Strategy 1

K-Value Points 3 & 5. Intitial and Final Centroids :

Initital

[[3.53350737 0.33198894]

[4.95728696 6.90897984]

[5.14167285 5.71626939]]

[5.07631894 3.30296197]

[1.87131855 3.43365823]

[5.02471033 8.23879873]

[8.46078528 2.85204573]

[7.68097556 0.83542043]

Final

A screenshot of a computer program

Description automatically generated

A graph of different types of lines

Description automatically generated with medium confidence

A line graph with numbers and a line

Description automatically generated

Strategy 2

K-Values 4&6 & Initial and Final Centroids :

Initial Centroids :

[2.37650624 8.15241778]

[2.38952606 7.22195564]

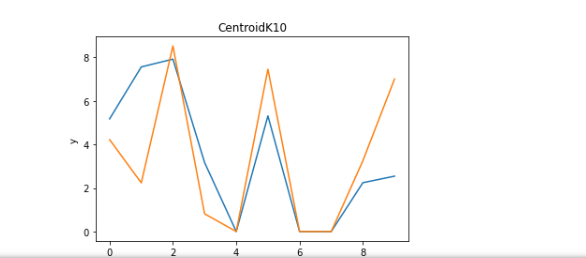
Final Centroids:

A screenshot of a computer program

Description automatically generated

A line graph with numbers and lines

Description automatically generated



A line graph with numbers and a white background

Description automatically generated

My immediate analysis of the two strategies is that we see that the consistency of using K++means, average maximum distance, creates more accuracy, and less objective loss vs the standard K means clustering of random points. We see that clustering is more effective, also when the number of centroids is increased, and that objective loss is also less. We noticed that as we utilize K++means and not standard random clustering we ‘lose’ or gain bad centroids, so we have to back fill such data with soft constraints to compensate for the “dead” data to gain good centroids and objective loss as we increase our centroid and cluster amount. We conclude that we would want to predict, and classify superior and accurate data that we would choose K-Means average max distance from centroid method with maximum allowable clusters to classify our data when we are not provided labels.