

Finance and Risk Analytics

Business Report



**Post Graduate Program in
Data Science and Business Analytics**

February 2022

Athisya Nadar

athisya@gmail.com

Contents

1.8 Random Forest Model (Train).....	4
1.9 Random Forest Model (Test).....	6
1.10 LDA Model (Train)	7
1.11 LDA Model (Test)	8
1.12 Logistics, Radom Forest and LDA comparison.....	10
1.13 Recommendations	12
2.1 Stock Price Graph.....	13
2.2 Stock Returns.....	18
2.3 Stock Means and Standard Deviation.....	19
2.4 Plot of Stock Means vs Standard Deviation.....	20
2.5 Conclusion and Recommendations.....	24

Problem Statement

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.

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.





Model comparison for credit risk

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

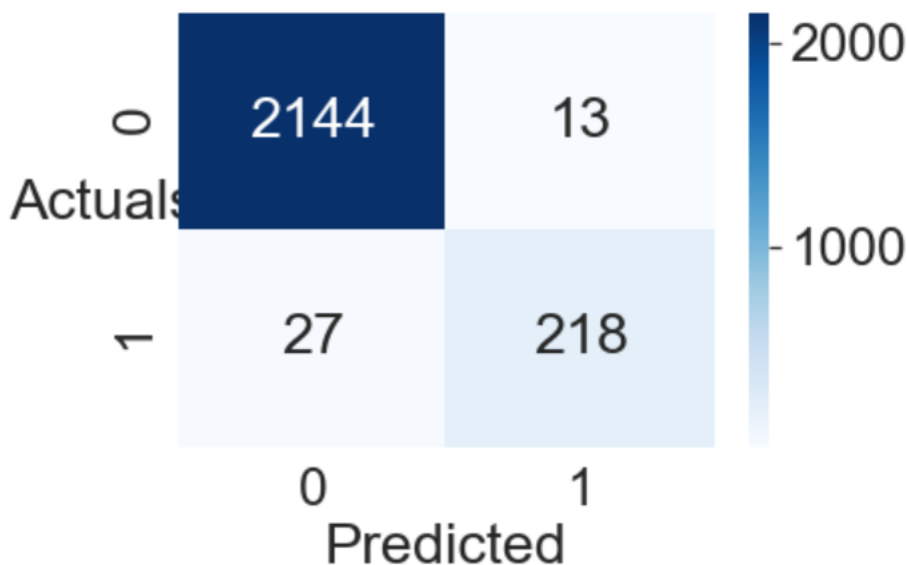
For Random forest model we have used RandomForestClassifier from klearn.ensemble package to build this model. Then we have used GridSearchCV from sklearn.model_selection to get the best parameters.

Let us look at the classification report after running the model:

	precision	recall	f1-score	support
0	0.99	0.99	0.99	2157
1	0.94	0.89	0.92	245
accuracy			0.98	2402
macro avg	0.97	0.94	0.95	2402
weighted avg	0.98	0.98	0.98	2402

We have got precision of 94% and recall of 89%

Now let us look at the confusion matrix:



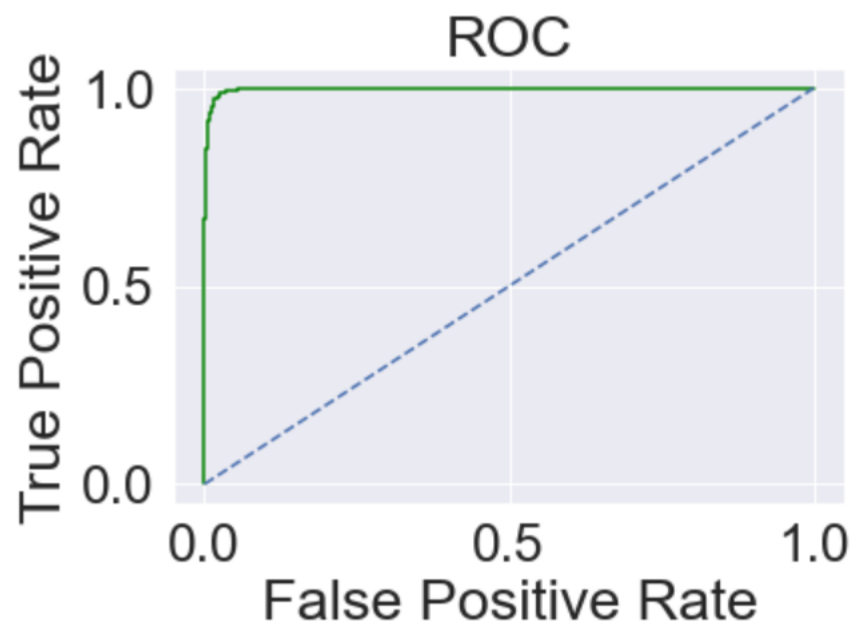
True Positive: 2144 non defaulters are classified correctly

True Negative: 218 defaulters are classified correctly

False Positive: 27 defaulters are classified as non-defaulters incorrectly

False Negative: 13 non defaulters are classified as defaulter incorrectly

This model has accuracy of 98.33% which is excellent



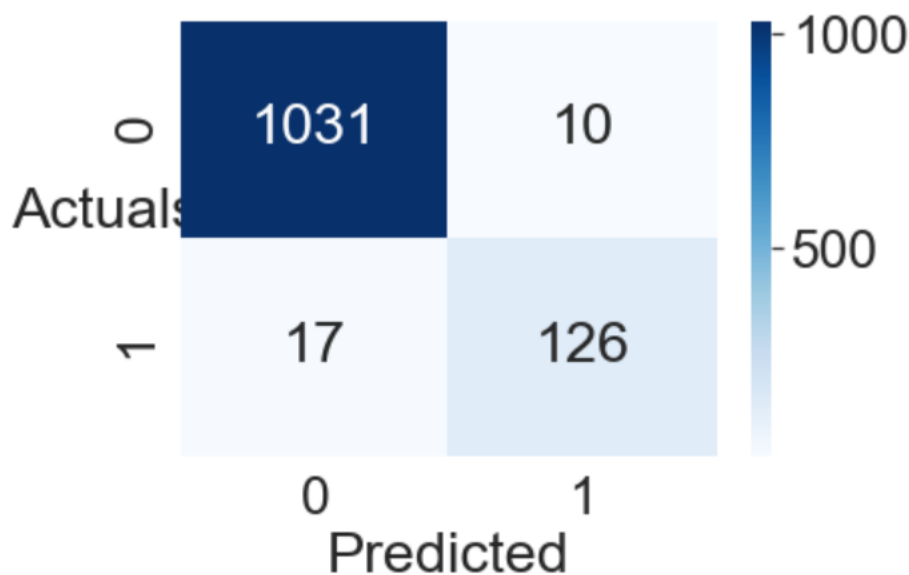
Area under Curve is 0.9978863311666808, overall this is a very good model.

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

Next we have validated the random forest model on test data

	precision	recall	f1-score	support
0	0.98	0.99	0.99	1041
1	0.93	0.88	0.90	143
accuracy			0.98	1184
macro avg	0.96	0.94	0.95	1184
weighted avg	0.98	0.98	0.98	1184

With 88% recall and 93% precision this model has performed good on test data. We can also check the confusion matrix to understand better:



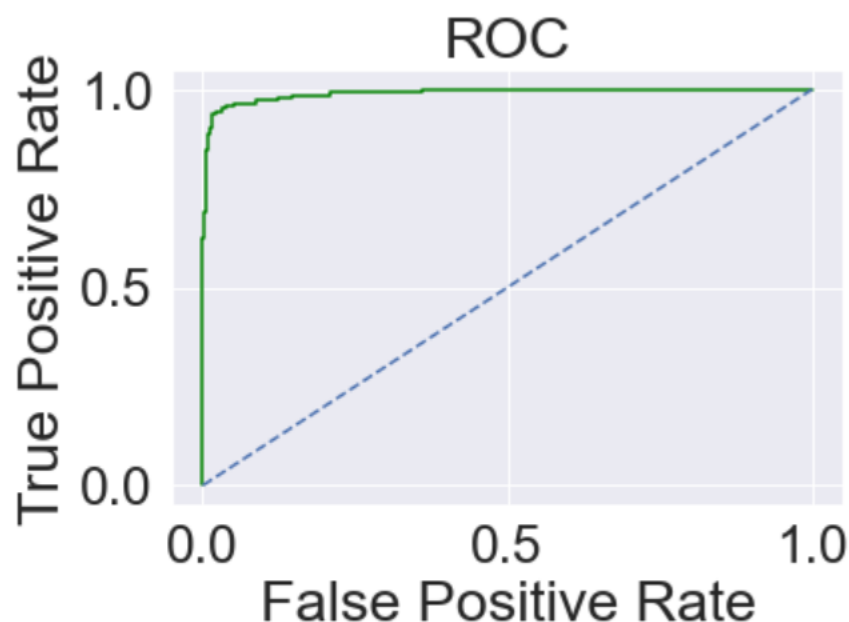
True Positive: 1031 non defaulters are classified correctly

True Negative: 126 defaulters are classified correctly

False Positive: 17 defaulters are classified as non-defaulters incorrectly

False Negative: 10 non defaulters are classified as defaulter incorrectly

Out of 143 defaulters we were able to predict 126 defaulters correctly. This model has accuracy of 97.71% which is excellent.



