Pidilite Stock Prediction Model for Closing Price

PROBLEM IDENTIFICATION

Detailed Background

Pidilite Industries is renowned for its stronghold in the adhesives and sealants market in India, with its flagship product, Fevicol, being nearly synonymous with adhesives in the Indian subcontinent. The company has established a wide array of products catering to various segments, including industrial and consumer goods, arts and crafts, and construction chemicals, to name a few. The adhesive giant has experienced consistent growth and market penetration, capitalizing on strong brand recall and an expansive distribution network. However, Pidilite Industries faces the challenge of navigating through a volatile market environment, influenced by various factors such as raw material prices, competition, economic cycles, and market sentiment. The stock prices of Pidilite, like any publicly-traded company, are subject to these fluctuations, reflecting the company's performance, market dynamics, investor confidence, and speculative activities, among other elements.

Problem Statement

The problem at hand is multifaceted. On the one hand, it involves understanding the historical movement of Pidilite's stock prices to glean insights into its financial health and market behavior. On the other hand, it extends to forecasting future stock price trends to aid in investment decision-making processes. This analysis holds significance for stakeholders, including investors, market analysts, and the company's financial strategists. This scenario is apt for time series analysis because stock prices, by their very nature, are sequential data points indexed in time order. Time series analysis will enable us to decompose the historical stock data of Pidilite into its constituent components, identify any underlying patterns such as trends or seasonality, and detect any anomalies or outliers that could indicate extraordinary events impacting the stock's performance.

ANALYSIS AND INTERPRETATION

Exploratory Data Analysis and Initial Interpretation

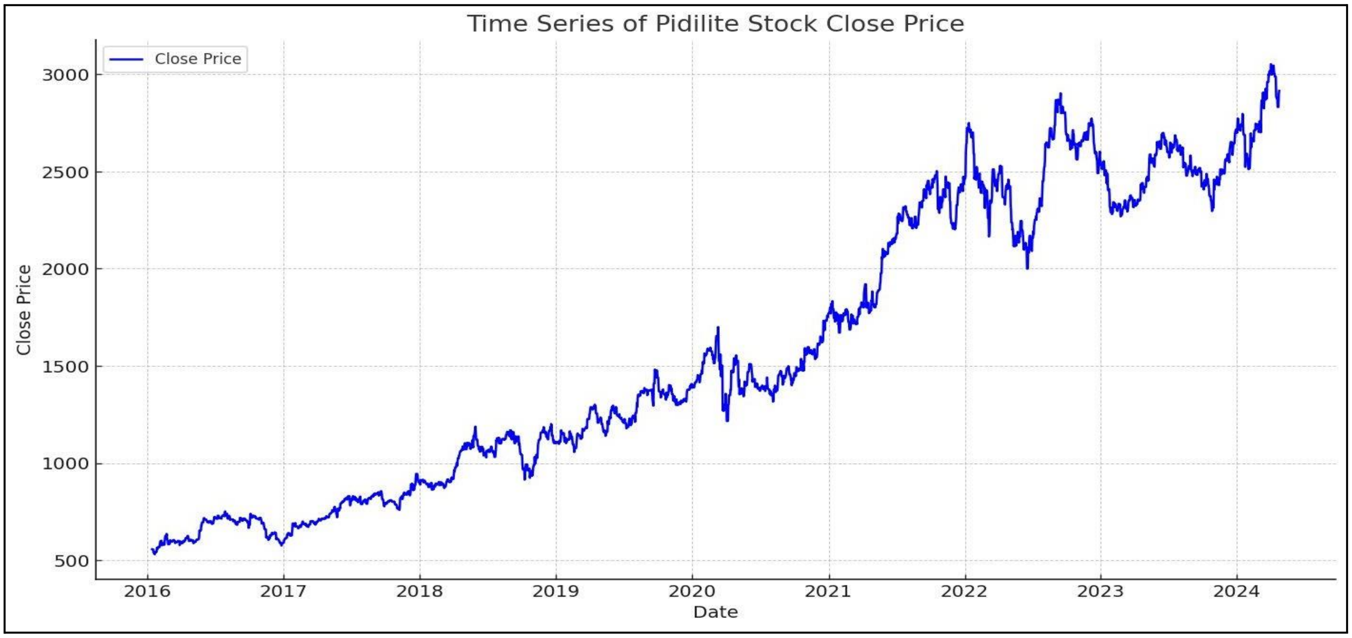


Fig - 1: Time Series of Pidilite Stock Close Price

The time series graph illustrates the stock price performance of Pidilite Industries Limited, a leading Indian manufacturer of adhesives and sealants. Over the span from 2016 to the beginning of 2024, the company's stock shows a notable long-term uptrend, reflecting investor confidence and a strong market position. This uptrend has been marked by periods of volatility, most significantly around 2020, likely influenced by the economic impact of the COVID-19 pandemic. Despite this, the stock has demonstrated recovery and growth post-pandemic. In early 2024, historical data from Yahoo Finance shows the stock price fluctuating but maintaining its general upward trajectory, with prices in late February hovering around the INR 2,700 mark. This indicates not only recovery but also potential resilience and growth in the face of market challenges. Investors might view this as a sign of a robust business model and strategic agility in adapting to market dynamics. The data points in the graph, when matched with historical stock price data and market capitalization information from sources like NSE India, would be integral for investors to evaluate the stock's performance and make informed decisions about their portfolios.

