Project -9 (Finance and Risk Analytics)

India credit risk(default) model

By Saleesh Satheeshchandran
PGP-BABI 2019-20 (G-6)

Objective

The project is done to build a India credit risk(default) model, using the data provided in the spreadsheet raw-data.xlsx, and validate it on validation_data.xlsx.

This project deals with some statistical modeling problems that are motivated by credit risk (default) analysis. The given data shows Central to credit risk is the default event, which occurs if the debtor is unable to meet its legal obligation according to the debt contract.

Assumptions

There are no particular assumptions.

Tool used for the analysis

RStudio Version 1.2.1335
R Version 3.6.0

Input Data

The data is available in a spreadsheet format with .xisx file extension.
The same is uploaded into the R Studio with the help of the function “read-xlsx.

Steps involved in the analysis

- Exploratory Data Analysis
- Univariate Analysis
- Bivariate Analysis
- Missing values
- Outliers and their treatment
- New Variables Creation (One ration for profitability, leverage, liquidity and company's size each)
- Multicollinearity
- Logistics Regression
- Analyze coefficient & their signs
- Predict accuracy of model on dev and validation datasets

Exploratory Data Analysis

```
> summary(IndCredData)
 Networth Next Year  Total assets       Net worth        Total income      Change in stock     Total expenses
 Min.   :-74265.6   Min.   :      0.1   Min.   :     0.0   Min.   :      0.0   Min.   :-3029.40   Min.   :     -0.1
 1st Qu.:     31.7   1st Qu.:     91.3   1st Qu.:    31.3   1st Qu.:    106.5   1st Qu.:   -1.80   1st Qu.:     95.8
 Median :    116.3   Median :    309.7   Median :   102.3   Median :    444.9   Median :    1.60   Median :    407.7
 Mean   :   1616.3   Mean   :   3443.4   Mean   :  1295.9   Mean   :   4582.8   Mean   :   41.49   Mean   :   4262.9
 3rd Qu.:    456.1   3rd Qu.:   1098.7   3rd Qu.:   377.3   3rd Qu.:   1440.9   3rd Qu.:   18.05   3rd Qu.:   1359.8
 Max.   :805773.4   Max.   :1176509.2   Max.   :613151.6   Max.   :2442828.2   Max.   :14185.50   Max.   :2366035.3
                                                          NA's   :198          NA's   :458        NA's   :139
 Profit after tax        PBDITA               PBT              Cash profit        PBDITA as % of total income PBT as % of total income
 Min.   : -3908.30   Min.   : -440.7   Min.   : -3894.80   Min.   : -2245.70   Min.   :-6400.000            Min.   :-21340.00
 1st Qu.:     0.50   1st Qu.:    6.9   1st Qu.:     0.70   1st Qu.:     2.90   1st Qu.:    5.000            1st Qu.:     0.55
 Median :     8.80   Median :   35.4   Median :    12.40   Median :    18.85   Median :    9.660            Median :     3.31
 Mean   :   277.36   Mean   :  578.1   Mean   :   383.81   Mean   :   392.07   Mean   :    4.571            Mean   :   -17.28
 3rd Qu.:    52.27   3rd Qu.:  150.2   3rd Qu.:    71.97   3rd Qu.:    93.20   3rd Qu.:   16.390            3rd Qu.:     8.80
 Max.   :119439.10   Max.   :208576.5  Max.   :145292.60   Max.   :176911.80   Max.   :  100.000            Max.   :   100.00
 NA's   :131         NA's   :131       NA's   :131         NA's   :131         NA's   :68                   NA's   :68
 PAT as % of total income Cash profit as % of total income PAT as % of net worth     Sales          Total capital
 Min.   :-21340.00        Min.   :-15020.000               Min.   :-748.72        Min.   :      0.1   Min.   :    0.1
 1st Qu.:     0.35        1st Qu.:     2.020               1st Qu.:   0.00        1st Qu.:    112.7   1st Qu.:   13.1
 Median :     2.34        Median :     5.640               Median :   7.92        Median :    453.1   Median :   42.1
 Mean   :   -19.20        Mean   :    -8.229               Mean   :  10.27        Mean   :   4549.5   Mean   :  216.6
 3rd Qu.:     6.34        3rd Qu.:    10.700               3rd Qu.:  20.19        3rd Qu.:   1433.5   3rd Qu.:  100.3
 Max.   :   150.00        Max.   :   100.000               Max.   :2466.67        Max.   :2384984.4   Max.   :78273.2
 NA's   :68               NA's   :68                                              NA's   :259         NA's   :4
 Reserves and funds   Borrowings        Current liabilities & provisions Shareholders funds Cumulative retained profits
 Min.   : -6525.9   Min.   :     0.10   Min.   :     0.1                 Min.   :     0.0   Min.   : -6534.3
 1st Qu.:     5.0   1st Qu.:    23.95   1st Qu.:    17.8                 1st Qu.:    32.0   1st Qu.:     1.1
 Median :    54.8   Median :    99.20   Median :    69.4                 Median :   105.6   Median :    37.1
 Mean   :  1163.8   Mean   :  1122.28   Mean   :   940.6                 Mean   :  1322.1   Mean   :   890.5
 3rd Qu.:   277.3   3rd Qu.:   352.60   3rd Qu.:   261.7                 3rd Qu.:   393.2   3rd Qu.:   202.3
 Max.   :625137.8   Max.   :278257.30   Max.   :352240.3                 Max.   :613151.6   Max.   :390133.8
 NA's   :85         NA's   :366         NA's   :96                                          NA's   :38

 Net fixed assets    Current assets      Net working capital Quick ratio (times) Current ratio (times) Debt to equity ratio (times)
 Min.   :     0.0   Min.   :     0.1   Min.   :-63839.0    Min.   :  0.000     Min.   :  0.00        Min.   :  0.00
 1st Qu.:    26.0   1st Qu.:    36.2   1st Qu.:    -1.1    1st Qu.:  0.410     1st Qu.:  0.93        1st Qu.:  0.22
 Median :    93.5   Median :   145.1   Median :    16.2    Median :  0.670     Median :  1.23        Median :  0.79
 Mean   :  1189.7   Mean   :  1293.4   Mean   :   138.6    Mean   :  1.401     Mean   :  2.13        Mean   :  2.78
 3rd Qu.:   344.9   3rd Qu.:   502.2   3rd Qu.:    84.2    3rd Qu.:  1.030     3rd Qu.:  1.71        3rd Qu.:  1.75
 Max.   :636604.6   Max.   :354815.2   Max.   : 85782.8    Max.   :341.000     Max.   :505.00        Max.   :456.00
 NA's   :118        NA's   :66         NA's   :32          NA's   :93          NA's   :93
 Cash to current liabilities (times) Cash to average cost of sales per day Creditors turnover Debtors turnover
 Min.   :  0.0000                    Min.   :     0.00                      Min.   :   0.000   Min.   :   0.00
 1st Qu.:  0.0200                    1st Qu.:     2.79                      1st Qu.:   3.700   1st Qu.:   3.76
 Median :  0.0700                    Median :     8.03                      Median :   6.095   Median :   6.32
