CredX Acquisition Analytics

BFS Final Submission

GROUP MEMBERS:

1. CHAYAN NASKAR
2. AVANISH KUMAR
3. ASA SINGH

Business Understanding

- CredX is a leading credit card provider that receives thousands of credit card applications every year.
- The CEO believes that the best strategy to mitigate credit risk is to '**acquire the right customers**'
- Problem Statement:
  - Help CredX identify the right customers using predictive models. Using past data of the bank's applicants,
    - Determine the factors affecting credit risk.
    - Create strategies to mitigate the acquisition risk.
    - Assess the financial benefit of your project.
- Assumptions made for revenue calculations:
  - CredX makes a revenue of $1,000 per good customer

Data Understanding

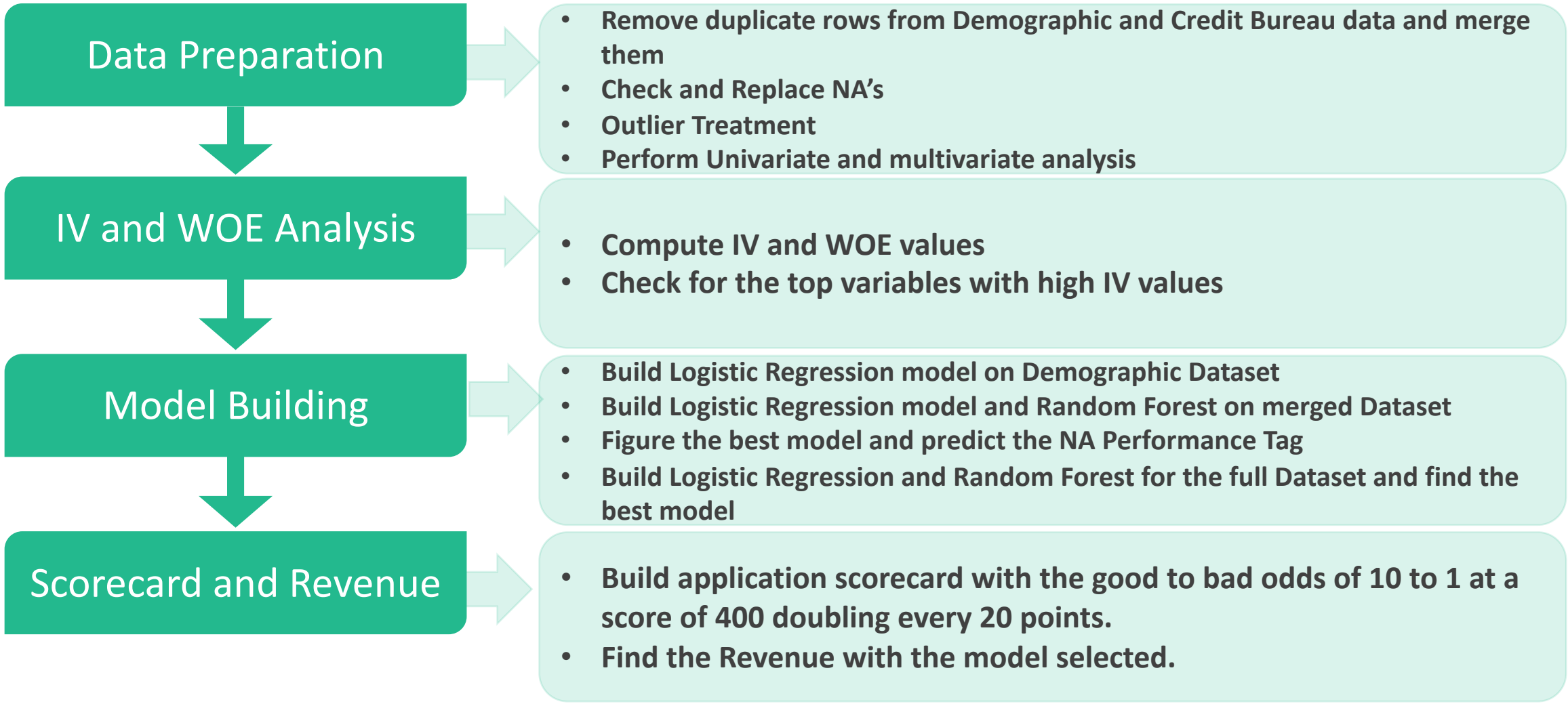
Demographic Data

- Age
- Gender
- Marital Status
- No of Dependents
- Income
- Education
- Profession
- Type of Residence
- No of months in current residence
- No of Months in current Company

