Predicting Apple Stock Prices

Introduction to Problem & Data

Problem Statement

Apple Inc., as one of the most valuable companies globally, is closely monitored by investors and analysts alike. Accurately predicting its stock prices can provide key insights into market behavior and assist in making informed financial decisions. For this project, the focus is on developing a robust time series predictive model to forecast Apple's daily closing stock prices. By analyzing historical trends from 2014 to 2023, the project seeks to identify patterns, seasonality, and key factors influencing price movements.This project will compare the performance of models like ARIMA, Prophet, and Long Short-Term Memory (LSTM), alongside simpler methods like Exponential Smoothing and Naive Forecaster. By evaluating these models with metrics such as mean squared error (MSE) and mean absolute error (MAE), the project aims to determine the best-performing approach for forecasting stock prices and generating actionable insights. This model may be useful for individual investors and financial institutions aiming to make informed investment decisions and optimize their portfolios. These predictions can help traders identify potential opportunities for buying or selling Apple stocks, ultimately maximizing their returns while mitigating risks. By leveraging key financial indicators and advanced forecasting methods, the model will not only improve predictive capabilities for stock prices but can also assist stakeholders in the financial industry with actionable insights.

Dataset Description

The data for this project is sourced from Kaggle in CSV format, providing comprehensive daily stock price information for Apple Inc. from January 2014 to December 2023. This dataset includes key financial metrics as well as derived technical indicators commonly used in stock analysis. While the dataset is rich in features, some preprocessing will be necessary to handle missing values, remove unnecessary columns, and prepare the data for time series modeling.

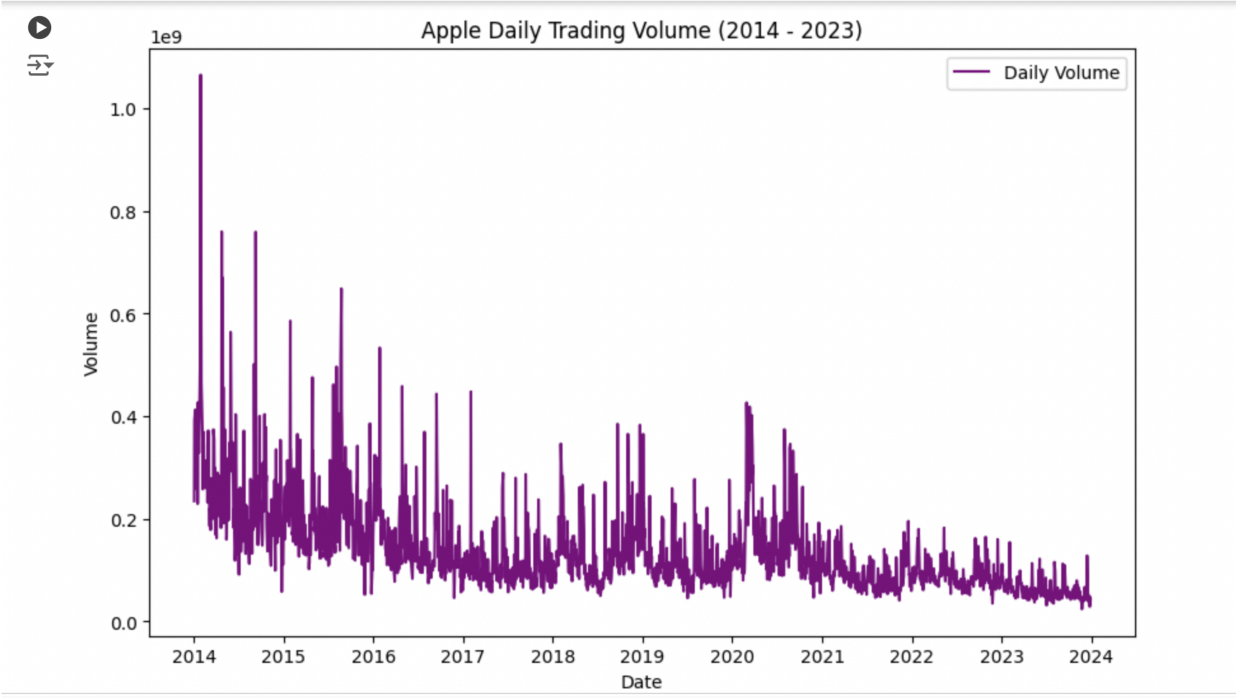
Challenges in constructing an accurate predictive model may arise due to the inherent volatility of financial markets and the noise in stock price movements. However, I believe that with appropriate modeling techniques and the inclusion of technical indicators, it is possible to capture temporal patterns and achieve reasonable accuracy in forecasting.

The dataset comprises 2,516 rows, with each row representing a trading day. It includes key attributes such as: Core Metrics: Daily open, high, low, and close prices, as well as trading volume. Technical Indicators: Features like RSI (Relative Strength Index), CCI (Commodity Channel Index), SMA (Simple Moving Average), EMA (Exponential Moving Average), MACD (Moving Average Convergence Divergence), Bollinger Bands, ATR (Average True Range), and True Range.

