

# Mini Project – Financial and Risk Analysis

Project Report

Jyosmitha M

## Contents

1.	Project Objective.....	2
2.	Assumptions.....	2
3.	Environment Set up and Data Import.....	2
3.1	Install necessary Packages and Invoke Libraries.....	2
3.2	Functions used in R-code: .....	3
3.3	Set up working Directory .....	3
3.4	Import and Read the Dataset.....	3
4.	Meta Data: .....	3
5.	Exploratory Data Analysis .....	5
5.1	Column names and number of observations in Train dataset:.....	5
5.3	Structure of the dataset:.....	7
5.4	Five point Summary: .....	11
5.5	Converting NetWorthNextYearColumn into categorical in Train dataset: .....	17
6.	Outlier Treatment: .....	18
7.	Missing Values Imputation: .....	20
8.	Feature Engineering/Feature Selection:.....	23
9.	MultiCollinearity: .....	24
10.	Univariate Analysis:.....	29
11.	Bivariate Analysis:.....	35
12.	Logistic Regression:.....	38
13.	Model Performance Measures: .....	44
14.	Deciles:.....	46
15.	Observation and inferences:.....	48
16.	Model Performance: .....	48

## **1. Project Objective**

To create an India credit risk(default) model, using the data provided in the spreadsheet raw-data.xlsx, and validate it on validation\_data.xlsx. Below are the items that will be

- Importing the dataset in R
- Understanding the structure of dataset and modifying the data into format
- Graphical exploration
- Descriptive statistics
- Insights from the dataset
- Check if the assumptions are met
- Check for multicollinearity
- Creating model using logistic regression and analyze coefficients
- Make predictions and evaluate with Confusion Matrix and ROC Curve
- Divide into deciles and check the performance.

## **2. Assumptions**

Data is highly imbalanced and there maybe be possible outliers and missing values. Data might not be following a normal distribution.

### **Logistic Regression:**

- No outliers should be present
- No missing values should be present
- There should not be a multi-collinearity between independent variables. (only a little collinearity is allowed)
- There should be a linear relationship between the link function and independent variables in logit model.
- Dependent variables need not be normally distributed
- Dataset should be large.
- Errors need to be independent and not normally distributed.

## **3. Environment Set up and Data Import**

### **3.1 Install necessary Packages and Invoke Libraries**

- library(mice)
- library(readxl)
- library(DataExplorer)
- library(stats)
- library(caret)
- library(pROC)
- library(car)
- library(caTools)
- library(dplyr)
- library(ggplot2)

- library(reshape)
- library(data.table())
- library(scales)

### 3.2 Functions used in R-code:

- median
- replace
- lapply
- glm
- confusionMatrix
- melt
- roc
- predict.glm
- vector
- percent
- aggregate
- quantile

### 3.3 Set 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

```
```{r setup, include=FALSE}
setwd("C:/Users/ammu/Desktop/Great Lakes/9. FRA/Project")
getwd()
````
```

### 3.4 Import and Read the Dataset

Train and Validation Data are excel format is imported into R environment.

```
```{r cars}
Train=read_excel("raw-data.xlsx")
Validation=read_excel("validation_data.xlsx")
````
```

## 4. Meta Data:

Churn is the predictor/response categorical variable. ContractRenewal and DataPlan are categorical and rest are Continuous variables

| Variable Name      | Description  |
|--------------------|--|
| Networth Next Year | Net worth of the customer in next year   |
| Total assets       | Total assets of customer   |
| Net worth          | Net worth of the customer of present year  |
| Total income       | Total income of the customer   |
| Change in stock    | difference between value of current stock and the value of stock in last trading day |
| Total expenses     | Total expense done by customer   |

|   |  |
|---|--|
| Profit after tax                            | Profit after tax deduction   |
| PBDITA                                      | Profit before depreciation, income tax and amortization                |
| PBT   | Profit before tax deduction  |
| Cash profit                                 | Total Cash profit  |
| PBDITA as % of total income                 | PBDITA / Total income  |
| PBT as % of total income                    | PBT / Total income   |
| PAT as % of total income                    | PAT / Total income   |
| Cash profit as % of total income            | Cash Profit / Total income   |
| PAT as % of net worth                       | PAT / Net worth  |
| Sales                                       | Sales done by customer   |
| Income from financial services              | Income from financial services   |
| Other income                                | Income from other sources  |
| Total capital                               | Total capital of the customer  |
| Reserves and funds                          | Total reserves and funds of the customer                               |
| Deposits (accepted by commercial banks)     | All blank values   |
| Borrowings                                  | Total amount borrowed by customer                                      |
| Current liabilities & provisions            | current liabilities of the customer                                    |
| Deferred tax liability                      | Future income tax customer will pay because of the current transaction |
| Shareholders funds                          | Amount of equity in a company, which is belong to shareholder          |
| Cumulative retained profits                 | Total cumulative profit retained by customer                           |
| Capital employed                            | Current asset minus current liabilities                                |
| TOL/TNW                                     | Total liabilities of the customer divided by Total net worth           |
| Total term liabilities / tangible net worth | Short + long term liabilities divided by tangible net worth            |
| Contingent liabilities / Net worth (%)      | Contingent liabilities / Net worth                                     |
| Contingent liabilities                      | Liabilities because of uncertain events                                |
| Net fixed assets                            | purchase price of all fixed assets                                     |
| Investments                                 | Total invested amount  |
| Current assets                              | Assets that are expected to be converted to cash within a year         |
| Net working capital                         | Difference of current liabilities and current assets                   |
| Quick ratio (times)                         | Total cash divided by current liabilities                              |
| Current ratio (times)                       | Current assets divided by current liabilities                          |
| Debt to equity ratio (times)                | Total liabilities divided by its shareholder equity                    |
| Cash to current liabilities (times)         | Total liquid cash divided by current liabilities                       |
| Cash to average cost of sales per day       | Total cash divided by average cost of the sales                        |
| Creditors turnover                          | Net credit purchase divided to average trade creditors                 |
| Debtors turnover                            | Net credit sales divided by average accounts receivable                |
| Finished goods turnover                     | Annual sales divided by average inventory                              |

|                       |   |
|-----------------------|---|
| WIP turnover          | The cost of goods sold for a period divided by the average inventory for that period  |
| Raw material turnover | Cost of goods sold is divided by the average inventory for the same period  |
| Shares outstanding    | Number of issued shares minus the number of share held in the company   |
| Equity face value     | cost of the equity at the time of issuing   |
| EPS                   | Net income divided by total number of outstanding share   |
| Adjusted EPS          | Adjusted net earning divided by the weighted average number of common share outstanding on a diluted basis during the plan year |
| Total liabilities     | Sum of all type of liabilities  |
| PE on BSE             | Company current stock price divided by its earning per share  |

## 5. Exploratory Data Analysis

### 5.1 Column names and number of observations in Train dataset:

Colum names and dimensions of dataset: There are **3541 rows and 52 columns** in the dataset.

Networth Next Year is the target variable. We can rename it as 'default'

```
> dim(Train)
[1] 3541   52
> names(Train)
[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"

```

In Validation dataset, there are **715 rows and 52 columns**. The “Default -1” is the target variable. We can change the column name as ‘Default’

```

> dim(Validation)
[1] 715  52
> names(Validation)
[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"
```

Below 2 variables can be dropped:

- Deposits (accepted by commercial banks) is having no data. Hence this can be dropped
- ID column does not give any info. Hence it can be dropped.

```
```{r}
#dropping Num and Deposits (accepted by commercial banks) and num |columns
Train=Train[, -c(1,22)]
Validation=Validation[, -c(1,22)]
```

```

### 5.3 Structure of the dataset:

All the variables are of numeric datatype. Some are numeric but have character as datatype.

#### Train Dataset:

```
> str(Train)
tibble [3,541 x 52] (S3: tbl_df/tbl/data.frame)
$ Num : num [1:3541] 1 2 3 4 5 6 7 8
9 10 ...
$ Networth Next Year : num [1:3541] 0 0 0 0 0 0 0 0
0 1 ...
```

|  |                                 |
|--|---------------------------------|
| \$ Total assets                            | : num [1:3541] 17512.3 941 232. |
| 8 2.7 478.5 ...                            |                                 |
| \$ Net worth                               | : num [1:3541] 7093.2 351.5 100 |
| .6 2.7 107.6 ...                           |                                 |
| \$ Total income                            | : num [1:3541] 24965 1527 477 N |
| A 1580 ...                                 |                                 |
| \$ Change in stock                         | : num [1:3541] 235.8 42.7 -5.2  |
| NA -17 ...                                 |                                 |
| \$ Total expenses                          | : num [1:3541] 23658 1455 479 N |
| A 1558 ...                                 |                                 |
| \$ Profit after tax                        | : num [1:3541] 1543.2 115.2 -6. |
| 6 NA 5.5 ...                               |                                 |
| \$ PBDITA                                  | : num [1:3541] 2860.2 283 5.8 N |
| A 31 ...                                   |                                 |
| \$ PBT                                     | : num [1:3541] 2417.2 188.4 -6. |
| 6 NA 6.3 ...                               |                                 |
| \$ Cash profit                             | : num [1:3541] 1872.8 158.6 0.3 |
| NA 11.9 ...                                |                                 |
| \$ PBDITA as % of total income             | : num [1:3541] 11.46 18.53 1.22 |
| 0 1.96 ...                                 |                                 |
| \$ PBT as % of total income                | : num [1:3541] 9.68 12.33 -1.38 |
| 0 0.4 ...                                  |                                 |
| \$ PAT as % of total income                | : num [1:3541] 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 [1:3541] 7.5 10.38 0.06 0 |
| 0.75 ...                                   |                                 |
| \$ PAT as % of net worth                   | : num [1:3541] 23.78 38.08 -6.3 |
| 5 0 5.25 ...                               |                                 |
| \$ Sales                                   | : num [1:3541] 24458 1504 476 N |
| A 1575 ...                                 |                                 |
| \$ Income from financial services          | : num [1:3541] 158 4 1.5 NA 3.9 |
| 6.4 NA NA 7.3 NA ...                       |                                 |
| \$ Other income                            | : num [1:3541] 297.2 15.9 0.2 N |
| A 0.9 ...                                  |                                 |
| \$ Total capital                           | : num [1:3541] 423.8 115.5 81.4 |
| 0.5 6.2 ...                                |                                 |
| \$ Reserves and funds                      | : num [1:3541] 6822.8 257.8 19. |
| 2 2.2 161.8 ...                            |                                 |
| \$ Deposits (accepted by commercial banks) | : logi [1:3541] NA NA NA NA NA  |
| NA ...                                     |                                 |
| \$ Borrowings                              | : num [1:3541] 14.9 272.5 35.4  |
| NA 193.1 ...                               |                                 |
| \$ Current liabilities & provisions        | : num [1:3541] 9965.9 210 96.8  |
| NA 112.8 ...                               |                                 |
| \$ Deferred tax liability                  | : num [1:3541] 284.9 85.2 NA NA |
| 4.6 ...                                    |                                 |
| \$ Shareholders funds                      | : num [1:3541] 7093.2 351.5 100 |
| .6 2.7 107.6 ...                           |                                 |
| \$ Cumulative retained profits             | : num [1:3541] 6263.3 247.4 32. |
| 4 2.2 82.7 ...                             |                                 |
| \$ Capital employed                        | : num [1:3541] 7108.1 624 136 2 |
| .7 300.7 ...                               |                                 |
| \$ TOL/TNW                                 | : num [1:3541] 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 [1:3541] 0 0.34 0.29 0 1.
59 0.37 0.03 0.94 0.3 0.54 ...
$ Contingent liabilities / Net worth (%): num [1:3541] 14.8 19.2 45.8 0
34.9 ...
$ Contingent liabilities: num [1:3541] 1049.7 67.6 46.1
NA 37.6 ...
$ Net fixed assets: num [1:3541] 1900.2 286.4 38.
7 2.5 94.8 ...
$ Investments: num [1:3541] 1069.6 2.2 4.3 N
A 7.4 ...
$ Current assets: num [1:3541] 13277.5 563.9 16
7.5 0.2 349.7 ...
$ Net working capital: num [1:3541] 3588.5 203.5 59.
6 0.2 215.8 ...
$ Quick ratio (times): num [1:3541] 1.18 0.95 1.11 N
A 1.41 0.48 NA 0.54 0.59 0.39 ...
$ Current ratio (times): num [1:3541] 1.37 1.56 1.55 N
A 2.54 1.27 NA 1.15 1.58 0.5 ...
$ Debt to equity ratio (times): num [1:3541] 0 0.78 0.35 0 1.
79 1.09 0.32 2.31 0.94 3.13 ...
$ Cash to current liabilities (times): num [1:3541] 0.43 0.06 0.21 N
A 0 0.11 NA 0.04 0.19 0 ...
$ Cash to average cost of sales per day: num [1:3541] 68.21 5.96 17.07
NA 0 ...
$ Creditors turnover: chr [1:3541] "3.62" "9.800000
000000007" "5.28" "0" ...
$ Debtors turnover: chr [1:3541] "3.85" "5.7" "5.
07" "0" ...
$ Finished goods turnover: chr [1:3541] "200.55" "14.21"
"9.24" NA ...
$ WIP turnover: chr [1:3541] "21.78" "7.49" "
0.23" NA ...
$ Raw material turnover: chr [1:3541] "7.71" "11.46" N
A "0" ...
$ Shares outstanding: chr [1:3541] "42381675" "1155
0000" "8149090" "52404" ...
$ Equity face value: chr [1:3541] "10" "10" "10" "
10" ...
$ EPS: num [1:3541] 35.52 9.97 -0.5
0 7.91 ...
$ Adjusted EPS: num [1:3541] 7.1 9.97 -0.5 0
7.91 ...
$ Total liabilities: num [1:3541] 17512.3 941 232.
8 2.7 478.5 ...
$ PE on BSE: chr [1:3541] "27.31" "8.17" "
-5.76" "NA" ...

```

Test Dataset structure:

```

tibble [715 x 52] (S3: tbl_df/tbl/data.frame)
$ Num: num [1:715] 1 2 3 4 5 6 7 8 9
10 ...
$ Default - 1: num [1:715] 0 0 1 0 0 0 0 0 0
0 ...
$ Total assets: num [1:715] 971 675 532 858
823 ...

```

|   |                                 |
|---|---------------------------------|
| \$ Net worth                                    | : num [1:715] 276 212 120 201   |
| 349 ...   |                                 |
| \$ Total income                                 | : num [1:715] 2185 819 564 3576 |
| 1034 ...  |                                 |
| \$ Change in stock                              | : num [1:715] 14.2 10.4 -28.1 - |
| 0.6 28.9 -0.5 NA -7.7 27.2 -0.2 ...             |                                 |
| \$ Total expenses                               | : num [1:715] 2099 810 578 3613 |
| 1042 ...  |                                 |
| \$ Profit after tax                             | : num [1:715] 100.2 19.7 -42.4  |
| -37.5 21.4 ...                                  |                                 |
| \$ PBDITA                                       | : num [1:715] 285.6 116 -31     |
| 68.2 90.1 ...                                   |                                 |
| \$ PBT  | : num [1:715] 152.1 33.7 -56    |
| 25.7 29.7 ...                                   |                                 |
| \$ Cash profit                                  | : num [1:715] 182.3 50.5 -35.3  |
| 37.3 62.7 ...                                   |                                 |
| \$ PBDITA as % of total income                  | : num [1:715] 13.07 14.16 -5.5  |
| 1.91 8.71 ...                                   |                                 |
| \$ PBT as % of total income                     | : num [1:715] 6.96 4.11 -9.94   |
| 0.72 2.87 ...                                   |                                 |
| \$ PAT as % of total income                     | : num [1:715] 4.59 2.4 -7.52 -  |
| 1.05 2.07 ...                                   |                                 |
| \$ Cash profit as % of total income             | : num [1:715] 8.34 6.16 -6.26   |
| 1.04 6.06 ...                                   |                                 |
| \$ PAT as % of net worth                        | : num [1:715] 42.11 10.66 -31.2 |
| 0 6.31 ...                                      |                                 |
| \$ Sales  | : num [1:715] 2171 817 552 3573 |
| 1027 ...  |                                 |
| \$ Income from financial services               | : num [1:715] 2.3 0.8 9.1 1 0.7 |
| ...   |                                 |
| \$ Other income                                 | : num [1:715] NA 0.2 2.1 1.5    |
| 2.3 0.1 NA NA 0.1 0.1 ...                       |                                 |
| \$ Total capital                                | : num [1:715] 48 114 47.1 50.5  |
| 33 ...  |                                 |
| \$ Reserves and funds                           | : num [1:715] 413.1 97.6 227.4  |
| 150.9 316.2 ...                                 |                                 |
| \$ Deposits (accepted by commercial banks)      | : logi [1:715] NA NA NA NA NA   |
| NA ...  |                                 |
| \$ Borrowings                                   | : num [1:715] 177.3 339.8 17.5  |
| 524.2 162.3 ...                                 |                                 |
| \$ Current liabilities & provisions             | : num [1:715] 328.5 100.5 240.1 |
| 75.2 299.6 ...                                  |                                 |
| \$ Deferred tax liability                       | : num [1:715] 3.7 23.1 NA 56.7  |
| 12.2 2.1 1.9 4.4 2.9 NA ...                     |                                 |
| \$ Shareholders funds                           | : num [1:715] 276 212 120 201   |
| 349 ...   |                                 |
| \$ Cumulative retained profits                  | : num [1:715] 227.8 97.6 69.9   |
| 150.9 316.2 ...                                 |                                 |
| \$ Capital employed                             | : num [1:715] 453 551 138 726   |
| 512 ...   |                                 |
| \$ TOL/TNW                                      | : num [1:715] 1.8 2.01 1.73     |
| 2.94 1.02 0.86 0.06 1.92 0.37 1.96 ...          |                                 |
| \$ Total term liabilities / tangible net worth: | num [1:715] 0.27 0.72 0.09      |
| 0.81 0.1 0.11 0.05 0.78 0 1.81 ...              |                                 |

|   |   |             |                     |                                    |
|---|---|-------------|---------------------|------------------------------------|
| \$ Contingent liabilities / Net worth (%) | : | num [1:715] | 112.94              | 5.77                               |
| 102.83 0.65 28.78 ...                     | : | num [1:715] | 311.5               | 12.2 123.6                         |
| \$ Contingent liabilities                 | : | num [1:715] |                     |                                    |
| 1.3 100.5 ...                             | : | num [1:715] | 332                 | 199 270 263                        |
| \$ Net fixed assets                       | : | num [1:715] |                     |                                    |
| 191 ...                                   | : | num [1:715] |                     |                                    |
| \$ Investments                            | : | num [1:715] | NA NA               | 0.7 NA NA                          |
| NA 17.3 2.6 NA NA ...                     | : | num [1:715] | 560                 | 407 148 536                        |
| \$ Current assets                         | : | num [1:715] |                     |                                    |
| 472 ...                                   | : | num [1:715] | 134.2               | 123.6 -97.1                        |
| \$ Net working capital                    | : | num [1:715] |                     |                                    |
| 99.6 75.3 ...                             | : | num [1:715] | 0.92                | 0.48 0.32                          |
| \$ Quick ratio (times)                    | : | num [1:715] | 0.51                | 0.58 0.97 166 0.52 0.88 0.6 ...    |
| \$ Current ratio (times)                  | : | num [1:715] | 1.23                | 1.19 1.86 166 1.56 1.19 0.55 ...   |
| \$ Debt to equity ratio (times)           | : | num [1:715] | 2.6                 | 0.46 0.32 0.05 1.24 0 1.81 ...     |
| \$ Cash to current liabilities (times)    | : | num [1:715] | 0.08                | 0.08 0 165 0.03 0.35 0.23 ...      |
| \$ Cash to average cost of sales per day  | : | num [1:715] | 3.71                | 11.15 ...                          |
| \$ Creditors turnover                     | : | chr [1:715] | "2.35"              | "NA" "5.94" "10.59"                |
| \$ Debtors turnover                       | : | chr [1:715] | "9.6"               | "NA" "5.74" "6.03"                 |
| \$ Finished goods turnover                | : | chr [1:715] | "8.23"              | "NA" "25.11" "28.96"               |
| \$ WIP turnover                           | : | chr [1:715] | "20.01000000000002" | "18.64999999999999" "6.6" "NA" ... |
| \$ Raw material turnover                  | : | chr [1:715] | "17.57999999999998" | "2.67" "3.77" "NA" ...             |
| \$ Shares outstanding                     | : | chr [1:715] | "11400000"          | "471285" "5050000" ...             |
| \$ Equity face value                      | : | chr [1:715] | "10"                | "10" "10" "100" ...                |
| \$ EPS                                    | : | num [1:715] | -7.09               | 5.9 18.6 1.65 -90.39               |
| \$ Adjusted EPS                           | : | num [1:715] | -7.09               | 5.9 18.6 1.65 -90.39               |
| \$ Total liabilities                      | : | num [1:715] | 823 ...             | 971 675 532 858                    |
| \$ PE on BSE                              | : | chr [1:715] | "-0.16" ...         | "NA" "NA" "-15.5"                  |

#### 5.4 Five point Summary:

- **NetWorthNextYear** should be changed into categorical variables and it is our predictor/target variable
- **Possible outliers** in most of the variables.
- **Null values** are present in many columns
- PE on BSE, Equity face value ,Shares outstanding ,Raw material turnover, WIP turnover ,Finished goods turnover ,Debtors turnover ,Creditors turnover **variables are of character datatype but they should be numeric.**

