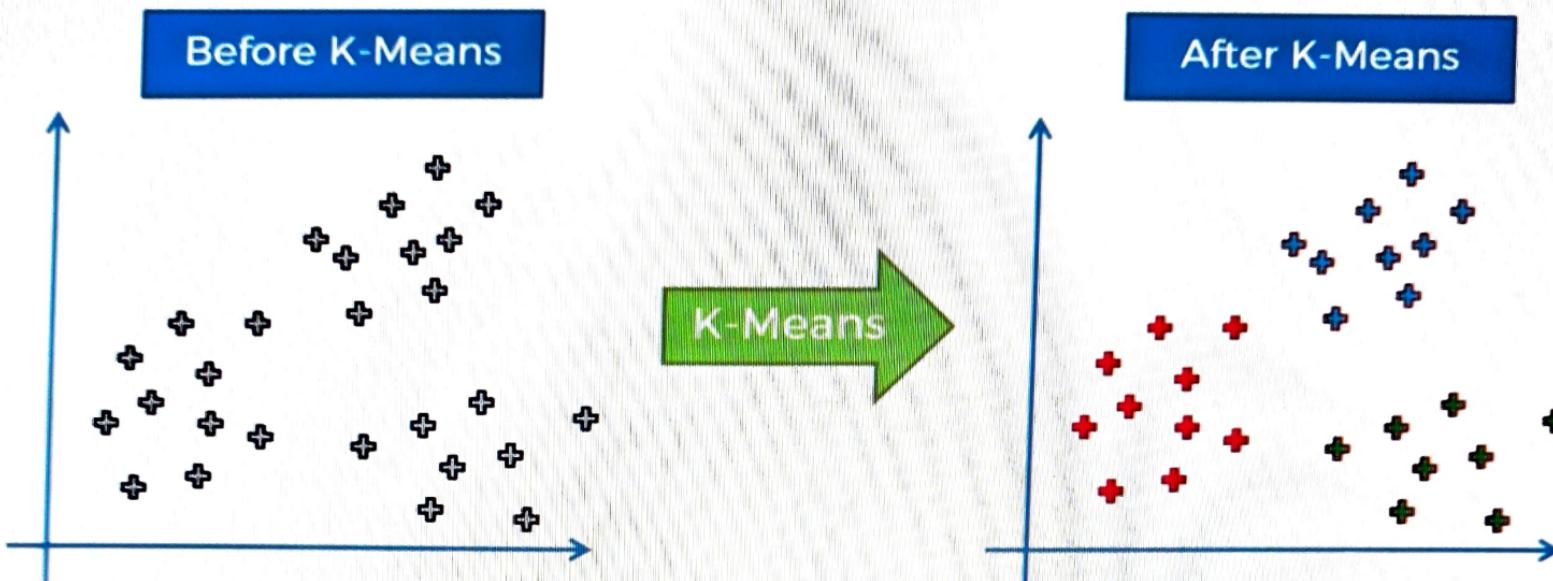


K-Means Intuition: Understanding K-Means

And in this section we're talking about the K means clustering algorithm.

