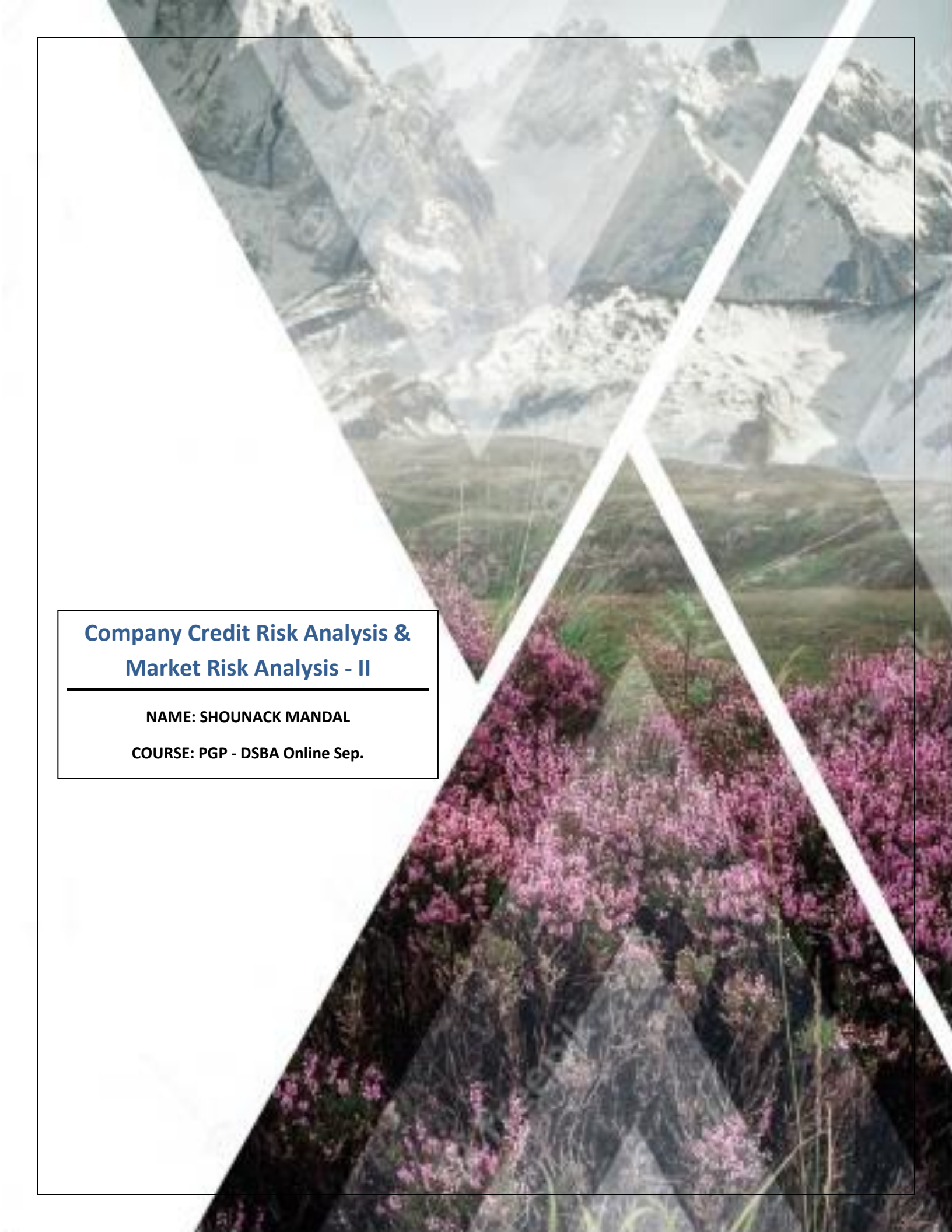


Company Credit Risk Analysis & Market Risk Analysis - II

NAME: SHOUNACK MANDAL

COURSE: PGP - DSBA Online Sep.

Table of contents

Problem Statement

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

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

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

1.8 Build a Random Forest Model on Train Dataset.

Using Scikit_Learn RandomisedSearchCV method, we can define a grid of hyperparameter ranges and randomly sample from the grid, performing K-Fold CV with each combination of values.

