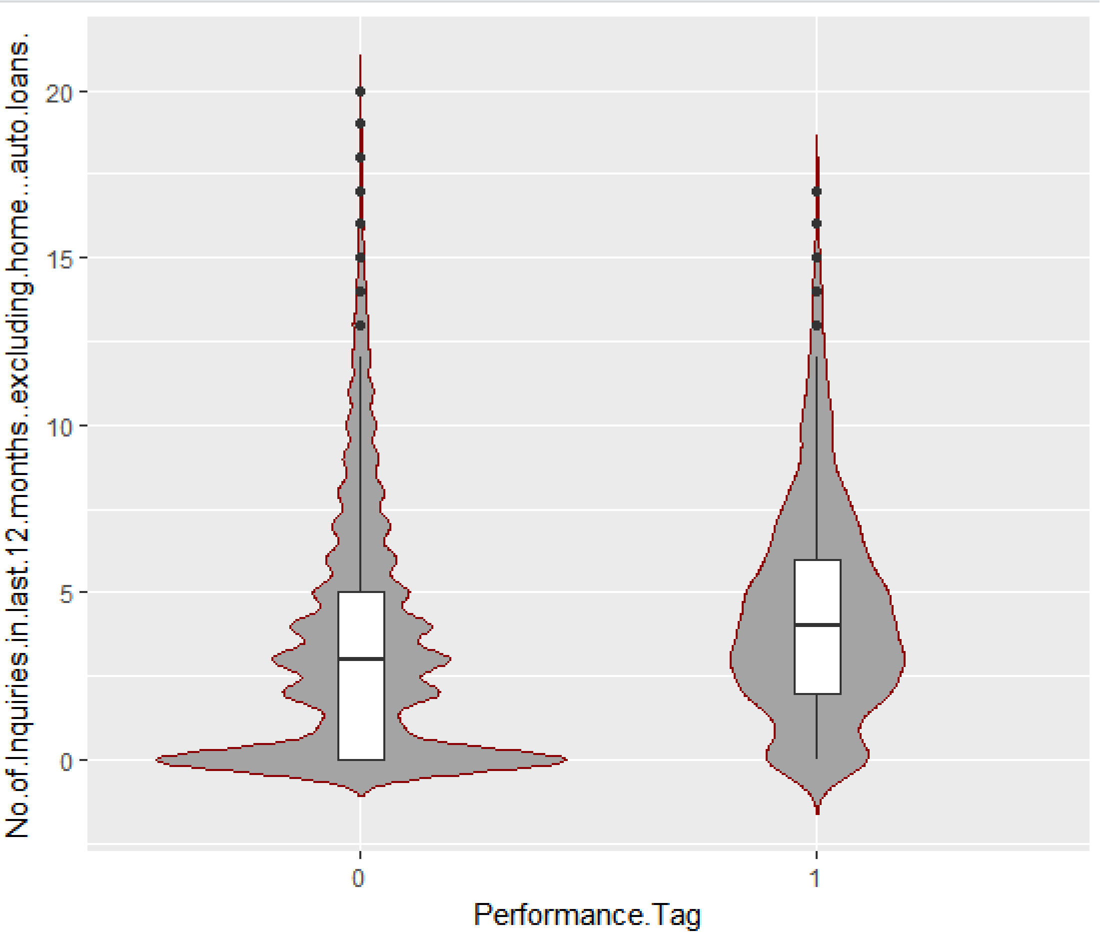




Exploratory Data Analysis – Bivariate Analysis

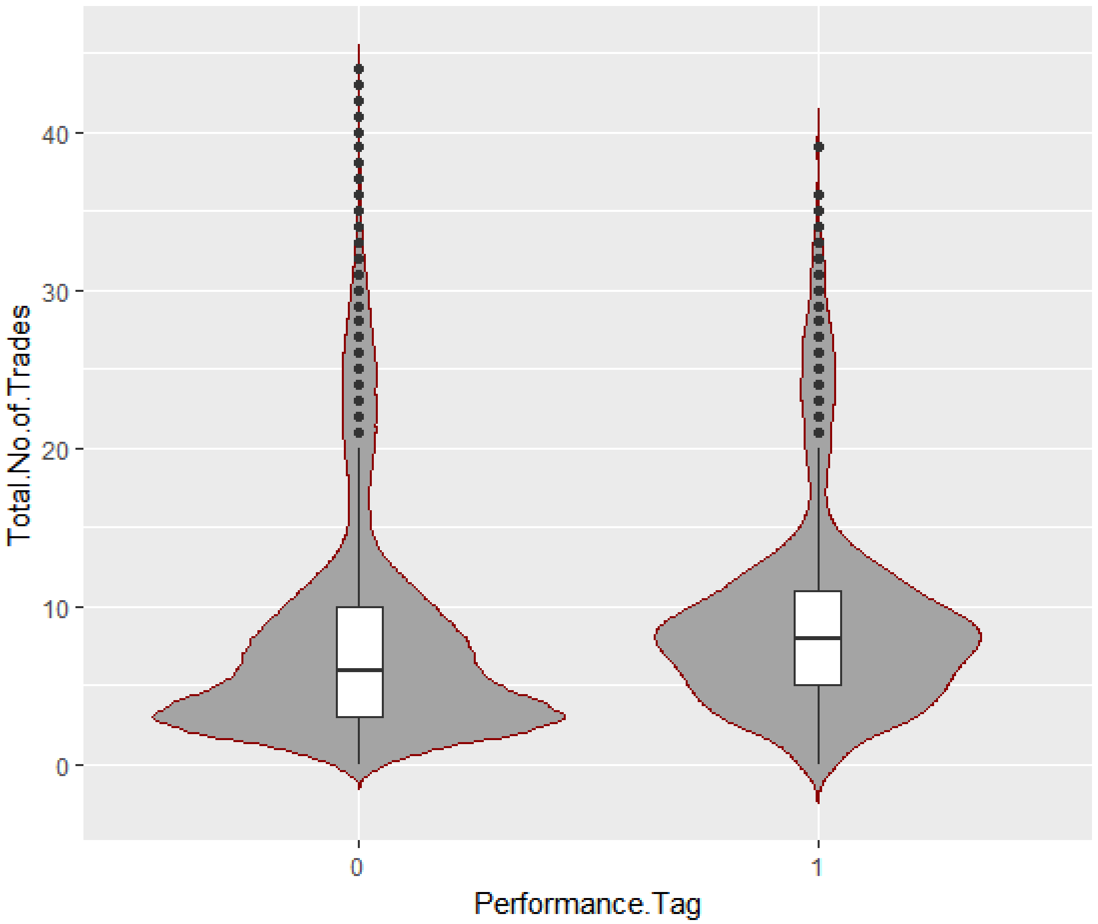
Credit Bureau Data – Performance Tag Vs No of Inquiries in last 6 & 12 months excluding home & auto loans

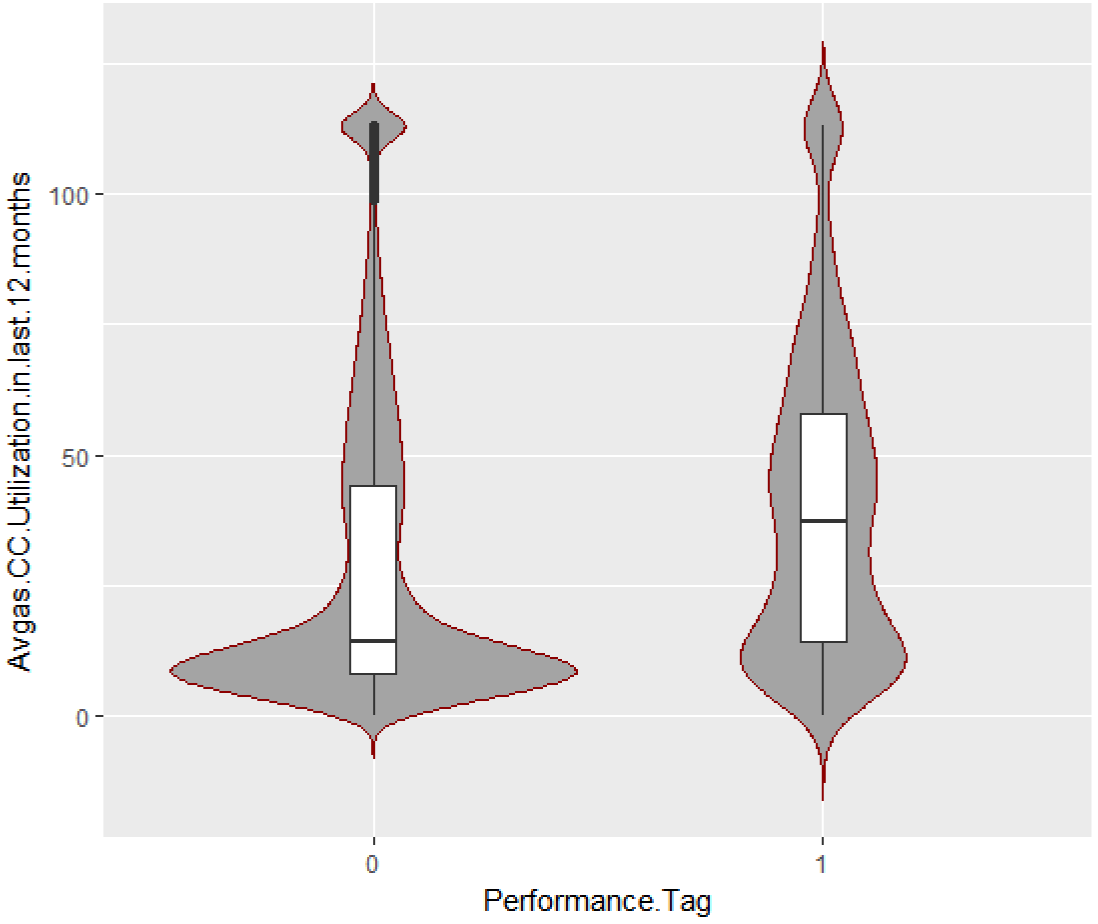




Exploratory Data Analysis – Bivariate Analysis

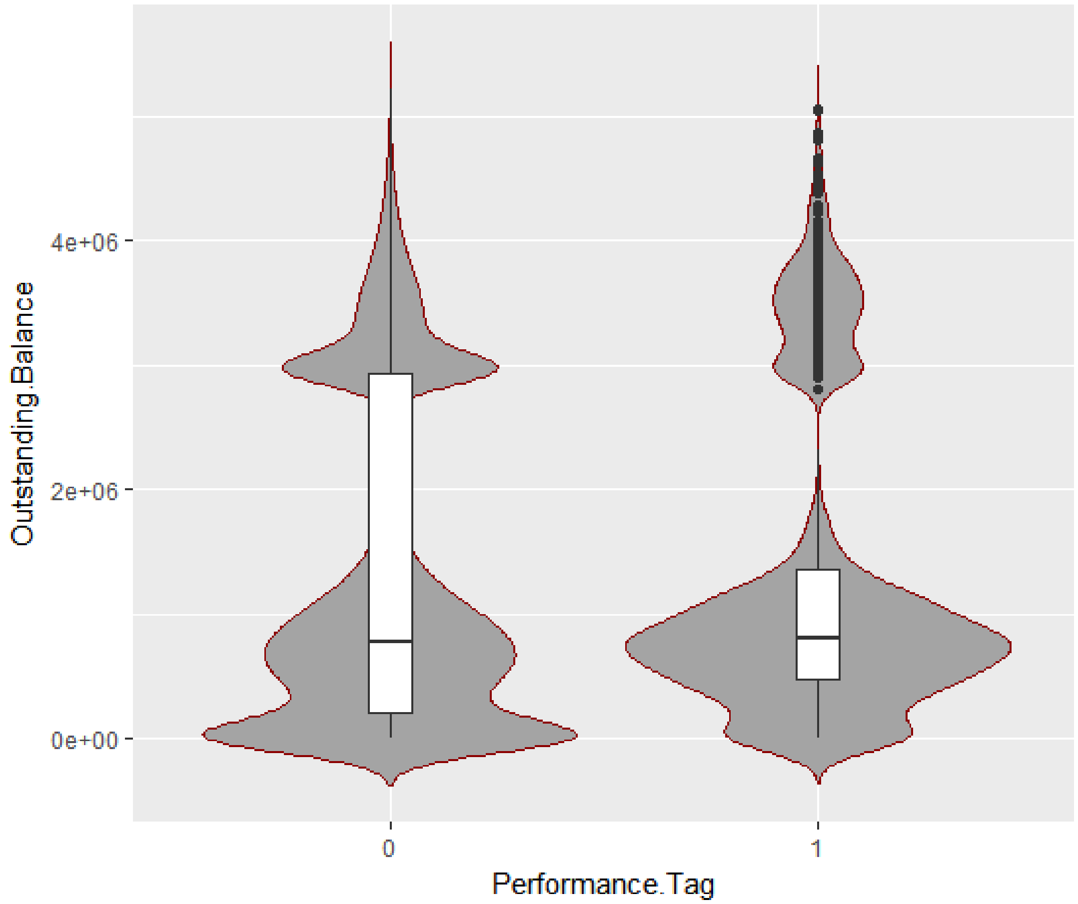
Credit Bureau Data – Performance Tag Vs Total No of Trades & Average Credit Card Utilization in last 12 months

