Report : Customer Segmentation / Clustering

1.Number of Clusters Formed:

- The dataset was segmented into 4 distinct clusters, as identified using K-Means Clustering. These clusters are based on customer features including total spend, transaction count, average purchase value, and average quantity.

2.Cluster Metrics:

- **Davies-Bouldin Index (DBI):**
    1. The Davies-Bouldin Index for the 4 clusters is 1.0226.
    2. The Davies-Bouldin Index measures the average similarity ratio of each cluster with its most similar cluster. Lower values indicate better-defined clusters, and in this case, a value close to 1 suggests moderate separation and cohesion**.**
- **Silhouette Score:**
    1. The Silhouette Score is 0.3141 for the 4 clusters.
    2. The Silhouette Score ranges from -1 to 1, where higher values (close to 1) indicate that the samples are well-clustered, and lower values (close to -1) suggest poor clustering. A score around 0.3141 indicates moderate cluster separation.

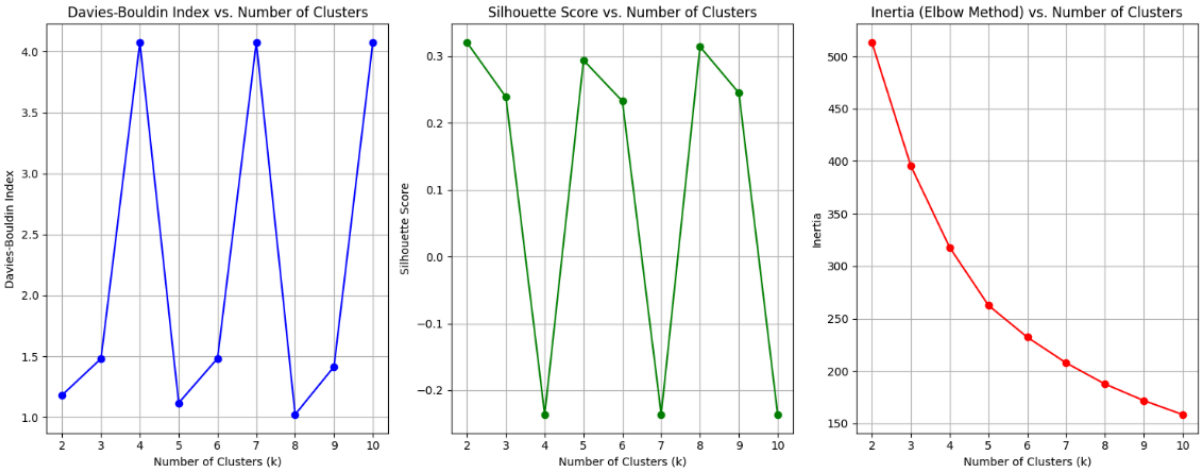
3.Cluster Sizes:

The number of samples in each cluster:

- **Cluster 0**: 53 samples
- **Cluster 1**: 89 samples
- **Cluster 2**: 39 samples
- **Cluster 3**: 19 samples

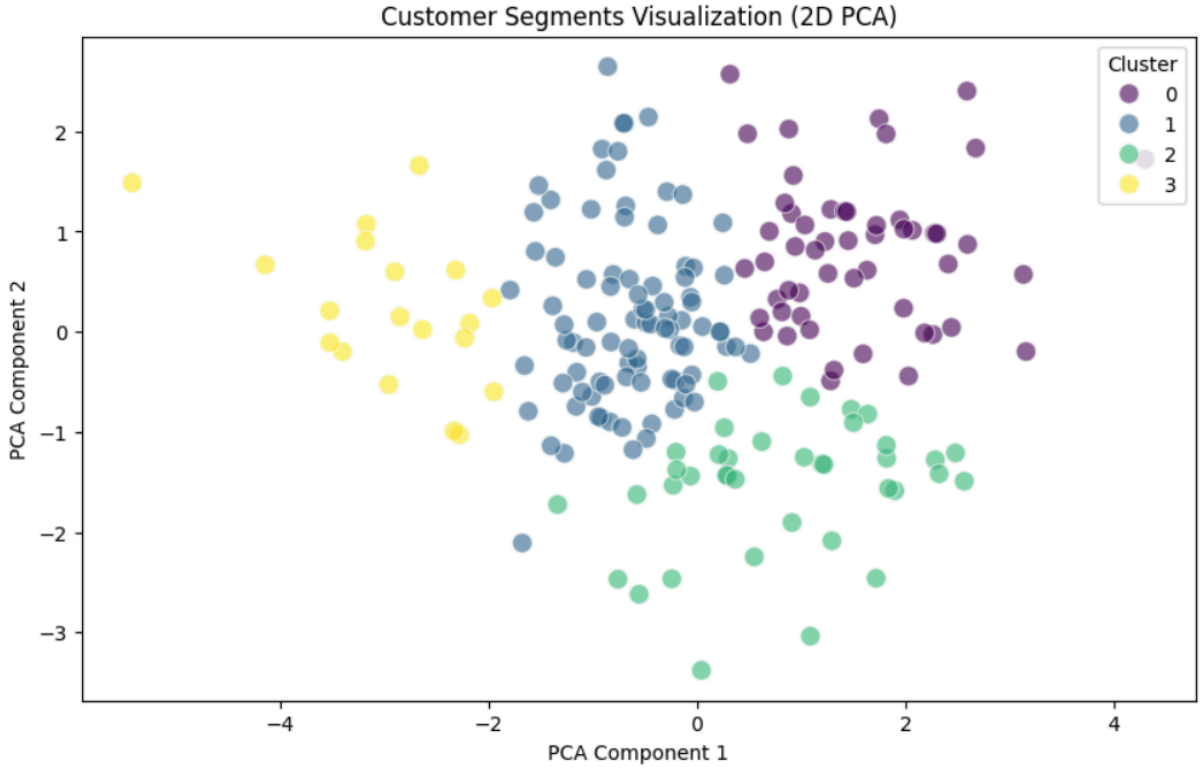
4.Visualization Results:

A. **Davies-Bouldin Index, Silhouette Score, and Inertia Plots:**
   These plots help in choosing the optimal number of clusters.
    - The Davies-Bouldin Index and Silhouette Score vs. cluster count (2-10) plots provide insights into how clustering performance varies as the number of clusters increases.
    - The Inertia (Elbow Method) plot helps to identify the point at which adding more clusters does not improve the model significantly.



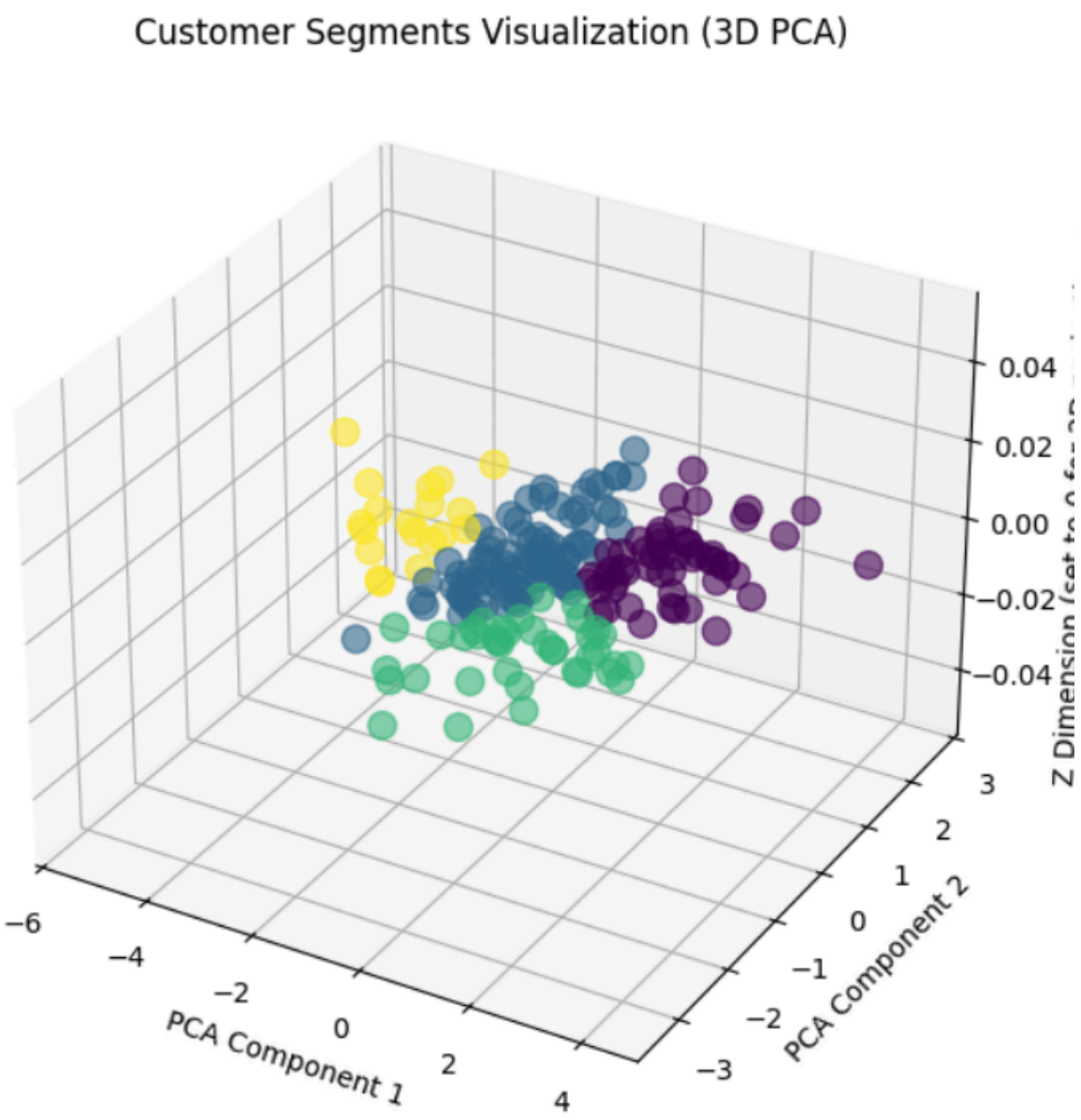
B. 2D Cluster Visualization (PCA-based):

The data was reduced to two dimensions using PCA and visualized as a scatter plot. Each point represents a customer, colored according to their assigned cluster. This plot provides an intuitive visual representation of the clustering.



C. 3D Cluster Visualization (PCA-based):

A 3D scatter plot was also created using the first two PCA components, providing an additional view of how well the clusters are distributed in a three-dimensional space.


Customer Segments Visualization (3D PCA)
PCA Component 1
PCA Component 2
0.04
0.02
0.00
−0.02
−0.04
−6
−4
−2
0
2
4
−3
−2
−1
0
1
2
3