TIME SERIES MODELLING CASE STUDY

Dhavalkumar Pithadiya
ID :- 22003118

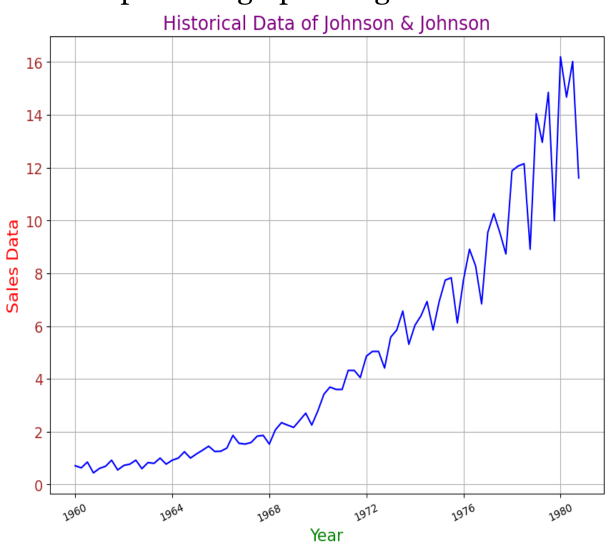
Time Series Analysis of Johnson & Johnson Sales and Amazon Stock Price

- Introduction

Nowadays, Time series analysis plays a crucial role in understanding and predicting trends in sales data and stock prices. In this report, we present a case study focusing on the time series analysis of Johnson & Johnson sales data and Amazon's closing share price data. The objective is to uncover patterns, make forecasts, and derive insights to inform strategic decisions.

First of all, we have imported all necessary libraries to simplify our code and facilitate the execution of our program.

Then we plot the graph for given historical data of j&j and Amazon which are as shown below:



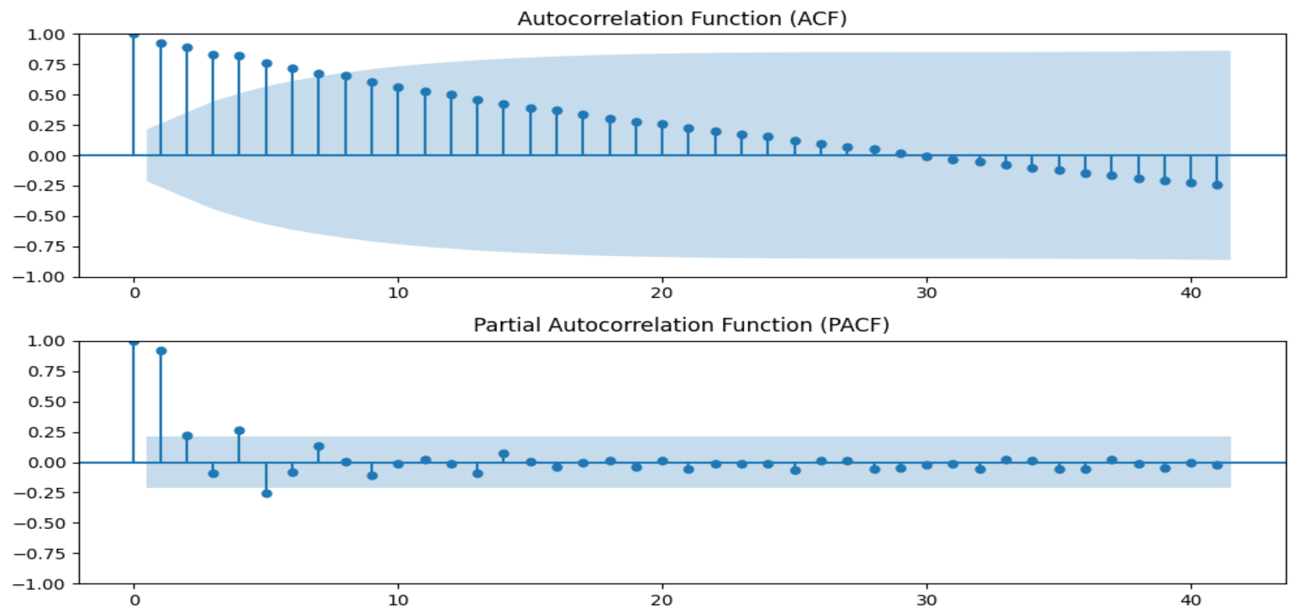
The Augmented Dickey-Fuller (ADF) test is a statistical hypothesis test used to determine whether a unit root is present in a time series dataset, which indicates the presence of non-stationarity. A unit root suggests that a time series has a stochastic trend and is not stationary. The ADF test is commonly employed in time series analysis to assess the stationarity of a dataset before applying further statistical techniques or models.

