Mini Project - FRA

Project Objective: The objective of the report is to develop a Credit Risk default model using raw-data.xlsx dataset in R and prepare a Managerial Report by explaining the following points.

This Managerial report consists of the following:

- Business Objective Statement
- Perform Exploratory data analysis(EDA)
- Outlier treatment
- Missing Value treatment
- New variables creation(Profitability, Liquidity, Leverage and Company's size)
- Check for multicollinearity
- Univariate and bivariate analysis
- Building Logistic Regression model
- Model Performance Measures

The sample Cars dataset is:

Num	Networth Next Year	Total asse	Net worth	Total income	Change in stock	Total expenses	Profit afte	PBDITA	PBT	Cash profi	PBDITA as	PBT as % c	PAT as % c	Cash profi	PAT as % c	Sales	Ir
1	8890.6	17512.3	7093.2	24965.2	235.8	23657.8	1543.2	2860.2	2417.2	1872.8	11.46	9.68	6.18	7.5	23.78	24458	
2	394.3	941	351.5	1527.4	42.7	1454.9	115.2	283	188.4	158.6	18.53	12.33	7.54	10.38	38.08	1504.3	
3	92.2	232.8	100.6	477.3	-5.2	478.7	-6.6	5.8	-6.6	0.3	1.22	-1.38	-1.38	0.06	-6.35	475.6	
4	2.7	2.7	2.7								0	0	0	0	0		
5	109	478.5	107.6	1580.5	-17	1558	5.5	31	6.3	11.9	1.96	0.4	0.35	0.75	5.25	1575.1	
6	688.6	2434.4	675.8	2648.6	62.3	2636.4	74.5	200.1	74.5	146.9	7.55	2.81	2.81	5.55	21.78	2639.5	
7	246	327.1	245.2								0	0	0	0	0		
8	13.7	80	12.7	153.6	-8.5	144	1.1	9.7	2	4	6.32	1.3	0.72	2.6	9.95	153.4	
9	291.5	573.8	238.6	582.6	31	565.3	48.3	110.1	68.5	82.6	18.9	11.76	8.29	14.18	21.65	573.9	
10	-7.3	88.6	19.6	83.4	-6.7	79.1	-2.4	0.3	-14.4	-10.5	0.36	-17.27	-2.88	-12.59	-11.57	83.4	
11	93.3	159.6	86.7	172.9	0.1	169.7	3.3	18.4	3.7	12.6	10.64	2.14	1.91	7.29	3.88	172.1	
12	2371.3	5669.1	2039.6	6910.1	74	6518.7	465.4	960.1	605.1	704.1	13.89	8.76	6.74	10.19	24.74	6834	
13	2164.4	3142.5	1882.8	5303.5	-36.1	4244.9	1022.5	1442.6	1380.9	1067.6	27.2	26.04	19.28	20.13	61.08	4740.1	
14	-7.4	46.8	3.9	11.8		13.5	-1.7	0	-1.7	0	0	-14.41	-14.41	0	0	11.6	
15	43.6	84.8	42.1	152.1	-6.5	145	0.6	8	1.2	3.6	5.26	0.79	0.39	2.37	0	151.9	
16	481.3	954.3	243.8	589	56.5	527.2	118.3	201.5	117.6	164.1	34.21	19.97	20.08	27.86	56.27	585.9	
17	261.5	660.1	223.4	1623.1	20	1577.2	65.9	128.6	83.5	74.2	7.92	5.14	4.06	4.57	34.59	1616	
18	58.4	141.1	53.7	101.9	0.2	97.7	4.4	16.5	5.1	8	16.19	5	4.32	7.85	8.54	101.6	
19	-2.6	120	6.2	13.9	-2.8	20.4	-9.3	-1.8	-9	-7.3	-12.95	-64.75	-66.91	-52.52	-85.71	13.9	
20	159.2	245.6	138.3	219.6	0.5	203.9	16.2	37.8	24.6	25.2	17.21	11.2	7.38	11.48	12.19	210.2	
21	73.9	166.5	71.8	68.4	7	79.8	-4.4	-3	-8.6	-5.7	-4.39	-12.57	-6.43	-8.33	-5.95	65	
22	1083.8	7394.3	1010	20952.3	-796.3	20143.7	12.3	490.5	19.4	112.9	2.34	0.09	0.06	0.54	1.19	20682.1	
23	433.8	1199.3	393.8	1188.2	132.5	1284.6	36.1	103.1	47	57.3	8.68	3.96	3.04	4.82	9.51	1158.7	
24	146.9	269.3	134.1	248	4.7	243	9.7	48.4	28.1	38	19.52	11.33	3.91	15.32	7.26	247.5	

raw data

Business Objective: Identify the companies which can default their credit in the given data set.

Exploratory Data Analysis(EDA):

EDA has been performed using R code.

Please refer Appendix A for Source Code.

