Predicting Stock Prices for three tech companies

(Apple, Google, Microsoft)

Sima Nikzad

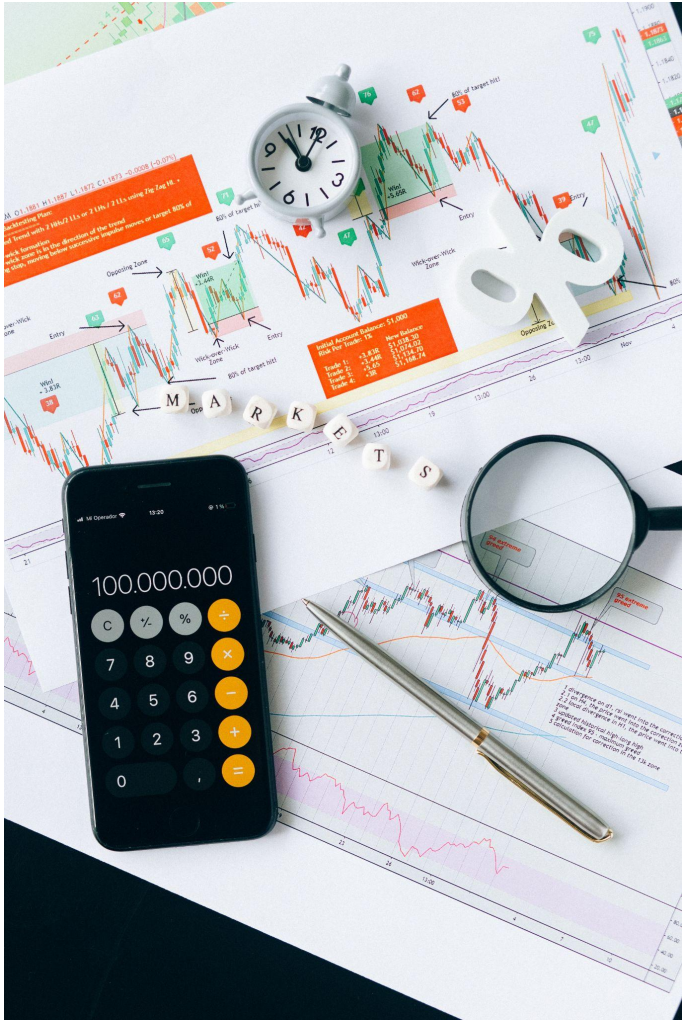
Data Science Capstone project

The Problem

Stock market prediction determines the future movement of the stock value of a financial exchange. The accurate prediction of share price movement will lead to more profit for investors.

Predicting how the stock market will move is one of the most challenging issues due to many factors that involved in the stock prediction, such as interest rates, politics, economic growth and ... which make the stock market volatile and very hard to predict accurately.

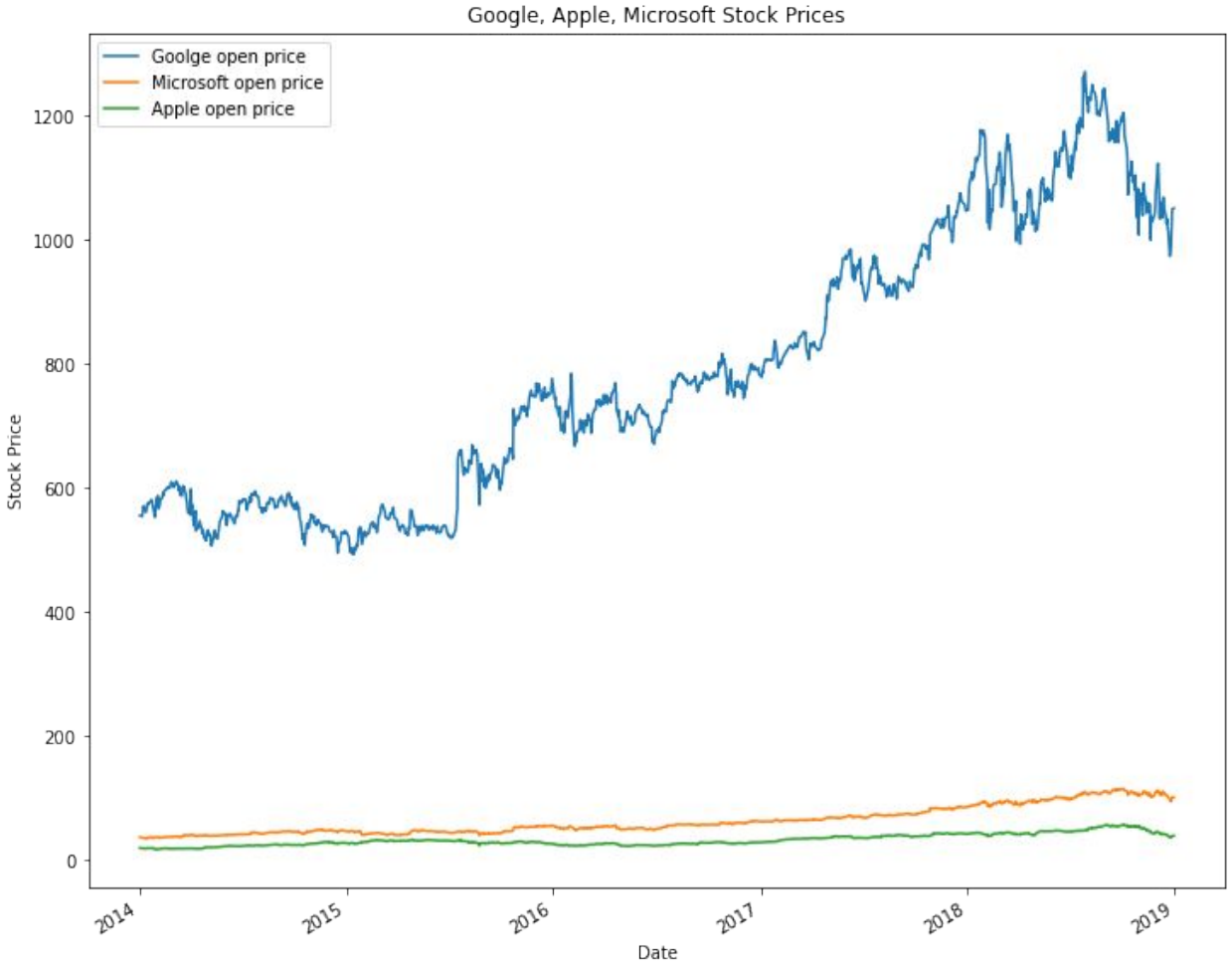
Company X wants to create a model to predict a future stock prices, there are three particular stock prices, will be analyzed in our presentation.



