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

Presentation Goals

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

Problem Statement

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

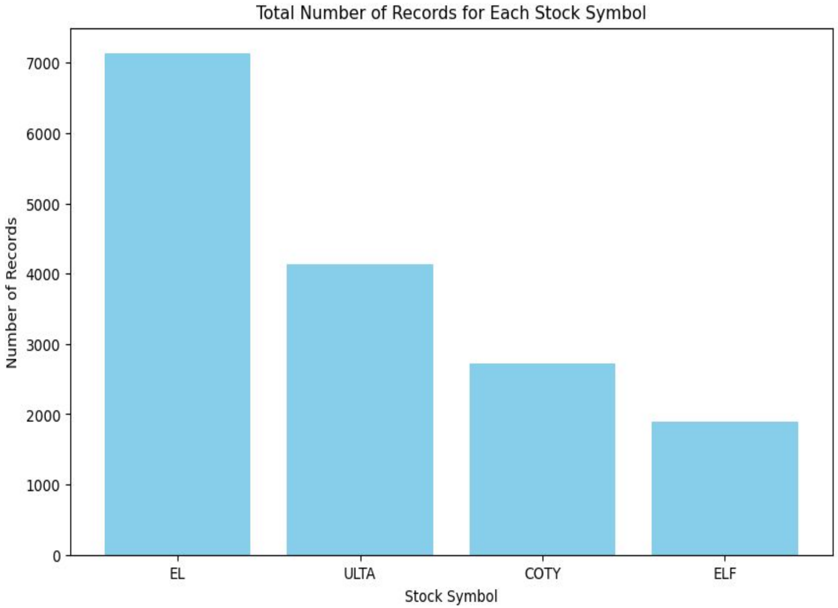
Challenges

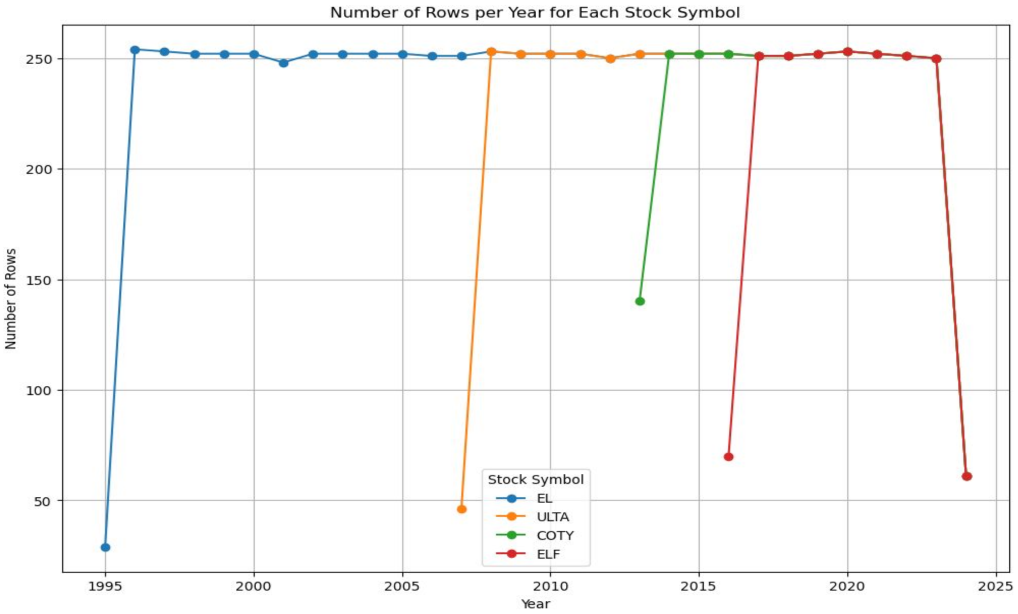
- Stock Market Volatility
- Lack of Accurate Prediction Tools
- Complexity of Data