```

> #summary
> summary(Train)

      Num      Networth Next Year  Total assets      Net worth
Min.   : 1      Min.   :0.00000  Min.   : 0.1  Min.   : 0.0
1st Qu.: 886    1st Qu.:0.00000  1st Qu.: 91.3  1st Qu.: 31.3
Median :1773    Median :0.00000  Median : 309.7  Median : 102.3
Mean   :1772    Mean   :0.06608  Mean   : 3443.4  Mean   : 1295.9
3rd Qu.:2658    3rd Qu.:0.00000  3rd Qu.: 1098.7 3rd Qu.: 377.3
Max.   :3545    Max.   :1.00000  Max.   :1176509.2 Max.   :613151.6

      Total income      Change in stock  Total expenses      Profit after tax
Min.   : 0.0      Min.   :-3029.40  Min.   : -0.1  Min.   : -3908.30
1st Qu.: 106.5    1st Qu.: -1.80  1st Qu.: 95.8  1st Qu.: 0.50
Median : 444.9    Median : 1.60  Median : 407.7  Median : 8.80
Mean   : 4582.8   Mean   : 41.49  Mean   : 4262.9  Mean   : 277.36
3rd Qu.: 1440.9   3rd Qu.: 18.05  3rd Qu.: 1359.8 3rd Qu.: 52.27
Max.   :2442828.2  Max.   :14185.50  Max.   :2366035.3 Max.   :119439.10
NA's   :198       NA's   :458     NA's   :139     NA's   :131

      PBDITA          PBT          Cash profit
Min.   :-440.7    Min.   :-3894.80  Min.   : -2245.70
1st Qu.: 6.9      1st Qu.: 0.70   1st Qu.: 2.90
Median : 35.4     Median : 12.40  Median : 18.85
Mean   : 578.1    Mean   : 383.81  Mean   : 392.07
3rd Qu.: 150.2    3rd Qu.: 71.97  3rd Qu.: 93.20
Max.   :208576.5   Max.   :145292.60  Max.   :176911.80
NA's   :131       NA's   :131     NA's   :131

PBDITA as % of total income PBT as % of total income PAT as % of total incom
e
      Min.   : -6400.000      Min.   : -21340.00      Min.   : -21340.00
1st Qu.: 5.000       1st Qu.: 0.55       1st Qu.: 0.35
Median : 9.660       Median : 3.31       Median : 2.34
Mean   : 4.571       Mean   : -17.28      Mean   : -19.20
3rd Qu.: 16.390       3rd Qu.: 8.80       3rd Qu.: 6.34
Max.   : 100.000      Max.   : 100.00      Max.   : 150.00
NA's   :68           NA's   :68           NA's   :68

      Cash profit as % of total income PAT as % of net worth  Sales
Min.   : -15020.000      Min.   : -748.72      Min.   : 0.1
1st Qu.: 2.020          1st Qu.: 0.00       1st Qu.: 112.7
Median : 5.640          Median : 7.92       Median : 453.1
Mean   : -8.229         Mean   : 10.27       Mean   : 4549.5
3rd Qu.: 10.700         3rd Qu.: 20.19      3rd Qu.: 1433.5
Max.   : 100.000         Max.   :2466.67      Max.   :2384984.4
NA's   :68               NA's   :68           NA's   :259

      Income from financial services Other income  Total capital
Min.   : 0.00          Min.   : 0.00       Min.   : 0.1
1st Qu.: 0.40          1st Qu.: 0.40       1st Qu.: 13.1
Median : 1.80          Median : 1.40       Median : 42.1
Mean   : 80.84          Mean   : 41.36      Mean   : 216.6
3rd Qu.: 9.68          3rd Qu.: 5.97       3rd Qu.: 100.3
Max.   :51938.20        Max.   :42856.70     Max.   :78273.2
NA's   :935             NA's   :1295       NA's   :4

      Reserves and funds Deposits (accepted by commercial banks) Borrowings
Min.   : -6525.9      Mode:logical          Min.   : 0.10
1st Qu.: 5.0          NA's:3541            1st Qu.: 23.95
Median : 54.8          NA's:3541            Median : 99.20

```

|   |  |                       |                |
|---|--|-----------------------|----------------|
| Mean : 1163.8                               |  | Mean : 1122.28        |                |
| 3rd Qu.: 277.3                              |  | 3rd Qu.: 352.60       |                |
| Max. :625137.8                              |  | Max. :278257.30       |                |
| NA's :85                                    |  | NA's :366             |                |
| Current liabilities & provisions            | Deferred tax liability                 | Shareholders funds    |                |
| Min. : 0.1                                  | Min. : 0.1                             | Min. : 0.0            |                |
| 1st Qu.: 17.8                               | 1st Qu.: 3.2                           | 1st Qu.: 32.0         |                |
| Median : 69.4                               | Median : 13.4                          | Median : 105.6        |                |
| Mean : 940.6                                | Mean : 227.2                           | Mean : 1322.1         |                |
| 3rd Qu.: 261.7                              | 3rd Qu.: 50.0                          | 3rd Qu.: 393.2        |                |
| Max. :352240.3                              | Max. :72796.6                          | Max. :613151.6        |                |
| NA's :96                                    | NA's :1140                             |                       |                |
| Cumulative retained profits                 | Capital employed                       | TOL/TNW               |                |
| Min. : -6534.3                              | Min. : 0.0                             | Min. : -350.480       |                |
| 1st Qu.: 1.1                                | 1st Qu.: 60.8                          | 1st Qu.: 0.600        |                |
| Median : 37.1                               | Median : 214.7                         | Median : 1.430        |                |
| Mean : 890.5                                | Mean : 2328.3                          | Mean : 3.994          |                |
| 3rd Qu.: 202.3                              | 3rd Qu.: 767.3                         | 3rd Qu.: 2.830        |                |
| Max. :390133.8                              | Max. :891408.9                         | Max. : 473.000        |                |
| NA's :38                                    |  |                       |                |
| Total term liabilities / tangible net worth | Contingent liabilities / Net worth (%) |                       |                |
| Min. : -325.600                             |  | Min. : 0.00           |                |
| 1st Qu.: 0.050                              |  | 1st Qu.: 0.00         |                |
| Median : 0.340                              |  | Median : 5.33         |                |
| Mean : 1.844                                |  | Mean : 53.94          |                |
| 3rd Qu.: 1.000                              |  | 3rd Qu.: 30.76        |                |
| Max. : 456.000                              |  | Max. : 14704.27       |                |
| Contingent liabilities                      | Net fixed assets                       | Investments           | Current assets |
| Min. : 0.1                                  | Min. : 0.0                             | Min. : 0.00           | Min. : 0       |
| .1  |  |                       |                |
| 1st Qu.: 6.3                                | 1st Qu.: 26.0                          | 1st Qu.: 1.00         | 1st Qu.: 36    |
| .2  |  |                       |                |
| Median : 38.0                               | Median : 93.5                          | Median : 8.35         | Median : 145   |
| .1  |  |                       |                |
| Mean : 932.9                                | Mean : 1189.7                          | Mean : 694.73         | Mean : 1293    |
| .4  |  |                       |                |
| 3rd Qu.: 192.7                              | 3rd Qu.: 344.9                         | 3rd Qu.: 64.30        | 3rd Qu.: 502   |
| .2  |  |                       |                |
| Max. :559506.8                              | Max. :636604.6                         | Max. :199978.60       | Max. :354815   |
| .2  |  |                       |                |
| NA's :1188                                  | 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 | Debtors turnover      |                 |
| Min. : 0.00                           | Length:3541        | Length:3541           |                 |
| 1st Qu.: 2.79                         | Class :character   | Class :character      |                 |
| Median : 8.03                         | Mode :character    | Mode :character       |                 |
| Mean : 158.44                         |                    |                       |                 |
| 3rd Qu.: 21.79                        |                    |                       |                 |
| Max. :128040.76                       |                    |                       |                 |
| NA's :85                              |                    |                       |                 |
| Finished goods turnover               | WIP turnover       | Raw material turnover |                 |
| Length:3541                           | Length:3541        | Length:3541           |                 |
| Class :character                      | Class :character   | Class :character      |                 |
| Mode :character                       | Mode :character    | Mode :character       |                 |
| <br>Shares outstanding                | Equity face value  | EPS                   | Adjusted EPS    |
| Length:3541                           | Length:3541        | Min. :-843181.8       | Min. :-843181.8 |
| Class :character                      | Class :character   | 1st Qu.: 0.0          | 1st Qu.: 0.0    |
| Mode :character                       | Mode :character    | Median : 1.4          | Median : 1.2    |
|                                       |                    | Mean : -220.3         | Mean : -221.5   |
|                                       |                    | 3rd Qu.: 9.6          | 3rd Qu.: 7.5    |
|                                       |                    | Max. : 34522.5        | Max. : 34522.5  |
| <br>Total liabilities                 | PE on BSE          |                       |                 |
| Min. : 0.1                            | Length:3541        |                       |                 |
| 1st Qu.: 91.3                         | Class :character   |                       |                 |
| Median : 309.7                        | Mode :character    |                       |                 |
| Mean : 3443.4                         |                    |                       |                 |
| 3rd Qu.: 1098.7                       |                    |                       |                 |
| Max. :1176509.2                       |                    |                       |                 |

### Summary of Validation Data:

|                  |                 |                 |                |
|------------------|-----------------|-----------------|----------------|
| Num              | Default - 1     | Total assets    | Net worth      |
| Min. : 1.0       | Min. :0.00000   | Min. : 0.1      | Min. : 0.1     |
| 1st Qu.:179.5    | 1st Qu.:0.00000 | 1st Qu.: 93.2   | 1st Qu.: 34.4  |
| Median :358.0    | Median :0.00000 | Median : 347.7  | Median : 120.9 |
| Mean :358.0      | Mean :0.07552   | Mean : 4218.6   | Mean : 1629.7  |
| 3rd Qu.:536.5    | 3rd Qu.:0.00000 | 3rd Qu.: 1315.3 | 3rd Qu.: 451.5 |
| Max. :715.0      | Max. :1.00000   | Max. :354727.3  | Max. :171840.0 |
| <br>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 : 14   |                                     | 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

```

## 5.5 Converting NetWorthNextYearColumn into categorical in Train dataset:

```

> ifelse(Train$`Networth Next Year` >=0,0,1)
data.frame(table(Train$`Networth Next Year`))
  Var1 Freq
1   0 3307
2   1  234

> data.frame((table(Validation$`Default - 1`)))
  Var1 Freq
1   0  661
2   1   54

```

- 3307 will have positive networthNetyear and hence will not default. 234 customers are most likely to default.
- From the above, we can understand that our dataset is highly imbalanced

8 columns are of character type, but they are numeric in nature. Hence converting them into num datatype

```

```{r}
#Converting Char columns into numeric of train dataset
Train$`Creditors turnover`=as.numeric(Train$`Creditors turnover`)
Train$`Equity face value`=as.numeric(Train$`Equity face value`)
Train$`Shares outstanding`=as.numeric(Train$`Shares outstanding`)
Train$`Raw material turnover`=as.numeric(Train$`Raw material turnover`)
Train$`WIP turnover`=as.numeric(Train$`WIP turnover`)
Train$`Finished goods turnover`=as.numeric(Train$`Finished goods turnover`)
Train$`Debtors turnover`=as.numeric(Train$`Debtors turnover`)
Train$`Creditors turnover`=as.numeric(Train$`Creditors turnover`)
Train$`PE on BSE`=as.numeric(Train$`PE on BSE`)

#Converting Char columns into numeric of test dataset
Test$`Creditors turnover`=as.numeric(Test$`Creditors turnover`)
Test$`Equity face value`=as.numeric(Test$`Equity face value`)
Test$`Shares outstanding`=as.numeric(Test$`Shares outstanding`)
Test$`Raw material turnover`=as.numeric(Test$`Raw material turnover`)
Test$`WIP turnover`=as.numeric(Test$`WIP turnover`)
Test$`Finished goods turnover`=as.numeric(Test$`Finished goods turnover`)
Test$`Debtors turnover`=as.numeric(Test$`Debtors turnover`)
Test$`Creditors turnover`=as.numeric(Test$`Creditors turnover`)
Test$`PE on BSE`=as.numeric(Test$`PE on BSE`)
```

```

## 6. Outlier Treatment:

Except below variables all other variables have outliers

|  |
|--|
| Contingent liabilities / Net worth (%) |
| Debt to equity ratio (times)           |
| Cash to current liabilities (times)    |
| Creditors turnover                     |
| Debtors turnover                       |

There are 7091 number of outlier values in all the columns in train dataset and 1398 in validation dataset. Details given below

For treating outlier's 5<sup>th</sup> quartile is taken as lower bound and 95<sup>th</sup> quartile is taken as upper bound for all the columns.

```

```{r}
#Outlier Treatment for Train Data. below 5th quartile and above 95 quartile
are chose
Outliers <- function(x){
  quantiles <- quantile( x, c(.05, .95 ),na.rm = TRUE )
  result <- length(which(x[ x < quantiles[1] ]| x[ x > quantiles[2] ]))
  cat("\nNumber of outliers:",result)
  x[ x < quantiles[1] ] <- quantiles[1]
  x[ x > quantiles[2] ] <- quantiles[2]
  cat("\nLower Bound",quantiles[1])
  cat("\nUpper Bound",quantiles[2])
  return(x)
}

#capping outliers with for all columns and checking before and after
imputation summary for each column for Train Data.
col=names(Train)
for (i in (2:length(col))) {
  cat("\n\nColumn Name:",col[i])
  cat("\n\nSummary Before Imputation:\n",summary(Train[i]))
  Train[i]=Outliers(Train[i])
  cat("\nSummary After Imputation:\n",summary(Train[i]))
}
```

```

**Sample summary after imputation of outliers is given below**

| Variable Name                    | Count of Outliers in Train Dataset | Count of Outliers in Validation Dataset | Variable Name                               | Count of Outliers in Train Dataset | Count of Outliers in Validation Dataset |
|----------------------------------|------------------------------------|---|---|------------------------------------|---|
| Total assets                     | 177                                | 36                                      | Investments                                 | 106                                | 22                                      |
| Net worth                        | 177                                | 36                                      | Current assets                              | 174                                | 35                                      |
| Total income                     | 168                                | 35                                      | Net working capital                         | 176                                | 36                                      |
| Change in stock                  | 155                                | 32                                      | Quick ratio (times)                         | 171                                | 36                                      |
| Total expenses                   | 171                                | 35                                      | Current ratio (times)                       | 173                                | 36                                      |
| Profit after tax                 | 171                                | 35                                      | Debt to equity ratio (times)                | 0                                  | 0                                       |
| PBDITA                           | 171                                | 35                                      | Cash to current liabilities (times)         | 0                                  | 0                                       |
| PBT                              | 171                                | 35                                      | Cash to average cost of sales per day       | 173                                | 35                                      |
| Cash profit                      | 171                                | 35                                      | Creditors turnover                          | 0                                  | 0                                       |
| PBDITA as % of total income      | 174                                | 36                                      | Debtors turnover                            | 0                                  | 0                                       |
| PBT as % of total income         | 174                                | 36                                      | Finished goods turnover                     | 140                                | 29                                      |
| PAT as % of total income         | 174                                | 36                                      | WIP turnover                                | 145                                | 30                                      |
| Cash profit as % of total income | 174                                | 36                                      | Raw material turnover                       | 159                                | 0                                       |
| PAT as % of net worth            | 177                                | 36                                      | Shares outstanding                          | 143                                | 30                                      |
| Sales                            | 165                                | 34                                      | Equity face value                           | 143                                | 30                                      |
| Income from financial services   | 131                                | 0                                       | EPS   | 177                                | 36                                      |
| Other income                     | 113                                | 23                                      | Adjusted EPS                                | 177                                | 36                                      |
| Total capital                    | 177                                | 36                                      | Total liabilities                           | 177                                | 36                                      |
| Reserves and funds               | 173                                | 36                                      | PE on BSE                                   | 68                                 | 15                                      |
| Borrowings                       | 159                                | 33                                      | Cumulative retained profits                 | 176                                | 36                                      |
| Current liabilities & provisions | 173                                | 35                                      | Capital employed                            | 177                                | 36                                      |
| Deferred tax liability           | 120                                | 25                                      | TOL/TNW                                     | 177                                | 36                                      |
| Shareholders funds               | 177                                | 36                                      | Total term liabilities / tangible net worth | 176                                | 36                                      |

|                        |     |    |   |      |      |
|------------------------|-----|----|---|------|------|
| Contingent liabilities | 118 | 25 | Contingent liabilities / Net worth (%)  | 0    | 0    |
| Net fixed assets       | 172 | 35 | Total Number of outliers in the dataset | 7091 | 1398 |

```

Column Name: Total.assets
Summary Before Imputation:
Min. : 0.1  1st Qu.: 91.3  Median : 309.7  Mean : 3443.4  3rd Qu.: 1098.7  Max. :1176509.2  longer object length is not a multiple
of shorter object length
Number of outliers: 177
Lower Bound 10.6
Upper Bound 8452.9
Summary After Imputation:
Min. : 10.6  1st Qu.: 91.3  Median : 309.7  Mean : 1226.9  3rd Qu.: 1098.7  Max. :8452.9

Column Name: Net.worth
Summary Before Imputation:
Min. : 0.0  1st Qu.: 31.3  Median : 102.3  Mean : 1295.9  3rd Qu.: 377.3  Max. :613151.6  longer object length is not a multiple of
shorter object length
Number of outliers: 177
Lower Bound 2.9
Upper Bound 3034.4
Summary After Imputation:
Min. : 2.9  1st Qu.: 31.3  Median : 102.3  Mean : 434.2  3rd Qu.: 377.3  Max. :3034.4

Column Name: Total.income
Summary Before Imputation:
Min. : 0.0  1st Qu.: 106.5  Median : 444.9  Mean : 4582.8  3rd Qu.: 1440.9  Max. :2442828.2  NA's :198  longer object length is
not a multiple of shorter object length
Number of outliers: 168
Lower Bound 5.7
Upper Bound 9339.68
Summary After Imputation:
Min. : 5.7  1st Qu.: 106.5  Median : 444.9  Mean : 1457.2  3rd Qu.: 1440.9  Max. :9339.7  NA's :198