Credit Bureau Data

- Average Credit Card Utilization
- DPD’s
- Outstanding Balance
- No of Trades
- Loan Inquiries
- PL Trades
- Open Home Loan
- Open Auto Loan

Problem Solving Methodology



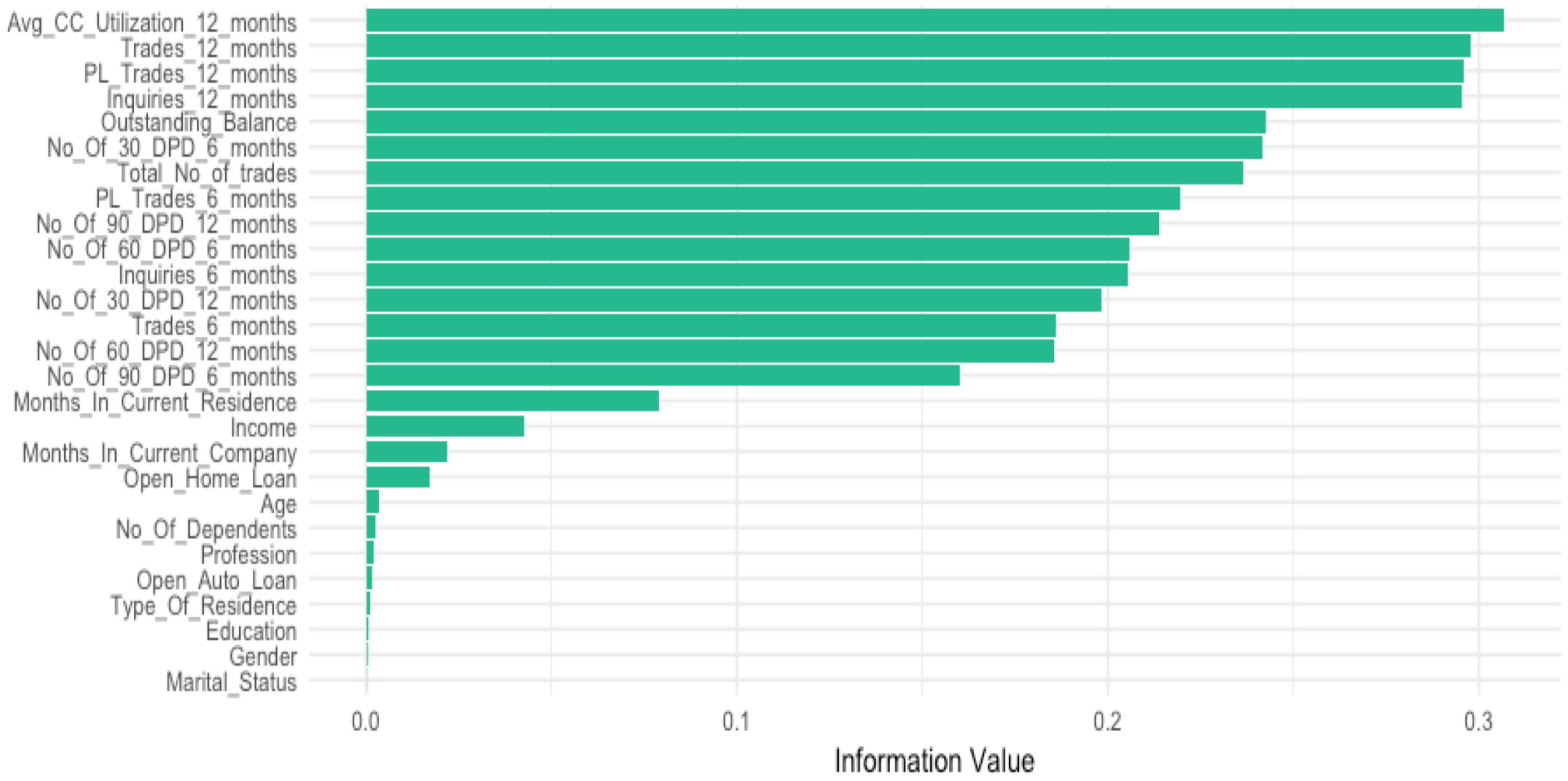
Information Value

The top 6 variables with high Information value are:

