Times_Series_Model_Report

INTRODUCTION

This project focuses on extracting data from the Vahan Parivahan dashboard and utilizing it to forecast vehicle sales for the next quarter. By analyzing historical registration data, the goal is to provide accurate predictions for the upcoming months, helping stakeholders in the automotive industry make informed decisions.

OBJECTIVE

To forecast the monthly sales of four-wheelers for the next quarter using deep learning models, leveraging historical sales data obtained through web scraping from the Vahan database.

WEB SCRAPING

For web scraping, Selenium was utilized due to its robustness in handling dynamic web content and its ability to interact with various web elements effectively. This made it an ideal choice for extracting comprehensive data from the Vahan Parivahan dashboard. The extracted data was then saved as a CSV file, serving as the foundation for further analysis.

DATASET DESCRIPTION

The extracted dataset consists of several key columns, including **Vehicle Class**, **Year**, and **S.No**, which uniquely identify each entry from the Vahan Parivahan website . Additionally, the dataset captures monthly sales data across the columns **JAN** through **DEC**, representing sales figures for each month of the year. The **TOTAL** column aggregates these monthly values, providing an annual sales figure for each vehicle class. This comprehensive dataset forms the basis for the predictive analysis conducted in this project.

DATA PRE-PROCESSING

During the preprocessing stage, unnecessary columns such as **Vehicle_Class**, **S.No**, and **TOTAL** were dropped from the dataset to streamline the data for analysis. The DataFrame was then melted to transform the monthly sales data into a more suitable format. The new DataFrame now consists of columns **Vehicle Class**, **Year**, **Month**, and **Sale**, which is more appropriate for analysis and model development.

MODEL DEVELOPMENT

For this project, GRU (Gated Recurrent Unit) layers were chosen due to their promising performance compared to traditional RNNs and LSTMs. GRU layers effectively capture temporal dependencies in time-series data while being computationally efficient. The ReLU (Rectified Linear Unit) activation function was utilized for its ability to introduce non-linearity and enhance model performance. The Adam optimizer was employed for its efficient handling of gradient descent and adaptive learning rates.

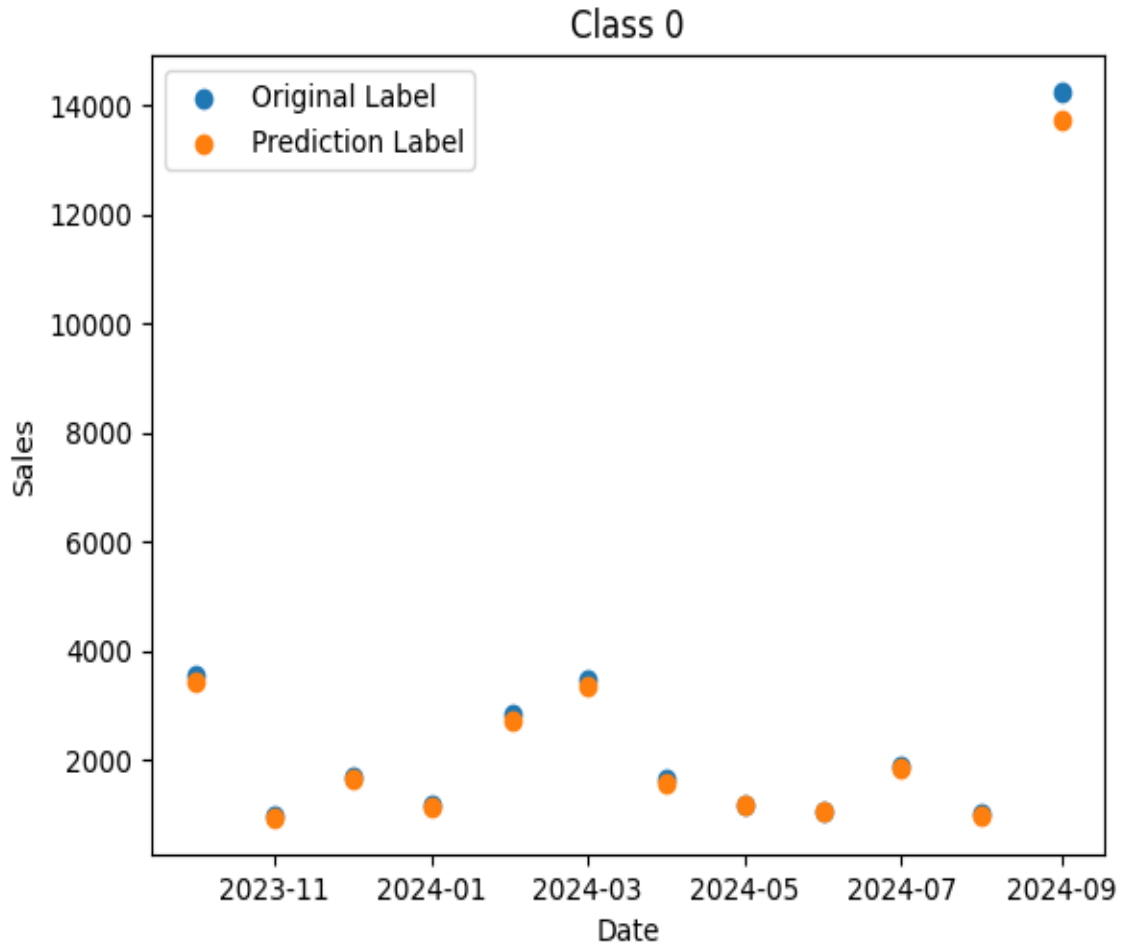
To address the unique characteristics of each vehicle class, six distinct GRU-based models were developed, one for each vehicle class in the dataset. This approach ensures that the model captures the specific patterns and trends associated with each vehicle class, leading to more accurate sales forecasts.

GAINED INSIGHTS

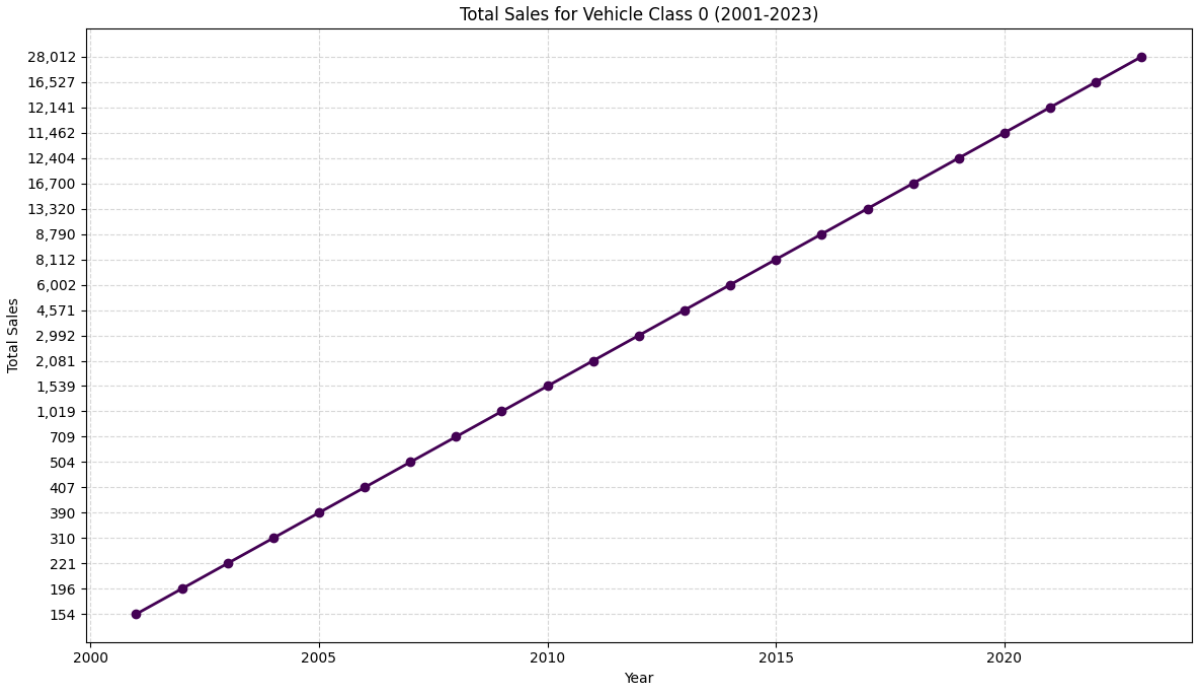
Model-0 [Adapted Vehicle]

The model's performance is highly accurate, as indicated by the R2 score of 0.9978, which suggests that the model explains approximately 99.78% of the variance in the data. The Root Mean Squared Error (RMSE) of 165.07 reflect the average difference between the predicted and actual values, with the RMSE giving an easily interpretable measure in the same units as the target variable. These results suggest that the model has successfully captured the underlying patterns in the data, with minimal error in predictions.

