**Car Sales Data Analysis and Resale Value Prediction using Regression.**

**Objective:**

Explore and understand car sales data, then build a model to predict a car's resale value based on key features.

**1. Load and Inspect Data:**

* What: Load car sales data from a file and show the first 5 rows to understand what the data looks like.
* Why: Get a quick overview of the dataset.

**2. Univariate Analysis for Categorical Data:**

* What: Create bar charts to show how different factors like manufacturers, vehicle types, and fuel efficiency relate to sales.
* Why: Understand which factors might influence car sales.

**3. Analysing Numerical Variables:**

* What: Make histograms to visualize the distribution of numerical data (like width, sales, engine size).
* Why: See the range and patterns in numerical data.

A graph of a graph showing a graph of cars

Description automatically generated with medium confidence

**4. Data Preprocessing:**

* What: Check for missing values and remove rows that have incomplete data.
* Why: Clean the data for accurate analysis.

**5. Normalizing Numerical Data:**

* What: Use statistical methods to transform numerical data for better model performance.
* Why: Ensure fair treatment of different numerical variables.

**6. Correlation Analysis:**

* What: Create a heatmap to show how numerical variables relate to each other.
* Why: Identify which features are strongly related.

A close-up of a graph

Description automatically generated

**7. Building a Regression Model:**

* What: Construct a mathematical model that predicts resale value based on selected features (price, curb weight, fuel capacity).
* Why: Understand how these features contribute to the resale value.
* A graph showing the relationship between a couple of numbers

  Description automatically generated with medium confidence

**8. Testing the Model:**

* What: Split the dataset into training and testing sets, build a model on the training set, and test its accuracy on the testing set.
* Why: Ensure the model works well on new, unseen data.

**9. Model Evaluation:**

* What: Calculate Root Mean Squared Error (RMSE) to see how accurate the predictions are.
* Why: Understand how close the predicted values are to the actual values.

**10. Visualization:**

* What: Plot a scatter plot of actual vs. predicted resale values.
* Why: Visualize how well the model predictions align with the actual data.
* A graph with a red line

  Description automatically generated