Mall Customers Segmentation through Clustering

Overview

In this project, student teams will analyze mall customer records to gain insights into shopper behavior, spending patterns, and segment differences. The dataset includes basic demographic and spending attributes for individual customers. This project provides an opportunity to explore patterns and relationships in retail data, such as how income and age relate to spending behavior, how distinct customer groups emerge, and how those groups can guide marketing or service strategies. By applying data analysis, clustering, and visualization techniques, students can uncover actionable insights that deepen their understanding of both unsupervised learning and customer analytics.

Project Requirements

Datasets

• Mall Customers:

Primary source from https://www.kaggle.com/datasets/shwetabh123/mall-customers

• Description:

This is a tabular dataset (≈200 rows) with five fields—CustomerID, Gender, Age, Annual Income ($), and Spending Score (1–100). Its clean schema make it suitable for clustering, internal evaluation, and 2D visualization

Tasks

A. Data Acquisition, Understanding and Processing

a) Download the dataset and document the schema (fields, types, size).
b) Perform basic cleaning (handle missing values/duplicates), scale numerical features appropriately and select effective features.

B. Clustering with Data Visualization:

Clustering on datasets reveals latent structure in customer behavior, visualizations help validate separability and communicate segment differences.

a) Implement at least two clustering algorithms and compare results..
b) Provide a 2D visualization (e.g., PCA projection) to inspect potential group structures.
c) Evaluate how clustering hyperparameters affect results (e.g., K for K-Means).
d) Choose one clustering method and assign human-readable labels to each cluster (brief profile per segment).

C. Modeling Using AI-Related Tools (Optional)

Apply AI-related tools or LLM to create useful applications and gain deeper insights.

a) Implement at least two clustering algorithms and compare results.

b) Use internal metrics for model selection and comparison.

c) Perform a brief stability check (e.g., different random seeds once).

d) Provide concise cluster interpretation (names/profiles) and 1–2 actionable suggestions.

Expected Learning Outcome

1. Technical Skills:

a) Proficiency in Python programming and data manipulation.

b) Ability to use Python libraries for data manipulation and visualization.

c) Experience in applying AI-related tools and models (if attempted).

2. Critical and Creative Thinking Skills:

a) Ability to critically analyze unstructured data and identify meaningful patterns.

b) Creativity in exploring data and interpreting insights.

3. Collaboration and Communication:

a) Strengthening of teamwork abilities.

b) Improved presentation skills by summarizing and sharing findings.