|  |  |
| --- | --- |
| Risk Analytics: A CRO's POV | News | Management Events Finance Risk Analytics Project***Modelling India Credit Risk******Submitted By******Shweta Gupta (PGP BABI)*** |  |

**BACKGROUND**

* Create an India credit risk(default) model, using the data provided in the spreadsheet raw-data.xlsx, and validate it on validation\_data.xlsx.
* Use the logistic regression framework to develop the credit default model**.**

**KEY TASKS TO BE PERFORMED**

* Outlier Treatment
* Missing Value Treatment
* New Variables Creation (One ration for profitability, leverage, liquidity and company's size each )
* Check for multicollinearity
* Univariate & bivariate analysis
* Check for multicollinearity
* Build Logistic Regression Model on most important variables
* Analyze coefficient & their signs
* Predict accuracy of model on dev and validation datasets
* Sort the data in descending order based on probability of default and then divide into 10 deciles based on probability & check how well the model has performed

**DATA DICTIONARY**

**Two sets of data are provided**

1. **Raw data –** 3541 observations across 52 variables **for *training the model***
2. **Validation data. –** 715 observations across 52 variables **for *testing the accuracy of model***
3. Net worth next year **is the dependent variable using which we will create a binary variable Default to use in our model**

|  |  |  |
| --- | --- | --- |
| Variable Name | Discreption | Bucket |
| **Networth Next Year** | **Net worth of the customer in next year** | **Depedent Variable** |
| Total assets | Total assets of customer | Size |
| Net worth | Net worth of the customer of present year | Size |
| Total income | Total income of the customer | Size |
| Change in stock | difference between value of current stock and the value of stock in last trading day | change in size |
| Total expenses | Total expense done by customer | size/costs |
| Profit after tax | Profit after tax deduction | profit |
| PBDITA | Profit before depreciation, income tax and amortization | profit |
| PBT | Profit before tax deduction | profit |
| Cash profit | Total Cash profit | profit |
| PBDITA as % of total income | PBDITA / Total income | profit |
| PBT as % of total income | PBT / Total income | profit |
| PAT as % of total income | PAT / Total income | profit |
| Cash profit as % of total income | Cash Profit / Total income | profit |
| PAT as % of net worth | PAT / Net worth | profit |
| Sales | Sales done by customer | size |
| Income from financial services | Income from financial services | profit |
| Other income | Income from other sources | profit |
| Total capital | Total capital of the customer | size |
| Reserves and funds | Total reserves and funds of the customer | profit |
| Deposits (accepted by commercial banks) | All blank values | profits/size |
| Borrowings | Total amount borrowed by customer | leverage |
| Current liabilities & provisions | current liabilities of the customer | liquidity |
| Deferred tax liability | Future income tax customer will pay because of the current transaction | liquidity |
| Shareholders funds | Amount of equity in a company, which is belong to shareholder | size |
| Cumulative retained profits | Total cumulative profit retained by customer | profit |
| Capital employed | Current asset minus current liabilities | size |
| TOL/TNW | Total liabilities of the customer divided by Total net worth | leverage |
| Total term liabilities / tangible net worth | Short + long term liabilities divided by tangible net worth | leverage |
| Contingent liabilities / Net worth (%) | Contingent liabilities / Net worth | leverage |
| Contingent liabilities | Liabilities because of uncertain events | liquidity |
| Net fixed assets | purchase price of all fixed assets | size |
| Investments | Total invested amount | size |
| Current assets | Assets that are expected to be converted to cash within a year | size |
| Net working capital | Difference of current liabilities and current assets | liquidity/size |
| Quick ratio (times) | Total cash divided by current liabilities | liquidity |
| Current ratio (times) | Current assets divided by current liabilities | leverage |
| Debt to equity ratio (times) | Total liabilities divided by its shareholder equity | liquidity |
| Cash to current liabilities (times) | Total liquid cash divided by current liabilities | liquidity |
| Cash to average cost of sales per day | Total cash divided by average cost of the sales | liquidity |
| Creditors turnover | Net credit purchase divided to average trade creditors | liquidity |
| Debtors turnover | Net credit sales divided by average accounts receivable | liquidity |
| Finished goods turnover | Annual sales divided by average inventory | liquidity |
| WIP turnover | The cost of goods sold for a period divided by the average inventory for that period | liquidity |
| Raw material turnover | Cost of goods sold is divided by the average inventory for the same period | liquidity |
| Shares outstanding | Number of issued shares minus the number of share held in the company | size |
| Equity face value | cost of the equity at the time of issuing | size |
| EPS | Net income divided by total number of outstanding share | profitability |
| Adjusted EPS | Adjusted net earning divided by the weighted average number of common share outstanding on a diluted basis during the plan year | profitability |
| Total liabilities | Sum of all type of liabilities | leverage |
| PE on BSE | Company current stock price divided by its earning per share | market sentiment' |