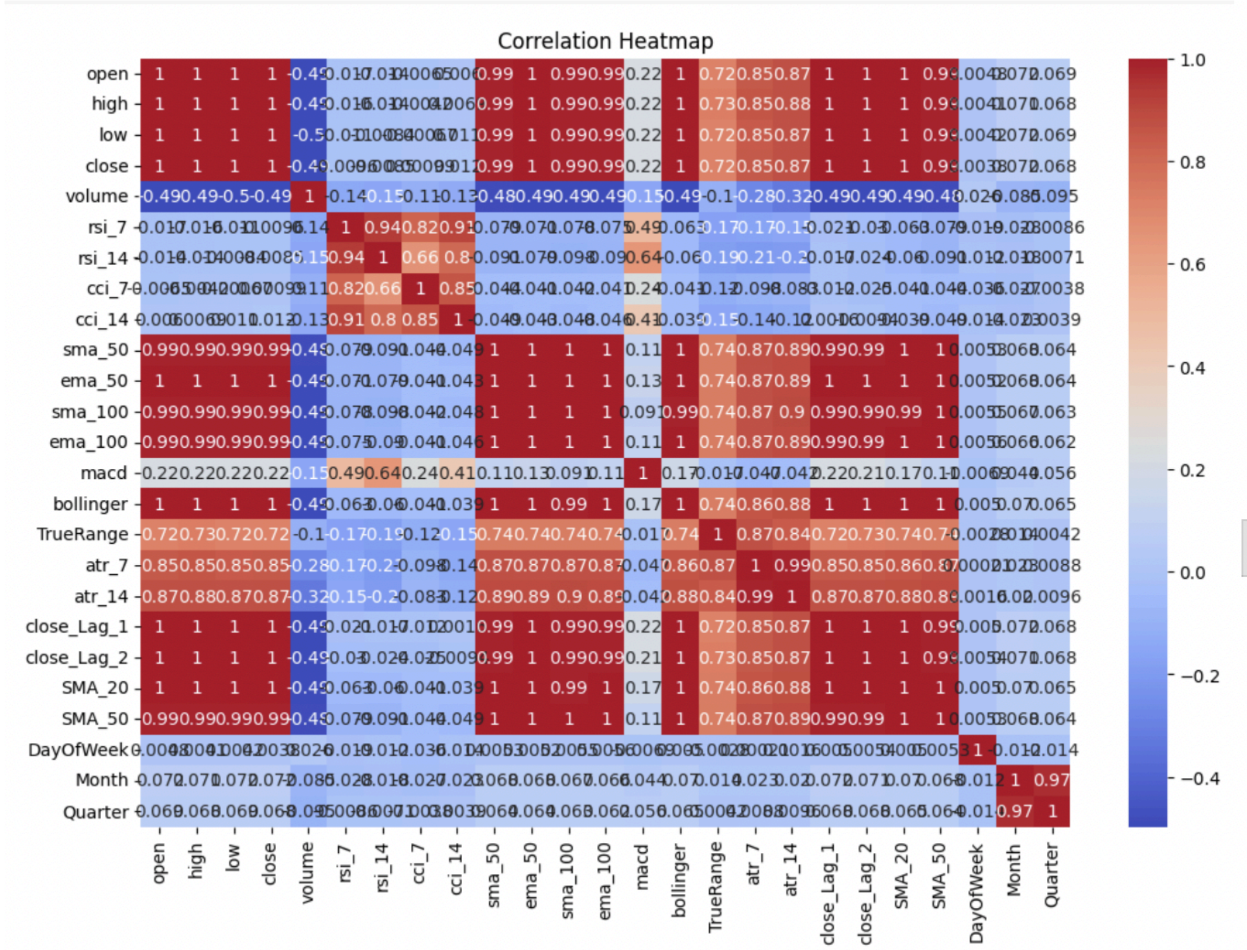
These features will be used to train and evaluate models for predicting Apple's future closing prices and analyzing price trends.

Exploratory Data Analysis (EDA)

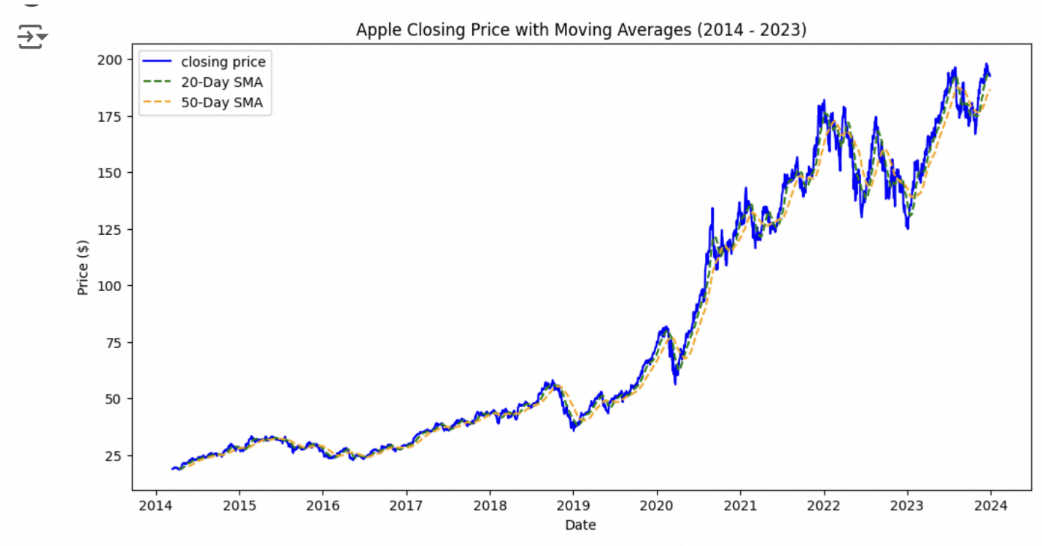
This graph shows the steady upward trend in Apple's closing stock price from 2014 to 2023, with notable spikes and dips, particularly around 2020, reflecting periods of increased volatility potentially linked to market events. Overall, the long-term trend indicates significant growth in value over this period.



