CredX Credit Card Acquisition Analytics Project

Manjula Raman

Srinivas Panganamala

Vishnu Das

Objectives

- To create strategies to mitigate the acquisition risk
  - Use past data of the bank’s applicants and credit bureau data and determine the factors affecting credit risk
  - Use predictive models to measure the probability of default
  - Assign Application scores for each applicant and use it as a basis to grant or reject credit
  - Financial Benefit assessment of the model

Key Steps

- Data Understanding and Cleaning
- Exploratory Data Analysis
  - Univariate Analysis
  - Bivariate Analysis
  - WOE and IV Analysis
- Modelling
  - Data Balancing
  - Modelling using only demographic data
    - Logistic Regression
  - Modelling using complete data (demographic and credit bureau data)
    - Logistic Regression
    - Binary trees
    - Random Forests
- Model Evaluation
- Financial Benefit Assessment

Data Understanding and Cleaning

- Data is from 2 csv files demographic_data and credit_bureau data
    - Demographic data relates to applicant’s personal data E.g. Age, Income, Education, months in current company and residence Etc
    - Credit bureau data relates to applicant’s credit history E.g. Total Trades, Presence of Auto/Home loans, PL trades, Outstanding Balance, Utilization etc
- Merging by Application.ID
    - Two common columns in both the files are 1) Application.ID 2) Performance. Tag (1- default; 0- non-default) . The files are merged by Application.ID to create the master_frame
- Duplicate values corrected
    - Duplicate numbers in Application.ID corrected by assigning new Application.IDs
- Filtering out applicants who were denied credit
    - Performance.Tag had 1425 rows with NA in demographic and credit bureau data. These correspond to the applicants who are rejected; These rows are filtered out

Data Understanding and Cleaning

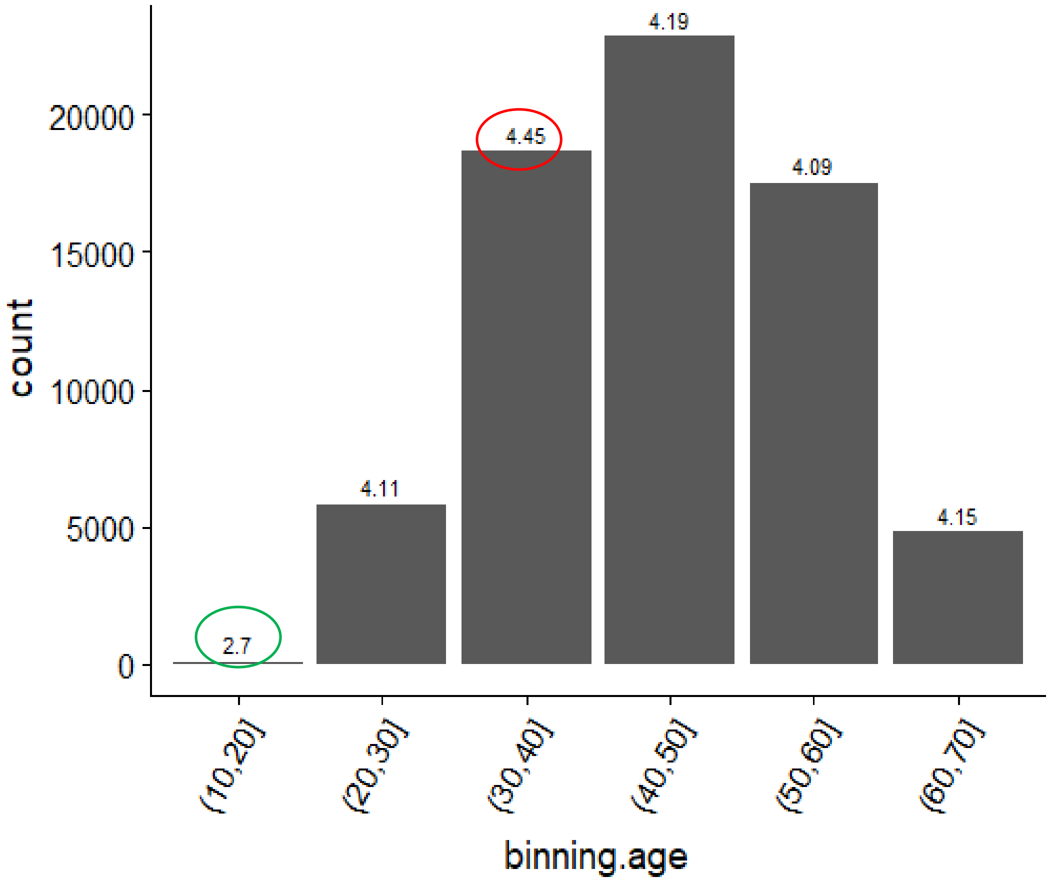
- Blank/NA value handling
  - Blank/NA values imputed for columns having few missing entries by assigning the value of majority of applicants
  - However ,the following 4 columns have significant missing data
    - Education, Utilization, Presence.of.open.home.loan and Outstanding.balance
    - Using the “information” package, WOE was generated
    - Missing data replaced with the value whose WOE was closest to the missing data WOE
- Incorrect data handling
  - Applicants with age less than 18 dropped
  - 107 rows having negative and zero income dropped
  - All values of Avgas.CC.Utilization.in.last.12.months > 100% capped at 100%

Exploratory Data Analysis :Univariate Analysis

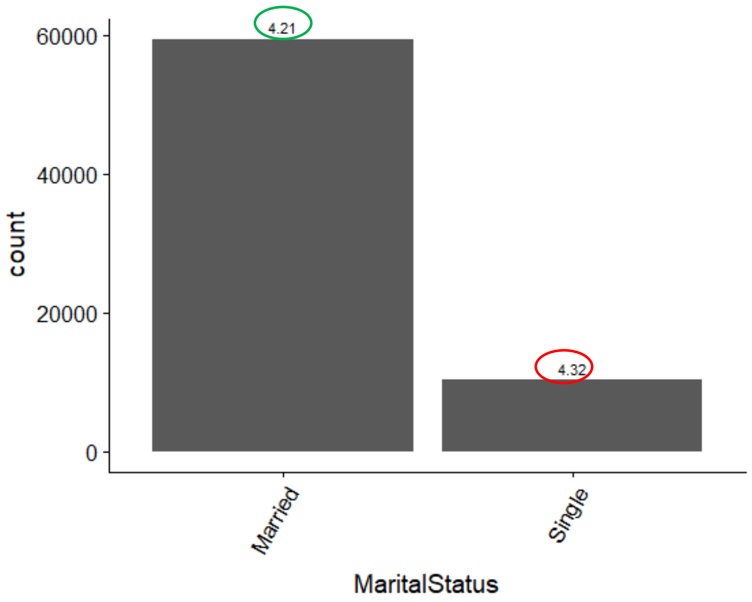
The independent variables are classified Numerical and categorical variables as follows

- <u>Numeric Variables</u>
  - Age
  - No. of.dependents
  - Income
  - No.of.months.in.current.residence
  - No.of.months.in.current.company
  - No.of.times.90.DPD.or.worse.in.last.6.months
  - No.of.times.60.DPD.or.worse.in.last.6.months
  - 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.12.months
  - No.of.times.30.DPD.or.worse.in.last.12.months
  - Avgas.CC.Utilization.in.last.12.months
  - No.of.trades.opened.in.last.6.months
  - No.of.trades.opened.in.last.12.months
  - No.of.PL.trades.opened.in last.6.months
  - No.of.PL.trades.opened.in.last.12.months
  - No.of Inquiries.in.last.6.months..excluding.home…auto.loans.
  - No.of Inquiries.in.last.12.months..excluding.home…auto.loans.
  - Outstanding.Balance
  - Total.No.of.Trades
- <u>Categorical Variables</u>
  - Gender
  - Marital.Status..at.the.time.of.application
  - Education
  - Profession
  - Type.of.residence
  - Presence.of.open.home.loan
  - Presence.of.open.auto.loan

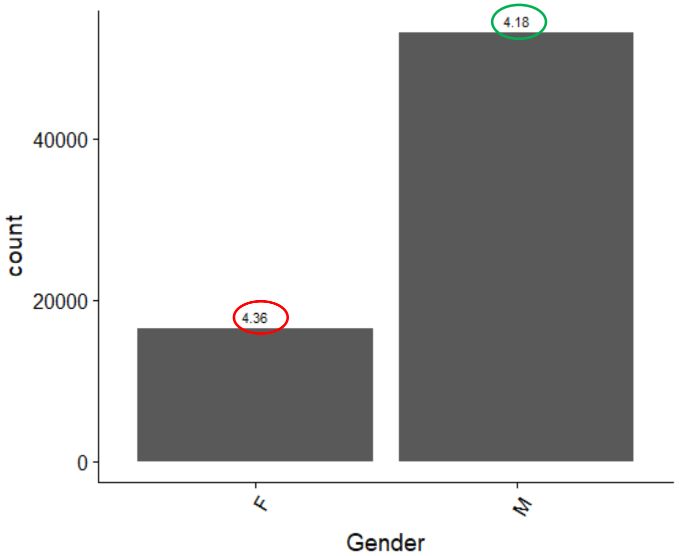
Age, Gender and Marital Status



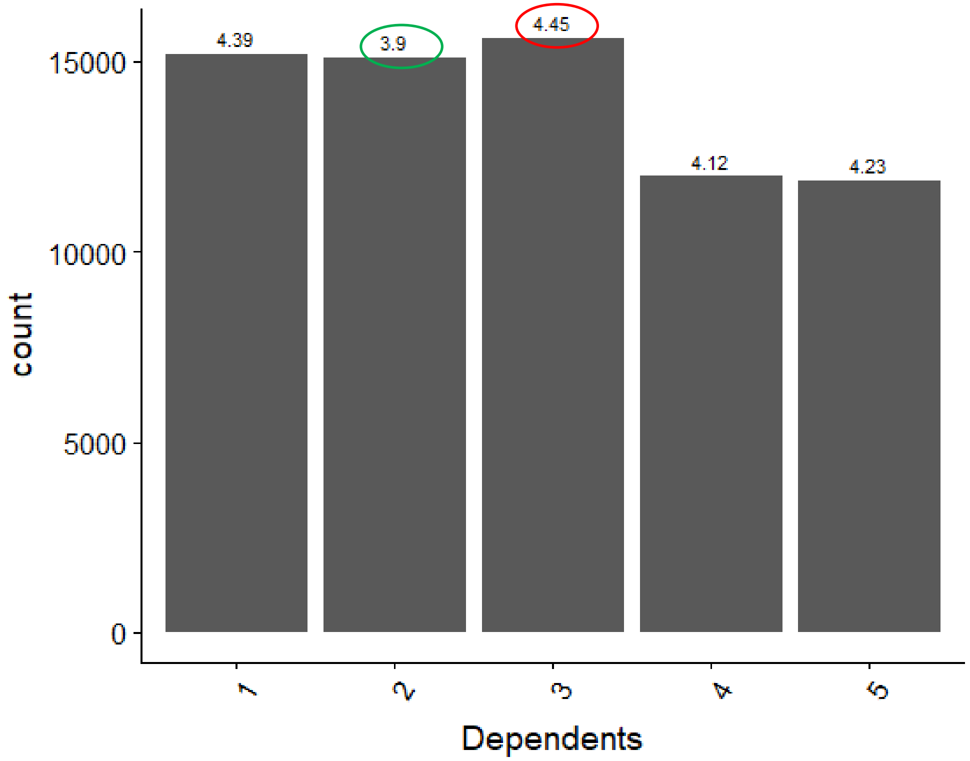
Highest default rate: 30-40 bin (4.45%)
Lowest default rate : 10-20 bin (2.7%)



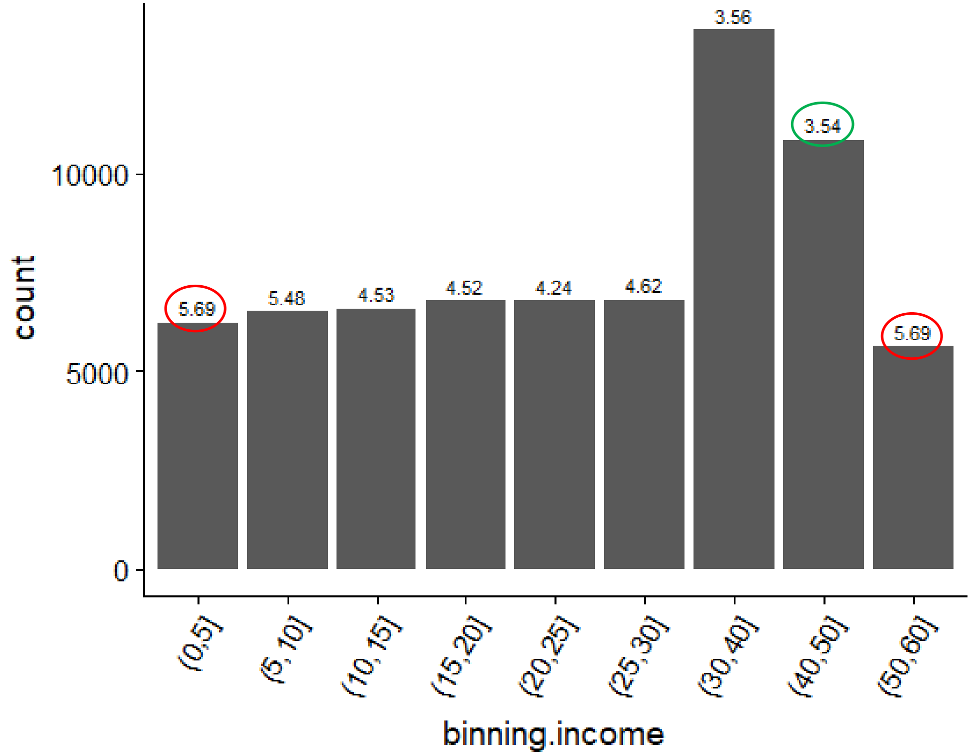
Marital Status and Gender are not significant influencers.



Number of dependents & Income

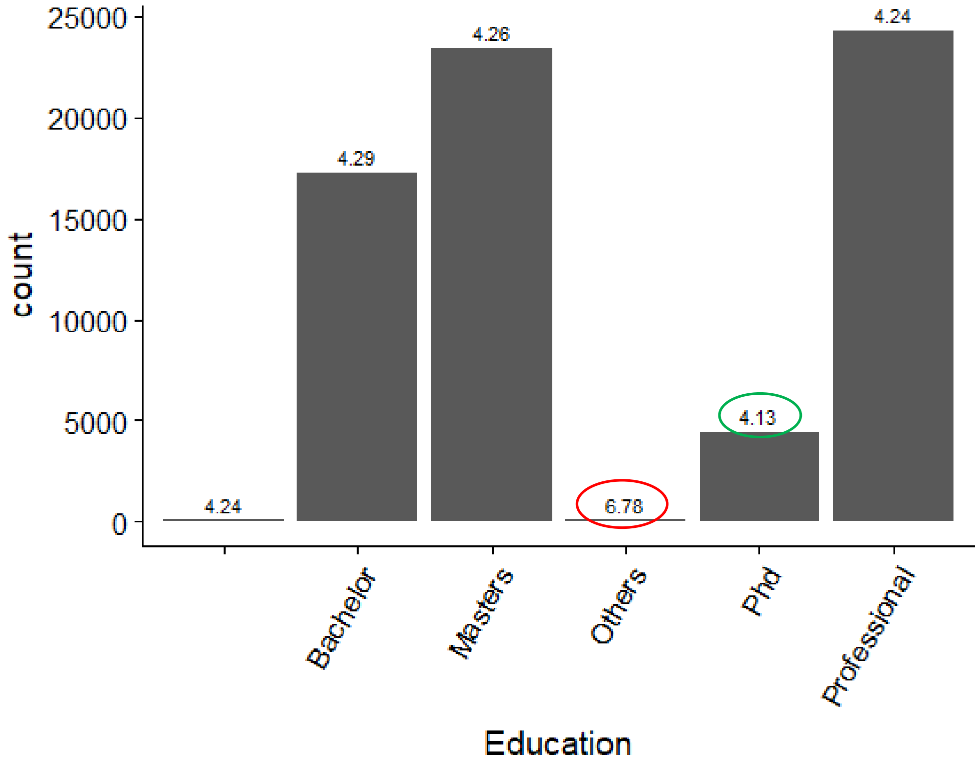


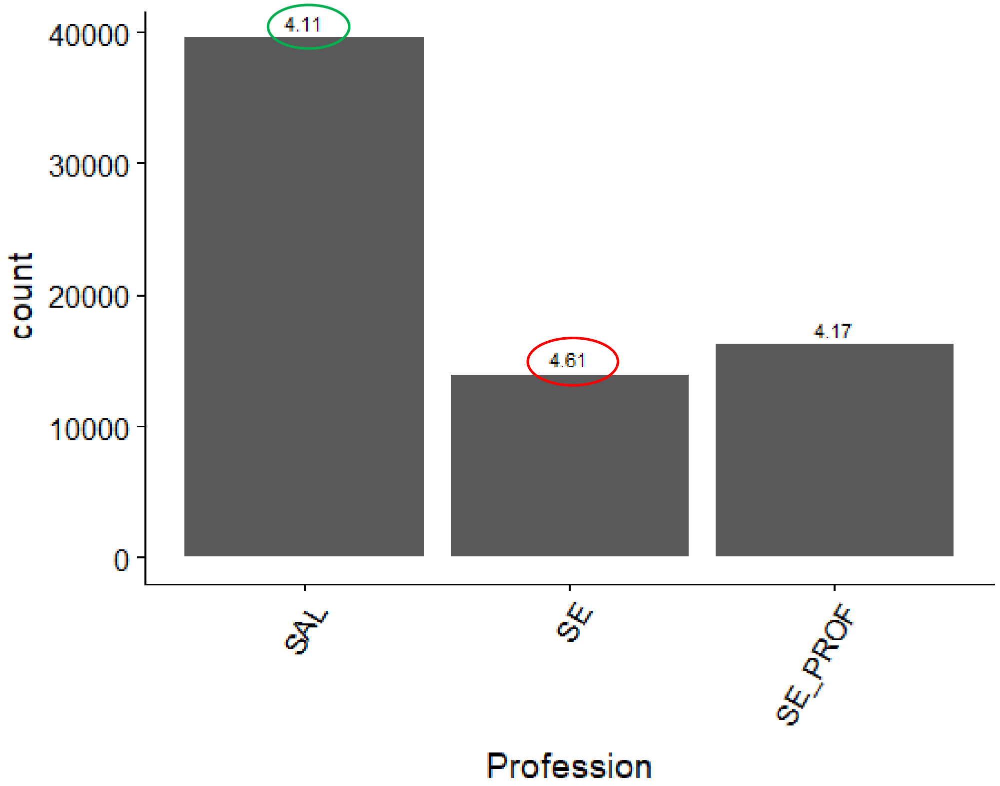
The variable has medium significance
Highest default rate: 3 dependents(4.45%)
Lowest default rate: 2 dependents(3.89%)



