FINANCE & RISK ANALYTICS
MILESTONE - 1
BUSINESS REPORT

RHEA.S.M

PGPDSBA Online Sep_B 2021

Table of Contents

List of Figures

List of Tables

1. Business Problem

1.1. Objective

- The objective the problem is to build a logistic regression model, to predict which company will default if they are unable to keep up with their debt obligations on the basis of the given information.
- We additionally also have to validate the model on Test dataset and state interpretations from the model.

1.2. Descriptive and Exploratory Data Analysis

Background: 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 Networth of the company in the following year (2016) is provided which can be used to drive the labeled field.

Explanation of data fields available in Data Dictionary, 'Credit Default Data Dictionary.xlsx'

Data Dictionary:

Table-1: Data Dictionary

Field Name	Description	New Field Name
Co_Code	Company Code	Co_Code
Co_Name	Company Name	Co_Name
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
Equity Paid Up	Amount that has been received by the company through the issue of shares to the shareholders	Equity_Paid_Up
Networth	Value of a company as on 2015 - Current Year	Networth
Capital Employed	Total amount of capital used for the acquisition of profits by a company	Capital_Employed
Total Debt	The sum of money borrowed by the company and is due to be paid	Total_Debt
Gross Block	Total value of all of the assets that a company owns	Gross_Block
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
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
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
Total Assets/Liabilities	Ratio of total assets to liabailities of the company	Total_Assets_to_Liab
Gross Sales	The grand total of sale transactions within the accounting period	Gross_Sales

Net Sales	Gross sales minus returns, allowances, and discounts	Net_Sales
Other Income	Income realized from non-business activities (e.g. sale of long term asset)	Other_Income
Value Of Output	Product of physical output of goods and services produced by company and its market price	Value_Of_Output
Cost of Production	Costs incurred by a business from manufacturing a product or providing a service	Cost_of_Prod
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
PBIDT	Profit Before Interest, Depreciation & Taxes	PBIDT
PBDT	Profit Before Depreciation and Tax	PBDT
PBIT	Profit before interest and taxes	PBIT
PBT	Profit before tax	PBT
PAT	Profit After Tax	PAT
Adjusted PAT	Adjusted profit is the best estimate of the true profit	Adjusted_PAT
CP	Commercial paper , a short-term debt instrument to meet short-term liabilities.	CP
Revenue earnings in forex	Revenue earned in foreign currency	Rev_earn_in_forex
Revenue expenses in forex	Expenses due to foreign currency transactions	Rev_exp_in_forex
Capital expenses in forex	Long term investment in forex	Capital_exp_in_forex
Book Value (Unit Curr)	Net asset value	Book_Value_Unit_Curr
Book Value (Adj.) (Unit Curr)	Book value adjusted to reflect asset's true fair market value	Book_Value_Adj_Unit_Curr
Market Capitalisation	Product of the total number of a company's outstanding shares and the current market price of one share	Market_Capitalisation
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
Cash Flow From Operating Activities	Use of cash from ongoing regular business activities	Cash_Flow_From_Opr
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
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
ROG-Net Worth (%)	Rate of Growth - Networth	ROG_Net_Worth_perc
ROG-Capital Employed (%)	Rate of Growth - Capital Employed	ROG_Capital_Employed_perc
ROG-Gross Block (%)	Rate of Growth - Gross Block	ROG_Gross_Block_perc
ROG-Gross Sales (%)	Rate of Growth - Gross Sales	ROG_Gross_Sales_perc
ROG-Net Sales (%)	Rate of Growth - Net Sales	ROG_Net_Sales_perc
ROG-Cost of Production (%)	Rate of Growth - Cost of Production	ROG_Cost_of_Prod_perc
ROG-Total Assets (%)	Rate of Growth - Total Assets	ROG_Total_Assets_perc
ROG-PBIDT (%)	Rate of Growth- PBIDT	ROG_PBIDT_perc

ROG-PBDT (%)	Rate of Growth- PBDT	ROG_PBDT_perc
ROG-PBIT (%)	Rate of Growth- PBIT	ROG_PBIT_perc
ROG-PBT (%)	Rate of Growth- PBT	ROG_PBT_perc
ROG-PAT (%)	Rate of Growth- PAT	ROG_PAT_perc
ROG-CP (%)	Rate of Growth- CP	ROG_CP_perc
ROG-Revenue earnings in forex (%)	Rate of Growth - Revenue earnings in forex	ROG_Rev_earn_in_forex_perc
ROG-Revenue expenses in forex (%)	Rate of Growth - Revenue expenses in forex	ROG_Rev_exp_in_forex_perc
ROG-Market Capitalisation (%)	Rate of Growth - Market Capitalisation	ROG_Market_Capitalisation_perc
Current Ratio[Latest]	Liquidity ratio, company's ability to pay short-term obligations or those due within one year	Curr_Ratio_Latest
Fixed Assets Ratio[Latest]	Solvency ratio, the capacity of a company to discharge its obligations towards long-term lenders indicating	Fixed_Assets_Ratio_Latest
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
Debtors Ratio[Latest]	Measures how quickly cash debtors are paying back to the company	Debtors_Ratio_Latest
Total Asset Turnover Ratio[Latest]	The value of a company's revenues relative to the value of its assets	Total_Asset_Turnover_Ratio_Latest
Interest Cover Ratio[Latest]	Determines how easily a company can pay interest on its outstanding debt	Interest_Cover_Ratio_Latest
PBIDTM (%)[Latest]	Profit before Interest Depreciation and Tax Margin	PBIDTM_perc_Latest
PBITM (%)[Latest]	Profit Before Interest Tax Margin	PBITM_perc_Latest
PBDTM (%)[Latest]	Profit Before Depreciation Tax Margin	PBDTM_perc_Latest
CPM (%)[Latest]	Cost per thousand (advertising cost)	CPM_perc_Latest
APATM (%)[Latest]	After tax profit margin	APATM_perc_Latest
Debtors Velocity (Days)	Average days required for receiving the payments	Debtors_Vel_Days
Creditors Velocity (Days)	Average number of days company takes to pay suppliers	Creditors_Vel_Days
Inventory Velocity (Days)	Average number of days the company needs to turn its inventory into sales	Inventory_Vel_Days
Value of Output/Total Assets	Ratio of Value of Output (market value) to Total Assets	Value_of_Output_to_Total_Assets
Value of Output/Gross Block	Ratio of Value of Output (market value) to Gross Block	Value_of_Output_to_Gross_Block

