

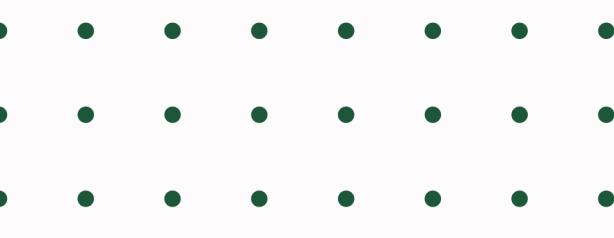
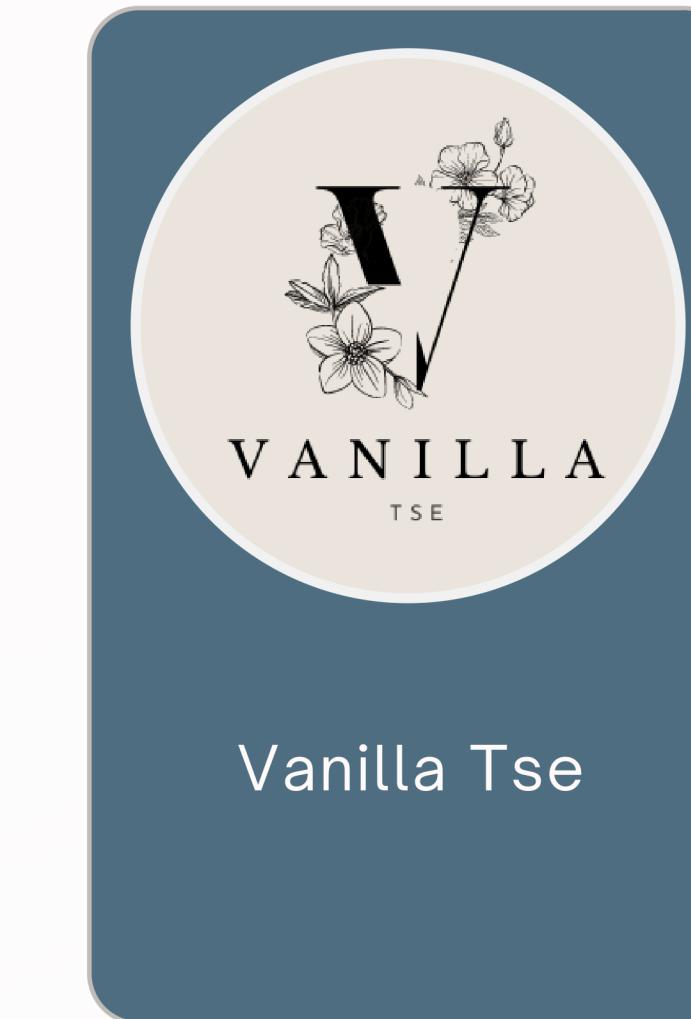
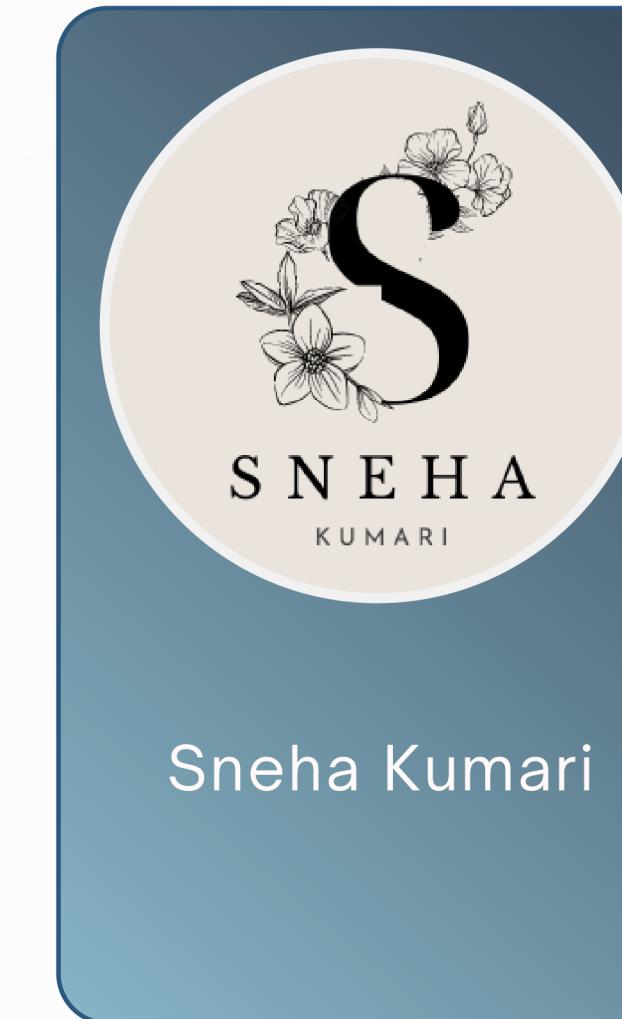
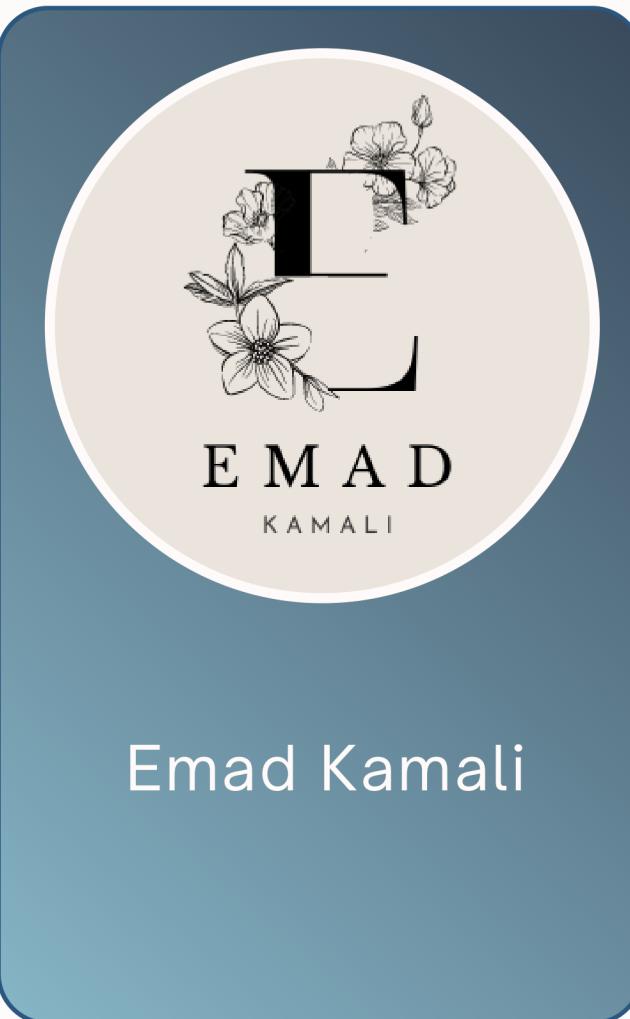


STOCK ANALYSIS

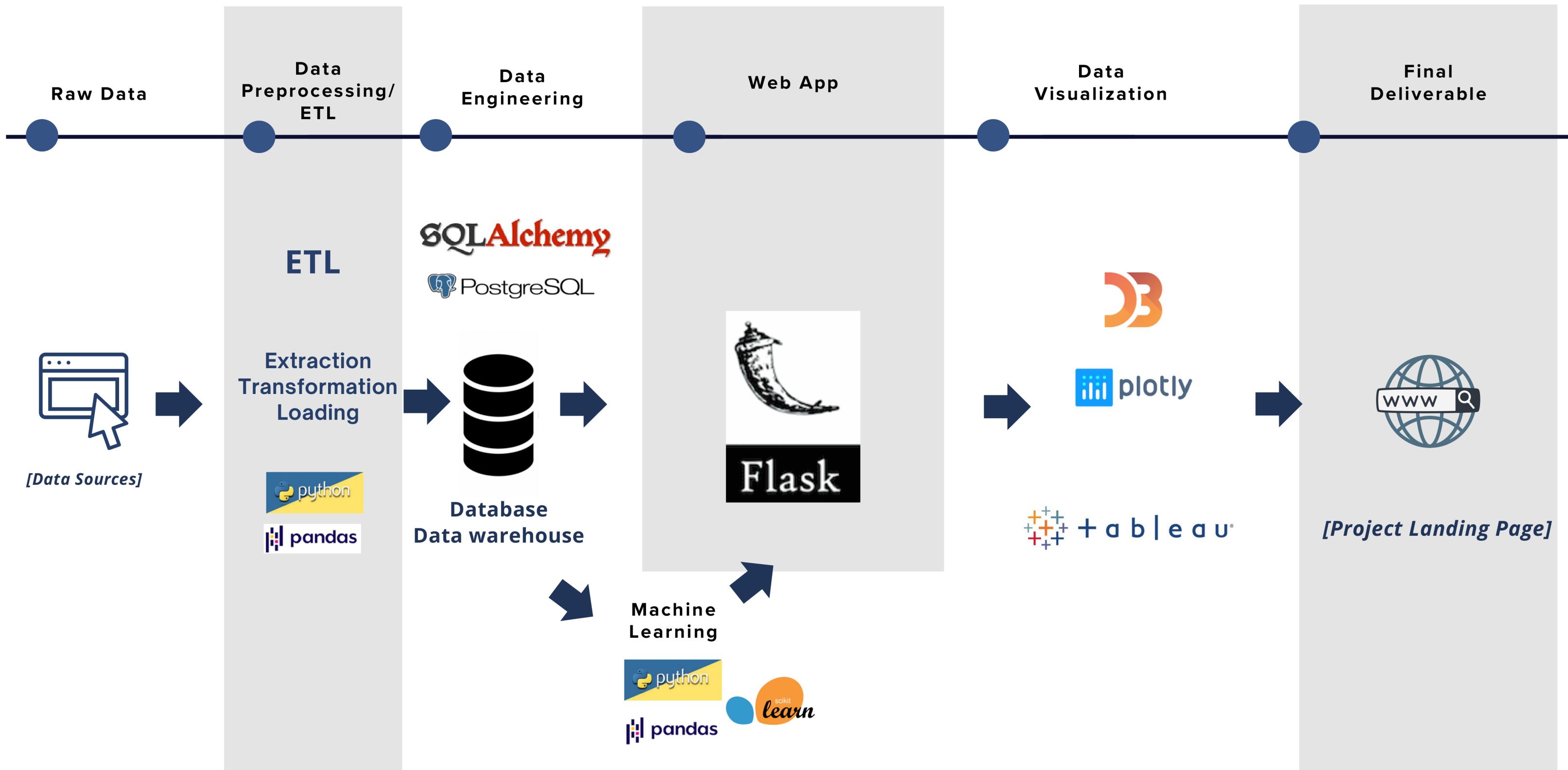
STOCK ANALYSIS AND PREDICTION

Comprehensive Project Analysis and Comparison of Models

OUR TEAM



DATA PIPELINE



GRADIENT BOOSTING

Machine Learning Model

Model Selection

- **Model Choice Rationale:**
 - Gradient Boosting excels with time series data.
 - Decision trees' strength in handling complex data patterns.
- **Data Characteristics:**
 - Thorough analysis of overall distribution and characteristics.
 - Determination of decision tree efficacy for stock analysis.

Dataset and Metrics

- **Dataset Partitioning:** 90-10
- **Target Variable:** Previous day's closing price as the prediction target
- **Evaluation Metrics:**
- **Primary:** Mean Squared Error (MSE)
- **Supplementary:** R-squared value for comprehensive evaluation

GRADIENT BOOSTING MODEL

Model Configurations

- **Initial Model:**
 - 300 estimators
 - Learning rate of 0.1
- **Subsequent Model:**
 - 1000 estimators
 - Reduced learning rate of 0.01

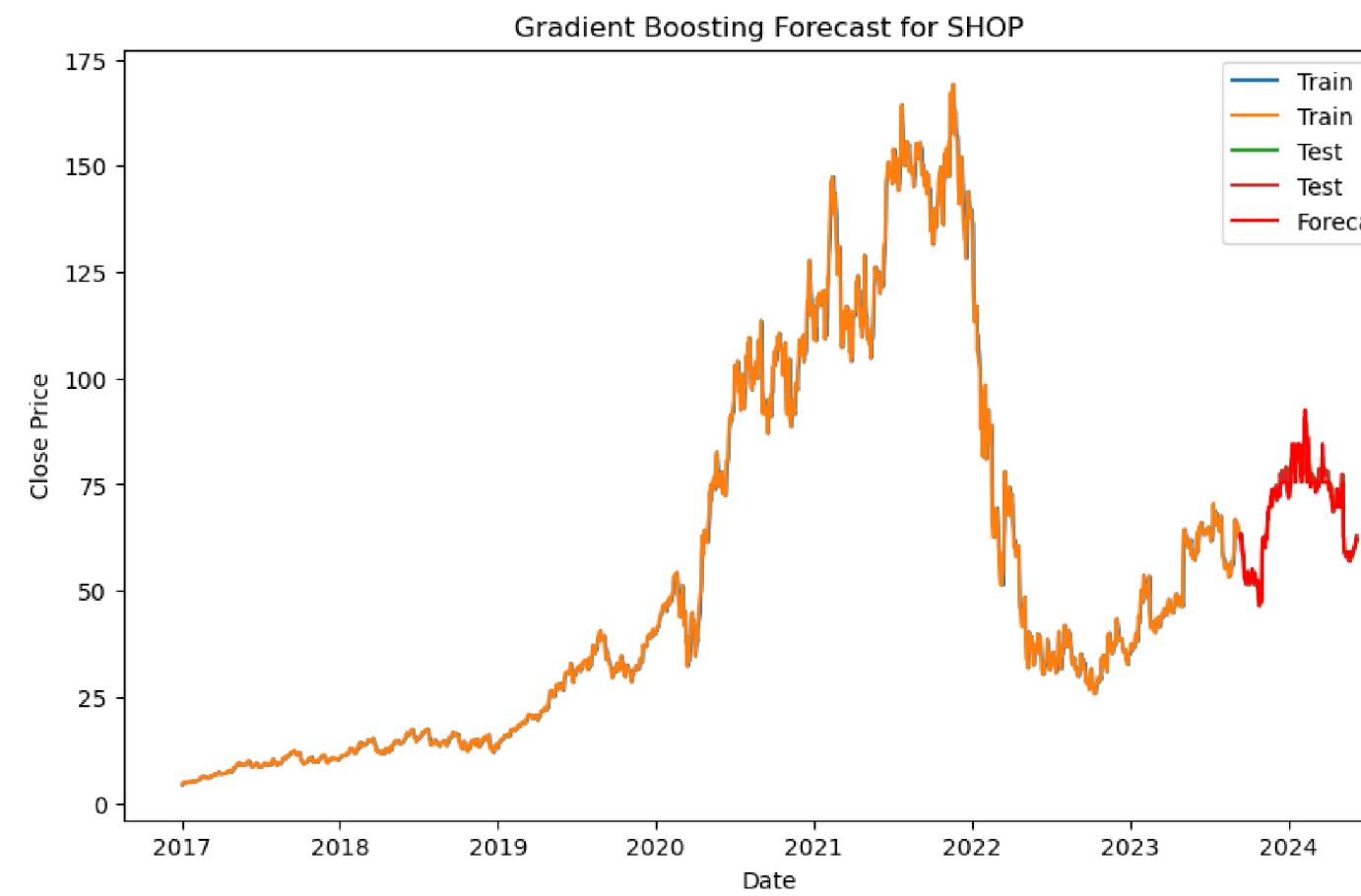
Model Performance Evaluation

- **Performance Metrics:**
 - Evaluation based on Mean Squared Error (MSE)
- **Visual Examination:**
 - Plotting predictions against actual closing prices
 - Facilitates visual comprehension of model performance