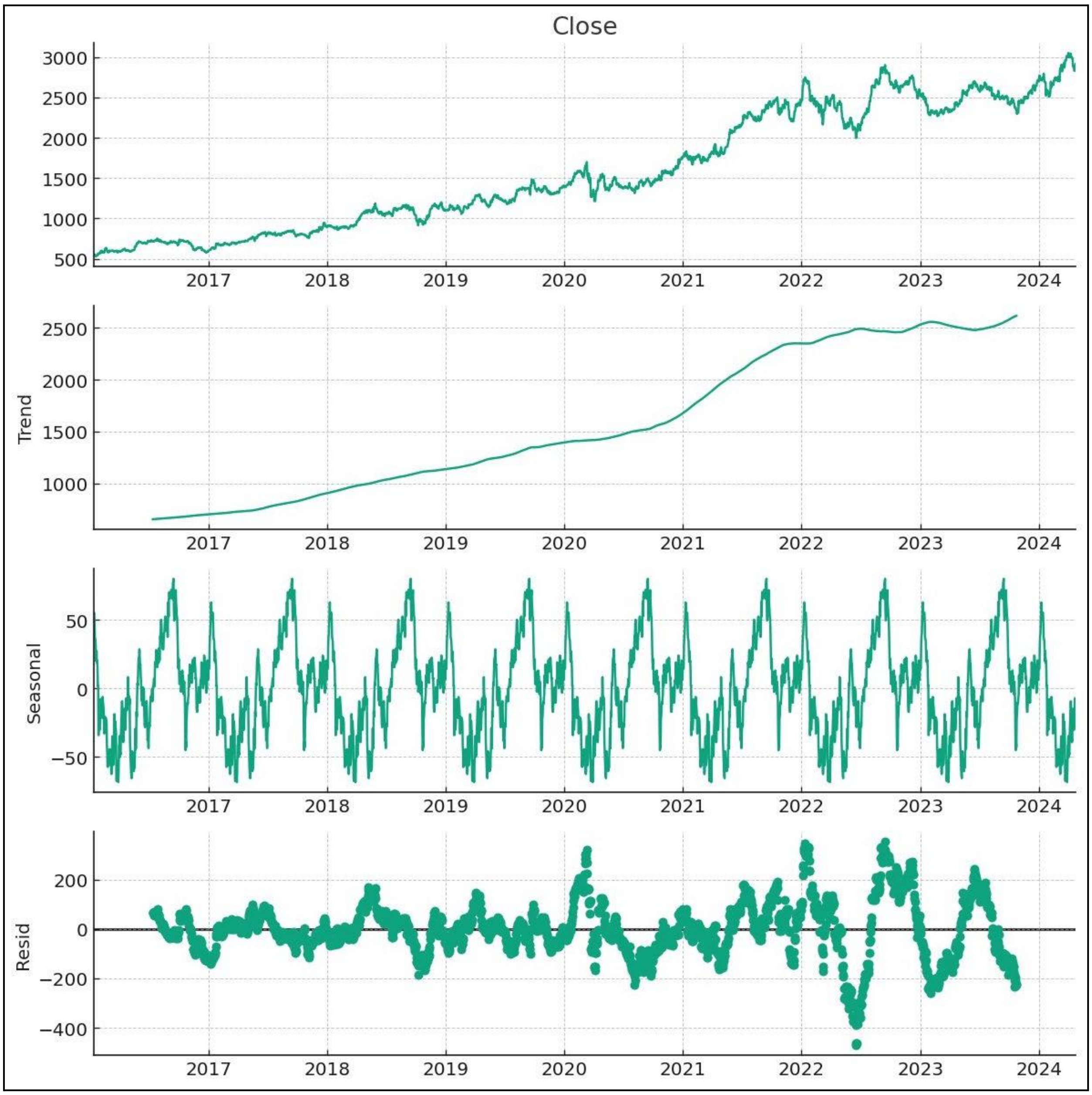


Fig - 2 : Seasonal Decomposition

Initial interpretations from the seasonal decomposition suggest that Pidilite's stock shows a strong and persistent upward trend, indicative of a growing company with potentially increasing profitability and market share. The seasonal fluctuations may imply periodic factors affecting the stock, which could be associated with the company's earnings cycle, budgetary announcements affecting the sector, or seasonal demand variations for its products. The residual component, which exhibits some structure, points towards complex market dynamics that a simple additive model might not be capturing. This warrants a deeper investigation, perhaps with more sophisticated time series modeling techniques, to improve the understanding and forecasting of Pidilite's stock price movements.

The seasonal decomposition of the time series data has provided us with four plots, each representing a different component of the time series:

- **Observed**: This is the actual 'Close' time series data as it was observed.
- **Trend**: This plot shows the long-term progression of the series, illustrating a clear upward trend over time. This indicates that the average value of the series is increasing.
- **Seasonal**: The seasonal component captures the regular pattern of variability within the time series. In this case, you can observe fluctuations that seem to repeat annually, which could be due to factors like seasonality in the market or other annual cycles affecting the 'Close' values.
- **Residual**: The residuals represent the noise or randomness of the series after the trend and seasonal components have been removed. Ideally, these should resemble white noise, which would indicate that the trend and seasonal components have successfully captured the systematic information in the time series. In this case, there appears to be some structure in the residuals, suggesting that the additive model might not be capturing all the patterns in the data, especially in the latter part of the series where the volatility seems to increase.

Applying a log transformation to Pidilite's stock data makes sense because it helps make the data easier to work with. It evens out repeating patterns like seasonal changes, and lets us see changes as percentages instead of raw numbers. Since the stock shows a steady climb and some tricky patterns in leftover data, using the log transformation helps tone down extreme swings and makes it easier to predict future movements. This step is important for better understanding how Pidilite's stock behaves and for making more accurate forecasts.

