FRA Project (Milestone-2)

Name: EDIL MONICA S
PGP-DSBA Online
Date: 12/12/2021

Table of contents

Business Report

Objective:

- Build Random Forest model and LDA model and Compare the performance metrics with Logistic Regression, Random forest and LDA. Plot ROC curve and compare AUC metrics as well. And finally recommend a model to the investor.
- Plot Stock price graph for 2 stocks and find returns of all stocks.
- Calculate and plot stock mean and standard deviation of all stocks and their interpretation. Finally recommendations.

Description: Businesses or companies can fall prey to default if they are not able to keep up their debt obligations. Defaults will lead to a lower credit rating for the company which in turn reduces its chances of getting credit in the future and may have to pay higher interests on existing debts as well as any new obligations. From an investor's point of view, he would want to invest in a company if it is capable of handling its financial obligations, can grow quickly, and is able to manage the growth scale.

A balance sheet is a financial statement of a company that provides a snapshot of what a company owns, owes, and the amount invested by the shareholders. Thus, it is an important tool that helps evaluate the performance of a business.

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

Data Dictionary:

New Field Name	Description
Co_Code	Company Code
Co_Name	Company Name
Networth_Next_Year	Value of a company as on 2016 - Next Year(difference between the value of total assets and total liabilities)
Equity_Paid_Up	Amount that has been received by the company through the issue of shares to the shareholders
Networth	Value of a company as on 2015 - Current Year
Capital_Employed	Total amount of capital used for the acquisition of profits by a company
Total_Debt	The sum of money borrowed by the company and is due to be paid
Gross_Block	Total value of all of the assets that a company owns
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).

Curr_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_Liab_and_Prov	Short-term financial obligations that are due within one year (includes amount that is set aside cover a future liability)
Total_Assets_to_Liab	Ratio of total assets to liabilities of the company
Gross_Sales	The grand total of sale transactions within the accounting period
Net_Sales	Gross sales minus returns, allowances, and discounts
Other_Income	Income realized from non-business activities (e.g. sale of long term asset)
Value_Of_Output	Product of physical output of goods and services produced by company and its market price
Cost_of_Prod	Costs incurred by a business from manufacturing a product or providing a service
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)
PBIDT	Profit Before Interest, Depreciation & Taxes
PBDT	Profit Before Depreciation and Tax
PBIT	Profit before interest and taxes
PBT	Profit before tax
PAT	Profit After Tax
Adjusted_PAT	Adjusted profit is the best estimate of the true profit
CP	Commercial paper, a short-term debt instrument to meet short-term liabilities.
Rev_earn_in_forex	Revenue earned in foreign currency
Rev_exp_in_forex	Expenses due to foreign currency transactions
Capital_exp_in_forex	Long term investment in forex
Book_Value_Unit_Curr	Net asset value

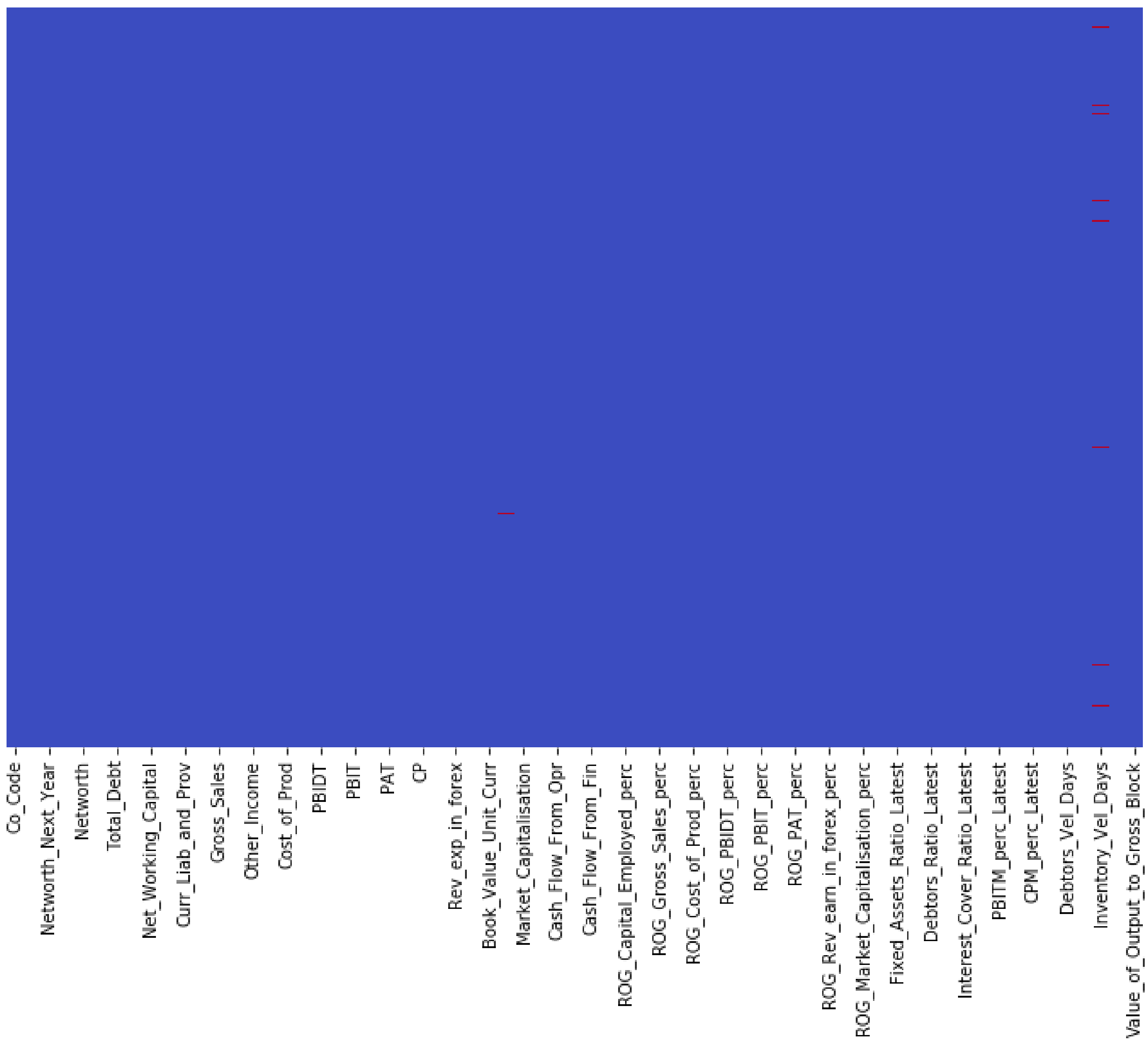
Book_Value_Adj_Unit_Curr	Book value adjusted to reflect asset's true fair market value
Market_Capitalisation	Product of the total number of a company's outstanding shares and the current market price of one share
CEPS_annualised_Unit_Curr	Cash Earnings per Share, profitability ratio that measures the financial performance of a company by calculating cash flows on a per share basis
Cash_Flow_From_Opr	Use of cash from ongoing regular business activities
Cash_Flow_From_Inv	Cash used in the purchase of non-current assets–or long-term assets– that will deliver value in the future
Cash_Flow_From_Fin	Net flows of cash that are used to fund the company (transactions involving debt, equity, and dividends)
ROG_Net_Worth_perc	Rate of Growth - Net worth
ROG_Capital_Employed_perc	Rate of Growth - Capital Employed
ROG_Gross_Block_perc	Rate of Growth - Gross Block
ROG_Gross_Sales_perc	Rate of Growth - Gross Sales
ROG_Net_Sales_perc	Rate of Growth - Net Sales
ROG_Cost_of_Prod_perc	Rate of Growth - Cost of Production
ROG_Total_Assets_perc	Rate of Growth - Total Assets
ROG_PBIDT_perc	Rate of Growth- PBIDT
ROG_PBDT_perc	Rate of Growth- PBDT
ROG_PBIT_perc	Rate of Growth- PBIT
ROG_PBT_perc	Rate of Growth- PBT
ROG_PAT_perc	Rate of Growth- PAT
ROG_CP_perc	Rate of Growth- CP
ROG_Rev_earn_in_forex_perc	Rate of Growth - Revenue earnings in forex

ROG_Rev_exp_in_forex_perc	Rate of Growth - Revenue expenses in forex
ROG_Market_Capitalisation_perc	Rate of Growth - Market Capitalization
Curr_Ratio_Latest	Liquidity ratio, company's ability to pay short-term obligations or those due within one year
Fixed_Assets_Ratio_Latest	Solvency ratio, the capacity of a company to discharge its obligations towards long-term lenders indicating
Inventory_Ratio_Latest	Activity ratio, specifies the number of times the stock or inventory has been replaced and sold by the company
Debtors_Ratio_Latest	Measures how quickly cash debtors are paying back to the company
Total_Asset_Turnover_Ratio_Latest	The value of a company's revenues relative to the value of its assets
Interest_Cover_Ratio_Latest	Determines how easily a company can pay interest on its outstanding debt
PBIDTM_perc_Latest	Profit before Interest Depreciation and Tax Margin
PBITM_perc_Latest	Profit Before Interest Tax Margin
PBDTM_perc_Latest	Profit Before Depreciation Tax Margin
CPM_perc_Latest	Cost per thousand (advertising cost)
APATM_perc_Latest	After tax profit margin
Debtors_Vel_Days	Average days required for receiving the payments
Creditors_Vel_Days	Average number of days company takes to pay suppliers
Inventory_Vel_Days	Average number of days the company needs to turn its inventory into sales
Value_of_Output_to_Total_Assets	Ratio of Value of Output (market value) to Total Assets
Value_of_Output_to_Gross_Block	Ratio of Value of Output (market value) to Gross Block

EDA

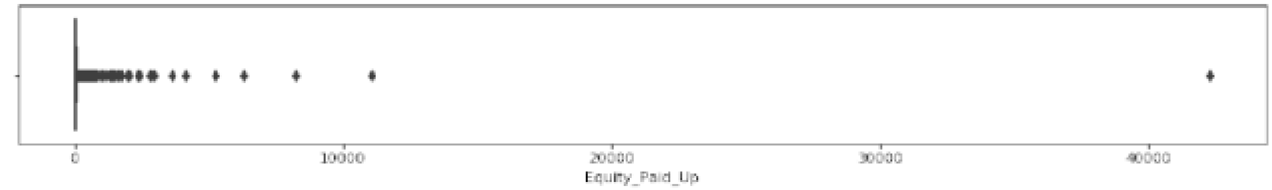
The Financial statement of the companies were imported and below are the observations.

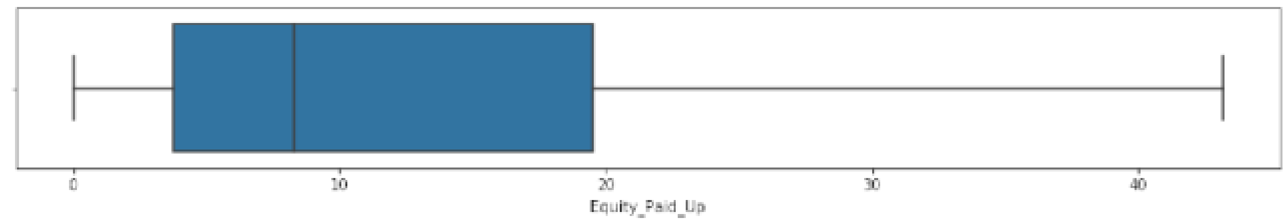
- There are 3586 rows and 67 columns
- There are no duplicated rows
- There are few missing values which were negligible. (118 values of 240144 which is 0.0004 percent of data.)
  Visual Presentation of missing values: Figure1

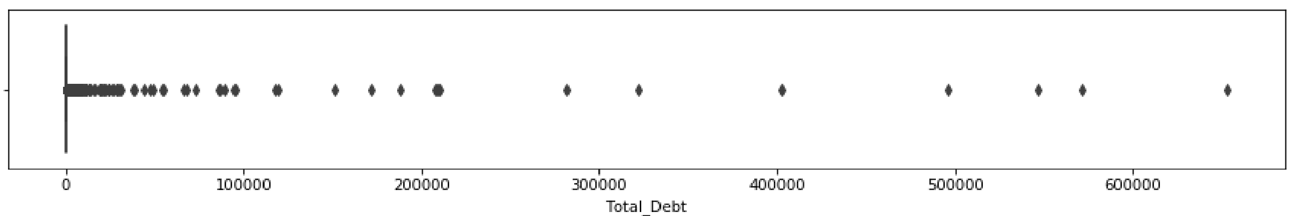


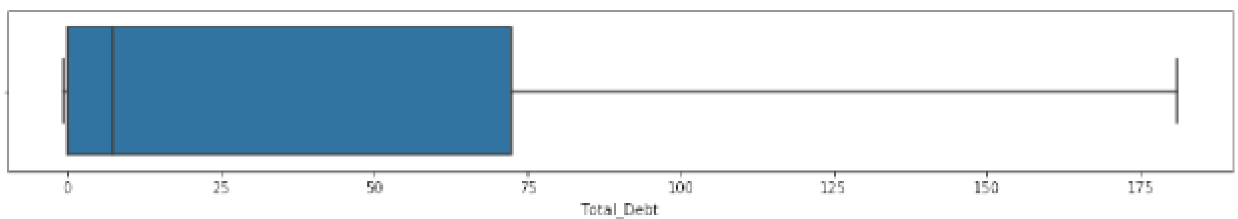
- The missing values was imputed by median.
- Total of 118 values were imputed by median.

