STOCK CLOSING PRICE PREDICTION USING MACHINE LEARNING TECHNIQUES

Presented By:
Sunakshi Meena
Vishal Singh
Yatharth Nagdeve

CONTENTS

INTRODUCTION

- Stock market prediction is a complex task due to its dynamic, non-linear, and unpredictable nature.
- It is influenced by various factors such as company performance, political events, and economic conditions.
- Two primary methods are used: **Technical analysis** (based on historical stock data) and **Fundamental analysis** (based on qualitative data such as news and market sentiment).
- Recent advancements in machine learning have significantly improved forecasting accuracy. This study applies Artificial Neural Networks (ANN) and Long Short-Term Memory (LSTM) models using technical indicators to predict next-day stock closing prices and evaluates their performance using RMSE, MAPE and MBE.

MOTIVATION

- Traditional linear models (e.g., Linear Regression) lack predictive power in highly non-linear and time series stock data.
- The need for **more accurate and automated** prediction tools that can handle **complex, high-volume financial data.**
- Machine learning offers a promising way to **reduce prediction error** and **aid investors** in making informed decisions.
- Once trained, **ML models** automate forecasting with minimal human intervention, saving time and resources.
- Accurate forecasts help investors manage downside risks more effectively.

OBJECTIVE

Our main objective is to evaluate and compare the effectiveness of **Artificial Neural Networks** (**ANN**) and **Long Short-Term Memory** (**LSTM**) models in **predicting next-day stock closing prices** using technical indicators derived from historical OHLC (Open, High, Low, Close) data.

The study aims to identify which model performs better in terms of **accuracy and bias**, measured using **RMSE, MAPE, and MBE**, across five companies from different sectors and to also study the applicability of machine learning in enhancing prediction accuracy for stock market applications.

RESEARCH METHODS

I. Obtained stock data of 5 different companies which are to be used for stock prediction.

II. Cleaned the data, added new variables and removed any empty (NaN) values.

III. Build and deployed our machine learning models on each of the companies for the stock prediction.

IV. Obtained the values of RMSE, MAPE and MBE.

V. Compared the values to find the best model for the stock value prediction.

Info:

- **RMSE(Root Mean Square Error) - shows average prediction error**
- **MAPE(Mean Absolute Percentage Error) - shows average percentage error**
- **MBE(Mean Bias Error) - shows whether predictions are generally too high or too low**

INFO

1. RMSE (Root Mean Square Error): shows average prediction error

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

2. MAPE(Mean Absolute Percentage Error): shows average percentage error

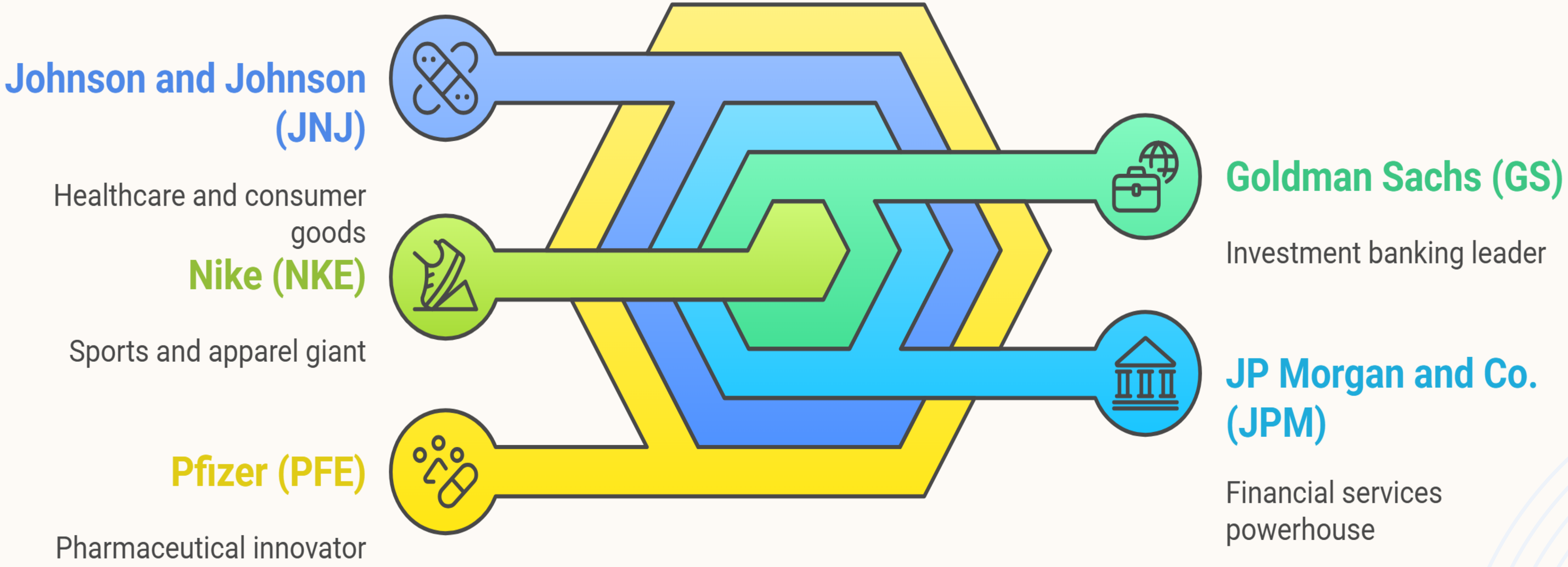
$$\frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| = \text{MAPE}$$

3. MBE(Mean Bias Error) - shows whether predictions are in general too high or too low

$$\text{MBE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)$$

DATA SOURCE

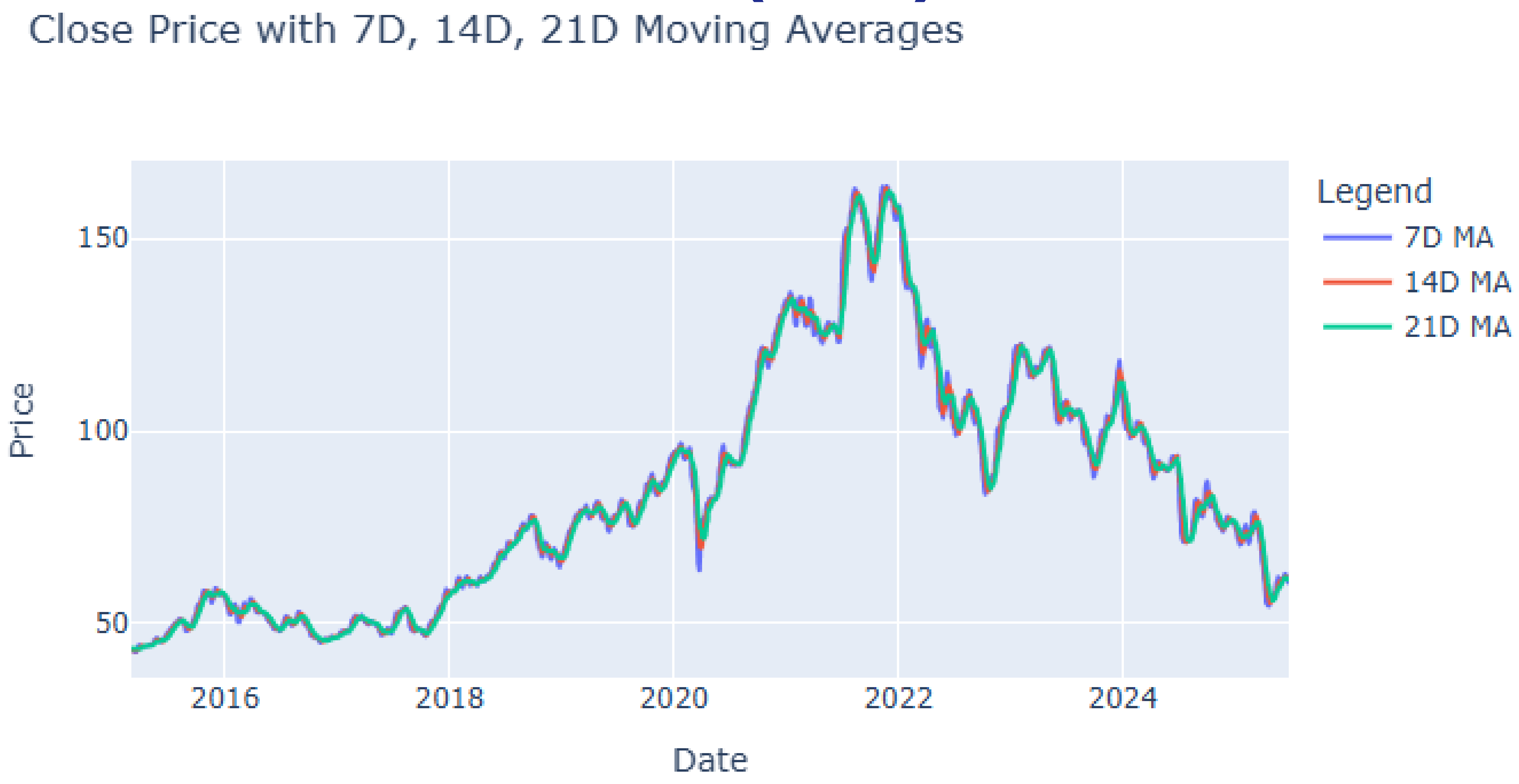
Source: Yahoo Finance(https://finance.yahoo.com/)
Timeline: 2015-2025
Total companies: 5



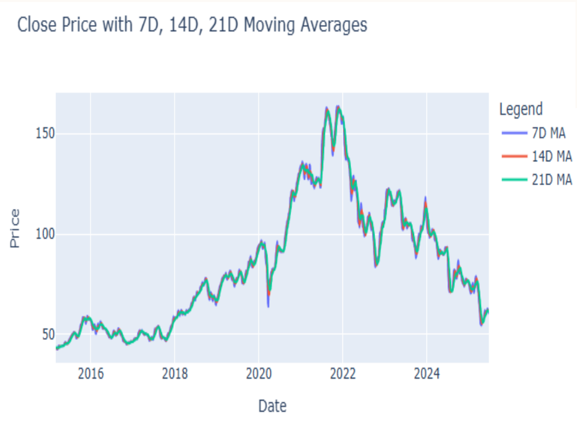
NIKE (NKE)

