# Tesla Stock Price Prediction

# Project Overview


## Abstract:

This report explores the usage of ARIMA, SARIMA, and LSTM neural networks to predict the opening stock prices of Tesla Inc. By integrating traditional statistical models with cutting-edge deep learning techniques, we aim to significantly enhance the predictive accuracy and reliability of financial market forecasts, addressing the complexities inherent in volatile stock price movements.


## Introduction:

Accurately predicting stock prices is a formidable challenge due to the dynamic and unpredictable nature of financial markets. This project adopts a hybrid modeling approach, combining ARIMA and SARIMA models—renowned for their statistical accuracy in capturing linear relationships and seasonality - with LSTM networks, which are adept at understanding long-term dependencies in sequential data. This combination allows for a robust analysis of both linear and non-linear patterns, providing a comprehensive tool for financial forecasting.

## Objective:

The primary objective of this analysis is to develop a predictive model using time series techniques like ARIMA, SARIMA, and LSTM techniques. Our goal is to refine investment decision-making processes by providing high-accuracy predictions of Tesla's stock prices, thereby aiding investors in navigating the complexities of the stock market more effectively.

# Methodology

## Data Collection:

*Source and Integrity:* We procured an extensive dataset of Tesla's stock prices from the 'YFinance' module (Yahoo Finance), spanning from April 15, 2014, to April 11, 2024. This dataset is characterized by its completeness, with no missing values across all essential stock attributes, ensuring a reliable basis for our analysis.

```
Open         0
High         0
Low          0
Close        0
Adj Close    0
Volume       0
dtype: int64
0
Open         float64
High         float64
Low          float64
Close        float64
Adj Close    float64
Volume         int64
dtype: object
```

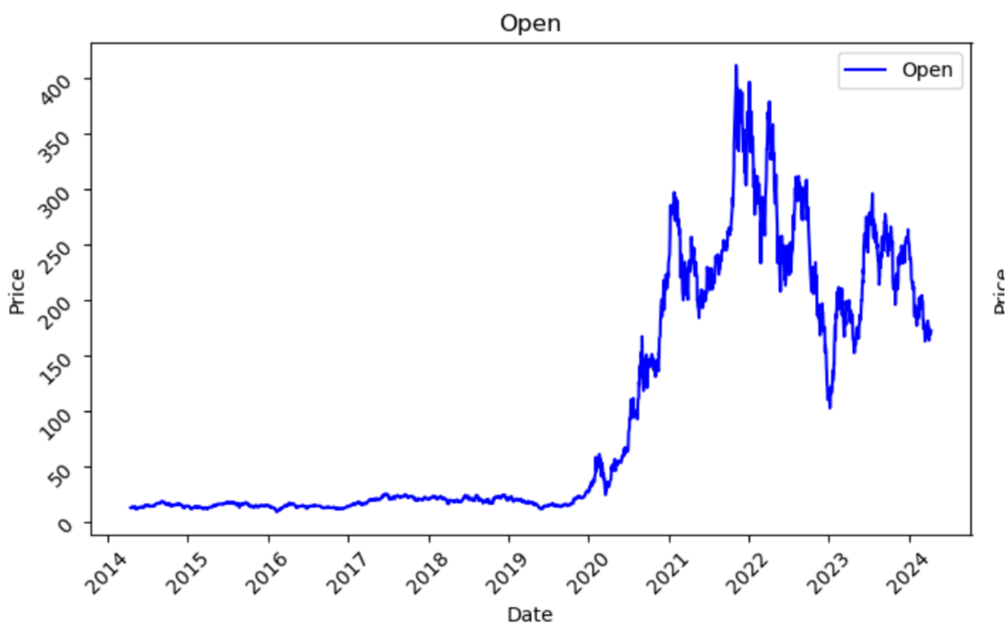
## Data Cleaning and Preprocessing:

*Date Handling***:** We converted text-based date representations into datetime objects, which facilitates more sophisticated time-series analyses by allowing for precise temporal indexing and manipulation.

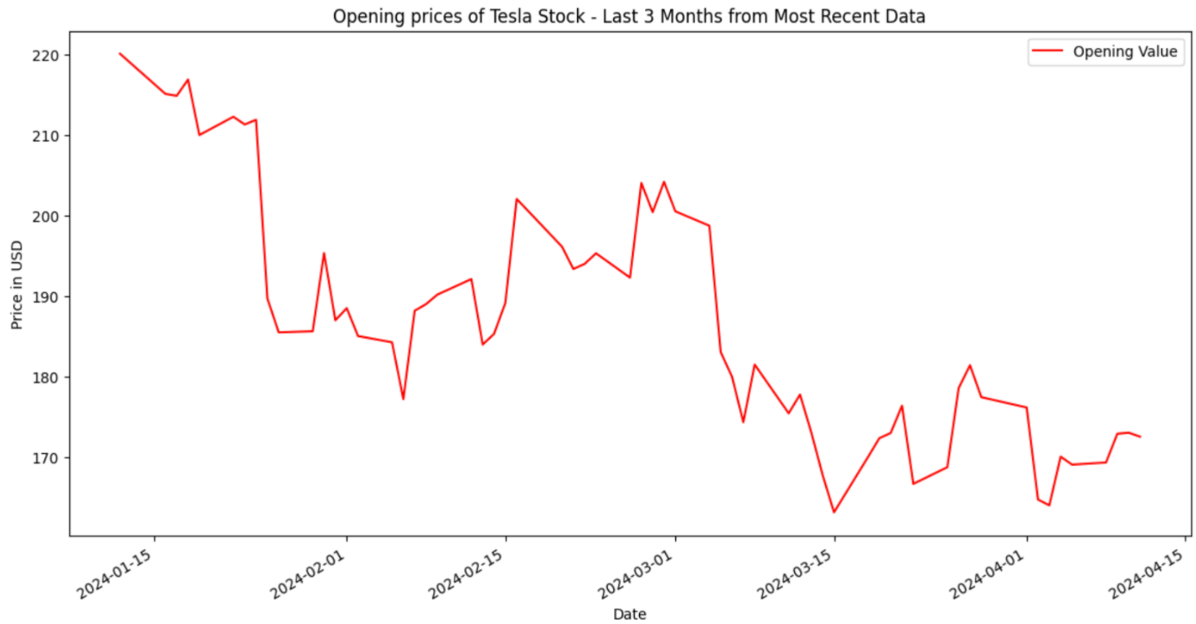
*Normalization Strategy***:** Utilizing the MinMaxScaler, we normalized financial indicators to a [0,1] range. This normalization is critical for the LSTM component of our study, as it standardizes input features, optimizing the network's ability to converge during training.