```

## 7. Missing Values Imputation:

There are many missing values in the dataset.

```

> anyNA(Train)
[1] TRUE
> sum(is.na(Train))
[1] 13548
> sum(rowSums(is.na(Train)))
[1] 13548
> sum(colSums(is.na(Train)))
[1] 13548
> sum(colSums(is.na(Train)))
[1] 13548

```

Validation Dataset also have null values:

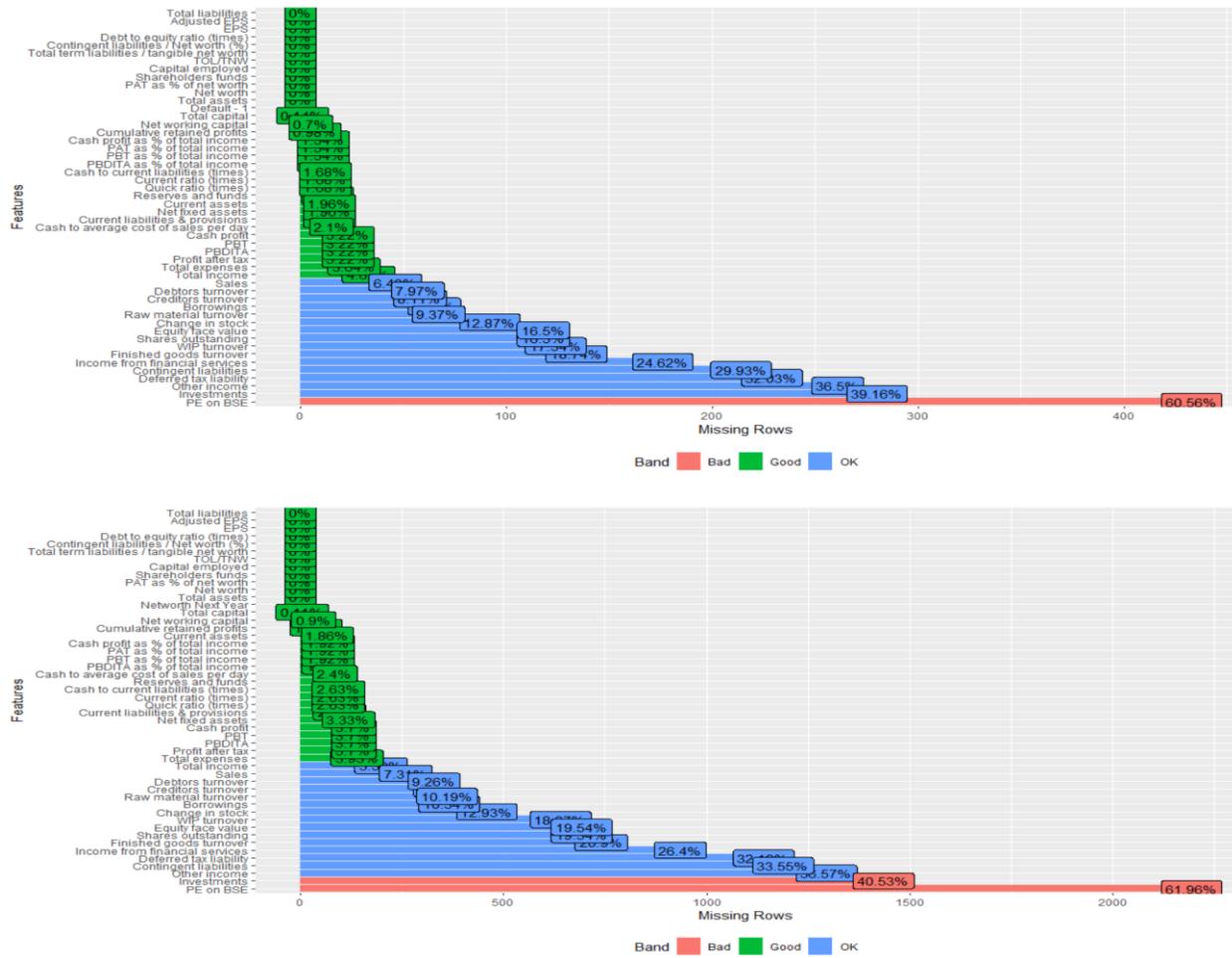
```

> anyNA(Validation)
[1] TRUE
> sum(is.na(Validation))
[1] 2593
> sum(rowSums(is.na(Validation)))
[1] 2593
> sum(colSums(is.na(Validation)))
[1] 2593

```

**Column wise missing values in Train and Validation datasets:**

| Variable Name                    | Train | Validation | Variable Name                               | Train | Validation |
|----------------------------------|-------|------------|---|-------|------------|
| Num                              | 0     | 0          | Deposits (accepted by commercial banks)     | 3541  | 715        |
| Networth Next Year               | 0     | 0          | Borrowings                                  | 366   | 65         |
| Total assets                     | 0     | 0          | Current liabilities & provisions            | 96    | 14         |
| Net worth                        | 0     | 0          | Deferred tax liability                      | 1140  | 229        |
| Total income                     | 198   | 33         | Shareholders funds                          | 0     | 0          |
| Change in stock                  | 458   | 92         | Cumulative retained profits                 | 38    | 7          |
| Total expenses                   | 139   | 26         | Capital employed                            | 0     | 0          |
| Profit after tax                 | 131   | 23         | TOL/TNW                                     | 0     | 0          |
| PBDITA                           | 131   | 23         | Total term liabilities / tangible net worth | 0     | 0          |
| PBT                              | 131   | 23         | Contingent liabilities / Net worth (%)      | 0     | 0          |
| Cash profit                      | 131   | 23         | Contingent liabilities                      | 1188  | 214        |
| PBDITA as % of total income      | 68    | 11         | Net fixed assets                            | 118   | 14         |
| PBT as % of total income         | 68    | 11         | Investments                                 | 1435  | 280        |
| PAT as % of total income         | 68    | 11         | Current assets                              | 66    | 14         |
| Cash profit as % of total income | 68    | 11         | Net working capital                         | 32    | 5          |
| PAT as % of net worth            | 0     | 0          | Quick ratio (times)                         | 93    | 12         |
| Sales                            | 259   | 46         | Current ratio (times)                       | 93    | 12         |
| Income from financial services   | 935   | 176        | Debt to equity ratio (times)                | 0     | 0          |
| Other income                     | 1295  | 261        | Cash to current liabilities (times)         | 93    | 12         |
| Total capital                    | 4     | 1          | Cash to average cost of sales per day       | 85    | 15         |
| Reserves and funds               | 85    | 13         | Creditors turnover                          | 47    | 9          |
| Shares outstanding               | 0     | 0          | Debtors turnover                            | 42    | 8          |
| Equity face value                | 0     | 0          | Finished goods turnover                     | 454   | 85         |
| EPS                              | 0     | 0          | WIP turnover                                | 354   | 75         |
| Adjusted EPS                     | 0     | 0          | Raw material turnover                       | 75    | 18         |
| Total liabilities                | 0     | 0          | PE on BSE                                   | 23    | 6          |



```
```{r}
#imputing missing values with median for train and validation dataset

NA2Median <- function(x) replace(x, is.na(x), median(x, na.rm = TRUE))
Train=replace(Train, TRUE, lapply(Train, NA2Median))
Validation=replace(Validation, TRUE, lapply(Validation, NA2Median))

```

```{r}
#converting variables as factors
Train$Default =as.factor(Train$Default)
Validation$Default=as.factor(Validation$Default)
```

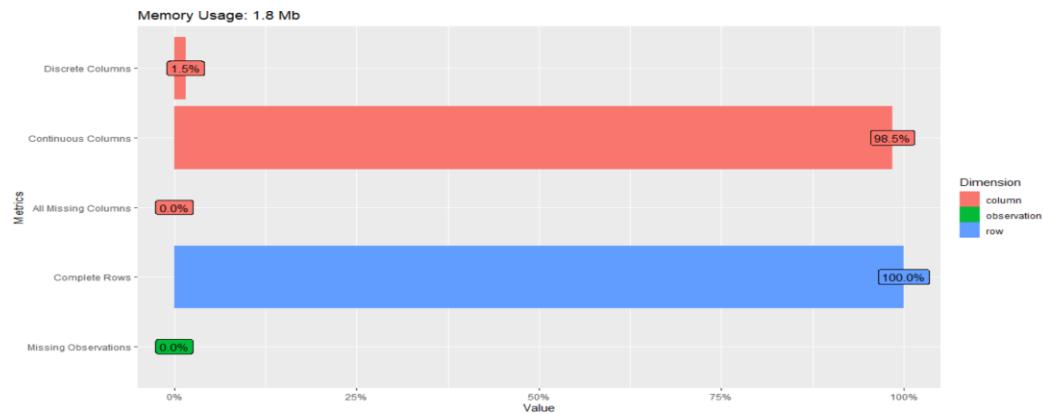
```{r}
#checking for existence of missing values
anyNA(Train)
anyNA(Validation)
```


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

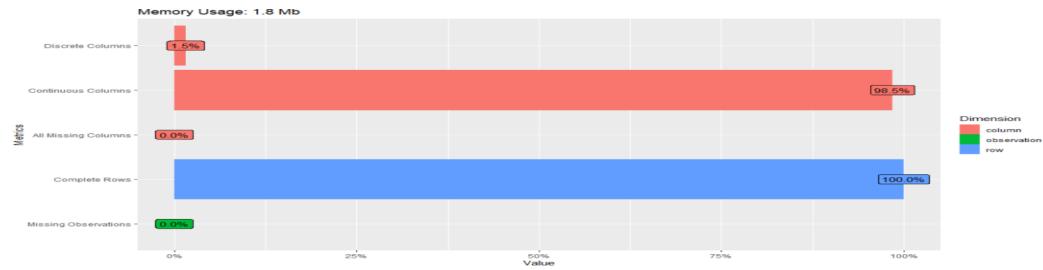

```

The missing values are imputed with median of the respective columns

### Glance on Train Dataset:



### Glance on Validation dataset:



## 8. Feature Engineering/Feature Selection:

There are many variables which have huge correlation among them. Some variables can be removed by taking their ratios

A few variables that can be created as ratios in terms of Profitability, Leverage, Liquidity, Company's Size

1. **Profitability:** This is the profit earned by company on its sale after deductions
2. **Leverage:** This is the amount of borrowed capital used to increase the potential return of an investment
3. **Liquidity:** Liquidity refers to the ease with which an asset, or security, can be converted into ready cash without affecting its market price.
4. **Company's size:** Small companies will have less capital and profits when compared to large companies. So it is ideal to take ratios so that scaled version of metrics can be used.

Below are the variables created. They can be seen in the snapshot of both train and validation dataset

### New variable Creation:

```
#profitability
Train$Profitability=Train$Profit.after.tax/Train$Sales
Train$PriceperShare=Train$EPS*Train$PE.on.BSE

#Liquidity
Train$NetWC2TotAsset=Train$Net.working.capital /Train$Total.assets
Train$Currentliabilities2Assets=Train$Current.liabilities...provisions/Trai
n$Total.assets
Train$CapitalEmployed2TotalAssets=Train$Capital.employed/Train$Total.assets
Train$TotalLiability2TotalAssets=Train$Total.liabilities/Train$Total.assets
Train$Liability2Asset=Train$Current.liabilities...provisions
/Train$Total.assets

#Leverage
Train$TotalIncome2TotalAssets=Train$Total.income/Train$Total.assets
Train$NetWC2TotAsset=Train$Net.working.capital /Train$Total.assets

#Size
Train$NetWorth2TotAssets=Train$Net.worth/Train$Total.assets|
Train$Total.expenses2TotalAssets=Train$Total.expenses/Train$Total.assets
Train$PAT2TotalAssets=Train$Profit.after.tax/Train$Total.assets
Train$PBT2TotalAssets=Train$PBT/Train$Total.assets
Train$Sales2TotalAssets=Train$Sales/Train$Total.assets
Train$NetfixedAssets2TotalAssets=Train$Net.fixed.assets/Train$Total.assets
Train$Investments2TotalAssets=Train$Investments/Train$Total.assets
```

### On Validation Dataset:

```
#profitability
Validation$Profitability=Validation$Profit.after.tax/Validation$Sales
Validation$PriceperShare=Validation$EPS*Validation$PE.on.BSE

#Liquidity
Validation$NetWC2TotAsset=Validation$Net.working.capital
/Validation$Total.assets
Validation$Currentliabilities2Assets=Validation$Current.liabilities...provi
sions/Validation$Total.assets
Validation$CapitalEmployed2TotalAssets=Validation$Capital.employed/Validati
on$Total.assets
Validation$TotalLiability2TotalAssets=Validation$Total.liabilities/Validati
on$Total.assets
Validation$Liability2Asset=Validation$Current.liabilities...provisions
/Validation$Total.assets

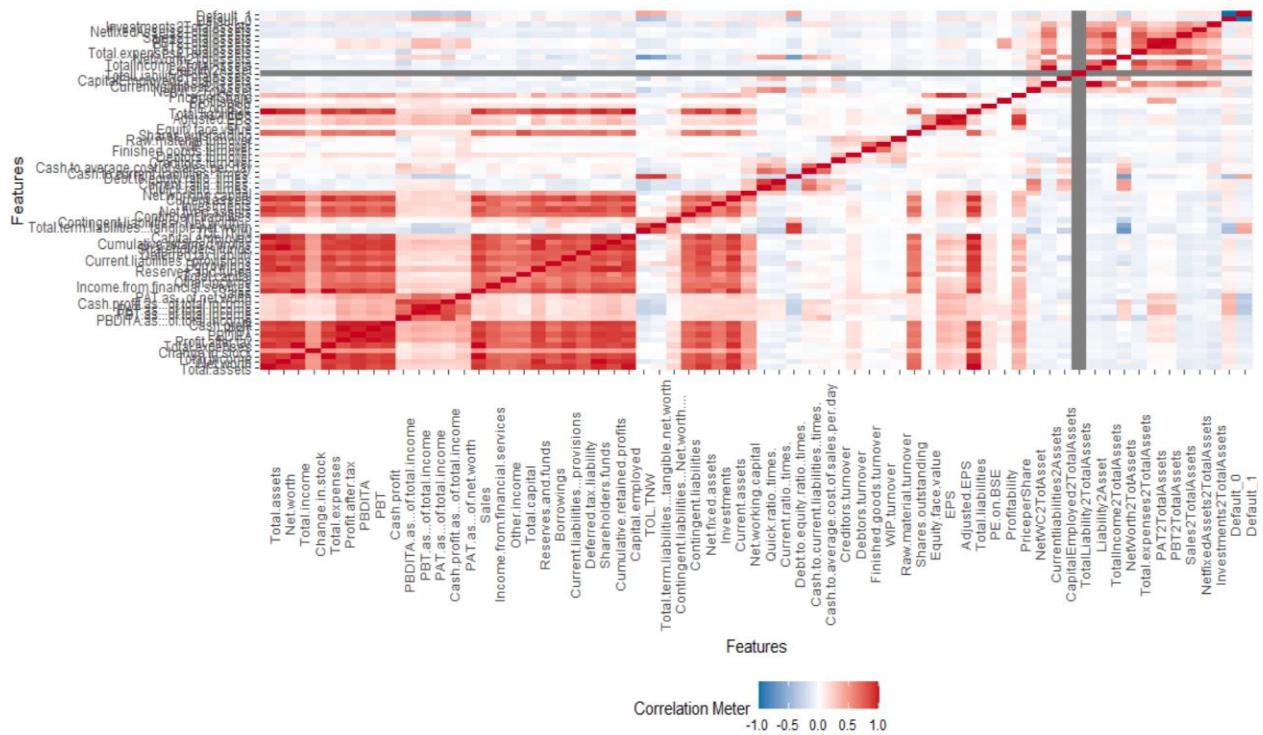
#Leverage
Validation$TotalIncome2TotalAssets=Validation$Total.income/Validation$Total
.assets
Validation$NetWC2TotAsset=Validation$Net.working.capital
/Validation$Total.assets

#Size
Validation$NetWorth2TotAssets=Validation$Net.worth/Validation$Total.assets
Validation$Total.expenses2TotalAssets=Validation$Total.expenses/Validation$T
otal.assets
Validation$PAT2TotalAssets=Validation$Profit.after.tax/Validation$Total.ass
ets
Validation$PRT2TotalAssets=Validation$PRT/Validation$Total.assets
```

### 9. MultiCollinearity:

We see there is a strong multi collinearity among the variables from correlation plot and VIF values as shown below. For logistic regression, variables should not be showing multi collinearity else model will not give proper predictions. Hence it is mandatory to treat multi collinearity first and then build the model.

