11/13/2022

FRA Milestone 1

Great Learning

Shailesh Pande
PUNE

Contents

1 Problem Statement

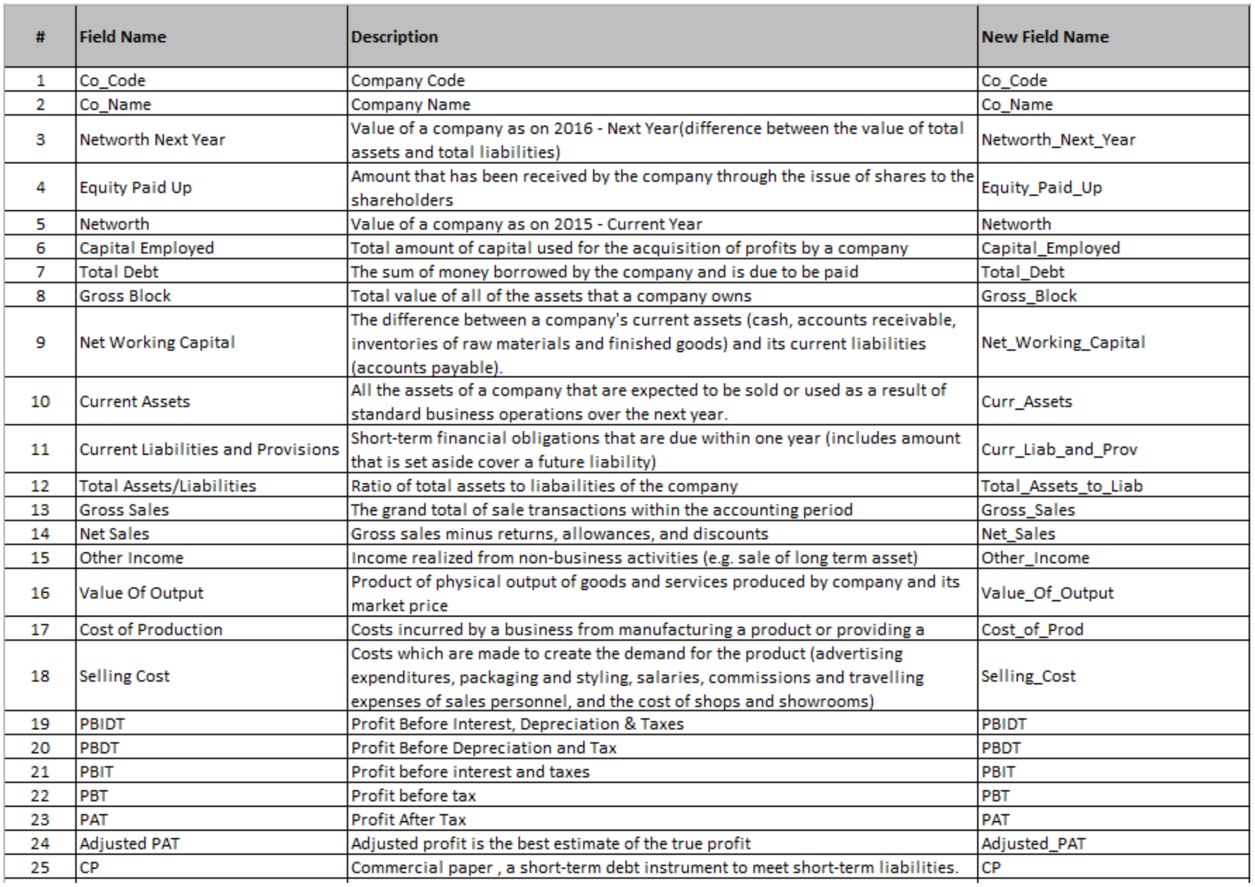
Businesses or companies can fall prey to default if they are not able to keep up their debt obligations. Defaults will lead to a lower credit rating for the company, which in turn reduces its chances of getting credit in the future and may have to pay higher interests on existing debts as well as any 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.

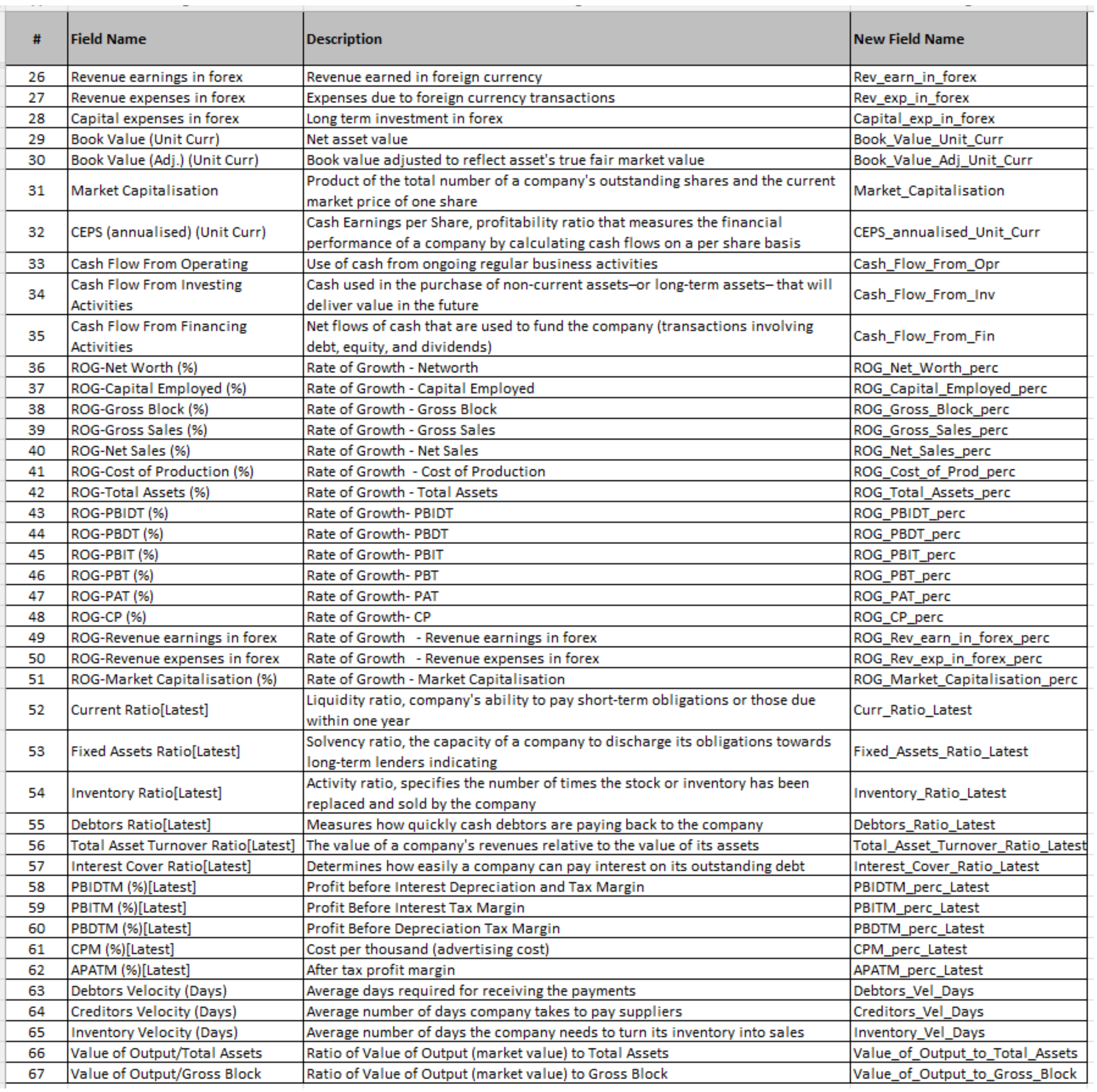
Data that is available includes information from the financial statement of the companies for the previous year (2015). Also, information about the Net worth of the company in the following year (2016) is provided which can be used to drive the labeled field.

Data Dictionary is as under

Table 1: Data Dictionary



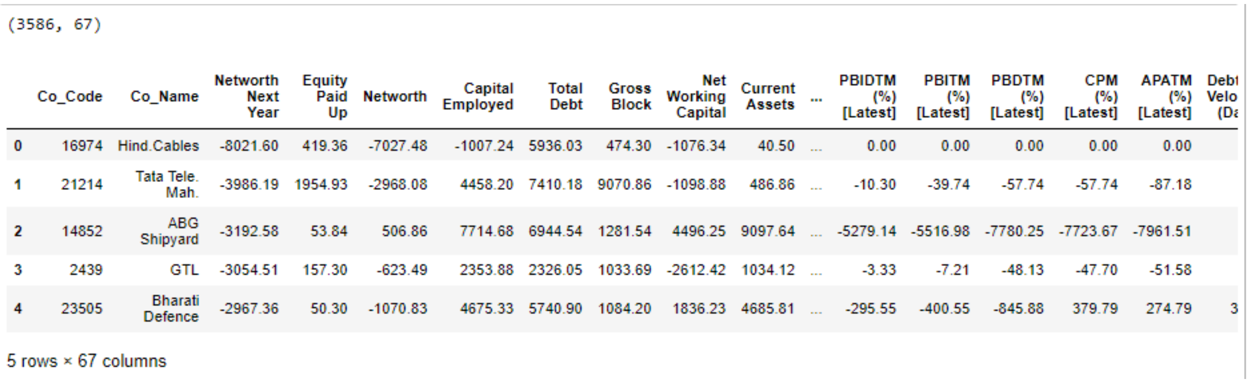
#	Field Name	Description	New Field Name
1	Co_Code	Company Code	Co_Code
2	Co_Name	Company Name	Co_Name
3	Networth Next Year	Value of a company as on 2016 - Next Year(difference between the value of total assets and total liabilities)	Networth_Next_Year
4	Equity Paid Up	Amount that has been received by the company through the issue of shares to the shareholders	Equity_Paid_Up
5	Networth	Value of a company as on 2015 - Current Year	Networth
6	Capital Employed	Total amount of capital used for the acquisition of profits by a company	Capital_Employed
7	Total Debt	The sum of money borrowed by the company and is due to be paid	Total_Debt
8	Gross Block	Total value of all of the assets that a company owns	Gross_Block
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
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
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
12	Total Assets/Liabilities	Ratio of total assets to liabailities of the company	Total_Assets_to_Liab
13	Gross Sales	The grand total of sale transactions within the accounting period	Gross_Sales
14	Net Sales	Gross sales minus returns, allowances, and discounts	Net_Sales
15	Other Income	Income realized from non-business activities (e.g. sale of long term asset)	Other_Income
16	Value Of Output	Product of physical output of goods and services produced by company and its market price	Value_Of_Output
17	Cost of Production	Costs incurred by a business from manufacturing a product or providing a	Cost_of_Prod
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
19	PBIDT	Profit Before Interest, Depreciation & Taxes	PBIDT
20	PBDT	Profit Before Depreciation and Tax	PBDT
21	PBIT	Profit before interest and taxes	PBIT
22	PBT	Profit before tax	PBT
23	PAT	Profit After Tax	PAT
24	Adjusted PAT	Adjusted profit is the best estimate of the true profit	Adjusted_PAT
25	CP	Commercial paper , a short-term debt instrument to meet short-term liabilities.	CP



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

- The data (top 5 rows) is as under

Table 2 : Top 5 rows of raw data



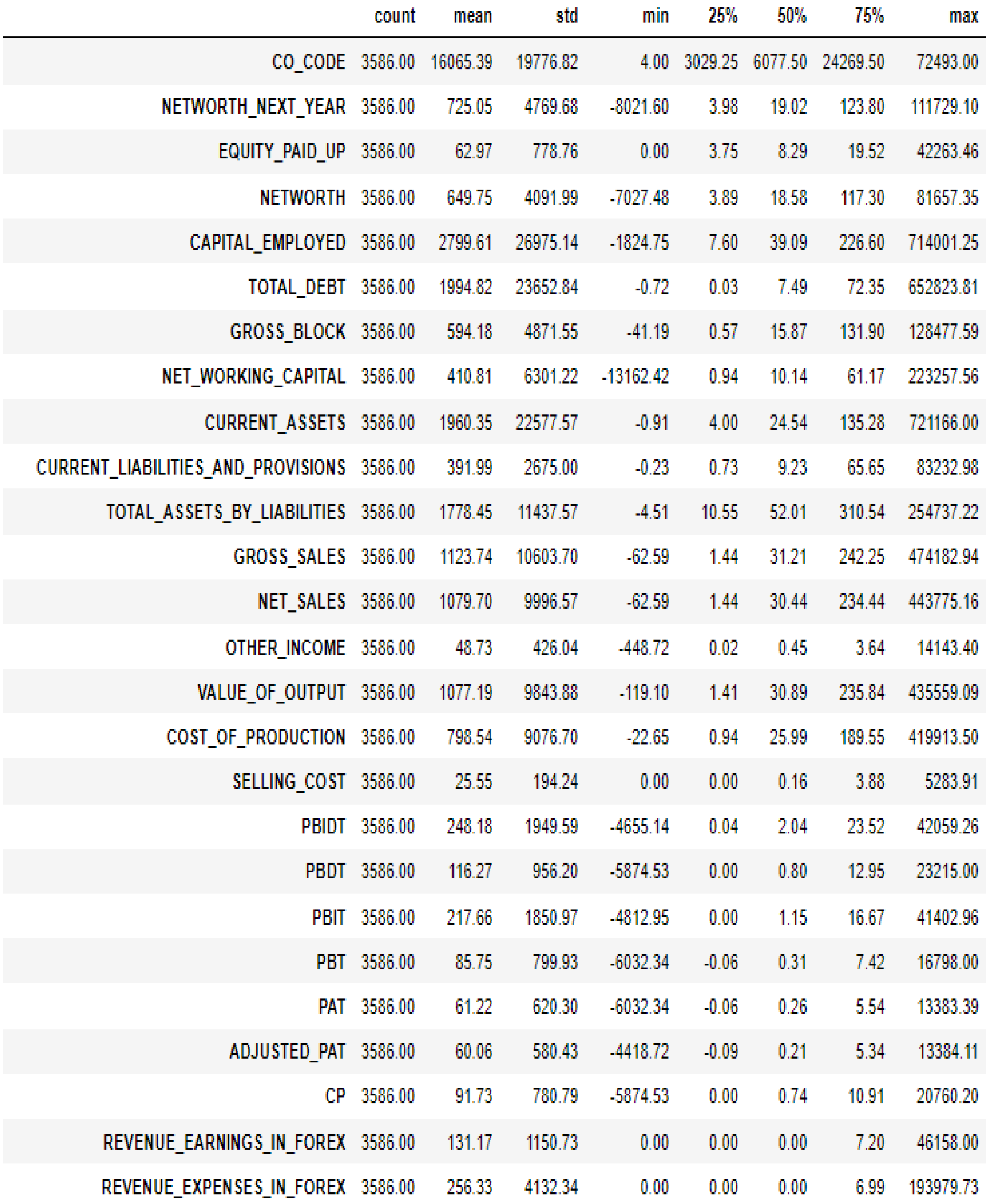
(3586, 67)