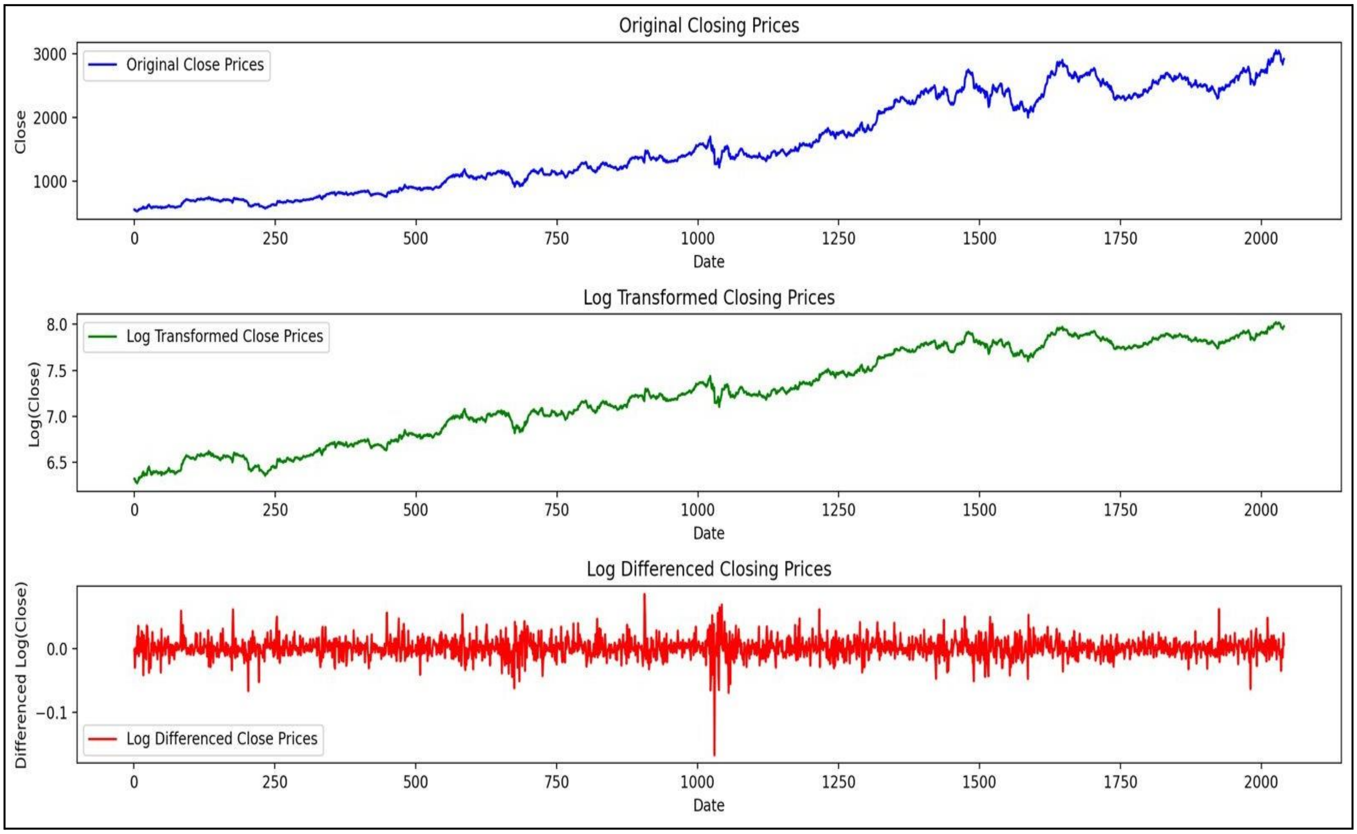


Fig - 3: Log Transformation Plots

The second plot shows the time series after the log transformation. The exponential growth visible in the original data has been linearized, indicating a steady trend over time. However, the presence of a trend implies that the series is not yet stationary, which is a key assumption for many time series forecasting methods.

To achieve stationarity, we apply differencing to the log-transformed data. Differencing is a method that subtracts the previous observation from the current observation. In the case of stock prices, a first-order difference can remove trends and cycles, leaving behind a stationary series that fluctuates around a constant mean.

The third plot exhibits the log-differenced stock prices, where we can observe that the trend has been removed, and the transformed data appears to fluctuate around a mean of zero with constant variance over time. This stationary series is what we aim for before moving forward with any further time series analysis or forecasting. It indicates that the series is now mostly composed of noise, and any autocorrelation present can be attributed to random fluctuations rather than systematic changes in the data.

Detailed Analysis

To capture the underlying patterns and dynamics in the 'Close' prices of the dataset, an ARIMA (AutoRegressive Integrated Moving Average) model was employed. However, instead of directly modeling the 'Close' prices, the natural logarithm of the 'Close' prices was utilized. This transformation helps stabilize the variance of the time series and ensures that the model captures relative changes rather than absolute values.
Following the utilization of the ARIMA model to analyze the log of 'Close' prices, we transitioned to assessing the autoregressive (p) and moving average (q) parameters using ACF and PACF plots. These diagnostic tools offer insights into the correlation between observations at various lags. Significant autocorrelation in the ACF plot suggests the inclusion of lag observations in the AR term, while significant partial autocorrelation in the PACF plot indicates the inclusion of lag observations in the MA term. Since the series was stationary after differencing once, we opted for d=1. By analyzing these plots, we identified optimal values for p, q, and d, enhancing the accuracy and reliability of the ARIMA model.

