PEOPLEIogy
TRANSFORMING NATION THROUGH PEOPLE

Forecasting Market Volatility for GBP/USD Using Time Series Analysis (TSA)

Fahmi Roslan

What is Volatility?

“Volatility is the frequency and magnitude of price movements, ups and down. The bigger and more frequent the price swings, the more volatile the market”.

There are few factors that influence volatility:

- Timing : it can be hours or days that there will be more participants
- Sentiment: such as war
- Government policies: Announcement of Interest rate etc

Note – use APA style referencing for in-text citations (refer to link above)

How volatility being measured?

In technical analysis the are volume reading for each time according to each timeframe.

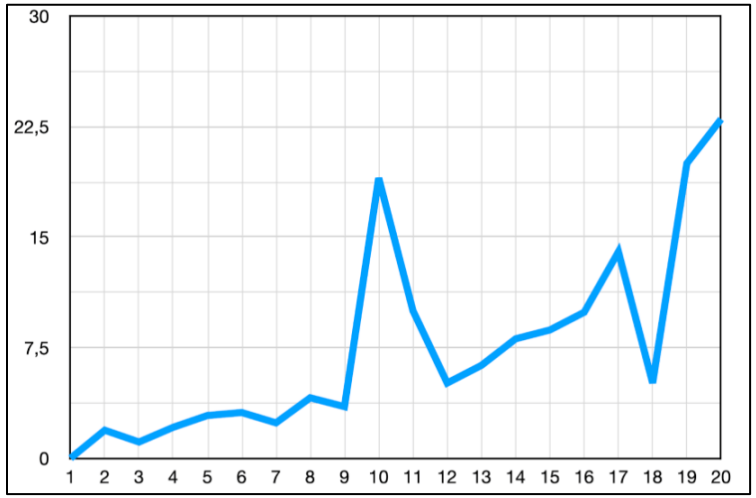
It can be measure in every instrument

every stocks, every commodities

or even every forex pairs

Note – use APA style referencing for in-text citations (refer to link above)

The Data Sets

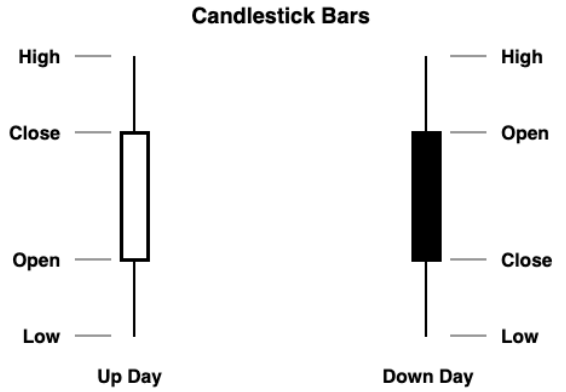


The Data : Daily Recording from 2000-2023

Data source : Kaggle

Date	Open	High	Low	Close	Volume
2000-01-03	1.6146	1.6400	1.6138	1.6361	4444
2000-01-04	1.6359	1.6415	1.6310	1.6373	6141
2000-01-05	1.6376	1.6450	1.6354	1.6419	6504
2000-01-06	1.6421	1.6511	1.6411	1.6470	6473
2000-01-07	1.6476	1.6499	1.6360	1.6394	4754

As mentioned earlier the volume is the readings of participant. What is OHLC:



OHLC refers to Open, High, Low Close level. However in this analysis we will not focus on that components

Problem Statement

- This project will only focus on single forex pair data which is GBP/USD

"To understand the long term volatility trend and forecast the short term volume to enhance better trading or investment performance"