## Correlation plot:



## VIF for checking the multicollinearity:

We see there are some variables showing high VIF values. Vif value should be brought back to below 10 atleast to resolve multi collinearity by dropping unnecessary and insignificant variables.

| Variable Name                    | VIF Value | Variable Name                               | VIF Value |
|----------------------------------|-----------|---|-----------|
| PBT2TotalAssets                  | 342.1348  | Total.term.liabilities...tangible.net.worth | 5.824495  |
| PAT2TotalAssets                  | 268.2384  | Quick.ratio..times.                         | 5.321089  |
| Total.income                     | 237.2525  | Investments                                 | 5.012082  |
| Total.assets                     | 155.4381  | Current.ratio..times..                      | 4.495332  |
| Total.expenses                   | 152.6424  | Cash.profit.as...of.total.income            | 3.980182  |
| Capital.employed                 | 151.6393  | Shares.outstanding                          | 3.720206  |
| PBT                              | 104.8223  | Cash.to.current.liabilities..times..        | 3.690432  |
| Profit.after.tax                 | 94.41379  | Reserves.and.funds                          | 3.666688  |
| PBT.as...of.total.income         | 49.4798   | NetWorth2TotAssets                          | 3.355113  |
| PAT.as...of.total.income         | 48.28467  | Profitability                               | 3.230809  |
| Sales                            | 47.44151  | CapitalEmployed2TotalAssets                 | 3.068736  |
| Borrowings                       | 36.14901  | Currentliabilities2Assets                   | 2.890653  |
| Current.liabilities...provisions | 27.87275  | Contingent.liabilities                      | 2.823868  |
| Shareholders.funds               | 21.70413  | Cash.profit                                 | 2.749733  |
| Current.assets                   | 18.09921  | Income.from.financial.services              | 2.694208  |

|                              |          |  |          |
|------------------------------|----------|--|----------|
| PBDITA                       | 16.1107  | Net.working.capital                    | 2.636078 |
| Total.expenses2TotalAssets   | 14.22449 | PBDITA.as...of.total.income            | 2.548979 |
| TotalIncome2TotalAssets      | 13.63574 | NetfixedAssets2TotalAssets             | 2.414152 |
| EPS                          | 12.51899 | WIP.turnover                           | 2.386911 |
| Net.fixed.assets             | 11.90802 | Investments2TotalAssets                | 2.322969 |
| Net.worth                    | 10.77864 | Cash.to.average.cost.of.sales.per.day  | 2.288799 |
| Adjusted.EPS                 | 10.7126  | Finished.goods.turnover                | 2.236348 |
| Sales2TotalAssets            | 9.585214 | Other.income                           | 2.188571 |
| Debt.to.equity.ratio..times. | 8.638438 | PAT.as...of.net.worth                  | 1.990407 |
| PriceperShare                | 8.536316 | NetWC2TotAsset                         | 1.91208  |
| TOL.TNW                      | 6.724457 | Cumulative.retained.profits            | 1.91017  |
| Total.capital                | 6.0188   | Deferred.tax.liability                 | 1.812317 |
| Creditors.turnover           | 1.584956 | Change.in.stock                        | 1.681404 |
| Raw.material.turnover        | 1.505791 | Contingent.liabilities...Net.worth.... | 1.66659  |
| Equity.face.value            | 1.26802  | Debtors.turnover                       | 1.620009 |
| PE.on.BSE                    | 1.217888 |  |          |

```
```{r}
#checking p value for each variable against Default|
col=names(finalTrain)
for (i in (2:length(col)))
{
  LogModel=glm(Default~.,data=finalTrain[ ,i] ,family="binomial")
  print(summary(LogModel))
}
```

```

### Summary of the model:

A few variables are showing as insignificant when considered their p-values. These can be discarded as and when necessary after performing Univariate and Bivariate analysis

```
Call:
glm(formula = Default ~ ., family = "binomial", data = Train[,
 -c(49, 64, 65)])

Deviance Residuals:
    Min      1Q  Median      3Q     Max 
-2.7018 -0.1908 -0.0726 -0.0078  3.8134 

Coefficients:
                                         Estimate Std. Error z value
(Intercept)                         -1.644e+00  6.225e-01 -2.641
Total.assets                           8.287e-04  8.936e-04  0.927
Net.worth                             -1.295e-03  1.165e-03 -1.112
Total.income                            7.170e-04  1.320e-03  0.543
Change.in.stock                         6.777e-03  4.542e-03  1.492
Total.expenses                         -2.183e-04  1.054e-03 -0.207
```

|   |            |           |        |
|---|------------|-----------|--------|
| Profit.after.tax                            | -8.737e-03 | 1.694e-02 | -0.516 |
| PBDITA                                      | -2.763e-03 | 2.859e-03 | -0.966 |
| PBT   | 1.371e-02  | 1.412e-02 | 0.971  |
| Cash.profit                                 | -1.060e-02 | 6.224e-03 | -1.704 |
| PBDITA.as...of.total.income                 | -2.217e-02 | 1.710e-02 | -1.296 |
| PBT.as...of.total.income                    | 4.810e-03  | 6.011e-02 | 0.080  |
| PAT.as...of.total.income                    | -4.110e-02 | 6.370e-02 | -0.645 |
| Cash.profit.as...of.total.income            | -1.324e-02 | 2.479e-02 | -0.534 |
| PAT.as...of.net.worth                       | -3.250e-02 | 7.419e-03 | -4.380 |
| Sales                                       | -6.413e-04 | 6.247e-04 | -1.027 |
| Income.from.financial.services              | 1.250e-02  | 9.917e-03 | 1.261  |
| Other.income                                | 2.145e-02  | 1.645e-02 | 1.304  |
| Total.capital                               | -4.833e-04 | 1.708e-03 | -0.283 |
| Reserves.and.funds                          | -7.956e-04 | 9.587e-04 | -0.830 |
| Borrowings                                  | -1.614e-03 | 1.142e-03 | -1.413 |
| Current.liabilities...provisions            | -1.305e-04 | 1.387e-03 | -0.094 |
| Deferred.tax.liability                      | -1.540e-03 | 3.444e-03 | -0.447 |
| Shareholders.funds                          | -4.426e-04 | 1.366e-03 | -0.324 |
| Cumulative.retained.profits                 | -1.007e-02 | 3.192e-03 | -3.157 |
| Capital.employed                            | 2.050e-04  | 1.293e-03 | 0.159  |
| TOL.TNW                                     | 4.456e-02  | 6.159e-02 | 0.724  |
| Total.term.liabilities...tangible.net.worth | -3.404e-03 | 1.341e-01 | -0.025 |
| Contingent.liabilities...Net.worth....      | 2.399e-03  | 2.246e-03 | 1.068  |
| Contingent.liabilities                      | -7.440e-04 | 7.831e-04 | -0.950 |
| Net.fixed.assets                            | 9.451e-05  | 6.928e-04 | 0.136  |
| Investments                                 | 1.383e-03  | 1.800e-03 | 0.768  |
| Current.assets                              | -1.909e-06 | 8.202e-04 | -0.002 |
| Net.working.capital                         | 2.089e-03  | 1.706e-03 | 1.225  |
| Quick.ratio..times.                         | -4.689e-01 | 3.353e-01 | -1.398 |
| Current.ratio..times.                       | -4.500e-02 | 2.254e-01 | -0.200 |
| Debt.to.equity.ratio..times.                | 1.667e-01  | 1.044e-01 | 1.596  |
| Cash.to.current.liabilities..times.         | 1.392e+00  | 6.060e-01 | 2.297  |
| Cash.to.average.cost.of.sales.per.day       | 2.717e-03  | 2.654e-03 | 1.024  |
| Creditors.turnover                          | -4.314e-03 | 1.370e-02 | -0.315 |
| Debtors.turnover                            | -9.870e-03 | 1.124e-02 | -0.878 |
| Finished.goods.turnover                     | 1.516e-03  | 3.543e-03 | 0.428  |
| WIP.turnover                                | -1.064e-02 | 9.081e-03 | -1.172 |
| Raw.material.turnover                       | -2.611e-03 | 1.440e-02 | -0.181 |
| Shares.outstanding                          | -1.500e-08 | 1.568e-08 | -0.957 |
| Equity.face.value                           | -1.394e-03 | 4.417e-03 | -0.316 |
| EPS   | 4.426e-02  | 4.574e-02 | 0.968  |
| Adjusted.EPS                                | -7.301e-02 | 4.673e-02 | -1.562 |
| PE.on.BSE                                   | -1.621e-02 | 1.014e-02 | -1.599 |
| Profitability                               | 1.469e-02  | 4.397e-02 | 0.334  |
| PriceperShare                               | 9.509e-04  | 4.124e-03 | 0.231  |
| NetWC2TotAsset                              | -1.220e+00 | 4.908e-01 | -2.487 |
| NetWorth2TotAssets                          | -4.251e+00 | 1.021e+00 | -4.165 |
| TotalIncome2TotalAssets                     | 1.861e-02  | 2.771e-02 | 0.672  |
| Total.expenses2TotalAssets                  | 9.930e-03  | 3.391e-02 | 0.293  |
| PAT2TotalAssets                             | 9.914e+00  | 4.994e+00 | 1.985  |
| PBT2TotalAssets                             | -8.066e+00 | 4.335e+00 | -1.861 |
| Sales2TotalAssets                           | 1.462e-02  | 2.145e-02 | 0.682  |
| Currentliabilities2Assets                   | 8.407e-02  | 9.824e-02 | 0.856  |
| CapitalEmployed2TotalAssets                 | 6.151e-01  | 8.480e-01 | 0.725  |
| NetfixedAssets2TotalAssets                  | 1.231e-01  | 5.937e-02 | 2.073  |

|   |              |            |       |
|---|--------------|------------|-------|
| Investments2TotalAssets                     | 8.938e-01    | 5.056e-01  | 1.768 |
| (Intercept)                                 | Pr(> z )     | 0.00827 ** |       |
| Total.assets                                | 0.35368      |            |       |
| Net.worth                                   | 0.26632      |            |       |
| Total.income                                | 0.58706      |            |       |
| Change.in.stock                             | 0.13573      |            |       |
| Total.expenses                              | 0.83600      |            |       |
| Profit.after.tax                            | 0.60598      |            |       |
| PBDITA                                      | 0.33385      |            |       |
| PBT   | 0.33145      |            |       |
| Cash.profit                                 | 0.08842 .    |            |       |
| PBDITA.as...of.total.income                 | 0.19481      |            |       |
| PBT.as...of.total.income                    | 0.93622      |            |       |
| PAT.as...of.total.income                    | 0.51879      |            |       |
| Cash.profit.as...of.total.income            | 0.59335      |            |       |
| PAT.as...of.net.worth                       | 1.19e-05 *** |            |       |
| Sales                                       | 0.30465      |            |       |
| Income.from.financial.services              | 0.20742      |            |       |
| Other.income                                | 0.19226      |            |       |
| Total.capital                               | 0.77715      |            |       |
| Reserves.and.funds                          | 0.40660      |            |       |
| Borrowings                                  | 0.15757      |            |       |
| Current.liabilities...provisions            | 0.92501      |            |       |
| Deferred.tax.liability                      | 0.65478      |            |       |
| Shareholders.funds                          | 0.74598      |            |       |
| Cumulative.retained.profits                 | 0.00160 **   |            |       |
| Capital.employed                            | 0.87402      |            |       |
| TOL.TNW                                     | 0.46936      |            |       |
| Total.term.liabilities...tangible.net.worth | 0.97975      |            |       |
| Contingent.liabilities...Net.worth....      | 0.28566      |            |       |
| Contingent.liabilities                      | 0.34208      |            |       |
| Net.fixed.assets                            | 0.89149      |            |       |
| Investments                                 | 0.44223      |            |       |
| Current.assets                              | 0.99814      |            |       |
| Net.working.capital                         | 0.22066      |            |       |
| Quick.ratio..times.                         | 0.16203      |            |       |
| Current.ratio..times.                       | 0.84172      |            |       |
| Debt.to.equity.ratio..times.                | 0.11044      |            |       |
| Cash.to.current.liabilities..times.         | 0.02162 *    |            |       |
| Cash.to.average.cost.of.sales.per.day       | 0.30604      |            |       |
| Creditors.turnover                          | 0.75285      |            |       |
| Debtors.turnover                            | 0.37984      |            |       |
| Finished.goods.turnover                     | 0.66864      |            |       |
| WIP.turnover                                | 0.24118      |            |       |
| Raw.material.turnover                       | 0.85607      |            |       |
| Shares.outstanding                          | 0.33866      |            |       |
| Equity.face.value                           | 0.75234      |            |       |
| EPS   | 0.33320      |            |       |
| Adjusted.EPS                                | 0.11818      |            |       |
| PE.on.BSE                                   | 0.10991      |            |       |
| Profitability                               | 0.73832      |            |       |
| PriceperShare                               | 0.81762      |            |       |
| NetWC2TotAsset                              | 0.01289 *    |            |       |
| NetWorth2TotAssets                          | 3.12e-05 *** |            |       |

```

TotalIncome2TotalAssets          0.50182
Total.expenses2TotalAssets      0.76967 *
PAT2TotalAssets                 0.04715 *
PBT2TotalAssets                  0.06281 .
Sales2TotalAssets                0.49534
Currentliabilities2Assets       0.39210
CapitalEmployed2TotalAssets     0.46821
NetfixedAssets2TotalAssets       0.03816 *
Investments2TotalAssets         0.07709 .

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1723.67 on 3540 degrees of freedom
Residual deviance: 819.04 on 3479 degrees of freedom
AIC: 943.04

Number of Fisher Scoring iterations: 11

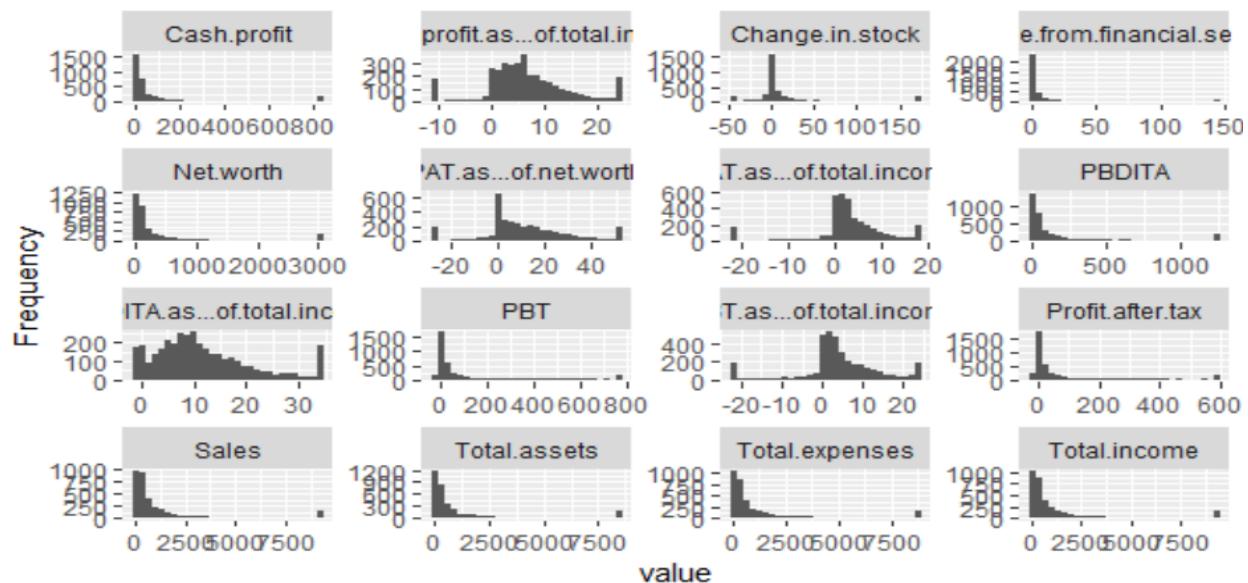
```

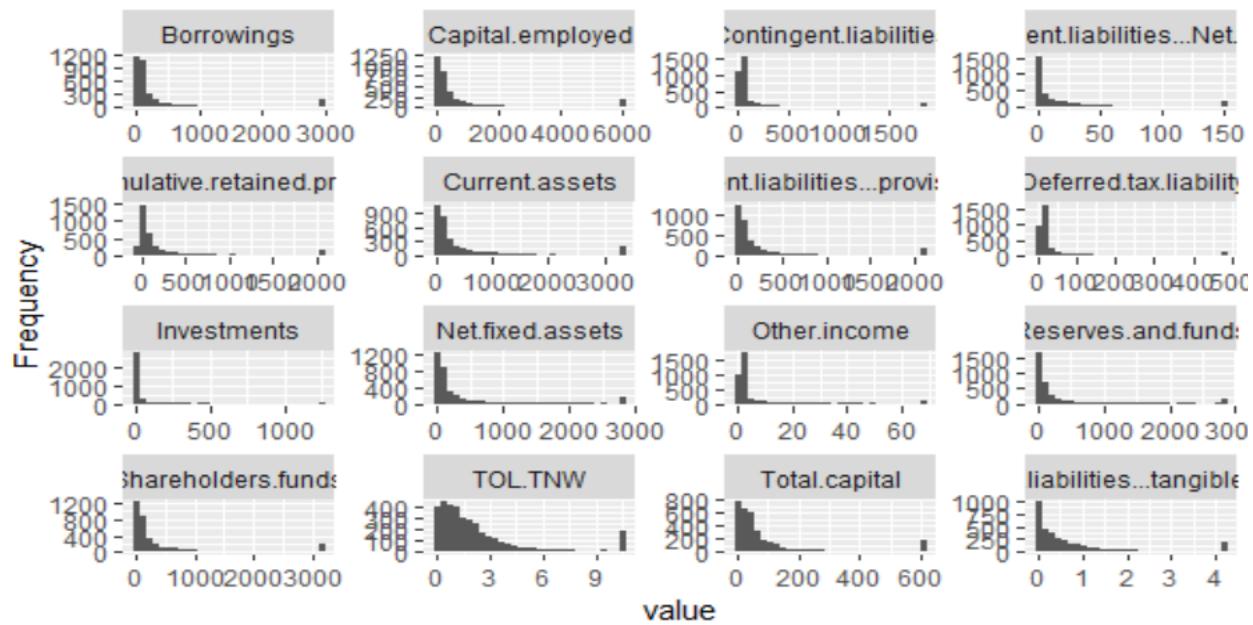
There are many variables which are not significant. To treat multi collinearity, VIF values should be below 10. After Univariate and Bivariate Analysis, we will remove few variables iteratively and bring down the VIF value below 10. (multi collinearity is treated after Univariate and Bivariate Analysis)