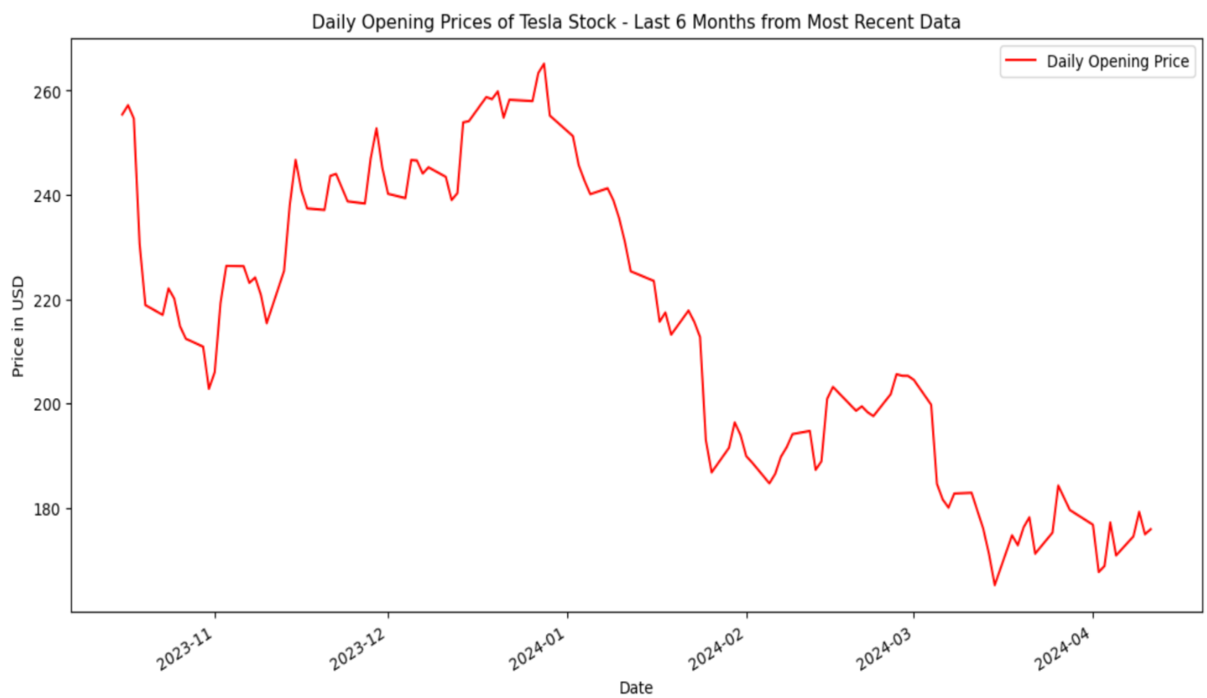
## Exploratory Data Analysis:

- Tesla's stock price experienced initial stability and a gradual uptrend from 2014-2019.
- The rapid growth phase from 2019-2021 saw a surge in price, likely due to milestones like production scaling and electric vehicle expansion.
- The high volatility and market speculation period from 2021-2023 saw intense peaks and troughs, possibly due to speculative trading and investor reactions.
- The recent downturn, post-January 2024, was influenced by market corrections, company-specific news, macroeconomic factors, and technical adjustments.
- The overall variance in Tesla's stock price reflects its unpredictable nature in a highly innovative yet competitive market.