The Best Times to Trade the Forex Markets

By TROY SEGAL Updated February 28, 2024

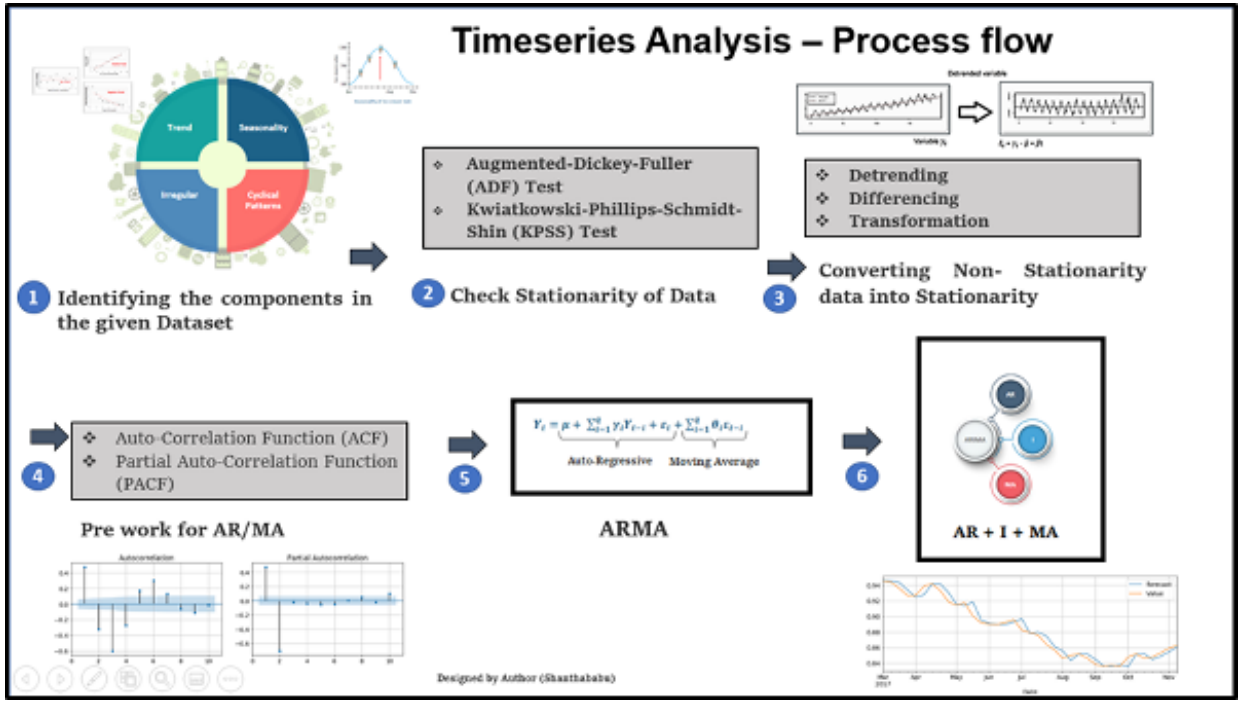
Reviewed by ERIKA RASURE

Fact checked by YARILET PEREZ

Many first-time forex traders hit the market running. They watch various economic calendars and trade voraciously on every release of data, viewing the 24-hours-a-day, five-days-a-week foreign exchange market as a convenient way to trade all day long. Not only can this strategy deplete a trader's reserves quickly, but it can burn out even the most persistent trader. Unlike Wall Street, which runs on regular business hours, the forex market runs on the normal business hours of four different parts of the world and their respective time zones, which means trading lasts all day and night.

Investopedia link : https://www.investopedia.com/articles/forex/08/forex-trading-schedule-trading-times

Time Series Analysis



- Summary of TSA components use for this capstone project:

1. Read data set and start the processing
2. Need to undergo ADF Test (Augmented Dickey Fuller test) to obtain the P-Value
3. Converting the nonstationary into stationary with eliminate the trend and seasonality by differencing method
4. Using ARIMA (Autoregression, Integration & Moving Average)
5. Check accuracy of forecast using accuracy metrics

Part 1 : Exploratory Data (EDA)

1) Data Processing

- Importing data
- Identify components within dataset
- Set up the index

```
In [7]: import numpy as np # linear algebra
        import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
        import seaborn as sns
        import matplotlib.pyplot as plt
```

```
In [8]: # Load data
        df = pd.read_csv('gbpusd.csv')
        df.head()
```

Out[8]:

	Date	Open	High	Low	Close	Volume
0	2000.01.03	1.6146	1.6400	1.6138	1.6361	4444
1	2000.01.04	1.6359	1.6415	1.6310	1.6373	6141
2	2000.01.05	1.6376	1.6450	1.6354	1.6419	6504
3	2000.01.06	1.6421	1.6511	1.6411	1.6470	6473
4	2000.01.07	1.6476	1.6499	1.6360	1.6394	4754

```
In [10]: # Change date column to be datetime dtype
         df['Date'] = pd.to_datetime(df['Date'])
```

```
In [11]: # Set Date to be in the index
         df.set_index('Date', inplace=True)
```

```
In [12]: df.head()
```

Out[12]:

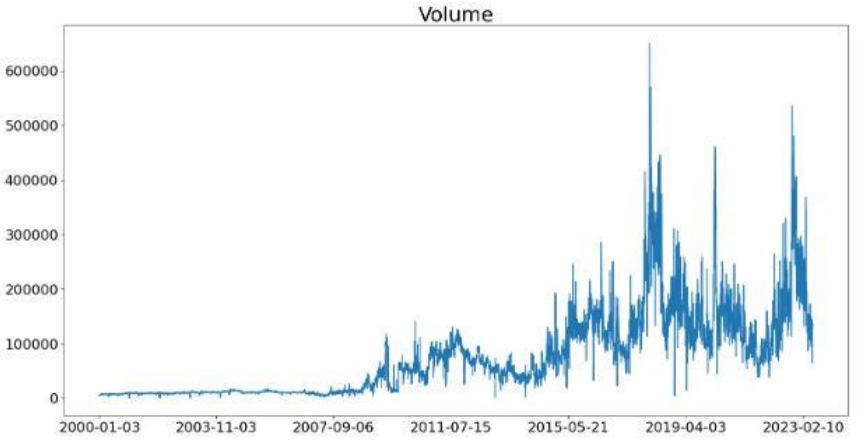
Date	Open	High	Low	Close	Volume
2000-01-03	1.6146	1.6400	1.6138	1.6361	4444
2000-01-04	1.6359	1.6415	1.6310	1.6373	6141
2000-01-05	1.6376	1.6450	1.6354	1.6419	6504
2000-01-06	1.6421	1.6511	1.6411	1.6470	6473
2000-01-07	1.6476	1.6499	1.6360	1.6394	4754

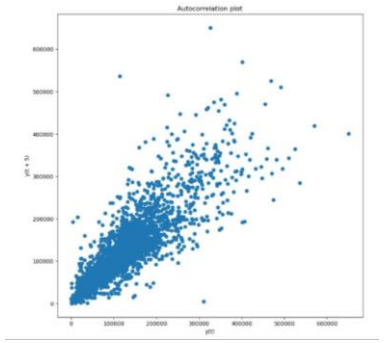
Part 1 : Exploratory Data (EDA)

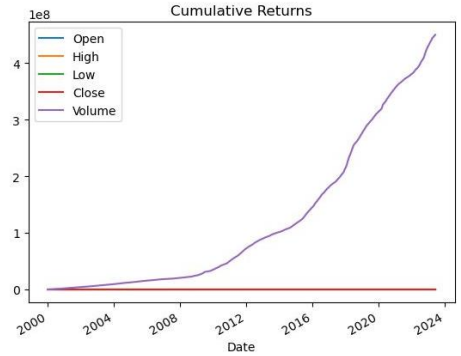
1) Check Stationary of Data

- Plotting the visualization of chosen variable which is 'Volume'
- Check in the visual data (mean/variance/covariance) over time
- Plot autocorrelation of the variable

```
In [15]: # Generate a time plot of our data.
         plot_series(df, ['Volume'], title = "Volume", steps=1000)
```







Part 1 : Exploratory Data (EDA)

```
In [10]: # Import Augmented Dickey-Fuller test.
         from statsmodels.tsa.stattools import adfuller

         # Run ADF test on original (non-differenced!) data.

         # Import Augmented Dickey-Fuller test.
         from statsmodels.tsa.stattools import adfuller

         # Run ADF test on original (non-differenced!) data.

         adfuller(df['Volume'])

Out[10]: (-2.943655895014861,
          0.0405018542219686,
          34,
          6044,
          {'1%': -3.4314324089136767,
           '5%': -2.8620183258811283,
           '10%': -2.567024610317412},
          137339.01168246148)

In [11]: # Code written by Joseph Nelson.

         def interpret_dftest(dftest):
             dfoutput = pd.Series(dftest[0:2], index=['Test Statistic','p-value'])
             return dfoutput

In [12]: # Run ADF test on original (non-differenced!) data.
         # Run ADF test on original (non-differenced!) data.

         interpret_dftest(adfuller(df['Volume']))

Out[12]: Test Statistic   -2.943656
         p-value           0.040502
         dtype: float64
```