	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

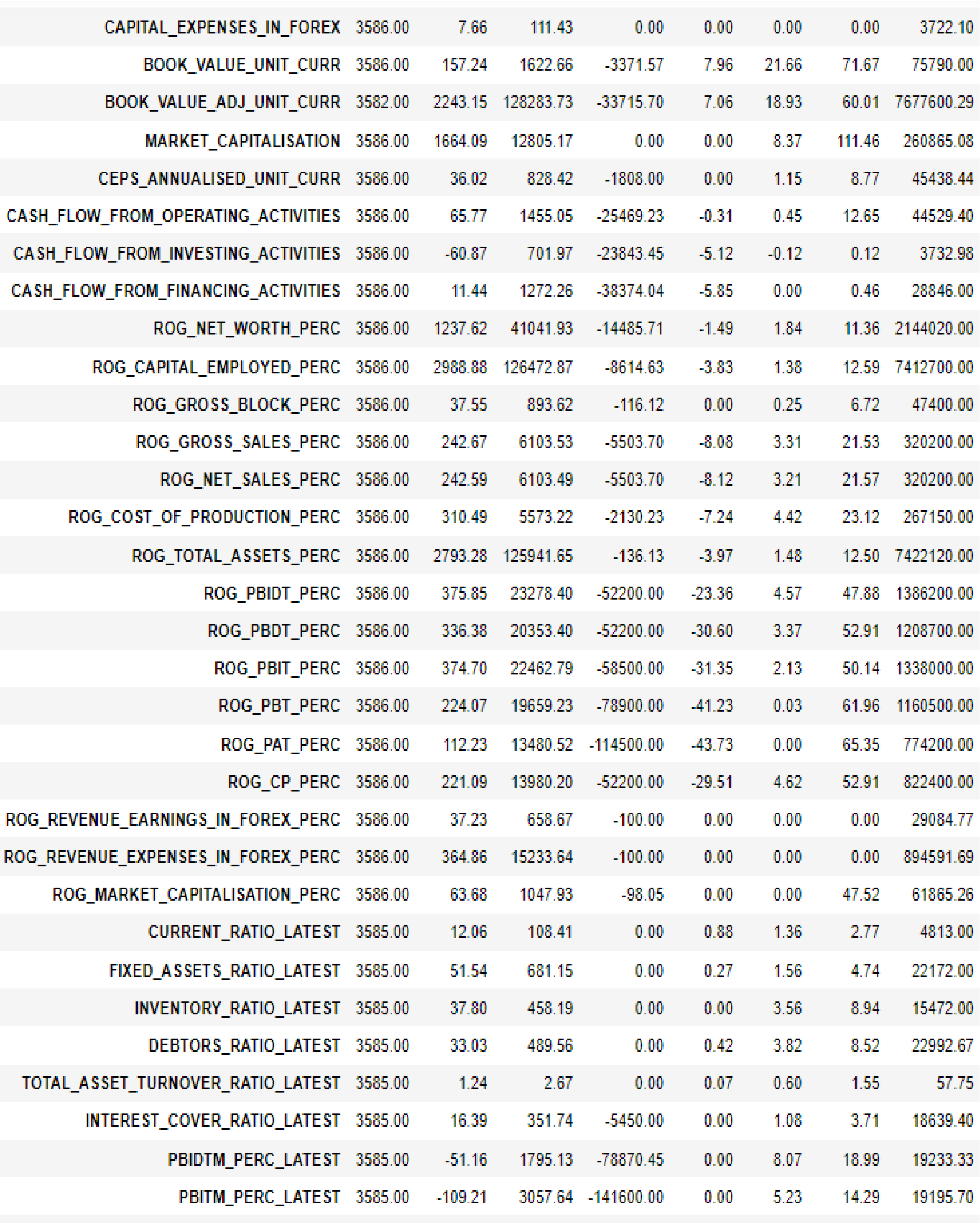
5 rows × 67 columns

- There are 3586 rows and 67 columns
- There are a total of 118 missing values
- Column names are cleaned up and are made into upper case for ease of workability
- There are no duplicate rows
- Out of the 67 columns , only two columns namely CO_CODE and CO_NAME is categorical in nature . All other features are numeric
- The aforementioned categorical features are dropped as they will be redundant for our study.
- The Statistical Summary of the data

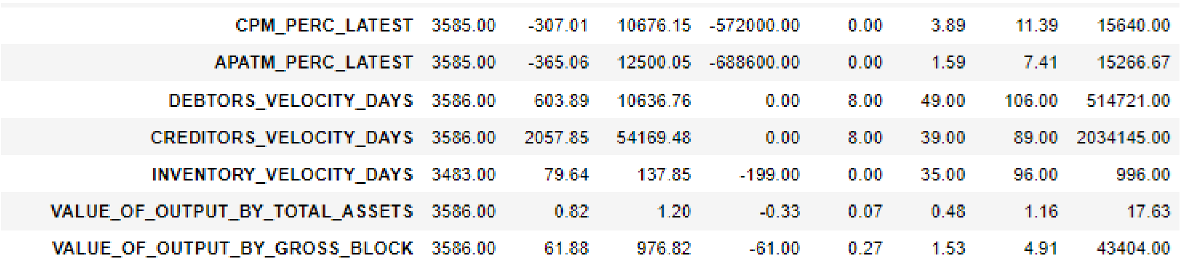
Table 3 : 5-point Statistical Summary of Features



	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



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

2 Outlier Treatment

Any data lying outside the upper bound and lower bound as defined here is

construed as an outlier

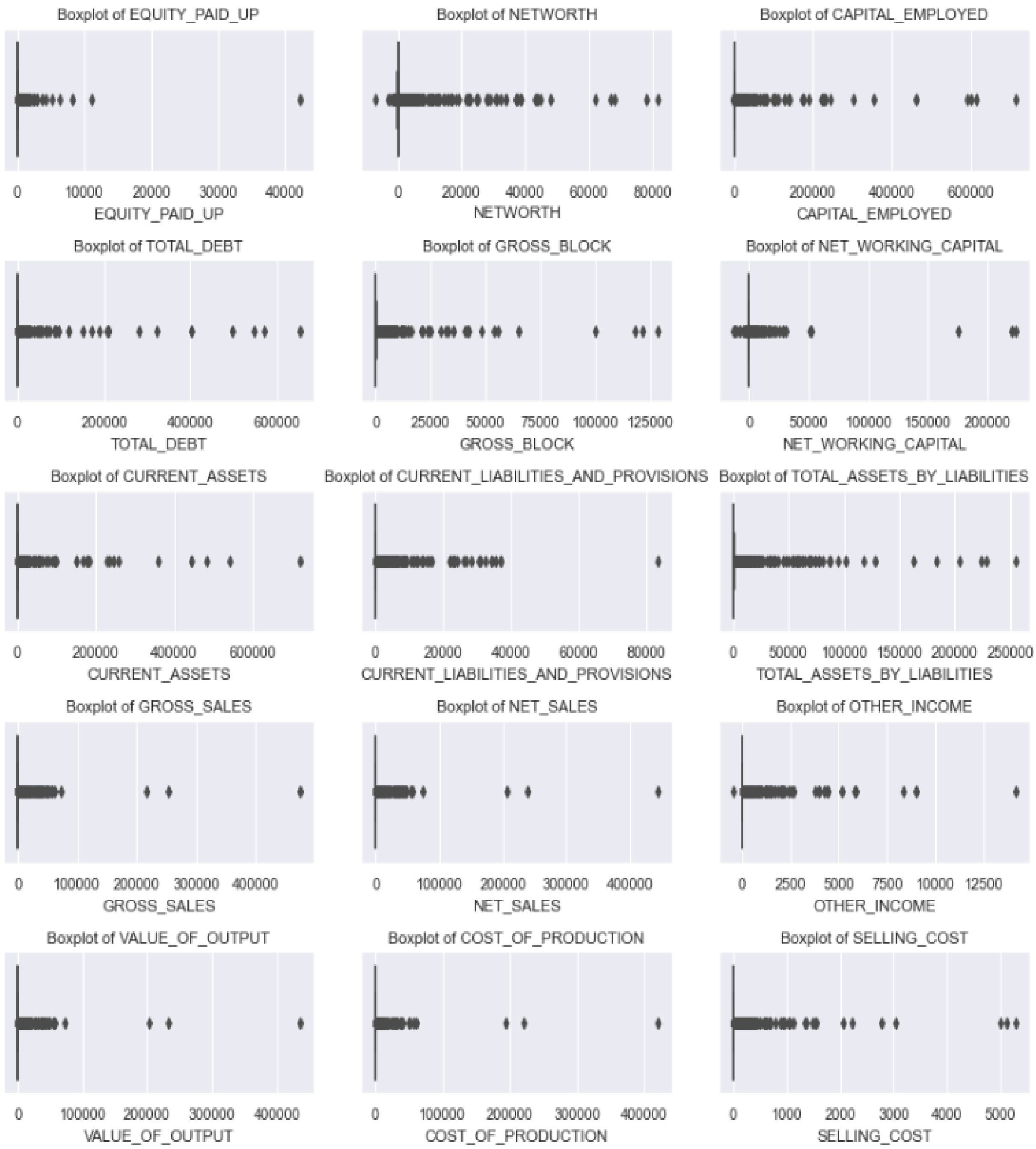
Upper Limit = Q3 + 1.5 x IQR

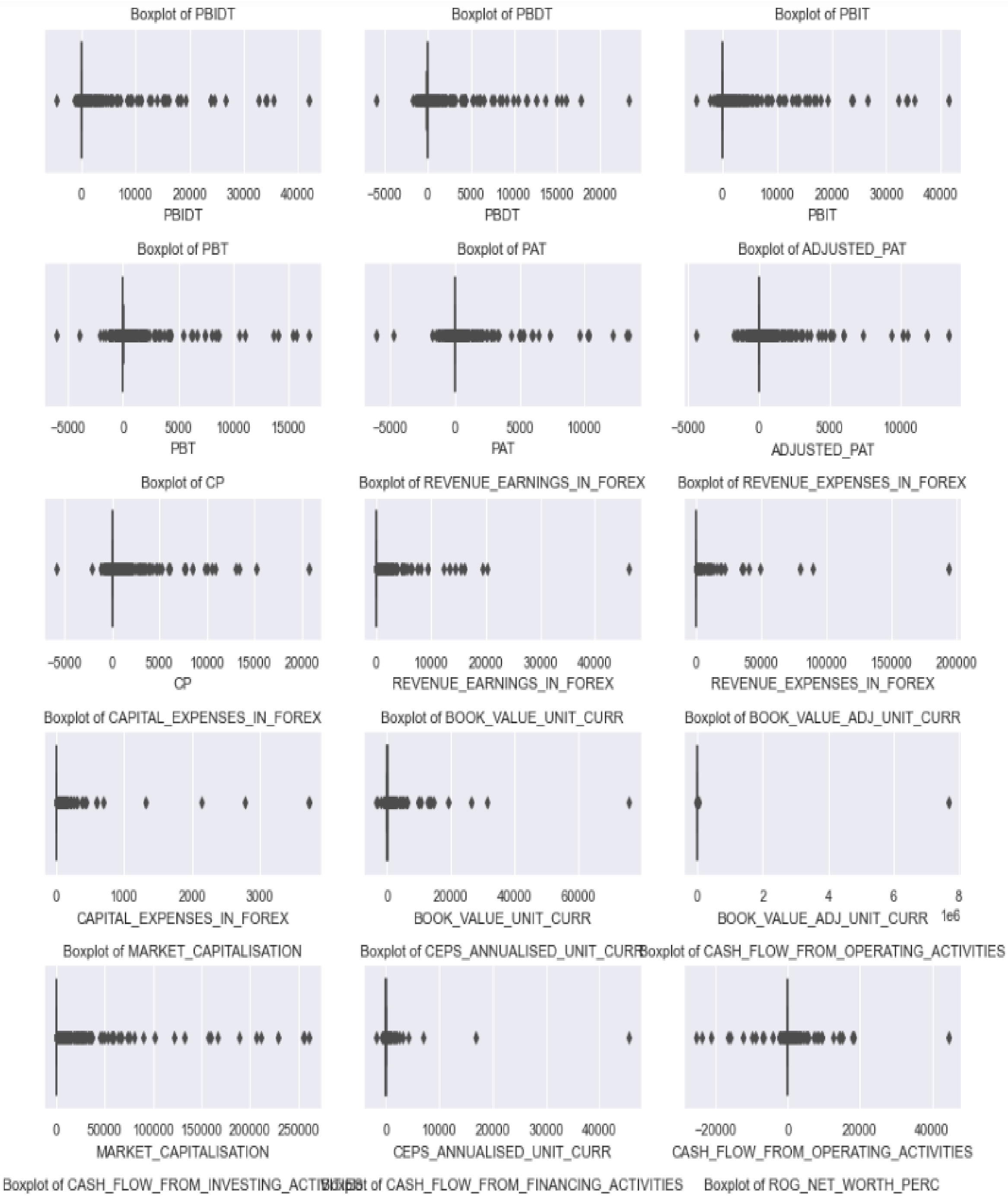
Lower Limit = Q1 – 1.5xIQR

Where Q1 / Q3 = First and Third Quartile

IQR = Inter Quartile Range = Q3 – Q1

- Out lier Visualisation


Boxplot of EQUITY_PAID_UP
0
10000
20000
30000
40000
EQUITY_PAID_UP
Boxplot of NETWORTH
0
20000
40000
60000
80000
NETWORTH
Boxplot of CAPITAL_EMPLOYED
0
200000
400000
600000
CAPITAL_EMPLOYED
Boxplot of TOTAL_DEBT
0
200000
400000
600000
TOTAL_DEBT
Boxplot of GROSS_BLOCK
0
25000
50000
75000
100000
125000
GROSS_BLOCK
Boxplot of NET_WORKING_CAPITAL
0
50000
100000
150000
200000
NET_WORKING_CAPITAL
Boxplot of CURRENT_ASSETS
0
200000
400000
600000
CURRENT_ASSETS
Boxplot of CURRENT_LIABILITIES_AND_PROVISIONS
0
20000
40000
60000
80000
CURRENT_LIABILITIES_AND_PROVISIONS
Boxplot of TOTAL_ASSETS_BY_LIABILITIES
0
50000
100000
150000
200000
250000
TOTAL_ASSETS_BY_LIABILITIES
Boxplot of GROSS_SALES
0
100000
200000
300000
400000
GROSS_SALES
Boxplot of NET_SALES
0
100000
200000
300000
400000
NET_SALES
Boxplot of OTHER_INCOME
0
2500
5000
7500
10000
12500
OTHER_INCOME
Boxplot of VALUE_OF_OUTPUT
0
100000
200000
300000
400000
VALUE_OF_OUTPUT
Boxplot of COST_OF_PRODUCTION
0
100000
200000
300000
400000
COST_OF_PRODUCTION
Boxplot of SELLING_COST
0
1000
2000
3000
4000
5000
SELLING_COST


