**Forecasting Stock Prices with ARIMA and Prophet Time Series Analyses**

**Problem Statement:**

The Stock Market is one of many types of investment. It offers high gains as well as high losses, to companies and individuals. The US stock market offers many lucrative investment opportunities for both institutional and individual investors, barring any disastrous or unforeseeable changes with the economy or any specific publicly traded company/companies. The stocks offering the most returns to investors usually are from companies who have and continue to trend positively, both in financial performance and future outlook. Being able to forecast a stock’s future trend and quantify its future value empowers investors to make more informed decisions whether to sell a stock now to maximize returns or buy more of it for even greater returns.

**Data**

We start with exploring the data we collected.

We collected our data from Yahoo Finance ranging from April 2012 to Dec 2019 roughly counting to 2011 instances of data.

**Data Describtion :**

**Date** : Date of trading

**Open** : Price at which security first trades

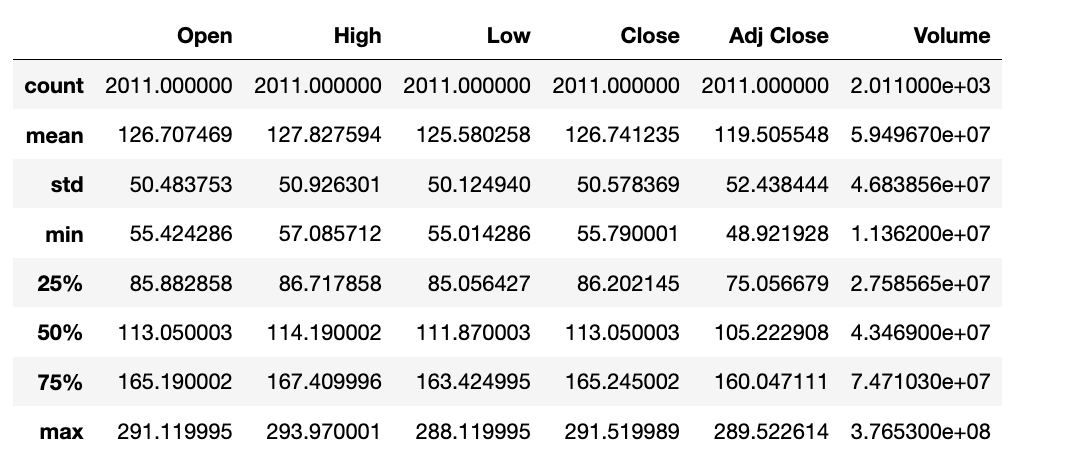
**High** : Highest Price of the trading day

**Low** : Lowest Price of the trading day

**Close** : Last Price the stock traded during the trading day

**Adj Close** : Price that is adjusts Corporate Actions on Closing Price

**Volume** : Number of Shares that changed hands during the trading day



**Project Goals:**

The goal of this project will be to use stock market data to determine the future trend and closing price of the Apple Inc. stock (denoted as AAPL). Through the Arima and Prophet time series models.

**Dataset Description:**

I will be retrieving data from Yahoo! Finance for this project. The data downloaded provides dates, open price, volume, close price, high price, low price, close price, adjusted close price and volume. It will be downloaded ten years worth of Apple’s stock market data, from October 2012 to September 2019.

