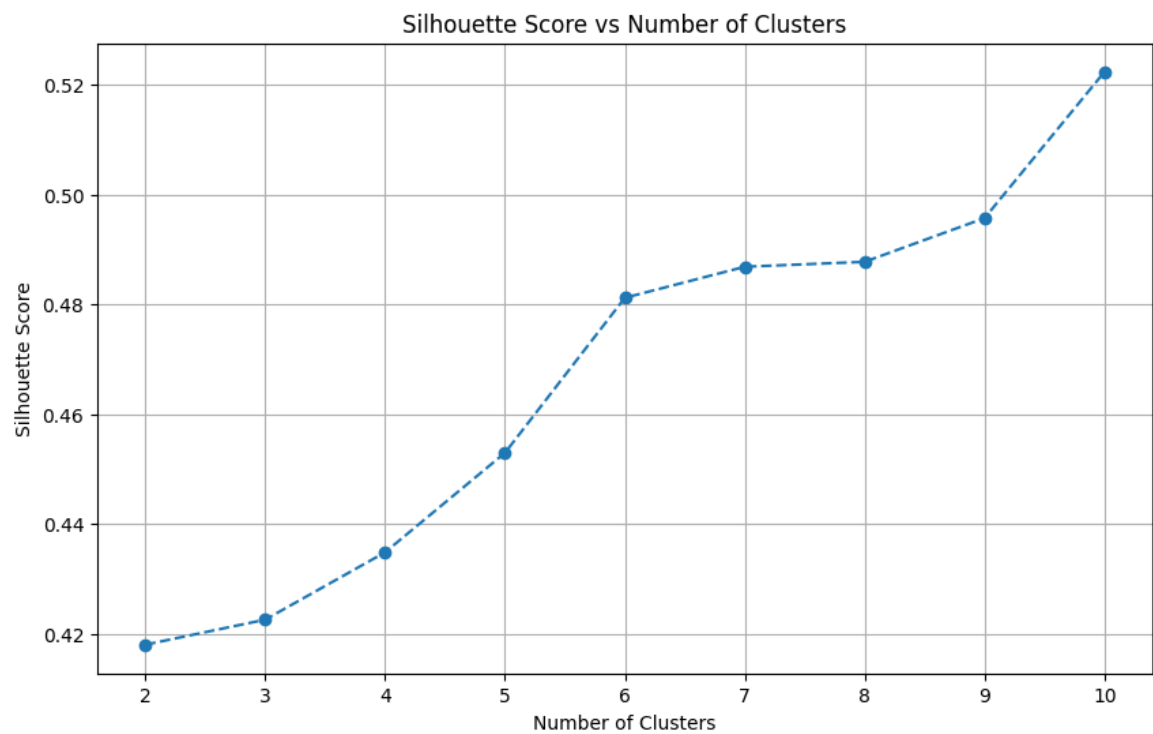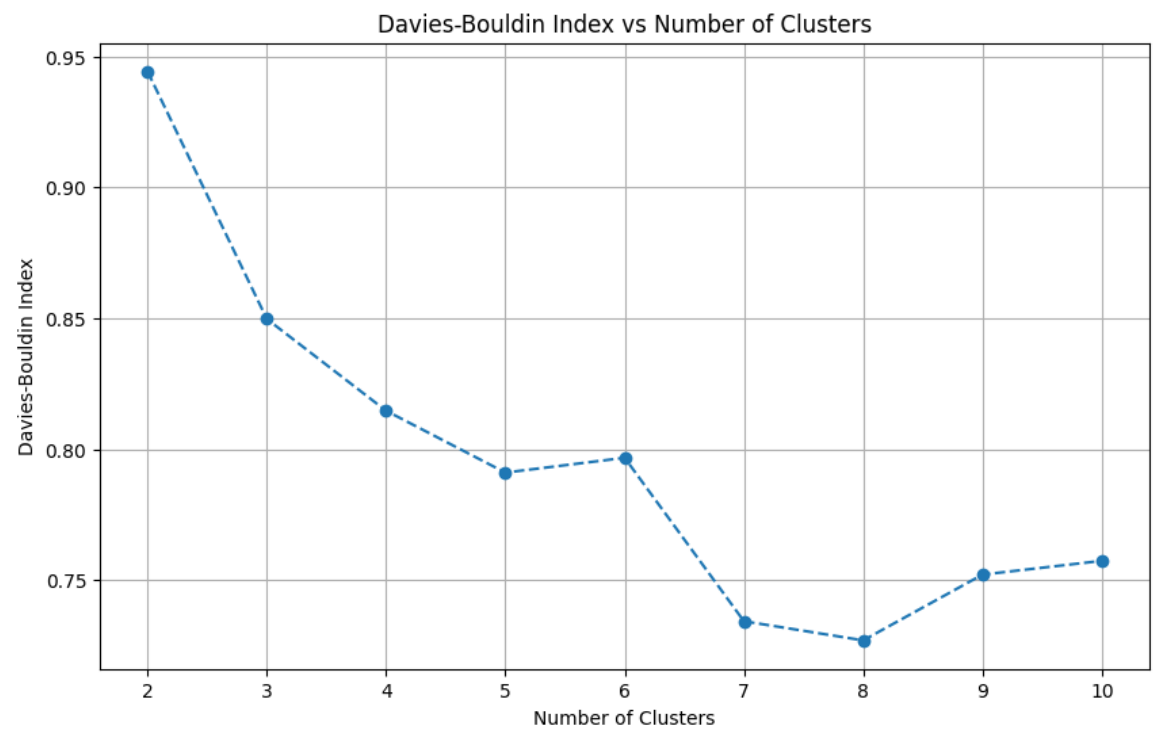# Report on Clustering Results