**EXPLORATORY DATA ANALYSIS**

**TOTAL MISSING VALUES**

* DEVELOPMENT DATA - 18533
* VALIDATION DATA. - 3501

**Our data contains significant number of missing values**

**FIVE POINT SUMMARY – DEVELPOMENT DATA**

**Num Networth.Next.Year Total.assets Net.worth**

**Min. : 1 Min. :-74265.6 Min. : 0.1 Min. : 0.0**

**1st Qu.: 886 1st Qu.: 31.7 1st Qu.: 91.3 1st Qu.: 31.3**

**Median :1773 Median : 116.3 Median : 309.7 Median : 102.3**

**Mean :1772 Mean : 1616.3 Mean : 3443.4 Mean : 1295.9**

**3rd Qu.:2658 3rd Qu.: 456.1 3rd Qu.: 1098.7 3rd Qu.: 377.3**

**Max. :3545 Max. :805773.4 Max. :1176509.2 Max. :613151.6**

**Total.income Change.in.stock Total.expenses Profit.after.tax**

**Min. : 0.0 Min. :-3029.40 Min. : -0.1 Min. : -3908.30**

**1st Qu.: 106.5 1st Qu.: -1.80 1st Qu.: 95.8 1st Qu.: 0.50**

**Median : 444.9 Median : 1.60 Median : 407.7 Median : 8.80**

**Mean : 4582.8 Mean : 41.49 Mean : 4262.9 Mean : 277.36**

**3rd Qu.: 1440.9 3rd Qu.: 18.05 3rd Qu.: 1359.8 3rd Qu.: 52.27**

**Max. :2442828.2 Max. :14185.50 Max. :2366035.3 Max. :119439.10**

**NA's :198 NA's :458 NA's :139 NA's :131**

**PBDITA PBT Cash.profit**

**Min. : -440.7 Min. : -3894.80 Min. : -2245.70**

**1st Qu.: 6.9 1st Qu.: 0.70 1st Qu.: 2.90**

**Median : 35.4 Median : 12.40 Median : 18.85**

**Mean : 578.1 Mean : 383.81 Mean : 392.07**

**3rd Qu.: 150.2 3rd Qu.: 71.97 3rd Qu.: 93.20**

**Max. :208576.5 Max. :145292.60 Max. :176911.80**

**NA's :131 NA's :131 NA's :131**

**PBDITA.as...of.total.income PBT.as...of.total.income PAT.as...of.total.income**

**Min. :-6400.000 Min. :-21340.00 Min. :-21340.00**

**1st Qu.: 5.000 1st Qu.: 0.55 1st Qu.: 0.35**

**Median : 9.660 Median : 3.31 Median : 2.34**

**Mean : 4.571 Mean : -17.28 Mean : -19.20**

**3rd Qu.: 16.390 3rd Qu.: 8.80 3rd Qu.: 6.34**

**Max. : 100.000 Max. : 100.00 Max. : 150.00**

**NA's :68 NA's :68 NA's :68**

**Cash.profit.as...of.total.income PAT.as...of.net.worth Sales**

**Min. :-15020.000 Min. :-748.72 Min. : 0.1**

**1st Qu.: 2.020 1st Qu.: 0.00 1st Qu.: 112.7**

**Median : 5.640 Median : 7.92 Median : 453.1**

**Mean : -8.229 Mean : 10.27 Mean : 4549.5**

**3rd Qu.: 10.700 3rd Qu.: 20.19 3rd Qu.: 1433.5**

**Max. : 100.000 Max. :2466.67 Max. :2384984.4**

**NA's :68 NA's :259**

**Income.from.financial.services Other.income Total.capital**

**Min. : 0.00 Min. : 0.00 Min. : 0.1**

**1st Qu.: 0.40 1st Qu.: 0.40 1st Qu.: 13.1**

**Median : 1.80 Median : 1.40 Median : 42.1**

**Mean : 80.84 Mean : 41.36 Mean : 216.6**

**3rd Qu.: 9.68 3rd Qu.: 5.97 3rd Qu.: 100.3**

**Max. :51938.20 Max. :42856.70 Max. :78273.2**

**NA's :935 NA's :1295 NA's :4**

**Reserves.and.funds Deposits..accepted.by.commercial.banks. Borrowings**

**Min. : -6525.9 Mode:logical Min. : 0.10**

**1st Qu.: 5.0 NA's:3541 1st Qu.: 23.95**

**Median : 54.8 Median : 99.20**

**Mean : 1163.8 Mean : 1122.28**

**3rd Qu.: 277.3 3rd Qu.: 352.60**

**Max. :625137.8 Max. :278257.30**

**NA's :85 NA's :366**

**Current.liabilities...provisions Deferred.tax.liability Shareholders.funds**

**Min. : 0.1 Min. : 0.1 Min. : 0.0**

**1st Qu.: 17.8 1st Qu.: 3.2 1st Qu.: 32.0**

**Median : 69.4 Median : 13.4 Median : 105.6**

**Mean : 940.6 Mean : 227.2 Mean : 1322.1**

**3rd Qu.: 261.7 3rd Qu.: 50.0 3rd Qu.: 393.2**

**Max. :352240.3 Max. :72796.6 Max. :613151.6**

**NA's :96 NA's :1140**

**Cumulative.retained.profits Capital.employed TOL.TNW**

**Min. : -6534.3 Min. : 0.0 Min. :-350.480**

**1st Qu.: 1.1 1st Qu.: 60.8 1st Qu.: 0.600**

**Median : 37.1 Median : 214.7 Median : 1.430**

**Mean : 890.5 Mean : 2328.3 Mean : 3.994**

**3rd Qu.: 202.3 3rd Qu.: 767.3 3rd Qu.: 2.830**

**Max. :390133.8 Max. :891408.9 Max. : 473.000**

**NA's :38**

**Total.term.liabilities...tangible.net.worth**

**Min. :-325.600**

**1st Qu.: 0.050**

**Median : 0.340**

**Mean : 1.844**

**3rd Qu.: 1.000**

**Max. : 456.000**

**Contingent.liabilities...Net.worth.... Contingent.liabilities Net.fixed.assets**

**Min. : 0.00 Min. : 0.1 Min. : 0.0**

**1st Qu.: 0.00 1st Qu.: 6.3 1st Qu.: 26.0**

**Median : 5.33 Median : 38.0 Median : 93.5**

**Mean : 53.94 Mean : 932.9 Mean : 1189.7**

**3rd Qu.: 30.76 3rd Qu.: 192.7 3rd Qu.: 344.9**

**Max. :14704.27 Max. :559506.8 Max. :636604.6**

**NA's :1188 NA's :118**

**Investments Current.assets Net.working.capital Quick.ratio..times.**

**Min. : 0.00 Min. : 0.1 Min. :-63839.0 Min. : 0.000**

**1st Qu.: 1.00 1st Qu.: 36.2 1st Qu.: -1.1 1st Qu.: 0.410**

**Median : 8.35 Median : 145.1 Median : 16.2 Median : 0.670**

**Mean : 694.73 Mean : 1293.4 Mean : 138.6 Mean : 1.401**

**3rd Qu.: 64.30 3rd Qu.: 502.2 3rd Qu.: 84.2 3rd Qu.: 1.030**

**Max. :199978.60 Max. :354815.2 Max. : 85782.8 Max. :341.000**

**NA's :1435 NA's :66 NA's :32 NA's :93**

**Current.ratio..times. Debt.to.equity.ratio..times.**

**Min. : 0.00 Min. : 0.00**

**1st Qu.: 0.93 1st Qu.: 0.22**

**Median : 1.23 Median : 0.79**

**Mean : 2.13 Mean : 2.78**

**3rd Qu.: 1.71 3rd Qu.: 1.75**

**Max. :505.00 Max. :456.00**

**NA's :93**

**Cash.to.current.liabilities..times. Cash.to.average.cost.of.sales.per.day**

**Min. : 0.0000 Min. : 0.00**

**1st Qu.: 0.0200 1st Qu.: 2.79**

**Median : 0.0700 Median : 8.03**

**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**

**Creditors.turnover Debtors.turnover Finished.goods.turnover WIP.turnover**

**Min. : 0.000 Min. : 0.00 Min. : -0.09 Min. : -0.18**

**1st Qu.: 3.700 1st Qu.: 3.76 1st Qu.: 8.20 1st Qu.: 5.10**

**Median : 6.095 Median : 6.32 Median : 17.27 Median : 9.76**

**Mean : 15.446 Mean : 17.04 Mean : 87.08 Mean : 27.93**

**3rd Qu.: 11.490 3rd Qu.: 11.68 3rd Qu.: 40.35 3rd Qu.: 20.24**

**Max. :2401.000 Max. :3135.20 Max. :17947.60 Max. :5651.40**

**NA's :333 NA's :328 NA's :740 NA's :640**

**Raw.material.turnover Shares.outstanding Equity.face.value EPS**

**Min. : -2.00 Min. :-2.147e+09 Min. :-999999 Min. :-843181.8**

**1st Qu.: 2.99 1st Qu.: 1.316e+06 1st Qu.: 10 1st Qu.: 0.0**

**Median : 6.40 Median : 4.672e+06 Median : 10 Median : 1.4**

**Mean : 19.09 Mean : 2.207e+07 Mean : -1334 Mean : -220.3**

**3rd Qu.: 11.85 3rd Qu.: 1.065e+07 3rd Qu.: 10 3rd Qu.: 9.6**

**Max. :21092.00 Max. : 4.130e+09 Max. : 100000 Max. : 34522.5**

**NA's :361 NA's :692 NA's :692**

**Adjusted.EPS Total.liabilities PE.on.BSE**

**Min. :-843181.8 Min. : 0.1 Min. :-1116.64**

**1st Qu.: 0.0 1st Qu.: 91.3 1st Qu.: 3.27**

**Median : 1.2 Median : 309.7 Median : 9.10**

**Mean : -221.5 Mean : 3443.4 Mean : 63.91**

**3rd Qu.: 7.5 3rd Qu.: 1098.7 3rd Qu.: 17.79**

**Max. : 34522.5 Max. :1176509.2 Max. :51002.74**

**NA's :2194**

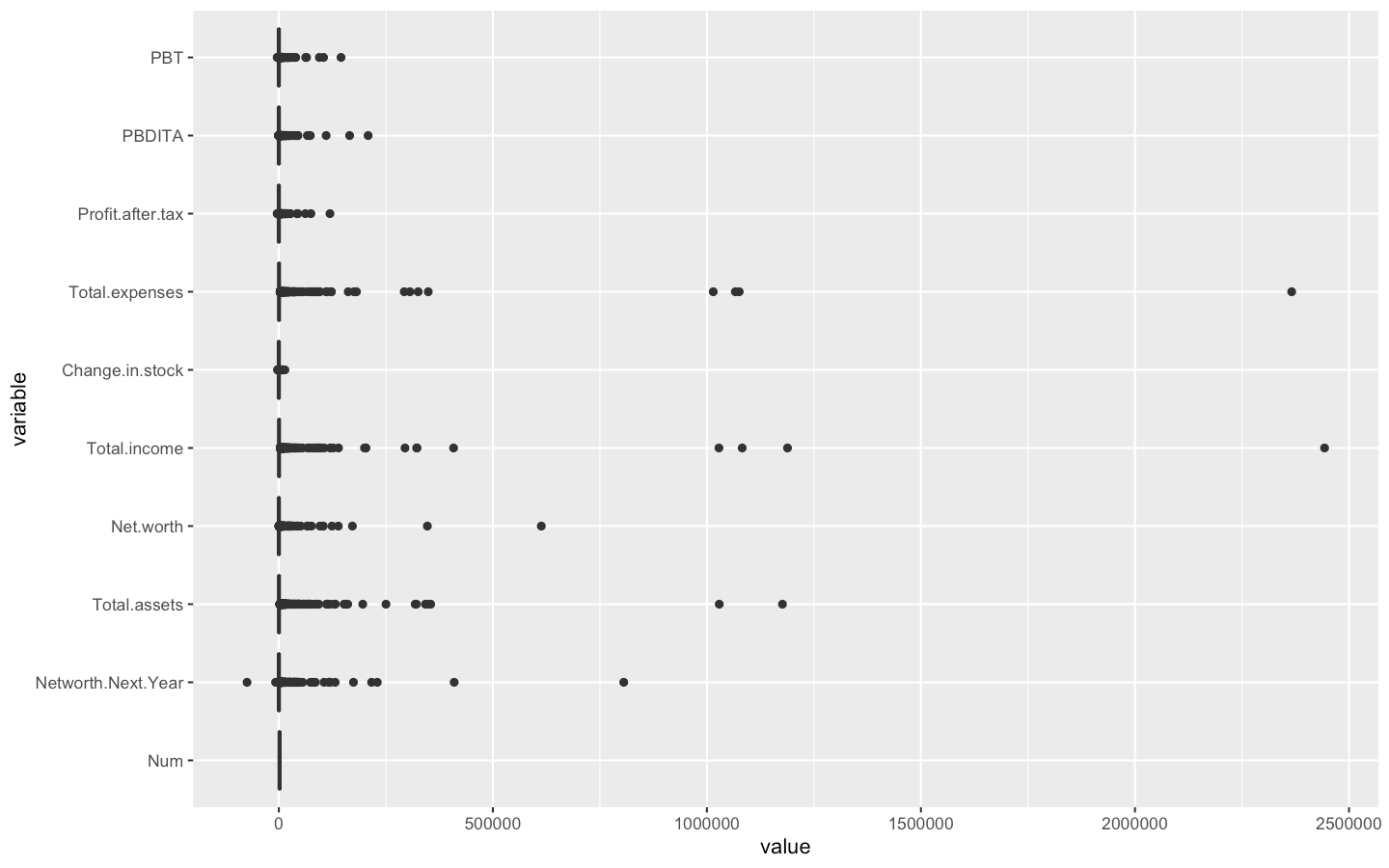
**KEY TAKE OUTS FROM FIVE POINT SUMMARY**

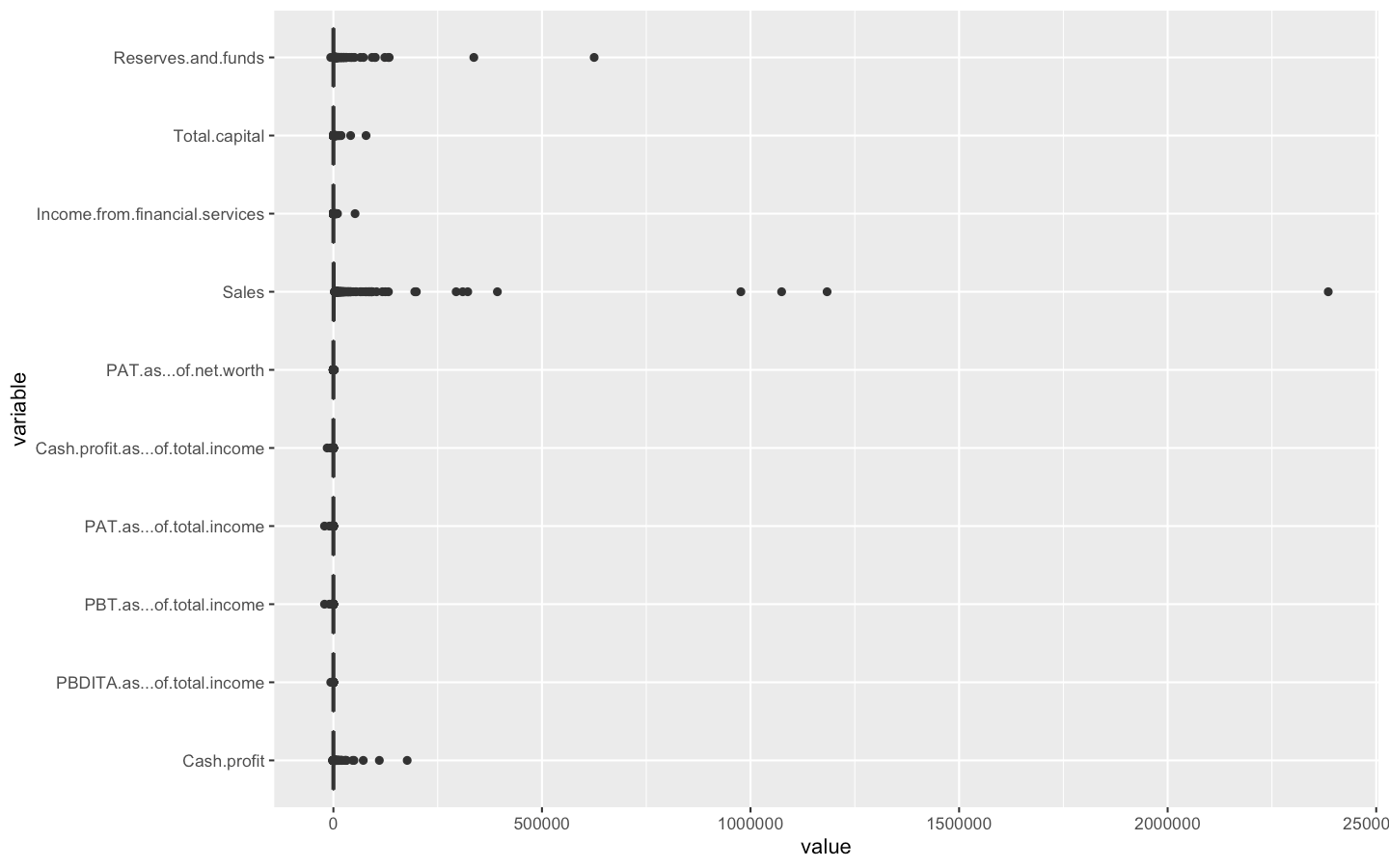
* Almost all variables have outliers
* Most of the variables have missing values
* Deposits variables has no data
* Variables with more than 30% missing values
  + Contingent liabilities -1188 NAs (34%)
  + Investments -1435 NAs (41%)
  + Deferred Tax liability- 1140 NAs (31%)
  + Other income – 1295 NAs (32%)
  + PE on BSE – 2194 NAs (62%)

