Data Science Assignment: eCommerce Transactions Dataset

VIGHNESH M S

232104@iiitt.ac.in

8088012208

Assignment Tasks:

Task 3: Customer Segmentation / Clustering

Perform **customer segmentation** using clustering techniques. Use both **profile information**
(from Customers.csv) and **transaction information** (from Transactions.csv).

- You have the flexibility to choose any clustering algorithm and any number of clusters in
  between(2 and 10)
- Calculate clustering metrics, including the **DB Index(Evaluation will be done on this)**.
- Visualise your clusters using relevant plots.

Deliverables:

- A report on your clustering results, including:
  - The number of clusters formed.
  - DB Index value.
  - Other relevant clustering metrics.
- A Jupyter Notebook/Python script containing your clustering code.

The customer segmentation analysis was performed to identify distinct customer groups based on their profile and transaction data. The clustering results provide actionable insights into customer behavior, helping to tailor marketing strategies and improve overall customer engagement.

Clustering Results

1. **Optimal Number of Clusters**:
   Based on the Davies-Bouldin Index (DBI), the optimal number of clusters identified was **8**.
2. **Davies-Bouldin Index (DBI)**:
   The DB Index for the 8 clusters is **1.1224**, indicating a good balance between cluster separation and compactness.
3. **Silhouette Score**:
   The Silhouette Score for the optimal clustering was **0.2559**, suggesting moderate cluster quality. This indicates that some clusters are distinct, while others may slightly overlap.
4. **Inertia (WCSS)**:
   The Within-Cluster Sum of Squares (WCSS) for the 8 clusters was **13.7024**, reflecting the compactness of the clusters.

Cluster Insights

Further analysis of the clusters (details in `Customer_Segments.csv`) revealed:

1. Each cluster represents a distinct customer group with unique spending patterns, product preferences, and transaction frequency.
2. **High-value customers** are concentrated in specific clusters, which can be targeted with loyalty programs or premium offers.
3. Clusters with **low transaction frequency** represent potential opportunities for re-engagement through targeted campaigns.
4. **Regional differences** in customer behavior are evident, highlighting the need for localized marketing strategies.
5. Clusters with dominant spending in specific product categories (e.g., Books, Electronics) provide an opportunity for category-focused promotions.