**Car Price Prediction Project**

**1. Introduction**

**Problem Statement**

The objective of this project is to develop a machine learning model that can predict the selling price of a car based on various features such as the year of manufacture, present price, kilometers driven, fuel type, selling type, transmission type, and number of previous owners.

**2. Data Overview**

**Data Source**

The dataset used in this project is provided in a CSV file named `car data.csv`. It contains information about various cars including their attributes and selling price.

**Data Description**

The dataset contains the following columns:

* `Car\_Name`: Name of the car
* `Year`: Year of manufacture
* `Selling\_Price`: Price at which the car is being sold (in lakhs)
* `Present\_Price`: Current ex-showroom price of the car (in lakhs)
* `Driven\_kms`: Number of kilometers driven
* `Fuel\_Type`: Type of fuel used by the car (Petrol/Diesel/CNG)
* `Selling\_type`: Whether the car is new or used
* `Transmission`: Type of transmission (Manual/Automatic)
* `Owner`: Number of previous owners

**Data Preprocessing**

* Load the data using pandas.
* Check for null values and duplicates.
* Drop duplicate rows.
* Encode categorical variables using `LabelEncoder`.
* Drop unnecessary columns (`Car\_Name`).

**3. Exploratory Data Analysis (EDA)**

**Visualizations**

* Count plots for `Fuel\_Type`, `Selling\_type`, and `Transmission`.
* Scatter plots to show the relationship between `Selling\_Price` and other numerical features.
* Pair plot to visualize the relationships between features.
* Heatmap to show the correlation between different attributes.

**4. Model Development**

**Feature Selection**

The following features were selected for model development:

* `Year`
* `Present\_Price`
* `Driven\_kms`
* `Fuel\_Type`
* `Selling\_type`
* `Transmission`
* `Owner`

**Target Variable**

* `Selling\_Price`

**Data Splitting**

The data was split into training and testing sets using `train\_test\_split` from sklearn with a test size of 33%.

**Model Training**

A Linear Regression model was used for this project. The model was trained on the training set and evaluated on the testing set.

**5. Model Evaluation**

Metrics

To assess our machine learning model's performance, we utilized:

Mean Absolute Error (MAE): Measures the average magnitude of errors in predictions, calculated as the average of the absolute differences between predicted and actual values. It provides a straightforward interpretation of the prediction error.

Mean Squared Error (MSE): Measures the average of the squares of the errors, calculated by squaring the differences between predicted and actual values and then averaging them. MSE gives higher weight to larger errors, useful for identifying significant prediction errors.

These metrics provide insights into the model's accuracy and generalization capabilities.

**6. Prediction Function**

To make the model practical, we developed a function predict\_car\_price that estimates the selling price of a car based on user inputs.

**Function Description**

Inputs:

Year: Year of manufacture.

Present\_Price: Current ex-showroom price (in lakhs).

Driven\_kms: Total kilometers driven.

Fuel\_Type: Type of fuel (Petrol/Diesel/CNG).

Selling\_type: New or used car.

Transmission: Type of transmission (Manual/Automatic).

Owner: Number of previous owners.

**Process**

Encoding Categorical Variables: Converts categorical inputs to numerical values using predefined mappings, as the model was trained on numerical data.

Creating a DataFrame: Constructs a new DataFrame with the input values, matching the structure of the training data.

Making Predictions: Utilizes the trained linear regression model to predict the car's selling price based on the input features and returns the estimated price.

This function allows users to input specific car details and receive an estimated selling price, making the model accessible and user-friendly**.**

**7. Conclusion**

The project successfully developed a linear regression model to predict car prices with reasonable accuracy. Further improvements can be made by experimenting with more advanced machine learning algorithms and fine-tuning the model.