**Company Data**

**Analysis**

Name: Saurabh Dharmadhikari

PGP-DSBA Online January’ 21

Table of Contents:

Problem Statement……………………...................................................................................................................................3

Data Description ............................................................................................................................................................4

Sample Data. …………………………………………………………………………………………………………………………………………….............5

Data Summary: ……………………….…………………………………………………………………………………………………………………………….7

Outliers ………………………..………………………………………………………………......…………………………………………………………........10

Missing value treatment. …………………………………….....................................................................................................13

Univariant and bivariant analysis. ..............................................................................................................................14

Scatter Plot:. …………………………………...........................................................................................................................17

Heat Map of Correlation. ............................................................................................................................................21

Train and Test Split …………………………………………………………………………........................................................................22

Linear Regression Model. …………….............................................................................................................................22

Model 1 ………………………………………………………………………………………………………………………………………………………………..24

Model 2 Using REF …………………………….………………………………………………………………………………………………………………….24

Model 3 Using SMOTE (75:25) ……………..……………………………………………………………………………………………………………….25

Model 4 SMOTE (65:35) ………………………………………………………………………………………………………………………………………..26

Model Validation …..………………………………………………………………………………………………………………………………………………26

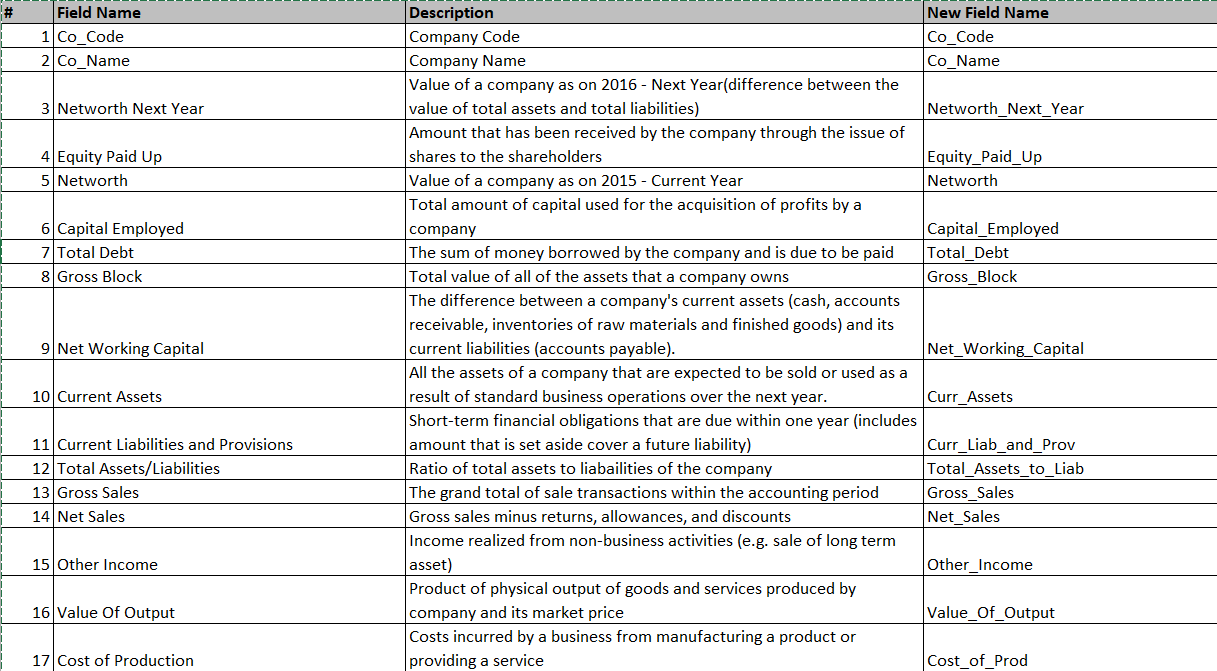
Problem Statement

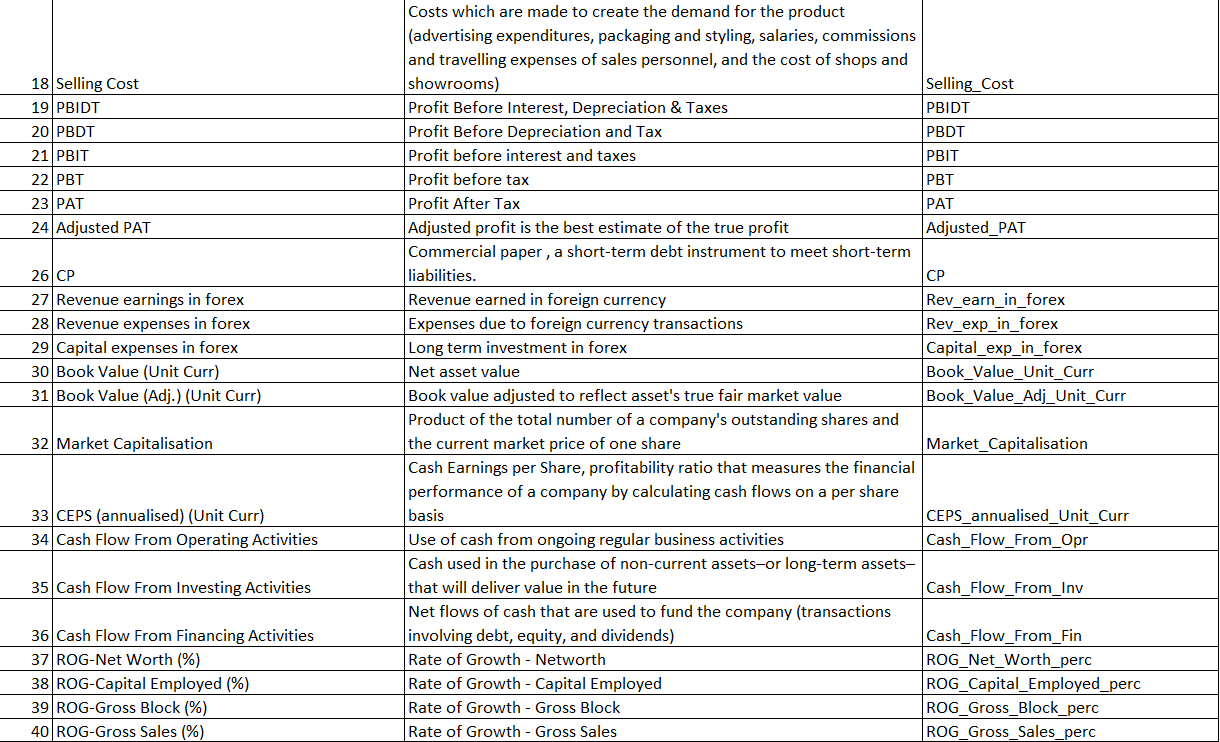
Businesses or companies can fall prey to default if they are not able to keep up their debt obligations. Defaults will lead to a lower credit rating for the company which in turn reduces its chances of getting credit in the future and may have to pay higher interests on existing debts as well as any new obligations. From an investor's point of view, he would want to invest in a company if it is capable of handling its financial obligations, can grow quickly, and is able to manage the growth scale.

