Financial Risk Analysis

Submission Date: Apr 26, 2020

Author: Ayush Jain
Mentor: Deepak Gupta

Table of Content

Table of Contents

1 Project Objective

This report is to build an India Credit Risk Model that would help bank on Financial Risk Analysis.

Financial risk analysis is the analysis of evaluation of credit risk attached to the Creditor from the Borrower. It effectively assesses a borrower's ability to repay a loan to a creditor. Investors are at risk of loss of principal or loss of financial award if borrowers fail to pay back a loan or meet the terms of a contract. Credit risk goes hand in hand with a borrower expectation to use future income to pay off an already existing debt.

The analysis of the Credit risk is done based on several Parameters that banks takes into consideration at the time accepting the application.

The four Major factor that bank relies on for the Analysis are:

- Size of the Company

  The size of the company is determined by the Revenue that Company generates, the assets that company owns and the net Worth of the Company.

- Profitability of the Company:

  Profitability of the company is the degree to which a business or activity yields profit or financial gain.

- Leverage of the Company

  The leverage is determined as the ratio of a company's loan capital (debt) to the value of its common stock (equity).

- Liquidity of the Company

  The liquidity of the company is the availability of liquid assets to company. It includes the Cash available, Current Assets etc.

These factors are used to determine the capability of the company to repay it's debts

Using the data provided in the spreadsheet raw-data.xlsx we build the Logistic regression framework to develop the credit default Model and validate that model on validation_data.xlsx.

During this process of Model building we undergoes the step of Data Cleaning and Data Modelling.

We first investigate and Analyze the data to gain the insights and the understand the readiness of the data for modelling.

Once the Data is ready, we build the Model on the Original Data and Validate the model on Validation (Unseen) data.

Original Data contains 52 variables and 3541 observations and Validation dataset have 52 variables and 715 observations.

2 Exploratory Data Analysis – Step by step approach

Exploratory Data Analysis is one of the important phases in the data Analysis in understanding the significance and accuracy of the data. It usually consists of setting up the environment to work in R, loading the data and checking the validity of data loaded.

The typical data exploration activity consists of the following steps:

- Environment Set up and Data Import.
    - Install Necessary Package in R.
    - Setting Up Working Directory.
    - Reading Dataset in R.
    - Treating missing values
    - Performing Univariate Analysis on independent variables.
    - Performing Bi-variate Analysis.
    - Checking for Multicollinearity within the Independent variables.
    - Treating Outliers.
    - Preparing the Data for Model Building.
- Variable Identification

2.1 Environment Set up and Data Import

2.1.1 Deploying necessary Packages in R.

In this section, we will install and invoke the necessary Packages and Libraries that are going to be the part of our work throughout the project. Having all the packages at the same places increases code readability and Understandability.

```
# Deploying the Libraries

library(readxl)
library(mice)
library(corrplot)
library(DMwR)
library(e1071)
library(car)
library(blorr)
library(MASS)
library(caret)
library(psych)
library(tidyverse)
```

2.1.2 Setting Up Working Directory.

Setting a working directory on starting of the R session makes importing and exporting data files and code files easier. Basically, working directory is the location/ folder on the PC where you have the data, codes etc. related to the project. This helps maintain the code readability and avoid unwanted errors.

```
# Setting up the Working Directory.

setwd("D:/Great Learning/Finance and Risk Analytics")
```

2.1.3 Reading Dataset in R.

The given dataset is in .xlsx format. Hence, the command 'read.xslx' from readxl package is used for importing the file.

```
# Reading the Raw data.

raw_data <- read_xlsx("raw-data.xlsx")
```

2.1.4 Performing basic Data checks.

This section of the report checks for the basic steps to ensure that the data is imported properly and also checks the Structure of the dataset and Summary to have the basic understanding of the Data.

We here transform the names of the variables to ensure that they make more sense

```
# Transforming the names of the Variables

colnames(raw_data)<- make.names(colnames(raw_data))
```

The dataset is provided with the variable 'Net worth next year', we assume if the value of net worth next year is less then 0, the company is most likely to be default and is at high risk. Using this we create a new factor variable Default which will have value of 1 if the company is likely to be default else 0.

```
# Creating Default variable basis of Net Worth Next Year

raw_data$Default <- ifelse(raw_data$Networth.Next.Year <0,1,0)
```

Here we will perform a few basic checks to ensure that the data is read correctly and understand the basics of the data.

- Reading the first 10 observations

```r
#Perfroming Exploratory Data Analysis
head(raw_data,n=10)
```

```
## # A tibble: 10 x 53
##      Num Networth.Next.Y~ Total.assets Net.worth Total.income
##    <dbl>            <dbl>        <dbl>     <dbl>        <dbl>
##  1     1           8891.       17512.     7093.       24965.
##  2     2            394.         941       352.        1527.
##  3     3             92.2        233.      101.         477.
##  4     4              2.7          2.7       2.7          NA
##  5     5            109          478.      108.        1580.
##  6     6            689.        2434.      676.        2649.
##  7     7            246          327.      245.          NA
##  8     8             13.7         80        12.7        154.
##  9     9            292.         574.      239.         583.
## 10    10             -7.3         88.6      19.6         83.4
## # ... with 48 more variables: Change.in.stock <dbl>, Total.expenses <dbl>,
## #   Profit.after.tax <dbl>, PBDITA <dbl>, PBT <dbl>, Cash.profit <dbl>,
## #   PBDITA.as...of.total.income <dbl>, PBT.as...of.total.income <dbl>,
## #   PAT.as...of.total.income <dbl>,
## #   Cash.profit.as...of.total.income <dbl>, PAT.as...of.net.worth <dbl>,
## #   Sales <dbl>, Income.from.financial.services <dbl>, Other.income <dbl>,
## #   Total.capital <dbl>, Reserves.and.funds <dbl>,
## #   Deposits..accepted.by.commercial.banks. <lgl>, Borrowings <dbl>,
## #   Current.liabilities...provisions <dbl>, Deferred.tax.liability <dbl>,
## #   Shareholders.funds <dbl>, Cumulative.retained.profits <dbl>,
## #   Capital.employed <dbl>, TOL.TNW <dbl>,
## #   Total.term.liabilities...tangible.net.worth <dbl>,
## #   Contingent.liabilities...Net.worth.... <dbl>,
## #   Contingent.liabilities <dbl>, Net.fixed.assets <dbl>,
## #   Investments <dbl>, Current.assets <dbl>, Net.working.capital <dbl>,
## #   Quick.ratio..times. <dbl>, Current.ratio..times. <dbl>,
## #   Debt.to.equity.ratio..times. <dbl>,
## #   Cash.to.current.liabilities..times. <dbl>,
## #   Cash.to.average.cost.of.sales.per.day <dbl>, Creditors.turnover <chr>,
## #   Debtors.turnover <chr>, Finished.goods.turnover <chr>,
## #   WIP.turnover <chr>, Raw.material.turnover <chr>,
## #   Shares.outstanding <chr>, Equity.face.value <chr>, EPS <dbl>,
## #   Adjusted.EPS <dbl>, Total.liabilities <dbl>, PE.on.BSE <chr>,
## #   Default <dbl>
```

- Checking the dimensions of the data

```r
dim(raw_data)
```

```
## [1] 3541   53
```

- Understanding the structure of the data

```
str(raw_data)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    3541 obs. of  53 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
1580 ...
##  $ Change.in.stock                            : num  235.8 42.7 -5.2 NA -
17 ...
##  $ Total.expenses                             : num  23658 1455 479 NA
1558 ...
##  $ 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
11.9 ...
##  $ 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
0.35 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
1575 ...
##  $ 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
193.1 ...
##  $ Current.liabilities...provisions           : num  9965.9 210 96.8 NA
```

```
112.8 ...
##  $ 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
300.7 ...
##  $ 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
37.6 ...
##  $ 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
1.41 0.48 NA 0.54 0.59 0.39 ...
##  $ Current.ratio..times.                       : num  1.37 1.56 1.55 NA
2.54 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"
"9.8000000000000007" "5.28" "0" ...
##  $ Debtors.turnover                            : chr  "3.85" "5.7" "5.07"
"0" ...
##  $ Finished.goods.turnover                     : chr  "200.55" "14.21"
"9.24" 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
```

We observe that there are a few variables that have character datatype but contains the value as numeric, we will further be converting these to numeric so as to perform any calculations.

- Analyzing the basics of dataset

```
summary(raw_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            Default
##  Length:3541        Min.   :0.00000
##  Class :character   1st Qu.:0.00000
##  Mode  :character   Median :0.00000
##                     Mean   :0.06608
##                     3rd Qu.:0.00000
##                     Max.   :1.00000
##
```

From the summary performed we observed the below points.

- In the data, we have a massive number of missing vales that are required to be addressed.
- We encounter a great number of outliers that are available in data and will affect the model accuracy if not treated.
- Variables like “Num” and “Deposits” can be dropped as they makes no sense in the data.

2.1.5 Treating the missing values.

Missing Values: These are the null values in the dataset that were not provided by the customers and needs to be treated to get the proper accuracy of the Model.

Missing values can be treated by either dropping them or imputing them using various algorithms.

In this case we will be using mice “Multivariate Imputation via Chained Equations” using method ‘pmm’ “Predictive Mean Matching” method.

Before treating the missing values, we drop the variables where we have missing values greater then 25%.

```
# Removing the unnecessary variables and variables with missing value greater
then 25%
names(raw_data)

data <- raw_data[,-c(1,2,18,19,22,25,32,34,42:47)]

# Converting datatypes of the variables
data$Equity.face.value <- as.numeric(data$Equity.face.value)

data$PE.on.BSE <- as.numeric(data$PE.on.BSE)

data$Default <- as.factor(data$Default)

# Imputing the missing Values.

data1 <- mice(data,m=5,maxit=10,meth='pmm',seed=500)
```

```
data_new <- complete(data1,1)
colSums(is.na(data_new))
```

```
data_new <- na.omit(data_new)
dim(data_new)

## [1] 3282   39
```

2.1.6 Performing Univariate Analysis.

Univariate analysis is perhaps the simplest form of statistical analysis. Like other forms of statistics, it can be inferential or descriptive. The key fact is that only one variable is involved.

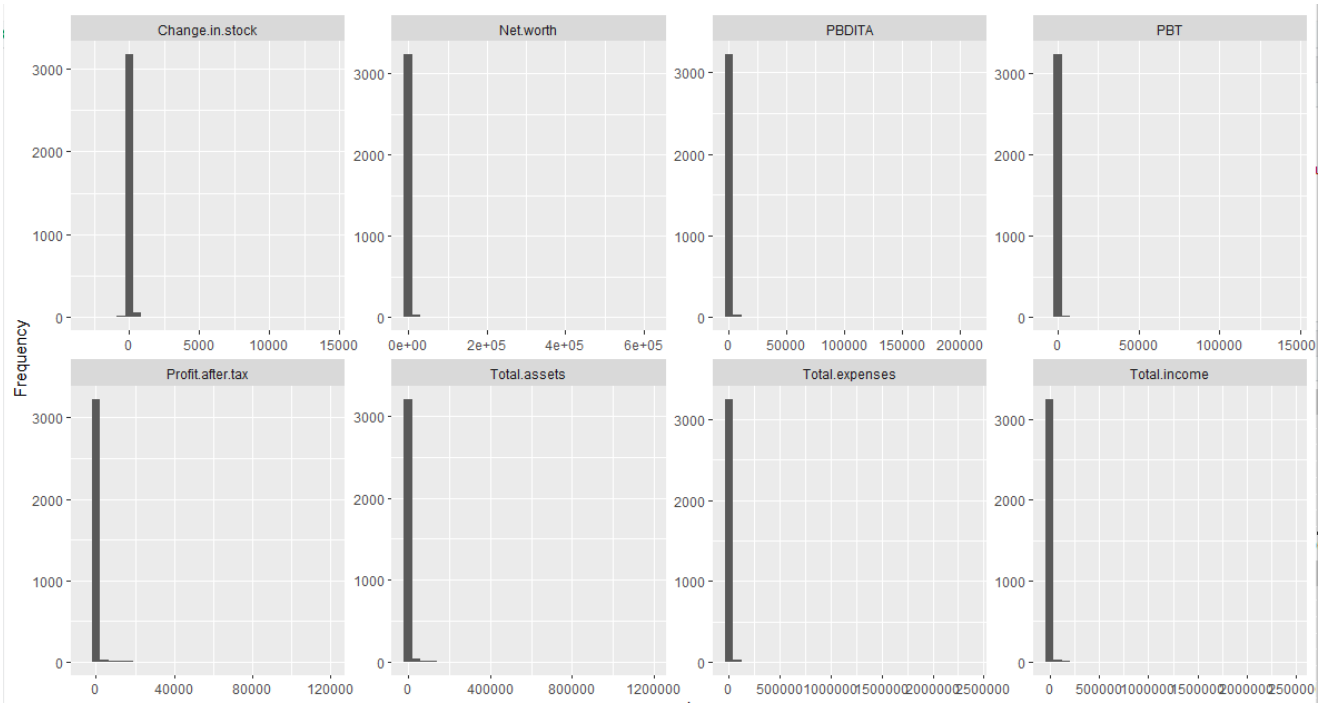
For Numeric variables, default plot is histogram and boxplot while for Categorical variables it is Bar plot.

Histogram: A histogram is an accurate representation of the distribution of numerical data. It is an estimate of the probability distribution of a continuous variable.

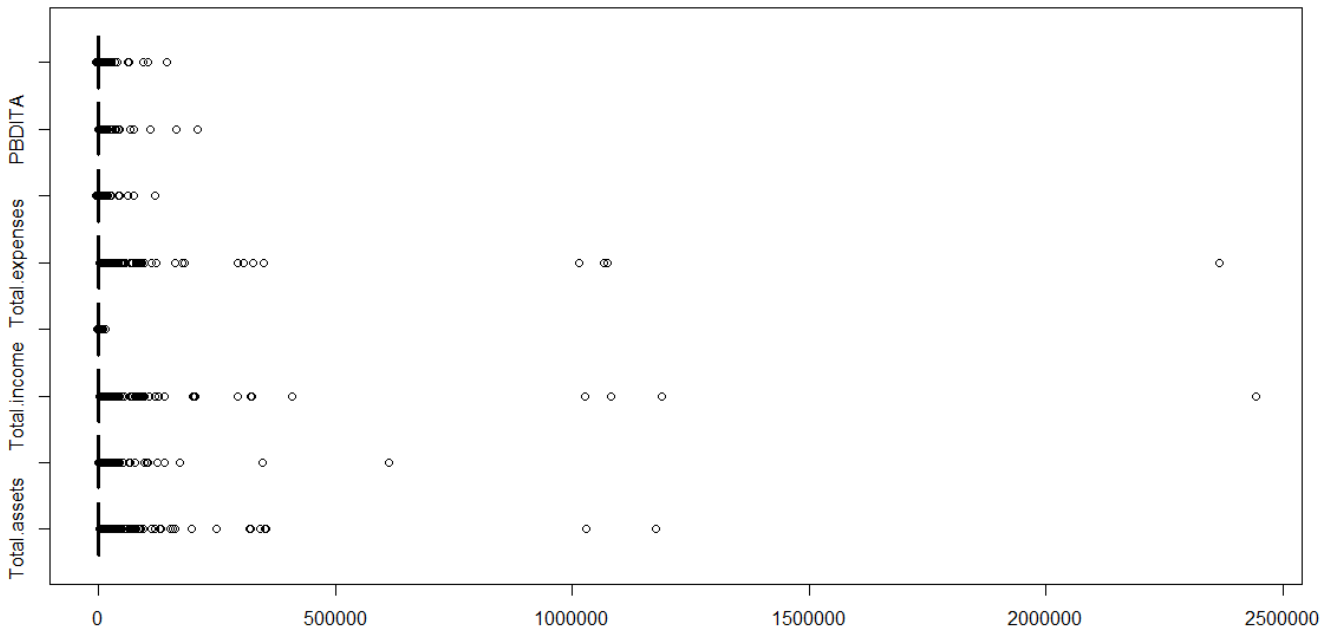
Boxplot: A box plot or boxplot is a method for graphically depicting groups of numerical data through their quartiles. Outliers may be plotted as individual points.

In the problem given, we will be using the above two plotting functions to perform the Univariate analysis on the dataset and identify any outliners present in the data.

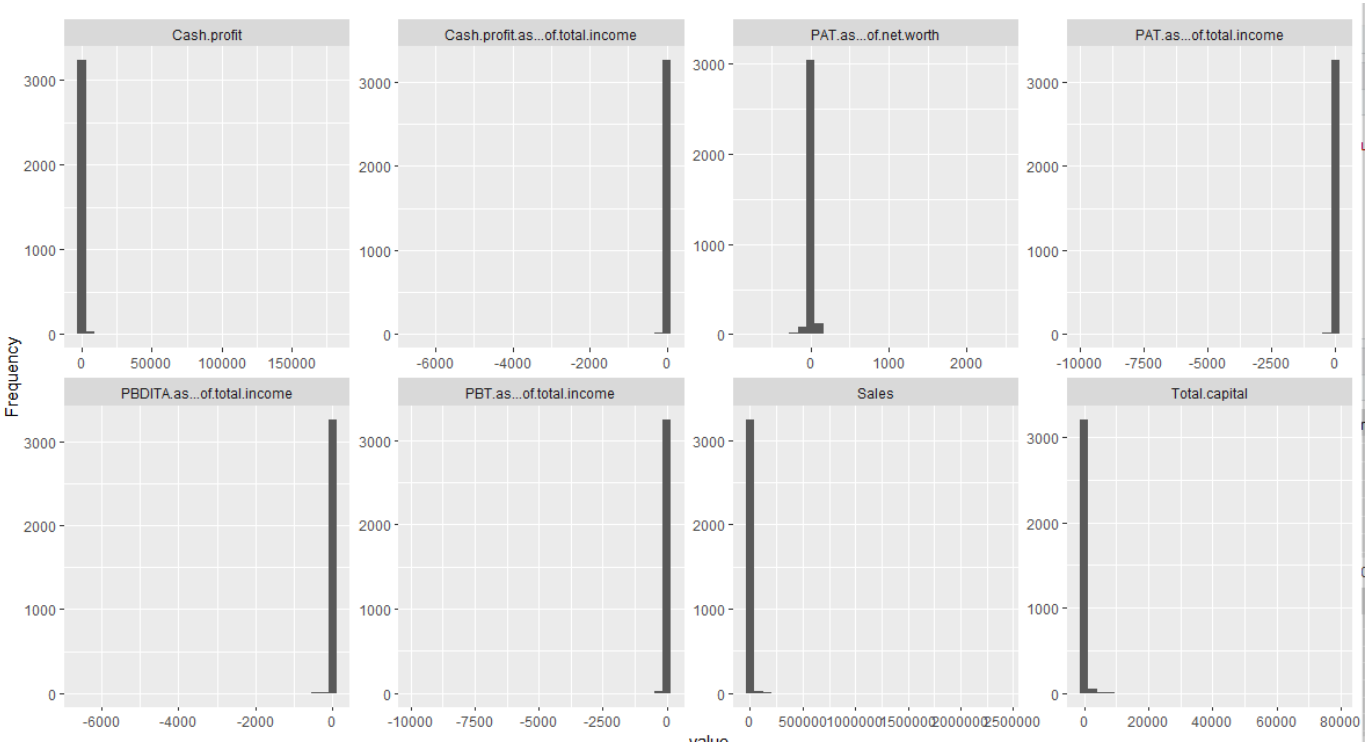
```
#Perfroming Univariate Analysis
dev.off()

## null device
##           1

plot_histogram(data_new[1:8])
```



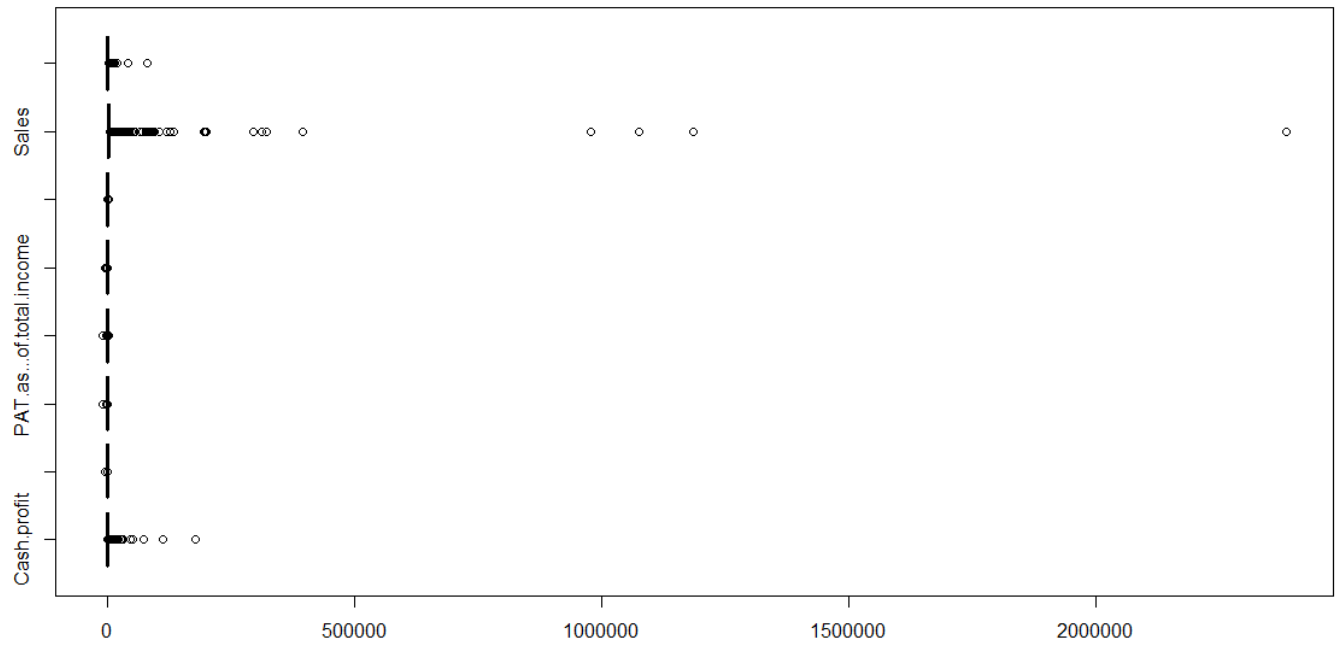

```
boxplot(data_new[1:8],horizontal = TRUE)
```



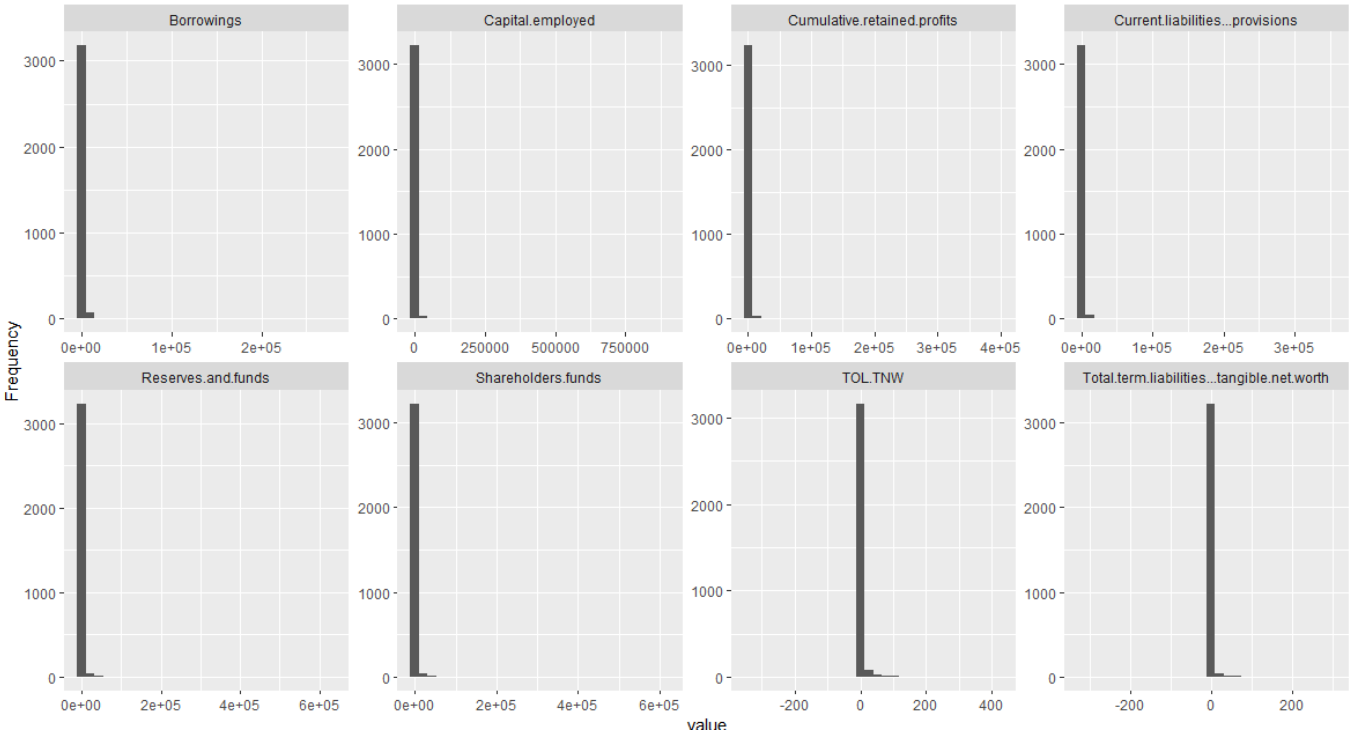
```
plot_histogram(data_new[9:16])
```



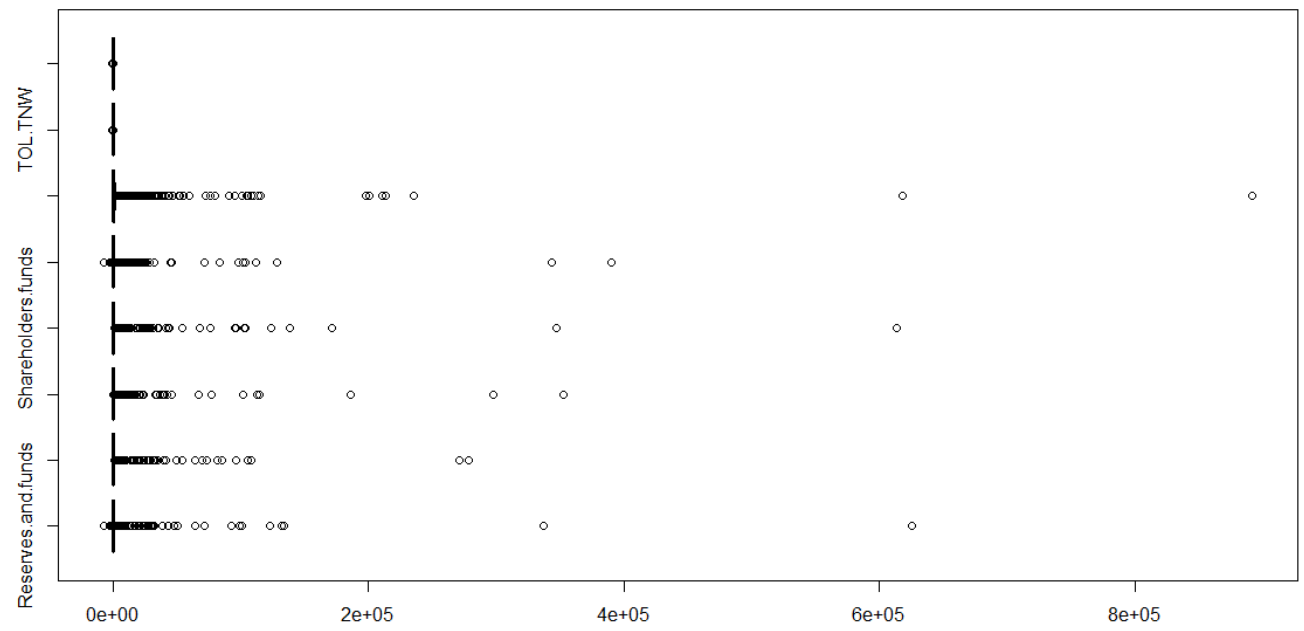

```
boxplot(data_new[9:16],horizontal = TRUE)
```



