Time Series Forecasting for Toyota Motors

Riti Dahal

University of South Dakota

Instructor: Dr. Tiahrt

Data Mining for Competitive Advantage (DSCI 725)

May 9, 2025

Table of Contents

1 Introduction

For many years, the stock market has been seen as a "predictor" or indication of the state of the economy. While many think that significant increase in stock prices predict future economic development, significant declines in stock prices are indicative of a future recession (Comincioli, 1996). Thus, in the global economy, the stock market serves as a key indicator for financial stability. Forecasting stock price has become crucial for financial analysts and investors to make informed decisions in an unpredictable market. In order to forecast future stock prices and market trends, historical data must be analyzed and statistical models must be used.

Established in 1937, Toyota Motor Corporation has evolved from a domestic Japanese automaker to a world leader in automobile manufacturing and innovation. Toyota has consistently raised the bar for the global car industry based on quality engineering and lean manufacturing. Being a prominent player in both the automotive industry and the worldwide financial markets, it was first listed on the Tokyo Stock Exchange in the year 1949 and the New York Stock Exchange (NYSE) in 1999 (News, 1999).

The main purpose of this project is to explore the historical data of Toyota and forecast the future trends by using different time series forecasting techniques. Using statistical models including Naïve Forecasting as the baseline model, Exponential Smoothing (ETS), Time Series Linear Models (TSLM), ARIMA, and Random Walk with drift. Further, to evaluate the model’s performance, the project uses different key metrics like the mean absolute scaled error (MASE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE).

2 Background Research

Stock price forecasting is a crucial area of study in financial research as it helps in risk management, investment planning, and financial decision-making. However, it is challenging to forecast stock prices due to market fluctuations and global events. With technological advancements and statistical methodology, we can use historical data and implement advanced time series forecasting.

The Naïve model is a baseline technique used in previous studies to predict future values that represent the most recent observed data point. Exponential Smoothing (ETS) approaches provide flexibility by taking seasonality and trend. ARIMA also known as Autoregressive Integrated Moving Average model is widely used to handle non-stationary data as it incorporate auto regression, differencing, and moving average components. According to (A. Ariyo & Adewumi, 2014) ARIMA model is one of the potential models for short-term prediction.

Similarly, Time Series Linear Models (TSLM) is helpful when adding explanatory variables, like economic indicators to make it simpler to incorporate external regressors (Sonkavde & Dharrao , 2023). Random Walk with Drift models, which assume that the value of the series fluctuates randomly over time but shows an upward or downward trend, also offer a probabilistic framework. Therefore, the project uses Toyota's historical data and explores different forecasting methods to evaluate its usefulness and robustness.

3 Problem Presentation

The stock markets are extremely volatile. Various factors affect the stock prices, such as economic factors, industry, global events, business-specific factors, etc. Economic factors, including the interest rate, inflation, and consumers' purchasing behavior, can impact an investor's sentiments and market behavior. For instance, if the interest rate is extremely high, making car loans less affordable, which can have a direct impact on car sales in the case of Toyota models. Global events such as war, geopolitical tension, and pandemics have also affected the stock market significantly. During the COVID-19 pandemic, which dropped Toyota car sales were the supply chain disruption, shutdown of manufacturing, and liquidity issues (Fu, 2023). Despite being one of the leading companies with a strong operational structure, Toyota faced these challenges, and the overall uncertainty was seen in its stock price.

Further business-specific events, such as quarterly earnings, strategic alliances, or leadership changes can also have both positive and negative impacts on Toyota's stock market. Likewise, industry-specific factors are also important. The increasing trend towards electric vehicles, fuel price fluctuations, environmental regulations, etc., highly influence the automobile industry. In such cases, companies like Toyota have to constantly adjust to unpredictable circumstances, and investor trust is often based on how well the company innovates and adapts. Thus, forecasting such fluctuations is important for financial analysts and investors to mitigate financial risks and make data-driven decisions. In such a case, understanding Toyota’s stock market is important for both short and long-term trading given its global impact in the automotive industry. The project uses different statistical models among which the best model is chosen based on its evaluation metrics.

4 Specification and Design

The paper uses historical data from 2021 to 2025, to forecast the daily closing prices for Toyota Motors' shares. The data was processed in R. Data transformation, trend and seasonality analysis, stationarity testing, and model validation were all part of the forecasting pipeline. The best forecasting model was chosen by iteratively applying many time series models, including Naïve, STL + ETS, ARIMA, and Random Walk with Drift. Therefore, performance was evaluated using RMSE, MAPE, and MASE.

5 Data Acquisition

The data set used in the project is obtained from Google Finance using the formula:

= GOOGLEFINANCE("TM", "all", DATE(2021,1,4), TODAY(), "DAILY") which extracts Toyota Motors Corporation stock price from the year 2021 to 2025. The data set includes variables such as Date, Open, High, Low, Close, and Volume. The predictive purpose of the project is to predict the future trend of Toyota Motors Corporation based on historical data and for that 'Closing Price' is selected as the primary variable for forecasting.

6 Data Exploration and Transformation

Initially, the data exploration was done to check the structure and quality of the data set. R functions such as str (), head (), summary (), and colSums(is.na()) are used to check the missing values, data types, and duplicate values, and format the date columns. The summary of the data showed its completeness for modeling. Further, to prepare data for time series forecasting I have transformed the date column from character to date class using as.POSIXct and as.Date. An additional variable 'Year' was extracted to simplify visual segmentation by year and a time series object was created with a frequency of 252 to represent the number of trading days in a year.

Once the data exploration was done exploratory data analysis (EDA) was carried out to analyze the general structure and fluctuations in Toyota's stock price over time. The main idea behind this was to identify the long-term patterns, yearly variations, and potential trends. The patterns and trends are further depicted through plots below:

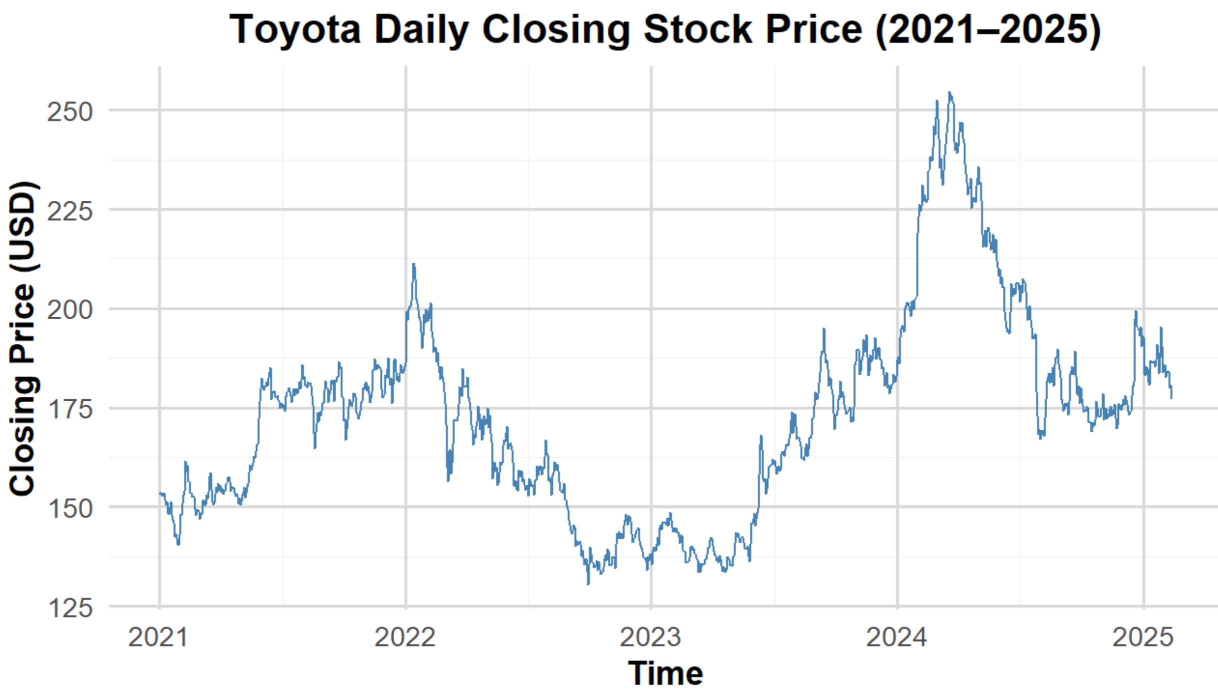

Figure 1: Time Series Plot of Daily Closing Price

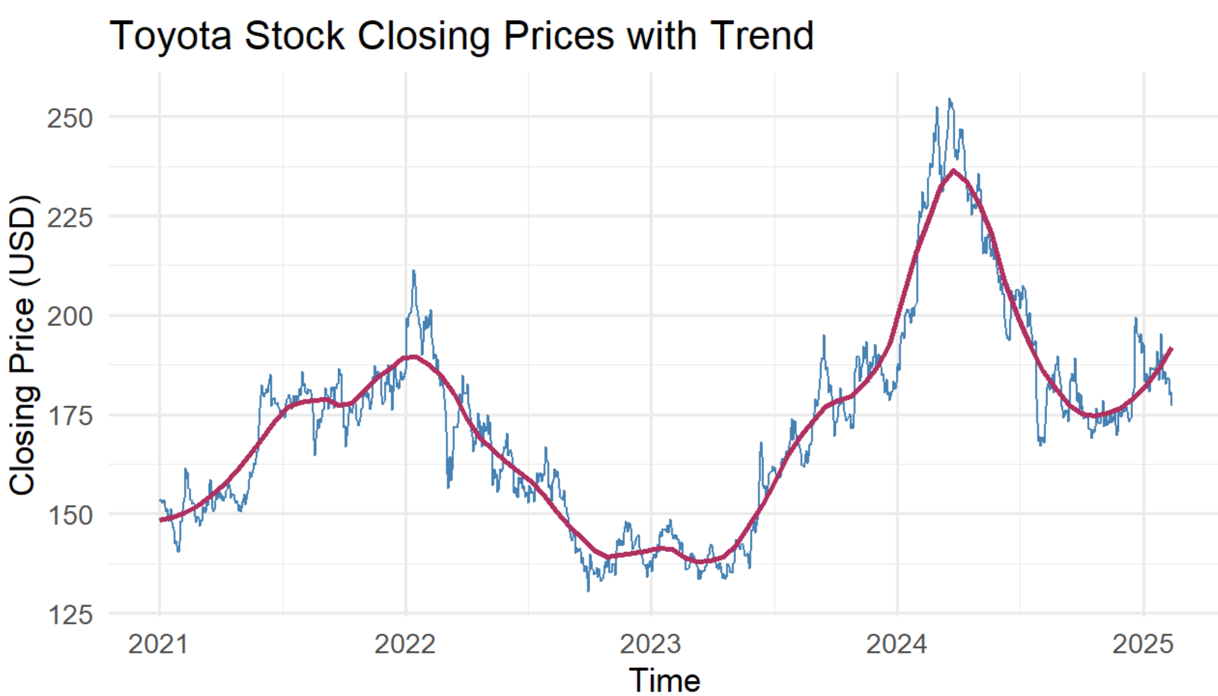

Figure 2: Toyota Stock Closing Prices with Trend

Figure 2 shows a Toyota Stock Closing Prices with Trend. As we can see from the figure the third-order polynomial trend is supported by the curved pattern that Toyota's stock price plot displays as it rises and falls over time. Further, the plot also shows that the additive seasonality may be present since the ups and downs surrounding the trend remain roughly the same size. This indicates that seasonal variations do not increase or decrease in parallel with the trend. Thus, the time series plot exhibits both seasonal patterns and a wavy long-term trend.

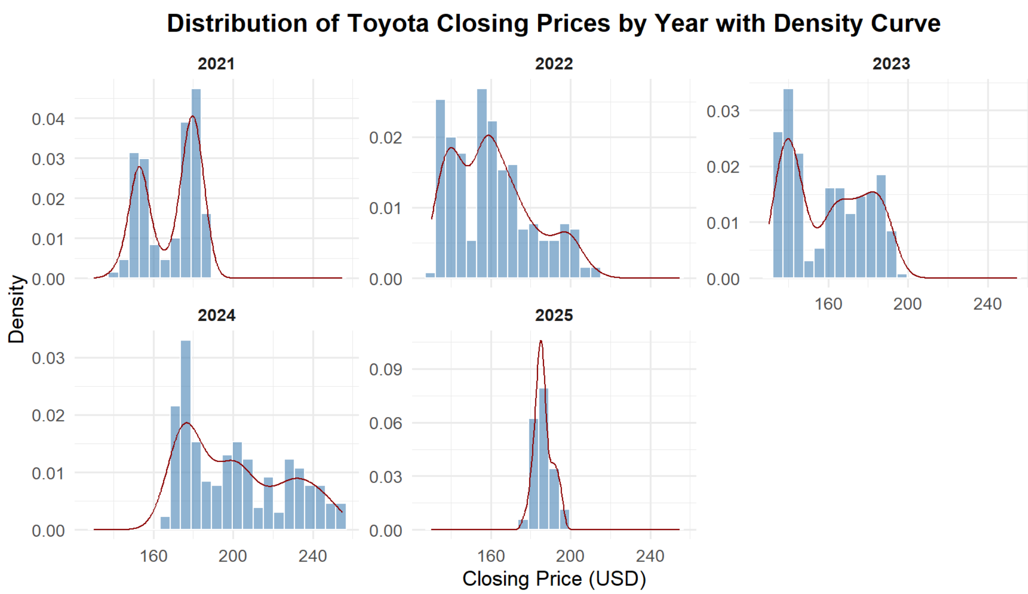

Figure 3: Distribution Plot with Density Curve

Figure 3 shows the distribution of Toyota Closing Prices from the year 2021 to 2025. In the years 2021 and 2022, the data is left-skewed which indicates the higher prices with sporadic declines. The stock market in 2023 seems more balanced and symmetrical. Likewise, in 2024, the overall trend shifts to the right which indicates a surge in prices with some extremely high values.