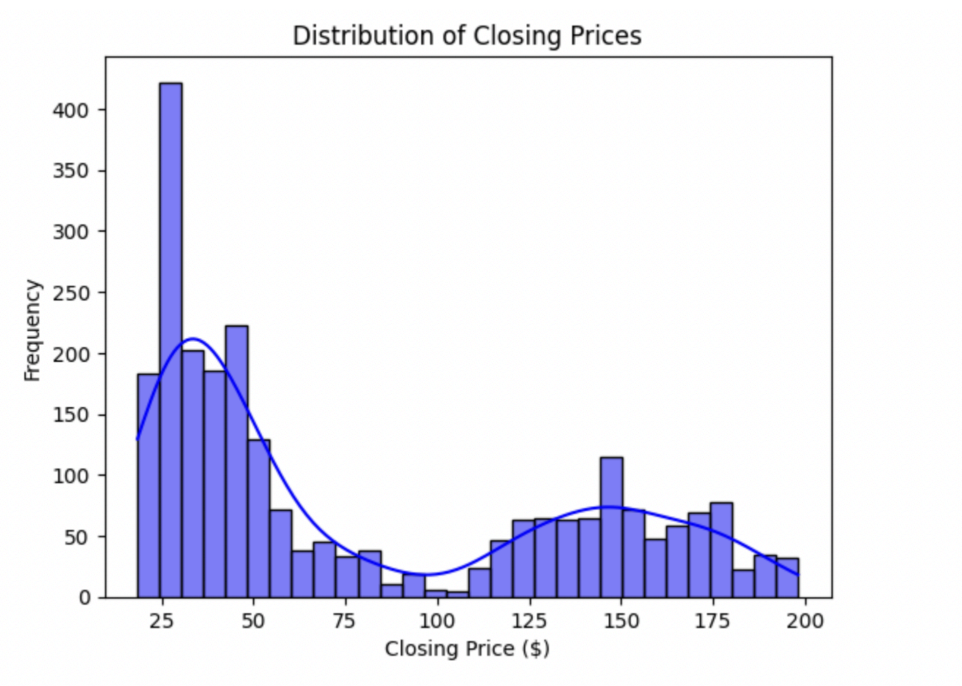
This graph shows Apple's daily trading volume from 2014 to 2023, with a clear decline in volume over time. The early years exhibit significant spikes, suggesting periods of high trading activity, while recent years indicate lower and steadier trading volumes, potentially reflecting changing investor behaviors or market dynamics.



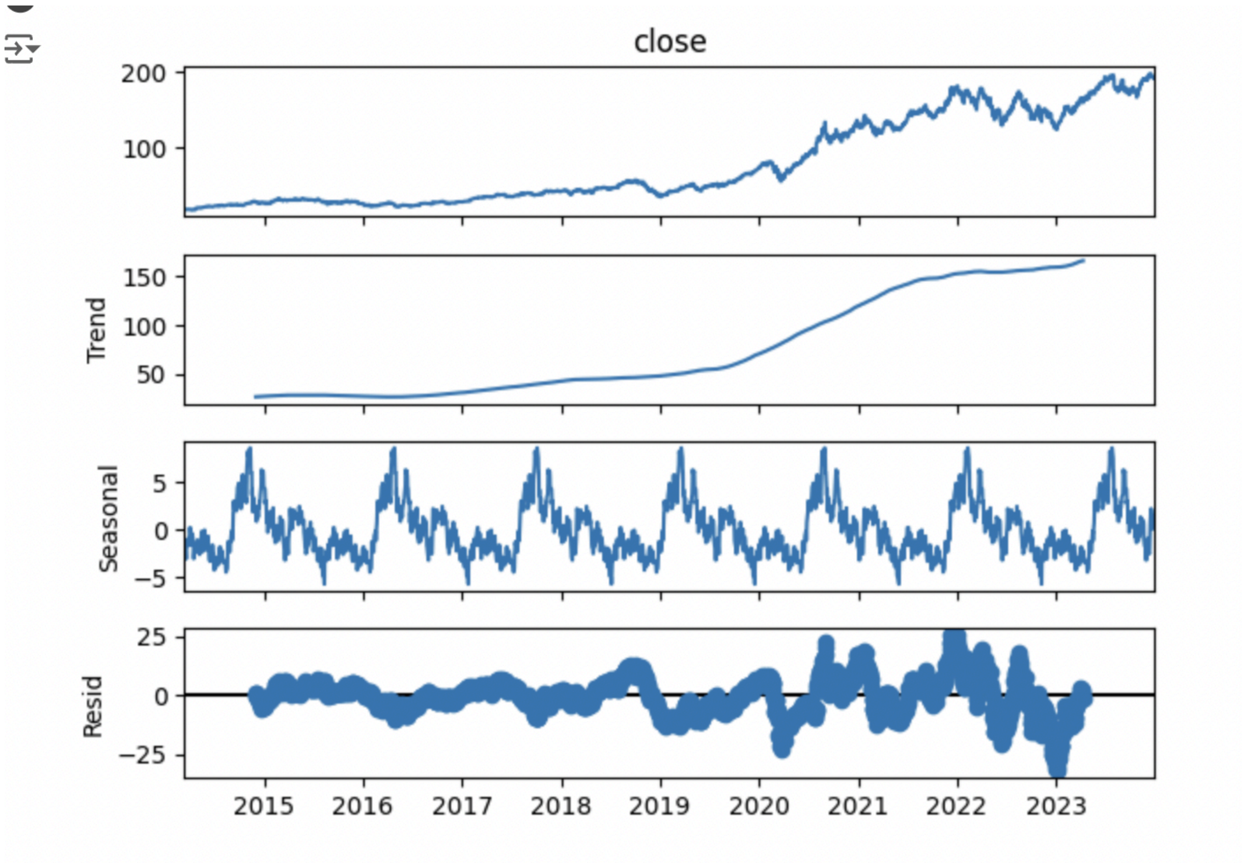
The heatmap shows strong positive correlations among features derived from the closing price, such as moving averages (SMA_20, SMA_50, close_Lag_1), reflecting their dependence on the underlying price. Volume exhibits a weaker correlation with closing price-related features, indicating it may provide additional, less redundant information.



