Coursera Specialized Models: Time Series and Survival Analysis

NAME: LIM CHIN EN

COMPLETION DATE: 21ST FEB 2022

Data Selected

In this assignment, the stock price of Apple (Ticker Symbol: AAPL) has been selected as the time series data. Time series analysis will be applied on the stock price of Apple, and we will evaluate various techniques to analyze as well as predicts the stock price of Apple.

We can use the yfinance library to download the daily stock price of AAPL from 1st Jan 2021 to 9th Feb 2022 directly onto Python. (Please find the notebook for more detail)

***Disclaimer: All investment strategies and investments involve risk of loss. Nothing in this assignment constitutes professional and/or financial advice.**

Data Selected

```
#see your data
stock.tail(5)
```

Date	Open	High	Low	Close	Volume	Dividends	Stock Splits
2022-02-02	174.527647	175.656214	173.109456	175.616257	84914300	0.00	0
2022-02-03	174.257984	176.015754	171.900986	172.679993	89418100	0.00	0
2022-02-04	171.679993	174.100006	170.679993	172.389999	82391400	0.22	0
2022-02-07	172.860001	173.949997	170.949997	171.660004	77251200	0.00	0
2022-02-08	171.729996	175.350006	171.429993	174.830002	74829200	0.00	0

In this assignment, the column of concern will be the ‘Close’ column.

Data Selected

Data Analysis

Before conduct any predictive analysis on the data, lets perform a data analysis on the time series to understand the data.

ANALYSIS 1: VISUAL INSPECTION

- We can see that the stock price of Apple has a clear uptrend.
- The variance looks pretty consistent at first glance, we will confirm this quantitatively later.
- There is seasonality in stock price movement. This is because:
  - ❖Sales of the underlying company may fluctuate according to holiday season, earning report release or macroeconomic factor.
  - ❖According to statistic, stock prices are more likely to fall on Monday and rise on Friday.

  However, the seasonality of stock price is very difficult to spot. The magnitude and the period of the seasonality varies across the timeline.

Due to the presence of an uptrend, it seems like this is not a stationary series. Let's confirm this by inspect the times series quantitatively