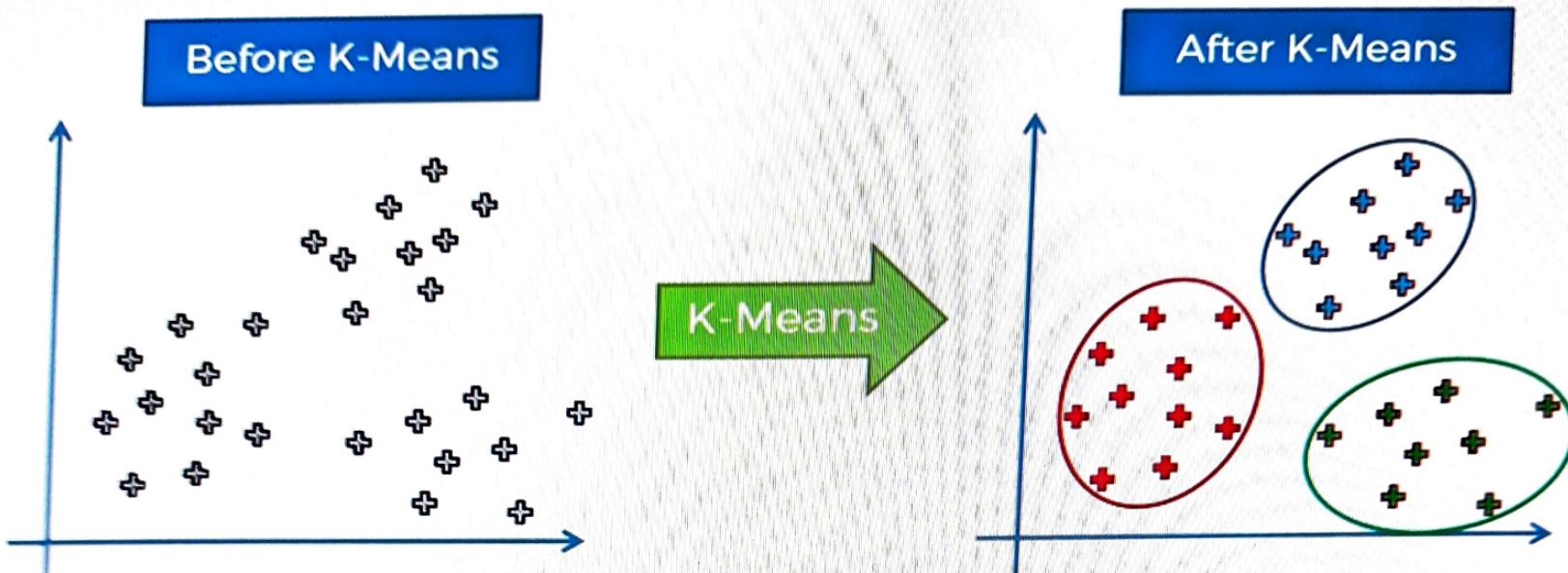
What K-Means does for you



205 people have written a note here.

allows you to very easily identify those clusters are actually called
clusters of data points in your

What K-Means does for you



205 people have written a note here.

How did it do that ?

STEP 1: Choose the number K of clusters



STEP 2: Select at random K points, the centroids (not necessarily from your dataset)



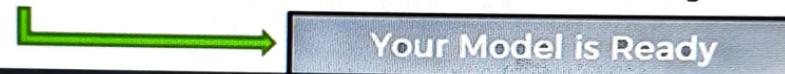
STEP 3: Assign each data point to the closest centroid → That forms K clusters



STEP 4: Compute and place the new centroid of each cluster



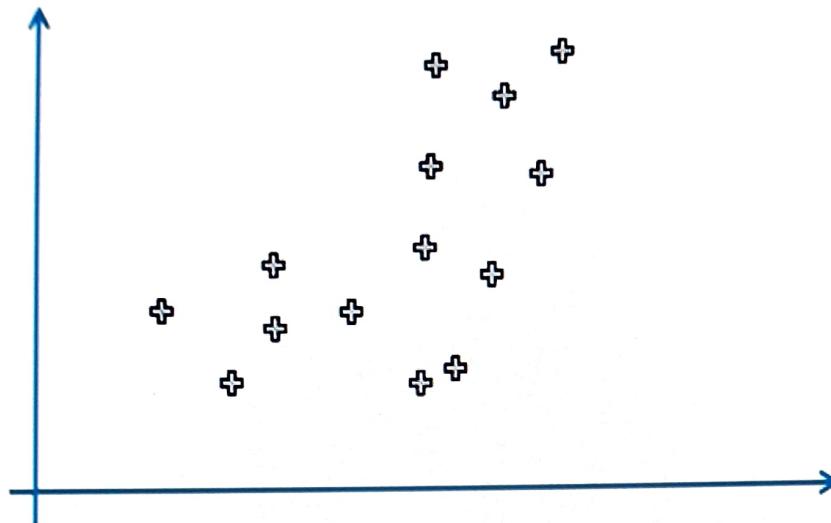
STEP 5: Reassign each data point to the new closest centroid.
If any reassignment took place, go to STEP 4, otherwise go to FIN.



