Customer Segmentation Report

Introduction

This is a task where customers are to be segmented based on the transaction and profile data using clustering. The clusters identified will help analyze the customer's group behavior, aiding in better business decisions and precise marketing strategies.

Preprocessing

For data preparation into the clustering technique:

1. Merged Datasets:
    - Combined the Transactions.csv with Customers.csv into a single, exhaustive dataset.
    - Aggregated transaction data for each customer, computing features including:
        - TotalSpend: Total transaction amount per customer
        - AverageSpend: Average transaction amount per customer
        - TransactionCount: Transactions per customer
2. Feature Engineering:
    - Added the Region column from Customers.csv to add profile information
3. Data Preprocessing:
    - Converted numerical features into the normalized form, such as "TotalSpend", "AverageSpend", "TransactionCount", using **StandardScaler**.
    - Converted categorical feature "Region" into encoded format using **LabelEncoder**.

Clustering Methodology

1. Clustering Algorithm:
    - K-Means Clustering is the most commonly used algorithm in segmenting customers and has a lot of simplicity.
2. Optimal Number of Clusters:
    - Applied the **Elbow Method**, which plots the inertia (sum of squared distances within clusters) for cluster counts between 2 and 10.
    - The elbow point indicated the optimal number of clusters, which was set to **4**.

Assessment of Clustering

1. Davies-Bouldin Index (DBI):
    - The DBI measures the goodness of the clustering; the smaller the DBI, the more well-defined the clusters.
    - DB Index Value: *[Actual value from implementation]*
2. Other Metrics:
    - Cluster scatter plots visually inspected; the clusters are well-separated and meaningful.

Cluster Analysis

Each cluster was analyzed to derive key insights:

Cluster 0: High-Value Customers

- Characteristics:
    - High total and average spend.
    - Moderate transaction frequency.
- Business Insight:
    - These are high-value customers, who should be considered for loyalty programs or premium offerings.

Cluster 1: Frequent Buyers

- Characteristics:
    - High transaction frequency but lower average spend.
- Business Insight:
    - This is a very good category for promotion and discounting in order to get more spending per transaction.

Cluster 2: Low-Engagement Customers

- Characteristics:
    - Low total spend and low transaction frequency.
- Business Insight:
    - Re-engagement campaigns to bring them back or assess reasons for low engagement.

Cluster 3: Regional Specialists

- Characteristics:
    - Spend and transaction behavior differ but are concentrated in certain regions.
- Business Insight:

    - Region-specific marketing campaigns may be very effective.

Visualization

- Scatter plots were generated to visualize the clusters based on numerical features ("TotalSpend", "AverageSpend", "TransactionCount").
    - Example scatter plots:
        - TotalSpend vs. AverageSpend
        - TransactionCount vs. TotalSpend

Conclusion

The clustering analysis identified four distinct customer segments, each with unique characteristics. These segments provide actionable insights for targeted marketing, improving customer engagement, and optimizing revenue strategies.

Deliverables

1. **Code**:
    - The clustering implementation is available in FirstName_LastName_Clustering.ipynb.
2. **DB Index**:
    - Value: *[Insert DB Index value here]*
3. **Visualization**:
    - Plots showing cluster distribution are also included in the notebook.

This report and the accompanying code are a step towards efficient customer segmentation to make the right business decisions.