In our analysis of Johnson & Johnson sales data, we utilized the ADF test to evaluate the stationarity of the sales time series. The ADF test provides us with several key statistics to interpret the stationarity of the data and here we took different correlative function as well

The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are tools used in time series analysis to understand the relationships between observations and their lagged values. The ACF measures correlation at different lags, helping identify serial correlation, while the PACF focuses on direct relationships between observations, aiding in the selection of model terms.

In our analysis of Johnson & Johnson sales data, we used these functions to identify significant correlations, inform model selection, and improve forecast accuracy.

p-value: The p-value associated with the ADF statistic is compared to a chosen significance level (e.g., 0.05) to determine the statistical significance of the test. A lower p-value indicates stronger evidence against the null hypothesis of non-stationarity.



But we got value 1 and 0.45297 so it's indicating non-stationary hypothesis so we have to make data stationary. In order to do that we will use Boxcox Transformation so it will become stationary data.

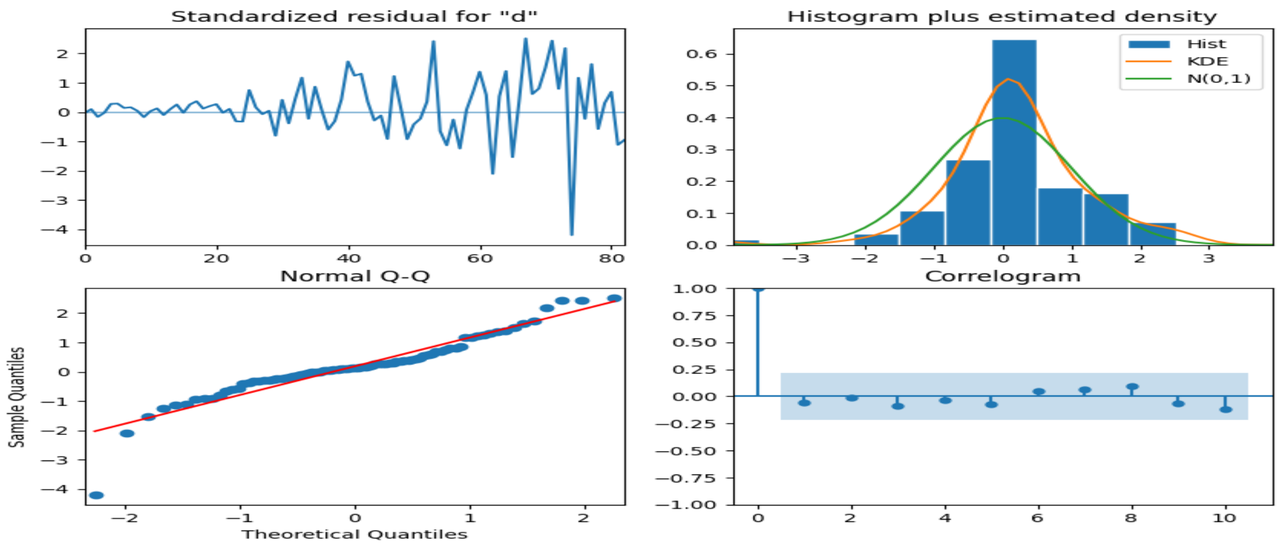
Employing the ADF test and Boxcox Transformation, we attained stationarity for Johnson & Johnson sales data (p = 0.0003200619768274673) and Amazon's share price (p = 7.63428151147939e-26), enabling robust time series analysis.

➢ ARMA Model

In my exploration of Johnson & Johnson sales data, I employed ARIMA modelling to identify the most suitable parameters for forecasting. Utilizing Python's statsmodels library, I systematically searched for the optimal combination of autoregressive (p), differencing (d), and moving average (q) terms, aiming to minimize the Akaike Information Criterion (AIC). I pursued this objective through two approaches: loop wise iteration over a predefined range of parameter values and leveraging the automated model selection capabilities of the pmdarima library. Despite encountering challenges, such as convergence issues during the loop wise iteration, I persisted and successfully uncovered the best-fitting ARIMA model configuration. This process not only enhanced my understanding of time series analysis but also sharpened my proficiency in Python programming and data analysis techniques.

