Financial Risk & Analytics Assignment

Mohamed Yusuf S - PGPBABI Aug 19 Batch

Exploratory Data Analysis

Basic Statistics

Raw Counts

Name	Value
Rows	3,541
Columns	52
Discrete columns	8
Continuous columns	43
All missing columns	1
Missing observations	13,548
Complete Rows	0
Total observations	184,132
Memory allocation	2.1 Mb

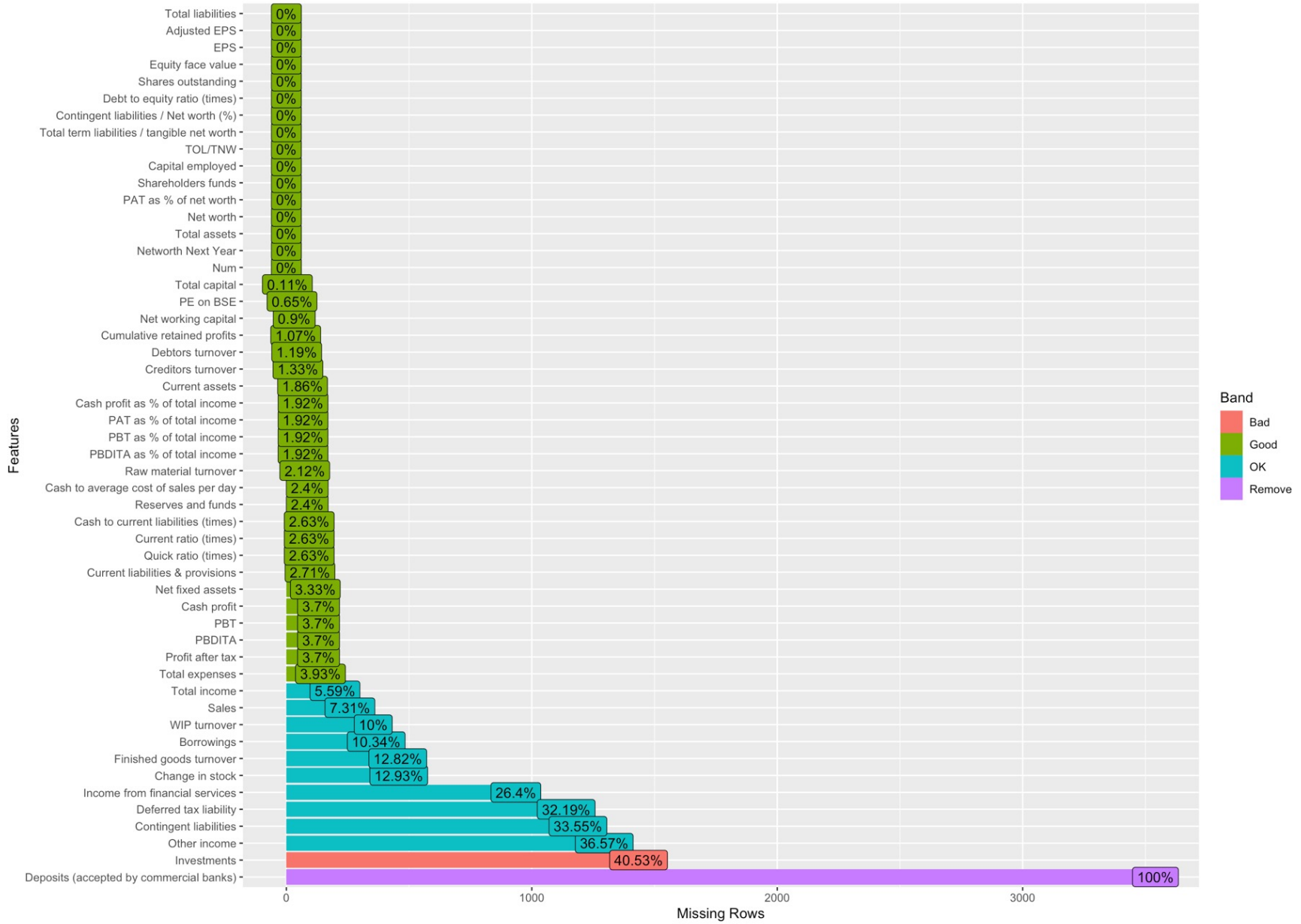
```
── Data Summary ────────────────────────
                           Values
Name                       coydefault
Number of rows             3541
Number of columns          52
_____________________________
Column type frequency:
  character                8
  logical                  1
  numeric                  43
_____________________________
Group variables            None
```

