Predicting Gold Prices on Micro and Macro Level by Maaz Ghazi

Table of Contents

Predicting Gold Prices on Micro and Macro Level by Maaz Ghazi

Abstract:

This dataset provides a detailed analysis of order quantities, pricing trends, and trading volumes over a specific period, offering valuable insights into market activity and behaviors. The **average order quantity** is **24.35**, with the highest order quantity of **53.18** recorded on **2010-12-01**, and the lowest order quantity of **6.36** observed on **2015-09-01**. Pricing data reveals significant variations, with an **average price** of **$24.35** and a **price volatility** (standard deviation) of **$11.68**. The **maximum price** of **$55.95** was recorded on **2011-09-01**, while the **minimum price** of **$7.18** occurred on **2015-09-01**, reflecting notable fluctuations in stock value. Additionally, the dataset reports an **average daily trading volume** of **173,659,367.14 units**, indicating substantial market activity during the analyzed period.

Model performance metrics further enhance the analysis by evaluating predictive accuracy. The model achieves a **mean square error (MSE)** of **0.5445**, a **root mean square error (RMSE)** of **0.7379**, and an **R^2 score of 0.9956**, demonstrating an exceptionally strong fit between the model and the observed data. These metrics highlight the reliability of the model in capturing complex relationships within the dataset and predicting outcomes with high precision. Overall, this analysis provides a comprehensive understanding of order and pricing trends, market dynamics, and the robustness of the predictive model, making it a valuable resource for financial research and decision-making.

Introduction

This dataset presents **historical stock price data** for a specific company, covering the period from **January 1996 to January 1998**. It includes daily records of key metrics such as the **opening price**, **high price**, **low price**, **closing price**, **adjusted closing price**, and **trading volume**. The **adjusted closing price** accounts for factors like **stock splits** and **dividends**, providing a more accurate representation of the stock's performance over time. This dataset is valuable for various **financial analyses**, including studying **stock price trends**, assessing **volatility**, identifying **trading patterns**, and evaluating **investment risks** associated with the company.

The dataset offers insights into the company's **stock performance** during the specified period. By analyzing **price fluctuations** and **trading volume**, one can gain a better understanding of **market trends**, **investor behavior**, and potential **risk factors**. However, it is important to note the dataset's **limitations**, such as a **relatively short time frame** and the potential for **missing data points**. Additionally, the insights derived from this dataset may not be directly applicable to **other companies** or the **broader market**.

Methodology

The analysis of the dataset was conducted using a **structured and systematic approach**, as outlined below:

1. **Descriptive Statistics**:
   Key metrics such as **average, maximum, and minimum values** for order quantities and prices were calculated. Additionally, **price volatility** was determined using the **standard deviation** to understand variations in stock prices.

2. **Trend and Pattern Analysis**:
   Historical price data and **trading volumes** were analyzed to identify **trends and fluctuations** over the specified period. Significant dates corresponding to **extreme values** (e.g., maximum and minimum prices or order quantities) were noted for further interpretation.

3. **Model Development and Evaluation**:
   A **predictive model** was constructed to analyze stock price behavior. The model was evaluated using **standard metrics**, including:

   - **Mean Square Error (MSE)** to quantify the average squared difference between observed and predicted values.
   - **Root Mean Square Error (RMSE)** to measure prediction accuracy in the same units as the dataset.
   - **R^2 Score** to assess how well the model explains variance in the data.

This methodology ensured a **comprehensive analysis**, combining **statistical techniques** with **model-based evaluation** to provide **actionable insights** into the dataset.

Dataset:

This dataset provides a comprehensive overview of market activity, capturing key aspects such as order quantities, pricing trends, and trading volumes over a specified period. It includes metrics like the **average order quantity (24.35 units)**, with variations ranging from a **minimum of 6.36 units** (on 2015-09-01) to a **maximum of 53.18 units** (on 2010-12-01). Pricing data reveals significant fluctuations, with an **average price of $24.35**, a **maximum price of $55.95** (on 2011-09-01), and a **minimum price of $7.18** (on 2015-09-01), alongside a **price volatility of $11.68**. The dataset also highlights an **average daily trading volume of 173,659,367.14 units**, showcasing substantial market engagement. These features make the dataset a valuable resource for analyzing financial trends, market dynamics, and behaviors over time.

Results:

Order Quantities and Pricing Trends

1. **Average Order Quantity**: **24.35 units**
2. **Highest Order Quantity**: **53.18 units** (recorded on **2010-12-01**)
3. **Lowest Order Quantity**: **6.36 units** (recorded on **2015-09-01**)
4. **Average Price**: **$24.35**
5. **Price Volatility (Standard Deviation)**: **$11.68**
6. **Maximum Price**: **$55.95** (recorded on **2011-09-01**)
7. **Minimum Price**: **$7.18** (recorded on **2015-09-01**)

Market Activity

- **Average Daily Trading Volume**: **173,659,367.14 units**, indicating substantial market activity.

