FRA Milestone 1 Business Project Report

TEJAS PADEKAR
PGP-DSBA Online
Date: 04/12/2022

CONTENTS:

List of Figures

List of Tables

PROBLEM 1

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.

Problem Statement 1:

Accurately predict the defaulters using Statsmodel approach in Logistic Regression. Data that is available includes information from the financial statement of the companies for the previous year (2015). Also, information about the Networth of the company in the following year (2016) is provided which is to be used to drive the labeled field.

We need to create a default variable that should take the value of 1 when net worth next year is negative & 0 when net worth next year is positive. We have created the necessary label field 'default' using the projected company networth next year for 1 year in future ie 2016 as requested.

Data Dictionary:

	#	Field Name	Description	New Field Name
0	1	Co_Code	Company Code	Co_Code
1	2	Co_Name	Company Name	Co_Name
2	3	Networth Next Year	Value of a company as on 2016 - Next Year(difference between the value of total assets and total liabilities)	Networth_Next_Year
3	4	Equity Paid Up	Amount that has been received by the company through the issue of shares to the shareholders	Equity_Paid_Up
4	5	Networth	Value of a company as on 2015 - Current Year	Networth
5	6	Capital Employed	Total amount of capital used for the acquisition of profits by a company	Capital_Employed
6	7	Total Debt	The sum of money borrowed by the company and is due to be paid	Total_Debt
7	8	Gross Block	Total value of all of the assets that a company owns	Gross_Block
8	9	Net Working Capital	The difference between a company's current assets (cash, accounts receivable, inventories of raw materials and finished goods) and its current liabilities (accounts payable).	Net_Working_Capital
9	10	Current Assets	All the assets of a company that are expected to be sold or used as a result of standard business operations over the next year.	Curr_Assets
10	11	Current Liabilities and Provisions	Short-term financial obligations that are due within one year (includes amount that is set aside cover a future liability)	Curr_Liab_and_Prov
11	12	Total Assets/Liabilities	Ratio of total assets to liabailities of the company	Total_Assets_to_Liab
12	13	Gross Sales	The grand total of sale transactions within the accounting period	Gross_Sales
13	14	Net Sales	Gross sales minus returns, allowances, and discounts	Net_Sales
14	15	Other Income	Income realized from non-business activities (e.g. sale of long term asset)	Other_Income
15	16	Value Of Output	Product of physical output of goods and services produced by company and its market price	Value_Of_Output
16	17	Cost of Production	Costs incurred by a business from manufacturing a product or providing a service	Cost_of_Prod
17	18	Selling Cost	Costs which are made to create the demand for the product (advertising expenditures, packaging and styling, salaries, commissions and travelling expenses of sales personnel, and the cost of shops and showrooms)	Selling_Cost
18	19	PBIDT	Profit Before Interest, Depreciation & Taxes	PBIDT
19	20	PBDT	Profit Before Depreciation and Tax	PBDT
20	21	PBIT	Profit before interest and taxes	PBIT
21	22	PBT	Profit before tax	PBT
22	23	PAT	Profit After Tax	PAT
23	24	Adjusted PAT	Adjusted profit is the best estimate of the true profit	Adjusted_PAT
24	26	CP	Commercial paper , a short-term debt instrument to meet short-term liabilities.	CP
25	27	Revenue earnings in forex	Revenue earned in foreign currency	Rev_earn_in_forex
26	28	Revenue expenses in forex	Expenses due to foreign currency transactions	Rev_exp_in_forex
27	29	Capital expenses in forex	Long term investment in forex	Capital_exp_in_forex
28	30	Book Value (Unit Curr)	Net asset value	Book_Value_Unit_Curr

29	31	Book Value (Adj.) (Unit Curr)	Book value adjusted to reflect asset's true fair market value	Book_Value_Adj_Unit_Curr
30	32	Market Capitalisation	Product of the total number of a company's outstanding shares and the current market price of one share	Market_Capitalisation
31	33	CEPS (annualised) (Unit Curr)	Cash Earnings per Share, profitability ratio that measures the financial performance of a company by calculating cash flows on a per share basis	CEPS_annualised_Unit_Curr
32	34	Cash Flow From Operating Activities	Use of cash from ongoing regular business activities	Cash_Flow_From_Opr
33	35	Cash Flow From Investing Activities	Cash used in the purchase of non-current assets–or long-term assets– that will deliver value in the future	Cash_Flow_From_Inv
34	36	Cash Flow From Financing Activities	Net flows of cash that are used to fund the company (transactions involving debt, equity, and dividends)	Cash_Flow_From_Fin
35	37	ROG-Net Worth (%)	Rate of Growth - Networth	ROG_Net_Worth_perc
36	38	ROG-Capital Employed (%)	Rate of Growth - Capital Employed	ROG_Capital_Employed_perc
37	39	ROG-Gross Block (%)	Rate of Growth - Gross Block	ROG_Gross_Block_perc
38	40	ROG-Gross Sales (%)	Rate of Growth - Gross Sales	ROG_Gross_Sales_perc
39	41	ROG-Net Sales (%)	Rate of Growth - Net Sales	ROG_Net_Sales_perc
40	42	ROG-Cost of Production (%)	Rate of Growth - Cost of Production	ROG_Cost_of_Prod_perc
41	43	ROG-Total Assets (%)	Rate of Growth - Total Assets	ROG_Total_Assets_perc
42	44	ROG-PBIDT (%)	Rate of Growth- PBIDT	ROG_PBIDT_perc
43	45	ROG-PBDT (%)	Rate of Growth- PBDT	ROG_PBDT_perc
44	46	ROG-PBIT (%)	Rate of Growth- PBIT	ROG_PBIT_perc
45	47	ROG-PBT (%)	Rate of Growth- PBT	ROG_PBT_perc
46	48	ROG-PAT (%)	Rate of Growth- PAT	ROG_PAT_perc
47	49	ROG-CP (%)	Rate of Growth- CP	ROG_CP_perc
48	50	ROG-Revenue earnings in forex (%)	Rate of Growth - Revenue earnings in forex	ROG_Rev_earn_in_forex_perc
49	51	ROG-Revenue expenses in forex (%)	Rate of Growth - Revenue expenses in forex	ROG_Rev_exp_in_forex_perc
50	52	ROG-Market Capitalisation (%)	Rate of Growth - Market Capitalisation	ROG_Market_Capitalisation_perc
51	53	Current Ratio[Latest]	Liquidity ratio, company's ability to pay short-term obligations or those due within one year	Curr_Ratio_Latest
52	54	Fixed Assets Ratio[Latest]	Solvency ratio, the capacity of a company to discharge its obligations towards long-term lenders indicating	Fixed_Assets_Ratio_Latest
53	55	Inventory Ratio[Latest]	Activity ratio, specifies the number of times the stock or inventory has been replaced and sold by the company	Inventory_Ratio_Latest
54	56	Debtors Ratio[Latest]	Measures how quickly cash debtors are paying back to the company	Debtors_Ratio_Latest
55	57	Total Asset Turnover Ratio[Latest]	The value of a company's revenues relative to the value of its assets	Total_Asset_Turnover_Ratio_Latest
56	58	Interest Cover Ratio[Latest]	Determines how easily a company can pay interest on its outstanding debt	Interest_Cover_Ratio_Latest
57	59	PBIDTM (%)[Latest]	Profit before Interest Depreciation and Tax Margin	PBIDTM_perc_Latest
58	60	PBITM (%)[Latest]	Profit Before Interest Tax Margin	PBITM_perc_Latest
59	61	PBDTM (%)[Latest]	Profit Before Depreciation Tax Margin	PBDTM_perc_Latest
60	62	CPM (%)[Latest]	Cost per thousand (advertising cost)	CPM_perc_Latest
61	63	APATM (%)[Latest]	After tax profit margin	APATM_perc_Latest
62	64	Debtors Velocity (Days)	Average days required for receiving the payments	Debtors_Vel_Days
63	65	Creditors Velocity (Days)	Average number of days company takes to pay suppliers	Creditors_Vel_Days
64	66	Inventory Velocity (Days)	Average number of days the company needs to turn its inventory into sales	Inventory_Vel_Days
65	67	Value of Output/Total Assets	Ratio of Value of Output (market value) to Total Assets	Value_of_Output_to_Total_Assets
66	68	Value of Output/Gross Block	Ratio of Value of Output (market value) to Gross Block	Value_of_Output_to_Gross_Block

Data Description in brief.

Dataset Head (Snapshot of Top 5 rows):

	Co_Code	Co_Name	Networth Next Year	Equity Paid Up	Networth	Capital Employed	Total Debt	Gross Block	Net Working Capital	Current Assets	...	PBIDTM (%) [Latest]	PBITM (%) [Latest]	PBDTM (%) [Latest]	CPM (%) [Latest]	APATM (%) [Latest]	Debt Velo (Da
0	16974	Hind.Cables	-8021.60	419.36	-7027.48	-1007.24	5936.03	474.30	-1076.34	40.50	...	0.00	0.00	0.00	0.00	0.00	
1	21214	Tata Tele. Mah.	-3986.19	1954.93	-2968.08	4458.20	7410.18	9070.86	-1098.88	486.86	...	-10.30	-39.74	-57.74	-57.74	-87.18	
2	14852	ABG Shipyard	-3192.58	53.84	506.86	7714.68	6944.54	1281.54	4496.25	9097.64	...	-5279.14	-5516.98	-7780.25	-7723.67	-7961.51	
3	2439	GTL	-3054.51	157.30	-623.49	2353.88	2326.05	1033.69	-2612.42	1034.12	...	-3.33	-7.21	-48.13	-47.70	-51.58	
4	23505	Bharati Defence	-2967.36	50.30	-1070.83	4675.33	5740.90	1084.20	1836.23	4685.81	...	-295.55	-400.55	-845.88	379.79	274.79	3

Above is a small snippet of the dataset. There are a total of 3586 rows and 67 columns in the dataset. The dataset contains mixed information across columns ranging from ratios to figures to percentages of a company's financial record.

