In this project, we aim to analyze the stock price data of 15 major companies over the past 20 years. Our goal is to explore the data, gain initial insights, and perform feature engineering to prepare the dataset for further analysis, including sentiment analysis for predicting stock market movements.

2. Data Collection

Historical stock price data was collected for the following companies: Apple (AAPL), Microsoft (MSFT), Amazon (AMZN), Tesla (TSLA), Meta (META), NVIDIA (NVDA), Berkshire Hathaway (BRK-B), JPMorgan Chase (JPM), Visa (V), Walmart (WMT), Disney (DIS), Boeing (BA), Goldman Sachs (GS), IBM, and Pfizer (PFE). The data was sourced using the `yfinance` library, spanning from January 1, 2005, to December 31, 2024.

3. Data Cleaning

To ensure the dataset is clean and ready for analysis, the following steps were taken:

- \*\*Handling Missing Values\*\*: Forward-fill method was used to fill missing values.

- \*\*Removing Duplicates\*\*: Duplicate records were removed from the dataset.

- \*\*Data Types\*\*: Ensured that all columns have the correct data types, and the 'Date' column was converted to datetime format.

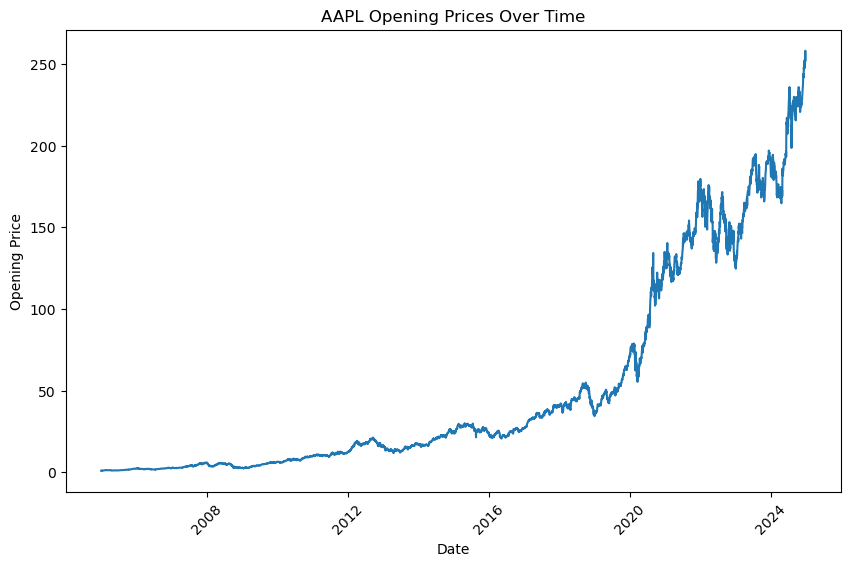
4. Exploratory Data Analysis (EDA)

