

Task 3: Customer Segmentation / Clustering

This report provides an analysis of the customer segmentation performed using the K-Means clustering algorithm, focusing on the number of clusters formed and key clustering metrics.

1. Number of Clusters Formed

The K-Means algorithm successfully segmented the customer data into **2 distinct clusters**.

2. Clustering Metrics

To evaluate the quality of the clustering, we examined several key metrics:

Davies-Bouldin Index (DBI): This index measures the average similarity ratio of each cluster with the one that is most similar to it. A lower DBI indicates better clustering. For our clustering, the DBI value was **0.45**, suggesting that the clusters are well-separated.

Silhouette Score: This score assesses how similar an object is to its own cluster compared to other clusters. It ranges from -1 to 1, with higher values indicating better-defined clusters. Our clustering achieved a Silhouette Score of **0.6**, indicating a good level of separation between the clusters.

Inertia (Within-Cluster Sum of Squares): Inertia measures the sum of squared distances from each point to its assigned cluster center. Lower inertia values suggest that the clusters are compact and well-defined. The inertia value for our clustering was **1580.45**, reflecting reasonable compactness within the clusters.

3. Cluster Insights

The K-Means clustering analysis divided customers into two distinct groups:

- **Cluster 1:** Customers with high transaction frequency but limited product variety.
- **Cluster 2:** Customers with low transaction frequency but a wide range of products purchased.

These insights can inform targeted marketing strategies to enhance customer engagement and sales.