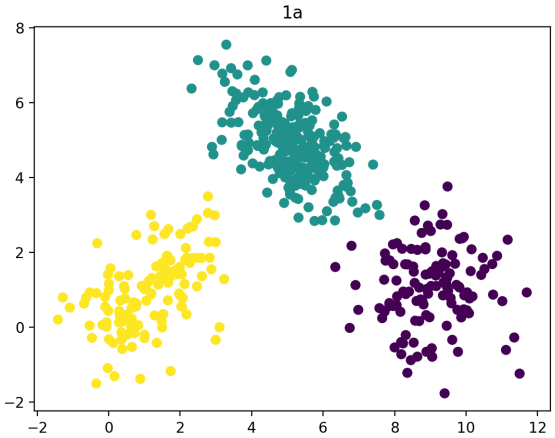
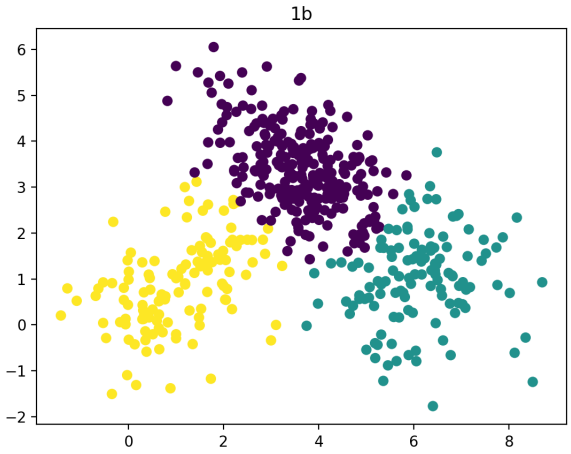
Below is the result of Q1:

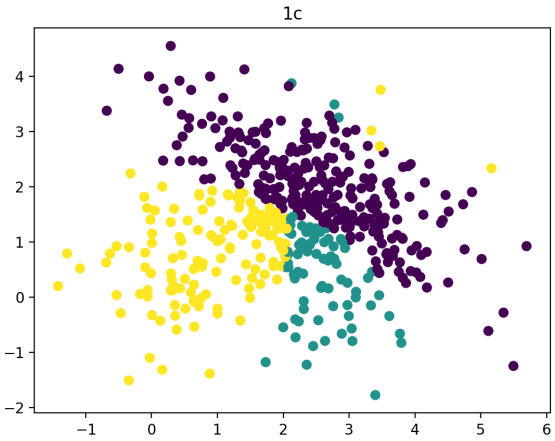
```
1a. mu: [[9.03993609 1.06310294]
 [5.11695814 4.8819475 ]
 [1.13037882 0.95010449]]
1a. sigma: [[[ 1.01558242e+00  6.28527214e-04]
  [ 6.28527214e-04  1.07416971e+00]]

 [[ 8.80287360e-01 -4.55010898e-01]
  [-4.55010898e-01  8.30052948e-01]]

 [[ 1.08854861e+00  5.47149330e-01]
  [ 5.47149330e-01  1.04922055e+00]]]
1a. P: [0.24869293 0.50103393 0.25027314]
```




```
1b. mu: [[3.64626857 3.360813  ]
 [6.02381355 1.02292134]
 [1.14151426 0.97292703]]
1b. sigma: [[[ 0.94138273 -0.47602571]
  [-0.47602571  0.84835602]]

 [[ 1.09308922  0.02283348]
  [ 0.02283348  1.0715613 ]]

 [[ 1.16171947  0.59624386]
  [ 0.59624386  1.08422172]]]
1b. P: [0.50805402 0.24184845 0.25009752]
```




```
1c. mu: [[2.44085985 1.92150479]
 [2.47320053 1.20976248]
 [1.42506833 1.08971477]]
1c. sigma: [[[ 1.45955738 -0.81090577]
  [-0.81090577  0.92933154]]

 [[ 0.45949192 -0.1845796 ]
  [-0.1845796   1.20372979]]

 [[ 1.37434053  0.53715228]
  [ 0.53715228  0.90186254]]]
1c. P: [0.39383242 0.25733522 0.34883236]
```



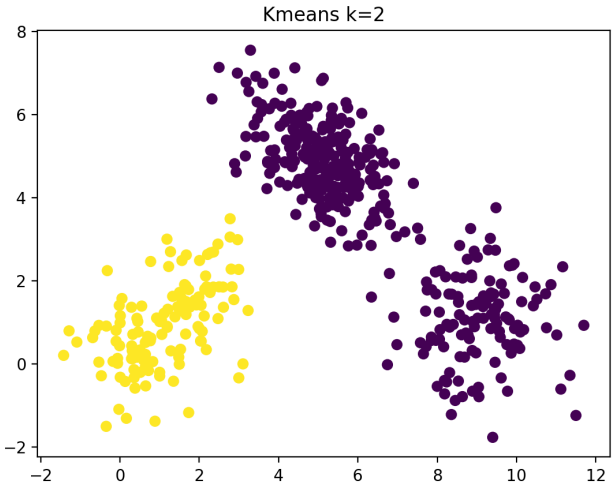
Experiment Setting:

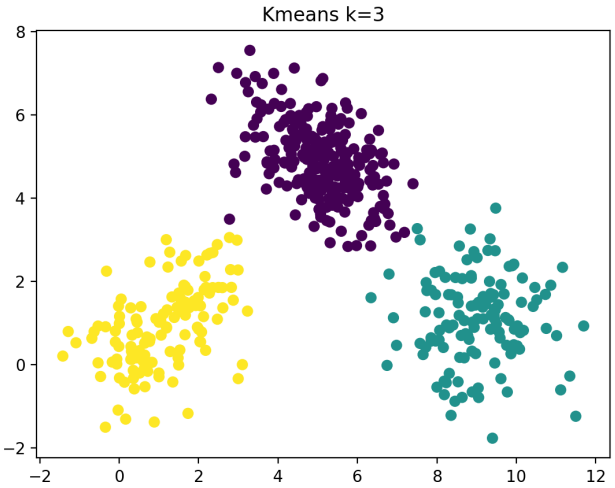
- Initialization: Randomly initialize the parameters.
- Stopping Criterion: EM iterations will stop when the lower bound average gain is below 1e-4.

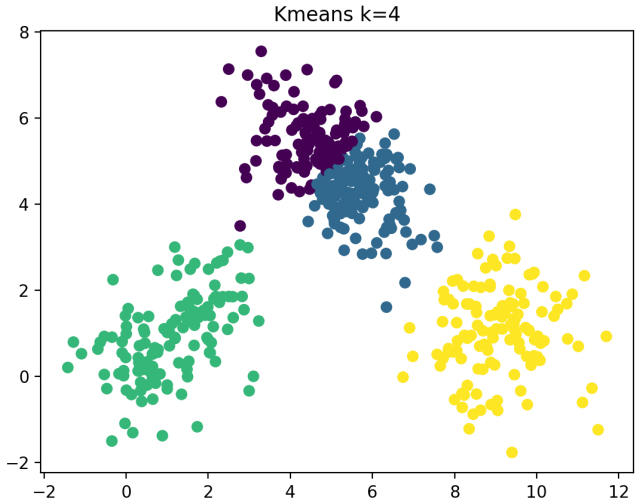
Result:

- As we can see, EM performs better in 1a than in 1b, and it performs better in 1b than in 1c since the distribution of 1c overlaps a lot which we can easily see.

Below is the result of Q2:

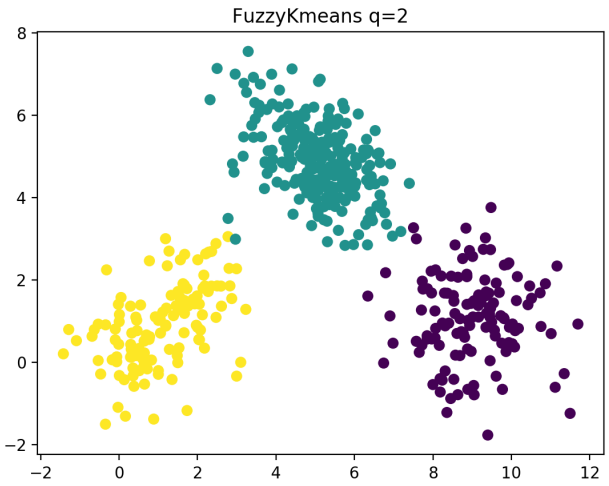


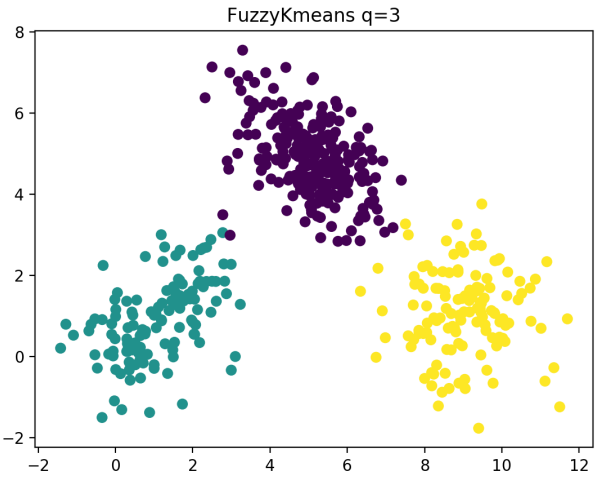




Result: As we can see, k=3 performs better than k=2 and k=4 since the distributions are composed of three different Gaussian distributions. Of course the cluster seems to be much better in k=3 than the other two.

Below is the result of Q3:





Result: Basically, under k=3 condition, q=2 and q=3 perform equally well in this experiment.