****

**CS713 Project Report:**

**Analyzing and Forecasting Zomato Stock Prices**

Name: Dwij Siyal

Student ID: 200492977

**Abstract**

This project addresses the challenge of volatility in Zomato's stock price since its Initial Public Offering (IPO) in July 2021. Investors face difficulties in making informed decisions due to fluctuating nature of stock prices and a lack of comprehensive analysis. This study aims to bridge this gap by conducting a detailed analysis of Zomato’s stock price data from July 2021 to February 2024. The analysis involves four key phases: Exploratory Data Analysis (EDA), Anomaly Detection, Time Series Forecasting, and Result Visualization.

The EDA phase focuses on understanding the stock price distribution, volatility, and identifying long-term trends and cyclical patterns using visual tools like line charts and histograms. Anomaly detection employs techniques such as the Sign Method and PCOut Method to identify outliers in the stock price data, which may indicate unusual market events or errors. Time series forecasting uses the AutoRegressive Integrated Moving Average (ARIMA) model to predict future stock prices and volumes. This phase emphasizes optimizing model parameters to achieve accurate predictions, evaluated using metrics like Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE).

The project culminates in result visualization, presenting the findings through clear charts and graphs that highlight data trends, anomalies, and forecasted stock price movements. These insights are crucial for investors to make informed decisions, minimize risks, and optimize their investment strategies. The expected outcomes include a comprehensive understanding of stock price trends, identification of anomalies, and the development of an accurate forecasting model, all conveyed through effective visualizations.

Table of Contents

[**Introduction** 4](#_Toc169820502)

[**Problem Statement** 4](#_Toc169820503)

[**Data Description** 5](#_Toc169820504)

[**Methodology** 6](#_Toc169820505)

[**Results and Discussion** 10](#_Toc169820506)

[**Conclusion** 29](#_Toc169820507)

[**Appendices** 31](#_Toc169820508)

**List of Figures and Tables**

[Figure 1. 1 Dataset Head 11](#_Toc169733551)

[Figure 1. 2 Dataset Tail 11](#_Toc169733552)

[Figure 1. 3 General Schema of Dataset 12](#_Toc169733553)

[Figure 1. 4 Check for nulls 12](#_Toc169733554)

[Figure 1. 5 Descriptive Statistics 13](#_Toc169733555)

[Figure 1. 6 Colum Distributions 16](#_Toc169733556)

[Figure 1. 7 Pairwise Scatter Plots 17](#_Toc169733557)

[Figure 1. 8 Volume vs Price change 18](#_Toc169733558)

[Figure 1. 9 Difference of prices relationship 18](#_Toc169733559)

[Figure 1. 10 Close Price over time 19](#_Toc169733560)

[Figure 1. 11 Candlestick chart 19](#_Toc169733561)

[Figure 2. 1 Sign1 method 22](#_Toc169771275)

[Figure 2. 2 Sign2 method 22](#_Toc169771276)

[Figure 2. 3 PCOut method 23](#_Toc169771277)

[Figure 3. 1 Stationary Series 25](#_Toc169771252)

[Figure 3. 2 Autocorrelation 25](#_Toc169771253)

[Figure 3. 3 Partial Autocorrelation 26](#_Toc169771254)

[Figure 3. 4 Predicted and Actual values 27](#_Toc169771255)

[Figure 3. 5 RMSE and MAPE 27](#_Toc169771256)

[Figure 3. 6 Forecasted Values 28](#_Toc169771257)

## **Introduction**

Zomato, founded in 2008 by Deepinder Goyal and Pankaj Chaddah, is a prominent Indian multinational corporation specializing in restaurant aggregation and food delivery services. By 2022–23, Zomato had expanded its operations to over 1,000 cities across India, offering comprehensive restaurant information, menus, user reviews, and food delivery from partner establishments. In July 2021, Zomato's Initial Public Offering (IPO) saw its shares priced at INR 115 on the Bombay Stock Exchange (BSE) and INR 116 on the National Stock Exchange (NSE) of India, attracting significant investor interest because of its promising returns and inherent market volatility.

This project focuses on a comprehensive analysis of Zomato's stock price data from July 2021 to February 2024, sourced from Kaggle. The key objectives are to detect anomalies and develop predictive models using the AutoRegressive Integrated Moving Average (ARIMA) methodology. The analysis will proceed in stages: Exploratory Data Analysis (EDA), anomaly detection, and time series forecasting.

Understanding the trends and anomalies in Zomato's stock prices will provide crucial insights for investors, helping them make informed decisions and enhancing strategic planning by identifying patterns indicative of future stock performance, thus mitigating risks and optimizing investment outcomes.

## **Problem Statement**

Despite Zomato's rapid expansion and significant market presence in the food delivery industry, its stock price has exhibited notable volatility since its Initial Public Offering (IPO) in July 2021. The fluctuating stock prices pose a challenge for investors seeking to make informed decisions.

Currently, there is a lack of comprehensive analysis of Zomato’s stock performance that addresses the detection of price anomalies and the forecasting of future trends. Without accurate predictive models and a thorough understanding of the factors influencing stock price movements, investors may struggle to identify optimal investment opportunities and mitigate potential risks.

This project aims to fill this gap by conducting an in-depth analysis of Zomato’s stock price data, covering the period from July 2021 to February 2024. The analysis will utilize advanced techniques such as Exploratory Data Analysis (EDA), anomaly detection, and time series forecasting with ARIMA models. The goal is to provide actionable insights that will help investors make better-informed decisions, reduce the risk of losses, and enhance their investment strategies in a volatile market environment.

