Выбор метода кластеризации

Алгоритмы

Рассмотренные нами:

- K-средних
- EM-алгоритм
- Аггломеративная иерархическая кластеризация
- DBSCAN

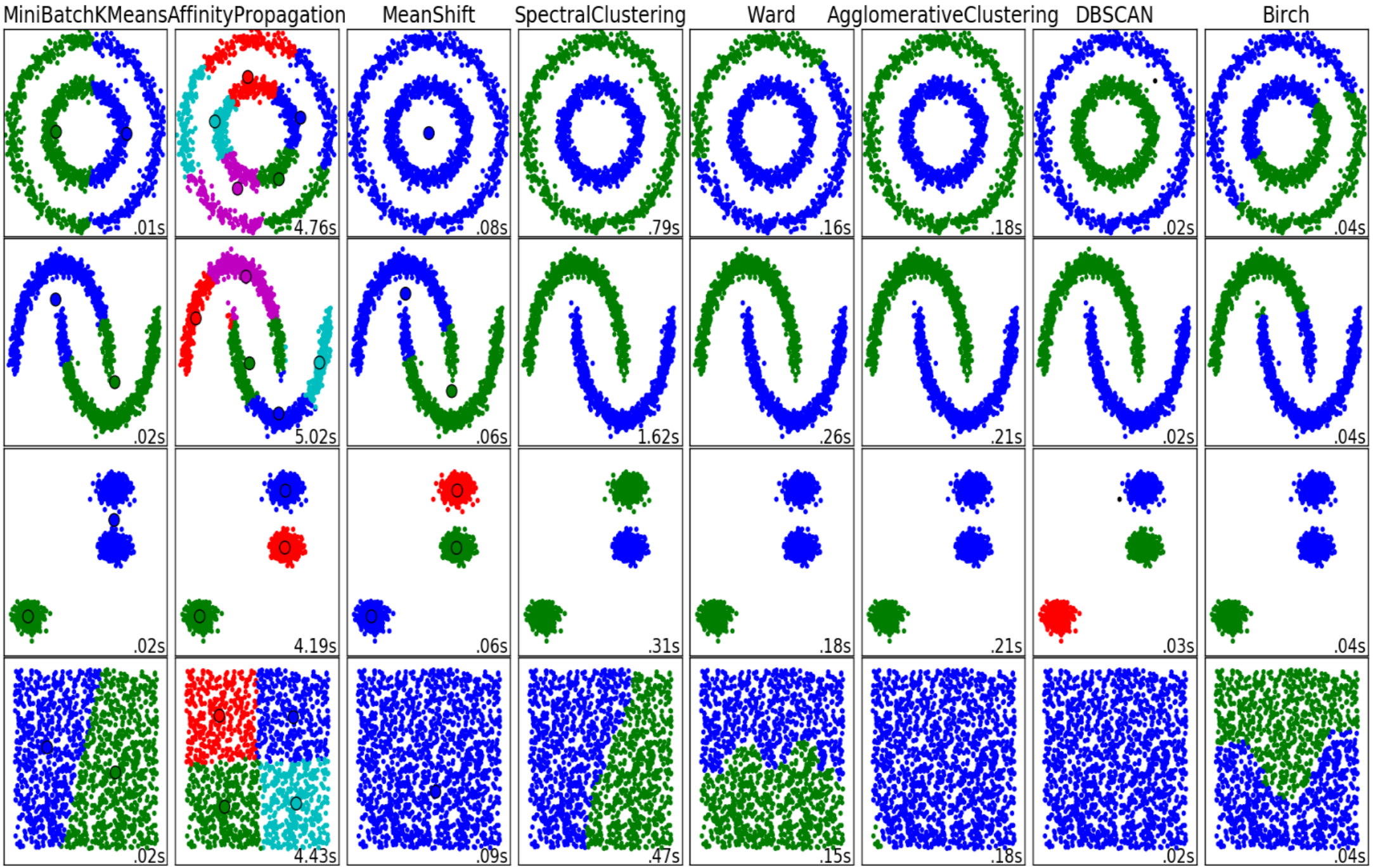
Алгоритмы

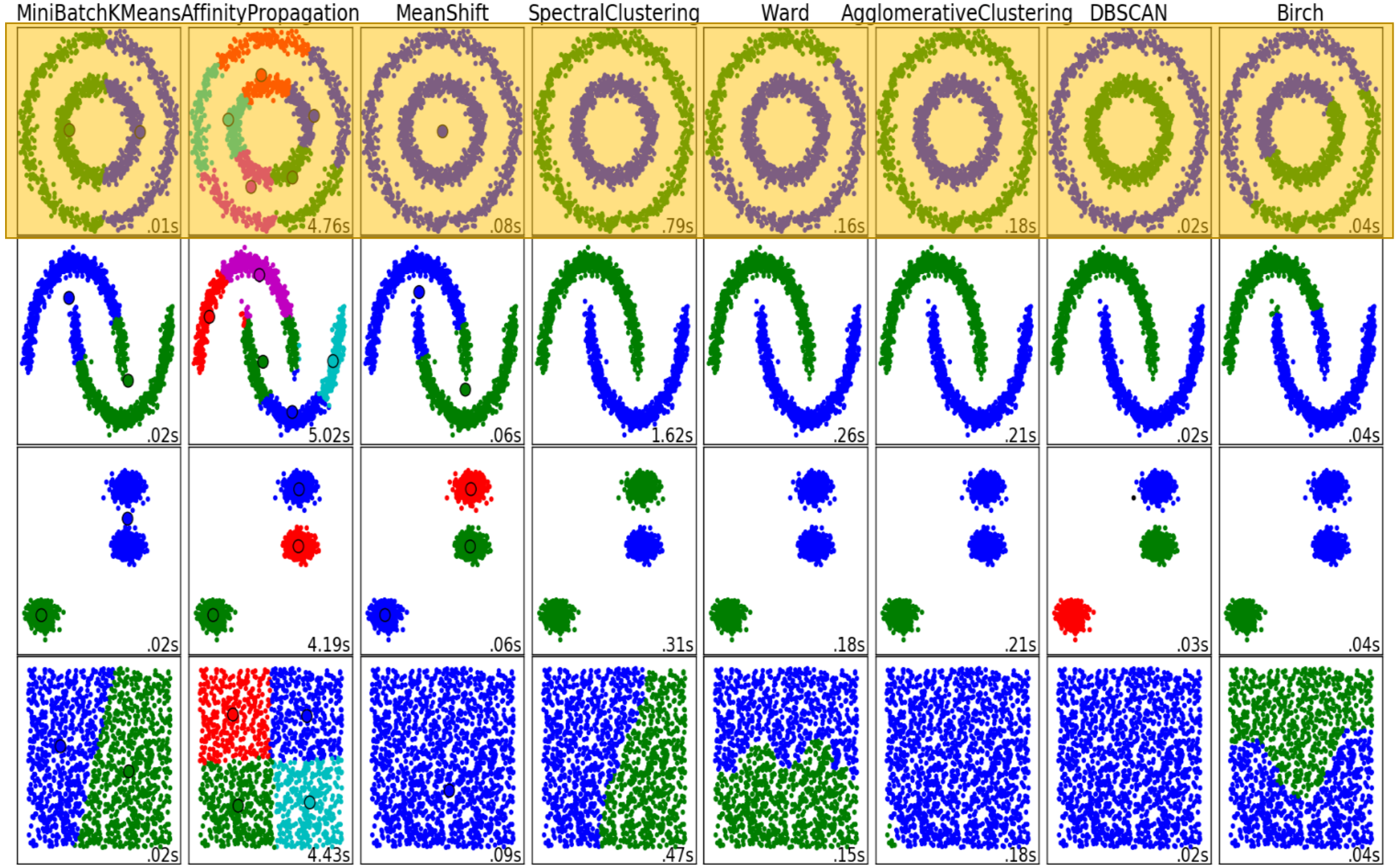
Рассмотренные нами:

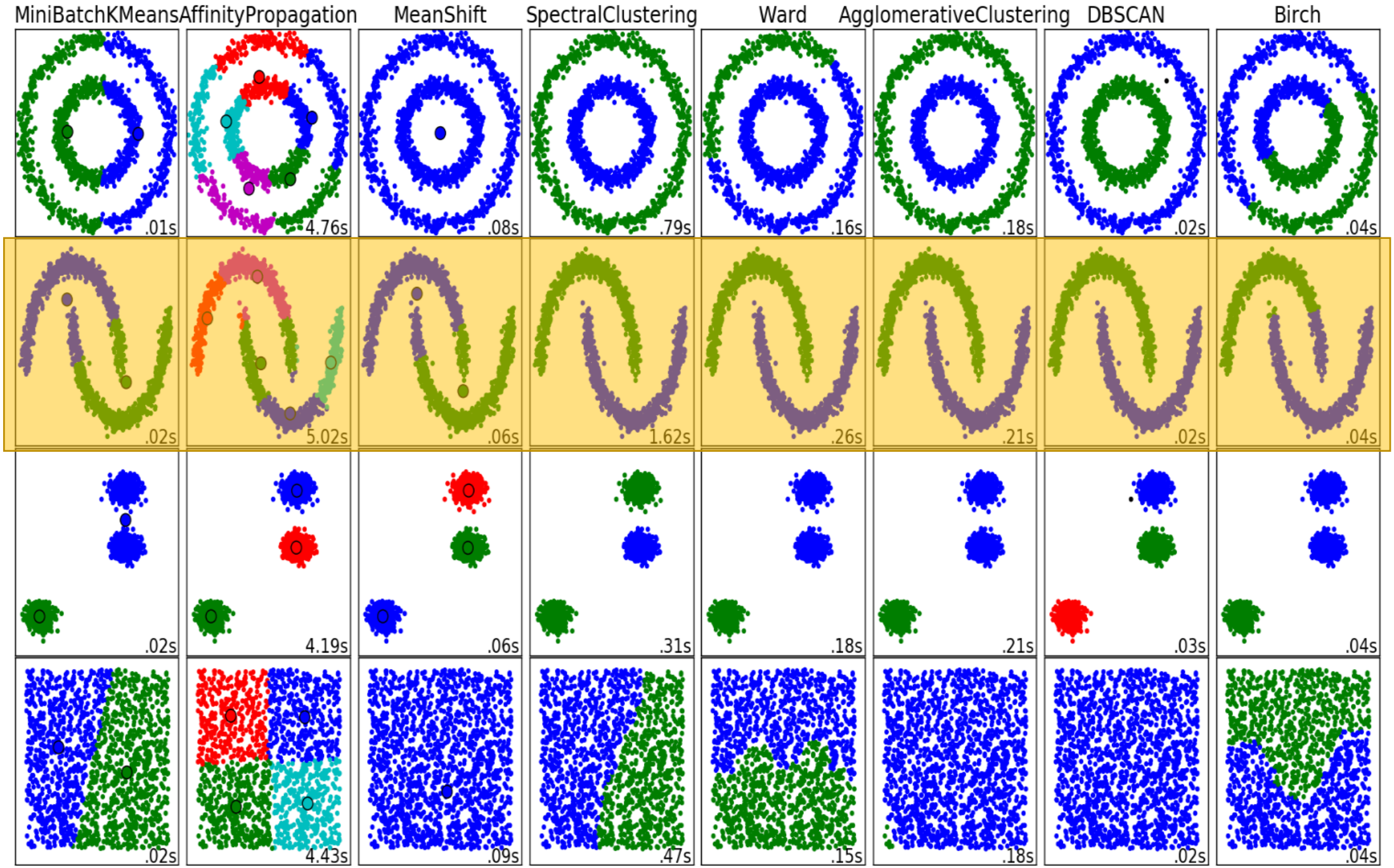
- K-средних
- EM-алгоритм
- Агломеративная иерархическая кластеризация
- DBSCAN

