Business Report

SMDM Project Business Report DSBA

Sanjay Srinivasan
PGP-DSBA Online
JULY' 21 Batch
Date: 15-05-2022

INDEX

List Of Figures

Problem - 1

Summary

The data is gathered based on the company financial balance sheet, which deals with the company finances. This dataset has financial statements for 3586 company with 67 variables. For investing in the company, to analyse from the investor's point of view, to predict that the company is capable of handling the financial obligation, can grow quickly and manage the growth scale.

Introduction

The purpose of this exercise is to find the company with good credit rating and handling the financial obligation.

Data Description

#	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
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

Sample of the dataset:

	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

Fig 1.1 Dataset Sample Before Changing Column Names

	Networth_Next_Year	Equity_Paid_Up	Networth	Capital_Employed	Total_Debt	Gross_Block	Net_Working_Capital	Curr_Assets	Curr_Liab_and_Prov
0	-8021.60	419.36	-7027.48	-1007.24	5936.03	474.30	-1076.34	40.50	1116.85
1	-3986.19	1954.93	-2968.08	4458.20	7410.18	9070.86	-1098.88	486.86	1585.74
2	-3192.58	53.84	506.86	7714.68	6944.54	1281.54	4496.25	9097.64	4601.39
3	-3054.51	157.30	-623.49	2353.88	2326.05	1033.69	-2612.42	1034.12	3646.54
4	-2967.36	50.30	-1070.83	4675.33	5740.90	1084.20	1836.23	4685.81	2849.58

Fig 1.2 Dataset Sample After Changing Column Names

Exploratory Data Analysis

Let us check the types of variables in the data frame.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3586 entries, 0 to 3585
Data columns (total 67 columns):
Co_Code                                3586 non-null int64
Co_Name                                3586 non-null object
Networth Next Year                     3586 non-null float64
Equity Paid Up                         3586 non-null float64
Networth                               3586 non-null float64
Capital Employed                       3586 non-null float64
Total Debt                             3586 non-null float64
Gross Block                            3586 non-null float64
Net Working Capital                    3586 non-null float64
Current Assets                         3586 non-null float64
Current Liabilities and Provisions     3586 non-null float64
Total Assets/Liabilities               3586 non-null float64
Gross Sales                            3586 non-null float64
Net Sales                              3586 non-null float64
Other Income                           3586 non-null float64
Value Of Output                        3586 non-null float64
Cost of Production                     3586 non-null float64
Selling Cost                           3586 non-null float64
PBIDT                                  3586 non-null float64
PBDT                                   3586 non-null float64
PBIT                                   3586 non-null float64
PBT                                    3586 non-null float64
PAT                                    3586 non-null float64
Adjusted PAT                           3586 non-null float64
CP                                     3586 non-null float64
Revenue earnings in forex              3586 non-null float64
Revenue expenses in forex              3586 non-null float64
Capital expenses in forex              3586 non-null float64
Book Value (Unit Curr)                 3586 non-null float64
Book Value (Adj.) (Unit Curr)          3582 non-null float64
Market Capitalisation                  3586 non-null float64
CEPS (annualised) (Unit Curr)          3586 non-null float64
Cash Flow From Operating Activities    3586 non-null float64
Cash Flow From Investing Activities    3586 non-null float64
```

Fig- 1.3. Sample Datatypes of the variable with null values

There are total 3586 rows and 67 columns in the dataset.

1.1 Outlier Treatment

The boxplot is plotted for all the variable without treating the outliers.

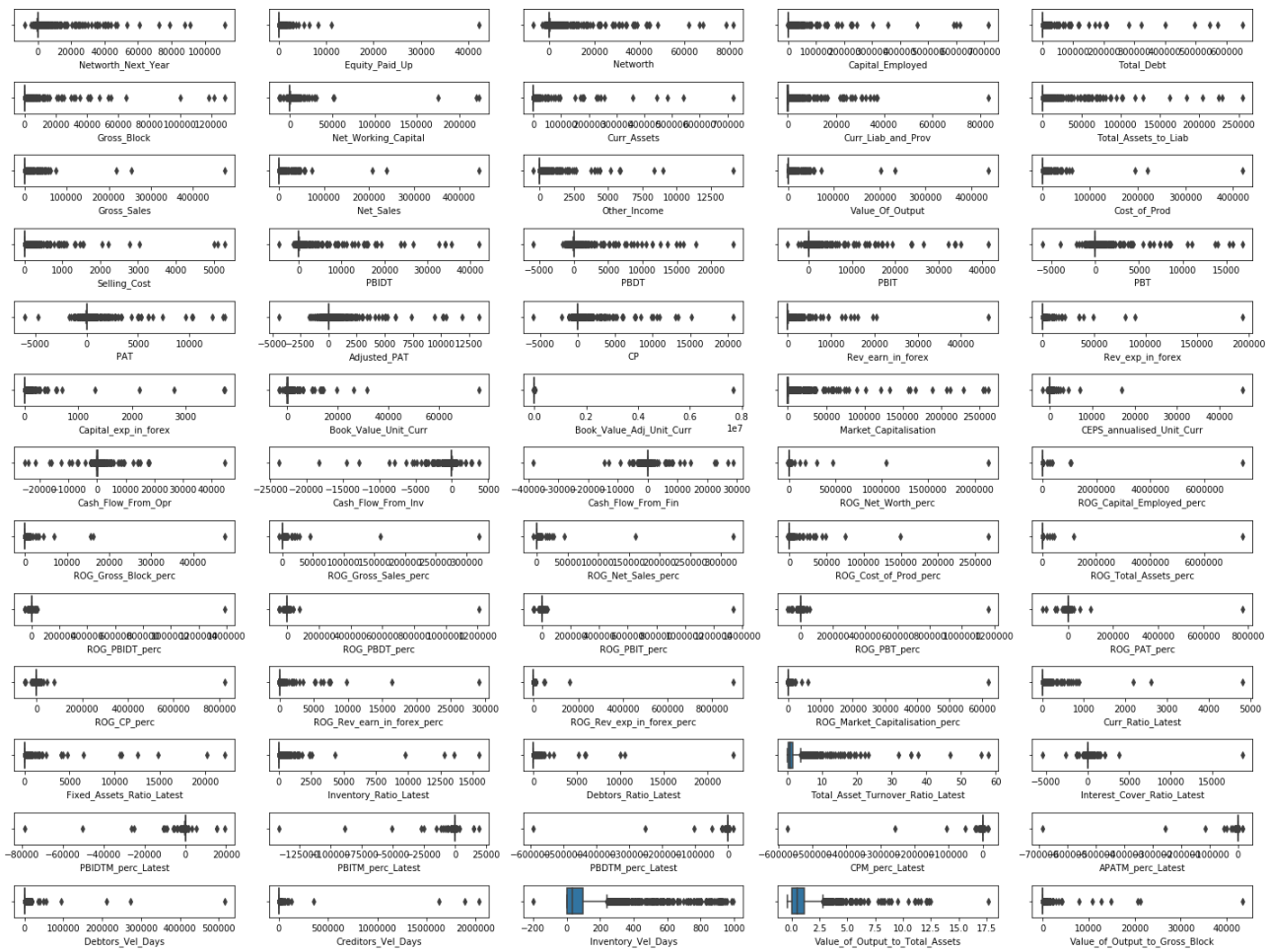


Fig- 1.4 Shape before Outliers Treatment

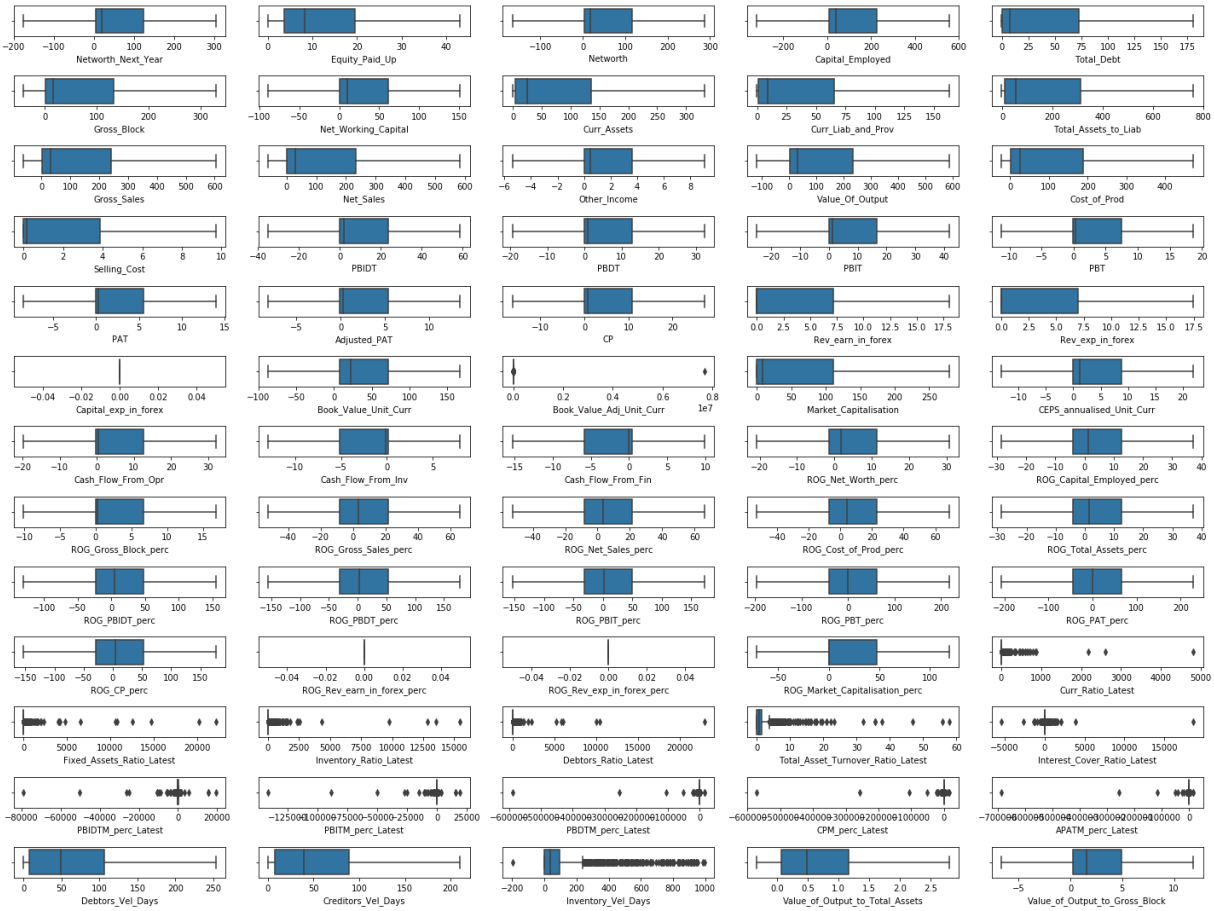


Fig- 1.5 Shape After Outliers Treatment

1.2 Missing Value Treatment

Fig- 1.4 Shape After Outliers Treatment

Co_Code	0
Co_Name	0
Networth_Next_Year	0
Equity_Paid_Up	0
Networth	0
Capital_Employed	0
Total_Debt	0
Gross_Block	0
Net_Working_Capital	0
Curr_Assets	0
Curr_Liab_and_Prov	0
Total_Assets_to_Liab	0
Gross_Sales	0
Net_Sales	0
Other_Income	0
Value_Of_Output	0
Cost_of_Prod	0
Selling_Cost	0
PBIDT	0
PBDT	0
PBIT	0
PBT	0
PAT	0
Adjusted_PAT	0
CP	0
Rev_earn_in_forex	0
Rev_exp_in_forex	0
Capital_exp_in_forex	0
Book_Value_Unit_Curr	0
Book_Value_Adj_Unit_Curr	4
...	
ROG_Gross_Block_perc	0
ROG_Gross_Sales_perc	0
ROG_Net_Sales_perc	0
ROG_Cost_of_Prod_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_Rev_earn_in_forex_perc	0

Fig- 1.6 Before Treating Missing value

Networth_Next_Year	0
Equity_Paid_Up	0
Networth	0
Capital_Employed	0
Total_Debt	0
Gross_Block	0
Net_Working_Capital	0
Curr_Assets	0
Curr_Liab_and_Prov	0
Total_Assets_to_Liab	0
Gross_Sales	0
Net_Sales	0
Other_Income	0
Value_Of_Output	0
Cost_of_Prod	0
Selling_Cost	0
PBIDT	0
PBDT	0
PBIT	0
PBT	0
PAT	0
Adjusted_PAT	0
CP	0
Rev_earn_in_forex	0
Rev_exp_in_forex	0
Capital_exp_in_forex	0
Book_Value_Unit_Curr	0
Book_Value_Adj_Unit_Curr	0
Market_Capitalisation	0
CEPS_annualised_Unit_Curr	0
..	
ROG_Gross_Block_perc	0
ROG_Gross_Sales_perc	0
ROG_Net_Sales_perc	0
ROG_Cost_of_Prod_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_Rev_earn_in_forex_perc	0

Fig- 1.7 After Treating Missing value

1.3 Transform Target variable into 0 and 1.

Target value 'Networth_Next_year' is transform into 0's and 1's.

Networth_Next_year < 0 (negative) then target or default variable = 1
Networth_Next_year > 0 (positive) then target or default variable = 0

1 - Company might default.
0 – Company might not default.

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

Fig – 1.8 Default count.

