**INTERNSHIP PROJECT REPORT**

**ON**

**“STOCK MARKET ANALYSIS”**

**SUBMITTED BY:**

**PRAYAS SAMAL**

**EXECUTIVE SUMMARY**

The daily stock data for four significant technological companies—Apple (AAPL), Microsoft (MSFT), Netflix (NFLX), and Google (GOOG)—covering the period from February 7, 2023, to May 5, 2023, is quantitatively analyzed in this study. The project's main objectives were statistical testing, exploratory data analysis, and creating a highly precise linear regression model for price prediction.

**Key Analytical Results**

Market Volatility: Netflix (NFLX) showed the highest average daily volatility, while Apple (AAPL) showed the lowest, according to the examination of daily price ranges.  
Trading Volume and Trends: During the analysis window, AAPL had the greatest total trading volume out of the four tickers. Over the course of the period, its closing price movement demonstrated a steady and evident increasing trajectory.

**Relationships in Statistics:**

A statistically significant difference between the mean closing prices of AAPL and GOOG was confirmed using a two-sample t-test.  
There was no statistically significant relationship between AAPL's daily closing price and trading volume, according to a Pearson correlation test.

**Modeling Predictively (AAPL Focus)**

To forecast the close price of AAPL, a Linear Regression model was used, utilizing features like Open, High, Low, and Volume. The model's outstanding performance confirmed these daily indicators' high predictive power:  
R-Squared: 0.9854  
This suggests that the input properties of the model account for about 98.54% of the closing price variance.  
MSE: 1.0942  
The model's forecasts closely match the actual observed closing prices, as indicated by the low MSE.

In summary, the information supports the different price patterns and trading patterns of the examined equities. Based on intraday market data, the created Linear Regression model for AAPL is very successful at short-term price predicting.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **SERIAL NO** | **PARTICULAR** | **PAGE NO** |
| 1. | EXECUTIVE SUMMARY | 2 |
| 2. | OBJECTIVES | 4 |
| 3. | DATA ANALYSIS | 5-15 |
| 4. | CONCLUSION | 16-17 |

**OBJECTIVES**

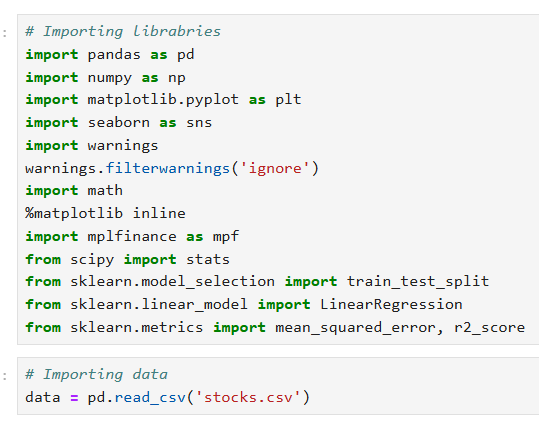
Conducting a thorough, data-driven examination of historical stock price and trade volume data for a few big technological companies (AAPL, MSFT, NFLX, and GOOG) covering the period from February 7, 2023, to May 5, 2023 is the main goal of this project module. Three main objectives are intended to be accomplished by this analysis:

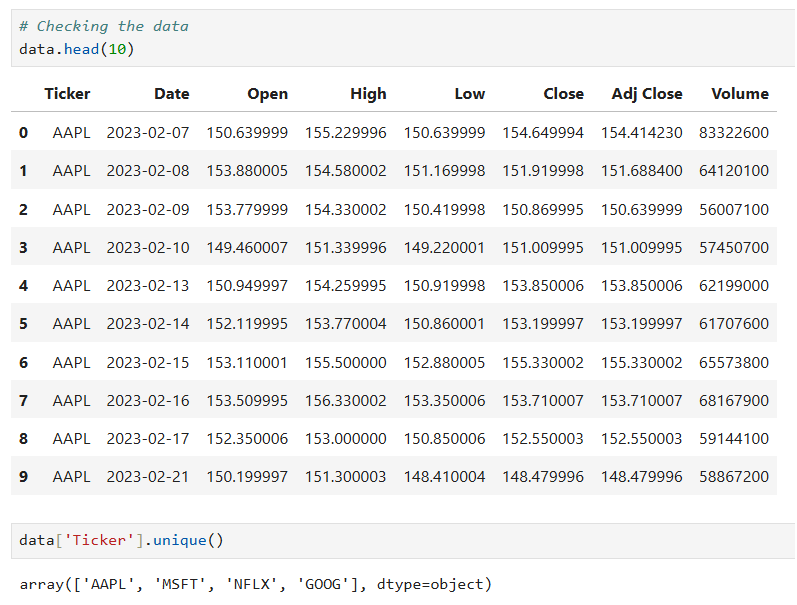
Exploratory Data Insight: To conduct thorough Exploratory Data Analysis (EDA) in order to describe market dynamics, such as price distributions, total trading volume by ticker, volatility of the daily price range, and general price trends over the observation period.

