CSE454 DATA MINIG

ASSIGNMENT-1 REPORT

161044005 – Feyza Nur Akyol

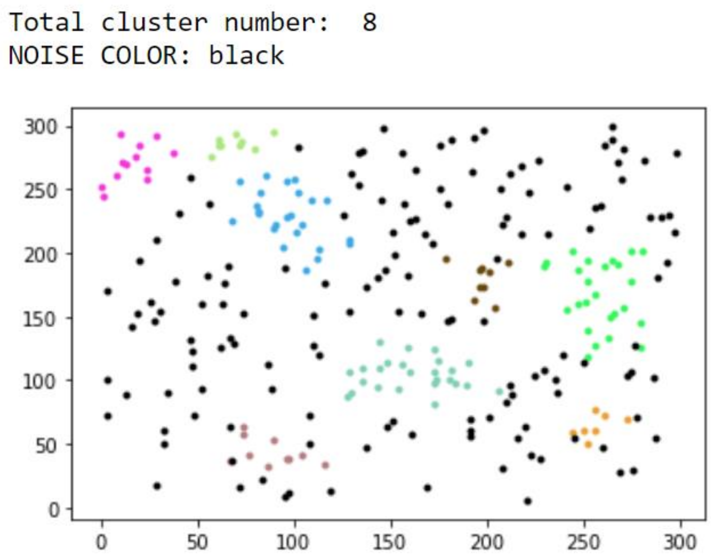
1. ***Implement DB-Scan model. You must use the algorithm mentioned in the book. You can use any programming language. Find a dataset to present the results. (You can not use any code from anywhere.)***

   I used a Python programming language to implement the algorithm. Also I created my own dataset. My dataset contains 300 random coordinates between 0 and 300 or 0 and 200.

   Also if I want to try different dataset, I can create easily new one.

2. ***Prepare an assignment report showing extracted clusters for at least 3 values of each parameter.***

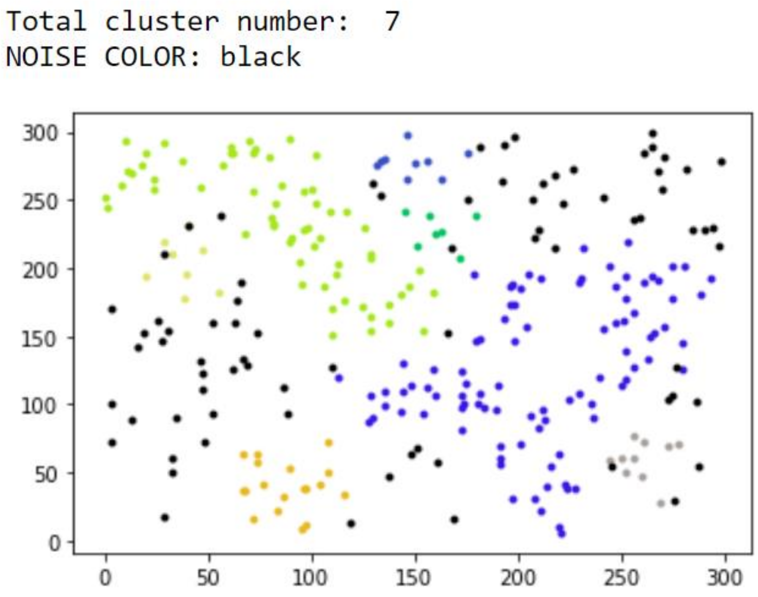
<u>*First I tried to change eps;*</u>



```
Total cluster number:  8
NOISE COLOR: black
```