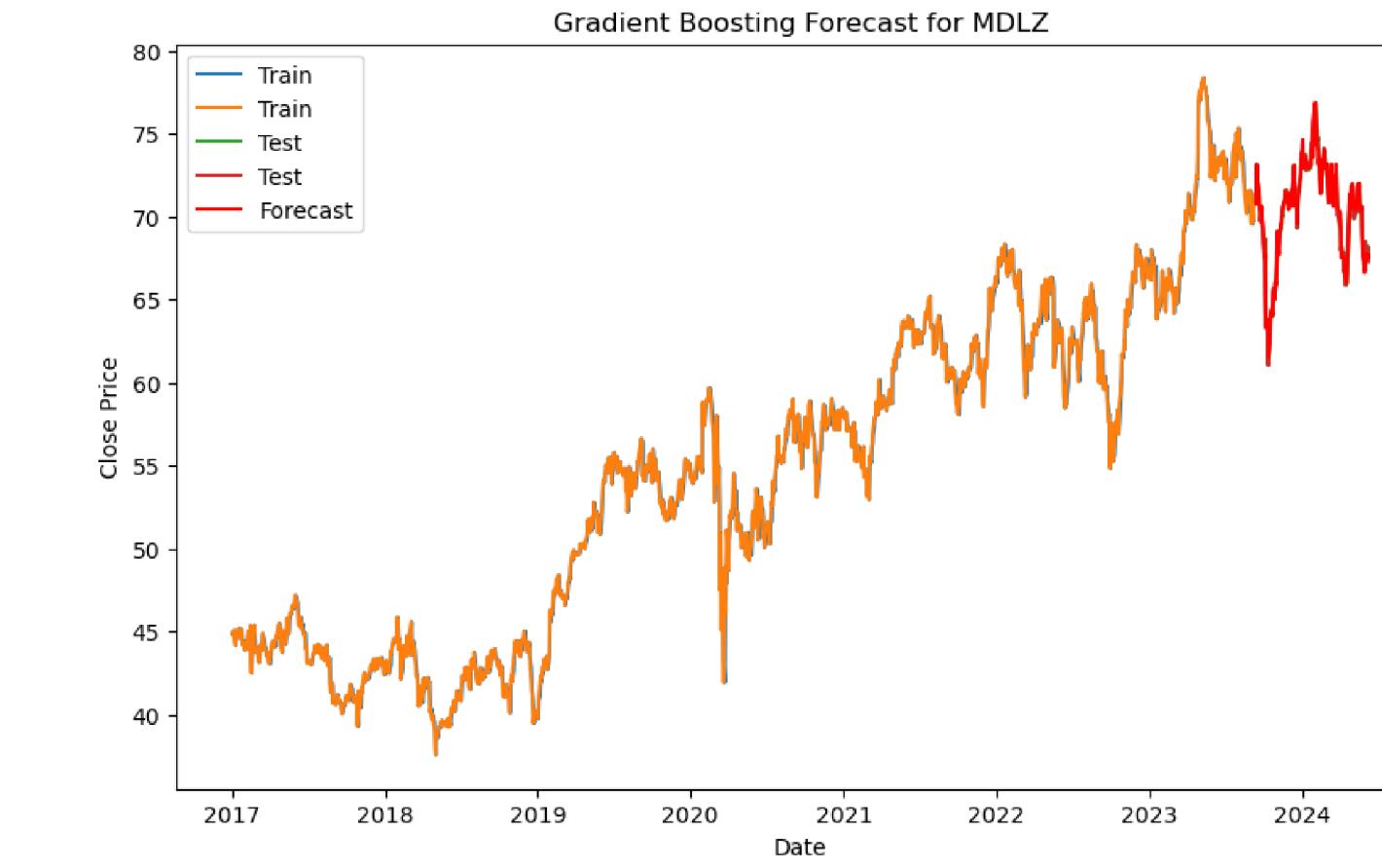
Conclusion

- **Key Insights:**
 - Gradient Boosting's robust performance in stock price prediction
 - Effectiveness of decision tree-based ensemble models in time series data
- **Future Directions:**
 - Potential model refinements
 - Exploration of additional features or alternative models for improved accuracy

