GREAT LEARNING'S

FRA Milestone 1 -BUSINESS REPORT

This Business Report shall provide detailed explanation of how we approached each problem given in the assignment. It shall also provide relative resolution and explanation with regards to the problems

PRABHJOT KAUR CHAHAL (PGP-DSBA -Online Mar_C 2021)

2021

GREAT LEARNING'S

Contents

MILESTONE:1:Overview and Credit Risk

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:

	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 service	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
26	CP	Commercial paper , a short-term debt instrument to meet short-term liabilities.	CP
27	Revenue earnings in forex	Revenue earned in foreign currency	Rev_earn_in_forex
28	Revenue expenses in forex	Expenses due to foreign currency transactions	Rev_exp_in_forex
29	Capital expenses in forex	Long term investment in forex	Capital_exp_in_forex
30	Book Value (Unit Curr)	Net asset value	Book_Value_Unit_Curr
31	Book Value (Adj.) (Unit Curr)	Book value adjusted to reflect asset's true fair market value	Book_Value_Adj_Unit_Curr
32	Market Capitalisation	Product of the total number of a company's outstanding shares and the current market price of one share	Market_Capitalisation

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

Dataset for Problem: Credit Risk Dataset , Data Dictionary

SUMMARIZING BUSINESS PROBLEM

This report includes a classification model of a company's financial data using logistic regression. It is planned to determine if a certain company is in good financial standing and whether is valued positive net worth next year or not. We used Python to code.

IMPORTING AND READING THE DATASET:

Dataset has 67 variables of which 63 are of float datatype, 3 are integer type and 1 is object type.

```
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
```

refer jupyter notebook

The head of the dataset is as below:

	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

Descriptive statistics / 5 point summary is shown below:

	count	mean	std	min	25%	50%	75%	max
Co_Code	3586.0	16065.388734	19776.817379	4.00	3029.2500	6077.500	24269.5000	72493.00
Networth_Next_Year	3586.0	725.045251	4769.681004	-8021.60	3.9850	19.015	123.8025	111729.10
Equity_Paid_Up	3586.0	62.966584	778.761744	0.00	3.7500	8.290	19.5175	42263.46
Networth	3586.0	649.746299	4091.988792	-7027.48	3.8925	18.580	117.2975	81657.35
Capital_Employed	3586.0	2799.611054	26975.135385	-1824.75	7.6025	39.090	226.6050	714001.25
...	...	...	...	...	...	...	...	...
Debtors_Velocity_Days	3586.0	603.894032	10636.759580	0.00	8.0000	49.000	106.0000	514721.00
Creditors_Velocity_Days	3586.0	2057.854992	54169.479197	0.00	8.0000	39.000	89.0000	2034145.00
Inventory_Velocity_Days	3483.0	79.644559	137.847792	-199.00	0.0000	35.000	96.0000	996.00
Value_of_Output_to_Total_Assets	3586.0	0.819757	1.201400	-0.33	0.0700	0.480	1.1600	17.63
Value_of_Output_to_Gross_Block	3586.0	61.884548	976.824352	-61.00	0.2700	1.530	4.9100	43404.00

66 rows × 8 columns

We performed the descriptive summary for the company data. Since most of the column data is continuous, we can see the mean, standard deviation and percentile details for all the columns.

- The data has 3586 Rows and 67 Columns.
- No duplicate data is present in the data set.
- We dropped unrequited columns like Co_Code and Co_Name since they do not add value to the analysis

- NULL VALUES: 118 Null values present for below variables

Inventory_Velocity_Days	103
Book_Value_Adj._Unit_Curr	4
Interest_Cover_Ratio_Latest_	1
PBITM_perc_Latest_	1
Fixed_Assets_Ratio_Latest_	1
Inventory_Ratio_Latest_	1
Debtors_Ratio_Latest_	1
Total_Asset_Turnover_Ratio_Latest_	1
PBIDTM_perc_Latest_	1
PBDTM_perc_Latest_	1
CPM_perc_Latest_	1
APATM_perc_Latest_	1
Current_Ratio_Latest_	1

1.1 Outlier Treatment

*For better view, please refer python notebook

We used 3 times the IQR range as the criteria to determine the outliers. Our analysis gave significant chunk of outliers in the data. Below are boxplots which were plotted to analyze this data.

