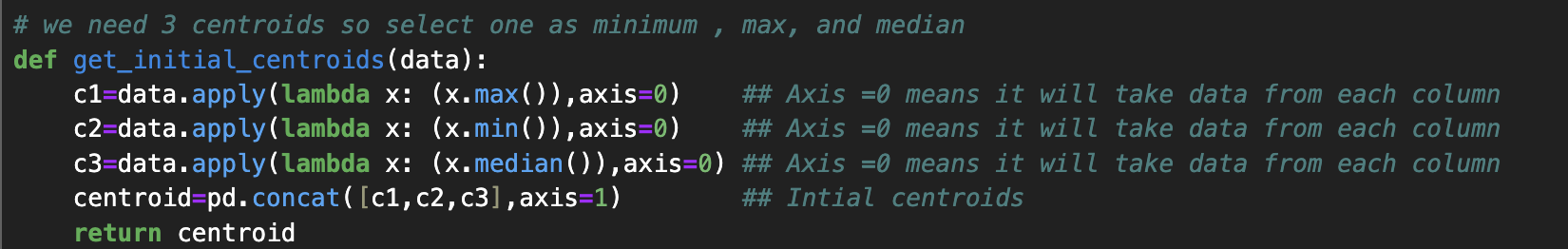
Assignment 1

K-Means and K-medoids

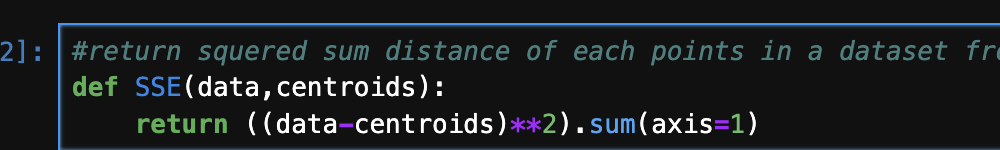
List of user defined functions: -

1-

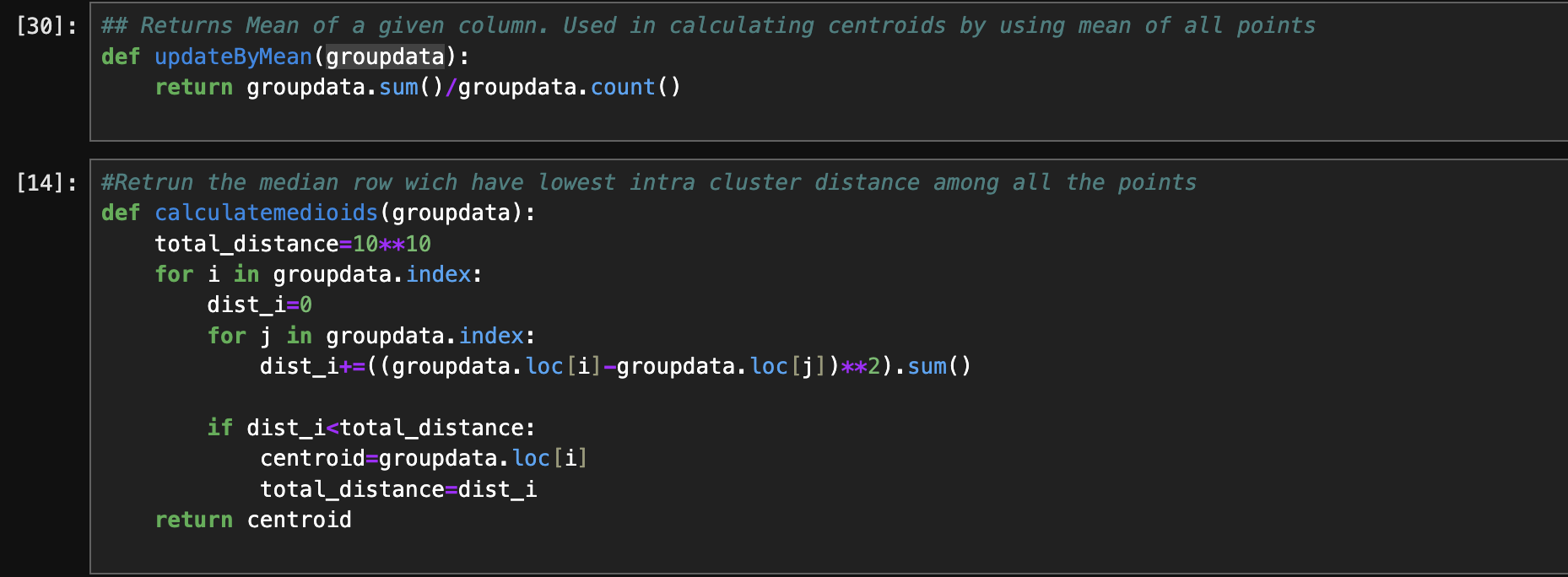


This function initializes 3 centroids which looks like this: -



2-

It finds SSE from one given centroid and given pandas data frame

4-

This two-function used to calculate medoids. Given a pandas data frame 1st function calculate mean of each column/feature. And other function returns the median of the data frame given.

It takes data frame of each cluster and returns the new centroids.

5-