## **10.Univariate Analysis:**

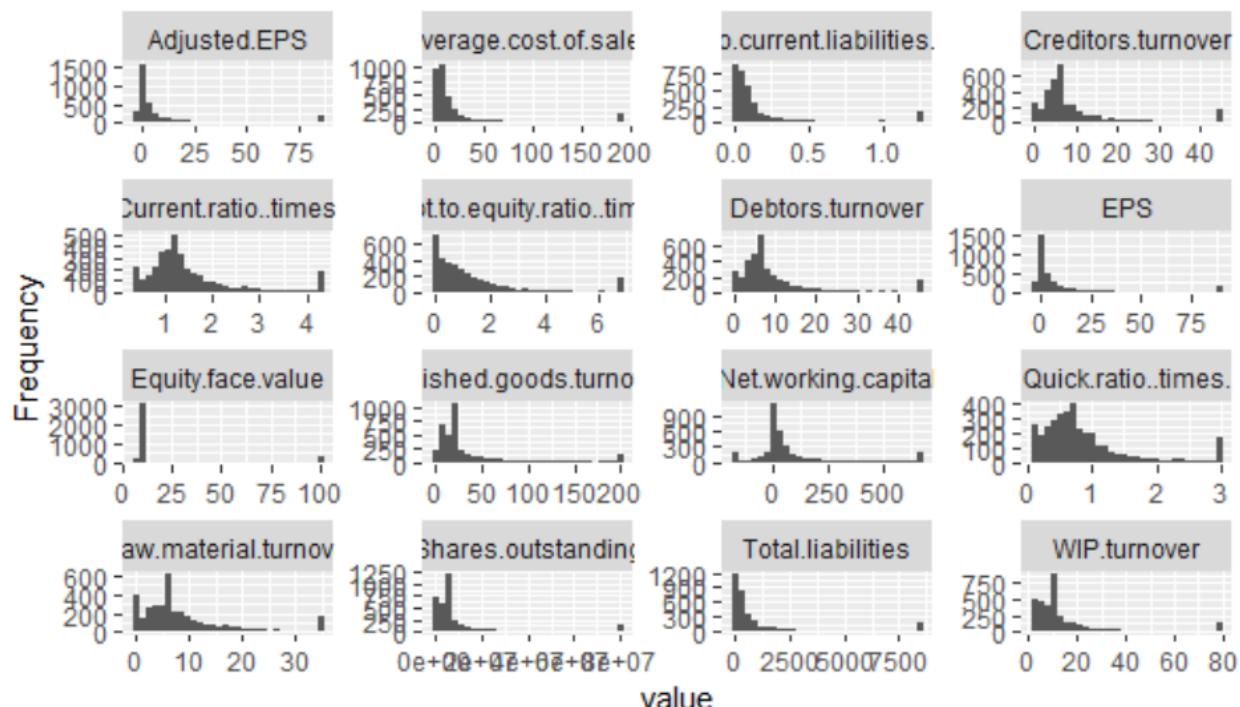
### **a. Histograms:**

We see all the variables are **highly right skewed/positive skewed** except Netwc2TotalAssets column which is close to normal distribution

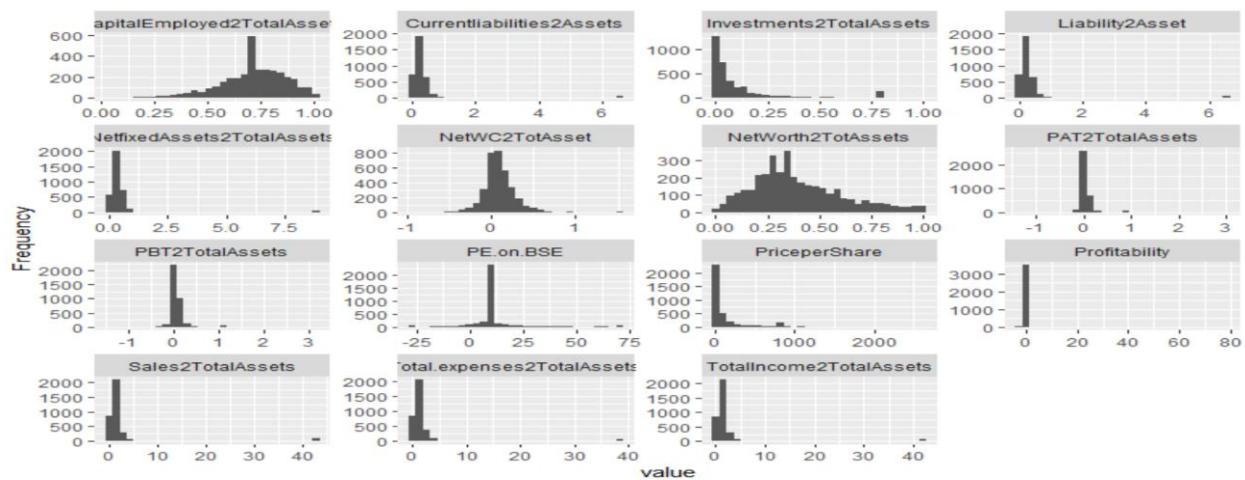




Page 2



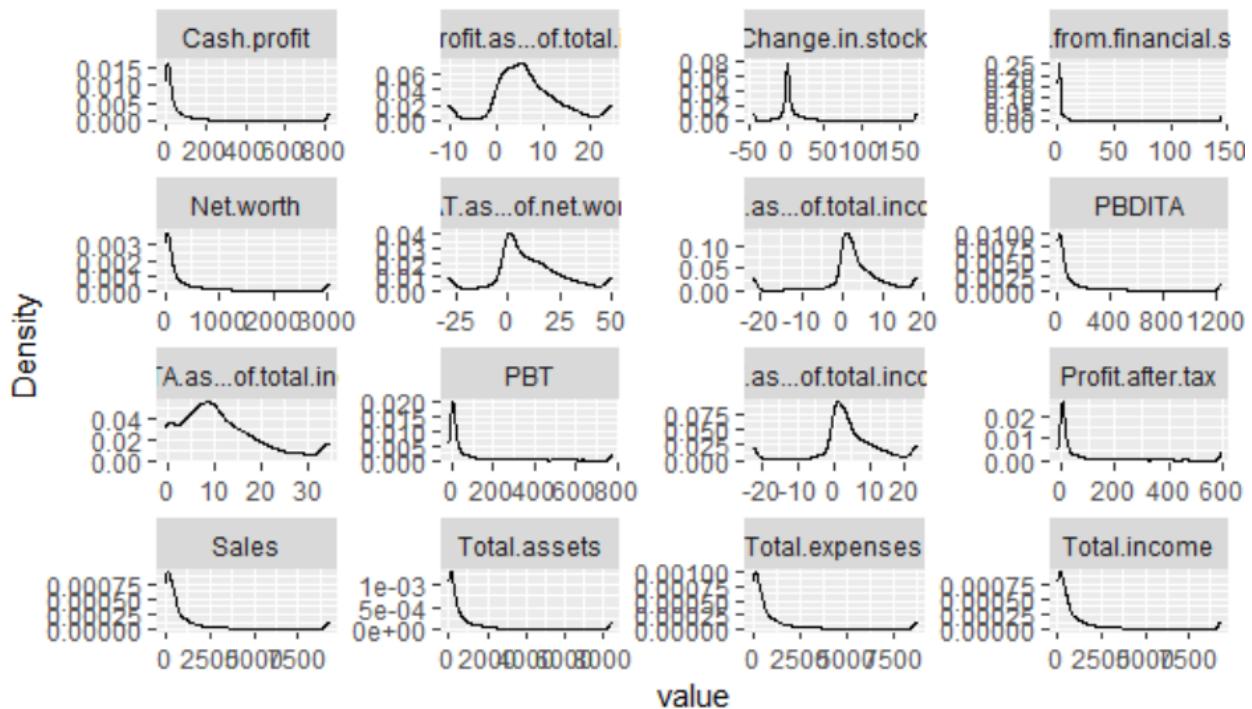
Page 3



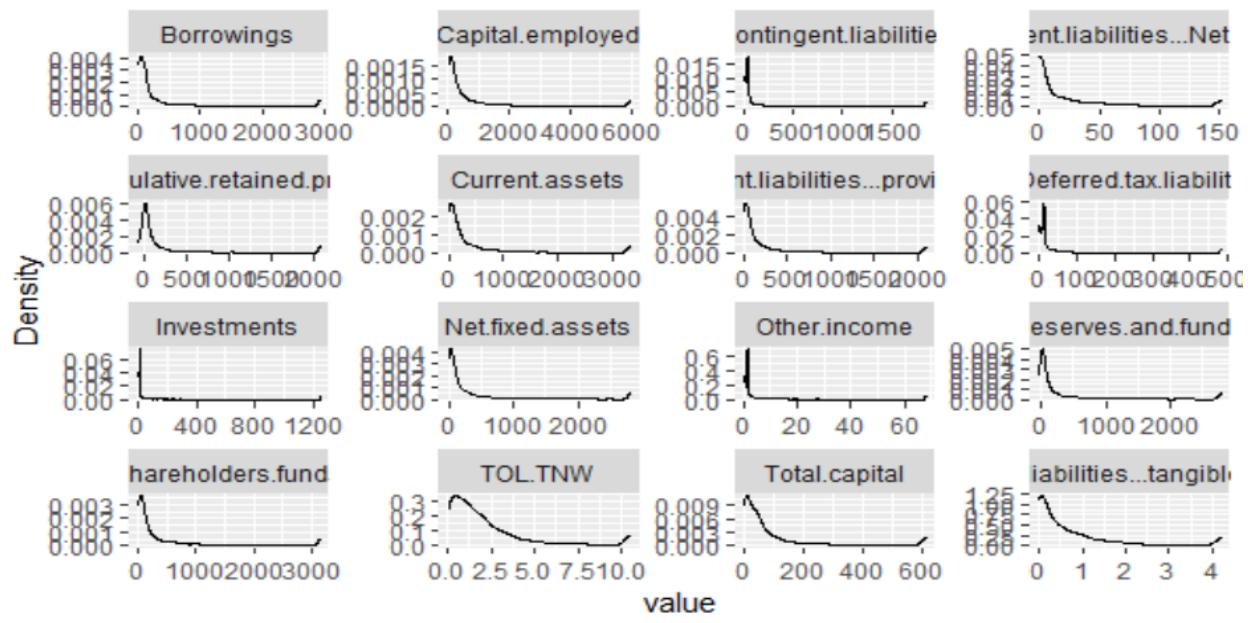
Page 4

### b. Density Plots:

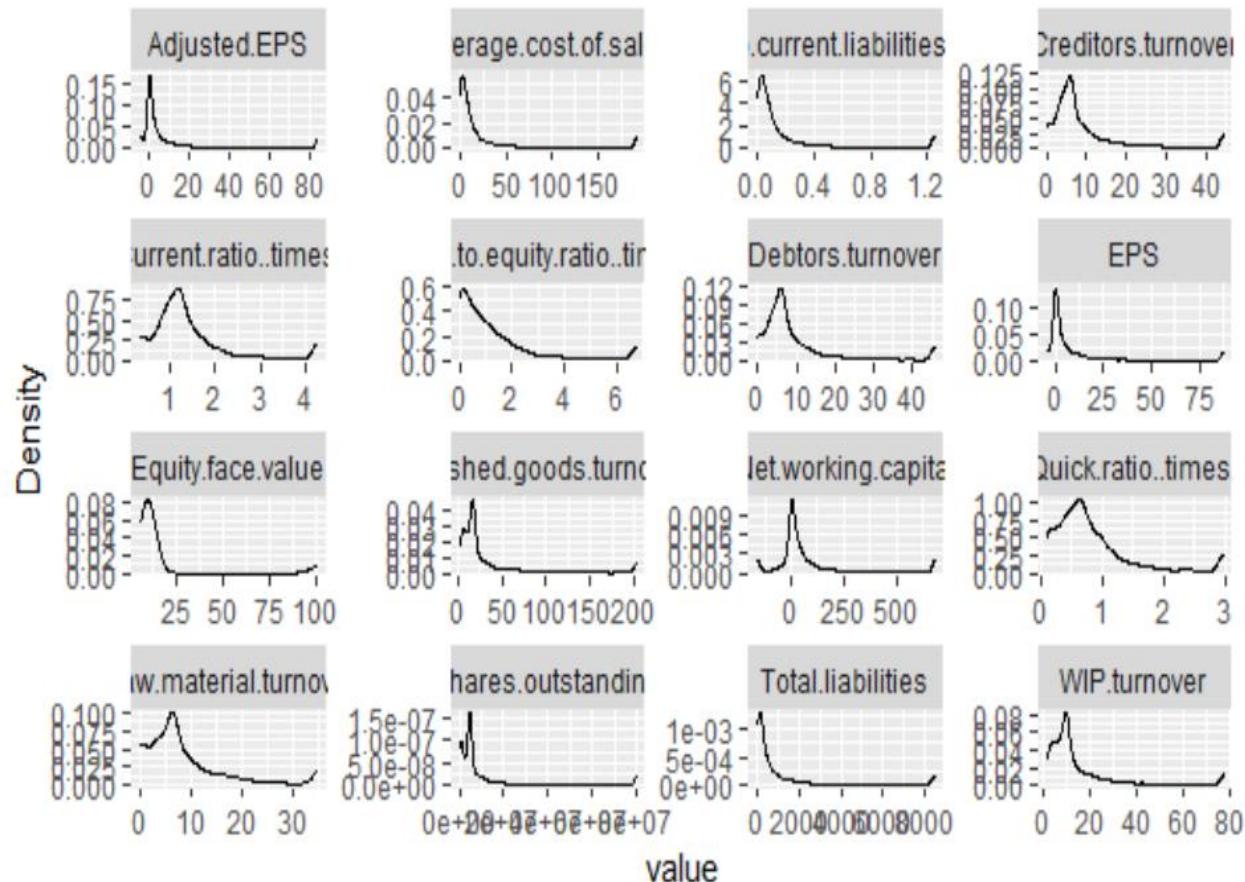
From density plots also we got the same insights as we got from the histogram plots



