

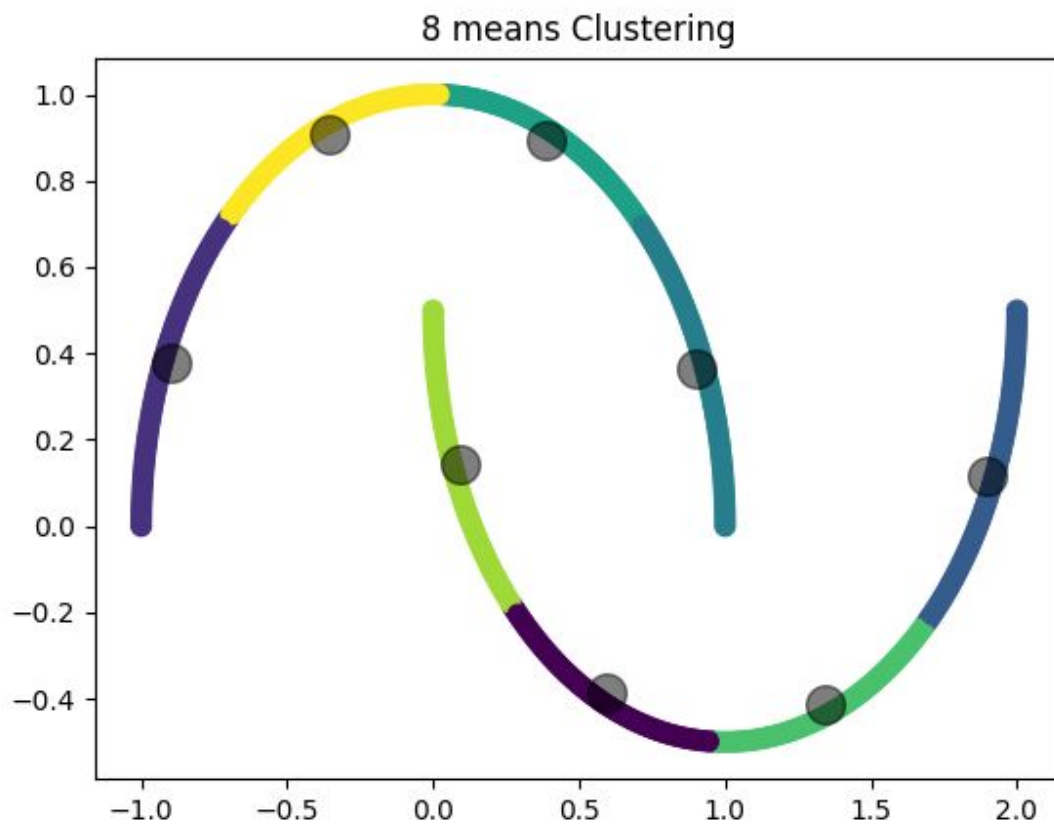
REPORT HW4

Question3

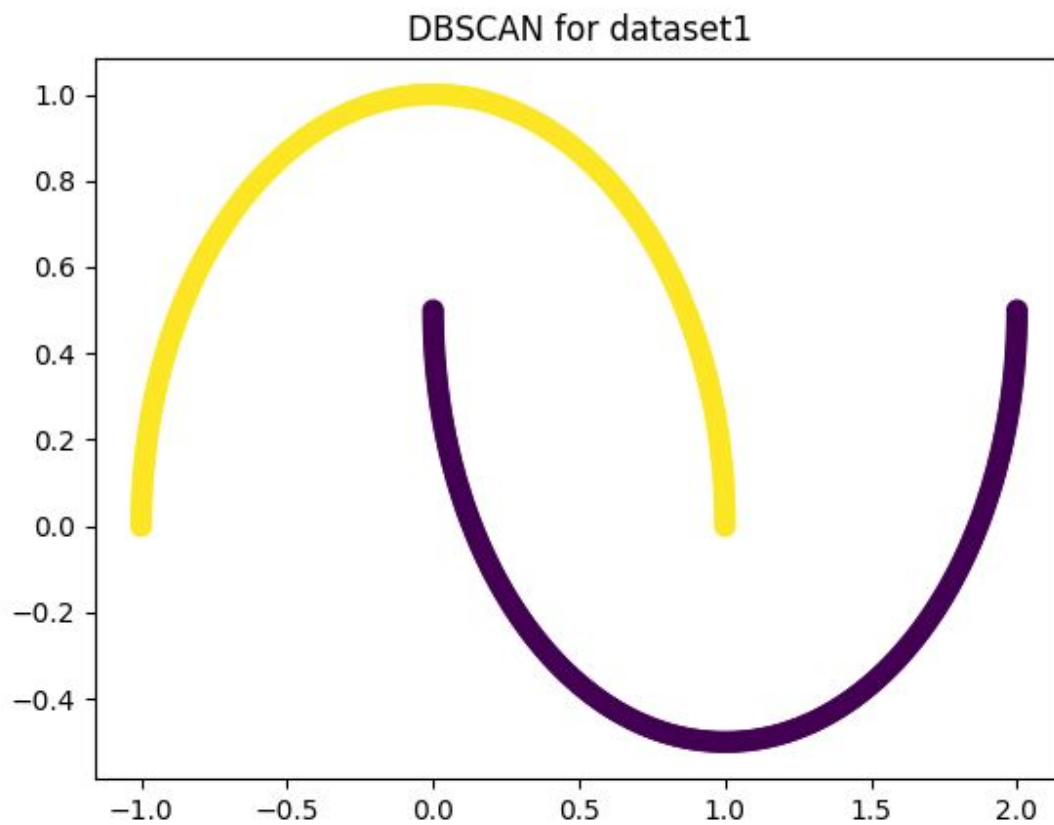
3.py contains KMeans code for dataset1 and dataset2.