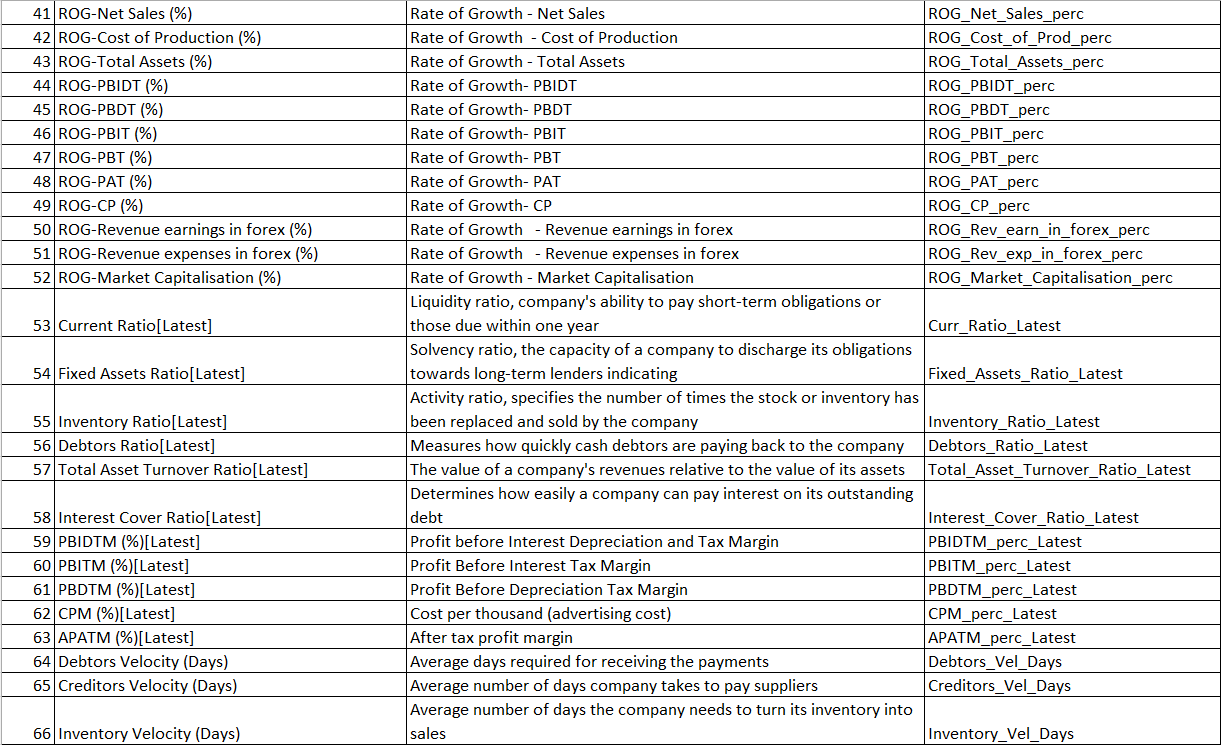
A balance sheet is a financial statement of a company that provides a snapshot of what a company owns, owes, and the amount invested by the shareholders. Thus, it is an important tool that helps evaluate the performance of a business.

Data that is available includes information from the financial statement of the companies for the previous year (2015). Also, information about the Net worth of the company in the following year (2016) is provided which can be used to drive the labelled field.

Data Description:

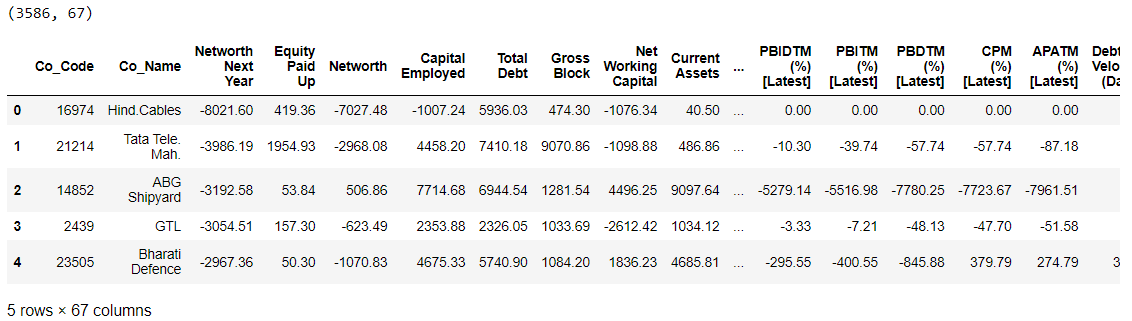








Sample Data



There is a total of 3586entries in our dataset and 67 columns.

There are no duplicates in this dataset.

A total of 118 null values are there.

RangeIndex: 3586 entries, 0 to 3585

Data columns (total 67 columns):