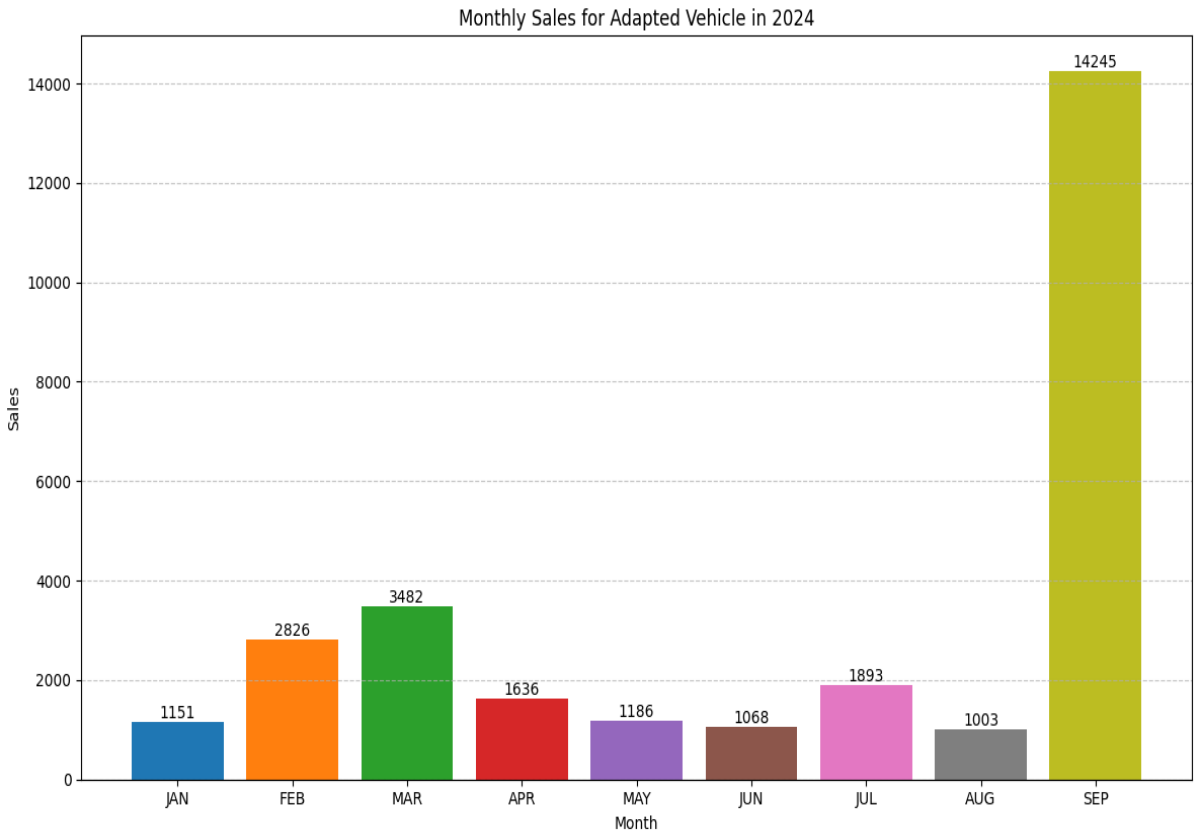
Actual vs Predicted



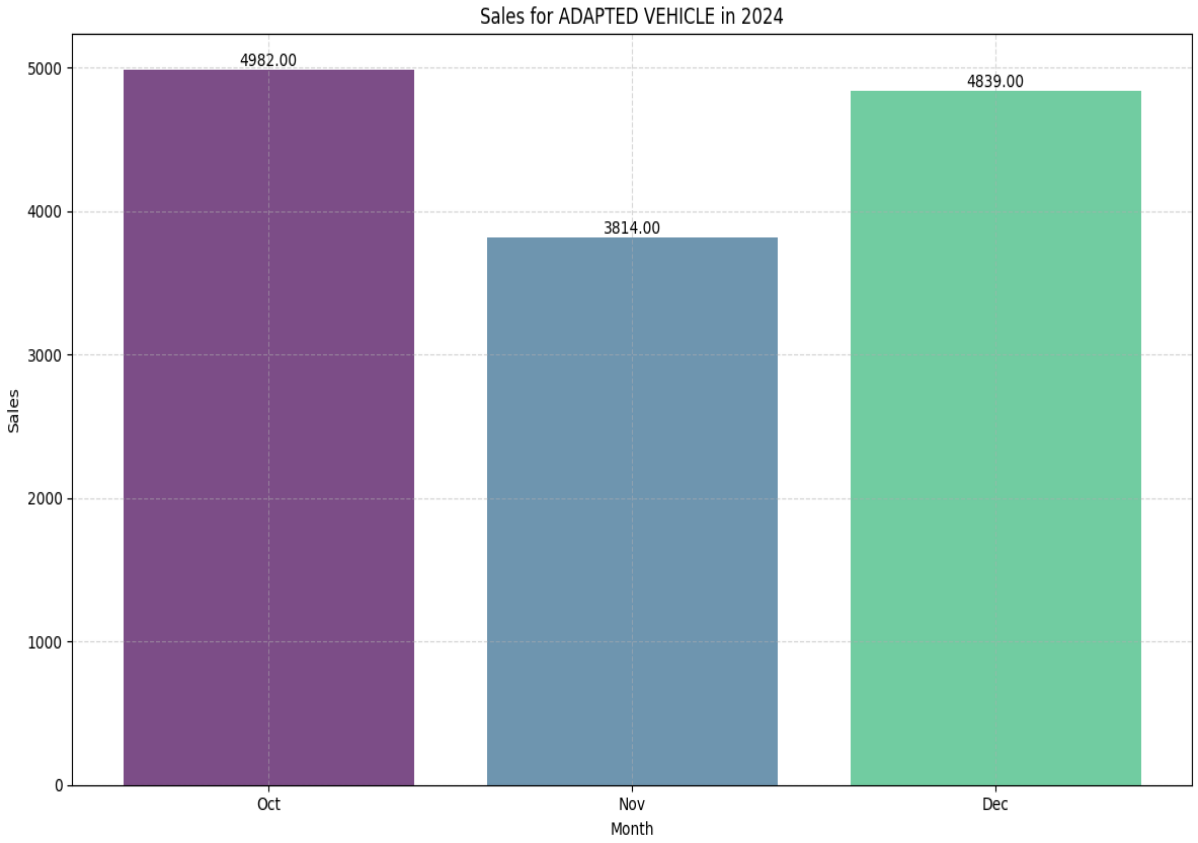
Total Vehicle Sale of Adapted Vehicle



Sale of 2024



Predicted Sale of month Oct, Nov, Dec 2024



The bar chart above illustrates the monthly sales figures for the "Adapted Vehicle" class for the year 2024. The data reveals the following:

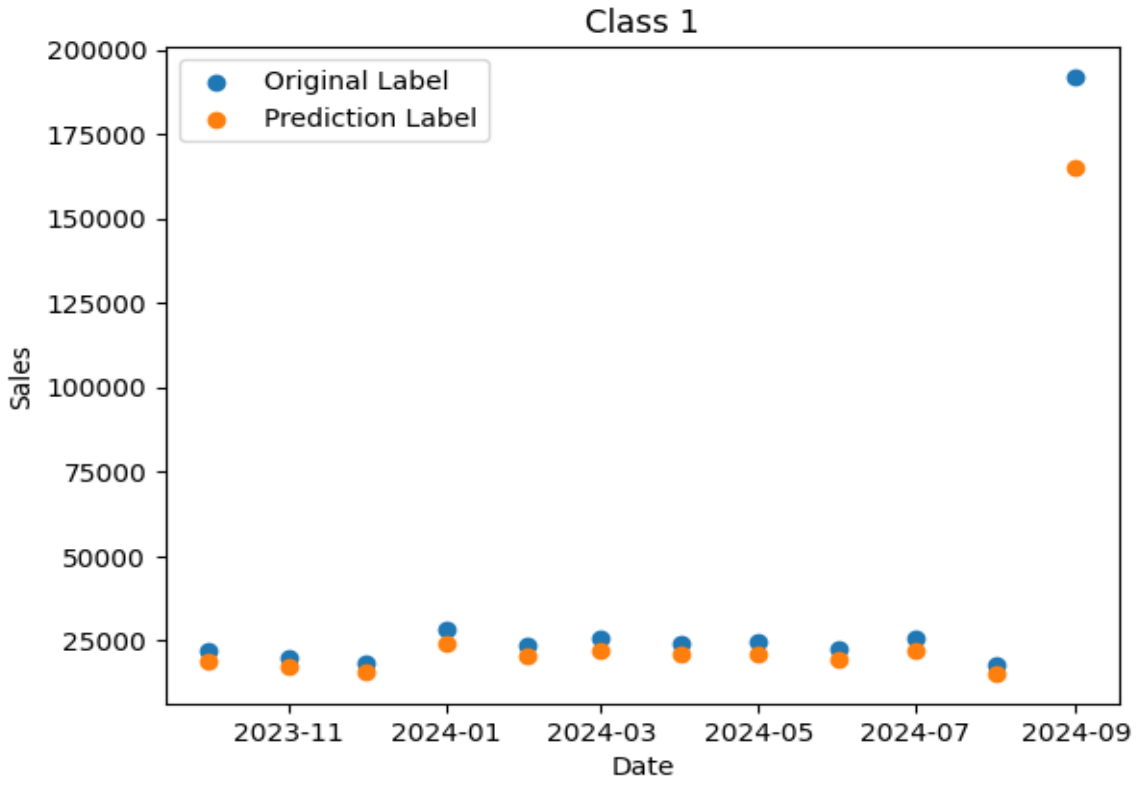
- **October**: 4,982
- **November**: 3,814
- **December**: 4,839

This trend can provide insights into seasonal demand and inform future strategic planning for the vehicle class.

