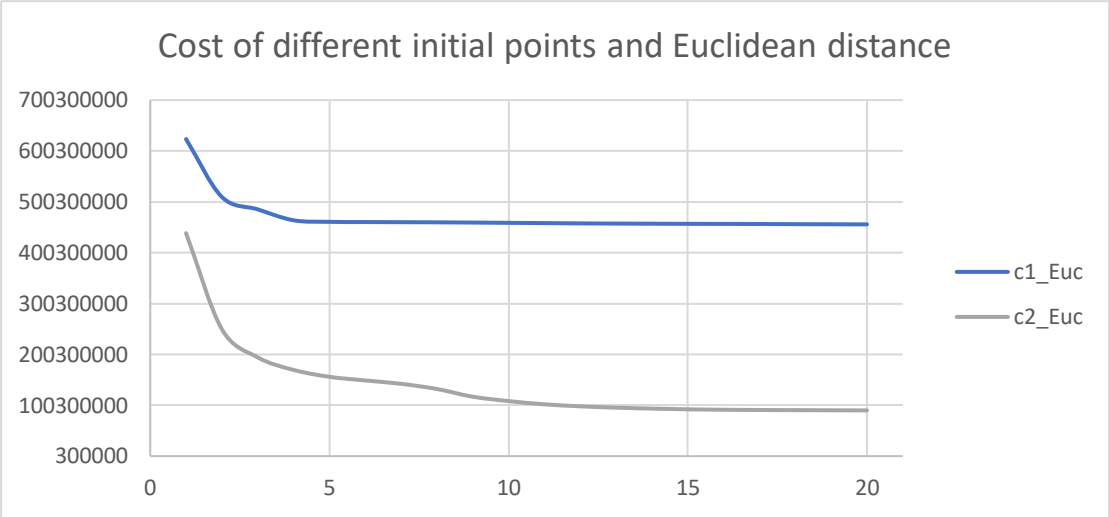
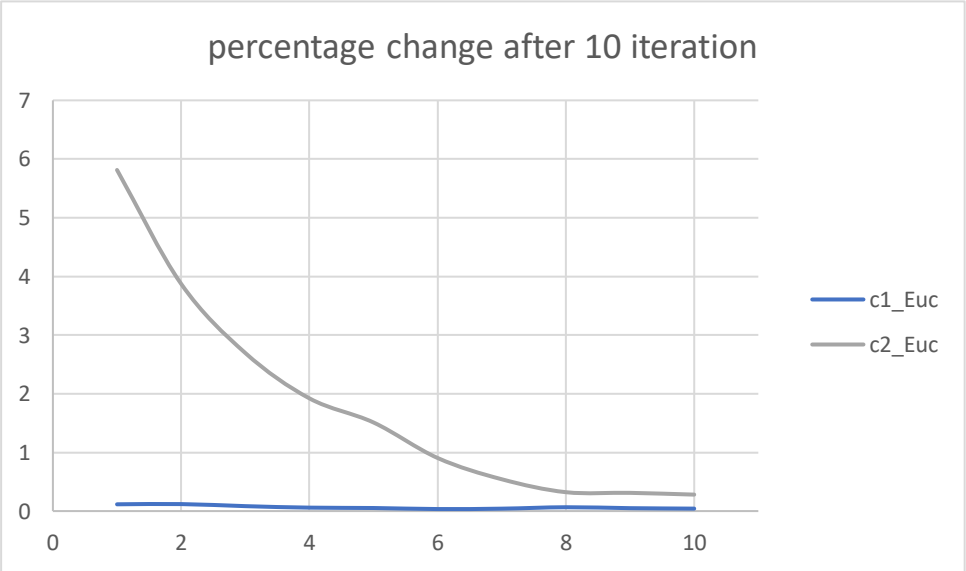
K-Means

高振晏 108061578

(a) A plot of cost vs. iteration for 2 initialization strategies(c1 and c2)