Exploratory Data Analysis

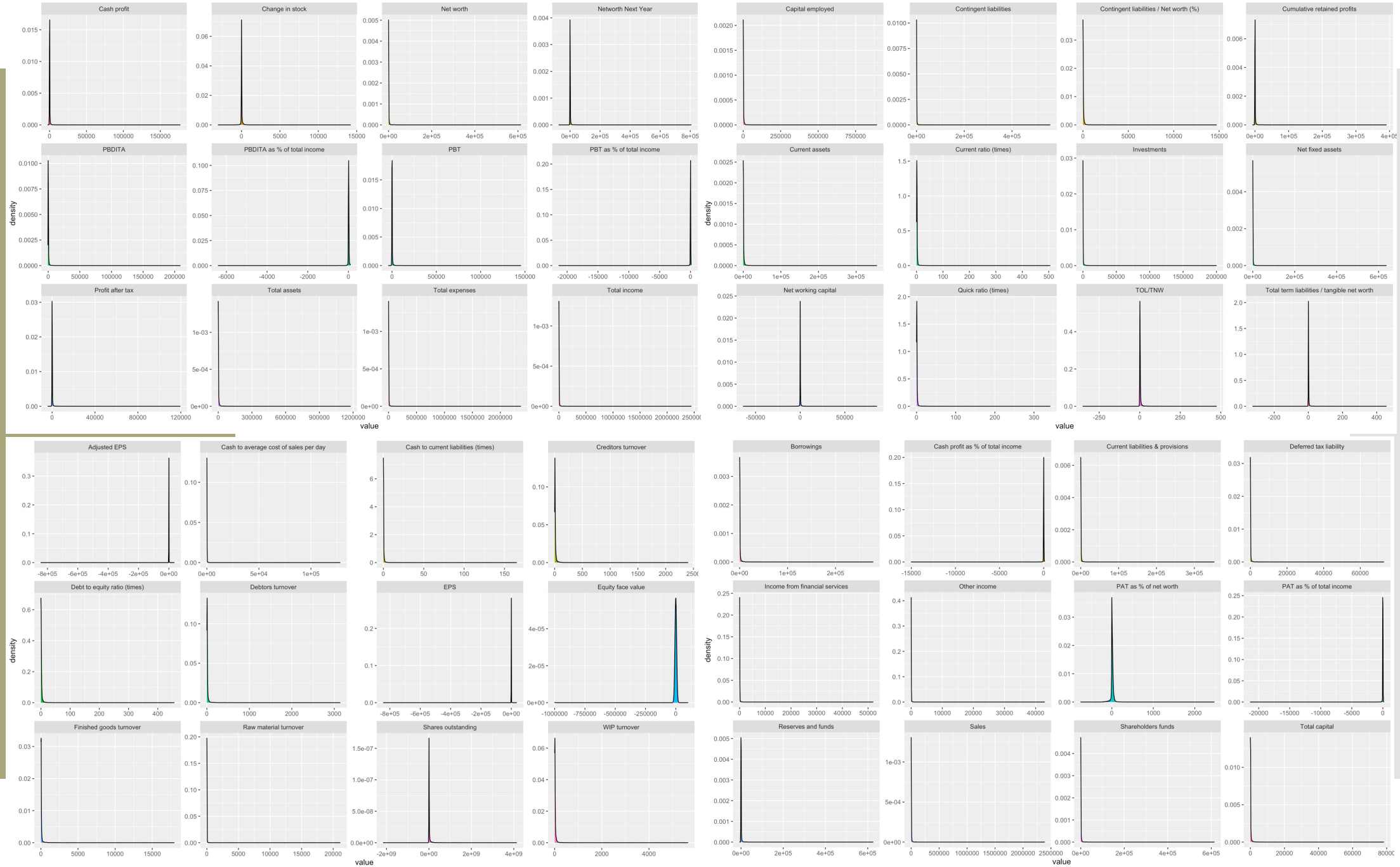
- The given dataset contains 3541 records and 5 2 variables on it. The column NUM can be removed which specifies only the record count.
- All the variables are converted to numeric type.
- Variable Name deposited is totally incomplete and it can be eliminated.
- More than 5 variables have missing values above 26% and investment variable has 40% of missing values to it.

