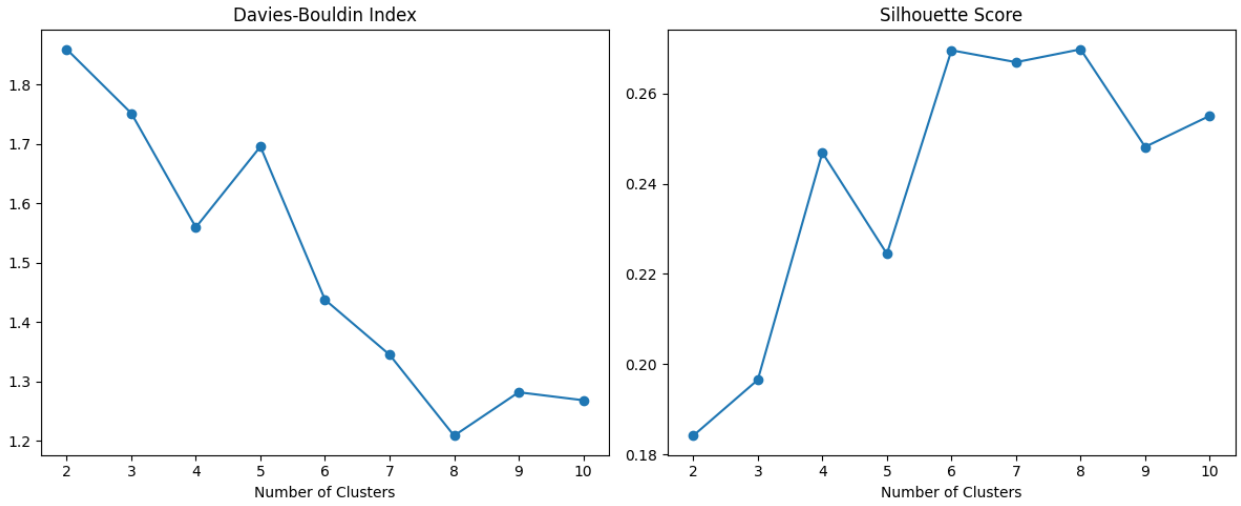
Clustering Report

1. Number of Clusters

- The analysis produced **4 clusters**, as seen in the clustering summary and visualization. This was determined based on evaluation metrics such as the Davies-Bouldin Index and Silhouette Score.

2. Clustering Evaluation Metrics

- **Davies-Bouldin Index (DB Index)**:
  - The optimal number of clusters was identified where the DB Index is minimized.
  - The DB Index for this clustering is **1.2**, indicating compact and well-separated clusters.
- **Silhouette Score**:
  - The Silhouette Score, which measures the separation distance between clusters, was also taken into consideration. Higher values suggest better-defined clusters.
  - Clusters in the range of **4-6** had higher scores, supporting the choice of 4 clusters.



3. Cluster Summary

Each cluster has been characterized based on aggregated metrics. Below is the summary:

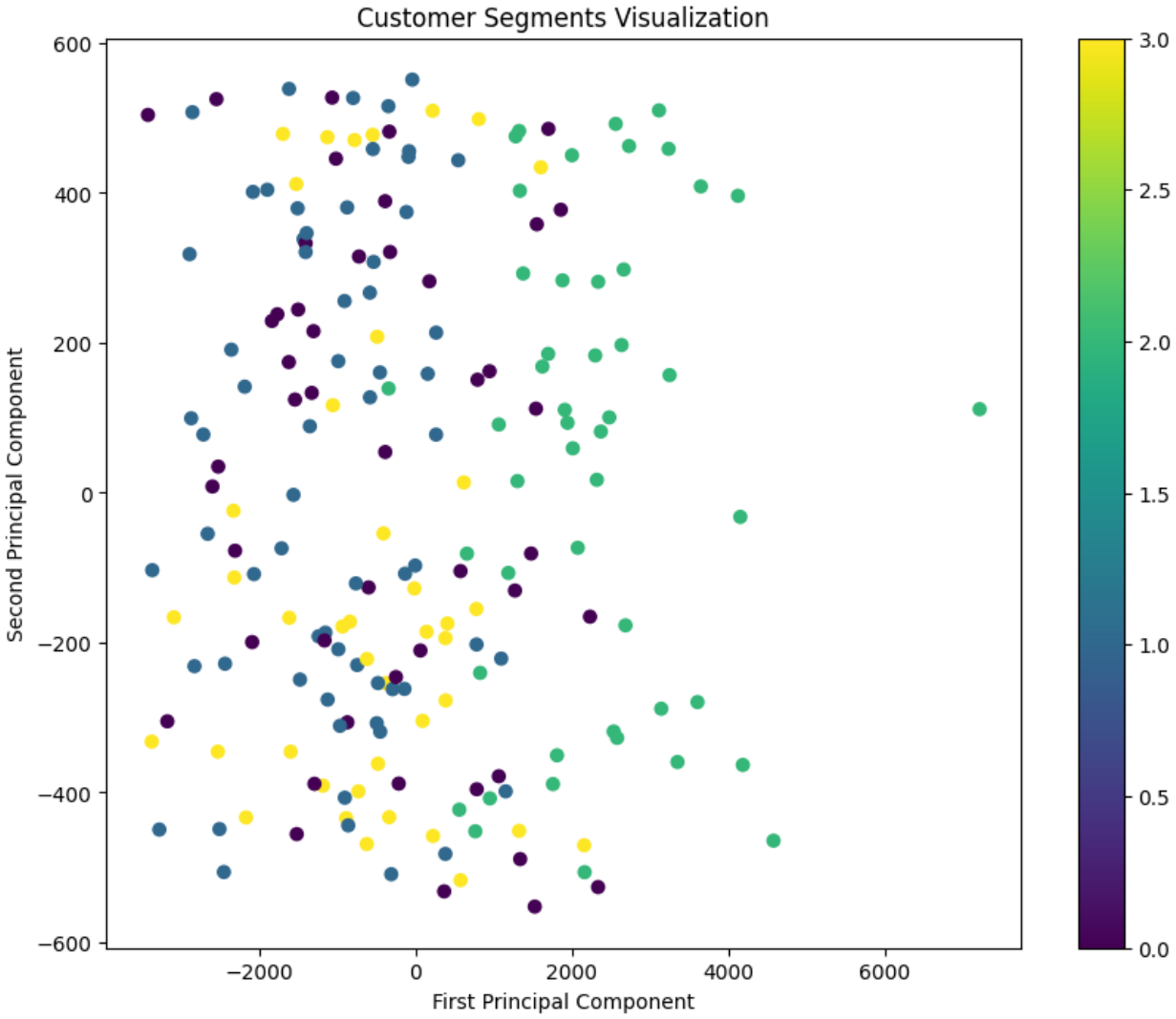
Cluster	TotalValue (Sum)	Transactions (Count)	Quantity (Sum)	Days Since Signup (Average)
0	3054.94	4.43	11.00	548.00
1	2405.67	3.84	9.01	551.36
2	5779.46	7.40	20.69	574.90
3	2890.35	4.78	11.29	411.98

4. Key Insights

- **Cluster 2**:
    - Represents the most valuable customer segment with the highest total value (5779.46), the most transactions (7.4), and the highest quantity of items purchased (20.69). However, their days since signup (574.90) are higher than Cluster 3.
- **Cluster 3**:
    - Has customers with moderately high value (2890.35) and a lower average time since signup (411.98), suggesting more recent activity.
- **Cluster 0 and 1**:
    - Represent smaller, less active customer segments, with lower total values, fewer transactions, and smaller quantities of items purchased.

5. Visualizations

- **Principal Component Analysis (PCA) Visualization**:
    - Customers are visually segmented into clusters based on the first two principal components.
    - Clear separations are seen, indicating effective clustering.
- **Davies-Bouldin Index and Silhouette Score Graphs**:
    - Show the performance of clustering across various cluster counts, confirming that the choice of 4 clusters balances both metrics.


Customer Segments Visualization
600
400
200
0
−200
−400
−600
Second Principal Component
−2000
0
2000
4000
6000
First Principal Component
3.0
2.5
2.0
1.5
1.0
0.5
0.0