

# FINANCE AND RISK ANALYTICS PROJECT REPORT 2

Akshaya Nallathambi

13th February, 2022

---



FINANCE AND RISK  
ANALYTICS  
—

---

---

# Table Of Contents

## Problem 1

Problem statement	4
Data Description	5
Sample of the dataset	8
1.8 Build a Random Forest Model on Train Dataset. Also showcase your model building approach	9
1.9 Validate the Random Forest Model on test Dataset and state the performance matrices. Also state interpretation from the model	9
1.10 Build a LDA Model on Train Dataset. Also showcase your model building approach	11
1.11 Validate the LDA Model on test Dataset and state the performance matrices. Also state interpretation from the model	11
1.12 Compare the performances of Logistics, Radom Forest and LDA models (include ROC Curve)	14
1.13 State Recommendations from the above models	15

## Problem 2

Problem statement	16
2.1 Draw Stock Price Graph(Stock Price vs Time) for any 2 given stocks with inference	17
2.2 Calculate Returns for all stocks with inference	19
2.3 Calculate Stock Means and Standard Deviation for all stocks with inference	20
2.4 Draw a plot of Stock Means vs Standard Deviation and state your inference	22

---

2.5 Conclusion and Recommendations	23
<b>List of Figures</b>	
FIGURE 1	8
FIGURE 2	8
FIGURE 3	9
FIGURE 4	10
FIGURE 5	12
FIGURE 6	13
FIGURE 7	16
FIGURE 8	19
FIGURE 9	19
FIGURE 10	20
FIGURE 11	21
FIGURE 12	22
<b>List of Graphs</b>	
GRAPH 1	10
GRAPH 2	11
GRAPH 3	12
GRAPH 4	13
GRAPH 5	17
GRAPH 6	18
GRAPH 7	22

---

---

# Problem 1

## ***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.

## Data Description-

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).
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.
Current Liabilities and Provisions	Short-term financial obligations that are due within one year (includes amount that is set aside cover a future liability)
Total Assets/Liabilities	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 Production	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.
Revenue earnings in forex	Revenue earned in foreign currency
Revenue expenses in forex	Expenses due to foreign currency transactions
Capital expenses 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 Operating Activities	Use of cash from ongoing regular business activities
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 Financing Activities	Net flows of cash that are used to fund the company (transactions involving debt, equity, and dividends)
ROG-Net Worth (%)	Rate of Growth - Network
ROG-Capital Employed (%)	Rate of Growth - Capital Employed
ROG-Gross Block (%)	Rate of Growth - Gross Block
ROG-Gross Sales (%)	Rate of Growth - Gross Sales
ROG-Net Sales (%)	Rate of Growth - Net Sales
ROG-Cost of Production (%)	Rate of Growth - Cost of Production
ROG-Total Assets (%)	Rate of Growth - Total Assets
ROG-PBIDT (%)	Rate of Growth- PBIDT
ROG-PBDT (%)	Rate of Growth- PBDT

ROG-PBIT (%)	Rate of Growth- PBIT
ROG-PBT (%)	Rate of Growth- PBT
ROG-PAT (%)	Rate of Growth- PAT
ROG-CP (%)	Rate of Growth- CP
ROG-Revenue earnings in forex (%)	Rate of Growth - Revenue earnings in forex
ROG-Revenue expenses in forex (%)	Rate of Growth - Revenue expenses in forex
ROG-Market Capitalisation (%)	Rate of Growth - Market Capitalisation
Current 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
PBITM (%) [Latest]	Profit before Interest Depreciation and Tax Margin
PBITM (%) [Latest]	Profit Before Interest Tax Margin
PBDTM (%) [Latest]	Profit Before Depreciation Tax Margin
CPM (%) [Latest]	Cost per thousand (advertising cost)
APATM (%) [Latest]	After tax profit margin
Debtors Velocity (Days)	Average days required for receiving the payments
Creditors Velocity (Days)	Average number of days company takes to pay suppliers
Inventory Velocity (Days)	Average number of days the company needs to turn its inventory into sales
Value of Output/Total Assets	Ratio of Value of Output (market value) to Total Assets
Value of Output/Gross Block	Ratio of Value of Output (market value) 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]
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
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
14852	ABG Shipyards	-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
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
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

