Project 8 – Finance & Risk Analytics

Assignment Report

- By Samrat Mallik

Table of Contents

1.Project Objective

- Outlier Treatment - Outlier Treatment
- Missing Value Treatment
- New Variables Creation (One ratio for profitability, leverage, liquidity and company's size)
- Check for multicollinearity
- Univariate & bivariate analysis
- Build Logistic Regression Model on most important variables
- Analyze coefficient & their signs
- Predict accuracy of model on dev and validation datasets
- Sort the data in descending order based on probability of default and then divide into 10 deciles based on probability & check how well the model has performed

We need to build the model on the raw dataset and check the model performance measures on the validation dataset.

2.Assumptions

- Normally distributed
- Linear relationship
- Multivariate normality
- No or little multicollinearity
- No auto-correlation
- Homoscedasticity

3.Exploratory Data Analysis

3.1.Environment Setup and Data Import

3.1.1.Installing Necessary Packages and Invoking Libraries

```r
library(readxl)
library(mice)

## Warning: package 'mice' was built under R version 3.6.1

## Loading required package: lattice

##
## Attaching package: 'mice'

## The following objects are masked from 'package:base':
##
##     cbind, rbind


library(ggplot2)

## Registered S3 methods overwritten by 'ggplot2':
##   method         from
##   [.quosures     rlang
##   c.quosures     rlang
##   print.quosures rlang


library(ggcorrplot)

## Warning: package 'ggcorrplot' was built under R version 3.6.1


library(ellipse)

## Warning: package 'ellipse' was built under R version 3.6.1

##
## Attaching package: 'ellipse'
## The following object is masked from 'package:graphics':
##
##     pairs
```

```r
library(RColorBrewer)
library(nFactors)
```

```
## Warning: package 'nFactors' was built under R version 3.6.1

## Loading required package: MASS

## Loading required package: psych

##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha

## Loading required package: boot

##
## Attaching package: 'boot'

## The following object is masked from 'package:psych':
##
##     logit

## The following object is masked from 'package:lattice':
##
##     melanoma

##
## Attaching package: 'nFactors'

## The following object is masked from 'package:lattice':
##
##     parallel
```

```r
library(psych)
library(lattice)
library(caTools)
```

```
## Warning: package 'caTools' was built under R version 3.6.1
```

```r
library(rpart)
library(rpart.plot)
```

```
## Warning: package 'rpart.plot' was built under R version 3.6.1
```

```r
library(rattle)
```

```
## Warning: package 'rattle' was built under R version 3.6.1

## Rattle: A free graphical interface for data science with R.
## Version 5.2.0 Copyright (c) 2006-2018 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.

library(data.table)

## Warning: package 'data.table' was built under R version 3.6.1

library(ROCR)

## Warning: package 'ROCR' was built under R version 3.6.1

## Loading required package: gplots

## Warning: package 'gplots' was built under R version 3.6.1

##
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':
##
##     lowess

library(ineq)
library(StatMeasures)

## Warning: package 'StatMeasures' was built under R version 3.6.1

library(htmlwidgets)
library(DataExplorer)

## Warning: package 'DataExplorer' was built under R version 3.6.1

library(corrplot)

## Warning: package 'corrplot' was built under R version 3.6.1

## corrplot 0.84 loaded

library(partykit)

## Warning: package 'partykit' was built under R version 3.6.1

## Loading required package: grid

## Loading required package: libcoin

## Warning: package 'libcoin' was built under R version 3.6.1

## Loading required package: mvtnorm

library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:data.table':
##
##     between, first, last
```

```
## The following object is masked from 'package:MASS':
##
##     select
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(purrr)
```

```
##
## Attaching package: 'purrr'
```

```
## The following object is masked from 'package:data.table':
##
##     transpose
```

```r
library(InformationValue)
```

```
## Warning: package 'InformationValue' was built under R version 3.6.1
```

```r
library(car)
```

```
## Warning: package 'car' was built under R version 3.6.1
```

```
## Loading required package: carData
```

```
## Registered S3 methods overwritten by 'car':
##   method                          from
##   influence.merMod                lme4
##   cooks.distance.influence.merMod lme4
##   dfbeta.influence.merMod         lme4
##   dfbetas.influence.merMod        lme4
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:purrr':
##
##     some
## The following object is masked from 'package:dplyr':
##
##     recode

## The following object is masked from 'package:boot':
##
##     logit

## The following object is masked from 'package:psych':
##
##     logit

## The following object is masked from 'package:ellipse':
##
##     ellipse

library(ROCR)
library(MASS)
library(e1071)

## Warning: package 'e1071' was built under R version 3.6.1

library(class)

## Warning: package 'class' was built under R version 3.6.1

library(caret)

## Warning: package 'caret' was built under R version 3.6.1

##
## Attaching package: 'caret'

## The following objects are masked from 'package:InformationValue':
##
##     confusionMatrix, precision, sensitivity, specificity

## The following object is masked from 'package:purrr':
##
##     lift

library(DMwR)

## Warning: package 'DMwR' was built under R version 3.6.1

## Registered S3 method overwritten by 'xts':
##   method     from
##   as.zoo.xts zoo

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

library(ipred)
```

```
## Warning: package 'ipred' was built under R version 3.6.1
```

3.1.2. Setting Up Working Directory and Importing the data

```
setwd("C:/Users/Samrat/Documents/R/Directories/")
getwd()
```

```
## [1] "C:/Users/Samrat/Documents/R/Directories"
```

```
data = read_excel("raw-data.xlsx")
```

