

# Predicting Future Stock Prices

Heather Adler

Data Science Intensive  
Capstone Project



## Presentation Purpose

- Proven Skincare is a personalized skin-care brand that is planning to go public through a Reg A+ mini-IPO. This offering allows private companies with less than \$1 billion in earnings to go public with lower accounting and disclosure standards compared to traditional IPOs. Proven Skincare aims to raise approximately \$42 million and has plans to expand into new markets, introduce new products, and increase its marketing efforts

## Problem Statement

- How can predictive analytics drive Proven Skincare's success in the stock market by understanding leading beauty and wellness stock performance?

## Presentation Goals

- Provide strategic insights and recommendations for Proven Skincare as it transitions to a publicly traded company.

## Challenges

- Stock Market Volatility
- Lack of Accurate Prediction Tools
- Complexity of Data

# Data Details

The dataset for this project was downloaded from Kaggle and has been filtered and cleaned to only include the following 4 beauty and wellness stock data:

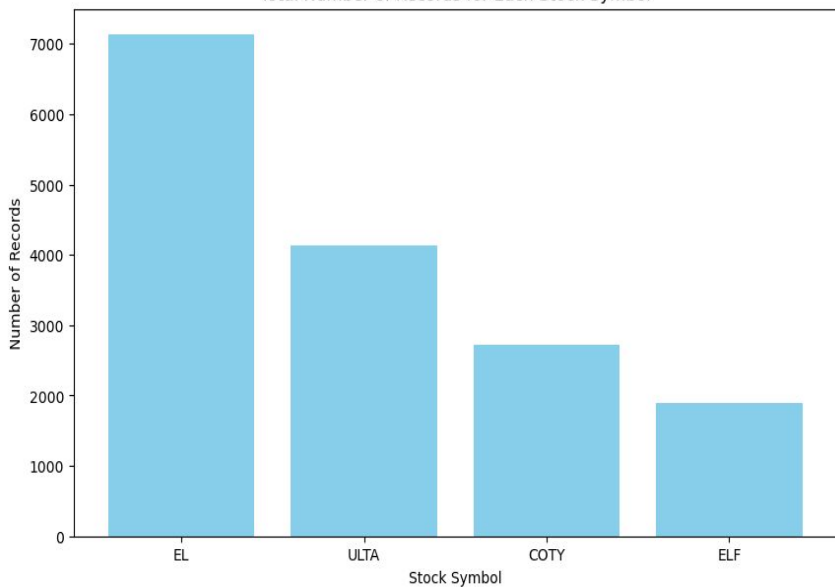
1) The Estée Lauder Companies Inc. (EL)

2) Ulta Beauty, Inc. (ULTA)

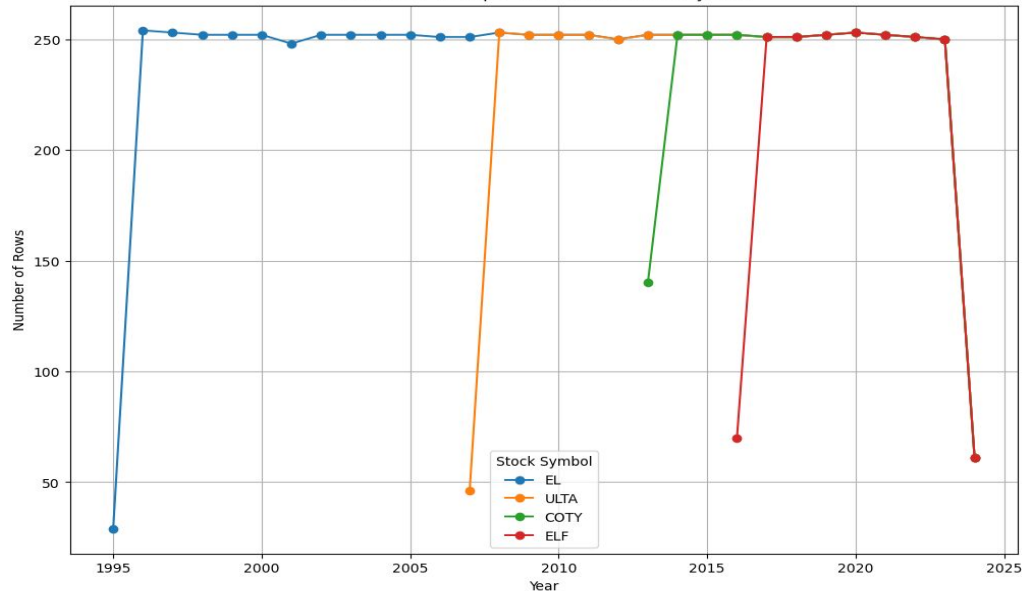
3) COTY (COTY)

4) e.l.f. Beauty, Inc. (ELF)

Total Number of Records for Each Stock Symbol



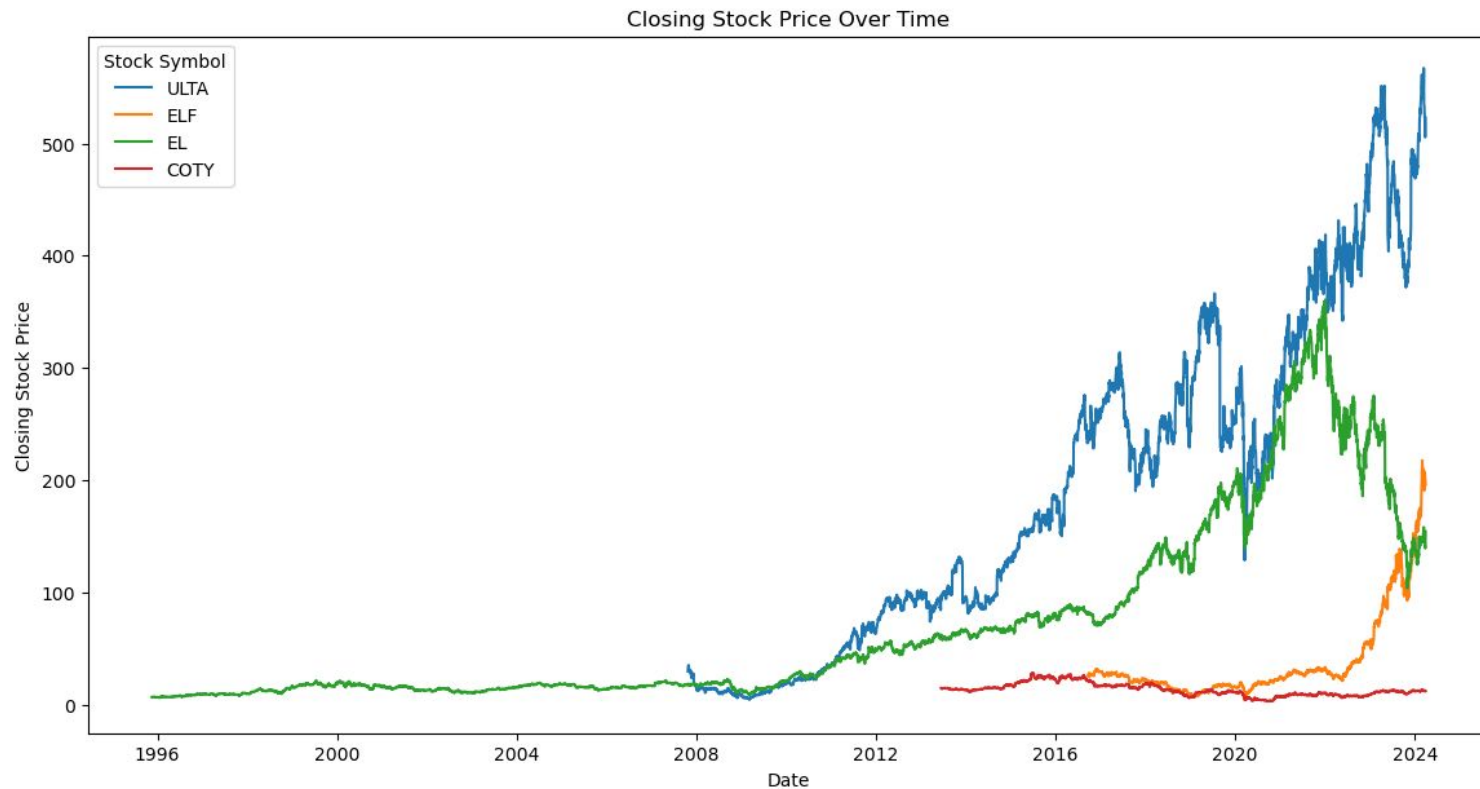
Number of Rows per Year for Each Stock Symbol



