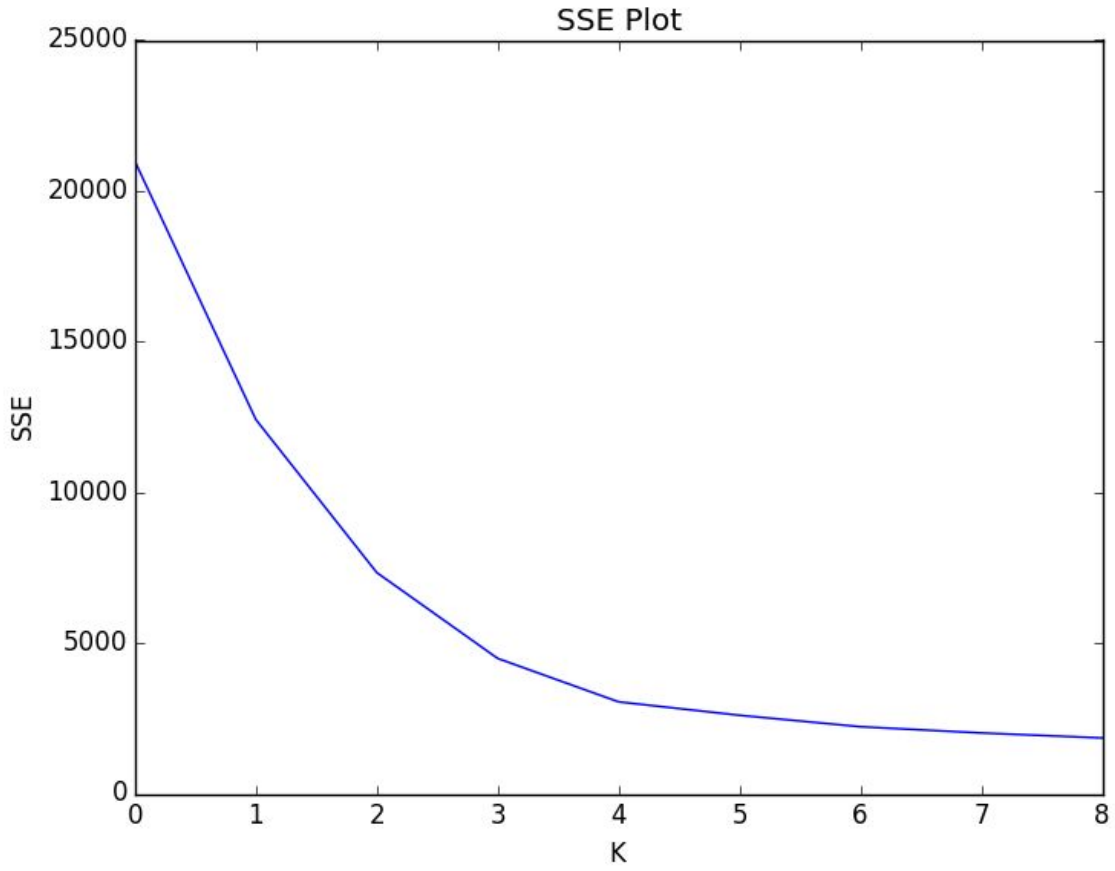
Tyler Paulsen
10/12/20015
csci-420-01
Homework 5

1) Estimate of four hours to complete the homework

4)

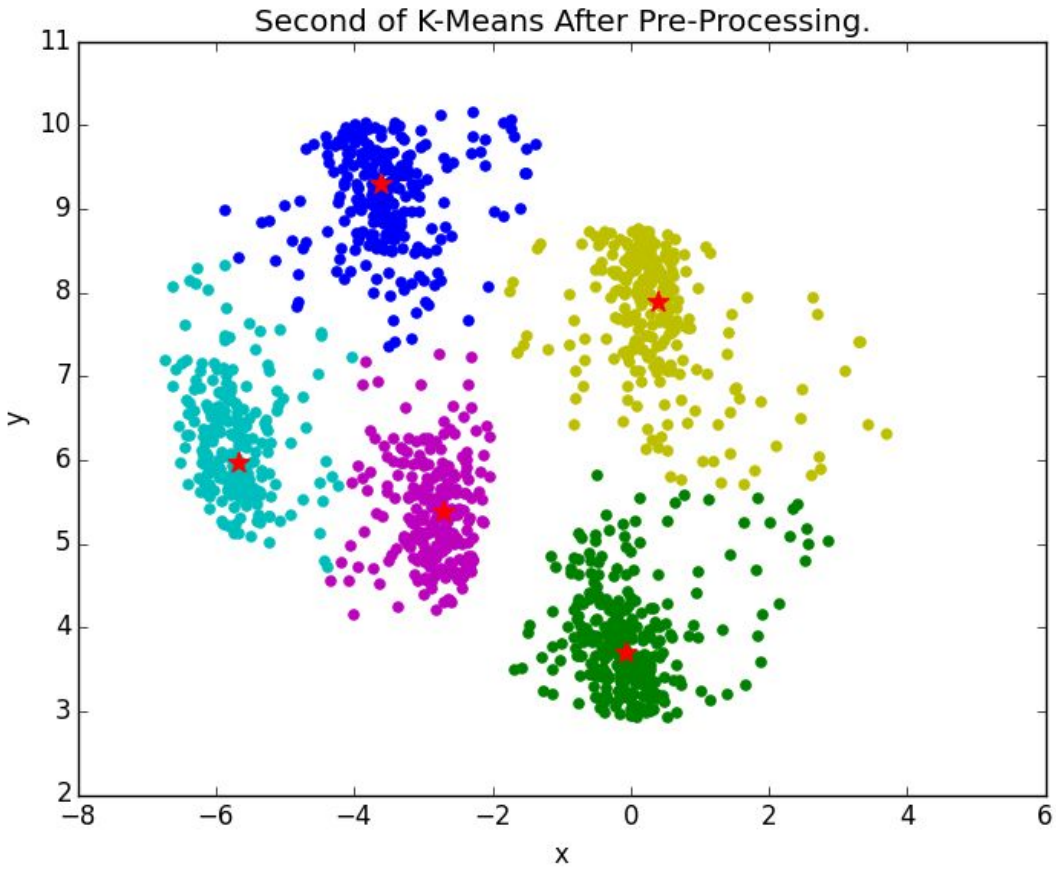
a. Yes, the x and y values were rotated 45 degrees to make the data axially aligned.
b. Yes, the k-means clustering algorithm was ran twice. First time to get an estimate for the cluster position, then the distance average and standard deviation for each cluster was calculated. The list was then iterated through, and any value that had a distance from its corresponding cluster that was further than 1 standard deviation from the mean was removed.
c. The euclidian distance was used.
d. The centroid was used for the prototype.
e. Plot of the Sum of Squared Error.



f. The value of k that worked the best was 5 with a SSE value close to 1461.32
g. The sorted cluster size is as follows: [220, 238, 245, 253, 283]

Tyler Paulsen
10/12/20015
csci-420-01
Homework 5

h. The k-Means color map per cluster.



i. By far the hardest thing was working with the python lists to get this algorithm working. Eventually, the program was converted to using numpy arrays in order to decrease computation time, and increase readability.

5) The homework took 8 hours.