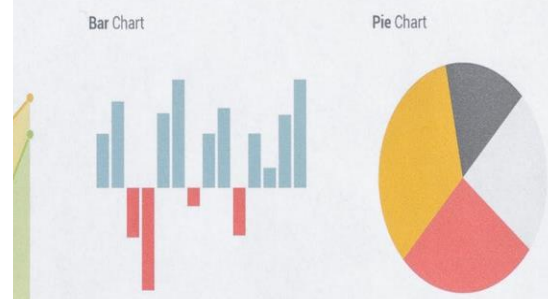
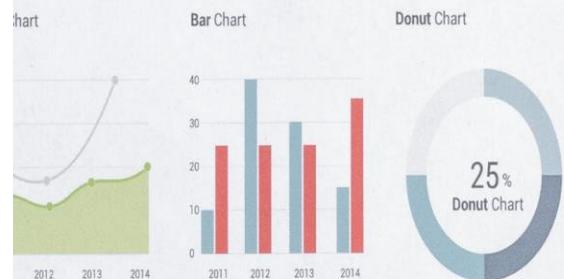


Credit Risk Analysis Report 2020

Feb 15, 2020

Credit Risk Analysis Report
Authored By: Deepti Lobo





CONTENTS

Project Objective	3
Known Facts	3
Exploratory Data Analysis (EDA).....	3
Missing Value Treatment	4
Outlier Treatment	6
Descriptive Analysis	9
Data Visualization	23
Variable Creation.....	26
Model Building	28
Confusion Matrix	35
Decile.....	36
Conclusion.....	37
Recommendation	37

Project Objective

The objective of the project is to create India credit risk(default) model using the given training dataset and validate it on the testing dataset. Logistic Regression framework is to be used, to develop the credit default model.

Known Facts

The training dataset has data on 3541 records and test dataset has 715 records. The data has information about the assets, income, expense, Profit before and after tax, sales, shares etc. There is a total of 51 variables provided.

Exploratory Data Analysis (EDA)

The given training dataset consists of 3541 observations and 52 variables. The testing dataset has 715 observations and 52 variables.

#Names of the columns

```
names(trainDS)
```

```
## [1] "Num" "Networth Next Year"
## [3] "Total assets" "Net worth"
## [5] "Total income" "Change in stock"
## [7] "Total expenses" "Profit after tax"
## [9] "PBDITA" "PBT"
## [11] "Cash profit" "PBDITA as % of total income"
## [13] "PBT as % of total income" "PAT as % of total income"
## [15] "Cash profit as % of total income" "PAT as % of net worth"
## [17] "Sales" "Income from financial services"
## [19] "Other income" "Total capital"
## [21] "Reserves and funds"
## [22] "Deposits (accepted by commercial banks)"
## [23] "Borrowings"
## [24] "Current liabilities & provisions" "Deferred tax liability"
## [26] "Shareholders funds" "Cumulative retained profits"
## [28] "Capital employed" "TOL/TNW"
## [30] "Total term liabilities / tangible net worth"
## [31] "Contingent liabilities / Net worth (%)"
## [32] "Contingent liabilities" "Net fixed assets"
## [34] "Investments" "Current assets"
## [36] "Net working capital" "Quick ratio (times)"
## [38] "Current ratio (times)" "Debt to equity ratio (times)"
## [40] "Cash to current liabilities (times)"
## [41] "Cash to average cost of sales per day"
## [42] "Creditors turnover" "Debtors turnover"
```

```
## [44] "Finished goods turnover"      "WIP turnover"
## [46] "Raw material turnover"        "Shares outstanding"
## [48] "Equity face value"           "EPS"
## [50] "Adjusted EPS"                 "Total liabilities"
## [52] "PE on BSE"
```

Attributes Details

All the variables are numeric in nature and are of Interval type.

```
#Display the first six rows
head(trainDS)
```

```
##      Num `Networth Next ~ `Total assets` `Net worth` `Total income`
## 1      1      8891.      17512.      7093.      24965.
## 2      2      394.      941      352.      1527.
## 3      3      92.2      233.      101.      477.
## 4      4      2.7      2.7      2.7      NA
## 5      5      109      478.      108.      1580.
## 6      6      689.      2434.      676.      2649.
```

As there is no default variable, so the default variable is created by splitting the “Networth Next Year” variable. The negative observations will be 1 and positive observations will be 0.

```
#Adding the default variable
trainDS$Default = ifelse(trainDS$`Networth Next Year` < 0, 1,0)
```

Missing Value Treatment

```
#Is there any values missing?
anyNA(trainDS)
```

```
## [1] TRUE
```

```
colSums(is.na(trainDS))
```

```
## Num      Total assets
## 0      0
## Net worth  Total income
## 0      198
## Change in stock  Total expenses
## 458      139
## Profit after tax  PBDITA
## 131      131
## PBT      Cash profit
## 131      131
## PBDITA as % of total income  PBT as % of total income
```

## 68	68
## PAT as % of total income	Cash profit as % of total income
## 68	68
## PAT as % of net worth	Sales
## 0	259
## Income from financial services	Other income
## 935	1295
## Total capital	Reserves and funds
## 4	85
## Deposits (accepted by commercial banks)	
## 3541	
## Borrowings	Current liabilities & provisions
## 366	96
## Deferred tax liability	Shareholders funds
## 1140	0
## Cumulative retained profits	Capital employed
## 38	0
## TOL/TNW	
## 0	
## Total term liabilities / tangible net worth	
## 0	
## Contingent liabilities / Net worth (%)	
## 0	
## Contingent liabilities	Net fixed assets
## 1188	118
## Investments	Current assets
## 1435	66
## Net working capital	Quick ratio (times)
## 32	93
## Current ratio (times)	Debt to equity ratio (times)
## 93	0
## Cash to current liabilities (times)	
## 93	
## Cash to average cost of sales per day	
## 85	
## Creditors turnover	Debtors turnover
## 47	42
## Finished goods turnover	WIP turnover
## 454	354
## Raw material turnover	Shares outstanding
## 75	0
## Equity face value	EPS
## 0	0
## Adjusted EPS	Total liabilities
## 0	0
## PE on BSE	Default
## 23	0

Since all the values are missing in “Deposits (accepted by commercial banks)”, removing this column from the data set.

#Replacing NA's with median for that columns

```
trainDS<-as.data.frame(trainDS)
```

```
for (i in 1:ncol(trainDS)) {  
  trainDS[,i] = as.numeric(trainDS[,i])  
  trainDS[is.na(trainDS[,i]),i] = median(trainDS[,i], na.rm = TRUE)  
}
```

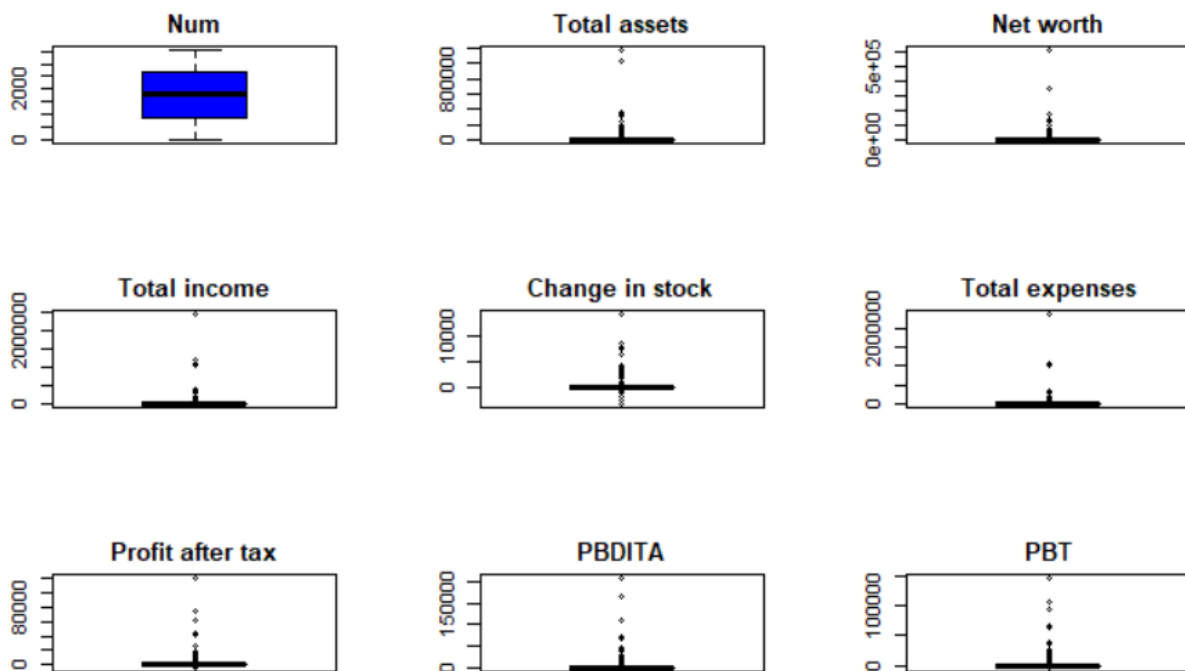
The missing values are treated by replacing the NA's with the median value.

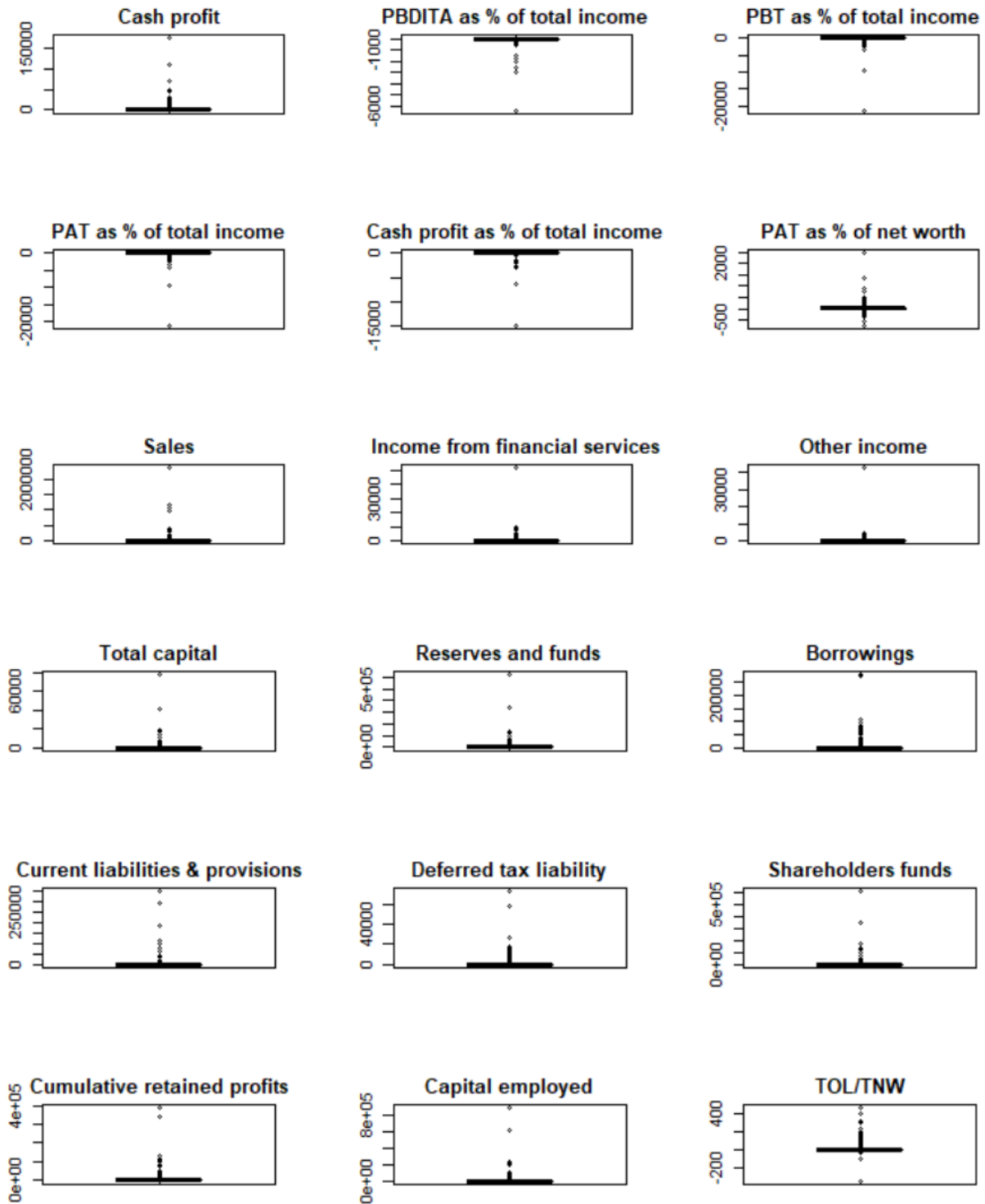
Outlier Treatment

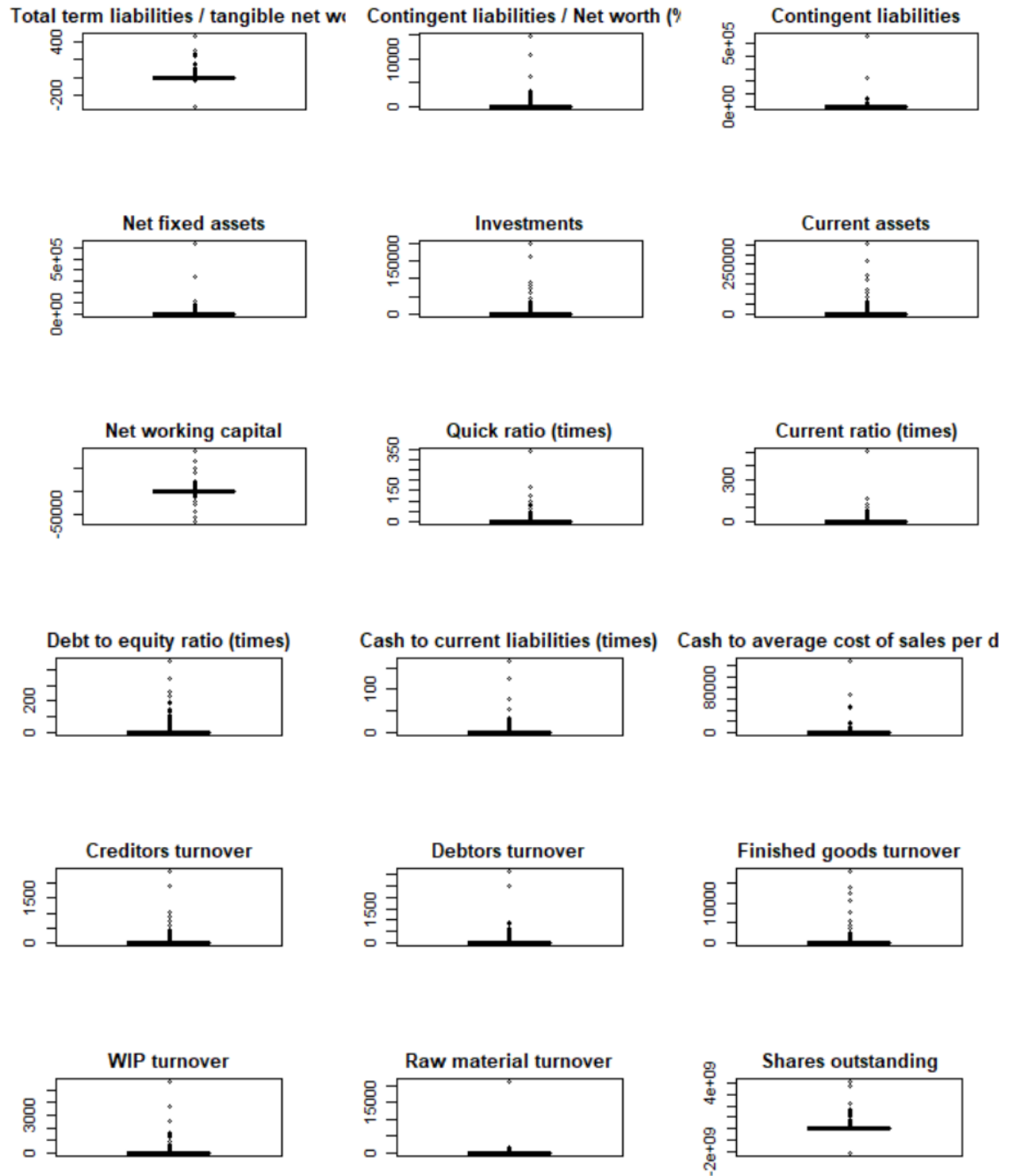
#Boxplot:

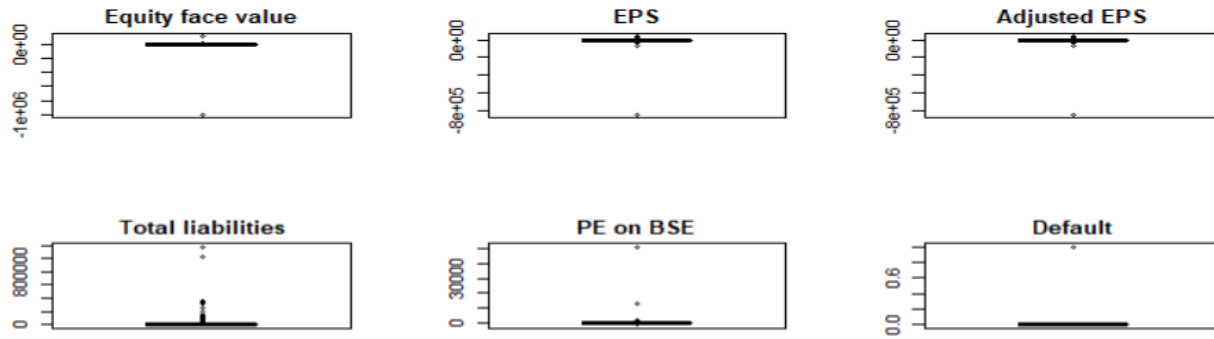
```
par(mfrow = c(3,3)) #reset plotting space
```

```
for(i in 1:length(trainDS)) {  
  boxplot(trainDS[,i], main = names(trainDS[i]), col = "blue", type = "l")  
}
```









All the continuous variables have outliers. Flooring and capping have been applied on all the continuous variables. We are using 1 and 99 percentiles for outlier treatment.

Descriptive Analysis

#Data types of all the columns

`str(trainDS)`

```
## 'data.frame':  3541 obs. of  51 variables:
## $ Num : num  1 2 3 4 5 6 7 8 9 10
## ...
## $ Total assets : num  17512.3 941 232.8 25.
5 478.5 ...
## $ Net worth : num  7093.2 351.5 100.6 8.
4 107.6 ...
## $ Total income : num  24965 1527 477 445 15
80 ...
## $ Change in stock : num  235.8 42.7 -5.2 1.6 -
12.9 ...
## $ Total expenses : num  23658 1455 479 408 15
58 ...
## $ Profit after tax : num  1543.2 115.2 -2.7 8.8
5.5 ...
## $ PBDITA : num  2860.2 283 5.8 35.4 3
1 ...
## $ PBT : num  2417.2 188.4 -2.7 12.
4 6.3 ...
## $ Cash profit : num  1872.8 158.6 0.3 18.9
11.9 ...
## $ PBDITA as % of total income : num  11.46 18.53 1.22 0.38
1.96 ...
## $ PBT as % of total income : num  9.68 12.33 -1.38 0 0.
4 ...
## $ PAT as % of total income : num  6.18 7.54 -1.38 0 0.3
5 2.81 0 0.72 8.29 -2.88 ...
## $ Cash profit as % of total income : num  7.5 10.38 0.06 0 0.75
...
## ...
```

```

## $ PAT as % of net worth : num 23.78 38.08 -4.94 0 5
.25 ...
## $ Sales : num 24458 1504 476 453 15
75 ...
## $ Income from financial services : num 158 4 1.5 1.8 3.9 6.4
1.8 1.8 7.3 1.8 ...
## $ Other income : num 286.2 15.9 0.2 1.4 0.
9 ...
## $ Total capital : num 423.8 115.5 81.4 3.6
6.2 ...
## $ Reserves and funds : num 6822.8 257.8 19.2 2.2
161.8 ...
## $ Borrowings : num 14.9 272.5 35.4 99.2
193.1 ...
## $ Current liabilities & provisions : num 9965.9 210 96.8 69.4
112.8 ...
## $ Deferred tax liability : num 284.9 85.2 13.4 13.4
4.6 ...
## $ Shareholders funds : num 7093.2 351.5 100.6 8.
6 107.6 ...
## $ Cumulative retained profits : num 6263.3 247.4 32.4 2.2
82.7 ...
## $ Capital employed : num 7108.1 624 136 17.2 3
00.7 ...
## $ TOL/TNW : num 1.33 1.23 1.44 0.15 2
.83 1.8 0.15 5.17 1.05 3.25 ...
## $ Total term liabilities / tangible net worth: num 0 0.34 0.29 0 1.59 0.
37 0.03 0.94 0.3 0.54 ...
## $ Contingent liabilities / Net worth (%) : num 14.8 19.2 45.8 0 34.9
...
## $ Contingent liabilities : num 1049.7 67.6 46.1 38 3
7.6 ...
## $ Net fixed assets : num 1900.2 286.4 38.7 7.6
94.8 ...
## $ Investments : num 1069.6 2.2 4.3 8.35 7
.4 ...
## $ Current assets : num 13277.5 563.9 167.5 6
.6 349.7 ...
## $ Net working capital : num 3588.5 203.5 59.6 0.2
215.8 ...
## $ Quick ratio (times) : num 1.18 0.95 1.11 0.67 1
.41 0.48 0.67 0.54 0.59 0.39 ...
## $ Current ratio (times) : num 1.37 1.56 1.55 1.23 2
.54 1.27 1.23 1.15 1.58 0.62 ...
## $ Debt to equity ratio (times) : num 0 0.78 0.35 0 1.79 1.
09 0.32 2.31 0.94 3.13 ...
## $ Cash to current liabilities (times) : num 0.43 0.06 0.21 0.07 0
.01 0.11 0.07 0.04 0.19 0.01 ...
## $ Cash to average cost of sales per day : num 68.21 5.96 17.07 8.02