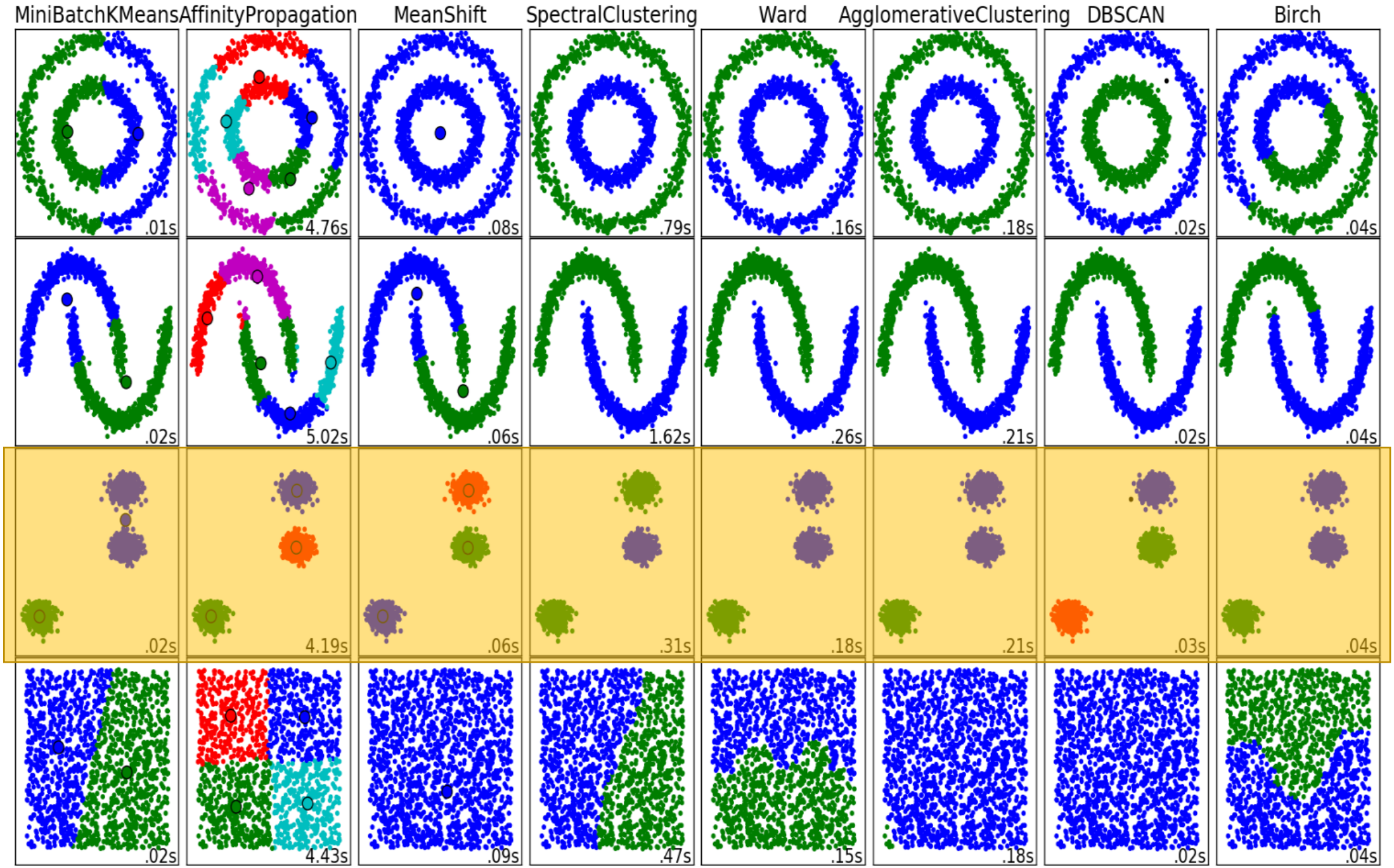
В scikit-learn:

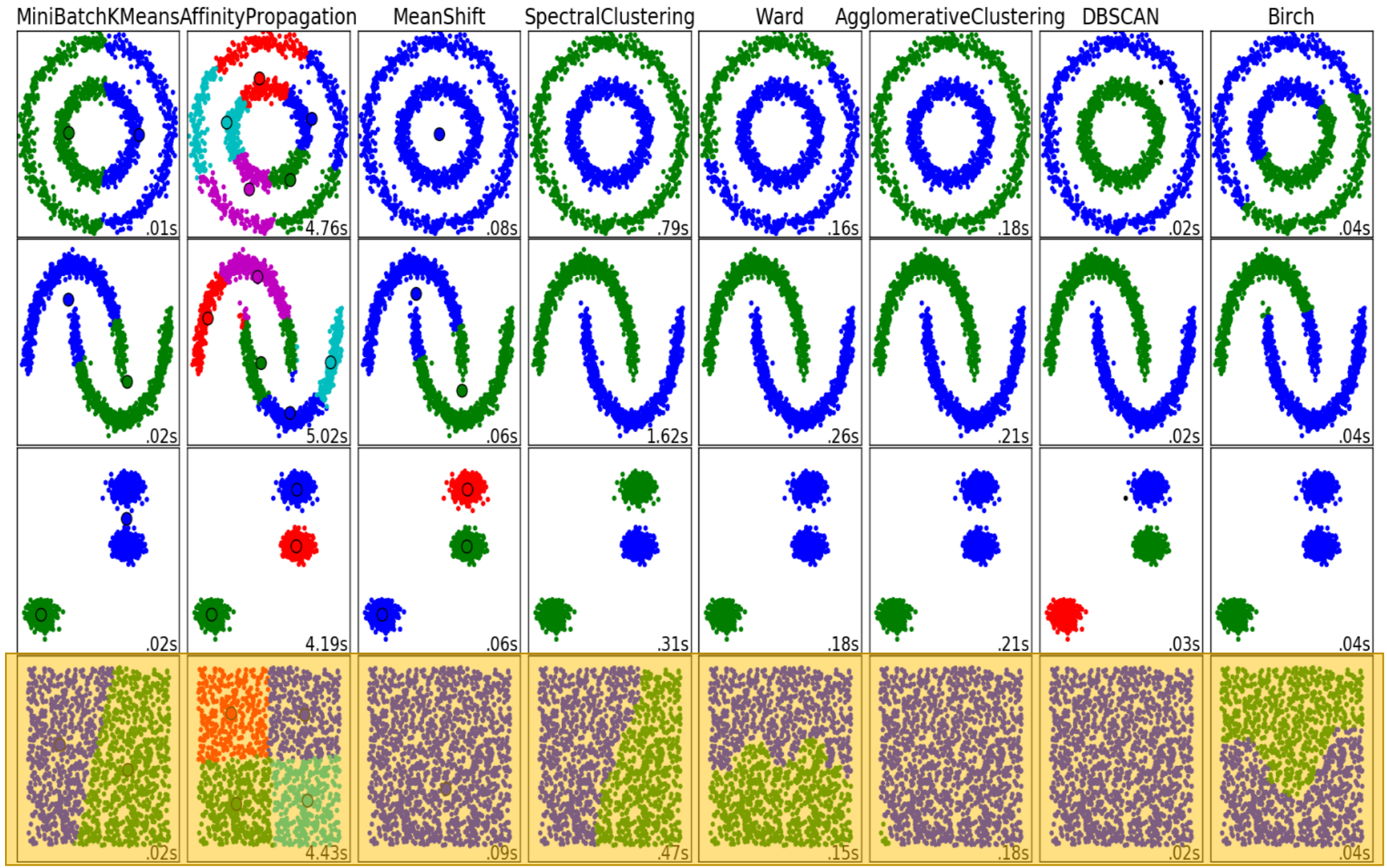
KMeans, MiniBatchKMeans, GaussianMixture, AgglomerativeClustering, Ward, DBSCAN, MeanShift, AffinityPropagation, SpectralClustering, Birch

