Week 8 Assignment – AutoMPG Data Analysis

Overview

In this assignment, you enhance your code from Week 7 to include more analysis and cleaning of the data in the Auto-MPG data set. We will also improve the data output.

Please follow the steps outlined below.

Preparation

1. Copy the data files from the Week 7 assignment.
2. Copy **autompg2.py** from the Week 7 assignment and rename it **autompg3.py**.

Step 1: Clean make data

In the data file, there are a number of automobile makes that are typos or otherwise duplicates. This is common with real data. Enhance your data parsing code to fix up the make values as the data are read in so that the following corrections are made:

Incorrect	Correct
chevroelt	chevrolet
chevy	chevrolet
maxda	mazda
mercedes-benz	mercedes
toyouta	toyota
vokswagen	volkswagen
vw	volkswagen

Step 2: Add analysis to AutoMPGData class

Enhance the AutoMPGData class by adding the following methods:

- **mpg_by_year** – This method should return a dictionary where the keys are the years that are present in the data set and the values are the average MPG for all cars in that year. A **defaultdict** is a good match for this functionality.
- **mpg_by_make** – This method should return a dictionary where the keys are the makes that are present in the data and the values are the average MPG for all cars of that make.

Step 3: Enhance command-line parsing

In **autompg3.py** add support for the following command line arguments:

- -o | –ofile <outfile> – This option allows the user to specify the name of a file to which output should be written. If this is not specified, output should be sent to **sys.stdout**.
- -p | –plot – This option allows the user to specify that graphical output using **matplotlib** should be created.
- There should be two additional command options implemented: **mpg_by_year** and **mpg_by_make**.

Please implement these options and adjust the **main** function as follows:

- The data sorting options (implemented in Week 7) should only take effect if the "print" command is being executed.
- The "print" command should honor the –ofile option and should also output the data in CSV format using the **csv** module with one column per attribute.
- The new commands **mpg_by_year** and **mpg_by_make** should execute the corresponding function on the AutoMPGData object and should format the output so that the data are shown sorted by key. The output should be in CSV format and should honor the –ofile option.
- If the –plot option is used, the **mpg_by_year** and **mpg_by_make** commands should also produce a simple matplotlib output.

Upload

Please put **autompg3.py** into a ZIP file.