****

|  |  |
| --- | --- |
| Credit Risk  and  Market Risk  FRA Project -2 | Debsmita Chakraborty  Batch-July ‘C’ |

Table of Contents

CREDIT ANALYSIS

Problem Statement………………………………………………………………………………………..……Pg-2

Summary……………………………………………………………………………………………………………Pg-2-3

1.8 Build a Random Forest Model on Train Dataset. Also showcase your model building approach……………………………………………………………………………………………..……………..Pg-4-5

1.9 Validate the Random Forest Model on test Dataset and state the performance matrices. Also state interpretation from the model……………………………………………………………………..…………..Pg-6-7

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

model building approach………………………………………………………………………………………….Pg-7

1.11 Validate the LDA Model on test Dataset and state the performance matrices. Also state interpretation from the model……………………………………………………………………………………Pg-8

1.12 Compare the performances of Logistics, Radom Forest and LDA models (include ROC Curve) ………………………………………………………………………………………………………………………..Pg-9-9

1.13 State Recommendations from the above models…………………………………………..………..Pg-9-10

MARKET ANALYSIS

* 1. Introduction…………………………………………………………………………………..…………….Pg-11
  2. Data Summary……………………………………………………………………………………………..Pg-11-12

2.1) Draw Stock Price Graph (Stock Price vs Time) for any 2 given stocks with inference…………………………………………………………………………………………………..……..Pg-12-13

2.2) Calculate Returns for all stocks with inference……………………………...…………..………….Pg-13

2.3) Calculate Stock Means and Standard Deviation for all stocks with inference………………………………………………………………………………………………………..…..Pg-14

2.4) Draw a plot of Stock Means vs Standard Deviation

and state your inference……………………………………………………………………………………………Pg-15

Figures and Plot

Data Figures [Fig.1(a-c)]…………………………………………………………………………..Pg-2-3

Random Forest [Fig.2-4]……………………………………………………………………………Pg-5-6

LDA Model [Fig.5-6]…………………………………………………………………………………Pg-6-8

Model Comparisons [Fig.7a-7b]……………………………………………………………………Pg-8-10

MARKET RISK

Data Figures [Fig.1-3]………………………………………………………………………………Pg-11-12

Stock Prices[Fig.4.1-4.2]……………………………………………………………….………….Pg-13

Returns [Fig.5.1-5.3]…………………………………………………………………………………Pg13-Pg-14

Standard Deviation and Mean [Fig.6]……………………………………………………………..Pg-15

Problem Statement:

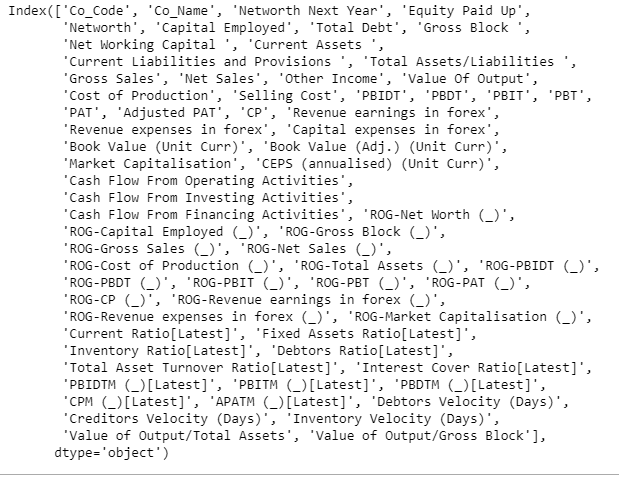
Businesses or companies can fall prey to default if they are not able to keep up their debt obligations. Defaults will lead to a lower credit rating for the company which in turn reduces its chances of getting credit in the future and may have to pay higher interests on existing debts as well as any 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 Net worth of the company in the following year (2016) is provided which can be used to drive the labelled field.