On Plotting Tesla’s opening prices over last 3 months and last 6 months,

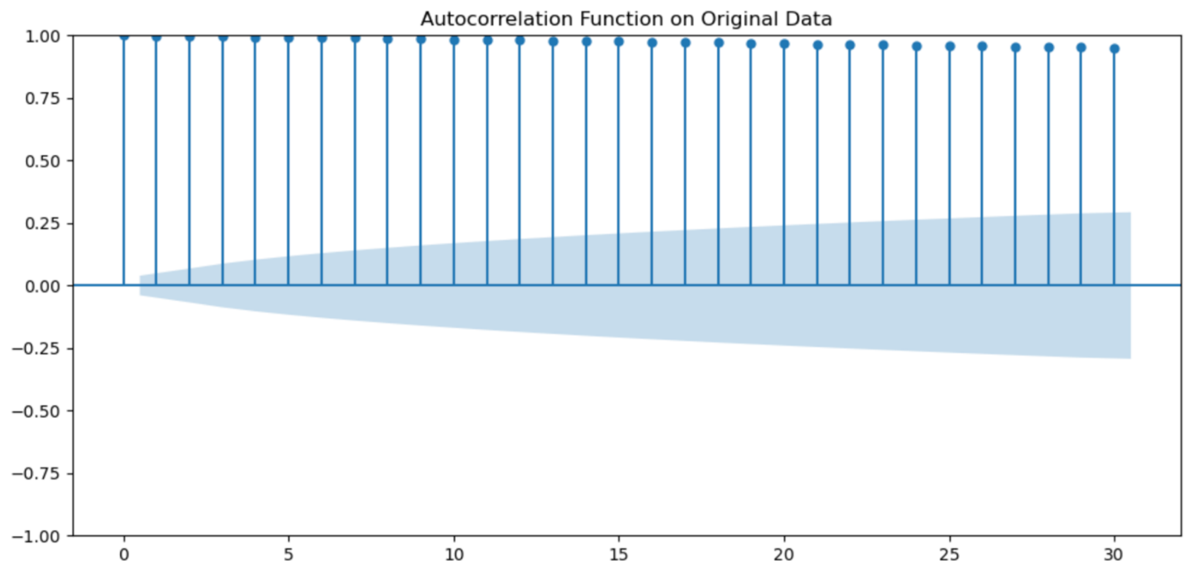


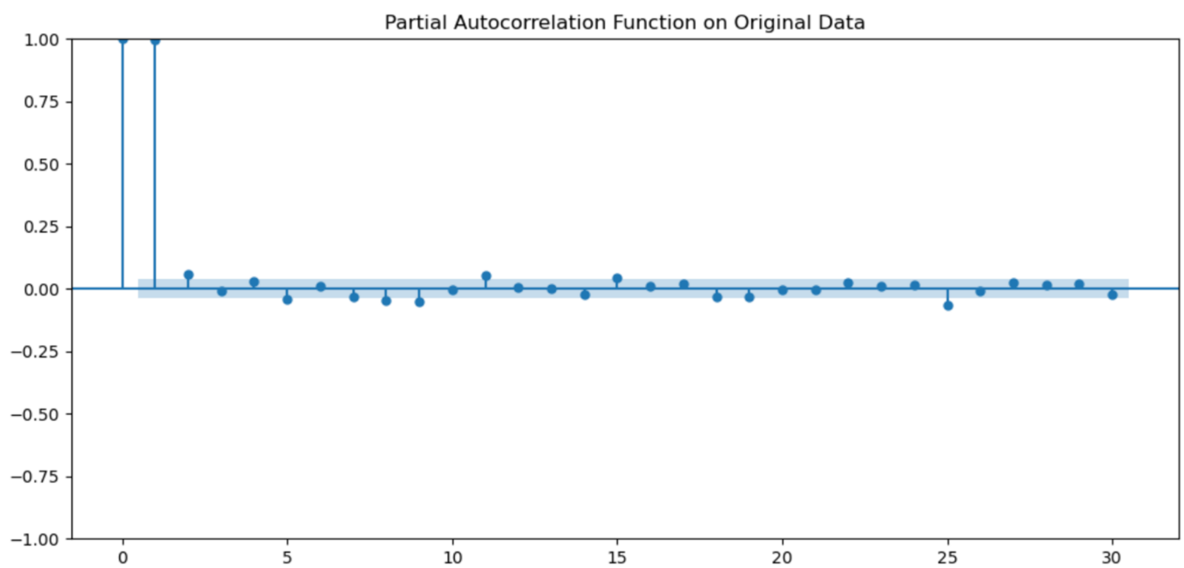




Over the past three months, Tesla's stock prices have consistently demonstrated a downward trend. In a broader six-month analysis, stock prices initially showed an increasing trend, peaking in January 2024. However, post-January, a clear decline is observed, punctuated by intermittent spikes that suggest temporary recoveries but generally align with an overall downward trajectory.

The observed variance in stock prices, characterized by fluctuations between high and low values during these periods, likely reflects the impact of multiple factors. These include Tesla's internal company performance, significant economic news, and shifts in broader market trends. Such variability is indicative of the volatile nature of the stock market, which can be influenced by a range of external conditions and events.

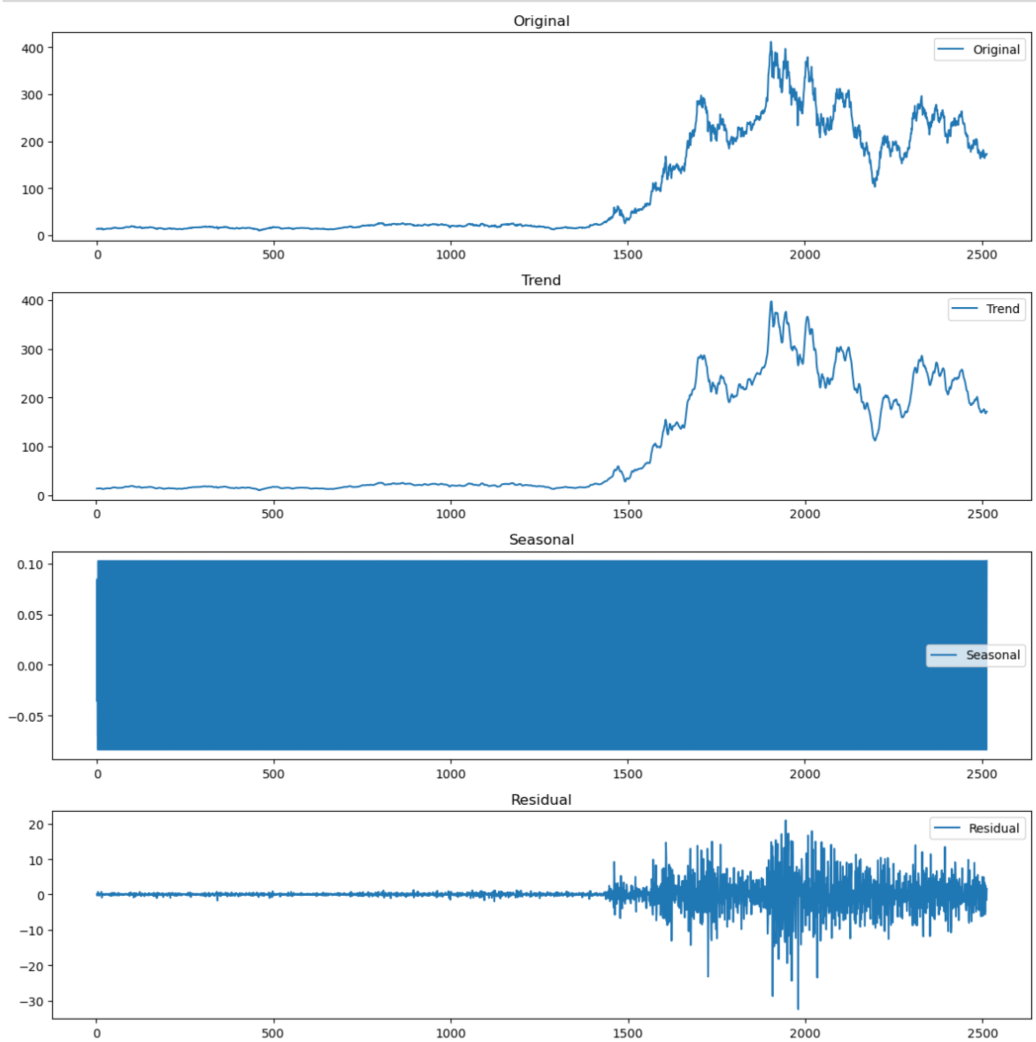
***Statistical Diagnostics*****:** Our initial ACF and PACF plots revealed non-stationarity, which we addressed through differencing, transforming the series into one that displays stationary characteristics suitable for ARIMA-based modelin**g.**