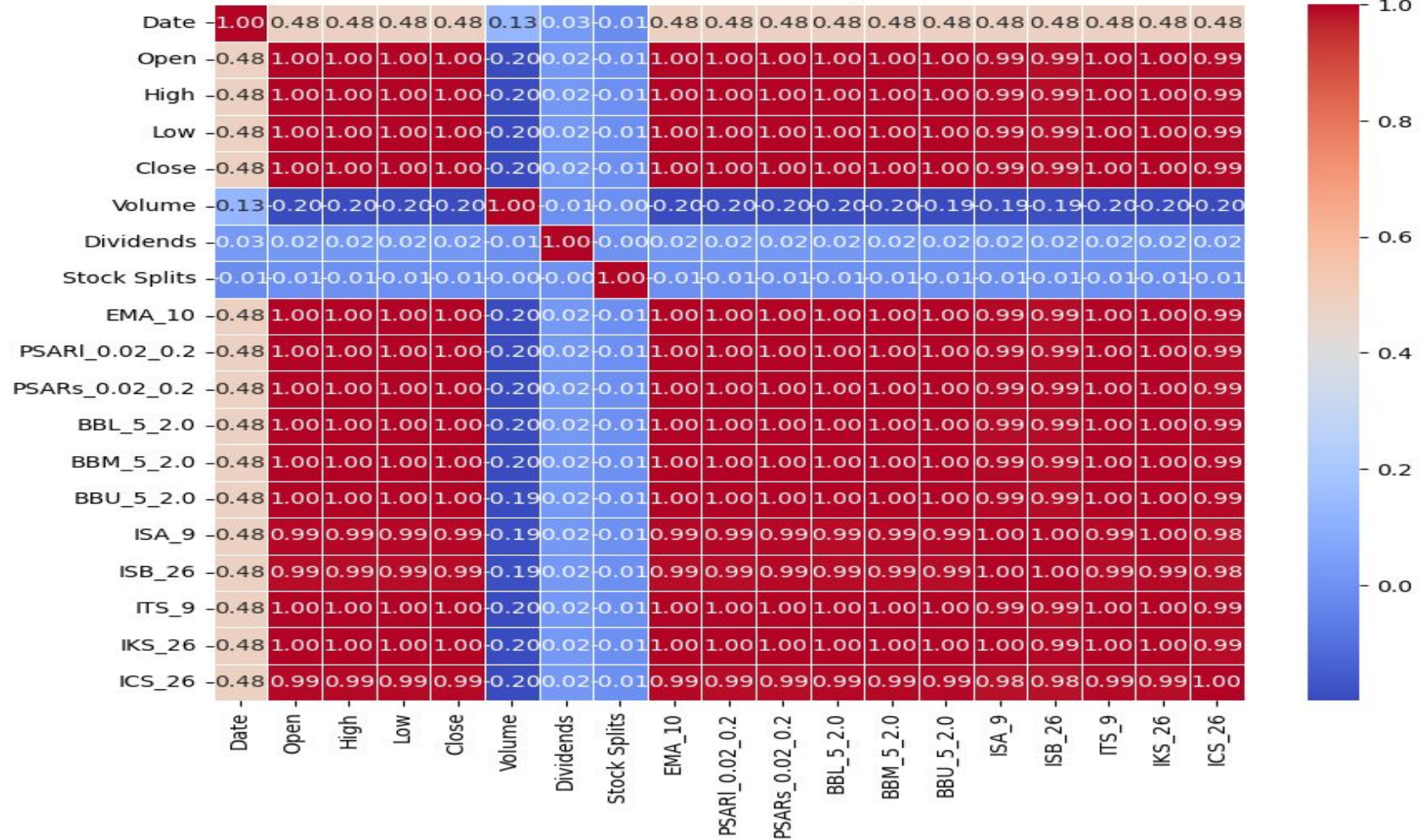


# Data Exploration

# Closing Stock Price Over Time



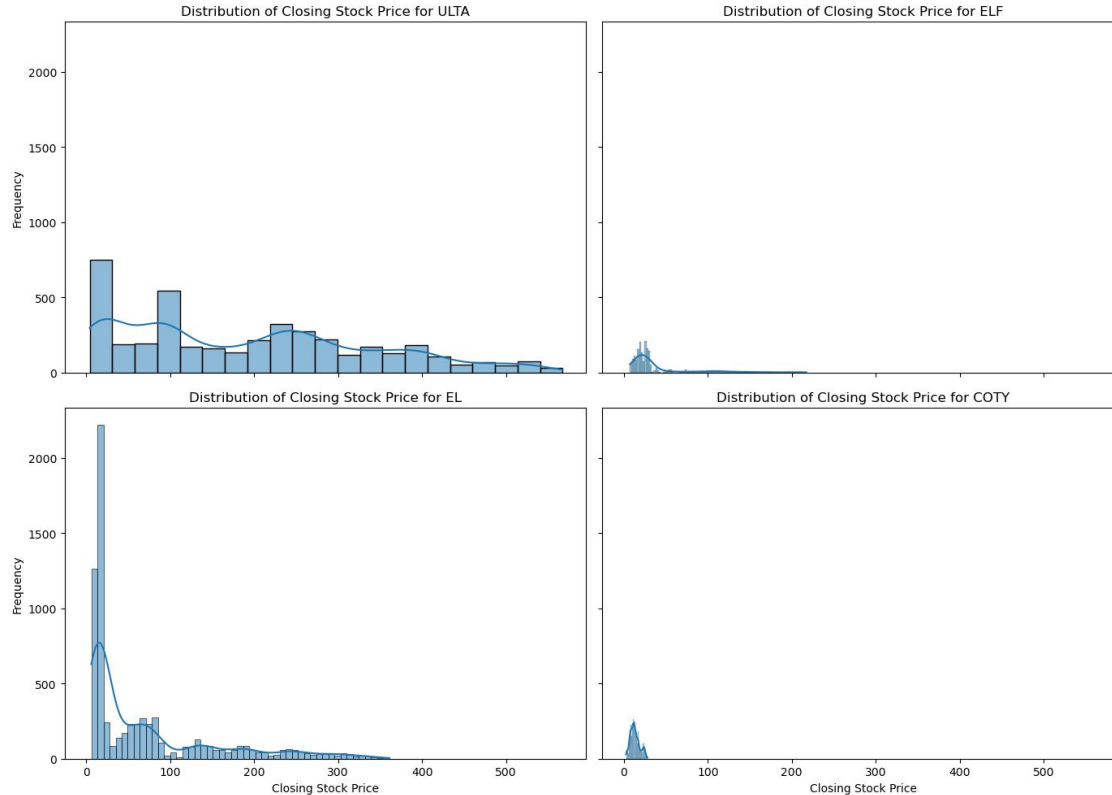
Correlation Heatmap



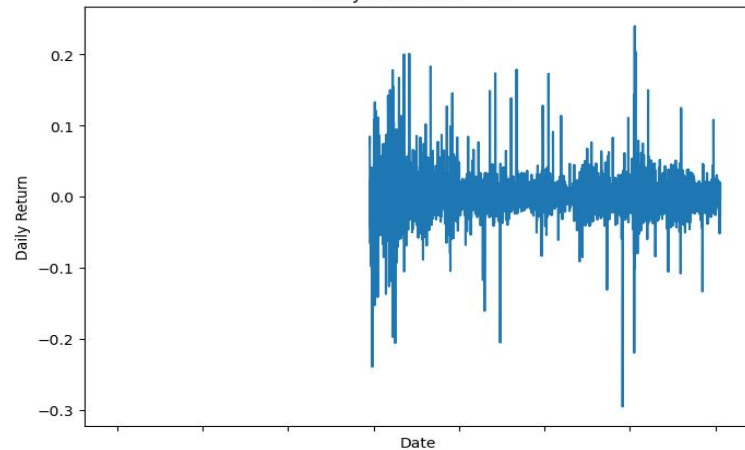


# Visualization of the Target Variable (Close Price)

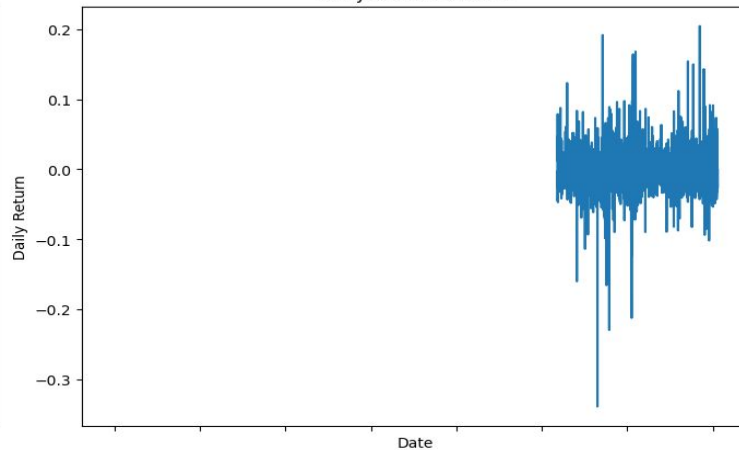
The Target variable in this case is the closing stock price. The below charts show the frequency for which the closing stock price occurred and shows that ULTA and EL had the highest closing stock prices at a higher frequency. This suggests that these companies may have experienced relatively strong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