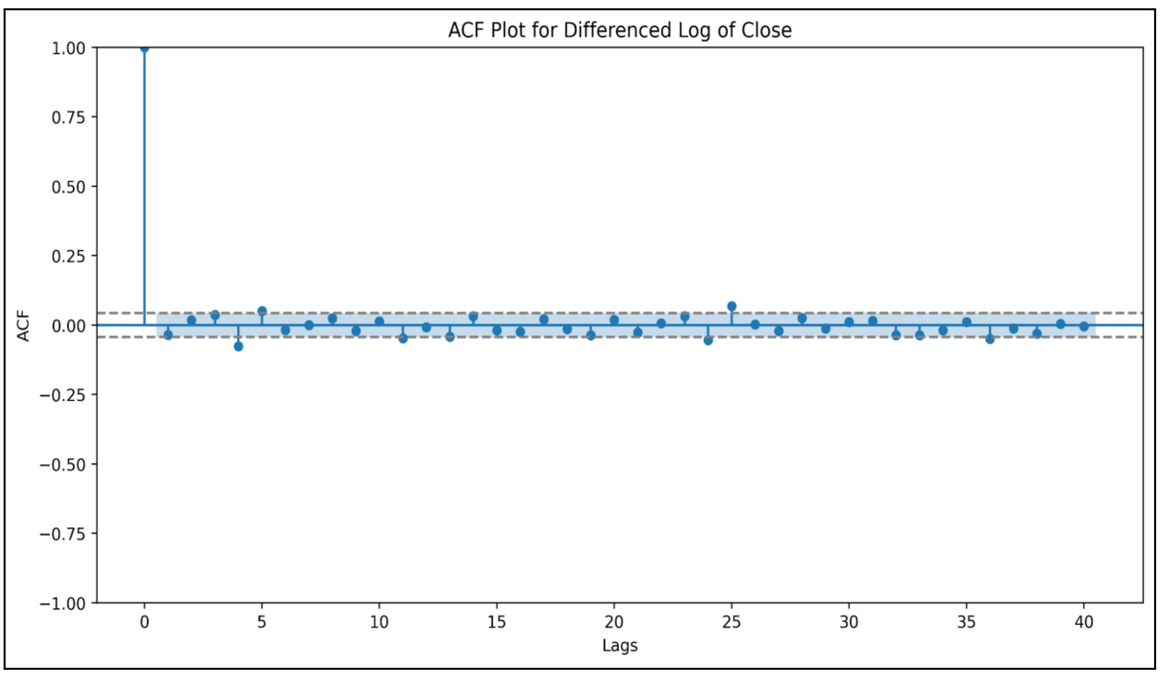


Fig - 4 : ACF Plot for Differenced Log of Close

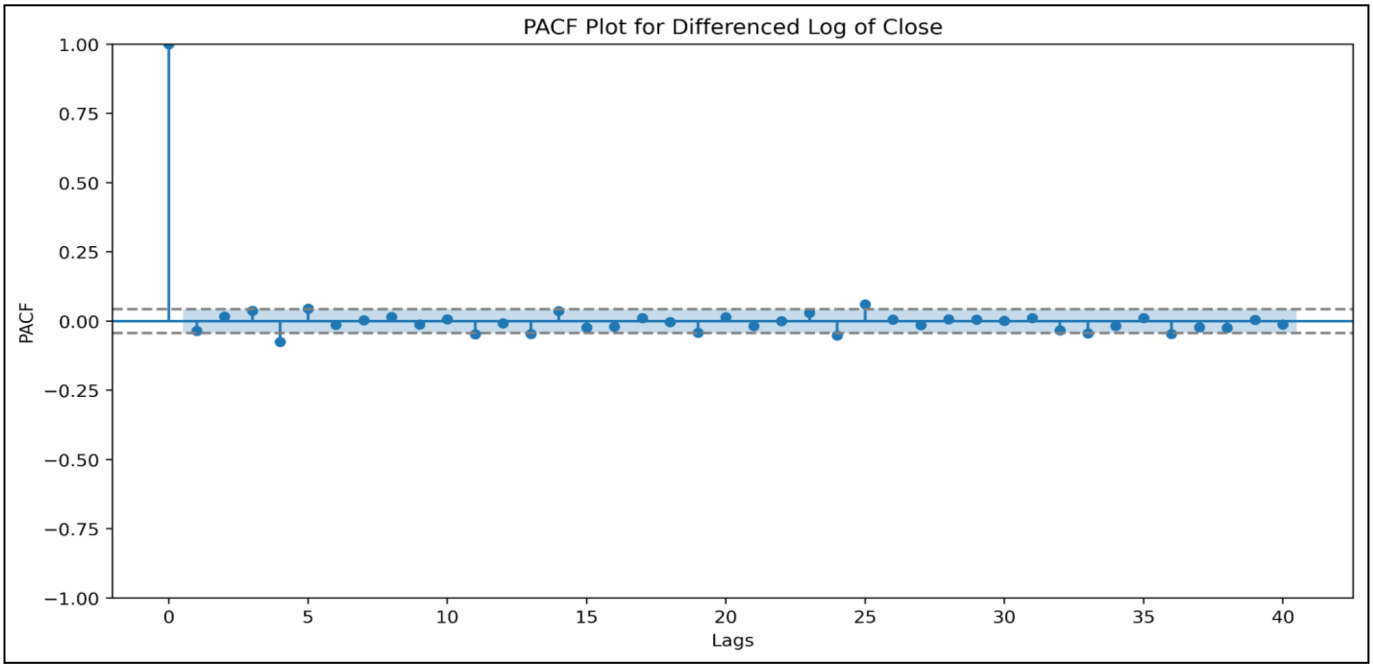


Fig - 5: PACF Plot for Differenced Log of Close

These figures represent the ACF and PACF plots for a time series dataset. The ACF plot illustrates the correlation of the time series with its own past values. The first lag at zero is inherently equal to 1, representing a perfect correlation with itself, which is a standard feature of ACF plots and typically not included in the analysis. The remaining lags in the ACF plot show that autocorrelations quickly drop off, indicating that there is little to no correlation between the time series data and its past values beyond the immediate lag.

The PACF plot, on the other hand, illustrates the partial correlation of the series with its own lagged values, controlling for the values of the time series at all shorter lags. The sharp drop after the first non-trivial lag suggests that the series is possibly a moving average process of order one, MA(1), since the partial autocorrelations after the first lag are essentially zero. This indicates that past values have no linear predictive power on future values of the time series beyond the first lag. Both plots reveal a time series with minimal autocorrelation, suggesting that the data may be close to white noise. This absence of significant peaks in both ACF and PACF, except for the excluded initial lag, implies that the time series does not exhibit strong seasonal or trend components that would otherwise be indicated by distinct spikes at specific lags

Model Training for ARIMA

To determine the optimal ARIMA model for the log-transformed stock price data, a comprehensive stepwise search was conducted across various combinations of autoregressive (AR), differencing (I), and moving average (MA) components. The goal was to minimize the Akaike Information Criterion (AIC) value, which is a widely-used metric for model selection where a lower AIC indicates a model with a better balance between goodness of fit and complexity. Each model configuration was systematically assessed to estimate its parameters with an intercept, capturing the mean level of the series after accounting for autoregressive and moving average dynamics. The AIC values from this process guide us toward the most parsimonious model that adequately describes the underlying stochastic process of the log-transformed time series data.

