AML ASSIGNMENT 3 REPORT

1) Kmeans

KMeans clustering with n_clusters = 8
The visualization of the clustered data.

2)DBSCAN

DBSCAN clustering with n_clusters = 2
The visualization of the clustered data.
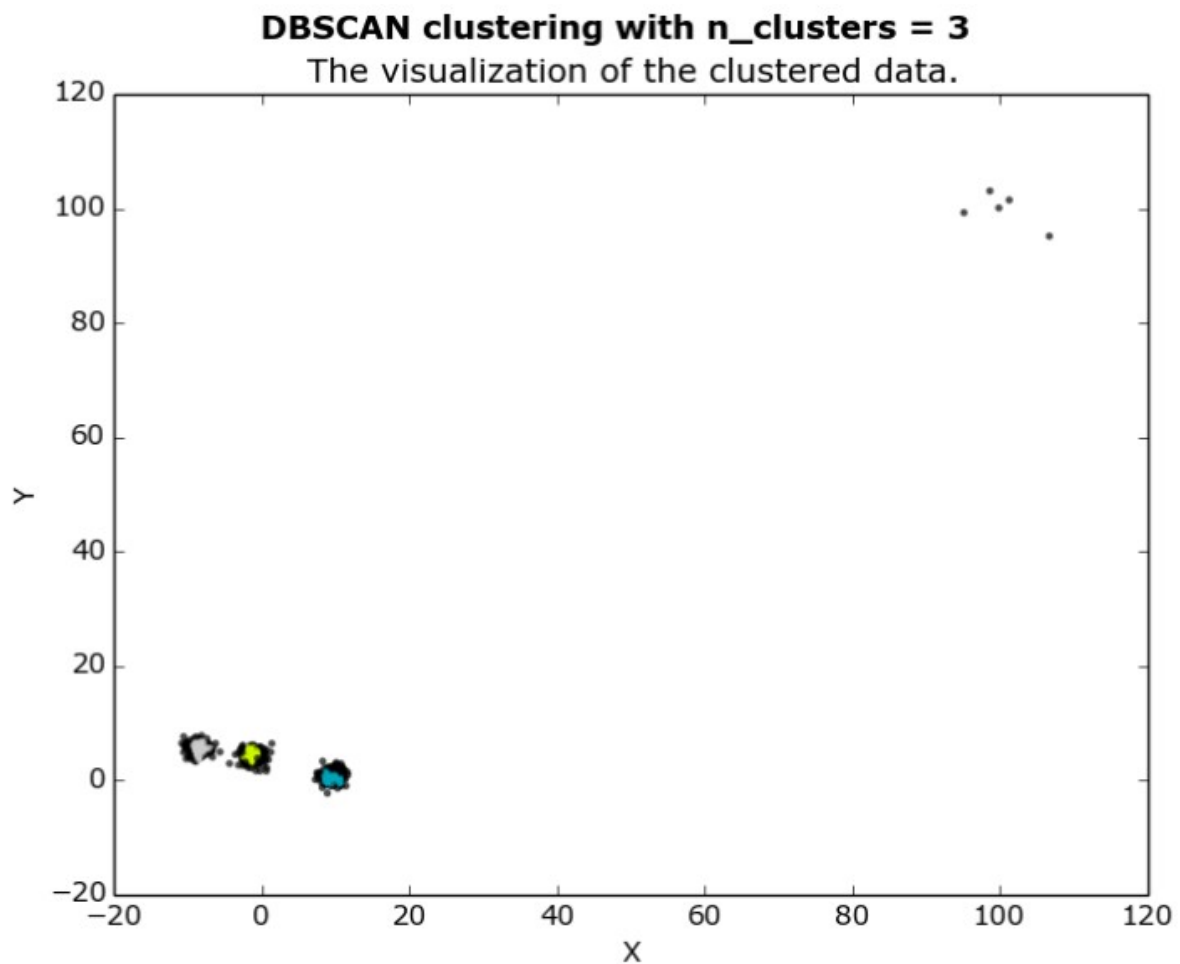
2) Dataset 2

