CLUSTERING RESULTS REPORT

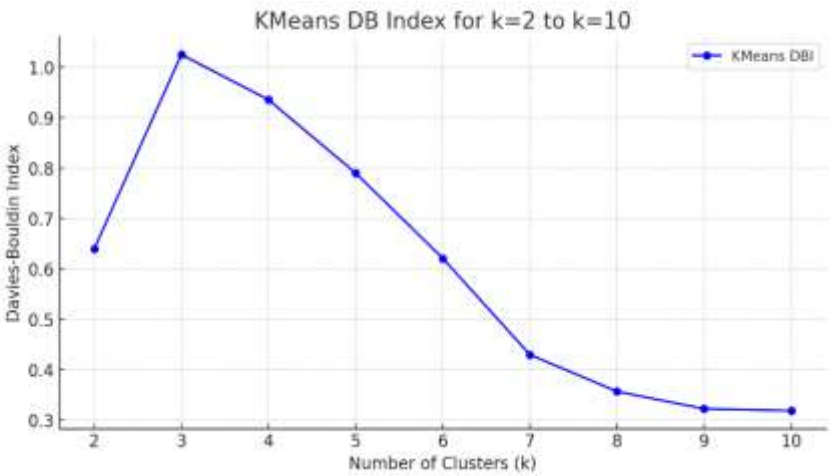
This report presents the clustering performance of KMeans, Hierarchical and DBSCAN Clustering algorithms on the customer data, using two important evaluation metrics: **Davies-Bouldin Index** and **Silhouette Score**. The results cover the clustering performance for different values of k (number of clusters), from 2 to 10.

KMeans Clustering Results

Davies-Bouldin Index for KMeans (k=2 to k=10):

The **Davies-Bouldin Index (DBI)** is a metric used to evaluate the clustering quality, where a lower score indicates better clustering. Below are the DBI values for different values of k:

k (Number of Clusters)	DB Index
2	0.6396
3	1.0255
4	0.9357
5	0.7903
6	0.6210
7	0.4298
8	0.3566
9	0.3223
10	0.3183

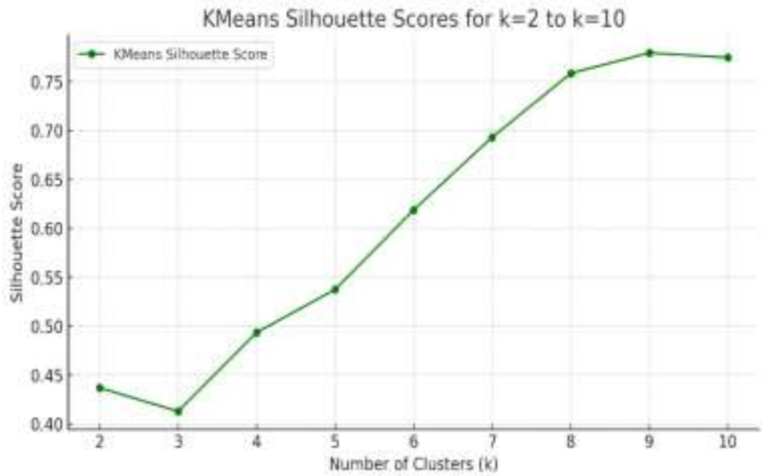


- **Interpretation**: As the number of clusters increases from k=2 to k=10, the Davies-Bouldin Index steadily decreases, with the lowest value observed at k=10 (0.3183). This indicates that clustering quality improves as the number of clusters increases. The **DBI** reaches its minimum at k=10, suggesting that this is the most well-separated set of clusters.

Silhouette Scores for KMeans (k=2 to k=10):

The **Silhouette Score** measures how similar each point is to its own cluster (cohesion) compared to other clusters (separation). The values range from -1 to +1, with higher values indicating better clustering.

k (Number of Clusters)	Silhouette Score
2	0.4369
3	0.4128
4	0.4935
5	0.5373
6	0.6188
7	0.6932
8	0.7585
9	0.7794
10	0.7747