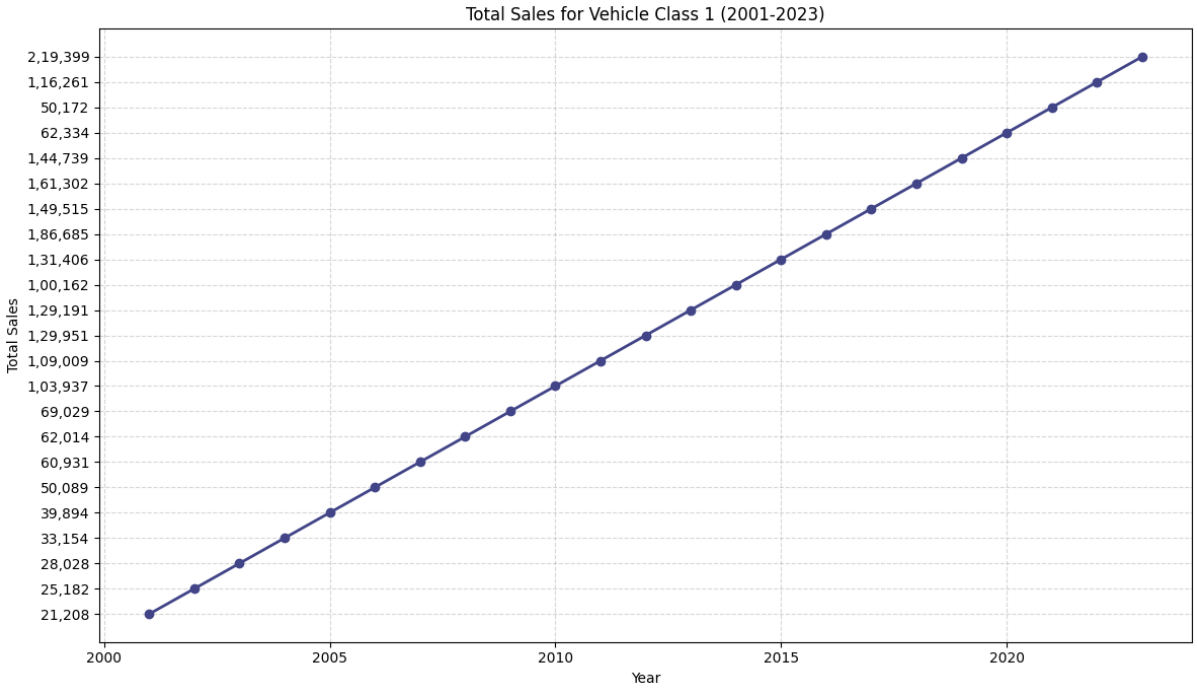
Model-1 [MOTOR CAB]

The model demonstrated impressive performance with an R^2 score of 0.968, indicating a high level of explanatory power and a strong fit between the predicted and actual values. The Root Mean Squared Error (RMSE) was 8,352.25, showing the average magnitude of the prediction errors. These metrics collectively suggest that the model is highly accurate in forecasting vehicle sales, with relatively low prediction errors.

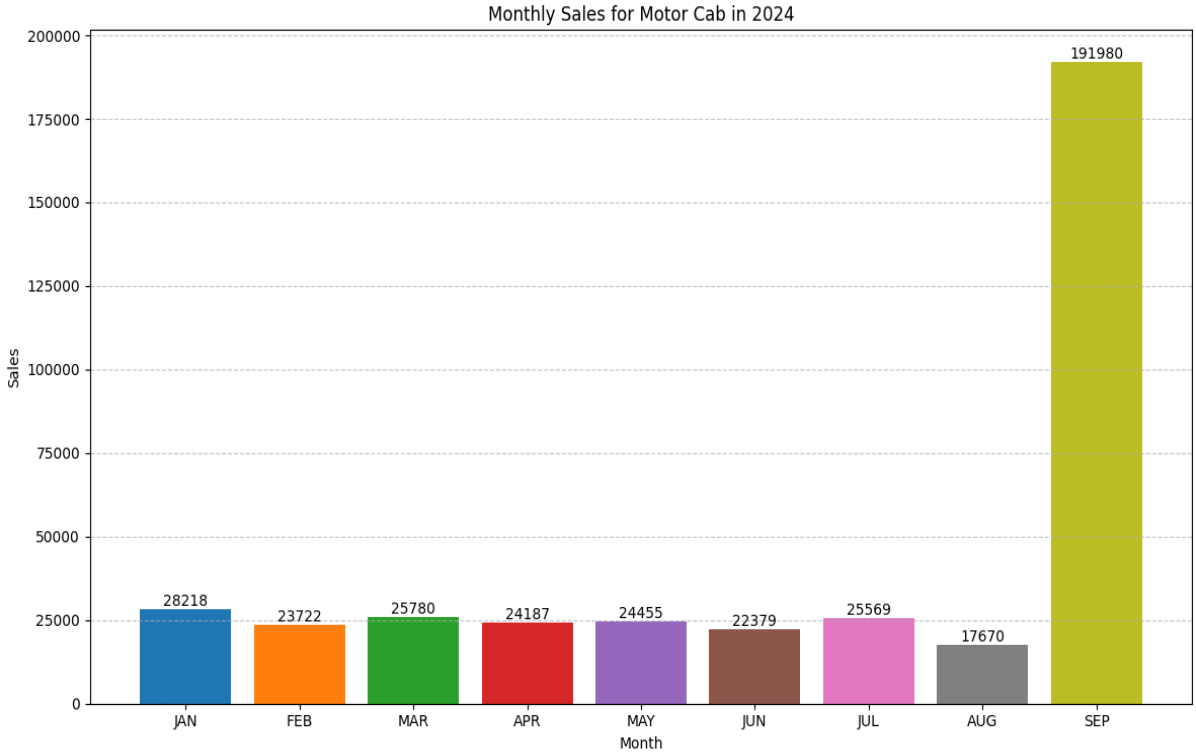
Actual vs Predicted



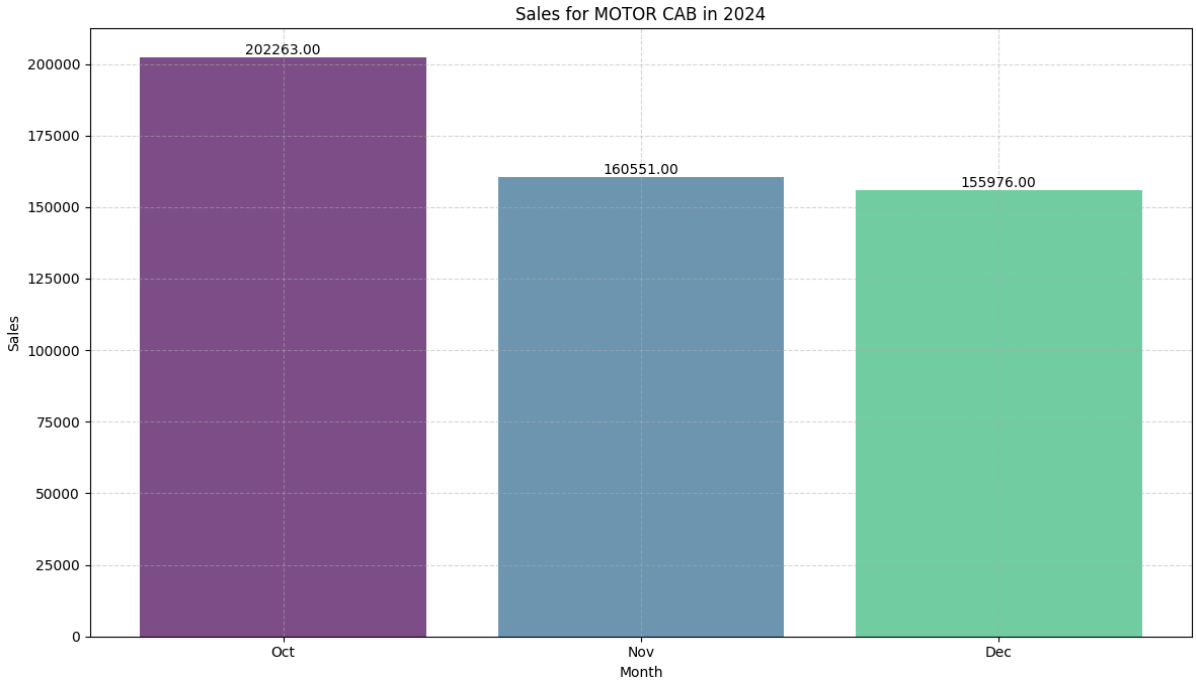
Total Vehicle Sale of Motor Cab



Sale of 2024



Predicted Sale of month Oct, Nov, Dec 2024



The bar chart above depicts the monthly sales figures for the "Motor Cab" class for the year 2024. The data shows:

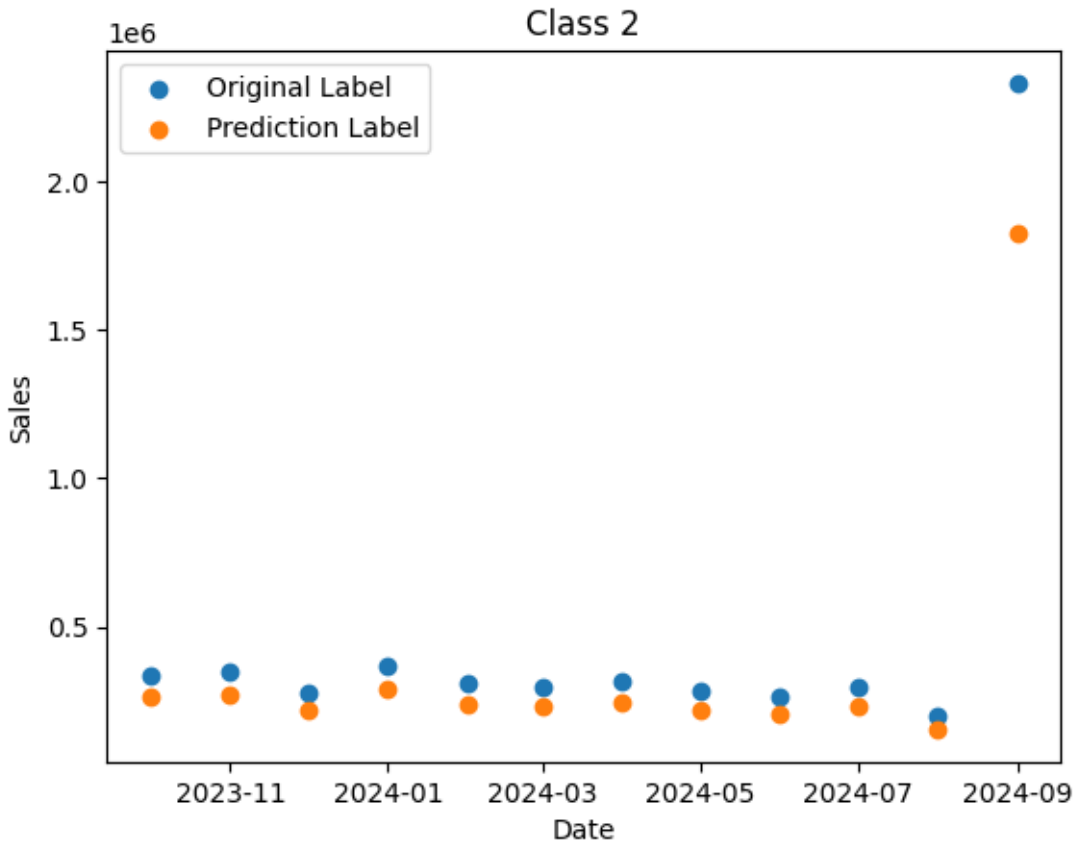
- **October:** 202,263
- **November:** 160,551
- **December:** 155,976

This trend offers valuable insights into seasonal demand and can guide strategic planning for the "Motor Cab" class, aiding in more accurate forecasting and decision-making.

