### **Regression Report**

#### **1. Introduction**

* **Dataset Description**: Describe the pakwheels\_used\_cars.csv dataset, which includes features like engine\_cc, mileage, and the target variable price.
* **Objective**: Explain the goal of predicting car prices using various regression algorithms based on the given features.

#### **2. Data Cleaning and Preparation**

* **Loading the Data**: Load the dataset and display the first few rows.
* **Handling Missing Values**:
  + Identify missing values and print the results.
  + Fill missing values in numerical columns with the mean.
  + For categorical columns, fill missing values with the mode.
* **Encoding Categorical Variables**: Convert categorical features into numerical values using Label Encoding.
* **Scaling Numerical Features**: Normalize numerical features using StandardScaler.

#### **3. Data Analysis and Visualization**

* **Summary Statistics**: Generate and display summary statistics for the dataset.
* **Histograms**: Create histograms to visualize the distribution of numerical features.
* **Scatter Plots**: Generate scatter plots to explore relationships between features and the target variable.
* **Box Plots**: Use box plots to visualize the distribution and detect outliers in the features.
* **Correlation Heatmaps**: Create heatmaps to show the correlation between numerical features and the target variable.

#### **4. Model Building**

* **Splitting Data**: Split the data into training and testing sets.
* **Applying Regression Models**:
  + **Linear Regression**: Train and evaluate a Linear Regression model.
  + **Decision Tree Regression**: Train and evaluate a Decision Tree Regressor.
  + **Random Forest Regression**: Train and evaluate a Random Forest Regressor.

#### **5. Model Evaluation**

* **Performance Metrics**: Evaluate and compare the performance of each regression model using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.
* **Performance Comparison**:
  + Create a DataFrame to compare the performance of the regression models.
  + Plot a bar chart to visualize the performance comparison.

#### **6. Conclusion**

* **Summary of Findings**: Summarize the findings from the regression models, highlighting which model performed best and why.
* **Future Work**: Suggest areas for improvement, such as tuning model parameters, incorporating more features, or using advanced regression techniques.

The graphs for pakwheels regression:





