Customer Segmentation Using Clustering Techniques

Introduction:

This report details the results of customer segmentation using clustering techniques. Profile information from 'Customers.csv' and transaction data from 'Transactions.csv' were utilized. The goal was to identify meaningful customer clusters, evaluate clustering performance using metrics such as Davies-Bouldin Index, and visualize the clusters.

Methodology :

1. Data Loading and Merging: Data from Customers.csv and Transactions.csv were merged to create a unified dataset.

2. Feature Engineering: Features like total spending, transaction frequency, average transaction value, and days since last transaction were derived.

3. Preprocessing: Features were standardized using StandardScaler. Regions were encoded using LabelEncoder.

4. Clustering: K-Means clustering was performed for 2-10 clusters, and metrics like Davies-Bouldin Index and silhouette score were calculated.
5. Visualization: PCA was used for dimensionality reduction, and clusters were visualized.

Results:

Clustering was performed for cluster counts ranging from 2 to 10. Metrics were calculated to evaluate clustering performance. Below are the results:

Clustering Metrics:

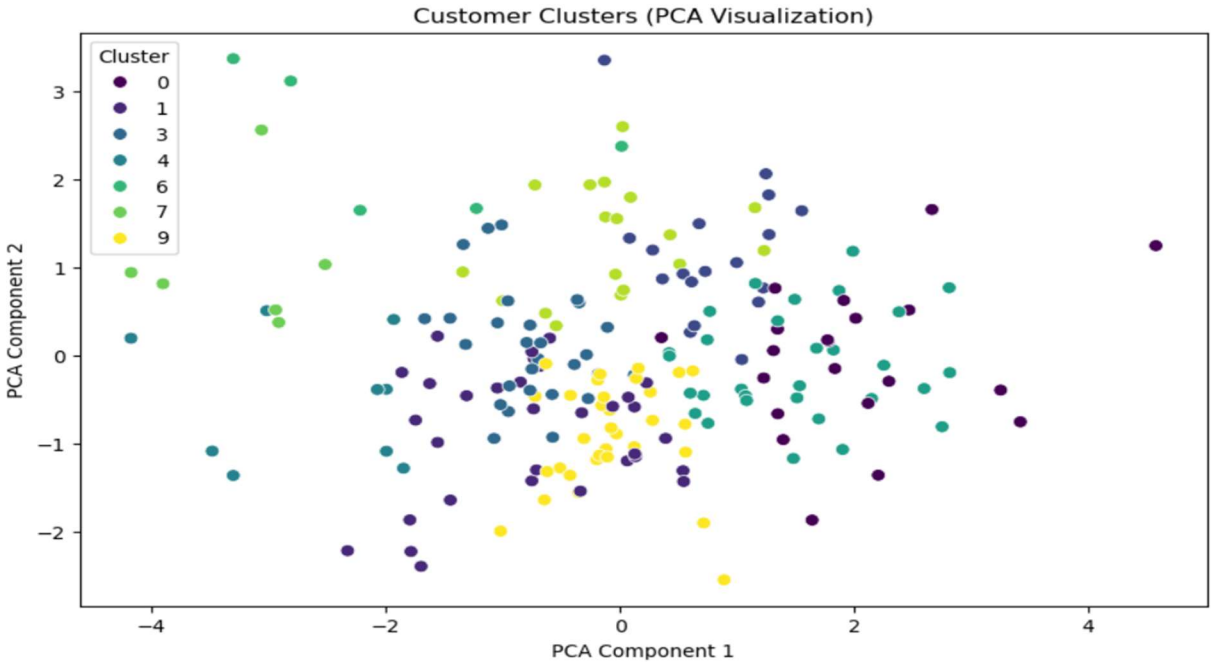
NumClusters	DBIndex	SilhouetteScore
10	1.248022	0.219577

Best Number of Clusters: 10

Davies-Bouldin Index for Best Clusters: 1.248022

Visualization:

Clusters were visualized using PCA, reducing features to two components. The scatter plot below highlights the clusters formed using the best number of clusters (10).



Conclusion:

The clustering analysis successfully segmented customers into 10 clusters with a Davies-Bouldin Index of 1.248. These results provide valuable insights into customer behavior and can help in targeted marketing strategies.