Assignment 5:

Name: Rohit Bhabire

Roll No: 281031 Batch: A2

Problem Statement

Apply clustering algorithms on the Mall Customers dataset to identify groups of profitable customers. Use *Spending Score* as a primary feature to segment customer types based on their shopping behavior.

Objectives

1. Perform data preprocessing and transformation on the dataset.
2. Apply unsupervised learning methods to identify customer segments.
3. Use two different clustering algorithms (e.g., K-Means and Hierarchical Clustering).
4. Visualize the resulting clusters for interpretability.

Resources Used

- **Software:** Google Colab
- **Libraries:** pandas, matplotlib, seaborn, sklearn, scipy

Theory

Clustering is an **unsupervised learning** technique used to group similar data points together. Unlike supervised learning, clustering does not use labeled data.

K-Means Clustering

- Partitions data into *K clusters* based on feature similarity.
- Minimizes the sum of squared distances between points and their respective cluster centroids.
- Requires predefining the number of clusters (K).

Hierarchical Clustering

- Builds a tree-like structure of nested clusters.
- Uses a dendrogram to decide the optimal number of clusters.
- No need to predefine the number of clusters.

Methodology

1. Data Preprocessing

- Loaded the dataset using Pandas.
- Checked for null or missing values.
- Selected relevant features: *Annual Income* and *Spending Score*.
- Scaled features using Standard Scaler to ensure uniformity.

2. Applying Clustering Algorithms

A. K-Means Clustering

- Used KMeans from sklearn.cluster.
- Applied the **Elbow Method** to find the optimal number of clusters.
- Visualized clusters using scatter plots.

B. Hierarchical Clustering

- Used scipy.cluster.hierarchy for dendrogram plotting.
- Chose the optimal cluster number based on dendrogram cut.
- Applied AgglomerativeClustering from sklearn.cluster.

3. Visualization

- Plotted the results of clustering using matplotlib and seaborn.
- Color-coded clusters to represent distinct customer groups.

Conclusion

- **K-Means** and **Hierarchical Clustering** successfully grouped mall customers into meaningful segments.
- These clusters help identify high-spending customers who are likely to be more profitable.
- Clustering insights can guide marketing and personalized offers to targeted segments, enhancing business strategy.