Data Details

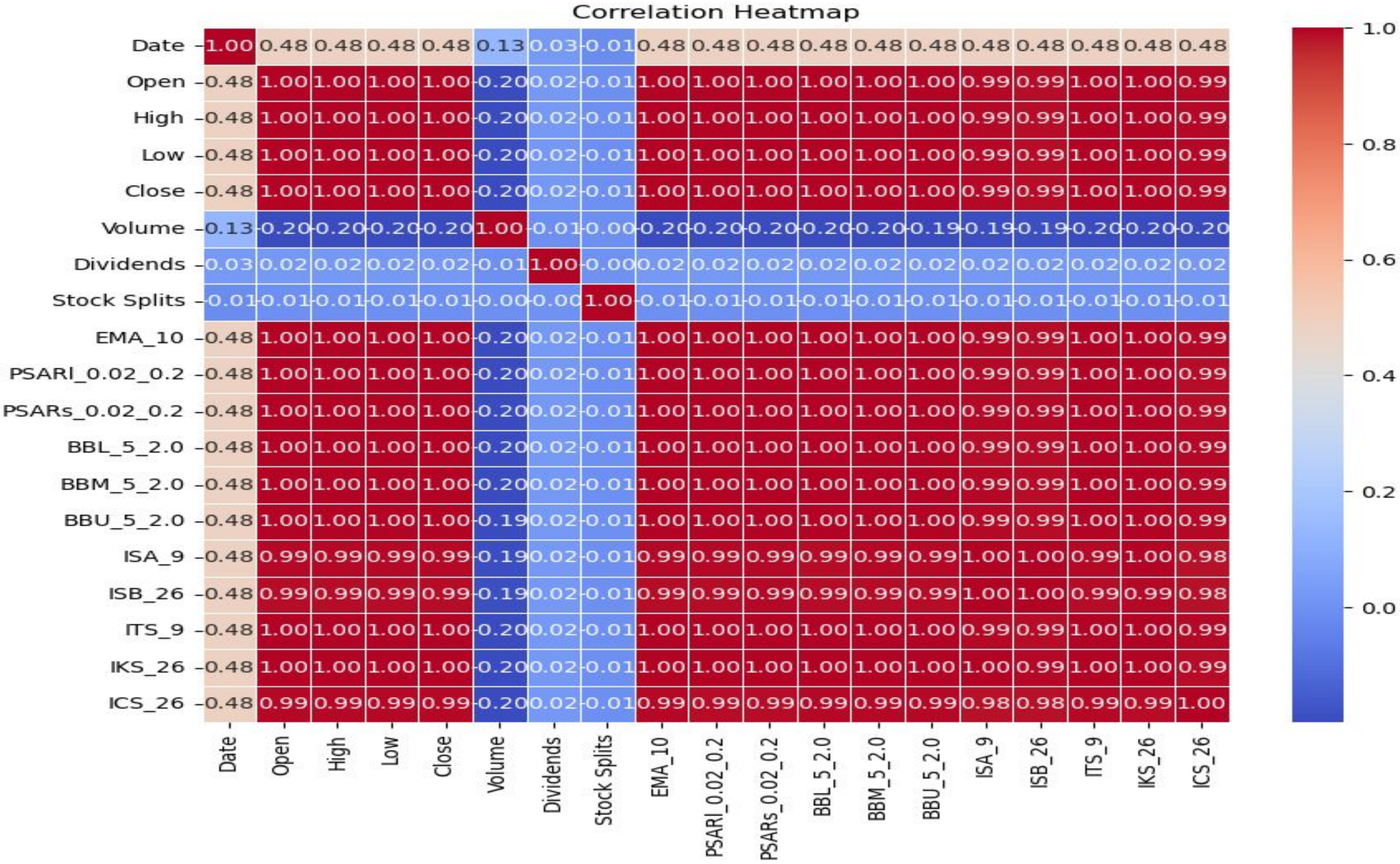
The dataset for this project was downloaded from Kaggle and has been filtered and cleaned to only include the following 4 beauty and wellness stock data:

1) The Estée Lauder Companies Inc. (EL)
2) Ulta Beauty, Inc. (ULTA)
3) COTY (COTY)
4) e.l.f. Beauty, Inc. (ELF)



