**Assignment 3: Clustering Algorithm Self-Study**  
**Student: Aryan Rathod**

**Objective**

The goal of this assignment was to explore, apply, and compare three clustering algorithms: DBSCAN, k-Means, and Hierarchical Clustering, across datasets with different cluster shapes and densities. By visually analyzing the clustering results and compiling a comparison table, we aimed to understand where each algorithm performs best.

**Datasets Used**

1. make\_moons: A dataset with two interlocking moon-shaped clusters. Ideal for testing how well an algorithm handles non-linear shapes.
2. make\_blobs (with varying densities): A dataset with three blob-like clusters, each having different standard deviations, making it a challenge for algorithms that assume equal cluster spread.

**Clustering Results and Analysis**

DBSCAN:

On the make\_moons dataset, DBSCAN performed very well. It correctly identified the curved clusters and handled noise effectively.

On the make\_blobs dataset, DBSCAN struggled due to the varying densities. Some points were misclassified or marked as noise.

k-Means:

On make\_moons, k-Means did not perform well. It assumes spherical clusters, which doesn't work for curved shapes.

On make\_blobs, k-Means worked well, successfully identifying the three blobs even with different variances.

Hierarchical Clustering:

On make\_moons, it performed better than k-Means, capturing the general structure, but still wasn't perfect.

On make\_blobs, the results were reasonable, but the method is slower and sensitive to how clusters are linked.

**Algorithm Comparison Table**

Criterion:  
Cluster Shape Assumption

* k-Means: Spherical
* Hierarchical: Flexible
* DBSCAN: Arbitrary shapes

Noise Handling

* k-Means: Poor
* Hierarchical: Poor
* DBSCAN: Excellent

Parameter Sensitivity

* k-Means: Sensitive to k
* Hierarchical: Sensitive to linkage method
* DBSCAN: Sensitive to eps and min\_samples

Scalability

* k-Means: Fast
* Hierarchical: Slower
* DBSCAN: Moderate

Best Use Case

* k-Means: Well-separated, equal density clusters
* Hierarchical: Nested clusters, small datasets
* DBSCAN: Spatial or noisy datasets

**Plots**

Six plots were generated, showing the performance of each algorithm on both datasets. These have been saved in the 'plots' directory with appropriate filenames for each clustering result.