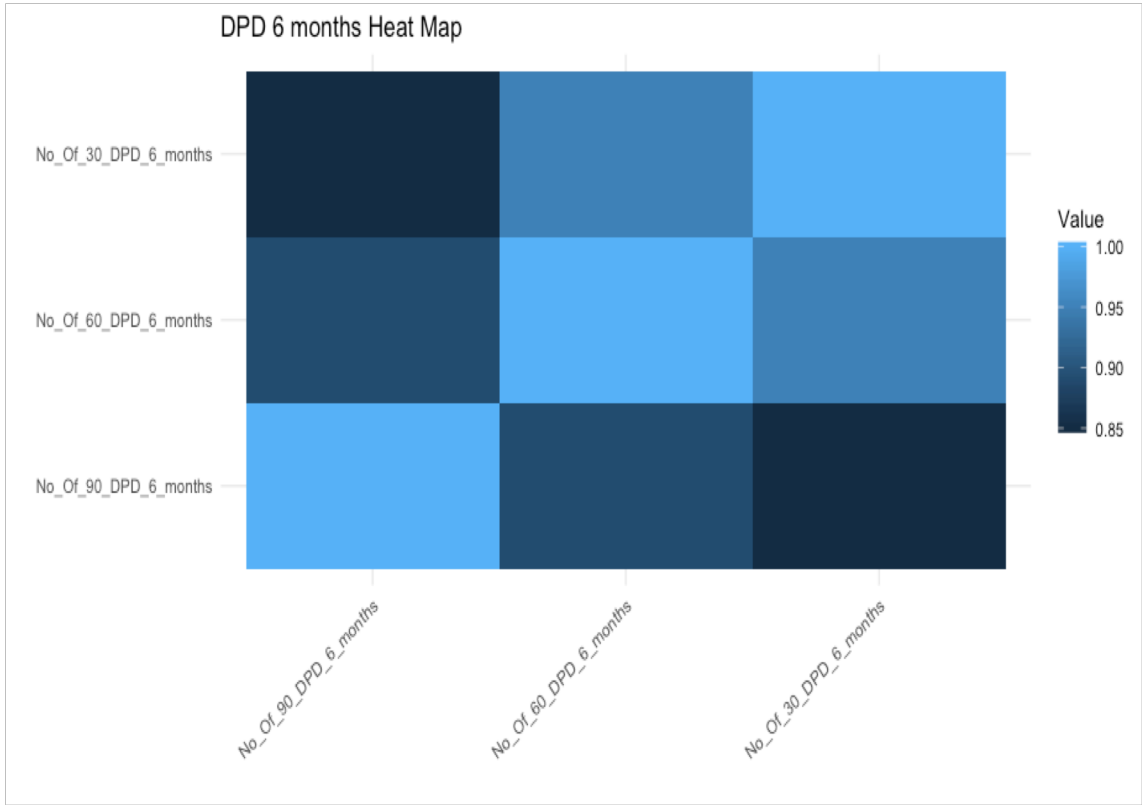
- Average Credit Card utilization
- Trades 12 months
- Pl Trades 12 months
- Inquiries 12 months
- Outstanding Balance
- 30 DPD in 6 months

