Customer Segmentation Report

Objective

This analysis aims to segment customers into distinct clusters based on their profile and transaction history using clustering techniques. This segmentation aims to provide actionable insights for targeted marketing, personalized customer engagement, and operational efficiency.

Clustering Methodology

- **Clustering Algorithms Tested**:
  - KMeans Clustering
  - Agglomerative (Hierarchical) Clustering
  - DBSCAN
- **Evaluation Metrics**:
  - Davies-Bouldin (DB) Index
  - Silhouette Score
  - Calinski-Harabasz Index
- **Number of Clusters**:
  - **7 clusters** were chosen as the optimal configuration after testing with 3, 5, 7, and 9 clusters. The 7-cluster configuration provided better separation and coherence compared to 3 or 5 clusters, and results were comparable to 9 clusters without overfitting.

Best Performing Method

- **Hierarchical Clustering (Agglomerative)** was determined to be the best method based on the evaluation metrics.

Evaluation Metrics

1. **Davies-Bouldin Index**: **0.9569**
   - A lower DB Index indicates better-defined clusters. This score reflects good separation and cohesion in the clusters.
2. **Silhouette Score**: **0.3707**
   - Indicates the extent to which a data point belongs to its cluster. While this score is moderate, it shows that the clusters are reasonably well-defined.
3. **Calinski-Harabasz Index**: **77.566**
   - Measures the ratio of between-cluster dispersion to within-cluster dispersion. A higher score reflects better clustering quality, and this score is indicative of distinct clusters.

Cluster Insights

The segmentation process grouped customers into 7 distinct clusters based on:

1. **Profile Information**:
    - Region
    - Signup Date
2. **Transaction History**:
    - Total Transactions
    - Total Spending
    - Average Transaction Value

Each cluster represents a unique segment of customers with similar behaviors and profiles.

Why Hierarchical Clustering?

- Hierarchical Clustering outperformed KMeans and DBSCAN in terms of the Davies-Bouldin Index, which is the primary evaluation metric.
- It captures customer relationships in a hierarchical structure, allowing for better-defined clusters in this dataset.
- The dendrogram visualizations confirmed meaningful separations between clusters.

Conclusion

The clustering analysis identified **7 well-defined customer segments** using Hierarchical Clustering with the following evaluation scores:

- **DB Index**: 0.9569
- **Silhouette Score**: 0.3707
- **Calinski-Harabasz Index**: 77.566