3.2. Variable Identification

```
summary(data)
```

```
##       Num       Networth Next Year  Total assets         Net worth
##  Min.   :   1   Min.   :-74265.6   Min.   :      0.1   Min.   :     0.0
##  1st Qu.: 886   1st Qu.:    31.7   1st Qu.:     91.3   1st Qu.:    31.3
##  Median :1773   Median :   116.3   Median :    309.7   Median :   102.3
##  Mean   :1772   Mean   :  1616.3   Mean   :   3443.4   Mean   :  1295.9
##  3rd Qu.:2658   3rd Qu.:   456.1   3rd Qu.:   1098.7   3rd Qu.:   377.3
##  Max.   :3545   Max.   :805773.4   Max.   :1176509.2   Max.   :613151.6
##
##   Total income       Change in stock    Total expenses
##  Min.   :      0.0   Min.   :-3029.40   Min.   :     -0.1
##  1st Qu.:    106.5   1st Qu.:   -1.80   1st Qu.:     95.8
##  Median :    444.9   Median :    1.60   Median :    407.7
##  Mean   :   4582.8   Mean   :   41.49   Mean   :   4262.9
##  3rd Qu.:   1440.9   3rd Qu.:   18.05   3rd Qu.:   1359.8
##  Max.   :2442828.2   Max.   :14185.50   Max.   :2366035.3
##  NA's   :198         NA's   :458        NA's   :139
##  Profit after tax        PBDITA              PBT
##  Min.   : -3908.30   Min.   : -440.7   Min.   : -3894.80
##  1st Qu.:     0.50   1st Qu.:    6.9   1st Qu.:     0.70
##  Median :     8.80   Median :   35.4   Median :    12.40
##  Mean   :   277.36   Mean   :  578.1   Mean   :   383.81
##  3rd Qu.:    52.27   3rd Qu.:  150.2   3rd Qu.:    71.97
##  Max.   :119439.10   Max.   :208576.5   Max.   :145292.60
##  NA's   :131         NA's   :131        NA's   :131
##   Cash profit        PBDITA as % of total income PBT as % of total income
##  Min.   : -2245.70   Min.   :-6400.000           Min.   :-21340.00
##  1st Qu.:     2.90   1st Qu.:    5.000           1st Qu.:     0.55
##  Median :    18.85   Median :    9.660           Median :     3.31
##  Mean   :   392.07   Mean   :    4.571           Mean   :   -17.28
##  3rd Qu.:    93.20   3rd Qu.:   16.390           3rd Qu.:     8.80
##  Max.   :176911.80   Max.   :  100.000           Max.   :   100.00
##  NA's   :131         NA's   :68                  NA's   :68
##  PAT as % of total income Cash profit as % of total income
##  Min.   :-21340.00        Min.   :-15020.000
##  1st Qu.:     0.35        1st Qu.:     2.020
##  Median :     2.34        Median :     5.640
##  Mean   :   -19.20        Mean   :    -8.229
##  3rd Qu.:     6.34        3rd Qu.:    10.700
##  Max.   :   150.00        Max.   :   100.000
##  NA's   :68               NA's   :68
##  PAT as % of net worth     Sales           Income from financial services
##  Min.   :-748.72       Min.   :      0.1   Min.   :    0.00
##  1st Qu.:   0.00       1st Qu.:    112.7   1st Qu.:    0.40
##  Median :   7.92       Median :    453.1   Median :    1.80
##  Mean   :  10.27       Mean   :   4549.5   Mean   :   80.84
##  3rd Qu.:  20.19       3rd Qu.:   1433.5   3rd Qu.:    9.68
##  Max.   :2466.67       Max.   :2384984.4   Max.   :51938.20
##                        NA's   :259         NA's   :935
##   Other income      Total capital     Reserves and funds
##  Min.   :    0.00   Min.   :    0.1   Min.   : -6525.9
##  1st Qu.:    0.40   1st Qu.:   13.1   1st Qu.:     5.0
##  Median :    1.40   Median :   42.1   Median :    54.8
##  Mean   :   41.36   Mean   :  216.6   Mean   :  1163.8
##  3rd Qu.:    5.97   3rd Qu.:  100.3   3rd Qu.:   277.3
##  Max.   :42856.70   Max.   :78273.2   Max.   :625137.8
##  NA's   :1295       NA's   :4         NA's   :85
##  Deposits (accepted by commercial banks)   Borrowings
##  Mode:logical                            Min.   :     0.10
##  NA's:3541                               1st Qu.:    23.95
##                                          Median :    99.20
##                                          Mean   :  1122.28
##                                          3rd Qu.:   352.60
##                                          Max.   :278257.30
##                                          NA's   :366
##  Current liabilities & provisions Deferred tax liability
##  Min.   :     0.1                 Min.   :    0.1
##  1st Qu.:    17.8                 1st Qu.:    3.2
##  Median :    69.4                 Median :   13.4
##  Mean   :   940.6                 Mean   :  227.2
##  3rd Qu.:   261.7                 3rd Qu.:   50.0
##  Max.   :352240.3                 Max.   :72796.6
##  NA's   :96                       NA's   :1140
##  Shareholders funds Cumulative retained profits Capital employed
##  Min.   :     0.0   Min.   : -6534.3            Min.   :     0.0
##  1st Qu.:    32.0   1st Qu.:     1.1            1st Qu.:    60.8
##  Median :   105.6   Median :    37.1            Median :   214.7
##  Mean   :  1322.1   Mean   :   890.5            Mean   :  2328.3
##  3rd Qu.:   393.2   3rd Qu.:   202.3            3rd Qu.:   767.3
##  Max.   :613151.6   Max.   :390133.8            Max.   :891408.9
##                      NA's   :38
##     TOL/TNW        Total term liabilities / tangible net worth
##  Min.   :-350.480   Min.   :-325.600
##  1st Qu.:   0.600   1st Qu.:   0.050
##  Median :   1.430   Median :   0.340
##  Mean   :   3.994   Mean   :   1.844
##  3rd Qu.:   2.830   3rd Qu.:   1.000
##  Max.   : 473.000   Max.   : 456.000
##
##  Contingent liabilities / Net worth (%) Contingent liabilities
##  Min.   :    0.00                       Min.   :     0.1
##  1st Qu.:    0.00                       1st Qu.:     6.3
##  Median :    5.33                       Median :    38.0
##  Mean   :   53.94                       Mean   :   932.9
##  3rd Qu.:   30.76                       3rd Qu.:   192.7
##  Max.   :14704.27                       Max.   :559506.8
##                                         NA's   :1188
##  Net fixed assets    Investments        Current assets
##  Min.   :     0.0   Min.   :     0.00   Min.   :     0.1
##  1st Qu.:    26.0   1st Qu.:     1.00   1st Qu.:    36.2
##  Median :    93.5   Median :     8.35   Median :   145.1
##  Mean   :  1189.7   Mean   :   694.73   Mean   :  1293.4
##  3rd Qu.:   344.9   3rd Qu.:    64.30   3rd Qu.:   502.2
##  Max.   :636604.6   Max.   :199978.60   Max.   :354815.2
##  NA's   :118        NA's   :1435        NA's   :66
##  Net working capital Quick ratio (times) Current ratio (times)
##  Min.   :-63839.0    Min.   :  0.000     Min.   :  0.00
##  1st Qu.:    -1.1    1st Qu.:  0.410     1st Qu.:  0.93
##  Median :    16.2    Median :  0.670     Median :  1.23
##  Mean   :   138.6    Mean   :  1.401     Mean   :  2.13
##  3rd Qu.:    84.2    3rd Qu.:  1.030     3rd Qu.:  1.71
##  Max.   : 85782.8    Max.   :341.000     Max.   :505.00
##  NA's   :32          NA's   :93          NA's   :93
##  Debt to equity ratio (times) Cash to current liabilities (times)
##  Min.   :  0.00               Min.   :  0.0000
##  1st Qu.:  0.22               1st Qu.:  0.0200
##  Median :  0.79               Median :  0.0700
##  Mean   :  2.78               Mean   :  0.4904
##  3rd Qu.:  1.75               3rd Qu.:  0.1900
##  Max.   :456.00               Max.   :165.0000
##                               NA's   :93
##  Cash to average cost of sales per day Creditors turnover
##  Min.   :     0.00                     Length:3541
##  1st Qu.:     2.79                     Class :character
##  Median :     8.03                     Mode  :character
##  Mean   :   158.44
##  3rd Qu.:    21.79
##  Max.   :128040.76
##  NA's   :85
##  Debtors turnover   Finished goods turnover WIP turnover
##  Length:3541        Length:3541             Length:3541
##  Class :character   Class :character        Class :character
##  Mode  :character   Mode  :character        Mode  :character
##
##
##
##
##  Raw material turnover Shares outstanding Equity face value
##  Length:3541           Length:3541        Length:3541
##  Class :character      Class :character   Class :character
##  Mode  :character      Mode  :character   Mode  :character
##
##
##
##
##       EPS              Adjusted EPS       Total liabilities
##  Min.   :-843181.8   Min.   :-843181.8   Min.   :      0.1
##  1st Qu.:      0.0   1st Qu.:      0.0   1st Qu.:     91.3
##  Median :      1.4   Median :      1.2   Median :    309.7
##  Mean   :   -220.3   Mean   :   -221.5   Mean   :   3443.4
##  3rd Qu.:      9.6   3rd Qu.:      7.5   3rd Qu.:   1098.7
##  Max.   :  34522.5   Max.   :  34522.5   Max.   :1176509.2
##
##   PE on BSE
##  Length:3541
##  Class :character
##  Mode  :character
##
##
##
##
```

```
str(data)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    3541 obs. of  52 variables:
##  $ Num                                        : num  1 2 3 4 5 6 7 8 9 10
...
##  $ Networth Next Year                         : num  8890.6 394.3 92.2 2.7
109 ...
##  $ Total assets                               : num  17512.3 941 232.8 2.7
478.5 ...
##  $ Net worth                                  : num  7093.2 351.5 100.6 2.
7 107.6 ...
##  $ Total income                               : num  24965 1527 477 NA 158
0 ...
##  $ Change in stock                            : num  235.8 42.7 -5.2 NA -1
7 ...
##  $ Total expenses                             : num  23658 1455 479 NA 155
8 ...
##  $ Profit after tax                           : num  1543.2 115.2 -6.6 NA
```

```
5.5 ...
##  $ PBDITA                                     : num  2860.2 283 5.8 NA 31
...
##  $ PBT                                        : num  2417.2 188.4 -6.6 NA
6.3 ...
##  $ Cash profit                                : num  1872.8 158.6 0.3 NA 1
1.9 ...
##  $ PBDITA as % of total income                : num  11.46 18.53 1.22 0 1.
96 ...
##  $ PBT as % of total income                   : num  9.68 12.33 -1.38 0 0.
4 ...
##  $ PAT as % of total income                   : num  6.18 7.54 -1.38 0 0.3
5 2.81 0 0.72 8.29 -2.88 ...
##  $ Cash profit as % of total income           : num  7.5 10.38 0.06 0 0.75
...
##  $ PAT as % of net worth                      : num  23.78 38.08 -6.35 0 5
.25 ...
##  $ Sales                                      : num  24458 1504 476 NA 157
5 ...
##  $ Income from financial services             : num  158 4 1.5 NA 3.9 6.4
NA NA 7.3 NA ...
##  $ Other income                               : num  297.2 15.9 0.2 NA 0.9
...
##  $ Total capital                              : num  423.8 115.5 81.4 0.5
6.2 ...
##  $ Reserves and funds                         : num  6822.8 257.8 19.2 2.2
161.8 ...
##  $ Deposits (accepted by commercial banks)    : logi  NA NA NA NA NA NA ..
.
##  $ Borrowings                                 : num  14.9 272.5 35.4 NA 19
3.1 ...
##  $ Current liabilities & provisions           : num  9965.9 210 96.8 NA 11
2.8 ...
##  $ Deferred tax liability                     : num  284.9 85.2 NA NA 4.6
...
##  $ Shareholders funds                         : num  7093.2 351.5 100.6 2.
7 107.6 ...
##  $ Cumulative retained profits                : num  6263.3 247.4 32.4 2.2
82.7 ...
##  $ Capital employed                           : num  7108.1 624 136 2.7 30
0.7 ...
##  $ TOL/TNW                                    : num  1.33 1.23 1.44 0 2.83
1.8 0.03 5.17 1.05 3.25 ...
##  $ Total term liabilities / tangible net worth: num  0 0.34 0.29 0 1.59 0.
37 0.03 0.94 0.3 0.54 ...
##  $ Contingent liabilities / Net worth (%)     : num  14.8 19.2 45.8 0 34.9
...
##  $ Contingent liabilities                     : num  1049.7 67.6 46.1 NA 3
7.6 ...
##  $ Net fixed assets                           : num  1900.2 286.4 38.7 2.5
```

```
94.8 ...
##  $ Investments                                : num  1069.6 2.2 4.3 NA 7.4
...
##  $ Current assets                             : num  13277.5 563.9 167.5 0
.2 349.7 ...
##  $ Net working capital                        : num  3588.5 203.5 59.6 0.2
215.8 ...
##  $ Quick ratio (times)                        : num  1.18 0.95 1.11 NA 1.4
1 0.48 NA 0.54 0.59 0.39 ...
##  $ Current ratio (times)                      : num  1.37 1.56 1.55 NA 2.5
4 1.27 NA 1.15 1.58 0.5 ...
##  $ Debt to equity ratio (times)               : num  0 0.78 0.35 0 1.79 1.
09 0.32 2.31 0.94 3.13 ...
##  $ Cash to current liabilities (times)        : num  0.43 0.06 0.21 NA 0 0
.11 NA 0.04 0.19 0 ...
##  $ Cash to average cost of sales per day      : num  68.21 5.96 17.07 NA 0
...
##  $ Creditors turnover                         : chr  "3.62" "9.80000000000
00007" "5.28" "0" ...
##  $ Debtors turnover                           : chr  "3.85" "5.7" "5.07" "
0" ...
##  $ Finished goods turnover                    : chr  "200.55" "14.21" "9.2
4" NA ...
##  $ WIP turnover                               : chr  "21.78" "7.49" "0.23"
NA ...
##  $ Raw material turnover                      : chr  "7.71" "11.46" NA "0"
...
##  $ Shares outstanding                         : chr  "42381675" "11550000"
"8149090" "52404" ...
##  $ Equity face value                          : chr  "10" "10" "10" "10" .
..
##  $ EPS                                        : num  35.52 9.97 -0.5 0 7.9
1 ...
##  $ Adjusted EPS                               : num  7.1 9.97 -0.5 0 7.91
...
##  $ Total liabilities                          : num  17512.3 941 232.8 2.7
478.5 ...
##  $ PE on BSE                                  : chr  "27.31" "8.17" "-5.76
" "NA" ...
```

```
dim(data)
```

```
## [1] 3541   52
```

```
names(data)
```

```
##  [1] "Num"
##  [2] "Networth Next Year"
##  [3] "Total assets"
##  [4] "Net worth"
##  [5] "Total income"
##  [6] "Change in stock"
##  [7] "Total expenses"
##  [8] "Profit after tax"
##  [9] "PBDITA"
## [10] "PBT"
## [11] "Cash profit"
## [12] "PBDITA as % of total income"
## [13] "PBT as % of total income"
## [14] "PAT as % of total income"
## [15] "Cash profit as % of total income"
## [16] "PAT as % of net worth"
## [17] "Sales"
## [18] "Income from financial services"
## [19] "Other income"
## [20] "Total capital"
## [21] "Reserves and funds"
## [22] "Deposits (accepted by commercial banks)"
## [23] "Borrowings"
## [24] "Current liabilities & provisions"
## [25] "Deferred tax liability"
## [26] "Shareholders funds"
## [27] "Cumulative retained profits"
## [28] "Capital employed"
## [29] "TOL/TNW"
## [30] "Total term liabilities / tangible net worth"
## [31] "Contingent liabilities / Net worth (%)"
## [32] "Contingent liabilities"
## [33] "Net fixed assets"
## [34] "Investments"
## [35] "Current assets"
## [36] "Net working capital"
## [37] "Quick ratio (times)"
## [38] "Current ratio (times)"
## [39] "Debt to equity ratio (times)"
## [40] "Cash to current liabilities (times)"
## [41] "Cash to average cost of sales per day"
## [42] "Creditors turnover"
## [43] "Debtors turnover"
## [44] "Finished goods turnover"
## [45] "WIP turnover"
## [46] "Raw material turnover"
## [47] "Shares outstanding"
## [48] "Equity face value"
## [49] "EPS"
## [50] "Adjusted EPS"
## [51] "Total liabilities"
## [52] "PE on BSE"
```

```
colnames(data) = make.names(colnames(data))
```

```
attach(data)

