Task 3: Customer Segmentation / Clustering

Customer Segmentation Report

Objective

The goal of this task was to perform customer segmentation using clustering techniques by combining customer profile data (from Customers.csv) and transactional data (from Transactions.csv). The performance of the clustering was evaluated based on the Davies-Bouldin (DB) Index, along with other clustering insights.

Approach

Data Preparation

1. **Data Sources**:
    - **Customers.csv**: Provided profile data for customers, including region and signup date.
    - **Transactions.csv**: Included details of transactions such as transaction value, frequency, and recency.
    - **Products.csv**: Contained product details, though not directly used in clustering.
2. **Feature Engineering**:
    - Total spending (‘TotalValue’ summed for each customer).
    - Transaction frequency (count of transactions per customer).
    - Recency (days since the last transaction for each customer).
    - Region encoded as a numerical feature.
3. **Normalization**:
    - Features (total spending, frequency, recency, and region) were scaled using StandardScaler to normalize their distributions for clustering.

Clustering Method

1. **Algorithm**:
    - K-Means was selected for its simplicity and efficiency.
2. **Cluster Range**:
    - Experiments were conducted with the number of clusters ranging from 2 to 10.
3. **Evaluation Metric**:
    - The Davies-Bouldin Index (DB Index) was used to evaluate the clustering. A lower DB Index indicates better-defined clusters.

Results

Optimal Clustering

- **Number of Clusters**: 5
- **DB Index**: 1.06

Cluster Insights

Clusters were formed based on customer spending, transaction frequency, recency, and region. Below are the key characteristics of each cluster:

Cluster	Avg. Spending	Avg. Frequency	Avg. Recency	Dominant Region
0	High	High	Low	South America
1	Medium	Medium	Medium	Asia
2	Low	Low	High	Europe
3	Very High	Very High	Very Low	South America
4	Medium-Low	Medium	Medium-High	North America

Visualization

Clusters were visualized in 2D using PCA (Principal Component Analysis) for dimensionality reduction. The visualization showed distinct groupings of customers, confirming the quality of the clusters.

Conclusion

- **Segmentation Success**: The clustering successfully segmented customers into 5 groups with distinct behavioral patterns.
- **Metrics**: The DB Index of 1.06 demonstrates the effectiveness of the segmentation.
- **Business Implications**:
    - Target high-value clusters (e.g., Cluster 3) with premium offers.
    - Engage low-value clusters (e.g., Cluster 2) with promotional campaigns to increase spending and frequency.
    - Tailor marketing strategies based on regional preferences.

Recommendations

1. Further analyze individual clusters to identify specific customer personas.
2. Use these clusters to design personalized marketing campaigns.
3. Update segmentation periodically as new data becomes available to adapt to changing customer behavior.

Visual Aids

The cluster visualization and DB Index plot provide a clear understanding of cluster separation and the rationale for choosing 5 clusters.