The summary of the data is

```
> summary(companydefault)
      Num        Networth Next Year  Total assets          Net worth          Total income       Change in stock
 Min.   :   1   Min.   :-74265.6   Min.   :     0.1   Min.   :     0.0   Min.   :      0.0   Min.   :-3029.40
 1st Qu.: 886   1st Qu.:    31.7   1st Qu.:    91.3   1st Qu.:    31.3   1st Qu.:    106.5   1st Qu.:   -1.80
 Median :1773   Median :   116.3   Median :   309.7   Median :   102.3   Median :    444.9   Median :    1.60
 Mean   :1772   Mean   :  1616.3   Mean   :  3443.4   Mean   :  1295.9   Mean   :   4582.8   Mean   :   41.49
 3rd Qu.:2658   3rd Qu.:   456.1   3rd Qu.:  1098.7   3rd Qu.:   377.3   3rd Qu.:   1440.9   3rd Qu.:   18.05
 Max.   :3545   Max.   :805773.4   Max.   :1176509.2   Max.   :613151.6   Max.   :2442828.2   Max.   :14185.50
                                                                         NA's   :198         NA's   :458
 Total expenses      Profit after tax        PBDITA              PBT             Cash profit
 Min.   :     -0.1   Min.   : -3908.30   Min.   :  -440.7   Min.   : -3894.80   Min.   : -2245.70
 1st Qu.:     95.8   1st Qu.:     0.50   1st Qu.:     6.9   1st Qu.:     0.70   1st Qu.:     2.90
 Median :    407.7   Median :     8.80   Median :    35.4   Median :    12.40   Median :    18.85
 Mean   :   4262.9   Mean   :   277.36   Mean   :   578.1   Mean   :   383.81   Mean   :   392.07
 3rd Qu.:   1359.8   3rd Qu.:    52.27   3rd Qu.:   150.2   3rd Qu.:    71.97   3rd Qu.:    93.20
 Max.   :2366035.3   Max.   :119439.10   Max.   :208576.5   Max.   :145292.60   Max.   :176911.80
 NA's   :139         NA's   :131         NA's   :131        NA's   :131         NA's   :131
 PBDITA as % of total income PBT as % of total income PAT as % of total income Cash profit as % of total income
 Min.   :-6400.000           Min.   :-21340.00        Min.   :-21340.00        Min.   :-15020.000
 1st Qu.:    5.000           1st Qu.:     0.55        1st Qu.:     0.35        1st Qu.:     2.020
 Median :    9.660           Median :     3.31        Median :     2.34        Median :     5.640
 Mean   :    4.571           Mean   :   -17.28        Mean   :   -19.20        Mean   :    -8.229
 3rd Qu.:   16.390           3rd Qu.:     8.80        3rd Qu.:     6.34        3rd Qu.:    10.700
 Max.   :  100.000           Max.   :   100.00        Max.   :   150.00        Max.   :   100.000
 NA's   :68                  NA's   :68               NA's   :68               NA's   :68
 PAT as % of net worth     Sales           Income from financial services  Other income      Total capital
 Min.   :-748.72       Min.   :      0.1   Min.   :    0.00               Min.   :     0.00   Min.   :     0.1
 1st Qu.:   0.00       1st Qu.:    112.7   1st Qu.:    0.40               1st Qu.:     0.40   1st Qu.:    13.1
 Median :   7.92       Median :    453.1   Median :    1.80               Median :     1.40   Median :    42.1
 Mean   :  10.27       Mean   :   4549.5   Mean   :   80.84               Mean   :    41.36   Mean   :   216.6
 3rd Qu.:  20.19       3rd Qu.:   1433.5   3rd Qu.:    9.68               3rd Qu.:     5.97   3rd Qu.:   100.3
 Max.   :2466.67       Max.   :2384984.4   Max.   :51938.20               Max.   :42856.70   Max.   :78273.2
                       NA's   :259         NA's   :935                    NA's   :1295       NA's   :4
```

```
Reserves and funds   Borrowings         Current liabilities & provisions Deferred tax liability Shareholders funds
Min.    : -6525.9   Min.    :     0.10   Min.    :     0.1               Min.    :     0.1       Min.    :     0.0
1st Qu.:     5.0    1st Qu.:    23.95   1st Qu.:    17.8               1st Qu.:     3.2       1st Qu.:    32.0
Median :    54.8    Median :    99.20   Median :    69.4               Median :    13.4       Median :   105.6
Mean   :  1163.8    Mean   :  1122.28   Mean   :   940.6               Mean   :   227.2       Mean   :  1322.1
3rd Qu.:   277.3    3rd Qu.:   352.60   3rd Qu.:   261.7               3rd Qu.:    50.0       3rd Qu.:   393.2
Max.   :625137.8    Max.   :278257.30   Max.   :352240.3               Max.   :72796.6       Max.   :613151.6
NA's   :85          NA's   :366         NA's   :96                     NA's   :1140
Cumulative retained profits Capital employed      TOL/TNW          Total term liabilities / tangible net worth
Min.    : -6534.3           Min.    :     0.0   Min.    :-350.480   Min.    :-325.600
1st Qu.:     1.1            1st Qu.:    60.8   1st Qu.:    0.600   1st Qu.:    0.050
Median :    37.1            Median :   214.7   Median :    1.430   Median :    0.340
Mean   :   890.5            Mean   :  2328.3   Mean   :    3.994   Mean   :    1.844
3rd Qu.:   202.3            3rd Qu.:   767.3   3rd Qu.:    2.830   3rd Qu.:    1.000
Max.   :390133.8            Max.   :891408.9   Max.   :  473.000   Max.   :  456.000
NA's   :38
Contingent liabilities / Net worth (%) Contingent liabilities Net fixed assets    Investments
Min.    :     0.00                     Min.    :     0.1       Min.    :     0.0   Min.    :     0.00
1st Qu.:     0.00                     1st Qu.:     6.3       1st Qu.:    26.0   1st Qu.:     1.00
Median :     5.33                     Median :    38.0       Median :    93.5   Median :     8.35
Mean   :    53.94                     Mean   :   932.9       Mean   :  1189.7   Mean   :   694.73
3rd Qu.:    30.76                     3rd Qu.:   192.7       3rd Qu.:   344.9   3rd Qu.:    64.30
Max.   :14704.27                      Max.   :559506.8       Max.   :636604.6   Max.   :199978.60
                                      NA's   :1188           NA's   :118        NA's   :1435
Current assets     Net working capital Quick ratio (times) Current ratio (times) Debt to equity ratio (times)
Min.    :     0.1  Min.    :-63839.0   Min.    :  0.000   Min.    :  0.00       Min.    :  0.00
1st Qu.:    36.2  1st Qu.:    -1.1     1st Qu.:  0.410   1st Qu.:  0.93       1st Qu.:  0.22
Median :   145.1  Median :    16.2     Median :  0.670   Median :  1.23       Median :  0.79
Mean   :  1293.4  Mean   :   138.6     Mean   :  1.401   Mean   :  2.13       Mean   :  2.78
3rd Qu.:   502.2  3rd Qu.:    84.2     3rd Qu.:  1.030   3rd Qu.:  1.71       3rd Qu.:  1.75
Max.   :354815.2  Max.   : 85782.8     Max.   :341.000   Max.   :505.00       Max.   :456.00
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                    Length:3541        Length:3541
1st Qu.:  0.0200                    1st Qu.:     2.79                    Class :character   Class :character
Median :  0.0700                    Median :     8.03                    Mode  :character   Mode  :character
Mean   :  0.4904                    Mean   :   158.44
3rd Qu.:  0.1900                    3rd Qu.:    21.79
Max.   :165.0000                    Max.   :128040.76
NA's   :93                          NA's   :85
Finished goods turnover WIP turnover       Raw material turnover Shares outstanding Equity face value
Length:3541             Length:3541        Length:3541           Length:3541        Length:3541
Class :character        Class :character   Class :character      Class :character   Class :character
Mode  :character        Mode  :character   Mode  :character      Mode  :character   Mode  :character


     EPS              Adjusted EPS       Total liabilities    PE on BSE
Min.    :-843181.8   Min.    :-843181.8   Min.    :      0.1   Length:3541
1st Qu.:      0.0    1st Qu.:      0.0    1st Qu.:     91.3   Class :character
Median :      1.4    Median :      1.2    Median :    309.7   Mode  :character
Mean   :   -220.3    Mean   :   -221.5    Mean   :   3443.4
3rd Qu.:      9.6    3rd Qu.:      7.5    3rd Qu.:   1098.7
Max.   :  34522.5    Max.   :  34522.5    Max.   :1176509.2
```

The internal structure of the data is:

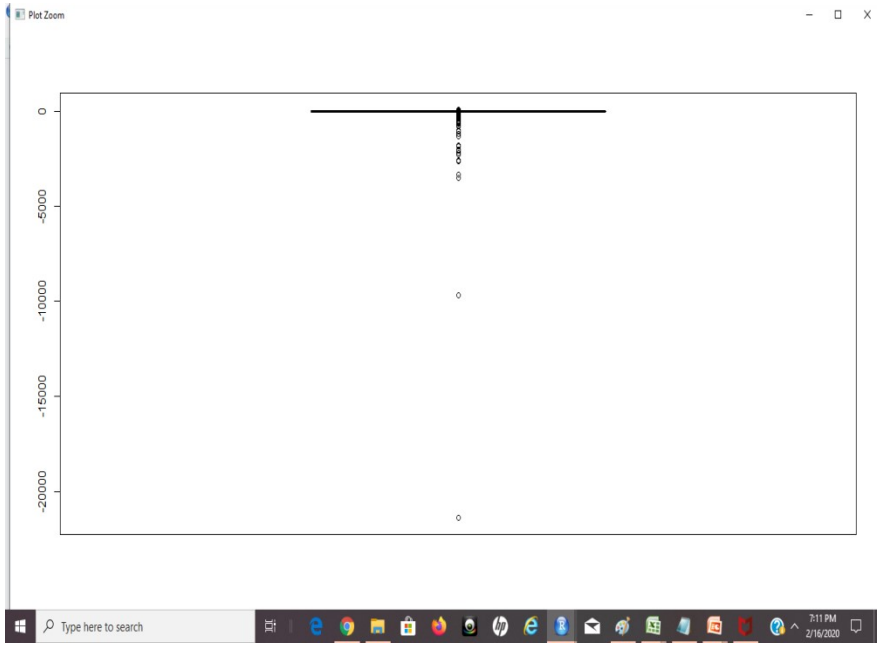
```
Classes 'tbl_df', 'tbl' and 'data.frame':       3541 obs. of  51 variables:
 $ Num                                          : num  1 2 3 4 5 6 7 8 9 10 ...
 $ Networth Next Year                           : num  8890.6 394.3 92.2 2.7 109 ...
 $ Total assets                                 : num  17512.3 941 232.8 2.7 478.5 ...
 $ Net worth                                    : num  7093.2 351.5 100.6 2.7 107.6 ...
 $ Total income                                 : num  24965 1527 477 NA 1580 ...
 $ Change in stock                              : num  235.8 42.7 -5.2 NA -17 ...
 $ Total expenses                               : num  23658 1455 479 NA 1558 ...
 $ Profit after tax                             : num  1543.2 115.2 -6.6 NA 5.5 ...
 $ PBDITA                                       : num  2860.2 283 5.8 NA 31 ...
 $ PBT                                          : num  2417.2 188.4 -6.6 NA 6.3 ...
 $ Cash profit                                  : num  1872.8 158.6 0.3 NA 11.9 ...
 $ PBDITA as % of total income                  : num  11.46 18.53 1.22 0 1.96 ...
 $ PBT as % of total income                     : num  9.68 12.33 -1.38 0 0.4 ...
 $ PAT as % of total income                     : num  6.18 7.54 -1.38 0 0.35 2.81 0 0.72 8.29 -2.88
...
 $ Cash profit as % of total income             : num  7.5 10.38 0.06 0 0.75 ...
 $ PAT as % of net worth                        : num  23.78 38.08 -6.35 0 5.25 ...
 $ Sales                                        : num  24458 1504 476 NA 1575 ...
 $ Income from financial services              : num  158 4 1.5 NA 3.9 6.4 NA NA 7.3 NA ...
 $ Other income                                 : num  297.2 15.9 0.2 NA 0.9 ...
 $ Total capital                                : num  423.8 115.5 81.4 0.5 6.2 ...
 $ Reserves and funds                           : num  6822.8 257.8 19.2 2.2 161.8 ...
 $ Borrowings                                   : num  14.9 272.5 35.4 NA 193.1 ...
 $ Current liabilities & provisions             : num  9965.9 210 96.8 NA 112.8 ...
 $ Deferred tax liability                       : num  284.9 85.2 NA NA 4.6 ...
 $ Shareholders funds                           : num  7093.2 351.5 100.6 2.7 107.6 ...
 $ Cumulative retained profits                  : num  6263.3 247.4 32.4 2.2 82.7 ...
 $ Capital employed                             : num  7108.1 624 136 2.7 300.7 ...
 $ TOL/TNW                                      : num  1.33 1.23 1.44 0 2.83 1.8 0.03 5.17 1.05 3.25
...
 $ Total term liabilities / tangible net worth: num  0 0.34 0.29 0 1.59 0.37 0.03 0.94 0.3 0.54 ...
 $ Contingent liabilities / Net worth (%)       : num  14.8 19.2 45.8 0 34.9 ...
 $ Contingent liabilities                       : num  1049.7 67.6 46.1 NA 37.6 ...
 $ Net fixed assets                             : num  1900.2 286.4 38.7 2.5 94.8 ...
 $ Investments                                  : num  1069.6 2.2 4.3 NA 7.4 ...
 $ Current assets                               : num  13277.5 563.9 167.5 0.2 349.7 ...
 $ Net working capital                          : num  3588.5 203.5 59.6 0.2 215.8 ...
 $ Quick ratio (times)                          : num  1.18 0.95 1.11 NA 1.41 0.48 NA 0.54 0.59 0.39
...
 $ Current ratio (times)                        : num  1.37 1.56 1.55 NA 2.54 1.27 NA 1.15 1.58 0.5 .
..
 $ Debt to equity ratio (times)                 : num  0 0.78 0.35 0 1.79 1.09 0.32 2.31 0.94 3.13 ..
.
 $ Cash to current liabilities (times)          : num  0.43 0.06 0.21 NA 0 0.11 NA 0.04 0.19 0 ...
 $ Cash to average cost of sales per day        : num  68.21 5.96 17.07 NA 0 ...
 $ Creditors turnover                           : chr  "3.62" "9.8000000000000007" "5.28" "0" ...
 $ Debtors turnover                             : chr  "3.85" "5.7" "5.07" "0" ...
 $ Finished goods turnover                      : chr  "200.55" "14.21" "9.24" NA ...
 $ WIP turnover                                 : chr  "21.78" "7.49" "0.23" NA ...
 $ Raw material turnover                        : chr  "7.71" "11.46" NA "0" ...
 $ Shares outstanding                           : chr  "42381675" "11550000" "8149090" "52404" ...
 $ Equity face value                            : chr  "10" "10" "10" "10" ...
 $ EPS                                          : num  35.52 9.97 -0.5 0 7.91 ...
 $ Adjusted EPS                                 : num  7.1 9.97 -0.5 0 7.91 ...
 $ Total liabilities                            : num  17512.3 941 232.8 2.7 478.5 ...
 $ PE on BSE                                    : chr  "27.31" "8.17" "-5.76" "NA" ...
```