FIGURE 1

Debtors Velocity (Days)	Creditors Velocity (Days)	Inventory Velocity (Days)	Value of Output/Total Assets	Value of Output/Gross Block
0	0	45.0	0.00	0.00
29	101	2.0	0.31	0.24
97	558	0.0	-0.03	-0.26
93	63	2.0	0.24	1.90
3887	346	0.0	0.01	0.05

FIGURE 2

There are 67 variables, out of which 1 variable is categorical and 66 are continuous.. The data given is for 3586 individual companies. There are null values that require processing.



---

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

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

Random Forest:

Random forest is a supervised learning algorithm which is used for both classification as well as regression. But it is mainly used for classification problems. As we know that a forest is made up of trees and more trees means more robust forest. Similarly, a random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result. <sup>[1]</sup>

**Train set**

The below is the classification report of the training data-

	precision	recall	f1-score	support
0.0	0.99	0.99	0.99	2143
1.0	0.93	0.89	0.91	259
accuracy			0.98	2402
macro avg	0.96	0.94	0.95	2402
weighted avg	0.98	0.98	0.98	2402

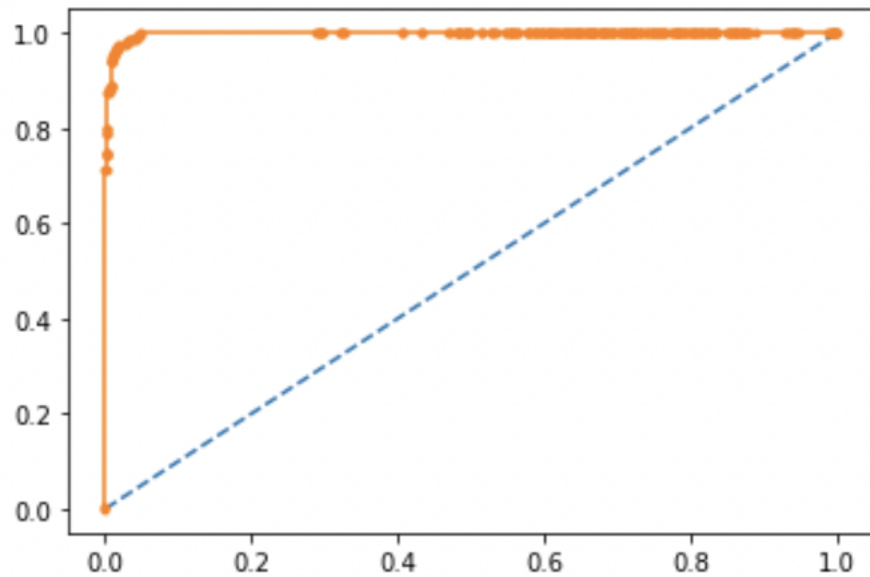
**FIGURE 3**

The training data accuracy is 98%.

---

ROC curve and AUC score for training data-

AUC: 0.998



**GRAPH 1**

### Test set

The below is the classification report of the testing data-

	precision	recall	f1-score	support
0.0	0.98	0.99	0.99	1056
1.0	0.90	0.87	0.88	128
accuracy			0.98	1184
macro avg	0.94	0.93	0.94	1184
weighted avg	0.98	0.98	0.98	1184