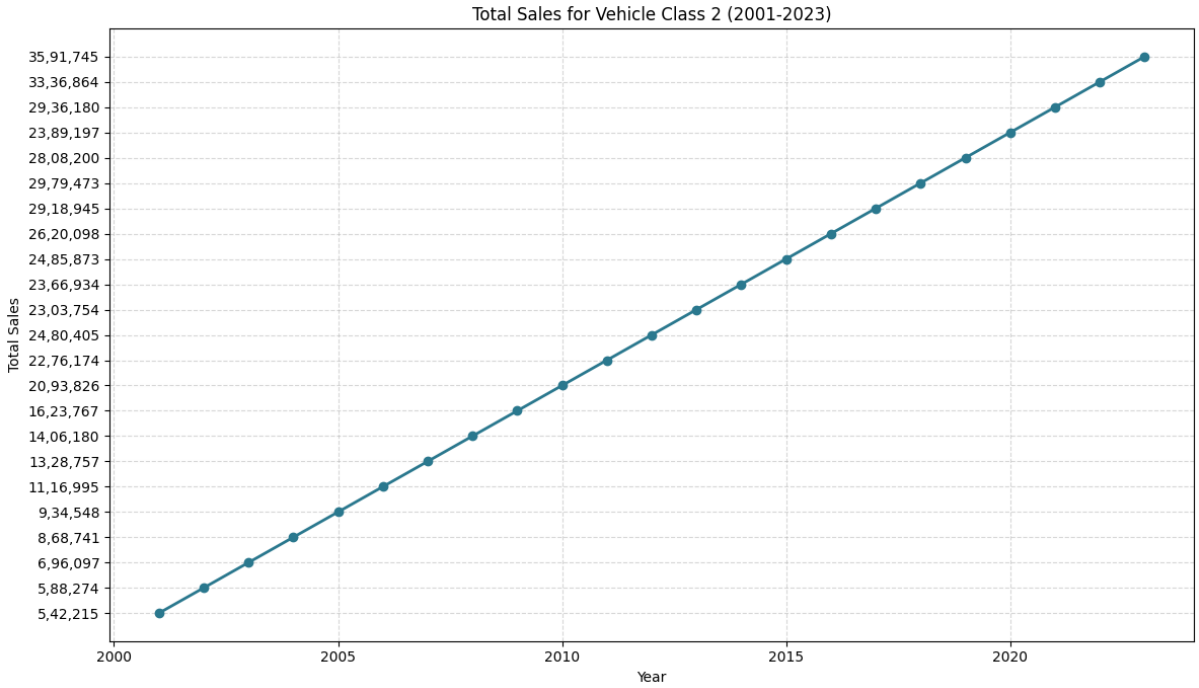
Model-2 [MOTOR CAR]

The model's performance metrics indicate a high level of accuracy in predictions. The R^2 score of 0.92 reflects a strong fit of the model to the data, suggesting that a significant proportion of the variance in sales can be explained by the model. The Root Mean Squared Error (RMSE) is 158,540.03. Despite the large RMSE, the high R^2 score suggests that the model is generally effective in capturing the underlying patterns in the data, although there may be some variation in individual predictions.

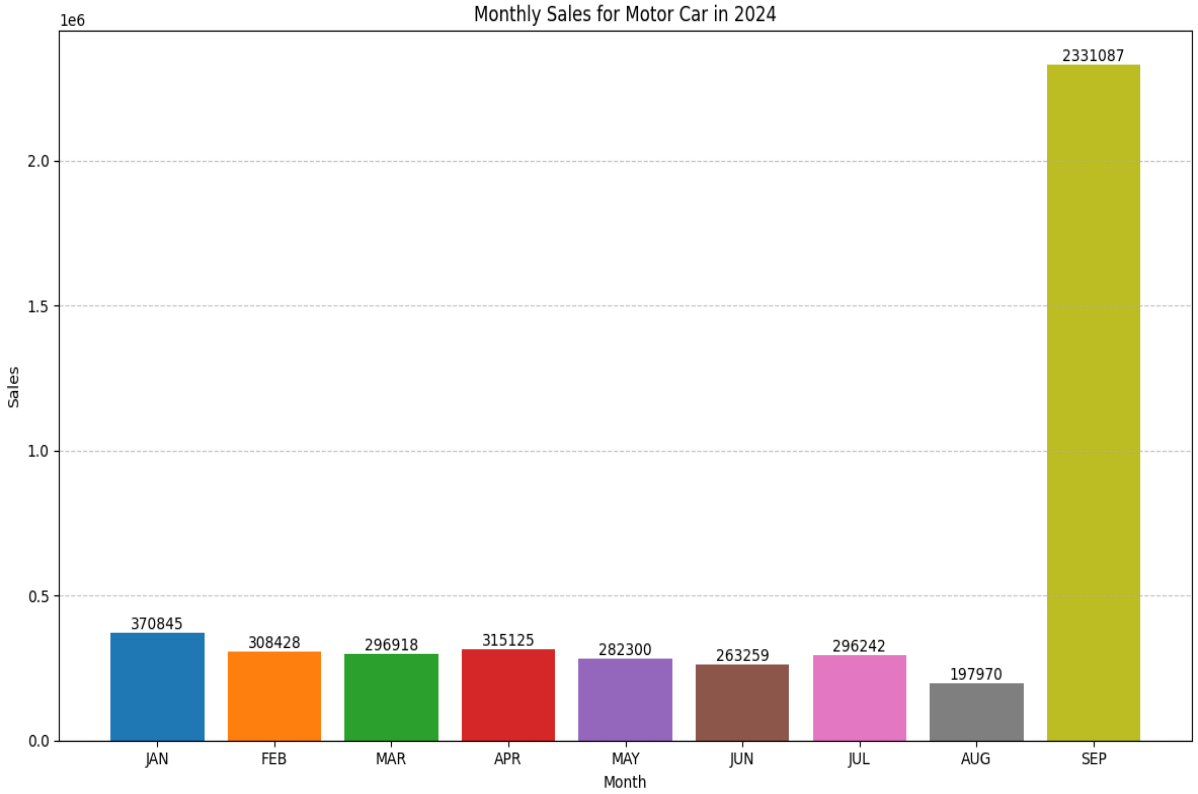
Actual vs Predicted



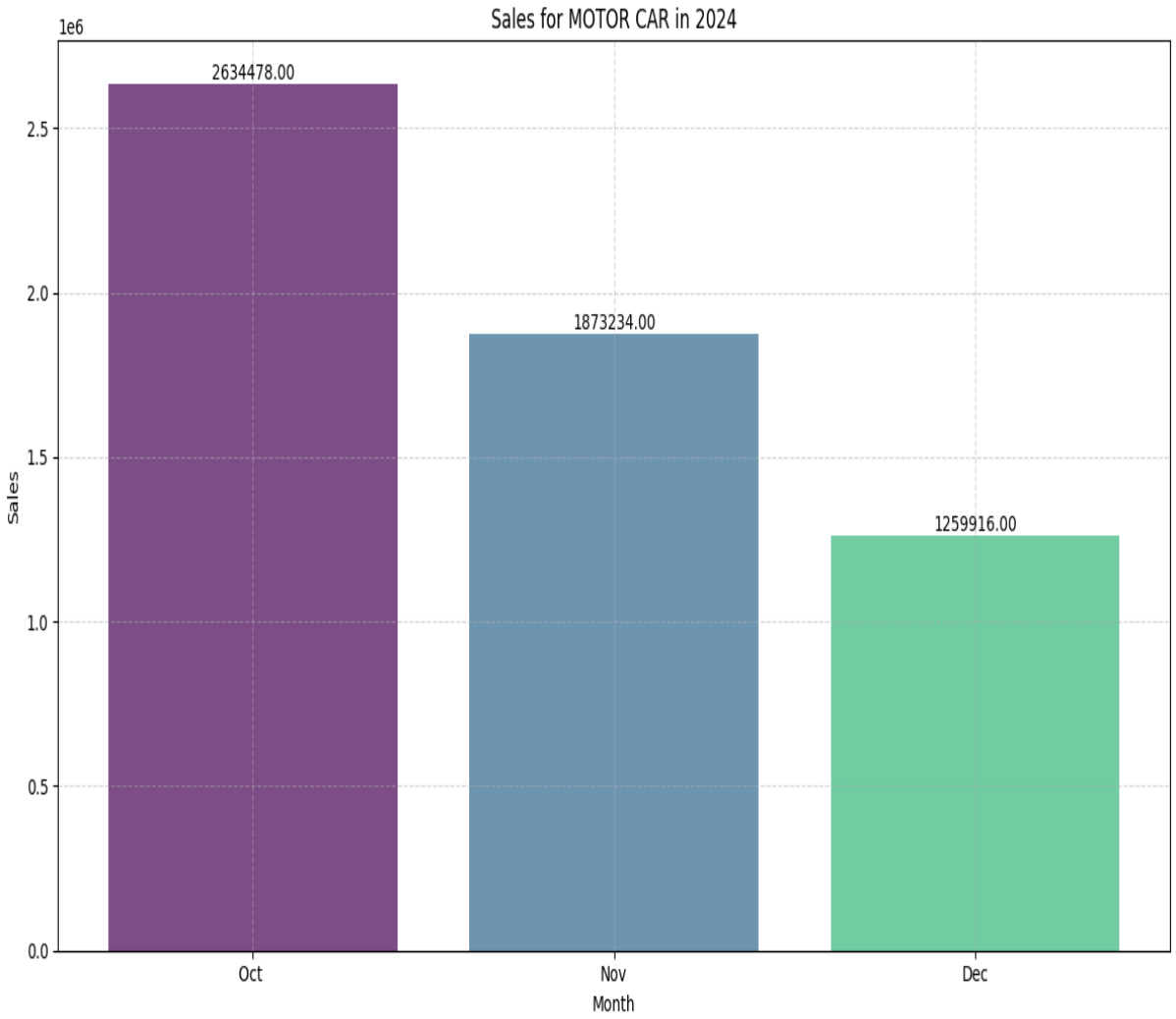
Total Vehicle Sale of Motor Car



Sale of 2024



Predicted Sale of month Oct, Nov, Dec 2024



The bar chart above presents the projected monthly sales figures for the "Motor Car" class for the year 2024. The data reveals the following predictions:

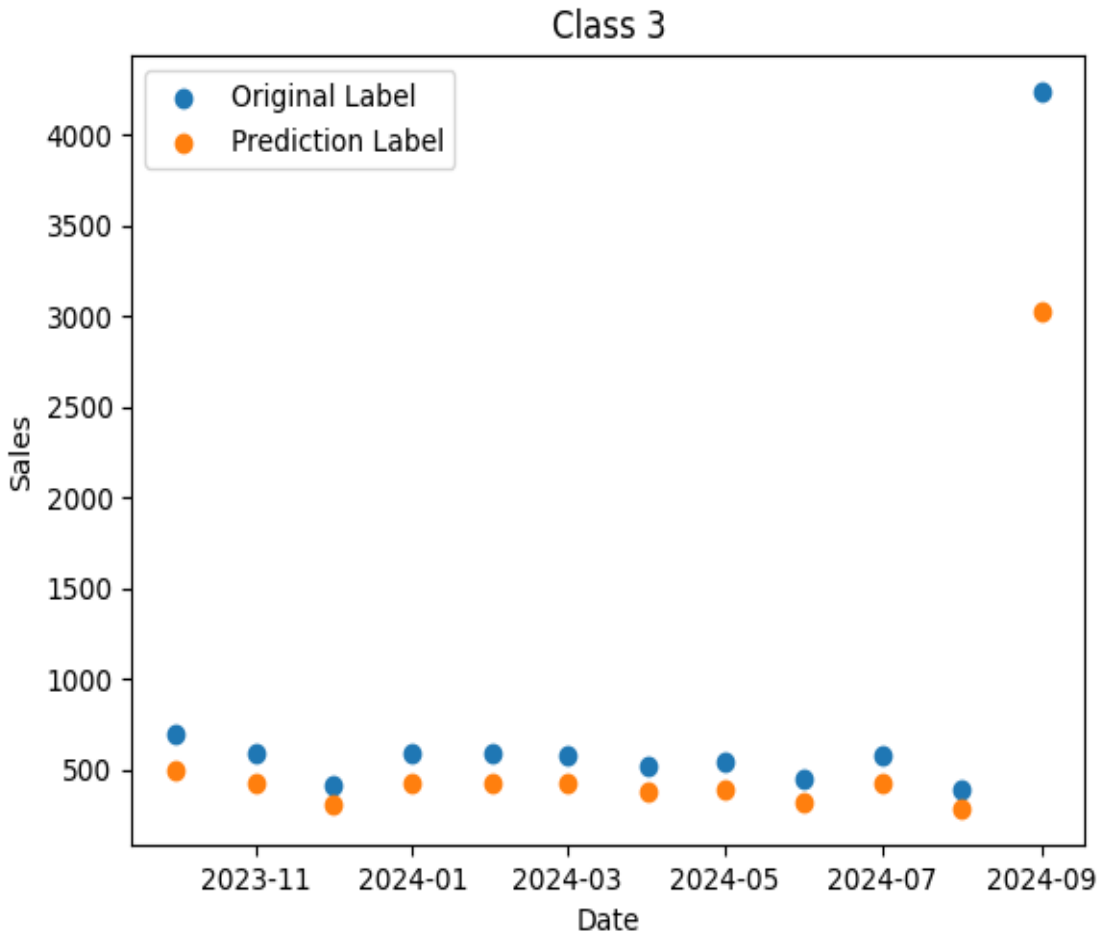
- **October**: 2,634,478
- **November**: 1,873,234
- **December**: 1,259,916

This trend offers valuable insights into seasonal demand and can guide strategic planning for the "Motor Car" class, aiding in more accurate forecasting and decision-making.

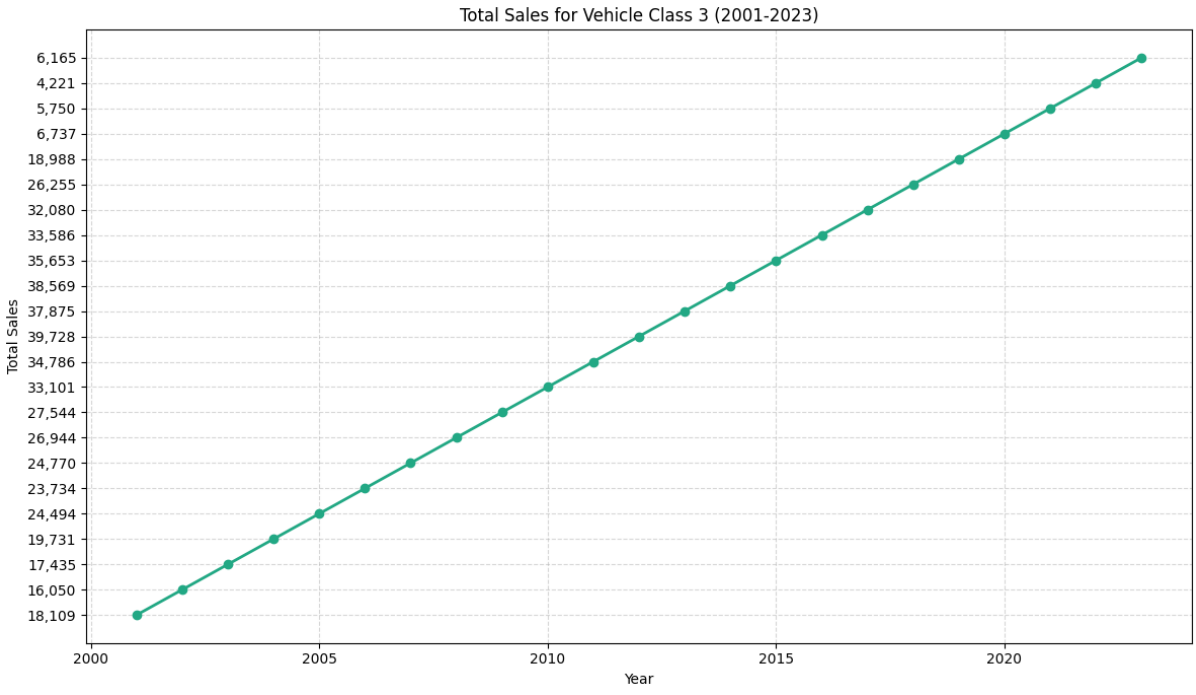
Model-3 [OMNI BUS]

The model's performance for predicting sales values is reflected in the following metrics: The R2 score is 0.86, suggesting that the model explains approximately 86% of the variance in the sales data. The Root Mean Squared Error (RMSE) is 379.84, showing the standard deviation of the prediction errors. These metrics collectively demonstrate that the model has a strong fit and predictive accuracy, effectively capturing the underlying patterns in the sales data.

