Customer Segmentation / Clustering

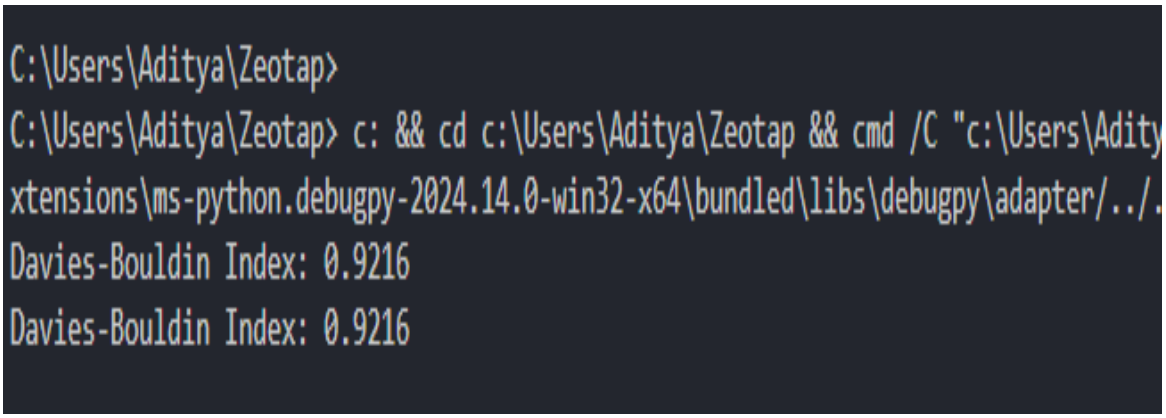
1. Number of Clusters Formed

The clustering analysis was performed using the **K-Means clustering algorithm**. Based on the analysis, the number of clusters formed was **5**.

2. Clustering Metrics

Davies-Bouldin Index

- **Value**: The Davies-Bouldin Index for the clustering is **0.9216**
- **Interpretation**: A lower Davies-Bouldin Index indicates that the clusters are compact and well-separated. This value reflects the overall performance of the clustering algorithm.



Silhouette Score

- **Value**: The Silhouette Score for the clustering is **0.3760**
- **Interpretation**: A higher Silhouette Score indicates better-defined clusters. Values closer to 1 suggest that data points are well matched within their cluster and poorly matched to other clusters.

Calinski-Harabasz Score

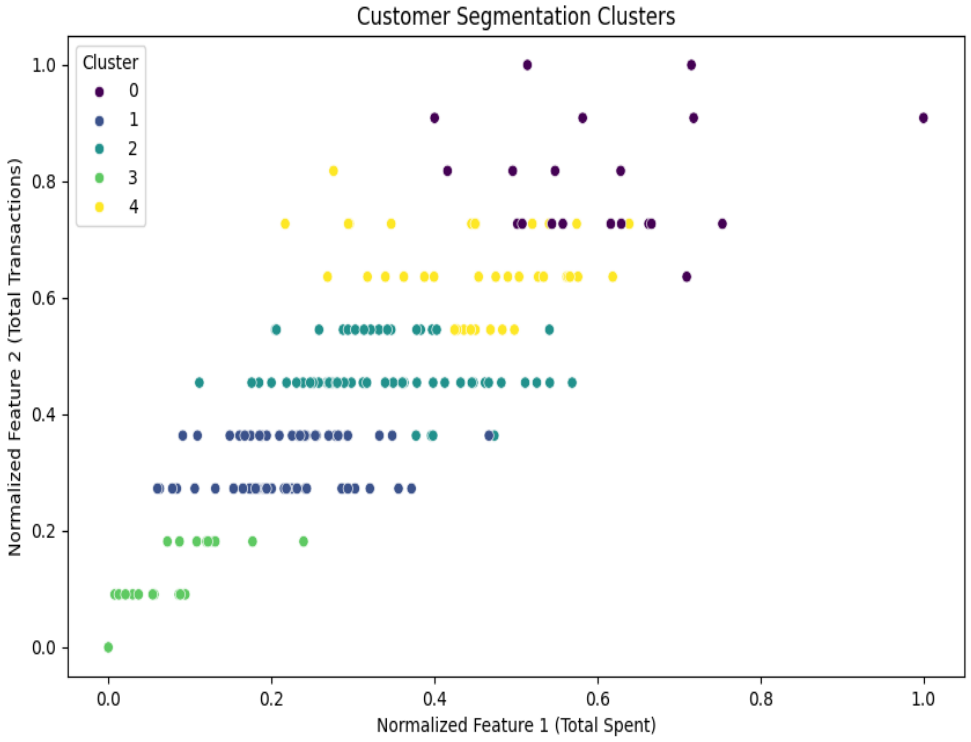
- **Value**: The Calinski-Harabasz Score for the clustering is **278.6607**
- **Interpretation**: A higher Calinski-Harabasz Score indicates that the clusters are dense and well-separated. This metric is particularly useful for assessing the overall clustering quality.

```
Columns in Customers.csv: ['CustomerID', 'CustomerName', 'Region', 'SignupDate']
Davies-Bouldin Index: 0.9216 (Lower is better)
Silhouette Score: 0.3760 (Higher is better)
Calinski-Harabasz Score: 278.6607 (Higher is better)
--- Clustering Metrics and Evaluation ---
Number of Clusters: 5
Davies-Bouldin Index: 0.9216 (Lower is better; indicates compact and well-separated clusters)
Silhouette Score: 0.3760 (Higher is better; indicates cohesion and separation)
Calinski-Harabasz Score: 278.6607 (Higher is better; indicates dense and well-separated clusters)
--- Visualizations ---
1. Scatter plot showing clusters using normalized features.
2. Pairplot of all features with cluster labels for detailed insights.
```