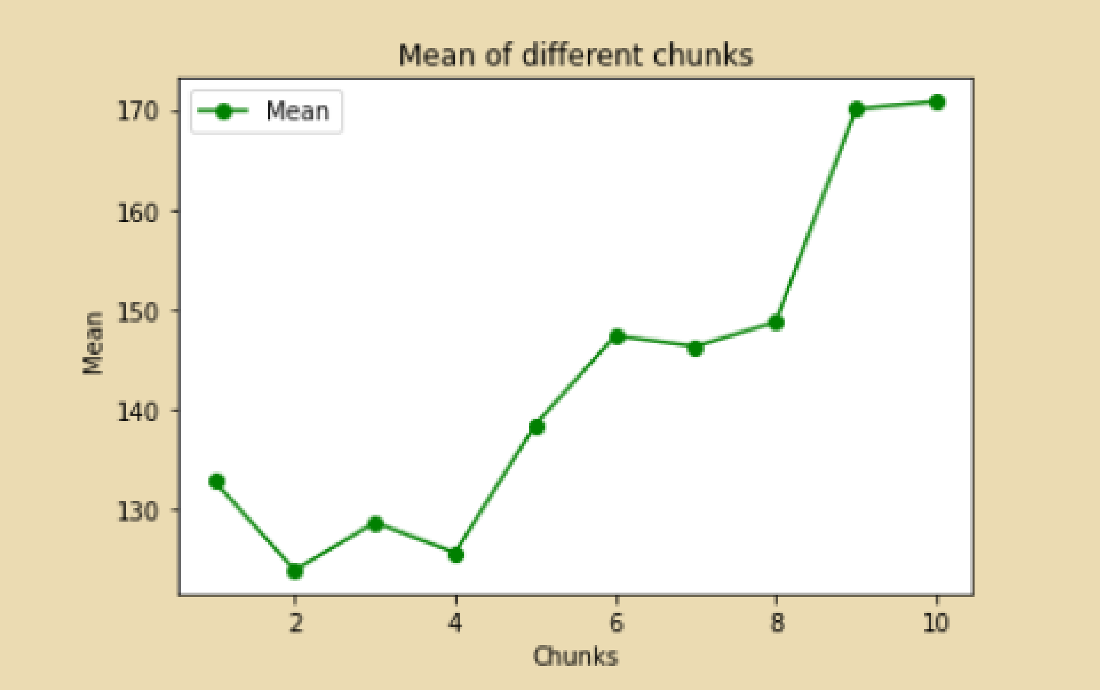
Data Analysis

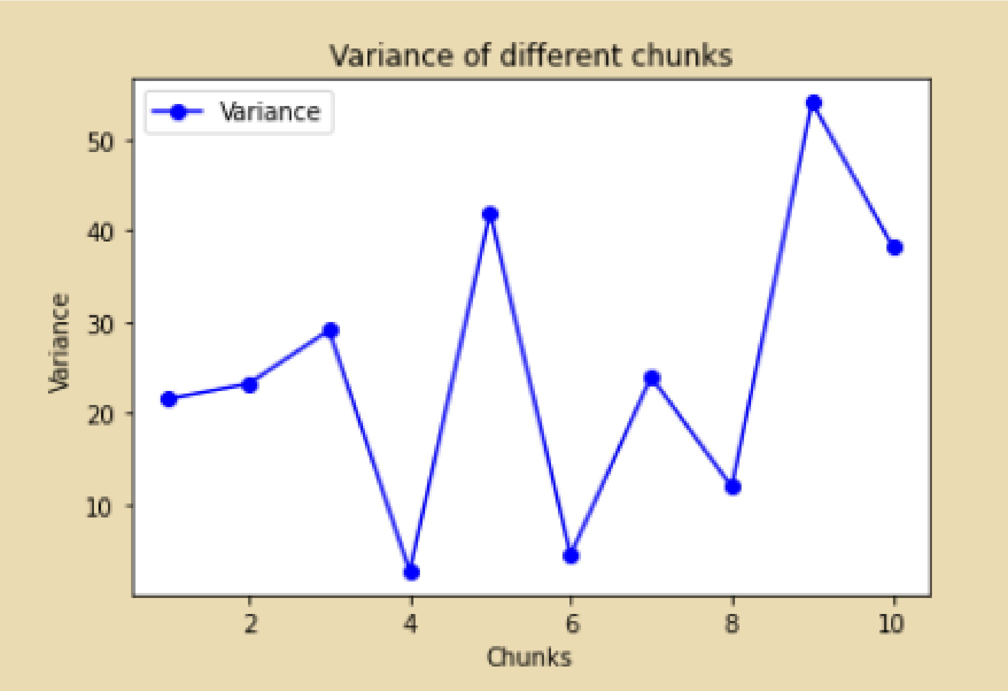
ANALYSIS 2: Summary Statistics & Plots

We are going to split the stock price into 10 chucks and inspect the mean and variance of each chunks.

From the 2 graphs below, we can see that the variance and mean fluctuate significantly across different chunks. The time series is not stationary

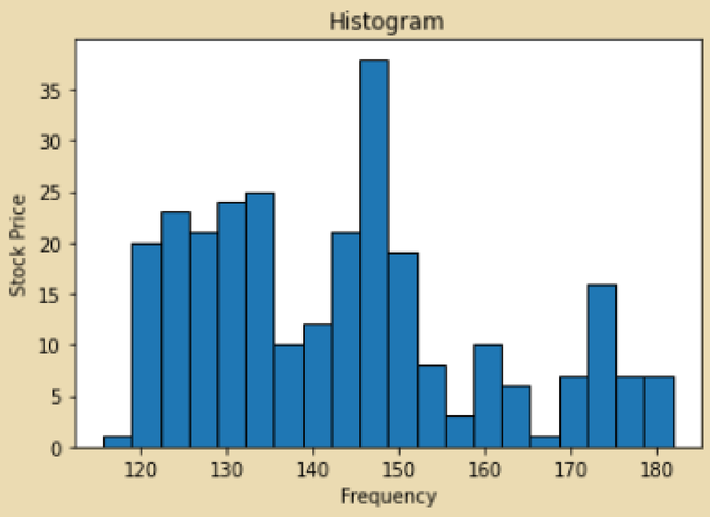
Data Analysis





Data Analysis

```
plt.hist(df_stock, edgecolor = "black", bins = 20)
plt.xlabel('Frequency')
plt.ylabel('Stock Price')
plt.title('Histogram')
plt.show()
```



ANALYSIS 3: Histogram

It is not normally distributed, which further confirms that this is a non-stationary data

Data Analysis

ANALYSIS 4: Augmented Dicky Fuller Test

```
In [7]:  from statsmodels.tsa.stattools import adfuller


         adf, pvalue, usedlag, nobs, critical_values, icbest = adfuller(df_stock, regression='c')
         print("ADF: ", adf)
         print("p-value:", pvalue)

         ADF:  -0.4894364343474657
         p-value: 0.894061472825354
```

The p-value is larger than 0.05 which concludes that this is not a stationary data.

Predictive Analysis

Next, we will try to build forecast model of AAPL stock price. We will establish the model using 3 different approach and investigate the accuracy of the forecast of each model:

- Triple Exponential Smoothing
- SARIMA Model
- LSTM Deep Learning Model

Triple Exponential Smoothing

Triple exponential smoothing can capture the trend & seasonality, which is demonstrated by the AAPL stock price. Let's check the performance of triple exponential smoothing on this time series.