300 random coordinates between 0-300

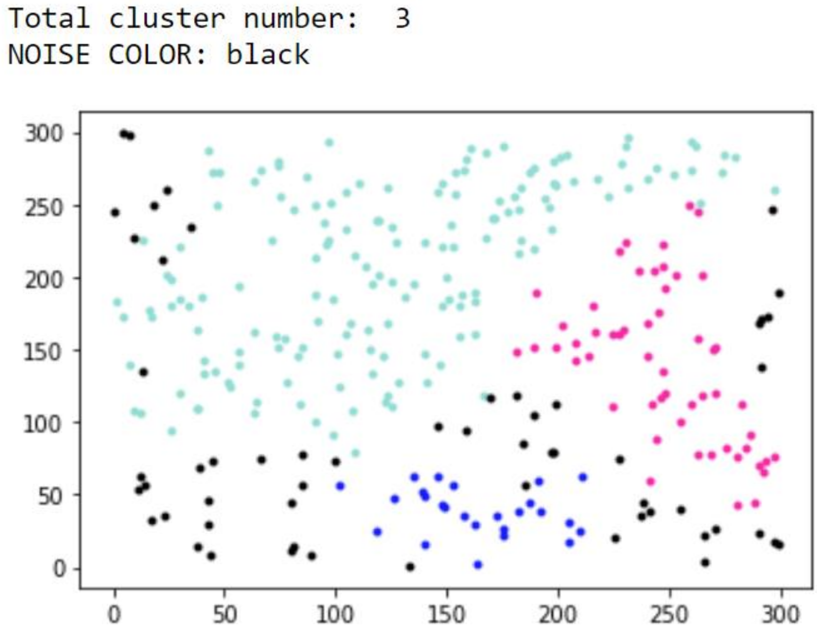
<u>eps = 20</u> minPts =7



```
Total cluster number:  7
NOISE COLOR: black
```

300 random coordinates between 0-300

eps = 23   minPts =7

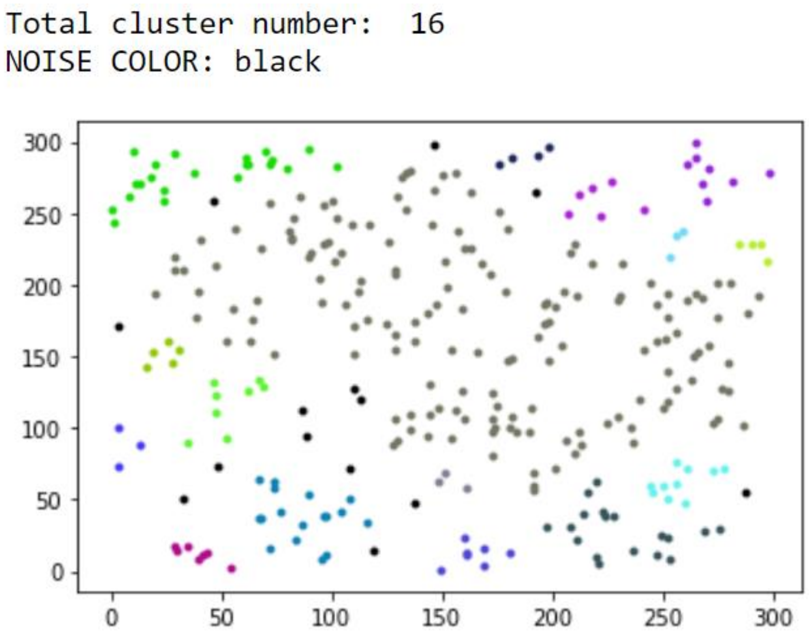


```
Total cluster number:  3
NOISE COLOR: black
```

300 random coordinates between 0-300

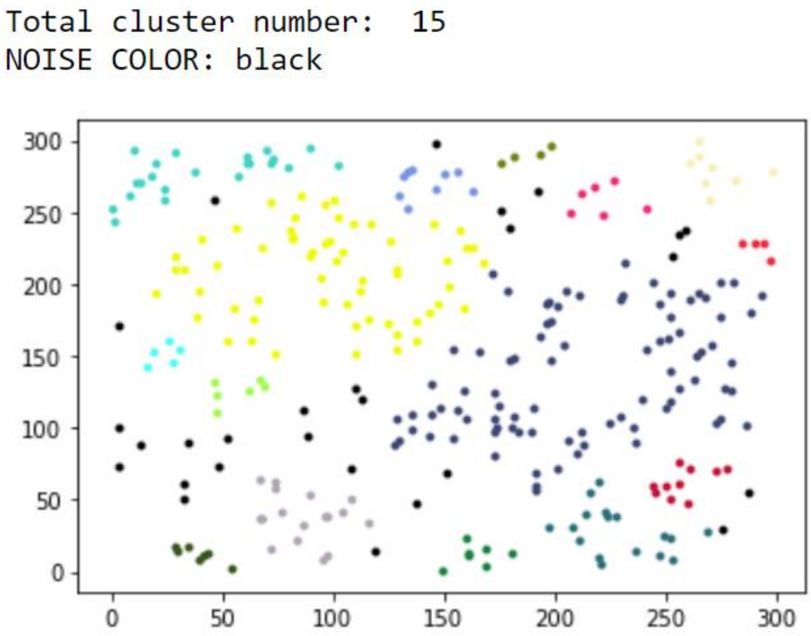
eps = 26   minPts =7

Then I change the minPts;



```
Total cluster number:  16
NOISE COLOR: black
```

