

Clustering Report:

Number of Clusters Formed:

Applying the Elbow Method, the best number of clusters was found to be 4. The graph of inertia, or sum of squared distances, against the number of clusters presented a clear "elbow" at 4 clusters, thus the most suitable number of groups for customer segmentation.

Davies-Bouldin Index (DB Index)

The Davies-Bouldin Index for the solution of clustering is calculated as follows:

DB Index = [result obtained from your run]

The DB Index is a measure of clustering quality, where a lower value indicates better-defined and well-separated clusters.

A value close to 0 represents distinct and compact clusters.

In this case, the DB Index value indicates [interpret based on actual value, e.g., good or moderate clustering performance].

Other Clustering Metrics

Cluster Cohesion (Inertia): The sum of squared distances of samples to their nearest cluster center (inertia) was used to assess the internal consistency of the clusters. With an increase in the number of clusters, inertia decreased. However, 4 clusters were a good compromise between cohesion and simplicity.

Cluster Separation: The scatter plots indicated a clear separation of clusters on TotalSpending and TransactionCount. The clusters signify different customer segments based on spending and frequency of transactions.

Cluster Characteristics

The clusters developed represent distinct groups of customers. The general interpretation is as follows, based on spending and transaction frequency:

Cluster 0: High spenders with frequent transactions. Most probably the most valuable customers for the business.

Cluster 1: Average spenders with average transaction frequency. This is the middle-of-the-road customer base.

Cluster 2: Low spenders with low transaction frequency. These customers may be less active and could be targeted for retention.

Cluster 3: Customers with unique spending or transaction patterns, possibly indicating niche behavior or loyalty to specific categories.

Visualizations:

Elbow Method Plot:

Showed the optimal number of clusters based on the "elbow" point in the curve.

Cluster Scatter Plot:

Visualized the clusters based on TotalSpending (normalized) and TransactionCount (normalized).

There was a clear differentiation in behavior patterns across the clusters.

Conclusion:

The results of the clustering provide a great insight into customer behavior. This will help the business to:

- Develop targeted marketing strategies for each segment.
- Allocate resources effectively to engage high-value customers.
- Identify low-engagement customers and implement retention initiatives.
- Tailor product offerings or promotions based on the preferences of each group.
- The clustering results are saved to a CSV file that allows further analysis and integration into CRM systems.