Boxplot of PBIDT
0
10000
20000
30000
40000
PBIDT
Boxplot of PBDT
−5000
0
5000
10000
15000
20000
PBDT
Boxplot of PBIT
0
10000
20000
30000
40000
PBIT
Boxplot of PBT
−5000
0
5000
10000
15000
PBT
Boxplot of PAT
−5000
0
5000
10000
PAT
Boxplot of ADJUSTED_PAT
−5000
0
5000
10000
ADJUSTED_PAT
Boxplot of CP
−5000
0
5000
10000
15000
20000
CP
Boxplot of REVENUE_EARNINGS_IN_FOREX
0
10000
20000
30000
40000
REVENUE_EARNINGS_IN_FOREX
Boxplot of REVENUE_EXPENSES_IN_FOREX
0
50000
100000
150000
200000
REVENUE_EXPENSES_IN_FOREX
Boxplot of CAPITAL_EXPENSES_IN_FOREX
0
1000
2000
3000
CAPITAL_EXPENSES_IN_FOREX
Boxplot of BOOK_VALUE_UNIT_CURR
0
20000
40000
60000
BOOK_VALUE_UNIT_CURR
Boxplot of BOOK_VALUE_ADJ_UNIT_CURR
0
2
4
6
8
1e6
BOOK_VALUE_ADJ_UNIT_CURR
Boxplot of MARKET_CAPITALISATION
0
50000
100000
150000
200000
250000
MARKET_CAPITALISATION
Boxplot of CEPS_ANNUALISED_UNIT_CURR
0
10000
20000
30000
40000
CEPS_ANNUALISED_UNIT_CURR
Boxplot of CASH_FLOW_FROM_OPERATING_ACTIVITIES
−20000
0
20000
40000
CASH_FLOW_FROM_OPERATING_ACTIVITIES
Boxplot of CASH_FLOW_FROM_INVESTING_ACTIVITIES
Boxplot of CASH_FLOW_FROM_FINANCING_ACTIVITIES
Boxplot of ROG_NET_WORTH_PERC

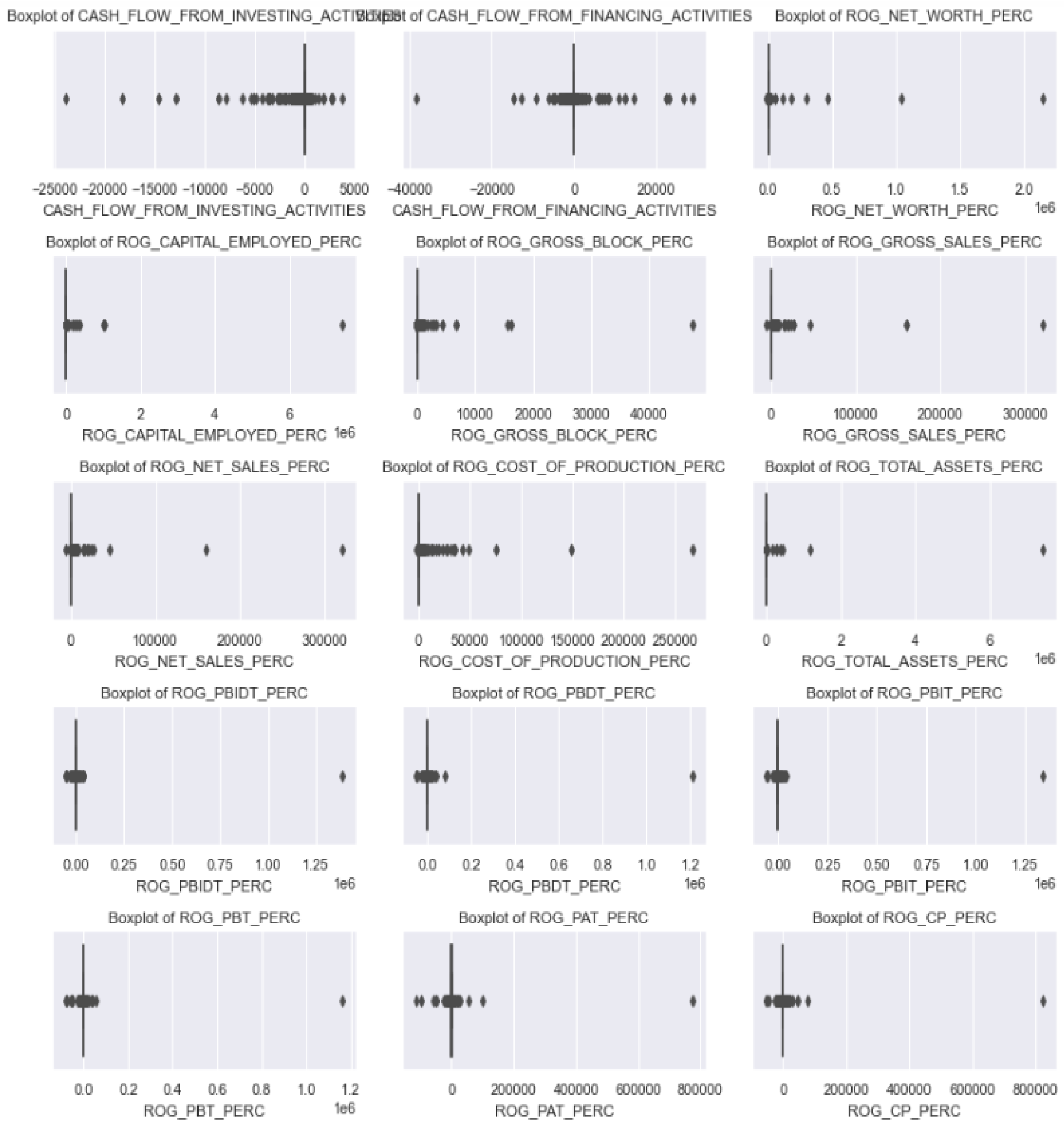
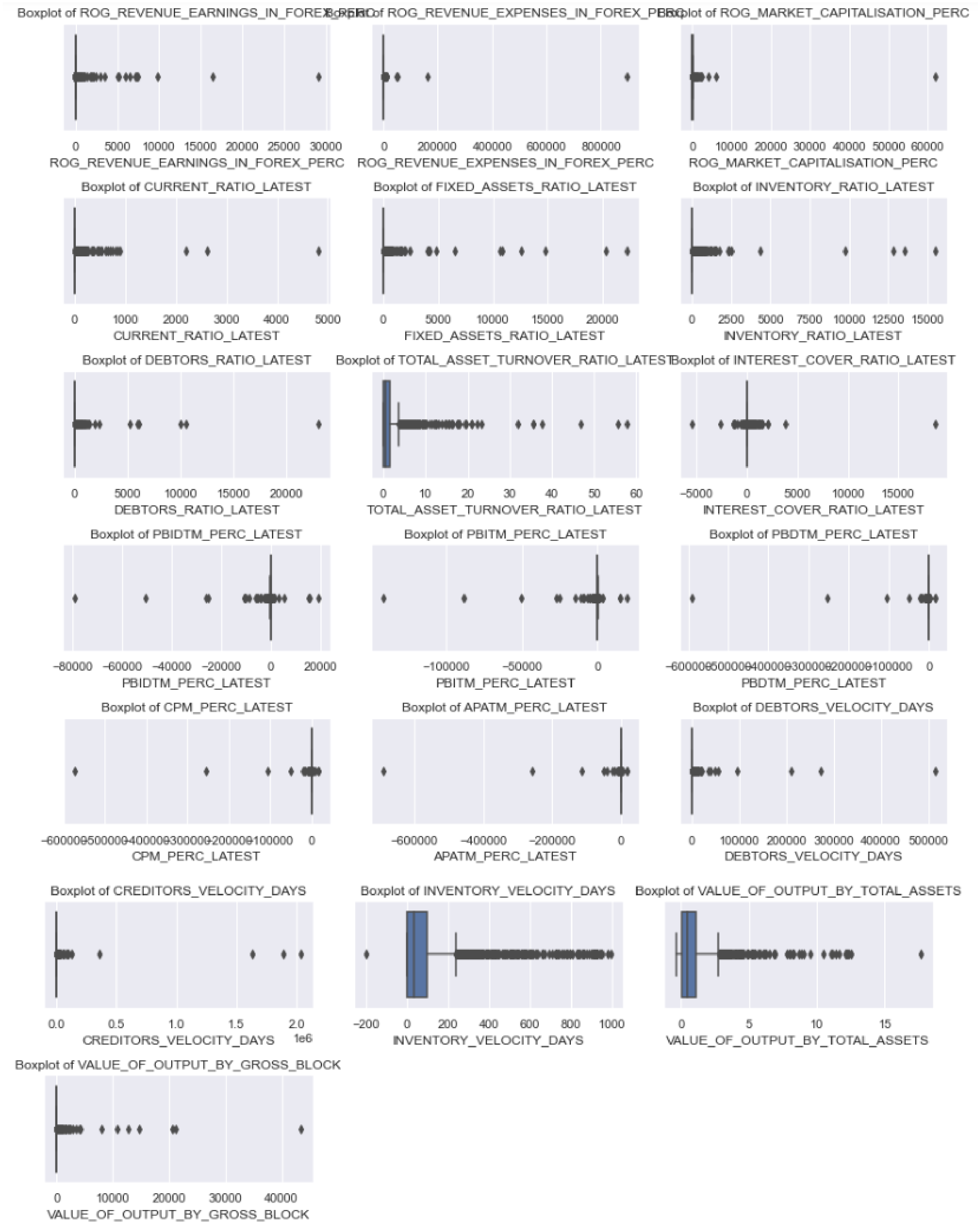
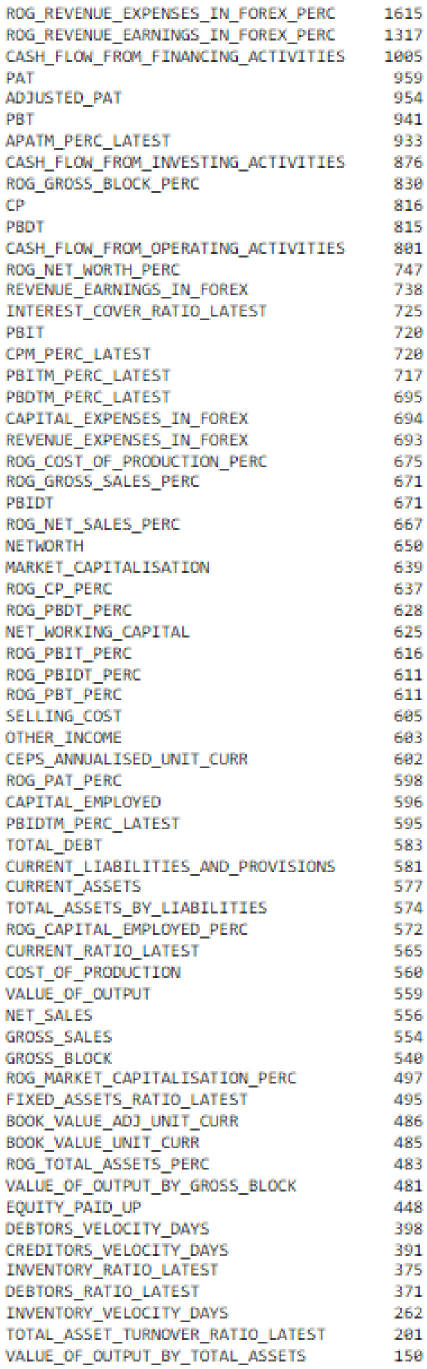


Figure 1: Visualization of Outliers in Features



- We note that there are outliers in every feature. The summary of the outliers is as under