```
0    0.891801
1    0.108199
Name: default, dtype: float64
```

Fig – 1.9 Default count in percentage.

Latest	APATM_perc_Latest	Debtors_Vel_Days	Creditors_Vel_Days	Inventory_Vel_Days	Value_of_Output_to_Total_Assets	Value_of_Output_to_Gross_Block	default
0.00	0.00	0.0	0.0	45.0	0.00	0.00	1
-57.74	-87.18	29.0	101.0	2.0	0.31	0.24	1
723.67	-7961.51	97.0	210.5	0.0	-0.03	-0.26	1
-47.70	-51.58	93.0	63.0	2.0	0.24	1.90	1
379.79	274.79	253.0	210.5	0.0	0.01	0.05	1

Fig – 1.10 Sample data after Transformation.

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)

Uni-Variate Analysis:

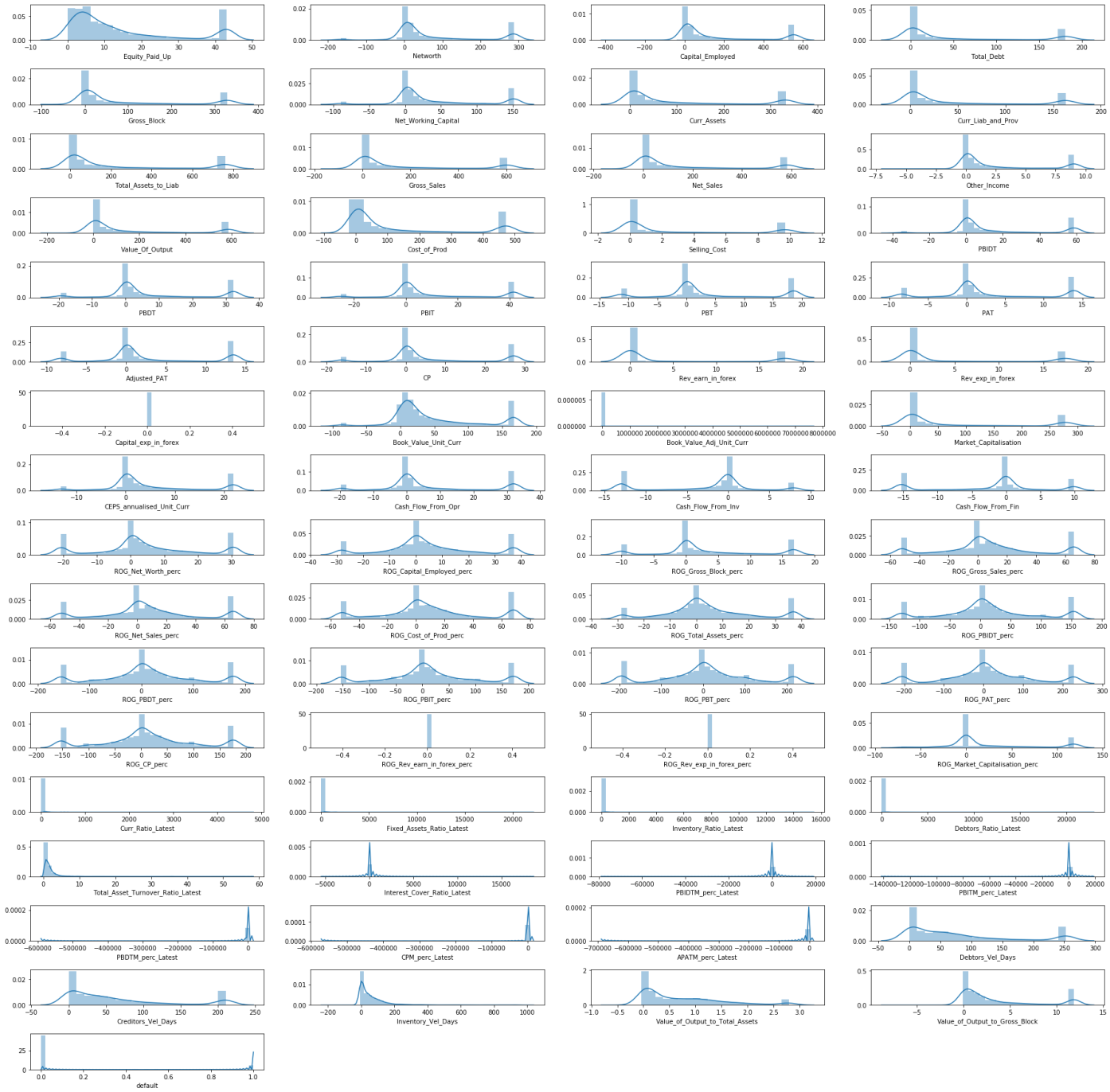


Fig – 1.11 Univariate Analysis

From the above chart (displot and boxplot), there are outliers present in the economic.cond.national and economic.cond.household data. We can infer that there is no trend or pattern that it follows a normal distribution.

Bi – variate Analysis:

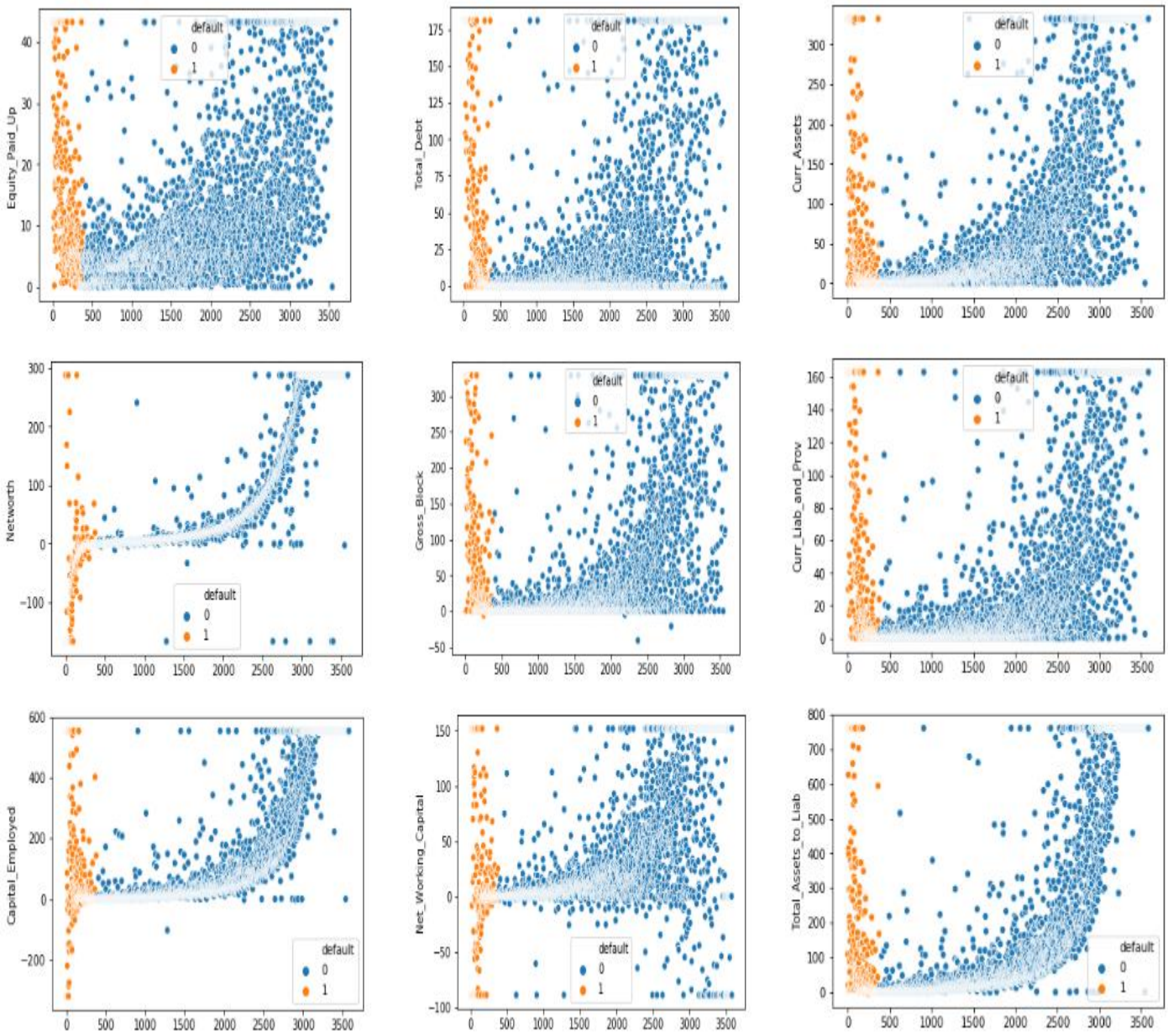


Fig – 1.12 Scatterplot for Bivariate Analysis

Multi – variate Analysis:

	Equity_Paid_Up	Networth	Capital_Employed	Total_Debt	Gross_Block	Net_Working_Capital	Curr_Assets	Curr_Liab_and_Prov
Equity_Paid_Up	1.000000	0.575311	0.678049	0.573822	0.615089	0.388607	0.631646	0.646325
Networth	0.575311	1.000000	0.873306	0.535012	0.665375	0.623119	0.760024	0.683618
Capital_Employed	0.678049	0.873306	1.000000	0.781394	0.824900	0.688619	0.902837	0.834573
Total_Debt	0.573822	0.535012	0.781394	1.000000	0.781595	0.574341	0.798859	0.780363
Gross_Block	0.615089	0.665375	0.824900	0.781595	1.000000	0.527678	0.814709	0.850400
Net_Working_Capital	0.388607	0.623119	0.688619	0.574341	0.527678	1.000000	0.761698	0.554177
Curr_Assets	0.631646	0.760024	0.902837	0.798859	0.814709	0.761698	1.000000	0.912895
Curr_Liab_and_Prov	0.646325	0.683618	0.834573	0.780363	0.850400	0.554177	0.912895	1.000000
Total_Assets_to_Liab	0.695357	0.836822	0.977779	0.806901	0.856130	0.657365	0.934370	0.906441
