

Customer Segmentation Using Clustering – Project Report

1. Introduction

Customer segmentation is a crucial strategy for businesses to understand their customer base and design targeted marketing campaigns. By grouping customers with similar characteristics, companies can improve customer satisfaction, loyalty, and revenue.

This project applies **unsupervised learning (clustering)** to the **Mall Customer Segmentation Dataset** to identify meaningful customer groups based on demographics and spending behavior.

Objectives:

- Explore and analyze customer demographics and spending data.
 - Perform clustering using **K-Means** and/or **Hierarchical Clustering**.
 - Visualize customer groups.
 - Derive actionable business insights for personalized marketing.
-

2. Dataset

The dataset used is the **Mall Customer Segmentation Data** from Kaggle.

Columns:

- `CustomerID` – Unique ID for each customer
- `Gender` – Male/Female
- `Age` – Customer age
- `Annual Income (k$)` – Customer income (in \$1000s)
- `Spending Score (1-100)` – Score assigned by the mall based on spending behavior

Dataset Overview:

- Number of Customers: 200
- Gender Distribution: ~56% Female, ~44% Male
- Age Range: 18 – 70 years
- Annual Income Range: 15k – 137k
- Spending Score Range: 1 – 100

3. Methodology

3.1 Exploratory Data Analysis (EDA)

- Checked data types, missing values, and distributions.
- Plotted histograms and boxplots for **Age**, **Income**, and **Spending Score**.
- Observed clusters visually with **scatter plots** (e.g., Income vs. Spending Score).

3.2 Feature Scaling

- StandardScaler/MinMaxScaler applied to normalize **Age**, **Income**, **Spending Score** for clustering.

3.3 Clustering Models

1. **K-Means Clustering**
 - Applied **Elbow Method** and **Silhouette Score** to choose optimal number of clusters (k).
 - Selected **k = 5 clusters**.
 2. **Hierarchical Clustering (Agglomerative)**
 - Used dendrograms to visualize merging of clusters.
 - Produced similar groupings as K-Means.
-

4. Results

4.1 K-Means Clusters

After applying K-Means with k=5, the customers were grouped as follows:

Cluster 1 – High Income, High Spending

- Premium customers, loyal and spend heavily.
- Ideal for **exclusive offers & VIP programs**.

Cluster 2 – High Income, Low Spending

- Affluent but conservative in spending.
- Require **personalized promotions** to increase engagement.

Cluster 3 – Average Income, Average Spending

- Middle-class customers, moderate shopping.
- Can be targeted with **seasonal discounts**.

Cluster 4 – Low Income, High Spending

- Budget-limited but spend disproportionately high.
- Likely **young customers** (trend-driven).
- Respond well to **discounted trendy products**.

Cluster 5 – Low Income, Low Spending

- Minimal engagement, unlikely to generate high revenue.
- Maintain engagement with **basic loyalty programs**.

4.2 Visualizations

- **Elbow Curve** showing optimal $k=5$.
- **Scatter Plots** (Annual Income vs. Spending Score) with clusters color-coded.
- **Dendrogram** for hierarchical clustering.

5. Business Insights

1. **Target High-Value Customers (Cluster 1):**
Offer premium memberships, early access to new collections, and exclusive services.
2. **Upsell to Wealthy Low Spenders (Cluster 2):**
Use personalized recommendations, luxury promotions, and one-on-one consulting.
3. **Engage Average Spenders (Cluster 3):**
Introduce seasonal promotions and loyalty rewards.
4. **Attract Young Trend-Seekers (Cluster 4):**
Focus on trendy, affordable products, and use social media campaigns.
5. **Low Priority Customers (Cluster 5):**
Keep minimal engagement, but not a key revenue source.

6. Conclusion

This project demonstrates the power of **unsupervised learning** for customer segmentation. Using **K-Means and Hierarchical clustering**, we identified 5 distinct customer groups.

Key Outcomes:

- Segments provide actionable strategies for targeted marketing.

- High-income high-spenders are the most valuable segment.
- Proper personalization can convert high-income low-spenders into loyal customers.

Future Work:

- Incorporate more features (purchase history, visit frequency, online activity).
- Use advanced clustering methods like **DBSCAN or Gaussian Mixture Models**.
- Deploy results in a **real-time recommendation system**.