# Column Non-Null Count Dtype

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

0 CO\_CODE 3586 non-null int64

1 CO\_NAME 3586 non-null object

2 NETWORTH\_NEXT\_YEAR 3586 non-null float64

3 EQUITY\_PAID\_UP 3586 non-null float64

4 NETWORTH 3586 non-null float64

5 CAPITAL\_EMPLOYED 3586 non-null float64

6 TOTAL\_DEBT 3586 non-null float64

7 GROSS\_BLOCK 3586 non-null float64

8 NET\_WORKING\_CAPITAL 3586 non-null float64

9 CURRENT\_ASSETS 3586 non-null float64

10 CURRENT\_LIABILITIES\_AND\_PROVISIONS 3586 non-null float64

11 TOTAL\_ASSETS\_BY\_LIABILITIES 3586 non-null float64

12 GROSS\_SALES 3586 non-null float64

13 NET\_SALES 3586 non-null float64

14 OTHER\_INCOME 3586 non-null float64

15 VALUE\_OF\_OUTPUT 3586 non-null float64

16 COST\_OF\_PRODUCTION 3586 non-null float64

17 SELLING\_COST 3586 non-null float64

18 PBIDT 3586 non-null float64

19 PBDT 3586 non-null float64

20 PBIT 3586 non-null float64

21 PBT 3586 non-null float64

22 PAT 3586 non-null float64

23 ADJUSTED\_PAT 3586 non-null float64

24 CP 3586 non-null float64

25 REVENUE\_EARNINGS\_IN\_FOREX 3586 non-null float64

26 REVENUE\_EXPENSES\_IN\_FOREX 3586 non-null float64

27 CAPITAL\_EXPENSES\_IN\_FOREX 3586 non-null float64

28 BOOK\_VALUE\_UNIT\_CURR 3586 non-null float64

29 BOOK\_VALUE\_ADJ\_UNIT\_CURR 3582 non-null float64

30 MARKET\_CAPITALISATION 3586 non-null float64

31 CEPS\_ANNUALISED\_UNIT\_CURR 3586 non-null float64

32 CASH\_FLOW\_FROM\_OPERATING\_ACTIVITIES 3586 non-null float64

33 CASH\_FLOW\_FROM\_INVESTING\_ACTIVITIES 3586 non-null float64

34 CASH\_FLOW\_FROM\_FINANCING\_ACTIVITIES 3586 non-null float64

35 ROG\_NET\_WORTH\_PERC 3586 non-null float64

36 ROG\_CAPITAL\_EMPLOYED\_PERC 3586 non-null float64

37 ROG\_GROSS\_BLOCK\_PERC 3586 non-null float64

38 ROG\_GROSS\_SALES\_PERC 3586 non-null float64

39 ROG\_NET\_SALES\_PERC 3586 non-null float64

40 ROG\_COST\_OF\_PRODUCTION\_PERC 3586 non-null float64

41 ROG\_TOTAL\_ASSETS\_PERC 3586 non-null float64

42 ROG\_PBIDT\_PERC 3586 non-null float64

43 ROG\_PBDT\_PERC 3586 non-null float64

44 ROG\_PBIT\_PERC 3586 non-null float64

45 ROG\_PBT\_PERC 3586 non-null float64

46 ROG\_PAT\_PERC 3586 non-null float64

47 ROG\_CP\_PERC 3586 non-null float64

48 ROG\_REVENUE\_EARNINGS\_IN\_FOREX\_PERC 3586 non-null float64

49 ROG\_REVENUE\_EXPENSES\_IN\_FOREX\_PERC 3586 non-null float64

50 ROG\_MARKET\_CAPITALISATION\_PERC 3586 non-null float64

51 CURRENT\_RATIO\_LATEST 3585 non-null float64

52 FIXED\_ASSETS\_RATIO\_LATEST 3585 non-null float64

53 INVENTORY\_RATIO\_LATEST 3585 non-null float64

54 DEBTORS\_RATIO\_LATEST 3585 non-null float64

55 TOTAL\_ASSET\_TURNOVER\_RATIO\_LATEST 3585 non-null float64

56 INTEREST\_COVER\_RATIO\_LATEST 3585 non-null float64

57 PBIDTM\_PERC\_LATEST 3585 non-null float64

58 PBITM\_PERC\_LATEST 3585 non-null float64

59 PBDTM\_PERC\_LATEST 3585 non-null float64

60 CPM\_PERC\_LATEST 3585 non-null float64

61 APATM\_PERC\_LATEST 3585 non-null float64

62 DEBTORS\_VELOCITY\_DAYS 3586 non-null int64

63 CREDITORS\_VELOCITY\_DAYS 3586 non-null int64

64 INVENTORY\_VELOCITY\_DAYS 3483 non-null float64

65 VALUE\_OF\_OUTPUT\_BY\_TOTAL\_ASSETS 3586 non-null float64

66 VALUE\_OF\_OUTPUT\_BY\_GROSS\_BLOCK 3586 non-null float64

dtypes: float64(63), int64(3), object(1)

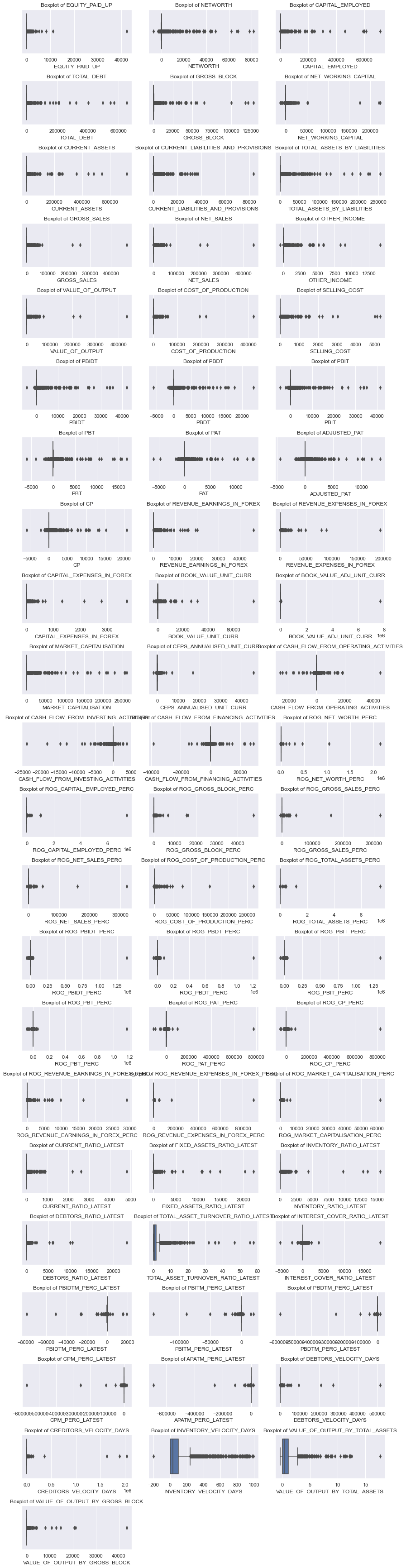
There is only 1 categorical column all other columns are numeric in nature.