test.data = read_excel("validation_data.xlsx")

summary(test.data)
```

```
##       Num          Default - 1       Total assets        Net worth
##  Min.   :  1.0   Min.   :0.00000   Min.   :     0.1   Min.   :     0.1
##  1st Qu.:179.5   1st Qu.:0.00000   1st Qu.:    93.2   1st Qu.:    34.4
##  Median :358.0   Median :0.00000   Median :   347.7   Median :   120.9
##  Mean   :358.0   Mean   :0.07552   Mean   :  4218.6   Mean   :  1629.7
##  3rd Qu.:536.5   3rd Qu.:0.00000   3rd Qu.:  1315.3   3rd Qu.:   451.5
##  Max.   :715.0   Max.   :1.00000   Max.   :354727.3   Max.   :171840.0
##
##   Total income       Change in stock   Total expenses
##  Min.   :      0.0   Min.   :-488.10   Min.   :      0.0
##  1st Qu.:    110.8   1st Qu.:  -1.90   1st Qu.:    104.1
##  Median :    536.0   Median :   1.80   Median :    511.1
##  Mean   :   5204.7   Mean   :  54.66   Mean   :   4817.3
##  3rd Qu.:   1727.1   3rd Qu.:  19.35   3rd Qu.:   1642.3
##  Max.   :1028087.4   Max.   :7540.00   Max.   :1014813.1
##  NA's   :33          NA's   :92        NA's   :26
##  Profit after tax       PBDITA             PBT
##  Min.   : -998.00   Min.   :  -393.90   Min.   : -993.90
##  1st Qu.:    0.68   1st Qu.:     7.15   1st Qu.:    1.00
##  Median :   10.20   Median :    42.20   Median :   14.25
##  Mean   :  382.22   Mean   :   743.35   Mean   :  540.59
##  3rd Qu.:   68.95   3rd Qu.:   192.82   3rd Qu.:   90.50
##  Max.   :62022.90   Max.   :110557.10   Max.   :94565.20
##  NA's   :23         NA's   :23          NA's   :23
##   Cash profit       PBDITA as % of total income PBT as % of total income
##  Min.   : -894.60   Min.   :-6400.000           Min.   :-9700.000
##  1st Qu.:    3.27   1st Qu.:    4.702           1st Qu.:    0.622
##  Median :   22.05   Median :    9.780           Median :    3.450
##  Mean   :  488.11   Mean   :   -3.681           Mean   :  -22.725
##  3rd Qu.:  120.30   3rd Qu.:   16.753           3rd Qu.:    9.725
##  Max.   :71581.60   Max.   :  100.000           Max.   :  100.000
##  NA's   :23         NA's   :11                  NA's   :11
##  PAT as % of total income Cash profit as % of total income
##  Min.   :-9700.000        Min.   :-6400.000
##  1st Qu.:    0.390        1st Qu.:    1.930
##  Median :    2.405        Median :    5.835
##  Mean   :  -24.147        Mean   :  -12.929
##  3rd Qu.:    6.790        3rd Qu.:   10.982
##  Max.   :  100.000        Max.   :  100.000
##  NA's   :11               NA's   :11
##  PAT as % of net worth     Sales          Income from financial services
##  Min.   :-194.520      Min.   :     0.1   Min.   :   0.10
##  1st Qu.:   0.000      1st Qu.:   120.8   1st Qu.:   0.50
##  Median :   8.710      Median :   552.5   Median :   2.00
##  Mean   :   9.666      Mean   :  5117.5   Mean   :  83.86
##  3rd Qu.:  20.215      3rd Qu.:  1721.3   3rd Qu.:  10.10
##  Max.   : 441.670      Max.   :976884.0   Max.   :8097.20
##                        NA's   :46         NA's   :176
##   Other income      Total capital     Reserves and funds
##  Min.   :    0.00   Min.   :    0.1   Min.   : -1125.00
##  1st Qu.:    0.32   1st Qu.:   14.1   1st Qu.:     7.33
##  Median :    1.65   Median :   45.3   Median :    57.45
##  Mean   :  128.16   Mean   :  263.9   Mean   :  1440.70
##  3rd Qu.:    7.25   3rd Qu.:  121.1   3rd Qu.:   334.80
##  Max.   :42856.70   Max.   :41304.0   Max.   :133684.20
##  NA's   :261        NA's   :1         NA's   :13
##  Deposits (accepted by commercial banks)   Borrowings
##  Mode:logical                            Min.   :     0.20
##  NA's:715                                1st Qu.:    25.93
##                                          Median :   105.50
##                                          Mean   :  1439.86
##                                          3rd Qu.:   391.82
##                                          Max.   :105175.30
##                                          NA's   :65
##  Current liabilities & provisions Deferred tax liability
##  Min.   :     0.1                 Min.   :    0.10
##  1st Qu.:    16.8                 1st Qu.:    3.10
##  Median :    75.2                 Median :   14.70
##  Mean   :  1058.9                 Mean   :  270.45
##  3rd Qu.:   300.4                 3rd Qu.:   62.42
##  Max.   :112712.7                 Max.   :27077.90
##  NA's   :14                       NA's   :229
##  Shareholders funds Cumulative retained profits Capital employed
##  Min.   :     0.1   Min.   : -2582.4            Min.   :     0.10
##  1st Qu.:    35.5   1st Qu.:     0.8            1st Qu.:    64.35
##  Median :   124.0   Median :    40.6            Median :   246.10
##  Mean   :  1646.0   Mean   :  1168.1            Mean   :  2954.96
##  3rd Qu.:   478.4   3rd Qu.:   244.5            3rd Qu.:   913.65
##  Max.   :171840.0   Max.   :128183.1            Max.   :235389.50
##                     NA's   :7
##     TOL/TNW         Total term liabilities / tangible net worth
##  Min.   :-350.480   Min.   :-325.600
##  1st Qu.:   0.595   1st Qu.:   0.060
##  Median :   1.400   Median :   0.350
##  Mean   :   4.181   Mean   :   1.906
##  3rd Qu.:   2.800   3rd Qu.:   1.005
##  Max.   : 411.270   Max.   : 292.020
##
##  Contingent liabilities / Net worth (%) Contingent liabilities
##  Min.   :  0.00                         Min.   :    0.1
##  1st Qu.:  0.00                         1st Qu.:    5.1
##  Median :  5.52                         Median :   37.5
##  Mean   : 64.47                         Mean   : 1022.0
##  3rd Qu.: 31.49                         3rd Qu.:  217.1
##  Max.   :6295.24                        Max.   :72620.8
##                                         NA's   :214
##  Net fixed assets    Investments      Current assets
##  Min.   :     0.1   Min.   :    0.0   Min.   :     0.1
##  1st Qu.:    27.2   1st Qu.:    0.9   1st Qu.:    38.9
##  Median :    95.0   Median :    7.8   Median :   165.6
##  Mean   :  1306.2   Mean   :  853.2   Mean   :  1632.9
##  3rd Qu.:   409.2   3rd Qu.:   61.6   3rd Qu.:   578.0
##  Max.   :115737.5   Max.   :88047.8   Max.   :196614.6
##  NA's   :14         NA's   :280       NA's   :14
##  Net working capital Quick ratio (times) Current ratio (times)
##  Min.   :-41908.3    Min.   :  0.000     Min.   :  0.000
##  1st Qu.:    -1.3    1st Qu.:  0.410     1st Qu.:  0.920
##  Median :    20.1    Median :  0.660     Median :  1.230
##  Mean   :   283.0    Mean   :  1.968     Mean   :  2.880
##  3rd Qu.:    99.2    3rd Qu.:  1.020     3rd Qu.:  1.725
##  Max.   : 85782.8    Max.   :341.000     Max.   :505.000
##  NA's   :5           NA's   :12          NA's   :12
##  Debt to equity ratio (times) Cash to current liabilities (times)
##  Min.   :  0.000              Min.   :  0.0000
##  1st Qu.:  0.220              1st Qu.:  0.0300
##  Median :  0.800              Median :  0.0800
##  Mean   :  3.327              Mean   :  0.7149
##  3rd Qu.:  1.700              3rd Qu.:  0.1900
##  Max.   :341.180              Max.   :165.0000
##                               NA's   :12
##  Cash to average cost of sales per day Creditors turnover
##  Min.   :    0.000                     Length:715
##  1st Qu.:    3.248                     Class :character
##  Median :    8.130                     Mode  :character
##  Mean   :   79.565
##  3rd Qu.:   22.645
##  Max.   :15999.170
##  NA's   :15
##  Debtors turnover   Finished goods turnover WIP turnover
##  Length:715         Length:715              Length:715
##  Class :character   Class :character        Class :character
##  Mode  :character   Mode  :character        Mode  :character
##
##
##
##
##  Raw material turnover Shares outstanding Equity face value
##  Length:715            Length:715         Length:715
##  Class :character      Class :character   Class :character
##  Mode  :character      Mode  :character   Mode  :character
##
##
##
##
```

```
##       EPS              Adjusted EPS       Total liabilities
##  Min.   :-72750.00   Min.   :-72750.00   Min.   :     0.1
##  1st Qu.:     0.00   1st Qu.:     0.00   1st Qu.:    93.2
##  Median :     1.83   Median :     1.50   Median :   347.7
##  Mean   :   -76.87   Mean   :   -78.74   Mean   :  4218.6
##  3rd Qu.:    11.46   3rd Qu.:     8.35   3rd Qu.:  1315.3
##  Max.   :  8784.00   Max.   :  8784.00   Max.   :354727.3
##
##   PE on BSE
##  Length:715
##  Class :character
##  Mode  :character
##
##
##
##
```

```
str(test.data)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    715 obs. of  52 variables:
##  $ Num                                        : num  1 2 3 4 5 6 7 8 9 10
...
##  $ Default - 1                                : num  0 0 1 0 0 0 0 0 0 0 .
..
##  $ Total assets                               : num  971 675 532 858 823 .
..
##  $ Net worth                                  : num  276 212 120 201 349 .
..
##  $ Total income                               : num  2185 819 564 3576 103
4 ...
##  $ Change in stock                            : num  14.2 10.4 -28.1 -0.6
28.9 -0.5 NA -7.7 27.2 -0.2 ...
##  $ Total expenses                             : num  2099 810 578 3613 104
2 ...
##  $ Profit after tax                           : num  100.2 19.7 -42.4 -37.
5 21.4 ...
##  $ PBDITA                                     : num  285.6 116 -31 68.2 90
.1 ...
##  $ PBT                                        : num  152.1 33.7 -56 25.7 2
9.7 ...
##  $ Cash profit                                : num  182.3 50.5 -35.3 37.3
62.7 ...
##  $ PBDITA as % of total income                : num  13.07 14.16 -5.5 1.91
8.71 ...
##  $ PBT as % of total income                   : num  6.96 4.11 -9.94 0.72
2.87 ...
##  $ PAT as % of total income                   : num  4.59 2.4 -7.52 -1.05
2.07 ...
##  $ Cash profit as % of total income           : num  8.34 6.16 -6.26 1.04
6.06 ...
```

```
##  $ PAT as % of net worth                      : num  42.11 10.66 -31.2 0 6
.31 ...
##  $ Sales                                      : num  2171 817 552 3573 102
7 ...
##  $ Income from financial services             : num  2.3 0.8 9.1 1 0.7 ...
##  $ Other income                               : num  NA 0.2 2.1 1.5 2.3 0.
1 NA NA 0.1 0.1 ...
##  $ Total capital                              : num  48 114 47.1 50.5 33 .
..
##  $ Reserves and funds                         : num  413.1 97.6 227.4 150.
9 316.2 ...
##  $ Deposits (accepted by commercial banks)    : logi  NA NA NA NA NA NA ..
.
##  $ Borrowings                                 : num  177.3 339.8 17.5 524.
2 162.3 ...
##  $ Current liabilities & provisions           : num  328.5 100.5 240.1 75.
2 299.6 ...
##  $ Deferred tax liability                     : num  3.7 23.1 NA 56.7 12.2
2.1 1.9 4.4 2.9 NA ...
##  $ Shareholders funds                         : num  276 212 120 201 349 .
..
##  $ Cumulative retained profits                : num  227.8 97.6 69.9 150.9
316.2 ...
##  $ Capital employed                           : num  453 551 138 726 512 .
..
##  $ TOL/TNW                                    : num  1.8 2.01 1.73 2.94 1.
02 0.86 0.06 1.92 0.37 1.96 ...
##  $ Total term liabilities / tangible net worth: num  0.27 0.72 0.09 0.81 0
.1 0.11 0.05 0.78 0 1.81 ...
##  $ Contingent liabilities / Net worth (%)     : num  112.94 5.77 102.83 0.
65 28.78 ...
##  $ Contingent liabilities                     : num  311.5 12.2 123.6 1.3 
100.5 ...
##  $ Net fixed assets                           : num  332 199 270 263 191 .
..
##  $ Investments                                : num  NA NA 0.7 NA NA NA 17
.3 2.6 NA NA ...
##  $ Current assets                             : num  560 407 148 536 472 .
..
##  $ Net working capital                        : num  134.2 123.6 -97.1 99.
6 75.3 ...
##  $ Quick ratio (times)                        : num  0.92 0.48 0.32 0.51 0
.58 0.97 166 0.52 0.88 0.6 ...
##  $ Current ratio (times)                      : num  1.31 1.39 0.6 1.23 1.
19 1.86 166 1.56 1.19 0.55 ...
##  $ Debt to equity ratio (times)               : num  0.64 1.61 0.15 2.6 0.
46 0.32 0.05 1.24 0 1.81 ...
##  $ Cash to current liabilities (times)        : num  0.09 0.03 0.04 0.08 0
.08 0 165 0.03 0.35 0.23 ...
##  $ Cash to average cost of sales per day      : num  7.56 3.88 4.63 3.71 1
```

```
1.15 ...
##  $ Creditors turnover                          : chr  "5.94" "10.59" "2.35"
"NA" ...
##  $ Debtors turnover                            : chr  "5.74" "6.03" "9.6" "
NA" ...
##  $ Finished goods turnover                     : chr  "25.11" "28.96" "8.23
" "NA" ...
##  $ WIP turnover                                : chr  "20.010000000000002"
"18.649999999999999" "6.6" "NA" ...
##  $ Raw material turnover                       : chr  "17.579999999999998"
"2.67" "3.77" "NA" ...
##  $ Shares outstanding                          : chr  "4800000" "11400000"
"471285" "5050000" ...
##  $ Equity face value                           : chr  "10" "10" "100" "10"
...
##  $ EPS                                         : num  18.6 1.65 -90.39 -7.0
9 5.9 ...
##  $ Adjusted EPS                                : num  18.6 1.65 -90.39 -7.0
9 5.9 ...
##  $ Total liabilities                           : num  971 675 532 858 823 .
..
##  $ PE on BSE                                   : chr  "NA" "NA" "-15.5" "-0
.16" ...
```

```
names(test.data)
```

```
##  [1] "Num"
##  [2] "Default - 1"
##  [3] "Total assets"
##  [4] "Net worth"
##  [5] "Total income"
##  [6] "Change in stock"
##  [7] "Total expenses"
##  [8] "Profit after tax"
##  [9] "PBDITA"
## [10] "PBT"
## [11] "Cash profit"
## [12] "PBDITA as % of total income"
## [13] "PBT as % of total income"
## [14] "PAT as % of total income"
## [15] "Cash profit as % of total income"
## [16] "PAT as % of net worth"
## [17] "Sales"
## [18] "Income from financial services"
## [19] "Other income"
## [20] "Total capital"
## [21] "Reserves and funds"
## [22] "Deposits (accepted by commercial banks)"
## [23] "Borrowings"
## [24] "Current liabilities & provisions"
## [25] "Deferred tax liability"
## [26] "Shareholders funds"
## [27] "Cumulative retained profits"
## [28] "Capital employed"
## [29] "TOL/TNW"
## [30] "Total term liabilities / tangible net worth"
## [31] "Contingent liabilities / Net worth (%)"
## [32] "Contingent liabilities"
## [33] "Net fixed assets"
## [34] "Investments"
## [35] "Current assets"
## [36] "Net working capital"
## [37] "Quick ratio (times)"
## [38] "Current ratio (times)"
## [39] "Debt to equity ratio (times)"
## [40] "Cash to current liabilities (times)"
## [41] "Cash to average cost of sales per day"
## [42] "Creditors turnover"
## [43] "Debtors turnover"
## [44] "Finished goods turnover"
## [45] "WIP turnover"
## [46] "Raw material turnover"
## [47] "Shares outstanding"
## [48] "Equity face value"
## [49] "EPS"
## [50] "Adjusted EPS"
## [51] "Total liabilities"
## [52] "PE on BSE"

colnames(test.data) = make.names(colnames(test.data))
```

