Stock Data Analysis

Introduction

This project is aimed at analyzing historical stock data to predict and forecast the pattern of stock in the future. Forecasting is the process of predicting the future values based on historical data and analyzing the trend of current data. By running simulations of future states based on present states, we can foresee the trend of stock market. To be specific, I chose to analyze stock data pertaining to car manufacturers such as Tesla, BMW, Ford, General Motors, Hyundai and Honda.

Problem Statement

In this project, we attempted to predict stocks using Linear Regression and KNN to improve the recommendation on choices of actions the user shall take in stock trading.

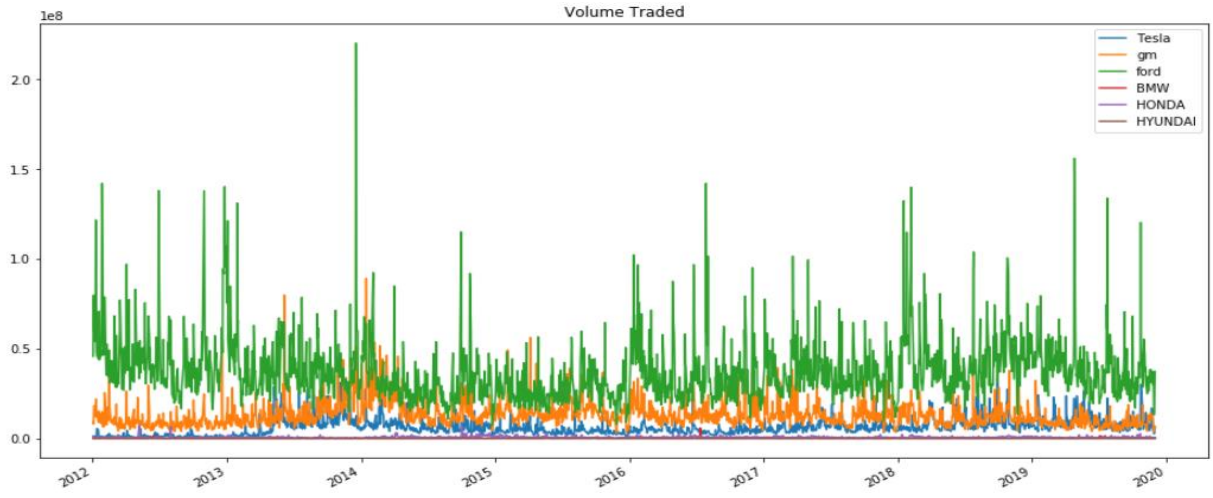
Data Collection

Data collection is done by extracting latest stocks data using pandas web-data reader and Yahoo Finance. The daily stocks for the six car manufacturers for the time in between 1-1-2012 and 12-1-2019 are being collected. The data consists of 1992 rows and 6 columns for each manufacturer.