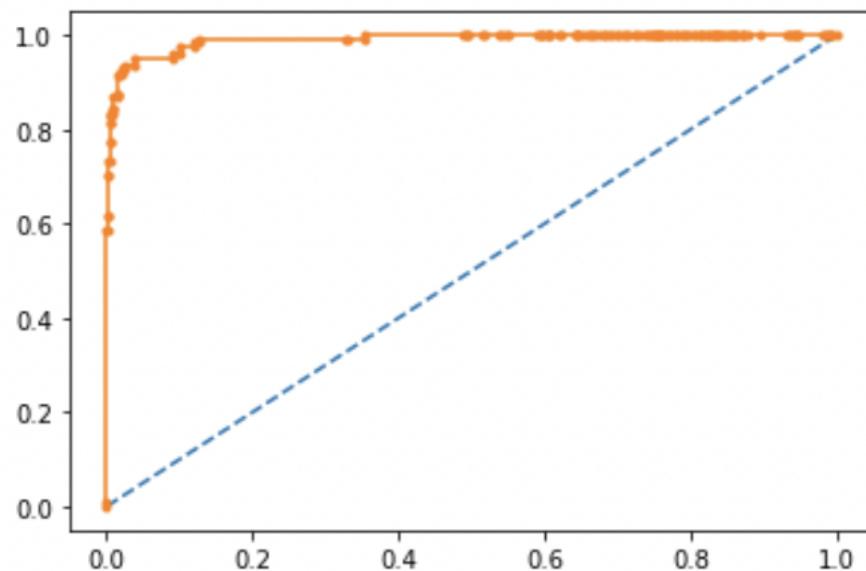
**FIGURE 4**

---

The testing data accuracy is 98%.

ROC curve and AUC score for testing data-

AUC: 0.990



**GRAPH 2**

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

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

### LINEAR DISCRIMINANT ANALYSIS:

Linear Discriminant Analysis (LDA) is a dimensionality reduction technique. As the name implies dimensionality reduction techniques reduce the number of dimensions (i.e. variables) in a dataset while retaining as much information as possible. For instance, suppose that we plotted the relationship between two variables where each color represents a different class. <sup>[2]</sup>

---

## Train set

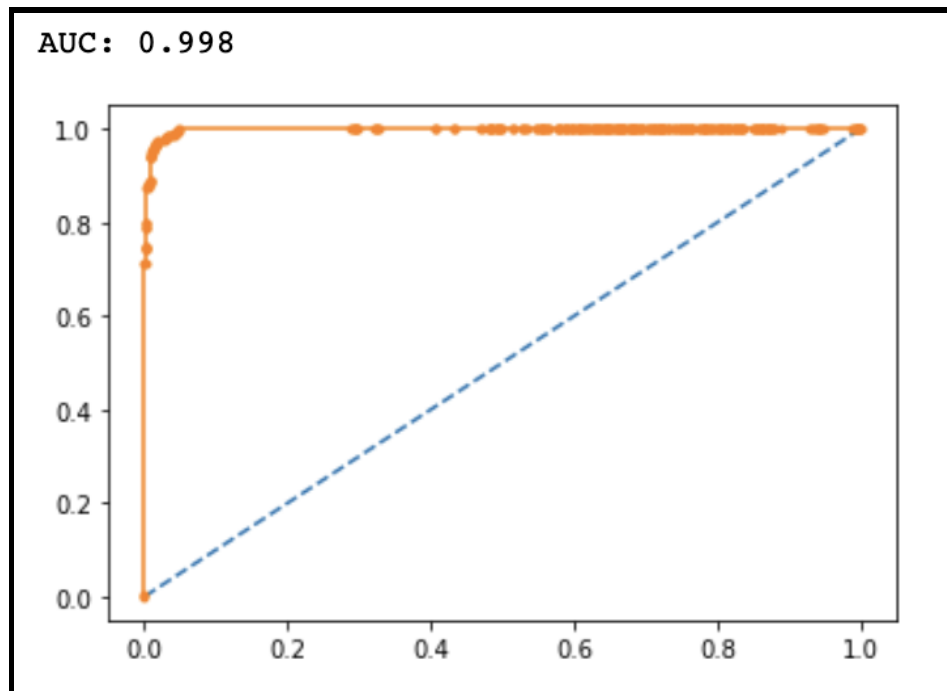
The below is the classification report of the training data-

	precision	recall	f1-score	support
0.0	0.93	0.98	0.95	2143
1.0	0.71	0.37	0.49	259
accuracy			0.92	2402
macro avg	0.82	0.68	0.72	2402
weighted avg	0.90	0.92	0.90	2402

**FIGURE 5**

The training data accuracy is 92%.

