Methodology

Data Preprocessing:

1.Merged customer profile and transaction datasets.

2.Handled missing values by filling them with column means.

3.Removed non-numeric columns and converted categorical columns into dummy variables.

4.Standardized the data using StandardScaler.

Clustering Algorithms:

K-Means was used for clustering with the number of clusters ranging from 2 to 10.

Metrics such as Silhouette Score and Davies-Bouldin Index were calculated for each configuration.

Cluster Evaluation:

Optimal clusters were selected based on the lowest Davies-Bouldin Index and high Silhouette Scores.

Visualization:

PCA (Principal Component Analysis) was applied to reduce dimensions for cluster visualization.

Optimal Number of Clusters

Number of Clusters Formed: **4**

Clustering Metrics:

Davies-Bouldin Index (K-Means with 4 clusters): **5.606518467696473**
Silhouette Score (K-Means with 4 clusters): **-0.00516449169188199**