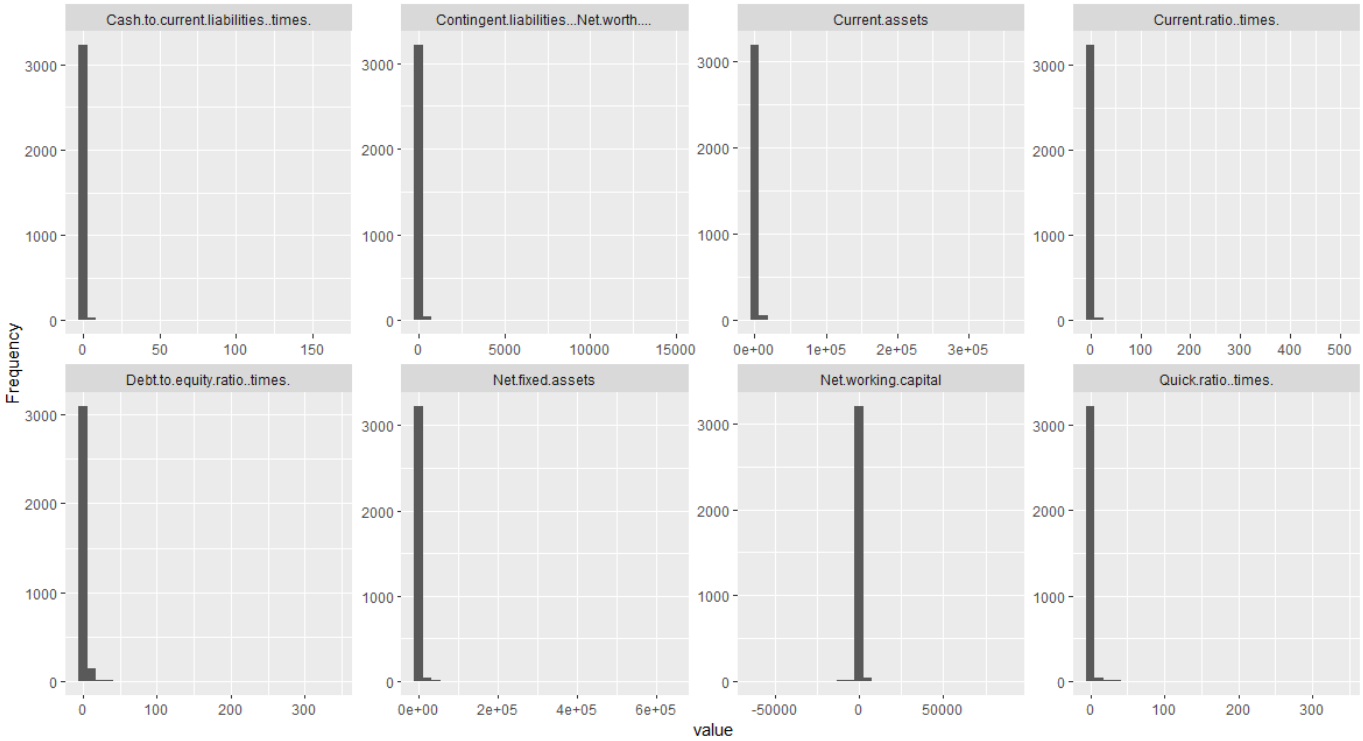
```
plot_histogram(data_new[17:24])
```



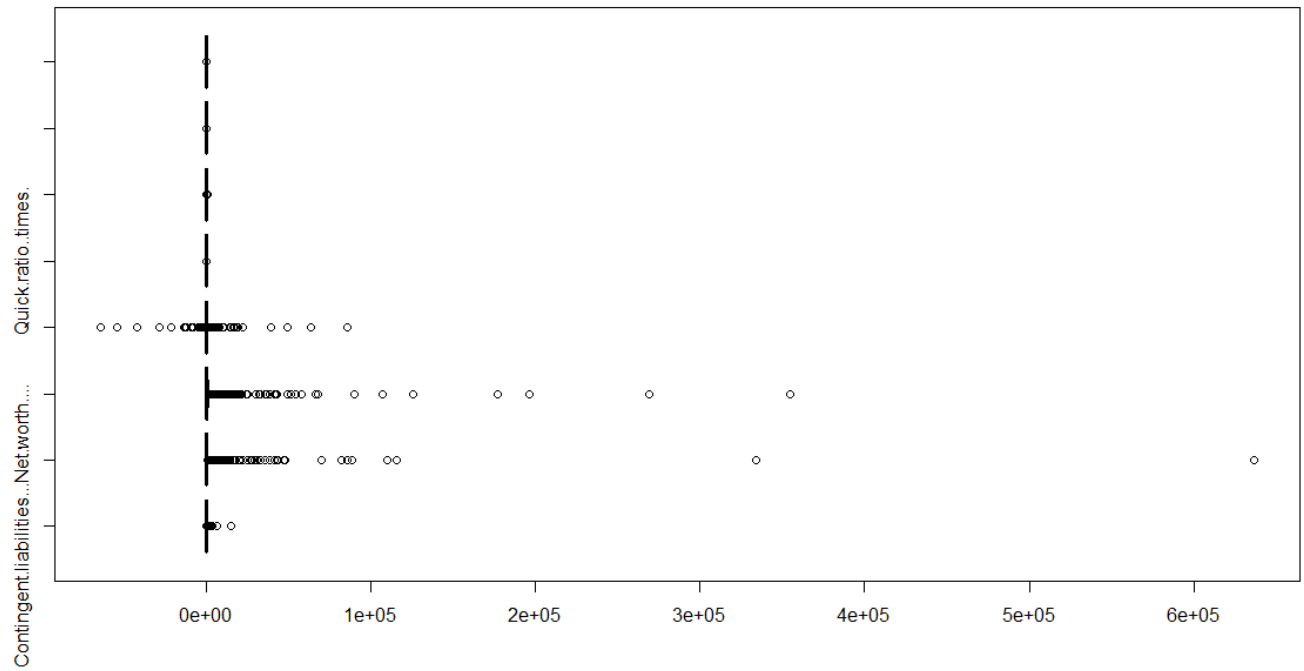

```
boxplot(data_new[17:24],horizontal = TRUE)
```



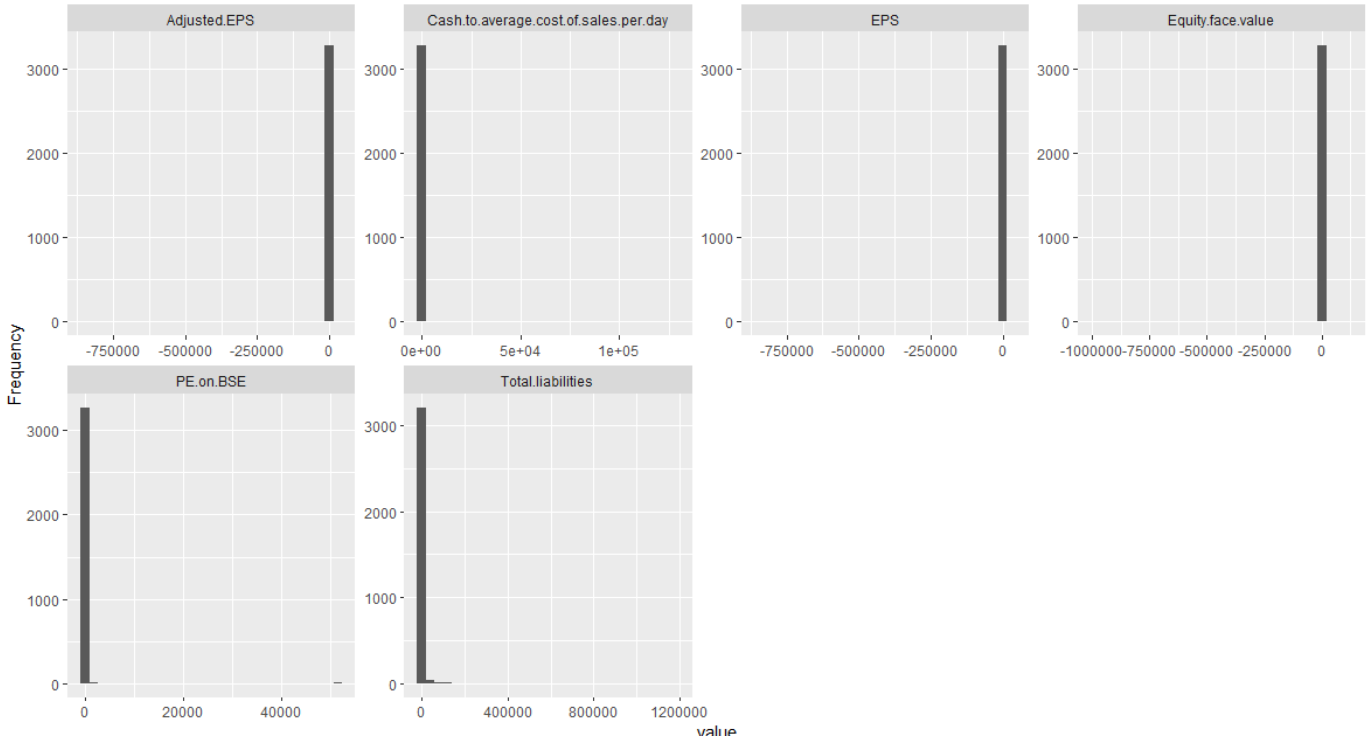
```
plot_histogram(data_new[25:32])
```



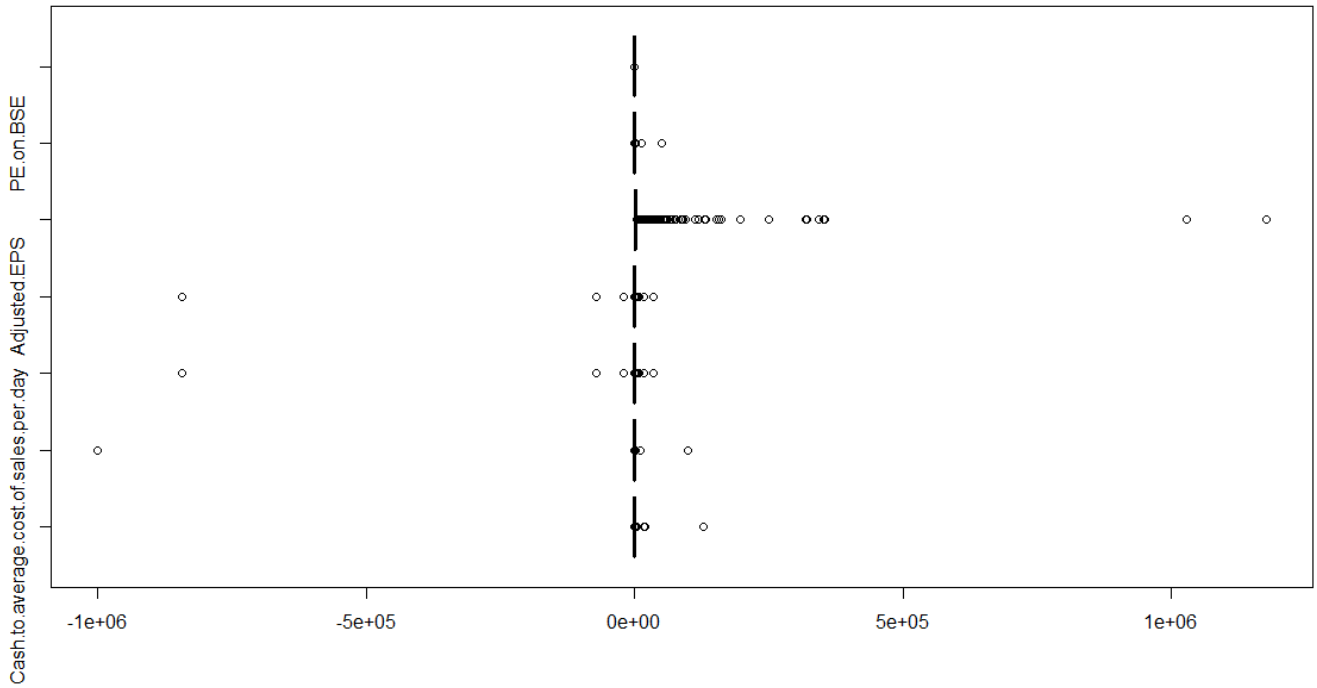

```
boxplot(data_new[25:32],horizontal = TRUE)
```



```
plot_histogram(data_new[33:39])
```




```
boxplot(data_new[33:39],horizontal = TRUE)
```



From the above plots we can conclude the below:

- From the Histogram and Box plot we can observe that there are certain higher values available in most of the variables.

- These values are well above the respective ‘mean’ in those categories.

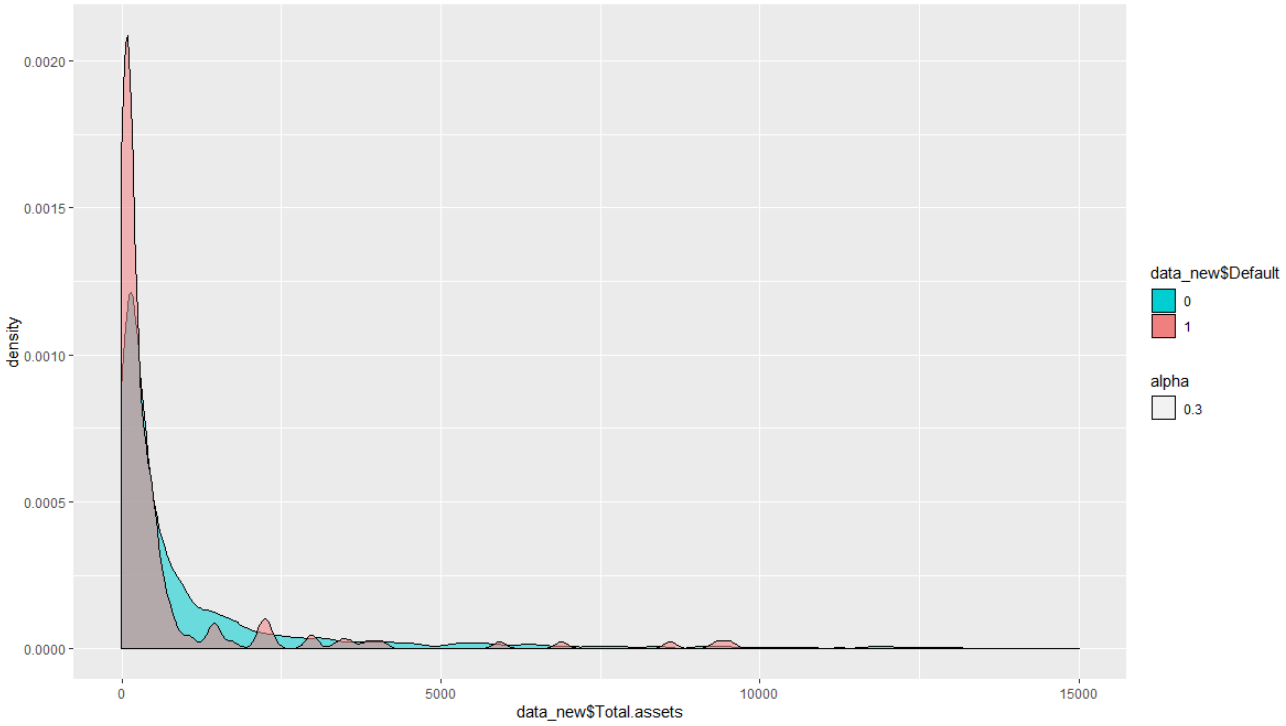
2.1.7 Performing Bi-Variate Analysis

Multivariate analysis is a set of techniques used for analysis of data sets that contain more than one variable, and the techniques are especially valuable when working with correlated variables. The techniques provide an empirical method for information extraction, regression, or classification.

For Multivariate analysis, the default plot is the Scatter Plot or Density Plot. We will be plotting the correlation between the different variables with Churn to understand the relation between the dependent variable Churn with the Independent variables.

- Density Plot of Total Assets with respect to Default.

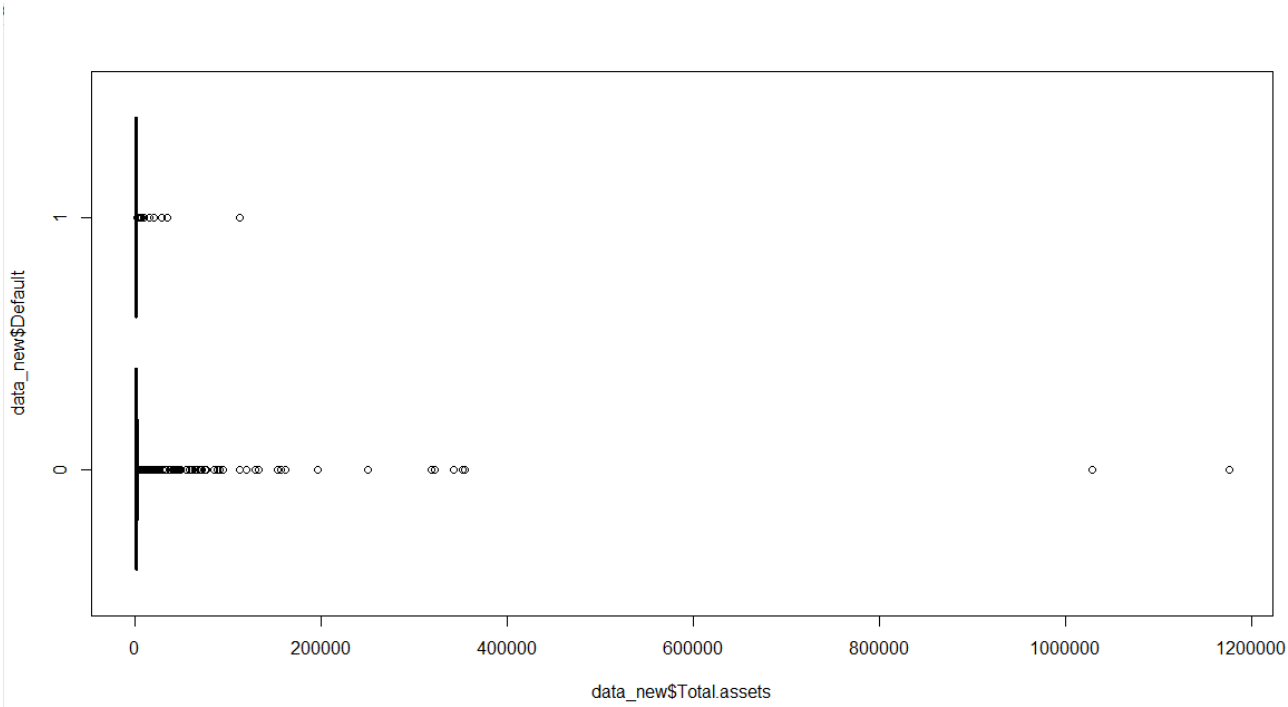
```
#Performing Bi-Variate Analaysis

ggplot(data_new, aes(x=data_new$Total.assets)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,15000)

## Warning: Removed 100 rows containing non-finite values (stat_density).
```



The above plot is a density plot for Total Assets of the company with respect to the Default ratio. From this we can infer that Majority of the values lies under 5000. Also, there are extreme values present in the data would require our attention to ensure the accuracy of the Model. We can also see that the most the defaulters are those with lesser Total Asset Size.

- Box plot of Total Asset with respect to Default

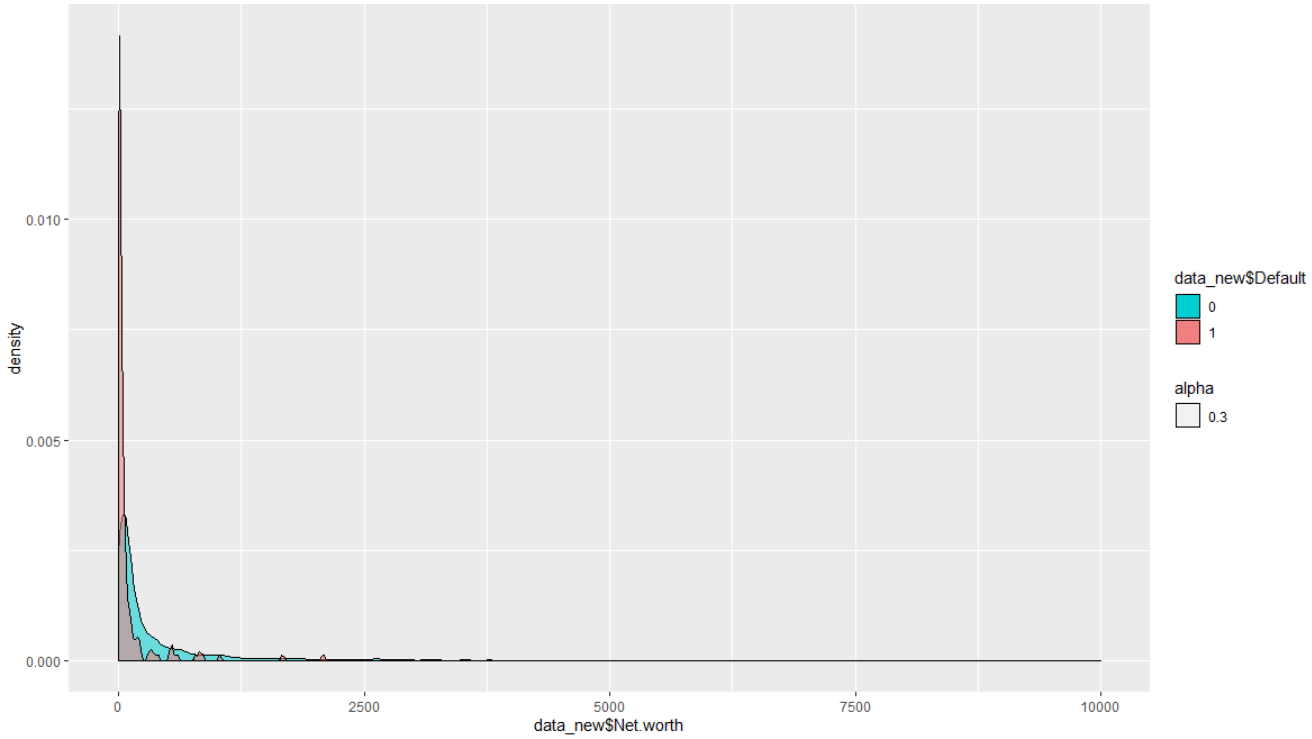
```
boxplot(data_new$Total.assets~data_new$Default, horizontal = TRUE)
```



This plot tells us the presence of extreme in Total Assets based on Default ratio.

- Density Plot of Net Worth with respect to Default.

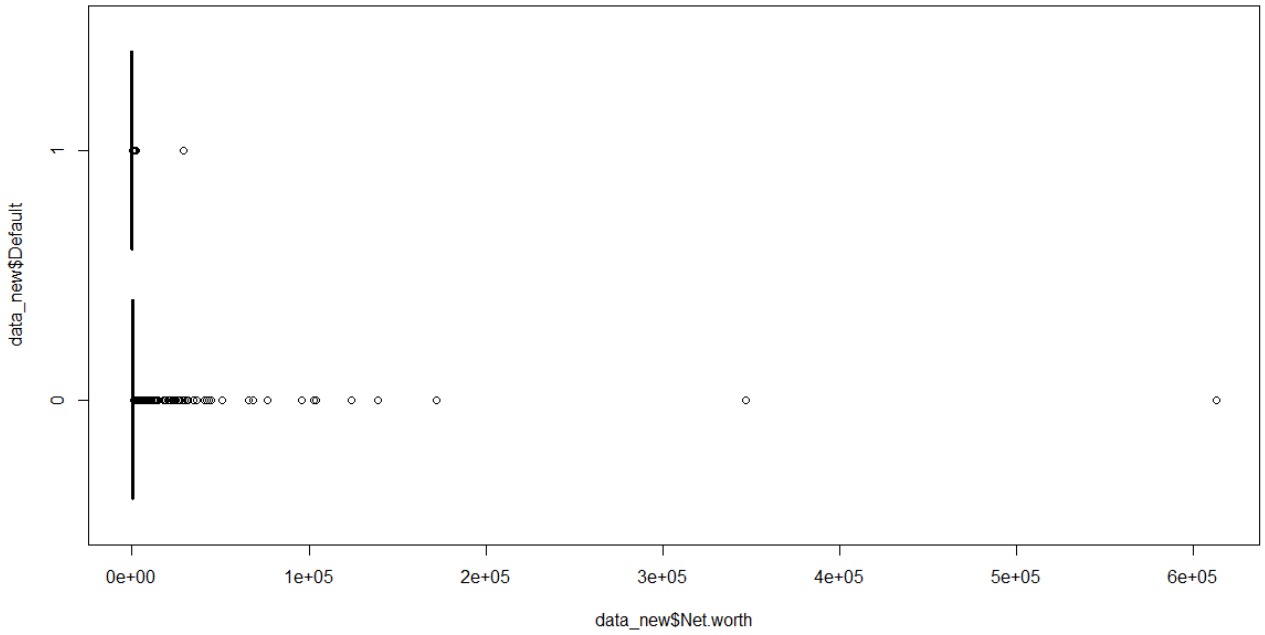
```
ggplot(data_new, aes(x=data_new$Net.worth)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,10000)
```



The above plot is a density plot for Net Worth of the company with respect to the Default ratio. From this we can infer that Majority of the values lies under 2500. Also, there are extreme values present in the data would require our attention to ensure the accuracy of the Model. We can also see that the most the defaulters are those with lesser Net Worth.

- Box plot of Net Worth with respect to Default

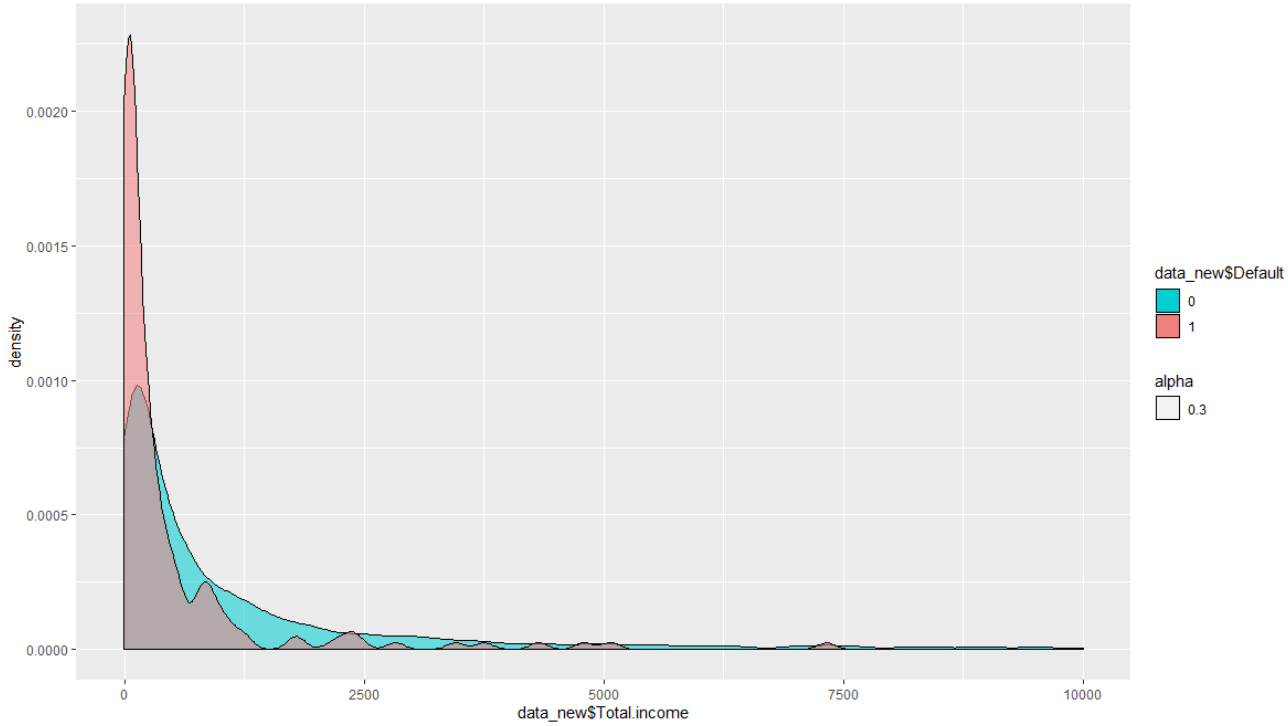
```
boxplot(data_new$Net.worth~data_new$Default, horizontal = TRUE)
```



This plot tells us the presence of extreme in Net Worth based on Default ratio.

- Density Plot of Total Income with respect to Default.

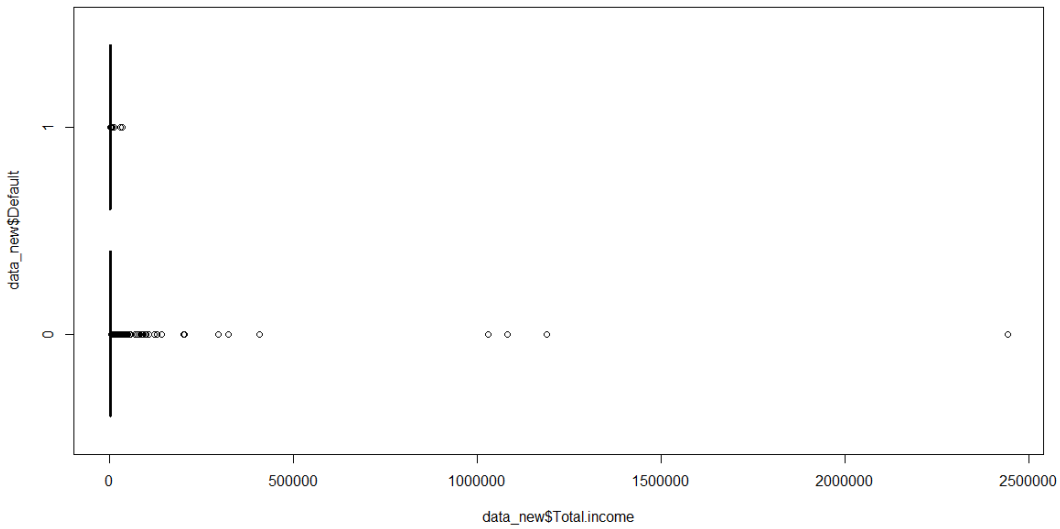
```
ggplot(data_new, aes(x=data_new$Total.income)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,10000)
```



The above plot is a density plot for Total Income of the company with respect to the Default ratio. From this we can infer that Majority of the values lies under 10000. Also, there are extreme values present in the data would require our attention to ensure the accuracy of the Model. We can also see that the most the defaulters are those with lesser Total Income.