Model	**AIC**
ARIMA(1,1,1)	-11160.365
ARIMA(0,1,0)	-11162.256
ARIMA(1,1,0)	-11162.773
ARIMA(0,1,1)	-11162.680
ARIMA(2,1,0)	-11161.302
ARIMA(2,1,1)	-11159.344

Fig - 6

The table shows the AIC values for several ARIMA model configurations with an intercept. It is evident that the models with a single differencing term are competitive, reflecting the importance of differencing in stabilizing the variance in the log-transformed series. The ARIMA(1,1,0) model, in particular, emerges with the lowest AIC, suggesting it as the optimal model among those considered. This model's

parsimony and performance indicate it may be a suitable candidate for forecasting future stock price movements on the log scale.
Here is the Model Summary for ARIMA(1,1,0):

Metric	Value
Dependent Variable	Log_Close
No. Observations	2,041
Model	SARIMAX(1, 1, 0)
Log Likelihood	5581.164
AIC	-11156.328
BIC	-11139.467

Fig - 7

$$\boxed{Y_t = 0.0008 + Y_{t-1} - 0.0351(Y_{t-1} - Y_{t-2}) + \varepsilon_t}$$

This equation represents an ARIMA(1, 1, 0) model with a constant term.
let's interpret the coefficients in the ARIMA(1, 1, 0) model:

1. **Intercept (0.0008):**
   The intercept represents the expected value of the dependent variable (in this case, Y_t) when all independent variables are zero. However, in a time series context like this, it might not have a direct interpretation as "when all independent variables are zero" might not make sense. Instead, it could represent a baseline or starting value for the series.

2. **ARIMA(1) Coefficient (-0.0351):**
   This coefficient is associated with the first-order autoregressive term. It indicates the impact of the previous time step's value Y_{t-1} on the current value Y_t . In this

case, a coefficient of -0.0351 suggests that there is a negative relationship between the previous time step's value and the current value. Specifically, it suggests that for every unit increase in Y_{t-1}, Y_t is expected to decrease by approximately 0.0351 units, all else being equal.

3. **Moving Average (MA) Coefficient:**
   In this model, there is no moving average term (MA) included, so there is no coefficient for the MA term to interpret.

Overall, this ARIMA(1, 1, 0) model suggests that there is a negative relationship between the previous time step's value and the current value, with a small intercept term to adjust the starting value of the series.

The plot initially depicted the data using a logarithmic scale, effectively compressing the range of values by applying a logarithmic transformation to the y-axis. This transformation is beneficial when visualizing data with a wide range of values, as it helps to emphasize relative changes across the entire spectrum of the data. However, to provide a clearer interpretation of the absolute magnitude of the values, the plot was subsequently converted to a regular scale. By reverting to a regular scale, each unit along the y-axis represents a consistent interval in the original measurement scale, enabling a straightforward interpretation of the absolute values.

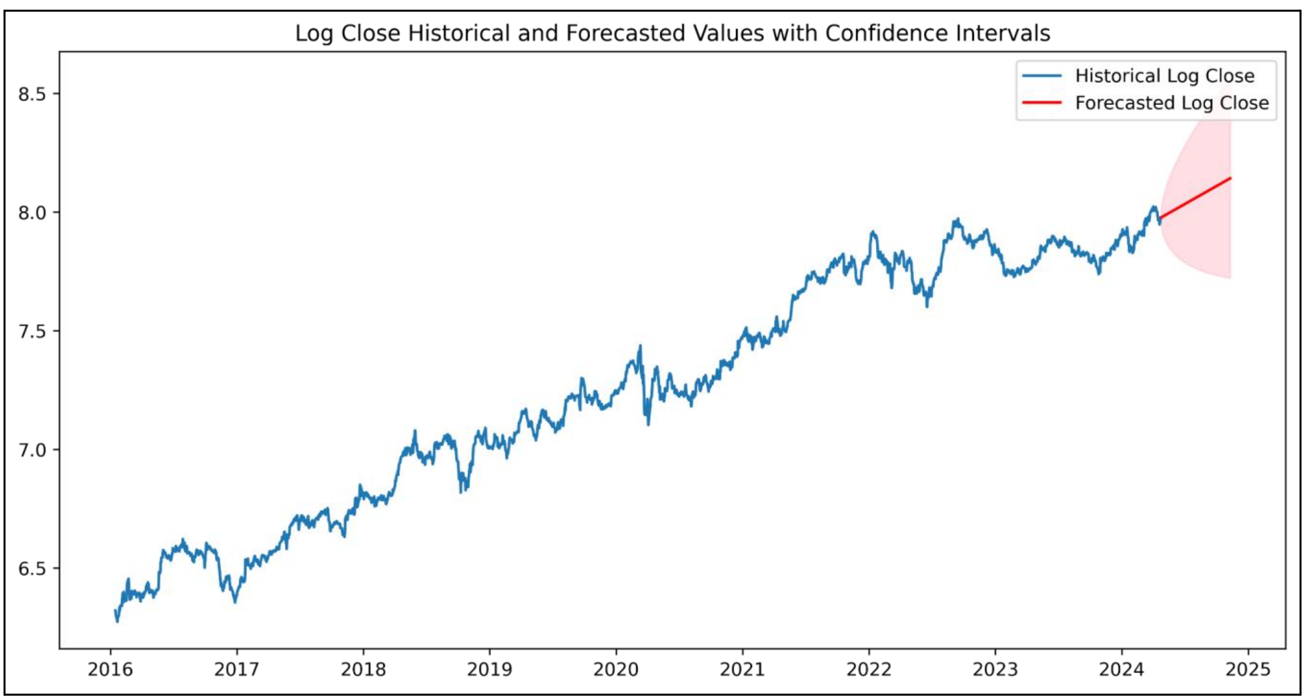


Fig - 8: Log CloseHistorical and Forecasted Values with Confidence Intervals

This transformation from a logarithmic scale to a regular scale allows for a more intuitive understanding of the data, particularly when conveying the precise numerical values and trends present in the dataset.

