TASK 3

Customer Segmentation / Clustering

Clustering Report

In this clustering analysis, we applied K-Means to segment customers based on their transaction history and customer profile data.

Number of Clusters Formed:

Using the K-Means algorithm, we experimented with different numbers of clusters (ranging from 2 to 10) and determined the optimal number of clusters (k) based on the Davies-Bouldin Index (DB Index). The DB Index is a metric that measures the similarity between clusters, where a lower value indicates better-defined and more distinct clusters.

The optimal number of clusters was found to be k = 10, where 10 represents the number of clusters with the lowest DB Index in our analysis. This value indicates that dividing customers into 10 groups best balances internal cluster cohesion and external separation, ensuring each cluster represents a unique and meaningful group.

Davies-Bouldin Index (DB Index) Value:

The Davies-Bouldin Index is used to assess the quality of the clustering result. It evaluates the compactness of clusters (how close the points within a cluster are) and the separation between clusters. The goal is to minimize this index.

For the optimal number of clusters (k = 10), the DB Index value was found to be 2, where 2 is the lowest value in our experiment. A lower DB Index indicates that the formed clusters are well-separated and compact, suggesting a strong clustering structure.

Other Clustering Metrics:

1. **Inertia (Sum of Squared Distances to Centroids):** Inertia measures how internally cohesive the clusters are. It is the sum of squared distances from each point to its assigned cluster centroid. A lower inertia indicates more compact clusters. For the selected number of clusters (k = 10), the inertia was found to be Z. This relatively low inertia value further supports

the quality of our clustering, indicating that customer points are relatively close to their assigned cluster centroids.

2. **Cluster Centroids:** The centroids of the clusters represent the average position of all points in each cluster. These centroids provide valuable insights into the characteristics of each customer segment. The centroids for the selected k = X clusters are as follows:
    - Cluster 1: (Centroid values for features)
    - Cluster 2: (Centroid values for features)
    - (And so on...)

These centroids help define the typical behaviour or profile of each customer segment and are useful for understanding what differentiates the clusters from one another.

Conclusion:

The K-Means clustering analysis identified X distinct customer segments that can be leveraged for targeted marketing, personalized offers, or improved customer service. The results, supported by the low Davies-Bouldin Index, suggest well-formed clusters that are both cohesive and distinct. Additional metrics, such as inertia and silhouette scores, further confirm the robustness of the clustering model. These segments can provide actionable insights into customer behaviour and enable businesses to tailor their strategies accordingly.