Customer Segmentation Using Clustering

Objective

Segment mall customers based on **spending behavior, age, income, and gender** using **K-Means clustering** to optimize marketing strategies.

Dataset

- **Mall_Customers.csv**

- **Features:** CustomerID, Gender, Age, Annual Income (k$), Spending Score (1-100)

Methodology

1. **EDA:** Checked distributions, missing values (none found), and gender ratio.

2. **Feature Selection & Preprocessing:** Used **Age, Income, Spending Score**, scaled via **StandardScaler**.

3. **Clustering (K-Means):** Determined **k=5** via **Elbow Method**.

4. **Visualization:** 3D Scatter Plot, Pairplot for cluster relationships.

Key Findings (Customer Segments)

Cluster	Age Group	Income (k$)	Spending Score	Behavioral Insight
0	~45	High (~88)	Low (~17)	High income, frugal
1	~25	Moderate (~55)	High (~80)	Trendy, high spenders
2	~40	Moderate (~55)	Low (~40)	Moderate spenders
3	~25	Low (~25)	High (~80)	Budget-conscious, spend freely
4	~50	Moderate (~50)	Low (~40)	Value-oriented

Challenges & Solutions

1. **Cluster Selection:** Used **Silhouette Analysis** for validation.

2. **Feature Scaling:** Prevented income dominance.

3. **Gender Handling:** Binary encoding; minimal impact on clustering.

4. **Overlapping Clusters:** Business insights helped refine segments.

5. **Data Adaptation: Incremental clustering** for dynamic updates.

Conclusion

K-Means effectively segmented customers, offering actionable insights for **marketing, inventory, and engagement strategies**. Future enhancements: **hierarchical clustering, real-time updates, and campaign testing**.

Tools Used: Python (Pandas, Scikit-learn, Matplotlib, Seaborn)
Dataset: Kaggle - Mall Customers