Stock Market Forecasting

OBJECTIVE

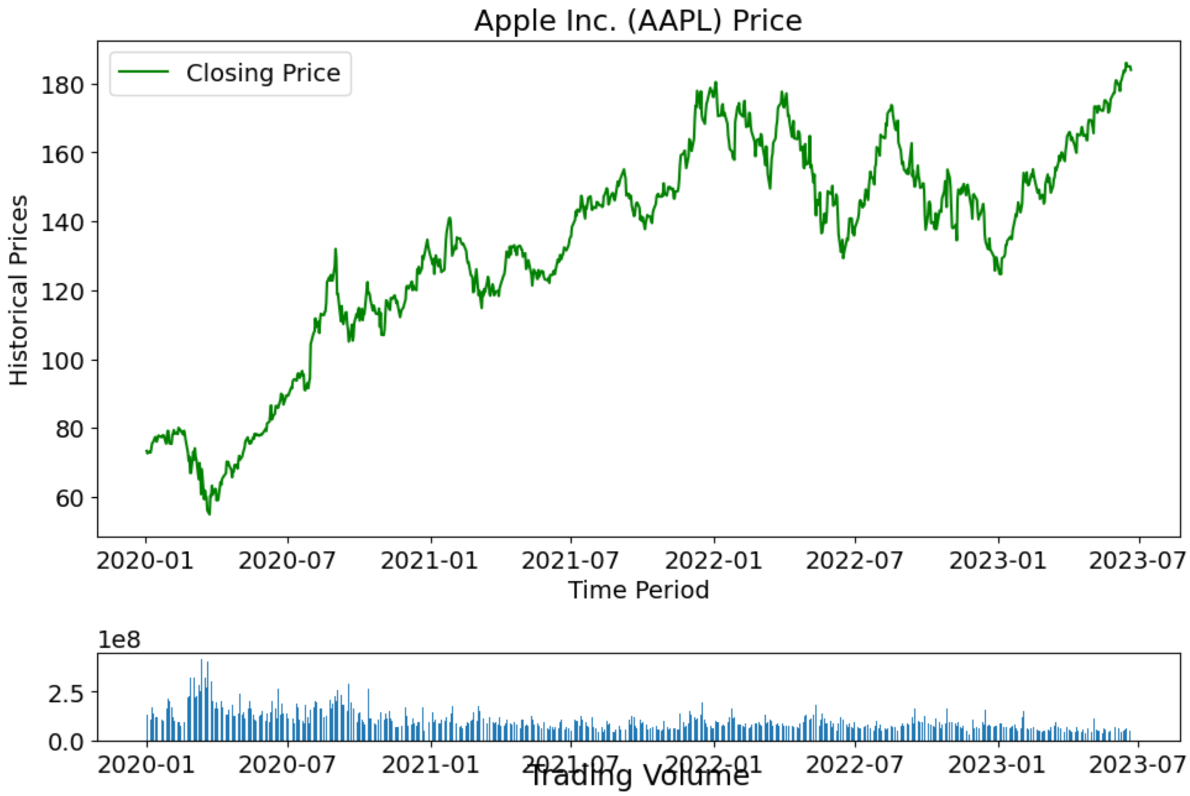
Stock market works through a network of exchanges enabling buyers and sellers to negotiate price and make trades. The supply and demand help in determining stock price at which the investors and traders are willing to buy or sell. Stock markets has tendency to be unpredictable and analysing it is a perpetual and hard to grasp process. Since stock prices can be treated as a discrete-time series model, machine learning can be used to ease this complicated process by analysing huge numbers of data, spotting the significant patters, and generating output based on previous stock data. This analysis's target is to find the best tool for forecasting the future values or trend of the stock price since strategic forecasting plays in profitable investment and trading activities.

