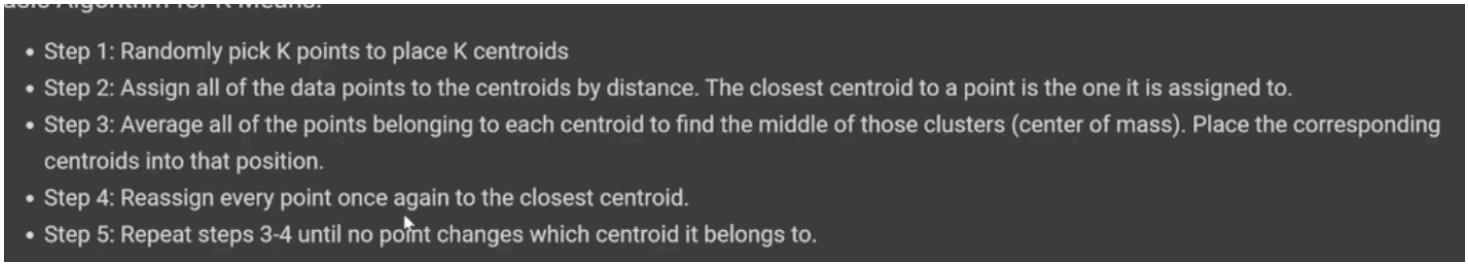
- ***Clustering***
  - -> unsupervised
  - -> this only works for specific problems
    - you have input features, but no labels / output information
    - -> clusters of like datapoints and telling you their location
    - -> you can pick the number of clusters you want
    - -> you give it all of the information of the algorithm generates the clsuters
- ***K Means algorithm***



  - -> the centroid is the centre of the cluster
  - -> you have n data points -> e.g on a graph
  - -> randomly picking k centroids -> the centre of the clusters is randomly chosen
    - -> this is an algorithm
  - -> k is the number of clusters (the number of those points)
  - -> then one of the data points is picked and its distance to each of the centroids is calculated -> it is allocated the number of whichever centroid it's closest to
    - -> this is repeated for all of the points in the dataset
  - -> then the centroid is moved to the centre of mass of all of the datapoints which are labelled as belonging to that centroid
  - -> then this is repeated until the clusters appear and it converges on them
    - -> the centroids are in the centre of the clusters
    - -> clusters
  - -> then when you add more new datapoints to the data set -> it can classify them according to which cluster it belongs in
    - but you have to train the model first
  - -> you need to know the number of clusters which you want -> there are some algorithms which can find it