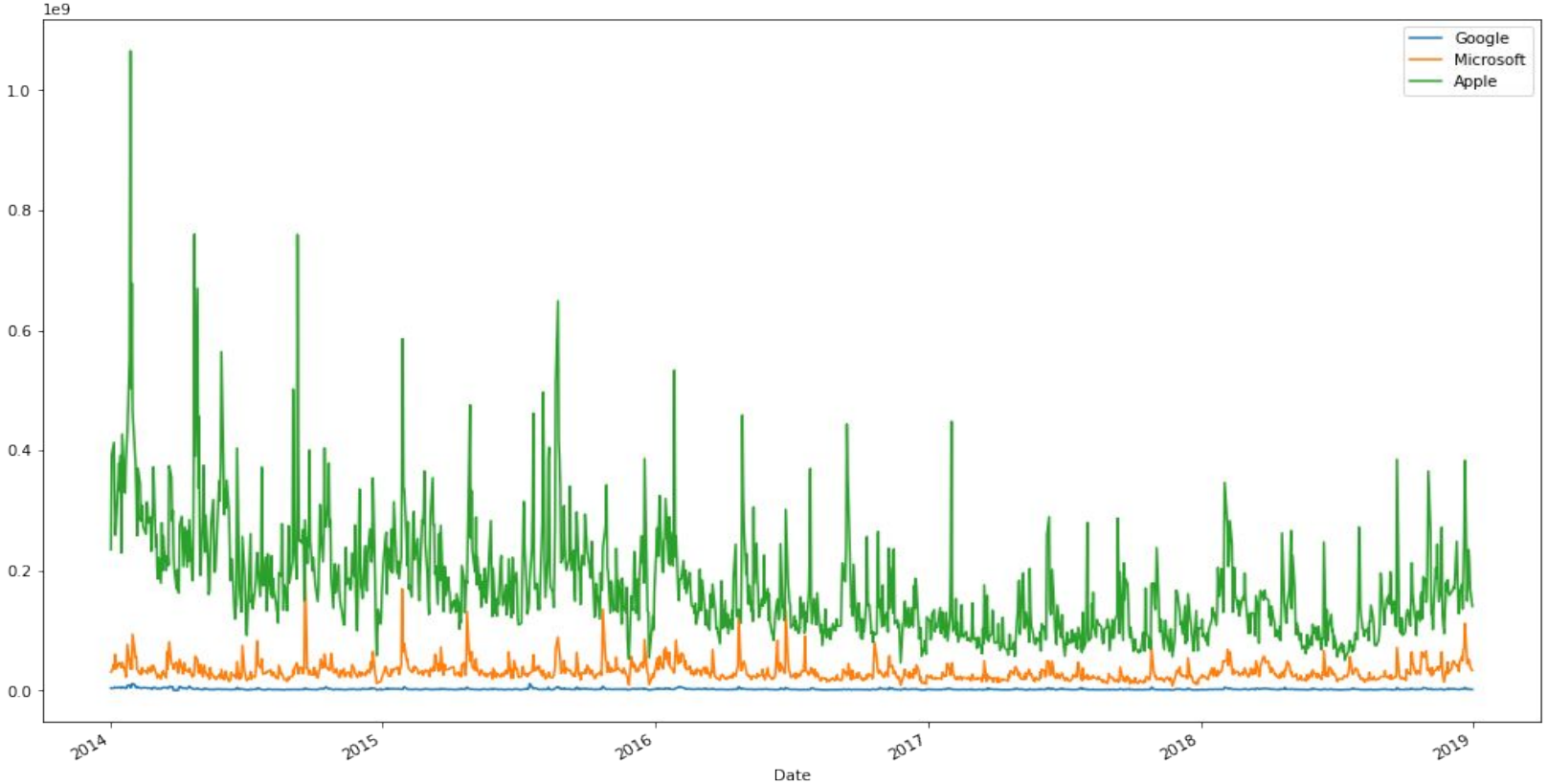
First Section

- Data Visualization and analysis using Python packages, Pandas, Matplotlib, Numpy
- Comparing three tech stock prices (Apple, Microsoft, Google) over 5 years from 2014 till the end of 2018 using plots
- Comparing performance, long term and short term returns, correlation, volatility of the above stocks and visualize them

