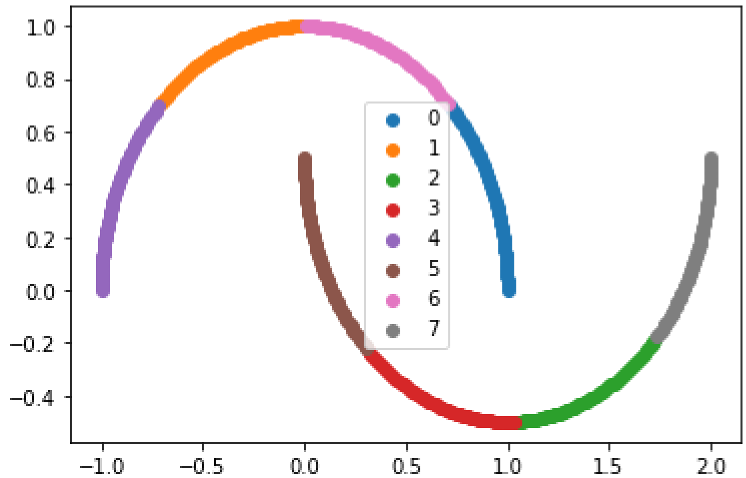
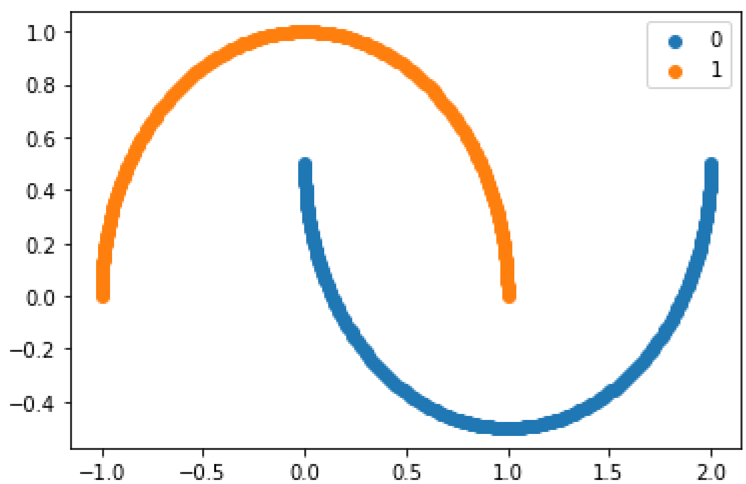
Question 4:

(a.) Using the elbow curve method for determining the number of clusters in k-means clustering, the number of clusters was determined to be 8. Following is the plot for 8 clusters.



(b.) The code for DBSCAN can be found in the ipynb file in question 4 folder. The following is the plot for the DBSCAN clustering:



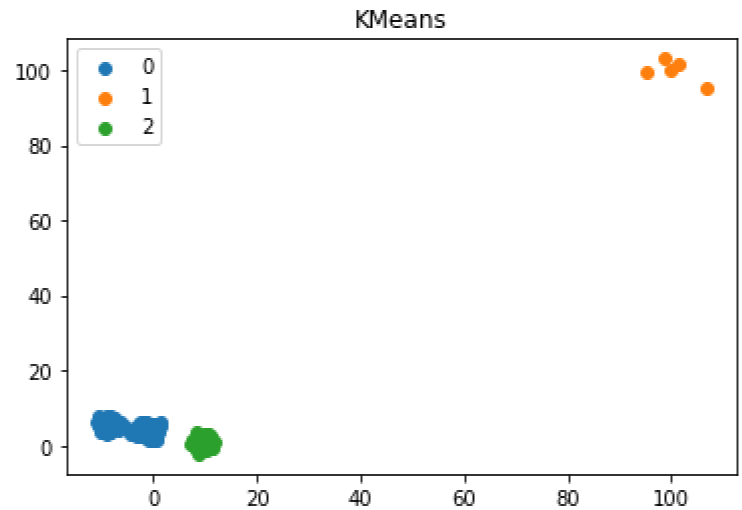
(c.) Differences between K-Means and DBSCAN:

- KMeans did not perform well on dataset1. DBSCAN performed well on dataset1. KMeans failed to capture the true structure of the data unlike DBSCAN.
- KMeans predicted the number of clusters to be 8 whereas DBSCAN predicted the number of clusters to be 2.

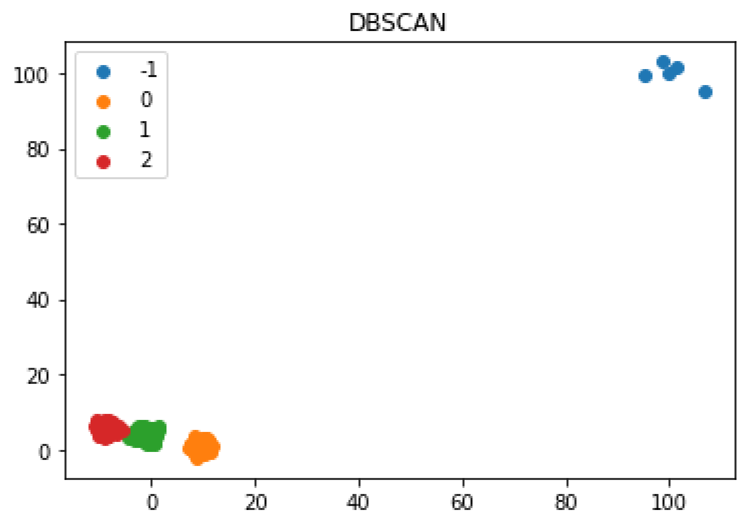
These differences arise because the K-Means clustering algorithm works based on proximity of each point from each cluster whereas DBSCAN works based on local density of points. Thus

DBSCAN works well on well separated density connected clusters whereas kmeans does not work well on non-spherical clusters.

