Entities: A(-1,0), B(2,0), C(1,1), D(0,3), E(3,2) and F(3,3)

• Agglomerative Hierarchical Clustering

➢ Single Method

-Distance Matrix

	A	B	C	D	E	F
A	0	3	√5	√10	2√5	5
B		0	√2	√13	√5	√10
C			0	√5	√5	2√2
D				0	√10	3
E					0	1
F						0

As a result, we will merge E and F

-Distance Matrix

	A	B	C	D	EF
A	0	3	√5	√5	2√5
B		0	√2	√13	√5
C			0	√5	√5
D				0	3
EF					0

As a result, we will merge B and C

-Distance Matrix

	A	BC	D	EF
A	0	√5	√10	2√5
BC		0	√5	√5
D			0	3
EF				0

We can merge either A and BC, or D and BC or BC and EF.

So we will merge A and BC

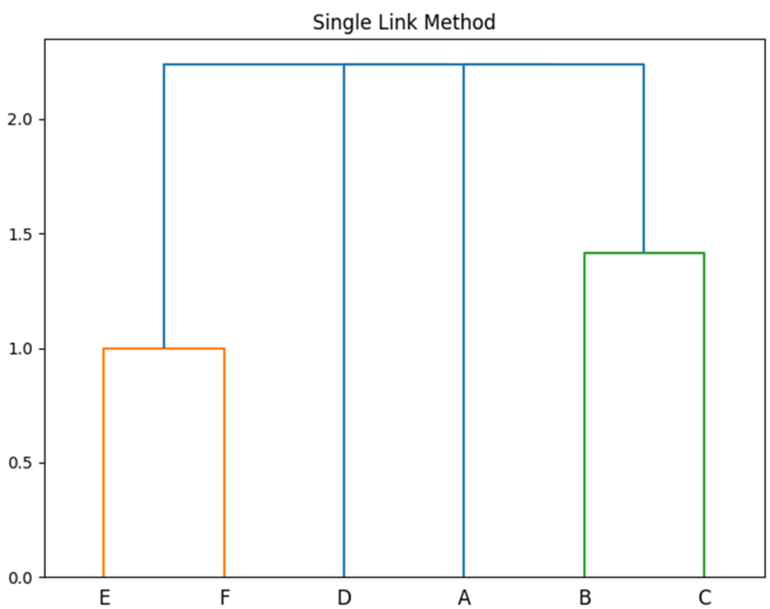
-Distance Matrix

	ABC	D	EF
ABC	0	√5	√5
D		0	3
EF			0

We can merge either ABC and D or ABC and EF

So, we will merge ABC and D

And lastly, we will merge ABCD with EF



➢ Complete Method

-Distance Matrix

	A	B	C	D	E	F
A	0	3	√5	√10	2√5	5
B		0	√2	√13	√5	√10
C			0	√5	√5	2√2
D				0	√10	3
E					0	1
F						0

As a result, we will merge E and F

-Distance Matrix

	A	B	C	D	EF

A	0	3	√5	√10	5
B		0	√2	√13	√10
C			0	√5	2√2
D				0	√10
EF					0

As a result, we will merge B and C

-Distance Matrix

	A	BC	D	EF
A	0	3	√10	5
BC		0	√13	√10
D			0	√10
EF				0

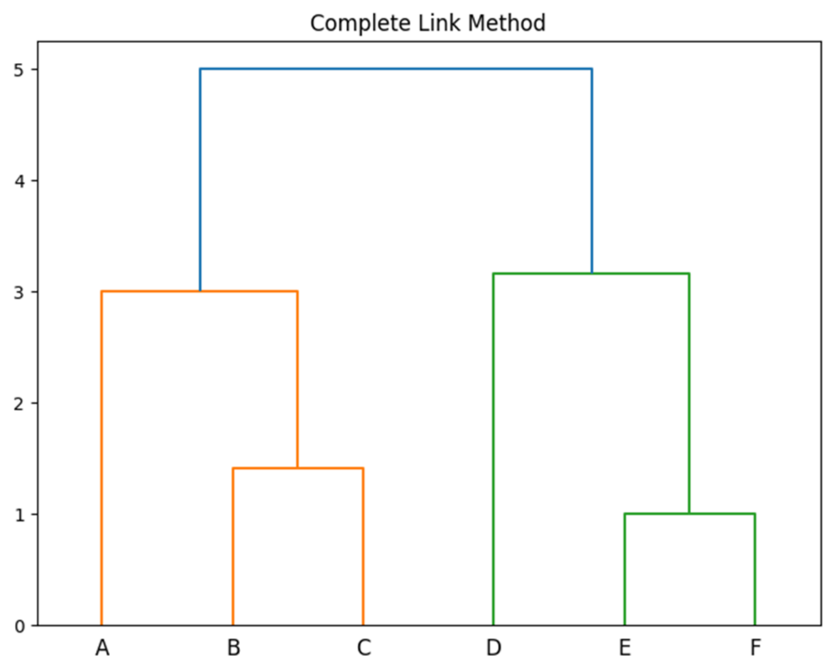
As a result, we will merge A and BC

-Distance Matrix

	ABC	D	EF
ABC	0	√13	5
D		0	√10
EF			0

As a result, we will merge D and EF

And lastly, we will merge ABC and DEF



➢ Average Method

-Distance Matrix

	A	B	C	D	E	F
A	0	3	√5	√10	2√5	5
B		0	√2	√13	√5	√10
C			0	√5	√5	2√2
D				0	√10	3
E					0	1
F						0

As a result, we will merge E and F

-Distance Matrix

	A	B	C	D	EF
A	0	3	√5	√10	4.736
B		0	√2	√13	2.7
C			0	√5	2.53
D				0	3.08
EF					0

As a result, we will merge B and C

-Distance Matrix

	A	BC	D	EF
A	0	2.618	√10	4.736

BC		0	2.92	2.615
D			0	3.08
EF				0

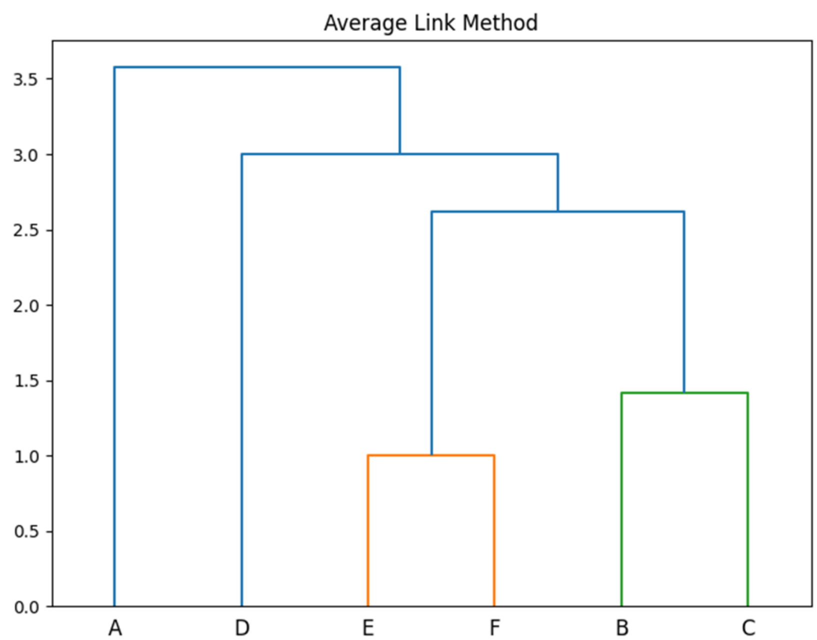
As a result, we will merge BC and EF

-Distance Matrix

	A	BCEF	D
A	0	3.677	√3
BCEF		0	3
D			0

As a result, we will merge BCEF and D

And lastly, we will merge A with BCDEF



Observations:

Although the 1st and 2nd clusters are the same for the three methods we can observe that intriguing differences arise as the clustering process progresses. These distinct patterns in cluster formation underlines the impact of different linkage methods on the hierarchical clustering process. With the single method, we observe that the clusters are tend to be formed by entities that are nearest neighbors to each other. The complete method creates clusters based on the entities that have the largest