Gross_Sales	0.564579	0.721942	0.825233	0.727858	0.833787	0.645461	0.885011	0.866521
Net_Sales	0.565780	0.723721	0.827319	0.728285	0.832207	0.646346	0.885998	0.866431
Other_Income	0.547215	0.663716	0.741598	0.600497	0.721992	0.509973	0.744745	0.739434
Value_Of_Output	0.565715	0.727349	0.827751	0.727520	0.830015	0.647768	0.886111	0.866309
Cost_of_Prod	0.537617	0.673156	0.792087	0.718135	0.836093	0.634337	0.864821	0.847112
Selling_Cost	0.449257	0.593682	0.665321	0.595529	0.715865	0.513265	0.705795	0.700329
PBIDT	0.455246	0.787430	0.766376	0.573574	0.691599	0.576149	0.743035	0.702935
PBDT	0.318126	0.696443	0.593961	0.354797	0.533308	0.471505	0.571838	0.521224
PBIT	0.382343	0.741600	0.689721	0.486828	0.587191	0.540258	0.672483	0.624369
PBT	0.235951	0.618249	0.488696	0.234827	0.392171	0.412861	0.476762	0.417439
PAT	0.236287	0.619903	0.492140	0.238806	0.393956	0.412756	0.478623	0.418445
Adjusted_PAT	0.235459	0.616226	0.478390	0.224134	0.378879	0.402873	0.471071	0.407710
CP	0.324023	0.705317	0.606486	0.368987	0.547914	0.477619	0.581137	0.529975
Rev_earn_in_forex	0.296983	0.449846	0.508536	0.444829	0.565897	0.437216	0.538245	0.512520
Rev_exp_in_forex	0.382239	0.533567	0.604157	0.527873	0.654415	0.506922	0.646190	0.633391
Capital_exp_in_forex	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Book_Value_Unit_Curr	0.074402	0.592468	0.470570	0.241273	0.341511	0.394028	0.422223	0.347588

Fig – 1.13 Sample Multivariate analysis for correlation

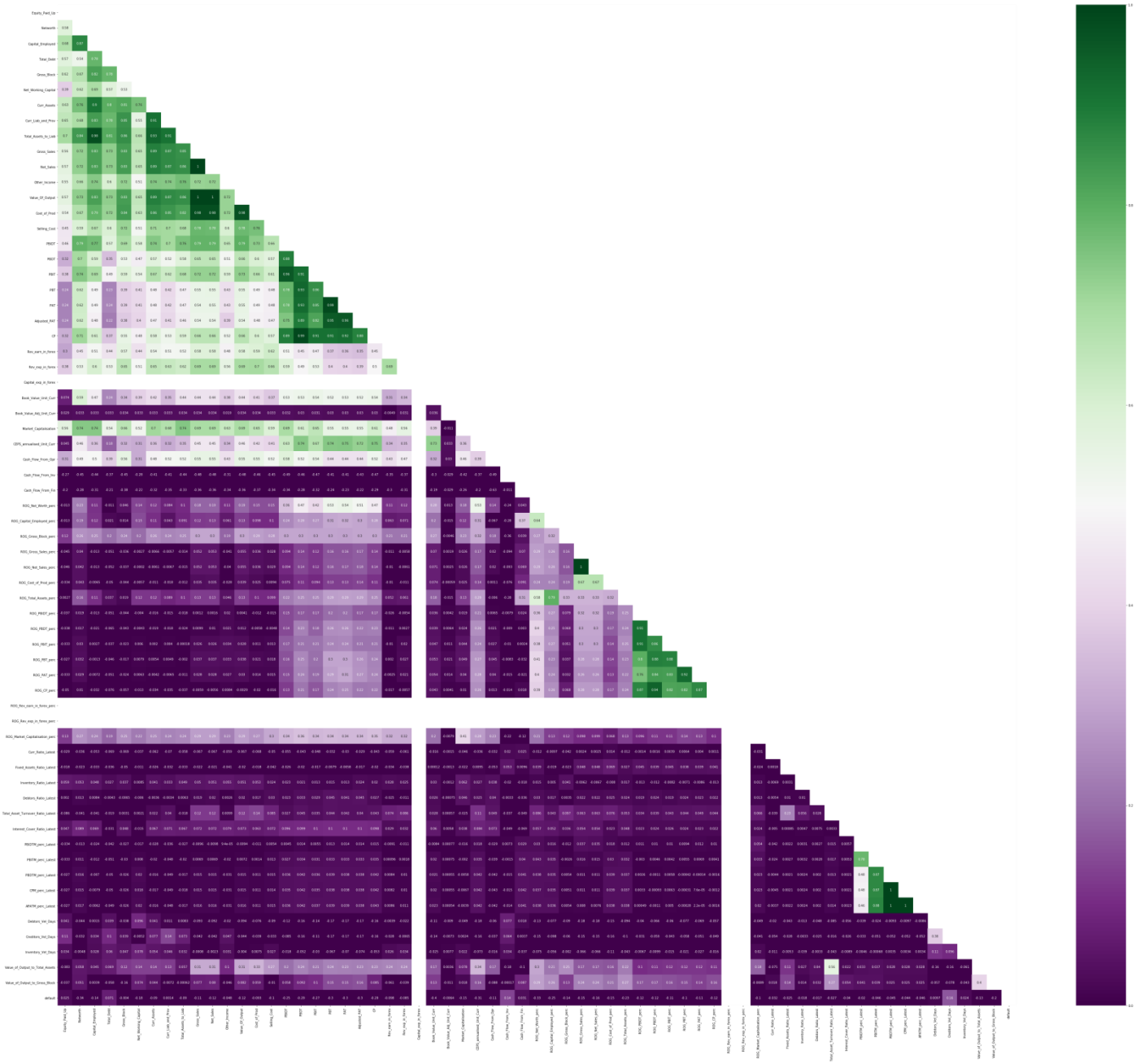


Fig – 1.14 Multivariate analysis of plotting correlation in heatmap

From this Heatmap we can infer that 3 variables do not have any correlation and do not contribute on the output. So, dropping the insignificant variables.

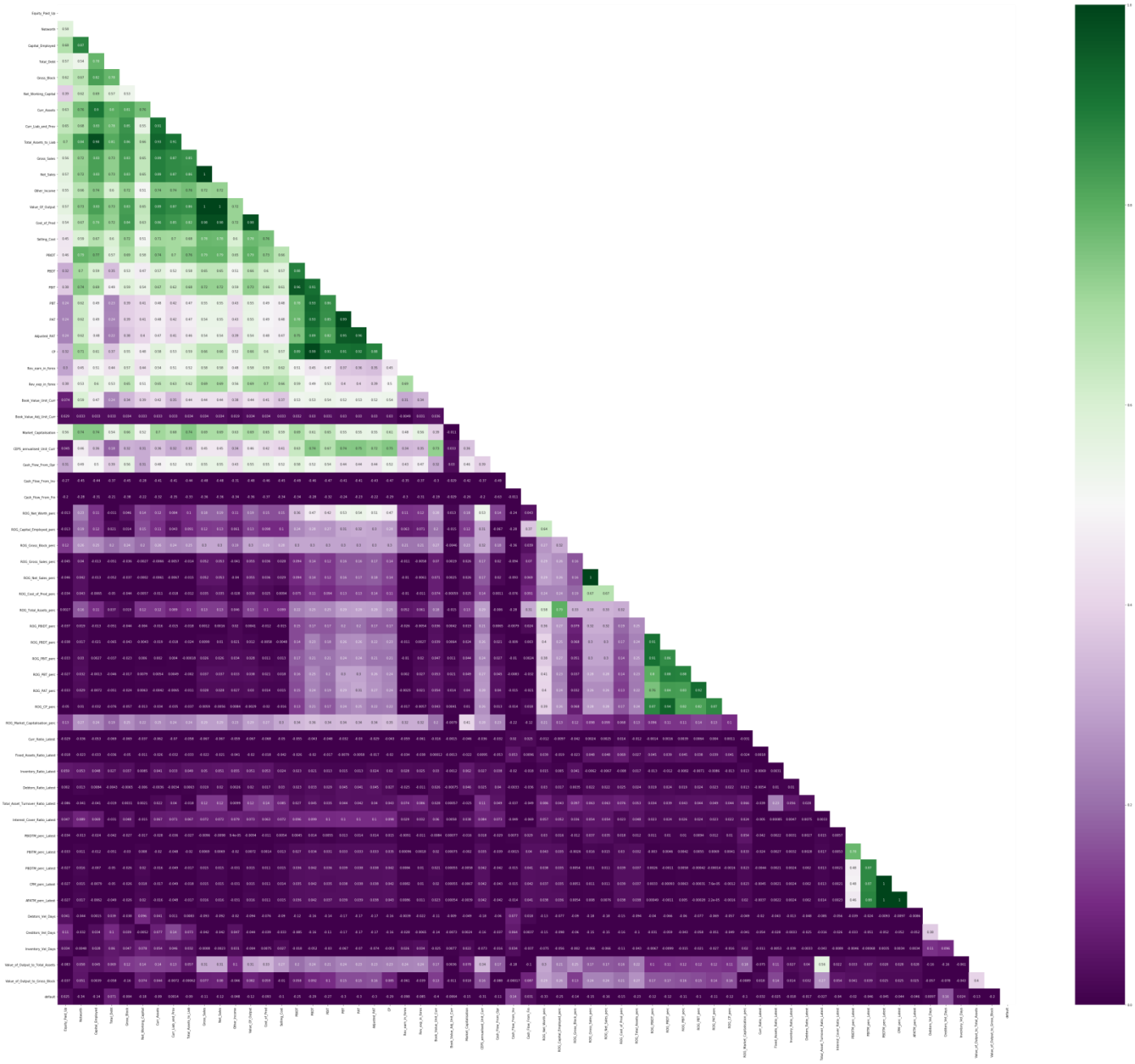


Fig – 1.15 Multivariate analysis of plotting correlation in heatmap after dropping insignificant variable.