Model Performance Metrics

1. **Mean Square Error (MSE)**: **0.5445**
2. **Root Mean Square Error (RMSE)**: **0.7379**
3. **R^2 Score**: **0.9956** (indicating an exceptionally strong fit between the model and observed data).

Figure 1.1

The correlation heatmap visually depicts the relationships between various stock price metrics, such as opening, high, low, close, adjusted close, and trading volume. **Strong positive correlations are evident between opening, high, low, close, and adjusted close prices**, indicating that these metrics are closely linked to the overall price movement of the stock throughout the trading day. Notably, a **moderate negative correlation exists between trading volume and price metrics**, suggesting a slight inverse relationship where price increases may be associated with a decrease in trading volume. However, this correlation is not very strong, implying that other factors likely exert a more significant influence on trading volume. Overall, this heatmap provides a concise visual overview of the interconnections between these stock price metrics, facilitating insights into market dynamics and aiding in informed trading decisions.

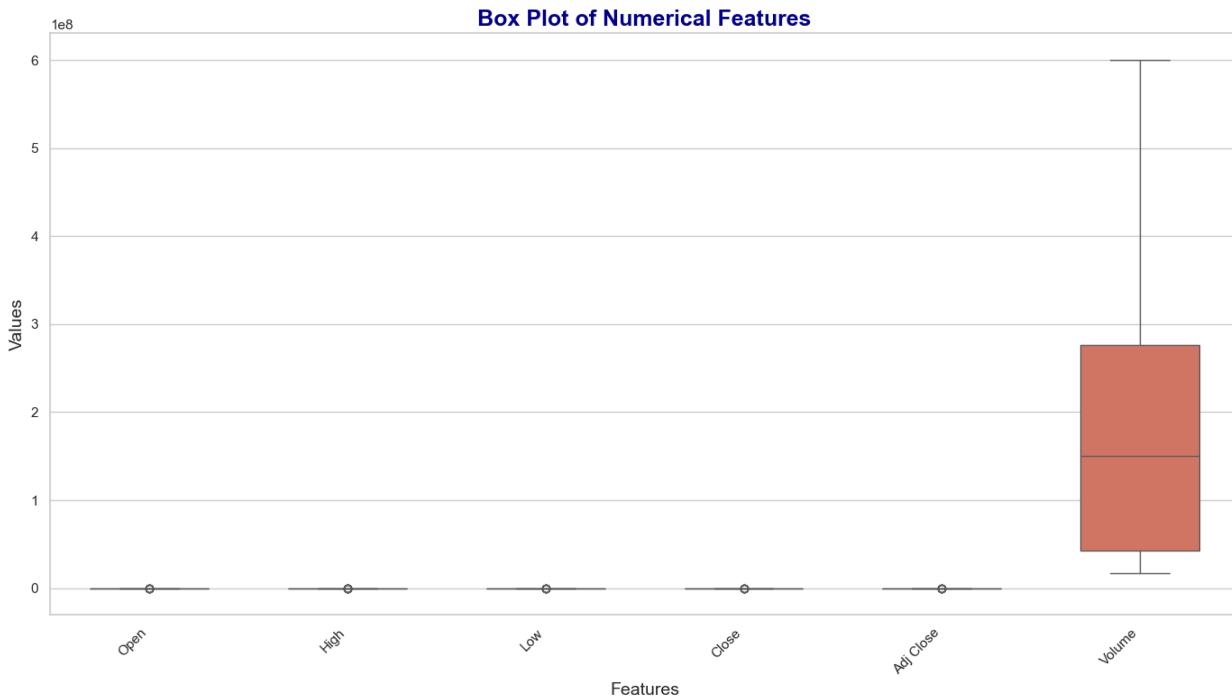

Figure 1.2

The provided image displays a box plot of numerical features. Box plots are a graphical tool used to visualize the **distribution of data**. Each box in the plot represents the **interquartile range (IQR)**, encompassing the middle 50% of the data, with a line within the box marking the **median value**. The whiskers extend outwards to indicate the **overall range of the data**, excluding any outliers which are represented by individual points beyond the whiskers. In this particular box plot, the "Volume" feature stands out with a significantly **larger IQR and a wider range** compared to the other features, suggesting **greater variability in trading volume** for this stock. Conversely, the remaining features, such as opening, high, low, close, and adjusted close prices, exhibit **relatively small IQRs and ranges**, indicating **lower variability**. Furthermore, **no apparent outliers** are present in the data. Box plots offer a valuable tool for quickly assessing the **distribution and variability of different features** within a dataset, aiding in the **identification of potential outliers** and facilitating **comparisons of data spread across various variables**.