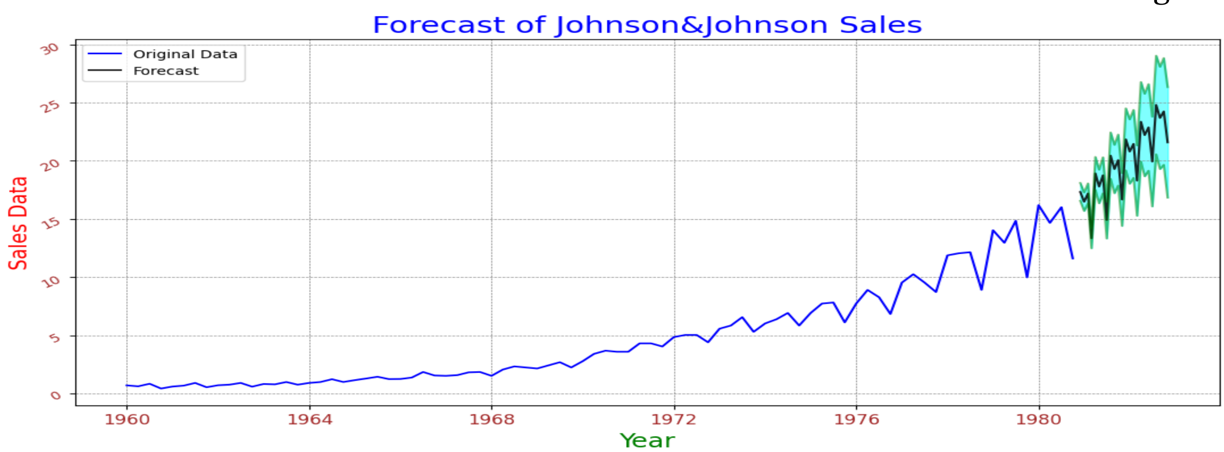
After conducting an exhaustive search for the best ARIMA model, the optimal configuration was found to be ARIMA(6, 1, 3), resulting in an AIC value of 115.30. This outcome signifies that the model with these parameters achieved the lowest AIC score among all the tested combinations, indicating its superior fit to the data compared to alternative models.

Here given graph represents the diagnostics of Johnson & Johnson sales data



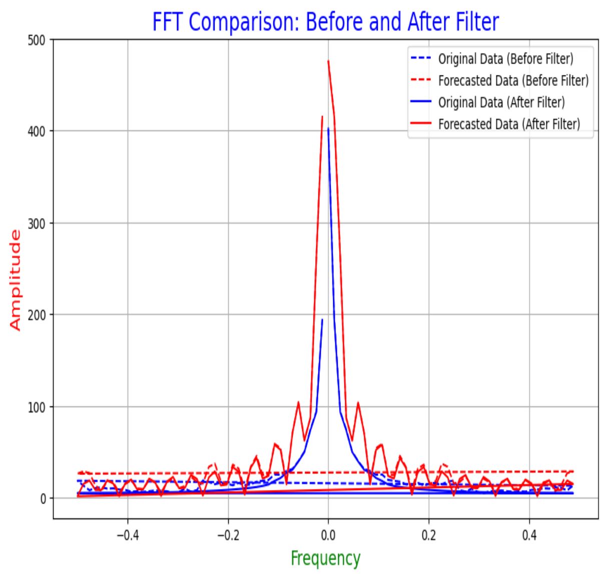
- Forecasting for 24 months

Now, we are going to make prediction for the sales data with the help of ARIMA model (with best parameters) for next 24 months. In order to make prediction I have used get_forecast() and we have also added confidence intervals to our forecasting value so the value can varies between that confidence interval. the below 2 graph represents forecasting of Johnson & Johnson Sales and Amazon's Share Closing Price.



➢ Fourier transforms

We filtered the original and forecasted data using a Butterworth low-pass filter to remove high-frequency noise and emphasize underlying trends. By setting a cut-off frequency and applying the filter, we smoothed the data while preserving important features. Comparing the FFTs of the original and filtered data revealed a reduction in high-frequency components, indicating successful noise reduction. This pre-processing step enhances signal clarity, contributing to more accurate forecasts by reducing the impact of irrelevant fluctuations.