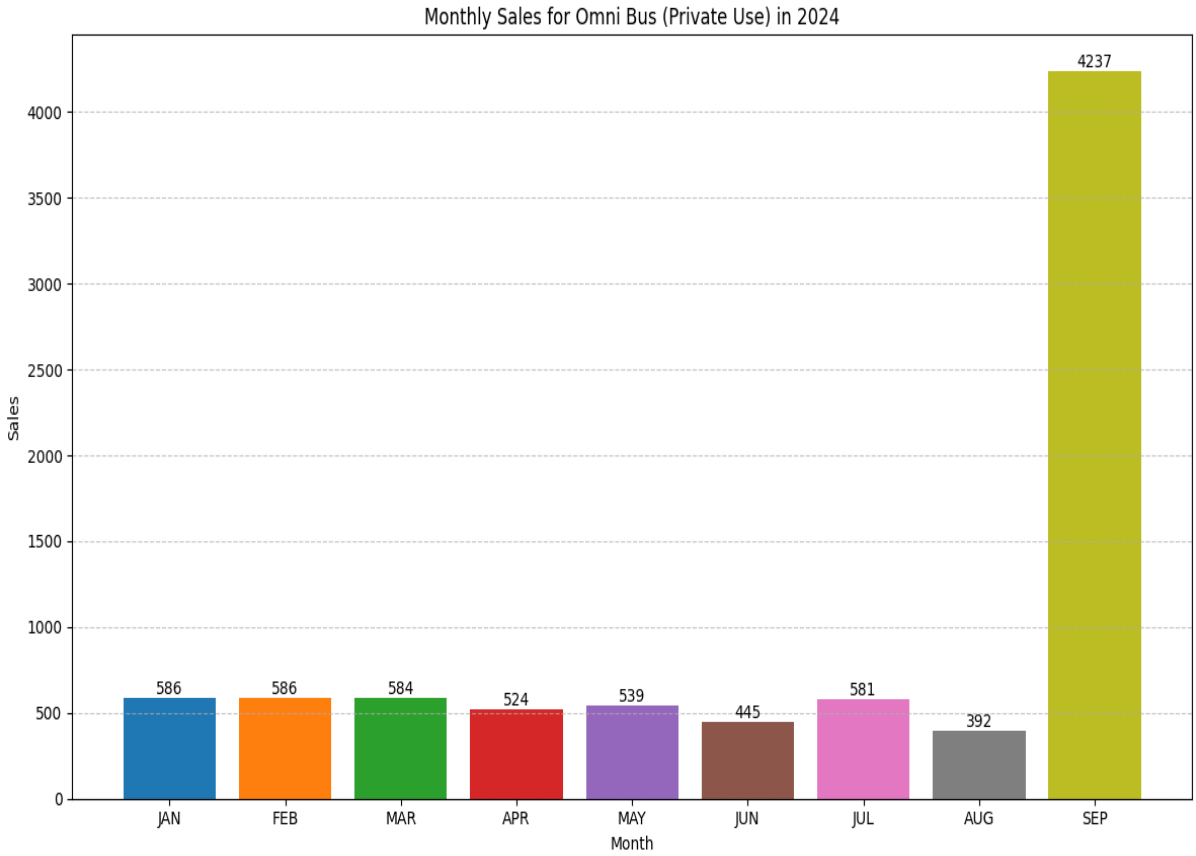
Actual vs Predicted



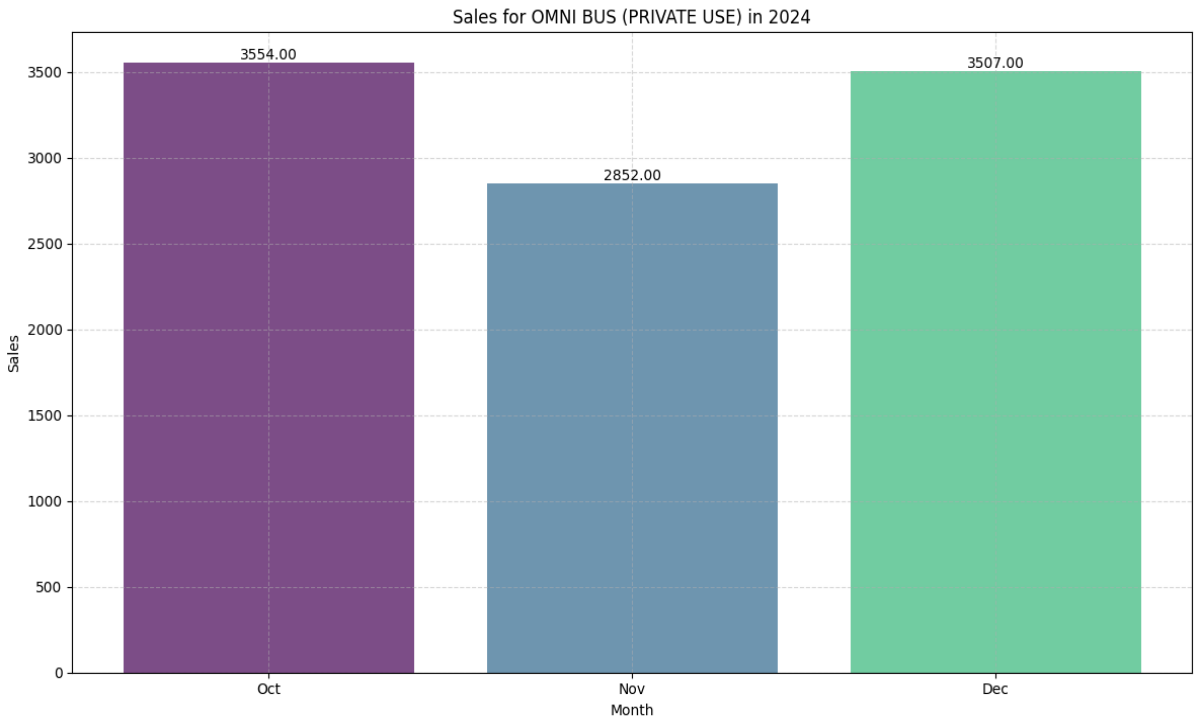
Total Vehicle Sale of Omni Bus



Sales of 2024



Predicted Sale of month Oct, Nov, Dec 2024



The bar chart above presents the monthly sales figures for the "Omni Bus (Private Use)" class for 2024. The data reveals the following:

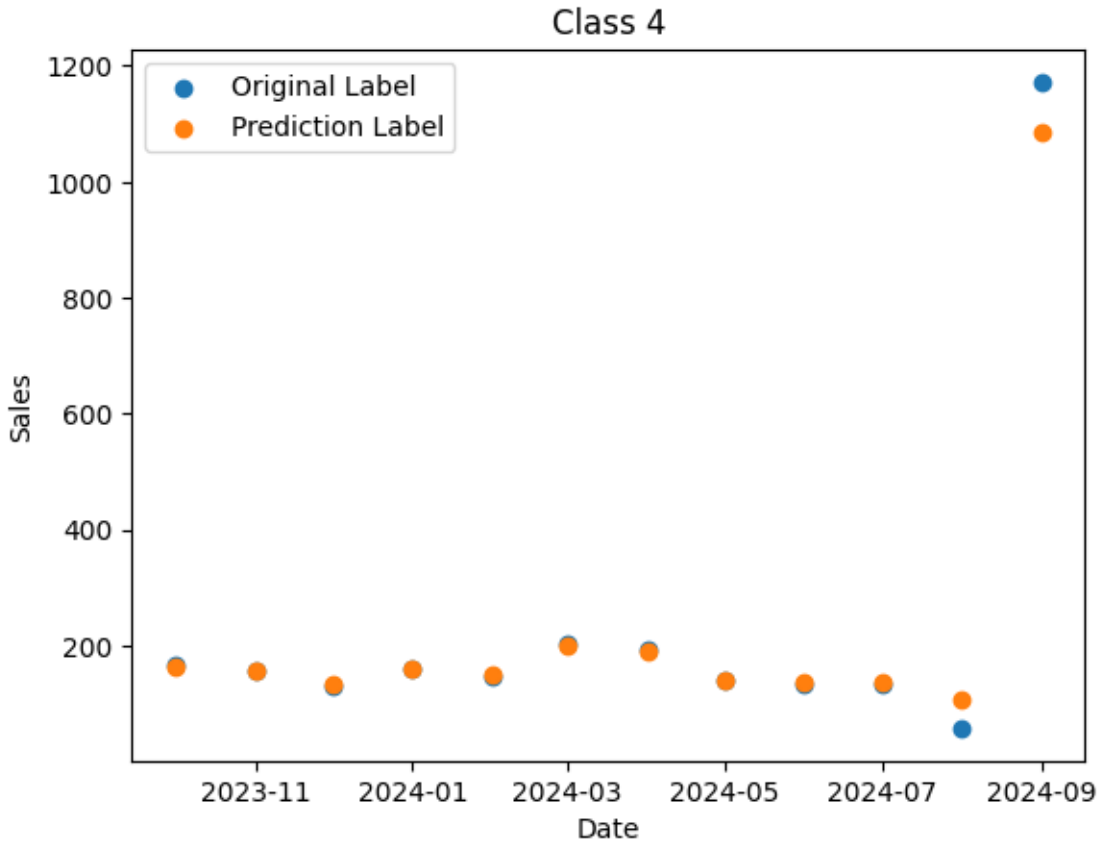
- **October:** 3,554
- **November:** 2,852
- **December:** 3,507

This trend offers valuable insights into seasonal demand and can guide strategic planning for the "Omni Bus" class, aiding in more accurate forecasting and decision-making.

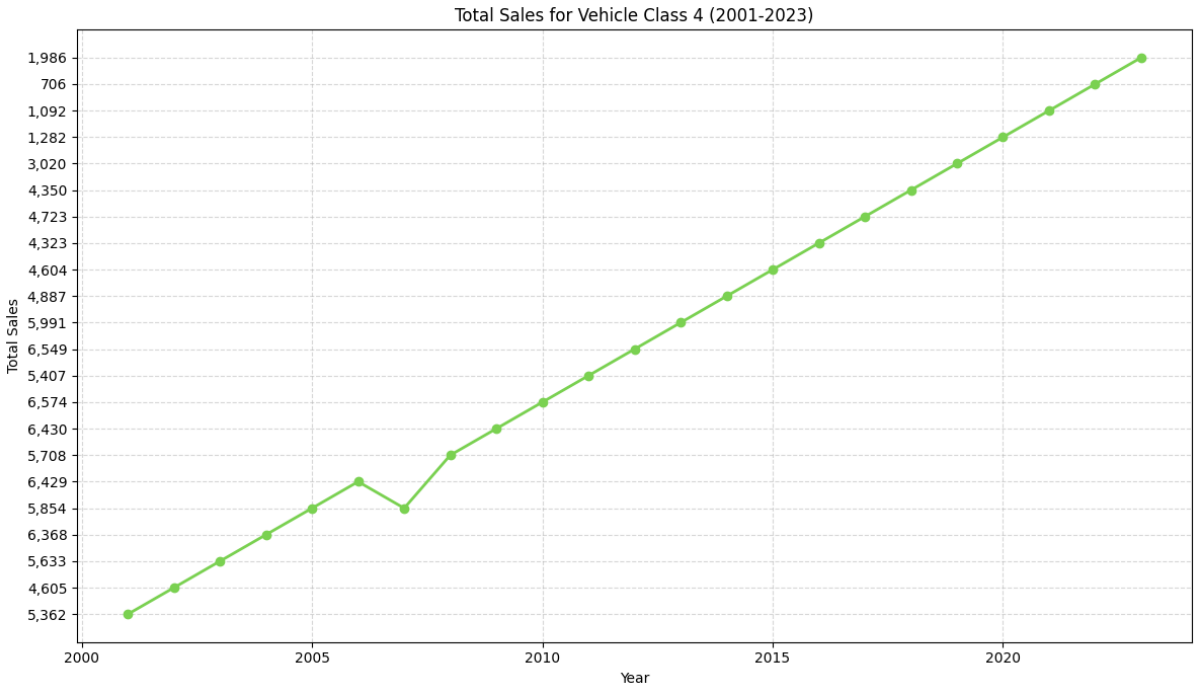
Model-4 [PRIVATE SERVICE VEHICLE]

The model's performance for the " Private Service Vehicle " class demonstrates impressive accuracy. With an R^2 score of 0.9897, the model effectively explains nearly 98% of the variance in the sales data. The Root Mean Squared Error (RMSE) is 28.94, reflecting a high degree of precision in the predictions. These results indicate that the model provides a highly reliable forecast for the " Private Service Vehicle " class, making it a valuable tool for predicting future sales and aiding in strategic planning.