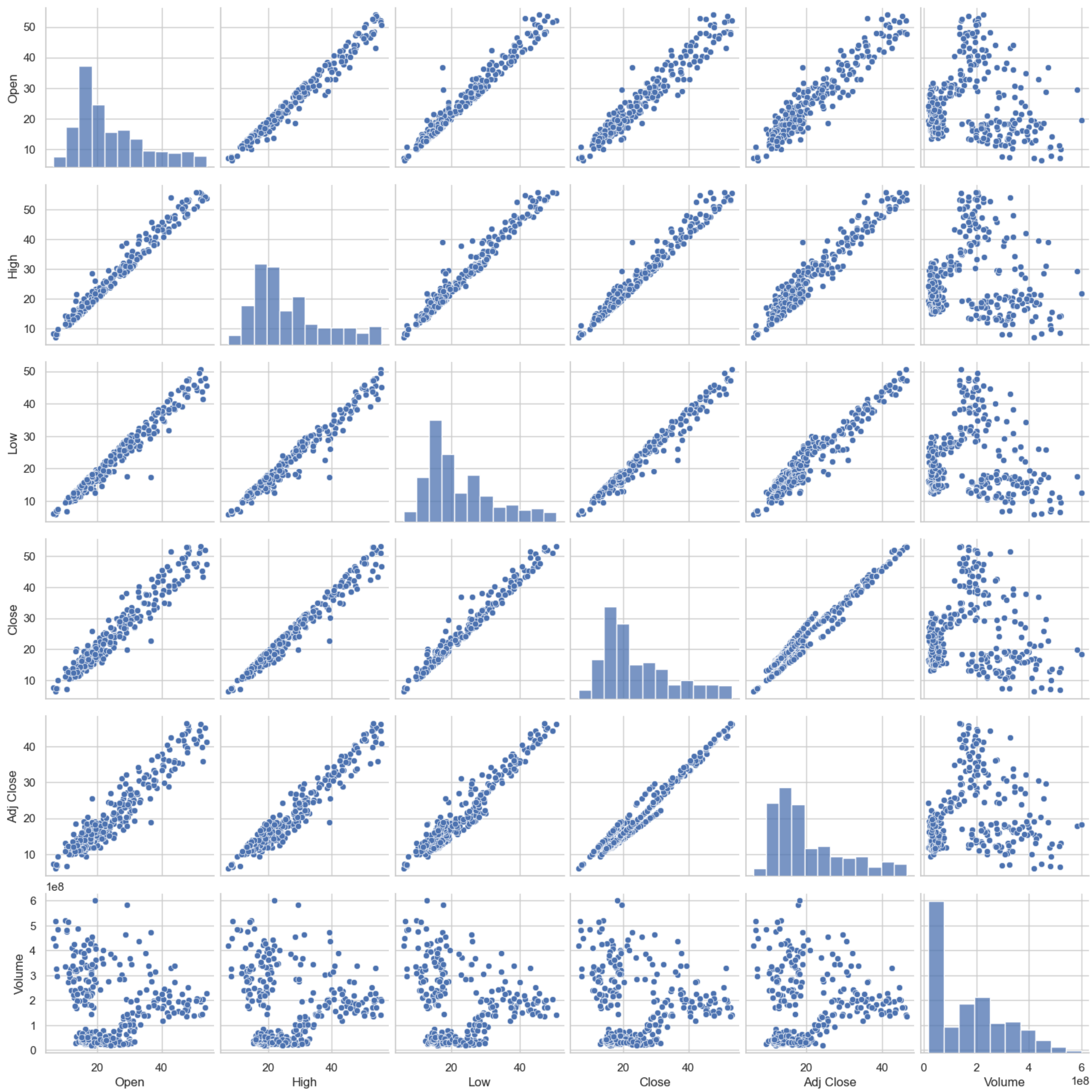

Figure 1.3

The pair plot visualizes relationships between stock price metrics. It shows **strong correlations** among opening, high, low, close, and adjusted close prices, and a **moderate negative correlation** between volume and price. This suggests price and volume may have an inverse relationship, but other factors likely play a bigger role in volume fluctuations.