3b.py contains DBSCAN algorithm for dataset1 and dataset2

3a) Number of clusters= 8

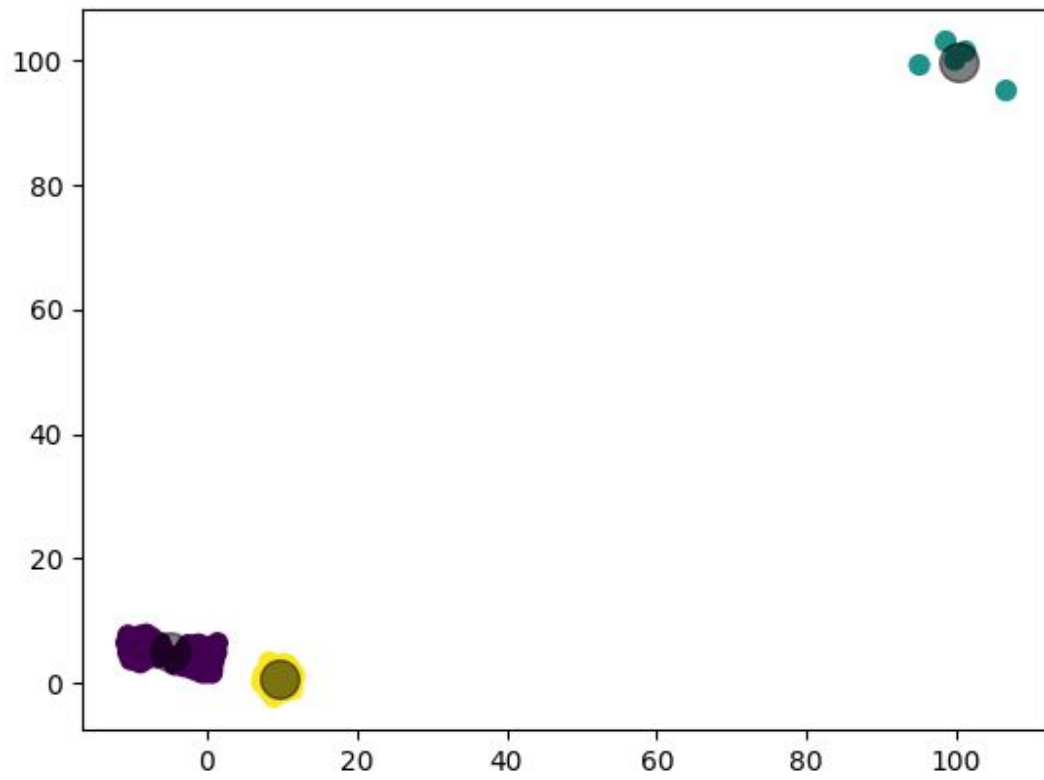


b) Epsilon = 0.1, Minimum number of points = 10

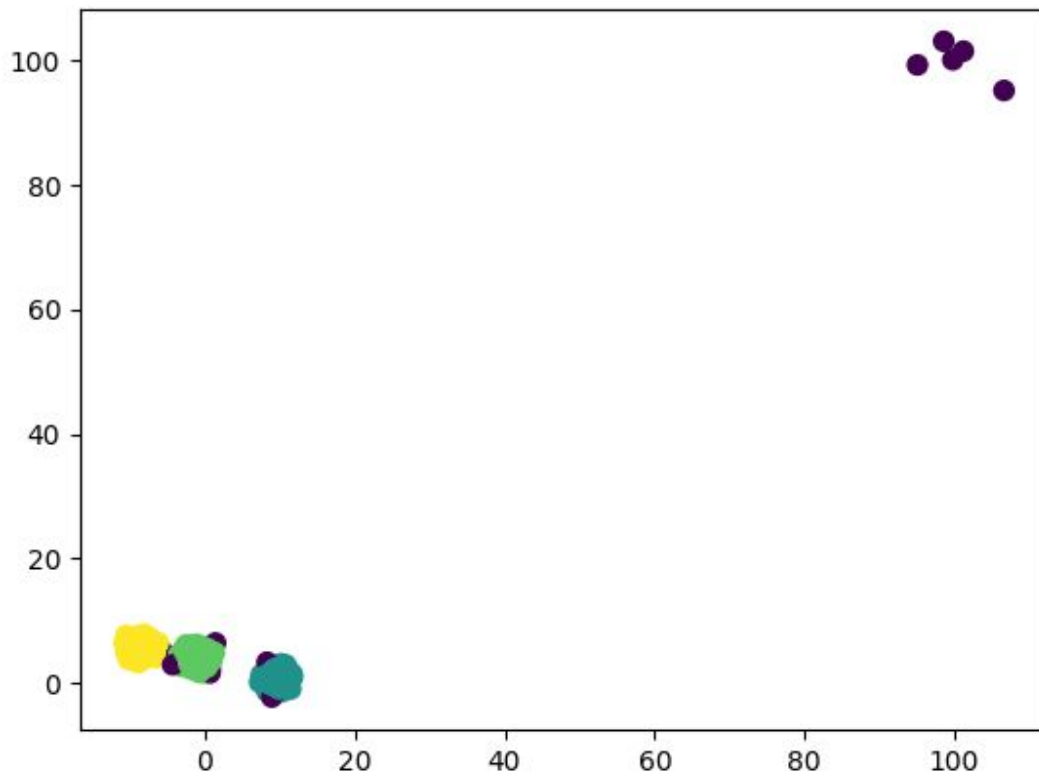


c) DBSCAN depends on the values of parameters. Great change can be seen on changing epsilon and minimum number of points. K means need the number of clusters whereas DBSCAN doesn't depend on k.

d) K-Means clustering, number of clusters = 3



Using DBSCAN:



K-means clustering is sensitive to outliers. DBSCAN can identify points that are not part of any cluster.

Pros:

- 1) DBScan doesn't need number of clusters
- 2) DBScan can capture clusters of complex shape.
- 3) KMeans is faster than DBScan.

Cons:

- 1) DBScan doesn't work well over clusters with different densities
- 2) DBScan needs a careful selection of its parameters