# Linear Regression

The objective of this lab assignment is to perform linear regression on the given "Car Price Prediction" dataset. In this lab, participants will explore the dataset, preprocess the data, split it into training and testing sets, build a linear regression model, train the model, make predictions, and evaluate its performance.
Dataset: You have been provided with the "car_price_data.csv" dataset. This dataset contains various features of cars, such as mileage, horsepower, number of cylinders, and other specifications, along with their corresponding prices.

**Tasks:**

**1. Data Preparation:**
- Import the necessary libraries (e.g., pandas, numpy, matplotlib, and scikit-learn).
- Load the "car_price_data.csv" dataset into a pandas DataFrame.
- Explore the dataset: check the dimensions, data types, summary statistics, and the presence of any missing values.

**2. Data Preprocessing:**
- Handle missing values, if any, using appropriate techniques (e.g., mean/median imputation).
- Perform feature scaling (e.g., Min-Max scaling) to standardize the numerical features.

**3. Handling Categorical Features:**
- Identify the categorical features in the dataset. These are the features that represent non-numeric data, such as car make, model, or fuel type.
- Hint: You can use the `pd.get_dummies()` function from pandas to perform one-hot encoding on the categorical features. This function will create binary vectors for each category within a categorical feature.

**4. Concatenate Transformed Data:**
- After one-hot encoding the categorical features, you will have additional binary columns in the DataFrame.
- Hint: Use the `pd.concat()` function to concatenate the transformed data with the original DataFrame.

**5. Data Splitting:**
- Split the dataset into training and testing sets using an 80-20 or 70-30 ratio.

- Separate the target variable (car prices) from the rest of the features in both the training and testing sets.

**6. Model Building:**

- Create an instance of the linear regression model from scikit-learn.
- Train the model using the training data.

**7. Model Evaluation:**

- Make predictions on the testing data using the trained model.
- Evaluate the model's performance using appropriate metrics (e.g., Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared score).
- Interpret the results and discuss the model's performance.

**8. Predictions:**

- Use the trained model to predict the prices of cars with new feature values. For example, predict the price of a car with specific mileage, horsepower, fuel type, etc.

**Conclusion:**

Summarize your findings from the lab, including insights gained from the data exploration, preprocessing, model training, and evaluation. Discuss the impact of one-hot encoding on the model's performance and how it handles categorical features effectively. Also, mention any challenges faced during the preprocessing or modeling stages and potential ways to overcome them.

**Note to students:**

Ensure to include comments in your code for better understanding and readability. Experiment with different settings, feature engineering techniques, or test/train split ratio to gain a better understanding of linear regression.