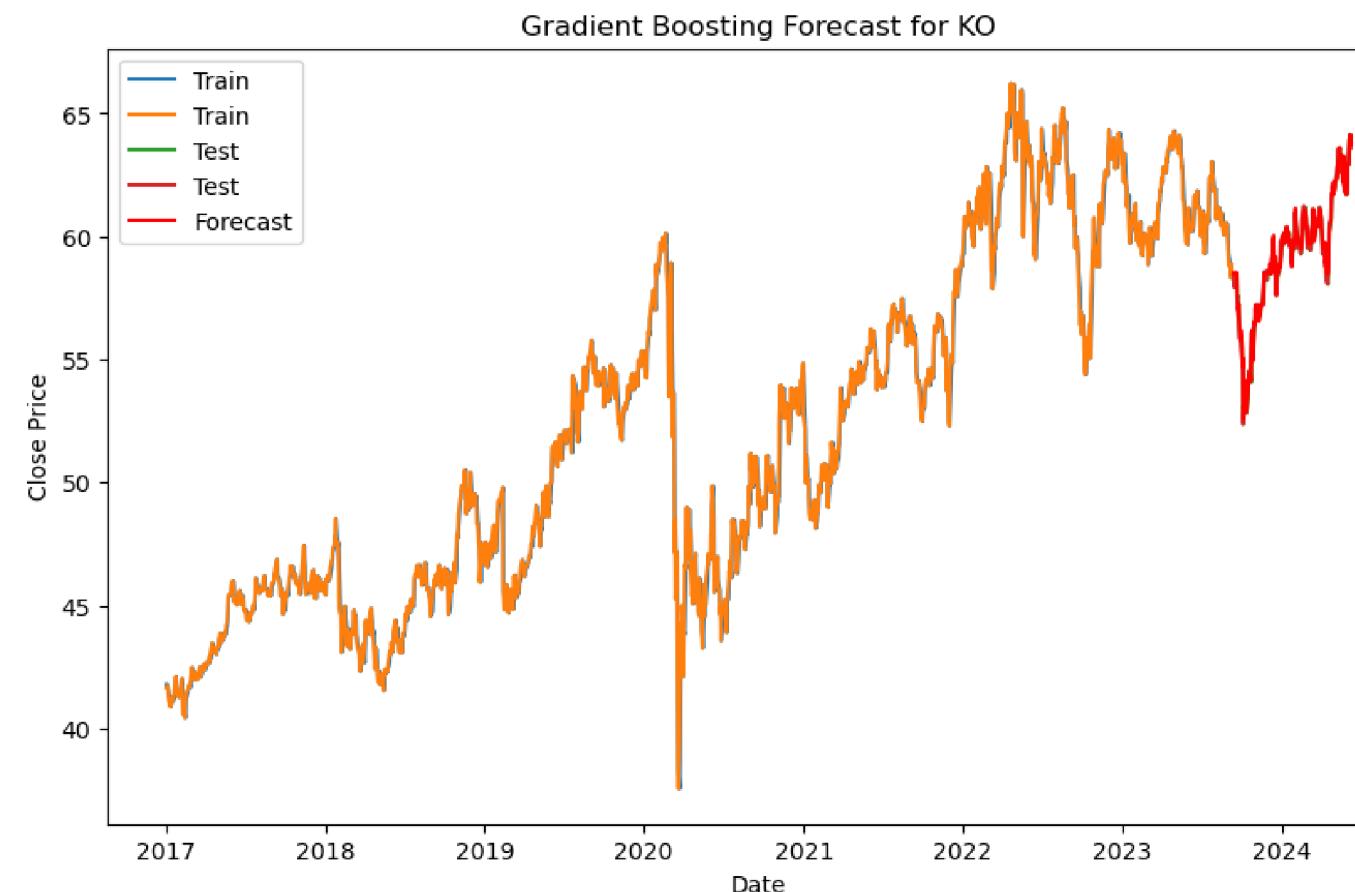
MODEL PERFORMANCE EVALUATION



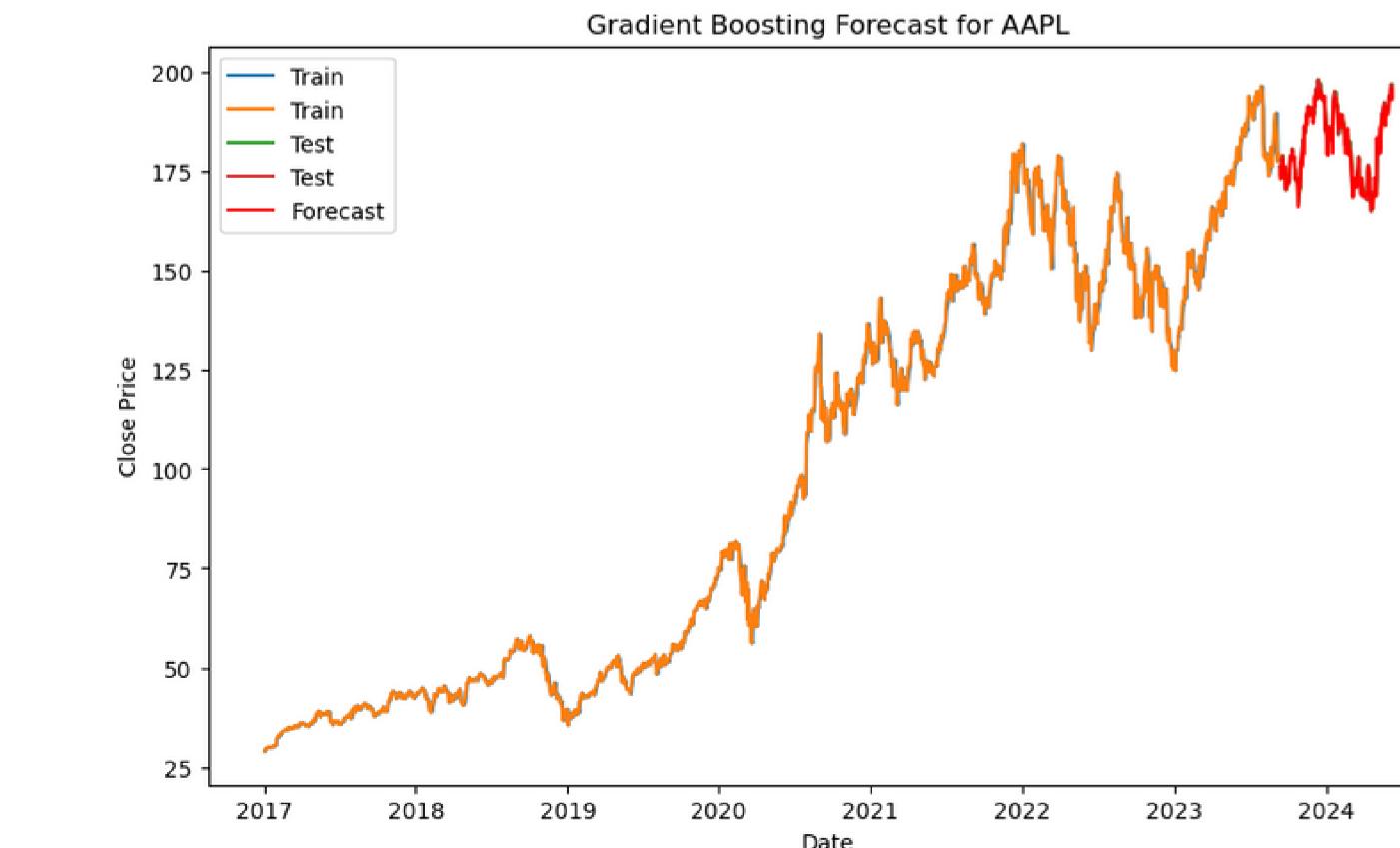
SHOP
MSE: 6.82
RMSE: 2.61



KO
MSE: 0.26
RMSE: 0.513

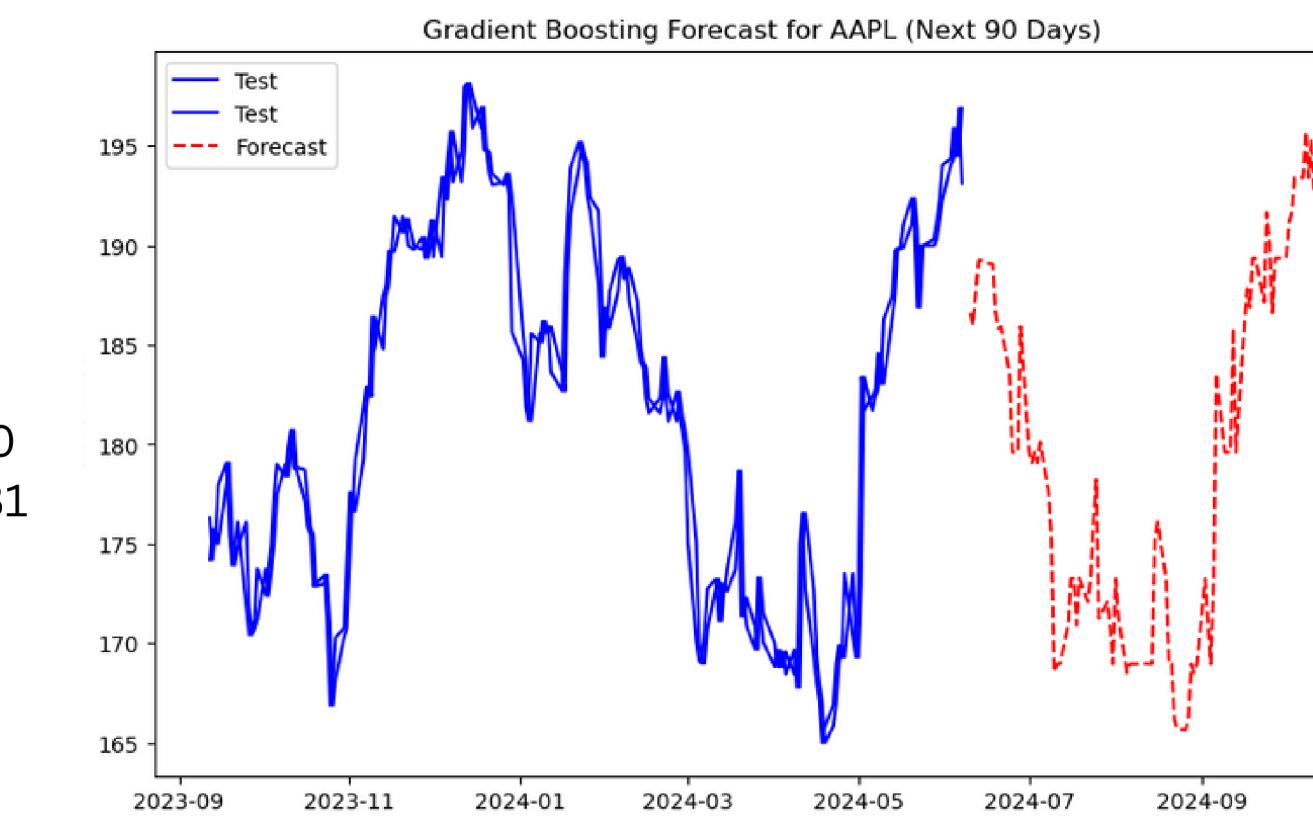
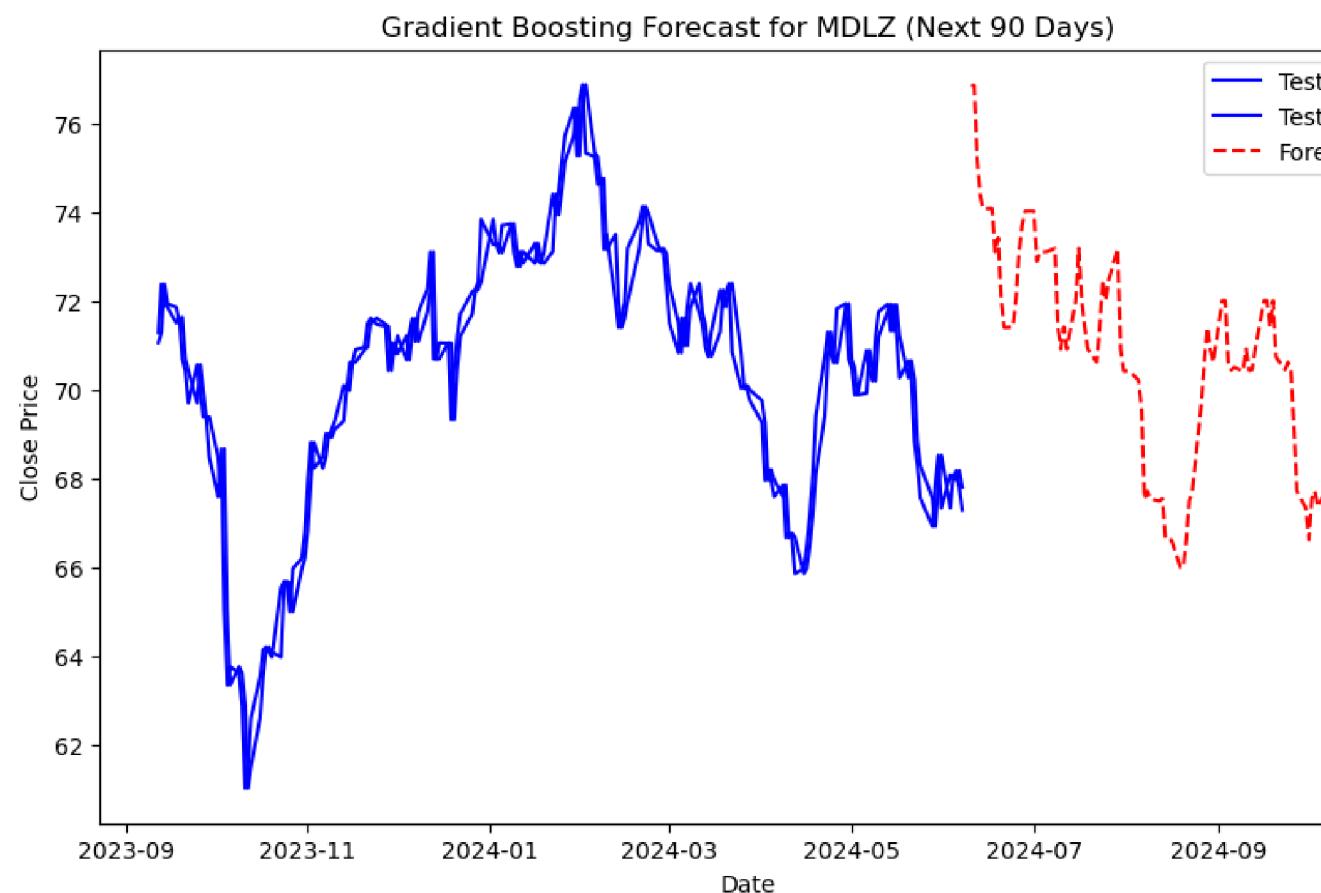
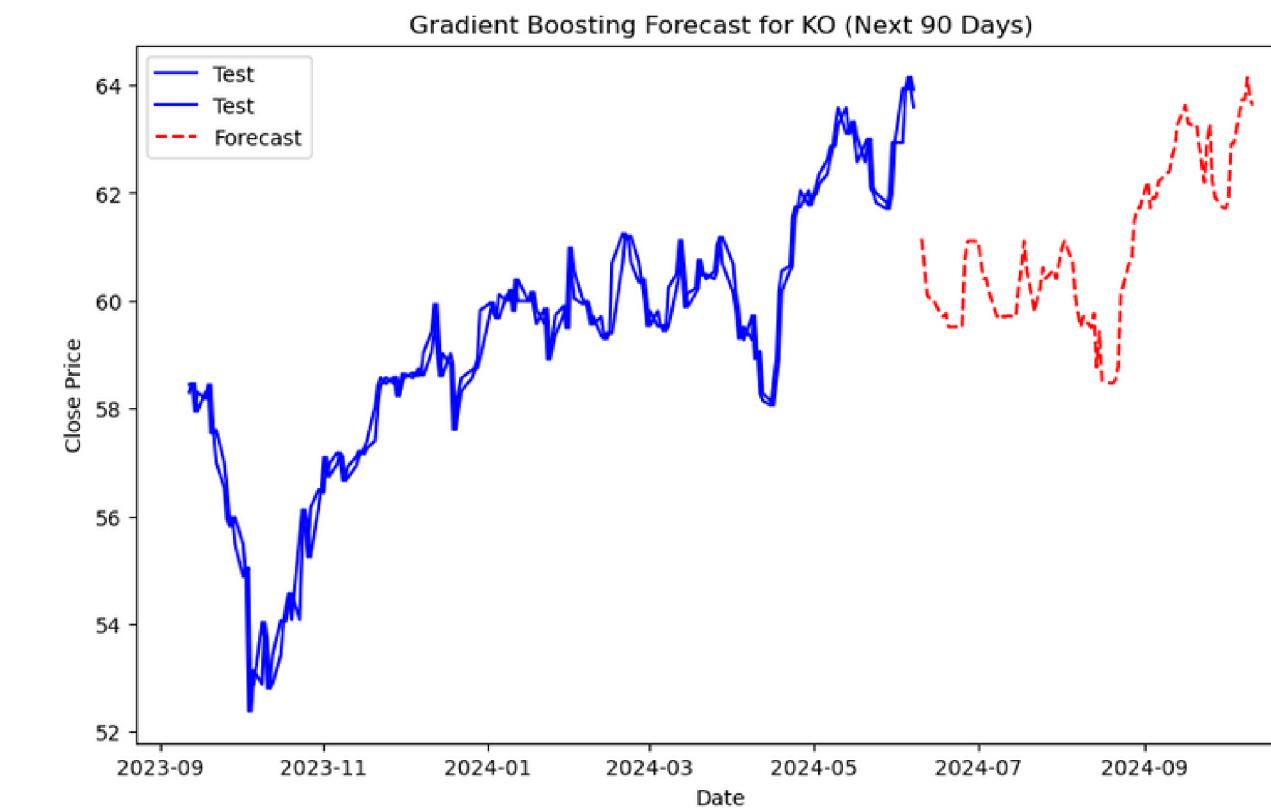
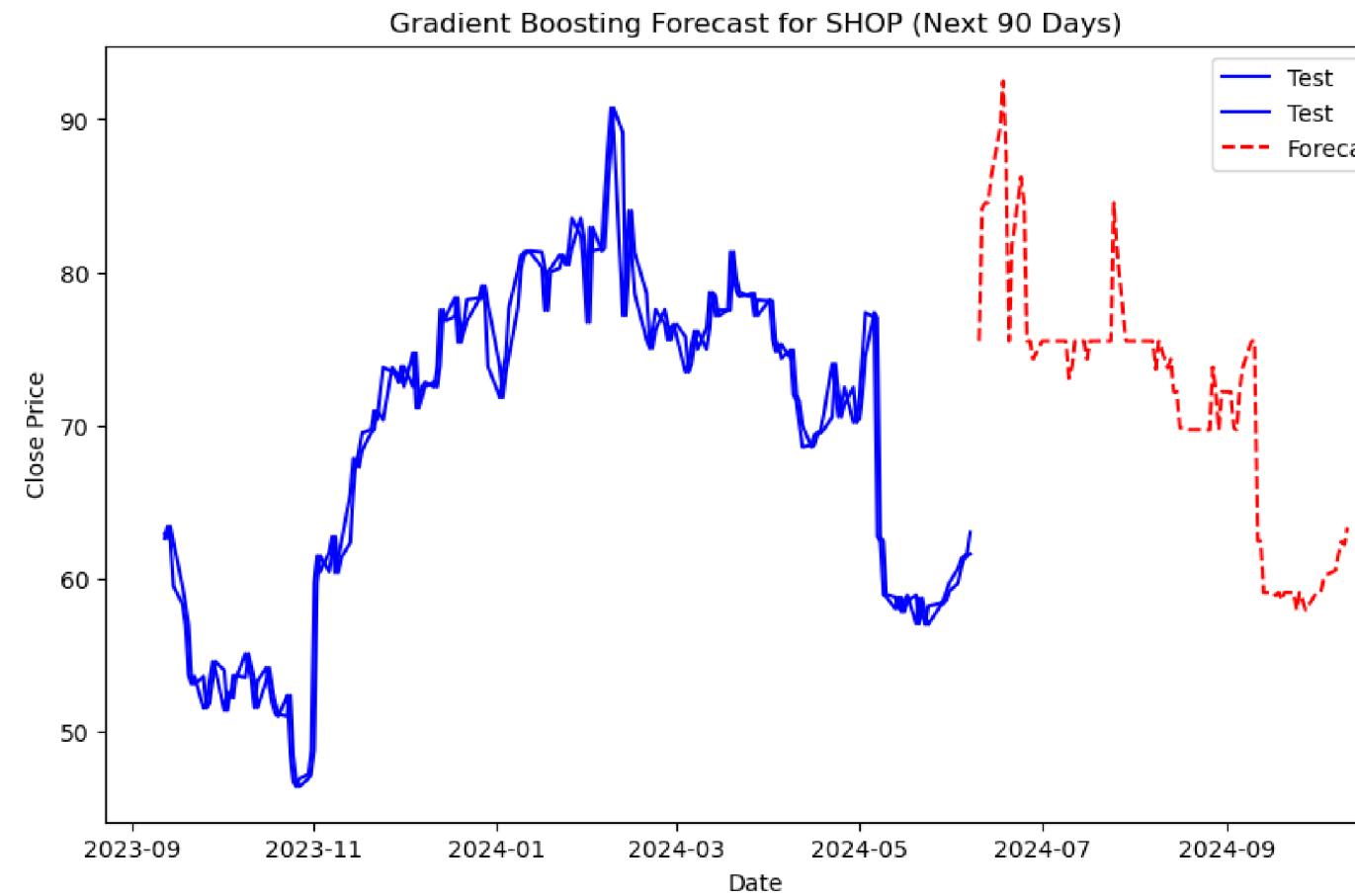


KO
MSE: 0.66
RMSE: 0.81



AAPL
MSE : 6.81
RMSE : 2.61

Predications for next 90 days using GBL model:

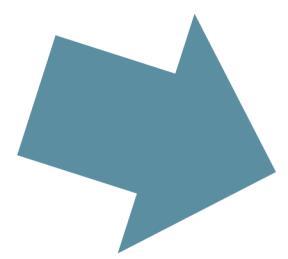


THE DATASET

1) Yahoo! Finance data

A screenshot of the Yahoo! Finance website. At the top, there's a navigation bar with links like News, Taxes, Watchlists, My Portfolio, My Screeners, Market Data, Industry News, and Real E. Below the navigation is a banner with market indices: S&P/TSX (22,007.00), S&P 500 (5,346.99), DOW (38,798.99), and CAD/USD (0.7265). The main content area shows Apple Inc. (AAPL) with a price of 196.89, up +2.41 (+1.24%) at close: June 7 04:00PM EDT. Below this is a 'Historical Data' section with a table of daily price data from Dec 11, 1980, to Jun 07, 2024.

