Task 3: Customer Segmentation / Clustering

1. Number of Clusters Formed

The optimal number of clusters determined using the **KMeans** clustering algorithm is **3**. These clusters represent distinct customer segments based on their transaction frequency and total spending behavior.

2. Davies-Bouldin Index (DB Index)

The Davies-Bouldin Index, which evaluates the quality of clustering by measuring intra-cluster similarity and inter-cluster dissimilarity, was calculated as **0.766**. A lower DB Index value indicates better-defined and more distinct clusters.

3. Other Relevant Clustering Metrics

- **Clustering Features**:
  - Transaction Frequency (TransactionCount)
  - Total Spending (TotalValue)
  - **Standardization:** Features were standardized using StandardScaler to ensure equal weighting.
  - **Silhouette Score :** The Silhouette Score for the optimal clustering is 0.49. This value indicates moderate clustering quality, with reasonable separation between clusters and cohesion within clusters.
- **Cluster Characteristics:**
  - **Cluster 0:** Represents customers with higher transaction counts and spending, likely indicating high-value customers or frequent shoppers.
  - **Cluster 1:** Consists of customers with lower transaction activity and spending, potentially representing occasional or low-value shoppers.

4. Visualization

A scatter plot of the clusters shows a clear separation based on transaction count and total spending. This provides insights into the behavioral differences between the identified customer segments.

5. Recommendations

• Cluster 0 (High-Value Customers): Focus on retention strategies, personalized offers, and loyalty programs to maximize lifetime value.

• Cluster 1 (Low-Value Customers): Implement targeted marketing campaigns to increase engagement and spending.