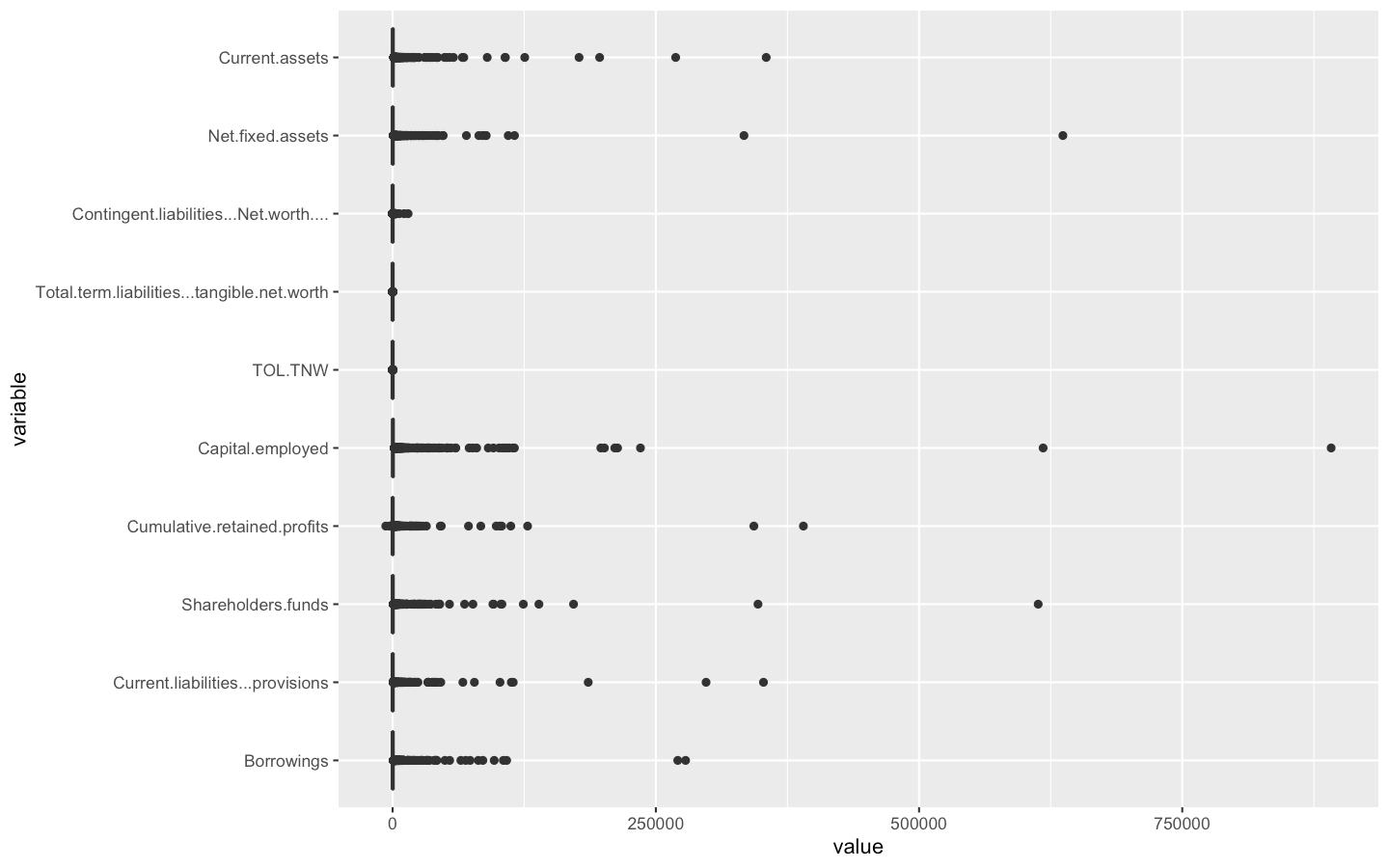
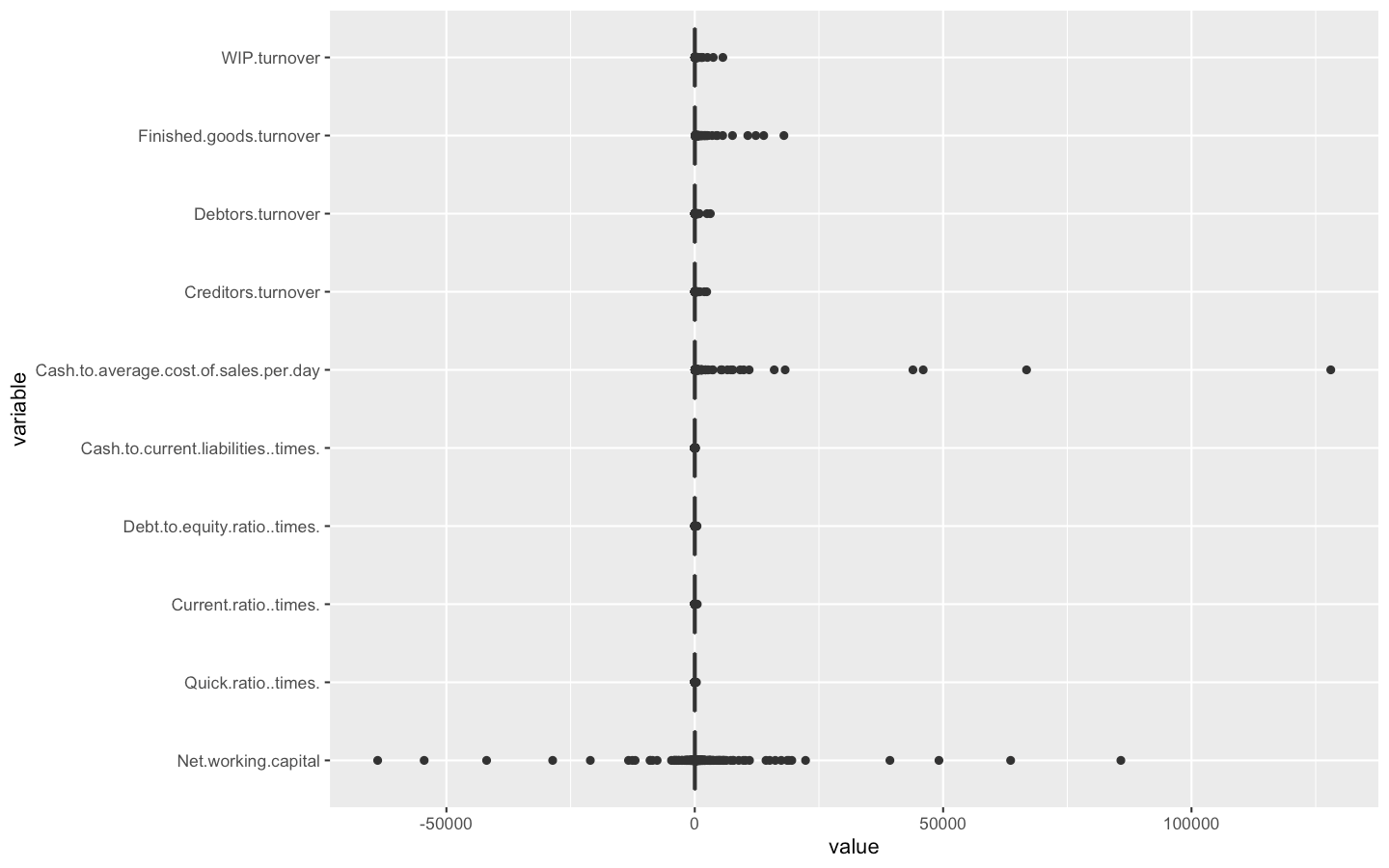
**KEY ACTIONS TAKEN ON THE DATA AT THIS STAGE**

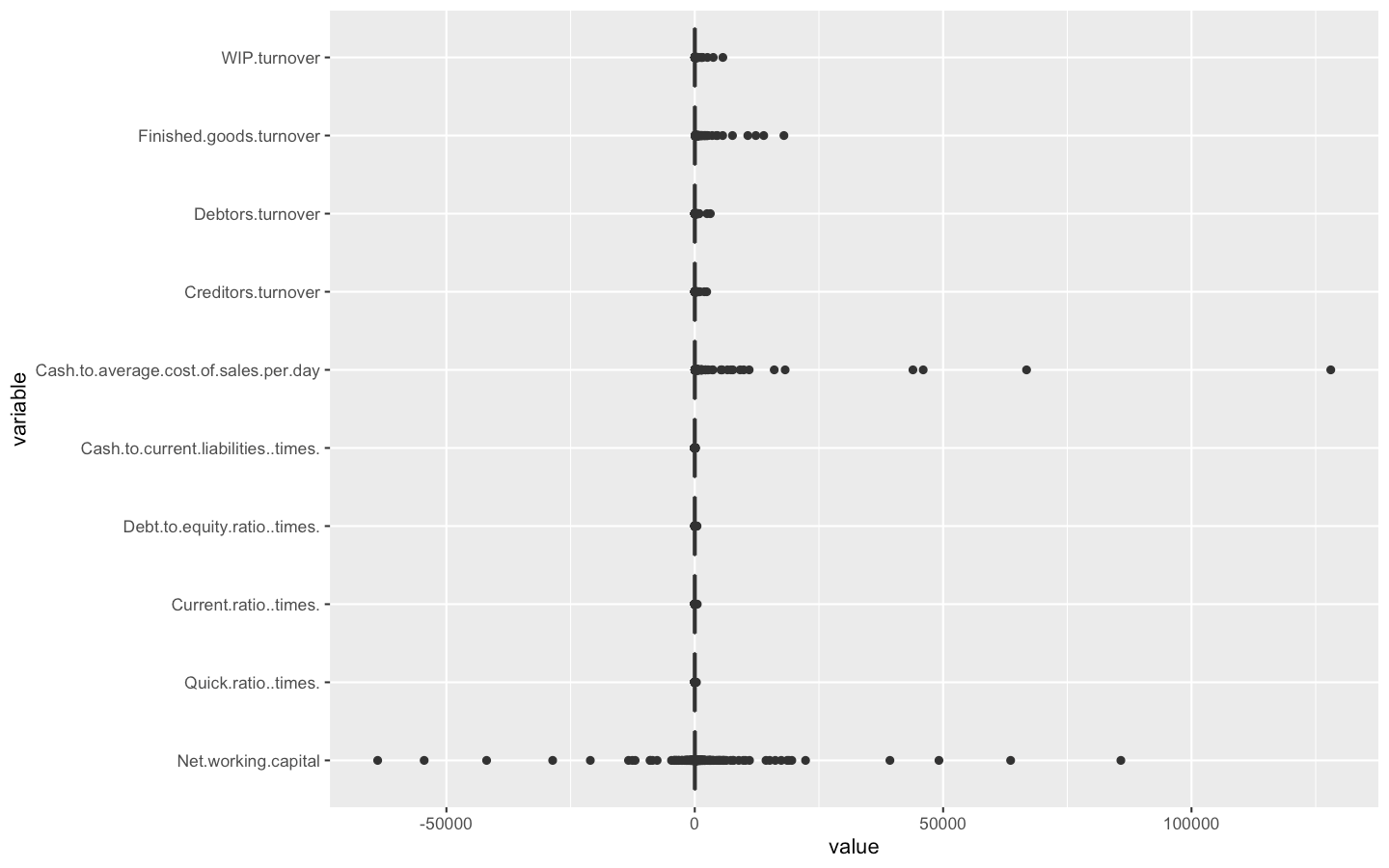
* We remove the below variables from our data on account of significant presence of missing values
  + - Deposits – 100% NAs
    - Contingent liabilities -1188 NAs (34%)
    - Investments -1435 NAs (41%)
    - Deferred Tax liability- 1140 NAs (31%)
    - Other income – 1295 NAs (32%)
    - PE on BSE – 2194 NAs (62%)
* Post this step the number of missing values come down to 7740 from earlier 18533 in our development data

**INVESTIGATING AND TREATING OUTLIERS**

**Boxplots of all variables**

****

****

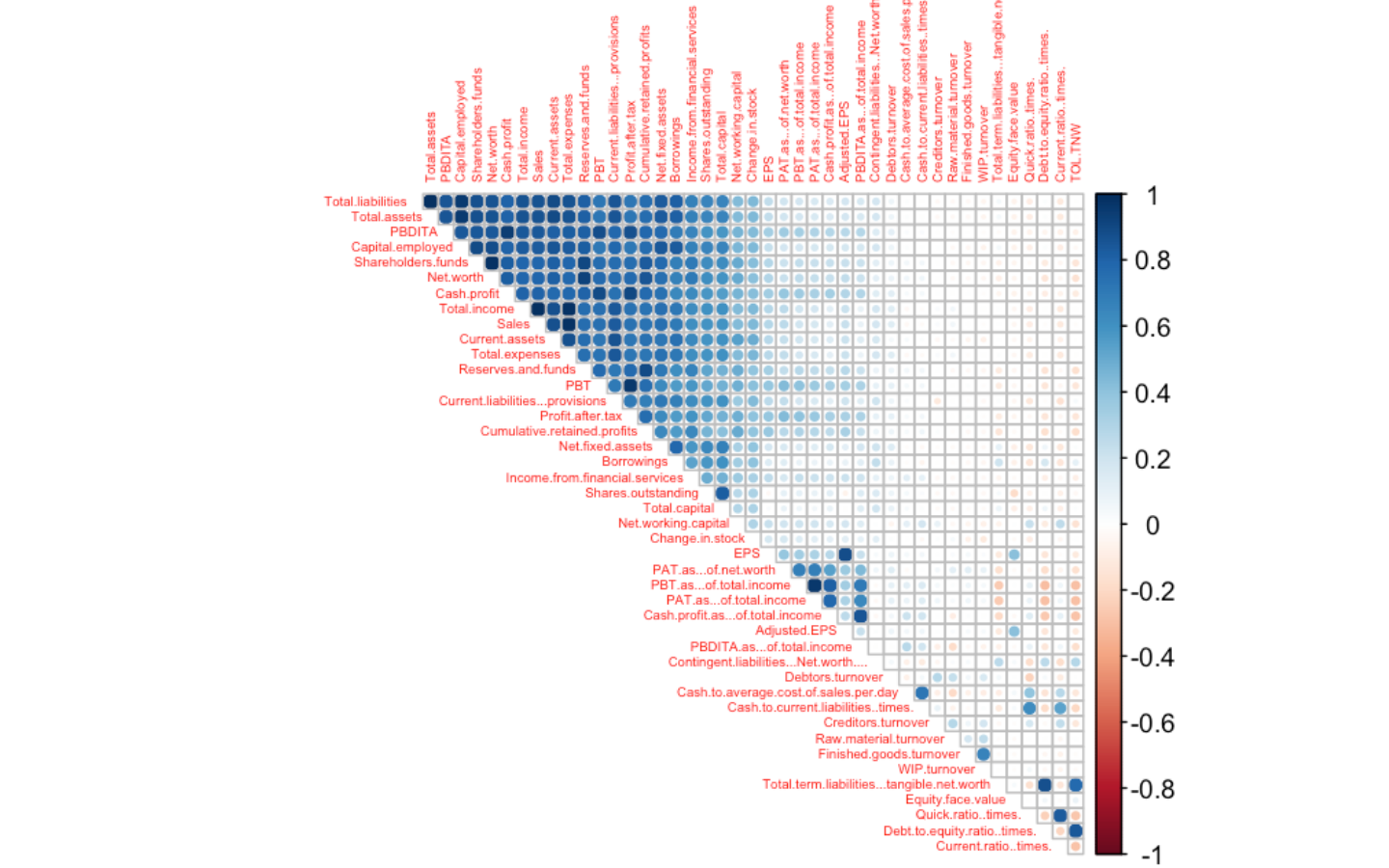
****

* From the above boxplots its evident that except for a few variables, most of our variables have outliers

