CLUSTERING

Objective

The code performs customer segmentation using clustering techniques, aiming to group customers based on their transactional behavior for better understanding and targeted marketing.

A] **Data Loading and Preprocessing**

- The datasets Customers.csv, Products.csv, and Transactions.csv are loaded.
- Date columns (SignupDate and TransactionDate) are converted to datetime format to ensure compatibility for time-based analyses.

B] **Feature Engineering**

- Transaction data is aggregated by CustomerID to compute two key metrics:
  - TotalValue: Total monetary value of transactions.
  - Quantity: Total number of items purchased.
- These metrics are merged with customer data, filling missing values with zero for consistency.

C] **Normalization**

- The TotalValue and Quantity metrics are normalized using StandardScaler to ensure equal weighting during clustering.

D] **Clustering**

- **K-Means Clustering**:
  - The normalized data is used to create 5 clusters (n_clusters=5) using the K-Means algorithm.

- Each customer is assigned to a cluster, which is appended as a new column (Cluster) in the dataset.

E] **Visualization**

- A scatterplot is generated to visualize the clusters:
  - **X-Axis**: TotalValue
  - **Y-Axis**: Quantity
  - **Hue**: Cluster (represented by different colors)

F] **Output**

- The function returns a DataFrame containing CustomerID and their corresponding Cluster.