ROC curve and AUC score for training data-



**GRAPH 3**

---

## Test set

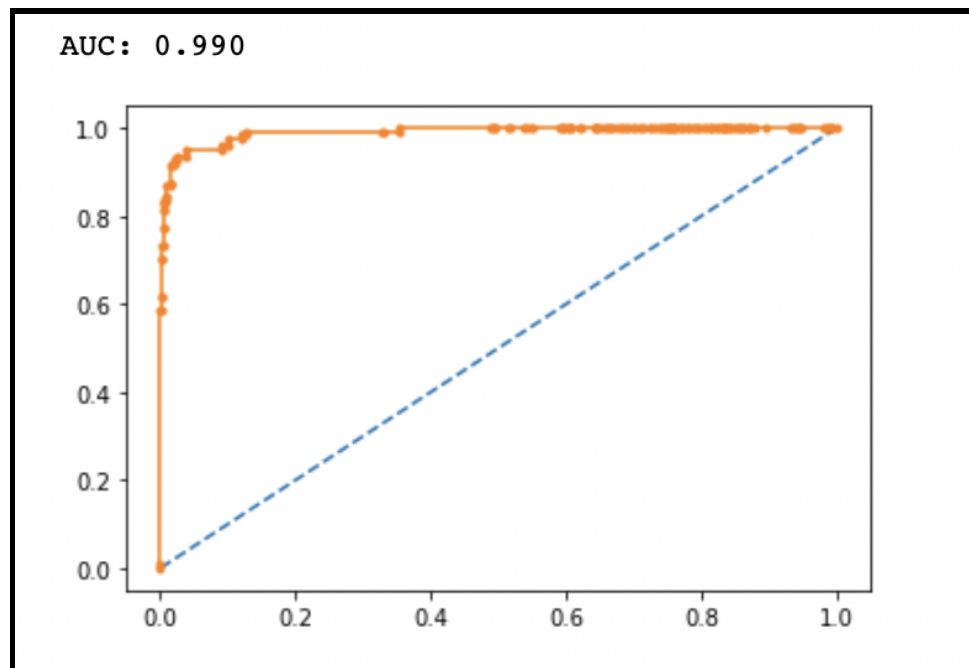
The below is the classification report of the testing data-

	precision	recall	f1-score	support
0.0	0.93	0.98	0.95	1056
1.0	0.70	0.39	0.50	128
accuracy			0.92	1184
macro avg	0.82	0.69	0.73	1184
weighted avg	0.91	0.92	0.91	1184

**FIGURE 6**

The testing data accuracy is 92%.

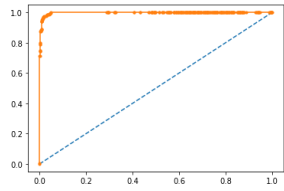
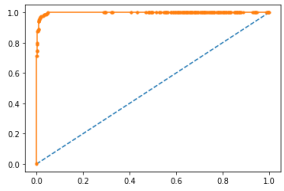
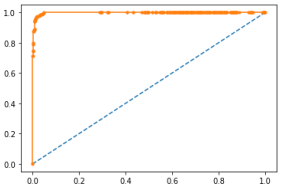
ROC curve and AUC score for testing data-



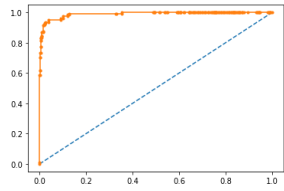
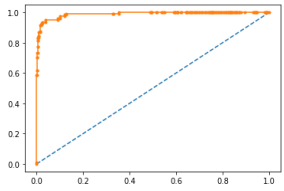
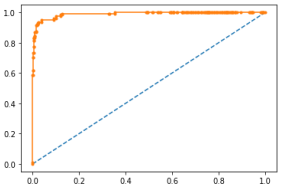
**GRAPH 4**

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

Training data-

	Logistics	Random Forest	LDA
Precision	0.178	0.93	0.71
Recall	0.919	0.89	0.37
F1 score	0.298	0.91	0.49
Accuracy	53.4%	98%	92%
ROC curve			

Testing data-

	Logistics	Random Forest	LDA
Precision	0.176	0.90	0.70
Recall	0.945	0.87	0.39
F1 score	0.297	0.88	0.50
Accuracy	51.6%	98%	92%
ROC curve			

---

### Inference:

Based on the table of comparison displayed above, it is clear that the Random Forest model has the best accuracy in both training and testing data. The F1 score is high which means both type 1 and type 2 errors are reduced. Also, it has the highest Precision value which means type 1 error is low in this.