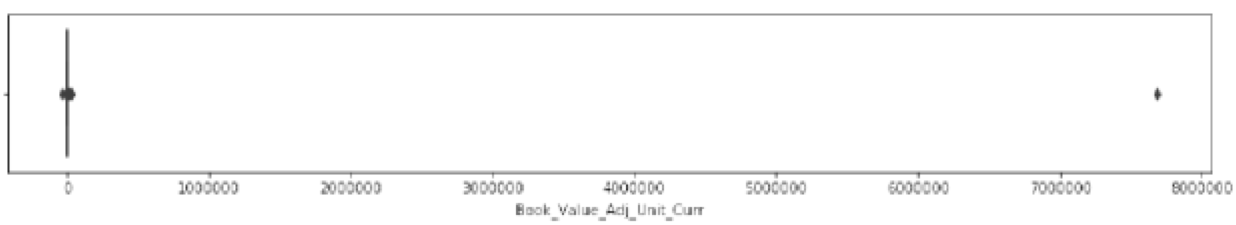
Outlier Treatment: Lower and upper capping done.

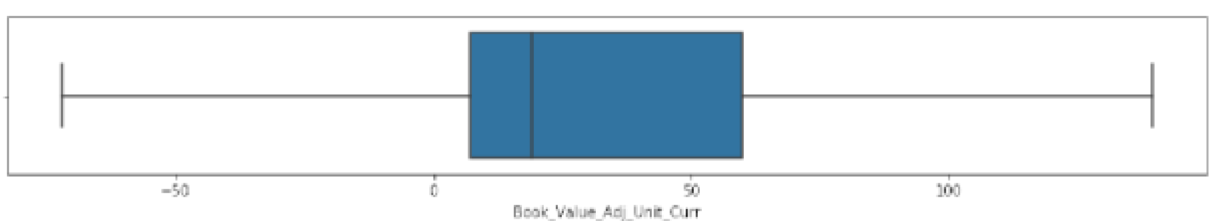


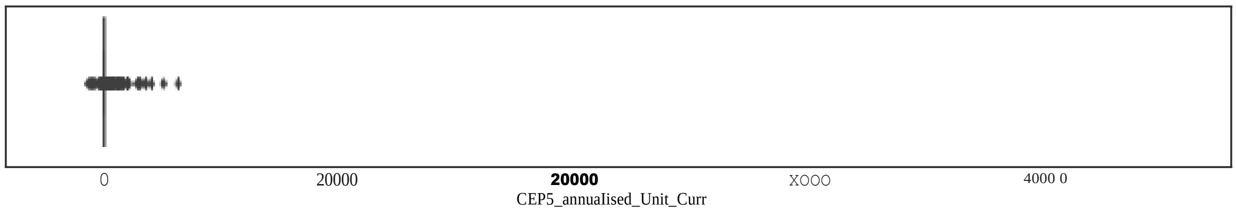


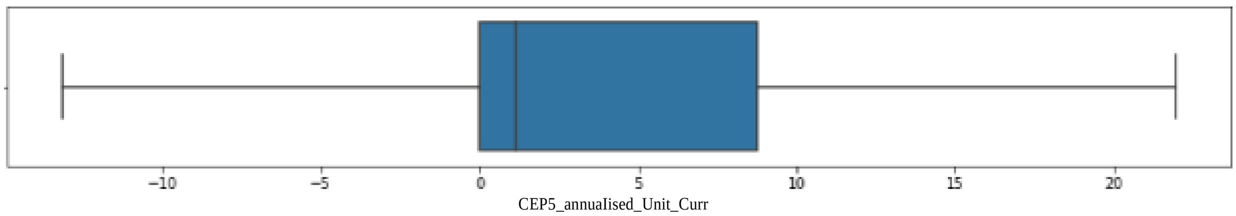


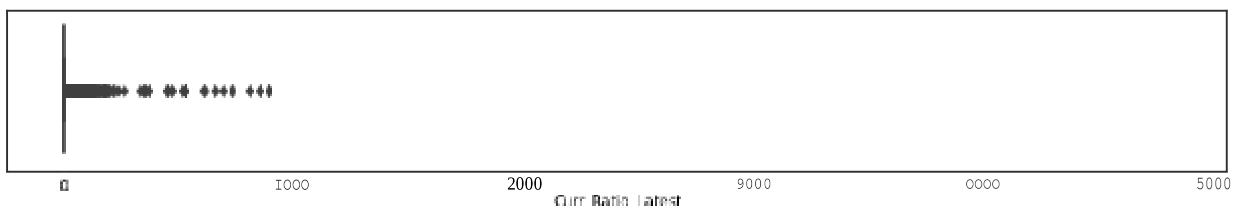


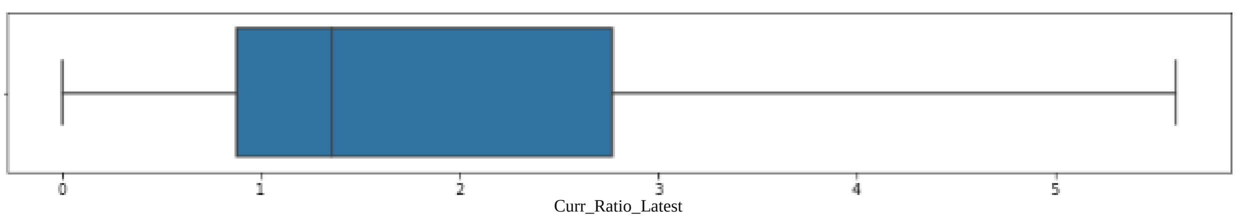


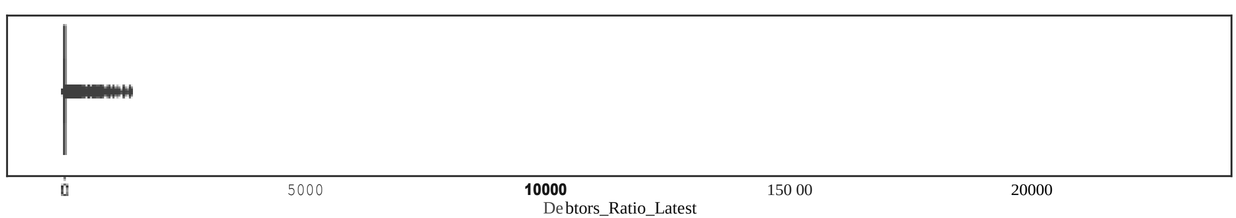


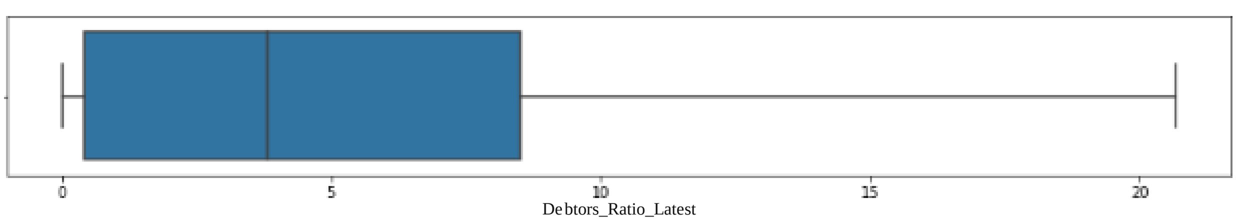

0
20000
20000
XOOO
4000 0
CEP5_annualised_Unit_Curr

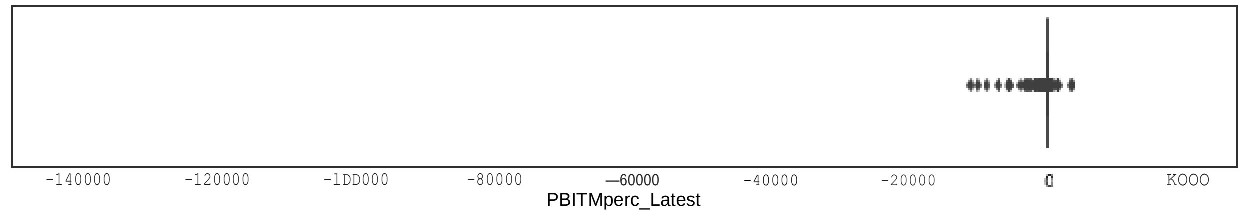

−10
−5
0
5
10
15
20
CEP5_annualised_Unit_Curr

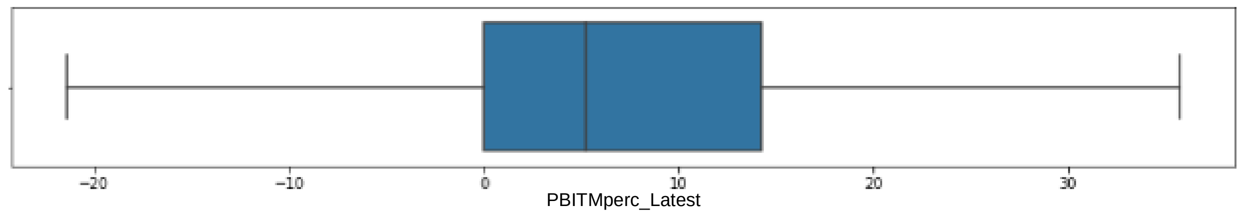

0
1000
2000
9000
0000
5000
Curr Ratio Latest

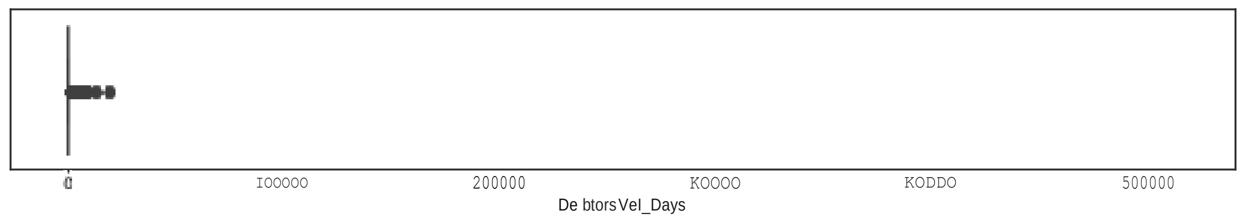

0
1
2
3
4
5
Curr_Ratio_Latest

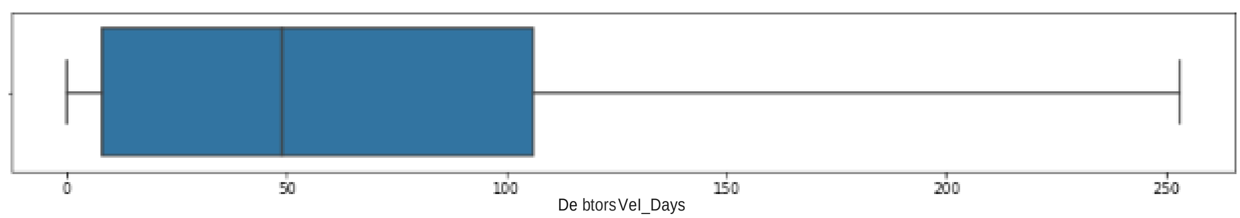

0
5000
10000
150 00
20000
De btors_Ratio_Latest

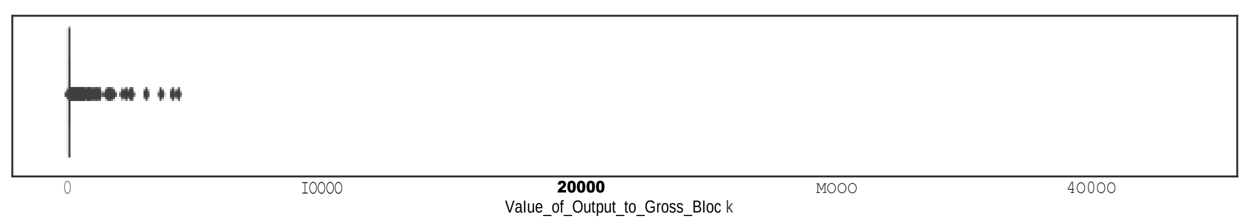

0
5
10
15
20
De btors_Ratio_Latest

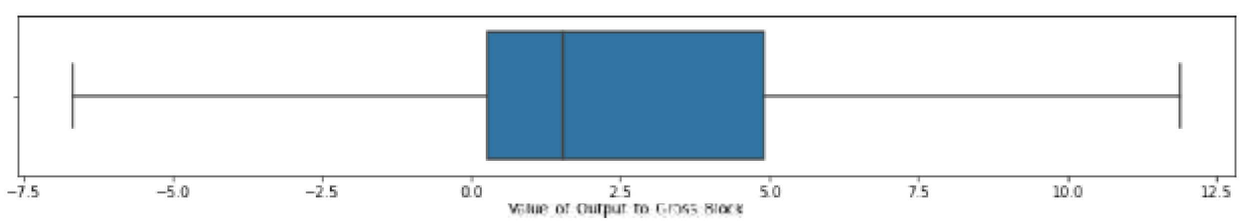

-140000
-120000
-1DD000
-80000
—60000
-40000
-20000
0
KOOO
PBITMperc_Latest


−20
−10
0
10
20
30
PBITMperc_Latest


0
IOOOOO
200000
KOOOO
KODDO
500000
De btorsVel_Days


0
50
100
150
200
250
De btorsVel_Days


0
IOOOO
20000
MOOO
4OOOO
Value_of_Output_to_Gross_Bloc k

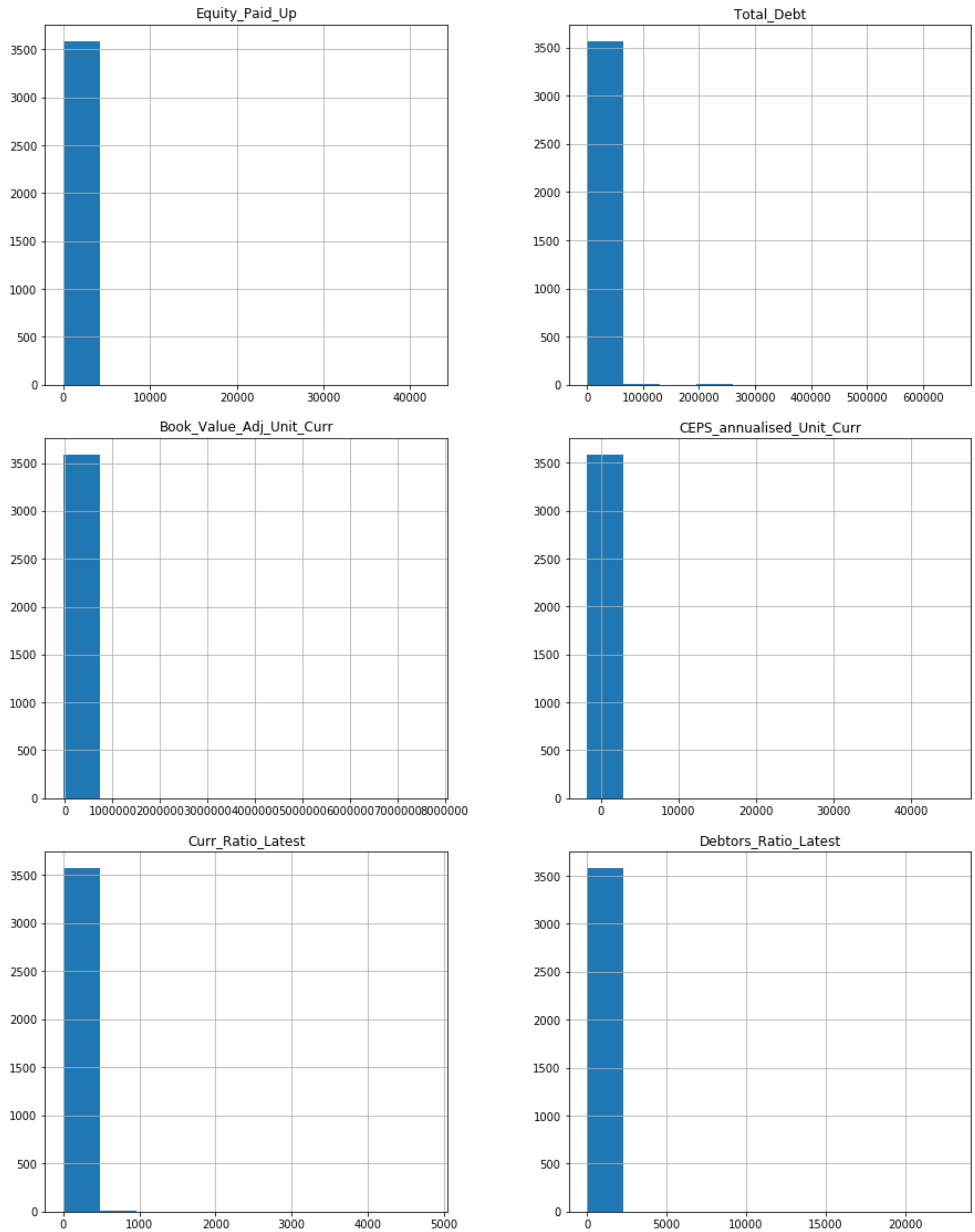

−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
12.5

Univariate Analysis

- List of Significant Variables used for model building are below :

New Field Name	Description
Equity_Paid_Up	Amount that has been received by the company through the issue of shares to the shareholders
Total_Debt	The sum of money borrowed by the company and is due to be paid
Book_Value_Adj_Unit_Curr	Book value adjusted to reflect asset's true fair market value
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
Curr_Ratio_Latest	Liquidity ratio, company's ability to pay short-term obligations or those due within one year