**Data Summary:**

|  | **count** | **mean** | **std** | **min** | **25%** | **50%** | **75%** | **max** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CO\_CODE** | 3586.00 | 16065.39 | 19776.82 | 4.00 | 3029.25 | 6077.50 | 24269.50 | 72493.00 |
| **NETWORTH\_NEXT\_YEAR** | 3586.00 | 725.05 | 4769.68 | -8021.60 | 3.98 | 19.02 | 123.80 | 111729.10 |
| **EQUITY\_PAID\_UP** | 3586.00 | 62.97 | 778.76 | 0.00 | 3.75 | 8.29 | 19.52 | 42263.46 |
| **NETWORTH** | 3586.00 | 649.75 | 4091.99 | -7027.48 | 3.89 | 18.58 | 117.30 | 81657.35 |
| **CAPITAL\_EMPLOYED** | 3586.00 | 2799.61 | 26975.14 | -1824.75 | 7.60 | 39.09 | 226.60 | 714001.25 |
| **TOTAL\_DEBT** | 3586.00 | 1994.82 | 23652.84 | -0.72 | 0.03 | 7.49 | 72.35 | 652823.81 |
| **GROSS\_BLOCK** | 3586.00 | 594.18 | 4871.55 | -41.19 | 0.57 | 15.87 | 131.90 | 128477.59 |
| **NET\_WORKING\_CAPITAL** | 3586.00 | 410.81 | 6301.22 | -13162.42 | 0.94 | 10.14 | 61.17 | 223257.56 |
| **CURRENT\_ASSETS** | 3586.00 | 1960.35 | 22577.57 | -0.91 | 4.00 | 24.54 | 135.28 | 721166.00 |
| **CURRENT\_LIABILITIES\_AND\_PROVISIONS** | 3586.00 | 391.99 | 2675.00 | -0.23 | 0.73 | 9.23 | 65.65 | 83232.98 |
| **TOTAL\_ASSETS\_BY\_LIABILITIES** | 3586.00 | 1778.45 | 11437.57 | -4.51 | 10.55 | 52.01 | 310.54 | 254737.22 |
| **GROSS\_SALES** | 3586.00 | 1123.74 | 10603.70 | -62.59 | 1.44 | 31.21 | 242.25 | 474182.94 |
| **NET\_SALES** | 3586.00 | 1079.70 | 9996.57 | -62.59 | 1.44 | 30.44 | 234.44 | 443775.16 |
| **OTHER\_INCOME** | 3586.00 | 48.73 | 426.04 | -448.72 | 0.02 | 0.45 | 3.64 | 14143.40 |
| **VALUE\_OF\_OUTPUT** | 3586.00 | 1077.19 | 9843.88 | -119.10 | 1.41 | 30.89 | 235.84 | 435559.09 |
| **COST\_OF\_PRODUCTION** | 3586.00 | 798.54 | 9076.70 | -22.65 | 0.94 | 25.99 | 189.55 | 419913.50 |
| **SELLING\_COST** | 3586.00 | 25.55 | 194.24 | 0.00 | 0.00 | 0.16 | 3.88 | 5283.91 |
| **PBIDT** | 3586.00 | 248.18 | 1949.59 | -4655.14 | 0.04 | 2.04 | 23.52 | 42059.26 |
| **PBDT** | 3586.00 | 116.27 | 956.20 | -5874.53 | 0.00 | 0.80 | 12.95 | 23215.00 |
| **PBIT** | 3586.00 | 217.66 | 1850.97 | -4812.95 | 0.00 | 1.15 | 16.67 | 41402.96 |
| **PBT** | 3586.00 | 85.75 | 799.93 | -6032.34 | -0.06 | 0.31 | 7.42 | 16798.00 |
| **PAT** | 3586.00 | 61.22 | 620.30 | -6032.34 | -0.06 | 0.26 | 5.54 | 13383.39 |
| **ADJUSTED\_PAT** | 3586.00 | 60.06 | 580.43 | -4418.72 | -0.09 | 0.21 | 5.34 | 13384.11 |
| **CP** | 3586.00 | 91.73 | 780.79 | -5874.53 | 0.00 | 0.74 | 10.91 | 20760.20 |
| **REVENUE\_EARNINGS\_IN\_FOREX** | 3586.00 | 131.17 | 1150.73 | 0.00 | 0.00 | 0.00 | 7.20 | 46158.00 |
| **REVENUE\_EXPENSES\_IN\_FOREX** | 3586.00 | 256.33 | 4132.34 | 0.00 | 0.00 | 0.00 | 6.99 | 193979.73 |
| **CAPITAL\_EXPENSES\_IN\_FOREX** | 3586.00 | 7.66 | 111.43 | 0.00 | 0.00 | 0.00 | 0.00 | 3722.10 |
| **BOOK\_VALUE\_UNIT\_CURR** | 3586.00 | 157.24 | 1622.66 | -3371.57 | 7.96 | 21.66 | 71.67 | 75790.00 |
| **BOOK\_VALUE\_ADJ\_UNIT\_CURR** | 3582.00 | 2243.15 | 128283.73 | -33715.70 | 7.06 | 18.93 | 60.01 | 7677600.29 |
| **MARKET\_CAPITALISATION** | 3586.00 | 1664.09 | 12805.17 | 0.00 | 0.00 | 8.37 | 111.46 | 260865.08 |
| **CEPS\_ANNUALISED\_UNIT\_CURR** | 3586.00 | 36.02 | 828.42 | -1808.00 | 0.00 | 1.15 | 8.77 | 45438.44 |
| **CASH\_FLOW\_FROM\_OPERATING\_ACTIVITIES** | 3586.00 | 65.77 | 1455.05 | -25469.23 | -0.31 | 0.45 | 12.65 | 44529.40 |
| **CASH\_FLOW\_FROM\_INVESTING\_ACTIVITIES** | 3586.00 | -60.87 | 701.97 | -23843.45 | -5.12 | -0.12 | 0.12 | 3732.98 |
| **CASH\_FLOW\_FROM\_FINANCING\_ACTIVITIES** | 3586.00 | 11.44 | 1272.26 | -38374.04 | -5.85 | 0.00 | 0.46 | 28846.00 |
| **ROG\_NET\_WORTH\_PERC** | 3586.00 | 1237.62 | 41041.93 | -14485.71 | -1.49 | 1.84 | 11.36 | 2144020.00 |
| **ROG\_CAPITAL\_EMPLOYED\_PERC** | 3586.00 | 2988.88 | 126472.87 | -8614.63 | -3.83 | 1.38 | 12.59 | 7412700.00 |
| **ROG\_GROSS\_BLOCK\_PERC** | 3586.00 | 37.55 | 893.62 | -116.12 | 0.00 | 0.25 | 6.72 | 47400.00 |
| **ROG\_GROSS\_SALES\_PERC** | 3586.00 | 242.67 | 6103.53 | -5503.70 | -8.08 | 3.31 | 21.53 | 320200.00 |
| **ROG\_NET\_SALES\_PERC** | 3586.00 | 242.59 | 6103.49 | -5503.70 | -8.12 | 3.21 | 21.57 | 320200.00 |
| **ROG\_COST\_OF\_PRODUCTION\_PERC** | 3586.00 | 310.49 | 5573.22 | -2130.23 | -7.24 | 4.42 | 23.12 | 267150.00 |
| **ROG\_TOTAL\_ASSETS\_PERC** | 3586.00 | 2793.28 | 125941.65 | -136.13 | -3.97 | 1.48 | 12.50 | 7422120.00 |
| **ROG\_PBIDT\_PERC** | 3586.00 | 375.85 | 23278.40 | -52200.00 | -23.36 | 4.57 | 47.88 | 1386200.00 |
| **ROG\_PBDT\_PERC** | 3586.00 | 336.38 | 20353.40 | -52200.00 | -30.60 | 3.37 | 52.91 | 1208700.00 |
| **ROG\_PBIT\_PERC** | 3586.00 | 374.70 | 22462.79 | -58500.00 | -31.35 | 2.13 | 50.14 | 1338000.00 |
| **ROG\_PBT\_PERC** | 3586.00 | 224.07 | 19659.23 | -78900.00 | -41.23 | 0.03 | 61.96 | 1160500.00 |
| **ROG\_PAT\_PERC** | 3586.00 | 112.23 | 13480.52 | -114500.00 | -43.73 | 0.00 | 65.35 | 774200.00 |
| **ROG\_CP\_PERC** | 3586.00 | 221.09 | 13980.20 | -52200.00 | -29.51 | 4.62 | 52.91 | 822400.00 |
| **ROG\_REVENUE\_EARNINGS\_IN\_FOREX\_PERC** | 3586.00 | 37.23 | 658.67 | -100.00 | 0.00 | 0.00 | 0.00 | 29084.77 |
| **ROG\_REVENUE\_EXPENSES\_IN\_FOREX\_PERC** | 3586.00 | 364.86 | 15233.64 | -100.00 | 0.00 | 0.00 | 0.00 | 894591.69 |
| **ROG\_MARKET\_CAPITALISATION\_PERC** | 3586.00 | 63.68 | 1047.93 | -98.05 | 0.00 | 0.00 | 47.52 | 61865.26 |
| **CURRENT\_RATIO\_LATEST** | 3585.00 | 12.06 | 108.41 | 0.00 | 0.88 | 1.36 | 2.77 | 4813.00 |
| **FIXED\_ASSETS\_RATIO\_LATEST** | 3585.00 | 51.54 | 681.15 | 0.00 | 0.27 | 1.56 | 4.74 | 22172.00 |
| **INVENTORY\_RATIO\_LATEST** | 3585.00 | 37.80 | 458.19 | 0.00 | 0.00 | 3.56 | 8.94 | 15472.00 |
| **DEBTORS\_RATIO\_LATEST** | 3585.00 | 33.03 | 489.56 | 0.00 | 0.42 | 3.82 | 8.52 | 22992.67 |
| **TOTAL\_ASSET\_TURNOVER\_RATIO\_LATEST** | 3585.00 | 1.24 | 2.67 | 0.00 | 0.07 | 0.60 | 1.55 | 57.75 |
| **INTEREST\_COVER\_RATIO\_LATEST** | 3585.00 | 16.39 | 351.74 | -5450.00 | 0.00 | 1.08 | 3.71 | 18639.40 |
| **PBIDTM\_PERC\_LATEST** | 3585.00 | -51.16 | 1795.13 | -78870.45 | 0.00 | 8.07 | 18.99 | 19233.33 |
| **PBITM\_PERC\_LATEST** | 3585.00 | -109.21 | 3057.64 | -141600.00 | 0.00 | 5.23 | 14.29 | 19195.70 |
| **PBDTM\_PERC\_LATEST** | 3585.00 | -311.57 | 10921.59 | -590500.00 | 0.00 | 4.69 | 14.11 | 15640.00 |
| **CPM\_PERC\_LATEST** | 3585.00 | -307.01 | 10676.15 | -572000.00 | 0.00 | 3.89 | 11.39 | 15640.00 |
| **APATM\_PERC\_LATEST** | 3585.00 | -365.06 | 12500.05 | -688600.00 | 0.00 | 1.59 | 7.41 | 15266.67 |
| **DEBTORS\_VELOCITY\_DAYS** | 3586.00 | 603.89 | 10636.76 | 0.00 | 8.00 | 49.00 | 106.00 | 514721.00 |
| **CREDITORS\_VELOCITY\_DAYS** | 3586.00 | 2057.85 | 54169.48 | 0.00 | 8.00 | 39.00 | 89.00 | 2034145.00 |
| **INVENTORY\_VELOCITY\_DAYS** | 3483.00 | 79.64 | 137.85 | -199.00 | 0.00 | 35.00 | 96.00 | 996.00 |
| **VALUE\_OF\_OUTPUT\_BY\_TOTAL\_ASSETS** | 3586.00 | 0.82 | 1.20 | -0.33 | 0.07 | 0.48 | 1.16 | 17.63 |
| **VALUE\_OF\_OUTPUT\_BY\_GROSS\_BLOCK** | 3586.00 | 61.88 | 976.82 | -61.00 | 0.27 | 1.53 | 4.91 | 43404.00 |

