4.3) **K-Means Clustering**

Pseudocode:

Repeat for different values of K

       Centroids ← Randomly select K centroids for the dataset

       Previous Centroids ← None

       While Previous Centroids NOT EQUAL TO Centroids AND Number_of_iterations LESS THAN 1000

              Previous Centroids ← Centroids

   Assign all data points a cluster based on the centroid the points are closest to

              Centroids ← Re-Compute the centroids for all the clusters using average values

              Number_of_iterations ← Number_of_iterations + 1

       F(mu, C) ← Compute potential function as a sum of square of distance from the point to its centroid for all data points

Plot the graph of value of K vs value of Potential function

Graph:

Value of Potential function vs Value of K
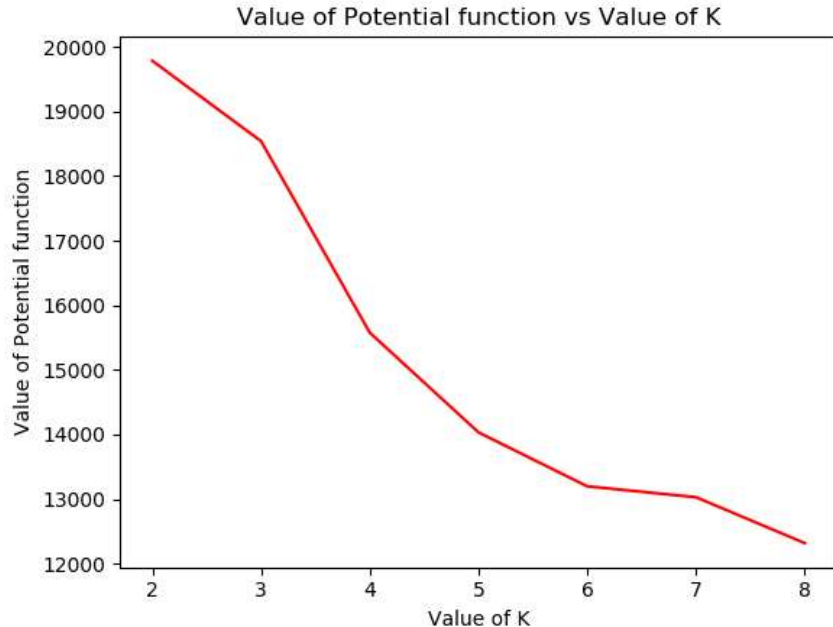
If I were to pick the optimal value of K based on this curve, I would pick the one with the lowest value of the potential function. It is because low value of potential function suggests that points in the clusters are close to their centroids and hence the clusters are tight.