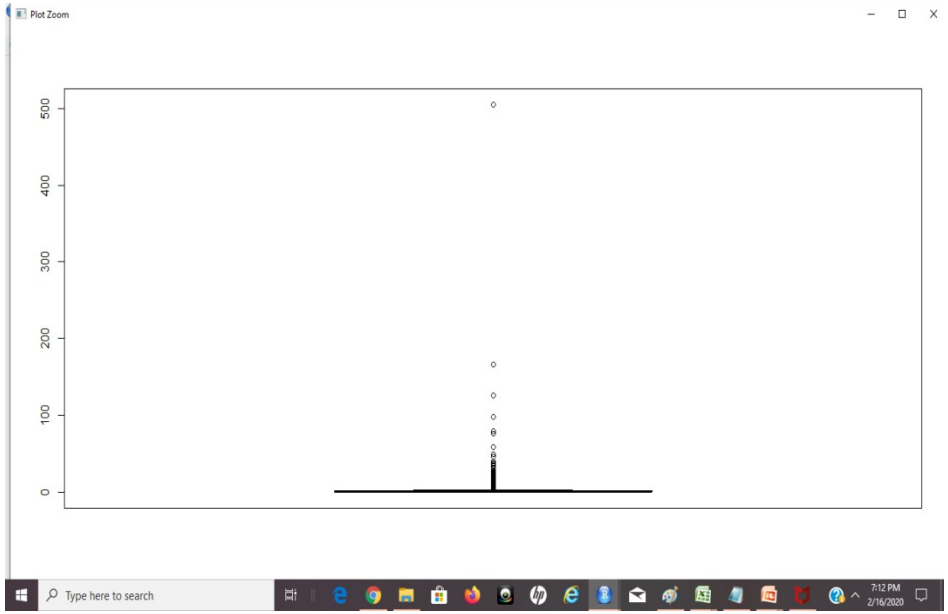
Outlier Treatment:

Box plots can be drawn for the important variables to check if any outliers exists.

Box plot for ‘PBT as % of total income’ is shown below, which does not have considerable outliers.

Similarly box plot of ‘Current ratio (times)` also shows no considerable outliers.





Missing Value Treatment:

Given dataset has many missing values, hence replaced the missing values with 0.

```
> #Check for null values
> sum(is.na(companydefault))
[1] 10007
```

Since the given data has different variable types, converted all the variables to numeric data type.

Internal structure of the data after converting all the variables to numeric type is:

```
 $ Num                                                : num  1 2 3 4 5 6 7 8 9 10 ...
 $ Networth Next Year                                 : num  8890.6 394.3 92.2 2.7 109 ...
 $ Total assets                                       : num  17512.3 941 232.8 2.7 478.5 ...
 $ Net worth                                          : num  7093.2 351.5 100.6 2.7 107.6 ...
 $ Total income                                       : num  24965 1527 477 0 1580 ...
 $ Change in stock                                    : num  235.8 42.7 -5.2 0 -17 ...
 $ Total expenses                                     : num  23658 1455 479 0 1558 ...
 $ Profit after tax                                   : num  1543.2 115.2 -6.6 0 5.5 ...
 $ PBDITA                                             : num  2860.2 283 5.8 0 31 ...
 $ PBT                                                : num  2417.2 188.4 -6.6 0 6.3 ...
 $ Cash profit                                        : num  1872.8 158.6 0.3 0 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 0 1575 ...
 $ Income from financial services                    : num  158 4 1.5 0 3.9 6.4 0 0 7.3 0 ...
 $ Other income                                       : num  297.2 15.9 0.2 0 0.9 ...
 $ Total capital                                      : num  423.8 115.5 81.4 0.5 6.2 ...
 $ Reserves and funds                                 : num  6822.8 257.8 19.2 2.2 161.8 ...
 $ Borrowings                                         : num  14.9 272.5 35.4 0 193.1 ...
 $ Current liabilities & provisions                   : num  9965.9 210 96.8 0 112.8 ...
 $ Deferred tax liability                             : num  284.9 85.2 0 0 4.6 ...
 $ Shareholders funds                                 : num  7093.2 351.5 100.6 2.7 107.6 ...
 $ Cumulative retained profits                        : num  6263.3 247.4 32.4 2.2 82.7 ...
 $ Capital employed                                   : num  7108.1 624 136 2.7 300.7 ...
 $ TOL/TNW                                            : num  1.33 1.23 1.44 0 2.83 1.8 0.03 5.17 1.05 3.25 ...
 $ Total term liabilities / tangible net worth        : num  0 0.34 0.29 0 1.59 0.37 0.03 0.94 0.3 0.54 ...
 $ Contingent liabilities / Net worth (%)             : num  14.8 19.2 45.8 0 34.9 ...
 $ Contingent liabilities                             : num  1049.7 67.6 46.1 0 37.6 ...

 $ Net fixed assets                                   : num  1900.2 286.4 38.7 2.5 94.8 ...
 $ Investments                                        : num  1069.6 2.2 4.3 0 7.4 ...
 $ Current assets                                     : num  13277.5 563.9 167.5 0.2 349.7 ...
 $ Net working capital                                : num  3588.5 203.5 59.6 0.2 215.8 ...
 $ Quick ratio (times)                                : num  1.18 0.95 1.11 0 1.41 0.48 0 0.54 0.59 0.39 ...
 $ Current ratio (times)                              : num  1.37 1.56 1.55 0 2.54 1.27 0 1.15 1.58 0.5 ...
 $ Debt to equity ratio (times)                       : num  0 0.78 0.35 0 1.79 1.09 0.32 2.31 0.94 3.13 ...
 $ Cash to current liabilities (times)                : num  0.43 0.06 0.21 0 0 0 0.11 0 0.04 0.19 0 ...
 $ Cash to average cost of sales per day              : num  68.21 5.96 17.07 0 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 0 12.68 ...
 $ WIP turnover                                       : num  21.78 7.49 0.23 0 7.9 ...
 $ Raw material turnover                              : num  7.71 11.46 0 0 17.03 ...
 $ Shares outstanding                                 : num  42381675 11550000 8149090 52404 619635 ...
 $ Equity face value                                  : num  10 10 10 10 10 10 10 0 10 10 ...
 $ EPS                                                : num  35.52 9.97 -0.5 0 7.91 ...
 $ Adjusted EPS                                       : num  7.1 9.97 -0.5 0 7.91 ...
 $ Total liabilities                                  : num  17512.3 941 232.8 2.7 478.5 ...
 $ PE on BSE                                          : num  27.31 8.17 -5.76 0 0 ...
>
```