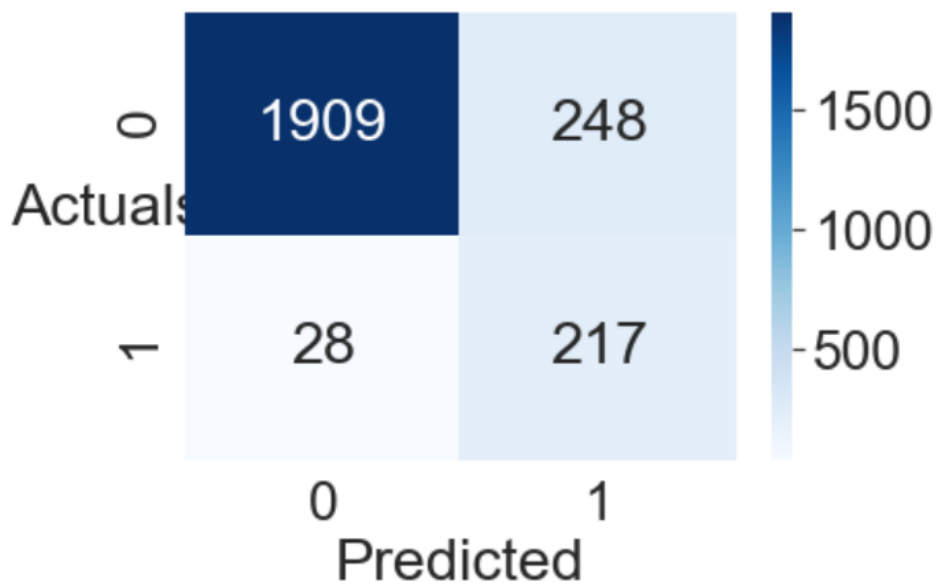
Area under Curve is 0.9907095786058322

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

We have used `LinearDiscriminantAnalysis` from `sklearn.discriminant_analysis` for building the model.

	precision	recall	f1-score	support
0	0.986	0.885	0.933	2157
1	0.467	0.886	0.611	245
accuracy			0.885	2402
macro avg	0.726	0.885	0.772	2402
weighted avg	0.933	0.885	0.900	2402

With 58% recall and 80% precision, we have not got a good model



True Positive: 1909 non defaulters are classified correctly

True Negative: 217 defaulters are classified correctly

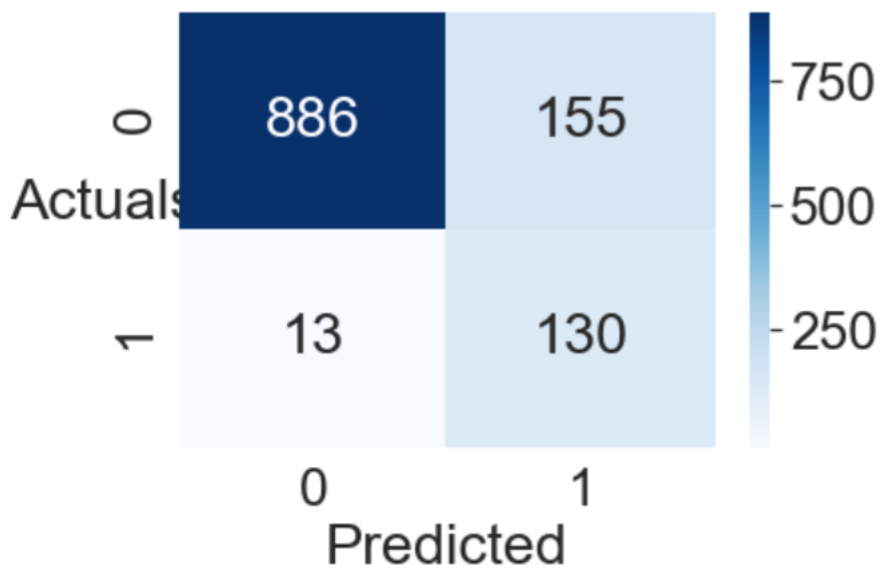
False Positive: 28 defaulters are classified as non-defaulters incorrectly

False Negative: 248 non defaulters are classified as defaulter incorrectly

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

Now we can run the model on test data and look at classification report:

	precision	recall	f1-score	support
0	0.986	0.851	0.913	1041
1	0.456	0.909	0.607	143
accuracy			0.858	1184
macro avg	0.721	0.880	0.760	1184
weighted avg	0.922	0.858	0.876	1184

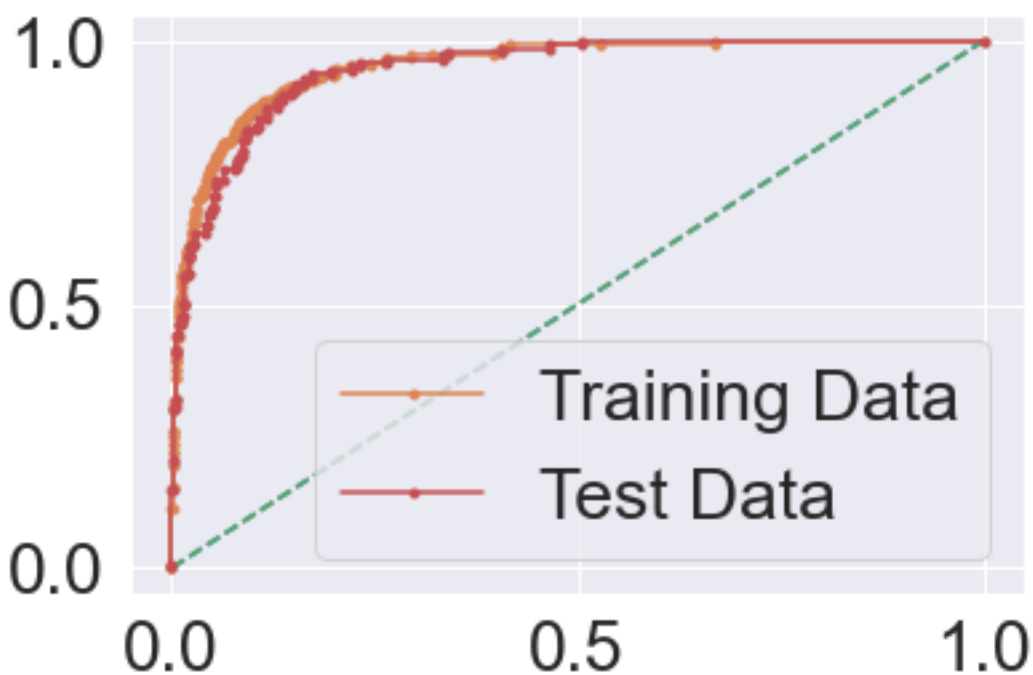


True Positive: 886 non defaulters are classified correctly

True Negative: 130 defaulters are classified correctly

False Positive: 13 defaulters are classified as non-defaulters incorrectly

False Negative: 155 non defaulters are classified as defaulter incorrectly



AUC for the Training Data: 0.954

AUC for the Test Data: 0.948

AUC of 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and more than 0.9 is considered outstanding.