So this slide will be a great reference point after we discussed this visual exercise that we're going

K-Means algorithm

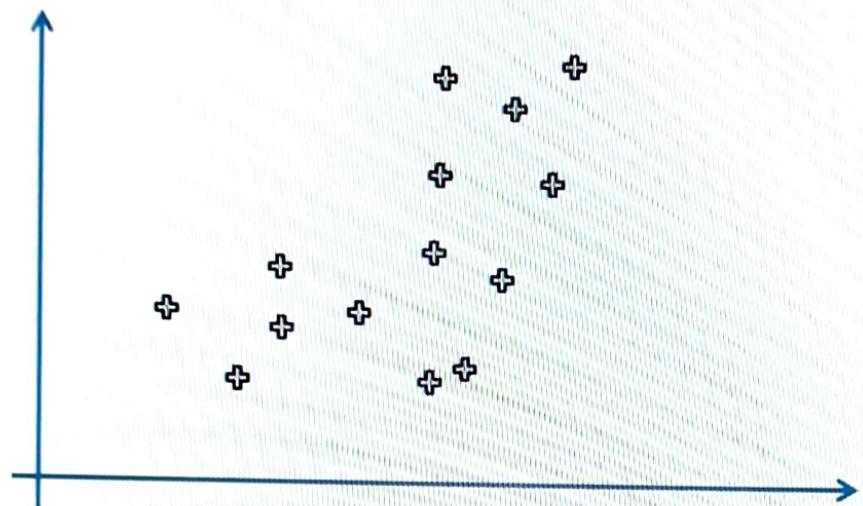
STEP 1: Choose the number K of clusters: $K = 2$



But in this case just so that we can get very comfortable with this algorithm.

K-Means algorithm

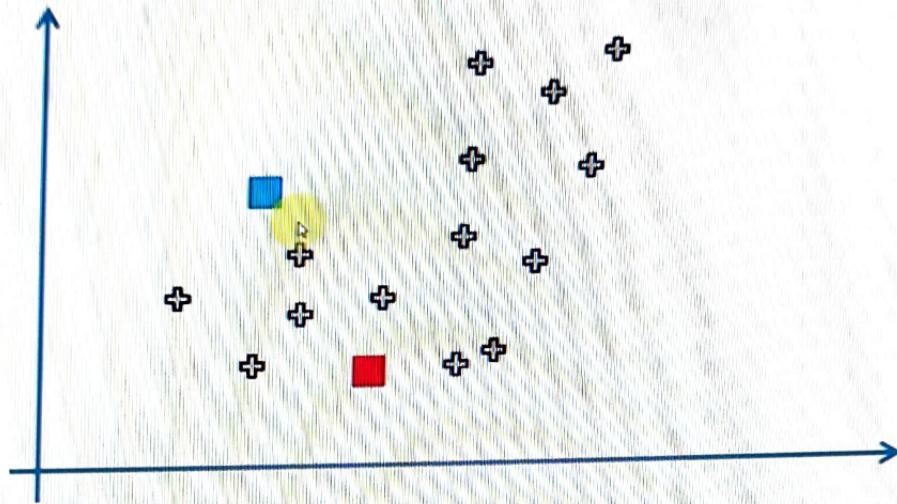
STEP 2: Select at random K points, the centroids (not necessarily from your dataset)



And then a step to select at random k points which will be the centroid of your clusters and not necessarily