## **Data Description**

The dataset utilized for this project is sourced from Kaggle, a well-known platform for data science and machine learning resources. This dataset provides a comprehensive overview of Zomato's stock prices over a specified period and comprises seven columns of information, each encapsulating critical aspects of the stock's daily trading activities. The schema of the dataset is as follows:

* **Date (mm-dd-yyyy)**: This column records the date for each entry, following the format of month, day, and year.
* **High (decimal)**: This column captures the highest trading price of Zomato's stock on a given day, recorded as a decimal.
* **Open (decimal)**: This field indicates the opening price of the stock for the day, represented in decimal form.
* **Close (decimal)**: This entry shows the closing price of the stock at the end of the trading day, also recorded as a decimal.
* **Low (decimal)**: This column details the lowest trading price of the stock for the day, formatted as a decimal.
* **Adj Close (decimal)**: This field reflects the adjusted closing price, which accounts for the adjusted closing price for the trading day, presented in decimal format.
* **Volume (decimal)**: This column denotes the total number of shares traded on that day, represented in decimal form.

Before proceeding with any analysis, it is crucial to address the integrity of the dataset. The initial step involves checking for null or missing values across all columns. Incomplete or missing data can significantly affect the accuracy and reliability of subsequent analyses and forecasts. To address any missing values, the forward fill technique will be employed. This method involves propagating the last observed valid value forward to fill any subsequent gaps. By ensuring that the dataset is complete and free from missing values, we set a robust foundation for the accurate analysis and modelling of Zomato's stock price data.

## **Methodology**

1. **Exploratory Data Analysis (EDA):**

The steps undertaken for conducting Exploratory Data Analysis (EDA) are outlined as follows:

**Data Inspection, Understanding, and Cleaning**: Initial examination of the dataset to identify any inconsistencies, missing values, or outliers, followed by data cleaning procedures to ensure data integrity and readiness for analysis.

**Calculation of Descriptive Statistics**: Computation of fundamental statistical measures such as mean, median, minimum, maximum, and count to summarize the central tendency, variability, and distribution of the dataset.

**Plotting Column Distributions**: Visual representation of the distribution of individual columns to understand their spread and identify any potential anomalies or patterns.

**Scatter Plot Analysis**: Creation of scatter plots for various combinations including:

**Open vs. Close Prices**

**Price Change (|Close - Open|) vs. Volume**

**Date vs. Close Price**

**Date vs. Open Price**

**Close - Open vs. High - Low**

**Close - High vs. Open - Low**

These plots are utilized to discern relationships and trends between variables.

**Visualization Tools**: Primarily, Python’s matplotlib and seaborn libraries are employed to generate these visualizations, providing insightful graphical representations of the data.

**Analyzed Metrics (using Descriptive statistics)**:

**Average Daily Price**: The mean stock price over a given period, offering insight into the overall price level.

**Price Volatility**: Measured through the standard deviation of stock prices, indicating the degree of fluctuation and uncertainty.

**Trading Volume**: The total number of shares traded, which reflects market activity and liquidity.

**Price Trends**: Examination of metrics such as moving averages and growth rates to assess the direction and strength of price movements, often represented graphically.

These steps collectively facilitate a comprehensive understanding of the data, enabling the identification of key trends and anomalies that inform subsequent analysis.

1. **Anomaly Detection**

Anomaly detection is a method used to identify outliers within a dataset, which allows for the management of these anomalies by either removing them or replacing them with average values. However, in the context of this project, anomalies will not be removed because stock market data inherently includes such outliers, which are crucial for accurately forecasting future values due to the market's inherent volatility.

The techniques employed in this project for anomaly detection are the Sign Method (sign1, sign2) and the PCOut method. The Sign Method utilizes Mahalanobis distance and principal components to calculate distances, making it particularly effective in detecting anomalies within multivariate datasets. These methods are chosen for their proficiency in identifying irregularities in complex data structures, thereby ensuring robust anomaly detection.

1. **Time Series Forecasting**

The AutoRegressive Integrated Moving Average (ARIMA) model is a prominent statistical technique for time series forecasting. It integrates three key components: AutoRegressive (AR), which regresses the variable against its own previous values; Integrated (I), which applies differencing to the data to achieve stationarity; and Moving Average (MA), which models the dependency between an observation and the residual errors from a moving average model applied to lagged observations. ARIMA is notably effective in capturing and predicting linear patterns and trends in time series data, making it a robust tool for forecasting future values based on historical data.

To determine the appropriate values for the ARIMA parameters (p, d, q), Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) graphs were utilized. The effectiveness of the model was evaluated using performance metrics such as Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE).

The steps followed in this analysis were as follows:

* **Extract Data**: Data was extracted from the dataset to be used for training the model.
* **Data Splitting**: The dataset was divided into training and testing sets, with 80% allocated to training and 20% to testing.
* **Parameter Selection**: The values for p, d, and q were determined using PACF and ACF graphs for the High, Low, Open, and Close price series.
* **Model Construction**: An ARIMA model was built using the identified parameters.
* **Model Evaluation**: The model's performance was assessed to ensure accuracy and reliability.
* **Forecasting and Comparison**: Future values were forecasted using the model, and these predictions were compared against actual observed values to evaluate accuracy.

1. **Result Visualization**

The visualization tools employed in this project include Python's **matplotlib** library and R's **ggplot2** library. Additionally, tables were created using Microsoft Word to provide a clear and accessible summary of the data.

The types of visualizations utilized in this analysis encompass line charts, histograms, scatter plots, and candlestick charts. These visual representations are used to illustrate various relationships within the data and to present the forecasts generated by our models effectively.

## **Results and Discussion**

1. **Findings from EDA:**

Exploratory Data Analysis (EDA) is a statistical approach used to analyze and summarize the main characteristics of a dataset, often employing visual methods to uncover patterns, trends, and relationships.