DATASET

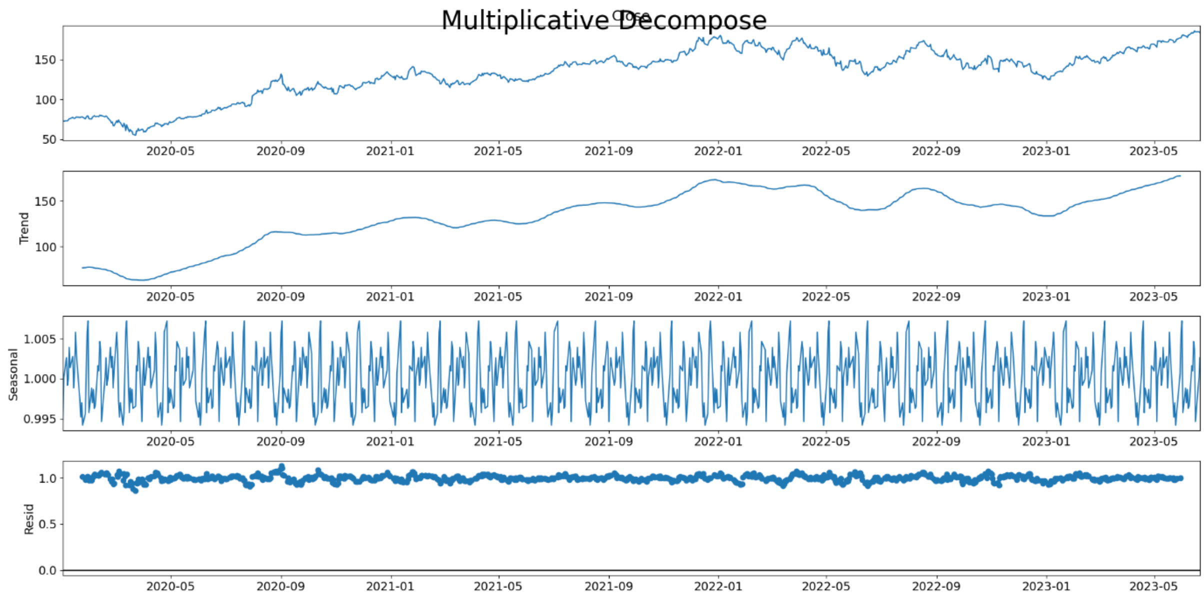
The data used in this analysis was sourced from Yahoo Finance that contains daily historical data from earlier 2020 until mid-2023. Data for Apple Inc. stock quote, shortened as **AAPL**, was downloaded as a DataFrame, which provides **Date**, **Open** (daily opening stock price), **High** (daily maximum stock price), **Low** (daily minimum stock price), **Close** (daily closing stock price), and **Volume** (daily total number of shares traded). Below is the table containing the mentioned data.

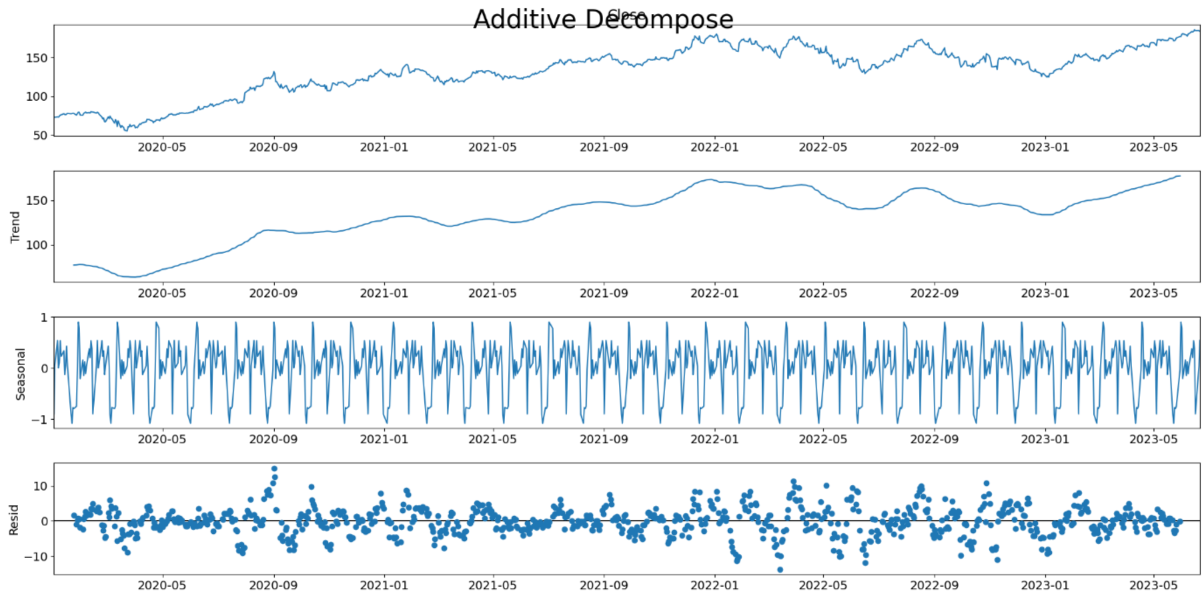
	Date	Open	High	Low	Close	Volume
Date						
2020-01-02	18263.0	72.344246	73.408998	72.087831	73.347946	135480400
2020-01-03	18264.0	72.566472	73.404104	72.407738	72.634850	146322800
2020-01-06	18267.0	71.745929	73.252690	71.491951	73.213615	118387200
2020-01-07	18268.0	73.223391	73.482251	72.647063	72.869293	108872000
2020-01-08	18269.0	72.568906	74.346741	72.568906	74.041481	132079200
...	...	...	...	...	...	...
2023-06-14	19522.0	183.369995	184.389999	182.020004	183.949997	57462900
2023-06-15	19523.0	183.960007	186.520004	183.779999	186.009995	65433200
2023-06-16	19524.0	186.729996	186.990005	184.270004	184.919998	101235600
2023-06-20	19528.0	184.410004	186.100006	184.410004	185.009995	49799100
2023-06-21	19529.0	184.899994	185.410004	182.589996	183.960007	49477400