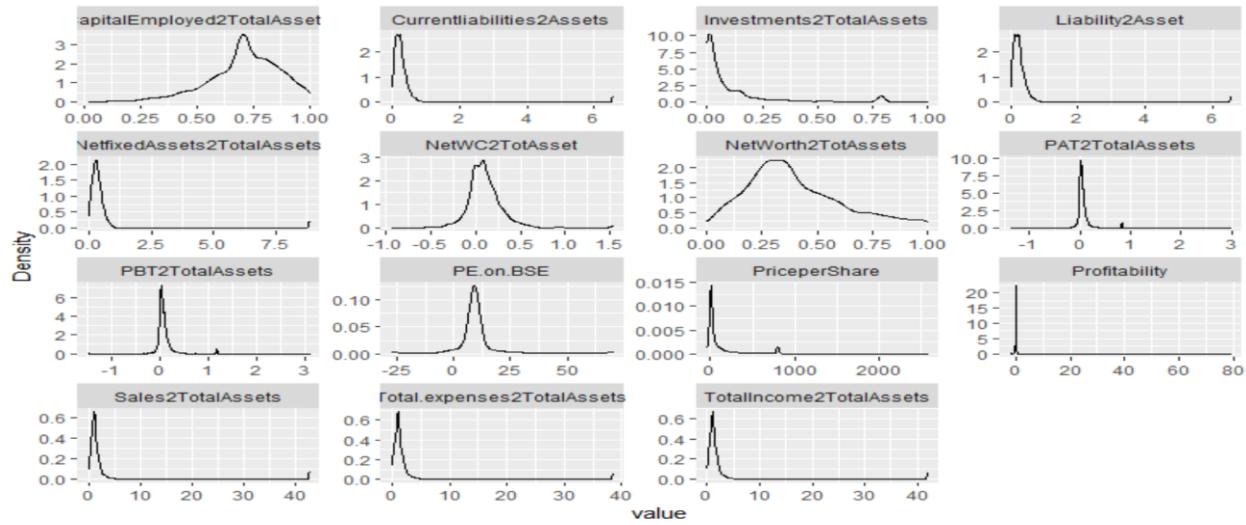
Page 1



Page 2



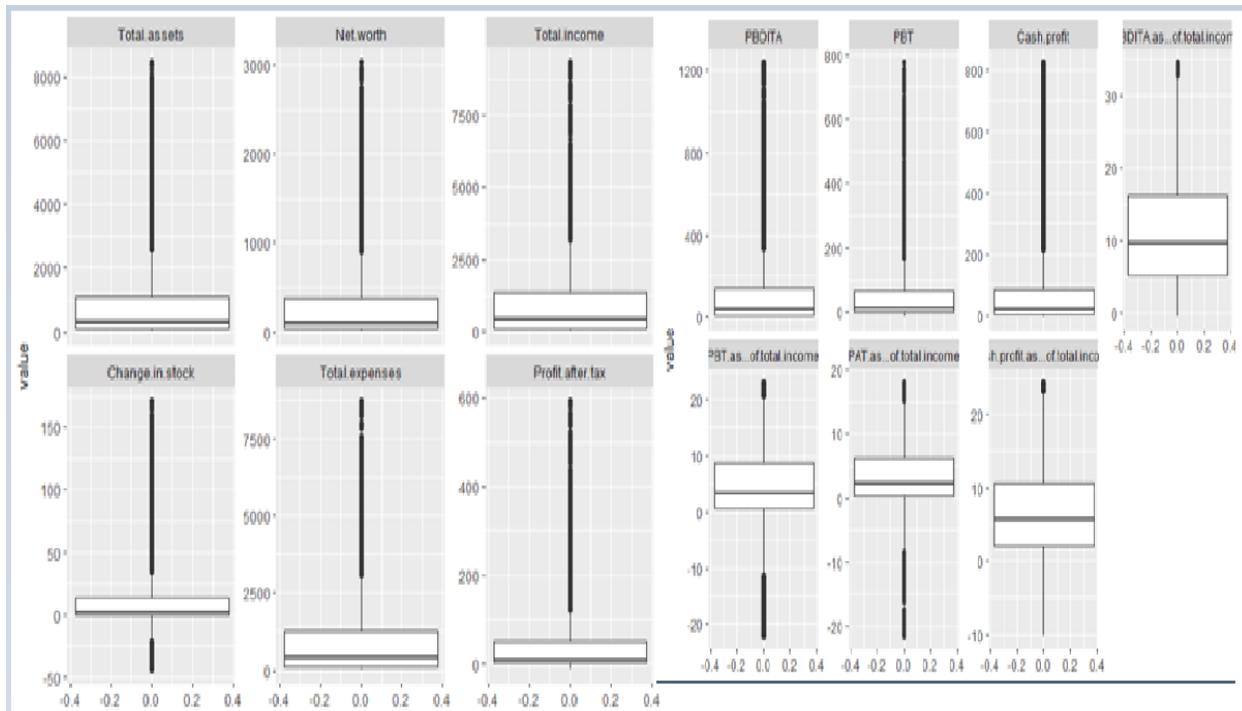
Page 3

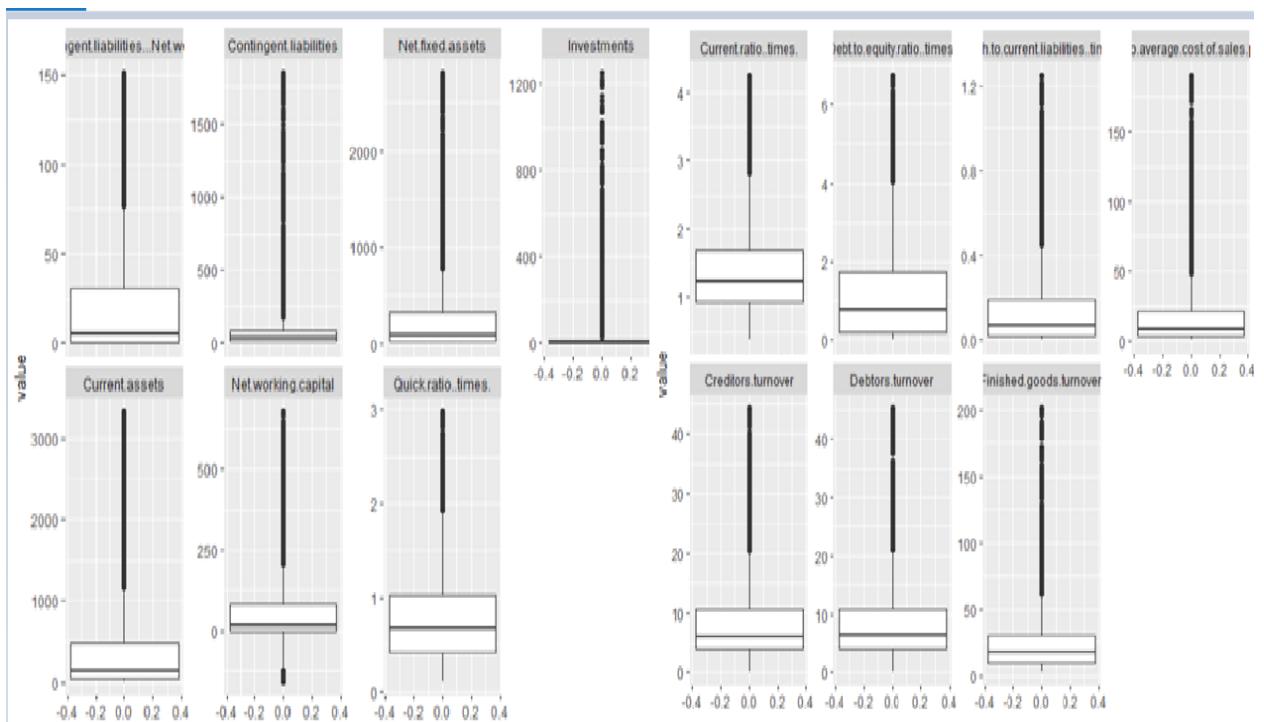
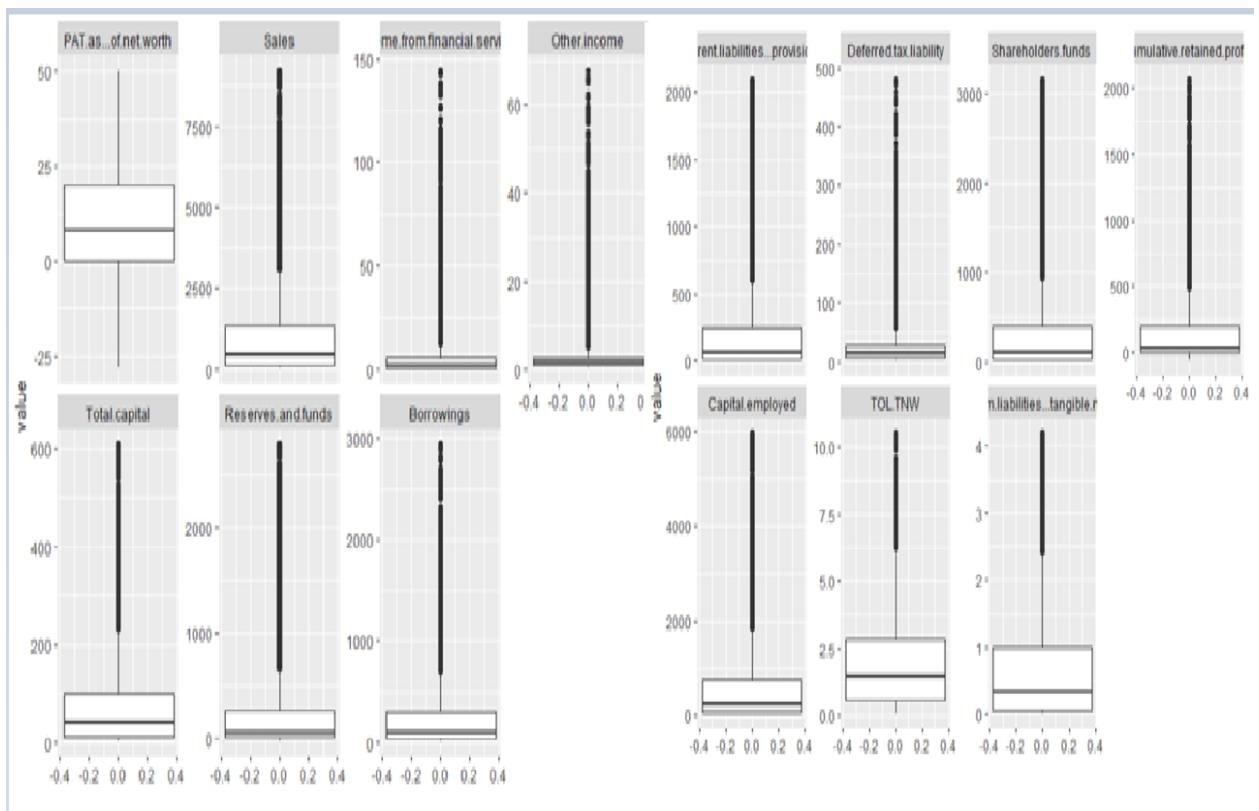


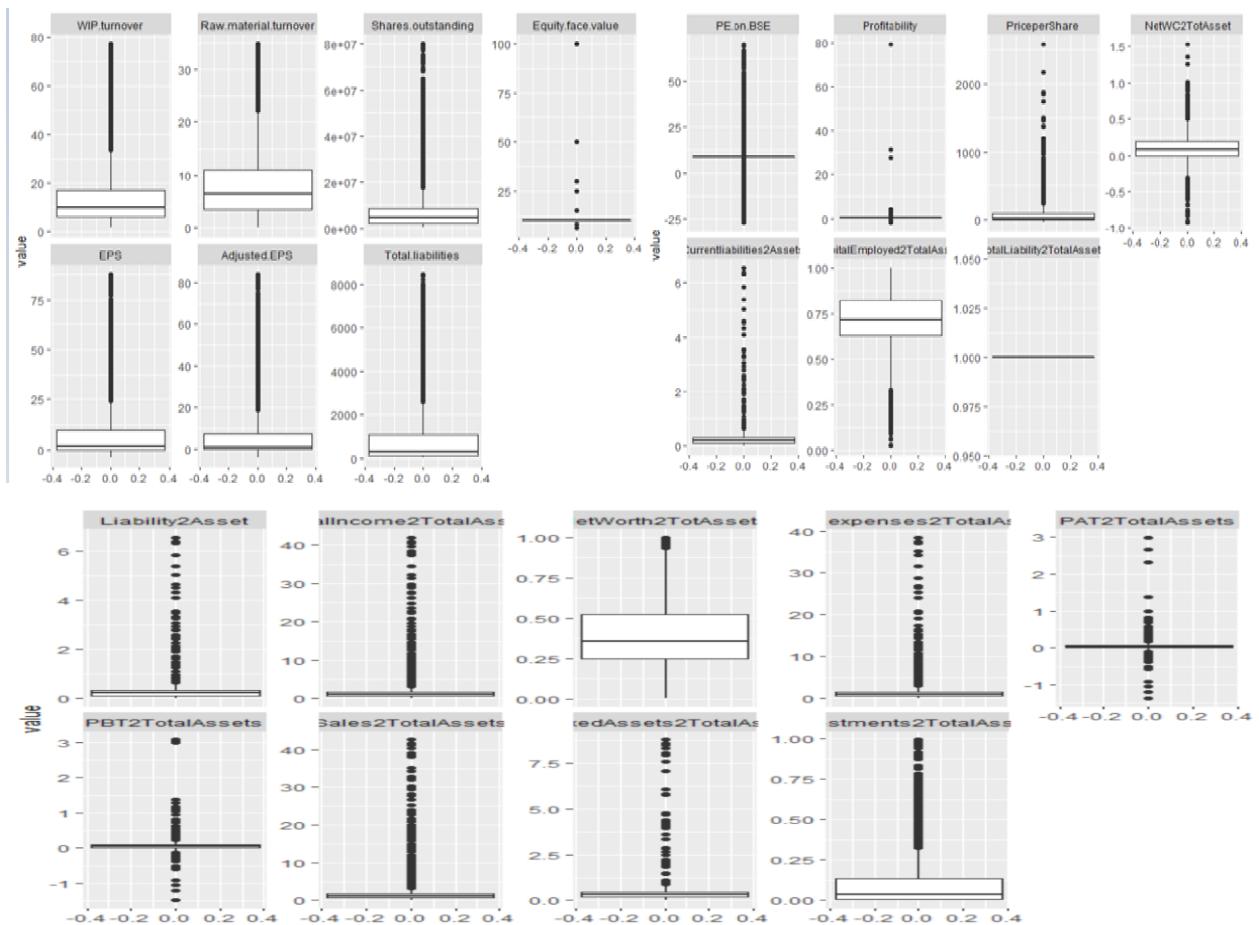
Page 4

### c. Boxplots:

We have imputed outliers which are more than 99<sup>th</sup> quartile and under 5<sup>th</sup> quartile with higher and lower bounds respectively for all the columns. Some variables which are in terms of ratios are not coming out great when boxplots are plotted on them.



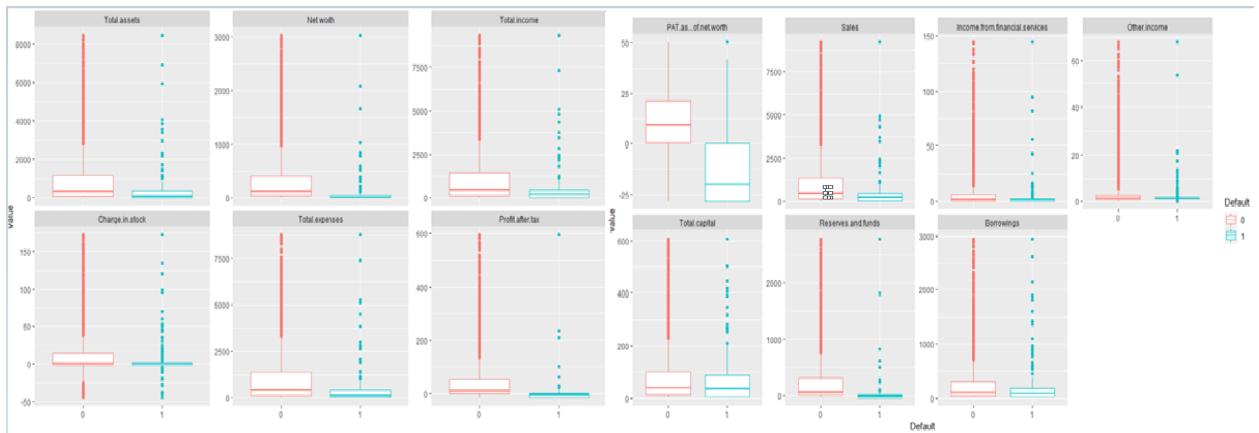


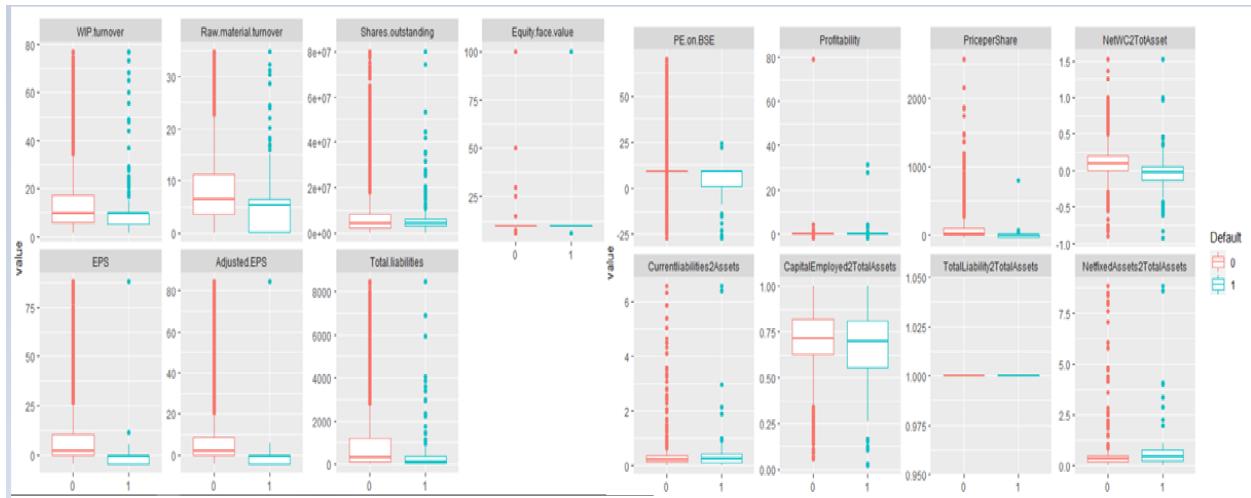
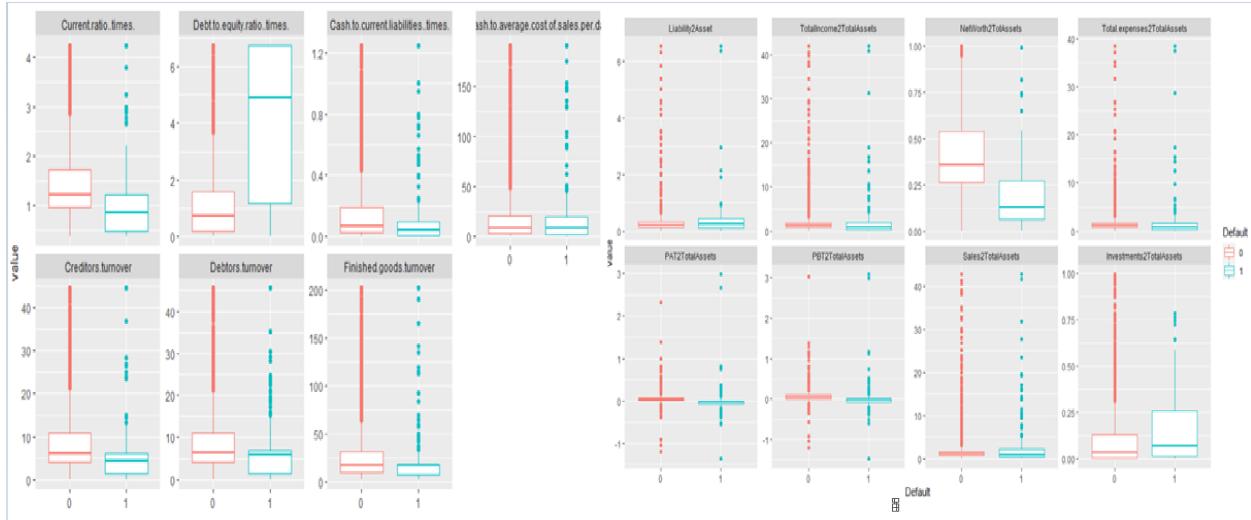
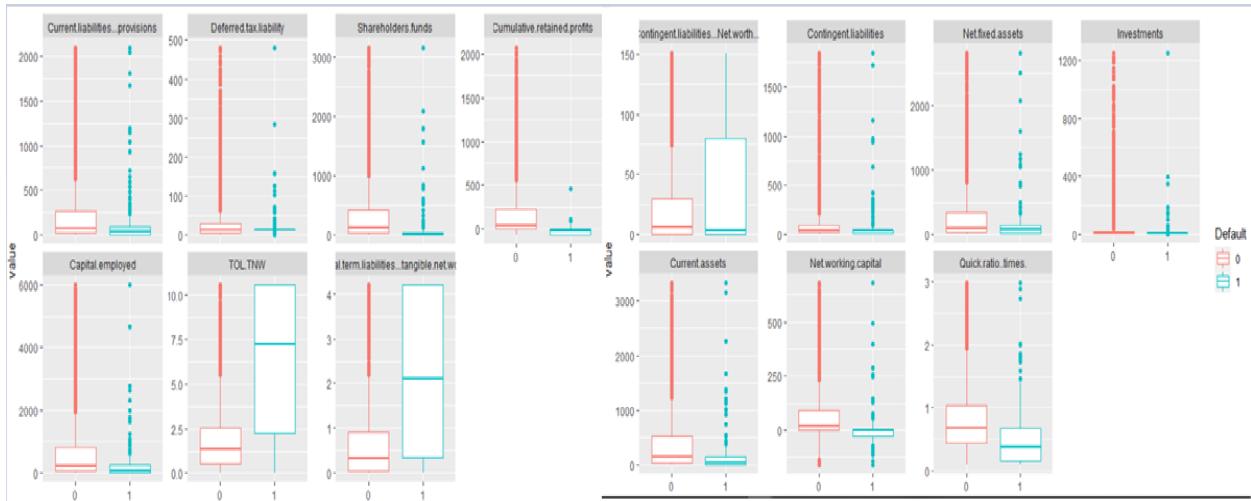


## **11. Bivariate Analysis:**

### a. Boxplot:

profit after tax networth ratio, TOL/TNW, currentliabilities/tangibleNetworth, Quickratio, times, networth2total assets, debittoequityratio are few columns out of many which are influencing the credit default variable.