New variables creation(Profitability, Liquidity, Leverage and Company's size):

The following new variables have been created and added to the dataset.

```
ROE.Profitability <- companydefault$`Profit after tax`/companydefault$`Total assets`
leverageRatio <- companydefault$`Capital employed`/companydefault$`Total assets`
companySize <- log(companydefault$`Total assets`)
```

Structure of the dataset after adding new variables is:

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4716  -0.7453  -0.0589   0.6872   8.4904

Coefficients:
                                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                                    2.208e+00  3.365e-01   6.560 5.38e-11 ***
`Change in stock`                              2.346e-03  1.607e-03   1.460 0.144335
`Profit after tax`                            -1.078e-02  1.773e-03  -6.082 1.19e-09 ***
`PBDITA as % of total income`                  1.272e-03  6.927e-04   1.837 0.066258 .
`PBT as % of total income`                    -7.777e-04  4.639e-04  -1.676 0.093659 .
`PAT as % of net worth`                       -1.702e-02  2.145e-03  -7.932 2.15e-15 ***
`Income from financial services`               5.665e-03  3.984e-03   1.422 0.155033
`Other income`                                -1.474e-02  4.844e-03  -3.044 0.002336 **
`Total capital`                                2.781e-03  5.125e-04   5.426 5.77e-08 ***
`Total term liabilities / tangible net worth`  1.160e-02  3.324e-03   3.490 0.000483 ***
`Contingent liabilities / Net worth (%)`       5.987e-04  3.307e-04   1.811 0.070184 .
`Contingent liabilities`                      -3.673e-03  9.930e-04  -3.699 0.000216 ***
`Net working capital`                         -1.415e-03  5.507e-04  -2.570 0.010158 *
`Quick ratio (times)`                         -4.369e-02  2.504e-02  -1.744 0.081086 .
`Cash to average cost of sales per day`        1.923e-04  8.161e-05   2.356 0.018493 *
`Creditors turnover`                          -6.238e-03  3.017e-03  -2.068 0.038675 *
`Debtors turnover`                             3.123e-03  7.990e-04   3.908 9.29e-05 ***
`Finished goods turnover`                      4.862e-04  1.365e-04   3.563 0.000367 ***
`WIP turnover`                                -5.268e-03  2.028e-03  -2.597 0.009406 **
`Raw material turnover`                       -5.683e-03  3.639e-03  -1.562 0.118353
`Shares outstanding`                          -9.447e-09  5.393e-09  -1.752 0.079842 .
`Equity face value`                            1.931e-05  1.166e-05   1.656 0.097789 .
`Adjusted EPS`                                -5.561e-03  1.847e-03  -3.011 0.002605 **
`PE on BSE`                                   -4.427e-03  2.459e-03  -1.800 0.071840 .
ROE.Profitability                              1.141e-01  6.566e-02   1.738 0.082238 .
leverageRatio                                 -1.783e+00  3.378e-01  -5.277 1.31e-07 ***
companySize                                   -2.817e-01  4.166e-02  -6.762 1.36e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Check for multicollinearity:

In the given dataset, ‘Networth Next Year’ is considered as dependant variable which actually shows a positive value as non default and a negative value as default.

Hence, converted ‘Networth Next Year’ as a factor variable which shows 0 for non default and 1 for default.

The ratios of default and non default in the given data set is:

```
> prop.table(table(companydefault$`Default - 1`))

         0          1
0.93137532 0.06862468
```

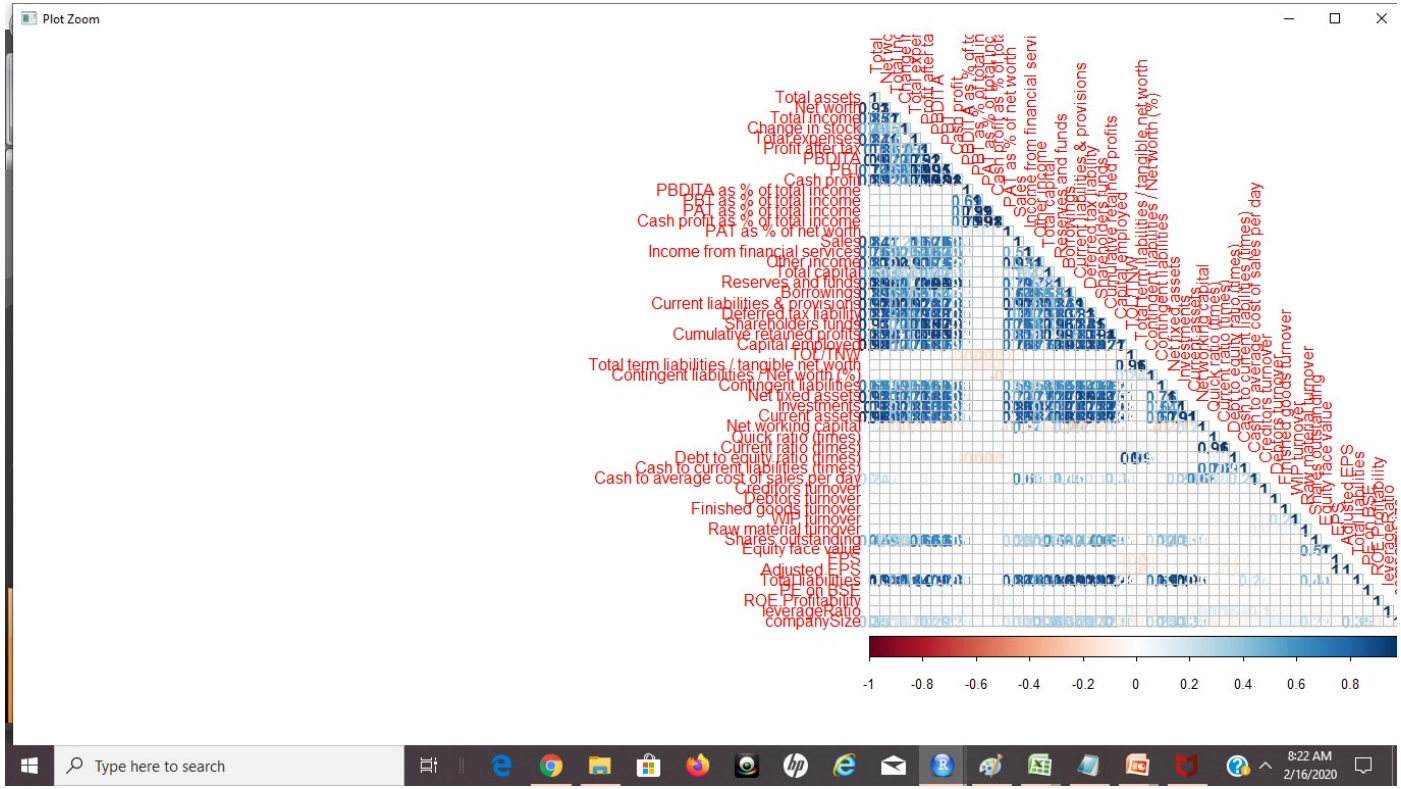
Since, the default percentage is only 6.8%, applied SMOTE test to balance the data.

The ratios of default and non default after SMOTE test is:

```
> prop.table(table(default.smote$`Default - 1`))

        0         1
0.5714286 0.4285714
```

Now, tested multicollinearity between the independent variables in the balanced data set.

