For this homework you are no longer given a code skeleton. It is up to you to decide how to structure your code. (PLEASE NOTE, YOU ARE EXPECTED TO IMPLEMENT EVERYTHING IN JUPYTER NOTEBOOK AND SUBMIT THE NOTEBOOK AS YOUR SOLUTION).

**Cluster Mania!**

You are given a training dataset in CSV format (hw4\_cluster.csv). The files each contain 40 rows with 2 columns. Column 1 & 2 are the features. There are no labels for this dataset. Your goal for this assignment is to implement different clustering algorithms and run them on this dataset. For this assignment you can assume the distance function is the Euclidean distance.

1) (35 points) Implement a generalized K-means algorithm. You should have a single function that takes in as input the data points, K, and some other hyperparameters, specified below. The function should return K sets of data points. Each set corresponding to one cluster.

The hyperparameters your functions should support and the values they can take are:

-Max\_iter: max number of iterations to run the algorithm.

-K: number of clusters

Note that your stopping condition should have two parts:

1) stop if you reach the max iterations

2) stop if no change is made to the clusters in the last step.

You will be running this code in question 3 of the assignment. For this part you just need to

implement the function.

2) (15 points) SSE score.

In this part of the assignment, you are implementing a function that calculates the SSE for a list of clusters. The function should take in a list of clusters (such as the output of the last function you implemented) and return a single SSE score.

3) (15 points) Finding best K.

Run the code you implemented in question 1 for k=2,3,4,5. Set the Max\_iterations to be 100.

Calculate the SSE for each K using the function in question 2 and use these scores to pick the best K. What is the best K?