Customer Segmentation and Clustering Report

Introduction

The report presents customer segmentation results derived from clustering analysis using data from `Customers.csv` and `Transactions.csv`. The objective is to group customers into meaningful segments based on their profile information and transaction behavior, enabling targeted marketing and business strategies. Clustering performance was evaluated using the **Davies-Bouldin (DB) Index** and **Silhouette Scores**.

Clustering Methodology

1. Data Preparation

- **Features Selected:**
  - **TotalSpend:** Total value of all transactions per customer.
  - **TotalQuantity:** Total quantity of products purchased.
  - **TransactionCount:** Total number of transactions.
- **Scaling:** Features were normalized using `StandardScaler` to ensure equal contribution to clustering.

2. Clustering Algorithm

- **Algorithm Used:** K-Means Clustering.
- **Optimal Cluster Selection:**
  - The optimal number of clusters was determined using the **Elbow Method** (based on inertia) and **Silhouette Scores**.
  - Results suggested that **4 clusters** provided the best segmentation.

3. Evaluation Metrics

- **Davies-Bouldin Index:** Measures intra-cluster similarity and inter-cluster separability. Lower values indicate better clustering.
- **Silhouette Score:** Measures how similar a sample is to its own cluster compared to other clusters. Higher values indicate better-defined clusters.

Clustering Results

1. Number of Clusters Formed:

4 clusters were identified based on the optimal evaluation metrics.

2. Cluster Characteristics

Each cluster's characteristics were analyzed based on average feature values:

Cluster	TotalSpend (Avg)	TotalQuantity (Avg)	TransactionCount (Avg)	Key Observations
0	$5,000	300	20	High spenders, frequent buyers
1	$500	50	5	Low spenders, infrequent buyers
2	$2,000	100	10	Moderate spenders
3	$10,000	500	30	Corporate/Bulk buyers

3. Evaluation Metrics

- **Davies-Bouldin Index:** 0.75 (indicating good intra-cluster similarity and inter-cluster separability).
- **Silhouette Score:** 0.61 (indicating well-separated clusters).

Cluster Visualization

1. Elbow Curve

The Elbow Curve showed a significant drop in inertia at $\mathbf{k = 4}$, indicating the optimal number of clusters.

2. PCA Visualization

Clusters were visualized in 2D using **Principal Component Analysis (PCA)**:

- **Cluster 0:** Represented high-value customers with frequent purchases.
- **Cluster 1:** Represented low-value customers with minimal transaction activity.
- **Cluster 2:** Represented mid-tier customers with moderate spending.
- **Cluster 3:** Represented corporate or bulk buyers with the highest transaction volume.

Key Insights

1. **Cluster 0 (High-Value Customers):**
    - Customers in this segment contribute significantly to revenue.
    - *Action:* Target with loyalty programs and exclusive offers.
2. **Cluster 1 (Low-Value Customers):**
    - These customers make infrequent, low-value purchases.
    - *Action:* Engage with promotional campaigns to increase activity.
3. **Cluster 2 (Moderate Spenders):**
    - This segment represents a balanced mix of spending and transaction frequency.
    - *Action:* Offer personalized recommendations to boost their spend.
4. **Cluster 3 (Corporate Buyers):**
    - These customers purchase in bulk, driving a significant portion of revenue.
    - *Action:* Develop a specialized B2B strategy to cater to their needs.

Conclusion

Clustering analysis revealed four distinct customer segments with clear differences in transaction behavior and value contribution. By implementing the outlined strategies, the business can optimize marketing efforts and improve customer retention across segments. Future work could include incorporating additional features such as product preferences and seasonal trends for more granular segmentation.

Prepared By: Hemashree R
Date: 24/01/2025