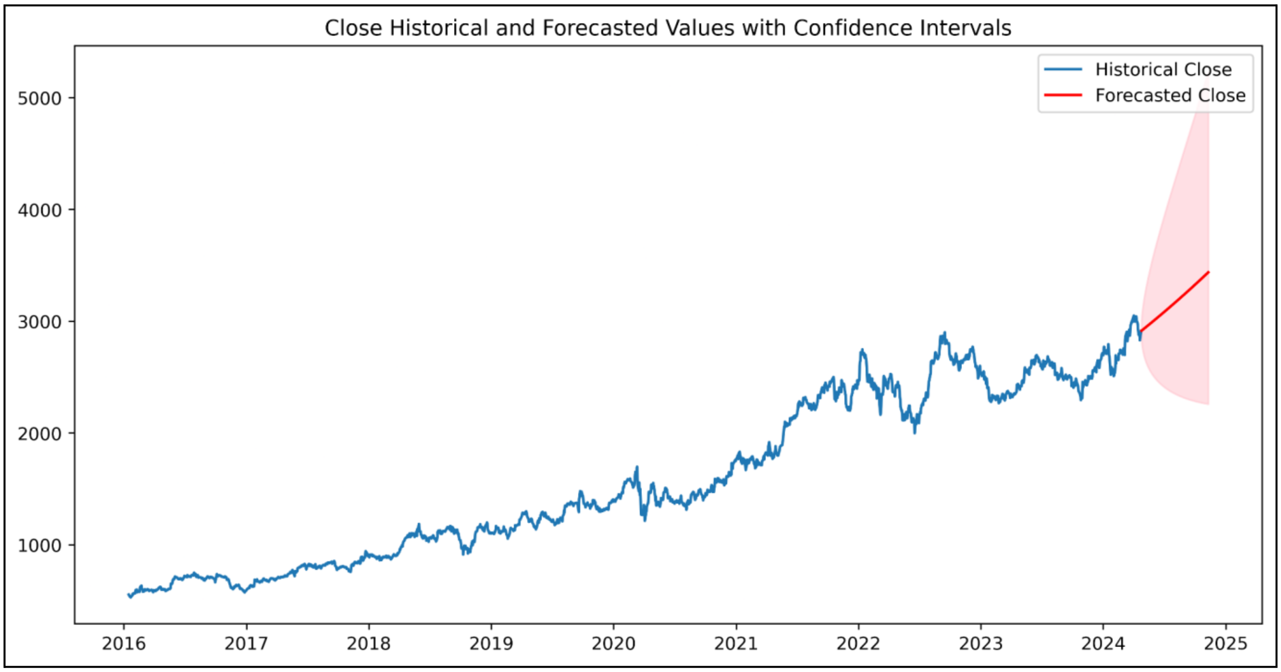


Fig - 9: Close Historical and Forecasted Values with Confidence Intervals

The graph visualizes stock price movements, with the blue line showing actual closing prices that exhibit a general increasing trend with some volatility. The red line extends from the last actual data point, representing the predicted future closing prices.

Notably, the shaded pink area around the forecasted prices indicates a 95% confidence interval, illustrating the range within which future prices are expected to lie with 95% certainty. As we move further into the future, this confidence interval widens, reflecting greater uncertainty in the prediction as time extends away from the last observed data point.

Diagnostic Checks for ARIMA

Test	Statistic	P-Value	Interpretation
Ljung-Box (L1) (Q)	0.00	0.99	No autocorrelation (good fit)
Jarque-Bera (JB)	5878.74	0.00	Distribution is not normal
Heteroskedasticity(H)	0.88	0.10	No heteroskedasticity (good fit)
Skew	-0.48	-	Slight negative skew
Kurtosis	11.26	-	High peak compared to normal dist

Fig - 10

Following the application of the ARIMA (AutoRegressive Integrated Moving Average) model, an investigation into Holt-Winters' double exponential smoothing method was undertaken to further explore its efficacy in time series forecasting. ARIMA, while robust, primarily addresses stationary data and may not fully capture trend dynamics present in the dataset. Recognizing Holt-Winters' capability to handle trend components alongside level fluctuations, it emerged as a promising alternative for modeling the 'Close' column data. This method's ability to discern trend patterns without the requirement of seasonality aligns well with the observed characteristics of the dataset. By delving into Holt-Winters' double exponential smoothing, we sought to capitalize on its capacity to provide nuanced insights into the data's evolving trends, thereby enhancing our forecasting capabilities.

Through this exploration, we aimed to leverage the strengths of both ARIMA and Holt-Winters methodologies to gain a more comprehensive understanding of the dataset's temporal dynamics and improve the accuracy of future projections.