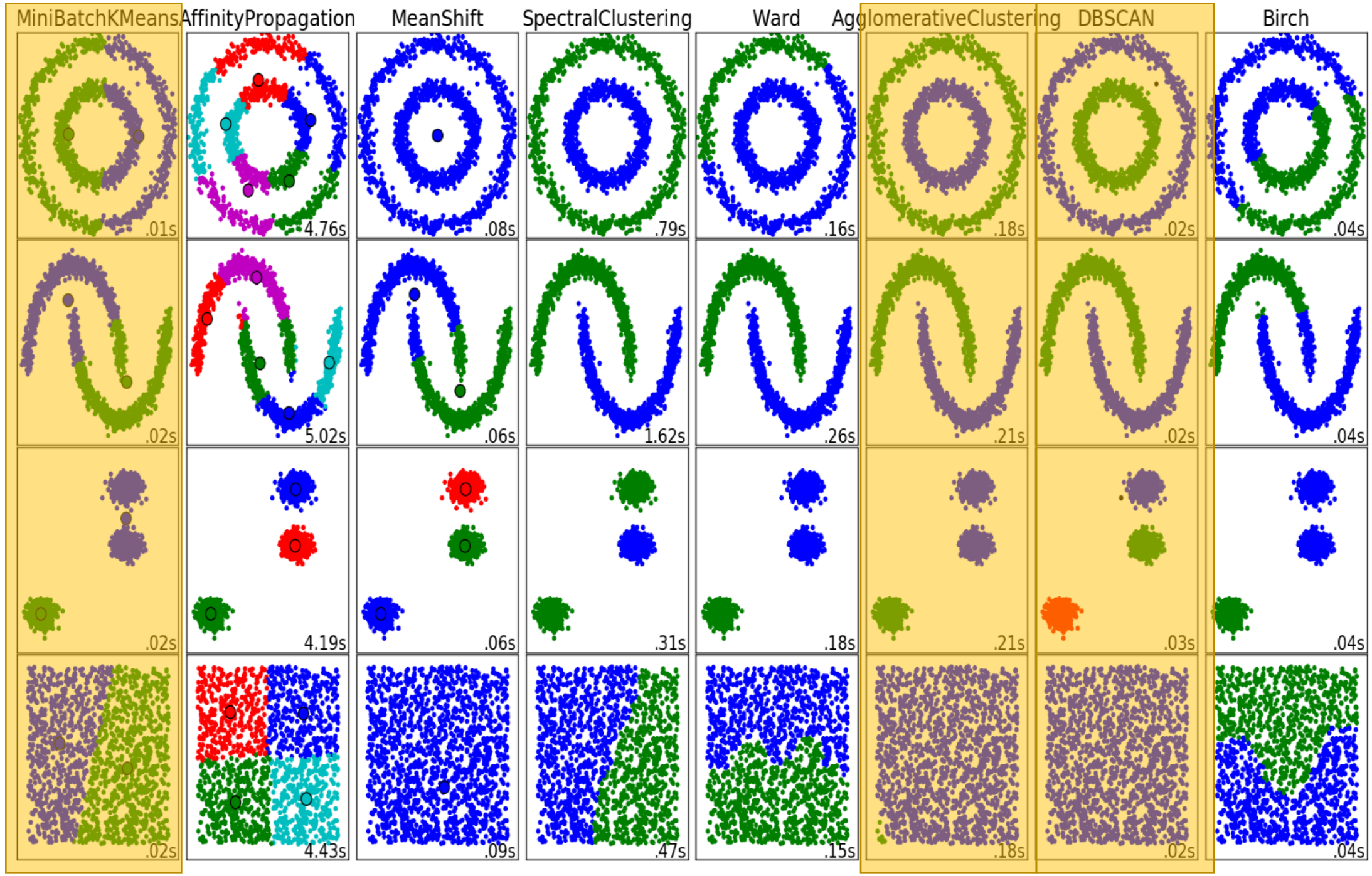

MiniBatchKMeans
AffinityPropagation
MeanShift
SpectralClustering
Ward
AgglomerativeClustering
DBSCAN
Birch
.01s
4.76s
.08s
.79s
.16s
.18s
.02s
.04s
.02s
5.02s
.06s
1.62s
.26s
.21s
.02s
.04s
.02s
4.19s
.06s
.31s
.18s
.21s
.03s
.04s
.02s
4.43s
.09s
.47s
.15s
.18s
.02s
.04s


MiniBatchKMeans
AffinityPropagation
MeanShift
SpectralClustering
Ward
AgglomerativeClustering
DBSCAN
Birch
.01s
4.76s
.08s
.79s
.16s
.18s
.02s
.04s
.02s
5.02s
.06s
1.62s
.26s
.21s
.02s
.04s
.02s
4.19s
.06s
.31s
.18s
.21s
.03s
.04s
.02s
4.43s
.09s
.47s
.15s
.18s
.02s
.04s