Yahoo! finance
.csv manual download



The yfinance 0.2.40 library page. It features a Python logo icon, the text 'yfinance 0.2.40', and a button to 'pip install yfinance'. Below this is a sub-section titled 'Download market data from Yahoo! Finance API'.

yfinance Library
Python 3.10 or lower



The YH Finance API Specification page. It shows the URL https://ytapi.net - Base URL and a list of API endpoints under the 'API' section:

- GET /v6/finance/quote
- GET /v7/finance/options/{symbol}
- GET /v8/finance/spark
- GET /v11/finance/quoteSummary/{symbol}
- GET /v8/finance/chart/{ticker}
- GET /v6/finance/recommendationsbySymbol/{symbol}
- GET /us/screeners/v1/finance/screeners/predefined/saved

YH Finance API
Pricing with rate Limit

RESEARCH QUESTIONS

1. How accurately can we predict the future stock prices of a particular company based on historical stock prices, trading volume, and relevant market indicators?
2. What are the most influential factors affecting the stock prices of companies, and how can we leverage machine learning algorithms to forecast these prices?
3. Can sentiment analysis of news articles, social media posts, and financial reports be used to predict short-term fluctuations in stock prices, and how does this predictive power vary across different industries?
4. Does stock perform generally better in the first half of the year/ later half in the year
5. How can the accuracy of stock price predictions be visually assessed using graphs, and what specific features in the graph would indicate that the model is performing well?

VISUALS/ DATA EXPLORATION

Data Summary

- **Categories:** Consumer Discretionary, Consumer Staples, Information Technology, Health care, and Industrial
- **Metrics:** Avg close, high, open prices, and volume

Key Metrics

- **Avg Opening Price:** IT shows sharp rise post-2018
- **Avg Closing Price:** Significant rise in IT post-2019
- **Avg High Price:** Consistent upward trend in IT

Detailed Analysis

- **Candlestick Graph:** Daily price movements over the last three months
- **Avg Volume Per Stock:** IT leads in trading volume, especially post-2020

Insights

- **Price Trends:** Highest growth in Information Technology
- **Volume Trends:** IT dominates market activity

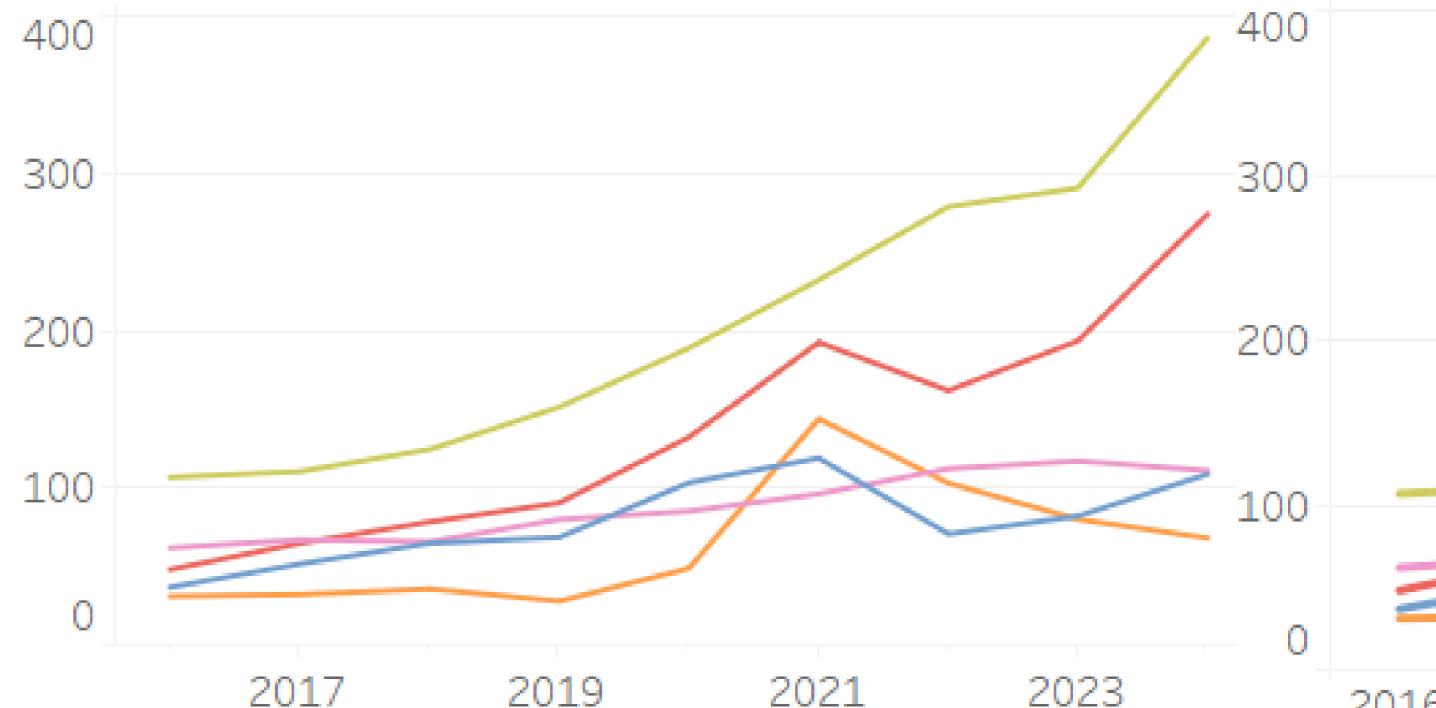


Stock Analysis Dashboard

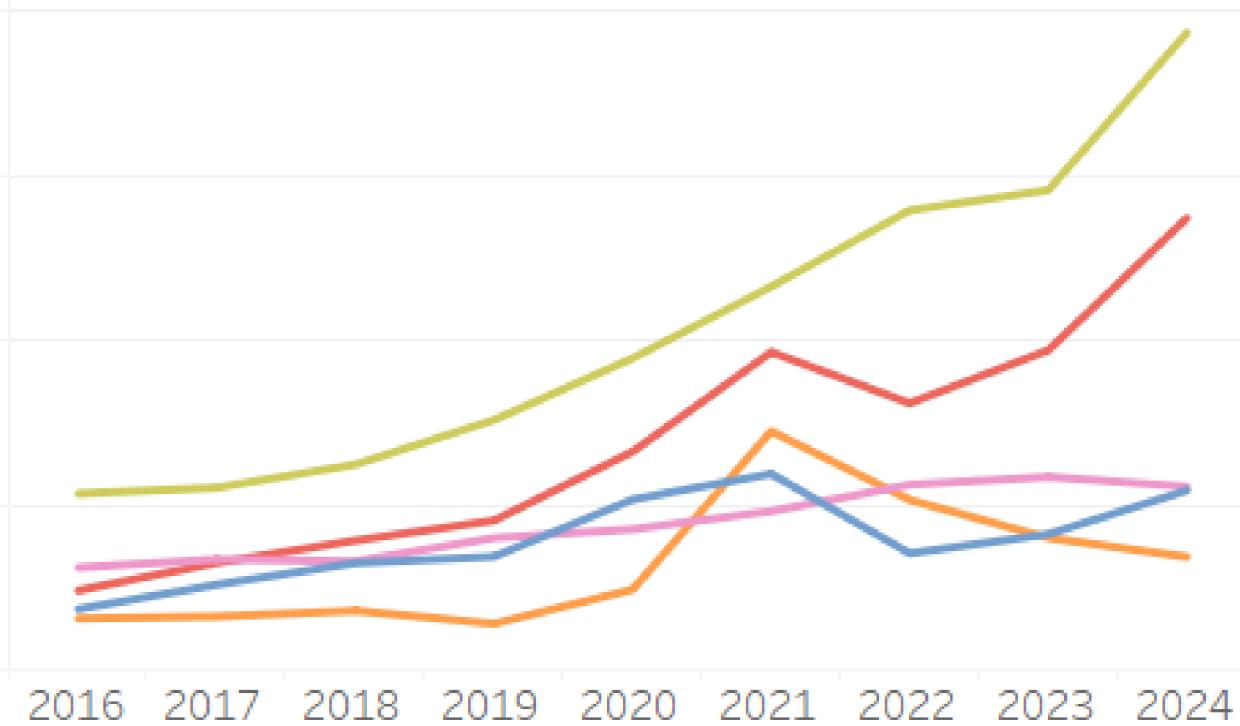
Stocks Category	Avg. Close	Avg. High	Avg. Open	Avg. Volume
Consumer Discret..	63	64	63	32,800,807
Consumer Staples	71	72	71	11,228,509
Health Care	56	57	56	26,406,640
Industrial	144	145	144	2,348,238
Information Tech..	87	88	87	82,748,245

Stocks Category	(All)
Consumer Discretionary	
Consumer Staples	
Health Care	
Industrial	
Information Technology	