NKE Close Price Over Time
Close
150
100
50
2016
2018
2020
2022
2024
Date

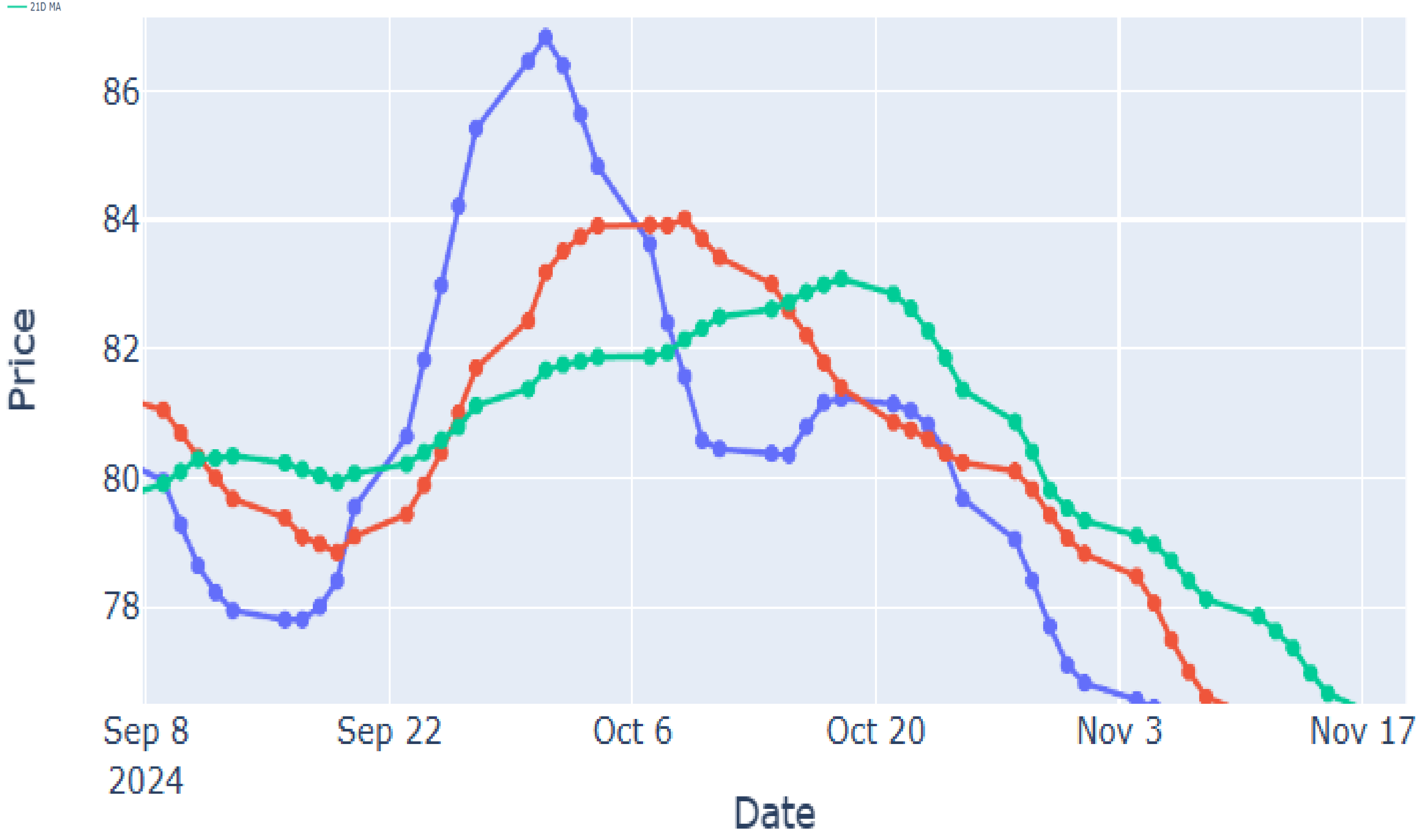
NIKE (NKE)


Close Price with 7D, 14D, 21D Moving Averages
Legend
7D MA
14D MA
21D MA
Price
150
100
50
2016
2018
2020
2022
2024
Date

NIKE (NKE)



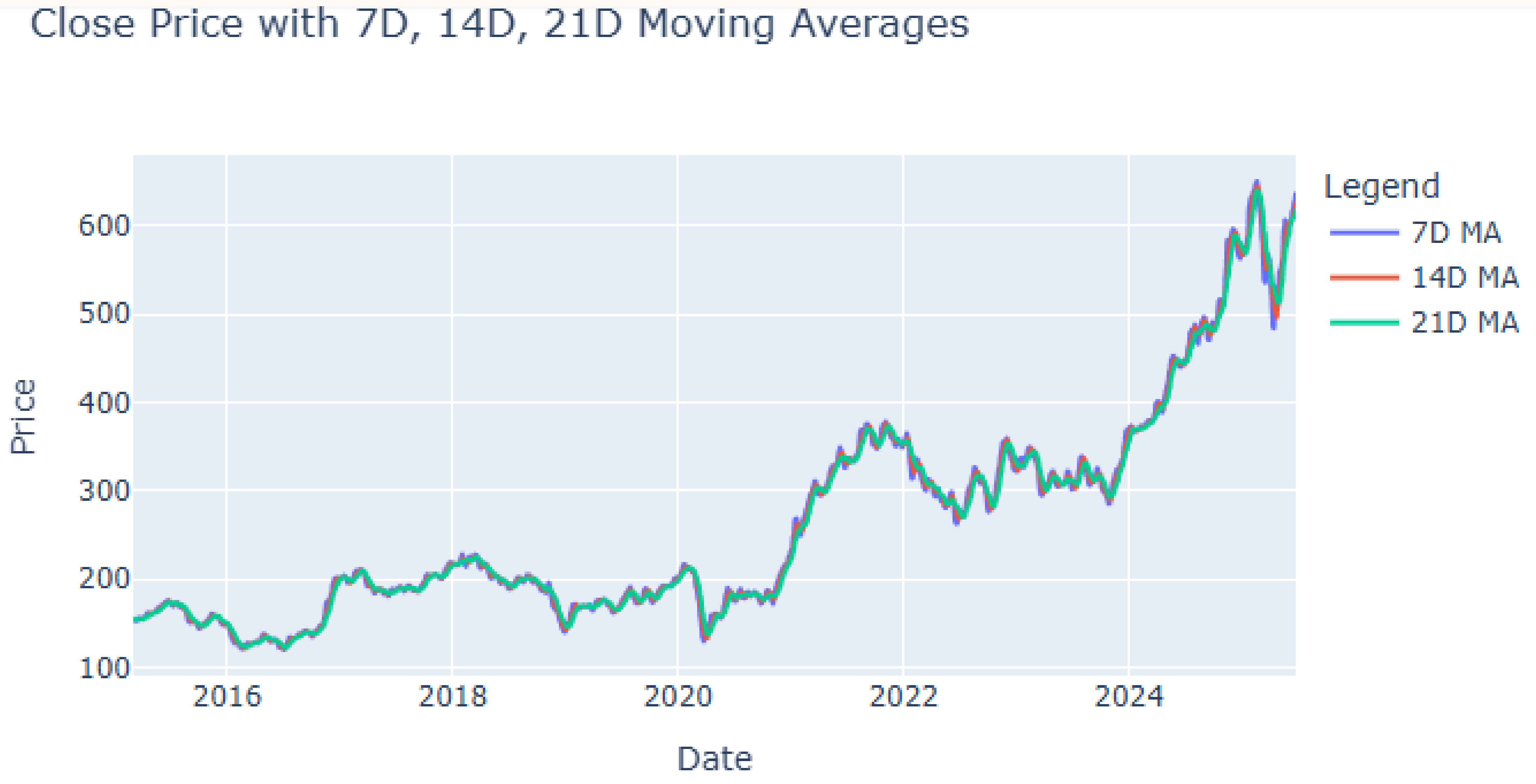
Close Price with 7D, 14D, 21D Moving Averages
Legend
7D MA
14D MA
21D MA
150
100
50
Price
2016
2018
2020
2022
2024
Date



