K-Means++

- In some usecases random initilization is not good as we get bad clusters.
- k-means++ solves that problem by selecting the cluster centroids in initial stage and all the remaining steps are as K-Means.
- Steps:

1. The first cluster is choosen randomly as k-means.
2. next we compute the distance between each point and the choosen centroid.
3. Then we choose the next centroid from the data with the probability of x being proportional to all the distances.
4. we repeat step 2 and 3 until we choose k centroids