distances between them. The average method takes into account the average distance between entities.

• Clustering using K-Means

Initial centroids: C1(0,0) and C2(3,3)

- **1st step: We Compute the distances between the points and the centroids**

- **Distance of the points from C1**

d(A,C1)=1

d(B,C1)=2

d(C,C1)=1,414

d(D,C1)=3

d(E,C1)=3.605

d(F,C1)=4.2416

- **Distance of the points from C2**

d(A,C2)=5

d(B,C2)=3.162

d(C,C2)=2.828

d(D,C2)=3

d(E,C2)=1

d(F,C2)=0

So the clusters will be A, B, C, D and E, F with centroids C1'(0.5,1) and C2'(3,2.5) respectively.

Note that D has equal distance from C1 and C2 so will assign it randomly to the first cluster.

- **2^{nd} step: We compute the distances between the points and the centroids**

- **Distance of the points from C1'**

d(A,C1')=1.802

d(B,C1')=1.802

d(C,C1')=0.5

d(D,C1')=2.06

d(E,C1')=2.692

d(F,C1')=3.201

- **Distance of the points from C2'**

d(A,C2')=4.71

d(B,C2')=2.69

d(C,C2')=2.5

d(D,C2')=3.04

d(E,C2')=0.5

d(F,C2')=0.5

So the clusters will be A, B, C, D and E, F with centroids C1''(0.5,1) and C2''(3,2.5) respectively.

As we observe the clusters do not change and also the centroids are remaining the same.

So the final clusters are A, B, C, D and E, F with centroids C1(0.5,1) and C2(3,2.5) respectively.

Observation-Discussion:

We observe that in both single method of agglomerative hierarchical clustering and the K-Means clustering the entities A, B, C, D were clustered together and the entities E, F were clustered together. The same results are produced in these methods because both methods are based on proximity between the data points. The fact that the final

centroids are positioned close to the final centroids suggests that the K-Means algorithm successfully converged to a solution where the centroids remained in proximity to their initial locations, and it also suggests that the initial centroids provided a good starting point, allowing the algorithm to converge to a relatively stable solution without significant centroid relocations. Given that the K-Means algorithm successfully converged to two distinct clusters and the final centroids are located close to their initial positions, it suggests that the data can be naturally divided into two separate clusters(selection of k=2).