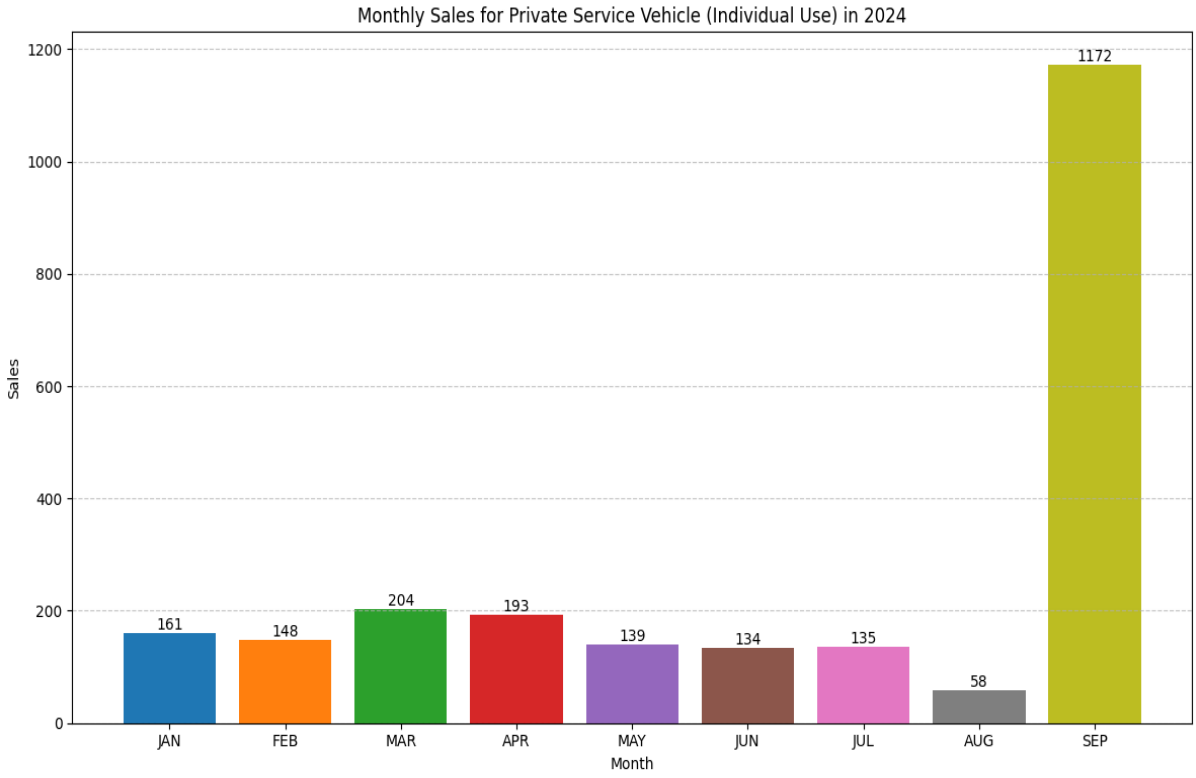
Actual vs Predicted



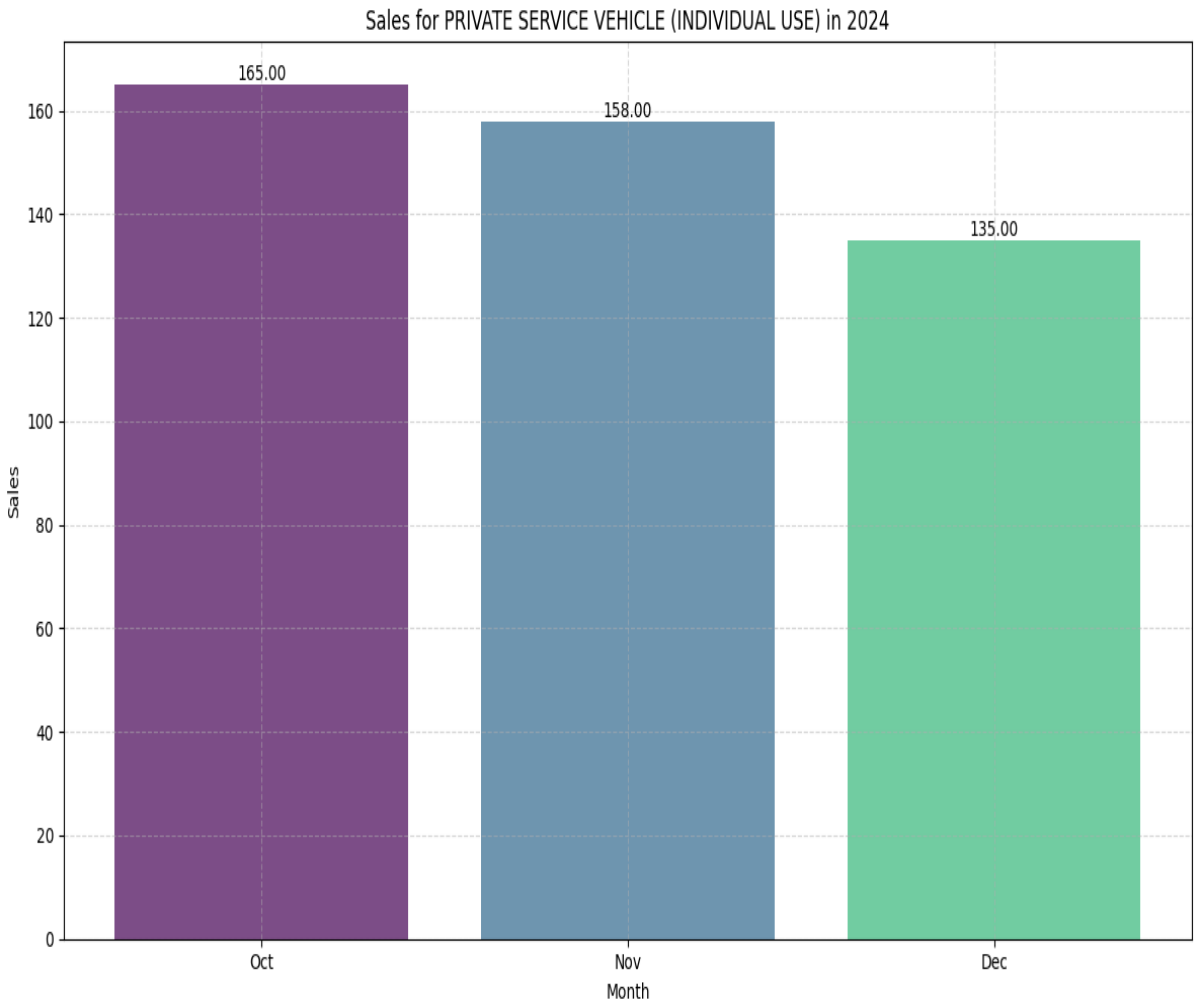
Total Vehicle Sale of Private Service Vehicle



Sales of 2024



Predicted Sale of month Oct, Nov, Dec 2024



The bar chart above illustrates the monthly sales figures for the " Private Service Vehicle " class for the year 2024. The data reveals the following:

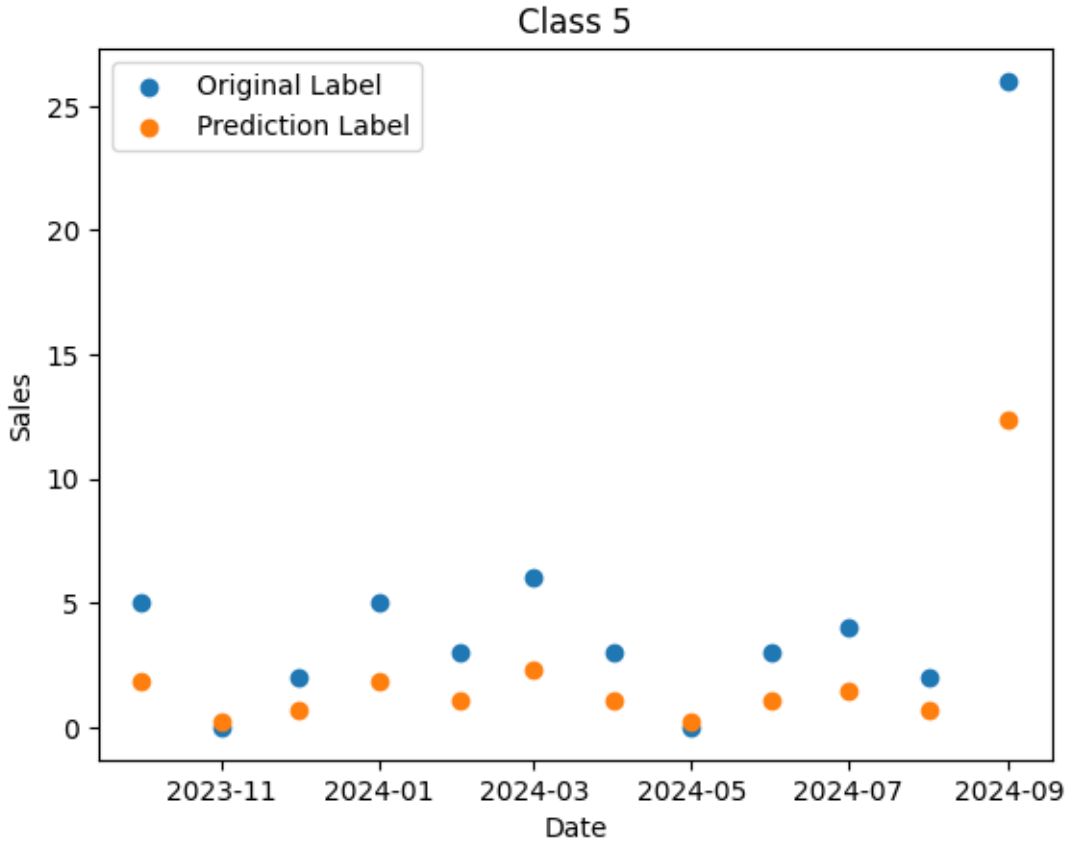
- **October**: 165
- **November**: 158
- **December**: 135

This trend offers valuable insights into seasonal demand and can guide strategic planning for the " Private Service Vehicle " class, aiding in more accurate forecasting and decision-making.