- Box plot of Total Income with respect to Default

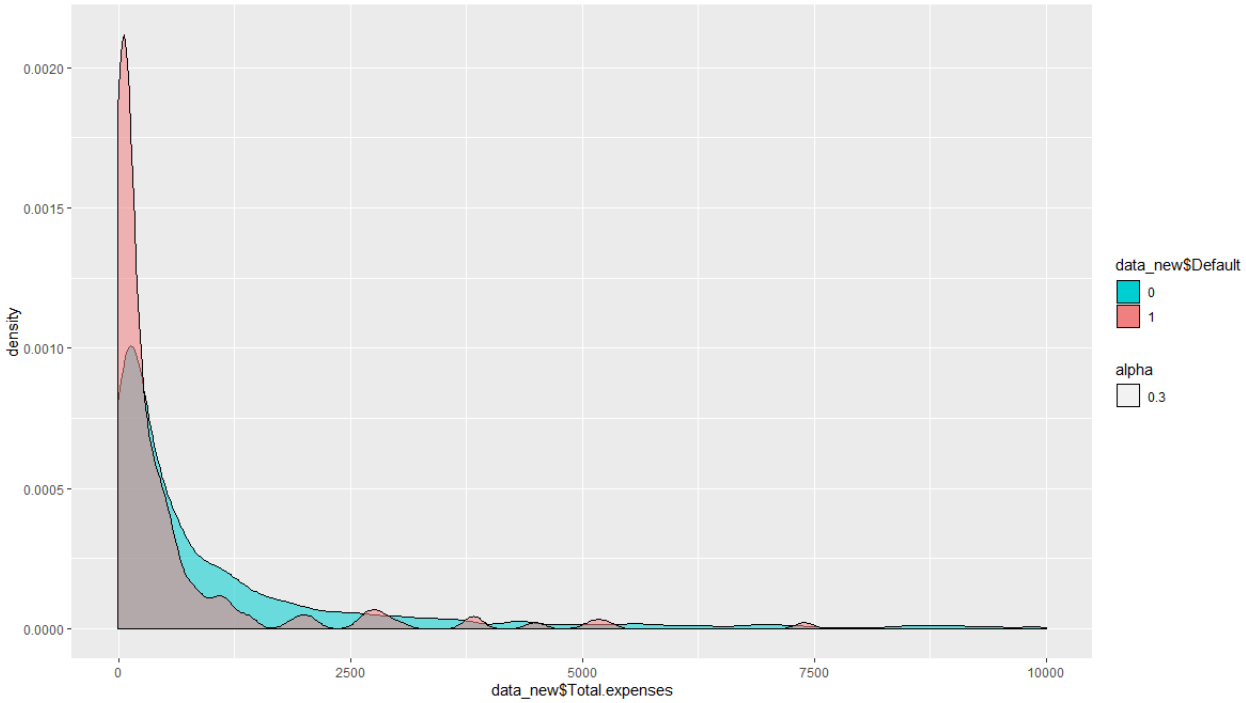
```
boxplot(data_new$Total.income~data_new$Default, horizontal = TRUE)
```



This plot tells us the presence of extreme in Total Income based on Default ratio.

- Density Plot of Total Expense with respect to Default.

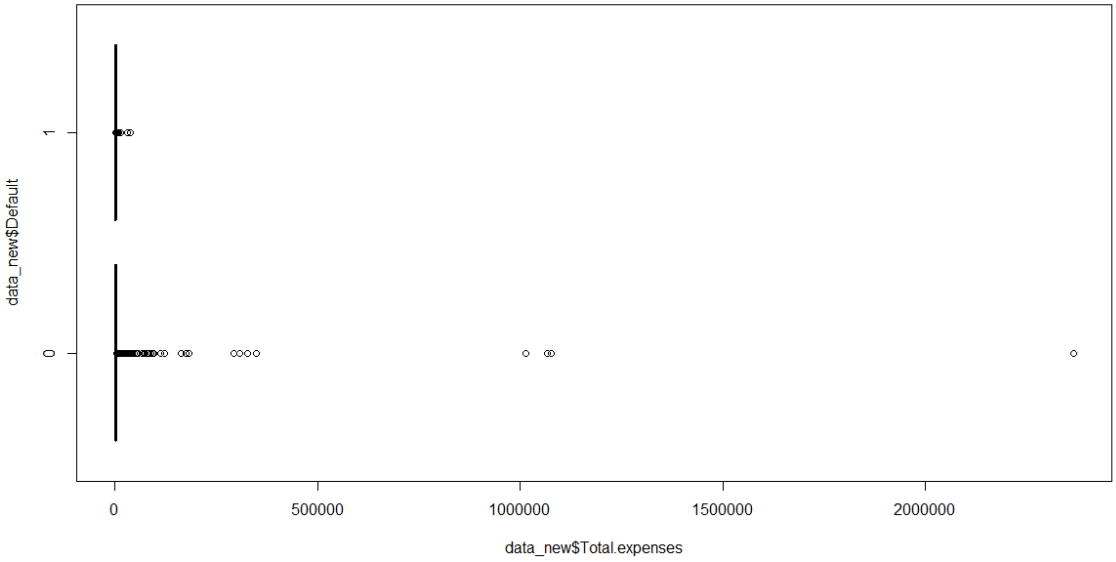
```
ggplot(data_new, aes(x=data_new$Total.expenses)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,10000)
```



The above plot is a density plot for Total Expense of the company with respect to the Default ratio. From this we can infer that Majority of the values lies under 10000. Also, there are extreme values present in the data would require our attention to ensure the accuracy of the Model. We can also see that the most the defaulters are those with lesser Total Expense.

- Box plot of Total Expense with respect to Default

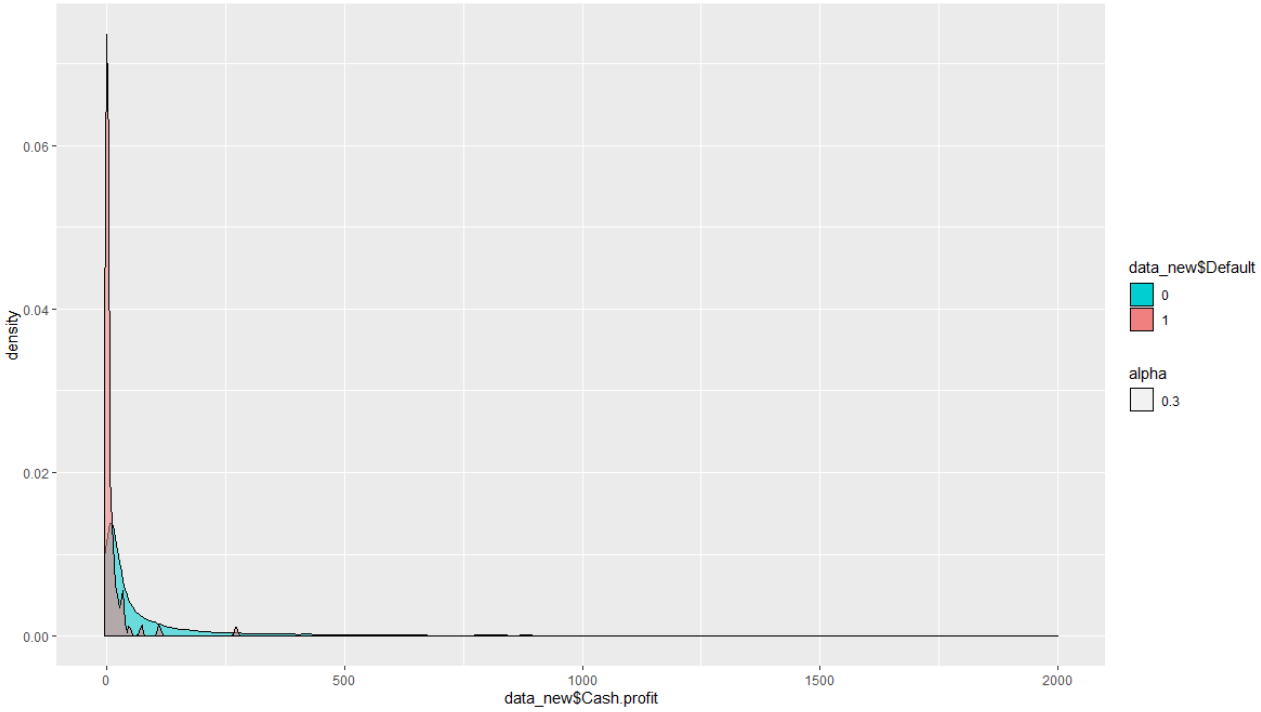
```
boxplot(data_new$Total.expenses~data_new$Default, horizontal = TRUE)
```



This plot tells us the presence of extreme in Total Expense based on Default ratio.

- Density Plot of Cash Profit with respect to Default.

```
ggplot(data_new, aes(x=data_new$Cash.profit)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,2000)
```

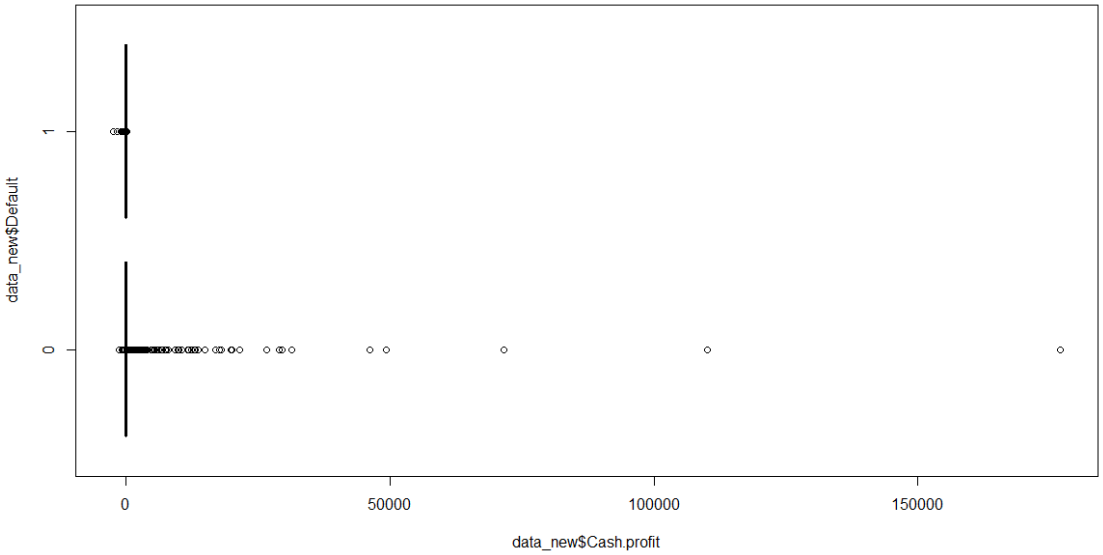


The above plot is a density plot for Cash Profit of the company with respect to the Default ratio. From this we can infer that Majority of the values lies under 500. Also, there are extreme

values present in the data would require our attention to ensure the accuracy of the Model. We can also see that the most the defaulters are those with lesser Cash Profit.

- Box plot of Cash Profit with respect to Default

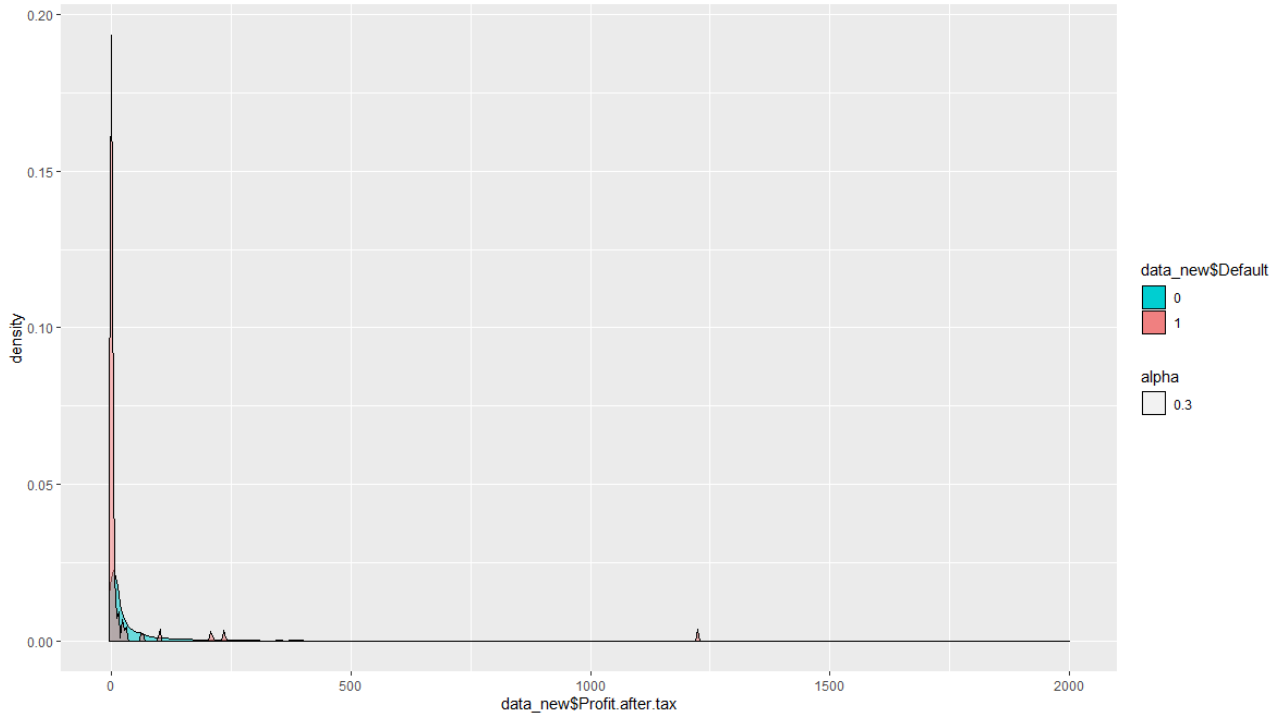
```
boxplot(data_new$Cash.profit~data_new$Default, horizontal = TRUE)
```



This plot tells us the presence of extreme in Cash Profit based on Default ratio.

- Density Plot of Profit After Tax with respect to Default.

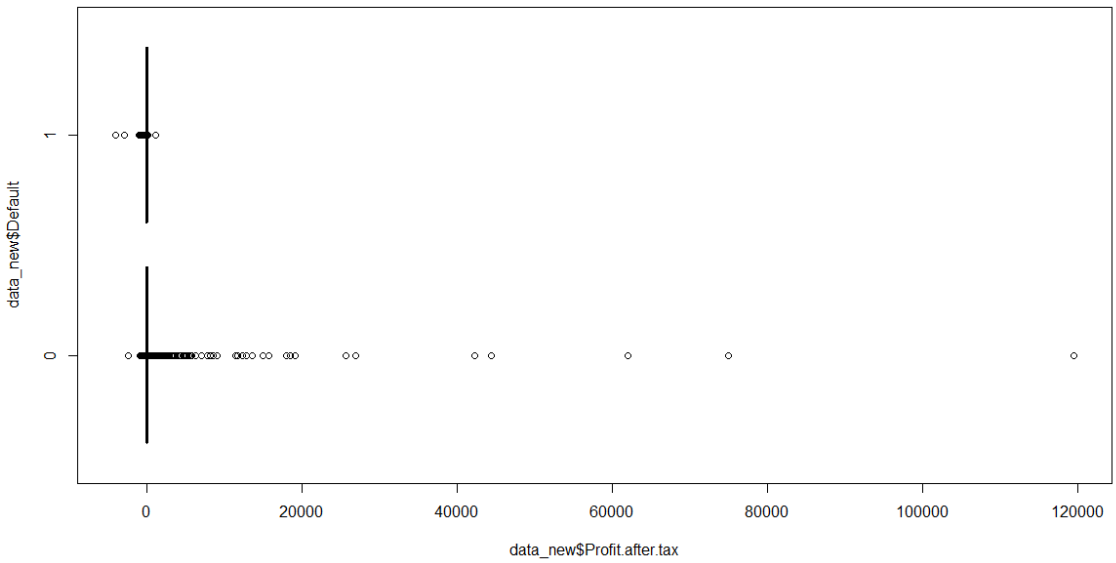
```
ggplot(data_new, aes(x=data_new$Profit.after.tax)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,2000)
```



The above plot is a density plot for Profit After Tax of the company with respect to the Default ratio. From this we can infer that Majority of the values lies under 500. Also, there are extreme values present in the data would require our attention to ensure the accuracy of the Model. We can also see that the most the defaulters are those with lesser Profit After Tax.

- Box plot of Profit After Tax with respect to Default

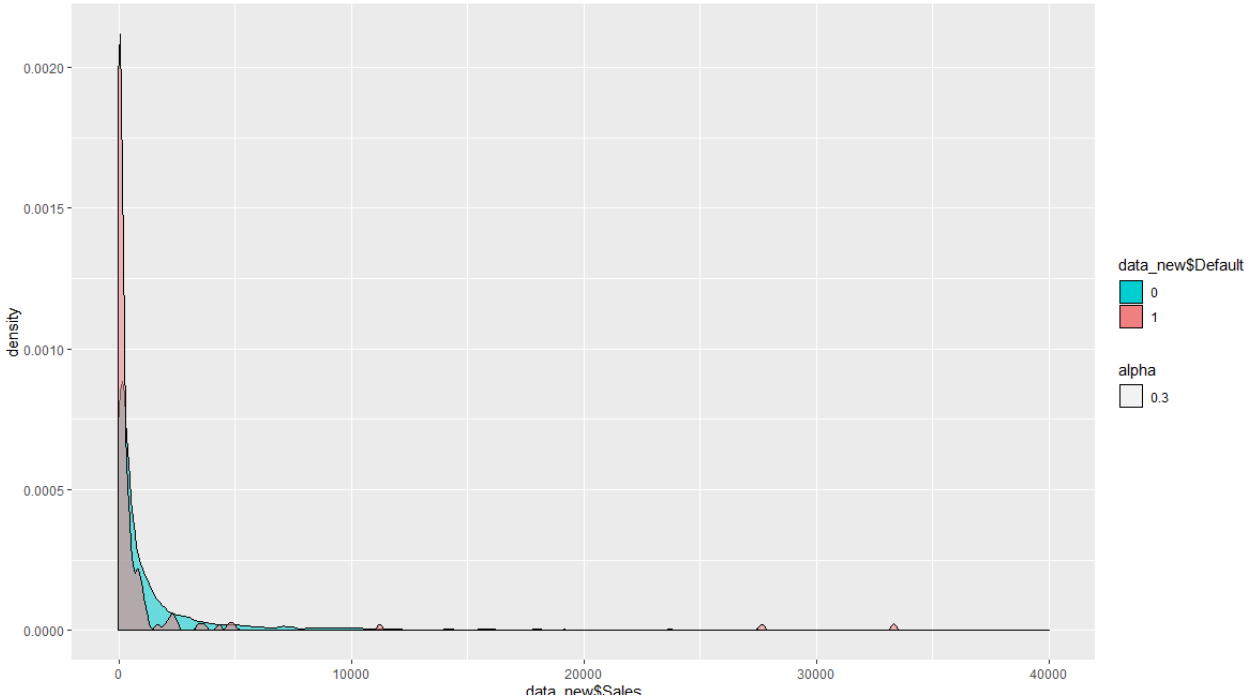
```
boxplot(data_new$Profit.after.tax~data_new$Default, horizontal = TRUE)
```



This plot tells us the presence of extreme in Profit After Tax based on Default ratio.

- Density Plot of Sales with respect to Default.

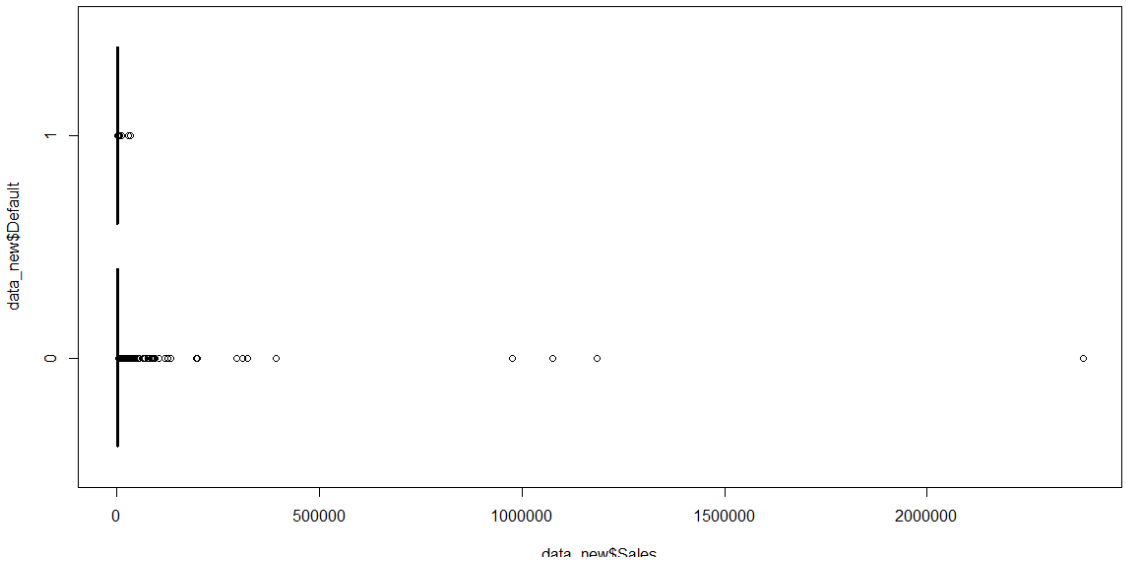
```
ggplot(data_new, aes(x=data_new$Sales)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,40000)
```



The above plot is a density plot for Sales of the company with respect to the Default ratio. From this we can infer that Majority of the values lies under 10000. Also, there are extreme values present in the data would require our attention to ensure the accuracy of the Model. We can also see that the most the defaulters are those with lesser Sales.

- Box plot of Sales with respect to Default

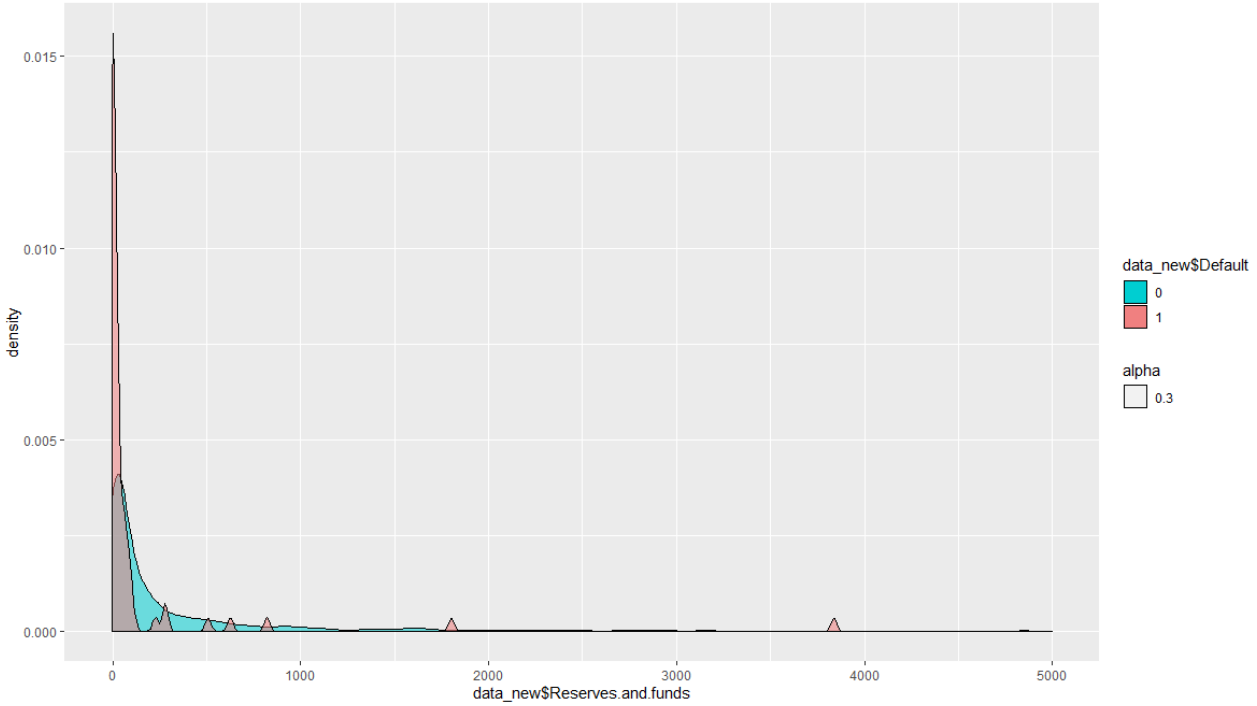
```
boxplot(data_new$Sales~data_new$Default, horizontal = TRUE)
```



This plot tells us the presence of extreme in Sales based on Default ratio.

- Density Plot of Reserves and funds with respect to Default.

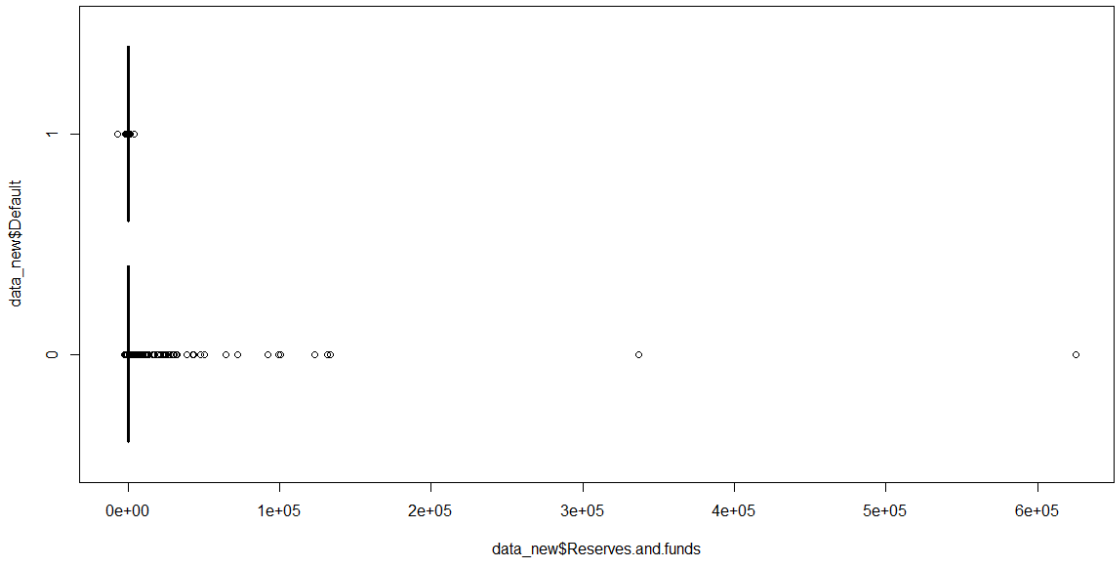
```
ggplot(data_new, aes(x=data_new$Reserves.and.funds)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,5000)
```



The above plot is a density plot for Reserves and Funds of the company with respect to the Default ratio. From this we can infer that Majority of the values lies under 2000. Also, there are extreme values present in the data would require our attention to ensure the accuracy of the Model. We can also see that the most the defaulters are those with lesser Reserves and Funds.

- Box plot of Reserves and Funds with respect to Default

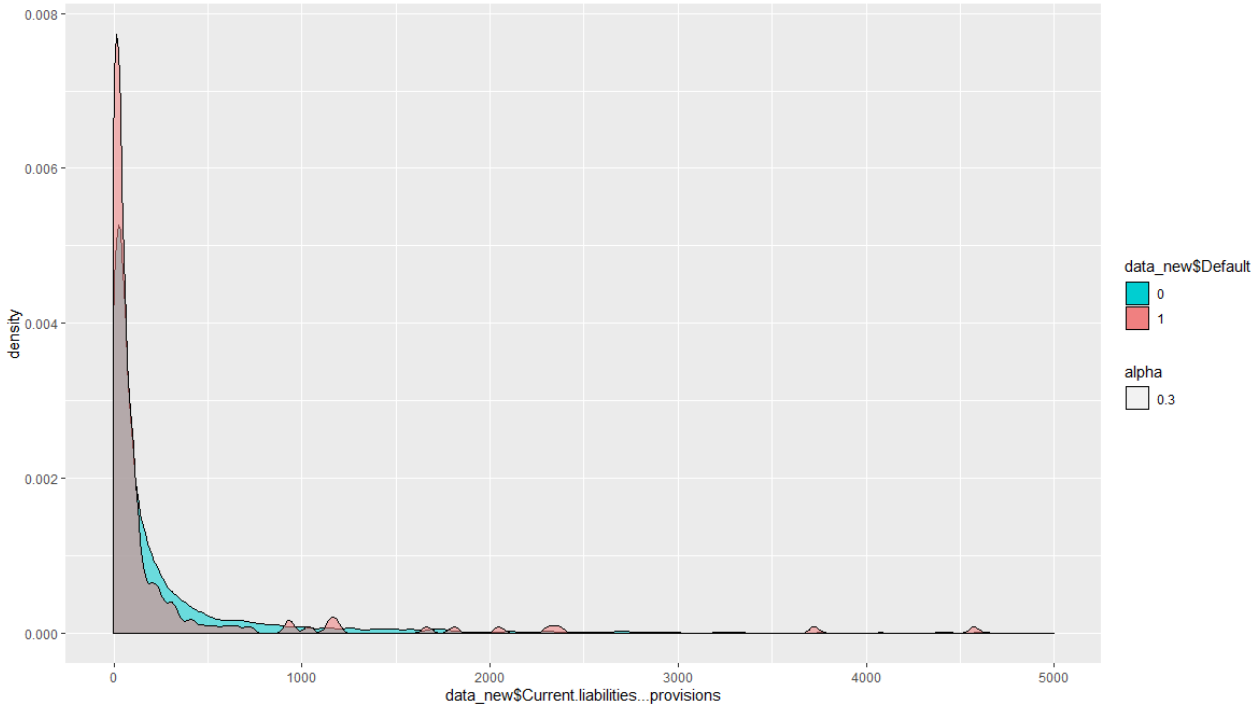
```
boxplot(data_new$Reserves.and.funds~data_new$Default, horizontal = TRUE)
```



This plot tells us the presence of extreme in Sales based on Default ratio.