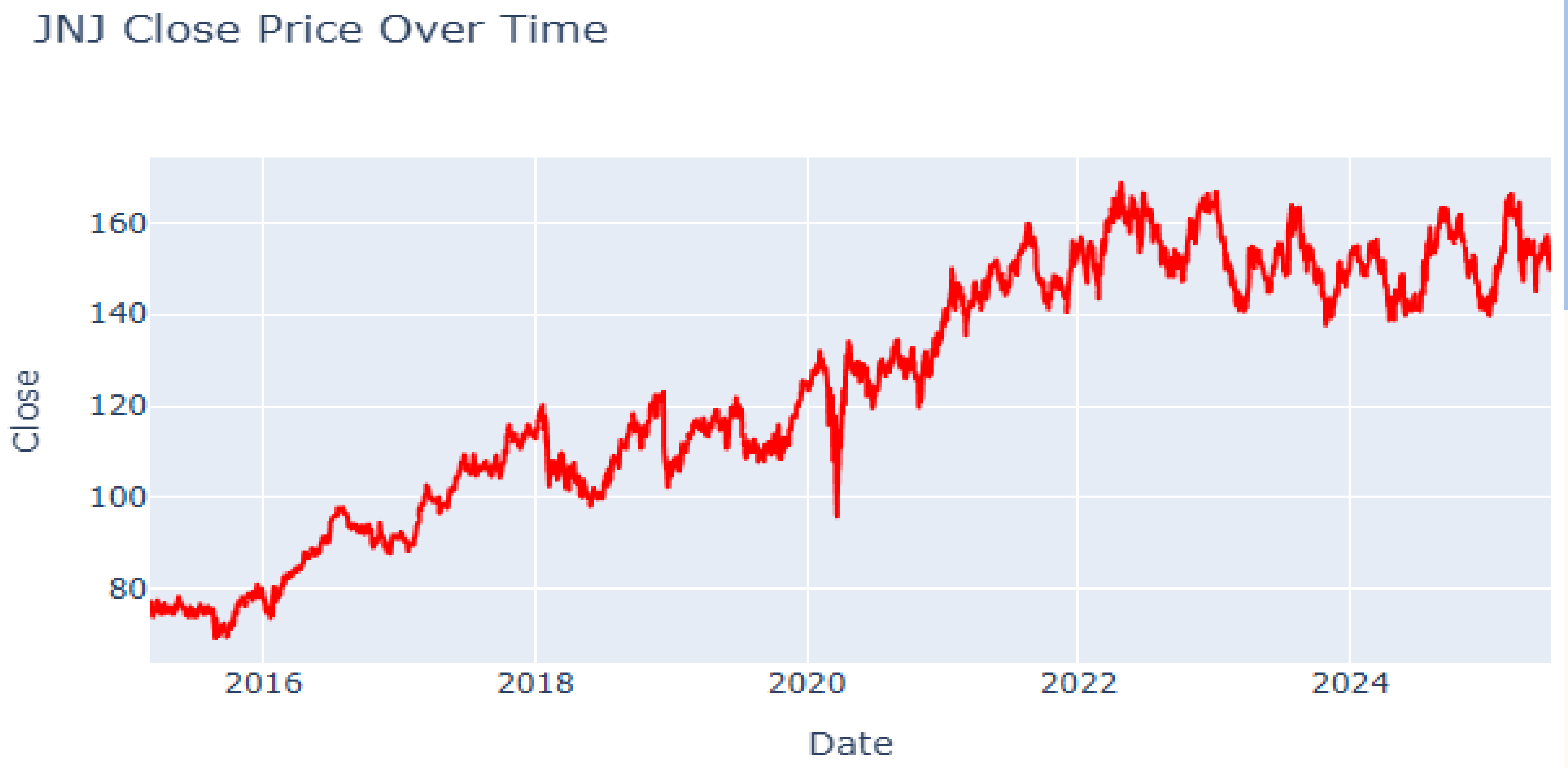
86
84
82
80
78
Price
Sep 8
2024
Sep 22
Oct 6
Oct 20
Nov 3
Nov 17
Date

GOLDMAN SACHS (GS)

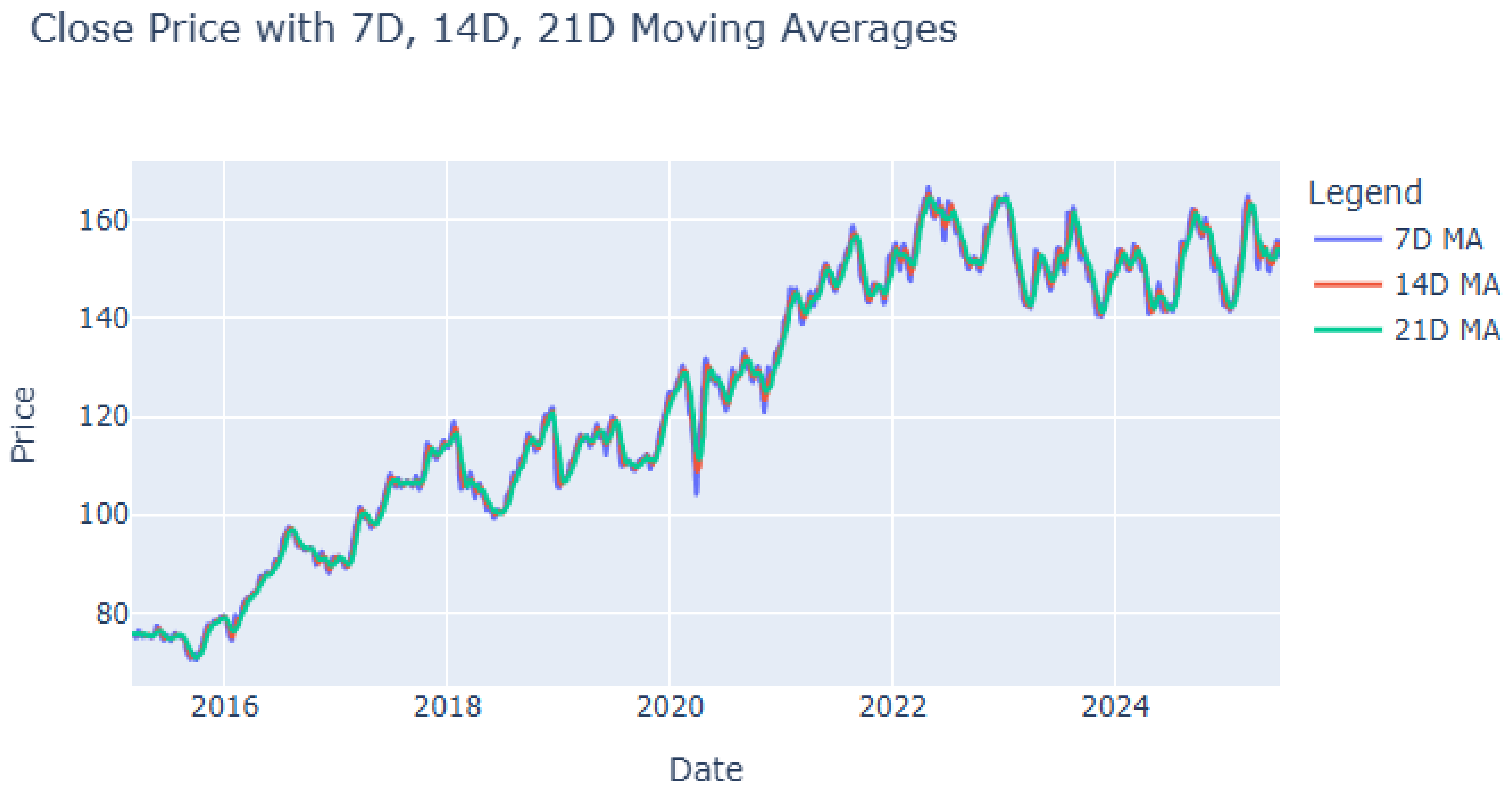
GOLDMAN SACHS (GS)



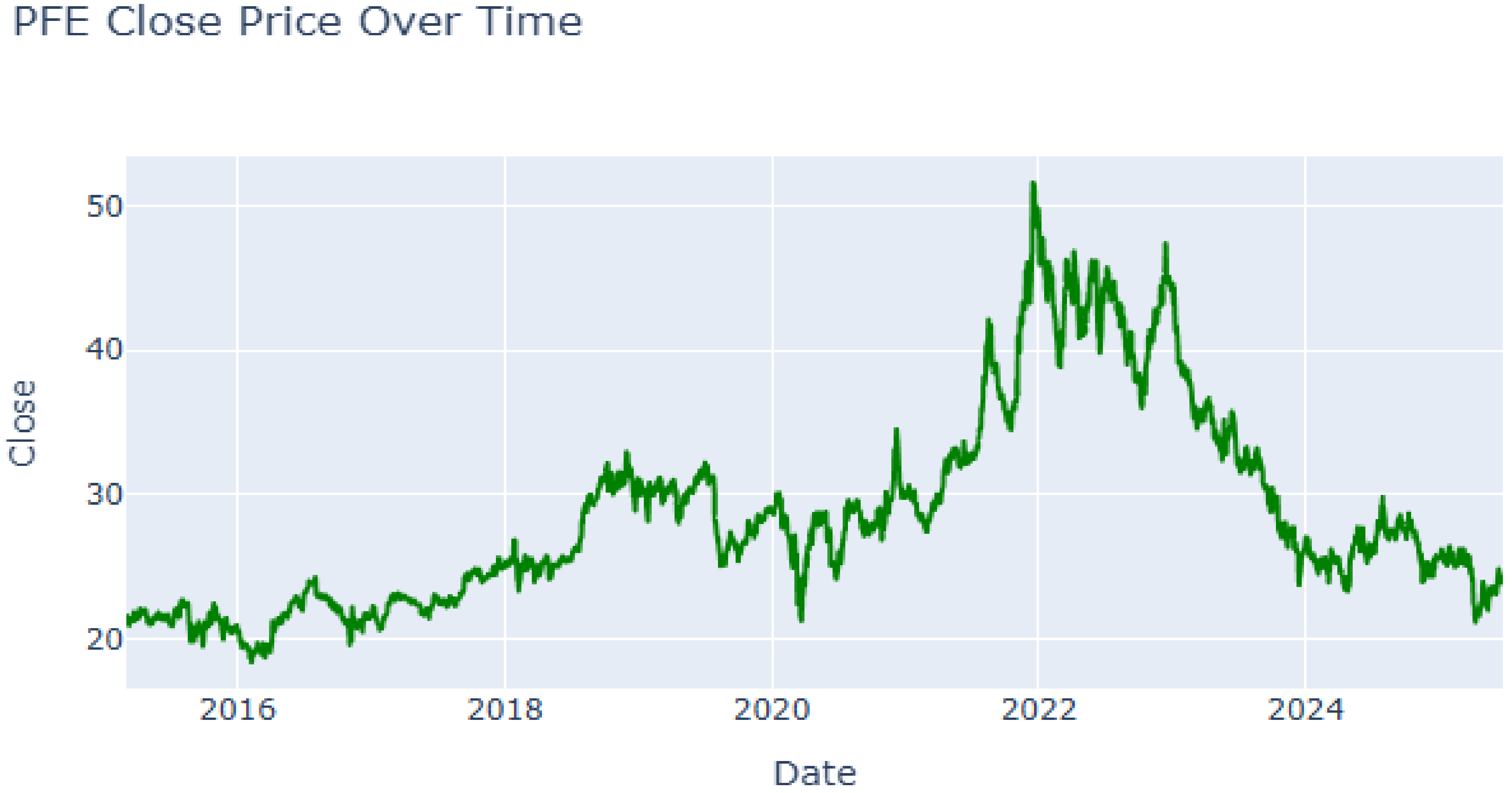
JOHNSON AND JOHNSON (JNJ)