- The ACF plot tails off slowly and shows a gradual decline as the lags increase, but there's a significant correlation at each lag. This could suggest a non-stationary series.
- PACF plot has a sharp drop after lag 1, with other lags' partial correlations being insignificant

***Decomposition Analysis****:* We applied seasonal decomposition techniques to parse out underlying trends, seasonal effects, and irregular components within the data, finding minimal seasonal influences but notable short-term irregularities likely tied to market-specific events.

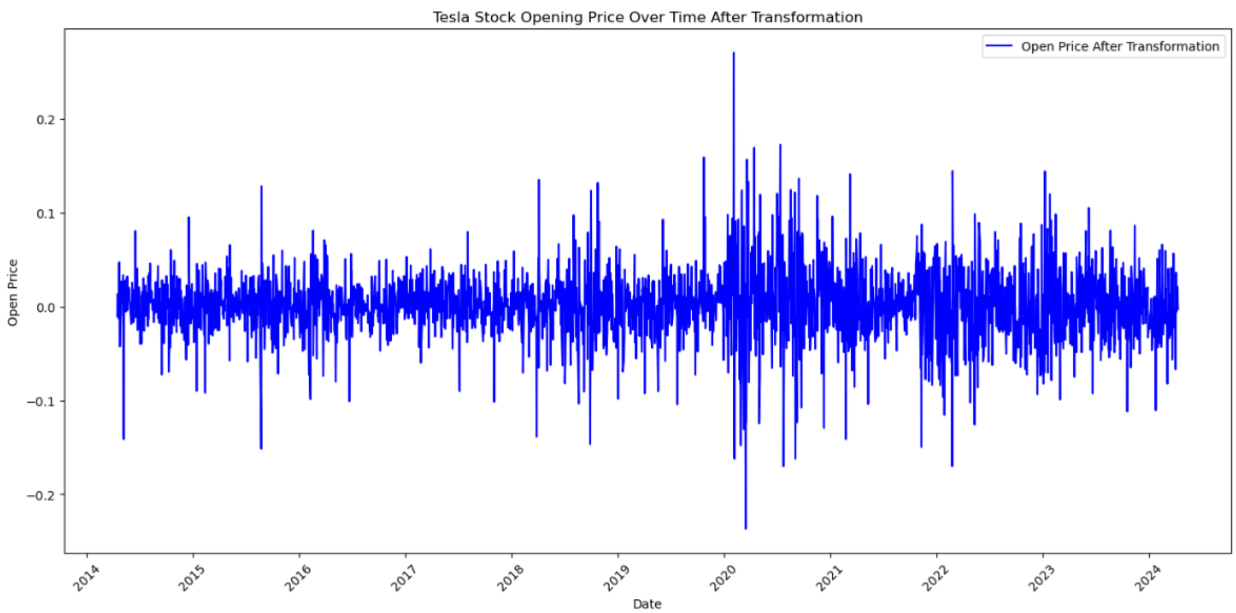



**Seasonality**:

- It appears to be a constant line with no visible fluctuation.
- Sometimes, Certain financial series like stock prices, may not exhibit clear seasonal patterns due to being influenced more by irregular events.

