Clustering Results Report3

Introduction

This report summarizes the results of customer segmentation performed using clustering techniques on the provided eCommerce dataset. The goal of this task was to segment customers into distinct groups based on their profile and transaction history, enabling targeted marketing and strategy development.

Clustering Results

1. Number of Clusters Formed

- The optimal number of clusters was determined to be **10**, based on the evaluation of clustering metrics, particularly the Davies-Bouldin Index (DB Index) and the Silhouette Score.

2. Davies-Bouldin Index (DB Index)

- **DB Index**: **0.8928**
- A lower DB Index indicates better-defined and well-separated clusters. The achieved value of 0.8928 suggests that the clusters are reasonably compact and distinct, providing meaningful segmentation.

3. Silhouette Score

- **Silhouette Score**: **0.3781**
- The Silhouette Score measures how well-separated the clusters are, with values closer to 1 indicating better-defined clusters. A score of 0.3781 suggests moderate separation among the clusters.

4. Characteristics of the Clusters

- Each cluster represents a distinct customer segment based on their transaction behavior and profile information.
- Clusters show variations in key metrics such as:
    - **Total Spending**: High-spending customers are grouped separately from low-spending customers.
    - **Transaction Frequency**: Customers with frequent purchases form distinct clusters compared to occasional buyers.
    - **Product Preferences**: Clusters exhibit differences in preferences for specific product categories (e.g., Electronics, Home Appliances).

    - **Geographic Distribution**: Region-based segmentation is evident, with customers from similar regions grouped together.

Evaluation of Clustering Performance

1. DB Index

- The DB Index of **0.8928** is indicative of well-formed clusters with minimal overlap.
- This metric highlights the compactness of clusters and the separation between them.

2. Silhouette Score

- The score of **0.3781** indicates moderate separation between clusters. While there is room for improvement, this score is acceptable for the given dataset and feature set.

3. Visual Inspection

- Pairplots and other visualizations of the clusters demonstrate clear groupings in some dimensions (e.g., spending behavior, region), validating the numerical metrics.

Actionable Insights

1. **High-Value Customer Targeting**:
    - Customers in clusters with high spending and frequent transactions can be prioritized for loyalty programs or exclusive offers.
2. **Region-Specific Campaigns**:
    - Clusters formed based on regional preferences can inform localized marketing strategies and inventory planning.
3. **Product Recommendations**:
    - Understanding product category preferences within clusters enables personalized product recommendations, increasing customer satisfaction and sales.
4. **Improvement Opportunities**:
    - Additional features (e.g., demographic data, customer feedback) could further refine the clustering model and improve separation.

Conclusion

The clustering analysis successfully segmented customers into 10 distinct groups based on their transaction history and profile information. The moderate DB Index (0.8928) and Silhouette Score (0.3781) suggest reasonable cluster quality. These clusters can drive targeted marketing,

personalized recommendations, and region-specific strategies, ultimately enhancing business outcomes.