Task 2
Clustering Result:

Understanding customer behavior and crafting personalized strategies to improve engagement and revenue relies heavily on customer segmentation. We utilized K-Means Clustering to categorize the eCommerce dataset into 10 separate clusters, based on customer profiles and their transactional data. The goal of this clustering was to identify patterns and insights that would assist the business in refining its strategies.

The clustering process consisted of several key steps:

1. **Data Integration and Feature Selection:** We combined data from Customers.csv and Transactions.csv to form a detailed dataset.

    Important features included:
    1. **Total Spending:** The total amount spent by each customer.
    2. **Transaction Frequency:** How often each customer made transactions.
    3. **Average Transaction Value**: The average amount spent per transaction.
    4. **Demographic Data:** Information on region and signup date.

2. **Feature Scaling:** To ensure that all features had an equal impact on the clustering process, we normalized the numerical features using StandardScaler.

3. **Clustering Algorithm:** We employed K-Means Clustering, selecting the number of clusters by minimizing the Davies-Bouldin Index (DB Index), which assesses the compactness and separation of clusters.

4. **Optimization:** We found that the **optimal number of clusters was 10**, with a **DB Index of 1.535,** indicating that the clusters were well-defined and distinct. The **Silhouette Score was calculated at 0.154**, highlighting the complexity of customer behavior.

To visualize the clusters, we applied Principal Component Analysis (PCA) for dimensionality reduction. The resulting scatter plot of the first two principal components displayed clear separations among most clusters, although some overlap occurred because of the high-dimensional feature space.