Mini Project: Predicting Car Prices Using Multivariate Linear Regression

Objective:

The goal of this project is to build a model that predicts the price of a car based on various features like the car's make, model, year, engine size, mileage, and other relevant factors.

Problem Statement:

You are provided with a dataset containing information about different cars. Your task is to predict the price of a car given its features using multivariate linear regression.

Dataset Features:

- **Make**: The brand of the car (e.g., Toyota, Honda, BMW).
- **Model**: The specific model of the car.
- **Year**: The year the car was manufactured.
- **Mileage**: The total miles driven by the car.
- **Engine Size**: The engine capacity in liters.
- **Fuel Type**: The type of fuel used (e.g., petrol, diesel, electric).
- **Transmission**: Type of transmission (e.g., manual, automatic).
- **Horsepower**: The car's horsepower.
- **Number of Doors**: The number of doors in the car.
- **Price (Target Variable)**: The price of the car in dollars.

Steps:

1. **Data Collection and Exploration**: car_prices_dataset.csv is attached. Explore the dataset to understand the relationships between features and the target variable. 2. **Data Preprocessing**:
    - ○ Handle missing values.
    - ○ Convert categorical variables (e.g., make, model, fuel type) into numerical values using techniques like one-hot encoding.
    - ○ Scale the features to standardize them, if necessary.
3. **Feature Selection**: Identify which features have the most significant impact on car prices using correlation analysis or other techniques.
4. **Model Development**:
    - ○ Split the dataset into training and testing sets.
    - ○ Build a multivariate linear regression model using the training data. ○ Train the model and evaluate its performance using metrics like Mean Squared Error (MSE) and R^2.
5. **Model Evaluation and Interpretation**:
    - ○ Analyze how well the model performs on the test data.
    - ○ Interpret the coefficients to understand the impact of each feature on car prices.
6. **Model Improvement**:
    - ○ Experiment with feature engineering or add polynomial features to improve the model's accuracy.

7. **Conclusion and Insights**: Summarize your findings and explain which features are the most influential in predicting car prices.

Expected Outcome:

By the end of this project, you should have a good understanding of how multivariate linear regression works and how it can be applied to real-world datasets for predictive analysis.