1.2.1. Descriptive Data analysis:

- Provided data set consists of total 67 variables out of which one is a dependent variable (default) has been created taking the total up to 68.
    a) As provided in the hint, 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. This is our target variable.
    b) The Net worth Next Year can be dropped now since it has been converted to a binary form now known as the 'default' column.
    c) The Company Code and Name column also have been dropped for ease of calculation purposes
- Data set contains total of 3586 entries among which 67 are integer type variables and 1 object type variables.
- The size of the dataset is 243848.
- The following Table 1 consists the head(), tail(), info() and description both normal and statistical of the dataset at hand before the default variable was created in the steps mentioned above.

Table-2: Summary of Descriptive statistics information

Head of the dataset before fixing messy column names:

	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

Info of dataset:

```
<class 'pandas.core.frame.DataFrame'>
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_to_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_to_Total_Assets         3586 non-null   float64
 66  Value_of_Output_to_Gross_Block          3586 non-null   float64
dtypes: float64(63), int64(3), object(1)
memory usage: 1.8+ MB
```

Head and Tail of the dataset after fixing messy columns:

	Co_Code	Co_Name	Networth_Next_Year	Equity_Paid_Up	Networth	Capital_Employed	Total_Debt	Gross_Block_	Net_Working_Capital_	Current_Assets_
0	16974	Hind.Cables	-8021.60	419.36	-7027.48	-1007.24	5936.03	474.30	-1076.34	40.50
1	21214	Tata Tele. Mah.	-3986.19	1954.93	-2968.08	4458.20	7410.18	9070.86	-1098.88	486.86
2	14852	ABG Shipyard	-3192.58	53.84	506.86	7714.68	6944.54	1281.54	4496.25	9097.64
3	2439	GTL	-3054.51	157.30	-623.49	2353.88	2326.05	1033.69	-2612.42	1034.12
4	23505	Bharati Defence	-2967.36	50.30	-1070.83	4675.33	5740.90	1084.20	1836.23	4685.81

5 rows × 67 columns

	Co_Code	Co_Name	Networth_Next_Year	Equity_Paid_Up	Networth	Capital_Employed	Total_Debt	Gross_Block_	Net_Working_Capital_	Current_Assets_
3581	4987	HDFC Bank	72677.77	501.30	62009.42	590576.00	496009.19	8463.30	0.00	444633.50
3582	502	Vedanta	79162.19	296.50	34057.87	71906.06	37643.79	29848.44	2503.86	11554.45
3583	12002	I O C L	88134.31	2427.95	67969.97	140686.75	55245.01	121643.45	6376.84	89609.82
3584	12001	NTPC	91293.70	8245.46	81657.35	173099.14	85995.34	128477.59	11449.79	42353.59
3585	15542	Bharti Airtel	111729.10	1998.70	78270.80	104241.00	21569.70	100084.90	-12145.30	11947.10

5 rows × 67 columns

Describe function on dataset:

	Co_Code	Networth_Next_Year	Equity_Paid_Up	Networth	Capital_Employed	Total_Debt	Gross_Block_	Net_Working_Capital_	Current_Assets_	Current_
count	3586.00	3586.00	3586.00	3586.00	3586.00	3586.00	3586.00	3586.00	3586.00	
mean	16065.39	725.05	62.97	649.75	2799.61	1994.82	594.18	410.81	1960.35	
std	19776.82	4769.68	778.76	4091.99	26975.14	23652.84	4871.55	6301.22	22577.57	
min	4.00	-8021.60	0.00	-7027.48	-1824.75	-0.72	-41.19	-13162.42	-0.91	
25%	3029.25	3.98	3.75	3.89	7.60	0.03	0.57	0.94	4.00	
50%	6077.50	19.02	8.29	18.58	39.09	7.49	15.87	10.14	24.54	
75%	24269.50	123.80	19.52	117.30	226.60	72.35	131.90	61.17	135.28	
max	72493.00	111729.10	42263.46	81657.35	714001.25	652823.81	128477.59	223257.56	721166.00	

8 rows × 66 columns

Null or Missing Values:

- There are null values present in the dataset.
- Variables with more than 30% null values shall be dropped as we progress and the rest of the missing values will be imputed as deemed necessary

Shape of the Dataset : (3586, 68)

Size of the Dataset: 243848

1.2.2. Missing Values and Outlier Treatment:

- There are a significant number of null values and Outliers present in the dataset. In a real world scenario we would not be expected to do much in order to alter the raw dataset provided, unless specifically told to do so by the client.
- It has been requested in this business report to treat outliers and impute the missing values so the same has been done.
- For mere purpose of ease, we have put in a formulae to calculate the outliers in the dataset and converted these outliers into null/missing values.
- Once this has been done we move to create a Heatmap stating all the values missing in the dataset in ‘red’.
- Those variables with more than 30% missing variables have been dropped. These variables are 'ROG_Revenue_expenses_in_forex_perc', 'ROG_Revenue_earnings_in_forex_perc'.

