**FRA PROJECT (MILESTONE-2)**

**BUSINESS REPORT**

**SULOCHANA**

Table of contents:

* 1. **Build a Random Forest Model on Train Dataset. Also showcase your model building approach……………………………………………............ 8**
  2. **Validate the Random Forest Model on test Dataset and state the performance matrices. Also state interpretation from the model…………………………………………… 10**
  3. **Build a LDA Model on Train Dataset. Also showcase your model building approach………………………… 12**
  4. **Validate the LDA Model on test Dataset and state the performance matrices. Also state interpretation from the model………………………………………………………. 14**
  5. **Compare the performances of Logistics, Radom Forest and LDA models (include ROC Curve)………………….. 16**
  6. **State Recommendations from the above models…… 17**
  7. **Draw Stock Price Graph (Stock Price vs Time) for any 2 given stocks with inference……………………………… 20**
  8. **Calculate Returns for all stocks with inference………. 21**
  9. **Calculate Stock Means and Standard Deviation for all stocks with inference………………………………………22**
  10. **Draw a plot of Stock Means vs Standard Deviation and state your inference………………………………………… 23**
  11. **Conclusion and Recommendations………………………24**

**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.

Explanation of data fields available in Data Dictionary, 'Credit Default Data Dictionary.xlsx'

Dependent variable - We need to create a default variable that should take the value of 1 when net worth next year is negative & 0 when net worth next year is positive.