Price change over 5 years

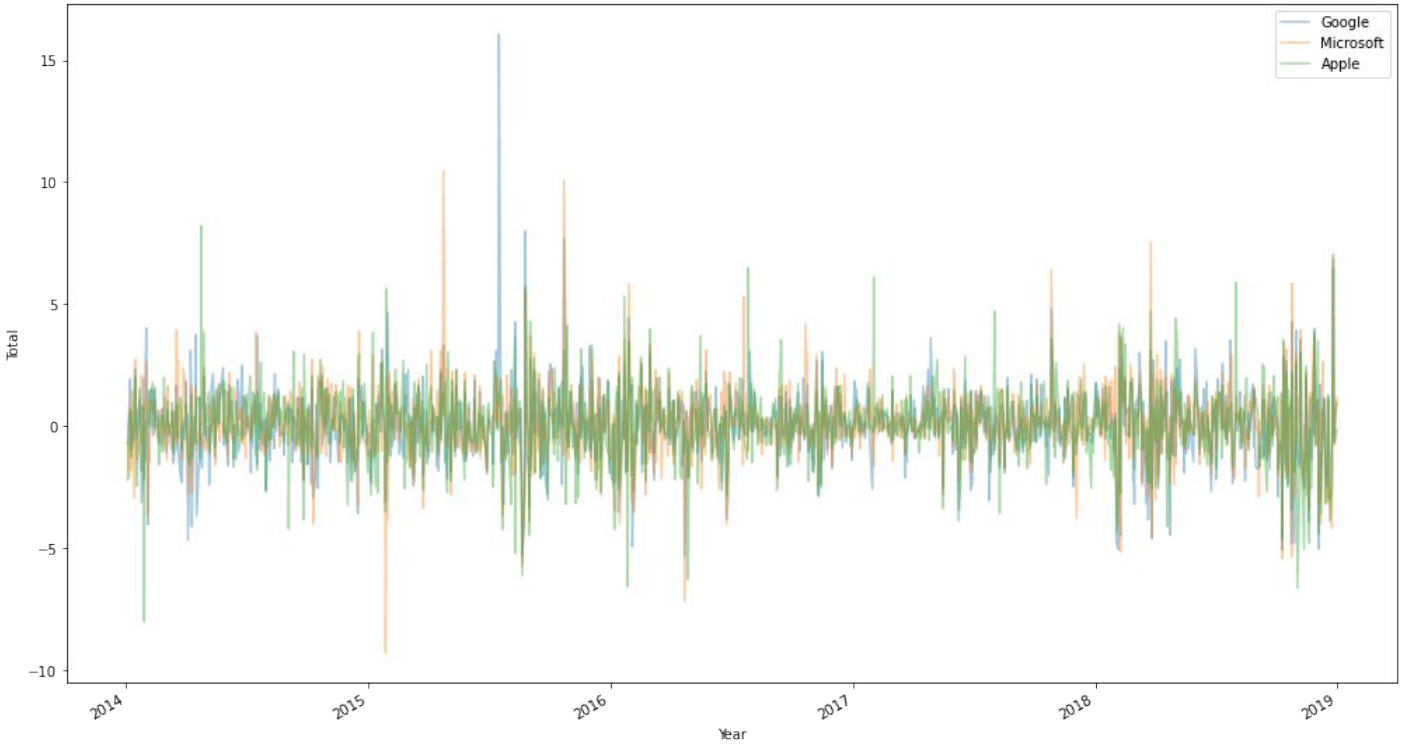


Volume Traded



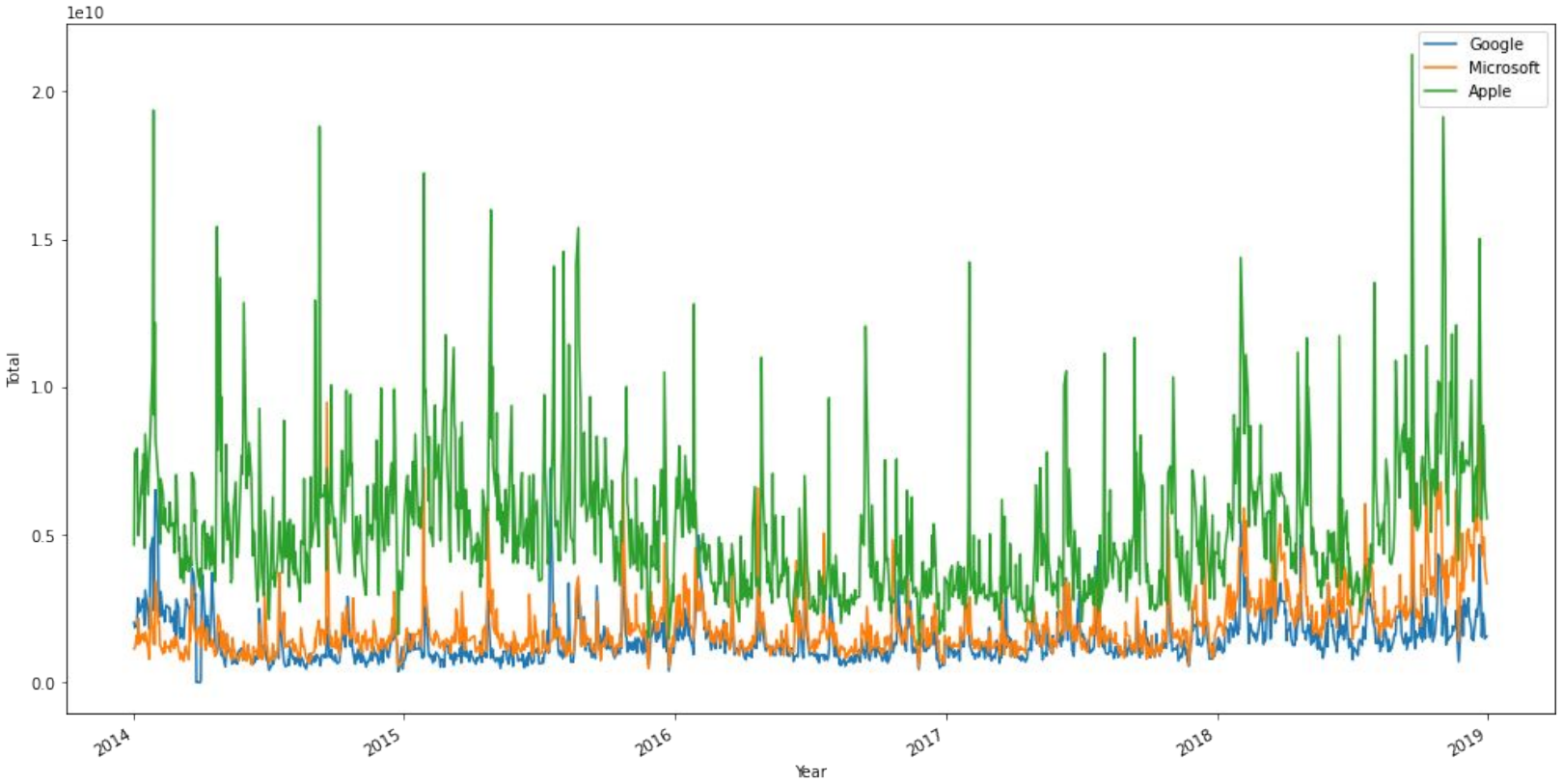
Apple consistently has the highest volume traded, over the 5 years