Variables of 'Co_Code' & 'Co_Name' will not add any significance to our model and hence, we will be dropping these variables.

Target field 'default' is to be created separately using the features for the variable 'Networth_Next_Year'. Once we create the 'default' field, we will also drop 'Networth_Next_Year' as it will create a bias in our model and pose a high weightage amongst the predictors.

Data Info:

```
 29  Book_Value_Adj_Unit_Curr          3582 non-null   float64
 51  Curr_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
 64  Inventory_Vel_Days                3483 non-null   float64
dtypes: float64(63), int64(3), object(1)
```

There are 66 continuous and 1 categorical variable. 63 variables are float type, 3 are integers and 1 is an object. There are null values present in a few variables which are mentioned above. Let us check these missing values data.

Missing Value Count:

```
118
```

```
df.size
```

```
240262
```

```
(118/240262)*100
```

```
0.04911305158535265
```

```
Inventory_Vel_Days                   103
Book_Value_Adj_Unit_Curr               4
Inventory_Ratio_Latest                 1
Interest_Cover_Ratio_Latest            1
Curr_Ratio_Latest                      1
Fixed_Assets_Ratio_Latest              1
Debtors_Ratio_Latest                   1
Total_Asset_Turnover_Ratio_Latest      1
PBIDTM_perc_Latest                     1
PBITM_perc_Latest                      1
PBDTM_perc_Latest                      1
CPM_perc_Latest                        1
APATM_perc_Latest                      1
```

There are in total 118 missing points of the total 240262 features in our dataset which is very insignificant at 0.05% . Also, there are no duplicate records in the dataset. Hence, the data can be said to be relatively good and we shall impute the missing data appropriately as part of our data preprocessing.

Treating missing values appropriately is an important exercise for regression and hence, these would need to be imputed appropriately. We will not be dropping them as there are not large in numbers compared to the overall data and the other features in the same rows could be valuable to our analysis and we must avoid losing them.

Dataset Description (Continuous Variables):

	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	226.61	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
Curr_Assets	3586.00	1960.35	22577.57	-0.91	4.00	24.54	135.28	721166.00
Curr_Liab_and_Prov	3586.00	391.99	2675.00	-0.23	0.73	9.23	65.65	83232.98
Total_Assets_to_Liab	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	3.63	14143.40
Value_Of_Output	3586.00	1077.19	9843.88	-119.10	1.41	30.89	235.84	435559.09
Cost_of_Prod	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
Rev_earn_in_forex	3586.00	131.17	1150.73	0.00	0.00	0.00	7.20	46158.00
Rev_exp_in_forex	3586.00	256.33	4132.34	0.00	0.00	0.00	6.99	193979.73
Capital_exp_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_Opr	3586.00	65.77	1455.05	-25469.23	-0.31	0.45	12.65	44529.40
Cash_Flow_From_Inv	3586.00	-60.87	701.97	-23843.45	-5.12	-0.12	0.12	3732.98
Cash_Flow_From_Fin	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_Prod_perc	3586.00	310.49	5573.22	-2130.23	-7.24	4.42	23.12	267150.00
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	-29.50	4.62	52.91	822400.00
ROG_Rev_earn_in_forex_perc	3586.00	37.23	658.67	-100.00	0.00	0.00	0.00	29084.77
ROG_Rev_exp_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
Curr_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_Vel_Days	3586.00	603.89	10636.76	0.00	8.00	49.00	106.00	514721.00
Creditors_Vel_Days	3586.00	2057.85	54169.48	0.00	8.00	39.00	89.00	2034145.00
Inventory_Vel_Days	3483.00	79.64	137.85	-199.00	0.00	35.00	96.00	996.00
Value_of_Output_to_Total_Assets	3586.00	0.82	1.20	-0.33	0.07	0.48	1.16	17.63
Value_of_Output_to_Gross_Block	3586.00	61.88	976.82	-61.00	0.27	1.53	4.91	43404.00

We have been provided with a very comprehensive list of very important financial ratios and calculations. However, the data has good amount of outliers and zero or close to zero values. Zero values could also be present as the particular ratio may not be relevant to the specific business or industry.

The mean and median figures for almost all variables have a huge difference. The std deviation for a majority of the financial records is high. By the range of Market Capitalization, we can see that the portfolio of clientele is diverse catering to large as well as mid and small businesses.

1.1 Outlier Treatment

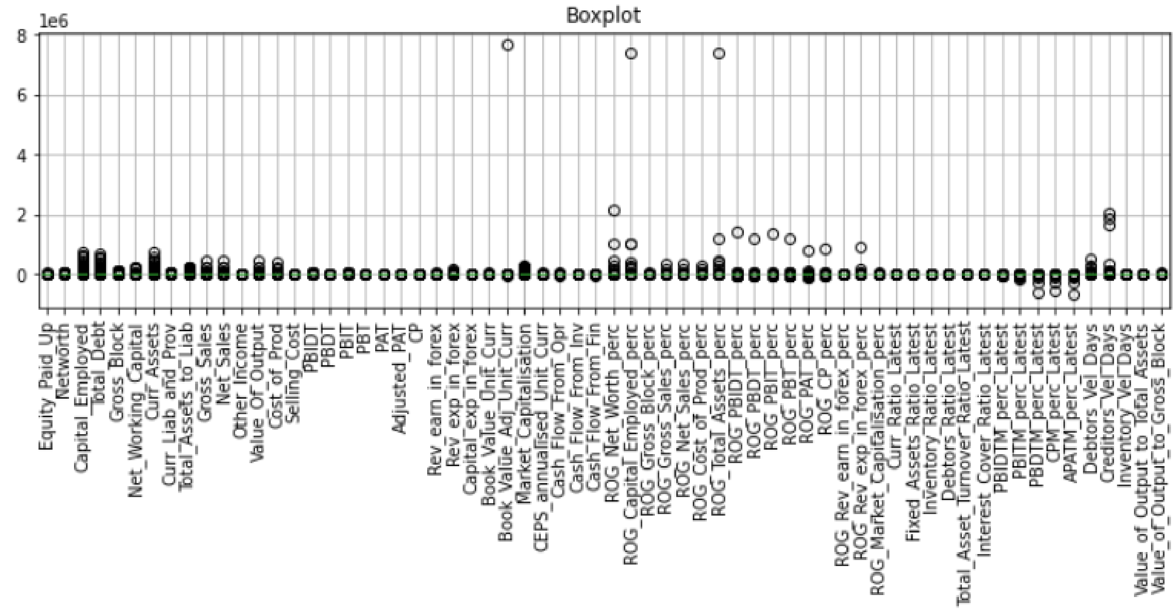


Figure 1 – Outliers Pre Treatment

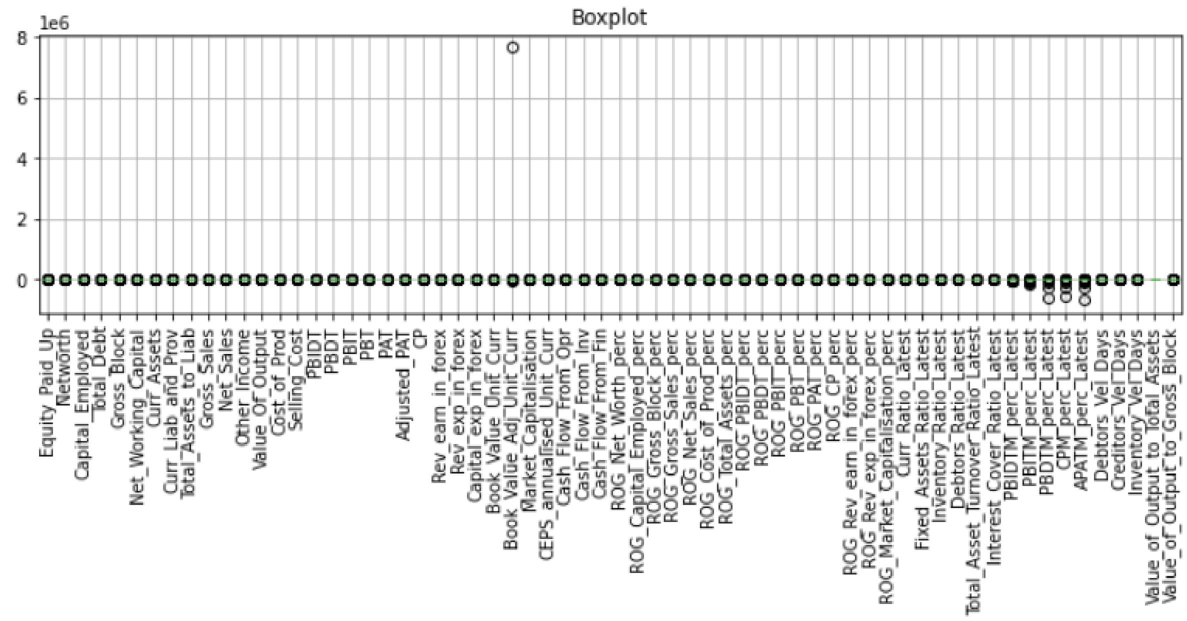


Figure 2 – Outliers Post Treatment

We used the capping and treating technique to treat the outliers where any outlier < Lower Limit (q25-(1.5*IQR) is capped at 5 percentile and any outlier > Upper Limit (q75+(1.5*IQR) per is capped at 95 percentile. In the above image, we can see the outliers have been eliminated as per our capping command.