The Logistics model has the highest recall score in both training and testing which means type 2 error is low in it.

### **1.13 State Recommendations from the above models**

- Amongst all the models we tried in this case Random Forest seems to be best aligned to our objective as it has the highest accuracy.
- On the business standpoint, using this model we can find the potential defaulters and we can send out early reminders to the companies to avoid any loss.
- These companies can also be given a higher interest rate to tally the losses.
- At the same time, the companies who are coming under the non-defaulters can be given more benefits in terms of loan approvals.

---

## Problem 2

### ***Problem statement-***

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. You are expected to do the Market Risk Analysis using Python.

### ***Sample of the dataset-***

Market+Risk+Dataset										
Date	Infosys	Indian Hotel	Mahindra & Mahindra	Axis Bank	SAIL	Shree Cement	Sun Pharma	Jindal Steel	Idea Vodafone	Jet Airways
31-03-2014	264	69	455	263	68	5543	555	298	83	278
07-04-2014	257	68	458	276	70	5728	610	279	84	303
14-04-2014	254	68	454	270	68	5649	607	279	83	280
21-04-2014	253	68	488	283	68	5692	604	274	83	282
28-04-2014	256	65	482	282	63	5582	611	238	79	243
05-05-2014	248	64	502	300	67	5646	606	244	80	239
12-05-2014	253	68	508	324	75	6041	596	269	84	244
19-05-2014	244	83	528	344	86	6394	567	299	82	289
26-05-2014	234	82	563	339	84	6709	593	296	81	241

**FIGURE 7**

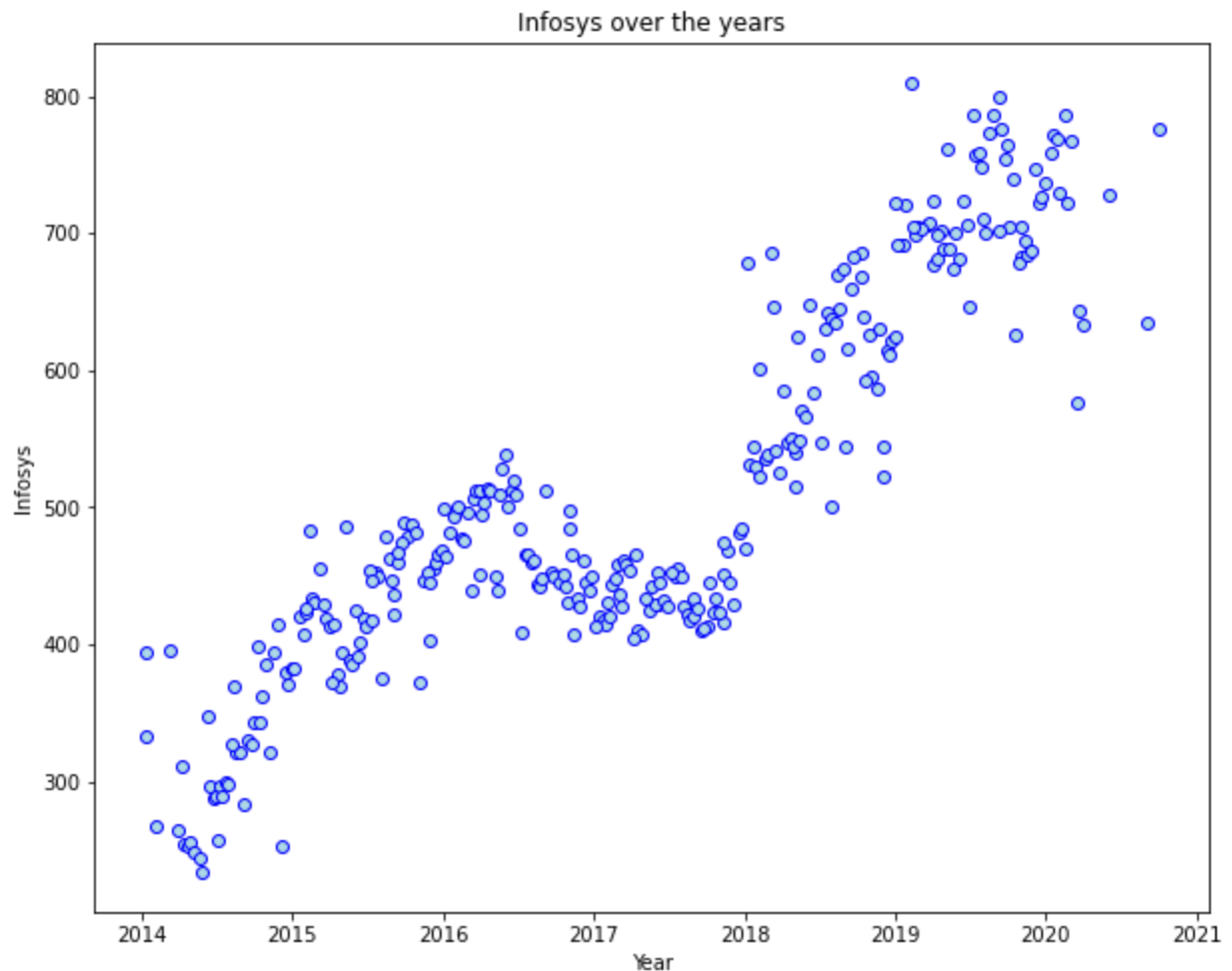
There are 11 variables and all of them are numeric. The data given is for 10 individual companies. There are no null values.