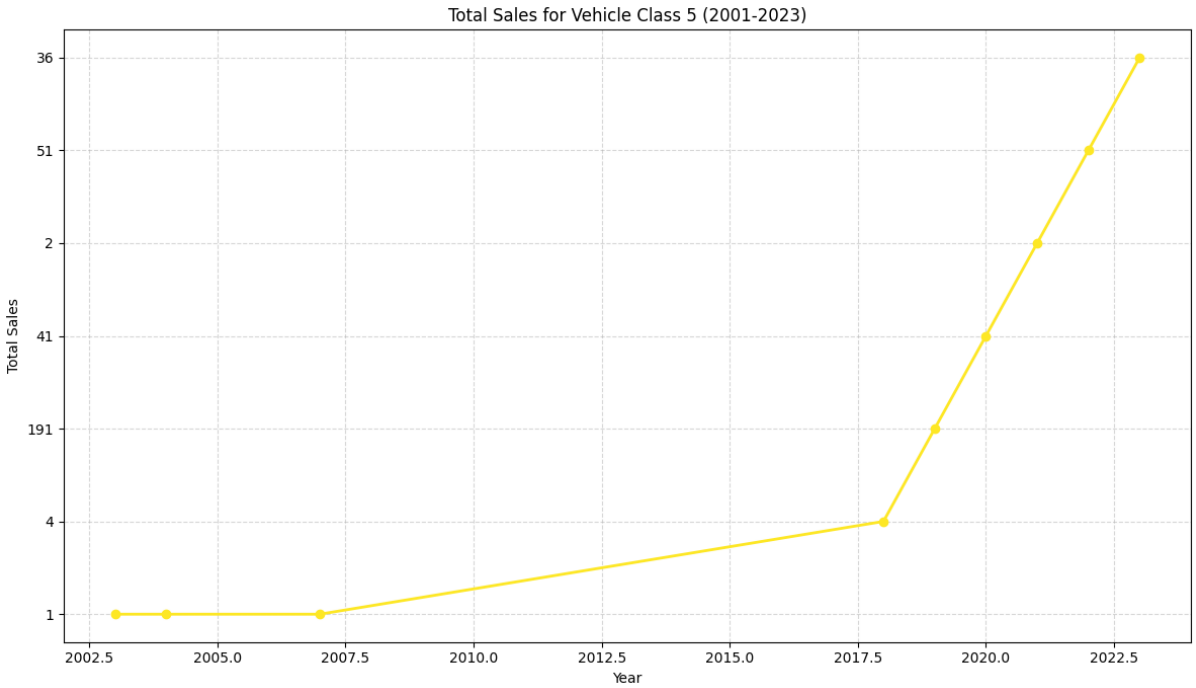
Model-5 [QUADRICYCLE]

The model's performance, reflected by an R2 score of 0.5406 and an RMSE of 4.47, indicates moderate predictive accuracy. The presence of low sales value in the dataset likely contributed to this outcome, as the lack of sales data on certain dates can challenge the model's ability to make accurate predictions. This suggests that the model might need further refinement or additional data to handle such inconsistencies better.

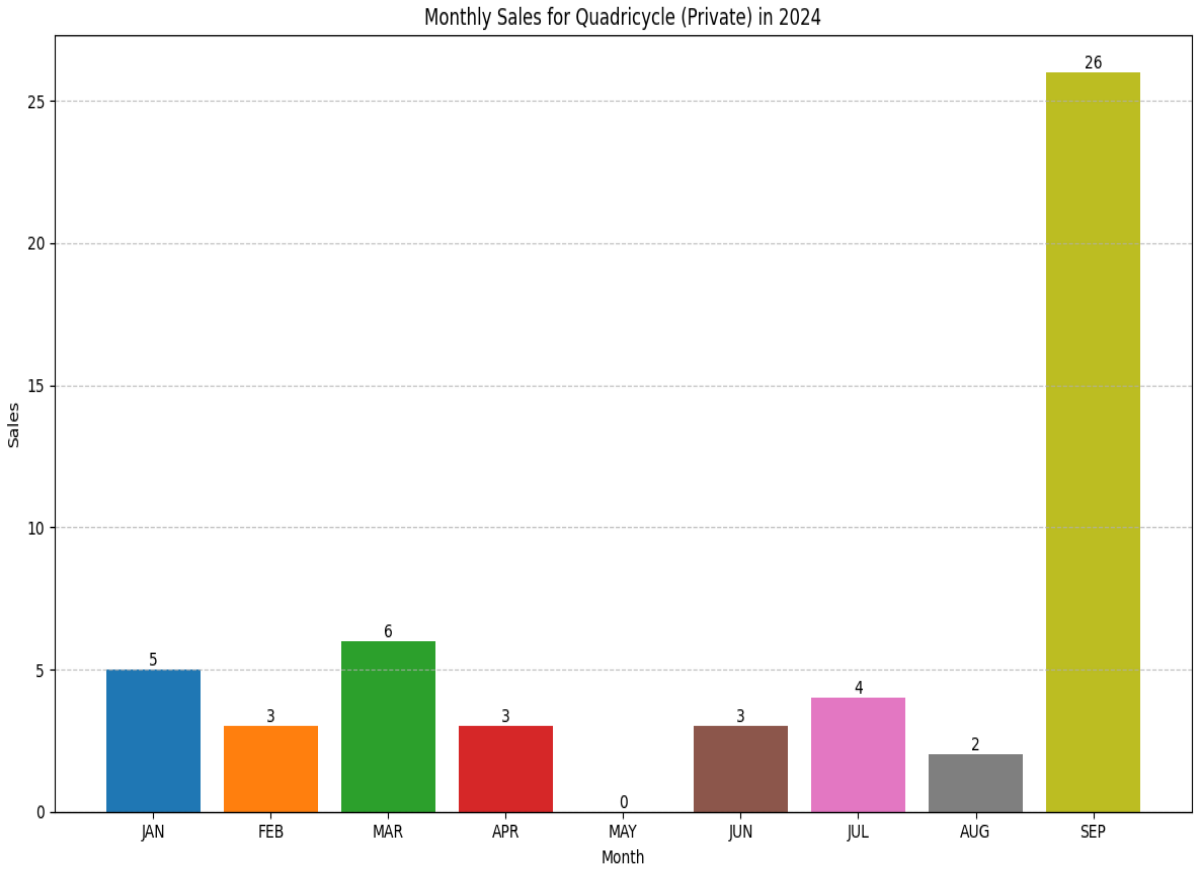
Actual vs Predicted



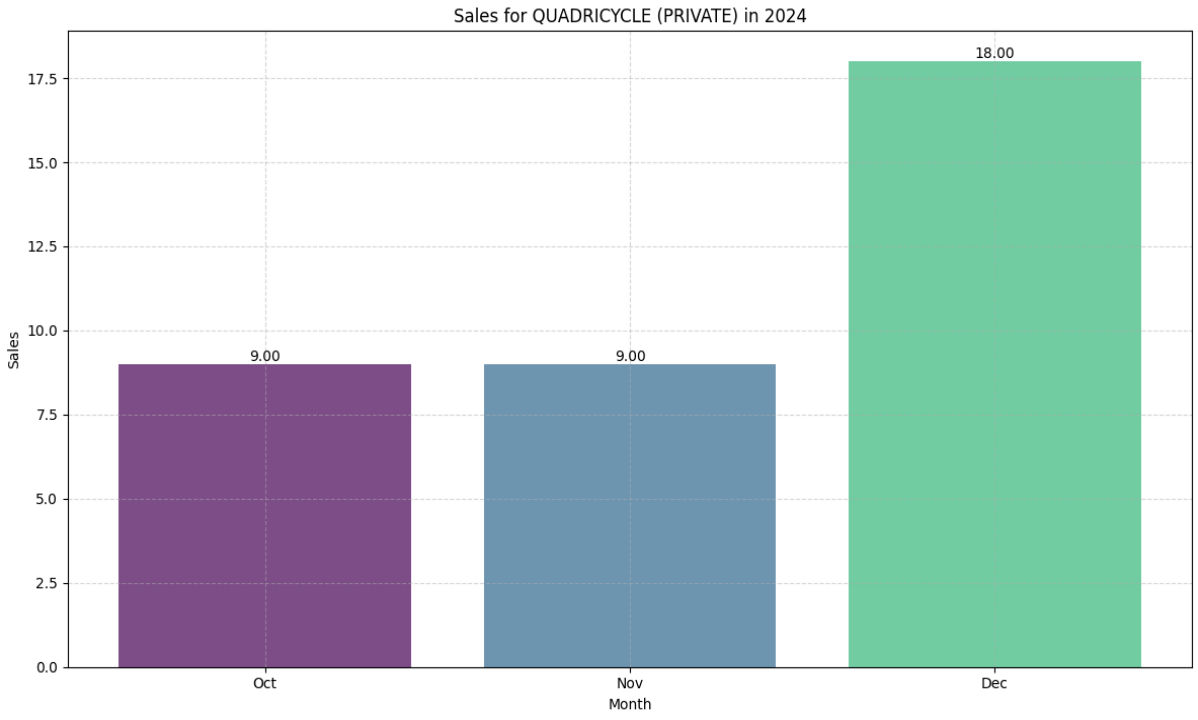
Total Vehicle Sale of Quadricycle



Sales of 2024



Predicted Sale of month Oct, Nov, Dec 2024



The above bar chart for the Quadricycle vehicle class shows the following predicted sales for the year 2024:

- October: 9
- November: 9
- December: 18

This trend offers valuable insights into seasonal demand and can guide strategic planning for the " Quadricycle " class, aiding in more accurate forecasting and decision-making.