# TCS Stock Market Forecasting

1. Introduction

The dataset used for this analysis consists of minute-wise stock data for Tata Consultancy Services (TCS). This data is crucial for understanding the minute-to-minute fluctuations in stock prices, which can be pivotal for traders and investors. The primary objectives of this analysis are to identify trends within the data, visualize key metrics, and forecast future stock prices.

2. Data Preprocessing

- Loading the Data:
  The dataset comprises several columns, including timestamp, open, high, low, close, and volume. The timestamp column is essential for tracking the chronological order of price movements.
- Handling Missing Values:
  In this analysis, missing values and duplicates are removed to ensure the integrity of the dataset, providing a clean slate for subsequent analyses.
- Feature Extraction"
  From the timestamp, various date and time components are extracted, including year, month, day, hour, minute, and second. This helps in analyzing trends over specific periods.

3. Exploratory Data Analysis (EDA)

- Summary Statistics
  A detailed description of summary statistics for numerical columns such as open, close, high, low, and volume was generated to understand the central tendency and dispersion of the data.
- Correlation Matrix and Heatmap
  A correlation analysis was conducted to explore relationships between numerical features. The heatmap visualizes these correlations, helping to identify which variables are positively or negatively related.
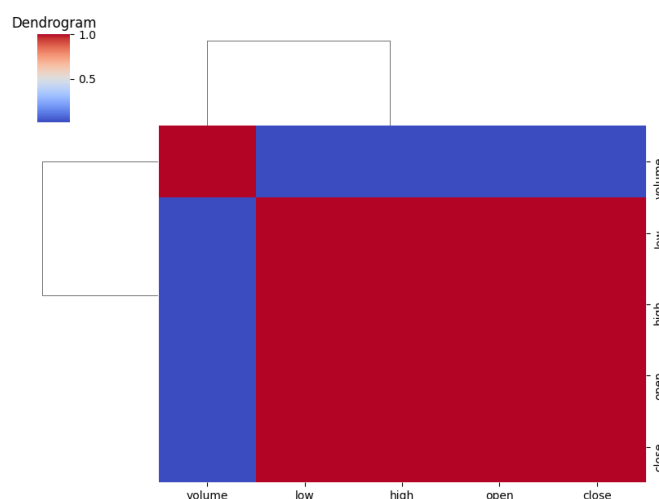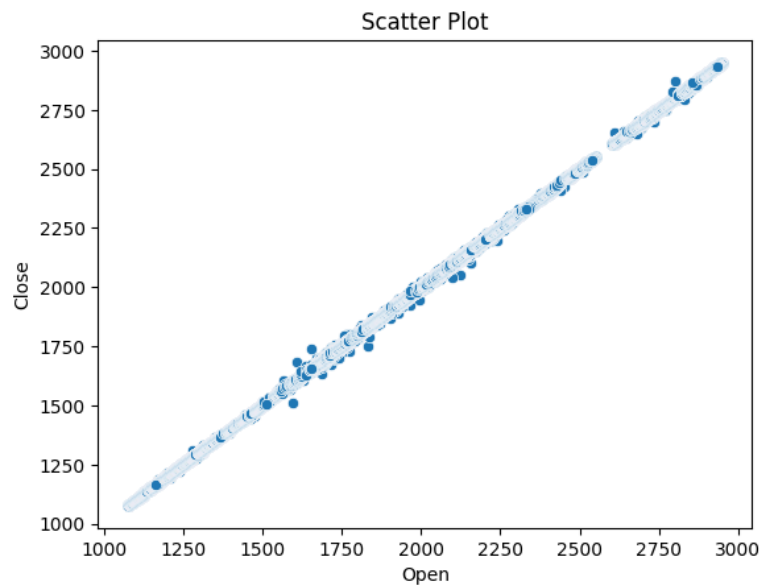

- Histograms
  Histograms were plotted for the numerical columns to visualize the distribution of values for open, close, high, low, and volume.

Histograms
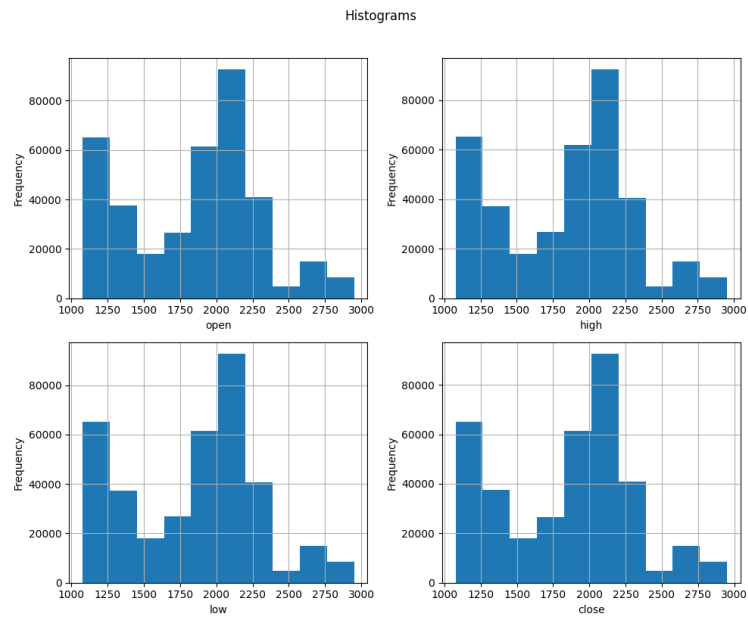
- Scatter Plot

    Scatter plots were created to demonstrate the relationship between open and close prices, allowing for a visual inspection of how closely related these two metrics are.

