**Exploratory Data Analysis (EDA) of Automobile Dataset**

**Details of the dataset:**

The dataset contains the following features:

1. mpg: Miles per gallon (target variable).
2. cylinders: Number of cylinders in the vehicle.
3. displacement: Engine displacement (in cubic inches).
4. horsepower: Horsepower of the vehicle.
5. weight: Weight of the vehicle (in pounds).
6. acceleration: Time taken to accelerate from 0 to 60 mph (in seconds).
7. model year: Model year of the car (numeric).
8. origin: Origin of the vehicle (1 = USA, 2 = Europe, 3 = Japan).
9. name: The name of the car.

**Task 1: Load and Inspect the Data (15-20 minutes)**

* Load the dataset and check the first few rows using head().
* Check for any missing values and decide how to handle them (isnull() and dropna() or imputation).
* Inspect the data types and convert columns as needed (dtype).

**Task 2: Data Cleaning and Preprocessing (20-25 minutes)**

* **Handle Missing Values**:
  + Check for missing values in columns like horsepower and mpg. Discuss strategies for handling them (e.g., fill missing values with median, drop rows, etc.).
* **Feature Engineering**:
  + Convert the origin column into a categorical column and assign labels (USA = 'USA', Europe = 'Europe', Japan = 'Japan').
  + Create a new column: weight\_per\_cylinder = weight / cylinders. This can provide an insight into the relationship between weight and the number of cylinders.
  + Create a new column: mpg\_per\_horsepower = mpg / horsepower, which could give a ratio of fuel efficiency relative to engine power.

**Task 3: Univariate and Bivariate Exploration (30-40 minutes)**

* **Univariate Analysis**:
  + Plot the distribution of numerical features like mpg, horsepower, and weight (e.g., using histograms and boxplots).
  + Check the distribution of categorical features like cylinders and origin (e.g., using bar plots).
* **Bivariate Analysis**:
  + Explore relationships between the target variable (mpg) and other numerical variables:
    - Plot mpg vs horsepower, mpg vs weight, and mpg vs displacement (scatter plots, boxplots, or violin plots).
  + Explore categorical variables:
    - Compare mpg across different origin or cylinders using bar plots.
* **Correlation Analysis**:
  + Use a heatmap to visualize the correlation matrix between numerical variables (mpg, horsepower, weight, etc.).
  + Identify which features are most strongly correlated with mpg.

**Task 4: Advanced Data Aggregation and Grouping (30 minutes)**

* **Group the Data**:
  + Calculate the average mpg by origin and cylinders. Which countries or cylinder groups have the most fuel-efficient cars?
  + Explore the relationship between mpg and model year to see how fuel efficiency has changed over the years.
* **Aggregate with Multiple Conditions**:
  + Group by origin and cylinders to calculate the average horsepower, weight, and mpg.
  + Compare the relationship between horsepower and weight across different origins (e.g., USA vs Japan vs Europe).

**Task 5: Data Visualization (15-20 minutes)**

* **Advanced Visualizations**:
  + Create **pair plots** or **scatter matrix plots** to explore relationships between multiple numerical variables.
  + Use **seaborn's facet grids** to create multiple visualizations for comparing mpg across different car origins or cylinder counts.
  + Visualize the relationship between multiple features using a **3D scatter plot** for features like mpg, horsepower, and weight using matplotlib or plotly (challenge: figure out how to create a 3D scatter plot by referring to the documentation)