Income is an important factor.
Trend : Default rate decreases as income increases.
Highest default rate 5-10 (5.688%) and 50-60 (5.688%)
(highest income bin is also likely to default)
Lowest default rate: Income range 30-40(3.542%)

Education & Profession

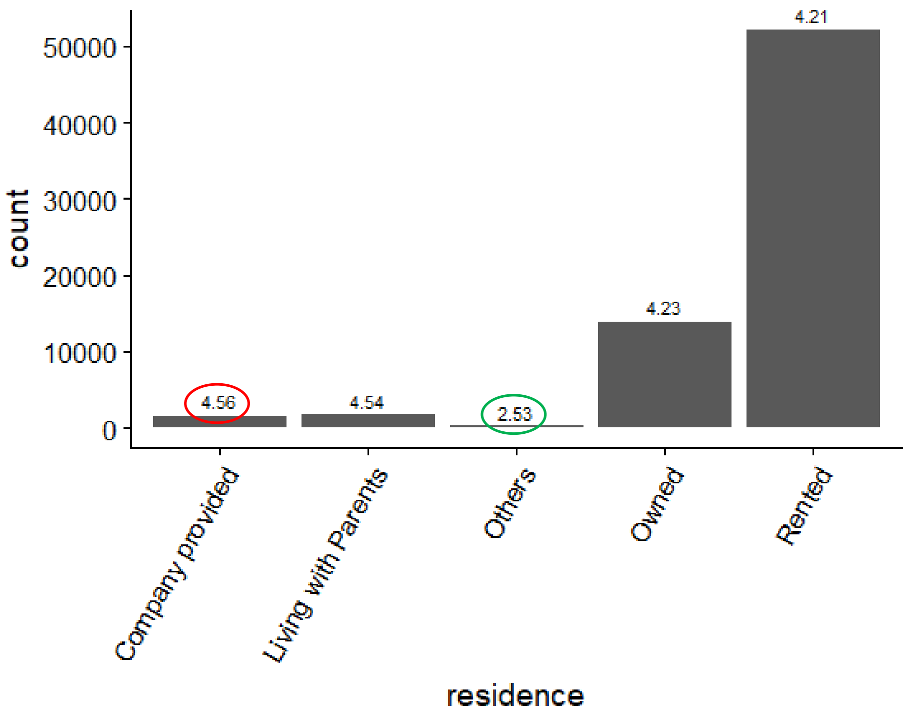




Highest default rate: "Others" (6.77%)
Lowest default rate: "Professional" (4.23%)
The NA group has a default rate similar to the Professional group(4.24%).

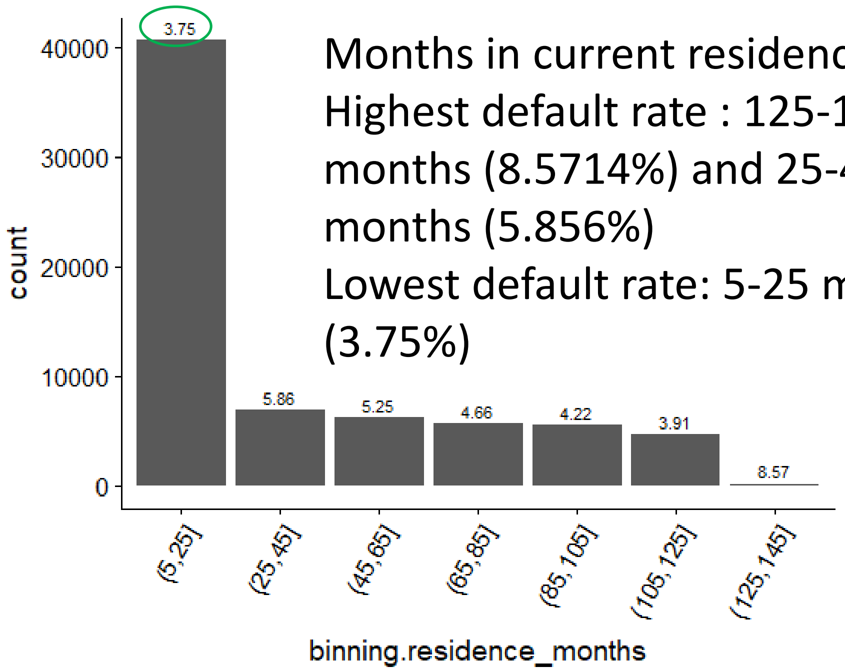
There is not much variation in Profession.
Highest default rate: SE (4.612%)
Lowest default rate: SAL (4.11%)

Type.of.residence , No.of.months.in.current.residence No.of.months.in.current.company

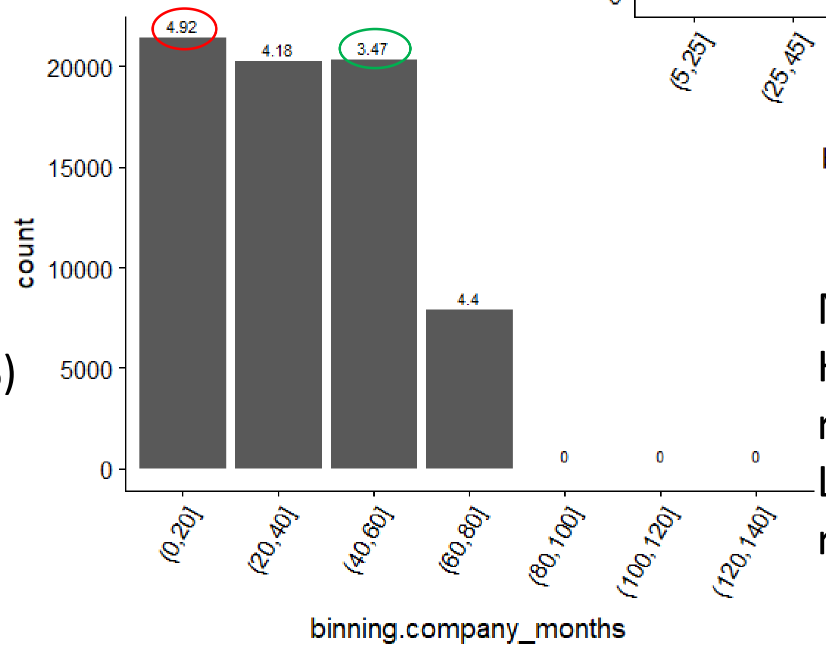


Highest default rate: Company provided (4.5625%)

Lowest default rate: Others (2.525%)

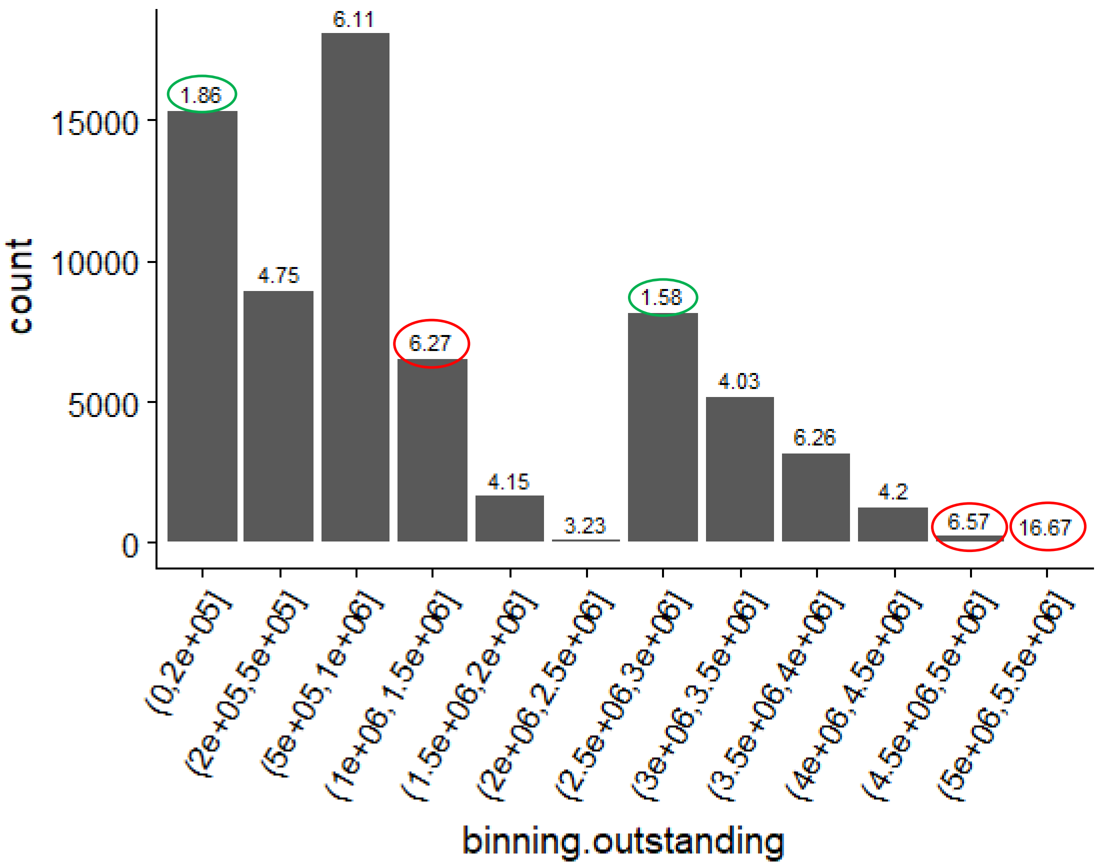


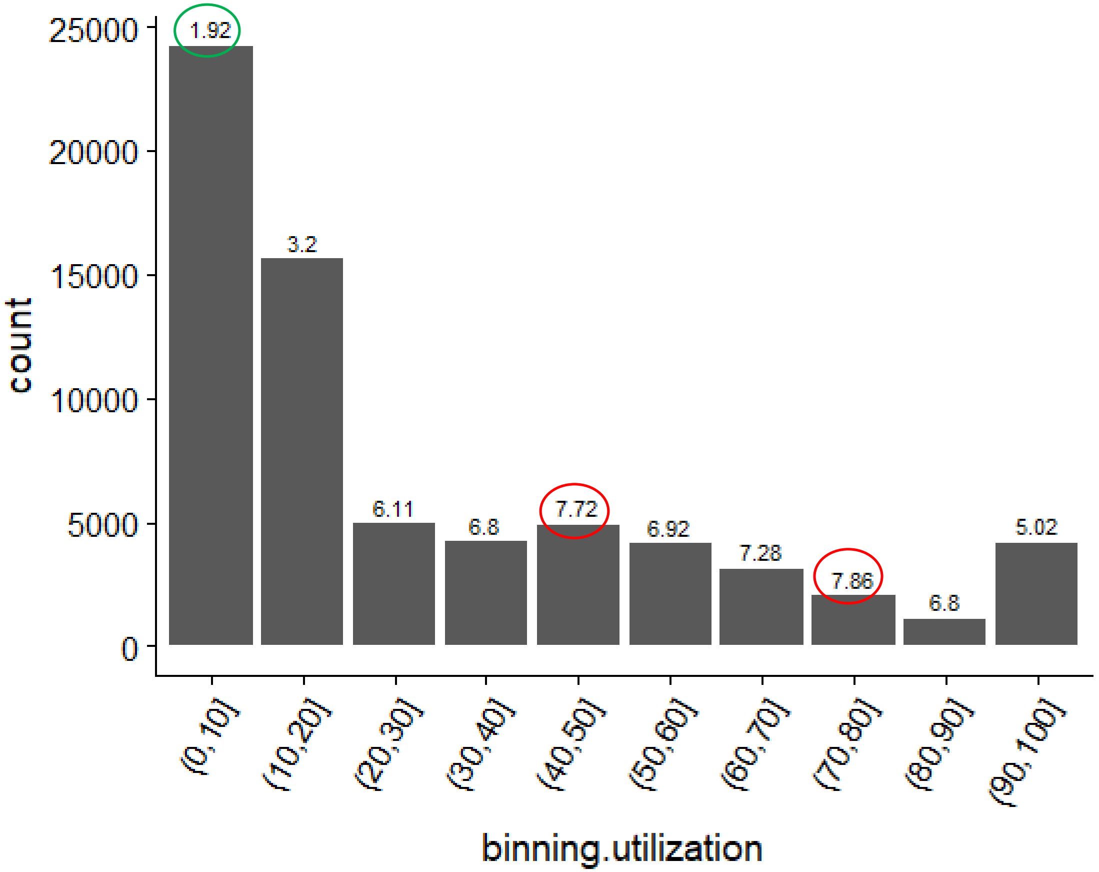
Months in current residence:
Highest default rate : 125-145 months (8.5714%) and 25-45 months (5.856%)
Lowest default rate: 5-25 months (3.75%)



Months in current company:
Highest default rate : 0-20 months range (4.92%)
Lowest default rate : 40-60 months range (3.472%)

Outstanding Balance and Utilization



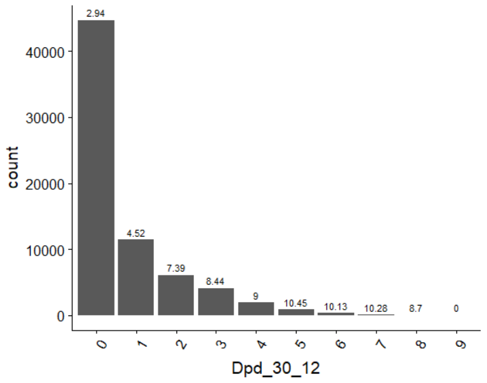


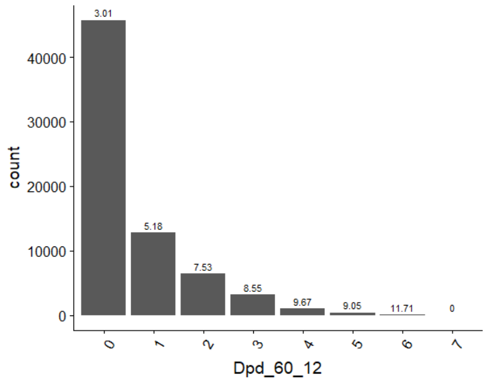
Outstanding_balance has a significant impact on default rate
High default rates : 50-55L (16.6%) ,45-50L (6.2%)
Low default rates : 0-2L (1.856%) ,25L-30L (1.583%)

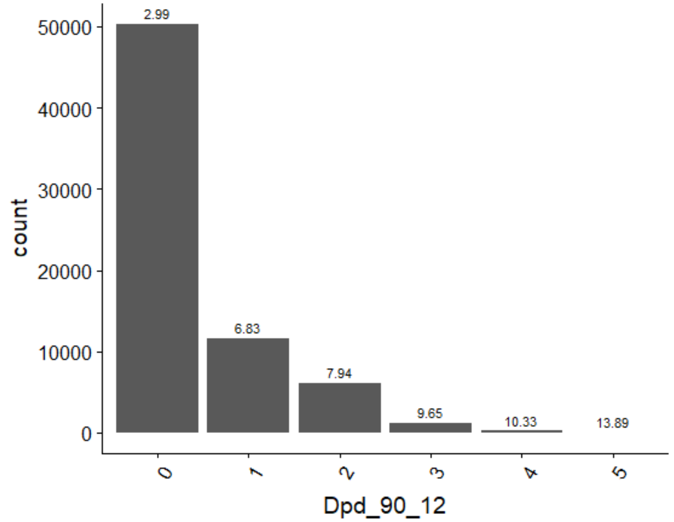
Utilization is also a significant factor
Highest default rates : 40-50% utilization (7.72%),70-80% utilization (7.86%)
Lowest default rates: 0-10% utilization(1.92%)