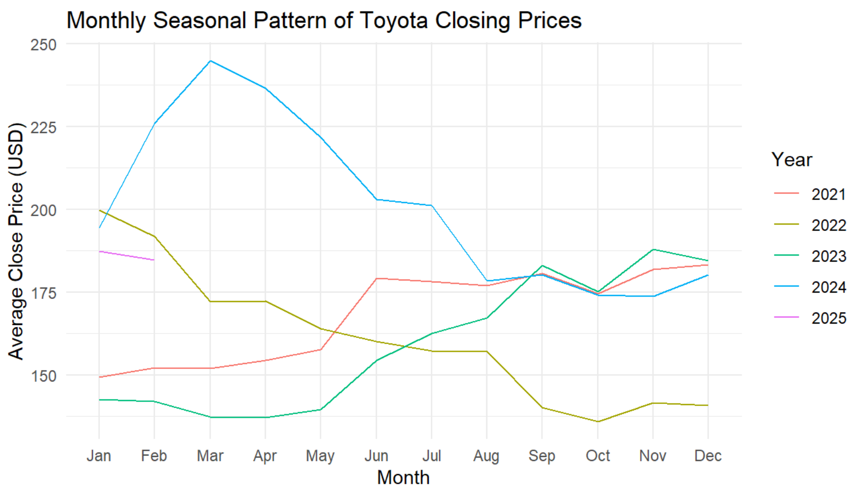

Figure 4: Monthly Seasonal Pattern

The monthly seasonal pattern shows that there was a steady decline throughout the second half of the year 2024. On the other hand, in the year 2022, there is a significant drop from January to October, which indicates the unstable stock market. Lastly, with a moderate increase in the second half, 2023 and 2021 displayed more consistent trends.

In the next step, we decompose the time series into its components using the additive method.

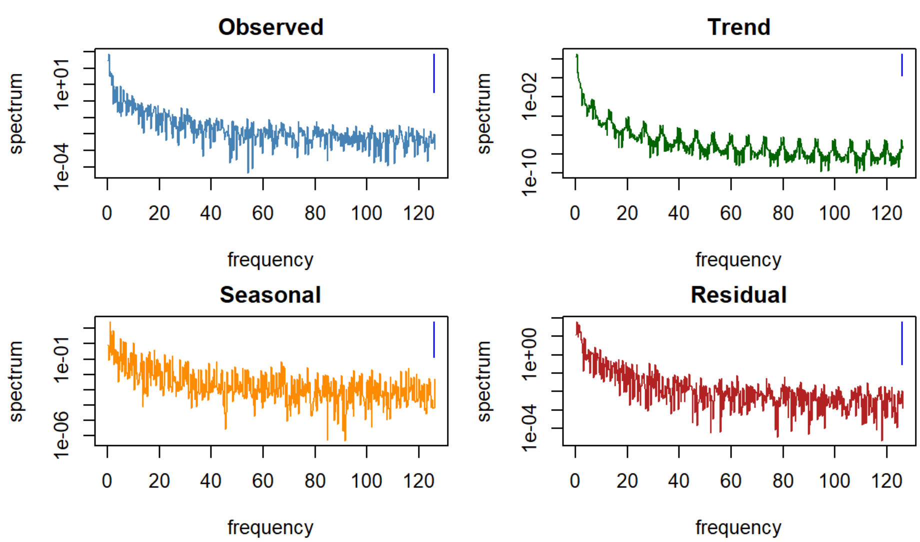

Figure 5: Time Series Component Plots

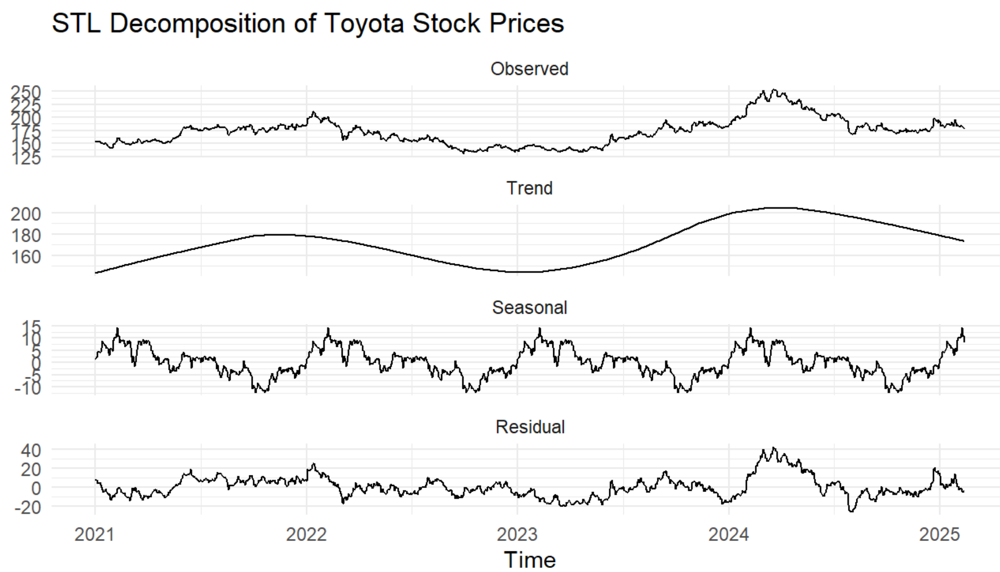

Figure 6: STL Decomposition of Toyota Stock Prices

The four main components help in obtaining trends, patterns, and errors within the data. The four components of the time series based on the Figure 6. are explained below:

Observed: It shows that long-term patterns like trend and seasonality predominate the original time series data.

Trend: The spectrum of the trend component shows that it resides on extremely low frequencies. This indicates the upward or downward movements in the stock prices over time.

Seasonal: From the graph we can see a periodic increase that aligns with regular seasonal trends in the stock data. This indicates that the seasonal pattern might be affected by production cycles or customer behaviour.

Residual: The remainder, which approximates a white noise technique, indicates fluctuations in the short term and unpredictable movements after trend and seasonality are removed.

7 Augmented Dickey-Fuller Test

The initial data exploration result shows that the Toyota Stock data contains both trend and seasonality. Thus, the Augmented Dickey-Fuller test is applied to achieve stationarity. The raw data shows that the time series was non-stationary with a statistic test of (-1.7456) and a p-value of (0.686). Since the p-value is greater than 0.05, we fail to reject the null hypothesis.

To address the non-stationarity of the series, the first-order differencing was applied to transform the data into stationarity. The result shows a statistical test of (-10.28) and a p-value of (0.01). Since the p-value is smaller than 0.05, we reject the null hypothesis.