 Mean   :  0.4904                    Mean   :   158.44                      Mean   :  15.446   Mean   :  17.04
 3rd Qu.:  0.1900                    3rd Qu.:    21.79                      3rd Qu.:  11.490   3rd Qu.:  11.68
 Max.   :165.0000                    Max.   :128040.76                      Max.   :2401.000   Max.   :3135.20
 NA's   :93                          NA's   :85                             NA's   :333        NA's   :328
 Finished goods turnover  WIP turnover      Raw material turnover Shares outstanding    Equity face value      EPS
 Min.   :   -0.09        Min.   :  -0.18   Min.   :   -2.00      Min.   :-2.147e+09   Min.   :-999999   Min.   :-843181.8
 1st Qu.:    8.20        1st Qu.:   5.10   1st Qu.:    2.99      1st Qu.: 1.316e+06   1st Qu.:     10   1st Qu.:      0.0
 Median :   17.27        Median :   9.76   Median :    6.40      Median : 4.672e+06   Median :     10   Median :      1.4
 Mean   :   87.08        Mean   :  27.93   Mean   :   19.09      Mean   : 2.207e+07   Mean   :  -1334   Mean   :   -220.3
 3rd Qu.:   40.35        3rd Qu.:  20.24   3rd Qu.:   11.85      3rd Qu.: 1.065e+07   3rd Qu.:     10   3rd Qu.:      9.6
 Max.   :17947.60        Max.   :5651.40   Max.   :21092.00      Max.   : 4.130e+09   Max.   : 100000   Max.   :  34522.5
 NA's   :740             NA's   :640       NA's   :361           NA's   :692          NA's   :692
  Adjusted EPS       Total liabilities
 Min.   :-843181.8   Min.   :      0.1
 1st Qu.:      0.0   1st Qu.:     91.3
 Median :      1.2   Median :    309.7
 Mean   :   -221.5   Mean   :   3443.4
 3rd Qu.:      7.5   3rd Qu.:   1098.7
 Max.   :  34522.5   Max.   :1176509.2
```

```
> str(IndCredData)
Classes ‘tbl_df’, ‘tbl’ and 'data.frame':       3541 obs. of  44 variables:
 $ Networth Next Year                          : num  8890.6 394.3 92.2 2.7 109 ...
 $ Total assets                                : num  17512.3 941 232.8 2.7 478.5 ...
 $ Net worth                                   : num  7093.2 351.5 100.6 2.7 107.6 ...
 $ Total income                                : num  24965 1527 477 NA 1580 ...
 $ Change in stock                             : num  235.8 42.7 -5.2 NA -17 ...
 $ Total expenses                              : num  23658 1455 479 NA 1558 ...
 $ Profit after tax                            : num  1543.2 115.2 -6.6 NA 5.5 ...
 $ PBDITA                                      : num  2860.2 283 5.8 NA 31 ...
 $ PBT                                         : num  2417.2 188.4 -6.6 NA 6.3 ...
 $ Cash profit                                 : num  1872.8 158.6 0.3 NA 11.9 ...
 $ PBDITA as % of total income                 : num  11.46 18.53 1.22 0 1.96 ...
 $ PBT as % of total income                    : num  9.68 12.33 -1.38 0 0.4 ...
 $ PAT as % of total income                    : num  6.18 7.54 -1.38 0 0.35 2.81 0 0.72 8.29 -2.88 ...
 $ Cash profit as % of total income            : num  7.5 10.38 0.06 0 0.75 ...
 $ PAT as % of net worth                       : num  23.78 38.08 -6.35 0 5.25 ...
 $ Sales                                       : num  24458 1504 476 NA 1575 ...
 $ Total capital                               : num  423.8 115.5 81.4 0.5 6.2 ...
 $ Reserves and funds                          : num  6822.8 257.8 19.2 2.2 161.8 ...
 $ Borrowings                                  : num  14.9 272.5 35.4 NA 193.1 ...
 $ Current liabilities & provisions            : num  9965.9 210 96.8 NA 112.8 ...
 $ Shareholders funds                          : num  7093.2 351.5 100.6 2.7 107.6 ...
 $ Cumulative retained profits                 : num  6263.3 247.4 32.4 2.2 82.7 ...
 $ Capital employed                            : num  7108.1 624 136 2.7 300.7 ...
 $ TOL/TNW                                     : num  1.33 1.23 1.44 0 2.83 1.8 0.03 5.17 1.05 3.25 ...
 $ Total term liabilities / tangible net worth: num  0 0.34 0.29 0 1.59 0.37 0.03 0.94 0.3 0.54 ...
 $ Contingent liabilities / Net worth (%)      : num  14.8 19.2 45.8 0 34.9 ...
 $ Net fixed assets                            : num  1900.2 286.4 38.7 2.5 94.8 ...
 $ Current assets                              : num  13277.5 563.9 167.5 0.2 349.7 ...
 $ Net working capital                         : num  3588.5 203.5 59.6 0.2 215.8 ...
 $ Quick ratio (times)                         : num  1.18 0.95 1.11 NA 1.41 0.48 NA 0.54 0.59 0.39 ...
 $ Current ratio (times)                       : num  1.37 1.56 1.55 NA 2.54 1.27 NA 1.15 1.58 0.5 ...
 $ Debt to equity ratio (times)                : num  0 0.78 0.35 0 1.79 1.09 0.32 2.31 0.94 3.13 ...
 $ Cash to current liabilities (times)         : num  0.43 0.06 0.21 NA 0 0.11 NA 0.04 0.19 0 ...
 $ Cash to average cost of sales per day       : num  68.21 5.96 17.07 NA 0 ...
 $ Creditors turnover                          : num  3.62 9.8 5.28 0 13 ...
 $ Debtors turnover                            : num  3.85 5.7 5.07 0 9.46 ...
 $ Finished goods turnover                     : num  200.55 14.21 9.24 NA 12.68 ...
 $ WIP turnover                                : num  21.78 7.49 0.23 NA 7.9 ...
 $ Raw material turnover                       : num  7.71 11.46 NA 0 17.03 ...
 $ Shares outstanding                          : num  42381675 11550000 8149090 52404 619635 ...
 $ Equity face value                           : num  10 10 10 10 10 10 10 NA 10 10 ...
 $ EPS                                         : num  35.52 9.97 -0.5 0 7.91 ...
 $ Adjusted EPS                                : num  7.1 9.97 -0.5 0 7.91 ...
 $ Total liabilities                           : num  17512.3 941 232.8 2.7 478.5 ...
> dim(IndCredData)
[1] 3541   44
> colSums(is.na(IndCredData))
```

Converting Data Types

```
IndCredData$`Creditors turnover`=as.numeric(IndCredData$`Creditors turnover`)
IndCredData$`Debtors turnover`=as.numeric(IndCredData$`Debtors turnover`)
IndCredData$`Finished goods turnover`=as.numeric(IndCredData$`Finished goods turnover`)
IndCredData$`WIP turnover`=as.numeric(IndCredData$`WIP turnover`)
IndCredData$`Raw material turnover`=as.numeric(IndCredData$`Raw material turnover`)
IndCredData$`Shares outstanding`=as.numeric(IndCredData$`Shares outstanding`)
IndCredData$`Equity face value`=as.numeric(IndCredData$`Equity face value`)
IndCredData$`PE on BSE`=as.numeric(IndCredData$`PE on BSE`)
```

New/Complex

The following variables are created variables available.