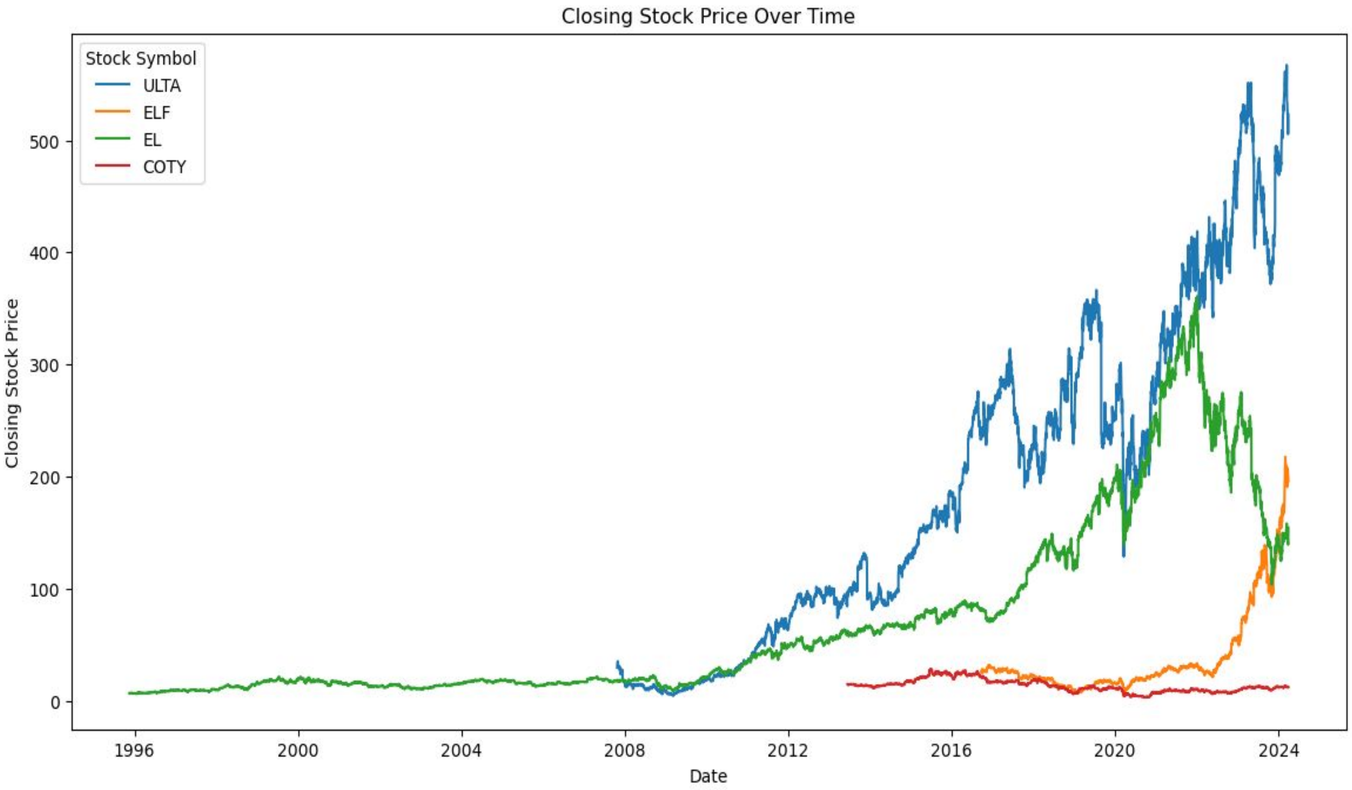


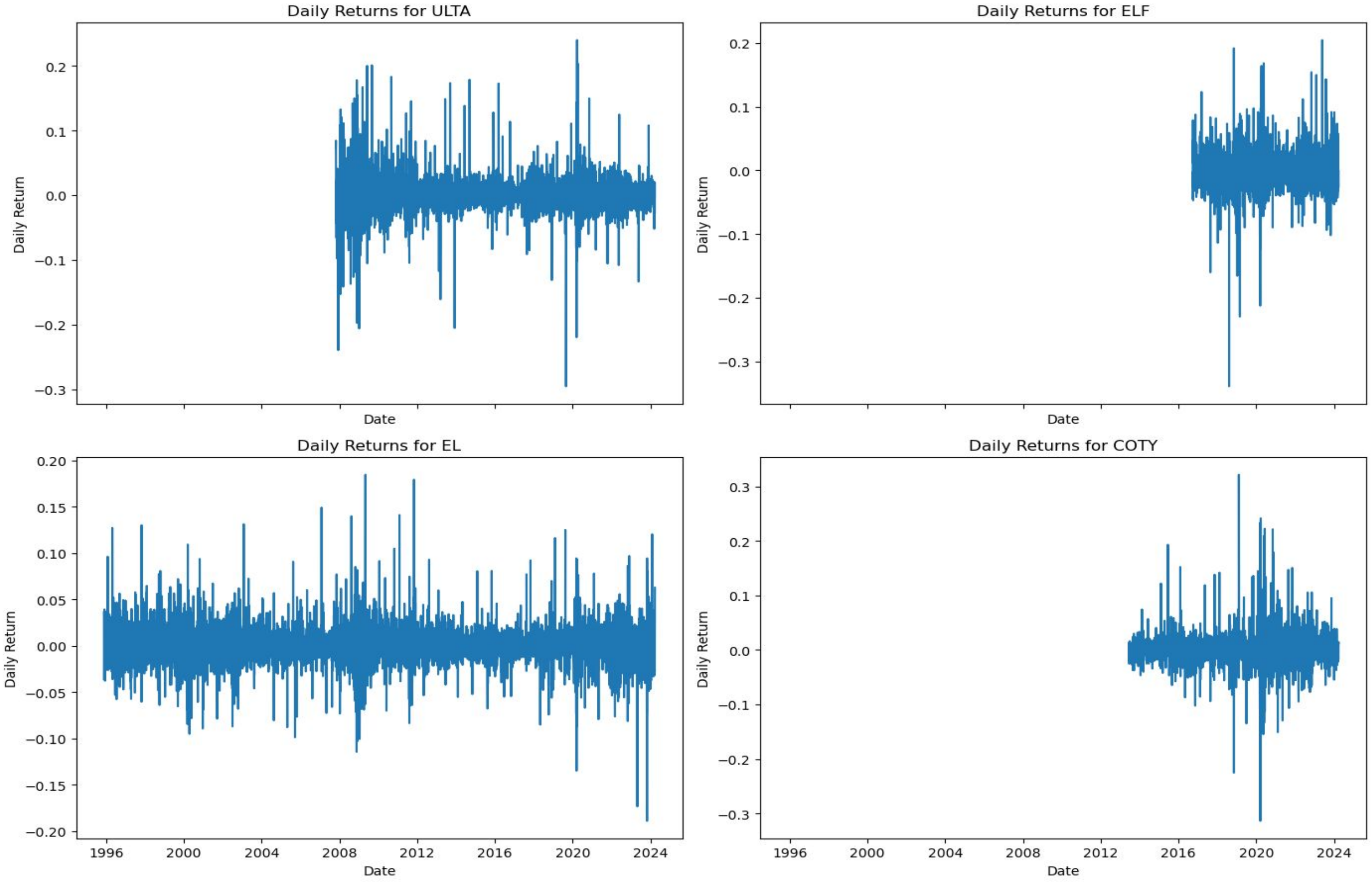
Data Exploration

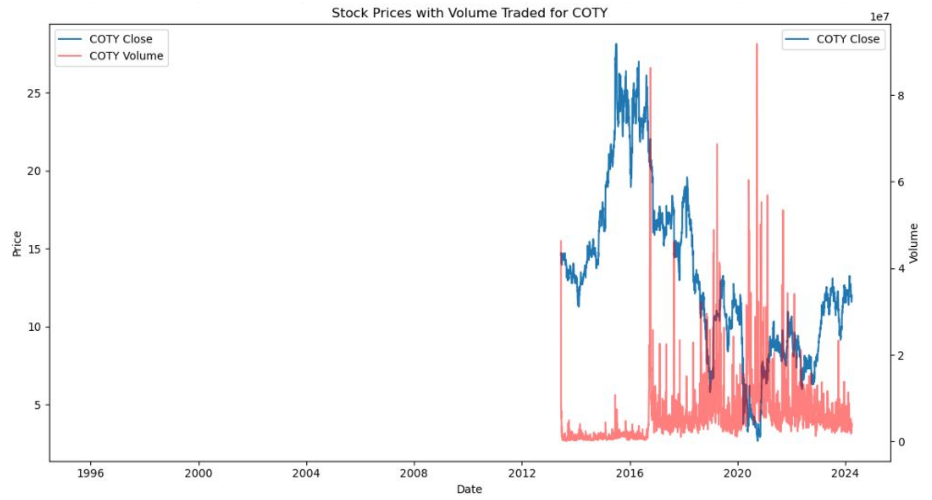

Correlation Heatmap
Date 1.00 0.48 0.48 0.48 0.48 0.13 0.03 -0.01 0.48 0.48 0.48 0.48 0.48 0.48 0.48 0.48 0.48 0.48 0.48
Open 0.48 1.00 1.00 1.00 1.00 -0.20 0.02 -0.01 1.00 1.00 1.00 1.00 1.00 1.00 0.99 0.99 1.00 1.00 0.99
High 0.48 1.00 1.00 1.00 1.00 -0.20 0.02 -0.01 1.00 1.00 1.00 1.00 1.00 1.00 0.99 0.99 1.00 1.00 0.99
Low 0.48 1.00 1.00 1.00 1.00 -0.20 0.02 -0.01 1.00 1.00 1.00 1.00 1.00 1.00 0.99 0.99 1.00 1.00 0.99
Close 0.48 1.00 1.00 1.00 1.00 -0.20 0.02 -0.01 1.00 1.00 1.00 1.00 1.00 1.00 0.99 0.99 1.00 1.00 0.99
Volume 0.13 -0.20 -0.20 -0.20 -0.20 1.00 -0.01 -0.00 -0.20 -0.20 -0.20 -0.20 -0.20 -0.19 -0.19 -0.19 -0.20 -0.20 -0.20
Dividends 0.03 0.02 0.02 0.02 0.02 -0.01 1.00 -0.00 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02