**Data Pre-Processing:**

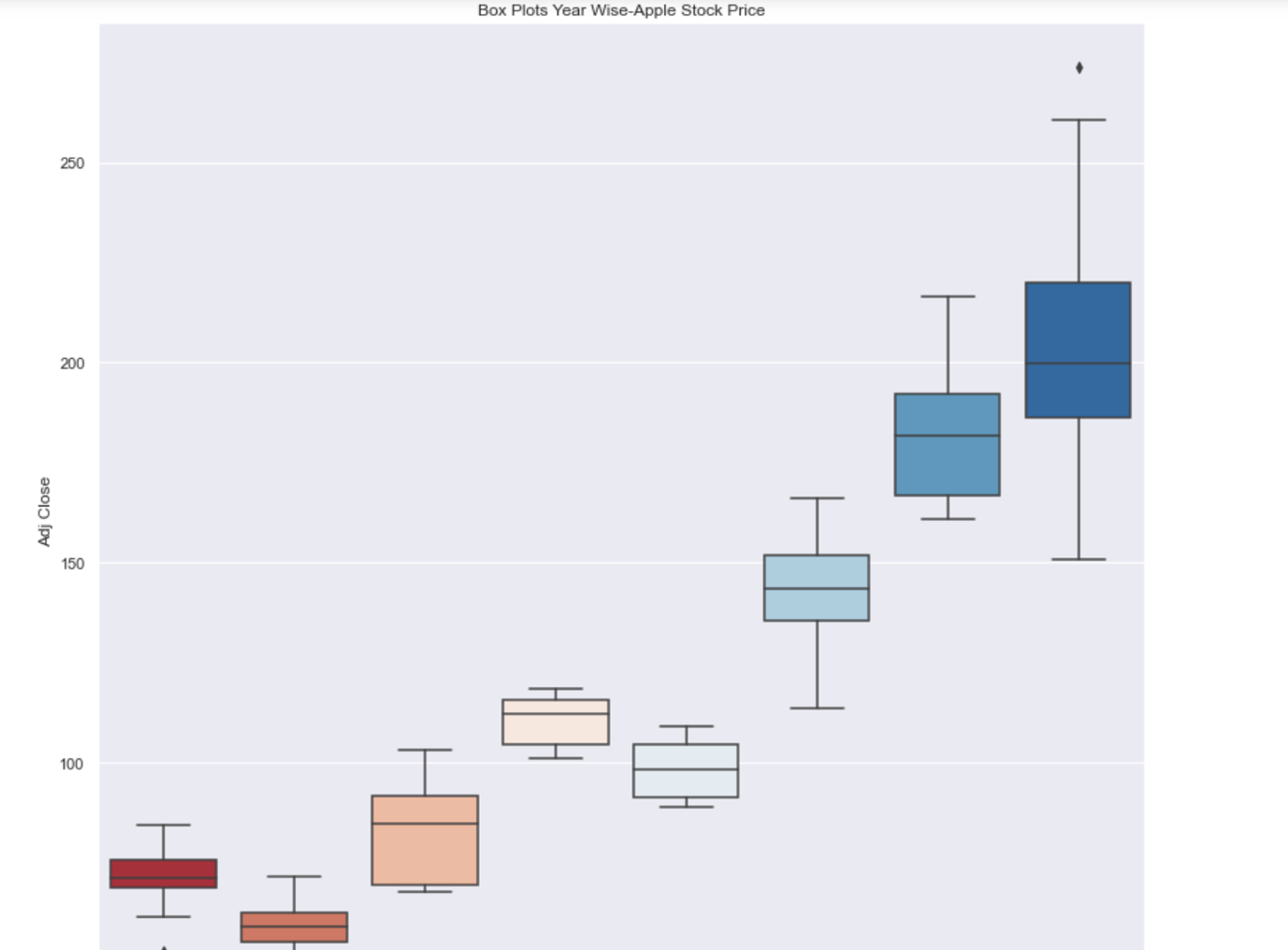
The following steps were done:

1. Convert Date into DateTime Index
2. Checked if there were any missing values
3. Eliminated Features like “Open”, ”High”, ”Close” as they were Multicollinear with “Adj Close”.

We consider “Adj Close” as our target variables as it accounts for all corporate decisions like stock split and dividends. Volume filtered out as it was less correlated to target variable.



**Exploratory Data Analysis**





When Data is Not Normal Inter-Quartile Range(IQR) is Better Variability Metric than Standard Deviation as IQR is not affected by outliers.

