**Case study – Clustering of cars into mini, prime, and sedan**

**Introduction**

This case study demonstrates the application of cluster analysis using python program to categorize car models into three distinct classes: mini, prime, and sedan. The analysis utilizes the mtcars dataset to segment 32 car models based on their performance and specification attributes. Key variables such as mpg, hp, am, and gear are used to derive meaningful groupings, providing insights into the characteristics of each car category.

**Objective**

To perform cluster analysis to divide car models into three classes: mini, prime, and sedan.

**Dataset link**

<https://drive.google.com/file/d/1_GLs7mStA8pMfIsern3vHCDfXib3SG2w/view?usp=sharing>

**Understanding the data**

The dataset contains information about 32 different car models and their various specifications. Here's a breakdown of the 12 columns:

Car Details

name (object) - Name of the car model (e.g., "Mazda RX4", "Datsun 710")

Performance & Specs

* mpg (Miles per Gallon) - Fuel efficiency of the car
* cyl – Number of Cylinders (4, 6, or 8 cylinders)
* disp – Engine displacement, indicating engine size
* hp – Engine horsepower
* drat – Gear ratio in the rear axle
* wt – Weight (1000 lbs)
* qsec – 1/4 Mile Time

Engine & Transmission

* vs – Type (0 = Fuel, 1 = EV)
* am – Transmission (0 = Automatic, 1 = Manual)
* gear – Number of Forward Gears (usually 3, 4, or 5)
* carb – Number of Carburetors

**Procedure for coding**

* Import necessary libraries
* Load dataset
* Check the number of rows and columns
* Check for missing values
* Define independent (X) and dependent (y) variables
* Standardize independent variables
* Compute distance matrices (euclidean, cityblock, minkowski having p value 3 & 4) for each pair
* Sort distances in ascending order
* Plot dendrogram
* Assign clusters
* Label class as Mini, Prime, Sedan
* Display the number of car models in each class
* Create separate data frames for each class

**Code File Link**

<https://github.com/Ishita2003M/Clustering-of-cars-into-mini-prime-and-sedan/blob/main/car_clust.ipynb>

**Interpretation and conclusion**

1. The goal of the analysis was to segment car models into meaningful categories based on performance and configuration attributes using hierarchical clustering.
2. The analysis was conducted on the classic mtcars dataset having 32 rows and 12 columns, using four key features:
   * mpg (miles per gallon)
   * hp (horsepower)
   * am (transmission type)
   * gear (number of gears)  
     These variables were selected to reflect a balance between fuel efficiency, engine power, and mechanical configuration.
3. All features were standardized to ensure comparability, given their varying scales and units.
4. Multiple distance metrics—Euclidean, Cityblock (Manhattan), and Minkowski with p=3 and p=4—were computed to assess similarity between car models. To enhance robustness, the minimum distance across all metrics for each pair was selected to form a combined distance matrix.
5. Hierarchical clustering was performed using Ward’s linkage method, which minimizes variance within clusters. A dendrogram was plotted to visualize the clustering structure.
6. The model was configured to form three distinct clusters, each reflecting a unique segment of cars based on the selected features.
7. The clusters were labeled to represent intuitive vehicle categories:
   * Mini: Compact, likely fuel-efficient models
   * Prime: Mid-range, balanced-performance models
   * Sedan: Larger or more powerful vehicles
8. The car models were distributed across the clusters as follows:
   * Prime: 12 models
   * Sedan: 12 models
   * Mini: 8 models
9. The relatively even distribution between Prime and Sedan suggests a balanced dataset with a slight skew towards mid- and high-performance vehicles. The Mini segment, though smaller, highlights a distinct cluster of more compact, likely economy-oriented cars.