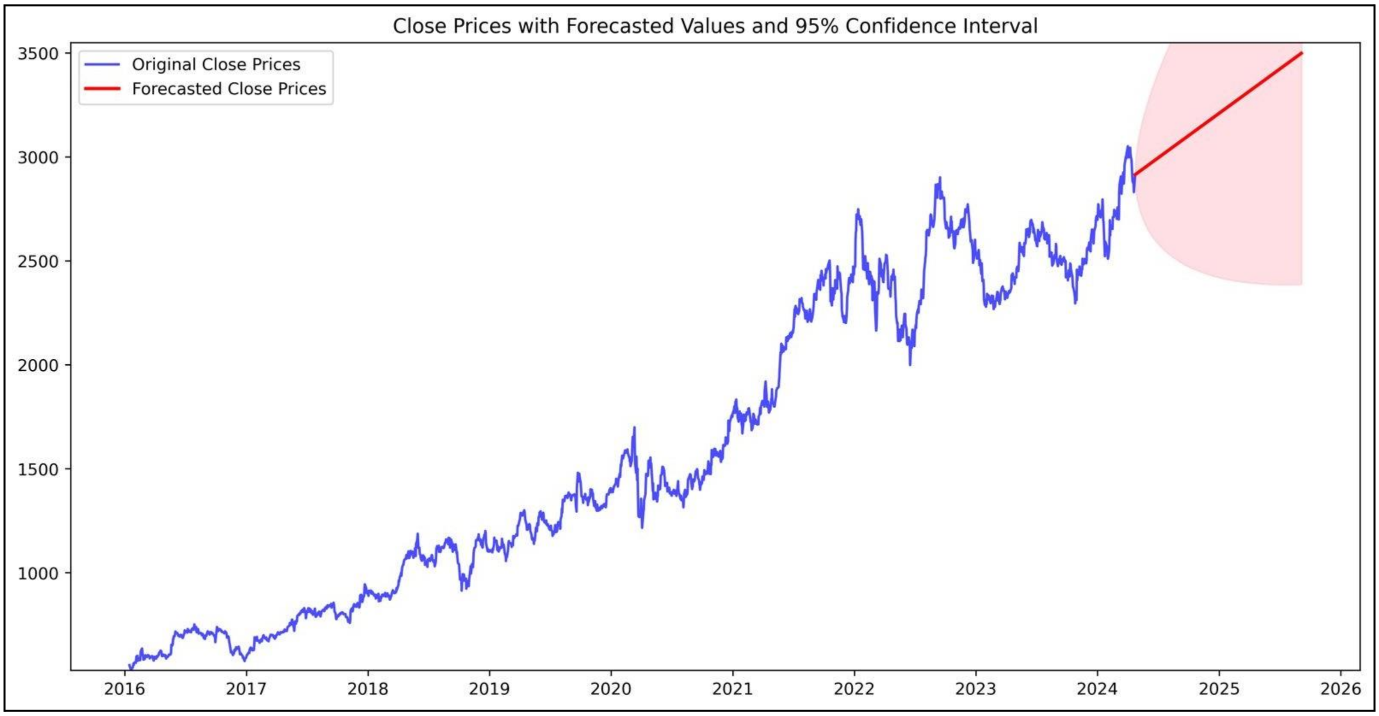


Fig - 11: Close prices with Forecasted Values and 95% Confidence Interval

The plot exhibits the historical and projected closing stock prices for Pidilite Industries, leveraging a Holt-Winters forecasting model. Pidilite's historical data, represented by the blue line, indicates a robust upward trajectory, suggesting consistent growth in the company's valuation. The model's strength lies in capturing and projecting trends evident in the historical data, which is crucial for a stock like Pidilite's that has shown long-term appreciation.

As the plot transitions from historical to forecasted data, the red line indicates the expected future closing prices. This projection continues the established upward trend, suggesting that the model anticipates further growth in Pidilite's stock value. The accompanying shaded pink area, broadening with the forecast horizon, represents the model's 95% confidence interval.

This expanding confidence interval reflects greater uncertainty in the longer-term price predictions, a common occurrence in forecast models due to compounding potential variances over time.

Results	Value
Dep. Variable	Close
No. Observations	2040
Model	Exponential Smoothing
SSE	1365591.132
AIC	13281.041
Trend	Additive
BIC	13303.524
Seasonal	None
AICC	13281.083

Fig - 12

Component	Coefficient	Optimized
Smoothing Level	0.9711557	TRUE
Smoothing Trend	9.9944E-05	TRUE
Initial Level	534.74383	TRUE
Initial Trend	1.1672139	TRUE

Fig - 13

Smoothing Level: The coefficient for the smoothing level represents the weight assigned to the current observed value when updating the level component of the model. In this case, the coefficient is approximately 0.971, indicating that the model heavily weighs the most recent observation when updating the level component.

Smoothing Trend: The coefficient for the smoothing trend represents the weight assigned to the current observed trend when updating the trend component of the model. Here, the coefficient is approximately 9.9944e-05 (or 0.000099944), indicating that the model assigns very little weight to the most recent observed trend when updating the trend component.

Initial Level: This coefficient represents the initial level estimate of the time series. In this case, the initial level estimate is approximately 534.744.

Initial Trend: The coefficient for the initial trend represents the initial trend estimate of the time series. Here, the initial trend estimate is approximately 1.167.

Comparison of Models

Model	RMSE	MAE	MAPE
ARIMA	28.66239594090730	17.658103060783800	1.1645772334106300
Holt-Winters	25.872908950614500	17.3734147255636	1.115198782599440

Fig - 14

To analyze the results of the RMSE, MAE, and MAPE for comparison between the ARIMA and Holt-Winters models:

1. **RMSE (Root Mean Squared Error):**
   - ARIMA: 28.66
   - Holt-Winters: 25.87

The RMSE measures the average magnitude of the errors between predicted and observed values. A lower RMSE indicates better accuracy. In this case, the Holt-Winters model has a lower RMSE compared to ARIMA, suggesting that it provides slightly better predictive performance.

2. **MAE (Mean Absolute Error):**
- ARIMA: 17.66
- Holt-Winters: 17.37

The MAE measures the average magnitude of the errors between predicted and observed values, regardless of direction. Similar to RMSE, a lower MAE indicates better accuracy. Here, the Holt-Winters model also has a lower MAE, further indicating slightly better predictive performance compared to ARIMA.

3. **MAPE (Mean Absolute Percentage Error):**
- ARIMA: 1.16%
- Holt-Winters: 1.12%

The MAPE measures the average percentage difference between predicted and observed values. A lower MAPE indicates better accuracy, with values closer to zero indicating better performance. Both models have relatively low MAPE values, indicating good predictive accuracy. However, the Holt-Winters model has a slightly lower MAPE, suggesting slightly better performance in terms of percentage accuracy compared to ARIMA.

In summary, while both models perform reasonably well in terms of predictive accuracy, the Holt-Winters model generally outperforms ARIMA based on all three metrics (RMSE, MAE, and MAPE). However, the differences in performance between the two models are relatively small.

Therefore, the choice between ARIMA and Holt-Winters may depend on other factors such as ease of interpretation, computational efficiency, and the specific characteristics of the dataset.