The Correlation matrix is:



The highly correlated values are:

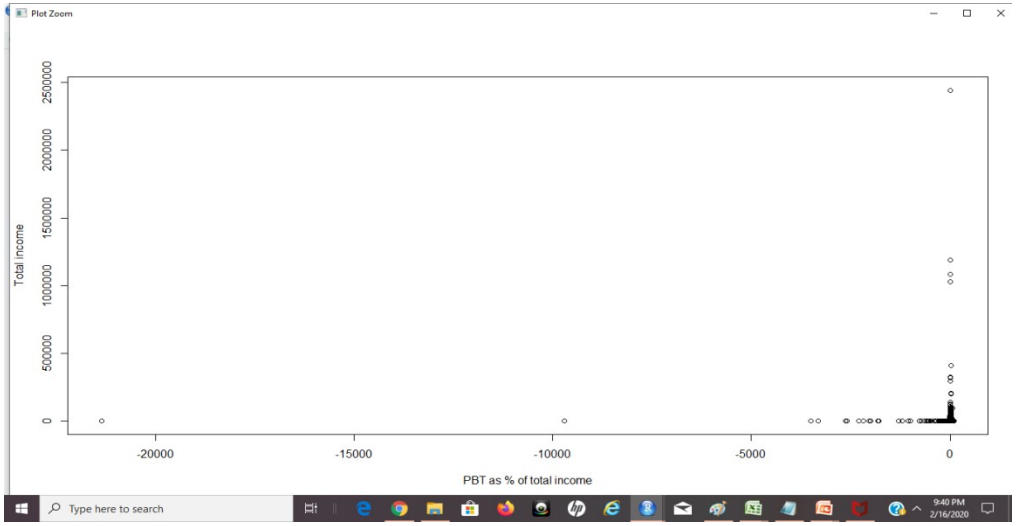
```
> print(highlyCorrelated)
 [1]  1 48 23  2  7 25 32 21 19  9 30 24 22 31  3 15  5  8 20 26 36 13 12 35 37 46
> highlyCorCol <- colnames(default.smote.cor)[highlyCorrelated]
> highlyCorCol
 [1] "Total assets"                      "Total liabilities"
 [3] "Shareholders funds"                "Net worth"
 [5] "PBDITA"                            "Capital employed"
 [7] "Current assets"                    "current liabilities & provisions"
 [9] "Reserves and funds"                "Cash profit"
[11] "Net fixed assets"                  "Cumulative retained profits"
[13] "Deferred tax liability"            "Investments"
[15] "Total income"                      "Sales"
[17] "Total expenses"                    "PBT"
[19] "Borrowings"                        "TOL/TNW"
[21] "Debt to equity ratio (times)"      "Cash profit as % of total income"
[23] "PAT as % of total income"          "Current ratio (times)"
[25] "Cash to current liabilities (times)" "EPS"
```

Univariate and Bivariate Analysis:

The following variables have been observed using R Code:

```
#Univariate Analysis
#Converting Networth Next Year to be the dependant variable
companydefault$`Networth Next Year` = ifelse(companydefault$`Networth Next Year` > 0,0,1)
summary(companydefault$`Networth Next Year`)
names(companydefault)[names(companydefault) == "Networth Next Year"] <- "Default - 1"
names(companydefault)
#BiVariate Analysis
plot(`PBT as % of total income`, `Total income` )
```

➢summary(companydefault$`Networth Next Year`) Min. 1st Qu. Median Mean 3rd Qu. Max. 0.00000 0.00000 0.00000 0.06862 0.00000 1.00000



Building Logistic Regression Model:

After removing the highly correlated variables from the data set, Linear Regression Model has been run on the dataset to identify the significant variables to determine the companies who can be most likely default their credit.

```
#Run the logistic regression model on the data
model=glm(newCompanyDefault$`Default - 1`~.,data = newCompanyDefault,family = binomial)
summary(model)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4716  -0.7453  -0.0589   0.6872   8.4904

Coefficients:
                                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                              2.208e+00  3.365e-01   6.560 5.38e-11 ***
`Change in stock`                        2.346e-03  1.607e-03   1.460 0.144335
`Profit after tax`                      -1.078e-02  1.773e-03  -6.082 1.19e-09 ***
`PBDITA as % of total income`            1.272e-03  6.927e-04   1.837 0.066258 .
`PBT as % of total income`              -7.777e-04  4.639e-04  -1.676 0.093659 .
`PAT as % of net worth`                 -1.702e-02  2.145e-03  -7.932 2.15e-15 ***
`Income from financial services`         5.665e-03  3.984e-03   1.422 0.155033
`Other income`                          -1.474e-02  4.844e-03  -3.044 0.002336 **
`Total capital`                          2.781e-03  5.125e-04   5.426 5.77e-08 ***
`Total term liabilities / tangible net worth` 1.160e-02  3.324e-03   3.490 0.000483 ***
`Contingent liabilities / Net worth (%)`  5.987e-04  3.307e-04   1.811 0.070184 .
`Contingent liabilities`                -3.673e-03  9.930e-04  -3.699 0.000216 ***
`Net working capital`                   -1.415e-03  5.507e-04  -2.570 0.010158 *
`Quick ratio (times)`                   -4.369e-02  2.504e-02  -1.744 0.081086 .
`Cash to average cost of sales per day`  1.923e-04  8.161e-05   2.356 0.018493 *
`Creditors turnover`                    -6.238e-03  3.017e-03  -2.068 0.038675 *
`Debtors turnover`                       3.123e-03  7.990e-04   3.908 9.29e-05 ***
`Finished goods turnover`                4.862e-04  1.365e-04   3.563 0.000367 ***
`WIP turnover`                          -5.268e-03  2.028e-03  -2.597 0.009406 **
`Raw material turnover`                 -5.683e-03  3.639e-03  -1.562 0.118353
`Shares outstanding`                    -9.447e-09  5.393e-09  -1.752 0.079842 .
`Equity face value`                      1.931e-05  1.166e-05   1.656 0.097789 .
`Adjusted EPS`                          -5.561e-03  1.847e-03  -3.011 0.002605 **
`PE on BSE`                             -4.427e-03  2.459e-03  -1.800 0.071840 .
ROE.Profitability                        1.141e-01  6.566e-02   1.738 0.082238 .
leverageRatio                           -1.783e+00  3.378e-01  -5.277 1.31e-07 ***
companySize                             -2.817e-01  4.166e-02  -6.762 1.36e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2323.3  on 1700  degrees of freedom
Residual deviance: 1446.5  on 1674  degrees of freedom
AIC: 1500.5

Number of Fisher Scoring iterations: 14
```

The VIF values of the model are:

```
> vif(model)
                                `Change in stock`                                  `Profit after tax`
                                         1.153756                                            1.639683
                        `PBDITA as % of total income`                       `PBT as % of total income`
                                         6.746464                                            6.835257
                              `PAT as % of net worth`                  `Income from financial services`
                                         1.326457                                           35.993928
                                   `Other income`                                     `Total capital`
                                         2.190268                                            4.443725
`Total term liabilities / tangible net worth`       `Contingent liabilities / Net worth (%)`
                                         1.156246                                            1.307869
                             `Contingent liabilities`                           `Net working capital`
                                         2.093729                                           33.294679
                                `Quick ratio (times)`       `Cash to average cost of sales per day`
                                         1.498381                                            6.231998
                                 `Creditors turnover`                               `Debtors turnover`
                                         1.093685                                            1.219015
                           `Finished goods turnover`                                  `WIP turnover`
                                         1.201066                                            1.179831
                            `Raw material turnover`                           `Shares outstanding`
                                         1.062652                                           81.431321
                                `Equity face value`                                `Adjusted EPS`
                                        79.368621                                            1.029628
                                       `PE on BSE`                           ROE.Profitability
                                         1.023365                                            1.088140
                                     leverageRatio                                       companySize
                                         1.151605                                            1.450124
>
```

Variables with VIF values >4 have been discarded from the dataset and again Logistic regression model has been run on the remaining data.

