Data Mining
Assignment 2

-Suraj Prathik Kumar (2016101)

Assumptions for the Code - Input: Number of Clusters(k) and Number of data Points (Including the Points)

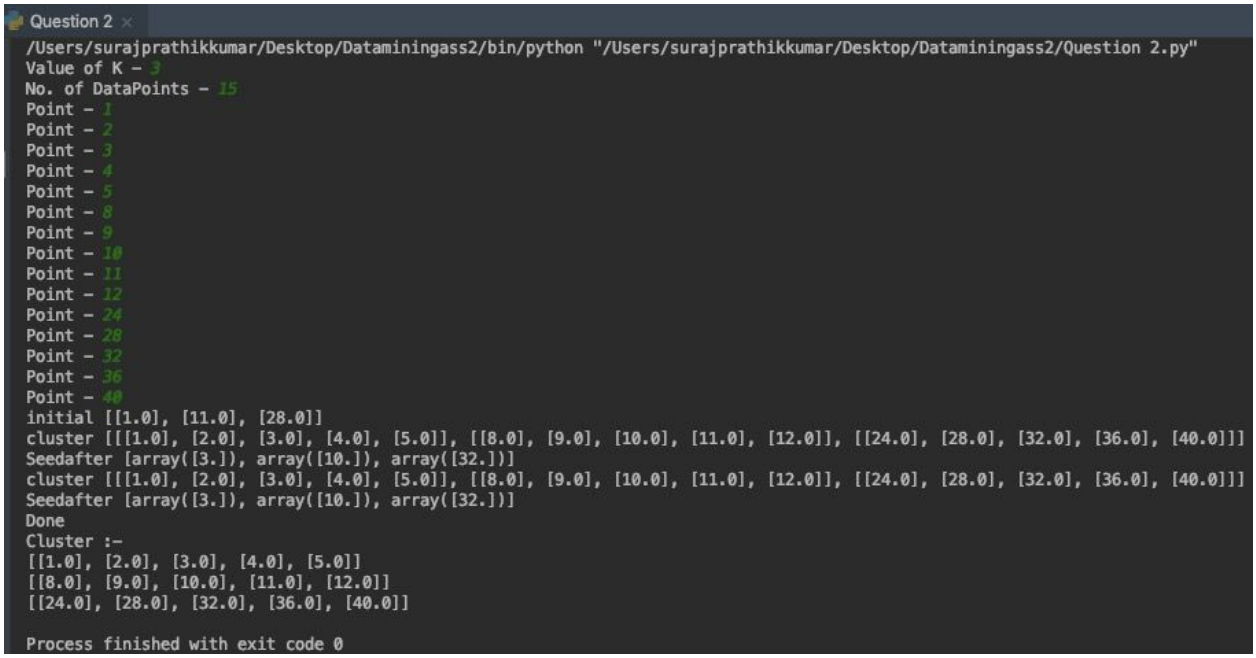
Question 1: Code Submitted : File Name - Question 1.py

Question 2: Code Submitted : File Name - Question 2.py

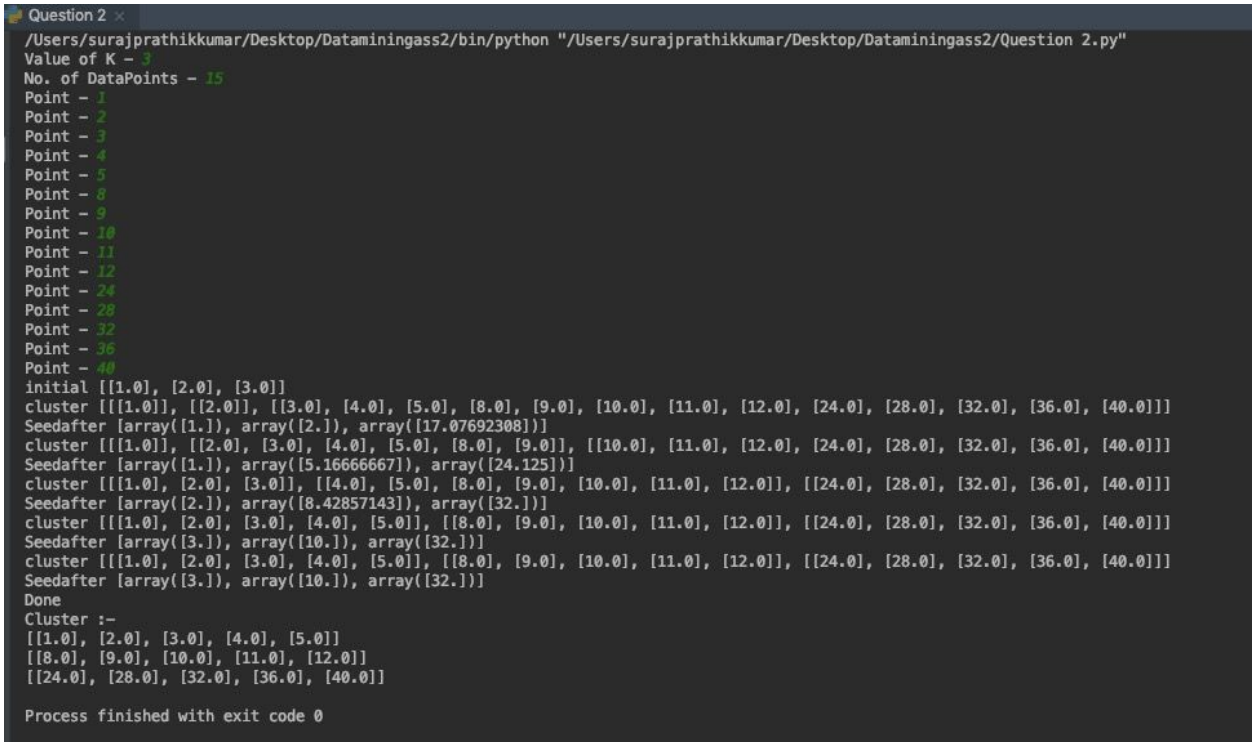
Question 3: Code Submitted : File Name - Question 3.py

Question 4:

a) Yes, the algorithm is able to find the right clusters given in the Question



b) Yes, the algorithm is able to find the right clusters given in the Question



c) In First and Second part, Initial Seeds are different {1, 11, 28} and {1, 2, 3} respectively.The final Cluster Formed are same but the number of it iteration are less in First part as compared to Second part.Choice of seed play in important factor in determining Faster Convergence. The seeds should not be very close to each other and should fall in the range of similar datapoints.
Example: Here there are 3 range [1-5] ,[8-11] and [24-40]
Selecting seed from each of the Range will give us faster Convergence and take Less iterations than selecting all the seeds from same range.