```

```

0.85 ...
## $ Creditors turnover          : num  3.62 9.8 5.28 1.74 13
...
## $ Debtors turnover           : num  3.85 5.7 5.07 1.69 9.
46 ...
## $ Finished goods turnover    : num  200.55 14.21 9.24 17.
27 12.68 ...
## $ WIP turnover               : num  21.78 7.49 3 9.76 7.9
...
## $ Raw material turnover      : num  7.71 11.46 6.4 0 17.0
3 ...
## $ Shares outstanding         : num  42381675 11550000 814
9090 350000 619635 ...
## $ Equity face value         : num  10 10 10 10 10 10 10
10 10 10 ...
## $ EPS                       : num  35.52 9.97 -0.5 0 7.9
1 ...
## $ Adjusted EPS              : num  7.1 9.97 -0.5 0 7.91
...
## $ Total liabilities         : num  17512.3 941 232.8 25.
5 478.5 ...
## $ PE on BSE                 : num  27.31 8.17 3.7 9.1 9.
1 ...
## $ Default                   : num  0 0 0 0 0 0 0 0 0 1 #

```

summary of the dataset
summary(trainDS)

```

##      Num      Total assets      Net worth      Total income
## Min.   : 1      Min.   : 25.5      Min.   : 8.4      Min.   : 21.4
## 1st Qu.: 886    1st Qu.: 91.3      1st Qu.: 31.3     1st Qu.: 121.2
## Median :1773    Median : 309.7      Median : 102.3     Median : 444.9
## Mean   :1772    Mean   : 2052.4      Mean   : 754.5     Mean   : 2042.0
## 3rd Qu.:2658    3rd Qu.: 1098.7     3rd Qu.: 377.3     3rd Qu.: 1340.3
## Max.   :3545    Max.   :51658.8      Max.   :20920.8     Max.   :42282.8
## Change in stock Total expenses      Profit after tax      PBDITA
## Min.   :-12.9    Min.   : 16.6      Min.   : -2.7      Min.   : 0.6
## 1st Qu.: -0.7    1st Qu.: 104.1     1st Qu.: 0.6       1st Qu.: 7.3
## Median : 1.6     Median : 407.7     Median : 8.8       Median : 35.4
## Mean   : 25.3     Mean   : 1918.8     Mean   : 130.3     Mean   : 285.9
## 3rd Qu.: 13.4     3rd Qu.: 1284.6     3rd Qu.: 48.1      3rd Qu.: 139.1
## Max.   :600.2     Max.   :38283.8     Max.   :4069.9     Max.   :7370.1
##      PBT      Cash profit      PBDITA as % of total income
## Min.   : -2.7    Min.   : -0.10     Min.   : 0.38
## 1st Qu.: 0.9     1st Qu.: 3.10     1st Qu.: 5.07
## Median : 12.4     Median : 18.85     Median : 9.66
## Mean   : 176.4     Mean   : 189.22     Mean   :12.63
## 3rd Qu.: 67.5     3rd Qu.: 86.80     3rd Qu.:16.15
## Max.   :5421.4     Max.   :5453.78     Max.   :80.81
## PBT as % of total income PAT as % of total income

```

```

## Min.      :-3.44          Min.      :-3.180
## 1st Qu.: 0.60           1st Qu.: 0.390
## Median : 3.31           Median : 2.340
## Mean    : 5.97           Mean    : 4.441
## 3rd Qu.: 8.63           3rd Qu.: 6.250
## Max.     :51.66          Max.     :43.550
## Cash profit as % of total income PAT as % of net worth Sales
## Min.      : 0.000          Min.      :-4.94      Min.      : 25.6
## 1st Qu.: 2.090           1st Qu.: 0.00      1st Qu.: 133.3
## Median : 5.640           Median : 7.92      Median : 453.1
## Mean    : 7.972           Mean    :13.32      Mean    : 1989.1
## 3rd Qu.:10.560           3rd Qu.:20.19      3rd Qu.: 1314.7
## Max.     :56.244          Max.     :97.36      Max.     :40605.1
## Income from financial services Other income Total capital
## Min.      : 0.2           Min.      : 0.200    Min.      : 3.6
## 1st Qu.: 0.7             1st Qu.: 0.800    1st Qu.: 13.1
## Median : 1.8             Median : 1.400    Median : 42.1
## Mean    : 23.4           Mean    : 9.497    Mean    : 150.0
## 3rd Qu.: 5.4            3rd Qu.: 2.500    3rd Qu.: 100.3
## Max.     :769.9          Max.     :286.160   Max.     :2936.7
## Reserves and funds Borrowings Current liabilities & provisions
## Min.      : -10.7        Min.      : 5.2      Min.      : 3.9
## 1st Qu.: 5.8            1st Qu.: 29.7      1st Qu.: 18.7
## Median : 54.8           Median : 99.2      Median : 69.4
## Mean    : 614.2         Mean    : 601.0     Mean    : 473.4
## 3rd Qu.: 263.2         3rd Qu.: 296.0     3rd Qu.: 249.1
## Max.     :17416.7        Max.     :14803.4   Max.     :11092.0
## Deferred tax liability Shareholders funds Cumulative retained profits
## Min.      : 1.40         Min.      : 8.6      Min.      : -20.5
## 1st Qu.: 6.70           1st Qu.: 32.0      1st Qu.: 1.3
## Median : 13.40          Median : 105.6     Median : 37.1
## Mean    : 76.59         Mean    : 770.0     Mean    : 456.4
## 3rd Qu.: 26.90          3rd Qu.: 393.2     3rd Qu.: 199.4
## Max.     :2226.60        Max.     :20920.8   Max.     :13027.3
## Capital employed TOL/TNW
## Min.      : 17.2        Min.      : 0.150
## 1st Qu.: 60.8          1st Qu.: 0.600
## Median : 214.7         Median : 1.430
## Mean    : 1402.5        Mean    : 3.223
## 3rd Qu.: 767.3         3rd Qu.: 2.830
## Max.     :34914.6       Max.     :55.958
## Total term liabilities / tangible net worth
## Min.      : 0.000
## 1st Qu.: 0.050
## Median : 0.340
## Mean    : 1.243
## 3rd Qu.: 1.000
## Max.     :29.464
## Contingent liabilities / Net worth (%) Contingent liabilities

```

## Min. : 0.00		Min. : 2.4	
## 1st Qu.: 0.00		1st Qu.: 16.0	
## Median : 5.33		Median : 38.0	
## Mean : 36.06		Mean : 247.5	
## 3rd Qu.: 30.76		3rd Qu.: 84.2	
## Max. : 773.79		Max. : 6177.7	
## Net fixed assets	Investments	Current assets	Net working capital
## Min. : 7.6	Min. : 0.40	Min. : 6.6	Min. : -53.2
## 1st Qu.: 27.3	1st Qu.: 4.60	1st Qu.: 37.2	1st Qu.: -1.0
## Median : 93.5	Median : 8.35	Median : 145.1	Median : 16.2
## Mean : 648.9	Mean : 147.13	Mean : 779.1	Mean : 144.7
## 3rd Qu.: 328.8	3rd Qu.: 16.10	3rd Qu.: 485.9	3rd Qu.: 81.6
## Max. : 16862.2	Max. : 5024.04	Max. : 17377.1	Max. : 3688.2
## Quick ratio (times)	Current ratio (times)	Debt to equity ratio (times)	
## Min. : 0.190	Min. : 0.620	Min. : 0.000	
## 1st Qu.: 0.420	1st Qu.: 0.940	1st Qu.: 0.220	
## Median : 0.670	Median : 1.230	Median : 0.790	
## Mean : 1.064	Mean : 1.774	Mean : 1.954	
## 3rd Qu.: 1.020	3rd Qu.: 1.690	3rd Qu.: 1.750	
## Max. : 14.000	Max. : 18.920	Max. : 37.104	
## Cash to current liabilities (times)	Cash to average cost of sales per day		
## Min. : 0.0100	Min. : 0.850		
## 1st Qu.: 0.0200	1st Qu.: 2.890		
## Median : 0.0700	Median : 8.025		
## Mean : 0.2891	Mean : 47.483		
## 3rd Qu.: 0.1900	3rd Qu.: 21.150		
## Max. : 5.7780	Max. : 1277.500		
## Creditors turnover	Debtors turnover	Finished goods turnover	
## Min. : 1.740	Min. : 1.69	Min. : 5.04	
## 1st Qu.: 3.940	1st Qu.: 4.01	1st Qu.: 10.21	
## Median : 6.095	Median : 6.32	Median : 17.27	
## Mean : 11.736	Mean : 12.83	Mean : 46.58	
## 3rd Qu.: 10.550	3rd Qu.: 10.82	3rd Qu.: 30.72	
## Max. : 131.374	Max. : 182.90	Max. : 857.22	
## WIP turnover	Raw material turnover	Shares outstanding	
## Min. : 3.00	Min. : 0.00	Min. : 350000	
## 1st Qu.: 5.93	1st Qu.: 3.41	1st Qu.: 2209860	
## Median : 9.76	Median : 6.40	Median : 4672063	
## Mean : 18.76	Mean : 10.06	Mean : 14455333	
## 3rd Qu.: 16.94	3rd Qu.: 10.92	3rd Qu.: 8320000	
## Max. : 218.88	Max. : 97.82	Max. : 268509911	
## Equity face value	EPS	Adjusted EPS	Total liabilities
## Min. : 10.00	Min. : -0.69	Min. : -0.66	Min. : 25.5
## 1st Qu.: 10.00	1st Qu.: 0.00	1st Qu.: 0.00	1st Qu.: 91.3
## Median : 10.00	Median : 1.43	Median : 1.18	Median : 309.7
## Mean : 18.24	Mean : 27.08	Mean : 25.92	Mean : 2052.4
## 3rd Qu.: 10.00	3rd Qu.: 9.62	3rd Qu.: 7.48	3rd Qu.: 1098.7
## Max. : 190.00	Max. : 896.14	Max. : 896.14	Max. : 51658.8
## PE on BSE	Default		

```
## Min.   : 3.70   Min.   :0.00000
## 1st Qu.: 9.10   1st Qu.:0.00000
## Median : 9.10   Median :0.00000
## Mean    : 12.65  Mean    :0.06608
## 3rd Qu.: 9.10   3rd Qu.:0.00000
## Max.    :156.05  Max.    :1.00000
```

##Summary Statistics Measure of central tendency and dispersion (Univariate Analysis)

```
describe(trainDS,na.rm = TRUE,
          quant = c(0.01,0.05,0.10,0.25,0.75,0.90,0.95,0.99),IQR=TRUE,check=TRUE)
```

##	vars	n	mean
## Num	1	3541	1772.45
## Total assets	2	3541	2052.41
## Net worth	3	3541	754.55
## Total income	4	3541	2042.04
## Change in stock	5	3541	25.30
## Total expenses	6	3541	1918.78
## Profit after tax	7	3541	130.30
## PBDITA	8	3541	285.86
## PBT	9	3541	176.44
## Cash profit	10	3541	189.22
## PBDITA as % of total income	11	3541	12.63
## PBT as % of total income	12	3541	5.97
## PAT as % of total income	13	3541	4.44
## Cash profit as % of total income	14	3541	7.97
## PAT as % of net worth	15	3541	13.32
## Sales	16	3541	1989.11
## Income from financial services	17	3541	23.40
## Other income	18	3541	9.50
## Total capital	19	3541	149.99
## Reserves and funds	20	3541	614.18
## Borrowings	21	3541	600.99
## Current liabilities & provisions	22	3541	473.40
## Deferred tax liability	23	3541	76.59
## Shareholders funds	24	3541	770.05
## Cumulative retained profits	25	3541	456.43
## Capital employed	26	3541	1402.52
## TOL/TNW	27	3541	3.22
## Total term liabilities / tangible net worth	28	3541	1.24
## Contingent liabilities / Net worth (%)	29	3541	36.06
## Contingent liabilities	30	3541	247.46
## Net fixed assets	31	3541	648.93
## Investments	32	3541	147.13
## Current assets	33	3541	779.06
## Net working capital	34	3541	144.69
## Quick ratio (times)	35	3541	1.06
## Current ratio (times)	36	3541	1.77