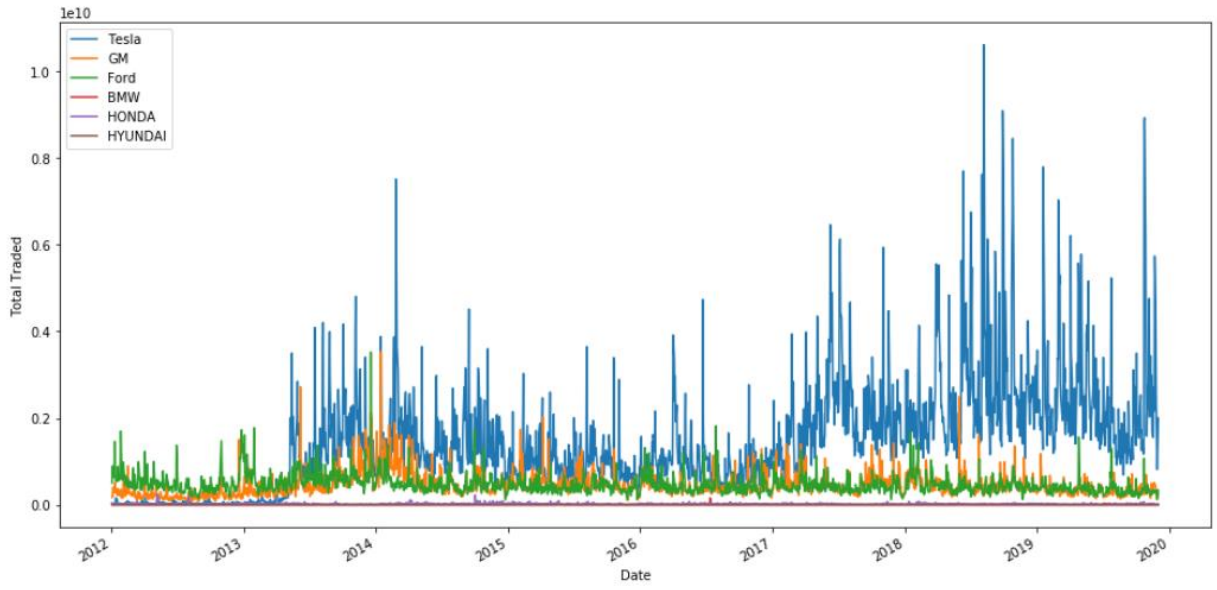
	High	Low	Open	Close	Volume	Adj Close
Date						
2012-01-03	24.120001	23.750000	23.750000	23.879999	102000.0	17.555784
2012-01-04	24.250000	24.010000	24.110001	24.129999	42500.0	17.739580
2012-01-05	24.100000	23.790001	24.010000	24.040001	24400.0	17.673412
2012-01-06	23.840000	23.320000	23.840000	23.510000	14400.0	17.283775
2012-01-09	24.309999	23.930000	24.280001	24.290001	32700.0	17.857206

Data Analytics

The plot of the open prices for all the manufacturers depict that Tesla has the highest stock on 06-26-2017 at 386.69.



The plot of the Volume Traded for all the manufacturers depict that Ford has the highest volume of stock traded on 12-18-2013 at 220363000.0



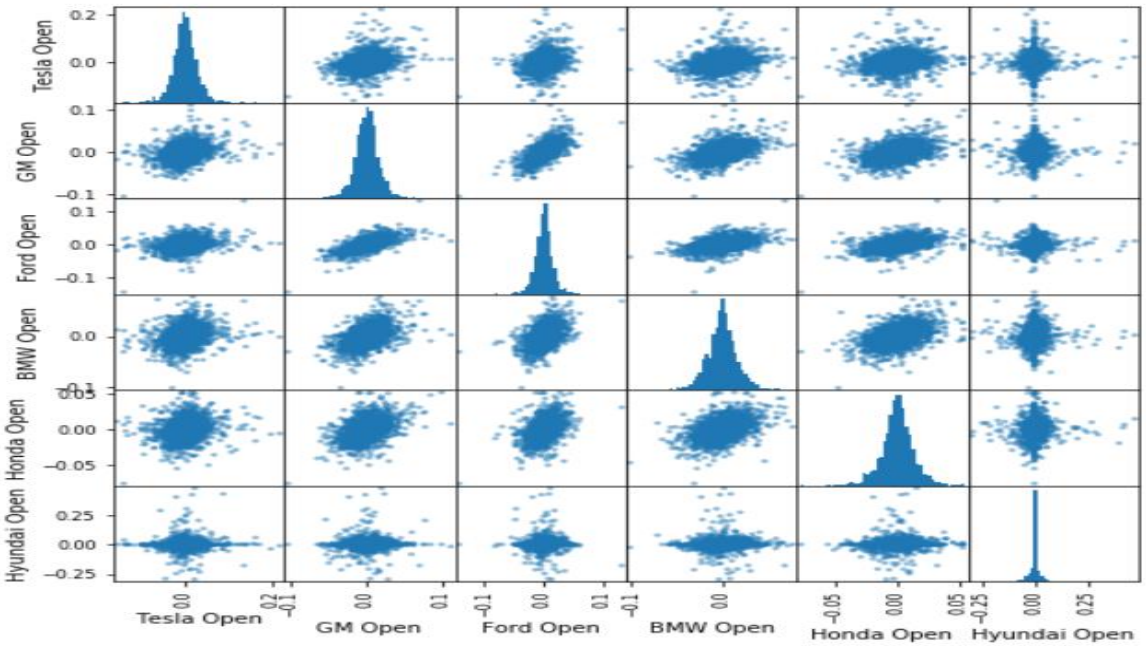
The plot of the Total Trade for all the manufacturers depict that Tesla has the highest total of stock traded on 08-07-2018 at 10616334958.92 .