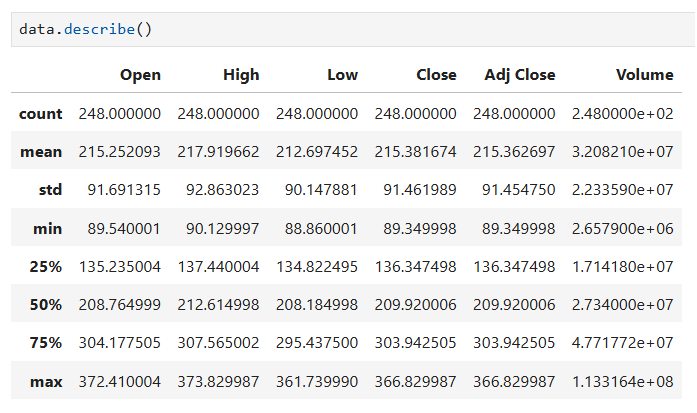
Statistical Validation: To statistically validate observable market behaviors, such as substantial differences in mean closing prices of large stocks and the relationship between trading volume and price movements, by using hypothesis testing (e.g., Pearson correlation and two-sample t-test).

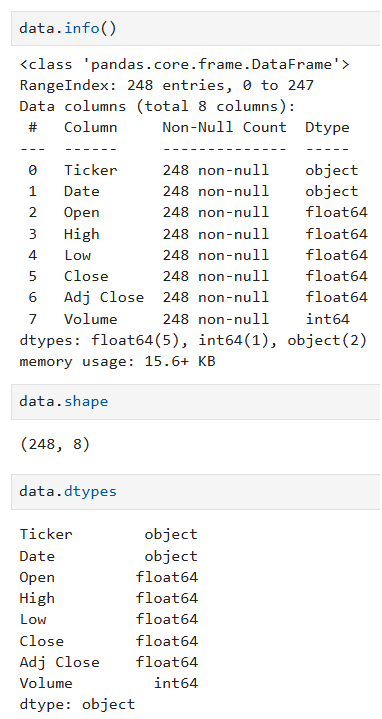
Predictive Model Development: To build, train, and assess a Linear Regression model that can reliably estimate daily closing prices using the knowledge gathered from the EDA and statistical validation. This will create a baseline for short-term forecasting abilities.