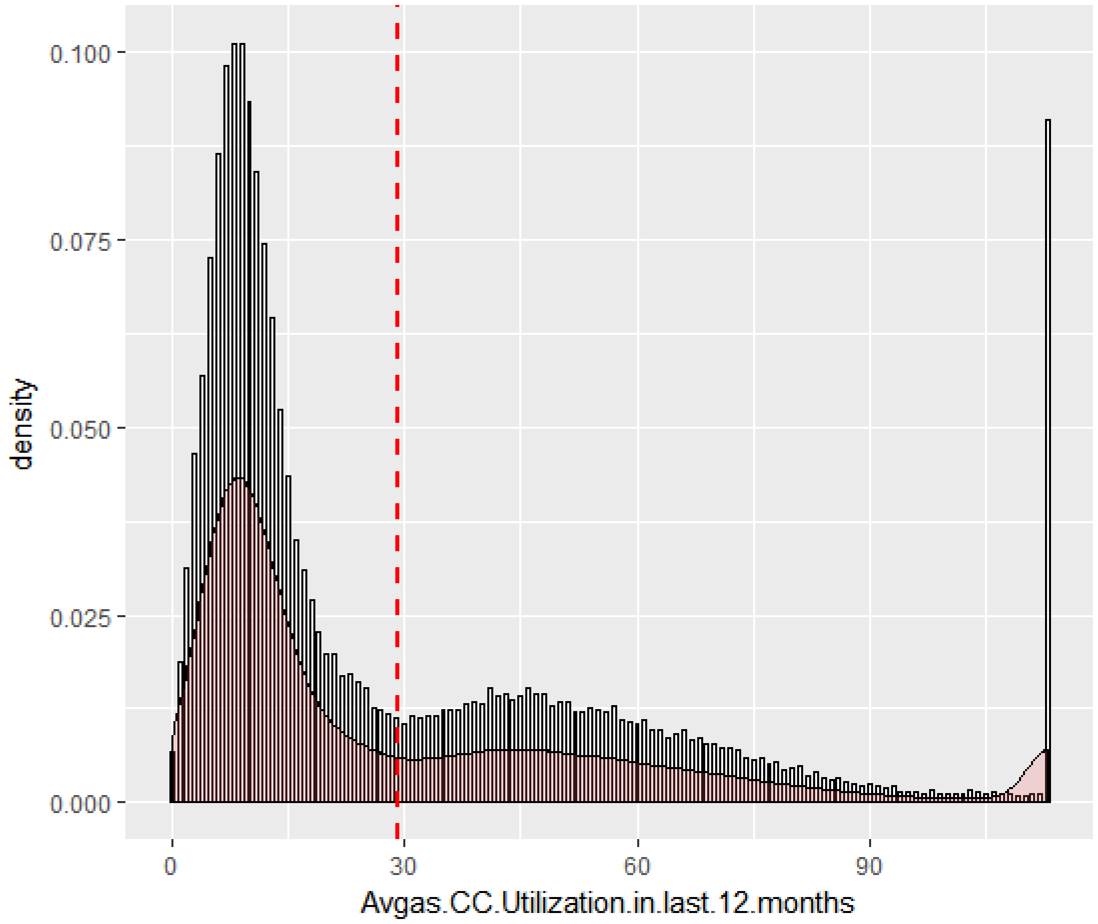




Exploratory Data Analysis – Bivariate Analysis

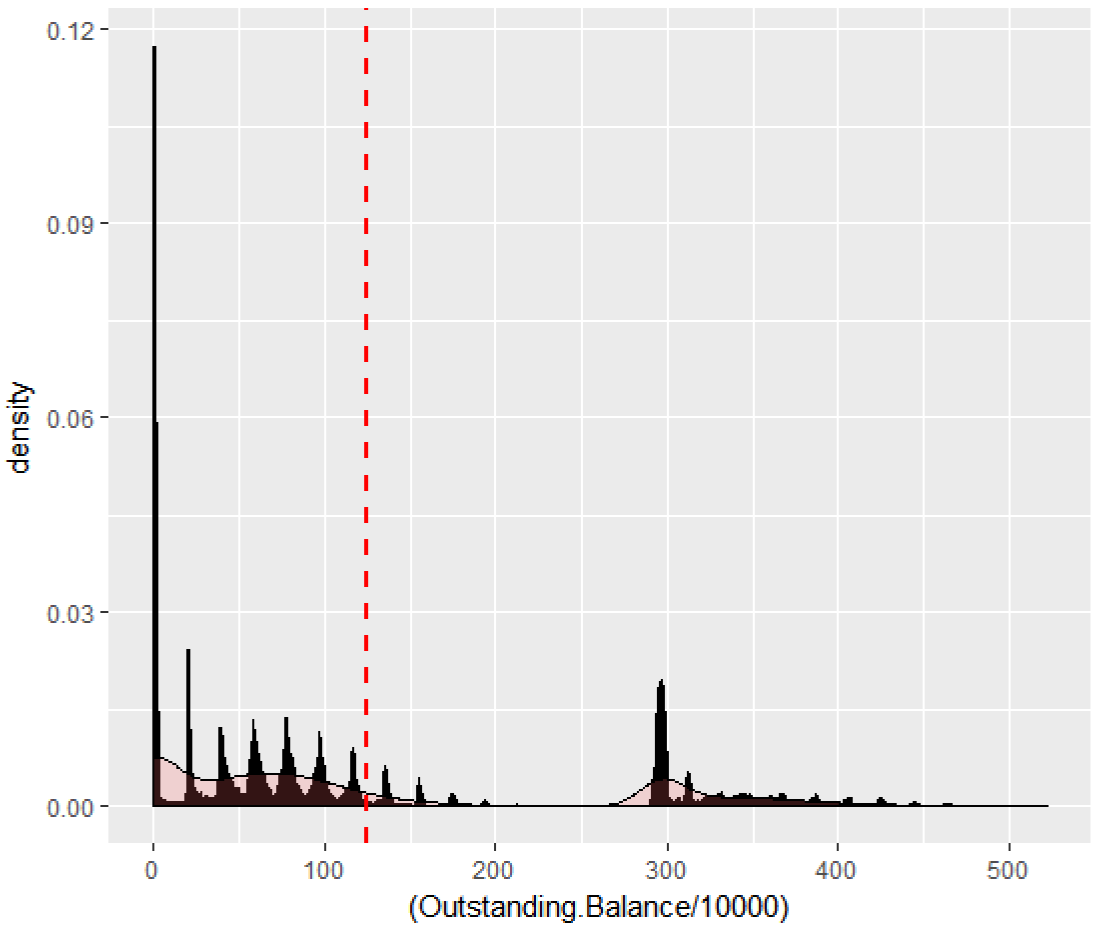
Credit Bureau Data – Performance Tag Vs Outstanding Balance & Average Credit Card Utilization in last 12 months Vs Density