**OUTLIERS TREATMENT AND MISSING VALUES TREATMENT**

* Treat Outliers by capping the outliers at 5 percentile and 95 percentile values as below
  + X < First quartile - 1.5 IQR -> capped at 5 percentile value of X
  + X > Third quartile + 1.5 IQR -> capped at 95 percentile value of X
* After treating outliers treat Missing Values using KNN imputation
* Arrive at new dataset post performing the above steps without outliers and missing values

**TEST MULTICOLLINEARITY**



* **From the above plot we can see that we have highly correlated data**

**NEXT STEPS**

* Feature engineering – create new ratios inspired by Altmans Z score model:
  + Liquidity - Working capital/Total assets
  + Leverage - Retained earnings/Total assets
  + Profitability - Earning before interest and taxes (EBIT)/Total assets
  + Profitability - Net sales/Total assets (measures asset turnover)
  + Size - Equity/Total assets
* Feature engineering – create new variable Default – based on Net worth Next year which will be our dependent variable for modelling :-
  + NNY < = 0 - Default or 1
  + NNY > 1 - Not default or 0
* Create a new dataset which includes only ratios.
  + Now we have two datasets
    - One with all variables
    - One with only ratio data variables
* Treat multicollinearity in data with all variables - using PCA analysis create factors
* Do same treatment as development data on the validation dataset

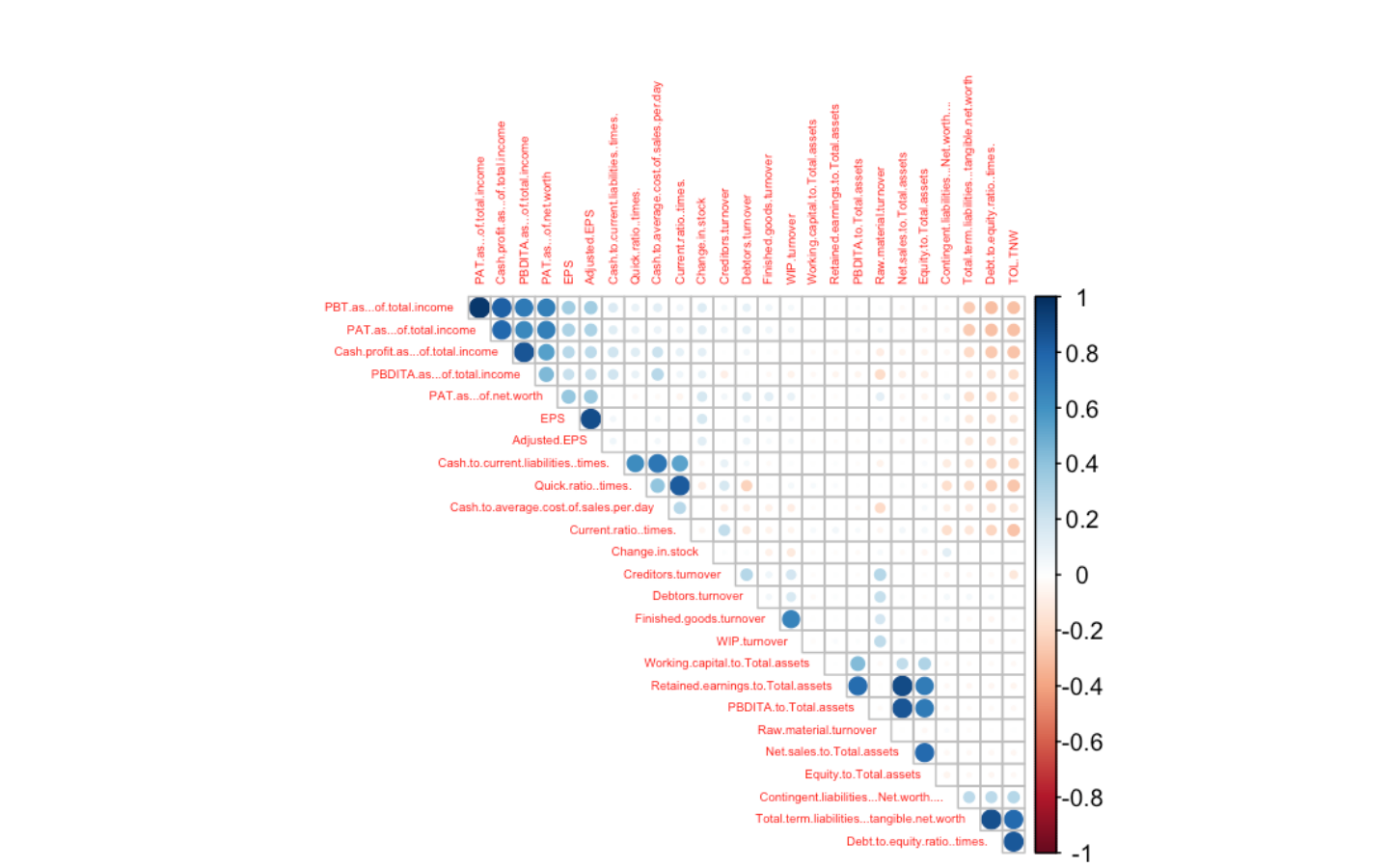
**RATIO DATASET**

data.frame': 3541 obs. of 27 variables:

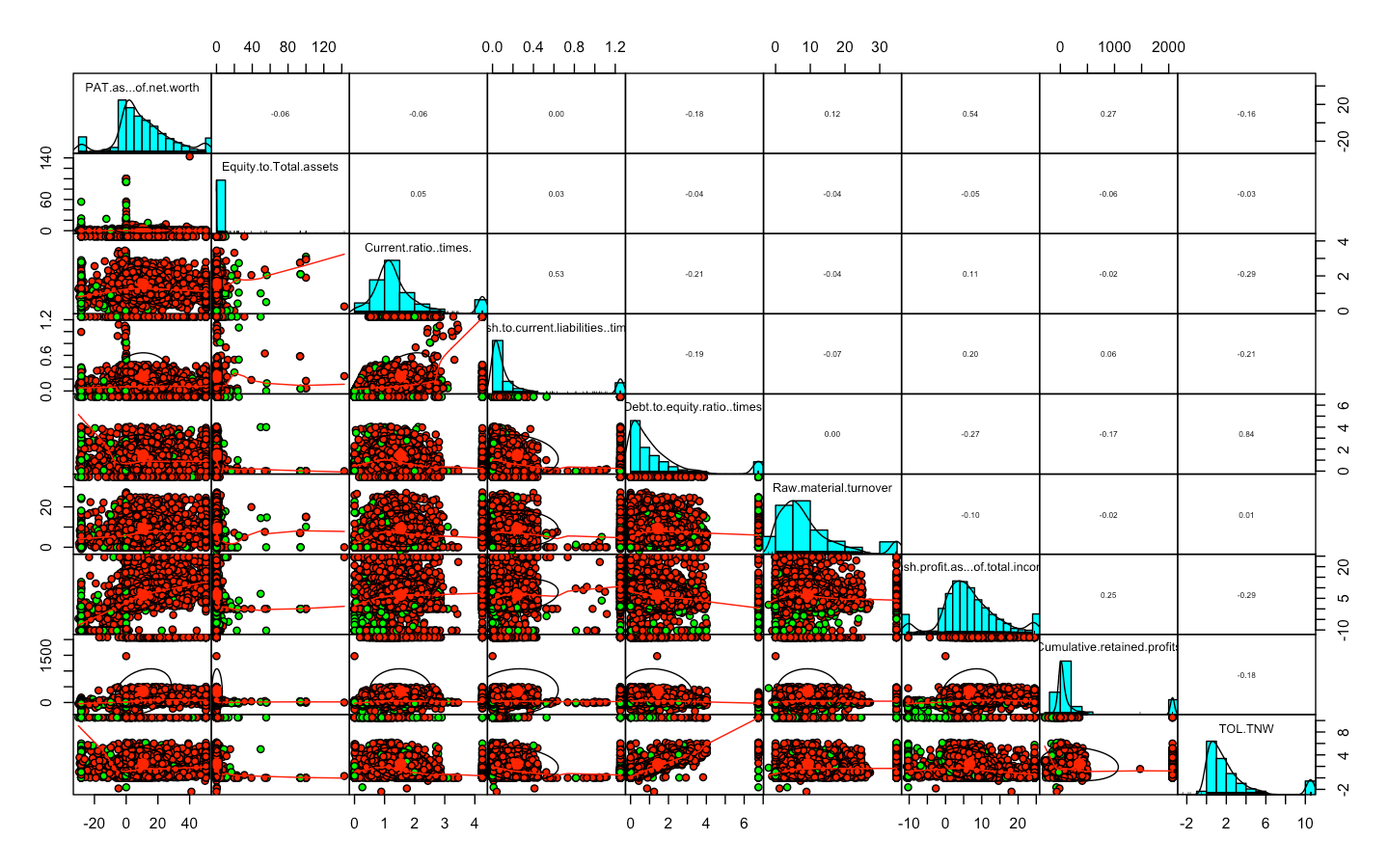
* + **$ Change.in.stock : num 171.68 42.7 -5.2 -2.63 -17 ...**
  + **$ PBDITA.as...of.total.income : num 11.46 18.53 1.22 0 1.96 ...**
  + **$ PBT.as...of.total.income : num 9.68 12.33 -1.38 0 0.4 ...**
  + **$ PAT.as...of.total.income : num 6.18 7.54 -1.38 0 0.35 2.81 0 0.72 8.29 -2.88 ...**
  + **$ 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 ...**
  + **$ 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 ...**
  + **$ Quick.ratio..times. : num 1.18 0.95 1.11 0.998 1.41 ...**
  + **$ Current.ratio..times. : num 1.37 1.56 1.55 1.91 2.54 ...**
  + **$ 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 0.304 0 ...**
  + **$ Cash.to.average.cost.of.sales.per.day : num 190.48 5.96 17.07 46.2 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 201.99 14.21 9.24 9.08 12.68 ...**
  + **$ WIP.turnover : num 21.78 7.49 0.23 5.54 7.9 ...**
  + **$ Raw.material.turnover : num 7.71 11.46 6.58 0 17.03 ...**
  + **$ EPS : num 87.71 9.97 -0.5 0 7.91 ...**
  + **$ Adjusted.EPS : num 7.1 9.97 -0.5 0 7.91 ...**
  + **$ Default : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 2 ...**
  + **$ Working.capital.to.Total.assets : num 0.0803 0.2163 0.256 0.0741 1.4189 ...**
  + **$ Retained.earnings.to.Total.assets : num 0.245 0.263 0.139 0.815 0.173 ...**
  + **$ PBDITA.to.Total.assets : num 0.1464 0.3007 0.0249 2.8157 0.0648 ...**
  + **$ Net.sales.to.Total.assets : num 1.09 1.6 2.04 56.04 3.29 ...**
  + **$ Equity.to.Total.assets : num 0.00118 0.01063 0.04296 3.7037 0.0209 ...**

