For Part A and B, use the Aggregation dataset that has been attached.

PART A. Distance-Based (k-means)

1- Implement the k-means model.
2- Run the algorithm on the dataset for k = {1,2,3,4,5,6,7,8,9,10}.
3- Obtain the cost function for every 10 modes of the previous section.
4- Plot the cost function corresponding to the previous section
5- Find the optimal k in this problem.
6- From the 10 models obtained in Part 2, plot the classification result by 5 models with better accuracy.

PART B. Density-Based (DBSCAN)

1- Consider at least 10 different values for MinPts and 10 different values for ε of your choice.
2- Run the DBSCAN model on the data, corresponding to the two-by-two values of the parameters in the previous section. (Note that at least 10 * 10 = 100 models are created)
3- Using the mean squared error (MSE) method, plot the MSE diagram corresponding to the models obtained in the previous section for different parameters. (plot a 3D diagrams: MinPts, e, MSE).
4- From the models obtained in Section 2, plot the classification result by 5 models with better accuracy.
5- Compare the result obtained in the previous section with what was obtained in Section 6 of exercise A and analyze the results.

PART C. Kmeans and DBSCAN and Agglomerative Clustering

In this homework, you will work with 4 datasets as name of **noisy_circles, noisy_moons, blobs, no_structure.**
It shows characteristics of different clustering algorithms on 2-dimensional datasets, with the exception of the last dataset.

1- Load and plot each dataset to visualize the dataset.
2- Train a Kmeans on your datasets.
3- Train a DBSCAN on your datasets and find best epsilon for each dataset.
4- Train an Agglomerative clustering on your datasets.
5- Plot and compare your results for different clustering algorithms and different datasets.

Good Luck.