Table 4 : Quantum of Outliers per feature



```
ROG_REVENUE_EXPENSES_IN_FOREX_PERC       1615
ROG_REVENUE_EARNINGS_IN_FOREX_PERC       1317
CASH_FLOW_FROM_FINANCING_ACTIVITIES      1005
PAT                                       959
ADJUSTED_PAT                              954
PBT                                       941
APATM_PERC_LATEST                         933
CASH_FLOW_FROM_INVESTING_ACTIVITIES       876
ROG_GROSS_BLOCK_PERC                      830
CP                                        816
PBDT                                      815
CASH_FLOW_FROM_OPERATING_ACTIVITIES       801
ROG_NET_WORTH_PERC                        747
REVENUE_EARNINGS_IN_FOREX                 738
INTEREST_COVER_RATIO_LATEST               725
PBIT                                      720
CPM_PERC_LATEST                           720
PBITM_PERC_LATEST                         717
PBDTM_PERC_LATEST                         695
CAPITAL_EXPENSES_IN_FOREX                 694
REVENUE_EXPENSES_IN_FOREX                 693
ROG_COST_OF_PRODUCTION_PERC               675
ROG_GROSS_SALES_PERC                      671
PBIDT                                     671
ROG_NET_SALES_PERC                        667
NETWORTH                                  650
MARKET_CAPITALISATION                     639
ROG_CP_PERC                               637
ROG_PBDT_PERC                             628
NET_WORKING_CAPITAL                       625
ROG_PBIT_PERC                             616
ROG_PBIDT_PERC                            611
ROG_PBT_PERC                              611
SELLING_COST                              605
OTHER_INCOME                              603
CEPS_ANNUALISED_UNIT_CURR                 602
ROG_PAT_PERC                              598
CAPITAL_EMPLOYED                          596
PBIDTM_PERC_LATEST                        595
TOTAL_DEBT                                583
CURRENT_LIABILITIES_AND_PROVISIONS        581
CURRENT_ASSETS                            577
TOTAL_ASSETS_BY_LIABILITIES               574
ROG_CAPITAL_EMPLOYED_PERC                 572
CURRENT_RATIO_LATEST                      565
COST_OF_PRODUCTION                        560
VALUE_OF_OUTPUT                           559
NET_SALES                                 556
GROSS_SALES                               554
GROSS_BLOCK                               540
ROG_MARKET_CAPITALISATION_PERC            497
FIXED_ASSETS_RATIO_LATEST                 495
BOOK_VALUE_ADJ_UNIT_CURR                  486
BOOK_VALUE_UNIT_CURR                      485
ROG_TOTAL_ASSETS_PERC                     483
VALUE_OF_OUTPUT_BY_GROSS_BLOCK            481
EQUITY_PAID_UP                            448
DEBTORS_VELOCITY_DAYS                     398
CREDITORS_VELOCITY_DAYS                   391
INVENTORY_RATIO_LATEST                    375
DEBTORS_RATIO_LATEST                      371
INVENTORY_VELOCITY_DAYS                   262
TOTAL_ASSET_TURNOVER_RATIO_LATEST         201
VALUE_OF_OUTPUT_BY_TOTAL_ASSETS           150
```

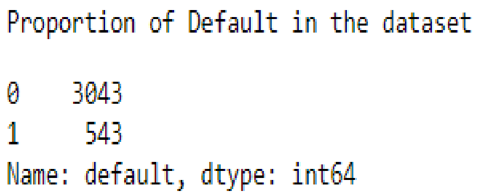
- Total number of rows are 3586 and out of that above are the outliers in every feature
- 18 % of the entire data provided is in the outlier category.
- All the outliers are replaced by null values,
- Total null values (outliers + missing values) is 18 %

3 Missing Value Treatment

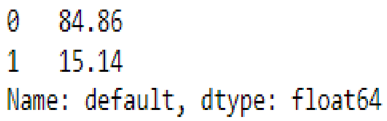
- The Outliers are converted into null values and treated together with the original missing values.
- However, before any treatment of the data , we convert the target variable – ‘NETWORTH_NEXT_YEAR’ to a binary variable under a new feature named ‘default’.
- ‘default’ is 0 if NETWORTH_NEXT_YEAR’ is >= 0
- ‘default’ is 1 if NETWORTH_NEXT_YEAR’ is < 0
- We also assess the features that have more than 30 % of records missing and
  - ➢ ROG_REVENUE_EXPENSES_IN_FOREX_PERC
  - ➢ ROG_REVENUE_EARNINGS_IN_FOREX_PERC

  are removed from the study
- Missing values are also studied across rows . Companies that have more than 10 % records missing in a row constitute a whopping 55 % of all the Company records.
- In view of the very high % of data that may need to be discarded which has a null value , we decided to treat the missing data and retain it.
- Before the data is treated, we split the data into Train Data and Test Data in the ratio of 66 % and 33 %, ensuring that the proportion of ‘default’ in both the test data and the train data is same as in the original data
- The data is scaled first by the Standard Scaler
- The missing data is then imputed by the KNN (k-nearest neighbor) tool.

4 Transform Target variable into 0 and 1



```
Proportion of Default in the dataset

0    3043
1     543
Name: default, dtype: int64
```



```
0   84.86
1   15.14
Name: default, dtype: float64
```

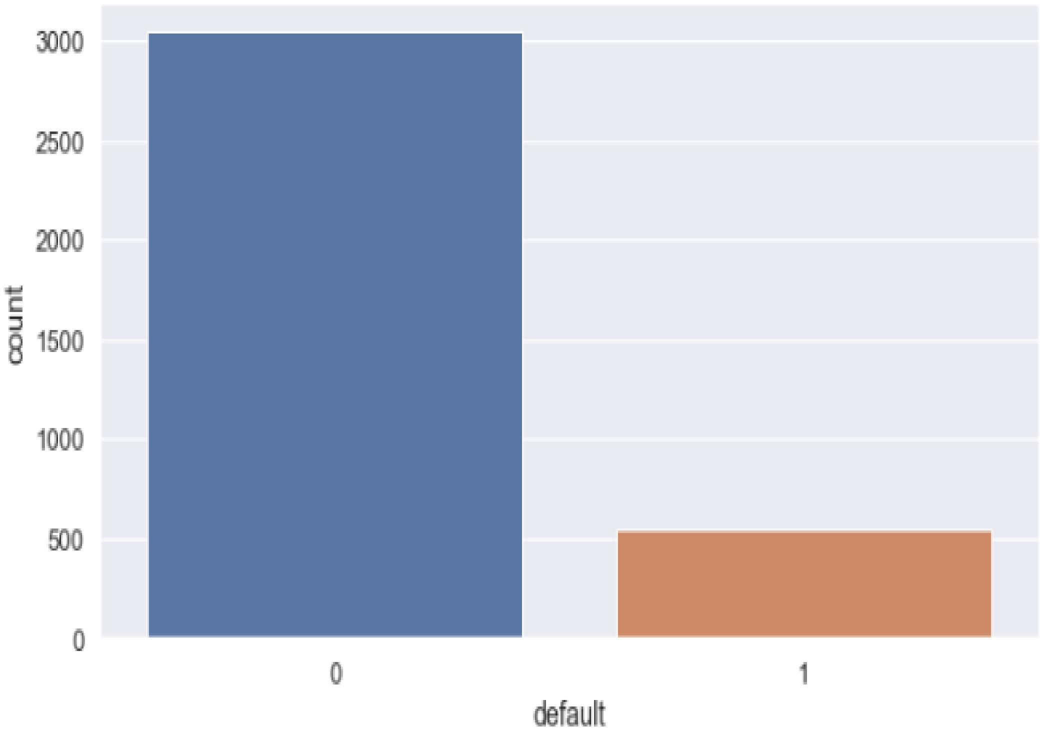


Figure 2: Proportion of Target variable

5 Univariate (4 marks) & Bivariate (6marks) analysis with proper interpretation.

- Univariate Analysis – the count plot visualization of features

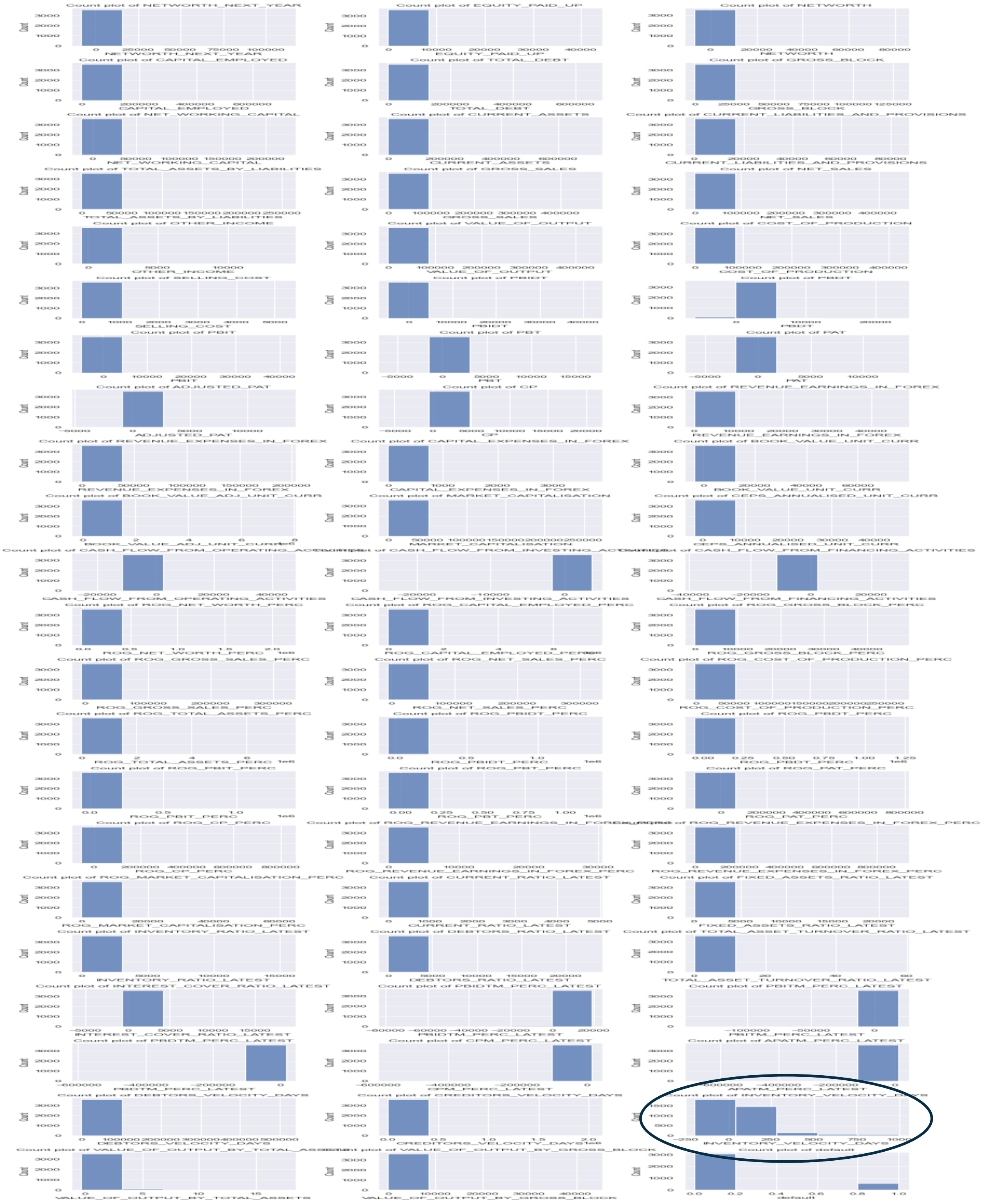


Figure 3: Count Plot of Features

- All the Companies are grouped together in a bin in most of the features other than INVENTORY_VELOCITY_DAYS, where they are segregated in multiple bins .