```
These are the highly correlated variables in the dataset with more than 90% of correlation

['Curr_Assets', 'Curr_Liab_and_Prov', 'Total_Assets_to_Liab', 'Net_Sales', 'Value_Of_Output', 'Cost_of_Prod', 'PBIT', 'PBT', 'P
AT', 'Adjusted_PAT', 'CP', 'ROG_Net_Sales_perc', 'ROG_PBDT_perc', 'ROG_PBIT_perc', 'ROG_PAT_perc', 'ROG_CP_perc', 'CPM_perc_Lat
est', 'APATM_perc_Latest']
```

Fig – 1.16 Taking Highly correlated variables.

1.5 Train Test Split

The shape of the dataset after splitting the train and test data. The train data and test data are splitted in the ratio of 67:33 with the random state – 42.

```
(2402, 43)
```

Fig – 1.17 X-train shape.

```
(1184, 43)
```

Fig – 1.18 X-test shape.

```
(2402,)
```

Fig – 1.19 Y-train shape

```
(1184,)
```

Fig – 1.20 Y-test shape

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.

	Feature	Rank
1	Networth	1
2	Capital_Employed	1
3	Total_Debt	1
8	Selling_Cost	1
9	PBIDT	1
12	Rev_exp_in_forex	1
13	Book_Value_Unit_Curr	1
15	Market_Capitalisation	1
16	CEPS_annualised_Unit_Curr	1
20	ROG_Net_Worth_perc	1
29	Curr_Ratio_Latest	1
31	Inventory_Ratio_Latest	1
32	Debtors_Ratio_Latest	1
34	Interest_Cover_Ratio_Latest	1
42	Value_of_Output_to_Gross_Block	1

Fig – 1.21 Selecting the feature with rank 1

Model - 1

Logit Regression Results

Dep. Variable:	default	**No. Observations:**	2402
Model:	Logit	**Df Residuals:**	2386
Method:	MLE	**Df Model:**	15
Date:	Sun, 08 May 2022	**Pseudo R-squ.:**	0.5980
Time:	15:33:23	**Log-Likelihood:**	-331.01
converged:	True	**LL-Null:**	-823.47
		LLR p-value:	2.301e-200

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-7.5240	0.505	-14.887	0.000	-8.515	-6.533
Networth	-0.6488	0.417	-1.557	0.119	-1.465	0.168
Capital_Employed	-0.5778	0.566	-1.020	0.308	-1.688	0.532
Total_Debt	1.3585	0.375	3.622	0.000	0.623	2.094
Selling_Cost	-0.3243	0.278	-1.168	0.243	-0.868	0.220
PBIDT	-0.5765	0.329	-1.752	0.080	-1.221	0.068
Rev_exp_in_forex	0.3099	0.226	1.372	0.170	-0.133	0.753
Book_Value_Unit_Curr	-6.0852	0.643	-9.460	0.000	-7.346	-4.824
Market_Capitalisation	-0.5763	0.307	-1.880	0.060	-1.177	0.024
CEPS_annualised_Unit_Curr	-0.4984	0.354	-1.406	0.160	-1.193	0.196
ROG_Net_Worth_perc	-0.4011	0.132	-3.045	0.002	-0.659	-0.143
Curr_Ratio_Latest	-0.6999	0.651	-1.074	0.283	-1.977	0.577
Inventory_Ratio_Latest	-1.5284	1.127	-1.356	0.175	-3.738	0.681
Debtors_Ratio_Latest	-1.1137	1.821	-0.612	0.541	-4.683	2.455
Interest_Cover_Ratio_Latest	-0.4314	0.329	-1.312	0.190	-1.076	0.213
Value_of_Output_to_Gross_Block	-0.4849	0.160	-3.027	0.002	-0.799	-0.171

Possibly complete quasi-separation: A fraction 0.31 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

Fig – 1.22 Model-1 summary report

	variables	VIF
1	Capital_Employed	10.551510
0	Networth	7.834042
4	PBIDT	4.958060
8	CEPS_annualised_Unit_Curr	3.867756
2	Total_Debt	3.726912
6	Book_Value_Unit_Curr	2.883423
7	Market_Capitalisation	2.637387
3	Selling_Cost	2.561778
5	Rev_exp_in_forex	2.035821
9	ROG_Net_Worth_perc	1.669240
14	Value_of_Output_to_Gross_Block	1.119548
13	Interest_Cover_Ratio_Latest	1.058831
12	Debtors_Ratio_Latest	1.013745
11	Inventory_Ratio_Latest	1.013311
10	Curr_Ratio_Latest	1.007535

Fig – 1.23 Variance Inflation Factor.of Model-1

The capital Employed has the highest vif and p-value is greater than the alpha value(0.05), capital_employed variable is dropped.

Model-2

Logit Regression Results

Dep. Variable:	default	**No. Observations:**	2402
Model:	Logit	**Df Residuals:**	2387
Method:	MLE	**Df Model:**	14
Date:	Sun, 08 May 2022	**Pseudo R-squ.:**	0.5974
Time:	15:33:24	**Log-Likelihood:**	-331.54
converged:	True	**LL-Null:**	-823.47
		LLR p-value:	4.540e-201

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-7.5470	0.510	-14.811	0.000	-8.546	-6.548
Networth	-0.8686	0.352	-2.470	0.014	-1.558	-0.179
Total_Debt	1.0725	0.242	4.431	0.000	0.598	1.547
Selling_Cost	-0.3760	0.279	-1.345	0.178	-0.924	0.172
PBIDT	-0.5928	0.330	-1.799	0.072	-1.239	0.053
Rev_exp_in_forex	0.3181	0.225	1.411	0.158	-0.124	0.760
Book_Value_Unit_Curr	-6.1148	0.640	-9.549	0.000	-7.370	-4.860
Market_Capitalisation	-0.6607	0.300	-2.204	0.028	-1.248	-0.073
CEPS_annualised_Unit_Curr	-0.4813	0.352	-1.368	0.171	-1.171	0.208
ROG_Net_Worth_perc	-0.3968	0.131	-3.026	0.002	-0.654	-0.140
Curr_Ratio_Latest	-0.7180	0.657	-1.093	0.274	-2.005	0.569
Inventory_Ratio_Latest	-1.7831	1.198	-1.488	0.137	-4.132	0.566
Debtors_Ratio_Latest	-1.0703	1.814	-0.590	0.555	-4.626	2.485
Interest_Cover_Ratio_Latest	-0.4291	0.331	-1.297	0.195	-1.078	0.219
Value_of_Output_to_Gross_Block	-0.4794	0.159	-3.006	0.003	-0.792	-0.167

Possibly complete quasi-separation: A fraction 0.32 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

Fig – 1.24 Model-2 Summary Report

Debtors_Ratio_Latest has the highest p-value and is insignificant, therefore, we need to eliminate it.

Model 3:

Logit Regression Results

Dep. Variable:	default	**No. Observations:**	2402
Model:	Logit	**Df Residuals:**	2388
Method:	MLE	**Df Model:**	13
Date:	Sun, 08 May 2022	**Pseudo R-squ.:**	0.5971
Time:	15:33:24	**Log-Likelihood:**	-331.81
converged:	True	**LL-Null:**	-823.47
		LLR p-value:	6.691e-202

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-7.5077	0.502	-14.955	0.000	-8.492	-6.524
Networth	-0.8494	0.351	-2.419	0.016	-1.538	-0.161
Total_Debt	1.0688	0.243	4.403	0.000	0.593	1.545
Selling_Cost	-0.3801	0.278	-1.366	0.172	-0.925	0.165
PBIDT	-0.5753	0.330	-1.742	0.081	-1.222	0.072
Rev_exp_in_forex	0.3218	0.225	1.431	0.152	-0.119	0.762
Book_Value_Unit_Curr	-6.1340	0.641	-9.568	0.000	-7.391	-4.877
Market_Capitalisation	-0.6840	0.294	-2.325	0.020	-1.261	-0.107
CEPS_annualised_Unit_Curr	-0.4861	0.353	-1.379	0.168	-1.177	0.205
ROG_Net_Worth_perc	-0.4018	0.131	-3.065	0.002	-0.659	-0.145
Curr_Ratio_Latest	-0.7118	0.654	-1.088	0.277	-1.994	0.571
Inventory_Ratio_Latest	-1.8011	1.171	-1.538	0.124	-4.097	0.495
Interest_Cover_Ratio_Latest	-0.4319	0.329	-1.314	0.189	-1.076	0.212
Value_of_Output_to_Gross_Block	-0.4848	0.158	-3.070	0.002	-0.794	-0.175

Possibly complete quasi-separation: A fraction 0.32 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

Fig – 1.25 Model-3 Summary Report

Curr_Ratio_Latest has the highest p-value and is insignificant, therefore, we need to eliminate it.

Model 4:

Logit Regression Results

Dep. Variable:	default	**No. Observations:**	2402
Model:	Logit	**Df Residuals:**	2389
Method:	MLE	**Df Model:**	12
Date:	Sun, 08 May 2022	**Pseudo R-squ.:**	0.5958
Time:	15:33:24	**Log-Likelihood:**	-332.83
converged:	True	**LL-Null:**	-823.47
		LLR p-value:	1.986e-202

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-7.5509	0.506	-14.928	0.000	-8.542	-6.560
Networth	-0.8387	0.353	-2.379	0.017	-1.530	-0.148
Total_Debt	1.0896	0.243	4.480	0.000	0.613	1.566
Selling_Cost	-0.3819	0.280	-1.362	0.173	-0.932	0.168