**Data Inspection, Understanding, and Cleaning**

An initial inspection of the dataset can be conducted by examining the first and last 10 rows. This inspection reveals that the "Adj Close" (Adjusted Close) and "Close" columns contain identical information. Consequently, it is advisable to remove the "Adj Close" column for the sake of efficiency.

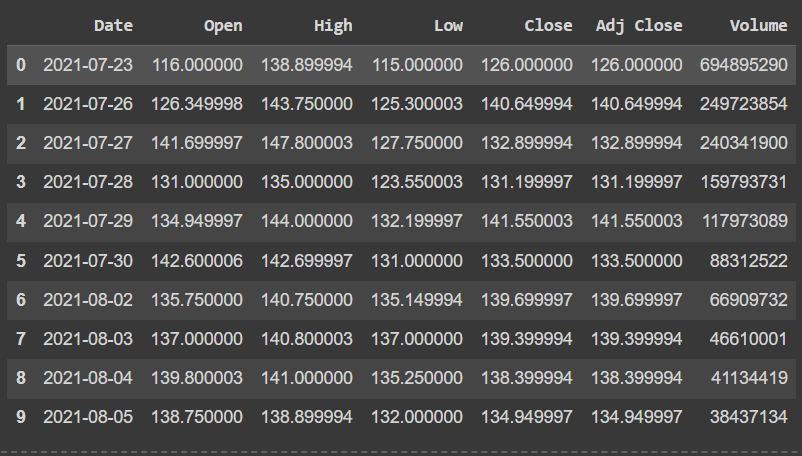


Figure 1. 1 Dataset Head

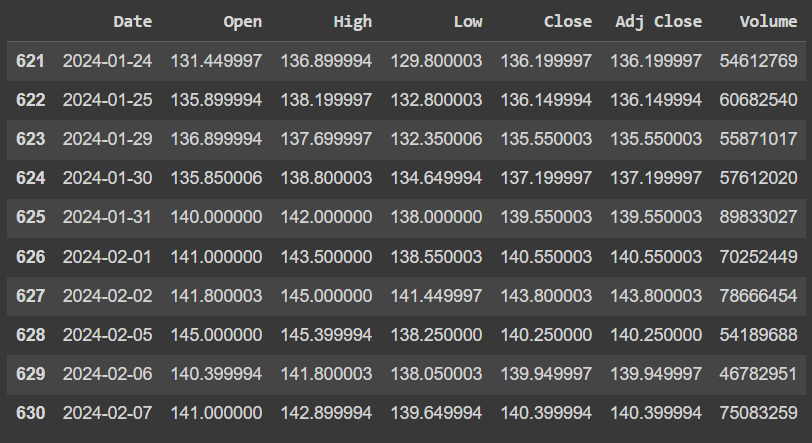


Figure 1. 2 Dataset Tail

Next, we will examine the general information of the dataset, including data types and column names, and verify the presence of any null values. If null values are detected, appropriate measures will be taken to address them. The subsequent figures indicate that there are no null values in the dataset. Therefore, the dataset is deemed suitable for further analysis using descriptive statistics and graphical methods.

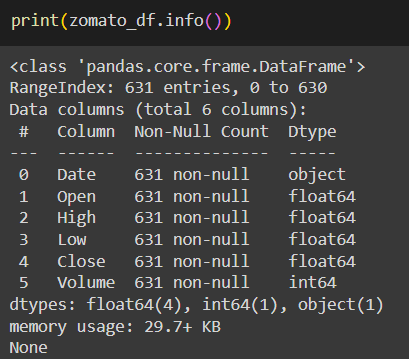


Figure 1. 3 General Schema of Dataset

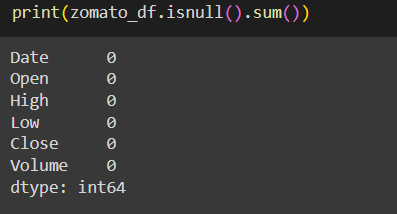
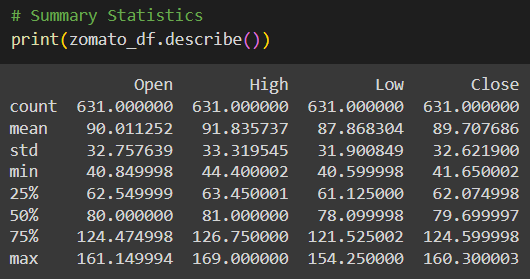


Figure 1. 4 Check for nulls

**Calculation of Descriptive Statistics**

Descriptive statistics encompass essential statistical measures such as mean, standard deviation, count, maximum, and minimum, which offer a summary of the central tendencies, variability, and overall characteristics of a dataset. In this analysis, these measures were computed using the describe method of the pandas DataFrame to provide a detailed understanding of Zomato's stock data. It was observed, however, that the describe method inaccurately calculated the volume metrics, prompting a manual computation to ensure precision.

The descriptive statistics reveal key insights into Zomato's stock performance. The average daily stock price is approximately 90 INR, with a standard deviation of 32 units, indicating a moderate level of price volatility. The price range extends from 40 to 160 INR, highlighting the variability in stock prices over the observed period. The dataset comprises a total of 631 records, providing a substantial sample for analysis. These statistics are crucial for understanding the underlying trends and patterns in Zomato's stock prices and offer a foundation for further detailed analysis.



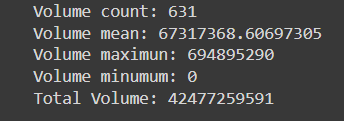


Figure 1. 5 Descriptive Statistics

**Plotting Column Distributions**

Examining column distributions is a crucial technique for understanding the underlying structure of each dataset column, identifying anomalies, and discerning patterns. This process is typically facilitated through the use of histograms or box plots. These graphical representations allow for the visualization of normal distributions and the detection of skewness, whether it be right or left skewness. In this analysis, we have chosen to utilize histograms due to their clarity and ease of interpretation.