Debtors_Ratio_Latest	Measures how quickly cash debtors are paying back to the company
PBITM_perc_Latest	Profit Before Interest Tax Margin
Debtors_Vel_Days	Average days required for receiving the payments
Value_of_Output_to_Gross_Block	Ratio of Value of Output (market value) to Gross Block

	count	mean	std	min	25%	50%	75%	max
Equity_Paid_Up	3586	62.966584	778.761744	0	3.75	8.29	19.5175	42263.46
Total_Debt	3586	1994.823779	23652.84275	-0.72	0.03	7.49	72.35	652823.81
Book_Value_Adj_Unit_Curr	3582	2243.153	128283.7	-33715.7	7.06	18.925	60.01	7677600
CEPS_annualised_Unit_Curr	3586	36.018709	828.420796	-1808	0	1.145	8.7725	45438.44
Curr_Ratio_Latest	3585	12.056603	108.410131	0	0.88	1.36	2.77	4813
Debtors_Ratio_Latest	3585	33.026996	489.563498	0	0.42	3.82	8.52	22992.67
PBITM_perc_Latest	3585	-109.213414	3057.63587	-141600	0	5.23	14.29	19195.7
Debtors_Vel_Days	3586	603.894032	10636.75958	0	8	49	106	514721
Value_of_Output_to_Gross_Block	3586	61.884548	976.824352	-61	0.27	1.53	4.91	43404


Equity_Paid_Up
0
500
1000
1500
2000
2500
3000
3500
0
10000
20000
30000
40000
Total_Debt
0
500
1000
1500
2000
2500
3000
3500
0
100000
200000
300000
400000
500000
600000
Book_Value_Adj_Unit_Curr
0
500
1000
1500
2000
2500
3000
3500
0
1000000
2000000
3000000
4000000
5000000
6000000
7000000
8000000
CEPS_annualised_Unit_Curr
0
500
1000
1500
2000
2500
3000
3500
0
10000
20000
30000
40000
Curr_Ratio_Latest
0
500
1000
1500
2000
2500
3000
3500
0
1000
2000
3000
4000
5000
Debtors_Ratio_Latest
0
500
1000
1500
2000
2500
3000
3500
0
5000
10000
15000
20000

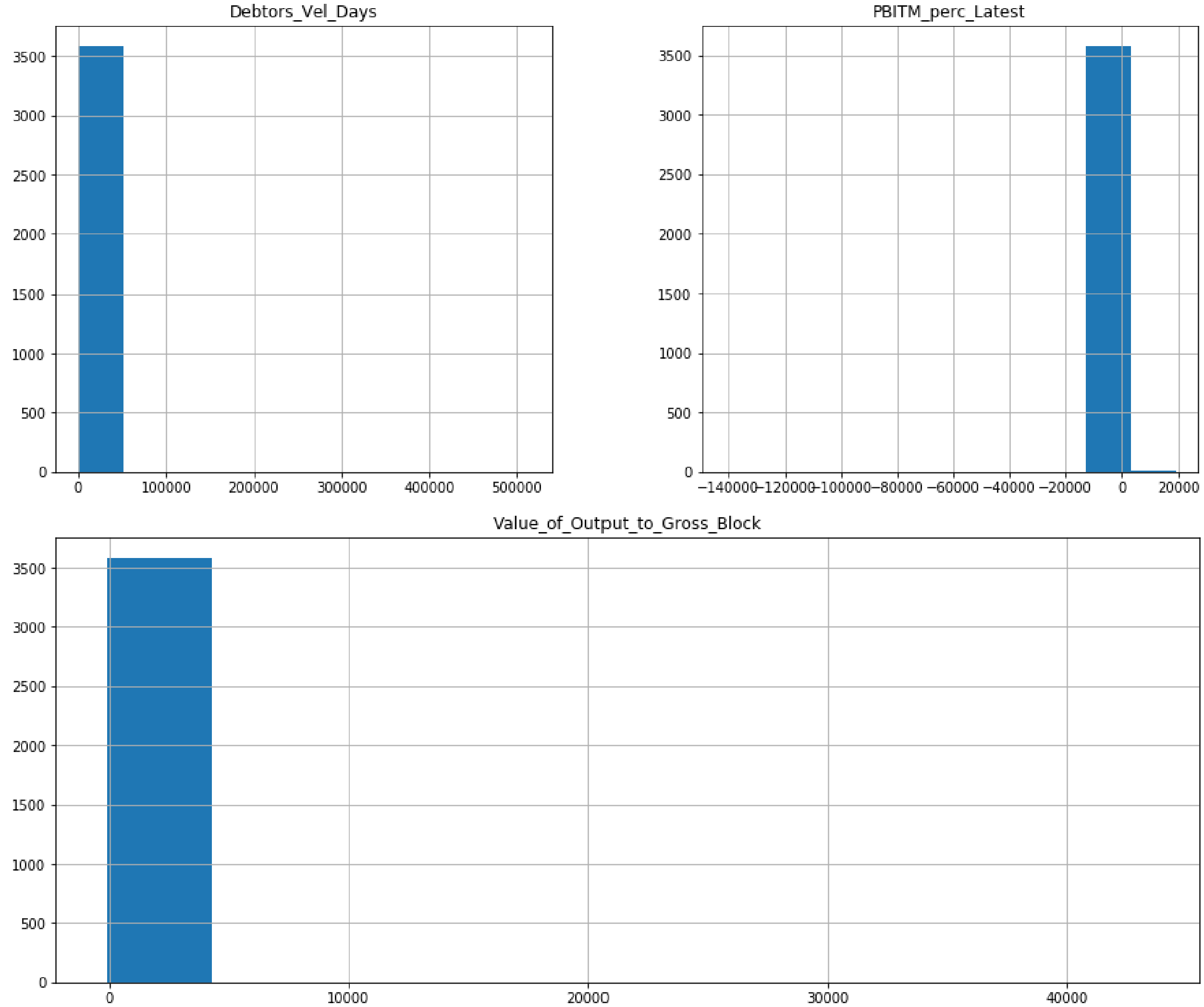
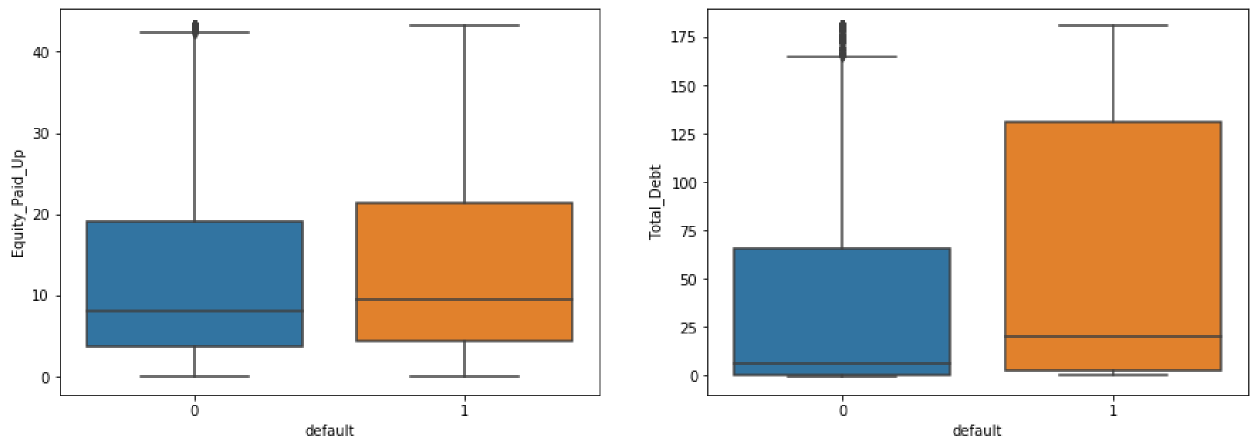


Figure2 Histogram of all Independent Variables to highlight the univariate data distributions.