As observed with BoxPlot 2014 and 2019 are the most volatile Years for Apple Stock

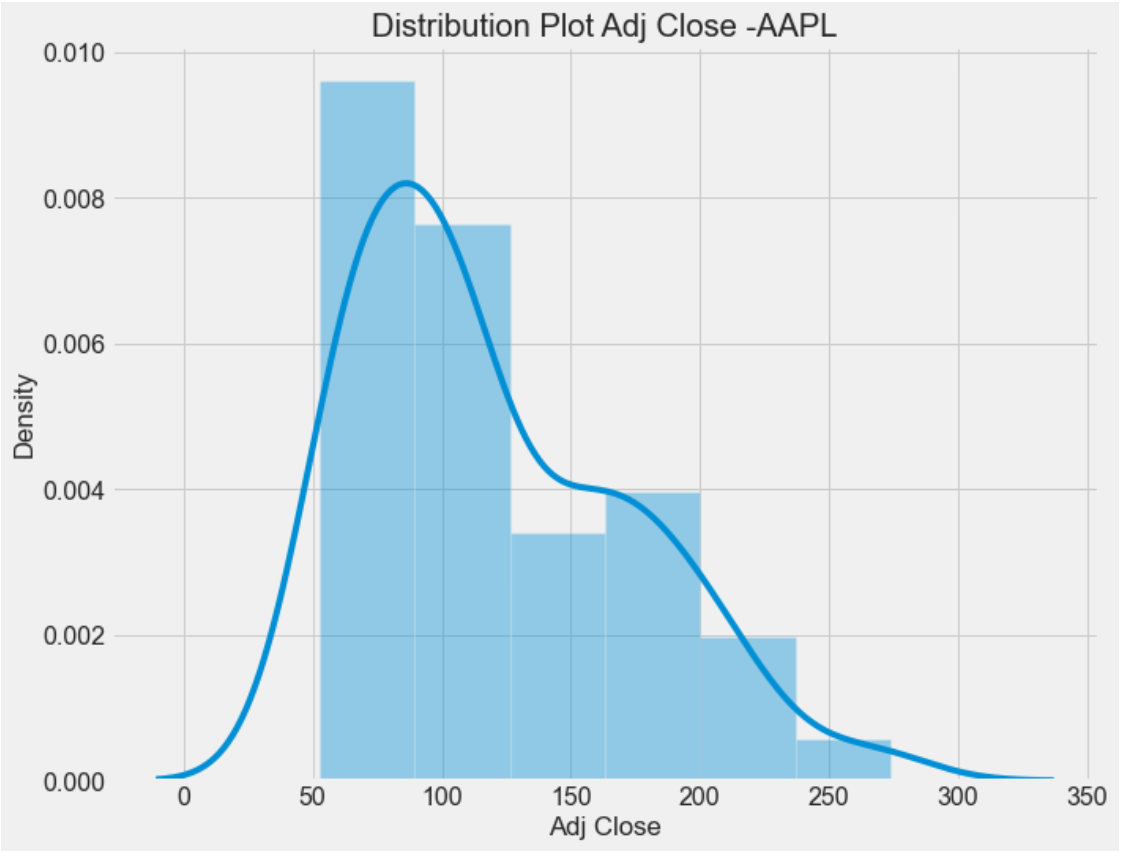


According to Mean price by Years, 2013 and 2016 are the only years where Mean price is lower than previous Year.

Average Stock Price is lower at start of the week in comparision to the end of the week.