Grouping up stocks into single data frame to check for correlation amongst the manufacturers.

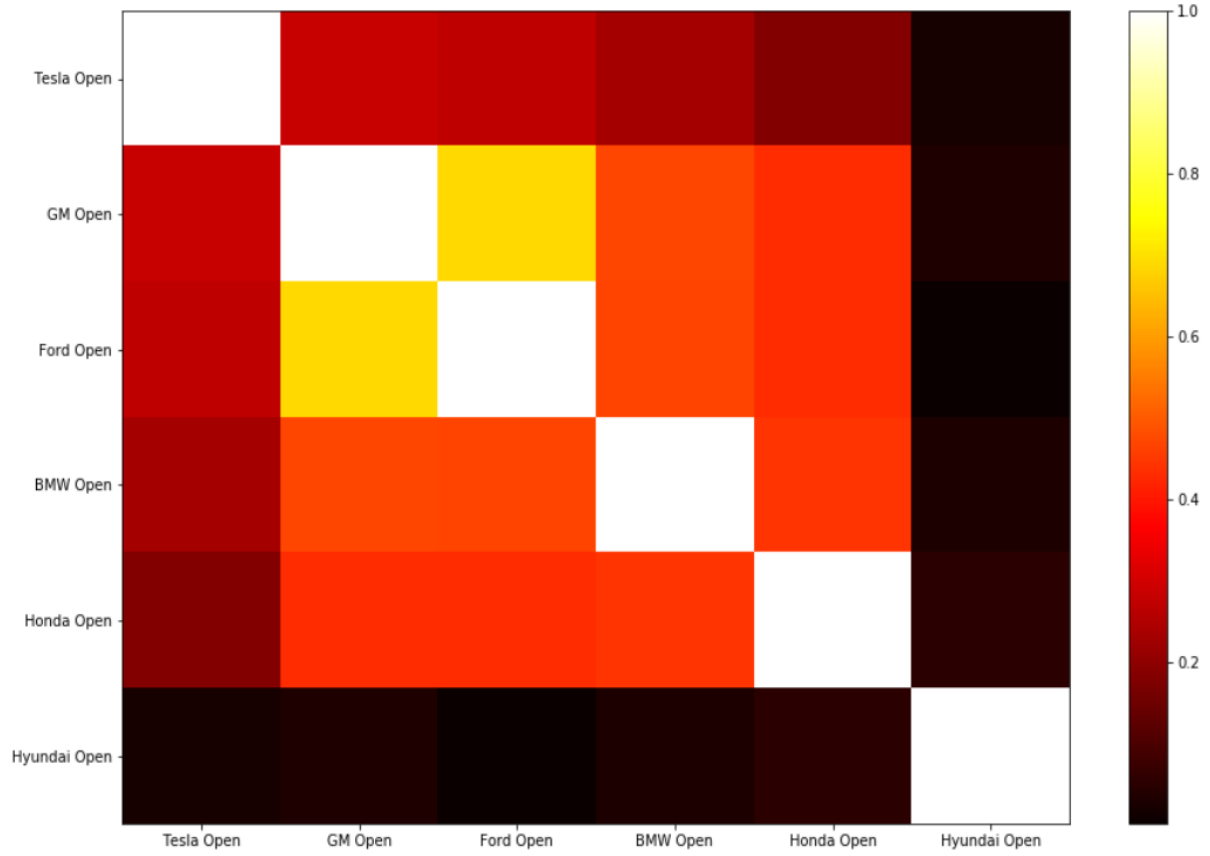
	Tesla Open	GM Open	Ford Open	BMW Open	Honda Open	Hyundai Open
Date						
2012-01-03	28.940001	20.830000	11.00	23.750000	31.180000	25.250000
2012-01-04	28.209999	21.049999	11.15	24.110001	31.650000	27.799999
2012-01-05	27.760000	21.100000	11.33	24.010000	31.650000	28.350000
2012-01-06	27.200001	22.260000	11.74	23.840000	31.990000	28.500000
2012-01-09	27.000000	23.200001	11.83	24.280001	32.099998	28.750000

