
Predicting Future Stock Prices
Heather Adler
Data Science Intensive
Capstone Project
Adobe Stock | #180008799

The Problem

Stock Market Volatility:

- Inherent market fluctuations and unpredictability.
- Rapid price changes and market uncertainty.

Lack of Accurate Prediction Tools:

- Traditional methods often fail to predict accurately.
- Potential financial losses due to inaccurate predictions.

Complexity of Data:

- Vast amounts of historical prices, trading volumes, and economic indicators.
- Difficulty in analyzing and interpreting data effectively.

Impact on Investors:

- Financial losses from poor investment decisions.
- Challenges in making timely and informed decisions.
- Deterrence of potential investors, reducing market liquidity.

Key Questions the Project Aims to Answer:

1) **Can we accurately predict future stock prices using historical data?**

2) **What is the expected accuracy and reliability of the predictive model?**

Who might care?

Individual Investors

Robinhood

E*TRADE

Financial Advisors and Portfolio Managers

Edward Jones
MAKING SENSE OF INVESTING

charles SCHWAB

Hedge Funds and Institutional Investors

BRIDGEWATER

Vanguard®

Financial Analysts and Researchers

Bloomberg

THOMSON REUTERS®

MORNINGSTAR®

Data Details

Estée Lauder Companies Inc. (EL)	Ulta Beauty Inc. (ULTA)	Coty Inc. (COTY)	e.l.f. Beauty Inc. (ELF)
• No of records: 7,138	• No of records: 4,134	• No of records: 2,717	• No of records: 1,891

Data Fields for Each Stock Symbol:

• stock_symbol	• Date	• Open	• High
• Low	• Close	• Volume	• Dividends
• Stock_Splits	• EMA_10	• PSARl_0.02_0.2	• PSARs_0.02_0
• BBL_5_2.0	• BBM_5_2.0	• BBU_5_2.0	• ISA_9
• ISB_26	• ITS_9	• IKS_26	• ICS_26

Data Field Definitions

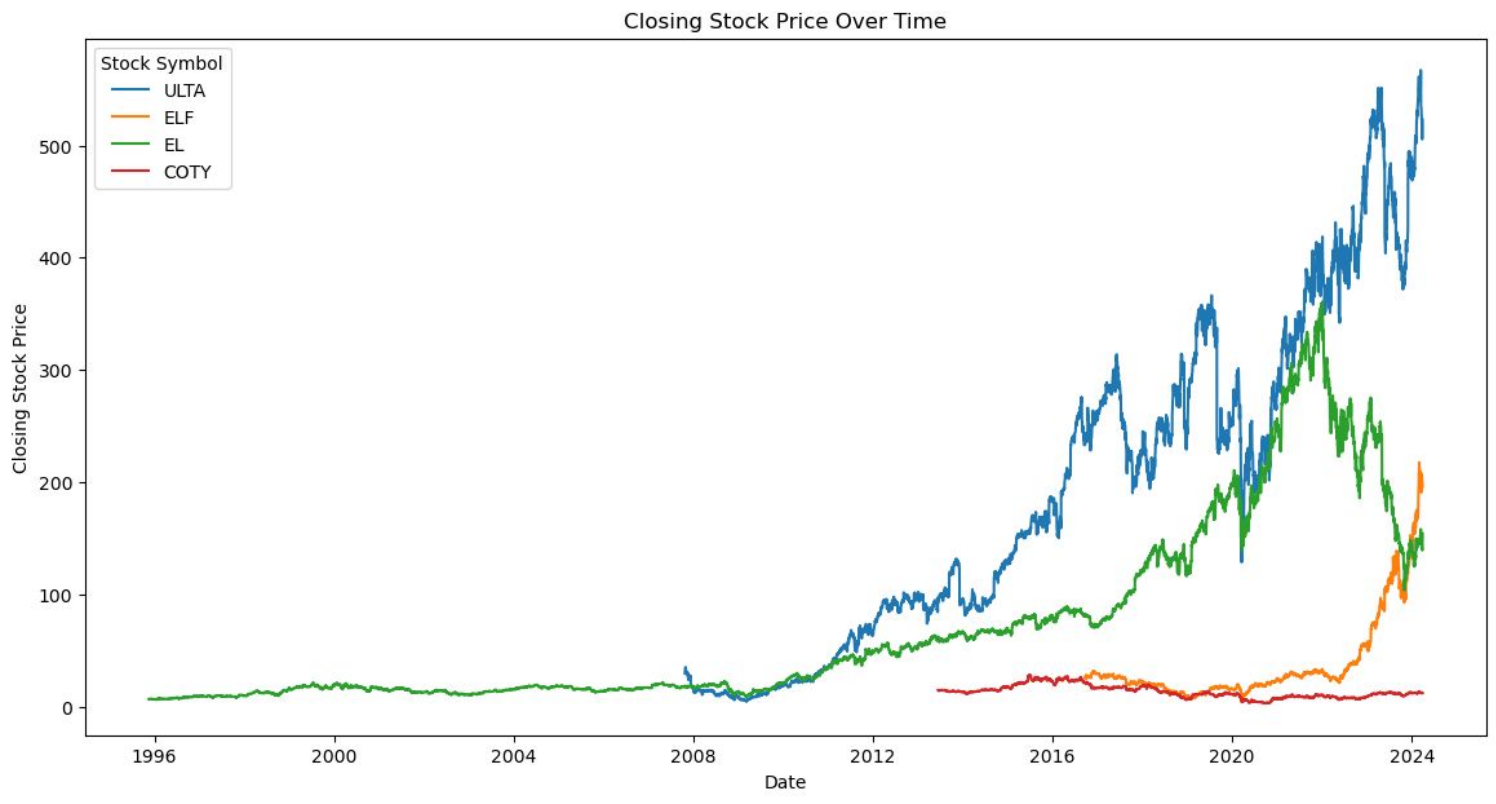
- **EMA_10: T**his stands for Exponential Moving Average. Unlike a simple moving average (SMA), which gives equal weight to all data points in the period, the EMA gives more weight to recent data points. This weighting is exponentially decreasing, with the most recent data points having the greatest impact on the average. The '10' represents the period used for calculating the EMA. In this case, it's 10 periods, meaning the EMA is calculated based on the closing prices of the last 10 periods.
- **PSARl_0.02_0.2:** This stands for Parabolic Stop and Reverse for long positions. It indicates that the PSAR value being calculated is specifically for long positions. The '.02' represents the acceleration factor or step used in the PSAR calculation. The acceleration factor determines the rate at which the PSAR moves closer to the price. In this case, it's set to 0.02. The '0.2' represents the maximum acceleration factor. Once the PSAR reaches this value, it stops increasing its acceleration and remains at this maximum value. In this case, it's set to 0.2.
- **PSARs_0.02_0.2:** Parabolic Stop and Reverse Short. It is similar to PSAR Long but used for short positions. TZhe parameters (0.02, 0.2) refer to the acceleration factor and maximum acceleration factor used in the PSAR calculation- The acceleration factor (AF) starts at 0.02 and increases by 0.02 each time a new extreme point (EP) is reached. The maximum acceleration factor limits the AF to 0.2.
- **BBL_5_2.0, BBM_5_2.0, BBU_5_2.0: B**ollinger Bands with parameters (5, 2.0). They consist of a middle band (BBM), upper band (BBU), and lower band (BBL). They are used to identify potential overbought or oversold conditions and measure volatility.
    a. The parameters (5, 2.0) typically represent:
    b. The period (5): This is the number of periods used in calculating the moving average. In this case, it's 5 periods.
    c. The standard deviation multiplier (2.0): This is the number of standard deviations used to calculate the width of the bands. In this case, it's 2.0, meaning the bands will be placed 2 standard deviations above and below the moving average.
        i. **BBL_5_2.0 specifically refers to the lower Bollinger Band with a period of 5 and a standard deviation multiplier of 2.0. It helps traders identify potential support levels based on volatility around the moving average.**
        ii. **BBM_5_2.0 denotes the Middle Bollinger Band with a period of 5 and a standard deviation multiplier of 2.0. This band serves as a reference level for the price movement and can help traders identify potential trends.**
        iii. **BBU_5_2.0 denotes the Upper Bollinger Band with a period of 5 and a standard deviation multiplier of 2.0. The Upper Band serves as an upper boundary or resistance level, helping traders identify potential overbought conditions or areas where prices may revert to the mean.**

