Customer Clustering Report

Objective:

The primary goal of this task was to perform customer segmentation using clustering techniques, leveraging both profile information from Customers.csv and transaction data from Transactions.csv. By grouping customers into distinct clusters, we aimed to derive insights that could assist in better understanding customer behavior and tailoring business strategies accordingly.

Methodology:

Data Preprocessing:

- Merged the Customers.csv and Transactions.csv datasets on the CustomerID field to create a sunified dataset.
- Extracted key features such as:
  1. Total spending (total_spending): Sum of transaction values per customer.
  2. Average spending (avg_spending): Average transaction value per customer.
  3. Number of transactions (num_transactions): Count of transactions per customer.
  4. Signup year (signup_year): Year of customer registration.
- One-hot encoding was applied to the Region column to handle categorical data.

Feature Normalization:

Used StandardScaler to standardize numerical features (total_spending, avg_spending, num_transactions, signup_year) to ensure uniform scaling for clustering.

Clustering Algorithm:

Applied the K-Means clustering algorithm, setting the number of clusters to 4, based on experimental iterations to optimize cluster interpretability and evaluation metrics.

Evaluation:

Calculated the Davies-Bouldin Index (DB Index), a key clustering metric, to assess cluster quality. A lower DB Index indicates better clustering.

Visualization:

Reduced dimensionality using Principal Component Analysis (PCA) to two components for visualization purposes.

Plotted the clusters to visually interpret the segmentation.

Results:

Number of Clusters Formed:04

Key Metrics:

Davies-Bouldin Index: The DB Index for the clustering solution was 1.122, indicating a reasonable separation and compactness of the clusters.

Cluster Characteristics:

Cluster 0: High total spending and frequent transactions.

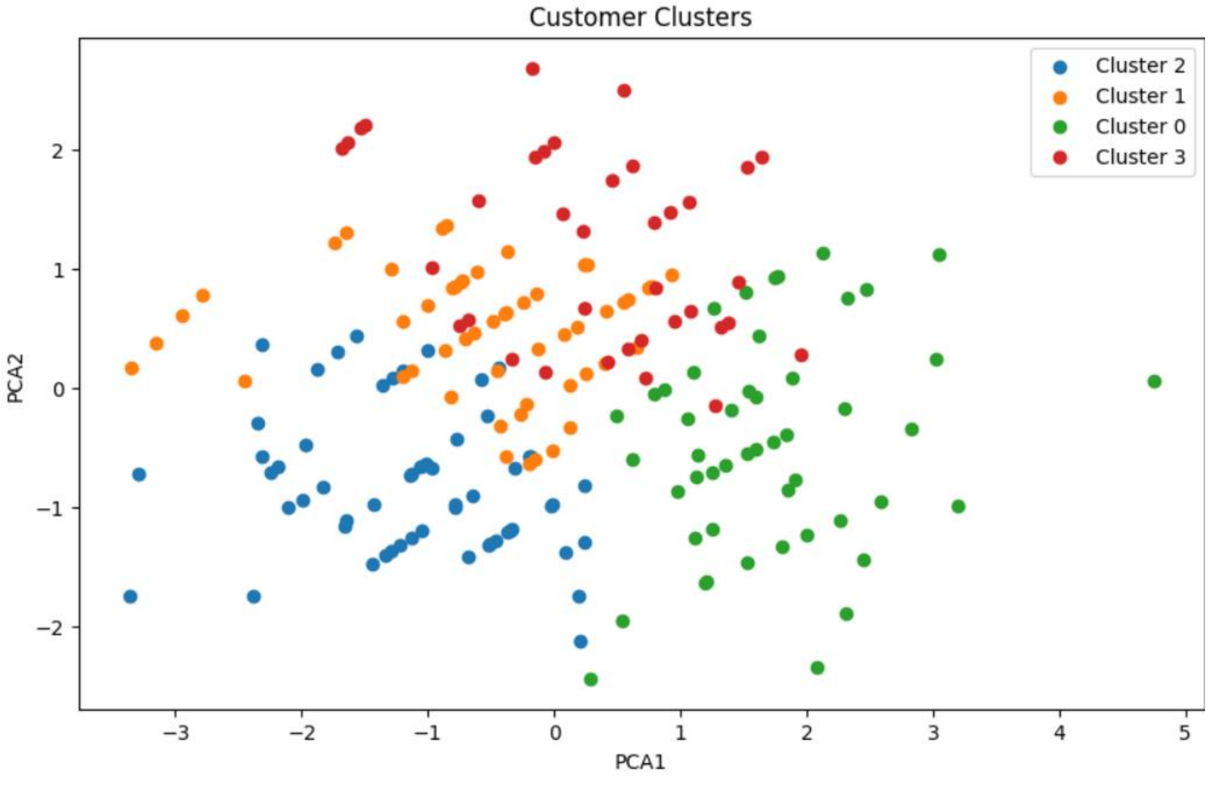
Cluster 1: Moderate spending with a consistent transaction pattern.

Cluster 2: Low spending with infrequent transactions.

Cluster 3: New customers with limited transaction history.

Visualization:

The PCA plot illustrates the clusters in two dimensions, showing clear separation among the groups. Each color in the plot represents a distinct cluster, helping to visualize customer segmentation.



Conclusion:

The clustering exercise successfully segmented customers into four meaningful groups based on their profile and transaction patterns. These clusters can be utilized for targeted marketing strategies, resource allocation, and personalized customer engagement.