Let us see the outliers in our dataset.

**Outliers:**



There are outliers in all our features.

We will be considering an outlier as 3rd quartile + 1.5 \*IQR and 1st quartile – 1.5\*IQR.

Inter Quartile = Q3 – Q1

Quantum of outliers for features:

ROG\_REVENUE\_EXPENSES\_IN\_FOREX\_PERC 45.04

ROG\_REVENUE\_EARNINGS\_IN\_FOREX\_PERC 36.73

CASH\_FLOW\_FROM\_FINANCING\_ACTIVITIES 28.03

PAT 26.74

ADJUSTED\_PAT 26.60

PBT 26.24

APATM\_PERC\_LATEST 26.05

CASH\_FLOW\_FROM\_INVESTING\_ACTIVITIES 24.43

ROG\_GROSS\_BLOCK\_PERC 23.15

CP 22.76

PBDT 22.73

CASH\_FLOW\_FROM\_OPERATING\_ACTIVITIES 22.34

ROG\_NET\_WORTH\_PERC 20.83

REVENUE\_EARNINGS\_IN\_FOREX 20.58

INTEREST\_COVER\_RATIO\_LATEST 20.25

CPM\_PERC\_LATEST 20.11

PBIT 20.08

PBITM\_PERC\_LATEST 20.02

PBDTM\_PERC\_LATEST 19.41

CAPITAL\_EXPENSES\_IN\_FOREX 19.35

REVENUE\_EXPENSES\_IN\_FOREX 19.33

ROG\_COST\_OF\_PRODUCTION\_PERC 18.82

ROG\_GROSS\_SALES\_PERC 18.71

PBIDT 18.71

ROG\_NET\_SALES\_PERC 18.60

NETWORTH 18.13

MARKET\_CAPITALISATION 17.82

ROG\_CP\_PERC 17.76

ROG\_PBDT\_PERC 17.51

NET\_WORKING\_CAPITAL 17.43

ROG\_PBIT\_PERC 17.18

ROG\_PBIDT\_PERC 17.04

ROG\_PBT\_PERC 17.04

SELLING\_COST 16.87

OTHER\_INCOME 16.82

CEPS\_ANNUALISED\_UNIT\_CURR 16.79

ROG\_PAT\_PERC 16.68

CAPITAL\_EMPLOYED 16.62

PBIDTM\_PERC\_LATEST 16.62

TOTAL\_DEBT 16.26

CURRENT\_LIABILITIES\_AND\_PROVISIONS 16.20

CURRENT\_ASSETS 16.09

TOTAL\_ASSETS\_BY\_LIABILITIES 16.01

ROG\_CAPITAL\_EMPLOYED\_PERC 15.95

CURRENT\_RATIO\_LATEST 15.78

COST\_OF\_PRODUCTION 15.62

VALUE\_OF\_OUTPUT 15.59

NET\_SALES 15.50

GROSS\_SALES 15.45

GROSS\_BLOCK 15.06

ROG\_MARKET\_CAPITALISATION\_PERC 13.86

FIXED\_ASSETS\_RATIO\_LATEST 13.83

BOOK\_VALUE\_ADJ\_UNIT\_CURR 13.66

BOOK\_VALUE\_UNIT\_CURR 13.52

ROG\_TOTAL\_ASSETS\_PERC 13.47

VALUE\_OF\_OUTPUT\_BY\_GROSS\_BLOCK 13.41

EQUITY\_PAID\_UP 12.49

DEBTORS\_VELOCITY\_DAYS 11.10

CREDITORS\_VELOCITY\_DAYS 10.90

INVENTORY\_RATIO\_LATEST 10.49

DEBTORS\_RATIO\_LATEST 10.37

INVENTORY\_VELOCITY\_DAYS 10.18

TOTAL\_ASSET\_TURNOVER\_RATIO\_LATEST 5.63

VALUE\_OF\_OUTPUT\_BY\_TOTAL\_ASSETS 4.18

dtype: float64

Thus, above are all outliers in our columns. All outliers are replaced with null value.

**Missing value treatment:**

There is a lot of missing values in our dataset so we will have treated the data and replace the missing values.

We will also convert our target variable to binary variable (1 & 0). 0 is Net\_worth\_next\_year >= 0 and 1 is Net\_worth\_next\_year < 0.

We are removing 2 columns with max missing values.

ROG\_REVENUE\_EXPENSES\_IN\_FOREX\_PERC 45.04

ROG\_REVENUE\_EARNINGS\_IN\_FOREX\_PERC 36.73

Missing values have been treated using KNN.

Target variable after conversion:

Proportion of Default in the dataset

0 3043

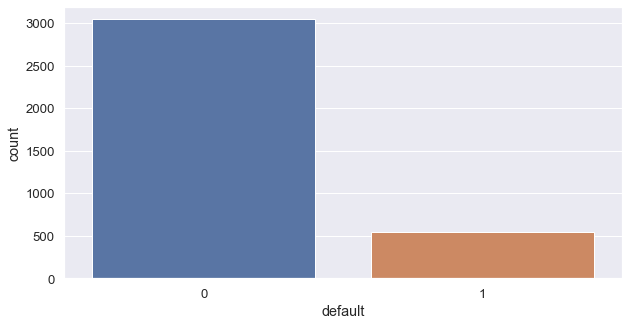
1 543

Name: default, dtype: int64

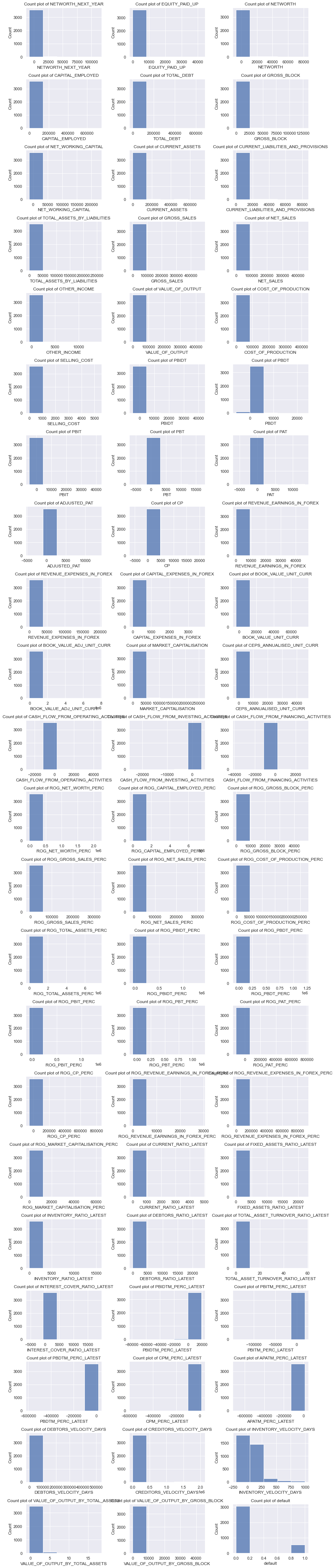
0 84.86

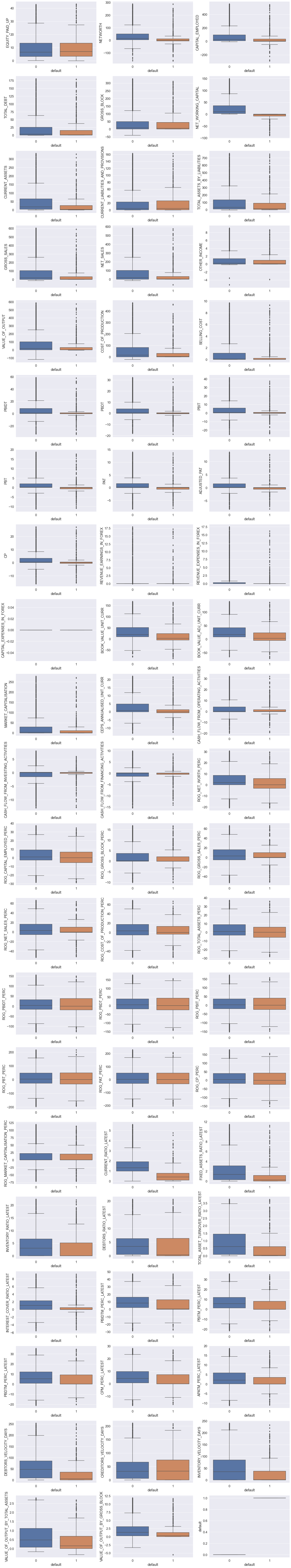
1 15.14

Name: default, dtype: float64



**Univariant and bivariant analysis:**





We can make out with univariant chart that most data is very skewed.

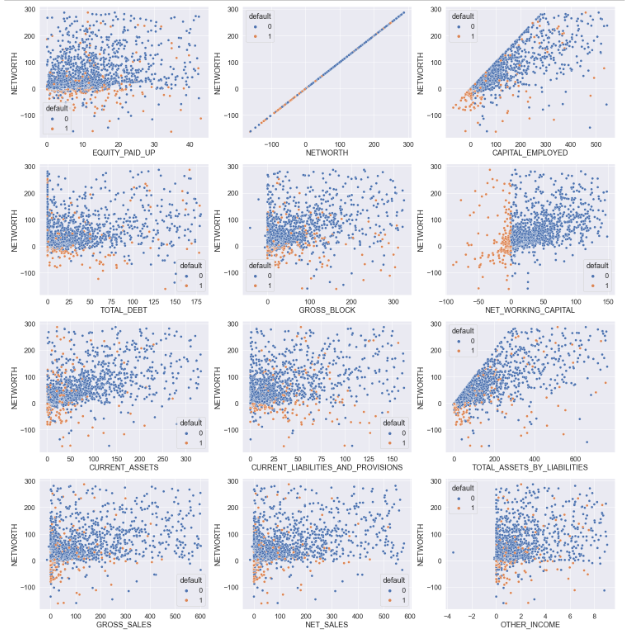
Most companies are able to convert their inventory to sales in short period.

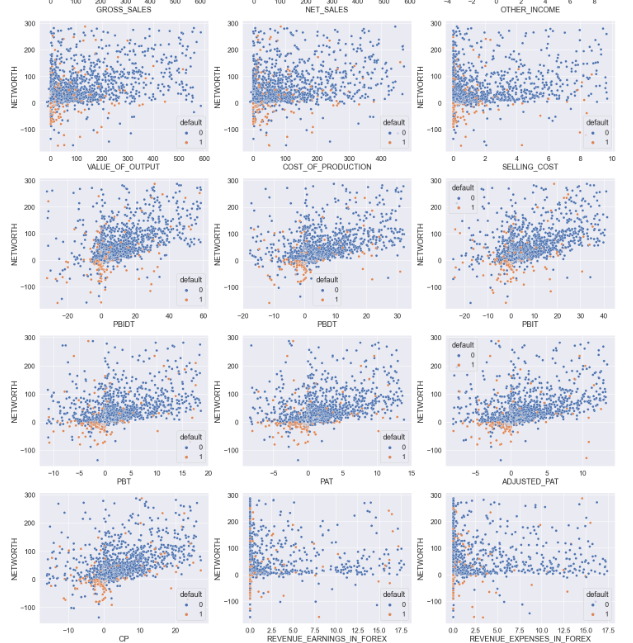
With the bivariant analysis charts we can easily see that the difference in companies that are doing well and those that are not doing well and how features can indicate the defaulting companies.

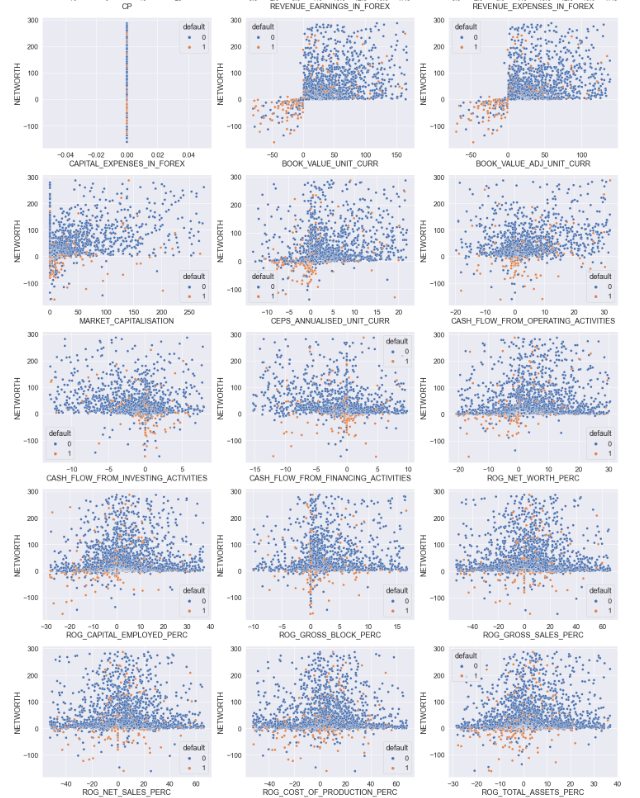
Companies that may default have low values in:

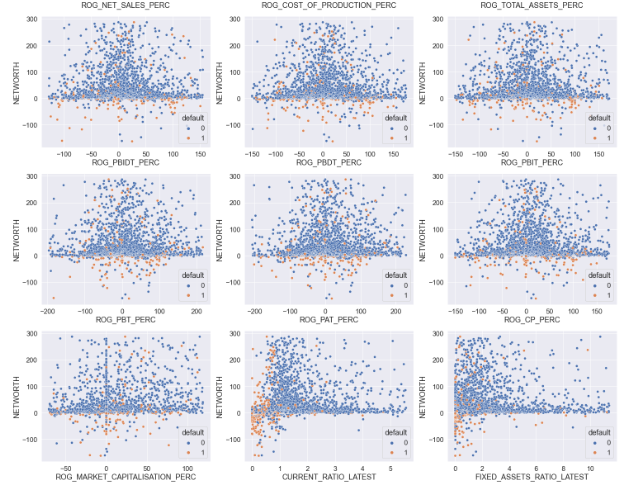
* Net Worth
* Capital Employed
* Net Working Capital
* Current Assets
* Total Assets by Liabilities
* Gross Sales
* Net Sales
* Selling Cost
* Value of Output
* Cost of Production
* PBIDT
* PBT
* Adjusted PAT
* CP
* Book Value of Adjusted Unit Currency

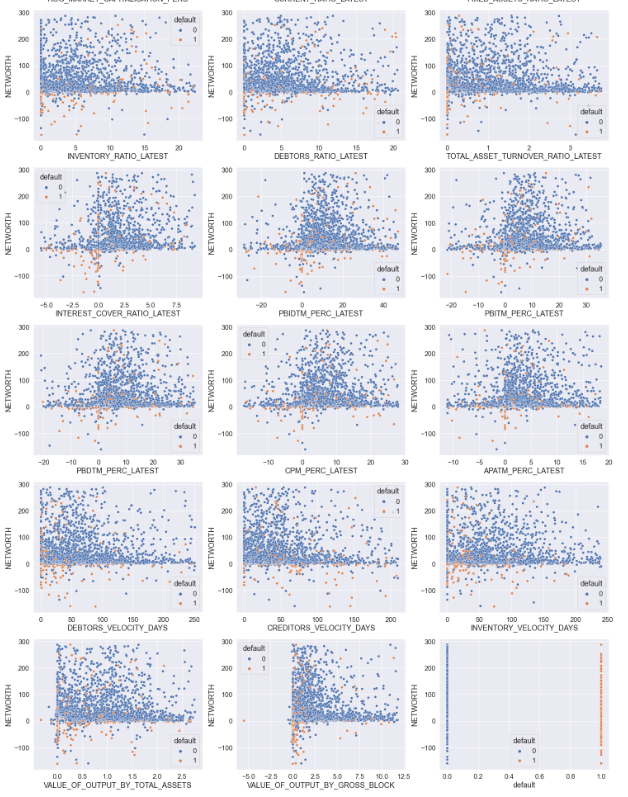
**Scatter Plot:**







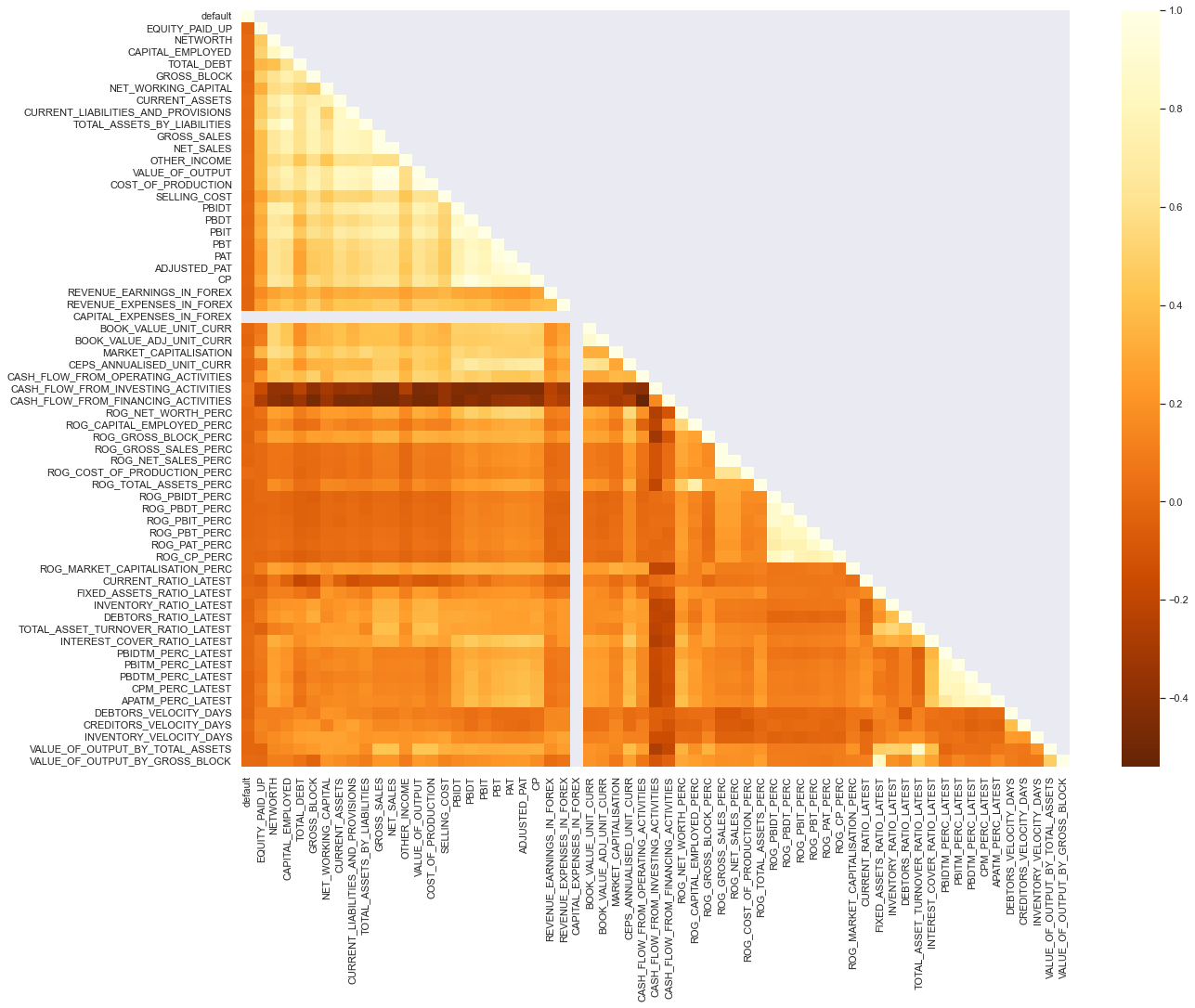




Here orange dot indicates defaulting companies in comparison to their current net worth.

We can see that orange dots for defaulting companies are at the bottom of the charts.

**Heat Map of correlation:**



There is a lot of positive correlation between the features in the data.

**Train and Test:**

We have split the data in 67:33 ratio.

