Clustering

This project applies unsupervised machine learning using K-Means Clustering to group customers based on their purchasing behaviour. By analyzing features such as Annual Income and Spending Score, we group customers into distinct clusters that represent different types of consumers — for example, high-income low-spending, low-income high-spending, young impulsive buyers, or conservative spenders.

Use the file: K-Means Clustering.ipynb to explore and run the project.

Dataset

The dataset (Mall_Customers.csv) contains the following columns:

- **CustomerID**: Unique ID assigned to each customer
- **Gender**: Gender of the customer
- **Age**: Age of the customer
- **Annual Income (k$)**: Annual income in thousands
- **Spending Score (1-100)**: Score assigned based on purchasing behaviour and spending nature

Notebook: K-Means Clustering.ipynb

This notebook demonstrates how to:

1. Load and preview the dataset
2. Import necessary libraries
3. Perform exploratory data analysis (EDA)
4. Visualize data distributions (e.g., Age vs Spending Score)
5. Apply K-Means Clustering to segment customers
6. Determine optimal number of clusters using the Elbow Method
7. Visualize the resulting clusters in 2D
8. Analyze customer segments for marketing strategies

Use Cases

- Targeted marketing strategies

- Product recommendations for different customer groups
- Loyalty program segmentation

References

1. https://**scikit-learn.org**/stable/modules/generated/sklearn.cluster.KMeans.html
2. https://www.geeksforgeeks.org/clustering-in-machine-learning/
3. Dataset: Mall Customers