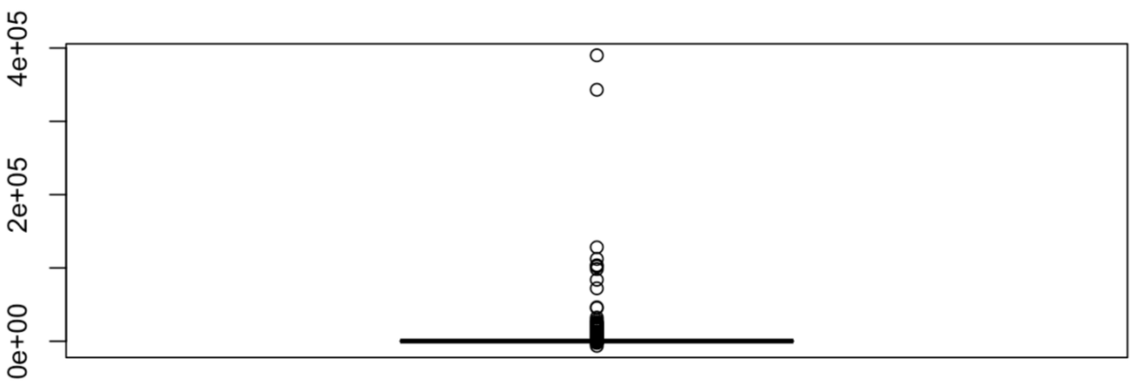
1. PBDITA / Total income
2. PBT / Total income
3. PAT / Total income
4. Cash Profit / Total income
5. PAT / Net worth
6. Total term liabilities / tangible net worth
7. Contingent liabilities / Net worth (%)

Identifying Right variables to fall in the Mix of desired Categories.

1. Profitability
- Cumulative retained profit
- EPS
2. Leverage
- Total Liabilities
- Current Ratio
3. Liquidity
- Debt to Equity Ratio,
- Current Liabilities and Provisions
4. Size
- Equity Face Value
- Capital Employed

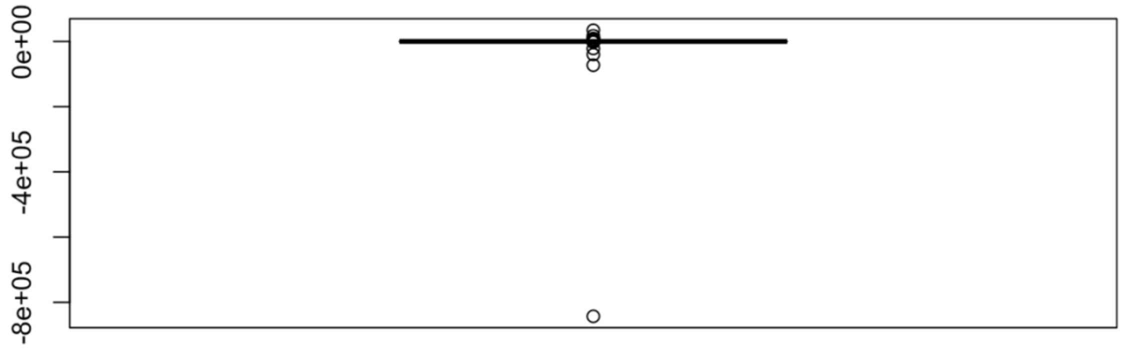
Outlier Treatment

1. boxplot(`Cumulative retained profits`)- ----



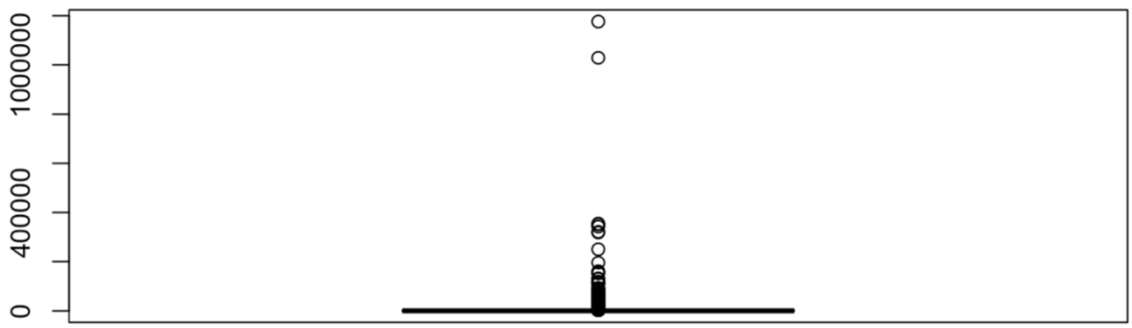
- Shows Outliers in the upper region
- Does not need to be treated.

2. EPS



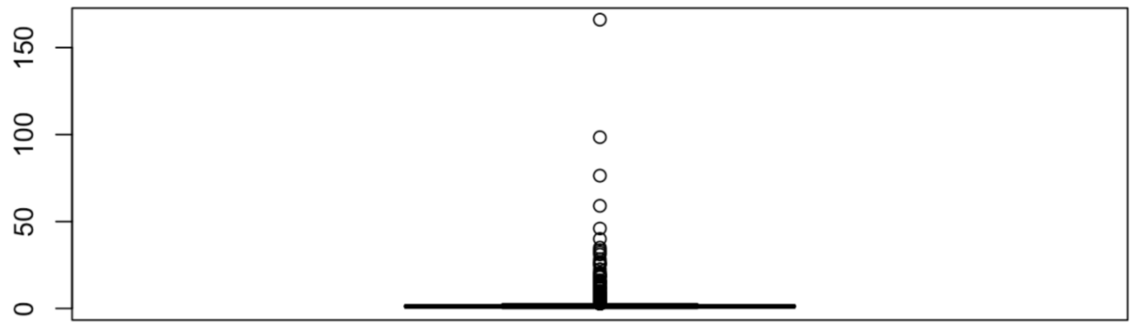
- Shows Outliers in the lower region
- Does not need to be treated.

3. Total liabilities



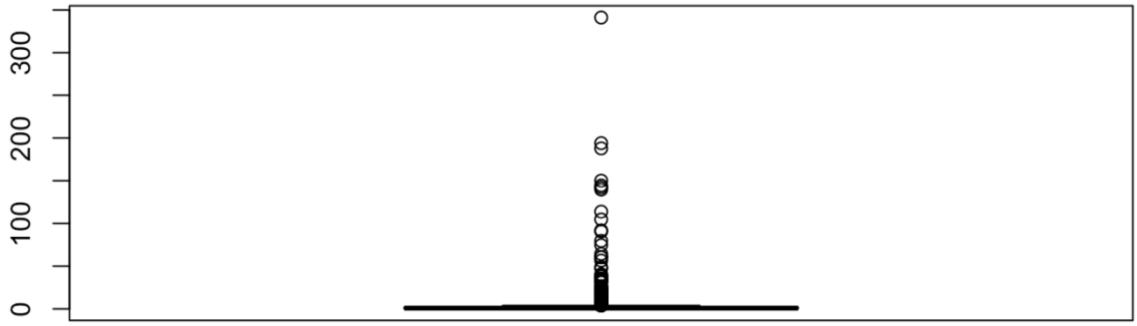
- Shows Outliers in the upper region
- Does not need to be treated.