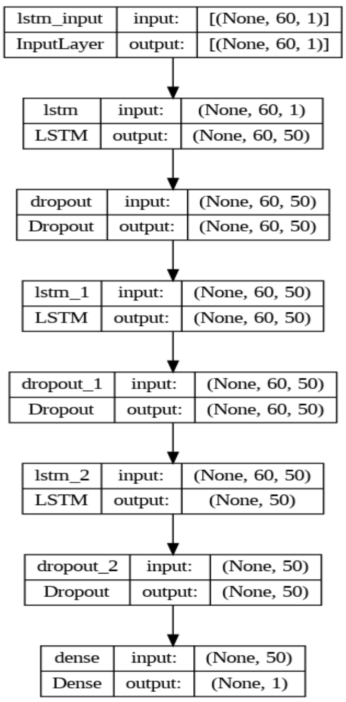
Here the graph represents the comparison before filter and after filter for original data and forecasted for the Johnson & Johnson.

We can check it out that after filter data is way clear to frequency as compare to before filter and Amplitude picks high value as well.

Furthermore, employing ARIMA and LSTM models, our analysis forecasts Amazon's closing share prices, offering valuable insights for strategic investment decisions in the stock market.

➢ Create a LSTM Model

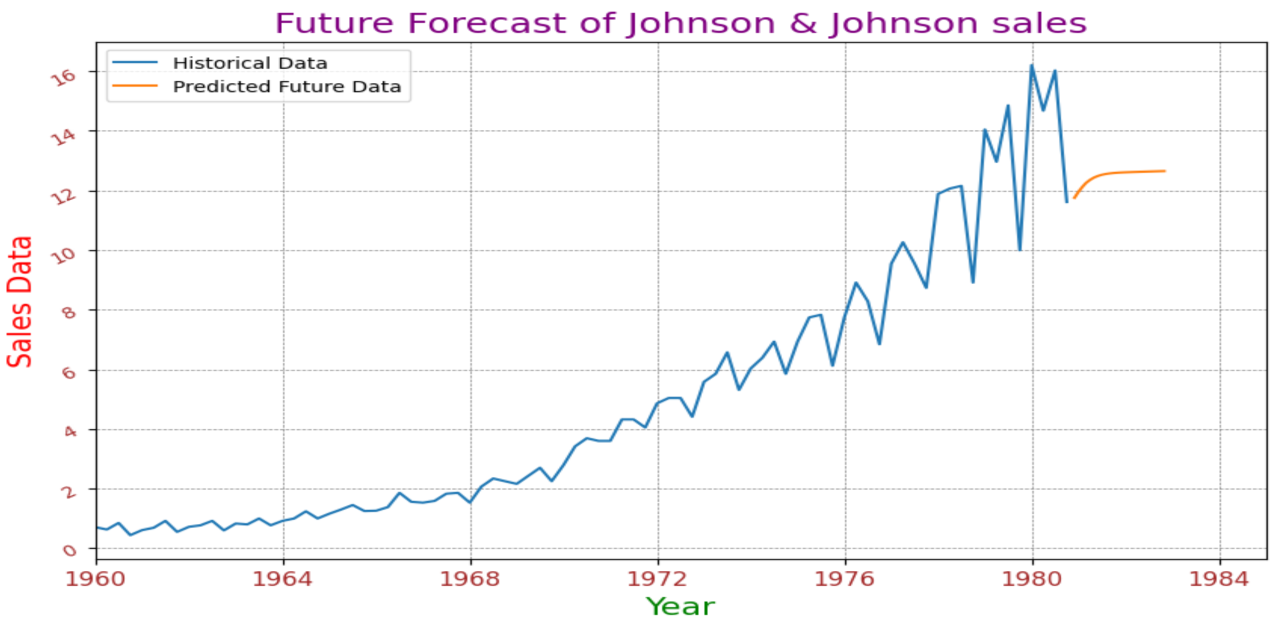
This code begins by bringing in time-based data from a CSV file and getting it ready for analysis. It does this by understanding the 'date' column as dates and arranging the data based on these dates. Then, it adjusts the data so that all features are on the same scale using a method called MinMaxScaler. After that, it organizes the data into chunks: each group contains 60 consecutive data points as input, and the next single data point as output.

