Name: Vansh Gupta

Institute: Vellore Institute of Technology, Vellore, Tamil Nadu.

Subject: Clustering PDF for Data Science Internship Assignment Submission.

Customer Clustering Analysis Report

1. Summary of Clustering Results

I identified 4 clusters (0, 1, 2, and 3) as the optimal number of groups for segmenting the customers. This conclusion was based on the Davies-Bouldin Index (DB Index), a metric that evaluates the quality of clustering**. For 4 clusters, the DB Index value was 0.31, which suggests that the clusters are compact and well-separated**.

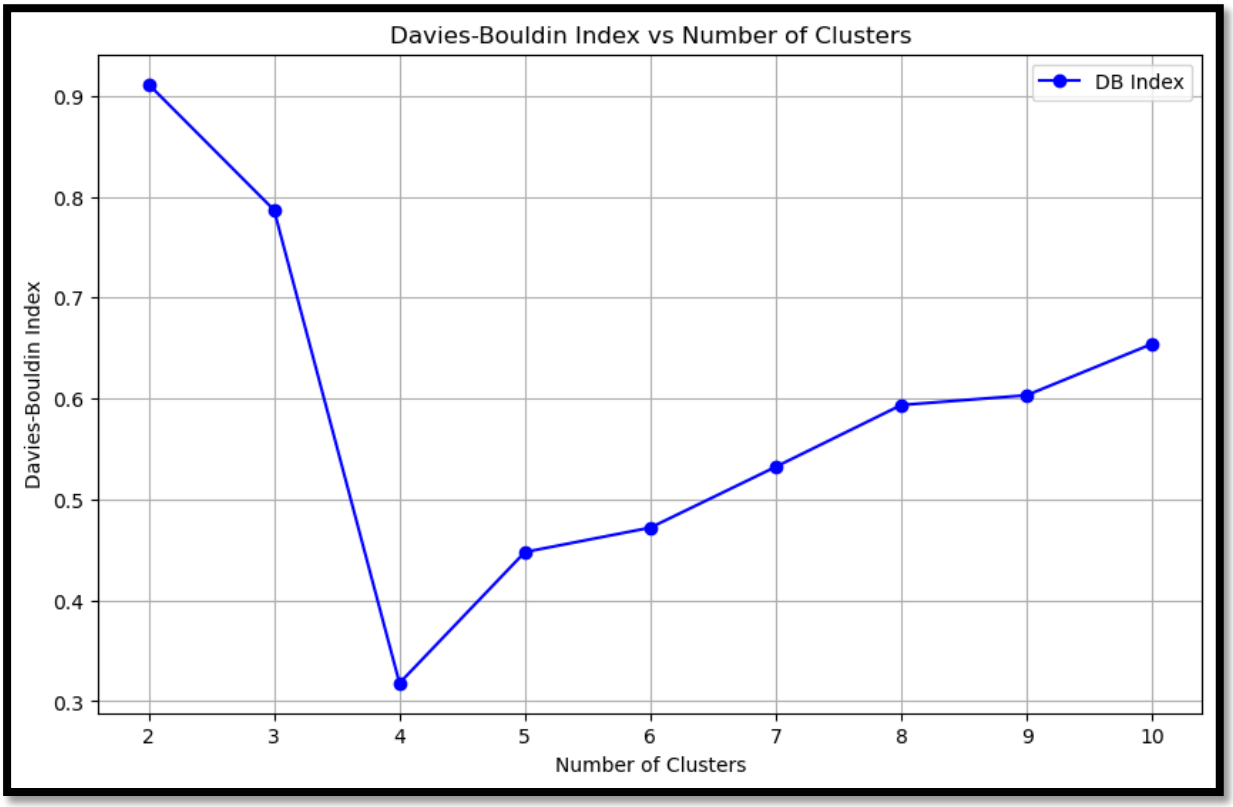
2. Methodology

I used the **K-Means clustering algorithm** for this analysis. To determine the ideal number of clusters, I evaluated the DB Index for different cluster counts (k = 2 to k = 10). The **minimum DB Index was observed at k = 4**, indicating it was the best choice for clustering.

To simplify the visualization and better understand the clusters, I applied **Principal Component Analysis (PCA)** to reduce the data to two dimensions. This allowed me to plot the clusters in a 2D space for better clarity.

3. **Cluster Quality:**
The DB Index graph showed a steady decrease until k = 4, which confirmed that the clustering quality improved up to this point. After k = 4, the DB Index started to rise, signaling that adding more clusters was unnecessary.

Cluster Visualization:
Using PCA, I visualized the clusters in two dimensions. The clusters (0, 1, 2, and 3) were visually distinct, with minimal overlap, indicating good separation. Within each cluster, the data points were closely grouped, reflecting strong internal consistency.