4. Current ratio (times)



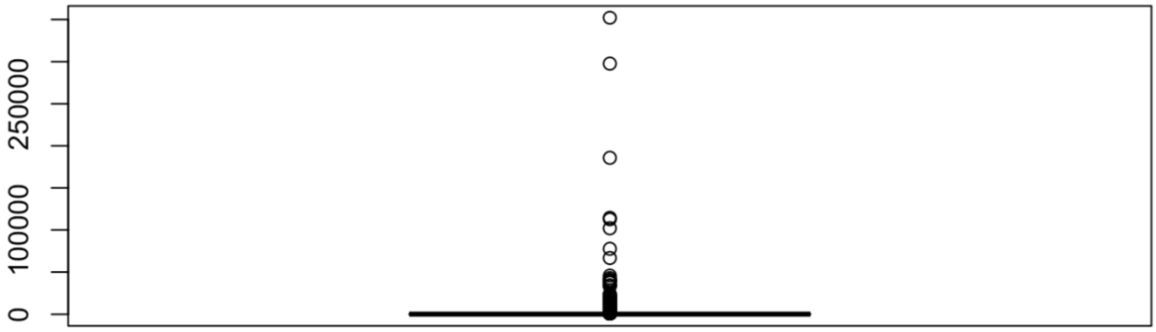
- Shows Outliers in the upper region
- Does not need to be treated.

5. Debt to equity ratio (times)



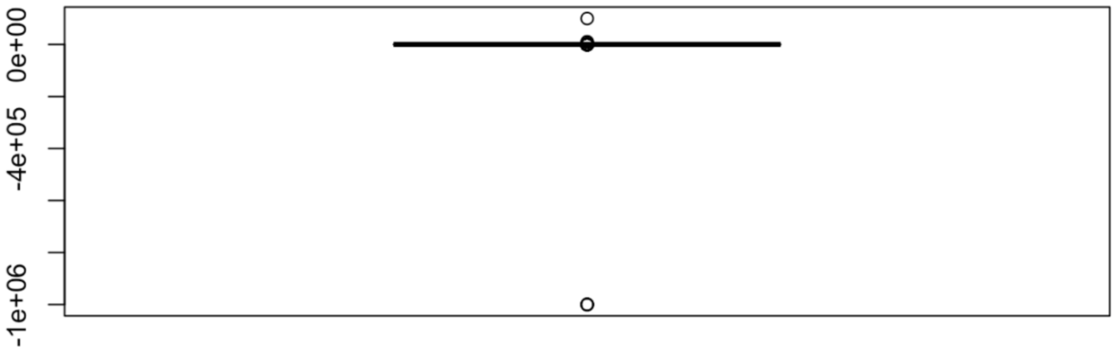
- Shows Outliers in the upper region
- Does not need to be treated.

6. Current liabilities & provisions



- Shows Outliers in the upper region
- Does not need to be treated.

7. Equity face value



- Shows Outliers in the lower region
- Does not need to be treated.

Checking for Missing Values and treating them

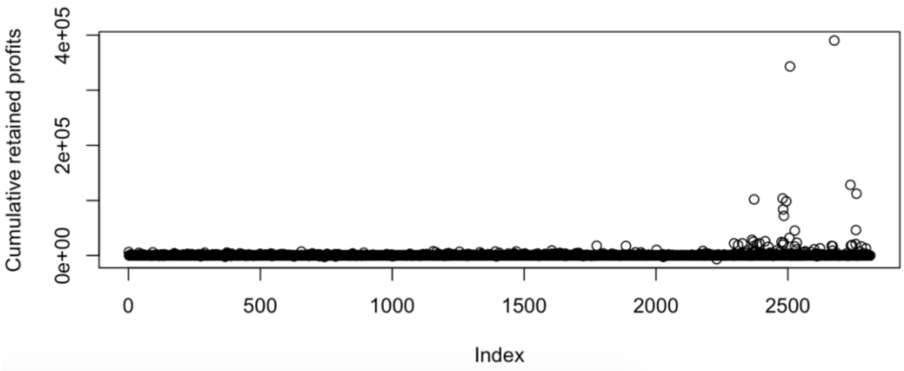
```
> sum(is.na(SelCreData))
[1] 919
> colSums(is.na(SelCreData))
               Networth Next Year Current liabilities & provisions
                                0                               96
         Cumulative retained profits                Capital employed
                               38                                0
             Current ratio (times)    Debt to equity ratio (times)
                               93                                0
                 Equity face value                             EPS
                              692                                0
                 Total liabilities
                                0
> NoNAData=na.omit(SelCreData)
> colSums(is.na(IndCredData))
```

	Networth Next Year	Current liabilities & provisions	Cumulative retained profits	Capital employed	Current ratio (times)	Debt to equity ratio (times)	Equity face value	EPS	To lial
1	8890.6	9965.9	6263.3	7108.1	1.37	0.00	10.0	35.52	17
2	394.3	210.0	247.4	624.0	1.56	0.78	10.0	9.97	94
3	92.2	96.8	32.4	136.0	1.55	0.35	10.0	-0.50	23
4	109.0	112.8	82.7	300.7	2.54	1.79	10.0	7.91	47
5	688.6	555.9	317.7	1415.3	1.27	1.09	10.0	30.57	24
6	291.5	75.1	173.8	462.7	1.58	0.94	10.0	12.69	57
7	-7.3	2.3	-30.4	81.0	0.50	3.13	10.0	-0.48	88
8	93.3	33.4	5.3	126.2	3.73	0.46	10.0	0.42	15
9	2371.3	1165.0	1986.7	4288.8	1.00	1.10	10.0	12.16	56
10	2164.4	1161.2	1881.0	1904.7	0.55	0.01	10.0	66.92	31
11	-7.4	1.4	-15.5	45.4	2.11	10.64	10.0	-2.10	46
12	481.3	58.7	161.6	891.2	2.09	2.66	10.0	19.80	95
13	261.5	94.5	191.2	528.7	0.78	1.37	10.0	20.57	66
14	58.4	11.5	39.3	101.9	1.07	0.90	10.0	1.29	14
15	159.2	31.5	100.4	201.2	1.01	0.45	10.0	6.43	24
16	73.9	58.5	70.6	87.1	0.78	0.21	10.0	-4.53	16
17	1083.8	3663.9	412.8	3386.6	1.18	2.35	10.0	-0.52	73