1.2 Missing Value Treatment

Imputing Missing Values:

```
default                      0
Cash_Flow_From_Fin           0
Cash_Flow_From_Opr           0
CEPS_annualised_Unit_Curr    0
Market_Capitalisation        0
                            ..
ROG_Cost_of_Prod_perc        0
ROG_Net_Sales_perc           0
ROG_Gross_Sales_perc         0
ROG_Gross_Block_perc         0
Equity_Paid_Up               0
Length: 65, dtype: int64
```

We had 118 missing values as per mentioned earlier and have imputed these missing Nan values with KNN imputer having 5 nearest neighbours as parameter for imputation. We chose this method as it is considered as a rather basic and more effective approach than using imputation through mean or median values.

1.3 Transform Target variable into 0 and 1

Target Variable Description:

```
0    3198
1     388
Name: default, dtype: int64

0   0.89
1   0.11
Name: default, dtype: float64
```

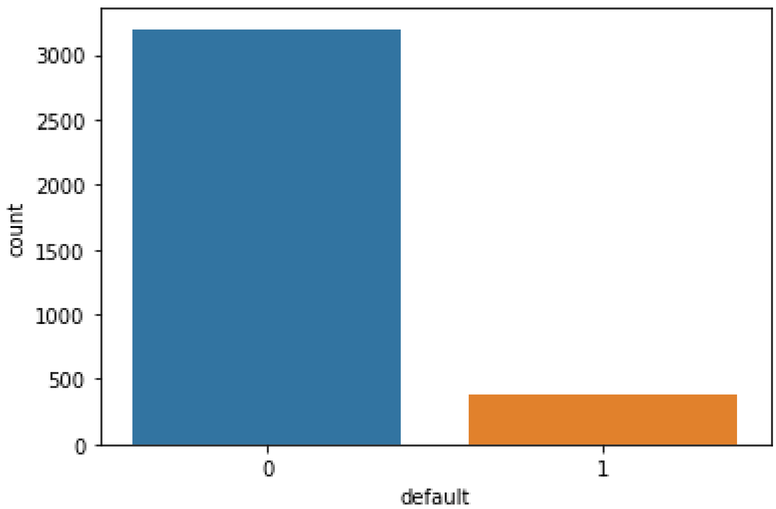


Figure 3: Proportion of Defaulters and Non Defaulters

We have 388 defaulters and 3198 non defaulters in the dataset which means around 11% of the total 3586 clients are defaulters.

Default & Non – Default to Net Sales:

A. Defaulters

```
count     388.00
mean      134.62
std       431.22
min       -62.59
25%         0.00
50%         5.39
75%        60.71
max      5243.07
Name: Net_Sales, dtype: float64
```

B. Non-Defaulters

```
count      3198.00
mean       1194.37
std       10579.01
min         -14.59
25%           1.93
50%          36.59
75%         264.61
max      443775.16
Name: Net Sales, dtype: float64
```

We can see that the mean net sales for defaulters is 135 Cr and median is 5 Cr, while the mean for non-defaulters is 1194 Cr and median is 37 Cr showing the huge gap in the way sales growth is affecting the businesses growth of defaulters.

Default & Non – Default to Profit Before Taxes:

A. Defaulters

```
count      388.00
mean       -47.88
std        156.76
min      -1267.23
25%        -18.09
50%         -1.61
75%         -0.07
max        433.97
```

B. Non-Defaulters

count	3198.00
mean	101.97
std	843.88
min	-6032.34
25%	0.00
50%	0.58
75%	10.36
max	16798.00

We can see that the mean profit before taxes for defaulters is a negative 48 Cr and median is negative 1.6 Cr, while the mean for non-defaulters is 101 Cr and median is 58 lakhs. This helps understand that majority of the defaulters have not been able to generate profits from their businesses.

Default & Non – Default to Total Debt:

A. Defaulters

count	388.00
mean	275.20
std	877.15
min	0.00
25%	2.42
50%	20.02
75%	130.91
max	7410.18

B. Non-Defaulters

count	3198.00
mean	2203.46
std	25037.16
min	-0.72
25%	0.00
50%	6.10
75%	66.05
max	652823.81

We can see that the median total debt for defaulters is a 20 Cr, while the median for non-defaulters is 6 Cr clearly indicating that the non – defaulters are aware of their debt obligations and have been reducing or restricting activity towards the same.

Default & Non – Default to Networth:

A. Defaulters

count	388.00
mean	-113.27
std	472.57
min	-7027.48
25%	-58.35
50%	-8.36
75%	-1.06
max	847.44

B. Non-Defaulters

```
count     3198.00
mean       742.32
std       4320.91
min       -441.86
25%          6.35
50%         25.52
75%        147.89
max      81657.35
```

We can see that the mean networth for defaulters is negative 113 Cr and the median is negative 8 Cr, while for non-defaulters the mean is 742 Cr and median is at 25 Cr. This huge difference between the two and negative mean and median for defaulters indicate that these businesses are struggling to generate equity and capital.

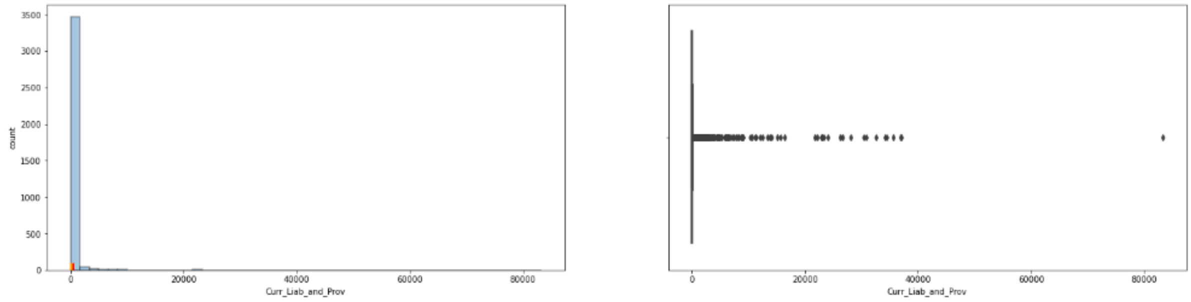
1.4 Univariate & Bivariate Analysis

Univariate Analysis:

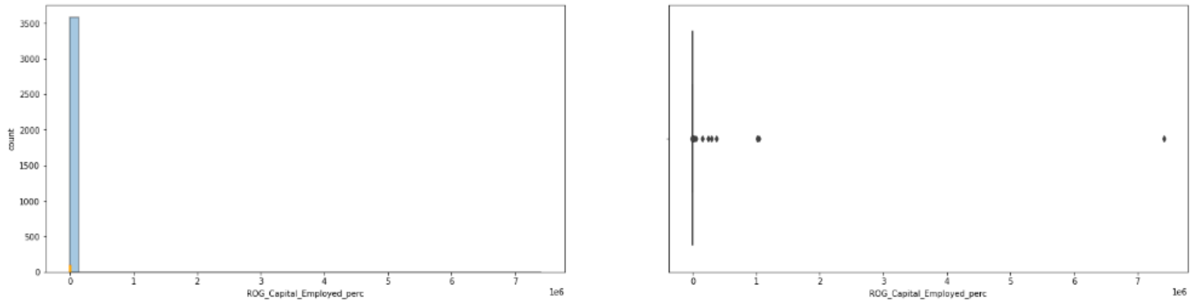
	variables	VIF
1	Curr_Liab_and_Prov	3.36
8	ROG_Capital_Employed_perc	3.35
10	ROG_Total_Assets_perc	3.20
6	CEPS_annualised_Unit_Curr	3.19
4	Book_Value_Unit_Curr	3.07
0	Total_Debt	2.51
2	Selling_Cost	2.34
5	Market_Capitalisation	2.19
3	Rev_exp_in_forex	2.16
7	ROG_Net_Worth_perc	1.90
13	Value_of_Output_to_Gross_Block	1.26
9	ROG_Net_Sales_perc	1.21
12	Debtors_Vel_Days	1.10
11	Interest_Cover_Ratio_Latest	1.03

The above is the snap shot of the 14 variables which based on our analysis suggested were good predictors to determine the defaulters. We shall share analysis, images and references majorly for these variables further in this report.