**DATA DICTIONARY:**

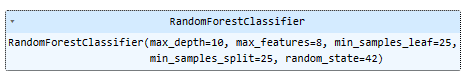
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **Field Name** | **Description** | **New Field Name** |  |
| 1 | Co\_Code | Company Code | Co\_Code |  |
| 2 | Co\_Name | Company Name | Co\_Name |  |
| 3 | Networth Next Year | Value of a company as on 2016 - Next Year(difference between the value of total assets and total liabilities) | Networth\_Next\_Year |  |
| 4 | Equity Paid Up | Amount that has been received by the company through the issue of shares to the shareholders | Equity\_Paid\_Up |  |
| 5 | Networth | Value of a company as on 2015 - Current Year | Networth |  |
| 6 | Capital Employed | Total amount of capital used for the acquisition of profits by a company | Capital\_Employed |  |
| 7 | Total Debt | The sum of money borrowed by the company and is due to be paid | Total\_Debt |  |
| 8 | Gross Block | Total value of all of the assets that a company owns | Gross\_Block |  |
| 9 | Net Working Capital | The difference between a company's current assets (cash, accounts receivable, inventories of raw materials and finished goods) and its current liabilities (accounts payable). | Net\_Working\_Capital |  |
| 10 | Current Assets | All the assets of a company that are expected to be sold or used as a result of standard business operations over the next year. | Curr\_Assets |  |
| 11 | Current Liabilities and Provisions | Short-term financial obligations that are due within one year (includes amount that is set aside cover a future liability) | Curr\_Liab\_and\_Prov |  |
| 12 | Total Assets/Liabilities | Ratio of total assets to liabailities of the company | Total\_Assets\_to\_Liab |  |
| 13 | Gross Sales | The grand total of sale transactions within the accounting period | Gross\_Sales |  |
| 14 | Net Sales | Gross sales minus returns, allowances, and discounts | Net\_Sales |  |
| 15 | Other Income | Income realized from non-business activities (e.g. sale of long term asset) | Other\_Income |  |
| 16 | Value Of Output | Product of physical output of goods and services produced by company and its market price | Value\_Of\_Output |  |
| 17 | Cost of Production | Costs incurred by a business from manufacturing a product or providing a service | Cost\_of\_Prod |  |
| 18 | Selling Cost | Costs which are made to create the demand for the product (advertising expenditures, packaging and styling, salaries, commissions and travelling expenses of sales personnel, and the cost of shops and showrooms) | Selling\_Cost |  |
| 19 | PBIDT | Profit Before Interest, Depreciation & Taxes | PBIDT |  |
| 20 | PBDT | Profit Before Depreciation and Tax | PBDT |  |
| 21 | PBIT | Profit before interest and taxes | PBIT |  |
| 22 | PBT | Profit before tax | PBT |  |
| 23 | PAT | Profit After Tax | PAT |  |
| 24 | Adjusted PAT | Adjusted profit is the best estimate of the true profit | Adjusted\_PAT |  |
| 26 | CP | Commercial paper , a short-term debt instrument to meet short-term liabilities. | CP |  |
| 27 | Revenue earnings in forex | Revenue earned in foreign currency | Rev\_earn\_in\_forex |  |
| 28 | Revenue expenses in forex | Expenses due to foreign currency transactions | Rev\_exp\_in\_forex |  |
| 29 | Capital expenses in forex | Long term investment in forex | Capital\_exp\_in\_forex |  |
| 30 | Book Value (Unit Curr) | Net asset value | Book\_Value\_Unit\_Curr |  |
| 31 | Book Value (Adj.) (Unit Curr) | Book value adjusted to reflect asset's true fair market value | Book\_Value\_Adj\_Unit\_Curr |  |
| 32 | Market Capitalisation | Product of the total number of a company's outstanding shares and the current market price of one share | Market\_Capitalisation |  |
| 33 | CEPS (annualised) (Unit Curr) | Cash Earnings per Share, profitability ratio that measures the financial performance of a company by calculating cash flows on a per share basis | CEPS\_annualised\_Unit\_Curr |  |
| 34 | Cash Flow From Operating Activities | Use of cash from ongoing regular business activities | Cash\_Flow\_From\_Opr |  |
| 35 | Cash Flow From Investing Activities | Cash used in the purchase of non-current assets–or long-term assets– that will deliver value in the future | Cash\_Flow\_From\_Inv |  |
| 36 | Cash Flow From Financing Activities | Net flows of cash that are used to fund the company (transactions involving debt, equity, and dividends) | Cash\_Flow\_From\_Fin |  |
| 37 | ROG-Net Worth (%) | Rate of Growth - Networth | ROG\_Net\_Worth\_perc |  |
| 38 | ROG-Capital Employed (%) | Rate of Growth - Capital Employed | ROG\_Capital\_Employed\_perc |  |
| 39 | ROG-Gross Block (%) | Rate of Growth - Gross Block | ROG\_Gross\_Block\_perc |  |
| 40 | ROG-Gross Sales (%) | Rate of Growth - Gross Sales | ROG\_Gross\_Sales\_perc |  |
| 41 | ROG-Net Sales (%) | Rate of Growth - Net Sales | ROG\_Net\_Sales\_perc |  |
| 42 | ROG-Cost of Production (%) | Rate of Growth - Cost of Production | ROG\_Cost\_of\_Prod\_perc |  |
| 43 | ROG-Total Assets (%) | Rate of Growth - Total Assets | ROG\_Total\_Assets\_perc |  |
| 44 | ROG-PBIDT (%) | Rate of Growth- PBIDT | ROG\_PBIDT\_perc |  |
| 45 | ROG-PBDT (%) | Rate of Growth- PBDT | ROG\_PBDT\_perc |  |
| 46 | ROG-PBIT (%) | Rate of Growth- PBIT | ROG\_PBIT\_perc |  |
| 47 | ROG-PBT (%) | Rate of Growth- PBT | ROG\_PBT\_perc |  |
| 48 | ROG-PAT (%) | Rate of Growth- PAT | ROG\_PAT\_perc |  |
| 49 | ROG-CP (%) | Rate of Growth- CP | ROG\_CP\_perc |  |
| 50 | ROG-Revenue earnings in forex (%) | Rate of Growth - Revenue earnings in forex | ROG\_Rev\_earn\_in\_forex\_perc |  |
| 51 | ROG-Revenue expenses in forex (%) | Rate of Growth - Revenue expenses in forex | ROG\_Rev\_exp\_in\_forex\_perc |  |
| 52 | ROG-Market Capitalisation (%) | Rate of Growth - Market Capitalisation | ROG\_Market\_Capitalisation\_perc |  |
| 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 |  |

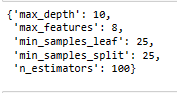
**1.8 Build a Random Forest Model on Train Dataset. Also showcase your model building approach**

Here we build random forest model based on train and test data split of 67:33.