Exploratory Data Analysis

Missing Data Profile



EDA – Data Distribution

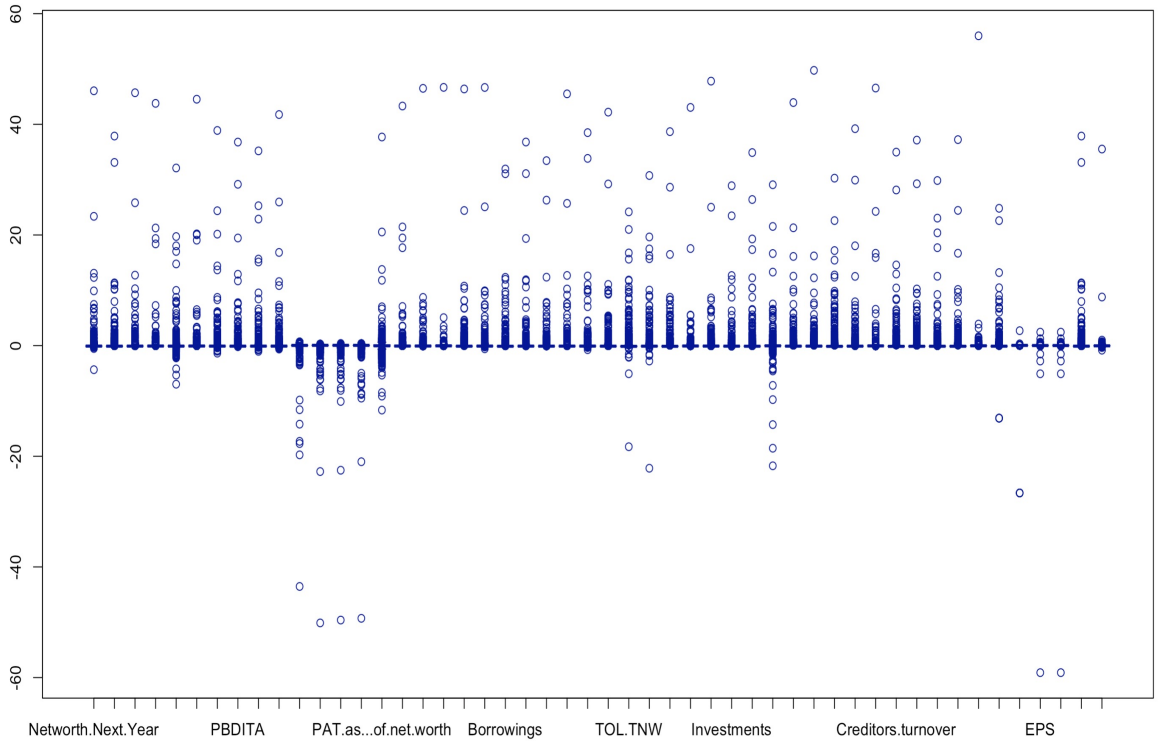


EDA

Outliers' detection

- From the below plot 1.01 , we could see lot of outliers exist in most of the variables.
- Hence it has to be tested statistically and we shall do an IQR test to statistically identify the outliers.
- The plot 1.02 shows the Inter Quartile Range and existence of outliers.
- So let caps and floor the outliers to avoid extreme values.