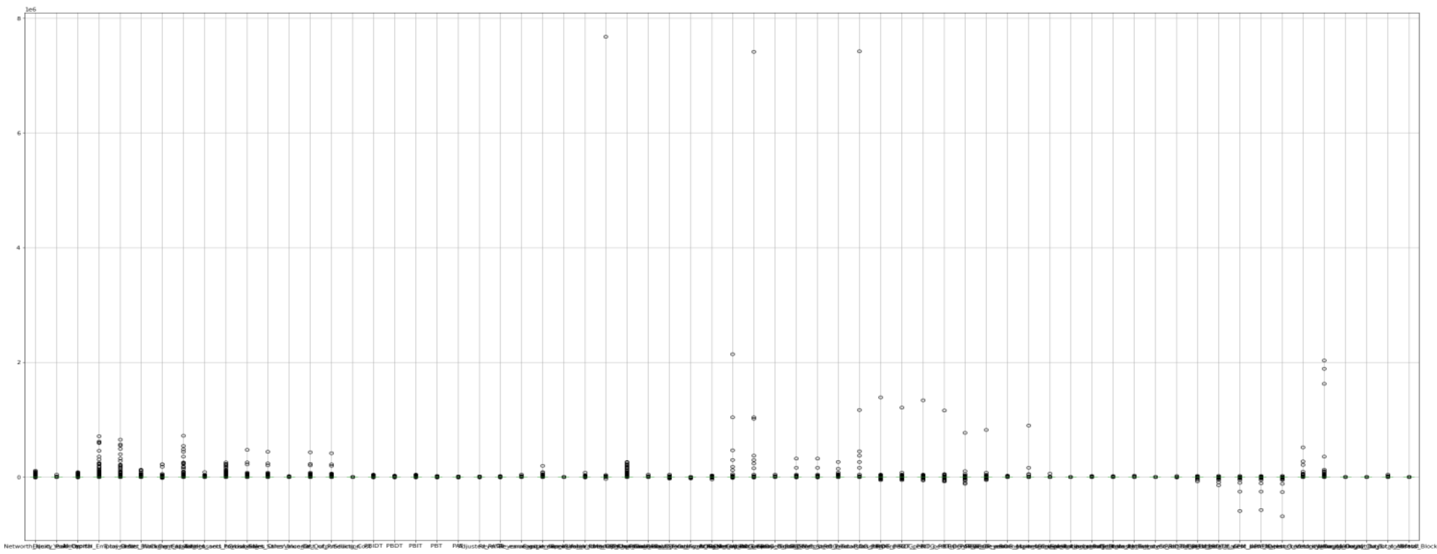


Fig1: Before outliers Treatment

- Significant number of outliers was present for almost all the variables.

<u>Treating outlier by using Inter Quartile range for each of numerical column.</u>
Values greater than Upper quartile range - capped with 75% of quartile value
Values lesser than Lower quartile range - capped with 25% of quartile value

The outliers would be replaced with Upper Quartile values or lower. And post outlier treatment the numerical variables in boxplot:

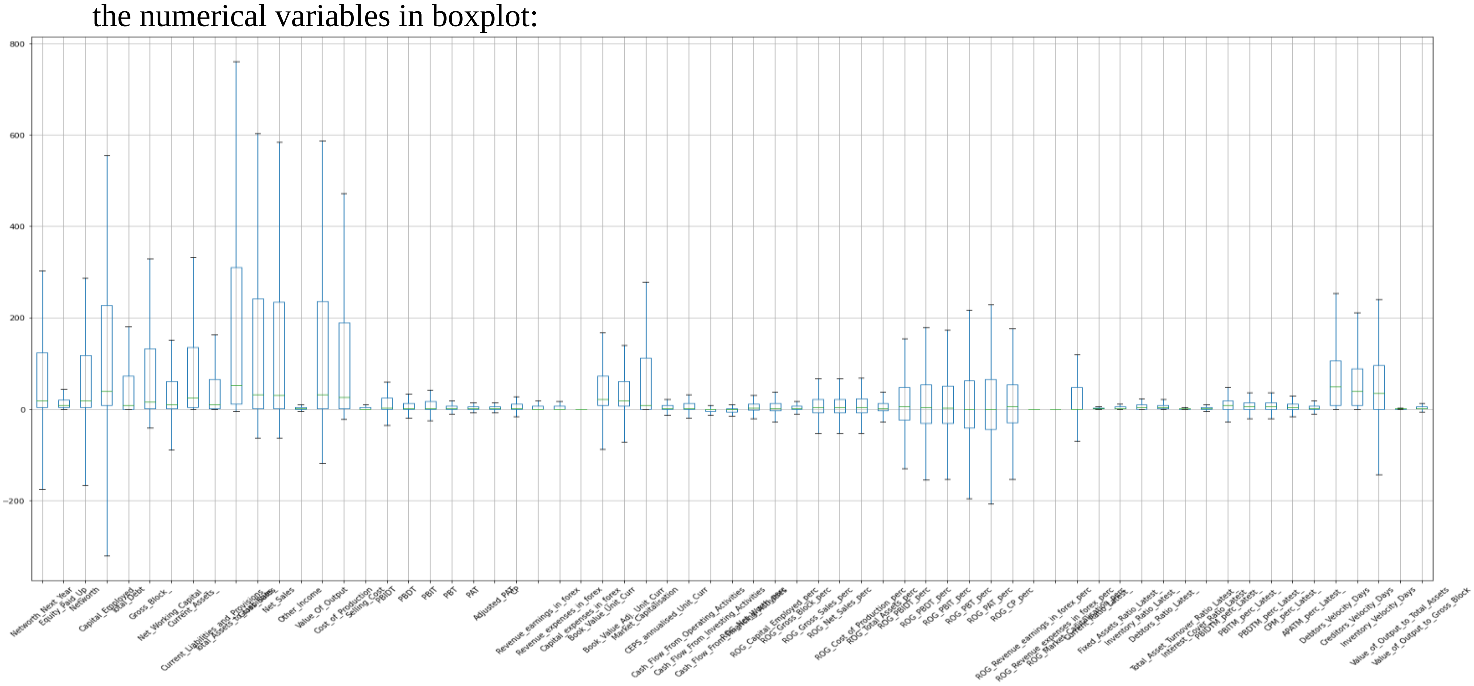


Fig2: After outliers Treatment

1.2 Missing Value Treatment

Checking for null values: Given the size of the data set i.e. 3586 rows, there were not many missing values to start with. There were a total of 118 missing records observed in the entire data.

Inventory_Velocity_Days	103
Book_Value_Adj._Unit_Curr	4
Interest_Cover_Ratio_Latest_	1
PBITM_perc_Latest_	1
Fixed_Assets_Ratio_Latest_	1
Inventory_Ratio_Latest_	1
Debtors_Ratio_Latest_	1
Total_Asset_Turnover_Ratio_Latest_	1
PBIDTM_perc_Latest_	1
PBDTM_perc_Latest_	1
CPM_perc_Latest_	1
APATM_perc_Latest_	1
Current_Ratio_Latest_	1

- Null values were present in many columns, however significant number was present in "Inventory_Vel_Days" column. This is the one which we treated.
- The missing values are of numeric nature and imputed using KNN imputer.
- Imputation is done by predicting the missing value based on values of 10 nearest neighbors of the same variable.