Graphs:

1. Number of Clusters Formed

- **Optimal Number of Clusters: 8**

- **Why 8 is Optimal:**
  - Based on the Davies-Bouldin Index (DB Index) graph, the DB Index reaches a minimum value at 8 clusters, which indicates an optimal balance between intra-cluster cohesion and inter-cluster separation.
  - The Silhouette Score graph shows a consistent increase, with higher scores indicating better-defined clusters. While 10 clusters show the highest silhouette score, the DB Index favors 8 clusters, striking a balance between compactness and separability.
  - The PCA visualization further validates this choice, showing clear, distinct clusters in a 2D projection. Each cluster is well-separated, with minimal overlap.

2. Davies-Bouldin Index (DB Index)

- **DB Index Value for 8 Clusters: 0.7271**

- **DB Index Range:**
  - The DB Index typically ranges from 0 (ideal clustering) to infinity. Lower values are better, as they represent:
    - High intra-cluster similarity (points within the same cluster are close to each other).
    - Low inter-cluster similarity (clusters are distinct and well-separated).
  - Our DB Index value of 0.7271 is significantly low, indicating high-quality clustering for the selected 8 clusters.

3. Other Relevant Clustering Metrics

Customer Segmentation Clusters (PCA Projection)
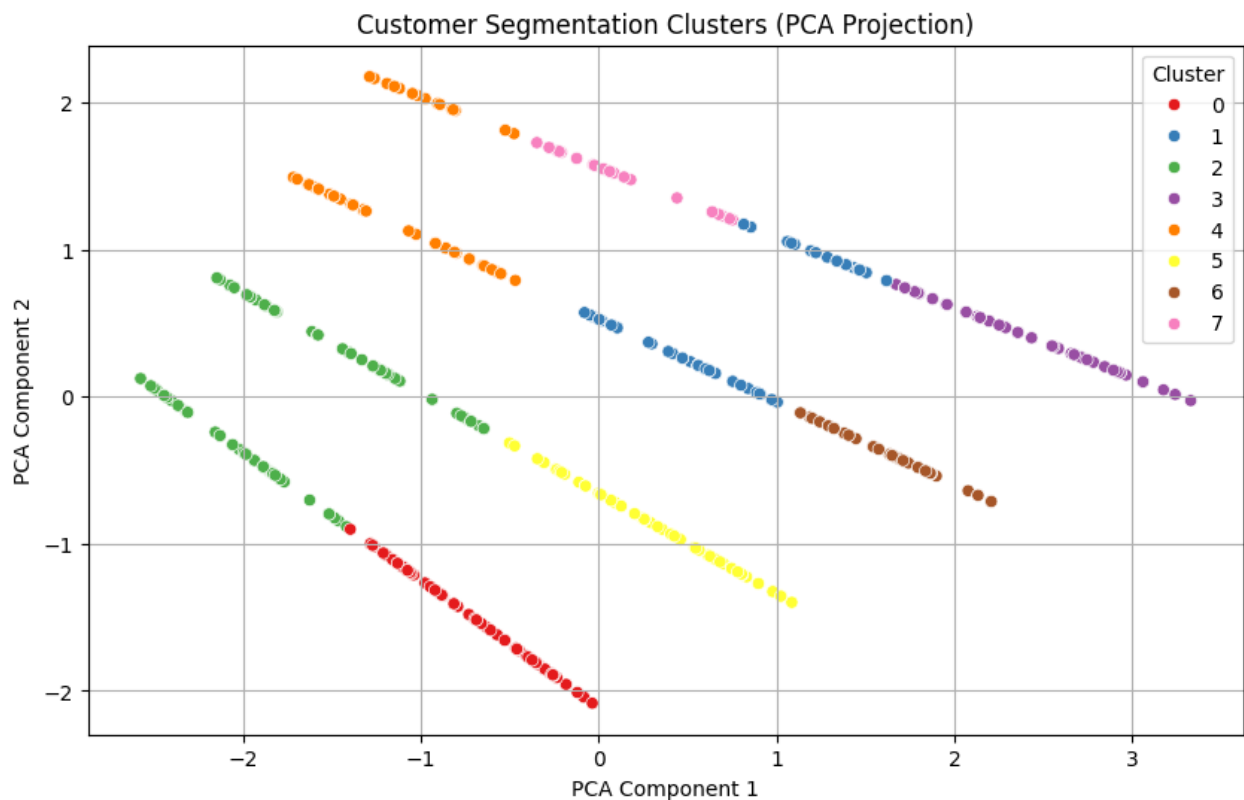
- **Silhouette Score:**
  - Value for 8 Clusters: 0.4927 (moderate to good clustering).
  - Range:
    - The silhouette score ranges from -1 to 1:
      - 1: Perfectly compact and well-separated clusters.