**Residual graph**: It shows variability and is likely due to short-term market events.

## Transformation for Stationary Data: Log Differencing

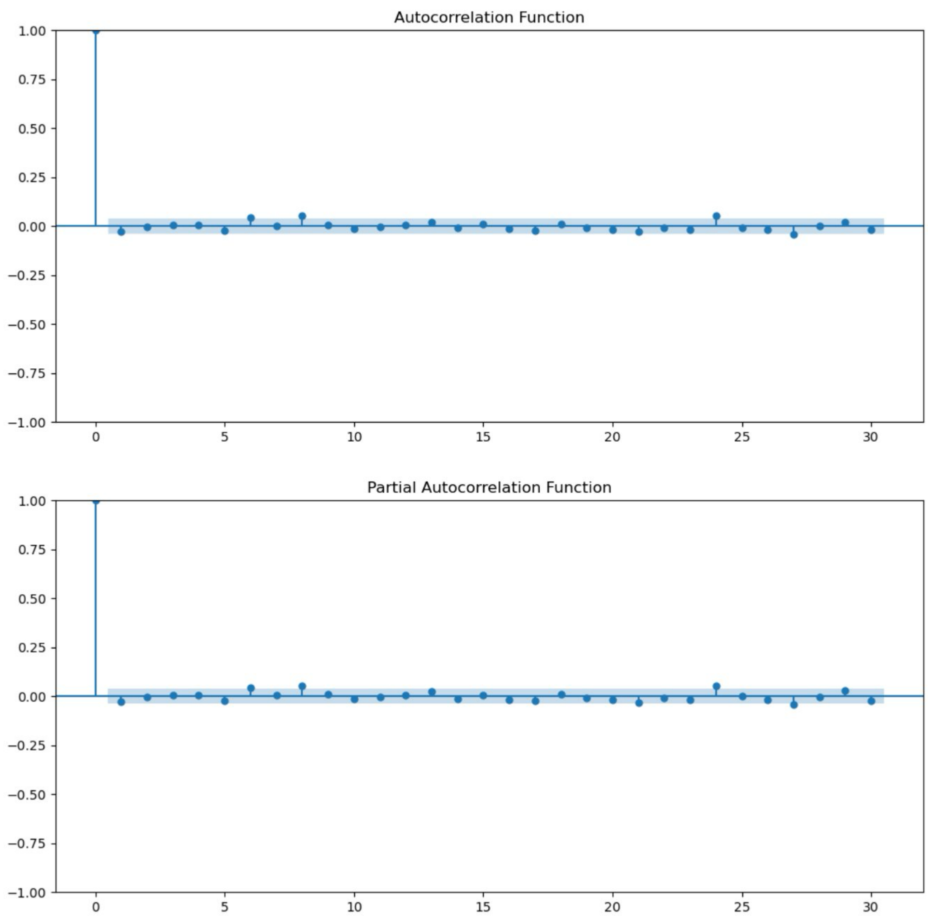



- Applied logarithmic transformation to stabilize the variance across the time series data, from the above plot, we can see that mean and variance are constant over time and the data is ready and stationary to move forward with forecasting.

- Again performed ADF test statistic to see if the critical values are below the test statistic. The test statistic of approximately -16.34 is far below the critical values at the 1%, 5%, and 10% levels, which indicates that we can reject the null hypothesis of a unit root at these levels of significance. Moreover, the p-value is extremely small (approx. 2.95e-29), which further supports the rejection of the null hypothesis, confirming that the data does not have a unit root and is stationary.
- This transformation and subsequent ADF test are essential steps in preparing the data for modeling, particularly for models like ARIMA or LSTM, which require stationarity to produce reliable forecasts. The result of the ADF test suggests that the data, once transformed, are suitable for further time series analysis.

```
Test Statistic                  -1.634762e+01
p-value                          2.953817e-29
#Lags Used                       7.000000e+00
Number of Observations Used      2.506000e+03
Critical Value (1%)             -3.432962e+00
Critical Value (5%)             -2.862694e+00
Critical Value (10%)            -2.567384e+00
dtype: float64
```

# ACF vs PACF after Log Transformation




- From the ACF plot: Sharp cut-off after 1st lag, indicates a moving average component of order one (MA(1)).

- From PACF Plot: we see a similar pattern, a significant spike at the first lag, followed by a cutoff. This is indicative of an autoregressive component of order one (AR(1)).

