

Customer Segmentation Report

Objective

The objective was to segment customers using clustering techniques based on combined profile information (Customers.csv) and transaction data (Transactions.csv). The segmentation helps in identifying meaningful groups to optimize customer engagement strategies.

Approach

1. Data Preprocessing:

- Merged Customers.csv and Transactions.csv datasets.
- Key features: Recency (days since signup), Total Spending, and Frequency (number of transactions).
- Scaled all numerical features using MinMaxScaler for uniformity.

2. Clustering Algorithm:

- **Algorithm Used:** K-Means Clustering.
- **Optimal Number of Clusters:** 4 (determined using Davies-Bouldin Index and Silhouette Score).

3. Metrics:

- **Final DB Index:** 0.79 (lower value indicates better cluster separation).
- **Final Silhouette Score:** 0.42 (indicating moderate cohesion and separation).

4. Visualization:

- A 2D scatter plot of clusters was created using Recency and Total Spending (scaled). The clusters are color-coded, clearly showing distinct groupings.

Results and Insights

1. Cluster 0 (Low-Spending, Inactive Customers):

- Customers with low spending and high recency values.
- **Action:** Focus on re-engagement through offers or loyalty incentives.

2. Cluster 1 (Low-Spending, Recent Customers):

- Customers with moderate recency and low spending.
- **Action:** Strengthen onboarding and engagement strategies.

3. **Cluster 2 (Moderately Engaged Customers):**

- Customers with moderate recency and spending levels.
- **Action:** Offer loyalty rewards and cross-sell related products or services.

4. **Cluster 3 (High-Spending, New Customers):**

- High-value customers with low recency values (recent signups).
- **Action:** Prioritize retention strategies with exclusive benefits or personalized offers.

Deliverables

1. **Clustering Metrics:**

- **DB Index:** 0.79 (indicates reasonably good cluster separation).
- **Silhouette Score:** 0.42 (moderate cluster cohesion).

2. **Visualizations:**

- A scatter plot visually demonstrates well-separated clusters based on customer behavior.

3. **Code:**

- A Python script is provided, containing preprocessing, clustering logic, evaluation metrics, and visualization steps.

Conclusion

The clustering segmented customers into four distinct groups, providing actionable insights to tailor customer strategies. While the Silhouette Score of 0.42 reflects moderate cluster cohesion, the DB Index of 0.79 confirms reasonable separation, supporting the effectiveness of the clustering approach.