**Name: Muhammad Abdullah**

**STD ID: STU-DS-251-930**

**Customer Segmentation Using Clustering Algorithms**

**Project Overview**

I have done this project by using dataset of customer segmentation and also use preprocessing techniques and method and also apply PCA for visualization and work on the models like K-means, DB-Scan, Hierarchical Clustering and also doing Model evaluations for these algorithms

**A. Data Preparation & Preprocessing**

**1. Loading and Exploring the Dataset**

I loaded the dataset using pandas and explored its structure, checking for data type s, column names, and summary statistics. This helped me understand the distribution of data and identify any inconsistencies.

**2. Handling Missing or Duplicate Values**

I checked for missing values using isnull().sum() and for duplicates using duplicated().sum(). There were no missing or duplicate values, so no rows were removed.

**3. Encoding Categorical Features**

The dataset included a categorical feature Gender. I applied label encoding to convert it into numerical form using LabelEncoder from sklearn.

**4. Feature Scaling**

I used StandardScaler to normalize the features so that each has a mean of 0 and standard deviation of 1. This ensured that all features contributed equally to distance calculations in clustering.

**B. Dimensionality Reduction**

**1. Principal Component Analysis (PCA)**

I applied PCA to reduce the dataset to 2 dimensions for easier visualization. I analyzed the explained variance ratios to ensure most of the variance was retained in the first two components.

**2. Visualization of PCA**

I visualized the PCA-transformed data using a scatter plot, which helped me understand the structure and distribution of the data points.

**C. Clustering Implementation**

**1. K-Means Clustering**

I applied the K-Means algorithm with different values of K. I used the Elbow Method by plotting inertia scores to find the optimal number of clusters. Then I applied K-Means with the optimal K and visualized the clusters.

**2. DBSCAN**

I implemented DBSCAN for density-based clustering. I tuned the eps and min\_samples parameters by analyzing the neighborhood distances. DBSCAN helped in identifying noise and handling non-globular clusters.

**3. Hierarchical Clustering**

I used Agglomerative Hierarchical Clustering and visualized the dendrogram to decide the number of clusters. After identifying the cluster count, I visualized the final clusters.

**D. Evaluation**

I compared the results of all three clustering methods. K-Means performed best with well-separated and compact clusters. DBSCAN effectively detected outliers and irregular clusters. Hierarchical Clustering offered a clear visual understanding of the clustering hierarchy.

---------------------------------------------------------------------------------------------------------------------

**Conclusion**

Through this project, I successfully segmented mall customers based on key behavioral features. I applied multiple clustering techniques and evaluated their performance using industry-standard metrics. The insights gained from this segmentation can be utilized by businesses to personalize marketing strategies, improve customer engagement, and increase overall profitability.