The resulting graphs reveal that the most frequent values for the columns "Low," "High," "Open," and "Close" predominantly fall within the range of 120 to 140. This indicates a concentrated clustering of stock prices within this interval. On the other hand, the "Volume" column displays a right-skewed distribution, signifying that a majority of the trading volumes are on the lower side, with fewer occurrences of higher trading volumes. This asymmetry in the volume data suggests that while lower volumes are more common, there are occasional spikes in trading activity.

|  |
| --- |
|  |
|  |
|  |
|  |
|  |

Figure 1. 6 Column Distributions

**Scatter Plot Analysis**

This type of analysis is an effective approach for uncovering the relationships between two or more variables within a dataset. We will investigate the relationships among each variable and its counterparts. By employing pair plots, we can discern that the relationships between the "High," "Low," "Close," and "Open" prices exhibit a linear and positive correlation. This suggests that as one of these price metrics increases, the others tend to increase as well, indicating a consistent trend in stock price movements.

In contrast, when examining the relationship with "Volume," no discernible pattern emerges. This lack of a clear relationship suggests that trading volume does not follow the same linear trends as the price metrics and may be influenced by different factors.

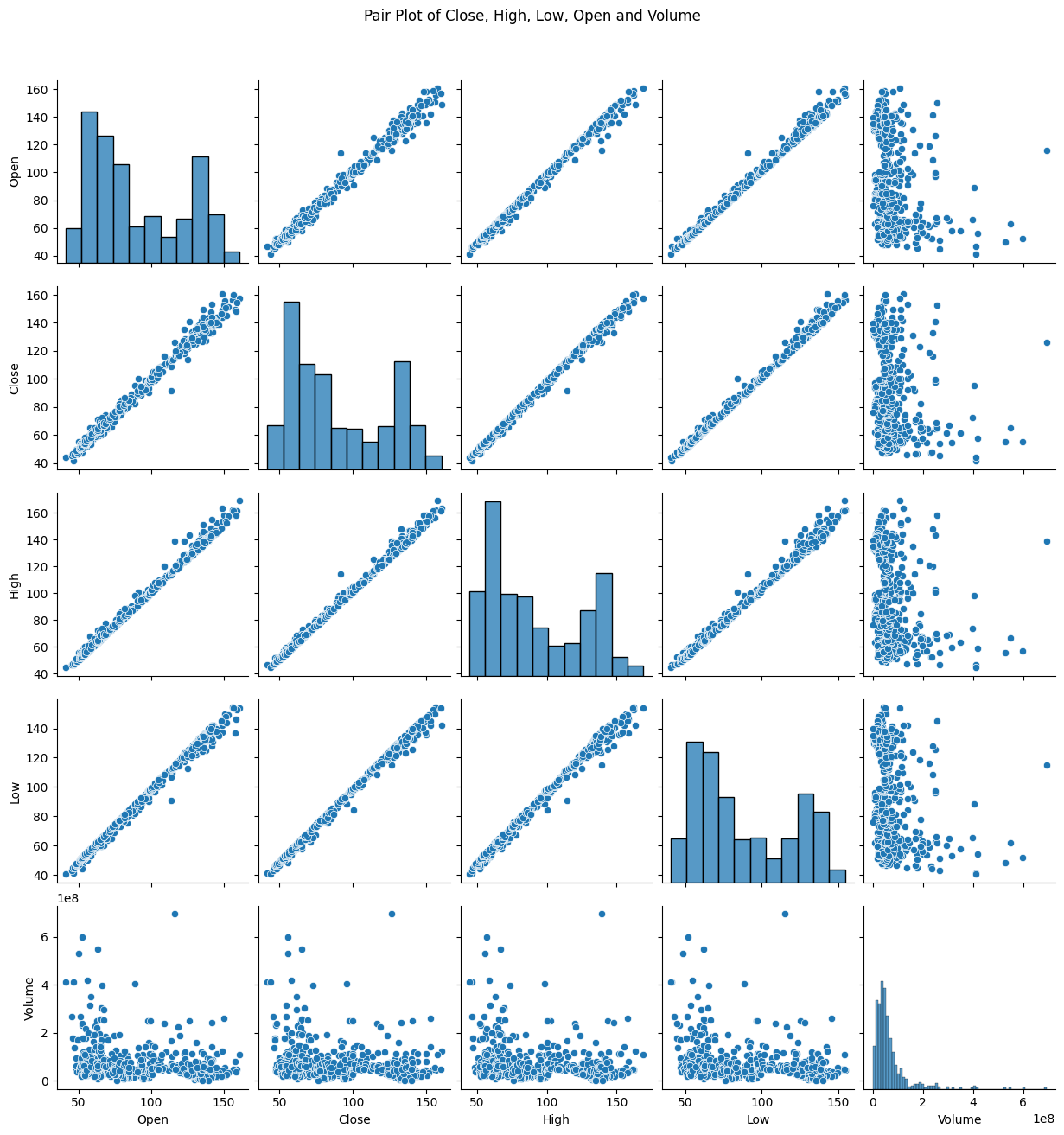


Figure 1. 7 Pairwise Scatter Plots

Furthermore, let us investigate the relationship between price change and volume (where price change = |Close - Open|). However, no meaningful relationship between them is apparent.

