



# FINANCIAL RISK ASSESSMENT (FRA) BUSINESS REPORT

## ABSTRACT

Dataset 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).

**Yashveer Kothari. A**

Post Graduate Programme in Data Science & Business Analytics

LIST OF CONTENTS		
CHAPTER/ QUESTION#	CONTENTS	PAGE#
FINANCIAL RISK ASSESSMENT	ABOUT FINANCIAL RISK ASSESSMENT	5
	TYPES OF FINANCIAL RISK ASSESSMENT	5
	3 ASPECTS OF CREDIT RISK	6
PROBLEM	PROBLEM STATEMENT	7
	DATA DICTIONARY	7
	Q.1.1. OUTLIER TREATMENT	10
	Q.1.2. MISSING VALUE TREATMENT	12
	Q.1.3 TRANSFORM TARGET VARIABLE TO 0 AND 1	16
	Q.1.4. UNIVARIATE AND BI VARIATE ANALYSIS	17
	Q.1.5. TRAIN TEST SPLIT	50
	Q.1.6. BUILD LOGISTIC REGRESSION MODEL ....	50
	Q.1.7 VALIDATE THE MODEL ON TEST DATASET ...	53

LIST OF FIGURES		
TABLE#	TABLE NAME	PAGE#
1	MISSING DATA IN PERCENTAGE	13
2	MISSING VALUES	14
3	MISSING VALUES AFTER TREATMENT	15
4	TRANSFORMING TARGET VARIABLE	16
5	PROPORTION OF DEFAULT	17
6	SKEWNESS OF IMPORTANT FEATURES	33
7	VARAINCE INFLATION FACTOR	49
8	MODEL 13	52
9	PREDICTION TRAIN DATA	53
10	PREDICTION TEST DATA	53
11	CLASSIFICATION REPORT TRAIN DATA	54
12	CLASSIFICATION REPORT TEST DATA	55

LIST OF FIGURES		
FIGURE#	FIGURE NAME	PAGE#
1	OUTLIERS IN THE DATSET	10
2	OUTLIERS AFTER TREATMENT	11
3	VISUALIZING THE MISSING VALUES	12
4	DISTRIBUTION & BOXPLOT OF NET WORKING_CAPITAL_	17
5	DISTRIBUTION & BOXPLOT OF BOOK VALUE UNIT CURR	18
6	DISTRIBUTION & BOXPLOT OF ROG NET WORTH PERC	19
7	DISTRIBUTION & BOXPLOT OF ROG CAPITAL EMPLOYED PERC	20
8	DISTRIBUTION & BOXPLOT OF ROG TOTAL ASSETS PERC	21
9	DISTRIBUTION & BOXPLOT OF CURRENT RATIO LATEST	22
10	DISTRIBUTION & BOXPLOT OF FIXED ASSET RATIO LATEST	23
11	DISTRIBUTION & BOXPLOT OF INVENTORY RATIO LATEST	24
12	DISTRIBUTION & BOXPLOT OF DEBTORS RATIO LATEST	25
13	DISTRIBUTION & BOXPLOT OF TOTAL ASSET TURNOVER RATIO LATEST	26
14	DISTRIBUTION & BOXPLOT OF INTEREST COVER RATIO LATEST	27
15	DISTRIBUTION & BOXPLOT OF ROG MARKET CAPITALISATION PERC	28
16	DISTRIBUTION & BOXPLOT OF ROG COST OF PRODUCTION PERC	29
17	COUNTPLOT OF TARGET VARIABLE AND DEFAULT	31
18	DISTRIBUTION PLOT OF IMPORTANT FEATURES	32
19	PLOT OF DEFAULT WITH CURRENT RATIO	35
20	PLOT OF DEFAULT WITH TOTAL ASSET TURNOVER RATIO	36
21	PLOT OF DEFAULT WITH FIXED ASSET TURNOVER RATIO	37

22	PLOT OF DEFAULT WITH ROG MARKET CAPITALIZATION PERC	38
23	PLOT OF DEFAULT WITH INTEREST COVER RATIO	39
24	PLOT OF DEFAULT WITH INVENTORY RATIO	40
25	PLOT OF DEFAULT WITH DEBTORS RATIO	41
26	PLOT OF DEFAULT WITH NET WORKING CAPITAL	42
27	PLOT OF DEFAULT WITH NET WORKING CAPITAL	43
28	PLOT OF DEFAULT WITH ROG CAPITAL EMPLOYED PERC	44
29	PLOT OF DEFAULT WITH BOOK VALUE UNIT CURR	45
30	PLOT OF DEFAULT WITH ROG TOTAL ASSETS PERC	46
31	HEATMAP	47
32	CONFUSION MATRIX OF TRAIN DATA	54
33	CONFUSION MATRIX OF TEST DATA	55

## ABOUT FINANCIAL RISK ASSESMENT

- A financial risk assessment is an evaluation of the probability that a financial asset or investment could lose value.
- A financial risk assessment can help you identify the risk level specific to your business, prioritize those risks, develop ways to avoid them and outline steps to manage them should they happen. A financial risk assessment also helps you understand your appetite for risk.
- Some common financial risks are credit, operational, foreign investment, legal, equity, and liquidity risks.

## TYPES OF FINANCIAL RISKS



1. **Credit Risk:** This type of risk arises when one fails to fulfil their obligations towards their counterparties. Credit risk can be classified into Sovereign Risk and Settlement Risk. Sovereign risk usually arises due to difficult foreign exchange policies. Settlement risk, on the other hand, arises when one party makes the payment while the other party fails to fulfil the obligations
2. **Liquidity Risk:** This type of risk arises out of an inability to execute transactions. Liquidity risk can be classified into Asset Liquidity Risk and Funding Liquidity Risk. Asset Liquidity risk arises either due to insufficient buyers or insufficient sellers against sell orders and buys orders respectively.
3. **Systematic Risk:** It refers to the risk inherent to the entire market or market segment. Systematic risk, also known as “undiversifiable risk,” “volatility” or “market risk,” affects the overall market, not just a particular stock or industry.
4. **Market Risk:** This type of risk arises due to the movement in prices of financial instrument. Market risk can be classified as Directional Risk and Non-Directional Risk. Directional risk is caused due to movement in stock price, interest rates and more. Non-Directional risk, on the other hand, can be volatility risks.
5. **Operational Risk:** This type of risk arises out of operational failures such as mismanagement or technical failures. Operational risk can be classified into Fraud Risk and Model Risk. Fraud risk arises due to the lack of controls and Model risk arises due to incorrect model application.

### 3 ASPECTS OF CREDIT RISK

1. **Probability of Default (PD):** Chance of Non-Payment (The percentage the portfolio is at risk).  $0 \leq PD \leq 1$ .
2. **Exposure at Default (EAD):** How much of the portfolio is at risk.
3. **Loss given Default (LGD):** The percentage of which the defaulted money is lost and how much percentage is recovered.  $100\% \text{ LGD} = \text{No Recovery}$ ;  $50\% \text{ LGD} = \text{Half loan recovered by sale of collateral/security}$ .

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

## DATA DICTIONARY

#	FIELD NAME	DESCRIPTION
1	Co_Code	Company Code
2	Co_Name	Company Name
3	Networth Next Year	Value of a company as on 2016 - Next Year (difference between the value of total assets and total liabilities)
4	Equity Paid Up	Amount that has been received by the company through the issue of shares to the shareholders
5	Networth	Value of a company as on 2015 - Current Year
6	Capital Employed	Total amount of capital used for the acquisition of profits by a company
7	Total Debt	The sum of money borrowed by the company and is due to be paid
8	Gross Block	Total value of all of the assets that a company owns
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).
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.
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)
12	Total Assets/Liabilities	Ratio of total assets to liabilities of the company
13	Gross Sales	The grand total of sale transactions within the accounting period
14	Net Sales	Gross sales minus returns, allowances, and discounts



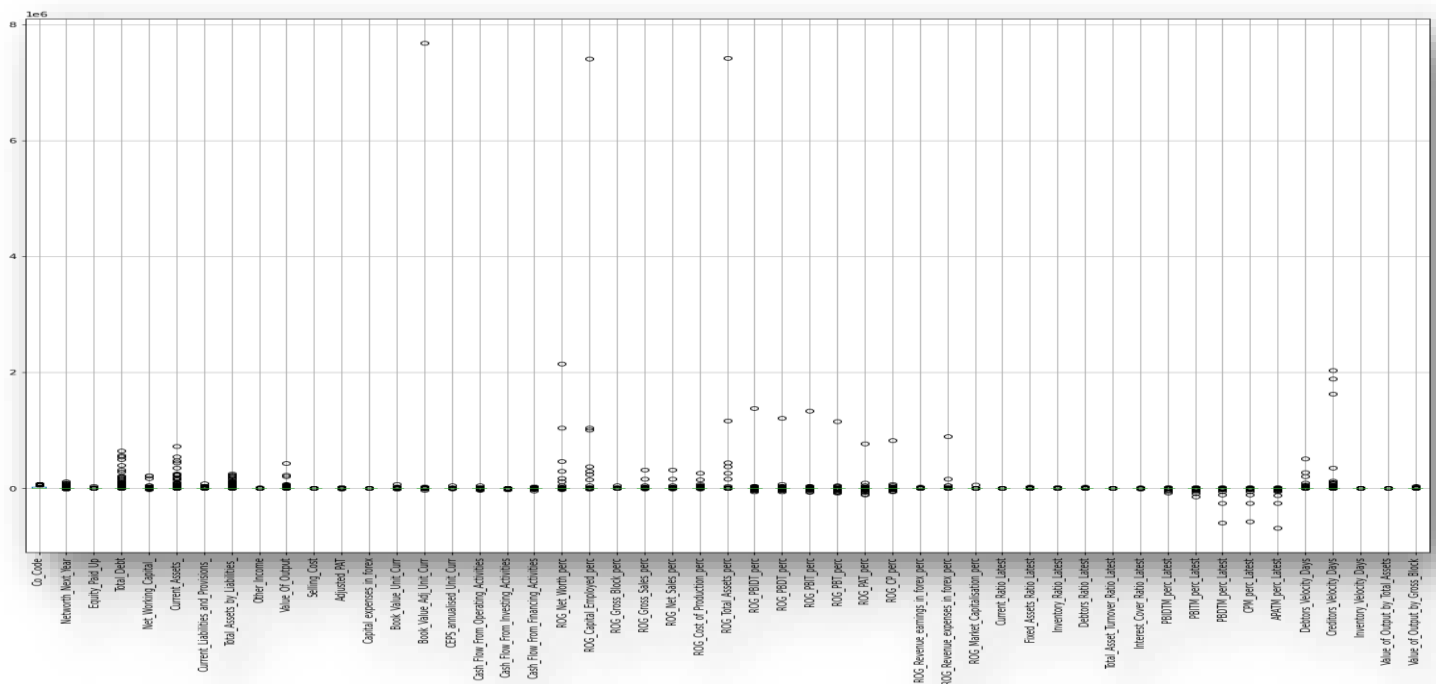
15	Other Income	Income realized from non-business activities (e.g., sale of long-term asset)
16	Value Of Output	Product of physical output of goods and services produced by company and its market price
17	Cost of Production	Costs incurred by a business from manufacturing a product or providing a service
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)
19	PBIDT	Profit Before Interest, Depreciation & Taxes
20	PBDT	Profit Before Depreciation and Tax
21	PBIT	Profit before interest and taxes
22	PBT	Profit before tax
23	PAT	Profit After Tax
24	Adjusted PAT	Adjusted profit is the best estimate of the true profit
26	CP	Commercial paper, a short-term debt instrument to meet short-term liabilities.
27	Revenue earnings in forex	Revenue earned in foreign currency
28	Revenue expenses in forex	Expenses due to foreign currency transactions
29	Capital expenses in forex	Long term investment in forex
30	Book Value (Unit Curr)	Net asset value
31	Book Value (Adj.) (Unit Curr)	Book value adjusted to reflect asset's true fair market value
32	Market Capitalisation	Product of the total number of a company's outstanding shares and the current market price of one share
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
34	Cash Flow From Operating Activities	Use of cash from ongoing regular business activities
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
36	Cash Flow From Financing Activities	Net flows of cash that are used to fund the company (transactions involving debt, equity, and dividends)
37	ROG-Net Worth (%)	Rate of Growth - Net worth
38	ROG-Capital Employed (%)	Rate of Growth - Capital Employed
39	ROG-Gross Block (%)	Rate of Growth - Gross Block
40	ROG-Gross Sales (%)	Rate of Growth - Gross Sales
41	ROG-Net Sales (%)	Rate of Growth - Net Sales
42	ROG-Cost of Production (%)	Rate of Growth - Cost of Production
43	ROG-Total Assets (%)	Rate of Growth - Total Assets
44	ROG-PBIDT (%)	Rate of Growth- PBIDT
45	ROG-PBDT (%)	Rate of Growth- PBDT
46	ROG-PBIT (%)	Rate of Growth- PBIT
47	ROG-PBT (%)	Rate of Growth- PBT
48	ROG-PAT (%)	Rate of Growth- PAT
49	ROG-CP (%)	Rate of Growth- CP

50	ROG-Revenue earnings in forex (%)	Rate of Growth - Revenue earnings in forex
51	ROG-Revenue expenses in forex (%)	Rate of Growth - Revenue expenses in forex
52	ROG-Market Capitalisation (%)	Rate of Growth - Market Capitalisation
53	Current Ratio[Latest]	Liquidity ratio, company's ability to pay short-term obligations or those due within one year
54	Fixed Assets Ratio[Latest]	Solvency ratio, the capacity of a company to discharge its obligations towards long-term lenders indicating
55	Inventory Ratio[Latest]	Activity ratio, specifies the number of times the stock or inventory has been replaced and sold by the company
56	Debtors Ratio[Latest]	Measures how quickly cash debtors are paying back to the company
57	Total Asset Turnover Ratio[Latest]	The value of a company's revenues relative to the value of its assets
58	Interest Cover Ratio[Latest]	Determines how easily a company can pay interest on its outstanding debt
59	PBIDTM (%) [Latest]	Profit before Interest Depreciation and Tax Margin
60	PBITM (%) [Latest]	Profit Before Interest Tax Margin
61	PBDTM (%) [Latest]	Profit Before Depreciation Tax Margin
62	CPM (%) [Latest]	Cost per thousand (advertising cost)
63	APATM (%) [Latest]	After tax profit margin
64	Debtors Velocity (Days)	Average days required for receiving the payments
65	Creditors Velocity (Days)	Average number of days company takes to pay suppliers
66	Inventory Velocity (Days)	Average number of days the company needs to turn its inventory into sales
67	Value of Output/Total Assets	Ratio of Value of Output (market value) to Total Assets
68	Value of Output/Gross Block	Ratio of Value of Output (market value) to Gross Block

## Q1.1. OUTLIER TREATMENT

1. The Inter-quartile range method is use top remove the outliers form the dataset.
2. The dataset before the removal of outliers is as follows:

FIGURE 1: OUTLIERS IN THE DATASET



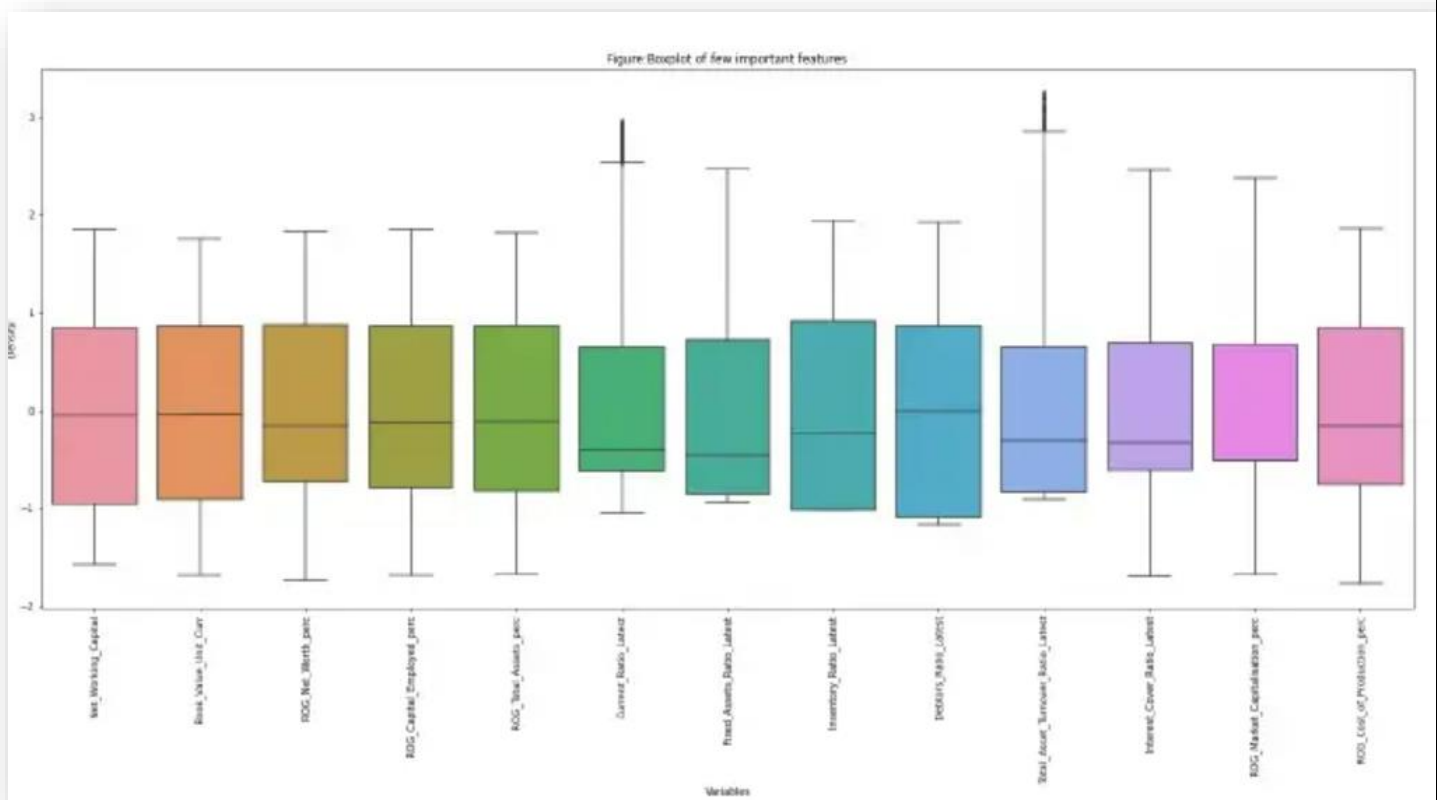
- Using the below formula we remove the outliers from the dataset

$$\text{IQR} = Q3 - Q1$$

$$\text{UL} = \text{Q3} + (1.5 * \text{IQR})$$

$$LL = Q1 - (1.5 * IQR)$$

## FIGURE 2: OUTLIERS AFTER IQR TREATMENT

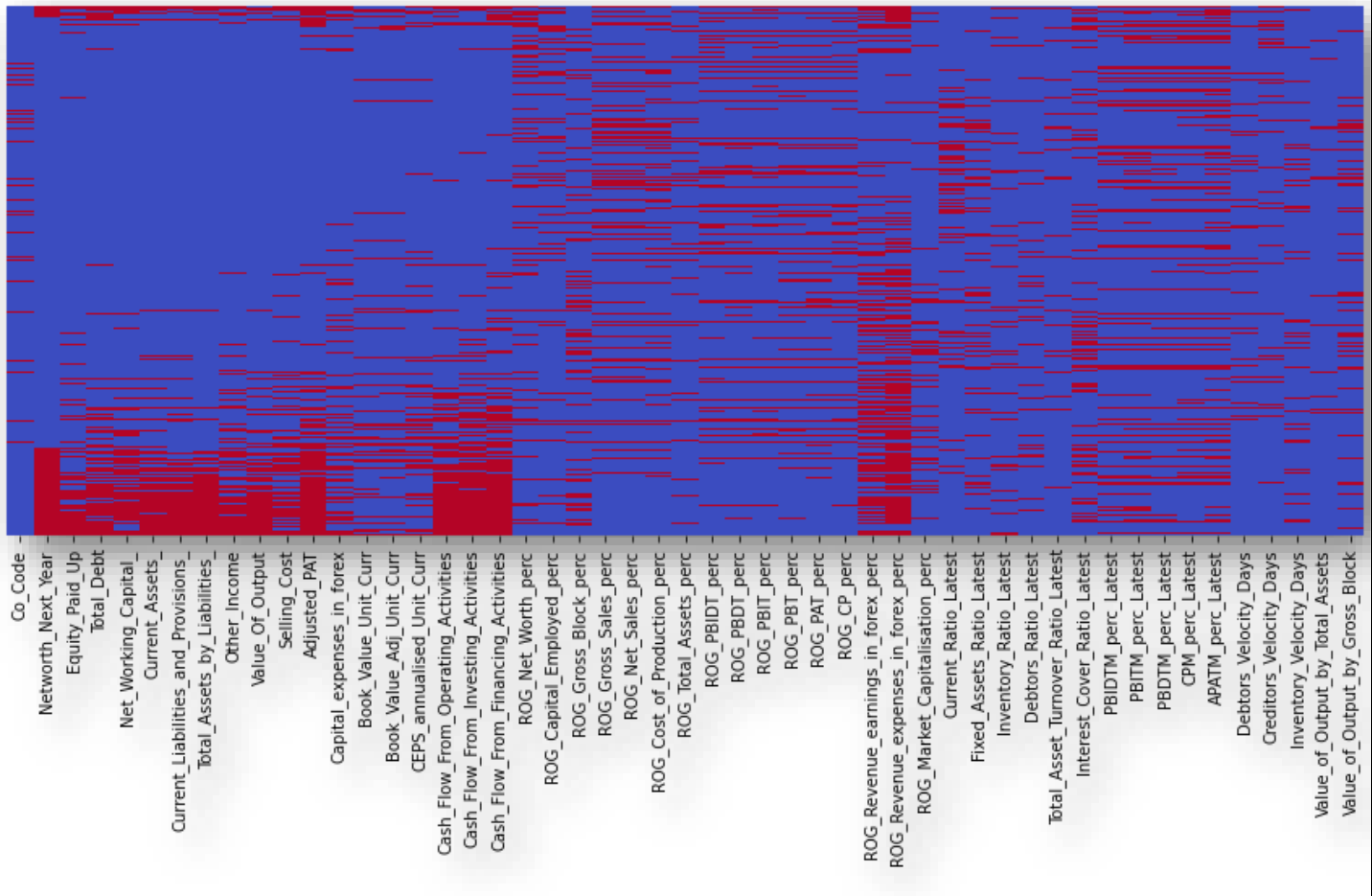


- Variable 'Current\_Ratio\_Latest' and 'Total\_asset\_Turnover\_Ratio\_Latest' still have some extreme values.