Stock Splits -0.01 -0.01 -0.01 -0.01 -0.01 -0.00 -0.00 1.00 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01
EMA_10 0.48 1.00 1.00 1.00 1.00 -0.20 0.02 -0.01 1.00 1.00 1.00 1.00 1.00 1.00 0.99 0.99 1.00 1.00 0.99
PSARl_0.02_0.2 0.48 1.00 1.00 1.00 1.00 -0.20 0.02 -0.01 1.00 1.00 1.00 1.00 1.00 1.00 0.99 0.99 1.00 1.00 0.99
PSARs_0.02_0.2 0.48 1.00 1.00 1.00 1.00 -0.20 0.02 -0.01 1.00 1.00 1.00 1.00 1.00 1.00 0.99 0.99 1.00 1.00 0.99
BBL_5_2.0 0.48 1.00 1.00 1.00 1.00 -0.20 0.02 -0.01 1.00 1.00 1.00 1.00 1.00 1.00 0.99 0.99 1.00 1.00 0.99
BBM_5_2.0 0.48 1.00 1.00 1.00 1.00 -0.20 0.02 -0.01 1.00 1.00 1.00 1.00 1.00 1.00 0.99 0.99 1.00 1.00 0.99
BBU_5_2.0 0.48 1.00 1.00 1.00 1.00 -0.19 0.02 -0.01 1.00 1.00 1.00 1.00 1.00 1.00 0.99 0.99 1.00 1.00 0.99
ISA_9 0.48 0.99 0.99 0.99 0.99 -0.19 0.02 -0.01 0.99 0.99 0.99 0.99 0.99 0.99 1.00 1.00 0.99 1.00 0.98