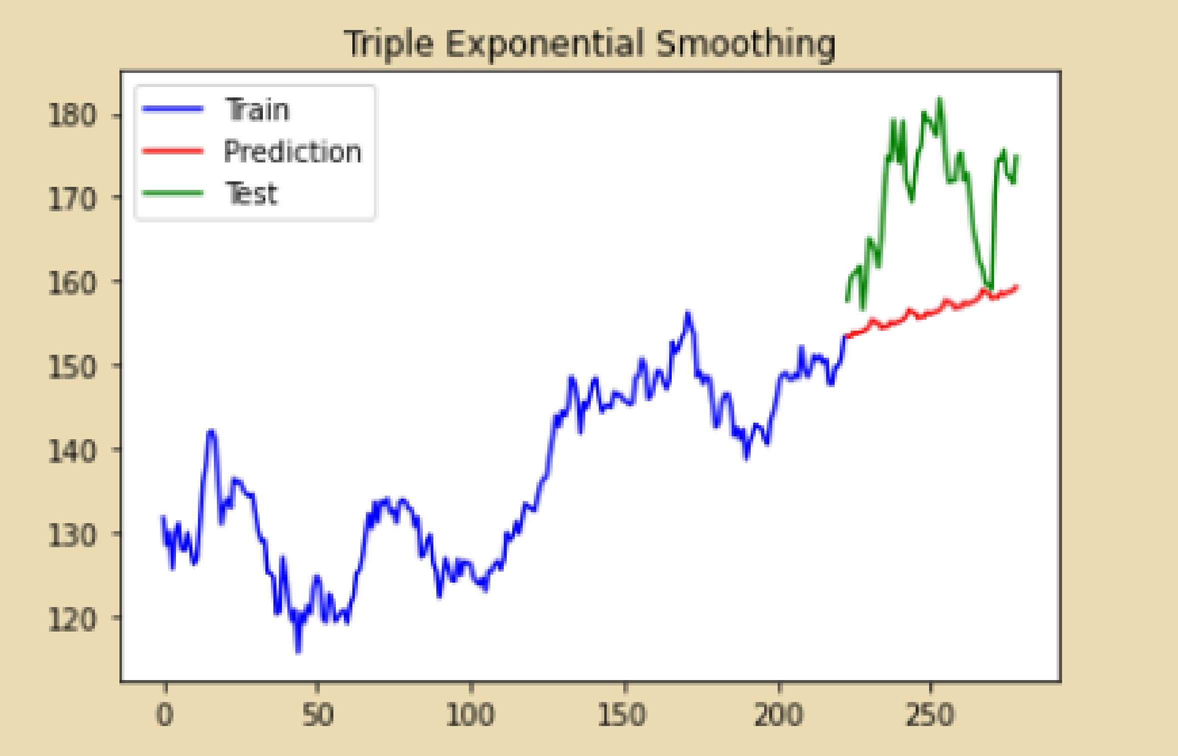
The time series have a total of 279 instances. We will take 20% (56) instances as test data and 80% (223) of the instances as train data.

It seems like the magnitudes of the seasonal values are independent of trend. So, this will be an additive model.

Unfortunately, it is very hard to judge the seasonal period of the stock price as the seasonal period is not consistent across the time series. We assume a seasonal period of 12 days first.

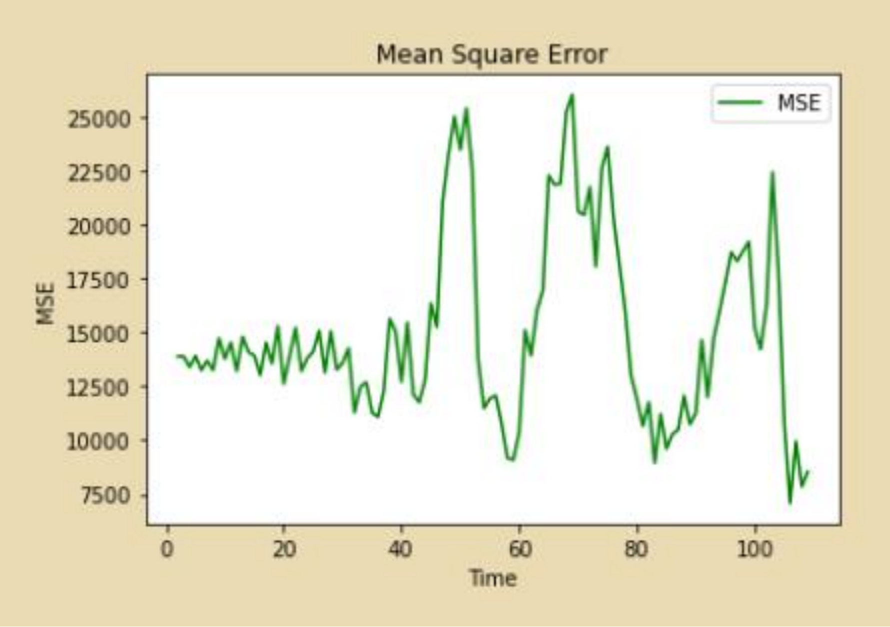
Triple Exponential Smoothing

The graph besides shows that the model fits poorly to the time series as it cannot capture the volatility of the stock well enough.



