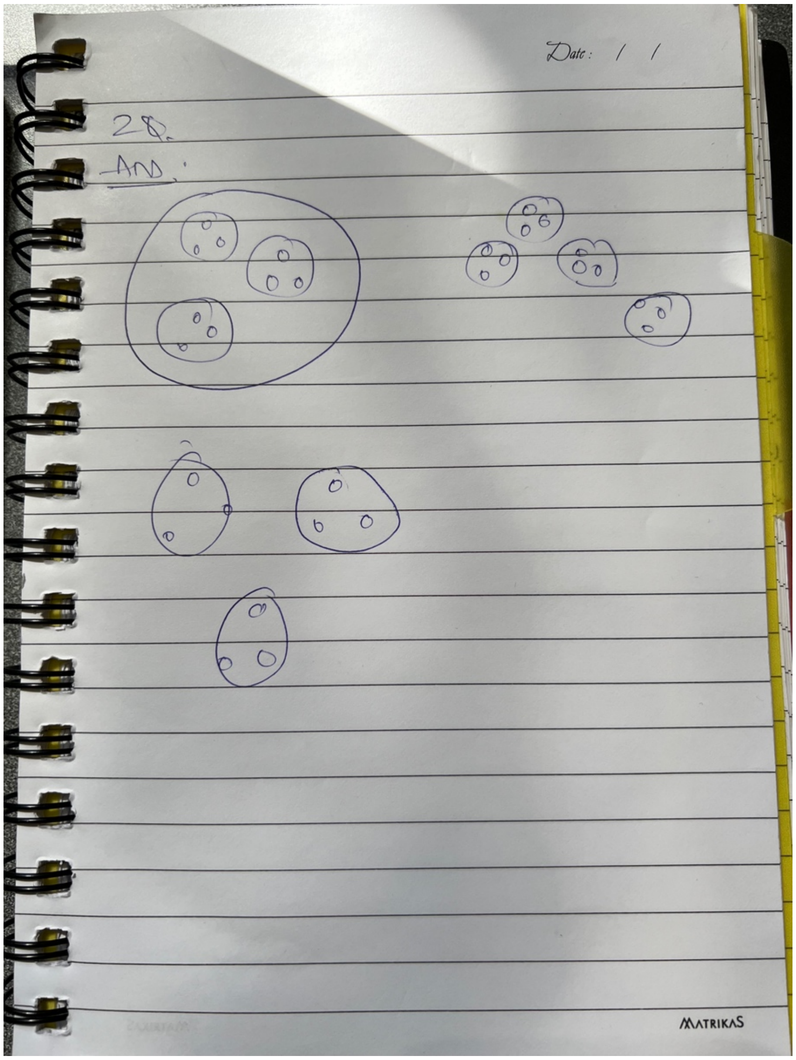
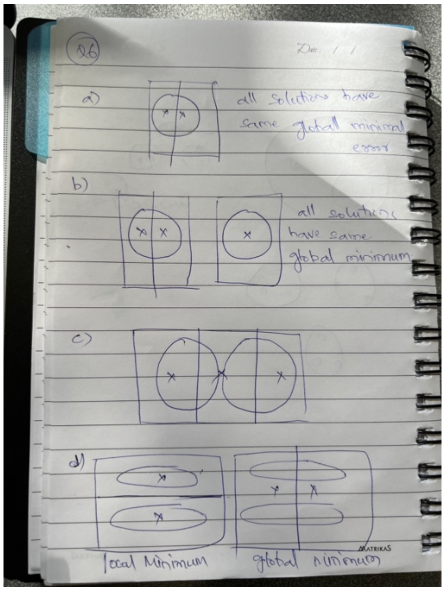
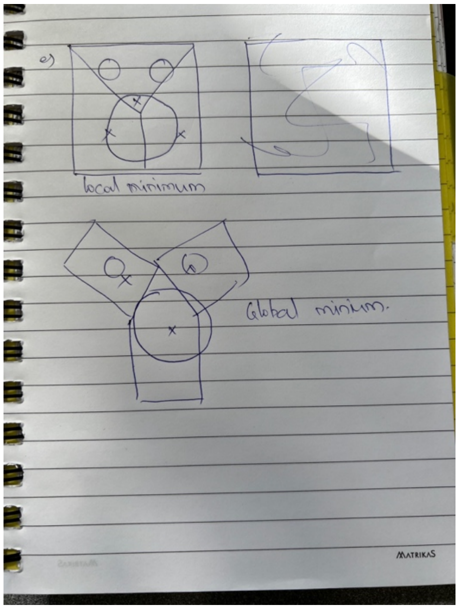
1. Exercises

1.1
Q2)
Ans



Q6)





Q7)
Ans:

(c). Less dense regions require more centroids if the squared error is to be minimized.