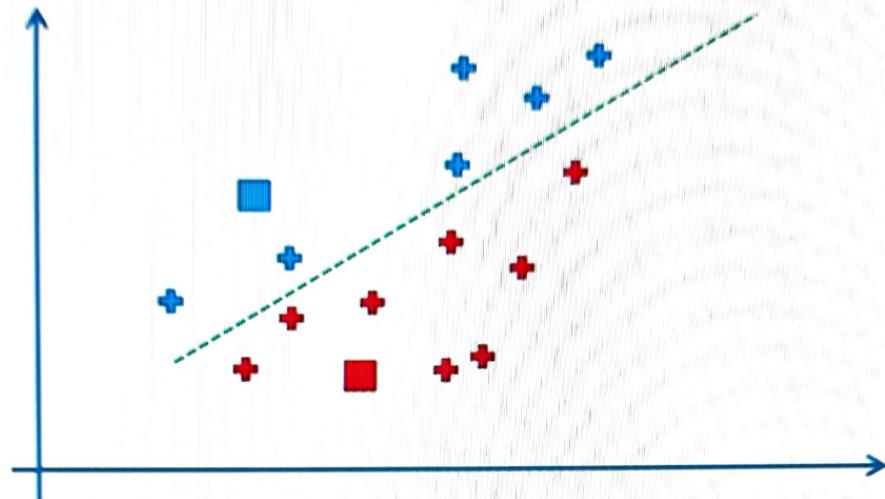
K-Means algorithm

STEP 3: Assign each data point to the closest centroid → That forms K clusters



K-Means algorithm

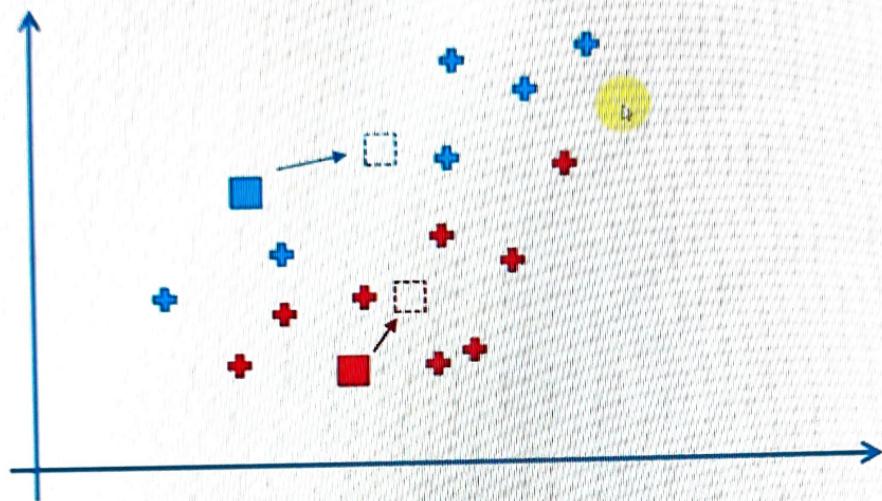
STEP 3: Assign each data point to the closest centroid \rightarrow That forms K clusters



Right but at the same time in mathematics and in data science
there's lots of different type of distances

K-Means algorithm

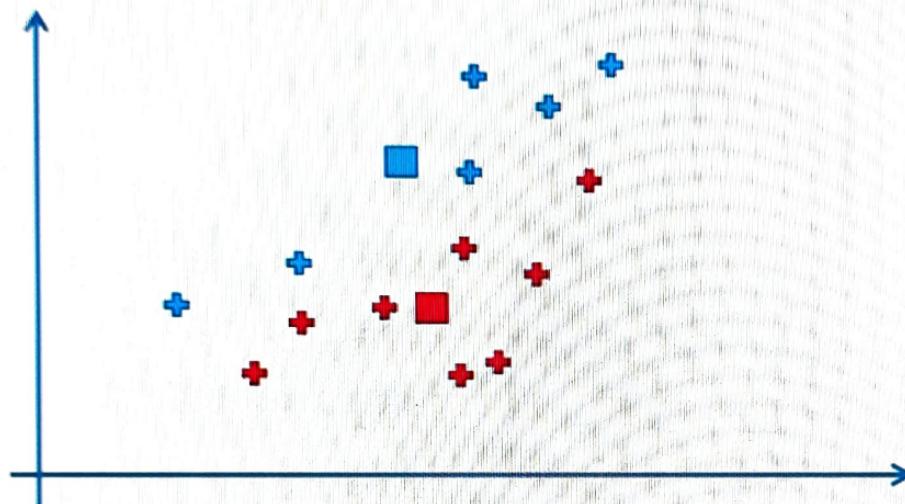
STEP 4: Compute and place the new centroid of each cluster



And we need to find out where the actual centroid the new
centroid for the Blue Points is and the new

K-Means algorithm

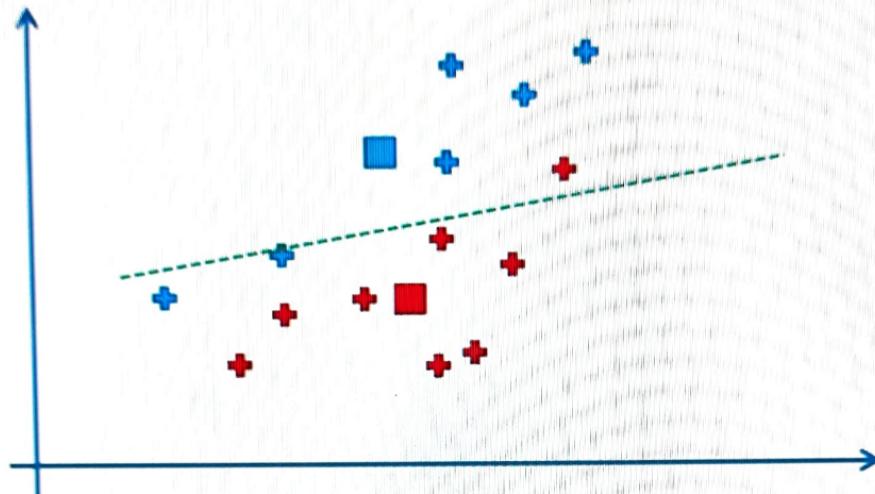
STEP 5: Reassign each data point to the new closest centroid.
If any reassignment took place, go to STEP 4, otherwise go to FIN.