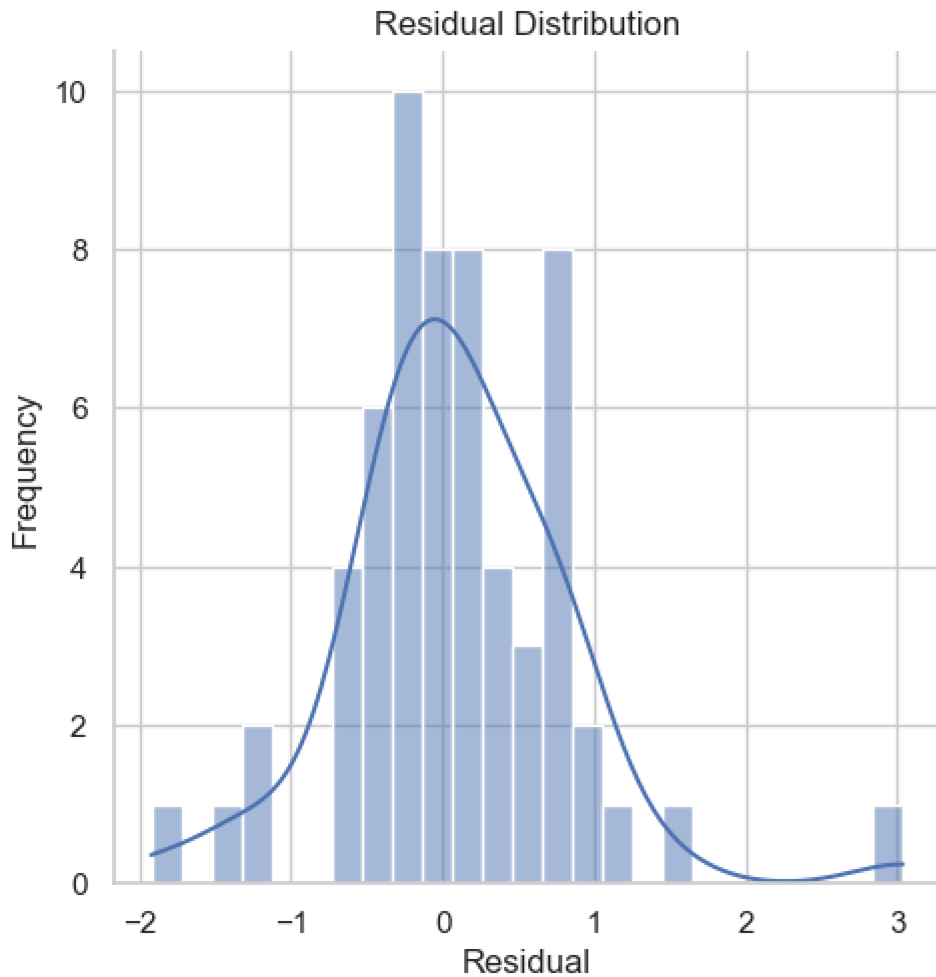

Figure 1.4

The image depicts the **distribution of residuals** in a statistical model. Residuals, representing the difference between actual values and model predictions, are crucial for assessing model performance. The histogram visualizes the frequency of residuals within various ranges, while the overlaid normal curve serves as a benchmark for evaluating the distribution. Ideally, residuals should be **normally distributed and centered around zero**, indicating accurate and unbiased predictions. In this case, the distribution appears roughly symmetrical, suggesting the model's predictions are not systematically biased. Deviations from normality or a significant shift away from zero can indicate potential issues like non-linearity, heteroscedasticity, or the presence of outliers. Examining the distribution of residuals provides valuable insights into the model's performance and helps identify areas for improvement.