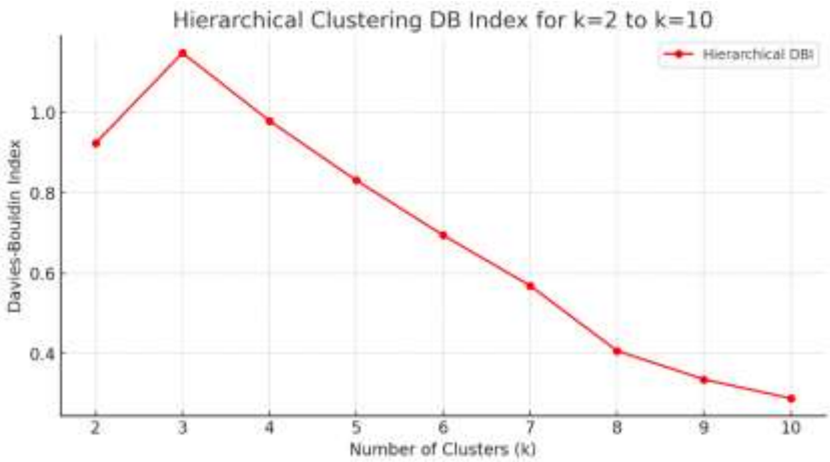
- **Interpretation**: The **Silhouette Score** increases as the number of clusters grows, peaking at k=9 with a value of **0.7794**. This suggests that the clustering becomes more cohesive and well-separated as the number of clusters increases, with the highest score at k=9. However, the Silhouette Score slightly decreases at k=10, indicating that adding more clusters beyond this point does not significantly improve clustering quality.

Hierarchical Clustering Results

Davies-Bouldin Index for Hierarchical Clustering (k=2 to k=10):

The **Davies-Bouldin Index (DBI)** for hierarchical clustering shows the following trend for different values of k:

k (Number of Clusters)	DB Index
2	0.9229
3	1.1478
4	0.9791
5	0.8310
6	0.6946
7	0.5691
8	0.4072
9	0.3361
10	0.2887

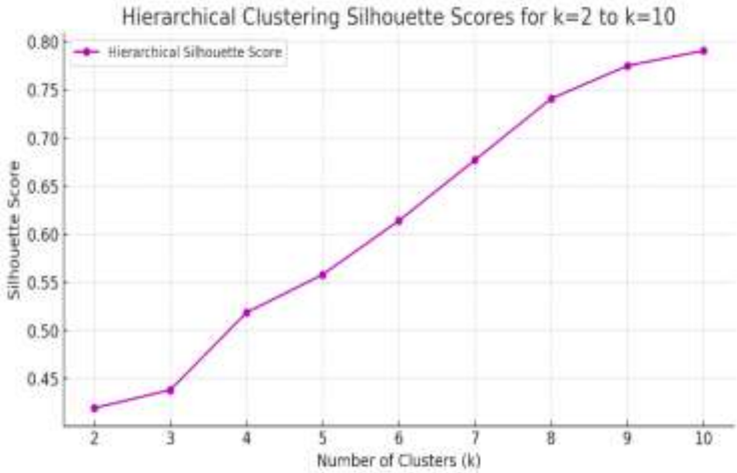


- **Interpretation**: Similar to KMeans, the **Davies-Bouldin Index** for hierarchical clustering decreases as the number of clusters increases, reaching its minimum value at k=10 (0.2887). This indicates that as more clusters are formed, the clusters become increasingly well-separated.

Silhouette Scores for Hierarchical Clustering (k=2 to k=10):

The **Silhouette Score** for hierarchical clustering is presented below:

k (Number of Clusters)	Silhouette Score
2	0.4194
3	0.4382
4	0.5186
5	0.5581
6	0.6141
7	0.6772
8	0.7409
9	0.7750
10	0.7907

