# Assignment -10

**Problem 1: Customer Segmentation using K-Means**

**Objective**: Segment mall customers into distinct groups based on their shopping behavior.

**Instructions**:

1. Load the dataset and handle any missing or irrelevant values.
2. Select relevant features for clustering (e.g., Annual Income (k$) and Spending Score (1-100)).
3. Normalize or scale the data appropriately.
4. Use the **Elbow Method** to determine the optimal number of clusters.
5. Apply **K-Means clustering** and visualize the clusters.
6. Label each customer with their cluster group and analyze the characteristics of each cluster.

**Questions**:

- How many clusters does the Elbow Method suggest?
- Describe each cluster's characteristics in terms of spending habits.
- What insights can mall management derive from this segmentation?

**Problem 2: Density-Based Clustering with DBSCAN**

**Objective**: Discover customer clusters of varying density using DBSCAN.

**Instructions**:

1. Use the same features as in Assignment 1.
2. Normalize the data.
3. Use **DBSCAN** to perform clustering. Experiment with different eps and min_samples values.
4. Visualize the clusters and identify outliers.
5. Compare DBSCAN results with K-Means.

**Questions**:

- What are the advantages of DBSCAN in this context?
- How many core points and outliers were detected?
- Does DBSCAN reveal any patterns that K-Means missed?

**Problem 3: Gaussian Mixture Models (GMM) for Soft Clustering**

**Objective**: Use GMM to probabilistically assign customers to clusters.

**Instructions**:

1. Reuse the preprocessed data from Assignment 1.
2. Fit a **Gaussian Mixture Model** to the dataset.
3. Determine the optimal number of components using **BIC/AIC**.
4. Visualize the probability-based clusters.
5. Compare GMM results with K-Means and DBSCAN.

**Questions**:

- What number of clusters is suggested by the BIC score?
- How does soft clustering provide more flexibility than hard clustering?
- Are any customers assigned significant probabilities to more than one cluster?

**Problem 4 : Dimensionality Reduction + Clustering with PCA**

**Objective**: Reduce feature dimensionality using PCA and apply clustering.

**Instructions**:

1. Select at least 4 features (e.g., Age, Gender (as numeric), Income, Spending Score).
2. Normalize and apply **PCA** to reduce to 2 principal components.
3. Visualize the explained variance.
4. Apply **K-Means** or **GMM** on the PCA-reduced data.
5. Compare results with the original feature space clustering.

**Questions**:

- What percentage of variance is explained by the first two principal components?
- Does clustering on PCA-transformed data give different results?
- What are the trade-offs when using PCA before clustering?

---xxx----