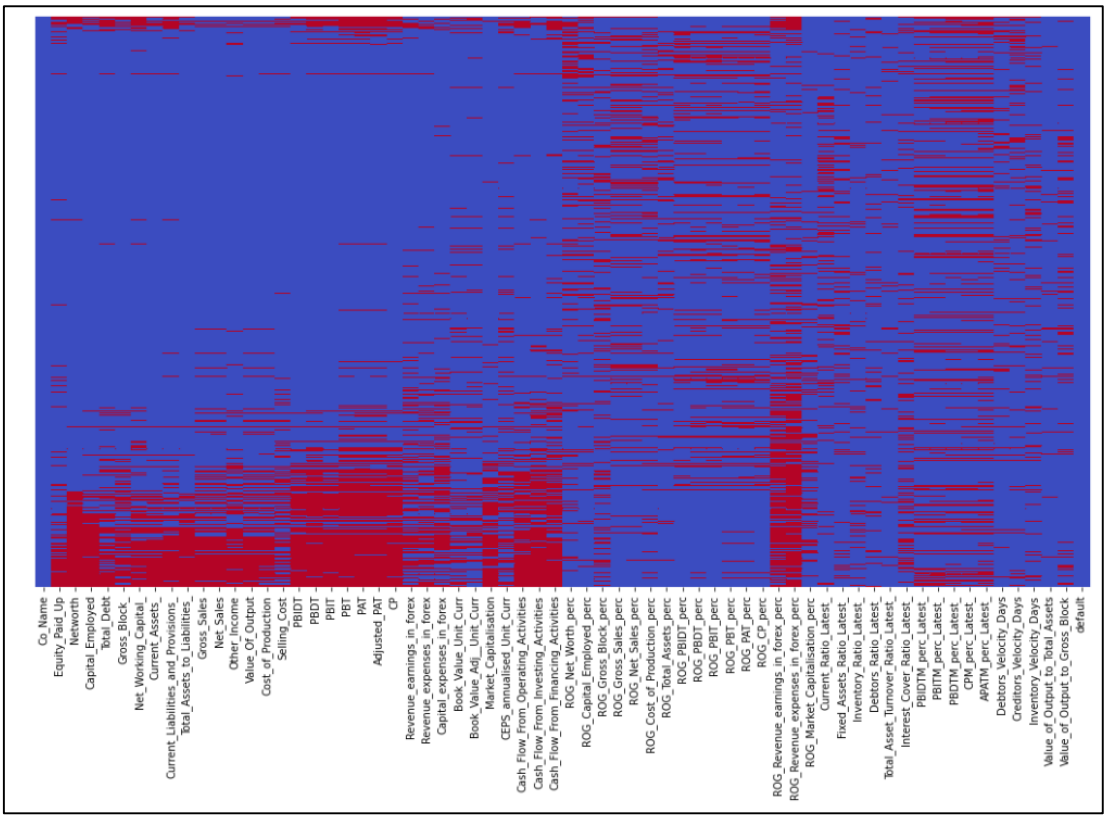


Figure-1(a) Heatmap showcasing missing variables in the dataset

```
ROG_Revenue_expenses_in_forex_perc    0.45
ROG_Revenue_earnings_in_forex_perc    0.37
Cash_Flow_From_Financing_Activities   0.28
PAT                                   0.27
Adjusted_PAT                          0.27
                                      ...
Inventory_Velocity_Days               0.10
Total_Asset_Turnover_Ratio_Latest_    0.06
Value_of_Output_to_Total_Assets       0.04
Co_Name                               0.00
default                               0.00
Length: 66, dtype: float64
```

Figure-1(b) Snapshot of column wise missing value %

- The missing values have been imputed using KNN Imputer, with n_neighbors being 10.
- As we move forward the boxplots are showcased with outliers so as to maintain the essence of the raw data. The treated outliers can be seen in the Jupyter notebook.

1.2.3. Univariate and Bivariate data Analysis:

- Univariate analysis is the simplest form of analyzing data. Univariate data requires to analyze each variable separately. While, a Bivariate analysis will measure the correlations between two variables.
- Figure-2 shows individual distributions for select continuous and ordinal variables present in the data set. We can see that most of the histograms are either normally distributed or right skewed.
- Out of the 67 features, there were 15 that have been found to be the highest ranked variables with Recursive Feature Elimination (RFE). From those 15 here are select distributions of 12 distinct variables.
- The data dictionary provided talks plenty about the different variables in the dataset at hand.

