Data Science Assignment: eCommerce Transactions Dataset

Task 3: Customer Segmentation / Clustering

Objective

The objective is to segment customers into distinct groups using clustering techniques. These segments will help the business understand customer behavior and enable targeted marketing strategies.

Approach and Methodology

1. Data Preparation

Datasets Used:

- Customers.csv: Contains customer demographics and signup information.

- Transactions.csv: Contains purchase history, including total transaction values and product preferences.

Steps:

- Merge Customers.csv with aggregated transaction data using CustomerID.

- Aggregate features such as total transaction value, frequency of purchases, and preferred product categories.

- Normalize numerical features to ensure uniform scaling across dimensions.

2. Clustering Technique

Algorithm: KMeans clustering.

Range of Clusters: Experimented with 2 to 10 clusters to identify the optimal number of groups. Evaluation

Metrics:

- Davies-Bouldin Index (DB Index): Measures cluster compactness and separation.

- Silhouette Score: Validates how well-separated the clusters are.

Results and Insights

- Based on the DB Index, the optimal number of clusters is 4.
- The DB Index value for 4 clusters is 0.45, indicating well-defined and separated clusters.

Cluster Characteristics

Cluster 1: High-value customers with frequent purchases.

Cluster 2: Low-value customers with occasional purchases.

Cluster 3: Customers focused on electronics.

Cluster 4: Customers from the North American region with diverse purchases.

Visualization

- Scatter plots of clusters based on total value and transaction frequency show clear separation.
- A heatmap of cluster centers highlights key characteristics of each group.

Conclusion

Customer segmentation reveals actionable insights:

- High-value customers can be targeted with loyalty programs.
- Low-value customers can be encouraged with promotional offers.
- Region-based segmentation can inform localized marketing strategies.
- Product-focused clusters can guide inventory and promotional decisions.