90,60 and 30 DPD in last 12 months

- **90 DPD in 12 months is a strong indicator of default as suggested in problem statement**
- **The lowest default rates are at 0 DPD and increases as number of DPD increases**
- **Trend is same in 30 DPD and 60 DPD**





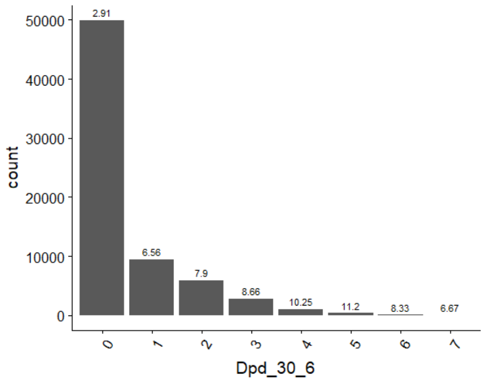


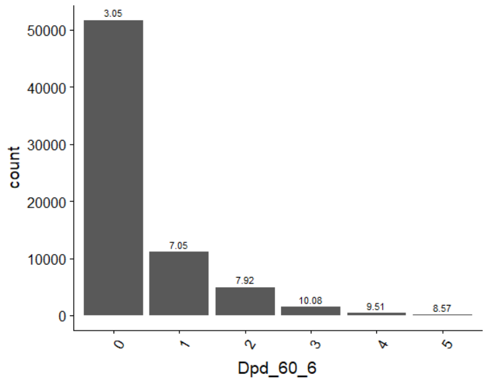
90,60 and 30 DPD in last 6 months

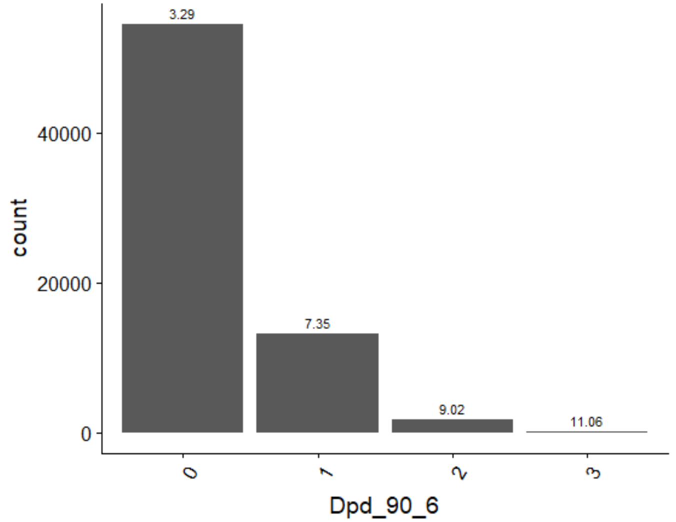
90 DPD in last six months is a strong indicator of default

30 and 60 DPD follow same trend as 90 DPD and can be said to be early indicators

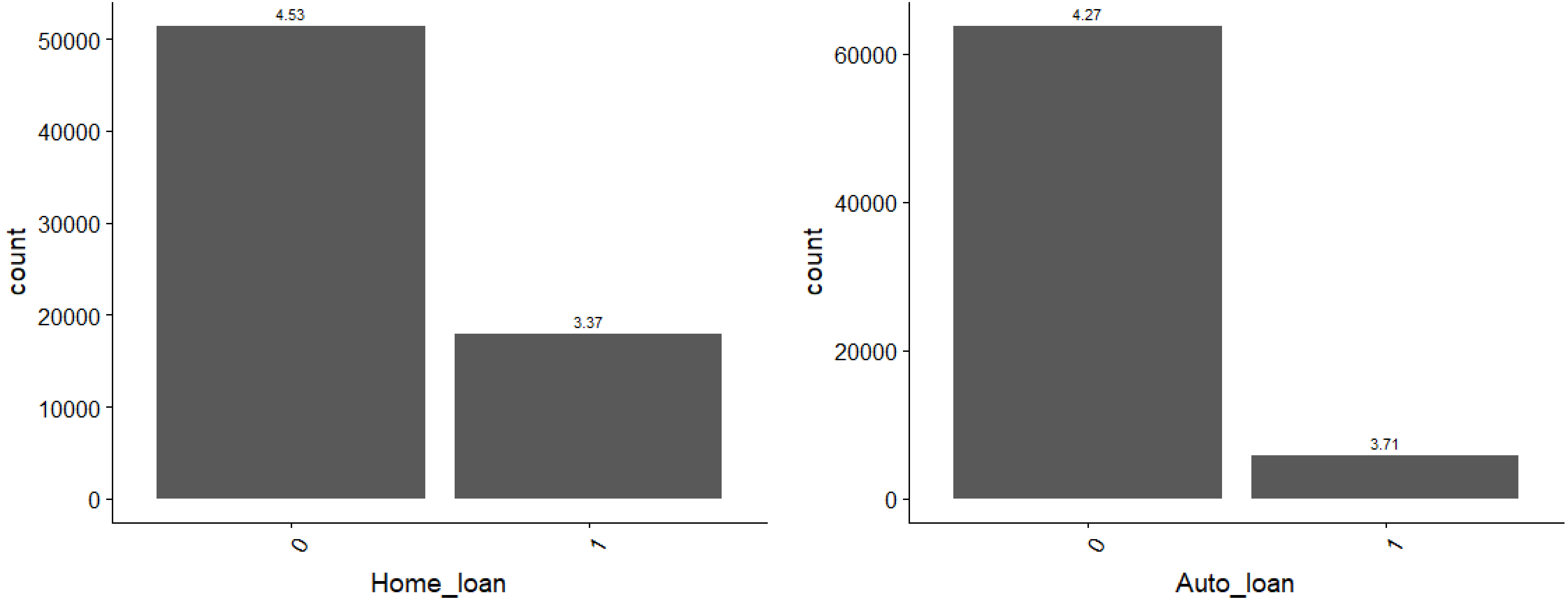
DPDs in last 6 months is an early indicator of DPDs in the last 12 months





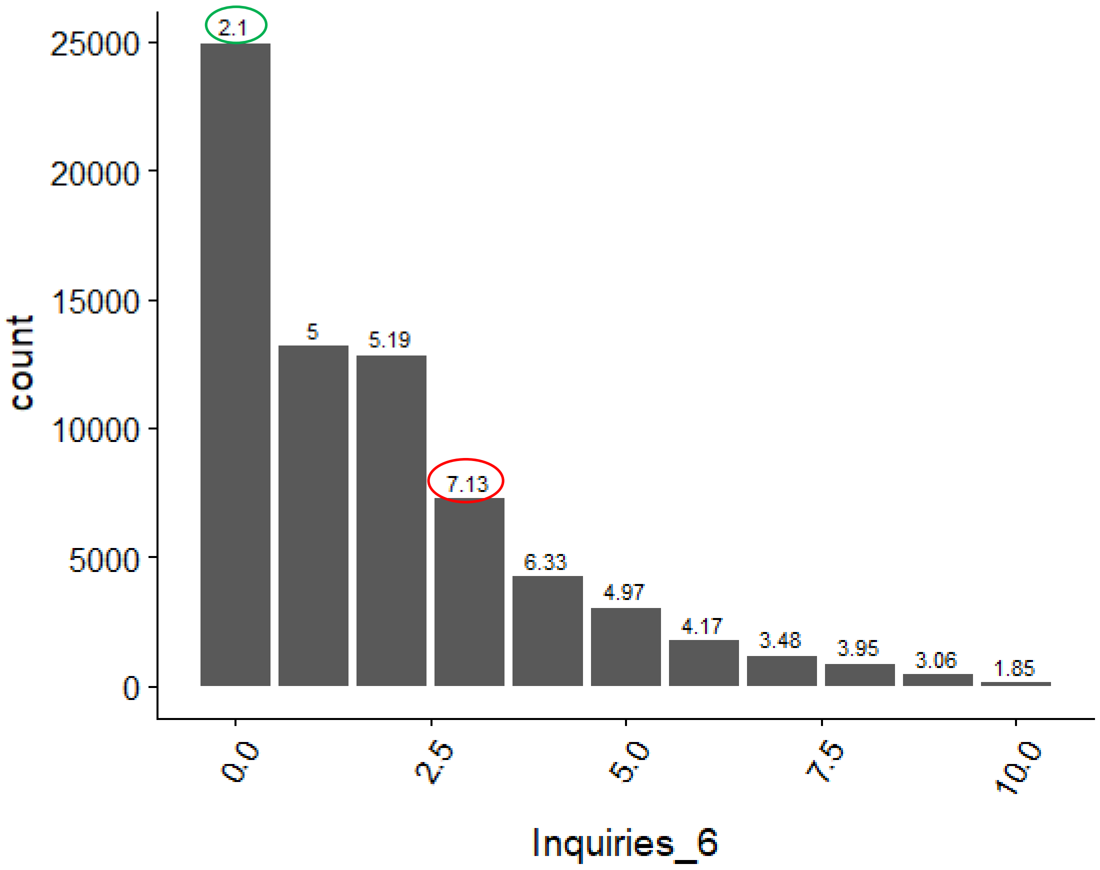


Presence of Home and Auto Loan

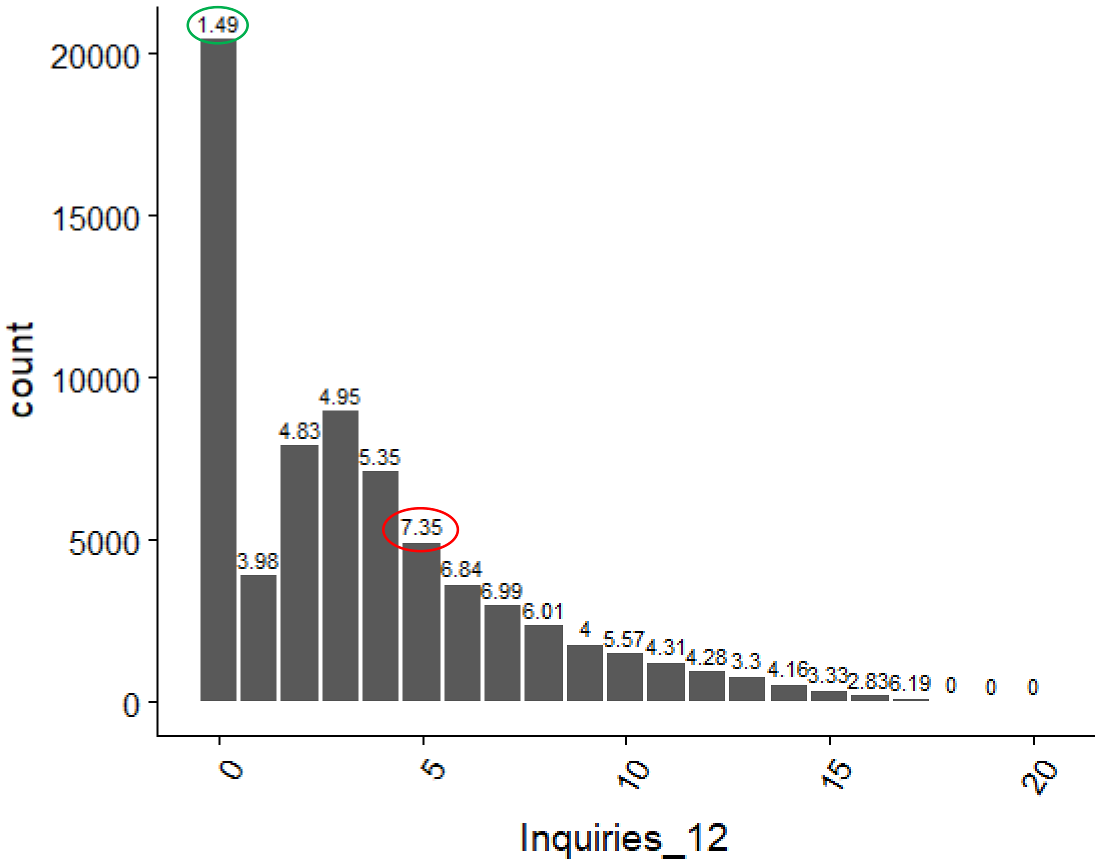


The non loan takers have higher default rate. Same trend seen with home loan and auto loan takers.
Loan takers are less likely to default

Number of Inquiries for loan in 12 and 6 months

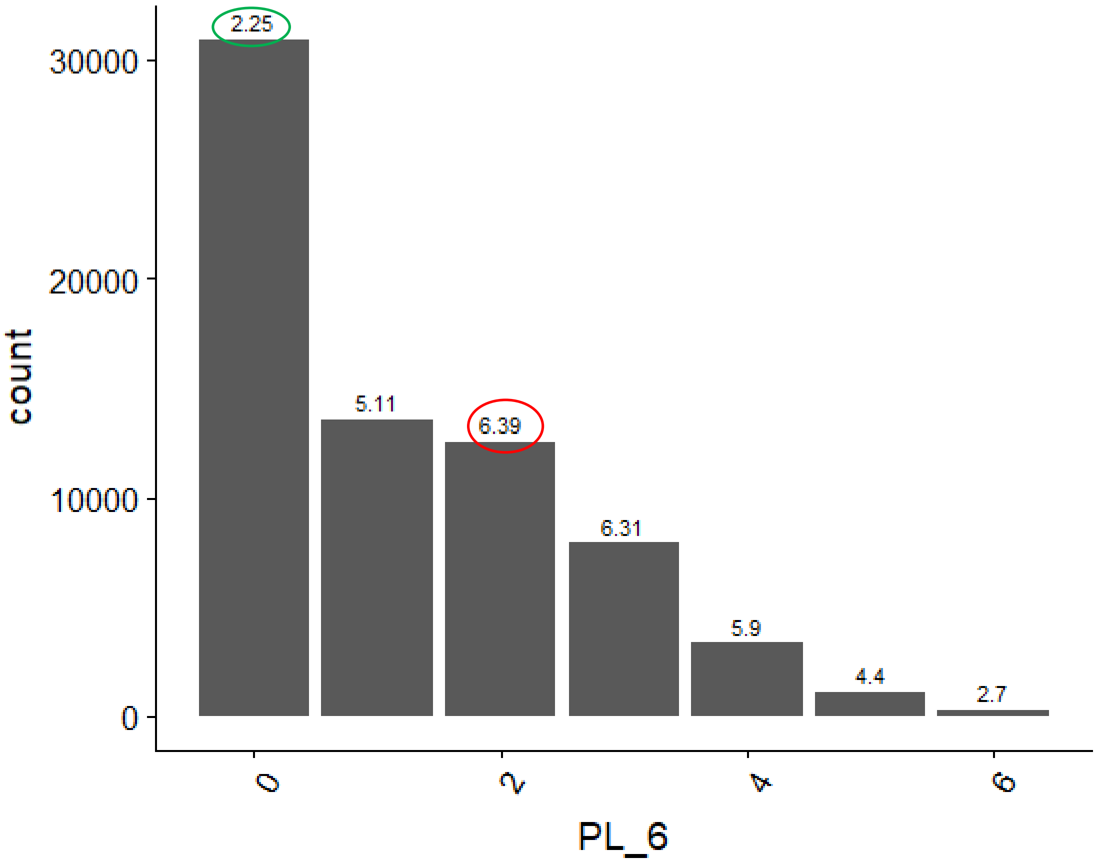


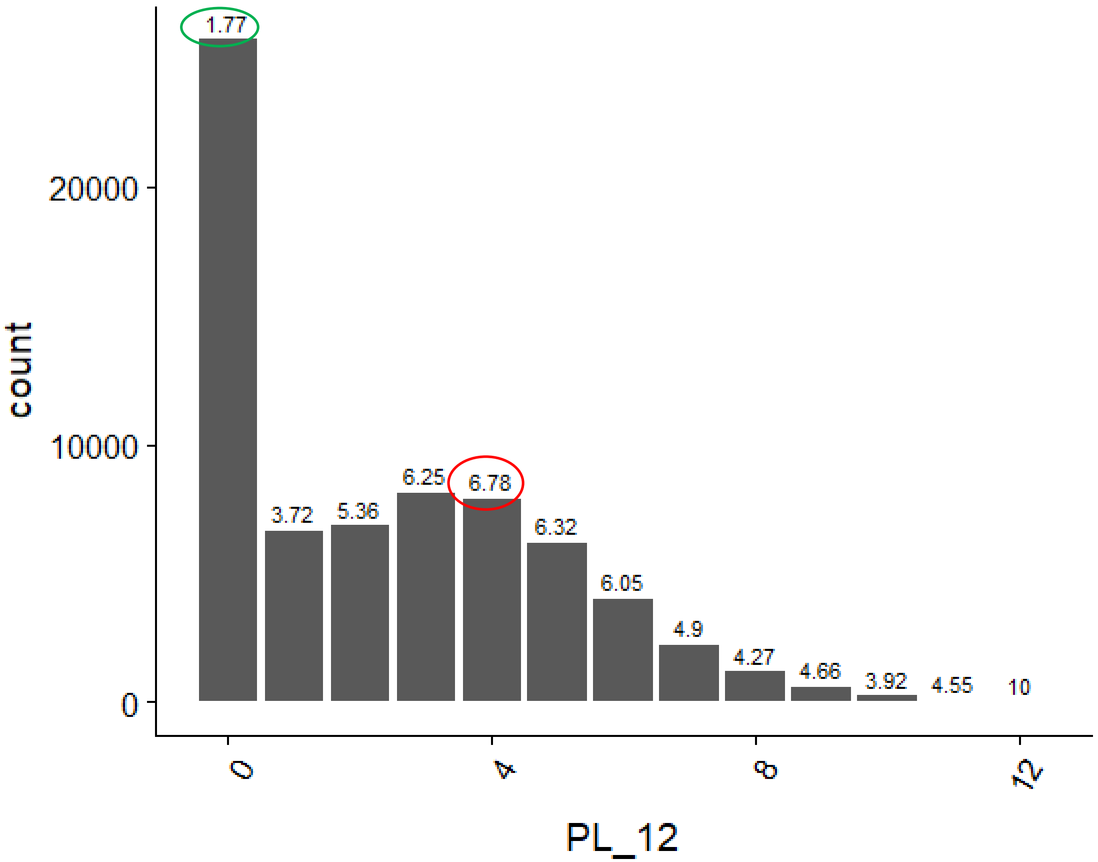
Default rate of those with 0 inquiries is the lowest with 2.1% while with 4 inquiries highest at 7.13%; **This trend continues in the 12 months data**