Transformation	**Statistic Test**	**p-value**
Before differencing	-1.7456	0.686
First order differencing	-10.28	0.01

Table 1: ADF Test Result

8 Auto-Correlation Function

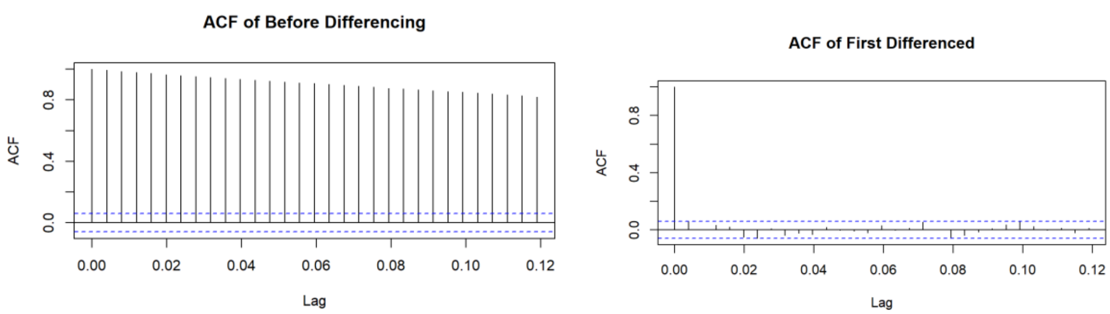

Figure 7: ACF chart (before and after differencing)

The ACF chart on the left shows there is a high autocorrelation at all lags. This indicates that the original data series has non-stationarity. On the other hand, the chart on the right shows a significant drop after lag 1 and stays within the confidence limit. This indicates that the data series is stationary and appropriate for time series modelling.

9 Data Partitioning

Before diving into forecasting models, I have partitioned my data into a training and a test set. The training set consists of the data set from January 2021 to December 2024, and the test set consists of the data from January 2025 onward. I have excluded the 2025 data to prevent data leakage and ensure an unbiased evaluation of the forecasting models. The figure below shows the data partitioning of the Toyota Stock Data.

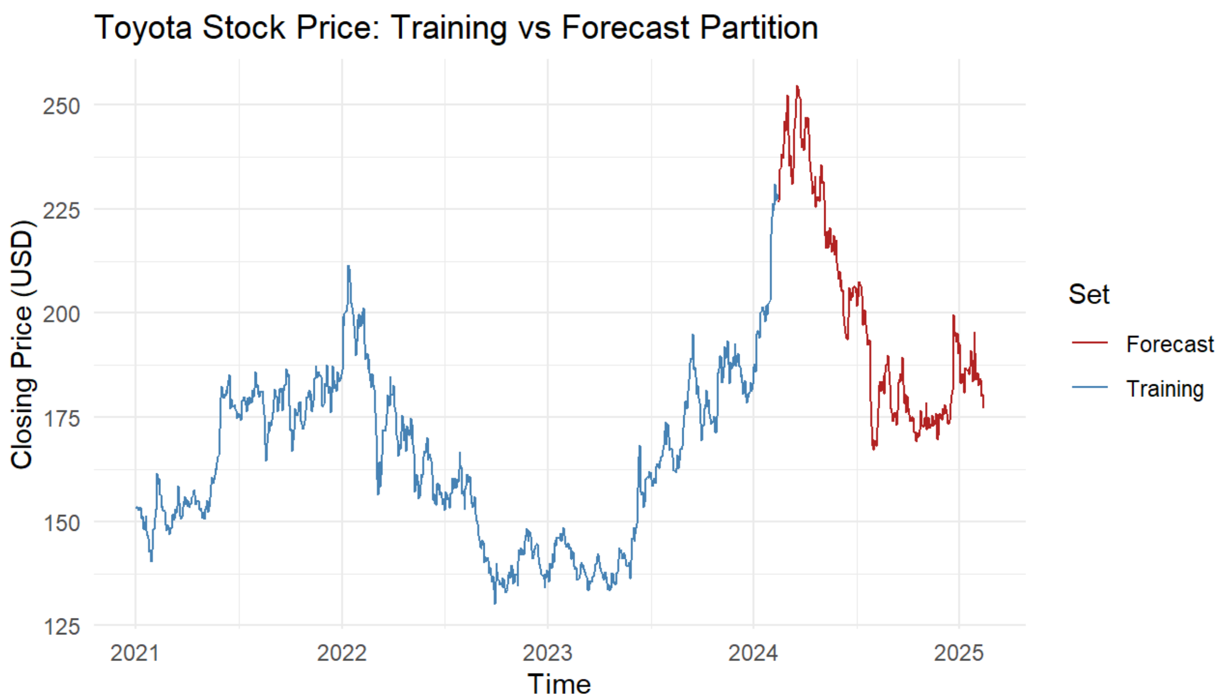

Figure 8: Data Partitioning

10 Motivation Based Theoretical Consideration and Iterative Results

Once the data partitioning is done, this part highlights different forecasting techniques that have been used to forecast the stock price. The primary purpose is to forecast fluctuations in stock prices by employing both basic and complex models that reflect the basic patterns of the data. The model that I have used for this project is explained below:

10.1 Naive Forecasting

The first model that has been chosen for time-series forecasting is the Naive Model. The naive model's simplicity and presumption that the most recent observation is the best predictor resulted in its selection as a baseline method. Although it provides a solid foundation, its test results showed forecast errors (RMSE = 38.07, MAPE = 18.18, MASE = 1.02). This result indicates that it has limited predictive power when there are underlying trends or structural changes.

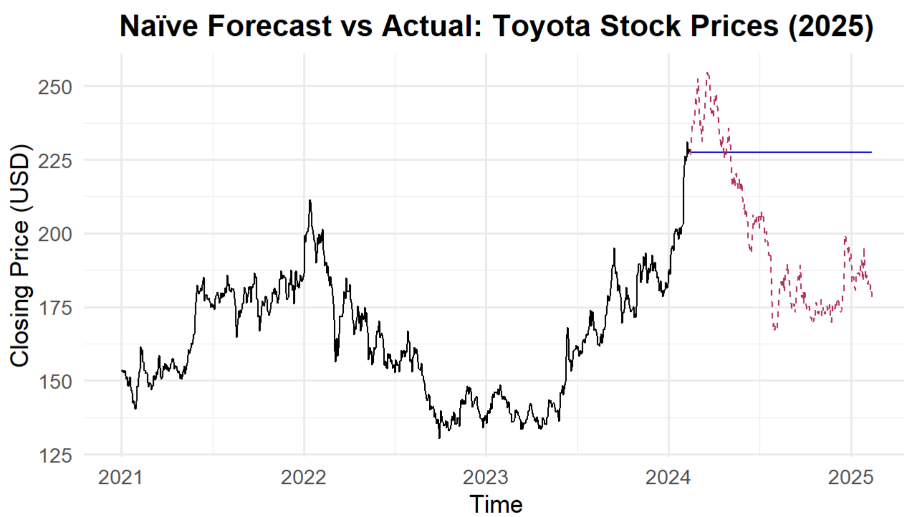

Figure 9 : Naive Forecasting

The figure 10 illustrates how the actual and expected stock values differ over time for Toyota. Despite the presence of some noticeable spikes and fluctuations, the residuals are scattered around the value of zero, suggesting that there is likely some autocorrelation alongside some degree of randomness. This means the naïve model lacks the necessary complexity to capture additional useful patterns such as seasonality.