PBIDT	-0.5714	0.332	-1.722	0.085	-1.222	0.079
Rev_exp_in_forex	0.3336	0.227	1.472	0.141	-0.111	0.778
Book_Value_Unit_Curr	-6.2499	0.641	-9.747	0.000	-7.507	-4.993
Market_Capitalisation	-0.6961	0.293	-2.375	0.018	-1.271	-0.122
CEPS_annualised_Unit_Curr	-0.4864	0.355	-1.370	0.171	-1.182	0.209
ROG_Net_Worth_perc	-0.4182	0.132	-3.178	0.001	-0.676	-0.160
Inventory_Ratio_Latest	-1.7889	1.194	-1.498	0.134	-4.129	0.551
Interest_Cover_Ratio_Latest	-0.4436	0.324	-1.369	0.171	-1.078	0.191
Value_of_Output_to_Gross_Block	-0.4809	0.157	-3.061	0.002	-0.789	-0.173

Possibly complete quasi-separation: A fraction 0.32 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

Fig – 1.26 Model-4 Summary Report

Selling_Cost has the highest p-value and is insignificant, therefore, we need to eliminate it.

Model – 5:

Logit Regression Results

Dep. Variable:	default	**No. Observations:**	2402
Model:	Logit	**Df Residuals:**	2390
Method:	MLE	**Df Model:**	11
Date:	Sun, 08 May 2022	**Pseudo R-squ.:**	0.5946
Time:	15:33:24	**Log-Likelihood:**	-333.80
converged:	True	**LL-Null:**	-823.47
		LLR p-value:	5.379e-203

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-7.5288	0.505	-14.918	0.000	-8.518	-6.540
Networth	-0.8268	0.353	-2.343	0.019	-1.519	-0.135
Total_Debt	0.9580	0.225	4.258	0.000	0.517	1.399
PBIDT	-0.5998	0.329	-1.824	0.068	-1.244	0.045
Rev_exp_in_forex	0.2174	0.213	1.022	0.307	-0.199	0.634
Book_Value_Unit_Curr	-6.3105	0.645	-9.783	0.000	-7.575	-5.046
Market_Capitalisation	-0.7353	0.293	-2.511	0.012	-1.309	-0.161
CEPS_annualised_Unit_Curr	-0.4640	0.352	-1.319	0.187	-1.153	0.226
ROG_Net_Worth_perc	-0.3982	0.130	-3.070	0.002	-0.653	-0.144
Inventory_Ratio_Latest	-1.9713	1.267	-1.555	0.120	-4.455	0.513
Interest_Cover_Ratio_Latest	-0.4464	0.328	-1.361	0.173	-1.089	0.196
Value_of_Output_to_Gross_Block	-0.4859	0.159	-3.059	0.002	-0.797	-0.175

Possibly complete quasi-separation: A fraction 0.32 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

Fig – 1.27 Model-5 Summary Report

Rev_exp_in_forex has the highest p-value and is insignificant, therefore, we need to eliminate it.

Model 6:

Logit Regression Results

Dep. Variable:	default	**No. Observations:**	2402
Model:	Logit	**Df Residuals:**	2391
Method:	MLE	**Df Model:**	10
Date:	Sun, 08 May 2022	**Pseudo R-squ.:**	0.5940
Time:	15:33:25	**Log-Likelihood:**	-334.32
converged:	True	**LL-Null:**	-823.47
		LLR p-value:	8.798e-204

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-7.5167	0.504	-14.910	0.000	-8.505	-6.529
Networth	-0.8178	0.349	-2.344	0.019	-1.502	-0.134
Total_Debt	1.0583	0.202	5.238	0.000	0.662	1.454
PBIDT	-0.5462	0.323	-1.688	0.091	-1.180	0.088
Book_Value_Unit_Curr	-6.2788	0.642	-9.777	0.000	-7.537	-5.020
Market_Capitalisation	-0.7145	0.287	-2.490	0.013	-1.277	-0.152
CEPS_annualised_Unit_Curr	-0.4606	0.351	-1.312	0.190	-1.149	0.228
ROG_Net_Worth_perc	-0.4063	0.130	-3.128	0.002	-0.661	-0.152
Inventory_Ratio_Latest	-2.0299	1.288	-1.576	0.115	-4.554	0.494
Interest_Cover_Ratio_Latest	-0.4404	0.332	-1.325	0.185	-1.092	0.211
Value_of_Output_to_Gross_Block	-0.4748	0.157	-3.018	0.003	-0.783	-0.166

Possibly complete quasi-separation: A fraction 0.31 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

Fig – 1.28 Model-6 Summary Report

CEPS_annualised_Unit_Curr has the highest p-value and is insignificant, therefore, we need to eliminate it.

Model 7:

Logit Regression Results

Dep. Variable:	default	**No. Observations:**	2402
Model:	Logit	**Df Residuals:**	2392
Method:	MLE	**Df Model:**	9
Date:	Sun, 08 May 2022	**Pseudo R-squ.:**	0.5930
Time:	15:33:25	**Log-Likelihood:**	-335.18
converged:	True	**LL-Null:**	-823.47
		LLR p-value:	1.928e-204

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-7.3533	0.484	-15.183	0.000	-8.303	-6.404
Networth	-0.8102	0.353	-2.292	0.022	-1.503	-0.118
Total_Debt	1.1250	0.199	5.641	0.000	0.734	1.516
PBIDT	-0.7685	0.286	-2.685	0.007	-1.329	-0.208
Book_Value_Unit_Curr	-6.2193	0.638	-9.750	0.000	-7.470	-4.969
Market_Capitalisation	-0.7431	0.286	-2.596	0.009	-1.304	-0.182
ROG_Net_Worth_perc	-0.4692	0.122	-3.847	0.000	-0.708	-0.230
Inventory_Ratio_Latest	-1.9878	1.292	-1.538	0.124	-4.521	0.545
Interest_Cover_Ratio_Latest	-0.4351	0.326	-1.336	0.182	-1.073	0.203
Value_of_Output_to_Gross_Block	-0.4700	0.158	-2.977	0.003	-0.779	-0.161

Possibly complete quasi-separation: A fraction 0.31 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

Fig – 1.29 Model-7 Summary Report

Interest_Cover_Ratio_Latest has the highest p-value and is insignificant, therefore, we need to eliminate it.

Model 8:

Logit Regression Results

Dep. Variable:	default	**No. Observations:**	2402
Model:	Logit	**Df Residuals:**	2393
Method:	MLE	**Df Model:**	8
Date:	Sun, 08 May 2022	**Pseudo R-squ.:**	0.5926
Time:	15:33:25	**Log-Likelihood:**	-335.46
converged:	True	**LL-Null:**	-823.47
		LLR p-value:	2.242e-205

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-7.3385	0.483	-15.186	0.000	-8.286	-6.391
Networth	-0.8121	0.354	-2.297	0.022	-1.505	-0.119
Total_Debt	1.1298	0.199	5.677	0.000	0.740	1.520
PBIDT	-0.7743	0.286	-2.712	0.007	-1.334	-0.215
Book_Value_Unit_Curr	-6.2276	0.640	-9.732	0.000	-7.482	-4.973
Market_Capitalisation	-0.7470	0.286	-2.608	0.009	-1.308	-0.186
ROG_Net_Worth_perc	-0.4691	0.122	-3.852	0.000	-0.708	-0.230
Inventory_Ratio_Latest	-1.9615	1.283	-1.529	0.126	-4.475	0.552
Value_of_Output_to_Gross_Block	-0.4720	0.158	-2.990	0.003	-0.781	-0.163

Possibly complete quasi-separation: A fraction 0.31 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

Fig – 1.30 Model-8 Summary Report

Inventory_Ratio_Latest has the highest p-value and is insignificant, therefore, we need to eliminate it.

Model – 9:

Logit Regression Results

Dep. Variable:	default	**No. Observations:**	2402
Model:	Logit	**Df Residuals:**	2394
Method:	MLE	**Df Model:**	7
Date:	Sun, 08 May 2022	**Pseudo R-squ.:**	0.5914
Time:	15:33:25	**Log-Likelihood:**	-336.44
converged:	True	**LL-Null:**	-823.47
		LLR p-value:	4.858e-206

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-7.2051	0.466	-15.468	0.000	-8.118	-6.292
Networth	-0.7906	0.351	-2.250	0.024	-1.479	-0.102
Total_Debt	1.1148	0.198	5.620	0.000	0.726	1.504
PBIDT	-0.7480	0.285	-2.625	0.009	-1.306	-0.190
Book_Value_Unit_Curr	-6.2362	0.639	-9.761	0.000	-7.488	-4.984
Market_Capitalisation	-0.7612	0.284	-2.680	0.007	-1.318	-0.205
ROG_Net_Worth_perc	-0.4728	0.121	-3.894	0.000	-0.711	-0.235
Value_of_Output_to_Gross_Block	-0.4841	0.157	-3.080	0.002	-0.792	-0.176

Possibly complete quasi-separation: A fraction 0.30 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

Fig – 1.31 Model-9 Summary Report

Now, all the variables are significant and p-value is less than the alpha value 0.05. Therefore, we don't need to eliminate the other variables.

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