Default rate of those with 0 inquiries is the lowest with 1.48% while with 5 inquiries highest at 7.35%

Total PL trades opened in 12 and 6 months



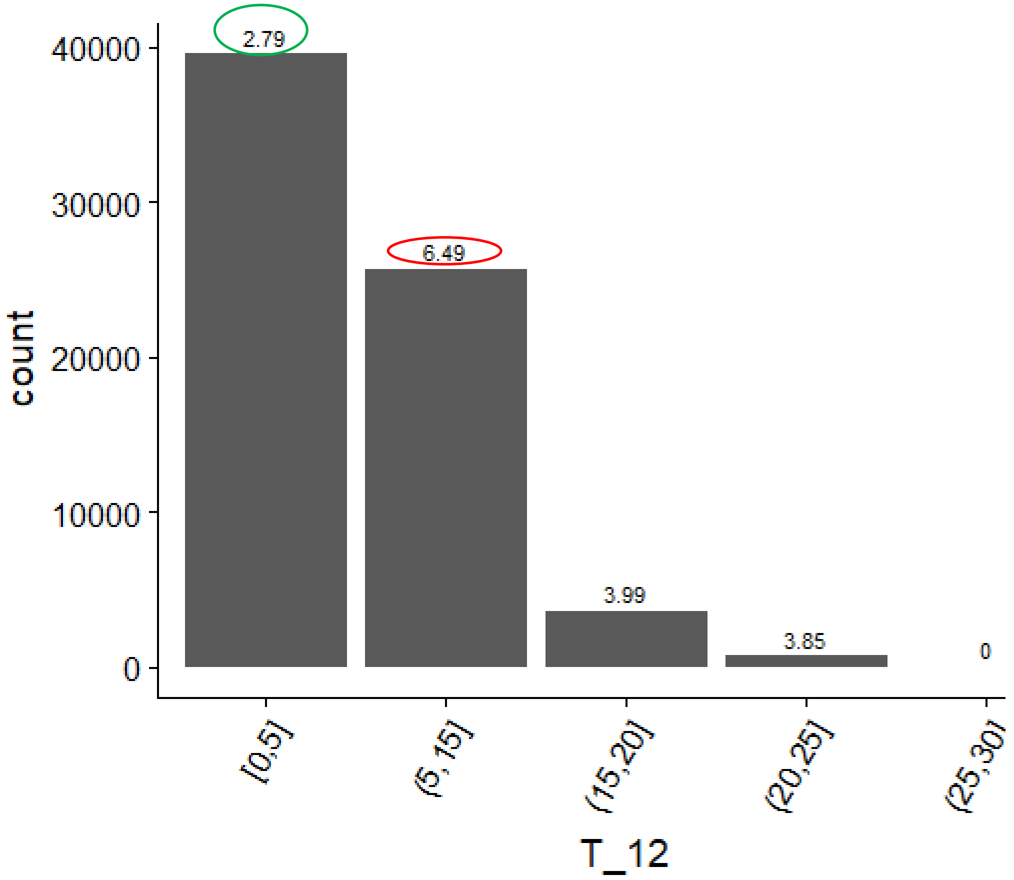


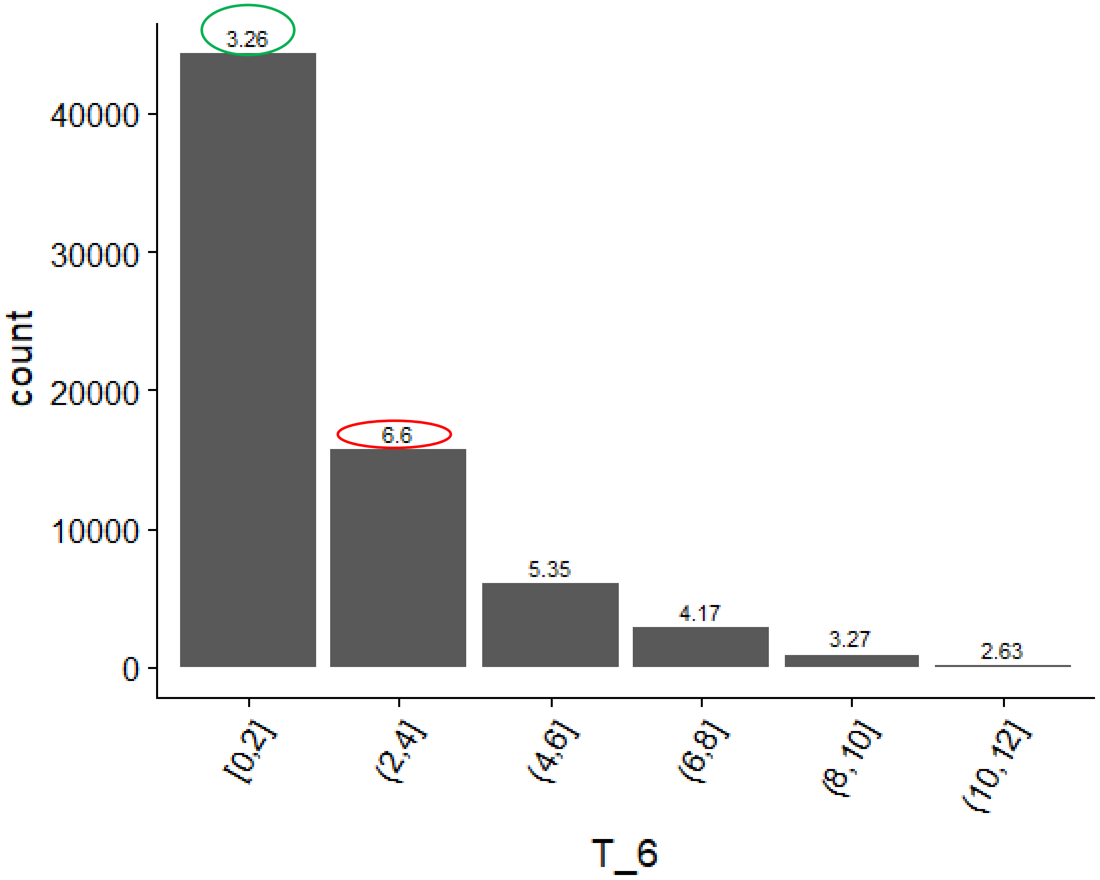
Default rate of those with 0 PL trades is lowest at 2.25% and those with 2 is highest at 6.39%. The same trend continues in the 12 months data

Default rate of those with 0 PL trades is lowest at 1.76% and those with 5 is highest at 6.77%

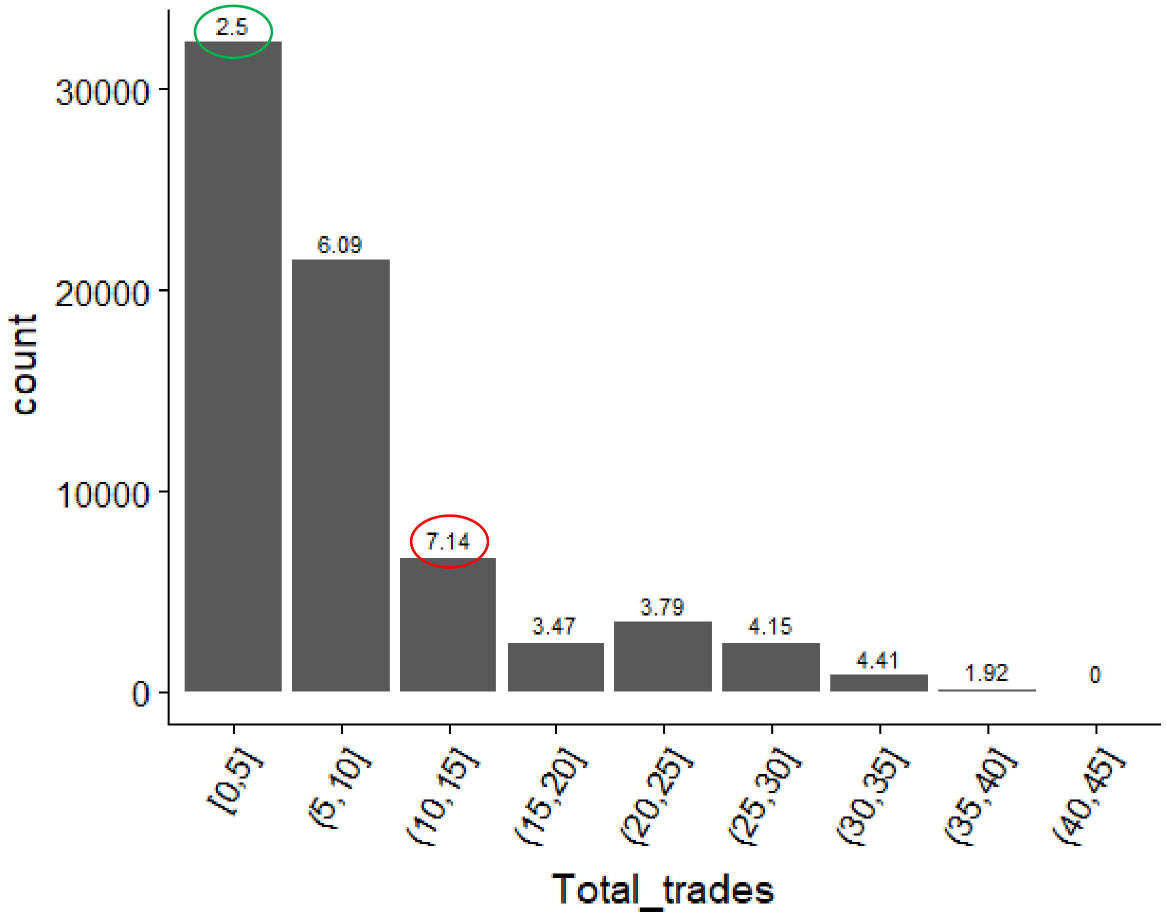
Those with 0 PL trades have very low default rate

Total Trades opened in 12 and 6 months





Total trades



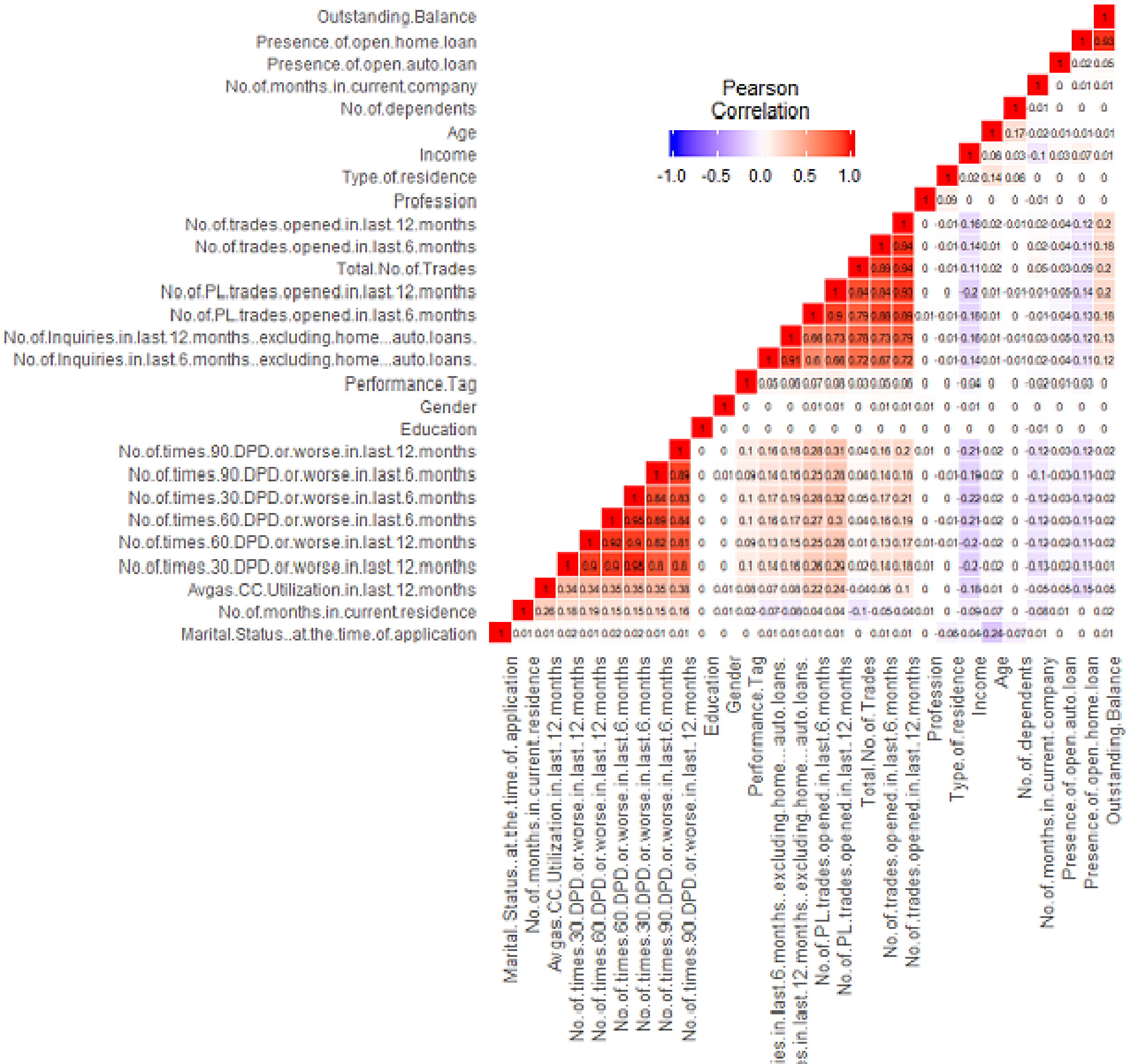
Similar trend in case of all trades variables; Lowest default rate in the lower number of trades

Trades in 12 months: Default rate of those with 0-5 trades is lowest at 2.863% and those with 5-15 is highest at 6.39%

Trades in 6 months :Default rate of those with 0-2 buckets trades is lowest at 3.26% and those with 2-4 is highest at 6.59%

Total Trades : Default rates lowest in the bucket 0-5 trades 2.442 and are highest in the 10-15 trades bucket 7.14%

Multivariate analysis: Correlation Map



Highly corelated variables are

No.of.times.30.DPD.or.worse.in.last.12.months
No.of.times.60.DPD.or.worse.in.last.12.months
No.of.times.90.DPD.or.worse.in.last.12.months
No.of.times.30.DPD.or.worse.in.last.6.months
No.of.times.60.DPD.or.worse.in.last.6.months
No.of.times.90.DPD.or.worse.in.last.6.months

No.of.trades.opened.in.last.12.months
No.of.trades.opened.in.last.6.months
Total.No.of.Trades
No.of.PL.trades.opened.in.last.12.months
No.of.PL.trades.opened.in.last.6.months
No.of.Inquiries.in.last.12.months..excluding.home...auto.loans.
No.of.Inquiries.in.last.12.months..excluding.home...auto.loans.