The development dataset has 3541 rows and 52 columns. The validation dataset has 715 rows and 52 columns. The columns are same except for the second one. In the dev data it is “Networth Next Year”, which is a continuous variable with numerical values, whereas in the val data it is “Default – 1”, a factor variable with 0 and 1 values. There are also a lot of missing values.

3.3. Treating NA’s and Outliers

```
sum(is.na(data))
```

```
## [1] 13548
```

```
imputed.data = mice(data[,-c(1,22,42,43,44,45,46,47,48,52)], method = "pmm")
```

```
## 
##  iter imp variable
##   1   1  Change.in.stock  Total.expenses  Profit.after.tax  PBDITA  PBT  C
ash.profit  PBDITA.as...of.total.income  PBT.as...of.total.income  PAT.as...o
f.total.income  Cash.profit.as...of.total.income  Income.from.financial.servi
ces  Other.income  Total.capital  Reserves.and.funds  Borrowings  Current.lia
bilities...provisions  Deferred.tax.liability  Cumulative.retained.profits  C
ontingent.liabilities  Net.fixed.assets  Investments  Current.assets  Net.wor
king.capital  Quick.ratio..times.  Current.ratio..times.  Cash.to.current.lia
bilities..times.  Cash.to.average.cost.of.sales.per.day
##   1   2  Change.in.stock  Total.expenses  Profit.after.tax  PBDITA  PBT  C
ash.profit  PBDITA.as...of.total.income  PBT.as...of.total.income  PAT.as...o
f.total.income  Cash.profit.as...of.total.income  Income.from.financial.servi
ces  Other.income  Total.capital  Reserves.and.funds  Borrowings  Current.lia
bilities...provisions  Deferred.tax.liability  Cumulative.retained.profits  C
ontingent.liabilities  Net.fixed.assets  Investments  Current.assets  Net.wor
bilities...provisions  Deferred.tax.liability  Cumulative.retained.profits  C
ontingent.liabilities  Net.fixed.assets  Investments  Current.assets  Net.wor
king.capital  Quick.ratio..times.  Current.ratio..times.  Cash.to.current.lia
ces  Other.income  Total.capital  Reserves.and.funds  Borrowings  Current.lia
bilities...provisions  Deferred.tax.liability  Cumulative.retained.profits  C
ontingent.liabilities  Net.fixed.assets  Investments  Current.assets  Net.wor
king.capital  Quick.ratio..times.  Current.ratio..times.  Cash.to.current.lia
bilities..times.  Cash.to.average.cost.of.sales.per.day
##   5   4  Change.in.stock  Total.expenses  Profit.after.tax  PBDITA  PBT  C
ash.profit  PBDITA.as...of.total.income  PBT.as...of.total.income  PAT.as...o
f.total.income  Cash.profit.as...of.total.income  Income.from.financial.servi
ces  Other.income  Total.capital  Reserves.and.funds  Borrowings  Current.lia
bilities...provisions  Deferred.tax.liability  Cumulative.retained.profits  C
ontingent.liabilities  Net.fixed.assets  Investments  Current.assets  Net.wor
king.capital  Quick.ratio..times.  Current.ratio..times.  Cash.to.current.lia
bilities..times.  Cash.to.average.cost.of.sales.per.day
##   5   5  Change.in.stock  Total.expenses  Profit.after.tax  PBDITA  PBT  C
ash.profit  PBDITA.as...of.total.income  PBT.as...of.total.income  PAT.as...o
f.total.income  Cash.profit.as...of.total.income  Income.from.financial.servi
ces  Other.income  Total.capital  Reserves.and.funds  Borrowings  Current.lia
bilities...provisions  Deferred.tax.liability  Cumulative.retained.profits  C
ontingent.liabilities  Net.fixed.assets  Investments  Current.assets  Net.wor
king.capital  Quick.ratio..times.  Current.ratio..times.  Cash.to.current.lia
bilities..times.  Cash.to.average.cost.of.sales.per.day
```

```
## Warning: Number of logged events: 680
```

```
summary(imputed.data)

## Class: mids
## Number of multiple imputations:  5 
## Imputation methods:
##                          Networth.Next.Year 
##                                          "" 
##                                Total.assets 
##                                          "" 
##                                   Net.worth 
##                                          "" 
##                                Total.income 
##                                          "" 
##                             Change.in.stock 
##                                       "pmm" 
##                              Total.expenses 
##                                       "pmm" 
##                            Profit.after.tax 
##                                       "pmm" 
##                                      PBDITA 
##                                       "pmm" 
##                                         PBT 
##                                       "pmm" 
##                            Capital.employed 
##                                          "" 
##                                     TOL.TNW 
##                                          "" 
## Total.term.liabilities...tangible.net.worth 
##                                          "" 
##      Contingent.liabilities...Net.worth.... 
##                                          "" 
##                      Contingent.liabilities 
##                                       "pmm" 
##                            Net.fixed.assets 
##                                       "pmm" 
##                                 Investments 
##                                       "pmm" 
##                              Current.assets 
##                                       "pmm" 
##                         Net.working.capital 
##                                       "pmm" 
##                         Quick.ratio..times. 
##                                       "pmm" 
##                       Current.ratio..times. 
##                                       "pmm" 
##                Debt.to.equity.ratio..times. 
##                                          "" 
##         Cash.to.current.liabilities..times. 
##                                       "pmm" 
##       Cash.to.average.cost.of.sales.per.day 
##                                        "pmm"
##                                          EPS
##                                           ""
##                                 Adjusted.EPS
##                                           ""
##                            Total.liabilities
##                                           ""
## PredictorMatrix:
##                     Networth.Next.Year Total.assets Net.worth Total.income
## Networth.Next.Year                  0            1         1            0
## Total.assets                        1            0         1            0
## Net.worth                           1            1         0            0
## Total.income                        0            0         0            0
## Change.in.stock                     1            1         1            0
## Total.expenses                      1            1         1            0
##                     Change.in.stock Total.expenses Profit.after.tax PBDITA
## Networth.Next.Year                1              1                1      1
## Total.assets                      1              1                1      1
## Net.worth                         1              1                1      1
## Total.income                      0              0                0      0
## Change.in.stock                   0              1                1      1
## Total.expenses                    1              0                1      1
## 3  0  0                 collinear
## 4  0  0                 collinear
## 5  0  0                 collinear
## 6  1  1 Change.in.stock       pmm
##
out
## 1
Shareholders.funds
## 2
Adjusted.EPS
## 3
Total.liabilities
## 4
Total.income
## 5
Sales
## 6 Total.assets, Total.expenses, PBDITA.as...of.total.income, PAT.as...of.t
otal.income, Cash.profit.as...of.total.income
```

```
complete.data = complete(imputed.data,1)
summary(complete.data)
```

```
##  Networth.Next.Year  Total.assets         Net.worth
##  Min.   :-74265.6   Min.   :      0.1   Min.   :     0.0
##  1st Qu.:    31.7   1st Qu.:     91.3   1st Qu.:    31.3
##  Median :   116.3   Median :    309.7   Median :   102.3
##  Mean   :  1616.3   Mean   :   3443.4   Mean   :  1295.9
##  3rd Qu.:   456.1   3rd Qu.:   1098.7   3rd Qu.:   377.3
##  Max.   :805773.4   Max.   :1176509.2   Max.   :613151.6
##
##   Total.income       Change.in.stock    Total.expenses
##  Min.   :      0.0   Min.   :-3029.40   Min.   :     -0.1
##  1st Qu.:    106.5   1st Qu.:   -1.90   1st Qu.:     96.3
##  Median :    444.9   Median :    1.40   Median :    421.8
##  Mean   :   4582.8   Mean   :   38.09   Mean   :   4197.0
##  3rd Qu.:   1440.9   3rd Qu.:   17.60   3rd Qu.:   1383.7
##  Max.   :2442828.2   Max.   :14185.50   Max.   :2366035.3
##  Median :     8.13                     Median :      1.4
##  Mean   :   165.50                     Mean   :   -220.3
##  3rd Qu.:    22.59                     3rd Qu.:      9.6
##  Max.   :128040.76                     Max.   :  34522.5
##
##   Adjusted.EPS       Total.liabilities
##  Min.   :-843181.8   Min.   :      0.1
##  1st Qu.:      0.0   1st Qu.:     91.3
##  Median :      1.2   Median :    309.7
##  Mean   :   -221.5   Mean   :   3443.4
##  3rd Qu.:      7.5   3rd Qu.:   1098.7
##  Max.   :  34522.5   Max.   :1176509.2
##
```

```
new.data = complete.data[,-c(4,16)]
```

```
summary(new.data)
```

```
##  Networth.Next.Year  Total.assets         Net.worth
##  Min.   :-74265.6   Min.   :      0.1   Min.   :     0.0
##  1st Qu.:    31.7   1st Qu.:     91.3   1st Qu.:    31.3
##  Median :   116.3   Median :    309.7   Median :   102.3
##  Mean   :  1616.3   Mean   :   3443.4   Mean   :  1295.9
##  3rd Qu.:   456.1   3rd Qu.:   1098.7   3rd Qu.:   377.3
##  Max.   :805773.4   Max.   :1176509.2   Max.   :613151.6
##  Change.in.stock     Total.expenses      Profit.after.tax
##  Min.   :-3029.40   Min.   :     -0.1   Min.   : -3908.3
##  1st Qu.:   -1.90   1st Qu.:     96.3   1st Qu.:     0.4
##  Median :    1.40   Median :    421.8   Median :     8.9
##  Mean   :   38.09   Mean   :   4197.0   Mean   :   278.4
##  Max.   :613151.6   Max.   :390133.8           Max.   :891408.9
##     TOL.TNW         Total.term.liabilities...tangible.net.worth
##  Min.   :-350.480   Min.   :-325.600
##  1st Qu.:   0.600   1st Qu.:   0.050
##  Median :   1.430   Median :   0.340
##  Mean   :   3.994   Mean   :   1.844
##  3rd Qu.:   2.830   3rd Qu.:   1.000
##  Max.   : 473.000   Max.   : 456.000
##  Contingent.liabilities...Net.worth.... Contingent.liabilities
##  Min.   :    0.00                        Min.   :     0.1
##  1st Qu.:    0.00                        1st Qu.:     5.9
##  Median :    5.33                       Median :    32.7  
##  Mean   :   53.94                       Mean   :   661.2  
##  3rd Qu.:   30.76                       3rd Qu.:   151.4  
##  Max.   :14704.27                       Max.   :559506.8  
##  Net.fixed.assets    Investments       Current.assets  
##  Min.   :     0.0   Min.   :     0.0   Min.   :     0.1  
##  1st Qu.:    26.0   1st Qu.:     0.8   1st Qu.:    36.2  
##  Median :    94.6   Median :     5.6   Median :   145.5  
##  Mean   :  1164.5   Mean   :   449.1   Mean   :  1278.9  
##  3rd Qu.:   346.1   3rd Qu.:    44.6   3rd Qu.:   502.9  
##  Max.   :636604.6   Max.   :199978.6   Max.   :354815.2  
##  Net.working.capital Quick.ratio..times. Current.ratio..times.
##  Min.   :-63839.0    Min.   :  0.000     Min.   :  0.000      
##  1st Qu.:    -1.3    1st Qu.:  0.410     1st Qu.:  0.930      
##  Median :    16.1    Median :  0.670     Median :  1.230      
##  Mean   :   135.0    Mean   :  1.387     Mean   :  2.112      
##  3rd Qu.:    84.2    3rd Qu.:  1.030     3rd Qu.:  1.710      
##  Max.   : 85782.8    Max.   :341.000     Max.   :505.000      
##  Debt.to.equity.ratio..times. Cash.to.current.liabilities..times.
##  Min.   :  0.00               Min.   :  0.000                    
##  1st Qu.:  0.22               1st Qu.:  0.020                    
##  Median :  0.79               Median :  0.070                    
##  Mean   :  2.78               Mean   :  0.491                    
##  3rd Qu.:  1.75               3rd Qu.:  0.190                    
##  Max.   :456.00               Max.   :165.000                    
##  Cash.to.average.cost.of.sales.per.day      EPS           
##  Min.   :     0.00                     Min.   :-843181.8  
##  1st Qu.:     2.82                     1st Qu.:      0.0  
##  Median :     8.13                     Median :      1.4  
##  Mean   :   165.50                     Mean   :   -220.3  
##  3rd Qu.:    22.59                     3rd Qu.:      9.6  
##  Max.   :128040.76                     Max.   :  34522.5  
##   Adjusted.EPS       Total.liabilities
##  Min.   :-843181.8   Min.   :      0.1  
##  1st Qu.:      0.0   1st Qu.:     91.3  
##  Median :      1.2   Median :    309.7  
##  Mean   :   -221.5   Mean   :   3443.4  
##  3rd Qu.:      7.5   3rd Qu.:   1098.7  
##  Max.   :  34522.5   Max.   :1176509.2
```

```
imputed.data2 = mice(test.data[,-c(1,2,22,42,43,44,45,46,47,48,52)], method =
"pmm")
```

```
## 
##  iter imp variable
##   1   1  Total.income  Change.in.stock  Total.expenses  Profit.after.tax
PBDITA  PBT  Cash.profit  PBDITA.as...of.total.income  PBT.as...of.total.inco
me  Cash.profit.as...of.total.income  Income.from.financial.services  Other.i
ncome  Total.capital  Reserves.and.funds  Borrowings  Current.liabilities...p
rovisions  Deferred.tax.liability  Cumulative.retained.profits  Contingent.li
```

```
abilities  Net.fixed.assets  Investments  Current.assets  Net.working.capital
Quick.ratio..times.  Current.ratio..times.  Cash.to.current.liabilities..time
s.  Cash.to.average.cost.of.sales.per.day
##   1   2  Total.income  Change.in.stock  Total.expenses  Profit.after.tax
PBDITA  PBT  Cash.profit  PBDITA.as...of.total.income  PBT.as...of.total.inco
me  Cash.profit.as...of.total.income  Income.from.financial.services  Other.i
ncome  Total.capital  Reserves.and.funds  Borrowings  Current.liabilities...p
rovisions  Deferred.tax.liability  Cumulative.retained.profits  Contingent.li
abilities  Net.fixed.assets  Investments  Current.assets  Net.working.capital
Quick.ratio..times.  Current.ratio..times.  Cash.to.current.liabilities..time
ncome  Total.capital  Reserves.and.funds  Borrowings  Current.liabilities...p
rovisions  Deferred.tax.liability  Cumulative.retained.profits  Contingent.li
abilities  Net.fixed.assets  Investments  Current.assets  Net.working.capital
Quick.ratio..times.  Current.ratio..times.  Cash.to.current.liabilities..time
s.  Cash.to.average.cost.of.sales.per.day
##   5   4  Total.income  Change.in.stock  Total.expenses  Profit.after.tax
PBDITA  PBT  Cash.profit  PBDITA.as...of.total.income  PBT.as...of.total.inco
me  Cash.profit.as...of.total.income  Income.from.financial.services  Other.i
ncome  Total.capital  Reserves.and.funds  Borrowings  Current.liabilities...p
rovisions  Deferred.tax.liability  Cumulative.retained.profits  Contingent.li
abilities  Net.fixed.assets  Investments  Current.assets  Net.working.capital
Quick.ratio..times.  Current.ratio..times.  Cash.to.current.liabilities..time
s.  Cash.to.average.cost.of.sales.per.day
##   5   5  Total.income  Change.in.stock  Total.expenses  Profit.after.tax
PBDITA  PBT  Cash.profit  PBDITA.as...of.total.income  PBT.as...of.total.inco
me  Cash.profit.as...of.total.income  Income.from.financial.services  Other.i
ncome  Total.capital  Reserves.and.funds  Borrowings  Current.liabilities...p
rovisions  Deferred.tax.liability  Cumulative.retained.profits  Contingent.li
abilities  Net.fixed.assets  Investments  Current.assets  Net.working.capital
Quick.ratio..times.  Current.ratio..times.  Cash.to.current.liabilities..time
s.  Cash.to.average.cost.of.sales.per.day

## Warning: Number of logged events: 680
```

```r
summary(imputed.data2)
```

```
## Class: mids
## Number of multiple imputations:  5
## Imputation methods:
##                                 Total.assets
##                                           ""
##                                    Net.worth
##                                           ""
##                                 Total.income
##                                        "pmm"
##                              Change.in.stock
##                                        "pmm"
##                               Total.expenses
##                                        "pmm"
## Profit.after.tax                        0                                1
##                  PAT.as...of.net.worth Sales
##                  Contingent.liabilities...Net.worth....
## Total.assets                                          1
## Net.worth                                             1
## Total.income                                          1
## Change.in.stock                                       1
## Total.expenses                                        1
## Change.in.stock                      1                            1
## Total.expenses                       1                            1
## Profit.after.tax                     1                            1
##                  Cash.to.current.liabilities..times.
## Total.assets                                      1
## Net.worth                                         1
## Total.income                                      1
## Change.in.stock                                   1
## Total.expenses                                    1
## Profit.after.tax                                  1
##                  Cash.to.average.cost.of.sales.per.day EPS Adjusted.EPS
## Total.assets                                        1   1            0
## Net.worth                                           1   1            0
## Total.income                                        1   1            0
## Change.in.stock                                     1   1            0
## Total.expenses                                      1   1            0
## Profit.after.tax                                    1   1            0
##                  Total.liabilities
## Total.assets                     0
## Net.worth                        0
## Total.income                     0
## Change.in.stock                  0
## Total.expenses                   0
## Profit.after.tax                 0
## Number of logged events:  680
##   it im          dep      meth
## 1  0  0              collinear
## 2  0  0              collinear
## 3  0  0              collinear
## 4  0  0              collinear
## 5  0  0              collinear
## 6  1  1 Total.income       pmm
##
out
## 1                                                                    Shar
eholders.funds
## 2
Adjusted.EPS
## 3                                                                     Tot
al.liabilities
## 4                                                              PAT.as...o
f.total.income
## 5
```

```
Sales
## 6 Total.assets, Change.in.stock, Total.expenses, PBT, Cash.profit, PAT.as.
..of.net.worth
```

```
complete.data2 = complete(imputed.data2,1)
summary(complete.data2)
```

```
##   Total.assets        Net.worth         Total.income     
##  Min.   :     0.1   Min.   :     0.1   Min.   :      0.0  
##  1st Qu.:    93.2   1st Qu.:    34.4   1st Qu.:    109.1  
##  Median :   347.7   Median :   120.9   Median :    536.7  
##  Mean   :  4218.6   Mean   :  1629.7   Mean   :   5019.0  
##  3rd Qu.:  1315.3   3rd Qu.:   451.5   3rd Qu.:   1721.8  
##  Max.   :354727.3   Max.   :171840.0   Max.   :1028087.4  
##                                                           
##  Change.in.stock   Total.expenses      Profit.after.tax 
##  Min.   :-488.10   Min.   :      0.0   Min.   : -998.0  
##  1st Qu.:  -1.90   1st Qu.:    104.4   1st Qu.:    0.6  
##  Median :   1.80   Median :    511.1   Median :    9.8  
##  Mean   :  57.35   Mean   :   4678.6   Mean   :  376.4  
##  3rd Qu.:  20.40   3rd Qu.:   1627.1   3rd Qu.:   68.5  
##  Max.   :7540.00   Max.   :1014813.1   Max.   :62022.9  
##                                                         
##      PBDITA               PBT           Cash.profit    
##  Min.   :  -393.90   Min.   : -993.9   Min.   : -894.6  
##  1st Qu.:     6.85   1st Qu.:    1.0   1st Qu.:    3.0  
##                                                         
##  Cash.to.average.cost.of.sales.per.day      EPS          
##  Min.   :    0.00                      Min.   :-72750.00  
##  1st Qu.:    3.35                      1st Qu.:     0.00  
##  Median :    8.38                      Median :     1.83  
##  Mean   :   78.86                      Mean   :   -76.87  
##  3rd Qu.:   23.39                      3rd Qu.:    11.46  
##  Max.   :15999.17                      Max.   :  8784.00  
##                                                           
##   Adjusted.EPS       Total.liabilities
##  Min.   :-72750.00   Min.   :     0.1  
##  1st Qu.:     0.00   1st Qu.:    93.2  
##  Median :     1.50   Median :   347.7  
##  Mean   :   -78.74   Mean   :  4218.6  
##  3rd Qu.:     8.35   3rd Qu.:  1315.3  
##  Max.   :  8784.00   Max.   :354727.3  
##                                        
```

```
names(complete.data2)
```

```
##  [1] "Total.assets"
##  [2] "Net.worth"
##  [3] "Total.income"
##  [4] "Change.in.stock"
##  [5] "Total.expenses"
##  [6] "Profit.after.tax"
##  [7] "PBDITA"
##  [8] "PBT"
##  [9] "Cash.profit"
## [10] "PBDITA.as...of.total.income"
## [11] "PBT.as...of.total.income"
## [12] "PAT.as...of.total.income"
## [13] "Cash.profit.as...of.total.income"
## [14] "PAT.as...of.net.worth"
## [15] "Sales"
## [16] "Income.from.financial.services"
## [17] "Other.income"
## [18] "Total.capital"
## [19] "Reserves.and.funds"
## [20] "Borrowings"
## [21] "Current.liabilities...provisions"
## [22] "Deferred.tax.liability"
## [23] "Shareholders.funds"
## [24] "Cumulative.retained.profits"
## [25] "Capital.employed"
## [26] "TOL.TNW"
## [27] "Total.term.liabilities...tangible.net.worth"
## [28] "Contingent.liabilities...Net.worth...."
## [29] "Contingent.liabilities"
## [30] "Net.fixed.assets"
## [31] "Investments"
## [32] "Current.assets"
## [33] "Net.working.capital"
## [34] "Quick.ratio..times."
## [35] "Current.ratio..times."
## [36] "Debt.to.equity.ratio..times."
## [37] "Cash.to.current.liabilities..times."
## [38] "Cash.to.average.cost.of.sales.per.day"
## [39] "EPS"
## [40] "Adjusted.EPS"
## [41] "Total.liabilities"
```

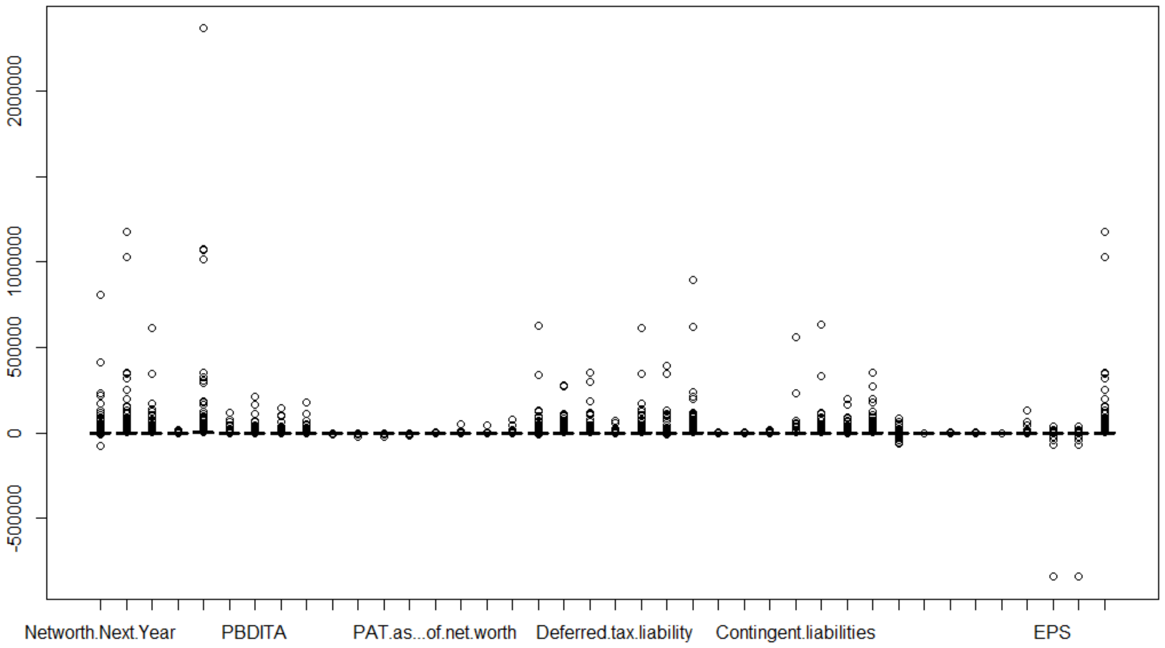
```
new.test.data = complete.data2[,c(10,35,36,39)]
summary(new.test.data)
```

```
##  PBDITA.as...of.total.income Current.ratio..times.
##  Min.   :-6400.000           Min.   :  0.00
##  1st Qu.:    4.770           1st Qu.:  0.92
##  Median :    9.770           Median :  1.24
##  Mean   :   -3.416           Mean   :  2.89
##  3rd Qu.:   16.765           3rd Qu.:  1.73
##  Max.   :  100.000           Max.   :505.00
##  Debt.to.equity.ratio..times.      EPS
##  Min.   :  0.000               Min.   :-72750.00
##  1st Qu.:  0.220               1st Qu.:     0.00
##  Median :  0.800              Median :     1.83
##  Mean   :  3.327              Mean   :   -76.87
##  3rd Qu.:  1.700              3rd Qu.:    11.46
##  Max.   :341.180              Max.   :  8784.00
```

```
boxplot(new.data)
```

We use the mice package in R to impute the missing values in the dataset by using the predictive mean matching method. We also clean up the data to remove unnecessary variables and any underlying missing values.