Scaling is not required for random forest model because it is tree based approach where distance matrix is not required. However I have performed this model on scaling data. Data normalization won’t affect the output for random forest model.

We build model on train data after that we validate with test data.

****

****

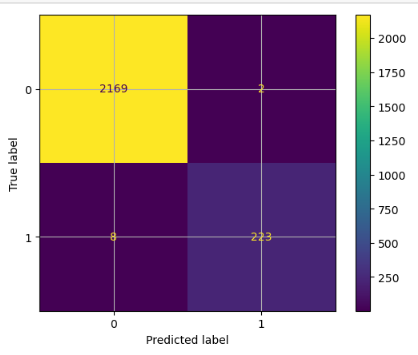
Initially, we performed grid search on data and indentified best parameters. We have used the above parameters as grid parameters.

**Train data**

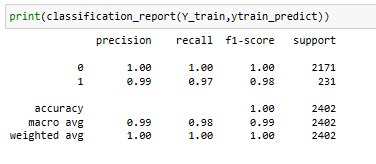
**RF Model evaluation on Train Data**

## Confusion Matrix on train data set:

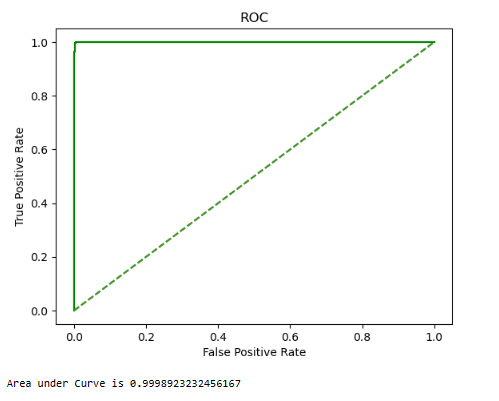
****

****

**Classification report on train data:**

****

## ROC Curve

****

Using the random forest model on train dataset, we obtain

Accuracy = 1.00

Recall = 0.97

Precision = 0.99

F1 score = 0.98

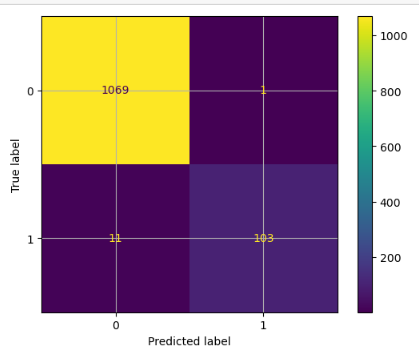
Auc score = 0.99

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

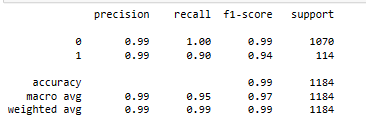
### RF Model evaluation on Test Data

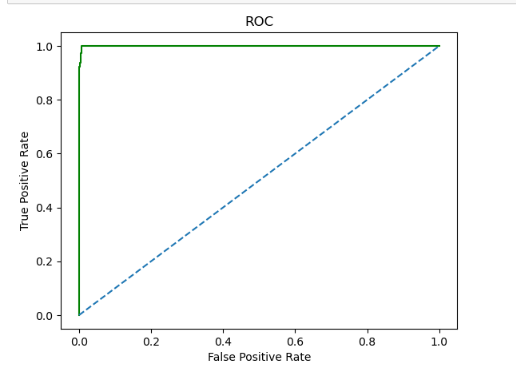
**Confusion Matrix on test data set:**





**Classification report on test data:**

****

****

Using the random forest model on test dataset, we obtain

Accuracy = 0.99

Recall = 0.90

Precision = 0.99

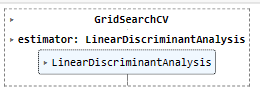
F1 score = 0.94

Auc score = 0.99

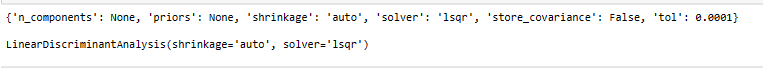
**1.10 Build a LDA Model on Train Dataset. Also showcase your model building approach**