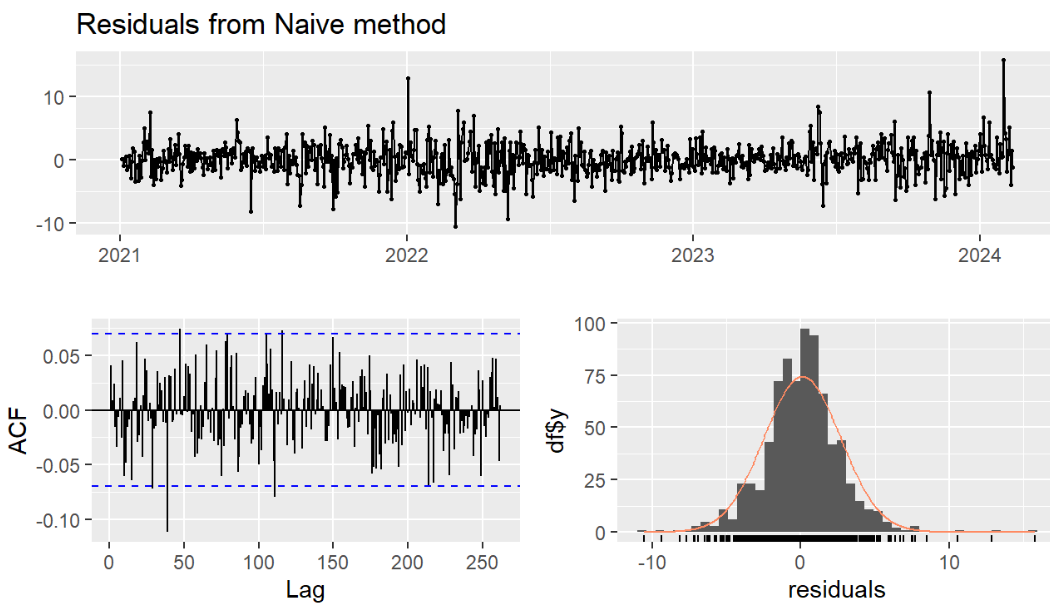

Figure 10: Residuals from Naïve Forecast Model

10.2 STL and ETS Model

STL and ETS have been used as the second forecasting model to analyze the trend and seasonality captured in Toyota's stock price data. From the plot, the model captures the historical trend and produces the best estimates for the future with utmost precision.

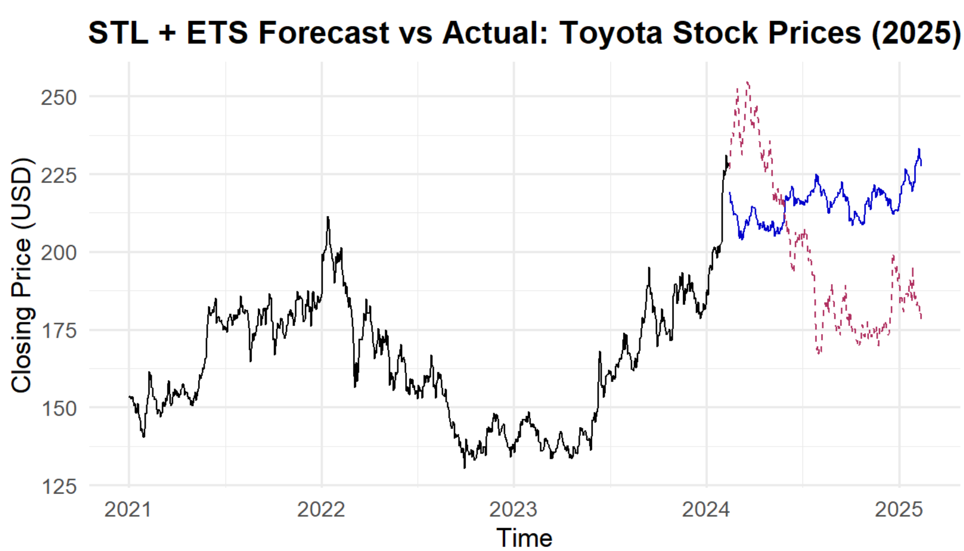

Figure 11: STL and ETS Forecast

The evaluation metric of the model shows a Test RMSE of 33.18 which corresponds to 25.32 in the training set, MAPE of 30.71 which was 22.8 in the training set and MAPE was 16.09%. These numbers show that the model is not optimally tuned and there is moderate error in predictions.

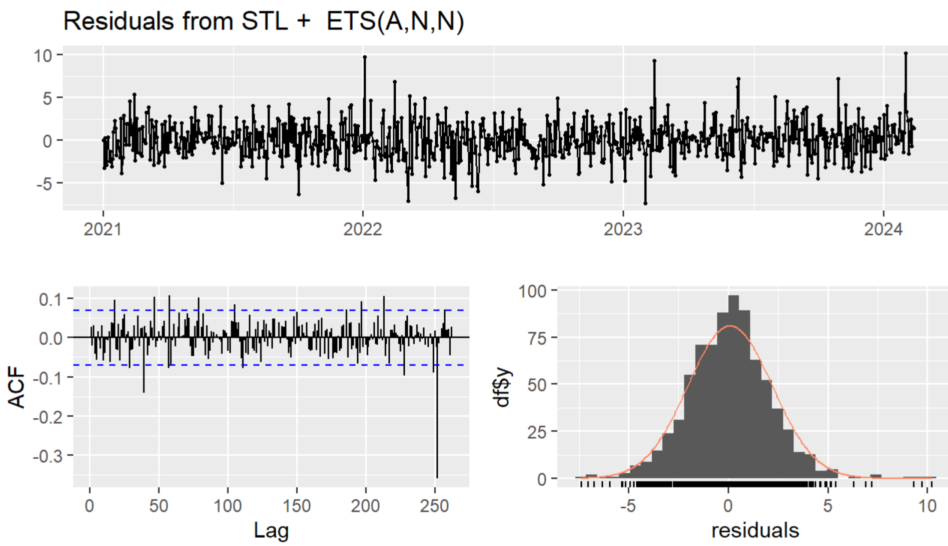

Figure 12: Residuals for STL and ETS

From the figure 12, the residuals for Toyota stock price data seem to be evenly distributed around zero, which means that the model has adequately explained the data. All of the ACF plot spikes seem to be within the confidence interval, thus supporting the hypothesis that the residuals are uncorrelated. Further, the histogram shows a normal distribution. These results illustrate that the STL + ETS model captures the trend and seasonality components of the data without overwriting the residuals, and therefore, the model can be used for forecasting.