Presence.of.open.home.loan
Outstanding Balance

These variables that are highly correlated are checked for multicollinearity during regression

Information Value and WOE

- “create_infotables” command from the “Information” package used to generate information value table. The variables are classified as

Information value table

	Variable	IV
33	binning.utilization	0.31294145566
17	Avgas.CC.Utilization.in.last.12.months	0.31033940329
19	No.of.trades.opened.in.last.12.months	0.29840457459
21	No.of.PL.trades.opened.in.last.12.months	0.29600248642
23	No.of.Inquiries.in.last.12.months..excluding.home...auto.loans.	0.29549940031
32	binning.outstanding.balance	0.26641539692
25	Outstanding.Balance	0.24574159704
13	No.of.times.30.DPD.or.worse.in.last.6.months	0.24442842211
26	Total.No.of.Trades	0.23724711796
20	No.of.PL.trades.opened.in.last.6.months	0.21948965290
16	No.of.times.30.DPD.or.worse.in.last.12.months	0.21850426189
14	No.of.times.90.DPD.or.worse.in.last.12.months	0.21588238599

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

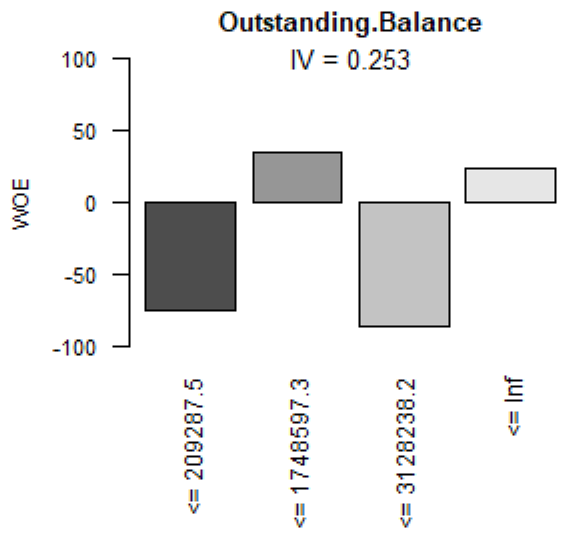
Information Value and WOE

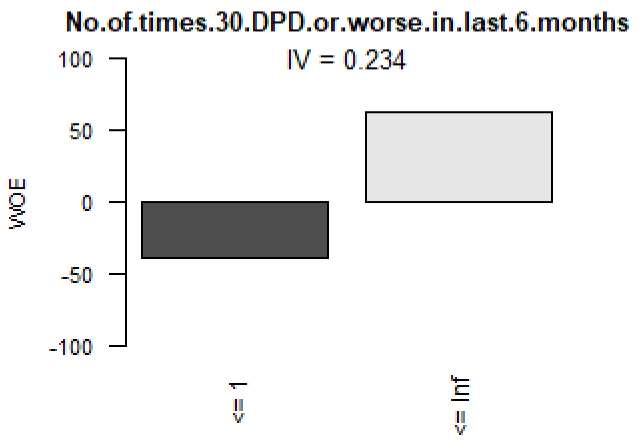
12	No.of.times.60.DPD.or.worse.in.last.6.months	0.21144795950
22	No.of.Inquiries.in.last.6.months..excluding.home...auto.loans.	0.20494812821
36	binning.total.trades	0.20348225588
15	No.of.times.60.DPD.or.worse.in.last.12.months	0.18846092230
18	No.of.trades.opened.in.last.6.months	0.18592399079
34	binning.trades12	0.18062311844
11	No.of.times.90.DPD.or.worse.in.last.6.months	0.16287100438
35	binning.trades6	0.11003299761
9	No.of.months.in.current.residence	0.07913479156
5	Income	0.04311878120
29	binning.income	0.04140231290
30	binning.residence_months	0.03025928604
10	No.of.months.in.current.company	0.02142614451
31	binning.company_months	0.01956801054
24	Presence.of.open.home.loan	0.01742740811
1	Age	0.00343990656
4	No.of.dependents	0.00261236946
7	Profession	0.00223993052
27	Presence.of.open.auto.loan	0.00160852785
28	binning.age	0.00138312210
8	Type.of.residence	0.00094846678
6	Education	0.00075247413
2	Gender	0.00033903584
3	Marital.Status..at.the.time.of.application.	0.00009124951

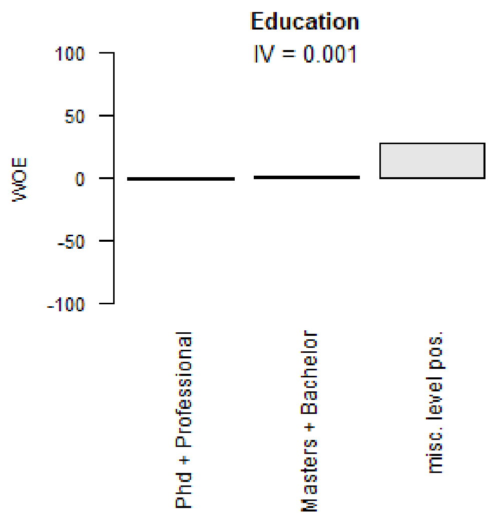
- **The credit bureau data variables have higher IV than demographic data**

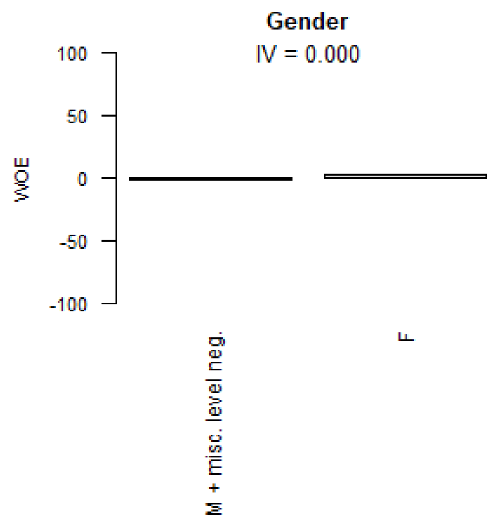
Coarse binning and WOE replacement for modelling

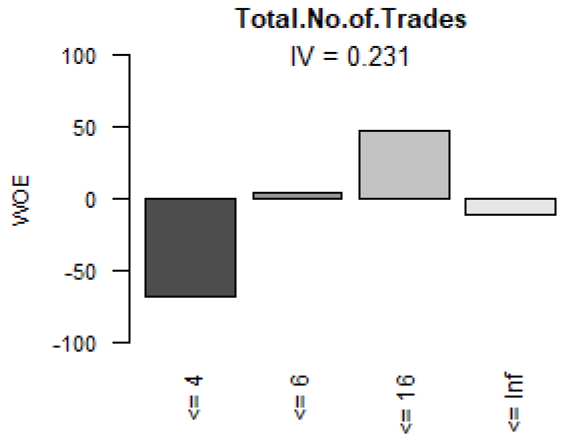
“woeBinning” package used for automatic binning of numeric variables and factors with respect to Performance.Tag

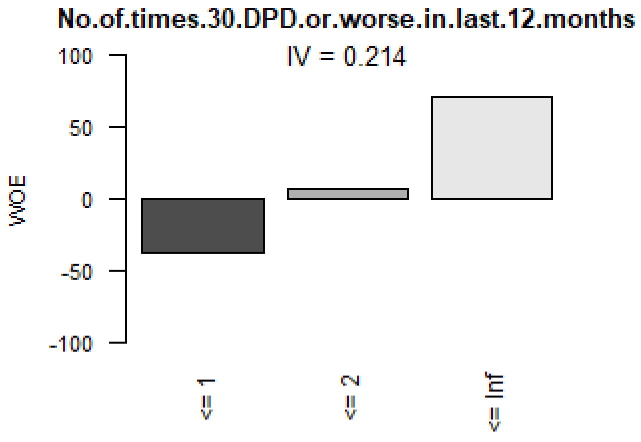




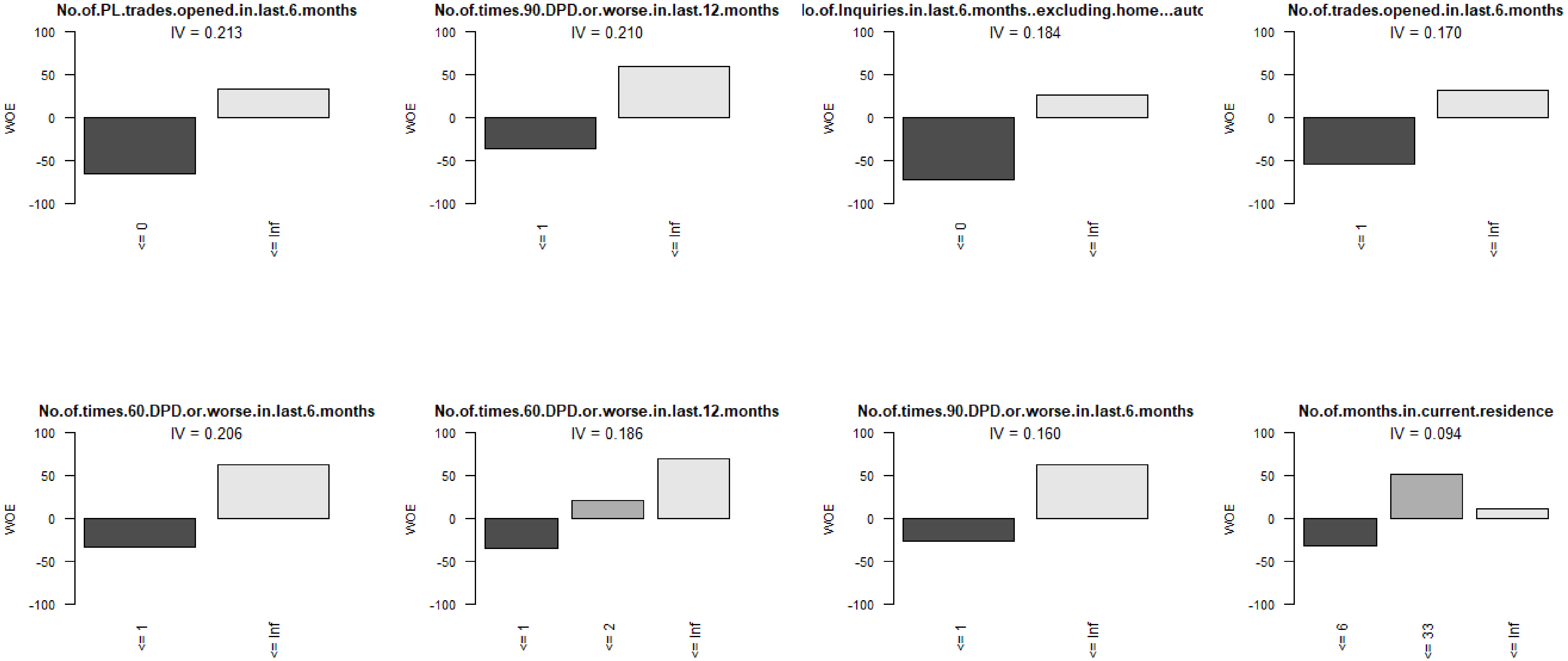




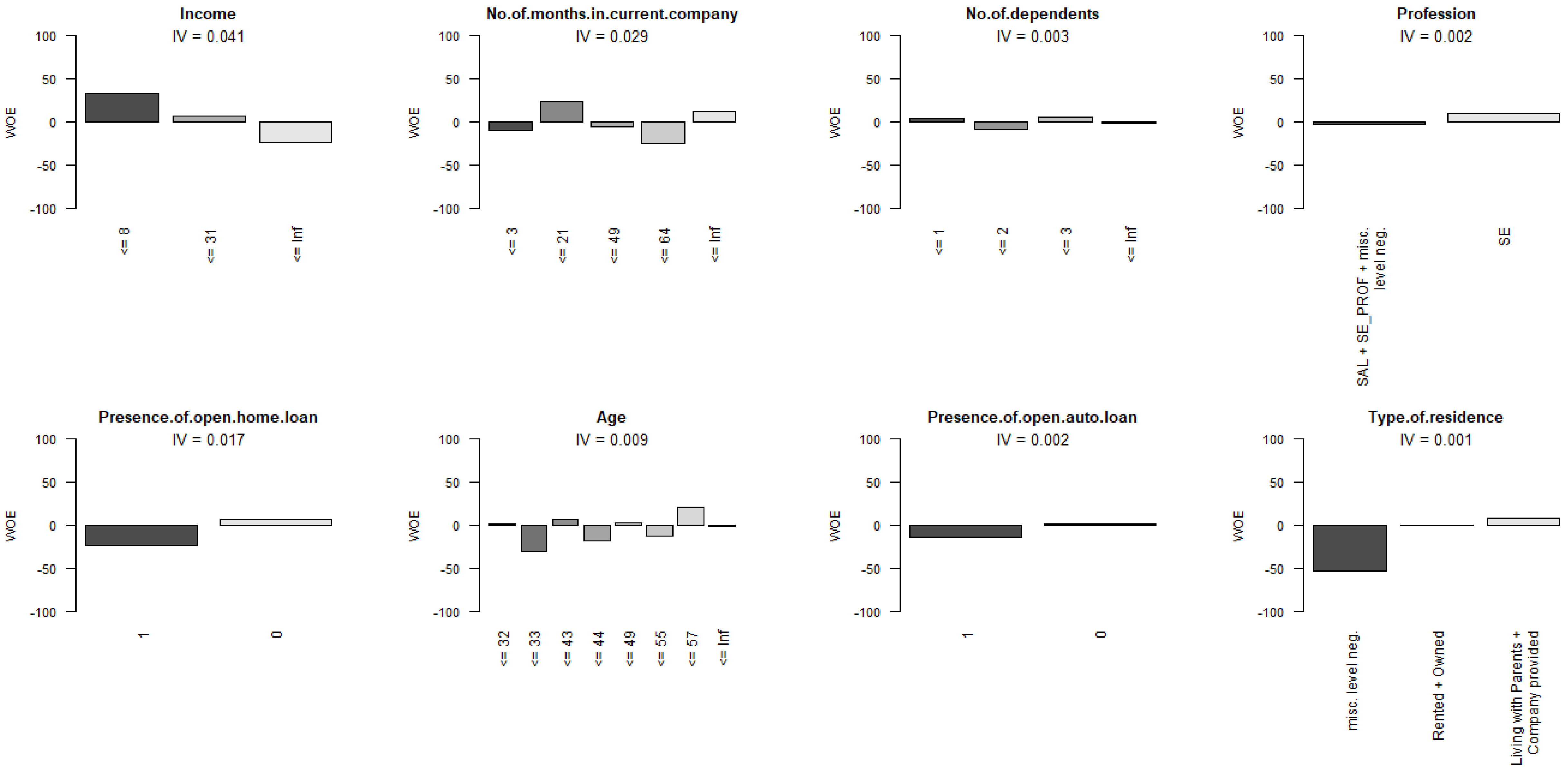




Coarse binning and WOE replacement for modelling



Coarse binning and WOE replacement for modelling



Data Balancing

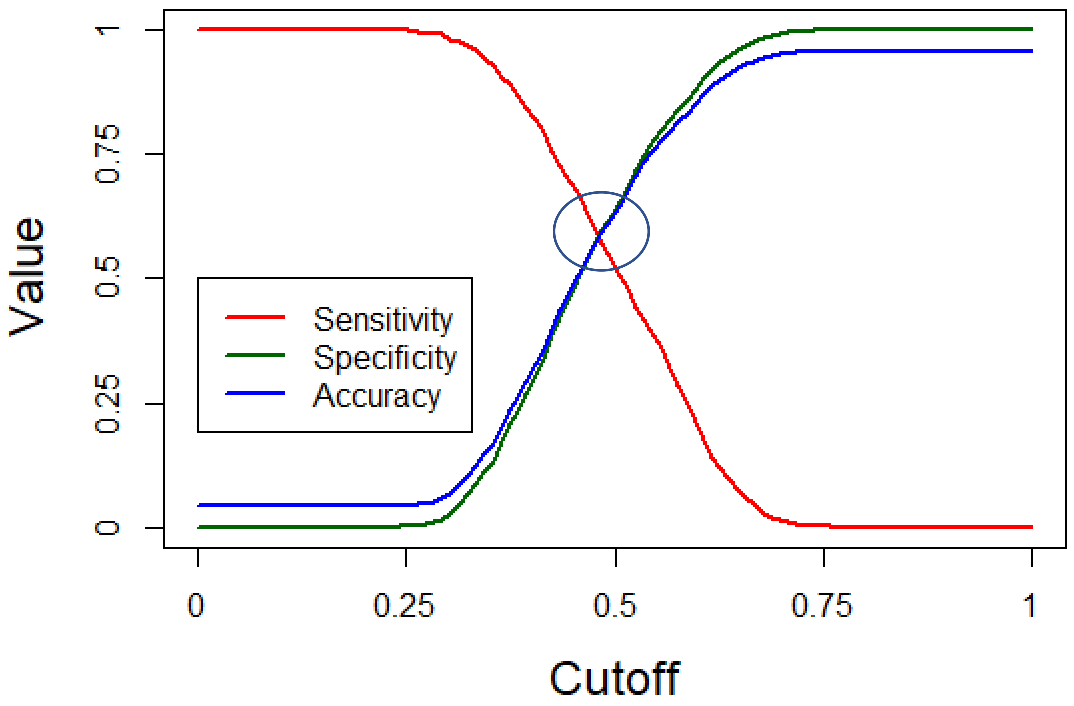
- The Data is split into train and test data in the 70:30 ratio
- The number of 1s in the Dependent variable, Performance.Tag is 2062 while 0s are 46728. We see 4.41% 1s and 95.58% 0s.
- The training data needs to be balanced to get a better model
- SMOTE algorithm is used to oversample the 1s and undersample 0s
  - train_master_smote <- SMOTE(Performance.Tag ~ .,train_master_file, perc.over = 800,perc.under=120)
- There are 51.61% 0s and 48.59% 1s in the train data after applying SMOTE