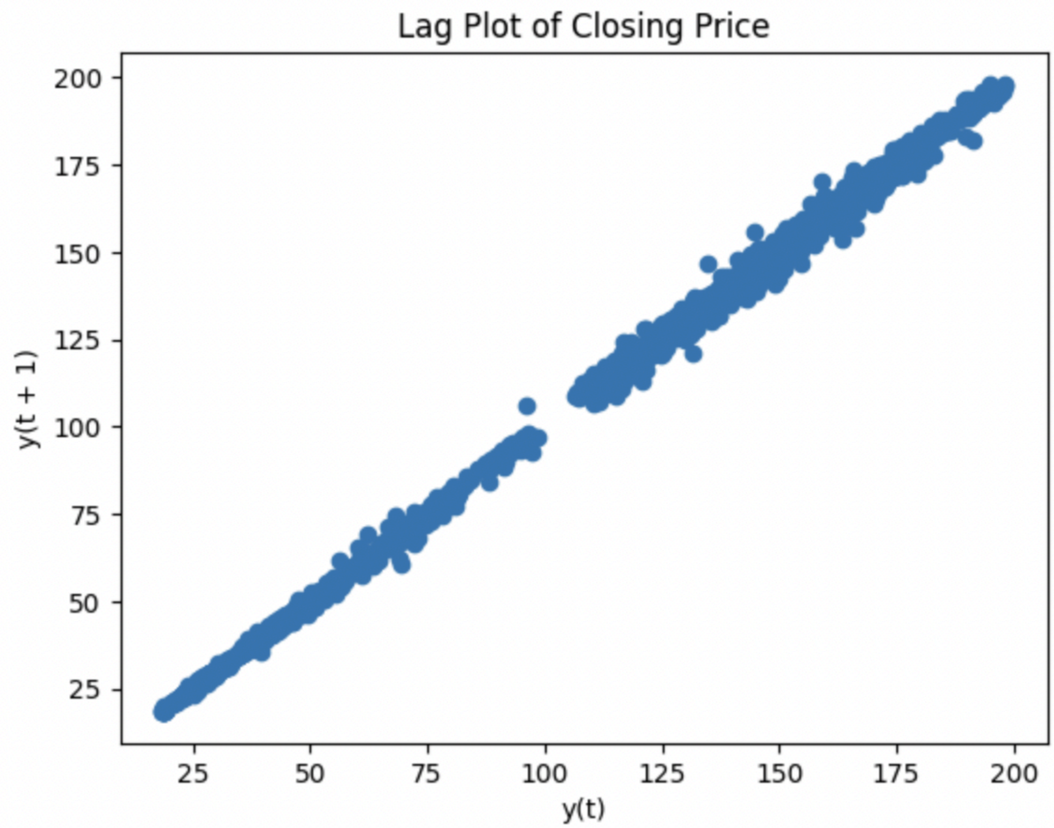
This graph illustrates Apple's closing price alongside 20-day and 50-day simple moving averages (SMAs) from 2014 to 2023. The SMAs smooth out short-term fluctuations, with the 20-day SMA closely tracking the closing price, while the 50-day SMA provides a broader view of trends. The alignment of these averages with the closing price highlights trends and potential reversal points in Apple’s stock performance.



This histogram shows a bimodal distribution of Apple's closing prices from 2014 to 2023, with a significant concentration in the 20-50 dollar range and another peak around $150.



This decomposition of Apple's closing price time series reveals a strong upward trend, capturing the overall growth in stock price from 2014 to 2023. The seasonal component shows recurring patterns likely influenced by periodic market behaviors, while the residual component highlights irregular fluctuations and noise, indicating potential short-term volatility not explained by trend or seasonality.



