Assignment-6: Clustering

1. What is Clustering? Explain with an example in what way can it be helpful?
2. What are the different types of clustering? (explain with examples)
3. Explain at least two problems/challenges with k-means clustering and how we can solve them.
4. **K-means clustering:** Assume We have the following dataset:

Point	**X**	**Y**
K1	2	10
K2	2	5
K3	8	4
K4	5	8
K5	7	5
K6	6	4
K7	1	2
K8	4	9

A. Calculate the distance matrix based on the Euclidean distance **$d(k1,k2)=sqrt((x_1-x_2)^2+(y_1-y_2)^2))$**
B. n_cluster=3 and centroids = K1, K4, and K7. Show (in 3 iterations) how the dataset is clustering into n_cluster.
C. Plot the result for each iteration.

5. Use the previous dataset and apply hierarchical clustering (single-link, complete-link, average-link agglomerative clustering). You need to show/explain your steps and the dendrograms.
6. Apply the k-Means clustering from scratch on the previous dataset. Use the initial centroids and the number of clusters used in Q1. You must add comments to the code.

What to Submit: Canvas Classroom.

1. Assignment6YourName.pdf (e.g., Assignment-6AliAburas.pdf). That contains all the answers to the above questions + screen-shots of your code! Note: Dont zip the pdf file!
2. Python code