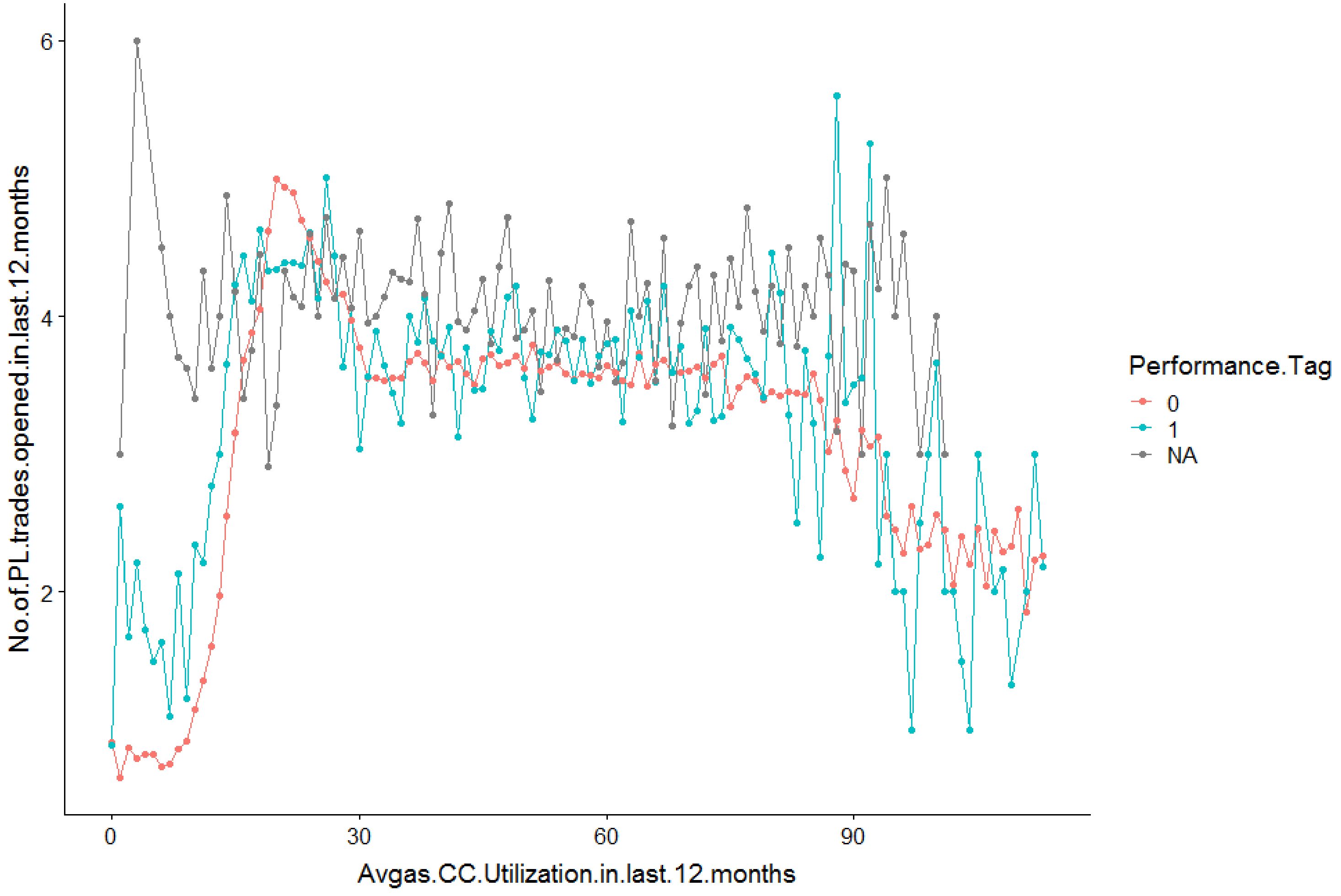


Exploratory Data Analysis – Bivariate Analysis

Credit Bureau Data – Outstanding Balance/ 10000 Vs Density

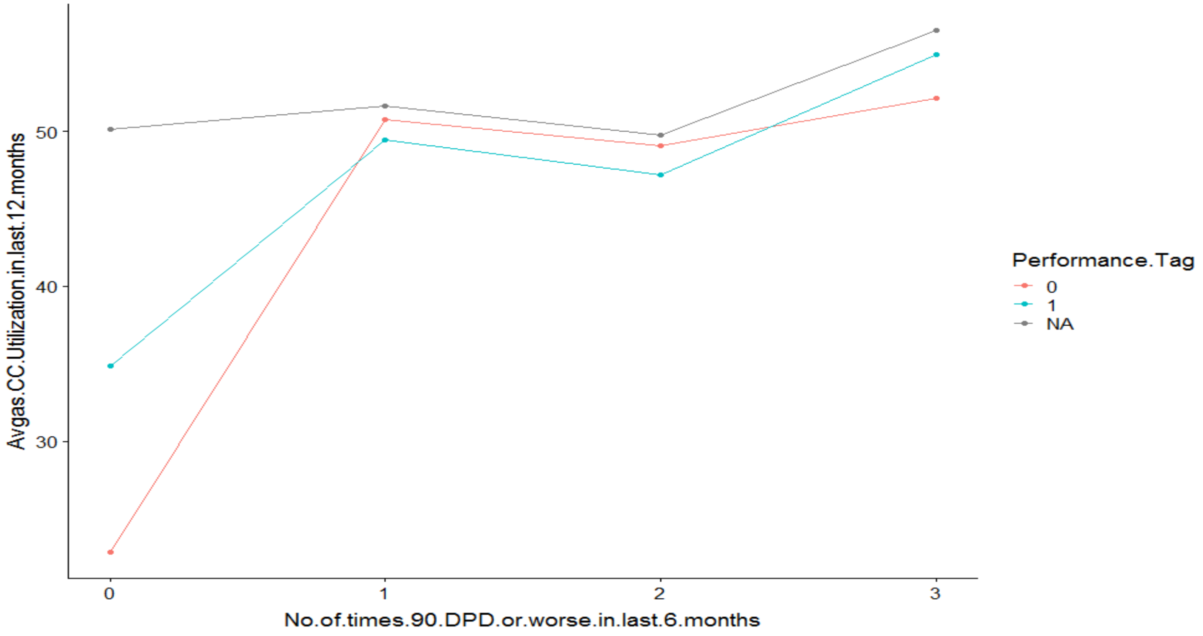


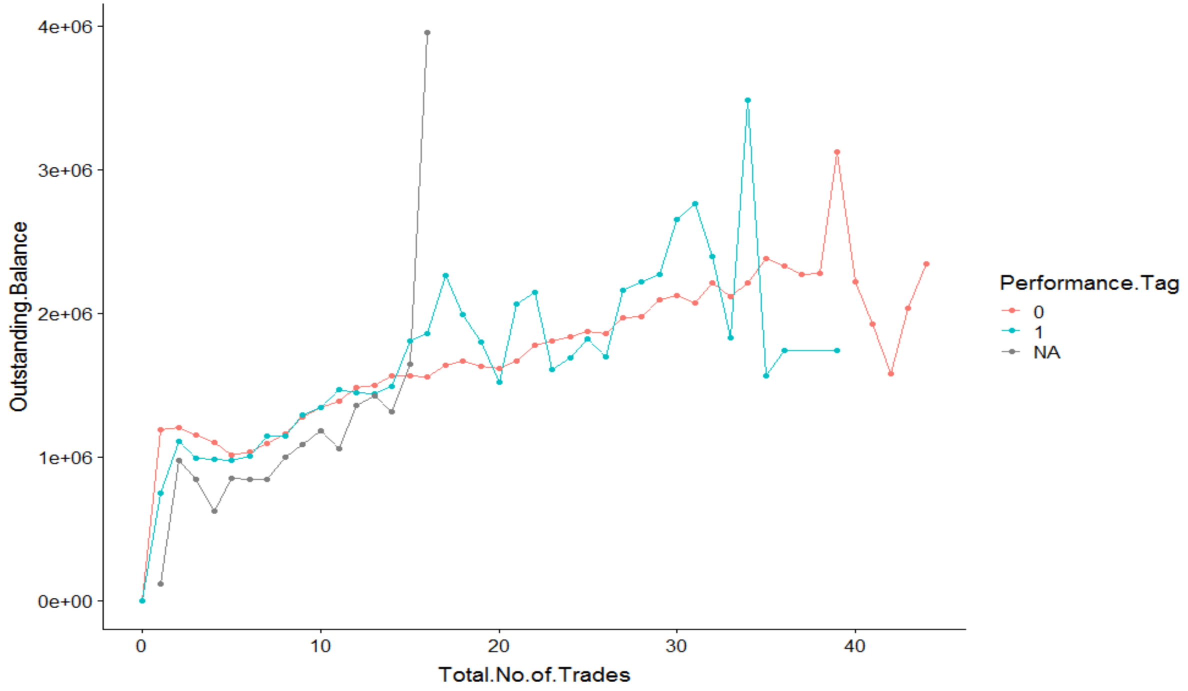
Exploratory Data Analysis – Analysis on merged data



Number of PL-trades opened is relatively higher for default users

Exploratory Data Analysis – Analysis on merged data



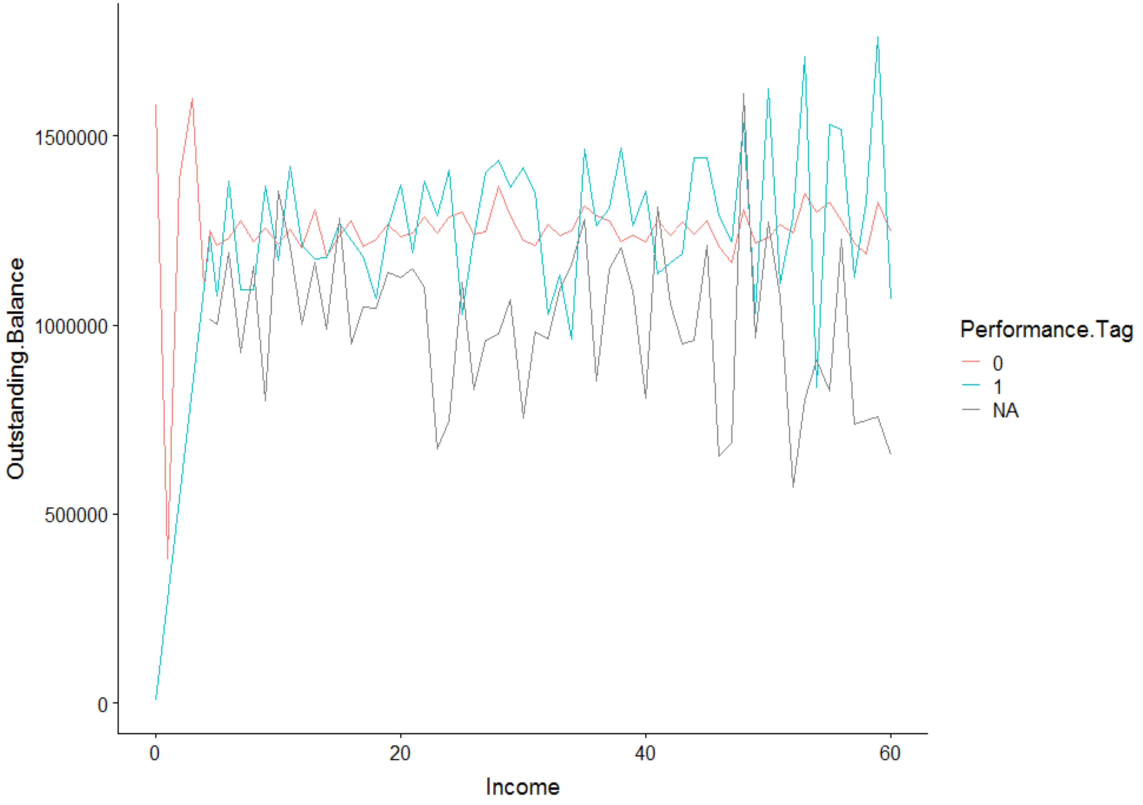


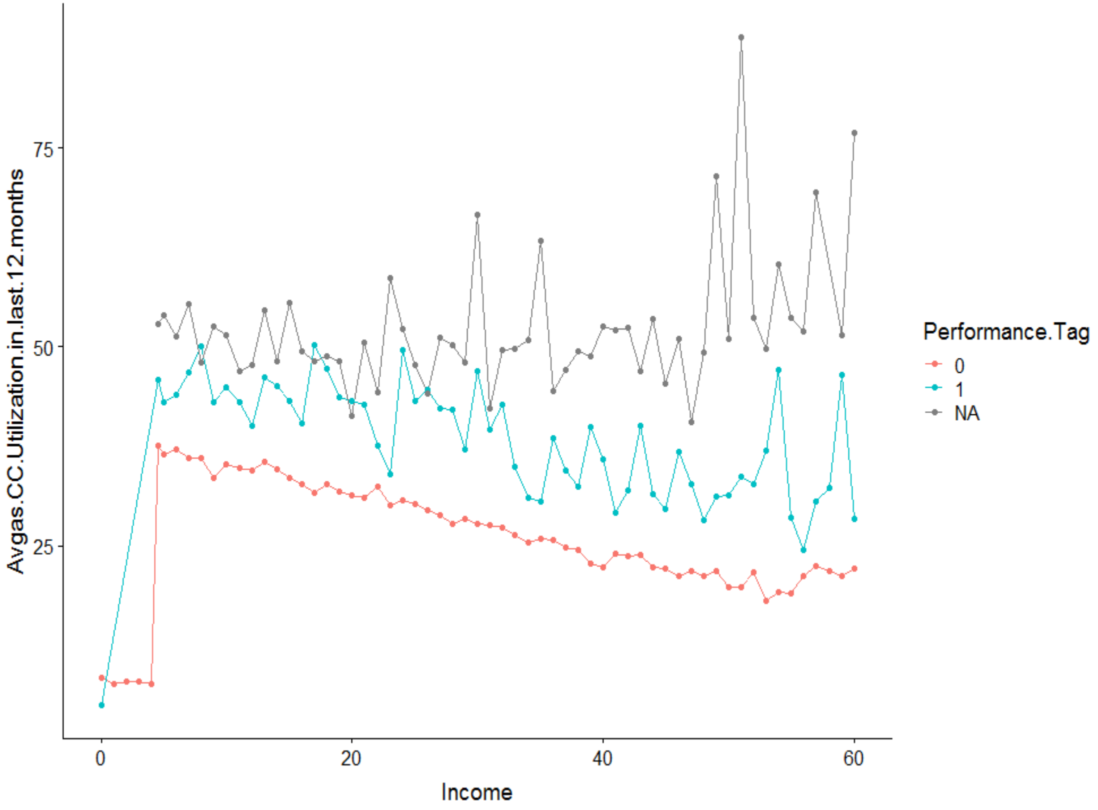
For default users Average-CC-utilization is overall higher , Also CC-usage is going high with increasing DPD values.

Total no of trades is overall in higher numbersfor default users.

Outstanding balance is relatively higher for most of default users.