MiniBatchKMeans
AffinityPropagation
MeanShift
SpectralClustering
Ward
AgglomerativeClustering
DBSCAN
Birch
.01s
4.76s
.08s
.79s
.16s
.18s
.02s
.04s
.02s
5.02s
.06s
1.62s
.26s
.21s
.02s
.04s
.02s
4.19s
.06s
.31s
.18s
.21s
.03s
.04s
.02s
4.43s
.09s
.47s
.15s
.18s
.02s
.04s


MiniBatchKMeans
AffinityPropagation
MeanShift
SpectralClustering
Ward
AgglomerativeClustering
DBSCAN
Birch
.01s
4.76s
.08s
.79s
.16s
.18s
.02s
.04s
.02s
5.02s
.06s
1.62s
.26s
.21s
.02s
.04s
.02s
4.19s
.06s
.31s
.18s
.21s
.03s
.04s
.02s
4.43s
.09s
.47s
.15s
.18s
.02s
.04s


MiniBatchKMeans
AffinityPropagation
MeanShift
SpectralClustering
Ward
AgglomerativeClustering
DBSCAN
Birch
.01s
4.76s
.08s
.79s
.16s
.18s
.02s
.04s
.02s
5.02s
.06s
1.62s
.26s
.21s
.02s
.04s
.02s
4.19s
.06s
.31s
.18s
.21s
.03s
.04s
.02s
4.43s
.09s
.47s
.15s
.18s
.02s
.04s


MiniBatchKMeans
AffinityPropagation
MeanShift
SpectralClustering
Ward
AgglomerativeClustering
DBSCAN
Birch
.01s
4.76s
.08s
.79s
.16s
.18s
.02s
.04s
.02s
5.02s
.06s
1.62s
.26s
.21s
.02s
.04s
.02s
4.19s
.06s
.31s
.18s
.21s
.03s
.04s
.02s
4.43s
.09s
.47s
.15s
.18s
.02s
.04s

Метод	Параметры	Масштабируемость	Use-case	Геометрия
KMeans	Число кластеров	Очень много примеров (MiniBatch), среднее число кластеров	Выпуклые, примерно одинаковые кластеры	Евклидово расстояние

Метод	Параметры	Масштабируемость	Use-case	Геометрия
KMeans	Число кластеров	Очень много примеров (MiniBatch), среднее число кластеров	Выпуклые, примерно одинаковые кластеры	Евклидово расстояние
GaussianMixture	Веса, векторы средних, матрицы ковариаций	-	Восстановление плотности, выпуклые кластеры	Обобщение евклидовой метрики (с весами)

Метод	Параметры	Масштабируемость	Use-case	Геометрия
KMeans	Число кластеров	Очень много примеров (MiniBatch), среднее число кластеров	Выпуклые, примерно одинаковые кластеры	Евклидово расстояние
GaussianMixture	Веса, векторы средних, матрицы ковариаций	-	Восстановление плотности, выпуклые кластеры	Обобщение евклидовой метрики (с весами)
Agglomerative Clustering	Число кластеров, linkage, метрика	Много примеров и много кластеров	Много кластеров, нужно задавать метрику	Любая метрика/функция близости

Метод	Параметры	Масштабируемость	Use-case	Геометрия
KMeans	Число кластеров	Очень много объектов (MiniBatch), среднее число кластеров	Выпуклые, примерно одинаковые кластеры	Евклидово расстояние
GaussianMixture	Веса, векторы средних, матрицы ковариаций	-	Восстановление плотности, выпуклые кластеры	Обобщение евклидовой метрики (с весами)
Agglomerative Clustering	Число кластеров, linkage, метрика	Много объектов и много кластеров	Много кластеров, нужно задавать метрику/близость (например, косинусную)	Любая метрика/функция близости, для евклидовой - Ward
DBSCAN	Радиус окрестности, число соседей	Много объектов, среднее число кластеров	Неравные невыпуклые кластеры, выбросы,	Евклидово расстояние

Резюме

- K-средних
- EM с нормальным распределением
- Иерархическая агломеративная кластеризация
- DBSCAN