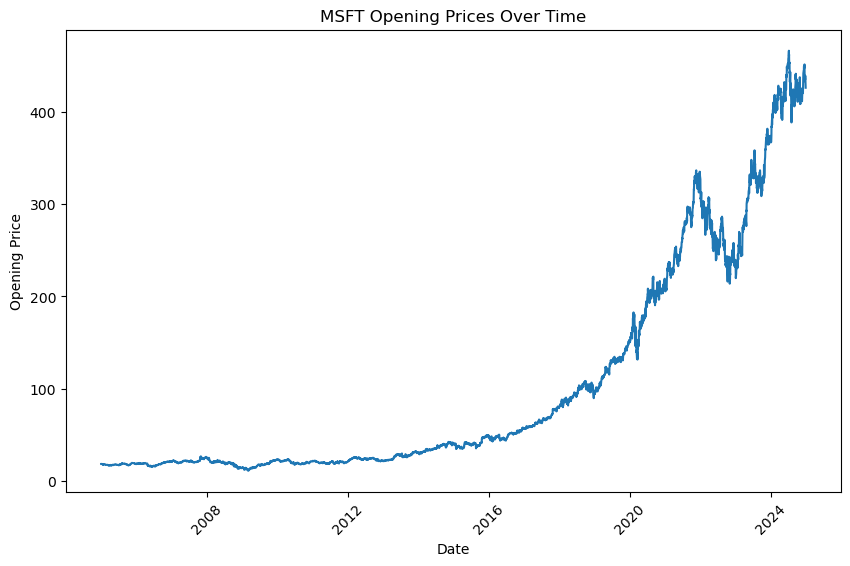
##### Visualizations and What They Mean

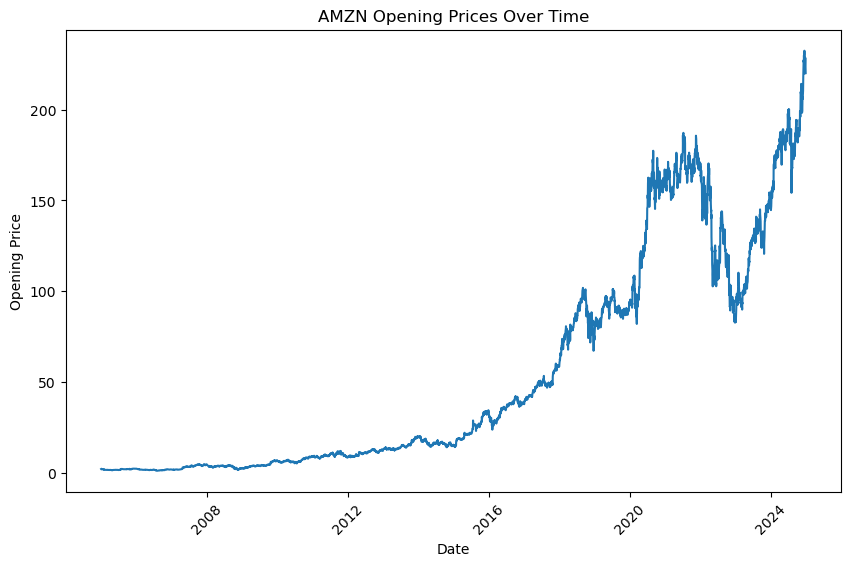
1. \*\*Line Plots for Opening Prices\*\*:

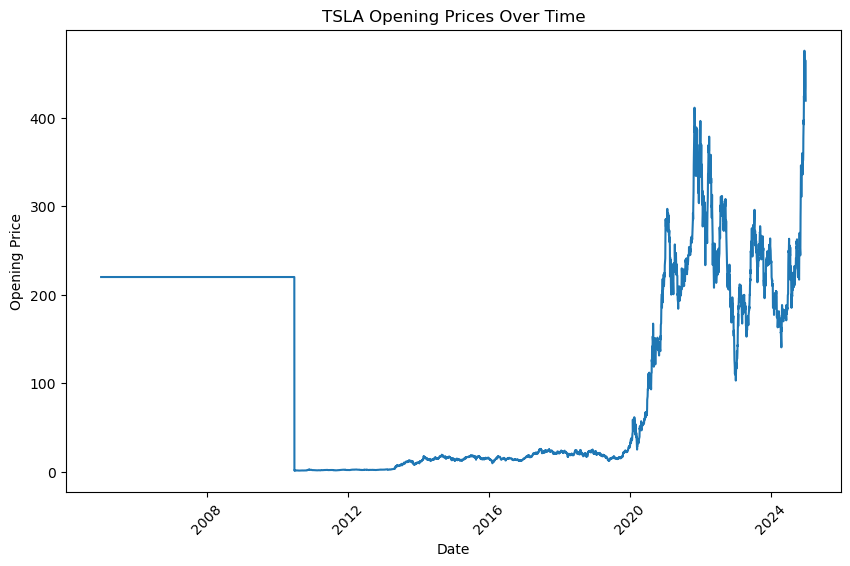
- \*\*Purpose\*\*: To visualize the opening prices of each company over time.

- \*\*Interpretation\*\*: These graphs show how the opening prices have changed over the years for each company. Trends can indicate how the company's stock has performed historically, highlighting periods of growth, decline, or stability.