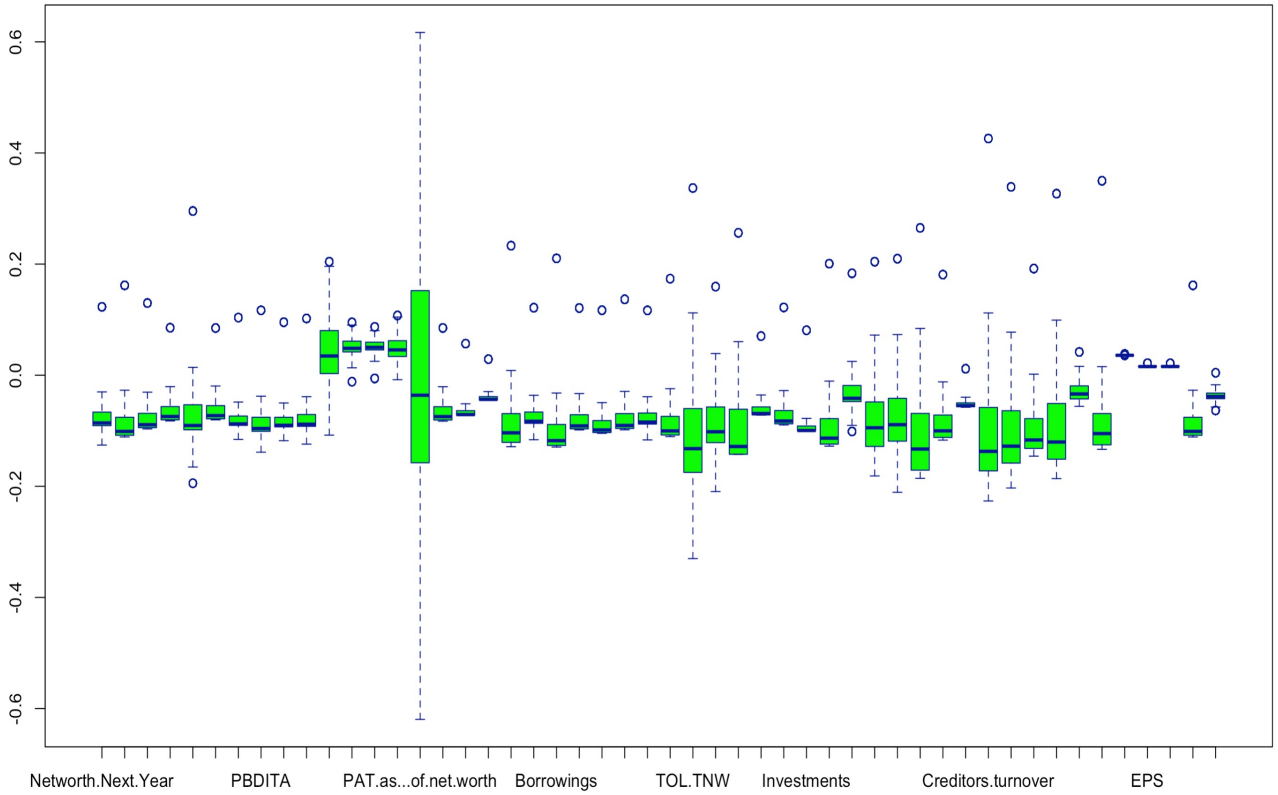
Plot : 1.01



Before Outlier Treatment

Post Outlier Treatment

Plot : 1.02

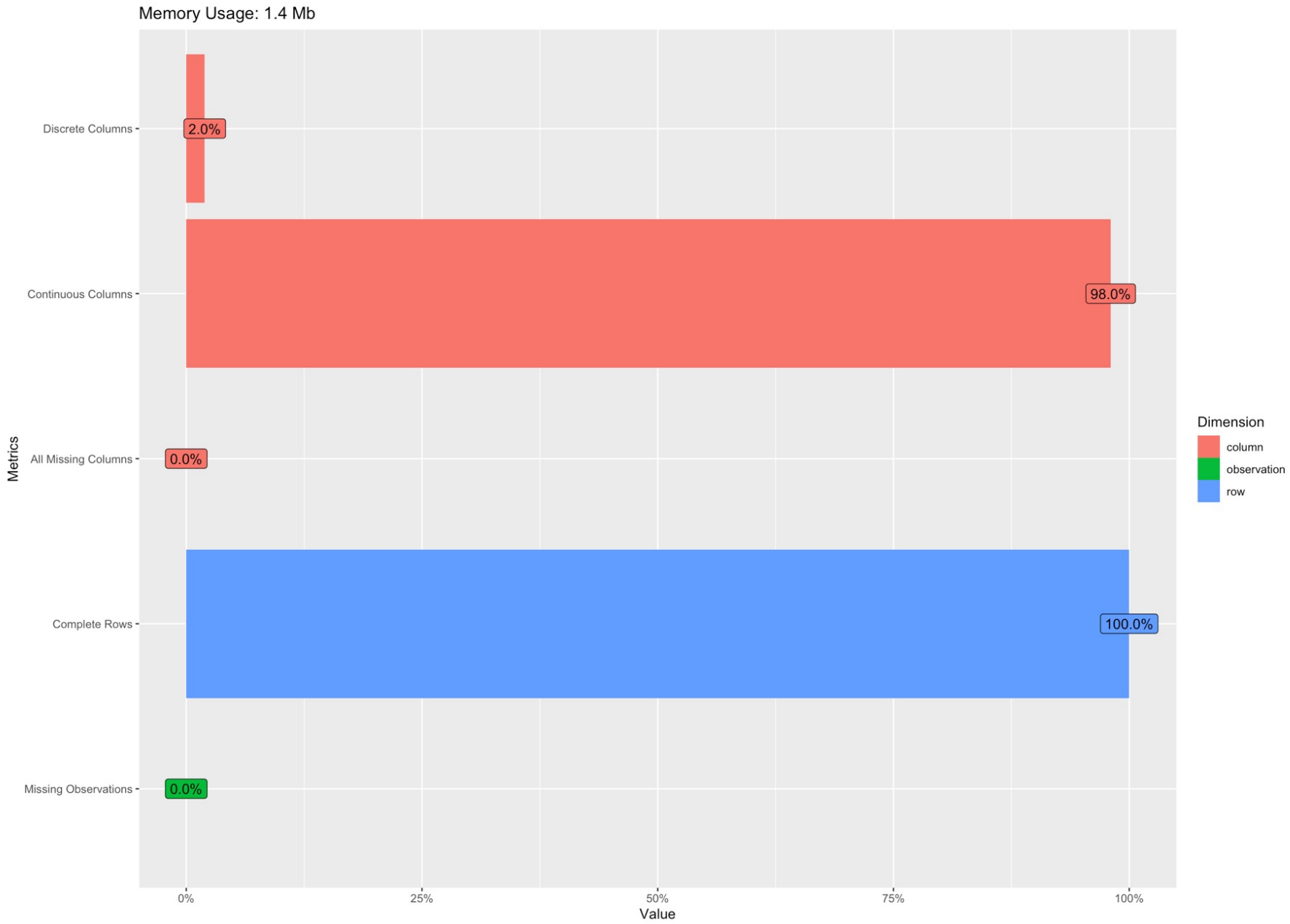


EDA

Missing
Value Imputation

using MICE

Percentages



New Variables Addition

We have created four new variables namely **Profitability**, **Leverage**, **Liquidity**, and **Company's Size**

1. Networking Capital / Current Liabilities explains how much portion of all liabilities of company can be met from the current working capital.

