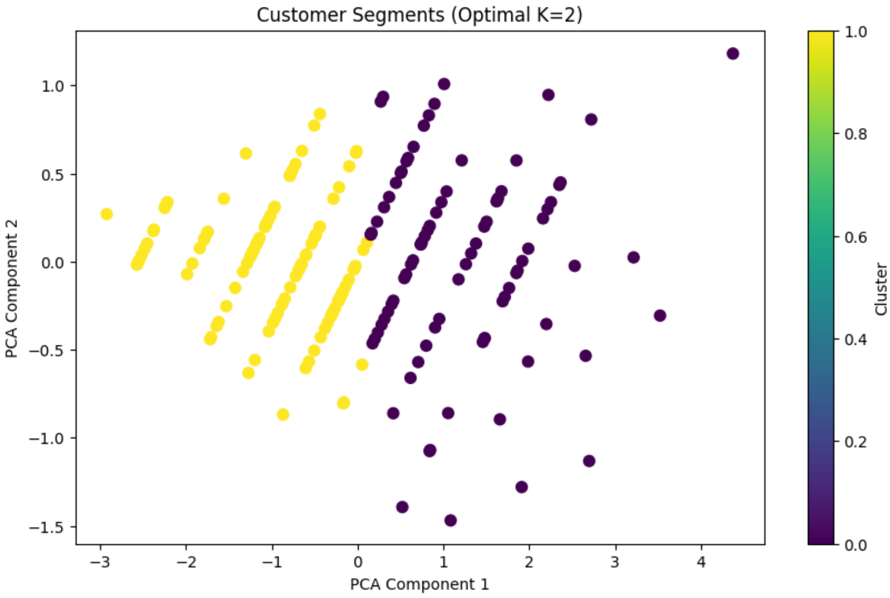
Clustering Results Report

1. Number of Clusters:

- The K-Means clustering algorithm was performed for k values ranging from 2 to 10. We computed the clustering metrics for each value of k, such as the **Davies-Bouldin Index** and the **Silhouette Score**.



2. Clustering Metrics:

- **Davies-Bouldin Index (DBI)**: This index measures the average similarity ratio of each cluster with the most similar one. Lower values indicate better-defined clusters.
- **Silhouette Score**: This metric evaluates how well-separated the clusters are. A score closer to +1 indicates that the points are well-clustered, while scores closer to -1 suggest that clusters are poorly defined.

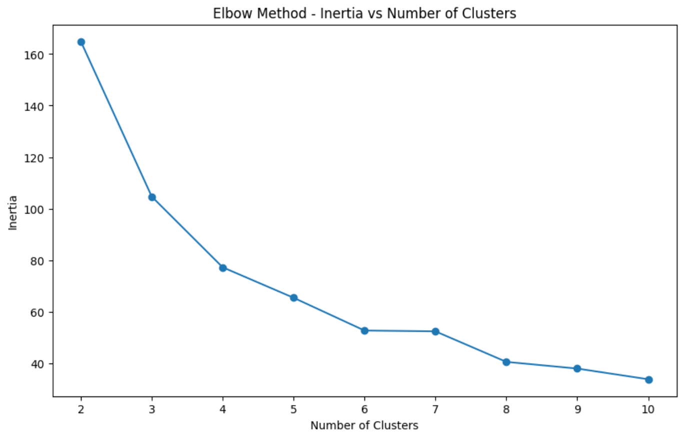
3. Clustering Results:

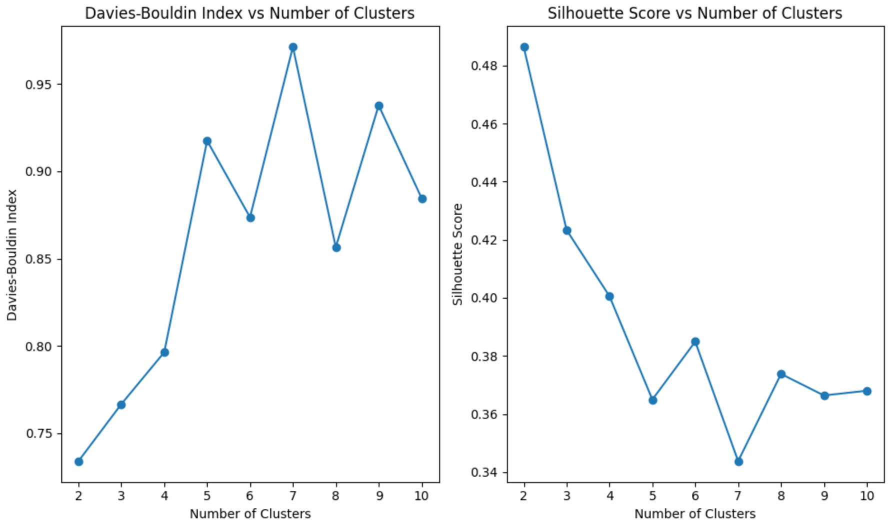
Here are the results of the clustering process, including the Davies-Bouldin Index and Silhouette Score for each number of clusters (k):

Number of Clusters	Davies-Bouldin Index (DBI)	Silhouette Score
2	0.734	0.486
3	0.767	0.423
4	0.796	0.401
5	0.918	0.365
6	0.874	0.385
7	0.971	0.344
8	0.857	0.374
9	0.938	0.366
10	0.885	0.368

4. Interpretation of Results:

- **Davies-Bouldin Index (DBI)**:
  - The DBI values tend to increase as the number of clusters increases. The lowest DBI value is for k=2 (0.734), which suggests that this might be the best number of clusters in terms of separation between clusters. After k=2, the DBI values increase, indicating that adding more clusters might lead to worse separation between the clusters.
- **Silhouette Score**:
  - The highest Silhouette Score is also observed for k=2 (0.486). While this is not a very high value, it indicates that the clusters are somewhat well-separated. For higher values of k, the Silhouette Score decreases, suggesting that the clusters are becoming less distinct and the model might be overfitting.





5. Optimal Number of Clusters:

- **Based on DBI**: The optimal number of clusters appears to be k=2 as it gives the lowest DBI value, indicating the best cluster separation.
- **Based on Silhouette Score**: Similarly, the highest Silhouette Score also corresponds to k=2, suggesting that the clusters are more distinct when there are only two clusters.

6. Conclusion:

- The **optimal number of clusters** for this dataset is **2** based on both the Davies-Bouldin Index and the Silhouette Score. This indicates that the customer segments are best represented by two distinct clusters, and further increasing the number of clusters does not improve the clustering quality.
- It is worth noting that while the Silhouette Score for k=2 is moderate (0.486), it suggests that the clusters are somewhat well-separated but could still benefit from further exploration.

7. Recommendations:

- **Exploration of Additional Features**: Consider including more features (e.g., customer demographics, product categories, etc.) to see if that can improve clustering and cluster separation.
- **Consider Alternative Algorithms**: While K-Means performed well for this task, you may explore other clustering methods such as **DBSCAN** or **Agglomerative Hierarchical Clustering** to see if they yield better separation of customer segments.