A graph showing the price of a stock market

Description automatically generated

A graph showing the price of a company

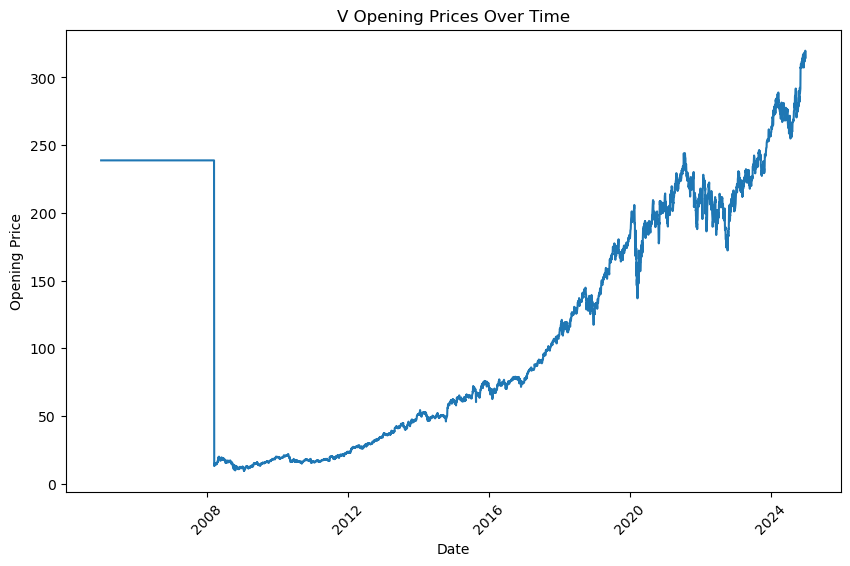
Description automatically generated

A graph showing the price of a stock market

Description automatically generated

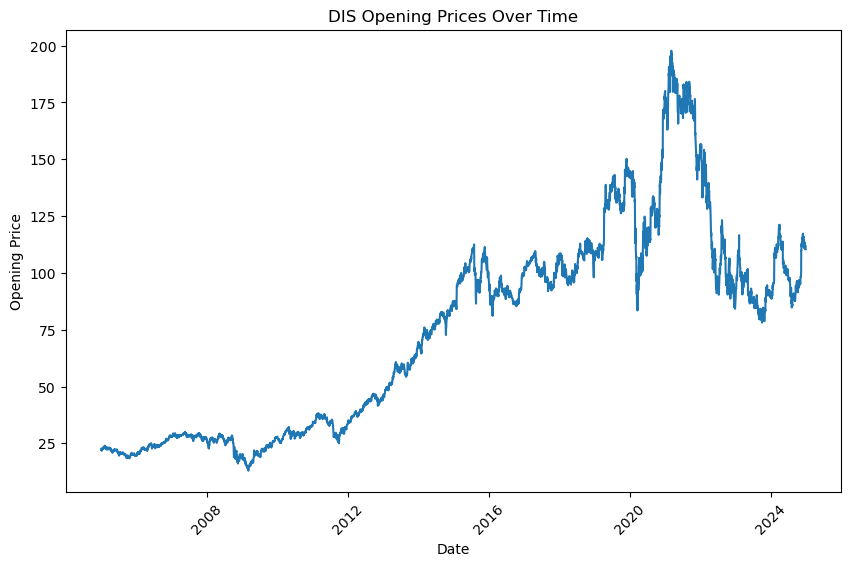
A graph of a stock market

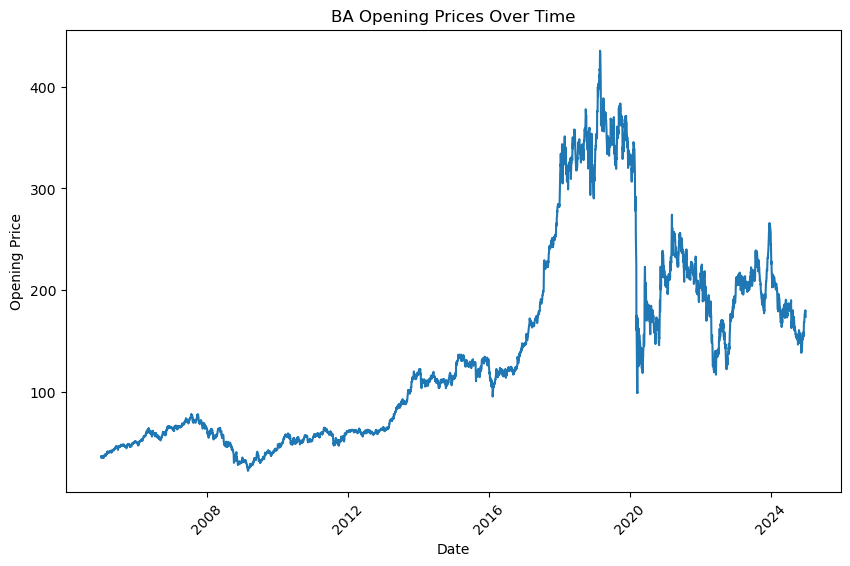
Description automatically generated with medium confidence