1.8.1 Showcase your model building approach

We now fit our model to the GridSearchCV for Random Forest model by training the model with our independent variable and dependent variables

- n_estimators = number of trees in the forest
- max_features = max number of features considered for splitting a node
- max_depth = max number of levels in each decision tree
- min_samples_split = min number of data points placed in a node before the node is split
- min_samples_leaf = min number of data points allowed in a leaf node

```
param_grid = {
    'max_depth': [15,20],
    'min_samples_leaf': [10,20],
    'min_samples_split': [ 50,100],
    'n_estimators': [301,401,701]
```

The probability on the training and test set;

	0	1
0	0.98	0.02
1	0.99	0.01
2	0.93	0.07
3	0.95	0.05
4	1.00	0.00

	0	1
0	0.99	0.01
1	0.97	0.03
2	0.86	0.14
3	0.21	0.79
4	0.94	0.06

Feature importance;

Top three are most important features, Book_Value_Unit_Curr, networth, Book_Value_Adj_Unit_Curr.

	IMP		
Book_Value_Unit_Curr	0.21	Gross_Sales	0.00
Networth	0.21	Net_Sales	0.00
Book_Value_Adj_Unit_Curr	0.19	Fixed_Assets_Ratio_Latest	0.00
Curr_Ratio_Latest	0.06	Creditors_Vel_Days	0.00
Capital_Employed	0.04	Market_Capitalisation	0.00
CP	0.03	ROG_Gross_Sales_perc	0.00
PBIDT	0.03	Value_Of_Output	0.00
CEPS_annualised_Unit_Curr	0.02	Cash_Flow_From_Inv	0.00
Net_Working_Capital	0.02	ROG_Capital_Employed_perc	0.00
PBDT	0.02	Debtors_Vel_Days	0.00
Adjusted_PAT	0.01	Value_of_Output_to_Total_Assets	0.00
Total_Asset_Turnover_Ratio_Latest	0.01	Selling_Cost	0.00
PBITM_perc_Latest	0.01	ROG_Net_Sales_perc	0.00
PAT	0.01	Equity_Paid_Up	0.00
Total_Debt	0.01	ROG_PBDT_perc	0.00
Interest_Cover_Ratio_Latest	0.01	Debtors_Ratio_Latest	0.00
PBIT	0.01	Rev_exp_in_forex	0.00
ROG_Net_Worth_perc	0.01	Inventory_Vel_Days	0.00
PBIDTM_perc_Latest	0.01	Cash_Flow_From_Opr	0.00
PBT	0.01	Cash_Flow_From_Fin	0.00
Value_of_Output_to_Gross_Block	0.01	ROG_PBIDT_perc	0.00
CPM_perc_Latest	0.00	ROG_CP_perc	0.00
Total_Assets_to_Liab	0.00	Other_Income	0.00
PBDTM_perc_Latest	0.00	ROG_Gross_Block_perc	0.00
APATM_perc_Latest	0.00	ROG_PAT_perc	0.00
Curr_Liab_and_Prov	0.00	ROG_PBIT_perc	0.00
Gross_Block	0.00	Inventory_Ratio_Latest	0.00
ROG_Total_Assets_perc	0.00	ROG_Market_Capitalisation_perc	0.00
Curr_Assets	0.00	ROG_PBT_perc	0.00
ROG_Cost_of_Prod_perc	0.00	Rev_earn_in_forex	0.00
Cost of Prod	0.00	Capital_exp_in_forex	0.00

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

1.9.1 Confusion matrix on the training and test data

Inference

Training data:

- True Negative : 2143
- False Positive : 14
- False Negative : 59
- True Positive : 186

Test data:

- True Negative : 1034
- False Positive : 7
- False Negative : 23
- True Positive : 120