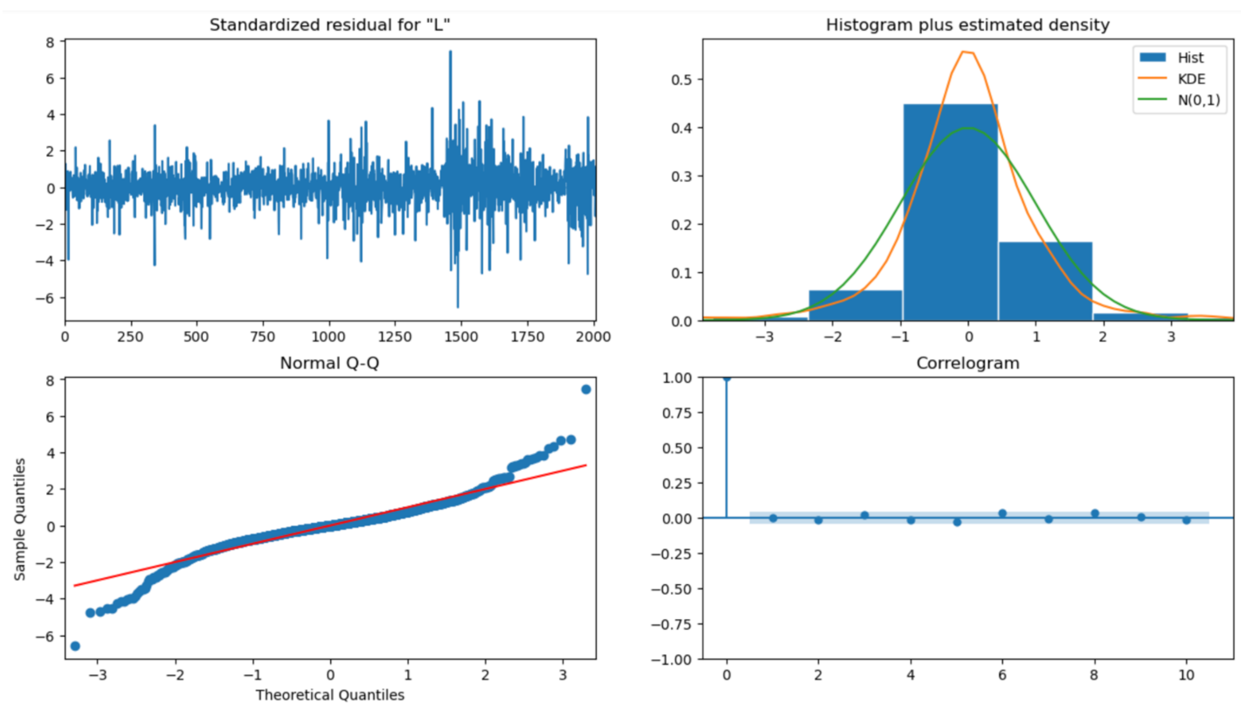
# ARIMA-SARIMA Model

## ARIMA/SARIMA Configuration:

*ARIMA (1,0,1):* We start with a basic ARIMA model based on ACF vs PACF model components. This model choice was predicated on the autocorrelation and partial autocorrelation analyses, which suggested a single autoregressive and a single moving average component would be apt for capturing the stock's time series properties and after evaluation of AIC and BIC criteria, effectively capturing the data's non-seasonal aspects.

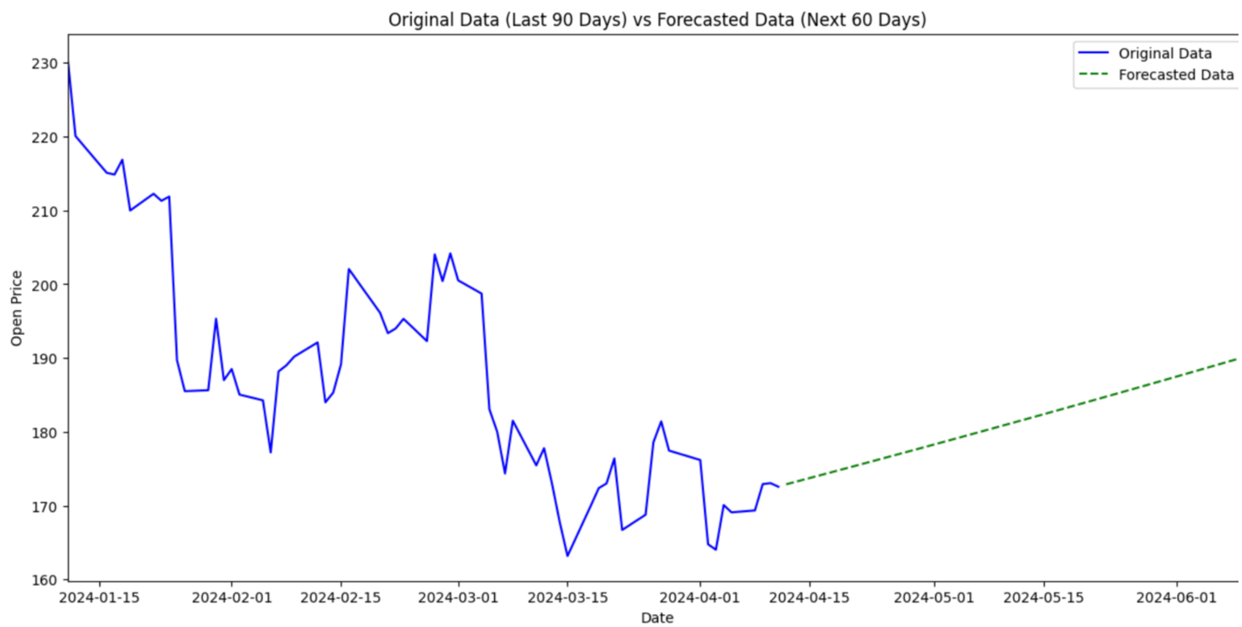
The ARIMA model was meticulously trained on a partitioned set comprising 80% of the available data, reserving the remaining 20% as a test set to assess the model's forecasting prowess. The training phase was guided by the principles of parsimony, ensuring that the model complexity was just sufficient to capture the essential dynamics of the data without overfitting.

Upon training, our ARIMA model achieved a Root Mean Squared Error (RMSE) of 0.036 on the training dataset, reflecting the model's accuracy in fitting the historical price data. The robustness of the model was further evidenced by its test RMSE of 0.038, indicating its proficiency in predicting unseen data while maintaining consistency in performance metrics between training and testing phases.




The model's forecasts, extending 60 days (about 2 months) beyond the last known data point, reveal a projected upswing in the stock's opening price. This forecast is visualized alongside the

original data for the last 90 days (about 3 months), providing a comparative perspective that aligns the model's predictions with recent trends.
Notably, the model projects an optimistic outlook despite the recent downturn, suggesting a rebound or a positive market response in the near term.




The contrast between the actual data and forecasted data is striking—the model anticipates a reversal of the recent downward trend, which may reflect latent market **f**actors or Tesla's business prospects that have yet to be fully realized by the market.

***Seasonal ARIMA (1,0,1) (1,1,1,5)**:* These models build upon the ARIMA model structure by incorporating seasonal elements, thus providing a more comprehensive representation of time series data that may exhibit periodicity along with trend and cycle components.

SARIMA Model Construction and Selection:

For the SARIMA model configuration, we first identified a seasonal pattern with a periodicity of five days, hypothesizing a weekly influence on stock prices. We then selected the SARIMA(1,0,1) (1,1,1,5) model after evaluating several seasonal periods. The parameter selection was driven by a systematic approach to capturing the essence of the data's underlying seasonal behavior, as well as optimizing information criteria such as Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC), which balance model fit with complexity.