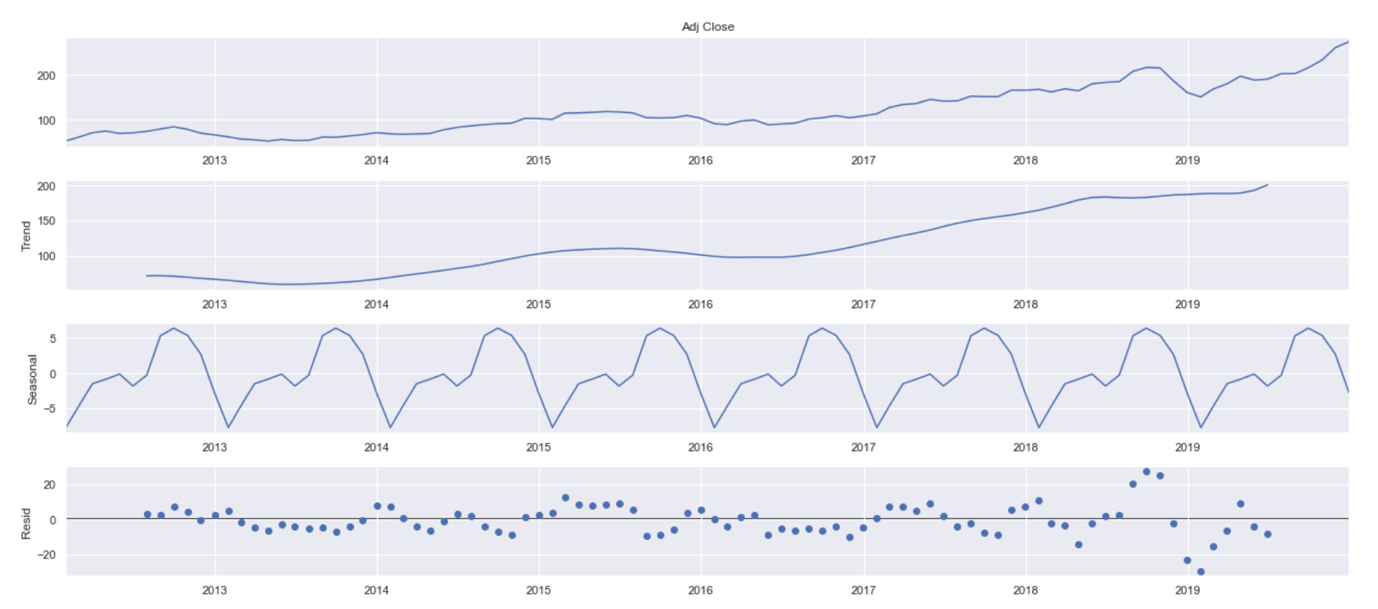
The Average Price is Highest in the Month of November.

Q4 is the best for Apple according to average stock price. By sales figures Q4 has always been strong for Apple since the new product cycle takes place and it’s the Holiday period. We also observe this as a seasonal effect for Apple.



**Seasonal Decomposition of Apple’s Closing Price**

Apple’s stock price performance can be broken down and explained by three main time series components-- trend, seasonality and noise (as illustrated below).



**Inferences By Decomposition :**

Trend : Overall an Upward Trend

Seasonality: There appears to be seasonality, AAPL has rallied during the Holiday season as expected. Since Holiday period has good sales for Apple Over the Years.