## Linear Discriminant Analysis

**Build LDA Model & fit on training data:**

****

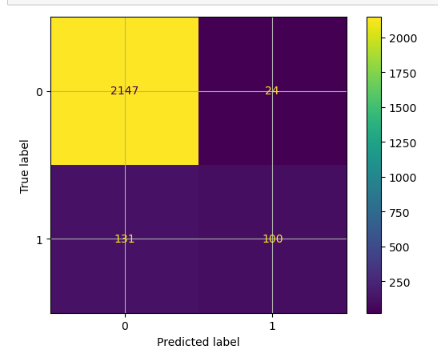
**Best parameters:**

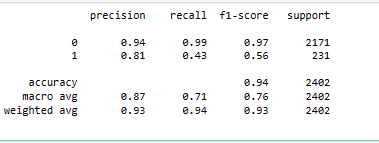
****

## Model Evaluation

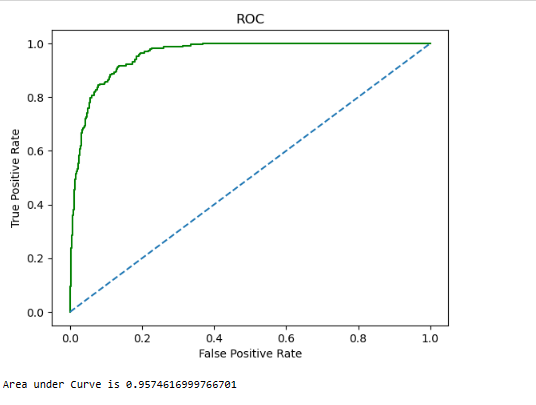
## Training Data

## Confusion Matrix & Classification Report Metrics

****

****

## ROC Curve

****

Using the LDA model on train dataset, we obtain

Accuracy = 0.93

Recall = 0.43

Precision = 0.80

F1 score = 0.56

Auc score = 0.95

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

## Test Data

## Confusion Matrix & Classification Report Metrics

## 

## 

## ROC Curve

## 

Using the LDA model on test dataset, we obtain

Accuracy = 0.92

Recall = 0.39

Precision = 0.73

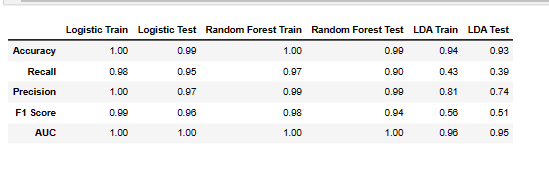
F1 score = 0.51

Auc score = 0.94

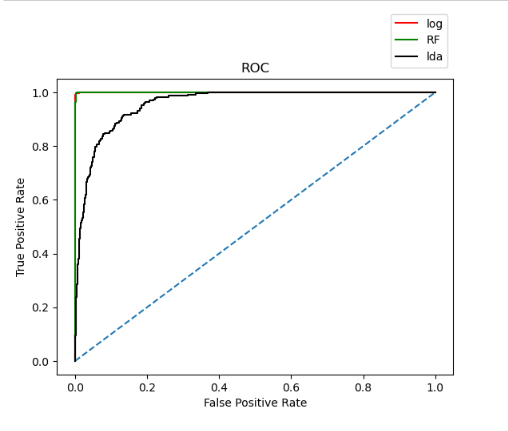
**1.12 Compare the performances of Logistics, Radom Forest and LDA models (include ROC Curve)**

Due to imbalanced data we are considering the recall value along with the precision value for our performance metrics.