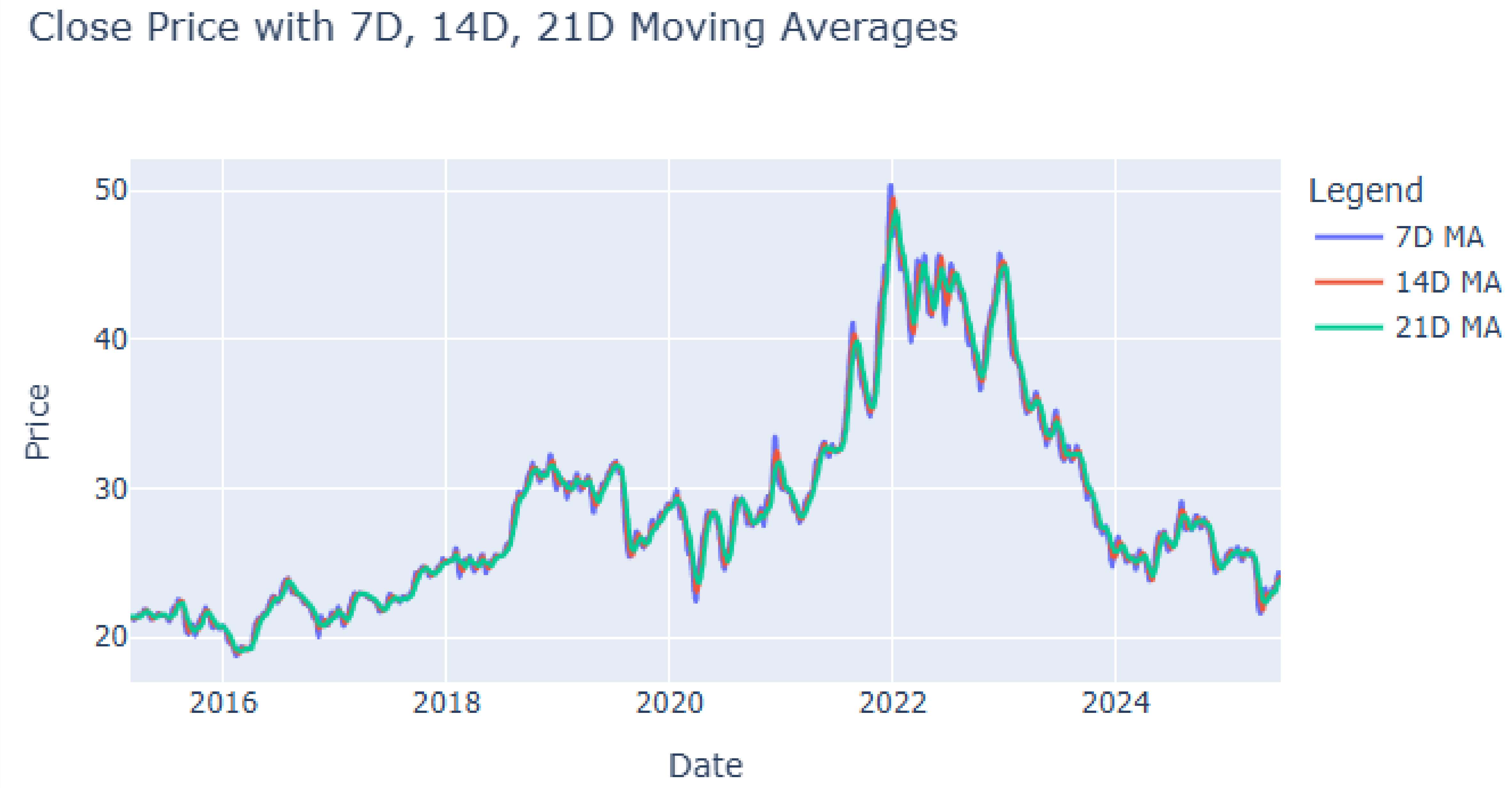
JOHNSON AND JOHNSON (JNJ)



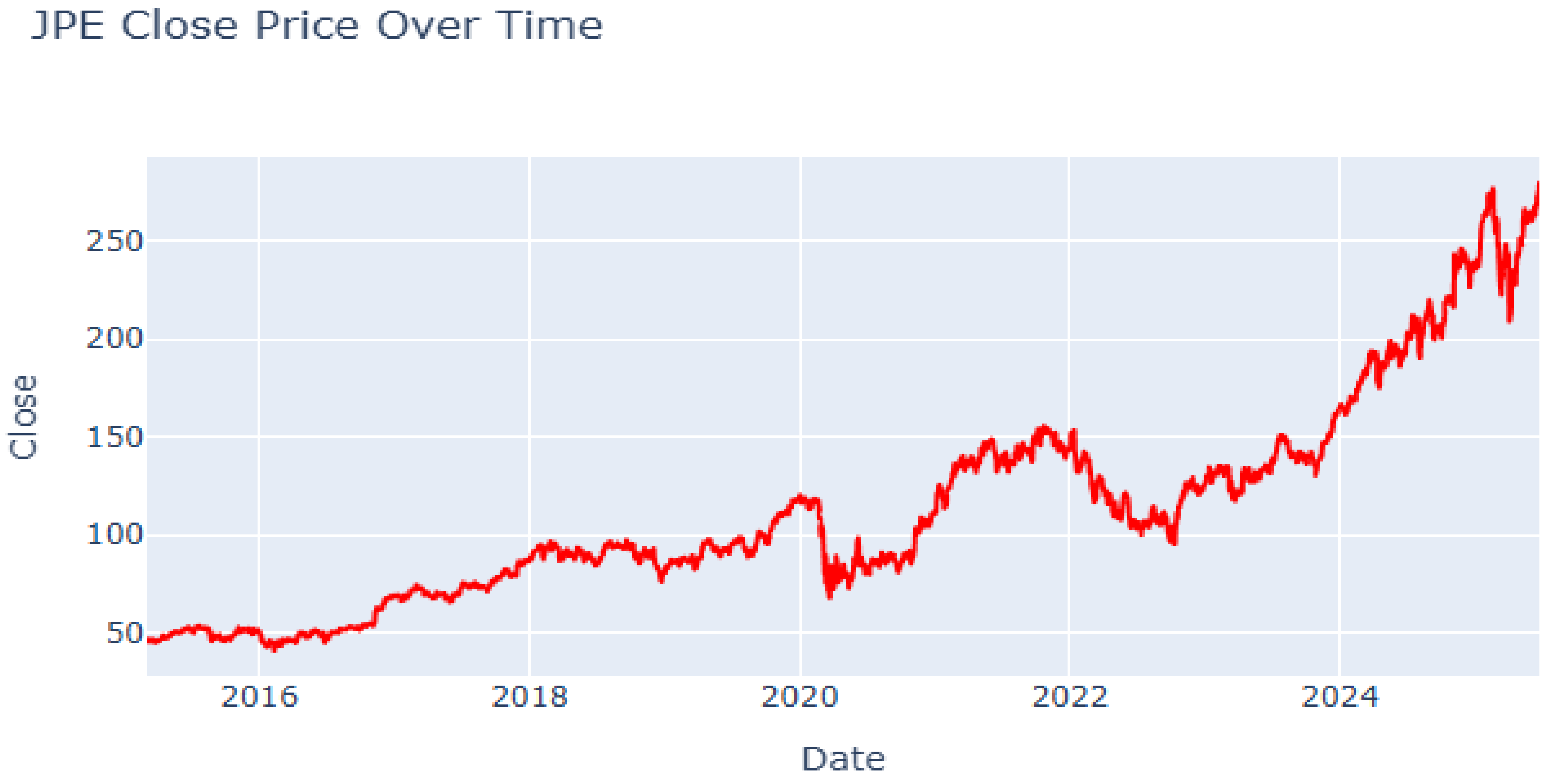
PFIZER (PFE)


