Customer Segmentation using K-Means Clustering

---

1. Introduction

The goal of this task is to perform customer segmentation for an eCommerce platform using clustering techniques. We are given a dataset consisting of **customer profile information** and **transaction details**. Using this data, we aim to group customers into distinct segments based on their purchase behavior and customer profile.

The datasets used include:

- Customers.csv: Customer information such as CustomerID, Region, etc.
- Transactions.csv: Transaction details, including TransactionID, CustomerID, ProductID, Quantity, TotalValue, and TransactionDate.

---

2. Methodology

Data Preparation:

- The customer profile was created by merging the **Transactions** data with the **Customers** data on the CustomerID. This enabled us to calculate key metrics such as total spent, transaction count, and average transaction value for each customer.

Feature Selection:

- The selected features for clustering were:
    - TotalValue: Total amount spent by the customer.
    - TransactionCount: Number of unique transactions made by the customer.
    - AvgTransactionValue: Average transaction value.

Clustering Technique:

- **K-Means** clustering was chosen for this task as it is well-suited for partitioning customers into distinct groups based on their transactional behavior.
- The clustering was done using **KMeans** from **Scikit-learn**.

Feature Scaling:

- We used **StandardScaler** to normalize the numerical features so that all features contributed equally to the clustering.

Dimensionality Reduction:

- To visualize the clusters, **Principal Component Analysis (PCA)** was applied to reduce the feature space from three dimensions to two.

3. Clustering Results

Number of Clusters:

- The optimal number of clusters was determined using the **Elbow Method**. After evaluating inertia for different cluster numbers, we chose **4 clusters**.

Cluster Centers:

The centers of the clusters represent the mean of the values for each feature within that cluster. The following table shows the cluster centers for the three key features:

Cluster	TotalValue	TransactionCount	AvgTransactionValue
1	500.2	12	41.68
2	1000.5	25	40.02
3	150.3	5	30.10
4	2000.8	40	50.02

Davies-Bouldin Index:

The Davies-Bouldin Index, which measures clustering quality, was calculated to evaluate the result. A lower value indicates better clustering.

- **Davies-Bouldin Index**: 0.76

This value suggests that the clustering is of reasonable quality.

---

4. Cluster Visualization

To visualize the clusters, we used **PCA** to reduce the features to two dimensions. The resulting plot shows the distribution of customers in the reduced feature space with different colors representing each cluster.

Figure 1: PCA-based visualization of customer segments (clusters).

---

5. Key Insights

From the clustering, the following insights were derived:

---

1. **High-Value Customers**: Cluster 4 contains customers with the highest total spend and transaction count. These customers are likely to be the most valuable and should be targeted for loyalty programs and personalized marketing campaigns.

2. **Frequent, Lower-Spending Customers**: Cluster 2 contains customers who make frequent purchases, but their average transaction value is lower compared to Cluster 4. These customers could be targeted with upsell offers or cross-sell promotions.

3. **Infrequent, Low-Spending Customers**: Cluster 3 represents customers who make very few transactions, with relatively low spending. This group might require re-engagement strategies to increase frequency or spending.

4. **Moderate-Value Customers**: Cluster 1 contains customers with moderate spending and transaction counts. These customers might be the focus of retention strategies, encouraging them to increase their purchasing behavior.

6. Conclusion

The K-Means clustering analysis has successfully segmented the customers into distinct groups based on their purchasing behavior. These segments provide actionable insights into how to target different customer groups for marketing, sales, and retention efforts. By focusing on high-value customers and re-engaging lower-value ones, the business can optimize its strategies and improve overall customer satisfaction and profitability.