**Clustering**

**Mall customers clustering**

You own a supermarket mall and through membership cards , you have some basic data about your customers like Customer ID, age, gender, annual income and spending score. Spending Score is something you assign to the customer based on your defined parameters like customer behavior and purchasing data.

**Problem:**

You own the mall and want to understand the customers like who can be easily converge [Target Customers] so that the sense can be given to marketing team and plan the strategy accordingly.

**Dataset:**

<https://drive.google.com/file/d/15M9juMcLRJ8-_ogruSWgfwdo3Ywqg7SO/view?usp=share_link>

**Dataset Description**

The dataset contains the following attributes:

1. **Customer ID**: Unique identifier (not used for clustering).
2. **Age**: Age of the customer (in years).
3. **Gender**: Male/Female.
4. **Annual Income (k$)**: Customer’s annual income (in $1000).
5. **Spending Score (1-100)**: Score based on customer behavior and purchasing patterns.

**Steps to Follow**

**1. Data Loading and Exploration**

* Load the dataset using Python.
* Inspect the dataset for structure, missing values, and descriptive statistics.

**2. Data Preprocessing**

* Encode categorical attributes (e.g., **Gender**) using label encoding or one-hot encoding.
* Normalize numerical attributes (e.g., **Age**, **Annual Income**, **Spending Score**) using Min-Max scaling or StandardScaler.

**3. Apply Clustering Techniques**

1. **KMeans Clustering**
   * Use the Elbow Method to find the optimal number of clusters (plot WCSS vs. clusters).
   * Fit the model and assign cluster labels.
2. **Hierarchical Clustering**
   * Use Agglomerative Clustering with different linkage methods (e.g., ward, complete, average).
   * Plot dendrograms to determine the optimal number of clusters.
3. **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**
   * Tune parameters like eps and min\_samples to identify clusters.

**4. Cluster Evaluation**

* Calculate and compare clustering quality using:
  + **Davies-Bouldin Index (DBI)**: Lower DBI indicates better clustering.
  + **Dunn Index**: Higher Dunn Index indicates better clustering.

**5. Visualization**

* Plot clusters for each technique:
  + Use 2D scatter plots (e.g., **Annual Income** vs. **Spending Score**) to visualize clusters.
  + Include noise points for DBSCAN.

**6. Compare and Analyze Results**

* Compare the number of clusters and their characteristics across techniques.
* Discuss which technique performs better based on DBI, Dunn Index, and interpretability.

**7. Conclusion**

* Summarize the strengths and weaknesses of each clustering technique.
* Discuss the implications of customer segmentation in real-world business scenarios.