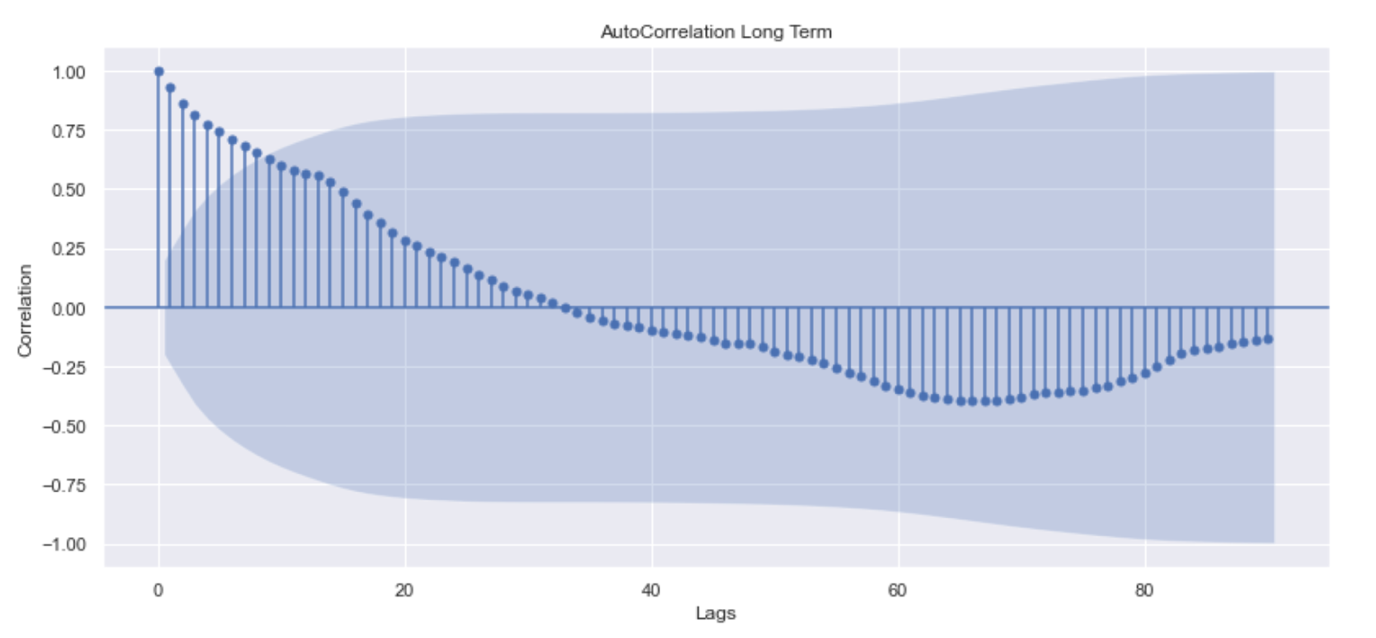
**Stationarity Test of Time Series**

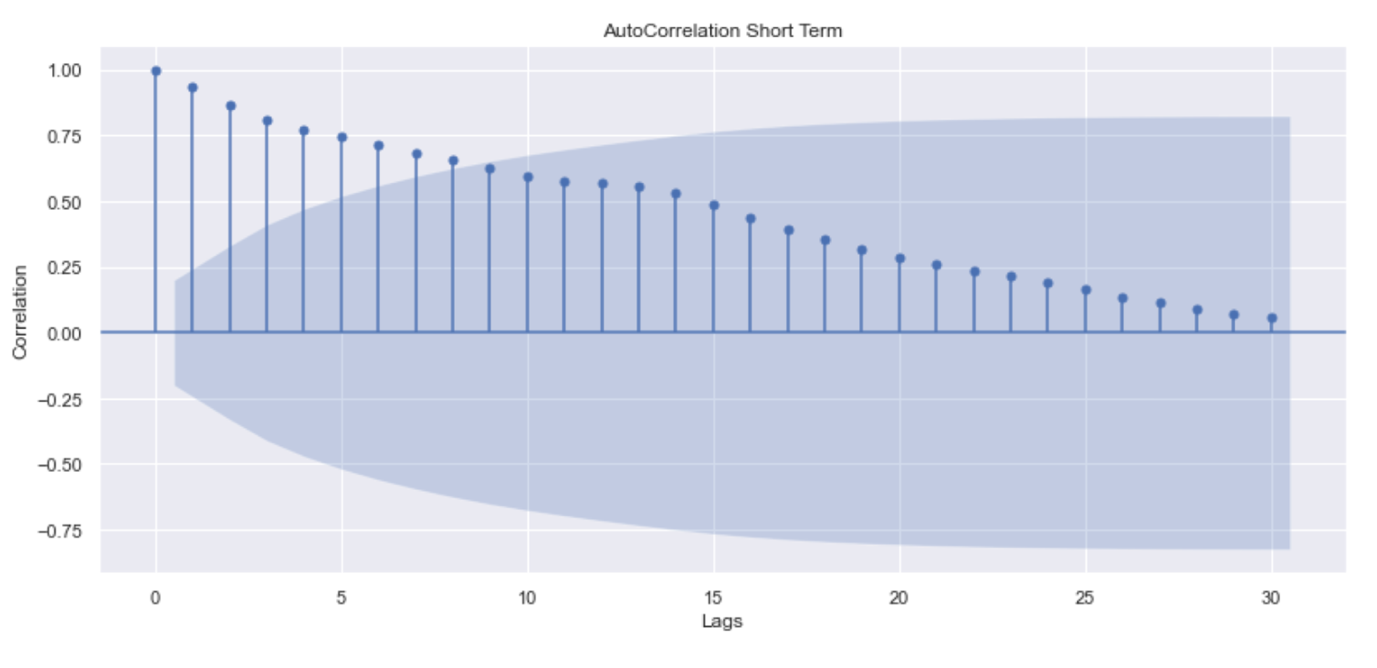
Using Augmented Dickey-Fuller(ADF) Test

**Null Hypothesis** : Time series has a unit root -*It is non-stationary*

**Alternate Hypothesis** : Time series does not have a unit root -*It is stationary*

Time Series is Stationary if we have constant mean, constant variance and No Trend and Seasonality.