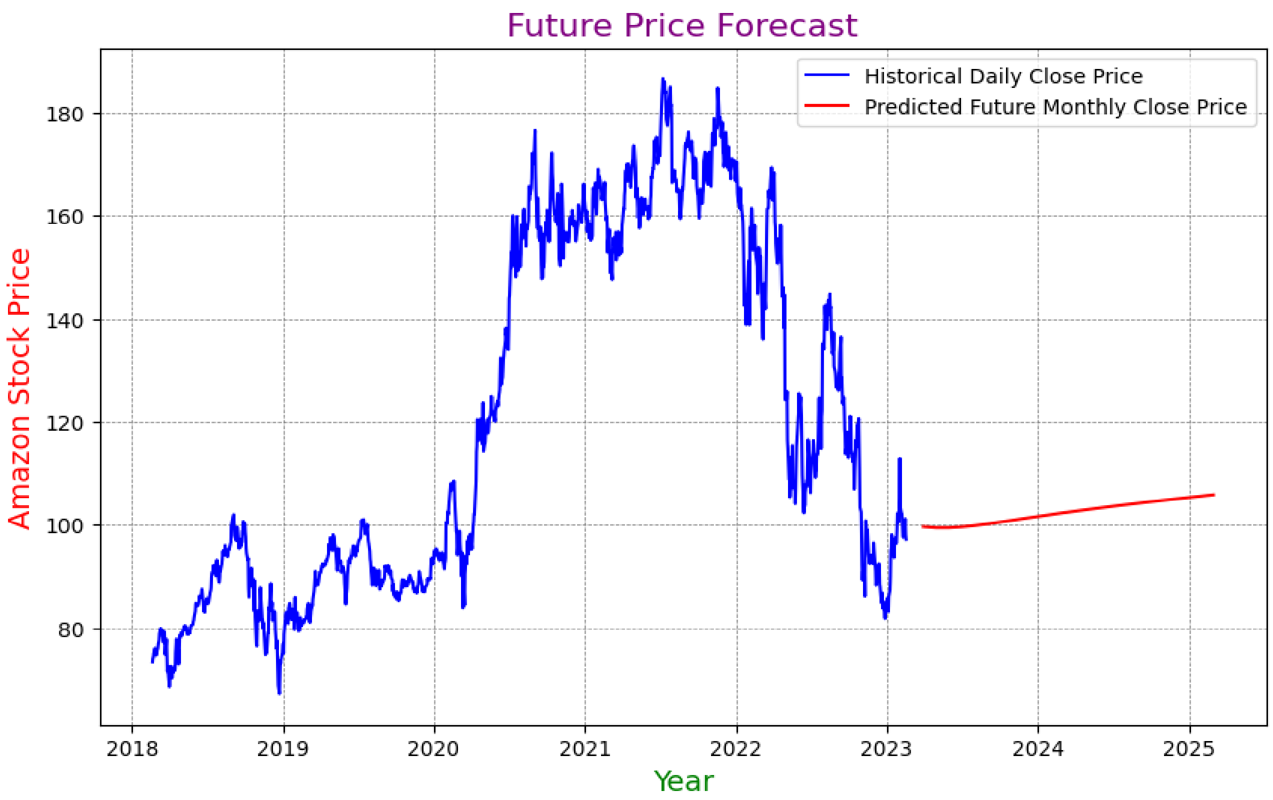


In this code, the LSTM model trained earlier is used to make predictions on both the training and testing datasets. These predictions are then inverse-transformed to their original scale using the MinMaxScaler. Root Mean Squared Error (RMSE) is calculated to evaluate the model's performance on both datasets, as well as for the entire dataset. Additionally, a future forecast is generated for 24 months ahead by iteratively predicting each future time step based on the previous predictions. The resulting forecasted sales data is then visualized alongside the historical data in a time series plot.

LSTM (Long Short-Term Memory) networks are particularly effective for time series forecasting tasks due to their ability to capture long-term dependencies and handle sequences with varying time lags. Unlike traditional feedforward neural networks, LSTM networks have memory cells that can maintain information over time, making them suitable for modelling complex temporal patterns present in time series data.

In the provided graph, the historical sales data of Johnson & Johnson and the closing share prices of Amazon are plotted alongside their respective predicted future values. This visualization facilitates a comprehensive comparison between observed and forecasted trends for both companies, offering stakeholders valuable insights for decision-making based on the model's predictions. Additionally, the root mean square error (RMSE) for the entire dataset of Johnson & Johnson sales is 2.45, indicating the model's accuracy in predicting sales data. Similarly, for Amazon's closing share prices, the RMSE for the entire dataset is 5.19, highlighting the predictive performance of the model.





- Result

In Conclusion, the report endeavours to delve into the realm of time series forecasting, focusing on two distinct datasets: sales data for Johnson & Johnson and closing share prices for Amazon. By employing a range of analytical techniques such as ARIMA modelling, LSTM neural networks, and filtering methods, the study aims to unveil the predictive potential of each approach. Through a meticulous examination of key performance indicators including RMSE, the report will not only showcase the effectiveness of these methodologies but also provide valuable insights into their comparative strengths and limitations. Ultimately, the findings are poised to offer actionable intelligence for stakeholders navigating the dynamic landscape of financial forecasting and strategic decision-making.