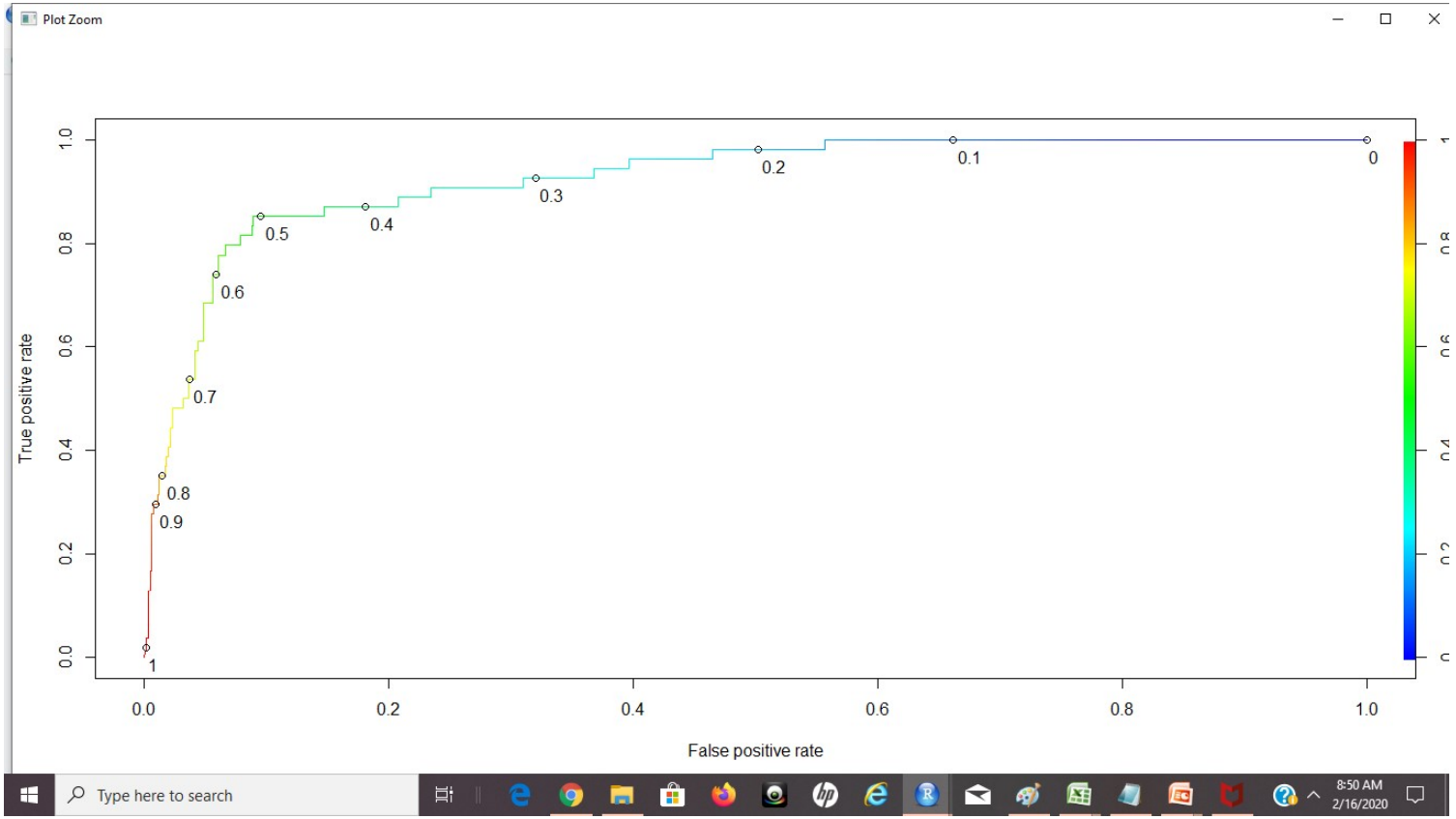
```
> model1=glm(newCompanyDefault$`Default - 1`~ `Change in stock`+`Profit after tax`
+            +`PAT as % of net worth`+`Other income`
+            +`Total term liabilities / tangible net worth`+`Contingent liabilities / Net worth (%)`
+            +`Contingent liabilities`+`Quick ratio (times)`+`Creditors turnover`
+            +`Debtors turnover`+`Finished goods turnover`+`WIP turnover`
+            +`Raw material turnover`+`Adjusted EPS`+`PE on BSE`+ROE.Profitability
+            +leverageRatio+companySize,data = newCompanyDefault,family = binomial)
Warning message:
glm.fit: fitted probabilities numerically 0 or 1 occurred
> vif(model1)
                            `Change in stock`                              `Profit after tax`
                                     1.136486                                       1.614149
                        `PAT as % of net worth`                                `Other income`
                                     1.283039                                       1.170008
`Total term liabilities / tangible net worth`      `Contingent liabilities / Net worth (%)`
                                     1.204799                                       1.214926
                         `Contingent liabilities`                       `Quick ratio (times)`
                                     1.410640                                       1.036539
                            `Creditors turnover`                           `Debtors turnover`
                                     1.094408                                       1.169641
                       `Finished goods turnover`                              `WIP turnover`
                                     1.180903                                       1.100882
                        `Raw material turnover`                              `Adjusted EPS`
                                     1.067050                                       1.030657
                                    `PE on BSE`                            ROE.Profitability
                                     1.013278                                       1.083212
                                  leverageRatio                                  companySize
                                     1.109832                                       1.311789
>
```

As per the above result, these variables are considered significant and hence ran the model performance test on this model.

Model Performance Measures:

```
> #Creation of confusion matrix to assess model performance measures
> tab.LR= table(newCompanyDefault$`Default - 1`,predTrain>0.3)
> tab.LR

    FALSE TRUE
  0   622  350
  1    64  665
> sum(diag(tab.LR))/sum(tab.LR)
[1] 0.7566138
>
> tab.LR= table(newCompanyDefault$`Default - 1`,predTrain>0.2)
> tab.LR

    FALSE TRUE
  0   441  531
  1    34  695
> sum(diag(tab.LR))/sum(tab.LR)
[1] 0.6678424
>
> tab.LR= table(newCompanyDefault$`Default - 1`,predTrain>0.5)
> tab.LR

    FALSE TRUE
  0   873   99
  1   163  566
> sum(diag(tab.LR))/sum(tab.LR)
[1] 0.845973
```

We can see that as the threshold value increases, the specificity also increases. In order to find the optimum threshold value, ROC curve has been drawn as shown in the picture.



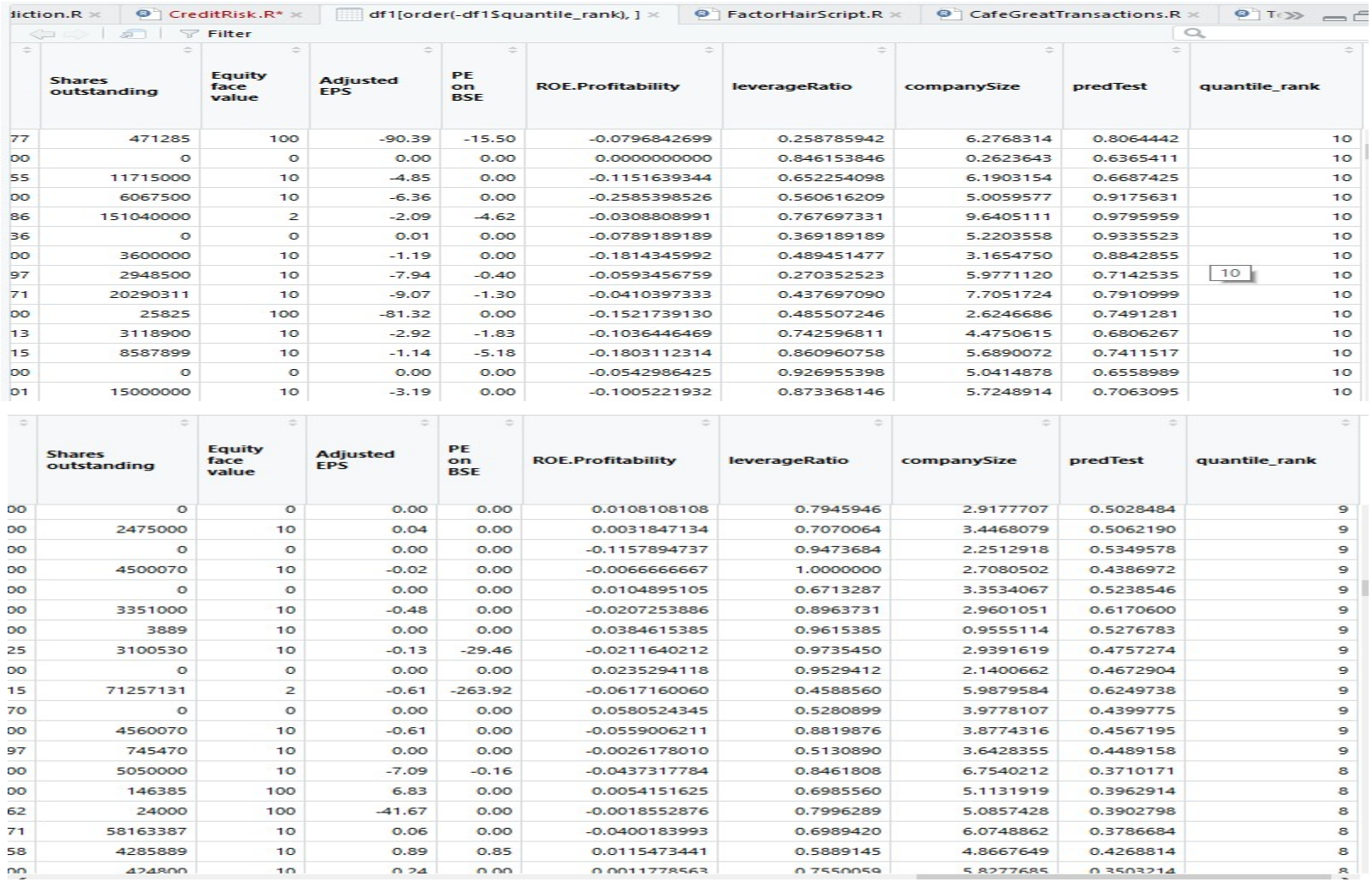
The threshold decreases as we move from (0,0) to (1,1). At the point (0, 0.5), we're correctly labeling about 80% of the poor care cases with a very small false positive rate. On the other hand, at the point (0.2, 0.9), we're correctly labeling about 90% of the poor care cases, but have a false positive rate of 55%. In the middle, around (0.3, 0.8), we're correctly labeling about 80% of the poor care cases, with a 35% false-positive rate. Hence, a threshold of 0.5 is good in this case.