## Q1.2. MISSING VALUE TREATMENT

1. There are a total of 118 missing value in the dataset.

**FIGURE 3: VISUALIZING THE MISSING VALUES**



2. Noticeable, presence of missing values in some variables can be observed. Blue color in the heatmap shows the occupied cell and the red color indicates the missing values in the dataset.
3. Typically, if data missing is less than 30% of our data and at row level data is atleast at 90% complete, we do not drop the data. Here, we will first check the completeness of the data and then decide the technique.
4. In order to check the missing values we will have to find the missing values in each row.
5. Trying to target companies which completes atleast 90% of the data in each row i.e. we will filter out companies where there are at least 5 or less missing values to identify the reliable data until this point.

The number rows and columns in the dataset are (3586, 49) respectively after dropping columns having more than 30% missing values

**TABLE 1: MISSING DATA IN PERCENTAGE**

ROG_Revenue_expenses_in_forex_perc	0.450
ROG_Revenue_earnings_in_forex_perc	0.367
Cash_Flow_From_Financing_Activities	0.280
Adjusted_PAT	0.266
APATM_perc_Latest	0.260
Cash_Flow_From_Investing_Activities	0.244
ROG_Gross_Block_perc	0.231
Cash_Flow_From_Operating_Activities	0.223
ROG_Net_Worth_perc	0.208
Interest_Cover_Ratio_Latest	0.202
CPM_perc_Latest	0.201
PBITM_perc_Latest	0.200
PBDTM_perc_Latest	0.194
Capital_expenses_in_forex	0.194
Networth_Next_Year	0.189
ROG_Cost_of_Production_perc	0.188
ROG_Gross_Sales_perc	0.187
ROG_Net_Sales_perc	0.186
ROG_CP_perc	0.178
ROG_PBDT_perc	0.175
Net_Working_Capital_	0.174
ROG_PBIT_perc	0.172
ROG_PBT_perc	0.170
ROG_PBITD_perc	0.170

Selling_Cost	0.169
Other_Income	0.168
CEPS_annualised_Unit_Curr	0.168
ROG_PAT_perc	0.167
PBITM_perc_Latest	0.166
Total_Debt	0.163
Current_Liabilities_and_Provisions_	0.162
Current_Assets_	0.161
Total_Assets_by_Liabilities_	0.160
ROG_Capital_Employed_perc	0.160
Current_Ratio_Latest	0.158
Value_Of_Output	0.156
ROG_Market_Capitalisation_perc	0.139
Fixed_Assets_Ratio_Latest	0.138
Book_Value_Adj_Unit_Curr	0.137
Book_Value_Unit_Curr	0.135
ROG_Total_Assets_perc	0.135
Value_of_Output_by_Gross_Block	0.134
Equity_Paid_Up	0.125
Debtors_Velocity_Days	0.111
Creditors_Velocity_Days	0.109
Inventory_Ratio_Latest	0.105
Debtors_Ratio_Latest	0.104
Inventory_Velocity_Days	0.102
Co_Code	0.081
Total_Asset_Turnover_Ratio_Latest	0.056
Value_of_Output_by_Total_Assets	0.042
dtype: float64	

6. **Dropping columns with more than 30% missing values:** We are sorting the proportion of missing values by dividing number of missing values by number of applicable rows. We will eliminate anything that is more than 30%. Noticeably “ROG\_Revenue\_expenses\_in\_forex\_perc” and “ROG\_Revenue\_earnings\_in\_forex\_perc” are the only values which are more than 30%. Therefore, we can eliminate them.

## TABLE 2: MISSING VALUES

```

Co_Code 0
Networth_Next_Year 0
Equity_Paid_Up 0
Total_Debt 0
Net_Working_Capital_ 0
Current_Assets_ 0
Current_Liabilities_and_Provisions_ 0
Total_Assets_by_Liabilities_ 0
Other_Income 0
Value_Of_Output 0
Selling_Cost 0
Adjusted_PAT 0
Capital_expenses_in_forex 0
Book_Value_Unit_Curr 0
Book_Value_Adj_Unit_Curr 4
CEPS_annualised_Unit_Curr 0
Cash_Flow_From_Operating_Activities 0
Cash_Flow_From_Investing_Activities 0
Cash_Flow_From_Financing_Activities 0
ROG_Net_Worth_perc 0
ROG_Capital_Employed_perc 0
ROG_Gross_Block_perc 0
ROG_Gross_Sales_perc 0
ROG_Net_Sales_perc 0
ROG_Cost_of_Production_perc 0

```

```

ROG_Total_Assets_perc 0
ROG_PBIDT_perc 0
ROG_PBDT_perc 0
ROG_PBIT_perc 0
ROG_PBT_perc 0
ROG_PAT_perc 0
ROG_CP_perc 0
ROG_Revenue_earnings_in_forex_perc 0
ROG_Revenue_expenses_in_forex_perc 0
ROG_Market_Capitalisation_perc 0
Current_Ratio_Latest 1
Fixed_Assets_Ratio_Latest 1
Inventory_Ratio_Latest 1
Debtors_Ratio_Latest 1
Total_Asset_Turnover_Ratio_Latest 1
Interest_Cover_Ratio_Latest 1
PBIDTM_perc_Latest 1
PBITM_perc_Latest 1
PBDTM_perc_Latest 1
CPM_perc_Latest 1
APATM_perc_Latest 1
Debtors_Velocity_Days 0
Creditors_Velocity_Days 0
Inventory_Velocity_Days 103
Value_of_Output_by_Total_Assets 0
Value_of_Output_by_Gross_Block 0
dtype: int64

```

7. The missing values are of numeric in nature. Hence, can be imputed using KNN Imputer function from the impute module of the sklearn. This imputer utilizes the k-nearest neighbours method to replace the missing values in the datasets by finding the nearest neighbours with Euclidean distance matrix.
8. The critical point here is that KNN imputer is a distance based imputation method and requires that the data should be normalized, because different scales of our data can lead the function to give a biased replacements for the values missing. Hence we use the standard scaler method which will scale the variables to have values between 0 and 1
9. The data is split into predictor and response variables.
10. The imputation is done by predicting the missing value based on values of 10 nearest neighbours of the same variable. Such that all the missing values are replaced based on the nearest neighbours value

# TABLE 3: MISSING VALUES AFTER TREATMENT

Co_Code	0	ROG_Total_Assets_perc	0
Equity_Paid_Up	0	ROG_PBIDT_perc	0
Total_Debt	0	ROG_PBDT_perc	0
Net_Working_Capital_	0	ROG_PBIT_perc	0
Current_Assets_	0	ROG_PBT_perc	0
Current_Liabilities_and_Provisions_	0	ROG_PAT_perc	0
Total_Assets_by_Liabilities_	0	ROG_CP_perc	0
Other_Income	0	ROG_Market_Capitalisation_perc	0
Value_Of_Output	0	Current_Ratio_Latest	0
Selling_Cost	0	Fixed_Assets_Ratio_Latest	0
Adjusted_PAT	0	Inventory_Ratio_Latest	0
Capital_expenses_in_forex	0	Debtors_Ratio_Latest	0
Book_Value_Unit_Curr	0	Total_Asset_Turnover_Ratio_Latest	0
Book_Value_Adj_Unit_Curr	0	Interest_Cover_Ratio_Latest	0
CEPS_annualised_Unit_Curr	0	PBIDTM_perc_Latest	0
Cash_Flow_From_Operating_Activities	0	PBITM_perc_Latest	0
Cash_Flow_From_Investing_Activities	0	PBDTM_perc_Latest	0
Cash_Flow_From_Financing_Activities	0	CPM_perc_Latest	0
ROG_Net_Worth_perc	0	APATM_perc_Latest	0
ROG_Capital_Employed_perc	0	Debtors_Velocity_Days	0
ROG_Gross_Block_perc	0	Creditors_Velocity_Days	0
ROG_Gross_Sales_perc	0	Inventory_Velocity_Days	0
ROG_Net_Sales_perc	0	Value_of_Output_by_Total_Assets	0
ROG_Cost_of_Production_perc	0	Value_of_Output_by_Gross_Block	0
		Networth_Next_Year	0
		dtype: int64	

11.Hence there are no missing value in the dataset.



### Q.1.3. TRANSFORM TARGET VARIABLE INTO 0 AND 1

1. There is no target variable defined, but since the objective is to build a model for investors to decode, which company to invest in, we use variable `Networth_Next_Year_` could be used to transform into target variable.
2. We create a default variable where, value = 1, when `Networth_Next_Year_` is negative and value = 0, when `Networth_Next_Year_` is positive.
  - 2.1. `Networth_Next_Year_` is positive: the company is returning good to there investors and is transformed as 1.
  - 2.2. `Networth_Next_Year_` is negative: the company is likely not to return good to investor and is transformed to 0.
3. Hence a binary taget variable is created using `Networth_Next_Year_`.

TABLE 4: TRANSFORMING TARGET VARIABLE

	default	Networth_Next_Year
0	0	18.601
1	0	16.505
2	0	104.304
3	0	26.609
4	0	26.419
5	1	-4.055
6	0	3.349
7	1	-45.285
8	0	24.496
9	0	149.660

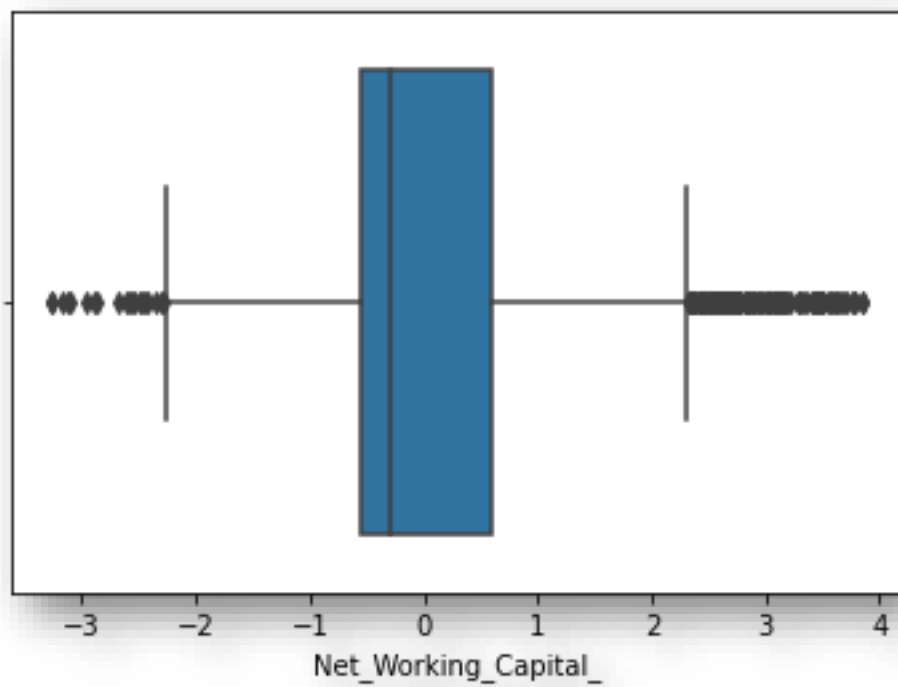
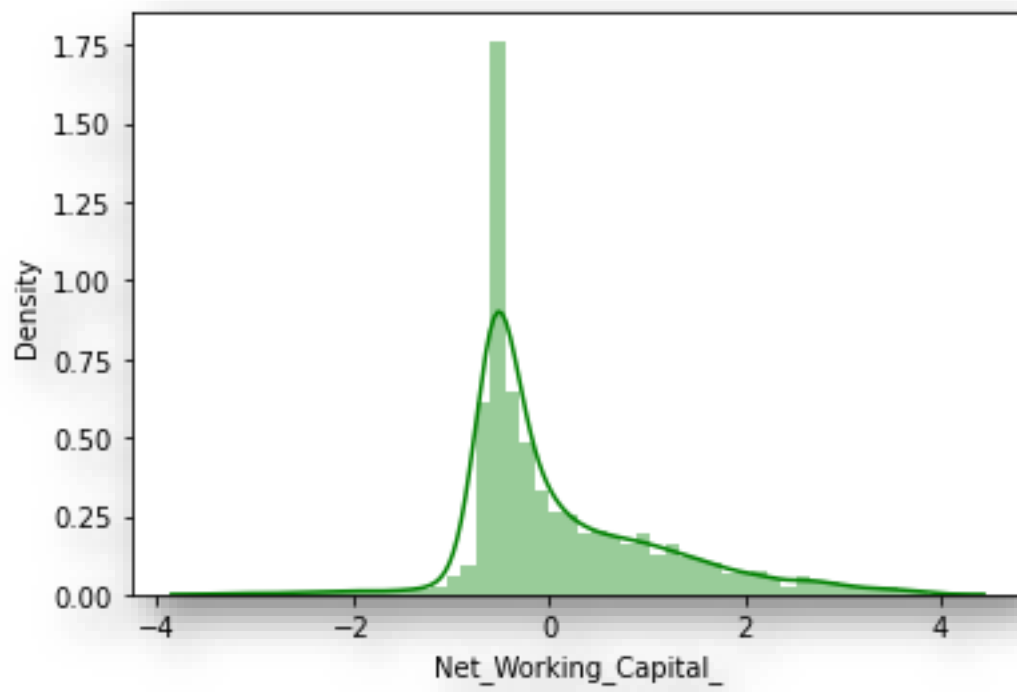
## TABLE 5: PROPORTION OF DEFAULT

```
0    0.905
1    0.095
Name: default, dtype: float64
```

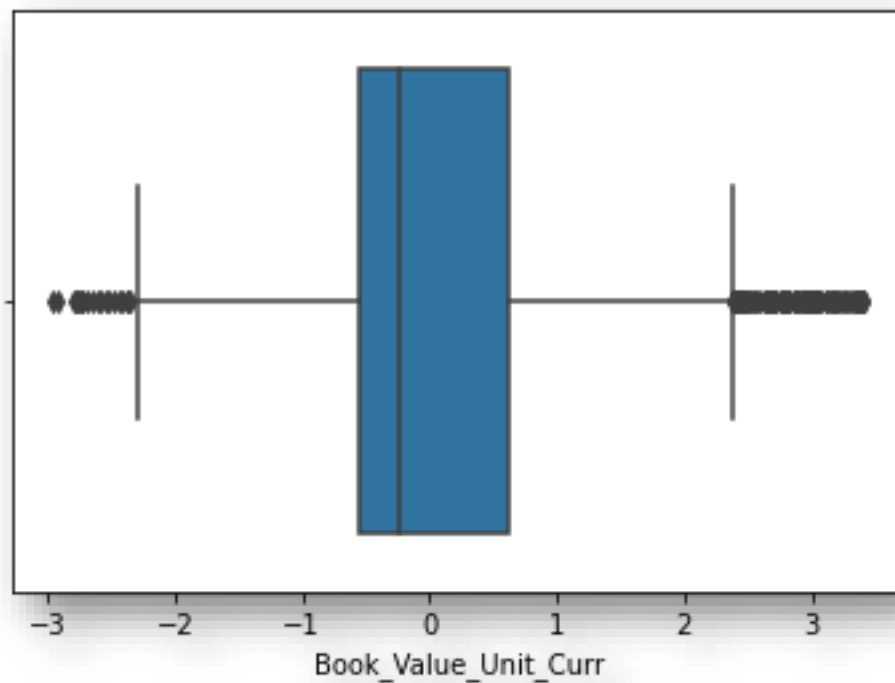
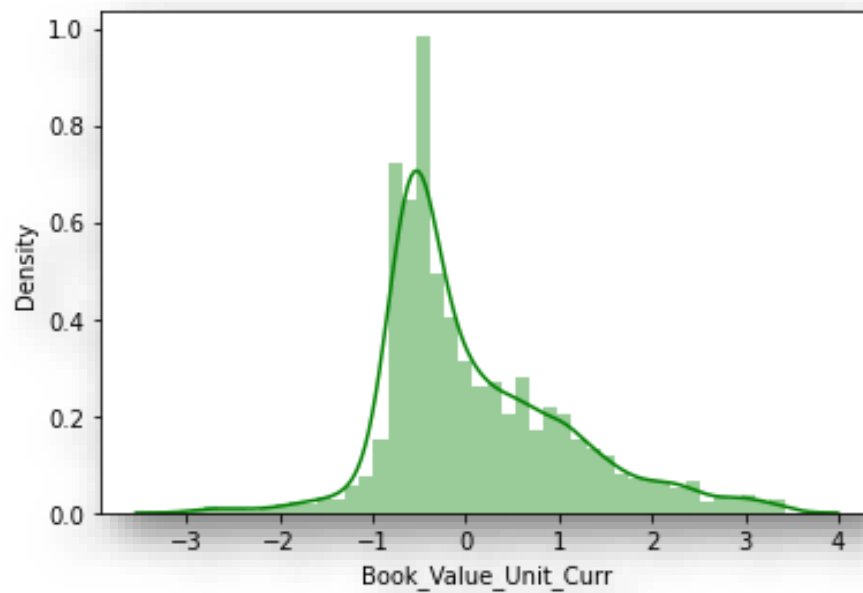
- Noticeably, approximately 10% of the companies from the dataset are going to default and these are the companies in which the companies should not or avoid investing.