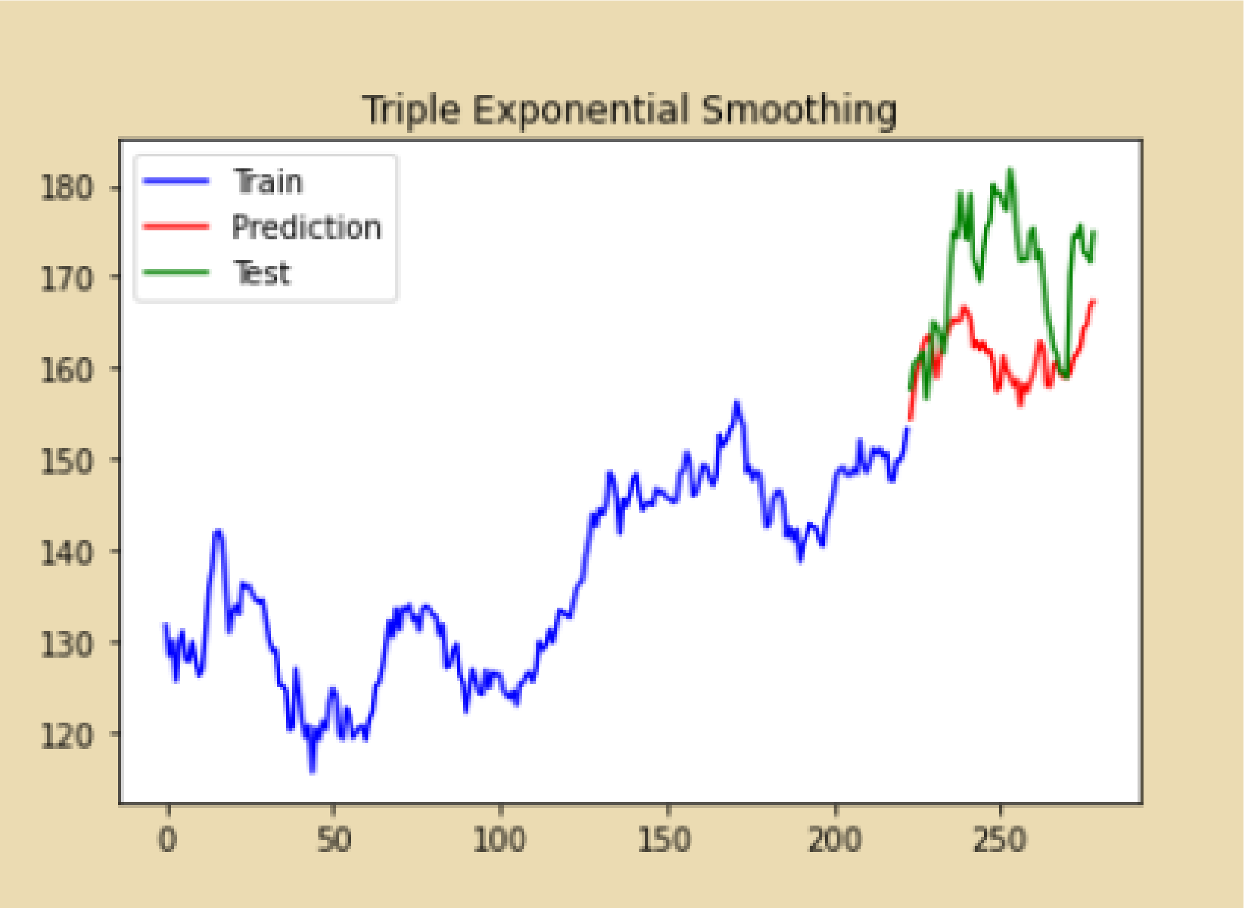
Triple Exponential Smoothing

To find the optimum seasonal period, we perform a grid search to identify the seasonal period that gives the lowest mean square error (MSE). Looks like 106 is the optimum seasonal period.



Triple Exponential Smoothing

The graph for seasonal period = 106 is shown at right. The prediction made is not that close to the test data. However at least the prediction has the “fluctuation” that typically seen in stock price movement.



SARIMA Model

SARIMA is short for **Seasonal ARIMA**, an extension of ARIMA models to address seasonality.

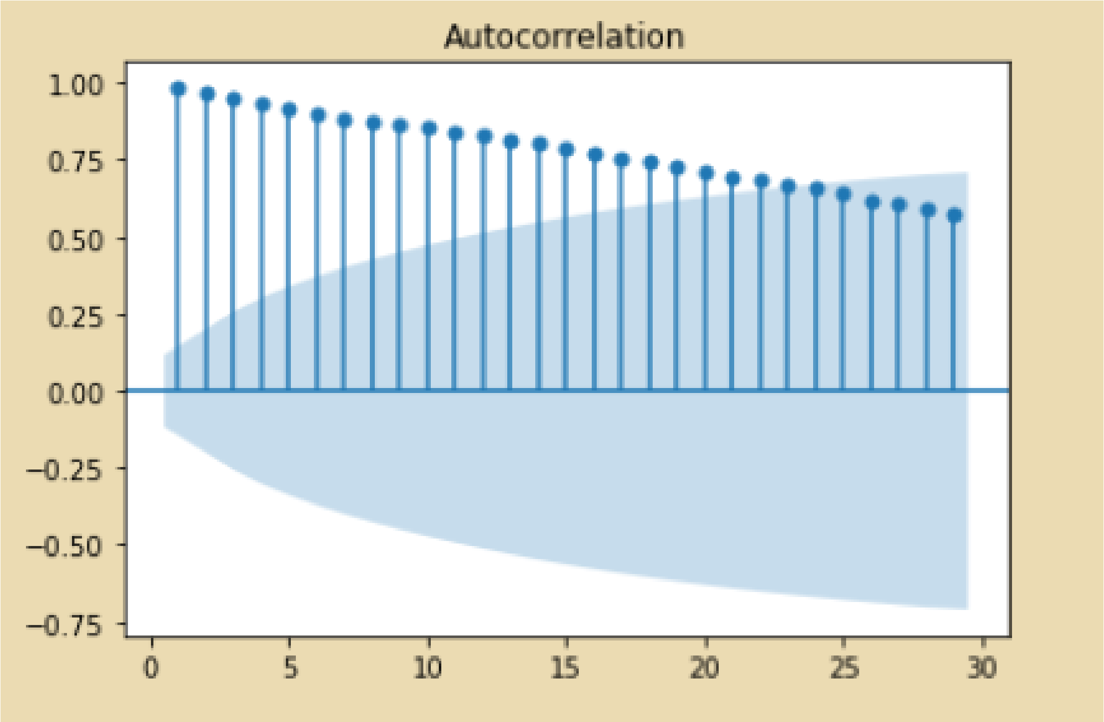
This model is used to remove seasonal components.