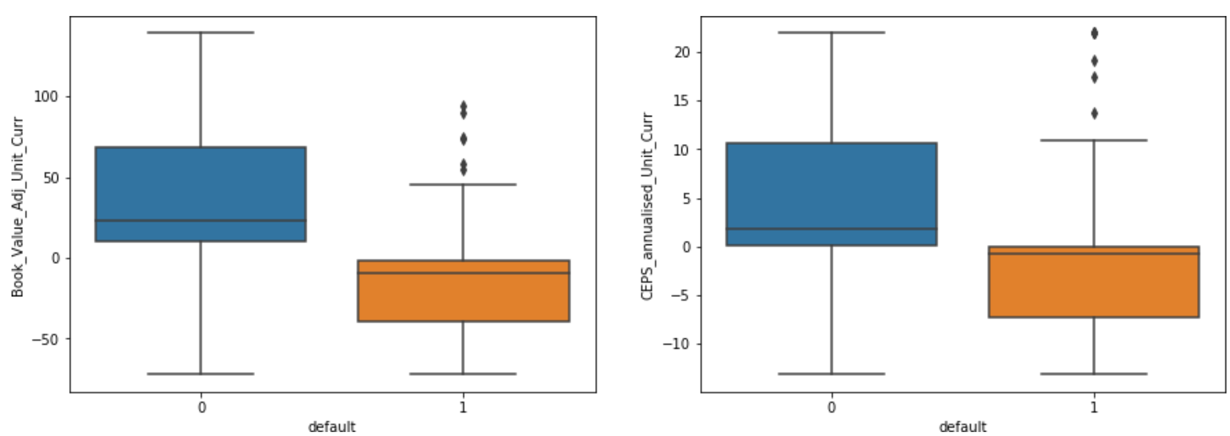
- All Significant Variables taken for building model are skewed.
- The **Equity Paid Up** variable's distribution is left skewed indicating that very few companies have maximum equity borrowing of 42,263. The mean Equity borrowing is 62.9. The 75^{th} percentile is around 19.5 indicating majority lies between 3.7 to 8.2.
- The **Total Debt** variable's distribution is left skewed indicating that very few companies have maximum borrowing of 65, 2823. The mean borrowing is 1994.
- Debtors Vel Days - Few companies are yet to receive payments for more than 2 years.

Bivariate Analysis with Target Variable (Defaulters – Negative Net worth Next year)

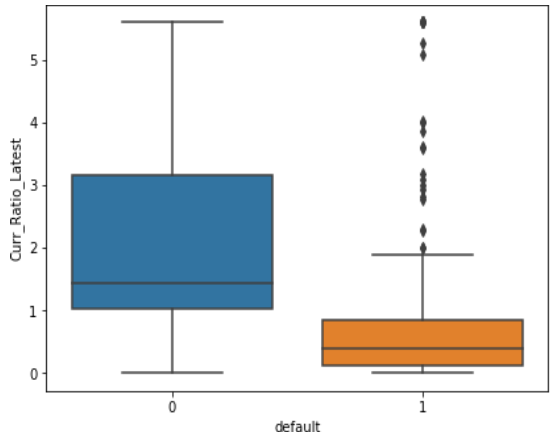
- Defaulter's Median of Equity borrowings are higher compared to non-defaulters.
- Defaulter's Median of debt is higher compared to non-defaulters.

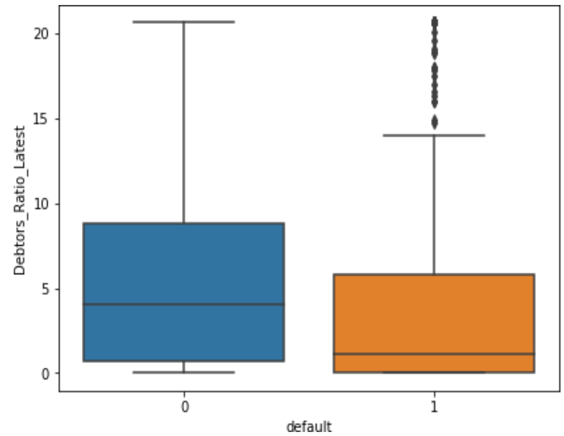


- Book value of a defaulted company valued currently at a fair price is far mor lesser than a non-defaulted company.
- Median of Cash earnings per share for a defaulting company is negative.

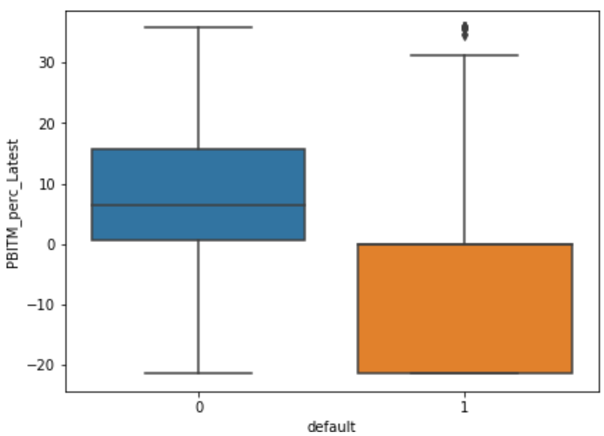


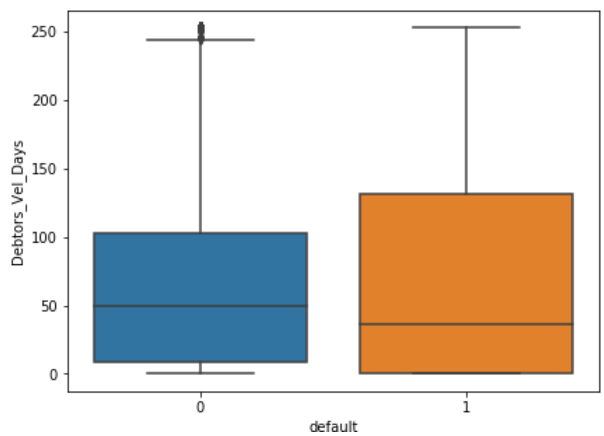
- For defaulting companies the liquidity ratio is lesser than the non-defaulters.
- The debtors to the company in case of defaulting companies, they pay lately as per the median below the non-defaulting company.



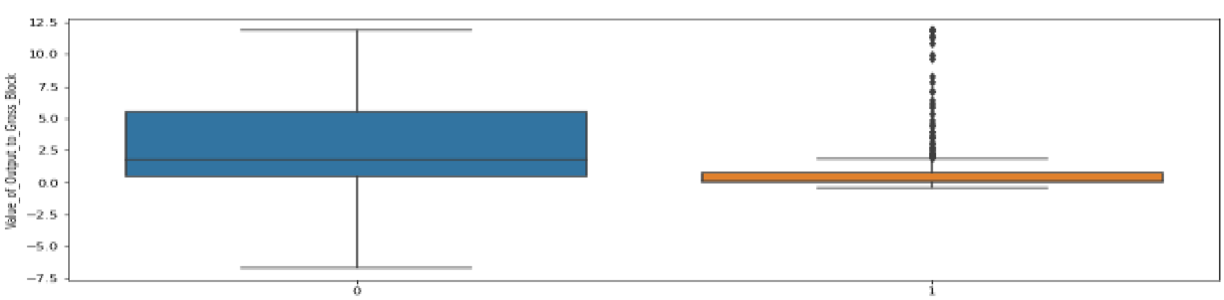


- The defaulting companies are not making profits – most of them.
- The days are more to receive payments in case of defaulting company compared to non-defaulters.





- Most of the defaulting company's market value ratio to gross median is less than the non-defaulters.



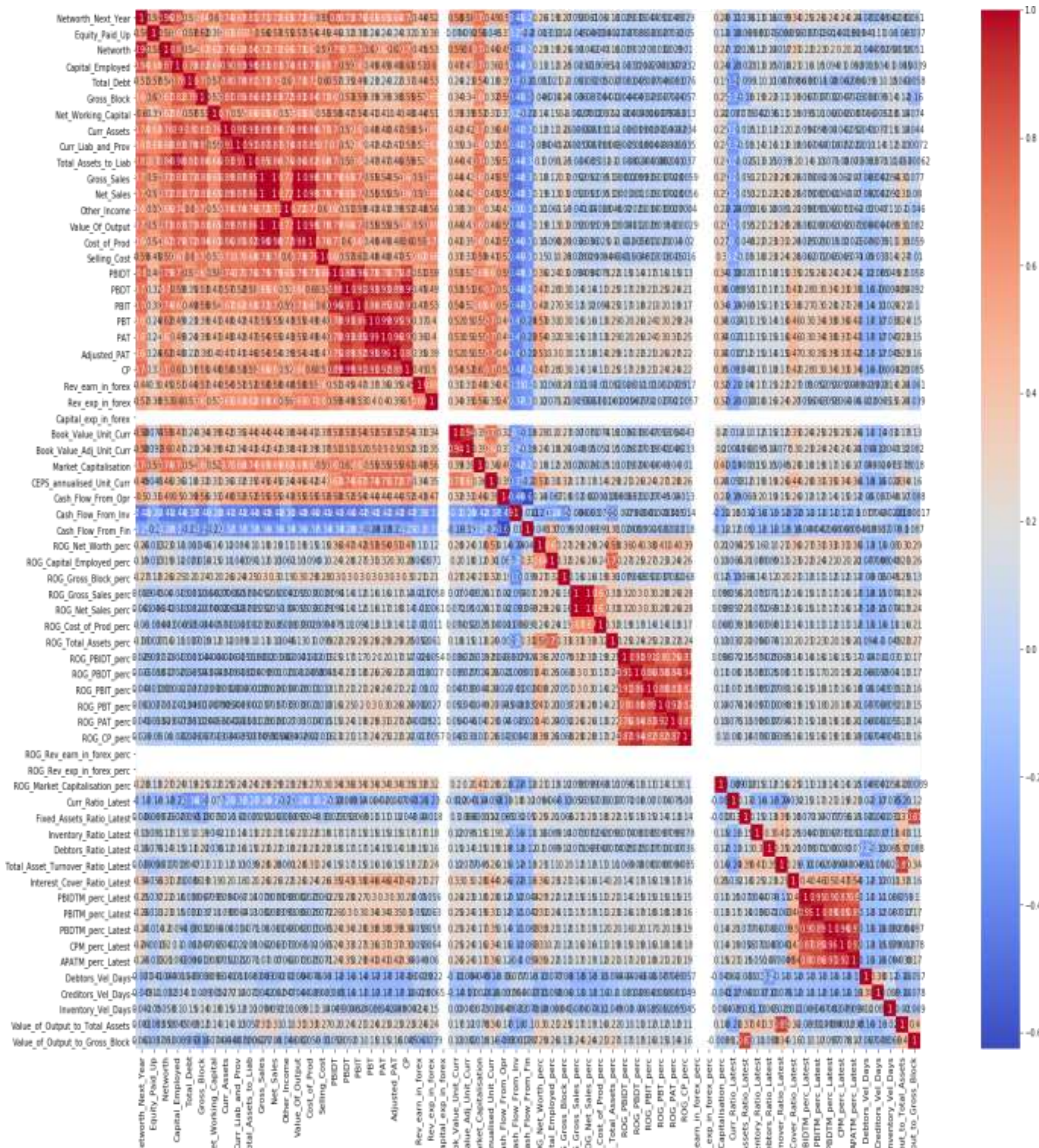


Figure 3.1 - All variables in the dataset.