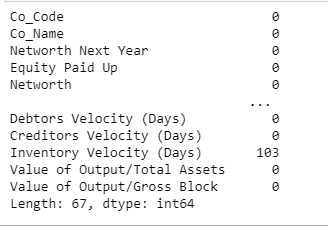
Summary:

* The data set contains 3586 rows and 67 columns with the headings:



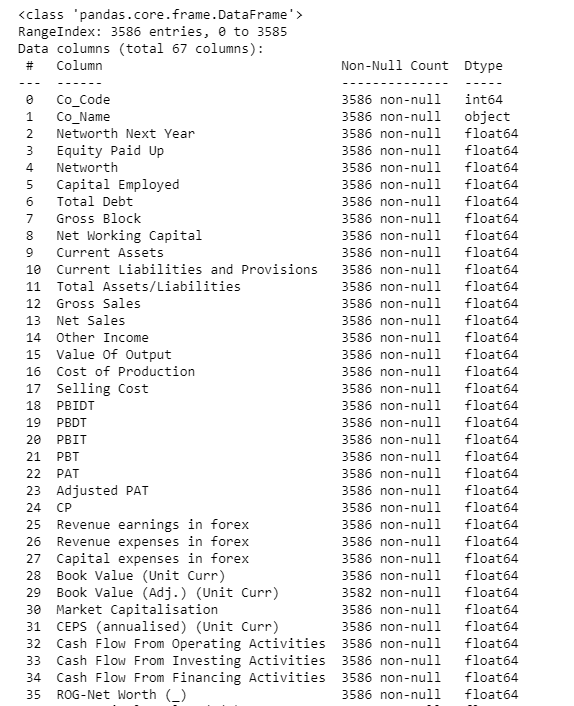
1.a Data Headings

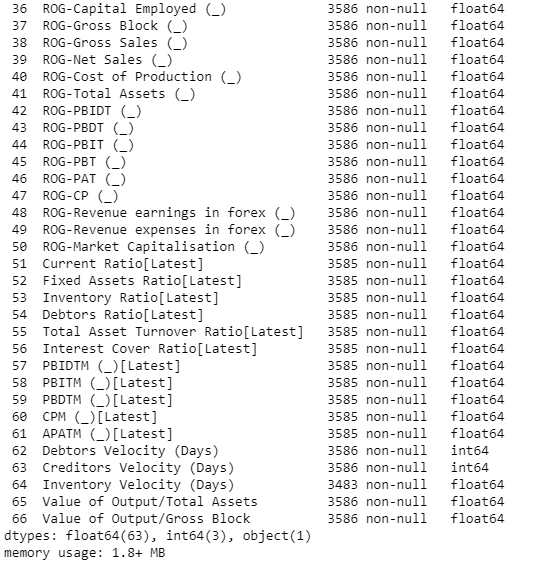
* The data has 545 duplicated values but a lot of null values.



1.b Data Null values

* The data has the below data information:





1.c Data Information

1.b Data Null values

* The data has 63 float values, 1 object or categorical value and 3 integer values
* The data uses 1.8+ MB of memory.

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

Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression. The algorithm operates by constructing a multitude of decision trees at training time and outputting the mean/mode of prediction of the individual trees.

Results of Train Data Set:



Fig2-Best Grid Score

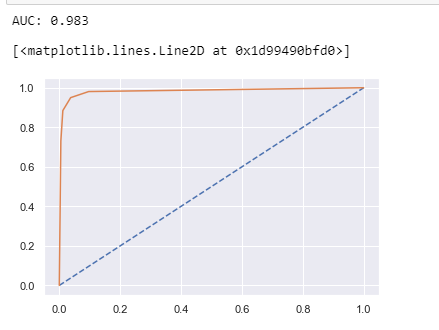
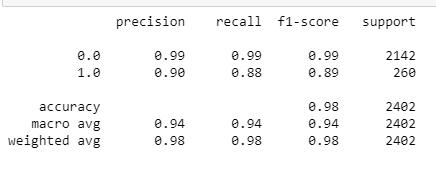
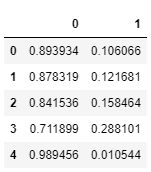
 

Fig2-Classification Matrix

Fig2-AUC-ROC Curve for Train data

Inference:

* As seen in the image we have AUC is 0.983. Since the accuracy is high, we can say the below matrix of 0 was maximum predicted correctly as 0 and 1 as 1:



* F1 is 0.89, Recall is 0.88 and precision is 0.90.
* F1 can be seen at 0.89 which is a good score as higher F1 better a model. Precision and recall has a good score to hence train data has good scores overall.
* The accuracy score is 98% which makes the model a good model.