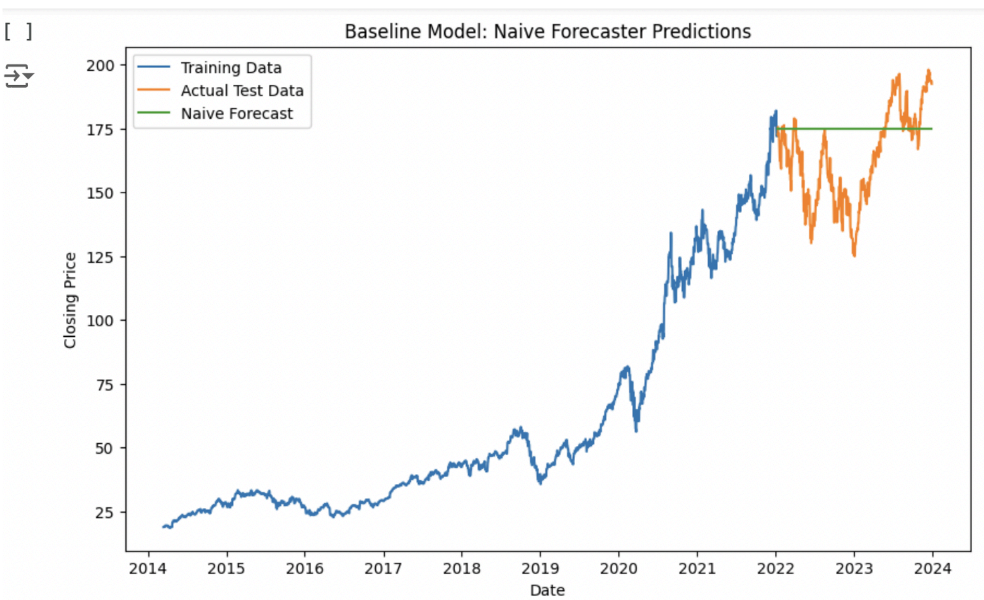
This lag plot of Apple's closing price shows a strong linear relationship between successive days' prices, indicating high autocorrelation. This suggests that the stock's closing price on a given day is strongly dependent on its price from the previous day, a key characteristic of time series data.

Modeling & Interpretations

To forecast Apple stock closing prices, I employed various time series forecasting models to analyze and predict future trends. Starting with a naive forecaster as a baseline, I then incorporated advanced methods such as ARIMA, Exponential Smoothing (Holt Winters'), Prophet to compare their predictive capabilities. Each model was evaluated using metrics like Mean Absolute Error (MAE) and Mean Squared Error (MSE) through temporal train-test splits.

Baseline Model (Naive Forecaster)

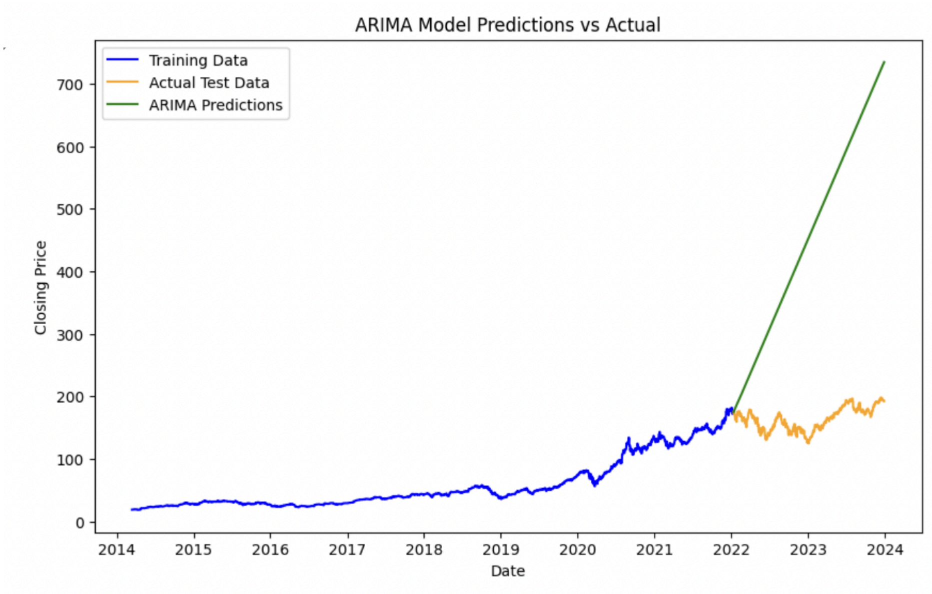
In order to establish a baseline for evaluating the performance of my time series models, I implemented a Naive Forecaster that uses the last observed value of the training data to predict the test data. This simple yet effective baseline provides a reference point for comparing the accuracy of more advanced forecasting methods.



The baseline model using the Naive Forecaster predicts the future closing prices as a flat line, accurately capturing the recent trend for short-term forecasting but failing to adapt to fluctuations in the actual test data.