- Density Plot of Current Liabilities and Provisions with respect to Default.

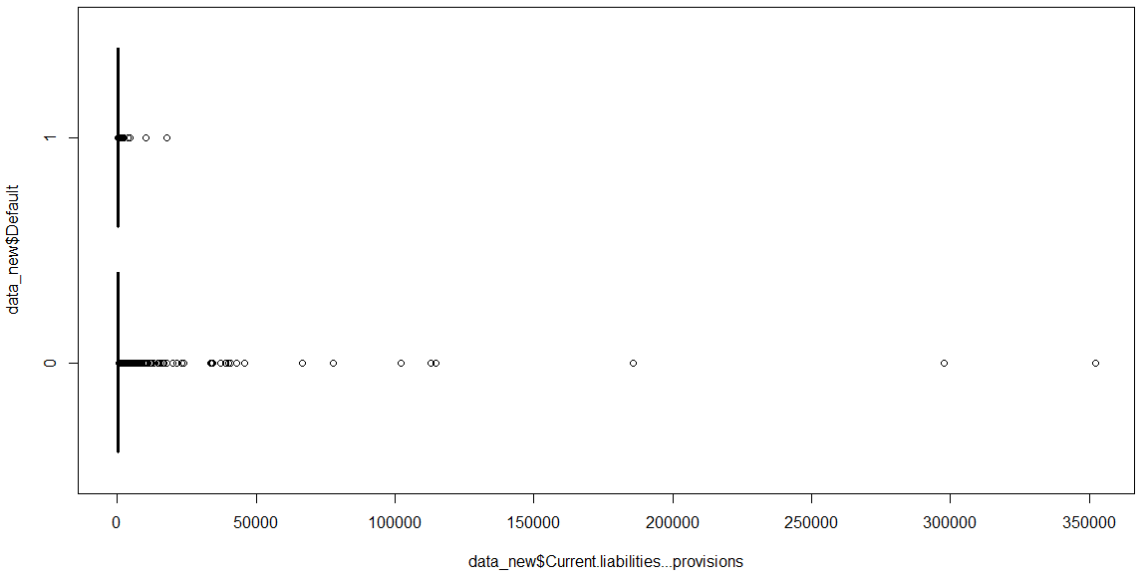
```
ggplot(data_new, aes(x=data_new$Current.liabilities...provisions)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,5000)
```



The above plot is a density plot for Current Liabilities and provisions of the company with respect to the Default ratio. From this we can infer that Majority of the values lies under 2000. Also, there are extreme values present in the data would require our attention to ensure the accuracy of the Model.

- Box plot of Current Liabilities and Provisions with respect to Default

```
boxplot(data_new$Current.liabilities...provisions~data_new$Default,
horizontal = TRUE)
```

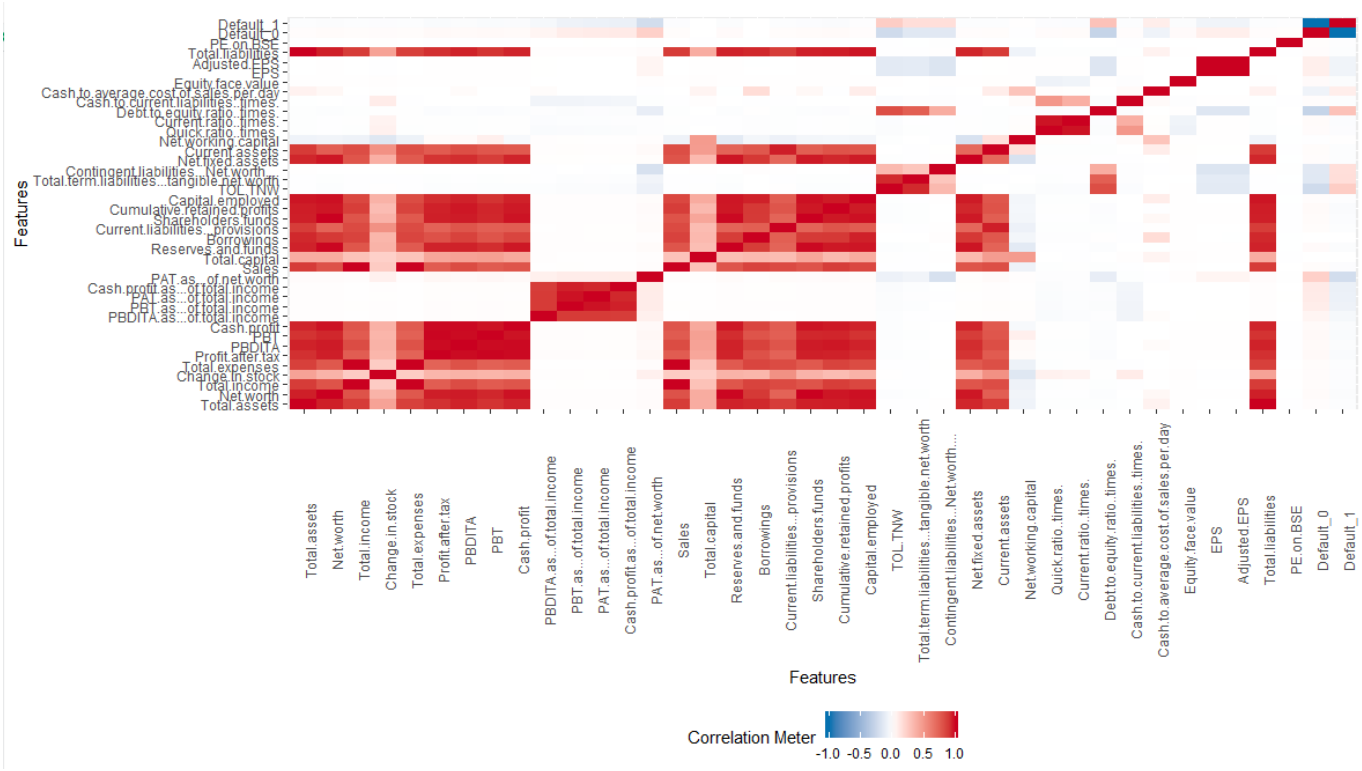


This plot tells us the presence of extreme in Current Liabilities based on Default ratio.

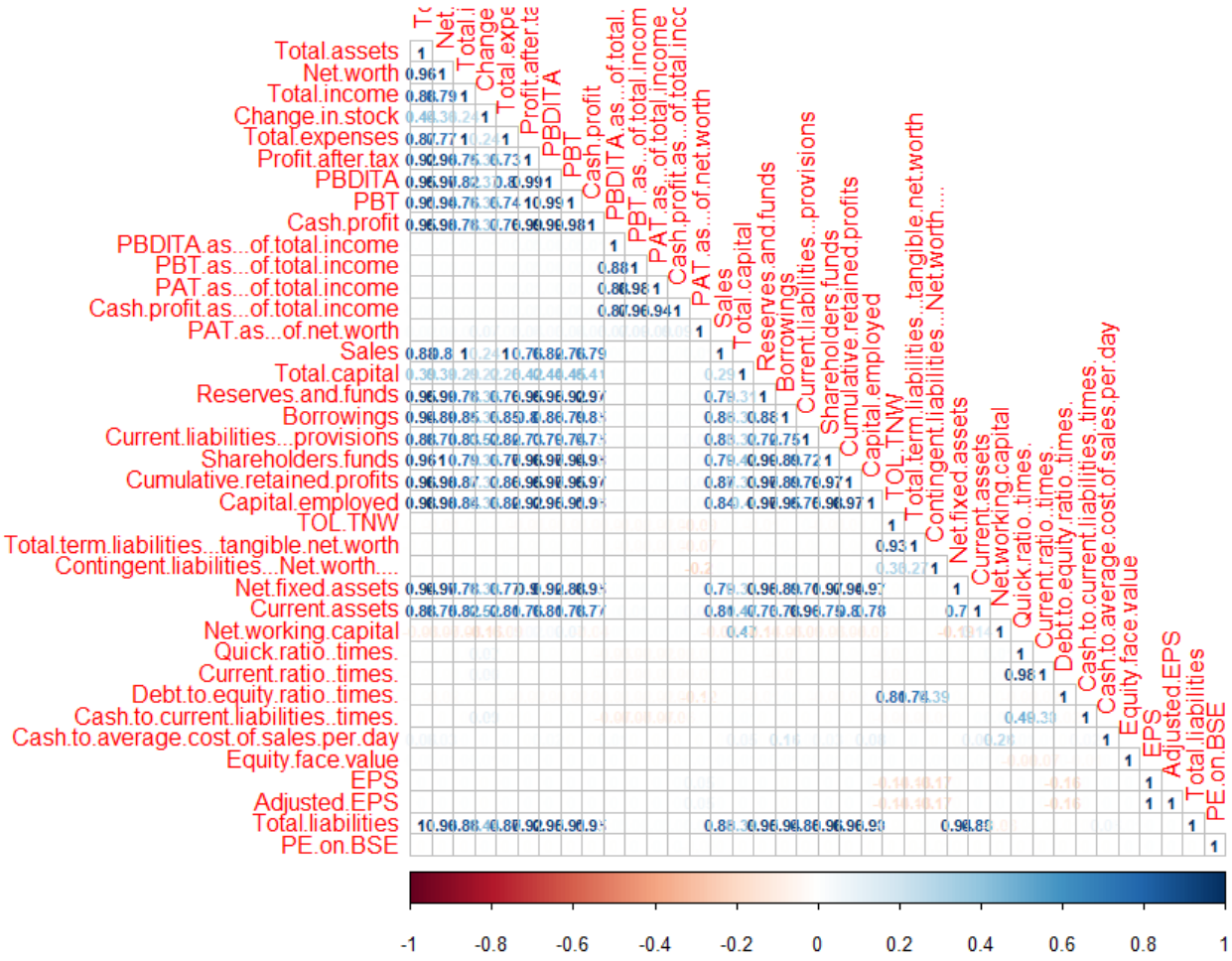
2.1.8 Performing Multicollinearity Plot.

Multicollinearity is a state of very high intercorrelations or inter-associations among the independent variables. It is therefore a type of disturbance in the data, and if present in the data the statistical inferences made about the data may not be reliable.

```
#Checking for the Multicollinearity

plot_correlation(data_new)
```




```
data <- data_new
dim(data)

## [1] 3282   39

corrplot(cor(data[,-39]),method = "number", type = "lower", number.cex = 0.7)
```



From the above two plots we infer that there exist a very high correlation in the variables, which is required to be handled. We will further be checking the Varience Inflation Factor to understand the collinearity of and make the necessary adjustments accordingly.

2.1.9 Outlier Treatment

An outlier is an observation that lies an abnormal distance from other values in a random sample from a population. Examination of the data for unusual observations that are far removed from the mass of data. These points are often referred to as outliers.

From the above Univariate and Bi-Variate Analysis, we identified that there are a large number of extreme values present in that data. We will be treating those extreme values using the Capping and Flooring method of Outlier Treatment.

Since we have a large number of variables present in the dataset, we create a for loop to treat the outliers in a single run.

```
dataset2 <- dataset
a <- c(1:38,40:68)

for(val in a){
  qnt<- quantile(dataset2[,val],probs = c(0.25,0.75))
  cap<- quantile(dataset2[,val],probs = c(0.05,0.95))

  h= 1.5*IQR(dataset2[,val])
  dataset2[,val][dataset2[,val]>(qnt[2]+h)]<- cap[2]
  dataset2[,val][dataset2[,val]<(qnt[1]-h)]<- cap[1]
}
```

At the above chunk, using the quantile function we generate the 1st and 3rd Quantile of variables and then generate the Upper limit and the lower limit. We then cap the values that are above the Upper limit with a 95% value and Floor the values that are below the lower limit at 5%.

2.1.10 Preparing the data for Model Building.

In this part of the report we will preparing the data for Modelling purpose.

As we have seen in the Overview section, that financial Institutes takes the four factors into consideration for checking the risk attached for a company to go bankrupt or Default. These factors also help identify the financial Institutes to understand what percentage of capital can be regained in case the Company becomes insolvent. This helps bank Identify the risk attached to their capital.

Using the crude variables provided in the dataset, we will be creating new ratios taking the four factors into consideration. Below are the new variables segregated on Size, Profitability, Liquidity, Leverage.

Size

- Capital.Employed.per.total.income
- Current.assets.per.total.asset
- Current.liabilities...provisions.per.total.asset
- Current.liabilities...provisions.per.current.asset
- Net.fixed.assets.per.total.asset
- Net.working.capital.per.total.asset
- Net.working.capital.per.total.capital
- Net.working.capital.per.sales
- Shareholder.fund.per.total.asset
- Shareholder.fund.per.total.capital

Profitability

- Change.in.stock.per.total.Income
- Cumulative.retained.profits.per.sale
- PAT.per.total.asset
- PBDITA.per.Sales
- PBT.per.Sales
- PBT.per.total.asset
- PBT.per.Total.Capital
- Sales.per.total.asset

- Total.Income.per.Shareholder.fund
- Total.income.per.total.asset
- Total.income.per.Sale
- Total.income.per.total.expense

Liquidity

- PAT.per.Sales
- Total.asset.per.Current.Liability
- Cash.profit.per.total.asset

Leverage

- Borrowings.per.total.Capital
- Borrowings.per.total.asset
- Total.liabilities.per.shareholder.fund

```
# Creating new variables using the standalone Variables

data$Borrowings.per.total.Capital <- data$Borrowings/data$Total.capital
data$Borrowings.per.total.asset <- data$Borrowings/data$Total.assets
data$capital.Employed.per.total.income <-
data$Capital.employed/data$Total.income
data$cash.profit.per.total.asset <- data$Cash.profit/data$Total.assets
data$Change.in.stock.per.total.Income <-
data$Change.in.stock/data$Total.income
data$Cumulative.retained.profits.per.sale <-
data$Cumulative.retained.profits/data$Sales
data$Current.assets.per.total.asset <- data$Current.assets/data$Total.assets
data$Current.liabilities...provisions.per.total.asset <-
data$Current.liabilities...provisions/data$Total.assets
data$Current.liabilities...provisions.per.current.asset <-
data$Current.liabilities...provisions/data$Current.assets
data$Net.fixed.assets.per.total.asset <-
data$Net.fixed.assets/data$Total.assets
data$Net.working.capital.per.total.asset <-
data$Net.working.capital/data$Total.assets
data$Net.working.capital.per.total.capital <-
data$Net.working.capital/data$Total.capital
data$PAT.per.Sales <- data$Profit.after.tax/data$Sales
data$PAT.per.total.asset <- data$Profit.after.tax/data$Total.assets
data$PBDITA.per.Sales <- data$PBDITA/data$Sales
data$PBT.per.Sales <- data$PBT/data$Sales
data$PBT.per.total.asset <- data$PBT/data$Total.assets
data$PBT.per.Total.Capital <- data$PBT/data$Total.capital
data$Sales.per.total.asset <- data$Sales/data$Total.assets
data$Net.working.capital.per.sales <- data$Net.working.capital/data$Sales
data$Shareholder.fund.per.total.asset <-
data$Shareholders.funds/data$Total.assets
data$Shareholder.fund.per.total.capital <-
data$Shareholders.funds/data$Total.capital
data$Total.asset.per.Current.Liability <-
data$Total.assets/data$Current.liabilities...provisions
data$Total.Income.per.Shareholder.fund <-
data$Total.income/data$Shareholders.funds
data$Total.income.per.total.asset <- data$Total.income/data$Total.assets
data$Total.income.per.Sale <- data$Total.income/data$Sales
data$Total.income.per.total.expense <- data$Total.income/data$Total.expenses
data$Total.liabilities.per.shareholder.fund <-
data$Total.liabilities/data$Shareholders.funds
```

The new dimensions of our dataset are as below.

```
dim(dataset2)

## [1] 3282   68
```

On the new dataset, we check the distribution of Solvent vs Insolvent.

```
# Checking the Proportion on Default.

table(dataset2$Default)

##
##    0    1
## 3104  178
```

We identified that the distribution is 94.5 is to 5.5 in % which signifies that the is highly imbalanced.

We here balance the data using SMOTE algorithm for unbalanced classification problems.

```
bal2 <- SMOTE(dataset2$Default~., data = dataset2,k=5)

table(bal2$Default)

##
##   0   1
## 712 534

prop.table(table(bal2$Default))

##
##         0         1
## 0.5714286 0.4285714
```

Once the data is balanced, we see that the new distribution is 57.1% is to 42.9%

This is our final dataset, that we will be using to build the Model and performing validation on the Unseen data.

2.2 Variable Identification

This section holds the Methods that are used during the Analysis of the problem.
Below are the Functions that we have used for the Analysis.

- **setwd()**
  Set the working directory to dir.

- **read_xlsx()**
  Reads a .xls file in table format and creates a data frame from it.

- **colnames()**
  Retrieve or set the row or column names of a matrix-like object..

- **make.names()**
  Make syntactically valid names out of character vectors.

- **head()**
  Returns the first parts of a vector, matrix, table, data frame or function.

- **str()**
  Compactly display the internal Structure of an R object.

- **summary()**
  Summary is a generic function used to produce result summaries of the results of various model fitting functions.

- **sum()**
  Sum returns the sum of all the values present in its arguments.

- **colsum()**
  Form row and column sums and means for numeric arrays.

- **is.null()**
  NULL is often returned by expressions and functions whose value is undefined. is.null returns TRUE if its argument's value is NULL and FALSE otherwise.

- **na.omit()**
  This function removed the rows from the dataset that contains null values.

- **boxplot()**
  It is plotting technique, which is used to identify if there any outliners are present in the data.

- **plot_histogram()**
  Plot histogram for each continuous feature on a single area.

- **cor()**
  cor compute the variance of x and the covariance or correlation of x and y if these are vectors. If x and y are matrices then the covariances (correlations) between the columns of x and the columns of y are computed.

- **corrplot()**
  This is used to plot the correlation matrix for better visualization and presentation.

- **quantile()**
  quantile produces sample quantiles corresponding to the given probabilities.

- **mice()**
  The mice package implements a method to deal with missing data.

- **glm()**

glm is used to fit generalized linear models, specified by giving a symbolic description of the linear predictor and a description of the error distribution. Here we will be building the Logistic Model using this.

- **vif()**
  Calculates variance-inflation and generalized variance-inflation factors for linear, generalized linear, and other models.

- **lrtest()**
  lrtest is a generic function for carrying out likelihood ratio tests. The default method can be employed for comparing nested (generalized) linear models (see details below).

- **pR2()**
  Compute various pseudo-R2 measures for various GLMs.

- **predict()**
  predict is a generic function for predictions from the results of various model fitting functions. The function invokes particular methods which depend on the class of the first argument.

- **table()**
  table uses the cross-classifying factors to build a contingency table of the counts at each combination of factor levels.

- **confusionmatrix()**
  Calculate the confusion matrix for the fitted values for a logistic regression model.

- **prop.table()**
  This is really sweep(x, margin, margin.table(x, margin), "/") for newbies, except that if margin has length zero, then one gets x/sum(x).

- **ifelse()**
  ifelse returns a value with the same shape as test which is filled with elements selected from either yes or no depending on whether the element of test is TRUE or FALSE.

- **SMOTE()**
  This function handles unbalanced classification problems using the SMOTE method. Namely, it can generate a new "SMOTEd" data set that addresses the class unbalance problem. Alternatively, it can also run a classification algorithm on this new data set and return the resulting model.

- **mutate()**
  This function is very similar to transform but it executes the transformations iteratively so that later transformations can use the columns created by earlier transformations.

- **cbind()**

Take a sequence of vector, matrix or data-frame arguments and combine by columns or rows, respectively. These are generic functions with methods for other R classes.

3 Conclusion

Once the Data is ready for the Model building we will use Logistic Regression Algorithm to build the Model for Credit Risk Analysis for Financial Institutes.

3.1 Logistic Model

In statistics, the logistic model (or logit model) is used to model the probability of a certain class or event existing such as pass/fail, win/lose, alive/dead or healthy/sick. This can be extended to model several classes of events such as determining whether an image contains a cat, dog, lion, etc. Each object being detected in the image would be assigned a probability between 0 and 1 and the sum adding to one.

3.1.1 **Building Logistic Model on Train Data**

```
#Building the Model with the dataset after performing the EDA.

attach(bal2)

glmModel<- glm(bal2$Default~.,data = bal2, family = "binomial")

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

The Logistic Model is created on the all the variables available in the dataset, we will further fine tune the model using Stepwise AIC on both the directions and Variable Inflation Factor.

```
summary(glmModel)
```

```
##
## Call:
## glm(formula = bal2$Default ~ ., family = "binomial", data = bal2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.6205  -0.1663   0.0000   0.1172   2.8098
##
## Coefficients: (1 not defined because of singularities)
##                                                  Estimate Std. Error
## (Intercept)                                    -1.845e+00  6.141e+00
## Total.assets                                    3.206e-04  3.253e-04
## Net.worth                                      -1.190e-03  2.006e-03
## Total.income                                    2.819e-03  1.498e-03
## Change.in.stock                                 8.808e-03  5.265e-03
## Total.expenses                                 -1.287e-04  1.656e-03
## Profit.after.tax                               -7.861e-03  1.436e-02
## PBDITA                                         -4.867e-03  3.534e-03
## PBT                                             6.685e-03  6.266e-03
## Cash.profit                                    -1.632e-02  9.371e-03
## PBDITA.as...of.total.income                     3.355e-01  2.694e-01
## PBT.as...of.total.income                        6.444e-01  4.412e-01
## PAT.as...of.total.income                       -1.217e+00  6.038e-01
## Cash.profit.as...of.total.income                1.466e-02  6.535e-02
## PAT.as...of.net.worth                          -7.910e-02  1.590e-02
## Sales                                          -2.155e-03  2.069e-03
## Total.capital                                  -7.639e-03  2.362e-03
## Reserves.and.funds                             -6.436e-04  1.696e-03
## Borrowings                                     -8.074e-04  7.520e-04
## Current.liabilities...provisions                1.033e-03  1.140e-03
## Shareholders.funds                              8.798e-04  2.777e-03
## Cumulative.retained.profits                    -1.556e-02  4.422e-03
## Capital.employed                                1.891e-03  1.604e-03
## TOL.TNW                                         2.373e-01  1.359e-01
## Total.term.liabilities...tangible.net.worth    -9.422e-01  2.698e-01
## Contingent.liabilities...Net.worth....          5.378e-03  2.778e-03
## Net.fixed.assets                               -1.620e-04  6.596e-04
## Current.assets                                 -2.290e-03  1.107e-03
## Net.working.capital                             6.038e-03  1.977e-03
## Quick.ratio..times.                             1.585e-01  5.048e-01
## Current.ratio..times.                          -9.610e-01  4.554e-01
## Debt.to.equity.ratio..times.                    5.158e-01  2.153e-01
## Cash.to.current.liabilities..times.             1.092e+00  8.296e-01
## Cash.to.average.cost.of.sales.per.day          -5.126e-03  5.049e-03
## Equity.face.value                               6.284e-04  6.539e-03
## EPS                                            -3.704e-02  6.627e-02
## Adjusted.EPS                                    4.371e-02  6.735e-02
## Total.liabilities                                      NA         NA
```

```
## PE.on.BSE                                          -1.107e-02  4.789e-03
## Borrowings.per.total.Capital                       -7.442e-02  4.511e-02
## Borrowings.per.total.asset                          6.700e+00  1.449e+00
## capital.Employed.per.total.income                  -4.268e-02  2.074e-01
## cash.profit.per.total.asset                        -1.682e+00  8.437e+00
## Change.in.stock.per.total.Income                   -1.035e+00  3.753e+00
## Cumulative.retained.profits.per.sale               -2.098e-01  4.977e-01
## Current.assets.per.total.asset                      1.654e+00  1.581e+00
## Current.liabilities...provisions.per.total.asset    6.276e+00  2.549e+00
## Current.liabilities...provisions.per.current.asset -4.748e-01  8.599e-01
## Net.fixed.assets.per.total.asset                    4.979e+00  1.139e+00
## Net.working.capital.per.total.asset                -2.627e-01  2.251e+00
## Net.working.capital.per.total.capital               4.239e-02  9.668e-02
## PAT.per.Sales                                       9.909e+01  5.642e+01
## PAT.per.total.asset                                 6.790e+01  1.904e+01
## PBDITA.per.Sales                                   -3.134e+01  2.536e+01
## PBT.per.Sales                                      -5.031e+01  4.146e+01
## PBT.per.total.asset                                -4.566e+01  1.476e+01
## PBT.per.Total.Capital                              -4.526e-01  2.649e-01
## Sales.per.total.asset                              -2.160e+01  5.275e+00
## Net.working.capital.per.sales                       9.853e-01  1.233e+00
## Shareholder.fund.per.total.asset                    8.741e-01  1.647e+00
## Shareholder.fund.per.total.capital                  5.621e-02  4.427e-02
## Total.asset.per.Current.Liability                  -4.859e-02  4.344e-02
## Total.Income.per.Shareholder.fund                  -5.616e-02  7.825e-02
## Total.income.per.total.asset                        2.211e+01  5.217e+00
## Total.income.per.Sale                              -3.069e+00  4.444e+00
## Total.income.per.total.expense                     -2.876e+00  4.035e+00
## Total.liabilities.per.shareholder.fund              9.551e-02  1.638e-01
##                                                   z value Pr(>|z|)
## (Intercept)                                        -0.300 0.763833
## Total.assets                                        0.986 0.324343
## Net.worth                                          -0.593 0.553046
## Total.income                                        1.882 0.059802 .
## Change.in.stock                                     1.673 0.094348 .
## Total.expenses                                     -0.078 0.938058
## Profit.after.tax                                   -0.548 0.583947
## PBDITA                                             -1.377 0.168515
## PBT                                                 1.067 0.286075
## Cash.profit                                        -1.741 0.081675 .
## PBDITA.as...of.total.income                         1.245 0.212963
## PBT.as...of.total.income                            1.461 0.144101
## PAT.as...of.total.income                           -2.015 0.043917 *
## Cash.profit.as...of.total.income                    0.224 0.822450
## PAT.as...of.net.worth                              -4.976 6.51e-07 ***
## Sales                                              -1.042 0.297596
## Total.capital                                      -3.234 0.001222 **
## Reserves.and.funds                                 -0.379 0.704359
## Borrowings                                         -1.074 0.282968
## Current.liabilities...provisions                    0.907 0.364468
```

```
## Shareholders.funds                                      0.317 0.751391
## Cumulative.retained.profits                            -3.519 0.000434 ***
## Capital.employed                                        1.179 0.238579
## TOL.TNW                                                 1.747 0.080663 .
## Total.term.liabilities...tangible.net.worth            -3.493 0.000478 ***
## Contingent.liabilities...Net.worth....                  1.936 0.052912 .
## Net.fixed.assets                                       -0.246 0.806025
## Current.assets                                         -2.069 0.038541 *
## Net.working.capital                                     3.054 0.002255 **
## Quick.ratio..times.                                     0.314 0.753491
## Current.ratio..times.                                  -2.110 0.034854 *
## Debt.to.equity.ratio..times.                            2.395 0.016608 *
## Cash.to.current.liabilities..times.                     1.316 0.188208
## Cash.to.average.cost.of.sales.per.day                  -1.015 0.309959
## Equity.face.value                                       0.096 0.923450
## EPS                                                    -0.559 0.576204
## Adjusted.EPS                                            0.649 0.516369
## Total.liabilities                                          NA       NA
## PE.on.BSE                                              -2.311 0.020833 *
## Borrowings.per.total.Capital                           -1.650 0.099009 .
## Borrowings.per.total.asset                              4.625 3.74e-06 ***
## capital.Employed.per.total.income                      -0.206 0.836944
## cash.profit.per.total.asset                            -0.199 0.841940
## Change.in.stock.per.total.Income                       -0.276 0.782749
## Cumulative.retained.profits.per.sale                   -0.422 0.673350
## Current.assets.per.total.asset                          1.046 0.295504
## Current.liabilities...provisions.per.total.asset        2.463 0.013797 *
## Current.liabilities...provisions.per.current.asset     -0.552 0.580839
## Net.fixed.assets.per.total.asset                        4.372 1.23e-05 ***
## Net.working.capital.per.total.asset                    -0.117 0.907089
## Net.working.capital.per.total.capital                   0.438 0.661088
## PAT.per.Sales                                           1.756 0.079037 .
## PAT.per.total.asset                                     3.567 0.000362 ***
## PBDITA.per.Sales                                       -1.236 0.216440
## PBT.per.Sales                                          -1.213 0.225031
## PBT.per.total.asset                                    -3.093 0.001982 **
## PBT.per.Total.Capital                                  -1.708 0.087567 .
## Sales.per.total.asset                                  -4.095 4.22e-05 ***
## Net.working.capital.per.sales                           0.799 0.424140
## Shareholder.fund.per.total.asset                        0.531 0.595610
## Shareholder.fund.per.total.capital                      1.270 0.204157
## Total.asset.per.Current.Liability                      -1.119 0.263303
## Total.Income.per.Shareholder.fund                      -0.718 0.472921
## Total.income.per.total.asset                            4.238 2.25e-05 ***
## Total.income.per.Sale                                  -0.691 0.489779
## Total.income.per.total.expense                         -0.713 0.476040
## Total.liabilities.per.shareholder.fund                  0.583 0.559897
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1701.81  on 1245  degrees of freedom
## Residual deviance:  441.78  on 1180  degrees of freedom
## AIC: 573.78
##
## Number of Fisher Scoring iterations: 12
```

Predicting the Model on the Trained dataset.

