Customer Segmentation / Clustering

Objective: Perform customer segmentation using clustering techniques to identify distinct customer groups.

Datasets Used:

- Customers.csv: Customer profile data.
- Transactions.csv: Transaction data.

Feature Creation: Generated features like TotalSpending and TransactionFrequency based on transaction data.

Normalization: Scaled the features using **StandardScaler** to ensure all features contribute equally to clustering.

K-Means Clustering: Applied K-means clustering to segment customers based on transaction behavior and spending patterns.

- **Optimal Number of Clusters**: Determined using the elbow method.
- **Clusters Formed**: 4 clusters were chosen as the optimal number based on the elbow curve.

Davies-Bouldin Index (DBI): Used to evaluate the clustering quality.

Silhouette Score: Measures how well-separated the clusters are.

Clustering Results:

- DB Index: 0.799
- Silhouette Score: 0.402

2D Scatter Plot: PCA was used to reduce the dimensions and plot customer clusters in 2D.

Cluster Analysis: Provided insights into each cluster based on spending behavior and transaction frequency.