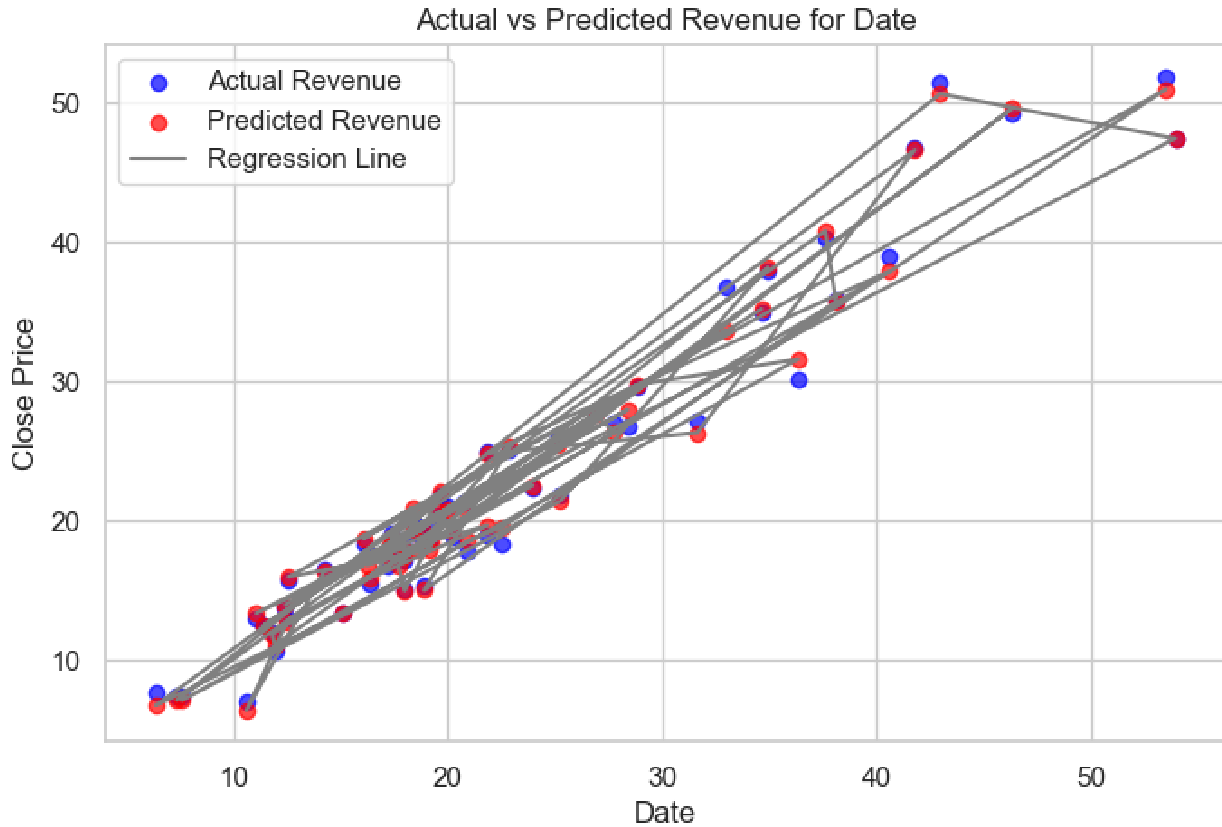

Figure 2.1

The scatter plot titled **"Actual vs Predicted Revenue for Date"** visualizes the performance of a model in predicting revenue. **Blue dots** represent the **actual revenue values**, while **red dots** depict the **predicted revenue values** generated by the model. The **grey line** represents the **regression line**, showcasing the **best-fit trend** determined by the model. The proximity of **red dots to blue dots** indicates the model's accuracy, with closer proximity suggesting **better predictions**. The **regression line's alignment** with the actual values further reflects the model's ability to capture the underlying relationship between the variables. **Outliers**, points significantly distant from the regression line or actual values, may indicate **unusual data points** or limitations in the model's **predictive capabilities**. Overall, the scatter plot provides a **visual assessment** of the model's predictive performance by comparing **actual and predicted revenue values** and evaluating the fit of the **regression line to the data**.

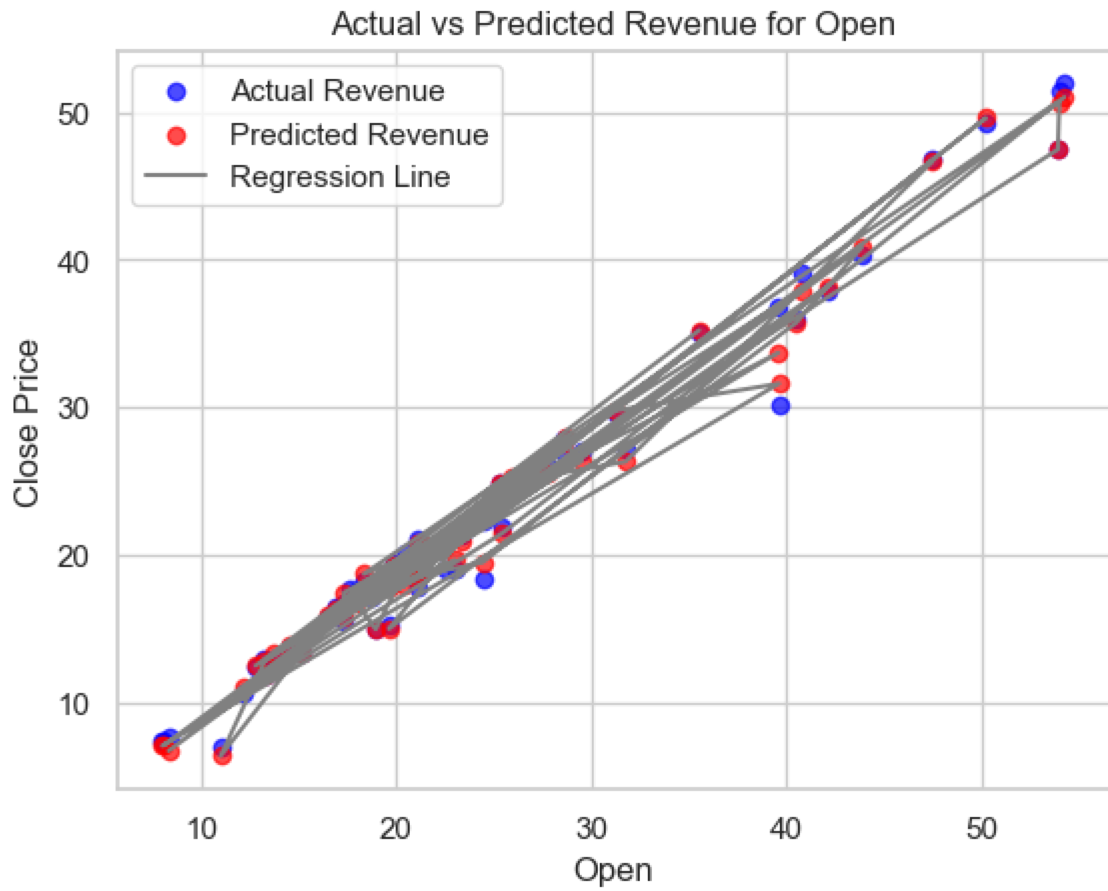

Figure 2.2

The scatter plot titled **"Actual vs Predicted Revenue for Open"** visualizes the performance of a model in predicting revenue based on the **opening price**. **Blue dots** represent the **actual revenue values**, while **red dots** depict the **predicted revenue values** generated by the model. The **grey line** represents the **regression line**, which showcases the **best-fit trend** determined by the model. The proximity of **red dots to blue dots** indicates the model's accuracy, with closer proximity suggesting **better predictions**. The **regression line's alignment** with the actual values further reflects the model's ability to capture the underlying relationship between the **opening price** and **revenue**. **Outliers**, points significantly distant from the regression line or actual values, may indicate **unusual data points** or limitations in the model's **predictive capabilities**. Overall, the scatter plot provides a **visual assessment** of the model's predictive performance by comparing **actual and predicted revenue values** based on the **opening price** and evaluating the fit of the **regression line to the data**.