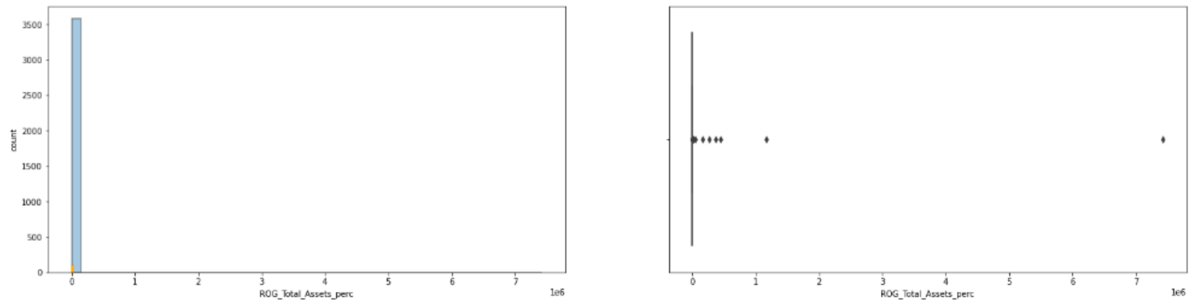
Curr_Liab_and_Prov
Skew: 15.29



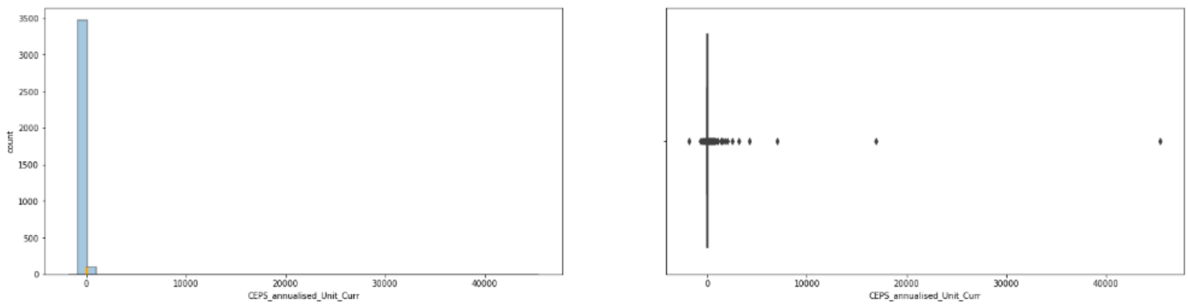
ROG_Capital_Employed_perc
Skew: 56.44



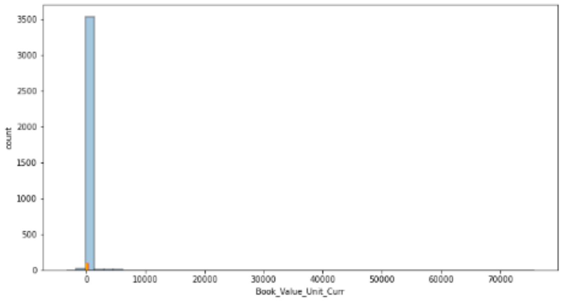
ROG_Total_Assets_perc
Skew: 57.3

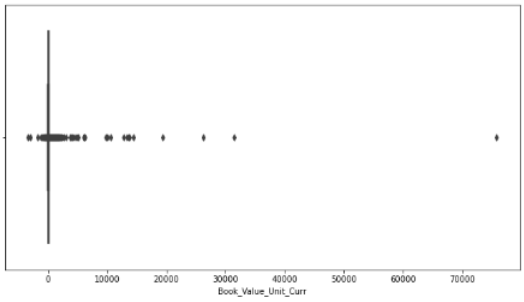


CEPS_annualised_Unit_Curr
Skew: 48.53

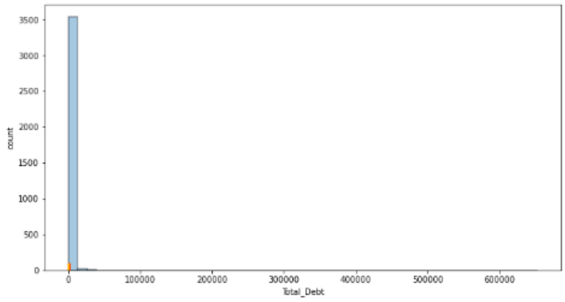


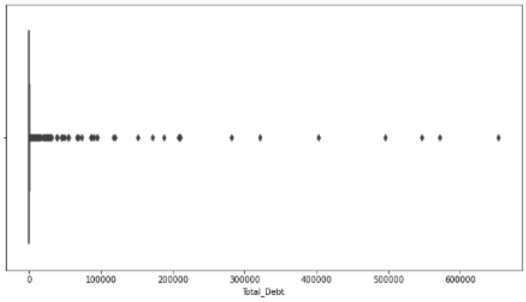
Book_Value_Unit_Curr
Skew: 32.98



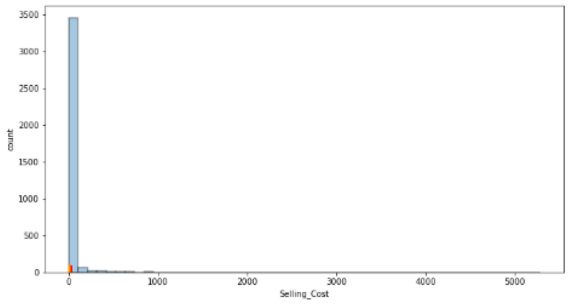


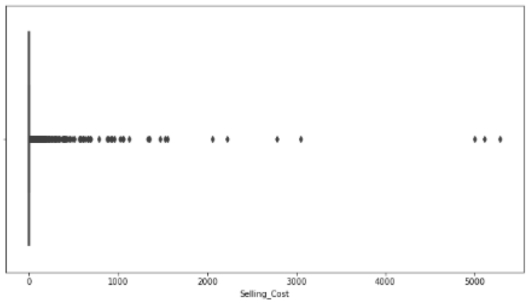
Total_Debt
Skew: 19.42



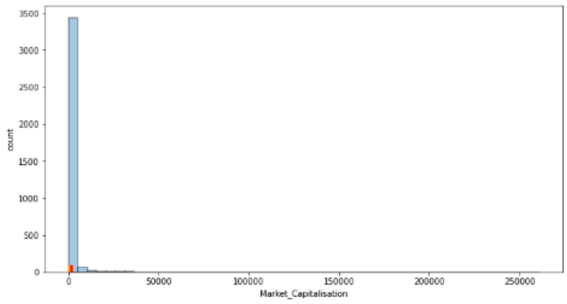


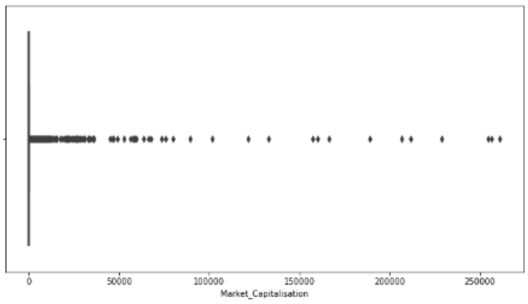
Selling_Cost
Skew: 18.88



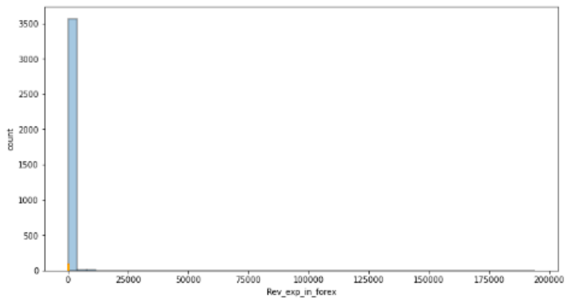


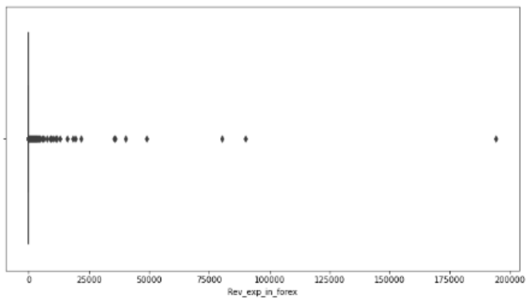
Market_Capitalisation
Skew: 14.39



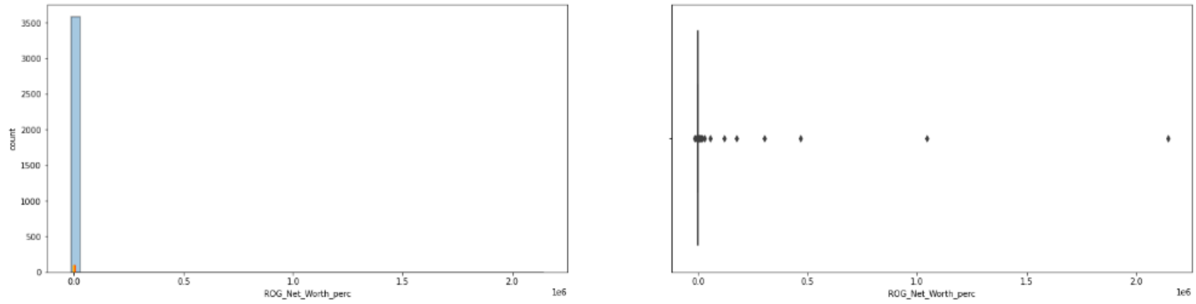


Rev_exp_in_forex
Skew: 34.84

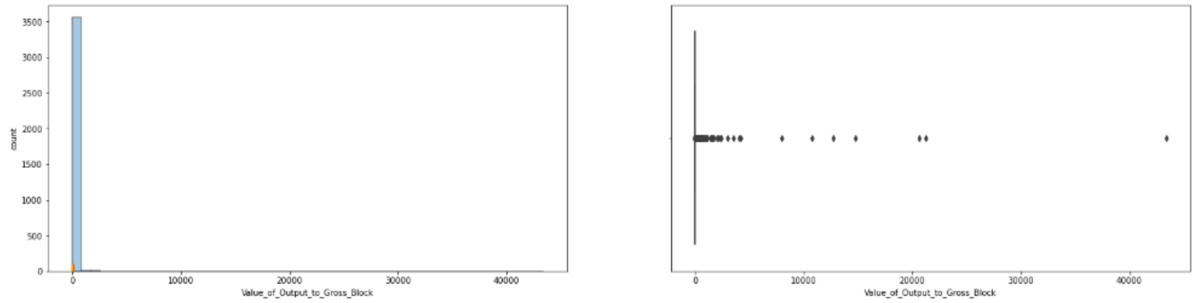




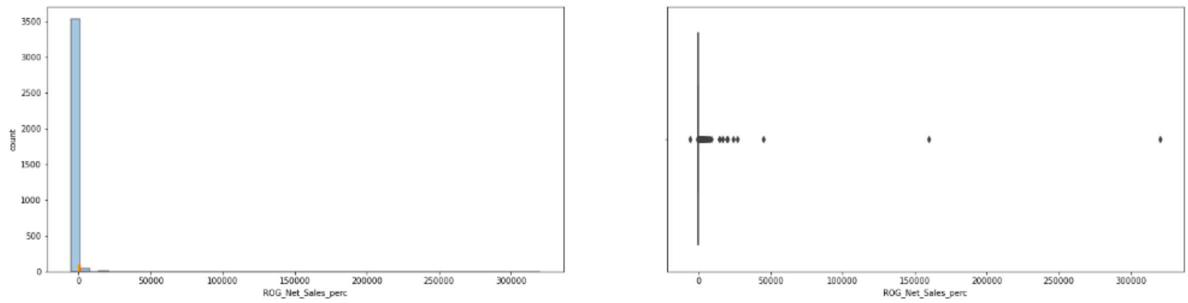
ROG_Net_Worth_perc
Skew: 44.83



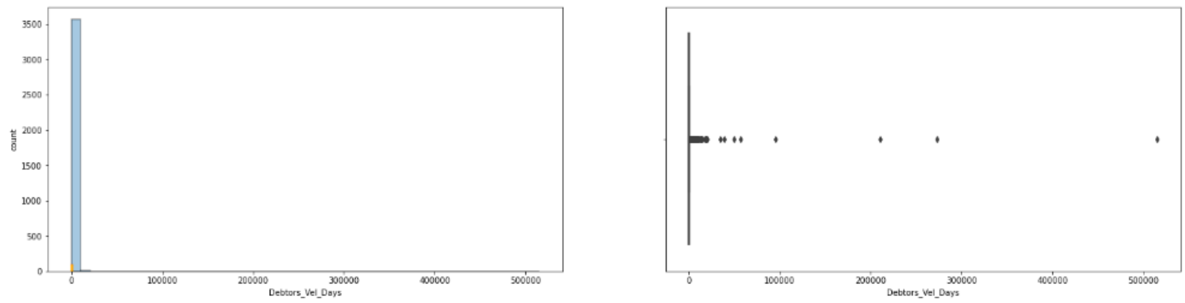
Value_of_Output_to_Gross_Block
Skew: 32.0