---

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

The below graph is a scatter plot between Stock Price and Time for Infosys -



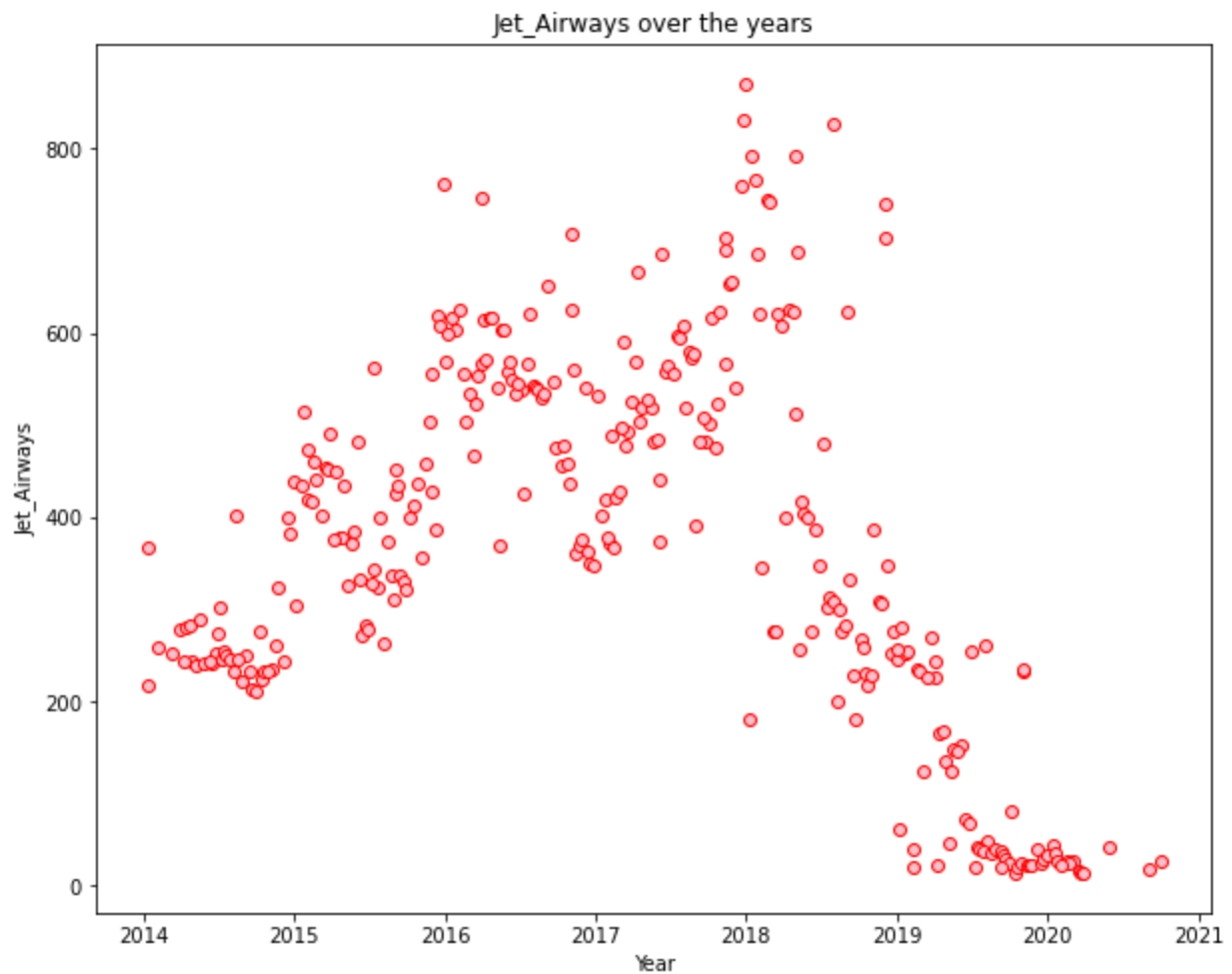
**GRAPH 5**

### Inference:

The trend of the price seems to have a positive direction over the years (i.e) there is a gradual increase in the stock price from 2014 through 2021.

---

The below graph is a scatter plot between Stock Price and Time for Jet Airways -



**GRAPH 6**

**Inference:**

The trend of the price seems to have a positive direction till the year 2018 and then plummeted after 2018 (i.e) there is a gradual increase in the stock price till 2018 and then falls after 2018.

## 2.2 Calculate Returns for all stocks with inference

A return, also known as a financial return, in its simplest terms, is the money made or lost on an investment over some period of time.

A return can be expressed nominally as the change in dollar value of an investment over time. A return can also be expressed as a percentage derived from the ratio of profit to investment. Returns can also be presented as net results (after fees, taxes, and inflation) or gross returns that do not account for anything but the price change. <sup>[3]</sup>

The steps for calculating returns from prices:

- (i) Take logarithms
- (ii) Take differences

The below image shows the sample data of the returns calculated -