A black screen with blue text

Description automatically generated

This function will return labels as per the given data frame and the 3 centroids.

Here for each centroid, we apply SSE to whole dataset. For each record we will get sse w.r.t to each cluster as shown below:-

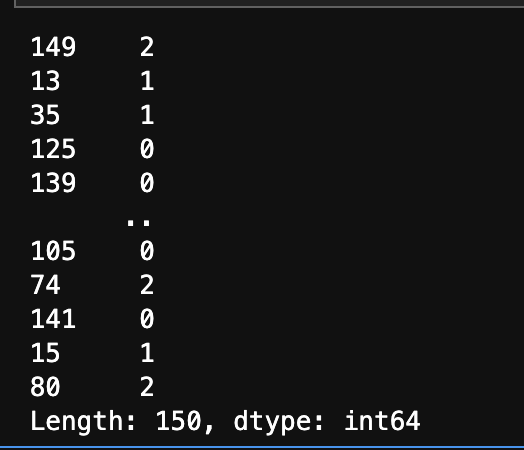
Distance dataframe look like this:-

A screenshot of a black screen

Description automatically generated

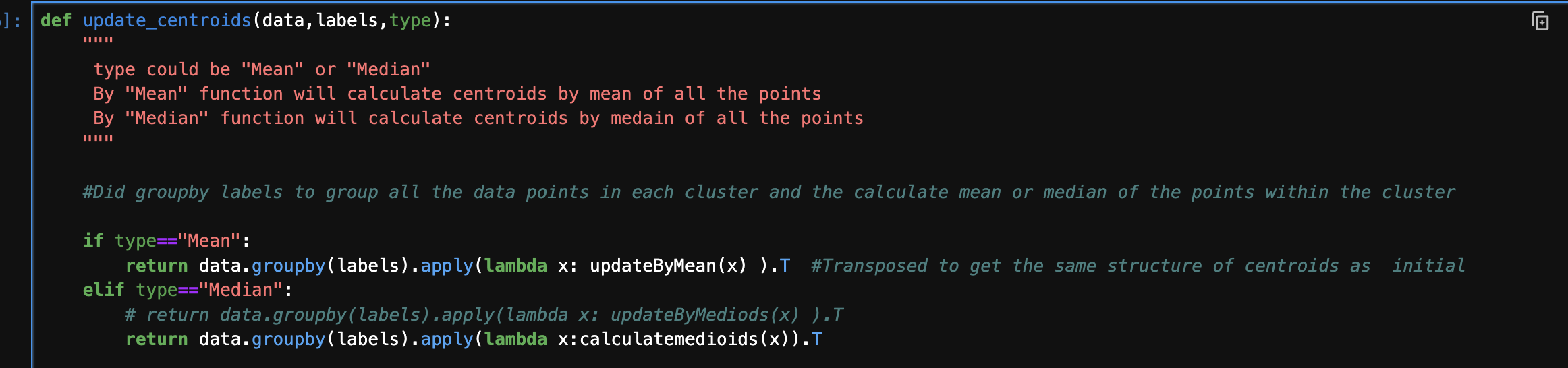
Idxmin will return the index with lowest value (lowest SSE value) along axis =1 ,for each record which will be its cluster (0,1,2)