1.9.2 Classification report for training and test data

	precision	recall	f1-score	support
0.0	0.97	0.99	0.98	2157
1.0	0.92	0.77	0.84	245
accuracy			0.97	2402
macro avg	0.95	0.88	0.91	2402
weighted avg	0.97	0.97	0.97	2402

	precision	recall	f1-score	support
0.0	0.98	0.99	0.99	1041
1.0	0.94	0.84	0.89	143
accuracy			0.97	1184
macro avg	0.96	0.92	0.94	1184
weighted avg	0.97	0.97	0.97	1184

Inferences

Train Data:

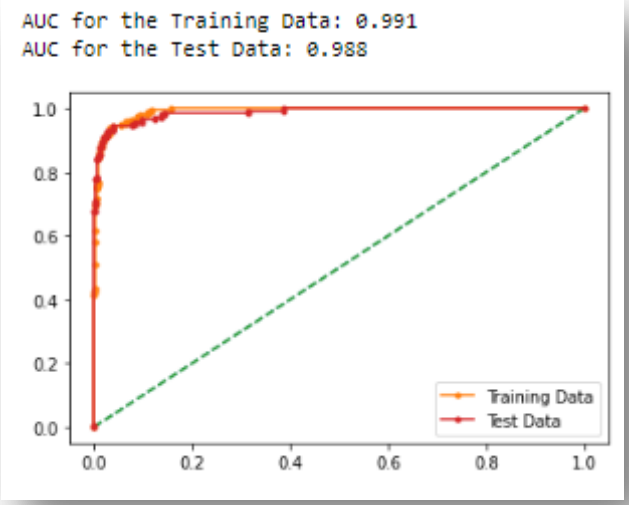
- Accuracy: 97%
- precision : 93%
- recall : 76%
- f1 :84%

Test Data:

- Accuracy: 97%
- precision: 94%
- recall : 84%
- f1 : 89%

AUC and ROC for the training data and test data

AUC for the Training Data: 0.991
AUC for the Test Data: 0.988



Here, recall has increased to 84% from 76% in test data even F1 Score is also increased to 89% with precision of 94%. It is comparatively good model.

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

- Linear Discriminant Analysis (LDA) is a dimensionality reduction technique which is commonly used for the supervised classification problems.
- It is used for modeling differences in groups i.e., separating two or more classes. It is used to project the features in higher dimension space into a lower dimension space.
- library used in LDA is sklearn
- Using GridsearchCV, we input various parameters like 'max_iter', 'penalty',solver', 'tol' which will helps us to find best grid for prediction of the better model

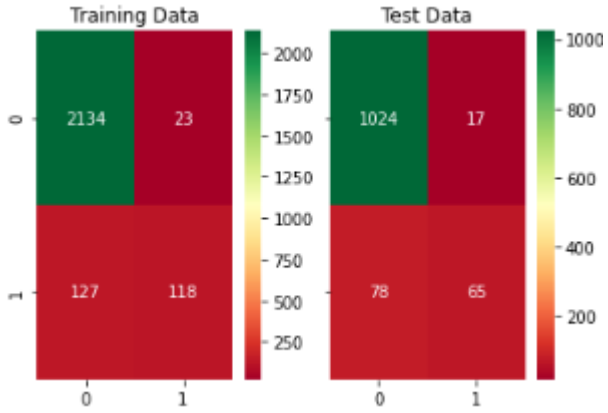
The probability on the training and test set;

	0	1
0	0.96	0.04
1	1.00	0.00
2	0.55	0.45
3	1.00	0.00
4	1.00	0.00

	0	1
0	0.98	0.02
1	0.95	0.05
2	0.86	0.14
3	0.92	0.08
4	1.00	0.00

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

Confusion matrix on the training and test data



Inference

Training data:

- True Negative : 2134
- False Positive : 23
- False Negative : 127
- True Positive : 118

Test data:

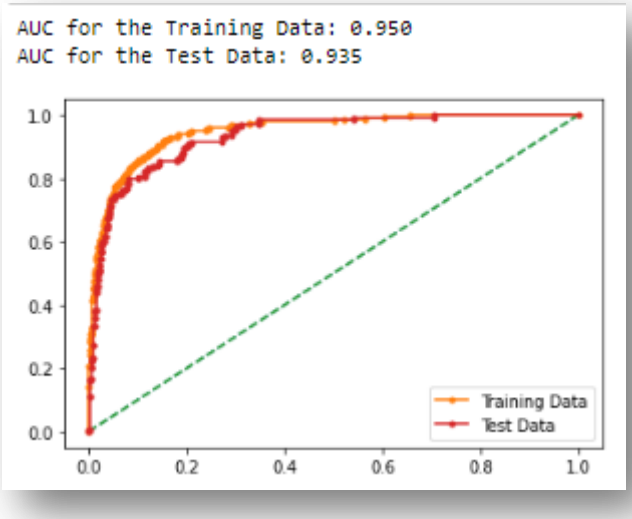
- True Negative : 1024
- False Positive : 17
- False Negative : 78
- True Positive : 65

1.11.1 Classification report for training and test data

	precision	recall	f1-score	support
0.0	0.94	0.99	0.97	2157
1.0	0.84	0.48	0.61	245
accuracy			0.94	2402
macro avg	0.89	0.74	0.79	2402
weighted avg	0.93	0.94	0.93	2402

	precision	recall	f1-score	support
0.0	0.93	0.98	0.96	1041
1.0	0.79	0.45	0.58	143
accuracy			0.92	1184
macro avg	0.86	0.72	0.77	1184
weighted avg	0.91	0.92	0.91	1184

1.11.2 AUC and ROC for the training and test data.



Inference

Train Data:

- AUC: 95%
- Accuracy: 94%
- precision : 84%
- recall :48%
- f1 :61%

Test Data:

- AUC: 93%
- Accuracy: 92%
- precision: 79%
- recall : 45%
- f1 : 58%

In this model recall is very low with 48%. It is not a good model,

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

Models	Data	Accuracy	Precision	Recal	F1 -Score	AUC
Random Forest	Train	97	93	76	84	99
	Test	97	94	84	89	98
Linear Discrimenant Analysis	Train	94	84	48	61	95
	Test	92	79	45	58	93
Logistic Regression	Train	95	86	64	95	96
	Test	95	82	69	75	96

Random forest with grid search performed well with highest recall and good f1 score. Roc Curve shows it's not unfitting or overfitting. While comparing other models, it is observed that Random Forest is best model for credit risk analysis with accuracy of 97%.

Conclusion

Credit report analysis provides information on the credit worthiness of a potential customer The model with selected features will predict a relatively high probability of default. Next step is to integrate with classification model where defaulters further classified into “very high risk”, “high risk”, “medium risk”, “low risk”, etc.

Market Risk Analysis

Problem:

Market Risk
The dataset contains 6 years of information (weekly stock information) on the stock prices of 10 different Indian Stocks. Calculate the mean and standard deviation on the stock returns and share insights.

Project Objective:

The Objective of the report is to explore the Market risk dataset in Python (JUPYTER NOTEBOOK) and generate insights about the dataset. This exploration report will consist of the following:

- Importing the dataset in jupyter notebook.
- Understanding the structure of dataset.
- Exploratory Data analysis
- Graphical exploration
- Calculate the mean and standard deviation on the stock returns
- Insights from the dataset

Data Exploration

Import the market risk data using pandas with Parse_date. We can use the parse_dates parameter to convince pandas to turn things into real datetime types. parse_dates take a list of columns (since you could want to parse multiple columns into datetimes.

	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

Data shape

```
The number of rows 314
The number of columns 11
```

Information on dataset

```
<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
```

