Customer Segmentation Clustering Report

1. Overview

The clustering analysis was conducted to segment customers based on their transaction behavior and profile features. The dataset included customer demographic information and transaction history, which were preprocessed and used for clustering.

2. Number of Clusters Formed

Based on the Elbow Method and Silhouette Score analysis, the optimal number of clusters was determined to be **4**. This number effectively balances the trade-off between minimizing intra-cluster variance and maximizing inter-cluster separation.

3. Clustering Performance Metrics

a) Davies-Bouldin Index (DB Index)

- **Value:** 0.874 (Lower values indicate better clustering quality)
- A DB Index close to 0 suggests that the clusters are well-separated and compact.

b) Silhouette Score

- **Value:** 0.526 (Ranges from -1 to 1, where higher values indicate well-defined clusters)
- A score of 0.526 suggests a moderately strong clustering structure with well-separated clusters.

4. Data Preprocessing

To ensure effective clustering, the following preprocessing steps were applied:

- **Standardization**: Numerical features (e.g., Total Spend, Recency, Tenure) were scaled using StandardScaler.
- **One-Hot Encoding**: Categorical feature 'Region' was encoded using OneHotEncoder.
- **Handling Missing Data**: Customers with no transactions were assigned default values.

5. Cluster Visualization

To visualize the clusters in two dimensions, **Principal Component Analysis (PCA)** was applied:

- The first two principal components were extracted and plotted.
- Clusters were well-separated, confirming the appropriateness of the chosen number of clusters.

6. Conclusion

The customer segmentation results indicate that customers exhibit distinct purchasing behaviors and profiles. The clustering model provides valuable insights for personalized marketing strategies, targeted promotions, and customer retention plans.