Otherwise you may finish algorithm meaning that it has
converged.

K-Means algorithm

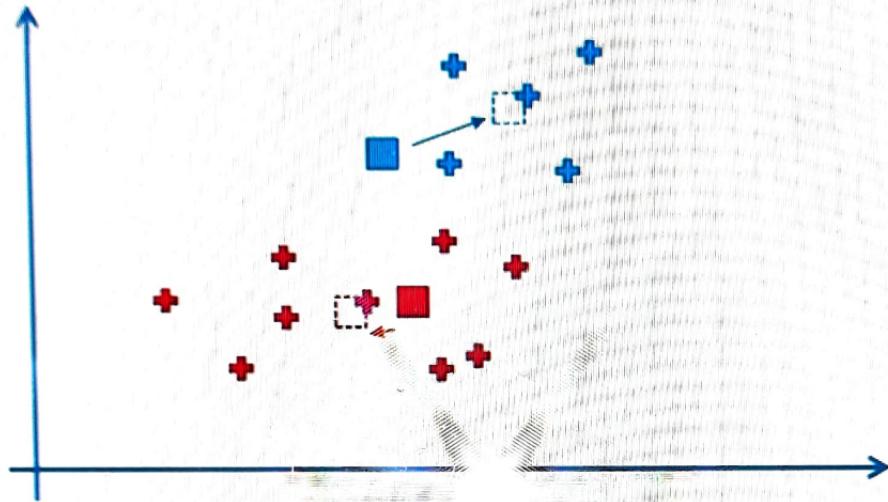
STEP 5: Reassign each data point to the new closest centroid.
If any reassignment took place, go to STEP 4, otherwise go to FIN.



So if we put our line through the scatterplot you'll see that there is
one point on it is actually three

K-Means algorithm

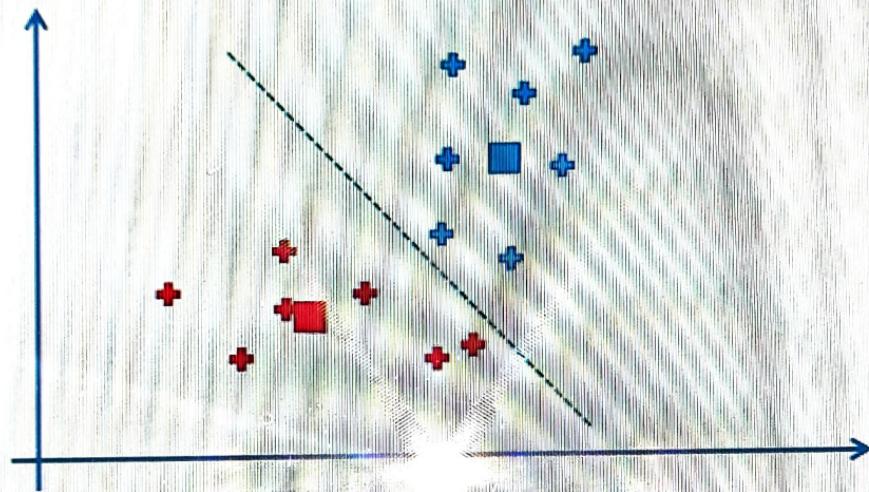
STEP 4: Compute and place the new centroid of each cluster



Find the center mass for this centroid find the center of mass for
this centroid place the centroid