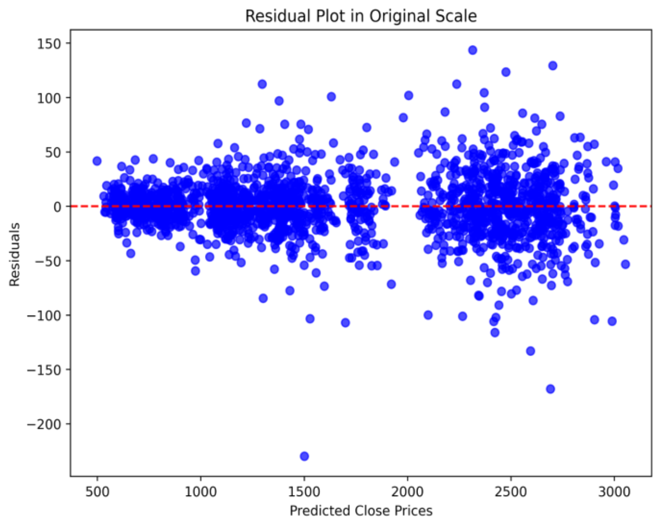
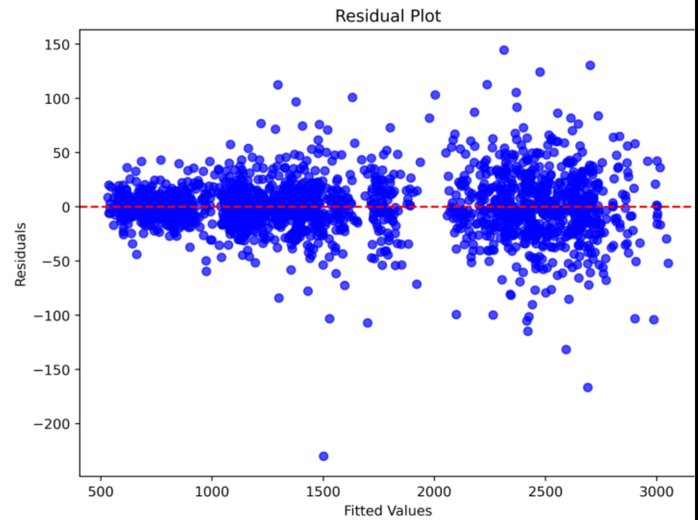
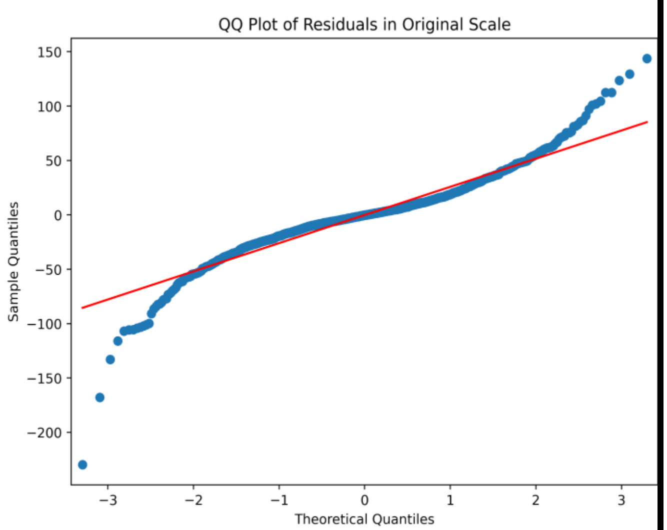
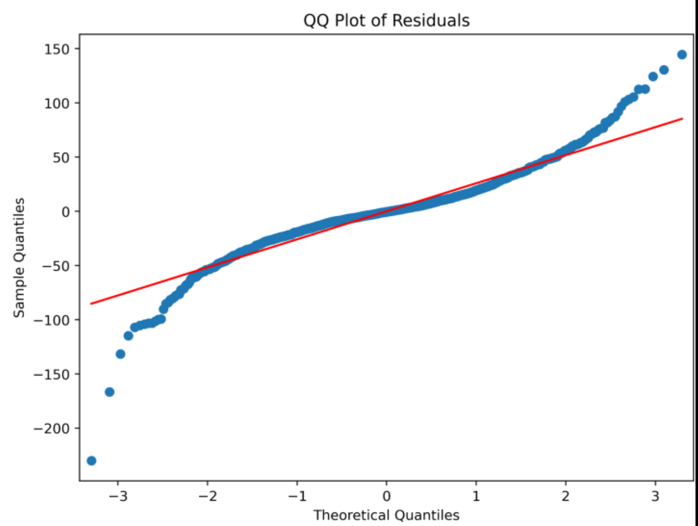
<table>
<tr><th>PLOTS</th><th>ARIMA (1,1,0)</th><th>HOLT WINTERS</th></tr>
<tr><td>RESIDUAL</td><td>
</td><td>
</td></tr>
<tr><td>QQ-PLOT</td><td>
</td><td>
</td></tr>
</table>

Fig - 15

The **Residual Plots** for both the ARIMA (1,1,0) and Holt-Winters models indicate that residuals are scattered around the zero line, which implies that neither model displays systematic bias. Outliers are present in both residual plots, with the Holt-Winters showing a slightly more pronounced spread. The **Q-Q Plots** for both models demonstrate a similar deviation from normality, particularly in the tails, indicating a heavier-than-normal distribution of residuals. This resemblance suggests that both models have a comparable capacity in handling the data, with similar limitations. The deviations in the Q-Q plots and the outliers in the residual plots may point to the presence of nonlinearity or extreme values in the time series that are not fully captured by either linear model.

CONCLUSION

In evaluating the performance of ARIMA (1,1,0) and Holt-Winters models on the Pidilite Industries stock price data, the residual plots suggest a competent capture of the general trend, with residuals dispersed around the zero line. Yet, the presence of outliers and the deviation from normality in the Q-Q plots indicate that both models may not fully account for the financial time series' volatility and extreme value conditions. The similarity in the residual patterns of both models implies a shared limitation in addressing these complexities.

The utilization of machine learning techniques such as Random Forests or Gradient Boosting Machines presents an opportunity to transcend these limitations. These methods excel in deciphering non-linear and complex relationships within data without the constraints of conventional distributional assumptions, offering robustness against anomalies and variable interdependencies.

For a more precise modeling of the volatility patterns often observed in financial markets, the application of Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models is advised. These models are particularly adept at capturing and forecasting dynamic variances, thereby directly addressing the volatility clustering that standard models might overlook.Moreover, robust statistical techniques like quantile regression could provide a comprehensive view of the potential influence of predictors across different points of the conditional distribution of stock prices, offering a more detailed depiction than mean-based models.

To encapsulate, while the ARIMA and Holt-Winters models have laid a solid foundation for understanding Pidilite Industries' stock price movements, it is recommended to embrace more advanced machine learning and econometric models, including GARCH, for enhanced analysis. These models are anticipated to handle the peculiar traits of financial time series with greater efficacy, leading to superior forecasting accuracy and a deeper insight into the stock's volatility, which is vital for informed decision-making in financial planning and risk assessment.

TOOLS USED

We used Python to generate the results and the outputs. The outputs were provided in the analysis:

- Time Series of Pidilite Stock Close Price
- Seasonal Decomposition
- Log Transformation Plots
- ACF Plot for Differenced Log of Close
- PACF Plot for Differenced Log of Close
- AIC Values for ARIMA Models
- Model Summary for ARIMA (1,1,0)
- Log Close Historical and Forecasted Values with Confidence Intervals
- Close Historical and Forecasted Values with Confidence Intervals
- Residual and QQ Plots for ARIMA (1,1,0) and Holt-Winters Models

These figures were generated using Python, which served as the primary analytical tool throughout the study.