4. Data Visualization

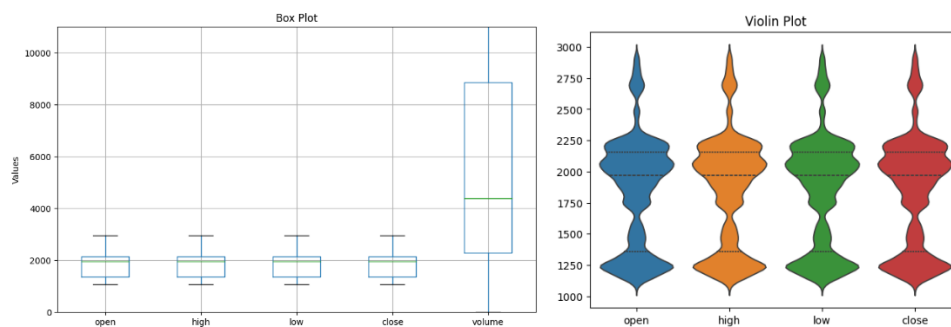- Line Plot:

    Line plots were generated for the open, high, low, close, and volume columns over time, showcasing the price movements and trading volumes.

Line Plot
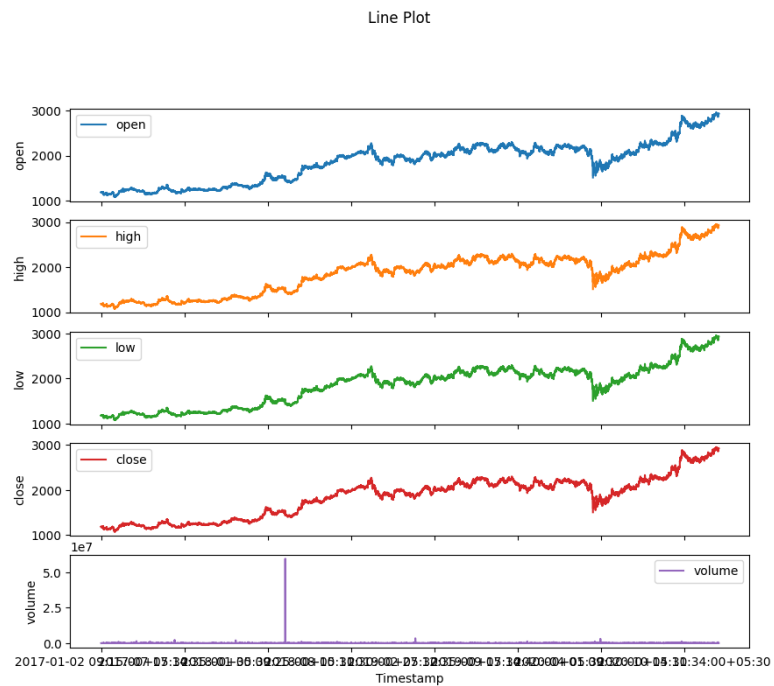
- Box Plot and Violin Plot:

  Box plots and violin plots were used to display the data distribution and identify outliers within the dataset.
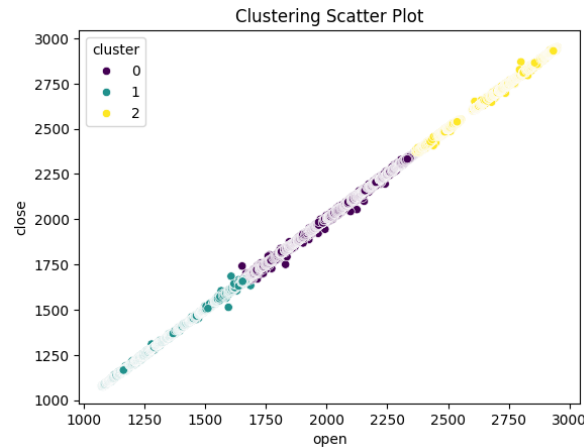

- Candlestick Chart:

  A candlestick chart was produced to visualize the stock's price movements over time, providing insights into market trends and volatility.

5. Clustering

K-Means Clustering:

        K-Means clustering was applied to identify patterns in stock prices, utilizing open-close and high-low values to classify different price movement behaviours.

6. Time Series Forecasting

- Stationarity Check (ADF Test):
  An Augmented Dickey-Fuller (ADF) test was performed to check the stationarity of the close price column, a crucial step for effective time series forecasting.

- ARIMA Modeling
  The data was differenced to achieve stationarity, followed by the selection of appropriate ARIMA model parameters (p, d, q). The model was then fitted to the data to forecast future stock prices.
- Forecasting
  Results of the forecasting were displayed, including a plot comparing historical values with forecasted values, allowing for an evaluation of the model's predictive power.

7. Conclusion

In summary, the analysis provided key insights into the TCS stock market, highlighting significant trends and patterns through EDA, clustering, and forecasting. Future work could focus on refining prediction models or exploring additional stock data for broader market insights.