2. percentage improvement

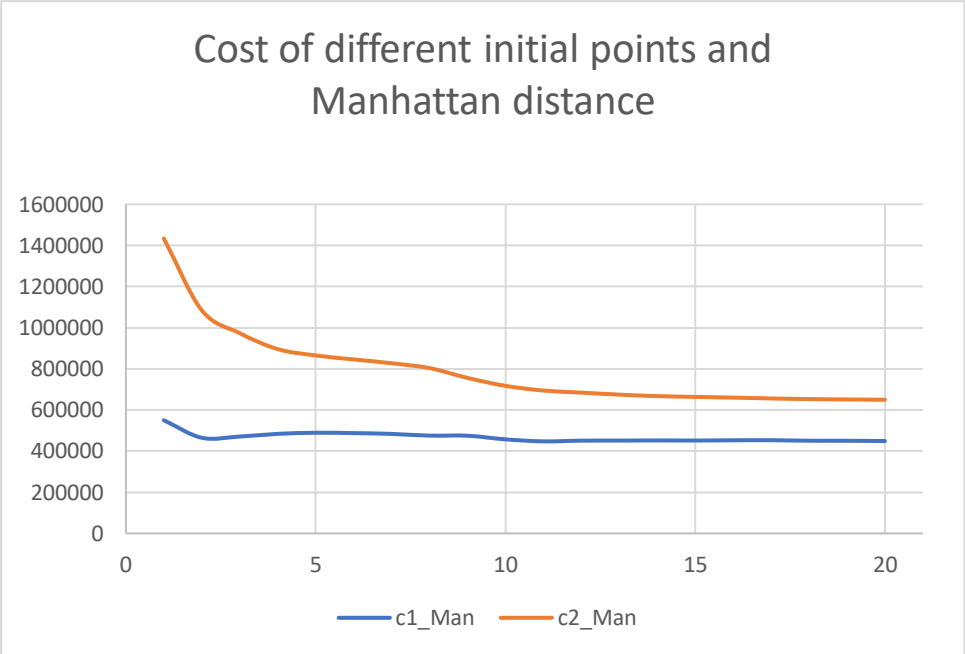


Analysis：we can see that for c2 we have better improvement even after 10 iterations, that is because we select initial points that are as far as possible making that may have clusters with closer size, instead of unbalanced cluster size.

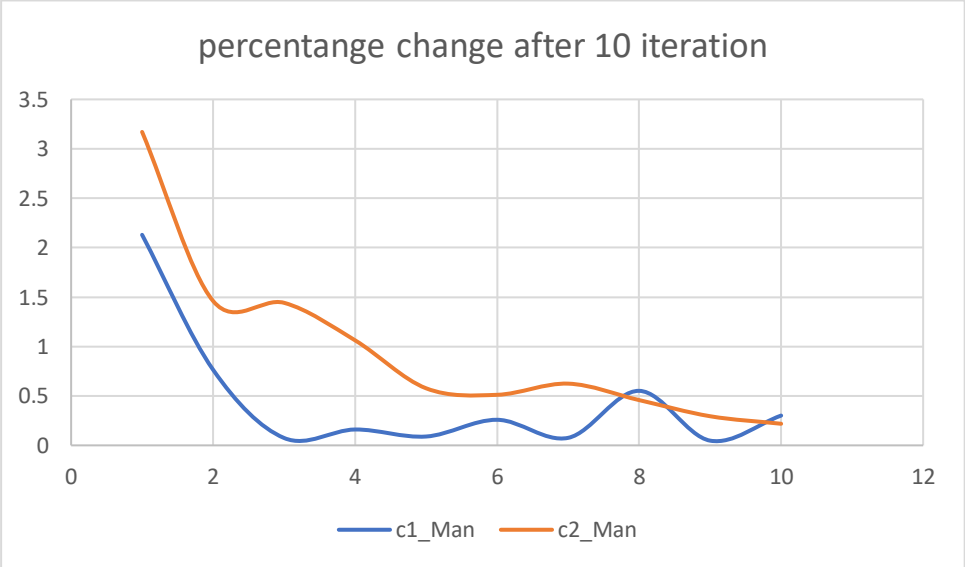
3. The Euclidean and Manhattan Distances for all pairs of centroids

Please look at the attached excel file of c1_Euc and c2_Euc

(b) A plot of cost vs. iteration for 2 initialization strategies(c1 and c2)



2. percentage improvement



Analysis：we can see that both of then still have great improvement after 10 iterations, and since we are using Manhattan distance so we are not punishing some points that are far from the center that much as Euclidean measures, meaning that it can distribute the centroids more evenly.

3. The Euclidean and Manhattan Distances for all pairs of centroids
Please look at the attached excel file of c1_Man and c2_Man