PFE Close Price Over Time
50
40
30
20
Close
2016
2018
2020
2022
2024
Date

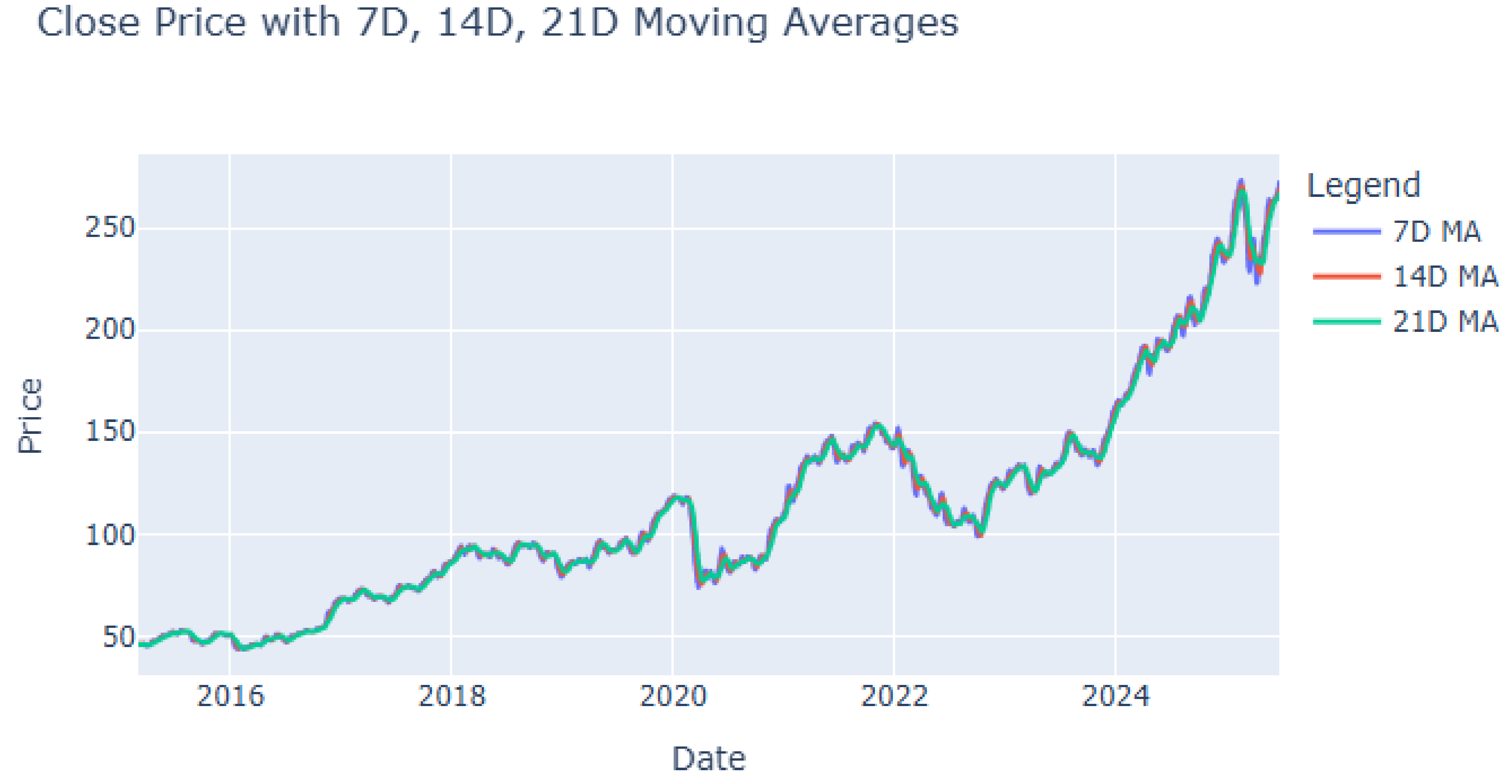
PFIZER (PFE)


Close Price with 7D, 14D, 21D Moving Averages
Legend
7D MA
14D MA
21D MA
Price
50
40
30
20
2016
2018
2020
2022
2024
Date

JP MORGAN AND C0. (JPE)



JP MORGAN AND C0. (JPE)



DATA AND FEATURES IN MODEL

Timeline of Stock data : **2015 to 2025**.

Training Data : **2015 to 2023**

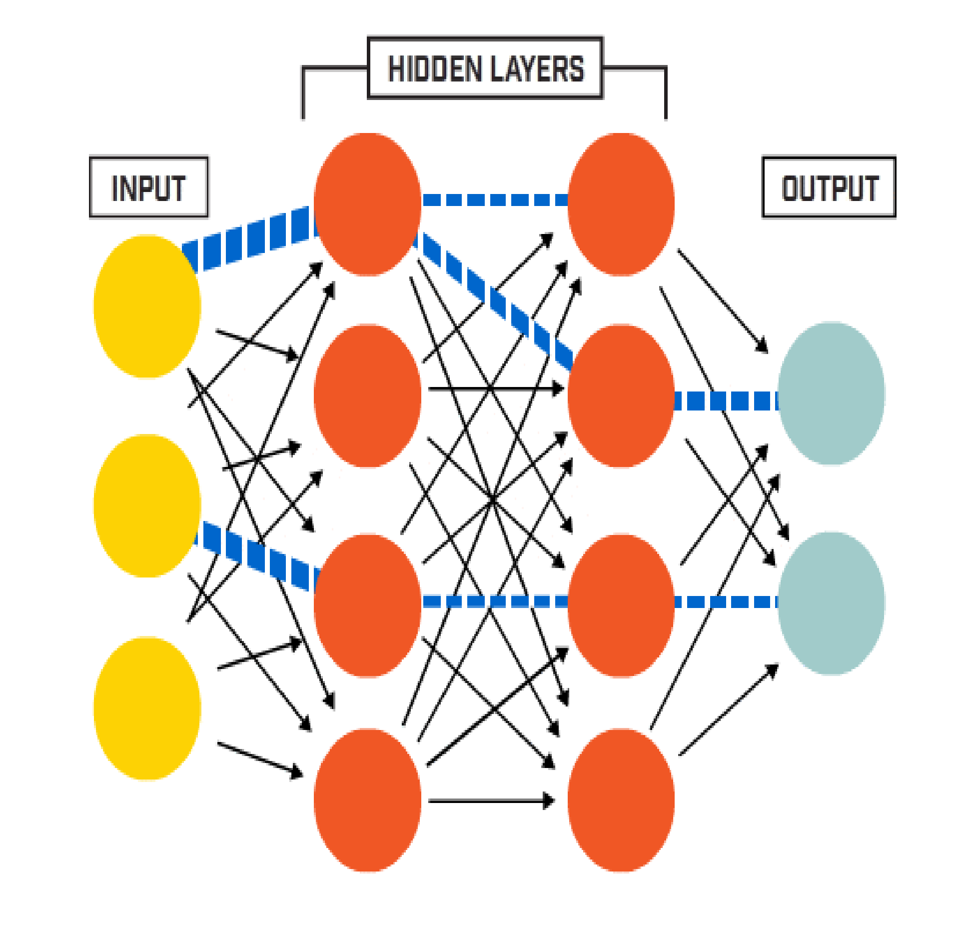
Testing Data : **2024 to 2025**

Features : Close (t-1), H-L, O-C, 7D MA, 14D MA, 21D MA, 7D STD DEV, Volume

Target : Close (t)

MinMaxScaler : It transforms features to a fixed range, usually [**0, 1**]. This prevents features with large values from dominating those with smaller values.

ARTIFICIAL NEURAL NETWORK (ANN) MODEL



- **Inspired by the human brain** with layers of interconnected neurons.
- **Takes inputs**, multiplies them by **weights**, adds a **bias**, and applies an **activation function**.
- **Layers**: Input layer, hidden layers, and output layer.
- **Learns from data** by adjusting weights using **backpropagation** and **gradient descent**.
- **Minimizes error** between actual and predicted output.
- **Capable of capturing complex patterns**, making it suitable for **stock prediction**, image processing, and classification tasks.

ANN MODEL

We have used a **Sequential** model with 2 hidden layers.

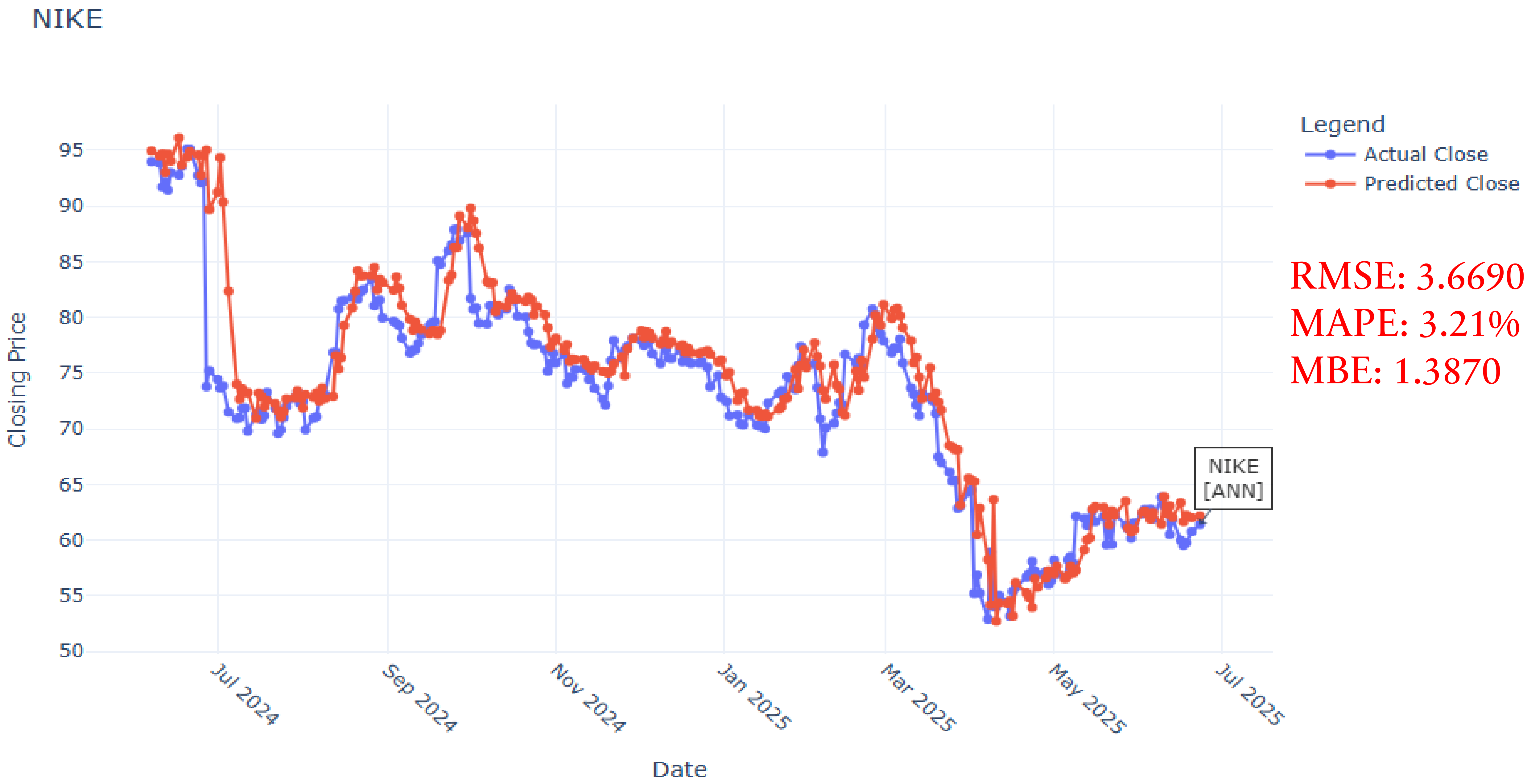
Dense (64) with ReLU activation
Dense (32) with ReLU activation