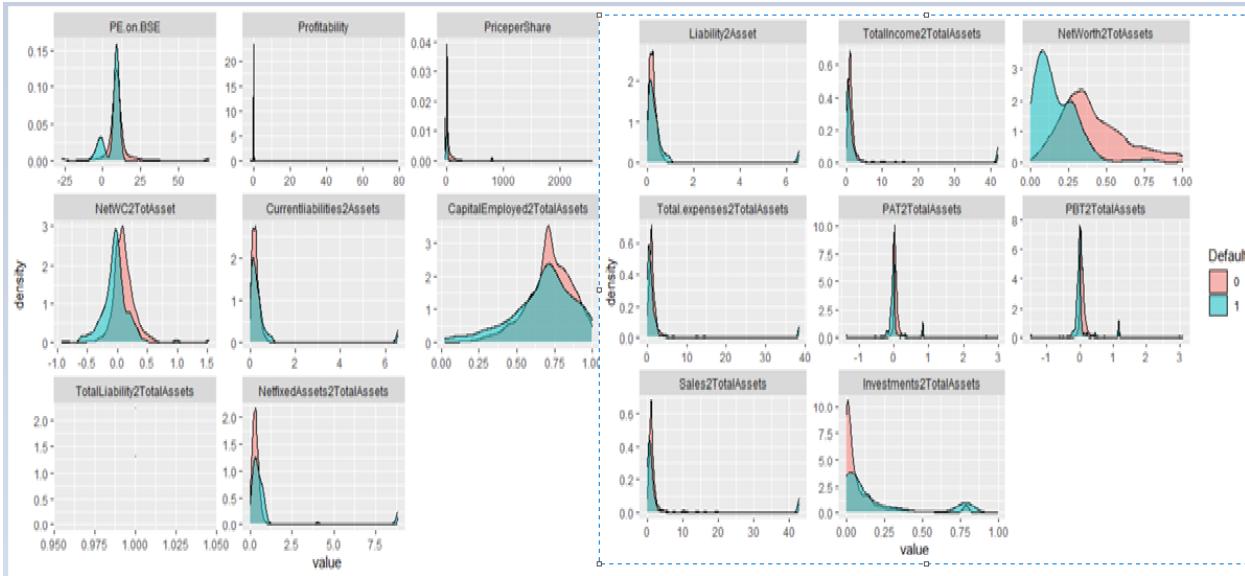




## b. Density plots:

From density plots also, we got similar insights as we got from boxplots





## **12.Logistic Regression:**

After carefully analyzing the variables from Bivariate analysis and checking on the multi collinearity, below are the variables shortlisted and the VIF values are <10

### **a. Important Variables:**

| Column Name                                 | VIF Value | Column Name                           | VIF Value |
|---|-----------|---------------------------------------|-----------|
| Net.worth                                   | 3.809427  | Creditors.turnover                    | 1.524403  |
| Change.in.stock                             | 1.432038  | Debtors.turnover                      | 1.463954  |
| Profit.after.tax                            | 3.847798  | Finished.goods.turnover               | 2.207426  |
| PBDITA                                      | 5.187744  | WIP.turnover                          | 2.338343  |
| Cash.profit                                 | 1.664086  | Raw.material.turnover                 | 1.452102  |
| PBDITA.as...of.total.income                 | 2.413272  | Adjusted.EPS                          | 9.451546  |
| PAT.as...of.total.income                    | 3.696946  | PE.on.BSE                             | 1.162881  |
| PAT.as...of.net.worth                       | 1.930808  | PriceperShare                         | 9.065594  |
| Sales                                       | 4.091222  | NetWC2TotAsset                        | 1.916602  |
| Other.income                                | 1.959579  | NetWorth2TotAssets                    | 3.244046  |
| Reserves.and.funds                          | 2.688437  | TotalIncome2TotalAssets               | 10.19342  |
| Current.liabilities...provisions            | 5.131612  | Total.expenses2TotalAssets            | 7.253689  |
| Deferred.tax.liability                      | 1.649401  | PAT2TotalAssets                       | 3.546271  |
| Cumulative.retained.profits                 | 1.459177  | Sales2TotalAssets                     | 8.387164  |
| TOL.TNW                                     | 6.368814  | Currentliabilities2Assets             | 2.428178  |
| Total.term.liabilities...tangible.net.worth | 5.377804  | CapitalEmployed2TotalAssets           | 2.667605  |
| Contingent.liabilities                      | 2.309631  | NetfixedAssets2TotalAssets            | 2.318043  |
| Net.fixed.assets                            | 5.612015  | Investments2TotalAssets               | 2.267522  |
| Investments                                 | 3.868602  | Cash.to.average.cost.of.sales.per.day | 1.662141  |

|                       |          |                                  |          |
|-----------------------|----------|----------------------------------|----------|
| Net.working.capital   | 1.715427 | Cash.profit.as...of.total.income | 3.730071 |
| Quick.ratio..times.   | 4.403852 | Debt.to.equity.ratio..times.     | 8.324668 |
| Current.ratio..times. | 4.617148 |                                  |          |

### b. Model Building:

```
```{r}
#Building Logistic Regression Model
attach(finalTrain)
LogModel=glm(Default~.,family="binomial",data=finalTrain)
summary(LogModel)
vif(LogModel)
```

```

### c. Equation:

Log(y)= -1.6669884 + (-0.0005171)\*Net.worth+(0.0065133)\*Change.in.stock+(0.0024627)\*Profit.after.tax+(0.0003144)\*PBDITA+(-0.0114943)\*Cash.profit+(-0.0229665)\*PBDITA.as...of.total.income+(-0.0288296)\*PAT.as...of.total.income+(-0.031991)\*PAT.as...of.net.worth+(-0.0001742)\*Sales+(0.0237627)\*Other.income+(-0.0001295)\*Reserves.and.funds+(0.0005216)\*Current.liabilities...provisions+(-0.0013699)\*Deferred.tax.liability+(-0.0095506)\*Cumulative.retained.profits+(0.0483636)\*TOL.TNW+(0.0130512)\*Total.term.liabilities...tangible.net.worth+(-0.0004079)\*Contingent.liabilities+(-0.0003327)\*Net.fixed.assets+(0.0004254)\*Investments+(0.0009522)\*Net.working.capital+(-0.1508673)\*Quick.ratio..times.+(+0.0358712)\*Current.ratio..times.+(+0.1645935)\*Debt.to.equity.ratio..times.+(-0.0004675)\*Creditors.turnover+(-0.0032909)\*Debtors.turnover+(0.0011516)\*Finished.goods.turnover+(-0.0105576)\*WIP.turnover+(-0.0075354)\*Raw.material.turnover+(-0.0350105)\*Adjusted.EPS+(-0.0147043)\*PE.on.BSE+(0.0019179)\*PriceperShare+(-1.1929566)\*NetWC2TotAsset+(-4.1929003)\*NetWorth2TotAssets+(0.0174048)\*TotalIncome2TotalAssets+(-0.0214709)\*Total.expenses2TotalAssets+(0.8399423)\*PAT2TotalAssets+(0.0083625)\*Sales2TotalAssets+(-0.0139434)\*Currentliabilities2Assets+(0.1568767)\*CapitalEmployed2TotalAssets+(0.1294064)\*NetfixedAssets2TotalAssets+(1.0914202)\*Investments2TotalAssets+(0.0059524)\*Cash.to.average.cost.of.sales.per.day+(-0.0128088)\*Cash.profit.as...of.total.income

```
glm.fit: fitted probabilities numerically 0 or 1 occurred
Call:
glm(formula = Default ~ ., family = "binomial", data = finalTrain)

Deviance Residuals:
    Min      1Q  Median      3Q     Max 
-2.5352 -0.1971 -0.0786 -0.0086  4.0081 

Coefficients:
                Estimate Std. Error
(Intercept) -1.6669884  0.5892984
Net.worth   -0.0005171  0.0006214
```

|   |            |              |
|---|------------|--------------|
| Change.in.stock                             | 0.0065133  | 0.0040340    |
| Profit.after.tax                            | 0.0024627  | 0.0035696    |
| PBDITA                                      | 0.0003144  | 0.0016528    |
| Cash.profit                                 | -0.0114943 | 0.0055292    |
| PBDITA.as...of.total.income                 | -0.0229665 | 0.0164879    |
| PAT.as...of.total.income                    | -0.0288296 | 0.0175473    |
| PAT.as...of.net.worth                       | -0.0319910 | 0.0072542    |
| Sales                                       | -0.0001742 | 0.0001833    |
| Other.income                                | 0.0237627  | 0.0145708    |
| Reserves.and.funds                          | -0.0001295 | 0.0007341    |
| Current.liabilities...provisions            | 0.0005216  | 0.0005705    |
| Deferred.tax.liability                      | -0.0013699 | 0.0031018    |
| Cumulative.retained.profits                 | -0.0095506 | 0.0027985    |
| TOL.TNW                                     | 0.0483636  | 0.0594201    |
| Total.term.liabilities...tangible.net.worth | 0.0130512  | 0.1277252    |
| Contingent.liabilities                      | -0.0004079 | 0.0006865    |
| Net.fixed.assets                            | -0.0003327 | 0.0004510    |
| Investments                                 | 0.0004254  | 0.0014860    |
| Net.working.capital                         | 0.0009522  | 0.0012941    |
| Quick.ratio..times.                         | -0.1508673 | 0.3010046    |
| Current.ratio..times.                       | 0.0358712  | 0.2285181    |
| Debt.to.equity.ratio..times.                | 0.1645935  | 0.1014634    |
| Creditors.turnover                          | -0.0004675 | 0.0132312    |
| Debtors.turnover                            | -0.0032909 | 0.0105514    |
| Finished.goods.turnover                     | 0.0011516  | 0.0035069    |
| WIP.turnover                                | -0.0105576 | 0.0089694    |
| Raw.material.turnover                       | -0.0075354 | 0.0141966    |
| Adjusted.EPS                                | -0.0350105 | 0.0430111    |
| PE.on.BSE                                   | -0.0147043 | 0.0097124    |
| PriceperShare                               | 0.0019179  | 0.0043245    |
| NetWC2TotAsset                              | -1.1929566 | 0.4883979    |
| NetWorth2TotAssets                          | -4.1929003 | 0.9899898    |
| TotalIncome2TotalAssets                     | 0.0174048  | 0.0235644    |
| Total.expenses2TotalAssets                  | -0.0214709 | 0.0240107    |
| PAT2TotalAssets                             | 0.8399423  | 0.5412538    |
| Sales2TotalAssets                           | 0.0083625  | 0.0198350    |
| Currentliabilities2Assets                   | -0.0139434 | 0.0899258    |
| CapitalEmployed2TotalAssets                 | 0.1568767  | 0.7790753    |
| NetfixedAssets2TotalAssets                  | 0.1294064  | 0.0576166    |
| Investments2TotalAssets                     | 1.0914202  | 0.4923305    |
| Cash.to.average.cost.of.sales.per.day       | 0.0059524  | 0.0022244    |
| Cash.profit.as...of.total.income            | -0.0128088 | 0.0239498    |
|   | z value    | Pr(> z )     |
| (Intercept)                                 | -2.829     | 0.004673 **  |
| Net.worth                                   | -0.832     | 0.405293     |
| Change.in.stock                             | 1.615      | 0.106398     |
| Profit.after.tax                            | 0.690      | 0.490263     |
| PBDITA                                      | 0.190      | 0.849148     |
| Cash.profit                                 | -2.079     | 0.037634 *   |
| PBDITA.as...of.total.income                 | -1.393     | 0.163642     |
| PAT.as...of.total.income                    | -1.643     | 0.100390     |
| PAT.as...of.net.worth                       | -4.410     | 1.03e-05 *** |
| Sales                                       | -0.950     | 0.341879     |
| Other.income                                | 1.631      | 0.102925     |
| Reserves.and.funds                          | -0.176     | 0.859990     |

```

Current.liabilities...provisions          0.914 0.360584
Deferred.tax.liability                  -0.442 0.658743
Cumulative.retained.profits            -3.413 0.000643 ***
TOL.TNW                                0.814 0.415687
Total.term.liabilities...tangible.net.worth 0.102 0.918612
Contingent.liabilities                 -0.594 0.552435
Net.fixed.assets                        -0.738 0.460793
Investments                            0.286 0.774656
Net.working.capital                   0.736 0.461862
Quick.ratio..times.                    -0.501 0.616222
Current.ratio..times.                  0.157 0.875266
Debt.to.equity.ratio..times.           1.622 0.104762
Creditors.turnover                     -0.035 0.971813
Debtors.turnover                       -0.312 0.755121
Finished.goods.turnover                0.328 0.742631
WIP.turnover                           -1.177 0.239168
Raw.material.turnover                  -0.531 0.595563
Adjusted.EPS                            -0.814 0.415652
PE.on.BSE                             -1.514 0.130032
PriceperShare                          0.443 0.657406
NetWC2TotAsset                         -2.443 0.014582 *
NetWorth2TotAssets                     -4.235 2.28e-05 ***
TotalIncome2TotalAssets                 0.739 0.460147
Total.expenses2TotalAssets              -0.894 0.371202
PAT2TotalAssets                         1.552 0.120699
Sales2TotalAssets                       0.422 0.673314
Currentliabilities2Assets              -0.155 0.876778
CapitalEmployed2TotalAssets             0.201 0.840415
NetfixedAssets2TotalAssets              2.246 0.024705 *
Investments2TotalAssets                2.217 0.026634 *
Cash.to.average.cost.of.sales.per.day 2.676 0.007451 **
Cash.profit.as...of.total.income      -0.535 0.592776
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1723.67 on 3540 degrees of freedom  
 Residual deviance: 836.27 on 3497 degrees of freedom  
 AIC: 924.27

Number of Fisher Scoring iterations: 11

#### d. Analyzing the coefficients:

After removing variables with High VIF values/ variables which are turned out as not important from bivariate analysis in iterative approach, below are the finalized variables.  
 When we built a logistic regression using these variables, we got coefficient for these variables

| Coefficients: | Estimate   | Std. Error | z value | Pr(> z )   |
|---------------|------------|------------|---------|------------|
| (Intercept)   | -1.6669884 | 0.5892984  | -2.829  | 0.004673** |

|   |            |           |        |              |
|---|------------|-----------|--------|--------------|
| Net.worth                                   | -0.0005171 | 0.0006214 | -0.832 | 0.405293     |
| Change.in.stock                             | 0.0065133  | 0.004034  | 1.615  | 0.106398     |
| Profit.after.tax                            | 0.0024627  | 0.0035696 | 0.69   | 0.490263     |
| PBDITA                                      | 0.0003144  | 0.0016528 | 0.19   | 0.849148     |
| Cash.profit                                 | -0.0114943 | 0.0055292 | -2.079 | 0.037634*    |
| PBDITA.as...of.total.income                 | -0.0229665 | 0.0164879 | -1.393 | 0.163642     |
| PAT.as...of.total.income                    | -0.0288296 | 0.0175473 | -1.643 | 0.10039      |
| PAT.as...of.net.worth                       | -0.031991  | 0.0072542 | -4.41  | 0.0000103*** |
| Sales                                       | -0.0001742 | 0.0001833 | -0.95  | 0.341879     |
| Other.income                                | 0.0237627  | 0.0145708 | 1.631  | 0.102925     |
| Reserves.and.funds                          | -0.0001295 | 0.0007341 | -0.176 | 0.85999      |
| Current.liabilities...provisions            | 0.0005216  | 0.0005705 | 0.914  | 0.360584     |
| Deferred.tax.liability                      | -0.0013699 | 0.0031018 | -0.442 | 0.658743     |
| Cumulative.retained.profits                 | -0.0095506 | 0.0027985 | -3.413 | 0.000643***  |
| TOL.TNW                                     | 0.0483636  | 0.0594201 | 0.814  | 0.415687     |
| Total.term.liabilities...tangible.net.worth | 0.0130512  | 0.1277252 | 0.102  | 0.918612     |
| Contingent.liabilities                      | -0.0004079 | 0.0006865 | -0.594 | 0.552435     |
| Net.fixed.assets                            | -0.0003327 | 0.000451  | -0.738 | 0.460793     |
| Investments                                 | 0.0004254  | 0.001486  | 0.286  | 0.774656     |
| Net.working.capital                         | 0.0009522  | 0.0012941 | 0.736  | 0.461862     |
| Quick.ratio..times.                         | -0.1508673 | 0.3010046 | -0.501 | 0.616222     |
| Current.ratio..times.                       | 0.0358712  | 0.2285181 | 0.157  | 0.875266     |
| Debt.to.equity.ratio..times.                | 0.1645935  | 0.1014634 | 1.622  | 0.104762     |
| Creditors.turnover                          | -0.0004675 | 0.0132312 | -0.035 | 0.971813     |
| Debtors.turnover                            | -0.0032909 | 0.0105514 | -0.312 | 0.755121     |
| Finished.goods.turnover                     | 0.0011516  | 0.0035069 | 0.328  | 0.742631     |
| WIP.turnover                                | -0.0105576 | 0.0089694 | -1.177 | 0.239168     |
| Raw.material.turnover                       | -0.0075354 | 0.0141966 | -0.531 | 0.595563     |
| Adjusted.EPS                                | -0.0350105 | 0.0430111 | -0.814 | 0.415652     |
| PE.on.BSE                                   | -0.0147043 | 0.0097124 | -1.514 | 0.130032     |
| PriceperShare                               | 0.0019179  | 0.0043245 | 0.443  | 0.657406     |
| NetWC2TotAsset                              | -1.1929566 | 0.4883979 | -2.443 | 0.014582*    |
| NetWorth2TotAssets                          | -4.1929003 | 0.9899898 | -4.235 | 0.0000228*** |
| TotalIncome2TotalAssets                     | 0.0174048  | 0.0235644 | 0.739  | 0.460147     |
| Total.expenses2TotalAssets                  | -0.0214709 | 0.0240107 | -0.894 | 0.371202     |
| PAT2TotalAssets                             | 0.8399423  | 0.5412538 | 1.552  | 0.120699     |
| Sales2TotalAssets                           | 0.0083625  | 0.019835  | 0.422  | 0.673314     |
| Currentliabilities2Assets                   | -0.0139434 | 0.0899258 | -0.155 | 0.876778     |
| CapitalEmployed2TotalAssets                 | 0.1568767  | 0.7790753 | 0.201  | 0.840415     |
| NetfixedAssets2TotalAssets                  | 0.1294064  | 0.0576166 | 2.246  | 0.024705*    |
| Investments2TotalAssets                     | 1.0914202  | 0.4923305 | 2.217  | 0.026634*    |
| Cash.to.average.cost.of.sales.per.day       | 0.0059524  | 0.0022244 | 2.676  | 0.007451**   |
| Cash.profit.as...of.total.income            | -0.0128088 | 0.0239498 | -0.535 | 0.592776     |

From the above table, we see that few coefficients are highlighted in green color and they are negative. It means that if there **one-unit decline** in the respective variable, it is going to cause a negative effect on the net worth next year value and hence likely to be turned out as credit card defaulter. The estimate given against each variable is the rate at which it affects our target variable in a negative way.

**Example:**

If there is one-unit decline in Net.worth then networth next year would decrease by -0.0005171 times keeping the other variables constant

Also, few coefficients which are not highlighted in the above table are having positive signs. This means that one-unit increase in the respective variable, it is going to cause a positive affect on the net worth next year value hence the customer is NOT likely turned out to be credit card defaulter. The estimate given against each variable is the rate at which affects our target variable in a positive way.

**Example:**

If there is one-unit increase in Investments2TotalAssets 0.0024627, then networth next year would increase by 1.0914202 times keeping the other variables constant.

**Top factors showing positive influence to be a non-defaulter:**

Top variables which are showing positive influence on networth next year are below. If these are good, then customer is NOT likely to be subjected to as defaulter

| Coefficients:                | Estimate  |
|------------------------------|-----------|
| Investments2TotalAssets      | 1.0914202 |
| PAT2TotalAssets              | 0.8399423 |
| Debt.to.equity.ratio..times. | 0.1645935 |
| CapitalEmployed2TotalAssets  | 0.1568767 |
| NetfixedAssets2TotalAssets   | 0.1294064 |

**Top factors showing positive influence to be a Defaulter:**

Top variables which are showing negative influence on networth next year are below. If these are good, then customer is likely to be subjected to as defaulter

| Coefficients:       | Estimate   |
|---------------------|------------|
| Adjusted.EPS        | -0.0350105 |
| Quick.ratio..times. | -0.1508673 |
| NetWC2TotAsset      | -1.1929566 |
| (Intercept)         | -1.6669884 |
| NetWorth2TotAssets  | -4.1929003 |

## **13. Model Performance Measures:**

### **Prediction on Train Dataset:**

The model which was built is used to make predictions on Train Dataset.

