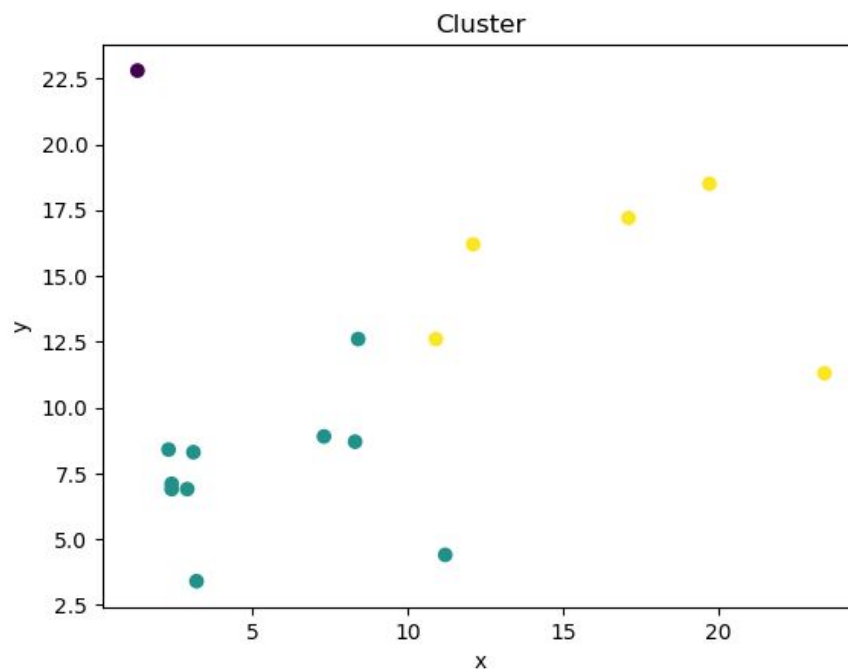


Question1.



From the result, it can be seen 3 clusters are classified, but from the distribution of raw data, the data may be classified 2 class is more reasonable. The point in the upper left corner may be noise data.

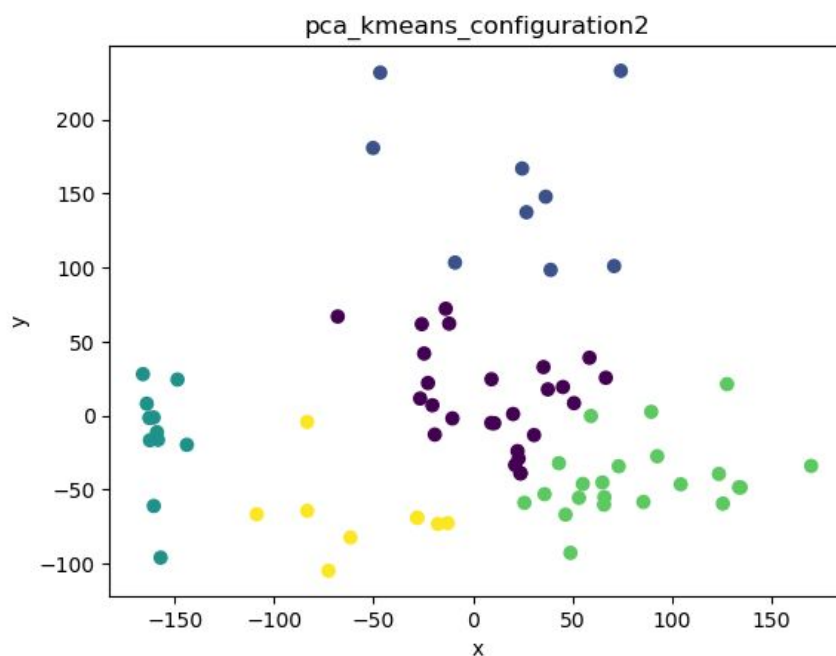
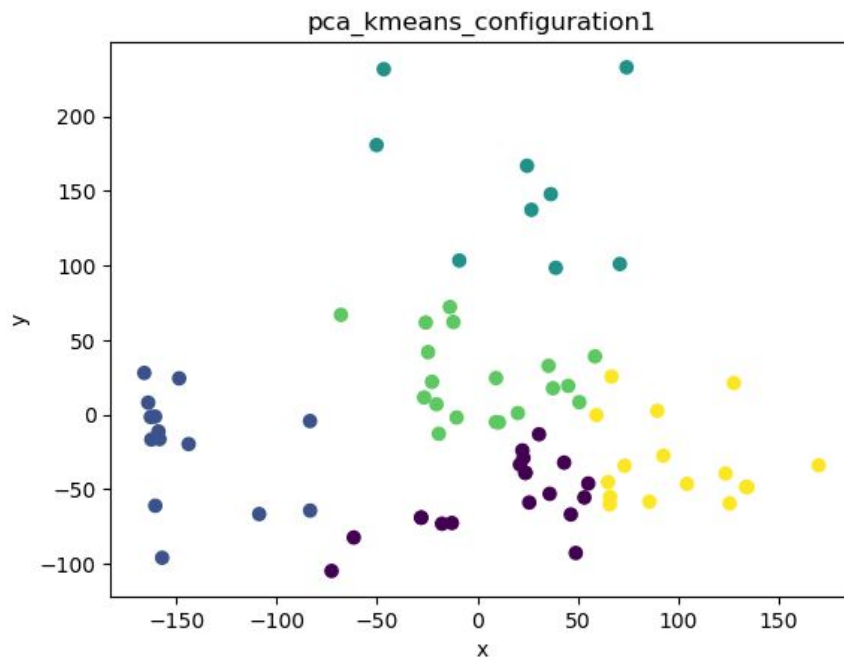
Question2.

