Name:- Abhishek Bayas

Roll NO :- 281020

PRN:- 22310349

**Assignment-05**

**Assignment 5: Advanced Breakdown**

**Objective**

Use clustering algorithms to segment customers based on spending habits and income. This helps mall owners target the right customers for **offers, loyalty programs, or marketing campaigns**.

**Dataset Overview**

**Features available:**

* CustomerID: Unique identifier
* Gender: Categorical (Male/Female)
* Age: Customer's age
* Annual Income (k$): Annual income in $1000s
* Spending Score (1-100): Score assigned by the mall based on customer behavior and spending nature

**Selected Features for Clustering**

We focus on:

* Annual Income (k$)
* Spending Score (1-100)

These are the **most relevant features** to determine customer spending behavior and affordability.

**Detailed Methodology**

**Step 1: Data Preprocessing**

* **Missing values:** Checked using isnull().sum() (usually clean in this dataset).
* **Gender Encoding:** Converted to 0 and 1 using LabelEncoder.
* **Feature Scaling:** Done using StandardScaler to normalize income and score range.

**Step 2: Data Preparation**

* Since clustering is **unsupervised**, train-test splitting is optional. But to follow best practices, we split into X\_train and X\_test (80-20) for evaluation.

**Step 3: KMeans Clustering**

**How KMeans works:**

* Assigns k cluster centers randomly.
* Iteratively assigns each point to the nearest cluster and updates centroids.
* Objective: Minimize **WCSS** (Within-Cluster Sum of Squares).

**Elbow Method:**

* Run for k=1 to 10 and plot WCSS.
* The **"elbow point"** (where WCSS starts decreasing slowly) gives optimal k. Usually, it's **5** for this dataset.

**Cluster Visualization:**

You’ll see clusters like:

* **Cluster 1:** Low income, low spending → Likely uninterested
* **Cluster 2:** High income, high spending → **Most Profitable**
* **Cluster 3:** Low income, high spending → **Impulse buyers**
* **Cluster 4:** High income, low spending → **Need engagement**
* **Cluster 5:** Average group

**Step 4: Hierarchical Clustering**

**How It Works:**

* Uses **agglomerative clustering** (bottom-up).
* Each point is a cluster → merges the closest ones using **Ward’s method**.
* Builds a **dendrogram** to visualize cluster formation.

**Hierarchical Clustering Insight:**

CODE- From Visual Studio Code.

**Business Insight from Clustering**

| **Cluster Type** | **Action** |
| --- | --- |
| High Income + High Spend | Loyalty programs, premium services |
| Low Income + High Spend | Offer budget-friendly packs |
| High Income + Low Spend | Promotions to encourage spending |
| Low Income + Low Spend | Exclude from intensive campaigns |

**Summary**

* **Both algorithms give consistent clusters**, but KMeans is more scalable.
* Helps understand customer base better and boost **targeted marketing**.
* Visualizations like **scatter plots, centroids, dendrograms** make interpretation easy.