**Microsoft Stock Price**

**Forecasting**

**Table of Contents**

[**Introduction** 1](#_Toc102491956)

[**Problem Statement** 1](#_Toc102491957)

[**Literature** 2](#_Toc102491958)

[**Data** 3](#_Toc102491959)

[**Basic Statistics** 3](#_Toc102491960)

[**Data Description** 4](#_Toc102491961)

[**Data Preprocessing** 4](#_Toc102491962)

[**Data Exploration** 4](#_Toc102491963)

[**1.** **Correlation among the variables** 4](#_Toc102491964)

[**2.** **Outliers Detection** 5](#_Toc102491965)

[**3.** **Exploring the Stock Volume Field** 5](#_Toc102491966)

[**Predictive Modeling** 6](#_Toc102491967)

[**1.** **Linear Regression** 6](#_Toc102491968)

[**2.** **ARIMA (Auto Regressive Integrated Moving Average)** 8](#_Toc102491969)

[**3.** **Exponential Smoothing Models** 12](#_Toc102491970)

[**4.** **ARIMAX** 15](#_Toc102491971)

[**Business Insights** 17](#_Toc102491972)

[**References -** 19](#_Toc102491973)

[**Appendix** 20](#_Toc102491974)

# **Introduction**

In the financial business, stock prediction is commonly utilized to assist market players in determining the future value of a stock. Primary stakeholders, such as the company's board of directors, will benefit from an accurate prediction of future stock values when budgeting capital expenditures. External users, like those of private investors and investment companies, will benefit from a good forecast as they'd be able to earn significantly from the stock market. As a result, analysts all over the world have invested a significant and tremendous amount of time and resources developing numerous techniques and principles to improve stock predictions.

Microsoft has been a global leader in personal computer operating systems for over three decades, with a market share of more than 90 percent. However, in 2014, they lost their market supremacy in smartphone operating systems to Android. However, their concentration on innovative digital cloud-based solutions for corporate systems, as well as the acquisition of LinkedIn, has helped them develop and dominate the digital arena. As a result of these changes, Microsoft's stock (MSFT) has become one of the most popular on the market.

We selected Microsoft since we were intrigued about the future trend and wanted to generate a stock prediction utilizing time series forecasting techniques. We targeted to apply several forecasting models in this project to produce data-driven decisions and reliable investing decisions. The goal was to build the best fitting model that can properly predict the Microsoft Stock price using historical data, which will eventually assist investors make investment decisions and prepare the company to better plan their future.

# **Problem Statement**

The stock market is a broad term that relates to the structured trading of securities on numerous exchanges, as well as the over-the-counter market. The stock market is amongst the most important aspects of a market economy since it allows businesses to raise funds by enabling investors to purchase shares of a company's ownership. Investors might potentially profit from a company's future success by purchasing shares in the company. While there are literally billions to be made by purchasing shares and then trading them for a positive return, not all investors succeed in generating a high return on investment, and even fewer succeed in making a fortune.

The stock market is highly volatile; any geo-political movement may have an influence on the price trend of stocks in the stock market; for example, we recently saw how Covid-19 impacted stock prices, so that is why doing a solid trend analysis on financial data is complicated.

General stock market study is insufficient for a novice investor to make an investing choice. Because the common trend in the stock market is very risky for investing, most people are unable to make decisions based on common trends. They are unable to grasp the intricacies of the data. The seasonal variation and consistent flow of any stock will assist both existing and new investors in understanding and deciding whether to invest in a certain stock. The most effective method to address this type of problem is to use data mining algorithms.

Our forecasting analysis will assist investors in determining whether to invest in Microsoft Stocks they should not invest at all, and it will also assist the company in making business choices based on the price predictions.

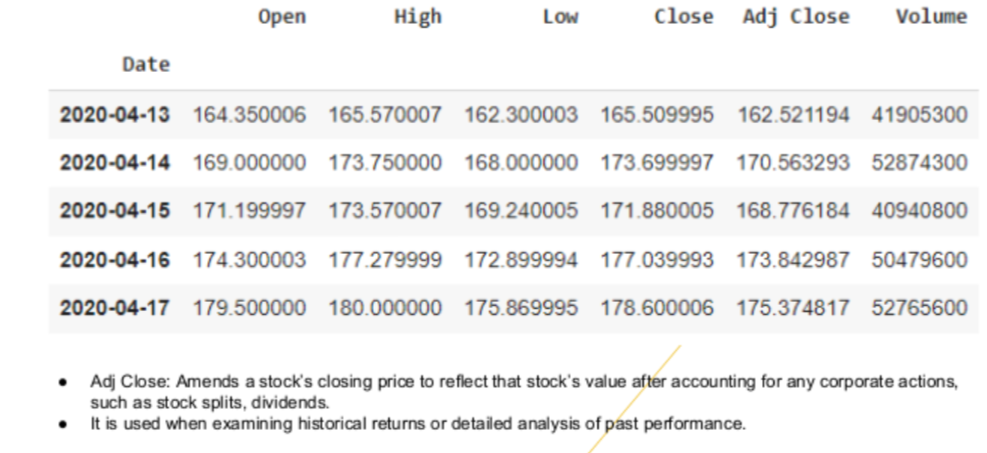
# **Literature**

Forecasting future market behavior has always been a hot and interesting topic. Stock forecasting is constantly popular in both scholars and professionals over the years to predict better returns on the investment over other asset classes. Theorists and researchers have worked hard to develop numerous approaches and instruments for conducting close to real predictions. In the initial stages, researchers used the traditional random walk theory and the Efficient Market Hypothesis (EMH). According to the EMH (Fama, 1991; Fama, Fisher, Jensen, Roll, 1969), the present stock price reflects all available information. As a result, the stock price follows a random walk model, making it challenging to anticipate the future price of the stock. While other scientists, meanwhile, found that the future stock price might be predicted by examining and evaluating the time series (Matsubara, Uehara, Akita, Yoshihara, 2016, Adewumi, Ariyo, Ayo, 2014). For share price predictions, statistical forecasting models such as the Autoregressive technique (AR), Moving average model (MA), Autoregressive Moving average (ARMA) model, and Autoregressive Integrated moving average (ARIMA) models and various other time series forecasting analysis have been employed. Experts have also explored improving the fundamental ARIMA model for greater accuracy. With these trials on the historical data of stock prices, the scholars discovered that the inclusion of additional features can significantly improve the data prediction accuracy.