ARIMA model

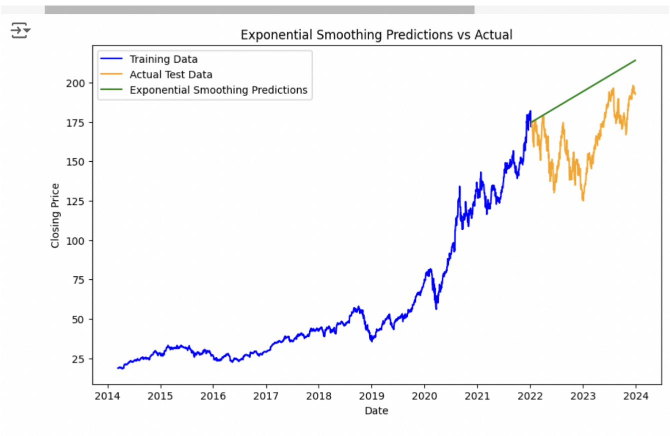
I decided to use the ARIMA (AutoRegressive Integrated Moving Average) model as my next metric because ARIMA captures linear trends and seasonality in data by modeling dependencies between past observations, which aligns well with the temporal nature of stock price data. By applying ARIMA, I aimed to see if incorporating these temporal relationships would result in more accurate predictions compared to the naive baseline model, which does not account for underlying patterns or seasonality in the dataset.



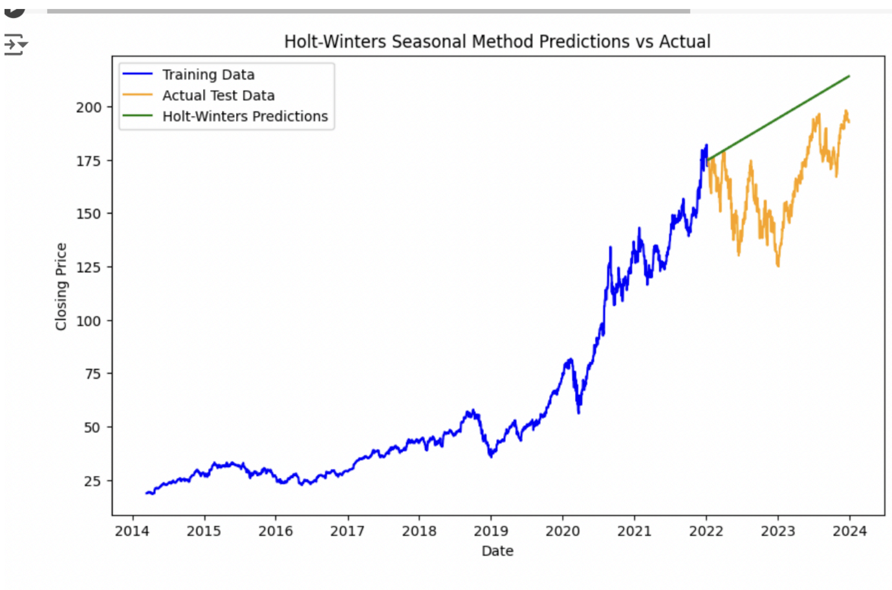
The ARIMA model demonstrated significantly higher errors compared to the baseline model, with a Mean Squared Error (MSE) of approximately 108,082 compared to the baseline's MSE of 455. This indicates that the ARIMA model's predictions deviated substantially from the actual values, suggesting it was not able to capture the underlying patterns effectively for this dataset.Additionally, the visual plot of ARIMA predictions shows a steep upward trend that diverges from the actual test data, reinforcing that the model's assumptions may not align with the stock price trends.

Exponential Smoothing and Holt-Winters' Seasonal Method

After evaluating the ARIMA model, which struggled to effectively capture the stock price patterns due to its limitations in handling seasonality, I chose Exponential Smoothing and Holt-Winters' Seasonal Method as the next metric. These models are well-suited for time series data with underlying trends and seasonal components, that can address the limitations of ARIMA. By explicitly incorporating trend and seasonal effects, these methods allow for more flexible and accurate forecasting of complex time series data.