Exploratory Data Analysis – Analysis on merged data

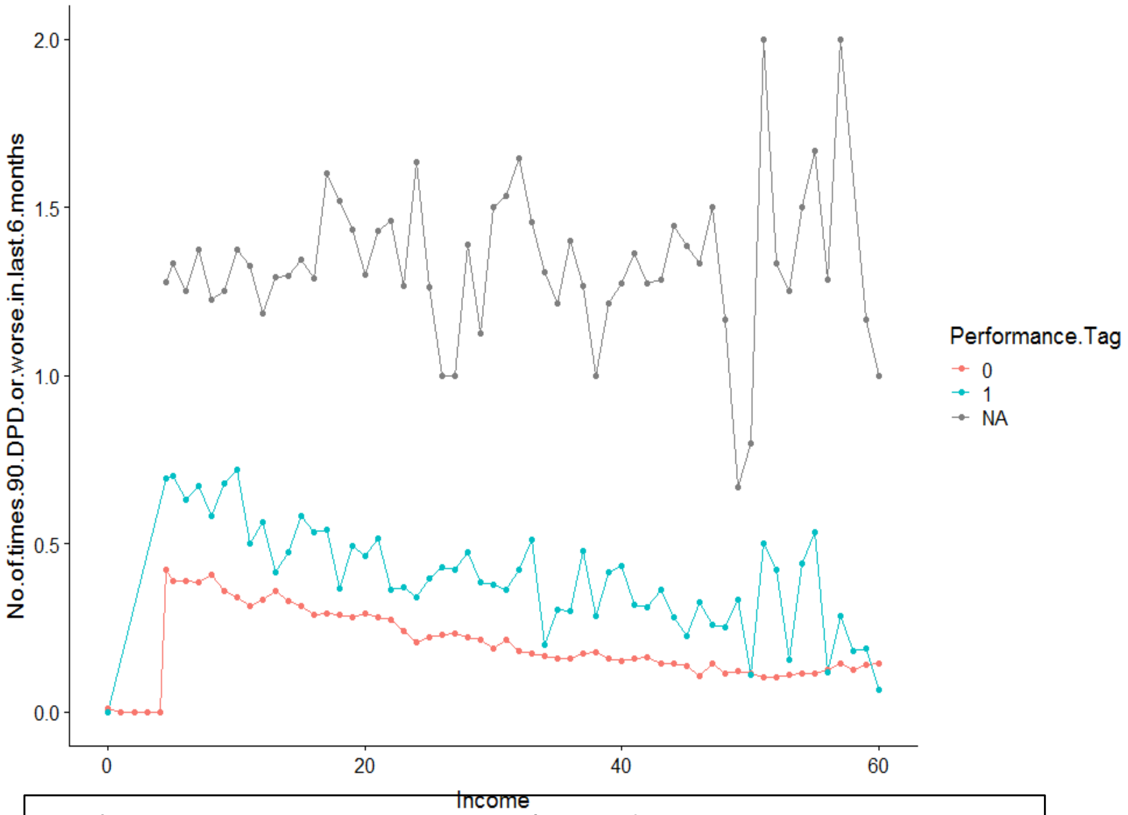


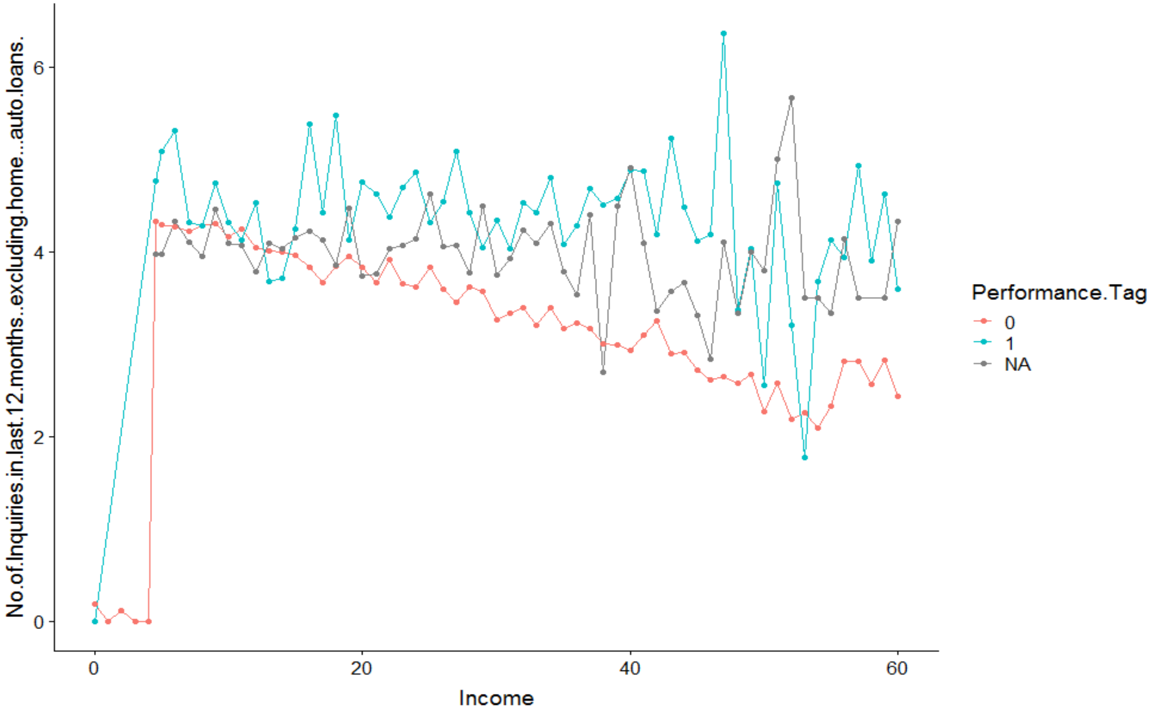


For defaulters Outstanding balance is higher.
No upward/downward trend for outstanding balance with increasing income.
If outstanding is more than 12.5lakh its a matter of concern.

With increasing income average-cc-usage decreases for whole population.
If average cc usage is >40 for a low income, >30 for middle income, >25 for higher income, they should be looked at.

Exploratory Data Analysis – Analysis on merged data

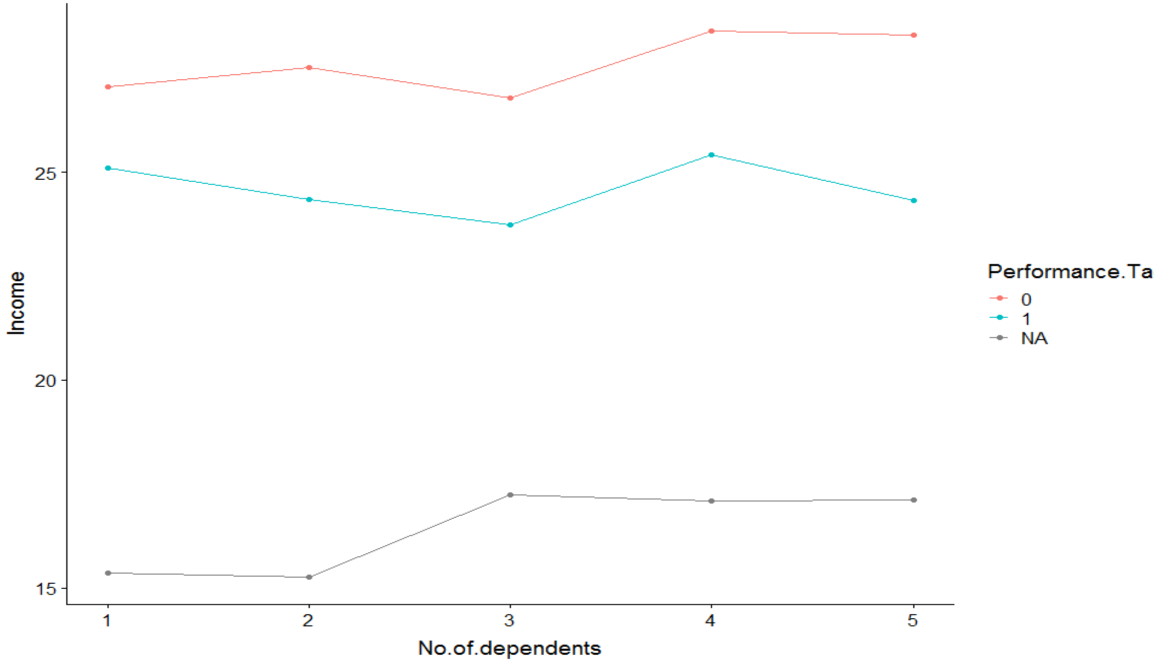


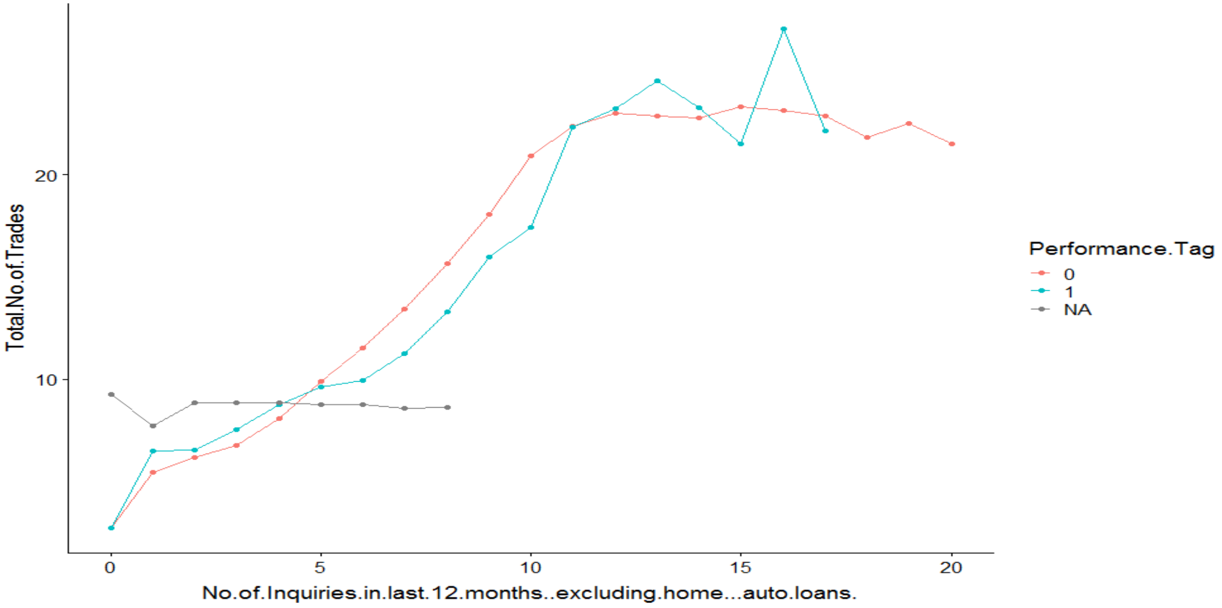


With increasing Income, DPD no's are decreasing.
Also for defaulting users DPD no's are way higher.
High no of defaulters are in lower to medium income range.

With increase in income no of inquiries are decreasing for non defaulters.
With increase in income no of inquiries relatively higher for defaulters.

Exploratory Data Analysis – Analysis on merged data

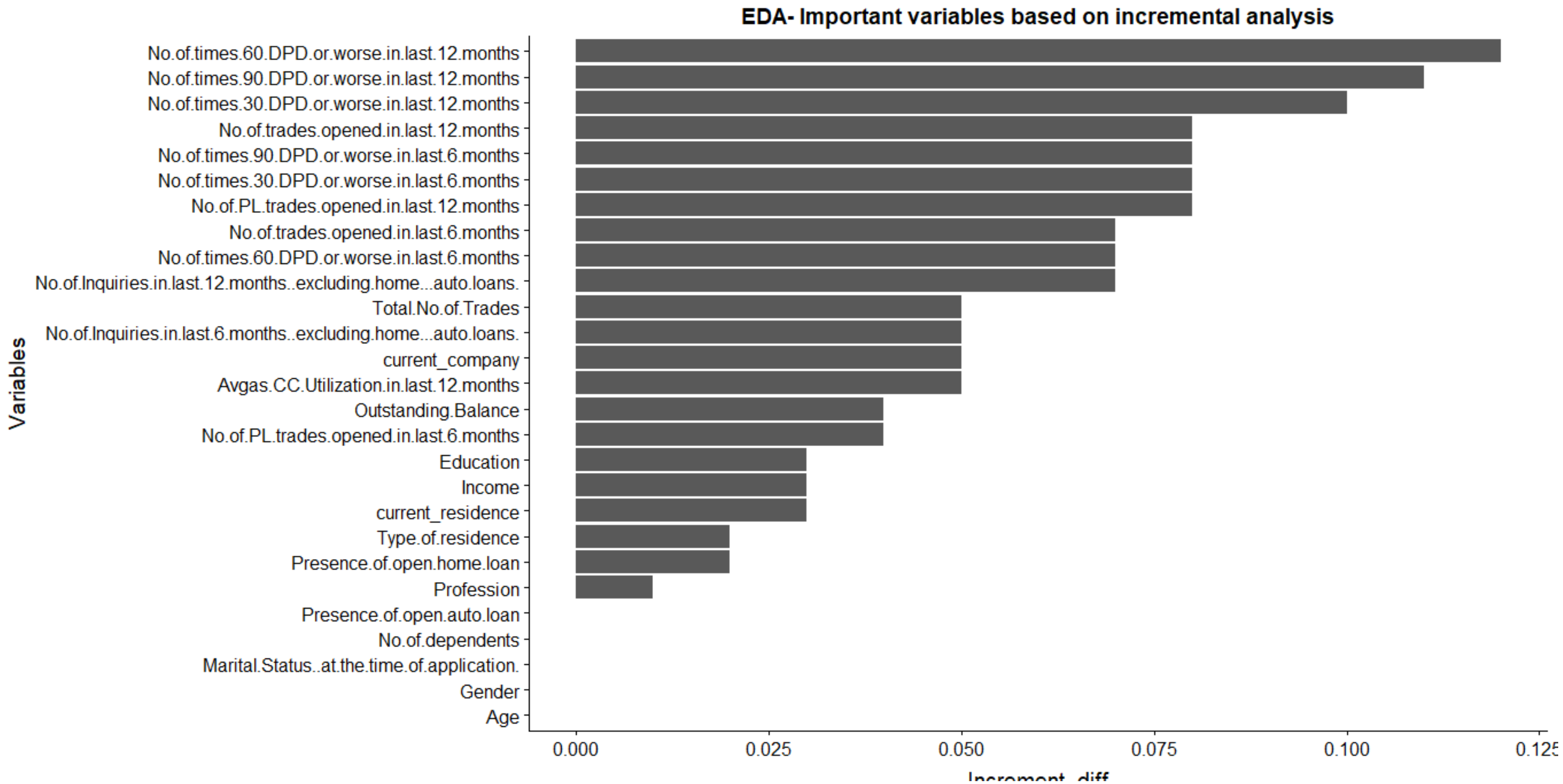




Income per no of dependants is very low for defaulters compared to non-defaulters.

