**Customer Segmentation Report**

Objective

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

---

Dataset Overview

The analysis used 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 provide a complete view of customer profiles and their transaction behaviors. Total spend, purchase frequency, and the total quantity of items bought were some of the key features used for clustering.

---

**Clustering Methodology**

**1.Feature Engineering:**

**Combined the transaction data by aggregating total spend, total quantity bought, and frequency of transactions.**

**Converted the categorical Region column into numerical form through one-hot encoding.**

**Standardized numerical features (TotalValue, Quantity, and TransactionID) with StandardScaler to provide equal weight to features.**

**2.Clustering Algorithm:**

**Applied KMeans Clustering with cluster numbers between 2 and 10.**

**Assessed the cluster quality based on Davies-Bouldin Index (DB Index) and Silhouette Score for every number of clusters.**

**3.Dimensionality Reduction:**

**Used Principal Component Analysis (PCA) to project the feature space to two dimensions for visualization.**

---

Clustering Results

Ideal Number of Clusters: From the DB Index, the ideal number of clusters was found to be 2. This finding suggests that the dataset has two clear groups of customers.

•Davies-Bouldin Index:

Measures intra-cluster similarity and inter-cluster separability. A smaller DB Index is evidence of good cluster separation.

The lowest DB Index was 0.46 at 2 clusters, demonstrating strong cluster quality.

Silhouette Score:

Quantifies how close an object is to its own cluster versus others. The value ranges between -1 and 1, with high value suggesting good clustering.

For 2 clusters, Silhouette Score was 0.52, which indicates fairly 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.