```
The optimum threshold is 0.16890979736726344
```

Fig – 1.32 Optimum threshold

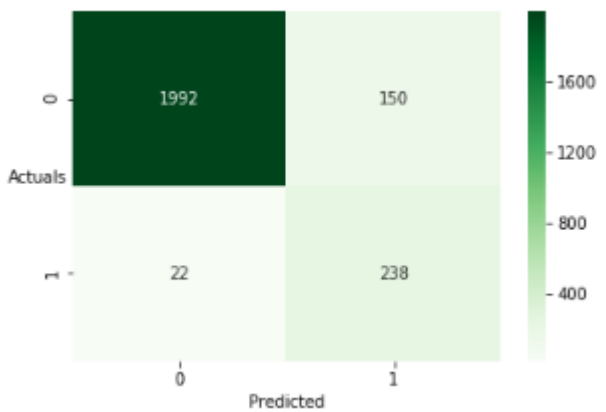


Fig – 1. 33 Confusion matrix for train data

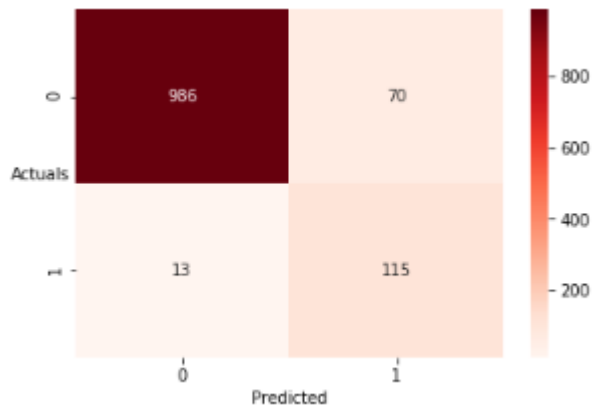


Fig – 1.34 Confusion matrix for test data

	precision	recall	f1-score	support
0	0.989	0.930	0.959	2142
1	0.613	0.915	0.735	260
accuracy			0.928	2402
macro avg	0.801	0.923	0.847	2402
weighted avg	0.948	0.928	0.934	2402

Fig – 1. 35 Classification report for train data

	precision	recall	f1-score	support
0	0.987	0.934	0.960	1056
1	0.622	0.898	0.735	128
accuracy			0.930	1184
macro avg	0.804	0.916	0.847	1184
weighted avg	0.947	0.930	0.935	1184

Fig – 1. 36 Classification report for test data

From the train data and test data we can infer that recall is good for both training and test data classification report .

The test data has 89.8% recall that company might default.

The precision of test data is slightly greater than the train data, test data is slightly over fitting.

1.8 Build a Random Forest Model on Train Dataset. Also showcase your model building approach.

```
RandomForestClassifier(max_depth=35, n_jobs=-1, oob_score=True, random_state=42)
```

Fig – 1. 37 Initializing Random Forest Classifier

	Feature	Rank
0	Networth	1
1	Capital_Employed	1
2	Total_Debt	1
4	PBIDT	1
6	Book_Value_Unit_Curr	1
7	Market_Capitalisation	1
8	CEPS_annualised_Unit_Curr	1
9	ROG_Net_Worth_perc	1
10	Curr_Ratio_Latest	1
12	Debtors_Ratio_Latest	1
13	Interest_Cover_Ratio_Latest	1
14	Value_of_Output_to_Gross_Block	1

Fig – 1.38. Taking features with Rank 1

```
0    2142
1     260
Name: default, dtype: int64
```

Fig – 1.39 value count of the target column.

<u>Model Building of Random forest model:</u>

<u>Model - 1</u>

```
Optimization terminated successfully.
         Current function value: 0.143538
         Iterations 11
```

Fig – 1.40 Initializing Model-1 using RF model

Logit Regression Results

Dep. Variable:	default	**No. Observations:**	2402
Model:	Logit	**Df Residuals:**	2392
Method:	MLE	**Df Model:**	9
Date:	Sun, 15 May 2022	**Pseudo R-squ.:**	0.5813
Time:	16:10:05	**Log-Likelihood:**	-344.78
converged:	True	**LL-Null:**	-823.47
		LLR p-value:	2.660e-200

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-7.2470	0.479	-15.123	0.000	-8.186	-6.308
Networth	-1.2523	0.361	-3.471	0.001	-1.959	-0.545
Capital_Employed	0.8079	0.279	2.897	0.004	0.261	1.355
PBIDT	-0.5091	0.298	-1.705	0.088	-1.094	0.076
Book_Value_Unit_Curr	-6.0603	0.619	-9.798	0.000	-7.273	-4.848
CEPS_annualised_Unit_Curr	-0.6261	0.343	-1.826	0.068	-1.298	0.046
ROG_Net_Worth_perc	-0.4156	0.129	-3.215	0.001	-0.669	-0.162
Curr_Ratio_Latest	-0.9498	0.715	-1.329	0.184	-2.350	0.451
Interest_Cover_Ratio_Latest	-0.4397	0.294	-1.493	0.135	-1.017	0.137
Value_of_Output_to_Gross_Block	-0.5023	0.150	-3.358	0.001	-0.795	-0.209

Possibly complete quasi-separation: A fraction 0.30 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

Fig – 1.41 Model-1 Summary.

	variables	VIF
3	PBIDT	4.937125
0	Networth	4.186395
7	CEPS_annualised_Unit_Curr	3.863618
5	Book_Value_Unit_Curr	2.882603
6	Market_Capitalisation	2.620160
2	Selling_Cost	2.558143
1	Total_Debt	2.056451
4	Rev_exp_in_forex	2.027065
8	ROG_Net_Worth_perc	1.652846
13	Value_of_Output_to_Gross_Block	1.118399
12	Interest_Cover_Ratio_Latest	1.048832
11	Debtors_Ratio_Latest	1.013579
10	Inventory_Ratio_Latest	1.013266
9	Curr_Ratio_Latest	1.007419

Fig – 1.42 Model-1 Variation Inflation Matrix (VIF)

Curr_Ratio_Latest has the highest p-value and is insignificant, therefore, we need to eliminate it.

Model - 2

```
Optimization terminated successfully.
         Current function value: 0.144188
         Iterations 11
```

Fig – 1.43 Initializing Model-2

Logit Regression Results

Dep. Variable:	default	**No. Observations:**	2402
Model:	Logit	**Df Residuals:**	2393
Method:	MLE	**Df Model:**	8
Date:	Sun, 15 May 2022	**Pseudo R-squ.:**	0.5794
Time:	16:10:05	**Log-Likelihood:**	-346.34
converged:	True	**LL-Null:**	-823.47
		LLR p-value:	1.111e-200

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-7.3006	0.485	-15.061	0.000	-8.251	-6.351
Networth	-1.2588	0.364	-3.456	0.001	-1.973	-0.545
Capital_Employed	0.8416	0.281	2.995	0.003	0.291	1.392
PBIDT	-0.5027	0.300	-1.675	0.094	-1.091	0.086
Book_Value_Unit_Curr	-6.2010	0.620	-10.003	0.000	-7.416	-4.986
CEPS_annualised_Unit_Curr	-0.6334	0.346	-1.830	0.067	-1.312	0.045
ROG_Net_Worth_perc	-0.4371	0.130	-3.362	0.001	-0.692	-0.182
Interest_Cover_Ratio_Latest	-0.4537	0.290	-1.565	0.117	-1.022	0.114
Value_of_Output_to_Gross_Block	-0.4973	0.149	-3.349	0.001	-0.788	-0.206

Possibly complete quasi-separation: A fraction 0.30 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

Fig – 1.44 Model-2 Summary

Interest_Cover_Ratio_Latest has the highest p-value and is insignificant, therefore, we need to eliminate it.

Model - 3

```
Optimization terminated successfully.
         Current function value: 0.144347
         Iterations 11
```

Fig – 1.45 Initializing Model-3

Logit Regression Results

Dep. Variable:	default	**No. Observations:**	2402
Model:	Logit	**Df Residuals:**	2394
Method:	MLE	**Df Model:**	7
Date:	Sun, 15 May 2022	**Pseudo R-squ.:**	0.5789
Time:	16:10:05	**Log-Likelihood:**	-346.72
converged:	True	**LL-Null:**	-823.47
		LLR p-value:	1.343e-201

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-7.2848	0.484	-15.051	0.000	-8.233	-6.336
Networth	-1.2642	0.365	-3.464	0.001	-1.979	-0.549
Capital_Employed	0.8500	0.281	3.025	0.002	0.299	1.401
PBIDT	-0.5105	0.300	-1.703	0.089	-1.098	0.077
Book_Value_Unit_Curr	-6.2101	0.622	-9.985	0.000	-7.429	-4.991
CEPS_annualised_Unit_Curr	-0.6336	0.346	-1.833	0.067	-1.311	0.044
ROG_Net_Worth_perc	-0.4367	0.130	-3.364	0.001	-0.691	-0.182
Value_of_Output_to_Gross_Block	-0.4992	0.149	-3.361	0.001	-0.790	-0.208

Possibly complete quasi-separation: A fraction 0.30 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

Fig – 1.46 Model-3 Summary

PBIDT has the highest p-value and is insignificant, therefore, we need to eliminate it.

Model - 4

```
Optimization terminated successfully.
         Current function value: 0.145004
         Iterations 11
```

Fig – 1.47 Initializing Model-4

Logit Regression Results

Dep. Variable:	default	**No. Observations:**	2402
Model:	Logit	**Df Residuals:**	2395
Method:	MLE	**Df Model:**	6
Date:	Sun, 15 May 2022	**Pseudo R-squ.:**	0.5770
Time:	16:10:06	**Log-Likelihood:**	-348.30
converged:	True	**LL-Null:**	-823.47
		LLR p-value:	4.908e-202

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-7.2406	0.478	-15.153	0.000	-8.177	-6.304
Networth	-1.1449	0.344	-3.326	0.001	-1.820	-0.470
Capital_Employed	0.6531	0.251	2.604	0.009	0.162	1.145
Book_Value_Unit_Curr	-6.2606	0.625	-10.023	0.000	-7.485	-5.036
CEPS_annualised_Unit_Curr	-0.9019	0.304	-2.965	0.003	-1.498	-0.306
ROG_Net_Worth_perc	-0.4620	0.130	-3.563	0.000	-0.716	-0.208
Value_of_Output_to_Gross_Block	-0.4935	0.147	-3.358	0.001	-0.782	-0.205

Possibly complete quasi-separation: A fraction 0.30 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

Fig – 1.48 Model-4 Summary

All features are having p-value less than the alpha value. So model building for the Random Forest is over.

1.9 Validate the Random Forest Model on test Dataset and state the performance matrices. Also state interpretation from the model.

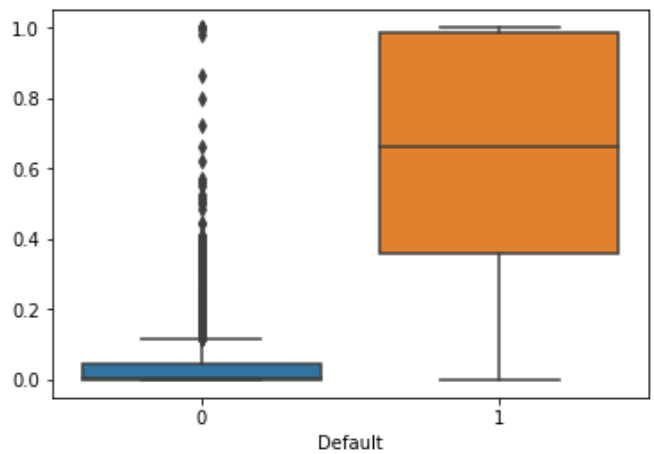


Fig – 1.49 Boxplot for Default variable.