**TEST MULTICOLLINEARITY IN RATIO DATASET**

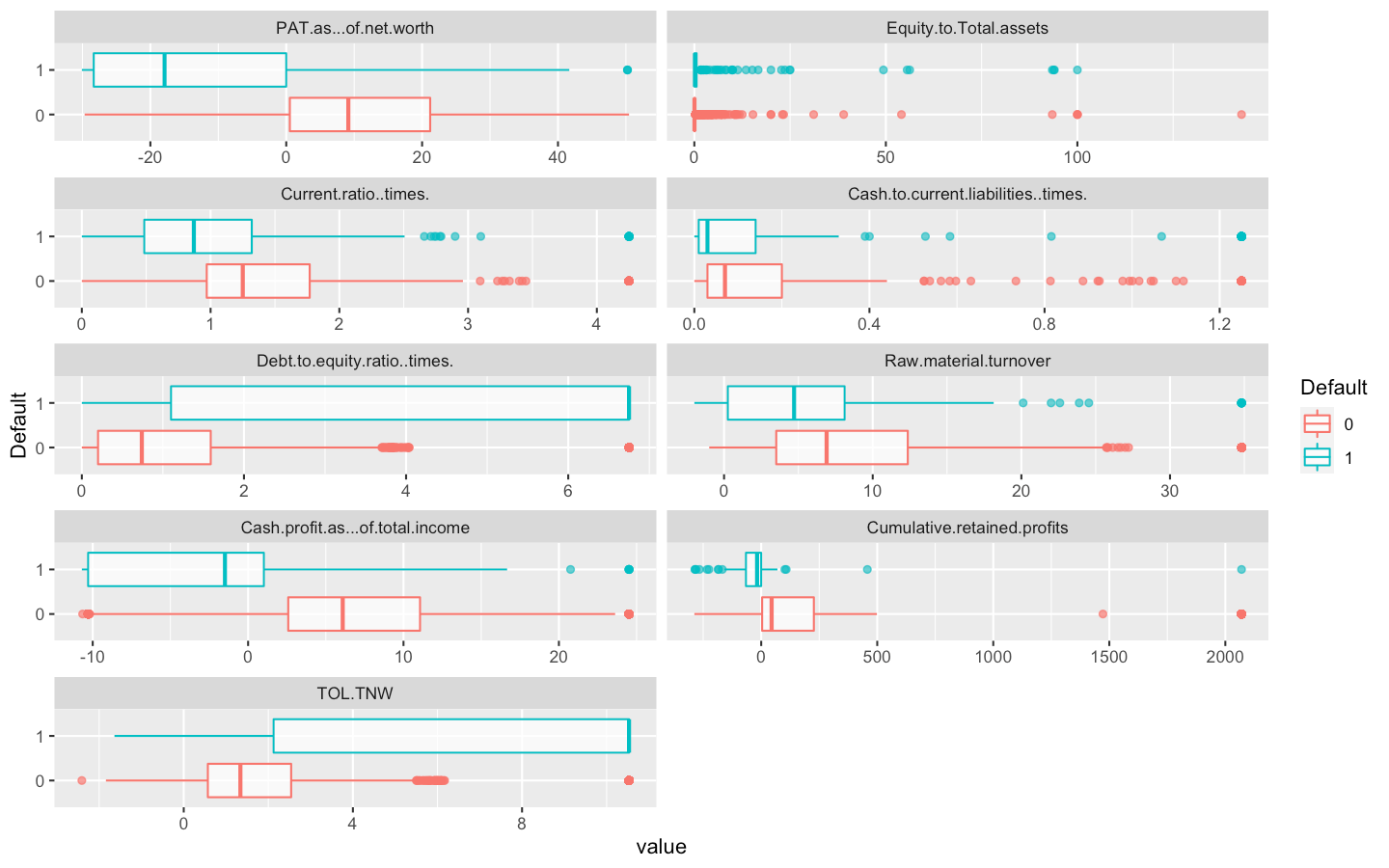
* From the graph below we can deduce that we still have highly correlated data even in our ratio dataset
* We will address multicollinearity during model building stage



***LETS ALSO LOOK AT HOW OUR DEPENDENT VARIABLE DEFAULT IS EXPLAINED BY OUR SIGNIFICANT INDEPENDENT VARIABLES(DERIVED FROM LOGISTIC REGRESSION AT LATER STAGE IN THE PROJECT)***

**PAIR PLOTS OF SIGNIFICANT VARIABLES**

* + All of the important independent variables are not correlated to each other
  + Many of them follow normal distribution such as PAT of net worth, Cash profit of total income, Tol.TNW, Raw material turnover, Current ratio times

**BOX PLOTS OF SIGNIFICANT VARIABLES WRT DEFAULT**

* As expected those companies who defaulted have a higher debt to equity ratio, TOL.TNW ratio and lower cash profit of total income, cumulative retained profits, raw material turnover, cash to current liabilities, Current ratio and PAT as of Net worth compared to companies those who did not Defaulted

**MODEL BUILDING**

**LOGISTIC REGRESSION ON RATIO DATASET**

**BASE MODEL WITH ALL RATIO VARIABLES**

**Call:**

**glm(formula = Default ~ ., family = binomial, data = ratio.dat)**

Call:

glm(formula = Default ~ ., family = binomial, data = ratio.dat)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.2002 -0.2357 -0.1542 -0.0766 3.4135

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -3.0643394 0.2762385 -11.093 < 2e-16

Change.in.stock -0.0015445 0.0019901 -0.776 0.437711

PBDITA.as...of.total.income 0.0013616 0.0137052 0.099 0.920863

PBT.as...of.total.income -0.0206467 0.0307984 -0.670 0.502614

PAT.as...of.total.income -0.0060017 0.0311538 -0.193 0.847234

Cash.profit.as...of.total.income -0.0381110 0.0216196 -1.763 0.077935

PAT.as...of.net.worth -0.0416681 0.0073289 -5.685 1.30e-08

TOL.TNW 0.1819715 0.0387234 4.699 2.61e-06

Total.term.liabilities...tangible.net.worth -0.1450548 0.1053454 -1.377 0.168529

Contingent.liabilities...Net.worth.... 0.0011989 0.0016213 0.739 0.459609

Quick.ratio..times. -0.3626669 0.2717056 -1.335 0.181949

Current.ratio..times. -0.2876134 0.1768350 -1.626 0.103854

Debt.to.equity.ratio..times. 0.2386450 0.0737291 3.237 0.001209

Cash.to.current.liabilities..times. 1.3222500 0.3626331 3.646 0.000266

Cash.to.average.cost.of.sales.per.day 0.0008898 0.0018589 0.479 0.632157

Creditors.turnover 0.0008681 0.0095529 0.091 0.927590

Debtors.turnover -0.0098545 0.0087063 -1.132 0.257684

Finished.goods.turnover 0.0014493 0.0025426 0.570 0.568670

WIP.turnover -0.0037254 0.0069791 -0.534 0.593490

Raw.material.turnover -0.0181008 0.0128996 -1.403 0.160554

EPS -0.0247329 0.0198431 -1.246 0.212608

Adjusted.EPS 0.0007237 0.0176971 0.041 0.967380

Working.capital.to.Total.assets 0.0118071 0.0237085 0.498 0.618476

Retained.earnings.to.Total.assets -0.0087806 0.0082328 -1.067 0.286184

PBDITA.to.Total.assets 0.0419191 0.0448964 0.934 0.350467

Net.sales.to.Total.assets 0.0006953 0.0013435 0.517 0.604816

Equity.to.Total.assets 0.0439975 0.0106152 4.145 3.40e-05

(Intercept) \*\*\*

Change.in.stock

PBDITA.as...of.total.income

PBT.as...of.total.income

PAT.as...of.total.income

Cash.profit.as...of.total.income .

PAT.as...of.net.worth \*\*\*

TOL.TNW \*\*\*

Total.term.liabilities...tangible.net.worth

Contingent.liabilities...Net.worth....

Quick.ratio..times.

Current.ratio..times.

Debt.to.equity.ratio..times. \*\*

Cash.to.current.liabilities..times. \*\*\*

Cash.to.average.cost.of.sales.per.day

Creditors.turnover

Debtors.turnover

Finished.goods.turnover

WIP.turnover

Raw.material.turnover

EPS

Adjusted.EPS

Working.capital.to.Total.assets

Retained.earnings.to.Total.assets

PBDITA.to.Total.assets

Net.sales.to.Total.assets

Equity.to.Total.assets \*\*\*

---

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1770.97 on 3540 degrees of freedom

Residual deviance: 979.82 on 3514 degrees of freedom

AIC: 1033.8

Number of Fisher Scoring iterations: 8

**VIF FOR BASE MODEL**

Change.in.stock

1.078248

PBDITA.as...of.total.income

2.031608

PBT.as...of.total.income

15.946560

PAT.as...of.total.income

15.236500

Cash.profit.as...of.total.income

3.623298

PAT.as...of.net.worth

2.064053

TOL.TNW

3.670263

Total.term.liabilities...tangible.net.worth

4.557649

Contingent.liabilities...Net.worth....

1.192191

Quick.ratio..times.

4.442727

Current.ratio..times.

3.764861

Debt.to.equity.ratio..times.

5.538063

Cash.to.current.liabilities..times.

2.864483

Cash.to.average.cost.of.sales.per.day

2.114208

Creditors.turnover

1.388477

Debtors.turnover

1.372025

Finished.goods.turnover

2.072297

WIP.turnover

2.268160

Raw.material.turnover

1.394500

EPS

3.806831

Adjusted.EPS

3.697017

Working.capital.to.Total.assets

1.212216

Retained.earnings.to.Total.assets

9.587798

PBDITA.to.Total.assets

5.221739

Net.sales.to.Total.assets

10.737335

Equity.to.Total.assets

2.640693

* From the above summary for base model, we can see that VIF >3.5 for many variables,
* Many insignificant variables are present in our base model as can be deduced from p value of independent variables in the model summary
* To optimize our model we run a stepwise AIC both analysis on our base model
  + both <- blr\_step\_aic\_both(default.model6, details = TRUE)
* Below is the summary of best model as estimated by stepwise AIC both analysis
* We run another model with all the independent variables as derived in step AIC model except for PBT.as.of .total income because of its low pvalue indicating it is not significant in the model

**STEP AIC SUMMARY**

Maximum Likelihood Estimates

------------------------------------------------------------------------------------------

Parameter DF Estimate Std. Error z value Pr(>|z|)

------------------------------------------------------------------------------------------

(Intercept) 1 -3.0077 0.2217 -13.5684 0.0000

PAT.as...of.net.worth 1 -0.0413 0.0071 -5.8356 0.0000

TOL.TNW 1 0.1748 0.0373 4.6870 0.0000

Equity.to.Total.assets 1 0.0478 0.0066 7.2133 0.0000

PBT.as...of.total.income 1 -0.0244 0.0138 -1.7710 0.0766

EPS 1 -0.0242 0.0109 -2.2137 0.0268

Current.ratio..times. 1 -0.4595 0.1095 -4.1952 0.0000

Cash.to.current.liabilities..times. 1 1.1417 0.2608 4.3779 0.0000

Debt.to.equity.ratio..times. 1 0.1702 0.0567 2.9984 0.0027

Raw.material.turnover 1 -0.0262 0.0114 -2.3038 0.0212

Cash.profit.as...of.total.income 1 -0.0407 0.0184 -2.2094 0.0271

**OPTIMISED MODEL BASED ON STEPWISE BOTH AIC METHOD**

**default.model8 = glm(Default~ PAT.as...of.net.worth + TOL.TNW + Equity.to.Total.assets+EPS + Current.ratio..times. +Cash.to.current.liabilities..times.+Debt.to.equity.ratio..times. + Raw.material.turnover+Cash.profit.as...of.total.income,data = ratio.dat,family = binomial)**

**VIF SUMMARY**

PAT.as...of.net.worth TOL.TNW

1.637730 3.343364

Equity.to.Total.assets EPS

1.054170 1.174277

Current.ratio..times. Cash.to.current.liabilities..times.

1.430761 1.452599

Debt.to.equity.ratio..times. Raw.material.turnover

3.244442 1.042300

Cash.profit.as...of.total.income

1.500739

* All VIF values are well with in tolerable range.
* Lets also look at the summary of this model to identify any further possibility of model optimization

**MODEL SUMMARY**

Call:

glm(formula = Default ~ PAT.as...of.net.worth + TOL.TNW + Equity.to.Total.assets +

EPS + Current.ratio..times. + Cash.to.current.liabilities..times. +

Debt.to.equity.ratio..times. + Raw.material.turnover + Cash.profit.as...of.total.income,

family = binomial, data = ratio.dat)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.0564 -0.2432 -0.1555 -0.0785 3.4392

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -2.886939 0.207656 -13.903 < 2e-16 \*\*\*

PAT.as...of.net.worth -0.046299 0.006594 -7.021 2.20e-12 \*\*\*

TOL.TNW 0.168199 0.036914 4.556 5.20e-06 \*\*\*

Equity.to.Total.assets 0.046901 0.006575 7.133 9.79e-13 \*\*\*

EPS -0.024493 0.011038 -2.219 0.02649 \*

Current.ratio..times. -0.451232 0.108992 -4.140 3.47e-05 \*\*\*

Cash.to.current.liabilities..times. 1.143397 0.256827 4.452 8.51e-06 \*\*\*

Debt.to.equity.ratio..times. 0.177720 0.056323 3.155 0.00160 \*\*

Raw.material.turnover -0.029222 0.011293 -2.588 0.00967 \*\*

Cash.profit.as...of.total.income -0.062368 0.014206 -4.390 1.13e-05 \*\*\*

---

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1770.97 on 3540 degrees of freedom

Residual deviance: 992.09 on 3531 degrees of freedom

AIC: 1012.1

Number of Fisher Scoring iterations: 8

* **We can further look at removing the EPS variable which is the least important in the model based on its P value**

**HENCE OUR FINAL RATIO MODEL**

Call:

glm(formula = Default ~ PAT.as...of.net.worth + TOL.TNW + Equity.to.Total.assets +

Current.ratio..times. + Cash.to.current.liabilities..times. +

Debt.to.equity.ratio..times. + Raw.material.turnover + Cash.profit.as...of.total.income,

family = binomial, data = ratio.dat)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.0793 -0.2400 -0.1592 -0.0977 3.2108

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -2.944825 0.208528 -14.122 < 2e-16 \*\*\*

PAT.as...of.net.worth -0.051583 0.006363 -8.106 5.22e-16 \*\*\*

TOL.TNW 0.165531 0.037157 4.455 8.39e-06 \*\*\*

Equity.to.Total.assets 0.047815 0.006589 7.257 3.96e-13 \*\*\*

Current.ratio..times. -0.455926 0.109882 -4.149 3.34e-05 \*\*\*

Cash.to.current.liabilities..times. 1.132283 0.255198 4.437 9.13e-06 \*\*\*

Debt.to.equity.ratio..times. 0.185383 0.056715 3.269 0.00108 \*\*

Raw.material.turnover -0.030582 0.011407 -2.681 0.00734 \*\*

Cash.profit.as...of.total.income -0.067114 0.014265 -4.705 2.54e-06 \*\*\*

---

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1771.0 on 3540 degrees of freedom

Residual deviance: 1001.1 on 3532 degrees of freedom

AIC: 1019.1

Number of Fisher Scoring iterations: 7

**VIF VALUES FOR FINAL MODEL**

PAT.as...of.net.worth TOL.TNW

1.556341 3.370649

Equity.to.Total.assets Current.ratio..times.

