Data Science Assignment: eCommerce Transactions Dataset

Task 3: Customer Segmentation / Clustering

Clustering Logic and Metrics:

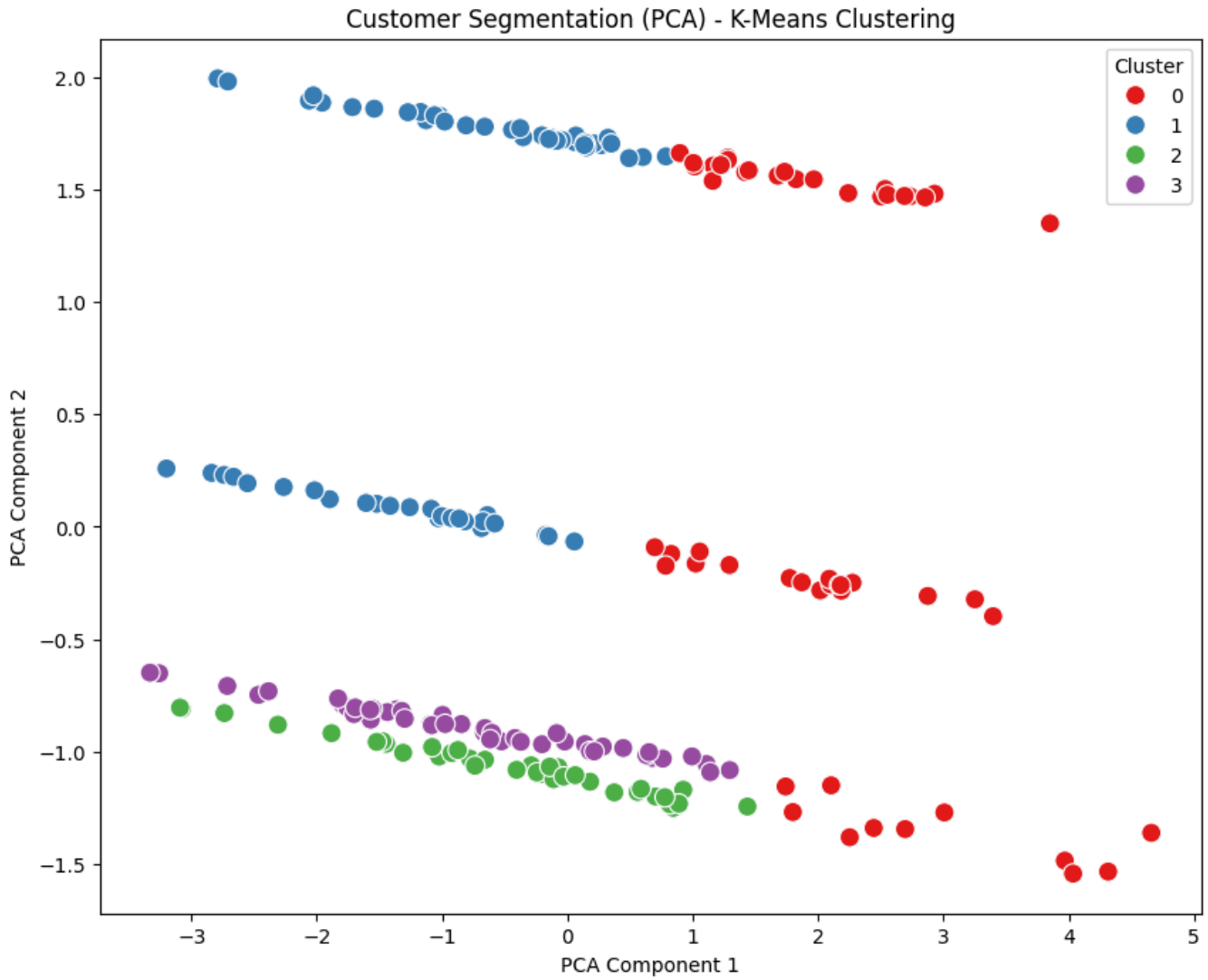
- Number of Clusters Formed: 4 clusters were created using the K-Means algorithm.
- Davies-Bouldin Index: The DB Index value is 0.9881, which suggests that the clusters are reasonably well separated, but there's room for improvement.
- Silhouette Score: The silhouette score is 0.396, which indicates moderate cohesion within clusters, but the clustering may not be very distinct. Values closer to 1 indicate better-defined clusters.

Visual Representation of Clusters:

- PCA Visualization: The scatter plot of the first two principal components (PCA1 and PCA2) shows how the customer clusters are distributed in a 2D space. Each point represents a customer, and the clusters are color-coded, allowing us to visualize the separation.
- Cluster Centers: The centers of the clusters in the original feature space are provided, giving insight into the typical customer profile for each cluster.

Cluster Summary Insights:

- Cluster 0: Customers in this cluster have the highest total transaction values and quantities. They also have the most transactions on average.
- Cluster 1: Customers in this cluster have relatively lower transaction values and quantities, with fewer total transactions.
- Cluster 2: These customers have moderately high total transaction values and quantities, with an average number of transactions.
- Cluster 3: This cluster is similar to Cluster 2 in terms of quantity and transaction count but has slightly higher transaction values.


Customer Segmentation (PCA) - K-Means Clustering
Cluster
0
1
2
3
PCA Component 2
PCA Component 1
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−3
−2
−1
0
1
2
3
4
5