      - 0: Overlapping clusters or points near cluster boundaries.
      - -1: Poor clustering where points are closer to other clusters.
  - Why Include Silhouette Score:
    - It complements the DB Index by evaluating how well each point fits within its cluster versus others.
    - While our score of 0.4927 indicates moderately well-separated clusters, it supports the DB Index in confirming the quality of the clustering.
- **Cluster Compactness and Separation:**
  - The PCA visualization shows clear separability of the clusters in 2D space, demonstrating that 8 clusters are distinct and compact.

4. Explanation of Hyperparameters

- **Range of K (Clusters):**
  - The range of 2–10 clusters was selected based on the domain context and practical considerations of interpretability. Too few clusters risk over-generalizing, while too many might overfit or split natural groupings.
  - Within this range, the DB Index and Silhouette Score effectively identified the optimal number of clusters.
- **Scaling:**
  - Features were scaled using StandardScaler to normalize numerical values and prevent dominance of features with larger ranges.

Summary

- The clustering process successfully identified 8 clusters as the optimal number, supported by a low DB Index (0.7271) and moderate silhouette score (0.4927).
- The DB Index minimum value and clear cluster separations in PCA visualization validate the clustering quality.
- These metrics and visualizations ensure a balanced evaluation of intra-cluster cohesion and inter-cluster separation, delivering actionable insights into customer segmentation.