Returned data frame looks like this:-



Labeled for each row/record.

6-



This function used to calculate centroid as per the inputs (mean or median).

I used group by to group data by labels so 3 groups will be created and will be passed one be one to calculatemedioids or updateByMean methods and will get 3 new centroids.

Final transposed it to make it dimension correct

7:-

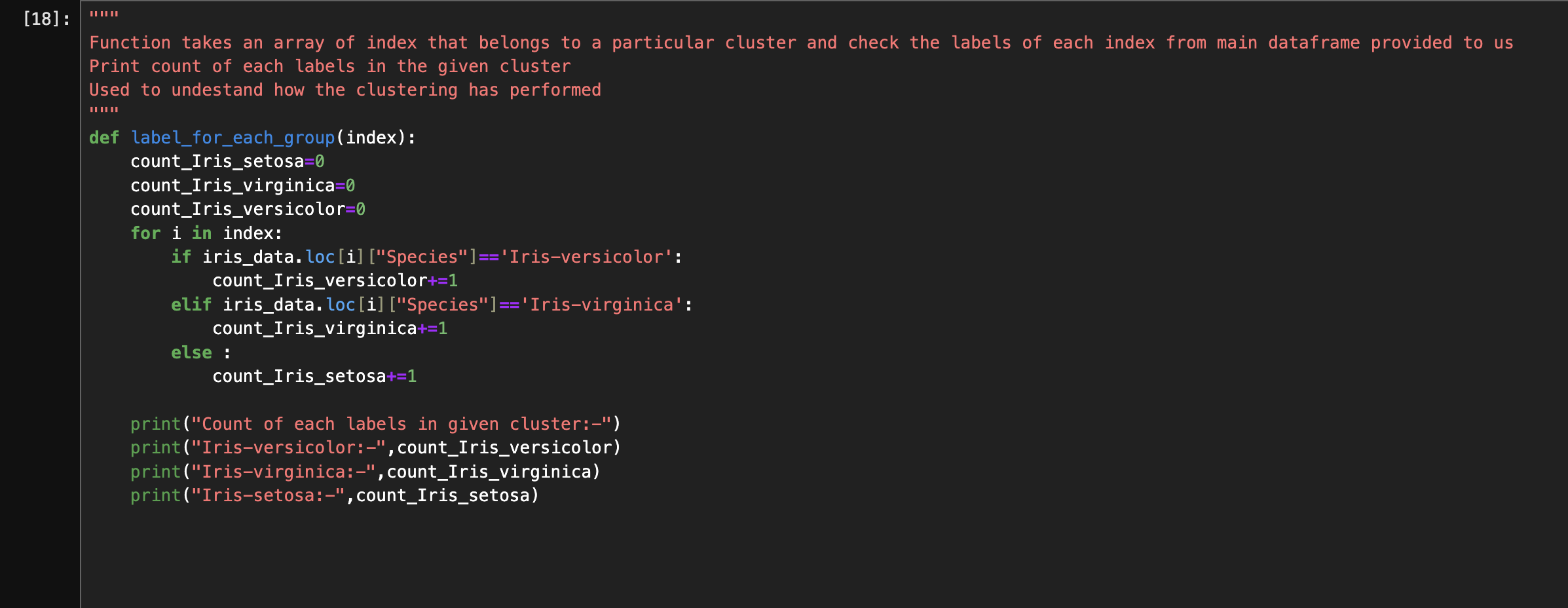


This function is used to plot the graphs. We have 4 features in 2d we can plot 6 types of graphs, but I only plotted 4 graphs in one figure for each iteration.