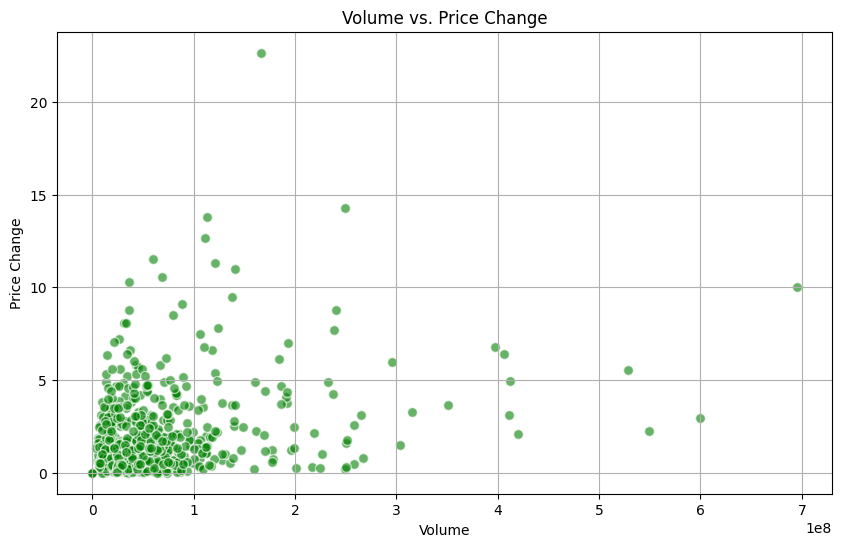


Figure 1. 8 Volume vs Price change

|  |  |
| --- | --- |
|  |  |

Figure 1. 9 Difference of prices relationship

Upon exploring the relationships between price differentials, it can be observed that they exhibit a somewhat weak positive correlation. Additionally, all price variables demonstrate a linear positive relationship with each other; for instance, when the open price rises, the close, high, and low prices also tend to increase. Therefore, these variables are expected to exhibit similar trends when plotted on a line graph.



Figure 1. 10 Close Price over time

Lastly, let us examine the candlestick chart, widely utilized by traders to forecast future market movements. Within the chart, the lines represent the moving averages calculated over 5 and 10 days.

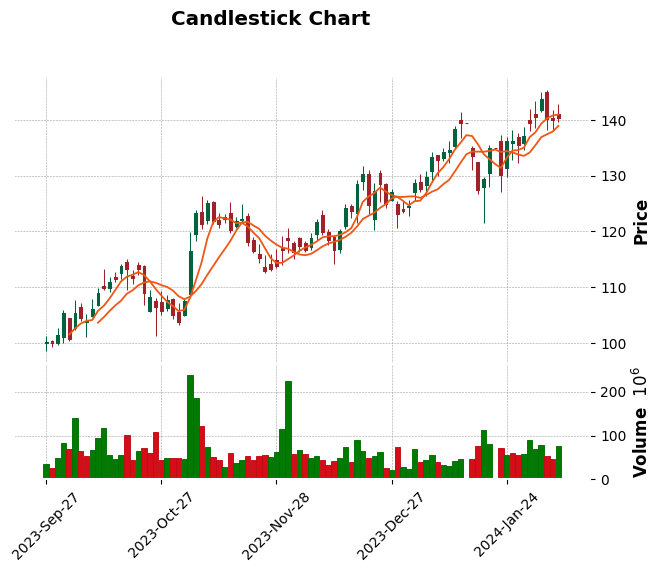


Figure 1. 11 Candlestick chart

**Conclusion**

In concluding the Exploratory Data Analysis (EDA), it is evident that the price variables exhibit a robust positive linear relationship with one another. This indicates that as one price metric (e.g., "Open," "High," "Low," or "Close") increases, the others tend to rise in a correlated manner, reflecting consistent market behavior. Specifically, the trend in the "Close" price over time shows a continuous upward trajectory throughout 2024. This positive trend highlights promising investment opportunities for potential investors who may seek to capitalize on the increasing stock value.

Additionally, the dataset comprises 631 rows, providing a substantial amount of data for analysis. It is important to note that rows with a volume of zero are indicative of non-trading days, which are days when the stock was not actively traded. The range of stock prices within this dataset spans from 40 INR to 160 INR, offering a comprehensive view of Zomato's stock performance over the analyzed period.

1. **Anomaly Detection Results**

Anomaly Detection is a method used to identify outliers in a dataset, which can either be removed or imputed with specific values to enhance model training efficiency. However, in the context of stock market data, where volatility is inherent, anomalies are expected and should not be removed or imputed.

Therefore, our approach involves examining anomalies to gain a deeper understanding of the dataset. We will employ the R programming language along with the 'mvoutlier' and 'ggplot2' libraries to detect and visualize outliers using methods such as Sign, Sign 1, Sign 2, and PCOut.

The Sign Method for anomaly detection utilizes statistical techniques to detect outliers in multivariate datasets by assessing deviations from a reference point, typically the median or mean. This method evaluates the direction and magnitude of deviations across multiple dimensions to identify data points that significantly deviate from the norm.

Specifically, the Sign 1 Method focuses on the direction of deviations between individual data points and the median or mean of the dataset. It is known for its simplicity and robustness in detecting anomalies based on the consistency of these deviations.

In our analysis, the Sign 1 method was implemented with q crit values (critical thresholds) of 0.975, 0.99, and 0.5. Anomalies are indicated by red points on the graph. It is observed that as the q crit value decreases, the number of detected anomalies increases

|  |  |  |
| --- | --- | --- |
|  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Result No.** | **Qcrit value** | **Number of outliers** | **Percentage of outliers** |
| Result 1 | 0.975 | 100 | 15.84 % |
| Result 2 | 0.99 | 86 | 13.62 % |
| Result 3 | 0.5 | 292 | 46.27 % |

Figure 2. 1 Sign1 method

The Sign 2 Method builds upon the Sign 1 approach by incorporating the magnitude of deviations to enhance the identification of anomalies. This method is particularly sensitive to the size of deviations and can effectively detect subtle anomalies. It utilizes principal components to calculate distances in the dataset.

