Clustering Report on Customer Segmentation

Introduction

Segmenting customers according to their demographic data and transactional behavior was the aim of this investigation. Using information from client transactions and profile data, the KMeans clustering method was used to accomplish the segmentation. Establishing discrete client groups for focused marketing or other business tactics was the aim.

METHODOLOGY

Data from the Customers and Transactions datasets were combined to do the segmentation, guaranteeing a thorough understanding of both customer profiles and transaction activity. The following characteristics were employed in our clustering:

- Total Value: The sum of the customer's expenditures for all transactions.
- Total Transactions: The sum of the customer's transactions.

Prior to utilizing the clustering process, these features were standardized using MinMaxScaler to make sure the values were on the same scale. After multiple testing, it was determined that four clusters was the ideal quantity.

CLUSTERING ALGORITHM

Data from the Transactions and Customers datasets were combined, and the Clustering Algorithm was used to carry out the segmentation.
For this segmentation task, the KMeans clustering technique was selected. By dividing the data into K clusters, this approach minimizes the variation within each cluster. Customers were categorized into one of the four clusters according to their behavior after the model was trained, guaranteeing a thorough understanding of both transaction behavior and client profiles. The following characteristics were employed in our clustering:

- Total Value: The sum of the customer's expenditures for all transactions.
- Total Transactions: The sum of the customer's transactions.

Prior to utilizing the clustering process, these features were standardized using MinMaxScaler to make sure the values were on the same scale. Four clusters were chosen as the number, as it was found to be optimal after several trials.

CLUSTERING METRICS

Using the Davies-Bouldin Index (DBI), the clustering's quality was assessed. A lower score denotes better clustering. This index assesses the compactness and separation of clusters. The clustering's computed DBI of 1.16 indicates that the clusters are compact and somewhat well-separated.
Apart from the DBI, further pertinent clustering metrics consist of:
The Silhouette Score calculates an object's similarity to its own cluster in relation to other clusters. Better-defined clusters are indicated by a higher silhouette score.
• Cluster Sizes: No cluster was notably bigger or smaller than the others, as seen by the fairly balanced consumer distribution across the clusters.

RESULT

Four separate clusters were created by the KMeans algorithm, each of which represented a unique set of consumer behaviors:
1. Cluster 1: Clients who make fewer purchases and spend less. These clients are probably less involved.
2. Cluster 2: Clients that spend moderately and make transactions frequently.
3. Cluster 3: Loyal or premium consumers are indicated by high-value clients with a high volume of transactions.
4. Cluster 4: Clients who spend a lot of money but don't buy things very often, which may be a sign that they purchase expensive goods.
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

The clustering findings were visualized using a scatter plot, where the Y-axis represented Total Transactions and the X-axis represented Total Spending (TotalValue). Customers were clearly grouped into four clusters, as indicated by the graphic, with each cluster being represented by a distinct hue.

CONCLUSION

Customers were successfully divided into discrete groups according to their transaction behaviour using the KMeans clustering methodology. The clusters offer insightful information about different kinds of customers, which can be used to tailor offerings or conduct targeted marketing. The efficacy of the segmentation process is confirmed by the clustering model's performance, which is measured by the DB Index and shows strong separation across clusters.The clustering findings were visualized using a scatter plot, where the Y-axis represented Total Transactions and the X-axis represented Total Spending

(TotalValue). Customers were clearly grouped into four clusters, as indicated by the graphic, with each cluster being represented by a distinct hue.