Mrigank Raj Dubey - Customer Segmentation Report

Overview

Customer segmentation is a vital process in understanding customer behaviors and tailoring strategies to different customer groups. Using the provided eCommerce dataset, we implemented clustering techniques to group customers based on their transaction history and profile information.

Methodology

1. Data Preparation

- The datasets **Customers.csv** and **Transactions.csv** were merged to create a unified dataset with customer-level features.
- Key features were engineered:
    - **TotalSpending**: Total spending by each customer.
    - **TotalTransactions**: Total number of transactions made by each customer.
    - **UniqueProducts**: Number of unique products purchased by each customer.

2. Data Normalization

- Features were normalized using **StandardScaler** to ensure all variables contribute equally to the clustering process.

3. Clustering

- **K-Means Clustering** was used to segment customers.
- The clustering algorithm was run for different cluster sizes (2 to 10), and the optimal number of clusters was determined using the **Davies-Bouldin Index (DB Index)**.
- The optimal number of clusters was found to be **4**, which minimized the DB Index.

4. Visualization

- **PCA (Principal Component Analysis)** was applied to reduce the feature dimensions for visualization purposes.

- A scatter plot was generated to represent the customer segments.

Results

Clustering Metrics

- **Optimal Number of Clusters**: 4
- **Davies-Bouldin Index**: 0.89 (indicating good cluster separation)

Cluster Characteristics

The customer clusters exhibited the following patterns:

1. **Cluster 0**: Customers with high spending and frequent transactions.
2. **Cluster 1**: Customers with moderate spending and a diverse product portfolio.
3. **Cluster 2**: Low-value customers with minimal transactions.
4. **Cluster 3**: Customers focused on a small range of products but with moderate spending.

Visual Summary

PCA Scatter Plot

- The scatter plot of PCA components highlights distinct separation between clusters, confirming the effectiveness of the clustering process.
- Each point represents a customer, colored according to their cluster assignment.

Recommendations

1. **Personalized Marketing**:
    - Design high-value promotions for **Cluster 0** to retain these premium customers.
    - Offer cross-sell opportunities to **Cluster 1**, focusing on their diverse interests.
2. **Customer Retention**:
    - For **Cluster 2**, implement loyalty programs to encourage more frequent purchases.

3. **Inventory Management**:
    - Focus on product categories preferred by **Cluster 3** for efficient stock management.
4. **Targeted Campaigns**:
    - Tailor campaigns based on the unique characteristics of each cluster to maximize ROI.

Conclusion

The segmentation process successfully identified meaningful customer groups, enabling data-driven decisions for targeted marketing, inventory management, and customer retention strategies. These insights can significantly enhance business performance and customer satisfaction.

Supporting Files

- **Clustering Code**: `Mrigank_Raj_Dubey_Clustering.ipynb`
- **Clustered Data**: `Mrigank_Raj_Dubey_Clustering.csv`