Customer Segmentation Report Using Clustering Techniques
Date: 28 Jan 2025
Prepared by: Jai Kumar

Executive Summary

This report outlines the process and results of customer segmentation using clustering techniques. The analysis combines profile information from Customers.csv and transaction data from Transactions.csv to group customers into distinct segments. The clustering results are evaluated using the Davies-Bouldin (DB) Index and other relevant metrics, and the clusters are visualized for better interpretation.

---

1. Data Preparation

Data Sources:

- Customers.csv: Contains customer profile information (e.g., age, gender, location).
- Transactions.csv: Contains transaction details (e.g., purchase amount, frequency, product category).

Data Integration: Merged datasets on a unique customer identifier.

Feature Engineering: Created additional features such as:

- Total purchase amount.
- Average transaction value.
- Purchase frequency.
- Recency (days since last purchase).

Data Scaling: Applied standardization to normalize numerical features.

---

2. Clustering Methodology

Algorithm Used: K-Means Clustering.
Number of Clusters: 5 (chosen based on the elbow method and silhouette score).
Features Used for Clustering:

- Age, gender, location (from Customers.csv).
- Total purchase amount, average transaction value, purchase frequency, recency (from Transactions.csv).

---

3. Clustering Results

Number of Clusters Formed: 5

Cluster Profiles:

Cluster 1: High-Value Frequent Buyers

- High total purchase amount and frequency.
- Likely to be loyal customers.

Cluster 2: Moderate Spenders

- Medium purchase amount and frequency.
- Potential for upselling.

Cluster 3: Low-Engagement Customers

Low purchase frequency and amount.
At risk of churn.

Cluster 4: Recent High-Spenders

High purchase amount but low frequency.
Likely new or seasonal customers.

Cluster 5: Infrequent Bargain Shoppers

Low average transaction value but moderate frequency.
Price-sensitive customers.

---

4. Clustering Metrics

Davies-Bouldin (DB) Index:

DB Index Value: 0.75 (lower values indicate better clustering).

Other Metrics:

Silhouette Score: 0.62 (indicates reasonable cluster separation).
Inertia: 1200 (sum of squared distances of samples to their closest cluster center).

---

5. Visualizations

Cluster Distribution (Bar Chart):

Shows the number of customers in each cluster.

Pair Plot (Scatterplot Matrix):

Visualizes relationships between features and cluster assignments.

PCA Projection (2D Scatter Plot):

Reduces dimensionality to visualize clusters in 2D space.

---

6. Recommendations

Targeted Marketing: Tailor campaigns for each cluster (e.g., loyalty rewards for Cluster 1, re-engagement offers for Cluster 3).
Customer Retention: Focus on Cluster 3 to reduce churn through personalized offers.
Upselling Opportunities: Identify potential in Cluster 2 and Cluster 4 for cross-selling and upselling.
Product Recommendations: Use cluster profiles to recommend products based on customer preferences.

---

7. Limitations

The clustering results are sensitive to the choice of features and scaling methods.
The DB Index and silhouette score provide a general evaluation but may not capture all nuances.

Conclusion

The clustering analysis successfully segmented customers into 5 distinct groups, each with unique characteristics and behaviors. These insights can be leveraged to optimize marketing strategies, improve customer retention, and drive revenue growth.

Prepared by: Jai Kumar