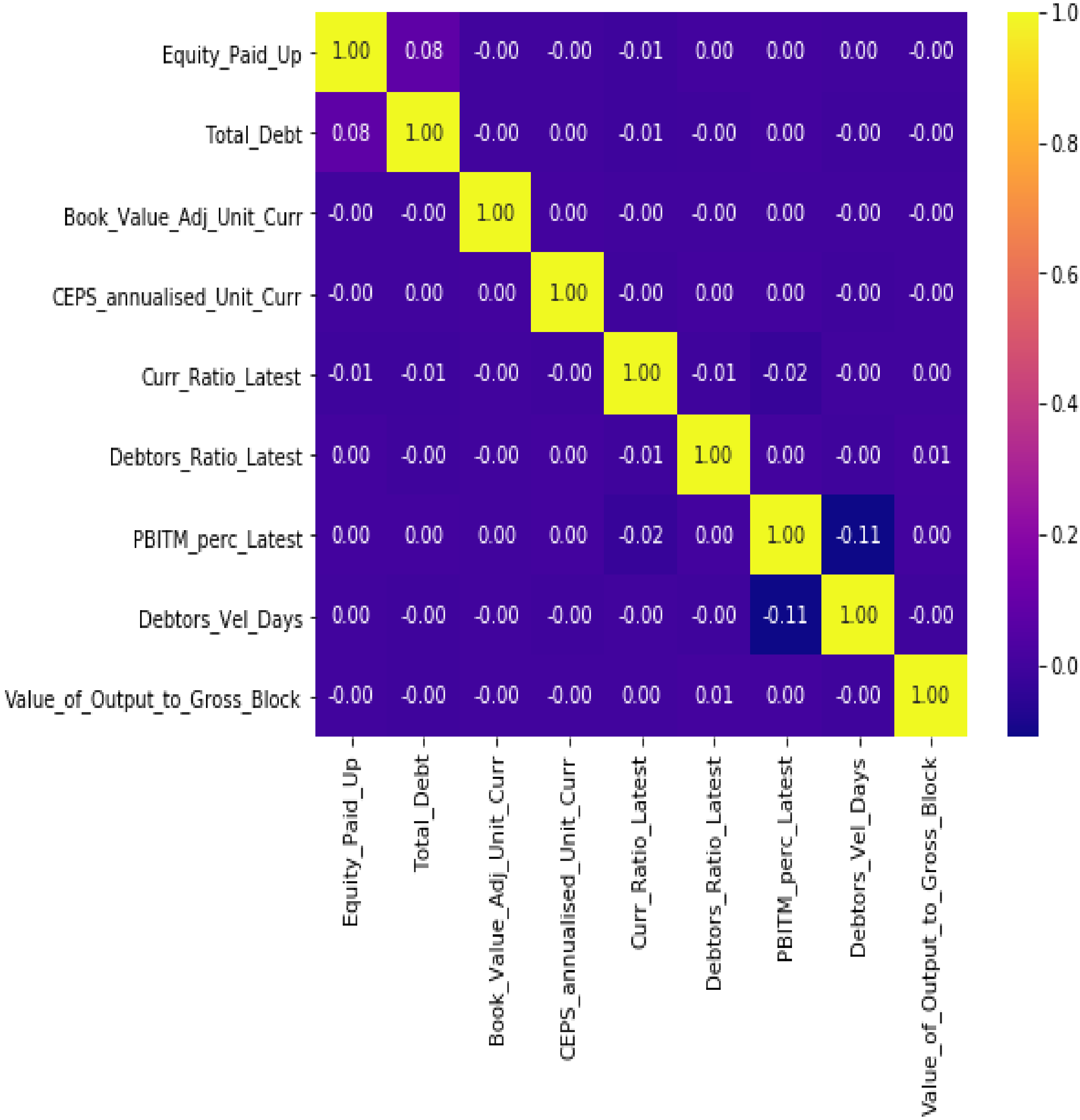


Figure 3.2 – Significant Variables having no correlation, suitable for the model.

Transform Target variable into 0 and 1:

- Using np.where function converted the Net worth next year variable into 0 and 1.
- 1 denoting the company net worth is negative and 0 when it's positive.
- Value counts was done and found that 89 percent of data is Positive and 10 percent is 10 percent odd is negative.
- Data is imbalanced.

Train and Test Split:

- From Sklearn Train test split function was imported and the split was done,

X_train - `(2402, 33)`
X_test - `(1184, 33)`

- Independent variables was split into train and test using ratio of 67:33 and random state 42.
- The training dataset has 2402 rows while the testing dataset has 1184 rows.
- 33 independent variables were selected after applying **variable inflation factor.** As variables were correlated and multicollinearity treatment was done.
- **Standard scaler** also was used to scale them as companies are from different industries and a common scale brought in for applying the model.
- **Smote function** was used to create a dataset **train smote** for to regularize the data as the data was imbalanced.

Logistic Regression Model approach & optimum Cutoff:

- Stats Model Library was used to build a logistic model.
- The P value was obtained by the summary and insignificant predictors were removed one by one to get 9 predictors which are significant.
- Below are the significant predictors which can predict Net worth negative next year.

Equity_Paid_Up
Total_Debt
Book_Value_Adj_Unit_Curr
CEPS_annualised_Unit_Curr
Curr_Ratio_Latest
Debtors_Ratio_Latest
PBITM_perc_Latest
Debtors_Vel_Days
Value_of_Output_to_Gross_Block

Out[167]:

Logit Regression Results

Dep. Variable:	default	No. Observations:	2402
Model:	Logit	Df Residuals:	2392
Method:	MLE	Df Model:	9
Date:	Sun, 05 Dec 2021	Pseudo R-squ.:	0.6361
Time:	14:55:48	Log-Likelihood:	-287.97
converged:	True	LL-Null:	-791.34
Covariance Type:	nonrobust	LLR p-value:	6.039e-211

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-6.9424	0.435	-15.954	0.000	-7.795	-6.089
Equity_Paid_Up	-0.4473	0.156	-2.860	0.004	-0.754	-0.141
Total_Debt	0.8182	0.176	4.638	0.000	0.472	1.164
Book_Value_Adj_Unit_Curr	-5.9915	0.535	-11.200	0.000	-7.040	-4.943
CEPS_annualised_Unit_Curr	-0.9580	0.257	-3.732	0.000	-1.461	-0.455
Curr_Ratio_Latest	-0.9629	0.158	-6.108	0.000	-1.272	-0.654
Debtors_Ratio_Latest	-0.3336	0.134	-2.484	0.013	-0.597	-0.070
PBITM_perc_Latest	-0.6933	0.134	-5.184	0.000	-0.955	-0.431
Debtors_Vel_Days	-0.2609	0.106	-2.458	0.014	-0.469	-0.053
Value_of_Output_to_Gross_Block	-0.3378	0.161	-2.104	0.035	-0.653	-0.023

- **The Coefficients are clearly showing a negative sign and the debt coefficient a positive to indicate the predictors are significant enough to predict the company Net worth next year is negative.**

Performance Metrics:

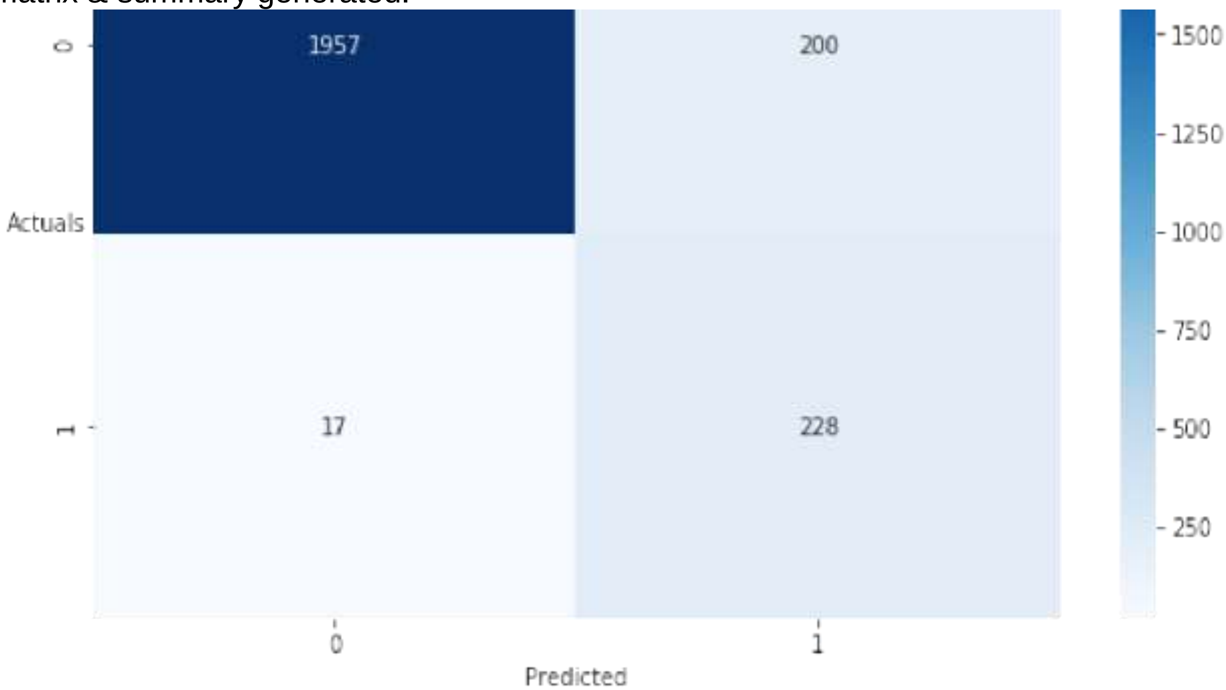
- With default cutoff at 0.5 applied to Train dataset the below Confusion matrix & summary generated.

	precision	recall	f1-score	support
0	0.965	0.991	0.978	2157
1	0.894	0.686	0.776	245
accuracy			0.960	2402
macro avg	0.929	0.838	0.877	2402
weighted avg	0.958	0.960	0.957	2402

Inference:

Recall at 68 percent and precision at 89 percent which only 68% is predicted correctly with a default cutoff 0.5. But Specificity 99 percent indicates that the most companies are ending up with a positive net worth next year.

- Optimum threshold was obtained using ROC curve from Sklearn metrics which is 0.11.
- With Optimum cutoff at 0.11 applied to Train dataset, the below Confusion matrix & summary generated.



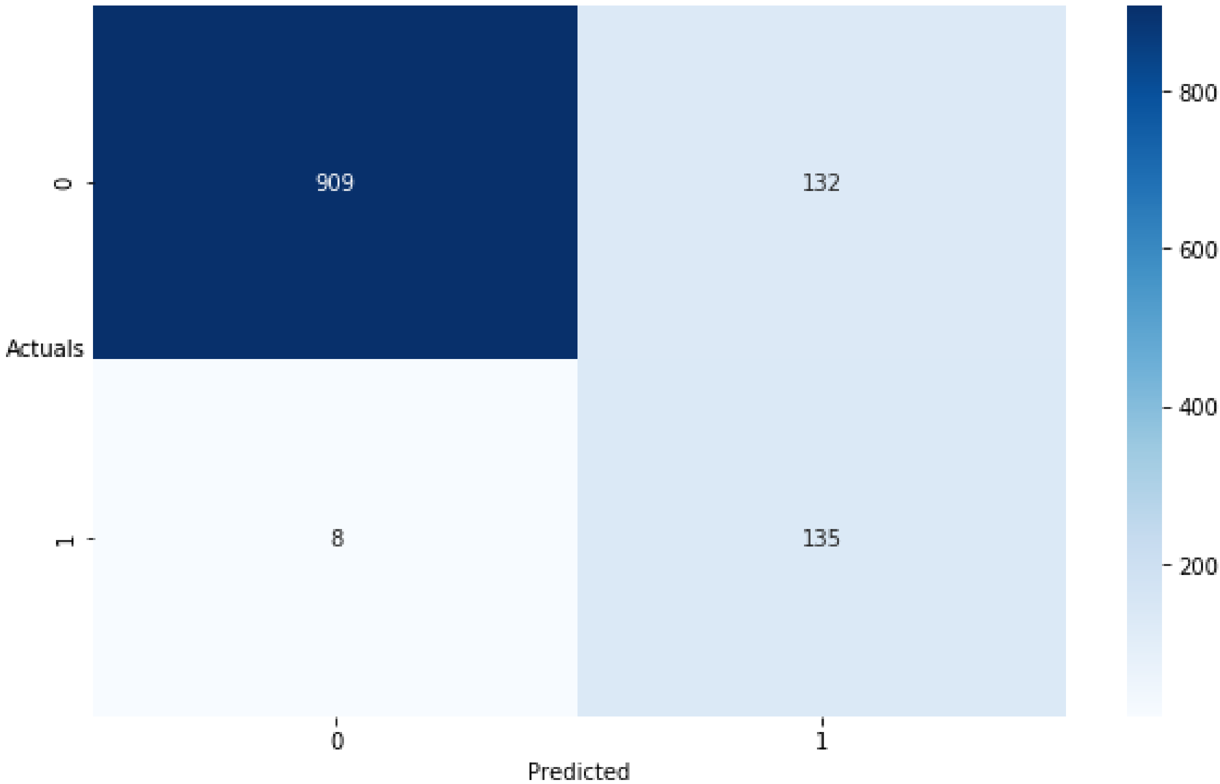
-

	Precision	recall	f1-score	support
0	0.991	0.907	0.947	2157
1	0.533	0.931	0.678	245
accuracy			0.910	2402
macro avg	0.762	0.919	0.813	2402
weighted avg	0.945	0.910	0.920	2402

Inference:

Recall at 93 percent and precision at 53 percent which 93% is predicted correctly with an optimum cutoff 0.11. Recall has improved lot. Specificity reduced to 90 percent indicates that the most companies are ending up with a positive net worth next year.

