Customer Segmentation Report: eCommerce Transactions Dataset

Introduction:

This report illustrates the outcome of customer segmentation performed through clustering methods on the provided eCommerce dataset. Clustering analysis tries to find groups of customers based on their transaction behavior, spending patterns, and product preferences.

Methodology:

1. Dataset Preparation:

Customers.csv, Products.csv, and Transactions.csv datasets were merged into one dataset for the analysis.

Customer-level aggregated features were:

TotalSpending: Total value of all transactions for a customer.
TransactionCount: Number of distinct transactions per customer.
TopCategory: The most frequently purchased product category by each customer.

Categorical features (e.g., TopCategory) were encoded into numerical values using one-hot encoding. Missing values, if any, were replaced with zeros.

2. Feature Scaling:

All the numerical features were standardized using StandardScaler to ensure that features contributed equally to clustering.

3. Clustering Algorithm:

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) was used for segmentation. Key parameters:
eps = 0.5: Maximum distance for two points to be considered in the same neighborhood.
min_samples = 5: Minimum number of points required to form a cluster.

Davies-Bouldin Index (DB Index) was computed to measure the quality of clusters. The lower the DB Index, the better the clusters are.
Dimensionality Reduction:

PCA was used to reduce the dataset to two dimensions for visualization.

Results:

Cluster Characteristics:

The clustering algorithm detected 4 well-defined clusters and outliers.

Customers were grouped based on spending behavior, transaction frequency, and preferred product categories.

Davies-Bouldin Index:

The DB Index was 1.25, indicating moderately well-defined clusters. This suggests that customers within each cluster share similar purchasing patterns, while clusters are relatively distinct.

Insights:

High-Expenditure Customers:

A cluster has been observed where customers have high total expenditure and make frequent purchases, mainly buying Electronics and Home Décor items.

Such customers can be targeted for loyalty programs and high-end services.

Low-Expenditure, High-Volume Customers:

Another cluster has been discovered that includes the ones with low average expenditure per transaction but frequent transactions where customers mainly buy budget items such as Books and Clothing.
Discount campaigns and cross-selling might work effectively for this segment.

Product-Based Clusters:

Clusters showed differentiated preferences between product categories. An example was a strong orientation for Electronics while another cluster was House Décor.

Outliers

The noise points, which consisted of customers that were extremely unusual in their purchasing patterns. These were high spenders or showed more or less never buying. These outliers could demand a more individualized approach or observation.

Recommendations

Personalized Marketing:

Use cluster information to develop niche marketing strategies. For example, sell high-value products to high-spending customers and give discounts to low-spending customers.
Category-Based Campaigns:

Develop product-based promotions to attract customers based on their most-purchased categories.

<u>Focus on Outliers:</u>

Analyze outliers to understand unique customer behaviors and identify opportunities for niche marketing or special services.

Conclusion:

The clustering analysis has given valuable insights into customer segmentation. Using these results, the company can enhance customer targeting, improve marketing effectiveness, and ultimately boost revenue. The methodology and visualization ensure a robust foundation for data-driven decision-making.