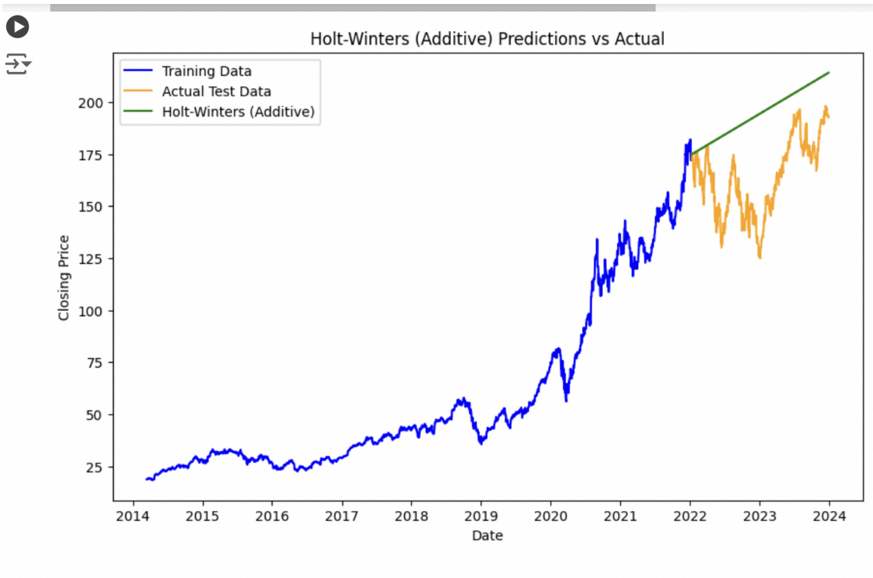


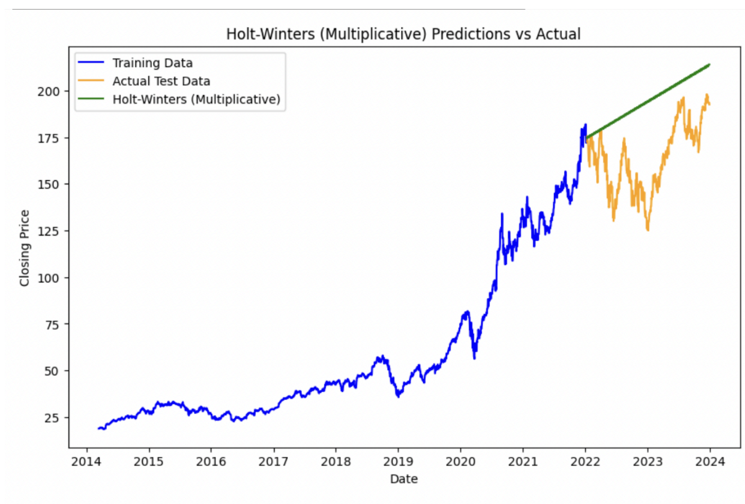
The Exponential Smoothing model outperformed the ARIMA model, achieving an MSE of 1,194 compared to ARIMA's 108,082, and produced a MAE of 31.08. Unlike ARIMA, which diverged significantly from the test data, Exponential Smoothing closely followed the stock price trends. While its performance was slightly less accurate than the baseline model, it provided a reasonable trend-based forecast, demonstrating its utility in capturing underlying patterns in stock prices. The visual plot for Exponential Smoothing shows a better alignment with the actual test data compared to ARIMA, making it more effective in capturing trends, though it still lacks the precision of the baseline model in this dataset.



The Holt-Winters' Seasonal Method achieved an MSE of 1,193.48 and an MAE of 31.06, similar to the Exponential Smoothing model. While its MSE is higher than the baseline model (MSE: 455), Holt-Winters better captures trends and seasonality, as shown in the predictions. Compared to the Naive Forecaster (MSE: 455, MAE: 17.34), Holt-Winters provides more nuanced, trend-aware forecasts but has higher error metrics due to its complexity. It significantly outperforms the ARIMA model (MSE: 108,082), aligning more closely with the actual data and avoiding ARIMA's divergence from trends. Overall, Holt-Winters' Seasonal Method performs comparably to Exponential Smoothing but provides additional insights through its explicit modeling of seasonality, which could be useful for datasets with pronounced seasonal components.

To further refine our analysis, we explored both additive and multiplicative seasonality in the Holt-Winters method to determine which best captures the seasonal patterns in Apple's stock prices.