```
Optimal threshold value of Random Forest model is 0.22423536888400453
```

Fig – 1.50 Optimum threshold value

```
842     0.080
1057    0.018
1595    0.189
100     0.888
1191    0.007
2163    0.000
2763    0.005
2701    0.000
2072    0.020
2349    0.001
1392    0.000
1621    0.032
1960    0.000
2148    0.000
571     0.000
1984    0.168
1592    0.006
3110    0.000
1564    0.000
2155    0.375
dtype: float64
```

Fig – 1.51 Predicted train values

```
251     7.840598e-01
3493    2.465653e-10
3063    5.228939e-07
2384    2.317520e-03
1679    2.060074e-02
dtype: float64
```

Fig – 1.52 Predicted test values

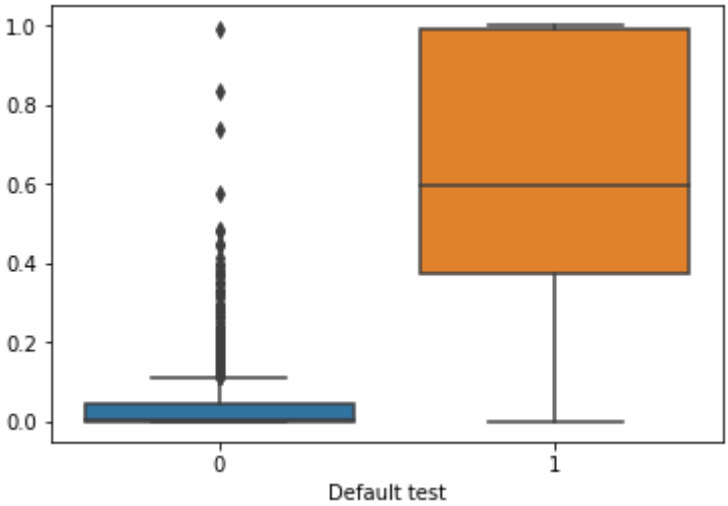
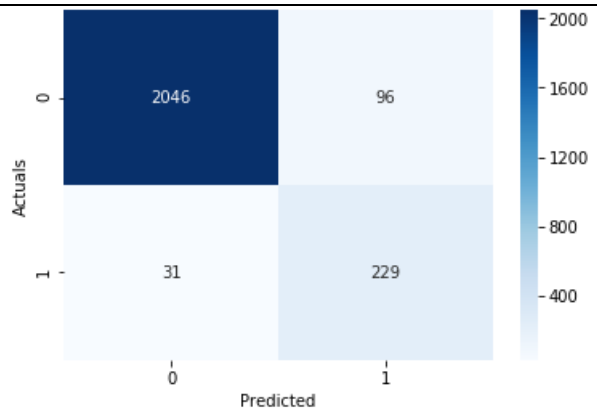
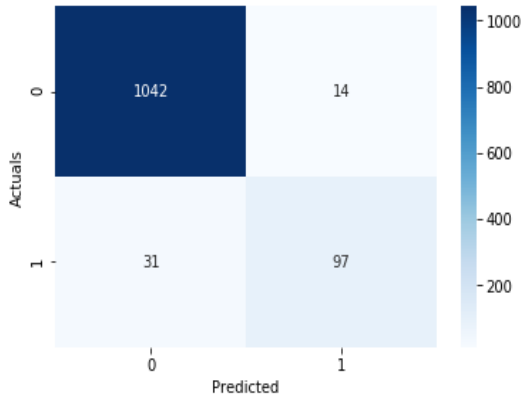
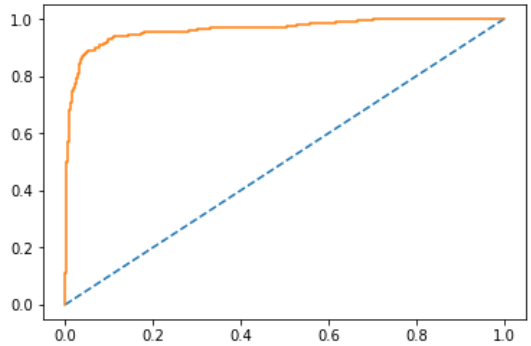
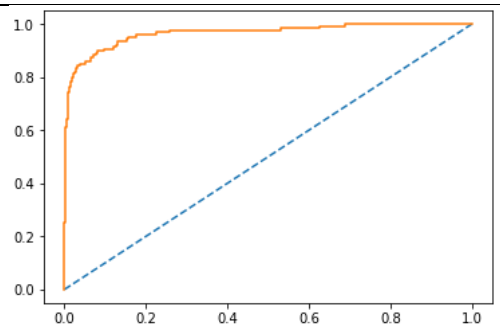


Fig – 1.53 Boxplot for test values

```
Optimal threshold value of Random Forest model test data is 0.36950413411928446
```

Fig – 1.54 Optimum threshold values for test data

Table	Train data	Test data
Confusion matrix	 Fig – 1.55 Confusion matrix train values	 Fig – 1.56 Confusion matrix test values
Classification report	precision recall f1-score support 0 0.985 0.955 0.970 2142 1 0.705 0.881 0.783 260 accuracy 0.947 2402 macro avg 0.845 0.918 0.876 2402 weighted avg 0.955 0.947 0.950 2402 Fig – 1.57 Classification report for train data	precision recall f1-score support 0 0.971 0.987 0.979 1056 1 0.874 0.758 0.812 128 accuracy 0.962 1184 macro avg 0.922 0.872 0.895 1184 weighted avg 0.961 0.962 0.961 1184 Fig – 1.58 Classification report for test data
AUC ROC Curve	 Fig – 1.59 ROC for train data AUC: for train data 0.963 Fig 1.60 AUC score for train data.	 Fig – 1.61 ROC for train data AUC for test data: 0.966 Fig 1.62 AUC score for test data.

1.10 Build a LDA Model on Train Dataset. Also showcase your model building approach.

```
LinearDiscriminantAnalysis(solver='eigen')
```

Fig 1.63 Initializing LDA model

	Feature	Rank
0	Networth	1
1	Capital_Employed	1
2	Total_Debt	1
4	PBIDT	1
6	Book_Value_Unit_Curr	1
7	Market_Capitalisation	1
8	CEPS_annualised_Unit_Curr	1
9	ROG_Net_Worth_perc	1
14	Value_of_Output_to_Gross_Block	1

Fig 1.64 Taking features with rank 1 for LDA model

```
0    1056
1     128
Name: default, dtype: int64
```

Fig 1.65 Value count for default variable

Model – 1

```
Optimization terminated successfully.
         Current function value: 0.139387
         Iterations 11
```

Fig 1.66 Model 1 Initializing

Logit Regression Results

Dep. Variable:	default	**No. Observations:**	2402
Model:	Logit	**Df Residuals:**	2392
Method:	MLE	**Df Model:**	9
Date:	Sun, 15 May 2022	**Pseudo R-squ.:**	0.5934
Time:	16:37:57	**Log-Likelihood:**	-334.81
converged:	True	**LL-Null:**	-823.47
		LLR p-value:	1.336e-204

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-7.3485	0.482	-15.245	0.000	-8.293	-6.404
Networth	-0.5328	0.408	-1.305	0.192	-1.333	0.268
Capital_Employed	-0.7052	0.548	-1.287	0.198	-1.779	0.369
Total_Debt	1.4287	0.357	3.997	0.000	0.728	2.129
PBIDT	-0.4808	0.321	-1.498	0.134	-1.110	0.148
Book_Value_Unit_Curr	-6.2383	0.645	-9.667	0.000	-7.503	-4.974
Market_Capitalisation	-0.6361	0.289	-2.197	0.028	-1.203	-0.069
CEPS_annualised_Unit_Curr	-0.4805	0.352	-1.366	0.172	-1.170	0.209
ROG_Net_Worth_perc	-0.4206	0.131	-3.221	0.001	-0.677	-0.165
Value_of_Output_to_Gross_Block	-0.4910	0.157	-3.130	0.002	-0.798	-0.184

Possibly complete quasi-separation: A fraction 0.31 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

Fig 1.67 Model -1 summary

	variables	VIF
1	Capital_Employed	10.348232
0	Networth	7.759803
3	PBIDT	4.738334
6	CEPS_annualised_Unit_Curr	3.815682
2	Total_Debt	3.464989
4	Book_Value_Unit_Curr	2.874226
5	Market_Capitalisation	2.531367
7	ROG_Net_Worth_perc	1.663652
8	Value_of_Output_to_Gross_Block	1.113905

Fig 1.68 Model -1 VIF

Capital_Employed has the highest VIF value and is insignificant, therefore, we need to eliminate it.

Model – 2

```
Optimization terminated successfully.
         Current function value: 0.139736
         Iterations 11
```

Fig 1.69 Model – 2 Intializing

Logit Regression Results

Dep. Variable:	default	**No. Observations:**	2402
Model:	Logit	**Df Residuals:**	2393
Method:	MLE	**Df Model:**	8
Date:	Sun, 15 May 2022	**Pseudo R-squ.:**	0.5924
Time:	16:37:57	**Log-Likelihood:**	-335.65
converged:	True	**LL-Null:**	-823.47
		LLR p-value:	2.697e-205

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-7.3574	0.485	-15.160	0.000	-8.309	-6.406
Networth	-0.7995	0.347	-2.302	0.021	-1.480	-0.119
Total_Debt	1.0527	0.201	5.241	0.000	0.659	1.446
PBIDT	-0.5350	0.322	-1.659	0.097	-1.167	0.097
Book_Value_Unit_Curr	-6.2904	0.643	-9.783	0.000	-7.551	-5.030
Market_Capitalisation	-0.7371	0.285	-2.589	0.010	-1.295	-0.179
CEPS_annualised_Unit_Curr	-0.4388	0.348	-1.261	0.207	-1.121	0.243
ROG_Net_Worth_perc	-0.4119	0.130	-3.179	0.001	-0.666	-0.158
Value_of_Output_to_Gross_Block	-0.4893	0.157	-3.124	0.002	-0.796	-0.182

Possibly complete quasi-separation: A fraction 0.31 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

Fig 1.70 Model – 2 Summary

CEPS_annualised_Unit_Curr has the highest p-value and is insignificant, therefore, we need to eliminate it.

Model – 3