With increasing no of inquiries in last 12months,
Total no of trades increases, then gradually it becomes constant.
For default users total no of trades is higher.

EDA based on Incremental Analysis



Information value of variables

```
                                              Variable           IV feedback
7                   Avgas.CC.Utilization.in.last.12.months 3.118158e-01   Strong
9                    No.of.trades.opened.in.last.12.months 2.992422e-01   Medium
11                No.of.PL.trades.opened.in.last.12.months 2.976330e-01   Medium
13 No.of.Inquiries.in.last.12.months..excluding.home...auto.loans. 2.965392e-01   Medium
15                                     Outstanding.Balance 2.469674e-01   Medium
3              No.of.times.30.DPD.or.worse.in.last.6.months 2.420549e-01   Medium
16                                      Total.No.of.Trades 2.378859e-01   Medium
10                 No.of.PL.trades.opened.in.last.6.months 2.203559e-01   Medium
4             No.of.times.90.DPD.or.worse.in.last.12.months 2.142245e-01   Medium
2              No.of.times.60.DPD.or.worse.in.last.6.months 2.062044e-01   Medium
12  No.of.Inquiries.in.last.6.months..excluding.home...auto.loans. 2.052807e-01   Medium
6             No.of.times.30.DPD.or.worse.in.last.12.months 1.987550e-01   Medium
8                     No.of.trades.opened.in.last.6.months 1.864486e-01   Medium
5             No.of.times.60.DPD.or.worse.in.last.12.months 1.858931e-01   Medium
1              No.of.times.90.DPD.or.worse.in.last.6.months 1.603274e-01   Medium
26                      No.of.months.in.current.residence 7.889157e-02     Weak
22                                                 Income 4.281923e-02     Weak
27                        No.of.months.in.current.company 2.213321e-02     Weak
14                             Presence.of.open.home.loan 1.732251e-02  Useless
18                                                    Age 3.427162e-03  Useless
21                                      No.of.dependents 2.779641e-03  Useless
24                                             Profession 2.007310e-03  Useless
17                             Presence.of.open.auto.loan 1.685850e-03  Useless
25                                      Type.of.residence 9.800132e-04  Useless
23                                              Education 7.581767e-04  Useless
19                                                 Gender 3.572072e-04  Useless
20              Marital.Status..at.the.time.of.application. 7.551032e-05  Useless
>
```

EDA Observations

- Age group between 40-55 tend to default most
- Males proportion of default is more than female
- Married applicants are more defaulters
- Applicants having Professional/ Masters degree has higher risk of defaulting
- Salaried default most of the times as they have higher frequency
- Rented one have high default chances
- High default rate between 3 to 7 inquiries
- 1 to 2 times passed the 90 days or worse have higher chances of default.
- For WOE and IV generation we would change the performance values from 0 to 1 and 1 to 0. because 1 in our dataset is a bad customer and 0 is a good customer.
- No demographic variables seem to be important for predicting default.
- These is no high correlation between features as maximum is 27 percent between number of months in current residence and avg credit utilization.

What we got to know from EDA ?

- The most important variables seem to be: as per this thumb rule

- # Useless if IV is < 0.02
- # Weak if IV is [0.02, 0.1)
- # Medium if IV is [0.1, 0.3)
- # Strong if IV is[0.3, 0.5) and suspicious thereafter

- Avgas.CC.Utilization.in.last.12.months
- No.of.trades.opened.in.last.12.months
- No.of.PL.trades.opened.in.last.12.months
- No.of.Inquiries.in.last.12.months..excluding.home...auto. Loans.
- Outstanding. Balance
- No.of.times.30.DPD.or.worse.in.last.6.months
- Total.No.of.Trades

Insights of EDA, Model Selection with its Evaluation

- Before Information value generation for variables, We would update the Performance Tag variable from 1 to 0 and 0 to 1.
- It is because, Information value treats 0 as wrong classification, but in our case 0 shown the good candidate.
- In the observation, the Information value of the Demographic variables is less than Credit Bureau Variables.
- For missing values treatment, Mean, median or mode is not required.
- Weight of evidence approach would replace the missing values by WOE values.
- Scaling or outlier treatment would also be handled by WOE values.
- We would be running the final model on all the columns replaced by WOE values along with the performance tag.
- ApplicationId would be removed from dataset while creating the model, since it is a primary key.
- We would split the data frame into two data frames Train and Test.
- With the final equation of logistic regression model, we would be able to determine the factors affecting the risk involved in customer acquisition.
- There is class imbalance in the data as default have only 4% share in the whole data.
- We would use the SMOTE package to simulate/generate random data to remove the class imbalance issue.

Data modeling

The two types of models We need to build are as follows:

- Demographic data model: Build a model to predict the likelihood of default using only the demographic data

- This will give you a good idea of the predictive power of the application data. Obviously, the final model will use the credit bureau data as well, though this model is an important part of understanding the predictive power of application data.

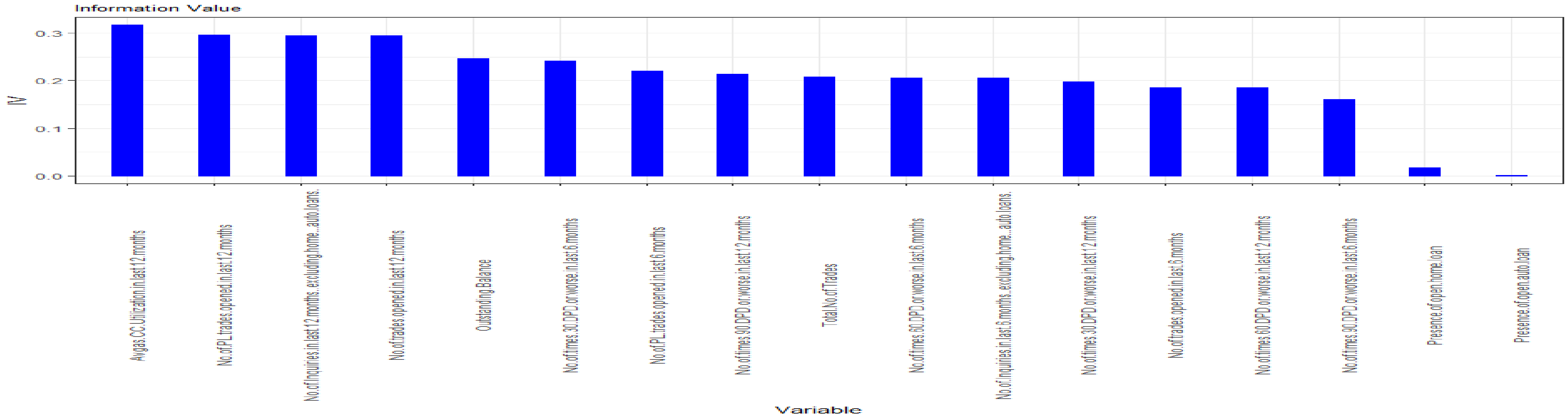
Logistic regression model

```
> confusionMatrix(y_pred,test$Performance.Tag)
Confusion Matrix and Statistics

          Reference
Prediction     0     1
         0 20059   902
         1     0     0

               Accuracy : 0.957
                 95% CI : (0.9541, 0.9597)
    No Information Rate : 0.957
    P-Value [Acc > NIR] : 0.5089

                  Kappa : 0
 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 1.000
            Specificity : 0.000
         Pos Pred Value : 0.957
         Neg Pred Value :   NaN
             Prevalence : 0.957
         Detection Rate : 0.957
   Detection Prevalence : 1.000
      Balanced Accuracy : 0.500

       'Positive' Class : 0
```

OBSERVATION

Accuracy : 0.9595
Sensitivity : 1.0000
Specificity : 0.0000

Very Bad model AS sensitivity is 1 and Specificity is 0.



Logistic regression model using both demographic and credit bureau with unbalanced data

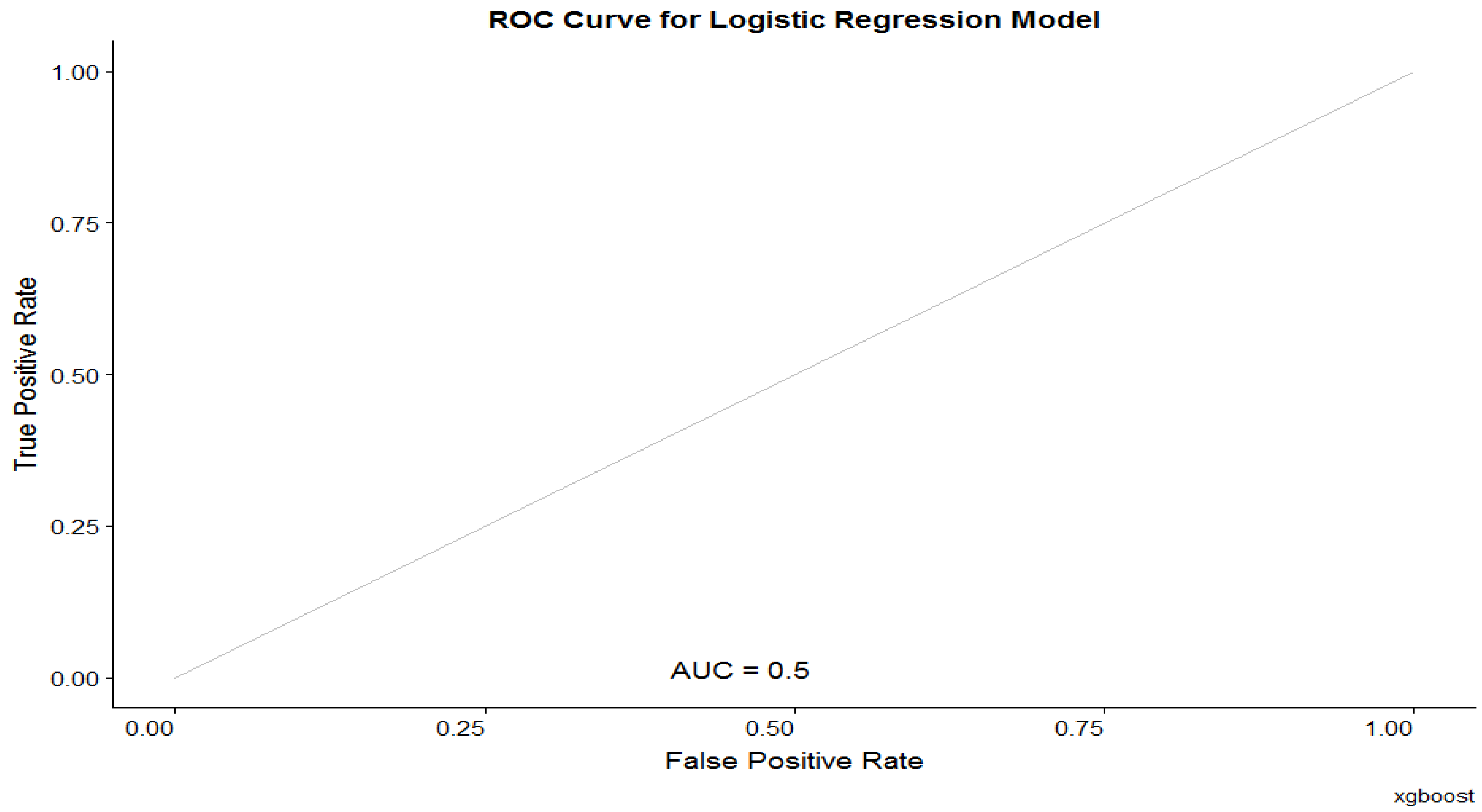
```
> conf_mtr_50_cutoff
Confusion Matrix and Statistics

          Reference
Prediction     0     1
         0 20078   884
         1     0     0

               Accuracy : 0.9578
                 95% CI : (0.955, 0.9605)
    No Information Rate : 0.9578
    P-Value [Acc > NIR] : 0.5089

                  Kappa : 0
 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 1.0000
            Specificity : 0.0000
         Pos Pred Value : 0.9578
         Neg Pred Value :    NaN
             Prevalence : 0.9578
         Detection Rate : 0.9578
   Detection Prevalence : 1.0000
      Balanced Accuracy : 0.5000

       'Positive' Class : 0
```