## Q.1.4. UNIVARIATE AND BIVARIATE ANALYSIS

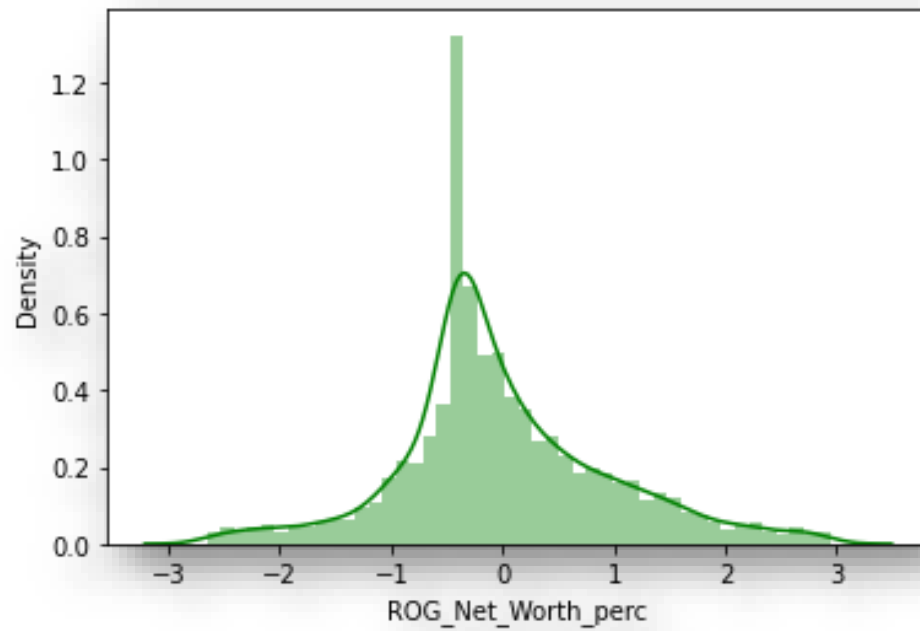
## FIGURE 4: DISTRIBUTION & BOXPLOT OF NET WORKING\_CAPITAL\_

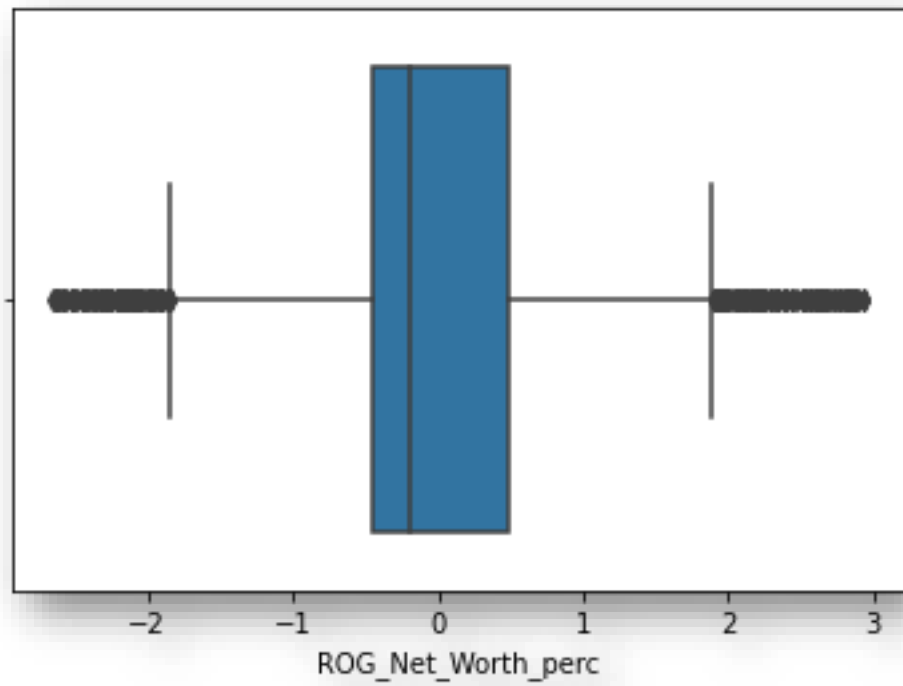


**FIGURE 5: DISTRIBUTION & BOXPLOT OF BOOK  
VALUE UNIT CURR**

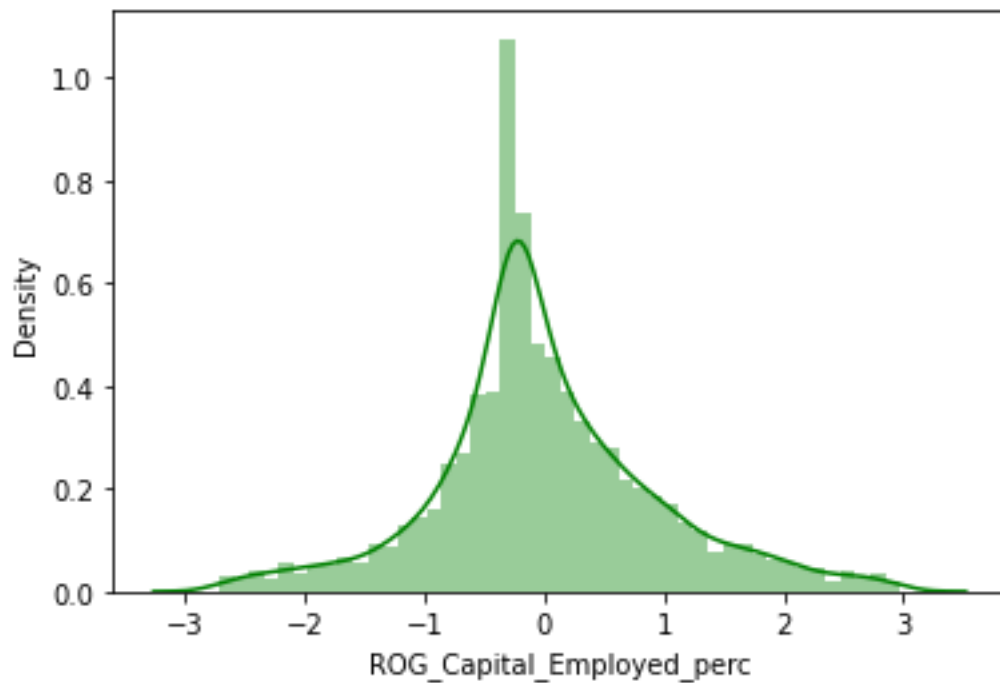


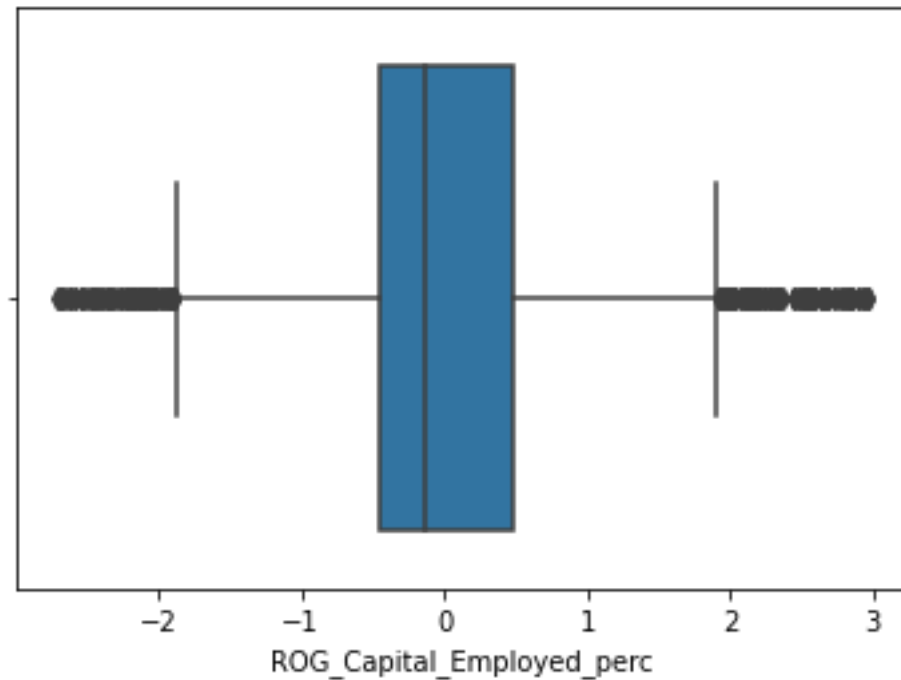
**FIGURE 6: DISTRIBUTION & BOXPLOT OF ROG  
NET WORTH PERC**