I will be calling this function in each iteration.

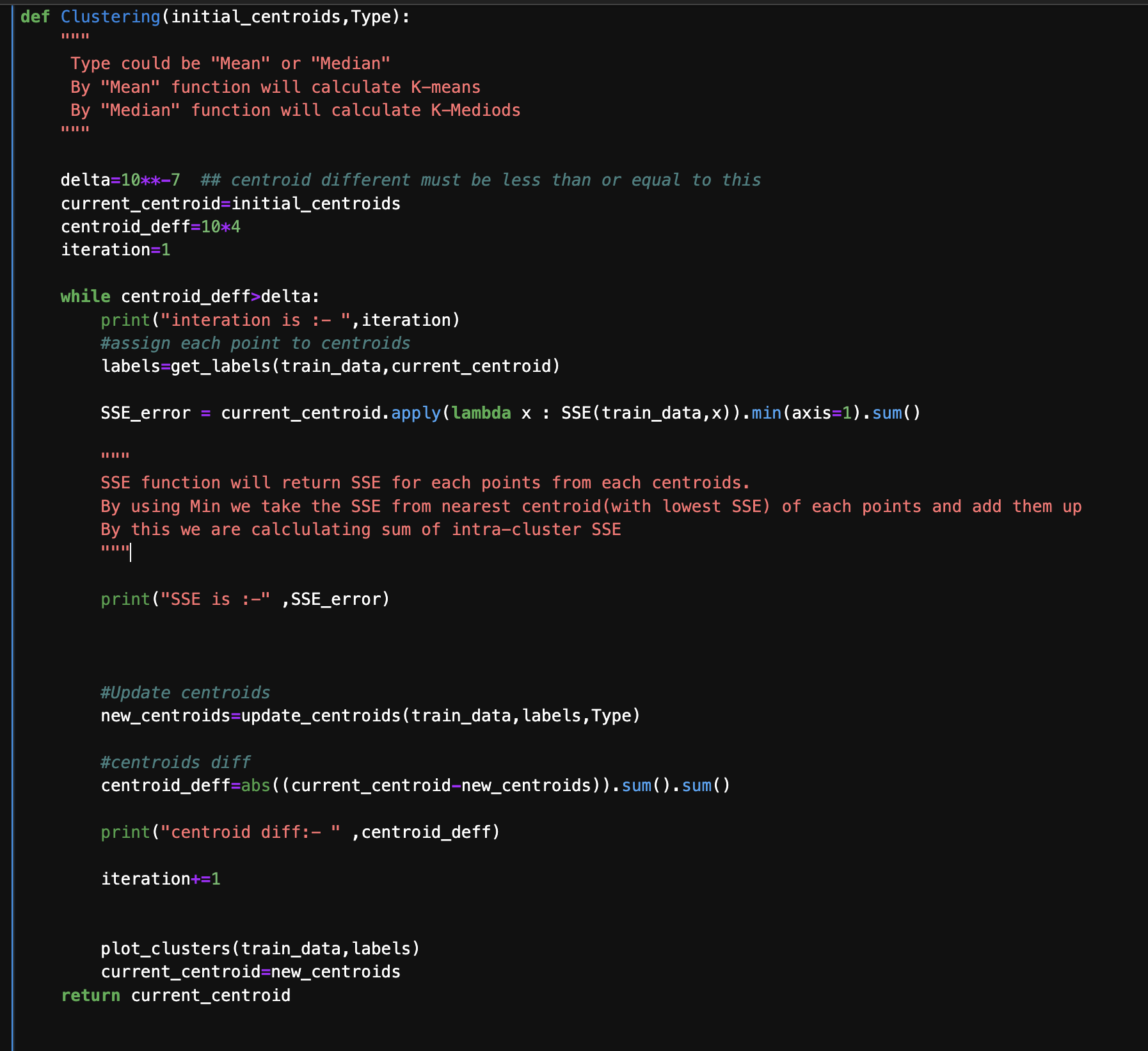
Took color (c) as labels so for each index in dataset it will pick color as per labels

8:-



As given data is pre labeled, we want to see how many records are correctly clustered.

