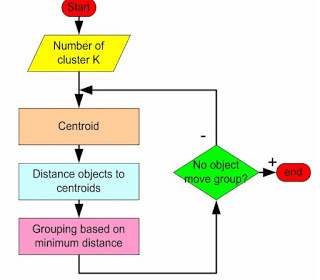
**K- MEANS CLUSTERING**

1. means clustering is a type of unsupervised learning, which is used when you have unlabeled data. The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable *K*. The algorithm works in a iterative manner to assign each data point to one of *K* groups based on the features that are provided.

**Flow chart for k-means clustering:**

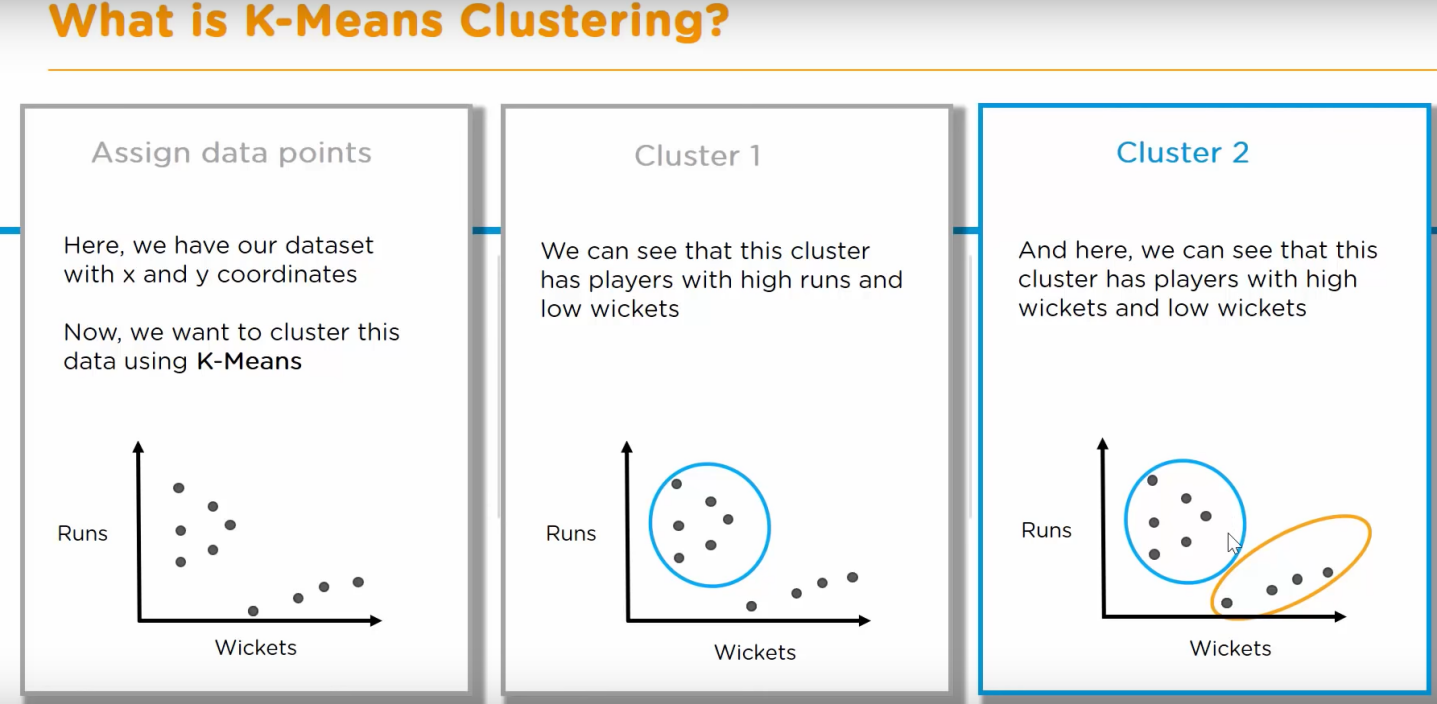


**Algorithm:**

 Assuming we have inputs x1,x2,x3,…xn and value of **K**

 **Step 1** - Pick K random points as cluster centers called centroids.

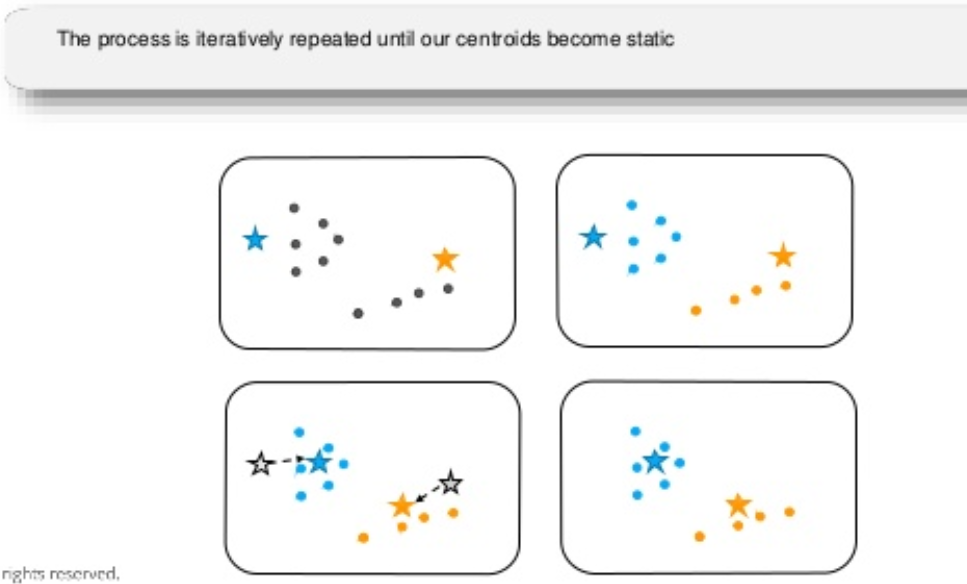