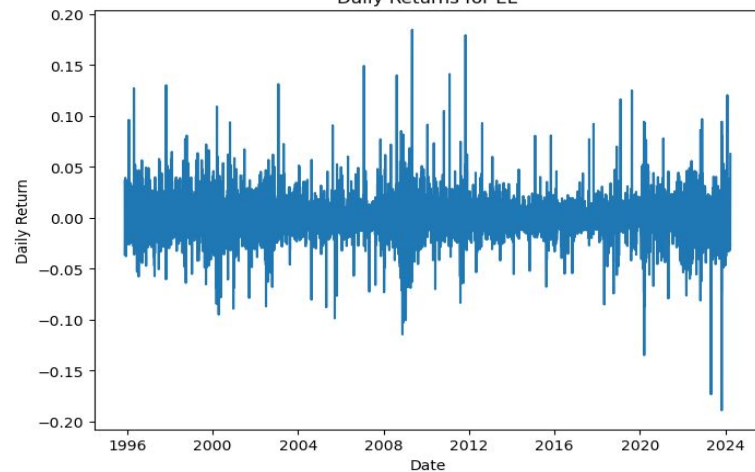
Daily Returns for ULTA



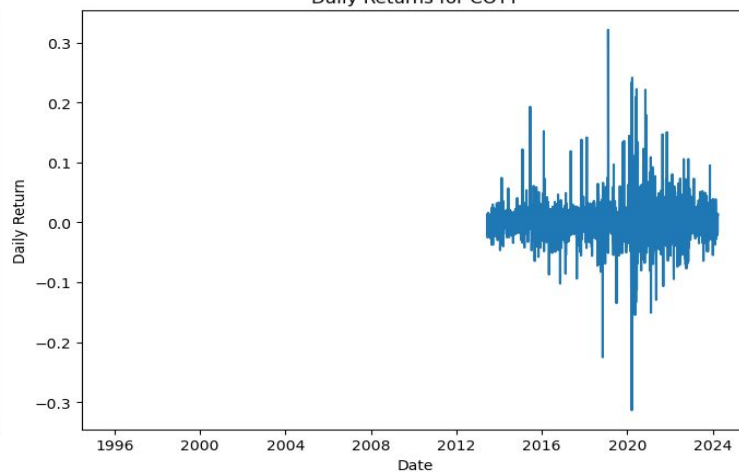
Daily Returns for ELF



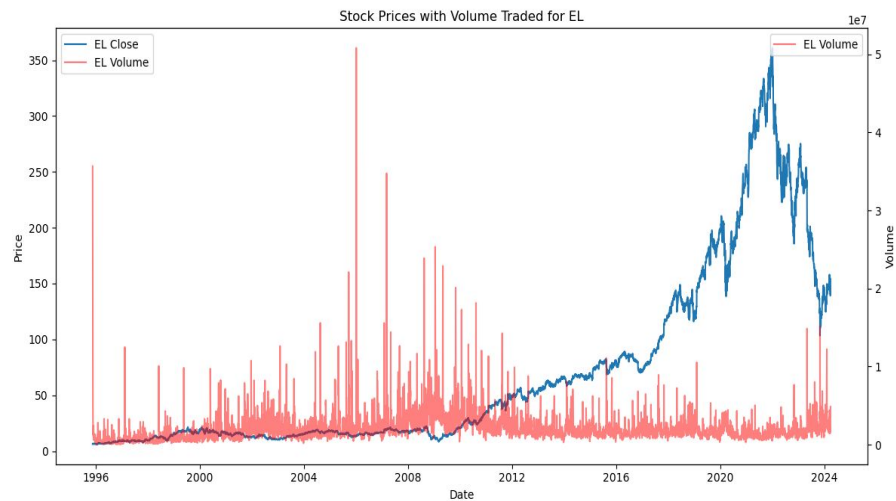
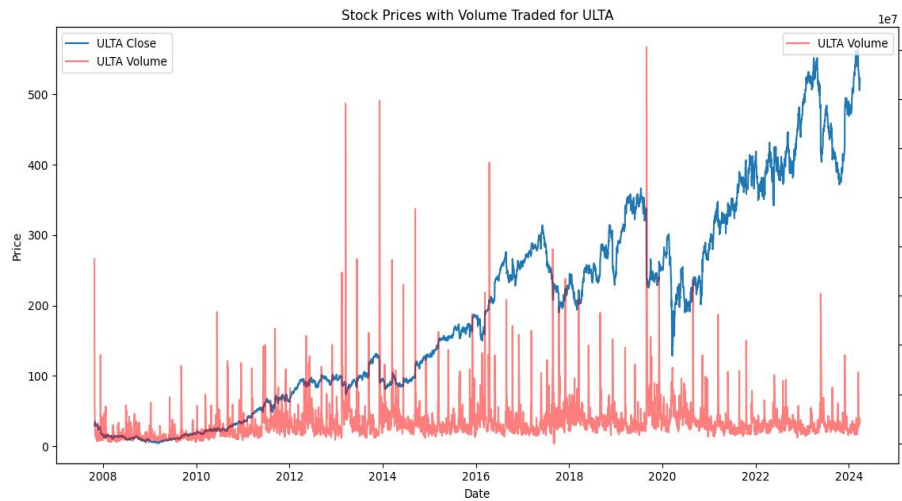
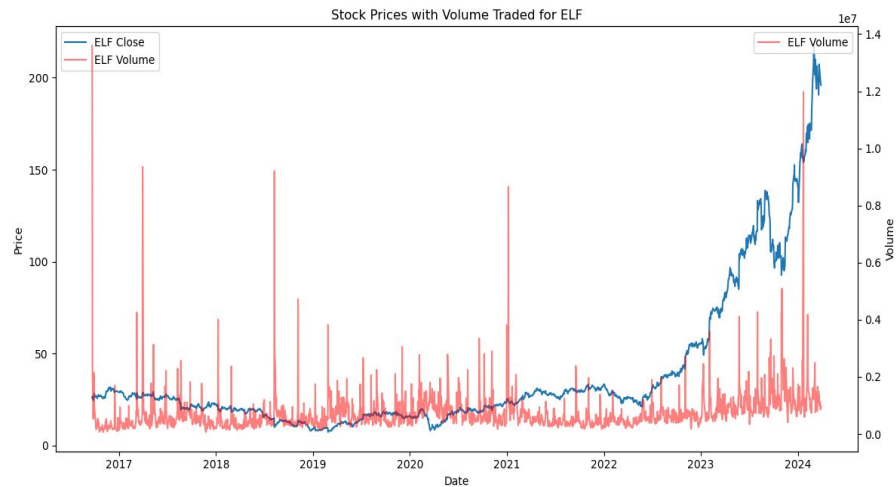
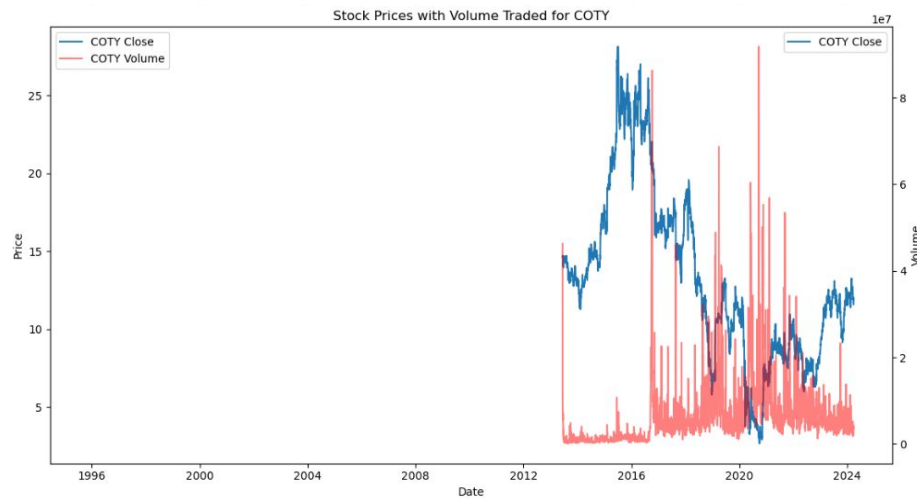
Daily Returns for EL



Daily Returns for COTY







# Data Modeling

# 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

# Reasoning for RMSE Metric for Evaluation

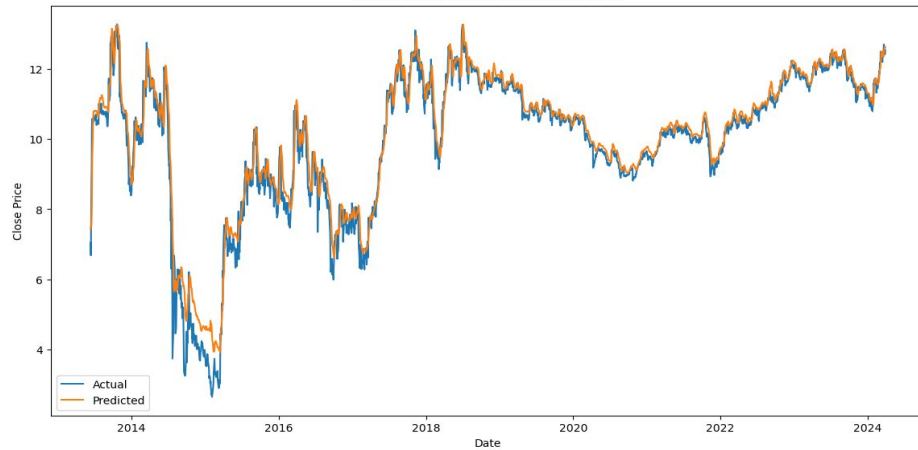
- **RMSE was chosen for model evaluation due to its advantages in interpretability, sensitivity to large errors, and suitability for optimization processes.**
- **Given the context of stock price prediction, where large deviations from actual values can significantly impact investment decisions, RMSE's penalization of large errors is particularly beneficial.**
- **Additionally, the need for a standard and widely understood metric facilitates easier comparison with other models and studies in the financial domain.**