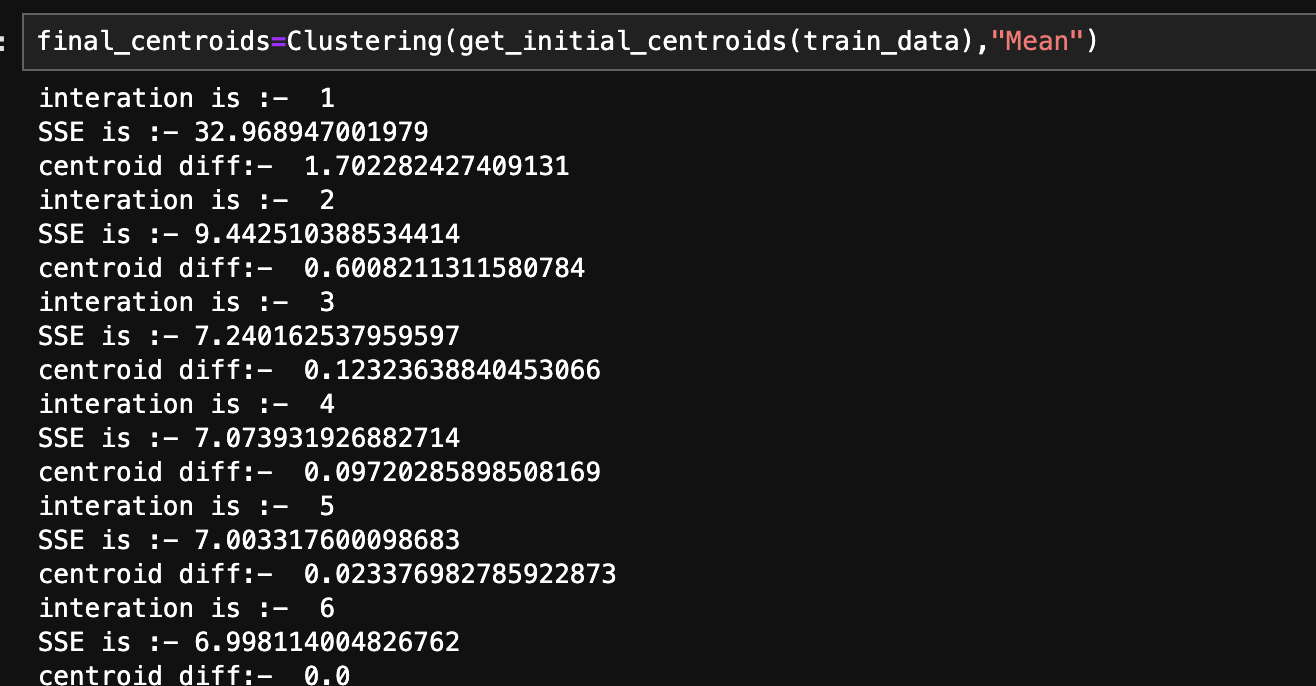
This function will take index of records belonging to a particular cluster and print all the labeled count present in that cluster.

9:-

This is main function to evaluate k-means or k-medoids algorithm.

