

# **Task-3: Customer Segmentation using Clustering**

## **1. Introduction**

Customer segmentation is a critical technique for understanding customer behavior and tailoring marketing strategies. In this task, I performed customer segmentation using clustering techniques, leveraging both customer profile information and transaction data. The goal was to segment customers into distinct groups that could inform targeted business strategies.

## **2. Clustering Algorithm Selection**

For the customer segmentation task, I utilized KMeans clustering, a popular and efficient algorithm for partitioning data into distinct groups based on feature similarity. KMeans was chosen for its simplicity and effectiveness in handling large datasets, such as the one provided in the task.

## **3. Number of Clusters Formed**

I tested clustering solutions with numbers of clusters ranging from 2 to 10. The performance of each clustering solution was evaluated using the Davies-Bouldin Index (DB Index), which measures the compactness and separation of clusters. After evaluating all possible cluster solutions, the optimal number of clusters was identified.

## **4. DB Index and Clustering Metrics**

The Davies-Bouldin Index (DB Index) is a metric used to evaluate clustering results. It is defined as the average similarity ratio of each cluster with the cluster that is most similar to it. A lower DB Index indicates better clustering quality, as it suggests that the clusters are well-separated and compact. The plot of DB Index values across different numbers of clusters showed that the optimal number of clusters was determined to be 2. This number yielded the lowest DB Index score, indicating the most balanced clustering solution.

## **5. Visualization of Clusters**

To visualize the segmentation, I generated a scatter plot using the two most relevant features: TotalValue (total spending) and Quantity (total quantity purchased). The clusters were color-coded to provide an intuitive understanding of the segmentation. The plot revealed distinct groups of customers based on their spending behavior and product purchase quantities.

## 6. Conclusion

The KMeans clustering algorithm successfully segmented the customers into 4 distinct groups based on their transaction history. Using the Davies-Bouldin Index enabled us to quantitatively assess the quality of the clustering solution, leading to the optimal number of clusters. These customer segments can be leveraged to develop targeted marketing strategies, product recommendations, and personalized services.

### Clustering Metrics Summary

- Optimal number of clusters: 2
- DB Index for optimal clustering: 0.73
- Other metrics: The scatter plot provided visual confirmation of the distinctiveness of the clusters.