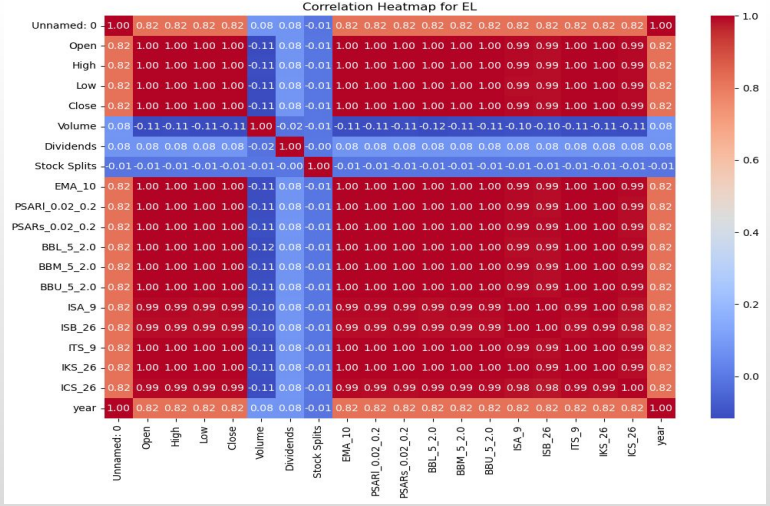
Data Field Definitions, Continued

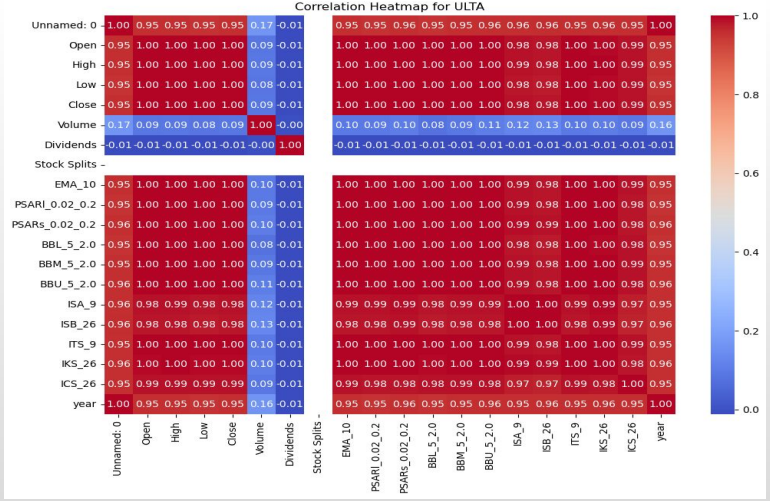
- **ISA_9, ISB_26, ITS_9, IKS_26, ICS_26:** Ichimoku Cloud components (Conversion Line, Base Line, Leading Span A, Leading Span B, and Lagging Span). They are used in the Ichimoku Cloud indicator to identify support, resistance, and trend direction.
    a. The Ichimoku Cloud indicator is a comprehensive trend-following indicator used in technical analysis. It consists of several components, including the Senkou Span A and Senkou Span B lines, which together form the "cloud" or "kumo." These lines are calculated based on specific periods and plotted ahead of the current price to indicate potential future support and resistance levels.
    b. **ISA_9- T**he Senkou Span A line, also known as "Leading Span A," is calculated as the average of the Tenkan-sen (Conversion Line) and Kijun-sen (Base Line) plotted forward by a certain number of periods. In the case of "ISA_9," it's calculated with a period of 9. Therefore, "ISA_9" represents the Ichimoku Senkou Span A with a period of 9 in the Ichimoku Cloud indicator. It's used to assess potential future support and resistance levels and to identify trend direction in the market.
    c. **ISB_26-** The Senkou Span B line, also known as "Leading Span B," is calculated similarly to Senkou Span A but with a longer period. It's typically calculated as the average of the highest high and lowest low over the past 26 periods, plotted forward by 26 periods. Therefore, "ISB_26" represents the Ichimoku Senkou Span B with a period of 26 in the Ichimoku Cloud indicator. It's used alongside Senkou Span A to provide additional insight into potential future support and resistance levels and to assess the overall trend direction in the market.