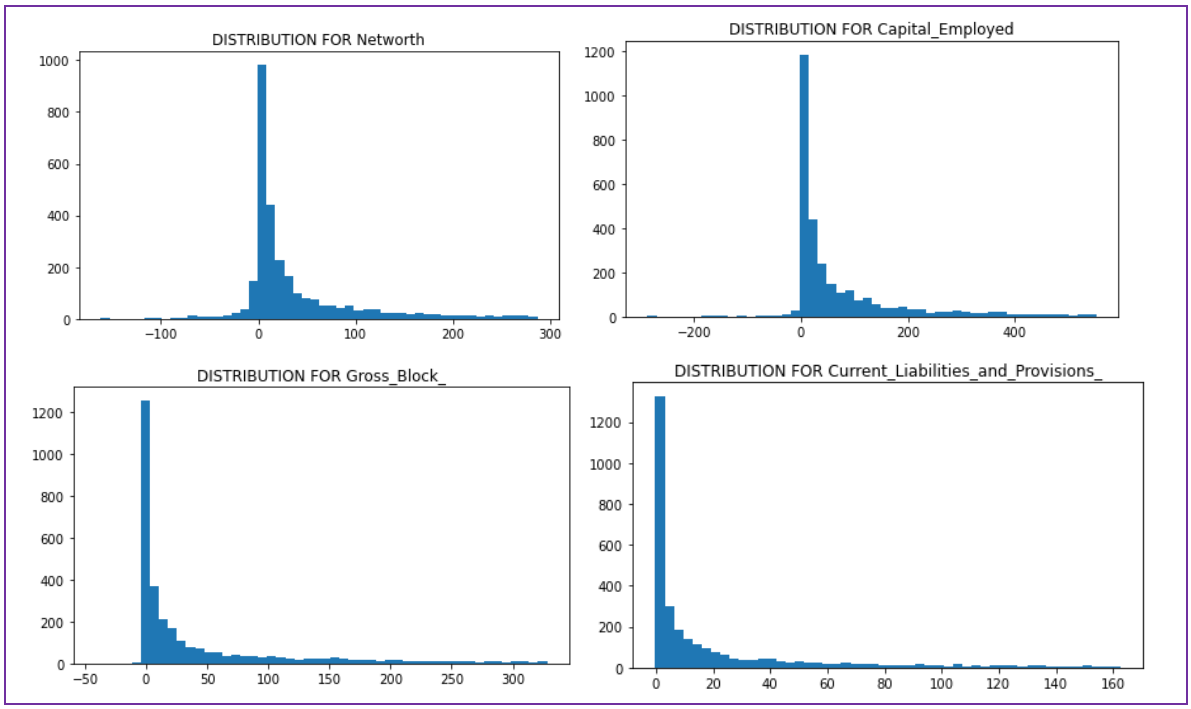


Figure-2(a) Histograms for select continuous and ordinal variables

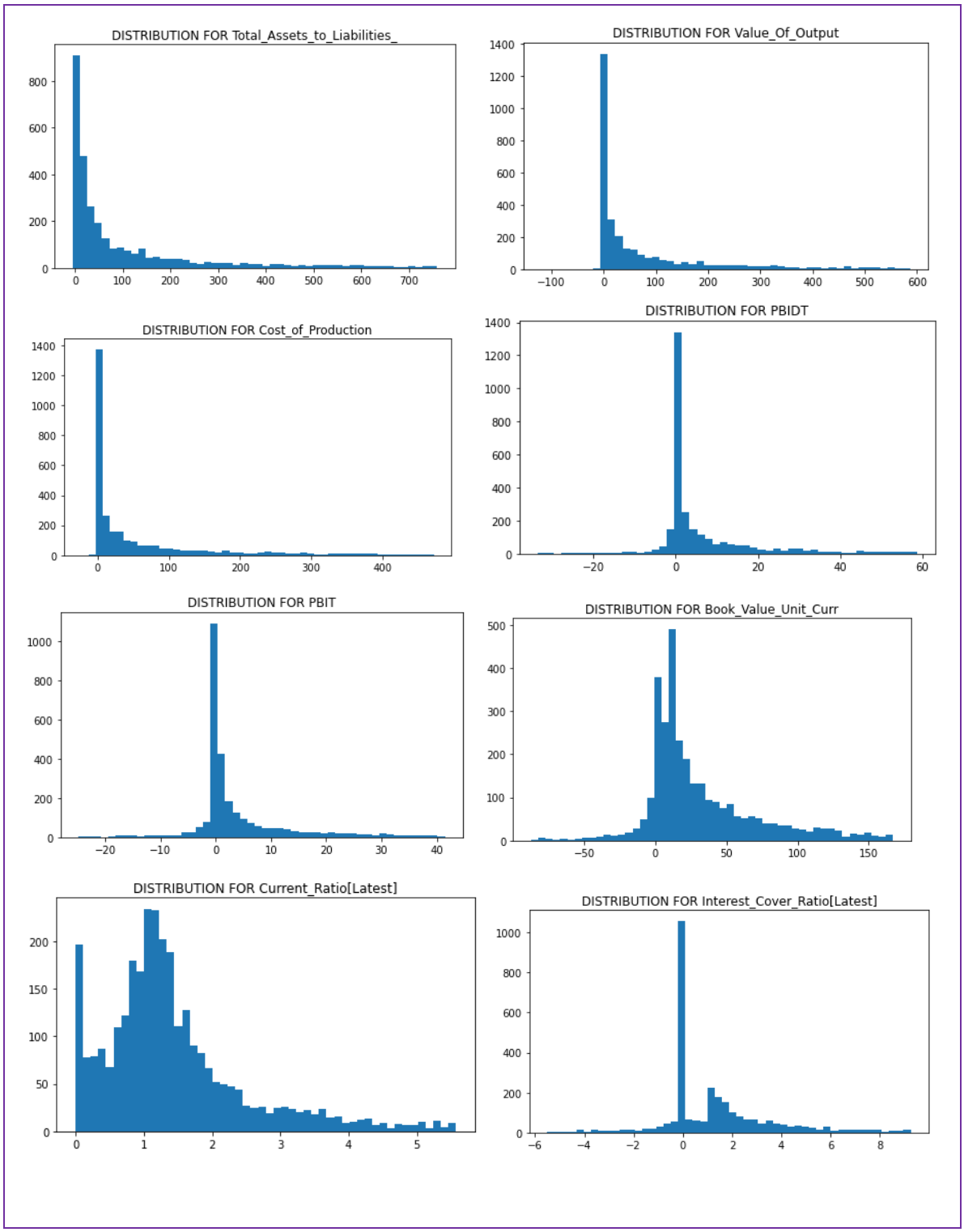


Figure-2(b) Histograms for select continuous and ordinal variables

- Figure-3 shows boxplots for select continuous and ordinal variables present in the data set. The selection process of the same has been deeply explained already above.
- We can see that the dataset has a lot of outliers these outliers have been treated and an explanation of how this has been done will soon follow.
- It is clear from the boxplots that the select variables are positively skewed, ie: the mean is greater than the median.

