NAME: Nischitha S G

EMAIL: nischithasg@gmail.com

COLLEGE: B.M.S. College of Engineering

Title: Stock Market Performance Analysis using Python

Objective:

The primary objective of this project is to analyse and compare the short-term stock performance of four major technology companies — Apple (AAPL), Microsoft (MSFT), Netflix (NFLX), and Google (GOOG) — over a three-month period using data analytics and machine learning techniques.

This analysis seeks to:

- Identify trends and patterns in historical stock price movements, helping to understand each company's market behaviour.
- Calculate key technical indicators, including moving averages and volatility, to evaluate price stability and potential trading signals.
- Perform correlation analysis between different company stocks to uncover interdependencies and similar market responses.
- Apply data visualization tools for intuitive understanding and explore ML models for predicting future price movements based on historical data.

Ultimately, this project aims to transform raw market data into actionable insights, supporting smarter investment decisions and deeper market comprehension.

Tools & Technologies Used

Category	Tools / Technologies
Programming Language	Python
Data Handling & Analysis	Pandas, NumPy, SQL
Data Visualization	Matplotlib, Seaborn, Pandas Profiling
Machine Learning & Modelling	Scikit-learn (Linear Regression, Model Evaluation), Keras (for advanced modelling)
Preprocessing & Feature Engineering	Scikit-learn (StandardScaler, train_test_split), Pandas Built-in Methods
IDE / Platform	Jupyter Notebook
Reporting & Documentation	Excel, PDF, HTML (via Pandas Profiling), MS Word (optional for report writing)

Dataset Information

The dataset used in this project contains historical stock data for four leading technology companies: Apple (AAPL), Microsoft (MSFT), Netflix (NFLX), and Google (GOOG). It covers a 3-month period from February to May 2023, with a total of 248 records.

Key Details:

- **Source**: Provided CSV (originally from Google Drive)
- **Companies**: AAPL, MSFT, NFLX, GOOG
- **Time Period**: Feb 2023 – May 2023 (approx. 3 months)
- **Total Records**: 248 (daily entries for each company)
- **Missing Data**: None (complete dataset)

Features in the Dataset:

- **Date** – Trading date
- **Ticker** – Company symbol (e.g., AAPL)
- **Open** – Opening price
- **High** – Highest price of the day
- **Low** – Lowest price of the day
- **Close** – Closing price
- **Adj Close** – Adjusted close price
- **Volume** – Number of shares traded

This dataset is suitable for analyzing short-term market behaviour, identifying trends, and building predictive models.

Exploratory Data Analysis (EDA)

The Exploratory Data Analysis phase aimed to understand the underlying structure and distribution of the stock data for the four selected companies. Various visualizations and statistical summaries were used to uncover patterns, trends, and relationships.

1. Distribution of Closing Prices

- A histogram and boxplot were used to visualize the distribution of closing prices across all companies.
- These plots highlighted the range, central tendency, and outliers in the closing price data.
- Most closing prices were concentrated between $100 and $300, with a few higher values, especially from Google and Microsoft.

2. Volume Comparison by Ticker

- A bar chart was plotted to compare the total trading volume of each stock.

- Apple (AAPL) had the highest trading volume over the period, followed by Microsoft (MSFT), indicating higher market activity and investor interest.

3. Volume vs. Closing Price

- A scatter plot was used to examine the relationship between trading volume and closing price.
- No strong linear relationship was observed, but the plot helped identify high-volume trading days and potential anomalies.

4. Correlation Matrix

- A heatmap of the correlation matrix was generated to study the relationships between numerical features like Open, High, Low, Close, Adj Close, and Volume.
- Strong positive correlations (>0.99) were observed between price-related features (Open, High, Low, Close), suggesting they move together.
- Volume showed a weaker, slightly negative correlation with price features.

5. Ticker-wise Analysis

- Each company (AAPL, MSFT, NFLX, GOOG) was individually filtered and analysed to examine:
    - Trends over time
    - Daily price fluctuations
    - Rolling statistics like moving averages and volatility (discussed in the next section)

This EDA phase provided a solid foundation for feature engineering and model development in the subsequent steps.