## Debt to equity ratio (times)	37 3541	1.95
## Cash to current liabilities (times)	38 3541	0.29
## Cash to average cost of sales per day	39 3541	47.48
## Creditors turnover	40 3541	11.74
## Debtors turnover	41 3541	12.83
## Finished goods turnover	42 3541	46.58
## WIP turnover	43 3541	18.76
## Raw material turnover	44 3541	10.06
## Shares outstanding	45 3541 14455333	39
## Equity face value	46 3541	18.24
## EPS	47 3541	27.08
## Adjusted EPS	48 3541	25.92
## Total liabilities	49 3541	2052.41
## PE on BSE	50 3541	12.65
## Default	51 3541	0.07
##	sd	median
## Num	1023.73	1773.00
## Total assets	6562.22	309.70
## Net worth	2564.08	102.30
## Total income	5597.88	444.90
## Change in stock	84.49	1.60
## Total expenses	5169.49	407.70
## Profit after tax	489.75	8.80
## PBDITA	943.34	35.40
## PBT	656.35	12.40
## Cash profit	667.62	18.85
## PBDITA as % of total income	12.85	9.66
## PBT as % of total income	9.23	3.31
## PAT as % of total income	7.52	2.34
## Cash profit as % of total income	9.18	5.64
## PAT as % of net worth	18.61	7.92
## Sales	5402.84	453.10
## Income from financial services	95.51	1.80
## Other income	36.14	1.40
## Total capital	392.96	42.10
## Reserves and funds	2154.30	54.80
## Borrowings	1895.43	99.20
## Current liabilities & provisions	1465.90	69.40
## Deferred tax liability	277.97	13.40
## Shareholders funds	2574.87	105.60
## Cumulative retained profits	1592.61	37.10
## Capital employed	4457.30	214.70
## TOL/TNW	7.23	1.43
## Total term liabilities / tangible net worth	3.61	0.34
## Contingent liabilities / Net worth (%)	96.55	5.33
## Contingent liabilities	817.57	38.00
## Net fixed assets	2135.82	93.50
## Investments	638.37	8.35
## Current assets	2254.96	145.10

## Net working capital	484.83	16.20
## Quick ratio (times)	1.72	0.67
## Current ratio (times)	2.31	1.23
## Debt to equity ratio (times)	4.69	0.79
## Cash to current liabilities (times)	0.77	0.07
## Cash to average cost of sales per day	161.12	8.02
## Creditors turnover	18.92	6.09
## Debtors turnover	24.73	6.32
## Finished goods turnover	114.17	17.27
## WIP turnover	30.38	9.76
## Raw material turnover	13.99	6.40
## Shares outstanding	37109523.84	4672063.00
## Equity face value	28.93	10.00
## EPS	111.58	1.43
## Adjusted EPS	111.33	1.18
## Total liabilities	6562.22	309.70
## PE on BSE	18.65	9.10
## Default	0.25	0.00
##	trimmed	mad
## Num	1772.31	1313.58
## Total assets	643.26	402.08
## Net worth	220.77	134.03
## Total income	780.60	583.70
## Change in stock	6.34	7.71
## Total expenses	743.29	546.78
## Profit after tax	28.39	14.38
## PBDITA	82.56	50.11
## PBT	39.36	19.72
## Cash profit	49.72	27.80
## PBDITA as % of total income	10.48	7.89
## PBT as % of total income	4.44	4.91
## PAT as % of total income	3.19	3.47
## Cash profit as % of total income	6.30	5.97
## PAT as % of net worth	10.21	11.74
## Sales	769.84	579.84
## Income from financial services	3.89	1.93
## Other income	2.06	1.04
## Total capital	61.31	49.82
## Reserves and funds	154.54	87.77
## Borrowings	181.92	124.83
## Current liabilities & provisions	147.02	93.11
## Deferred tax liability	19.53	12.16
## Shareholders funds	228.96	138.77
## Cumulative retained profits	114.19	69.98
## Capital employed	440.77	278.28
## TOL/TNW	1.74	1.48
## Total term liabilities / tangible net worth	0.53	0.50
## Contingent liabilities / Net worth (%)	15.04	7.90
## Contingent liabilities	63.93	38.40

## Net fixed assets	191.28	120.83
## Investments	16.69	6.75
## Current assets	280.39	194.37
## Net working capital	41.92	45.07
## Quick ratio (times)	0.73	0.44
## Current ratio (times)	1.32	0.50
## Debt to equity ratio (times)	1.00	1.02
## Cash to current liabilities (times)	0.11	0.07
## Cash to average cost of sales per day	13.44	9.38
## Creditors turnover	7.46	4.19
## Debtors turnover	7.53	4.34
## Finished goods turnover	21.89	12.65
## WIP turnover	11.78	6.82
## Raw material turnover	7.30	5.23
## Shares outstanding	5895081.51	4305563.80
## Equity face value	10.00	0.00
## EPS	4.93	2.95
## Adjusted EPS	4.02	2.46
## Total liabilities	643.26	402.08
## PE on BSE	9.17	0.00
## Default	0.00	0.00
##	min	max
## Num	1.00	3545.00
## Total assets	25.50	51658.78
## Net worth	8.40	20920.76
## Total income	21.40	42282.76
## Change in stock	-12.90	600.20
## Total expenses	16.60	38283.80
## Profit after tax	-2.70	4069.88
## PBDITA	0.60	7370.08
## PBT	-2.70	5421.40
## Cash profit	-0.10	5453.78
## PBDITA as % of total income	0.38	80.81
## PBT as % of total income	-3.44	51.66
## PAT as % of total income	-3.18	43.55
## Cash profit as % of total income	0.00	56.24
## PAT as % of net worth	-4.94	97.36
## Sales	25.60	40605.08
## Income from financial services	0.20	769.94
## Other income	0.20	286.16
## Total capital	3.60	2936.70
## Reserves and funds	-10.70	17416.72
## Borrowings	5.20	14803.38
## Current liabilities & provisions	3.90	11092.00
## Deferred tax liability	1.40	2226.60
## Shareholders funds	8.60	20920.76
## Cumulative retained profits	-20.50	13027.32
## Capital employed	17.20	34914.60
## TOL/TNW	0.15	55.96

## Total term liabilities / tangible net worth	0.00	29.46
## Contingent liabilities / Net worth (%)	0.00	773.79
## Contingent liabilities	2.40	6177.70
## Net fixed assets	7.60	16862.16
## Investments	0.40	5024.04
## Current assets	6.60	17377.14
## Net working capital	-53.20	3688.16
## Quick ratio (times)	0.19	14.00
## Current ratio (times)	0.62	18.92
## Debt to equity ratio (times)	0.00	37.10
## Cash to current liabilities (times)	0.01	5.78
## Cash to average cost of sales per day	0.85	1277.50
## Creditors turnover	1.74	131.37
## Debtors turnover	1.69	182.90
## Finished goods turnover	5.04	857.22
## WIP turnover	3.00	218.88
## Raw material turnover	0.00	97.82
## Shares outstanding	350000.00	268509911.20
## Equity face value	10.00	190.00
## EPS	-0.69	896.14
## Adjusted EPS	-0.66	896.14
## Total liabilities	25.50	51658.78
## PE on BSE	3.70	156.05
## Default	0.00	1.00
##	range skew kurtosis	
## Num	3544.00	0.00 -1.20
## Total assets	51633.28	5.98 38.93
## Net worth	20912.36	6.39 44.36
## Total income	42261.36	5.38 32.16
## Change in stock	613.10	4.99 27.52
## Total expenses	38267.20	5.26 30.64
## Profit after tax	4072.58	6.50 46.05
## PBDITA	7369.48	5.98 38.87
## PBT	5424.10	6.44 45.10
## Cash profit	5453.88	6.39 44.42
## PBDITA as % of total income	80.43	2.77 10.45
## PBT as % of total income	55.10	2.41 7.61
## PAT as % of total income	46.73	2.66 9.31
## Cash profit as % of total income	56.24	2.66 9.42
## PAT as % of net worth	102.30	2.00 5.07
## Sales	40579.48	5.36 31.85
## Income from financial services	769.74	6.45 44.28
## Other income	285.96	6.24 41.30
## Total capital	2933.10	5.26 30.57
## Reserves and funds	17427.42	6.21 42.15
## Borrowings	14798.18	5.75 36.40
## Current liabilities & provisions	11088.10	5.61 34.21
## Deferred tax liability	2225.20	6.29 42.12
## Shareholders funds	20912.16	6.32 43.52

## Cumulative retained profits	13047.82	6.20	42.44
## Capital employed	34897.40	5.95	38.55
## TOL/TNW	55.81	5.58	34.40
## Total term liabilities / tangible net worth	29.46	6.31	43.38
## Contingent liabilities / Net worth (%)	773.79	5.63	36.55
## Contingent liabilities	6175.30	5.58	33.66
## Net fixed assets	16854.56	5.96	38.51
## Investments	5023.64	6.31	41.82
## Current assets	17370.54	5.61	34.75
## Net working capital	3741.36	5.60	34.51
## Quick ratio (times)	13.81	5.65	36.40
## Current ratio (times)	18.30	5.64	35.96
## Debt to equity ratio (times)	37.10	5.75	36.91
## Cash to current liabilities (times)	5.77	5.31	31.32
## Cash to average cost of sales per day	1276.65	6.09	40.00
## Creditors turnover	129.63	4.32	21.07
## Debtors turnover	181.21	5.16	29.34
## Finished goods turnover	852.18	5.48	32.41
## WIP turnover	215.88	4.41	22.68
## Raw material turnover	97.82	3.96	19.17
## Shares outstanding	268159911.20	5.04	27.68
## Equity face value	180.00	3.71	14.16
## EPS	896.83	6.43	43.51
## Adjusted EPS	896.80	6.49	44.16
## Total liabilities	51633.28	5.98	38.93
## PE on BSE	152.35	6.16	41.25
## Default	1.00	3.49	10.20
##	se	IQR	Q0.01
## Num	17.20	1772.00	36.40
## Total assets	110.28	1007.40	25.50
## Net worth	43.09	346.00	8.40
## Total income	94.07	1219.10	21.40
## Change in stock	1.42	14.10	-12.90
## Total expenses	86.87	1180.50	16.60
## Profit after tax	8.23	47.50	-2.70
## PBDITA	15.85	131.80	0.60
## PBT	11.03	66.60	-2.70
## Cash profit	11.22	83.70	-0.10
## PBDITA as % of total income	0.22	11.08	0.38
## PBT as % of total income	0.16	8.03	-3.44
## PAT as % of total income	0.13	5.86	-3.18
## Cash profit as % of total income	0.15	8.47	0.00
## PAT as % of net worth	0.31	20.19	-4.94
## Sales	90.79	1181.40	25.60
## Income from financial services	1.60	4.70	0.20
## Other income	0.61	1.70	0.20
## Total capital	6.60	87.20	3.60
## Reserves and funds	36.20	257.40	-10.70
## Borrowings	31.85	266.30	5.20

