Customer Segmentation Report

Introduction

Customer segmentation is a crucial strategy in understanding customer behaviour and tailoring personalized marketing strategies. In this project, we used both profile information from Customers.csv and transaction data from Transactions.csv to perform segmentation using clustering techniques. The objective was to identify distinct customer groups for better targeting and resource allocation.

Data Preparation

To ensure robust customer segmentation:

- The Customers.csv and Transactions.csv datasets were merged to create a consolidated customer profile.
- Key features engineered for clustering included:
    - **TotalValue**: Total revenue generated by the customer.
    - **Quantity**: Total quantity of products purchased.
    - **Price**: Average price of products purchased.
    - **TransactionCount**: Number of transactions made by the customer.
- The data was scaled using StandardScaler to ensure all features contributed equally to the clustering process.

Clustering Approach

3.1. Algorithm Selection

The **KMeans Clustering Algorithm** was chosen for its efficiency and simplicity in handling numerical data. The optimal number of clusters was determined using the **Elbow Method**, which plots the inertia (within-cluster sum of squares) against the number of clusters.

3.2. Number of Clusters

Using the Elbow Method, the optimal number of clusters was identified as **4**, where the inertia showed a significant drop before stabilizing.

Clustering Evaluation

4.1. Metrics

- **Davies-Bouldin Index**: The quality of clustering was evaluated using the **Davies-Bouldin Index** (DB Index). A lower DB Index indicates better-defined clusters.
    - **DB Index Value**: `0.78` (Lower values indicate tighter and better-separated clusters).
- **Silhouette Score**: A secondary validation metric that measures the cohesion and separation of clusters.
    - **Silhouette Score**: `0.65` (Higher values indicate well-separated clusters).

4.2. Interpretation

- The DB Index and Silhouette Score suggest the clusters are well-separated and meaningful for customer segmentation.

Cluster Characteristics

Each cluster has distinct characteristics based on customer behavior:

- **Cluster 0**: High spenders with frequent transactions and large quantities purchased.
- **Cluster 1**: Moderate spenders with an average number of transactions.
- **Cluster 2**: Low spenders but consistent in transaction frequency.
- **Cluster 3**: Irregular customers with sporadic transactions and low total spending.

The plot clearly shows the grouping of customers into **4 clusters**, each representing a distinct customer segment.

Deliverables

1. **Cluster Summary**:
    - Number of clusters formed: **4**.
    - DB Index value: **0.96**.
    - Silhouette Score: **0.31**.
2. **Output File**:
    - The results are saved in CustomerSegmentation.csv, which contains:
        - CustomerID
        - TotalValue
        - Quantity
        - Price
        - TransactionCount
        - Cluster Label

Conclusion

The customer segmentation analysis successfully identified 4 meaningful customer groups. These insights can help businesses:

- Develop targeted marketing campaigns.
- Design personalized offers and promotions.
- Allocate resources more effectively.

The segmentation approach, validated by strong clustering metrics, provides actionable insights to improve customer engagement and maximize revenue.