

1. Number of Clusters Formed

Clustering Number: 5

The number of clusters was chosen through preliminary experimentation and assessment of the Elbow Method and Silhouette Score.

2. Davies-Bouldin Index (DB Index)

Value of DB Index : 0.74

Davies-Bouldin Index defines the compactness and separation properties of clusters. The smaller the DB Index, the better will be the well-defined cluster. The DB Index obtained using this number is 0.74, which is good for clustering.

3. Silhouette Score

Silhouette Score: 0.65

This score measures how well each point is assigned to its cluster. A score close to 1 suggests well-separated clusters and a score close to 0 represents overlapping clusters. The score of 0.65 translates to fair to good separation.

4. Clustering Method

Algorithm Used: KMeans Clustering

KMeans was chosen for simplicity and its efficiency in handling large datasets. The clustering results were further improved by normalizing features and one-hot encoding categorical data.

5. Feature Importance

These were the following features that were used to cluster:

Numerical Features:

Total spent (total_spent)

Average transaction value (avg_transaction_value)

Transaction count (transaction_count)

Days since signup (days_since_signup)

Categorical Feature:

Region (one-hot encoded)

These features capture several fundamental aspects of customer behavior, like transaction history and geographic distribution.

6. Assessment of Quality of Clusters

Cluster Distribution:

The sizes of the clusters were well-balanced, and every cluster would have a substantial amount of customers. It means the algorithm did not create outlier-sized clusters.

7. Cluster Visualization

Dimensionality Reduction: PCA (Principal Component Analysis) was applied to reduce the dataset into two dimensions to visualize.

Visualization Results: Clusters are well-separated in the PCA plot, with clear distinctions between groups. The separation indicates that customer profiles and transaction behavior are meaningfully segmented.