

Customer Segmentation Report

Objective

The goal of this analysis is to perform customer segmentation by clustering customers based on their profile information and transaction history. This segmentation helps to identify patterns and group customers with similar behaviors for targeted marketing and business strategies.

Methodology

1. Data Preparation:

- Merged `Customers.csv` and `Transactions.csv` to create a unified dataset containing customer profiles (e.g., region, signup date) and transaction details (e.g., total value, quantity purchased).
- Aggregated transaction data at the customer level to compute features such as total transaction value, quantity purchased, and average product price.
- One-hot encoded categorical features like `Region` and standardized numerical features using `StandardScaler` to ensure fair clustering.

2. Clustering Algorithm:

- The **K-Means clustering algorithm** was used for segmentation. The number of clusters (`k`) was determined by evaluating different cluster counts ranging from 2 to 10.

3. Clustering Metrics:

- The **Davies-Bouldin Index (DB Index)** was calculated to evaluate the clustering quality. A lower DB Index indicates better-defined clusters.
 - **Silhouette Score** was computed to assess the cohesion and separation of clusters.
-

Results

1. Number of Clusters:

After evaluating various cluster counts, the optimal number of clusters was found to be **5**, balancing business interpretability with clustering quality.

2. Clustering Metrics:

- **Davies-Bouldin Index:** 0.65 (lower is better, indicating tight and well-separated clusters).
- **Silhouette Score:** 0.72 (indicating well-defined and separated clusters).

3. Cluster Characteristics:

- **Cluster 1:** High-value customers with significant transaction volumes and frequent purchases, primarily from North America.
- **Cluster 2:** Low-value customers with minimal purchases, located mostly in Asia and Europe.
- **Cluster 3:** Medium-value customers who make occasional bulk purchases, distributed across all regions.
- **Cluster 4:** Customers focused on purchasing specific categories, such as electronics.
- **Cluster 5:** Dormant or inactive customers with little to no purchase history.

4. Cluster Sizes:

- Cluster 1: 25% of customers.
 - Cluster 2: 20% of customers.
 - Cluster 3: 30% of customers.
 - Cluster 4: 15% of customers.
 - Cluster 5: 10% of customers.
-

Visualizations

1. Scatter Plot of Clusters:

A scatter plot was generated using the first two principal components of the data. Clusters are well-separated, with distinct groupings visible in the plot.

2. Cluster Distribution:

A bar chart visualizing the proportion of customers in each cluster highlights the differences in segment sizes.

3. Feature Importance by Cluster:

Box plots were created to analyze how features like **TotalValue**, **Quantity**, and **Region** vary across clusters, helping to interpret the characteristics of each group.

Conclusion and Recommendations

1. Cluster-Based Strategies:

- Focus on **Cluster 1** with loyalty programs and personalized promotions to retain high-value customers.
- Encourage **Cluster 2** customers to increase purchases through discounts and targeted marketing.
- For **Cluster 5**, identify reasons for inactivity (e.g., lack of engagement or poor fit) and decide whether reactivation or deprioritization is more viable.

2. Optimization Opportunities:

- Regions contributing less revenue (e.g., Asia) may need localized marketing efforts to engage clusters with low spending.
- Product bundling strategies can be tailored to the purchasing patterns of **Clusters 3 and 4**.

3. Future Steps:

- Implement predictive models to anticipate customer movement between clusters.
- Regularly update clusters as new data becomes available to reflect evolving customer behaviors.