Our approach was to use past data to fit various models such as the Linear Regression Model, Double Exponential Smoothing Model, ARIMA, and ARIMAX model to predict the future price of Microsoft stock. The best model would be selected based on error terms and how well the model fits. The best model will then be used for prediction for the purpose of our project.

# **Data**

We sourced our data from Yahoo Finance. We collected data which was recorded on a daily basis for a period of two years from April 13, 2020, to April 11, 2022.

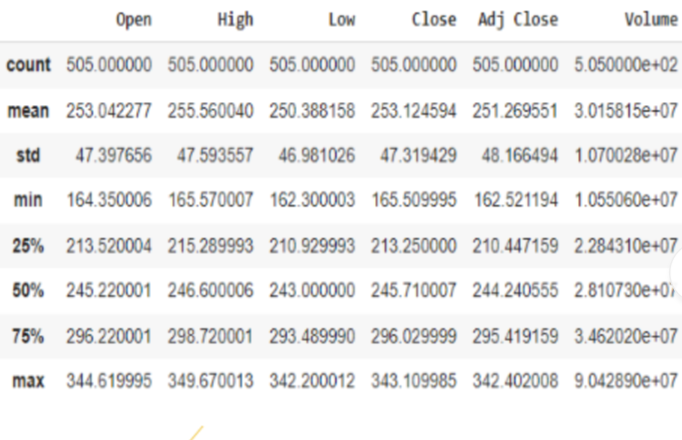


**Figure 1 - MSFT Data Set**

# **Basic Statistics**

Our target variable ‘Adj Close’ which is the amended value of the ‘Close’ field in our data set had a maximum value of 342.402008 in this period of two years of data that we had collected on a day-to-day basis of the stock.

The minimum value of our target variable that was recorded in this period was 162.521194 while the mean value recorded for the target variable for a period of two years was 251.269551.

****

**Figure 2 - Statistical Analysis of the Data**

# **Data Description**

We worked with 505 rows in our dataset. The dataset has the field ‘Date’ which is of the Date datatype, and this records the date information for the daily trading session. We have the ‘Open’ field which is of the numeric data type. This field records the opening price of the stock on a day-to-day basis when the market opens up. We have the ‘High’ field which is of the numeric data type and records the highest price of the stock during the trading session in a day. The numeric data field ‘Low’ records the lowest price of the stock in a day during the trading session.

The numeric field ‘Close’ records the lowest price of the stock during the trading sessions. ‘Volume’ field is the numeric data field which records the total number of shares traded in the market trading sessions during the day. Our target variable ‘Adj Close’ is also of the numeric data type which records the amended close price of the stock during the day. The amendment in the closing price of the stock can happen because of any corporate actions, such as stock splits, dividends.

We selected the ‘Adj Close’ field as our target variable as we considered that the other fields will help us determine a trend in this variable better.

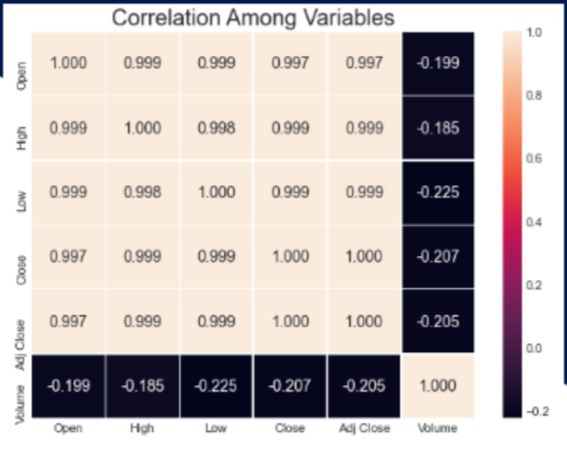
# **Data Preprocessing**

There were no missing values in our data set. We partitioned our data set into train and test. We considered the last one year’s dataset for our testing purposes.

# **Data Exploration**

## **Correlation among the variables**

We found that the predictor variable ‘Close’ and our Target Variable ‘Adj Close’ were highly correlated for obvious reasons and had a maximum positive correlation value of 1. Hence, we decided to drop the ‘Close’ field. We also found a maximum negative correlation value of -0.225 between the fields ‘Low’ and the ‘Volume’.

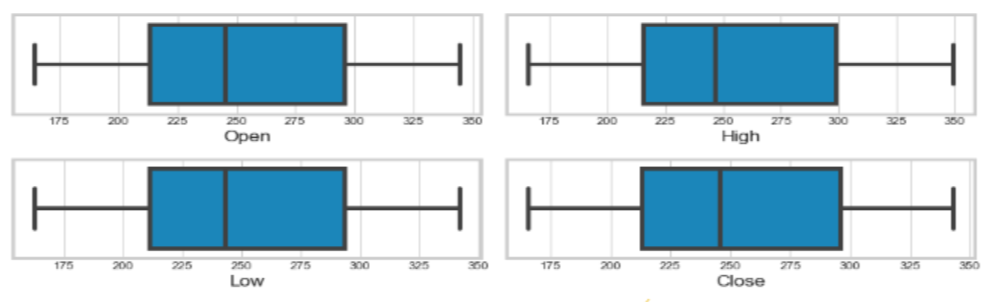


**Figure 3 - Correlation Analysis among variables**

## 

## **Outliers Detection**

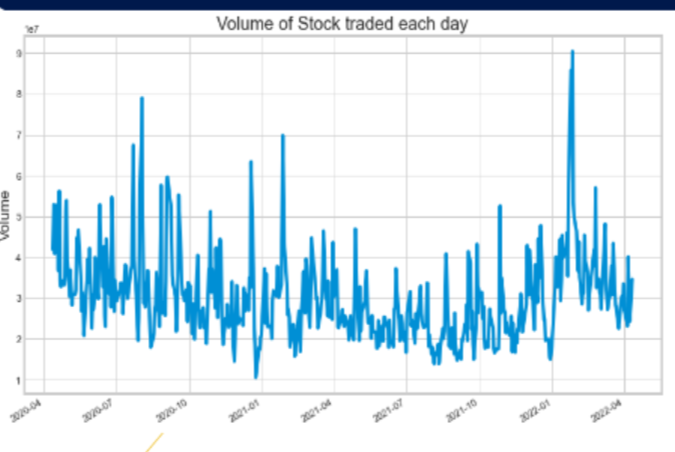
We built boxplots to inspect if we had any outliers. We found that all our data points were well within the boxplots and this clearly implied that our data did not have any outliers.



**Figure 4 - Outlier Analysis using Box plot**

## **Exploring the Stock Volume Field**

We built a trend line to observe the volume of the stock that was traded in this period of two years. On creating a trend line, we recorded an observation that the Maximum Number of shares were traded between the months of January to February of the year 2022. Also, there was another interesting observation that was recorded. We observed that the least number of shares were traded, and this was recorded between the months of December 2020 and January 2021.

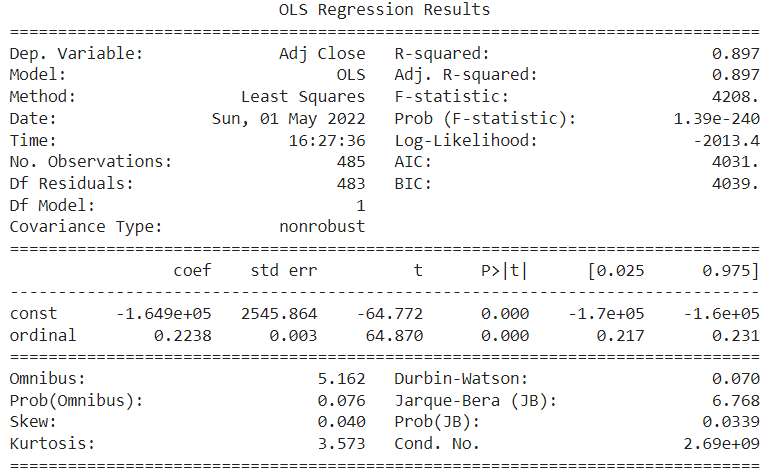


**Figure 5 - Stock trading volume Time Series Trend**

# **Predictive Modeling**

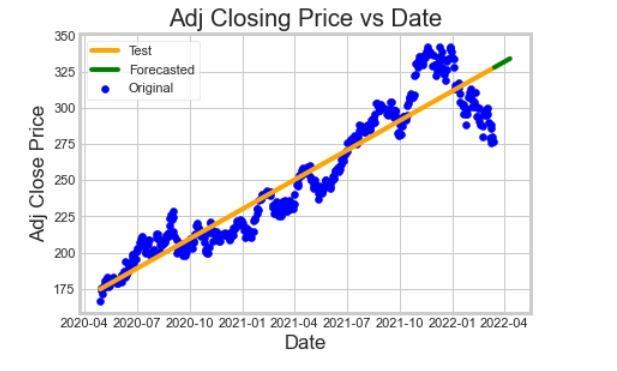
## **Linear Regression**

* Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data.
* One variable is an explanatory variable, and the other is a dependent variable.
* Before attempting to fit a linear model to observed data, a modeler should first determine whether there is a relationship between the variables of interest.
* This does not necessarily imply that one variable causes the other, but that there is some significant association between the two variables.
* A scatterplot can be a helpful tool in determining the strength of the relationship between two variables. If there appears to be no association between the proposed explanatory and dependent variables (i.e., The scatter plot does not indicate any increasing or decreasing trends), then fitting a linear regression model to the data probably will not provide a useful model.
* A valuable numerical measure of association between two variables is the correlation coefficient, which is a value between -1 and 1 indicating the strength of the association of the observed data for the two variables.
* For our purpose to predict the stock price of Microsoft, we started by fitting a linear regression model first.
* In the below image we can observe the details of the linear regression model



**Figure 6 - Linear Regression Model Results**

* On observing the statistics, we can conclude that the linear regression model has an R-Squared Value of 0.897
* The linear regression model has a Mean Squared Error of 860.33
* The linear regression model has a Root Mean Squared Error (RMSE) of 29.33

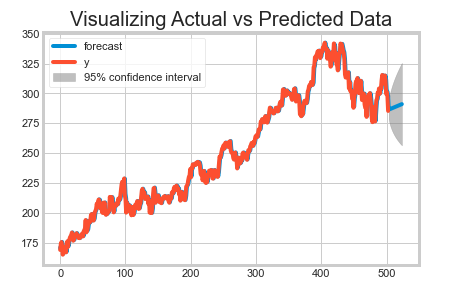


**Figure 7 - Adjusted stock closing price trend with date**

* After fitting the linear regression model, we plotted a graph between our target variable i.e., Adj Close Price vs Date and found out that the model initially showed a good fit as the actual price (blue dots) fit closely on the regression line (yellow).
* Later, the actual price was away from the predicted price, increasing the error of the predictions.