2. Total Borrowing / Total Asset explains how much borrowings can be paid from Total Asset the company owns.

3. Profit After Tax / Total Income helps to identify how a company is making profit as a percentage from the total income.

4. Net Cash Available / Total Asset helps to know how the company is holding as cash from Total Asset.

Multi-collinearity Check may not be used for a Logistics Model and hence let's build a model to identify the multi-collinearity with NA's

A **model 1** is built after removing the column 1 (Num – which represents sequential number) and column 22 (Deposits - With fully incomplete rows). In this model we have taken all the other variables.

```
                                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                             -4.153e+00  4.811e+00  -0.863 0.388049
Total.assets                             6.549e-03  5.457e+00   0.001 0.999042
Net.worth                                1.958e+00  6.494e+00   0.301 0.763063
Total.income                             2.655e+00  4.370e+00   0.608 0.543460
Change.in.stock                          7.640e-01  7.910e-01   0.966 0.334132
Total.expenses                          -3.383e+00  4.636e+00  -0.730 0.465586
Profit.after.tax                        -2.110e+00  4.768e+00  -0.443 0.658044
PBDITA                                  -3.032e+00  3.051e+00  -0.994 0.320231
PBT                                      1.116e+00  3.841e+00   0.291 0.771406
Cash.profit                             -8.121e-01  3.632e+00  -0.224 0.823048
PBDITA.as...of.total.income             -7.047e-01  2.028e+00  -0.348 0.728163
PBT.as...of.total.income                -6.368e+00  8.464e+00  -0.752 0.451856
PAT.as...of.total.income                -4.324e+00  8.674e+00  -0.499 0.618118
Cash.profit.as...of.total.income        -8.028e+00  6.134e+00  -1.309 0.190623
PAT.as...of.net.worth                   -2.851e+00  4.666e-01  -6.109    1e-09 ***
Sales                                    2.491e+00  3.690e+00   0.675 0.499684
Income.from.financial.services           2.517e+00  2.343e+00   1.075 0.282577
Other.income                             3.772e+00  4.043e+00   0.933 0.350810
Total.capital                           -3.496e+00  1.522e+00  -2.296 0.021660 *
Reserves.and.funds                       1.513e+00  2.654e+00   0.570 0.568597
Borrowings                               4.449e+00  2.648e+00   1.680 0.092892 .
Current.liabilities...provisions         4.107e+00  2.596e+00   1.582 0.113736
Deferred.tax.liability                   6.425e-01  1.520e+00   0.423 0.672504
Shareholders.funds                       4.199e+00  6.265e+00   0.670 0.502722
Cumulative.retained.profits             -1.538e+01  6.551e+00  -2.348 0.018887 *
Capital.employed                        -9.708e+00  5.256e+00  -1.847 0.064751 .
TOL.TNW                                  2.677e+00  7.930e-01   3.376 0.000735 ***
Total.term.liabilities...tangible.net.worth -1.592e+00  1.590e+00  -1.001 0.316634
Contingent.liabilities...Net.worth....   8.759e-01  6.490e-01   1.350 0.177109
Contingent.liabilities                  -4.725e+00  2.487e+00  -1.900 0.057399 .
Net.fixed.assets                        -3.315e-01  2.758e+00  -0.120 0.904329
Investments                             -3.665e-02  1.560e+00  -0.023 0.981253
Current.assets                          -4.421e+00  2.531e+00  -1.747 0.080669 .
Net.working.capital                      1.076e+00  2.453e+00   0.439 0.660915
Quick.ratio..times.                      4.114e-01  1.875e+00   0.219 0.826272
Current.ratio..times.                   -4.630e+00  1.687e+00  -2.744 0.006061 **
Debt.to.equity.ratio..times.             4.318e+00  1.162e+00   3.715 0.000203 ***
Cash.to.current.liabilities..times.      3.956e+00  1.433e+00   2.760 0.005781 **
Cash.to.average.cost.of.sales.per.day    3.125e+00  4.623e+00   0.676 0.499015
Creditors.turnover                      -5.278e-01  6.161e-01  -0.857 0.391544
Debtors.turnover                         5.248e-02  6.336e-01   0.083 0.933986
Finished.goods.turnover                  1.199e+00  9.967e-01   1.203 0.228869
WIP.turnover                            -1.150e+00  7.241e-01  -1.588 0.112317
Raw.material.turnover                   -4.289e+00  4.213e+00  -1.018 0.308653
Shares.outstanding                       1.633e+00  7.188e-01   2.271 0.023126 *
Equity.face.value                        1.446e+02  1.239e+02   1.166 0.243416
EPS                                     -3.030e+02  3.241e+02  -0.935 0.349700
Adjusted.EPS                            -6.891e+00  2.920e+02  -0.024 0.981173
Total.liabilities                               NA         NA      NA       NA
PE.on.BSE                               -1.085e+01  6.230e+00  -1.741 0.081613 .
NWCbyCL                                 -5.926e-02  1.578e-01  -0.376 0.707213
Borrowings_by_Total_asset                1.717e-01  1.708e-01   1.005 0.314687
profit_by_income                        -5.644e-01  2.021e-01  -2.792 0.005233 **
equity_by_asset                                 NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1770.97  on 3540  degrees of freedom
Residual deviance:  960.07  on 3489  degrees of freedom
AIC: 1064.1
```