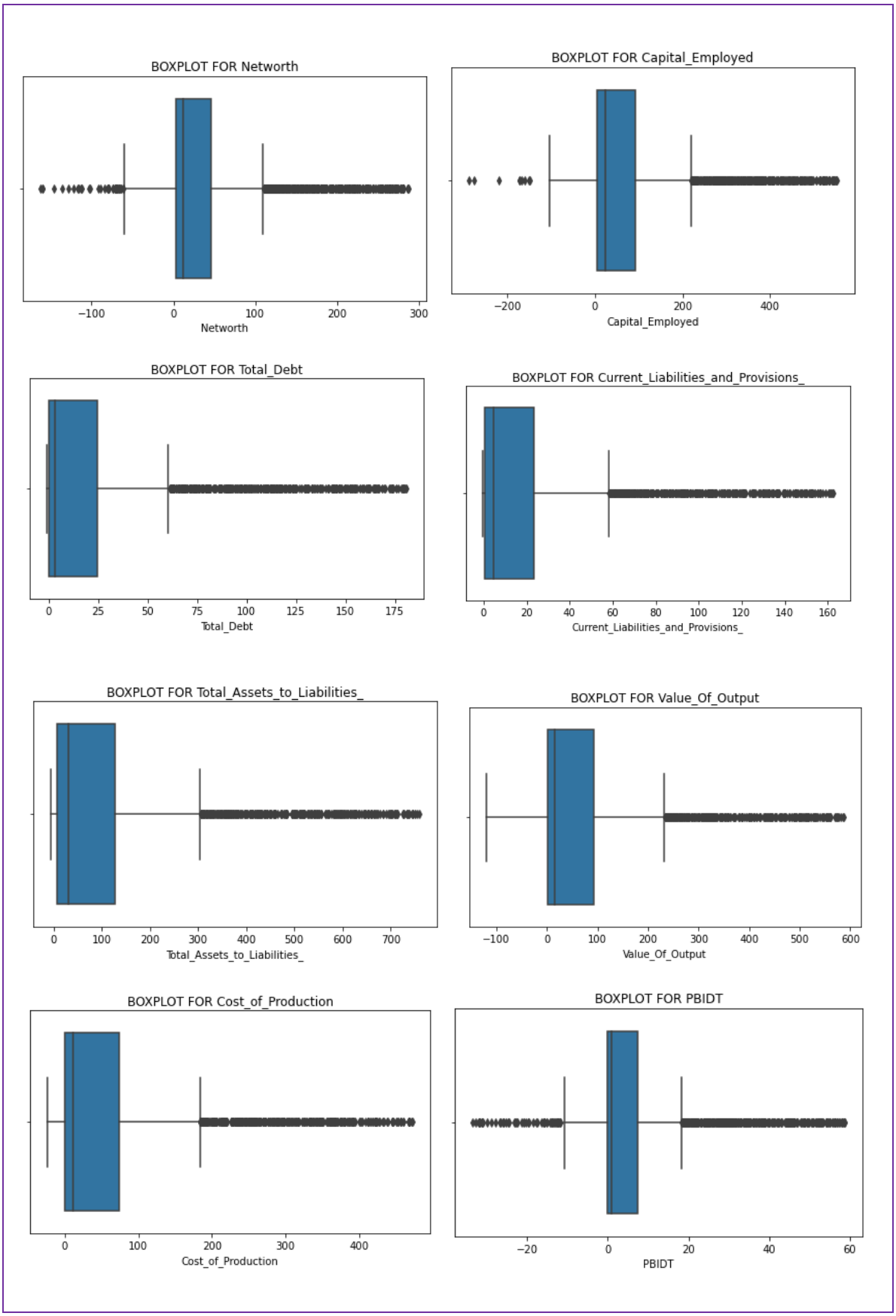


Figure-3(a) Boxplots for select continuous and ordinal variables

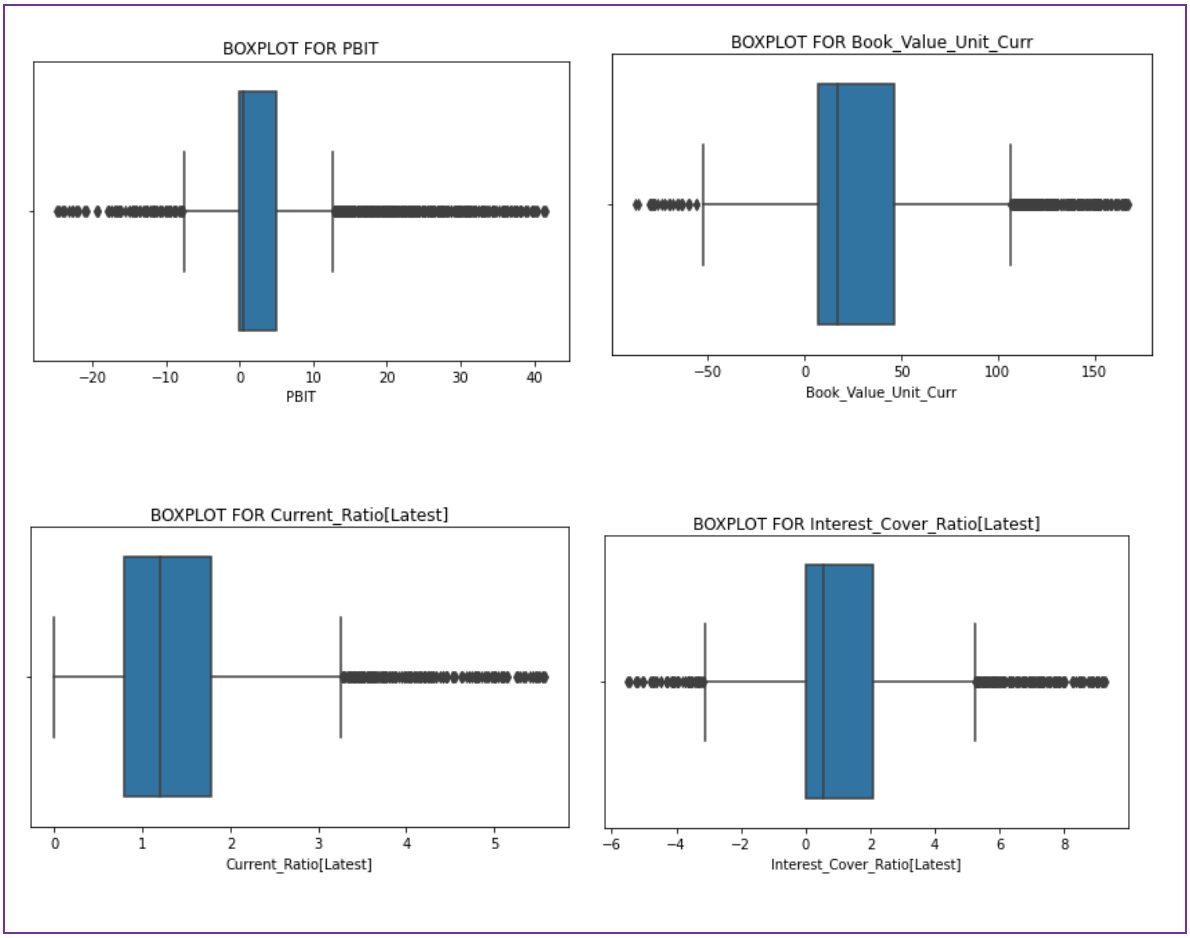


Figure-3(b) Boxplots for select continuous and ordinal variables

- Figure-4 shows count plot of dependent variable. From the figure it is observed that there are very few companies likely to default in proportion to those that won't default. This is good for the company's growth and from an investor point of view, chances of investment are higher for non-defaulters.
- Investors focus on how companies deal with the different financial obligations and their growth scales before making an informed decision on any investment.
- Creating a defaulters variable and categorizing based on net worth next year is hence a smart move.