    d. **ITS_9-** The Tenkan-sen, or Conversion Line, is one of the components of the Ichimoku Cloud indicator. It's calculated as the average of the highest high and lowest low over the past 9 periods. The result is then plotted on the chart. Therefore, "ITS_9" represents the Tenkan-sen with a period of 9 in the Ichimoku Cloud indicator. It's used to assess short-term momentum and potential trend reversals in the market.
- **IKS_26-** The Kijun-sen, or Base Line, is another key component of the Ichimoku Cloud indicator. It's calculated as the average of the highest high and lowest low over the past 26 periods. The result is then plotted on the chart. Therefore, "IKS_26" represents the Kijun-sen with a period of 26 in the Ichimoku Cloud indicator. It's used to assess medium-term momentum and potential trend reversals in the market.
- **ICS_26-** The Chikou Span is used to assess momentum and confirm trend direction. When the Chikou Span is above the price plot, it suggests bullish momentum, indicating potential upward movement in prices. Conversely, when the Chikou Span is below the price plot, it suggests bearish momentum, indicating potential downward movement in prices. Therefore, "ICS_26" represents the Chikou Span with a period of 26 in the Ichimoku Cloud indicator, and it's used to confirm the strength and direction of the prevailing trend.

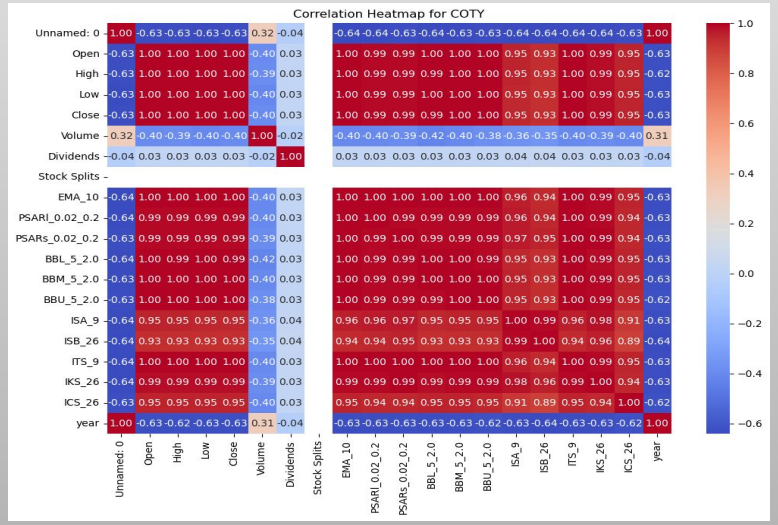

Data
Exploration
Adobe Stock | #180008799

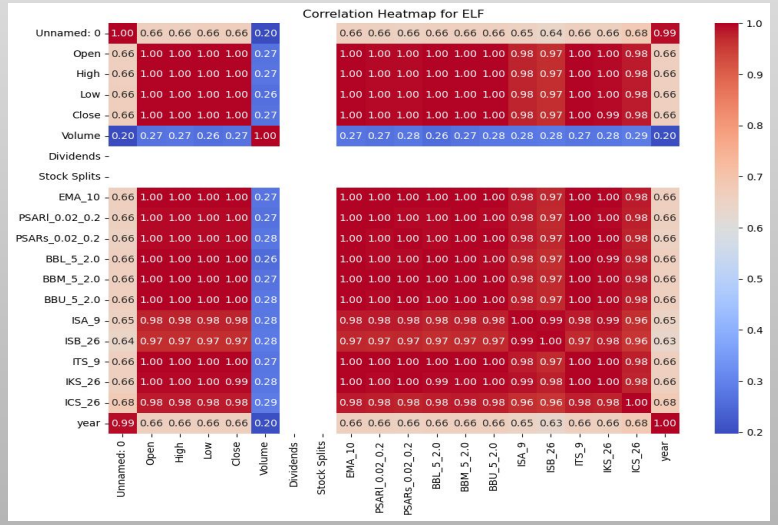
Closing Stock Price Over Time




Correlation Heatmap for EL
Unnamed: 0
Open
High
Low
Close
Volume
Dividends
Stock Splits
EMA_10
PSARl_0.02_0.2
PSARs_0.02_0.2
BBL_5_2.0
BBM_5_2.0
BBU_5_2.0
ISA_9
ISB_26
ITS_9
IKS_26
ICS_26
year