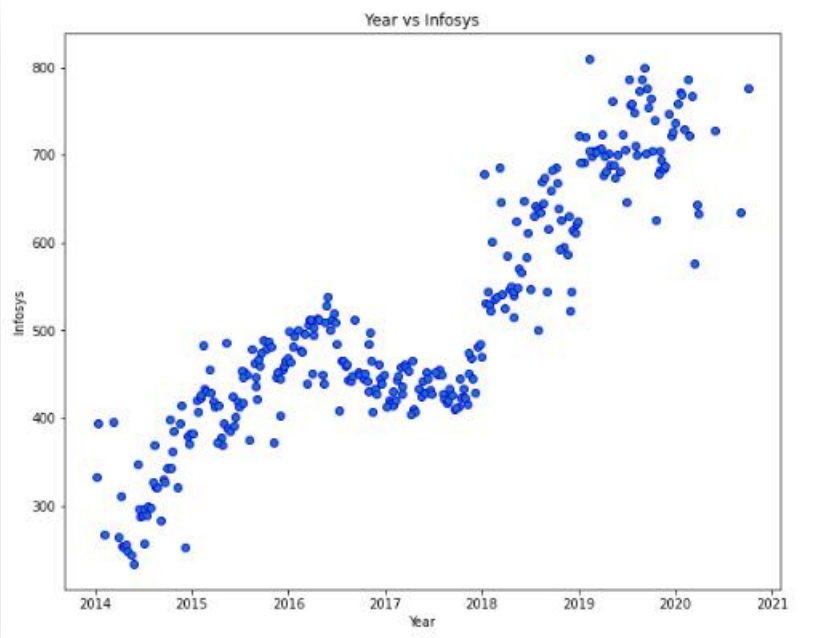
Summary of the dataset

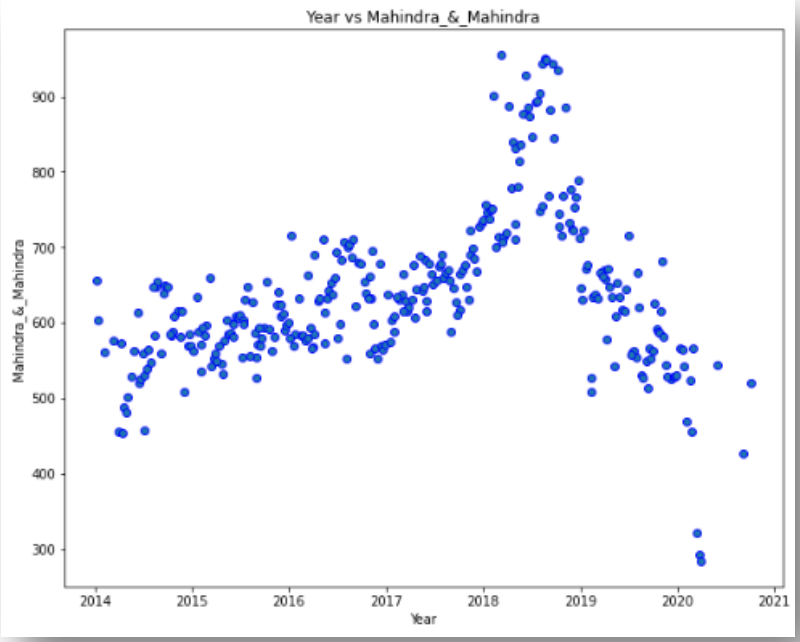
	Infosys	Indian_Hotel	Mahindra_&_Mahindra	Axis_Bank	SAIL	Shree_Cement	Sun_Pharma	Jindal_Steel	Idea_Vodafone	Jet_Airways
count	314.000000	314.000000	314.000000	314.000000	314.000000	314.000000	314.000000	314.000000	314.000000	314.000000
mean	511.340764	114.560510	636.678344	540.742038	59.095541	14806.410828	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%	486.500000	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

