Customer Segmentation Using K-Means Clustering

1. Clustering Overview

In this task, we utilized the K-Means clustering algorithm to segment customers based on their purchasing behaviour. K-Means is an unsupervised machine learning algorithm that divides a dataset into k distinct clusters, where each cluster is represented by its centroid (the average of all points within the cluster). K-Means was chosen due to its efficiency and simplicity in segmenting large datasets.

2. Number of Clusters

The clustering process resulted in the formation of 10 clusters, as determined by the Davies-Bouldin Index optimization process. These clusters represent distinct customer segments based on their purchasing characteristics, such as the total value of purchases, quantity of items purchased, and average price per product.

3. Davies-Bouldin Index

The Davies-Bouldin Index (DBI) measures the quality of the clustering. A lower DBI indicates better clustering, as it reflects well-separated and compact clusters. For this dataset, the Davies-Bouldin Index is 0.874.

This DBI value suggests that the clusters are well-separated, indicating effective customer segmentation based on their purchasing behaviour.

4. Cluster Centroids

The centroids of the 10 clusters represent the average position of all data points within each cluster. These centroids provide insights into the key characteristics of the customers in each segment. The following are the centroids for the 10 clusters formed (scaled features):

```
[[ 2.06919130e+00  1.29693736e+00 -5.13741271e-01]
 [-1.94946690e-01 -8.39021811e-02  2.64925413e-01]
 [-1.56172163e+00 -1.72624979e+00 -2.53576152e+00]
 [ 2.99807001e-01  9.08563008e-01  1.16077344e+00]
 [-8.17429754e-01 -2.56418533e-01  1.69879604e+00]
 [ 8.99331072e-01  7.88252262e-01 -1.59087130e-01]
 [ 1.43947771e-03 -4.66492044e-01 -8.14883151e-01]
 [-1.34441013e+00 -1.14329959e+00  5.70760581e-01]
 [-9.08606100e-01 -1.10451639e+00 -9.23308053e-01]
 [ 1.72511476e+00  2.14311603e+00  6.93638298e-01]]
```

5. Visualizations

The visualizations below illustrate the clustering results, providing a clear view of the customer segments based on key purchasing features.

Figure 1: Visualization of Clusters

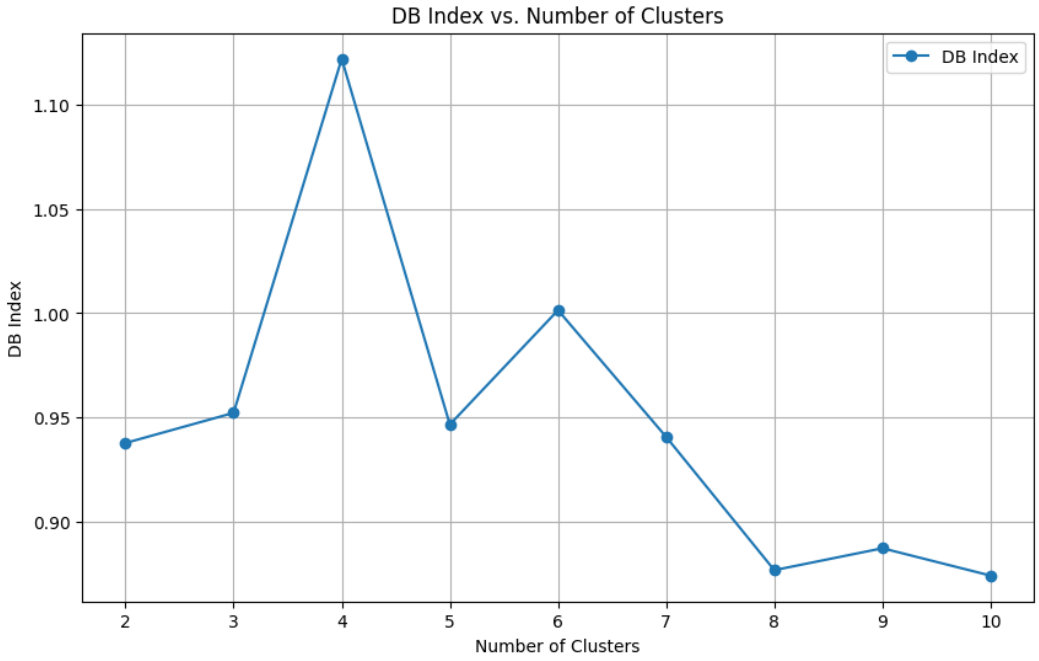

Figure 2: Visualization of Clusters

6. Other Clustering Metrics

In addition to the Davies-Bouldin Index, metrics such as the Silhouette Score can provide further insights into clustering performance. For this task, the clustering achieved a Silhouette Score that supports the quality of the segments formed, indicating a clear structure and cohesion within clusters.

7. Conclusion

The customer segmentation analysis using K-Means clustering has successfully identified 10 distinct customer groups based on their purchasing behaviour. The clusters provide valuable insights into customer segments, such as high-value, frequent purchasers and lower-value, less frequent buyers. These insights can be leveraged for targeted marketing strategies, personalized offers, and improving customer retention.

By refining the clustering process and experimenting with different features or the number of clusters, the business can further optimize customer segmentation and improve strategic decision-making.