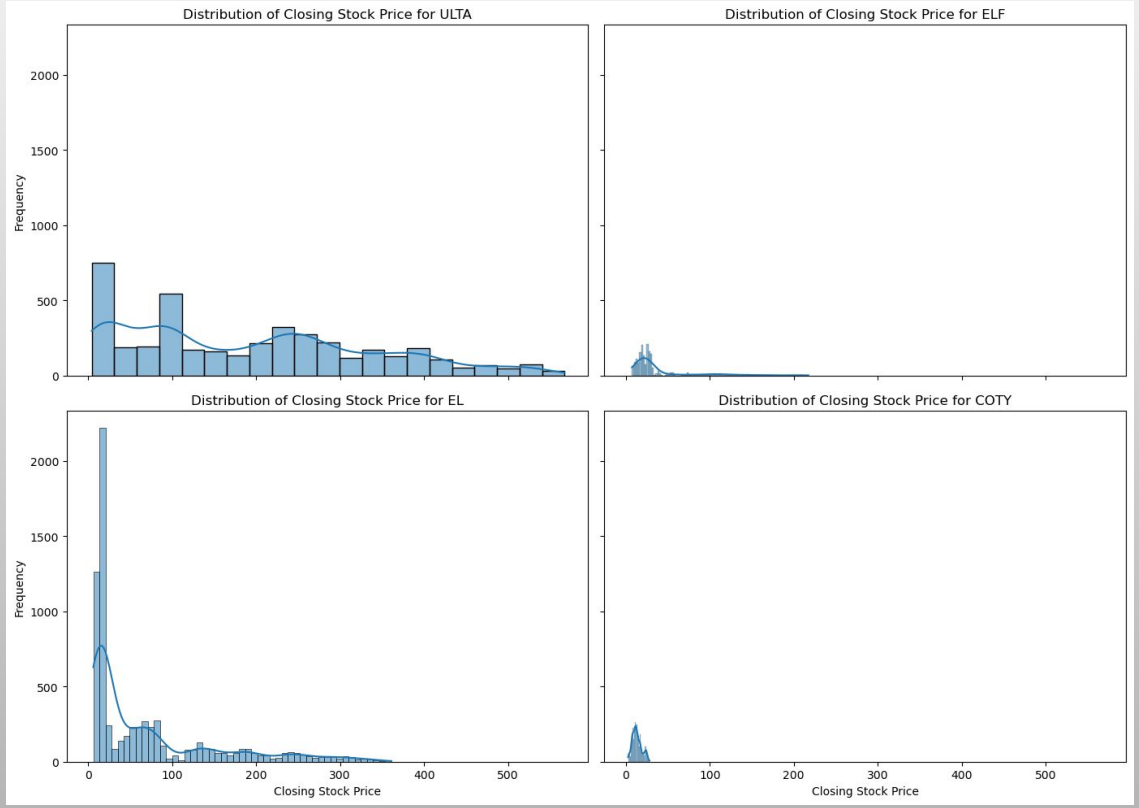

Correlation Heatmap for ULTA
Unnamed: 0
Open
High
Low
Close
Volume
Dividends
Stock Splits
EMA_10
PSARl_0.02_0.2
PSARs_0.02_0.2
BBL_5_2.0
BBM_5_2.0
BBU_5_2.0
ISA_9
ISB_26
ITS_9
IKS_26
ICS_26
year

