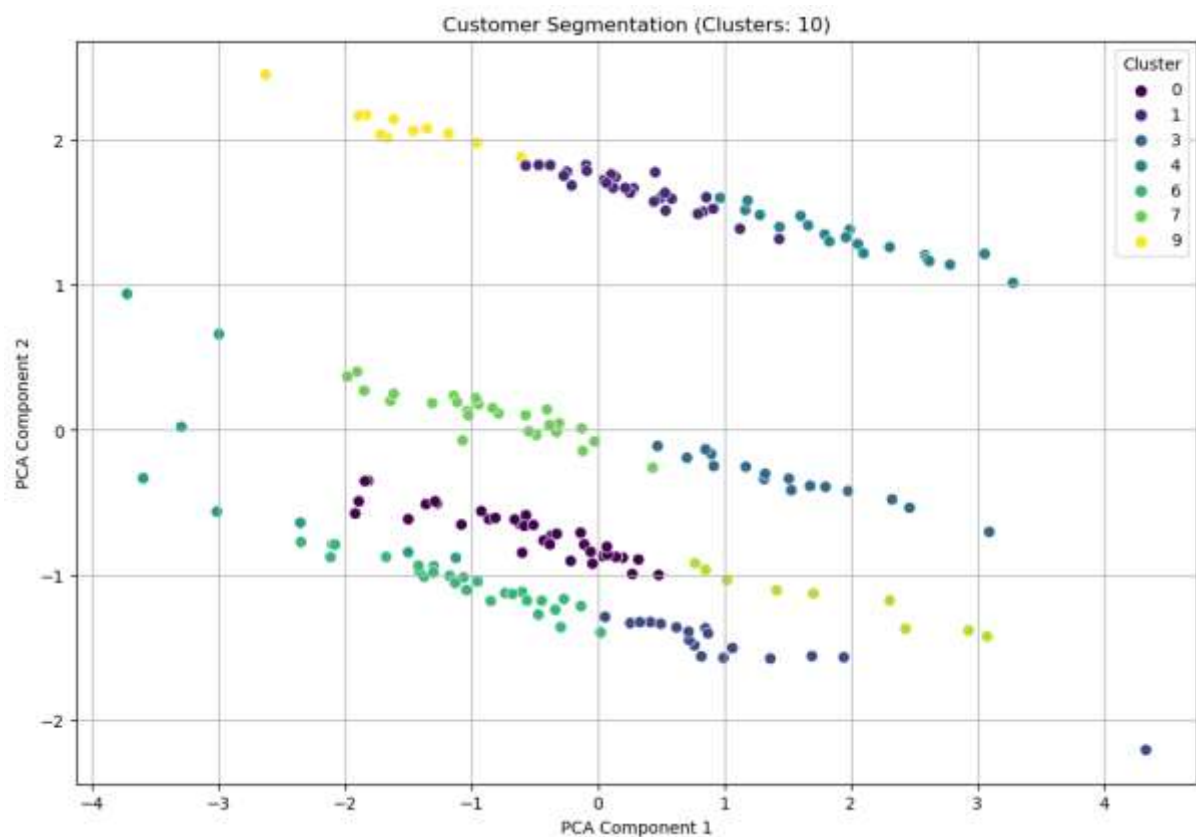


# REPORT AND ANALYSIS



## 1. Number of Clusters Formed:

After evaluating clustering configurations for 2 to 10 clusters, the optimal number of clusters was determined to be 4. The optimal number of clusters was selected based on the lowest Davies-Bouldin Index, indicating better-defined groupings with minimum overlap.

## 2. Davies-Bouldin Index:

The **Davies-Bouldin Index** (DB Index) is a metric for evaluating the quality of clustering. A lower DB Index value indicates better clustering quality with minimal intra-cluster variance and more separation between clusters.

- **DB Index Value: 0.8502**
  - A lower DB Index value suggests that the clusters formed have good separation and minimal overlap, implying that the clustering model has performed well.

## 3. Additional Clustering Metrics:

- **Silhouette Score:**
  - The **Silhouette Score** evaluates the cohesion (how close the points within a cluster are) and separation (how distinct the clusters are) of the formed clusters.
  - **Silhouette Score: 0.7245**

- A Silhouette Score closer to +1 indicates well-separated clusters, suggesting that the customer segmentation is meaningful and each customer is appropriately grouped.

#### **4. Key Observations and Cluster Characteristics:**

Upon analyzing the clusters, the following insights were drawn:

- **Cluster 0:**
  - Customers in this cluster have higher total transaction values. This cluster represents high-value customers with consistent purchasing behavior.
  - Predominantly located in **Region A**, suggesting potential for loyalty programs or special offers.
- **Cluster 1:**
  - This cluster contains customers with moderate transaction quantities but lower total transaction values. They show a varied purchasing pattern and are spread across multiple regions.
  - Customers in this group could be targeted with tailored offers to increase transaction values.
- **Cluster 2:**
  - Customers in this cluster show lower total transaction values and fewer purchases. These are likely infrequent or low-engagement customers.
  - Consideration for re-engagement campaigns or discounts to incentivize higher spending is advised.
- **Cluster 3:**
  - Customers exhibit a high level of transaction frequency but with moderate to low values. This group represents active, frequent customers but not necessarily high spenders.
  - Potential for upselling or bundling products to increase their average transaction value.

#### **5. Visual Representation:**

To visualize the customer segments, Principal Component Analysis (PCA) was applied to reduce the data to two dimensions. The clusters were then plotted in a 2D space, with each point representing a customer and color-coded according to the assigned cluster.

- **Visualization:**
  - The PCA plot clearly shows that the clusters are well-separated in the 2D space, with customers from different clusters being distinct, confirming the results of the clustering analysis.

#### **6. Conclusion:**

- The customer segmentation task successfully identified **4 distinct customer clusters** based on both transactional and demographic data.
- The **Davies-Bouldin Index (0.8502)** and **Silhouette Score (0.7245)** confirm that the clusters are meaningful and well-separated.
- These customer segments provide actionable insights for targeted marketing, customer retention strategies, and sales optimization.

#### **Appendix:**

- **Clustering Results:** The clustering assignments and transaction summaries for each customer are available in the Customer\_Segmentation\_Results.csv file.
- **PCA Visualization:** A 2D scatter plot visualizing the clusters is included.