Customer Segmentation using K-Means Clustering

1. Introduction

This project applies K-Means Clustering (Unsupervised Machine Learning) to segment mall customers based on Age, Annual Income, and Spending Score.

2. Problem Statement

Businesses need customer segmentation to improve marketing efficiency, increase revenue, and personalize offers without having labeled data.

3. Dataset

The dataset includes Age, Annual Income (k$), and Spending Score (1-100). No target variable is used since this is unsupervised learning.

4. Methodology

Steps performed: Data Cleaning, Exploratory Data Analysis, Feature Scaling, Elbow Method, Silhouette Score evaluation, and K-Means clustering.

5. Feature Scaling

StandardScaler was applied to normalize features because K-Means relies on Euclidean distance.

6. Optimal Clusters

Elbow Method and Silhouette Score were used to determine the optimal number of clusters (k=5).

7. Results

Customers were segmented into 5 distinct groups based on spending behavior and income levels.

8. Business Insights

High-income high-spending customers: Premium products. Moderate customers: Loyalty programs. Low-income customers: Discount strategies.

9. Technologies Used

Python, Pandas, NumPy, Matplotlib, Seaborn, Plotly, Scikit-learn.