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

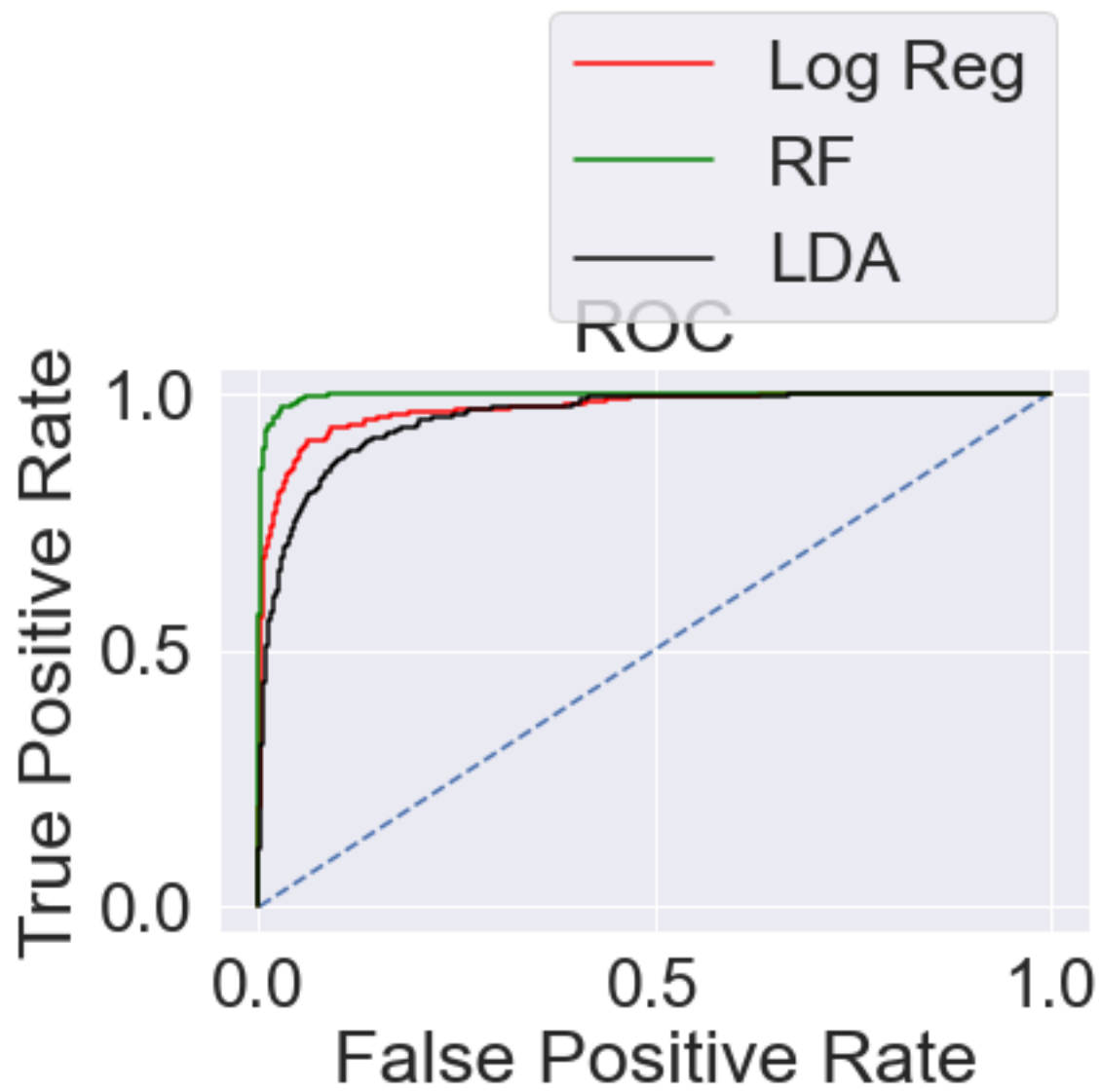
	Logistics		Random Forest		LDA	
	train	test	train	test	train	test
recall	0.873	0.902	0.89	0.88	0.886	0.909
precision	0.688	0.683	0.94	0.93	0.467	0.456
AUC	0.970	0.986	0.997	0.990	0.954	0.948
F1 score	0.77	0.77	0.91	0.90	0.611	0.607

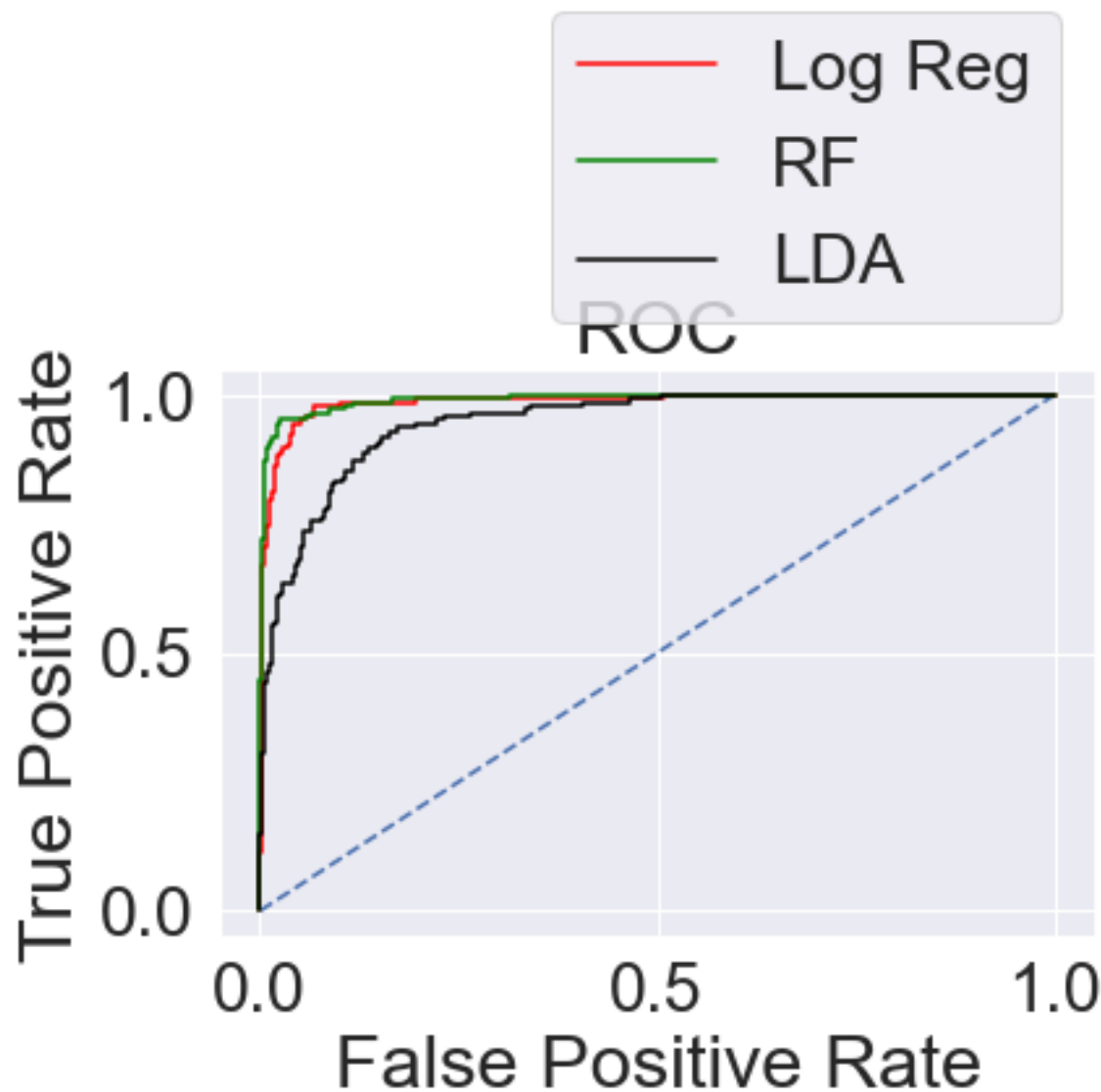
Precision can be seen as a measure of quality, and recall as a measure of quantity. Higher precision means that an algorithm returns more relevant results than irrelevant ones, and high recall means that an algorithm returns most of the relevant results.

On the test data, we can see that logistic regression has given the best recall, however random forest gives us a better precision.

ROC is a probability curve and AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. Higher the AUC, the better the model is at predicting 0 classes as 0 and 1 classes as 1.

AUC of 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and more than 0.9 is considered outstanding. We have got outstanding AUC for all 3 models.





1.13 State Recommendations from the above models.

After building logistic regression, random forest and LDA model we can conclude that random forest and logistic regression have yielded best results.

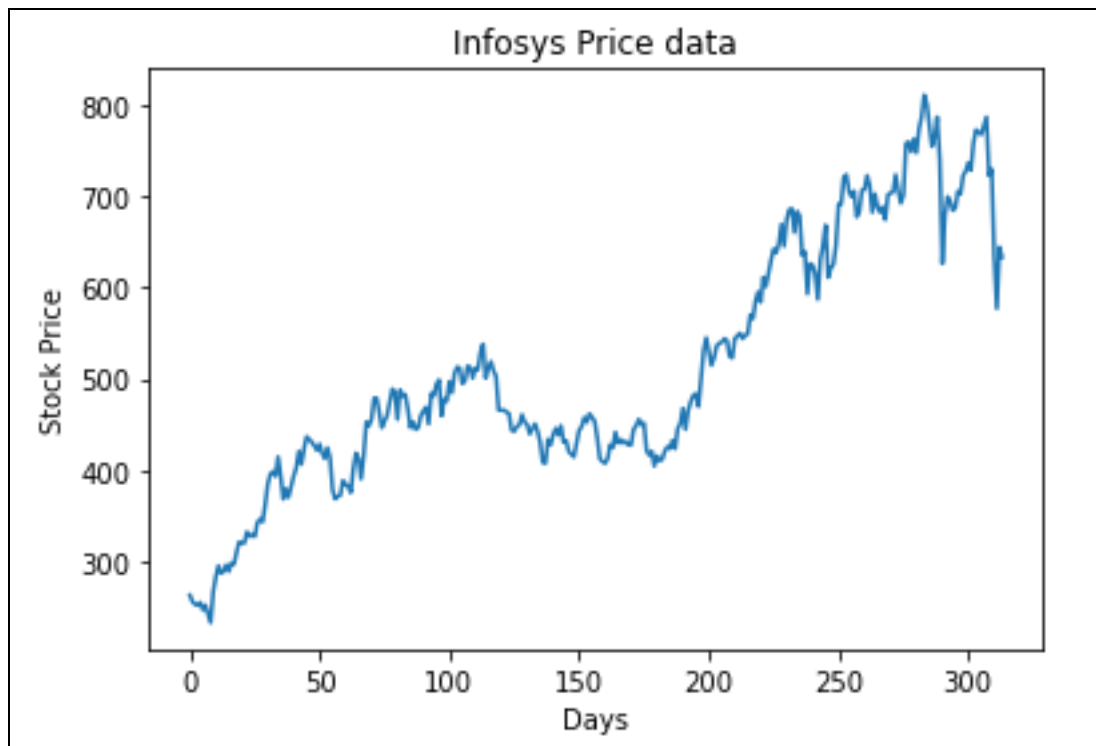
A balance sheet is a financial statement of a company that provides a snapshot of what a company owns, owes, and the amount invested by the shareholders. We have tried to predict if the company will default or not using various models.