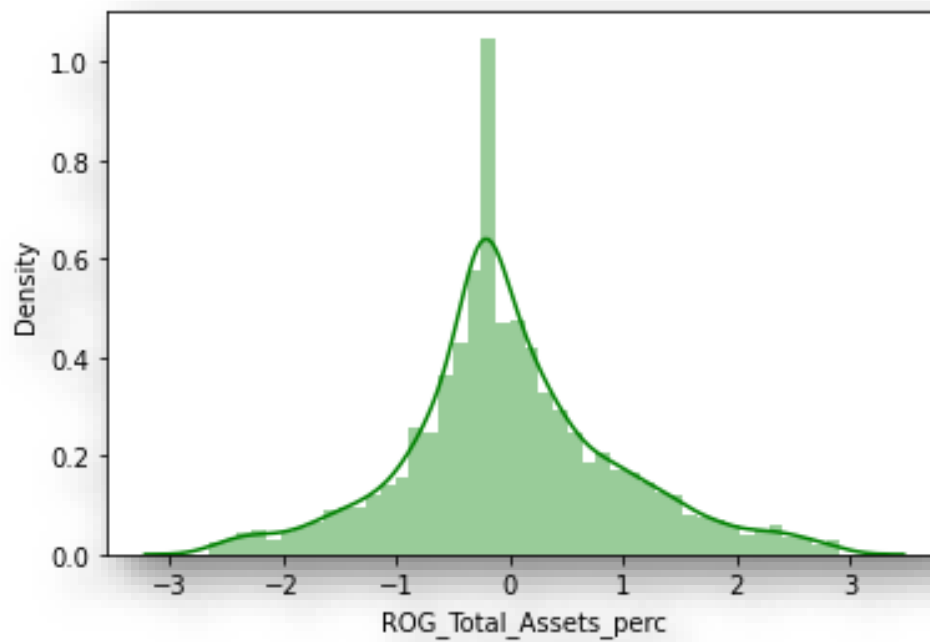


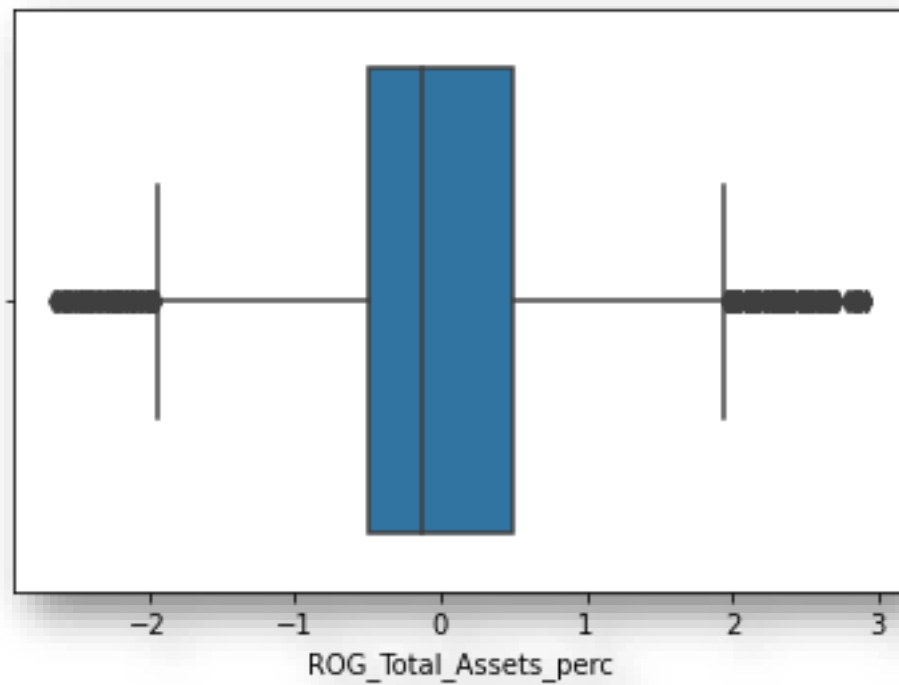
**FIGURE 7: DISTRIBUTION & BOXPLOT OF ROG CAPITAL EMPLOYED PERC**



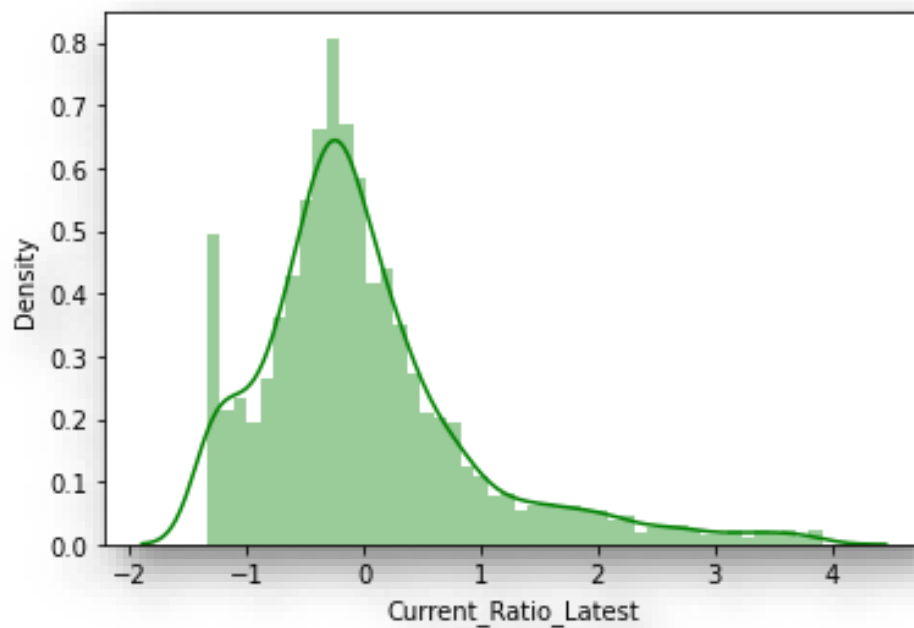


**FIGURE 8: DISTRIBUTION & BOXPLOT OF ROG  
TOTAL ASSETS PERC**

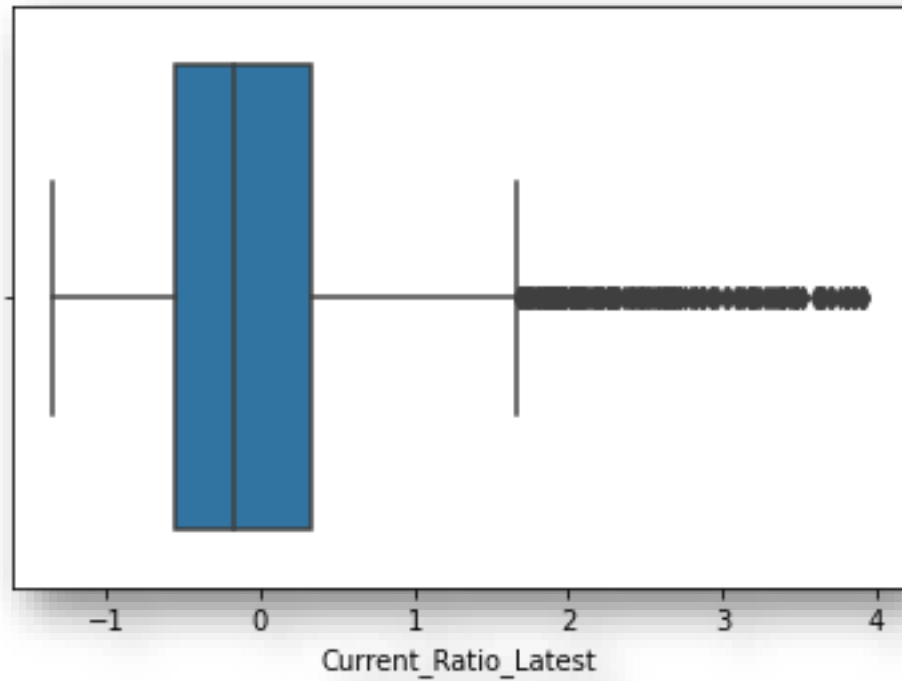




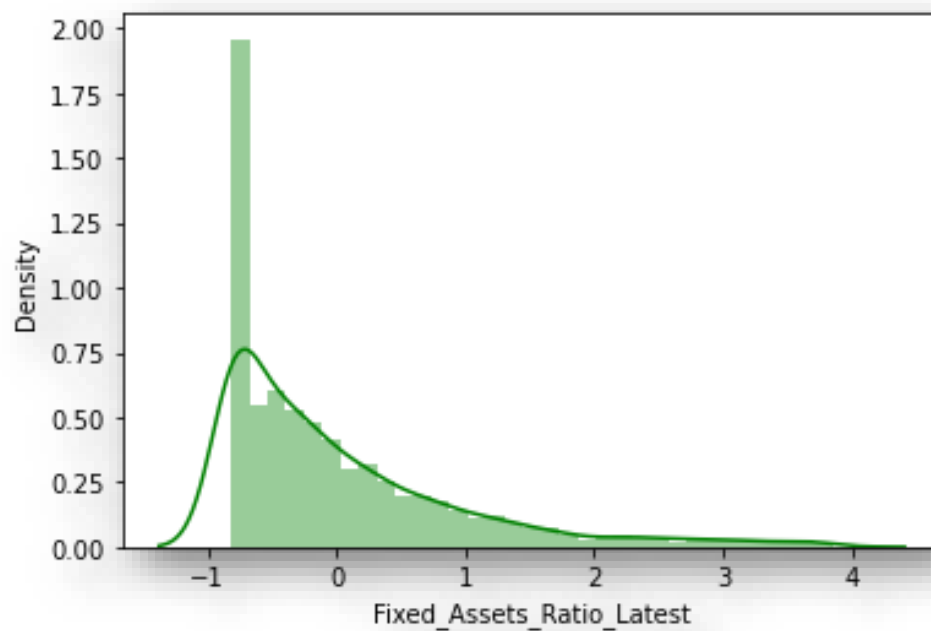
**FIGURE 9: DISTRIBUTION & BOXPLOT OF  
CURRENT RATIO LATEST**

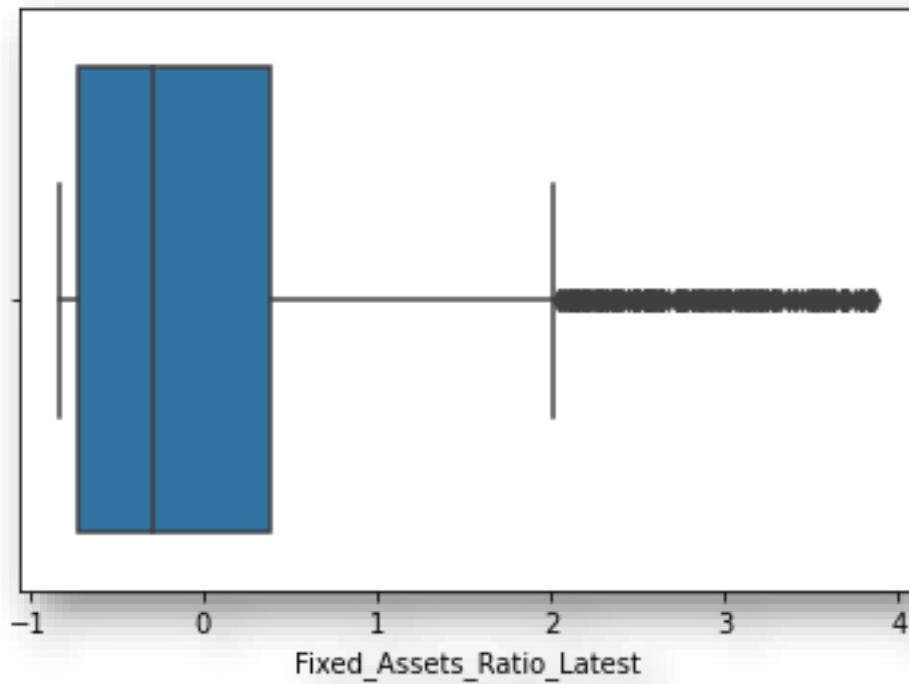




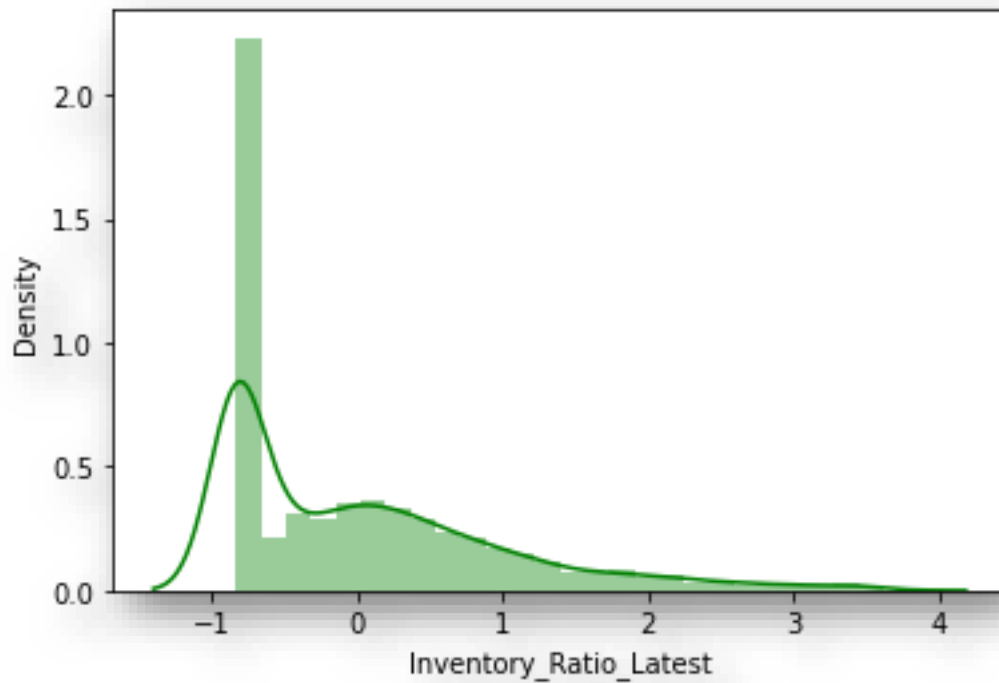


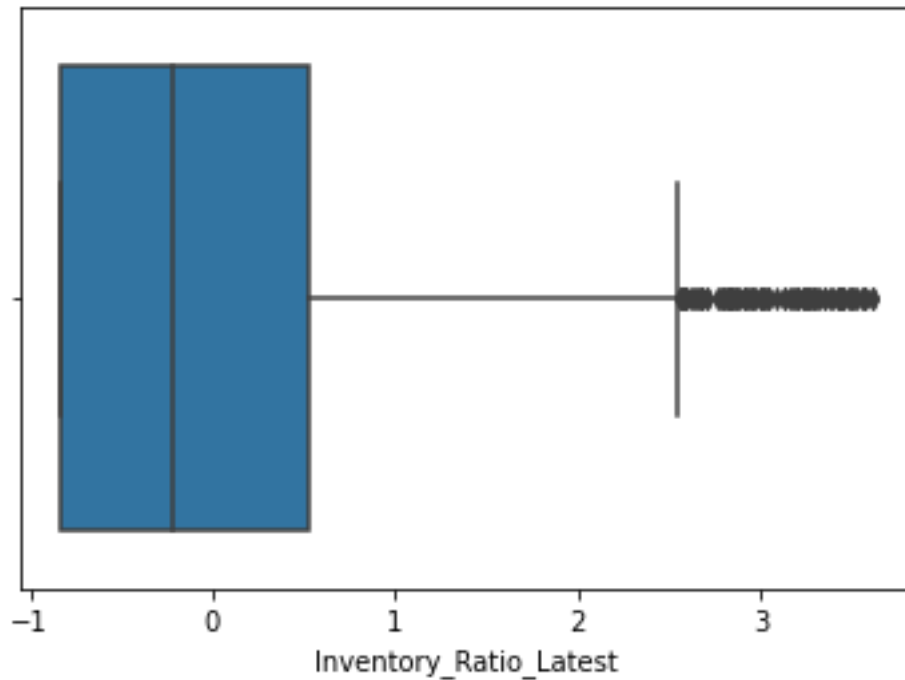
**FIGURE 10: DISTRIBUTION & BOXPLOT OF  
FIXED ASSET RATIO LATEST**



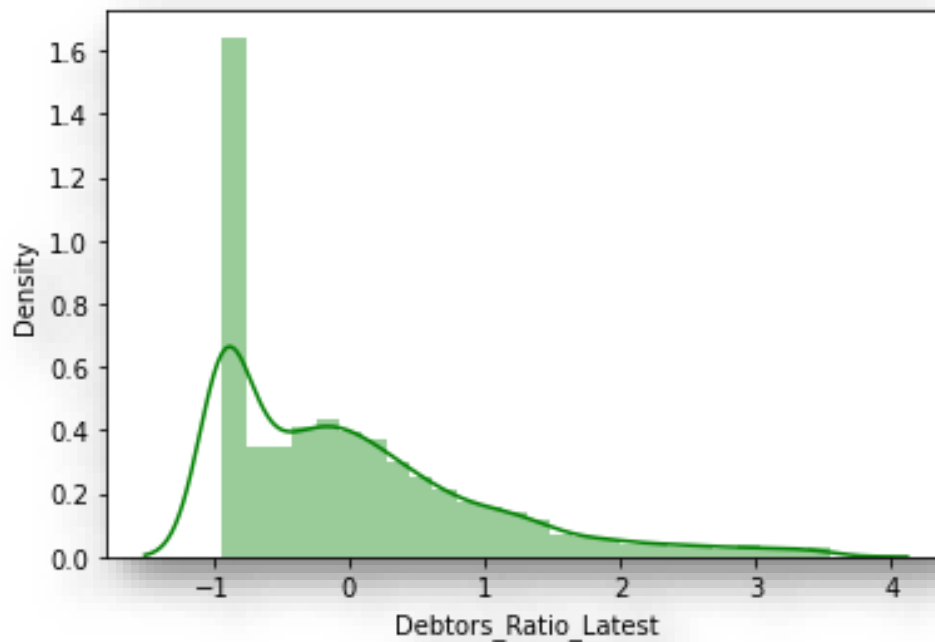


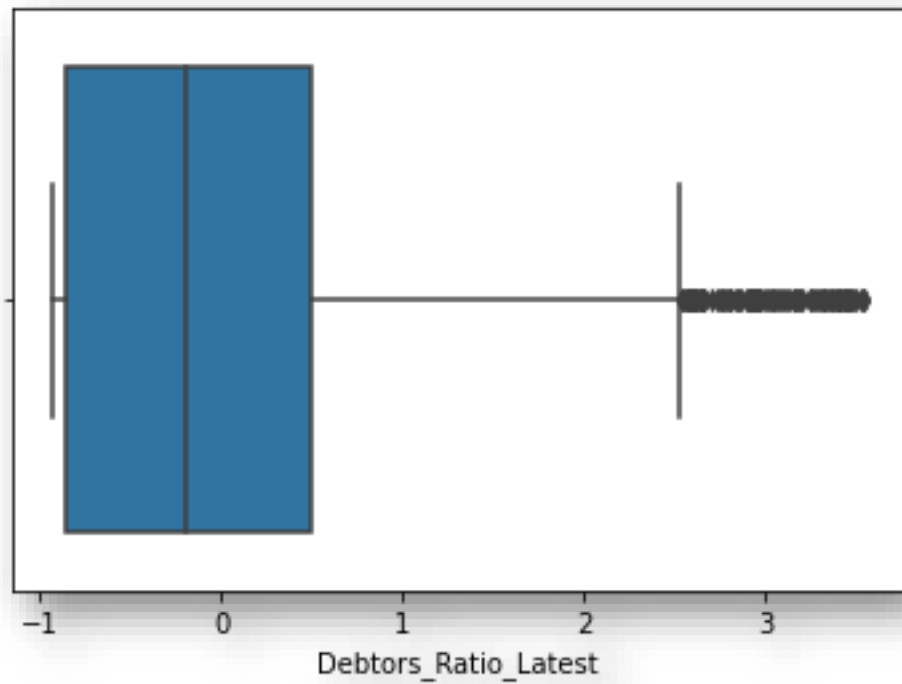
**FIGURE 11: DISTRIBUTION & BOXPLOT OF INVENTORY RATIO LATEST**



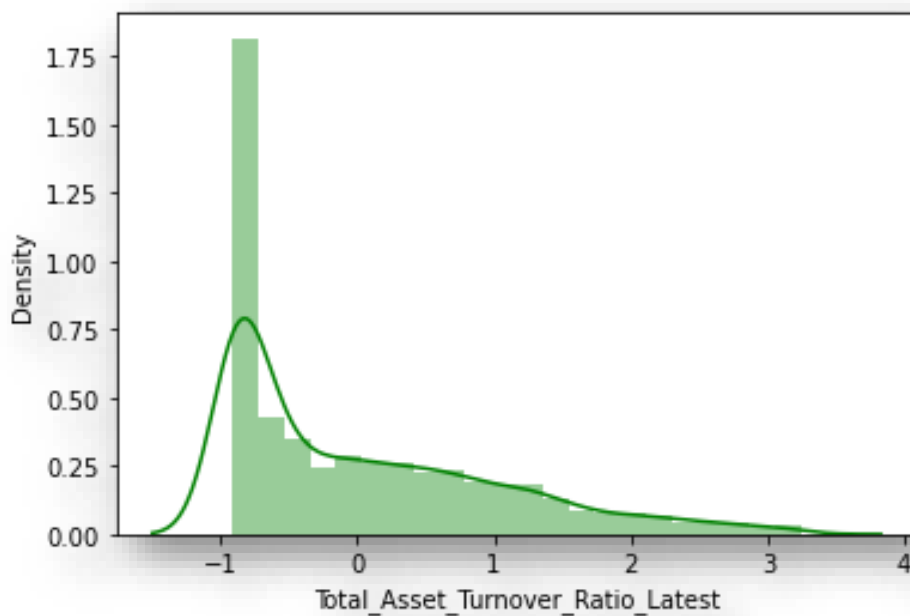


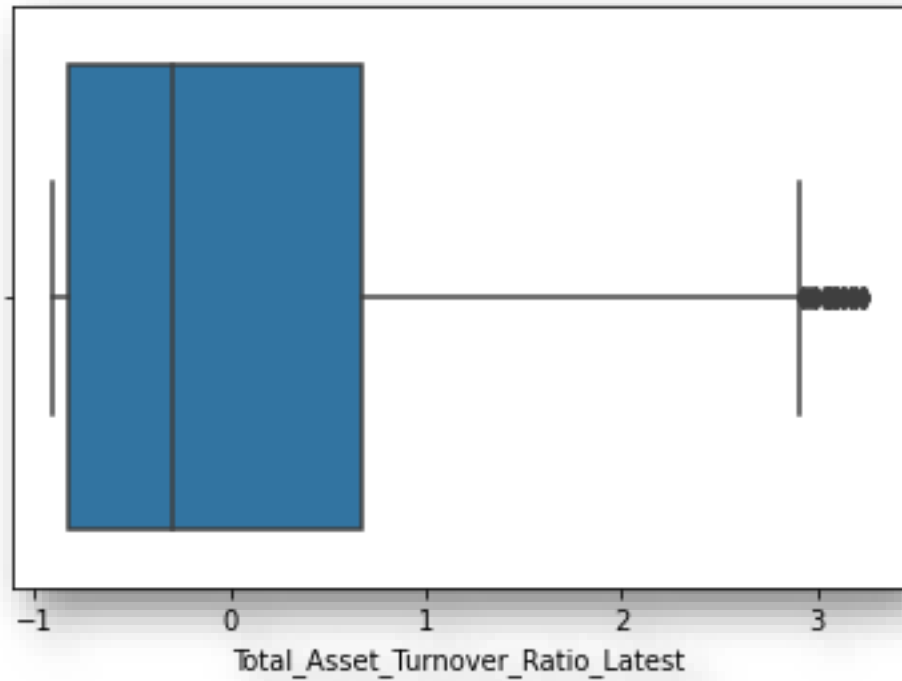
**FIGURE 12: DISTRIBUTION & BOXPLOT OF DEBTORS RATIO LATEST**



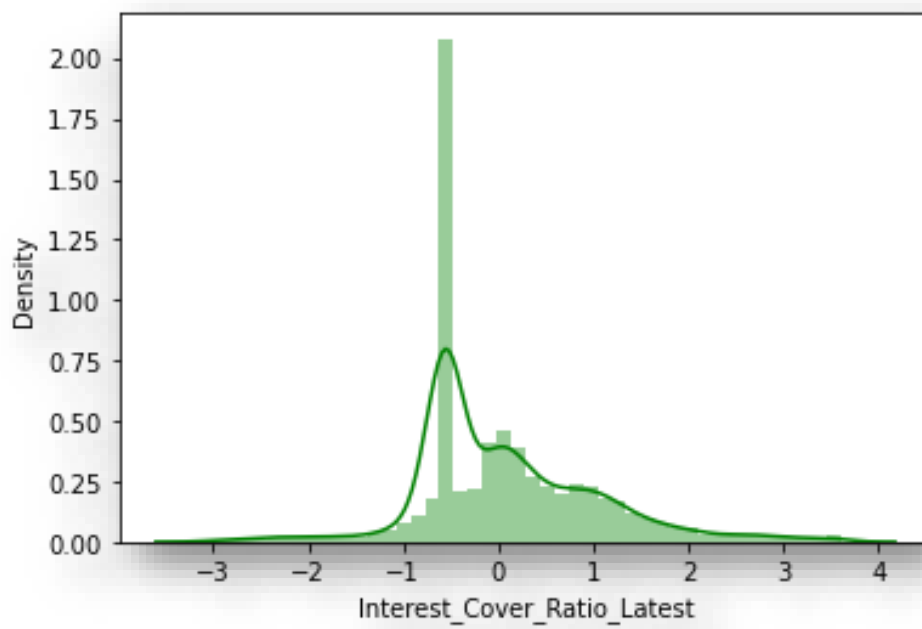


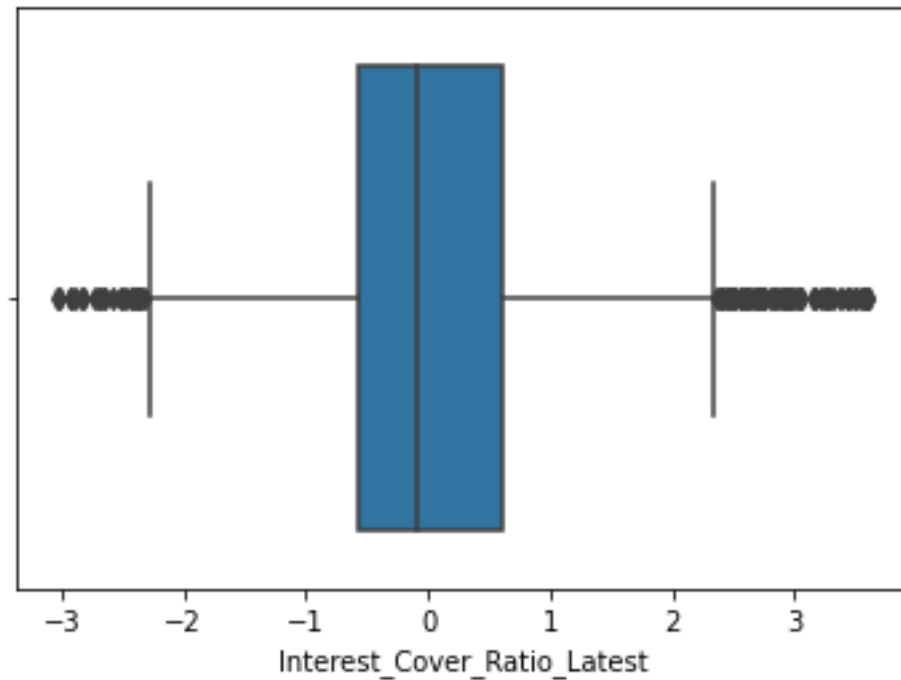
**FIGURE 13: DISTRIBUTION & BOXPLOT OF TOTAL ASSET TURNOVER RATIO LATEST**



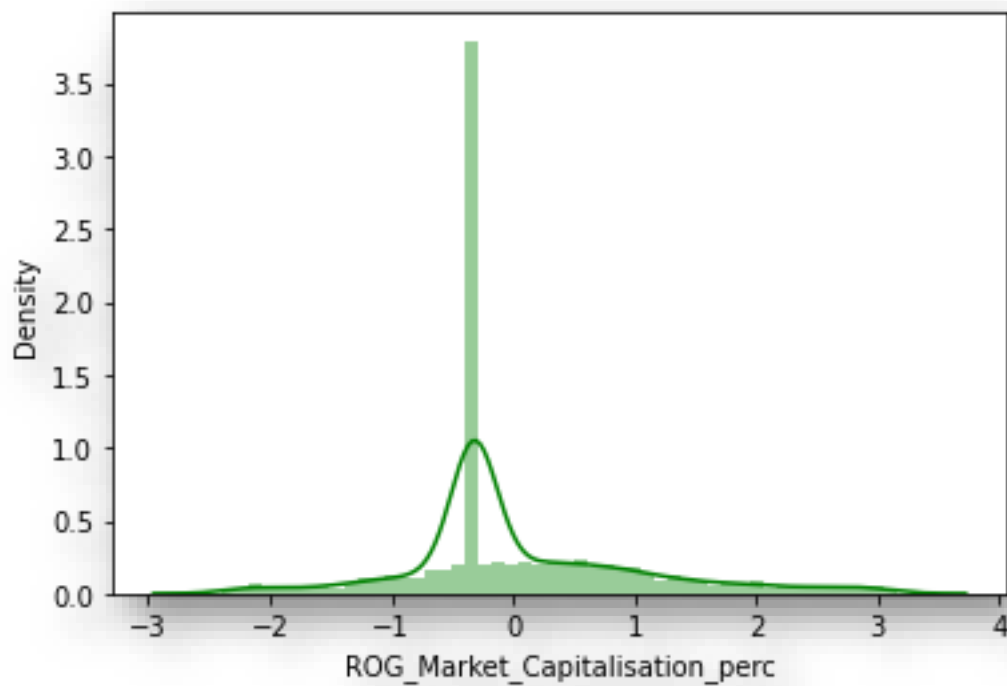


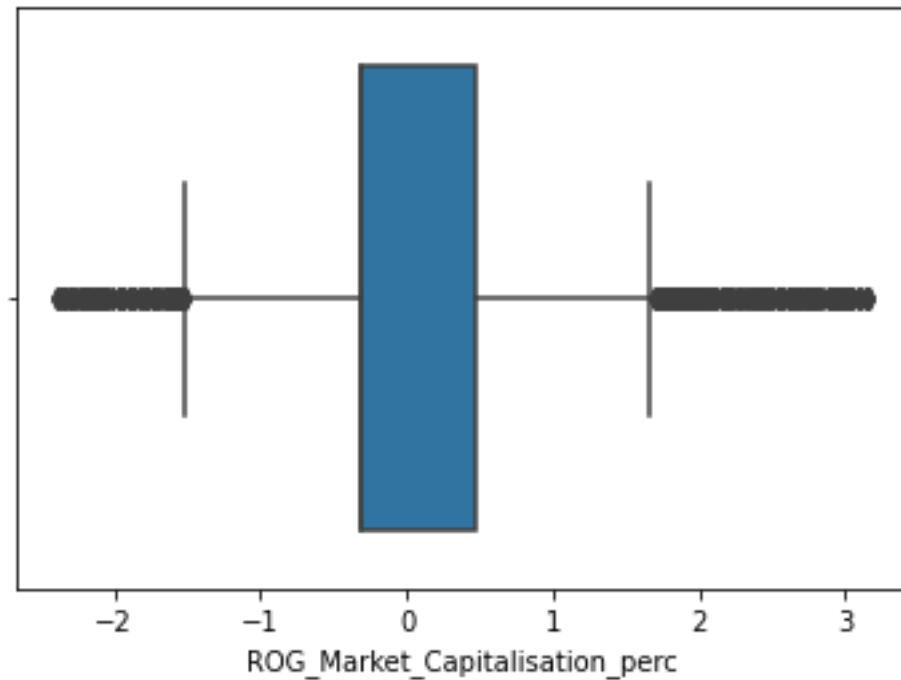
**FIGURE 14: DISTRIBUTION & BOXPLOT OF INTEREST COVER RATIO LATEST**



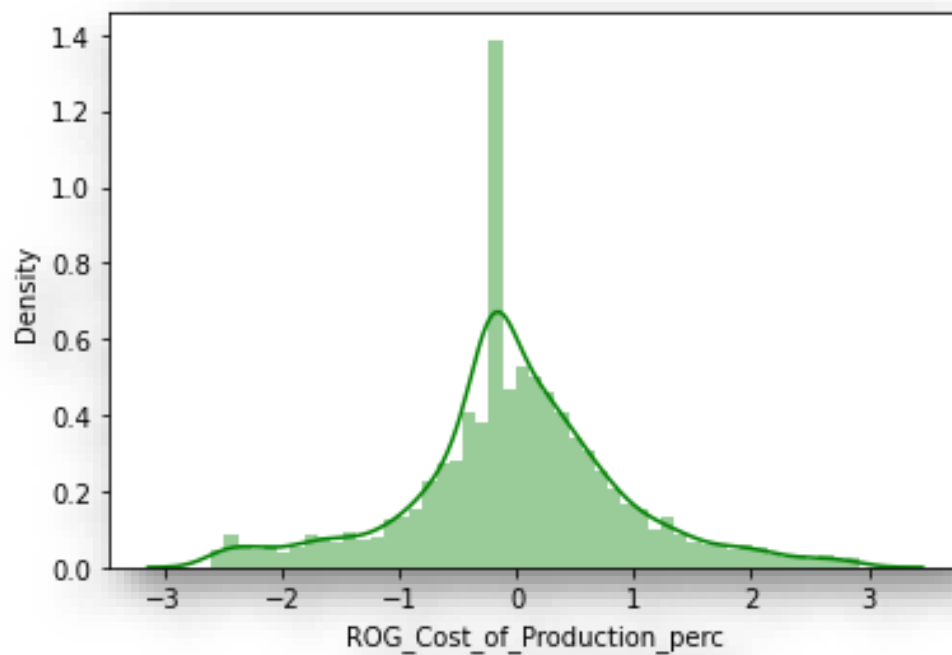


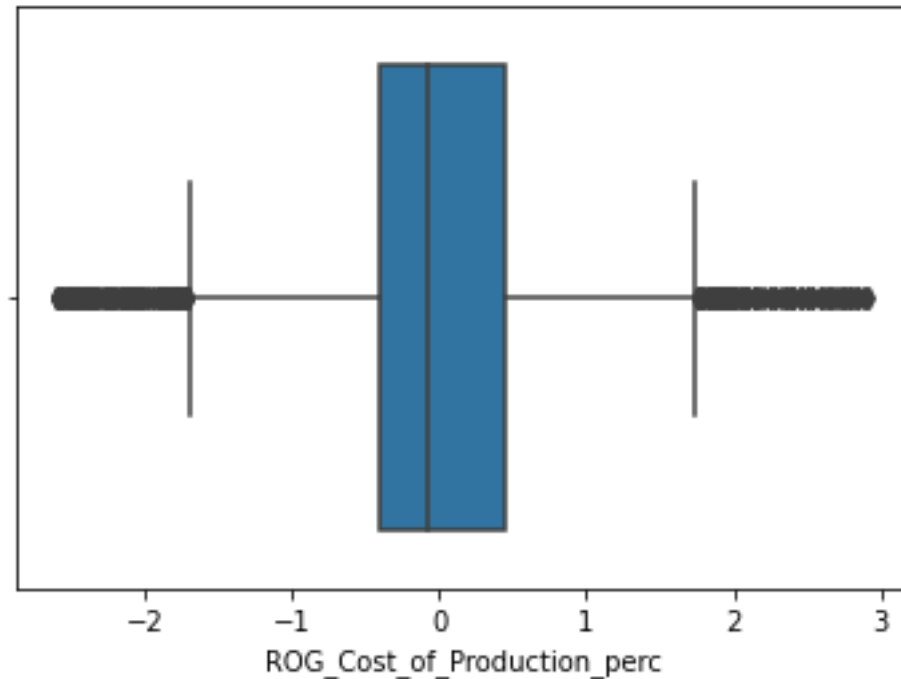
**FIGURE 15: DISTRIBUTION & BOXPLOT OF ROG MARKET CAPITALISATION PERC**





**FIGURE 16: DISTRIBUTION & BOXPLOT OF ROG COST OF PRODUCTION PERC**





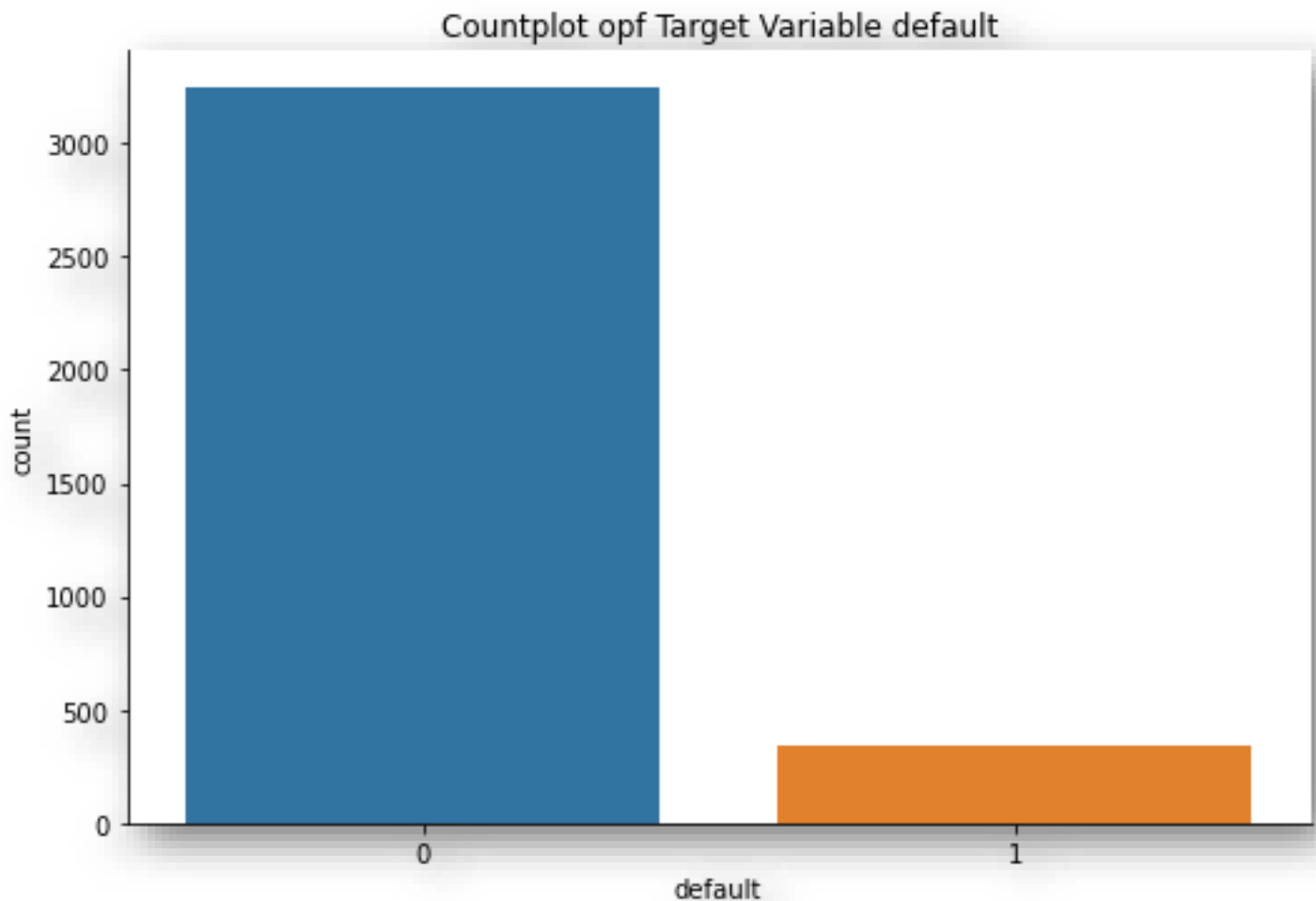
## INTERPRETATION OF THE ABOVE DISTRIBUTION AND BOXPLOT

1. Net working capital: The distribution plot is not normally distributed; it is not symmetrical. It is negative for a period of time and reaches maximum and is positive throughout, the boxplot has outliers and the both in the low and high whiskers most of it is negative. The outliers have been treated.
2. Book value unit curr: it is not distributed normally and is right skewed and has outliers which are treated above and most of them are around the mean.
3. ROG Net worth perc, ROG capital employed, ROG total assets perc is normally distributed and has outliers around both the whiskers.
4. Current Ratio, Fixed asset ratio latest, Inventory Ratio Latest, Debtors Ratio Latest, Total asset Turnover ratio Latest are right skewed and not distributed normally and have boxplot at there higher whisker, treated later.
5. ROG market capitalization perc is normally distributed and has outliers.
6. Interest cover ratio latest is not distributed normally and has outliers.
7. ROG cost of production is left skewed and has outliers.

NOTE: The outliers have been treated in the earlier question.



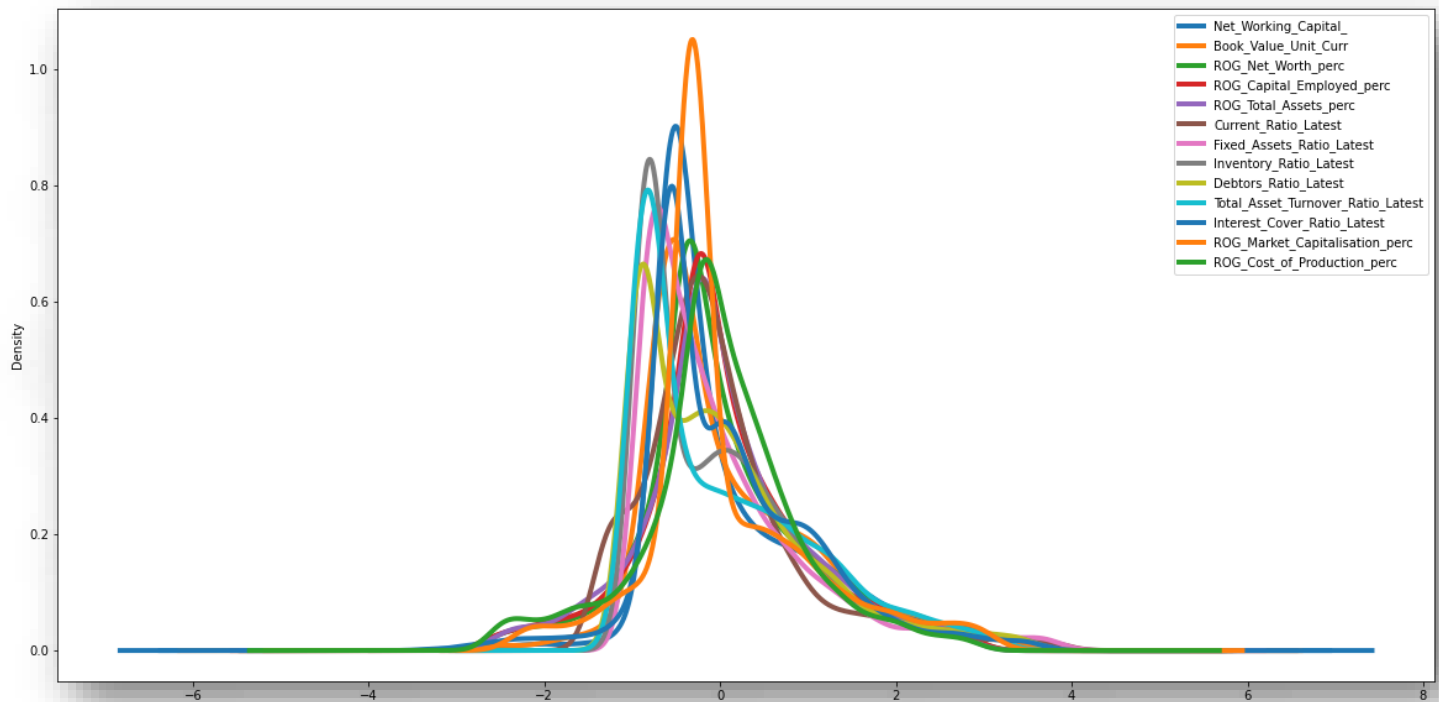
**FIGURE 17: COUNTPLOT OF TARGET VARIABLE  
AND DEFAULT**



