Capstone Three Report

1. Problem statement

Predict SP500_Index Future Price
Method: Time Series
Feature Engineering
Lags and Moving Average
Target Variable: Adj Close

2. Data preprocessing & feature engineering

Stock price data from Yahoo.com, don't have missing values and duplicates

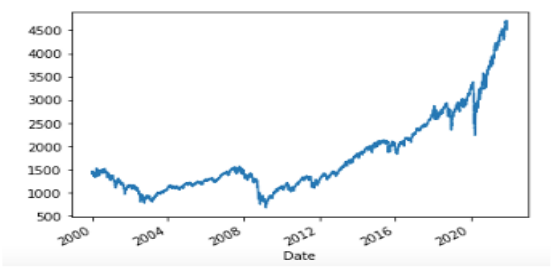
3. EDA

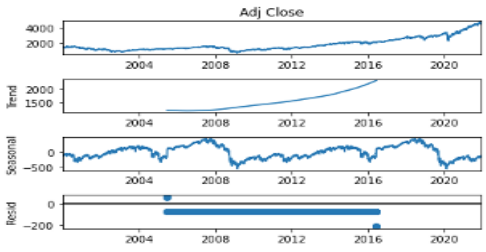
Histogram, data left skewed
Correlation matrix: moving average, open, high, low, standard deviation are high correlated, finally decided to use time, volume and moving average as independent variables

4. Apply Time Series on target variable: adj close price

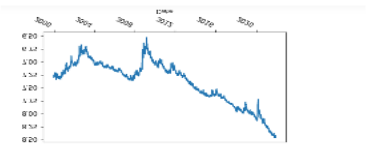
Original plot for adj close has increase trend and seasonality
After log transform and difference, I get a rather stationary plot of adj close, which means I get rid of the auto-correlation and independence issue for linear regression, also using Dicky Fuller test, I get the p-value less than 0.05, so I can apply my data for models

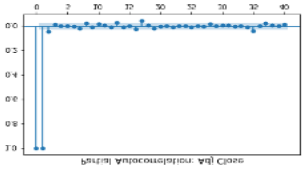
(original)

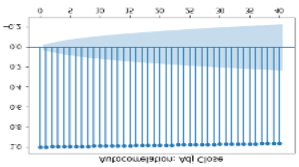




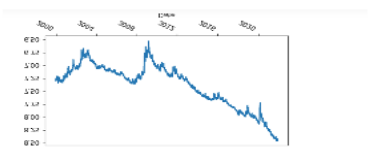
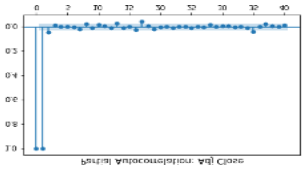
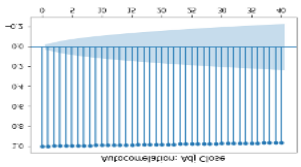
(After log transform)







(After difference)

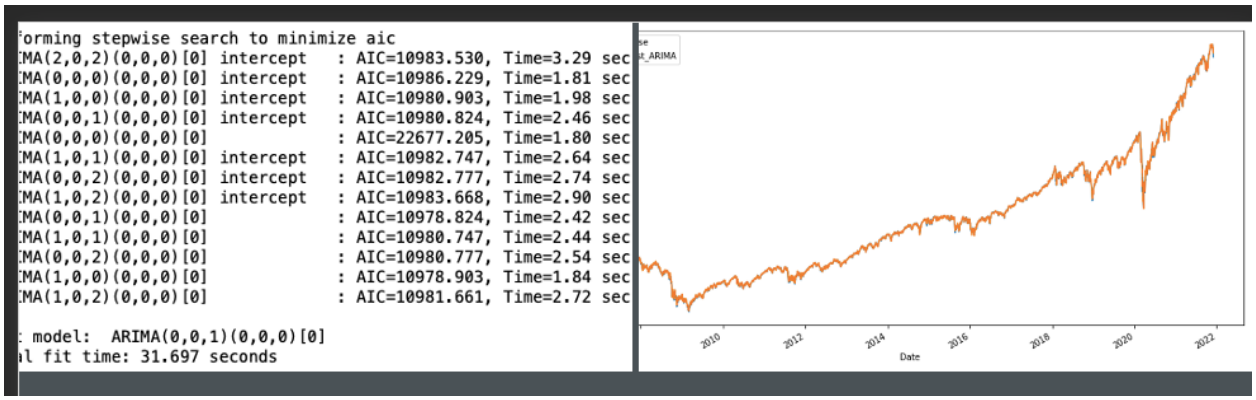




5. Time Series

Find out the best p, d, q for time series model

Evaluation Metrix: MAE: 5.56, RMSE: 8.65, MAPE: 0.003

Time Series Plot



6. Conclusion: SP500 Index will keep increasing in the future