CLUSTERING REPORT

The clustering results across **K-Means**, **DBSCAN**, **Gaussian Mixture Models (GMM)**, and **Hierarchical Clustering** are analysed below. Each method is evaluated using the **DB Index** and **Silhouette Score** to assess the quality of clusters.

1. K-Means Clustering

- K-Means clustering was evaluated for **2 to 10 clusters**.
- The **DB Index** improved as the number of clusters increased, with the **best score at 7 clusters (1.1218)**.
- The **Silhouette Score** remained consistently low, with the highest score at **2 clusters (0.2987)**.

Metrics:

Number of Clusters	DB Index	Silhouette Score
2	1.2779	0.2987
3	1.4398	0.2197
4	1.2426	0.2348
5	1.3140	0.2443
6	1.2497	0.2512
7	**1.1218**	**0.2538**
8	1.2031	a0.2538
9	1.1719	0.2401
10	1.1439	0.2414

Observations:

- Best Configuration: 7 Clusters
- The DB Index is the lowest, indicating better compactness and separation.

2. DBSCAN (Density-Based Spatial Clustering)

- DBSCAN identified clusters based on density, using eps = 1.6 and min_samples = 5.
- **DB Index: 0.6527** (best across all methods).
- **Silhouette Score: 0.3822** (highest among all methods).
- The Silhouette Score (0.3822) is higher than K-Means, indicating moderately well-separated clusters.

3. Gaussian Mixture Models (GMM)

- GMM assumes data follows a Gaussian distribution.
- Evaluated with **7 components** for direct comparison with K-Means.
- **DB Index**: **2.0535** (highest, indicating poor separation).
- **Silhouette Score**: **0.1470** (lowest, indicating overlapping clusters).
- Not ideal for this dataset.

4. Hierarchical Clustering

- Hierarchical Clustering was evaluated with 5 clusters based on dendrogram inspection.
- **DB Index: 1.4151.**
- **Silhouette Score: 0.1929.**
- The DB Index and Silhouette Score indicate weak clustering.
- Hierarchical Clustering can provide insights into the hierarchical structure but is less suitable for this dataset compared to DBSCAN.

Key Findings

Method	Number of Clusters	DB Index	Sulhouette Score	Remarks
K-Means	7	1.1218	0.2596	Best for simple clustering. Weak separation.
DBSCAN	Density-Based	0.6527	0.3822	Best overall performance. Overlapping clusters.
Gaussian Mixture	7	2.0535	0.1470	Poor performance. Overlapping clusters.
Hierarchical Clustering	5	1.4151	0.1929	Moderate clustering. Limited separation.

Conclusion

1. **Best Clustering Method**: **DBSCAN** achieved the best results:
    - **DB Index**: 0.6527 (lowest, indicating compact and well-separated clusters).
    - **Silhouette Score**: 0.3822 (highest, suggesting moderate cluster quality).
2. **K-Means**:

- Performs better with 7 clusters but suffers from weak separation (low Silhouette Score).

3. **GMM**:
    - Not suitable for this dataset due to overlapping clusters and poor metrics.
4. **Hierarchical Clustering**:
    - Provides limited insights but is less effective than DBSCAN.