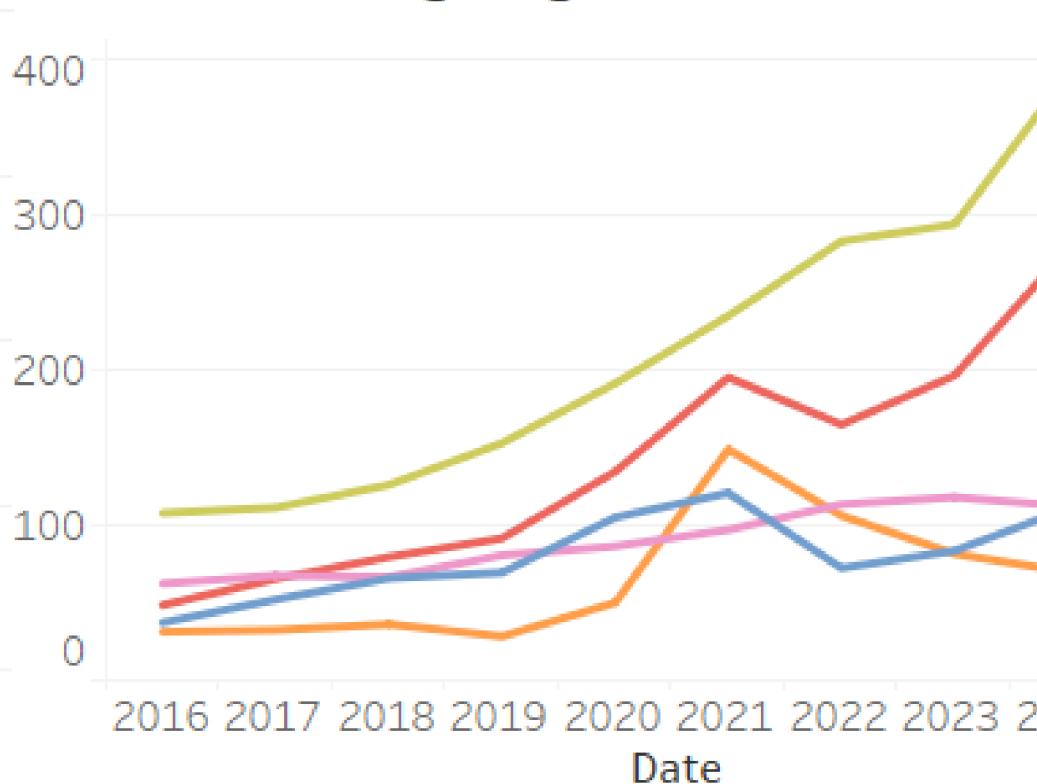
Avg Opening Price



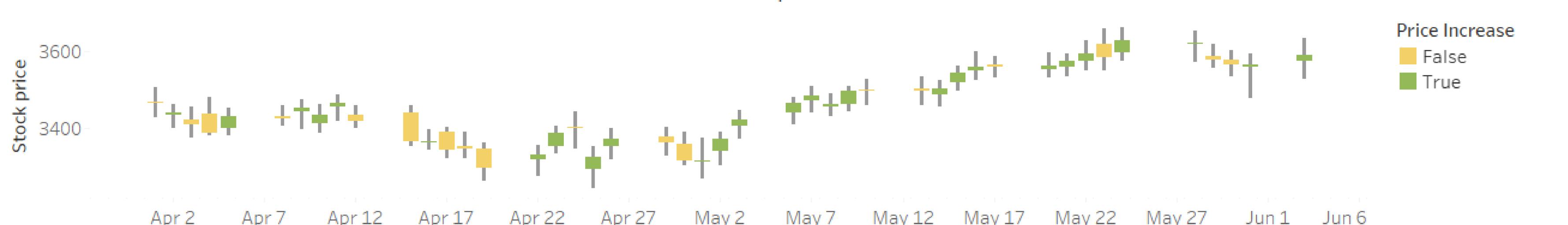
Avg Closing Price



Avg High Price



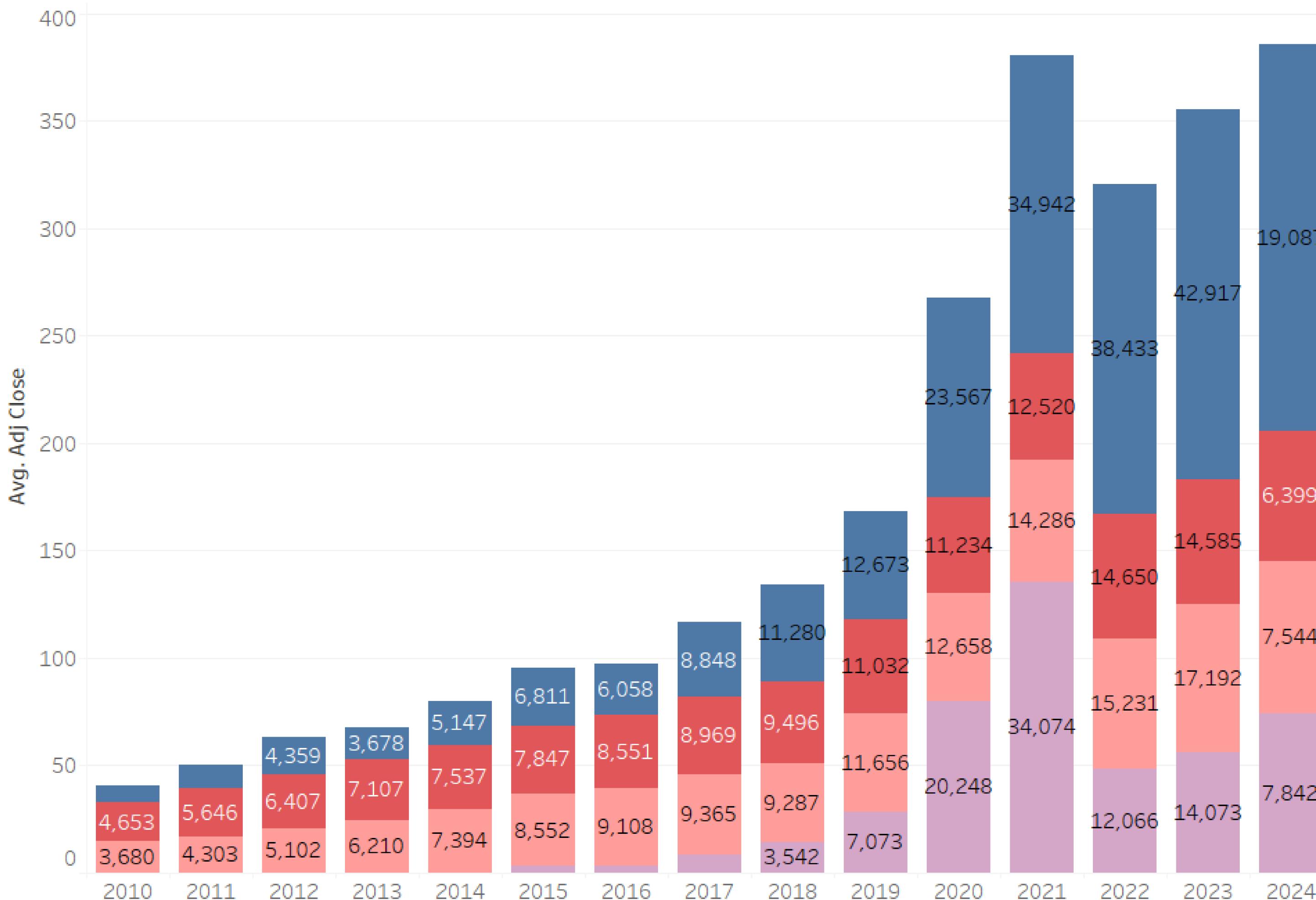
Candlestick Graph



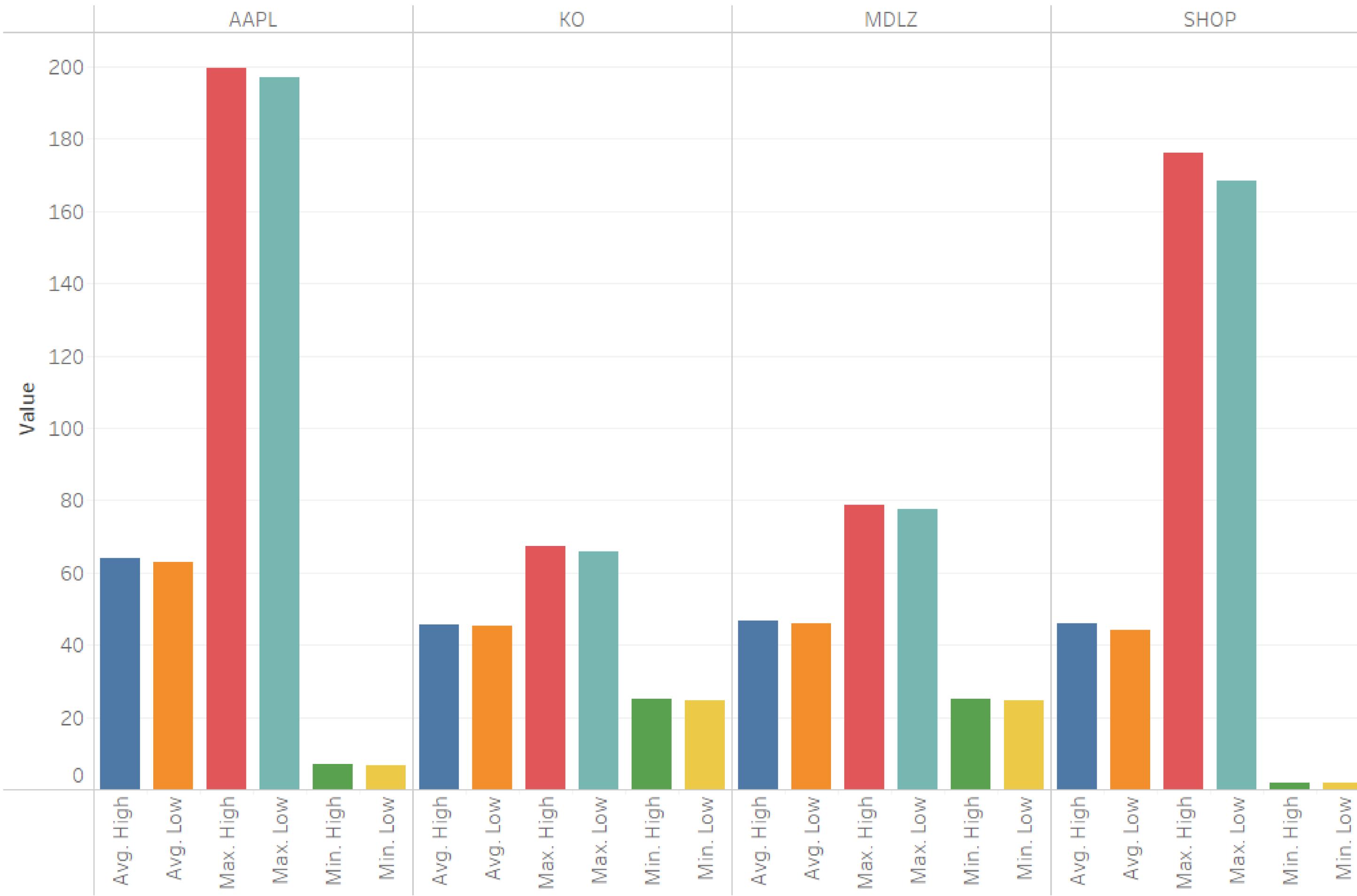
Avg Volume Per Category

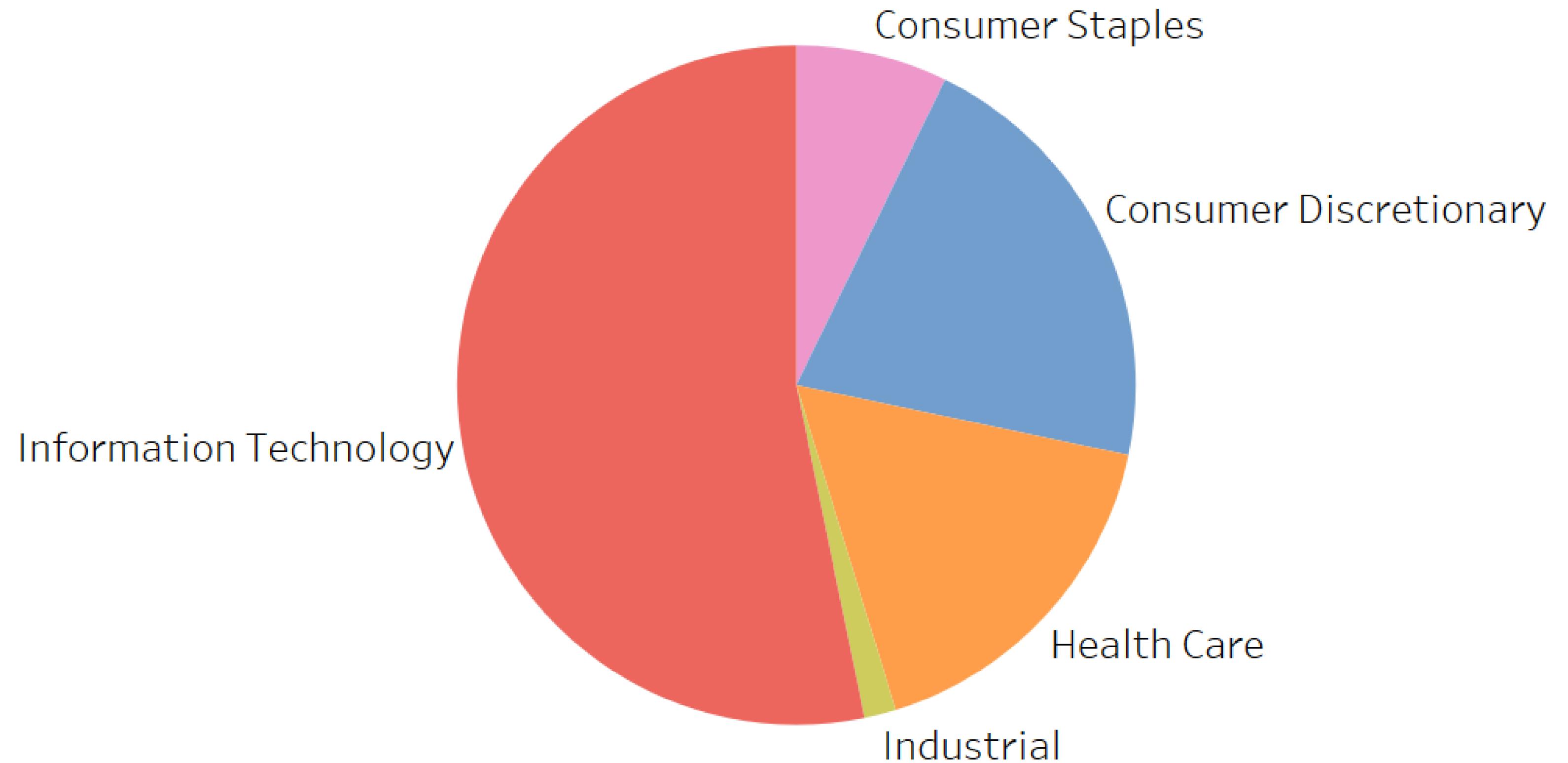
Information Technology AAPL 236,972,460	Information Technology AMD 46,586,253	Consumer Discretionary AMZN 84,969,910	Consumer Discretionary EBAY 16,684,964
		Consumer Discretionary BABA 19,096,920	
	Information Technology MSFT 36,814,688	Consumer Discretionary SHOP 17,281,310	
		Consumer Staples WMT 26,077,039	Consumer Staples KO 15,010,763
			Consumer Staples MDLZ
Information Technology GOOG 54,545,493	Information Technology META 30,135,763	Health Care PFE 33,302,031	