## **ARIMA (Auto Regressive Integrated Moving Average)**

* Auto Regressive Integrated Moving Average (ARIMA) model is a class of models that predicts a given time series based on its own past values, that is, its own lags and the lagged forecast errors.
* In time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model.
* Both models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting).
* ARIMA models are applied in some cases where data show evidence of non-stationarity in the sense of mean (but not variance/autocovariance), where an initial differencing step can be applied one or more times to eliminate the non-stationarity of the mean function.
* When the seasonality shows in a time series, the seasonal differencing could be applied to eliminate the seasonal component.
* After fitting the ARIMA Model we plotted the Actual and Predicted data and made a forecast (blue line) for the next 20 days.



**Figure 8 - Actual vs Predicted stock price plot for MSFT**

* From the below graph we can see the time series plot of the Microsoft Stock Price starting from April 2020.

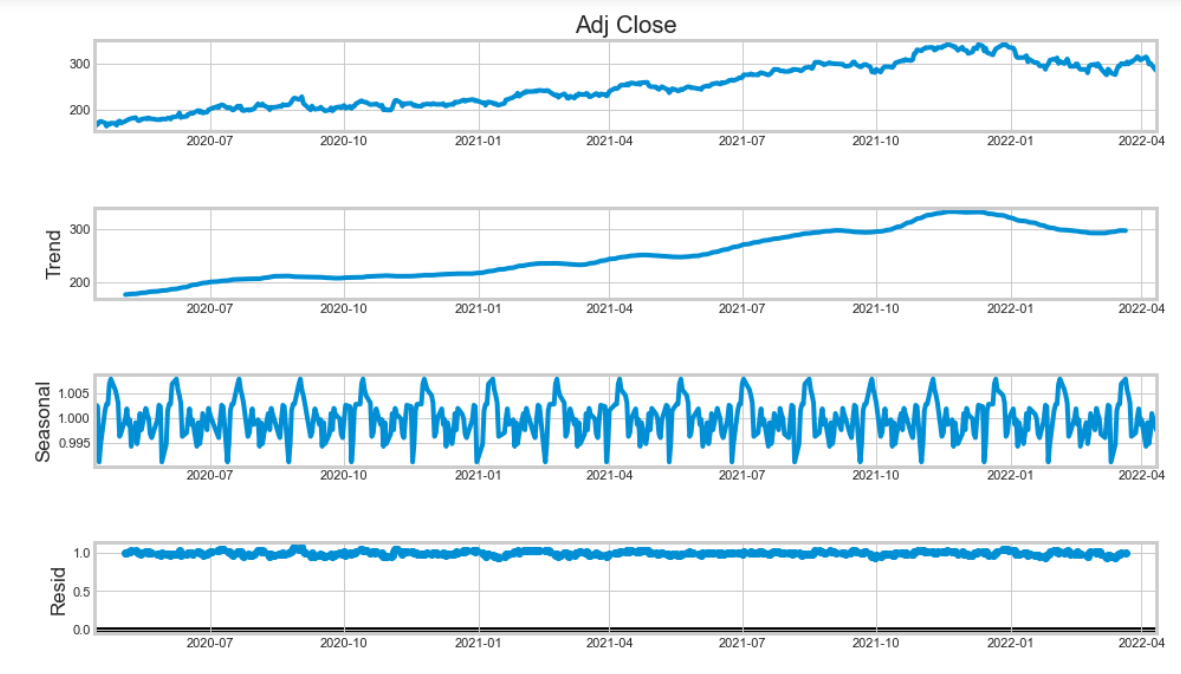


**Figure 9 - MSFT Stock Price Analysis Time Series Trend**

**Equation –**

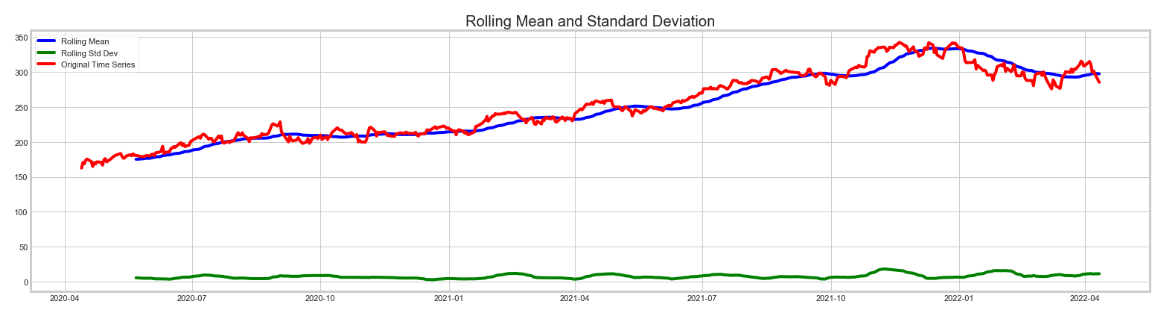
**y’(t) = c + ϕ1\* y′(t−1) +⋯ + ϕp\*y′(t−p) + θ1\*ε(t−1) +⋯ + θq\*ε(t−q) + εt**

* In order to separate the trend and seasonality from the time series of Microsoft Stock Price time series, a python code was added which was used to decompose the series.
* From the image we can see that the time series has been decomposed into 4 graphs which show: the original time series, trend in the time series, seasonality in the time series and residual in the time series.
* We can clearly deduce that starting from April 2020, there has been an upward trend in the Microsoft stock price.
* The time series has shown seasonality where we can see an upward trend being created at regular intervals.

