Customer Segmentation Report

Overview

This report details the customer segmentation performed using clustering techniques. The segmentation was based on both customer profile information and transactional data. The main goal was to divide customers into meaningful clusters for better insights and targeted strategies.

Data Preprocessing and Feature Engineering

Steps Taken:

1. **Merged Datasets:**

   o Combined Customers.csv and Transactions.csv using CustomerID.

2. **Datetime Conversion:**

   o Converted SignupDate and TransactionDate columns to datetime format.

3. **Feature Engineering:**

   o **Recency:** Calculated the number of days since the last transaction for each customer.

   o **TotalSpend:** Sum of all TotalValue for each customer.

   o **Frequency:** Count of transactions for each customer.

4. **Categorical Encoding:**

   o Encoded the Region column using Label Encoding.

5. **Feature Standardization:**

   o Standardized numerical features for better clustering results.

Clustering Approach

Algorithm:

- **K-Means Clustering**

Number of Clusters:

- Chosen: 5 (after evaluating metrics like Elbow Method and Silhouette Score).

Features Used:

1. TotalSpend
2. Frequency

3. Recency

4. Region

Evaluation Metrics

Davies-Bouldin Index (DB Index):

- **Value:** 1.0560 (Lower is better; indicates compact and well-separated clusters.)

Silhouette Score:

- **Value:** 0.2978 (Ranges from -1 to 1; higher indicates better-defined clusters.)

Visualization

A 2D visualization was generated using Principal Component Analysis (PCA) to reduce the dimensions of the dataset. The plot shows distinct clusters, supporting the segmentation results. The visualization is included in the accompanying Jupyter Notebook.