Task 3: Customer Segmentation / Clustering

Clustering logic and Results

The clustering logic in your code is focused on segmenting customers based on their spending and transaction behavior. Below is the steps for clustering process:

1. **Loading Data**:

    - The load data function reads customer and transaction data from CSV files. It loads two datasets:
    - Customers: Contains customer details like CustomerID.
    - Transactions: Contains transaction records, including CustomerID, transaction details, and the total value of each transaction.

2. **Data Preprocessing**:

    - The preprocess data function processes the raw transaction data by aggregating transaction details for each customer:
    - This aggregated data is then merged with the customers dataset on CustomerID.

3. **Finding Optimal Number of Clusters**:

    - Find optimal clusters function performs K-Means clustering with different numbers of clusters (from 2 to 10).
    - For each value of k, the function fits a K-Means model, predicts the cluster labels, and computes the Davies-Bouldin (DB) index, which is a measure of clustering quality. A lower DB index indicates better clustering.
    - The optimal number of clusters is selected as the one that minimizes the DB index.

4. **Clustering**:

    - Clustering function runs K-Means with the optimal number of clusters found in the previous step.
    - It adds a new column to the dataset called Cluster, which indicates the cluster assignment for each customer.

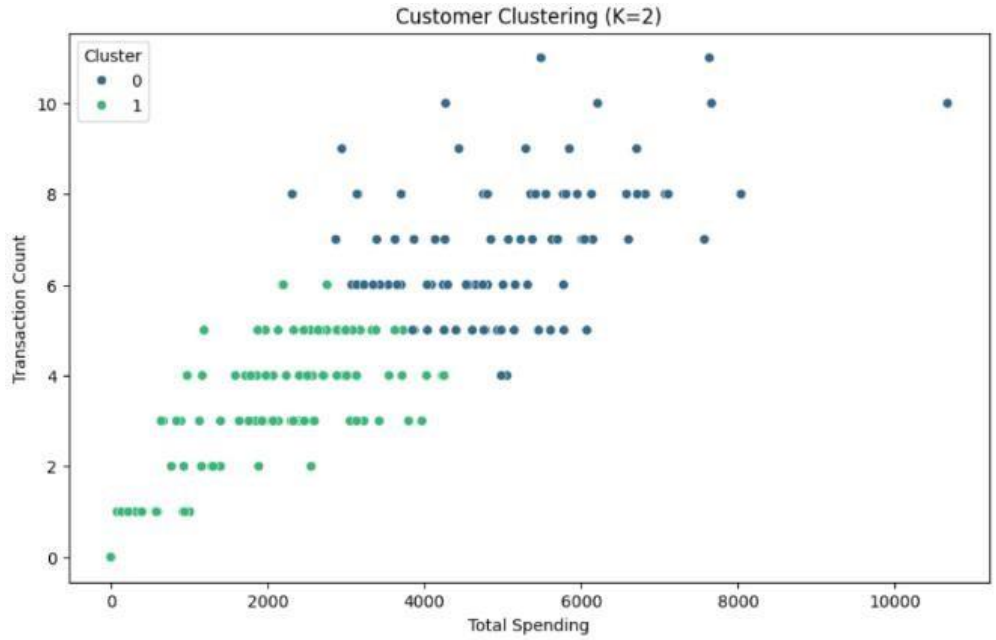
5. **Visualizing Clusters**:

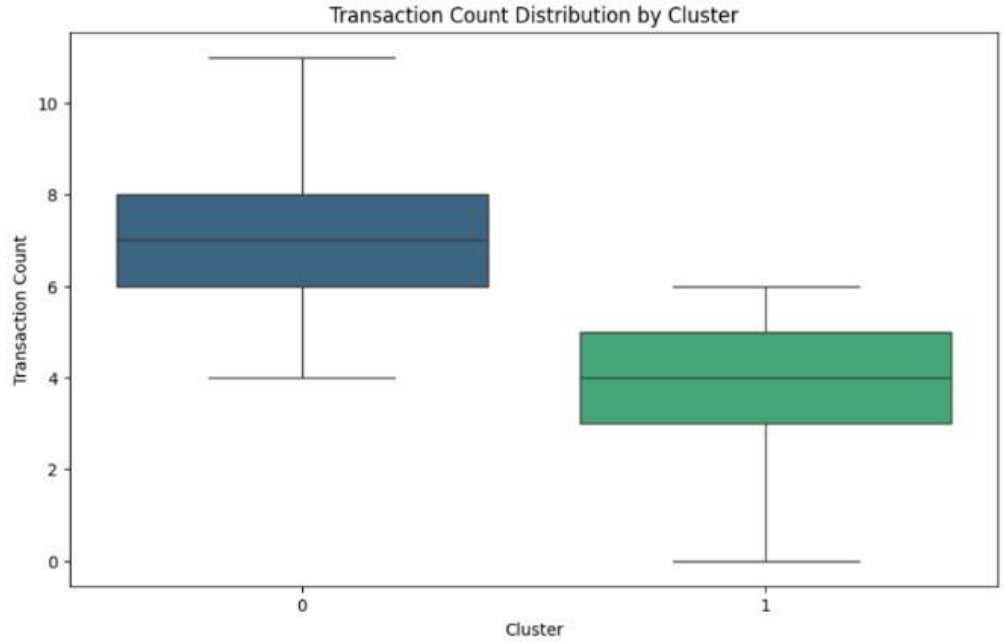
    - Visualize clusters function generates several plots to visualize the clustering results

6. **Results**:

    - The main function runs the entire process: loading data, preprocessing, finding the optimal clusters, performing clustering, visualizing the clusters, and printing the optimal number of clusters and the DB index.

Results






```
Optimal Clusters: 2, DB Index: 0.7340
```

The number of clusters formed = 2

DB Index value = 0.7340