Normalizing the stocks by calculating the percentage change in the stock by day.

	Tesla Open	GM Open	Ford Open	BMW Open	Honda Open	Hyundai Open
Date						
2012-01-03	NaN	NaN	NaN	NaN	NaN	NaN
2012-01-04	-0.025225	0.010562	0.013636	0.015158	0.015074	0.100990
2012-01-05	-0.015952	0.002375	0.016144	-0.004148	0.000000	0.019784
2012-01-06	-0.020173	0.054976	0.036187	-0.007080	0.010743	0.005291
2012-01-09	-0.007353	0.042228	0.007666	0.018456	0.003439	0.008772



The scatterplot depicts that there is some kind of relationship between General Motors and Ford. But it is not accurate enough to find through this plot.



So, we use heat map instead to find the relation. The heat map gives significant relation between Ford and General Motors than any other manufacturers.

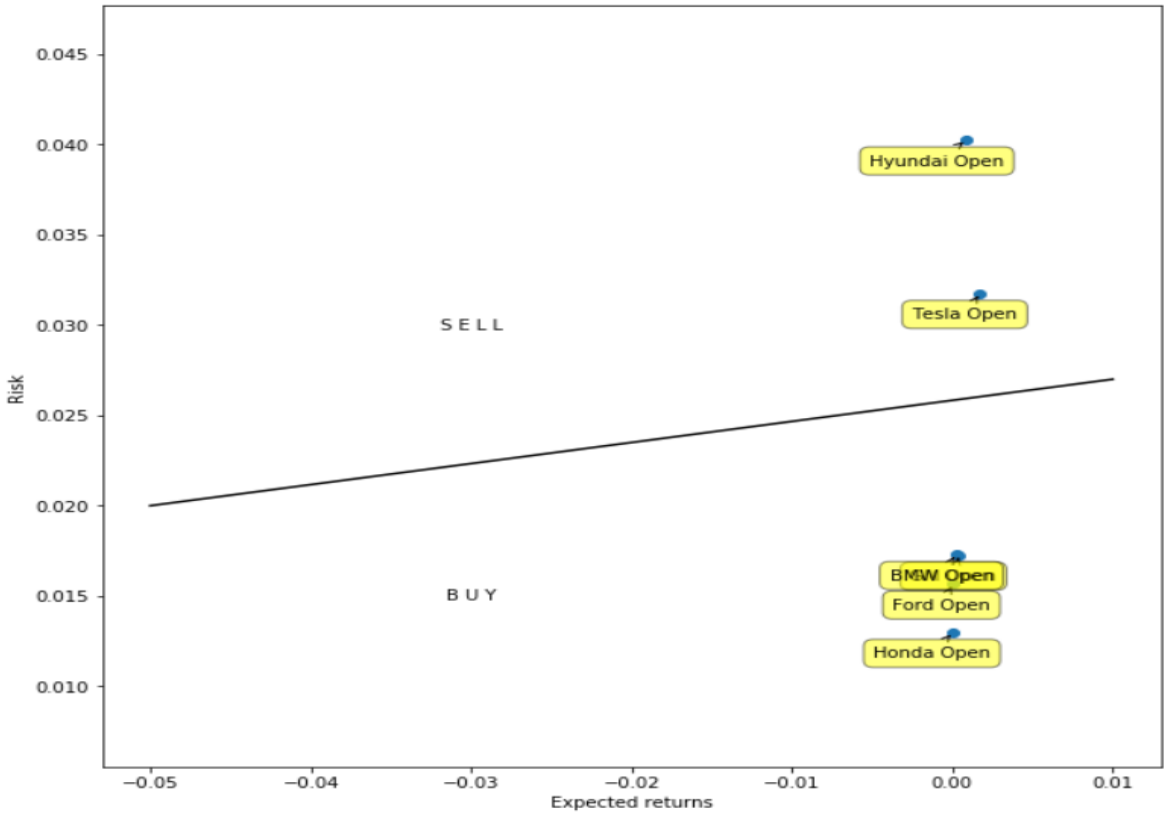
Hypothesis

Hypothesis: The forecasting of Tesla stock by the end of December 2019 is on a rise.

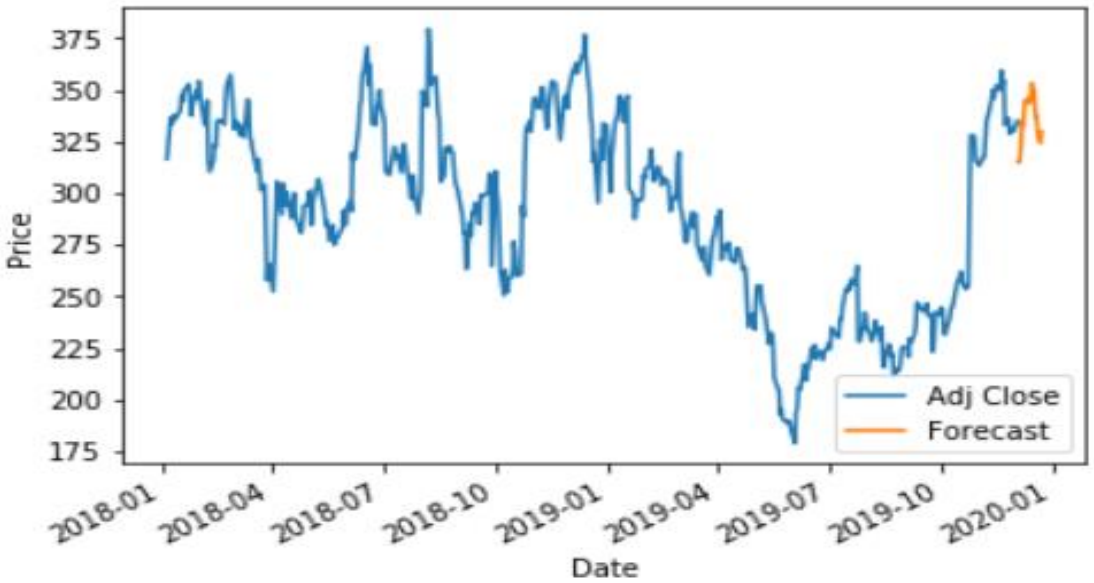
Null Hypothesis: These is no change in Tesla's stock by the end of December 2019.

Evaluating risks and Returns

The image below depicts the risk and returns plot where we differentiate the manufacturers based on how risky it is to buy the stock and best returns to expect from. We draw a line on the plot, where we buy the stock if it is below the line as it is less risky and has good returns and sell the stock above the line which is too risky and not good returns.



To predict the trend of stock based on historic data, we use Linear regression and KNN. Out of the two, regression shows better accuracy of 92% compared to that of the 88% with KNN. The final plot shows the stock forecast for Tesla. The forecast depicts rise for mid December 2019 and huge fall at the end of December 2019.



Hence our hypothesis of forecasting a rise in the stock of Tesla is evident, following with the Fall.

References

1. https://towardsdatascience.com/in-12-minutes-stocks-analysis-with-pandas-and-scikit-learn-a8d8a7b50ee7
2. https://home.cse.ust.hk/~yqsong/teaching/comp3211/projects/2017Fall/G26.pdf
3. https://www.analyticsvidhya.com/blog/2018/10/predicting-stock-price-machine-learningnd-deep-learning-techniques-python/
4. http://www.cse.scu.edu/~mwang2/projects/Predict_stockMarket_16w.pdf
5. https://www.slideshare.net/anilsth91/stock-market-analysis-and-prediction