OBSERVATION

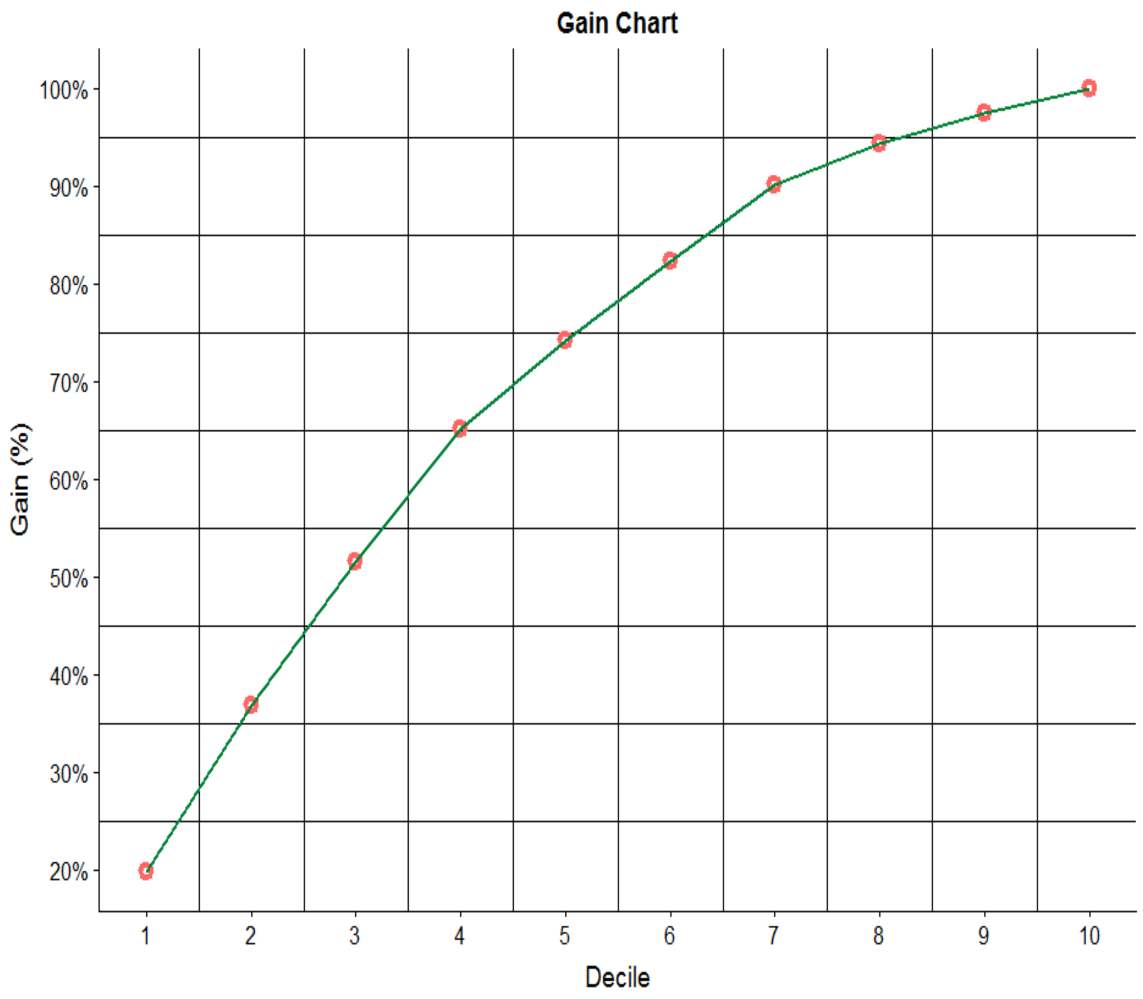
Accuracy : 0.9578
Sensitivity : 1.0000
Specificity : 0.0000

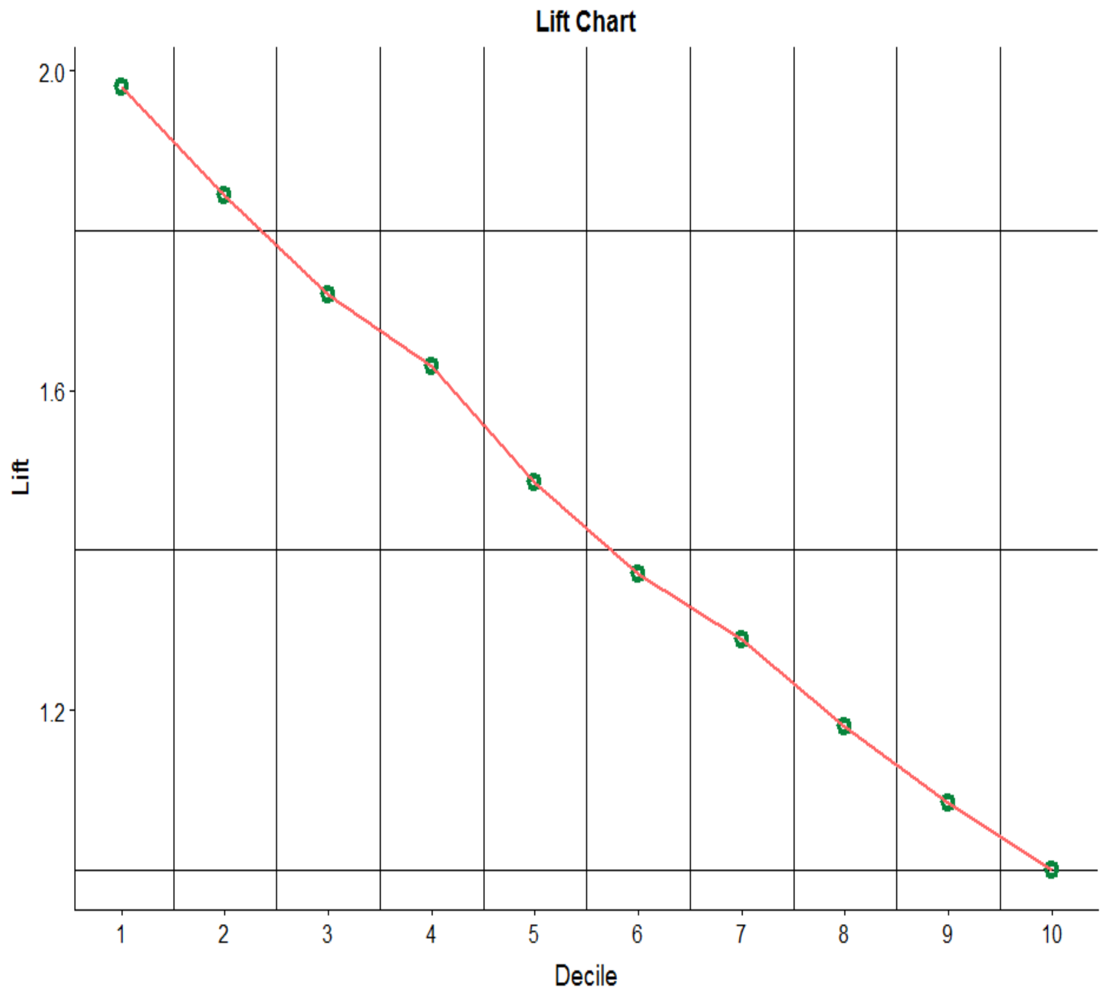
Logistic model is not good

ROC curve for logistic regression model



Plotting gain and lift chart





Building model for decision tree

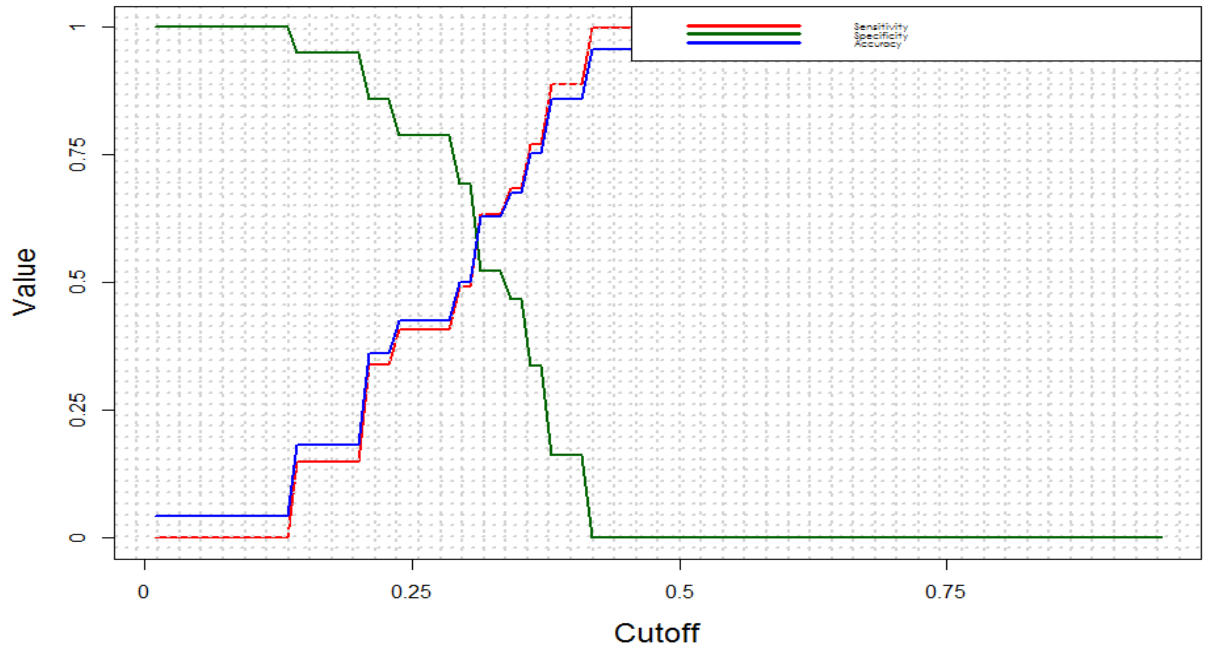
```
Confusion Matrix and Statistics

          Reference
Prediction     0     1
         0 20064   883
         1    14     1

               Accuracy : 0.9572
                 95% CI : (0.9544, 0.9599)
    No Information Rate : 0.9578
    P-Value [Acc > NIR] : 0.6801

                  Kappa : 8e-04
 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 1.131e-03
            Specificity : 9.993e-01
         Pos Pred Value : 6.667e-02
         Neg Pred Value : 9.578e-01
             Prevalence : 4.217e-02
         Detection Rate : 4.771e-05
   Detection Prevalence : 7.156e-04
      Balanced Accuracy : 5.002e-01

       'Positive' Class : 1
```