Logistics Regression

Model 2

A Model 2 is build based on predictor variables that are significant from the previous model 1. Few variables are removed considering less significant after running this model2.

```
Call:
glm(formula = defualt ~ PBT + PAT.as...of.net.worth + Borrowings +
    Cumulative.retained.profits + Capital.employed + TOL.TNW +
    Current.ratio..times. + Debt.to.equity.ratio..times. + Cash.to.current.liabilities..times. +
    Raw.material.turnover + profit_by_income, family = "binomial",
    data = p.data.final_rem_1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9892  -0.2636  -0.1649  -0.0706   3.3174

Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                       -4.6439     0.5642  -8.231  < 2e-16 ***
PBT                               -3.0267     3.2554  -0.930 0.352501
PAT.as...of.net.worth             -4.3168     0.3709 -11.639  < 2e-16 ***
Borrowings                         1.3781     1.3032   1.057 0.290315
Cumulative.retained.profits      -13.1696     6.2264  -2.115 0.034419 *
Capital.employed                  -5.8858     2.2502  -2.616 0.008905 **
TOL.TNW                            2.6471     0.6931   3.819 0.000134 ***
Current.ratio..times.             -4.7675     1.0755  -4.433 9.30e-06 ***
Debt.to.equity.ratio..times.       3.0718     0.8390   3.661 0.000251 ***
Cash.to.current.liabilities..times.  4.0022     1.0433   3.836 0.000125 ***
Raw.material.turnover             -8.2891     3.8323  -2.163 0.030545 *
profit_by_income                  -0.5213     0.1254  -4.156 3.24e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1771.0  on 3540  degrees of freedom
Residual deviance: 1029.6  on 3529  degrees of freedom
AIC: 1053.6

Number of Fisher Scoring iterations: 9
```

Logistics Regression

Model 3

A Model 3 is build with set of new predictors that are significant from previous model and finally most important and significant variables are found from this model 3.

```
Call:
glm(formula = defualt ~ PBT + TOL.TNW + Current.ratio..times. +
    Debt.to.equity.ratio..times. + Cash.to.current.liabilities..times. +
    Raw.material.turnover, family = "binomial", data = p.data.final_rem_1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7010  -0.3029  -0.2223  -0.1201   3.7912

Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                        -4.4928     0.3865 -11.624  < 2e-16 ***
PBT                               -17.1095     3.9806  -4.298 1.72e-05 ***
TOL.TNW                             3.8878     0.6051   6.425 1.31e-10 ***
Current.ratio..times.              -5.3251     1.1263  -4.728 2.27e-06 ***
Debt.to.equity.ratio..times.        3.2987     0.7126   4.629 3.67e-06 ***
Cash.to.current.liabilities..times.  4.4376     0.9761   4.546 5.46e-06 ***
Raw.material.turnover              -14.3083     3.7053  -3.862 0.000113 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1771.0  on 3540  degrees of freedom
Residual deviance: 1255.5  on 3534  degrees of freedom
AIC: 1269.5

Number of Fisher Scoring iterations: 8
```

Model Inferences

The Co-efficient and signs explains that

- With every single unit increase in **PBT**, there will a decrease in chance of default by 17.10 and the p value of very close to 0.0001 tells that its highly significant in explaining the company on default.
- With every single unit increase in **TOL.TNW** , there will be an increase in chance of default by 3.88. Also the p-value is very close to zero which determines the default of company.
- Similarly with every single unit increase in '**Current.ratio..times.** 'there will be a decrease in chance of default by 5.32. Also the p-value is very close to zero which determines the default of company.
- Similarly with every single unit increase in '**Debt.to.equity.ratio..times**.' there will be an increase in chance of default by 3.29. Also the p-value is very close to zero which determines the default of company.
- Similarly with every single unit increase in '**Cash.to.current.liabilities..times**.' there will be an increase in chance of default by 4.43. Also the p-value is very close to zero which determines the default of company.
- Similarly with every single unit increase in '**Raw.material.turnover**' there will be a decrease in chance of default by 14.3. Also the p-value is very close to zero which determines the default of company.