Since this is a univariate analysis, only closing price of the stock that would be used. The closing price and its trading volume is visualized in the following figure.

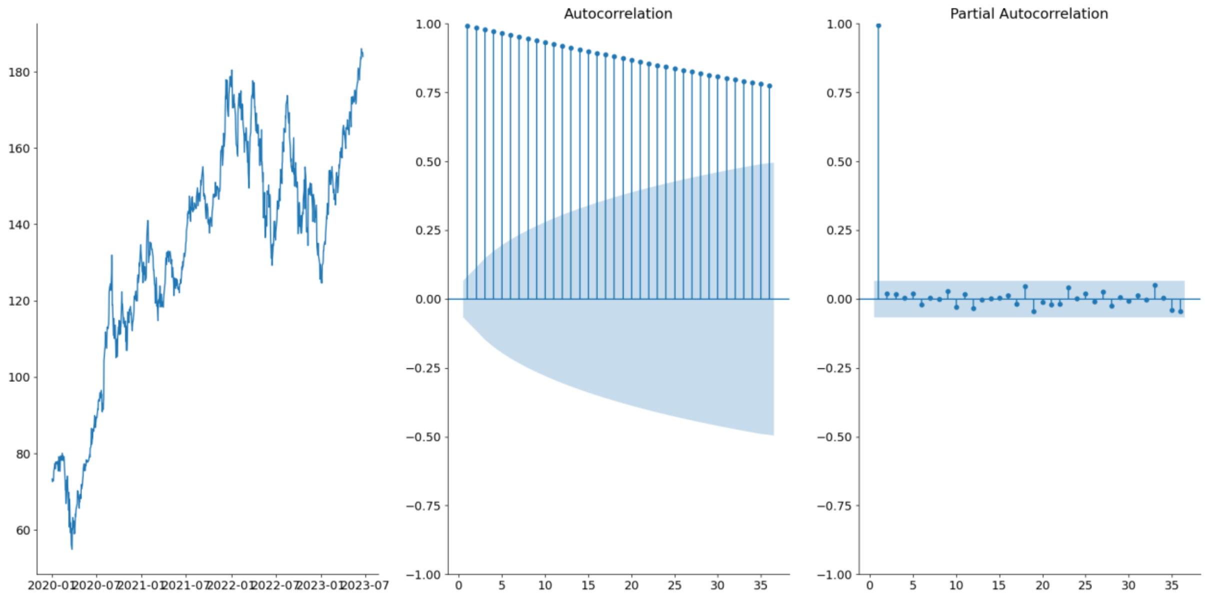


At glance, there is trend happening in apple stock price plot during the time. To be clear, the data is decomposed into trend, seasonal, and residual as shown below.





In both multiplicative and additive decompose plots, we can see that there is an upward trend. This time series also shows annual seasonality and the residual variance seems randomly scatter around in one value.