Daily percent change (short term profit/loss)



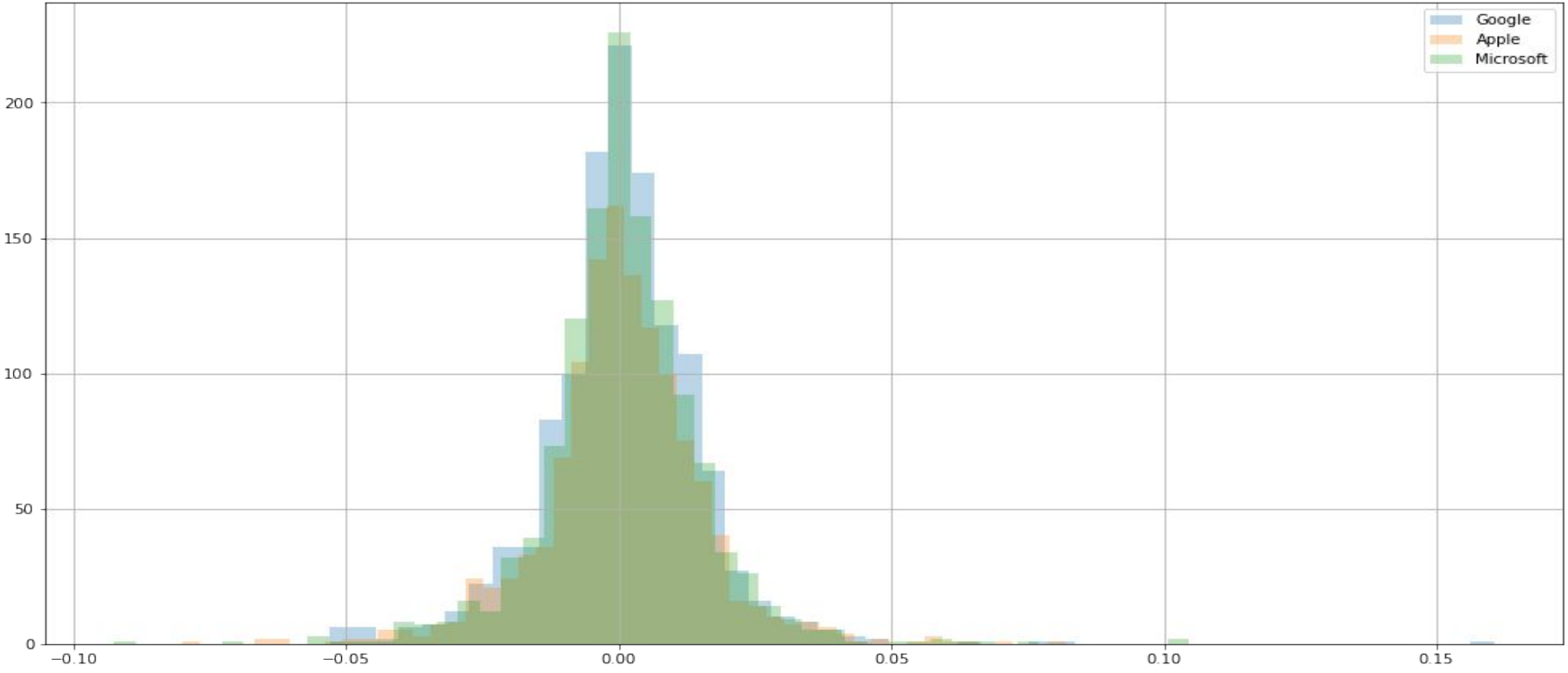
The plot show the percent change over the years for all stock are about the same

