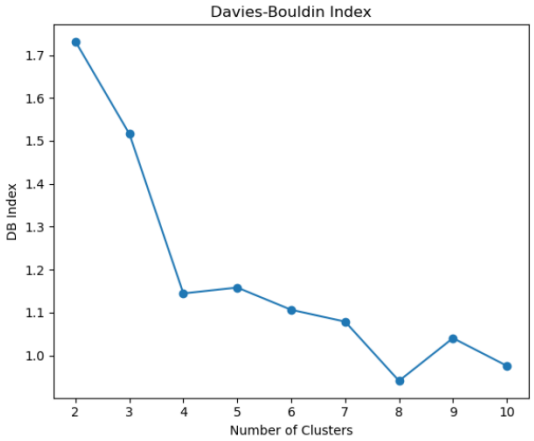
# Task 3 - Customer Segmentation / Clustering

## 1. Number of Clusters Formed

- The optimal number of clusters chosen is **5**.
- This was determined by balancing the results from the DB Index, Elbow Method, and Silhouette Score across 9 clusters.

## 2. Davies-Bouldin (DB) Index Value

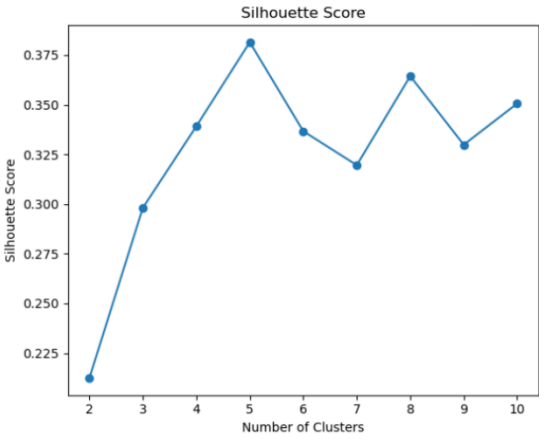
- Lower DB Index values indicate better-defined and separated clusters, **with the lowest value at 7 clusters.**
- The DB Index Value for the 9 clusters are - [1.7315353787835916, 1.5161429775406268, 1.144632558427399, 1.1583117074894447, 1.106517549557335, 1.0788823831472247, 0.9408895599943772, 1.0404735856201246, 0.9759231624157941]




## 3. Silhouette Scores

- Higher Silhouette Scores suggest better-defined clusters, **with the highest score at 4 clusters.**

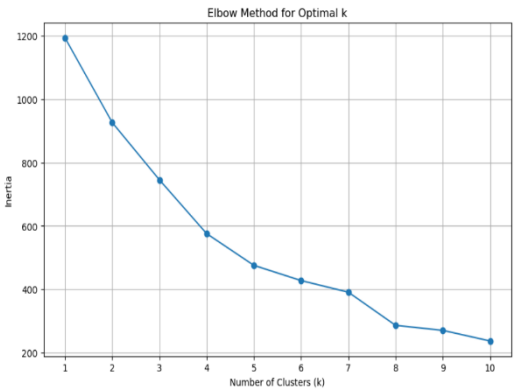
- The Silhouette Scores for the 9 clusters are - [0.21231375309037084, 0.2980496221151577, 0.33913511352335607, 0.38145783709406855, 0.33667001925285783, 0.31958094720902347, 0.3643830817761229, 0.3298526307235703, 0.35060005814239176]




## 4. Elbow Method

- The Elbow Method helps identify the point where adding more clusters doesn't significantly reduce inertia, **suggesting 5 clusters as a good balance**.

