Question K-Means algorithm

Create a cluster (K=3) on the Student-it Dataset and explain your steps how the algorithm process. (hint: partional clustering approach) (2 Marks)

	sample ID	Height cm	weight cm	Distance to centeroid 1	Distance to centeroid 2	Distance to centeroid 3	cluster	CLUSTER TAG
	1	125	61	48,37			1	
	2	178	90				1	
	3	178	92				1	
	4	180	83				1	
*	5	167	85				1	
	6	170	89				1	
	7	173	98				1	
*	8	135	40				1	
	9	120	35				1	
	10	145	70				1	
	11	125	50				1	
	1	125	61				2	
	2	178	90				2	
	3	178	92				2	
*	4	180	83				2	
	5	167	85				2	
	6	170	89				2	
	7	173	98				2	
	8	135	40				2	
	9	120	35				2	
	10	145	70				2	
	11	125	50				2	

1) Initial centroids = 3 (Random)

Centroid 1 = (167, 85), Centroid 2 = (135, 40), Centroid 3 = (110, 83)

2) Measure distance and assign samples

1) Euclidian Distance $d(x,y)^2 = \sum_{i=1}^{m} (x_i - y_i)^2 = \|x-y\|_2^2$

2) or

- $distance((x,y),(a,b)) = \sqrt{(x-a)^2 + (y-b)^2}$

- $distance((167,85),(125,61)) = \sqrt{(167-125)^2 + (85-61)^2} = 53$

- etc...

3) Re-calculate centroids (clusters)
$\bar{x}$ (Average of assigned clusters)

4) Reassign

5) Measure distance and assign sampler

6) Find centroids