Clustering Report

1. Objective

This analysis uses clustering techniques to segment customers based on their total spending and transaction count. **The Davies-Bouldin Index** determines the optimal number of clusters.

2. Data Used

- **Customers Dataset:** Contains customer ID, name, region, and signup date.
- **Transactions Dataset:** Includes transaction ID, customer ID, product ID, transaction date, quantity, total value, and price.

3. Data Preprocessing

- The datasets were merged based on CustomerID.
- Missing values in **TotalSpending** and **TransactionCount** were **filled with zeros**.
- Data was standardized using **StandardScaler** to ensure equal scaling of features.

4. Clustering Methodology

- **Feature Selection:**
  - TotalSpending: Total amount spent by a customer.
  - TransactionCount: Number of transactions made by a customer.
- **K-Means Clustering:**
  - Clusters were formed based on **standardized spending and transaction count**.
  - The optimal number of clusters was determined using **the Davies-Bouldin Index**.
- **Evaluation Metric:**
  - **The Davies-Bouldin Index** (DB Index) was used to assess **cluster cohesion and separation.**

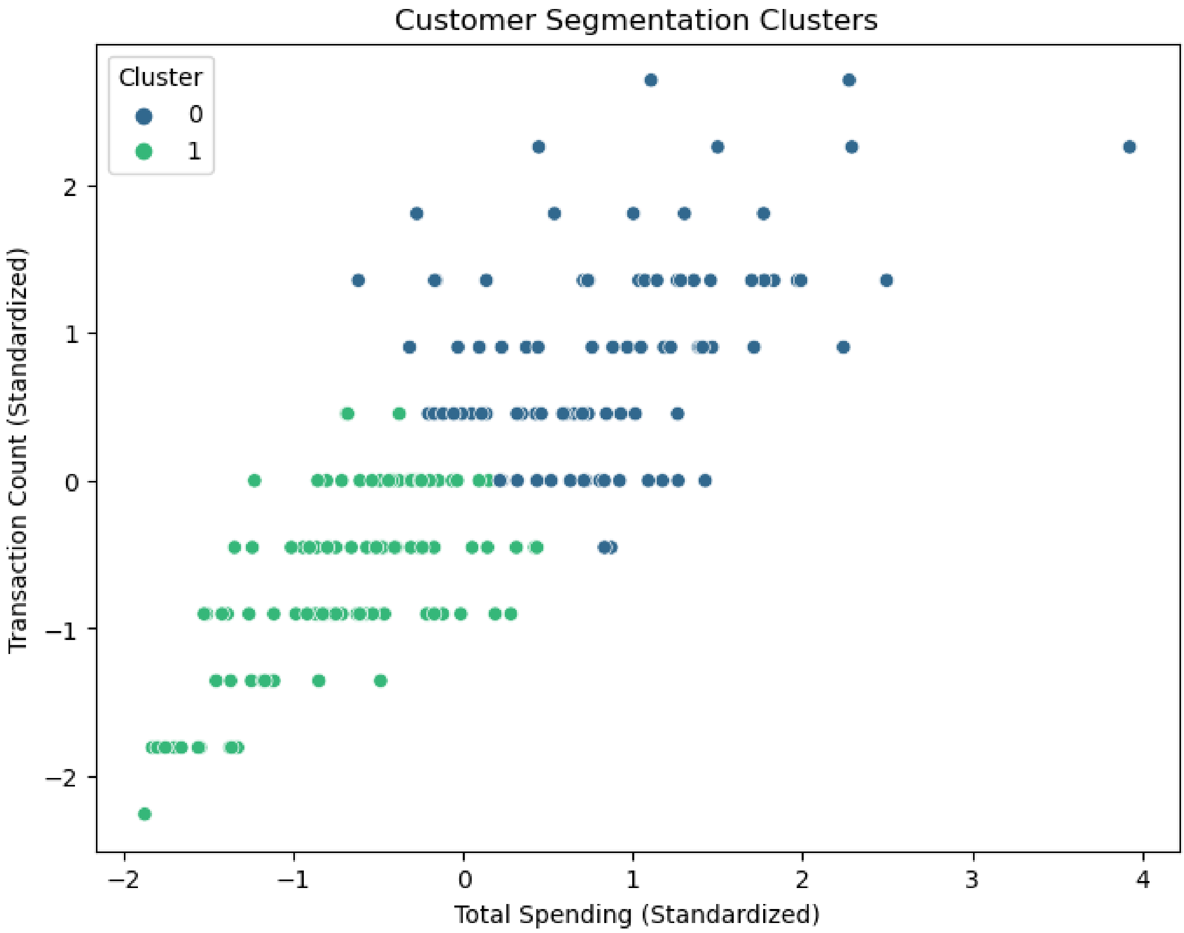
5. Results

- **Optimal Number of Clusters:** 2
- **Final DB Index for 2 Clusters:** 0.7339
- **Cluster Characteristics:**
  - **Cluster 0:** Higher spending and transaction count.
  - **Cluster 1:** Lower spending and transaction count.

Cluster	Total Spending (Avg)	Transaction Count (Avg)
0	6075.81	8.10
1	1711.75	2.84

6. Visualizations

- **Elbow Method Plot:**
  - Used to identify the optimal number of clusters.
  - Plotted DB Index values for clusters ranging from **2 to 10.**



- **Customer Segmentation Scatter Plot:**
  - Visualizes customer segmentation based **on total spending and transaction count**.
  - Different clusters are color-coded for **better analysis.**

7. Key Observations

- Customers in **cluster 0** are **high-value customers** with significantly **more spending and transactions.**
- **Cluster 1** customers exhibit **lower spending** behavior, potentially requiring **targeted marketing strategies.**
- The **DB Index value of 0.7339** suggests moderately well-separated clusters.

8. Additional Insights

- Performing further clustering with **3 clusters** resulted in a **DB Index of 0.7662**, indicating slightly poorer clustering quality. The additional cluster primarily **segments mid-level customers** who exhibit **moderate spending and transaction behavior**. However, the separation between clusters **is less distinct,** leading to an overlap in customer profiles and reducing the overall effectiveness of segmentation.
- The data highlights regional spending behavior, which can be leveraged for personalized marketing.

9. Conclusion

- The **optimal number** of clusters for **customer segmentation is 2.**
- The **3-cluster model** introduced **a middle segment** of customers with **moderate spending** behavior, but the clusters were **less distinct**, leading to potential challenges in precise targeting.
- The clustering results indicate that high-value customers **(cluster 0**) can be targeted with **exclusive offers,** while low-value customers **(cluster 1)** may need more **engagement efforts**.
- Further improvements can be achieved by incorporating additional features such as customer demographics and product preferences, which could enhance cluster separability and provide actionable business insights.

10. Recommendations

- Use segmentation results to **target high-value customers** with loyalty programs.
- Implement personalized marketing campaigns for **lower-spending customers** to increase their transaction frequency.