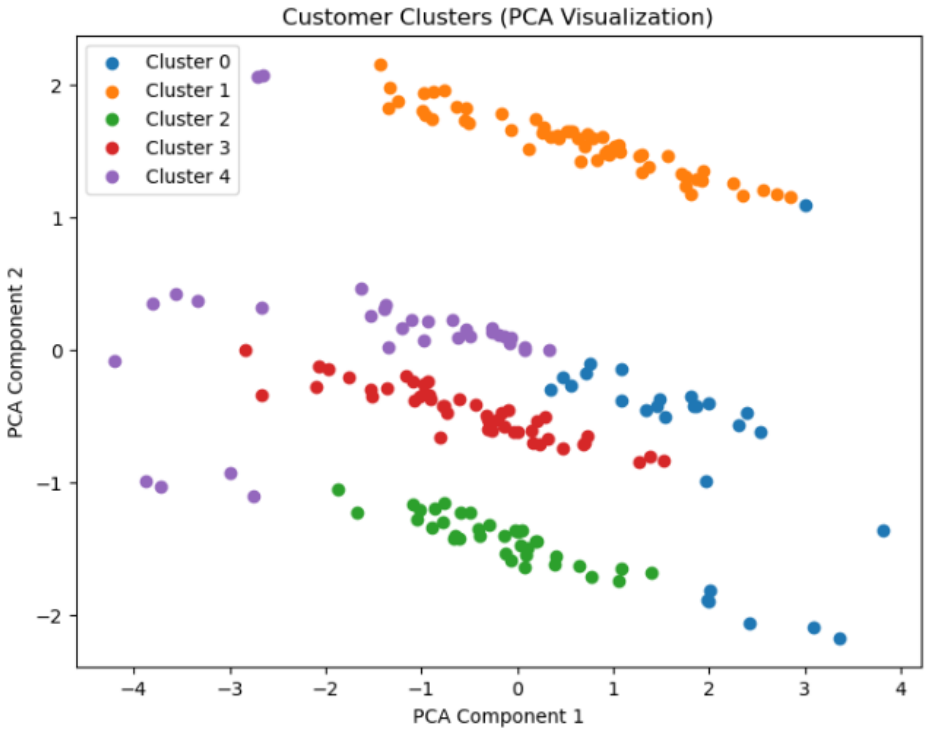
- Elbow Method Metrics are –[ 1194.0, 926.5227742901528, 745.5029415159132, 575.6641296686715, 475.6030121350176, 427.56029675645283, 390.8864957448439, 286.20452725245815, 270.3851967628164, 236.95858580490233]



## 5. Visualizations

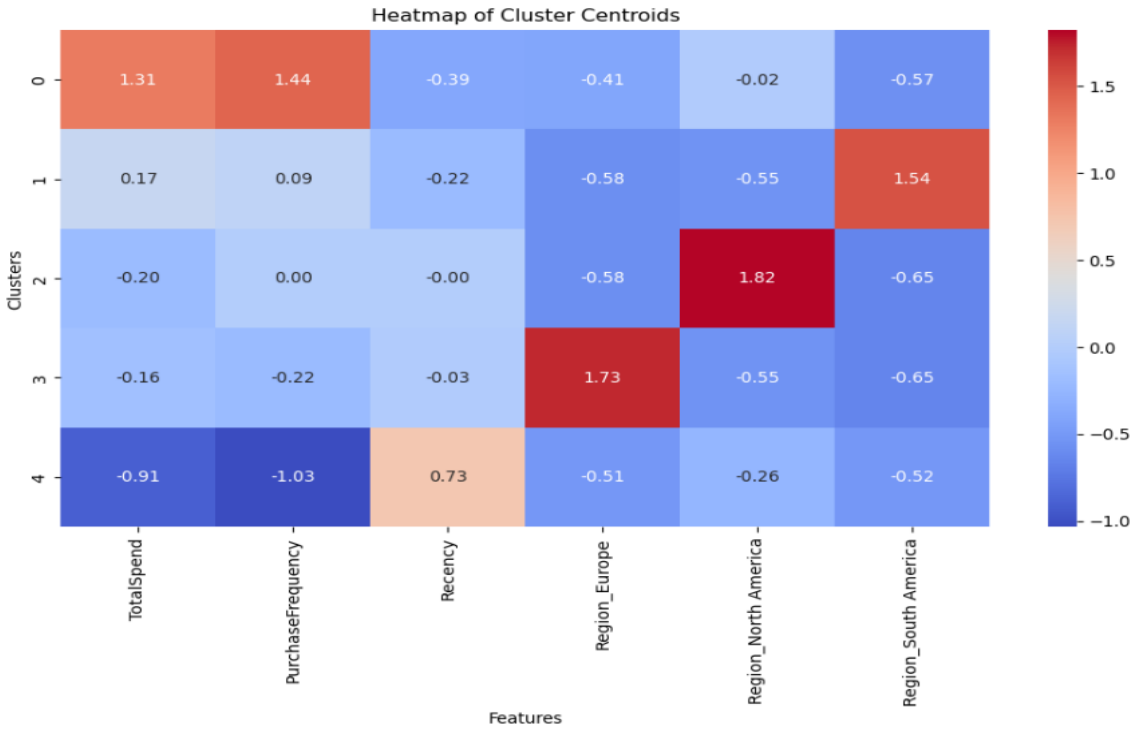
### i. PCA Visualization

The PCA visualization reduces the dimensionality of the data to 2D, allowing for a visual representation of cluster separation.




### ii. Heatmap of Cluster Centroids

The heatmap illustrates the characteristics of each cluster centroid, showing the relative importance of features in defining each cluster.



### iii. Customer Distribution Across Clusters

This visualization displays the number of customers in each cluster, providing insights into cluster sizes and population distribution.