2) ADF Test

Test the models with ADF/KPSS test

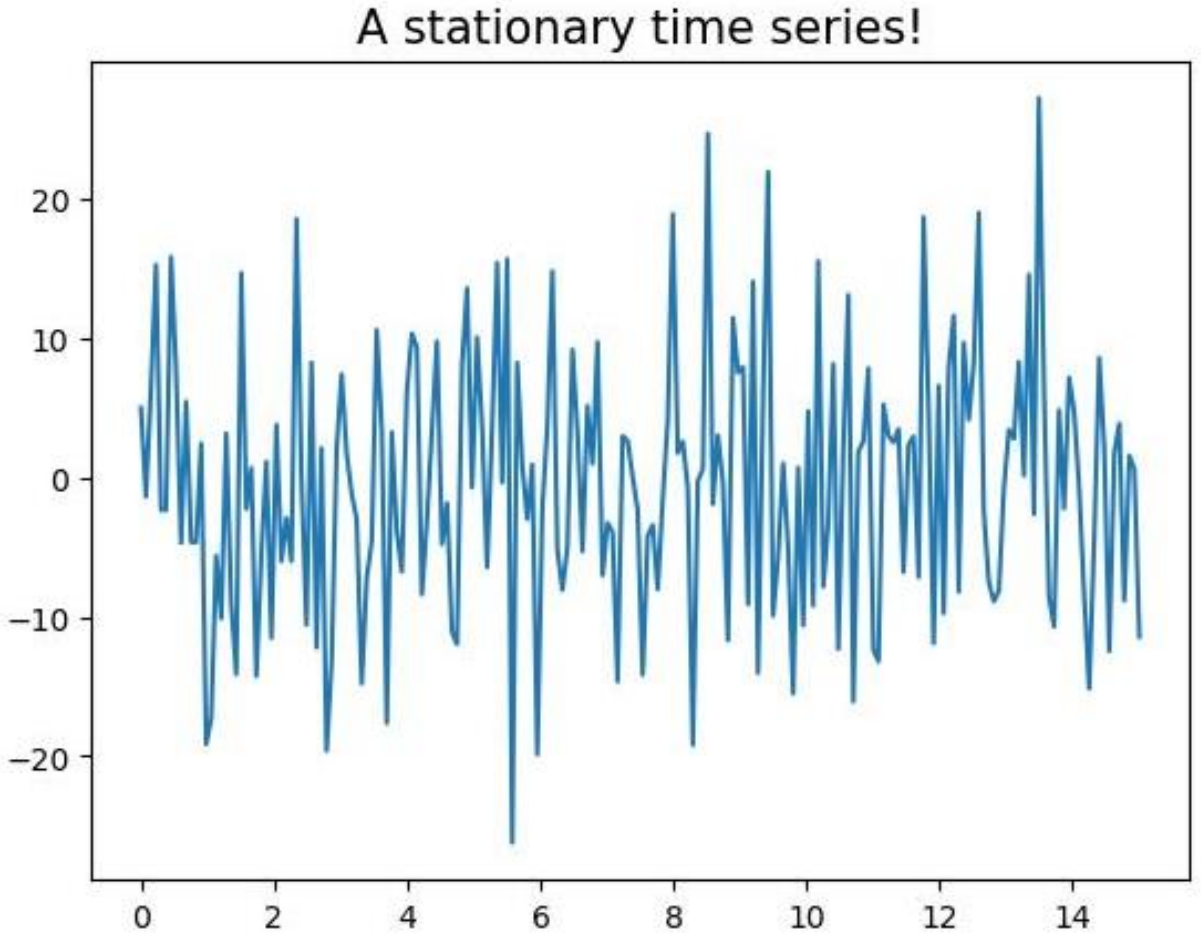
- Null Hypothesis : The series has a unit root (non stationary)
- Test statistic : -2.943656
- P-Value : 0.040502
- Number of lags : 34
- Observations : 6044

Conclusion : P-Value is smaller than 0.05, we reject the null hypothesis and the series is stationary. Hence the steps as below is not compulsory and the modelling can be start.