In the visual representation, anomalies are denoted by red points. As with the Sign 1 Method, decreasing the q crit value results in an increased detection of anomalies.

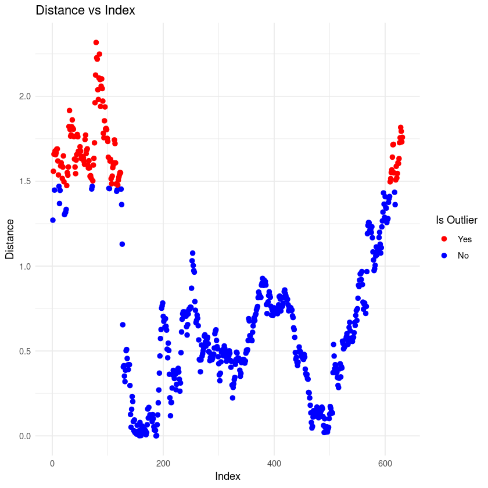
|  |  |  |
| --- | --- | --- |
|  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Result No.** | **Qcrit value** | **Number of outliers** | **Percentage of outliers** |
| Result 1 | 0.975 | 2 | 0.31 % |
| Result 2 | 0.95 | 14 | 2.21 % |
| Result 3 | 0.75 | 175 | 27.73 % |

Figure 2. 2 Sign2 method

The PCOut method is an advanced statistical technique employed to identify anomalies or outliers within multivariate datasets. PCOut stands for Principal Component Outlier and utilizes principal component analysis (PCA) to pinpoint unusual data points that deviate significantly from the underlying data structure.

In the graphical representation, anomalies are highlighted by red points, with a default q crit value of 0.975.



|  |  |  |  |
| --- | --- | --- | --- |
| **Result No.** | **Qcrit value** | **Number of outliers** | **Percentage of outliers** |
| Result 1 | 0.975 | 129 | 20.44 % |

Figure 2. 3 PCOut method

**Conclusion**

In conclusion, the discrepancies in anomalies detected by various methods stem from differences in the distance computation measures and the methodologies applied. This variability is evident in the varying number of outliers identified for the same q crit value across different techniques. Thus, this comparison of anomaly detection methods provides a clearer understanding of the dataset.

1. **Forecasting Results**

Forecasting is a method used to predict the future values of a particular variable within a time series. In this project, we will train an ARIMA model for each variable in the dataset, specifically Open, Close, High, and Low prices. The dataset is divided into train set and test set with 80-20 ratio. Therefore, train set contains 504 rows and test set contains 127 rows.

The initial step in constructing an ARIMA model involves detrending the time series if it is non-stationary. From the Exploratory Data Analysis (EDA) conducted earlier, we determined that the price variables are non-stationary.

Therefore, we will apply differencing to the time series up to the third order. The first and second-order differencing did not yield optimal results for forecasting, necessitating the use of third-order differencing.

|  |  |
| --- | --- |
|  |  |
|  |  |

Figure 3. 1 Stationary Series

We have selected the training parameters for the ARIMA model using traditional methods, specifically by plotting the Partial Autocorrelation Function (PACF) to determine p and the Autocorrelation Function (ACF) to determine q. From the graphs, it can be deduced that for the Open, Close, High, and Low price models, the parameters will be p=2, d=3, and q=3.

|  |  |
| --- | --- |
|  |  |
|  |  |

Figure 3. 2 Autocorrelation

|  |  |
| --- | --- |
|  |  |
|  |  |

Figure 3. 3 Partial Autocorrelation

We have trained our model using the ARIMA function from the statsmodels library and fit the model with the fit() function, which returns an ARIMAResultsWrapper object.

The model's performance will be evaluated using the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). We will obtain a set of predictions using walk-forward validation. Where, the model is trained recursively and forecast one step at a time throughout the test data. The following figures show the visualization of predictions and calculated results of evaluation metrics.

|  |  |
| --- | --- |
|  |  |
|  |  |

Figure 3. 4 Predicted and Actual values

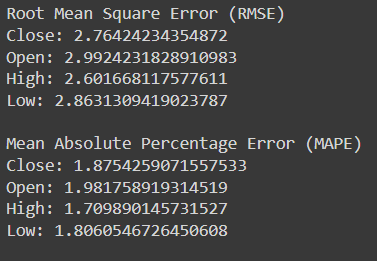


Figure 3. 5 RMSE and MAPE

Upon reviewing the evaluation results and visualizations, it appears that the model demonstrates a satisfactory fit.

Finally, we will perform forecasting using the trained model on the test dataset to facilitate comparison and analysis of the results. Forecasting was conducted using the `get\_forecast()` method, which returns a `PredictionResults` object containing predicted values and confidence intervals, simplifying the process of plotting forecasted data on a graph. The plotted graph includes training data, testing data, a line indicating the predicted prices, and a grey area representing the confidence intervals.

Analyzing the graphs we can observe that the trend in forecasted price and actual price matches in all the prices. Forecast made of 127 steps is shown in the graph which is equivalent to test set.

|  |  |
| --- | --- |
|  |  |
|  |  |

Figure 3. 6 Forecasted Values

**Interpretation of Results**

Based on the graphical analysis, it can be concluded that the stock price is projected to increase over time. This upward trend suggests that the stock holds potential for long-term growth. Therefore, it would be prudent for investors to consider holding their investments in this stock for an extended period. Such a strategy is likely to yield favourable returns as the stock value appreciates in the future.