```
glmPred <- predict(glmModel,data= bal2)
glmPred1 <- as.factor(ifelse(glmPred <0.6,0,1))
caret::confusionMatrix(glmPred1,as.factor(bal2$Default),positive='1')

## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 687  68
##          1  25 466
##
##                Accuracy : 0.9254
##                  95% CI : (0.9093, 0.9393)
##     No Information Rate : 0.5714
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8461
##
##  Mcnemar's Test P-Value : 1.329e-05
##
##             Sensitivity : 0.8727
##             Specificity : 0.9649
##          Pos Pred Value : 0.9491
##          Neg Pred Value : 0.9099
##              Prevalence : 0.4286
##          Detection Rate : 0.3740
##    Detection Prevalence : 0.3941
##       Balanced Accuracy : 0.9188
##
##        'Positive' Class : 1
##
```

We've achieved the accuracy of 92.5% with specificity of 96% and Sensitivity of 87%, since there is a huge collinearity in the data it is not a reliable Model to be considered. We will further be fine tuning this model using Stepwise AIC and VIF.

Performing the Stepwise AIC on the initial Model

```
# Refining the Model using Step Wise AIC.

stepAIC(glmModel,direction = "both")

##
## Call:  glm(formula = bal2$Default ~ Total.income + Change.in.stock +
##     PBDITA + Cash.profit + PBDITA.as...of.total.income +
PBT.as...of.total.income +
##     PAT.as...of.total.income + PAT.as...of.net.worth + Sales +
##     Total.capital + Cumulative.retained.profits + Capital.employed +
##     TOL.TNW + Total.term.liabilities...tangible.net.worth +
Contingent.liabilities...Net.worth.... +
##     Current.assets + Net.working.capital + Current.ratio..times. +
##     Debt.to.equity.ratio..times. + PE.on.BSE +
Borrowings.per.total.Capital +
##     Borrowings.per.total.asset + Current.assets.per.total.asset +
##     Current.liabilities...provisions.per.total.asset +
Net.fixed.assets.per.total.asset +
##     PAT.per.Sales + PAT.per.total.asset + PBDITA.per.Sales +
##     PBT.per.Sales + PBT.per.total.asset + Sales.per.total.asset +
##     Total.income.per.total.asset + Cash.to.current.liabilities..times.,
##     family = "binomial", data = bal2)
##
## Coefficients:
##                                     (Intercept)
##                                       -8.115847
##                                    Total.income
##                                        0.002068
##                                 Change.in.stock
##                                        0.009055
##                                          PBDITA
##                                       -0.004369
##                                    Cash.profit
##                                       -0.013698
##                       PBDITA.as...of.total.income
##                                        0.416451
##                          PBT.as...of.total.income
##                                        0.599947
##                          PAT.as...of.total.income
##                                       -1.095204
##                             PAT.as...of.net.worth
##                                       -0.073758
##                                           Sales
##                                       -0.001622
##                                   Total.capital
##                                       -0.006354
##                       Cumulative.retained.profits
##                                       -0.013563
##                                 Capital.employed
##                                        0.001271
```

```
##                                        TOL.TNW
##                                       0.190110
##      Total.term.liabilities...tangible.net.worth
##                                      -0.893462
##           Contingent.liabilities...Net.worth....
##                                       0.005185
##                                  Current.assets
##                                      -0.001153
##                              Net.working.capital
##                                       0.004789
##                           Current.ratio..times.
##                                      -0.501888
##                     Debt.to.equity.ratio..times.
##                                       0.527679
##                                     PE.on.BSE
##                                      -0.010747
##                     Borrowings.per.total.Capital
##                                      -0.054101
##                       Borrowings.per.total.asset
##                                       6.134866
##                   Current.assets.per.total.asset
##                                       1.933700
## Current.liabilities...provisions.per.total.asset
##                                       6.780646
##                 Net.fixed.assets.per.total.asset
##                                       4.722513
##                                   PAT.per.Sales
##                                      85.440010
##                               PAT.per.total.asset
##                                      61.937928
##                                  PBDITA.per.Sales
##                                     -39.766182
##                                    PBT.per.Sales
##                                     -44.599820
##                               PBT.per.total.asset
##                                     -46.528015
##                           Sales.per.total.asset
##                                     -20.745722
##                     Total.income.per.total.asset
##                                      21.059810
##              Cash.to.current.liabilities..times.
##                                       0.825638
##
## Degrees of Freedom: 1245 Total (i.e. Null);  1212 Residual
## Null Deviance:       1702
## Residual Deviance: 456.4     AIC: 524.4
```

Above are the final variables that we received from the Stepwise AIC, we will be creating the model using the above variables and refine it further if the VIF value is too high.

```
glmModel1 <- glm(bal2$Default~Total.assets+Profit.after.tax+Cash.profit+
                   PBT.as...of.total.income+Cash.profit.as...of.total.income+

PAT.as...of.net.worth+Total.capital+Borrowings+Current.liabilities...provisio
ns+

Cumulative.retained.profits+Contingent.liabilities...Net.worth....
                 +Net.working.capital+Debt.to.equity.ratio..times.+
Cash.to.current.liabilities..times.+

PE.on.BSE+Borrowings.per.total.asset+capital.Employed.per.total.income

+Change.in.stock.per.total.Income+Cumulative.retained.profits.per.sale+

Current.assets.per.total.asset+Current.liabilities...provisions.per.current.a
sset+

Net.fixed.assets.per.total.asset+Net.working.capital.per.total.asset+
                   Shareholder.fund.per.total.asset+Total.income.per.Sale+
                   Total.liabilities.per.shareholder.fund
                 ,data = bal2,
                 family = "binomial")
```

```
summary(glmModel1)
```

```
##
## Call:
## glm(formula = bal2$Default ~ Total.assets + Profit.after.tax +
##     Cash.profit + PBT.as...of.total.income +
Cash.profit.as...of.total.income +
##     PAT.as...of.net.worth + Total.capital + Borrowings +
Current.liabilities...provisions +
##     Cumulative.retained.profits + Contingent.liabilities...Net.worth.... +
##     Net.working.capital + Debt.to.equity.ratio..times. +
Cash.to.current.liabilities..times. +
##     PE.on.BSE + Borrowings.per.total.asset +
capital.Employed.per.total.income +
##     Change.in.stock.per.total.Income +
Cumulative.retained.profits.per.sale +
##     Current.assets.per.total.asset +
Current.liabilities...provisions.per.current.asset +
##     Net.fixed.assets.per.total.asset + Net.working.capital.per.total.asset
+
##     Shareholder.fund.per.total.asset + Total.income.per.Sale +
##     Total.liabilities.per.shareholder.fund, family = "binomial",
##     data = bal2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.1837  -0.2901   0.0000   0.2180   3.2004
##
## Coefficients:
##                                                 Estimate Std. Error
## (Intercept)                                   -1.305e+01  2.853e+00
## Total.assets                                   2.223e-04  1.527e-04
## Profit.after.tax                               8.148e-03  3.283e-03
## Cash.profit                                   -8.724e-03  2.911e-03
## PBT.as...of.total.income                      -7.911e-03  2.728e-02
## Cash.profit.as...of.total.income              -2.153e-02  3.134e-02
## PAT.as...of.net.worth                         -5.639e-02  8.938e-03
## Total.capital                                 -3.516e-03  1.470e-03
## Borrowings                                    -3.370e-04  4.246e-04
## Current.liabilities...provisions               4.542e-04  4.524e-04
## Cumulative.retained.profits                   -1.139e-02  2.485e-03
## Contingent.liabilities...Net.worth....         3.795e-03  2.217e-03
## Net.working.capital                            2.043e-03  1.069e-03
## Debt.to.equity.ratio..times.                   2.480e-01  1.041e-01
## Cash.to.current.liabilities..times.            4.107e-01  4.770e-01
## PE.on.BSE                                     -1.171e-02  3.700e-03
## Borrowings.per.total.asset                     4.170e+00  9.679e-01
## capital.Employed.per.total.income             -3.939e-01  1.281e-01
## Change.in.stock.per.total.Income               2.159e+00  2.129e+00
## Cumulative.retained.profits.per.sale          -1.114e+00  3.686e-01
```

```
## Current.assets.per.total.asset                        4.482e+00  1.040e+00
## Current.liabilities...provisions.per.current.asset  1.739e+00  4.373e-01
## Net.fixed.assets.per.total.asset                      2.978e+00  7.660e-01
## Net.working.capital.per.total.asset                -2.343e+00  1.166e+00
## Shareholder.fund.per.total.asset                      4.407e-01  1.220e+00
## Total.income.per.Sale                                 6.214e+00  2.352e+00
## Total.liabilities.per.shareholder.fund                6.508e-02  8.715e-02
##                                                     z value Pr(>|z|)
## (Intercept)                                          -4.574 4.78e-06 ***
## Total.assets                                          1.456 0.145437
## Profit.after.tax                                      2.482 0.013063 *
## Cash.profit                                          -2.997 0.002730 **
## PBT.as...of.total.income                             -0.290 0.771814
## Cash.profit.as...of.total.income                     -0.687 0.492030
## PAT.as...of.net.worth                                -6.309 2.82e-10 ***
## Total.capital                                        -2.392 0.016735 *
## Borrowings                                           -0.794 0.427311
## Current.liabilities...provisions                      1.004 0.315361
## Cumulative.retained.profits                          -4.585 4.55e-06 ***
## Contingent.liabilities...Net.worth....                1.712 0.086913 .
## Net.working.capital                                   1.911 0.055972 .
## Debt.to.equity.ratio..times.                          2.381 0.017252 *
## Cash.to.current.liabilities..times.                   0.861 0.389195
## PE.on.BSE                                            -3.165 0.001549 **
## Borrowings.per.total.asset                            4.309 1.64e-05 ***
## capital.Employed.per.total.income                    -3.075 0.002104 **
## Change.in.stock.per.total.Income                      1.014 0.310443
## Cumulative.retained.profits.per.sale                 -3.022 0.002512 **
## Current.assets.per.total.asset                        4.308 1.64e-05 ***
## Current.liabilities...provisions.per.current.asset    3.977 6.98e-05 ***
## Net.fixed.assets.per.total.asset                      3.888 0.000101 ***
## Net.working.capital.per.total.asset                  -2.008 0.044593 *
## Shareholder.fund.per.total.asset                      0.361 0.717905
## Total.income.per.Sale                                 2.642 0.008248 **
## Total.liabilities.per.shareholder.fund                0.747 0.455238
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1701.8  on 1245  degrees of freedom
## Residual deviance:  554.0  on 1219  degrees of freedom
## AIC: 608
##
## Number of Fisher Scoring iterations: 9
```

In the above model, majority of the variables appears to be the most significant variables, below are the most significant variables that contributes in the identification of Default i.e. higher the probability greater are the changes of company to be default.

- Total.income.per.Sale
- Net.fixed.assets.per.total.asset
- Current.liabilities...provisions.per.current.asset
- Current.assets.per.total.asset
- Change.in.stock.per.total.Income
- Borrowings.per.total.asset
- Cash.to.current.liabilities..times
- Shareholder.fund.per.total.asset

Below are the few variables that are negatively contributing to interpret the risk, i.e. lesser the probability higher the risk of company getting default.

- Current.liabilities...provisions
- Borrowings
- Cash Profit

- PBT as of Total Income
- Cumulative.retained.profits
- PAT.as...of.net.worth

We see, that the above variables contribute in identifying the Company to become Default

Performing the likelihood ratio to check the validity of the model

```
lr.Test <- lmtest::lrtest(glmModel1)
lr.Test

## Likelihood ratio test
##
## Model 1: bal2$Default ~ Total.assets + Profit.after.tax + Cash.profit +
##     PBT.as...of.total.income + Cash.profit.as...of.total.income +
##     PAT.as...of.net.worth + Total.capital + Borrowings +
Current.liabilities...provisions +
##     Cumulative.retained.profits + Contingent.liabilities...Net.worth.... +
##     Net.working.capital + Debt.to.equity.ratio..times. +
Cash.to.current.liabilities..times. +
##     PE.on.BSE + Borrowings.per.total.asset +
capital.Employed.per.total.income +
##     Change.in.stock.per.total.Income +
Cumulative.retained.profits.per.sale +
##     Current.assets.per.total.asset +
Current.liabilities...provisions.per.current.asset +
##     Net.fixed.assets.per.total.asset + Net.working.capital.per.total.asset
+
##     Shareholder.fund.per.total.asset + Total.income.per.Sale +
##     Total.liabilities.per.shareholder.fund
## Model 2: bal2$Default ~ 1
##   #Df LogLik  Df  Chisq Pr(>Chisq)
## 1  27 -277.0
## 2   1 -850.9 -26 1147.8  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the above likelihood ratio, we have achieved the Chisq value of 2.2e-16 which is too small from alpha hence the model is significant.

Checking the VIF values of the final Model build.

```
# Checking the Variable Inflation Factor

vif(glmModel1)
```

```
##                                       Total.assets
##                                           5.955498
##                                    Profit.after.tax
##                                           4.520845
##                                         Cash.profit
##                                           5.055048
##                               PBT.as...of.total.income
##                                           5.611406
##                        Cash.profit.as...of.total.income
##                                           4.182282
##                                 PAT.as...of.net.worth
##                                           2.066293
##                                      Total.capital
##                                           4.564983
##                                        Borrowings
##                                           5.471812
##                        Current.liabilities...provisions
##                                           3.123628
##                             Cumulative.retained.profits
##                                           2.035013
##                  Contingent.liabilities...Net.worth....
##                                           1.236799
##                                  Net.working.capital
##                                           1.691265
##                           Debt.to.equity.ratio..times.
##                                           4.015623
##                     Cash.to.current.liabilities..times.
##                                           1.639853
##                                         PE.on.BSE
##                                           1.129655
##                              Borrowings.per.total.asset
##                                           2.926413
##                        capital.Employed.per.total.income
##                                           2.441577
##                         Change.in.stock.per.total.Income
##                                           1.138035
##                     Cumulative.retained.profits.per.sale
##                                           1.886317
##                           Current.assets.per.total.asset
##                                           4.389126
## Current.liabilities...provisions.per.current.asset
##                                           2.952782
##                         Net.fixed.assets.per.total.asset
##                                           2.129210
##                      Net.working.capital.per.total.asset
##                                           3.407543
```

```
##                         Shareholder.fund.per.total.asset
##                                           4.531783
##                                  Total.income.per.Sale
##                                           1.441215
##                  Total.liabilities.per.shareholder.fund
##                                           5.349458
```

Above, we can see that the VIF value for the variables consumed by the Model are less then 6 hence there exist no collinearity in the variables.

Predicting the Model on Original Dataset.

```
# Predicting probability on Original Dataset

pred <- predict(glmModel1,data = bal2)
pred1 <- ifelse(pred<0.5,0,1)
pred1 <- as.factor(pred1)
actuals <- as.factor(bal2$Default)
```

Creating the confusion Matrix.

```
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 671  79
         1  41 455

               Accuracy : 0.9037
                 95% CI : (0.8859, 0.9195)
    No Information Rate : 0.5714
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.8016

 Mcnemar's Test P-Value : 0.0007312

            Sensitivity : 0.8521
            Specificity : 0.9424
         Pos Pred Value : 0.9173
         Neg Pred Value : 0.8947
             Prevalence : 0.4286
         Detection Rate : 0.3652
   Detection Prevalence : 0.3981
      Balanced Accuracy : 0.8972

       'Positive' Class : 1
```

We have achieved the accuracy of 90% with a sensitivity of 85% and specificity of 94%, hence the model is a significant model and a balanced model. We will be using this model to predict the probabilities on the validation set.

Validating the Model using Likelihood Ratio Test.

```
lr.Test <- lmtest::lrtest(glmModel)
lr.Test

## Likelihood ratio test
##
## Model 1: bal2$Default ~ Total.assets + Net.worth + Total.income +
Change.in.stock +
##     Total.expenses + Profit.after.tax + PBDITA + PBT + Cash.profit +
##     PBDITA.as...of.total.income + PBT.as...of.total.income +
##     PAT.as...of.total.income + Cash.profit.as...of.total.income +
##     PAT.as...of.net.worth + Sales + Total.capital + Reserves.and.funds +
##     Borrowings + Current.liabilities...provisions + Shareholders.funds +
##     Cumulative.retained.profits + Capital.employed + TOL.TNW +
##     Total.term.liabilities...tangible.net.worth +
Contingent.liabilities...Net.worth.... +
##     Net.fixed.assets + Current.assets + Net.working.capital +
##     Quick.ratio..times. + Current.ratio..times. +
Debt.to.equity.ratio..times. +
##     Cash.to.current.liabilities..times. +
Cash.to.average.cost.of.sales.per.day +
##     Equity.face.value + EPS + Adjusted.EPS + Total.liabilities +
##     PE.on.BSE + Borrowings.per.total.Capital + Borrowings.per.total.asset
+
##     capital.Employed.per.total.income + cash.profit.per.total.asset +
##     Change.in.stock.per.total.Income +
Cumulative.retained.profits.per.sale +
##     Current.assets.per.total.asset +
Current.liabilities...provisions.per.total.asset +
##     Current.liabilities...provisions.per.current.asset +
Net.fixed.assets.per.total.asset +
##     Net.working.capital.per.total.asset +
Net.working.capital.per.total.capital +
##     PAT.per.Sales + PAT.per.total.asset + PBDITA.per.Sales +
##     PBT.per.Sales + PBT.per.total.asset + PBT.per.Total.Capital +
##     Sales.per.total.asset + Net.working.capital.per.sales +
Shareholder.fund.per.total.asset +
##     Shareholder.fund.per.total.capital + Total.asset.per.Current.Liability
+
##     Total.Income.per.Shareholder.fund + Total.income.per.total.asset +
##     Total.income.per.Sale + Total.income.per.total.expense +
##     Total.liabilities.per.shareholder.fund
## Model 2: bal2$Default ~ 1
##   #Df  LogLik  Df Chisq Pr(>Chisq)
## 1  66 -220.89
## 2   1 -850.90 -65  1260  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the above likelihood ratio, we have achieved the Chisq value of 2.2e-16 which is too small from alpha hence the model is significant.

From this section, we will be importing the Validation dataset and prepare it for the prediction using the above final Model.

Importing the Validation dataset

```
#########Importing the Validation dataset###############
unseen <- read_xlsx('validation_data.xlsx')
```

Transforming the names of the variables

```
colnames(unseen)<- make.names(colnames(unseen))
unseen1 <- unseen
```

Converting the variable datatypes to numeric

```
# Converting the datatype of variables

unseen1$Creditors.turnover <- as.numeric(unseen1$Creditors.turnover)

## Warning: NAs introduced by coercion

unseen1$Debtors.turnover <- as.numeric(unseen1$Debtors.turnover)

## Warning: NAs introduced by coercion

unseen1$Finished.goods.turnover <-
as.numeric(unseen1$Finished.goods.turnover)

## Warning: NAs introduced by coercion

unseen1$WIP.turnover <- as.numeric(unseen1$WIP.turnover)

## Warning: NAs introduced by coercion

unseen1$Raw.material.turnover <- as.numeric(unseen1$Raw.material.turnover)

## Warning: NAs introduced by coercion

unseen1$Shares.outstanding <- as.numeric(unseen1$Shares.outstanding)

## Warning: NAs introduced by coercion

unseen1$Equity.face.value <- as.numeric(unseen1$Equity.face.value)

## Warning: NAs introduced by coercion

unseen1$PE.on.BSE <- as.numeric(unseen1$PE.on.BSE)

## Warning: NAs introduced by coercion
```

Imputing the null values in the validation dataset.

```
# Dropping variables that are dropped in Original to bring it to the same
level

data_unseen <- unseen1[,-c(1,18,19,22,25,32,34,42:47)]

# Imputing the missing values

unseen <- mice(data_unseen,m=5,meth = "pmm",maxit = 10,seed = 500)
```

```
unseen$imp

unseen <- complete(unseen,1)
```

```
unseen<-na.omit(unseen)
anyNA(unseen)

## [1] FALSE
```

Creating the same variables that were created in Original dataset.

```
# Creating new variables that are created in Original Dataset to create it of
the same level

unseen$Borrowings.per.total.Capital <- unseen$Borrowings/unseen$Total.capital
unseen$Borrowings.per.total.asset <- unseen$Borrowings/unseen$Total.assets
unseen$capital.Employed.per.total.income <-
unseen$Capital.employed/unseen$Total.income
unseen$cash.profit.per.total.asset <- unseen$Cash.profit/unseen$Total.assets
unseen$Change.in.stock.per.total.Income <-
unseen$Change.in.stock/unseen$Total.income
unseen$Cumulative.retained.profits.per.sale <-
unseen$Cumulative.retained.profits/unseen$Sales
unseen$Current.assets.per.total.asset <-
unseen$Current.assets/unseen$Total.assets
unseen$Current.liabilities...provisions.per.total.asset <-
unseen$Current.liabilities...provisions/unseen$Total.assets
unseen$Current.liabilities...provisions.per.current.asset <-
unseen$Current.liabilities...provisions/unseen$Current.assets
unseen$Net.fixed.assets.per.total.asset <-
unseen$Net.fixed.assets/unseen$Total.assets
unseen$Net.working.capital.per.total.asset <-
unseen$Net.working.capital/unseen$Total.assets
unseen$Net.working.capital.per.total.capital <-
unseen$Net.working.capital/unseen$Total.capital
unseen$PAT.per.Sales <- unseen$Profit.after.tax/unseen$Sales
unseen$PAT.per.total.asset <- unseen$Profit.after.tax/unseen$Total.assets
unseen$PAT.per.total.income <- unseen$Profit.after.tax/unseen$Total.income
unseen$PBDITA.per.Sales <- unseen$PBDITA/unseen$Sales
unseen$PBT.per.Sales <- unseen$PBT/unseen$Sales
unseen$PBT.per.total.asset <- unseen$PBT/unseen$Total.assets
unseen$PBT.per.Total.Capital <- unseen$PBT/unseen$Total.capital
unseen$Sales.per.total.asset <- unseen$Sales/unseen$Total.assets
unseen$Net.working.capital.per.sales <-
unseen$Net.working.capital/unseen$Sales
unseen$Shareholder.fund.per.total.asset <-
unseen$Shareholders.funds/unseen$Total.assets
unseen$Shareholder.fund.per.total.capital <-
unseen$Shareholders.funds/unseen$Total.capital
unseen$Total.asset.per.Current.Liability <-
unseen$Total.assets/unseen$Current.liabilities...provisions
unseen$Total.Income.per.Shareholder.fund <-
unseen$Total.income/unseen$Shareholders.funds
unseen$Total.income.per.total.asset <-
unseen$Total.income/unseen$Total.assets
unseen$Total.income.per.Sale <- unseen$Total.income/unseen$Sales
unseen$Total.income.per.total.expense <-
unseen$Total.income/unseen$Total.expenses
unseen$Total.liabilities.per.shareholder.fund <-
unseen$Total.liabilities/unseen$Shareholders.funds
```

Predicting the values in the validation dataset using the final model build.

```
# Predicting the probability on Validation dataset

pred_unseen <- predict(glmModel1,newdata = unseen)
unseen$probs <- pred_unseen
pred_unseen1 <- as.factor(ifelse(pred_unseen<0.6,0,1))
unseen$Predicted <- pred_unseen1
default <- as.factor(unseen$Defaul)
```

Creating the Confusion Matrix on the Validation dataset.

```
# Creating Confusion Matrix on Validation dataset

caret::confusionMatrix(pred_unseen1,default,positive='1')

## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 556   4
##          1  74  35
##
##                Accuracy : 0.8834
##                  95% CI : (0.8566, 0.9067)
##     No Information Rate : 0.9417
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.4235
##
##  Mcnemar's Test P-Value : 5.597e-15
##
##             Sensitivity : 0.89744
##             Specificity : 0.88254
##          Pos Pred Value : 0.32110
##          Neg Pred Value : 0.99286
##              Prevalence : 0.05830
##          Detection Rate : 0.05232
##    Detection Prevalence : 0.16293
##       Balanced Accuracy : 0.88999
##
##        'Positive' Class : 1
##
```

From the above confusion Matrix, we have achieved the accuracy of 88.3% with sensitivity of 89% and specificity of 88.2%, the model

Grouping the observations based on Probability in Descending Order

We here divide the data into 10 groups based on the probability in descending order.

```
# Decile the data on probability

prob <- seq (0,1, length = 11)
prob

##  [1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

qs_train<- quantile(unseen$probs,prob,na.rm = TRUE)
unseen$Decile <- cut(unseen$probs,unique(qs_train), include.lowest = TRUE,
right = FALSE)

unseen.decile <- unseen %>% mutate(quartile = ntile(-unseen$probs, 10))
view(unseen.decile)

table(unseen.decile$Default)

##
##   0   1
## 630  39
```

```
defaulter<- data.table::data.table(unseen.decile)
```

```
unseen1_decile <- defaulter[,list(`# Defaulter` <- sum(Default==1),
                                  Total  <-length(Default)) , by=
Decile][order(Decile)]
unseen1_decile
```

	Groups	Defaults	Non_Default	Total
1:	1	30	37	67
2:	2	7	60	67
3:	3	1	66	67
4:	4	1	66	67
5:	5	0	67	67
6:	6	0	67	67
7:	7	0	67	67
8:	8	0	67	67
9:	9	0	67	67
10:	10	0	66	66

Above, we group the data in the descending order of the probability and Rank the Group with highest probability as 1, 2 for the group and so on. From the above Matrix, we can conclude that the majority of defaulters lies in the group 1, where the probability is high.

At the end, we conclude that Total Income per Sales, Net fixed Asset per Total Assets, Current liabilities and provisions per Current Asset, Current asset per total Asset, Change in Stock per Total Income, Borrowings per Total Assets, Cash to current liabilities and Shareholder funds per Total Asset plays an important role in predicting the Companies risk of getting Default. We see that the Model built is working well and can be further used as a Credit Risk Model for the financial Institutes to understand the risk attached.

4 Appendix A – Source Code

```
# Deploying the Libraries

library(readxl)
library(mice)

##
## Attaching package: 'mice'

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

library(corrplot)

## corrplot 0.84 loaded

library(DMwR)

## Loading required package: lattice

## Loading required package: grid

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

library(e1071)
library(car)
```

```
## Loading required package: carData

library(blorr)
library(MASS)
library(caret)

## Loading required package: ggplot2

library(psych)

##
## Attaching package: 'psych'

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

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

library(tidyverse)

## -- Attaching packages ------------------------------ tidyverse 1.3.0 --

## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
## v purrr   0.3.3

## -- Conflicts --------------------------------- tidyverse_conflicts() --
## x psych::%+%()    masks ggplot2::%+%()
## x psych::alpha()  masks ggplot2::alpha()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x purrr::lift()   masks caret::lift()
## x dplyr::recode() masks car::recode()
## x dplyr::select() masks MASS::select()
## x purrr::some()   masks car::some()

library(DataExplorer)

# Setting up the Working Directory.

setwd("D:/Great Learning/Finance and Risk Analytics")

# Reading the Raw data.

raw_data <- read_xlsx("raw-data.xlsx")

# Transforming the names of the Variables

colnames(raw_data)<- make.names(colnames(raw_data))

# Creating Default variable basis of Net Worth Next Year

raw_data$Default <- ifelse(raw_data$Networth.Next.Year <0,1,0)
```

```r
#Perfroming Exploratory Data Analysis

# Checking the first 10 rows of the dataset

head(raw_data,n=10)
```

```
## # A tibble: 10 x 53
##      Num Networth.Next.Y~ Total.assets Net.worth Total.income
##    <dbl>            <dbl>        <dbl>     <dbl>        <dbl>
##  1     1           8891.       17512.     7093.       24965.
##  2     2            394.         941       352.        1527.
##  3     3             92.2        233.      101.         477.
##  4     4              2.7          2.7       2.7          NA
##  5     5            109          478.      108.        1580.
##  6     6            689.        2434.      676.        2649.
##  7     7            246          327.      245.          NA
##  8     8             13.7         80        12.7        154.
##  9     9            292.         574.      239.         583.
## 10    10             -7.3         88.6      19.6         83.4
## # ... with 48 more variables: Change.in.stock <dbl>, Total.expenses <dbl>,
## #   Profit.after.tax <dbl>, PBDITA <dbl>, PBT <dbl>, Cash.profit <dbl>,
## #   PBDITA.as...of.total.income <dbl>, PBT.as...of.total.income <dbl>,
## #   PAT.as...of.total.income <dbl>,
## #   Cash.profit.as...of.total.income <dbl>, PAT.as...of.net.worth <dbl>,
## #   Sales <dbl>, Income.from.financial.services <dbl>, Other.income <dbl>,
## #   Total.capital <dbl>, Reserves.and.funds <dbl>,
## #   Deposits..accepted.by.commercial.banks. <lgl>, Borrowings <dbl>,
## #   Current.liabilities...provisions <dbl>, Deferred.tax.liability <dbl>,
## #   Shareholders.funds <dbl>, Cumulative.retained.profits <dbl>,
## #   Capital.employed <dbl>, TOL.TNW <dbl>,
## #   Total.term.liabilities...tangible.net.worth <dbl>,
## #   Contingent.liabilities...Net.worth.... <dbl>,
## #   Contingent.liabilities <dbl>, Net.fixed.assets <dbl>,
## #   Investments <dbl>, Current.assets <dbl>, Net.working.capital <dbl>,
## #   Quick.ratio..times. <dbl>, Current.ratio..times. <dbl>,
## #   Debt.to.equity.ratio..times. <dbl>,
## #   Cash.to.current.liabilities..times. <dbl>,
## #   Cash.to.average.cost.of.sales.per.day <dbl>, Creditors.turnover <chr>,
## #   Debtors.turnover <chr>, Finished.goods.turnover <chr>,
## #   WIP.turnover <chr>, Raw.material.turnover <chr>,
## #   Shares.outstanding <chr>, Equity.face.value <chr>, EPS <dbl>,
## #   Adjusted.EPS <dbl>, Total.liabilities <dbl>, PE.on.BSE <chr>,
## #   Default <dbl>
```

```r
# Checking the dimensions of the data

dim(raw_data)
```

```
## [1] 3541   53
```

```r
# Understanding the structure of the dataset

str(raw_data)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    3541 obs. of  53 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 1580 ...
##  $ Change.in.stock                              : num  235.8 42.7 -5.2 NA -17 ...
##  $ Total.expenses                               : num  23658 1455 479 NA 1558 ...
##  $ 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 11.9 ...
##  $ 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 0.35 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 1575 ...
##  $ Income.from.financial.services               : num  158 4 1.5 NA 3.9 6.4 NA NA 7.3 N
A ...
##  $ 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 193.1 ...
##  $ Current.liabilities...provisions             : num  9965.9 210 96.8 NA 112.8 ...
##  $ 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 2.7 300.7 ...
##  $ 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 0.9
4 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 37.6 ...
##  $ 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 1.41 0.48 NA 0
.54 0.59 0.39 ...
##  $ Current.ratio..times.                        : num  1.37 1.56 1.55 NA 2.54 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 2.3
1 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" "9.8000000000000007" "5.2
8" "0" ...
##  $ Debtors.turnover                             : chr  "3.85" "5.7" "5.07" "0" ...
```

```
##  $ Finished.goods.turnover                    : chr  "200.55" "14.21" "9.24" 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 7.91 ...
##  $ 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" ...
##  $ Default                                    : num  0 0 0 0 0 0 0 0 0 1 ...