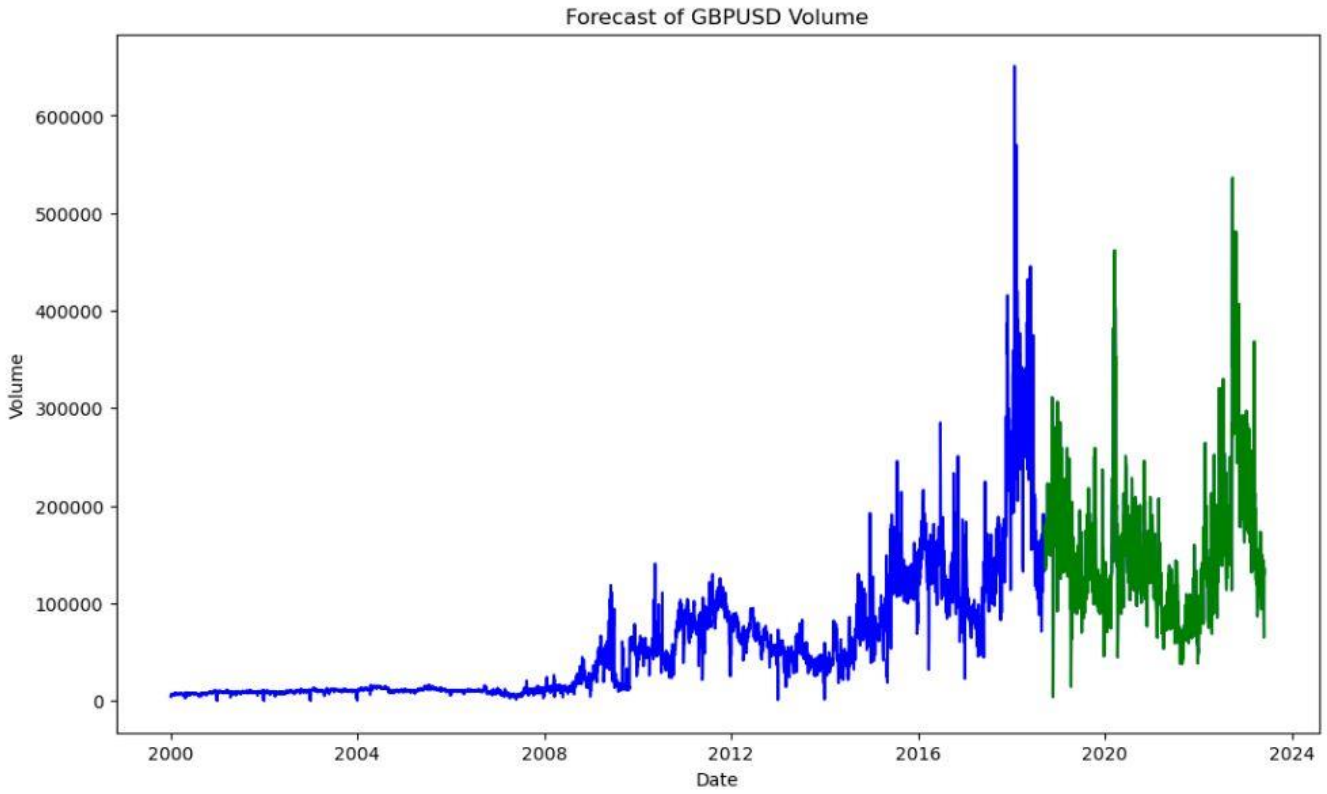
Part 2 : Data Modelling

Modelling has been carried as steps below:

- Plot raw data
- Fitting Model
- Identify parameters
- Manual GridSearch
- Get the p,d,q value of AR1MA



Part 2 : ARIMA



Plotting ARIMA (where green is the prediction on volume) :

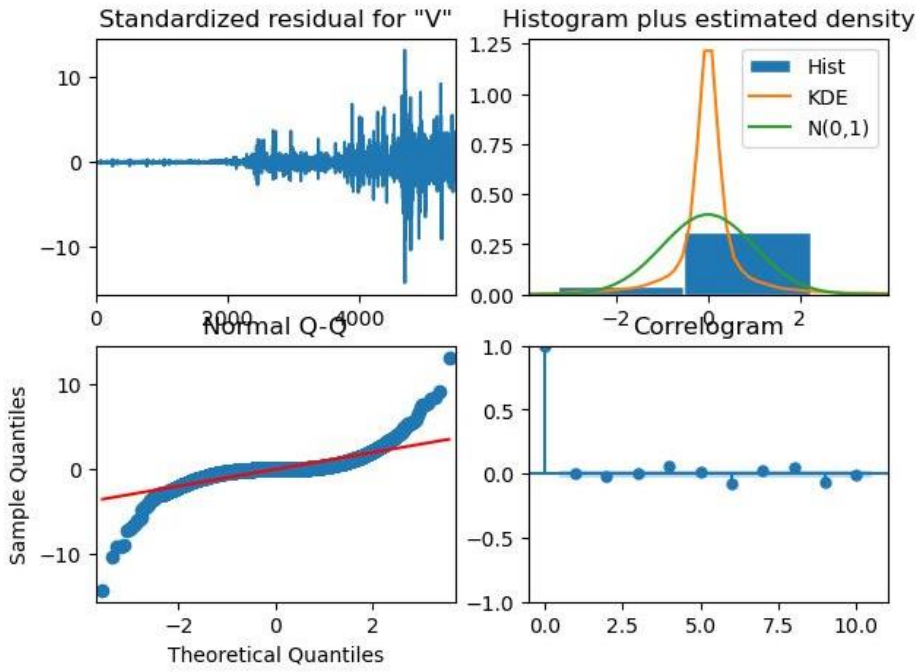
Based on manual GridSearch, we could find the the parameters and value of p-d—q.