	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

FIGURE 8

The below image shows the five point summary of the data for the returns-

	Infosys	Indian_Hotel	Mahindra_&_Mahindra	Axis_Bank	SAIL	Shree_Cement	Sun_Pharma	Jindal_Steel	Idea_Vodafone	Jet_Airways
count	313.000000	313.000000	313.000000	313.000000	313.000000	313.000000	313.000000	313.000000	313.000000	313.000000
mean	0.002794	0.000266	-0.001506	0.001167	-0.003463	0.003681	-0.001455	-0.004123	-0.010608	-0.009548
std	0.035070	0.047131	0.040169	0.045828	0.062188	0.039917	0.045033	0.075108	0.104315	0.097972
min	-0.167300	-0.236389	-0.285343	-0.284757	-0.251314	-0.129215	-0.179855	-0.283768	-0.693147	-0.458575
25%	-0.014514	-0.023530	-0.020884	-0.022473	-0.040822	-0.019546	-0.020699	-0.049700	-0.045120	-0.052644
50%	0.004376	0.000000	0.001526	0.001614	0.000000	0.003173	0.001530	0.000000	0.000000	-0.005780
75%	0.024553	0.027909	0.019894	0.028522	0.032790	0.029873	0.023257	0.037179	0.024391	0.036368
max	0.135666	0.199333	0.089407	0.127461	0.309005	0.152329	0.166604	0.243978	0.693147	0.300249

FIGURE 9

---

It is clear from the description that all the companies have hit a negative value. The company Idea\_Vodafone seems to have the maximum returns. But we cannot come to a conclusion just with the five point summary. The averages of all the companies vary between positive and negative values.

