Data Mining

Assignment # 2

Name: IQRA NAWAZ

Roll No: DT-22005

Batch: 2022

Department: CSIT

Specialization: DATA SCIENCE

Question 1. **Applying Agglomerative Hierarchical Clustering** to cluster the points mentioned below. Euclidean distance for the distance measure between objects and **complete link and Average link** for the distance between clusters. Show clearly all iterations and give **dendrogram** of clustering. Are the dendrogram of complete link and Average link different? Give reason to justify. **[CLO_2, C3]**

A(7,8), B(3,5), C(6,4), D(6,0), E(1,5), F(8,4)

ASSIGNMENT : 02

A(7,8), B(3,5), C(6,4), D(6,0), E(1,5), F(8,4)

COMPLETE LINK :—

	A	B	C	D	E	F
A	0					
B	5	0				
C	4.12	3.16	0			
D	8.06	5.83	4	0		
E	6.70	(2)	5.09	7.07	0	
F	4.12	5.09	2	4.47	7.07	0

Merge B and E :

	A	B, E	C	D	F
A	0				
B, E	6.70	0			
C	4.12	5.09	0		
D	8.06	7.07	4	0	
F	4.12	7.07	(2)	4.47	0

Merge C and F :

	A	B,E	C,F	D
A	0			
B,E	6.70	0		
C,F	4.12	7.07	0	
D	8.06	7.07	4.47	0

Merge A and C,F :

	A,C,F	B,E	D
A,C,F	0		
B,E	7.07	0	
D	8.06	7.07	0

Merge A,C,F and B,E :

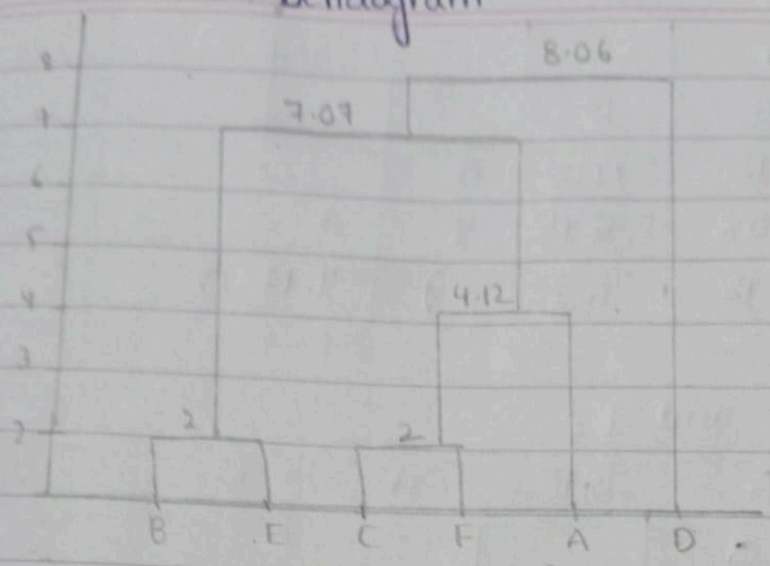
	A,C,F,B,E	D
A,C,F,B,E	0	
D	8.06	0

Merge D and A,C,F,B,E :

Now, all elements are merged so no more merging.

Final. A,B,C,D,E,F

"Dendogram"



AVERAGE LINK:

	A	B	C	D	E	F
A	0					
B	5	0				
C	4.12	3.16	0			
D	8.06	5.83	4	0		
E	6.70	2	5.09	7.07	0	
F	4.12	5.09	2	4.47	7.07	0

Merge B and E :-

	A	B, E	C	D	F
A	0				
B, E	5.85	0			
C	4.12	4.12	0		
D	8.06	6.45	4	0	
F	4.12	6.08	(2)	4.47	0

Merge C and F:

	A	B, E	C, F	D
A	0			
B, E	5.85	0		
C, F	4.12	5.102	0	
D	8.06	6.45	4.23	0

Merge A and C, F:

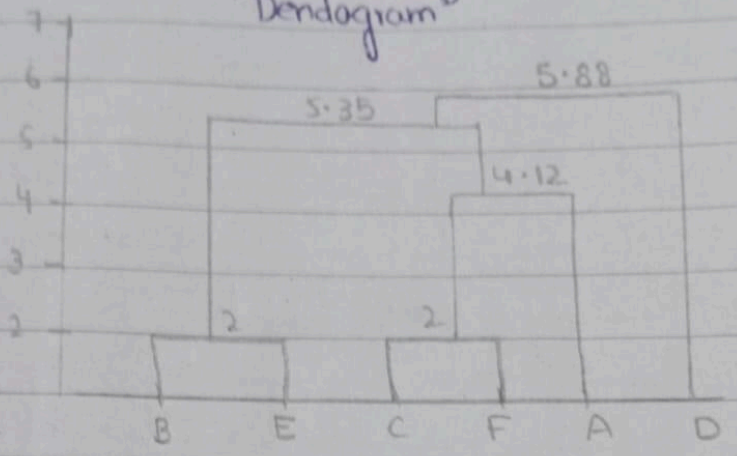
	A, C, F	B, E	D
A, C, F	0		
B, E	(5.35)	0	
D	5.51	6.45	0

Merge A, C, F and B, E :—

	A, C, F, B, E	D
A, C, F, B, E	0	5.88
D	5.88	0

Merge AC, F, B, E and D.

"Dendogram"



Yes, the dendogram of complete link and average link are different.

In complete link the merging happens based on max distances which leads to higher linkages compared of average link, where the average distances may result in earlier or more gradual merging.

Cluster shape and size :— Complete link tends to form compact clusters while average link can accomodate a wide range of shapes and sizes for clusters.