Such that all the missing values are replaced based on nearest neighbors value as follow:

Equity_Paid_Up	0
Networth	0
Capital_Employed	0
Total_Debt	0
Gross_Block_	0
Net_Working_Capital_	0
Current_Assets_	0
Current_Liabilities_and_Provisions_	0
Total_Assets_to_Liabilities_	0
Gross_Sales	0
Net_Sales	0
Other_Income	0
Value_Of_Output	0
Cost_of_Production	0
PBIDT	0
PBIT	0
Capital_expenses_in_forex	0
Book_Value_Unit_Curr	0
Book_Value_Adj._Unit_Curr	0
Market_Capitalisation	0
CEPS_annualised_Unit_Curr	0
ROG_Capital_Employed_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_Market_Capitalisation_perc	0
Current_Ratio_Latest_	0
Fixed_Assets_Ratio_Latest_	0
Inventory_Ratio_Latest_	0
Debtors_Ratio_Latest_	0
Total_Asset_Turnover_Ratio_Latest_	0
Interest_Cover_Ratio_Latest_	0
PBIDTM_perc_Latest_	0
PBITM_perc_Latest_	0
PBDTM_perc_Latest_	0
CPM_perc_Latest_	0
Debtors_Velocity_Days	0
Creditors_Velocity_Days	0
Value_of_Output_to_Total_Assets	0
Value_of_Output_to_Gross_Block	0
default	0
dtype: int64	

1.3 Transform Target variable into 0 and 1

A new dependent variable named "Default" was created based on the criteria given in the project notes.
There is no target variable defined – but since the objective is to build a model for investor to decode which company to invest in – the variable “Networth_Next_Year” could be used to transform into target variable.

CRITERIA –

1 – **DEFAULT**-If the Net worth Next Year is negative or less than 0 for the company. This means the company would likely not return a good investment to investor and transformed as 1.

0 - **NON-DEFAULT**-If the Net worth Next Year is positive or greater than 0 for the company.This means the company would continue to return good investment for investor and thus could be transformed as 0.

Made use of np.where function to achieve this.

We checked for the split of data based on this dependent variable, after generating a dependent column. Below is a bar plot showing the same.

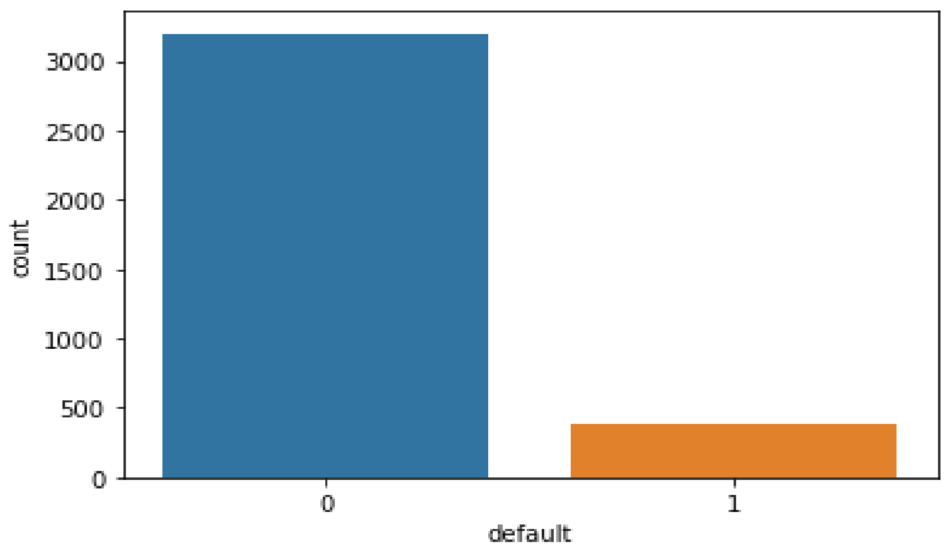


Fig 3: Overall distribution of the Default

Distinct values of the dependent variable – 0 and 1:

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

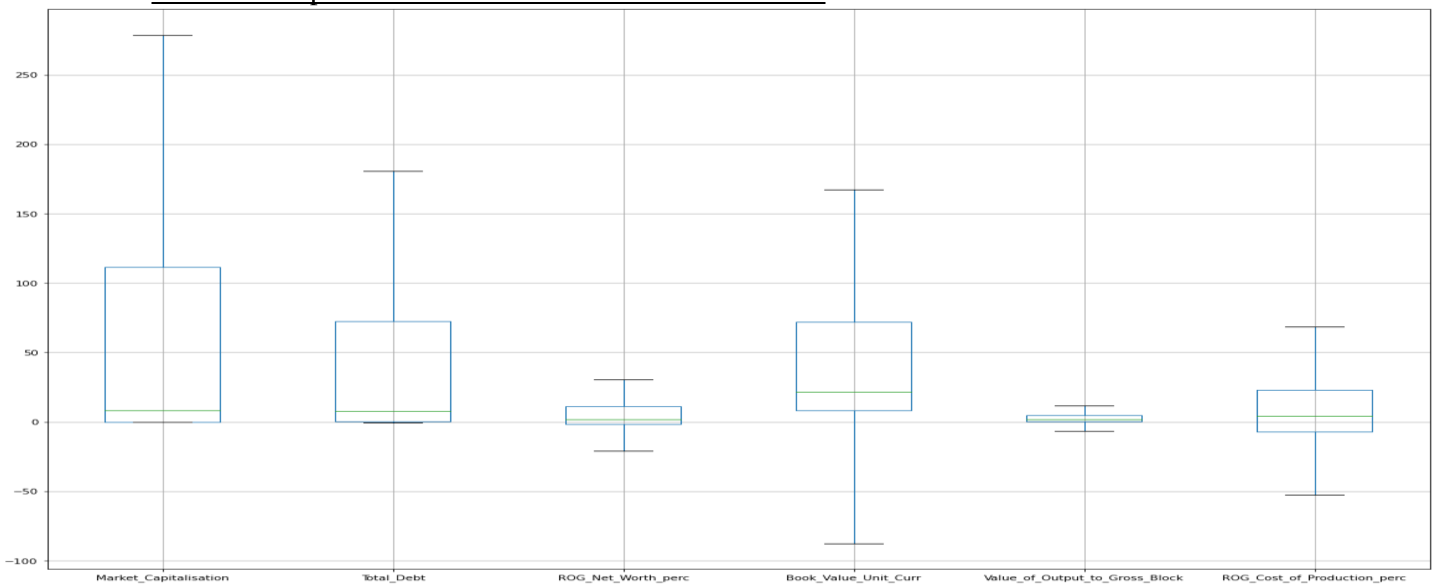
This interprets, 11% of the companies from the dataset are likely to be default and are ones the investor could avoid investing in.