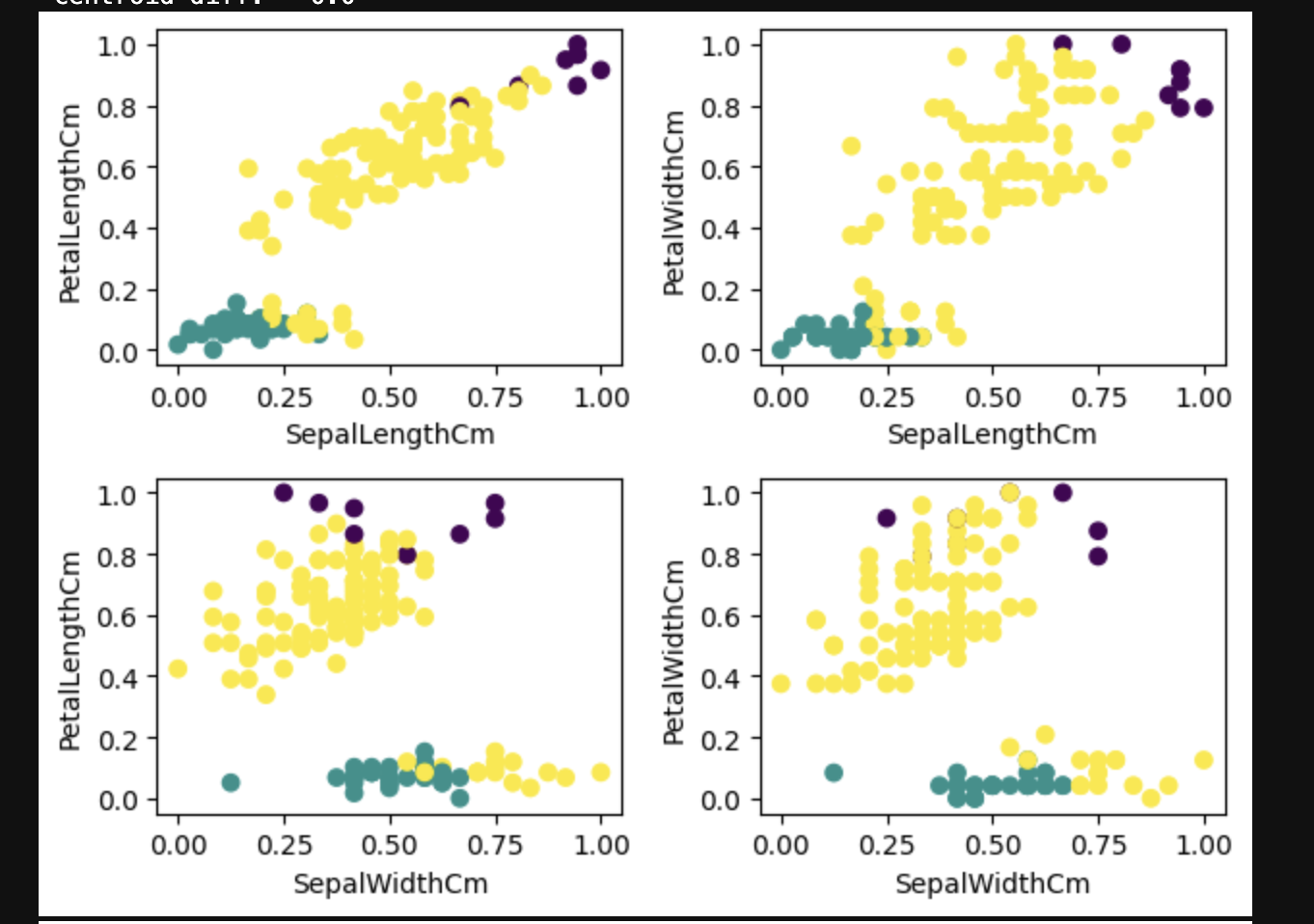
Uses all the helper function defined earlier and returns the final centroids, also prints SSE, plot clustered graph at each iterations.