```
#Making Predictions on Train Dataset
logPred <- predict.glm(LogModel, finalTrain, type="response")
values1= ifelse(logPred >0.5,1,0)
values1=as.factor(values1)
confusionMatrix(finalTrain$Default,values1)
roc.logit<-roc(finalTrain$Default,logPred ,print.auc=TRUE,
show.thres=TRUE,plot=TRUE,main="ROC Curve for Train
Dataset",col="Red")
```

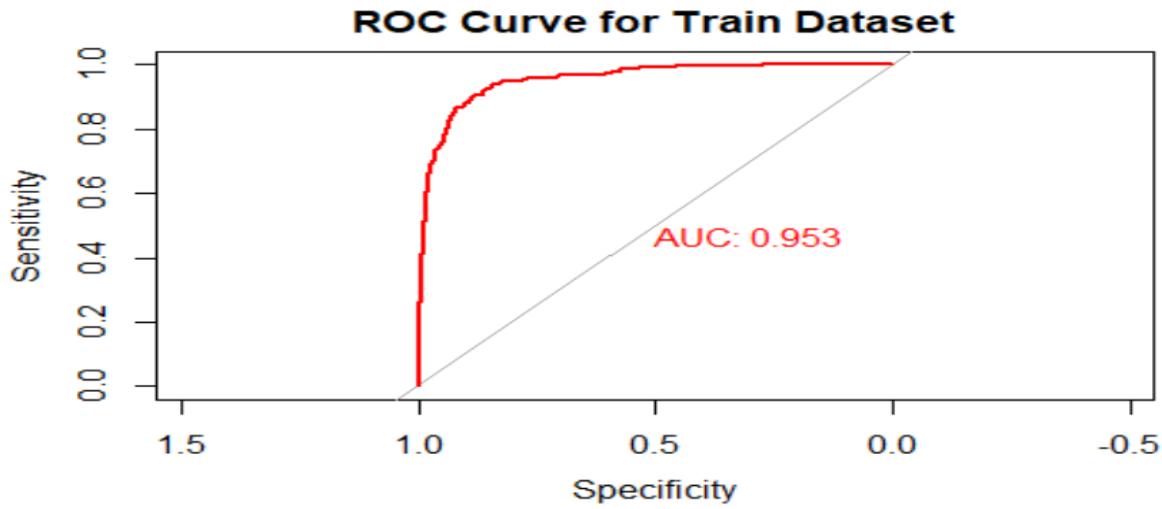
Confusion Matrix and Statistics

|            |   | Reference |     |
|------------|---|-----------|-----|
|            |   | 0         | 1   |
| Prediction | 0 | 3258      | 49  |
|            | 1 | 114       | 120 |

Accuracy : 0.954  
95% CI : (0.9465, 0.9606)  
No Information Rate : 0.9523  
P-Value [Acc > NIR] : 0.3358  
  
Kappa : 0.5718  
  
McNemar's Test P-Value : 5.362e-07  
  
Sensitivity : 0.9662  
Specificity : 0.7101  
Pos Pred Value : 0.9852  
Neg Pred Value : 0.5128  
Prevalence : 0.9523  
Detection Rate : 0.9201  
Detection Prevalence : 0.9339  
Balanced Accuracy : 0.8381  
  
'Positive' Class : 0

Here our metric measure should be specificity as we have to predict the Negative class (1 is negative class). We got 71% sensitivity.

We got 95.3% as Area under curve which is good. This is a striking balance sensitivity and specificity.



#### Prediction on Test Dataset:

```
```{r}
#Making Predictions on Validation Dataset
logPred <- predict.glm(LogModel, Validation, type="response")
values1= ifelse(logPred >0.5,1,0)
values1=as.factor(values1)
confusionMatrix(Validation$Default,values1)
roc.logit<-roc(Validation$Default,logPred ,print.auc=TRUE,
show.thres=TRUE,plot=TRUE,main="ROC Curve for Test Dataset",col="Red")
```

```

Confusion Matrix and Statistics

|            |   | Reference |    |
|------------|---|-----------|----|
|            |   | 0         | 1  |
| Prediction | 0 | 642       | 19 |
|            | 1 | 17        | 37 |

Accuracy : 0.9497  
95% CI : (0.931, 0.9645)  
No Information Rate : 0.9217  
P-Value [Acc > NIR] : 0.002092

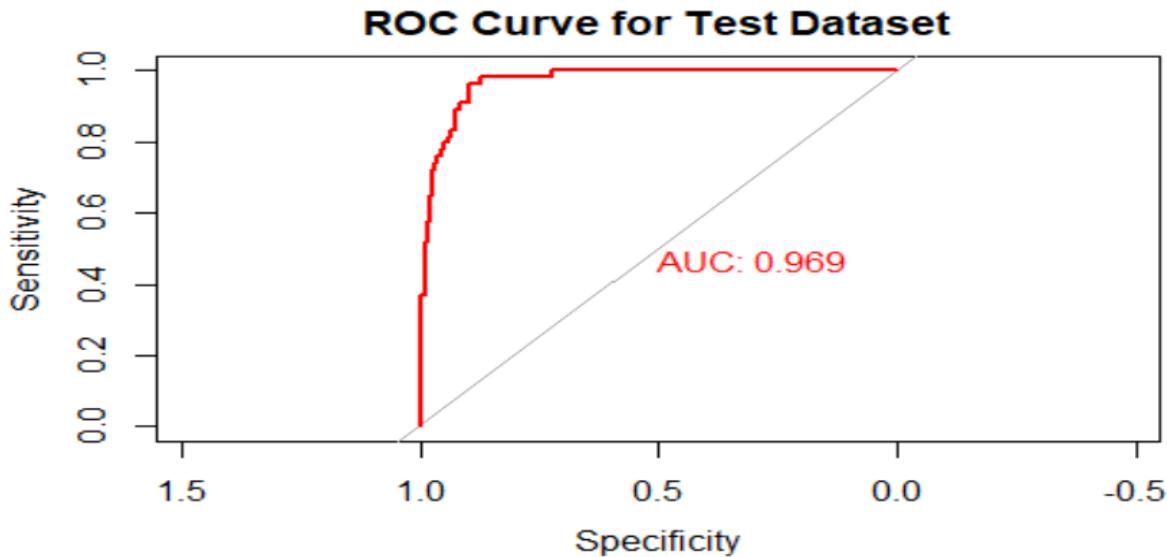
Kappa : 0.6455

Mcnemar's Test P-Value : 0.867632

Sensitivity : 0.9742  
Specificity : 0.6607  
Pos Pred Value : 0.9713  
Neg Pred Value : 0.6852  
Prevalence : 0.9217  
Detection Rate : 0.8979  
Detection Prevalence : 0.9245  
Balanced Accuracy : 0.8175

'Positive' Class : 0

Specificity is 66% on Testing Dataset. Area under curve is 96.9% which is better test dataset.



#### **14. Deciles:**

```
#creating deciles
decile <- function(x)
{
  #vector of 10 buckets(deciles)
  deciles <- vector(length=10)

  #10 deciles with from 0.1 to 1 with 0.1 interval. i.e 0.1, 0.2,0.3 etc
  for (i in seq(0.1,1,.1))
  {
    #converting deciles to %
    deciles[i*10] <- quantile(x, i)
  }
  return (
  #whch decile the probability falls into
  ifelse(x<deciles[1], 1,
         ifelse(x<deciles[2], 2,
                ifelse(x<deciles[3], 3,
                       ifelse(x<deciles[4], 4,
                              ifelse(x<deciles[5], 5,
                                     ifelse(x<deciles[6], 6,
                                         ifelse(x<deciles[7], 7,
                                               ifelse(x<deciles[8], 8,
                                                     ifelse(x<deciles[9], 9, 10
                                                       )))))))))
}
````
```

Dividing the predictions made into 10 deciles

```
```{r}
#adding deciles to dataframe and bucketing the probabilities into deciles|
finalTrain$deciles <- decile(finalTrain$pred)
Validation$deciles <- decile(Validation$pred)
````
```

Sorting deciles into descending order and counting the number of defaulters and non-defaulters

```
```{r}
#sorting in descending order based on probability and counting in each decile
#the number of defaulters and non defaulters|
rankTrain <- tmp_Train[, list(cnt=length(Default),
                                cnt_resp=sum(Default==1),
                                cnt_non_resp=sum(Default==0)),
                                by=deciles][order(-deciles)]
...```

```

### Train Deciles:

**Deciles:** the customers are divided into 10 buckets called deciles and sorted in descending order

**Cnt:** this is the count of customers falling into each decile based on their predicted probabilities

**Cnt\_resp:** this is the number of customers who will be defaulting in the respective decile

**Cnt\_non\_resp:** this is the number of customers who will NOT be defaulting in the respective decile

**Rate:**% of defaulters in each decile

**Cum\_resp:** cumulative count of number of defaulters in each decile

**Cum\_non\_resp:** cumulative count of number of NON-defaulters in each decile

**Cum\_rel\_resp:** This is the ratio of cum\_res/sum(cum\_resp). this gives the total % number of customers who will be defaulting overall

**Cum\_non\_rel\_resp:** This is the ratio of cum\_non\_res/sum(cum\_non\_resp). this gives the total % number of customers who will NOT defaulting overall

**Ks:** this is the % difference of cum\_non\_resp- cum\_rel\_non\_resp

### Train Decile:

	deciles	cnt	cum_resp	cum_non_resp	rrate	cum_resp	cum_non_resp	cum_rel_resp	cum_rel_non_resp	ks
1	10	355	179	176	50.42%	179	176	76.50%	5.3%	71.18
2	9	354	37	317	10.45%	216	493	92.31%	14.9%	77.40
3	8	353	8	345	2.27%	224	838	95.73%	25.3%	70.39
4	7	355	2	353	0.56%	226	1191	96.58%	36.0%	60.57
5	6	354	5	349	1.41%	231	1540	98.72%	46.6%	52.15
6	5	354	2	352	0.56%	233	1892	99.57%	57.2%	42.36
7	4	353	0	353	0.00%	233	2245	99.57%	67.9%	31.68
8	3	355	1	354	0.28%	234	2599	100.00%	78.6%	21.41
9	2	354	0	354	0.00%	234	2953	100.00%	89.3%	10.70
10	1	354	0	354	0.00%	234	3307	100.00%	100.0%	0.00

## Test Deciles:

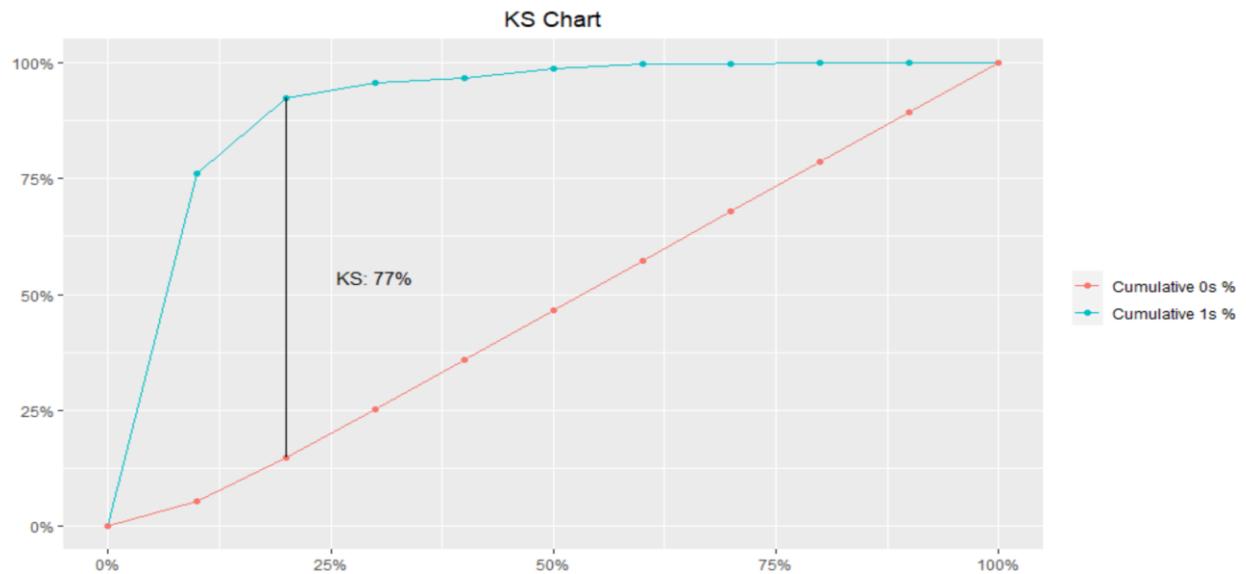
	deciles	cnt	cnt_resp	cnt_non_resp	rrate	cum_resp	cum_non_resp	cum_rel_resp	cum_rel_non_resp	ks
1	10	72	42	30	58.3%	42	30	77.8%	4.5%	73.24
2	9	71	11	60	15.5%	53	90	98.2%	13.6%	84.53
3	8	72	0	72	0.0%	53	162	98.2%	24.5%	73.64
4	7	71	1	70	1.4%	54	232	100.0%	35.1%	64.90
5	6	72	0	72	0.0%	54	304	100.0%	46.0%	54.01
6	5	71	0	71	0.0%	54	375	100.0%	56.7%	43.27
7	4	71	0	71	0.0%	54	446	100.0%	67.5%	32.53
8	3	72	0	72	0.0%	54	518	100.0%	78.4%	21.63
9	2	71	0	71	0.0%	54	589	100.0%	89.1%	10.89
10	1	72	0	72	0.0%	54	661	100.0%	100.0%	0.00

## 15. Observation and inferences:

From deciling we observe that we can target customer who are falling in 1-9 deciles. Customers who are falling in 10th decile are likely to be turned into a defaulter causing huge cost to the bank. Issuing credit cards to the customers who are falling in this decile would be leading to default credit risk.

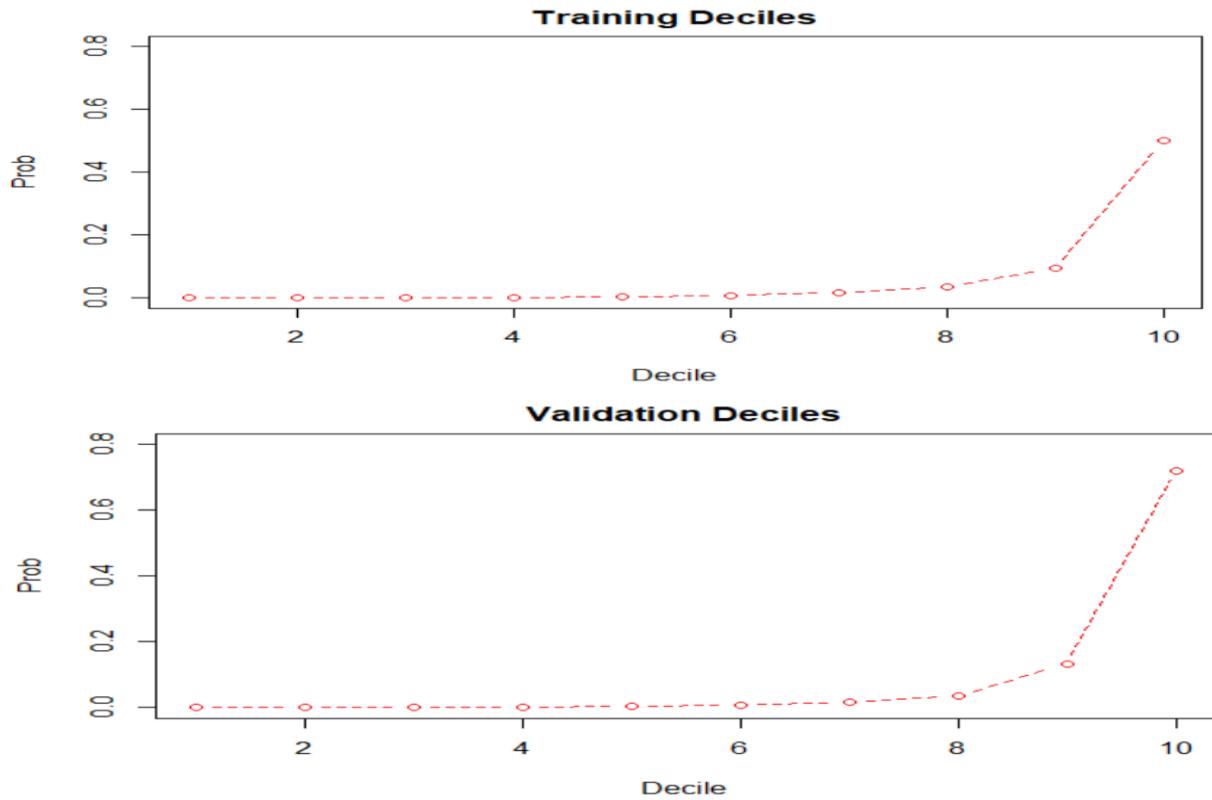
## 16. Model Performance:

To evaluate the mode performance and inferences we made, we can plot a KS chart to check the at which decile the maximum difference is coming out between cumulative default and non-default customers.



At the 9th decile we see KS state value is 77 % which is good, targeted customers are falling in 1-9 decile

**Plots for deciles:**



Source for decile: 4.3 - Model Performance Hands-on Part 1-2

\*\*\*\*\*THE END\*\*\*\*\*