1.4 Univariate (4 marks) & Bivariate (6marks) analysis with proper interpretation. (You may choose to include only those variables which were significant in the model building)

- Those variables which were significant in the model building:

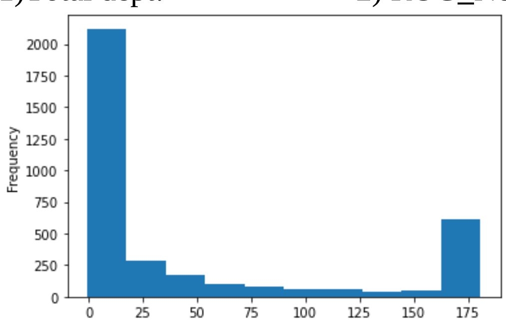
Market_Capitalisation, Total_Debt, ROG_Net_Worth_perc, Book_Value_Unit_Curr, Value_of_Output_to_Gross_Block, ROG_Cost_of_Production_perc

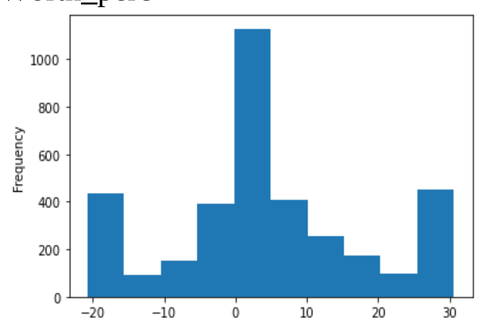
- <u>And the boxplot of them in same order is as follows:</u>



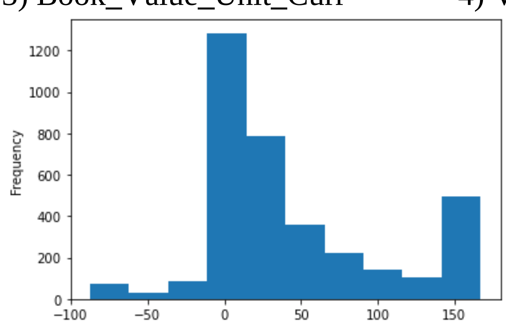
- <u>Similarly the histograms of the same are:</u>