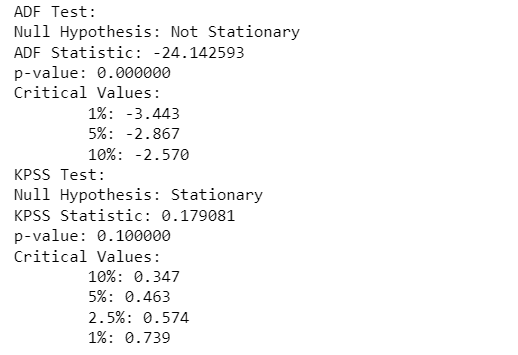


**Figure 10 - Time Series Analysis using Python Visualization**

* The graph below shows the rolling mean, rolling standard deviation and original time series from April 2020 till April 2022.
* It can be deduced that the original time series plot is in complete accordance with the rolling mean of the time series.

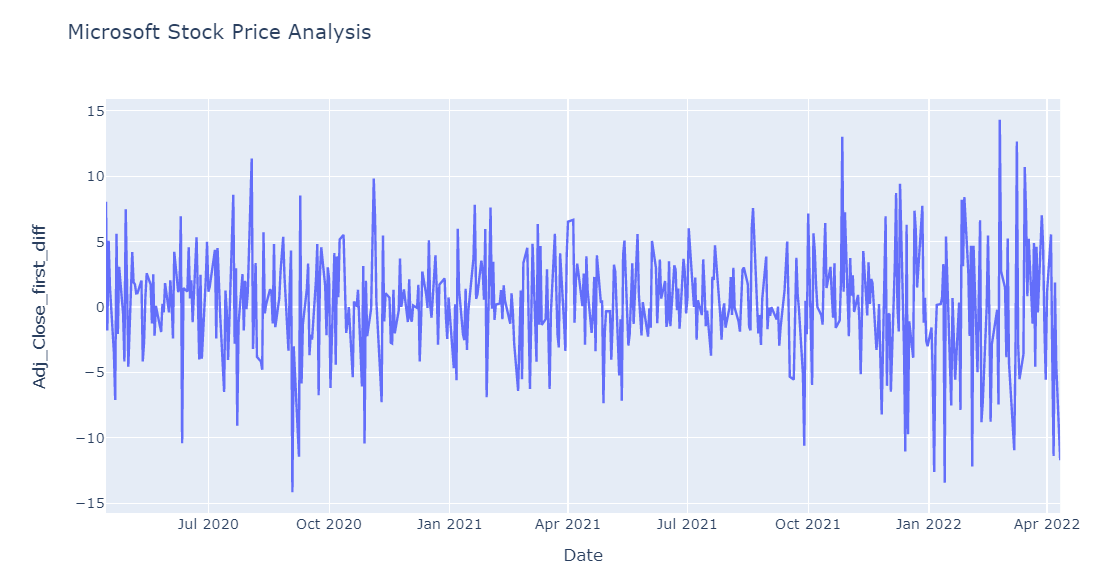


**Figure 11 - Rolling Mean and Standard Deviation analysis of MSFT**



**Figure 12 - ADF Test and KPSS Test Analysis Results in Python**

* Testing stationarity using ADF and KPSS tests.
* From the statistics, we can observe that the ARIMA Model assumes that the time series is stationary.
* The original series is not stationary as it clearly has an upward trend.
* As the p-value for ADF is less than alpha, we reject the null hypothesis and thus the series is stationary. Also, from KPSS we can infer that the series is stationary.

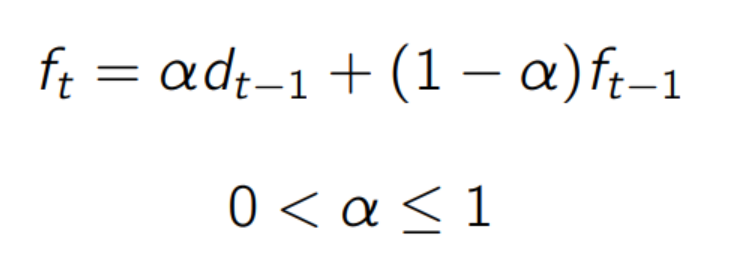
  
**Figure 13 - Time Series Trend of MSFT price**

* The graph above shows the effect of applying the first difference to the time series model.
* After first order differencing the series becomes stationary.
* The series can now be used for ARIMA modeling.

## **Exponential Smoothing Models**

**Single Exponential Smoothing:**

The idea of this model is to assume that the future will be more or less the same as the (recent) past. The weight put on each observation decreases exponentially over time. Most recent observation has the highest weight.



* **alpha** = importance the model will allocate to the most recent observation
* **alpha d(t-1)** = represents previous observation times the learning rate
* **(1-alpha) f(t-1)** = represents how much the model remembers from its previous forecast

**Double Exponential Smoothing:**

Double Exponential Smoothing is an extension to Exponential Smoothing that explicitly adds support for trends in the univariate time series. In addition to the alpha parameter for controlling smoothing factor for the level, an additional smoothing factor is added to control the decay of the influence of the change in trend called beta (b).

The method supports trends that change in different ways: an additive and a multiplicative, depending on whether the trend is linear or exponential respectively.

Double Exponential Smoothing with an additive trend is classically referred to as Holt’s linear trend model, named for the developer of the method Charles Holt.