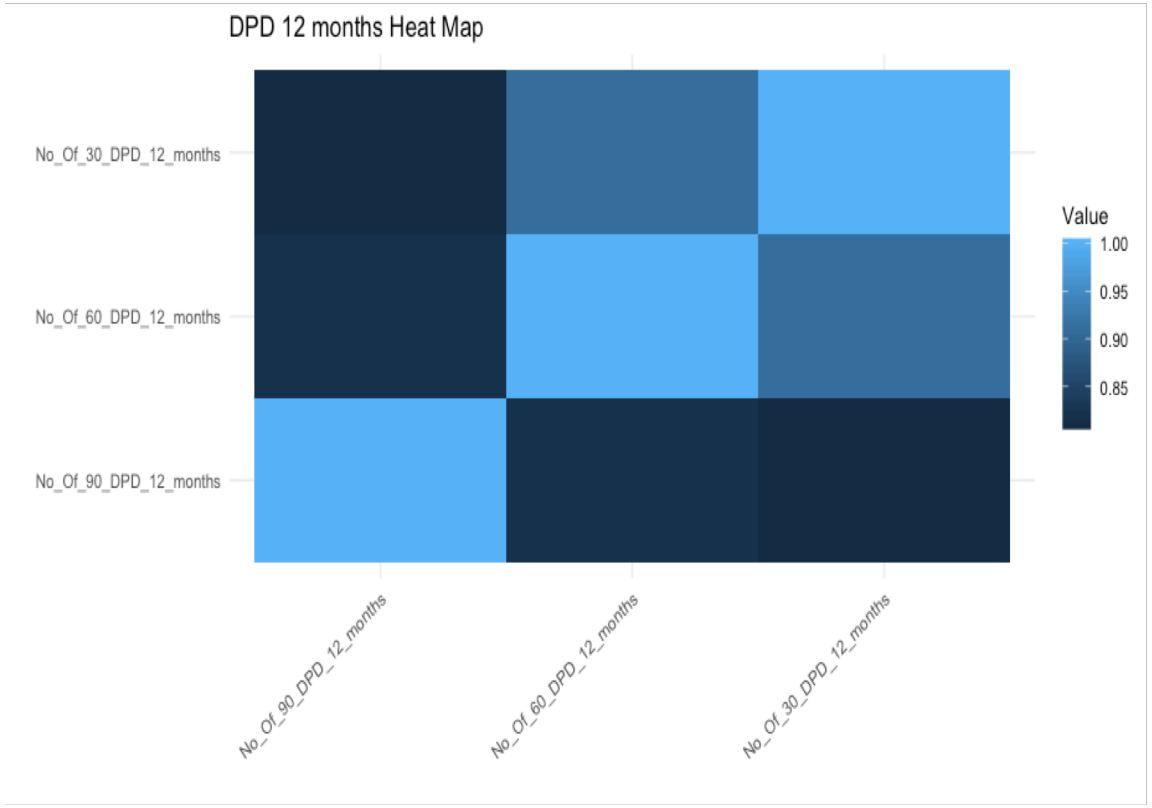
	Variable	IV
17	Avg_CC_Utilization_12_months	3.068152e-01
19	Trades_12_months	2.979571e-01
21	PL_Trades_12_months	2.958955e-01
23	Inquiries_12_months	2.954243e-01
25	Outstanding_Balance	2.428344e-01
13	No_Of_30_DPD_6_months	2.415627e-01
26	Total_No_of_trades	2.366049e-01
20	PL_Trades_6_months	2.197050e-01
14	No_Of_90_DPD_12_months	2.138748e-01
12	No_Of_60_DPD_6_months	2.058339e-01
22	Inquiries_6_months	2.051870e-01
16	No_Of_30_DPD_12_months	1.982549e-01
18	Trades_6_months	1.860015e-01
15	No_Of_60_DPD_12_months	1.854989e-01
11	No_Of_90_DPD_6_months	1.601169e-01
9	Months_In_Current_Residence	7.894353e-02
5	Income	4.241780e-02
10	Months_In_Current_Company	2.175441e-02
24	Open_Home_Loan	1.696972e-02
1	Age	3.349157e-03
4	No_Of_Dependents	2.647040e-03
7	Profession	2.228309e-03
27	Open_Auto_Loan	1.654820e-03
8	Type_Of_Residence	9.252553e-04
6	Education	7.822023e-04
2	Gender	3.255737e-04
3	Marital_Status	9.592186e-05