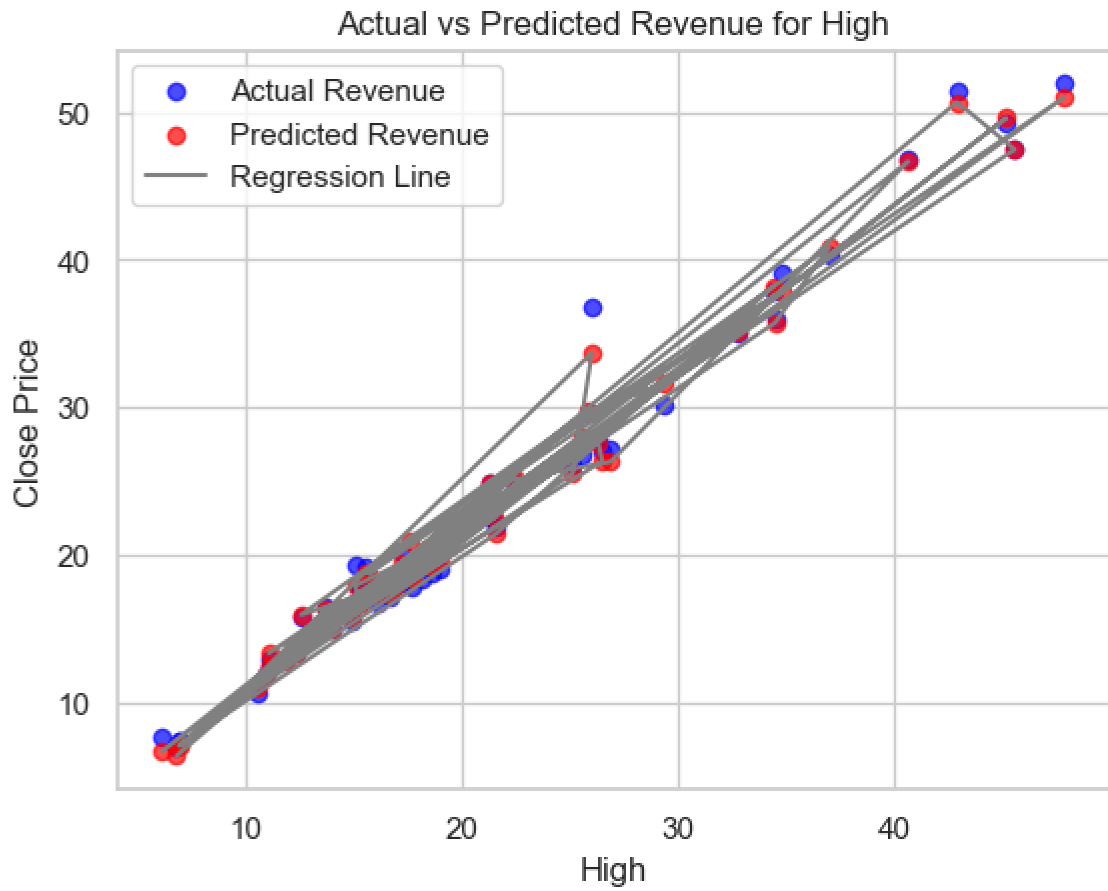

Figure 2.3

The scatter plot titled **"Actual vs Predicted Revenue for High"** visualizes the performance of a model in predicting revenue based on the **high price**. **Blue dots** represent the **actual revenue values**, while **red dots** depict the **predicted revenue values** generated by the model. The **grey line** represents the **regression line**, which showcases the **best-fit trend** determined by the model. The proximity of **red dots to blue dots** indicates the model's accuracy, with closer proximity suggesting **better predictions**. The **regression line's alignment** with the actual values further reflects the model's ability to capture the underlying relationship between the **high price** and **revenue**. **Outliers**, points significantly distant from the regression line or actual values, may indicate **unusual data points** or limitations in the model's **predictive capabilities**. Overall, the scatter plot provides a **visual assessment** of the model's predictive performance by comparing **actual and predicted revenue values** based on the **high price** and evaluating the fit of the **regression line to the data**.