ROG_Net_Sales_perc
Skew: 45.41



Debtors_Vel_Days
Skew: 38.66



Interest_Cover_Ratio_Latest
Skew: 40.82

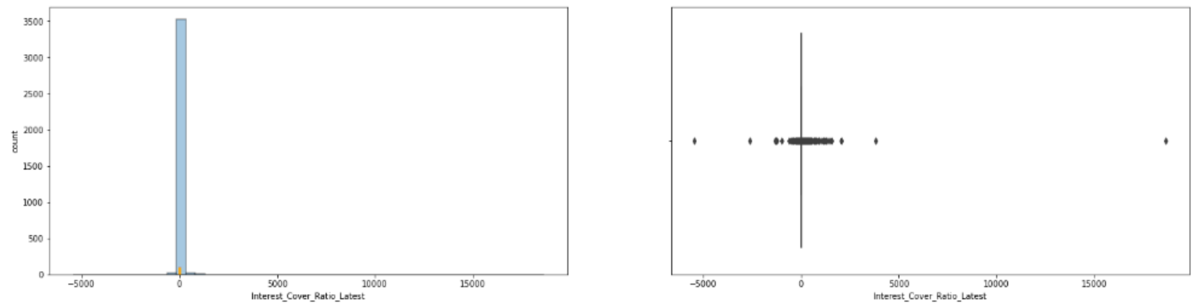


Figure 4: Univariate Analysis

The data is not normally distributed. Skewness is seen to be high with right skewness for all the variables and presence of excess kurtosis can be seen as well. Also, there are outliers present in all of the continuous variables which can have a huge effect on regression. Therefore, we will be treating them appropriately.

Bivariate Analysis:

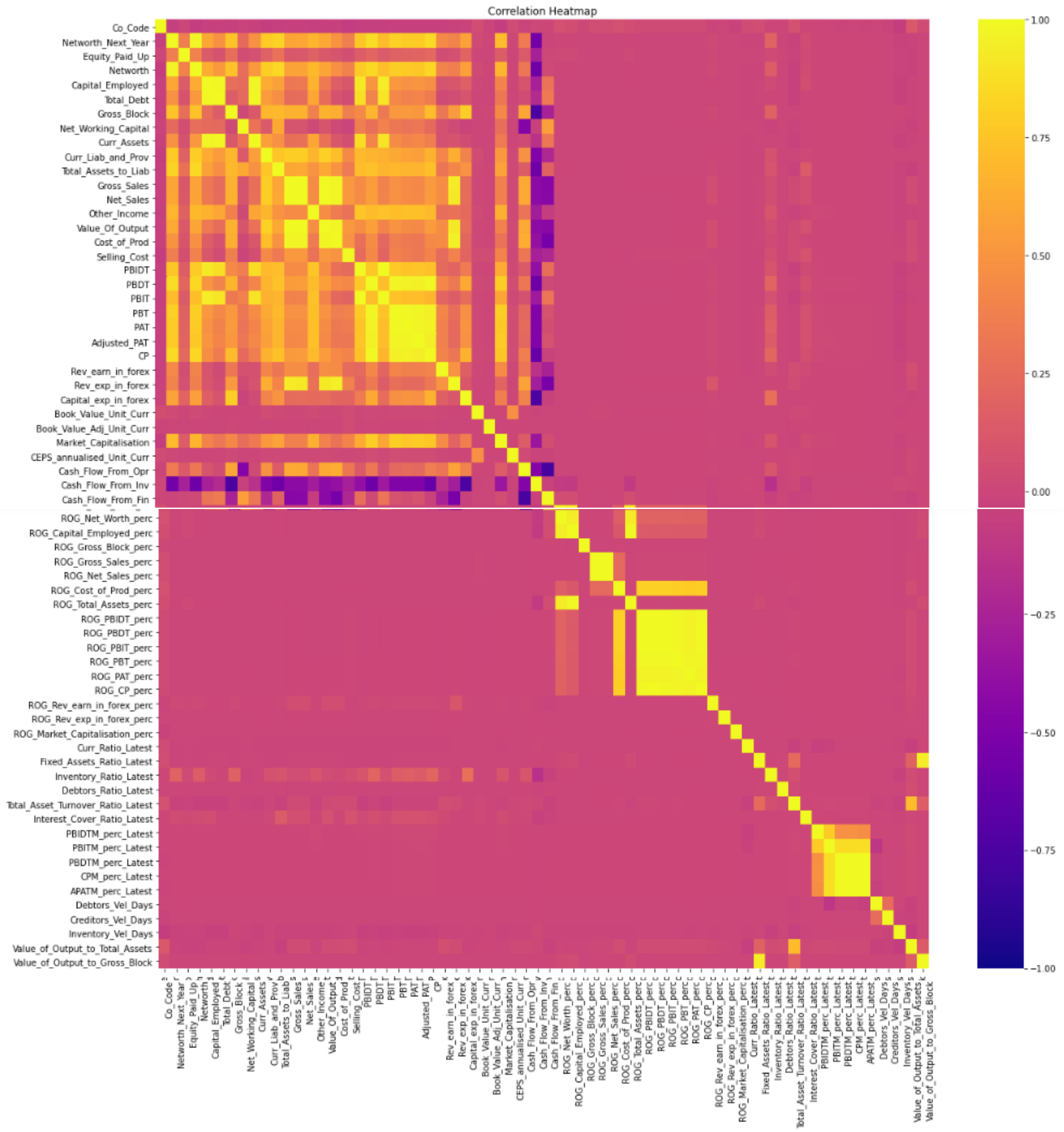


Figure 5 – Bivariate Analysis

It can be seen that a some of variables have strong correlation with each other which is understood given these are financial ratios and they are often interrelated. The cells highlighted in yellow indicates strong correlation, orange with medium correlation, purple with low correlation and violet indicates no correlation.

Let us check correlation for our significant variables:

- Current Assets and Liabilities can be seen having strong correlation with Sales, Profit, Capital, Book Value, Equity. It is understood that in case of increase in profits, capital the assets quality improves and vice versa Similarly, increase in assets benefits the book value, sales, equity etc.

- ROG Capital Employed has strong correlation with ROG Networth, ROG Gross Block, ROG Net Sales, ROG Total Assets and some correlation with other Rate of Growth parameters.

- ROG Total Assets has strong correlation with Cash Flow from Financing Activities & ROG Net_worth_perc

- Cash Earnings per Share and Book Value (Net Asset Value) are correlated with each other

- Total_Debt has strong correlation with Networth, Capital Employed, Current Assets, Current liabilities and provisions, Total Assets to Liabilities, Other Income, PBIDT, PBIT, PAT etc

- Selling_Cost has strong correlation with Networth, Gross Block, Gross Sales, Net Sales, Other Income, Value of Output, Cash Flow from Operations, Market Capitalization, PBT, PAT, CP, Revenue expenses in forex.

- Market_Capitalisation has strong correlation Selling Cost, CP, PAT, PBT, PBIT, PBIDT, Total Assets to Liabilities, Other Income, Networth, Gross Block, Capital Employed etc

- Rev_exp_in_forex has strong correlation Cash Flow from Operations, Net Sales, Gross Sales, Cost of Production, Value of Output, Current liabilities and provisions & Total Assets to Liabilities

- ROG_Net_Worth_perc has strong correlation ROG_Capital_Employed_perc, ROG_Total_Assets_perc

- Value_of_Output_to_Gross_Block has strong correlation Fixed_Asset_Ratio_Latest, Value_of_Output_to_Total_Assets & Total_Asset_Turnover_Ratio_Latest

- ROG_Net_Sales_perc has strong correlation ROG_Gross_Sales_perc & ROG_Cost_of_Prod_perc

- Debtors_Vel_Days has strong correlation with Creditors_Vel_Days

- Interest_Cover_Ratio_Latest has strong correlation Capital_Employed, Total Assets to Liabilities, networth, selling_cost

Also, many important parameters such as Networth, Profit, Sales, Rate of Growth etc had to be dropped from our model building due to high multicollinearity which are otherwise essential data points for any business to determine default probability.

As multicollinearity needs to be dealt with before applying Logistic Regression as otherwise it effects the significance of variables used in the prediction, we would be eliminating variables with high vif ie; Vif >5 using backward elimination approach.

Post dropping variables with VIF >5, we noticed that of 64 total continuous variables, 21 variables had VIF >= 5, which we can say that 1/3rd of the variables or financial parameters had multicollinearity present. Below is the list of balance 43 variables.