## Current liabilities & provisions	24.63	230.40	3.90
## Deferred tax liability	4.67	20.20	1.40
## Shareholders funds	43.27	361.20	8.60
## Cumulative retained profits	26.76	198.10	-20.50
## Capital employed	74.90	706.50	17.20
## TOL/TNW	0.12	2.23	0.15
## Total term liabilities / tangible net worth	0.06	0.95	0.00
## Contingent liabilities / Net worth (%)	1.62	30.76	0.00
## Contingent liabilities	13.74	68.20	2.40
## Net fixed assets	35.89	301.50	7.60
## Investments	10.73	11.50	0.40
## Current assets	37.89	448.70	6.60
## Net working capital	8.15	82.60	-53.20
## Quick ratio (times)	0.03	0.60	0.19
## Current ratio (times)	0.04	0.75	0.62
## Debt to equity ratio (times)	0.08	1.53	0.00
## Cash to current liabilities (times)	0.01	0.17	0.01
## Cash to average cost of sales per day	2.71	18.26	0.85
## Creditors turnover	0.32	6.61	1.74
## Debtors turnover	0.42	6.81	1.69
## Finished goods turnover	1.92	20.51	5.04
## WIP turnover	0.51	11.01	3.00
## Raw material turnover	0.24	7.51	0.00
## Shares outstanding	623623.42	6110140.00	350000.00
## Equity face value	0.49	0.00	10.00
## EPS	1.88	9.62	-0.69
## Adjusted EPS	1.87	7.48	-0.66
## Total liabilities	110.28	1007.40	25.50
## PE on BSE	0.31	0.00	3.70
## Default	0.00	0.00	0.00
##	00.05	00.1	00.25
## Num	178.00	355.00	886.00
## Total assets	25.50	25.50	91.30
## Net worth	8.40	8.40	31.30
## Total income	21.40	21.40	121.20
## Change in stock	-12.90	-12.90	-0.70
## Total expenses	16.60	16.60	104.10
## Profit after tax	-2.70	-2.70	0.60
## PBDITA	0.60	0.60	7.30
## PBT	-2.70	-2.70	0.90
## Cash profit	-0.10	-0.10	3.10
## PBDITA as % of total income	0.38	0.38	5.07
## PBT as % of total income	-3.44	-3.44	0.60
## PAT as % of total income	-3.18	-3.18	0.39
## Cash profit as % of total income	0.00	0.00	2.09
## PAT as % of net worth	-4.94	-4.94	0.00
## Sales	25.60	25.60	133.30
## Income from financial services	0.20	0.20	0.70
## Other income	0.20	0.20	0.80

## Total capital	3.60	3.60	13.10
## Reserves and funds	-10.70	-10.70	5.80
## Borrowings	5.20	5.20	29.70
## Current liabilities & provisions	3.90	3.90	18.70
## Deferred tax liability	1.40	1.40	6.70
## Shareholders funds	8.60	8.60	32.00
## Cumulative retained profits	-20.50	-20.50	1.30
## Capital employed	17.20	17.20	60.80
## TOL/TNW	0.15	0.15	0.60
## Total term liabilities / tangible net worth	0.00	0.00	0.05
## Contingent liabilities / Net worth (%)	0.00	0.00	0.00
## Contingent liabilities	2.40	2.40	16.00
## Net fixed assets	7.60	7.60	27.30
## Investments	0.40	0.40	4.60
## Current assets	6.60	6.60	37.20
## Net working capital	-53.20	-53.20	-1.00
## Quick ratio (times)	0.19	0.19	0.42
## Current ratio (times)	0.62	0.62	0.94
## Debt to equity ratio (times)	0.00	0.00	0.22
## Cash to current liabilities (times)	0.01	0.01	0.02
## Cash to average cost of sales per day	0.85	0.85	2.89
## Creditors turnover	1.74	1.74	3.94
## Debtors turnover	1.69	1.69	4.01
## Finished goods turnover	5.04	5.04	10.21
## WIP turnover	3.00	3.00	5.93
## Raw material turnover	0.00	0.00	3.41
## Shares outstanding	350000.00	350000.00	2209860.00
## Equity face value	10.00	10.00	10.00
## EPS	-0.69	-0.69	0.00
## Adjusted EPS	-0.66	-0.66	0.00
## Total liabilities	25.50	25.50	91.30
## PE on BSE	3.70	3.70	9.10
## Default	0.00	0.00	0.00
##	Q0.75	Q0.9	
## Num	2658.00	3191.00	
## Total assets	1098.70	3936.80	
## Net worth	377.30	1428.00	
## Total income	1340.30	4316.90	
## Change in stock	13.40	59.30	
## Total expenses	1284.60	4211.20	
## Profit after tax	48.10	245.20	
## PBDITA	139.10	558.50	
## PBT	67.50	328.50	
## Cash profit	86.80	360.30	
## PBDITA as % of total income	16.15	25.00	
## PBT as % of total income	8.63	15.93	
## PAT as % of total income	6.25	12.01	
## Cash profit as % of total income	10.56	17.22	
## PAT as % of net worth	20.19	35.65	

## Sales	1314.70	4183.90
## Income from financial services	5.40	31.00
## Other income	2.50	12.30
## Total capital	100.30	295.00
## Reserves and funds	263.20	1118.70
## Borrowings	296.00	1095.90
## Current liabilities & provisions	249.10	933.60
## Deferred tax liability	26.90	109.20
## Shareholders funds	393.20	1511.20
## Cumulative retained profits	199.40	922.60
## Capital employed	767.30	2634.60
## TOL/TNW	2.83	5.67
## Total term liabilities / tangible net worth	1.00	2.20
## Contingent liabilities / Net worth (%)	30.76	83.14
## Contingent liabilities	84.20	419.10
## Net fixed assets	328.80	1126.80
## Investments	16.10	167.30
## Current assets	485.90	1567.70
## Net working capital	81.60	325.60
## Quick ratio (times)	1.02	1.80
## Current ratio (times)	1.69	2.79
## Debt to equity ratio (times)	1.75	3.66
## Cash to current liabilities (times)	0.19	0.61
## Cash to average cost of sales per day	21.15	75.38
## Creditors turnover	10.55	23.68
## Debtors turnover	10.82	22.24
## Finished goods turnover	30.72	80.38
## WIP turnover	16.94	38.80
## Raw material turnover	10.92	20.51
## Shares outstanding	832000.00	26956800.00
## Equity face value	10.00	10.00
## EPS	9.62	36.33
## Adjusted EPS	7.48	31.26
## Total liabilities	1098.70	3936.80
## PE on BSE	9.10	16.78
## Default	0.00	0.00
##	00.95	00.99
## Num	3368.00	3509.60
## Total assets	8452.90	50240.79
## Net worth	3034.40	20892.42
## Total income	8870.10	41958.86
## Change in stock	142.90	599.96
## Total expenses	8586.90	38129.12
## Profit after tax	562.40	4022.41
## PBDITA	1219.10	7351.77
## PBT	735.00	5359.72
## Cash profit	803.00	5436.07
## PBDITA as % of total income	34.19	80.75
## PBT as % of total income	22.89	51.64

## PAT as % of total income	18.04	43.43
## Cash profit as % of total income	24.37	55.97
## PAT as % of net worth	50.23	97.35
## Sales	8764.10	40220.17
## Income from financial services	92.70	769.12
## Other income	36.40	285.42
## Total capital	607.60	2924.46
## Reserves and funds	2714.90	17293.55
## Borrowings	2534.10	14761.07
## Current liabilities & provisions	2005.60	11076.40
## Deferred tax liability	286.50	2213.40
## Shareholders funds	3160.00	20892.42
## Cumulative retained profits	2024.50	12987.67
## Capital employed	5988.70	34757.52
## TOL/TNW	10.53	55.61
## Total term liabilities / tangible net worth	4.20	29.37
## Contingent liabilities / Net worth (%)	151.04	750.66
## Contingent liabilities	1103.00	6158.50
## Net fixed assets	2671.00	16779.22
## Investments	544.60	5017.22
## Current assets	3300.40	17162.60
## Net working capital	673.80	3687.78
## Quick ratio (times)	2.91	14.00
## Current ratio (times)	4.20	18.87
## Debt to equity ratio (times)	6.75	36.57
## Cash to current liabilities (times)	1.22	5.75
## Cash to average cost of sales per day	184.92	1277.50
## Creditors turnover	39.73	131.34
## Debtors turnover	40.79	181.04
## Finished goods turnover	158.11	854.90
## WIP turnover	68.83	216.51
## Raw material turnover	32.05	97.50
## Shares outstanding	61380853.00	261116746.72
## Equity face value	100.00	154.00
## EPS	87.71	881.82
## Adjusted EPS	84.23	881.82
## Total liabilities	8452.90	50240.79
## PE on BSE	29.43	156.01
## Default	1.00	1.00

