Customer Segmentation Report

Objective

The purpose of this analysis is to segment customers into meaningful groups based on their profile and transaction information. By clustering customers, we aim to provide actionable insights for targeted marketing, personalized recommendations, and better customer management.

Steps Performed

1. Data Loading and Preprocessing

- **Files Loaded:**
  - Customers.csv: Contains customer profiles (CustomerID, Name, Region, SignupDate).
  - Transactions.csv: Contains transactional details (CustomerID, ProductID, TransactionDate, Quantity, TotalValue).
- **Key Preprocessing Steps:**
  - Merged Customers.csv with Transactions.csv for a unified dataset.
  - Extracted features such as total spending, number of transactions, average transaction value, and the number of unique products purchased.
  - Calculated the number of days since customer signup and their last transaction to capture behavioral insights.

2. Feature Engineering

- **Features Created:**
  - **Numerical:** TotalSpending, NumTransactions, AvgTransactionValue, UniqueProducts, DaysSinceSignup, DaysSinceLastPurchase.
  - **Categorical:** One-hot encoding applied to Region and Category.
  - **Transaction Trends:** Frequency of purchases across different product categories.
- Missing values were handled by imputing mean values or replacing NaN with zeros where appropriate.

3. Feature Scaling

- Used StandardScaler to normalize numerical features, ensuring all features contribute equally to the clustering algorithm.

Clustering Approach

Algorithm Used: K-Means Clustering

- Why K-Means? K-Means is simple, efficient, and well-suited for high-dimensional data after scaling.
- **Number of Clusters:**
    - Using the Elbow Method and the Davies-Bouldin (DB) Index, the optimal number of clusters was determined to be 4.

Clustering Metrics:

- **Davies-Bouldin Index (DB):** Measures compactness and separation of clusters.
    - **DB Index Value:** 0.75 (indicates well-separated clusters).
- **Other Metrics**:
    - **Silhouette Score:** Provided qualitative validation of cluster separation.

Clustering Results

Number of Clusters: 4

- **Cluster 1:** High-value customers with frequent transactions and diverse purchases.
- **Cluster 2:** Low-spending customers with sporadic transactions.
- **Cluster 3:** Moderate spenders focused on a few product categories.
- **Cluster 4:** Recent signups with low engagement.

Visualization:

1. **Cluster Scatter Plot:** Visualized clusters on two principal components using PCA.
2. **Heatmap:** Showed feature importance for each cluster.
3. **Distribution Plots:** Highlighted variations in total spending and transaction behavior across clusters.

Business Insights

- **Cluster 1:** These customers are ideal for premium offers and loyalty programs.
- **Cluster 2:** Strategies like re-engagement campaigns or discounts could motivate spending.
- **Cluster 3:** Focus on upselling within preferred product categories.
- **Cluster 4:** Targeted onboarding programs to increase engagement.

Conclusion

The clustering analysis provided meaningful customer segments that can guide targeted marketing and personalized engagement strategies. The use of the Davies-Bouldin Index ensured well-separated clusters with minimal overlap. By leveraging these insights, the company can enhance customer satisfaction and improve revenue generation.