Modelling

- Modelling Default using Demographic data only
  - All the independent variables are replaced by WOE
  - Logistic regression used to model default : Final model has 8 variables of significance

```
Coefficients:
                                             Estimate Std. Error z value      Pr(>|z|)    
(Intercept)                                -0.0646979  0.0104421  -6.196 0.00000000058 ***
woe.No.of.months.in.current.residence.binned  0.0092097  0.0003458  26.633       < 2e-16 ***
woe.Income.binned                           0.0087403  0.0005230  16.713       < 2e-16 ***
woe.No.of.months.in.current.company.binned  0.0097238  0.0006190  15.710       < 2e-16 ***
woe.Age.binned                              0.0099905  0.0011523   8.670       < 2e-16 ***
woe.No.of.dependents.binned                 0.0101902  0.0020853   4.887 0.00000102538 ***
woe.Profession.binned                       0.0053905  0.0022344   2.412       0.01584 *  
woe.Type.of.residence.binned                0.0106271  0.0036122   2.942       0.00326 ** 
woe.Gender.binned                           0.0095352  0.0056918   1.675       0.09388 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model Evaluation



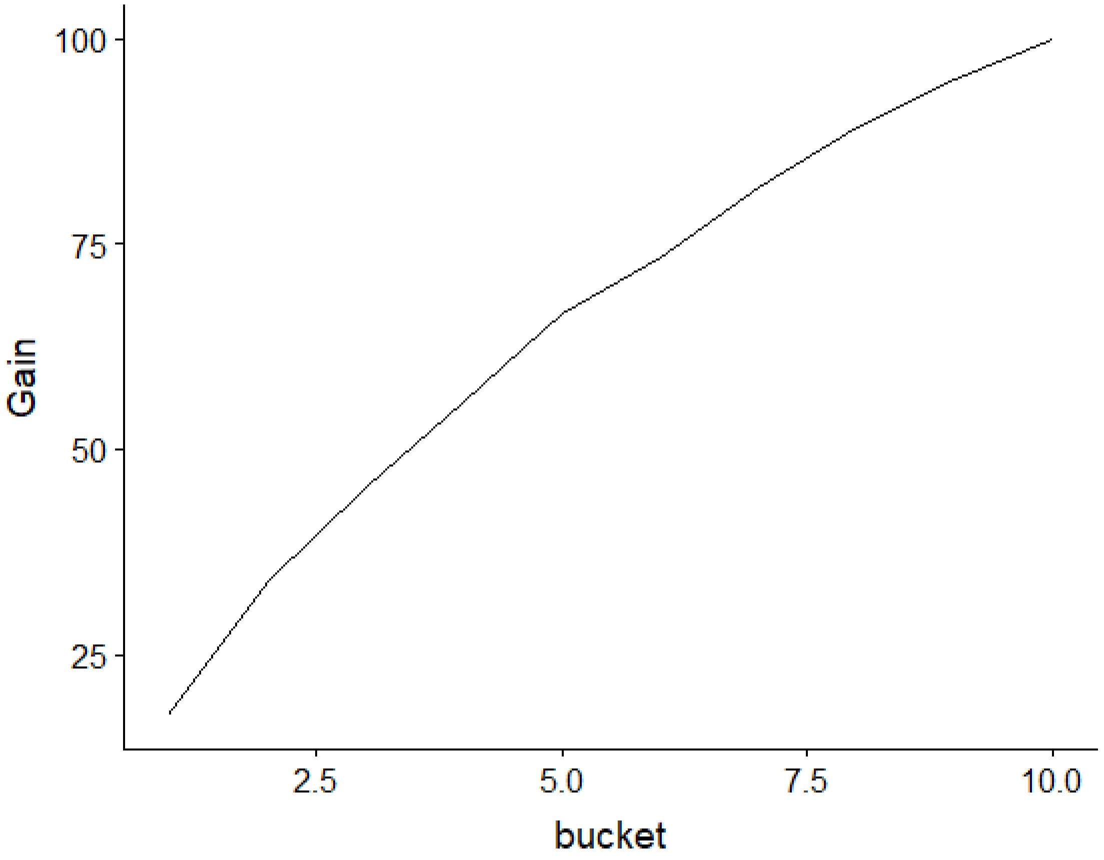
Choosing the cutoff probability as the intersection of the Sensitivity, Specificity and Accuracy: 48.48% , confusion matrix gives

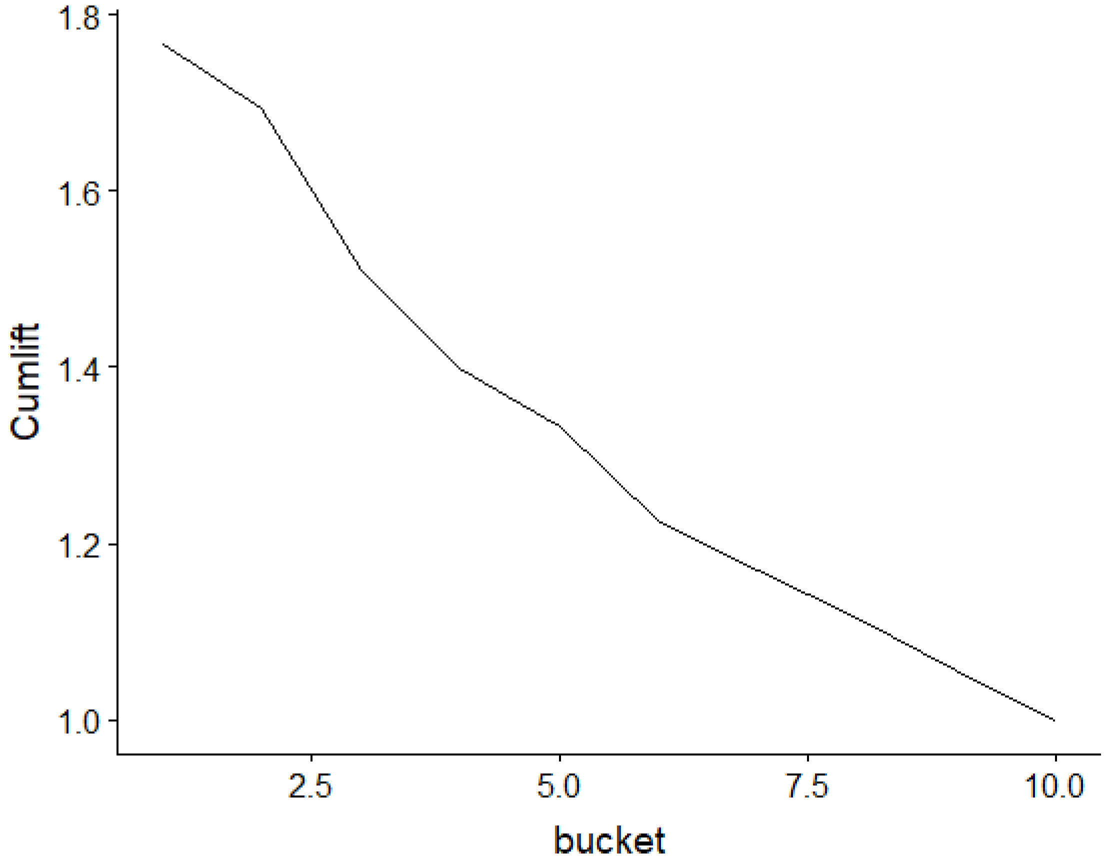
Accuracy	57.92%
Sensitivity	58.89%
Specificity	57.87%
KS Statistic	13.84%

Gain and lift table

```
# A tibble: 10 x 6
  bucket total totalresp Cumresp  Gain Cumlift
   <int> <int>     <dbl>   <dbl> <dbl>   <dbl>
1      1  2091       156     156  17.7    1.77
2      2  2091       143     299  33.9    1.69
3      3  2091       101     400  45.3    1.51
4      4  2091        94     494  55.9    1.40
5      5  2091        94     588  66.6    1.33
6      6  2091        60     648  73.4    1.22
7      7  2091        75     723  81.9    1.17
8      8  2091        65     788  89.2    1.12
9      9  2091        51     839  95.0    1.06
10    10  2090        44     883 100      1
```

Gain and Lift charts



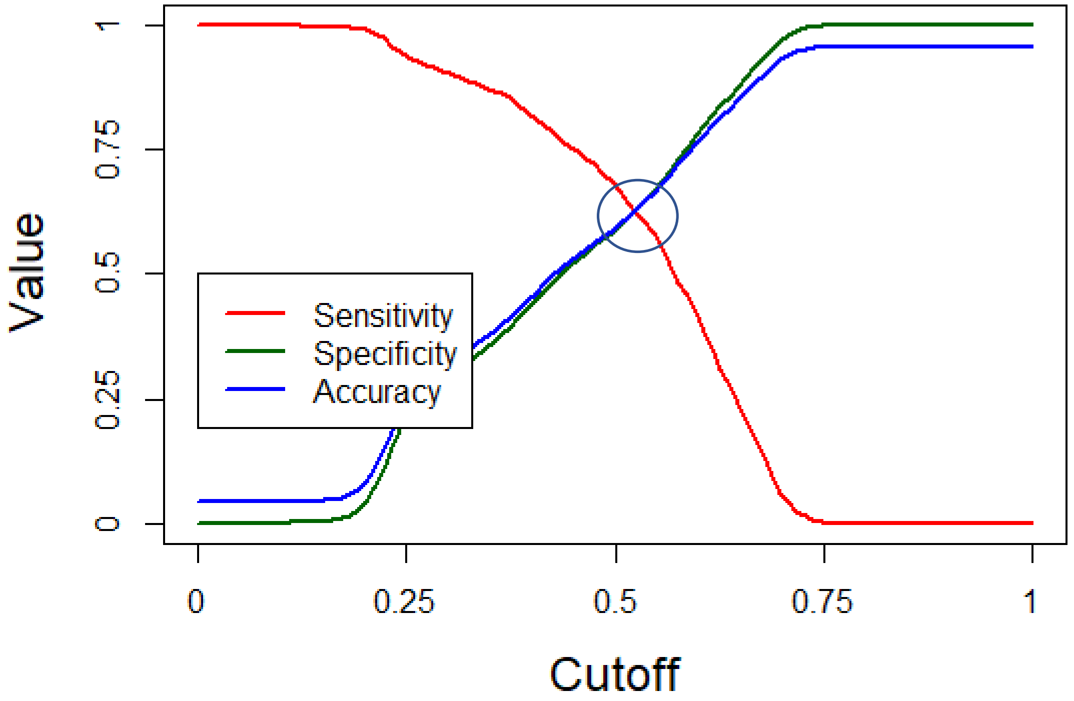


Modelling Default using complete data Logistic Regression

- All the independent variables are replaced by WOE
- Final model has 16 variables of 3* significance

```
Coefficients:
                                                                      Estimate Std. Error z value Pr(>|z|)    
(Intercept)                                                         -0.0756947  0.0108675  -6.965 3.28e-12 ***
woe.Avgas.CC.Utilization.in.last.12.months.binned                    0.0047096  0.0002738  17.200  < 2e-16 ***
woe.No.of.PL.trades.opened.in.last.12.months.binned                 -0.0026264  0.0005390  -4.873 1.10e-06 ***
woe.No.of.Inquiries.in.last.12.months..excluding.home...auto.loans..binned  0.0038411  0.0004472   8.589  < 2e-16 ***
woe.Outstanding.Balance.binned                                       0.0031578  0.0004410   7.161 8.02e-13 ***
woe.No.of.times.30.DPD.or.worse.in.last.12.months.binned             0.0041568  0.0005256   7.909 2.59e-15 ***
woe.No.of.PL.trades.opened.in.last.6.months.binned                   0.0016959  0.0004487   3.779 0.000157 ***
woe.No.of.times.60.DPD.or.worse.in.last.12.months.binned            -0.0026037  0.0006140  -4.240 2.23e-05 ***
woe.No.of.Inquiries.in.last.6.months..excluding.home...auto.loans..binned  0.0015693  0.0004469   3.512 0.000445 ***
woe.No.of.times.90.DPD.or.worse.in.last.6.months.binned              0.0029617  0.0004810   6.157 7.42e-10 ***
woe.Income.binned                                                    0.0028457  0.0005765   4.936 7.97e-07 ***
woe.No.of.months.in.current.company.binned                           0.0026176  0.0006734   3.887 0.000101 ***
woe.Age.binned                                                       0.0082286  0.0012072   6.816 9.33e-12 ***
woe.No.of.dependents.binned                                          0.0167523  0.0022443   7.464 8.38e-14 ***
woe.Presence.of.open.auto.loan.binned                                0.0272092  0.0030363   8.961  < 2e-16 ***
woe.Type.of.residence.binned                                         0.0156212  0.0037624   4.152 3.30e-05 ***
woe.Marital.Status..at.the.time.of.application.binned               -0.0474730  0.0116951  -4.059 4.92e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Model Evaluation



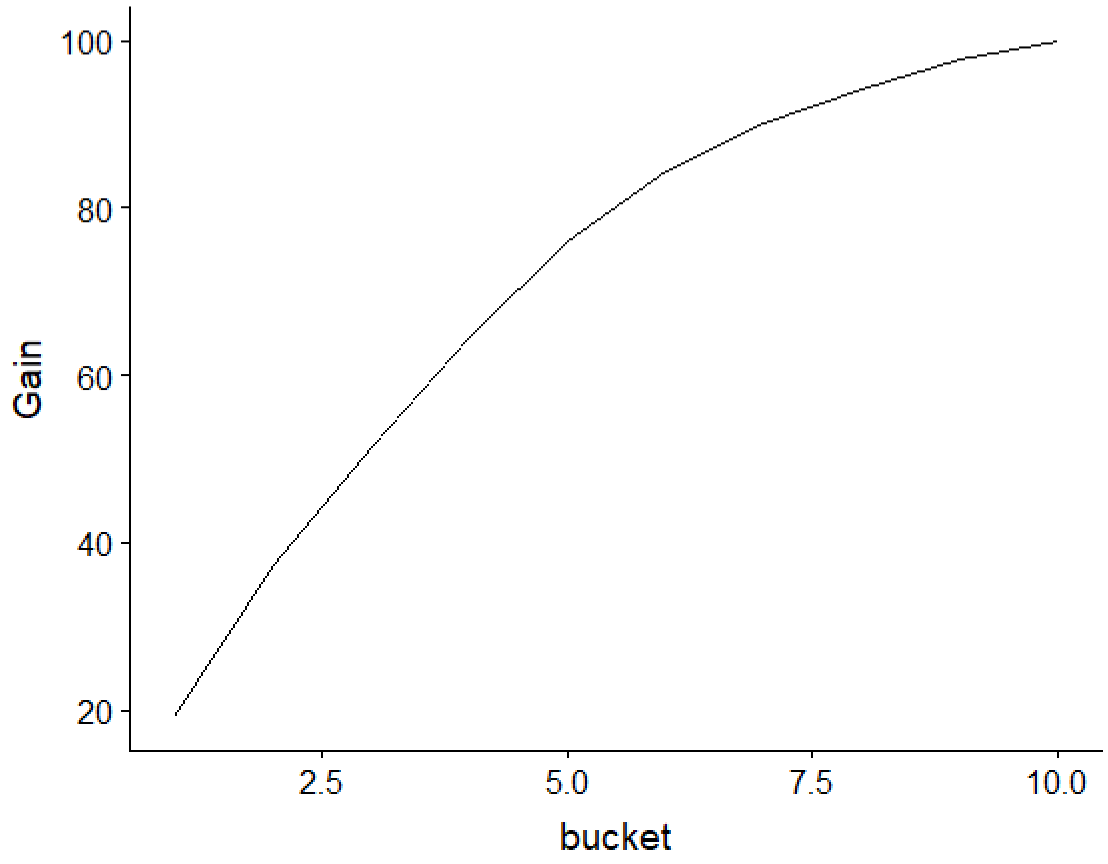
Choosing the cutoff probability as the intersection of the Sensitivity, Specificity and Accuracy: 52.02% , confusion matrix gives

