Customer Segmentation Clustering Report

Overview

This report summarizes the results of the customer segmentation analysis performed using clustering techniques on customer and transaction data. The goal was to identify distinct customer segments based on their purchasing behavior and demographic information.

Clustering Methodology

- Clustering Algorithm: K-Means Clustering
- Data Used: Customer profiles created from transaction data, including total spending, transaction count, unique products purchased, and region information.
- Dimensionality Reduction: Principal Component Analysis (PCA) was applied to reduce the feature space for better visualization and clustering performance.

Clustering Results

Number of Clusters Formed

- Optimal Number of Clusters: 8
- The clustering analysis determined that 8 clusters provided a good balance between granularity and interpretability.

Davies-Bouldin Index (DB Index)

- DB Index Value: 0.6316
- The Davies-Bouldin Index measures the average similarity ratio of each cluster with its most similar cluster. A lower value indicates better clustering. The value of 0.6316 suggests that the clusters are somewhat distinct but could be improved.

Other Relevant Clustering Metrics

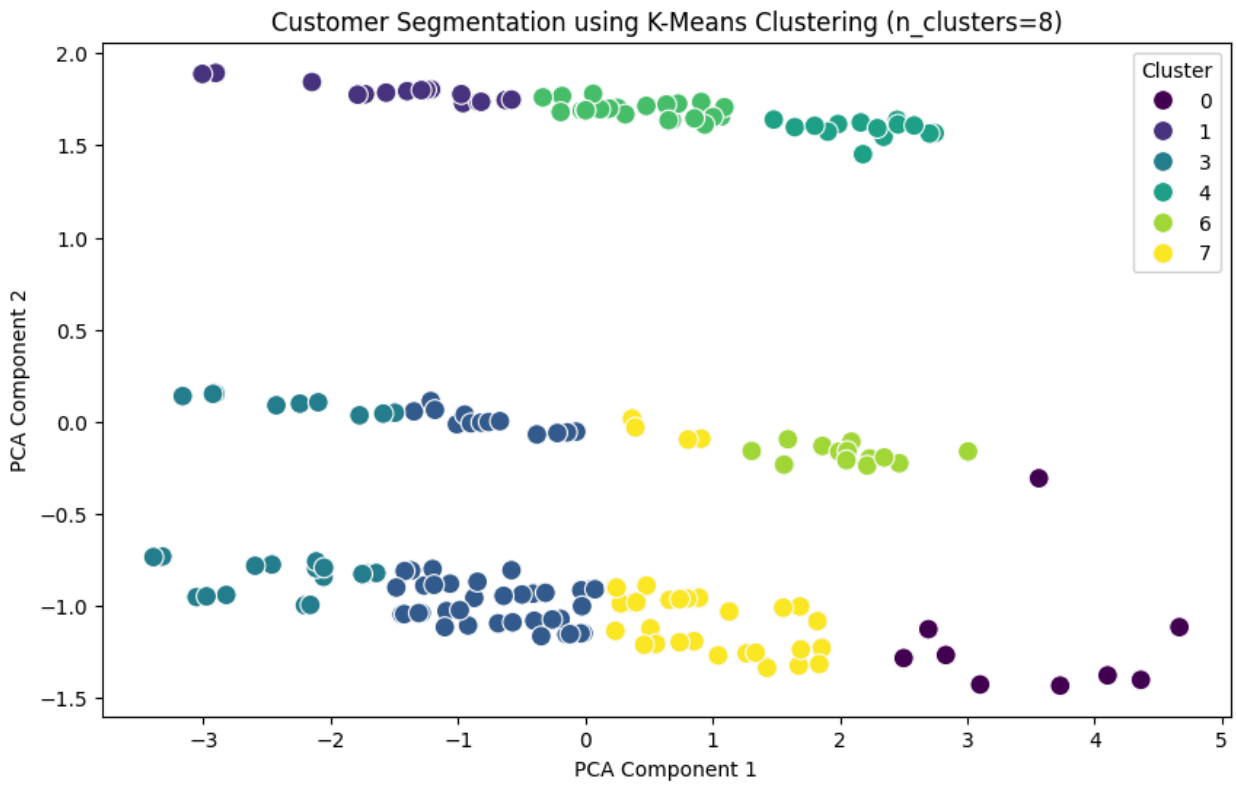
- Silhouette Score: 0.4568
- The Silhouette Score measures how similar an object is to its own cluster compared to other clusters. A score of 0.4568 indicates that the clusters are reasonably well-separated but still have room for improvement.

Explained Variance from PCA

- PCA Component Variance:
- The PCA components explained a significant portion of the variance in the data, indicating that the reduced dimensions retained essential information for clustering.

Visual Representation

- A scatter plot visualizing the clusters was generated, showing the distribution of customers across the two principal components derived from PCA. Each point represents a customer, colored by their assigned cluster.



Conclusion

The clustering analysis successfully identified 8 distinct customer segments based on their purchasing behavior and demographics. While the DB Index and Silhouette Score indicate that the clusters are reasonably well-defined, further refinement of the clustering approach (e.g., exploring different algorithms or feature sets) can yield improved results.