- The series model that minimizes AIC on the training data is the ARIMA(4,1,4)
- p = 4 # Autoregressive order
- d = 1 # Differencing order
- q = 4 # Moving average order

The AIC for ARIMA 4,1,4 is; 123274.08814598958

Part 3 : Model Evaluation (Residuals)

Evaluation based on residuals:



- Top left: The residual errors seem to fluctuate around a mean of zero and have a uniform variance.
-
- Top Right: The density plot suggest normal distribution with mean zero.
-
- Bottom left: All the dots should fall perfectly in line with the red line. Any significant deviations would imply the distribution is skewed.
-
- Bottom Right: The Correlogram, aka, ACF plot shows the residual errors are not autocorrelated. Any autocorrelation would imply that there is some pattern in the residual errors which are not explained in the model. So you will need to look for more X's (predictors) to the model

Part 3 : Model Evaluation (Metrics)

I have carried on few evaluation metrics to measure accuracy of this model as below:

- **Root Mean Squared Error (RMSE)**: This is the square root of the MSE and provides an interpretable measure of error in the same units as the original data.
- RMSE = sqrt(MSE)
- **RMSE: 20.73247619699088**

- **Mean Absolute Percentage Error (MAPE)**: This measures the average absolute percentage difference between the actual and predicted values, making it easy to understand the magnitude of error relative to the actual values.
- MAPE = Σ(|Actual - Predicted| / Actual) * 100 / n
- **# MAPE: 228.33332232842398**

- **Mean Absolute Error (MAE)**: This is the average of the absolute errors between the predicted values and the actual values. It provides a simple and interpretable measure of accuracy.
- MAE = Σ|Actual - Predicted| / n
- **# MAE: 49.99999601481055**

- **Mean Absolute Scaled Error (MASE)** : is a metric used to measure the accuracy of forecasts, particularly in time series forecasting, by comparing the forecast errors to a naive forecast. MASE is robust to different scales and is useful for comparing the accuracy of different forecasting methods on different datasets.
- **# MASE: 1.5584414342278614**

Findings

- After evaluating model I can conclude that this model is not very good based on my the accuracy metrics. My scale of data is quite a big range and the RMSE is not very near to 0. I decided to also use MAPE which involves several considerations, similar to evaluating other error metrics and resulting in high percentage (considered as negative reading.
- Hence, my problem statement cannot be answered using this model at this stage. However, the positive positive finding of this Time Series Analysis is actually very significant which is;
- **To develop a baseline model to be the anchor for my next time series forecasting model that can be use in forecasting any financial instrument over time. With this model I can assess different future developed model to enhance accuracy.**

Model Improvement

These are few ways to improve my next model with better accuracy:

- Data processing
- Assessing Autocorrelation
- Model Selection (SARIMA, LSTM etc)
- Parameter Tuning
- Cross Validation

By implementing these strategies and iteratively refining your models, you can improve the accuracy of your time series forecasts and make more informed decisions based on the predictions. Remember that achieving high accuracy may require a combination of data preprocessing, model selection, and careful parameter tuning.

Significant of this TSA

- I think there's no need for introduction on the audience of financial markets.
- I to I find it very interesting to go deep into Time Series Analysis as it is a method to forecast any variables over a period of time. When I persuade this project I can see this project might be very helpful for any investors and traders either retail or working for any professional organization.
- **However this model is still at its earliest stage to be an useful tools for those people or organization. This model can only be more useful with similar and improvise successful metric.**

Thank you

Appendix