## Model Fitting and Diagnostics:

The chosen SARIMA model was meticulously fitted to 70% of the logarithmically differenced data, revealing significant coefficients and passing various diagnostic tests:

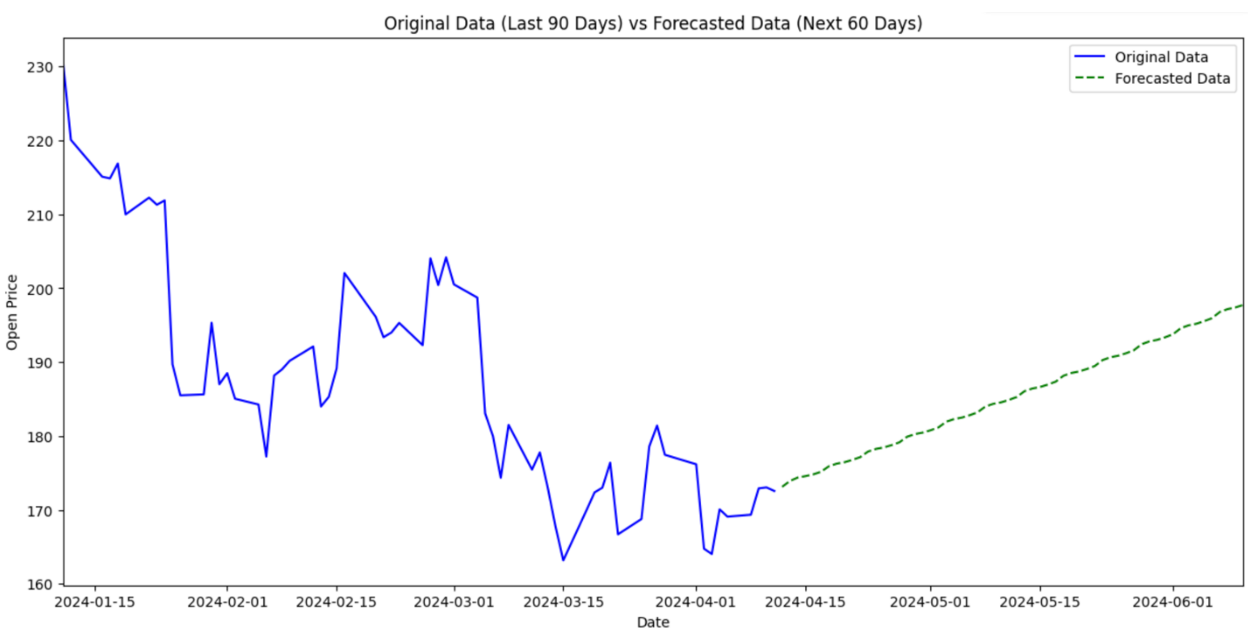
- The Ljung-Box test result was non-significant ($p > 0.05$), suggesting that the residuals are independently distributed.
- The low AIC and BIC values indicated a good fit relative to the complexity of the model.
- The model's standard error metrics underscored the reliability of the parameter estimates.

## Forecasting and Error Metrics:

The SARIMA model's forecasting performance was substantiated by its RMSE on the training data, calculated to be approximately 0.036. This demonstrates a high degree of accuracy in the model's in-sample predictions. Subsequently, a forecast extending 60 days beyond the last observed data point was generated, offering insights into the expected future trajectory of Tesla's stock opening prices.

## Visualization and Forecast Interpretation:

The plot contrasted the actual stock prices for the last 90 days (about 3 months) with the forecasted data for the next 60 days (about 2 months). It revealed a discernible discrepancy between the recent declining trend and the model's projections, which suggested an impending upturn. This optimistic forecast, depicted by the green dashed line, underscores the model's tendency to revert to the mean, potentially indicating anticipated positive developments or market corrections.



## SARIMA Model Comparison and Decision:

Further to our SARIMA model with a weekly seasonal component, a SARIMA model with a 21-day seasonal cycle was also evaluated to investigate bi-weekly influences. The comparison between the two models based on their AIC, BIC, and RMSE led to the selection of the weekly model due to its slightly superior performance metrics.

# LSTM Model

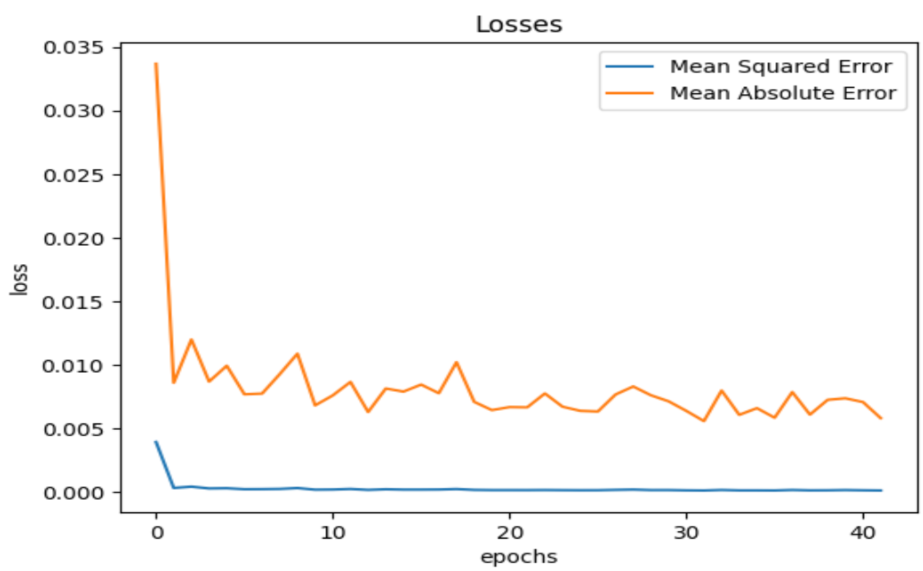
## Data Segmentation:

We divided the dataset into training (75%) and testing (25%) segments. A 60-day lookback window was used to construct sequential inputs, essential for training the LSTM to recognize and predict based on historical price patterns.