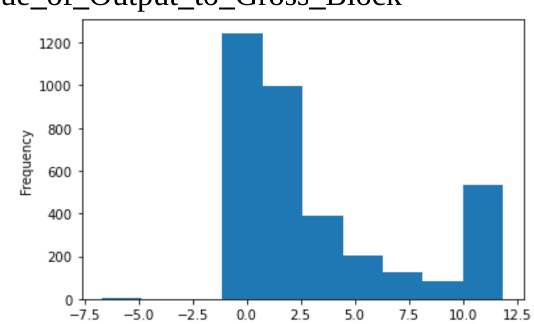
1)Total dept: 2) ROG_Net_Worth_perc



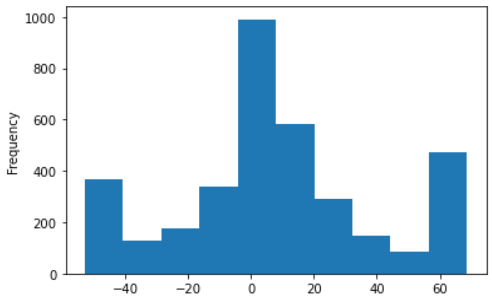


3) Book_Value_Unit_Curr 4) Value_of_Output_to_Gross_Block





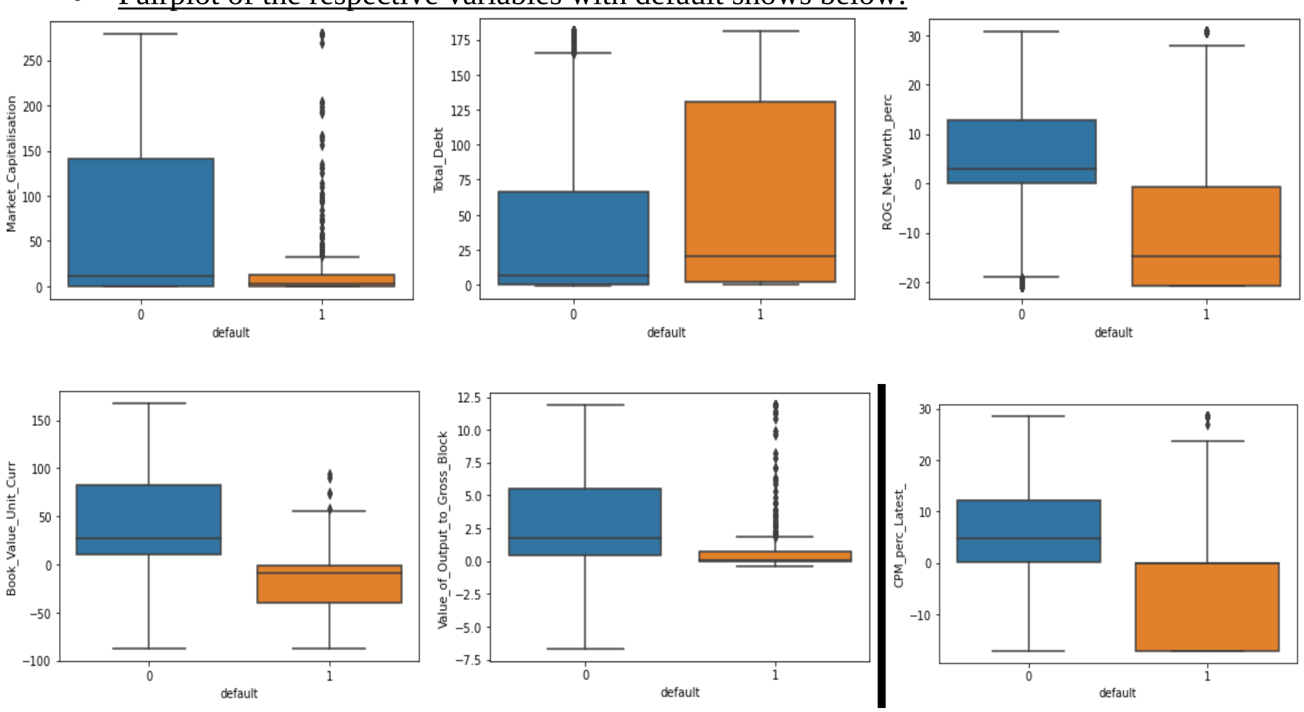
5) ROG_Cost_of_Production_perc



INFERENCE:

- There is one company which has borrowed a highest sum from market which is nearly 2000 units, and then some companies have taken dept of around 500-750.
- Most of the companies have dept lower than 250.
- The companies having high dept – may deploy high risk of not making profit next year.
- Net worth rate of growth – if it's more the company is likely to make more profit next year.
- The given dataset shows nearly half companies having positive growth rate of net worth and rest negative.
- Book Value for unit currency indicates Net asset value of company – if it's positive, the company would always have assets which can be used to capitalize should losses need to be covered.
- This is to say these companies have assets which can help bring in the credit rating in case of losses.
- There are 4 companies which have very good book values. And some has really negative book values.
- If these companies are having high depts. – then there is no way the losses could be covered with asset selling's.
- Almost 25% companies having good ratio's of market value and gross block – which
- Means these companies are likely not to default.
- Rate of growth of production depicts how much the company's growth rate is for production cost – it may mean, the company is more likely to have more operation cost or more market share or both.
- From the dataset – this rate is evenly distributed and the ones which are highest are mostly either emerging companies or performing really well.

- <u>Pairplot of the respective variables with default shows below:</u>



INFERENCE:

- The total debt of defaulters is high and market capitalization is less – which is to say more of money to pay than the company owns in market.
- Net worth percentage of the defaulter is less than those of non-defaulters and so are the asset values.
- The output values is defaulters is also less – which is to say less production for companies not likely to make profit next year.
- For whole data-Data is highly skewed and most of the data is found to be right skewed.
- A total of 61 variables were found having tails to the right and hence were right skewed.
- There were a total of 6 variables which were found to be left skewed i.e. they had a longer tail on the left hand side of the distribution.

- **Multivariate analysis:**

  We also performed multi variate analsysis on the data to see if there are any correlation that are observed within the data. Correlations function was used and seaborn clustermap was used to plot the correlations and to make better sense of the data.

  We observed that net worth and networth next year were highly correlated. Apart from this, we also found various Rate of Growth variables were highly correlated.

  This analysis tells us that there is a problem of collinearity with this data set.

