Clustering Report: Customer Segmentation

Objective

To segment customers based on their transaction behavior and profile information using clustering techniques and evaluate cluster quality using metrics like Davies-Bouldin Index (DB Index) and Silhouette Score.

Clustering Results

1. **Number of Clusters Formed:**
    - The optimal number of clusters was determined to be 4, based on the elbow method.
2. **Davies-Bouldin Index (DB Index):**
    - DB Index: 1.37
    - A lower DB Index indicates better clustering. The value of 1.37 suggests moderate separation and compactness of clusters.
3. **Silhouette Score:**
    - Silhouette Score: 0.22
    - A Silhouette Score closer to 0 indicates overlapping clusters or less-defined boundaries between clusters. The score of 0.22 suggests weakly defined clusters, potentially due to overlapping features or insufficient separation in the data.

Cluster Characteristics

1. **Cluster 1 (High Spenders):**
    - Customers with high total transaction value and frequent purchases.
    - Strategy: Focus on loyalty programs and exclusive offers.
2. **Cluster 2 (Moderate Spenders):**
    - Customers with average transaction value and diverse product interests.
    - Strategy: Design cross-sell and upsell campaigns.
3. **Cluster 3 (Occasional Buyers):**
    - Customers with low purchase frequency and minimal spending.
    - Strategy: Use promotional offers to encourage higher engagement.

4. **Cluster 4 (New Customers):**
    - Recently onboarded customers with little transaction data.
    - Strategy: Engage with onboarding campaigns and incentives.

Evaluation Metrics

1. **Davies-Bouldin Index (1.37):**
    - Indicates moderately well-separated clusters with some overlap.
2. **Silhouette Score (0.22):**
    - Suggests weakly defined clusters, potentially requiring feature engineering or different clustering techniques.
3. **Cluster Distribution:**
    - Review the number of customers in each cluster to ensure a balanced segmentation. Unbalanced clusters might require adjustments in feature scaling or algorithm parameters