## **Conclusion**

In this project, we conducted Exploratory Data Analysis (EDA) and Anomaly Detection to gain a comprehensive understanding of the dataset. Our analysis revealed a robust positive relationship among the price variables, indicating that when the opening price increases, the other three variables—high, low, and close prices—also tend to increase. In the anomaly detection phase, we quantified the percentage and number of outliers present in the dataset. However, given the inherent volatility of the stock market, we decided not to remove or impute these anomalies. Finally, we performed forecasting on four key variables: 'High,' 'Low,' 'Close,' and 'Open.' The forecasting results demonstrated that these values tend to increase over time, and the predicted prices closely align with the actual data. Additionally, the forecasted values exhibit trends that are consistent with the actual data, suggesting a positive outlook for the stock. This finding implies that investors may benefit from investing in the stock as it shows an upward trajectory. Future work could focus on enhancing the model's accuracy and developing a model capable of accurately predicting trading volumes.

**Limitations**

One significant limitation encountered during the modelling process was the poor fit of the ARIMA model when applied to the stock volume data. Despite various attempts to optimize the model parameters, the ARIMA model failed to provide accurate or reliable predictions for the stock's trading volume. This issue highlights the need for alternative approaches that can better capture the complexity and variability inherent in volume data. To improve the accuracy of volume predictions, it is essential to explore other advanced modelling techniques or hybrid approaches that may offer a more precise and robust solution for forecasting trading volumes.

**References**

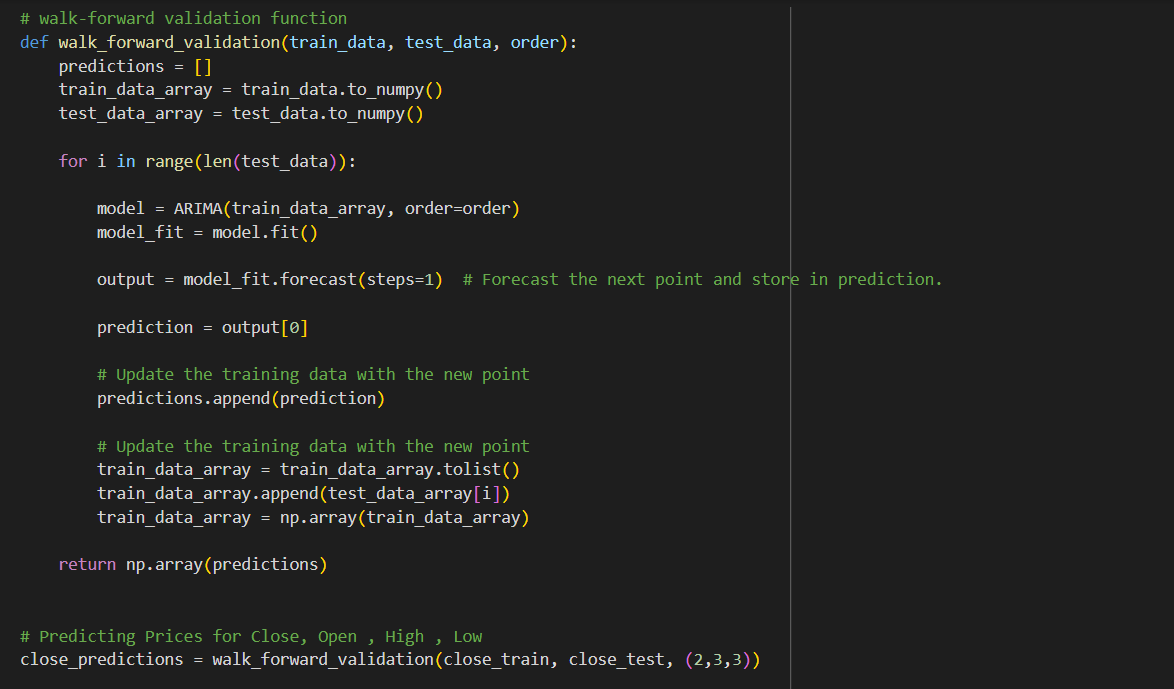
* Data Source - <https://www.kaggle.com/datasets/aradhakkandhari/zomato-stock-price-prediction>
* Plotting Forecast - <https://www.analyticsvidhya.com/blog/2021/07/stock-market-forecasting-using-time-series-analysis-with-arima-model/>
* Statmodels Library documentation - <https://www.statsmodels.org/stable/generated/statsmodels.tsa.arima.model.ARIMAResults.get_forecast.html#statsmodels.tsa.arima.model.ARIMAResults.get_forecast>
* Determining model parameters - <https://analyticsindiamag.com/quick-way-to-find-p-d-and-q-values-for-arima/#:~:text=Draw%20a%20partial%20autocorrelation%20graph,to%20the%20ACF%20is%20q>.

## **Appendices**

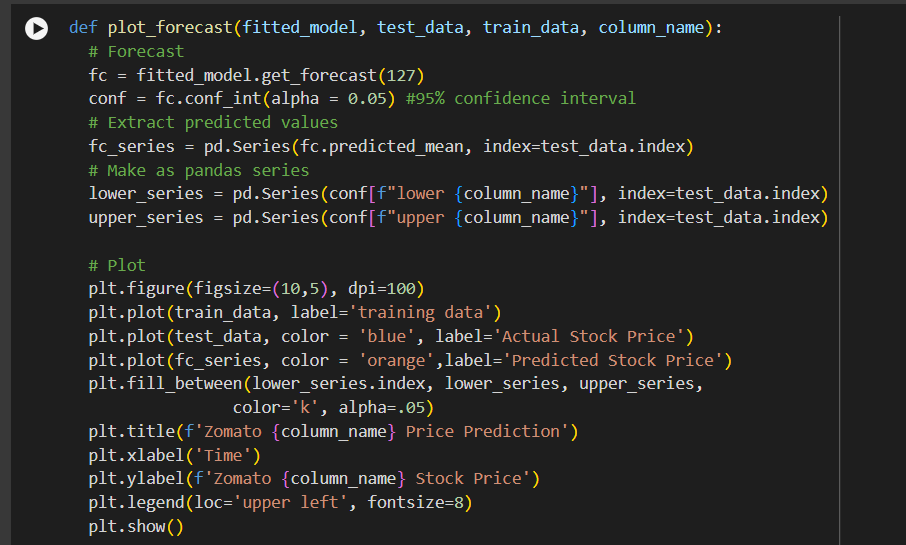
1. **Appendix A: Code Snippets**

**A1: Forecasting code snippets**

Walk Forward validation which is used to fetch the predictions per step.