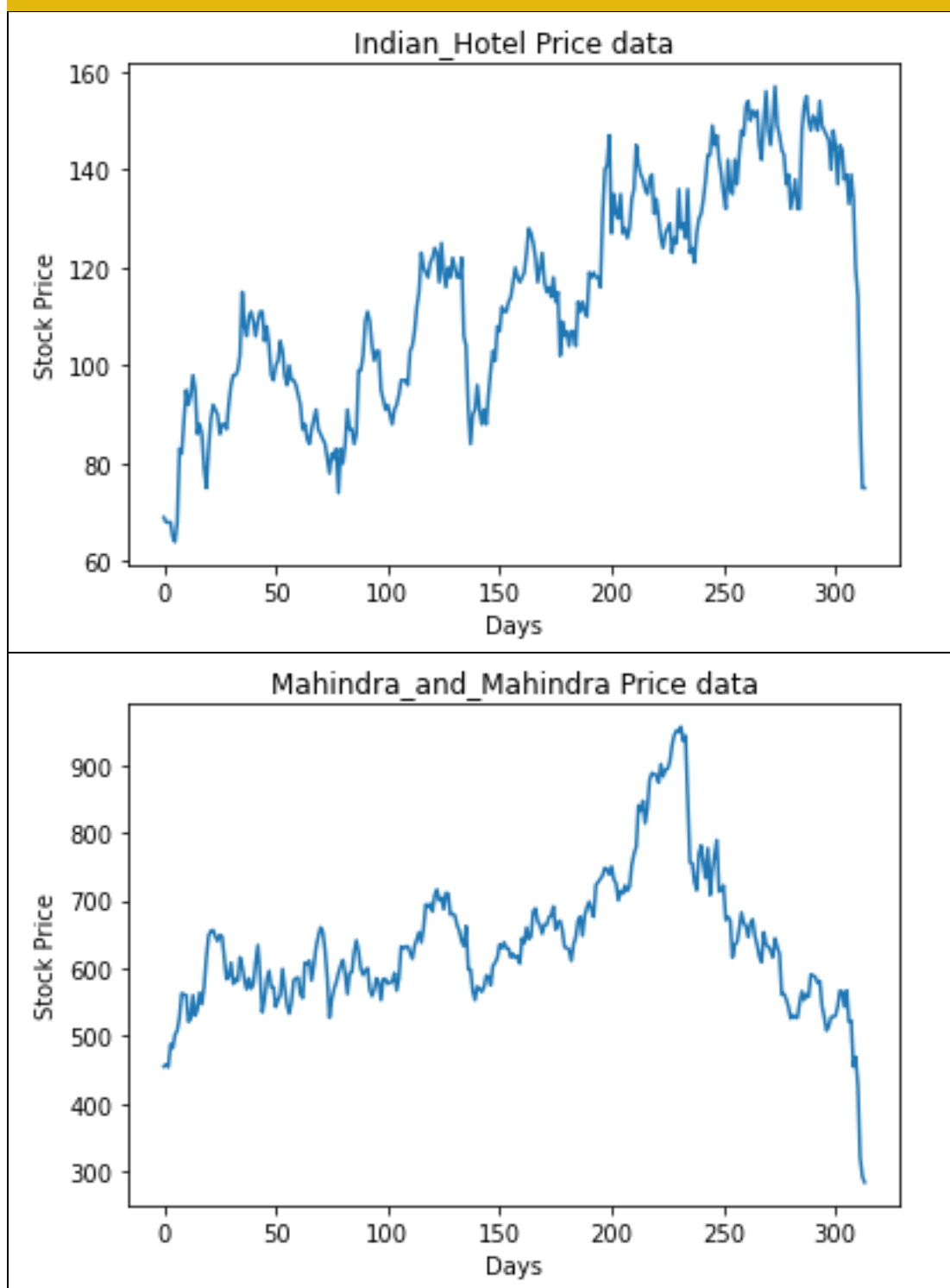
Credit risk plays a crucial role in deciding company's credit ratings. The company may drop its rating if they default. Credit rating agencies assign

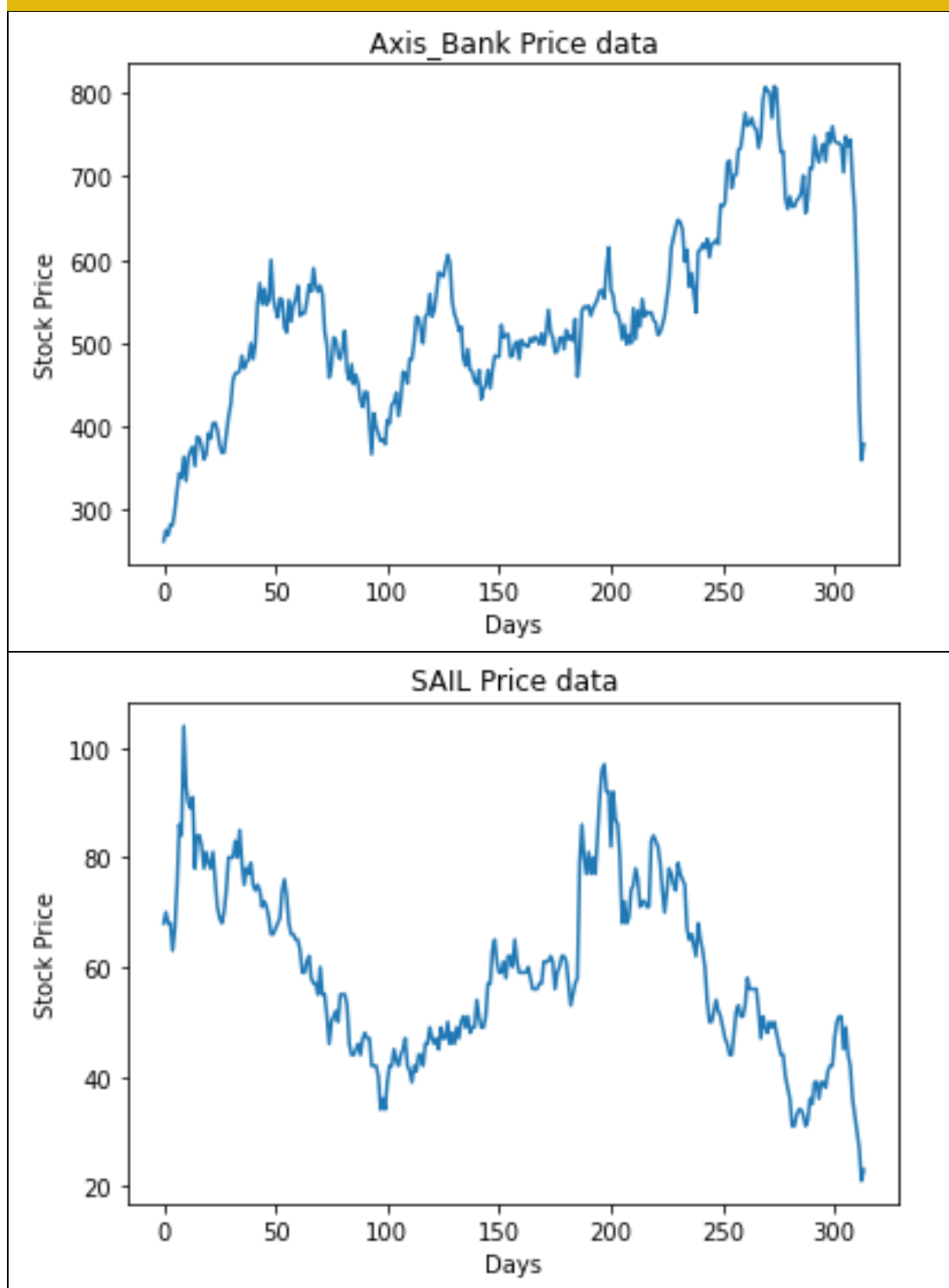
credit ratings based on a debtor's ability to pay back the debt by making timely principal and interest payments and the likelihood of default.