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

Stock Means: Average returns that the stock is making on a week to week basis

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

**FIGURE 10**

The company Shree Cement has the highest Stock Mean and the company Sun Pharma has the lowest Stock Mean with a negative value.

---

Stock Standard Deviation : It is a measure of volatility meaning the more a stock's returns vary from the stock's average return, the more volatile the stock

Infosys	0.035070
Indian_Hotel	0.047131
Mahindra_ & _Mahindra	0.040169
Axis_Bank	0.045828
SAIL	0.062188
Shree_Cement	0.039917
Sun_Pharma	0.045033
Jindal_Steel	0.075108
Idea_Vodafone	0.104315
Jet_Airways	0.097972
dtype:	float64

**FIGURE 11**

The company Idea Vodafone has the highest Stock Standard Deviation and the company Infosys has the lowest Stock Standard Deviation.

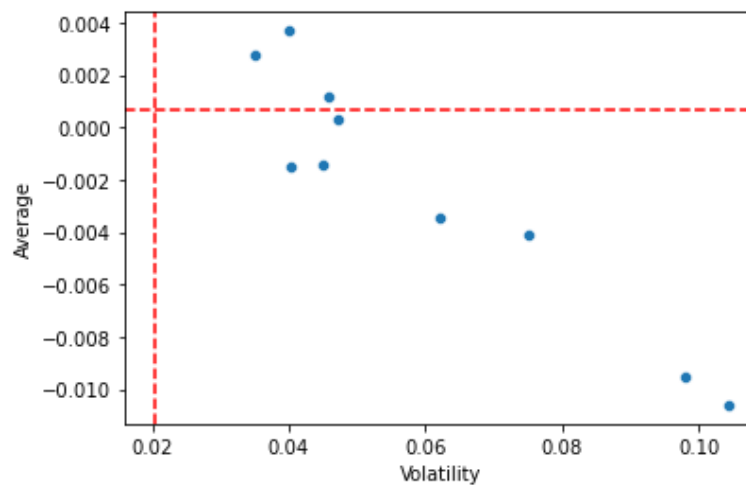
---

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

The Stock Means is renamed as Average and Stock Standard Deviation is renamed as Volatility.

	Average	Volatility
<b>Infosys</b>	0.002794	0.035070
<b>Indian_Hotel</b>	0.000266	0.047131
<b>Mahindra_&amp;_Mahindra</b>	-0.001506	0.040169
<b>Axis_Bank</b>	0.001167	0.045828
<b>SAIL</b>	-0.003463	0.062188
<b>Shree_Cement</b>	0.003681	0.039917
<b>Sun_Pharma</b>	-0.001455	0.045033
<b>Jindal_Steel</b>	-0.004123	0.075108
<b>Idea_Vodafone</b>	-0.010608	0.104315
<b>Jet_Airways</b>	-0.009548	0.097972

**FIGURE 12**



**GRAPH 7**

From the graph, it is clear that most of the stocks have higher volatility or higher risk. There are only three companies with high returns and low risk.

---

## 2.5 Conclusion and Recommendations

- Stock with a lower mean and higher standard deviation do not play a good role in the portfolio that has competing stock with more returns and less risk. Thus for the data we have here, we are left with only a few stocks with such a case
- Ones with higher return for a comparative or lower risk are considered better
- There are only three companies with high returns and low risk in the given data set.
- It is advised to invest more on companies with high returns and low risk
- The preferred companies would be Shree Cement, Infosys and Axis Bank as these have high returns and low risk compared to other companies.

---

## **Websites-**

- [1] [https://www.tutorialspoint.com/machine\\_learning\\_with\\_python/machine\\_learning\\_with\\_python\\_classification\\_algorithms\\_random\\_forest.htm](https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_random_forest.htm)
- [2] <https://towardsdatascience.com/linear-discriminant-analysis-in-python-76b8b17817c2>
- [3] <https://www.investopedia.com/terms/r/return.asp>

# End of Project