Clustering Report on Customer Segmentation

Introduction

Segmenting customers according to their demographic data and transactional behaviour is the aim of this analysis. KMeans clustering method was used to accomplish the segmentation using information from client transactions and profile data. The primary objective is to establish discrete client groups for focused marketing or other business tactics.

Methodology

Data from the Customers and Transactions datasets were combined to do the segmentation, guaranteeing a thorough understanding of both customer profiles and transaction behaviour. We used the following features for clustering:

- **Total Value**: The total spending of the customer across all transactions.
- **Total Transactions**: The total number of transactions made by the customer.

These features were normalised using MinMaxScaler to make sure the values were on the same scale. After lots of trials, it was determined that 4 clusters was the ideal number.

Clustering Algorithm

For this segmentation task, the KMeans clustering technique was selected. By dividing the data into K clusters, this approach minimises the variation within each cluster. After model training, clients' behaviour was used to assign them to one of the four clusters.

Clustering Metrics

“Davies-Bouldin Index (DBI)” was used to evaluate the quality of the clustering. A lower value indicates better clustering. This index assesses cluster compactness and separation. The clustering's computed DBI was **1.16**. This suggests that the clusters are compact and reasonably well-separated.

Apart from the DBI, further pertinent clustering metrics consist of:

The Silhouette Score, that calculates an object's similarity to its own cluster in relation to other clusters. Better-defined clusters are indicated by a higher silhouette score.

Sizes of Clusters: No cluster was excessively bigger or smaller than the others, as seen by the balanced customer distribution among the clusters.

Results

Four separate clusters were created by the KMeans algorithm, each of which represented a unique set of consumer behaviours:

1. **Cluster 1**: Customers with low spending and fewer transactions.
   These customers are likely less engaged.
2. **Cluster 2**: Customers with moderate spending and transaction frequency.
3. **Cluster 3**: High-value customers with many transactions, indicating loyal or premium customers.
4. **Cluster 4**: Customers with high spending but lower transaction frequency, probably indicating high-ticket item buyers.

These clusters are reasonably separated, as indicated by the DBI score of 1.16.

The clustering quality could be further confirmed by using additional measures, such as silhouette scores.

Visualization

To see the clustering findings, a scatter plot was made, with the X-axis standing for Total Spending (Total Value) and the Y-axis for Total Transactions.
The plot clearly displayed grouping of customers into four clusters, with each cluster being represented by different colours.

Conclusion

1) The KMeans clustering analysis segmented customers into distinct groups based on their transaction behavior.
2) The clusters provide valuable insights into customer types, which can be leveraged for targeted marketing or personalized offers.
3) The clustering model's performance, as evaluated by the DB Index, indicates good separation between clusters, confirming the effectiveness of the segmentation process.