OUTPUT of the code for K-means: -

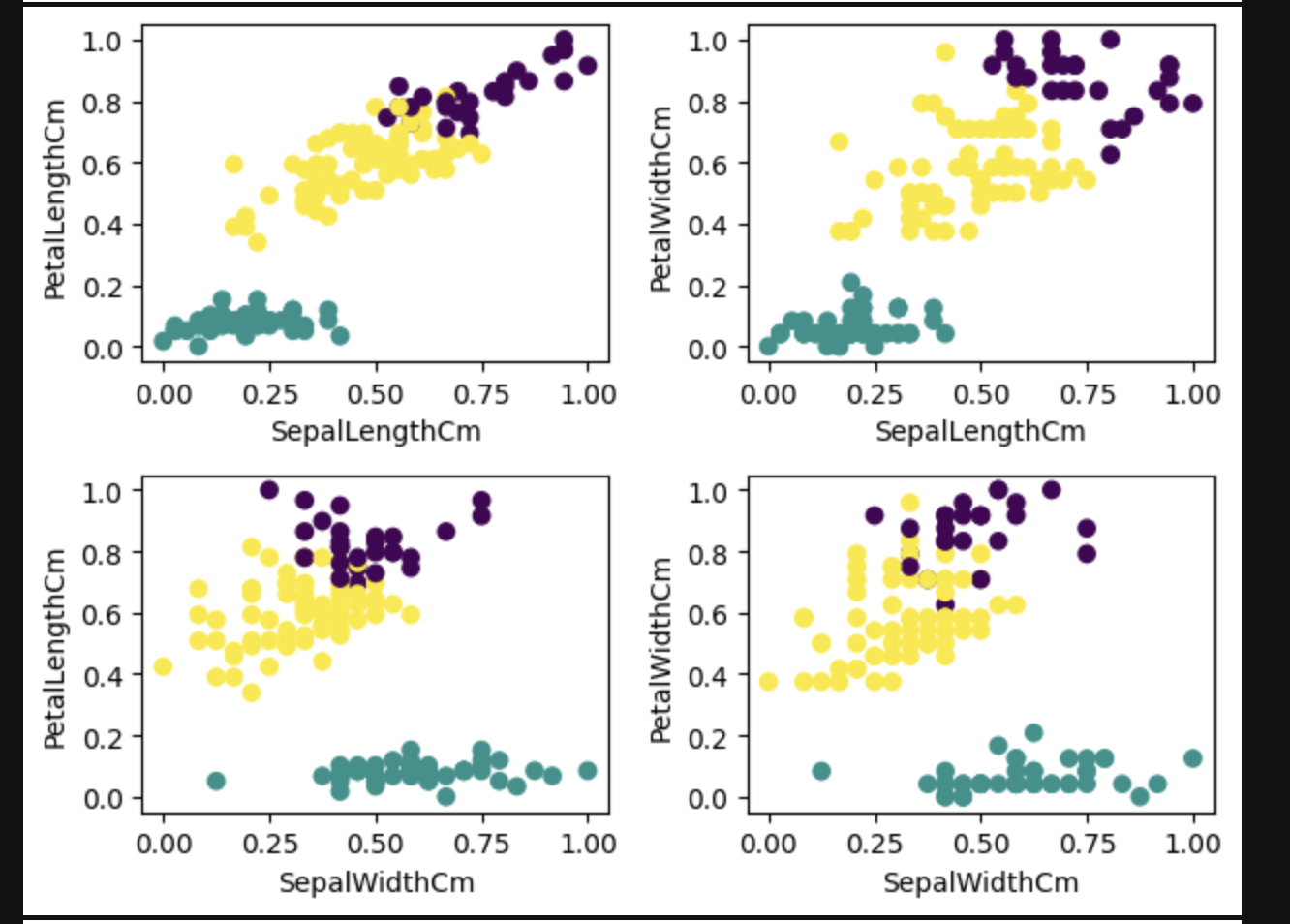


GRAPHS for each iteration: -

1st Iterations



2nd Iteration: -



3rd Iteration: -

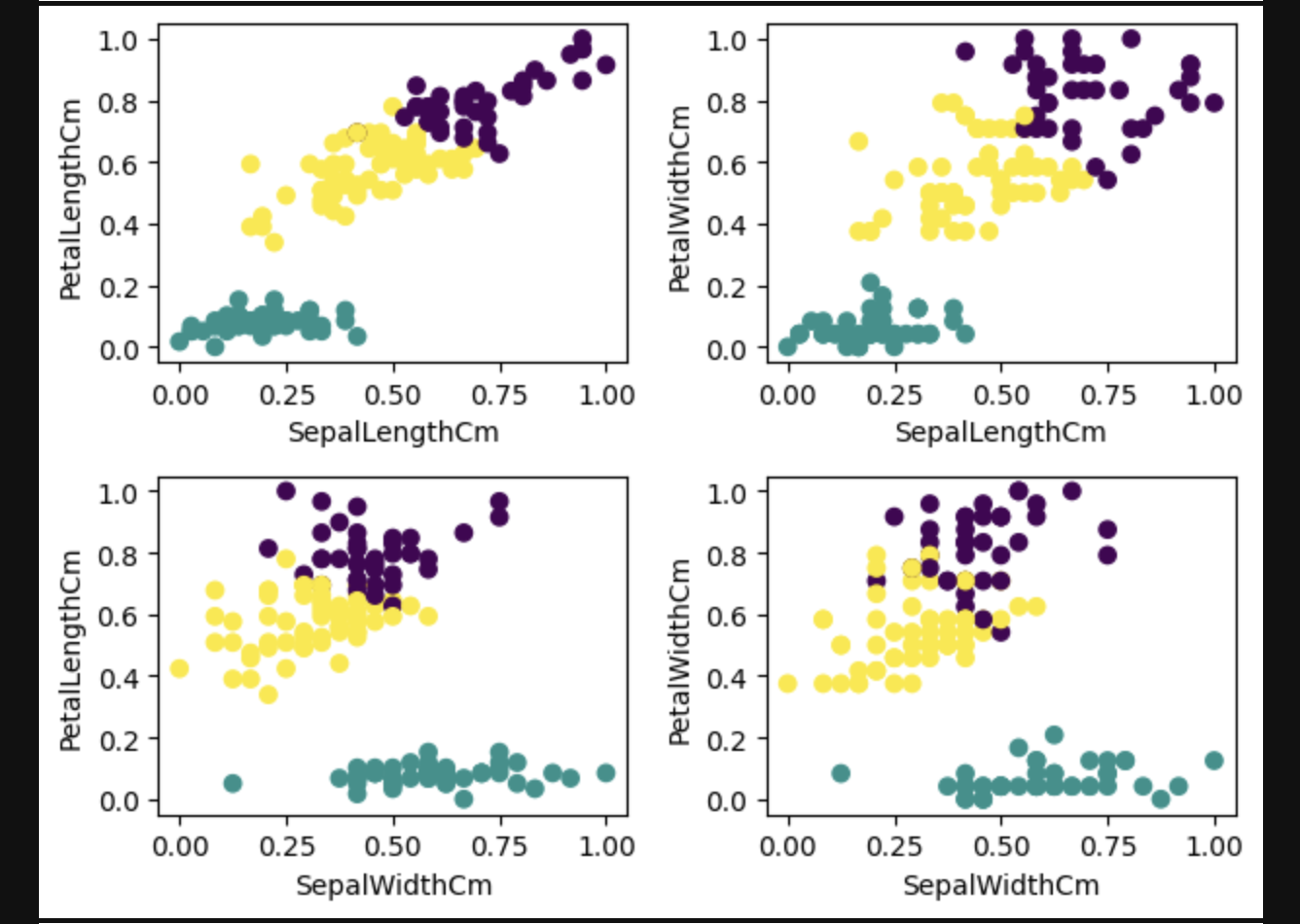
A group of colored dots

Description automatically generated with medium confidence

4th Iteration: -

A group of colored dots

Description automatically generated

5th :-

6th :-

A group of dots with different colors

Description automatically generated with medium confidence

OUTPUT of the code for K-medoids: -

A screenshot of a computer

Description automatically generated