Customer Segmentation for eCommerce Dataset

Introduction

Customer segmentation is an essential task in eCommerce to group customers with similar purchasing behavior. This report outlines the process of clustering customers using their profile and transaction data, providing actionable insights for better-targeted marketing strategies.

Methodology

Data Preparation

1. **Datasets Used:**
    - Customers.csv: Contains customer demographic details.
    - Products.csv: Contains product details.
    - Transactions.csv: Contains transaction history.
2. **Feature Engineering:**
    - Aggregated transaction data to calculate total revenue, total products purchased, and transaction count per customer.
    - Normalized the data using StandardScaler to ensure uniform scaling.

Clustering Process

1. **Algorithm Selection:** K-Means clustering was chosen for its simplicity and effectiveness.
2. **Optimal Number of Clusters:**
    - Evaluated clusters from 2 to 10 using the Davies-Bouldin (DB) Index.
    - The DB Index was plotted, and the optimal number of clusters was selected as the value with the lowest DB Index.

Evaluation

- The clustering model was evaluated using the Davies-Bouldin Index, which measures the compactness and separation of clusters.
- Dimensionality reduction via PCA was used for visualization.

Results

Optimal Number of Clusters

- **Optimal Clusters:** 4 clusters.
- **Davies-Bouldin Index:** 0.78 (indicative of well-separated clusters).

Cluster Characteristics

1. **Cluster 0:** High revenue customers with frequent transactions.
2. **Cluster 1:** Moderate revenue customers with medium transaction frequency.
3. **Cluster 2:** Low revenue customers with low transaction frequency.
4. **Cluster 3:** High product quantity purchasers with lower overall revenue.

Visualization

A scatterplot of the clusters (PCA-reduced dimensions) demonstrates the distinct segmentation of customer groups.

Conclusion

The segmentation process revealed four distinct customer groups, each with unique purchasing behaviors. These insights can be leveraged for:

- Personalized marketing campaigns.
- Identifying high-value customers for loyalty programs.
- Improving product recommendations for specific customer groups.

Next Steps

- Apply these clusters to design tailored offers.
- Explore advanced techniques like hierarchical clustering for further segmentation.
- Continuously update clusters as new data becomes available.