Goal: The ultimate goal of this training process is to create a model that can accurately predict the amount of CO2 emissions produced by a vehicle based on its engine size.

1.Importing Libraries:
The code starts by importing necessary libraries like `matplotlib.pyplot` (for visualization), `pandas` (for data manipulation), `numpy` (for numerical operations), and the `linear_model` module from `sklearn` (for linear regression).

2.Loading Data:
The code loads data from a CSV file named "FuelConsumption.csv" using `pd.read_csv()` into a pandas DataFrame named `df`.

3.Exploring Data:
The `head()` function is used to display the first few rows of the dataset to get an initial overview of its structure.
The `describe()` function provides summary statistics about the dataset, such as mean, median, quartiles, etc.

4.Selecting Data Columns:
A subset of the DataFrame containing specific columns (`ENGINESIZE`, `CYLINDERS`, `FUELCONSUMPTION_COMB`, `CO2EMISSIONS`) is extracted and stored in a new DataFrame called `cdf`.

Visualizing Data:

- The `hist()` function is used to create histograms for the selected columns, providing insights into the distribution of data.
- The `scatter()` function is used to create scatter plots of various feature combinations against CO2 emissions.

5.Data Splitting:
A random mask is generated using `np.random.rand(len(df)) < 0.8` to split the data into a training set (`train`) and a testing set (`test`).

6.Linear Regression:

- An instance of the `LinearRegression()` class is created using `linear_model.LinearRegression()`.
- The training feature data (`train_x`, engine sizes) and corresponding target data (`train_y`, CO2 emissions) are prepared.
- The `fit()` function is called to train the linear regression model using the provided training data.

7.Visualization of Linear Regression Line:

- The coefficients and intercept of the linear regression line are printed.
- A scatter plot of training data is created, and the linear regression line is plotted on top of it.

8.Model Evaluation:

- The `test_x` array is prepared using the engine sizes from the testing set.
- Predictions (`test_y_`) are made using the trained model on the testing feature data.
- Various evaluation metrics are calculated, including mean absolute error, mean squared error (MSE), and the R-squared score.

9.Printing Evaluation Metrics:

The calculated evaluation metrics (mean absolute error, MSE, and R-squared score) are printed to assess the performance of the linear regression model.

In summary, this code example demonstrates the process of loading, exploring, visualizing, splitting, training, and evaluating a linear regression model using the `scikit-learn` library to predict CO2 emissions based on engine size and other features. The code illustrates key steps in the machine learning pipeline, from data preprocessing to model evaluation.

Project Picture- Histogram of “ENGINESIZE”,”CO2 EMISSIONS”, CYLINDERS,

“FUELCONSUMPTION_COMB”

Histogram of 'CYLINDERS':

This histogram represents the distribution of the 'CYLINDERS' column from the dataset. It shows how many vehicles fall into different ranges of cylinder counts. The x-axis represents the number of cylinders, and the y-axis represents the frequency (or count) of vehicles having that number of cylinders.

Histogram of 'ENGINESIZE':

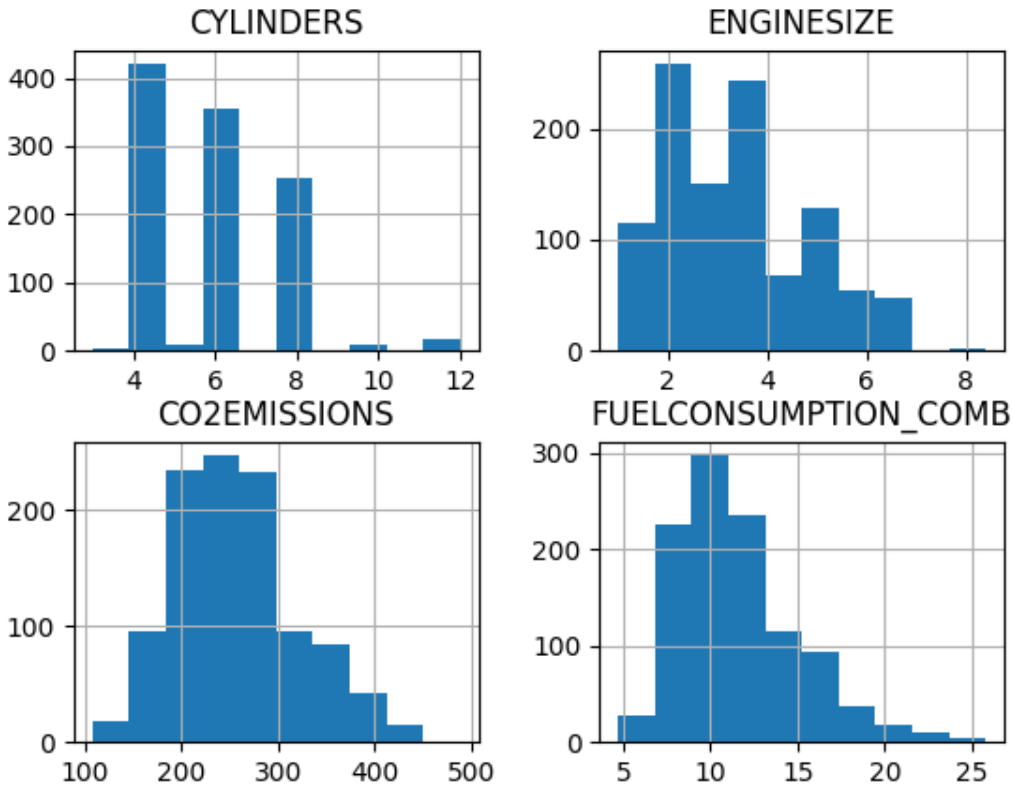
Similar to the first histogram, this one represents the distribution of the 'ENGINESIZE' column. It shows how many vehicles have different engine sizes. The x-axis represents engine sizes, and the y-axis represents the frequency of vehicles with those sizes.