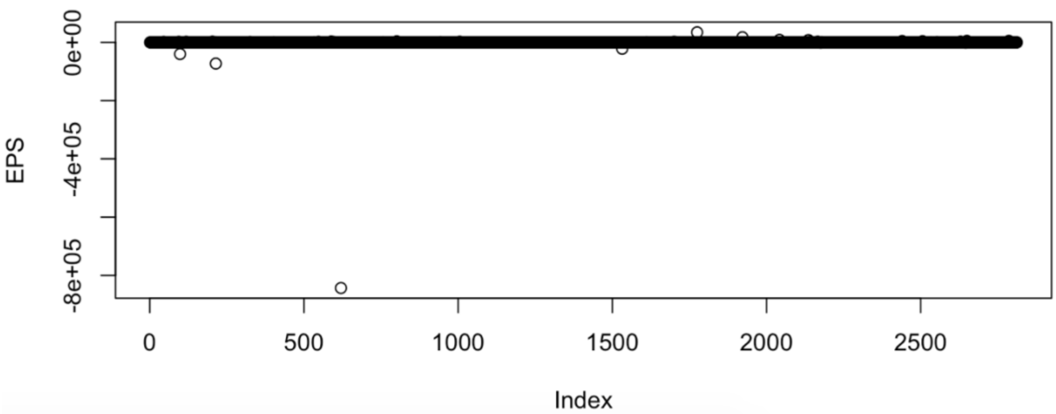
Univariate Analysis

1.

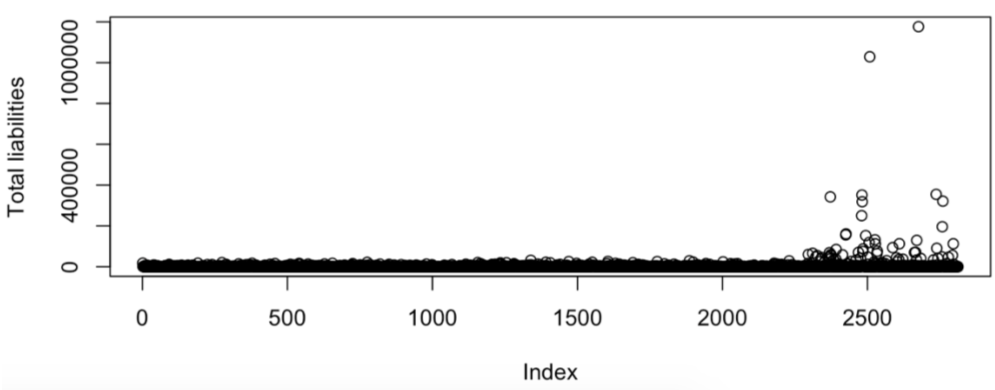
```
summary(`Cumulative retained profits`)
   Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-6534.3      5.6     59.5   1099.7    297.6 390133.8
```

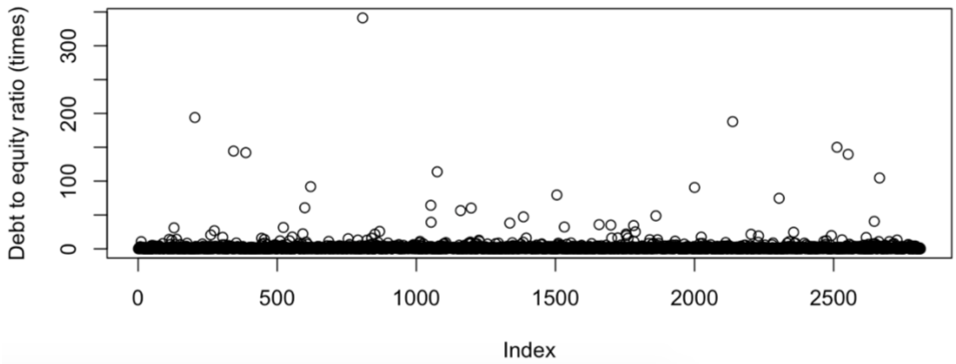


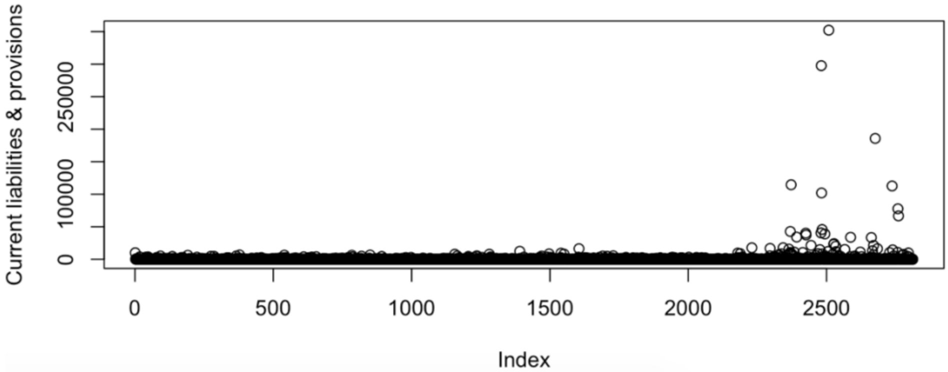

```
> summary(EPS)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-843181.8      0.2       3.4    -277.4      13.5   34522.5
```

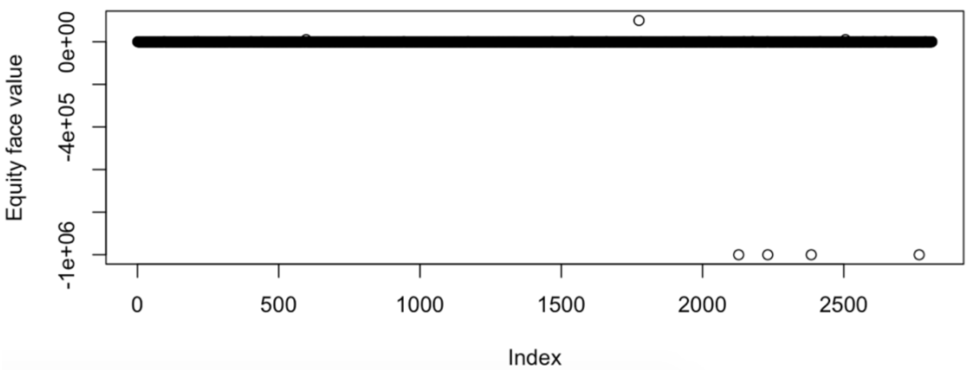


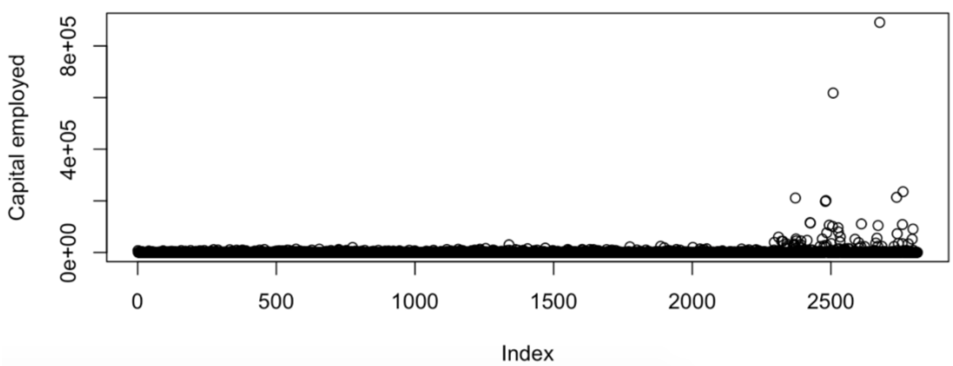

```
summary(`Total liabilities`)
   Min.  1st Qu.   Median     Mean  3rd Qu.      Max.
    0.5    147.1    444.1   4244.7   1455.7 1176509.2
```