Total volume traded amount($)



Apple has the highest volume traded amount (dollar amount) consistently over 5 years

Volatility for all 3 stocks (short term investment)



The width of the chart shows the volatility of each stock, seems like all 3 stocks are equally volatile

Cumulative Return (long term investment over 5 years)

The highest return on the investment, belongs to Microsoft, Apple is slightly lower and Google is the lowest. That means, if we have invested $1 in at the beginning of 2014, our investment would have increased to $2.6 (for Microsoft), $2 (for Apple), $1.7(for Google)

Second Section: Preprocessing and Modeling

Type: Time series forecasting

Tools: Python Tensorflow, Statsmodel packages

Preprocessing and Modeling steps

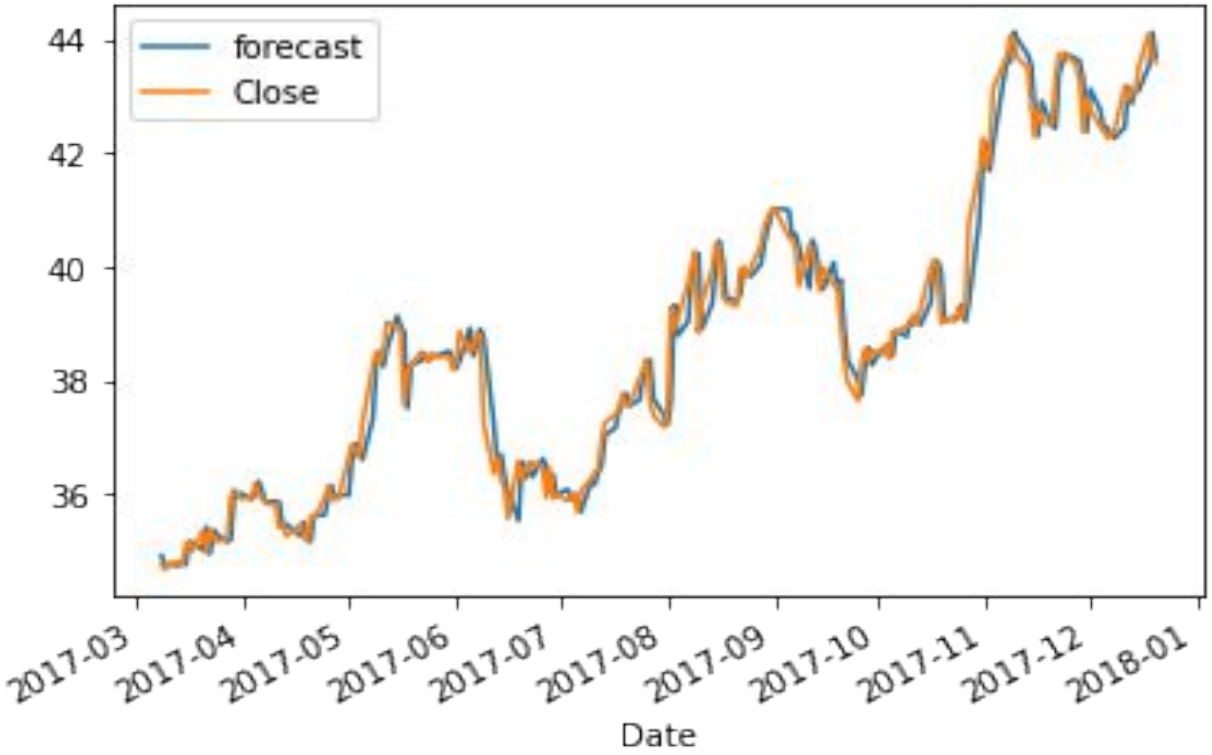
Building ARIMA Model:

1. Find out if the data is stationary
2. Find a way to make our data stationary (since it was not stationary) that is, by finding the returns of our time series data
3. Plot acf and pacf to find p and q for our model
4. Find d
5. Build our ARIMA model
6. Find the residuals to see the accuracy of our model
7. Visualize our model performance compare to the actual time series data