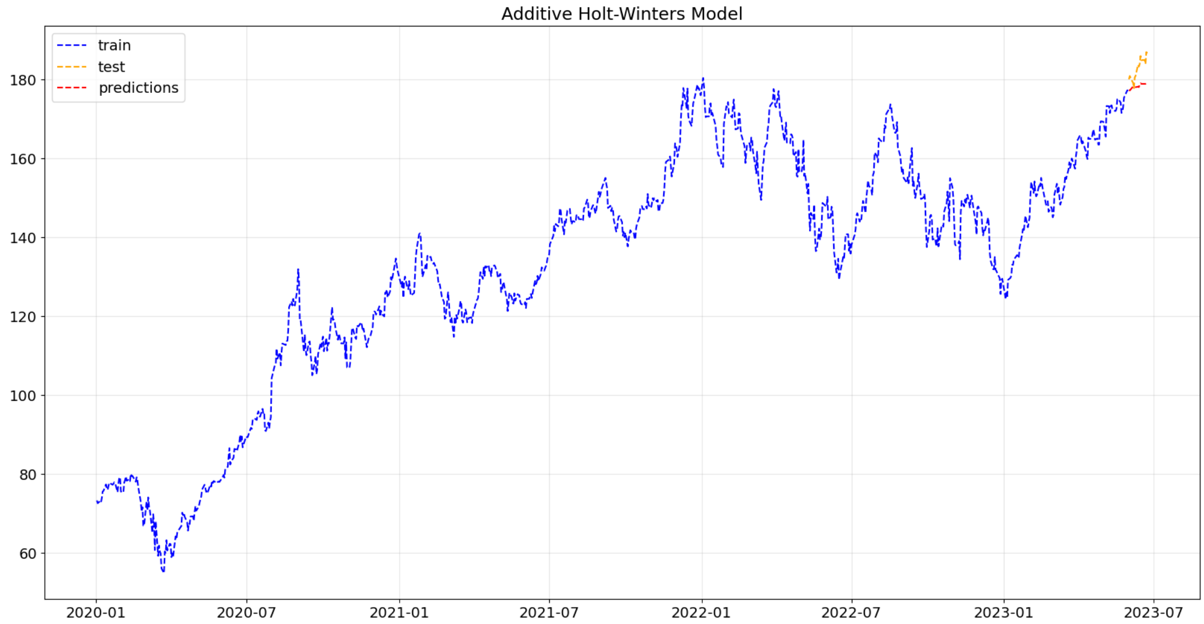


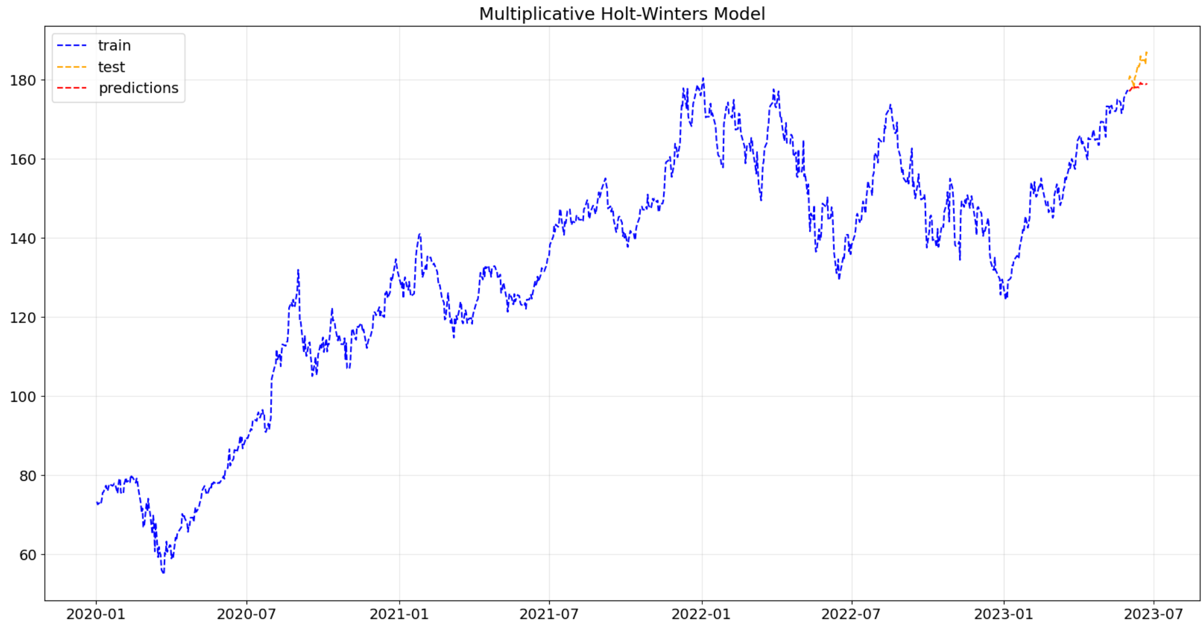
In addition, the above plots show that this time series is not stationary since its autocorrelation does not decrease to zero.

For this analysis, dataset is divided into 859 observations for train dataset and only 14 observations for test dataset.