Model Accuracy

by

Confusion Matrix

Model performance for Train Data with Accuracy of 93.96%

```
> confusionMatrix(p.data.final_rem_1$defualt,p.data.final_rem_1$pred)
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 3271   27
         1  187   56

               Accuracy : 0.9396
                 95% CI : (0.9312, 0.9472)
    No Information Rate : 0.9766
    P-Value [Acc > NIR] : 1

                  Kappa : 0.3198

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

            Sensitivity : 0.9459
            Specificity : 0.6747
         Pos Pred Value : 0.9918
         Neg Pred Value : 0.2305
             Prevalence : 0.9766
         Detection Rate : 0.9238
   Detection Prevalence : 0.9314
      Balanced Accuracy : 0.8103

       'Positive' Class : 0
```

Model performance for Test Data with Accuracy of 92.25%

Model Accuracy

by

Confusion Matrix

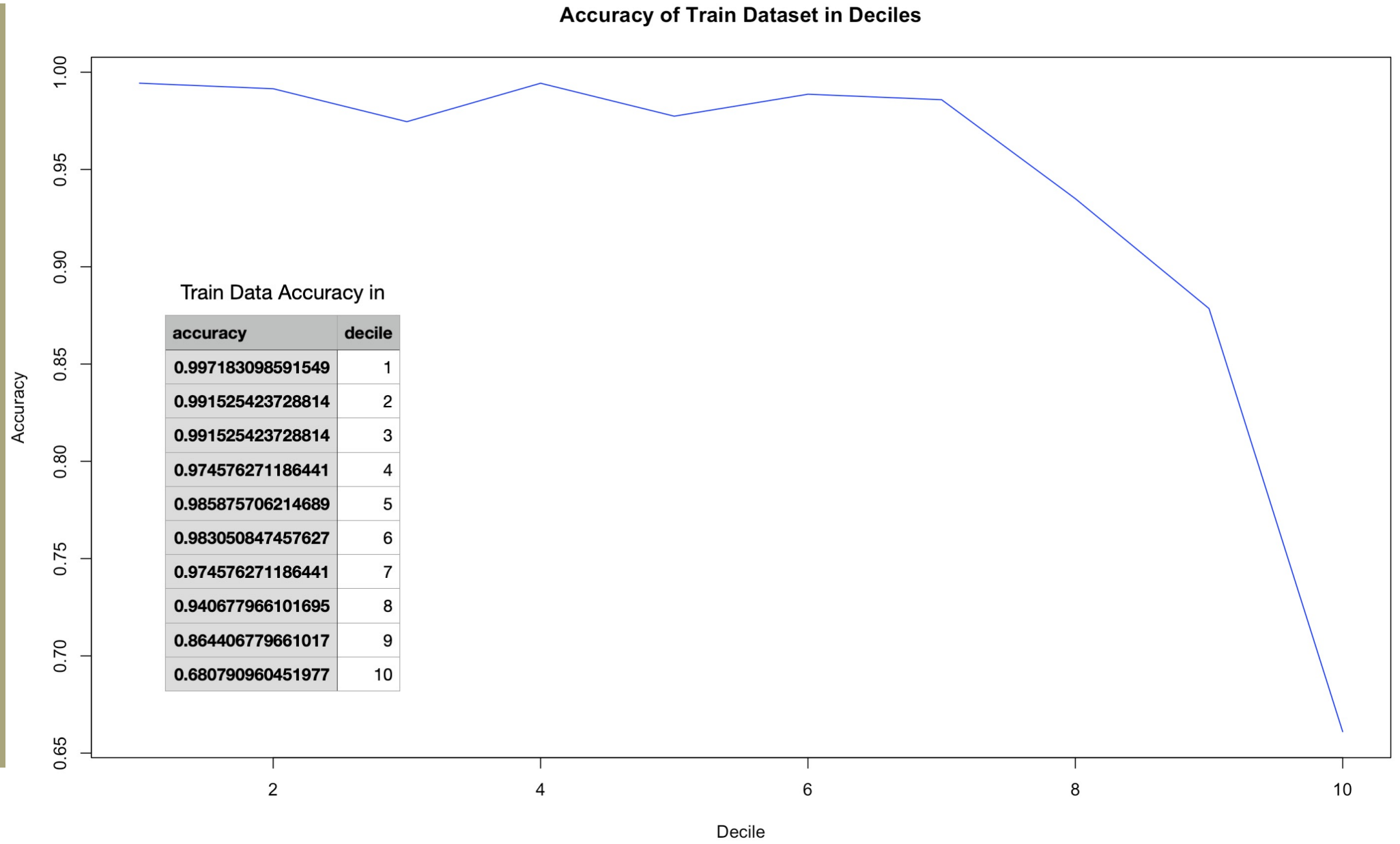
```
> confusionMatrix(coydef_data_test$default, coydef_data_test$pred)
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 546  40
         1   7  37

               Accuracy : 0.9254
                 95% CI : (0.902, 0.9447)
    No Information Rate : 0.8778
    P-Value [Acc > NIR] : 6.916e-05

                  Kappa : 0.5737

 Mcnemar's Test P-Value : 3.046e-06

            Sensitivity : 0.9873
            Specificity : 0.4805
         Pos Pred Value : 0.9317
         Neg Pred Value : 0.8409
             Prevalence : 0.8778
         Detection Rate : 0.8667
   Detection Prevalence : 0.9302
      Balanced Accuracy : 0.7339

       'Positive' Class : 0
```