1.053606 1.416715

Cash.to.current.liabilities..times. Debt.to.equity.ratio..times.

1.435904 3.276690

Raw.material.turnover Cash.profit.as...of.total.income

1.041526 1.509818

**INTERPRETATION OF FINAL RATIO MODEL**

* All variables are significant with a p value much lower than 0.05 in final model
* AIC value is also lower than our base model
* Our final model does not suffer from multicollinearity as indicated by VIF values of both models. VIF gauges how much the variance of regression coefficient is inflated due to multicolinearity.

**Coefficients/Odds Ratio final ratio model**

(Intercept)

0.05261124

PAT.as...of.net.worth

0.94972468

TOL.TNW

1.18001902

Equity.to.Total.assets

1.04897618

Current.ratio..times.

0.63386056

Cash.to.current.liabilities..times.

3.10273099

Debt.to.equity.ratio..times.

1.20367976

Raw.material.turnover

0.96988049

Cash.profit.as...of.total.income

0.93508863

**Probability final ratio model**

(Intercept)

0.04998164

PAT.as...of.net.worth

0.48710707

TOL.TNW

0.54128841

Equity.to.Total.assets

0.51195138

Current.ratio..times.

0.38795267

Cash.to.current.liabilities..times.

0.75625991

Debt.to.equity.ratio..times.

0.54621356

Raw.material.turnover

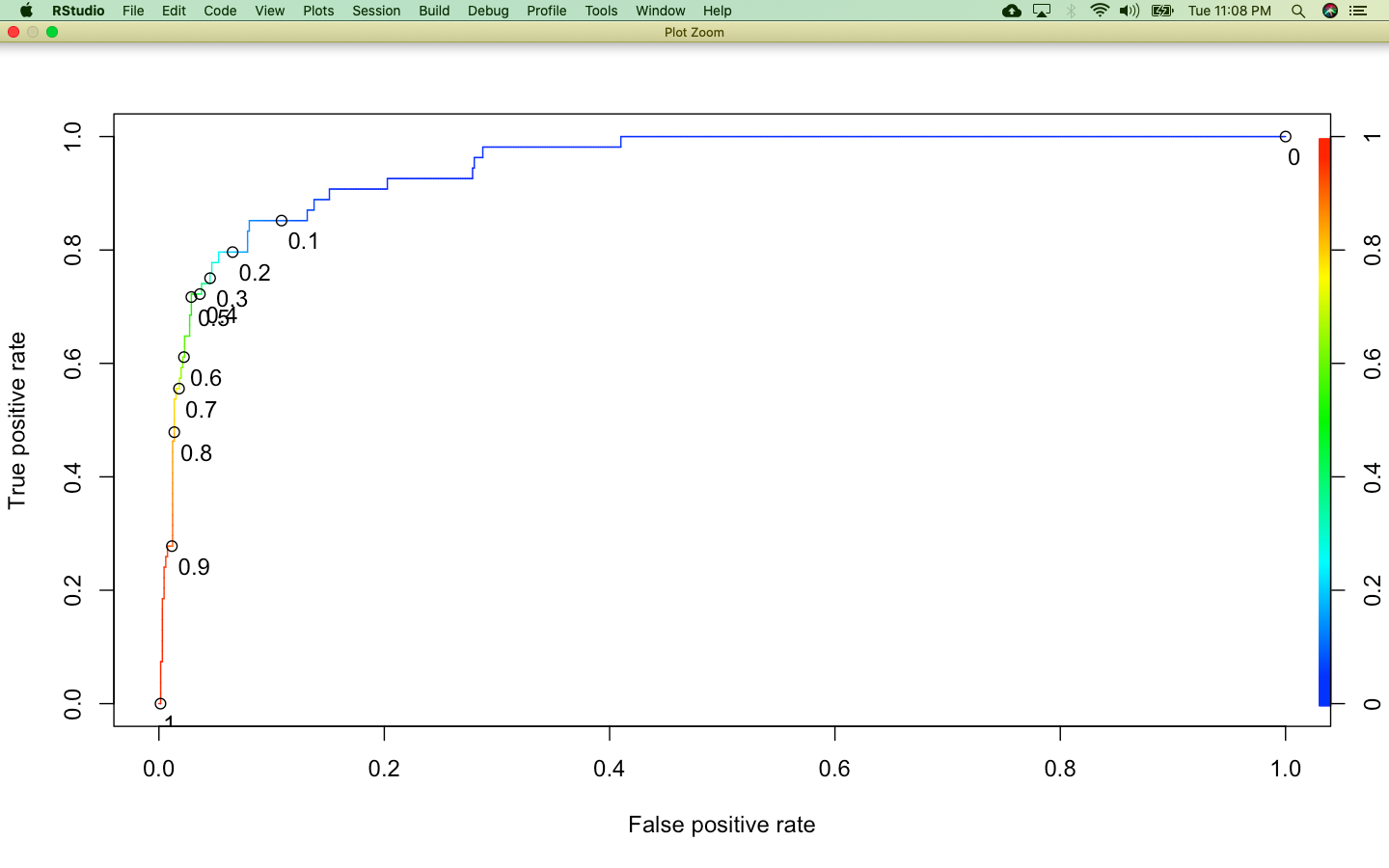
0.49235499

Cash.profit.as...of.total.income

0.48322780

* The odds ratio and probability are helpful in explaining the impact of each independent variable on the dependent variable. For example if Debt to Equity ratio is increased by 1 unit then probability of default = 1 will be increased by 54.6%.Similarly for cash to current liabilities , if it increases by 1 unit then then probability of default = 1 will be decreased by 75.6%,

**ROC CURVE TO DETERMINE THE THRESHOLD AND AUC OF FINAL RATIO MODEL**

****

AUC - AUC of a classifier is equal to the probability that the classifier will rank a randomly chosen positive data point higher than a randomly chosen negative data point. Higher the probability better is the classifier.

AUC

1] 0.9443884

**Threshold -** Since the objective is to maximize the sensitivity, based on the above ROC graph we take a threshold of 0.2 False Positive rate, we arrive at our threshold for cutoff as 0.10, which will maximize our sensitivity

**CONFUSION MATRIX AND PERFORMANCE EVALUATION BETWEEN TRAIN AND TEST DATA FOR OUR FINAL RATIO MODEL**

**Confusion Matrix and Statistics on Test data**

Reference

Prediction 0 1

0 589 8

1 72 46

Accuracy : 0.8881

95% CI : (0.8627, 0.9103)

No Information Rate : 0.9245

P-Value [Acc > NIR] : 0.9998

Kappa : 0.4811

Mcnemar's Test P-Value : 1.873e-12

Sensitivity : 0.85185

Specificity : 0.89107

Pos Pred Value : 0.38983

Neg Pred Value : 0.98660

Prevalence : 0.07552

Detection Rate : 0.06434

Detection Prevalence : 0.16503

Balanced Accuracy : 0.87146

'Positive' Class : 1

**Confusion Matrix and Statistics on Train Data**

Reference

Prediction 0 1

0 3039 58

1 259 185

Accuracy : 0.9105

95% CI : (0.9006, 0.9197)

No Information Rate : 0.9314

P-Value [Acc > NIR] : 1

Kappa : 0.4937

Mcnemar's Test P-Value : <2e-16

Sensitivity : 0.76132

Specificity : 0.92147

Pos Pred Value : 0.41667

Neg Pred Value : 0.98127

Prevalence : 0.06862

Detection Rate : 0.05225

Detection Prevalence : 0.12539

Balanced Accuracy : 0.84139

'Positive' Class : 1

**DEFINING THE CONFUSION MATRIX EVALUATION MEASURES**

* **Sensitivity** - Total no correct predictions of 1 out of total predictions of 1
* **Specificity** - Total no correct predictions of 0 out of total predictions of o
* **Accuracy** - Rati0 of correct predictions to total observations

**INTERPRETING THE CONFUSION MATRIX RESULTS**

* **Accuracy of 91% for Train and 88.8% for Test Model indicates a good fit model.**
* Even sensitivity **is well predicted for Test data at 85% and 76% for Train data**
* Sensitivity, Specificity, Accuracy, are all with in 10% difference for Test and Train data indicating a valid model

**DISTRIBUTING THE PREDICTED DEFAULT PROBABILITIES ON TEST DATA INTO DECILES AND RANKING THEM**

**DECILES**

0% 10% 20% 30% 40% 50%

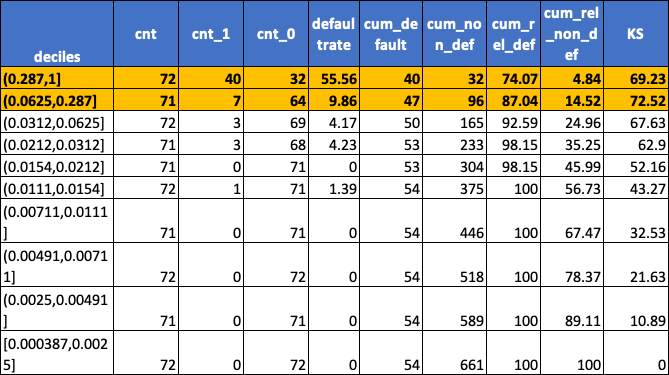
0.0003874815 0.0025049775 0.0049131393 0.0071148993 0.0110950484 0.0153577806

60% 70% 80% 90% 100%

0.0211833122 0.0311693251 0.0624587532 0.2867901876 0.9998907077

* We distributed our data in deciles from 0 to 100% at an interval of 10% based on predicted default probabilities to understand their distribution
* Only 10% of our test data have high predicted Default probabilities in the range of 28.6% to 99.9%

**RANK TABLE**

****

* We arranged the data as per decile distribution based on predicted Default values and calculated additional measures as below mentioned
* Default rate = (cnt\_1/ cnt)\*100
* cum\_Default = cumsum(cnt\_1)
* cum\_non\_Default = cumsum(cnt\_0)
* cum\_rel\_Default = cum\_Default/sum(cnt\_1)\*100
* cum\_rel\_non\_ Default = cum\_non\_Default/sum(cnt\_0)\*100
* KS = cum\_rel\_ Default - cum\_rel\_ non\_ Default
* In the topmost 90 – 100 % decile distribution of our Default probabilities, our Default rate is the highest - 56% .The companies in this decile range are most likely to default and hence should be addressed first
* The highest KS value is 72, which is the gap between cumulative default rate and cumulative non default rate. This is widest for top two deciles , indicating that we should prioritise to take action on companies within these top two deciles as they have the highest likelihood to Default in future