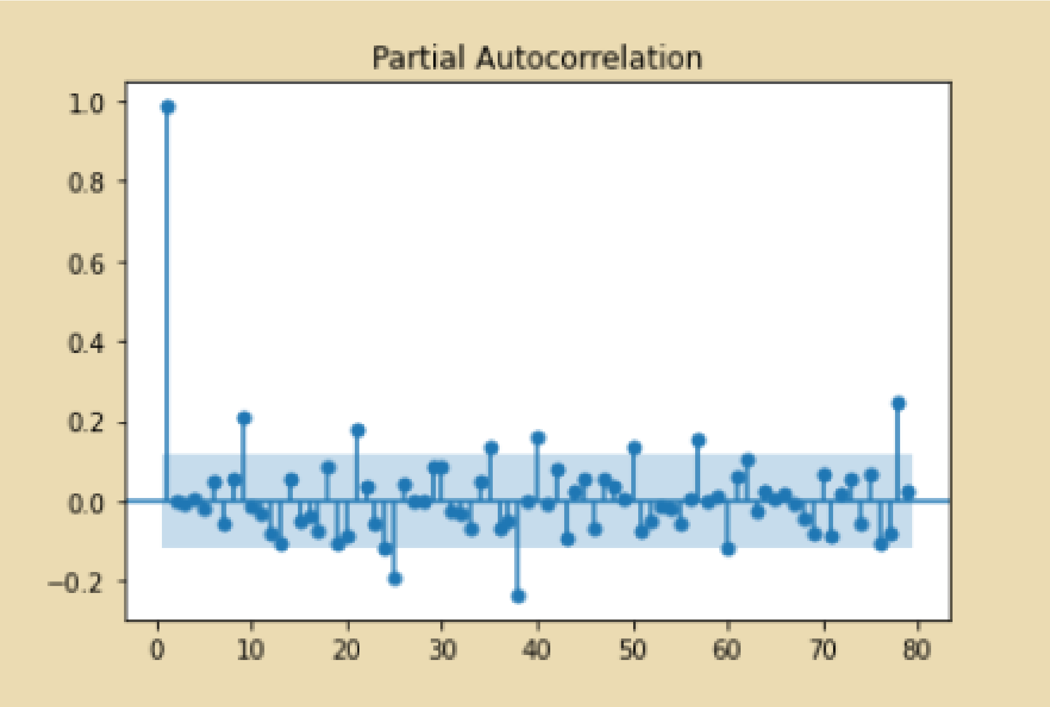
- The SARIMA model is denoted **SARIMA (p, d, q) (P, D, Q)**.
- **P, D, Q** represent the same as p, d, q but they are applied across a season.

First, we need to determine all the parameters required for training

SARIMA Model

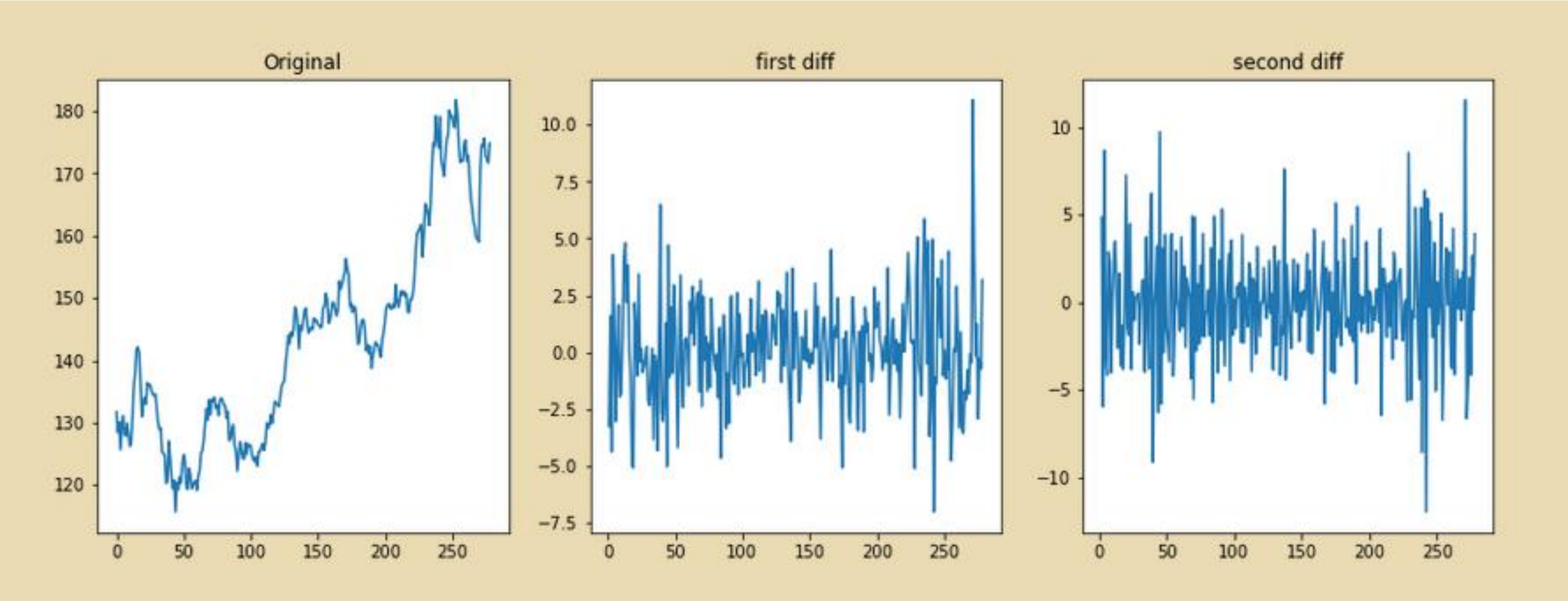
Plotting autocorrelation and partial correlation plots.