3. Visual Representations of Clusters

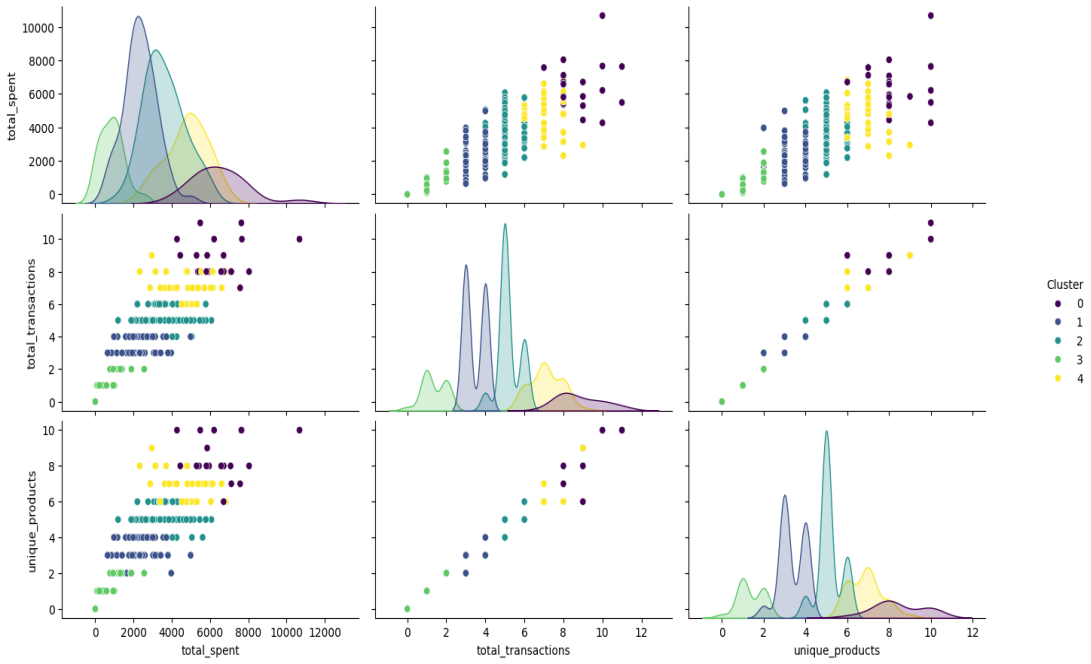
The clustering results were visualized using the following plots:

1. **Scatter Plot**:
    - Visualizes the distribution of clusters using the normalized features.
    - Highlights the separation and compactness of clusters.



2. **Pairplot**:

   - Displays pairwise relationships between all selected features along with the cluster labels.
   - Provides detailed insights into the distribution and overlap of data points across clusters.



4. Conclusion

- The clustering algorithm successfully segmented the customers into **5 clusters** based on their transaction history and profile information.
- The clustering metrics, including the Davies-Bouldin Index, Silhouette Score, and Calinski-Harabasz Score, indicate the overall quality and performance of the clustering.
- The visualizations confirm that the clusters are reasonably well-defined and separated.