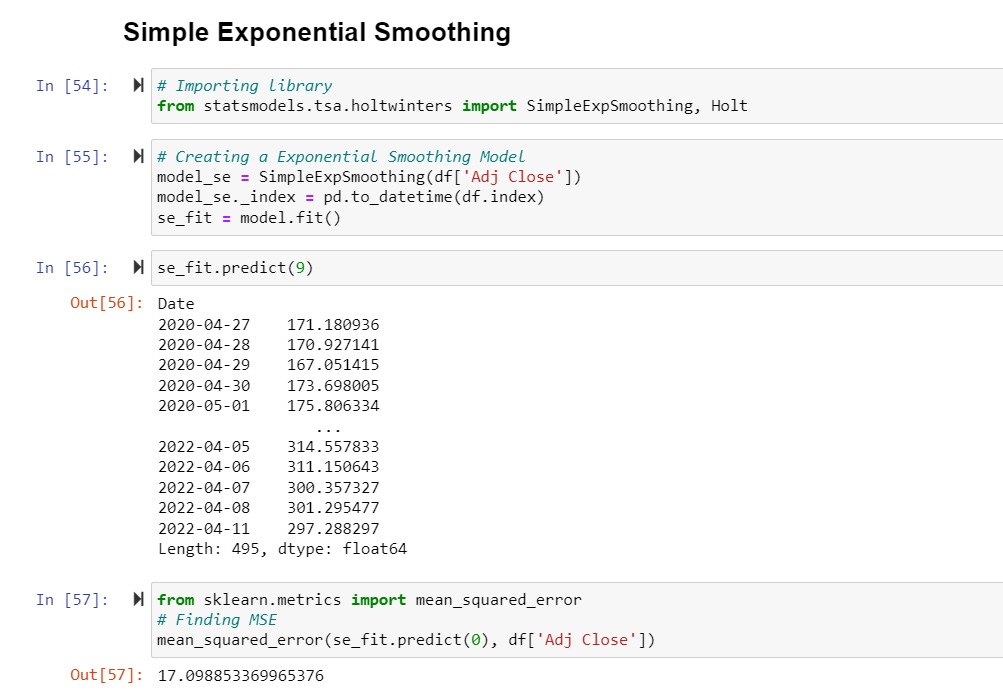
**Additive Trend:**

Double Exponential Smoothing with a linear trend.

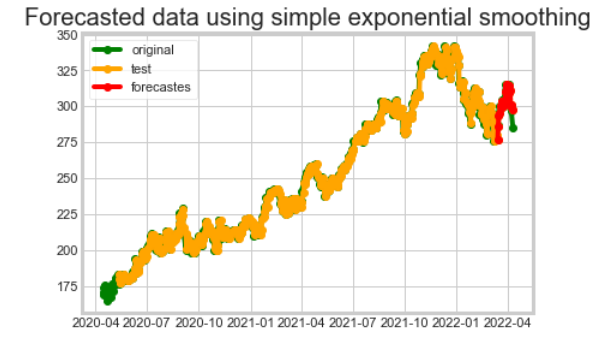
**Multiplicative Trend:**

Double Exponential Smoothing with an exponential trend.

* Using the stats model library in Python, our team imported the Simple Exponential Smoothing and Holt for fitting a double exponential smoothing model for the time series.
* From the below image, we can observe that after fitting the double exponential smoothing model, prediction is made for the behavior of the stock price.
* After fitting the model, we were also able to interpret the metrics and found that:
  + Mean Squared Error - 17.09
  + Root Mean Squared Error (RMSE) - 4.13

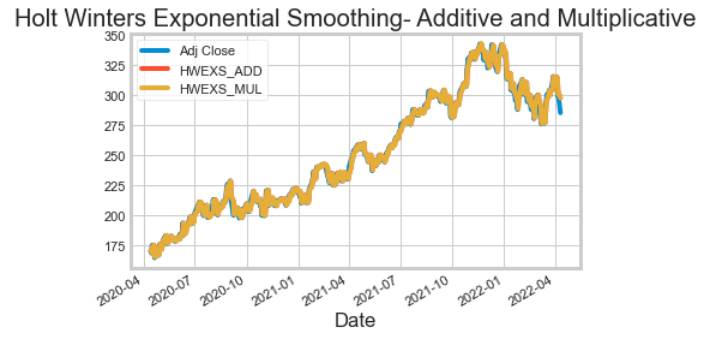


**Figure 13 - Simple Exponential Smoothing Results**



**Figure 14 - Forecasted MSFT Stock Price using Simple Exponential Smoothing**

* From the below graph, we can see the plot for the Double exponential smoothing additive and multiplicative trends.
* It can be observed that the double exponential smoothing model accurately predicts the Adj. Close price.

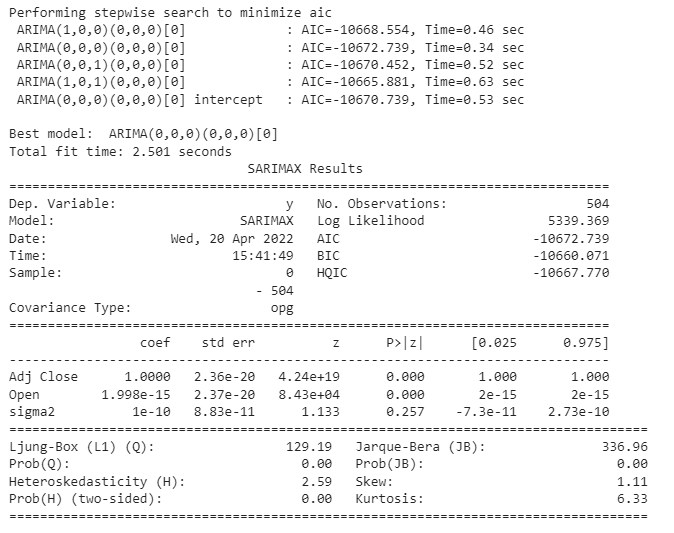
  
**Figure 15 - Holt Winters Exponential Smoothing of MSFT Price**

## **ARIMAX**

An **Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX)** model can be viewed as a multiple regression model with one or more autoregressive (AR) terms and/or one or more moving average (MA) terms. This method is suitable for forecasting when data is stationary/ nonstationary, and multivariate with any type of data pattern, i.e., level/ trend/ seasonality