Variables with VIF <5 (in descending order)

	variables	VIF
2	Gross_Block	4.65
4	Curr_Liab_and_Prov	4.39
15	Cash_Flow_From_Opr	3.96
14	CEPS_annualised_Unit_Curr	3.76
19	ROG_Capital_Employed_perc	3.59
1	Total_Debt	3.37
23	ROG_Total_Assets_perc	3.35
24	ROG_PBIT_perc	3.24
11	Book_Value_Unit_Curr	3.23
25	ROG_PBT_perc	3.19
7	Adjusted_PAT	3.17
26	ROG_CP_perc	3.14
13	Market_Capitalisation	3.05
41	Value_of_Output_to_Total_Assets	2.98
9	Rev_exp_in_forex	2.92
3	Net_Working_Capital	2.79
16	Cash_Flow_From_Inv	2.76
6	Selling_Cost	2.70
5	Other_Income	2.64
10	Capital_exp_in_forex	2.57
8	Rev_earn_in_forex	2.45
0	Equity_Paid_Up	2.36
17	Cash_Flow_From_Fin	2.35
18	ROG_Net_Worth_perc	2.09
21	ROG_Net_Sales_perc	1.96

34	Total_Asset_Turnover_Ratio_Latest	1.89
22	ROG_Cost_of_Prod_perc	1.87
42	Value_of_Output_to_Gross_Block	1.58
38	Debtors_Vel_Days	1.51
39	Creditors_Vel_Days	1.50
29	ROG_Market_Capitalisation_perc	1.47
37	CPM_perc_Latest	1.34
36	PBIDTM_perc_Latest	1.33
20	ROG_Gross_Block_perc	1.30
40	Inventory_Vel_Days	1.29
28	ROG_Rev_exp_in_forex_perc	1.17
31	Fixed_Assets_Ratio_Latest	1.10
27	ROG_Rev_earn_in_forex_perc	1.10
35	Interest_Cover_Ratio_Latest	1.04
32	Inventory_Ratio_Latest	1.03
30	Curr_Ratio_Latest	1.02
33	Debtors_Ratio_Latest	1.02
12	Book_Value_Adj_Unit_Curr	1.01

Figure 6 – Variables with VIF <5

1.5 Train Test Split

We have split the data into Train and Test dataset in a ratio of 67:33 and used random state =42 as informed. However, we have also used Statify to maintain the same proportion of target variable 'default' in both Train as well as Test datasets.

Train Dataset (Post Split)

```
(2402, 65)
```

Proportion of Default Variable in Train Dataset (Post Split)

```
0.00     2142
1.00      260
```

Test Dataset (Post Split)

```
(1184, 65)
```

Proportion of Default Variable in Test Dataset (Post Split)

```
0.00    1056
1.00     128
```

We can see that the split has divided the Train and Test dataset appropriately in 67:33 and also the proportion of target variable 'default' similarly.

P Value Significance and Logistic Regression Result:

We will now check the significance of the balance variables based on the p value to determine the final set of predictors. We start by fitting these variables in the logistic regression equation and basis on the P value we will further narrow down on the significant predictors.

Significant variables are those with P value < 0.05 (alpha). Hence, we will further eliminate variables with P value > 0.05 using backward elimination approach.

This being an iterative process, it had to be ran 31 times until we were left with 14 variables having P value < 0.05 and which form our Model 31 given below.

Logit Regression Results

Dep. Variable:	default	**No. Observations:**	2402
Model:	Logit	**Df Residuals:**	2387
Method:	MLE	**Df Model:**	14
Date:	Sun, 04 Dec 2022	**Pseudo R-squ.:**	0.6089
Time:	18:05:07	**Log-Likelihood:**	-322.07
converged:	True	**LL-Null:**	-823.47
Covariance Type:	nonrobust	**LLR p-value:**	3.913e-205

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-0.9924	0.136	-7.280	0.000	-1.260	-0.725
Total_Debt	0.0009	0.000	2.117	0.034	6.46e-05	0.002
Curr_Liab_and_Prov	0.0021	0.001	2.510	0.012	0.000	0.004
Selling_Cost	-0.0224	0.010	-2.179	0.029	-0.043	-0.002
Rev_exp_in_forex	0.0033	0.002	2.016	0.044	9.34e-05	0.007
Book_Value_Unit_Curr	-0.1521	0.011	-13.241	0.000	-0.175	-0.130
Market_Capitalisation	-0.0008	0.000	-3.358	0.001	-0.001	-0.000

CEPS_annualised_Unit_Curr	-0.0911	0.035	-2.613	0.009	-0.159	-0.023
ROG_Net_Worth_perc	-0.0139	0.004	-3.398	0.001	-0.022	-0.006
ROG_Capital_Employed_perc	0.0145	0.006	2.419	0.016	0.003	0.026
ROG_Net_Sales_perc	-0.0031	0.001	-2.157	0.031	-0.006	-0.000
ROG_Total_Assets_perc	-0.0146	0.007	-2.028	0.043	-0.029	-0.000
Interest_Cover_Ratio_Latest	-0.0019	0.001	-2.164	0.030	-0.004	-0.000
Debtors_Vel_Days	-0.0011	0.000	-2.356	0.018	-0.002	-0.000
Value_of_Output_to_Gross_Block	-0.0221	0.009	-2.419	0.016	-0.040	-0.004

Figure 7 – Logistic Regression Result for Model 31

We can see that all the above variables are having p value < 0.05 and hence, these 14 variables can be determined as significant predictors & may be useful to discriminate cases of default.

We see that adjusted pseudo R square of 0.591 is now close to pseudo R square of 0.608, thus suggesting presence of lesser insignificant variables in the model no 31 as well.

We now check the VIF for these variables below.

	variables	VIF
1	Curr_Liab_and_Prov	3.36
8	ROG_Capital_Employed_perc	3.35
10	ROG_Total_Assets_perc	3.20
6	CEPS_annualised_Unit_Curr	3.19
4	Book_Value_Unit_Curr	3.07
0	Total_Debt	2.51
2	Selling_Cost	2.34
5	Market_Capitalisation	2.19
3	Rev_exp_in_forex	2.16
7	ROG_Net_Worth_perc	1.90
13	Value_of_Output_to_Gross_Block	1.26
9	ROG_Net_Sales_perc	1.21
12	Debtors_Vel_Days	1.10
11	Interest_Cover_Ratio_Latest	1.03

The VIF is high for a few first variables but they are not > 5. Let us now plot the default variable on these variables.

Plotting Default Variable on Predicted Probability Values from Model 31:

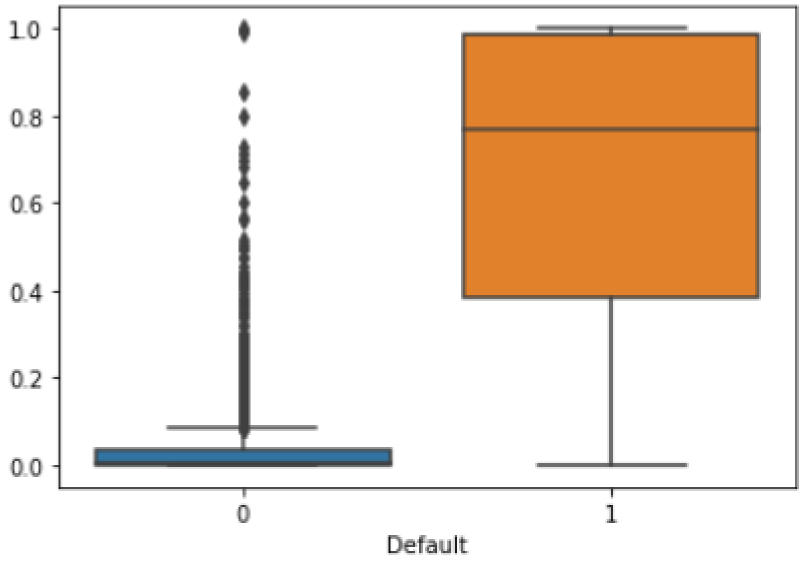


Figure 8 – Default Variable on Predicted Probability Values

We can see that the model is able to differentiate between the probability of defaulters and non – defaulters.

1.6 Build Logistic Regression Model (using statsmodel library) on most important variables on Train Dataset and choose the optimum cutoff. Also showcase your model building approach

Logistic Regression Model Building on Model 31

Model Building using 0.09 Threshold:

- **Confusion matrix for training set of 0.09 threshold**

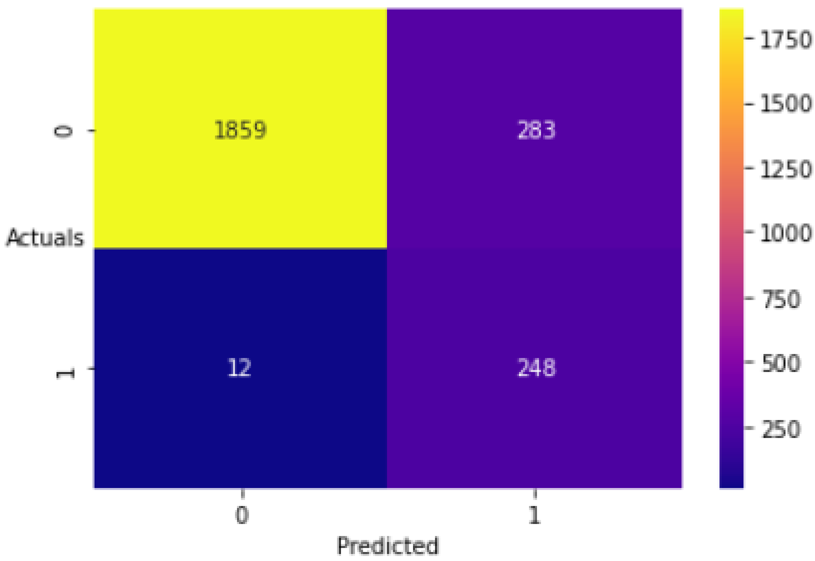


Figure 9 – Confusion matrix for training set of 0.09 threshold

- **Classification Report for training set of 0.09 threshold**

	precision	recall	f1-score	support
0.0	0.994	0.868	0.926	2142
1.0	0.467	0.954	0.627	260
accuracy			0.877	2402
macro avg	0.730	0.911	0.777	2402
weighted avg	0.937	0.877	0.894	2402

Figure 10 – Classification Report for training set of 0.09 threshold

As observed above, accuracy of the model i.e. percentage of overall correct predictions is 87.7% Sensitivity of the model is 95.4% i.e. 95.4% of those defaulted were correctly identified as defaulters by the model. Precision is seen at 46.7% for defaulters.