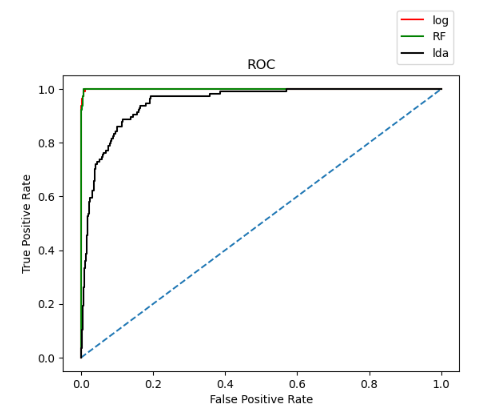
**Comparison of the performance metrics from the 3 models:**

****

# ROC Curve for the 3 models on the Training data



# ROC Curve for the 3 models on the Test data



From above comparison we see that logistic regression and random forest model performs well on this data with high recall and precision when compared to LDA models. We can say LDA performed worst on this data.

**1.13 State Recommendations from the above models**

**Below are recommendations from the above models:**

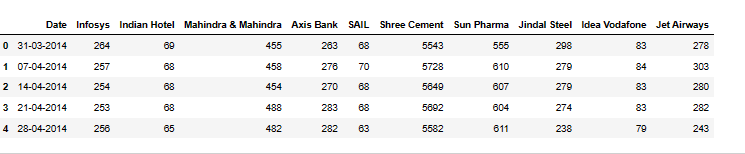
* **We are choosing the logistic and random forest models are our optimum models as we have higher performance metrics on all values (accuracy, recall, precision and f1 score).**
* **We set out to identify the potential customers in the bank who are predicted to default and our model should give us high recall value.**
* **Random model also gives high AUC value, which will give better results in identifying our potential defaulter and it is important for us as a bank to identify our defaulters.**

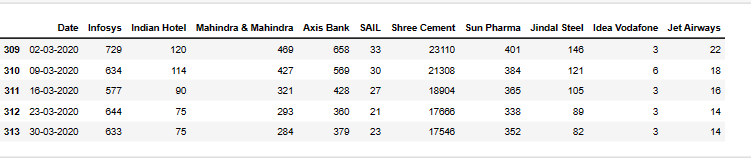
# Market Risk

### 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.

We have imported the dataset and we can see the stock values of various companies across the years 2014 and 2021.

## Glimpse of Data:

****

****

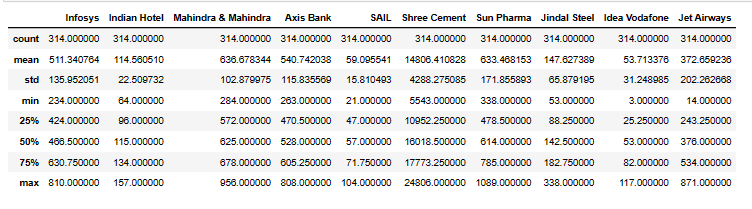
## Fixing messy column names (containing spaces) for ease of use