Building LSTM Model:

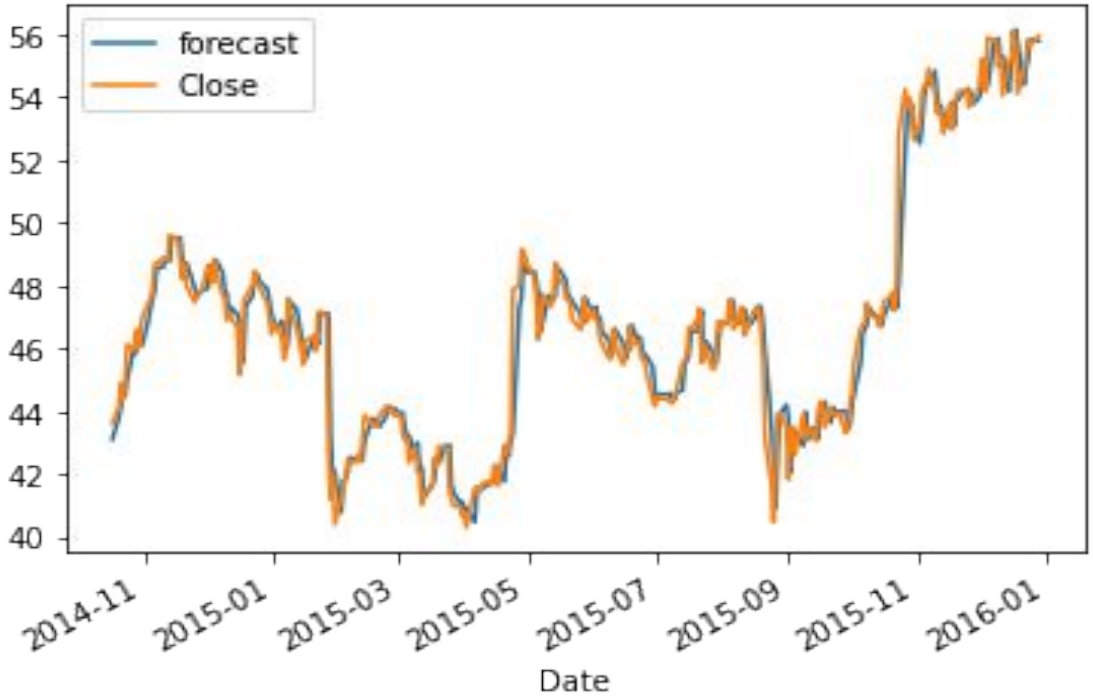
1. Scale the data for faster, better model
2. Split the data into test and train set (80% train and 20% test)
3. Reshape the data for our LSTM model as it needs to be 3 dimensional
4. Build the LSTM model
5. Fit the data to our model
6. Transform the data back to original
7. Find RMSE to compute the accuracy of our model
8. Visualize our model performance compare to the actual time series data

ARIMA Model Forecast for Apple stock



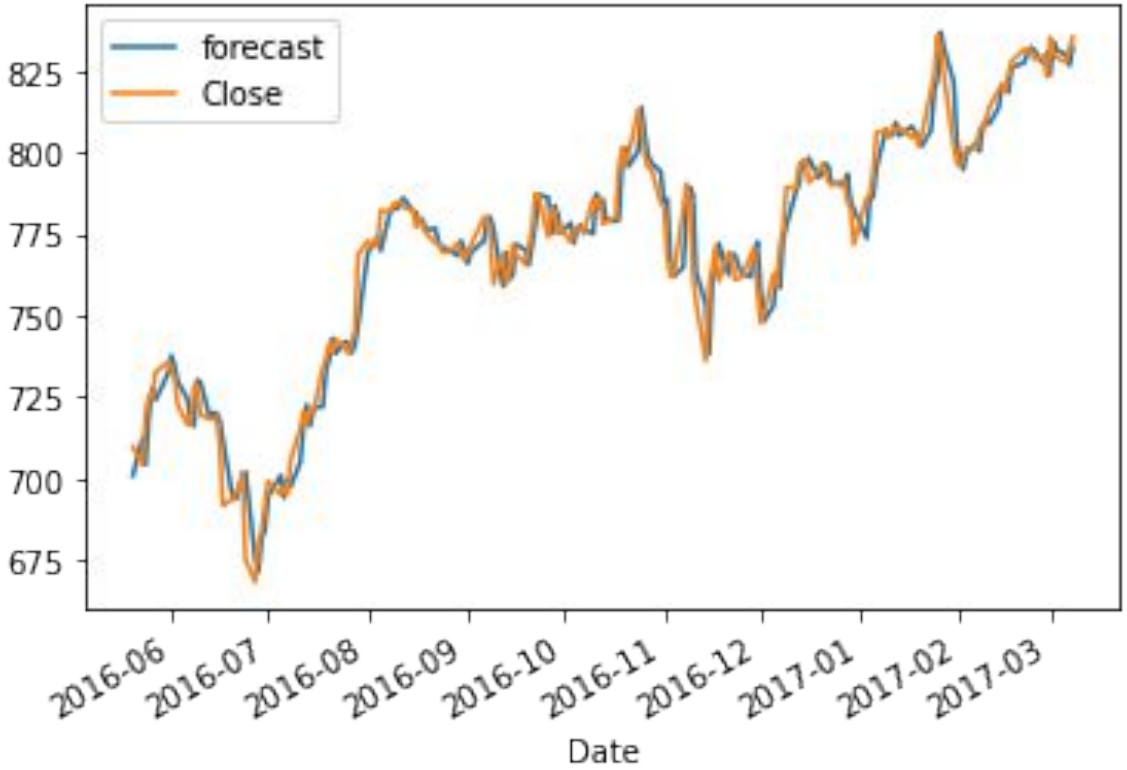
It can be seen that our model did a great job forecasting Apple stock for the above period, the blue line shows our model prediction and the orange line shows the actual price