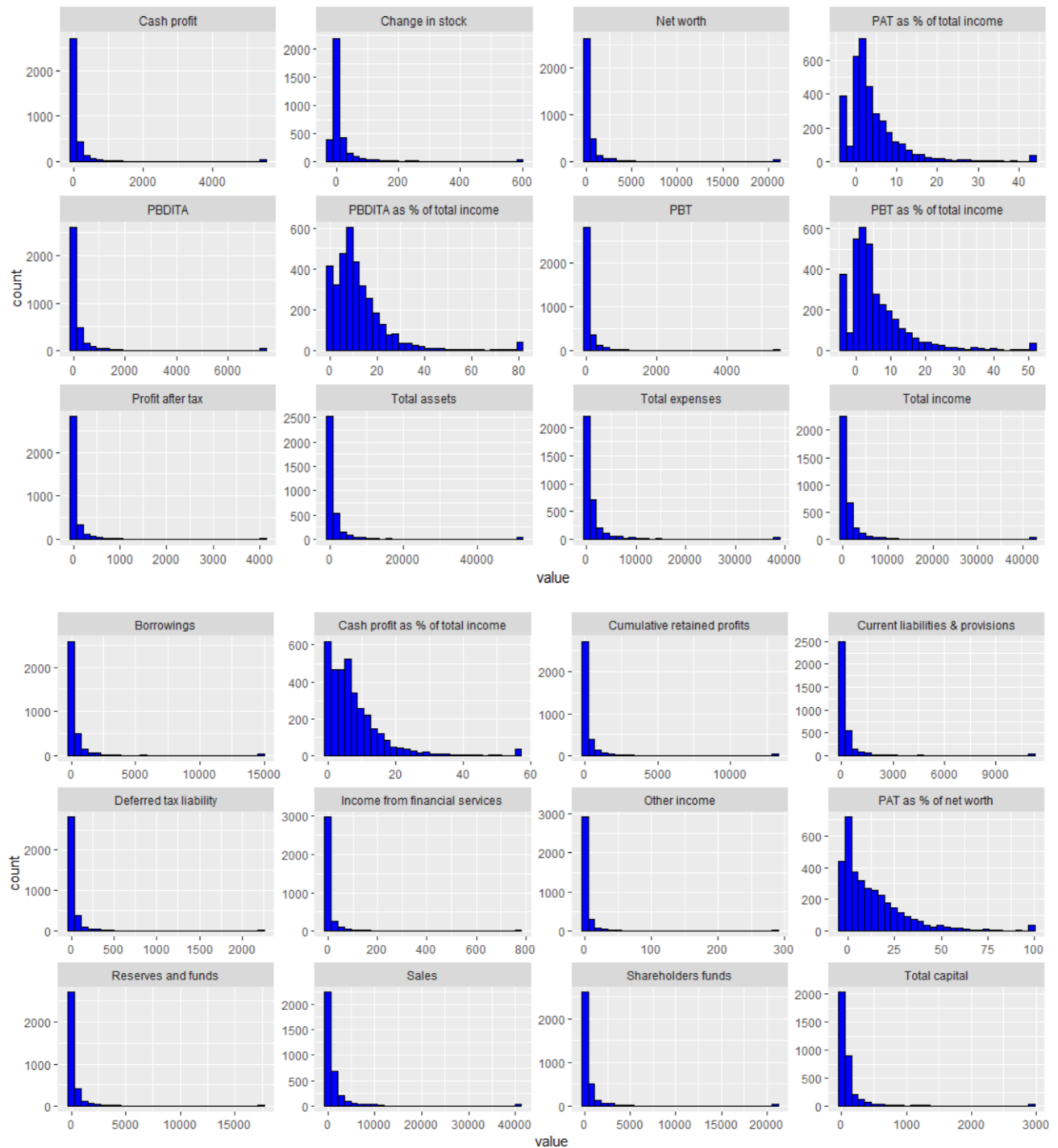
Data Visualization

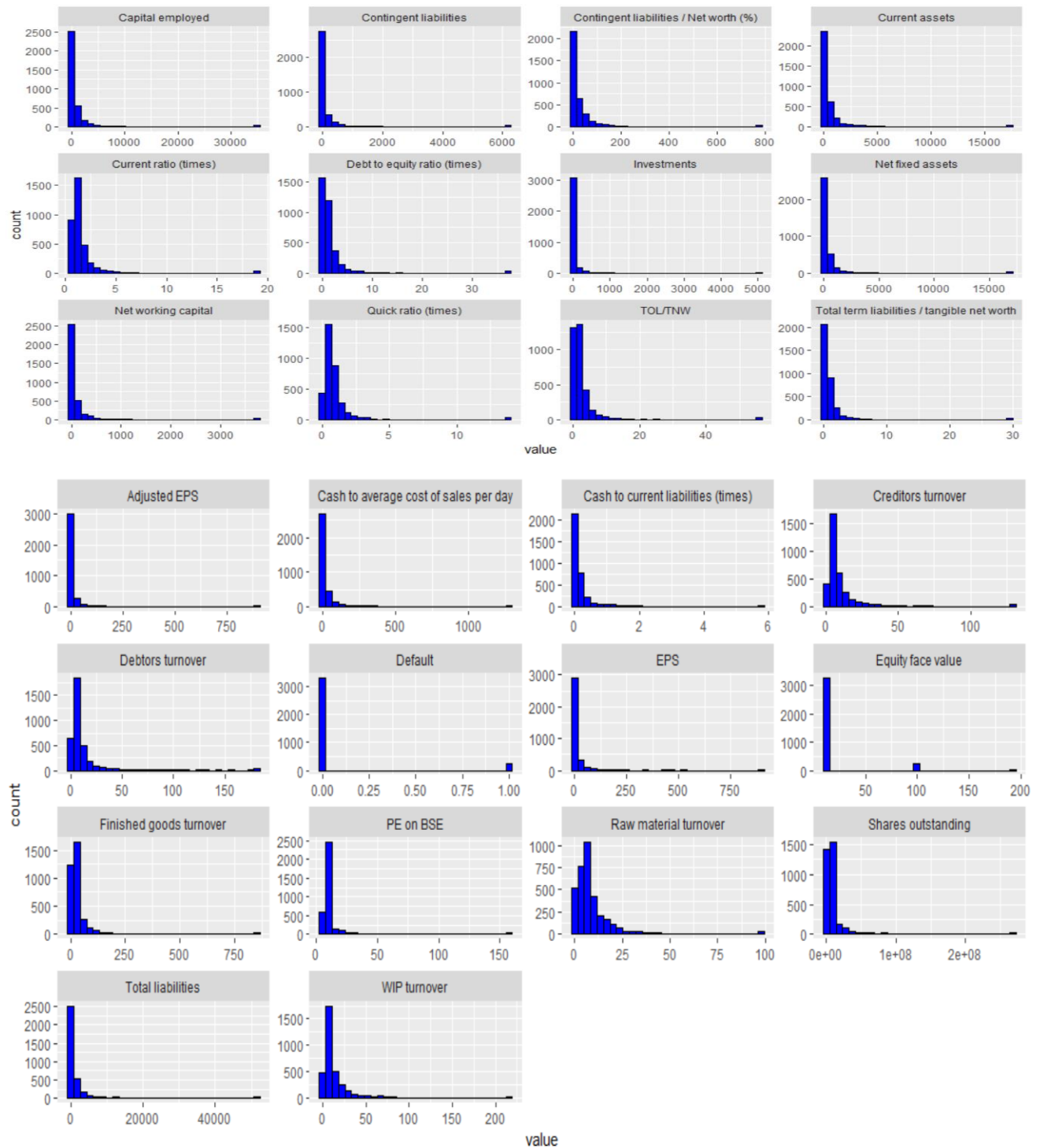
#Histogram for numerical variables (Univariate Analysis)

```
trainDS[,c(2:13)] %>% keep(is.numeric) %>% gather() %>%
```



```
ggplot(aes(value)) + facet_wrap(~key, scales = "free") + geom_histogram(col = "black", fill = "blue")
```

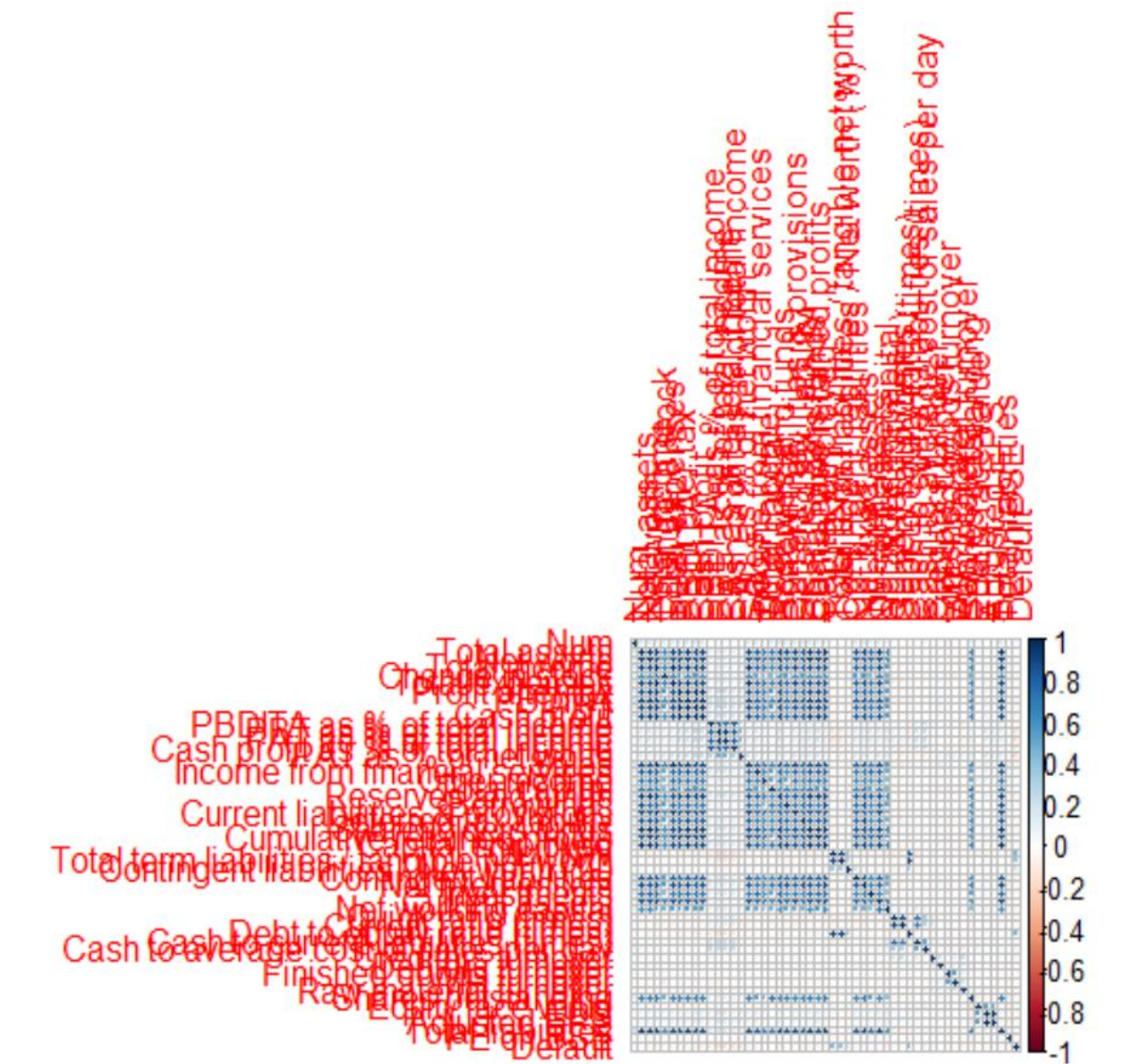




Observation

All the variables are slightly right skewed.

Correlation graph for Bivariate Analysis



The correlation graph shows the values in blue having the highest correlation, while in red having the lowest correlation. From the graph as well as the correlation matrix, we can see that most of the variables are highly correlated.

Variable Creation

Four new variables have been created.

1. **Profitability** – It is used to assess a business's ability to generate earnings relative to its revenue, operating costs etc using data from a specific point in time.

$$\text{Profitability} = \text{Profit after tax} / \text{Sales.}$$

2. **Leverage** – It is used to measure the portion of company's assets that are financed by stockholder's equity.

$$\text{Leverage} = \text{Total Assets} / (\text{Total Liabilities} / \text{Debit to equity ratio})$$

3. **Liquidity** – It is used to measure the ability of an organization to pay off its short-term obligations.

$$\text{Liquidity} = \text{Net Working Capital} / \text{Total Assets}$$

4. **Company Size** – It is used to analyze financial statements by expressing each line item as a percentage of the base amount for that period.

$$\text{Company size} = \text{Net worth} / \text{Total Asset}$$

#Profitability ratio is derived by dividing "Profit after tax" by "Sales"
`trainDS$Profitability = trainDS$`Profit after tax` / trainDS$Sales`

#Leverage ratio (Equity Multiplier) is derived by dividing Total Asset by Total Equity

#trainDS\$`Total Equity` = trainDS\$`Total Liabilities`/trainDS\$`Debt to equity ratio (times)`

`trainDS$`Equity Multiplier` = trainDS$`Total assets` / (trainDS$`Total liabilities`/trainDS$`Debt to equity ratio (times)`)`

#Liquidity ratio (Net working capital ratio) is derived by dividing Net Working Capital by Total Asset

`trainDS$`Net working Capital ratio` = trainDS$`Net working capital`/trainDS$`Total assets``

#Company size ratio is got by dividing the value of a column by the Total Assets

`trainDS$`Networth to Total Assets` = trainDS$`Net worth`/trainDS$`Total assets``

`trainDS$`Totalincome To Total Assets` = trainDS$`Total income`/trainDS$`Total assets``

`trainDS$`Totalexpenditures To Total Assets` = trainDS$`Total expenses`/trainDS$`Total assets``

Model Building

Logistic Regression is a predictive analysis model. It is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. As multicollinearity exists, the model needs to be treated using VIF.