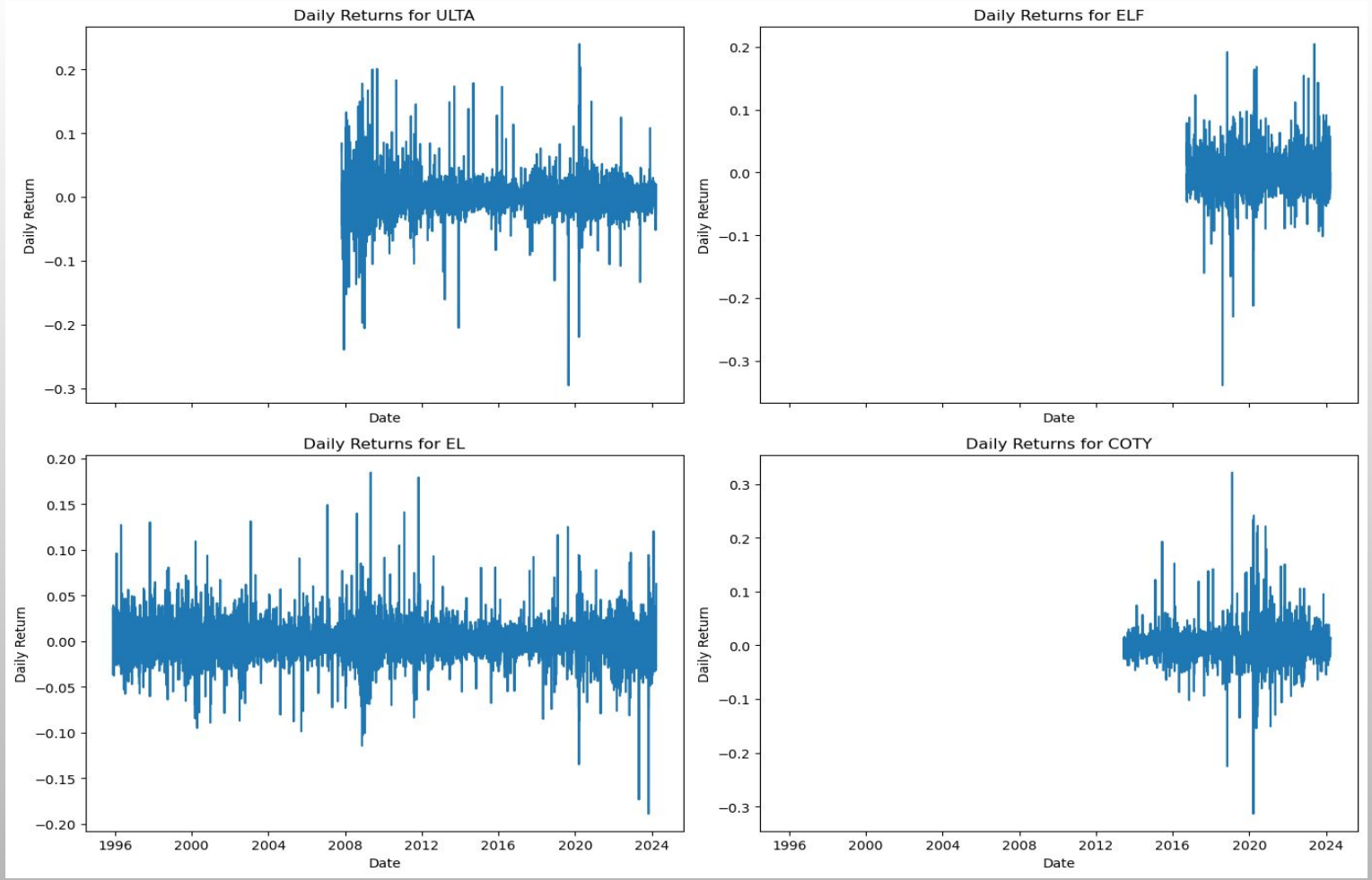

Correlation Heatmap for COTY
Unnamed: 0
Open
High
Low
Close
Volume
Dividends
Stock Splits
EMA_10
PSARl_0.02_0.2
PSARs_0.02_0.2
BBL_5_2.0
BBM_5_2.0
BBU_5_2.0
ISA_9
ISB_26
ITS_9
IKS_26
ICS_26
year

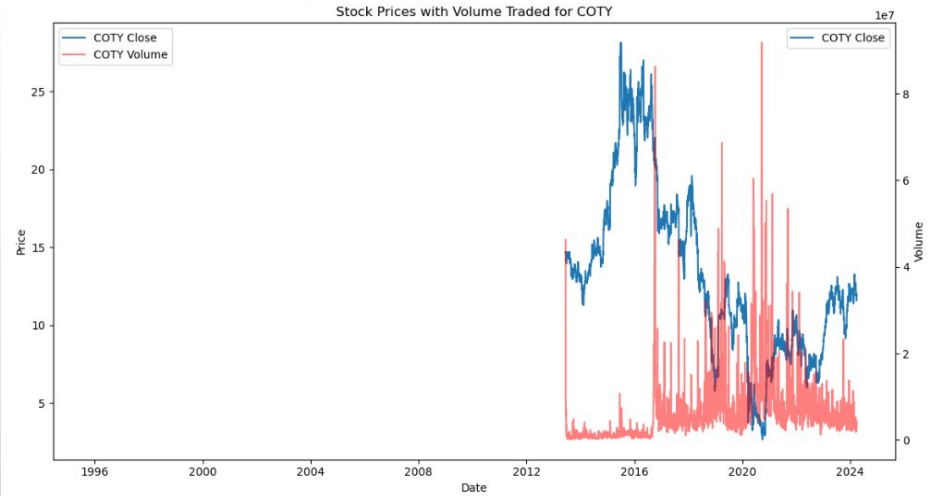

Correlation Heatmap for ELF
Unnamed: 0
Open
High
Low
Close
Volume
Dividends
Stock Splits
EMA_10
PSARl_0.02_0.2
PSARs_0.02_0.2
BBL_5_2.0
BBM_5_2.0
BBU_5_2.0
ISA_9
ISB_26
ITS_9
IKS_26
ICS_26
year

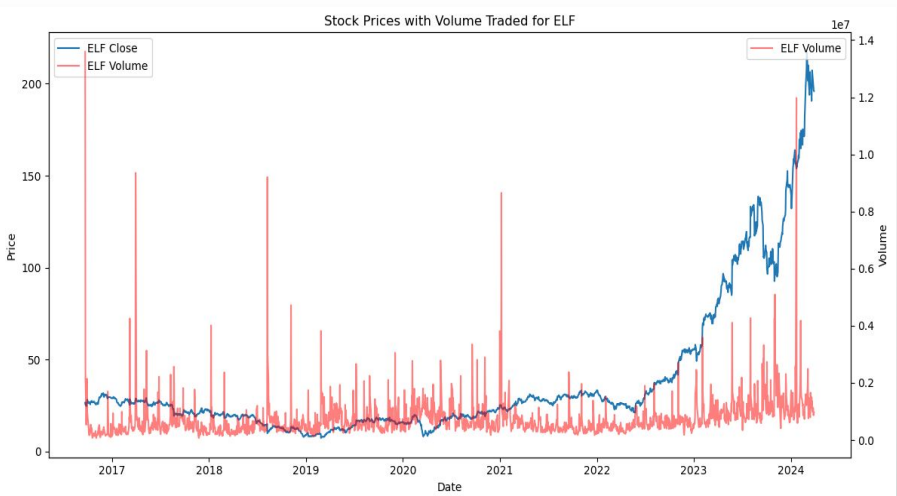
Visualization of the Target Variable (Close Price)

