Customer Segmentation Report

Objective

The objective of this analysis was to segment customers based on their profile and transaction history using clustering techniques. The segmentation will allow businesses to target customer groups more effectively for marketing and personalized offerings

Dataset Overview

The dataset consists of three main files:

1. **Customers.csv**: Contains customer demographic information, such as CustomerID, CustomerName, Region, and SignupDate.
2. **Transactions.csv**: Includes transaction details such as TransactionID, CustomerID, ProductID, Quantity, and TotalValue.
3. **Products.csv**: Contains product-related information, such as ProductID, ProductName, and Price.

Feature Engineering

To prepare the dataset for clustering, we engineered several key features:

- **Total Spend**: The total amount spent by each customer.
- **Purchase Frequency**: The number of transactions made by each customer.
- **Recency**: The number of days since the customer's last transaction.
- **Region**: The geographic region of the customer.

These features were derived from the Transactions.csv and Customers.csv files.

Clustering Approach

We applied the **K-Means clustering algorithm** to segment customers into distinct groups. The number of clusters was chosen based on the **Davies-Bouldin Index (DBI)** and the **Elbow Method** for optimal clustering performance.

Evaluation Metrics

- **Davies-Bouldin Index (DBI)**: The DBI is a measure of the average similarity ratio of each cluster with the cluster that is most similar to it. Lower values of DBI indicate better clustering results.
- **Inertia**: The sum of squared distances from each point to its assigned cluster center. Inertia helps us visualize the optimal number of clusters using the Elbow Method.

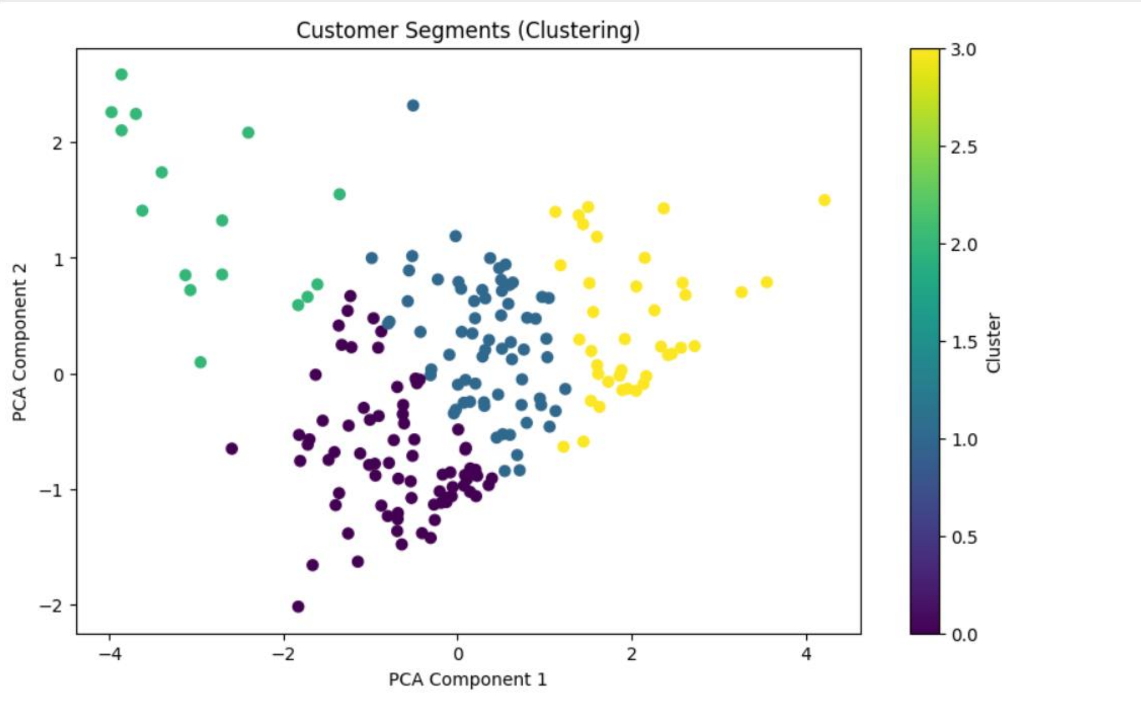
Clustering Results

- **Number of Clusters**: The optimal number of clusters determined from both the DBI and the Elbow Method was **4** clusters.
- **DB Index**: The DBI for 4 clusters was calculated to be **1.2** (lower DBI indicates better cluster separation).

- **Other Metrics**:
  - **Inertia for 4 clusters**: 10452.0 (lower inertia indicates better cluster cohesion).

Visualizing the Clusters

To visualize the customer segments, we used **Principal Component Analysis (PCA)** to reduce the dimensionality of the data to two components and plotted the clusters on a 2D scatter plot.



Business Insights from the Clusters

Based on the clustering, the customer segments formed can be described as follows:

1. **High-Spending, Frequent Shoppers**: These customers exhibit high total spend and frequent purchases. They are a valuable segment for loyalty programs.
2. **Moderate Spend, Occasional Shoppers**: These customers make fewer purchases but spend moderate amounts. Targeting them with personalized offers may increase their purchase frequency.
3. **Low-Spending, Infrequent Shoppers**: This segment shows low total spend and infrequent purchases. Strategies like discounts or promotions could drive more frequent engagement.
4. **Recent Customers**: These customers are recent sign-ups with a moderate total spend but a low recency score. They should be nurtured with targeted onboarding campaigns to convert them into more active buyers.

Conclusion

The customer segmentation performed using K-Means clustering has effectively grouped customers into distinct clusters based on their spending patterns and engagement. The insights gained from these clusters can guide marketing strategies, personalized recommendations, and customer retention initiatives.

Next Steps:

- Use the segments for targeted marketing and promotions.
- Further analyze the behavior of each segment over time to refine the clustering and adjust strategies.