# Understanding the basic summary of the dataset

summary(raw_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            Default
##  Length:3541        Min.   :0.00000
##  Class :character   1st Qu.:0.00000
##  Mode  :character   Median :0.00000
##                     Mean   :0.06608
##                     3rd Qu.:0.00000
##                     Max.   :1.00000
##
```

```
# Checking for the Missing Values in the dataset
colSums(is.na(raw_data))
```

```
##                                       Num
##                                         0
##                         Networth.Next.Year
##                                         0
##                               Total.assets
##                                         0
##                                  Net.worth
##                                         0
##                               Total.income
##                                       198
##                             Change.in.stock
##                                       458
##                              Total.expenses
##                                       139
##                            Profit.after.tax
##                                       131
##                                     PBDITA
##                                       131
##                                        PBT
##                                       131
##                                 Cash.profit
##                                       131
##                 PBDITA.as...of.total.income
##                                        68
##                    PBT.as...of.total.income
##                                        68
##                    PAT.as...of.total.income
##                                        68
##            Cash.profit.as...of.total.income
##                                        68
##                       PAT.as...of.net.worth
##                                         0
##                                     Sales
##                                       259
##               Income.from.financial.services
##                                       935
##                                Other.income
##                                      1295
```

```
##                                Total.capital
##                                            4
##                            Reserves.and.funds
##                                           85
##     Deposits..accepted.by.commercial.banks.
##                                         3541
##                                   Borrowings
##                                          366
##            Current.liabilities...provisions
##                                           96
##                        Deferred.tax.liability
##                                         1140
##                            Shareholders.funds
##                                            0
##                  Cumulative.retained.profits
##                                           38
##                               Capital.employed
##                                            0
##                                      TOL.TNW
##                                            0
## Total.term.liabilities...tangible.net.worth
##                                            0
##      Contingent.liabilities...Net.worth....
##                                            0
##                        Contingent.liabilities
##                                         1188
##                               Net.fixed.assets
##                                          118
##                                    Investments
##                                         1435
##                                 Current.assets
##                                           66
##                           Net.working.capital
##                                           32
##                           Quick.ratio..times.
##                                           93
##                         Current.ratio..times.
##                                           93
##                  Debt.to.equity.ratio..times.
##                                            0
##         Cash.to.current.liabilities..times.
##                                           93
##       Cash.to.average.cost.of.sales.per.day
##                                           85
##                              Creditors.turnover
##                                           47
##                                Debtors.turnover
##                                           42
##                        Finished.goods.turnover
##                                          454
##                                   WIP.turnover
##                                          354
##                           Raw.material.turnover
##                                           75
```

```
##                          Shares.outstanding
##                                           0
##                           Equity.face.value
##                                           0
##                                         EPS
##                                           0
##                                Adjusted.EPS
##                                           0
##                           Total.liabilities
##                                           0
##                                  PE.on.BSE
##                                          23
##                                    Default
##                                           0
```

```
# Removing the unnecessary variables and variables with missing value greater then 25%
names(raw_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"
## [53] "Default"

data <- raw_data[,-c(1,2,18,19,22,25,32,34,42:47)]

# Converting datatypes of the variables
data$Equity.face.value <- as.numeric(data$Equity.face.value)

## Warning: NAs introduced by coercion

data$PE.on.BSE <- as.numeric(data$PE.on.BSE)

## Warning: NAs introduced by coercion

data$Default <- as.factor(data$Default)

# Imputing the missing Values.

data1 <- mice(data,m=5,maxit=10,meth='pmm',seed=500)

data_new <- complete(data1,1)
colSums(is.na(data_new))

##                                Total.assets
##                                           0
##                                   Net.worth
##                                           0
##                                Total.income
##                                         198
##                             Change.in.stock
##                                           0
##                              Total.expenses
##                                           0
##                            Profit.after.tax
##                                           0
##                                      PBDITA
##                                           0
##                                         PBT
##                                           0
##                                  Cash.profit
##                                           0
##                 PBDITA.as...of.total.income
```

```
##                                           0
##                     PBT.as...of.total.income
##                                           0
##                     PAT.as...of.total.income
##                                           0
##             Cash.profit.as...of.total.income
##                                           0
##                        PAT.as...of.net.worth
##                                           0
##                                       Sales
##                                         259
##                                Total.capital
##                                           0
##                             Reserves.and.funds
##                                           0
##                                   Borrowings
##                                           0
##             Current.liabilities...provisions
##                                           0
##                             Shareholders.funds
##                                           0
##                  Cumulative.retained.profits
##                                           0
##                               Capital.employed
##                                           0
##                                      TOL.TNW
##                                           0
## Total.term.liabilities...tangible.net.worth
##                                           0
##      Contingent.liabilities...Net.worth....
##                                           0
##                               Net.fixed.assets
##                                           0
##                                 Current.assets
##                                           0
##                            Net.working.capital
##                                           0
##                            Quick.ratio..times.
##                                           0
##                         Current.ratio..times.
##                                           0
##                 Debt.to.equity.ratio..times.
##                                           0
##         Cash.to.current.liabilities..times.
##                                           0
##       Cash.to.average.cost.of.sales.per.day
##                                           0
##                              Equity.face.value
##                                           0
##                                         EPS
##                                           0
##                                  Adjusted.EPS
##                                           0
##                              Total.liabilities
```

```
##                                          0
##                                   PE.on.BSE
##                                          0
##                                    Default
##                                          0
```

```
str(data)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    3541 obs. of  39 variables:
##  $ 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 1580 ...
##  $ Change.in.stock                             : num  235.8 42.7 -5.2 NA -17 ...
##  $ Total.expenses                              : num  23658 1455 479 NA 1558 ...
##  $ 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 11.9 ...
##  $ 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 0.35 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 1575 ...
##  $ 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 ...
##  $ Borrowings                                  : num  14.9 272.5 35.4 NA 193.1 ...
##  $ Current.liabilities...provisions            : num  9965.9 210 96.8 NA 112.8 ...
##  $ 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 2.7 300.7 ...
##  $ 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 0.9
4 0.3 0.54 ...
##  $ Contingent.liabilities...Net.worth....      : num  14.8 19.2 45.8 0 34.9 ...
##  $ Net.fixed.assets                            : num  1900.2 286.4 38.7 2.5 94.8 ...
##  $ 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 1.41 0.48 NA 0
.54 0.59 0.39 ...
##  $ Current.ratio..times.                       : num  1.37 1.56 1.55 NA 2.54 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 2.3
1 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 ...
##  $ Equity.face.value                           : num  10 10 10 10 10 10 10 NA 10 10 ..
.
##  $ EPS                                         : num  35.52 9.97 -0.5 0 7.91 ...
##  $ 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                                       : num  27.31 8.17 -5.76 NA NA ...
##  $ Default                                         : Factor w/ 2 levels "0","1": 1 1 1 1 1
 1 1 1 1 2 ...

data_new <- na.omit(data_new)
dim(data_new)

## [1] 3282   39

names(data)

##  [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] "Total.capital"
## [17] "Reserves.and.funds"
## [18] "Borrowings"
## [19] "Current.liabilities...provisions"
## [20] "Shareholders.funds"
## [21] "Cumulative.retained.profits"
## [22] "Capital.employed"
## [23] "TOL.TNW"
## [24] "Total.term.liabilities...tangible.net.worth"
## [25] "Contingent.liabilities...Net.worth...."
## [26] "Net.fixed.assets"
## [27] "Current.assets"
## [28] "Net.working.capital"
## [29] "Quick.ratio..times."
## [30] "Current.ratio..times."
## [31] "Debt.to.equity.ratio..times."
## [32] "Cash.to.current.liabilities..times."
## [33] "Cash.to.average.cost.of.sales.per.day"
## [34] "Equity.face.value"
## [35] "EPS"
## [36] "Adjusted.EPS"
## [37] "Total.liabilities"
## [38] "PE.on.BSE"
## [39] "Default"

#Perfroming Univariate Analysis
dev.off()

## null device
##           1
```

```r
plot_histogram(data_new[1:8])

boxplot(data_new[1:8],horizontal = TRUE)

plot_histogram(data_new[9:16])

boxplot(data_new[9:16],horizontal = TRUE)

plot_histogram(data_new[17:24])

boxplot(data_new[17:24],horizontal = TRUE)

plot_histogram(data_new[25:32])

boxplot(data_new[25:32],horizontal = TRUE)

plot_histogram(data_new[33:39])

boxplot(data_new[33:39],horizontal = TRUE)


#Performing Bi-Variate Analaysis

ggplot(data_new, aes(x=data_new$Total.assets)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,15000)

## Warning: Removed 100 rows containing non-finite values (stat_density).

boxplot(data_new$Total.assets~data_new$Default, horizontal = TRUE)

ggplot(data_new, aes(x=data_new$Net.worth)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,10000)

## Warning: Removed 57 rows containing non-finite values (stat_density).

boxplot(data_new$Net.worth~data_new$Default, horizontal = TRUE)

ggplot(data_new, aes(x=data_new$Total.income)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,10000)

## Warning: Removed 153 rows containing non-finite values (stat_density).

boxplot(data_new$Total.income~data_new$Default, horizontal = TRUE)


ggplot(data_new, aes(x=data_new$Total.expenses)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
```

```
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,10000)

## Warning: Removed 138 rows containing non-finite values (stat_density).

boxplot(data_new$Total.expenses~data_new$Default, horizontal = TRUE)


ggplot(data_new, aes(x=data_new$Cash.profit)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,2000)

## Warning: Removed 248 rows containing non-finite values (stat_density).

boxplot(data_new$Cash.profit~data_new$Default, horizontal = TRUE)


ggplot(data_new, aes(x=data_new$Profit.after.tax)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,2000)

## Warning: Removed 327 rows containing non-finite values (stat_density).

boxplot(data_new$Profit.after.tax~data_new$Default, horizontal = TRUE)


ggplot(data_new, aes(x=data_new$Sales)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,40000)

## Warning: Removed 36 rows containing non-finite values (stat_density).

boxplot(data_new$Sales~data_new$Default, horizontal = TRUE)


ggplot(data_new, aes(x=data_new$Reserves.and.funds)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,5000)

## Warning: Removed 467 rows containing non-finite values (stat_density).

boxplot(data_new$Reserves.and.funds~data_new$Default, horizontal = TRUE)


ggplot(data_new, aes(x=data_new$Current.liabilities...provisions)) +
  geom_density(aes(fill =data_new$Default, alpha = 0.3)) +
  scale_color_manual(values = c("#868686FF", "#EFC000FF")) +
```

```
  scale_fill_manual(values = c("darkturquoise","lightcoral","lightgreen"))+
  xlim(-5,5000)
```

```
## Warning: Removed 77 rows containing non-finite values (stat_density).
```

```
boxplot(data_new$Current.liabilities...provisions~data_new$Default, horizontal = TRUE)
```

```
#Checking for the Multicollinearity
```

```
plot_correlation(data_new)
```

```
data <- data_new
dim(data)
```

```
## [1] 3282   39
```

```
corrplot(cor(data[,-39]),method = "number", type = "lower", number.cex = 0.7)
```

```
# Creating new variables using the standalone Variables
```

```
data$Borrowings.per.total.Capital <- data$Borrowings/data$Total.capital
data$Borrowings.per.total.asset <- data$Borrowings/data$Total.assets
data$capital.Employed.per.total.income <- data$Capital.employed/data$Total.income
data$cash.profit.per.total.asset <- data$Cash.profit/data$Total.assets
data$Change.in.stock.per.total.Income <- data$Change.in.stock/data$Total.income
data$Cumulative.retained.profits.per.sale <- data$Cumulative.retained.profits/data$Sales
data$Current.assets.per.total.asset <- data$Current.assets/data$Total.assets
data$Current.liabilities...provisions.per.total.asset <- data$Current.liabilities...prov
isions/data$Total.assets
data$Current.liabilities...provisions.per.current.asset <- data$Current.liabilities...pr
ovisions/data$Current.assets
data$Net.fixed.assets.per.total.asset <- data$Net.fixed.assets/data$Total.assets
data$Net.working.capital.per.total.asset <- data$Net.working.capital/data$Total.assets
data$Net.working.capital.per.total.capital <- data$Net.working.capital/data$Total.capita
l
data$PAT.per.Sales <- data$Profit.after.tax/data$Sales
data$PAT.per.total.asset <- data$Profit.after.tax/data$Total.assets
data$PBDITA.per.Sales <- data$PBDITA/data$Sales
data$PBT.per.Sales <- data$PBT/data$Sales
data$PBT.per.total.asset <- data$PBT/data$Total.assets
data$PBT.per.Total.Capital <- data$PBT/data$Total.capital
data$Sales.per.total.asset <- data$Sales/data$Total.assets
data$Net.working.capital.per.sales <- data$Net.working.capital/data$Sales
data$Shareholder.fund.per.total.asset <- data$Shareholders.funds/data$Total.assets
data$Shareholder.fund.per.total.capital <- data$Shareholders.funds/data$Total.capital
data$Total.asset.per.Current.Liability <- data$Total.assets/data$Current.liabilities...p
rovisions
data$Total.Income.per.Shareholder.fund <- data$Total.income/data$Shareholders.funds
data$Total.income.per.total.asset <- data$Total.income/data$Total.assets
data$Total.income.per.Sale <- data$Total.income/data$Sales
data$Total.income.per.total.expense <- data$Total.income/data$Total.expenses
data$Total.liabilities.per.shareholder.fund <- data$Total.liabilities/data$Shareholders.
funds
```

```r
dataset <- data
dim(dataset)
```

```
## [1] 3282   67
```

```r
names(dataset)
```

```
##  [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] "Total.capital"
## [17] "Reserves.and.funds"
## [18] "Borrowings"
## [19] "Current.liabilities...provisions"
## [20] "Shareholders.funds"
## [21] "Cumulative.retained.profits"
## [22] "Capital.employed"
## [23] "TOL.TNW"
## [24] "Total.term.liabilities...tangible.net.worth"
## [25] "Contingent.liabilities...Net.worth...."
## [26] "Net.fixed.assets"
## [27] "Current.assets"
## [28] "Net.working.capital"
## [29] "Quick.ratio..times."
## [30] "Current.ratio..times."
## [31] "Debt.to.equity.ratio..times."
## [32] "Cash.to.current.liabilities..times."
## [33] "Cash.to.average.cost.of.sales.per.day"
## [34] "Equity.face.value"
## [35] "EPS"
## [36] "Adjusted.EPS"
## [37] "Total.liabilities"
## [38] "PE.on.BSE"
## [39] "Default"
## [40] "Borrowings.per.total.Capital"
## [41] "Borrowings.per.total.asset"
## [42] "capital.Employed.per.total.income"
## [43] "cash.profit.per.total.asset"
## [44] "Change.in.stock.per.total.Income"
## [45] "Cumulative.retained.profits.per.sale"
## [46] "Current.assets.per.total.asset"
```

```
## [47] "Current.liabilities...provisions.per.total.asset"
## [48] "Current.liabilities...provisions.per.current.asset"
## [49] "Net.fixed.assets.per.total.asset"
## [50] "Net.working.capital.per.total.asset"
## [51] "Net.working.capital.per.total.capital"
## [52] "PAT.per.Sales"
## [53] "PAT.per.total.asset"
## [54] "PBDITA.per.Sales"
## [55] "PBT.per.Sales"
## [56] "PBT.per.total.asset"
## [57] "PBT.per.Total.Capital"
## [58] "Sales.per.total.asset"
## [59] "Net.working.capital.per.sales"
## [60] "Shareholder.fund.per.total.asset"
## [61] "Shareholder.fund.per.total.capital"
## [62] "Total.asset.per.Current.Liability"
## [63] "Total.Income.per.Shareholder.fund"
## [64] "Total.income.per.total.asset"
## [65] "Total.income.per.Sale"
## [66] "Total.income.per.total.expense"
## [67] "Total.liabilities.per.shareholder.fund"
```

```
# Treating the Outliers by Capping and Flooring method

dataset2 <- dataset
names(dataset2)
```

```
##  [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] "Total.capital"
## [17] "Reserves.and.funds"
## [18] "Borrowings"
## [19] "Current.liabilities...provisions"
## [20] "Shareholders.funds"
## [21] "Cumulative.retained.profits"
## [22] "Capital.employed"
## [23] "TOL.TNW"
## [24] "Total.term.liabilities...tangible.net.worth"
## [25] "Contingent.liabilities...Net.worth...."
## [26] "Net.fixed.assets"
## [27] "Current.assets"
## [28] "Net.working.capital"
```

```
## [29] "Quick.ratio..times."
## [30] "Current.ratio..times."
## [31] "Debt.to.equity.ratio..times."
## [32] "Cash.to.current.liabilities..times."
## [33] "Cash.to.average.cost.of.sales.per.day"
## [34] "Equity.face.value"
## [35] "EPS"
## [36] "Adjusted.EPS"
## [37] "Total.liabilities"
## [38] "PE.on.BSE"
## [39] "Default"
## [40] "Borrowings.per.total.Capital"
## [41] "Borrowings.per.total.asset"
## [42] "capital.Employed.per.total.income"
## [43] "cash.profit.per.total.asset"
## [44] "Change.in.stock.per.total.Income"
## [45] "Cumulative.retained.profits.per.sale"
## [46] "Current.assets.per.total.asset"
## [47] "Current.liabilities...provisions.per.total.asset"
## [48] "Current.liabilities...provisions.per.current.asset"
## [49] "Net.fixed.assets.per.total.asset"
## [50] "Net.working.capital.per.total.asset"
## [51] "Net.working.capital.per.total.capital"
## [52] "PAT.per.Sales"
## [53] "PAT.per.total.asset"
## [54] "PBDITA.per.Sales"
## [55] "PBT.per.Sales"
## [56] "PBT.per.total.asset"
## [57] "PBT.per.Total.Capital"
## [58] "Sales.per.total.asset"
## [59] "Net.working.capital.per.sales"
## [60] "Shareholder.fund.per.total.asset"
## [61] "Shareholder.fund.per.total.capital"
## [62] "Total.asset.per.Current.Liability"
## [63] "Total.Income.per.Shareholder.fund"
## [64] "Total.income.per.total.asset"
## [65] "Total.income.per.Sale"
## [66] "Total.income.per.total.expense"
## [67] "Total.liabilities.per.shareholder.fund"
```

```
a <- c(1:38,40:67)

for(val in a){
  qnt<- quantile(dataset2[,val],probs = c(0.25,0.75))
  cap<- quantile(dataset2[,val],probs = c(0.05,0.95))

  h= 1.5*IQR(dataset2[,val])
  dataset2[,val][dataset2[,val]>(qnt[2]+h)]<- cap[2]
  dataset2[,val][dataset2[,val]<(qnt[1]-h)]<- cap[1]
}


# Checking the Proportion on Default.

table(dataset2$Default)
```

```
##
##    0    1
## 3104  178
```

```
prop.table(table(dataset2$Default))
```

```
##
##          0          1
## 0.94576478 0.05423522
```

```
# Balancing the dataset.
anyNA(dataset2)
```

```
## [1] FALSE
```

```
bal2 <- SMOTE(dataset2$Default~., data = dataset2,k=5)
```

```
## Warning in names(data) == as.character(form[[2]]): longer object length is
## not a multiple of shorter object length
```

```
#bal2 <-ROSE(Default~., data=dataset2, seed=3)$data
table(bal2$Default)
```

```
##
##   0   1
## 712 534
```

```
prop.table(table(bal2$Default))
```

```
##
##         0         1
## 0.5714286 0.4285714
```

```
#Building the Model with the dataset after performing the EDA.

attach(bal2)
```

```
glmModel<- glm(bal2$Default~.,data = bal2, family = "binomial")
```

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

```
summary(glmModel)
```

```
##
## Call:
## glm(formula = bal2$Default ~ ., family = "binomial", data = bal2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.6205  -0.1663   0.0000   0.1172   2.8098
##
## Coefficients: (1 not defined because of singularities)
##                                                  Estimate Std. Error
## (Intercept)                                    -1.845e+00  6.141e+00
## Total.assets                                    3.206e-04  3.253e-04
## Net.worth                                      -1.190e-03  2.006e-03
## Total.income                                    2.819e-03  1.498e-03
```

```
## Change.in.stock                                          8.808e-03  5.265e-03
## Total.expenses                                          -1.287e-04  1.656e-03
## Profit.after.tax                                        -7.861e-03  1.436e-02
## PBDITA                                                  -4.867e-03  3.534e-03
## PBT                                                      6.685e-03  6.266e-03
## Cash.profit                                             -1.632e-02  9.371e-03
## PBDITA.as...of.total.income                              3.355e-01  2.694e-01
## PBT.as...of.total.income                                 6.444e-01  4.412e-01
## PAT.as...of.total.income                                -1.217e+00  6.038e-01
## Cash.profit.as...of.total.income                         1.466e-02  6.535e-02
## PAT.as...of.net.worth                                   -7.910e-02  1.590e-02
## Sales                                                   -2.155e-03  2.069e-03
## Total.capital                                           -7.639e-03  2.362e-03
## Reserves.and.funds                                      -6.436e-04  1.696e-03
## Borrowings                                              -8.074e-04  7.520e-04
## Current.liabilities...provisions                         1.033e-03  1.140e-03
## Shareholders.funds                                       8.798e-04  2.777e-03
## Cumulative.retained.profits                             -1.556e-02  4.422e-03
## Capital.employed                                         1.891e-03  1.604e-03
## TOL.TNW                                                  2.373e-01  1.359e-01
## Total.term.liabilities...tangible.net.worth             -9.422e-01  2.698e-01
## Contingent.liabilities...Net.worth....                   5.378e-03  2.778e-03
## Net.fixed.assets                                        -1.620e-04  6.596e-04
## Current.assets                                          -2.290e-03  1.107e-03
## Net.working.capital                                      6.038e-03  1.977e-03
## Quick.ratio..times.                                      1.585e-01  5.048e-01
## Current.ratio..times.                                   -9.610e-01  4.554e-01
## Debt.to.equity.ratio..times.                             5.158e-01  2.153e-01
## Cash.to.current.liabilities..times.                      1.092e+00  8.296e-01
## Cash.to.average.cost.of.sales.per.day                   -5.126e-03  5.049e-03
## Equity.face.value                                        6.284e-04  6.539e-03
## EPS                                                     -3.704e-02  6.627e-02
## Adjusted.EPS                                             4.371e-02  6.735e-02
## Total.liabilities                                               NA         NA
## PE.on.BSE                                               -1.107e-02  4.789e-03
## Borrowings.per.total.Capital                            -7.442e-02  4.511e-02
## Borrowings.per.total.asset                               6.700e+00  1.449e+00
## capital.Employed.per.total.income                       -4.268e-02  2.074e-01
## cash.profit.per.total.asset                             -1.682e+00  8.437e+00
## Change.in.stock.per.total.Income                        -1.035e+00  3.753e+00
## Cumulative.retained.profits.per.sale                    -2.098e-01  4.977e-01
## Current.assets.per.total.asset                           1.654e+00  1.581e+00
## Current.liabilities...provisions.per.total.asset         6.276e+00  2.549e+00
## Current.liabilities...provisions.per.current.asset      -4.748e-01  8.599e-01
## Net.fixed.assets.per.total.asset                         4.979e+00  1.139e+00
## Net.working.capital.per.total.asset                     -2.627e-01  2.251e+00
## Net.working.capital.per.total.capital                    4.239e-02  9.668e-02
## PAT.per.Sales                                            9.909e+01  5.642e+01
## PAT.per.total.asset                                      6.790e+01  1.904e+01
## PBDITA.per.Sales                                        -3.134e+01  2.536e+01
## PBT.per.Sales                                           -5.031e+01  4.146e+01
## PBT.per.total.asset                                     -4.566e+01  1.476e+01
## PBT.per.Total.Capital                                   -4.526e-01  2.649e-01
## Sales.per.total.asset                                   -2.160e+01  5.275e+00
```

```
## Net.working.capital.per.sales                          9.853e-01  1.233e+00
## Shareholder.fund.per.total.asset                       8.741e-01  1.647e+00
## Shareholder.fund.per.total.capital                     5.621e-02  4.427e-02
## Total.asset.per.Current.Liability                     -4.859e-02  4.344e-02
## Total.Income.per.Shareholder.fund                     -5.616e-02  7.825e-02
## Total.income.per.total.asset                           2.211e+01  5.217e+00
## Total.income.per.Sale                                 -3.069e+00  4.444e+00
## Total.income.per.total.expense                        -2.876e+00  4.035e+00
## Total.liabilities.per.shareholder.fund                 9.551e-02  1.638e-01
##                                                       z value Pr(>|z|)
## (Intercept)                                            -0.300 0.763833
## Total.assets                                            0.986 0.324343
## Net.worth                                              -0.593 0.553046
## Total.income                                            1.882 0.059802 .
## Change.in.stock                                         1.673 0.094348 .
## Total.expenses                                         -0.078 0.938058
## Profit.after.tax                                       -0.548 0.583947
## PBDITA                                                 -1.377 0.168515
## PBT                                                     1.067 0.286075
## Cash.profit                                            -1.741 0.081675 .
## PBDITA.as...of.total.income                             1.245 0.212963
## PBT.as...of.total.income                                1.461 0.144101
## PAT.as...of.total.income                               -2.015 0.043917 *
## Cash.profit.as...of.total.income                        0.224 0.822450
## PAT.as...of.net.worth                                  -4.976 6.51e-07 ***
## Sales                                                  -1.042 0.297596
## Total.capital                                          -3.234 0.001222 **
## Reserves.and.funds                                     -0.379 0.704359
## Borrowings                                             -1.074 0.282968
## Current.liabilities...provisions                        0.907 0.364468
## Shareholders.funds                                      0.317 0.751391
## Cumulative.retained.profits                            -3.519 0.000434 ***
## Capital.employed                                        1.179 0.238579
## TOL.TNW                                                 1.747 0.080663 .
## Total.term.liabilities...tangible.net.worth            -3.493 0.000478 ***
## Contingent.liabilities...Net.worth....                  1.936 0.052912 .
## Net.fixed.assets                                       -0.246 0.806025
## Current.assets                                         -2.069 0.038541 *
## Net.working.capital                                     3.054 0.002255 **
## Quick.ratio..times.                                     0.314 0.753491
## Current.ratio..times.                                  -2.110 0.034854 *
## Debt.to.equity.ratio..times.                            2.395 0.016608 *
## Cash.to.current.liabilities..times.                     1.316 0.188208
## Cash.to.average.cost.of.sales.per.day                  -1.015 0.309959
## Equity.face.value                                       0.096 0.923450
## EPS                                                    -0.559 0.576204
## Adjusted.EPS                                            0.649 0.516369
## Total.liabilities                                          NA       NA
## PE.on.BSE                                              -2.311 0.020833 *
## Borrowings.per.total.Capital                           -1.650 0.099009 .
## Borrowings.per.total.asset                              4.625 3.74e-06 ***
## capital.Employed.per.total.income                      -0.206 0.836944
## cash.profit.per.total.asset                            -0.199 0.841940
## Change.in.stock.per.total.Income                       -0.276 0.782749
```

```
## Cumulative.retained.profits.per.sale                -0.422 0.673350
## Current.assets.per.total.asset                       1.046 0.295504
## Current.liabilities...provisions.per.total.asset     2.463 0.013797 *
## Current.liabilities...provisions.per.current.asset  -0.552 0.580839
## Net.fixed.assets.per.total.asset                     4.372 1.23e-05 ***
## Net.working.capital.per.total.asset                 -0.117 0.907089
## Net.working.capital.per.total.capital                0.438 0.661088
## PAT.per.Sales                                        1.756 0.079037 .
## PAT.per.total.asset                                  3.567 0.000362 ***
## PBDITA.per.Sales                                    -1.236 0.216440
## PBT.per.Sales                                       -1.213 0.225031
## PBT.per.total.asset                                 -3.093 0.001982 **
## PBT.per.Total.Capital                               -1.708 0.087567 .
## Sales.per.total.asset                               -4.095 4.22e-05 ***
## Net.working.capital.per.sales                        0.799 0.424140
## Shareholder.fund.per.total.asset                     0.531 0.595610
## Shareholder.fund.per.total.capital                   1.270 0.204157
## Total.asset.per.Current.Liability                   -1.119 0.263303
## Total.Income.per.Shareholder.fund                   -0.718 0.472921
## Total.income.per.total.asset                         4.238 2.25e-05 ***
## Total.income.per.Sale                               -0.691 0.489779
## Total.income.per.total.expense                      -0.713 0.476040
## Total.liabilities.per.shareholder.fund               0.583 0.559897
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1701.81  on 1245  degrees of freedom
## Residual deviance:  441.78  on 1180  degrees of freedom
## AIC: 573.78
##
## Number of Fisher Scoring iterations: 12

glmPred <- predict(glmModel,data= bal2)
glmPred1 <- as.factor(ifelse(glmPred <0.6,0,1))
caret::confusionMatrix(glmPred1,as.factor(bal2$Default),positive='1')

## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 687  68
##          1  25 466
##
##                Accuracy : 0.9254
##                  95% CI : (0.9093, 0.9393)
##     No Information Rate : 0.5714
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8461
##
##  Mcnemar's Test P-Value : 1.329e-05
##
##             Sensitivity : 0.8727
```

```
##             Specificity : 0.9649
##          Pos Pred Value : 0.9491
##          Neg Pred Value : 0.9099
##              Prevalence : 0.4286
##          Detection Rate : 0.3740
##    Detection Prevalence : 0.3941
##       Balanced Accuracy : 0.9188
##
##        'Positive' Class : 1
##

lr.Test1 <- lmtest::lrtest(glmModel)
lr.Test1

## Likelihood ratio test
##
## Model 1: bal2$Default ~ Total.assets + Net.worth + Total.income + Change.in.stock +
##     Total.expenses + Profit.after.tax + PBDITA + PBT + Cash.profit +
##     PBDITA.as...of.total.income + PBT.as...of.total.income +
##     PAT.as...of.total.income + Cash.profit.as...of.total.income +
##     PAT.as...of.net.worth + Sales + Total.capital + Reserves.and.funds +
##     Borrowings + Current.liabilities...provisions + Shareholders.funds +
##     Cumulative.retained.profits + Capital.employed + TOL.TNW +
##     Total.term.liabilities...tangible.net.worth + Contingent.liabilities...Net.worth.
... +
##     Net.fixed.assets + Current.assets + Net.working.capital +
##     Quick.ratio..times. + Current.ratio..times. + Debt.to.equity.ratio..times. +
##     Cash.to.current.liabilities..times. + Cash.to.average.cost.of.sales.per.day +
##     Equity.face.value + EPS + Adjusted.EPS + Total.liabilities +
##     PE.on.BSE + Borrowings.per.total.Capital + Borrowings.per.total.asset +
##     capital.Employed.per.total.income + cash.profit.per.total.asset +
##     Change.in.stock.per.total.Income + Cumulative.retained.profits.per.sale +
##     Current.assets.per.total.asset + Current.liabilities...provisions.per.total.asset
 +
##     Current.liabilities...provisions.per.current.asset + Net.fixed.assets.per.total.a
sset +
##     Net.working.capital.per.total.asset + Net.working.capital.per.total.capital +
##     PAT.per.Sales + PAT.per.total.asset + PBDITA.per.Sales +
##     PBT.per.Sales + PBT.per.total.asset + PBT.per.Total.Capital +
##     Sales.per.total.asset + Net.working.capital.per.sales + Shareholder.fund.per.tota
l.asset +
##     Shareholder.fund.per.total.capital + Total.asset.per.Current.Liability +
##     Total.Income.per.Shareholder.fund + Total.income.per.total.asset +
##     Total.income.per.Sale + Total.income.per.total.expense +
##     Total.liabilities.per.shareholder.fund
## Model 2: bal2$Default ~ 1
##   #Df  LogLik  Df Chisq Pr(>Chisq)
## 1  66 -220.89
## 2   1 -850.90 -65  1260  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

rSq1<- pscl::pR2(glmModel)
rSq1
```

```
##          llh      llhNull           G2     McFadden         r2ML
## -220.8894360 -850.9034985 1260.0281249    0.7404060    0.6362391
##         r2CU
##    0.8542098

# Refining the Model using Step Wise AIC.

stepAIC(glmModel,direction = "both")

##
## Call:  glm(formula = bal2$Default ~ Total.income + Change.in.stock +
##     PBDITA + Cash.profit + PBDITA.as...of.total.income + PBT.as...of.total.income +
##     PAT.as...of.total.income + PAT.as...of.net.worth + Sales +
##     Total.capital + Cumulative.retained.profits + Capital.employed +
##     TOL.TNW + Total.term.liabilities...tangible.net.worth + Contingent.liabilities...
Net.worth.... +
##     Current.assets + Net.working.capital + Current.ratio..times. +
##     Debt.to.equity.ratio..times. + PE.on.BSE + Borrowings.per.total.Capital +
##     Borrowings.per.total.asset + Current.assets.per.total.asset +
##     Current.liabilities...provisions.per.total.asset + Net.fixed.assets.per.total.ass
et +
##     PAT.per.Sales + PAT.per.total.asset + PBDITA.per.Sales +
##     PBT.per.Sales + PBT.per.total.asset + Sales.per.total.asset +
##     Total.income.per.total.asset + Cash.to.current.liabilities..times.,
##     family = "binomial", data = bal2)
##
## Coefficients:
##                                       (Intercept)
##                                         -8.115847
##                                      Total.income
##                                          0.002068
##                                   Change.in.stock
##                                          0.009055
##                                            PBDITA
##                                         -0.004369
##                                      Cash.profit
##                                         -0.013698
##                          PBDITA.as...of.total.income
##                                          0.416451
##                             PBT.as...of.total.income
##                                          0.599947
##                             PAT.as...of.total.income
##                                         -1.095204
##                                PAT.as...of.net.worth
##                                         -0.073758
##                                             Sales
##                                         -0.001622
##                                    Total.capital
##                                         -0.006354
##                          Cumulative.retained.profits
##                                         -0.013563
##                                   Capital.employed
##                                          0.001271
##                                           TOL.TNW
##                                          0.190110
```

```
##      Total.term.liabilities...tangible.net.worth
##                                         -0.893462
##           Contingent.liabilities...Net.worth....
##                                          0.005185
##                                    Current.assets
##                                         -0.001153
##                               Net.working.capital
##                                          0.004789
##                             Current.ratio..times.
##                                         -0.501888
##                     Debt.to.equity.ratio..times.
##                                          0.527679
##                                         PE.on.BSE
##                                         -0.010747
##                     Borrowings.per.total.Capital
##                                         -0.054101
##                       Borrowings.per.total.asset
##                                          6.134866
##                   Current.assets.per.total.asset
##                                          1.933700
## Current.liabilities...provisions.per.total.asset
##                                          6.780646
##                 Net.fixed.assets.per.total.asset
##                                          4.722513
##                                     PAT.per.Sales
##                                         85.440010
##                               PAT.per.total.asset
##                                         61.937928
##                                  PBDITA.per.Sales
##                                        -39.766182
##                                     PBT.per.Sales
##                                        -44.599820
##                               PBT.per.total.asset
##                                        -46.528015
##                             Sales.per.total.asset
##                                        -20.745722
##                     Total.income.per.total.asset
##                                         21.059810
##              Cash.to.current.liabilities..times.
##                                          0.825638
##
## Degrees of Freedom: 1245 Total (i.e. Null);  1212 Residual
## Null Deviance:       1702
## Residual Deviance: 456.4     AIC: 524.4

glmModel1 <- glm(bal2$Default~Total.assets+Profit.after.tax+Cash.profit+
                   PBT.as...of.total.income+Cash.profit.as...of.total.income+
                   PAT.as...of.net.worth+Total.capital+Borrowings+Current.liabilities...
provisions+
                   Cumulative.retained.profits+Contingent.liabilities...Net.worth....
                 +Net.working.capital+Debt.to.equity.ratio..times.+ Cash.to.current.liab
ilities..times.+
                   PE.on.BSE+Borrowings.per.total.asset+capital.Employed.per.total.incom
e
                 +Change.in.stock.per.total.Income+Cumulative.retained.profits.per.sale+
```

```
                  Current.assets.per.total.asset+Current.liabilities...provisions.per.c
urrent.asset+
                  Net.fixed.assets.per.total.asset+Net.working.capital.per.total.asset+
                  Shareholder.fund.per.total.asset+Total.income.per.Sale+
                  Total.liabilities.per.shareholder.fund
                ,data = bal2,
                family = "binomial")

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

summary(glmModel1)

##
## Call:
## glm(formula = bal2$Default ~ Total.assets + Profit.after.tax +
##     Cash.profit + PBT.as...of.total.income + Cash.profit.as...of.total.income +
##     PAT.as...of.net.worth + Total.capital + Borrowings + Current.liabilities...provis
ions +
##     Cumulative.retained.profits + Contingent.liabilities...Net.worth.... +
##     Net.working.capital + Debt.to.equity.ratio..times. + Cash.to.current.liabilities.
.times. +
##     PE.on.BSE + Borrowings.per.total.asset + capital.Employed.per.total.income +
##     Change.in.stock.per.total.Income + Cumulative.retained.profits.per.sale +
##     Current.assets.per.total.asset + Current.liabilities...provisions.per.current.ass
et +
##     Net.fixed.assets.per.total.asset + Net.working.capital.per.total.asset +
##     Shareholder.fund.per.total.asset + Total.income.per.Sale +
##     Total.liabilities.per.shareholder.fund, family = "binomial",
##     data = bal2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.1837  -0.2901   0.0000   0.2180   3.2004
##
## Coefficients:
##                                                  Estimate Std. Error
## (Intercept)                                    -1.305e+01  2.853e+00
## Total.assets                                    2.223e-04  1.527e-04
## Profit.after.tax                                8.148e-03  3.283e-03
## Cash.profit                                    -8.724e-03  2.911e-03
## PBT.as...of.total.income                       -7.911e-03  2.728e-02
## Cash.profit.as...of.total.income               -2.153e-02  3.134e-02
## PAT.as...of.net.worth                          -5.639e-02  8.938e-03
## Total.capital                                  -3.516e-03  1.470e-03
## Borrowings                                     -3.370e-04  4.246e-04
## Current.liabilities...provisions                4.542e-04  4.524e-04
## Cumulative.retained.profits                    -1.139e-02  2.485e-03
## Contingent.liabilities...Net.worth....          3.795e-03  2.217e-03
## Net.working.capital                             2.043e-03  1.069e-03
## Debt.to.equity.ratio..times.                    2.480e-01  1.041e-01
## Cash.to.current.liabilities..times.             4.107e-01  4.770e-01
## PE.on.BSE                                      -1.171e-02  3.700e-03
## Borrowings.per.total.asset                      4.170e+00  9.679e-01
## capital.Employed.per.total.income              -3.939e-01  1.281e-01
## Change.in.stock.per.total.Income                2.159e+00  2.129e+00
```

```
## Cumulative.retained.profits.per.sale                -1.114e+00  3.686e-01
## Current.assets.per.total.asset                       4.482e+00  1.040e+00
## Current.liabilities...provisions.per.current.asset   1.739e+00  4.373e-01
## Net.fixed.assets.per.total.asset                     2.978e+00  7.660e-01
## Net.working.capital.per.total.asset                 -2.343e+00  1.166e+00
## Shareholder.fund.per.total.asset                     4.407e-01  1.220e+00
## Total.income.per.Sale                                6.214e+00  2.352e+00
## Total.liabilities.per.shareholder.fund               6.508e-02  8.715e-02
##                                                     z value Pr(>|z|)
## (Intercept)                                          -4.574 4.78e-06 ***
## Total.assets                                          1.456 0.145437
## Profit.after.tax                                      2.482 0.013063 *
## Cash.profit                                          -2.997 0.002730 **
## PBT.as...of.total.income                             -0.290 0.771814
## Cash.profit.as...of.total.income                     -0.687 0.492030
## PAT.as...of.net.worth                                -6.309 2.82e-10 ***
## Total.capital                                        -2.392 0.016735 *
## Borrowings                                           -0.794 0.427311
## Current.liabilities...provisions                      1.004 0.315361
## Cumulative.retained.profits                          -4.585 4.55e-06 ***
## Contingent.liabilities...Net.worth....               1.712 0.086913 .
## Net.working.capital                                  1.911 0.055972 .
## Debt.to.equity.ratio..times.                         2.381 0.017252 *
## Cash.to.current.liabilities..times.                  0.861 0.389195
## PE.on.BSE                                           -3.165 0.001549 **
## Borrowings.per.total.asset                           4.309 1.64e-05 ***
## capital.Employed.per.total.income                   -3.075 0.002104 **
## Change.in.stock.per.total.Income                     1.014 0.310443
## Cumulative.retained.profits.per.sale                -3.022 0.002512 **
## Current.assets.per.total.asset                       4.308 1.64e-05 ***
## Current.liabilities...provisions.per.current.asset   3.977 6.98e-05 ***
## Net.fixed.assets.per.total.asset                     3.888 0.000101 ***
## Net.working.capital.per.total.asset                 -2.008 0.044593 *
## Shareholder.fund.per.total.asset                     0.361 0.717905
## Total.income.per.Sale                                2.642 0.008248 **
## Total.liabilities.per.shareholder.fund               0.747 0.455238
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1701.8  on 1245  degrees of freedom
## Residual deviance:  554.0  on 1219  degrees of freedom
## AIC: 608
##
## Number of Fisher Scoring iterations: 9

lr.Test <- lmtest::lrtest(glmModel1)
lr.Test

## Likelihood ratio test
##
## Model 1: bal2$Default ~ Total.assets + Profit.after.tax + Cash.profit +
##     PBT.as...of.total.income + Cash.profit.as...of.total.income +
##     PAT.as...of.net.worth + Total.capital + Borrowings + Current.liabilities...provis
```

```
ions +
##     Cumulative.retained.profits + Contingent.liabilities...Net.worth.... +
##     Net.working.capital + Debt.to.equity.ratio..times. + Cash.to.current.liabilities.
.times. +
##     PE.on.BSE + Borrowings.per.total.asset + capital.Employed.per.total.income +
##     Change.in.stock.per.total.Income + Cumulative.retained.profits.per.sale +
##     Current.assets.per.total.asset + Current.liabilities...provisions.per.current.ass
et +
##     Net.fixed.assets.per.total.asset + Net.working.capital.per.total.asset +
##     Shareholder.fund.per.total.asset + Total.income.per.Sale +
##     Total.liabilities.per.shareholder.fund
## Model 2: bal2$Default ~ 1
##   #Df LogLik  Df  Chisq Pr(>Chisq)
## 1  27 -277.0
## 2   1 -850.9 -26 1147.8  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
rSq<- pscl::pR2(glmModel1)
rSq
```

```
##          llh      llhNull           G2     McFadden         r2ML
## -276.9993210 -850.9034985 1147.8083549    0.6744645    0.6019567
##         r2CU
##    0.8081825
```

```
# Checking the Variable Inflation Factor

vif(glmModel1)
```

```
##                                      Total.assets
##                                          5.955498
##                                   Profit.after.tax
##                                          4.520845
##                                       Cash.profit
##                                          5.055048
##                           PBT.as...of.total.income
##                                          5.611406
##                  Cash.profit.as...of.total.income
##                                          4.182282
##                              PAT.as...of.net.worth
##                                          2.066293
##                                    Total.capital
##                                          4.564983
##                                      Borrowings
##                                          5.471812
##                  Current.liabilities...provisions
##                                          3.123628
##                       Cumulative.retained.profits
##                                          2.035013
##             Contingent.liabilities...Net.worth....
##                                          1.236799
##                                Net.working.capital
##                                          1.691265
##                       Debt.to.equity.ratio..times.
```

```
##                                           4.015623
##                Cash.to.current.liabilities..times.
##                                           1.639853
##                                          PE.on.BSE
##                                           1.129655
##                        Borrowings.per.total.asset
##                                           2.926413
##                  capital.Employed.per.total.income
##                                           2.441577
##                   Change.in.stock.per.total.Income
##                                           1.138035
##               Cumulative.retained.profits.per.sale
##                                           1.886317
##                     Current.assets.per.total.asset
##                                           4.389126
## Current.liabilities...provisions.per.current.asset
##                                           2.952782
##                    Net.fixed.assets.per.total.asset
##                                           2.129210
##                Net.working.capital.per.total.asset
##                                           3.407543
##                   Shareholder.fund.per.total.asset
##                                           4.531783
##                               Total.income.per.Sale
##                                           1.441215
##             Total.liabilities.per.shareholder.fund
##                                           5.349458
```

```
# Predicting probability on Original Dataset

pred <- predict(glmModel1,data = bal2)
pred1 <- ifelse(pred<0.5,0,1)
pred1 <- as.factor(pred1)
actuals <- as.factor(bal2$Default)

# Creating Confusion Matrix on Original dataset

caret::confusionMatrix(pred1,actuals,positive='1')
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 679  84
##          1  33 450
##
##                Accuracy : 0.9061
##                  95% CI : (0.8885, 0.9217)
##     No Information Rate : 0.5714
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.806
##
##  Mcnemar's Test P-Value : 3.791e-06
##
```

```
##             Sensitivity : 0.8427
##             Specificity : 0.9537
##          Pos Pred Value : 0.9317
##          Neg Pred Value : 0.8899
##              Prevalence : 0.4286
##          Detection Rate : 0.3612
##    Detection Prevalence : 0.3876
##       Balanced Accuracy : 0.8982
##
##        'Positive' Class : 1
##

lr.Test <- lmtest::lrtest(glmModel)
lr.Test

## Likelihood ratio test
##
## Model 1: bal2$Default ~ Total.assets + Net.worth + Total.income + Change.in.stock +
##     Total.expenses + Profit.after.tax + PBDITA + PBT + Cash.profit +
##     PBDITA.as...of.total.income + PBT.as...of.total.income +
##     PAT.as...of.total.income + Cash.profit.as...of.total.income +
##     PAT.as...of.net.worth + Sales + Total.capital + Reserves.and.funds +
##     Borrowings + Current.liabilities...provisions + Shareholders.funds +
##     Cumulative.retained.profits + Capital.employed + TOL.TNW +
##     Total.term.liabilities...tangible.net.worth + Contingent.liabilities...Net.worth.
... +
##     Net.fixed.assets + Current.assets + Net.working.capital +
##     Quick.ratio..times. + Current.ratio..times. + Debt.to.equity.ratio..times. +
##     Cash.to.current.liabilities..times. + Cash.to.average.cost.of.sales.per.day +
##     Equity.face.value + EPS + Adjusted.EPS + Total.liabilities +
##     PE.on.BSE + Borrowings.per.total.Capital + Borrowings.per.total.asset +
##     capital.Employed.per.total.income + cash.profit.per.total.asset +
##     Change.in.stock.per.total.Income + Cumulative.retained.profits.per.sale +
##     Current.assets.per.total.asset + Current.liabilities...provisions.per.total.asset
 +
##     Current.liabilities...provisions.per.current.asset + Net.fixed.assets.per.total.a
sset +
##     Net.working.capital.per.total.asset + Net.working.capital.per.total.capital +
##     PAT.per.Sales + PAT.per.total.asset + PBDITA.per.Sales +
##     PBT.per.Sales + PBT.per.total.asset + PBT.per.Total.Capital +
##     Sales.per.total.asset + Net.working.capital.per.sales + Shareholder.fund.per.tota
l.asset +
##     Shareholder.fund.per.total.capital + Total.asset.per.Current.Liability +
##     Total.Income.per.Shareholder.fund + Total.income.per.total.asset +
##     Total.income.per.Sale + Total.income.per.total.expense +
##     Total.liabilities.per.shareholder.fund
## Model 2: bal2$Default ~ 1
##   #Df  LogLik  Df Chisq Pr(>Chisq)
## 1  66 -220.89
## 2   1 -850.90 -65  1260  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

rSq<- pscl::pR2(glmModel)
rSq
```

```
##          llh      llhNull           G2     McFadden         r2ML
## -220.8894360 -850.9034985 1260.0281249    0.7404060    0.6362391
##         r2CU
##    0.8542098

#########Importing the Validation dataset###############
unseen <- read_xlsx('validation_data.xlsx')

# Perfroming Exploratory data Analysis

summary(unseen)

##       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(unseen)
```

```
## 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 1034 ...
##  $ Change in stock                              : num  14.2 10.4 -28.1 -0.6 28.9 -0.5 N
A -7.7 27.2 -0.2 ...
##  $ Total expenses                               : num  2099 810 578 3613 1042 ...
##  $ 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 29.7 ...
##  $ 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 1027 ...
##  $ 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 0.0
6 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 0.0
5 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 N
A ...
##  $ 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 16
6 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 0.0
5 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 11.15 ...
##  $ 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" "18.6499999
99999999" "6.6" "NA" ...
##  $ Raw material turnover                        : chr  "17.579999999999998" "2.67" "3.7
7" "NA" ...
##  $ Shares outstanding                           : chr  "4800000" "11400000" "471285" "5
050000" ...
##  $ Equity face value                            : chr  "10" "10" "100" "10" ...
##  $ EPS                                          : num  18.6 1.65 -90.39 -7.09 5.9 ...
##  $ Adjusted EPS                                 : num  18.6 1.65 -90.39 -7.09 5.9 ...
##  $ Total liabilities                            : num  971 675 532 858 823 ...
##  $ PE on BSE                                    : chr  "NA" "NA" "-15.5" "-0.16" ...
```

```
colnames(unseen)<- make.names(colnames(unseen))
unseen1 <- unseen
str(unseen)
```

```
## 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 1034 ...
##  $ Change.in.stock                              : num  14.2 10.4 -28.1 -0.6 28.9 -0.5 N
A -7.7 27.2 -0.2 ...
##  $ Total.expenses                               : num  2099 810 578 3613 1042 ...
##  $ 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 29.7 ...
##  $ 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 1027 ...
##  $ 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 0.0
6 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 0.0
5 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 N
A ...
##  $ 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 16
6 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 0.0
5 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 11.15 ...
##  $ 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" "18.6499999
99999999" "6.6" "NA" ...
##  $ Raw.material.turnover                      : chr  "17.579999999999998" "2.67" "3.7
7" "NA" ...
##  $ Shares.outstanding                         : chr  "4800000" "11400000" "471285" "5
050000" ...
##  $ Equity.face.value                          : chr  "10" "10" "100" "10" ...
##  $ EPS                                        : num  18.6 1.65 -90.39 -7.09 5.9 ...
##  $ Adjusted.EPS                               : num  18.6 1.65 -90.39 -7.09 5.9 ...
##  $ Total.liabilities                          : num  971 675 532 858 823 ...
##  $ PE.on.BSE                                  : chr  "NA" "NA" "-15.5" "-0.16" ...
```

```r
# Convering the datatype of variables

unseen1$Creditors.turnover <- as.numeric(unseen1$Creditors.turnover)

## Warning: NAs introduced by coercion

unseen1$Debtors.turnover <- as.numeric(unseen1$Debtors.turnover)

## Warning: NAs introduced by coercion

unseen1$Finished.goods.turnover <- as.numeric(unseen1$Finished.goods.turnover)

## Warning: NAs introduced by coercion
```

```r
unseen1$WIP.turnover <- as.numeric(unseen1$WIP.turnover)
```

```
## Warning: NAs introduced by coercion
```

```r
unseen1$Raw.material.turnover <- as.numeric(unseen1$Raw.material.turnover)
```

```
## Warning: NAs introduced by coercion
```

```r
unseen1$Shares.outstanding <- as.numeric(unseen1$Shares.outstanding)
```

```
## Warning: NAs introduced by coercion
```

```r
unseen1$Equity.face.value <- as.numeric(unseen1$Equity.face.value)
```

```
## Warning: NAs introduced by coercion
```

```r
unseen1$PE.on.BSE <- as.numeric(unseen1$PE.on.BSE)
```

```
## Warning: NAs introduced by coercion
```

```r
names(unseen1)
```

```
##  [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"
```

```
# Dropping variables that are dropped in Original to bring it to the same level

data_unseen <- unseen1[,-c(1,18,19,22,25,32,34,42:47)]

# Imputing the missing values

unseen <- mice(data_unseen,m=5,meth = "pmm",maxit = 10,seed = 500)

unseen <- complete(unseen,1)
colSums(is.na(unseen))
```

```
##                                 Default...1
##                                           0
##                                Total.assets
##                                           0
##                                   Net.worth
##                                           0
##                                Total.income
##                                           0
##                              Change.in.stock
##                                           0
##                               Total.expenses
##                                           0
##                             Profit.after.tax
##                                           0
##                                      PBDITA
##                                           0
##                                         PBT
##                                           0
##                                Cash.profit
##                                           0
##                PBDITA.as...of.total.income
##                                           0
##                   PBT.as...of.total.income
##                                           0
##                   PAT.as...of.total.income
##                                          11
```

```
##            Cash.profit.as...of.total.income
##                                           0
##                       PAT.as...of.net.worth
##                                           0
##                                       Sales
##                                          46
##                                Total.capital
##                                           0
##                            Reserves.and.funds
##                                           0
##                                  Borrowings
##                                           0
##            Current.liabilities...provisions
##                                           0
##                            Shareholders.funds
##                                           0
##                 Cumulative.retained.profits
##                                           0
##                               Capital.employed
##                                           0
##                                     TOL.TNW
##                                           0
## Total.term.liabilities...tangible.net.worth
##                                           0
##      Contingent.liabilities...Net.worth....
##                                           0
##                               Net.fixed.assets
##                                           0
##                                 Current.assets
##                                           0
##                            Net.working.capital
##                                           0
##                            Quick.ratio..times.
##                                           0
##                         Current.ratio..times.
##                                           0
##                 Debt.to.equity.ratio..times.
##                                           0
##         Cash.to.current.liabilities..times.
##                                           0
##       Cash.to.average.cost.of.sales.per.day
##                                           0
##                               Equity.face.value
##                                           0
##                                          EPS
##                                           0
##                                Adjusted.EPS
##                                           0
##                               Total.liabilities
##                                           0
##                                      PE.on.BSE
##                                           0

unseen<-na.omit(unseen)
anyNA(unseen)
```

```
## [1] FALSE

# Creating new variables that are created in Original Dataset to create it of the same l
evel

unseen$Borrowings.per.total.Capital <- unseen$Borrowings/unseen$Total.capital
unseen$Borrowings.per.total.asset <- unseen$Borrowings/unseen$Total.assets
unseen$capital.Employed.per.total.income <- unseen$Capital.employed/unseen$Total.income
unseen$cash.profit.per.total.asset <- unseen$Cash.profit/unseen$Total.assets
unseen$Change.in.stock.per.total.Income <- unseen$Change.in.stock/unseen$Total.income
unseen$Cumulative.retained.profits.per.sale <- unseen$Cumulative.retained.profits/unseen
$Sales
unseen$Current.assets.per.total.asset <- unseen$Current.assets/unseen$Total.assets
unseen$Current.liabilities...provisions.per.total.asset <- unseen$Current.liabilities...
provisions/unseen$Total.assets
unseen$Current.liabilities...provisions.per.current.asset <- unseen$Current.liabilities.
..provisions/unseen$Current.assets
unseen$Net.fixed.assets.per.total.asset <- unseen$Net.fixed.assets/unseen$Total.assets
unseen$Net.working.capital.per.total.asset <- unseen$Net.working.capital/unseen$Total.as
sets
unseen$Net.working.capital.per.total.capital <- unseen$Net.working.capital/unseen$Total.
capital
unseen$PAT.per.Sales <- unseen$Profit.after.tax/unseen$Sales
unseen$PAT.per.total.asset <- unseen$Profit.after.tax/unseen$Total.assets
unseen$PAT.per.total.income <- unseen$Profit.after.tax/unseen$Total.income
unseen$PBDITA.per.Sales <- unseen$PBDITA/unseen$Sales
unseen$PBT.per.Sales <- unseen$PBT/unseen$Sales
unseen$PBT.per.total.asset <- unseen$PBT/unseen$Total.assets
unseen$PBT.per.Total.Capital <- unseen$PBT/unseen$Total.capital
unseen$Sales.per.total.asset <- unseen$Sales/unseen$Total.assets
unseen$Net.working.capital.per.sales <- unseen$Net.working.capital/unseen$Sales
unseen$Shareholder.fund.per.total.asset <- unseen$Shareholders.funds/unseen$Total.assets
unseen$Shareholder.fund.per.total.capital <- unseen$Shareholders.funds/unseen$Total.capi
tal
unseen$Total.asset.per.Current.Liability <- unseen$Total.assets/unseen$Current.liabiliti
es...provisions
unseen$Total.Income.per.Shareholder.fund <- unseen$Total.income/unseen$Shareholders.fund
s
unseen$Total.income.per.total.asset <- unseen$Total.income/unseen$Total.assets
unseen$Total.income.per.Sale <- unseen$Total.income/unseen$Sales
unseen$Total.income.per.total.expense <- unseen$Total.income/unseen$Total.expenses
unseen$Total.liabilities.per.shareholder.fund <- unseen$Total.liabilities/unseen$Shareho
lders.funds

# Chnaging the name of the Default Variable

colnames(unseen)[colnames(unseen) == "Default...1"] <- "Default"

# Predicting the probability on Validation dataset

pred_unseen <- predict(glmModel1,newdata = unseen)
unseen$probs <- pred_unseen
summary(unseen)
```