Initially, all the variables are used as the predictors. The Default variable will be used as the response variable.

```
#check split consistency
sum(as.integer(as.character(trainDS$Default)))/nrow(trainDS)

## [1] 0.06608303
```

This is an imbalanced dataset.

```
set.seed(1000)
modell1 = glm(Default~.-`Total liabilities`-`Equity Multiplier`-Num, data = trainDS, family = binomial)

both_model = stepAIC(modell1,direction = "both")

summary(both_model)

## Call:
## glm(formula = Default ~ `Total assets` + `Total income` + `Total expenses` +
##      PBDITA + PBT + `Cash profit` + `PAT as % of net worth` +
##      `Other income` + `Reserves and funds` + `Current liabilities & provisions` +
##      `Cumulative retained profits` + `Capital employed` + `TOL/TNW` +
##      `Total term liabilities / tangible net worth` + `Contingent liabilities` +
##      `Net fixed assets` + `Debt to equity ratio (times)` + `Cash to average cost of sales per day` +
##      `PE on BSE` + `Net working Capital ratio` + `Networth to Total Assets` +
##      `Totalincome To Total Assets`, family = binomial, data = trainDS)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.7469  -0.2552  -0.0783  -0.0009   5.1985
##
## Coefficients:
##                                Estimate Std. Error
## (Intercept)                   -0.4508076   0.3746486
## `Total assets`                  0.0041811   0.0015556
## `Total income`                 -0.0018733   0.0008611
## `Total expenses`               0.0017653   0.0008375
```

```

## PBDITA -0.0075715 0.0031074
## PBT 0.0103016 0.0033464
## `Cash profit` -0.0107527 0.0058968
## `PAT as % of net worth` -0.0547582 0.0096222
## `Other income` 0.0230951 0.0105972
## `Reserves and funds` -0.0046234 0.0017572
## `Current liabilities & provisions` -0.0037080 0.0016125
## `Cumulative retained profits` -0.0154735 0.0043657
## `Capital employed` -0.0043600 0.0016168
## `TOL/TNW` 0.0362774 0.0173637
## `Total term liabilities / tangible net worth` -0.0761041 0.0404855
## `Contingent liabilities` -0.0015385 0.0006371
## `Net fixed assets` 0.0004317 0.0002538
## `Debt to equity ratio (times)` 0.1077656 0.0313323
## `Cash to average cost of sales per day` 0.0009460 0.0003832
## `PE on BSE` -0.0796252 0.0346838
## `Net working Capital ratio` -1.7963593 0.4645830
## `Networth to Total Assets` -3.9197846 0.7001173
## `Totalincome To Total Assets` 0.1089325 0.0197731
## z value Pr(>|z|)
## (Intercept) -1.203 0.228868
## `Total assets` 2.688 0.007192 **
## `Total income` -2.175 0.029602 *
## `Total expenses` 2.108 0.035046 *
## PBDITA -2.437 0.014826 *
## PBT 3.078 0.002081 **
## `Cash profit` -1.823 0.068232 .
## `PAT as % of net worth` -5.691 1.26e-08 ***
## `Other income` 2.179 0.029305 *
## `Reserves and funds` -2.631 0.008510 **
## `Current liabilities & provisions` -2.300 0.021475 *
## `Cumulative retained profits` -3.544 0.000394 ***
## `Capital employed` -2.697 0.007003 **
## `TOL/TNW` 2.089 0.036684 *
## `Total term liabilities / tangible net worth` -1.880 0.060138 .
## `Contingent liabilities` -2.415 0.015734 *
## `Net fixed assets` 1.701 0.088970 .
## `Debt to equity ratio (times)` 3.439 0.000583 ***
## `Cash to average cost of sales per day` 2.469 0.013557 *
## `PE on BSE` -2.296 0.021690 *
## `Net working Capital ratio` -3.867 0.000110 ***
## `Networth to Total Assets` -5.599 2.16e-08 ***
## `Totalincome To Total Assets` 5.509 3.61e-08 ***
## ---
## 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: 911.84 on 3518 degrees of freedom
## AIC: 957.84
##
## Number of Fisher Scoring iterations: 13

#We are considering the VIF values greater than 10 as having high multicollinearity.
vif(both_model)

## `Total assets`                `Total income`
## 990.705931                    51.408065
## `Total expenses`              PBDITA
## 56.245822                     41.449157
## PBT                           `Cash profit`
## 19.212265                     1.431703
## `PAT as % of net worth`       `Other income`
## 1.187423                      1.622050
## `Reserves and funds`
## 17.543570
## `Current liabilities & provisions`
## 37.398429
## `Cumulative retained profits`  `Capital employed`
## 1.323916                      565.655176
## `TOL/TNW`
## 4.733955
## `Total term liabilities / tangible net worth`
## 7.342550
## `Contingent liabilities`       `Net fixed assets`
## 2.066820                      4.405997
## `Debt to equity ratio (times)`
## 6.759146
## `Cash to average cost of sales per day`
## 1.215207
## `PE on BSE`                  `Net working Capital ratio`
## 1.111728                      1.138701
## `Networth to Total Assets`    `Totalincome To Total Assets`
## 1.476179                      1.499482

```

The model was built for forward, backward and both-model. Based on the AIC and the p-value of each variable, the both-model was chosen. The AIC is 957.84. The variables were removed from the default model based on the p-value one by one. The p-value was considered as less than 0.05. The importance of the variable can be identified by *. For example, PAT as % of Net Worth has a very high importance and Cash Profit has comparatively have lower importance.

We can use variance inflation factor (VIF) to get rid of redundant predictors or the variables that have high multicollinearity between them. Multicollinearity exists when two or more predictor variables are highly related to each other and then it becomes difficult to understand the impact of an independent variable on the dependent variable.

The Variance Inflation Factor (VIF) is used to measure the multicollinearity between predictor variables in a model. A predictor having a VIF of 10 or less is generally considered safe and it can be assumed that it is not correlated with other predictor variables. Higher the VIF, greater is the correlation of the predictor variable w.r.t other predictor variables. However, Predictors with high VIF may have high p-value (or highly significant), hence, we need to see the significance of the Predictor variable before removing it from our model.

Likelihood – Ratio Test

The **Likelihood-Ratio test** (sometimes called the likelihood-ratio chi-squared test) is a hypothesis test that helps you choose the “best” model between two nested models. Basically, the test compares the fit of two models. The null hypothesis is that the smaller model is the “best” model; It is rejected when the test statistic is large. In other words, if the null hypothesis is rejected, then the larger model is a significant improvement over the smaller one.

```
#Log Likelihood test: To ensure if Logit model is valid or not
lrtest(both_model)

## Likelihood ratio test
##
## Model 1: Default ~ `Total assets` + `Total income` + `Total expenses` +
##      PBDITA + PBT + `Cash profit` + `PAT as % of net worth` +
##      `Other income` + `Reserves and funds` + `Current liabilities & provisions` +
##      `Cumulative retained profits` + `Capital employed` + `TOL/TNW` +
##      `Total term liabilities / tangible net worth` + `Contingent liabilities` +
##      `Net fixed assets` + `Debt to equity ratio (times)` + `Cash to average
##      cost of sales per day` +
##      `PE on BSE` + `Net working Capital ratio` + `Networth to Total Assets`
##      +
##      `Total income To Total Assets`
## Model 2: Default ~ 1
##      #Df  LogLik  Df  Chisq Pr(>Chisq)
## 1   23 -455.92
## 2    1 -861.83 -22 811.82 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The full model has a score of -455.92, while the null model has -861.83. We need to choose the model with a lower score. Therefore, the null hypothesis is not rejected.

```
# To get the Logit R2 of goodness
pR2(both_model)
```

	llh	llhNull	G2	McFadden	r2ML
	-455.9212867	-861.8334541	811.8243348	0.4709868	0.2048815

```
##          r2CU
##      0.5316156

# Trust only McFadden since its conservative
#if my McFadden > is between 0 to 10 - Goodness of fit is weak
#if my McFadden > is between 10 to 20 - Goodness of fit is fare
#if my McFadden > is between 20 to 30 - Goodness of fit is Moderately is robu
st
#if my McFadden > is between 30 and above - Goodness of fit is reasonably rob
ust model
#Typical in non-linear model R2 will be less as against Linear regression
```

Since our McFadden value is around 47.09%, we can consider our model to be reasonably robust.

```
odds = exp(coef(both_model))

#for identifying the relative importance of variables we have to use ODDS ins
tead of PROB
prob=odds/(1+odds)
relativeImportance=(odds[-1]/sum(odds[-1]))*100
relativeImportance[order(relativeImportance)]

## `Networth to Total Assets`          `Net working Capital ratio`
## 0.09805098                          0.81967898
## `PE on BSE`                        `PAT as % of net worth`
## 4.56259526                          4.67747584
## `Total term liabilities / tangible net worth`
## 4.57868935
## `Cumulative retained profits`        `Cash profit`
## 4.86488640                          4.88790665
## PBDITA                             `Reserves and funds`
## 4.90348067                          4.91795811
## `Capital employed`                  `Current liabilities & provisions`
## 4.91925359                          4.92246204
## `Total income`                      `Contingent liabilities`
## 4.93150167                          4.93315277
## `Net fixed assets`                  `Total expenses`
## 4.94288168                          4.94947803
## `Cash to average cost of sales per day`
## 4.94542465
## `Total assets`                      PBT
## 4.96144955                          4.99190902
## `Other income`                      `TOL/TNW`
## 5.05618349                          5.12327677
## `Debt to equity ratio (times)`      `Totalincome To Total Assets`
## 5.50293970                          5.50936481
```



