2. Please refer K-means example to calculate 2-cluster K-means for the following subjects

Subject	A	B
1	1.5	1.0
2	1.0	2.0
3	2.0	3.5
4	5.0	6.0
5	3.5	4.0
6	4.5	5.0
7	2.5	4.5

Step 1: Data: the scores of two variables on each of seven individuals.

Subject	A	B
1	1.5	1
2	1	2
3	2	3.5
4	5	6
5	3.5	4
6	4.5	5
7	2.5	4.5

Note:

- Two known information before k-means clustering:
  - The data in matrix format
  - Assuming that the data set is to be grouped into 2 clusters.

Step 2: Initial Partition

Define the initial cluster means:

1. Calculate the centroid.

Subject	A	B	Centroid (A+B)/2
1	1.5	1.0	1.25
2	1.0	2.0	1.50
3	2.0	3.5	2.75
4	5.0	6.0	5.50
5	3.5	4.0	3.75
6	4.5	5.0	4.75
7	2.5	4.5	3.50

2. Find the minimum and maximum centroids.

The minimum centroid is **1.25** marked in Red.
The maximum centroid is **5.50** marked in Blue.

3. Let the A & B values of the two individuals furthest apart (using the Euclidean distance measure), define the initial cluster means.

	Individual	Mean Vector (centroid)
Group 1	1	(1.5, 1.0)
Group 2	4	(5.0, 6.0)

Step 3: First Clustering:

Process:

1. Calculate the distance of each subject and the 2 centroids.
   Example: For Point 1 (1.5,1) the Centroid is 1.25,
   Distance from Centroid 1.25 is 0
   Distance from Centroid 5.50 is (5.50-1.25) = 4.25

   Similarly fill the complete table.

Subject	A	B	Centroid (A+B)/2	Distance from Centroid 1.25	Distance from Centroid 5.50
1	1.5	1.0	1.25	0.00	4.25
2	1.0	2.0	1.50	0.25	4.00
3	2.0	3.5	2.75	1.50	2.75
4	5.0	6.0	5.50	4.25	0.00
5	3.5	4.0	3.75	2.50	1.75
6	4.5	5.0	4.75	3.50	0.75
7	2.5	4.5	3.50	2.25	2.00

2. The remaining individuals are now examined in sequence and allocated to the cluster to which they are closest, in terms of Euclidean distance to the cluster mean.

   Subject 1 is closer to Cluster 1, so it is allocated to Cluster 1. (0.00<4.25)
   Subject 2 is closer to Cluster 1, so it is allocated to Cluster 1. (0.25<4.00)
   Subject 3 is closer to Cluster 1, so it is allocated to Cluster 1. (1.50<2.75)
   Subject 4 is closer to Cluster 2, so it is allocated to Cluster 2. (4.25>0.00)
   Subject 5 is closer to Cluster 2, so it is allocated to Cluster 2. (2.50>1.75)
   Subject 6 is closer to Cluster 2, so it is allocated to Cluster 2. (3.50>0.75)
   Subject 7 is closer to Cluster 2, so it is allocated to Cluster 2. (2.25>2.00)

Subject	A	B	Centroid (A+B)/2	Distance from Centroid 1.25	Distance from Centroid 5.50
1	1.5	1.0	1.25	0.00	4.25
2	1.0	2.0	1.50	0.25	4.00
3	2.0	3.5	2.75	1.50	2.75
4	5.0	6.0	5.50	4.25	0.00
5	3.5	4.0	3.75	2.50	1.75
6	4.5	5.0	4.75	3.50	0.75
7	2.5	4.5	3.50	2.25	2.00

3. The mean vector is recalculated each time a new member is added.

	Cluster 1		Cluster 2	
Step	Individual	Mean Vector (centroid)	Individual	Mean Vector (centroid)
1	1	(1.5, 1.0)	4	(5.0, 6.0)
2	1, 2	(1.25, 1.5)	4	(5.0, 6.0)
3	1, 2, 3	(1.5, 2.17)	4	(5.0, 6.0)
4	1, 2, 3	(1.5, 2.17)	4, 5	(4.25, 5.0)
5	1, 2, 3	(1.5, 2.17)	4, 5, 6	(4.33, 5)
6	1, 2, 3	(1.5, 2.17)	4, 5, 6, 7	(3.875,4.875)

1.5 = (1.5 + 1.0 + 2.0) / 3
2.17 = (1.0 + 2.0 + 3.5) / 3

3.875 = (5.0 + 3.5 + 4.5 + 2.5) / 4
4.875 = (6.0 + 4.0 + 5.0 + 4.5) / 4

Step 4: Check the result of the new clustering:

Now the initial partition has changed, and the two clusters at this stage having the following characteristics:

	Individual	Mean Vector (centroid)
Cluster 1	1, 2, 3	(1.5, 2.17)
Cluster 2	4, 5, 6, 7	(3.875,4.875)

Step 5: Compare each individual's distance to the 2 clusters.

We cannot yet be sure that each individual has been assigned to the right cluster.

So, we compare each individual's distance to its own cluster mean and also to the opposite cluster.

For example,

- The distance between individual 1 and the centroid of Cluster 1 is

  sqrt((1.5 - 1.5)^2 + (2.17 - 1.0)^2) = 1.17

- The distance between individual 1 and the centroid of Cluster 2 is

  sqrt((3.875 - 1.5)^2 + (4.875 - 1.0)^2) = 4.54

Similarly find the distances between Individual and Cluster 1 and Cluster 2.

Individual	Distance to mean (centroid) of Cluster 1	Distance to mean (centroid) of Cluster 2
	(1.5, 2.17)	(3.875,4.875)
1	1.17	4.54
2	0.53	4.07
3	1.42	2.33
4	5.19	1.59
5	2.71	0.95
6	4.12	0.64
7	2.54	1.43

As each individual's distance to its own cluster mean is smaller than the distance to the other cluster's mean, the iteration stops as there are no more relocations choosing the latest partitioning as the final cluster solution.