Histogram of 'CO2EMISSIONS':

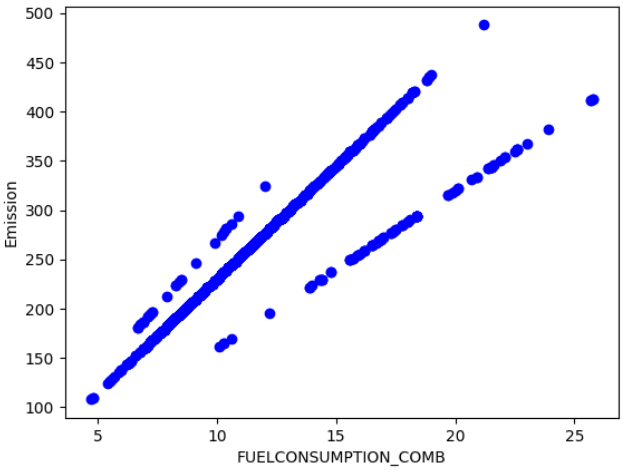
This histogram represents the distribution of CO2 emissions produced by vehicles. It provides insights into how frequently different levels of CO2 emissions occur. The x-axis represents emission levels, and the y-axis represents the frequency of vehicles emitting that amount of CO2.

Histogram of 'FUELCONSUMPTION_COMB':

This histogram represents the distribution of fuel consumption for combined driving (city and highway). It shows how many vehicles have different levels of fuel consumption. The x-axis represents fuel consumption values, and the y-axis represents the frequency of vehicles having that fuel consumption level.

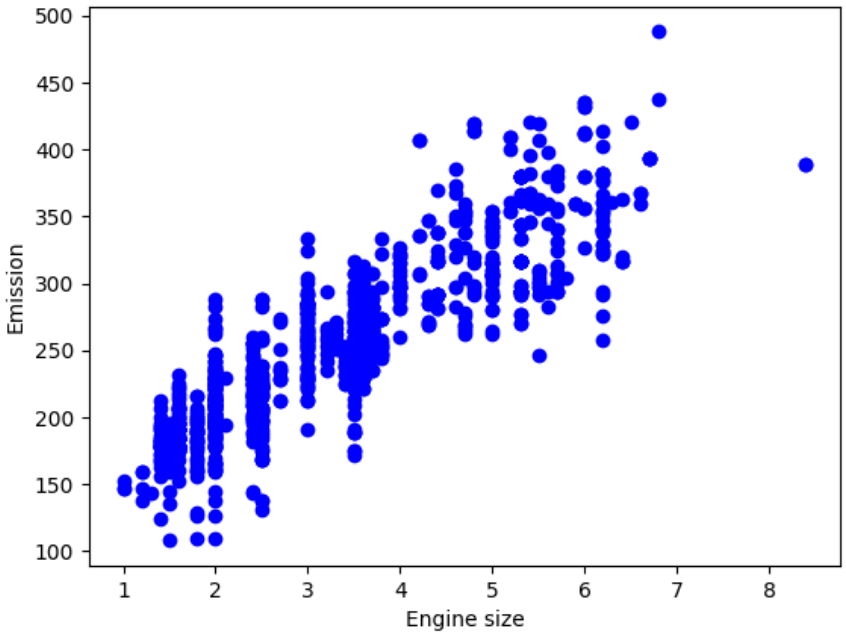


2. FUELCONSUMPTION_COMB VS EMISSION



- X-Axis ('FUELCONSUMPTION_COMB'): This axis represents the combined fuel consumption for vehicles. Each point on the x-axis corresponds to a specific value of fuel consumption.
- Y-Axis ('CO2EMISSIONS'): This axis represents the amount of CO2 emissions produced by the vehicles. Each point on the y-axis corresponds to a specific value of CO2 emissions.
- Showing the relationship between fuel consumption increases, CO2 emissions tend to increase as well.

3.ENGINE SIZE VS EMISSION



- X-Axis ('ENGINESIZE'): This axis represents the size of the vehicle's engine. Each point on the x-axis corresponds to a specific engine size value.
- Y-Axis ('CO2EMISSIONS'): This axis represents the amount of CO2 emissions produced by the vehicles. Each point on the y-axis corresponds to a specific value of CO2 emissions