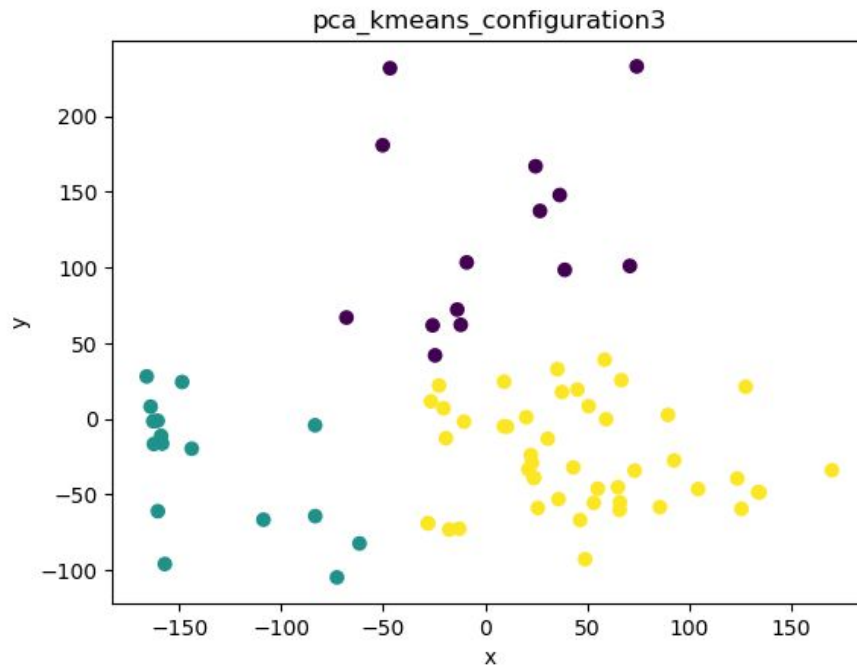
config1	config2
2	1
1	2
2	1
2	1
0	3
0	0
0	4
3	0
3	0
2	1
4	3
4	3
0	3

From the result, it can be seen the two different configurations produce different results. The main reason is using different parameter settings, n_init are 5 and 100 respectively. The number of times the algorithm was run with different initialization centroids. Since K-Means is

a locally optimal iterative algorithm whose results are affected by the initial values, it is necessary to run several times to select a better clustering effect.

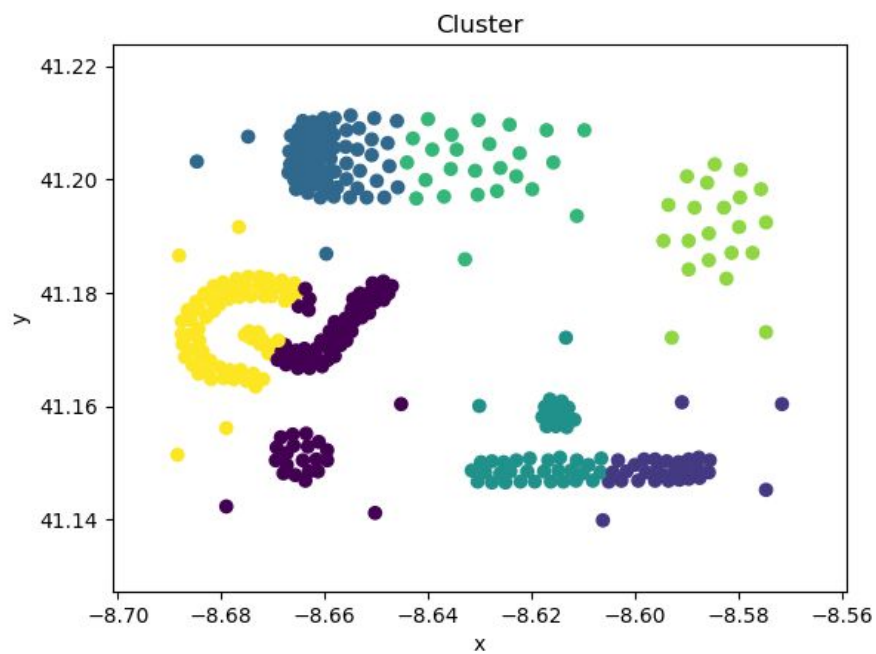
In order to evaluate which configurations is best, first, using the PCA to deal with data and then plot the result. The results are as follows:



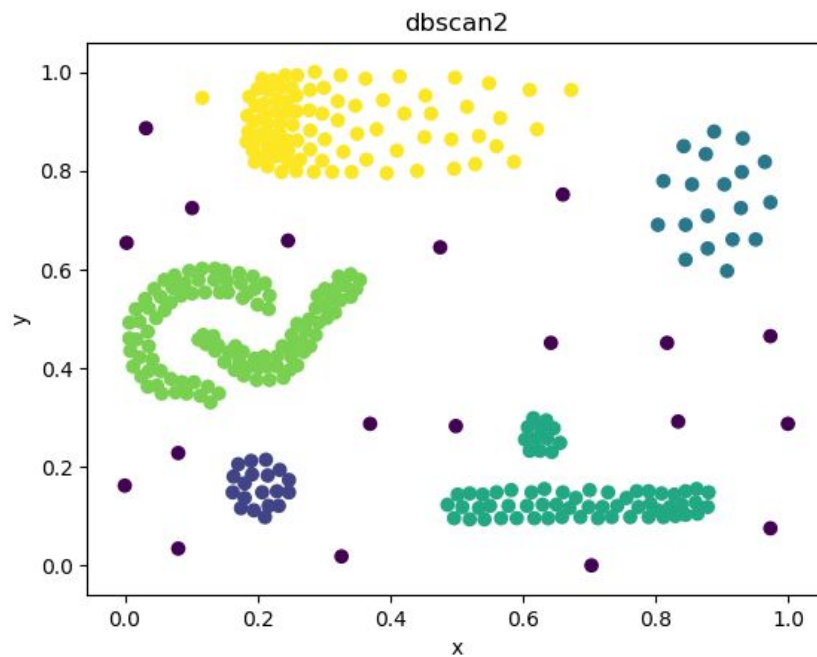
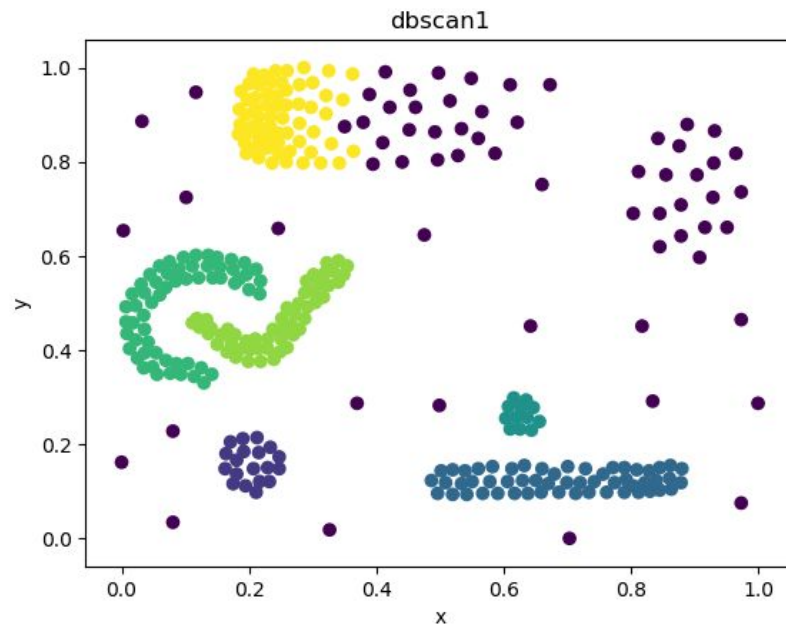


From the result, it can be seen the second configuration (100 maximum runs and 100 maximum optimization steps) can get the best result. The third configuration, it only has three classes, it cannot completely classify the data.

Question3.



From the result, it can be seen there are 7 classes are produced.



From the result, the third solution is the best, the main reason is K-means is a cluster based partition, DBSCAN is based on density. For the second solution, the esp is too small, it makes the model overfitting.