```

# Performance on TRAIN dataset
predTrain = predict(both_model, newdata = trainDS, type="response")
pt =table(trainDS$Default, predTrain>0.5)

sum(diag(pt)) / nrow(trainDS)
## [1] 0.9497317

(3273+90)/nrow(trainDS)
## [1] 0.9497317

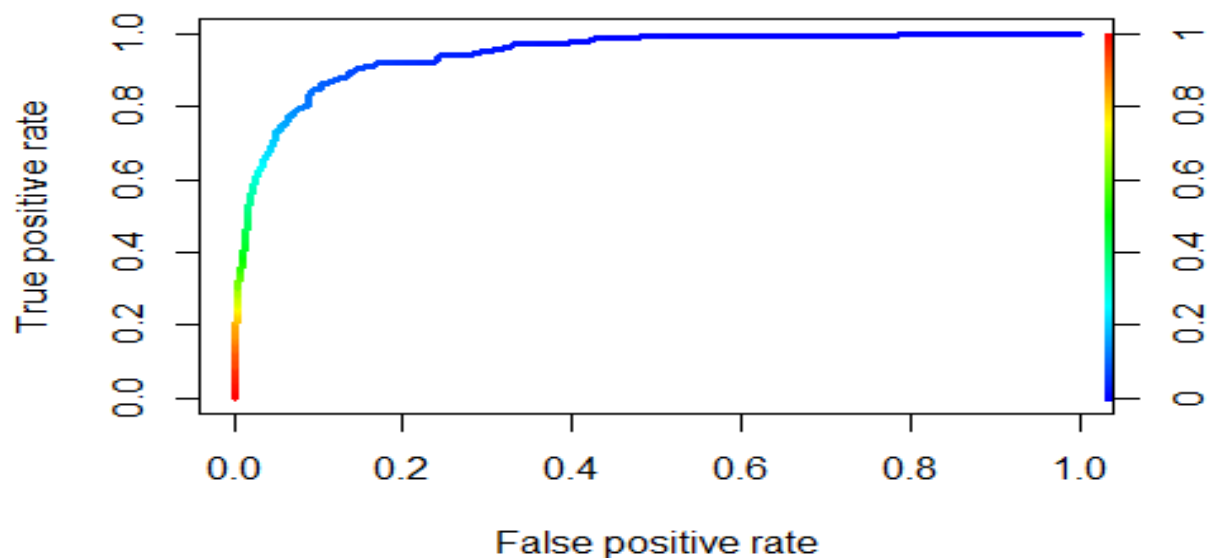
(144+34)/nrow(trainDS)
## [1] 0.05026829

ROCpred = prediction(predTrain, trainDS$Default)
as.numeric(performance(ROCpred, "auc")@y.values)
## [1] 0.9444768

perf = performance(ROCpred, "tpr", "fpr")
plot(perf,col="black",lty=2, lwd=2)

```

ROC stands for **Receiver Operating Characteristic**. ... The **ROC curve** does this by plotting sensitivity, the probability of predicting a real positive will be a positive, against 1-specificity, the probability of predicting a real negative will be a positive. The ROC for the train data set is 0.9445.



```

# Performance on TEST dataset
predTest = predict(both_model, newdata = testDS, type="response")
ptest = table(testDS$Default, predTest>0.5)

sum(diag(ptest)) / nrow(testDS)

## [1] 0.9286713

(616+48)/nrow(testDS)

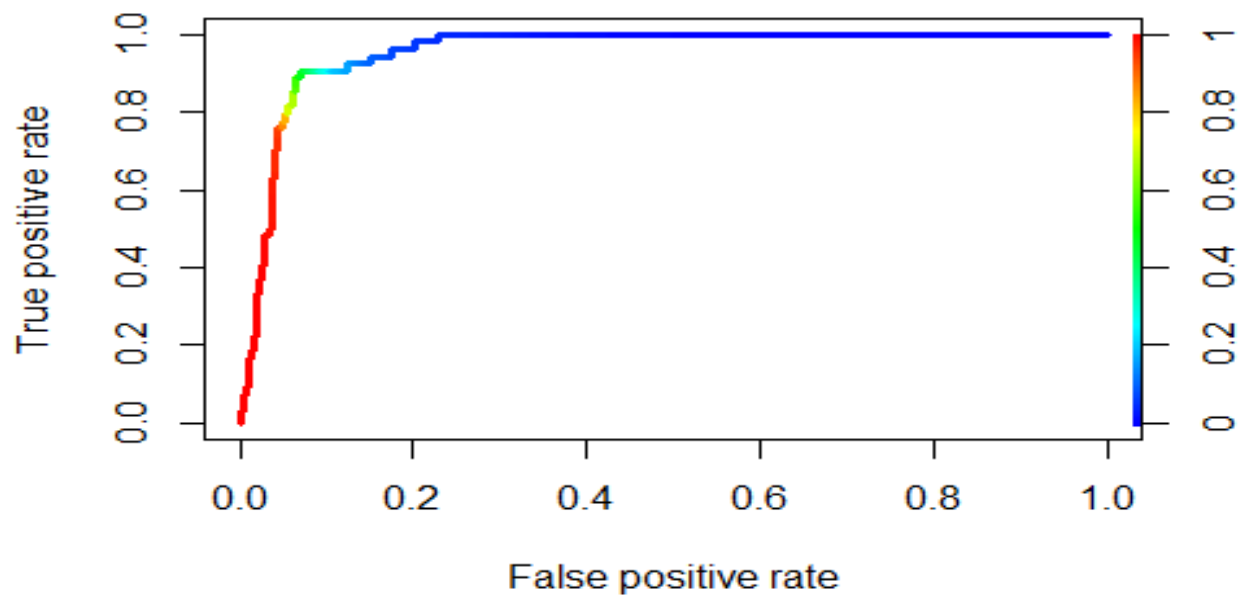
## [1] 0.9286713

ROCRpred = prediction(predTest, testDS$Default)
as.numeric(performance(ROCRpred, "auc")@y.values)

## [1] 0.9562251

perf = performance(ROCRpred, "tpr", "fpr")
plot(perf, col="black", lty=2, lwd=2)

```



The **ROC** for the test data set is 0.9562.

Confusion Matrix

A confusion matrix is a technique for summarizing the performance of a classification algorithm.

Matrix for Logistic Regression Model for train.

	False	True
0	3273	34
1	144	90

We can see that around 3273 out of 3417 are predicted as 0 while 90 out of 124 are predicted as 1. About 3363 are predicted correctly and 178 are having wrong predictions. The **False Negative or Type II error** is 34. The **False Positive or Type I error** is 144.

The **Accuracy** is 94.97%. The **Sensitivity** is 38.46%, while the **specificity** is 98.97%. The **error rate** is 5.03%.

Matrix for Logistic Regression Model for test.

	False	True
0	616	45
1	6	46

We can see that around 616 out of 622 are predicted as 0 while 46 out of 91 are predicted as 1. About 662 are predicted correctly and 51 are having wrong predictions. The **False Negative or Type II error** is 45. The **False Positive or Type I error** is 6.

The **Accuracy** is 92.87%. The **Sensitivity** is 88.88%, while the **specificity** is 93.19%. The **error rate** is 7.13%.

Decile

The train and test dataset are divided into 10 deciles based on the probability of default.

Train Decile

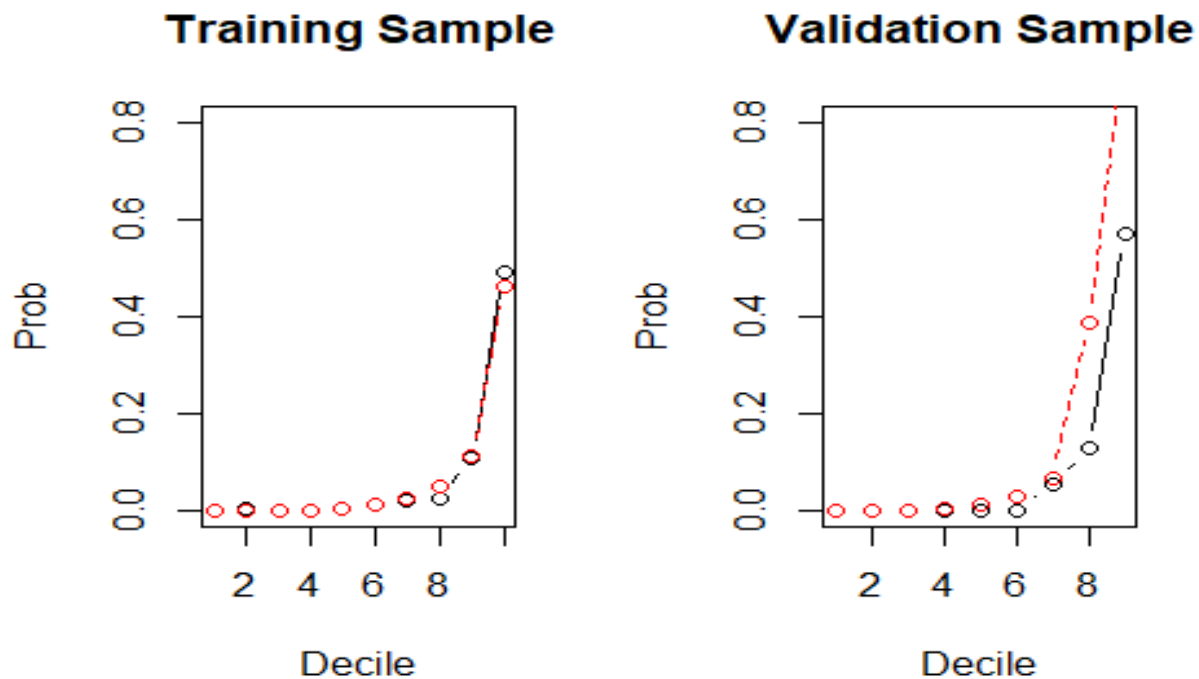
Deciles	cnt	Cnt_resp	Cnt_non_resp	rrate	Cum_resp	Cum_non_resp	Cum_rel_resp	Cum_rel_non_resp	ks
10	355	174	181	49%	174	181	74%	5.50%	68.89
9	354	38	316	11%	212	497	91%	15.00%	75.57
8	353	9	344	3%	221	841	94%	25.40%	69.01
7	355	7	348	2%	228	1189	97%	36.00%	61.49
6	354	4	350	1%	232	1539	99%	46.50%	52.61
5	354	1	353	0%	233	1892	100%	57.20%	42.36
4	353	0	353	0%	233	2245	100%	67.90%	31.68
3	355	0	355	0%	233	2600	100%	78.60%	20.95
2	354	1	353	0%	234	2953	100%	89.30%	10.7
1	354	0	354	0%	234	3307	100%	100.00%	0

The deciles are sorted in the descending order. The 10th decile has the maximum number of defaults in the form of cnt_resp. The response rate is gradually dropping as it is in descending order.

Test Decile

Deciles	cnt	Cnt_resp	Cnt_non_resp	rrate	Cum_resp	Cum_non_resp	Cum_rel_resp	Cum_rel_non_resp	ks
10	72	41	31	0.5694	41	31	76%	4.7%	71.24
9	71	9	62	0.1268	50	93	93%	14.1%	78.52
8	72	4	68	0.0556	54	161	100%	24.4%	75.64
7	71	0	71	0.0000	54	232	100%	35.1%	64.90
6	72	0	72	0.0000	54	304	100%	46.0%	54.01
5	71	0	71	0.0000	54	375	100%	56.7%	43.27
4	71	0	71	0.0000	54	446	100%	67.5%	32.53
3	72	0	72	0.0000	54	518	100%	78.4%	21.63
2	143	0	143	0.0000	54	661	100%	100.0%	0.00

The deciles are sorted in the descending order. The 10th decile has the maximum number of defaults in the form of cnt_resp. The response rate is gradually dropping as it is in descending order.



The mean is taken for both the training and test dataset, to differentiate the predicted and observed values. The plot shows that the model almost accurately predicted both the training and test dataset with an accuracy of almost 95% and 93% respectively.

Conclusion

The aim of India Credit risk model is to build a default credit model using logistic regression. Therefore, they need information about the connection between the variables and the given data.

The data was treated for missing values and outliers. New variables were added for Profitability, Leverage, Liquidity and Company size.

A logistic regression model was built on the train data and was validated against the test data. The model gave an accuracy of 92.87% on the test dataset. The Sensitivity is 88.88%, while the specificity is 93.19%. The error rate is 7.13%.

Recommendation

The model could be improved further by adding more new variables based on different financial ratios.