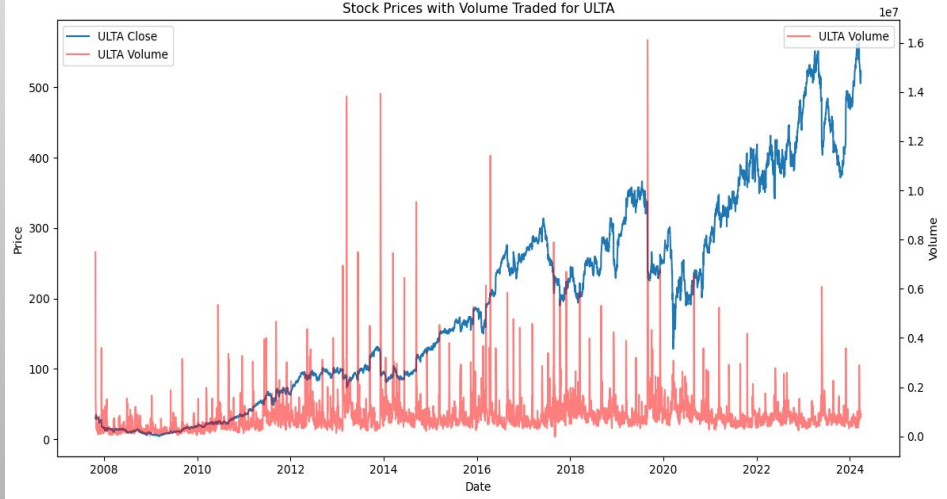
The Target variable in this case is the closing stock price. The below charts show the frequency for which the closing stock price occurred and shows that ULTA and EL had the highest closing stock prices at a higher frequency. This suggests that these companies may have experienced relatively strong performance or investor confidence during this period.

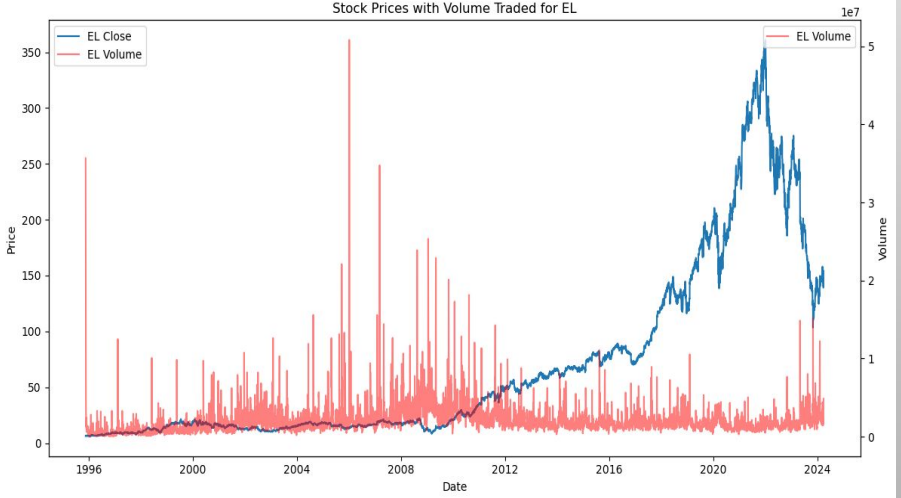



Daily Returns for ULTA
Daily Return
Date
Daily Returns for ELF
Daily Return
Date
Daily Returns for EL
Daily Return
Date
Daily Returns for COTY
Daily Return
Date


Stock Prices with Volume Traded for COTY
COTY Close
COTY Volume
1e7
Price
Volume
Date


Stock Prices with Volume Traded for ELF
ELF Close
ELF Volume
1e7
Price
Volume
Date


Stock Prices with Volume Traded for ULTA
ULTA Close
ULTA Volume
1e7
Price
Volume
Date


Stock Prices with Volume Traded for EL
EL Close
EL Volume
1e7
Price
Volume
Date


Data
Modeling
Adobe Stock | #180008799

Defining & Training the LSTM Model

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 50, 50)	11,000
dropout (Dropout)	(None, 50, 50)	0
lstm_1 (LSTM)	(None, 50)	20,200
dropout_1 (Dropout)	(None, 50)	0
dense (Dense)	(None, 1)	51

Total params: 31,253 (122.09 KB)

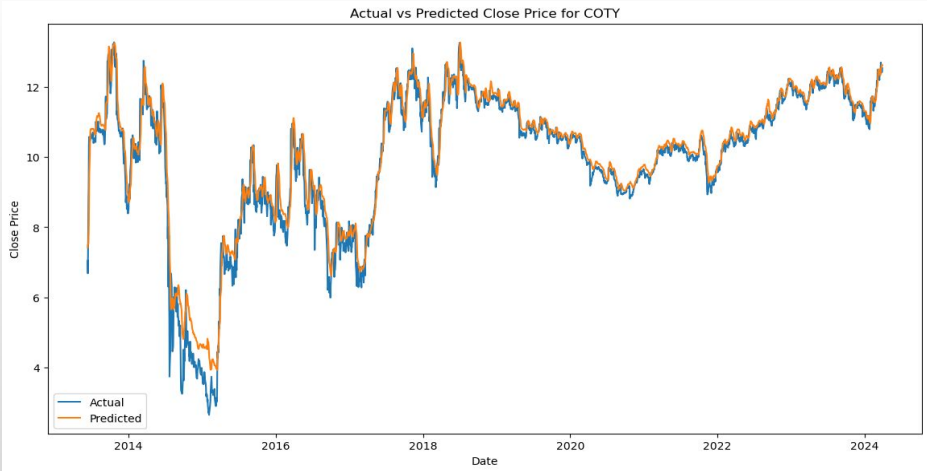
Trainable params: 31,251 (122.07 KB)