## Comparison with Other Metric

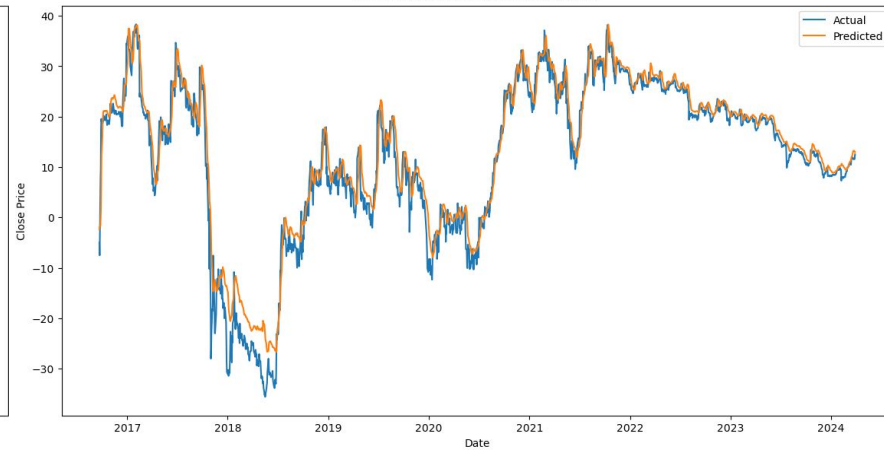
- **Mean Absolute Error (MAE):** MAE is another popular metric that calculates the average absolute differences between predicted and actual values. While MAE is less sensitive to outliers compared to RMSE, it does not penalize large errors as strongly. This can be both an advantage and a disadvantage depending on the specific application.
- **Mean Squared Error (MSE):** MSE is similar to RMSE but without taking the square root. While it also penalizes larger errors more heavily, its units are the square of the target variable, making it less interpretable compared to RMSE.
- **R-squared ( $R^2$ ):**  $R^2$  measures the proportion of variance in the target variable that is explained by the model. While useful for understanding the goodness of fit, it does not provide an intuitive measure of prediction error.



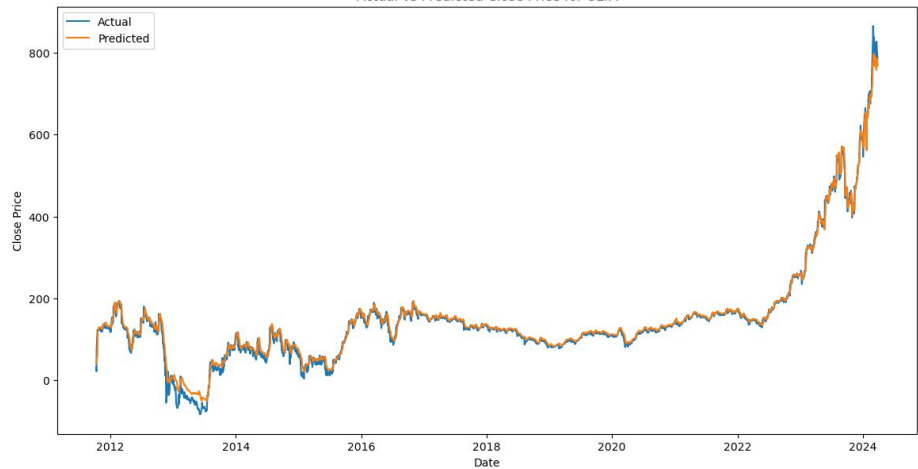
Actual vs Predicted Close Price for COTY