We also did Decision Tree:

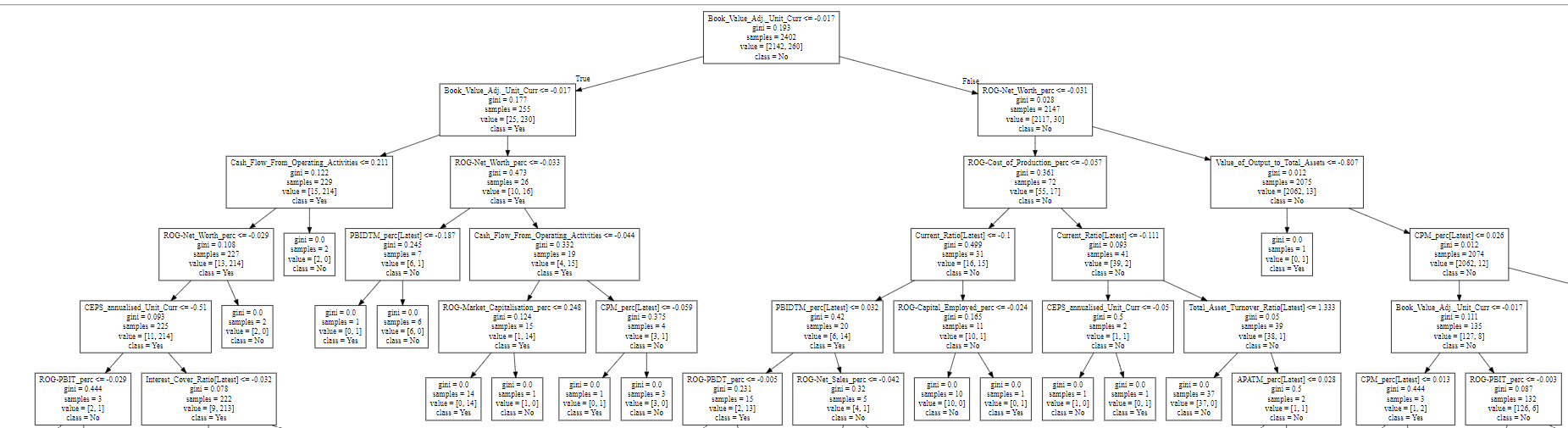
 

Fig3-Decision Tree

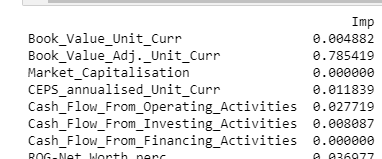
Inference:

* The below is the details of the decision tree:



Since the split is 30 could only show top 5 and last 5 splits.

* The top 5 features which came as important is:



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

Results of Test Data Set:

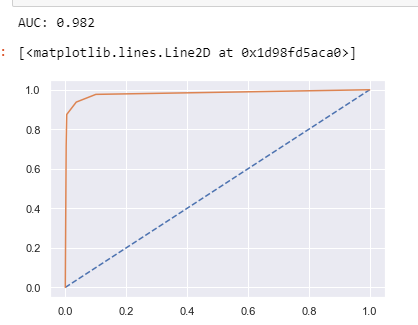
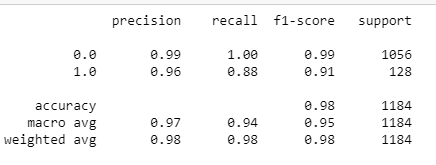
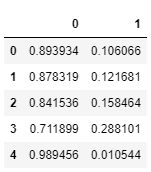


Fig4-AUC ROC Curve

Fig4-Classification Model

Inference:

* As seen in the image we have AUC is 0.982. Since the accuracy is high, we can say the below matrix of 0 was maximum predicted correctly as 0 and 1 as 1:



* F1 is 0.91, Recall is 0.88 and precision is 0.96.
* F1 can be seen at 0.89 which is a good score as higher F1 better a model. Precision and recall has a good score to hence train data has good scores overall.
* The accuracy score is 98% which makes the model a good model.

Random forest is a good model since train and test has no difference or 0.2 difference in there accuracy score. The model is also good because the AUC and ROC scores are high and similar.