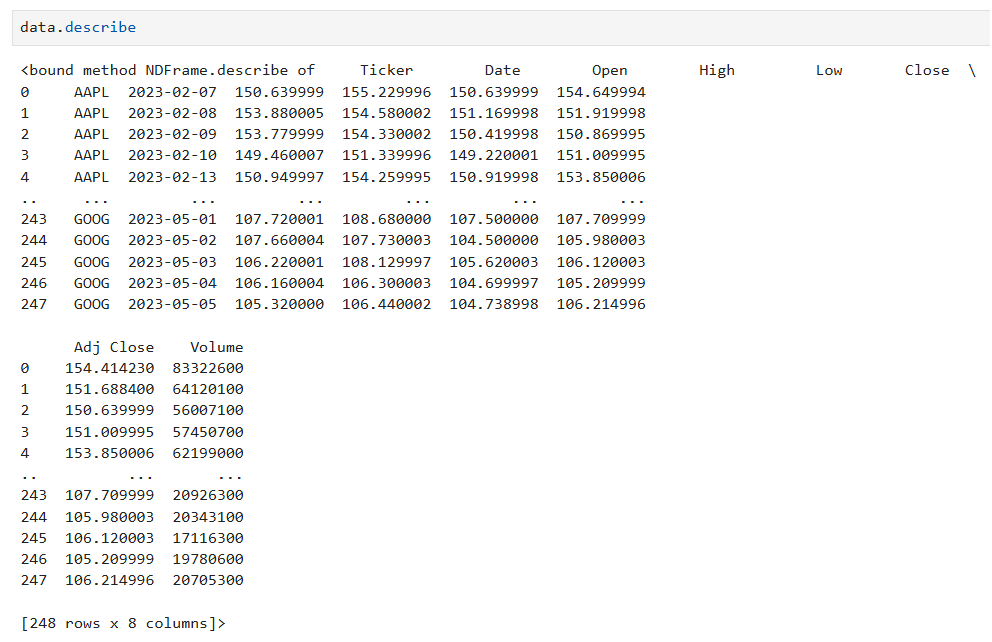
**DATA ANALYSIS**

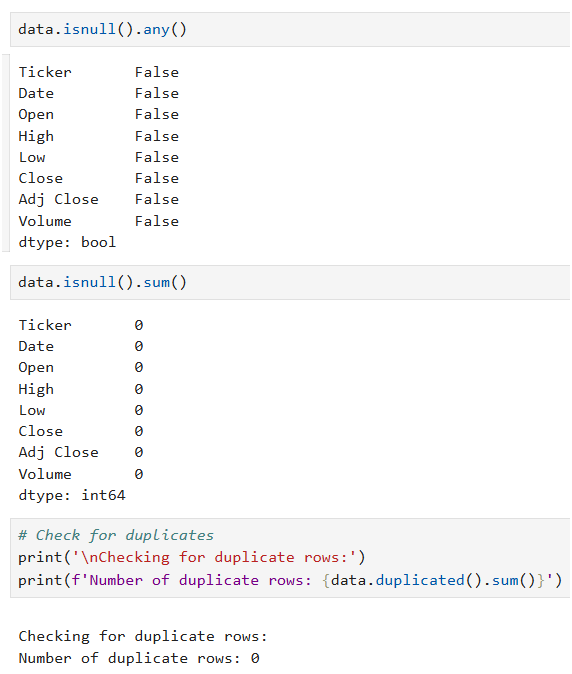
****

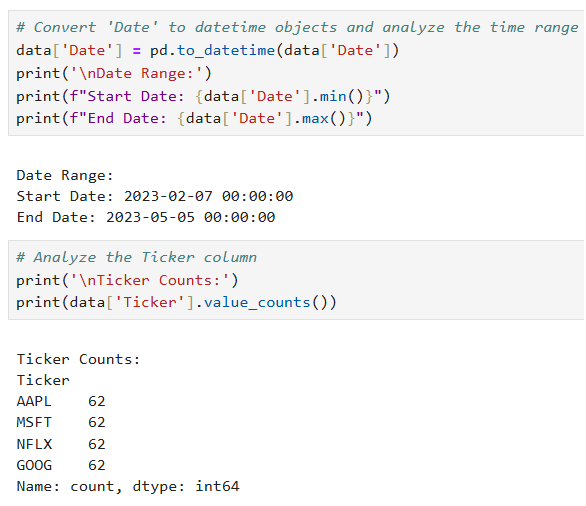
****

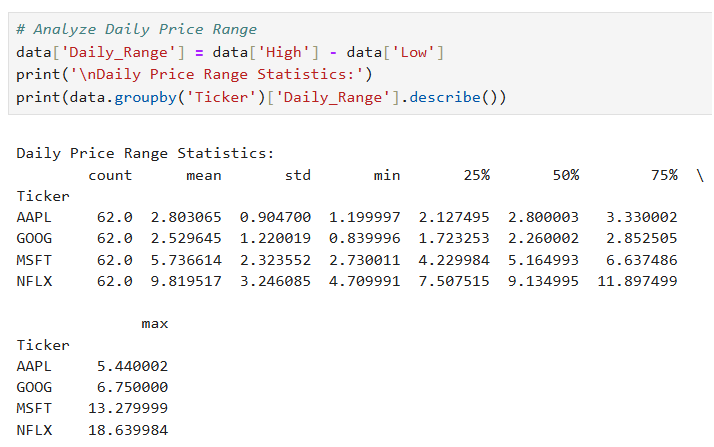
****

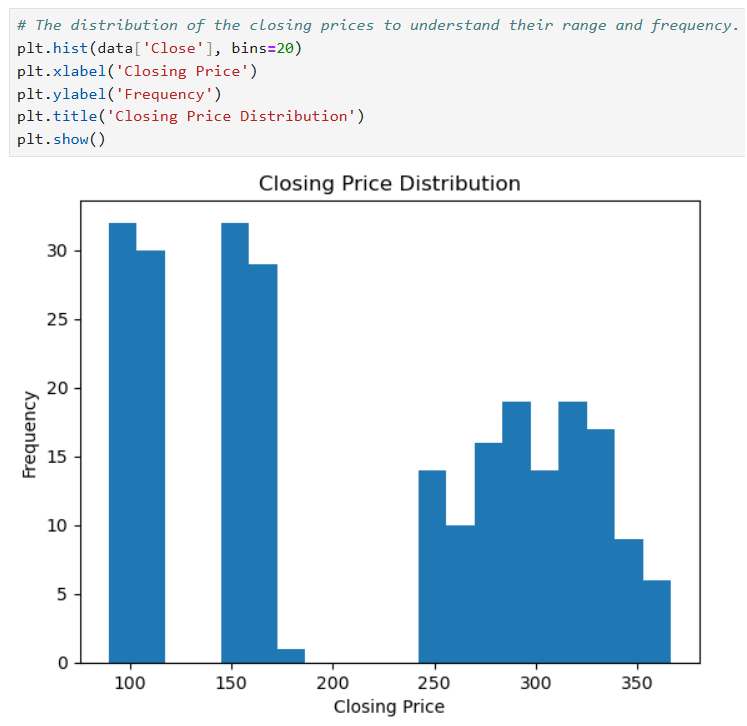
****

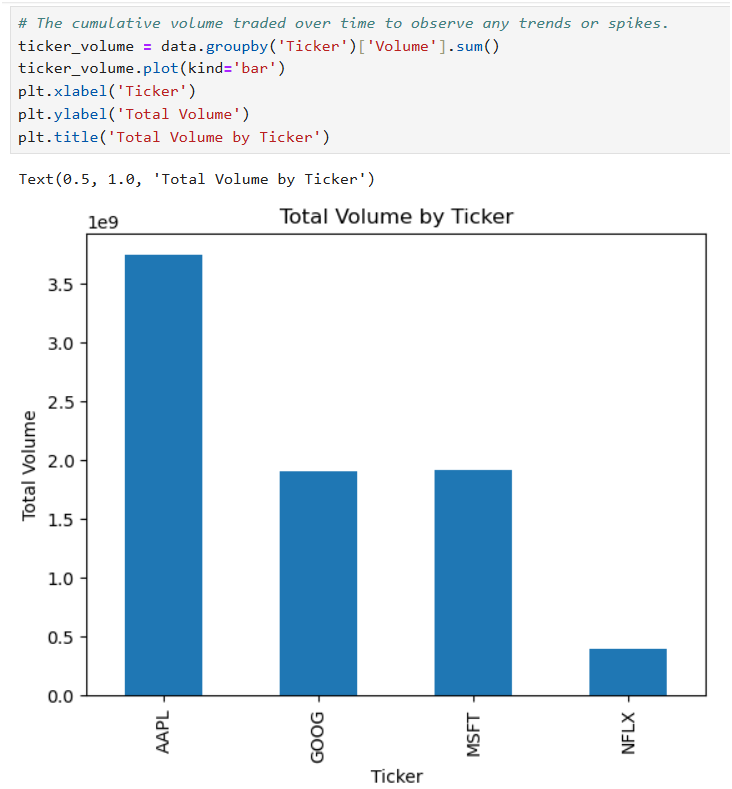
****

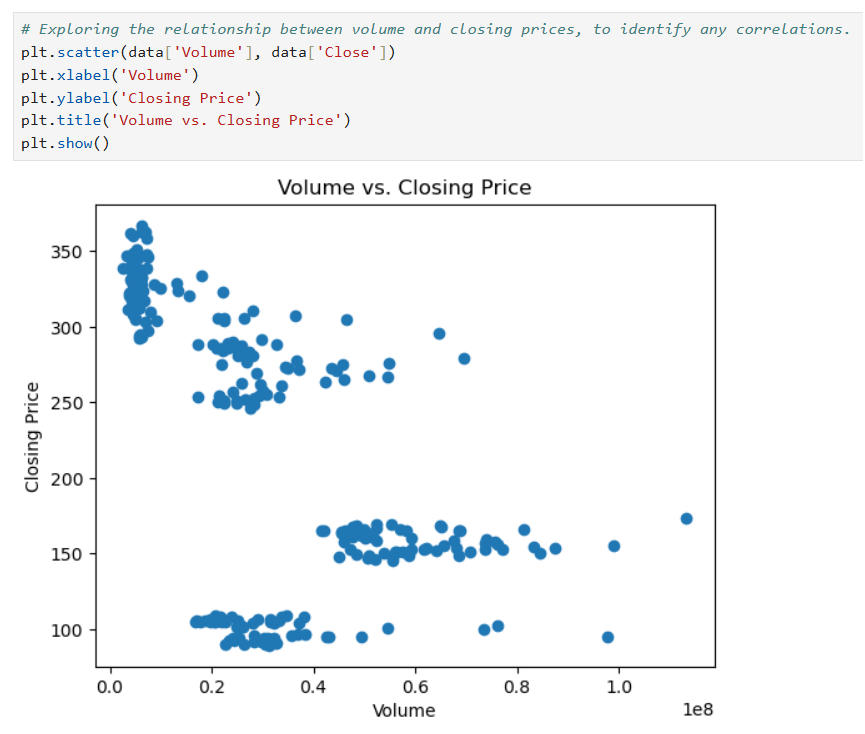
****

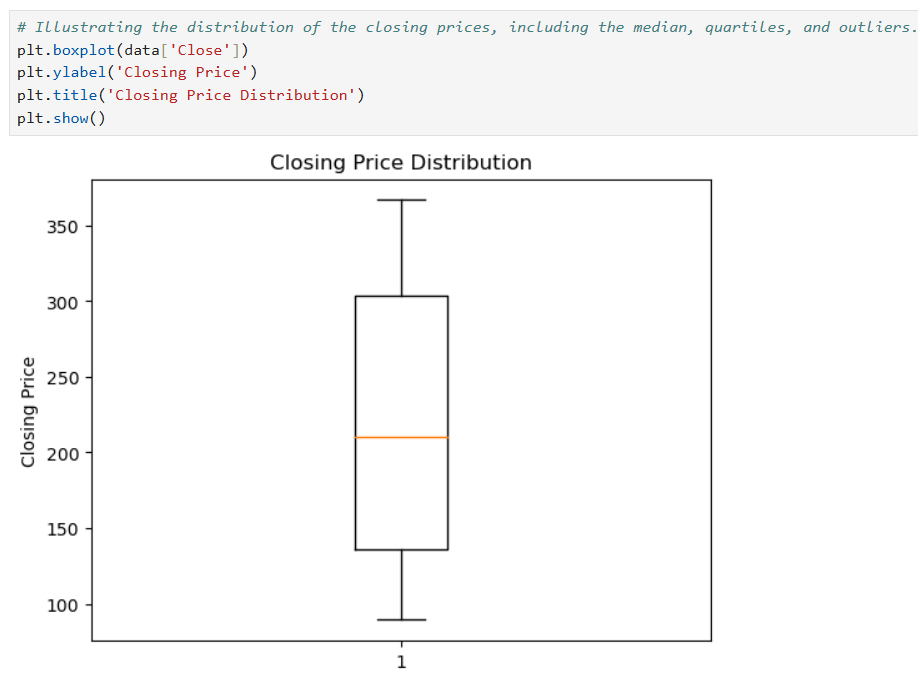
****

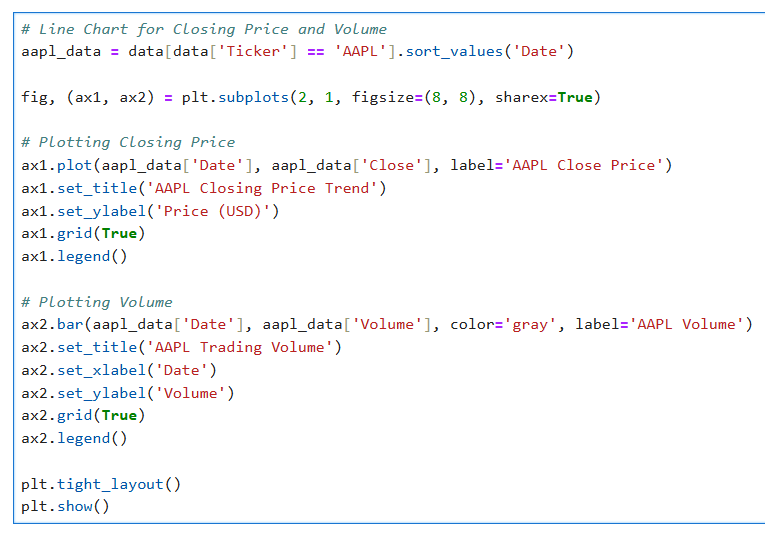
****

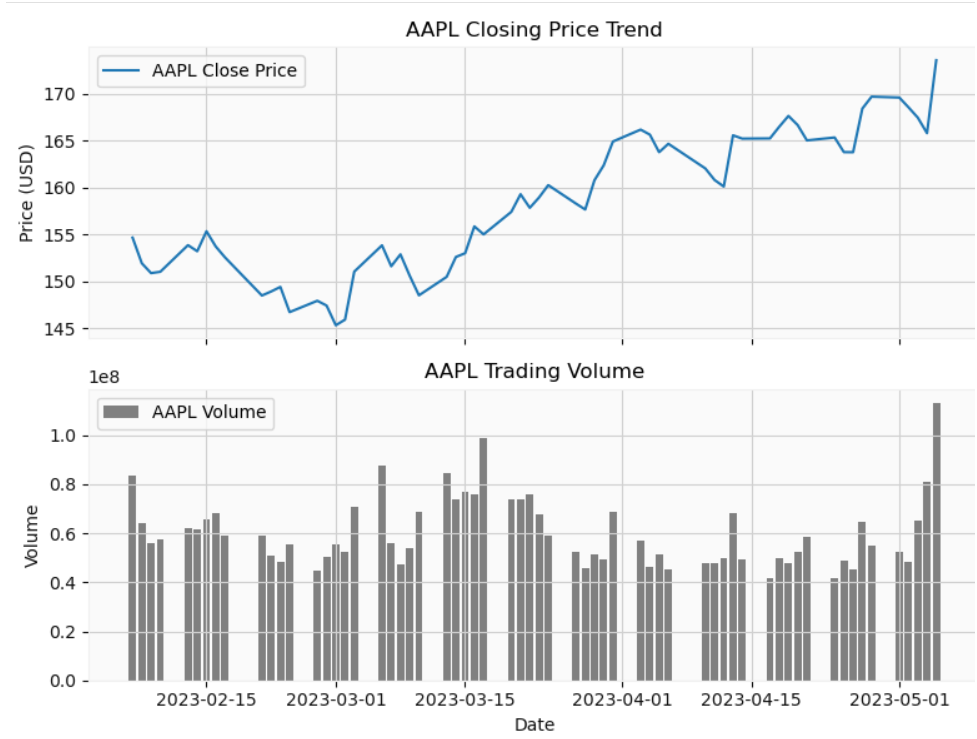
****

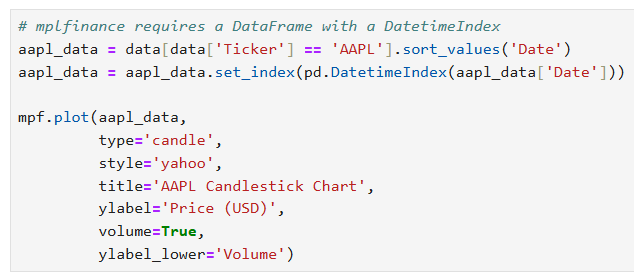
****

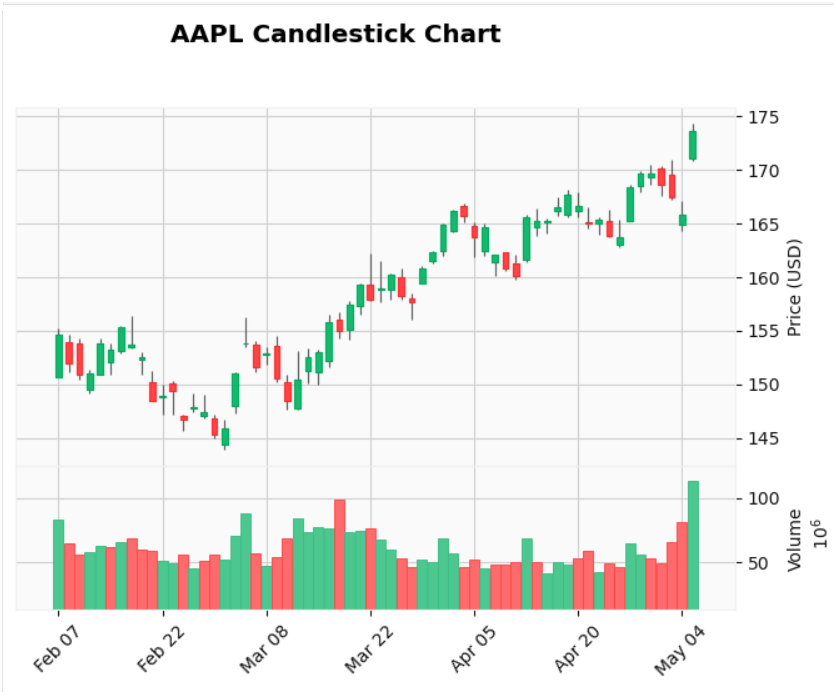