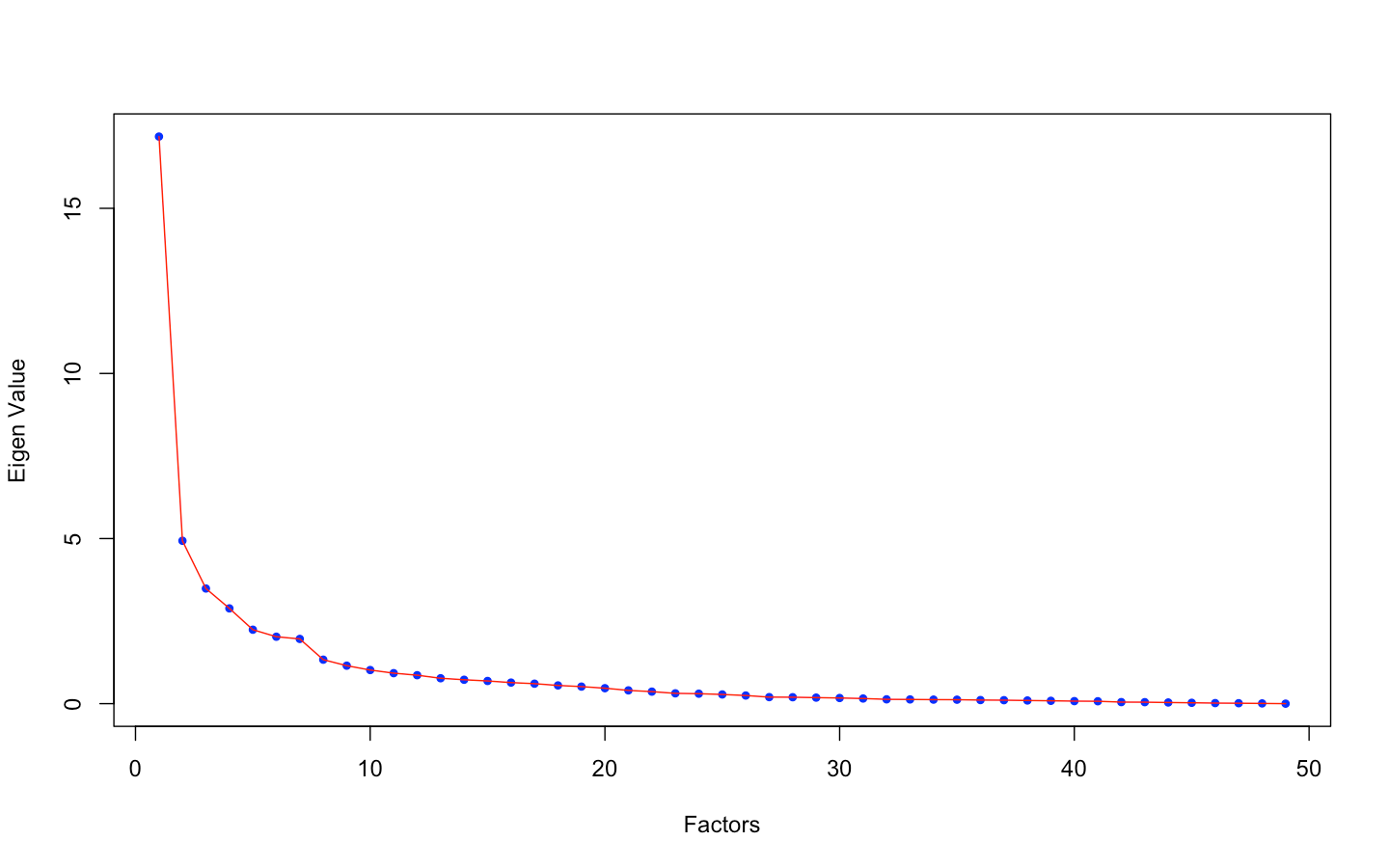
**MODEL BUILDING – LOGISTIC REGRESSION ON FULL DATA USING FACTORS**

**FACTOR ANALYSIS TO REMOVE MULTICOLLINEARITY**

**STEPS**

* Create factors using PCA varimax rotation on full data
* Check the scree plot and arrive at number of factors taking at cutoff at Eigen value >1 – in our case its 4
* Create a dataset with 4 factors and Default variable
* Apply the above steps on Test data too

**SCREE PLOT OF FACTORS**

****

* We will go ahead with 4 factors solution to align with the 4 buckets as per our understanding of financial measures in real world scenario

**VARIANCE EXPLAINED BY FACTOR SOLUTION**

MR1 MR2 MR3 MR4

SS loadings 16.41 4.21 3.24 3.22

Proportion Var 0.33 0.09 0.07 0.07

Cumulative Var 0.33 0.42 0.49 0.55

Proportion Explained 0.61 0.16 0.12 0.12

Cumulative Proportion 0.61 0.76 0.88 1.00

* 55% of variance is captured by our Factor solution which is good to go ahead

**FACTOR LOADINGS**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | MR1 | MR2 | MR3 | MR4 |
| <S3: AsIs> | <S3: AsIs> | <S3: AsIs> | <S3: AsIs> |
| Total.assets | 0.96 | 0.01 | -0.02 | -0.06 |
| Total.liabilities | 0.96 | 0.01 | -0.02 | -0.06 |
| Capital.employed | 0.94 | 0.01 | -0.02 | -0.04 |
| Shareholders.funds | 0.92 | 0.09 | -0.02 | 0.09 |
| Net.worth | 0.92 | 0.1 | -0.02 | 0.11 |
| Current.assets | 0.92 | 0.03 | -0.02 | -0.02 |
| Total.income | 0.92 | 0.06 | -0.01 | -0.08 |
| Sales | 0.91 | 0.06 | -0.01 | -0.08 |
| Total.expenses | 0.91 | 0.04 | -0.01 | -0.09 |
| PBDITA | 0.91 | 0.22 | -0.01 | -0.01 |
| Cash.profit | 0.87 | 0.27 | -0.01 | 0.02 |
| Reserves.and.funds | 0.87 | 0.15 | -0.01 | 0.11 |
| Current.liabilities...provisions | 0.87 | 0.05 | -0.02 | -0.1 |
| Net.fixed.assets | 0.84 | 0.02 | -0.02 | -0.12 |
| Cumulative.retained.profits | 0.82 | 0.2 | -0.01 | 0.12 |
| PBT | 0.81 | 0.33 | -0.01 | 0.05 |
| Profit.after.tax | 0.81 | 0.34 | 0 | 0.05 |
| Borrowings | 0.79 | -0.05 | -0.03 | -0.18 |
| Income.from.financial.services | 0.71 | 0.1 | -0.02 | 0.08 |
| Shares.outstanding | 0.69 | -0.07 | -0.01 | -0.01 |
| Total.capital | 0.68 | -0.09 | -0.04 | -0.06 |
| Net.working.capital | 0.5 | 0.07 | -0.02 | 0.28 |
| Change.in.stock | 0.48 | 0.07 | -0.02 | -0.05 |
| PBT.as...of.total.income | 0.17 | 0.89 | -0.02 | 0.19 |
| PAT.as...of.total.income | 0.15 | 0.86 | -0.01 | 0.18 |
| Cash.profit.as...of.total.income | 0.15 | 0.75 | -0.06 | 0.25 |
| PAT.as...of.net.worth | 0.21 | 0.71 | -0.02 | -0.02 |
| PBDITA.as...of.total.income | 0.15 | 0.63 | -0.09 | 0.19 |
| EPS | 0.26 | 0.46 | -0.02 | 0.02 |
| Adjusted.EPS | 0.2 | 0.46 | -0.02 | 0.03 |
| Finished.goods.turnover | -0.05 | 0.12 | 0.02 | -0.07 |
| WIP.turnover | -0.06 | 0.1 | 0.03 | -0.06 |
| Equity.face.value | -0.07 | 0.1 | 0 | -0.04 |
| Net.sales.to.Total.assets | -0.02 | -0.02 | 0.98 | 0.01 |
| PBDITA.to.Total.assets | -0.01 | 0.03 | 0.9 | -0.01 |
| Retained.earnings.to.Total.assets | 0 | 0 | 0.85 | 0.03 |
| Equity.to.Total.assets | -0.05 | -0.03 | 0.8 | 0.04 |
| Working.capital.to.Total.assets | -0.01 | -0.01 | 0.28 | 0.03 |
| Quick.ratio..times. | -0.06 | -0.09 | 0 | 0.77 |
| Current.ratio..times. | -0.07 | -0.1 | 0.03 | 0.71 |
| Cash.to.current.liabilities..times. | 0.03 | -0.01 | -0.02 | 0.65 |
| TOL.TNW | -0.04 | -0.27 | -0.06 | -0.6 |
| Debt.to.equity.ratio..times. | -0.01 | -0.29 | -0.06 | -0.57 |
| Total.term.liabilities...tangible.net.worth | 0.02 | -0.26 | -0.06 | -0.49 |
| Cash.to.average.cost.of.sales.per.day | 0.04 | 0.01 | -0.03 | 0.49 |
| Contingent.liabilities...Net.worth.... | 0.18 | -0.03 | -0.04 | -0.3 |
| Debtors.turnover | 0.09 | 0.13 | 0.02 | -0.14 |
| Creditors.turnover | -0.01 | 0.02 | 0 | 0.12 |
| Raw.material.turnover | -0.02 | 0.05 | 0 | -0.1 |

Basis above loadings we categorize all variables in 4 buckets as below:

**MR1 – Size**

**MR 2 – Profitability**

**MR 3 – Leverage**

**MR 4 - Liquidity**

**LOGISTIC REGRESSION ON FULL DATA USING FACTORS - MODEL SUMMARY**

Call:

glm(formula = Default ~ ., family = binomial, data = final\_data)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.1014 -0.2879 -0.1854 -0.0684 3.9583

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -4.08697 0.15218 -26.856 < 2e-16 \*\*\*

Company.size -1.05510 0.16728 -6.307 2.84e-10 \*\*\*

Profitability -1.52821 0.08677 -17.611 < 2e-16 \*\*\*

Leverage 0.21094 0.04011 5.259 1.45e-07 \*\*\*

Liquidity -1.21230 0.09820 -12.346 < 2e-16 \*\*\*

---

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1771.0 on 3540 degrees of freedom

Residual deviance: 1119.1 on 3536 degrees of freedom

AIC: 1129.1

Number of Fisher Scoring iterations: 7

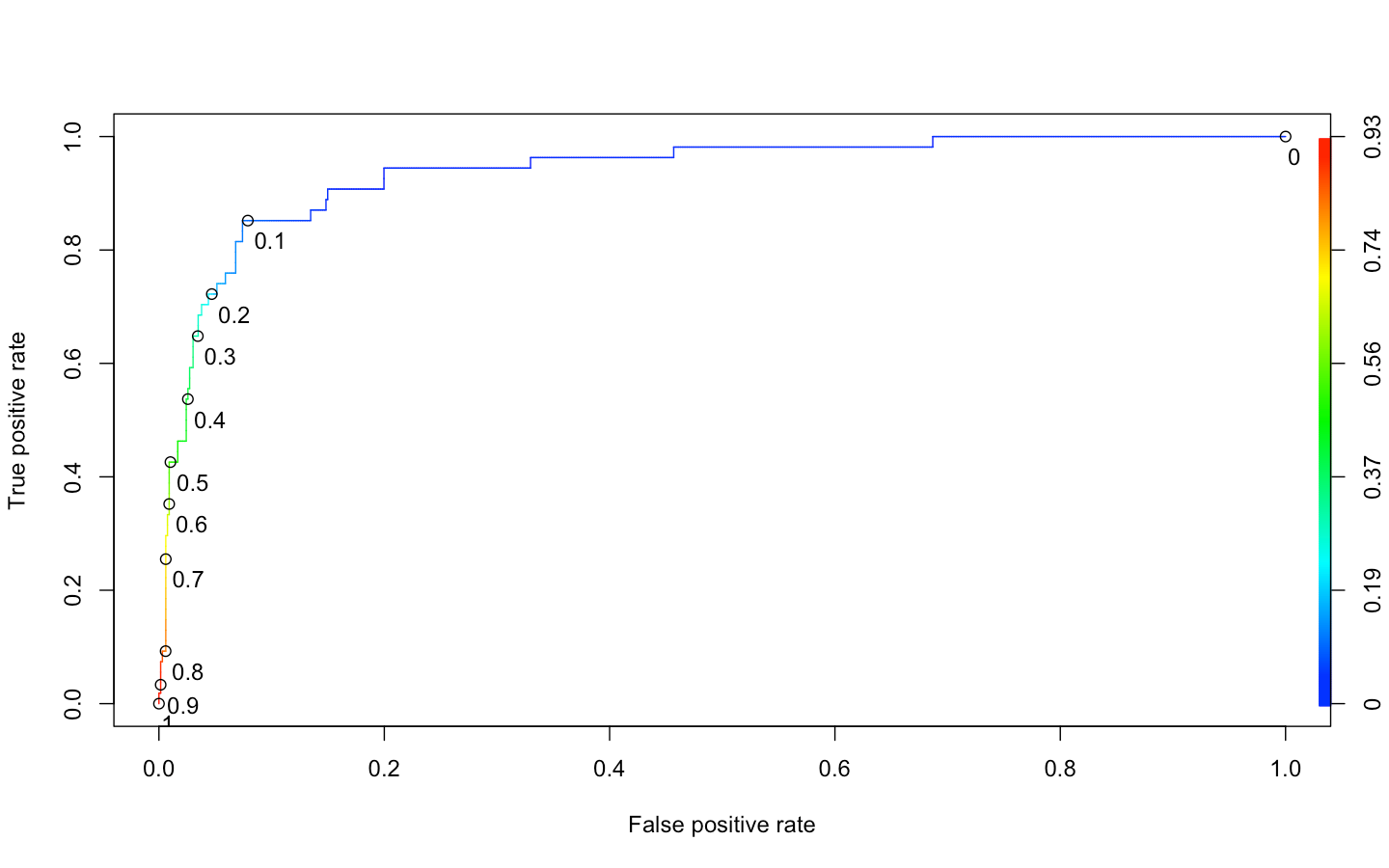
**VIF SUMMARY**

Company.size Profitability Leverage Liquidity

1.206728 1.187903 1.020899 1.089289

* From the above summary it is evident that all our variables are significant in explaining the model
* Also the VIF is well below 4, hence the multicollinearity is addressed by factors

