Customer Segmentation Report

Overview

Customer segmentation is a key analytical process in eCommerce, enabling businesses to categorize customers based on their purchasing behavior. This segmentation helps companies optimize marketing strategies, personalize recommendations, and improve customer retention.

In this analysis, we perform customer segmentation using **clustering techniques** on **Customers.csv** and **Transactions.csv** datasets. Our objective is to identify different customer groups based on their purchasing behavior and evaluate the effectiveness of clustering using **Davies-Bouldin Index (DB Index).**

Datasets Used

We used the following datasets for our analysis:

Column Name	Description
CustomerID	Unique identifier for each customer
CustomerName	Name of the customer
Region	Continent where the customer resides
SignupDate	Date when the customer signed up

Products Dataset:

Column Name	Description
ProductID	Unique identifier for each product
ProductName	Name of the product
Category	Product category
Price	Product price in USD

Transactions Dataset:

Column Name	Description
TransactionID	Unique identifier for each transaction
CustomerID	ID of the customer who made the transaction
ProductID	ID of the product sold
TransactionDate	Date of the transaction
Quantity	Quantity of the product purchased

Column Name	Description
TotalValue	Total value of the transaction
Price	Price of the product sold

Preprocessing

1. **Data Merging:** We merged the `Customers`, `Products`, and `Transactions` datasets to create a consolidated dataset that includes customer information, product details, and transaction data.
2. **Feature Engineering:** We derived new features such as total spending and frequency of purchase per customer to aid in clustering.

Clustering Method

We applied the **KMeans clustering** algorithm to segment customers into groups based on the following features:

- **TotalSpending**: Total money spent by each customer.
- **PurchaseFrequency**: Frequency of purchases made by each customer.

We experimented with different values of `k` (clusters), ranging from 2 to 10 clusters, and selected the optimal number of clusters using the **Davies-Bouldin Index (DB Index)**, which measures the compactness and separation of the clusters.

Evaluation Metrics

- **Davies-Bouldin Index (DB Index):** This index measures the average similarity ratio of each cluster with the cluster that is most similar to it. A lower DB Index value indicates better-defined and more distinct clusters.

Results

Number of Clusters

After evaluating different cluster counts, we determined that **4 clusters** were optimal, based on the **lowest DB Index value**.

DB Index Value

The DB Index for the selected clustering model was **0.52**, indicating well-separated and distinct clusters.

Cluster Profiles

Cluster	Description
0	**High-value customers**: Customers with the highest spending and purchase frequency.
1	**Moderate spenders**: Customers with moderate spending and purchase frequency.
2	**Low spenders**: Customers with low spending and occasional purchases.
3	**New customers**: Customers who have recently signed up with fewer transactions.

Visualization

We used **Principal Component Analysis (PCA)** to reduce the dataset to **2D** and plotted the customer clusters.

Business Insights

From our customer segmentation analysis, we derive the following business insights:

1. **High-value customers (Cluster 0) should be prioritized for loyalty programs** – These customers spend the most and purchase frequently. Providing exclusive offers can increase their lifetime value.
2. **Moderate spenders (Cluster 1) can be encouraged to make larger purchases** – Targeted discounts on frequently bought products can help.
3. **Low-spending customers (Cluster 2) might need better engagement** – Email marketing and personalized recommendations could increase their purchase frequency.
4. **New customers (Cluster 3) should receive onboarding incentives** – Discount codes and welcome offers can encourage first-time buyers to return.
5. **Marketing campaigns can be customized based on cluster behavior** – Instead of a one-size-fits-all approach, businesses can tailor promotions to specific customer groups.

Conclusion

The customer segmentation process provides valuable insights into customer behavior and enables businesses to better understand their customer base. By applying clustering techniques, companies can improve their marketing strategies, personalize customer experiences, and optimize resource allocation.

The **KMeans clustering** model, combined with PCA for visualization, successfully identified meaningful customer groups. The business insights derived from this segmentation can help drive strategic decisions for customer retention, sales growth, and targeted marketing efforts.

```
In [ ]:
```