ISB_26 0.48 0.99 0.99 0.99 0.99 -0.19 0.02 -0.01 0.99 0.99 0.99 0.99 0.99 0.99 1.00 1.00 0.99 0.99 0.98
ITS_9 0.48 1.00 1.00 1.00 1.00 -0.20 0.02 -0.01 1.00 1.00 1.00 1.00 1.00 1.00 0.99 0.99 1.00 1.00 0.99
IKS_26 0.48 1.00 1.00 1.00 1.00 -0.20 0.02 -0.01 1.00 1.00 1.00 1.00 1.00 1.00 1.00 0.99 1.00 1.00 0.99
ICS_26 0.48 0.99 0.99 0.99 0.99 -0.20 0.02 -0.01 0.99 0.99 0.99 0.99 0.99 0.99 0.98 0.98 0.99 0.99 1.00
Date
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
1.0
0.8
0.6
0.4
0.2
0.0

Visualization of the Target Variable (Close Price)

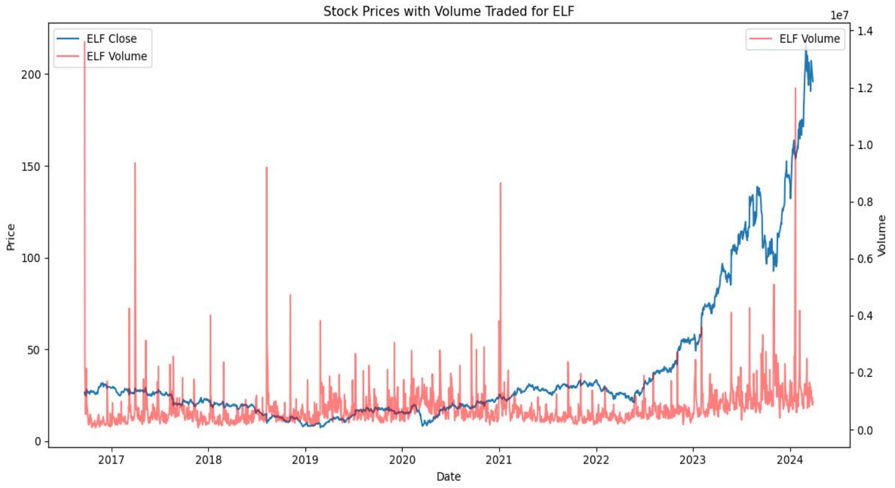
The Target variable in this case is the closing stock price. This charts shows the change in closing price over since the start of each stock symbol. ULTA and EL are strong growth stocks with long-term gains while ELF is an emerging stock with rapid recent growth. COTY shows the most volatility and has not recovered to its previous highs.