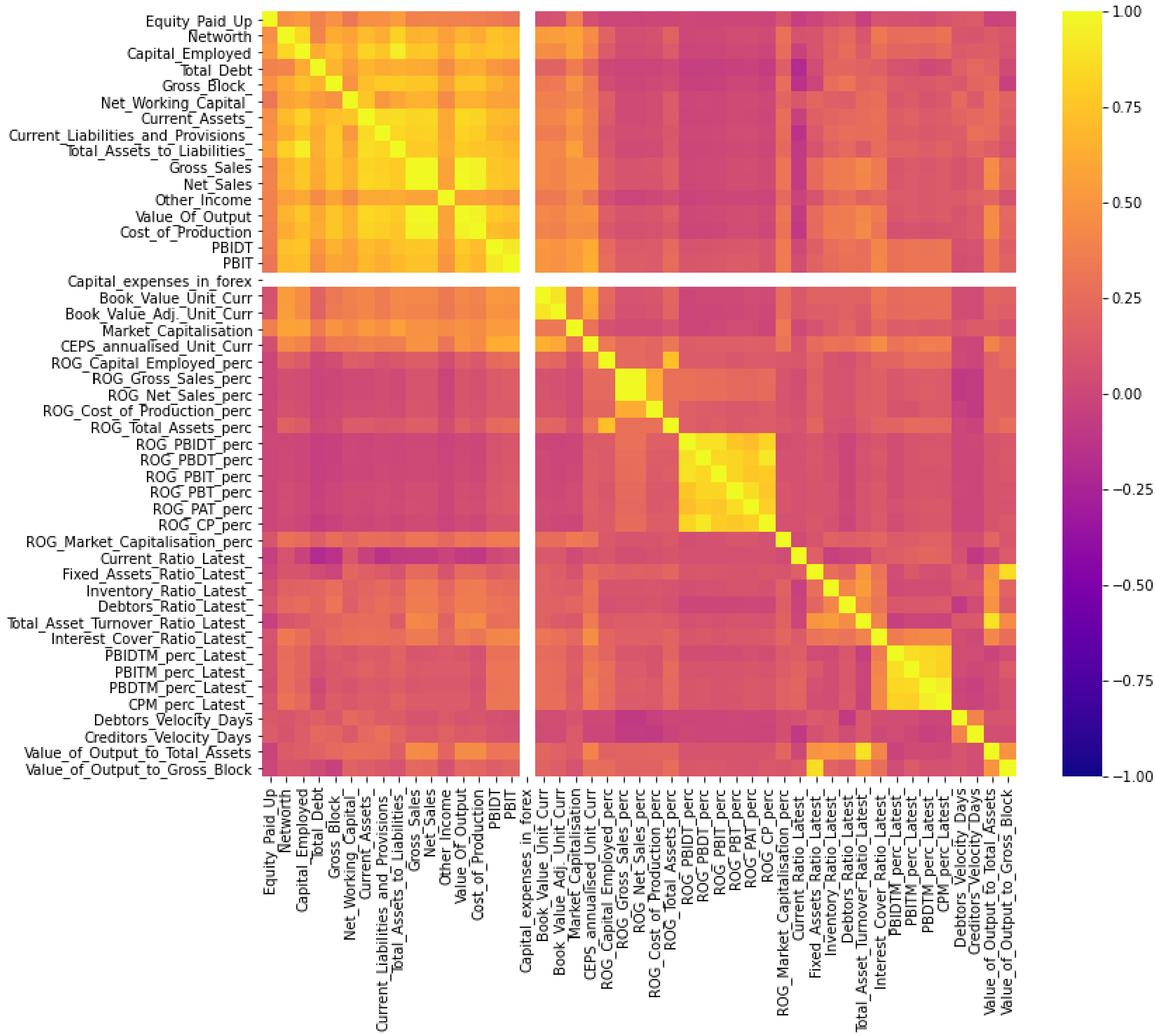


Fig 4: Heatmap

1.5 Train Test Split

```
shape of input - training set (2402, 47)
shape of input - testing set (1184, 47)
```

- The Target variable is Default
- We performed the splitting of training and testing sets in the ratio of 67: 33 and then we try to the fit the model into the testing and training sets and find out the performance of those sets.
- Seed value of 42 was used
- Stratified on default, to make sure both train data and test data have similar proportion of defaulters and non-defaulters.

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

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

Model Building using Logistic Regression for 'Probability at default'

The equation of the Logistic Regression by which we predict the corresponding probabilities and then go on predict a discrete target variable is

$$y = \frac{1}{1-e^{-z}}$$

Note: $z = \beta_0 + \sum_{i=1}^{n}(\beta_i X_1)$

- After importing statsmodels modules:
- Creating logistic regression equation & storing it in f_1
- Splitting arrays or matrices into random train and test subsets.
- Model will be fitted on train set and predictions will be made on the test set
- Train set consists of variables:

```
Index(['Equity_Paid_Up', 'Networth', 'Capital_Employed', 'Total_Debt',
       'Gross_Block_', 'Net_Working_Capital_', 'Current_Assets_',
       'Current_Liabilities_and_Provisions_', 'Total_Assets_to_Liabilities_',
       'Gross_Sales', 'Net_Sales', 'Other_Income', 'Value_Of_Output',
       'Cost_of_Production', 'PBIDT', 'PBIT', 'Capital_expenses_in_forex',
       'Book_Value_Unit_Curr', 'Book_Value_Adj._Unit_Curr',
       'Market_Capitalisation', 'CEPS_annualised_Unit_Curr',
       'ROG_Capital_Employed_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_', 'Inventory_Ratio_Latest_',
       'Debtors_Ratio_Latest_', 'Total_Asset_Turnover_Ratio_Latest_',
       'Interest_Cover_Ratio_Latest_', 'PBIDTM_perc_Latest_',
       'PBITM_perc_Latest_', 'PBDTM_perc_Latest_', 'CPM_perc_Latest_',
       'Debtors_Velocity_Days', 'Creditors_Velocity_Days',
       'Value_of_Output_to_Total_Assets', 'Value_of_Output_to_Gross_Block',
       'default'],
      dtype='object')
```

Model 1

Before starting model building, let's look at the problem of multicollinearity.
Multicollinearity occurs when two or more independent variables are highly correlated with one another in a regression model.
Prior to building the logistic regression model, we had to work on feature selection since there were too many columns to start with and we decided to eliminate a few of the columns using the Variation Inflation Factor i.e. VIF

	variables	VIF
33	Current_Ratio_Latest_	1.26
32	ROG_Market_Capitalisation_perc	1.27
43	Debtors_Velocity_Days	1.32
44	Creditors_Velocity_Days	1.36
36	Debtors_Ratio_Latest_	1.54
35	Inventory_Ratio_Latest_	1.59
0	Equity_Paid_Up	1.68
24	ROG_Cost_of_Production_perc	1.70
38	Interest_Cover_Ratio_Latest_	1.74
19	Market_Capitalisation	1.79
11	Other_Income	1.87
21	ROG_Capital_Employed_perc	2.32
25	ROG_Total_Assets_perc	2.36
3	Total_Debt	2.55
20	CEPS_annualised_Unit_Curr	2.84
5	Net_Working_Capital_	3.63
46	Value_of_Output_to_Gross_Block	4.39
4	Gross_Block_	4.50

1	Networth	5.25
18	Book_Value_Adj._Unit_Curr	5.72
30	ROG_PAT_perc	5.77
39	PBIDTM_perc_Latest_	5.82
17	Book_Value_Unit_Curr	5.94
40	PBITM_perc_Latest_	6.30
37	Total_Asset_Turnover_Ratio_Latest_	6.41
28	ROG_PBIT_perc	6.77
45	Value_of_Output_to_Total_Assets	6.77
7	Current_Liabilities_and_Provisions_	6.85
29	ROG_PBT_perc	7.06
26	ROG_PBIDT_perc	7.59
31	ROG_CP_perc	7.77
15	PBIT	8.28
42	CPM_perc_Latest_	8.39
14	PBIDT	9.78
6	Current_Assets_	10.45
27	ROG_PBDT_perc	10.57
41	PBDTM_perc_Latest_	11.58
2	Capital_Employed	11.61
8	Total_Assets_to_Liabilities_	13.97

7	Current_Liabilities_and_Provisions_	6.85
29	ROG_PBT_perc	7.06
26	ROG_PBIDT_perc	7.59
31	ROG_CP_perc	7.77
15	PBIT	8.28
42	CPM_perc_Latest_	8.39
14	PBIDT	9.78
6	Current_Assets_	10.45
27	ROG_PBDT_perc	10.57
41	PBDTM_perc_Latest_	11.58
2	Capital_Employed	11.61
8	Total_Assets_to_Liabilities_	13.97
13	Cost_of_Production	16.80
23	ROG_Net_Sales_perc	97.38
22	ROG_Gross_Sales_perc	97.43
9	Gross_Sales	107.10
12	Value_Of_Output	141.64
10	Net_Sales	241.77
16	Capital_expenses_in_forex	NaN

Here, we see that the value of VIF is high for many variables. Here, we may drop variables with VIF more than 3 (very high correlation) & build our model.