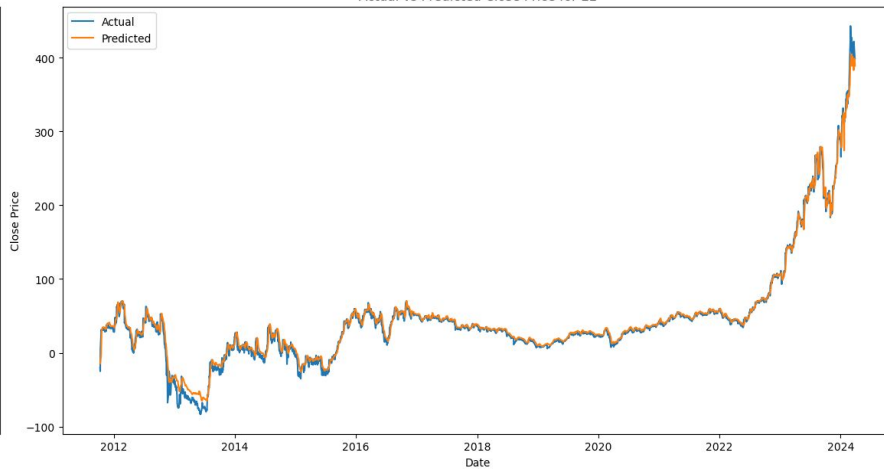
Actual vs Predicted Close Price for ELF



Actual vs Predicted Close Price for ULTA

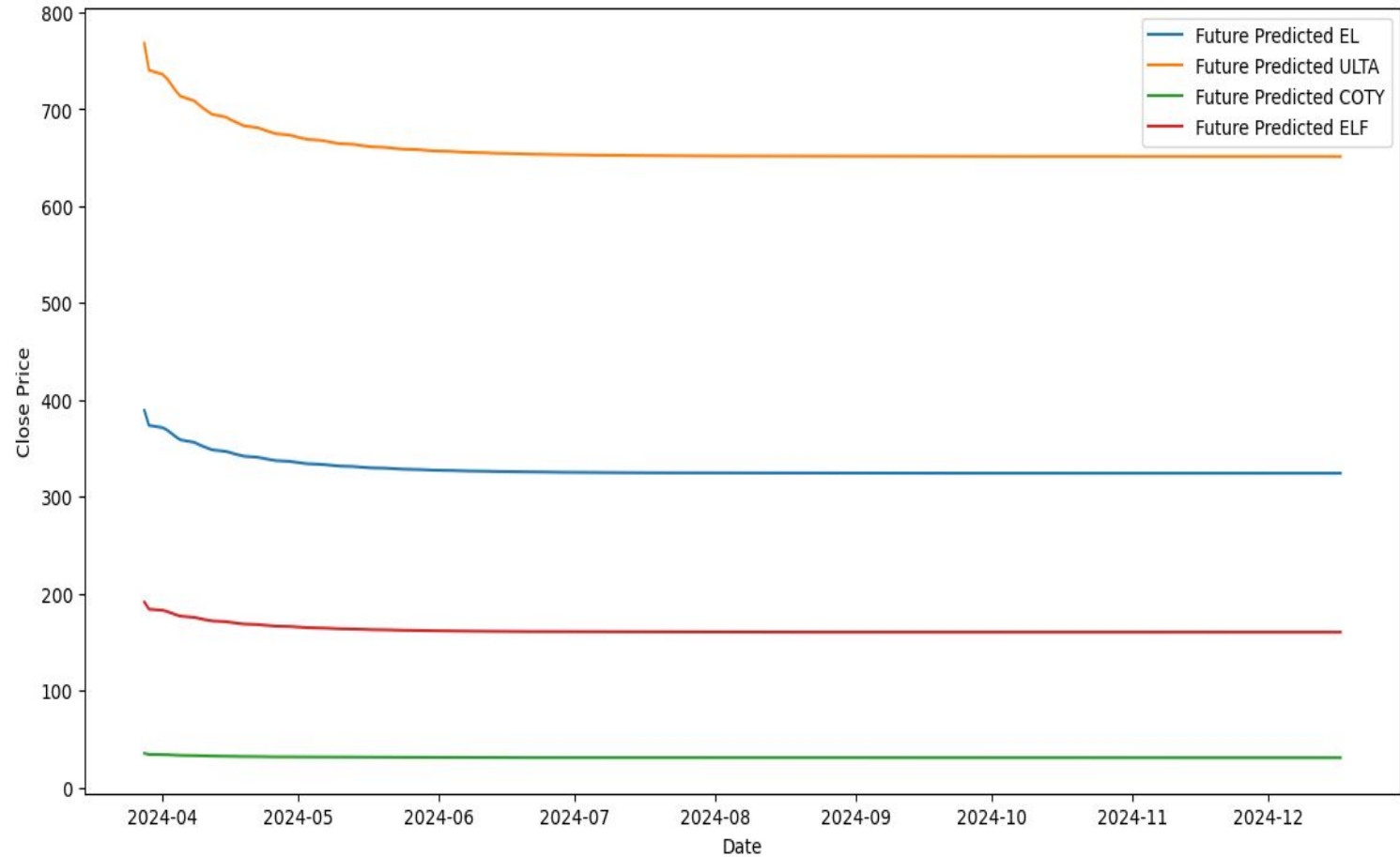


Actual vs Predicted Close Price for EL



# Future Stock Prediction

Future Predicted Close Price for Next 3 Quarters



# 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.



# 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.

# Appendix A: 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.