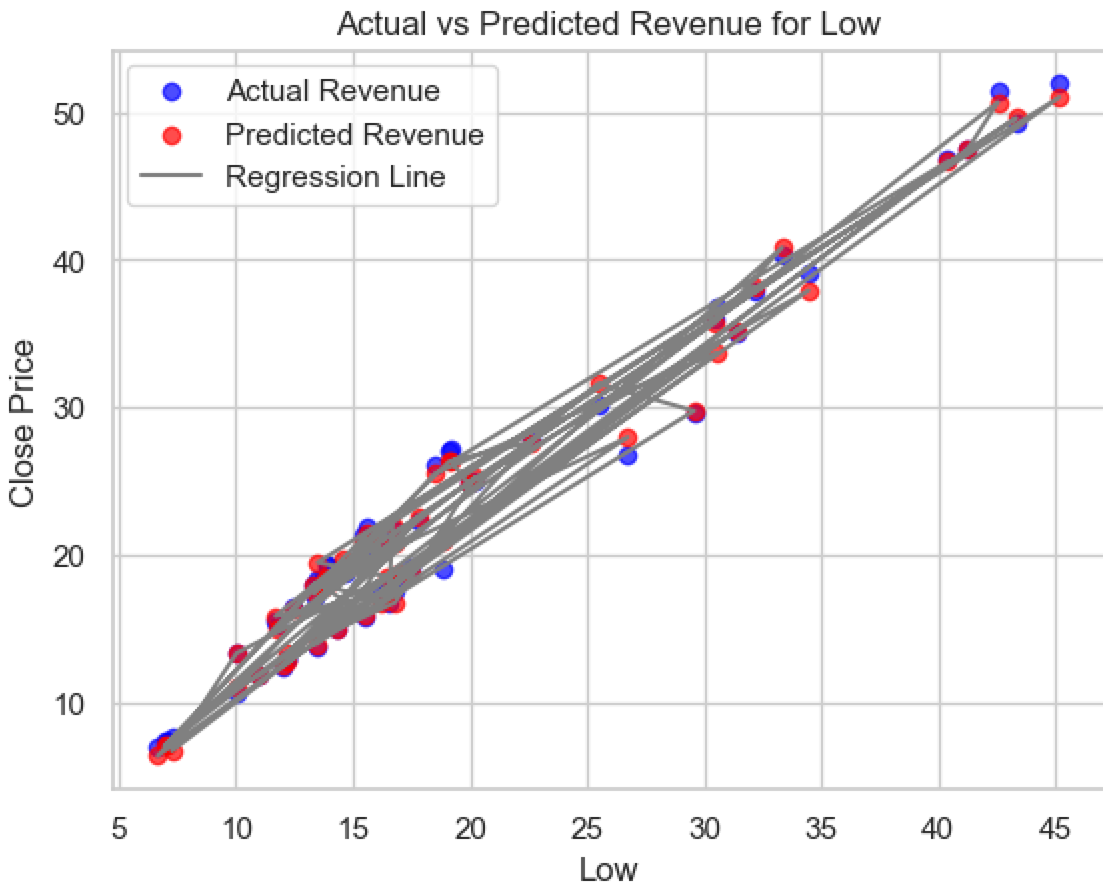

Figure 2.3

The scatter plot titled **"Actual vs Predicted Revenue for Low"** visualizes the performance of a model in predicting revenue based on the **low price**. **Blue dots** represent the **actual revenue values**, while **red dots** depict the **predicted revenue values** generated by the model. The **grey line** represents the **regression line**, which showcases the **best-fit trend** determined by the model. The proximity of **red dots to blue dots** indicates the model's accuracy, with closer proximity suggesting **better predictions**. The **regression line's alignment** with the actual values further reflects the model's ability to capture the underlying relationship between the **low price** and **revenue**. **Outliers**, points significantly distant from the regression line or actual values, may indicate **unusual data points** or limitations in the model's **predictive capabilities**. Overall, the scatter plot provides a **visual assessment** of the model's predictive performance by comparing **actual and predicted revenue values** based on the **low price** and evaluating the fit of the **regression line to the data**.