Feature Engineering

To enhance the dataset and extract more meaningful insights from the stock price movements, several new features were engineered for each company. These derived metrics help in understanding short-term trends, risk, and return potential of the stocks.

1. 5-Day Moving Average (MA)

- A 5-day simple moving average was calculated on the Close price to smooth out short-term fluctuations.
- This indicator highlights price trends and is commonly used in technical analysis to identify support and resistance levels.

2. Daily Return (%)

- Daily return was computed as the percentage change in the closing price from the previous trading day.
- This helps measure day-to-day performance and volatility of a stock, which is useful for understanding price momentum.

3. Rolling Volatility

- A rolling standard deviation of closing prices over a 5-day window was calculated as a measure of short-term volatility.
- This indicates how much the price fluctuates in a short period and helps assess risk associated with the stock.

These engineered features were later used in both the visual analysis and model-building phases to gain deeper insights into stock behaviour and performance patterns.

Model Building

To demonstrate the potential for predictive analysis, a Linear Regression model was implemented to estimate the closing price of a stock using key financial indicators.

Model Overview:

- The model used was Linear Regression from the scikit-learn library.
- The dataset was divided into training (80%) and testing (20%) sets using the train_test_split() function.

Objective:

The goal was to predict the **Close** price of a stock based on its opening price, highest and lowest prices of the day, and the trading volume.

Evaluation Metric:

- The model was evaluated using the R^2 Score (coefficient of determination), which indicates how well the predicted values match the actual values.
- A high R^2 score reflects that the model has captured most of the variance in the target variable.

Outcome:

- The model performed well, showing a strong correlation between the predicted and actual closing prices.
- This validates the earlier correlation analysis which showed that Open, High, and Low are closely linked to Close.

```
R² Score: 0.9854542039468671
```

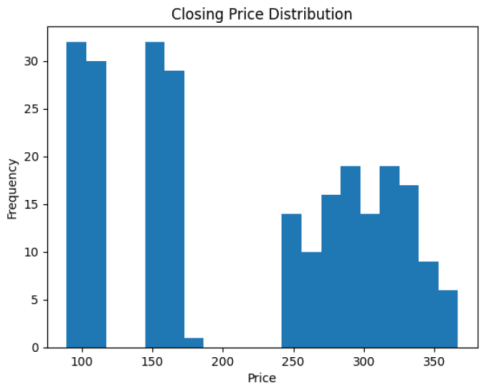
Insights and Observations

The analysis of stock data for Apple (AAPL), Microsoft (MSFT), Google (GOOG), and Netflix (NFLX) over a 3-month period revealed several key insights through visual exploration and statistical metrics.

1. Closing Price Distribution

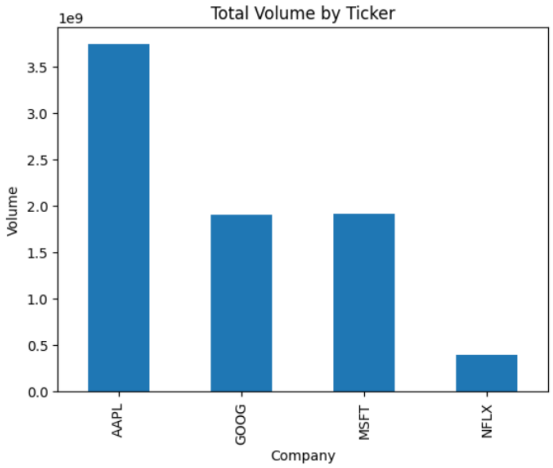
- The distribution of closing prices showed three distinct clusters: one for lower-valued stocks (like NFLX), one for mid-range prices (AAPL), and one for higher-value stocks (MSFT and GOOG).

- Most stock prices fall between $90–$370, with fewer entries in the $200 range, indicating varied price scales across the companies.