Next, we check for outliers.



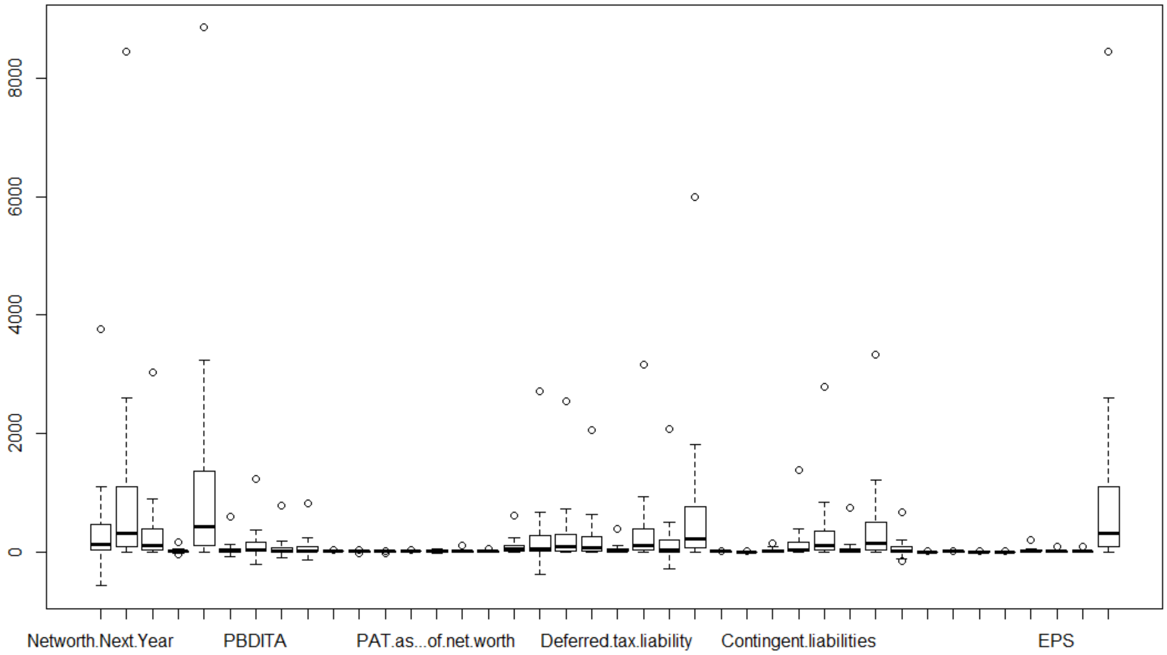
We clearly see multiple outliers in the data, which may introduce bias when executing a predictive model. Therefore, we proceed to cap the outliers within the 95th and 5th percentile of the data.

```
qnt = quantile(new.data[,1], probs = c(.25, .75),na.rm = T)
caps = quantile(new.data[,1], probs = c(.05, .95), na.rm = T)
H = 1.5 * IQR(new.data[,1])
new.data[,1][new.data[,1] < (qnt[1] - H)] = caps[1]
new.data[,1][new.data[,1] > (qnt[2] + H)] = caps[2]

qnt = quantile(new.data[,2], probs = c(.25, .75),na.rm = T)
caps = quantile(new.data[,2], probs = c(.05, .95), na.rm = T)
```

```
H = 1.5 * IQR(new.data[,2])
new.data[,2][new.data[,2] < (qnt[1] - H)] = caps[1]
new.data[,2][new.data[,2] > (qnt[2] + H)] = caps[2]

qnt = quantile(new.data[,3], probs = c(.25, .75),na.rm = T)
caps = quantile(new.data[,3], probs = c(.05, .95), na.rm = T)
H = 1.5 * IQR(new.data[,3])
new.data[,3][new.data[,3] < (qnt[1] - H)] = caps[1]
new.data[,3][new.data[,3] > (qnt[2] + H)] = caps[2]

qnt = quantile(new.data[,4], probs = c(.25, .75),na.rm = T)
caps = quantile(new.data[,4], probs = c(.05, .95), na.rm = T)
H = 1.5 * IQR(new.data[,4])
new.data[,4][new.data[,4] < (qnt[1] - H)] = caps[1]
new.data[,4][new.data[,4] > (qnt[2] + H)] = caps[2]

qnt = quantile(new.data[,36], probs = c(.25, .75),na.rm = T)
caps = quantile(new.data[,36], probs = c(.05, .95), na.rm = T)
H = 1.5 * IQR(new.data[,36])
new.data[,36][new.data[,36] < (qnt[1] - H)] = caps[1]
new.data[,36][new.data[,36] > (qnt[2] + H)] = caps[2]

qnt = quantile(new.data[,37], probs = c(.25, .75),na.rm = T)
caps = quantile(new.data[,37], probs = c(.05, .95), na.rm = T)
H = 1.5 * IQR(new.data[,37])
new.data[,37][new.data[,37] < (qnt[1] - H)] = caps[1]
new.data[,37][new.data[,37] > (qnt[2] + H)] = caps[2]

qnt = quantile(new.data[,38], probs = c(.25, .75),na.rm = T)
caps = quantile(new.data[,38], probs = c(.05, .95), na.rm = T)
H = 1.5 * IQR(new.data[,38])
new.data[,38][new.data[,38] < (qnt[1] - H)] = caps[1]
new.data[,38][new.data[,38] > (qnt[2] + H)] = caps[2]

qnt = quantile(new.data[,39], probs = c(.25, .75),na.rm = T)
caps = quantile(new.data[,39], probs = c(.05, .95), na.rm = T)
H = 1.5 * IQR(new.data[,39])
new.data[,39][new.data[,39] < (qnt[1] - H)] = caps[1]
new.data[,39][new.data[,39] > (qnt[2] + H)] = caps[2]

qnt = quantile(new.data[,40], probs = c(.25, .75),na.rm = T)
caps = quantile(new.data[,40], probs = c(.05, .95), na.rm = T)
H = 1.5 * IQR(new.data[,40])
new.data[,40][new.data[,40] < (qnt[1] - H)] = caps[1]
new.data[,40][new.data[,40] > (qnt[2] + H)] = caps[2]

summary(new.data)
```

```
##  Networth.Next.Year  Total.assets      Net.worth      Change.in.stock
##  Min.   :-579.6     Min.   :   0.1   Min.   :   0.0   Min.   :-41.70
##  1st Qu.:  31.7     1st Qu.:  91.3   1st Qu.:  31.3   1st Qu.: -1.40
##  Median : 116.3     Median : 309.7   Median : 102.3   Median :  0.90
##  Mean   : 681.7     Mean   :1553.9   Mean   : 559.3   Mean   : 19.84
##  3rd Qu.: 456.1     3rd Qu.:1098.7   3rd Qu.: 377.3   3rd Qu.: 14.70
##  Max.   :3764.4     Max.   :8452.9   Max.   :3034.4   Max.   :146.20
##  Total.expenses   Profit.after.tax     PBDITA            PBT
##  Min.   :  -0.1   Min.   :-70.90   Min.   :-158.1   Min.   :-97.5
##  1st Qu.:  79.8   1st Qu.:  0.30   1st Qu.:   5.3   1st Qu.:  0.4
##  Median : 370.9   Median :  7.60   Median :  32.2   Median : 10.5
##  Mean   :1585.7   Mean   : 97.01   Mean   : 218.9   Mean   :129.0
##  3rd Qu.:1300.1   3rd Qu.: 48.30   3rd Qu.: 140.1   3rd Qu.: 67.7
##  Max.   :8592.3   Max.   :562.40   Max.   :1219.1   Max.   :735.0
##   Cash.profit     PBDITA.as...of.total.income PBT.as...of.total.income
##  Min.   :-121.7   Min.   :-12.50              Min.   :-24.49
##  1st Qu.:   2.0   1st Qu.:  4.69              1st Qu.:  0.43
##  Median :  16.8   Median :  9.41              Median :  3.17
##  Mean   : 141.0   Mean   : 11.19              Mean   :  3.68
##  3rd Qu.:  87.2   3rd Qu.: 16.16              3rd Qu.:  8.64
##  Max.   : 803.0   Max.   : 34.21              Max.   : 22.89
##  PAT.as...of.total.income Cash.profit.as...of.total.income
##  Min.   :-24.490          Min.   :-11.200
##  1st Qu.:  0.250          1st Qu.:  1.820
##  Median :  2.270          Median :  5.490
##  Mean   :  2.117          Mean   :  6.443
##  3rd Qu.:  6.260          3rd Qu.: 10.560
##  Max.   : 18.110          Max.   : 24.370
##  PAT.as...of.net.worth Income.from.financial.services  Other.income
##  Min.   :-30.09        Min.   : 0.00                  Min.   : 0.00
##  1st Qu.:  0.00        1st Qu.: 0.20                  1st Qu.: 0.20
##  Median :  7.92        Median : 1.00                  Median : 0.70
##  Mean   : 10.66        Mean   :16.26                  Mean   : 6.82
##  3rd Qu.: 20.19        3rd Qu.: 6.40                  3rd Qu.: 3.50
##  Max.   : 50.46        Max.   :93.60                  Max.   :38.90
##  Total.capital   Reserves.and.funds   Borrowings
##  Min.   :  0.1   Min.   :-351.7     Min.   :   0.1
##  1st Qu.: 13.1   1st Qu.:   3.9     1st Qu.:  14.2
##  Median : 42.1   Median :  51.1     Median :  77.6
##  Mean   :118.9   Mean   : 479.3     Mean   : 466.4
##  3rd Qu.:100.3   3rd Qu.: 264.2     3rd Qu.: 301.5
##  Max.   :607.6   Max.   :2734.1     Max.   :2583.2
##  Current.liabilities...provisions Deferred.tax.liability
##  Min.   :   0.1                   Min.   :  0.10
##  1st Qu.:  15.3                   1st Qu.:  1.70
##  Median :  64.7                   Median :  8.10
##  Mean   : 365.9                   Mean   : 58.46
##  3rd Qu.: 249.1                   3rd Qu.: 36.90
##  Max.   :2005.6                   Max.   :340.80
##  Shareholders.funds Cumulative.retained.profits Capital.employed
##  Min.   :   0.0     Min.   :-288.2              Min.   :   0.0  
##  1st Qu.:  32.0     1st Qu.:   0.9              1st Qu.:  60.8  
##  Median : 105.6     Median :  36.2              Median : 214.7  
##  Mean   : 579.5     Mean   : 347.2              Mean   :1085.1  
##  3rd Qu.: 393.2     3rd Qu.: 199.4              3rd Qu.: 767.3  
##  Max.   :3160.0     Max.   :2024.5              Max.   :5988.7  
##     TOL.TNW       Total.term.liabilities...tangible.net.worth
##  Min.   :-2.410   Min.   :-1.2500                            
##  1st Qu.: 0.600   1st Qu.: 0.0500                            
##  Median : 1.430   Median : 0.3400                            
##  Mean   : 2.411   Mean   : 0.8236                            
##  3rd Qu.: 2.830   3rd Qu.: 1.0000                            
##  Max.   :10.530   Max.   : 4.2000                            
##  Contingent.liabilities...Net.worth.... Contingent.liabilities
##  Min.   :  0.00                         Min.   :   0.1        
##  1st Qu.:  0.00                         1st Qu.:   0.3        
##  Median :  5.33                         Median :   8.3        
##  Mean   : 27.87                         Mean   : 204.5        
##  3rd Qu.: 30.76                         3rd Qu.:  87.7        
##  Max.   :151.04                         Max.   :1158.2        
##  Net.fixed.assets  Investments    Current.assets   Net.working.capital
##  Min.   :   0.0   Min.   :  0.0   Min.   :   0.1   Min.   :-156.3     
##  1st Qu.:  23.3   1st Qu.:  0.4   1st Qu.:  33.0   1st Qu.:  -1.1     
##  Median :  87.6   Median :  3.5   Median : 138.6   Median :  15.7     
##  Mean   : 483.5   Mean   :118.9   Mean   : 608.8   Mean   : 102.7     
##  3rd Qu.: 329.7   3rd Qu.: 29.2   3rd Qu.: 488.0   3rd Qu.:  81.7     
##  Max.   :2689.5   Max.   :664.0   Max.   :3300.4   Max.   : 673.8     
##  Quick.ratio..times. Current.ratio..times. Debt.to.equity.ratio..times.
##  Min.   :0.0000      Min.   :0.00          Min.   :0.00                
##  1st Qu.:0.4000      1st Qu.:0.92          1st Qu.:0.22                
##  Median :0.6600      Median :1.22          Median :0.79                
##  Mean   :0.8794      Mean   :1.52          Mean   :1.44                
##  3rd Qu.:1.0300      3rd Qu.:1.71          3rd Qu.:1.75                
##  Max.   :2.9800      Max.   :4.34          Max.   :6.75                
##  Cash.to.current.liabilities..times. Cash.to.average.cost.of.sales.per.day
##  Min.   :0.0000                      Min.   :  0.00                       
##  1st Qu.:0.0200                      1st Qu.:  2.66                       
##  Median :0.0700                      Median :  7.95                       
##  Mean   :0.2393                      Mean   : 36.53                       
##  3rd Qu.:0.1900                      3rd Qu.: 22.03                       
##  Max.   :1.2500                      Max.   :199.72                       
##       EPS          Adjusted.EPS    Total.liabilities
##  Min.   :-14.24   Min.   :-10.88   Min.   :   0.1   
##  1st Qu.:  0.00   1st Qu.:  0.00   1st Qu.:  91.3   
##  Median :  1.43   Median :  1.18   Median : 309.7   
##  Mean   : 14.22   Mean   : 13.74   Mean   :1553.9   
##  3rd Qu.:  9.62   3rd Qu.:  7.48   3rd Qu.:1098.7   
##  Max.   : 87.71   Max.   : 84.23   Max.   :8452.9
```

```
boxplot(new.data)
```



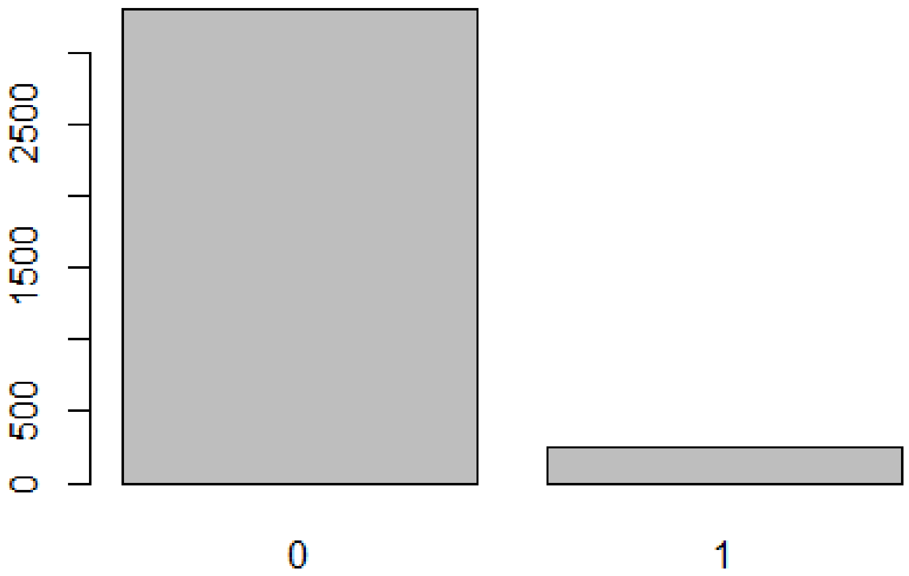
As evidenced by the two boxplots, the number of outliers have been brought down significantly enough to majorly remove bias that may have been present in the dataset.