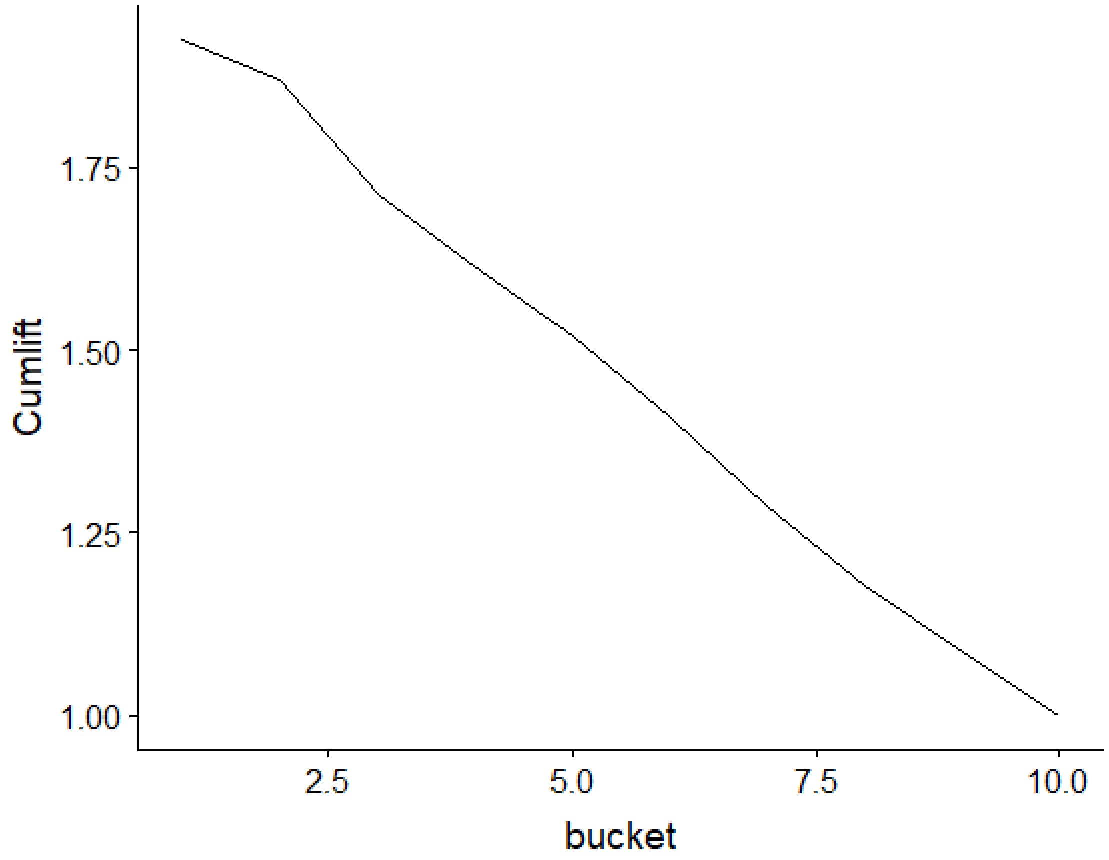
Accuracy	61.93%
Sensitivity	63.31%
Specificity	61.86%
KS Statistic	25.17%

Gain and Lift Table

```
# A tibble: 10 x 6
   bucket total totalresp Cumresp  Gain Cumlift
    <int> <int>     <dbl>   <dbl> <dbl>   <dbl>
 1      1  2091       170     170  19.3    1.93
 2      2  2091       160     330  37.4    1.87
 3      3  2091       124     454  51.4    1.71
 4      4  2091       116     570  64.6    1.61
 5      5  2091       101     671  76.0    1.52
 6      6  2091        74     745  84.4    1.41
 7      7  2091        50     795  90.0    1.29
 8      8  2091        37     832  94.2    1.18
 9      9  2091        32     864  97.8    1.09
10     10  2090        19     883 100      1
```

Gain and lift charts

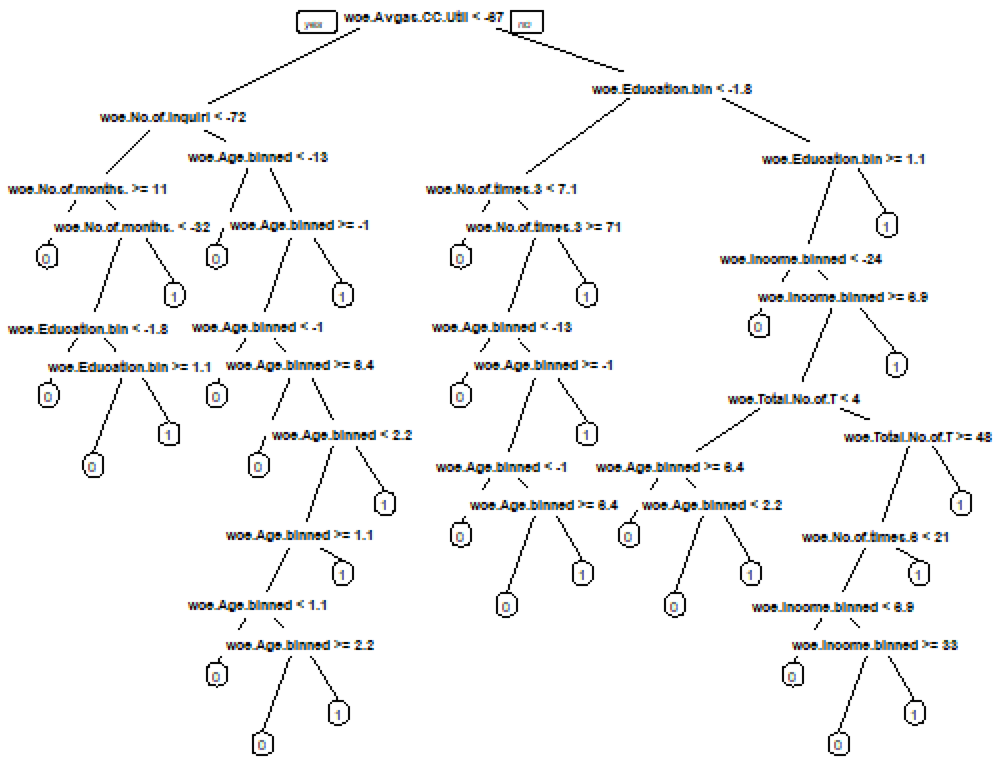




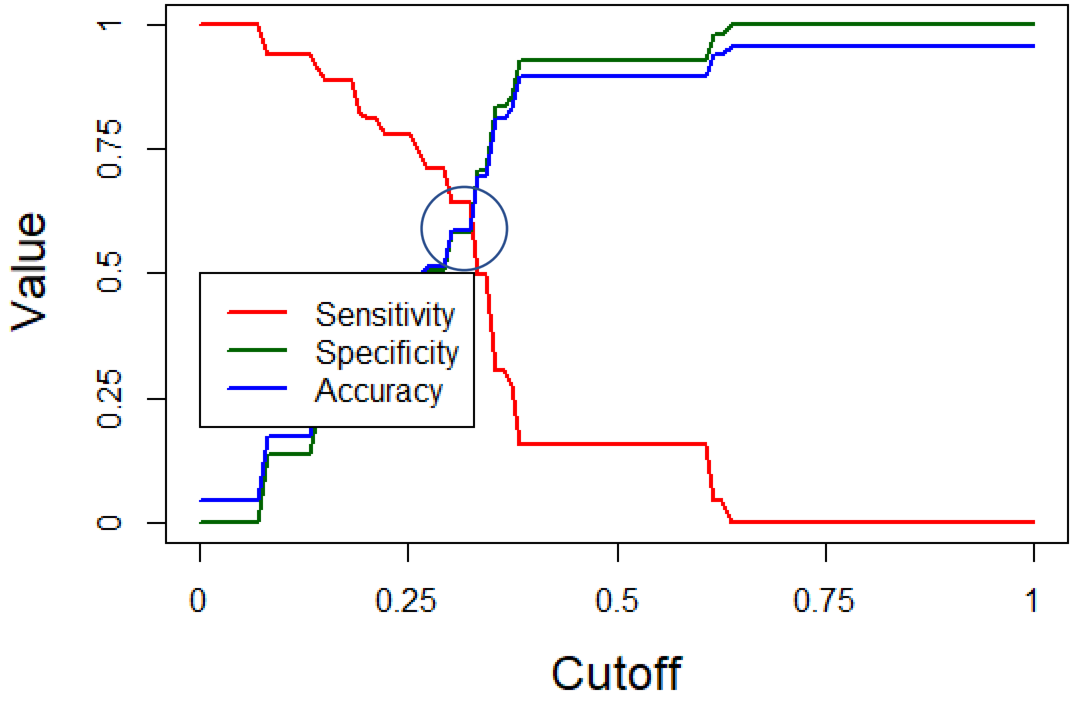
Modelling Default using complete data
Binary Trees

- Tried various trees by tuning hyperparameters minsplit and complexity parameter (cp)
  - First tree was with minsplit=20,cp = 0.0001. Highly Overfitted tree was generated; height = 77 ; size = 515
  - Second tree was with minsplit=50,cp=0.001. Tree was smaller , but still big; height = 33 ; size = 89
  - Third tree was with minsplit=70,cp=0.005 . Tree was average sized; height = 12 ; size= 31

Tree 3



Model Evaluation for Binary tree 3



Choosing the cutoff probability as the intersection of the Sensitivity, Specificity and Accuracy: 31.31%, confusion matrix gives

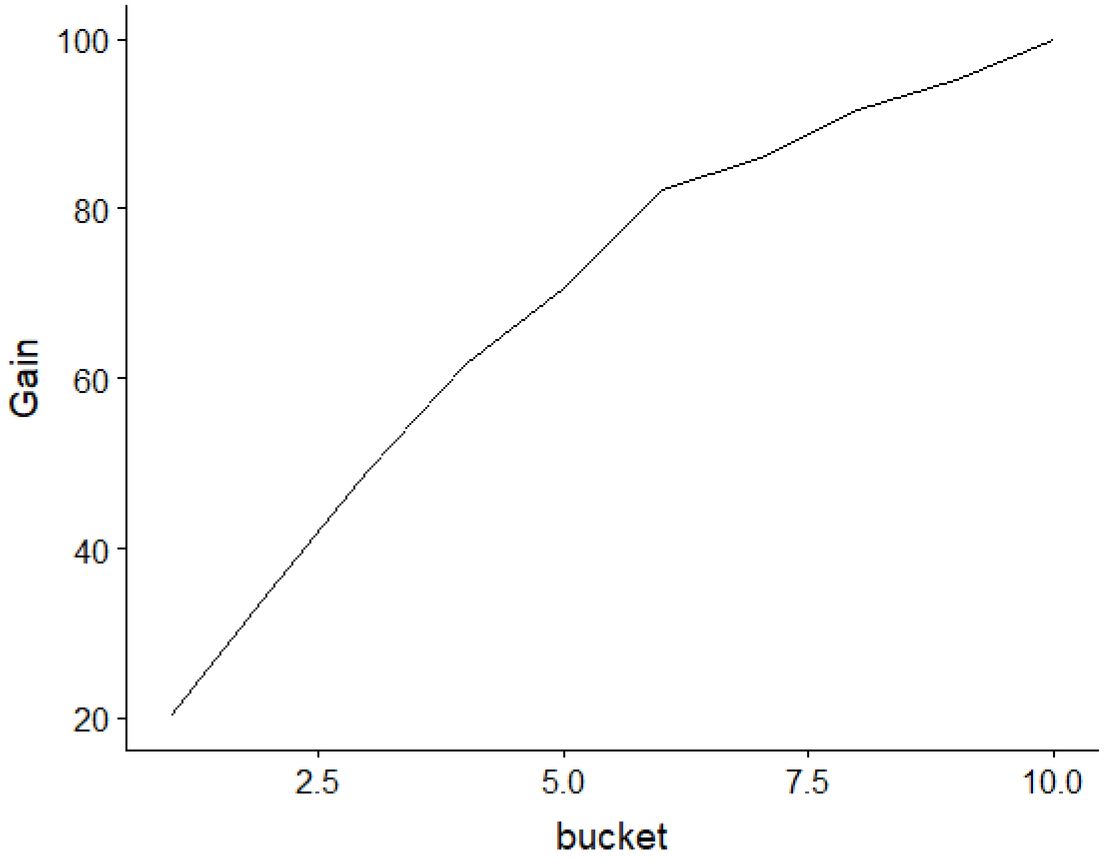
Accuracy	58.65%
Sensitivity	64.21%
Specificity	58.40%
KS Statistic	22.62%

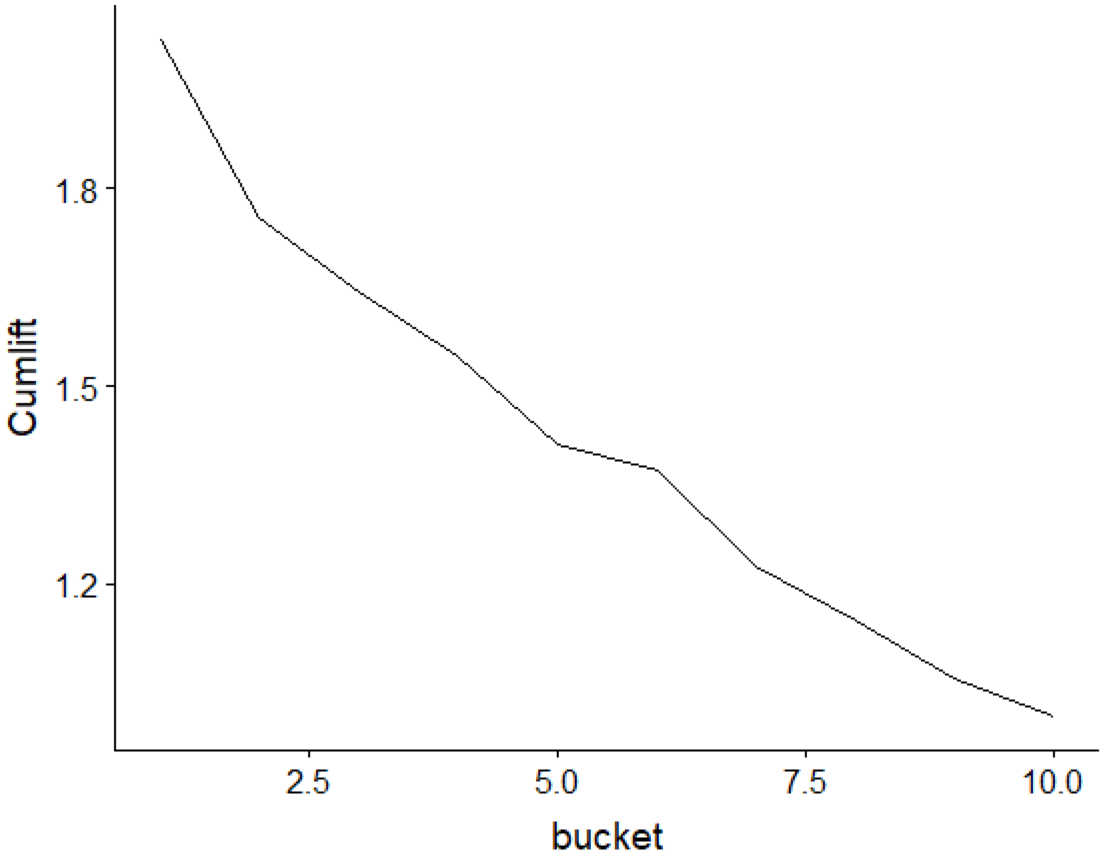
Gain and Lift Table

A tibble: 10 x 6

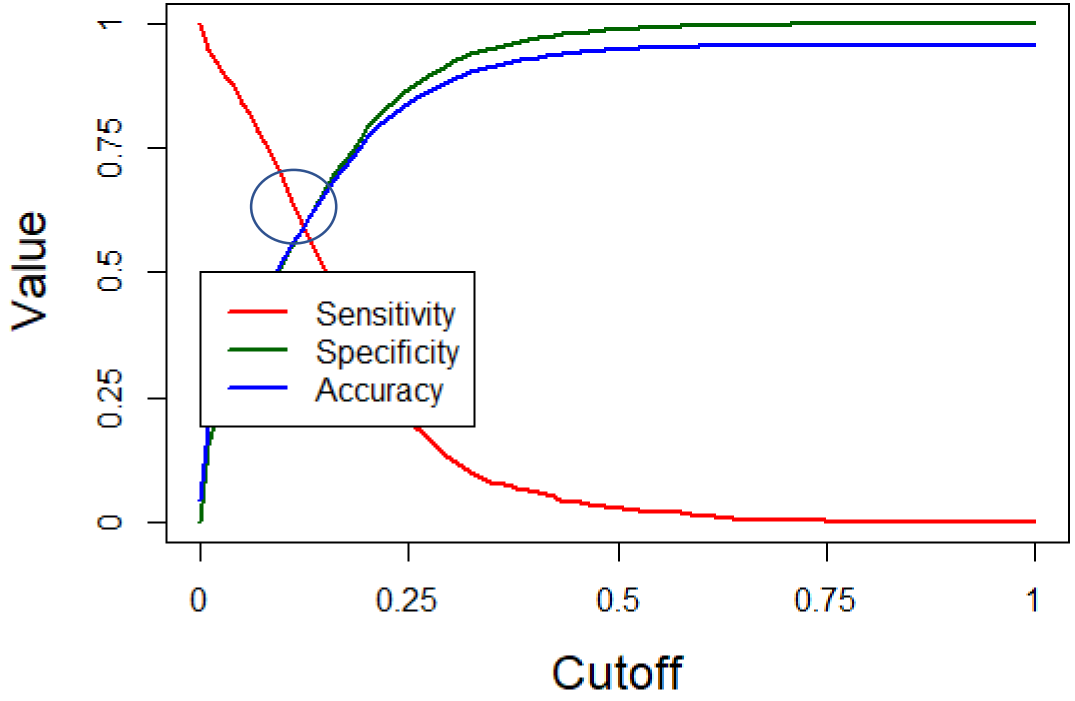
	bucket	total	totalresp	Cumresp	Gain	Cumlift
	<int>	<int>	<dbl>	<dbl>	<dbl>	<dbl>
1	1	2091	179	179	20.3	2.03
2	2	2091	131	310	35.1	1.76
3	3	2091	125	435	49.3	1.64
4	4	2091	111	546	61.8	1.55
5	5	2091	78	624	70.7	1.41
6	6	2091	103	727	82.3	1.37
7	7	2091	32	759	86.0	1.23
8	8	2091	50	809	91.6	1.15
9	9	2091	31	840	95.1	1.06
10	10	2090	43	883	100	1