We randomly pick **K** cluster centres (centroids). Let‘s assume these are c1, c2… ck and we can say that, C= c1, c2, ck is the set of all centroids. In the Following Example, k = 2.



 **Step 2** - Assign each xi to nearest cluster by calculating its distance to each centroid.

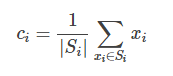
In this step we assign each input value to closest centre.

This is done by calculating Euclidean (L2) distance between the point and the each Centroid.



 **Step 3** - Find new cluster center by taking the average of the assigned points.

In this step, we find the new centroid by taking the average of all the points assigned to that cluster.

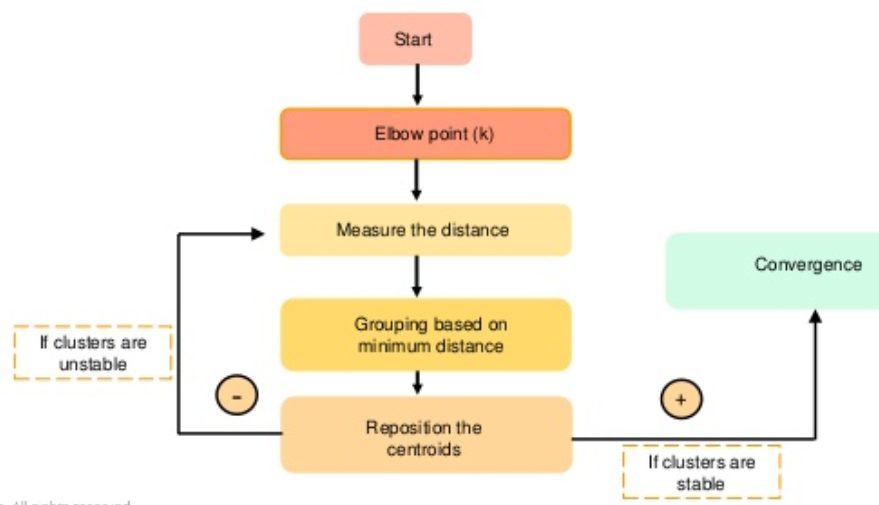


Si is the set of all points assigned to the ith cluster.

