Customer Clustering Report

The report is of the clustering analysis performed on the customer dataset using the K-Means clustering algorithm. Below is a detailed summary of the clustering process and its outcomes.

1. Number of Clusters Formed

- After testing the K-Means clustering algorithm with different numbers of clusters (from 2 to 10), the optimal number of clusters was determined to be **2**.
- This was based on the **Davies-Bouldin Index** (DB Index), a metric used to evaluate clustering quality. Lower DB Index values indicate tighter and well-separated clusters, signifying better clustering.

2. Davies-Bouldin Index

- The Davies-Bouldin Index evaluates the compactness and separation of clusters. It calculates the average similarity ratio between each cluster and the one most similar to it.
- **Findings**:
    - The **lowest DB Index value**, 0.627, was observed for **2 clusters**, making it the optimal choice.
    - For clusters greater than 2, the DB Index increased, indicating a decline in clustering quality.
- DB Index Scores for Different Cluster Counts

Number of Clusters	Davies-Bouldin Index
2	0.627
3	0.709
4	0.721
5	0.777
6	0.823
7	0.878
8	0.943
9	0.881
10	0.814

- **Interpretation**:
    - As the number of clusters increases beyond 2, the clustering quality deteriorates (higher DB Index).
    - This trend suggests that adding more clusters introduces overlap between clusters and reduces compactness.

3. Customer Features Used

- The clustering was performed using the following customer-specific features:
    - **Total Spend**: Total amount spent by the customer.
    - **Quantity Purchased**: Total number of items purchased by the customer.
- These features were extracted and aggregated from the transaction dataset.
- Sample of Aggregated Features

CustomerID	TotalSpend	Quantity
C0001	3354.52	12
C0002	1862.74	10
C0003	2725.38	14
C0004	5354.88	23
C0005	2034.24	7

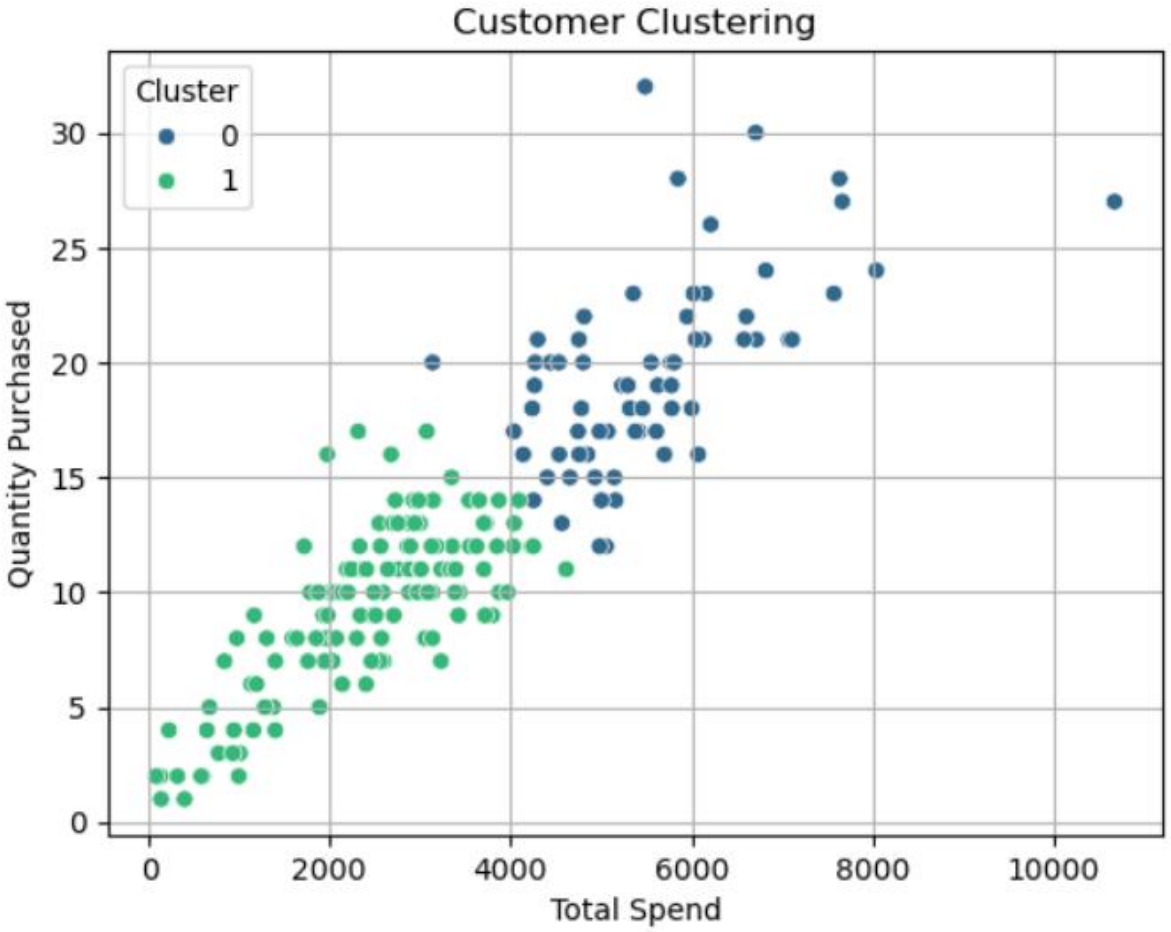
4. Visualization of Clusters

A scatter plot was created to visualize the clustering results:

- **Axes**:
    - X-axis: **Total Spend,** Y-axis: **Quantity Purchased**
- **Clusters**:
    - Each point represents a customer, with its color indicating the assigned cluster.
    - **Cluster 0 (High Spend, High Quantity)** and **Cluster 1 (Moderate Spend, Moderate Quantity)** are visually distinct.

Interpretation of the Scatter Plot:

- Customers in **Cluster 0** are concentrated in the top-right region, reflecting high spenders. Customers in **Cluster 1** are distributed in the mid-range, reflecting moderate spending and purchasing behavior.



5. Conclusion and Insights

- **Optimal Number of Clusters**: 2 clusters.
- **Clustering Quality**: The low **Davies-Bouldin Index (0.627)** indicates well-defined and separated clusters. Increasing the number of clusters beyond 2 resulted in overlapping and less-compact clusters.
- **Business Implications**:
    - **Cluster 0 (High Value Customers)**: Target with premium services, loyalty programs, and exclusive offers to maximize revenue.
    - **Cluster 1 (Moderate Value Customers)**: Focus on promotional campaigns or discounts to increase spending and engagement.