Optimizer: Adam with learning rate of 0.1% and loss of mean squared error

Batch Size: 32 (trained data split into 32 batches)

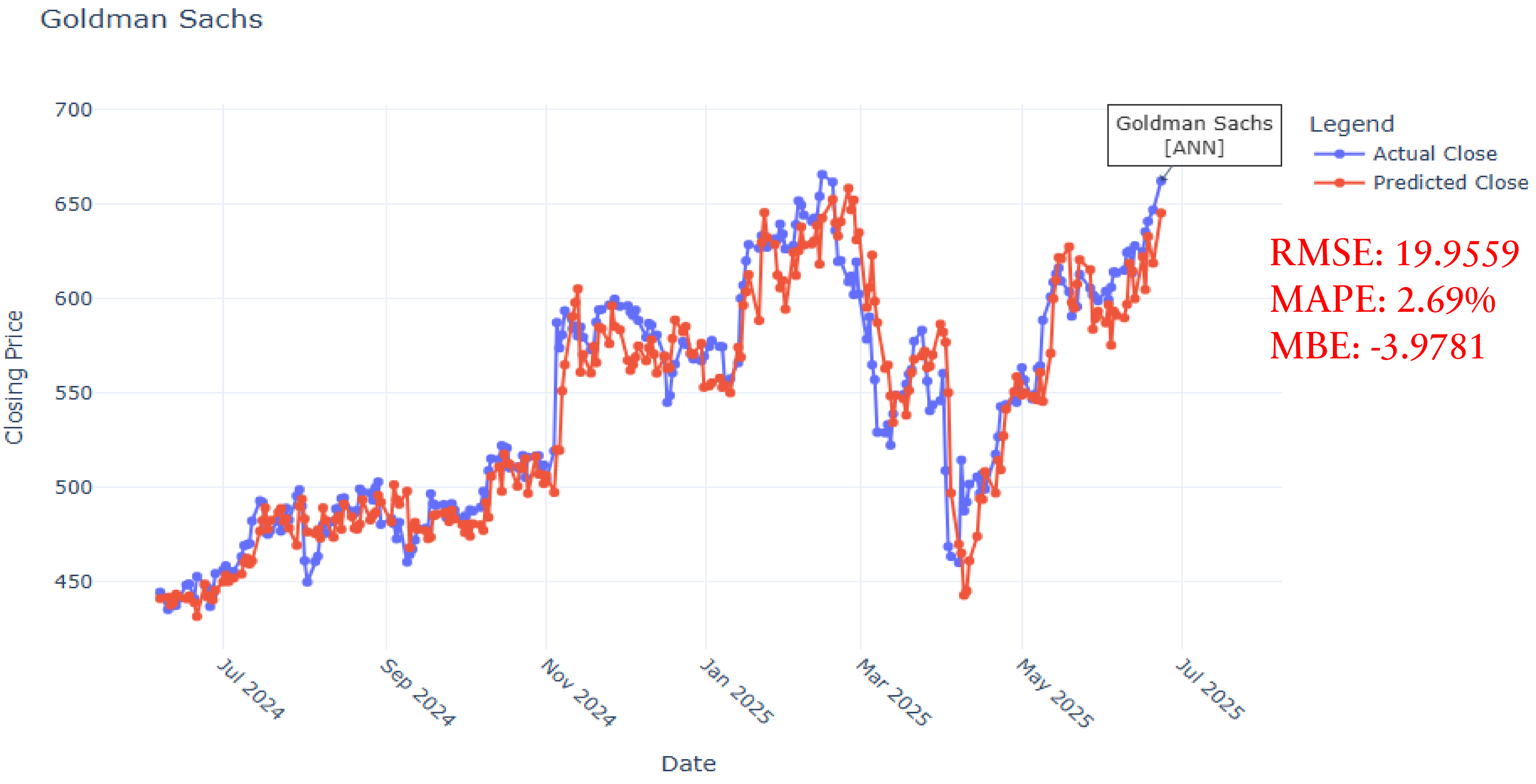
Epoch : 100 i.e., model go through entire dataset 100 times

NIKE (NKE)



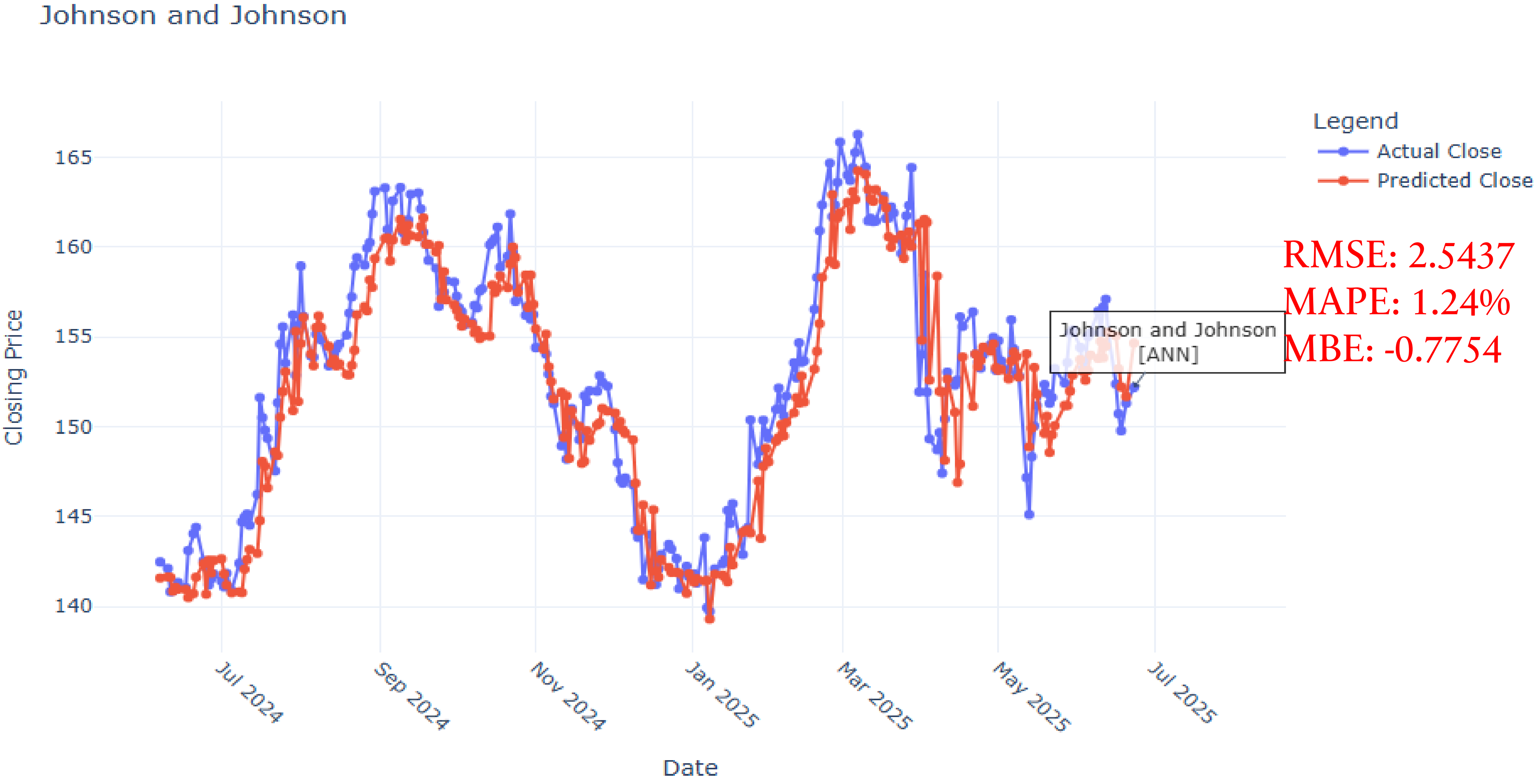
RMSE: 3.6690
MAPE: 3.21%
MBE: 1.3870

GOLDMAN SACHS (GS)