- The data has higher companies which are expected to have positive Net Worth Next Year, which in turn is good for investors for decision making.
- The following are the variables that are more likely contribute to the strength of the company's balance sheet:
  1. Net Working Capital: It measures company's liquidity and short-term financial health. A company will have negative NWC if its ratio of current assets to liabilities is less than one.
  2. Book Value Unit Curr: High book value per share (due to profits accumulated over the years) indicates a strong company.

3. ROG-Net Worth perc: Companies with low capital base (that don't need additional capital for growth) will show a higher ratio.
4. ROG-Capital Employed perc: Captures the profit generated on total capital employed (including debt). Companies with low capital base (those that don't need additional capital for growth) will display a higher ratio.
5. ROG-Total Assets perc: Captures the net profit generated on total assets.
6. Current Ratio Latest: It tells how cash rich a company is. It helps us gauge the short-term financial strength of a company.
7. Fixed Assets Ratio Latest: It reveals how efficient a company is at generating sales from its existing fixed assets.
8. Inventory Ratio Latest: Shows how efficiently the company manages its inventory.
9. Debtors Ratio Latest: A high debt to equity ratio is a warning signal, especially in situations like business downturns.
10. Total Asset Turnover Ratio Latest: Shows how efficiently the company manages its total assets.
11. Interest Cover Ratio Latest: measures a company's ability to handle its outstanding debt.
12. ROG-Market Capitalization perc: Company's worth as determined by the stock market.
13. ROG Cost of Production perc: Product costing is the process of tracking and studying all the various expenses that are accrued in the production and sale of a product.

## FIGURE 18: DISTRIBUTION PLOT OF IMPORTANT FEATURES



## TABLE 6: SKEWNESS OF IMPORTANT FEATURES

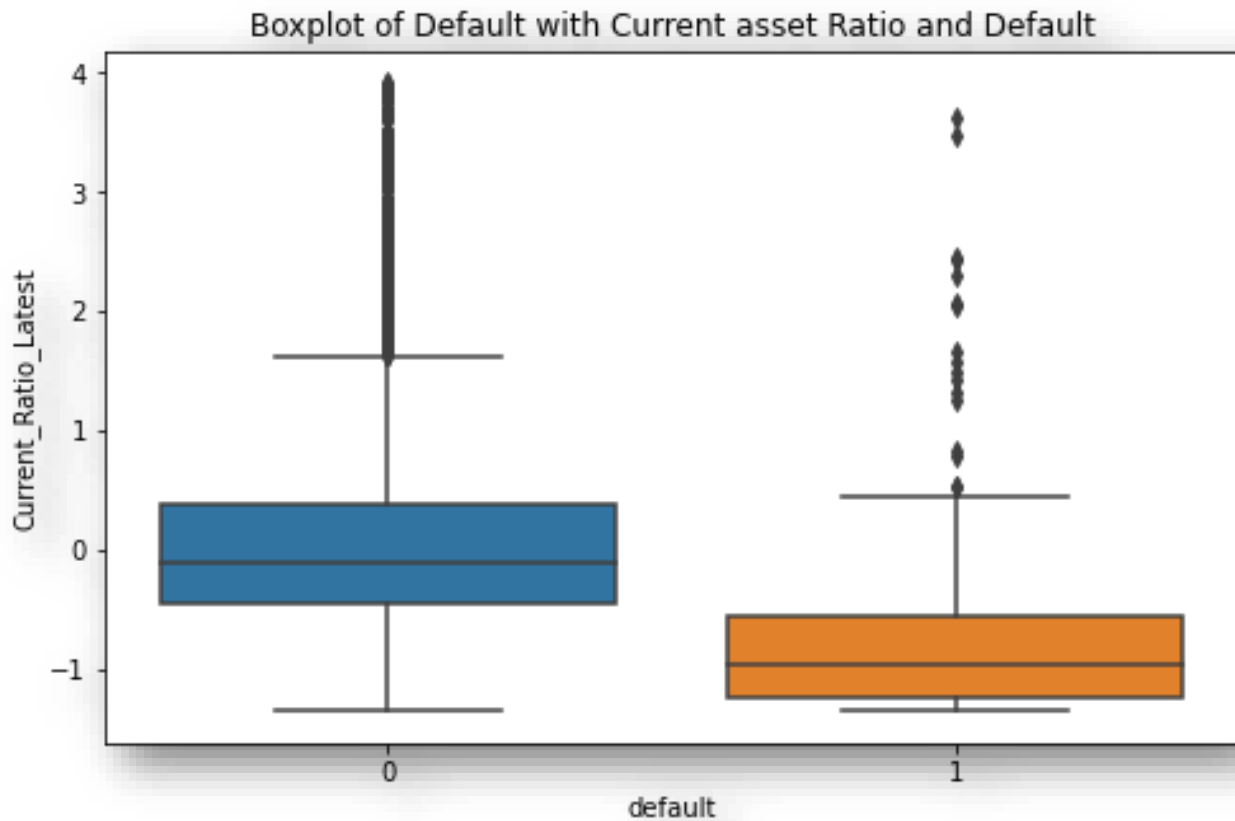
ROG_Cost_of_Production_perc	-0.007
ROG_Total_Assets_perc	0.280
ROG_Capital_Employed_perc	0.281
ROG_Net_Worth_perc	0.403
Interest_Cover_Ratio_Latest	0.762
ROG_Market_Capitalisation_perc	0.863
Book_Value_Unit_Curr	0.907
Total_Asset_Turnover_Ratio_Latest	1.031
Net_Working_Capital_	1.152
Debtors_Ratio_Latest	1.245
Inventory_Ratio_Latest	1.264
Current_Ratio_Latest	1.435
Fixed_Assets_Ratio_Latest	1.658
dtype: float64	

## INTERPRETATION OF THE DISTRIBUTION PLOT SKEWNESS OF IMPORTANT FEATURES

- Since, skewness is more than 1 indicating distribution is highly skewed for variables "Current Ratio Latest" and "Total Asset Turnover Ratio Latest"
- Data is moderately skewed for variables "Fixed \_Assets \_Ratio Latest", "ROG-Market Capitalization perc" and "Interest Cover Ratio Latest "
- Other variables look fairly symmetrical
- Note: The rule of thumb is:

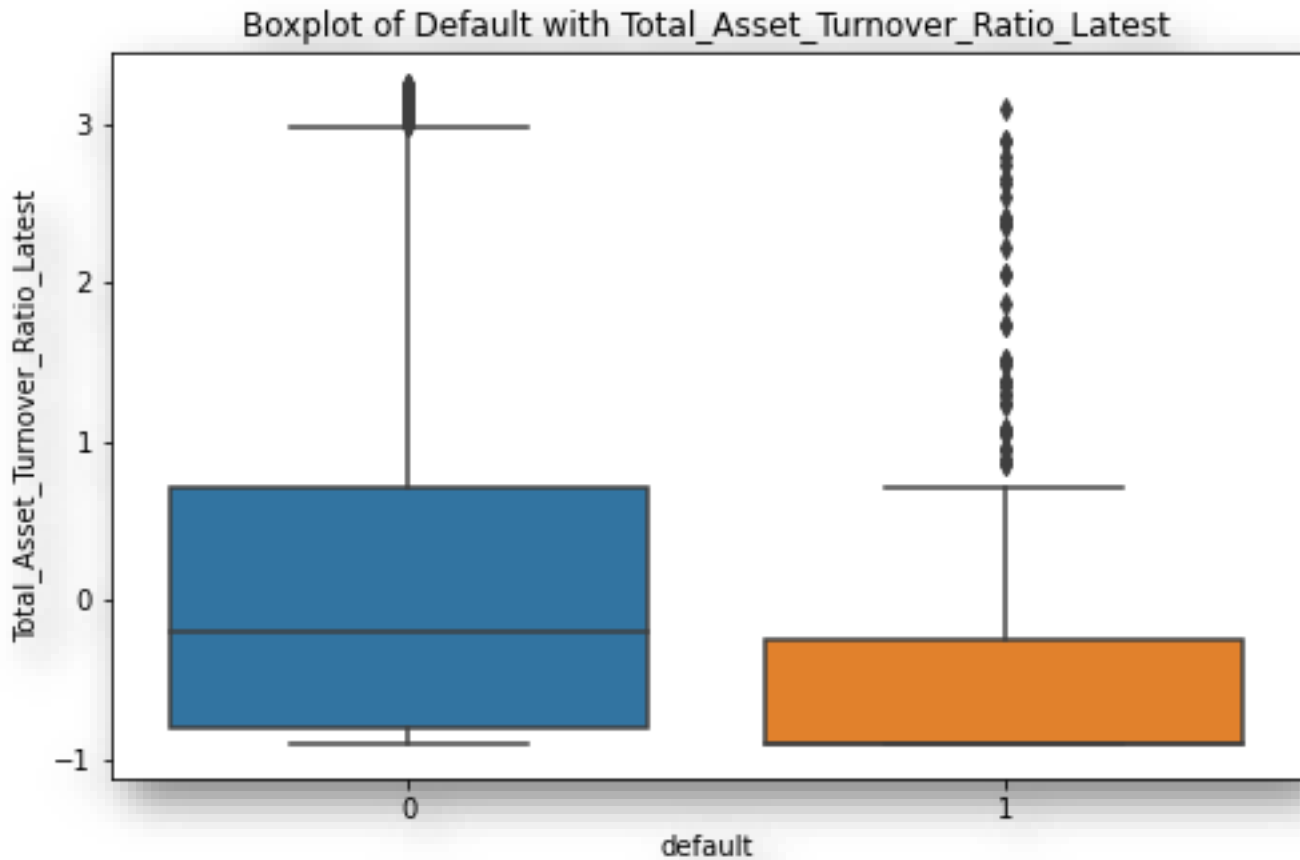
If the skewness is between -0.5 and 0.5, the data are fairly symmetrical. • If the skewness is between -1 and - 0.5 or between 0.5 and 1, the data are moderately skewed. • If the skewness is less than -1 or greater than 1, the data are highly skewed.

## FIGURE 19: PLOT OF DEFAULT WITH CURRENT RATIO



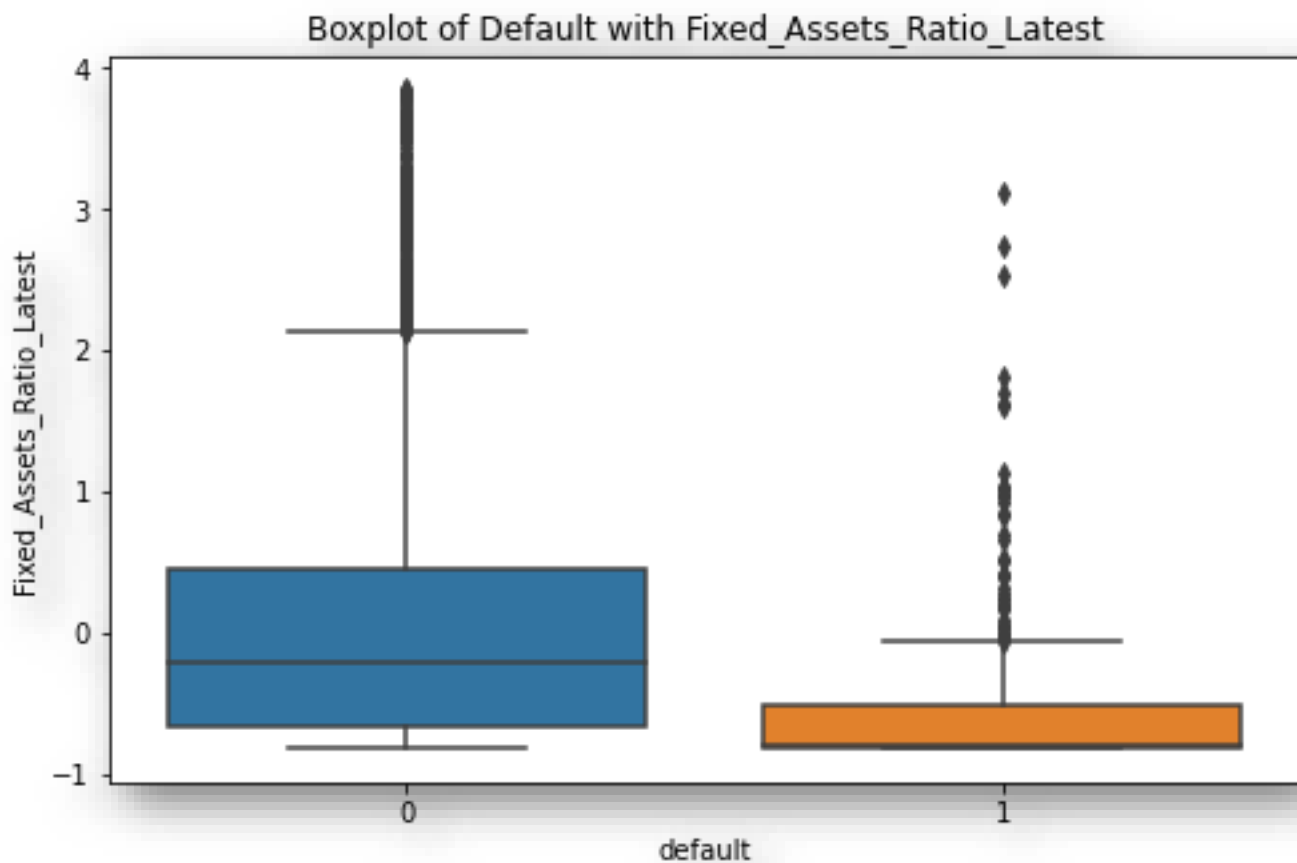
- The current asset ratio defines the link between itself and the current liabilities. The current assets are higher for non-defaulters when compared to the defaulters, the defaulters have a low current asset ratio, hence there are majorly higher non-defaulters and among them the current asset ratio is higher for the non-defaulters.

**FIGURE 20: PLOT OF DEFAULT WITH TOTAL ASSET TURNOVER RATIO**



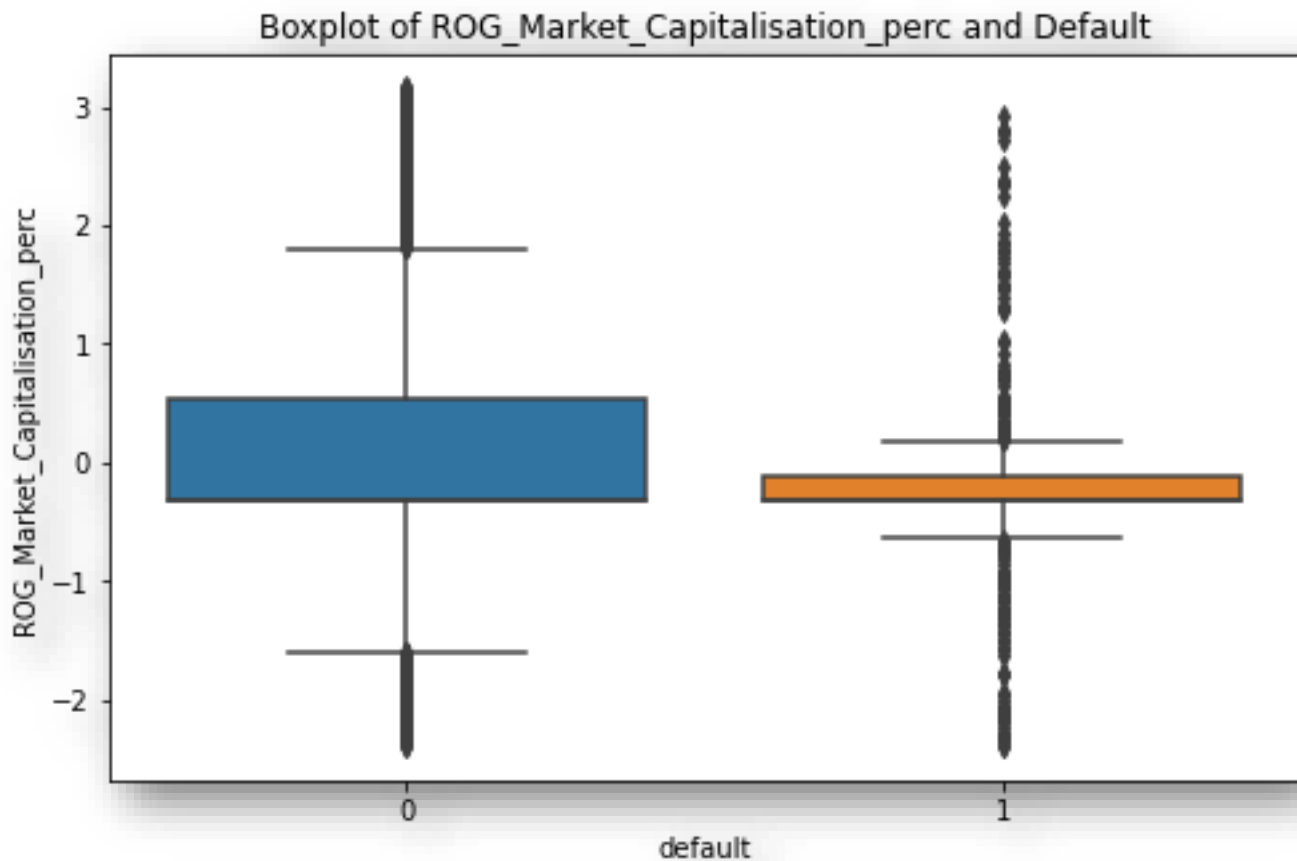
- The defaulter's total asset turnover is very low comparatively and the non-defaulter's total asset turnover ratio is higher but the median is negative. From this we get to know the companies (non-defaulting) are efficiently using or generating revenue from there total assets

**FIGURE 21: PLOT OF DEFAULT WITH FIXED ASSET TURNOVER RATIO**



- The non-defaulter has a higher Fixed asset turnover ratio compared to the defaulters. The defaulters have a negative fixed turnover ratio. A negative fixed asset turnover ratio means either the companies are over investing in assets or they are not able to generate revenue from the fixed assets they have invested in.

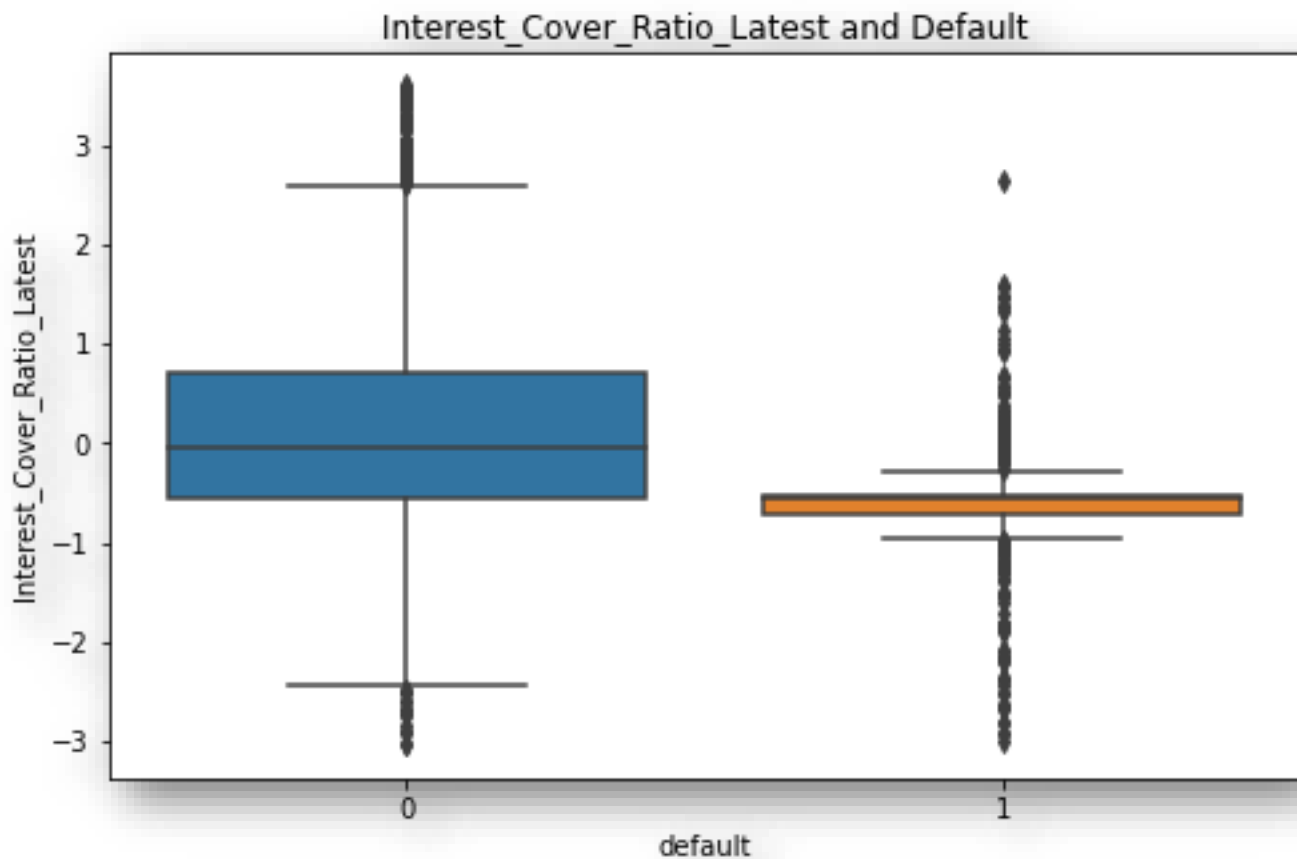
**FIGURE 22: PLOT OF DEFAULT WITH ROG  
MARKET CAPITALIZATION PERC**



- The defaulters and non-defaulters all most have the same level of market capitalization percentage, but the non-default companies have a higher edge compared to defaulters and thus their net worth for the next year is higher than defaulters.

