

Customer Segmentation using Clustering

Objective

The primary goal of this project was to build a **customer segmentation system** that helps businesses better understand and target their customers based on **demographics** (like age, gender, income) and **spending behavior**.

The dataset used was from Kaggle:

[Customer Segmentation Dataset](#)

Approach

1. Exploratory Data Analysis (EDA)

- Loaded and cleaned the dataset.
- Checked for missing values and data types.
- Visualized key distributions: Age, Annual Income, Spending Score.
- Analyzed correlations between features.
- Applied **label encoding** to convert categorical features like **Gender** to numerical format.

2. Feature Selection

- Selected relevant features: **Age**, **Annual Income (k\$)**, and **Spending Score (1-100)**.

3. K-Means Clustering

- Used the **Elbow Method** to determine the optimal number of clusters (k).
- Applied **KMeans** clustering with the optimal **k** (typically 4 or 5).
- Visualized clusters using 2D scatter plots and pair plots.
- Each cluster was labeled and interpreted based on behavior (e.g., High income, low spending).

4. Hierarchical Clustering

- Applied **Agglomerative Clustering** using Euclidean distance and Ward linkage.
- Generated a **dendrogram** to visualize the merging of clusters.
- Compared results with K-Means to validate consistency.

Challenges Faced

- **Categorical Encoding:** The original dataset contained categorical fields like gender, which caused issues with correlation plots. This was resolved using label encoding.
 - **Choosing the right number of clusters:** Required using the Elbow Method and dendrograms to ensure meaningful grouping.
 - **Visualizing High-Dimensional Data:** Limited features helped, but in real-world scenarios, dimensionality reduction techniques like PCA may be needed.
-

Model Performance & Insights

While clustering is an unsupervised method (no accuracy/F1-score), we evaluated performance based on:

- **Silhouette Score** (optional): Measured the compactness of clusters.
- **Visual separation** of clusters.
- **Interpretability** of each customer segment.

Example Cluster Insights:

- **Cluster 1:** Young, low income, high spending — potential impulse buyers.
- **Cluster 2:** Older, high income, low spending — likely conservative spenders.
- **Cluster 3:** Middle-aged, moderate income, moderate spending — balanced customers.

These segments allow marketing teams to create personalized strategies for each group