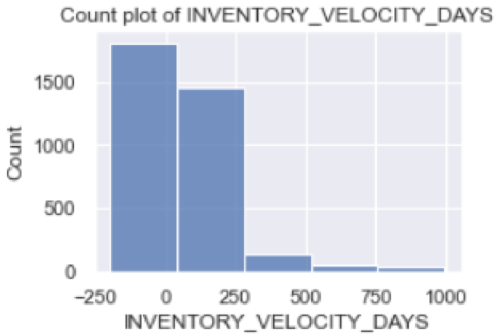
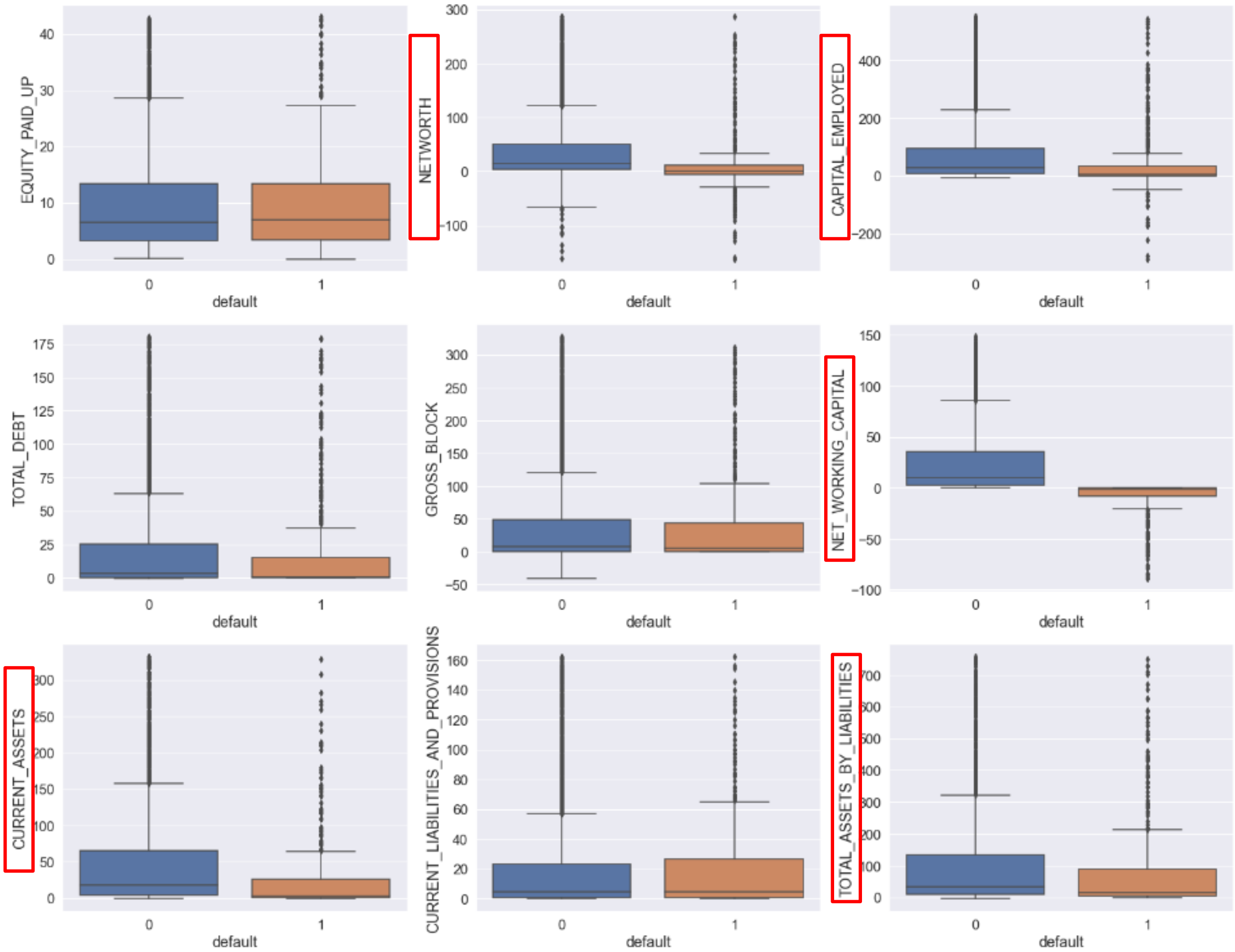
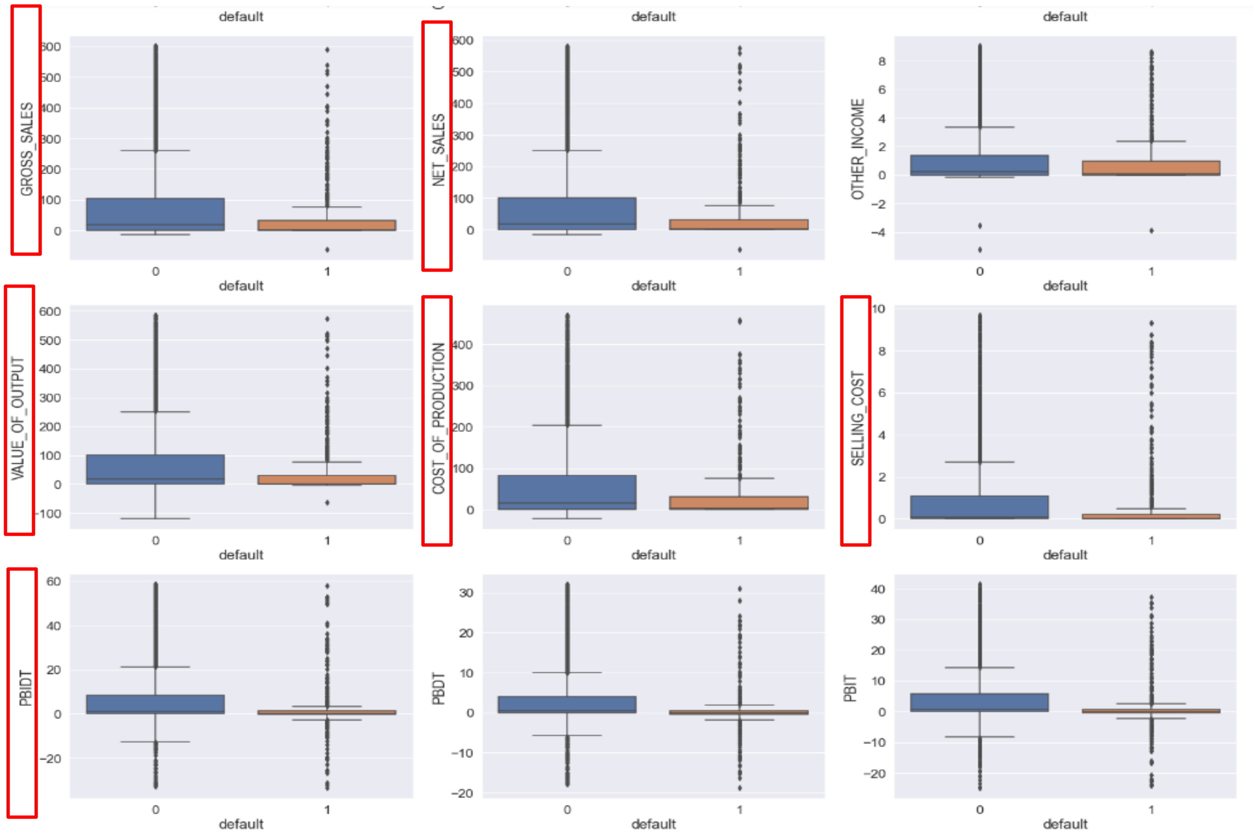


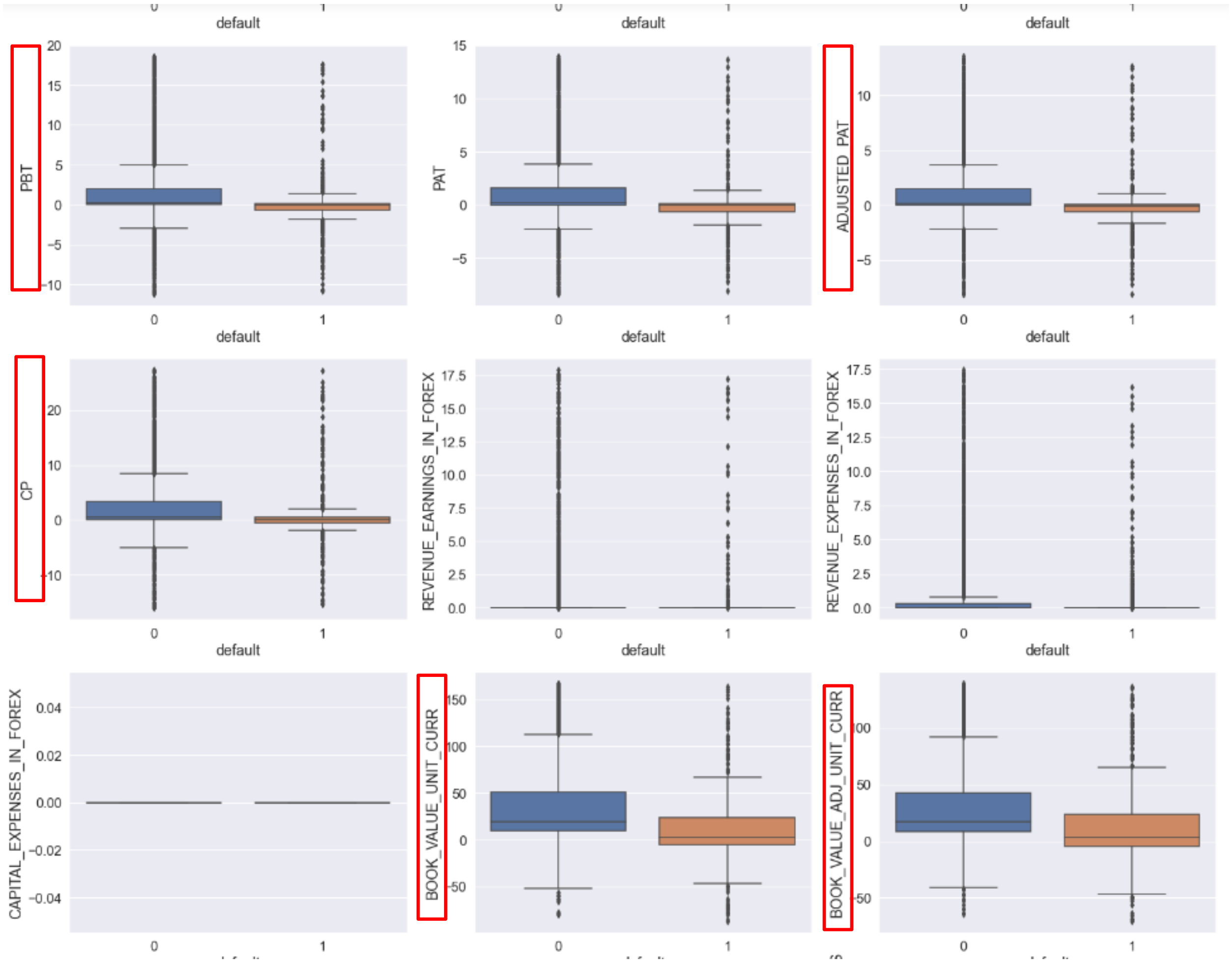
Figure 4: Count Plot of Inventory Velocity days

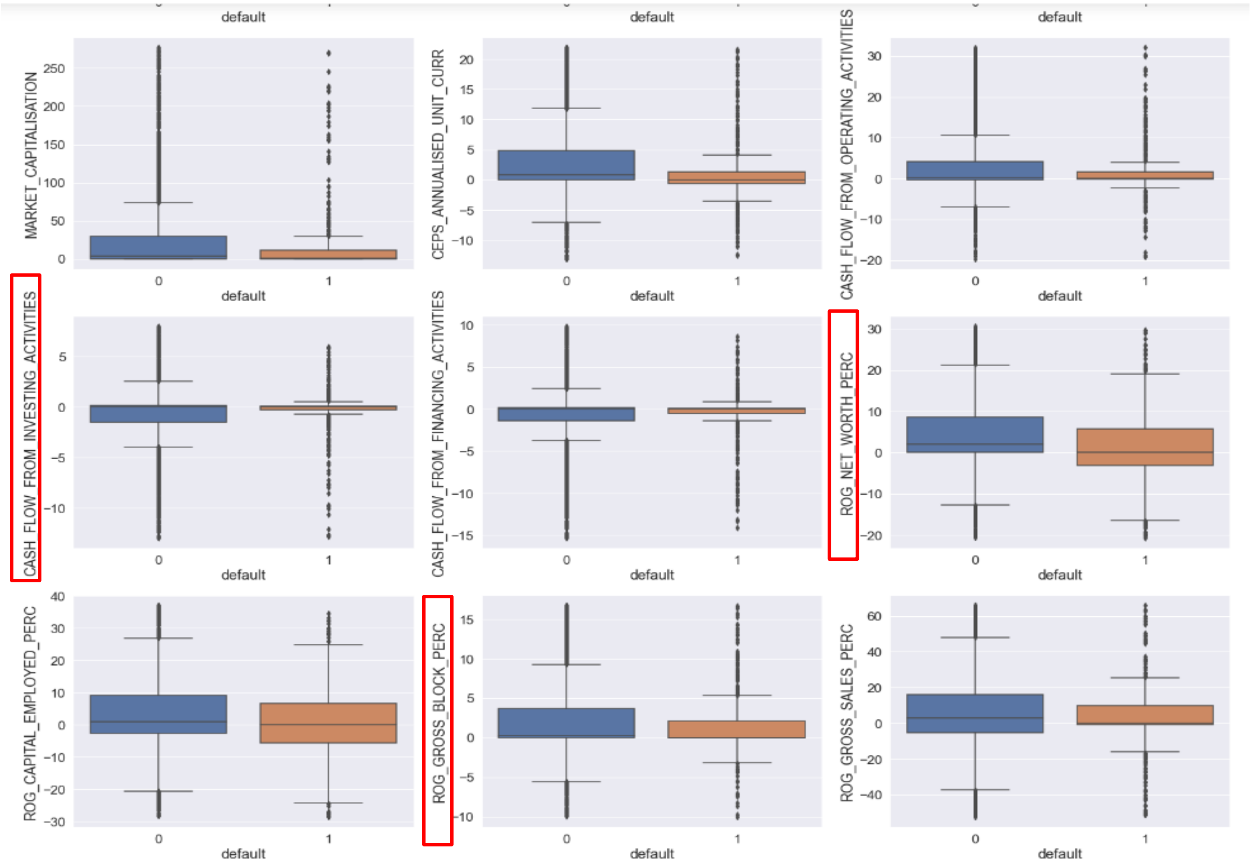
Most Cos convert their inventory into sales in less than 250 days. There are Cos which take up to 1000 days

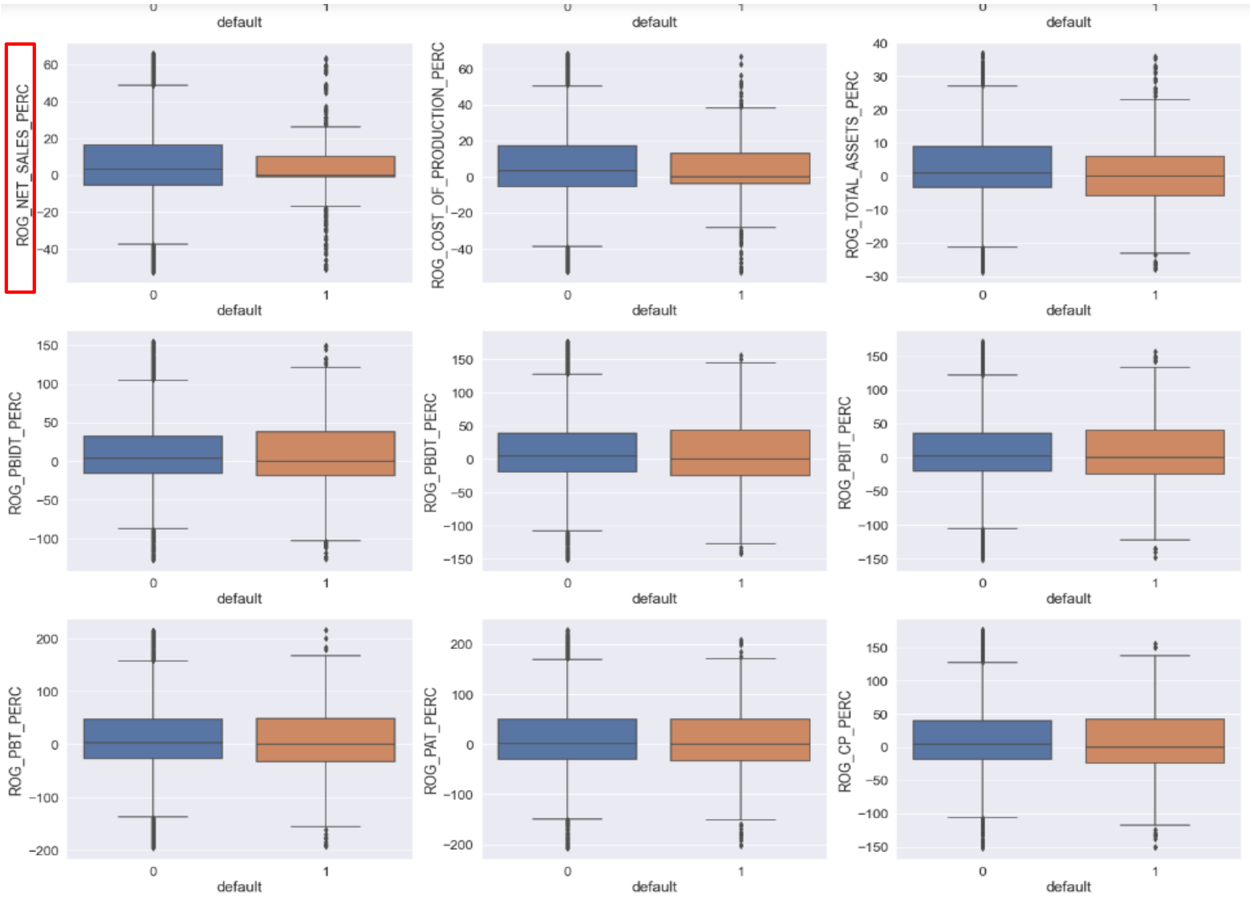
- We study the interaction of all the features with ‘ default ‘ vide boxplots