- With Optimum cutoff at 0.11 applied to Test dataset, the below Confusion matrix & summary generated.

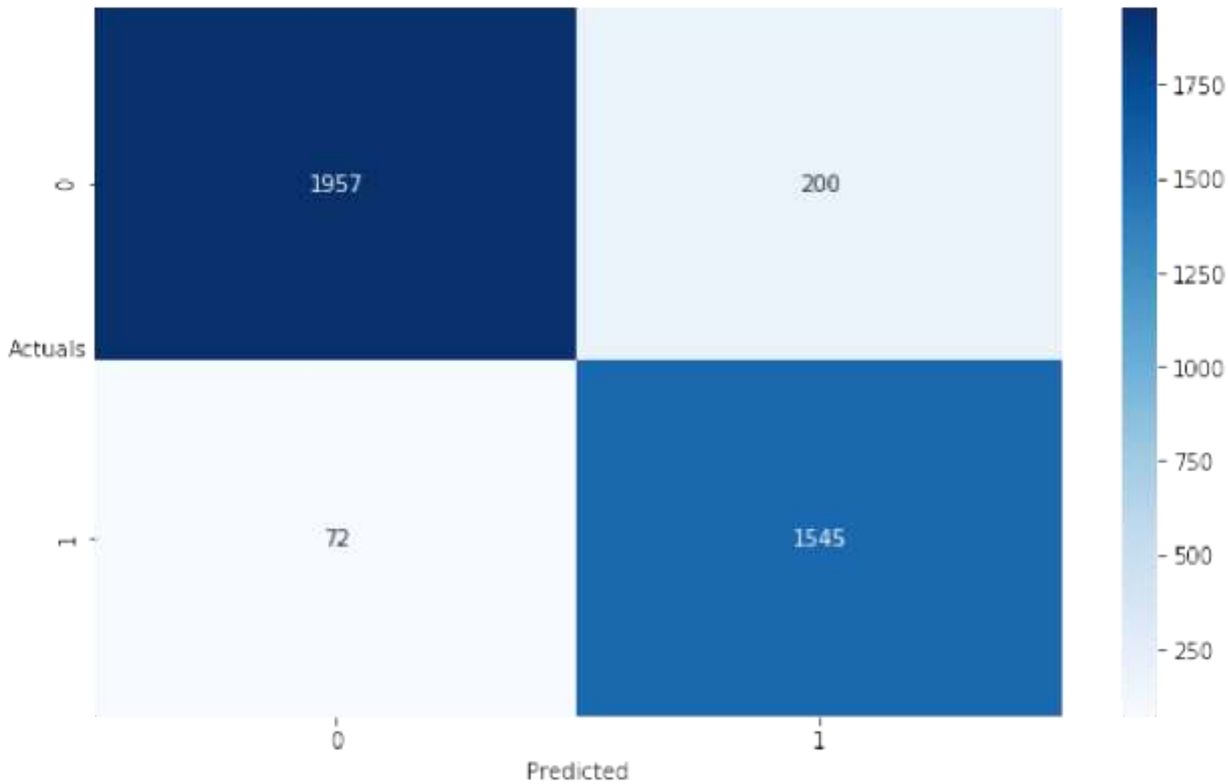


	precision	recall	f1-score	support
0	0.991	0.873	0.928	1041
1	0.506	0.944	0.659	143
accuracy			0.882	1184
macro avg	0.748	0.909	0.794	1184
weighted avg	0.933	0.882	0.896	1184

Inference:

Recall at 94 percent and precision at 50 percent which 94% is predicted correctly with an optimum cutoff 0.11. Recall has improved lot. Specificity at 87 percent and precision at 50 percent is a worrying factor.

- With Optimum cutoff at 0.11 applied to Train smote dataset, the below Confusion matrix & summary generated. The data is generalized.



	precision	recall	f1-score	support
0	0.965	0.907	0.935	2157
1	0.885	0.955	0.919	1617
accuracy			0.928	3774
macro avg	0.925	0.931	0.927	3774
weighted avg	0.931	0.928	0.928	3774

Inference:

Recall at 95 percent and precision at 88 percent which 95% is predicted correctly with an optimum cutoff 0.11 is a very good model when smote is applied. Recall is at maximum compared to past 3 summary. Both Recall and precision are high with a regularized data.

Final Interpretation on the Model:

From an Investor's Point of view, Companies having a Total debt Coefficient >0.81 can be avoided for investing. As it's likely to get into negative territory.

Book value of companies lesser than -5.99 can be taken as an Exit Call as the company is more likely to go into Degrowth phase.

Milestone 2

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

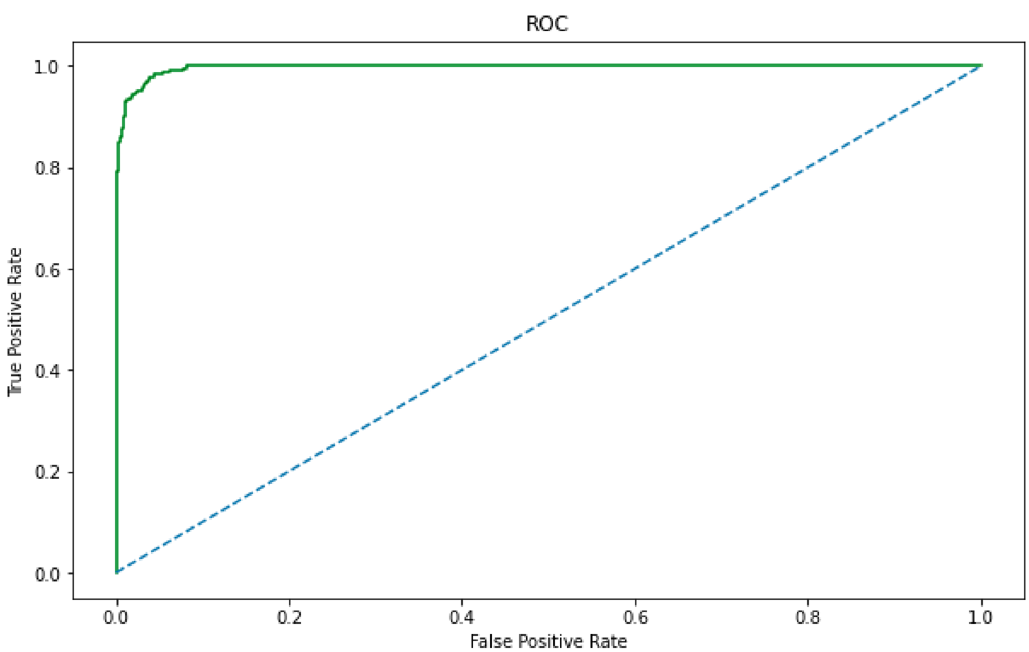
Best Params

```
Out[241]: {'max_depth': 7,
           'max_features': 8,
           'min_samples_leaf': 5,
           'min_samples_split': 50,
           'n_estimators': 800}
```

Train Data

AUC

```
Area under Curve is 0.9964160351205851
```



Confusion matrix

```
Out[422]: array([[2143,   14],
                 [  33,  212]], dtype=int64)
```

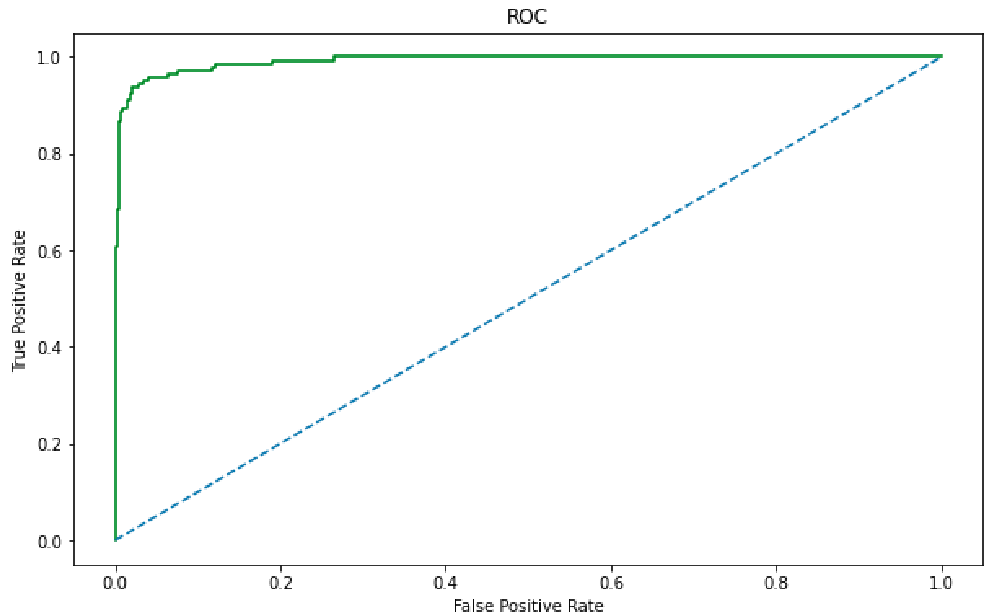
Classification report

```
              precision    recall  f1-score   support

           0       0.98      0.99      0.99      2157
           1       0.94      0.87      0.90       245

    accuracy                           0.98      2402
   macro avg       0.96      0.93      0.94      2402
weighted avg       0.98      0.98      0.98      2402
```

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

Test data

```
Area under Curve is 0.9913813372026627
```



Confusion matrix

```
array([[1031,   10],
       [  16,  127]], dtype=int64)
```

Classification report

	precision	recall	f1-score	support
0	0.98	0.99	0.99	1041
1	0.93	0.89	0.91	143
accuracy			0.98	1184
macro avg	0.96	0.94	0.95	1184
weighted avg	0.98	0.98	0.98	1184

Inference:

- From Sklearn imported grid search & random forest classifier used grid search to get the ideal hyper parameters and to fit the estimator.
- Fit it on to the Train and Test dataset.
- Got the best parameters.
- Computed confusion matrix, Summary and ROC curve AUC values.