3.4. Univariate Analysis

```
names(new.data)

##  [1] "Networth.Next.Year"
##  [2] "Total.assets"
##  [3] "Net.worth"
##  [4] "Change.in.stock"
##  [5] "Total.expenses"
##  [6] "Profit.after.tax"
##  [7] "PBDITA"
##  [8] "PBT"
##  [9] "Cash.profit"
## [10] "PBDITA.as...of.total.income"
## [11] "PBT.as...of.total.income"
## [12] "PAT.as...of.total.income"
## [13] "Cash.profit.as...of.total.income"
## [14] "PAT.as...of.net.worth"
## [15] "Income.from.financial.services"
## [16] "Other.income"
## [17] "Total.capital"
## [18] "Reserves.and.funds"
## [19] "Borrowings"
## [20] "Current.liabilities...provisions"
## [21] "Deferred.tax.liability"
## [22] "Shareholders.funds"
## [23] "Cumulative.retained.profits"
## [24] "Capital.employed"
## [25] "TOL.TNW"
## [26] "Total.term.liabilities...tangible.net.worth"
## [27] "Contingent.liabilities...Net.worth...."
## [28] "Contingent.liabilities"
## [29] "Net.fixed.assets"
## [30] "Investments"
## [31] "Current.assets"
## [32] "Net.working.capital"
## [33] "Quick.ratio..times."
## [34] "Current.ratio..times."
## [35] "Debt.to.equity.ratio..times."
## [36] "Cash.to.current.liabilities..times."
## [37] "Cash.to.average.cost.of.sales.per.day"
## [38] "EPS"
## [39] "Adjusted.EPS"
## [40] "Total.liabilities"
```

We create the new variable “Default” in the development data using the values of “Networth Next Year” variable. We take positive values as 0 and negative values as 1. We do this in order to be able to compare our model with the validation data.

```r
new.data$Default = ifelse(new.data$Networth.Next.Year > 0,0,1)
new.data$Default = as.factor(new.data$Default)


summary(new.data$Default)

##    0    1
## 3298  243

plot(new.data$Default)
```



We observe 243 defaulters and 3298 non-defaulters based on our estimation. The ratio of defaulters to non-defaulters is 7.36%. Data may be too imbalanced to give satisfactory results. May need to consider SMOTE.

```r
243/3298

## [1] 0.07368102

final.data = new.data[,c(10,11,12,13,14,20,25,26,27,33,34,35,36,37,38,39,41)]
summary(final.data)
```

```
##  PBDITA.as...of.total.income PBT.as...of.total.income
##  Min.   :-12.50              Min.   :-24.49
##  1st Qu.:  4.69              1st Qu.:  0.43
##  Median :  9.41              Median :  3.17
##  Mean   : 11.19              Mean   :  3.68
##  3rd Qu.: 16.16              3rd Qu.:  8.64
##  Max.   : 34.21              Max.   : 22.89
##  PAT.as...of.total.income Cash.profit.as...of.total.income
##  Min.   :-24.490          Min.   :-11.200
##  1st Qu.:  0.250          1st Qu.:  1.820
##  Median :  2.270          Median :  5.490
##  Mean   :  2.117          Mean   :  6.443
##  3rd Qu.:  6.260          3rd Qu.: 10.560
##  Max.   : 18.110          Max.   : 24.370
##  PAT.as...of.net.worth Current.liabilities...provisions    TOL.TNW
##  Min.   :-30.09        Min.   :   0.1                   Min.   :-2.410
##  1st Qu.:  0.00        1st Qu.:  15.3                   1st Qu.: 0.600
##  Median :  7.92        Median :  64.7                   Median : 1.430
##  Mean   : 10.66        Mean   : 365.9                   Mean   : 2.411
##  3rd Qu.: 20.19        3rd Qu.: 249.1                   3rd Qu.: 2.830
##  Max.   : 50.46        Max.   :2005.6                   Max.   :10.530
##  Total.term.liabilities...tangible.net.worth
##  Min.   :-1.2500
##  1st Qu.: 0.0500
##  Median : 0.3400
##  Mean   : 0.8236
##  3rd Qu.: 1.0000
##  Max.   : 4.2000
##  Contingent.liabilities...Net.worth.... Quick.ratio..times.
##  Min.   :  0.00                         Min.   :0.0000
##  1st Qu.:  0.00                         1st Qu.:0.4000
##  Median :  5.33                         Median :0.6600
##  Mean   : 27.87                         Mean   :0.8794
##  3rd Qu.: 30.76                         3rd Qu.:1.0300
##  Max.   :151.04                         Max.   :2.9800
##  Current.ratio..times. Debt.to.equity.ratio..times.
##  Min.   :0.00          Min.   :0.00
##  1st Qu.:0.92          1st Qu.:0.22
##  Median :1.22          Median :0.79
##  Mean   :1.52          Mean   :1.44
##  3rd Qu.:1.71          3rd Qu.:1.75
##  Max.   :4.34          Max.   :6.75
##  Cash.to.current.liabilities..times. Cash.to.average.cost.of.sales.per.day
##  Min.   :0.0000                      Min.   :  0.00
##  1st Qu.:0.0200                      1st Qu.:  2.66
##  Median :0.0700                      Median :  7.95
##  Mean   :0.2393                      Mean   : 36.53
##  3rd Qu.:0.1900                      3rd Qu.: 22.03
##  Max.   :1.2500                      Max.   :199.72
##       EPS          Adjusted.EPS    Default
##  Min.   :-14.24   Min.   :-10.88   0:3298
##  1st Qu.:  0.00   1st Qu.:  0.00   1: 243
##  Median :  1.43   Median :  1.18
##  Mean   : 14.22   Mean   : 13.74
##  3rd Qu.:  9.62   3rd Qu.:  7.48
##  Max.   : 87.71   Max.   : 84.23
```

```
str(final.data)
```

```
## 'data.frame':    3541 obs. of  17 variables:
##  $ PBDITA.as...of.total.income                : num  11.46 18.53 1.22 0 1.
96 ...
##  $ PBT.as...of.total.income                   : num  9.68 12.33 -1.38 0 0.
4 ...
##  $ PAT.as...of.total.income                   : num  6.18 7.54 -1.38 0 0.3
5 2.81 0 0.72 8.29 -2.88 ...
##  $ Cash.profit.as...of.total.income           : num  7.5 10.38 0.06 0 0.75
...
##  $ PAT.as...of.net.worth                      : num  23.78 38.08 -6.35 0 5
.25 ...
##  $ Current.liabilities...provisions           : num  2005.6 210 96.8 0.3 1
12.8 ...
##  $ TOL.TNW                                    : num  1.33 1.23 1.44 0 2.83
1.8 0.03 5.17 1.05 3.25 ...
##  $ Total.term.liabilities...tangible.net.worth: num  0 0.34 0.29 0 1.59 0.
37 0.03 0.94 0.3 0.54 ...
##  $ Contingent.liabilities...Net.worth....     : num  14.8 19.2 45.8 0 34.9
...
##  $ Quick.ratio..times.                        : num  1.18 0.95 1.11 0 1.41
0.48 0.42 0.54 0.59 0.39 ...
##  $ Current.ratio..times.                      : num  1.37 1.56 1.55 0 2.54
1.27 1.17 1.15 1.58 0.5 ...
##  $ Debt.to.equity.ratio..times.               : num  0 0.78 0.35 0 1.79 1.
09 0.32 2.31 0.94 3.13 ...
##  $ Cash.to.current.liabilities..times.        : num  0.43 0.06 0.21 0 0 0.
11 0.01 0.04 0.19 0 ...
##  $ Cash.to.average.cost.of.sales.per.day      : num  199.72 5.96 17.07 0 0
...
##  $ EPS                                        : num  87.71 9.97 -0.5 0 7.9
1 ...
##  $ Adjusted.EPS                               : num  7.1 9.97 -0.5 0 7.91
...
##  $ Default                                    : Factor w/ 2 levels "0","1"
: 1 1 1 1 1 1 1 1 1 2 ...
```

```
names(final.data)
```

```
##  [1] "PBDITA.as...of.total.income"
##  [2] "PBT.as...of.total.income"
##  [3] "PAT.as...of.total.income"
##  [4] "Cash.profit.as...of.total.income"
##  [5] "PAT.as...of.net.worth"
##  [6] "Current.liabilities...provisions"
##  [7] "TOL.TNW"
##  [8] "Total.term.liabilities...tangible.net.worth"
##  [9] "Contingent.liabilities...Net.worth...."
## [10] "Quick.ratio..times."
## [11] "Current.ratio..times."
## [12] "Debt.to.equity.ratio..times."
## [13] "Cash.to.current.liabilities..times."
## [14] "Cash.to.average.cost.of.sales.per.day"
## [15] "EPS"
## [16] "Adjusted.EPS"
## [17] "Default"

plot(final.data$PBDITA.as...of.total.income)
```

We do some more cleaning up of the dataset and select the four financial ratios based on their individual significance in prediciting the variability of "Default"

We take :

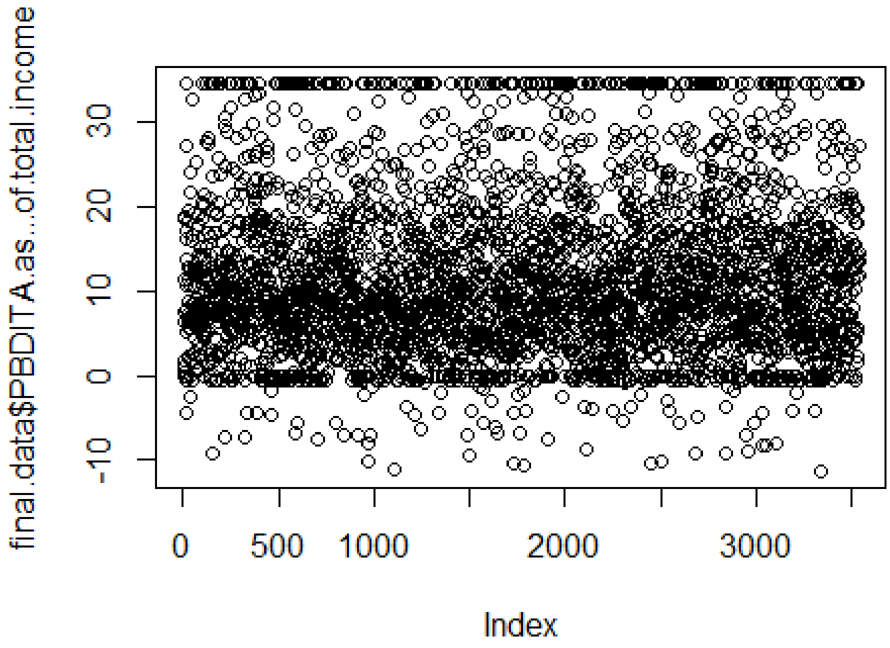
1.**PBDITA as a Percentage of total income** as the **Profitability ratio**,

2.**Current Ratio** as the **Liquidity Ratio**

3.**Debt to Equity Ratio** as the **Leverage ratio**

4.And **Earnings per Share** as the **common size ratio**

Next, we study these variables independently via scatter plots to understand their distribution.



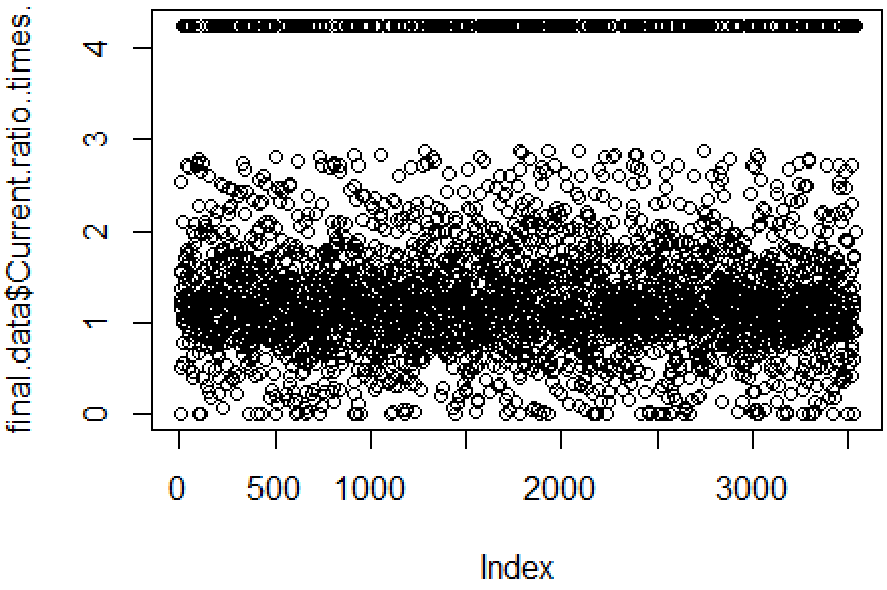
```r
glm(data = final.data, Default~ PBDITA.as...of.total.income, family = binomia
l)
```

```
## 
## Call:  glm(formula = Default ~ PBDITA.as...of.total.income, family = binom
ial, 
##     data = final.data)
## 
## Coefficients:
##                  (Intercept)  PBDITA.as...of.total.income  
##                      -1.7494                      -0.1033  
## 
## Degrees of Freedom: 3540 Total (i.e. Null);  3539 Residual
## Null Deviance:       1771 
## Residual Deviance: 1640  AIC: 1644
```

```r
summary(glm(data = final.data, Default~ PBDITA.as...of.total.income , family
= binomial))
```

```
## 
## Call:
## glm(formula = Default ~ PBDITA.as...of.total.income, family = binomial, 
##     data = final.data)
## 
## Deviance Residuals: 
##     Min       1Q   Median       3Q      Max  
## -0.9422  -0.4271  -0.3338  -0.2246   3.2664  
## 
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)    
## (Intercept)                 -1.74942    0.09099  -19.23   <2e-16 ***
## PBDITA.as...of.total.income -0.10330    0.01029  -10.04   <2e-16 ***
## ---
## 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: 1640.4  on 3539  degrees of freedom
## AIC: 1644.4
## 
## Number of Fisher Scoring iterations: 6
```

```r
plot(final.data$Current.ratio..times.)
```