****

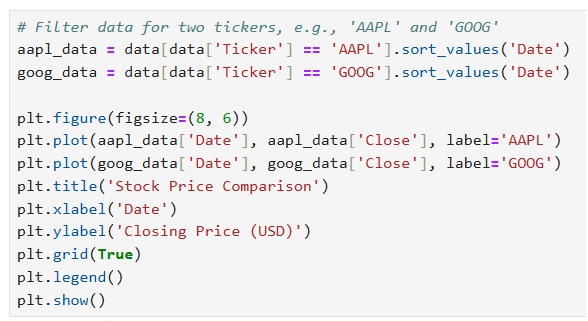
****

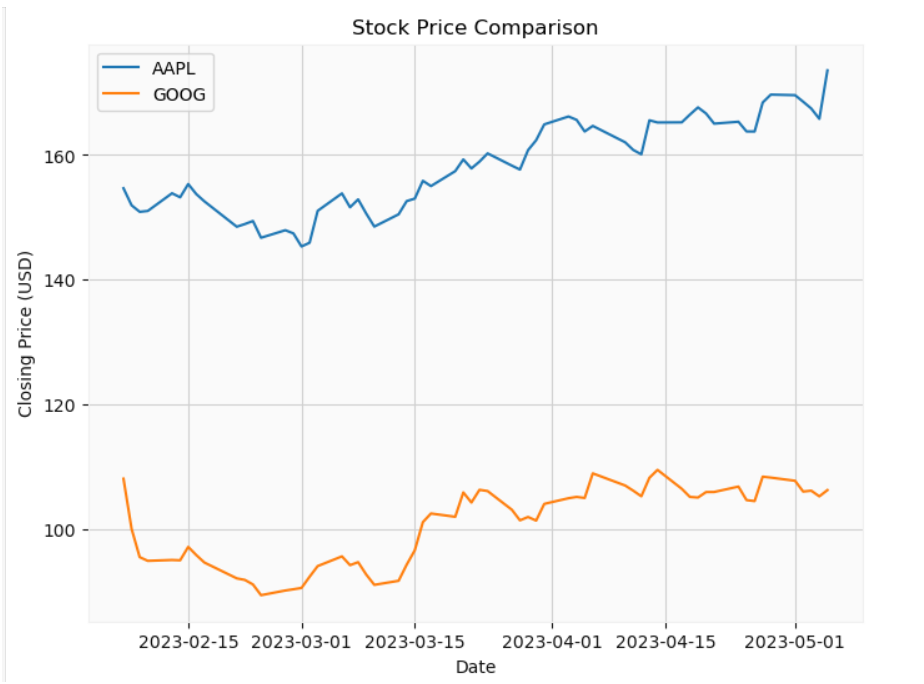
****

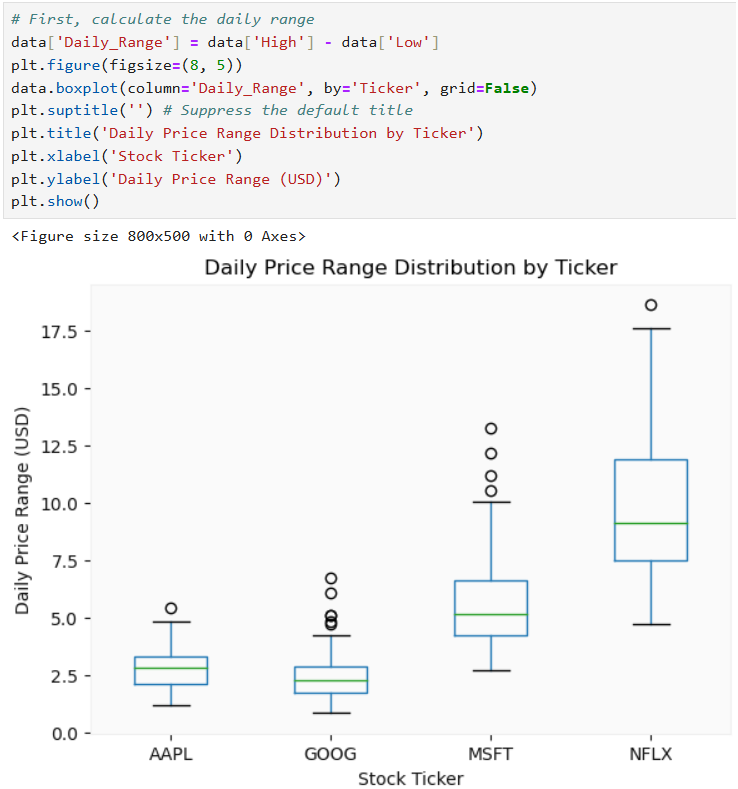
****

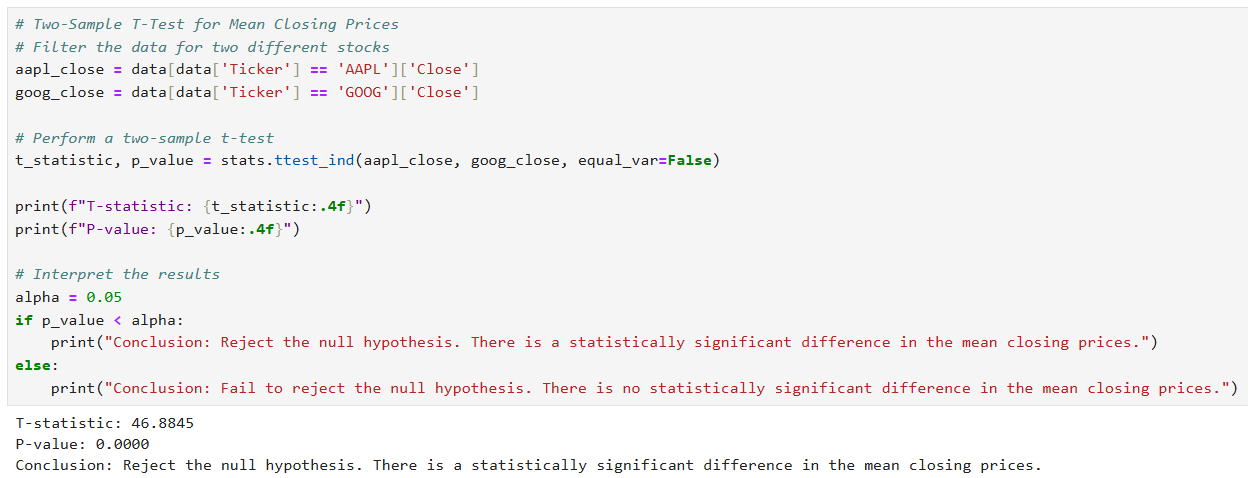
****

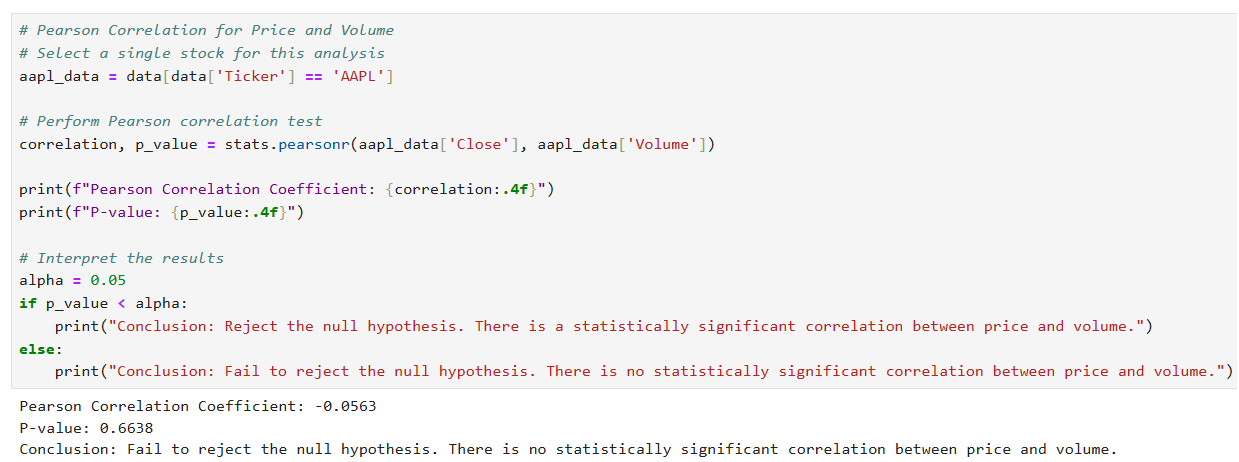
****

****

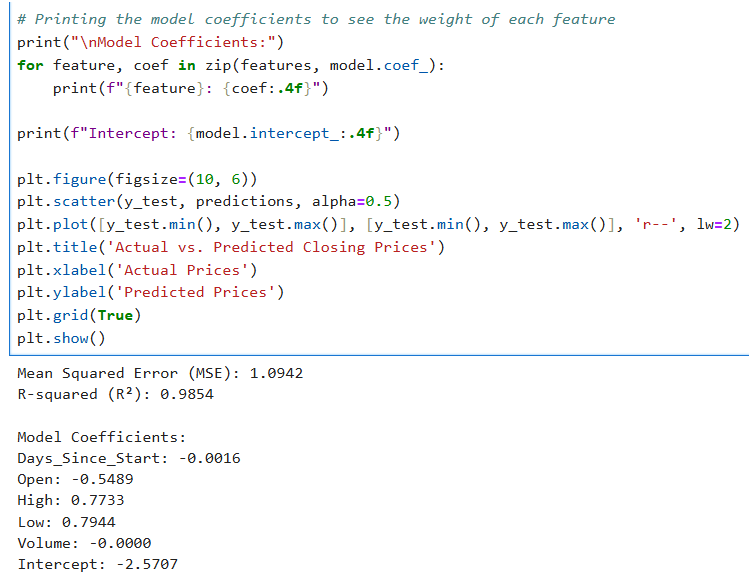
****

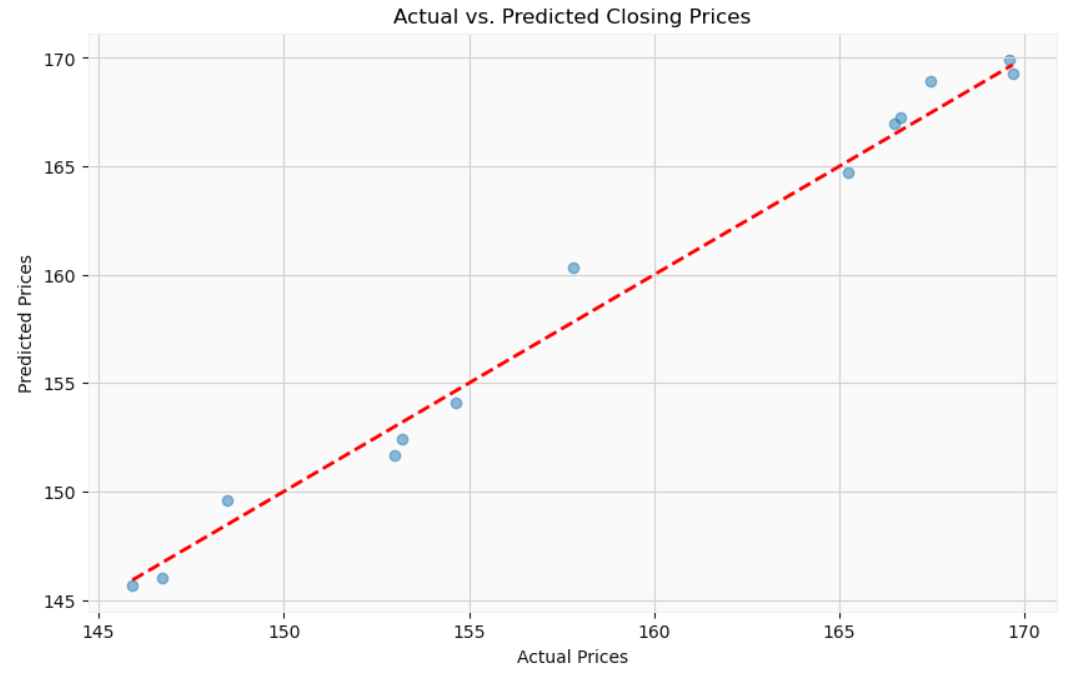
****

****

****

****

****

****

**CONCLUSION**

By creating a very effective baseline predictive model and offering a thorough quantitative analysis of stock data for AAPL, GOOG, MSFT, and NFLX, the analysis module effectively met its goals.