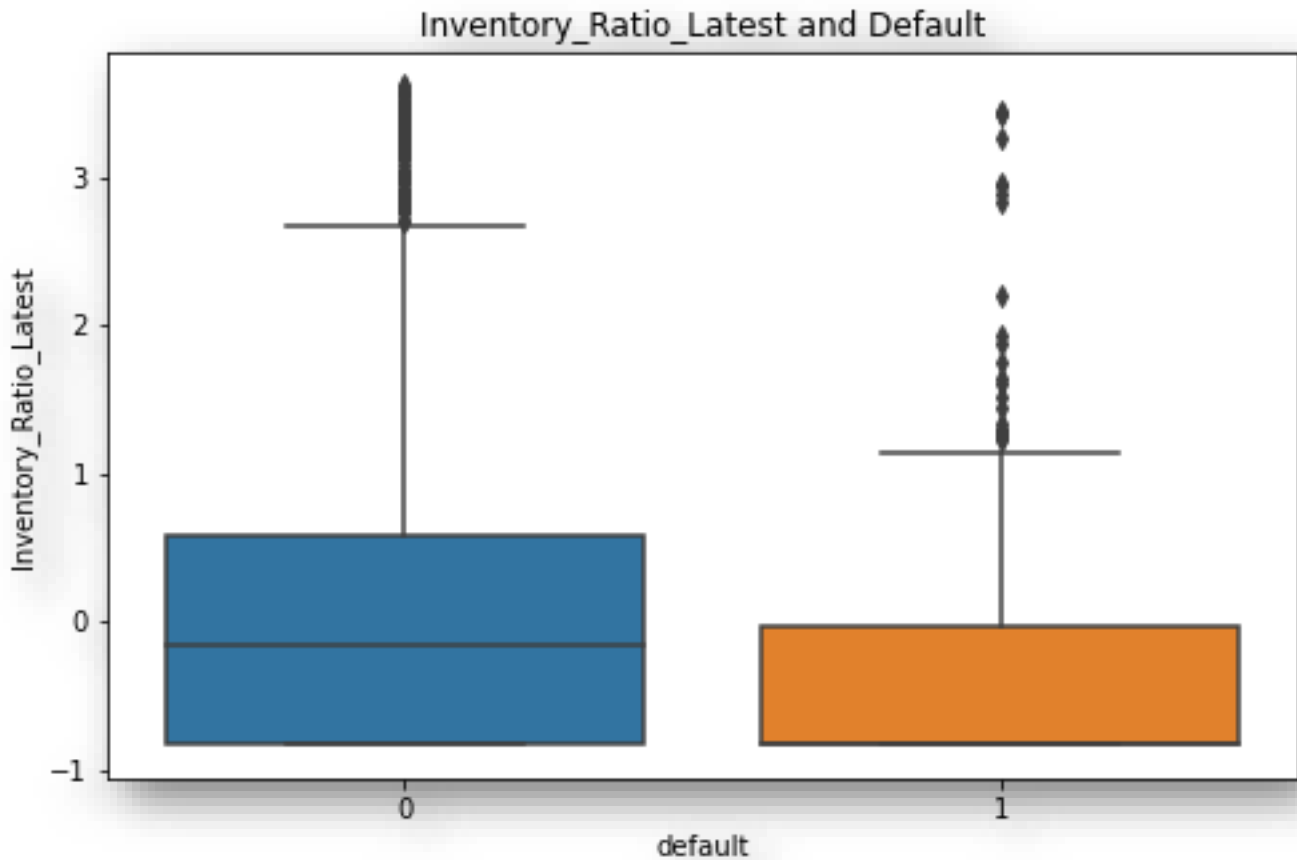


**FIGURE 23: PLOT OF DEFAULT WITH INTEREST COVER RATIO**



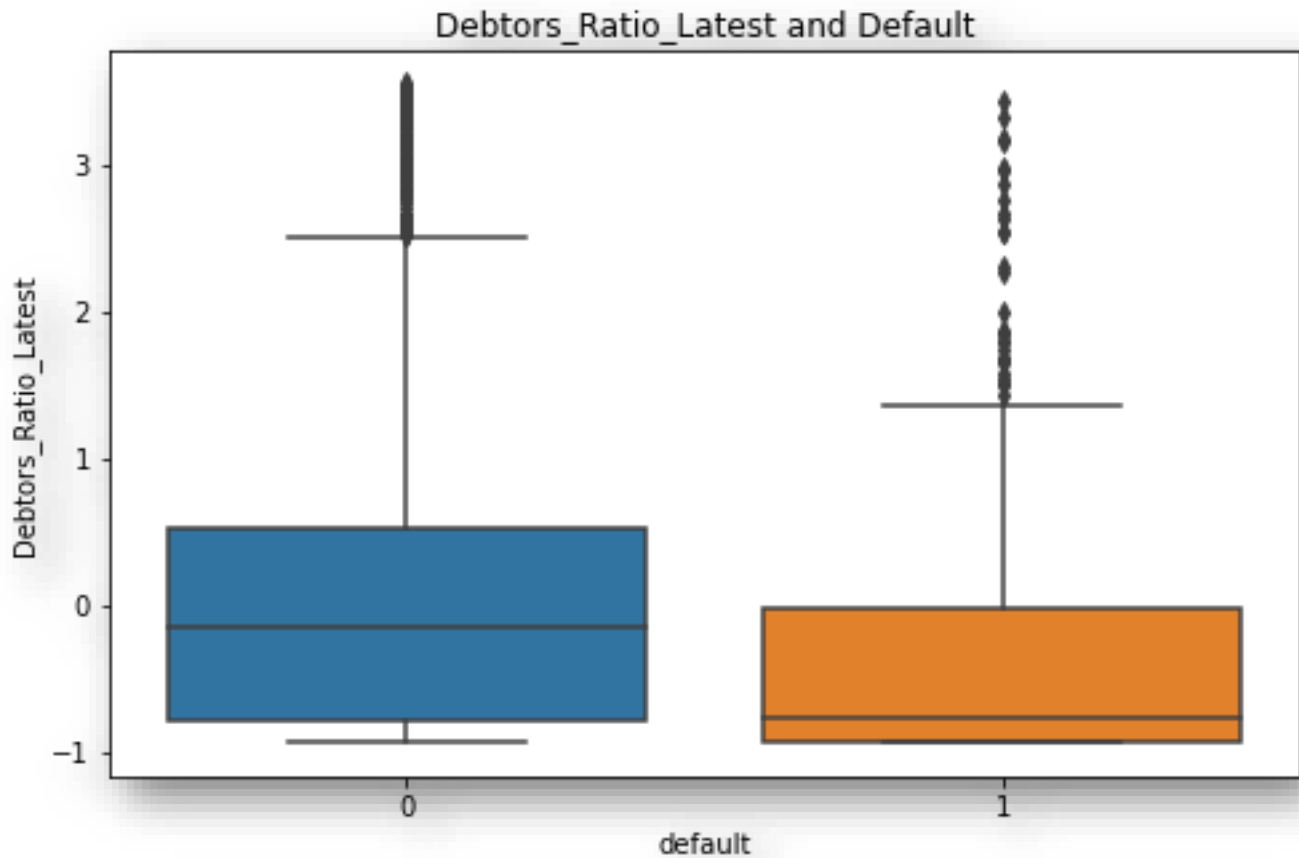
- The non-defaulter has higher interest and non-defaulter have a low interest coverage ratio, which means the non-defaulters have a higher EBIT and are able to pay their dues on time and in the case of defaulter it is opposite in fact negative.

**FIGURE 24: PLOT OF DEFAULT WITH INVENTORY RATIO**



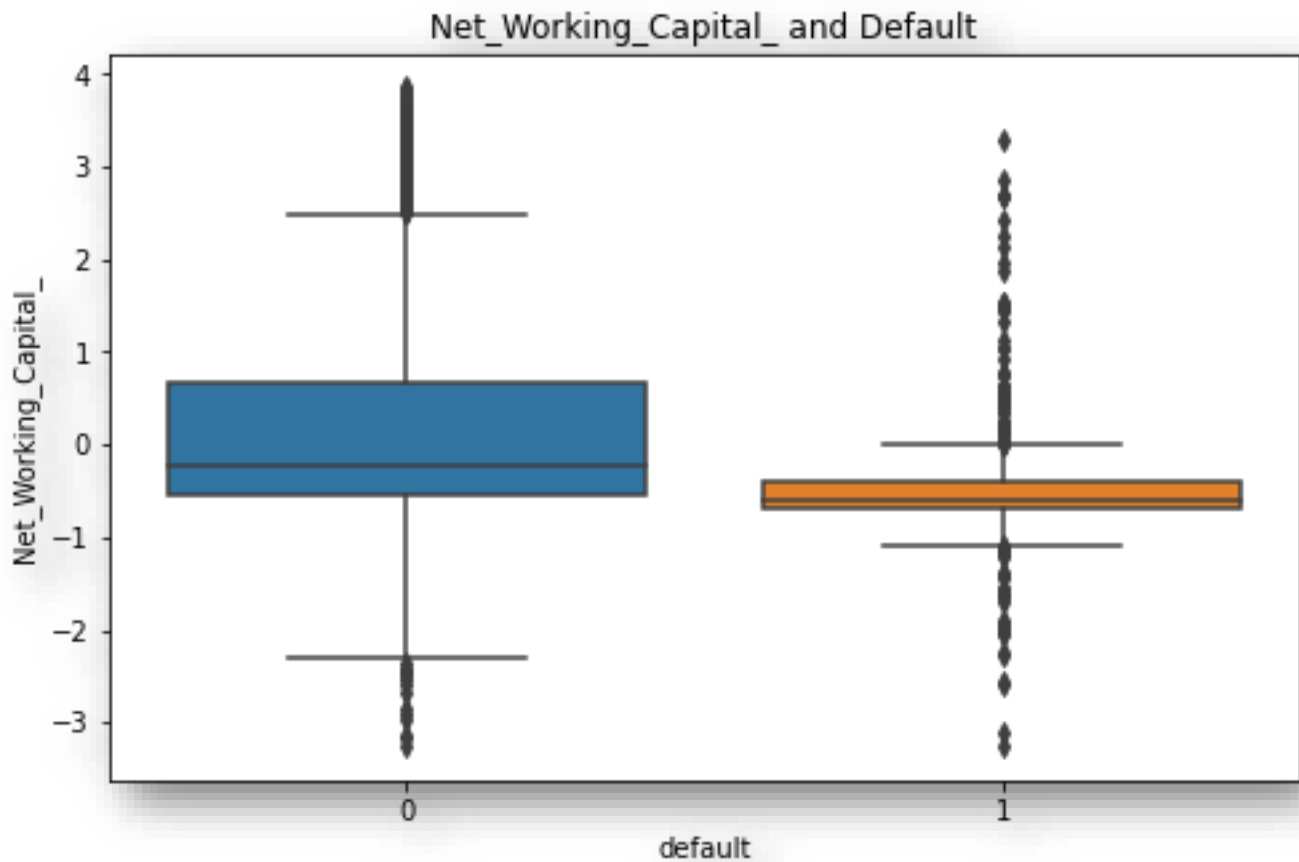
- The non-defaulter has a higher inventory ratio when compared to the non-defaulters, which means that the percentage of goods sold and replaced is higher, which means higher the sales. The non-defaulter will have a higher net worth for next year and help the investors in decision making.

**FIGURE 25: PLOT OF DEFAULT WITH DEBTORS RATIO**



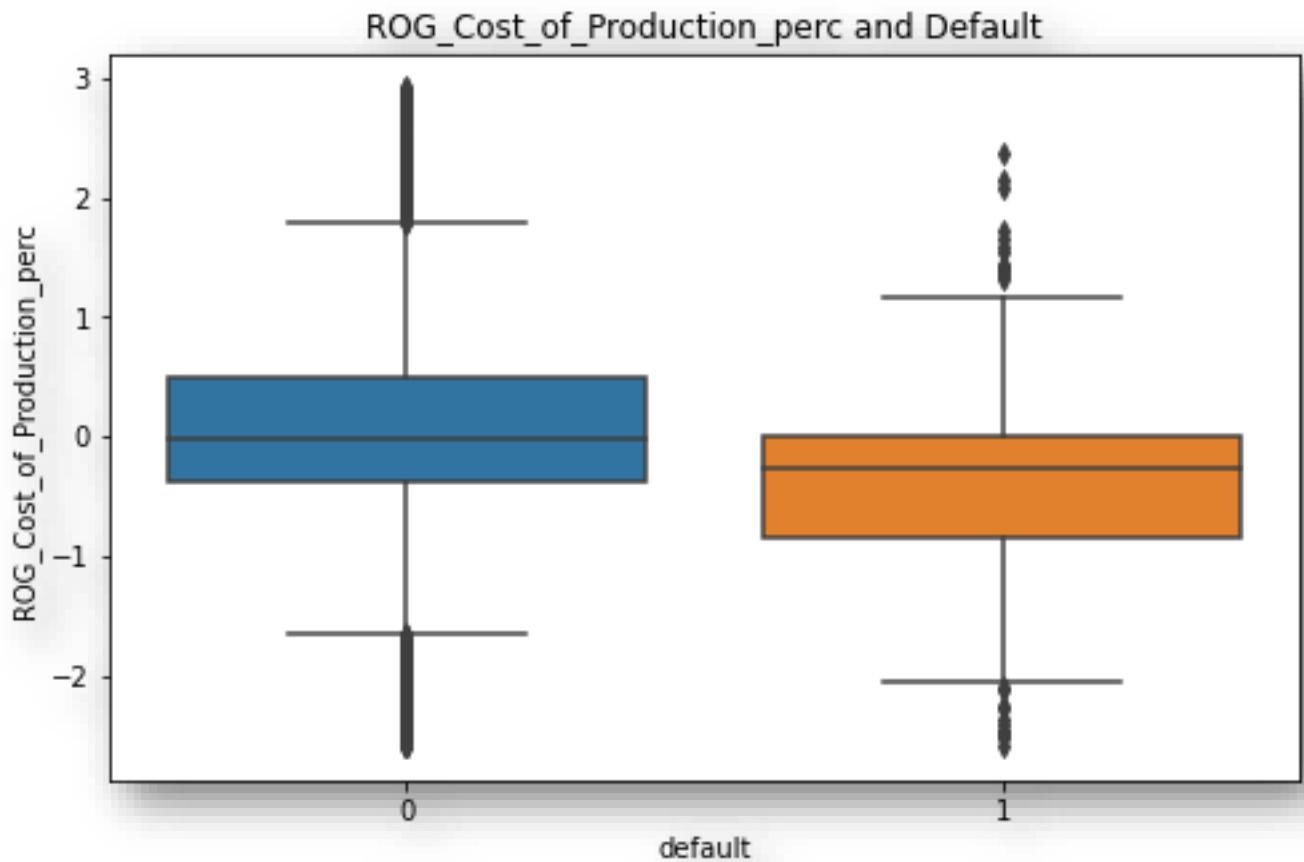
- The non-defaulters have a higher debtor's ratio comparatively which means that higher liquidity and more efficient is the management of the debtors. When comparing the debtors and the market capitalization, we can infer that the non-defaulters have a higher capitalized with the leverage taken and vice versa for defaulters.

**FIGURE 26: PLOT OF DEFAULT WITH NET WORKING CAPITAL**



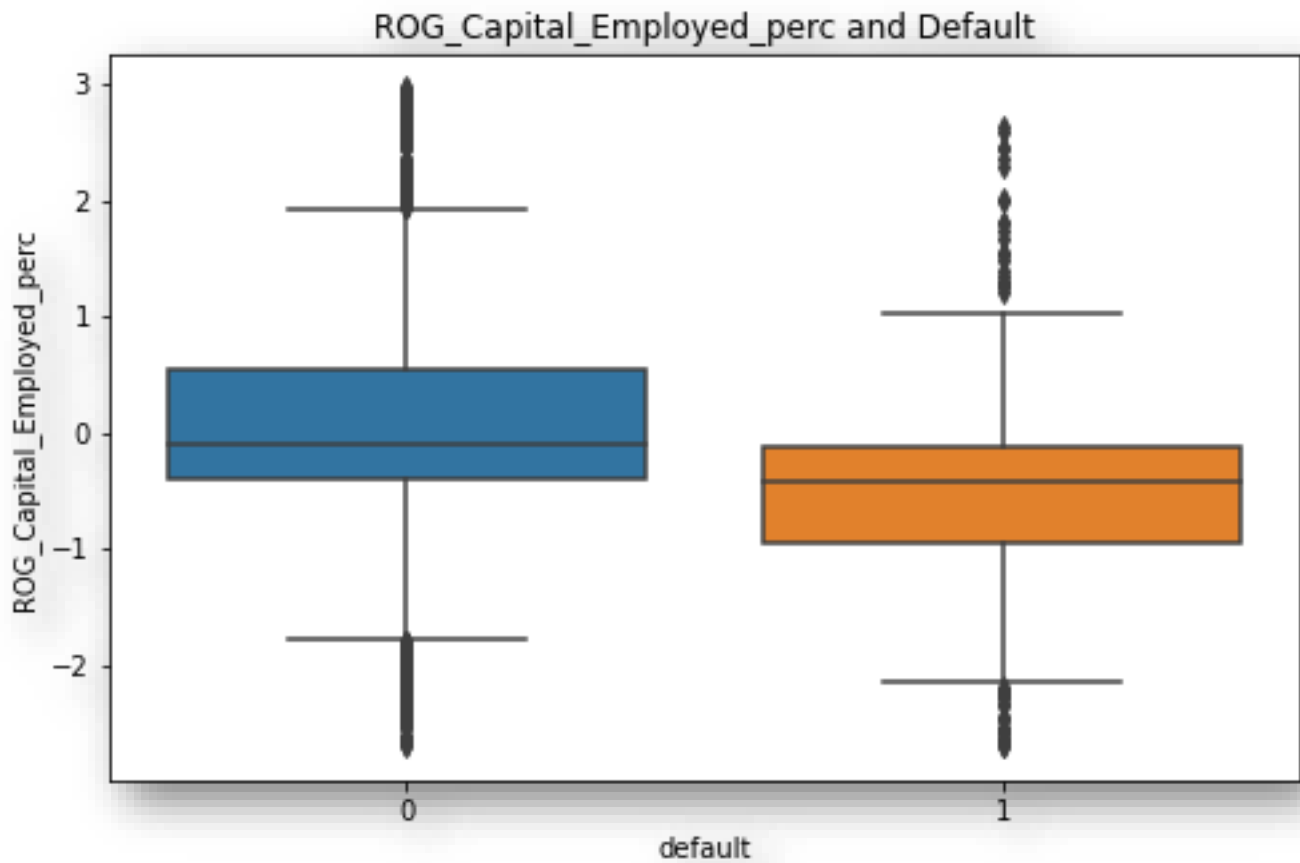
- The net working capital percentage indicates the cashflow for operating the business on a day-to-day basis. The higher the NWC, the efficiency of the company is higher.

**FIGURE 27: PLOT OF DEFAULT WITH NET WORKING CAPITAL**

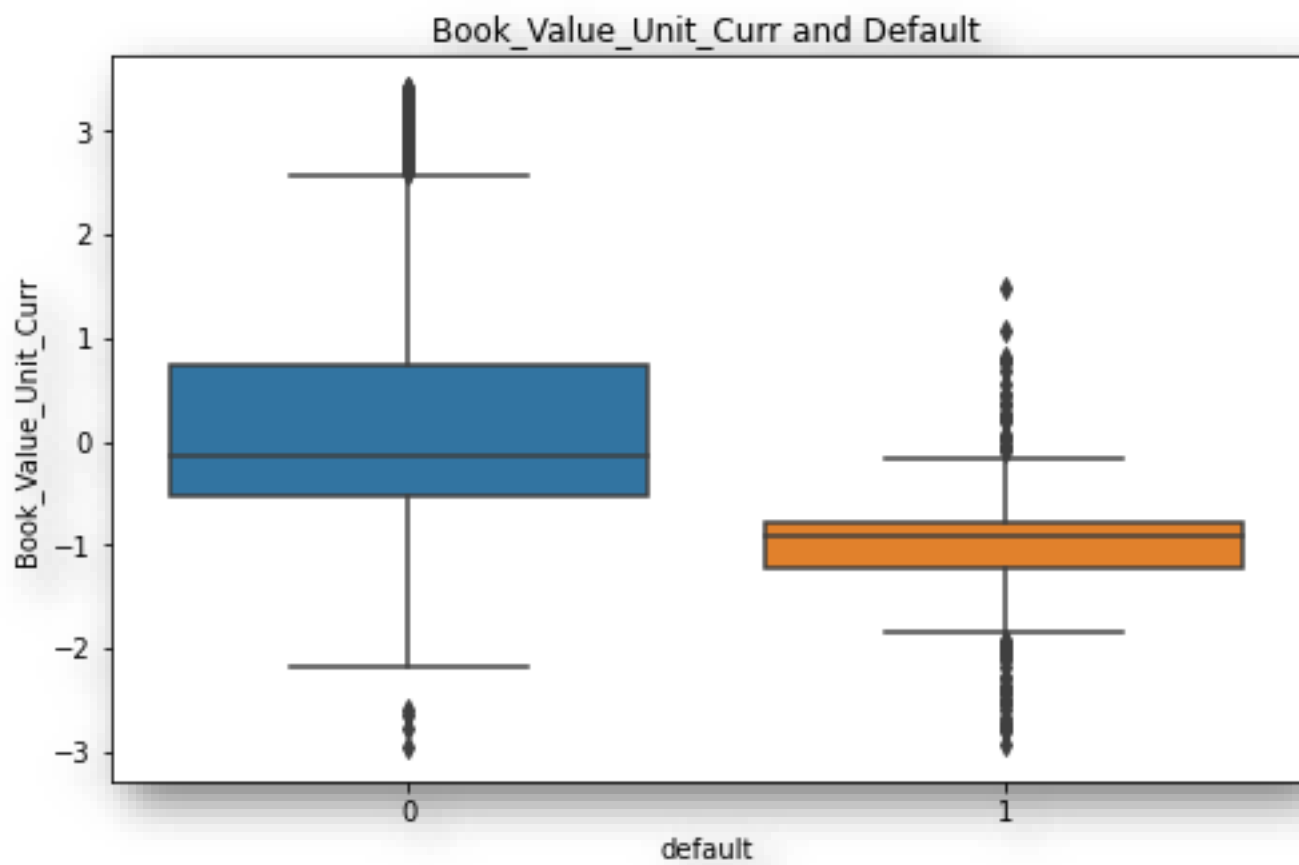


- The defaulters have less cost of production which is good when compared to non-defaulters, it may depend on the quality of production and hence may effect the next year net worth, will need more to infer this.

