Customer Segmentation Report

Objective

The goal of this analysis was to segment customers based on their transaction history and demographic profile using clustering techniques. We aimed to identify distinct customer groups for personalized marketing and targeted strategies.

Dataset Overview

The analysis utilized data from two sources:

1. **Customers.csv**: Included customer IDs, regions, and signup dates.
2. **Transactions.csv**: Contained transaction details such as total value, quantity purchased, and transaction dates.

The datasets were merged on CustomerID to create a comprehensive view of customer profiles and their transaction behaviors. Key features such as total spend (TotalValue), purchase frequency, and the total quantity of items bought were used for clustering.

Clustering Methodology

1. **Feature Engineering**:
    - Aggregated transaction data to calculate key metrics: total spend, total quantity purchased, and transaction frequency.
    - Encoded the categorical Region column into numerical form using one-hot encoding.
    - Standardized numerical features (TotalValue, Quantity, and TransactionID) using StandardScaler to ensure equal weight across features.
2. **Clustering Algorithm**:
    - Used **KMeans Clustering** with cluster counts ranging from 2 to 10.
    - Evaluated the quality of clusters using the **Davies-Bouldin Index (DB Index)** and **Silhouette Score** for each cluster count.
3. **Dimensionality Reduction**:
    - Applied **Principal Component Analysis (PCA)** to reduce the feature space to two dimensions for visualization.

Clustering Results

- **Optimal Number of Clusters**: Based on the DB Index, the optimal number of clusters was determined to be **2**. This result indicates that the dataset exhibits two distinct groups of customers.

- **Davies-Bouldin Index**:
    - Measures intra-cluster similarity and inter-cluster separability. A lower DB Index indicates better cluster separation.
    - The lowest DB Index was **0.46** at 2 clusters, demonstrating strong cluster quality.
- **Silhouette Score**:
    - Measures how similar an object is to its own cluster compared to others. Values range from -1 to 1, where higher values indicate better clustering.
    - For 2 clusters, the Silhouette Score was **0.52**, which confirms reasonably well-separated clusters.

Insights from Clustering

1. **Customer Segmentation**:
    - **Cluster 1** represents high-value customers with frequent purchases and larger transaction amounts. These customers are likely your most loyal and profitable segment.
    - **Cluster 2** includes lower-value customers who make infrequent or smaller purchases, suggesting opportunities for engagement or retention strategies.
2. **Geographic Trends**:
    - The regional encoding revealed that certain clusters have a strong geographic concentration. Marketing efforts can be tailored regionally to maximize effectiveness.
3. **Behavioral Patterns**:
    - High-value customers (Cluster 1) demonstrated a higher average spend and transaction frequency, indicating their preference for specific product categories or brands.
4. **Targeted Strategies**:
    - Cluster 1 can be targeted with premium offerings, loyalty programs, and personalized recommendations.
    - Cluster 2 might benefit from discounts, promotions, or awareness campaigns to increase engagement.
5. **Visualization**:
    - The PCA visualization shows clear separation between the two clusters, validating the clustering results. This indicates distinct customer behaviors and profiles that align well with the segmentation.

Recommendations

1. **Personalization**:
    - Implement personalized email campaigns and promotions based on cluster characteristics.
    - Leverage transaction data to recommend relevant products to high-value customers.
2. **Retention Efforts**:
    - Focus retention efforts on Cluster 2 by introducing loyalty incentives, referral programs, or targeted communication.
3. **Marketing Budget Allocation**:
    - Allocate more resources toward Cluster 1 customers to maximize return on investment while using engagement strategies to uplift Cluster 2.

Conclusion

The clustering analysis successfully segmented customers into two meaningful groups. The insights can be leveraged to drive customer engagement, retention, and revenue growth through data-driven strategies. By focusing on the distinct needs and behaviors of each cluster, the business can enhance customer satisfaction and profitability.