**Summary of Key Findings**

Exploratory Data Analysis (EDA) allowed us to identify significant variations in market behavior:

Volatility: Compared to the other tickers, especially Apple (AAPL), which showed the least amount of volatility, Netflix (NFLX) showed the biggest mean daily price range, indicating higher intraday volatility.

Price Trends: Throughout the observation period, AAPL's closing price trend showed a steady and robust increasing trajectory, in contrast to Google's somewhat flat trend.

Statistical Differences: The mean closing prices of AAPL and GOOG showed a statistically significant difference (alpha = 0.05) according to a two-sample t-test, confirming that both stocks have different pricing mechanisms. On the other hand, there was no statistically significant linear link between AAPL's closing price and trading volume over this time frame, according to the Pearson correlation test.

**Model Outcomes and Consequences**

A Mean Squared Error (MSE) of 1.0942 and a R-Square score of 0.9854 were two outstanding performance indicators for the constructed Linear Regression model used to forecast AAPL closing prices. The model's selected features (Open, High, Low, and Days\\_Since\\_Start) are very good predictors of the closing price, as seen by this remarkably high $R^2$ value, which makes the model a trustworthy instrument for short-term price estimation. Additionally, the coefficient analysis showed that the Low and High prices had the most influence on the Close price.

**Future Work**

Although the linear regression model has a high degree of short-term predictive power, future research should concentrate on expanding this analysis to include more complex methods in order to account for non-linear market dynamics:

Time Series Modeling: Use sophisticated time series models, like Prophet or ARIMA, to predict future prices that take into account temporal relationships and go beyond a single trading day.

Feature Engineering: To potentially increase model robustness and forecast accuracy, particularly for equities with higher volatility like NFLX, investigate other technical indicators (such as moving averages and RSI) as input features.

Comparative Predictive Analysis: To offer a performance comparison of the various stock types, apply the Linear Regression modeling to MSFT, NFLX, and GOOG.