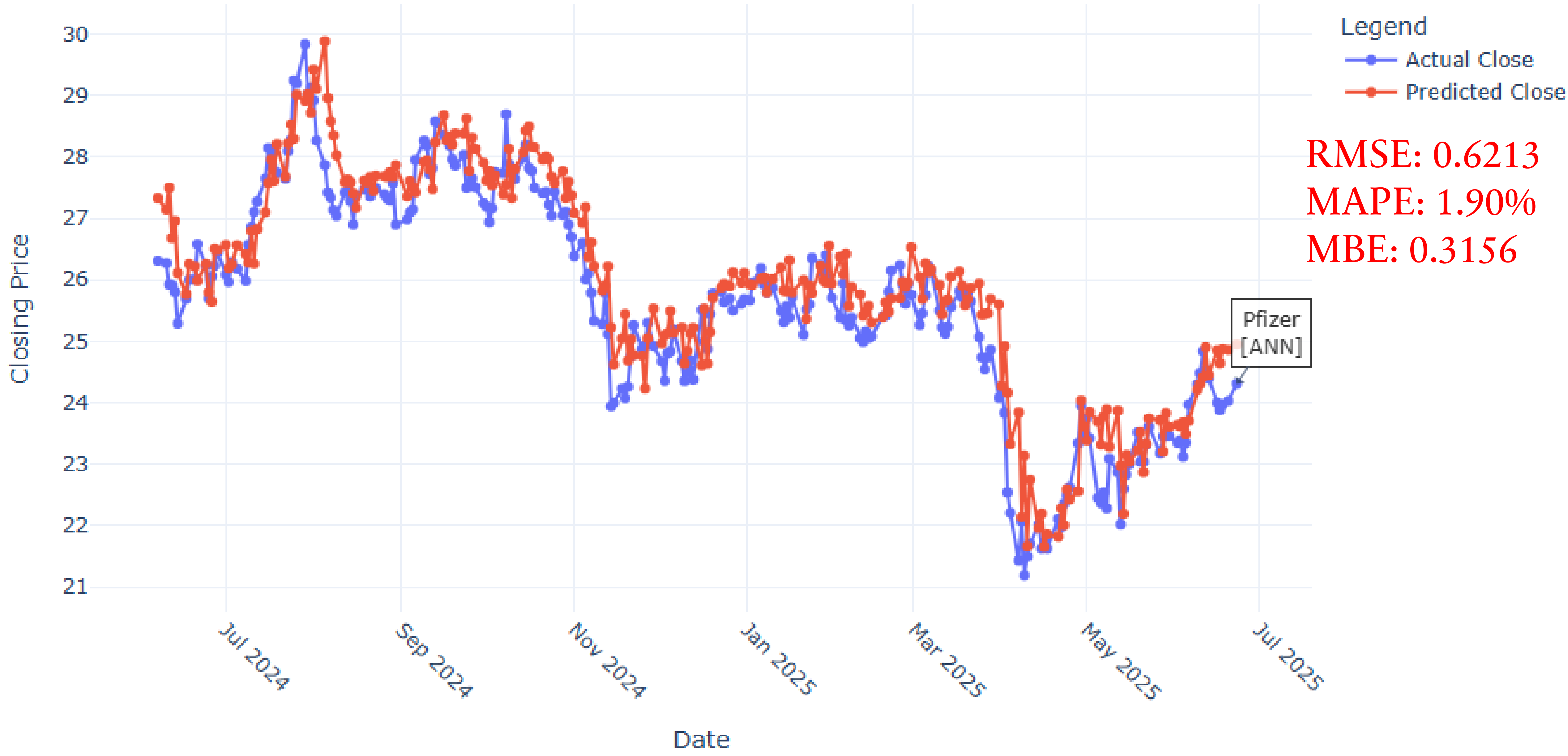
RMSE: 19.9559
MAPE: 2.69%
MBE: -3.9781

JOHNSON AND JOHNSON (JNJ)



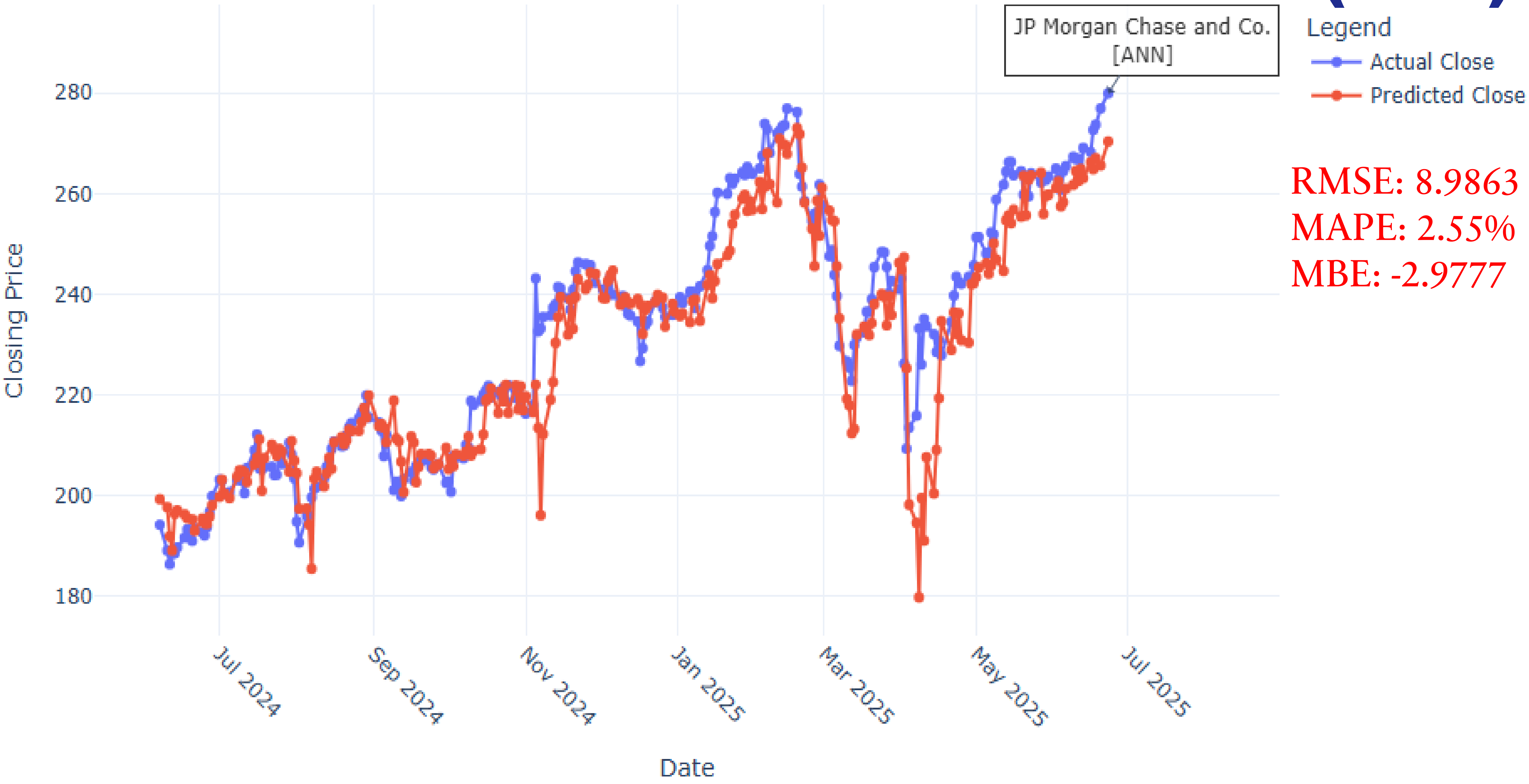
RMSE: 2.5437
MAPE: 1.24%
MBE: -0.7754

PFIZER (PFE)



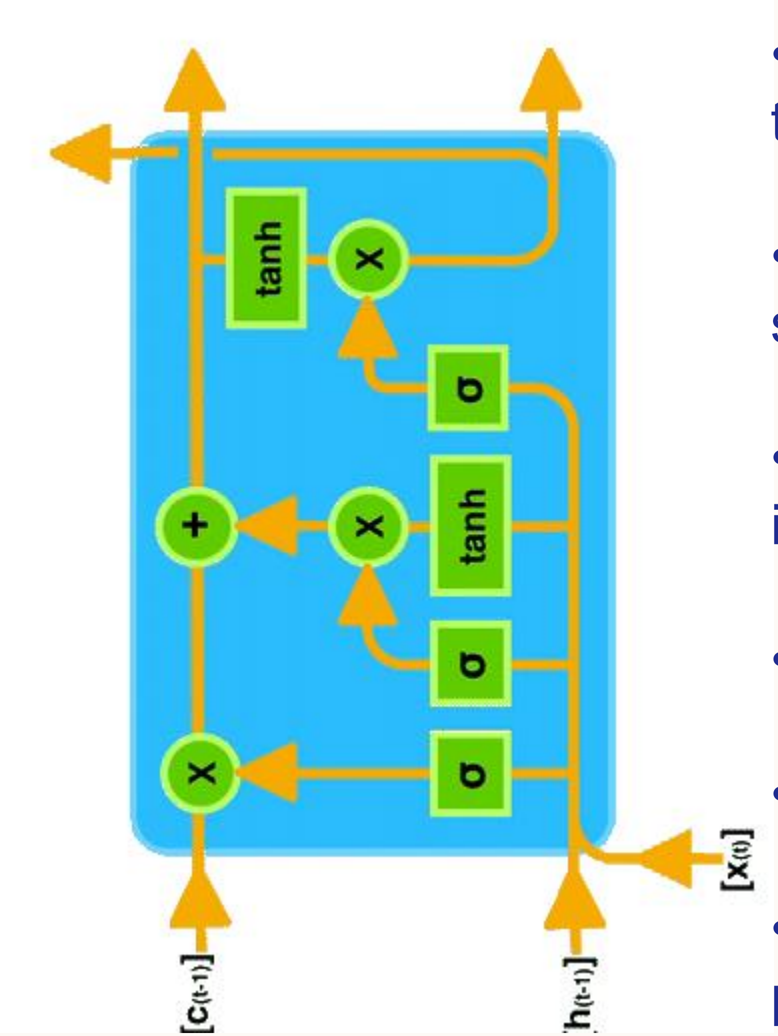
RMSE: 0.6213
MAPE: 1.90%
MBE: 0.3156

JP MORGAN (JPM)



RMSE: 8.9863
MAPE: 2.55%
MBE: -2.9777

LONG SHORT-TERM MEMORY (LSTM)



- A type of **Recurrent Neural Network (RNN)** designed to learn from sequential data.
- Maintains memory over longtime steps using special cell states.
- Uses gates (**input, forget, output**) to control the flow of information.
- Excellent for time-series forecasting, like stock prices.
- Learns temporal patterns and dependencies in data.
- Trained using backpropagation through time to minimize prediction error.

LSTM MODEL

We have used a **Sequential** model.