Fitting the logistic regression mode

```
Optimization terminated successfully.
         Current function value: 0.216778
         Iterations 8
```

Model 1 summary:

Logit Regression Results

Dep. Variable:	default	No. Observations:	2402
Model:	Logit	Df Residuals:	2387
Method:	MLE	Df Model:	14
Date:	Sun, 06 Feb 2022	Pseudo R-squ.:	0.3420
Time:	19:48:50	Log-Likelihood:	-520.70
converged:	True	LL-Null:	-791.34
Covariance Type:	nonrobust	LLR p-value:	1.614e-106

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-3.3594	0.144	-23.289	0.000	-3.642	-3.077
CEPS_annualised_Unit_Curr	-0.6952	0.123	-5.638	0.000	-0.937	-0.454
Total_Debt	0.3698	0.089	4.169	0.000	0.196	0.544
ROG_Total_Assets_perc	-0.3650	0.129	-2.822	0.005	-0.618	-0.111
ROG_Capital_Employed_perc	0.1413	0.126	1.123	0.261	-0.105	0.388
Market_Capitalisation	-0.5741	0.130	-4.420	0.000	-0.829	-0.320
Interest_Cover_Ratio_Latest_	-0.4513	0.123	-3.656	0.000	-0.693	-0.209
ROG_Cost_of_Production_perc	-0.1742	0.094	-1.845	0.065	-0.359	0.011
Equity_Paid_Up	0.1228	0.091	1.346	0.178	-0.056	0.302
Inventory_Ratio_Latest_	0.0356	0.101	0.351	0.725	-0.163	0.234
Debtors_Ratio_Latest_	-0.0881	0.097	-0.912	0.362	-0.277	0.101
Creditors_Velocity_Days	0.1747	0.083	2.105	0.035	0.012	0.337
Debtors_Velocity_Days	-0.0989	0.093	-1.067	0.286	-0.280	0.083
ROG_Market_Capitalisation_perc	-0.0592	0.095	-0.622	0.534	-0.246	0.127
Current_Ratio_Latest_	-1.7654	0.162	-10.870	0.000	-2.084	-1.447

We can see that few variables are insignificant & may not be useful to discriminate cases of deault
Let us look at the adjusted pseudo R-square value

```
The adjusted pseudo R-square value is 0.3243134162365523
```

Adjusted pseudo R-square seems to be lower than Pseudo R-square value which means there are insignificant variables present in the model. Lets try & remove variables whose p value is greater than 0.05 & rebuild our model

Model 2

Model 2 summary:

```
Optimization terminated successfully.
         Current function value: 0.302415
         Iterations 7
```

Logit Regression Results

Dep. Variable:	default	No. Observations:	2402
Model:	Logit	Df Residuals:	2394
Method:	MLE	Df Model:	7
Date:	Sun, 06 Feb 2022	Pseudo R-squ.:	0.08207
Time:	19:48:50	Log-Likelihood:	-726.40
converged:	True	LL-Null:	-791.34
Covariance Type:	nonrobust	LLR p-value:	6.644e-25

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-2.3824	0.080	-29.759	0.000	-2.539	-2.226
ROG_Market_Capitalisation_perc	-0.2486	0.080	-3.127	0.002	-0.405	-0.093
Debtors_Velocity_Days	-0.2268	0.083	-2.748	0.006	-0.388	-0.065
Debtors_Ratio_Latest_	-0.2167	0.089	-2.434	0.015	-0.391	-0.042
Inventory_Ratio_Latest_	-0.1995	0.093	-2.150	0.032	-0.381	-0.018
Equity_Paid_Up	0.1846	0.067	2.738	0.006	0.052	0.317
ROG_Capital_Employed_perc	-0.5740	0.082	-6.992	0.000	-0.735	-0.413
Creditors_Velocity_Days	0.3318	0.069	4.807	0.000	0.197	0.467

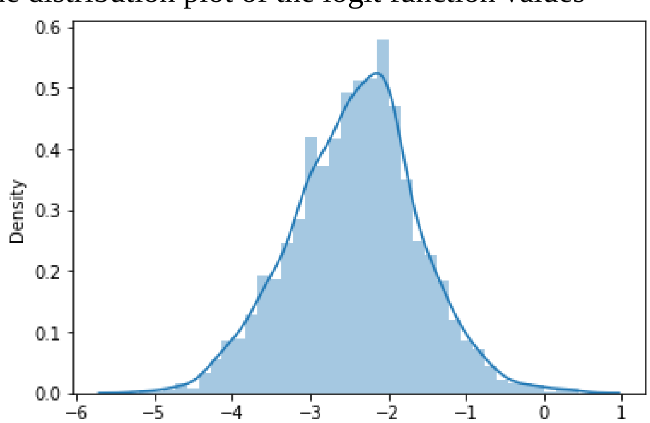
- We can see that all variables are significant & may be useful to discriminate cases of deault

```
The adjusted pseudo R-square value is 0.07322015707561758
```

- We see that adjusted R sq is now close to R sq, thus suggesting lesser insignificant variables in the model
- We also notice that current model has no insignificant variables and can be used for prediction purposes.
- Lets test the prediction of this model on train and test dataset

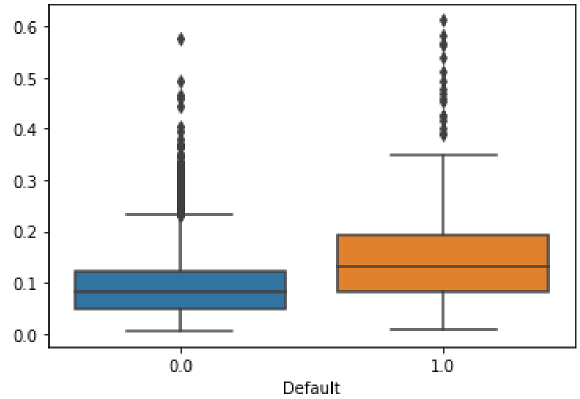
Prediction on the Data

Let us first check the distribution plot of the logit function values



From the boxplot, we need to decide on one such value of a cut-off which will give us the most reasonable descriptive power of the model.