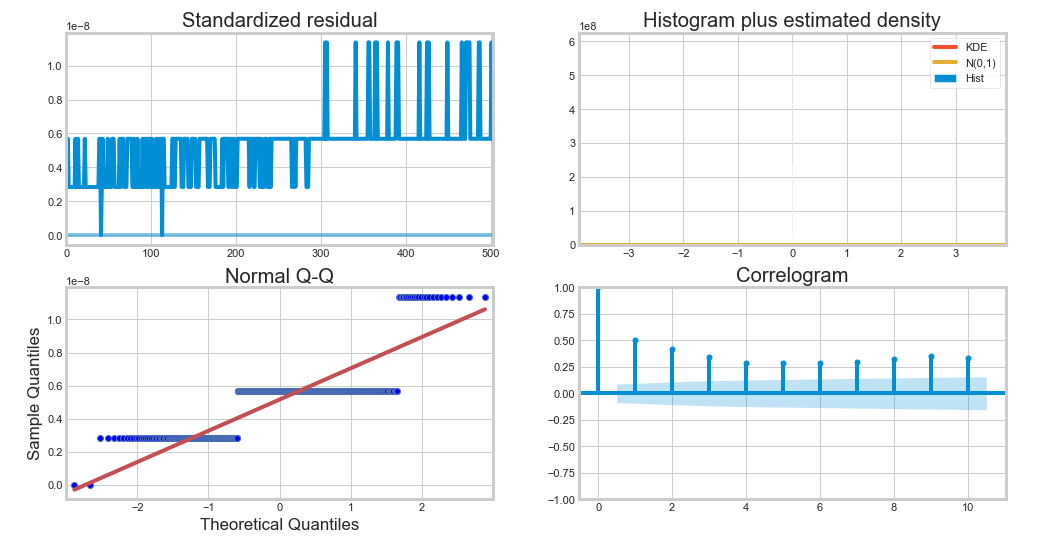
Our team has considered open price as proxy for exogenous variable

* The below statistics have been deduced after fitting the ARIMAX model.

****

**Figure 16 - ARIMAX Model Results**

* After finding out the statistics, we found out the error metrics for the model and found out that the Root Mean Squared Error (RMSE) value for the model is 88.269.
* The following diagnostics plots were plotted for the ARIMAX Model.

****

**Figure 17 - Interquartile Graphical Analysis in Python**

* The below graph shows the plot between closing price and the dates for the original time series and the predicted time series.
* The Addition of new variables disrupt the fit of the model.
* New external variables should be identified and incorporated into the model.

****

**Figure 18 - Projections Vs Actuals MSFT Stock price**

# **Business Insights**

As we stated earlier in our problem statement, stock prices are highly volatile in nature and are dependent on various factors. Even the stock specialists often fail to predict the stock market. However, for investors, it is essential to maximize their return on investments (ROI). In our analysis, we used recent past data for Microsoft Stock Price (MSP) from Yahoo to forecast the stock price.

For our investors, we recommend using the predictions from our final model - Double Exponential Smoothing (lowest Root Mean Square Error (RMSE) score of 4.13). Besides this, investors can find the patterns from the MSP stock variations to identify the trends for better investment planning. Although we cannot rely entirely on the model predictions of stock and trends/ patterns, there are some external factors we recommend investors to consider while making a decision. We have enlisted few factors that are critical to investors for better planning -

**Sentiment Analysis -** Stock sentiment analysis can be used to determine investors’ opinions of a specific stock or asset. Sentiment may at times hint at future price action. For example - Cristiano Ronaldo (famous football player) while speaking to the media in the Euro 2020 cup press conference, he removed the Coca-Cola bottles from the podium. The day after this, Coca-Cola experienced a drop in their market value by $4 Billion dollars. This is an example of how trading psychology can affect a market, assisting as a forecasting tool to determine possible future price changes in a particular asset. MSP also faced similar consequences where it was reportedly in focus following a complaint about anti-competitive behavior involving its cloud unit, Azure. While Microsoft has largely been able to stay out of the antitrust spotlight over the last few years, MSFT stock may now be getting roped into that discussion. Microsoft is reportedly combining Office and Windows together to move customers away from other large cloud companies such as Google Cloud and Amazon Web Services. While this might not seem like a big deal, critics are comparing it to Microsoft’s antitrust issues from the early 2000s, when it used similar business practices.

There are various other factors that influence stock sentiment, which include news (economic, political and industry related) and social media. These factors help influence stock sentiment as they impact stock market volatility, trading volume and company earnings.

**External Investments -** MSFT were able to maintain their stock price with multiple investments from time to time. Microsoft shocked the tech and gaming world on January 18th when it announced it would acquire Activision Blizzard in a $68.7 billion deal, by far the biggest ever in gaming. Microsoft will become the world’s third-largest gaming company by revenue, behind Tencent and Sony. Gaming is the most dynamic and exciting category in entertainment across all platforms today and will play a key role in the development of metaverse platforms. This indicates Microsoft’s ability to manage and maintain their stock price. Investors can take into consideration the strategy of Microsoft for better investment planning.