KMeans clustering with n_clusters = 3

The visualization of the clustered data.
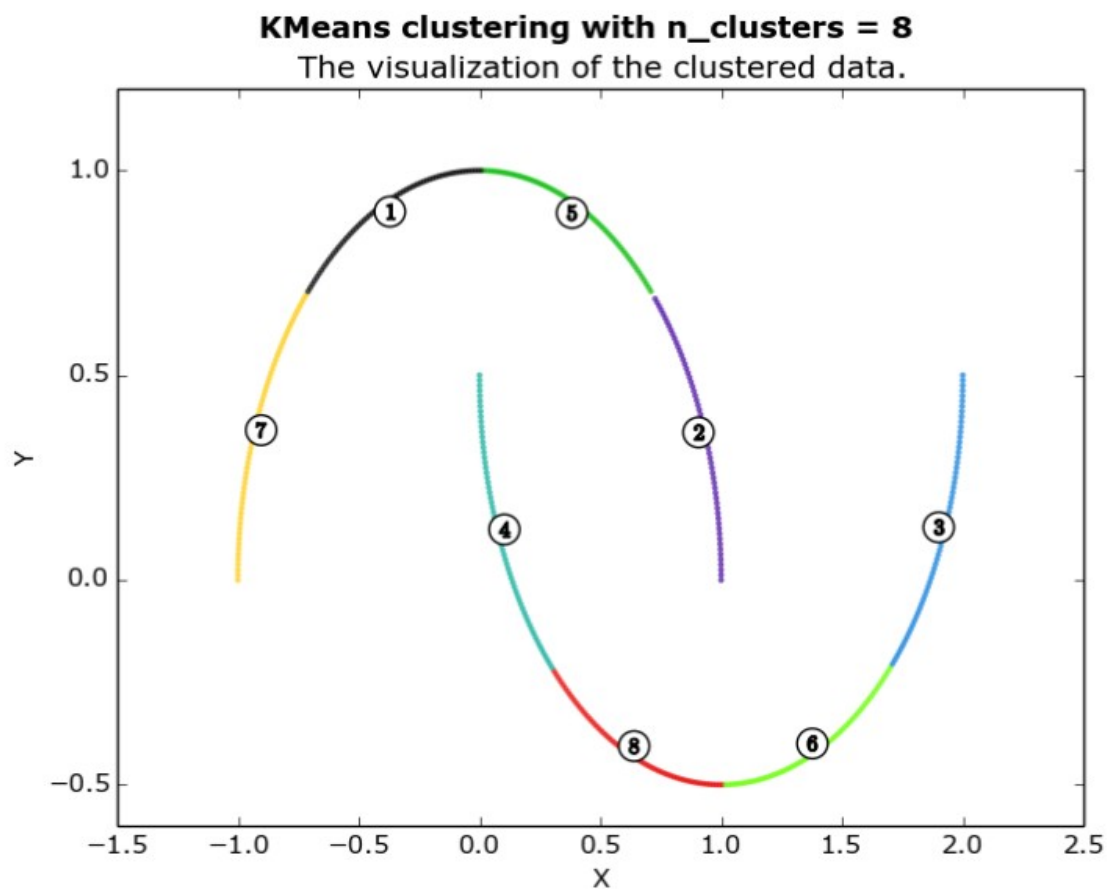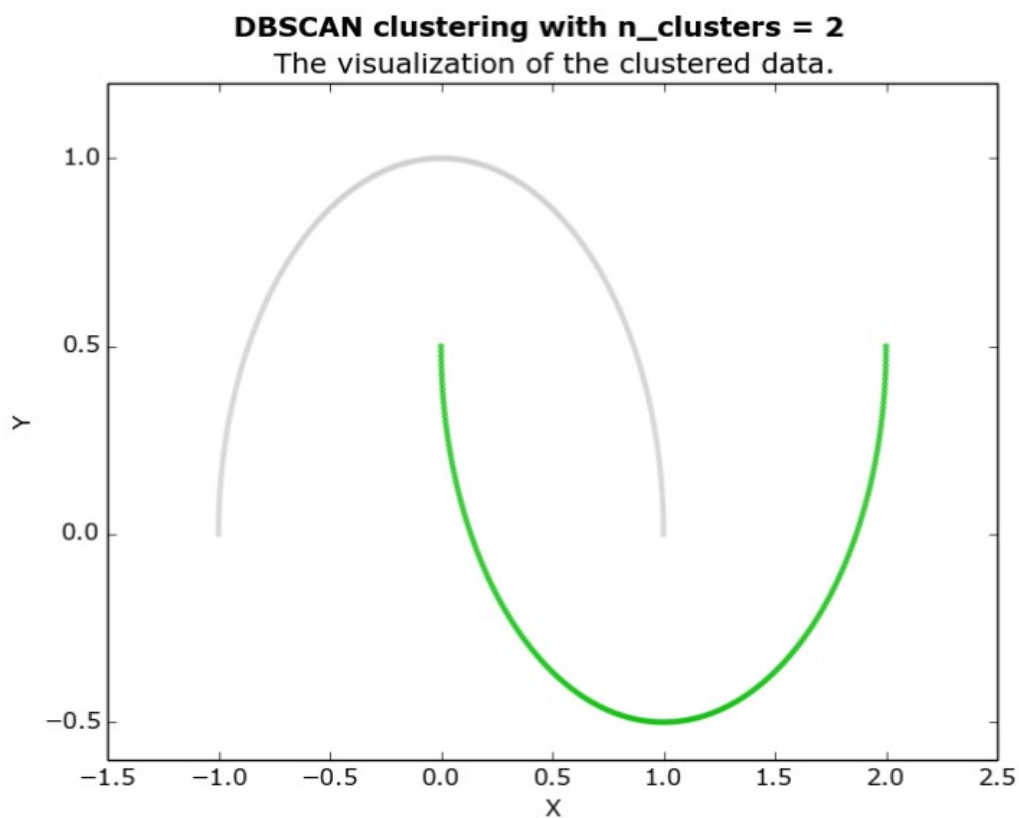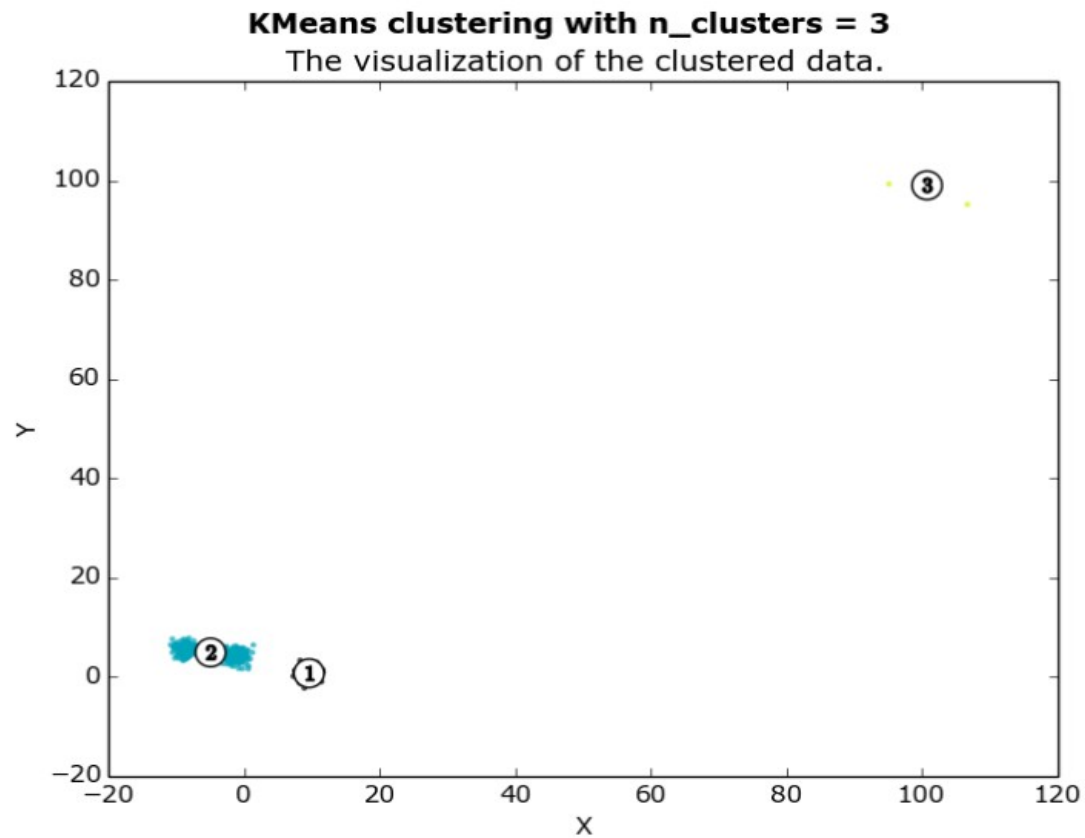
DBSCAN

c)

We observe that the number of clusters is more in case of Kmeans than in

DBSCAN clustering with n_clusters = 3

The visualization of the clustered data.

DBSCAN.
For the best possible case I got 4 clusters for DBSCAN but the other part can also be considred as noise.
K-Means works based on calculating the internal cluster distance and calculating centroids whereas DBSCAN works on cluster densities where it takes minimum distance between two points and minimum number of points required for clustering.

Advantages of K-means:
It is easy to implement as it calculates the cluster centroids using the minimum distance criteria.

Advantage of DBSCAN:
 The number of clusters need not not be specified.
 The parameters can be set to get best results.

Disadvantage of DBSCAN:
        DBSCAN scan can fail to perform well for datasets with large differences in densities
Finding best eps and min number of cluster points

2)
a)The assessment we make is that the PCA can range affect the dataset upto value of n. Hence it is generally taken care that n is comparable to m.

b)
 We observe that
For 50% variance we need 4 principal components
For 80% variance we need 20 principal components
For 90% variance we need 41 principal components

c)We observe that as the PCA increases, the detailing increases.

d) We observe that the detailing doesn't change much.

3)We observe that the DBSCAN gives only one cluster. Kmean performs better slightly.