ARIMA Model Forecast for Microsoft Stock



It can be seen that our model did a great job forecasting Microsoft stock for the above period as well, the blue line shows our model prediction and the orange line shows the actual price

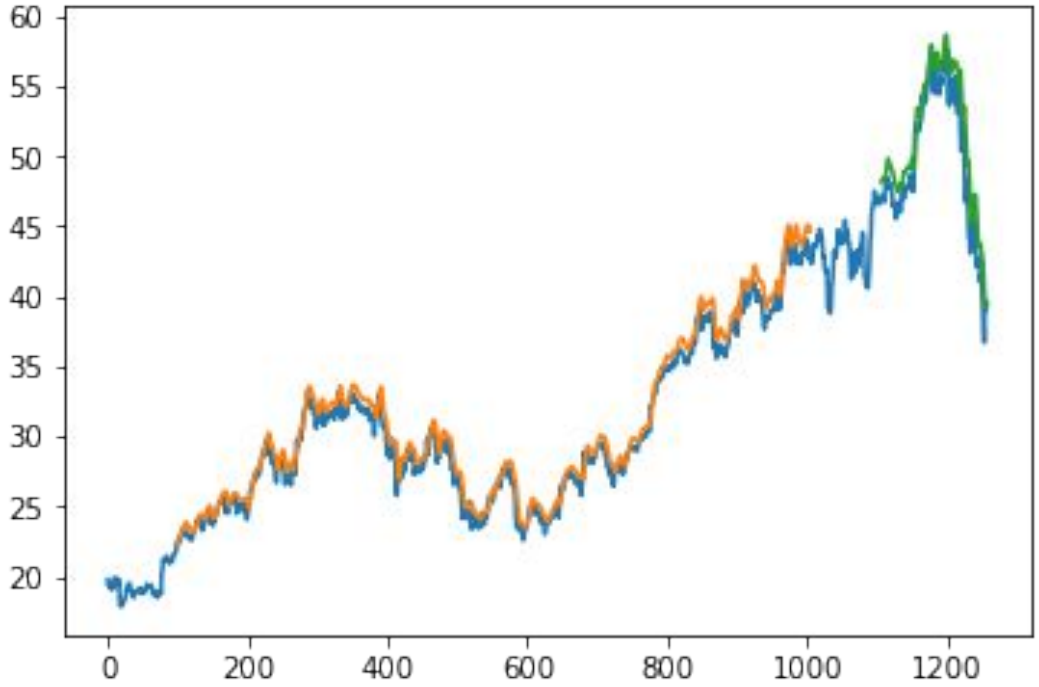
ARIMA Model Forecast for Google Stock



Just like Apple and Microsoft models, It can be seen that our model did a great job forecasting Microsoft stock for the above period as well, the blue line shows our model prediction and the orange line shows the actual price

LSTM Model Forecast for Apple Stock

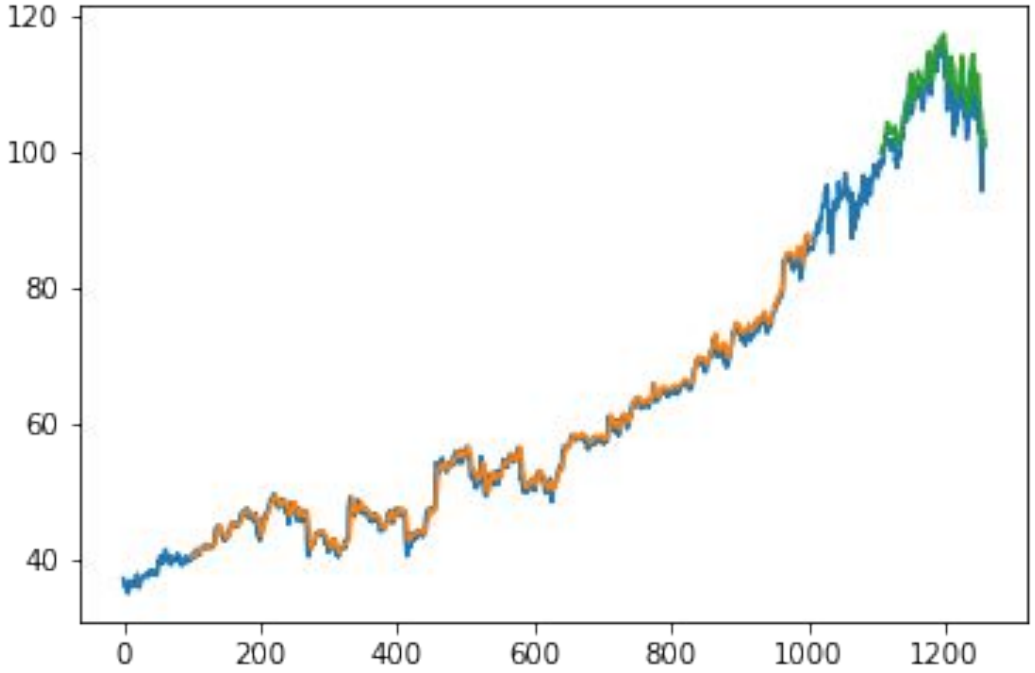
	Prediction	Actual
0	48.164997	47.145000
1	48.320568	46.974998
2	48.445816	46.875000
3	48.463799	46.717499
4	48.396873	47.560001
5	48.269321	47.957500
6	48.447556	48.327499
7	48.842003	48.494999
8	49.303482	48.365002
9	49.676147	47.924999