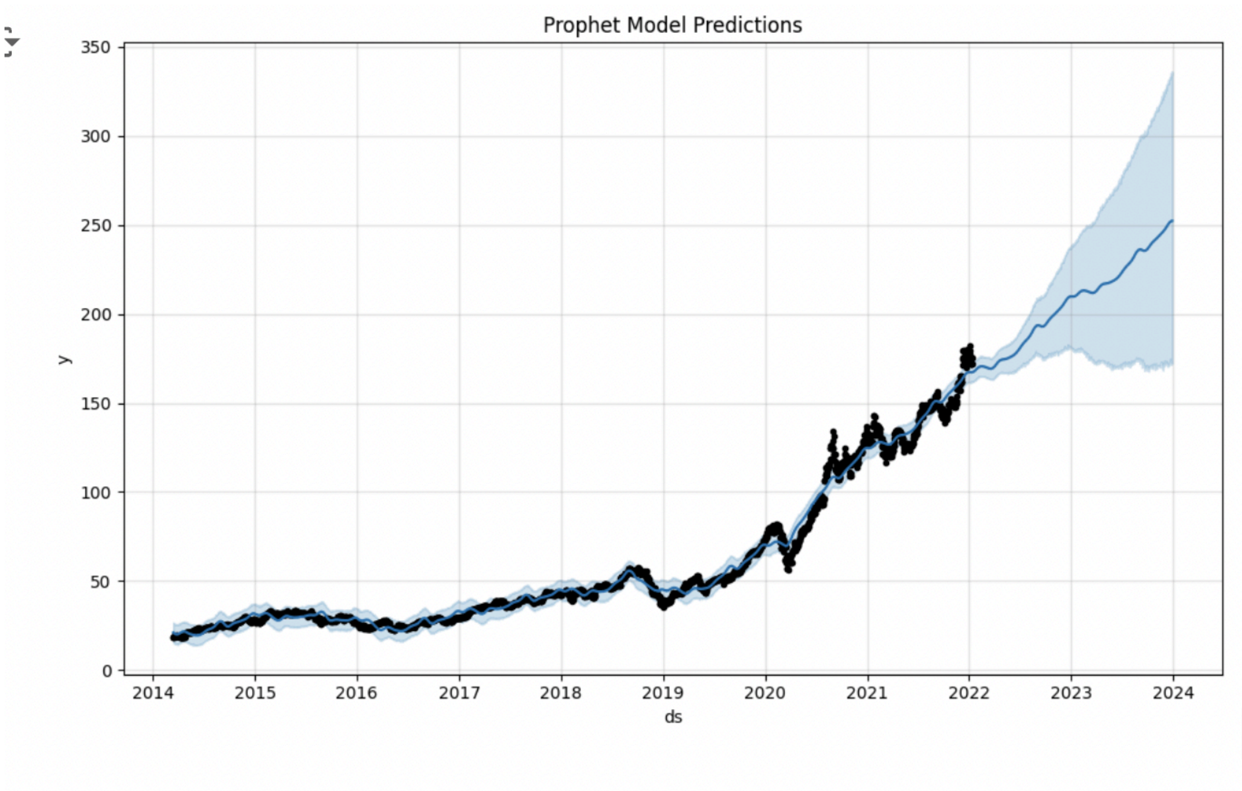




Comparing Holt-Winters models with additive and multiplicative seasonality shows that the multiplicative model slightly outperformed the additive model, achieving a lower Mean Absolute Error (MAE) of 30.95 and Mean Squared Error (MSE) of 1,186.57 compared to the additive model's MAE of 31.06 and MSE of 1,193.48. This suggests that multiplicative seasonality better captures the non-linear seasonal variations present in the stock price dataset. While the Exponential Smoothing model, which does not account for seasonality, performed similarly to the additive Holt-Winters model, the inclusion of multiplicative seasonality demonstrates a marginal improvement in accuracy. Among the smoothing-based methods, the multiplicative Holt-Winters model currently provides the most accurate predictions for this dataset.

Prophet model

I chose the Prophet model model to leverage its flexibility in handling complex trends and irregular seasonality since it is particularly useful for time series data with missing values or seasonal variations.



The Prophet model achieved a Mean Absolute Error (MAE) of 42.81 and a Mean Squared Error (MSE) of 2,262.71, which is higher than the performance of Exponential Smoothing (MAE: 31.08, MSE: 1,194.26) and Holt-Winters with multiplicative seasonality (MAE: 30.95, MSE: 1,186.57), indicating it did not outperform these models in this case. However, Prophet significantly outperformed ARIMA (MAE: 290.95, MSE: 108,082), which struggled with divergent predictions. While Prophet offers flexibility in handling seasonality and trends, its error metrics suggest that it is less effective than Exponential Smoothing and Holt-Winters for this dataset. The visual plot of Prophet predictions shows a generally consistent trend but includes higher prediction uncertainty, evident in its confidence intervals, which may contribute to the larger error values.

Next Steps & Discussion

Summary of Findings:

In my analysis of Apple stock price predictions, all models demonstrated varied performance compared to the baseline predictor, highlighting the importance of appropriate model selection for time series forecasting. The models ranked in terms of performance are as follows: Holt-Winters with multiplicative seasonality, Holt-Winters with additive seasonality, Exponential Smoothing, Prophet, ARIMA.

Key Findings:

1) Success of Holt-Winters Model with Multiplicative Seasonality: Among the models, the Holt-Winters model with multiplicative seasonality emerged as the most effective, showcasing the best predictive capabilities. Its ability to account for both trend and seasonal patterns led to consistently accurate predictions, outperforming other models.

2) Performance of Prophet Model:While the Prophet model demonstrated flexibility in handling complex trends, its Mean Absolute Error (MAE) and Mean Squared Error (MSE) were higher than those of the Holt-Winters models. However, its adaptability makes it a strong candidate for data with irregular seasonalities or missing values.

3: ARIMA and Naive Forecaster Limitations: The ARIMA model struggled with capturing the underlying trends of the stock prices, leading to overly divergent predictions. Similarly, the Naive Forecaster, relying solely on the last observed value, proved insufficient in modeling the complexities of stock price movements.

In conclusion, the Holt-Winters model with multiplicative seasonality provided the most accurate and robust forecasts, effectively leveraging both trend and seasonal components. These findings underscore the importance of incorporating seasonality and trend analysis in time series forecasting for stock prices, offering insights for future improvements in predictive modeling strategies.

Next Steps/Improvements:

To enhance the predictive capabilities of the time series models and achieve more nuanced insights into Apple's stock price trends, I would want to incoporate these additional features into my models:

- Adding External Factors: Including data like economic indicators, market trends, or major news related to Apple could provide additional context that helps the models better understand stock price movements. This would allow me to account for factors beyond just historical prices.
- Technical Indicators: Adding derived features like Relative Strength Index (RSI) or Bollinger Bands could help the models capture additional dimensions of market dynamics that are often used in technical analysis.
- Lag Features: Adding lagged values of closing prices as additional features could help models account for momentum and past trends, which are often predictive of future stock price movements.
- High and Low Prices: Leveraging the daily high and low prices in addition to the closing prices could allow the models to better understand intraday volatility and price range trends, offering a more comprehensive view of the stock's behavior.
- Exploring Deep Learning: Trying more advanced methods like Long Short-Term Memory (LSTM) networks could allow me to capture complex patterns and long-term dependencies in the stock price data. These models could open up new opportunities for more robust predictions.

By integrating these additional factors into the analysis, I believe my models will be better equipped to provide even more accurate forecasts and a better understanding of the factors driving Apple’s stock prices. These improvements could offer valuable insights for investors and decision-makers looking to navigate the complexities of the stock market.