Task 3: Customer Segmentation / Clustering

Task Overview

This report documents findings for Task 3: Customer Segmentation / Clustering. The task involved segmenting customers based on their profile and transaction data using clustering techniques.

Objectives

1. Segment customers using clustering techniques.
2. Use Customers.csv (profile information) and Transactions.csv (transaction data).
3. Calculate relevant clustering metrics, including the DB Index.
4. Visualize the clusters.

Methodology

Data Preparation

- Profile and transaction data were combined using customer IDs as the key.
- Derived features included Recency, Frequency, and Monetary (RFM) values.
- Features were scaled using standard normalization techniques for clustering.

Clustering Technique

- **Algorithm Used**: K-Means clustering.
- **Evaluation Metrics**:
    - **Davies-Bouldin (DB) Index**: Measured cluster compactness and separation.
    - **Silhouette Score**: Evaluated intra-cluster cohesion and inter-cluster separation.
    - **WCSS (Within-Cluster Sum of Squares)**: Used for determining the optimal cluster count (via Elbow Method).

Results

Optimal Number of Clusters

- Based on the analysis, the optimal number of clusters was determined to be **10**.

Clustering Metrics

1. **DB Index**: Achieved a value of **0.91**, indicating effective clustering with compact and well-separated clusters.
2. **Silhouette Score**: Reflects cluster cohesion and separation (not explicitly listed in the outputs).
3. **Cluster Insights**:
    - Clusters were analyzed for spending behavior, transaction frequency, and product diversity. Below are highlights:

Cluster	Avg. Spending	Avg. Transactions	Unique Products	Customer Count
0	1831.78	2.73	2.73	15
1	4180.90	5.81	5.78	27
2	3959.38	5.25	5.14	28
3	5379.35	7.50	7.06	18
4	2124.87	3.19	3.12	26
5	3470.24	5.80	5.80	25
6	1558.39	2.88	2.82	17
7	6964.94	9.71	9.43	7
8	1940.75	3.05	2.95	22
9	5856.95	7.71	7.21	14

Visualizations

1. **Cluster Scatter Plot**:
    - PCA was used to reduce dimensionality for 2D visualizations, clearly showing distinct clusters.
2. **Heatmaps**:
    - Highlighted feature distributions across clusters (e.g., spending, transactions).

3. **Centroid Bar Charts**:
    - Displayed average feature values for each cluster.

Conclusion

- Successfully segmented customers into **10 clusters** based on combined profile and transaction data.
- The **DB Index of 0.91** confirms effective clustering.
- Cluster insights reveal actionable customer behavior patterns for targeted marketing strategies.

Recommendations

1. Focus loyalty programs on high-spending clusters (e.g., Cluster 7).
2. Encourage low-frequency shoppers (e.g., Cluster 6) with personalized offers.
3. Design campaigns targeting clusters with high product diversity (e.g., Clusters 7 and 3).