The LSTM model was constructed with a layered approach, starting with an initial LSTM layer of 50 neurons that returned sequences to the next layer. This allowed for the capture of sequential dependencies in the data. The network progressed through an additional LSTM layer with 64 neurons, transitioning into a dense network topology that culminated in a single output neuron, signifying the predicted stock price.

## Training Dynamics and Model Optimization:

We used a training technique of, setting a batch size of 32 and monitoring for 100 epochs to ensure thorough learning while employing EarlyStopping to prevent overfitting. The training utilized Mean Squared Error (MSE) as a loss function and included Mean Absolute Error (MAE) as an additional metric to monitor the model's performance. The EarlyStopping callback monitored the loss, halting the training process after ten epochs without improvement, thus conserving computational resources and preventing overtraining**.**

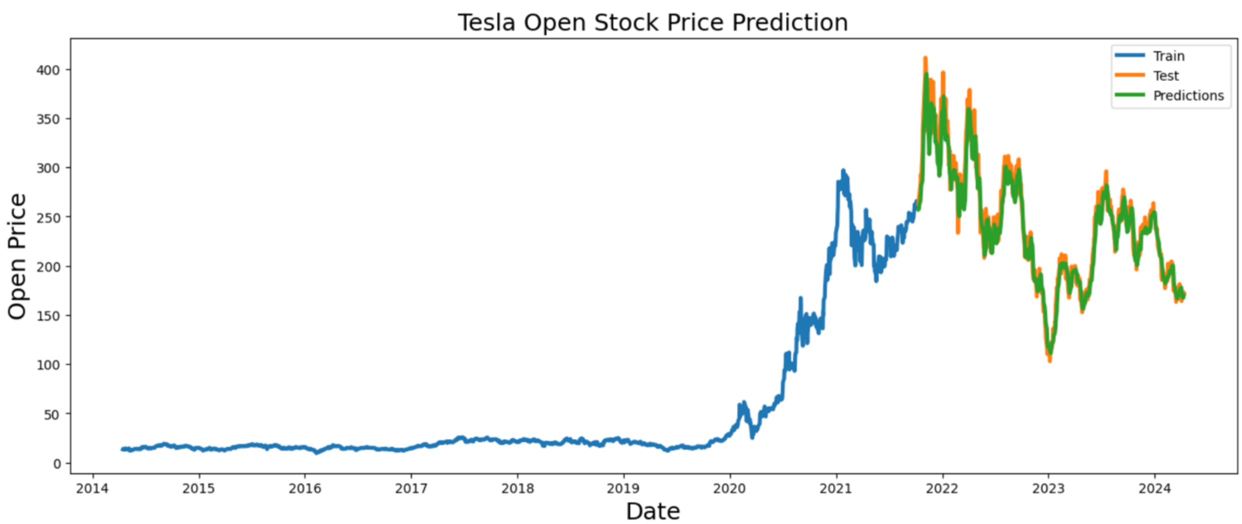


## Empirical Findings:

The LSTM model's performance on unseen data reveals that it has effectively learned from historical trends and closely aligns its predictions with actual stock prices, suggesting a strong capacity for generalization. This is evidenced by the close proximity of the model's predicted values to the true stock prices in the testing phase. The RMSE score of 4.02 further quantifies the model's accuracy, providing a concrete measure of the average deviation between the model's forecasts and the actual figures. An RMSE of this value in the context of stock market predictions is indicative of a well-performing model, as it demonstrates the model's ability to navigate and predict within the volatile realm of stock prices with a relatively small margin of error.

### Visualization of LSTM Predictions:

While there are discrepancies, the LSTM model tracked the overall trend and volatility of the stock prices with a commendable degree of accuracy.




## Results and Evaluation:

Our rigorous analysis employed ARIMA, SARIMA, and LSTM models to predict Tesla's stock prices, each revealing unique facets of the data's character. The ARIMA model (1,0,1), grounded in time series fundamentals, achieved an RMSE of 0.036 on the training set and 0.038 on the test set, highlighting its capability for linear trend and noise modeling.

The SARIMA model (1,0,1) (1,1,1,5), chosen for its ability to model seasonal patterns, showed an improved RMSE of 0.0362, suggesting a nuanced capture of both non-seasonal and seasonal dynamics over weekly periods. Its forecast indicated a potential rebound in stock prices, despite recent declines, pointing to a resilient growth trend.

In parallel, the LSTM model was able to effectively learn and represent the underlying trends, seasonality, and non-linear relationships in Tesla's opening price data.

# Conclusion

The application of ARIMA, SARIMA, and LSTM models provides a comprehensive analytical perspective on Tesla's stock price movements. Each model contributes uniquely to our understanding of the data, with ARIMA and SARIMA offering robustness in trend and seasonal analysis, and LSTM delivering insights into non-linear patterns and long-term dependencies.

The comparison of the models' forecasts with recent historical data underscores a central finding: while short-term fluctuations are prevalent, the overall trajectory suggests an underlying growth trend that may continue.

# Future Work

*Enhanced Data Handling:* Future iterations could incorporate high-frequency data and additional predictive features, such as news sentiment and macroeconomic indicators, potentially employing SARIMAX to integrate exogenous variables.

*Model Refinement:* Continued exploration and optimization of neural network architectures and parameters are recommended to enhance forecast accuracy and model robustness.