10.3 ARIMA Model

The moving average and autoregressive components of Toyota's stock prices were both captured by the ARIMA model. The model used the auto.arima() function to automatically choose the parameters that best match the AIC criterion. Since residual autocorrelation was appropriately addressed by differencing and the auto.arima() algorithm effectively lowered AIC, manual tuning was not explored.

Forecasts were produced following model fitting and contrasted with the actual stock prices during the test period. The model that was chosen was ARIMA(0,1,0). This model functions the same as a random walk model without drift. However, its forecasting performance on the test set was poor, as illustrated by the high RMSE of 38.08 and MAPE of 18.19%, which show significant prediction errors.

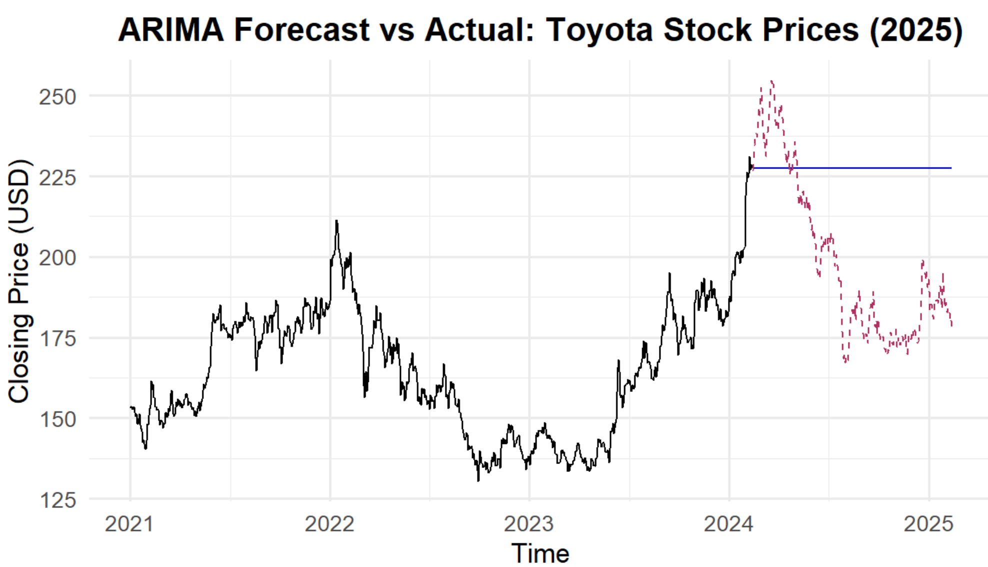

Figure 13: ARIMA Model

In contrast to a naive model, figure 14 shows a roughly normal distribution however, the high Theil's U value (12.44) indicates weak predictive power. Therefore, the ARIMA model fails to generalize well on unseen data, even though it is theoretically flexible.

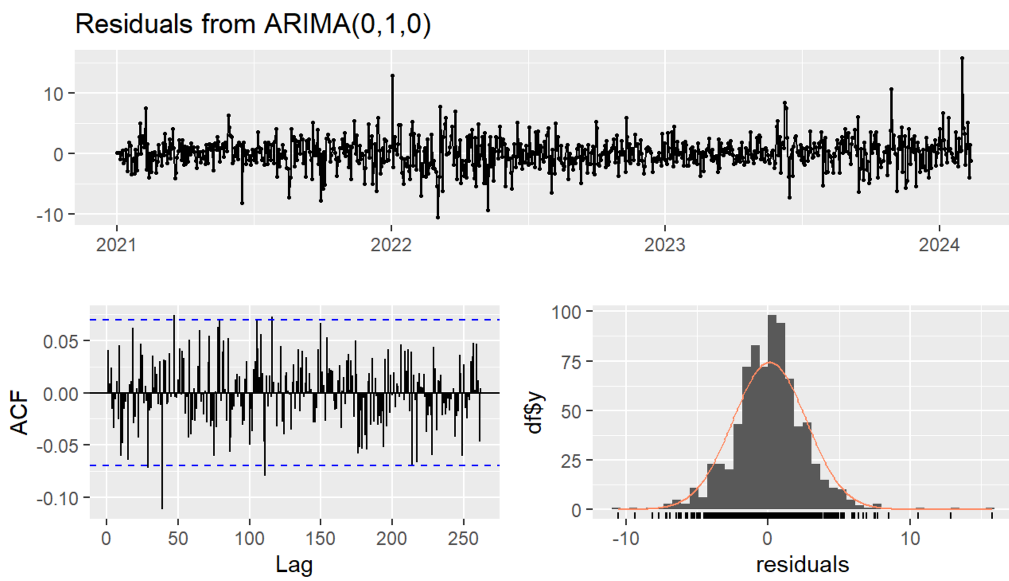

Figure 14: Residuals from ARIMA

10.4 Random Walk with Drift

Lastly, the Random Walk with Drift (RWF) was used for the forecasting model. It was chosen on the basis of the financial theory, which states that stock prices rise in value as they follow an unpredictable process with a consistent directional component. As a result, RWF is the ideal for stock data. In comparison to all other models evaluated, the RWF model stood out due to its predictive accuracy. In the test set (2025), RWF achieved a forecast accuracy of RMSE (10.53), and MASE (0.275) which indicates, that RWF outperformed as compared to other models. In addition, the mean absolute error of 9.07 and mean absolute percentage error of 3.78% show Toyota's stock price movements were tracked with high accuracy due to low deviation.

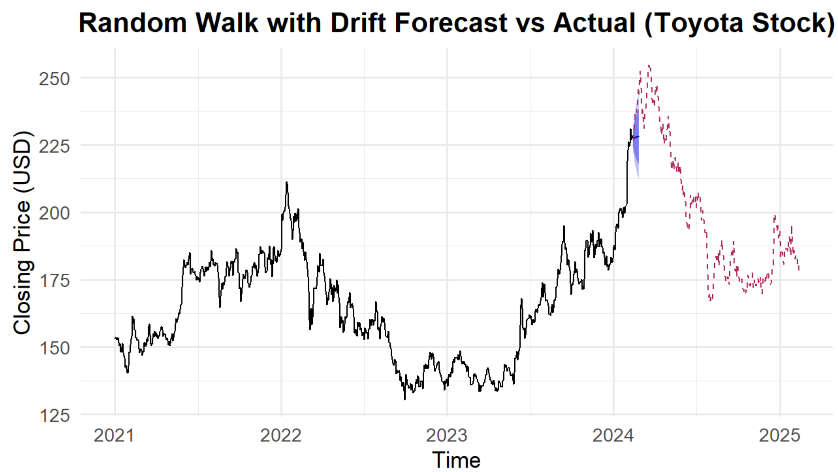

Figure 15: Random Walk with Drift

Although the residuals of the test set showed some degree of autocorrelation with an ACF1 score of 0.628, it significantly outperformed baseline approaches about predictive power. Therefore, RWF proved to be the most effective in tracking the ongoing trend in the stock prices of Toyota during the forecast horizon.