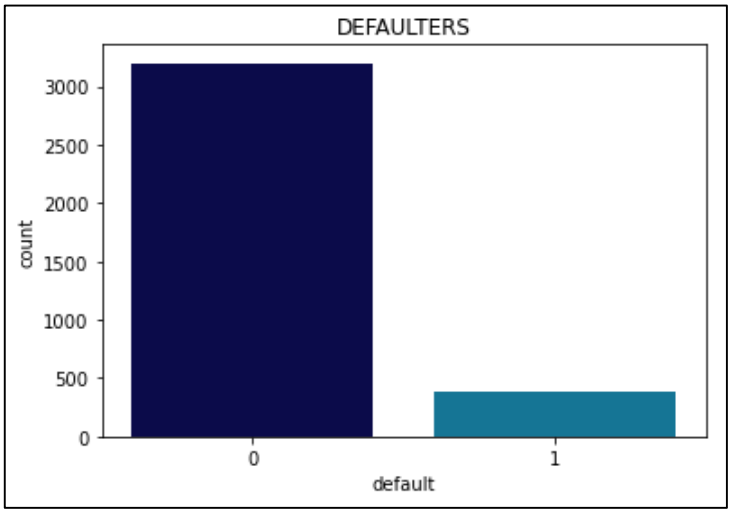


Figure-4 Count plot for Target variable

- Figure-5 shows us the bivariate analysis done for a few pairs of variables, these pairs have been created based on the financial perspective to the best of my knowledge.
- We see that the Gross Sales and Net Sales have a higher correlation with each other this may be because the gross sales are the value of business sales transactions over a specified period of time without accounting for any deductions, while Net Sales takes the deductions into consideration.
- PBT and PAT also have a positive correlation with one another, though there are points scattered from the line in the graph which is due to maybe outliers present in the dataset. PBT here is Profit before Tax and PAT is Profit after calculating tax.
- Cost of Production and Value of output is positively correlated to one another with a lot more noise in data. To the Cost of Production when fixed and variable costs the total value of the product at hand increases and so does the Value of the Output.
- The other graphs showcased clearly are not well correlated with one another. We can see the points are scattered all over not forming any line or shape of any kind. The scatter plots cannot definitely tell us much about the variables here. Maybe the correlation matrix may be of some help.