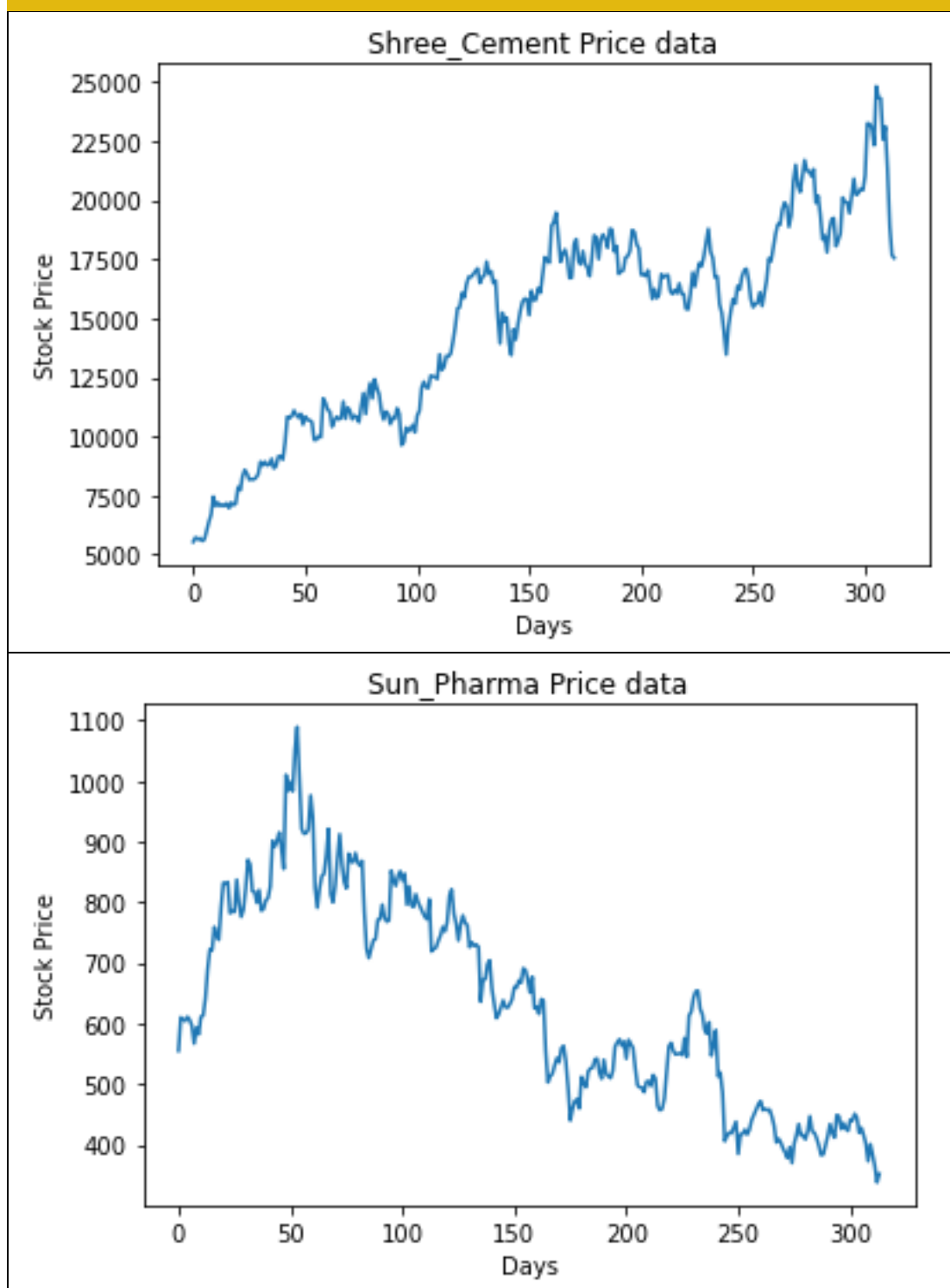
we conclude that statistical methods play a key role in credit risk modeling.

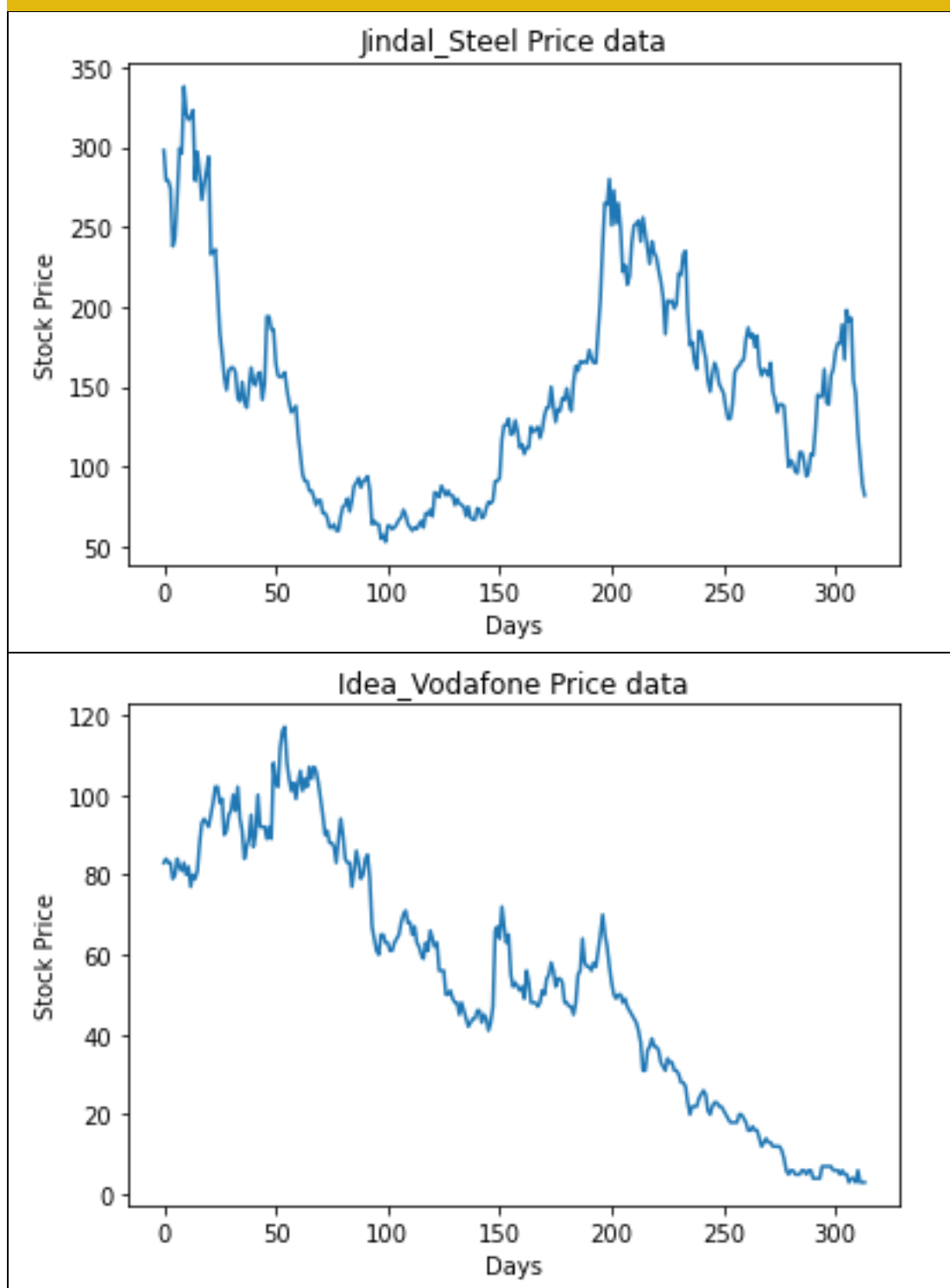
2.1 Draw Stock Price Graph(Stock Price vs Time) for any 2 given stocks with inference