Avg Volume Per Stock



Comparison of Average, Maximum, and Minimum Low Values for Selected Stocks



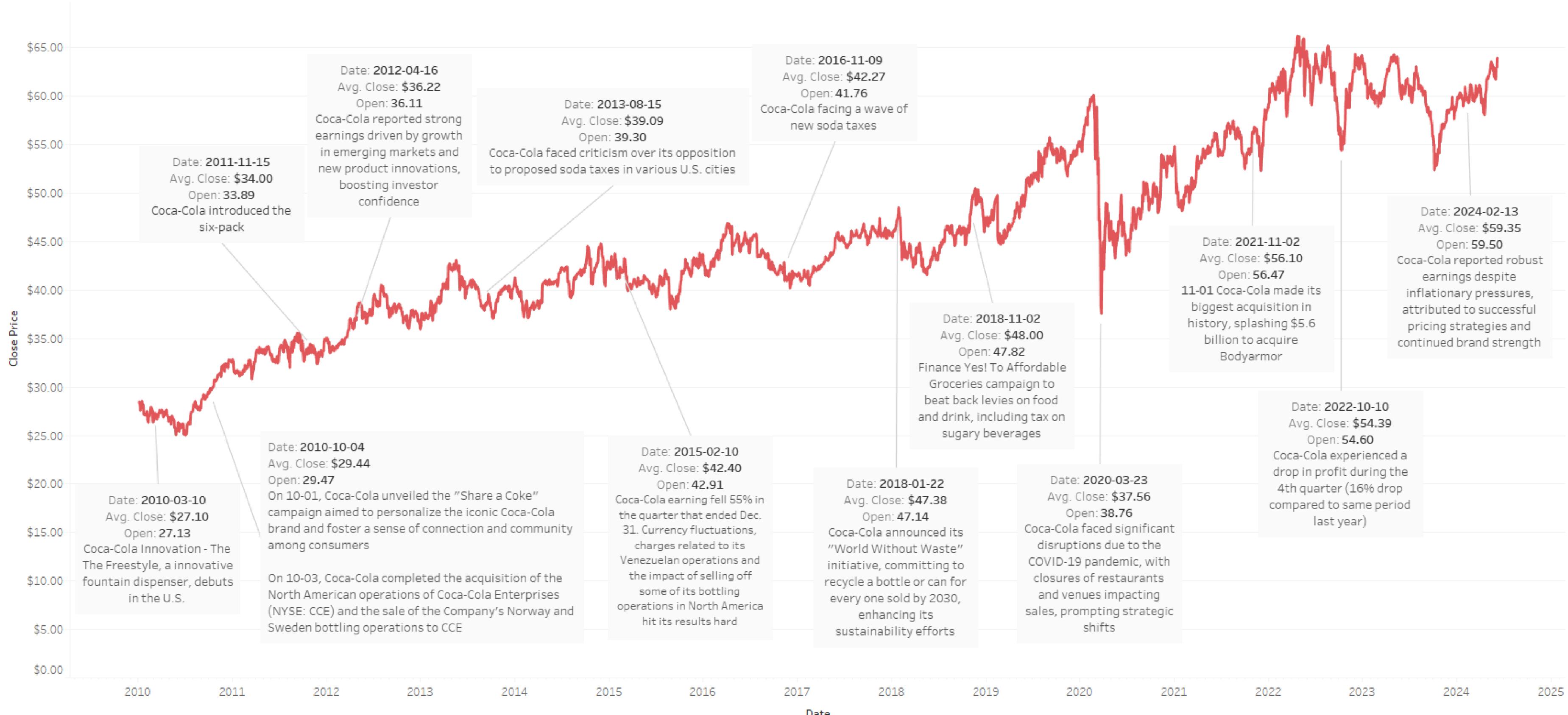


STOCKS FOR ML MODEL



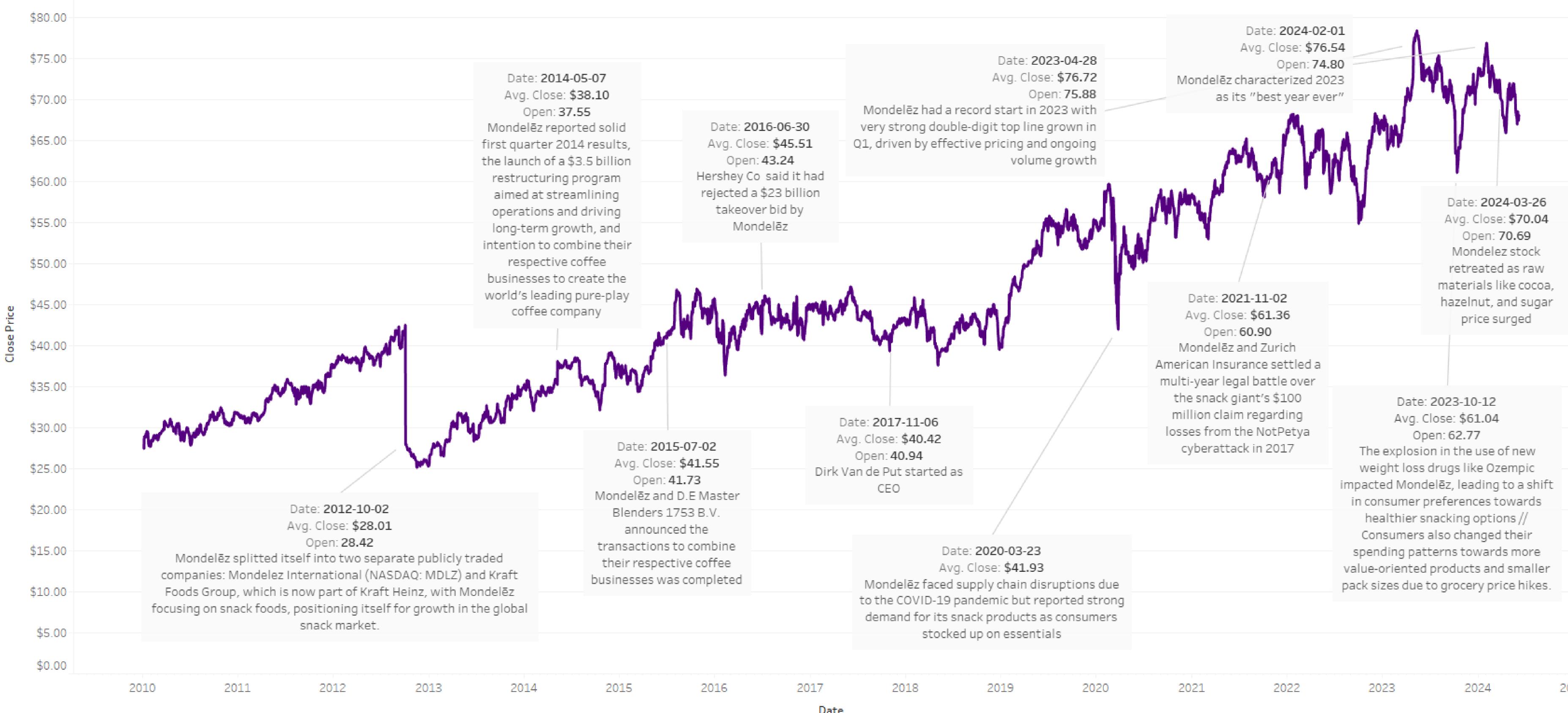
COCA-COLA CO

NYSE: KO



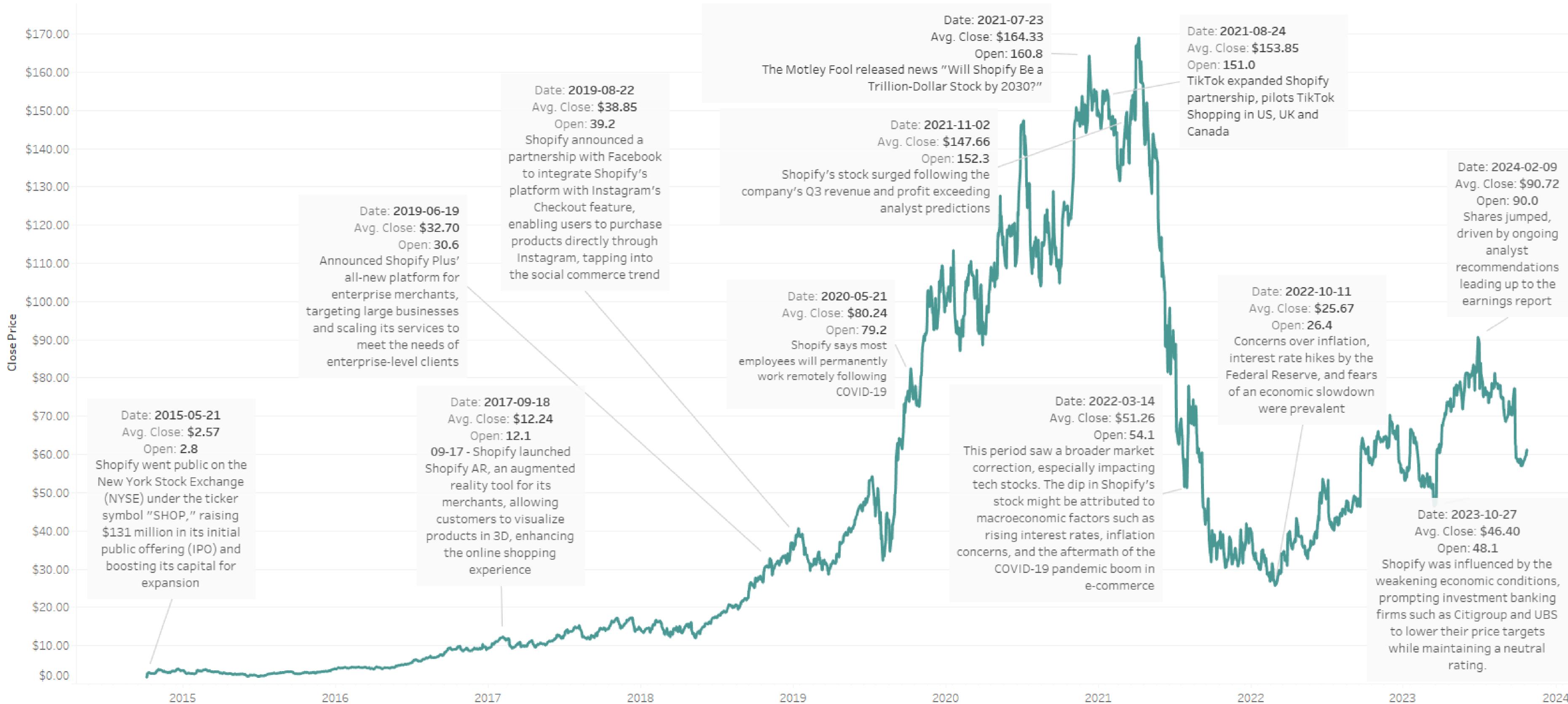
MONDELĒZ INTERNATIONAL, INC.

NASDAQ:MDLZ



SHOPIFY INC.

TSE:SHOP



APPLE INC.

NASDAQ: AAPL



ML FOR TIME SERIES ANALYSIS

Considerations when choosing the correct machine learning model for time series analysis

Nature of the Data

- Does the data exhibits seasonal patterns, trends, or cycles?
- Is the time series is stationary or if it needs differencing or transformation?

Model Complexity and Interpretability

- Is simple model like ARIMA sufficient for straightforward patterns?
- Can the model's predictions and parameters be easily understood and explained?

Handling Missing Data and Outliers

- Can the model effectively handle missing data and outliers without significant preprocessing?
- How the model manages periods with no data due to weekends, holidays, or other interruptions?

Scalability and Computational Efficiency

- Is the time and cost required for model training and forecasting manageable?
- Can the model be scale with the size of the dataset, especially for high-frequency time series?

Forecast Horizon and Accuracy

- Is the model suitable for the forecasting horizon, whether short-term or long-term?
- What metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), or R-squared (R^2) value, are we using to evaluate the the model's accuracy?



Machine Learning Model 2:

ADDICTIVE REGRESSION MODEL/ PROPHET LIBRARY

Prophet, developed by Facebook, is an additive regression model designed for forecasting time series data.

The model assumes that the time series $y(t)$ can be decomposed into several components: trend ($g(t)$), seasonality ($s(t)$), and holidays ($h(t)$), along with an error term (ϵ_t).

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$



Handling Seasonality and Trends

- Automatically adjusts for yearly, weekly, and daily patterns
- Includes changepoints to model shifts in stock price trends



Flexibility and Customizability

- Accounts for significant dates impacting stock prices
- Customizable settings to capture the nuances of stock price



Robust to Missing Data and Outliers

- Handles gaps in data due to market holidays and interruptions
- Effectively deals with occasional price spikes or drops



Ease of Use and Interpretability

- Simple functions for fitting models and making forecasts
- Built-in tools for visualizing trends and seasonality components



Scalability and Speed

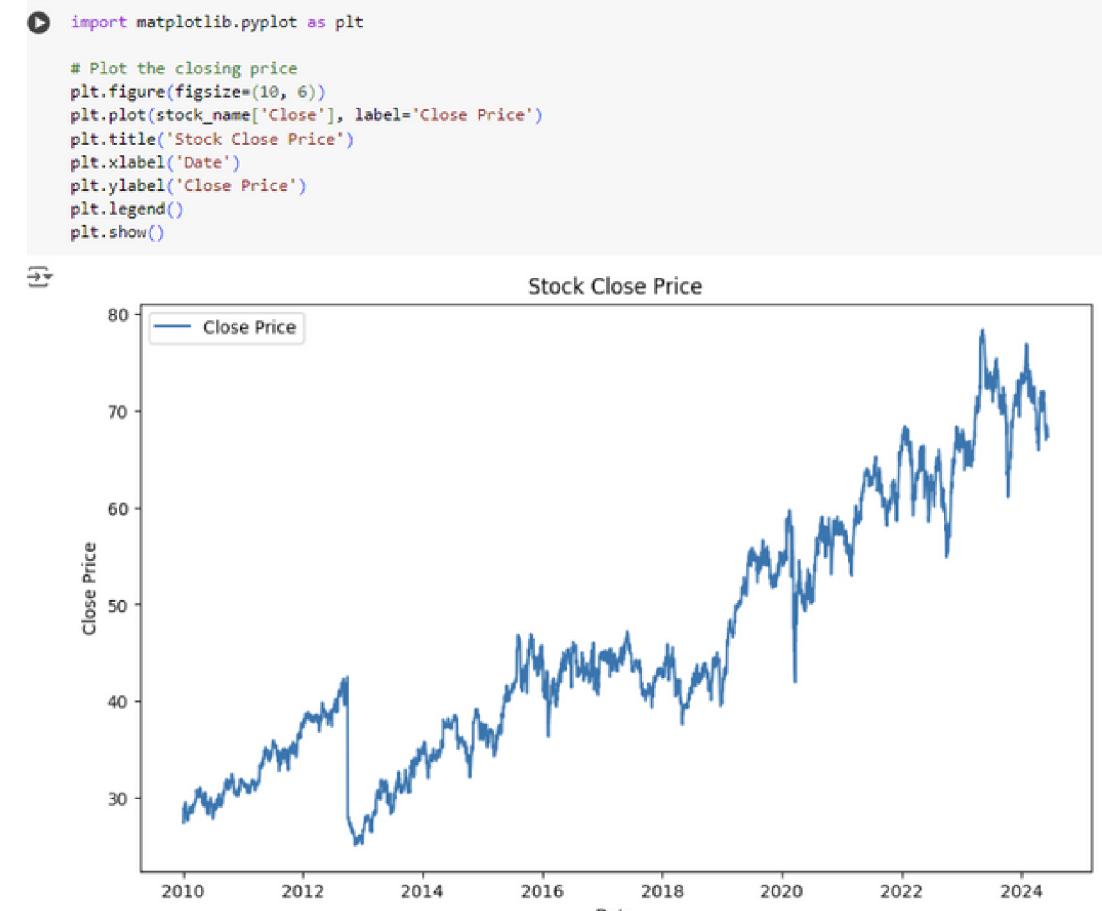
- Efficient with large datasets, ideal for long historical stock data
- Quick training and prediction updates with new data