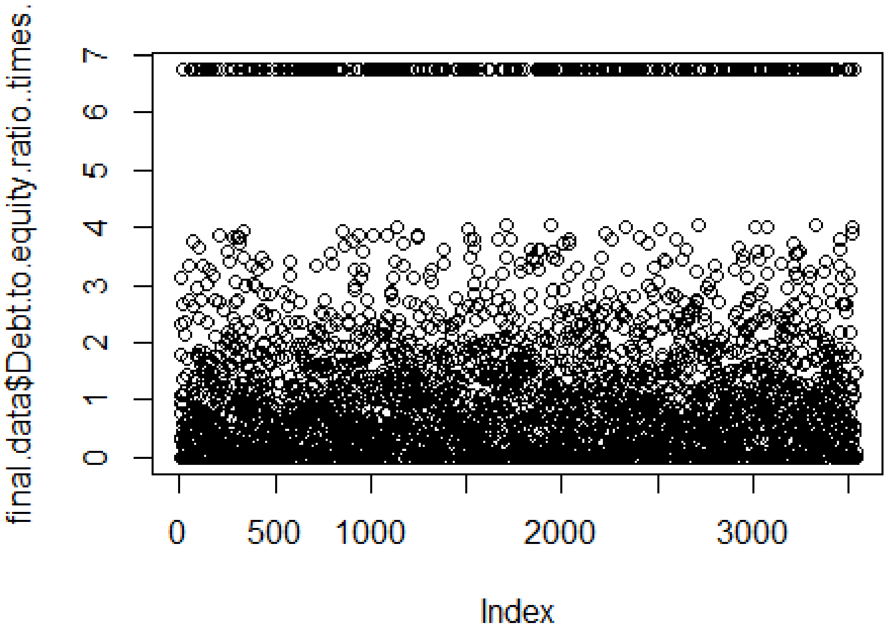


```
glm(data = final.data, Default ~ Current.ratio..times., family = binomial)

##
## Call:  glm(formula = Default ~ Current.ratio..times., family = binomial,
##     data = final.data)
##
## Coefficients:
##           (Intercept)  Current.ratio..times.
##               -1.7776                -0.6468
##
## Degrees of Freedom: 3540 Total (i.e. Null);  3539 Residual
## Null Deviance:       1771
## Residual Deviance: 1716  AIC: 1720
```

```
summary(glm(data = final.data, Default~ Current.ratio..times. , family = bino
mial))

##
## Call:
## glm(formula = Default ~ Current.ratio..times., family = binomial,
##     data = final.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.5589  -0.4132  -0.3775  -0.3039   3.0124
##
## Coefficients:
```

```
##                         Estimate Std. Error z value Pr(>|z|)    
## (Intercept)              -1.7776     0.1324 -13.428  < 2e-16 ***
## Current.ratio..times.    -0.6468     0.1025  -6.309  2.8e-10 ***
## ---
## 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: 1715.8  on 3539  degrees of freedom
## AIC: 1719.8
##
## Number of Fisher Scoring iterations: 6
```

```
plot(final.data$Debt.to.equity.ratio..times.)
```



```
glm(data = final.data, Default~ Debt.to.equity.ratio..times. , family = binom
ial)
```

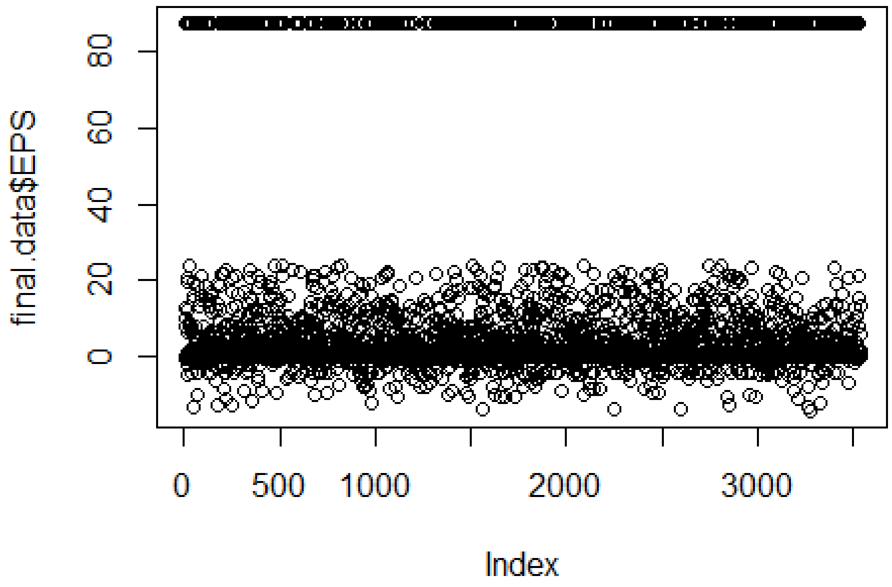
```
##
## Call:  glm(formula = Default ~ Debt.to.equity.ratio..times., family = bino
mial,
##     data = final.data)
##
## Coefficients:
##                  (Intercept)  Debt.to.equity.ratio..times.
##                      -3.8498                        0.5092
```

```
##
## Degrees of Freedom: 3540 Total (i.e. Null);  3539 Residual
## Null Deviance:       1771
## Residual Deviance: 1406  AIC: 1410
```

```
summary(glm(data = final.data, Default~ Debt.to.equity.ratio..times. , family
= binomial))
```

```
##
## Call:
## glm(formula = Default ~ Debt.to.equity.ratio..times., family = binomial,
##     data = final.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0078  -0.2989  -0.2417  -0.2105   2.7824
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -3.84976    0.11855  -32.47   <2e-16 ***
## Debt.to.equity.ratio..times.  0.50916    0.02652   19.20   <2e-16 ***
## ---
## 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: 1406.1  on 3539  degrees of freedom
## AIC: 1410.1
##
## Number of Fisher Scoring iterations: 6
```

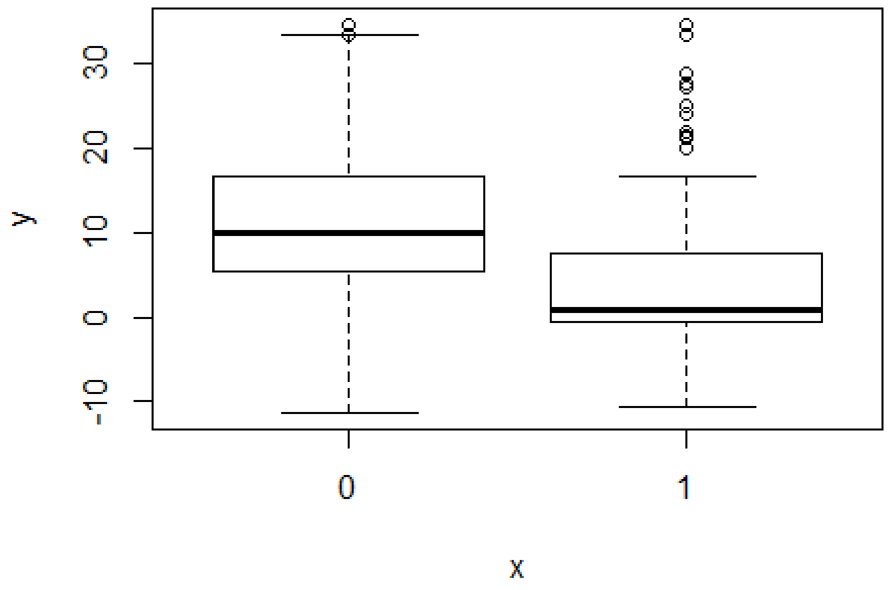
```
plot(final.data$EPS)
```



```
glm(data = final.data, Default~ EPS , family = binomial)

##
## Call:  glm(formula = Default ~ EPS, family = binomial, data = final.data)
##
## Coefficients:
## (Intercept)          EPS
##     -2.2901      -0.2193
##
## Degrees of Freedom: 3540 Total (i.e. Null);  3539 Residual
## Null Deviance:       1771
## Residual Deviance: 1515  AIC: 1519

summary(glm(data = final.data, Default~ EPS , family = binomial))

##
## Call:
## glm(formula = Default ~ EPS, family = binomial, data = final.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4119  -0.4392  -0.3341  -0.0829   6.5607
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.29012    0.07038  -32.54   <2e-16 ***
## EPS         -0.21926    0.01863  -11.77   <2e-16 ***
## ---
## 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: 1515.1  on 3539  degrees of freedom
## AIC: 1519.1
##
## Number of Fisher Scoring iterations: 8
```

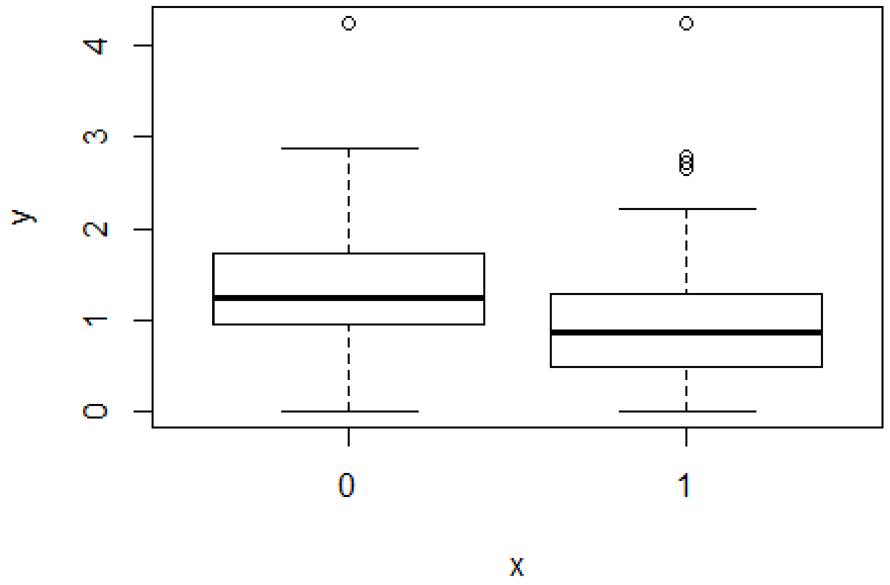
```
plot(final.data$Default,final.data$PBDITA.as...of.total.income)
```

3.5. Bi-variate Analysis



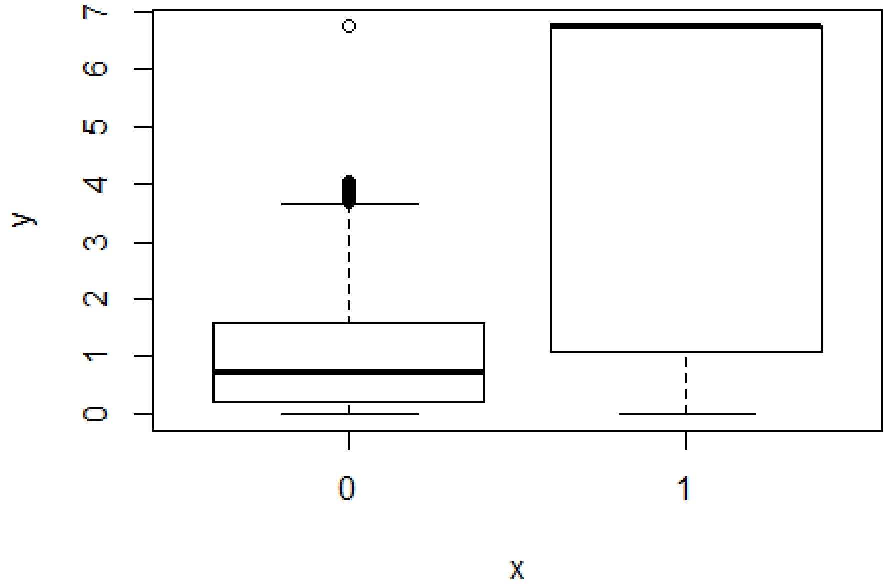
These relations between the x and y variables clearly depicts how the changes in one variable can affect the dependent variable. We see that entities having a higher financial ratio tend to default less and vice versa.

```
plot(final.data$Default,final.data$Current.ratio..times.)
```



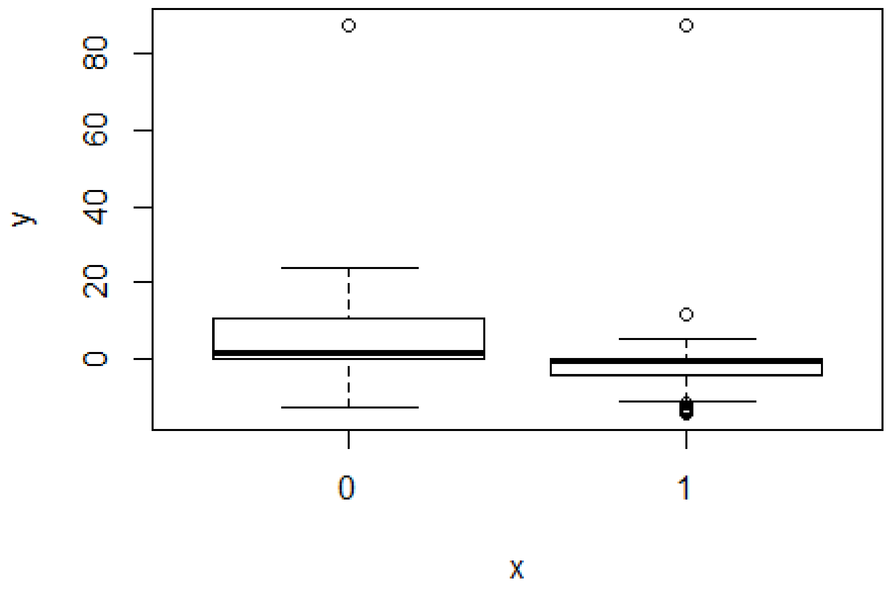
Current Ratio is the value of current assets per current liabilities. Understandably, this is also inversely related with “Default”. However, we can see that in one instance a default had occurred in spite of having a high current ratio.

```
plot(final.data$Default,final.data$Debt.to.equity.ratio..times.)
```



Debt to Equity ratio, which determines the total outstanding liabilities to equity share of the company is positively related to "Default" since, greater the liability higher the chance of default. Here also we see that an entity had not defaulted even though its debt to equity ratio was high.

```
plot(final.data$ Default,final.data$EPS)
```

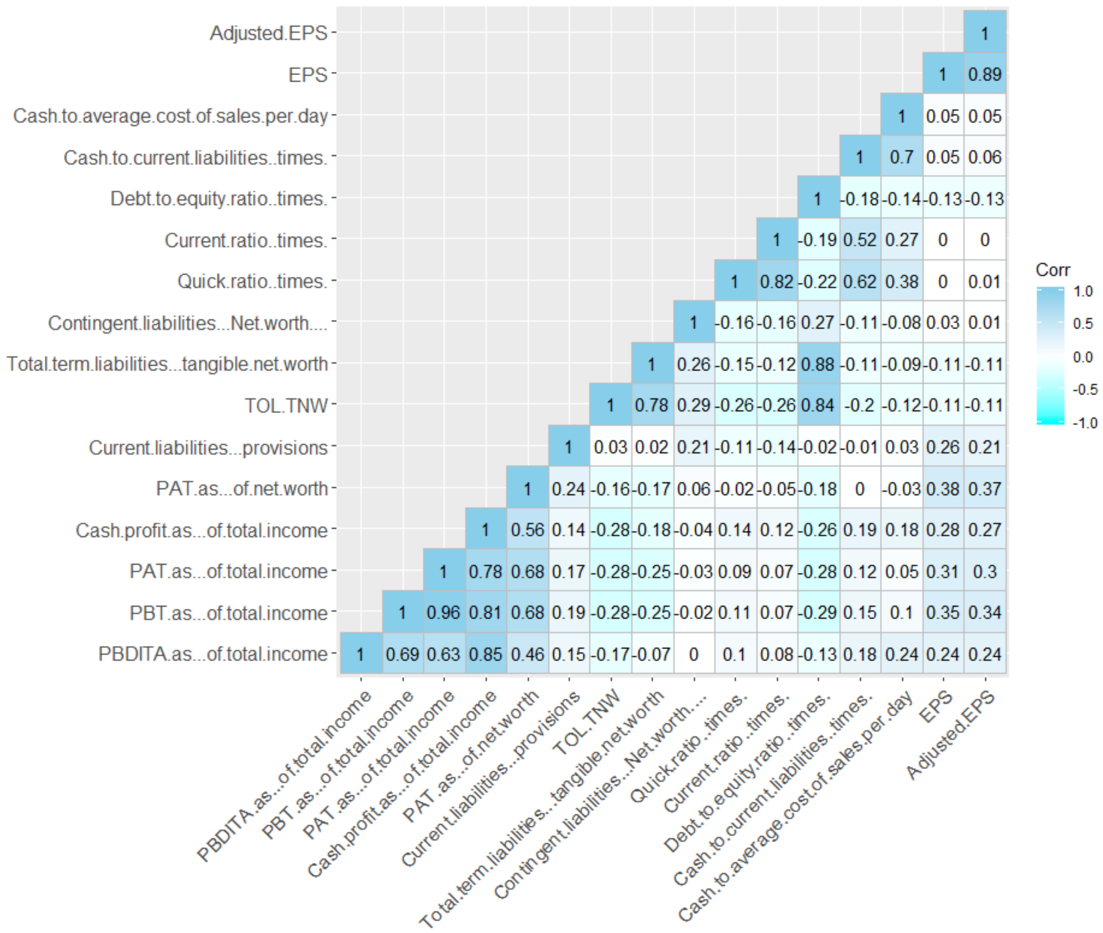


EPS or Earnings per Share also has an inverse proportionality with "Default". Lower the EPS, higher the chances of default. Here also we see a couple of exceptions.