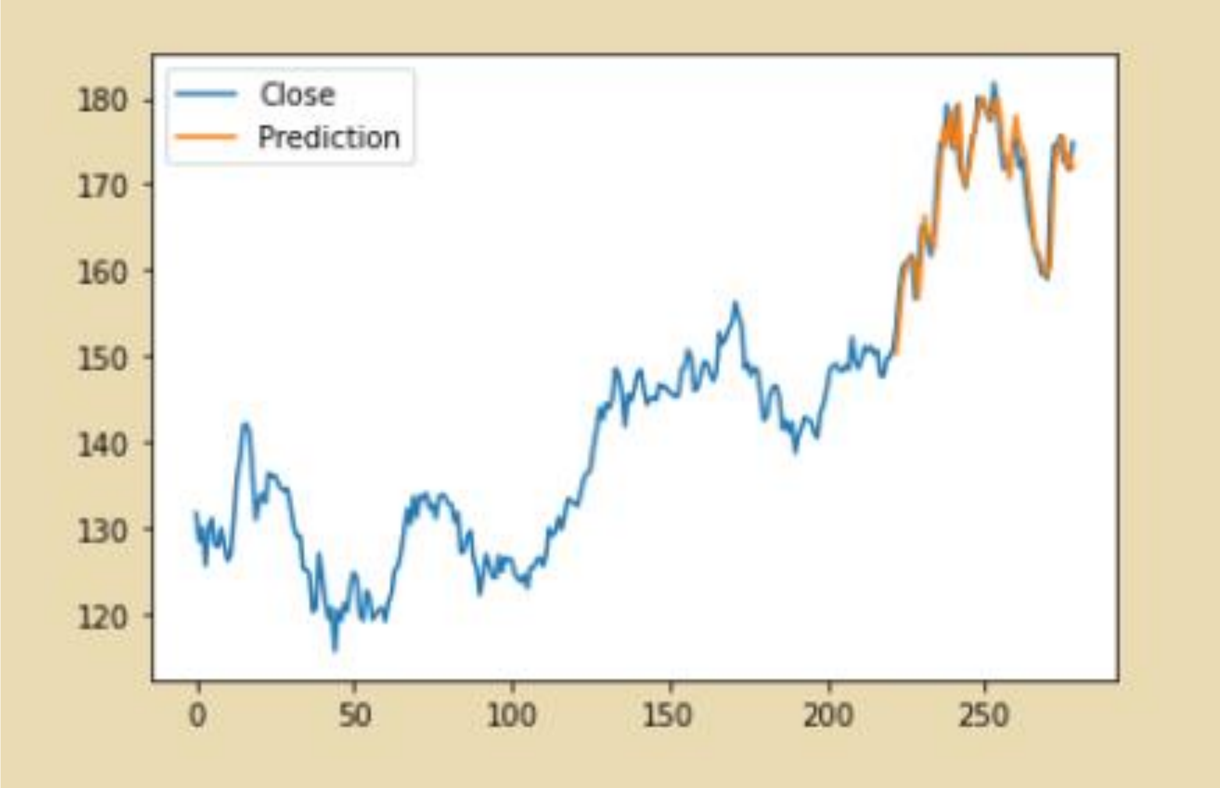
SARIMA Model

Check Differencing. Looks like first differencing is good.



