Assignment 5 (Group 3)

Iteration-1

					Distance from C1	Distance from C2	Distance from C3	Allocation to nearest cluster
C1	21	53	56	59				
C2	19	82	82	60				
C3	47	75	76	77				
S1	18	73	75	57	44	23	52	C2
S2	18	79	85	75	74	22	44	C2
S3	23	70	70	52	40	36	60	C2
S4	20	55	55	55	8	60	90	C1
S5	22	85	86	87	91	37	55	C2
S6	19	91	90	89	104	46	87	C2
S7	20	70	65	60	28	30	60	C1
S8	21	53	56	59	0	58	86	C1
S9	19	82	82	60	58	0	41	C2
S10	47	75	76	77	86	58	0	C3

Iteration-2(New Centroids)

					Distance from C1	Distance from C2	Distance from C3	Allocation to nearest cluster
C1	20.33	59.33	58.66	58				
C2	19.83	80	81.33	69.67				
C3	47	75	76	77				
S1	18	73	75	57	33.34	27.83	52	C2
S2	18	79	85	75	65.34	11.83	44	C2
S3	23	70	70	52	30.68	42.17	60	C1
S4	20	55	55	55	11.32	66.17	90	C1
S5	22	85	86	87	83.68	29.17	55	C2
S6	19	91	90	89	95.34	39.83	70	C2
S7	20	70	65	60	19.34	43.5	60	C1
S8	21	53	56	59	10.66	64.17	86	C1
S9	19	82	82	60	49.34	13.17	58	C2
S10	47	75	76	77	78.68	44.83	0	C3

Iteration-3(New Centroids)

					Distance from C1	Distance from C2	Distance from C3	Allocation to nearest cluster
C1	21	62	61.5	56.5				
C2	19.2	82	83.6	73.6				
C3	47	75	76	77				
S1	18	73	75	57	28	35.4	52	C1
S2	18	79	85	75	62	7	44	C2
S3	23	70	70	52	23	51	60	C1
S4	20	55	55	55	16	75	90	C1

S5	22	85	86	87	79	21.6	55	C2
S6	19	91	90	89	92	31	70	C2
S7	20	70	65	60	16	45	60	C1
S8	21	53	56	59	17	73	86	C1
S9	19	82	82	60	46	15.4	58	C2
S10	47	75	76	77	74	45.8	0	C3

Iteration-4(New Centroids)

C1	20.4	64.2	64.2	56.6	Distance from C1	Distance from C2	Distance from C3	Allocation to nearest cluster
C2	19.5	84.3	85.8	77.8				
C3	47	75	76	77				
S1	18	73	75	57	22.4	44.4	52	C1
S2	18	79	85	75	56.4	10.4	44	C2
S3	23	70	70	52	18.8	59.4	60	C1
S4	20	55	55	55	20.4	83.4	90	C1
S5	22	85	86	87	74.6	12.6	55	C2
S6	19	91	90	89	86.4	22.6	70	C2
S7	20	70	65	60	10.4	53.4	60	C1
S8	21	53	56	59	22.4	81.4	86	C1
S9	19	82	82	60	40.4	24.4	58	C2
S10	47	75	76	77	69.6	47.4	0	C3

Results - **Cluster 1 :- S1, S3, S4, S7 & S8, Cluster 2 :- S2, S5, S6 & S9, Cluster 3 :- S10**

Analysis :- Clearly the above results are different from the results obtained via first three objects as seeds - **Cluster 1: S1, S9; Cluster 2: S2, S5, S6, S10; Cluster 3: S3, S4, S7, S8**. Following are the observations.

a. K-means algorithm is sensitive to seed values and that is the reason it produces different set of clusters with different seed object values.

b. It also shows some data objects are very similar which will usually stay together i.e. S2, S5 & S6 are together in one cluster and they did not change for different set of seed values but other data members are not so similar and hence they changed clusters.

c. This algorithm produced better results by taking last three objects as seed values as it identified the outlier - S10 which will reduce distance within clusters but the convergence took 4 iterations compared to 2 iterations used for first three students as seeds.