TASK -3

Main Steps for Customer Segmentation using Clustering:

1. **Load and Merge Data**:

   - Import the customer profile data (Customers.csv) and transaction data (Transactions.csv).
   - Merge the two datasets on the common column CustomerID.

2. **Data Preprocessing**:

   - Clean the data by handling missing values (e.g., dropping or imputing).
   - Create new features that will be useful for clustering:
     - **Total Spent**: Sum of TotalValue for each customer.
     - **Purchase Frequency**: Count of TransactionID for each customer.

   - Drop duplicates (if necessary) after creating new features.

3. **Feature Engineering**:

   - Select relevant features for clustering (e.g., total_spent, purchase_frequency).
   - Standardize or normalize the features to avoid bias due to differences in scales.

4.

Clustering:

- Apply a clustering algorithm like **K-Means** to segment customers into clusters (e.g., 4 clusters).
- Add the cluster labels to your dataset for each customer.

5. **Evaluate Clustering**:

- Use metrics like **Davies-Bouldin Index** (DB Index) and **Silhouette Score** to evaluate the quality of the clustering.
- A lower DB Index and a higher Silhouette Score indicate better clustering.

6. **Visualization**:

- Use **PCA** (Principal Component Analysis) or **t-SNE** to reduce data to 2D or 3D for visualization.
- Plot the clusters to visually inspect how well the customers are grouped.

7. **Save Results**:

- Save the final dataset with cluster labels for further analysis or reporting.