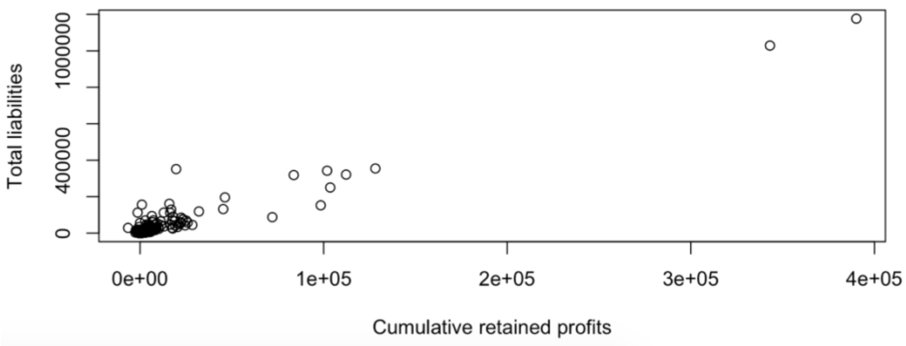

Debt to equity ratio (times)
0
100
200
300
0
500
1000
1500
2000
2500
Index

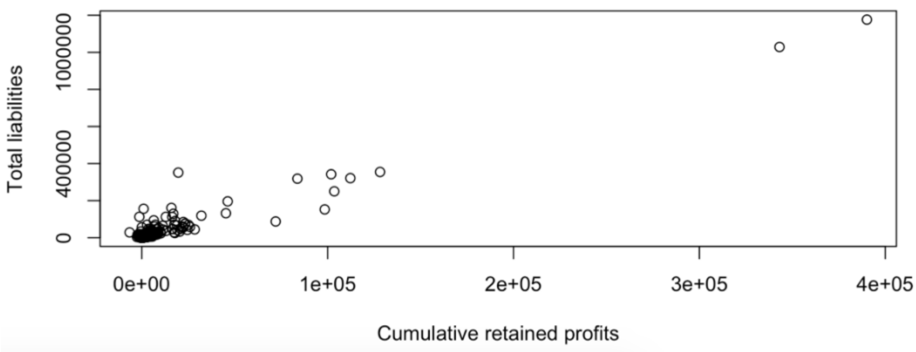

Current liabilities & provisions
0
100000
250000
0
500
1000
1500
2000
2500
Index

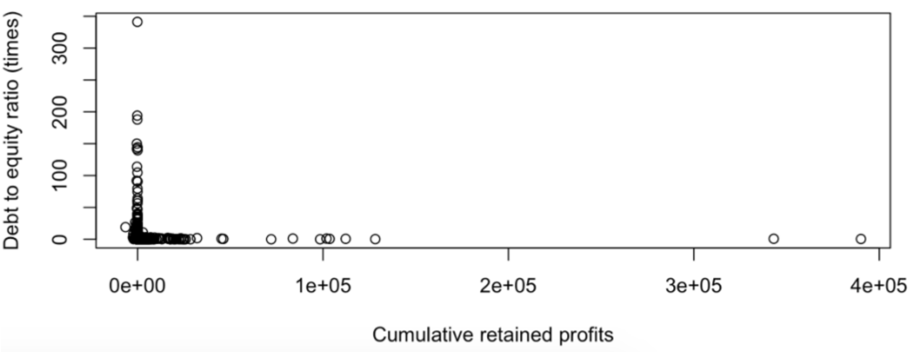

Equity face value
0e+00
-4e+05
-1e+06
0
500
1000
1500
2000
2500
Index

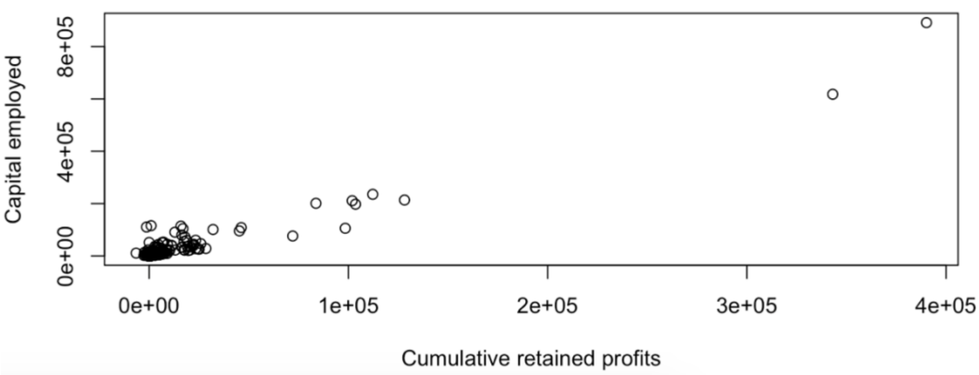

Capital employed
0e+00
4e+05
8e+05
0
500
1000
1500
2000
2500
Index

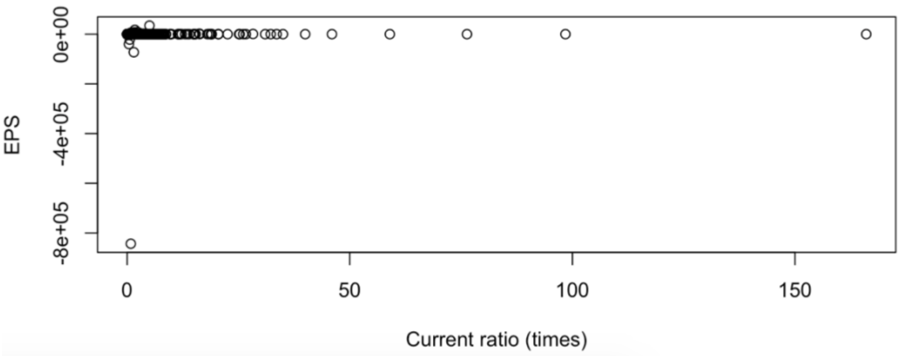
Bivariate Analysis

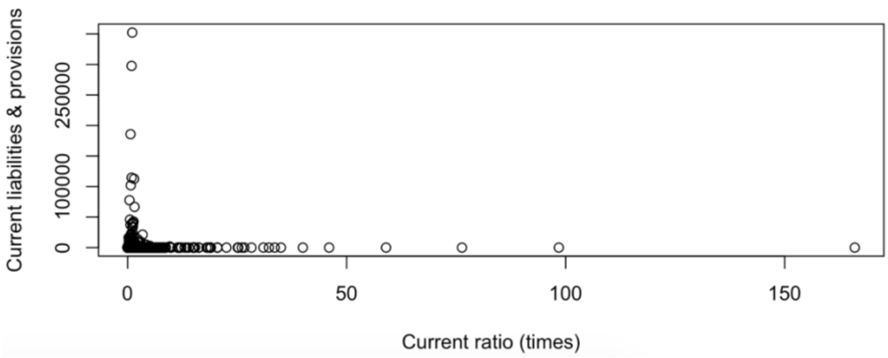


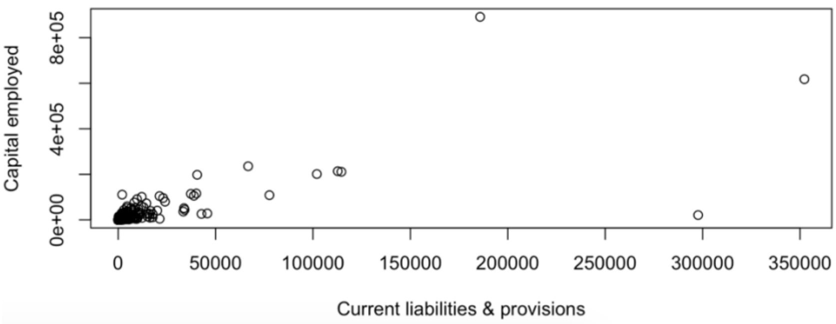


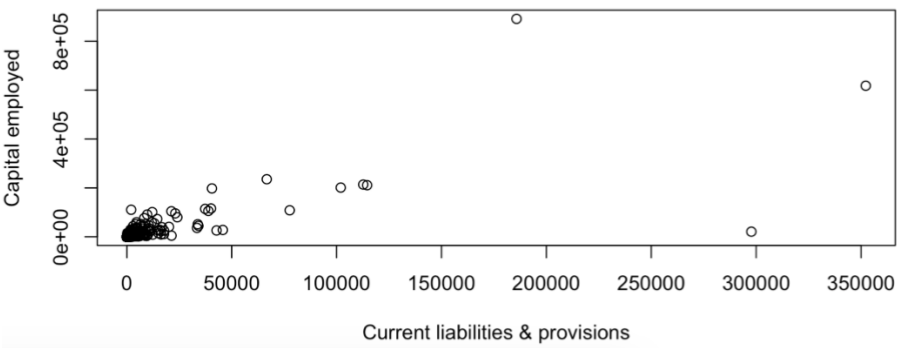





EPS
0e+00
-4e+05
-8e+05
0
50
100
150
Current ratio (times)


Current liabilities & provisions
0
100000
250000
0
50
100
150
Current ratio (times)


Capital employed
0e+00
4e+05
8e+05
0
50000
100000
150000
200000
250000
300000
350000
Current liabilities & provisions


Capital employed
0e+00
4e+05
8e+05
0
50000
100000
150000
200000
250000
300000
350000
Current liabilities & provisions

Creating New Binary Variable “Default”

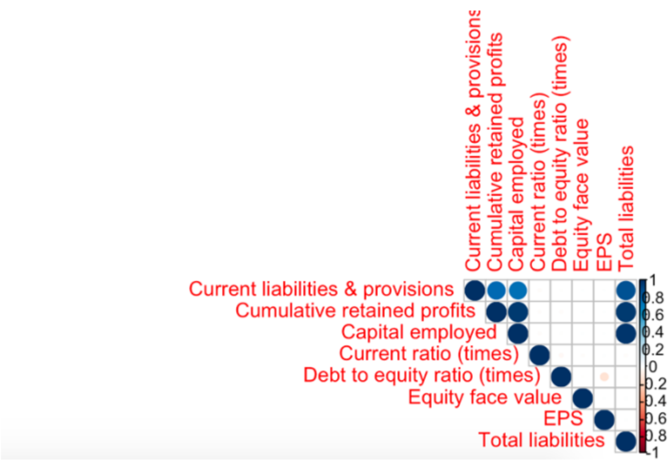
```
#Creating new binary variable "Default"
Default=ifelse(NoNAData$`Networth Next Year`>0, 0,1)
Default=as.factor(Default)
summary(Default)
prop.table(table(Default))

FinalData=cbind(NoNAData[ ,-1],Default)
attach(FinalData)
```

Checking Multicollinearity and treating it

The below process is followed to check multicollinearity.

1. See Correlation Values
2. Create a linear regression model with all variables.
3. Check the variable inflation factor.
4. Check the summary and identify variables showing considerable inflation.

```
> cor(FinalData[ ,-9])
                                Current liabilities & provisions Cumulative retained profits Capital employed Current ratio (times)
Current liabilities & provisions                    1.0000000000                 0.784995066      0.752523440          -0.016887963
Cumulative retained profits                         0.7849950665                 1.000000000      0.966178855          -0.011402111
Capital employed                                    0.7525234397                 0.966178855      1.000000000          -0.013706759
Current ratio (times)                              -0.0168879628                -0.011402111     -0.013706759           1.000000000
Debt to equity ratio (times)                       -0.0091106717                -0.016728541     -0.007058979          -0.030676179
Equity face value                                   0.0038649852                 0.005175393      0.005557976          -0.014440056
EPS                                                 0.0009238646                 0.004635809      0.001610825           0.005133951
Total liabilities                                   0.8750526108                 0.962747079      0.976948297          -0.015534586