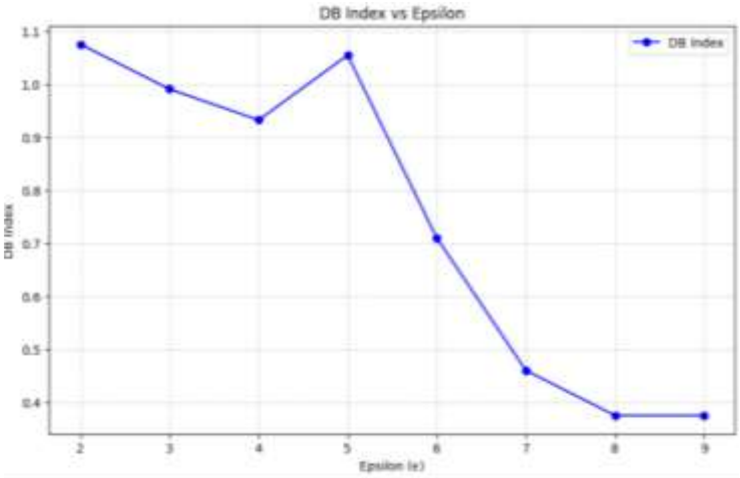


- **Interpretation**: The **Silhouette Score** increases with the number of clusters, peaking at k=10 with a score of **0.7907**. This suggests that hierarchical clustering improves as the number of clusters increases, achieving the highest cohesion and separation at k=10. The Silhouette Score for hierarchical clustering suggests that this model performs well when more clusters are created, with the best performance at k=10.

DBSCAN Clustering Results

Davies-Bouldin Index (DBI) for varying ϵ values:

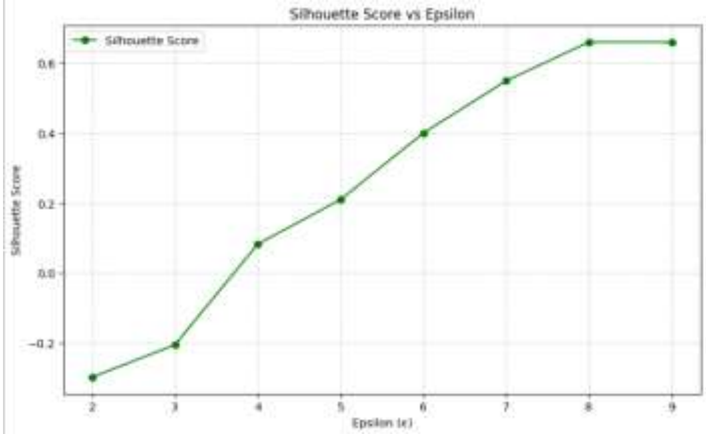
ϵ	DB Index
2	1.0758
3	0.9911
4	0.9326
5	1.0550
6	0.7101
7	0.4596
8	0.3746
9	0.3746



- **Interpretation**: Similar to KMeans, the **Davies-Bouldin Index** for DBSCAN decreases as the value of **ϵ** increases, reaching its minimum value at **ϵ**=8 (0.3746). This indicates that as more clusters are formed, the clusters become increasingly well-separated.

- **Silhouette Scores** for varying ϵ:

ϵ\epsilon	Silhouette Score
2	-0.2978
3	-0.2053
4	0.0827
5	0.2102
6	0.4005
7	0.5499
8	0.6607
9	0.6607



- **Interpretation**: Similar to KMeans, the **Silhouette Score** for DBSCAN increases as the value of **ϵ** increases, reaching its minimum value at **ϵ**=8 (0.6607). This indicates that as more clusters are formed, the clusters become increasingly cohesive and well-separated.

Comparison of All three Clustering methods

Metric	KMeans (Best k)	Hierarchical (Best k)	DBSCAN (ϵ)
Davies-Bouldin Index	0.3183 (k=10)	0.2887 (k=10)	0.3746(ϵ=8 or 9)
Silhouette Score	0.7794 (k=9)	0.7907 (k=10)	0.6607(ϵ=8 or 9)

- **Davies-Bouldin Index**: Both algorithms show a similar trend, with hierarchical clustering achieving a slightly lower DBI value at k=10, indicating that it may offer slightly better separation of clusters than KMeans at this point.
- **Silhouette Score**: Hierarchical clustering achieves a higher silhouette score than KMeans, peaking at **0.7907** at k=10 compared to KMeans' peak of **0.7794** at k=9. This suggests that hierarchical clustering produces better-defined clusters than KMeans at its optimal number of clusters.

Conclusion

- **Optimal Number of Clusters**: Based on the evaluation metrics of best performing algorithms, the **optimal number of clusters** for both KMeans and Hierarchical clustering is **10 clusters**. While both clustering algorithms show improvement with more clusters, hierarchical clustering produces the most well-defined clusters at k=10 according to both the Davies-Bouldin Index and the Silhouette Score.
- **Best Algorithm**: While both methods show good clustering quality, **Hierarchical Clustering** has a slight edge over KMeans in terms of Silhouette Score, indicating that it generates more cohesive and well-separated clusters at the optimal value of k=10.