Customer Segmentation Using Clustering

Objective: The goal of this analysis was to segment customers based on both their profile and transaction behavior. We utilized clustering techniques to group customers into distinct segments and assessed the clustering quality using the Davies-Bouldin Index.

Dataset Used:

- **Customers.csv:** Contains customer details including the CustomerID, CustomerName, Region, and SignupDate.
- **Products.csv:** Contains product details including ProductID, ProductName, Category, and Price.
- **Transactions.csv:** Contains transaction data including TransactionID, CustomerID, ProductID, TransactionDate, Quantity, TotalValue, and Price.

Data Preprocessing: The customer and transaction data were merged to create a consolidated dataset. Customer transaction behavior was aggregated based on total spending, number of transactions, average spending per transaction, and average quantity per transaction.

Clustering Algorithm: We applied the KMeans clustering algorithm with 5 clusters to segment the customers. KMeans is a centroid-based algorithm that groups customers into clusters by minimizing the variance within each cluster.

Evaluation Metric: We used the **Davies-Bouldin Index (DB Index)** to evaluate the quality of the clusters. The DB Index is a measure of the average similarity ratio of each cluster with the cluster that is most similar to it. A lower DB Index indicates better clustering.

Clustering Results:

- **Number of clusters formed:** 5 clusters
- **DB Index value:** 1.0534768450920309

Visualizations:

1. **Pairplot:** A pairplot was created to visualize the relationships between key metrics such as total spending, total transactions, and average spending per transaction across the clusters.
2. **PCA Scatter Plot:** A 2D scatter plot was generated using Principal Component Analysis (PCA) to reduce the dimensionality of the data, allowing for a clear visualization of the customer segments.

Conclusion: The segmentation successfully divided customers into meaningful clusters based on their transaction and profile data. The low DB Index suggests that the clusters are well-separated and distinct. These customer segments can be used for targeted marketing or further business strategy development.