                                Debt to equity ratio (times) Equity face value          EPS Total liabilities
Current liabilities & provisions                -0.009110672       0.003864985  0.0009238646       0.875052611
Cumulative retained profits                     -0.016728541       0.005175393  0.0046358086       0.962747079
Capital employed                                -0.007058979       0.005557976  0.0016108250       0.976948297
Current ratio (times)                           -0.030676179      -0.014440056  0.0051339514      -0.015534586
Debt to equity ratio (times)                     1.000000000      -0.001423964 -0.1493647288      -0.007969745
Equity face value                               -0.001423964       1.000000000  0.0014050103       0.005159252
EPS                                             -0.149364729       0.001405010  1.0000000000       0.001544812
Total liabilities                               -0.007969745       0.005159252  0.0015448124       1.000000000
> colSums(is.na(IndCredData))
```



```
glm(formula = Default ~ `Current liabilities & provisions` +
    `Cumulative retained profits` + `Capital employed` + `Current ratio (times)` +
    `Debt to equity ratio (times)` + `Equity face value` + EPS +
    `Total liabilities`, family = binomial("logit"), data = FinalData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.2436  -0.3668  -0.3188  -0.1611   3.4210

Coefficients:
                                    Estimate Std. Error z value Pr(>|z|)
(Intercept)                       -2.507e+00  1.344e-01 -18.655  < 2e-16 ***
`Current liabilities & provisions` -2.230e-04  6.191e-04  -0.360   0.7187
`Cumulative retained profits`     -3.289e-03  4.597e-04  -7.156 8.33e-13 ***
`Capital employed`                 3.020e-05  4.680e-04   0.065   0.9486
`Current ratio (times)`           -8.973e-02  6.171e-02  -1.454   0.1460
`Debt to equity ratio (times)`     3.392e-02  7.169e-03   4.731 2.23e-06 ***
`Equity face value`               -1.752e-06  1.179e-06  -1.486   0.1374
EPS                               -4.598e-05  2.424e-05  -1.897   0.0579 .
`Total liabilities`               -2.856e-05  4.689e-04  -0.061   0.9514
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1261.5  on 2811  degrees of freedom
Residual deviance: 1041.7  on 2803  degrees of freedom
AIC: 1059.7

Number of Fisher Scoring iterations: 12

>
> vif(model1)
`Current liabilities & provisions`      `Cumulative retained profits`                 `Capital employed`
                          9.828362                           1.283809                         486.504940
           `Current ratio (times)`     `Debt to equity ratio (times)`                 `Equity face value`
                          1.019725                           1.032200                           1.006816
                               EPS                `Total liabilities`
                          1.001057                         545.403846
> colSums(is.na(IndCredData))
```