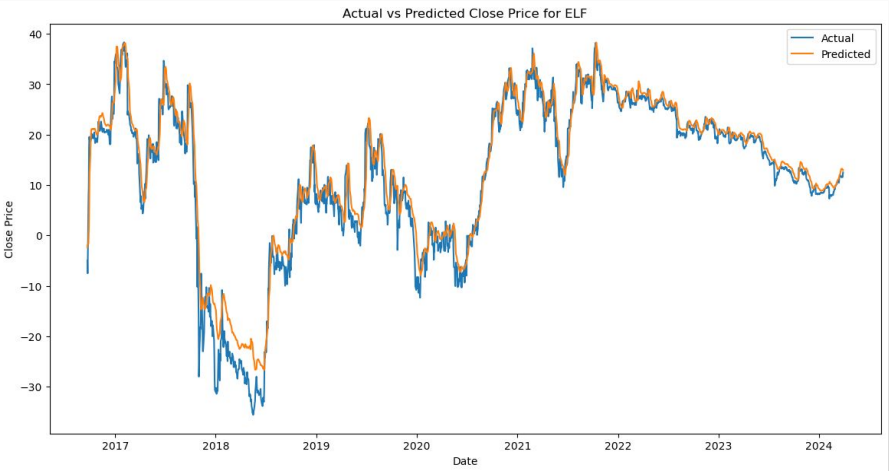
Non-trainable params: 0 (0.00 B)

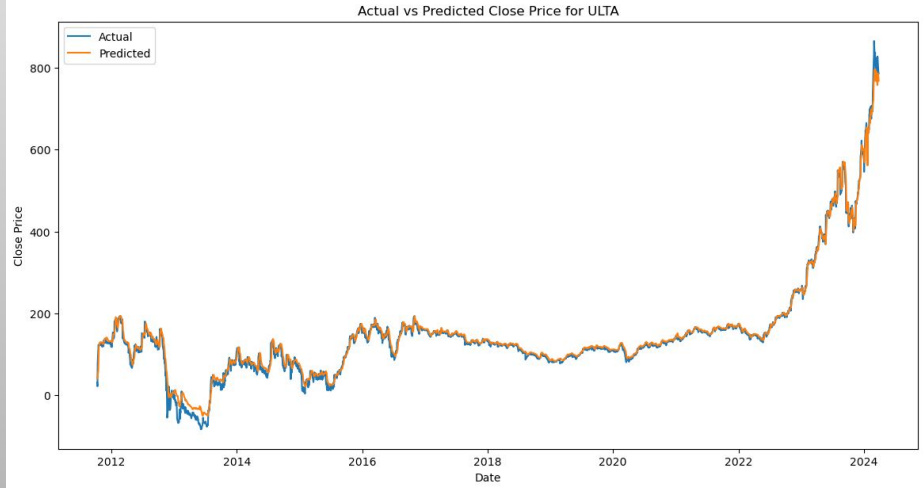
Optimizer params: 2 (12.00 B)

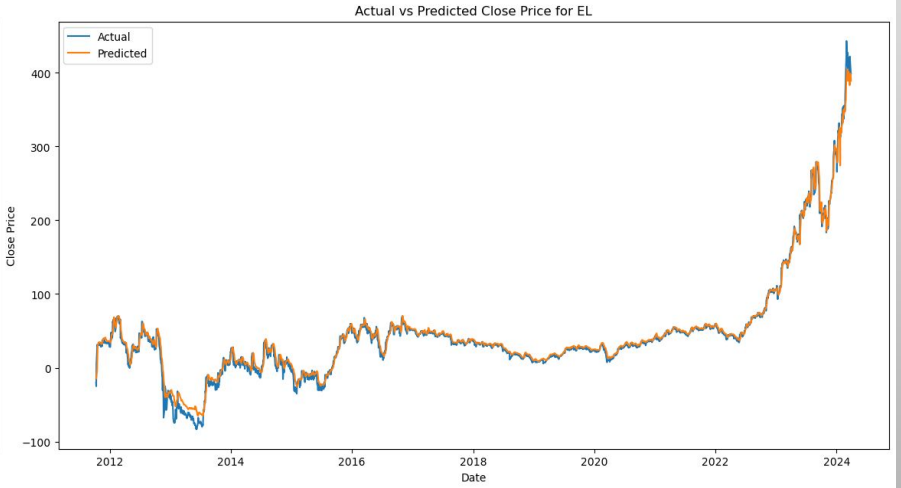
Model Evaluation and Prediction

```
Evaluating model for EL...
98/98 ━━━━━━━━━━━━━━━━━━━━ 1s 6ms/step
Root Mean Squared Error (RMSE) for EL: 6.9571
Evaluating model for ULTA...
98/98 ━━━━━━━━━━━━━━━━━━━━ 1s 7ms/step
Root Mean Squared Error (RMSE) for ULTA: 12.5352
Evaluating model for COTY...
98/98 ━━━━━━━━━━━━━━━━━━━━ 1s 9ms/step
Root Mean Squared Error (RMSE) for COTY: 0.4803
Evaluating model for ELF...
98/98 ━━━━━━━━━━━━━━━━━━━━ 1s 7ms/step
Root Mean Squared Error (RMSE) for ELF: 3.3409
```