(d.) Following is the plot for k means clustering:



Following is the plot for DBSCAN (-1 denotes noise):



Observations:

- It can be seen that the K Means algorithm did not perform good clustering. It did not separate two clusters but labelled outliers as another cluster.
- DBSCAN recognized outliers as noise and separated the three clusters as was intended.

Pros of K means:

- Is easy to tune as only one parameter (K) needs to be tuned.
- Works well on spherical clusters.

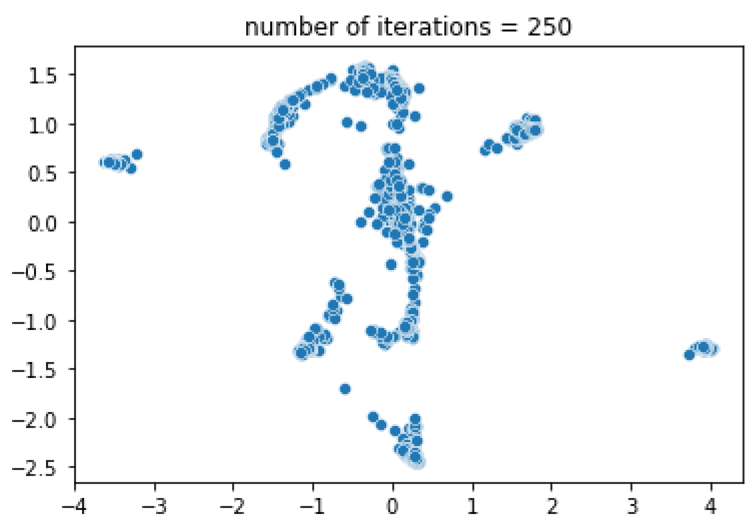
Cons of K means:
- Not robust to outliers
- Does not work well on non-spherical clusters
- Assumes that clusters are of same size

Pros of DBSCAN:
- Does not need number of clusters to be specified
- Robust to outliers
- Cluster shape is not hindering performance

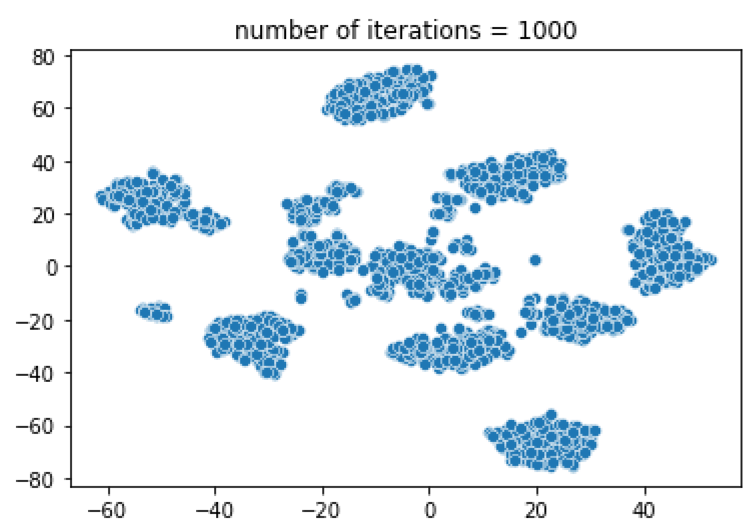
Cons of DBSCAN:
- Determining eps and MinPts can be difficult.
- If clusters are not well separated, then DBSCAN will not perform well

Question 5

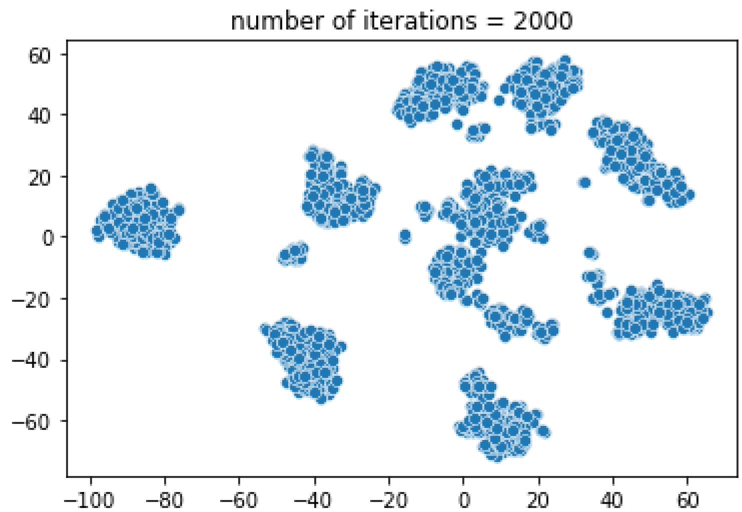
(a.) Plot for 250 iterations:



Plot for 1000 iterations:



Plot for 2000 iterations:



We can observe that the plot for 250 iterations is very different from plots for 1000 and 2000 iterations. We can observe no distinct clusters in the plot for 250 iterations. This is probably because at 250 iterations, the KL divergence has not converged yet.

Plots for 1000 and 2000 iterations are somewhat similar. This is probably because at 1000 iterations, the KL divergence is somewhere near convergence into the same minnima as that in 2000 iterations.

(b.) t-SNE uses gradient descent to minimize Kullback Liebler divergence. KL divergence is not convex. Thus it can have local minima. Since we use gradient descent, we might end up in different local minimas in different runs. Thus different runs of t-SNE with the same parameters can produce different results.