Logistic Regression

One of the best models for predictions when the variable to be determined is binary in nature is the logistic regression model.
Like all regression analyses, the logistic regression is a predictive analysis. Logistic regression 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.

```
#Bulding first logistic regression model.

model1=glm(Default~`Current liabilities & provisions`+`Cumulative retained profits`+
             `Capital employed`+`Current ratio (times)`+`Debt to equity ratio (times)`+`Equity face value`+EPS+
             `Total liabilities`,data=FinalData,family=binomial("logit"))
summary(model1)

vif(model1)
```

```
glm(formula = Default ~ `Current liabilities & provisions` +
    `Cumulative retained profits` + `Capital employed` + `Current ratio (times)` +
    `Debt to equity ratio (times)` + `Equity face value` + EPS +
    `Total liabilities`, family = binomial("logit"), data = FinalData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.2436  -0.3668  -0.3188  -0.1611   3.4210

Coefficients:
                                     Estimate Std. Error z value Pr(>|z|)
(Intercept)                        -2.507e+00  1.344e-01 -18.655  < 2e-16 ***
`Current liabilities & provisions` -2.230e-04  6.191e-04  -0.360   0.7187
`Cumulative retained profits`      -3.289e-03  4.597e-04  -7.156 8.33e-13 ***
`Capital employed`                  3.020e-05  4.680e-04   0.065   0.9486
`Current ratio (times)`            -8.973e-02  6.171e-02  -1.454   0.1460
`Debt to equity ratio (times)`      3.392e-02  7.169e-03   4.731 2.23e-06 ***
`Equity face value`                -1.752e-06  1.179e-06  -1.486   0.1374
EPS                                -4.598e-05  2.424e-05  -1.897   0.0579 .
`Total liabilities`                -2.856e-05  4.689e-04  -0.061   0.9514
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1261.5  on 2811  degrees of freedom
Residual deviance: 1041.7  on 2803  degrees of freedom
AIC: 1059.7

Number of Fisher Scoring iterations: 12

>
> vif(model1)
`Current liabilities & provisions`      `Cumulative retained profits`              `Capital employed`
                          9.828362                           1.283809                      486.504940
           `Current ratio (times)`     `Debt to equity ratio (times)`             `Equity face value`
                          1.019725                           1.032200                        1.006816
                               EPS                `Total liabilities`
                          1.001057                         545.403846
```

Analysing Coefficients and Signs

1. Intercept
   - The intercept has a negative value. Shows the major portion of the data has negative implication in causing default.
   - The intercept has high P value. Says it is very Significant
2. Current Liabilities and Provisions
   - Has negative value. Negative Correlation to the determinant variable.
   - The P value shows less significance.

3. Cumulative retained profits
   - Has negative value. Negative Correlation.
   - The P value shows high significance.

4. Capital employed
   - Has positive value. Positive Correlation.
   - The P value shows low significance.

5. Current ratio (times)`
   - Has negative value. Negative Correlation.
   - The P value shows low significance.

6. Debt to equity ratio (times)
   - Has positive value. Positive Correlation.
   - The P value shows high significance.

7. Equity face value
   - Has negative value. Negative Correlation.
   - The P value shows low significance.

8. Total liabilities
   - Has negative value. Negative Correlation.
   - The P value shows low significance.

Building Final Logistics Regression Model

```
glm(formula = Default ~ `Cumulative retained profits` + `Current ratio (times)` +
    `Debt to equity ratio (times)` + `Equity face value` + EPS,
    family = binomial("logit"), data = FinalData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.3286  -0.3615  -0.3196  -0.1752   3.3437

Coefficients:
                                 Estimate Std. Error z value Pr(>|z|)
(Intercept)                    -2.565e+00  1.256e-01 -20.416  < 2e-16 ***
`Cumulative retained profits`  -3.095e-03  4.076e-04  -7.593 3.12e-14 ***
`Current ratio (times)`        -8.300e-02  6.021e-02  -1.379   0.1680
`Debt to equity ratio (times)`  3.415e-02  7.167e-03   4.765 1.89e-06 ***
`Equity face value`            -1.749e-06  1.176e-06  -1.486   0.1372
EPS                            -4.672e-05  2.432e-05  -1.921   0.0548 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1261.5  on 2811  degrees of freedom
Residual deviance: 1043.5  on 2806  degrees of freedom
AIC: 1055.5

Number of Fisher Scoring iterations: 10

> vif(model2)
 `Cumulative retained profits`        `Current ratio (times)` `Debt to equity ratio (times)`
                      1.032692                       1.011915                       1.030847
           `Equity face value`                            EPS
                      1.006255                       1.000481
> -6534.3      5.6      59.5    1099.7     297.6 390133.8
```

```
#Preparing Validation data

ValiData=read_xlsx("Validationdata.xlsx")
ValiData$`Equity face value`=as.numeric(ValiData$`Equity face value`)


# Checking Default Ratio between Training Data and Validation Data
prop.table(table(Default))
prop.table(table(ValiData$`Default - 1`))
```

Predict accuracy of model on dev and validation datasets

```
# Checking Default Ratio between Training Data and Validation Data
prop.table(table(Default))
prop.table(table(ValiData$`Default - 1`))

#Model Accuracy
predTest=predict(model2,newdata = ValiData,type = "response")

#Creation of confusion matrix to assess model performance measures
tab.LR=table(ValiData$`Default - 1`,predTest > 0.5)
tab.LR
sum(diag(tab.LR))/sum(tab.LR)
```

```
> # Checking Default Ratio between Training Data and Validation Data
> prop.table(table(Default))
Default
         0          1
0.94096728 0.05903272
> prop.table(table(ValiData$`Default - 1`))

         0          1
0.92447552 0.07552448
>
> #Model Accuracy
> predTest=predict(model2,newdata = ValiData,type = "response")
>
> #Creation of confusion matrix to assess model performance measures
> tab.LR=table(ValiData$`Default - 1`,predTest > 0.5)
> tab.LR

    FALSE TRUE
  0   548    6
  1    34    2
> sum(diag(tab.LR))/sum(tab.LR)
[1] 0.9322034
> -6534.3       5.6      59.5    1099.7     297.6 390133.8
```

- The data shows the modal has a very high accuracy.

Sort the data in descending order based on probability of default and then divide into 10 dociles based on probability & check how well the model has performed