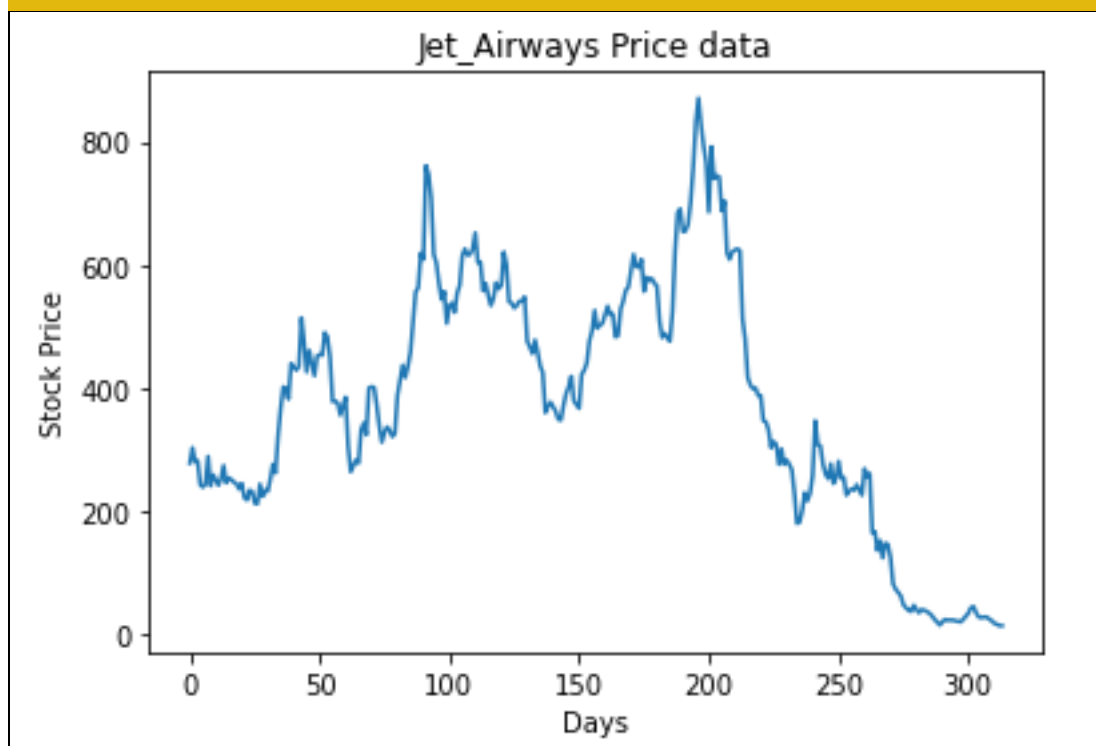










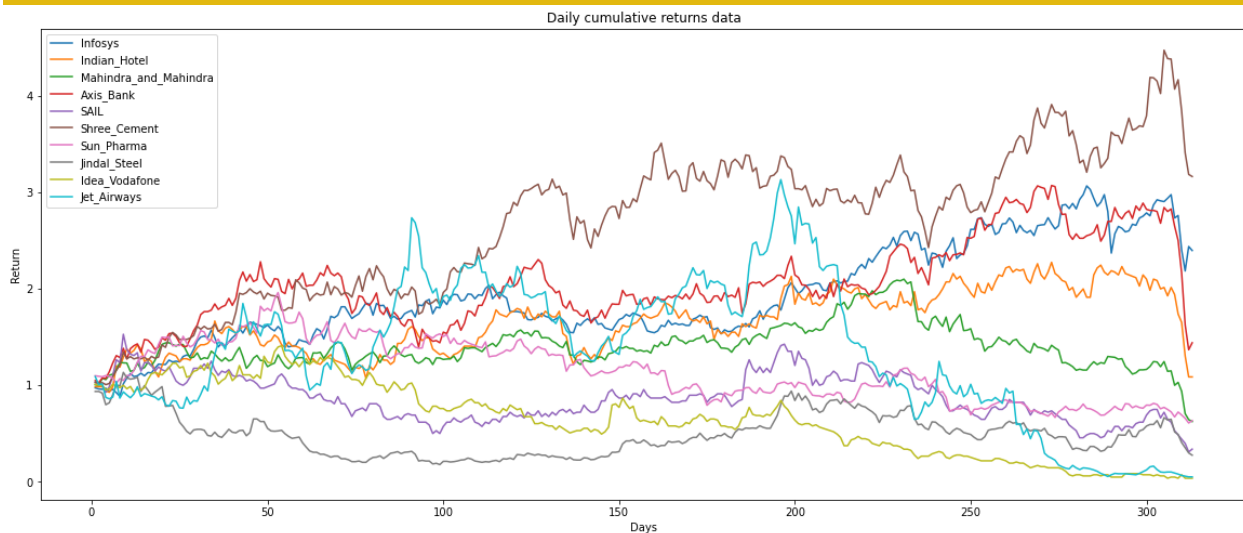


2.2 Calculate Returns for all stocks with inference

After using `cumprod` function in python to calculate cumulative returns we have plotted the returns:

Infosys	139.772727
Indian Hotel	8.695652
Mahindra	-37.582418
Axis bank	44.106464
SAIL	-66.176471
Shree Cement	216.543388
Sun Pharma	-36.576577
Jindal Steel	-72.483221
Vodafone Idea	-96.385542
Jet Airways	-94.964029

We can say only Infosys, Indian hotel, axis bank and shree cement have given good returns.



When we calculate returns Vodafone stock is given worst return and Infosys and given best returns.

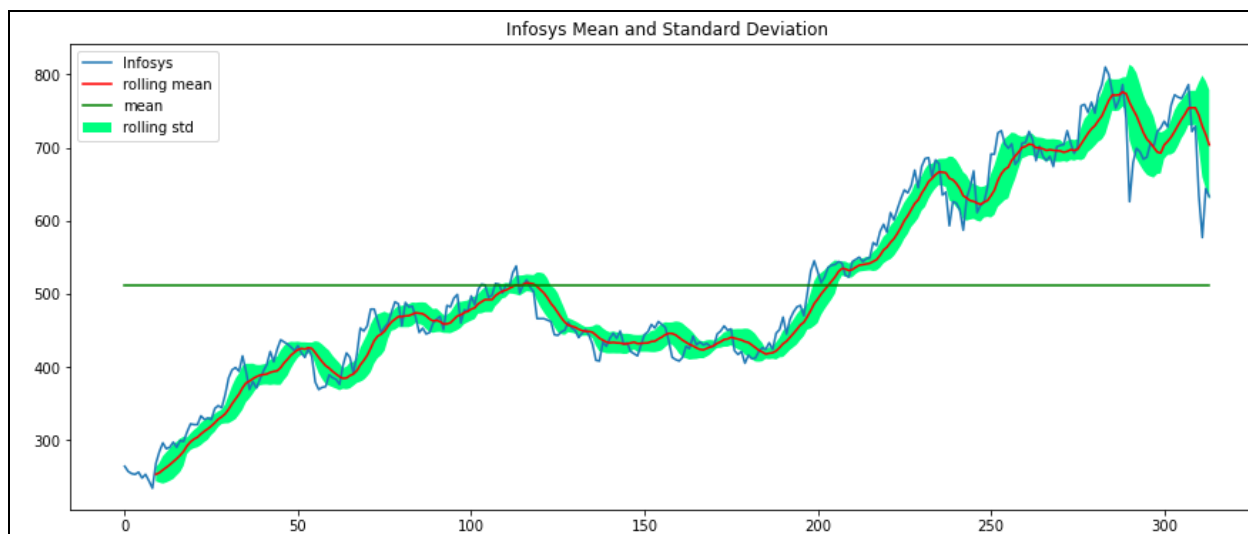
2.3 Calculate Stock Means and Standard Deviation for all stocks with inference

	minimum	maximum	mean	Standard deviation
Infosys	234.0	810.0	511.340764	135.952051
Indian_Hotel	64.0	157.0	114.56051	22.509732
Mahindra_and_Mahindra	284.0	956.0	636.678344	102.879975
Axis_Bank	263.0	808.0	540.742038	115.835569
SAIL	21.0	104.0	59.095541	15.810493
Shree_Cement	5543.0	24806.0	14806.410828	4288.275085
Sun_Pharma	338.0	1089.0	633.468153	171.855893
Jindal_Steel	53.0	338.0	147.627389	65.879195
Idea_Vodafone	3.0	117.0	53.713376	31.248985
Jet_Airways	14.0	871.0	372.659236	202.262668

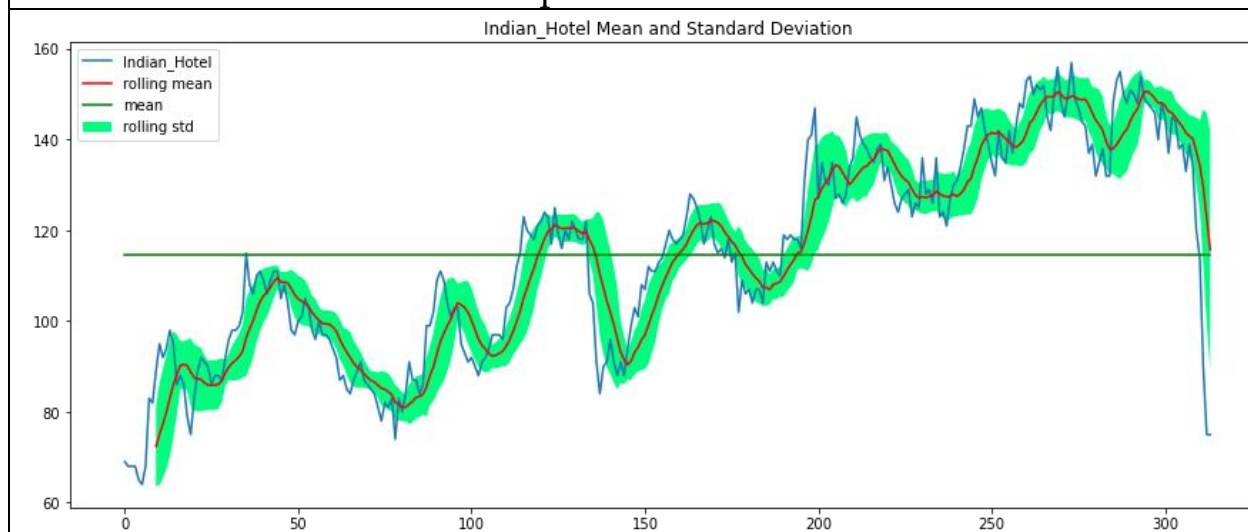
Standard deviation helps determine market volatility or the spread of asset prices from their average price. When prices move wildly, standard deviation is high, meaning an investment will be risky. Low standard deviation means prices are calm, so investments come with low risk.

mean is a good tool to evaluate the performance of a company or portfolio. When we look at the stock performance of jet airways, we can see that stand deviation is high which means it is a highly risky stock. however, Mahindra and Mahindra standard deviation from mean is less, hence it is less risky. Closer the mean to maximum values better the stock.