BASE MODE: ADDICTIVE REGRESSION MODEL

Prophet Library without technical indicators

1. Load and Summarize Dataset

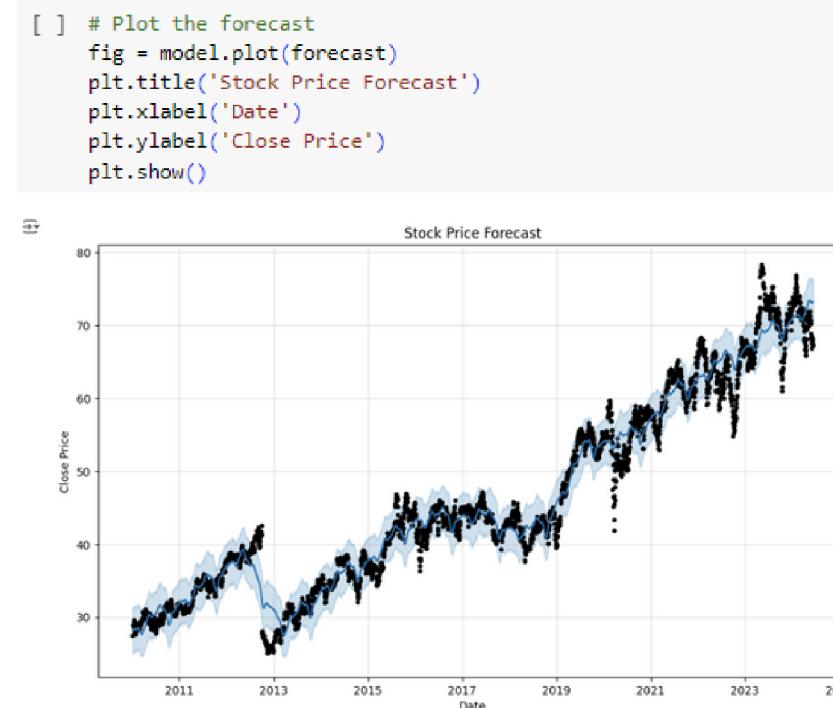
```
[ ] #!pip install prophet  
  
[ ] import prophet  
print('Prophet %s' % prophet.__version__)  
Prophet 1.1.5  
  
[ ] import yfinance as yf  
import pandas as pd  
  
# Load the data  
stock_name = yf.download('MDLZ', start='2010-01-01', end='2024-12-31')  
  
# Summarize the dataset  
stock_name.shape  
  
2. Plot the Data on Column 'Close'
```



3. Data Pre-processing and Fit the Prophet Model

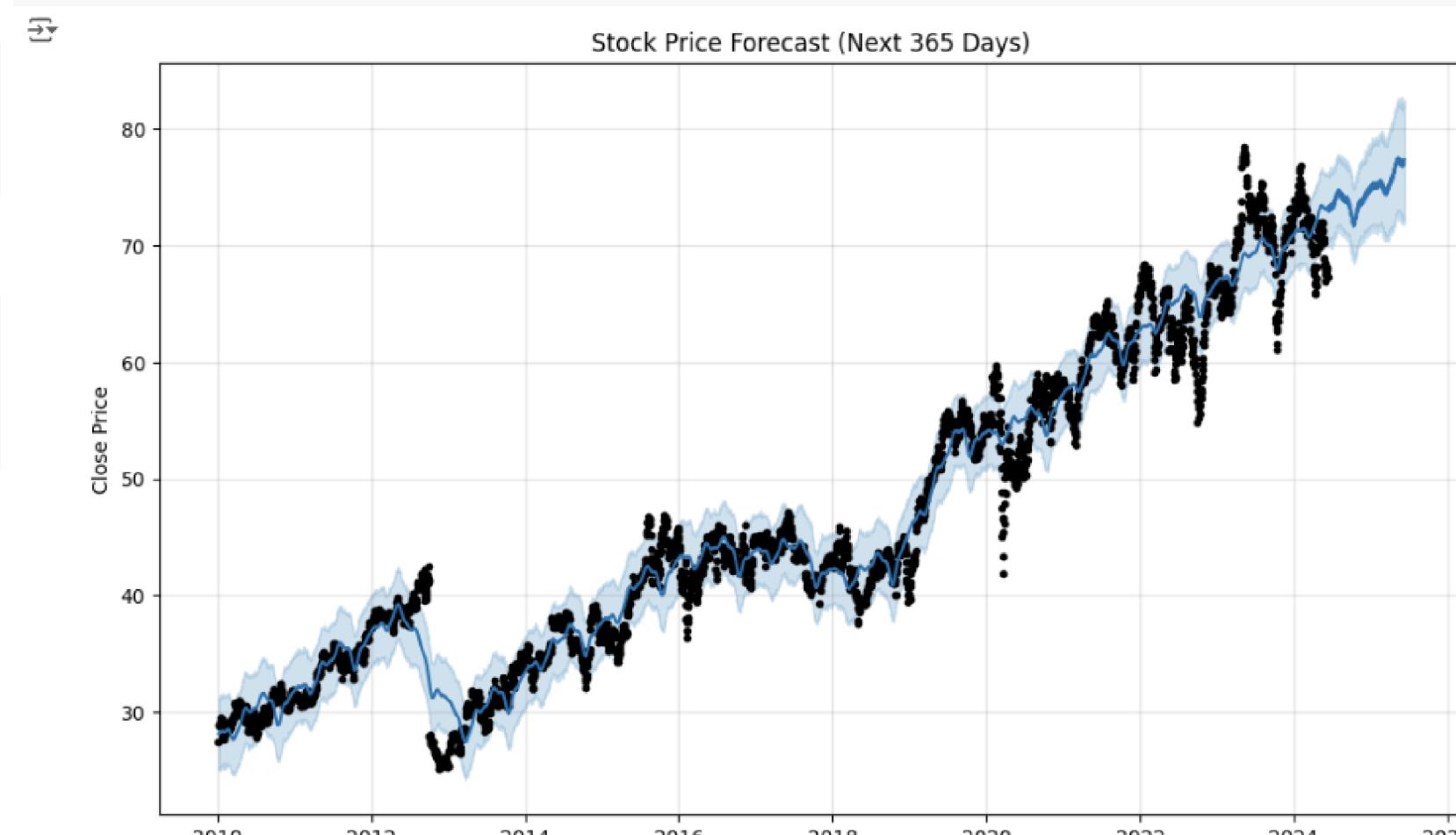
```
[ ] from prophet import Prophet  
  
# Prepare the expected column names  
df = stock_name[['Close']].reset_index()  
df.columns = ['ds', 'y']  
df['ds'] = pd.to_datetime(df['ds'])  
  
# Fit the Prophet model  
model = Prophet()  
model.fit(df)  
  
future = model.make_future_dataframe(periods=0)  
forecast = model.predict(future)  
  
# Summarize the forecast  
print(forecast[['ds', 'yhat', 'yhat_lower', 'yhat_upper']].head())
```

5. Visualize the Prediction in the Context of the Training Dataset



6. Make an Out-of-Sample Forecast (Next 365 Days)

```
[ ] # Extend the forecast for the next 365 days  
future = model.make_future_dataframe(periods=365)  
forecast = model.predict(future)  
  
# Plot the extended forecast  
fig = model.plot(forecast)  
plt.title('Stock Price Forecast (Next 365 Days)')  
plt.xlabel('Date')  
plt.ylabel('Close Price')  
plt.show()
```



BASE MODEL EVALUATION

Date: 2023-05-11
Avg. Close: 78.36
Open: 77.92

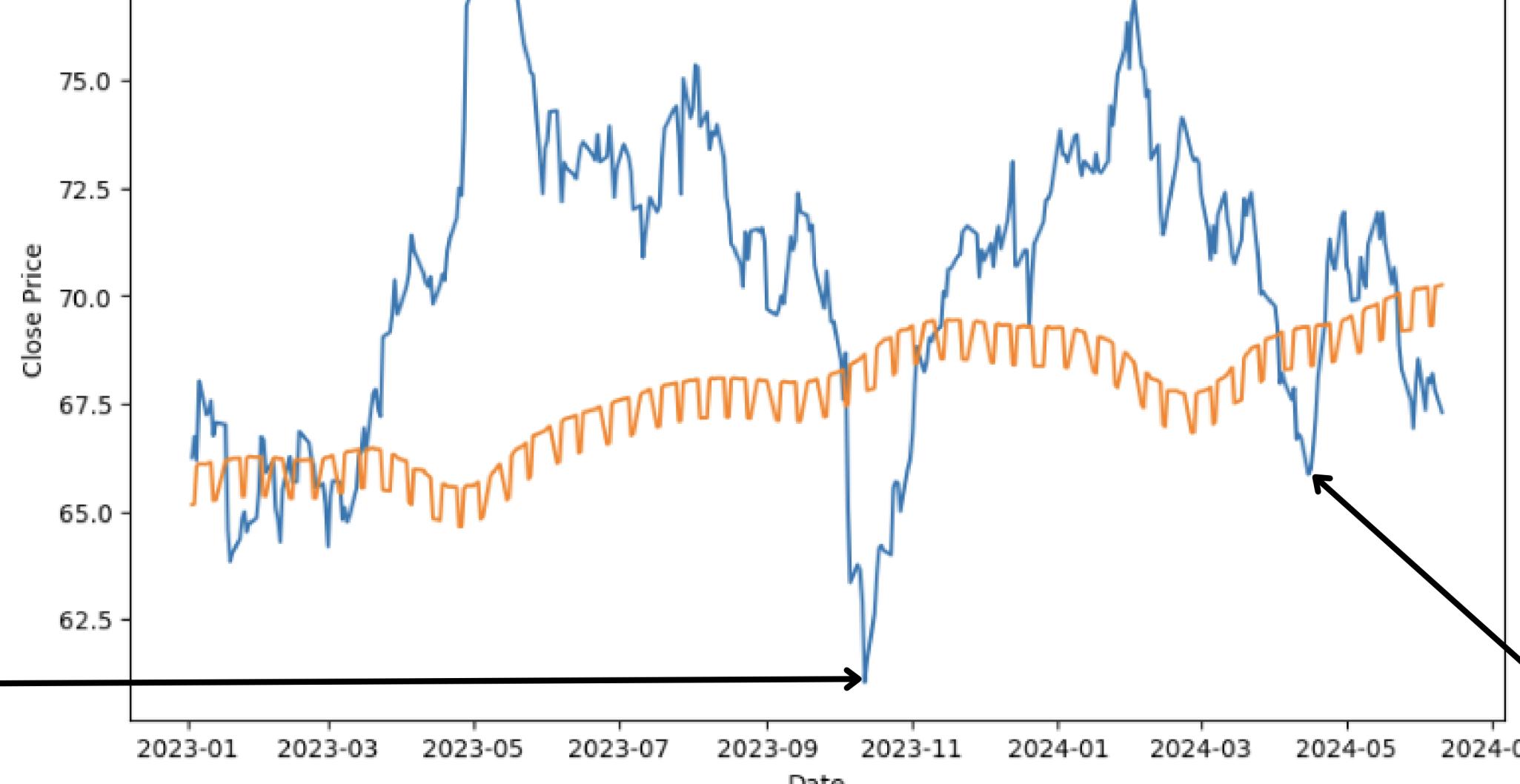
Mondelēz reported strong quarterly results, with a significant increase in net revenue and earnings per share

Date: 2023-10-12
Avg. Close: 61.04
Open: 62.77

The explosion in the use of new weight loss drugs like Ozempic impacted Mondelēz, leading to a shift in consumer preferences towards healthier snacking options // Consumers also changed their spending patterns towards more value-oriented products and smaller pack sizes due to grocery price hikes

Expected vs Predicted Close Prices

Actual
Predicted



Date: 2024-02-01
Avg. Close: 76.54
Open: 74.80

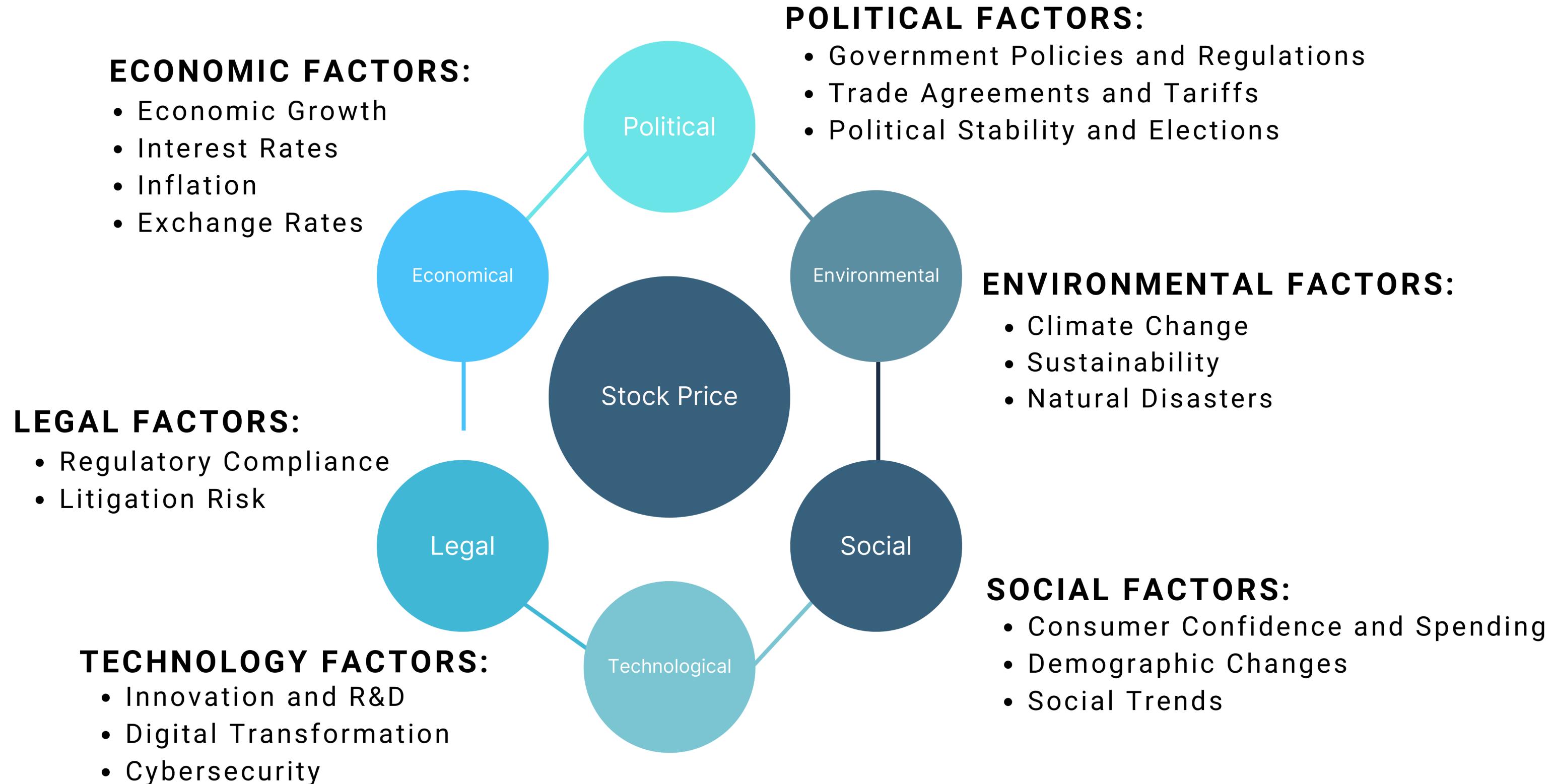
Mondelēz characterized 2023 as its "best year ever"