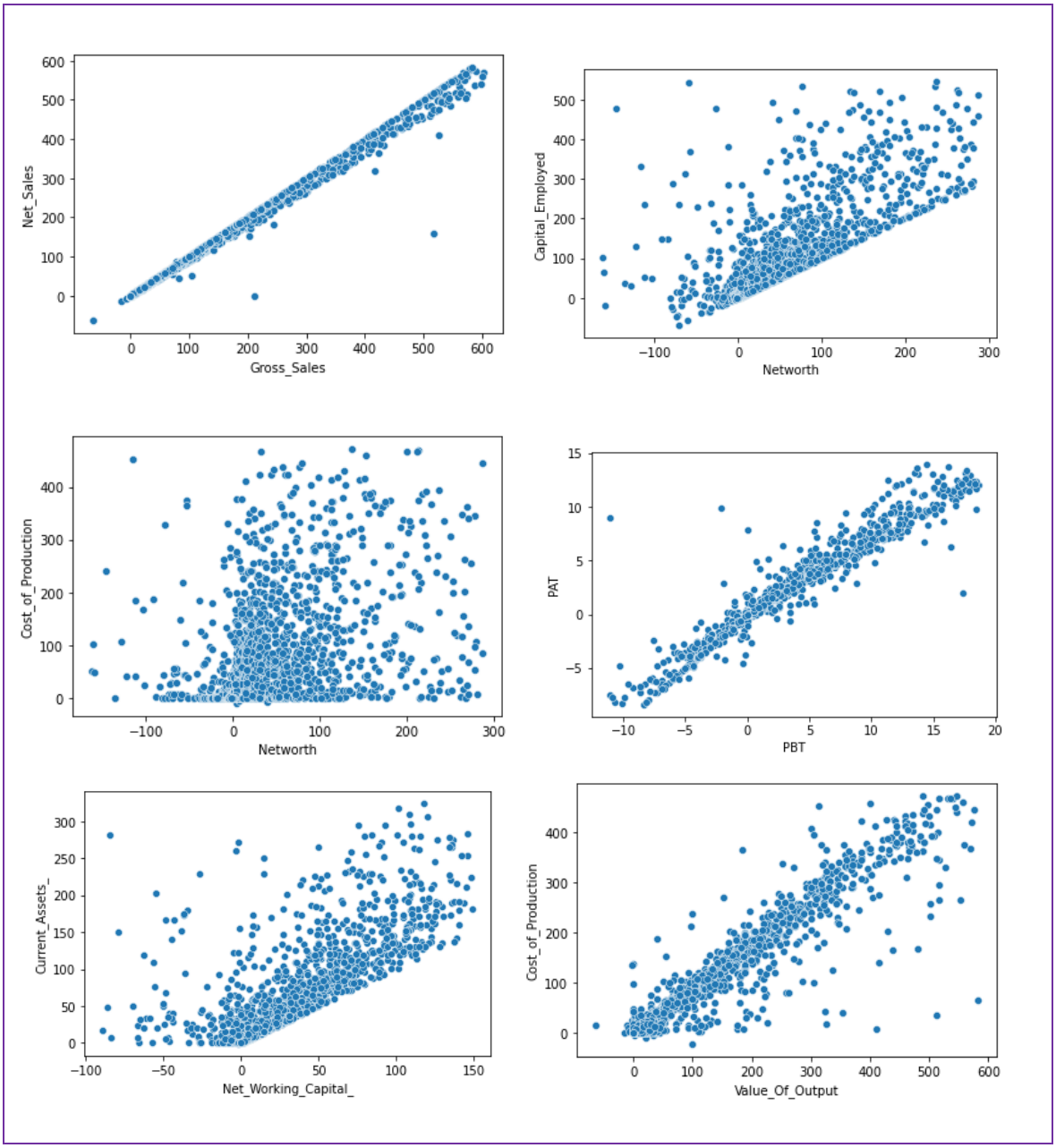


Figure-5 Bivariate Analysis through Scatterplots

1.2.4. Correlation analysis:

- Figure-6 demonstrates the heat map or correlation plot of variables. A heat map is a two-dimensional representation of data in which values are represented by colors. A simple heat map provides an immediate visual summary of information.
- As per figure it is observed that there is a weak correlation between the variables while mostly positive there are some negatively weak too.
- Those values that are lighter shades of blue and white are positively correlated. While those with darker hues are negatively correlated with one and other.
- Since there are so many variables is a little hard to promptly point out the variables distinctively.

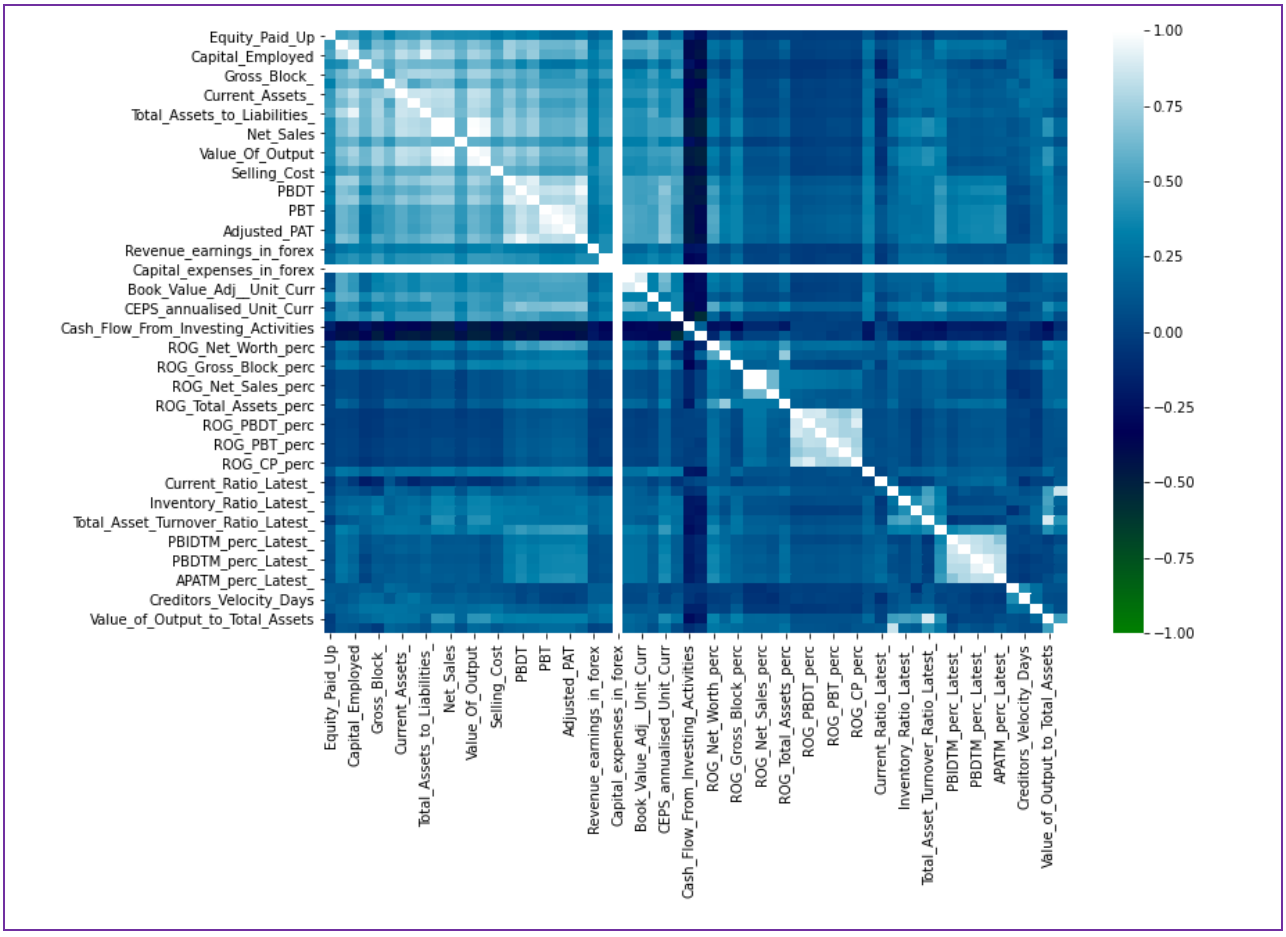


Figure-6 Heat map or Correlation plot for continuous/ordinal variables

1.3. Data Split:

- Data split was performed with 67:33 ratio of Train and Test data using defined random state=42.
- Total of 62 independent variable were present in the X data frame whereas Y contains the dependent variable which was ‘default’, since the aim here is to check how many will default.
- We have reached 63 variables are dropping necessary columns.

1.4. Logistic Regression Model

- Before we move on to use Stats model to build the logistic regression model we shall first eliminate unnecessary features using the Recursive Feature Elimination (RFE) Method. This will help save

us time, as using the Stats Model features can only be eliminated one by one based on the variable having the highest probability. This makes it a pretty time consuming process since we have 63 variables to look into.

- I have limited the no. of features to be selected to 15, hoping this is the best case scenario for the RFE.
- Now I have 15 highly ranked variables that can be put into the Stats Model formula directly, this method not only saves time but also labour.
- Due to use RFE before Stats Model, all the p values are above 0.05 hence none of the variables need to be dropped manually.