3.6. Checking for Multicollinearity

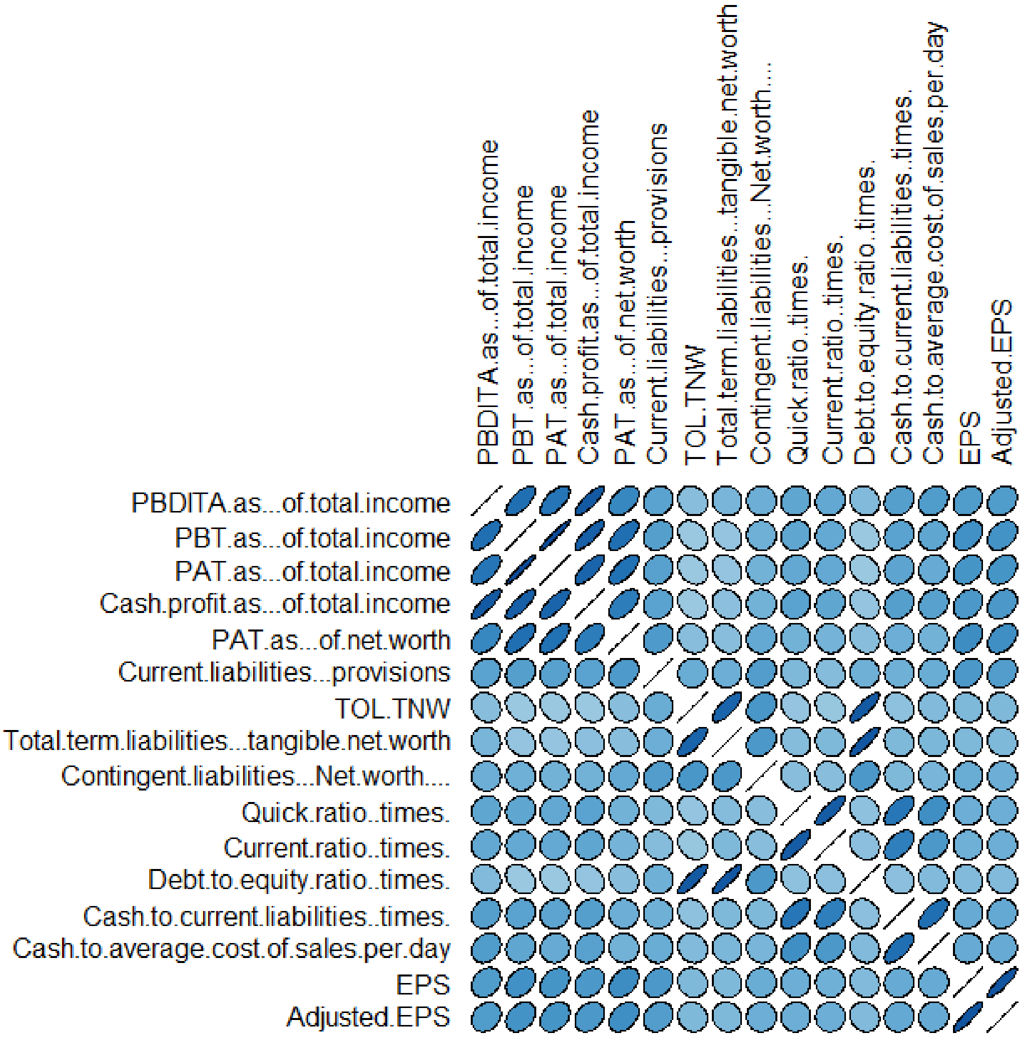
```
corr.matrix = round(cor(final.data[,-17]),3)
corr.matrix

##                                             PBDITA.as...of.total.income
## PBDITA.as...of.total.income                                       1.000
## PBT.as...of.total.income                                          0.692
## PAT.as...of.total.income                                          0.637
## Cash.profit.as...of.total.income                                  0.855
## PAT.as...of.net.worth                                             0.465
## Current.liabilities...provisions                                  0.148
## TOL.TNW                                                          -0.169
## Total.term.liabilities...tangible.net.worth                      -0.076
## Contingent.liabilities...Net.worth....                            0.001
## Quick.ratio..times.                                               0.103
## Current.ratio..times.                                             0.088
## Debt.to.equity.ratio..times.                                     -0.130
## Cash.to.current.liabilities..times.                               0.185
## Cash.to.average.cost.of.sales.per.day                             0.231
## EPS                                                               0.240
## Adjusted.EPS                                                      0.238
##                                              PBT.as...of.total.income
## PBDITA.as...of.total.income                                     0.692
## PBT.as...of.total.income                                        1.000
## PAT.as...of.total.income                                        0.958
## Cash.profit.as...of.total.income                                0.809
## PAT.as...of.net.worth                                           0.683
## Current.liabilities...provisions                                0.192
## TOL.TNW                                                        -0.283
## Total.term.liabilities...tangible.net.worth                    -0.248
## Contingent.liabilities...Net.worth....                         -0.018
## Quick.ratio..times.                                             0.098
## Current.ratio..times.                                           0.076
## Debt.to.equity.ratio..times.                                   -0.292
## Cash.to.current.liabilities..times.                             0.154
## Cash.to.average.cost.of.sales.per.day                           0.099
## EPS                                                             0.345
## Adjusted.EPS                                                    0.337
##                                              PAT.as...of.total.income
## PBDITA.as...of.total.income                                     0.637
## PBT.as...of.total.income                                        0.958
## PAT.as...of.total.income                                        1.000
## Cash.profit.as...of.total.income                                0.775
## PAT.as...of.net.worth                                           0.678
## Current.liabilities...provisions                                0.168
## TOL.TNW                                                        -0.288
## Total.term.liabilities...tangible.net.worth                    -0.256
## Contingent.liabilities...Net.worth....                         -0.027
## Quick.ratio..times.                                             0.093
## Current.ratio..times.                                           0.076
## Debt.to.equity.ratio..times.                                   -0.290
## Cash.to.current.liabilities..times.                             0.126
## Cash.to.average.cost.of.sales.per.day                           0.060
## EPS                                                             0.313
## Quick.ratio..times.                          0.006        0.014
## Current.ratio..times.                       -0.001        0.004
## Debt.to.equity.ratio..times.                -0.133       -0.133
## Cash.to.current.liabilities..times.          0.055        0.065
## Cash.to.average.cost.of.sales.per.day        0.046        0.049
## EPS                                          1.000        0.888
## Adjusted.EPS                                 0.888        1.000
```

```
ggcorrplot(corr.matrix, type = "lower", ggtheme = ggplot2::theme_gray,
          show.legend = TRUE, show.diag = TRUE, colors = c("cyan","white","s
ky blue"),
          lab = TRUE)
```



We use some plots to visually identify intercorelated variables. Since some of the variables are ratios of other variables occurrence of multicollinearity is evident in the dataset. Undoubtedly, we can see that the similar variables are correlated. These variables are: PBDITA.as....of.total.income, PBT.as....of.total.income, PAT.as....of.total.income, Cash.profit.as....of.total.income and PAT.as....of.net.worth. We only keep PBDITA.as....of.total.income from these five. Next, between TOL.TNW, Debt.to.equity.ratio...times and Total.term.liabilities....tangible.net.worth we take only Debt.to.equity.ratio...times.

```
my_colors = brewer.pal(7, "Blues")
my_colors = colorRampPalette(my_colors)(100)
plotcorr(corr.matrix , col=my_colors[corr.matrix*50+50] , mar=c(1,1,1,1), )
```



Between Quick.ratio…times, Current.ratio…times, Cash.to.current.liabilities…times and Cash.to.avergae.cost.of.sales.per.day we take only Current.ratio….times. Between EPS and Adjusted EPS only EPS is taken.

```
test.model = glm(final.data$Default ~ PBDITA.as...of.total.income + PBT.as...
of.total.income  + PAT.as...of.total.income + Cash.profit.as...of.total.incom
e + PAT.as...of.net.worth + Current.liabilities...provisions + TOL.TNW+ Total
.term.liabilities...tangible.net.worth + Contingent.liabilities...Net.worth..
.. + Quick.ratio..times. +  Current.ratio..times. + Debt.to.equity.ratio..tim
es. + Cash.to.current.liabilities..times. + Cash.to.average.cost.of.sales.per
.day + EPS + Adjusted.EPS, family = binomial)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
vif(test.model)
```

```
##                 PBDITA.as...of.total.income
##                                1.171362e+01
##                    PBT.as...of.total.income
##                                2.115600e+02
##                    PAT.as...of.total.income
##                                1.929854e+02
##            Cash.profit.as...of.total.income
##                                1.790042e+01
##                       PAT.as...of.net.worth
##                                1.130693e+00
##            Current.liabilities...provisions
##                                1.038799e+00
##                                     TOL.TNW
##                                1.268289e+01
## Total.term.liabilities...tangible.net.worth
##                                1.074438e+01
##      Contingent.liabilities...Net.worth....
##                                1.260240e+00
##                         Quick.ratio..times.
##                                1.357606e+02
##                       Current.ratio..times.
##                                9.181542e+01
##                Debt.to.equity.ratio..times.
##                                2.093233e+00
##         Cash.to.current.liabilities..times.
##                                8.000358e+01
##       Cash.to.average.cost.of.sales.per.day
##                                1.173887e+00
##                                         EPS
##                                2.053526e+06
##                                Adjusted.EPS
##                                2.053524e+06
```

```
final.data = final.data[,-c(2,3,4,5,7,8,9,10,13,16)]
```

We do a vif test to check for any more evidence of multicollinearity. We can say that we have removed multicollinarity from the data but reducing intercorrelation between the variablres.

4.Statistical Analysis

4.1. Logistic Regression

```
summary(glm(data = final.data, Default ~ PBDITA.as...of.total.income + Curren
t.ratio..times. + Debt.to.equity.ratio..times. + EPS , family = binomial))

##
## Call:
## glm(formula = Default ~ PBDITA.as...of.total.income + Current.ratio..times
. +
##     Debt.to.equity.ratio..times. + EPS, family = binomial, data = final.da
ta)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9657  -0.3209  -0.2091  -0.0816   5.4012
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -2.49828    0.17982 -13.893  < 2e-16 ***
## PBDITA.as...of.total.income  -0.05797    0.01023  -5.669 1.44e-08 ***
## Current.ratio..times.        -0.31009    0.09258  -3.350  0.00081 ***
## Debt.to.equity.ratio..times.  0.42360    0.02883  14.691  < 2e-16 ***
## EPS                          -0.12673    0.01999  -6.341 2.28e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1771  on 3540  degrees of freedom
## Residual deviance: 1221  on 3536  degrees of freedom
## AIC: 1231
##
## Number of Fisher Scoring iterations: 9

model1 = glm(data = final.data, Default ~ PBDITA.as...of.total.income + Curre
nt.ratio..times. + Debt.to.equity.ratio..times. + EPS , family = binomial)


prediction = ifelse(model1$fitted.values > 0.065,1,0)

table(model1$y,prediction)

##    prediction
##        0    1
##   0 2688  610
##   1   43  200

prediction1 = predict(model1, newdata = new.test.data)
```

```
cmLR = table(test.data$Default...1, prediction1 > 0.1)
cmLR

##
##     FALSE TRUE
##   0   634   27
##   1    21   33

sum(diag(cmLR))/sum(cmLR)

## [1] 0.9328671
```

Logistic regression model gives us an accuracy of 93.29%. However, sensitivity is only 61.11%. Since we are designing a credit risk model to predict defaulters, we should aim for higher sensitivity. Let us proceed to SMOTE.

4.2. SMOTE

```
set.seed(1000)
balanced.data = SMOTE(Default ~.,perc.over = 500 , final.data , k = 5, perc.u
nder = 900)
table(balanced.data$Default)

##
##     0     1
## 10935  1458

1458/10935

## [1] 0.1333333

model2 = glm(data = balanced.data, Default ~ PBDITA.as...of.total.income + Cu
rrent.ratio..times. + Debt.to.equity.ratio..times. + EPS , family = binomial)
summary(model2)

##
## Call:
## glm(formula = Default ~ PBDITA.as...of.total.income + Current.ratio..times
. +
##     Debt.to.equity.ratio..times. + EPS, family = binomial, data = balanced
.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3758  -0.4107  -0.2501  -0.0472   5.7866
##
## Coefficients:
##                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -1.875053   0.077062 -24.332   <2e-16 ***
## PBDITA.as...of.total.income  -0.059130   0.004507 -13.120   <2e-16 ***
## Current.ratio..times.        -0.346932   0.040615  -8.542   <2e-16 ***
## Debt.to.equity.ratio..times.  0.430573   0.013190  32.643   <2e-16 ***
## EPS                          -0.157668   0.009670 -16.305   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 8977.8  on 12392  degrees of freedom
## Residual deviance: 5800.9  on 12388  degrees of freedom
## AIC: 5810.9
##
## Number of Fisher Scoring iterations: 8
```

```
prediction = ifelse(model2$fitted.values > 0.13,1,0)
table(model2$y,prediction)
```

```
##    prediction
##        0    1
##   0 9431 1504
##   1  285 1173
```

```
prediction2 = predict(model2, newdata = new.test.data)
cmLR = table(test.data$Default...1, prediction2 > 0.1)
cmLR
```

```
##
##     FALSE TRUE
##   0   626   35
##   1    15   39
```

```
sum(diag(cmLR))/sum(cmLR)
```

```
## [1] 0.9300699
```

SMOTE results in a slightly reduced accuracy of 93% but in this case we get a better sensitivity score of 72.22% which we would call an improvement over the previous model.

```
new.test.data$Probability.of.Default = predict(model2, newdata = new.test.dat
a)
new.test.data$Decile.groups = decile(vector = new.test.data$Probability.of.De
fault, decreasing = TRUE )
new.test.data$Default = test.data$Default...1
new.test.data$Default.Prediction = prediction2 > 0.1



output.data = new.test.data[order(new.test.data$Probability.of.Default),]
```

```
View(output.data)

write.csv(output.data, file = "FRA.output.csv")
```

5.Conclusion

We conclude by saying that the logistic regression on SMOTE model performed better only in terms of sensitivity. However, in a default risk model, there should be more weightage towards identifying defaulters over non-defaulters. Since, one default would cause direct loss to the institution giving out the loan; it generally becomes more important to avoid a default than the risk involved in losing potential business. Keeping this in mind, we sort the data and divide into deciles with bucket 1 having the highest chance of default and bucket 10 having the lowest.