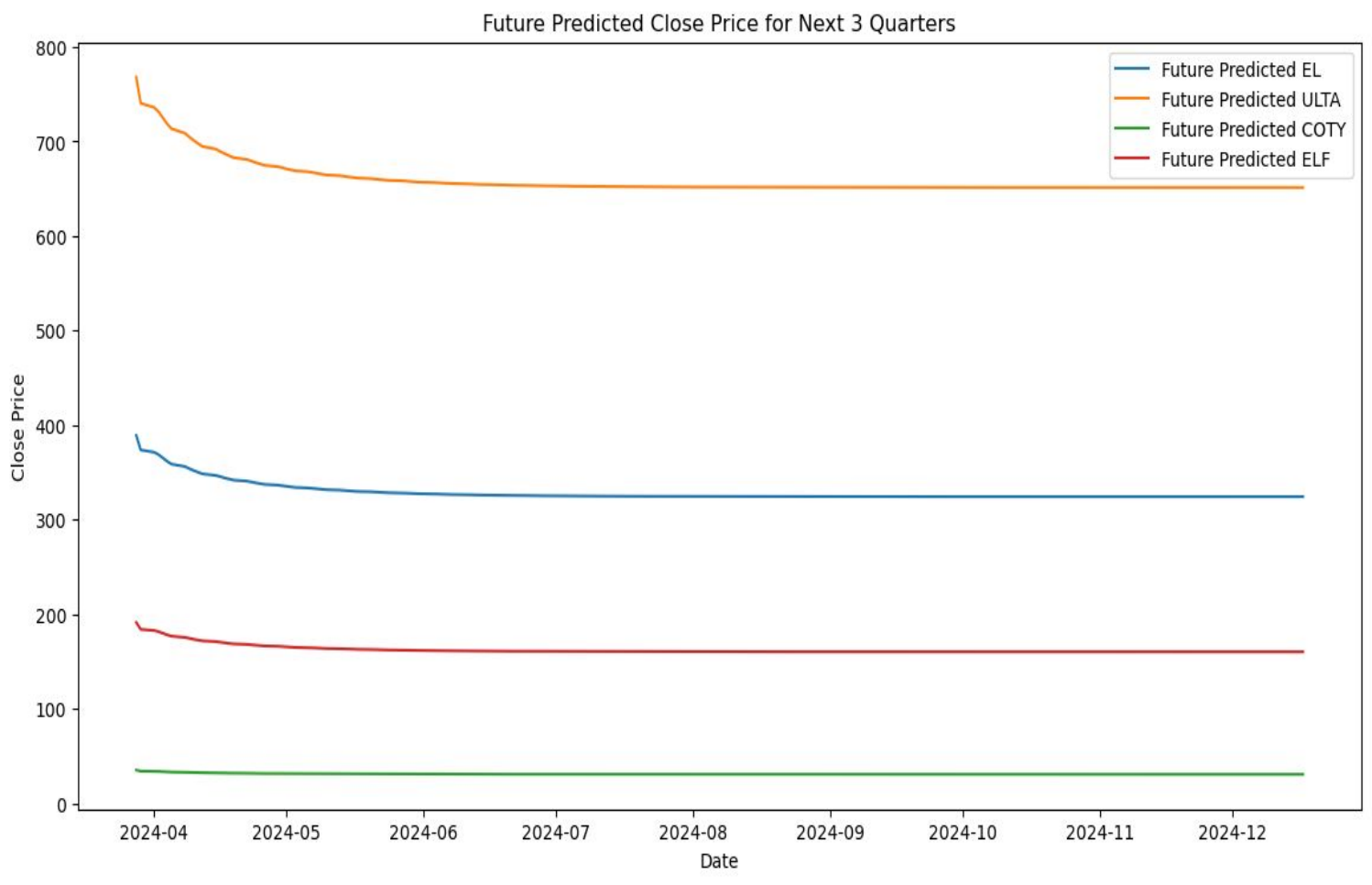

Actual vs Predicted Close Price for COTY
Close Price
Date
Actual
Predicted


Actual vs Predicted Close Price for ELF
Close Price
Date
Actual
Predicted


Actual vs Predicted Close Price for ULTA
Close Price
Date
Actual
Predicted


Actual vs Predicted Close Price for EL
Close Price
Date
Actual
Predicted


Future
Stock
Predictions
Adobe Stock | #180008799


Future Predicted Close Price for Next 3 Quarters
Future Predicted EL
Future Predicted ULTA
Future Predicted COTY
Future Predicted ELF
Close Price
800
700
600
500
400
300
200
100
0
2024-04
2024-05
2024-06
2024-07
2024-08
2024-09
2024-10
2024-11
2024-12
Date

Assumptions, Limitations and Disclaimers

- **Market Conditions:** The model assumes that future market conditions will follow similar patterns to historical data. This includes trends, volatility, and economic conditions.
- **Model Complexity:** While LSTM models can capture complex patterns, they also require significant computational resources and time for training, which may not be feasible for all users.
- **Short-Term Predictions:** The model is more suitable for short- to medium-term predictions due to the inherent unpredictability of the stock market over long periods.
- **External Factors:** The model does not account for sudden market shocks, geopolitical events, or changes in regulatory policies that can significantly impact stock prices.
- **No Financial Advice:** The predictions generated by the model are for informational purposes only and should not be construed as financial advice. Users should perform their own due diligence before making any investment decisions.
- **Performance Guarantees:** There is no guarantee that the model's predictions will be accurate. The stock market is influenced by numerous unpredictable factors that the model cannot account for.
- **Regular Updates Required:** The model's accuracy may degrade over time if it is not regularly updated with new data and retrained to account for recent market conditions.

More Ideas to Improve the Model in Future

- **Incorporate More Data Sources**
  - **Economic Indicators:**
    i. Integrate macroeconomic indicators such as GDP growth rates, unemployment rates, inflation rates, and interest rates to capture broader economic trends that affect stock prices.
  - **Sentiment Analysis:**
    i. Use natural language processing (NLP) to analyze news articles, social media posts, and financial reports to gauge market sentiment and its impact on stock prices.
- **Regular Updates and Retraining**
  - **Continuous Learning:**
    i. Implement a system for continuous learning where the model is regularly updated with new data to adapt to changing market conditions.
  - **Retraining Schedule:**
    i. Set up a retraining schedule to periodically retrain the model with the latest data, ensuring that it remains accurate and relevant.

Conclusions

- The LSTM model demonstrated strong predictive accuracy for the stock prices of EL, ULTA, COTY, and ELF, closely aligning predicted values with actual historical prices.
- The model maintained its accuracy even during periods of high volatility, indicating its robustness in handling fluctuating market conditions. This makes it a reliable tool for short- to medium-term stock price predictions.
- The model can be enhanced with the incorporation of new ideas in the future.