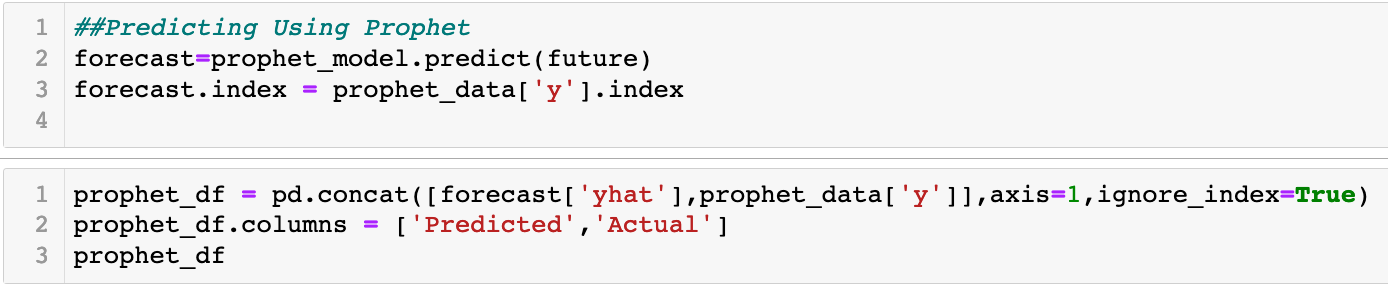
**Interpretations of Auto Correlation Funcion (ACF) Plot**:

* Slow Decay of correlation values indicates that the future values are heavily dependent on the lagged values . This shows that the series is not random and good for time series modelling .
* Also tells us series is Non-stationary
* Indicates Moving Average of 1

Arima Test Results

The mean value is close to zero but it suggests that there is some bias in the model. Overall the model has performed good for a problem like Stock Price Prediction being a difficult problem .