Train

- Recall – 87
- Precision - 94
- Accuracy – 98
- AUC – 99

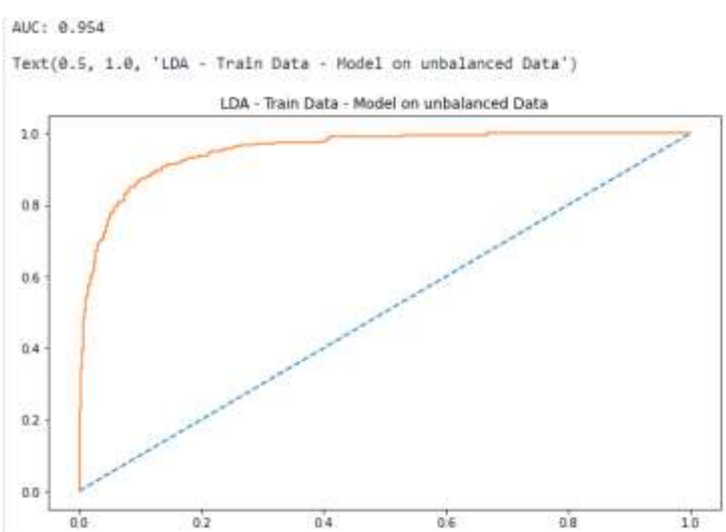
Test

- Recall – 89
- Precision - 93
- Accuracy – 98
- AUC - 99

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

Train Data

AUC



Confusion matrix

```
array([[1937,  220],
       [  31,  214]], dtype=int64)
```

Classification report

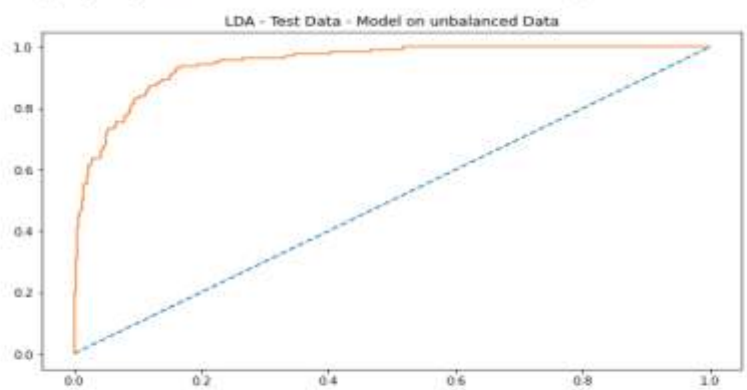
	precision	recall	f1-score	support
0	0.98	0.90	0.94	2157
1	0.49	0.87	0.63	245
accuracy			0.90	2402
macro avg	0.74	0.89	0.78	2402
weighted avg	0.93	0.90	0.91	2402

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

Test Data

AUC

```
AUC: 0.948
Text(0.5, 1.0, 'LDA - Test Data - Model on unbalanced Data')
```



Confusion matrix

```
array([[904, 137],
       [ 17, 126]], dtype=int64)
```

Classification report

```
              precision    recall  f1-score   support

           0       0.98      0.87      0.92      1041
           1       0.48      0.88      0.62       143

    accuracy                           0.87      1184
   macro avg       0.73      0.87      0.77      1184
weighted avg       0.92      0.87      0.89      1184
```

Inference:

- From Sklearn, imported linear discriminant analysis.
- Fit it on to the Train & Test dataset.
- Go the optimum value of 0.13
- Predicted on both train, test.
- Computed confusion matrix, Summary and ROC curve AUC values.

Train

- Recall – 87
- Precision - 49
- Accuracy – 90
- AUC –95

Test

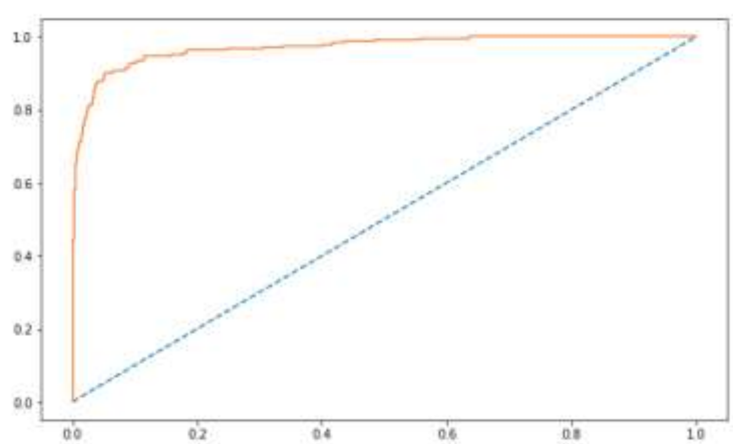
- Recall – 88
- Precision - 48
- Accuracy – 87
- AUC - 94

Logistic Regression

Train Data

AUC

```
AUC: 0.970
[<matplotlib.lines.Line2D at 0x1924a56ce80>]
```



Confusion matrix

```
array([[2064,   93],
       [  30,  215]], dtype=int64)
```

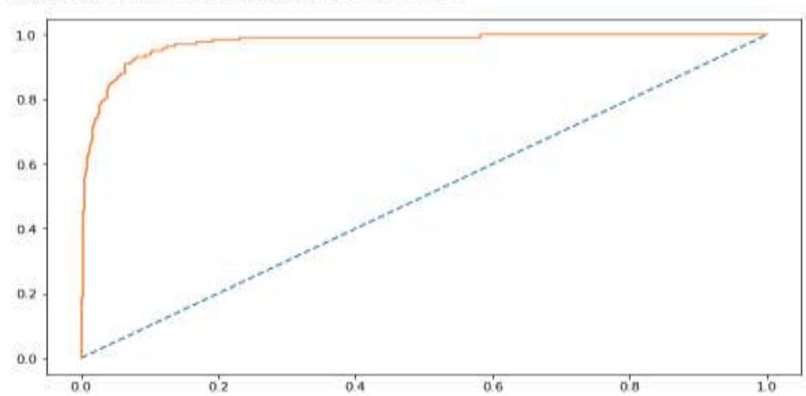
Classification Report

	precision	recall	f1-score	support
0	0.99	0.96	0.97	2157
1	0.70	0.88	0.78	245
accuracy			0.95	2402
macro avg	0.84	0.92	0.87	2402
weighted avg	0.96	0.95	0.95	2402

Test Data

AUC

```
AUC: 0.975
[<matplotlib.lines.Line2D at 0x1924b2d6b50>]
```



Confusion matrix

```
array([[975,  66],
       [ 16, 127]], dtype=int64)
```

Classification matrix

```
              precision    recall  f1-score   support

           0       0.98      0.94      0.96      1041
           1       0.66      0.89      0.76       143

    accuracy                           0.93      1184
   macro avg       0.82      0.91      0.86      1184
weighted avg       0.94      0.93      0.94      1184
```

Inference:

- From Sklearn, imported linear discriminant analysis.
- Predicted on both train, test and train with smote dataset.
- Computed confusion matrix, Summary and ROC curve AUC values.

Train

- Recall – 88
- Precision - 70
- Accuracy – 95
- AUC –97

Test

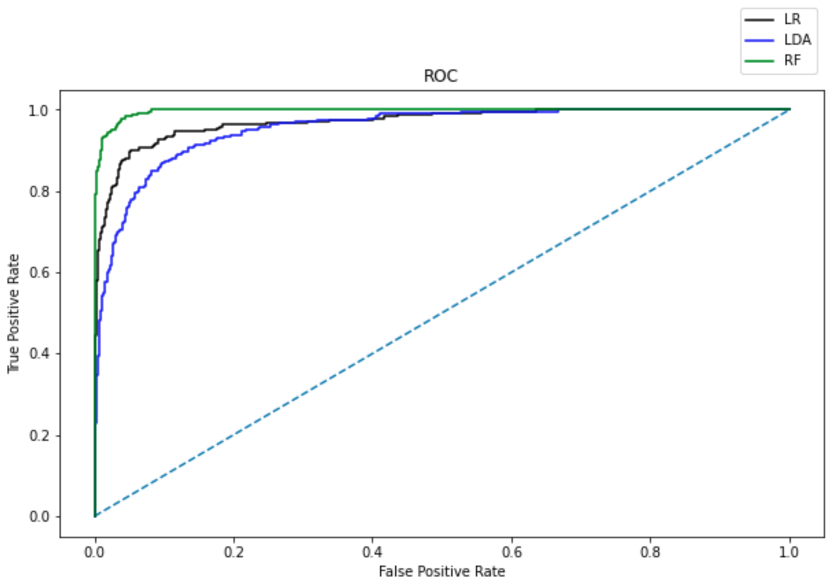
- Recall – 89
- Precision - 66
- Accuracy – 93
- AUC - 97

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

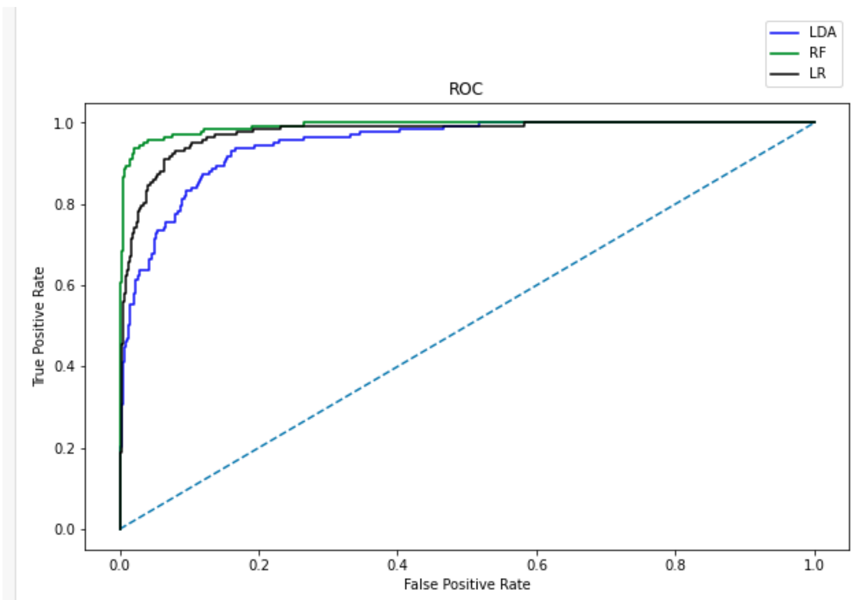
	LR Train	LR Test	Random Forest Train	Random Forest Test	LDA Train	LDA Test
Accuracy	0.96	0.95	0.98	0.98	0.94	0.93
AUC	0.97	0.97	1.00	0.99	0.95	0.95
Recall	0.88	0.89	0.87	0.89	0.87	0.88
Precision	0.70	0.66	0.94	0.93	0.49	0.48
F1 Score	0.78	0.76	0.90	0.91	0.63	0.62

ROC Curve of Train data

<matplotlib.legend.Legend at 0x19d921eeee0>



ROC Curve of Test data



1.13 State Recommendations from the above models

Inference and Recommendations:

- From the graph we can predict that "Random Forest" predicts the data best as it is the curve which is above all the curve in training and test data when compared to all the 3 models. The more it shifts to the left side i.e. true positive rate the better the model is.
- The Logistic Regression is to the left side when compared to Linear Discriminant Analysis so when compared to all 3 RF is best and second best model is LR.

- The data is all about the Credit Risk and the Businesses or companies can fall prey to default and we are supposed to find whether to invest or not.
- So I would like to suggest to invest in the companies as the Accuracy is best in the companies and default status is low.

Market Risk

2.1 Draw Stock Price Graph (Stock Price vs Time) for any 2 given stocks with inference

Read:

	Date	Infosys	Indian Hotel	Mahindra & Mahindra	Axis Bank	SAIL	Shree Cement	Sun Pharma	Jindal Steel	Idea Vodafone	Jet Airways
0	31-03-2014	264	69	455	263	68	5543	555	298	83	278
1	07-04-2014	257	68	458	276	70	5728	610	279	84	303
2	14-04-2014	254	68	454	270	68	5649	607	279	83	280
3	21-04-2014	253	68	488	283	68	5692	604	274	83	282
4	28-04-2014	256	65	482	282	63	5582	611	238	79	243

Checking data types of all columns:

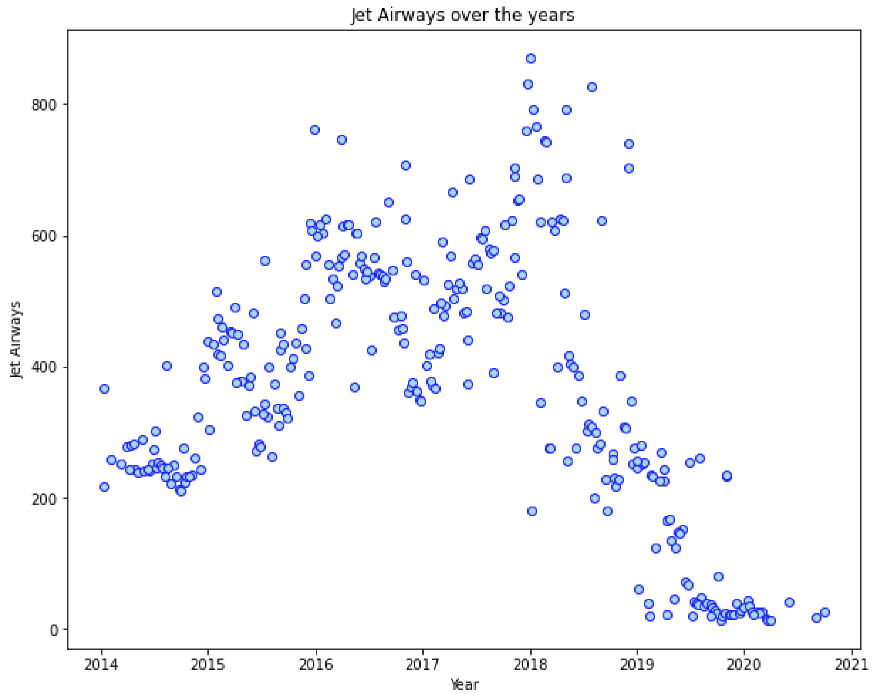
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 314 entries, 0 to 313
Data columns (total 11 columns):
 #   Column               Non-Null Count  Dtype 
---  ------               --------------  ----- 
 0   Date                 314 non-null    object
 1   Infosys              314 non-null    int64 
 2   Indian Hotel         314 non-null    int64 
 3   Mahindra & Mahindra  314 non-null    int64 
 4   Axis Bank            314 non-null    int64 
 5   SAIL                 314 non-null    int64 
 6   Shree Cement         314 non-null    int64 
 7   Sun Pharma           314 non-null    int64 
 8   Jindal Steel         314 non-null    int64 
 9   Idea Vodafone        314 non-null    int64 
 10  Jet Airways          314 non-null    int64 
dtypes: int64(10), object(1)
memory usage: 27.1+ KB
```

Descriptive stats:

	Infosys	Indian Hotel	Mahindra & Mahindra	Axis Bank	SAIL	Shree Cement	Sun Pharma	Jindal Steel	Idea Vodafone	Jet Airways
count	314.000000	314.000000	314.000000	314.000000	314.000000	314.000000	314.000000	314.000000	314.000000	314.000000
mean	511.340764	114.560510	636.678344	540.742038	59.095541	14806.410828	633.468153	147.627389	53.713376	372.659236
std	135.952051	22.509732	102.879975	115.835569	15.810493	4288.275085	171.855893	65.879195	31.248985	202.262668
min	234.000000	64.000000	284.000000	263.000000	21.000000	5543.000000	338.000000	53.000000	3.000000	14.000000
25%	424.000000	96.000000	572.000000	470.500000	47.000000	10952.250000	478.500000	88.250000	25.250000	243.250000
50%	466.500000	115.000000	625.000000	528.000000	57.000000	16018.500000	614.000000	142.500000	53.000000	376.000000
75%	630.750000	134.000000	678.000000	605.250000	71.750000	17773.250000	785.000000	182.750000	82.000000	534.000000
max	810.000000	157.000000	956.000000	808.000000	104.000000	24806.000000	1089.000000	338.000000	117.000000	871.000000

Let's us plot & see price trend over time for different companies

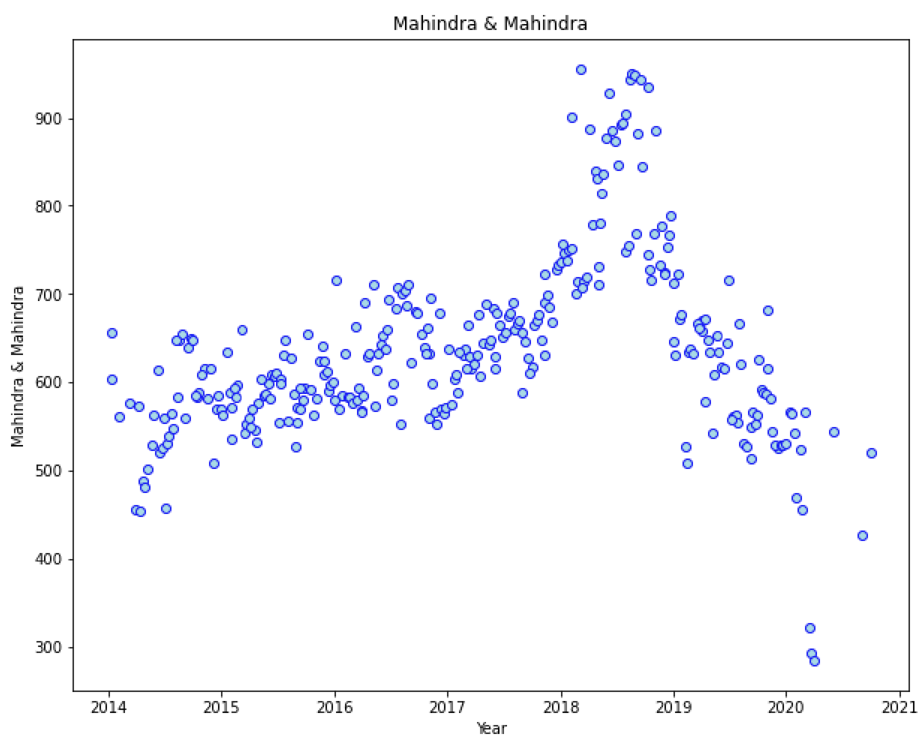
1. **Jet Airways**



Inference:

- Simple Line Chart of Infosys dated from 2014 to 2021.
- Trend in an upward trajectory.
- On early 2018, the stock price broke the resistance of 550 to inch higher.
- Made a High of 800 odd levels before falling to 550 levels during covid 19 initial days of early 2020's.

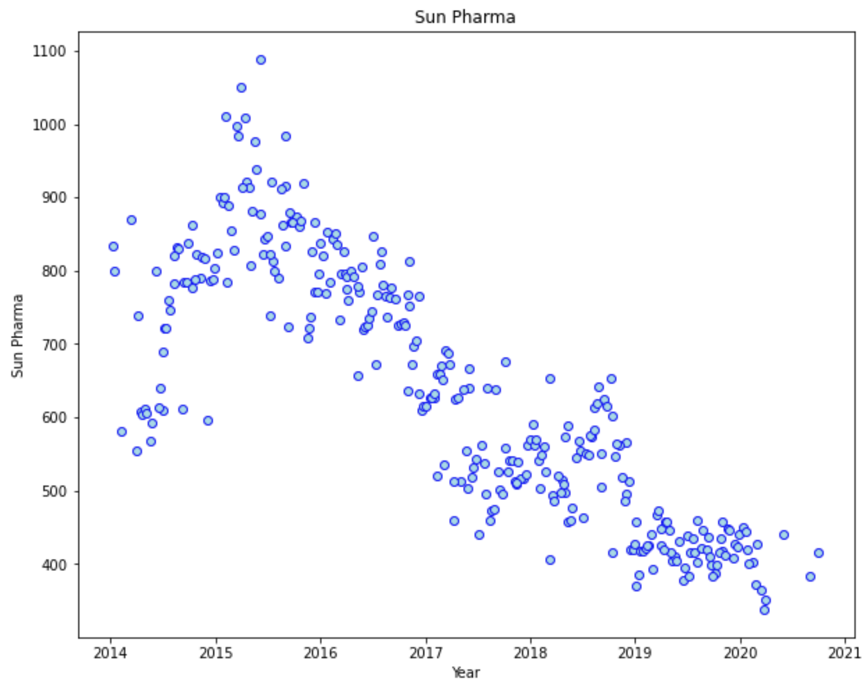
2. **Mahindra & Mahindra**



Inference:

- Simple Line Chart of Mahindra & Mahindra dated from 2014 to 2020.
- Trend in a downward trajectory.
- In between 2015 to 2016 the value of stock was high around 1100,
- Later there is fall from 2016 till 2021.

3. Sun Pharma



Inference:

- Simple Line Chart of Sun Pharma dated from 2014 to 2021.
- Trend in a downward trajectory.
- On early 2019, the stock price made a high of 900 per share.
- From 2019 to 2021 the stock price is fallen.

2.2 Calculate Returns for all stocks with inference

	Infosys	Indian Hotel	Mahindra & Mahindra	Axis Bank	SAIL	Shree Cement	Sun Pharma	Jindal Steel	Idea Vodafone	Jet Airways
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	-0.026873	-0.014599	0.006572	0.048247	0.028988	0.032831	0.094491	-0.065882	0.011976	0.086112
2	-0.011742	0.000000	-0.008772	-0.021979	-0.028988	-0.013888	-0.004930	0.000000	-0.011976	-0.078943
3	-0.003945	0.000000	0.072218	0.047025	0.000000	0.007583	-0.004955	-0.018084	0.000000	0.007117
4	0.011788	-0.045120	-0.012371	-0.003540	-0.076373	-0.019515	0.011523	-0.140857	-0.049393	-0.148846

Inference:

- Returns = Price(t) – Price(t-1)/ Price(t-1)
- Stock returns was calculated by taking logarithms & Differences.
- As the data is a weekly data of 6 yrs, the returns are for consecutive weeks.
- 1st Week of Infosys shows a negative 2.6 percent and 4th Week positive 1.1 percent returns compared to previous week.
- Likewise all stocks returns are calculated by the difference of the previous week, row wise.
- In 1st week some stock price is good with positive value and in 4th week most of the stock values are less with negative values.

2.3 Calculate Stock Means and Standard Deviation for all stocks with inference

Calculating stock means

```
Shree Cement            0.003681
Infosys                 0.002794
Axis Bank               0.001167
Indian Hotel            0.000266
Sun Pharma             -0.001455
Mahindra & Mahindra    -0.001506
SAIL                   -0.003463
Jindal Steel           -0.004123
Jet Airways            -0.009548
Idea Vodafone          -0.010608
dtype: float64
```

Calculating stock standard deviation

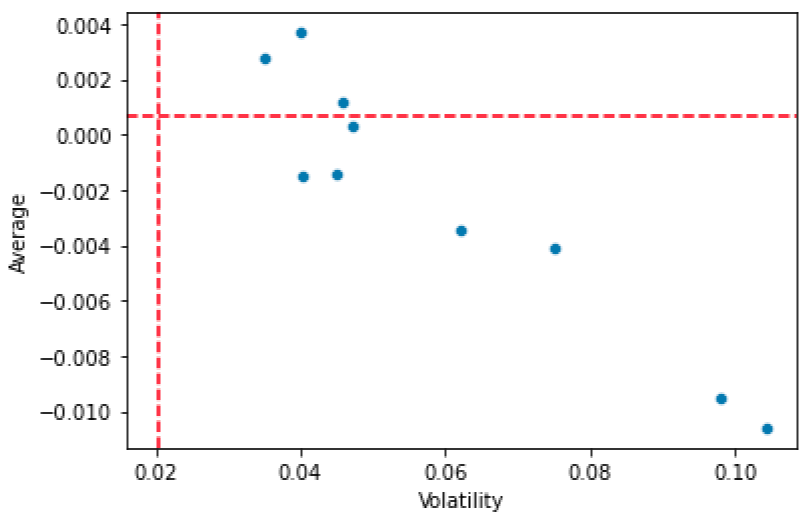
```
Idea Vodafone           0.104315
Jet Airways             0.097972
Jindal Steel            0.075108
SAIL                    0.062188
Indian Hotel            0.047131
Axis Bank               0.045828
Sun Pharma              0.045033
Mahindra & Mahindra     0.040169
Shree Cement            0.039917
Infosys                 0.035070
dtype: float64
```

Inference:

- Shree cement and Infosys are giving higher returns in the past 6 years.
- Idea Vodafone and Jet airways giving a negative returns.
- It's how much average return of the stock vary from the average return of the stock.
- Considered as risk.
- More Volatility stocks are often termed as risky.
- Infosys and Shree cement vary less with their average return.
- Idea Vodafone and Jet airways vary large with their average returns.

2.4 Draw a plot of Stock Means vs Standard Deviation and state your inference

	Average	Volatility
Infosys	0.002794	0.035070
Indian Hotel	0.000266	0.047131
Mahindra & Mahindra	-0.001506	0.040169
Axis Bank	0.001167	0.045828
SAIL	-0.003463	0.062188
Shree Cement	0.003681	0.039917
Sun Pharma	-0.001455	0.045033
Jindal Steel	-0.004123	0.075108
Idea Vodafone	-0.010608	0.104315
Jet Airways	-0.009548	0.097972



Inference:

- As per the data of 10 Stocks
- From the above Graph, the stocks which has high volatility is giving lesser returns. In case of Idea Vodafone & jet airways.
- More returns are visible with lesser volatility, in case of axis bank, shree cement, and Infosys & Indian hotel.

Conclusion & Recommendations:

- Normally higher returns with less risk is part of the portfolio. (Axis bank, shree cement, Mahindra & Mahindra)
- Lower mean and higher standard deviation are not preferred in portfolio. (Idea Vodafone, jet airways)