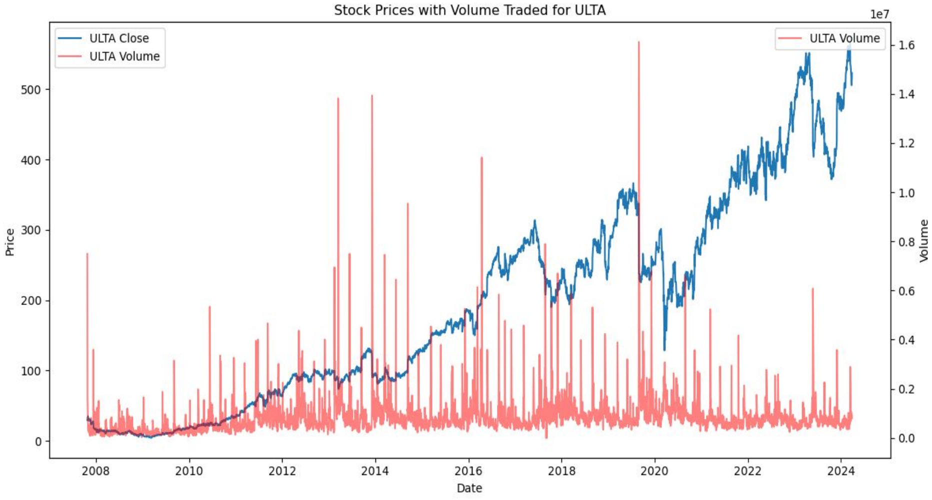

Daily Returns for ULTA
Daily Returns for ELF
Daily Returns for EL
Daily Returns for COTY
Daily Return
Date
1996
2000
2004
2008
2012
2016
2020
2024



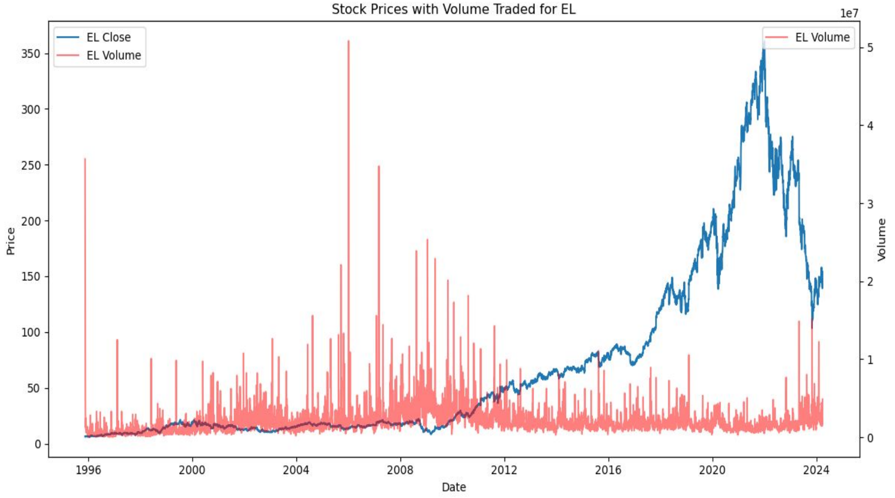
Stock Prices with Volume Traded for COTY
COTY Close
COTY Volume
COTY Close
1e7
Price
Volume
Date



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



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



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

Data Modeling

Defining & Training the Model

- **Model Used:** Long Short-Term Memory (LSTM) model
- LSTM is a type of recurrent neural network (RNN) designed to handle long-term dependencies
- LSTM was selected for the model because is particularly effective for time series prediction, which is crucial in stock market forecasting due to its ability to remember important information over extended periods.

LSTM Pros	LSTM Cons
• **Captures Long-Term Dependencies**: LSTMs are designed to overcome the limitations of traditional RNNs by capturing long-term dependencies in sequential data. This is crucial for stock price prediction, where past prices can influence future prices over long periods. • **Handles Non-linearities**: Stock prices often exhibit non-linear patterns. LSTMs can capture these complex relationships more effectively than linear models. • **Memory of Previous States**: The architecture of LSTMs allows them to remember important information over longer periods, making them more effective at understanding trends and patterns in historical stock price data. • **Handles Variable Length Sequences**: LSTMs can process sequences of varying lengths, making them adaptable to different time horizons and datasets without requiring fixed input sizes.	• **Computational Complexity**: LSTMs are computationally intensive and require significant processing power and memory, especially for large datasets. Training an LSTM model can be time-consuming and resource-intensive. • **Risk of Overfitting**: Due to their complexity, LSTMs can overfit the training data, especially if the dataset is small or not representative of future trends. Regularization techniques and careful tuning are necessary to mitigate this risk. • **Interpretability**: LSTM models internal workings are not easily interpretable. This can be a disadvantage when trying to explain model predictions to stakeholders or for understanding the underlying reasons for the predictions. • **Long Training Times**: Training LSTM models can take a long time, especially with large datasets and complex architectures. This can be a bottleneck in the model development process.