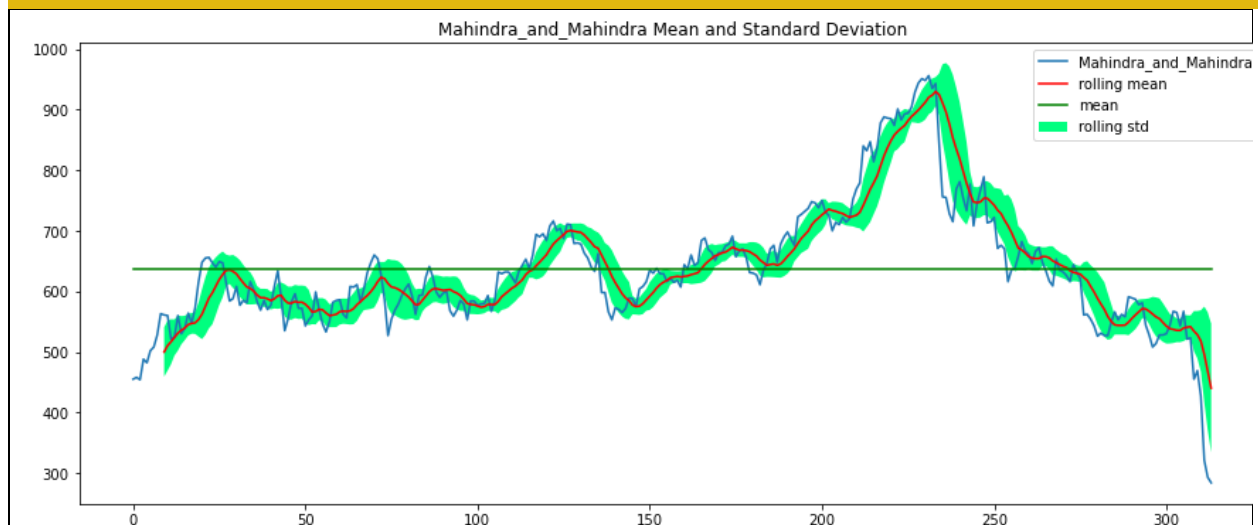
2.4 Draw a plot of Stock Means vs Standard Deviation and state your inference



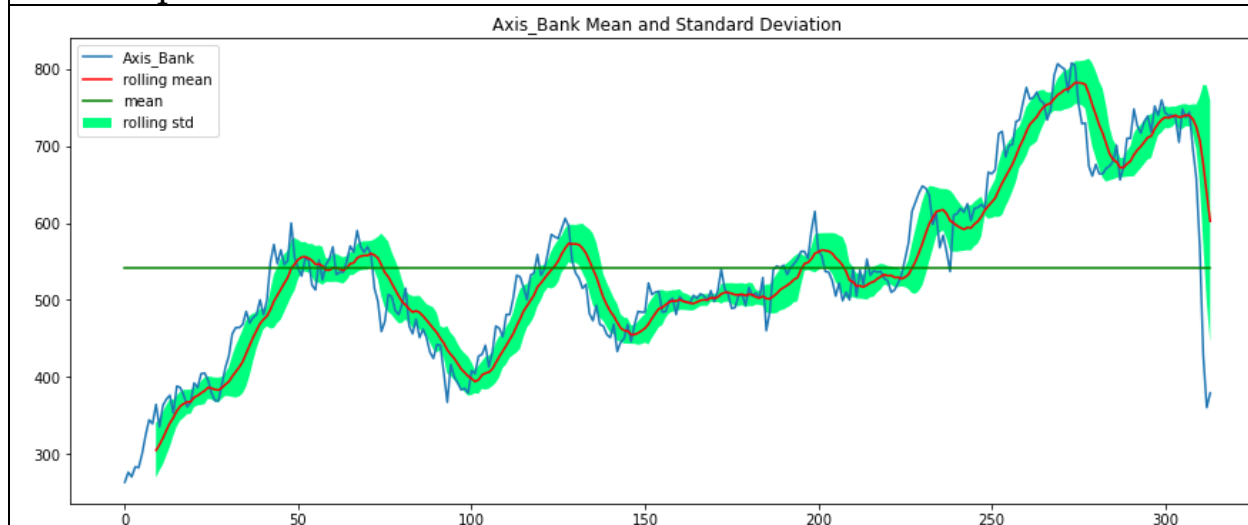
This graph shows mean of 314 days, rolling standard deviation and rolling mean across 10 days for Infosys stock. We can see an upward trend for this stock in this time period.



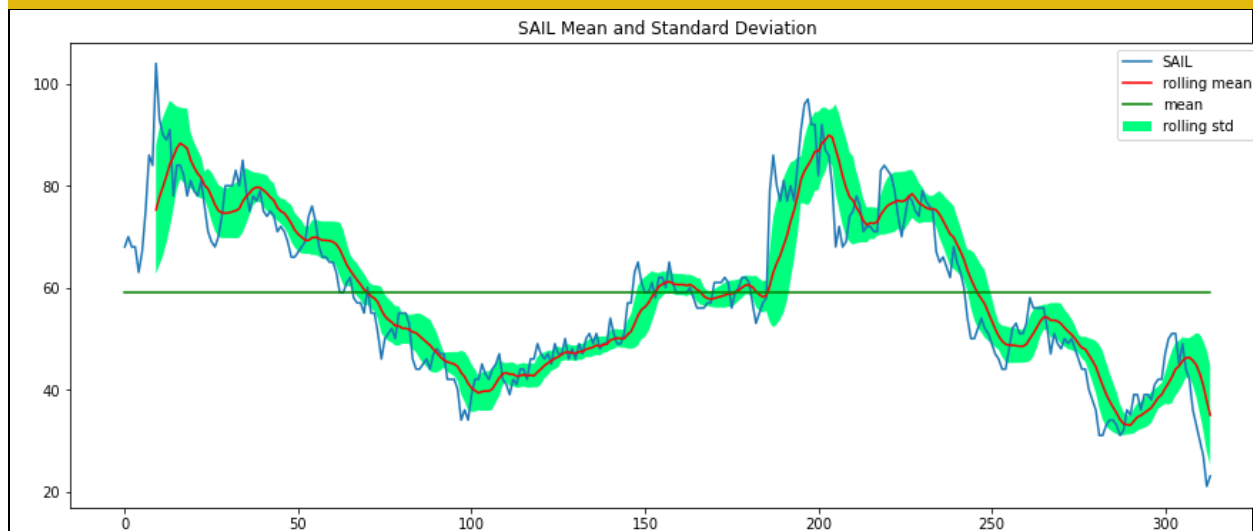
This graph shows mean of 314 days, rolling standard deviation and rolling mean across 10 days for Indian Hotel stock. We can see an upward trend for this stock in this time period.



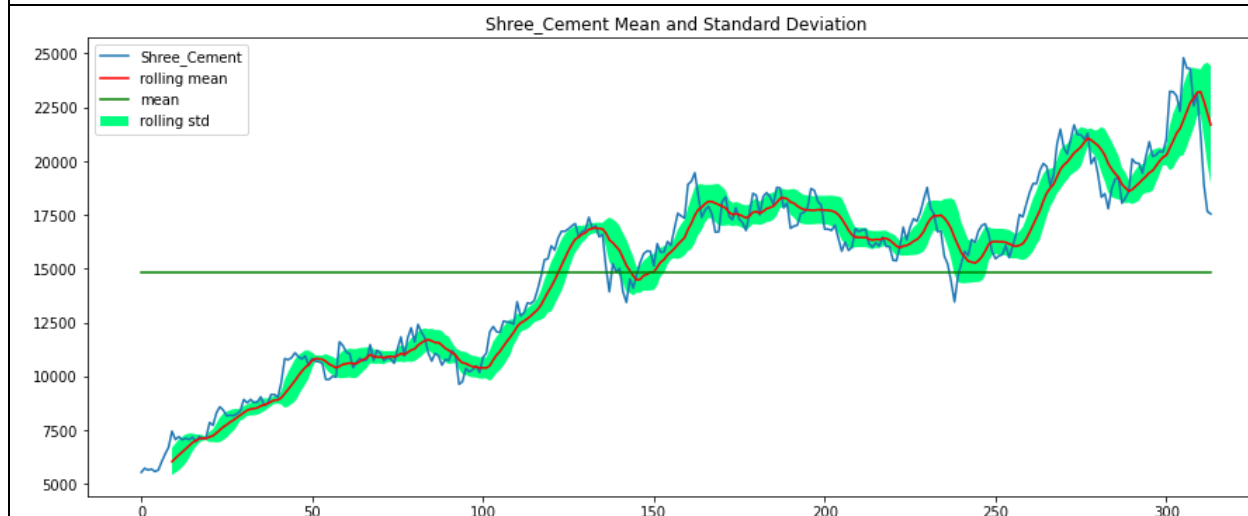
This graph shows mean of 314 days, rolling standard deviation and rolling mean across 10 days for Mahindra and mahindra stock. We can see an upward trend for this stock but in consistent around the mean.