It can be seen that the LSTM model perfectly predicted Apple prices, both on the training set and test set. The left one shows the first 10 prediction for the Green line

LSTM Model Forecast for Microsoft Stock

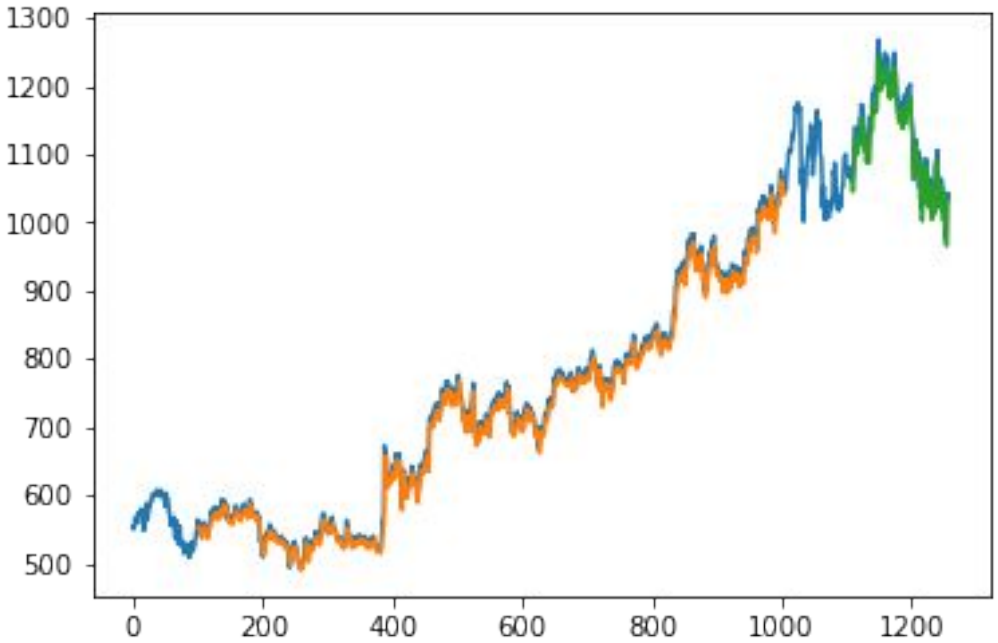
	Prediction	Actual
0	98.376732	98.360001
1	98.741219	98.010002
2	99.083641	98.949997
3	99.294876	98.839996
4	99.521927	100.790001
5	99.728477	101.669998
6	100.175629	102.190002
7	100.857483	102.489998
8	101.641754	100.879997
9	102.378960	101.629997



It can be seen that the LSTM model perfectly predicted Google prices, both on the training set and test set. The left one shows the first 10 prediction for the Green line

LSTM Model Forecast for Google Stock

	Prediction	Actual
0	1088.934937	1075.660034
1	1090.602539	1060.319946
2	1086.656982	1067.800049
3	1071.788574	1084.989990
4	1074.087891	1119.500000
5	1091.962891	1139.290039
6	1126.080078	1139.660034
7	1149.166016	1136.880005
8	1148.733521	1123.859985
9	1140.483032	1120.869995



It can be seen that the LSTM model perfectly predicted Apple prices, both on the training set and test set. The left one shows the first 10 prediction for the Green line

Conclusion

Predicting the stock market is a challenging task due to consistently changing stock values which are dependent on many different parameters which can form complex patterns. The historical dataset that was used from ‘Yahoo Finance’ consists of only few features like high, low, open, close, adj close value of stock prices, volume of shares traded etc., which may not be sufficient to build a perfect model.

To obtain the accuracy of our two models, we can see that for the ARIMA model the residual variance stays around zero, which means the forecast for this model is pretty good, also, after visualizing the plot, we can see that the model fits very well to our data. For the LSTM model, we measure RMSE and that gave us great score, after visualizing the data, it is confirmed that the model fits really well on the data as well. If we were to pick the best model, since the RMSE for the ARIMA model gives us a lower score, we can conclude that ARIMA model has higher accuracy score compare to LSTM model, thus, it is our best model. For future work, deep learning models could be developed by collecting more data such as, financial news articles, along with financial parameters such as profit and loss statements etc., for possibly better results.