Data is sorted in descending order based on probability of default and then divide into 10 dociles based on probability

Model Performance

Train Data

Performance of the model is drastically decreased on 8,9 and 10 docile



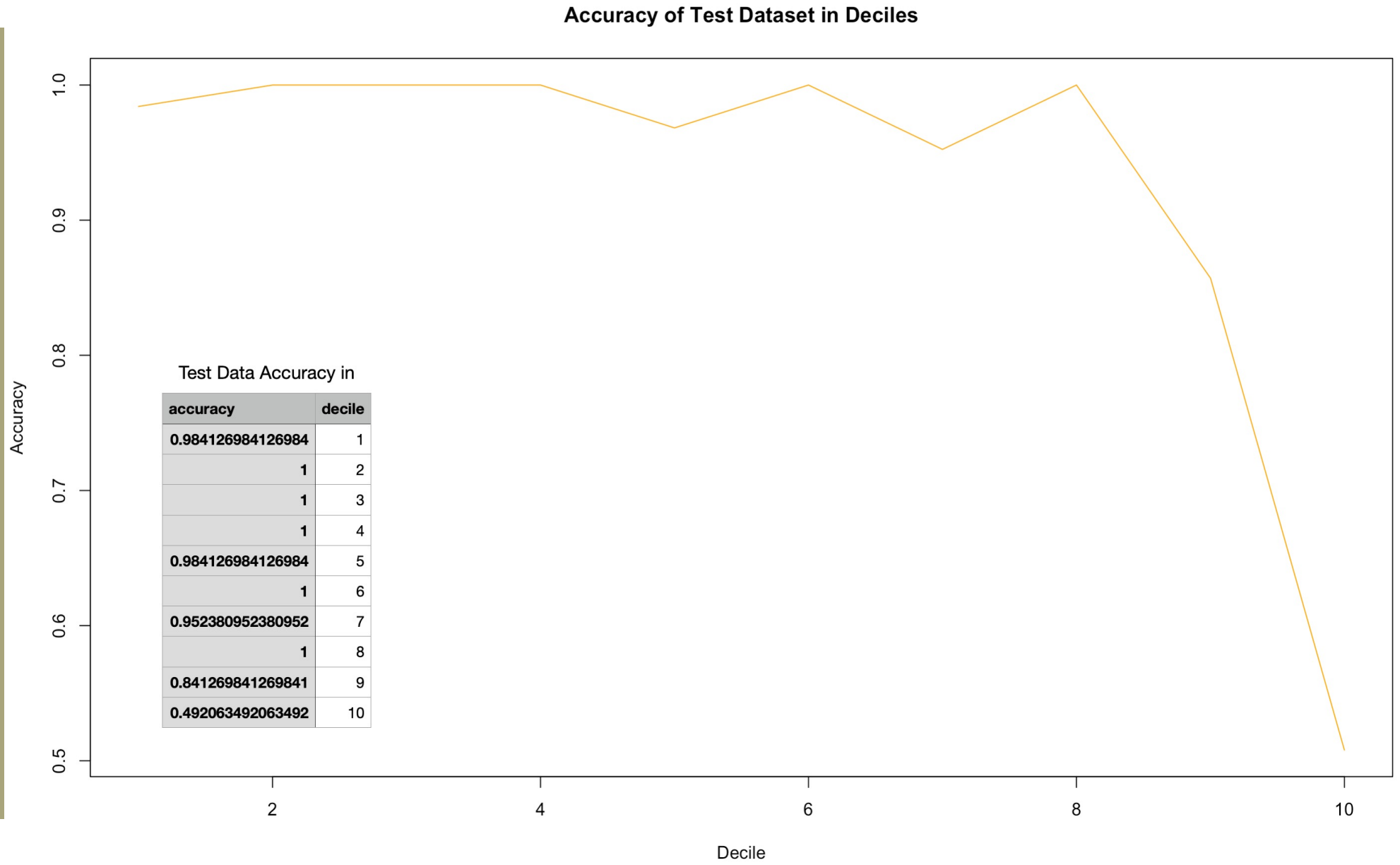
Train Data Accuracy in

accuracy	decile
0.997183098591549	1
0.991525423728814	2
0.991525423728814	3
0.974576271186441	4
0.985875706214689	5
0.983050847457627	6
0.974576271186441	7
0.940677966101695	8
0.864406779661017	9
0.680790960451977	10

Model Performance

Test Data

Performance of the model is drastically decreased on 9 and 10 docile



Test Data Accuracy in

accuracy	decile
0.984126984126984	1
1	2
1	3
1	4
0.984126984126984	5
1	6
0.952380952380952	7
1	8
0.841269841269841	9
0.492063492063492	10

FRA

Thank You