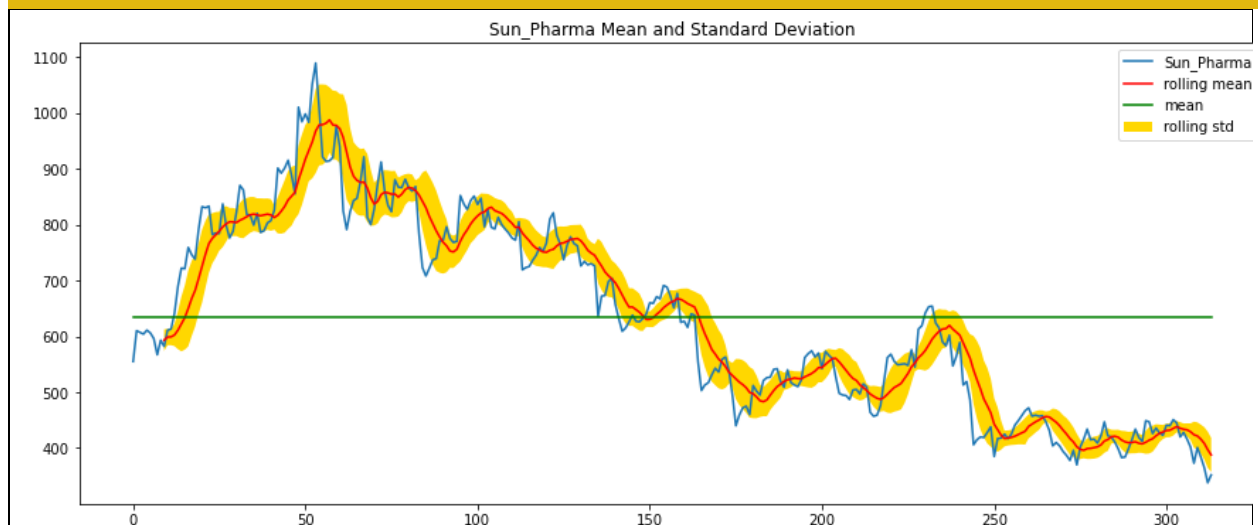
This graph shows mean of 314 days, rolling standard deviation and rolling mean across 10 days for Axis bank stock. We can see an upward trend for this stock during this time period.



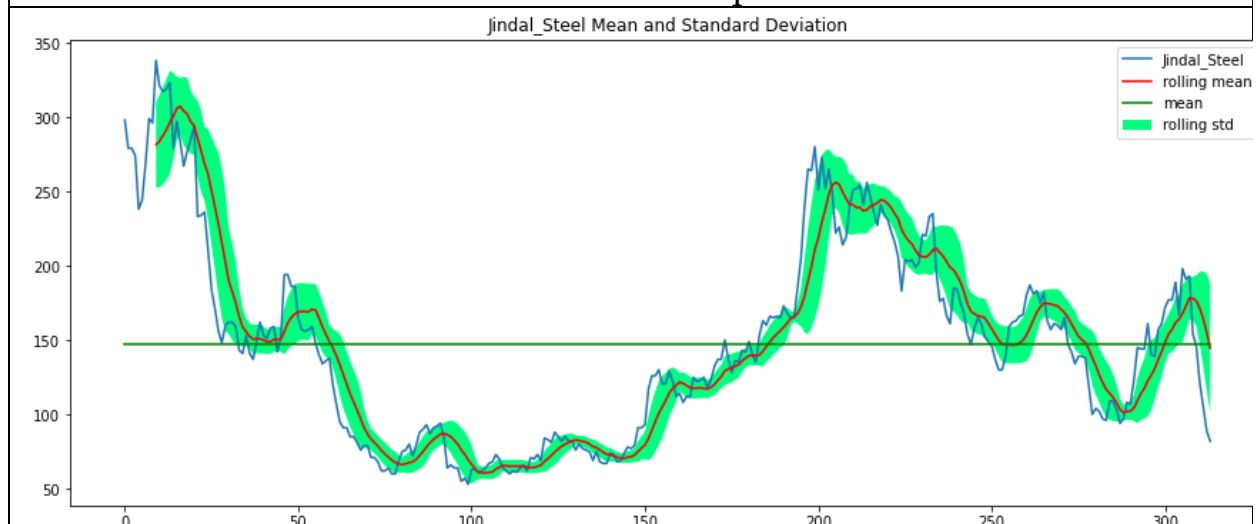
This graph shows mean of 314 days, rolling standard deviation and rolling mean across 10 days for SAIL stock. We can see a lot of volatility for this stock.



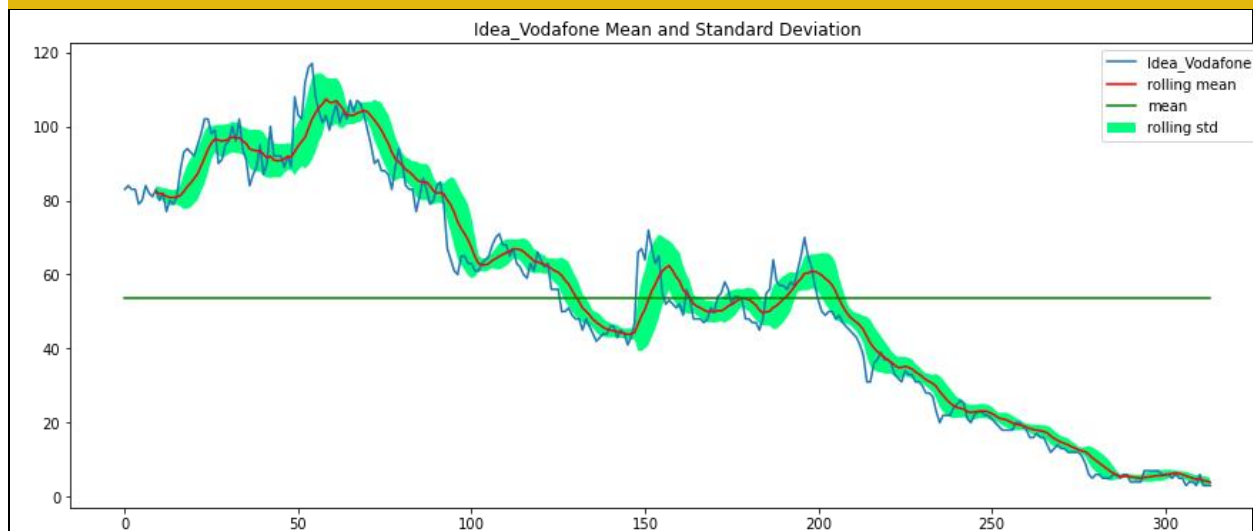
This graph shows mean of 314 days, rolling standard deviation and rolling mean across 10 days for Shree cement stock. We can see an upward trend for this stock.



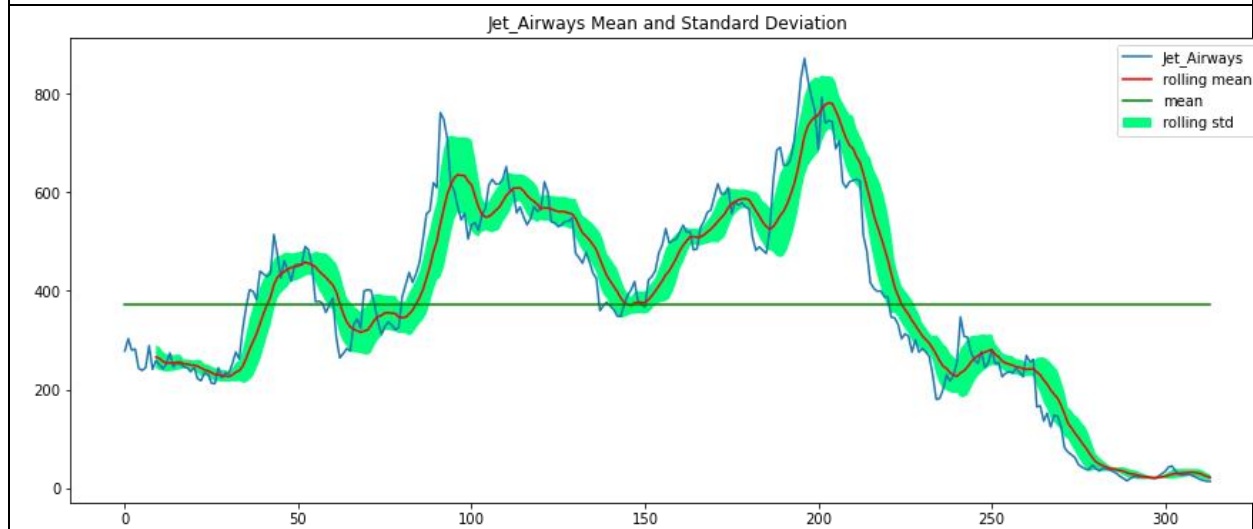
This graph shows mean of 314 days, rolling standard deviation and rolling mean across 10 days for Sun pharma stock. We can see an downward trend for this stock in this time period.



This graph shows mean of 314 days, rolling standard deviation and rolling mean across 10 days for Jindal steel stock. We can see lot of volatility for this stock in this time period because the there is lot of deviation from mean.



This graph shows mean of 314 days, rolling standard deviation and rolling mean across 10 days for idea vodafone stock. We can see lot of volatility and downward trend for this stock.



This graph shows mean of 314 days, rolling standard deviation and rolling mean across 10 days for jet airways stock. We can see lot of volatility and upward as well as downward trend for this stock. Also lot of deviation from the mean hence unpredictable and risky.

2.5 Conclusion and Recommendations

From the above analysis if I were to pick the best stock I would pick Infosys, shree cement, axis bank since these stocks show less volatility, consistency and upward trend across 314 days.

The worst stocks can be jet airways and idea Vodafone because they are highly volatile and unpredictable.

Investors who want to play safe, may stick to shree cement or Mahindra which are not very volatile, but consistent and steady.

Cumulative returns graph clearly identifies Infosys, axis bank and shree cement as the best performers which is something every investor would be interested in.





Thank You