Information Value Plot

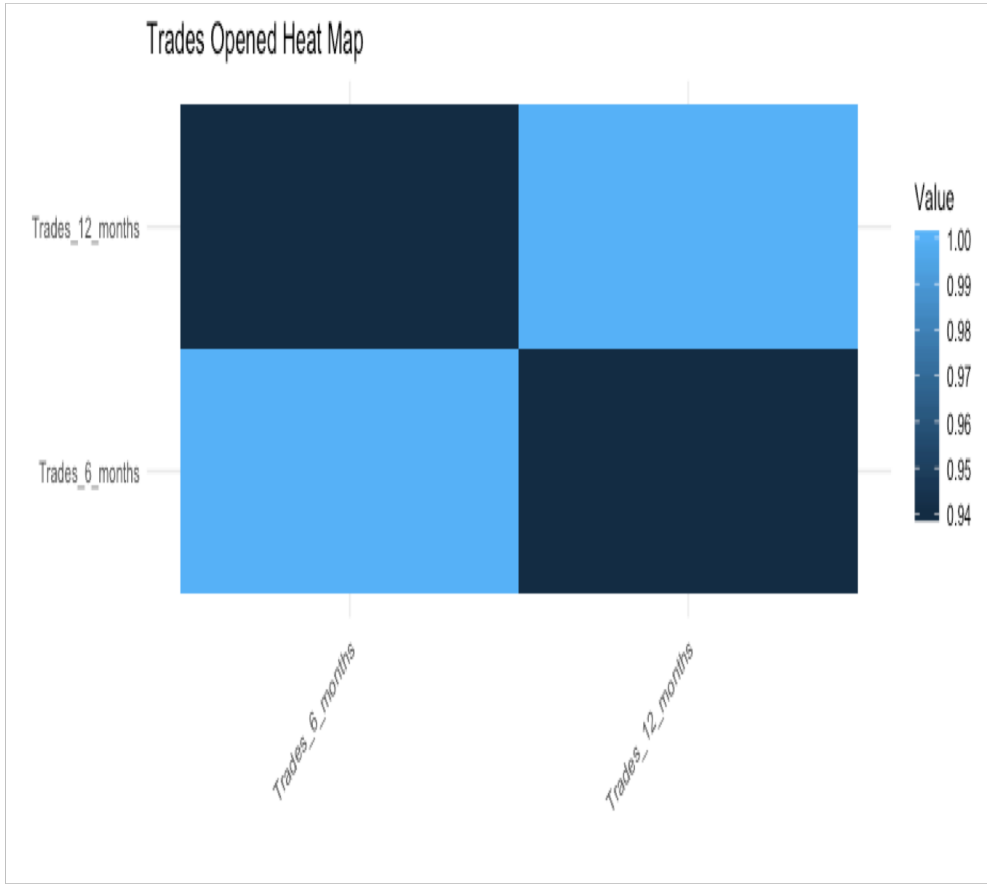


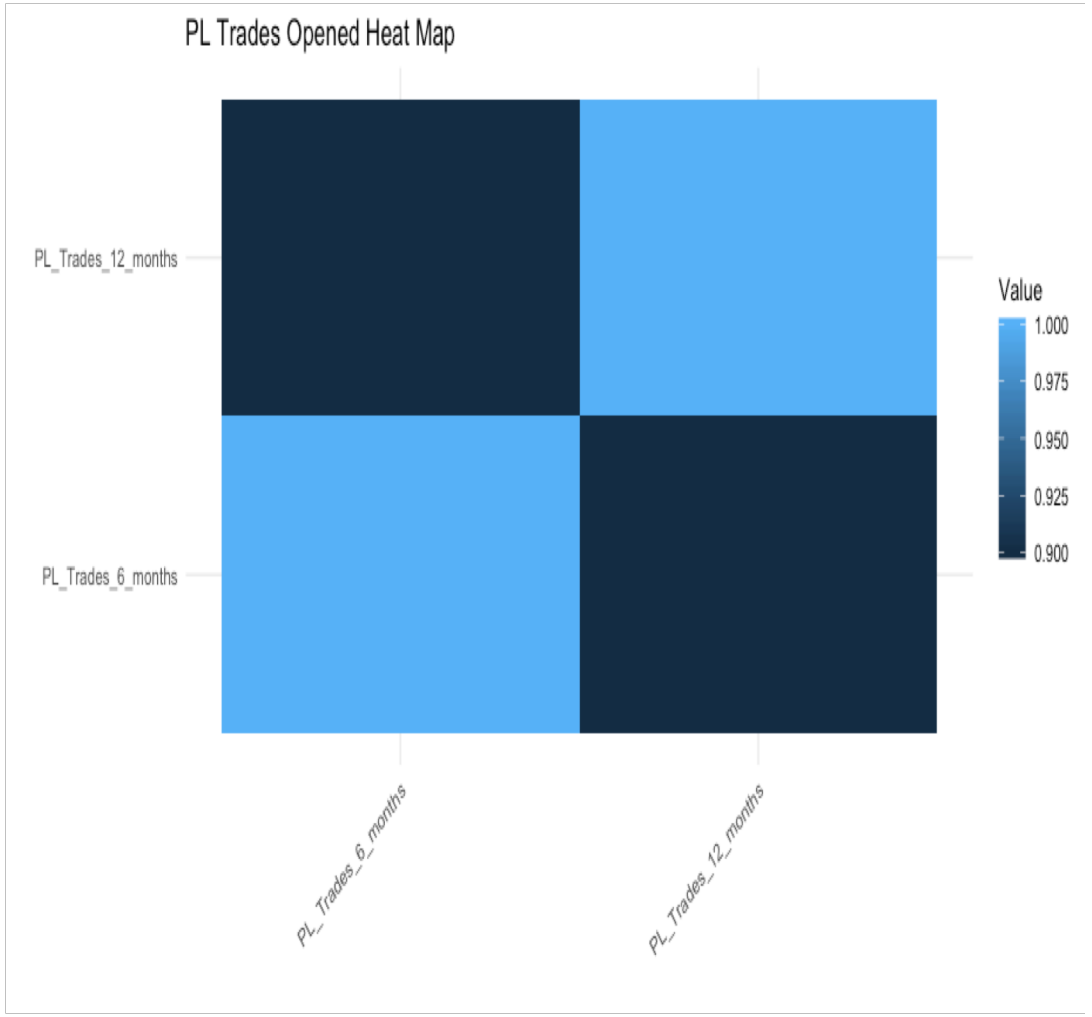
DPD’s Heat Map





Trades and PL Trades Heat Map





Predictive Modelling

- First the demographic dataset is used to build the model then full merged dataset is used for finding the best model.
- The best model is then used to predict the Performance for the dataset with Performance Tag as NA.
- The training data is being balanced using **'ROSE'**.

Demographic Dataset Logistic Regression

Model Metrics	Value
Accuracy	0.56
Sensitivity	0.54
Specificity	0.56

Merged Dataset Logistic Regression

Model Metrics	Value
Accuracy	0.63
Sensitivity	0.62
Specificity	0.63

Merged Dataset Random Forest

Model Metrics	Value
Accuracy	0.64
Sensitivity	0.62
Specificity	0.64

Final Model Used

Logistic regression is applied to dataset with data both NA's and Non-NA's Performance Tag.
The NA's were imputed using the prediction from the best of previous models.