```
##     Default        Total.assets        Net.worth
##  Min.   :0.0000   Min.   :     0.7   Min.   :     0.2
##  1st Qu.:0.0000   1st Qu.:   118.1   1st Qu.:    43.4
##  Median :0.0000   Median :   422.3   Median :   131.0
##  Mean   :0.0583   Mean   :  4494.4   Mean   :  1736.3
##  3rd Qu.:0.0000   3rd Qu.:  1399.2   3rd Qu.:   506.5
##  Max.   :1.0000   Max.   :354727.3   Max.   :171840.0
##   Total.income       Change.in.stock  Total.expenses
##  Min.   :      0.1   Min.   :-488.1   Min.   :      0.2
##  1st Qu.:    125.7   1st Qu.:  -1.9   1st Qu.:    130.4
##  Median :    563.5   Median :   1.8   Median :    538.8
##  Mean   :   5305.8   Mean   :  54.3   Mean   :   4961.3
##  3rd Qu.:   1760.0   3rd Qu.:  19.6   3rd Qu.:   1734.2
##  Max.   :1028087.4   Max.   :7540.0   Max.   :1014813.1
##  Profit.after.tax      PBDITA              PBT           Cash.profit
##  Min.   : -998.0   Min.   :  -393.9   Min.   : -993.9   Min.   : -894.6
##  1st Qu.:    1.0   1st Qu.:     8.0   1st Qu.:    1.4   1st Qu.:    4.0
##  Median :   11.9   Median :    47.6   Median :   16.1   Median :   25.3
##  Mean   :  395.4   Mean   :   768.9   Mean   :  559.2   Mean   :  504.9
##  3rd Qu.:   71.8   3rd Qu.:   198.3   3rd Qu.:  103.4   3rd Qu.:  125.6
##  Max.   :62022.9   Max.   :110557.1   Max.   :94565.2   Max.   :71581.6
##  PBDITA.as...of.total.income PBT.as...of.total.income
##  Min.   :-6400.00            Min.   :-9700.00
##  1st Qu.:    5.32            1st Qu.:    0.94
##  Median :   10.21            Median :    3.59
##  Mean   :    1.33            Mean   :  -14.16
##  3rd Qu.:   16.92            3rd Qu.:    9.89
##  Max.   :  100.00            Max.   :   97.74
##  PAT.as...of.total.income Cash.profit.as...of.total.income
##  Min.   :-9700.00         Min.   :-6400.00
##  1st Qu.:    0.63         1st Qu.:    2.39
##  Median :    2.69         Median :    6.22
##  Mean   :  -15.66         Mean   :   -4.43
##  3rd Qu.:    7.04         3rd Qu.:   11.11
##  Max.   :   97.74         Max.   :  100.00
##  PAT.as...of.net.worth     Sales          Total.capital
##  Min.   :-169.23       Min.   :     0.1   Min.   :    0.4
##  1st Qu.:   0.95       1st Qu.:   120.8   1st Qu.:   17.1
##  Median :  10.22       Median :   552.5   Median :   49.5
##  Mean   :  11.07       Mean   :  5117.5   Mean   :  276.2
##  3rd Qu.:  21.31       3rd Qu.:  1721.3   3rd Qu.:  129.6
##  Max.   : 441.67       Max.   :976884.0   Max.   :41304.0
##  Reserves.and.funds   Borrowings       Current.liabilities...provisions
##  Min.   : -1125.0   Min.   :     0.2   Min.   :     0.1
##  1st Qu.:    11.1   1st Qu.:    21.4   1st Qu.:    19.9
##  Median :    64.4   Median :    98.7   Median :    84.7
##  Mean   :  1512.1   Mean   :  1394.7   Mean   :  1109.9
##  3rd Qu.:   352.3   3rd Qu.:   365.1   3rd Qu.:   319.2
##  Max.   :133684.2   Max.   :105175.3   Max.   :112712.7
##  Shareholders.funds Cumulative.retained.profits Capital.employed
##  Min.   :     0.2   Min.   : -2582.4           Min.   :     0.2
##  1st Qu.:    45.2   1st Qu.:     3.8           1st Qu.:    83.3
##  Median :   138.1   Median :    48.3           Median :   284.1
##  Mean   :  1753.3   Mean   :  1237.2           Mean   :  3145.9
```

```
##  3rd Qu.:   523.9   3rd Qu.:   277.6            3rd Qu.:   987.3
##  Max.   :171840.0   Max.   :128183.1            Max.   :235389.5
##     TOL.TNW         Total.term.liabilities...tangible.net.worth
##  Min.   :-350.480   Min.   :-325.600
##  1st Qu.:   0.680   1st Qu.:   0.070
##  Median :   1.430   Median :   0.360
##  Mean   :   3.409   Mean   :   1.336
##  3rd Qu.:   2.710   3rd Qu.:   0.990
##  Max.   : 411.270   Max.   : 292.020
##  Contingent.liabilities...Net.worth.... Net.fixed.assets
##  Min.   :   0.00                        Min.   :     0.1
##  1st Qu.:   0.00                        1st Qu.:    31.1
##  Median :   7.27                        Median :   104.6
##  Mean   :  60.05                        Mean   :  1379.0
##  3rd Qu.:  31.84                        3rd Qu.:   465.9
##  Max.   :6295.24                        Max.   :115737.5
##  Current.assets     Net.working.capital Quick.ratio..times.
##  Min.   :     0.1   Min.   :-41908.3    Min.   :  0.000
##  1st Qu.:    48.4   1st Qu.:    -0.8    1st Qu.:  0.430
##  Median :   181.5   Median :    25.0    Median :  0.670
##  Mean   :  1715.2   Mean   :   300.7    Mean   :  1.961
##  3rd Qu.:   612.5   3rd Qu.:   105.1    3rd Qu.:  1.020
##  Max.   :196614.6   Max.   : 85782.8    Max.   :341.000
##  Current.ratio..times. Debt.to.equity.ratio..times.
##  Min.   :  0.000       Min.   :  0.000
##  1st Qu.:  0.950       1st Qu.:  0.250
##  Median :  1.240       Median :  0.810
##  Mean   :  2.877       Mean   :  2.807
##  3rd Qu.:  1.730       3rd Qu.:  1.680
##  Max.   :505.000       Max.   :341.180
##  Cash.to.current.liabilities..times. Cash.to.average.cost.of.sales.per.day
##  Min.   :  0.0000                    Min.   :    0.00
##  1st Qu.:  0.0300                    1st Qu.:    3.65
##  Median :  0.0800                    Median :    8.53
##  Mean   :  0.7272                    Mean   :   64.71
##  3rd Qu.:  0.1900                    3rd Qu.:   22.54
##  Max.   :165.0000                    Max.   :15999.17
##  Equity.face.value      EPS             Adjusted.EPS
##  Min.   :    1.0   Min.   :-72750.00   Min.   :-72750.00
##  1st Qu.:   10.0   1st Qu.:     0.00   1st Qu.:     0.00
##  Median :   10.0   Median :     2.56   Median :     1.95
##  Mean   :   44.8   Mean   :   -82.10   Mean   :   -84.11
##  3rd Qu.:   10.0   3rd Qu.:    12.90   3rd Qu.:     9.26
##  Max.   :10000.0   Max.   :  8784.00   Max.   :  8784.00
##  Total.liabilities    PE.on.BSE       Borrowings.per.total.Capital
##  Min.   :     0.7   Min.   :-263.92   Min.   :  0.0008
##  1st Qu.:   118.1   1st Qu.:   2.14   1st Qu.:  0.8015
##  Median :   422.3   Median :   9.03   Median :  2.4310
##  Mean   :  4494.4   Mean   :  21.88   Mean   :  7.9921
##  3rd Qu.:  1399.2   3rd Qu.:  14.07   3rd Qu.:  5.7691
##  Max.   :354727.3   Max.   :1478.42   Max.   :350.0000
##  Borrowings.per.total.asset capital.Employed.per.total.income
##  Min.   : 0.0002            Min.   :  0.0039
##  1st Qu.: 0.1452            1st Qu.:  0.3680
```

```
##  Median : 0.3003            Median :  0.5937
##  Mean   : 0.3972            Mean   :  2.6504
##  3rd Qu.: 0.4588            3rd Qu.:  1.0223
##  Max.   :37.7684            Max.   :409.0000
##  cash.profit.per.total.asset Change.in.stock.per.total.Income
##  Min.   :-0.67368            Min.   :-178.7500
##  1st Qu.: 0.03503            1st Qu.:  -0.0055
##  Median : 0.07406            Median :   0.0062
##  Mean   : 0.07529            Mean   :   1.9385
##  3rd Qu.: 0.11572            3rd Qu.:   0.0262
##  Max.   : 0.50503            Max.   : 530.6875
##  Cumulative.retained.profits.per.sale Current.assets.per.total.asset
##  Min.   :-760.0000                    Min.   :  0.0008
##  1st Qu.:   0.0238                    1st Qu.:  0.3073
##  Median :   0.1078                    Median :  0.4502
##  Mean   :  -2.1981                    Mean   :  0.9407
##  3rd Qu.:   0.2952                    3rd Qu.:  0.6013
##  Max.   :  33.7586                    Max.   :321.3793
##  Current.liabilities...provisions.per.total.asset
##  Min.   :  0.00022
##  1st Qu.:  0.12455
##  Median :  0.20939
##  Mean   :  0.41751
##  3rd Qu.:  0.34058
##  Max.   :112.72632
##  Current.liabilities...provisions.per.current.asset
##  Min.   :    0.000
##  1st Qu.:    0.304
##  Median :    0.526
##  Mean   :   17.308
##  3rd Qu.:    0.761
##  Max.   :10709.000
##  Net.fixed.assets.per.total.asset Net.working.capital.per.total.asset
##  Min.   : 0.00262                 Min.   :-0.90221
##  1st Qu.: 0.19173                 1st Qu.:-0.01067
##  Median : 0.31333                 Median : 0.09332
##  Mean   : 0.40834                 Mean   : 0.10383
##  3rd Qu.: 0.47115                 3rd Qu.: 0.20962
##  Max.   :23.75676                 Max.   : 0.92266
##  Net.working.capital.per.total.capital PAT.per.Sales
##  Min.   :-61.2188                      Min.   :-97.0000
##  1st Qu.: -0.0412                      1st Qu.:  0.0063
##  Median :  0.5935                      Median :  0.0271
##  Mean   :  2.8635                      Mean   :  0.6173
##  3rd Qu.:  2.0168                      3rd Qu.:  0.0722
##  Max.   :322.5208                      Max.   :542.3750
##  PAT.per.total.asset PAT.per.total.income PBDITA.per.Sales
##  Min.   :-1.217810   Min.   :-97.00000    Min.   :-64.0000
##  1st Qu.: 0.007492   1st Qu.:  0.00628    1st Qu.:  0.0533
##  Median : 0.035202   Median :  0.02690    Median :  0.1035
##  Mean   : 0.050946   Mean   : -0.15661    Mean   :  0.8341
##  3rd Qu.: 0.081773   3rd Qu.:  0.07042    3rd Qu.:  0.1761
##  Max.   : 4.000996   Max.   :  0.97736    Max.   :542.3750
##  PBT.per.Sales      PBT.per.total.asset PBT.per.Total.Capital
```

```
##  Min.   :-97.0000   Min.   :-1.21661    Min.   :-66.25000
##  1st Qu.:  0.0099   1st Qu.: 0.01303    1st Qu.:  0.04878
##  Median :  0.0365   Median : 0.04993    Median :  0.43846
##  Mean   :  0.6416   Mean   : 0.06934    Mean   :  2.08847
##  3rd Qu.:  0.1011   3rd Qu.: 0.10843    3rd Qu.:  1.81220
##  Max.   :542.3750   Max.   : 4.00100    Max.   :125.68000
##  Sales.per.total.asset Net.working.capital.per.sales
##  Min.   : 0.001133     Min.   :-223.00000
##  1st Qu.: 0.664970     1st Qu.:  -0.00823
##  Median : 1.140697     Median :   0.07013
##  Mean   : 1.335769     Mean   :  -0.39059
##  3rd Qu.: 1.709436     3rd Qu.:   0.17467
##  Max.   :11.570220     Max.   :  56.72414
##  Shareholder.fund.per.total.asset Shareholder.fund.per.total.capital
##  Min.   :0.004348                 Min.   :  0.010
##  1st Qu.:0.257235                 1st Qu.:  1.359
##  Median :0.378016                 Median :  3.048
##  Mean   :0.410813                 Mean   :  9.461
##  3rd Qu.:0.543684                 3rd Qu.:  8.053
##  Max.   :0.999778                 Max.   :287.120
##  Total.asset.per.Current.Liability Total.Income.per.Shareholder.fund
##  Min.   :   0.009                  Min.   :  0.00244
##  1st Qu.:   2.936                  1st Qu.:  1.72866
##  Median :   4.776                  Median :  3.14410
##  Mean   :  15.429                  Mean   :  5.92617
##  3rd Qu.:   8.029                  3rd Qu.:  5.71936
##  Max.   :4500.000                  Max.   :276.49231
##  Total.income.per.total.asset Total.income.per.Sale
##  Min.   : 0.002444            Min.   :  1.000
##  1st Qu.: 0.721828            1st Qu.:  1.002
##  Median : 1.163158            Median :  1.008
##  Mean   : 1.370888            Mean   :  1.912
##  3rd Qu.: 1.738560            3rd Qu.:  1.026
##  Max.   :11.844332            Max.   :554.938
##  Total.income.per.total.expense Total.liabilities.per.shareholder.fund
##  Min.   : 0.0102                Min.   :  1.000
##  1st Qu.: 0.9919                1st Qu.:  1.839
##  Median : 1.0238                Median :  2.645
##  Mean   : 1.1379                Mean   :  5.267
##  3rd Qu.: 1.0704                3rd Qu.:  3.888
##  Max.   :61.2345                Max.   :230.000
##      probs
##  Min.   :-1438.218
##  1st Qu.:   -7.799
##  Median :   -3.521
##  Mean   :   27.844
##  3rd Qu.:   -1.090
##  Max.   :19604.680

pred_unseen1 <- as.factor(ifelse(pred_unseen<0.6,0,1))
unseen$Predicted <- pred_unseen1
default <- as.factor(unseen$Defaul)

# Creating Confusion Matrix on Validation dataset
```

```r
caret::confusionMatrix(pred_unseen1,default,positive='1')
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 556   4
##          1  74  35
##
##                Accuracy : 0.8834
##                  95% CI : (0.8566, 0.9067)
##     No Information Rate : 0.9417
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.4235
##
##  Mcnemar's Test P-Value : 5.597e-15
##
##             Sensitivity : 0.89744
##             Specificity : 0.88254
##          Pos Pred Value : 0.32110
##          Neg Pred Value : 0.99286
##              Prevalence : 0.05830
##          Detection Rate : 0.05232
##    Detection Prevalence : 0.16293
##       Balanced Accuracy : 0.88999
##
##        'Positive' Class : 1
##
```

```r
# Deciling the data on probability

prob <- seq (0,1, length = 11)
prob
```

```
##  [1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
```

```r
qs_train<- quantile(unseen$probs,prob,na.rm = TRUE)
unseen$Decile <- cut(unseen$probs,unique(qs_train), include.lowest = TRUE, right = FALSE
)

unseen.decile <- unseen %>% mutate(quartile = ntile(-unseen$probs, 10))
view(unseen.decile)

table(unseen.decile$Default)
```

```
##
##   0   1
## 630  39
```

```r
defaulter<- data.table::data.table(unseen.decile)
defaulter
```

```
##      Default Total.assets Net.worth Total.income Change.in.stock
##   1:       0        970.6     275.8       2185.2            14.2
```

```
##   2:       0        675.0     211.6        819.2            10.4
##   3:       1        532.1     120.2        563.5           -28.1
##   4:       0        857.5     201.4       3576.5            -0.6
##   5:       0        823.3     349.2       1034.3            28.9
##  ---
## 665:       0        201.4      83.2        310.1            19.4
## 666:       0        463.1     211.1        429.2            -0.2
## 667:       0       4012.8    1640.6       4599.0            15.1
## 668:       0        602.2     301.9       2731.5            21.1
## 669:       0       6382.2    2609.4       3786.7           111.4
##      Total.expenses Profit.after.tax PBDITA    PBT Cash.profit
##   1:         2099.2            100.2  285.6  152.1       182.3
##   2:          809.9             19.7  116.0   33.7        50.5
##   3:          577.8            -42.4  -31.0  -56.0       -35.3
##   4:         3613.4            -37.5   68.2   25.7        37.3
##   5:         1041.8             21.4   90.1   29.7        62.7
##  ---
## 665:          315.0             14.5   30.3   14.9        19.9
## 666:          388.2             40.8   94.8   64.2        68.5
## 667:         4329.0            285.1  645.9  394.4       507.2
## 668:          343.2           2409.4 2470.6 2409.4        -0.7
## 669:         3795.9            102.2  518.9  109.1       340.5
##      PBDITA.as...of.total.income PBT.as...of.total.income
##   1:                       13.07                     6.96
##   2:                       14.16                     4.11
##   3:                       -5.50                    -9.94
##   4:                        1.91                     0.72
##   5:                        8.71                     2.87
##  ---
## 665:                        9.77                     4.80
## 666:                       22.09                    14.96
## 667:                       14.04                     8.58
## 668:                       90.45                    88.21
## 669:                       13.70                     2.88
##      PAT.as...of.total.income Cash.profit.as...of.total.income
##   1:                     4.59                            8.34
##   2:                     2.40                            6.16
##   3:                    -7.52                           -6.26
##   4:                    -1.05                            1.04
##   5:                     2.07                            6.06
##  ---
## 665:                     4.68                            6.42
## 666:                     9.51                           15.96
## 667:                     6.20                           11.03
## 668:                    88.21                           -0.03
## 669:                     2.70                            8.99
##      PAT.as...of.net.worth  Sales Total.capital Reserves.and.funds
##   1:                 42.11 2171.1          48.0              413.1
##   2:                 10.66  817.0         114.0               97.6
##   3:                -31.20  552.2          47.1              227.4
##   4:                  0.00 3573.3          50.5              150.9
##   5:                  6.31 1026.7          33.0              316.2
##  ---
## 665:                 20.51  309.6          85.9              -16.1
```

```
## 666:                 21.70  427.2          14.1              197.0
## 667:                 21.88 4586.6          46.6             1610.8
## 668:                  0.00  292.3         331.6              -29.7
## 669:                  3.99 3706.7         280.0             2329.4
##      Borrowings Current.liabilities...provisions Shareholders.funds
##   1:      177.3                              328.5               275.8
##   2:      339.8                              100.5               211.6
##   3:       17.5                              240.1               120.2
##   4:      524.2                               75.2               201.4
##   5:      162.3                              299.6               349.2
##  ---
## 665:       56.5                               61.3                83.2
## 666:        6.6                              244.0               211.1
## 667:     1383.9                              662.1              1640.6
## 668:      176.1                              124.2               301.9
## 669:     1466.0                             1723.3              2609.4
##      Cumulative.retained.profits Capital.employed TOL.TNW
##   1:                       227.8            453.1    1.80
##   2:                        97.6            551.4    2.01
##   3:                        69.9            137.7    1.73
##   4:                       150.9            725.6    2.94
##   5:                       316.2            511.5    1.02
##  ---
## 665:                       -16.5            139.7    1.41
## 666:                       197.0            217.7    1.02
## 667:                      1183.1           3024.5    1.61
## 668:                      -120.4            478.0    0.99
## 669:                      2286.4           4075.4    1.08
##      Total.term.liabilities...tangible.net.worth
##   1:                                        0.27
##   2:                                        0.72
##   3:                                        0.09
##   4:                                        0.81
##   5:                                        0.10
##  ---
## 665:                                        0.25
## 666:                                        0.01
## 667:                                        0.71
## 668:                                        0.31
## 669:                                        0.31
##      Contingent.liabilities...Net.worth.... Net.fixed.assets
##   1:                                 112.94            332.3
##   2:                                   5.77            199.1
##   3:                                 102.83            270.0
##   4:                                   0.65            262.8
##   5:                                  28.78            190.6
##  ---
## 665:                                  21.39             87.6
## 666:                                   0.00            158.4
## 667:                                  13.46           1691.1
## 668:                                   0.00            261.4
## 669:                                 133.18           3115.0
##      Current.assets Net.working.capital Quick.ratio..times.
##   1:          559.7               134.2                0.92
```

```
##   2:          407.3               123.6                0.48
##   3:          147.8               -97.1                0.32
##   4:          535.6                99.6                0.51
##   5:          471.8                75.3                0.58
##  ---
## 665:           98.7                 2.0                0.49
## 666:          228.2               -19.9                0.84
## 667:         1615.7               460.3                0.59
## 668:          290.1                87.6                0.67
## 669:         2091.7              -131.9                0.52
##      Current.ratio..times. Debt.to.equity.ratio..times.
##   1:                  1.31                        0.64
##   2:                  1.39                        1.61
##   3:                  0.60                        0.15
##   4:                  1.23                        2.60
##   5:                  1.19                        0.46
##  ---
## 665:                  1.02                        0.68
## 666:                  0.92                        0.03
## 667:                  1.40                        0.84
## 668:                  1.41                        0.58
## 669:                  0.94                        0.56
##      Cash.to.current.liabilities..times.
##   1:                                0.09
##   2:                                0.03
##   3:                                0.04
##   4:                                0.08
##   5:                                0.08
##  ---
## 665:                                0.02
## 666:                                0.14
## 667:                                0.06
## 668:                                0.35
## 669:                                0.06
##      Cash.to.average.cost.of.sales.per.day Equity.face.value    EPS
##   1:                                  7.56                10  18.60
##   2:                                  3.88                10   1.65
##   3:                                  4.63               100 -90.39
##   4:                                  3.71                10  -7.09
##   5:                                 11.15                10   5.90
##  ---
## 665:                                  2.30               100  -0.01
## 666:                                 34.93                10  62.77
## 667:                                  5.97                10  61.76
## 668:                                 90.93                10  -0.86
## 669:                                 14.22                 5   1.71
##      Adjusted.EPS Total.liabilities PE.on.BSE Borrowings.per.total.Capital
##   1:        18.60             970.6      2.64                    3.6937500
##   2:         1.65             675.0     -4.17                    2.9807018
##   3:       -90.39             532.1    -15.50                    0.3715499
##   4:        -7.09             857.5     -0.16                   10.3801980
##   5:         5.90             823.3     -7.10                    4.9181818
##  ---
## 665:        -0.01             201.4     12.26                    0.6577416
```

```
## 666:        62.77             463.1     -2.70                    0.4680851
## 667:        61.76            4012.8     37.46                   29.6974249
## 668:        -0.09             602.2    -38.59                    0.5310615
## 669:         0.34            6382.2     37.46                    5.2357143
##      Borrowings.per.total.asset capital.Employed.per.total.income
##   1:                 0.18267051                          0.2073494
##   2:                 0.50340741                          0.6730957
##   3:                 0.03288855                          0.2443656
##   4:                 0.61131195                          0.2028799
##   5:                 0.19713349                          0.4945374
##  ---
## 665:                 0.28053625                          0.4504998
## 666:                 0.01425178                          0.5072227
## 667:                 0.34487141                          0.6576430
## 668:                 0.29242776                          0.1749954
## 669:                 0.22970136                          1.0762405
##      cash.profit.per.total.asset Change.in.stock.per.total.Income
##   1:                 0.187821966                     0.0064982610
##   2:                 0.074814815                     0.0126953125
##   3:                -0.066340913                    -0.0498669033
##   4:                 0.043498542                    -0.0001677618
##   5:                 0.076156929                     0.0279416030
##  ---
## 665:                 0.098808342                     0.0625604644
## 666:                 0.147916217                    -0.0004659832
## 667:                 0.126395534                     0.0032833225
## 668:                -0.001162405                     0.0077246934
## 669:                 0.053351509                     0.0294187551
##      Cumulative.retained.profits.per.sale Current.assets.per.total.asset
##   1:                           0.10492377                      0.5766536
##   2:                           0.11946144                      0.6034074
##   3:                           0.12658457                      0.2777673
##   4:                           0.04222987                      0.6246064
##   5:                           0.30797701                      0.5730596
##  ---
## 665:                          -0.05329457                      0.4900695
## 666:                           0.46114232                      0.4927661
## 667:                           0.25794706                      0.4026366
## 668:                          -0.41190558                      0.4817336
## 669:                           0.61682899                      0.3277397
##      Current.liabilities...provisions.per.total.asset
##   1:                                       0.33845044
##   2:                                       0.14888889
##   3:                                       0.45123097
##   4:                                       0.08769679
##   5:                                       0.36390137
##  ---
## 665:                                       0.30436941
## 666:                                       0.52688404
## 667:                                       0.16499701
## 668:                                       0.20624377
## 669:                                       0.27001661
##      Current.liabilities...provisions.per.current.asset
##   1:                                           0.5869216
```

```
##   2:                                          0.2467469
##   3:                                          1.6244926
##   4:                                          0.1404033
##   5:                                          0.6350148
##  ---
## 665:                                          0.6210740
## 666:                                          1.0692375
## 667:                                          0.4097914
## 668:                                          0.4281282
## 669:                                          0.8238753
##      Net.fixed.assets.per.total.asset Net.working.capital.per.total.asset
##   1:                        0.3423655                          0.138264991
##   2:                        0.2949630                          0.183111111
##   3:                        0.5074234                         -0.182484495
##   4:                        0.3064723                          0.116151603
##   5:                        0.2315073                          0.091461193
##  ---
## 665:                        0.4349553                          0.009930487
## 666:                        0.3420428                         -0.042971281
## 667:                        0.4214264                          0.114707935
## 668:                        0.4340751                          0.145466622
## 669:                        0.4880762                         -0.020666855
##      Net.working.capital.per.total.capital PAT.per.Sales
##   1:                            2.79583333    0.04615172
##   2:                            1.08421053    0.02411261
##   3:                           -2.06157113   -0.07678377
##   4:                            1.97227723   -0.01049450
##   5:                            2.28181818    0.02084348
##  ---
## 665:                            0.02328289    0.04683463
## 666:                           -1.41134752    0.09550562
## 667:                            9.87768240    0.06215933
## 668:                            0.26417370    8.24290113
## 669:                           -0.47107143    0.02757169
##      PAT.per.total.asset PAT.per.total.income PBDITA.per.Sales
##   1:          0.10323511           0.04585393       0.13154622
##   2:          0.02918519           0.02404785       0.14198286
##   3:         -0.07968427          -0.07524401      -0.05613908
##   4:         -0.04373178          -0.01048511       0.01908600
##   5:          0.02599296           0.02069032       0.08775689
##  ---
## 665:          0.07199603           0.04675911       0.09786822
## 666:          0.08810192           0.09506058       0.22191011
## 667:          0.07104765           0.06199174       0.14082327
## 668:          4.00099635           0.88207944       8.45227506
## 669:          0.01601329           0.02698920       0.13998975
##      PBT.per.Sales PBT.per.total.asset PBT.per.Total.Capital
##   1:   0.070056653          0.15670719             3.1687500
##   2:   0.041248470          0.04992593             0.2956140
##   3:  -0.101412532         -0.10524338            -1.1889597
##   4:   0.007192231          0.02997085             0.5089109
##   5:   0.028927632          0.03607433             0.9000000
##  ---
## 665:   0.048126615          0.07398213             0.1734575
```

```
## 666:   0.150280899          0.13863097             4.5531915
## 667:   0.085989622          0.09828549             8.4635193
## 668:   8.242901129          4.00099635             7.2659831
## 669:   0.029433189          0.01709442             0.3896429
##      Sales.per.total.asset Net.working.capital.per.sales
##   1:             2.2368638                   0.061811985
##   2:             1.2103704                   0.151285190
##   3:             1.0377749                  -0.175842086
##   4:             4.1671137                   0.027873394
##   5:             1.2470545                   0.073341775
##  ---
## 665:             1.5372393                   0.006459948
## 666:             0.9224789                  -0.046582397
## 667:             1.1429924                   0.100357563
## 668:             0.4853869                   0.299692097
## 669:             0.5807872                  -0.035584212
##      Shareholder.fund.per.total.asset Shareholder.fund.per.total.capital
##   1:                        0.2841541                           5.7458333
##   2:                        0.3134815                           1.8561404
##   3:                        0.2258974                           2.5520170
##   4:                        0.2348688                           3.9881188
##   5:                        0.4241467                          10.5818182
##  ---
## 665:                        0.4131082                           0.9685681
## 666:                        0.4558411                          14.9716312
## 667:                        0.4088417                          35.2060086
## 668:                        0.5013285                           0.9104343
## 669:                        0.4088559                           9.3192857
##      Total.asset.per.Current.Liability Total.Income.per.Shareholder.fund
##   1:                          2.954642                           7.923133
##   2:                          6.716418                           3.871456
##   3:                          2.216160                           4.688020
##   4:                         11.402926                          17.758193
##   5:                          2.747997                           2.961913
##  ---
## 665:                          3.285481                           3.727163
## 666:                          1.897951                           2.033160
## 667:                          6.060716                           2.803243
## 668:                          4.848631                           9.047698
## 669:                          3.703476                           1.451177
##      Total.income.per.total.asset Total.income.per.Sale
##   1:                    2.2513909             1.006494
##   2:                    1.2136296             1.002693
##   3:                    1.0590115             1.020464
##   4:                    4.1708455             1.000896
##   5:                    1.2562857             1.007402
##  ---
## 665:                    1.5397219             1.001615
## 666:                    0.9267977             1.004682
## 667:                    1.1460825             1.002704
## 668:                    4.5358685             9.344851
## 669:                    0.5933221             1.021583
##      Total.income.per.total.expense Total.liabilities.per.shareholder.fund
##   1:                      1.0409680                                3.519217
```

```
##   2:                      1.0114829                                3.189981
##   3:                      0.9752510                                4.426789
##   4:                      0.9897880                                4.257696
##   5:                      0.9928009                                2.357675
##  ---
## 665:                      0.9844444                                2.420673
## 666:                      1.1056157                                2.193747
## 667:                      1.0623701                                2.445934
## 668:                      7.9589161                                1.994700
## 669:                      0.9975763                                2.445850
##           probs Predicted            Decile quartile
##   1:  -6.535538         0     [-9.56,-6.12)        8
##   2:  -2.903199         0      [-3.52,-2.6)        5
##   3:   1.103417         1     [-0.445,5.65)        2
##   4:  -1.964010         0       [-2.6,-1.7)        4
##   5:  -5.797700         0     [-6.12,-4.71)        7
##  ---
## 665:  -2.079130         0       [-2.6,-1.7)        4
## 666:  -5.906247         0     [-6.12,-4.71)        7
## 667: -17.846209         0     [-20.3,-9.56)        9
## 668:  71.049737         1   [5.65,1.96e+04]        1
## 669: -29.989202         0 [-1.44e+03,-20.3)       10

unseen1_decile <- defaulter[,list(`# Defaulter` <- sum(Default==1),
                                  Total  <-length(Default)) , by= Decile][order(Decile)]
```