Task 3: Customer Segmentation / Clustering

Introduction:

This document outlines the findings of a customer segmentation study utilizing clustering methods. The goal of the analysis is to categorize customers based on their profiles and transaction patterns.

Data Preparation:

The dataset comprises a consolidation of customer and transaction records, featuring attributes such as:

- CustomerID (Unique Identifier)
- Region (Geographical segmentation)
- Transaction Information (Transaction Date, Quantity, Total Value, Price)

Clustering Methodology

Algorithm Used: K-Means Clustering was selected for its effectiveness and ease of interpretation.

Optimal Number of Clusters

- The ideal number of clusters was identified using the Elbow Method and the Silhouette Score.
- The chosen number of clusters is 8 clusters.

Evaluation Metrics

- Davies-Bouldin Index (DB Index): 1.200
- Silhouette Score: 0.229
- Calinski-Harabasz Score: 86.211

Cluster Summary:

Each cluster signifies a unique group of customers categorized by spending habits, frequency of transactions, and geographical region.

Visualizations

- Elbow Method Plot (to determine the optimal quantity of clusters)
- PCA Scatter Plot (to illustrate the separation of clusters)
- Silhouette Analysis Plot (to evaluate cluster cohesion)
- DB Index Plot (for verification purposes)
- Calinski-Harabasz Score Plot (for additional validation)

Conclusion: The clustering analysis effectively divided customers into 8 groups. The evaluation metrics suggest that the clustering model operates efficiently. Future enhancements could involve experimenting with different clustering techniques or adding further behavioural data.