Logistic Regression on Full Dataset

Model Metrics	Value
Accuracy	0.69
Sensitivity	0.70
Specificity	0.69

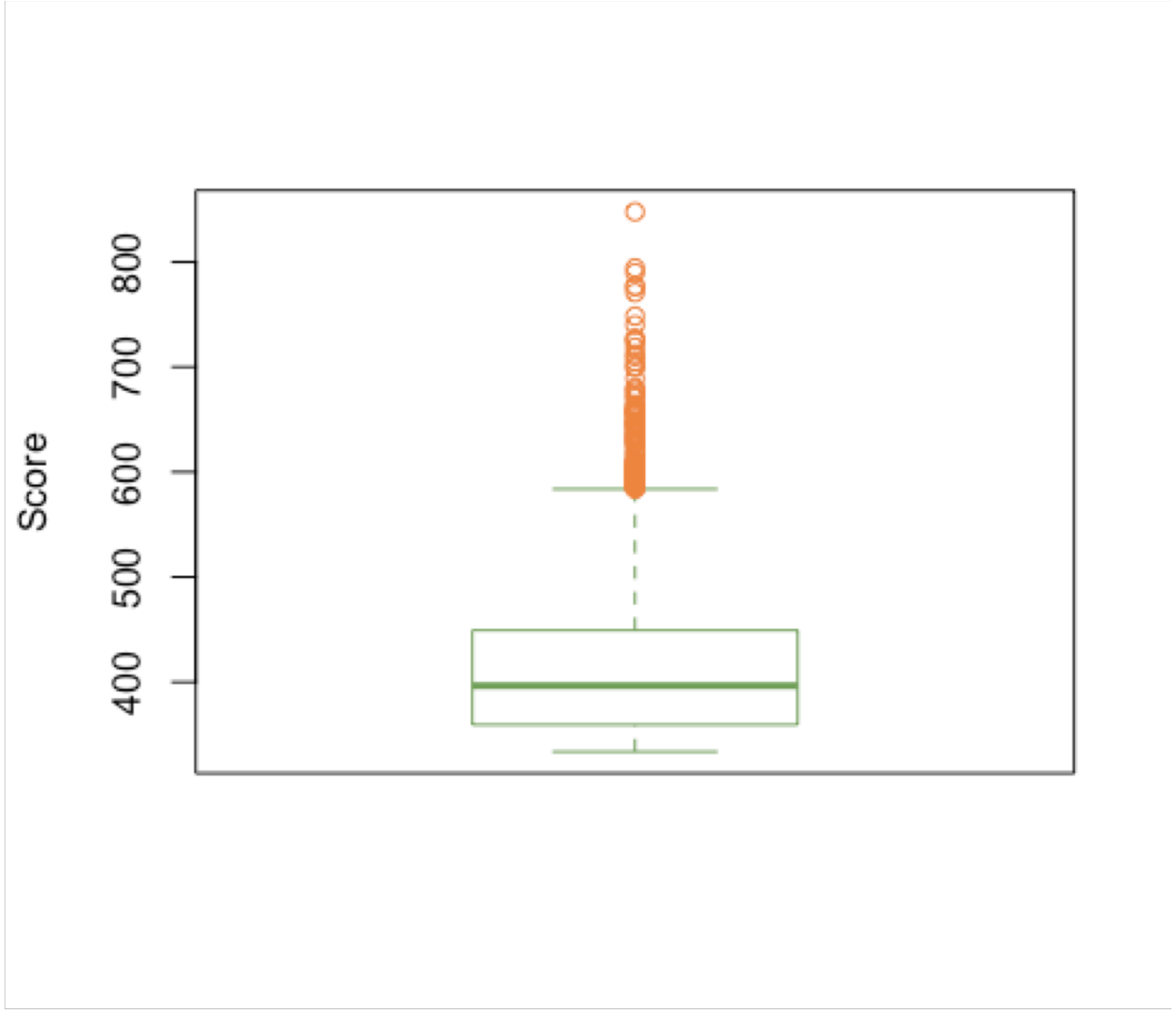
Random Forest on Full Dataset

Model Metrics	Value
Accuracy	0.62
Sensitivity	0.64
Specificity	0.62

Hence we can see the best model is Logistic Regression with Accuracy of about 69%, Sensitivity of about 70% and Specificity of about 69%.

Application Scorecard

- Cutoff-Score: 359
- Total rejected population: 17615
- Non-Defaulters acquired by the model: 97%
- Percentage of Defaulters acquired by the model: 3%



Revenue

The following are the revenue metrics by using the proposed model:

- Suppose the Bank makes $1,000 per good customer,
  - The revenue loss will be $ 14.9 million for rejecting good customer.
  - The revenue loss will be $ 1.6 million for acquiring defaulting customer.
  - Revenue gained will be $ 51.9 million for acquired non-defaulting customer.
- The default percentage for the model used is 3%.