	Feature	Rank
1	Networth	1
2	Capital_Employed	1
4	Gross_Block_	1
7	Current_Liabilities_and_Provisions_	1
8	Total_Assets_to_Liabilities_	1
12	Value_Of_Output	1
13	Cost_of_Production	1
15	PBIDT	1
17	PBIT	1
25	Book_Value_Unit_Curr	1
26	Book_Value_Adj__Unit_Curr	1
32	ROG_Net_Worth_perc	1
33	ROG_Capital_Employed_perc	1
46	Current_Ratio_Latest_	1
51	Interest_Cover_Ratio_Latest_	1

Figure-7 RFE Ranked

Logit Regression Results

Dep. Variable:	default	No. Observations:	2402
Model:	Logit	Df Residuals:	2386
Method:	MLE	Df Model:	15
Date:	Sun, 10 Jul 2022	Pseudo R-squ.:	0.5863
Time:	20:28:42	Log-Likelihood:	-327.37
converged:	True	LL-Null:	-791.34
Covariance Type:	nonrobust	LLR p-value:	3.686e-188

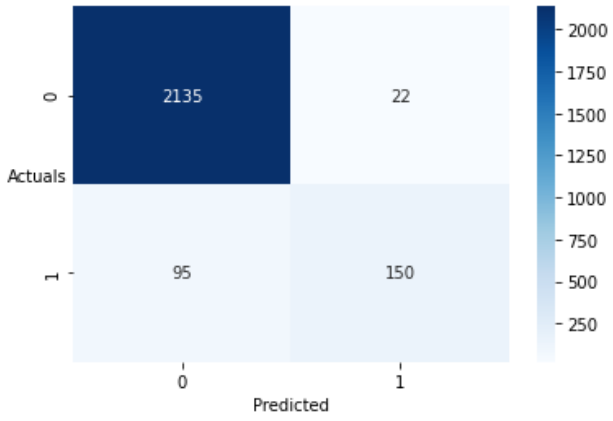
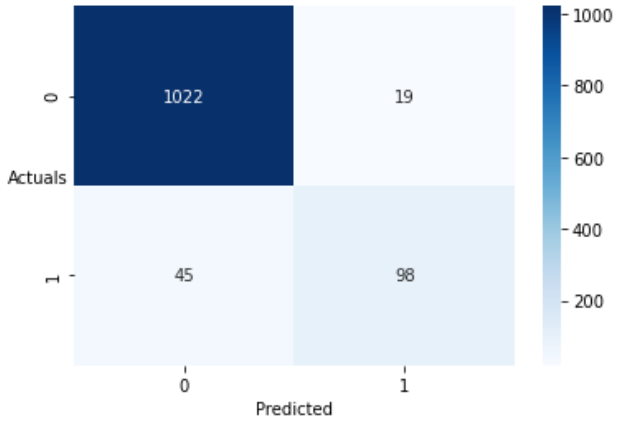
	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-5.2239	0.292	-17.872	0.000	-5.797	-4.651
Networth	-1.5555	0.334	-4.664	0.000	-2.209	-0.902
Capital_Employed	-0.7493	0.309	-2.424	0.015	-1.355	-0.143
Gross_Block_	0.8500	0.228	3.733	0.000	0.404	1.296
Current_Liabilities_and_Provisions_	0.7379	0.236	3.125	0.002	0.275	1.201
Total_Assets_to_Liabilities_	0.7680	0.306	2.509	0.012	0.168	1.368
Value_Of_Output	-1.8154	0.552	-3.290	0.001	-2.897	-0.734
Cost_of_Production	1.6849	0.489	3.447	0.001	0.727	2.643
PBIDT	-1.2197	0.257	-4.745	0.000	-1.724	-0.716
PBIT	0.9219	0.251	3.670	0.000	0.430	1.414
Book_Value_Unit_Curr	-2.0100	0.544	-3.693	0.000	-3.077	-0.943
Book_Value_Adj__Unit_Curr	-1.5899	0.539	-2.950	0.003	-2.646	-0.533
ROG_Net_Worth_perc	-0.5607	0.149	-3.768	0.000	-0.852	-0.269
ROG_Capital_Employed_perc	0.4830	0.132	3.672	0.000	0.225	0.741
Current_Ratio_Latest_	-1.0811	0.163	-6.639	0.000	-1.400	-0.762
Interest_Cover_Ratio_Latest_	-0.7117	0.167	-4.265	0.000	-1.039	-0.385

Figure-8 Logit Regression Stats Model

1.2.5. Logistic Regression Models Performance and Inference:

- As per Logistic Regression analysis the following summary metrics are presented. These metrics are Confusion Matrix and Classification Report.

Table-3 Logistic Regression Model Evaluation

Train Data Set	Test Data Set
Confusion Matrix:	**Confusion Matrix:**
	
Classification Report:	**Classification Report:**
precision recall f1-score support 0.0 0.957 0.990 0.973 2157 1.0 0.872 0.612 0.719 245 accuracy 0.951 2402 macro avg 0.915 0.801 0.846 2402 weighted avg 0.949 0.951 0.947 2402	precision recall f1-score support 0.0 0.958 0.982 0.970 1041 1.0 0.838 0.685 0.754 143 accuracy 0.946 1184 macro avg 0.898 0.834 0.862 1184 weighted avg 0.943 0.946 0.944 1184
• Overall 95.1% of correct predictions to total predictions were made by the model • 61% of those defaulted were correctly identified as defaulters by the model.	• Overall 94.6% of correct predictions to total predictions were made by the model • 68% of those defaulted were correctly identified as defaulters by the model. • The test data is performing better than the train.

- In general, a model fits the data well if the differences between the observed values and the model's predicted values are small and unbiased. The train and test model scores are not too far apart from one another hence this is a good fit model.
- Precision is the fraction of true positive examples among the examples that the model classified as positive. In other words, the number of true positives divided by the number of false positives plus true positives.
- Recall, also known as sensitivity, is the fraction of examples classified as positive, among the total number of positive examples. In other words, the number of true positives divided by the number of true positives plus false negatives.
- When both the recall and precision values are important we look at the F1-score, as it is the harmonic mean of precision and recall. It combines precision and recall into a single number. It is a measure of the models accuracy on the dataset. These scores are closer to 1 hence stating that they have a good accuracy.