OBSERVATION

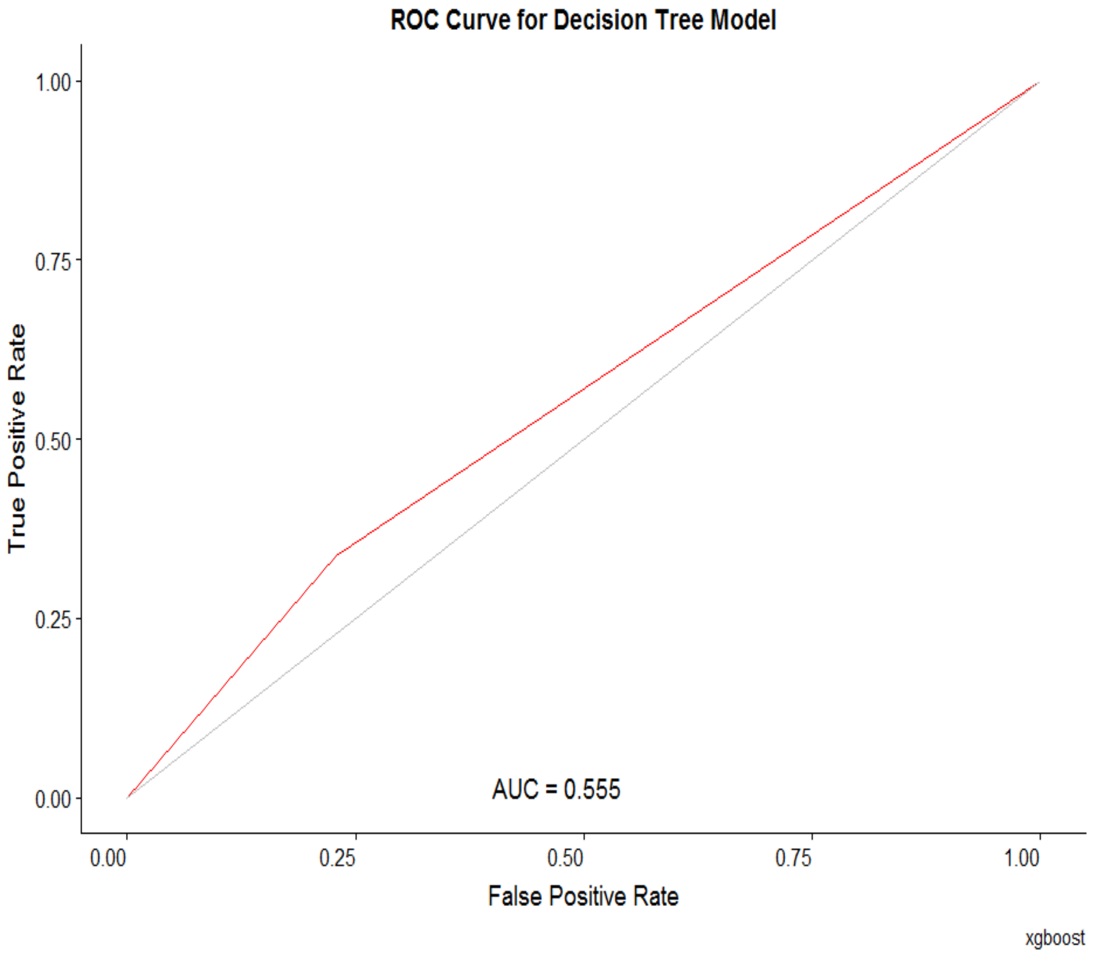
Accuracy : 0.9572

Sensitivity : 0.044118

Specificity : 0.975890



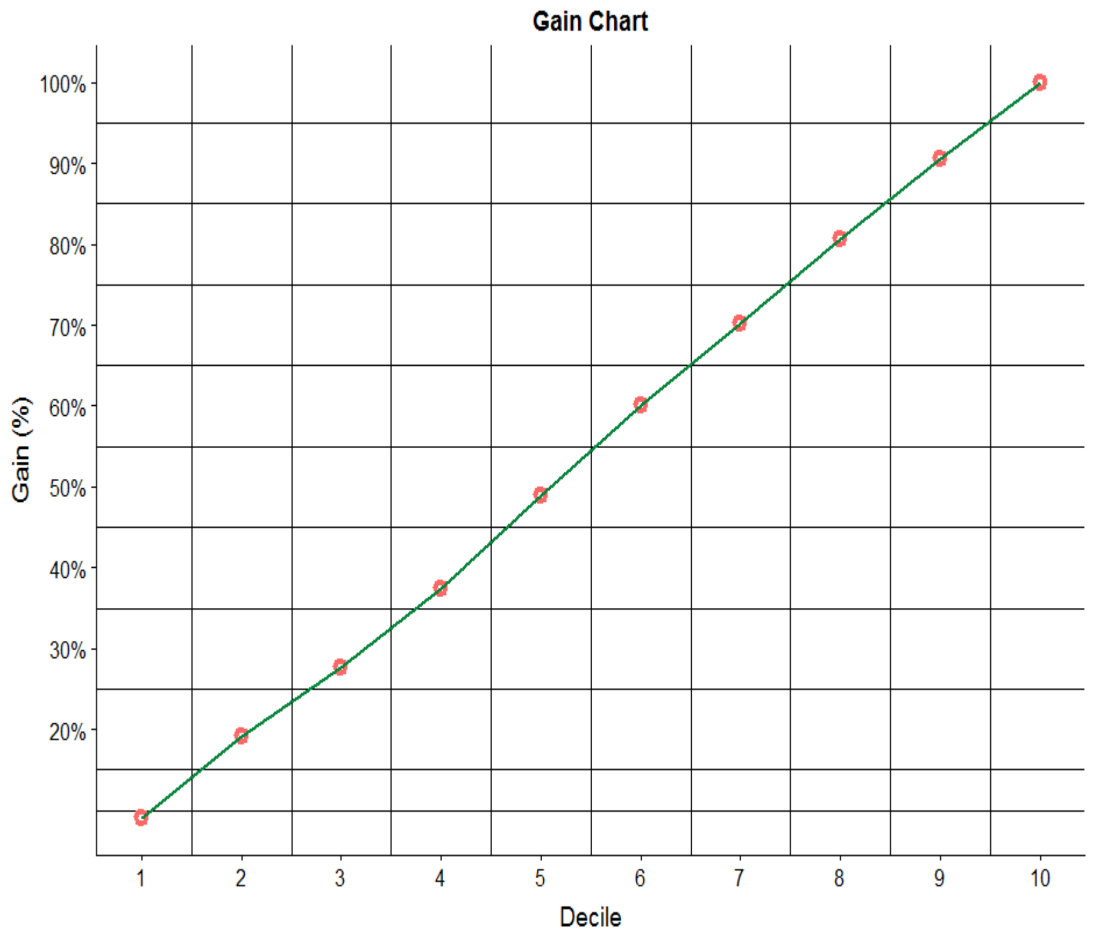
KS -statistic - Decision Tree - Test Data

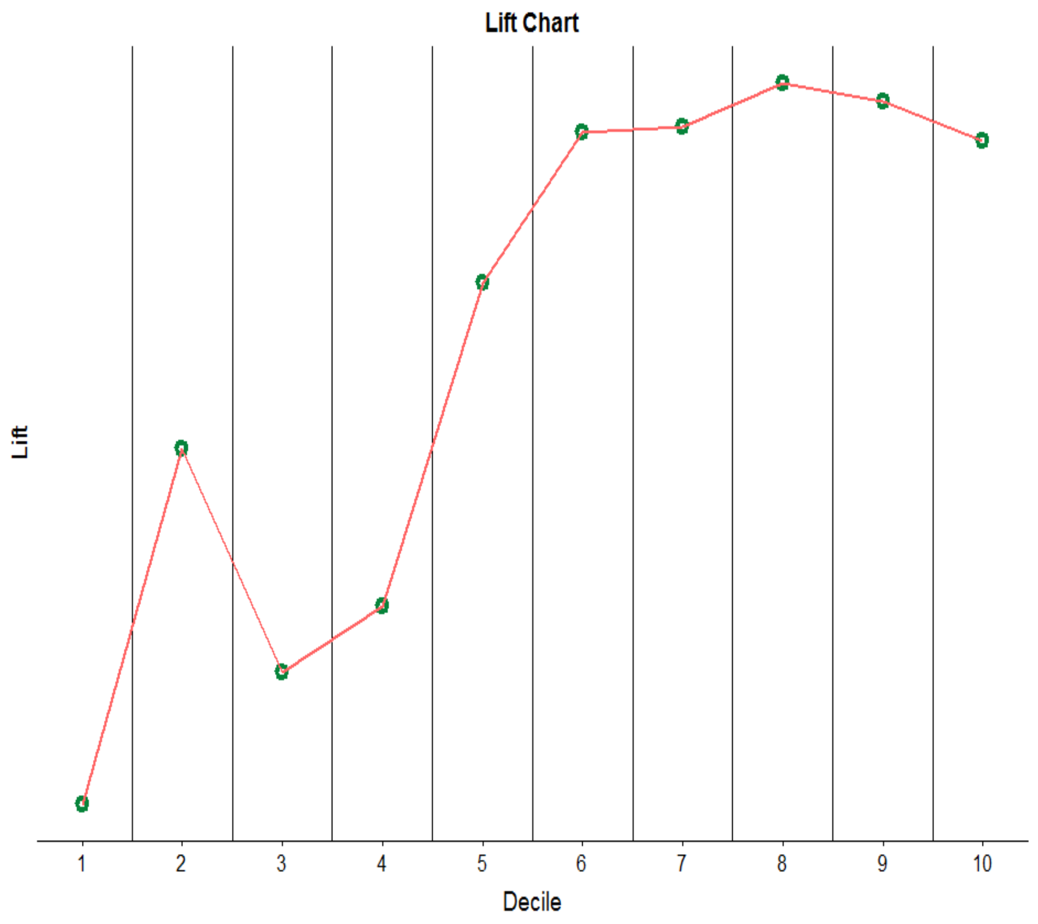


Ks-Statistics is 0%
Area under curve is :0.63
Gini:0.255077

bucket	total	totalresp	Cumresp	Gain	Cumlift
<int>	<int>	<dbl>	<dbl>	<dbl>	<dbl>
1	2097	80	80	9.05	0.905
2	2096	89	169	19.1	0.956
3	2096	76	245	27.7	0.924
4	2096	85	330	37.3	0.933
5	2096	103	433	49.0	0.980
6	2097	98	531	60.1	1.00
7	2096	89	620	70.1	1.00
8	2096	93	713	80.7	1.01
9	2096	87	800	90.5	1.01
10	2096	84	884	100	1

Plotting gain chart and lift chart for decision tree





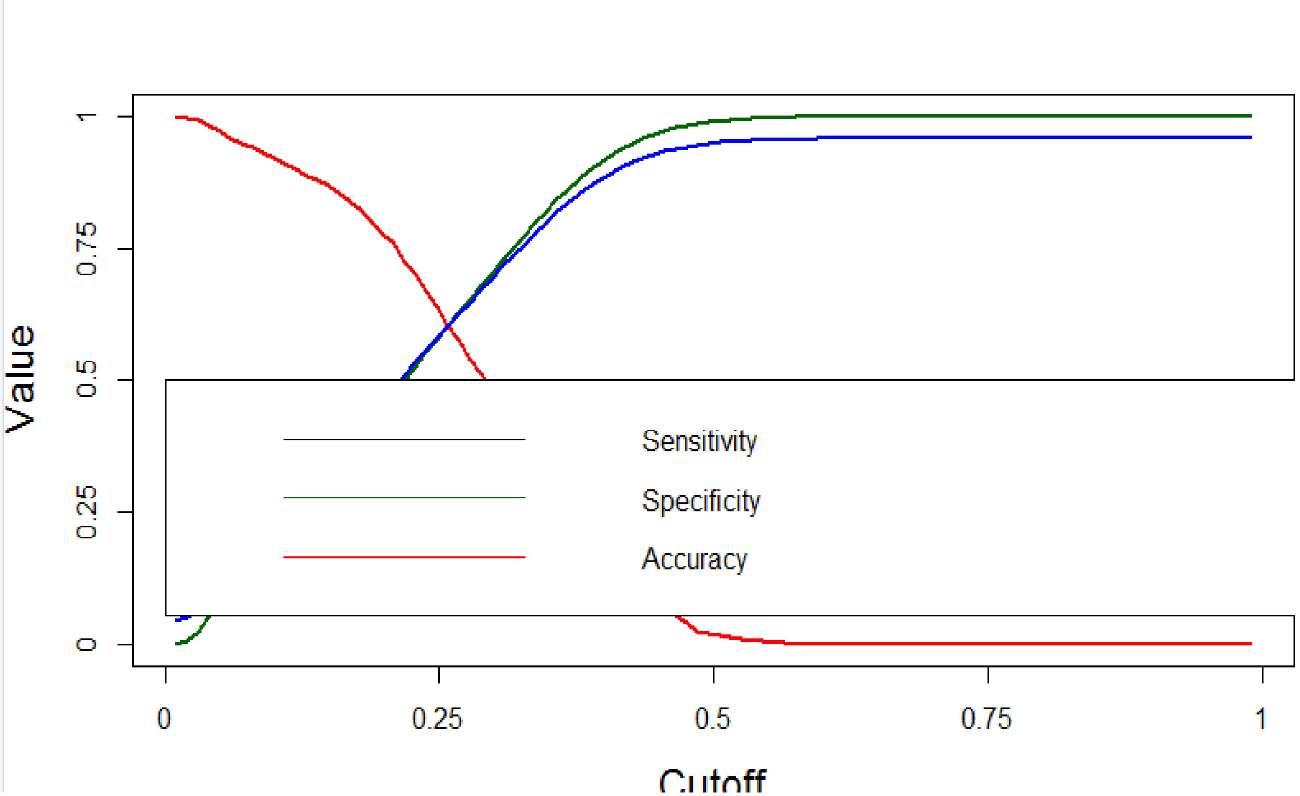
Random forest model

```
> summary(rf_FinalData)
                Length Class  Mode
call                7  -none- call
type                1  -none- character
predicted       17552  factor numeric
err.rate         3000  -none- numeric
confusion           6  -none- numeric
votes           35104  matrix numeric
oob.times       17552  -none- numeric
classes             2  -none- character
importance         27  -none- numeric
importanceSD        0  -none- NULL
localImportance     0  -none- NULL
proximity           0  -none- NULL
ntree               1  -none- numeric
mtry                1  -none- numeric
forest             14  -none- list
y               17552  factor numeric
test                0  -none- NULL
inbag               0  -none- NULL
terms               3  terms  call
```

```
Confusion Matrix and Statistics

          Reference
Prediction     0     1
         0 19872   868
         1   206    16

               Accuracy : 0.9488
                 95% CI : (0.9457, 0.9517)
    No Information Rate : 0.9578
    P-Value [Acc > NIR] : 1

                  Kappa : 0.0122
 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.0180995
            Specificity : 0.9897400
         Pos Pred Value : 0.0720721
         Neg Pred Value : 0.9581485
             Prevalence : 0.0421715
         Detection Rate : 0.0007633
   Detection Prevalence : 0.0105906
      Balanced Accuracy : 0.5039198

       'Positive' Class : 1
```

