Code summary :

The code performs customer segmentation using K-Means clustering, followed by evaluation and visualization of the results. Here's an overview of the key steps:

1. Data Merging: The code loads data from three CSV files (Customers.csv, Transactions.csv, Products.csv) and merges them to create a combined dataset of customer transactions.

2. Feature Aggregation:

   o Aggregates data for each customer by calculating:

      ▪ Total Spend: The sum of money spent by the customer.

      ▪ Transaction Count: The number of transactions made by the customer.

      ▪ Total Quantity: The total quantity of products purchased by the customer.

3. Normalization: Standardizes the aggregated features (TotalSpend, TransactionCount, TotalQuantity) to make them comparable for clustering.

4. K-Means Clustering:

   o Applies K-Means clustering for cluster sizes ranging from 2 to 10.

   o Evaluates the clustering performance using:

      ▪ Davies-Bouldin Index: A measure of cluster separation, with a lower value indicating better clustering.

      ▪ Silhouette Score: A metric that measures how well-defined the clusters are, with higher values being better.

5. Best Cluster Selection: The best clustering result is selected based on the lowest Davies-Bouldin Index.

6. PCA Visualization:

   o Uses Principal Component Analysis (PCA) to reduce the dimensionality of the features and visualize the clusters in a 2D scatter plot.

7. Output:

   o The best clustering result is displayed, including the number of clusters and the Davies-Bouldin Index.

   o A plot of the customer clusters is shown, making it easier to visualize the segmentation.

Key Points:

- The clustering is used to segment customers based on their spending behavior, transaction frequency, and total quantity purchased.

- The effectiveness of clustering is evaluated using the Davies-Bouldin Index and Silhouette Score.

- The resulting clusters are visualized in 2D using PCA for easier interpretation.

This approach is useful for businesses to understand customer segments, which can guide personalized marketing strategies or customer relationship management efforts.