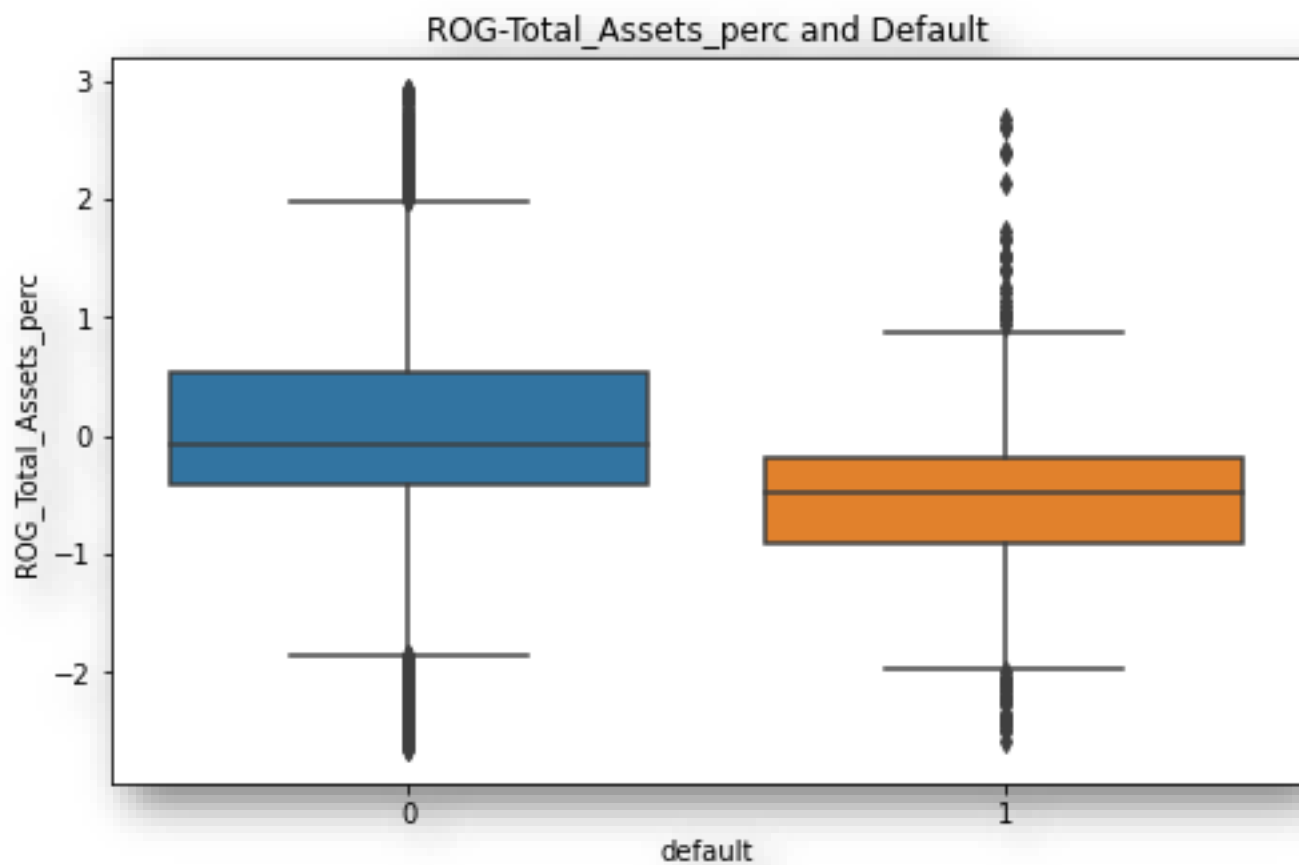
**FIGURE 28: PLOT OF DEFAULT WITH ROG  
CAPITAL EMPLOYED PERC**



**FIGURE 29: PLOT OF DEFAULT WITH BOOK  
VALUE UNIT CURR**



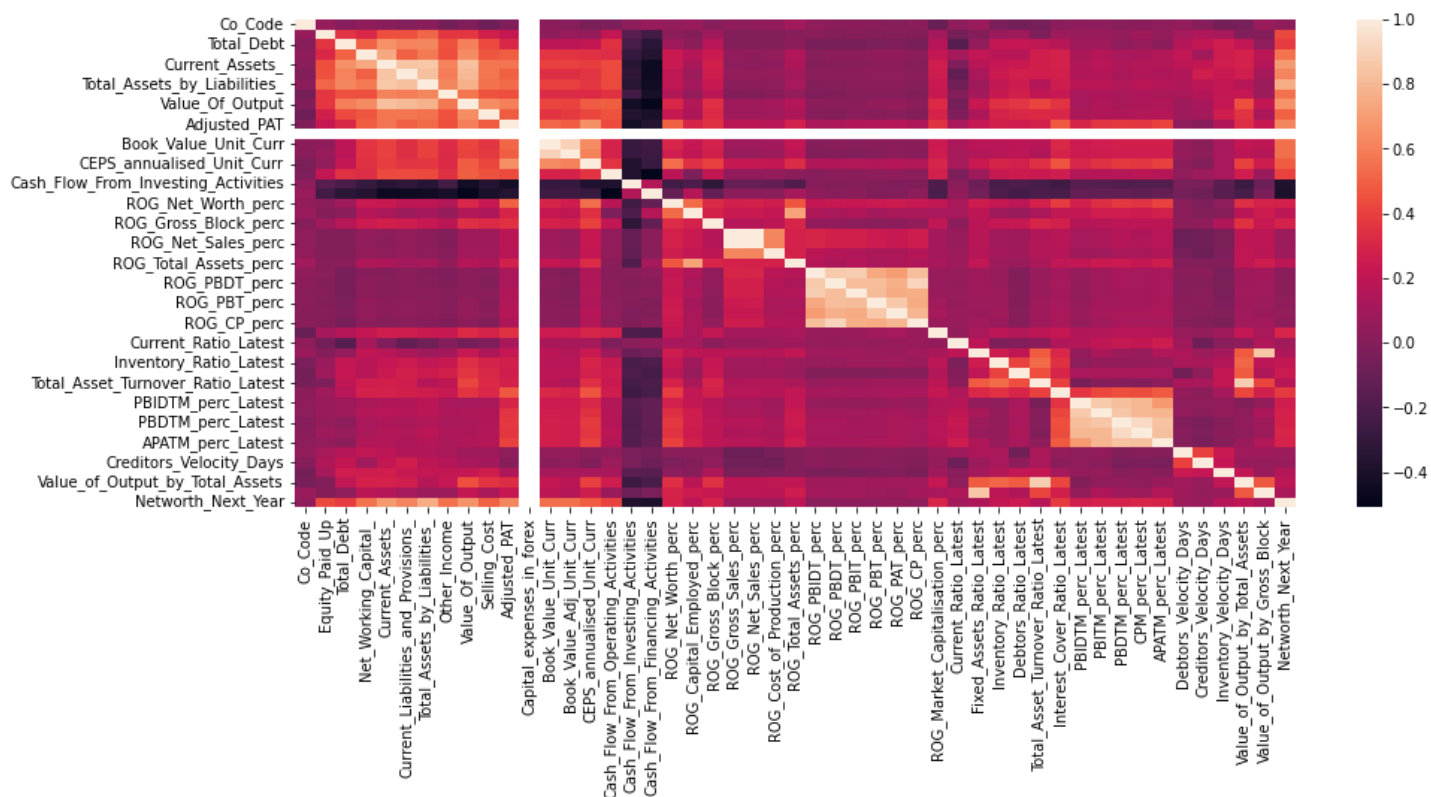
**FIGURE 30: PLOT OF DEFAULT WITH ROG  
TOTAL ASSETS PERC**





# FIGURE 31: HEATMAP

There



- There are a lot of red patches seen in the heatmap. This means that there is a high correlation between many variables.
- Higher correlated features cause multi-collinearity which affect the interpretability of logistic regression model. They are best removed
- We use Variance inflation factor method and remove all the variables with  $VIF > 5$ , done one by one. (Shown in Q.1.6)

## Q.1.5. TRAIN AND TEST SPLIT

- Split the data into Train and Test dataset in a ratio of 67:33 with the fixed random state as 42 to ensure uniformity across multiple systems and stratify on default to make sure both train and test data have similar proportion of defaulters and non-defaulters. This is done as the dataset is imbalanced and has more of non-defaulters. Before we do the train-test split, we will first separate independent (X) and dependent (y) variables (to perform Train-Test split).
- The train set has 2188 observations and test set has 214 observations.

## Q.1.6. BUILD LOGISTIC REGRESSION MODEL

Here, we will use Logistic regression Model to evaluate the relationship between one dependent binary variable and one or more independent variables. This model will help predicts the probability of occurrence of Default using a logit function.

Assumptions of Logistic Regression Model:

1. It assumes that there is minimal, or no multi-collinearity among the independent variables.
  2. It assumes that independent variables are linearly related to log of odds.
  3. It assumes a large sample for good prediction.
  4. It assumes that the observations are independent of each other.
  5. There are no influential values(outliers) in the continuous predictors (independent variables).
  6. Logistic Regression with 2 classes that the dependent variable is binary and the ordered Logistic Regression requires the dependent variable to be ordered.
- There are two methods to solve a Logistic Regression problem:
    1. Stats Model
    2. Scikit Learn

Note: Stats models provides a Logit () function for performing logistic regression. The Logit 0 function accepts y and X as parameters and returns the Logit object. The model is then fitted to the data. The logit function is simply the logarithm of the odds.

$$\text{logit}(x) = \log(x) / (1 - x)$$

The inverse of the logit function is the sigmoid function. The equation of the Logistic Regression by which we predict the corresponding probabilities and then go on predict a discrete target variable is  $Y=1/1+e^{-z}$ .

**TABLE 7: VARAINCE INFLATION FACTOR**

	variables	VIF
21	ROG_Gross_Sales_perc	87.900
22	ROG_Net_Sales_perc	87.613
40	PBDTM_perc_Latest	10.026
26	ROG_PBDT_perc	9.091
4	Current_Assets_	8.777
41	CPM_perc_Latest	7.954
30	ROG_CP_perc	7.132
25	ROG_PBITD_perc	6.716
28	ROG_PBT_perc	6.260
39	PBITM_perc_Latest	6.164
27	ROG_PBIT_perc	6.034
46	Value_of_Output_by_Total_Assets	5.988
5	Current_Liabilities_and_Provisions_	5.766
6	Total_Assets_by_Liabilities_	5.607
12	Book_Value_Unit_Curr	5.497
36	Total_Asset_Turnover_Ratio_Latest	5.478
8	Value_Of_Output	5.391
13	Book_Value_Adj_Unit_Curr	5.334

29	ROG_PAT_perc	5.331
38	PBITDM_perc_Latest	5.072
42	APATM_perc_Latest	4.360
33	Fixed_Assets_Ratio_Latest	4.089
47	Value_of_Output_by_Gross_Block	3.892
3	Net_Working_Capital_	3.031
10	Adjusted_PAT	2.874
14	CEPS_annualised_Unit_Curr	2.852
48	Networth_Next_Year	2.720
19	ROG_Capital_Employed_perc	2.580
24	ROG_Total_Assets_perc	2.317
2	Total_Debt	2.073
9	Selling_Cost	2.044
18	ROG_Net_Worth_perc	2.029
7	Other_Income	1.827
37	Interest_Cover_Ratio_Latest	1.815
17	Cash_Flow_From_Financing_Activities	1.774
15	Cash_Flow_From_Operating_Activities	1.760
23	ROG_Cost_of_Production_perc	1.680
1	Equity Paid Up	1.574

34	Inventory_Ratio_Latest	1.573
35	Debtors_Ratio_Latest	1.552
16	Cash_Flow_From_Investing_Activities	1.551
20	ROG_Gross_Block_perc	1.395
44	Creditors_Velocity_Days	1.365
43	Debtors_Velocity_Days	1.342
31	ROG_Market_Capitalisation_perc	1.264
32	Current_Ratio_Latest	1.256
45	Inventory_Velocity_Days	1.236
0	Co_Code	1.097
11	Capital_expenses_in_forex	NaN

- Here we see that the value of VIF is high for many variables. Hence dropping the variables which are more than 5 and build our model.

#### 1. MODEL 1:

We create a model summary using the significant variable after VIF

#### 2. MODEL 2:

Variable 'ROG\_Gross\_Block\_perc' has the highest p-value (96%) and is insignificant, therefore, we need to eliminate it.

#### 3. MODEL 3:

Variable 'Other\_Income' has the highest p-value (89.7%) and is insignificant, therefore, we need to eliminate it.

#### 4. MODEL 4:

Variable 'Value\_of\_Output\_by\_Gross\_Block' has the highest p-value (86.6%) and is insignificant, therefore, we need to eliminate it.

#### 5. MODEL 5:

Variable 'ROG\_CP\_perc' has the highest p-value (75.4%) and is insignificant, therefore, we need to eliminate it.

#### 6. MODEL 6:

Variable 'ROG\_PBIT\_perc' has the highest p-value (79.5%) and is insignificant, therefore, we need to eliminate it.

#### 7. MODEL 7:

Variable 'Selling Cost' has the highest p-value (72.5%) and is insignificant, therefore, we need to eliminate it.

#### **8. MODEL 8:**

Variable 'PBDTM\_perc\_Latest' has the highest p-value (63.7%) and is insignificant, therefore, we need to eliminate it.

#### **9. MODEL 9:**

Variable 'ROG\_Market\_Capitalisation\_perc' has the highest p-value (53.7%) and is insignificant, therefore, we need to eliminate it.

#### **10.MODEL 10:**

Variable 'ROG\_Total\_Assets\_perc' has the highest p-value (29.4%) and is insignificant, therefore, we need to eliminate it.

#### **11. MODEL 11:**

Variable 'Cash\_Flow\_From\_Financing\_Activities' has the highest p-value (24.1%) and is insignificant, therefore, we need to eliminate it.

#### **12.MODEL 12:**

Variable 'Debtors\_Velocity\_Days' has the highest p-value (23.3%) and is insignificant, therefore, we need to eliminate it.

#### **13.MODEL 13:**

Variable 'Inventory\_Ratio\_Latest' has the highest p-value (18.4%) and is insignificant, therefore, we need to eliminate it.

# TABLE 8: MODEL 13

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-5.1229	0.221	-23.165	0.000	-5.556	-4.689
Book_Value_Unit_Curr	-1.0029	0.355	-2.822	0.005	-1.700	-0.306
Book_Value_Adj_Unit_Curr	-2.5097	0.370	-6.780	0.000	-3.235	-1.784
Value_of_Output_by_Total_Assets	0.6619	0.177	3.733	0.000	0.314	1.009
Total_Asset_Turnover_Ratio_Latest	-0.3229	0.173	-1.862	0.063	-0.663	0.017
CPM_perc_Latest	-0.2082	0.115	-1.804	0.071	-0.434	0.018
Fixed_Assets_Ratio_Latest	-0.4221	0.161	-2.618	0.009	-0.738	-0.106
Adjusted_PAT	-0.2412	0.141	-1.709	0.087	-0.518	0.035
ROG_Capital_Employed_perc	0.1895	0.099	1.904	0.057	-0.006	0.384
ROG_Net_Worth_perc	-0.3321	0.118	-2.811	0.005	-0.564	-0.101
Interest_Cover_Ratio_Latest	-0.4380	0.144	-3.051	0.002	-0.719	-0.157
Debtors_Ratio_Latest	-0.1873	0.109	-1.725	0.085	-0.400	0.026
Cash_Flow_From_Investing_Activities	0.3178	0.135	2.361	0.018	0.054	0.582
Net_Working_Capital_	-0.5153	0.141	-3.649	0.000	-0.792	-0.239
Total_Debt	0.8154	0.119	6.861	0.000	0.582	1.048
Current_Ratio_Latest	-0.9807	0.130	-7.569	0.000	-1.235	-0.727
Equity_Paid_Up	-0.2937	0.102	-2.879	0.004	-0.494	-0.094

- Now all the variables are significant, therefore we don't need to eliminate any variable. Therefore, after many such iterations the variable are removed.

## Q.1.7. VALIDATE THE MODEL ON TEST DATSET AND STATE THE PERFORMANCE MATRICES

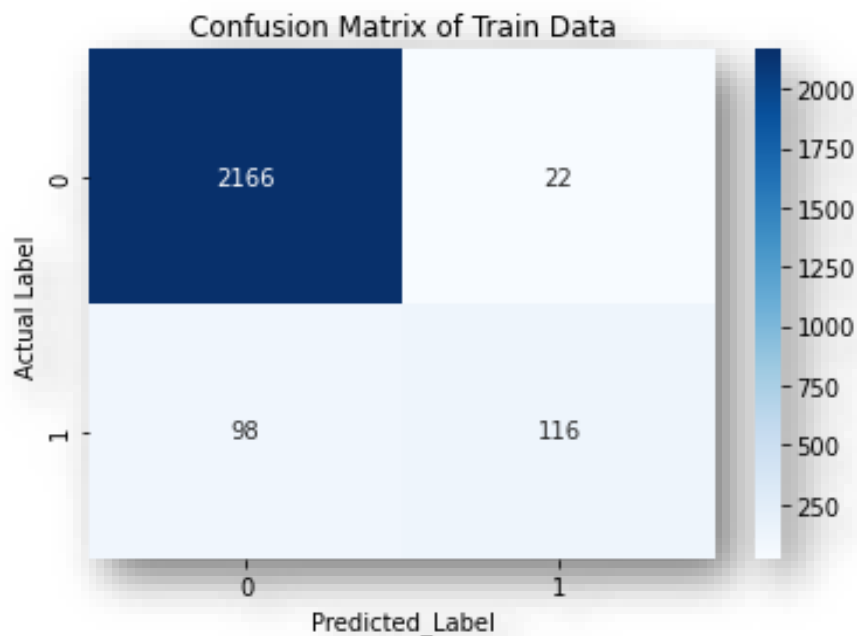
TABLE 9: PREDICTION TRAIN DATA

	0
662	0.109
1373	0.002
3268	0.009
3246	0.000
1456	0.007

TABLE 10: PREDICTION TEST DATA

	0
3163	0.003
3133	0.001
937	0.105
196	0.901
2852	0.023

**FIGURE 32: CONFUSION MATRIX OF TRAIN DATA**

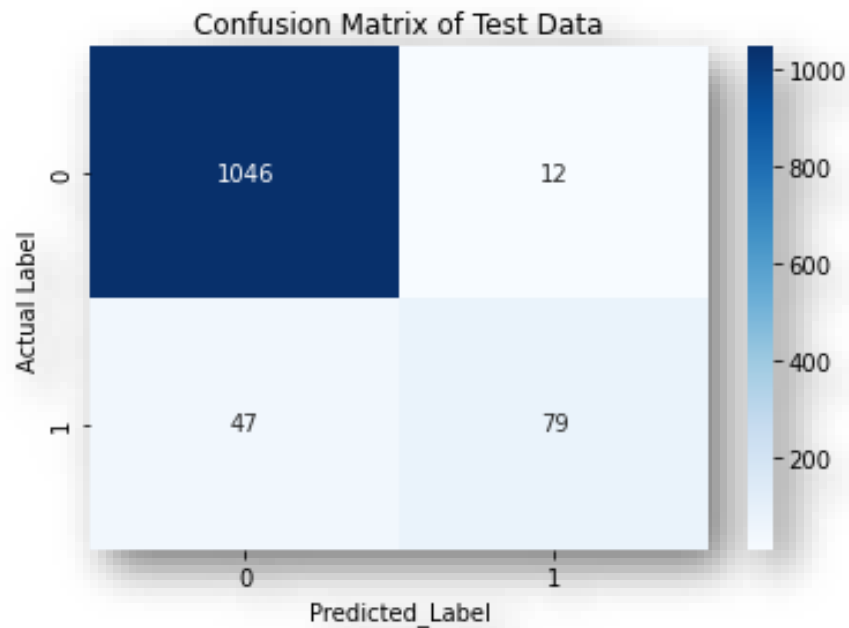


**TABLE 11: CLASSIFICATION REPORT TRAIN DATA**

	precision	recall	f1-score	support
0	0.96	0.99	0.97	2188
1	0.84	0.54	0.66	214
accuracy			0.95	2402
macro avg	0.90	0.77	0.82	2402
weighted avg	0.95	0.95	0.95	2402



**FIGURE 33: CONFUSION MATRIX OF TEST DATA**



**TABLE 11: CLASSIFICATION REPORT TEST DATA**

	precision	recall	f1-score	support
0	0.96	0.99	0.97	1058
1	0.87	0.63	0.73	126
accuracy			0.95	1184
macro avg	0.91	0.81	0.85	1184
weighted avg	0.95	0.95	0.95	1184

## INTERPRETATION

1. The precision and recall for both the train and test data are similar. Here the precision and recall mean the actual prediction which is true.
2. Of many variables - significantly only 6 variables contribute to the company being predicted as default or not from logistic regression point of view.
3. The model is likely to predict the 86% companies that could default correctly.
4. Which means only in 14% cases - it could happen that a company that is predicted as defaulter may not be a defaulter but from an investor point of view - it is ok to no invest money on company that could likely not default.
5. The precision is a bit less in this model - however still 68% times, the model will predict the defaulter company correctly.