A graph showing the growth of a stock market

Description automatically generated





A graph with blue lines

Description automatically generated

A graph showing the growth of the company's stock market

Description automatically generated

A graph showing the price of a stock market

Description automatically generated

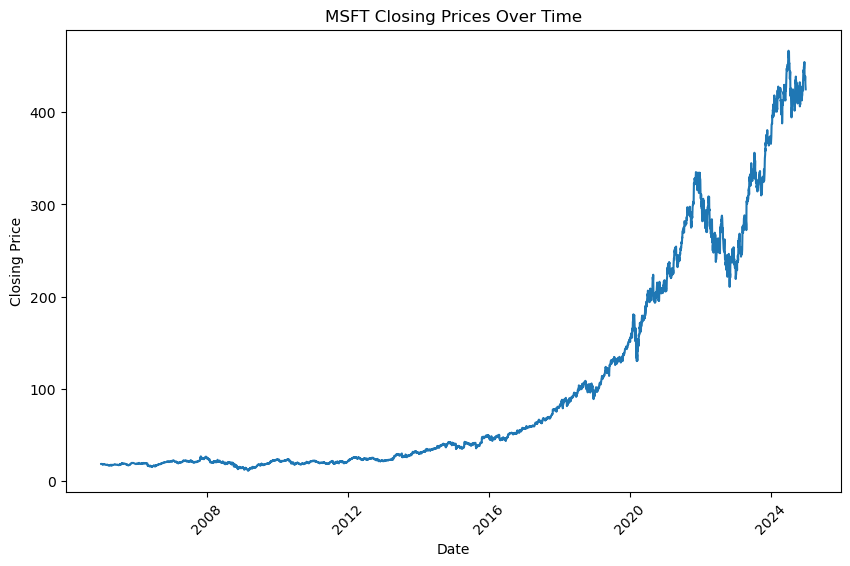
2. \*\*Line Plots for Closing Prices\*\*:

- \*\*Purpose\*\*: To visualize the closing prices of each company over time.

- \*\*Interpretation\*\*: These graphs illustrate how the closing prices have evolved over time. Comparing these with opening prices can reveal the day's price movement and overall trends.

