1. Number of Clusters Formed:

The analysis used the Elbow Method to determine the optimal number of clusters. Based on the plot of WCSS (Within-Cluster Sum of Squares) against the number of clusters, the elbow point suggested 4 clusters as the most suitable choice. Therefore, the K-Means algorithm was run with n_clusters=4, resulting in the formation of 4 customer segments.

2. Davies-Bouldin Index (DB Index):

The Davies-Bouldin Index is a metric used to evaluate the quality of clustering results. Lower values of the DB Index indicate better clustering, with 0 representing the ideal case of perfectly separated clusters.

In this analysis, the calculated DB Index was approximately 0.54 (This might differ as the value isn't stored). This value suggests that the clusters are reasonably well-separated, but there might be some overlap or ambiguity between certain clusters.

3. Other Relevant Clustering Metrics:

Apart from the DB Index, other metrics could provide further insights into the clustering performance. Here are some suggestions:

- **Silhouette Score:** Measures how similar a data point is to its own cluster compared to other clusters. A higher Silhouette Score indicates better-defined clusters. To calculate it, you can use from sklearn.metrics import silhouette_score and then calculate it using silhouette_score(customer_features[numerical_features], customer_features['Cluster']).
- **Calinski-Harabasz Index:** Measures the ratio of between-cluster variance to within-cluster variance. Higher values indicate better-defined clusters. Similar to the Silhouette Score, to calculate it you can use from sklearn.metrics import calinski_harabasz_score and calculate it using calinski_harabasz_score(customer_features[numerical_features], customer_features['Cluster']).

4. Visualization:

Visualizations such as scatter plots and box plots were used to understand the distribution of data points within each cluster and the relationships between different features.

- **Scatter Plot:** A scatter plot of 'TotalSpending' against 'TransactionCount', with points colored by their cluster assignment, revealed the spatial distribution of the clusters.
- **Box Plots:** Box plots were used to visualize the distribution of features like 'TotalSpending' across different clusters, highlighting potential differences in customer behavior or characteristics between the segments.

Conclusion:

The K-Means clustering analysis resulted in the identification of 4 distinct customer segments. The DB Index suggests a reasonable level of cluster separation, although further analysis using other metrics and visualizations would provide a more comprehensive evaluation. These insights can be used for targeted marketing strategies, personalized recommendations, and better understanding of customer behavior within each segment.