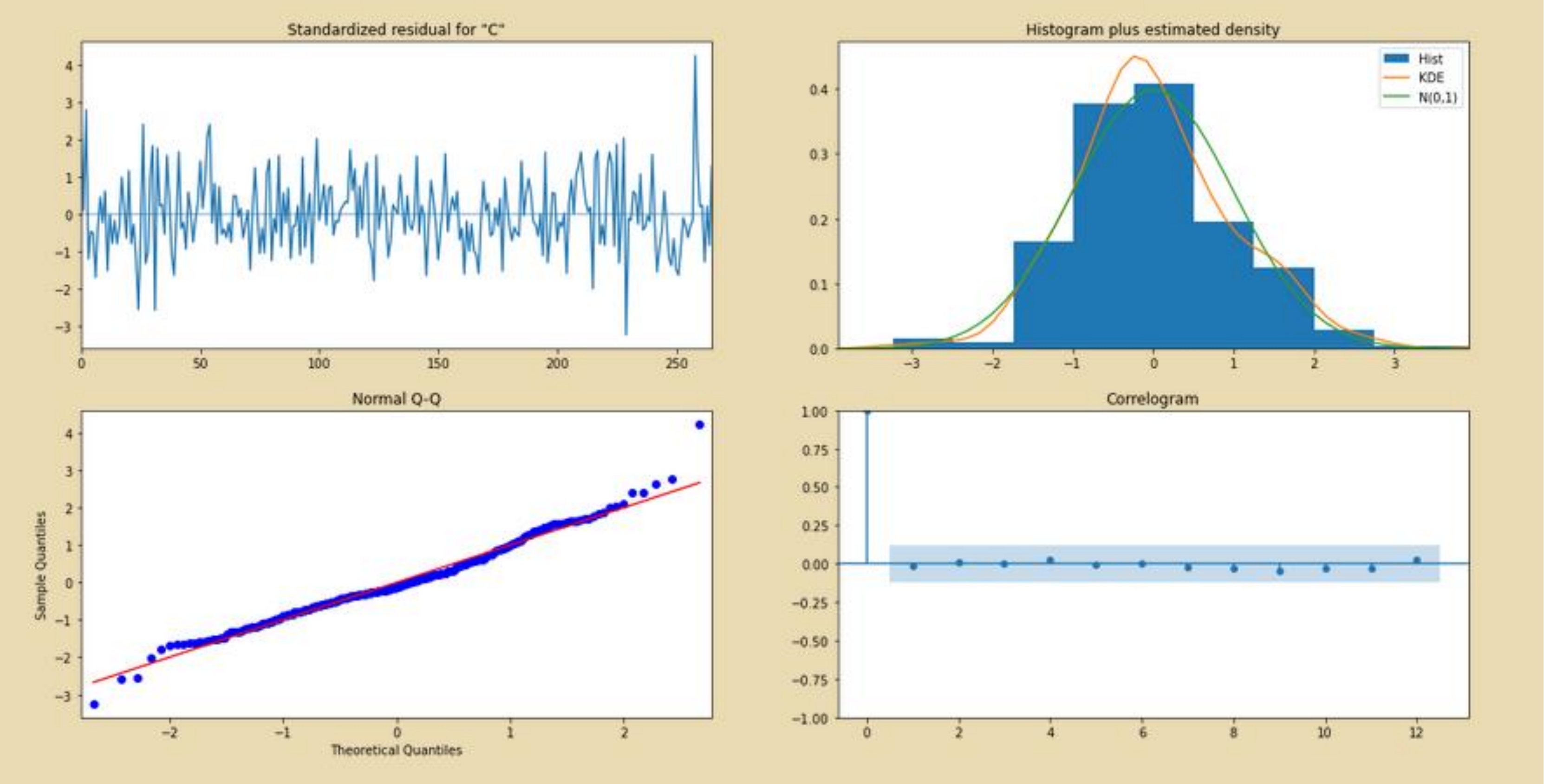
SARIMA Model

Train the model with (22,1,0)(0,1,1,12) yield the results at the right. The resultant MSE value is 7.5, which is much better than the exponential smoothing techniques.