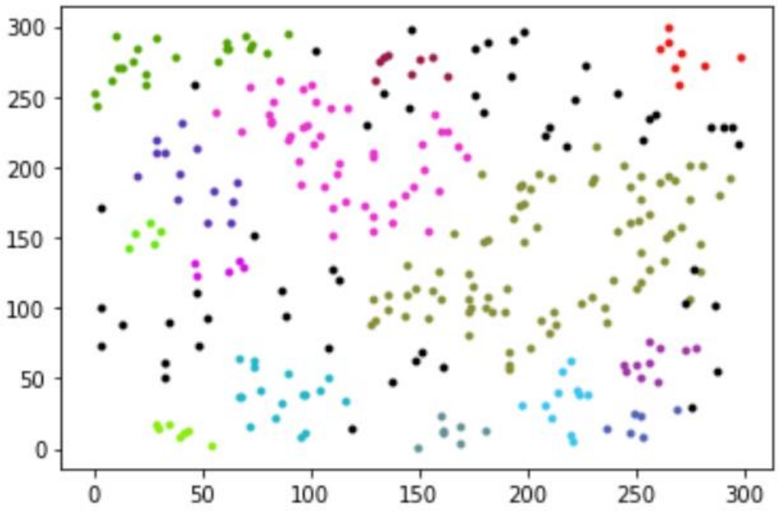
300 random coordinates between 0-300
eps = 20 minPts =3



```
Total cluster number:  15
NOISE COLOR: black
```

300 random coordinates between 0-300
eps = 20 minPts =4

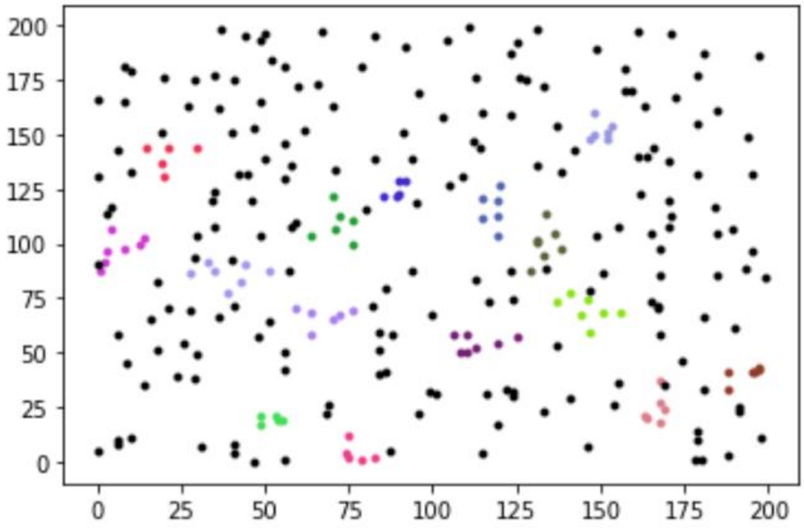
```
Total cluster number:  14
NOISE COLOR: black
```



300 random coordinates between 0-200
eps = 20 minPts =5

I tried a different dataset with 300 random coordinates between 0-200 and Eps = 5 MinPts = 10

```
Total cluster number:  15
NOISE COLOR: black
```



3. In the report, write a discussion about how the parameters effect the results.

Like I expected, when "eps" getting bigger, number of the cluster is decreasing. Also, when "minPoints" getting smaller, number of the cluster is increasing.

But I think, there is not always a linear increasing or decreasing.

4. In the report, give a technique to automatically decide on the parameters of DB-Dcan?

There are different methods to do that.

The determination of the MinPts parameter is very difficult, so it is often chosen experimantal depending on the datasets.

$$MinPts = \begin{cases} round(d_p + 0.5) \;\; for \;\; dim(X) == 2 \\ round(d_p - 0.5) \;\; for \;\; dim(X) > 2 \end{cases}, \tag{9}$$

In the Eps parameter, the main issue is to accurately determine the sharp increments of the distances.

(A NEW METHOD FOR AUTOMATIC DETERMINING OF THE DBSCAN PARAMETERS Artur Starczewski1,∗, Piotr Goetzen2, Meng Joo Er3)