default
GROSS_SALES
600
500
400
300
200
100
0
0
1
default
NET_SALES
600
500
400
300
200
100
0
0
1
default
OTHER_INCOME
8
6
4
2
0
−2
−4
0
1
default
VALUE_OF_OUTPUT
600
500
400
300
200
100
0
-100
0
1
default
COST_OF_PRODUCTION
400
300
200
100
0
0
1
default
SELLING_COST
10
8
6
4
2
0
0
1
default
PBIDT
60
40
20
0
-20
0
1
default
PBDT
30
20
10
0
-10
-20
0
1
default
PBIT
40
30
20
10
0
-10
-20
0
1
default


default
PBT
20
15
10
5
0
-5
-10
0
1
default
PAT
15
10
5
0
-5
0
1
default
ADJUSTED_PAT
10
5
0
-5
0
1
default
CP
20
10
0
-10
0
1
default
REVENUE_EARNINGS_IN_FOREX
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0
1
default
REVENUE_EXPENSES_IN_FOREX
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0
1
default
CAPITAL_EXPENSES_IN_FOREX
0.04
0.02
0.00
−0.02
−0.04
0
1
default
BOOK_VALUE_UNIT_CURR
150
100
50
0
-50
0
1
default
BOOK_VALUE_ADJ_UNIT_CURR
100
50
0
-50
0
1


default
default
default
MARKET_CAPITALISATION
250
200
150
100
50
0
0
1
default
CEPS_ANNUALISED_UNIT_CURR
20
15
10
5
0
−5
−10
0
1
default
CASH_FLOW_FROM_OPERATING_ACTIVITIES
30
20
10
0
−10
−20
0
1
default
CASH_FLOW_FROM_INVESTING_ACTIVITIES
5
0
−5
−10
0
1
default
CASH_FLOW_FROM_FINANCING_ACTIVITIES
10
5
0
−5
−10
−15
0
1
default
ROG_NET_WORTH_PERC
30
20
10
0
−10
−20
0
1
default
ROG_CAPITAL_EMPLOYED_PERC
40
30
20
10
0
−10
−20
−30
0
1
default
ROG_GROSS_BLOCK_PERC
15
10
5
0
−5
−10
0
1
default
ROG_GROSS_SALES_PERC
60
40
20
0
−20
−40
0
1
default


0
1
default
0
1
default
0
1
default
ROG_NET_SALES_PERC
60
40
20
0
−20
−40
0
1
default
ROG_COST_OF_PRODUCTION_PERC
60
40
20
0
−20
−40
0
1
default
ROG_TOTAL_ASSETS_PERC
40
30
20
10
0
−10
−20
−30
0
1
default
ROG_PBIDT_PERC
150
100
50
0
−50
−100
0
1
default
ROG_PBDT_PERC
150
100
50
0
−50
−100
−150
0
1
default
ROG_PBIT_PERC
150
100
50
0
−50
−100
−150
0
1
default
ROG_PBT_PERC
200
100
0
−100
−200
0
1
default
ROG_PAT_PERC
200
100
0
−100
−200
0
1
default
ROG_CP_PERC
150
100
50
0
−50
−100
−150
0
1
default

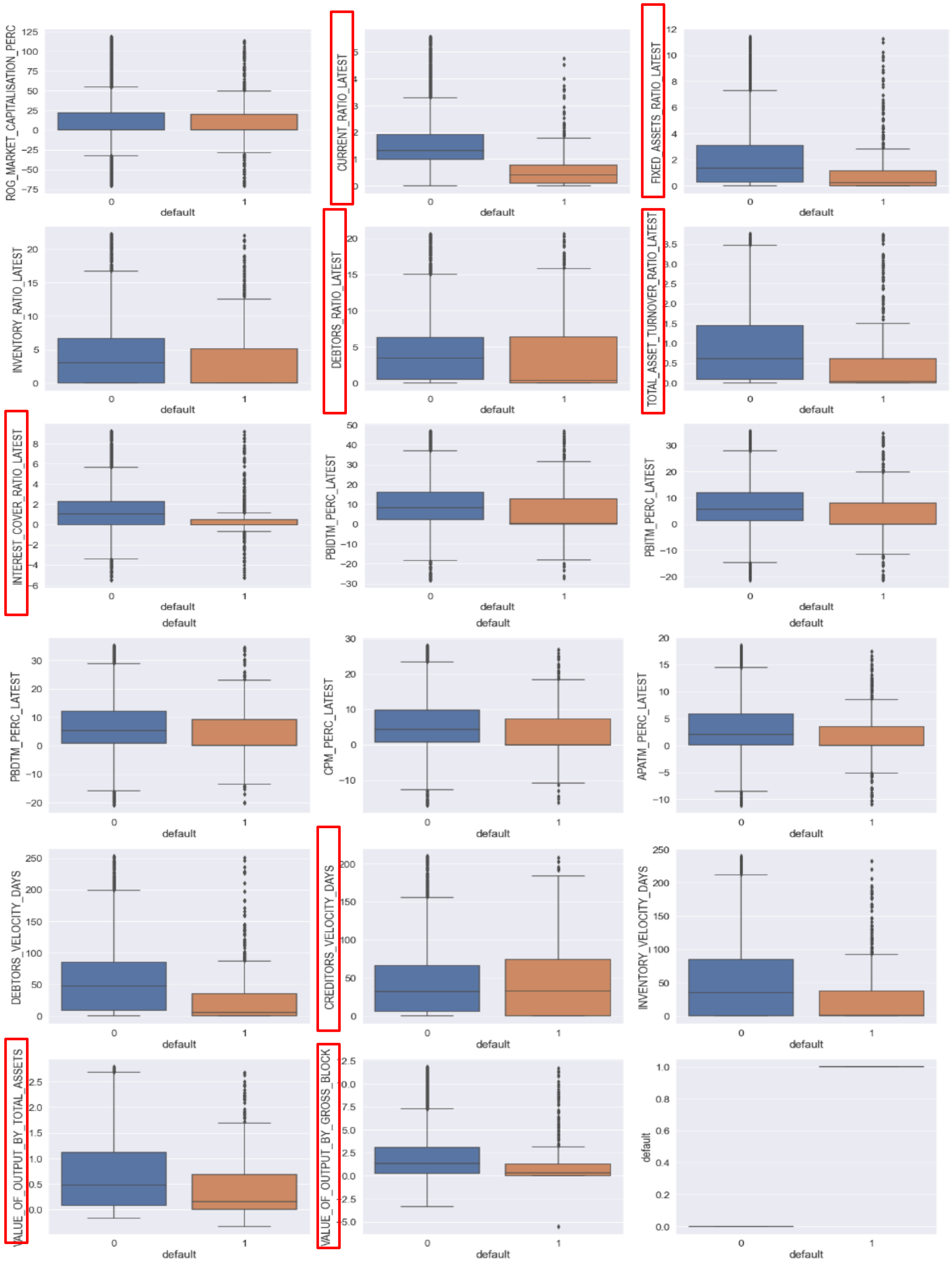


Figure 5 : Box Plot of Features vs Default

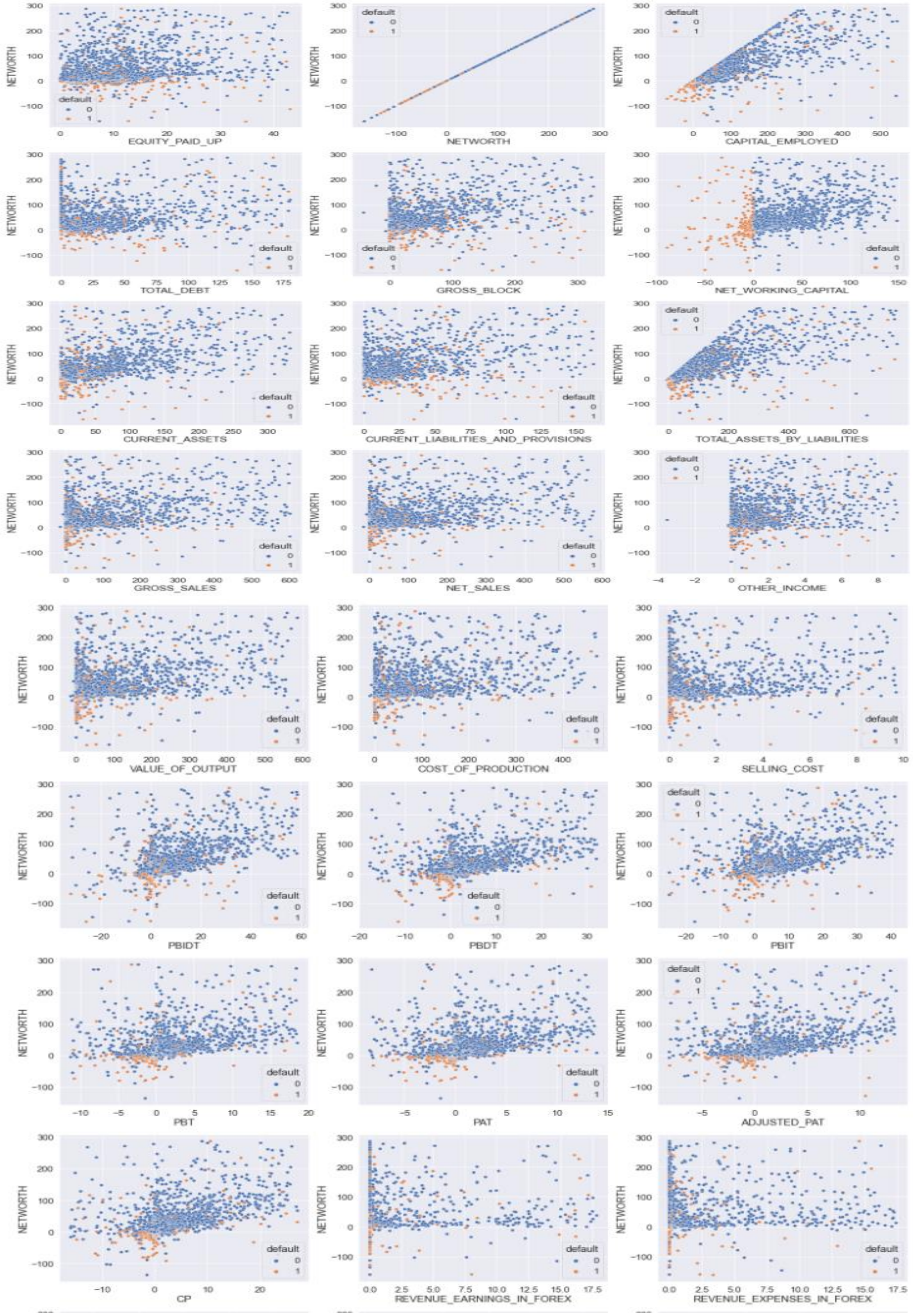
- The difference in the values of the features between the Companies which default and those who do not is very obvious from the above figure. The important features are highlighted in the red box in the figure above. These features clearly indicate the difference between healthy and unhealthy Companies.
    - ➢ Defaulting Companies have relatively lower values in

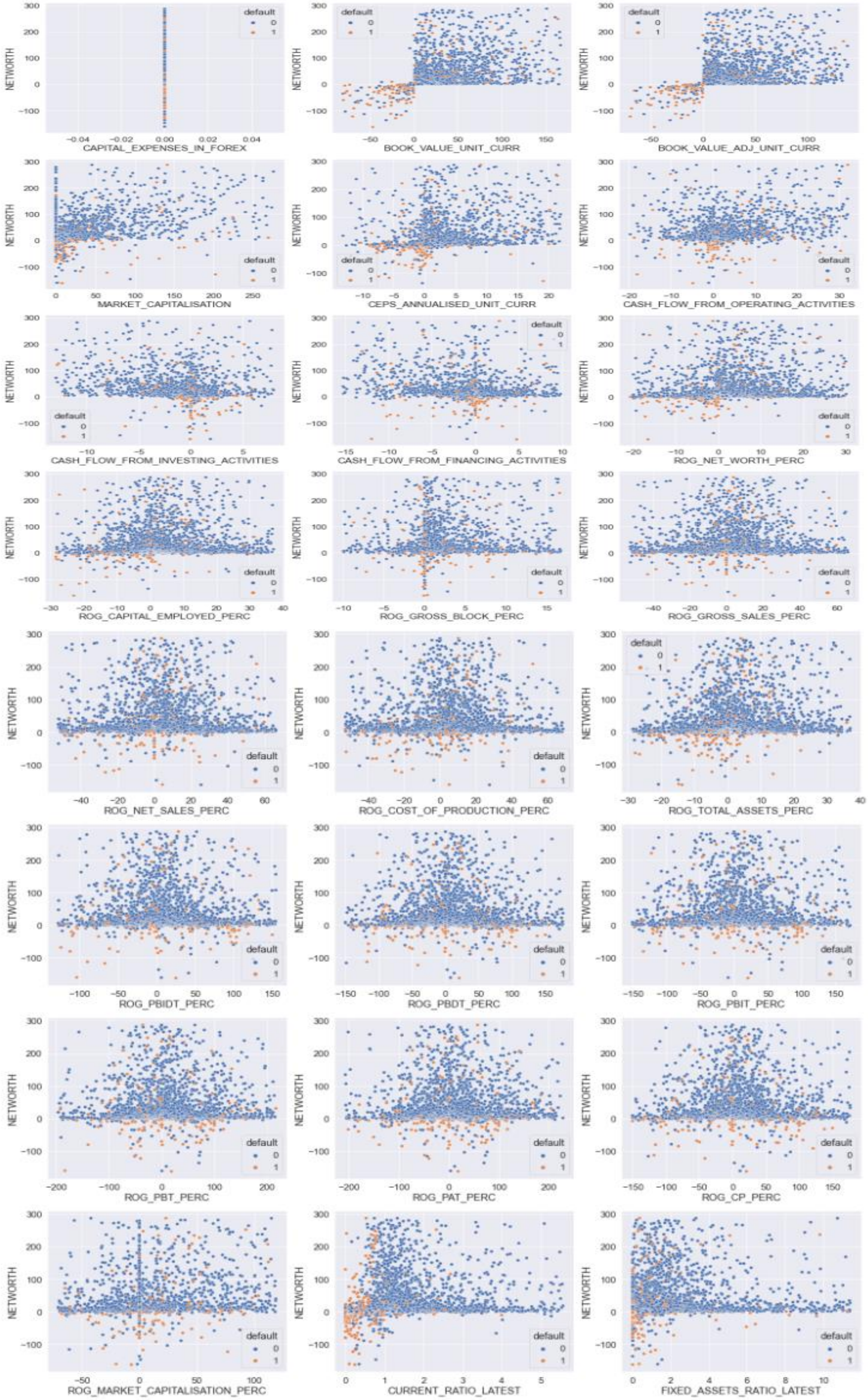
1. Net Worth
2. Capital Employed
3. Net Working Capital
4. Current Assets
5. Total Assets by Liabilities
6. Gross Sales
7. Net Sales
8. Selling Cost
9. Value of Output
10. Cost of production
11. PBIDT
12. PBT
13. Adjusted PAT
14. CP
15. Book Value of Adjusted Unit Currency

Etc. as features highlighted in red box. We note that for many features the IQR spread for defaulting Cos is very less than the Cos that do not default – such as

Interest Cover – Ratio Latest

- Scatter Plot of features vs default studied with Net worth of current year. Remember that Net worth for next year if negative has a default value of 1.