A graph showing the price of a stock market

Description automatically generated

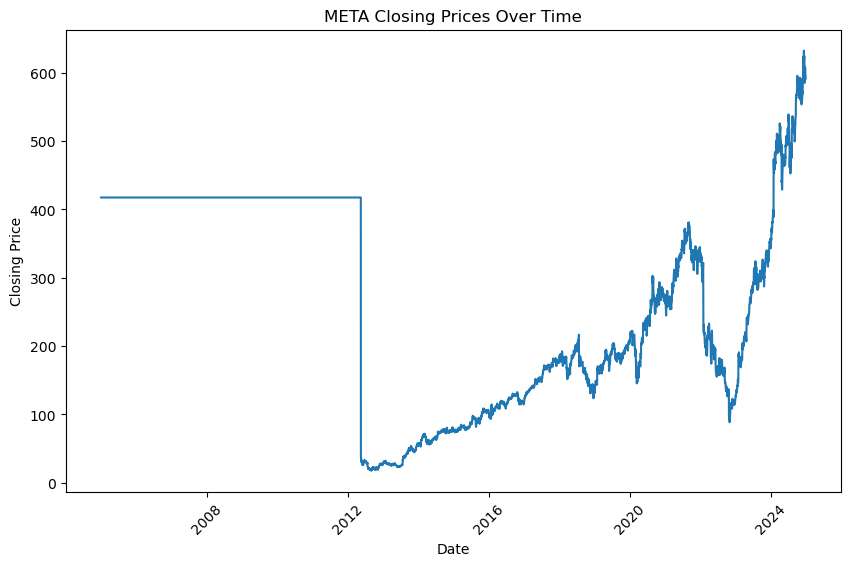


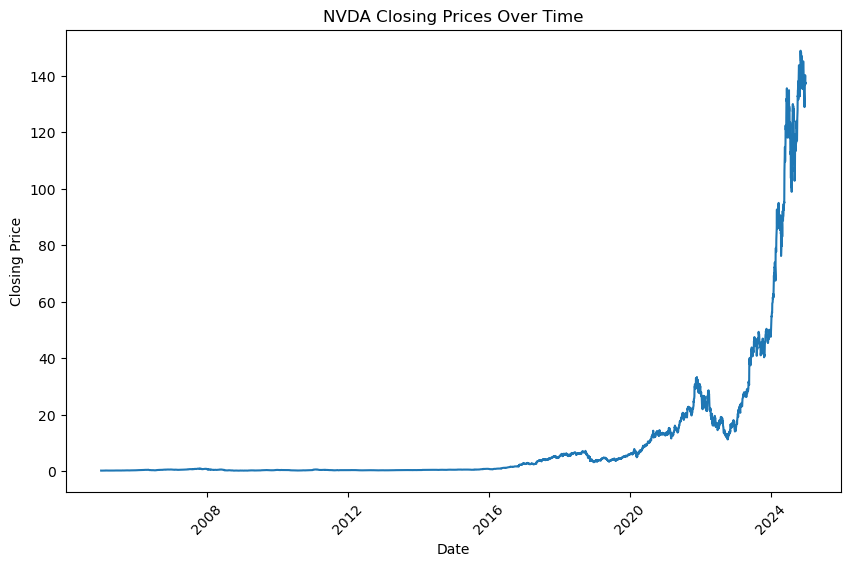
A graph showing the price of a stock market

