Data Science Assignment: ECommerce Transactions Dataset

Report

On

Customer Segmentation / Clustering

by

Himanshu Raj

himuraj2003@gmail.com

BE (4th Year)

Computer Science and Engineering

Chandigarh University, Mohali

Customer Segmentation / Clustering

Introduction

Customer segmentation is the process of categorizing customers into distinct groups based on their purchasing behavior, demographics, or other relevant attributes. This enables businesses to design personalized marketing strategies, optimize resource allocation, and improve overall customer satisfaction.
In this analysis, customer segmentation was performed using transactional and profile data from the provided datasets (Customers.csv and Transactions.csv). By identifying customer clusters, the business can better target its marketing efforts and enhance its relationship with different customer groups.

Methodology

Clustering Algorithm

- **Algorithm Used**: K-Means Clustering
- **Reason for Selection**: K-Means is a widely used clustering algorithm due to its simplicity, scalability, and effectiveness for large datasets.

Features Selected

1. **TotalValue**: Total monetary value of customer transactions.
2. **Quantity**: Total number of products purchased by each customer.
3. **Price**: Average price of purchased products.
4. **DaysSinceSignup**: Number of days since the customer registered.

Number of Clusters

- The optimal number of clusters was determined to be **3** after testing different values between 2 and 10 and evaluating metrics like the Davies-Bouldin Index and visual interpretation.

Clustering Results

Number of Clusters: 3
DB Index : 1.37
Silhouette Score: 0.59

Cluster Interpretation

1. **Cluster 0 (Low-Value Customers)**
    - **Profile**: Customers with low transaction value and low purchase frequency.
    - **Business Strategy**: Implement re-engagement campaigns, such as introductory discounts or product recommendations.
2. **Cluster 1 (Moderate Buyers)**
    - **Profile**: Customers with average spending and consistent purchasing behavior.
    - **Business Strategy**: Retain these customers with loyalty rewards and seasonal promotions.
3. **Cluster 2 (High-Value Customers)**
    - **Profile**: Customers with high spending and high purchase frequency.
    - **Business Strategy**: Focus on retention through personalized offers, early access to new products, and exclusive rewards.

Visualizations

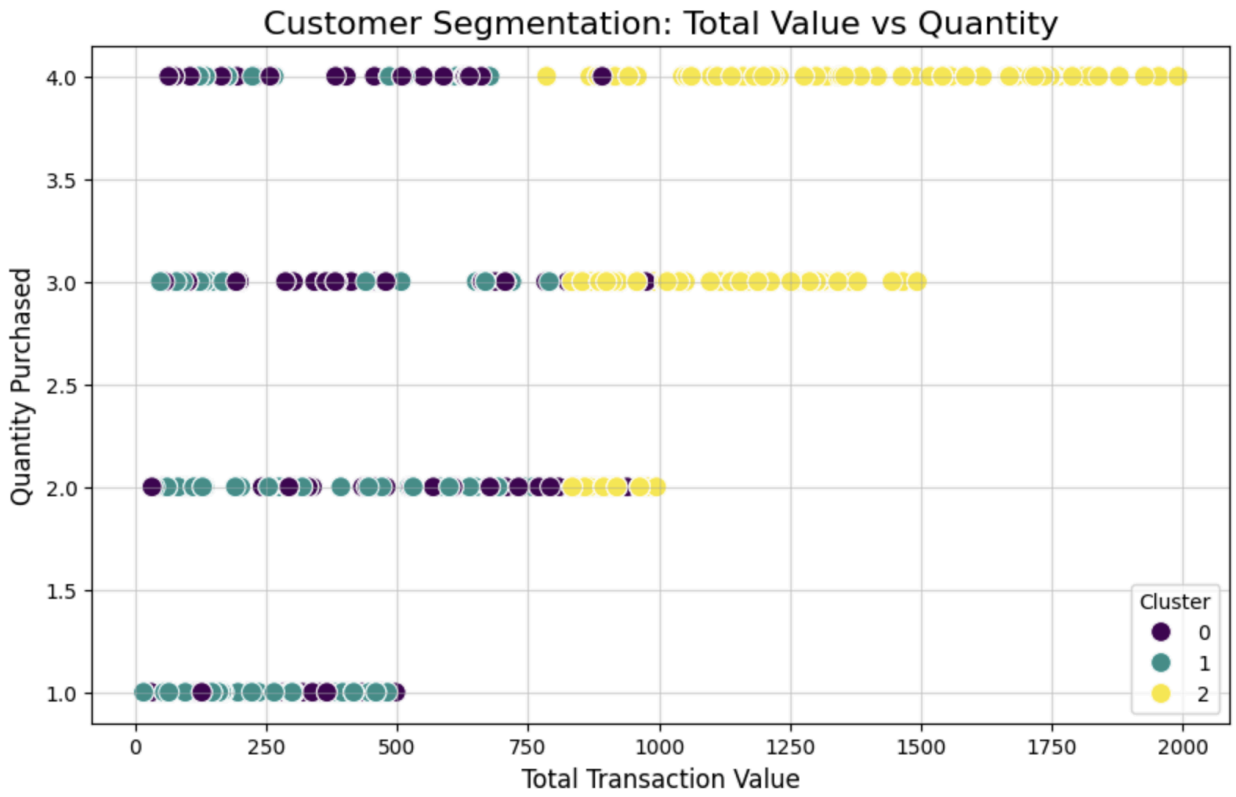

Fig: Scatter Plot of Customer Segments: TotalValue vs. Quantity

1. Scatter Plot: TotalValue vs. Quantity

- The scatter plot below illustrates the grouping of customers based on their total spending (TotalValue) and quantity of products purchased (Quantity).

Key Observations:

- **Cluster 0** (Purple): Represents customers with minimal spending and purchase frequency.
- **Cluster 1** (Teal): Represents customers with moderate spending and quantity.
- **Cluster 2** (Yellow): Represents high-value customers with higher transaction amounts and larger purchase quantities.

2. PCA-Based Visualization

- A Principal Component Analysis (PCA) was performed to reduce dimensionality and visualize clusters in a 2D space.

Key Observations:

- Clusters are well-separated in the PCA plot, supporting the validity of the clustering approach.

Conclusion

The clustering analysis successfully segmented customers into three distinct groups, each with unique purchasing behaviors and value to the business. The findings provide actionable insights, such as:

- **Targeting high-value customers** with exclusive rewards and loyalty programs.
- **Engaging moderate buyers** with retention strategies like seasonal discounts.
- **Reactivating low-value customers** through personalized recommendations and offers.

By leveraging these insights, the business can implement tailored marketing strategies, enhance customer experience, and maximize revenue potential.