NETWORTH
default
0
1
CAPITAL_EXPENSES_IN_FOREX
BOOK_VALUE_UNIT_CURR
BOOK_VALUE_ADJ_UNIT_CURR
MARKET_CAPITALISATION
CEPS_ANNUALISED_UNIT_CURR
CASH_FLOW_FROM_OPERATING_ACTIVITIES
CASH_FLOW_FROM_INVESTING_ACTIVITIES
CASH_FLOW_FROM_FINANCING_ACTIVITIES
ROG_NET_WORTH_PERC
ROG_CAPITAL_EMPLOYED_PERC
ROG_GROSS_BLOCK_PERC
ROG_GROSS_SALES_PERC
ROG_NET_SALES_PERC
ROG_COST_OF_PRODUCTION_PERC
ROG_TOTAL_ASSETS_PERC
ROG_PBIDT_PERC
ROG_PBDT_PERC
ROG_PBIT_PERC
ROG_PBT_PERC
ROG_PAT_PERC
ROG_CP_PERC
ROG_MARKET_CAPITALISATION_PERC
CURRENT_RATIO_LATEST
FIXED_ASSETS_RATIO_LATEST

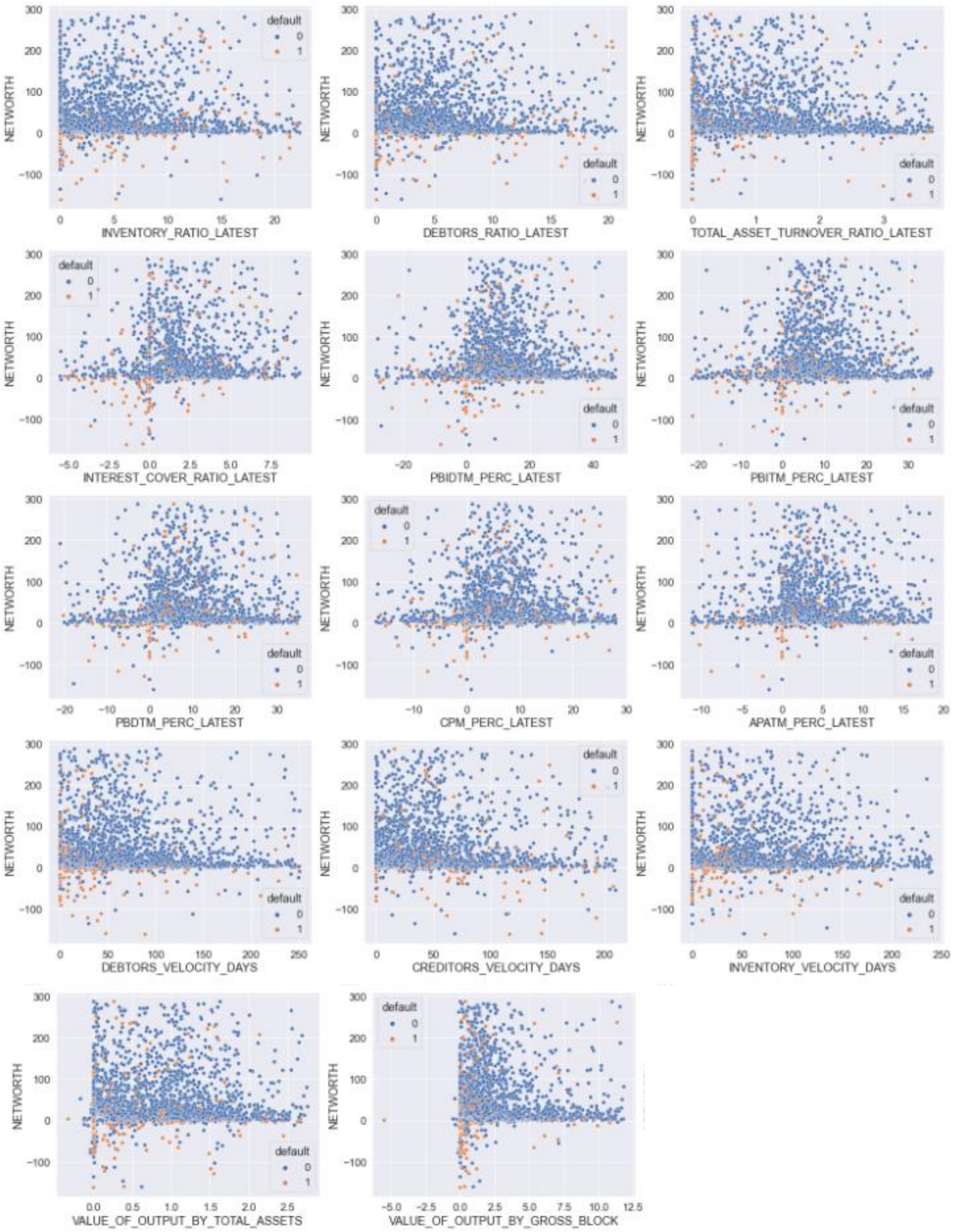


Figure 6 : Scatter Plot of Features vs Net worth of Current Year vs Default

- The yellow dots in the above plot show the defaulting Companies performance in the different features vs their current Net worth.

- We clearly see that the yellow dots are clustered near the left bottom in most performance criteria.

- Correlations between the features

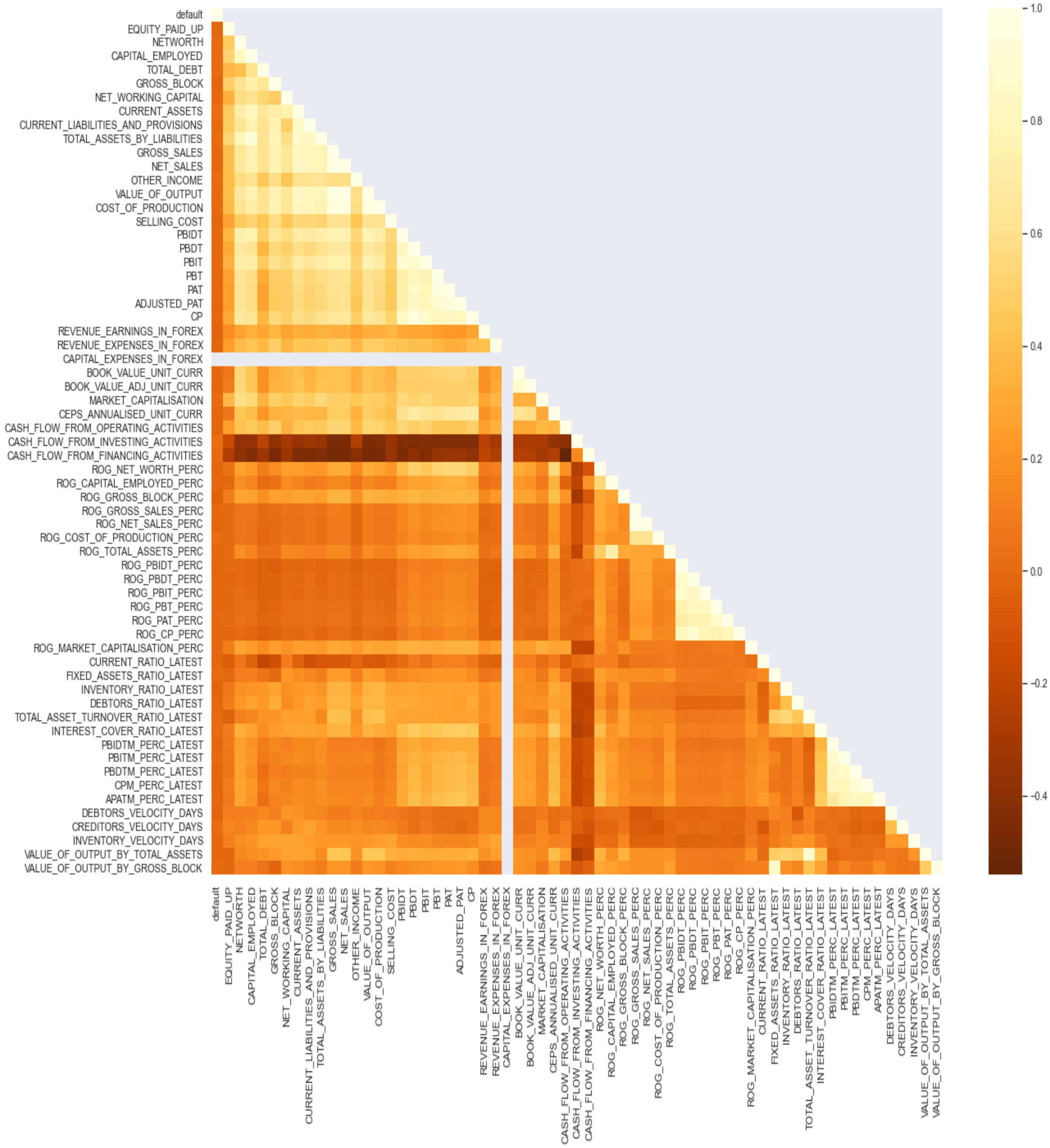
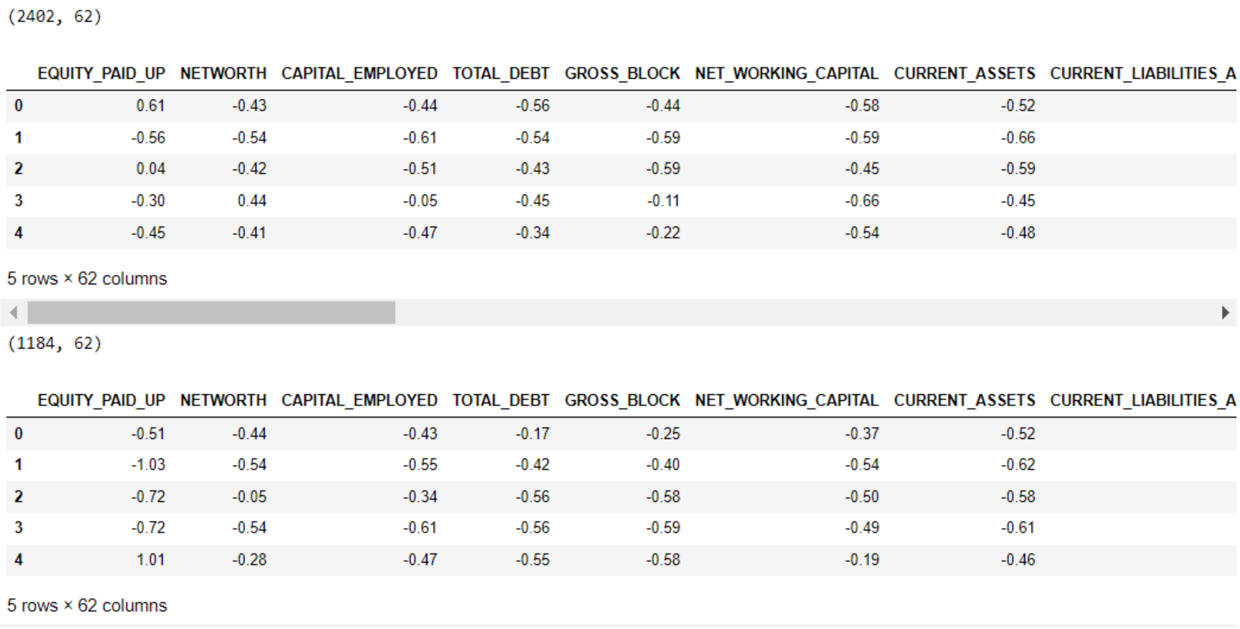


Figure 7: Correlation Heat Map

- We see that there is a lot of positive correlation between features. Multicollinearity will become an issue

6 Train Test Split

Table 5: Train and Test Data



```
(2402, 62)
```

	EQUITY_PAID_UP	NETWORTH	CAPITAL_EMPLOYED	TOTAL_DEBT	GROSS_BLOCK	NET_WORKING_CAPITAL	CURRENT_ASSETS	CURRENT_LIABILITIES_A
0	0.61	-0.43	-0.44	-0.56	-0.44	-0.58	-0.52	
1	-0.56	-0.54	-0.61	-0.54	-0.59	-0.59	-0.66	
2	0.04	-0.42	-0.51	-0.43	-0.59	-0.45	-0.59	
3	-0.30	0.44	-0.05	-0.45	-0.11	-0.66	-0.45	
4	-0.45	-0.41	-0.47	-0.34	-0.22	-0.54	-0.48	