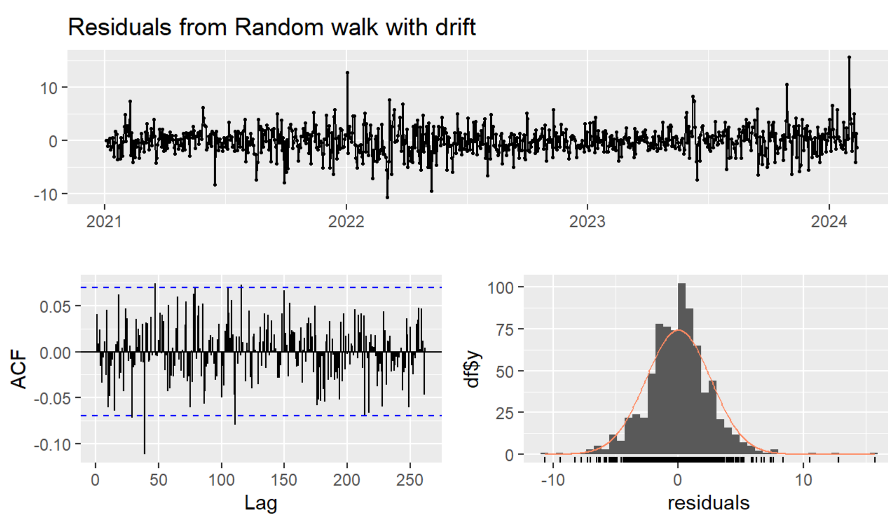

Figure 16 : Residuals from Random Walk with Drift

The project uses four different models i.e. Naïve, STL and ETS, ARIMA, and Random Walk with Drift. This shows an iterative process has been followed to obtain the best forecasting model. The initial assumptions related to trend and seasonality showed that the data series was non-stationary, leading to transformation, differencing, and iterative model selection.

11 Comparing Models and Results

For Training Set

Table 2: Performance Metrics for Training Set

Model	**RMSE**	**MAPE (%)**	**MASE**
Naive	2.596	1.15	0.058
Random Walk with Drift	2.595	1.148	0.058
STL + ETS	2.127	0.985	0.49
ARIMA	2.595	1.149	0.058

For Test Set

Table 3: Performance Metrics for Test Set

Model	**RMSE**	**MAPE (%)**	**MASE**
Naive	38.07	18.18	1.02
Random Walk with Drift	10.53	3.78	0.27
STL + ETS	33.18	16.09	0.93
ARIMA	38.075	18.19	1.02

For model evaluation, the project compares three different metric that are :

- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error (MAPE)
- Mean Absolute Scaled Error (MASE)

12 Performance Summary

The Naïve model performed well on the training set with RMSE (2.596), MAPE (1.15%) and MASE (0.058). However, on the test set, it performs poorly with RMSE (38.07), MAPE(18.18%), and MASE (1.02). This indicates that despite Naïve as a baseline model it showed a limited predictive power. ARIMA model has shown a similar performance as the Naïve Model. However, ARIMA was unable to make accurate projections on future data. STL and ETS had the strongest fit to historical data, with the lowest training RMSE (2.127) and MAPE (0.985%). The performance on the test set, however, was significantly lower indicating possible overfitting and low predictive power. Likewise, Random Walk with Drift outperformed on the test set with the lowest RMSE (10.53), MAPE (3.78%), and MASE (0.27). Thus, RWF was chosen to be the best forecasting model for Toyota Motors. The figure below shows the final plot of Random Walk with drift.

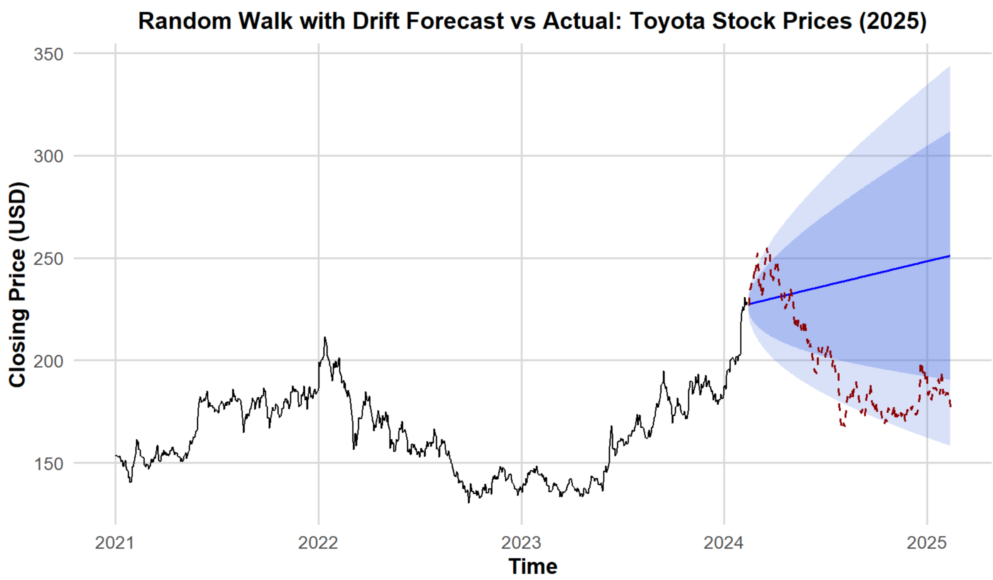

Figure 17: Final Plot

Figure 17 shows the Random Walk with Drift model's predictive power and limitations. Although the model is effective at capturing historical patterns, it might not be able to handle sudden changes which are common in stock markets. However, this is reduced by the addition of wide confidence intervals, which offer a reasonable range of expected values.

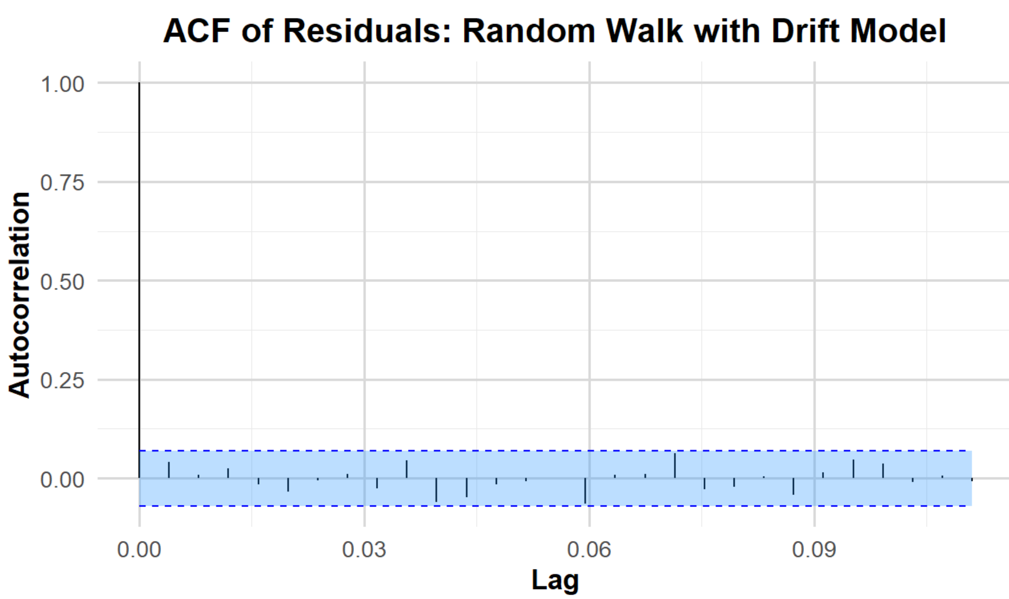

Figure 18: ACF of Residuals

Figure 18 shows an Autocorrelation Plot of Residuals generated by the Random Walk with Drift Model. Form the plot we can say that the model captures the structure in the data since the outcome shows that residuals act like white noise. Thus, one of the most important conditions for accurate forecasting, which is the independence of residuals, is indeed fulfilled by the Random Walk with Drift model.