Gain and Lift plot for the Binary tree 3





Modelling Default using complete data Random Forests: Model Evaluation



Choosing the cutoff probability as the intersection of the Sensitivity, Specificity and Acuuracy: 12.62%, confusion matrix gives

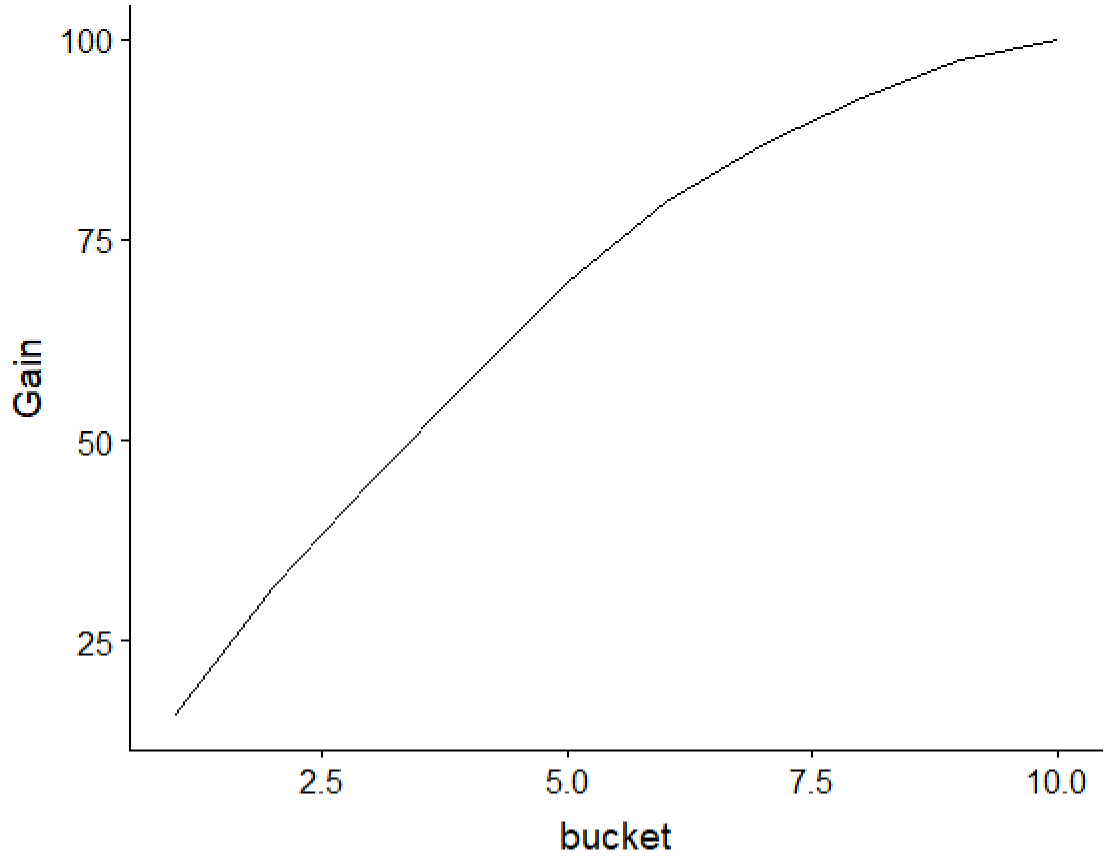
Accuracy	60.22%
Sensitivity	57.98%
Specificity	60.32%
KS Statistic	18.30%

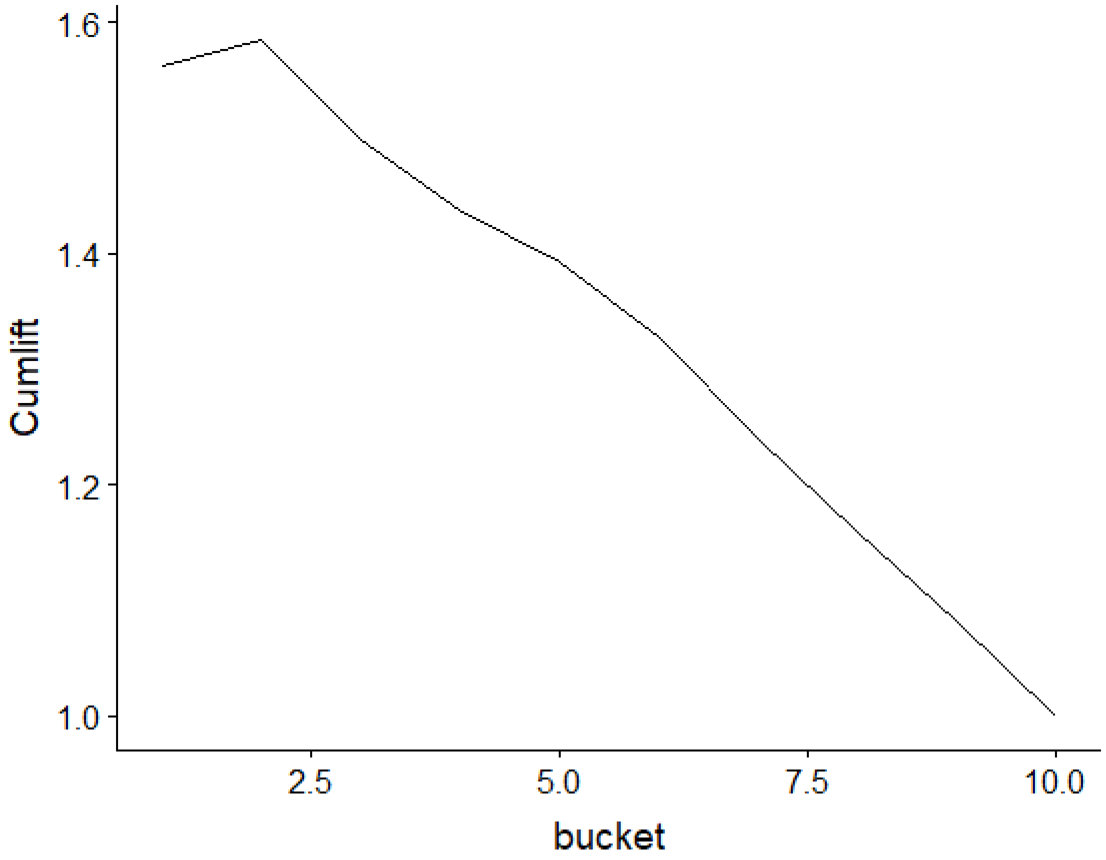
Gain and lift table

\# A tibble: 10 x 6

	bucket	total	totalresp	Cumresp	Gain	Cumlift
	<int>	<int>	<dbl>	<dbl>	<dbl>	<dbl>
1	1	2091	138	138	15.6	1.56
2	2	2091	142	280	31.7	1.59
3	3	2091	117	397	45.0	1.50
4	4	2091	111	508	57.5	1.44
5	5	2091	107	615	69.6	1.39
6	6	2091	88	703	79.6	1.33
7	7	2091	64	767	86.9	1.24
8	8	2091	52	819	92.8	1.16
9	9	2091	42	861	97.5	1.08
10	10	2090	22	883	100	1

Gain and Lift plots for random forests





Models’ Summary

Model	Accuracy	Sensitivity	Specificity	KS Statistic	Gain(first and second decile)	Lift(first and second decile)
Modelling using demographic data only:						
Logistic Regression	57.92%	58.89%	57.87%	13.84	17.7 33.9	1.77 1.69
Modelling using complete data:						
Logistic Regression	61.93%	63.31%	61.86%	25.17	19.3 37.4	1.93 1.87
Binary Tree	58.65%	64.21%	58.40%	22.62	20.3 35.1	2.03 1.76
Random Forest	60.22%	57.98%	60.32%	18.3	15.6 31.7	1.56 1.59

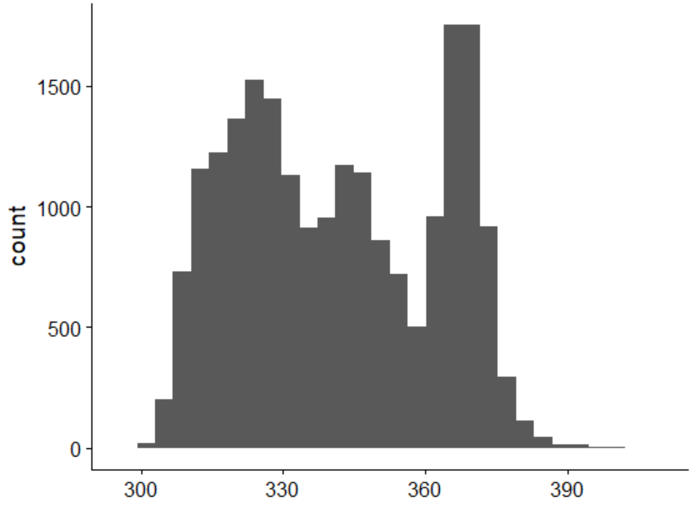
Models using complete data have more predictive power than the models using only demographic data

Model using Logistic regression gave the best Sensitivity (True positive is most important for default), Gain,lift and KS statistic and hence chosen as the best model

Application Scorecard Generation

- The application scorecard is built using the chosen model with the good to bad odds of 10 to 1 at a score of 400 doubling every 20 points.
- Linear relationship between ln(odds) and the application score helped derive the equation of the straight line
  - 400= m*ln10 +c
  - 420=mln20 + c
- Solving the above linear equations, m= 28.85 and c=333.56
- Application score = 28.85 * log(odds) + 333.56
- The Cutoff probability for the model is 52.02%
- Therefore cutoff odds is 0.9223 and the **Cutoff Application Score is 331.23**

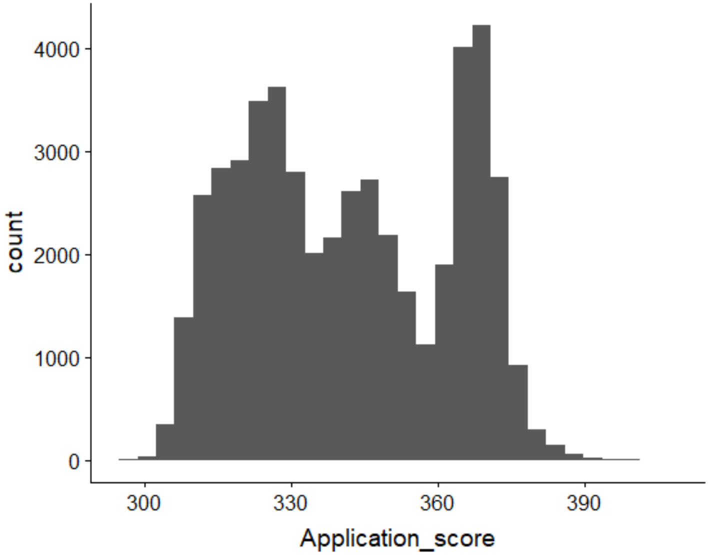
Application Scorecard charts for Test, Train and Rejected candidates



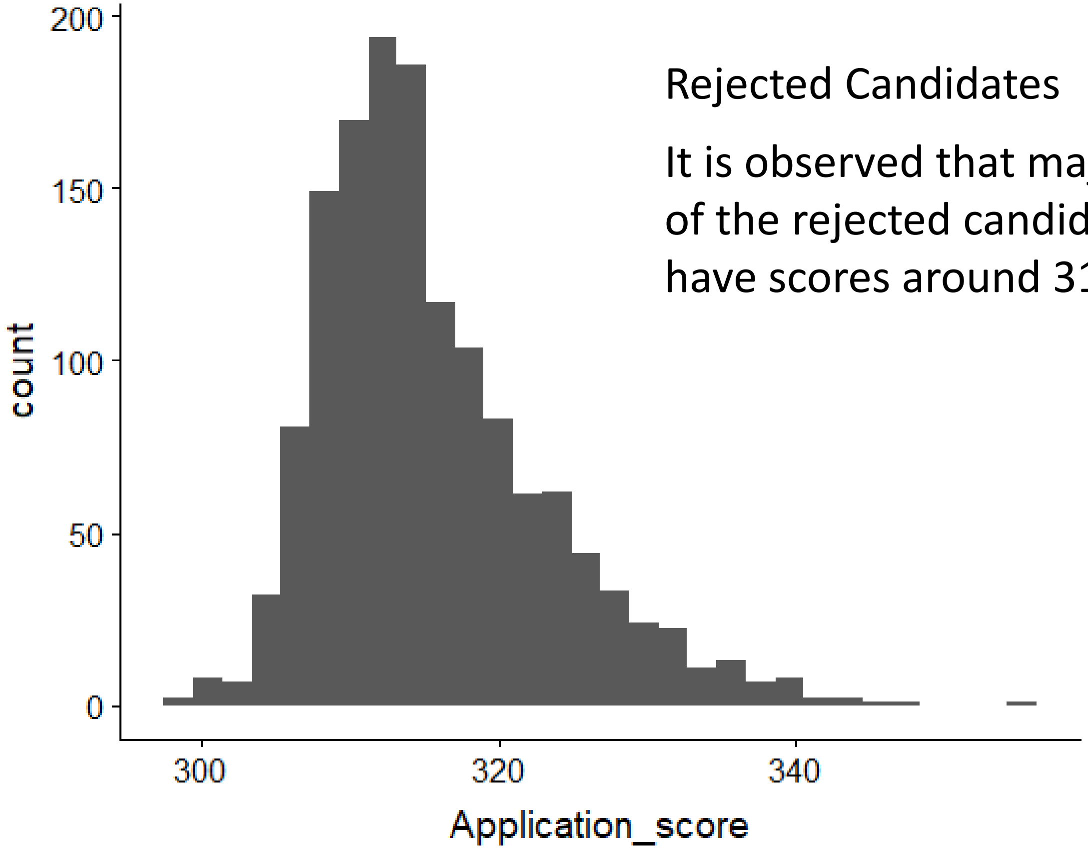
Test data

Majority of the test and train data applicants have scores around 370

Train data



Based on the cutoff probability, cutoff score is 331.23



Rejected Candidates

It is observed that majority of the rejected candidates have scores around 315

Financial benefit assessment

Based on the model, Percentage of applicants automatically accepted and rejected are calculated as follows

- Originally Rejected applicants =**1425**;
  - 1425/71124*100 = **2.003%**
- Considering all 3 test, train and rejected (originally) applicants
- Generating scorecard and counting number above cutoff , we get
- Percentage_rejected_with_model = Number of applicants below cutoff score in test,train and reject(originally) ÷ Total number of applicants in test,train and reject(originally) = (1362+8194+19036)/(1425+20909+48790) = **40.20%**
- Total number of candidates rejected by model = 28592
- Originally rejected candidates = 1425
- Additional candidates rejected by model = 28592-1425=27167
- Additional Percentage rejected by model = 38.20%

Financial Assessment

- Revenue loss: due to more rejected applicants
  - % Rejected applicants with no model = 1425/71124*100 = 2.003%
  - % Rejected applicants after model is applied = number with application score < 331.23 (test + train + originally rejected)/71124 = (1362+8194+19036)/71124*100 = 40.20%
  - **This leads to revenue loss of 38.20% due to additional customer rejections**
- Addition of revenue from good applicants in the originally rejected group
  - Number of applicants in originally rejected group (Performance.Tag == NA)= 1425
  - Applicants with score > 331.23 in this group = 63
  - **Revenue gain from 4.421% applicants**

Financial Assessment

- Credit Loss saved due to model
  - % Defaults without model = Defaults(number of applicants with Performance tag =1)/ Total approved applicants*100 =2945 /69699 *100 = 4.23%
  - % Defaults with model = Defaults after using model(score >= 331.23)/Total approved applicants *100 = 1070 /69699 *100 = 1.54%
  - **2.69% credit loss saved by model**
  - % of defaults avoided by using model = (Defaults with score < 331.23)/(Actual Defaults)*100 = (1875/2945)*100 = 63.67%
  - Average credit loss per default = Average outstanding balance of defaulters in test and train = 1260675
  - Amount saved by the model = Defaulters rejected by model* Average credit loss per default = 1875* 1260675 = 2363765384(2.36 bn)

Conclusions

- The significant influencers of default are (top 5)
  - Avgas.CC.Utilization.in.last.12.months
  - No.of.trades.opened.in.last.12.months
  - No.of.PL.trades.opened.in.last.12.months_woe
  - No.of.Inquiries.in.last.12.months..excluding.home...auto.loans._woe
  - Outstanding.Balance
- Logistic regression model is used to calculate probability of default and application scores
- Application score > 331.23 is the criterion for acceptance
- Implications of model
  - 2.69% credit loss saved by model
  - 63.67% of defaulters avoided
  - 4.42% good candidates from originally rejected group now accepted
  - 38.20% potential revenue loss due to additional rejection