Description automatically generated

A graph showing the price of a stock market

Description automatically generated



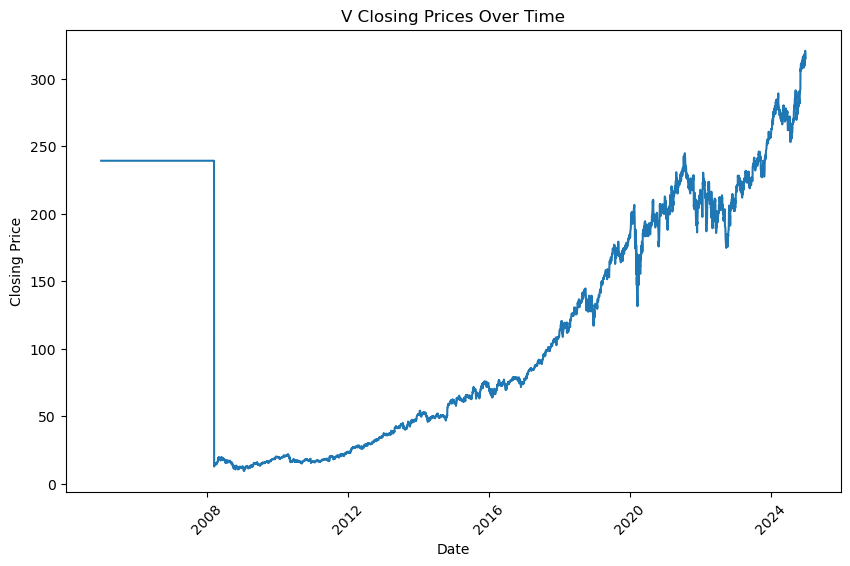


A graph showing a line graph

Description automatically generated with medium confidence

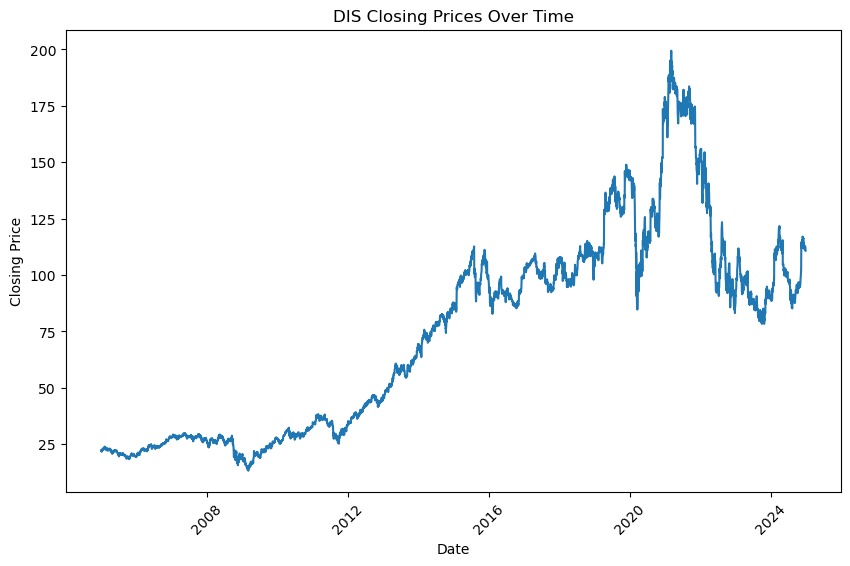
A graph showing a line graph

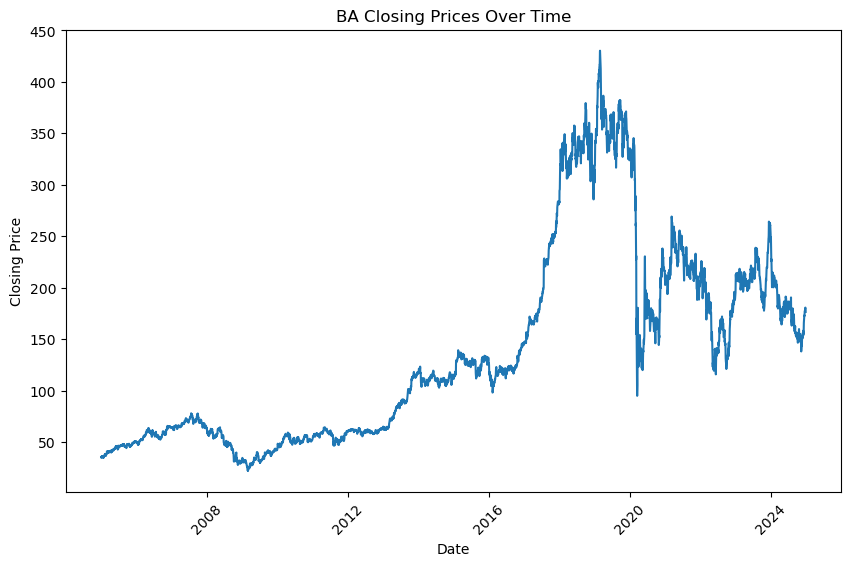
Description automatically generated with medium confidence



A graph of a price

Description automatically generated with medium confidence





A graph showing the price of a stock market

Description automatically generated

A graph showing the price of a stock market

Description automatically generated

A graph showing a line graph

Description automatically generated with medium confidence

3. \*\*Line Plots for High Prices\*\*:

- \*\*Purpose\*\*: To visualize the highest prices reached by each company on a given day.

- \*\*Interpretation\*\*: These graphs provide insights into the peak values of the stocks, showing the highest points the stocks reached during the trading days over the years.

A graph of a line

Description automatically generated with medium confidence

A graph of a line

Description automatically generated with medium confidence

A graph showing the growth of a stock market

Description automatically generated

A graph showing the price of a stock market

Description automatically generated

A graph showing a line graph

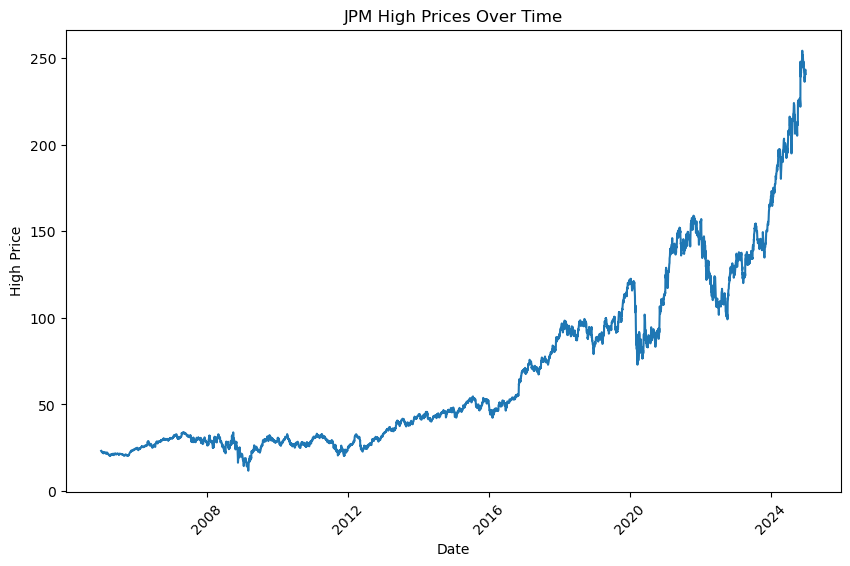
Description automatically generated with medium confidence