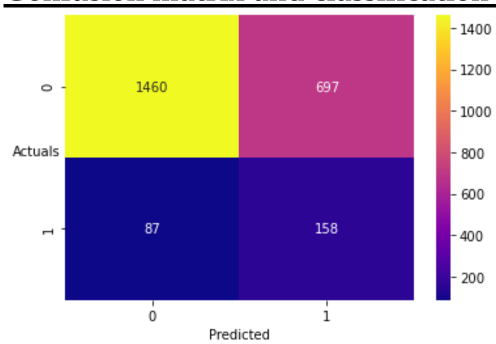
Let us take a cut-off of 0.11 and check.



Let us now see the predicted classes

Checking the accuracy of the model using confusion matrix for training set:

- **Confusion matrix and classification report for train data**



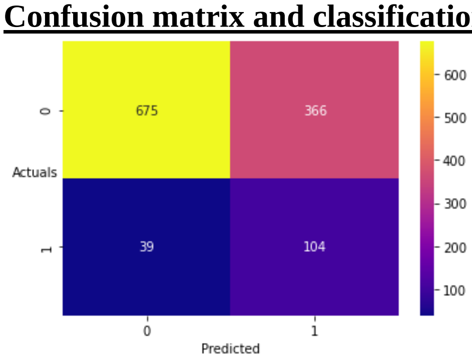
	precision	recall	f1-score	support
0.0	0.944	0.677	0.788	2157
1.0	0.185	0.645	0.287	245
accuracy			0.674	2402
macro avg	0.564	0.661	0.538	2402
weighted avg	0.866	0.674	0.737	2402

As observed above, accuracy of the model i.e. %overall correct predictions is 67%
Sensitivity of the model is 65% i.e. 65% of those defaulted were correctly identified as defaulters by the model

Prediction on Test set

Checking the accuracy of the model using confusion matrix for test set:

- **Confusion matrix and classification report for train data**



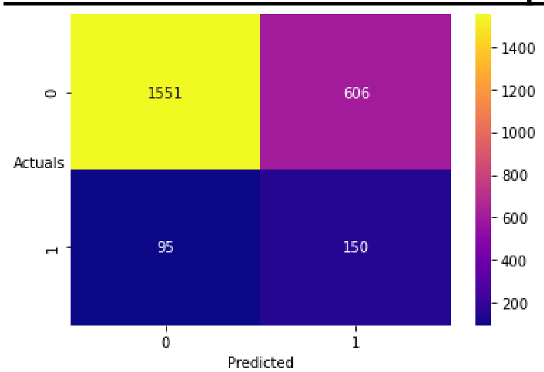
	precision	recall	f1-score	support
0.0	0.945	0.648	0.769	1041
1.0	0.221	0.727	0.339	143
accuracy			0.658	1184
macro avg	0.583	0.688	0.554	1184
weighted avg	0.858	0.658	0.717	1184

As observed above, accuracy of the model i.e. %overall correct predictions is 66%
Sensitivity of the model is 73% i.e. 73% of those defaulted were correctly identified as defaulters by the model

Let us take a cut-off of 0.1156 and check if our predictions have improved

Checking the accuracy of the model using confusion matrix for training set:

- **Confusion matrix and classification report for train data**

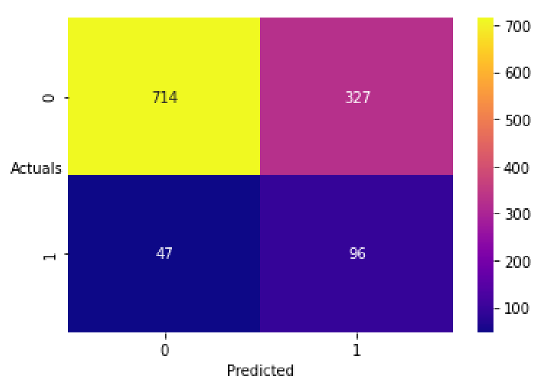


	precision	recall	f1-score	support
0.0	0.942	0.719	0.816	2157
1.0	0.198	0.612	0.300	245
accuracy			0.708	2402
macro avg	0.570	0.666	0.558	2402
weighted avg	0.866	0.708	0.763	2402

Accuracy of the model i.e. %overall correct predictions has increased from 67% to 71% but sensitivity of the model has dropped slightly from 65% to 61%

Prediction on Test set
Checking the accuracy of the model using confusion matrix for test set:

- **Confusion matrix and classification report for train data**



	precision	recall	f1-score	support
0.0	0.938	0.686	0.792	1041
1.0	0.227	0.671	0.339	143
accuracy			0.684	1184
macro avg	0.583	0.679	0.566	1184
weighted avg	0.852	0.684	0.738	1184

Accuracy of the model i.e. %overall correct predictions is 68% & sensitivity of the model stands at 67%
We may choose cutoff of 0.1156 as it gave higher model sensitivity & overall accuracy of the model in test dataset

Interpretation of model 2:

- Of many variables – significantly only 6 variables contribute to the company being predicted as default or not from logistic regression point of view.
- The model is likely to predict the 67% companies that could default correctly.
- Which means only in 33% cases – it could happen that a company that is predicted as defaulter may not be a defaulter but forms an investor point of view – it is ok to no invests money on company that could likely not default.
- The precision is a bit less in this model – however still 22% times, the model will predict the defaulter company correctly.
- From Multi-variate Analysis, we observed that many companies had good profit margins before considering taxes, interests, and other costs.
- But once all costs are considered along-with taxes and depreciation, majority of these companies slide to the bottom half in Profitability.
- These companies should focus on optimizing their bottom line.

THE END