**Unprecedented Crisis - Covid19 -** While the world was facing the Covid19 pandemic in April 2020, many organizations faced a drop in their stock prices. However, Shares of Microsoft Corporation (MSP) increased another 11.1% in June, according to data provided by S&P Global Market Intelligence. On June 17, the company declared a regular quarterly dividend, a reminder its business is still thriving even during tough economic times. With investments in personal computers, cloud, and video games, Microsoft continued to excel even while people were sheltering at home during the coronavirus pandemic. Unsurprisingly, its stock has greatly outperformed the S&P 500 during this time and over the past year.

**Inflation Rate -** Rising inflation can be costly for consumers, stocks and the economy. Value stocks perform better in high inflation periods and growth stocks perform better when inflation is low. Stocks tend to be more volatile when inflation is elevated.Microsoft Corp. fell sharply with the software company leading a decline among mega cap internet and technology stocks. The shares fell about 4.1% in their biggest intraday percentage decline since October 2020. The day’s drop erased more than $100 billion off its market valuation. Even with the decline, however, Microsoft remains up more than 45% this year.

**War Crisis -** Recent war crisis between Russia and Ukraine in Mar 2022 also led to change in policies for Microsoft. It is suspending all new sales of products and services in Russia, in a major addition to the flood of companies cutting ties over the invasion of Ukraine. Microsoft has maintained its reputation with its compelling decisions. Investors can perfectly rely on MSP stocks investments under such circumstances.

Time series forecasting is a very intriguing field to work with as we have realized during the phase of our project. There is a perception in the community that it’s a complex field, and while there is a grain of truth in there, it’s not so difficult once you get the hang of the basic techniques.

# **References -**

1. Data Source: [Yahoo Finance](https://finance.yahoo.com/quote/MSFT/)
2. Sentiment Analysis - Cristiano Ronaldo Example - <https://www.si.com/soccer/2021/06/16/coca-cola-shares-drop-four-billion-cristiano-ronaldo-snub>
3. Microsoft Acquisition of Blizzard -   
   <https://news.microsoft.com/2022/01/18/microsoft-to-acquire-activision-blizzard-to-bring-the-joy-and-community-of-gaming-to-everyone-across-every-device/>
4. Microsoft during Covid19 pandemic -   
   <https://www.thesoftwarereport.com/after-years-of-growth-microsoft-is-investing-in-keeping-its-upward-trajectory-over-the-next-five-years/>
5. Linear Regression - <http://www.stat.yale.edu/Courses/1997-98/101/linreg.html>
6. ARIMAX - <https://www.elegantjbi.com/blog/what-is-arimax-forecasting-and-how-is-it-used-for-enterprise-analysis.html>
7. ARIMA - <https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average>
8. Double Exponential Smoothing - https://machinelearningmastery.com/exponential-smoothing-for-time-series-forecasting-in-python/
9. Microsoft Share price during inflation  
   [https://www.bnnbloomberg.ca/microsoft-loses-100-billion-in-value-after-inflation-data-](https://www.bnnbloomberg.ca/microsoft-loses-100-billion-in-value-after-inflation-data-1.1695776#:~:text=(Bloomberg)%20%2D%2D%20Microsoft%20Corp.,percentage%20decline%20since%20October%202020)
10. Machine learning in the Stock market <https://www.analyticsvidhya.com/blog/2020/11/stock-market-price-trend-prediction-using-time-series-forecasting/>
11. Random walks in Stock market price

Fama, E. (1995). Random walks in stock market prices. Financial Analysts Journal, 51(1), 75- 80.<http://dx.doi.org/10.2469/faj.v51.n1.1861>

1. Financial Analysis - Fama, E. F. (1995). Random walks in stock market prices. Financial analysts journal, 51(1), 75-80
2. Fama, Eugene F., 1970, “Efficient Capital Markets: A Review of Theory and Empirical Work.”
3. <https://medium.com/analytics-vidhya/a-thorough-introduction-to-holt-winters-forecasting-c21810b8c0e6>

# **Appendix**

**Abbreviations -**

* MSP - Microsoft Stock Price
* MSFT - Microsoft Stock
* EMH - Efficient Market Hypothesis
* AR - Autoregressive Technique
* MA - Moving Average
* ARIMA - Autoregressive Integrated Moving Average
* ADF Test - Augmented Dickey–Fuller test
* KPSS Test - Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests
* MSE - Mean Squared Error
* RMSE - Root Mean Squared Error
* ARIMAX - Autoregressive Integrated Moving Average with Explanatory Variable
* ROI - Return on Investments

**Figures and Tables Descriptions -**

* Figure 1 - MSFT Data Set
* Figure 2 - Statistical Analysis of the Data
* Figure 3 - Correlation Analysis among variables
* Figure 4 - Outlier Analysis using Box plot
* Figure 5 - Stock trading volume Time Series Trend
* Figure 6 - Linear Regression Model Results
* Figure 7 - Adjusted stock closing price trend with date
* Figure 8 - Actual vs Predicted stock price plot for MSFT
* Figure 9 - MSFT Stock Price Analysis Time Series Trend
* Figure 10 - Time Series Analysis using Python Visualization
* Figure 11 - Rolling Mean and Standard Deviation analysis of MSFT
* Figure 12 - ADF Test and KPSS Test Analysis Results in Python
* Figure 13 - Simple Exponential Smoothing Results
* Figure 14 - Forecasted MSFT Stock Price using Simple Exponential Smoothing
* Figure 15 - Holt Winters Exponential Smoothing of MSFT Price
* Figure 16 - ARIMAX Model Results
* Figure 17 - Interquartile Graphical Analysis in Python
* Figure 18 - Projections Vs Actuals MSFT Stock price