Task 3: Customer Segmentation / Clustering

Choosing the Clustering Algorithm:

- Applied KMeans clustering, a popular unsupervised algorithm.
- Tested different numbers of clusters (2-10).

Optimal Number of Clusters:

- 10 clusters were selected based on the lowest Davies-Bouldin Index (DBI) score.
- The DBI score for 10 clusters was the lowest, indicating well-separated and distinct clusters.

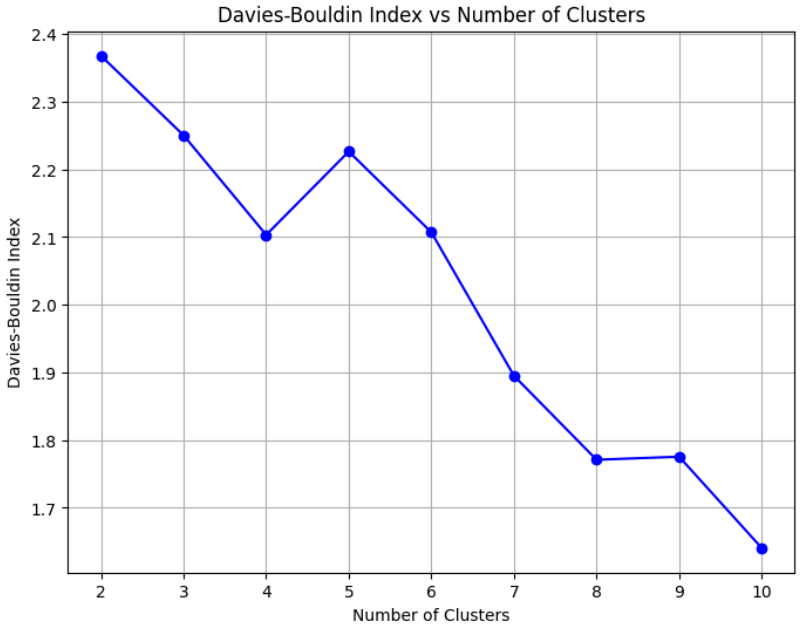
Evaluation Metrics:

- **Davies-Bouldin Index (DBI):** Achieved the lowest DBI score for 10 clusters, confirming the best cluster separation.
- **Silhouette Score:** Evaluated the compactness and separation of clusters. The score for 10 clusters was satisfactory.

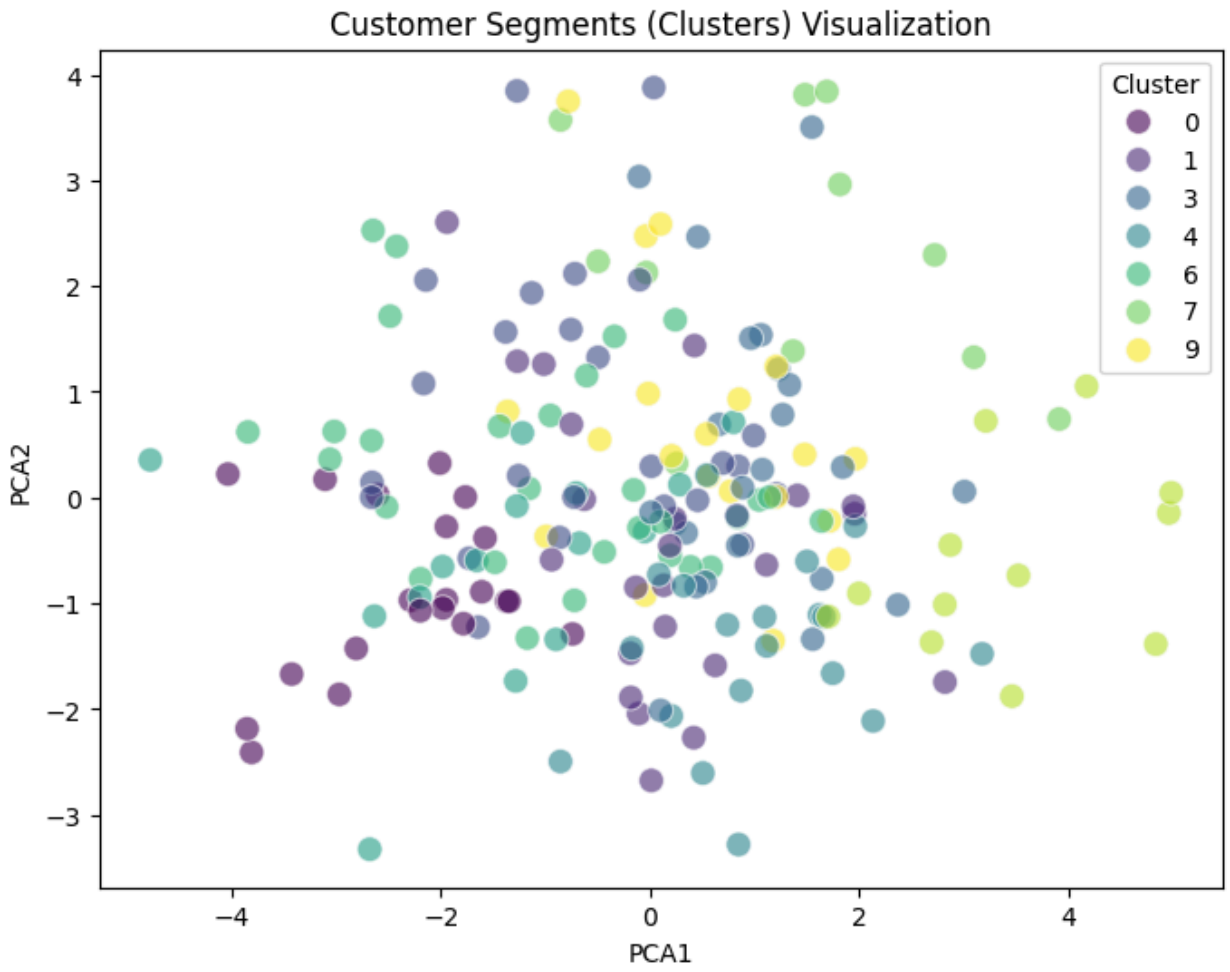
Visualization:

Visualized the clusters using PCA for dimensionality reduction (2D plot) to observe the separation between customer segments.

Results:




Silhouette Score vs Number of Clusters
Silhouette Score
0.140
0.135
0.130
0.125
0.120
0.115
2
3
4
5
6
7
8
9
10
Number of Clusters



Interpreting DBI:

- Lower DBI values indicate better clustering. It suggests that clusters are compact and well-separated.
- Higher DBI values indicate poor clustering, meaning the clusters are either too spread out or too close together.