Accuracy : 0.9479
This is at a standard cut-off Reference Prediction
no 19851 868
yes 224 16
Sensitivity : 0.0180995
Specificity : 0.9888418

Random forest model cut offs

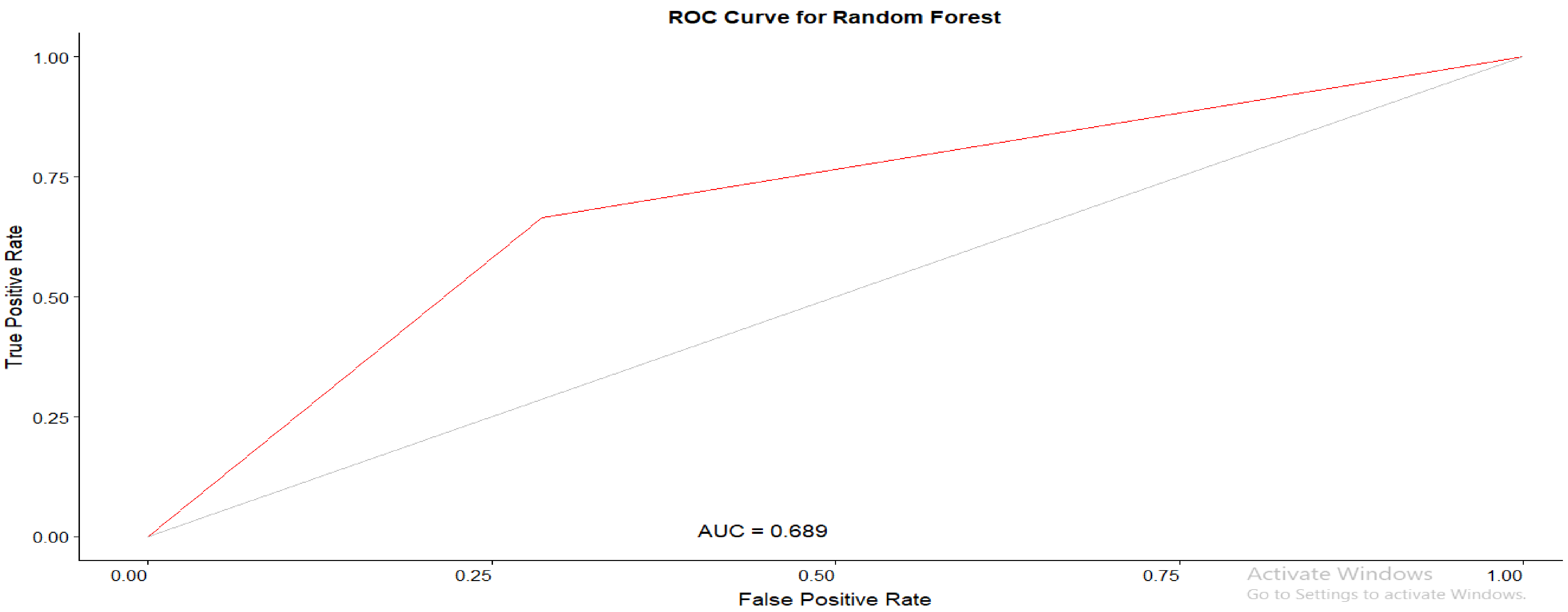



```
Confusion Matrix and Statistics

          Reference
Prediction     0     1
         0 12589   373
         1  7489   511

               Accuracy : 0.6249
                 95% CI : (0.6183, 0.6315)
    No Information Rate : 0.9578
    P-Value [Acc > NIR] : 1

                  Kappa : 0.0423
 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.57805
            Specificity : 0.62700
         Pos Pred Value : 0.06387
         Neg Pred Value : 0.97122
             Prevalence : 0.04217
         Detection Rate : 0.02438
   Detection Prevalence : 0.38164
      Balanced Accuracy : 0.60253

       'Positive' Class : 1
```

The plot shows that cut off value of around 17.8% optimises sensitivity and accuracy
The cut off is too low.

Accuracy : 0.6276
Sensitivity : 0.62330
Specificity : 0.62780

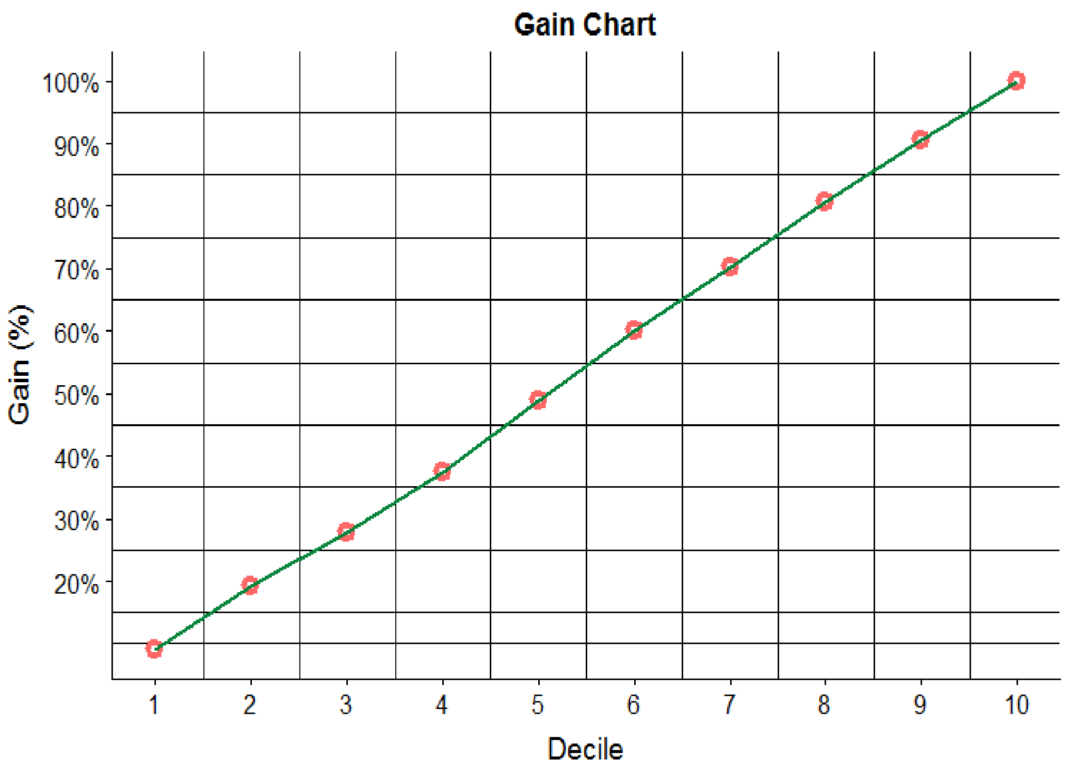
Roc curve for random forest

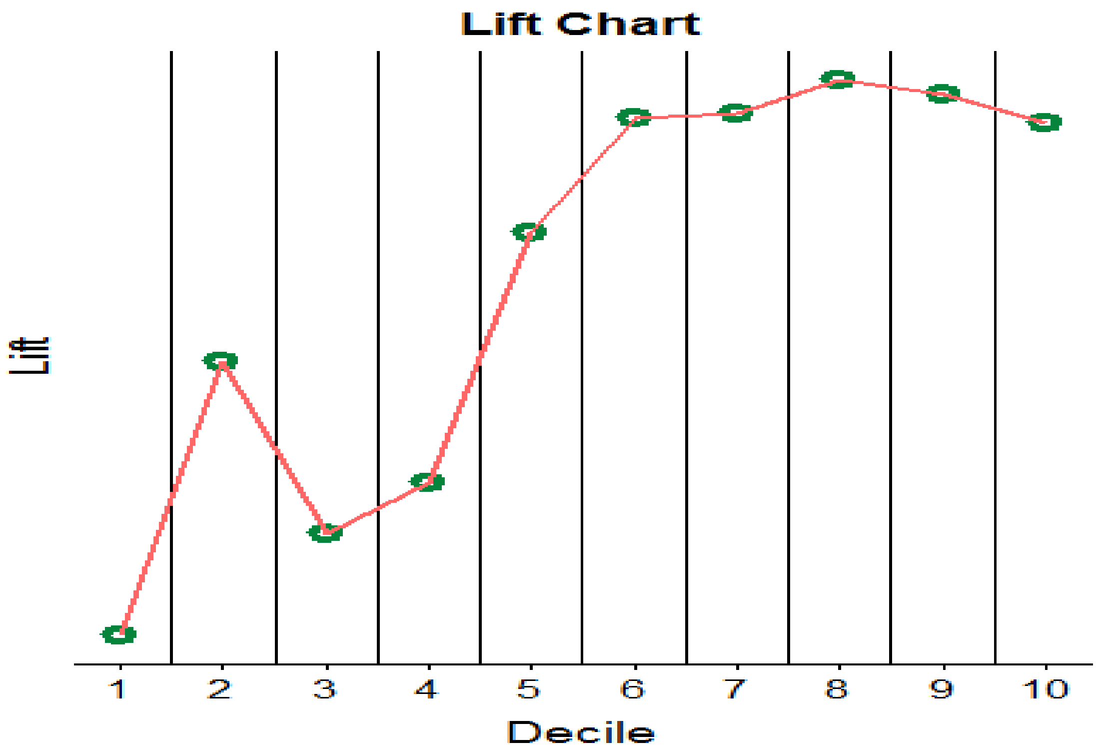


Area under curve is:0.6255495

Gini : 0.251

Plotting gain chart and lift chart for random forest





Confusion matrix and statistics

```
Confusion Matrix and Statistics

          Reference
Prediction   no  yes
       no     0    0
       yes   51 1374

               Accuracy : 0.9642
                 95% CI : (0.9532, 0.9732)
    No Information Rate : 0.9642
    P-Value [Acc > NIR] : 0.5372

                  Kappa : 0
 Mcnemar's Test P-Value : 2.534e-12

            Sensitivity : 1.0000
            Specificity : 0.0000
         Pos Pred Value : 0.9642
         Neg Pred Value :    NaN
             Prevalence : 0.9642
         Detection Rate : 0.9642
   Detection Prevalence : 1.0000
      Balanced Accuracy : 0.5000

       'Positive' Class : yes

> summary(as.factor(test_pred_optimal))
  no  yes
  51 1374
```

95.23 percent of the Defaulters detected very well which is very good

Assessing the financial benefit of project

- 2.5 %(36/1425) of rejected candidates rejected by the bank are accepted by the model
- 15.37 %(10689/69864) of candidates accepted by the bank are rejected by the model.
- Our model has rejected 15.37% of selected candidates.
- We need to find out how many candidates rejected by the model have defaulted.
- 2947/69864 = 4.21%

Credit loss saved

- Total number of candidates selected by the bank but defaulted - 10689/69864 = 15.29%
- No of candidates selected by the model and who defaulted - 548
- No of candidates selected by the model - 56228
- % of candidates selected by the model and defaulted - 548/69864 = 0.78%
- % of employees selected and defaulting before model= 4.21%
- % of employees selected and defaulting after model=0.78% .
- Credit loss saved = 3.43%

Revenue loss

- Count of candidates rejected by the model who didn't default - 10689
- Total count of candidates who didn't default - 66917
- Percentage of good candidates rejected by our model - 15.97%.
- So 15.97% percent is the revenue loss where we have identified good customers as bad .

Conclusion

Important variables that can be used to identify good customers from Random Forest Model:
Avgas.CC.Utilization.in.last.12.months Outstanding. Balance
No.of.times.30.DPD.or.worse.in.last.6.months
No.of.times.90.DPD.or.worse.in.last.12.months
No.of.times.60.DPD.or.worse.in.last.6.months
No.of.Inquiries.in.last.6.months..excluding.home...auto. Loans.

These variables are used while inquiring about customer before giving them loan.

Thus We can conclude that the model has accurately predicted approximately 84% of the performance Tag in the dataset.
After tuning the model we have achieved the required accuracy of the model and now model can be used to predict whether who will default and who will not default.
This could save a lot of resources of bank at the same time increasing efficiency.
With the help of this model we found out that credit loss % was decreased when we used this model from 4 to less than 1.
Model has performed accurately in rejecting the candidate who may default in future.
This can save a lot of hours, money of the bank and at the same time increase the efficiency and resources of the bank.