Clustering Results Report

Overview

This report summarizes the clustering analysis performed using KMeans on the provided dataset. The primary objectives were to identify natural groupings within the data and evaluate the quality of clustering using standard metrics.

Key Results

1. **Number of Clusters Formed**:
    - A total of **2 clusters** were identified using the devis index and implemented in KMeans clustering. This number was determined as optimal based on the minimum of dbi score.
2. **Clustering Quality Metrics**:
    - **Davies-Bouldin Index (DBI)**:
        - Value: **0.96**
        - Interpretation: Lower values indicate better-defined clusters. The value suggests moderately well-separated clusters.
    - **Silhouette Score**:
        - Value: **0.62**
        - Interpretation: The score indicates a reasonable level of cohesion within clusters and separation between clusters. While not perfect, the clusters demonstrate a distinguishable structure.
3. **Cluster Characteristics**:
    - Each cluster represents unique characteristics of the data. Patterns emerge based on spending behaviour, transaction volume, and other features. A detailed summary of these patterns can be provided upon further analysis.

Preprocessing Steps

- **Standardization**:
    - Data was scaled using standard normalization techniques to ensure equal weighting of features during clustering.
- **Feature Selection**:
    - Relevant features were selected to improve clustering quality and reduce noise in the dataset.

My Observations

1. **Cluster Interpretation**:
    - Further domain-specific analysis is recommended to understand the practical implications of each cluster.
    - Labelling clusters based on their defining characteristics can help in actionable decision-making.
2. **Model Evaluation**:
    - While the Silhouette Score and DBI suggest reasonable performance, exploring advanced clustering techniques DBSCAN could yield improved results.
3. **Data Insights**:
    - Insights derived from clusters can assist in targeted marketing, personalized recommendations, and resource allocation.