Customer Segmentation Report

Objective

The goal of this task was to perform customer segmentation using clustering techniques based on profile and transaction information. The analysis included:

1. Selecting an appropriate clustering algorithm.
2. Determining the optimal number of clusters between 2 and 10.
3. Evaluating the clustering performance using metrics such as the Davies-Bouldin Index (DB Index).
4. Visualizing the clusters for better interpretation.

Methodology

Data Preparation

1. Aggregated customer features from the transactional data:
    - TotalValue: Sum of transaction values for each customer.
    - Quantity: Total quantity purchased by each customer.
    - TransactionPrice: Average transaction price per customer.
2. Normalized the aggregated features using StandardScaler to ensure fair clustering.

Clustering Algorithm

- Used **K-Means Clustering** with 4 clusters (n_clusters=4), as determined during experimentation.

Evaluation Metric

- The **Davies-Bouldin Index** was used to evaluate the clustering performance. Lower values indicate better-defined clusters.

Dimensionality Reduction

- Applied **Principal Component Analysis (PCA)** to reduce the data to two dimensions for visualization.

Results

1. **Number of Clusters Formed**: 4
2. **Davies-Bouldin Index**: **1.1218**
3. **Cluster Visualization**:
    - The clusters were visualized in 2D space using PCA.
    - Each cluster was well-separated and could represent distinct customer segments.

Insights

1. Customers were grouped based on purchasing behavior (e.g., total value, quantity, and average price).
2. The clustering outcome can be used to:
    - Design targeted marketing strategies.
    - Identify high-value customers.
    - Develop loyalty programs tailored to specific segments.
3. The Davies-Bouldin Index value of **1.1218** indicates reasonably well-defined clusters.