Machine Learning Models

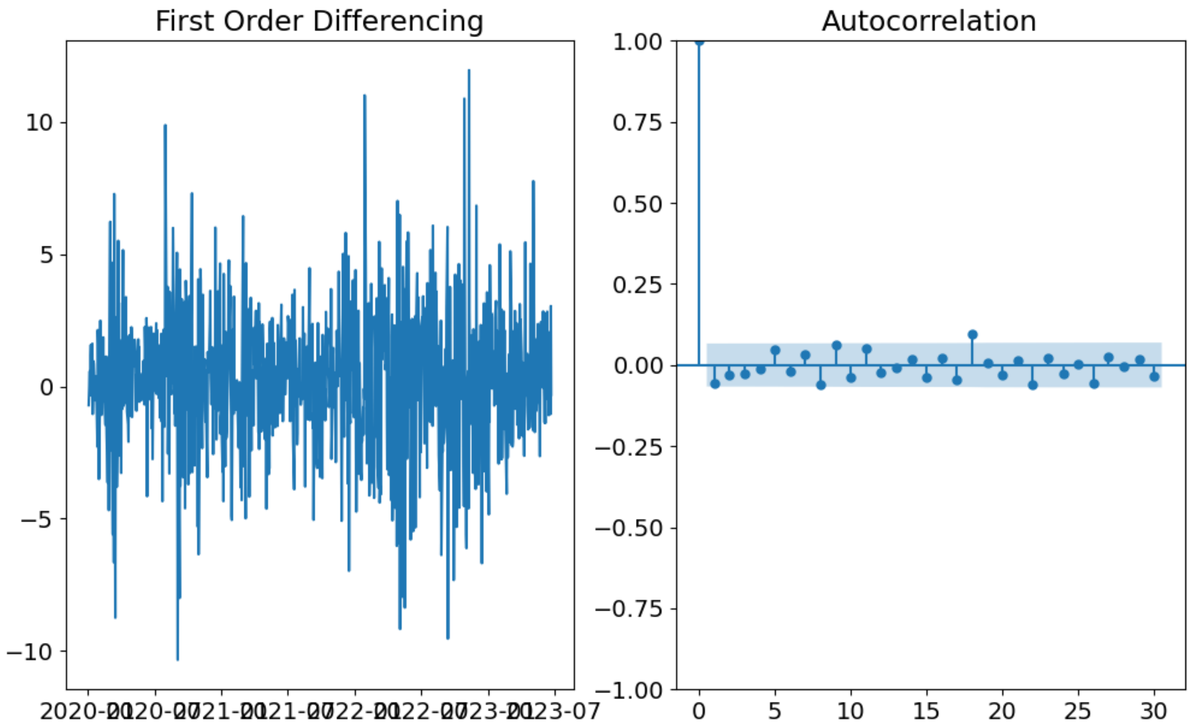
Since the dataset plot shows the existence of trend and seasonal, Holt-Winters or triple exponential smoothing model is chosen, compared to other smoothing models. There are two methods for this model, additive and multiplicative. The following plots show the results of both methods.





There is almost no difference between the above plots. Both models is able to generate some predictions that are almost accurate at first but then, continue missing out. Between two, additive Holt-Winters, with MSE value of 24.4261, is slightly better than multiplicative holt winters (MSE values is 24.831).

The next models are ARIMA and SARIMA. Since the dataset is not stationary, differencing is necessary method to do. It seems first order is enough to make the dataset stationary. The autocorrelation plot is show below.



At first, by setting the frequency to 7 (for daily dataset) and seasonality to False, result from using auto-arima is ARIMA(0,1,1). The result is also same when setting seasonality to True and let the function deciding the value of P, D, and Q. The MSE value for this model is 37.875. Meanwhile, setting D value to 1, result from auto-arima is SARIMA(1,0,0)(2,1,0,7) and its MSE value is 48.722. Hence, for this dataset ARIMA model is better than SARIMA model. Following is the plot of the ARIMA model prediction result.

However, as we can see that ARIMA model fails to capture the phenomena compared to additive Holt-Winters model. It also can be seen that the MSE value of ARIMA model is bigger than additive Holt-Winters. Hence, the best tool for forecasting the future value for this dataset is the additive Holt-Winters model.

For future analysis, other models such as deep learning methods need to be considered before deciding which model is better in forecasting future values of this dataset. It could also better if the number of observations in test dataset is less than the number of observations in this analysis since we dealing with black swan event, which is event that started since the beginning covid-19 pandemic. This event causes most of financial data hard to predict.