That is 85% of the cases are actually predicted as good.
Now , applying this model on the given test data, the result is as shown below.

```
> predTest=predict(model1, newdata = newTestData,type = "response")
> tab.LR1= table(newTestData$`Default - 1`,predTest>0.3)
> tab.LR1

    FALSE TRUE
  0   450  211
  1     4   50
> sum(diag(tab.LR1))/sum(tab.LR1)
[1] 0.6993007
>
> tab.LR1= table(newTestData$`Default - 1`,predTest>0.2)
> tab.LR1

    FALSE TRUE
  0   330  331
  1     1   53
> sum(diag(tab.LR1))/sum(tab.LR1)
[1] 0.5356643
>
> tab.LR1= table(newTestData$`Default - 1`,predTest>0.4)
> tab.LR1

    FALSE TRUE
  0   542  119
  1     7   47
> sum(diag(tab.LR1))/sum(tab.LR1)
[1] 0.8237762
```

This result is almost same as train data set result.
Hence, this model is considered as good model to identify defaulters in the given dataset.

Data has been sorted based on the probabilities of default and shown below.



jiction.R | CreditRisk.R* | df1[order(-df1$quantile_rank),] | FactorHairScript.R | CafeGreatTransactions.R

Filter

	Shares outstanding	Equity face value	Adjusted EPS	PE on BSE	ROE.Profitability	leverageRatio	companySize	predTest	quantile_rank
77	471285	100	-90.39	-15.50	-0.0796842699	0.258785942	6.2768314	0.8064442	10
00	0	0	0.00	0.00	0.0000000000	0.846153846	0.2623643	0.6365411	10
55	11715000	10	-4.85	0.00	-0.1151639344	0.652254098	6.1903154	0.6687425	10
00	6067500	10	-6.36	0.00	-0.2585398526	0.560616209	5.0059577	0.9175631	10
86	151040000	2	-2.09	-4.62	-0.0308808991	0.767697331	9.6405111	0.9795959	10
36	0	0	0.01	0.00	-0.0789189189	0.369189189	5.2203558	0.9335523	10
00	3600000	10	-1.19	0.00	-0.1814345992	0.489451477	3.1654750	0.8842855	10
97	2948500	10	-7.94	-0.40	-0.0593456759	0.270352523	5.9771120	0.7142535	10
71	20290311	10	-9.07	-1.30	-0.0410397333	0.437697090	7.7051724	0.7910999	10
00	25825	100	-81.32	0.00	-0.1521739130	0.485507246	2.6246686	0.7491281	10
13	3118900	10	-2.92	-1.83	-0.1036446469	0.742596811	4.4750615	0.6806267	10
15	8587899	10	-1.14	-5.18	-0.1803112314	0.860960758	5.6890072	0.7411517	10
00	0	0	0.00	0.00	-0.0542986425	0.926955398	5.0414878	0.6558989	10
01	15000000	10	-3.19	0.00	-0.1005221932	0.873368146	5.7248914	0.7063095	10

	Shares outstanding	Equity face value	Adjusted EPS	PE on BSE	ROE.Profitability	leverageRatio	companySize	predTest	quantile_rank
00	0	0	0.00	0.00	0.0108108108	0.7945946	2.9177707	0.5028484	9
00	2475000	10	0.04	0.00	0.0031847134	0.7070064	3.4468079	0.5062190	9
00	0	0	0.00	0.00	-0.1157894737	0.9473684	2.2512918	0.5349578	9
00	4500070	10	-0.02	0.00	-0.0066666667	1.0000000	2.7080502	0.4386972	9
00	0	0	0.00	0.00	0.0104895105	0.6713287	3.3534067	0.5238546	9
00	3351000	10	-0.48	0.00	-0.0207253886	0.8963731	2.9601051	0.6170600	9
00	3889	10	0.00	0.00	0.0384615385	0.9615385	0.9555114	0.5276783	9
25	3100530	10	-0.13	-29.46	-0.0211640212	0.9735450	2.9391619	0.4757274	9
00	0	0	0.00	0.00	0.0235294118	0.9529412	2.1400662	0.4672904	9
15	71257131	2	-0.61	-263.92	-0.0617160060	0.4588560	5.9879584	0.6249738	9
70	0	0	0.00	0.00	0.0580524345	0.5280899	3.9778107	0.4399775	9
00	4560070	10	-0.61	0.00	-0.0559006211	0.8819876	3.8774316	0.4567195	9
97	745470	10	0.00	0.00	-0.0026178010	0.5130890	3.6428355	0.4489158	9
00	5050000	10	-7.09	-0.16	-0.0437317784	0.8461808	6.7540212	0.3710171	8
00	146385	100	6.83	0.00	0.0054151625	0.6985560	5.1131919	0.3962914	8
62	24000	100	-41.67	0.00	-0.0018552876	0.7996289	5.0857428	0.3902798	8
71	58163387	10	0.06	0.00	-0.0400183993	0.6989420	6.0748862	0.3786684	8
58	4285889	10	0.89	0.85	0.0115473441	0.5889145	4.8667649	0.4268814	8
00	424800	10	0.24	0.00	0.0011778563	0.7550059	5.8277685	0.3503214	8

