Customer Segmentation Report

Overview: This report highlights how customer segmentation based on transactional and profile data can enhance targeted marketing efforts and improve customer retention. Clustering techniques were used to identify groups of customers with similar behaviors, enabling more effective strategic decision-making. The goal is to identify groups of customers with similar behaviors to facilitate targeted marketing and strategic decision-making.

Data Preparation:

1. **Datasets Used:**
    - **Customers.csv:** Contains CustomerID, Name, Region, and SignupDate.
    - **Transactions.csv:** Contains TransactionID, CustomerID, ProductID, Quantity, TotalValue, etc.
2. **Steps Taken:**
    - Aggregated transactional data to compute customer-level metrics:
        - Total Spend, Total Transactions, and Average Transaction Value. These metrics were chosen because they capture critical aspects of customer behavior: Total Spend reflects overall value, Total Transactions indicates purchase frequency, and Average Transaction Value helps identify spending patterns. Together, they contribute to clustering accuracy by highlighting distinct customer profiles.
    - Merged aggregated metrics with customer profiles.
    - One-hot encoded the "Region" column and normalized numeric features for uniform scaling.

Clustering Process:

1. **Clustering Algorithm:**
    - K-Means clustering was applied to segment customers.
2. **Evaluation Metric:**
    - The **Davies-Bouldin Index (DBI)** was used to evaluate the compactness and separation of clusters.
    - Optimal clusters were determined by minimizing the DBI score.

3. **Optimal Number of Clusters:**
    - The best number of clusters was found to be **4**, with a DBI score of **0.5824**.

Cluster Characteristics: The following cluster characteristics were derived from an analysis of key metrics, including Total Spend, Total Transactions, and Average Transaction Value. These insights highlight distinct customer behaviors and their significance for informed business decisions.

Cluster	Total Spend	Total Transactions	Avg. Transaction Value	Regions
0	High	High	Moderate	Predominantly Asia
1	Medium	Medium	High	North America
2	Low	Low	Low	South America
3	High	Low	High	Europe

Cluster Visualization:

- The dataset was reduced to two dimensions using PCA (Principal Component Analysis) for visualization. PCA was used because it simplifies high-dimensional data while preserving its essential structure, making it easier to identify and visualize patterns or separations between clusters.
- A scatter plot was created, showing clear separations between clusters.

Insights and Recommendations:

To effectively address customer needs and improve business strategies, the following recommendations are proposed based on identified customer behaviors:

- **Premium Product Promotions and Loyalty Programs:** High spenders with frequent transactions (Cluster 0, predominantly in Asia) can be targeted with premium product promotions and loyalty programs.
- **Cross-Selling Opportunities:** Moderate spenders with higher transaction values (Cluster 1, primarily in North America) offer an opportunity for cross-selling strategies.

- **Re-Engagement Campaigns:** Low-value customers with infrequent activity (Cluster 2, concentrated in South America) could benefit from re-engagement efforts to boost activity.
- **Seasonal or High-Value Offerings:** High spenders with infrequent transactions (Cluster 3, mostly in Europe) may respond well to targeted seasonal or high-value product offerings.
    - High spenders with frequent transactions in Asia.
    - Ideal for premium product promotions and loyalty programs.

2. **Cluster 1**:
    - Moderate spenders with higher transaction values in North America.
    - Focus on cross-selling opportunities.
3. **Cluster 2**:
    - Low-value customers in South America.
    - Re-engagement campaigns could help increase activity.
4. **Cluster 3**:
    - High spenders but infrequent transactions in Europe.
    - Target with seasonal or high-value product offerings.

Future Work:

1. Incorporate product-level data to refine segmentation.
2. Compare with other clustering algorithms (e.g., DBSCAN, Hierarchical Clustering).
3. Extend the analysis to predict customer lifetime value.

Key Metrics:

- Optimal Clusters: **4**
- Best Davies-Bouldin Index: **0.5824**

This segmentation framework can guide data-driven marketing strategies, boosting customer engagement and profitability.