LSTM units (i.e., memory cells): 64
Look back period : 90
Activation : ReLU

Optimizer: Adam with loss of mean squared error

Batch Size: 32 (trained data split into 32 batches)

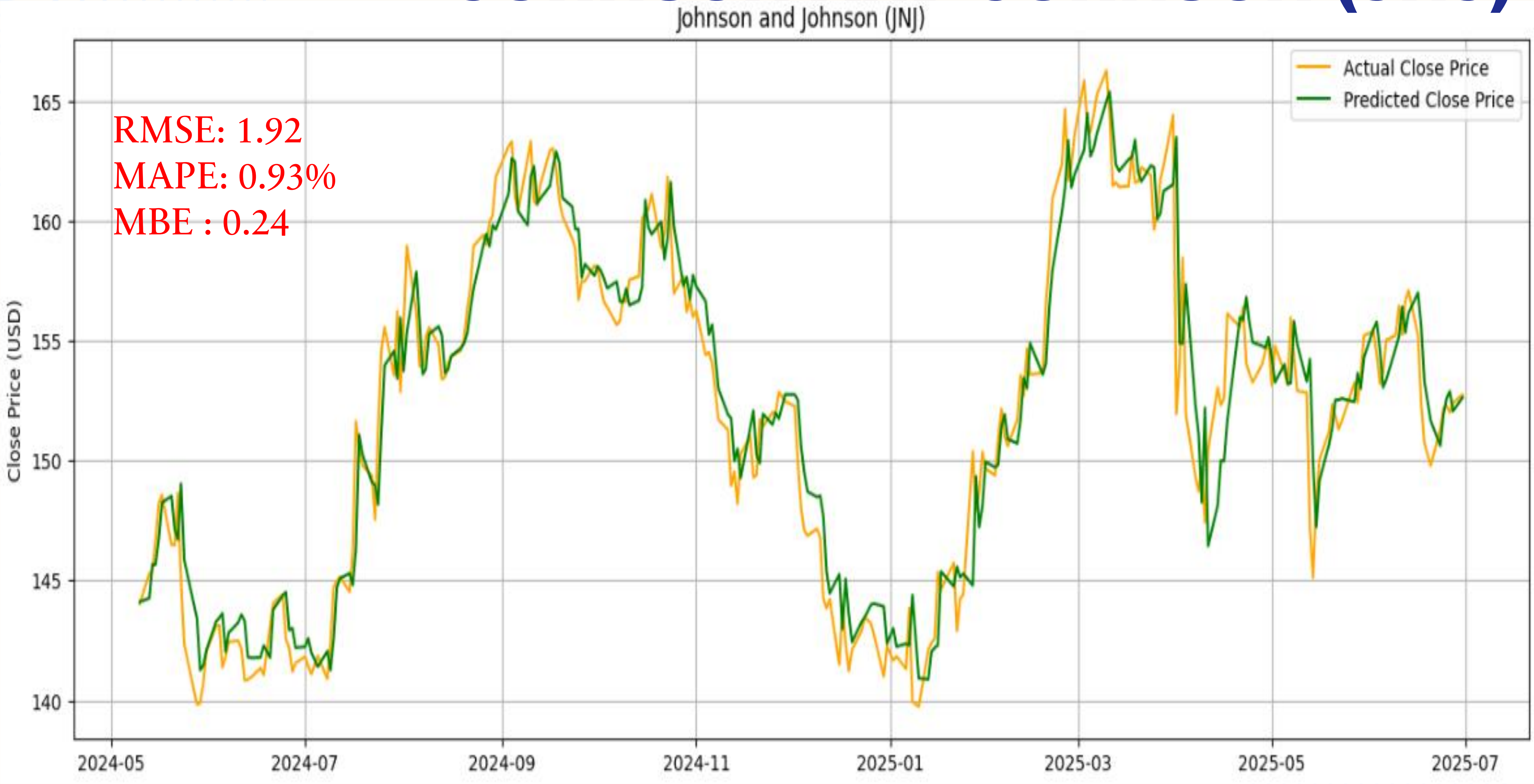
Epoch : 50 i.e., model go through entire dataset 50 times

NIKE (NKE)

RMSE: 2.21
MAPE: 1.94%
MBE : 0.60

GOLDMAN SACHS (GS)

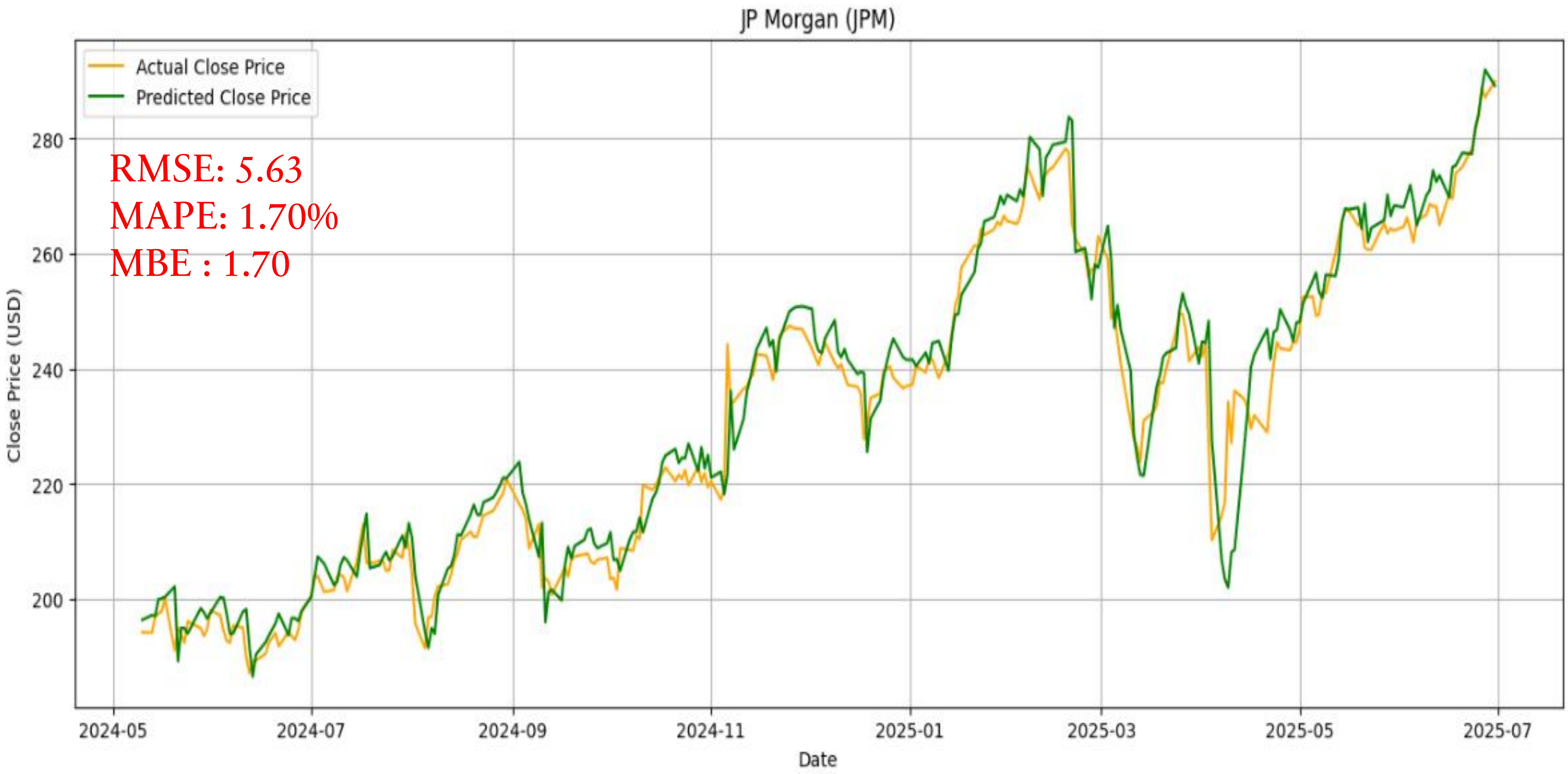
JOHNSON AND JOHNSON (JNJ)



PFIZER (PFE)

RMSE: 0.41
MAPE: 1.23%
MBE : 0.07

JP MORGAN (JPM)



COMPARISON B/W ANN AND LSTM MODEL

Company	ANN			LSTM		
	RMSE	MAPE	MBE	RMSE	MAPE	MBE
Nike	3.67	3.21%	1.38	2.21	1.94%	0.60
Goldman Sachs	19.95	2.69%	-3.98	13.45	1.73%	-4.80
Johnson & Johnson	2.54	1.24%	-0.78	1.92	0.93%	0.24
Pfizer Inc.	0.62	1.90%	0.32	0.41	1.23%	0.07
JP Morgan and Co.	8.99	2.55%	-2.98	5.63	1.70%	1.70

CONCLUSION

Based on the comparative performance metrics (RMSE, MAPE, and MBE) across five companies, the **LSTM model consistently outperforms the ANN model** in stock price prediction:

- **LSTM exhibits lower RMSE and MAPE values** for all companies, indicating higher predictive accuracy and reduced percentage error.
- The **MBE values** from the LSTM model are generally closer to zero, reflecting more **unbiased predictions**.
- Notably, for firms like **Goldman Sachs and JP Morgan**, the LSTM model delivers **substantially better results**, highlighting its effectiveness in capturing temporal dependencies inherent in financial time series data.

In conclusion, the **LSTM model demonstrates superior overall performance** and is better suited for stock price forecasting tasks compared to a standard ANN model.

FUTURE ENDEAVOURS

- **Hybrid Model Development**
  Combining LSTM with CNN or Attention-based mechanisms to enhance feature extraction and capturing complex temporal dependencies.

- **Multi-Stock & Portfolio-Level Forecasting**
  Expanding the scope to predict multiple correlated stocks simultaneously, supporting portfolio optimization and risk management.

- **Incorporate Alternative Data Sources**
  Enhancing model accuracy by integrating news sentiment, social media signals, and macroeconomic indicators.

- **Robustness to Market Anomalies**
  Training the model to adapt to unexpected events (e.g., economic crises, earnings announcements) using regime-switching models.

RESOURCES

1. Research Paper (link)
2. Yahoo finance (link)
3. LSTM Time series (link)
4. Price Prediction using LSTM (link)
5. Google, OpenAI, Claude AI, Napkin AI, Wikipedia

CODE LINK:

Colab.link (1. ANN Model)

Colab.link (2. LSTM Model)

GitHub link (click here)

THANK YOU