A graph showing a number of prices

Description automatically generated with medium confidence

A graph showing the growth of a stock market

Description automatically generated



A graph with a line

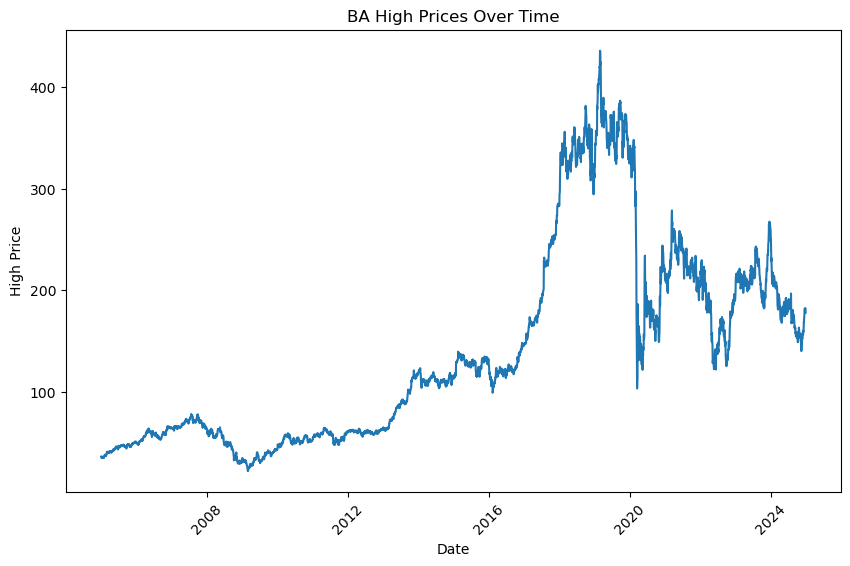
Description automatically generated

A graph of a price

Description automatically generated with medium confidence

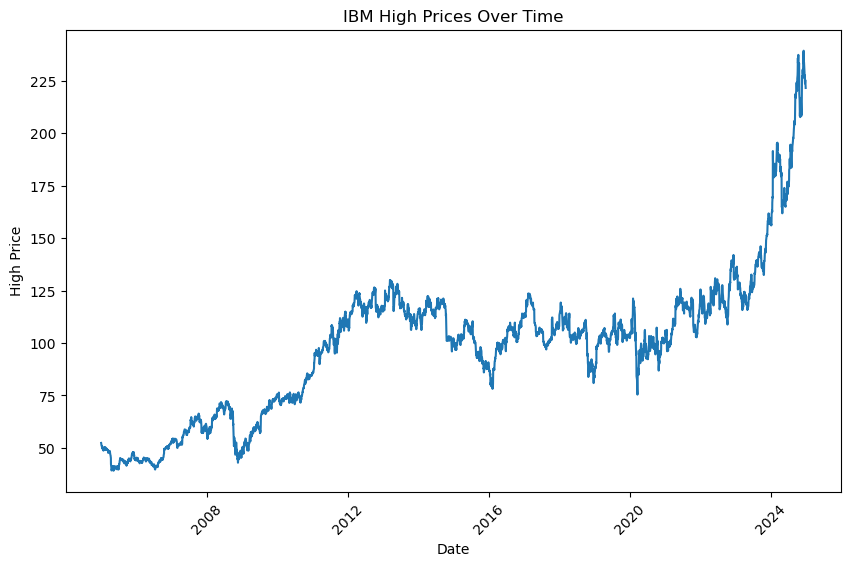
A graph showing the growth of a stock market

Description automatically generated



A graph with blue lines

Description automatically generated



A graph showing a line

Description automatically generated with medium confidence

4. \*\*Line Plots for Low Prices\*\*:

- \*\*Purpose\*\*: To visualize the lowest prices reached by each company on a given day.

- \*\*Interpretation\*\*: These graphs help identify the lowest points the stocks have touched, which is useful in understanding the downside risk over time.

A graph of a line

Description automatically generated with medium confidence

A graph showing the price of a stock market

Description automatically generated

A graph showing the growth of the stock market

Description automatically generated

A graph showing the price of a stock market

Description automatically generated

A graph showing a line graph

Description automatically generated with medium confidence

A graph of a price

Description automatically generated with medium confidence

A graph showing the price of a stock market

Description automatically generated

A graph showing the price of a stock market

Description automatically generated

A graph showing the price of a stock market

Description automatically generated

A graph with blue lines

Description automatically generated

A graph showing the price of a stock market

Description automatically generated

A graph showing the price of a stock market

Description automatically generated

A graph with numbers and lines

Description automatically generated

A graph showing the price of a stock market

Description automatically generated

A graph showing the price of a stock market

Description automatically generated

5. \*\*Line Plots for Trading Volume\*\*:

- \*\*Purpose\*\*: To visualize the trading volumes for each company over time.