5 rows × 62 columns

```
(1184, 62)
```

	EQUITY_PAID_UP	NETWORTH	CAPITAL_EMPLOYED	TOTAL_DEBT	GROSS_BLOCK	NET_WORKING_CAPITAL	CURRENT_ASSETS	CURRENT_LIABILITIES_A
0	-0.51	-0.44	-0.43	-0.17	-0.25	-0.37	-0.52	
1	-1.03	-0.54	-0.55	-0.42	-0.40	-0.54	-0.62	
2	-0.72	-0.05	-0.34	-0.56	-0.58	-0.50	-0.58	
3	-0.72	-0.54	-0.61	-0.56	-0.59	-0.49	-0.61	
4	1.01	-0.28	-0.47	-0.55	-0.58	-0.19	-0.46	

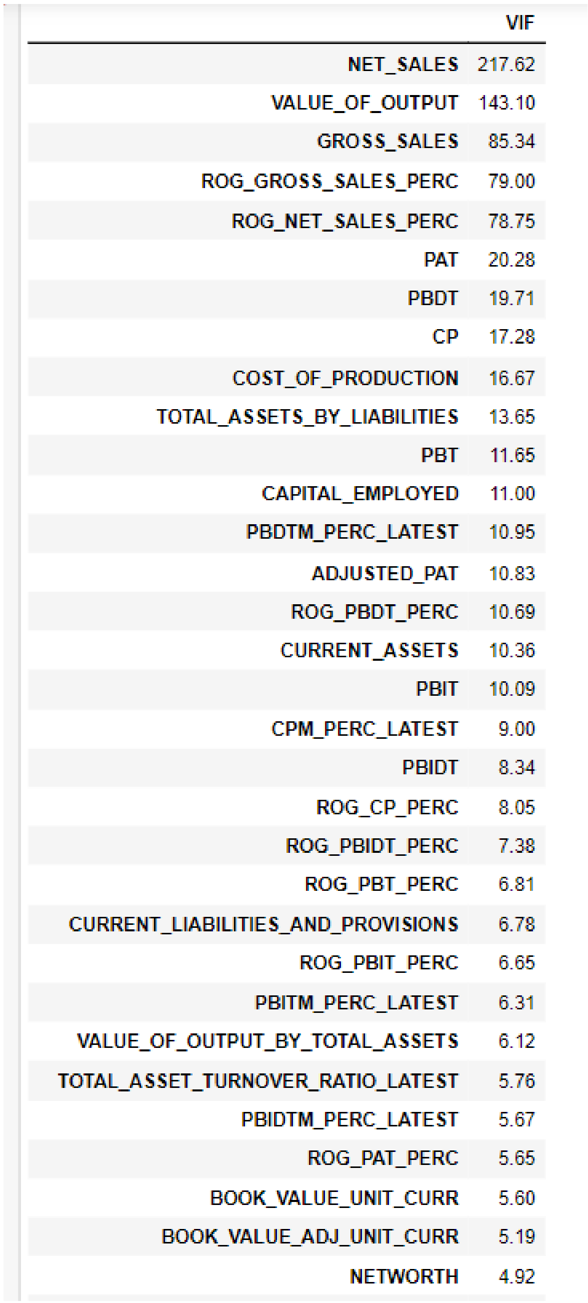
5 rows × 62 columns

- The data is split into Train and Test in 67:33 ratio.
- We split the data ensuring that the proportion of the 'defaults' are same in Train and Test data.

7 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

- We use the Variance Inflation Factor (VIF) method to find the features that contribute to multicollinearity.

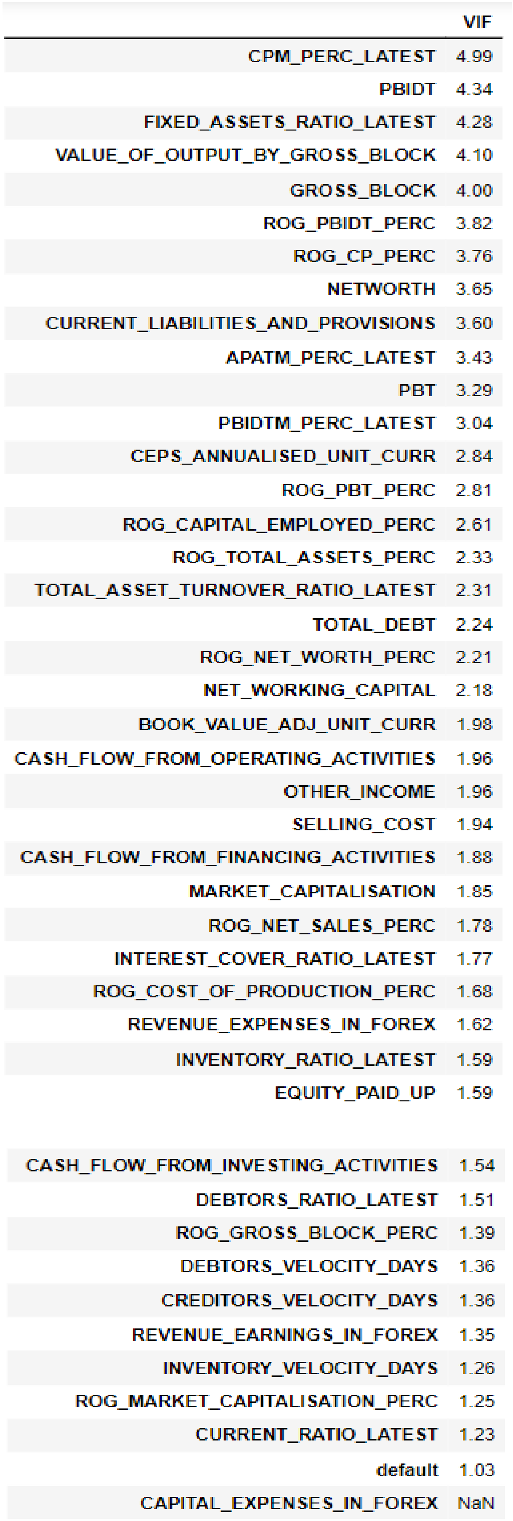
Table 6 : Features with High VIF values



	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

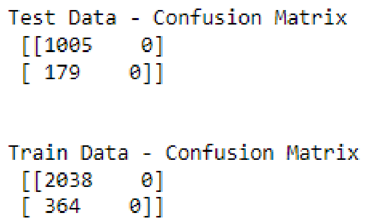
- All features that have a VIF factor more than 5 are eliminated from the model

Table 7: Features with VIF value less than 5



	VIF
CPM_PERC_LATEST	4.99
PBIDT	4.34
FIXED_ASSETS_RATIO_LATEST	4.28
VALUE_OF_OUTPUT_BY_GROSS_BLOCK	4.10
GROSS_BLOCK	4.00
ROG_PBIDT_PERC	3.82
ROG_CP_PERC	3.76
NETWORTH	3.65
CURRENT_LIABILITIES_AND_PROVISIONS	3.60
APATM_PERC_LATEST	3.43
PBT	3.29
PBIDTM_PERC_LATEST	3.04
CEPS_ANNUALISED_UNIT_CURR	2.84
ROG_PBT_PERC	2.81
ROG_CAPITAL_EMPLOYED_PERC	2.61
ROG_TOTAL_ASSETS_PERC	2.33
TOTAL_ASSET_TURNOVER_RATIO_LATEST	2.31
TOTAL_DEBT	2.24
ROG_NET_WORTH_PERC	2.21
NET_WORKING_CAPITAL	2.18
BOOK_VALUE_ADJ_UNIT_CURR	1.98
CASH_FLOW_FROM_OPERATING_ACTIVITIES	1.96
OTHER_INCOME	1.96
SELLING_COST	1.94
CASH_FLOW_FROM_FINANCING_ACTIVITIES	1.88
MARKET_CAPITALISATION	1.85
ROG_NET_SALES_PERC	1.78
INTEREST_COVER_RATIO_LATEST	1.77
ROG_COST_OF_PRODUCTION_PERC	1.68
REVENUE_EXPENSES_IN_FOREX	1.62
INVENTORY_RATIO_LATEST	1.59
EQUITY_PAID_UP	1.59
CASH_FLOW_FROM_INVESTING_ACTIVITIES	1.54
DEBTORS_RATIO_LATEST	1.51
ROG_GROSS_BLOCK_PERC	1.39
DEBTORS_VELOCITY_DAYS	1.36
CREDITORS_VELOCITY_DAYS	1.36
REVENUE_EARNINGS_IN_FOREX	1.35
INVENTORY_VELOCITY_DAYS	1.26
ROG_MARKET_CAPITALISATION_PERC	1.25
CURRENT_RATIO_LATEST	1.23
default	1.03
CAPITAL_EXPENSES_IN_FOREX	NaN

- **<u>Model 1 - Statsmodel</u>**
  - Logistic regression model using Statsmodel was not happening as it kept showing ‘Singular Matrix ‘ error.
- **<u>Model 2 – Sklearn</u>**
  - We build the model with Logistic regression from the Sklearn Library with the results as under

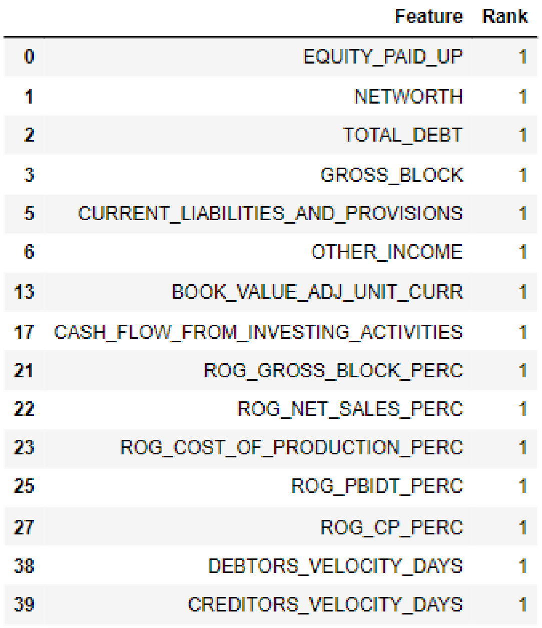


```
Test Data - Confusion Matrix
 [[1005    0]
 [ 179    0]]


Train Data - Confusion Matrix
 [[2038    0]
 [ 364    0]]
```

  - The results are very poor as the model fails to predict even a single default either in Train or Test data. As we see it , the model is basically predicting 100 % - no default.

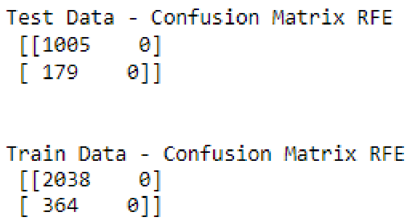
- **<u>Model 3</u>** - using RFE method (recursive feature elimination) to select features (15 features)
  - 15 Features selected by the RFE method

Table 8: RFE selected features



	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

- The Confusion matrix for the RFE model is as under

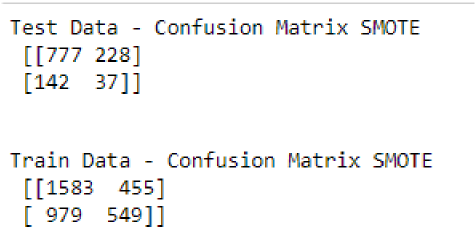


```
Test Data - Confusion Matrix RFE
 [[1005    0]
 [ 179    0]]


Train Data - Confusion Matrix RFE
 [[2038    0]
 [ 364    0]]
```

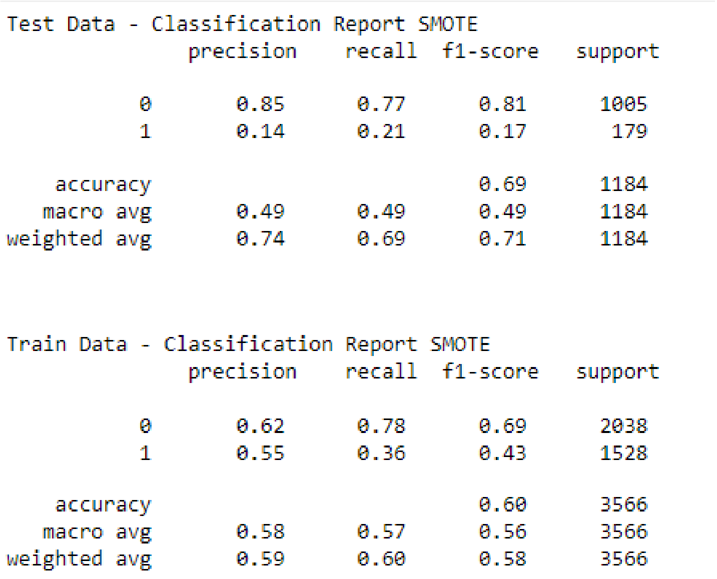
    - These results are exactly the same as the earlier model and therefore we discard the RFE model also.

- **Model 4** - Balance the data by SMOTE (sampling strategy 75:25)and using the features identified by the VIF model (vif values less than 5)

Table 9: Confusion Matrix using SMOTE 75:25



```
Test Data - Confusion Matrix SMOTE
 [[777 228]
 [142  37]]


Train Data - Confusion Matrix SMOTE
 [[1583  455]
 [ 979  549]]
```

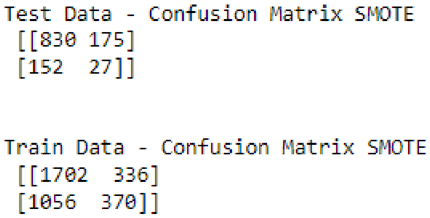


```
Test Data - Classification Report SMOTE
              precision    recall  f1-score   support

           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.

- **<u>Model 5</u>** – SMOTE (65 : 35)

Table 10 : Confusion Matrix - SMOTE (65 : 35)



```
Test Data - Confusion Matrix SMOTE
 [[830 175]
 [152  27]]


Train Data - Confusion Matrix SMOTE
 [[1702  336]
 [1056  370]]
```

    - The results are even poorer than the earlier SMOTE model and hence is discarded.

- **Changing the cut off probability to 0.7 for a predict of 1 (default) and also to 0.1 for a predict of 1 (default) does not improve the performance of the model**

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

- 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.