- **Confusion matrix for test set of 0.09 threshold**

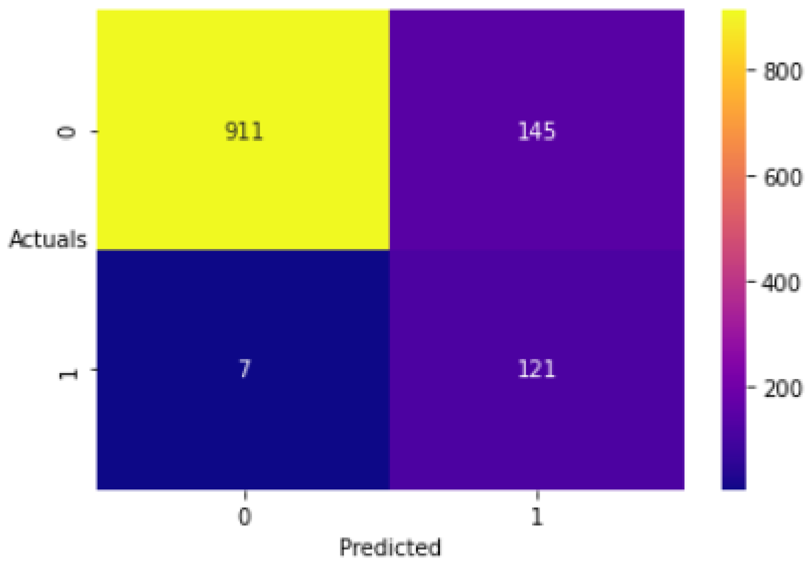


Figure 11 – Confusion matrix for test set of 0.09 threshold

- **Classification Report for test set of 0.09 threshold**

	precision	recall	f1-score	support
0.0	0.992	0.863	0.923	1056
1.0	0.455	0.945	0.614	128
accuracy			0.872	1184
macro avg	0.724	0.904	0.769	1184
weighted avg	0.934	0.872	0.890	1184

Figure 12 – Classification Report for test set of 0.09 threshold

As observed above, accuracy of the model i.e. percentage of overall correct predictions is 87.2% Sensitivity of the model is 94.5% i.e. 94.5% of those defaulted were correctly identified as defaulters by the model and the result is similar to the train data as well. Precision is 45.5 % for the defaulters.

Model Building using 0.05 Threshold:

- **Confusion matrix for training set of 0.05 threshold**

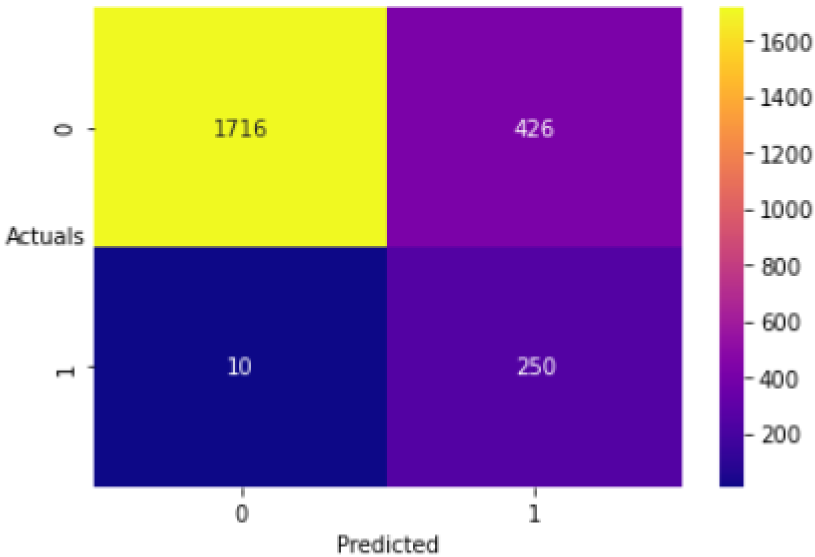


Figure 13 – Confusion matrix for training set of 0.05 threshold

- **Classification Report for training set of 0.05 threshold**

	precision	recall	f1-score	support
0.0	0.994	0.801	0.887	2142
1.0	0.370	0.962	0.534	260
accuracy			0.818	2402
macro avg	0.682	0.881	0.711	2402
weighted avg	0.927	0.818	0.849	2402

Figure 14 – Classification Report for training set of 0.05 threshold

As observed above, accuracy of the model i.e. percentage of overall correct predictions is 81.8% Sensitivity of the model is 96.2% i.e. 96.2% of those defaulted were correctly identified as defaulters by the model.

Whereas, the sensitivity is slightly better for defaulter compared to the previous model with threshold of 0.09, the accuracy has dropped and the precision has dropped drastically. Hence, overall the previous model with a threshold of 0.09 seems better. Let us check results for Test set as well.

- **Confusion matrix for test set of 0.05 threshold**

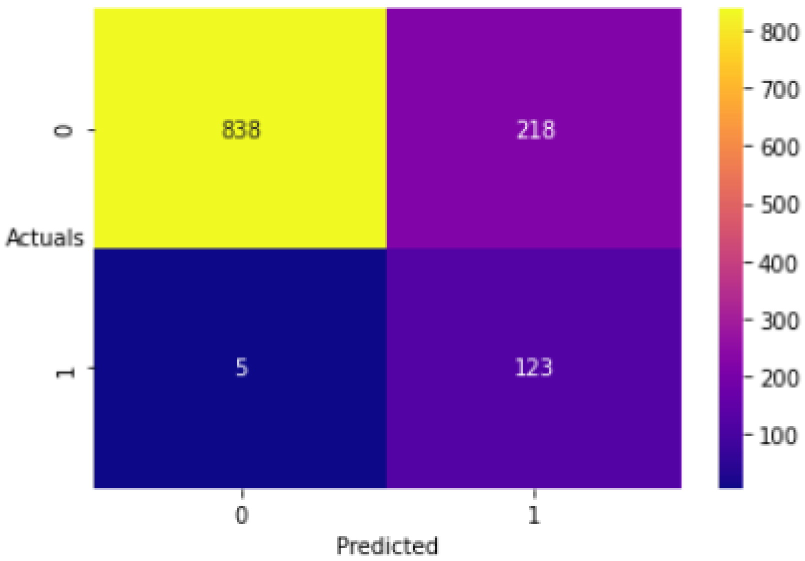


Figure 15 – Confusion matrix for test set of 0.05 threshold

- **Classification Report for test set of 0.05 threshold**

	precision	recall	f1-score	support
0.0	0.994	0.794	0.883	1056
1.0	0.361	0.961	0.525	128
accuracy			0.812	1184
macro avg	0.677	0.877	0.704	1184
weighted avg	0.926	0.812	0.844	1184

Figure 16 – Classification Report for test set of 0.05 threshold

As observed above, accuracy of the model i.e. percentage overall correct predictions is 81.2% Sensitivity of the model is 96.1% i.e. 96.1% of those defaulted were correctly identified as defaulters by the model and the result is similar to the train data as well. However, the precision has decreased to only 36.1%.

Whereas, the sensitivity is slightly better for defaulter compared to the previous model with threshold of 0.09 but the accuracy has dropped and the precision has dropped drastically. Hence, overall the previous model with a threshold of 0.09 seems better on test set as well.

We will now try to determine the model's result using optimum threshold to see if the performance increases.

Model Building using 0.178 Threshold:

- **Confusion matrix for training set of 0.178 optimum threshold**

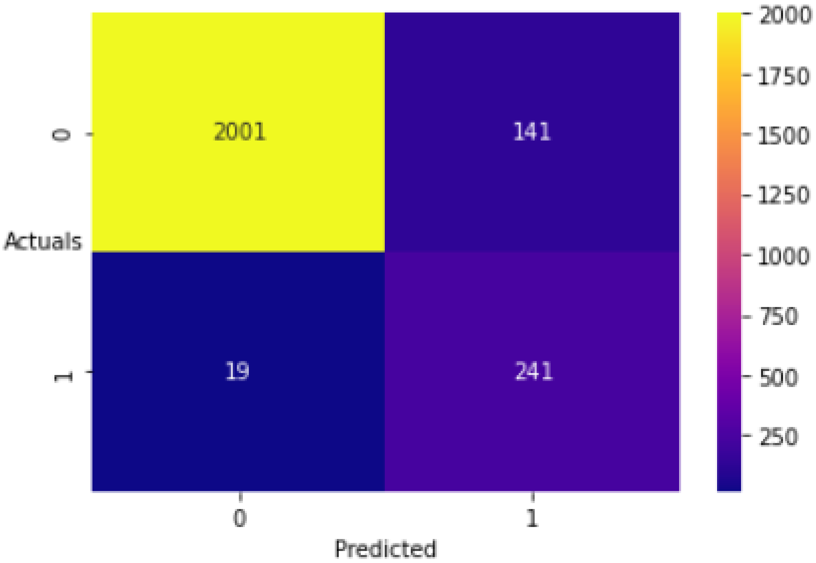


Figure 17 – Confusion matrix for training set of 0.178 threshold

- **Classification Report for training set of 0.178 optimum threshold**

	precision	recall	f1-score	support
0.0	0.991	0.934	0.962	2142
1.0	0.631	0.927	0.751	260
accuracy			0.933	2402
macro avg	0.811	0.931	0.856	2402
weighted avg	0.952	0.933	0.939	2402

Figure 18 – Classification Report for training set of 0.178 threshold

As observed above, accuracy of the model i.e. percentage of overall correct predictions is 93.3% which is the best so far. Sensitivity of the model is 92.7% i.e. 92.7% of those defaulted were correctly identified as defaulters by the model. Also, the precision has improved drastically to 63.1% for this model.