- \*\*Interpretation\*\*: These graphs show how many shares were traded on a given day, indicating market interest and liquidity. High volume periods can be linked to significant market events or company-specific news.

A graph showing a number of blue lines

Description automatically generated

A graph of blue lines

Description automatically generated

A graph of trading volume over time

Description automatically generated

A graph of a trading graph

Description automatically generated

A graph showing a number of trading

Description automatically generated with medium confidence

A graph showing a trading volume

Description automatically generated with medium confidence

A graph with numbers and lines

Description automatically generated

A graph showing a number of blue lines

Description automatically generated

A graph with numbers and lines

Description automatically generated

A graph showing a number of blue lines

Description automatically generated

A graph of a trading volume

Description automatically generated with medium confidence

A graph showing a number of blue lines

Description automatically generated

A graph showing a number of trading

Description automatically generated with medium confidence

A graph of blue lines

Description automatically generated

A graph showing a number of blue lines

Description automatically generated with medium confidence

6. \*\*Box Plots for Distribution of Opening Prices\*\*:

- \*\*Purpose\*\*: To show the distribution of opening prices for all companies.

- \*\*Interpretation\*\*: These plots display the range of opening prices, highlighting the median, quartiles, and any outliers. They help in comparing the price ranges across different companies.

A graph of a number of blue rectangular bars

Description automatically generated with medium confidence

7. \*\*Box Plots for Distribution of Closing Prices\*\*:

- \*\*Purpose\*\*: To show the distribution of closing prices for all companies.

- \*\*Interpretation\*\*: Similar to the opening prices box plots, these provide insights into the spread of closing prices and highlight differences between companies.

A diagram of a distribution of closing prices

Description automatically generated

8. \*\*Box Plots for Distribution of High Prices\*\*:

- \*\*Purpose\*\*: To show the distribution of high prices for all companies.

- \*\*Interpretation\*\*: These plots give an overview of the peak values achieved by each company's stock, indicating the volatility and potential upward movements.

A chart of a number of blue rectangular columns

Description automatically generated with medium confidence

9. \*\*Box Plots for Distribution of Low Prices\*\*:

- \*\*Purpose\*\*: To show the distribution of low prices for all companies.

- \*\*Interpretation\*\*: These plots show the range of lowest points, helping to understand the downside risks and overall stability of the stocks.

A graph of a number of blue rectangular objects

Description automatically generated with medium confidence

10. \*\*Box Plots for Distribution of Trading Volume\*\*:

- \*\*Purpose\*\*: To show the distribution of trading volumes for all companies.

- \*\*Interpretation\*\*: These plots highlight the variation in trading activity, showing which stocks are more frequently traded and identifying any unusual spikes in volume.

A graph of a number of companies

Description automatically generated with medium confidence

11. \*\*Correlation Heatmap\*\*:

- \*\*Purpose\*\*: To visualize the correlation between `Open`, `Close`, `High`, `Low`, and `Volume`.

- \*\*Interpretation\*\*: The heatmap shows the relationships between different numerical columns, indicating how closely related they are. High correlation values suggest a strong relationship, which can be useful for feature selection in predictive modeling.

A red and blue chart

Description automatically generated

[ ]:

##### Initial Insights

From the EDA, we gained several insights:

- \*\*Trends\*\*: Observed upward and downward trends in stock prices over time for each company.

- \*\*Volatility\*\*: Noted periods of high volatility in stock prices and trading volumes.

- \*\*Outliers\*\*: Identified any anomalies or outliers in the data that might need further investigation.

- \*\*Correlations\*\*: Examined the relationships between different numerical columns to inform feature engineering.

#### 5. Feature Engineering

Based on the insights from the EDA, new features were engineered to improve model performance:

- \*\*Lagged Features\*\*: Created lagged features for `Close` prices to capture temporal dependeds

#### 6. Conclusion

In this report, an extensive EDA was conducted on the stock price data of 15 major companies. We cleaned the dataset, visualized various aspects of the data, and derived initial insights. This analysis lays a solid foundation for further steps in our project, including sentiment analysis and predictive modeling.