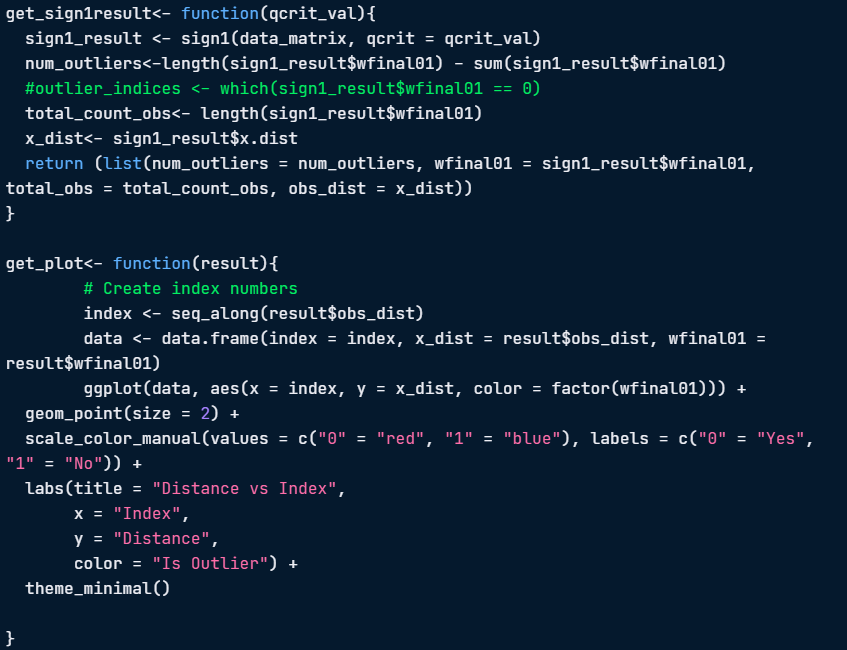


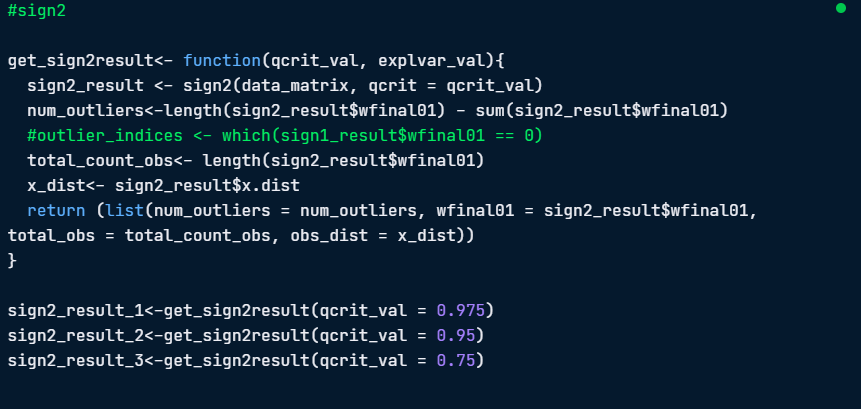
Forecasting 127 values that is equivalent to the test set records.

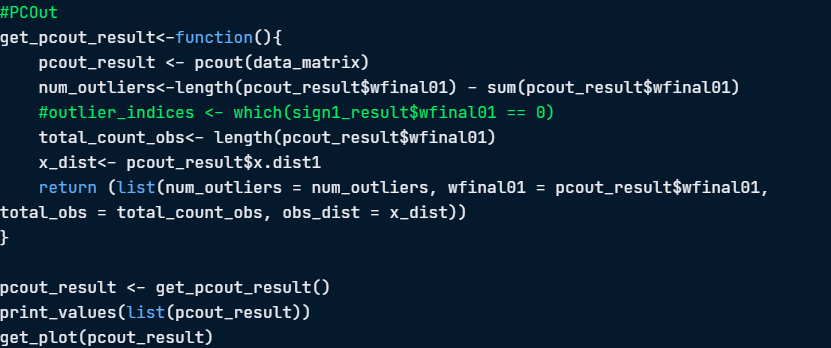


**A2: Anomaly Detection code snippets**

Detecting anomalies using sign1 sign2 and pcout method in mvoutlier library in R.







1. **Appendix B: Supplementary Material**

**B1: Detailed Description of Data Source**

The dataset used in this project was sourced from Kaggle, consisting of historical stock price data for Zomato. It includes variables such as Open, Close, High, Low, and Volume for the period from July 2021 to February 2024.

**B2: Overview of Anomaly Detection Techniques**

* **Sign 1 Method**: Utilizes the sign of deviations from a median reference point to identify anomalies. It is robust for multivariate data and highlights significant outliers.
* **Sign 2 Method**: Extends the Sign 1 method by incorporating the magnitude of deviations. It is particularly sensitive to subtle anomalies in the dataset.
* **PCOut Method**: Applies principal component analysis to detect outliers that deviate significantly from the principal components of the data.

**C3: Glossary of Terms**

* **ARIMA**: AutoRegressive Integrated Moving Average, a statistical model used for time series forecasting.
* **PACF**: Partial Autocorrelation Function, used to identify the order of the AR component in ARIMA.
* **ACF**: Autocorrelation Function, used to determine the order of the MA component in ARIMA.