- **Confusion matrix for test set of 0.178 optimum threshold**

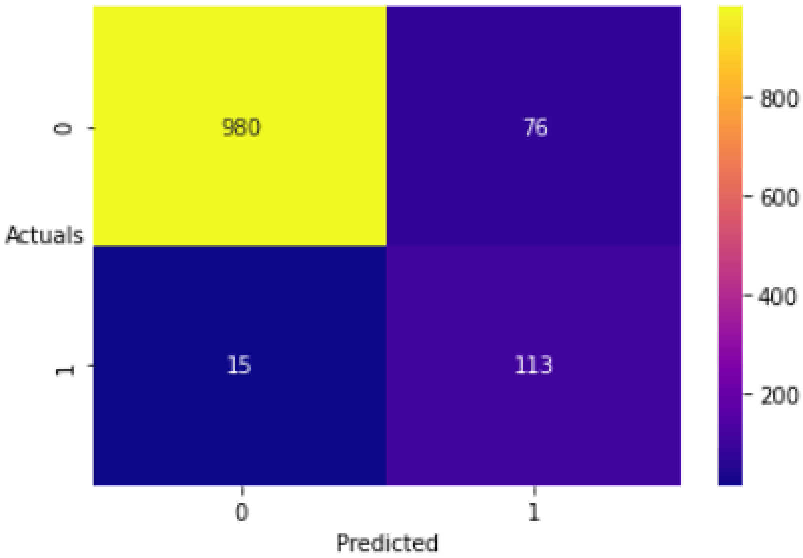


Figure 19 – Confusion matrix for test set of 0.178 threshold

- **Classification Report for test set of 0.178 optimum threshold**

	precision	recall	f1-score	support
0.0	0.985	0.928	0.956	1056
1.0	0.598	0.883	0.713	128
accuracy			0.923	1184
macro avg	0.791	0.905	0.834	1184
weighted avg	0.943	0.923	0.929	1184

Figure 20 – Classification Report for test set of 0.178 threshold

As observed above, accuracy of the model i.e. percentage of overall correct predictions is 92.3% Sensitivity of the model is 88.3% i.e. 88.3% of those defaulted were correctly identified as defaulters by the model but the result is not similar to the train data. The precision has also dropped to 59.8.1% compared to the training set. It is observed that the model has performed better on the train set but not on the test set.

In comparison with the previous model with 0.09 threshold, the overall accuracy and precision has increased in this model from 87.2% to 92.3% and 45.5% to 59.8% on the test sets respectively. However, the Sensitivity has decreased drastically from 94.5% to 88.3%.

1.7 Validate the Model on Test Dataset and state the performance matrices. Also state interpretation from the model

Our objective of the analysis is to correctly predict the defaulters which the model with 0.178 optimum threshold has not been able to perform compared to model with 0.09 threshold. Hence, we can determine that though it is a more balanced model, it is not a better model compared to the model with 0.09 threshold when predicting defaulters.

Hence, we determine that the model with 0.09 threshold performed better and is the optimum Logistic Regression model for analysis of this dataset. Below is the table with the summary of the performance metrices for all models build.

	ACCURACY (%)	PRECISION (%)	RECALL (%)	F SCORE (%)
Thres 0.09 TRAIN	**87.7**	**46.7**	**95.4**	**62.7**
Thres 0.09 TEST	**87.2**	**45.5**	**94.5**	**61.4**
Thres 0.05 TRAIN	**81.8**	**37**	**96.2**	**53.4**
Thres 0.05 TEST	**81.2**	**36.1**	**96.1**	**52.5**
Thres 0.178 TRAIN	**93.3**	**63.1**	**92.7**	**75.1**
Thres 0.178 TEST	**92.3**	**59.8**	**88.3**	**71.3**

Table 1 - Comparison Chart LR Models

Interpretations Basis Significant Variables:

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-0.9924	0.136	-7.280	0.000	-1.260	-0.725
Total_Debt	0.0009	0.000	2.117	0.034	6.46e-05	0.002
Curr_Liab_and_Prov	0.0021	0.001	2.510	0.012	0.000	0.004
Selling_Cost	-0.0224	0.010	-2.179	0.029	-0.043	-0.002
Rev_exp_in_forex	0.0033	0.002	2.016	0.044	9.34e-05	0.007
Book_Value_Unit_Curr	-0.1521	0.011	-13.241	0.000	-0.175	-0.130
Market_Capitalisation	-0.0008	0.000	-3.358	0.001	-0.001	-0.000
CEPS_annualised_Unit_Curr	-0.0911	0.035	-2.613	0.009	-0.159	-0.023
ROG_Net_Worth_perc	-0.0139	0.004	-3.398	0.001	-0.022	-0.006
ROG_Capital_Employed_perc	0.0145	0.006	2.419	0.016	0.003	0.026
ROG_Net_Sales_perc	-0.0031	0.001	-2.157	0.031	-0.006	-0.000
ROG_Total_Assets_perc	-0.0146	0.007	-2.028	0.043	-0.029	-0.000
Interest_Cover_Ratio_Latest	-0.0019	0.001	-2.164	0.030	-0.004	-0.000
Debtors_Vel_Days	-0.0011	0.000	-2.356	0.018	-0.002	-0.000
Value_of_Output_to_Gross_Block	-0.0221	0.009	-2.419	0.016	-0.040	-0.004

Figure 21 – Coefficients for Significant Variables

Interpretations:

- Our interpretations will focus on guiding business to take corrective measures basis on coefficients observed for the significant variables as per Figure 21 above.
- **Total Debt** has a positive coefficient indicating that any addition to debt would likely increase the possibility of default. Hence, businesses should focus on reducing debts and refrain from any further borrowings.
- **Current Liabilities and Provisions** has a positive coefficient indicating that any addition to the same would likely increase the possibility of default. Hence, businesses should focus on paying them off asap or reducing them systematically over a period of time.
- **Selling Cost** has a negative coefficient indicating that any addition to the same would likely decrease the possibility of default. Hence, businesses should focus on increasing the demand for its products. Raising sales targets and offering higher commissions for achievers such as sales personals, partners, distributors etc can be tried. Reward and Recognition programs can be participated in. Profit from Operations while reducing cost of operations can be focused on. Building strong sales, marketing and leadership teams can be prioritized. Engaging in promotional activities and testing results for the same can also work for some businesses. However, they should refrain from incurring significant expenses to achieve any of these targets.
- **Revenue expenses in forex** has a positive coefficient indicating that any addition to the same would likely increase the possibility of default. Businesses should focus on profitable supply demand engagements only and minimize any non-profitable businesses especially w.r.t overseas clients, overseas subsidiaries, partners etc.
- **Book Value (Unit Curr)** has a negative coefficient indicating that any addition to the same would likely decrease the possibility of default. Increasing assets and reducing liabilities and generating equity over a period of time would help businesses increase their Book Value.
- **Market Capitalisation** has a negative coefficient indicating that any addition to the same would likely decrease the possibility of default. Increasing assets, generating profits, reducing debts, rewarding shareholders could increase the demand for the business and shareholders which will help increase Market Cap.
- **Cash Earnings per Share (annualised)** has a negative coefficient indicating that any addition to the same would likely decrease the possibility of default. Business need to reduce debtor days, improve on receiving payments from customers at the earliest, reduce cash outflows, increase on receiving advance payments.
- **Rate of Growth – Networth** has a negative coefficient indicating that any addition to the same would likely decrease the possibility of default. Increasing assets and reducing debt consistently and regularly can help increase a company's net worth. Also, improving revenues and profits at a rapid pace by building superb revenue streams help in growth of net worth.
- **Rate of Growth - Capital Employed** has a positive coefficient indicating that any addition to the same would likely increase the possibility of default. Businesses need to reduce costs and cash outflows while increase working capital and cash inflows rapidly.
- **Rate of Growth - Net Sales** has a negative coefficient indicating that any addition to the same would likely decrease the possibility of default. Companies can boost their product portfolio, promotional activities, increase in sales targets, consideration towards employee engagement, revision to incentive and commission programs for employees and partners alike. Also, considering reward and recognition practices can help boost sales.
- **Rate of Growth - Total Assets** has a negative coefficient indicating that any addition to the same would likely decrease the possibility of default. Increasing revenue generating tangible

as well as non-tangible assets over a period of time and recurringly will improve the growth of total assets. Also, renting and/or leasing assets can be a better option to cut cost.

- **Interest Cover Ratio** has a negative coefficient indicating that any addition to the same would likely decrease the possibility of default. Business need to focus on debt reduction and cutting cost while improving revenues and profits. They need to better negotiate with lenders.
- **Debtors Velocity (Days)** has a negative coefficient indicating that any addition to the same would likely decrease the possibility of default. Business need to negotiate on reducing payment receipt days with their customers, creditors, etc. Demanding for advance payments especially for large future projects can also help. Plus focusing on improving good customer and creditors base is also equally important.
- **Value of Output to Gross Block** has a negative coefficient indicating that any addition to the same would likely decrease the possibility of default. Business need to focus on improving and retaining profitable assets and selling non profitable assets, businesses verticals. They can also loan their assets such as equipment, buildings, spaces to other businesses in return for rent.

THE END