

Overview

This document outlines the findings of a customer segmentation analysis conducted using clustering techniques on an eCommerce transaction dataset. The primary objective was to group customers into distinct segments based on their purchasing patterns, enabling the development of tailored marketing strategies and business insights.

Data Preparation

1. **Feature Aggregation:** Transactional data was aggregated to create meaningful features:
 - TotalValue: Total monetary expenditure.
 - Quantity: Total quantity of items purchased.
 - TransactionID count: Number of transactions conducted.
 - Unique ProductID count: Variety of products purchased by each customer.
2. **Normalization:** All features were normalized using the **MinMaxScaler** to ensure comparability across variables.
3. **Dimensionality Reduction:** Data dimensionality was reduced to two components using **Principal Component Analysis (PCA)** for enhanced visualization and computational efficiency.

Clustering Techniques

1. KMeans Clustering

- The optimal number of clusters (K) was determined by evaluating **Davies-Bouldin Index (DB Index)** and **Silhouette Score** for K values ranging from 2 to 10.
- **Outcome:** Three clusters were deemed optimal.

2. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

- Parameters used:
 - eps: 0.5
 - min_samples: 5
- Identified clusters of dense data points and flagged outliers as noise.

Evaluation Metrics

1. **Davies-Bouldin Index (DB Index):**
 - Measures the compactness and separation of clusters.
 - **Result: 0.6887**, suggesting well-separated and compact clusters.
2. **Silhouette Score:**
 - Assesses how well each point aligns with its assigned cluster versus neighboring clusters.
 - **Result: 0.5039**, indicating moderately strong clustering.

Key Findings

1. **Optimal Clusters Identified:** Using KMeans, three distinct customer groups were revealed:
 - **Cluster 1:** High-spending, frequent shoppers.
 - **Cluster 2:** Moderate spenders with average transaction frequency.
 - **Cluster 3:** Low spenders or infrequent buyers.

2. DBSCAN Observations:

- Identified outliers and potential niche customer behaviors.

Visualizations

1. Cluster Scatterplot:

- A 2D visualization of KMeans clusters post-PCA transformation, showing clear distinctions between groups.

2. Elbow Method Plot:

- Highlights the reduction in **Sum of Squared Errors (SSE)** with an increasing number of clusters, affirming the choice of $K = 3$.

Recommendations

Marketing Strategies

1. **Cluster 1:** Prioritize these high-value customers with premium offerings, personalized rewards, and loyalty incentives.
2. **Cluster 2:** Develop targeted campaigns to encourage greater spending and foster loyalty.
3. **Cluster 3:** Implement awareness campaigns and promotions aimed at increasing engagement and purchase frequency.

Further Analysis

- Examine the noise points identified by DBSCAN to uncover unique customer segments or niche behaviors.
- Incorporate additional features such as transaction recency and preferred product categories to refine clustering outcomes.

Model Refinement

- Explore alternative clustering methodologies, such as **Gaussian Mixture Models (GMM)** or **Hierarchical Clustering**, for enhanced insights.
- Fine-tune DBSCAN parameters (eps and min_samples) to achieve improved density-based clustering.

Conclusion

This analysis successfully segmented the customer base into three distinct groups, providing actionable insights for targeted marketing and strategic decision-making. Future iterations of this study can incorporate more granular data and advanced modeling techniques to further improve segmentation quality.

Metrics Summary

- **Optimal Cluster Count (K): 3**
- **Davies-Bouldin Index: 0.6887**
- **Silhouette Score: 0.5039**