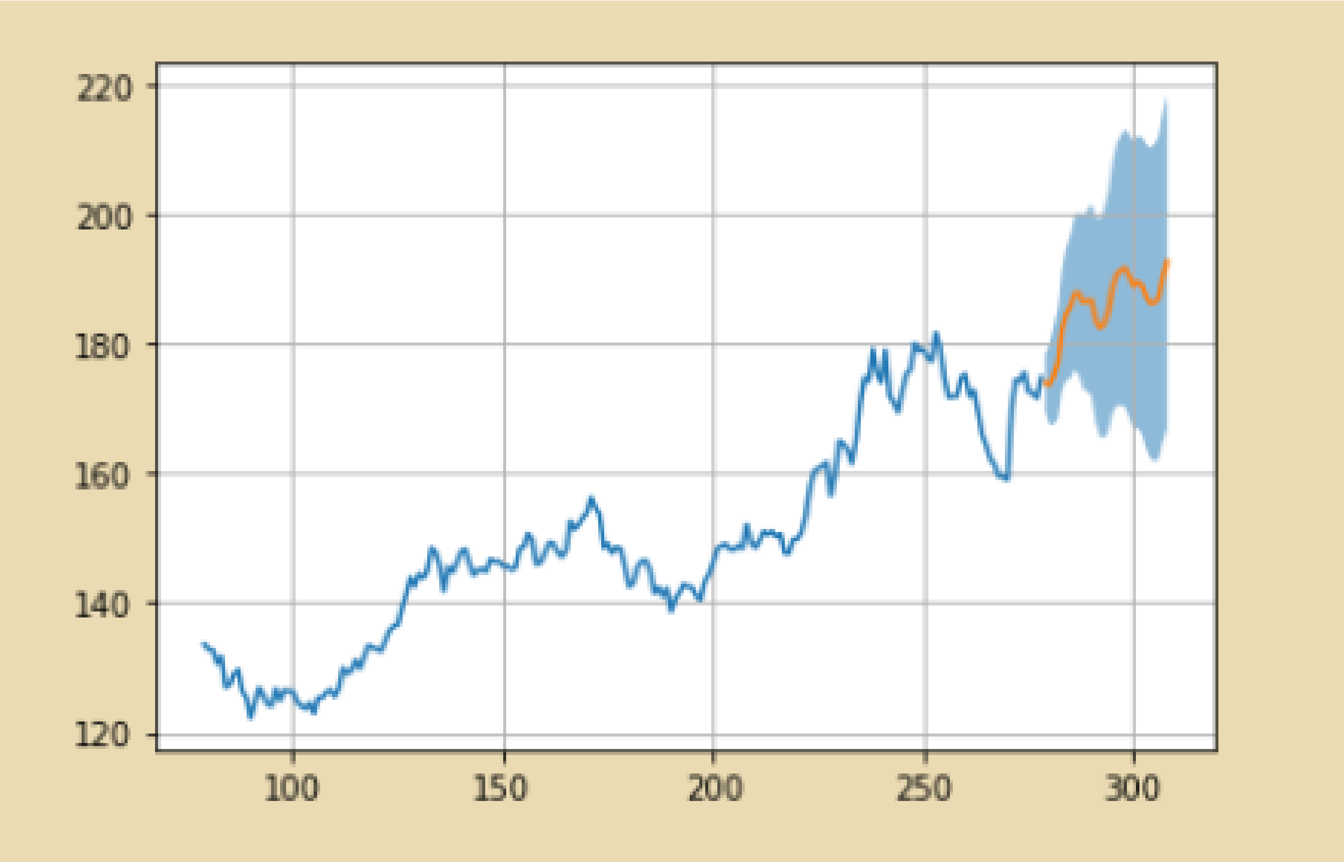
SARIMA Model

Plot Diagnostics



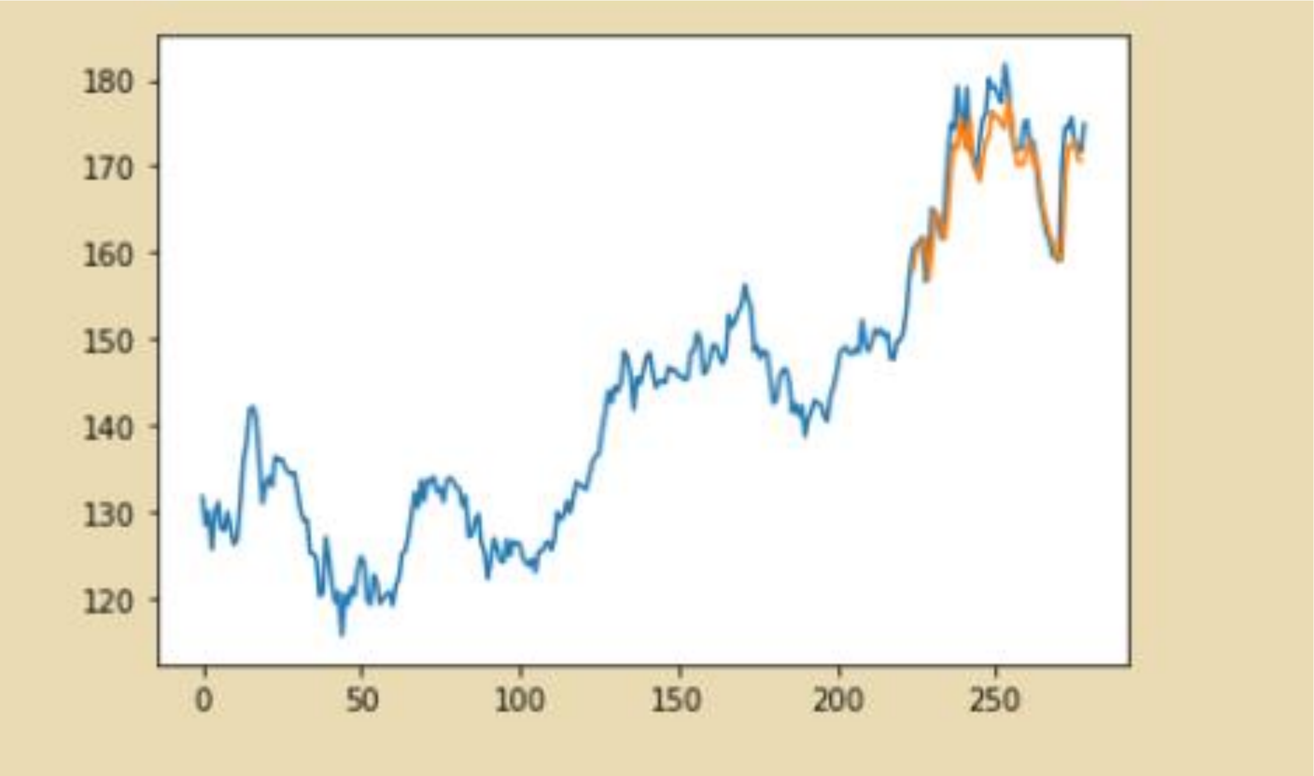
SARIMA Model

Forward forecasting for 30 days, with a level significance of 0.05.



LSTM Deep Learning

We will use LSTM with 4 hidden layer and epoch of 100. This give an MSE of 3.58



Conclusion

Both SARIMA and LSTM Model are very robust when it comes to predicting the movement of the stock price. Both of the give a very good MSE prediction compared to the exponential smoothing model.

However in practical application of the model, MSE will not be the only evaluation criteria. A model which give a more conservative prediction (i.e. model that trends to give lower stock price prediction) may be favoured as it gives a safer prediction. So the Exponential Smoothing model may be favoured for more conservative investor or trader.

The most difficult challenge of building this predictive model is the inconsistent seasonality of the stock price. One potential way to overcome this inconsistent seasonality is to build a multivariate time series model. By linking the time series to other external factors such as interest rates, inflation rates, market hype, Covid-19 cases and earning release dates allow us to overcome the inconsistent seasonality.