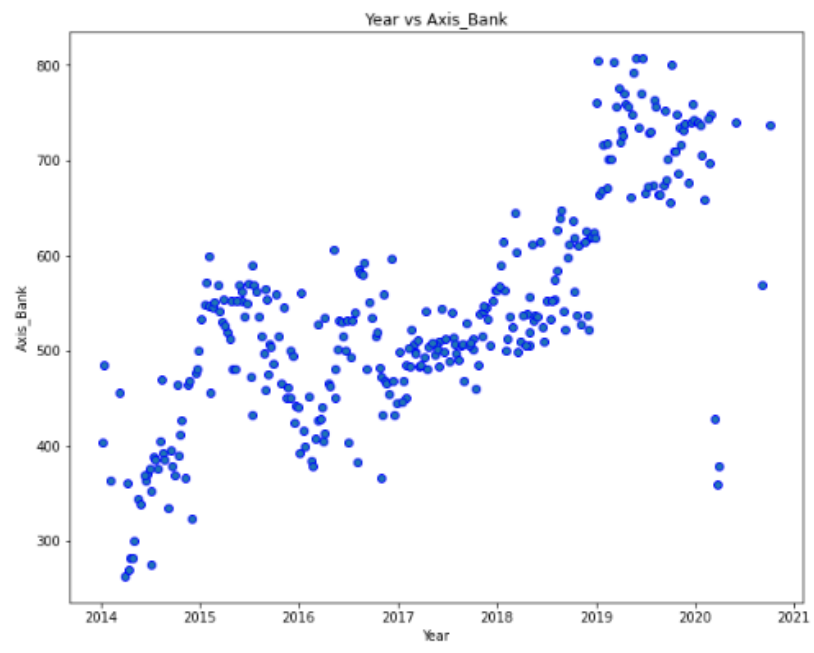
Inference

- Shree Cements have highest stock price
- SAIL Company have low stock price

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







For the first analysis of the price and time chosen stock is Infosys, shows increasing trend for 2014 to 2016, then started consolidating from 2016 to 2018 and then again showed upward trend from 2018 till 2021.

For M&M we see the gradual increase from the start of year till its peak (more than 900) of price in the year 2018 to 2019, afterwards started down fall till 2021.

For Axis bank there is up and down in every years but there is steady growth in the price from start of year till the last year 2021.

2.2 Calculate Returns for all stocks with inference

Head of returns for all stocks by logarithmic differences. Checking top 5 rows giving return values week for week.

	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

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

Stock Means and Standard Deviation for all stocks has been calculated below:-

Stock means:-

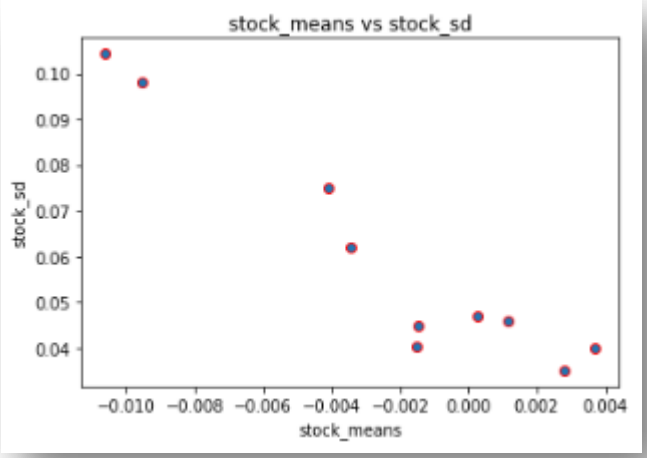
```
Shree_Cement           0.003681
Infosys                0.002794
Axis_Bank              0.001167
Indian_Hotel           0.000266
Sun_Pharma            -0.001455
Mahindra_&_Mahindra   -0.001506
SAIL                  -0.003463
Jindal_Steel          -0.004123
Jet_Airways           -0.009548
Idea_Vodafone         -0.010608
dtype: float64
```

Standard deviations:-

```
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
```

The highest risky stock is idea Vodafone and on the other hand Infosys has the least risk factors, so best in the stock list to invest.

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



Stocks higher up but on the far left indicate high volatility and low returns, while the stocks on the bottom right indicate low volatility and high returns.

During the investment, this graph is very useful in analyzing the risk from different companies.

2.5 Conclusion and Recommendations

Stocks with lower mean and higher SD do not play all the roles in selecting the stocks for portfolio that has competing stock with more return and less risk. Therefore from pure returns perspective, Shree_cement followed by Infosys and axis bank looks good in this dataset. From pure risk perspective as measured by SD, Infosys followed by Shree_Cement and Mahindra & Mahindra looks good in this dataset.

We recommend using the stock means vs standard deviation plot to assess the risk to reward ratio. More volatile stock might give short term gains but might not be a good stock for investment for long term. Whereas low volatile stock might not be a good investment in short term, but might be good return in long term.

Hence based on the type of investment that one is looking for a inference should be made from the above mentioned plot.