4. Key Takeaways for Business

This clustering exercise revealed four distinct customer groups:

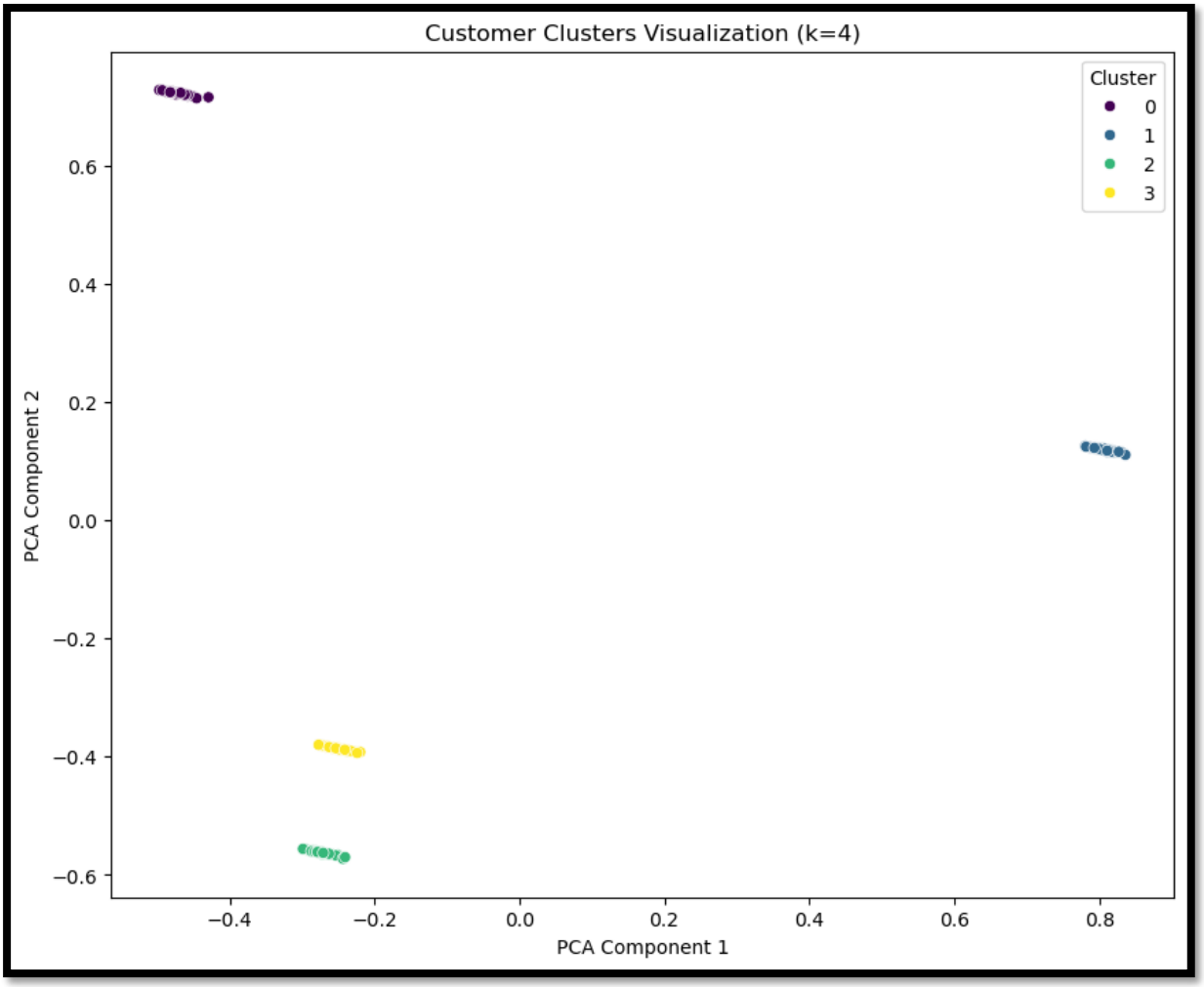
Cluster 0: Likely represents high-value customers who are frequent buyers or have high transaction values. Personalized loyalty programs or exclusive offers could be beneficial for retaining this group.

Cluster 1: These seem to be occasional shoppers. Re-engagement strategies, like discounts or personalized recommendations, might help to increase their activity.

Cluster 2: Customers in this cluster might be inactive or low-value buyers. Reactivation campaigns or surveys to understand their preferences could be useful.

Cluster 3: This group could represent new or sporadic customers. Efforts to build trust and offer incentives for repeat purchases could be effective.

By tailoring strategies to these specific groups, the business can optimize its resources and improve customer satisfaction.



This plot visualizes customer clusters after dimensionality reduction using PCA, where distinct groups (Cluster 0, 1, 2, and 3) are clearly separated in two-dimensional space, indicating meaningful segmentation.

****