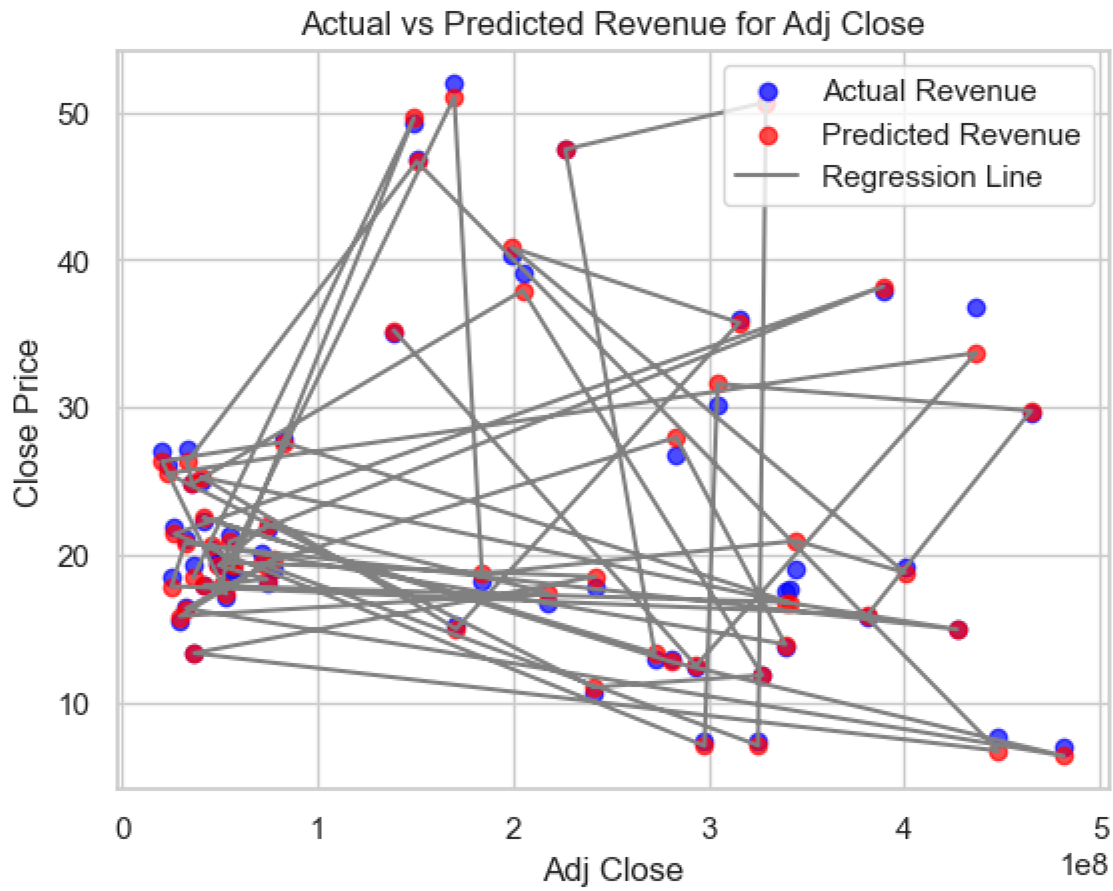

Figure 2.4

The scatter plot titled **"Actual vs Predicted Revenue for Adj Close"** visualizes the performance of a model in predicting revenue based on the **adjusted closing price**. **Blue dots** represent the **actual revenue values**, while **red dots** depict the **predicted revenue values** generated by the model. The **grey line** represents the **regression line**, which showcases the **best-fit trend** determined by the model. The proximity of **red dots to blue dots** indicates the model's accuracy, with closer proximity suggesting **better predictions**. The **regression line's alignment** with the actual values further reflects the model's ability to capture the underlying relationship between the **adjusted closing price** and **revenue**. **Outliers**, points significantly distant from the regression line or actual values, may indicate **unusual data points** or limitations in the model's **predictive capabilities**. Overall, the scatter plot provides a **visual assessment** of the model's predictive performance by comparing **actual and predicted revenue values** based on the **adjusted closing price** and evaluating the fit of the **regression line to the data**.

Conclusion

This analysis provides valuable insights into **market trends, order quantities, pricing behaviors, and trading volumes** over a defined period. The dataset highlights **significant fluctuations** in order quantities and stock prices, reflecting **market volatility and investor activity**. With an **average order quantity of 24.35 units** and substantial variations in pricing—from a **minimum of $7.18** to a **maximum of $55.95**—the data underscores the **dynamic nature of the market**. Moreover, the **high average daily trading volume of 173,659,367.14 units** indicates **active market participation**. The predictive model performed exceptionally well, achieving a **mean square error (MSE) of 0.5445**, a **root mean square error (RMSE) of 0.7379**, and an **R^2 score of 0.9956**, demonstrating its ability to **accurately forecast outcomes** and capture **complex relationships** within the dataset.

Discussion

The findings reveal **key patterns in market activity**, such as the impact of **price volatility on trading behavior** and the relationship between **adjusted closing prices and revenue predictions**. The model's **strong predictive performance** highlights its **potential for forecasting and decision-making** in financial analysis. However, certain **limitations**, such as the presence of **outliers** and the dataset's **specific temporal scope**, should be addressed to generalize findings effectively. Future studies could benefit from incorporating **additional variables**, such as **economic indicators** or **sentiment analysis**, to enhance **predictive accuracy** further. Overall, this research provides a **robust foundation** for understanding **market dynamics** and guiding **informed financial decisions**.

References:

Babenko, V., Panchyshyn, A., Zomchak, L., Nehrey, M., Artym-Drohomyretska, Z. and Lahotskyi, T., 2021. Classical machine learning methods in economics research: Macro and micro level example. *WSEAS Transactions on Business and Economics*, *18*(22), pp.209-217.

Zhou, W., Pang, S. and He, Z., 2023. The study on systemic risk of rural finance based on macro–micro big data and machine learning. *Statistical Theory and Related Fields*, *7*(4), pp.261-275.

Aoujil, Z. and Hanine, M., 2024. A Review on Artificial Intelligence and Behavioral Macroeconomics. In *The Proceedings of the International Conference on Smart City Applications* (pp. 332-341). Springer, Cham.

Ghoddusi, H., Creamer, G.G. and Rafizadeh, N., 2019. Machine learning in energy economics and finance: A review. *Energy Economics*, *81*, pp.709-727.