Date: 2024-03-26
Avg. Close: 70.04
Open: 70.69

Mondelez stock retreated as raw materials like cocoa, hazelnut, and sugar price surged

Mean Absolute Error: 3.7740146597497706
Mean Squared Error: 22.032384275050156
R-squared: -0.8154948378029072

EXTERNAL ANALYSIS



THE DATASET

2) News API

The screenshot shows the YH Finance API Specification page. At the top, there's a navigation bar with links for 'Specification', 'Tutorial', and 'Products'. Below it, a sub-navigation bar has 'YH Finance' and 'SA Finance' selected. The main content area is titled 'YH Finance API Specification 1.0.0 OAS3' and includes a link to 'yh-finance-api-specification.json'. A note states 'Real time low latency Finance API for stock market, crypto currencies, and currency exchange'. A 'Servers' dropdown is set to 'https://yfapi.net - Base URL'. The 'API' section lists several endpoints with their methods and URLs:

- GET /v6/finance/quote
- GET /v7/finance/options/{symbol}
- GET /v8/finance/spark
- GET /v11/finance/quoteSummary/{symbol}
- GET /v8/finance/chart/{ticker}
- GET /v6/finance/recommendationsbysymbol/{symbol}
- GET /ws/screeners/v1/finance/screeners/predefined/saved

2) Alpha Vantage

The screenshot shows the Alpha Vantage homepage. The header features the 'ALPHA VANTAGE' logo and navigation links for 'ABOUT', 'DOCUMENTATION', 'SPREADSHEETS', 'ACADEMY', and 'SUPPORT'. The main headline is 'Stock Market API, Reimagined'. Below it is a bulleted list of features:

- Realtime & historical stock market data APIs
- Options, forex, crypto & other asset classes
- 60+ technical & economic indicators
- Market news API & sentiments
- Global coverage

At the bottom, there are three buttons: 'STOCK MARKET API', 'GLOBAL NEWS API', and 'GET FREE API KEY'.

STOCK ANALYSIS & PREDICTION

Prophet Library 2 (With Technical Indicators):

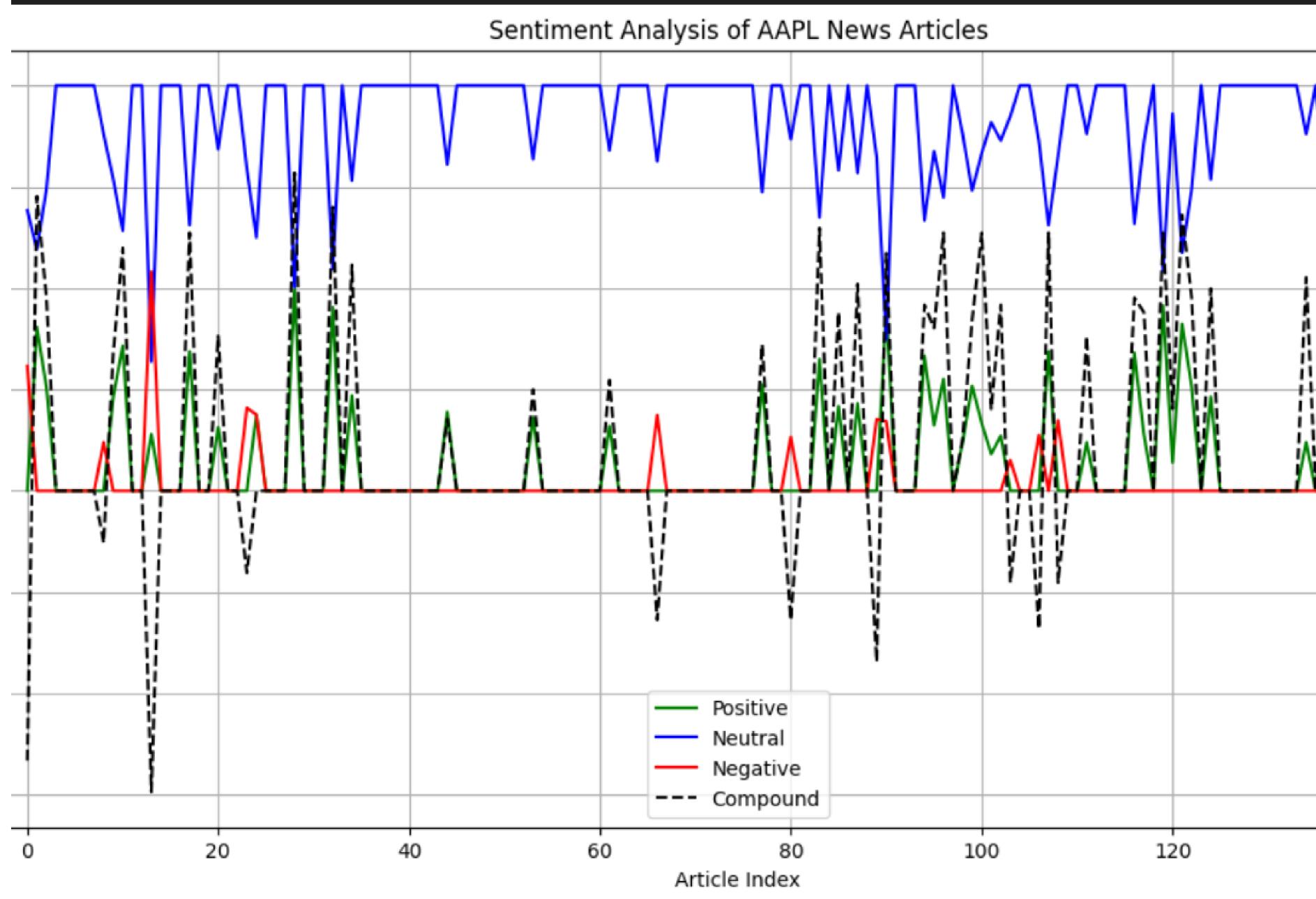
- **Enhanced Model:**
 - Uses SMA, EMA, RSI, and MACD
- **Robust Analysis:**
 - Comprehensive market analysis
- **Pros and Cons:**
 - Strengths and weaknesses explored
- **Performance:** Compared to other models
- **Ideal Use:** Best scenarios identified
- **Long-Term Focus:** For long-term investors
- **Market Insights:** Trends and patterns analyzed



SENTIMENT STUDY

Prophet Library 2 (With Sentimental Analysis):

```
Headline Sentiments:  
Date: 2024-06-05T20:29:18Z  
Headline: Apple Becomes $3 Trillion Company Again, But Nvidia Steals the Moment  
Sentiment: {'neg': 0.308, 'neu': 0.692, 'pos': 0.0, 'compound': -0.6652} (Category: Negative)  
-----  
Date: 2024-06-06T11:06:46Z  
Headline: Nvidia overtakes AAPL to exceed $3T value, wins an antitrust investigation  
Sentiment: {'neg': 0.0, 'neu': 0.596, 'pos': 0.404, 'compound': 0.7269} (Category: Positive)  
-----  
Date: 2024-06-10T14:17:22Z  
Headline: Apple wants to redefine AI as 'Apple intelligence' at WWDC  
Sentiment: {'neg': 0.0, 'neu': 0.744, 'pos': 0.256, 'compound': 0.4767} (Category: Positive)  
-----  
Date: 2024-05-16T10:27:55Z  
Headline: One hedge fund completely bailed out of AAPL, but another more than picked up the slack  
Sentiment: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0} (Category: Neutral)  
-----  
Date: 2024-05-14T14:25:16Z  
Headline: Google I/O Event Expectations Run High Amid OpenAI, Microsoft Rivalry  
Sentiment: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0} (Category: Neutral)  
-----  
Date: 2024-06-06T22:16:00Z  
Headline: NVIDIA Splits 10-to-1; Non-farm Payrolls on Deck for Friday  
Sentiment: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0} (Category: Neutral)  
-----  
...  
-----  
Mean Compound Score: 0.10325466666666668  
Median Compound Score: 0.0  
Standard Deviation of Compound Score: 0.27344816073902484
```



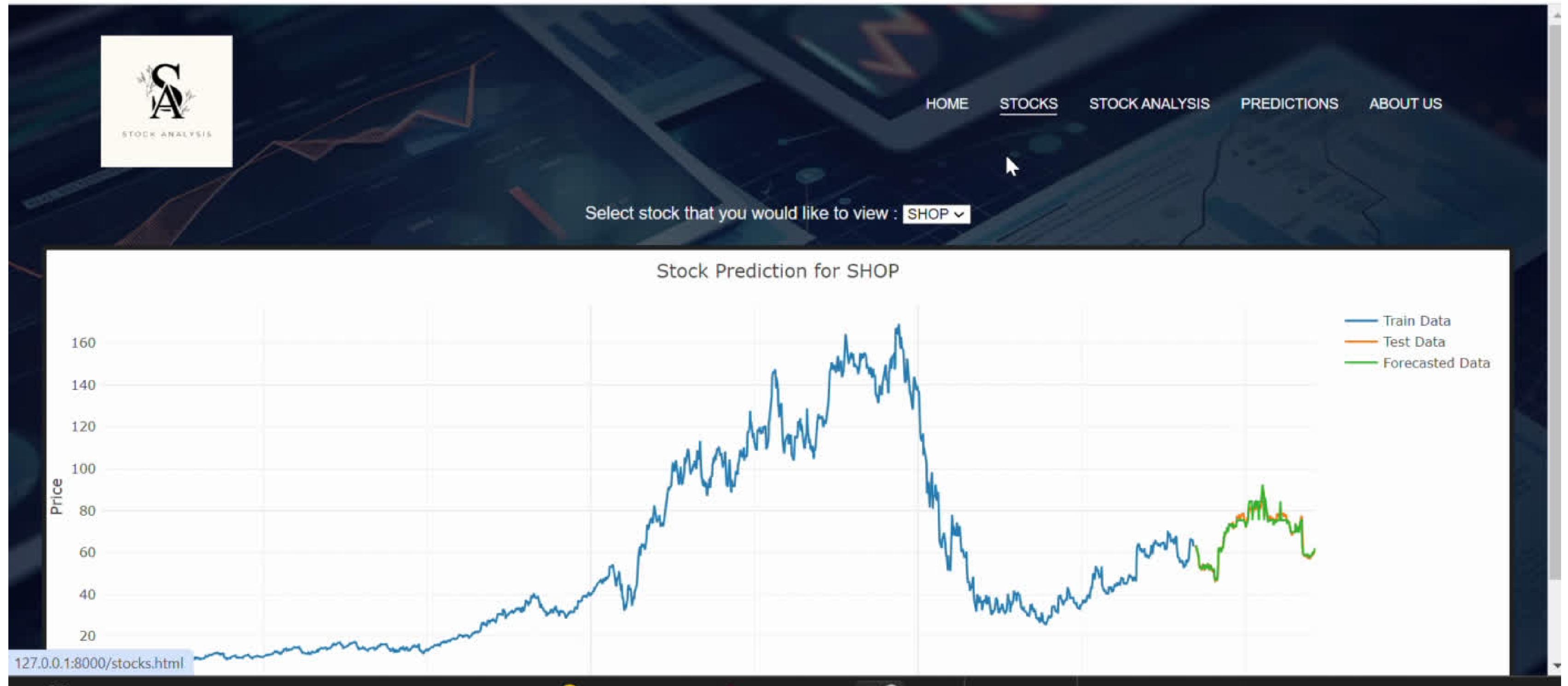


HOME STOCKS STOCK ANALYSIS PREDICTIONS ABOUT US

Stock Analysis University of Toronto

Group Contributors - Jasleen Shergill | Sneha Kumari | Emad Kamli | Vanilla Tse
text for stock analysis

Visit our Github repo



FUTURE SCOPE

Fine-tuning Models: Improve accuracy through hyperparameter tuning

Enhanced Web Interface: Interactive and responsive design improvements

Incorporating Additional Data: Integration of economic indicators and news sentiment

Ensemble Methods: Combine multiple models for enhanced predictive power

Deployment on Cloud: For scalability and accessibility



**THANK
YOU**

For future project, let's connect on LinkedIn!