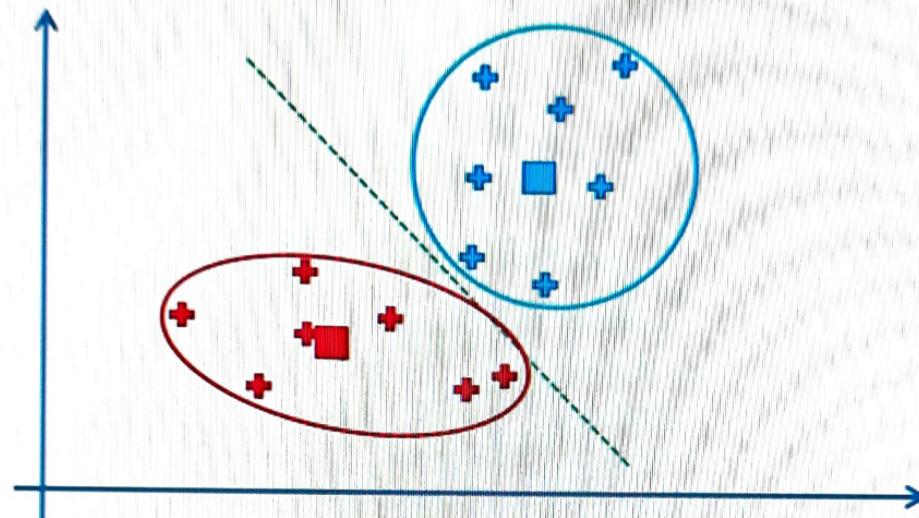
K-Means algorithm

STEP 5: Reassign each data point to the new closest centroid.
If any reassignment took place, go to STEP 4, otherwise go to FIN.



K-Means algorithm

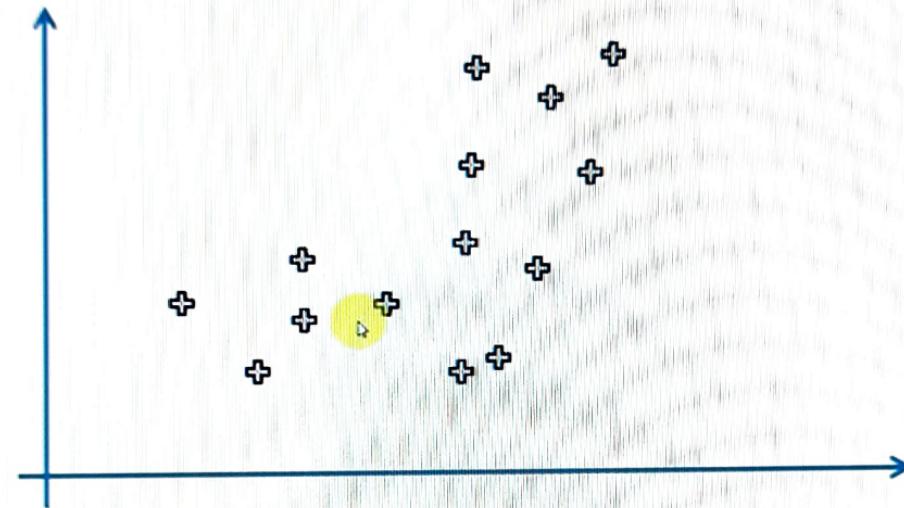
FIN: Your Model Is Ready



Those are our clusters and the model is ready and so now we can just remove the centroid and equidistant

K-Means algorithm

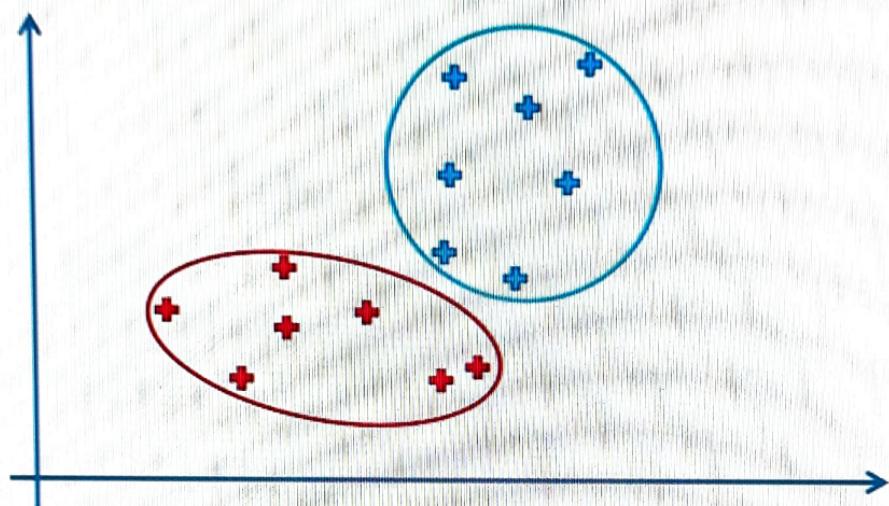
STEP 2: Select at random K points, the centroids (not necessarily from your dataset)



For instance you might have thought that maybe this could be a cluster on its own and this could be

K-Means algorithm

FIN: Your Model Is Ready



are not exactly very similar to each other but this is what the K means algorithm is suggesting hopefully