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
```

Interpretation of RMSE Results:

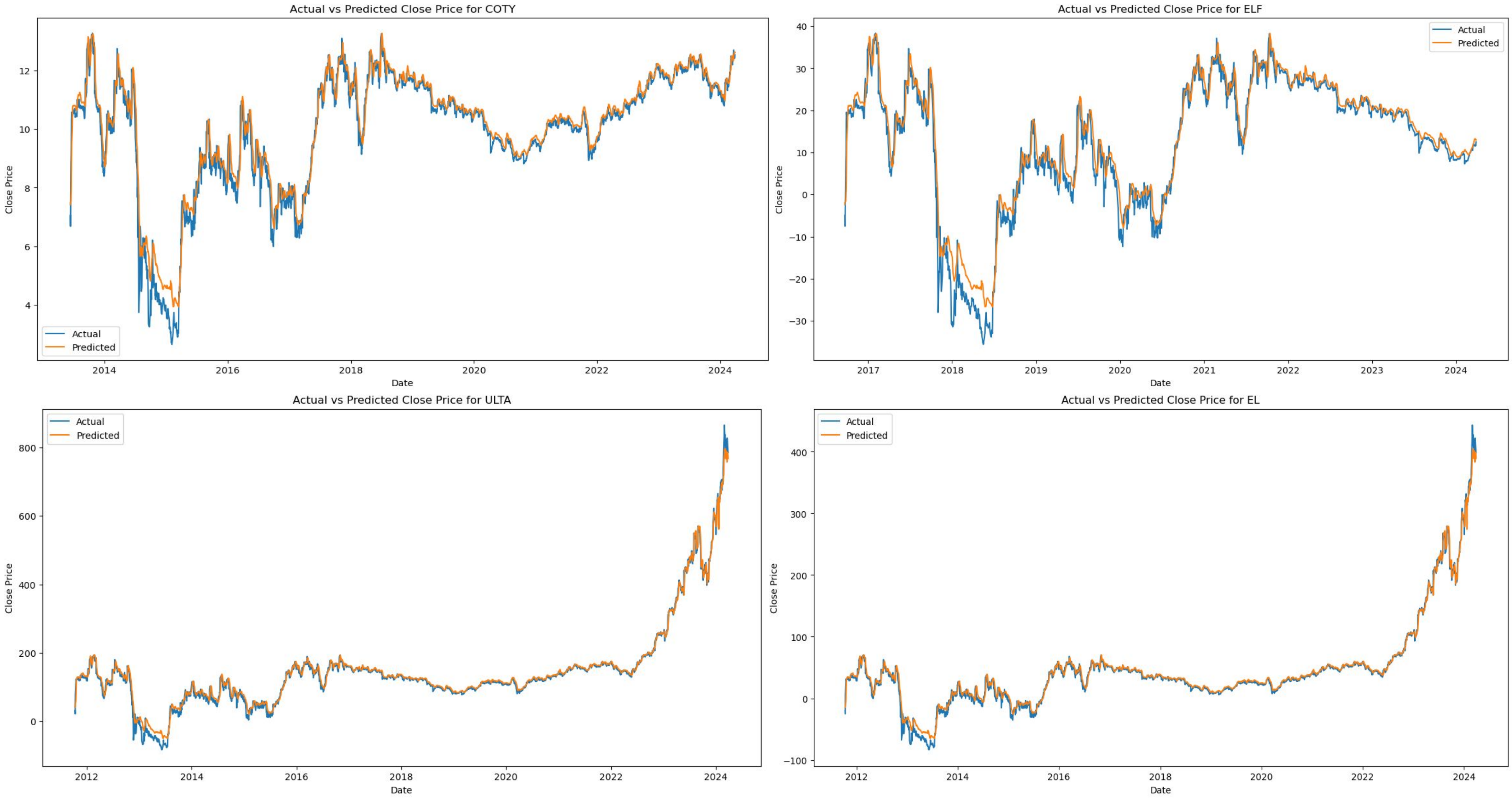
- **EL: RMSE**: 6.9571
    a. **Interpretation**: This shows moderately accurate results, meaning it is reasonably accurate but has some room for improvement.
- **ULTA: RMSE**: 12.5352
    a. **Interpretation**: This indicates the predictions are less accurate, suggesting that Ulta Beauty's stock prices are more challenging to predict accurately with our current model.
- **COTY: RMSE** .4803
    a. This means the predictions are very close to the actual prices, showing excellent accuracy.
- **ELF: RMSE** 3.3409
    a. **Interpretation**: This also shows moderately accurate results, meaning it is reasonably accurate but has some room for improvement.

Reasoning for RMSE Metric for Evaluation

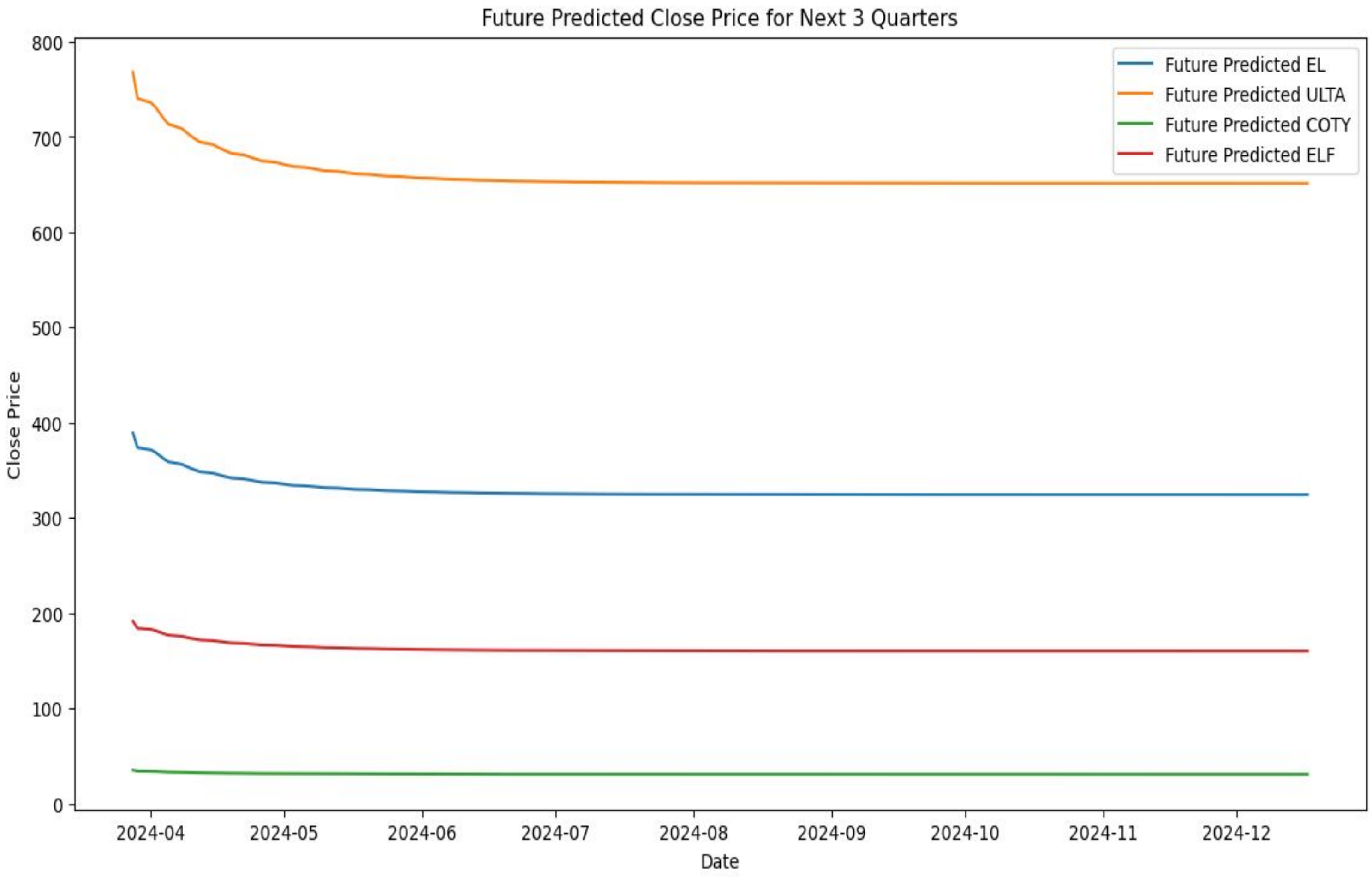
- RMSE was chosen for model evaluation due to its advantages in interpretability, sensitivity to large errors, and suitability for optimization processes.
- Given the context of stock price prediction, where large deviations from actual values can significantly impact investment decisions, RMSE's penalization 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
Actual
Predicted
Close Price
Date
2012
2014
2016
2017
2018
2019
2020
2021
2022
2023
2024

Future Stock Predictions


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

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.
- To enhance the model and validate predictions, conducting out-of-time sample testing in the future will be essential for ensuring the model's robustness and accuracy by testing it on data it has not been exposed to during training.
- The model can be enhanced with the incorporation of additional new ideas in the future.

More Ideas to Improve the Model in Future

- **Incorporate More Data Sources**
  - **Economic Indicators:**
    1. 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:**
    1. 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:**
    1. Implement a system for continuous learning where the model is regularly updated with new data to adapt to changing market conditions.
  - **Retraining Schedule:**
    1. 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.