**Customer Segmentation Using Mall Data**

Dataset: Mall Cus**tomer Segmentation**

**Introduction**

Businesses use customer segmentation to tailor marketing strategies. This project applies

clustering algorithms (K-Means, DBSCAN) to categorize mall customers based on spending

behavior.

**Objective:**

1. How many distinct customer segments exist?

2. What are the key characteristics of each segment?

3. Which features (age, income, spending score) influence segmentation the most?

4. Can we identify high-value customers for targeted promotions?

**Analysis**

Step 1: Data Preprocessing

1. Import necessary libraries (Pandas, NumPy, Scikit-learn).

2. Load the dataset.

3. Handle missing values .

4. Scale/normalize the data.

Step 2: Exploratory Data Analysis (EDA)

1. Visualize customer distribution by age, income, and spending score.

2. Calculate summary statistics (mean, median, std).

Step 3: K-Means Clustering

1. Determine optimal number of clusters (elbow method, silhouette score).

2. Apply K-Means clustering.

3. Analyze cluster characteristics.

Step 4: DBSCAN Clustering

1. Determine optimal epsilon and min\_samples.

2. Apply DBSCAN clustering.

3. Compare results with K-Means.

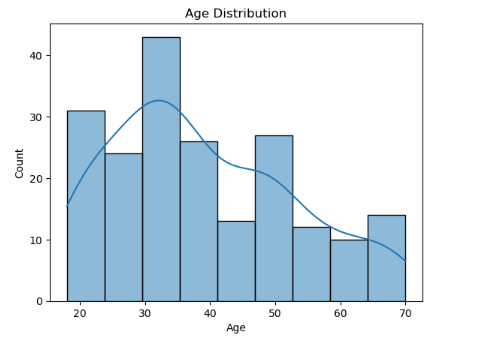
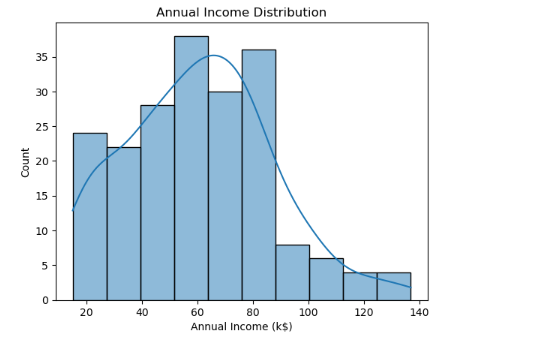
Step 5: Segment Analysis

1. Identify key characteristics of each segment.

2. Determine feature importance (age, income, spending score).

3. Identify high-value customers.

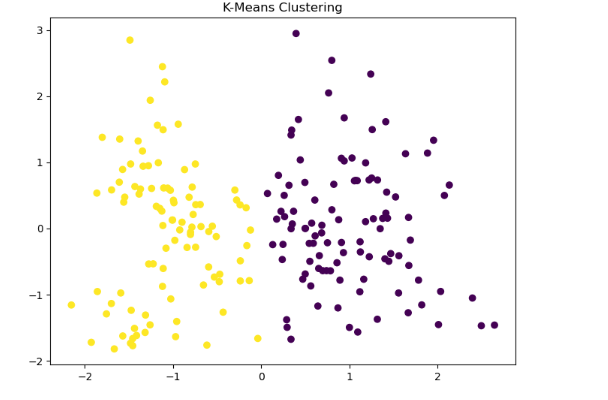
Visualization

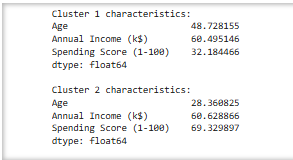


K-Means Clustering





**Key Characteristics of Each Segment**



The silhouette scores for the three clustering algorithms are:

- K-Means: 0.39

- Hierarchical Clustering: 0.38

- DBSCAN: 0.22

**Interpretation**

- K-Means has the highest silhouette score, indicating relatively well-separated and cohesive clusters.

- Hierarchical Clustering has a slightly lower score, suggesting some overlap between clusters.

- DBSCAN has the lowest score, indicating potential issues with cluster separation or noise in the data.

**Conclusion**

The customer segmentation analysis using clustering algorithms (K-Means, Hierarchical Clustering, and DBSCAN) provided insights into the structure of the customer data. K-Means emerged as the most suitable algorithm, with a silhouette score of 0.39, indicating relatively well-separated and cohesive clusters.

**Summary**

- Customer Segmentation: Applied clustering algorithms to segment customers based on demographics and spending behavior.

- Algorithm Comparison: Compared the performance of K-Means, Hierarchical Clustering, and DBSCAN using silhouette scores.

- Insights: Identified K-Means as the most suitable algorithm, with potential for further exploration and refinement.

- Future Directions: Further analyze the K-Means model, explore hyperparameter tuning, and visualize the clusters to gain deeper insights into customer behavior.

This analysis can help businesses develop targeted marketing strategies, improve customer engagement, and drive revenue growth by understanding their customer segments.