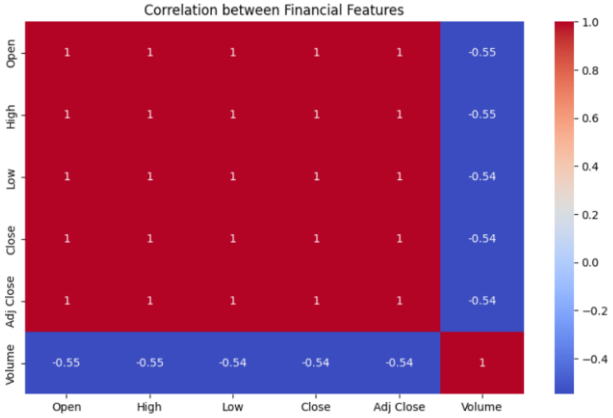
2. Total Volume by Ticker

- Apple (AAPL) had the highest total trading volume, significantly more than the other companies.
- Netflix (NFLX) had the lowest volume, indicating relatively lower trading activity.
- This suggests that AAPL is more actively traded, possibly due to its widespread popularity and market reach.



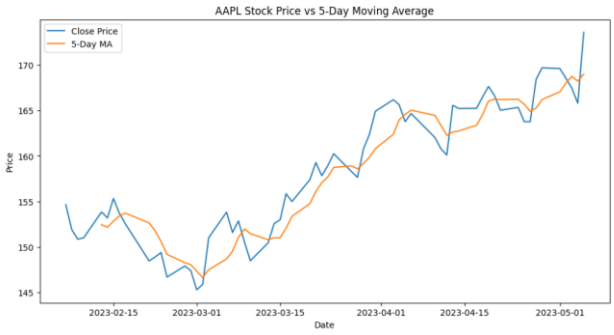
3. Correlation Between Financial Features

- A strong positive correlation (~1.0) exists between Open, High, Low, and Close prices, meaning these values move very closely together.
- Volume, however, shows a moderate negative correlation (~-0.54) with these features, indicating that higher volume days don't necessarily correspond to higher prices.



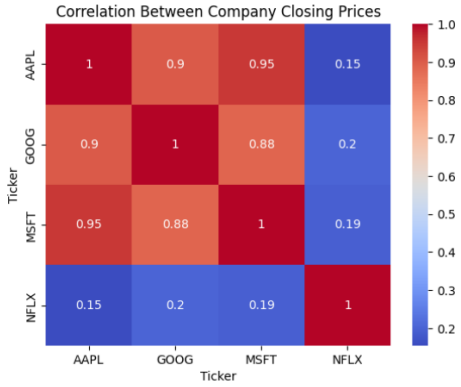
4. 5-Day Moving Average vs Close Price (AAPL)

- The 5-day moving average (MA) of Apple's stock tracks the closing price closely, but with a smoothing effect.
- This is useful for identifying price trends and potential support/resistance zones.
- The upward trend toward the end indicates positive price momentum for AAPL.



5. Correlation Between Company Closing Prices

- AAPL, MSFT, and GOOG show high correlation with each other (above 0.88), suggesting similar market behaviour and response to external factors.
- NFLX has very low correlation (~0.15–0.20) with the other three, indicating independent movement and possibly different market dynamics.



8. Conclusion

The stock market analysis successfully evaluated the short-term performance of four major technology companies — Apple, Microsoft, Google, and Netflix — using historical price and volume data. Through exploratory data analysis, feature engineering, and simple machine learning modelling, several key patterns were uncovered:

- Price-related features (Open, High, Low, Close) were highly correlated.
- Apple had the highest trading volume, indicating strong market interest.
- Netflix behaved more independently from the other stocks, with lower correlation.
- Technical indicators like moving averages and volatility helped visualize trends and risk levels.
- The linear regression model showed that stock prices could be reasonably predicted based on basic financial features.

This project demonstrates how data analytics can provide insights into stock behaviour and lay the foundation for future predictive modelling or algorithmic trading strategies.