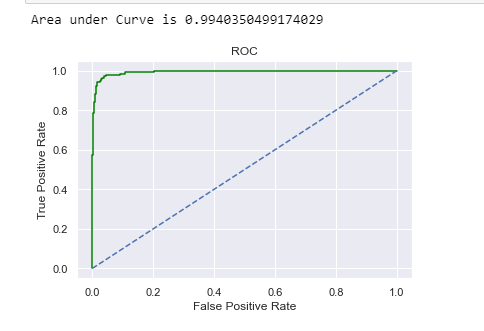
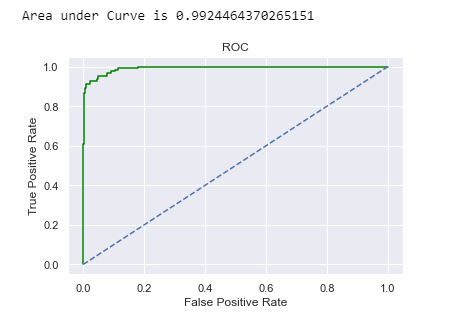
 

Fig 5- AUC AND ROC

TRAIN

TEST

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

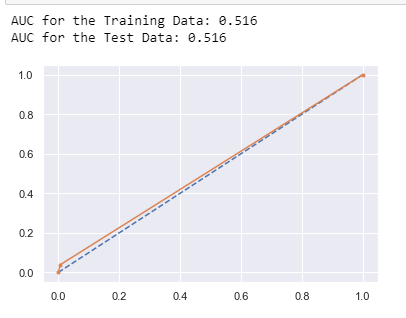
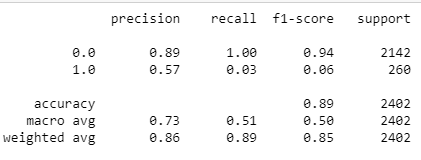


Fig 6 Classification Matrix and AUC/ROC scores

Inference:

* As seen in the image we have AUC is 0.749. Since the accuracy is high at 89%, we can say the model is a good model.
* F1 is 0.06, Recall is 0.03 and precision is 0.57.
* F1 can be seen at 0. 06which is a bad score as higher F1 better a model. Precision and recall has a bad score too hence train data has bad scores overall.

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

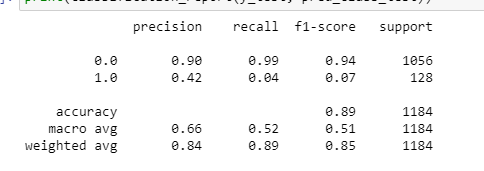
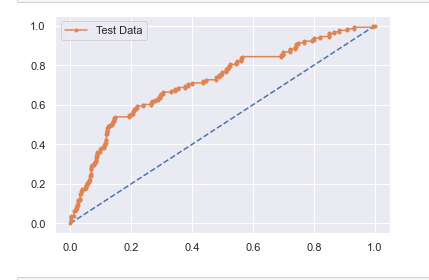


Fig 6 Classification Matrix and AUC/ROC scores

Inference:

* As seen in the image we have AUC is 0.516. Since the accuracy is high at 89%, we can say the model is an average model.
* F1 is 0.07, Recall is 0.04 and precision is 0.42.
* F1 can be seen at 0. 07 which is a bad score as higher F1 better a model. Precision and recall has a bad score too hence train data has bad scores overall.

LDA is not a fit model for this data set.

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

* Below is the models we tried to work on the data. However, the data seems to be bad data and we cannot get any good score:

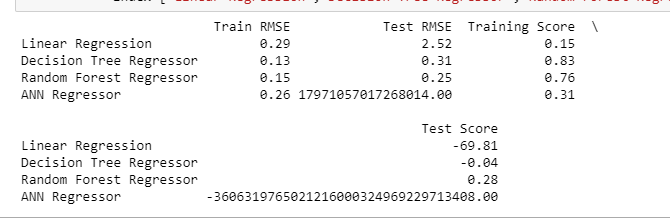


Fig-7a-Accuracy best parameter

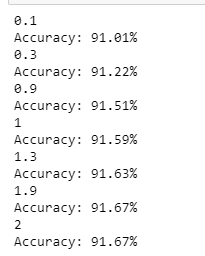
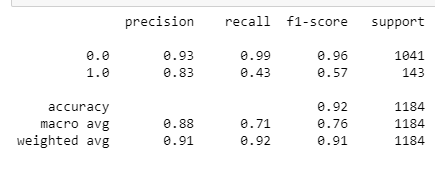
 

Fig-7b-Accuracy best parameter

Fig-7b-Classification matrix for test data

Interpretations:

* We have performed a lot of tests and found values coming in negatives which mean the data is failing to give us accurate results. We need to verify the data as it is bad data.
* We still do not see any great increase in precision score.
* The accuracy is at 92% which can be said with only test data we have good accuracy.
* The best can be seem from 1 and above as the parameter rises the score is becoming good.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Logistic Regression | Random Forest | Linear Discriminant Analysis |
| Test\_AUC | 0.95 | 0.982 | 0.516 |
| Train\_AUC | 0.94 | 0.983 | 0.749 |
| Train\_Accuracy | 0.93 | 0.89 | 0.89 |
| Test\_Accuracy | 0.94 | 0.98 | 0.89 |
| Train\_Precision | 0.9 | 0.9 | 0.57 |
| Test\_Precision | 0.81 | 0.96 | 0.42 |
| Train\_F1 | 0.62 | 0.89 | 0.06 |
| Test\_F1 | 0.59 | 0.91 | 0.07 |
| Train\_Recall | 0.47 | 0.88 | 0.03 |
| Test\_Recall | 0.47 | 0.88 | 0.04 |

The best model as per all the scores is logistic regression followed by random forest. LDA would be a bad model for this data set.

The accuracy scores of both train and test data is near by each other 93% and 94% in Logistic Regression which makes our model a best fit model. The predicted to actual us also at a good score with precision to recall for Logistic regression.

1.13 State Recommendations from the above models

* The networth next year is mostly negative and the same should be validated from the client as to why?
* There are a lot of outliers in the data which should also be fixed by removing bad and unnecessary data post asking the client.
* Hind Cables have the lowest Networth Next year so no investments should be made in there.
* Below companies have more assets so it can be a good option to invest in there:

|  |
| --- |
| Bank of Baroda |
| Bank of India |
| Punjab Natl.Bank |
| HDFC Bank |
| Axis Bank |
| IDBI Bank |
| Syndicate Bank |

* Debt ratio is more for below companies so no investment or careful investment should be made:

|  |
| --- |
| DFM Foods |
| KSE |
| PTL Enterprises |
| Sital Lease & Fn |
| Amrapali Capital |
| V2 Retail |
| Thangamayil Jew. |
| Capital Trade Lk |
| Hind Syntex |
| Binny |

Introduction:

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.

Data Summary

The data has 10 companies with its stock information from 31st March 2014 to 31st March 2020.

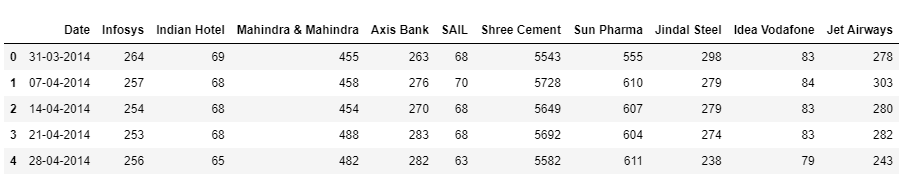


Fig 1. Data heading

The data has 11 columns and 314 rows which occupies a memory usage of 27.1 KB.



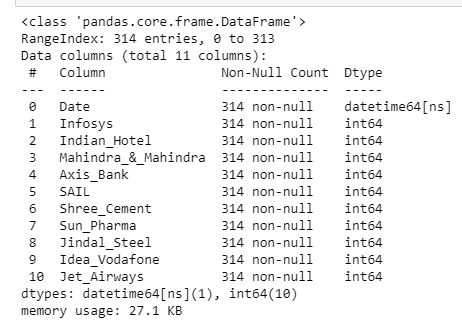


Fig 2. Data Information

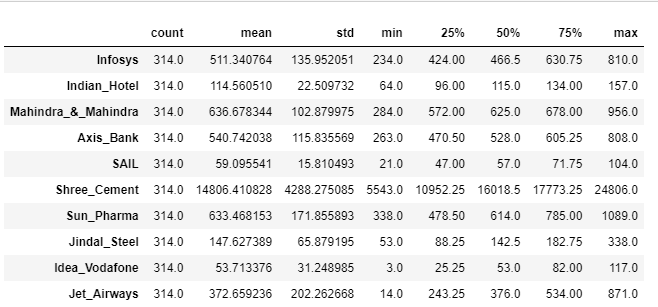


Fig 3. Data Description

Inference:

* The data has highest standard deviation for Shree\_Cement and the lowest for Indian\_Hotel.
* The mean is the highest for Shree\_Cement and lowest for Idea\_Vodafone.
* The minimum value is the highest for Shree\_Cement and lowest for Idea\_Vodafone.
* Overall even in the quartiles Shree\_Cement stands at high numbers.
* For quartiles Idea\_Vodafone is the lowest in 25th and 50th Quartile and Sail in 75th Quartile.
* Except for date all are integer values with no null columns

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

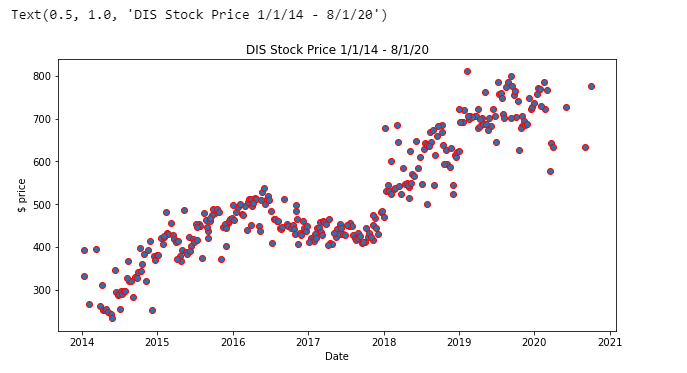


Fig 4.1 Stock price-Date (Infosys)

Inference:

* We see that there is an upward trend in the plot with each passing year.
* The data from 2014-2016 had a high stock rise with a bit of decline in 2017-2018 which then led to high rise from 2019-2020.
* We can see the highest stock price in the year 2020.

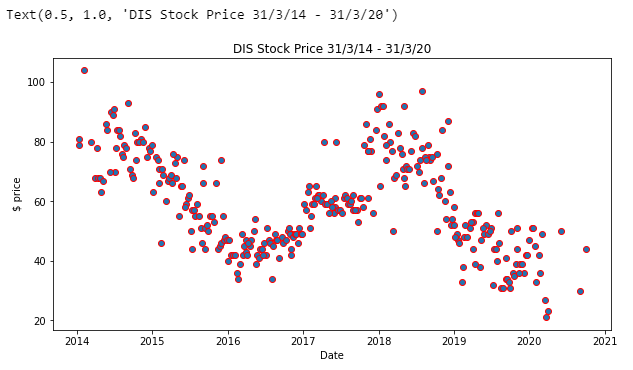


Fig 4.2 Stock price-Date (Sail)

Inference:

* We see the stock prices fluctuate with each year.
* The stock prices are high in 2014 declining until 2016, rising again from 2017-2017 and falling down low than all the years in 2020.
* Overall, there is no constant stock price but a lot of fluctuations.

2.2 Calculate Returns for all stocks with inference

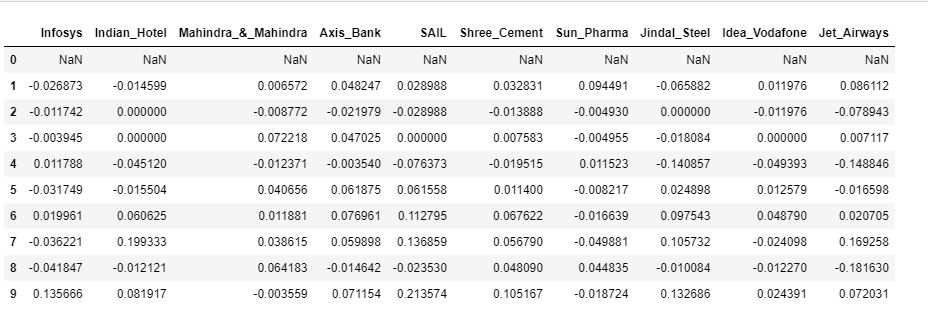


Fig 5.1 Stock Returns -top 10

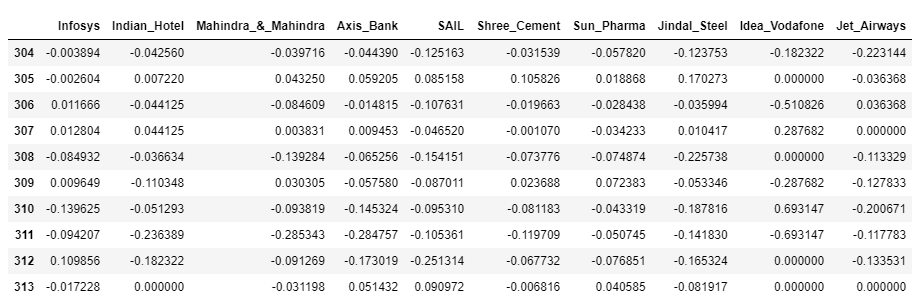


Fig 5.2 Stock Returns -Bottom 10

Inference:

* A negative rate of return is **a loss of the principal invested for a specific period of time**. The negative may turn into a positive in the next period, or the one after that. A negative rate of return is a paper loss unless the investment is cashed in.
* A negative return refers to a loss, either on an investment, a business's performance, or on invested projects.
* The Indian Hotel and Sail has the most negative returns along with Jet Airways.
* Shree Cement has the best returns.
* Infosys registers the least along with Jindal Steel.

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

Standard deviation is the statistical measure of market volatility, measuring how widely prices are dispersed from the average price. The smaller an investment's standard deviation, the less volatile it is. The larger the standard deviation, the more dispersed those returns are and thus the riskier the investment is.

Mean return, in securities analysis, is the expected value, or mean, of all the likely returns of investments comprising a portfolio. A mean return is also known as an expected return and can refer to how much a stock returns on a monthly basis.

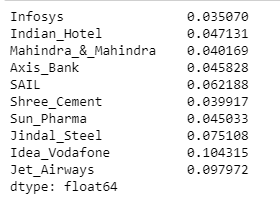
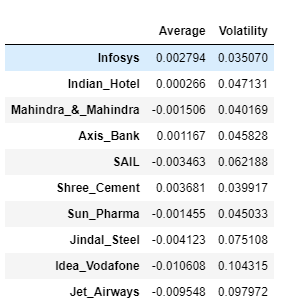
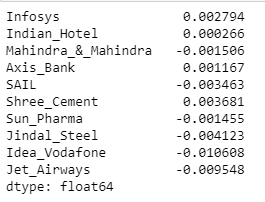
 

Fig 5.3 Data Frame

Fig 5.3 Stock Means

Fig 5.3 Stock Standard Deviations

Inference:

* From the stock standard deviation, we see the highest is for Idea Vodafone and the lowest is for Infosys at 0.035070.
* Average in data frame means ‘Mean’ and Volatility means ‘Standard Deviation’.
* Except for Infosys, Indian Hotel, Axis Bank and Shree Cement all the data has mean and standard deviation with negative and positive value which means data is not cluttered around mean.
* The highest Standard deviation is for Idea Vodafone which means investing in their stocks will be risky. Safest investment will be Infosys and Shree Cement.

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

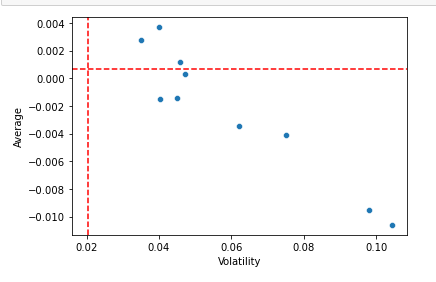


Fig 6 Stock Standard Deviations

Inference:

* Average = Mean; Volatility = Standard Deviation
* The line parallel to X axis or Volatility is our margin line.
* We see 3 dots beyond the line of margin which means we do have high returns. However with the data we have maximum is below the line stating with standard deviation and mean the return is low.
* 0.10 of volatility is the lowest hence the return is the least. Idea Vodafone has the lowest return is yet again confirmed from here.
* Mahindra, Infosys and Axis is a good return on investment.

2.5 Conclusion and Recommendations

* From the data that we have it will be a better option to check their investments and then invest. More the invest more will be the risk and which is why we see standard deviation as high is some cases
* The best stock investment is of Infosys and Shree Cement where chances of return is more as risk is less.
* Since we see Idea Vodafone as the decreasing stock prices it will be a bad investment on it.
* Jindal Steel has been fluctuating a lot with returns as share price rise and fall every year rapidly.
* Shree Hotel, Infosys, Indian Hotel and Axis bank have less market fluctuations with upward rising trend so returns will definitely be high.
* Sail, Jet Airways, Idea Vodafone and Sun Pharma has been decreasing stocks with each year and it would be a bad selection for investment.

THE END