Table 4: Forecasted Data Points

Month	Forecast	Lo_80	Hi_80	Lo_95	Hi_95	Month_Date
Feb-25	227.6431	222.0533	233.2328	219.0943	236.1919	01/02/2025
Mar-25	228.8697	215.6804	242.059	208.6984	249.041	01/03/2025
Apr-25	230.8984	210.1122	251.6846	199.1086	262.6881	01/04/2025
May-25	232.9742	206.3881	259.5604	192.3142	273.6342	01/05/2025
Jun-25	235.0029	203.5699	266.4359	186.9303	283.0755	01/06/2025
Jul-25	237.0787	201.1797	272.9778	182.1759	291.9816	01/07/2025
Aug-25	239.1546	199.1127	279.1965	177.9158	300.3934	01/08/2025
Sep-25	241.1833	197.3229	285.0436	174.1047	308.2619	01/09/2025
Oct-25	243.3063	195.6308	290.9818	170.3929	316.2196	01/10/2025
Nov-25	245.3349	194.1475	296.5223	167.0506	323.6193	01/11/2025
Dec-25	247.3636	192.7732	301.9541	163.8747	330.8525	01/12/2025

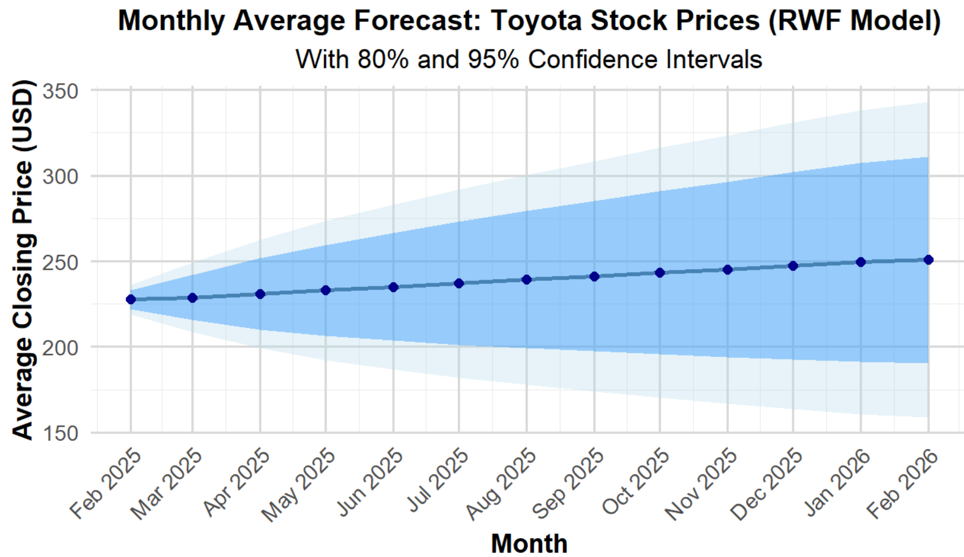

Figure 19: Monthly Average Forecast

The figure 18. shows the Random Walk with Drift model's monthly average forecast of Toyota's stock prices for the year 2025 along with the 80% and 95% confidence intervals. The forecast suggests that average stock prices will experience a steady increase and the expanding dark regions exhibit greater uncertainty over time.

13 Conclusion and Strategic Implications

Initially, the data was explored to check the data types and missing values. There were no missing values in the Toyota Stock Data. Further, to prepare data for time series forecasting, the date column was transformed from character to date class using the R function. An ADF test was done to transform the data series from non-stationarity to stationarity. In the second half, the data was partitioned into training and testing to perform different forecasting models such as Naïve, STL and ETS, ARIMA, and Random Walk with Drift. Therefore, among the four forecasting models, RWF outperformed the model with the lowest RMSE (10.53), MAPE (3.78%), and MASE (0.27). Lastly, the paper presents the forecasted data points along with a monthly average forecast of Toyota's stock prices for the year 2025.

The forecasted stock prices can help Toyota with better and accurate decision-making. If the stock prices are foreseen to decline, Toyota can make adjustments in its production, marketing, or other investment activities to minimize risks. For instance, Toyota Motors might slow down the production rate based on the prediction or put new projects on hold. Even though the Random Walk with Drift model worked well in this project, it has some limitations. It does not seem to cope with sharp shifts resulting from economic meltdowns or major events. Furthermore, it ignores significant external variables like inflation and interest rates that may be crucial in predicting stock prices. Therefore, incorporating such factors can increase their accuracy.

14 Appendix: Estimated and Actual Time

Category	**Estimated Time (in hrs)**	**Actual Time (in hrs)**
Conducting Initial Exploratory Analysis	1.5	2
Perform Descriptive Analysis , Create Plots and Analyze Results	3	3.5
Perform correlation analysis on predictive variables and transform variables	2	3
Split the data set into training and test data sets	1	2
Forecasting Models (Naïve, ARIMA, STL, ETS & Random Walk with Drift)	3	8
Evaluate the models on test data	4	4
Analyze the best performing model and tune the parameters	3	4
Conduct final analysis, choose the best model , write first draft of the project data analysis	5	8
Finalize the data analysis report	3	2

References

A. Ariyo, A., & Adewumi, A. (2014). Stock Price Prediction Using the ARIMA Model. *UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge, UK*, 106-112.

Comincioli, B. (1996). The stock market as a leading indicator: An application of granger causality. *University avenue undergraduate journal of economics*, 31-40.

Fu, H. (2023). Analysis of Toyota's Operations to Maintain its Brand Position During the COVID-19 Pandemic. *BCP Business & Management*.

News, B. (1999, September 8). *INTERNATIONAL BUSINESS; Toyota Set to List Shares in U.S. and Britain*. Retrieved from The New York Times: https://www.nytimes.com/1999/09/08/business/international-business-toyota-set-to-list-shares-in-us-and-britain.html

Sonkavde, G., & Dharrao , D. (2023). Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications. *International Journal of Financial Studies*, 11(3), 94.