Time Series Forecasting

Data Collection

Pandas Datareader : It's a python library that allows you to retrieve data from various online sources directly into a pandas DataFrame. It's often used for financial and economic data, making it useful for time series forecasting projects.

Used "`yfinance`" Package (Yahoo Finance API Wrapper) : The "`yfinance`" library is an alternative that is commonly used for fetching Yahoo Finance data and works well.

Data Description:

Open: The stock price when the market opened on that date.
High: The highest price during the trading session.
Low: The lowest price during the trading session.
Close: The price at market close.
Adj Close: The adjusted closing price, accounting for any corporate actions like stock splits or dividends.
Volume: The number of shares traded during that day.

Summary of the Tesla stock data, with key statistical measures for each of the columns: Open, High, Low, Close, Adj Close, and Volume

Column-wise Breakdown:

1. **Open, High, Low, Close, Adj Close:**

- **Count:** The total number of days (3576 days of data).
- **Mean:** The average stock price over the observed period. All values are around $76, indicating that the stock price has been rising substantially over the years (early values were very low compared to recent ones).
- **Standard Deviation (std):** The large standard deviation (~102) suggests high volatility in the Tesla stock price, with substantial fluctuations over time.
- **Min:** The lowest recorded value for each of the price columns is around $1, showing how Tesla's stock price has dramatically increased over the years.
- **25%, 50%, 75% (Percentiles):**
    - 25th Percentile: Around $11-$12, meaning 25% of the stock prices were below this value.
    - 50th Percentile (Median): Around $17.54, indicating that half of the stock prices were below this value.
    - 75th Percentile: Around $168, which shows that 75% of the data is below this threshold, and only 25% of the data involves prices above this value.
- **Max:** The maximum value for Tesla stock reached over $410 for all price columns, reflecting the peak Tesla hit during its market highs.

2. **Volume:**

- Mean Volume: Around 96.8 million shares traded on average per day.

- **Standard Deviation:** With a standard deviation of 78.3 million, the trading volume fluctuates widely from day to day.
- **Min Volume:** The minimum number of shares traded on any given day was 1.77 million.
- **25%, 50%, 75% Percentiles:**
    - 25th Percentile: About 48 million shares traded.
    - 50th Percentile: About 82 million shares traded, meaning 50% of the trading days saw more than 82 million shares traded.
    - 75th Percentile: About 122.5 million shares traded.
- **Max Volume:** The highest number of shares traded in a single day was 914 million, indicating a period of extremely high activity in the market, likely due to a major event related to Tesla or the market as a whole.

Insight:

- Tesla's stock has undergone substantial growth, as reflected by the significant difference between the minimum and maximum prices.
- The large standard deviation, particularly in the price data, highlights Tesla's high volatility, which is typical for growth stocks, especially in the tech and electric vehicle sectors.
- The volume data also reflects how actively Tesla shares are traded, with a massive range between quiet trading days (1.77 million) and extremely active.

Exploratory Data Analysis :

Plotted Tesla's high prices over time:

- **Key Observations:**

1. **Early Stability (2010 - 2019)**:
    - Tesla's stock price remained relatively stable and low between 2010 and 2019, hovering under $100.
    - This suggests that Tesla was still a growing company, with limited market influence during this period.
2. **Rapid Growth (2020 Onwards)**:
    - Around late 2019 and early 2020, there was a significant upward trend in the stock price, where Tesla's high prices surged dramatically past $400 at its peak.
    - This is likely tied to Tesla's strong financial performance, increased demand for electric vehicles (EVs), and investor enthusiasm.
3. **Volatility (2021 - 2024)**:
    - After reaching a peak, Tesla's stock experienced considerable volatility, with high swings both upward and downward.
    - There were sharp corrections, followed by recoveries, indicating uncertainty in the market.

Time Sampling: Time resampling in Pandas allows you to aggregate data over a new time frequency. For instance, you can convert daily stock prices to weekly, monthly, or yearly data by resampling and applying an aggregation function like mean, sum, or max.

Trying to identify the lowest values observed in each year: It's informative for trend analysis and understanding the lowest price points or trading volumes during each period.

Yearly Comparison Insight: Comparing these annual lows can help in understanding how the stock's resilience or volatility has changed over the years. For instance, the considerable drop in minimum prices from 2021 to 2022 might warrant a deeper investigation into external factors like market conditions, Tesla's business decisions, or global economic events at that time.

The maximum values of Tesla stock prices and volume for each year, from 2010 to 2024:

Insight:

- Tesla experienced rapid growth, particularly between 2020 and 2021, when its stock price more than doubled in value.
- After reaching its peak in 2021, the stock price saw a gradual correction but remained significantly higher than pre-2020 levels.
- The stock's trading volume also reflects major market activities, with substantial peaks in 2020, reflecting broader market trends and interest.

Cumulative Moving Average (CMA) of Tesla's opening stock prices from 2010 to 2024:

Key Observations:

- From 2010 to 2019, the cumulative average remains relatively low and stable, increasing very gradually. This indicates that Tesla's stock prices were relatively steady with small fluctuations during this period.
- After 2020, the curve exhibits a sharp upward trajectory, indicating a significant increase in the opening stock prices. This aligns with Tesla's growth during that time, where its stock saw massive gains due to increased production, demand, and market attention.

Exponential Weighted Moving Average (EWMA) of Tesla's opening stock prices from 2020 to 2024

Key Observations:

- The orange line (EMA) smooths out these fluctuations by giving more weight to recent prices while considering the stock's overall trend.
- It closely follows the actual price but with less volatility, reflecting the underlying trend more clearly without reacting as drastically to short-term price changes.

Data Modelling:

SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors)

- The model captures the overall trend well, but the individual AR and MA terms are not significantly different from zero. This suggests that the differencing term (d=1) might be doing most of the heavy lifting in terms of capturing the series' behavior.
- The residuals show no significant autocorrelation (good), but they are not normally distributed and show signs of heteroskedasticity (not ideal). This suggests that while the model captures the main trend, it might not fully capture all the dynamics in the data.
- The high kurtosis indicates that the data has more extreme values than would be expected in a normal distribution, which is common in financial time series.
- The model's good fit statistics suggest it could be useful for forecasting, but the non-normal residuals and heteroskedasticity mean that prediction intervals and other inferences should be interpreted with caution.
- Given the non-significant AR and MA terms, a simpler model (perhaps just differencing or a random walk with drift) might perform similarly well and be more parsimonious.