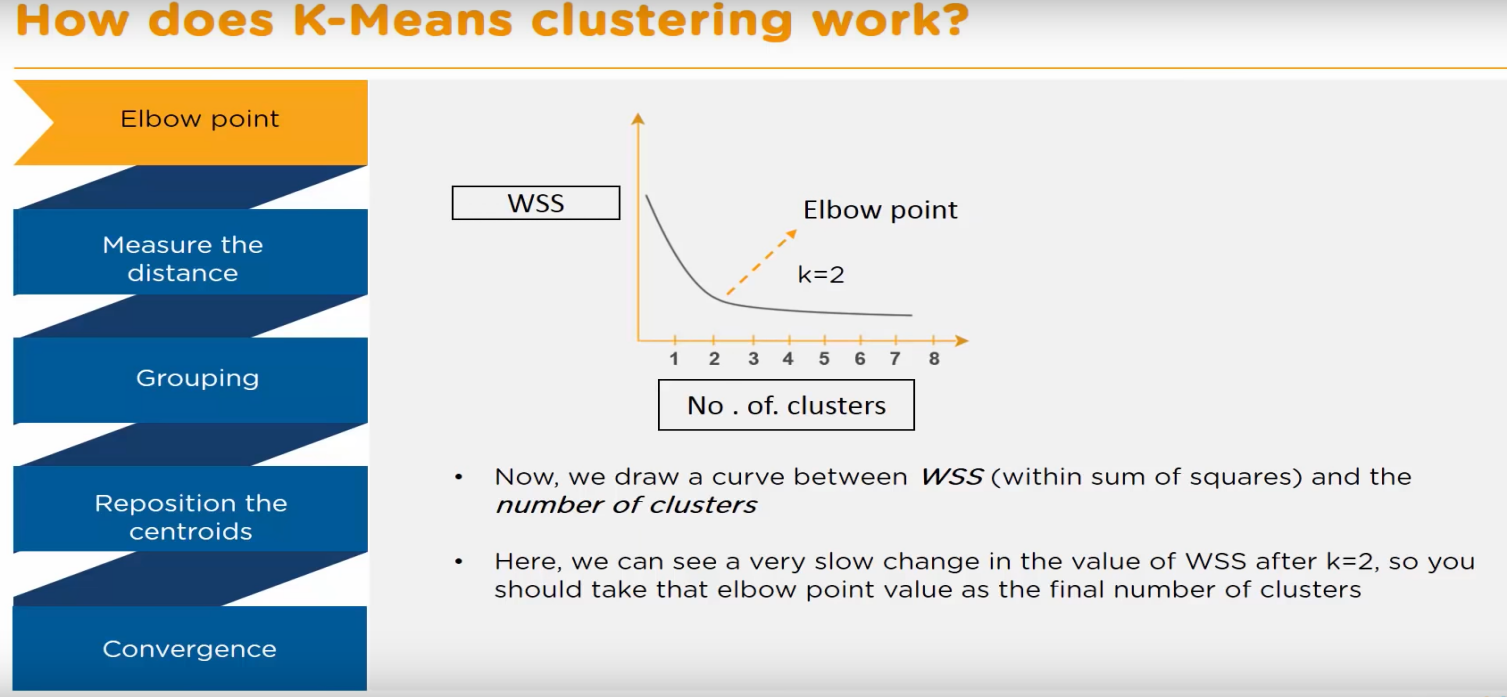
 **Step 4** - Repeat Step 2 and 3 until none of the cluster assignments change.

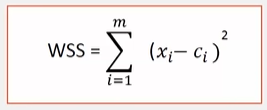
In this step, we repeat step 2 and 3 until none of the cluster assignments change.

That means until our clusters remain stable, we repeat the algorithm.



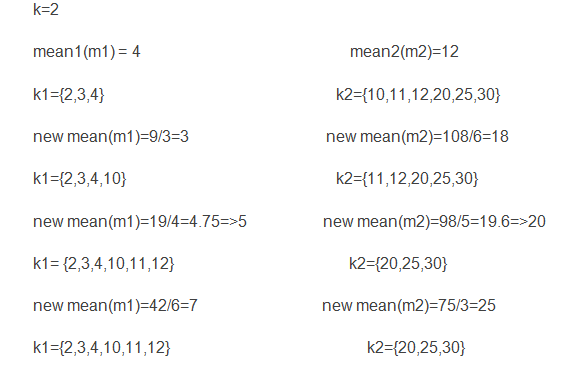
Note :- Elbow Method help us to find the optimal ‘K’.



Where , 

Ex :-

{ 2, 3, 4, 10, 11, 12, 20, 25, 30 }



**Means is widely used for many applications:**

* Image Segmentation
* Clustering Gene Segmentation Data
* News Article Clustering
* Clustering Languages
* Species Clustering