## 

## Data info:

## 

**Data description:**

****

### First, let us check the number of rows (observations) and the number of columns (variables)

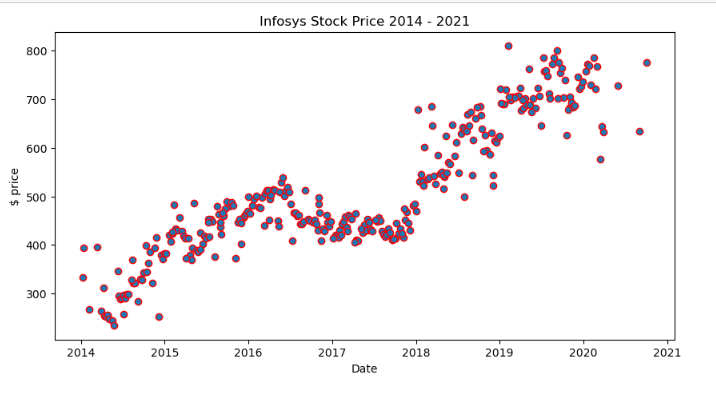


**2.1 Draw Stock Price Graph (Stock Price vs. Time) for any 2 given stocks with inference**

**Below is the stock price graph for two companies:**

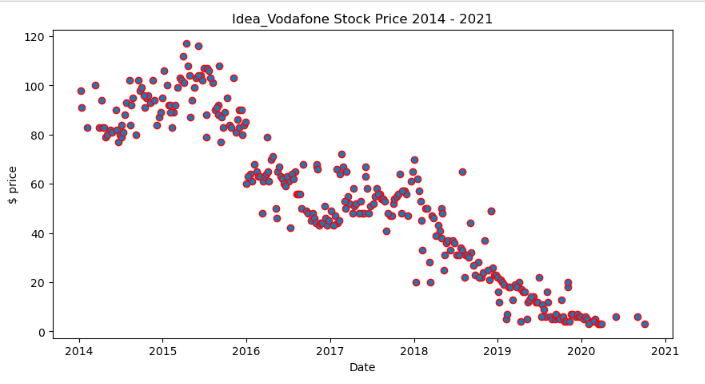
**Infosys:**

**For Infosys Company we can observe the stock price seems to be on up rise from the years 2014 to 2021.we can see a slight dip in stock prices in the year 2017 to 2018. From 2019 onwards there is steep increase in the stock price.**

****

**Idea-vodafone:**

**From below plot, stock price is decreasing over the years and we cloud see that the prices are inversely proportional to the year. We find the company has higher risk in terms of stock market investments.**

****

**2.2 Calculate Returns for all stocks with inference**

#### Returns for the all stocks i.e. difference of log of price at t and the log of price at t-1 are shown below.

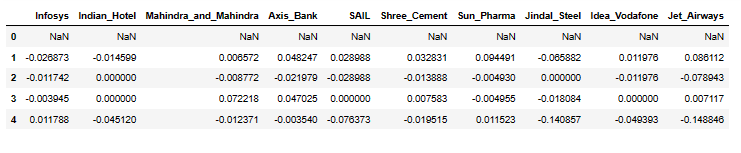
We calculate the returns by taking logarithmic value difference by subtracting the stock value from any day to its previous day value.

The first row will return a null value, as we don’t have any previous day stock price value.

**stock\_returns shape**

(314, 10)

**stock\_returns Head**

****

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

**Below are the stock mean and standard deviation values across the rows.**

## Calculating stock means

## 

## Calculating stock standard Deviation

## 

## Lesser the standard deviation we can say lesser in the risk in investing in that particular company and higher the mean value means that the stock price is on the higher side.

**Companies those have higher mean:**

**Shree cement –** 0.003681

**Infosys -** 0.002794

**Companies those have lesser standard deviation:**

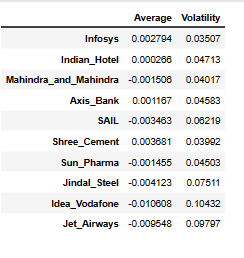
**Infosys -** 0.03507

**Shree cement –** 0.03992

**Mahindra & Mahindra -** 0.04017

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

### Creating dataframe for its mean value as average and the standard deviation as its volatility

****

## Let us plot & see what they say about stock prices

## 

**We can see that the companies Infosys and shree cement has higher stock price and less volatility.**

**2.5 Conclusion and Recommendations**

**Below are the conclusions and recommendations on the market risk analysis:**

**We can recommend to investors to invest in the companies Infosys and Shree Company, as we could see the company’s share price has been steadily increasing over the years and also we could see the mean value being the highest with lesser risk value.**

**We also encourage investing in other companies Indian hotel and axis bank with less investment shares. They may not the best in their stock prices, but they have better risk values than other companies**

**END**