**ROC CURVE TO DETERMINE THE THRESHOLD AND AUC**

****

* The cutoff threshold is determined at 10% from the above ROC curve

**AUC**

**[1] 0.9385331**

**CONFUSION MATRIX AND PERFORMANCE EVALUATION ON TEST DATA FOR OUR FACTOR SOLUTION MODEL**

**Confusion Matrix and Statistics on Test data**

Reference

Prediction 0 1

0 603 8

1 58 46

Accuracy : 0.9077

95% CI : (0.8841, 0.9279)

No Information Rate : 0.9245

P-Value [Acc > NIR] : 0.9583

Kappa : 0.5362

Mcnemar's Test P-Value : 1.625e-09

Sensitivity : 0.85185

Specificity : 0.91225

Pos Pred Value : 0.44231

Neg Pred Value : 0.98691

Prevalence : 0.07552

Detection Rate : 0.06434

Detection Prevalence : 0.14545

Balanced Accuracy : 0.88205

'Positive' Class : 1

* Accuracy and specificity of this model is slightly higher than our final ratio model

**LOGISTIC REGRESSION ON FULL DATA USING ALL VARIABLES TO IDENTIFY IMPORTANT VARIABLES**

default.model3 =glm(Default~.,data = data.wo.nas[,-c(1,2)],family = binomial)

**SIGNIFICANT VARIABLES**

(Intercept) -9.543 < 2e-16 \*\*\*

Total.assets 1.366 0.171974

Net.worth 0.056 0.955077

Total.income 1.321 0.186417

Change.in.stock 0.987 0.323700

Total.expenses -0.393 0.694431

Profit.after.tax 0.983 0.325510

PBDITA -1.823 0.068244 .

PBT 0.014 0.989009

Cash.profit -1.646 0.099672 .

PBDITA.as...of.total.income -0.307 0.758819

PBT.as...of.total.income -0.585 0.558669

PAT.as...of.total.income -0.369 0.712319

Cash.profit.as...of.total.income -0.887 0.374939

PAT.as...of.net.worth -5.548 2.89e-08 \*\*\*

Sales -0.787 0.431028

Income.from.financial.services 1.383 0.166629

Total.capital -1.882 0.059844 .

Reserves.and.funds -0.883 0.377261

Borrowings -0.948 0.342925

Current.liabilities...provisions 0.646 0.518201

Shareholders.funds 0.793 0.427738

Cumulative.retained.profits -2.686 0.007226 \*\*

Capital.employed -0.401 0.688424

TOL.TNW 2.986 0.002827 \*\*

Total.term.liabilities...tangible.net.worth -1.047 0.295183

Contingent.liabilities...Net.worth.... 1.266 0.205639

Net.fixed.assets -0.574 0.565833

Current.assets -1.628 0.103548

Net.working.capital 0.739 0.460120

Quick.ratio..times. -1.375 0.169006

Current.ratio..times. -1.790 0.073452 .

Debt.to.equity.ratio..times. 3.881 0.000104 \*\*\*

Cash.to.current.liabilities..times. 3.646 0.000266 \*\*\*

Cash.to.average.cost.of.sales.per.day 0.132 0.894683

Creditors.turnover 0.122 0.902575

Debtors.turnover -1.228 0.219282

Finished.goods.turnover 0.476 0.634420

WIP.turnover -0.678 0.497865

Raw.material.turnover -1.330 0.183438

Shares.outstanding 0.905 0.365730

Equity.face.value 0.726 0.467587

EPS -0.668 0.504061

Adjusted.EPS -0.142 0.886831

Total.liabilities NA NA

Working.capital.to.Total.assets 0.549 0.582942

Retained.earnings.to.Total.assets -0.691 0.489770

PBDITA.to.Total.assets 0.553 0.580192

Net.sales.to.Total.assets 0.592 0.553843

Equity.to.Total.assets 3.510 0.000448 \*\*\*

---

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

* We now create a logistic model using the significant variables at 95% confidence level and higher as identified in the above model

**FINAL MODEL ON FULL DATA WITH ONLY SIGNIFICANT VARIABLES**

Call:

glm(formula = Default ~ PAT.as...of.net.worth + Cumulative.retained.profits +

TOL.TNW + Debt.to.equity.ratio..times. + Cash.to.current.liabilities..times. +

Equity.to.Total.assets, family = binomial, data = data.wo.nas[,

-c(1, 2)])

Deviance Residuals:

Min 1Q Median 3Q Max

-2.0924 -0.2654 -0.1729 -0.0582 4.8837

Coefficients:

Estimate Std. Error z value

(Intercept) -3.602699 0.159057 -22.650

PAT.as...of.net.worth -0.061948 0.005559 -11.145

Cumulative.retained.profits -0.004271 0.001205 -3.545

TOL.TNW 0.181392 0.035302 5.138

Debt.to.equity.ratio..times. 0.135816 0.054412 2.496

Cash.to.current.liabilities..times. 0.482129 0.214729 2.245

Equity.to.Total.assets 0.043228 0.006504 6.647

Pr(>|z|)

(Intercept) < 2e-16 \*\*\*

PAT.as...of.net.worth < 2e-16 \*\*\*

Cumulative.retained.profits 0.000392 \*\*\*

TOL.TNW 2.77e-07 \*\*\*

Debt.to.equity.ratio..times. 0.012558 \*

Cash.to.current.liabilities..times. 0.024749 \*

Equity.to.Total.assets 2.99e-11 \*\*\*

---

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1771.0 on 3540 degrees of freedom

Residual deviance: 1026.5 on 3534 degrees of freedom

AIC: 1040.5

Number of Fisher Scoring iterations: 9

**VIF SUMMARY**

PAT.as...of.net.worth

1.139483

Cumulative.retained.profits

1.142531

TOL.TNW

3.209434

Debt.to.equity.ratio..times.

3.152927

Cash.to.current.liabilities..times.

1.072293

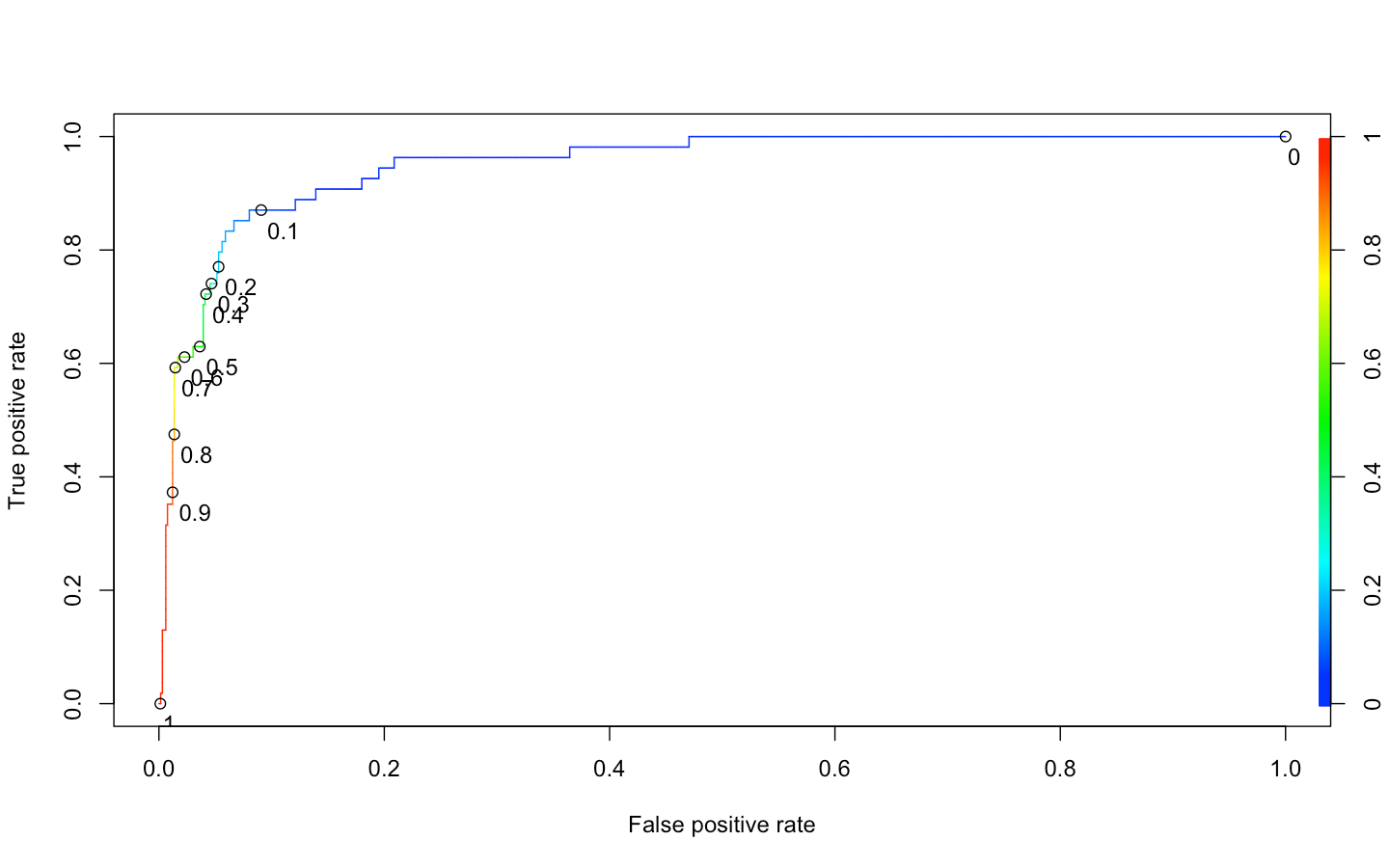
Equity.to.Total.assets

1.029765

* All our variables are significant in explaining the model
* The VIF is well below 4, hence the multicollinearity is addressed in the model

**ROC CURVE TO DETERMINE THE THRESHOLD AND AUC**

* The cutoff threshold is determined at 10% from the below ROC curve

****

**AUC**

**[1] 0.9508881**

**CONFUSION MATRIX AND PERFORMANCE EVALUATION ON TEST DATA FOR OUR FINAL FULL DATA MODEL**

**Confusion Matrix and Statistics**

Reference

Prediction 0 1

0 602 7

1 59 47

Accuracy : 0.9077

95% CI : (0.8841, 0.9279)

No Information Rate : 0.9245

P-Value [Acc > NIR] : 0.9583

Kappa : 0.5416

Mcnemar's Test P-Value : 3.437e-10

Sensitivity : 0.87037

Specificity : 0.91074

Pos Pred Value : 0.44340

Neg Pred Value : 0.98851

Prevalence : 0.07552

Detection Rate : 0.06573

Detection Prevalence : 0.14825

Balanced Accuracy : 0.89056

'Positive' Class : 1

* Sensitivity and AUC is highest for this model so far

|  |  |  |  |
| --- | --- | --- | --- |
| MEASURE | RATIO MODEL ON IMPORTANT RATIOS | FACTOR MODEL WITH 4 FACTORS | FULL DATA MODEL ON IMPORTANT VARIABLES |
| ACCURACY | 88.8 | 90.7 | 90.7 |
| SENSITIVITY | 85.1 | 85.1 | 87 |
| SPECIFICITY | 89.1 | 91.2 | 91 |
| AUC | 94.4 | 93.8 | 95 |

**SHORTLISTING THE BEST FIT MODEL – accuracy measures on Test data**

***data in %***

* All 3 models are very close to each other on performance.
* Full data model with important variables is slightly better performing then the other two models
* For the sake of simplicity, we would not like to go ahead with factor model as it becomes complicated to understand a model based on factors when it comes to explaining the coefficients of model equation
* Ratios do a better job in differentiating large companies, from mid and small companies , hence we can go ahead with any of the other two models as both of them have key ratios incorporated in their model equation

Thank you!

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_