```
Optimization terminated successfully.
         Current function value: 0.140067
         Iterations 11
```

Fig 1.71 Intializing Model -3

Logit Regression Results

Dep. Variable:	default	**No. Observations:**	2402
Model:	Logit	**Df Residuals:**	2394
Method:	MLE	**Df Model:**	7
Date:	Sun, 15 May 2022	**Pseudo R-squ.:**	0.5914
Time:	16:37:57	**Log-Likelihood:**	-336.44
converged:	True	**LL-Null:**	-823.47
		LLR p-value:	4.858e-206

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-7.2051	0.466	-15.468	0.000	-8.118	-6.292
Networth	-0.7906	0.351	-2.250	0.024	-1.479	-0.102
Total_Debt	1.1148	0.198	5.620	0.000	0.726	1.504
PBIDT	-0.7480	0.285	-2.625	0.009	-1.306	-0.190
Book_Value_Unit_Curr	-6.2362	0.639	-9.761	0.000	-7.488	-4.984
Market_Capitalisation	-0.7612	0.284	-2.680	0.007	-1.318	-0.205
ROG_Net_Worth_perc	-0.4728	0.121	-3.894	0.000	-0.711	-0.235
Value_of_Output_to_Gross_Block	-0.4841	0.157	-3.080	0.002	-0.792	-0.176

Possibly complete quasi-separation: A fraction 0.30 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

Fig 1.72 Model -3 Summary

All the features are having p-value less than the alpha value. So model building for the LDA model is completed.

1.11 Validate the LDA Model on test Dataset and state the performance matrices. Also state interpretation from the model.

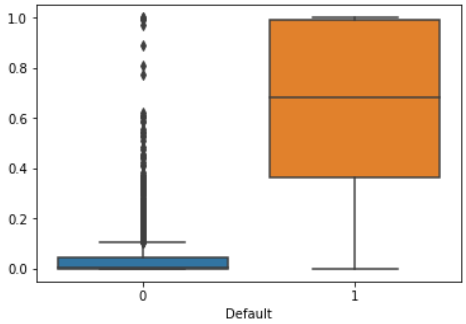


Fig 1.73 Default value for LDA train model

```
Optimum threshold value for LDA train model 0.16890979736726344
```

Fig 1.74 Optimum threshold value for LDA train model

```
842     0.080
1057    0.019
1595    0.035
100     0.861
1191    0.010
2163    0.000
2763    0.002
2701    0.000
2072    0.020
2349    0.001
1392    0.000
1621    0.039
1960    0.000
2148    0.000
571     0.000
1984    0.026
1592    0.008
3110    0.000
1564    0.000
2155    0.264
dtype: float64
```

Fig 1.75 Predicted value for LDA train model

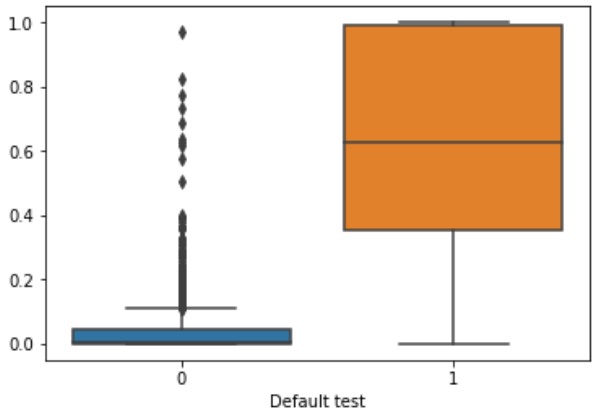
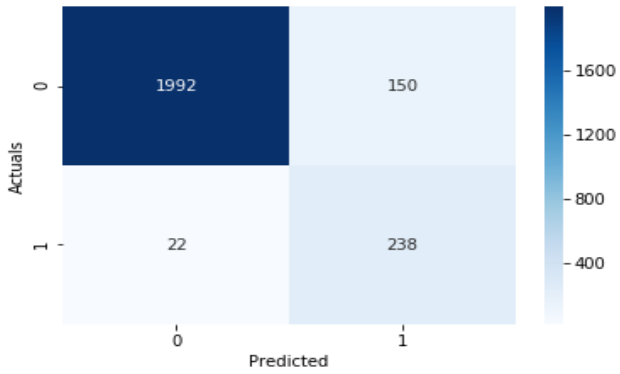
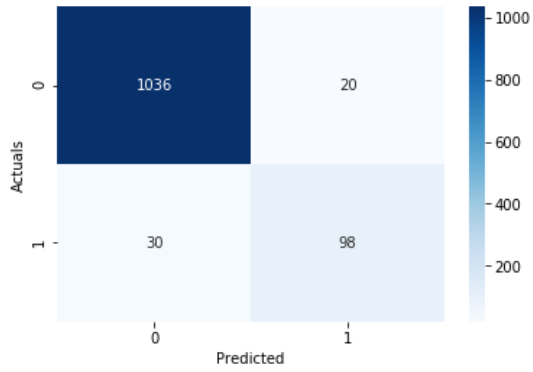
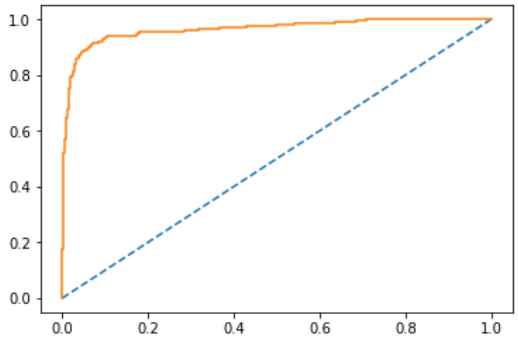
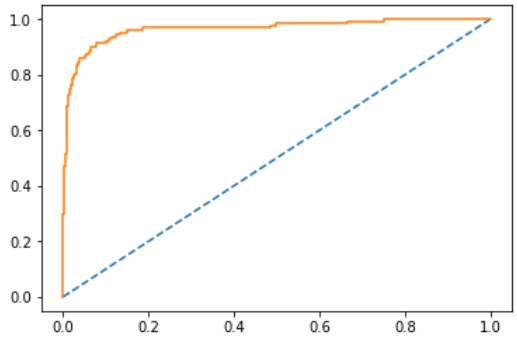


Fig 1.76 Default value for LDA test model

```
251     0.706
3493    0.000
3063    0.000
2384    0.001
1679    0.016
604     0.003
3434    0.000
2244    0.000
2523    0.000
2162    0.000
3102    0.000
1638    0.101
2046    0.000
1241    0.143
133     0.255
2294    0.001
2139    0.000
2844    0.011
1360    0.050
2896    0.000
dtype: float64
```

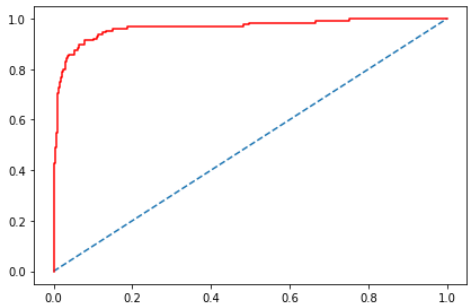
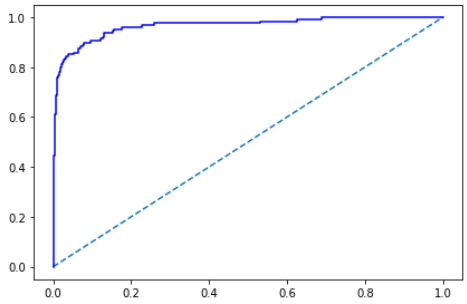
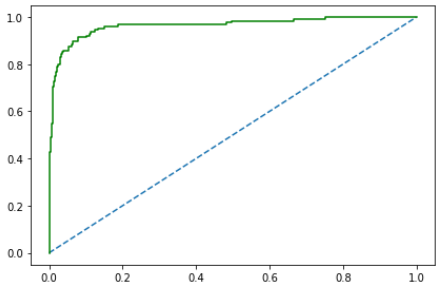
Fig 1.77 Predicted value for LDA test model

Table	Train data	Test data
Confusion matrix	 Fig – 1.78 Confusion matrix train values	 Fig – 1.79 Confusion matrix test values
Classification report	precision recall f1-score support 0 0.989 0.930 0.959 2142 1 0.613 0.915 0.735 260 accuracy 0.928 2402 macro avg 0.801 0.923 0.847 2402 weighted avg 0.948 0.928 0.934 2402 Fig – 1.80 Classification report for train data	precision recall f1-score support 0 0.972 0.981 0.976 1056 1 0.831 0.766 0.797 128 accuracy 0.958 1184 macro avg 0.901 0.873 0.887 1184 weighted avg 0.957 0.958 0.957 1184 Fig – 1.81 Classification report for test data
AUC ROC Curve	 Fig – 1.82 ROC for train data AUC for train data: 0.964 Fig 1.83 AUC score for train data.	 Fig – 1.84 ROC for train data AUC for test data: 0.965 Fig 1.85 AUC score for test data.

1.12 Compare the performances of Logistics, Radom Forest and LDA models (include ROC Curve).

	LR Test	RF Test	LDA Test
Accuracy	0.930	0.962	0.958
AUC	0.916	0.966	0.965
Recall	0.898	0.758	0.766
Precision	0.622	0.874	0.831
F1 Score	0.735	0.812	0.797

Fig 1.86 Comparison dataframe for LR,RF and LDA values.

LR ROC Curve	***RF ROC Curve***	***LDA ROC Curve***
 Fig 1.87 ROC curve for LR model	 Fig 1.88 ROC curve for RF model	 Fig 1.89 ROC curve for LDA model

From the above dataframe, Recall is higher in logistic Regression, precision is better in Random Forest classifier.
Overall Random Forest is better algorithm.

1.13 State Recommendations from the above models

From the above model, Random Forest model is the best model with higher precision and recall.

Company with the following details will lead the investor to invest in the company are

1. Increase in debtor's shows company turnover has increased.

2. Increase in debtor's / decrease in creditors will lead to get fresh loan with lower interest rate with good credit rating for the company.

3. Reduction in creditors shows that the company follows the strict/disciplined payment terms.

4. Change in debt equity ratios shows that the company is growing.

5. Growth in current asset and decrease in current liability.

6. Increase in Net worth will help the company to provide good dividends for the share/stake holders.