**Logistic Regression Model:**

We will use VIF method to know the features that have high multicollinearity.

|  | **VIF** |
| --- | --- |
| **NET\_SALES** | 217.62 |
| **VALUE\_OF\_OUTPUT** | 143.10 |
| **GROSS\_SALES** | 85.34 |
| **ROG\_GROSS\_SALES\_PERC** | 79.00 |
| **ROG\_NET\_SALES\_PERC** | 78.75 |
| **PAT** | 20.28 |
| **PBDT** | 19.71 |
| **CP** | 17.28 |
| **COST\_OF\_PRODUCTION** | 16.67 |
| **TOTAL\_ASSETS\_BY\_LIABILITIES** | 13.65 |
| **PBT** | 11.65 |
| **CAPITAL\_EMPLOYED** | 11.00 |
| **PBDTM\_PERC\_LATEST** | 10.95 |
| **ADJUSTED\_PAT** | 10.83 |
| **ROG\_PBDT\_PERC** | 10.69 |
| **CURRENT\_ASSETS** | 10.36 |
| **PBIT** | 10.09 |
| **CPM\_PERC\_LATEST** | 9.00 |
| **PBIDT** | 8.34 |
| **ROG\_CP\_PERC** | 8.05 |
| **ROG\_PBIDT\_PERC** | 7.38 |
| **ROG\_PBT\_PERC** | 6.81 |
| **CURRENT\_LIABILITIES\_AND\_PROVISIONS** | 6.78 |
| **ROG\_PBIT\_PERC** | 6.65 |
| **PBITM\_PERC\_LATEST** | 6.31 |
| **VALUE\_OF\_OUTPUT\_BY\_TOTAL\_ASSETS** | 6.12 |
| **TOTAL\_ASSET\_TURNOVER\_RATIO\_LATEST** | 5.76 |
| **PBIDTM\_PERC\_LATEST** | 5.67 |
| **ROG\_PAT\_PERC** | 5.65 |
| **BOOK\_VALUE\_UNIT\_CURR** | 5.60 |
| **BOOK\_VALUE\_ADJ\_UNIT\_CURR** | 5.19 |
| **NETWORTH** | 4.92 |
| **GROSS\_BLOCK** | 4.51 |
| **FIXED\_ASSETS\_RATIO\_LATEST** | 4.44 |
| **VALUE\_OF\_OUTPUT\_BY\_GROSS\_BLOCK** | 4.43 |
| **APATM\_PERC\_LATEST** | 4.12 |
| **NET\_WORKING\_CAPITAL** | 3.38 |
| **CEPS\_ANNUALISED\_UNIT\_CURR** | 3.34 |
| **ROG\_CAPITAL\_EMPLOYED\_PERC** | 2.63 |
| **TOTAL\_DEBT** | 2.46 |
| **ROG\_TOTAL\_ASSETS\_PERC** | 2.37 |
| **ROG\_NET\_WORTH\_PERC** | 2.24 |
| **OTHER\_INCOME** | 2.01 |
| **SELLING\_COST** | 1.99 |
| **CASH\_FLOW\_FROM\_OPERATING\_ACTIVITIES** | 1.99 |
| **CASH\_FLOW\_FROM\_FINANCING\_ACTIVITIES** | 1.90 |
| **MARKET\_CAPITALISATION** | 1.89 |
| **INTEREST\_COVER\_RATIO\_LATEST** | 1.80 |
| **ROG\_COST\_OF\_PRODUCTION\_PERC** | 1.71 |
| **EQUITY\_PAID\_UP** | 1.64 |
| **REVENUE\_EXPENSES\_IN\_FOREX** | 1.64 |
| **INVENTORY\_RATIO\_LATEST** | 1.62 |
| **CASH\_FLOW\_FROM\_INVESTING\_ACTIVITIES** | 1.61 |
| **DEBTORS\_RATIO\_LATEST** | 1.58 |
| **ROG\_GROSS\_BLOCK\_PERC** | 1.41 |
| **CREDITORS\_VELOCITY\_DAYS** | 1.40 |
| **DEBTORS\_VELOCITY\_DAYS** | 1.38 |
| **REVENUE\_EARNINGS\_IN\_FOREX** | 1.36 |
| **INVENTORY\_VELOCITY\_DAYS** | 1.28 |
| **ROG\_MARKET\_CAPITALISATION\_PERC** | 1.27 |
| **CURRENT\_RATIO\_LATEST** | 1.25 |
| **CAPITAL\_EXPENSES\_IN\_FOREX** | NaN |

We will eliminate features with VIF factor more than 5 from the mode.

Model 1:

Test Data - Confusion Matrix

[[1005 0]

[ 179 0]]

Train Data - Confusion Matrix

[[2038 0]

[ 364 0]]

This model is not good as it is predicting all of them as not default.

Model 2 using RFE:

|  | **Feature** | **Rank** |
| --- | --- | --- |
| **0** | EQUITY\_PAID\_UP | 1 |
| **1** | NETWORTH | 1 |
| **2** | TOTAL\_DEBT | 1 |
| **3** | GROSS\_BLOCK | 1 |
| **5** | CURRENT\_LIABILITIES\_AND\_PROVISIONS | 1 |
| **6** | OTHER\_INCOME | 1 |
| **13** | BOOK\_VALUE\_ADJ\_UNIT\_CURR | 1 |
| **17** | CASH\_FLOW\_FROM\_INVESTING\_ACTIVITIES | 1 |
| **21** | ROG\_GROSS\_BLOCK\_PERC | 1 |
| **22** | ROG\_NET\_SALES\_PERC | 1 |
| **23** | ROG\_COST\_OF\_PRODUCTION\_PERC | 1 |
| **25** | ROG\_PBIDT\_PERC | 1 |
| **27** | ROG\_CP\_PERC | 1 |
| **38** | DEBTORS\_VELOCITY\_DAYS | 1 |
| **39** | CREDITORS\_VELOCITY\_DAYS | 1 |

Model has selected above feature.

Test Data - Confusion Matrix RFE

[[1005 0]

[ 179 0]]

Train Data - Confusion Matrix RFE

[[2038 0]

[ 364 0]]

Again, the results are that no company will default.

Model 3 Using SMOTE (75:25):

Test Data - Classification Report SMOTE

precision recall f1-score support

0 0.85 0.77 0.81 1005

1 0.14 0.21 0.17 179

accuracy 0.69 1184

macro avg 0.49 0.49 0.49 1184

weighted avg 0.74 0.69 0.71 1184

Train Data - Classification Report SMOTE

precision recall f1-score support

0 0.62 0.78 0.69 2038

1 0.55 0.36 0.43 1528

accuracy 0.60 3566

macro avg 0.58 0.57 0.56 3566

weighted avg 0.59 0.60 0.58 3566

We see that the recall for default in the test data only 21 %. That is out of the total default cases , our model will be able to identify only 21 % of defaults .

Precision for default is 14 % , that is out of all the predictions that we make for default – only 14 % are likely to be correct.

The same figures for default in Train data Recall at 36 % and Precision at 56 %.

There is a big difference between the train and test data figures . This is a case of under fitment.

Model 4 SMOTE (65:35):

Test Data - Confusion Matrix SMOTE

[[830 175]

[152 27]]

Train Data - Confusion Matrix SMOTE

[[1702 336]

[1056 370]]

The results are not better than the earlier model.

**Model Validation:**

The model performance is not satisfactory.

More data needs to be collected which is clean and without as many outliers.

Meaningful models can be built only then