**Forecasting Prophet-Basic Model**



Result Metrics for Prophet Basic-Train Data

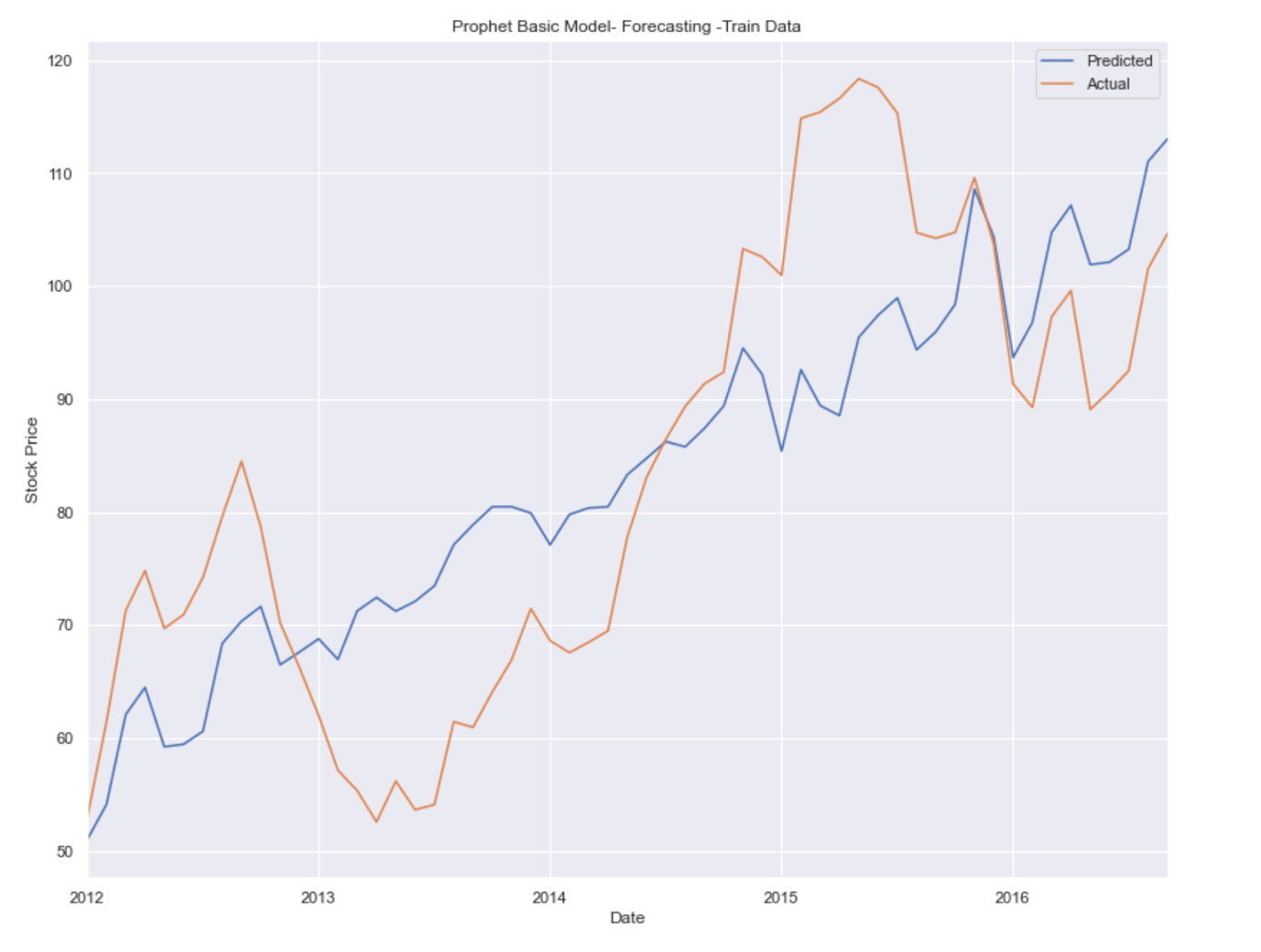
R2 Score : 0.598

Mean Squared Error : 157.648

Mean Absolute Error : 10.794

Mean Absolute Percentage Error 13.785

Accuracy(100-MAPE) of Model is 86%



Problem with Prophet is its Overfits quite easily for out dataset. Therefore we try some other hyperparamters with Hit and Trial.

After Hyper-parameter tuning the model is not able to capture the seasonality and sudden jump in time series in the Year 2017 onwards. As we can see the Prophet Model is easily overfitted but can be useful at capturing the trend. Seasonal ARIMA shown superior to Prophet by creating extra regressors we can maybe improve the Results in future projects.

Project Summary

Seasonal ARIMA Vs Facebook's Prophet

Advantages of Prophet includes very easy to implement, fast , and less statistical know-how model . In Seasonal ARIMA we had to follow lot of tests and process to generate predictions. Seasonal ARIMA is better at capturing the seasonality part .

Prophet has a overfitting problem. Overall Both models are robust. Prophet is better at dealing with outliers.

We have found Seasonal ARIMA is much better our at prediction problem. More confidence when predicting with Seasonal-ARIMA since its backed by Mathematical and Statistical tests. Accuracy of SARIMA is 89% and 80% for Prophet(Both on Out of Sample Data). Anyway after the analysis we can conclude that the best thing to do is to buy Apple stocks.

**Conclusion**

The robustness of this time series modeling will provide the most critical information to investing or divesting from a certain stock, as it provides tangible and quantifiable results that empower asset/money managers to make informed decisions on behalf of their clients. Moreover, it allows such decision makers to avoid relying on price speculation, investor sentiment and market volatility-- all of which are intangible factors that cannot not be confidently quantified or used to make a responsible investment. Other methods, however can be applied to predict time series, probably more sophisticated ones, but for a starting point this is fairly enough to have a solid understanding of how to approach this kind of challenge.