Appendix A

```r
#================================================================================
# Data Analysis - Credit Risk
#================================================================================
#Environment Set up and Data Import
#Set up working Directory
setwd("C:/Users/Radhika/Desktop/R Programming/Project_FRA")
getwd()
#
rm(list = ls(all.names = TRUE))
#install.packages("readxl")
library(readxl)
#Read the input file
companydefault=read_excel("raw-data.xlsx", sheet = "raw data")
attach(companydefault)
summary(companydefault)
#Check for null values
sum(is.na(companydefault))
#Replace null values with 0
companydefault[is.na(companydefault)] <- 0
sum(is.na(companydefault))
summary(companydefault)
#Checking outliers
boxplot(`PBT as % of total income`)
boxplot(`Current ratio (times)`)
#check the data types of the variables
str(companydefault)
length(companydefault)
#Convert the character type variables to numeric type variables
companydefault$`Creditors turnover`[companydefault$`Creditors turnover`=="NA"]<-"0"
companydefault$`Creditors turnover` = as.numeric(companydefault$`Creditors turnover`)
companydefault$`Debtors turnover`[companydefault$`Debtors turnover`=="NA"]<-"0"
companydefault$`Debtors turnover` = as.numeric(companydefault$`Debtors turnover`)
companydefault$`Finished goods turnover`[companydefault$`Finished goods turnover`=="NA"]<-"0"
companydefault$`Finished goods turnover` = as.numeric(companydefault$`Finished goods turnover`)
companydefault$`WIP turnover`[companydefault$`WIP turnover`=="NA"]<-"0"
```

```r
companydefault$`WIP turnover` = as.numeric(companydefault$`WIP turnover`)
companydefault$`Raw material turnover`[companydefault$`Raw material turnover`=="NA"]<-"0"
companydefault$`Raw material turnover` = as.numeric(companydefault$`Raw material turnover`)
companydefault$`Shares outstanding`[companydefault$`Shares outstanding`=="NA"]<-"0"
companydefault$`Shares outstanding` = as.numeric(companydefault$`Shares outstanding`)
companydefault$`PE on BSE`[companydefault$`PE on BSE`=="NA"]<-"0"
companydefault$`PE on BSE` = as.numeric(companydefault$`PE on BSE`)
companydefault$`Equity face value`[companydefault$`Equity face value`=="NA"]<-"0"
companydefault$`Equity face value` = as.numeric(companydefault$`Equity face value`)
str(companydefault)
#Univariate Analysis
#Converting Networth Next Year to be the dependant variable
companydefault$`Networth Next Year` = ifelse(companydefault$`Networth Next Year` > 0,0,1)
summary(companydefault$`Networth Next Year`)
names(companydefault)[names(companydefault) == "Networth Next Year"] <- "Default - 1"
names(companydefault)
#BiVariate Analysis
plot(`PBT as % of total income`, `Total income` )
#check the percentage of 0's and 1's
prop.table(table(companydefault$`Default - 1`))
#as per the result, % of negative networth data is only 6.8% of total data
#hence, we need to balance the data
#Creating new columns of Profitability ratio, Liquidity Ratio and CompanySize
ROE.Profitability <- companydefault$`Profit after tax`/companydefault$`Total assets`
leverageRatio <- companydefault$`Capital employed`/companydefault$`Total assets`
companySize <- log(companydefault$`Total assets`)
companydefault = cbind(companydefault, ROE.Profitability)
companydefault = cbind(companydefault, leverageRatio)
companydefault = cbind(companydefault, companySize)
str(companydefault)
length(companydefault)
#balance the data
#install.packages("DMwR")
library(DMwR)
attach(companydefault)
names(companydefault)
```

```r
companydefault = as.data.frame(companydefault)
companydefault$`Default - 1` = as.factor(companydefault$`Default - 1`)
summary(companydefault$`Default - 1`)
str(companydefault$`Default - 1`)
default.smote = SMOTE(`Default - 1` ~., data=companydefault)
#companydefault$`Default - 1` = as.numeric(companydefault$`Default - 1`)
summary(companydefault$`Default - 1`)
prop.table(table(default.smote$`Default - 1`))

library(caTools)
library(caret)
library(ROCR)
library(corrplot)
library(car)
#checking for multicollinearity
#remove the number column and dependant variable column while
#checking multicollinearity
default.smote.cor = default.smote[,-1]
default.smote.cor = default.smote.cor[,-1]
length(default.smote.cor)
names(default.smote.cor)
#default.smote.cor$Debt.to.capital.leverage = round(default.smote.cor$Debt.to.capital.leverage, digits = 2)
#default.smote.cor = default.smote.cor[,-51]
str(default.smote.cor)

correlation=corrplot(cor(default.smote.cor),method = "number",type = "lower")
highlyCorrelated <- findCorrelation(cor(default.smote.cor), cutoff=0.7)
print(highlyCorrelated)
highlyCorCol <- colnames(default.smote.cor)[highlyCorrelated]
highlyCorCol
#Remove the highly correlated columns from the data set
newCompanyDefault <- default.smote[, -which(colnames(default.smote) %in% highlyCorCol)]
length(newCompanyDefault)
#remove the number column from the new data set
#as it is not required for generating a model
newCompanyDefault = newCompanyDefault[,-1]
```

```
newCompanyDefault = newCompanyDefault[,-1]
str(newCompanyDefault)
#Run the logistic regression model on the data
model=glm(newCompanyDefault$`Default - 1`~.,data = newCompanyDefault,family = binomial)
summary(model)
vif(model)
#removing columns of `Total capital`, `Shares outstanding` as
#their VIF values are > 4
#newCompanyDefault = newCompanyDefault[, -7]
#newCompanyDefault = newCompanyDefault[, -18]
names(newCompanyDefault)
#Run the logistic regression model on the new data after
#removing columns of VIF values > 4
model1=glm(newCompanyDefault$`Default - 1`~ `Change in stock`+`Profit after tax`
           +`PAT as % of net worth`+`Other income`
           +`Total term liabilities / tangible net worth`+`Contingent liabilities / Net worth (%)`
           +`Contingent liabilities`+`Quick ratio (times)`+`Creditors turnover`
           +`Debtors turnover`+`Finished goods turnover`+`WIP turnover`
           +`Raw material turnover`+`Adjusted EPS`+`PE on BSE`+ROE.Profitability
           +leverageRatio+companySize,data = newCompanyDefault,family = binomial)
summary(model1)
vif(model1)
#Predit the train data model
predTrain=predict(model1, newdata = newCompanyDefault,type = "response")

#Creation of confusion matrix to assess model performance measures
tab.LR= table(newCompanyDefault$`Default - 1`,predTrain>0.3)
tab.LR
sum(diag(tab.LR))/sum(tab.LR)

tab.LR= table(newCompanyDefault$`Default - 1`,predTrain>0.2)
tab.LR
sum(diag(tab.LR))/sum(tab.LR)

tab.LR= table(newCompanyDefault$`Default - 1`,predTrain>0.5)
tab.LR
```

```
sum(diag(tab.LR))/sum(tab.LR)
#Preparing test data
#Creating new columns as in Raw data
testData=read_excel("validation_data.xlsx", sheet = "valdata")
attach(testData)
summary(testData)
sum(is.na(testData))
testData[is.na(testData)] <- 0
sum(is.na(testData))
summary(testData)
str(testData)
#Converting character type variables to numeric type variables
testData$`Creditors turnover`[testData$`Creditors turnover`=="NA"]<-"0"
testData$`Creditors turnover` = as.numeric(testData$`Creditors turnover`)
testData$`Debtors turnover`[testData$`Debtors turnover`=="NA"]<-"0"
testData$`Debtors turnover` = as.numeric(testData$`Debtors turnover`)
testData$`Finished goods turnover`[testData$`Finished goods turnover`=="NA"]<-"0"
testData$`Finished goods turnover` = as.numeric(testData$`Finished goods turnover`)
testData$`WIP turnover`[testData$`WIP turnover`=="NA"]<-"0"
testData$`WIP turnover` = as.numeric(testData$`WIP turnover`)
testData$`Raw material turnover`[testData$`Raw material turnover`=="NA"]<-"0"
testData$`Raw material turnover` = as.numeric(testData$`Raw material turnover`)
testData$`Shares outstanding`[testData$`Shares outstanding`=="NA"]<-"0"
testData$`Shares outstanding` = as.numeric(testData$`Shares outstanding`)
testData$`PE on BSE`[testData$`PE on BSE`=="NA"]<-"0"
testData$`PE on BSE` = as.numeric(testData$`PE on BSE`)
testData$`Equity face value`[testData$`Equity face value`=="NA"]<-"0"
testData$`Equity face value` = as.numeric(testData$`Equity face value`)
str(testData)
length(testData)
#Creating new ratios of Profitability, liverage and company size
ROE.Profitability <- testData$`Profit after tax`/testData$`Total assets`
leverageRatio <- testData$`Capital employed`/testData$`Total assets`
companySize <- log(testData$`Total assets`)
testData = cbind(testData, ROE.Profitability)
testData = cbind(testData, leverageRatio)
```

```r
testData = cbind(testData, leverageRatio)
testData = cbind(testData, companySize)
str(testData)
length(testData)
#removing highly correlated columns from test data
newTestData <- testData[, -which(colnames(testData) %in% highlyCorCol)]
length(newTestData)
names(newTestData)
#removing number column from test data
newTestData = newTestData[,-1]
#Predicting test data with train model
predTest=predict(model1, newdata = newTestData,type = "response")
predTest
#Creation of confusion matrix to assess model performance measures
tab.LR1= table(newTestData$`Default - 1`,predTest>0.3)
tab.LR1
sum(diag(tab.LR1))/sum(tab.LR1)

tab.LR1= table(newTestData$`Default - 1`,predTest>0.2)
tab.LR1
sum(diag(tab.LR1))/sum(tab.LR1)

tab.LR1= table(newTestData$`Default - 1`,predTest>0.5)
tab.LR1
sum(diag(tab.LR1))/sum(tab.LR1)

library(ROCR)
ROCRpred = prediction(predTest, newTestData$`Default - 1`)
as.numeric(performance(ROCRpred, "auc")@y.values)
perf = performance(ROCRpred, "tpr","fpr")
plot(perf)
#Plot ROC curve
plot(perf, colorize =T, print.cutoffs.at= seq(0, 1, .1), text.adj = c(-.2, 1.7))
```

```r
#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
library(dplyr)
dim(newTestData)
decileData = newTestData
decileData
decileData = cbind(decileData,predTest)
df1 = mutate(decileData, quantile_rank = ntile(decileData$predTest,10))
str(df1)
View(df1[order(-df1$quantile_rank),])
```