11Q)
Ans:

SSE for one variable is low for all clusters: It means the variable is a constant.

Low for just one cluster: If the SSE is low for only one cluster, then that variable helps in defining the cluster.

High for all clusters: It indicates that the variable is noise.

High for just one cluster: It opposes the data given by the low SSE that defined the cluster. The variable doesn't help in defining the cluster

How could you use the per variable SSE information to improve your clustering: We need to eliminate attributes that have poor distinguishing power

12Q)

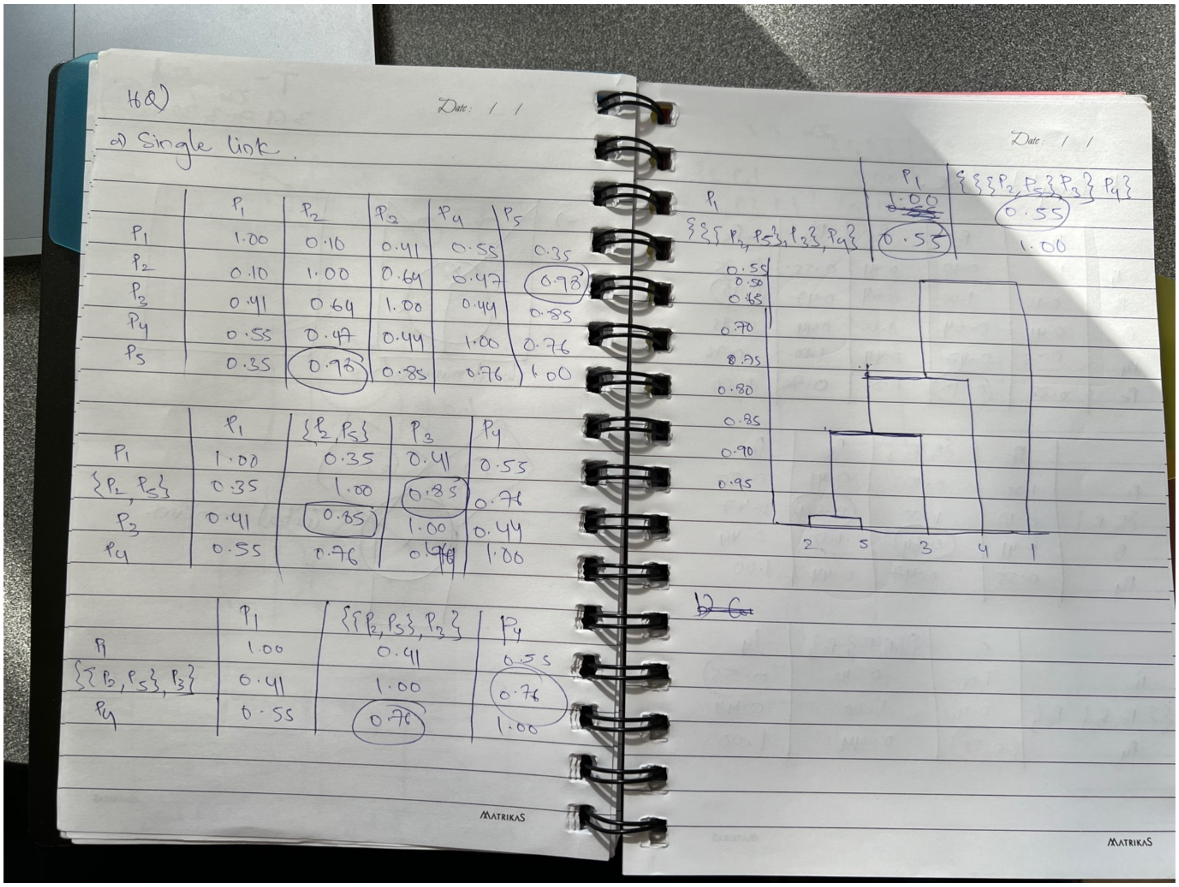
Ans:

a. The leader algorithm is computationally efficient as the data is scanned once. Even though the leader algo is order dependent it will always result in same set of clusters. We can't have a fixed number of resulting clusters directly beforehand like k means.
b. The leader algorithm should be modified to perform clustering for many thresholds.

16Q)

Ans:

a.



16Q)

a) Single link.

	P_1	P_2	P_3	P_4	P_5
P_1	1.00	0.10	0.41	0.55	0.35
P_2	0.10	1.00	0.64	0.47	0.98
P_3	0.41	0.64	1.00	0.44	0.85
P_4	0.55	0.47	0.44	1.00	0.76
P_5	0.35	0.98	0.85	0.76	1.00

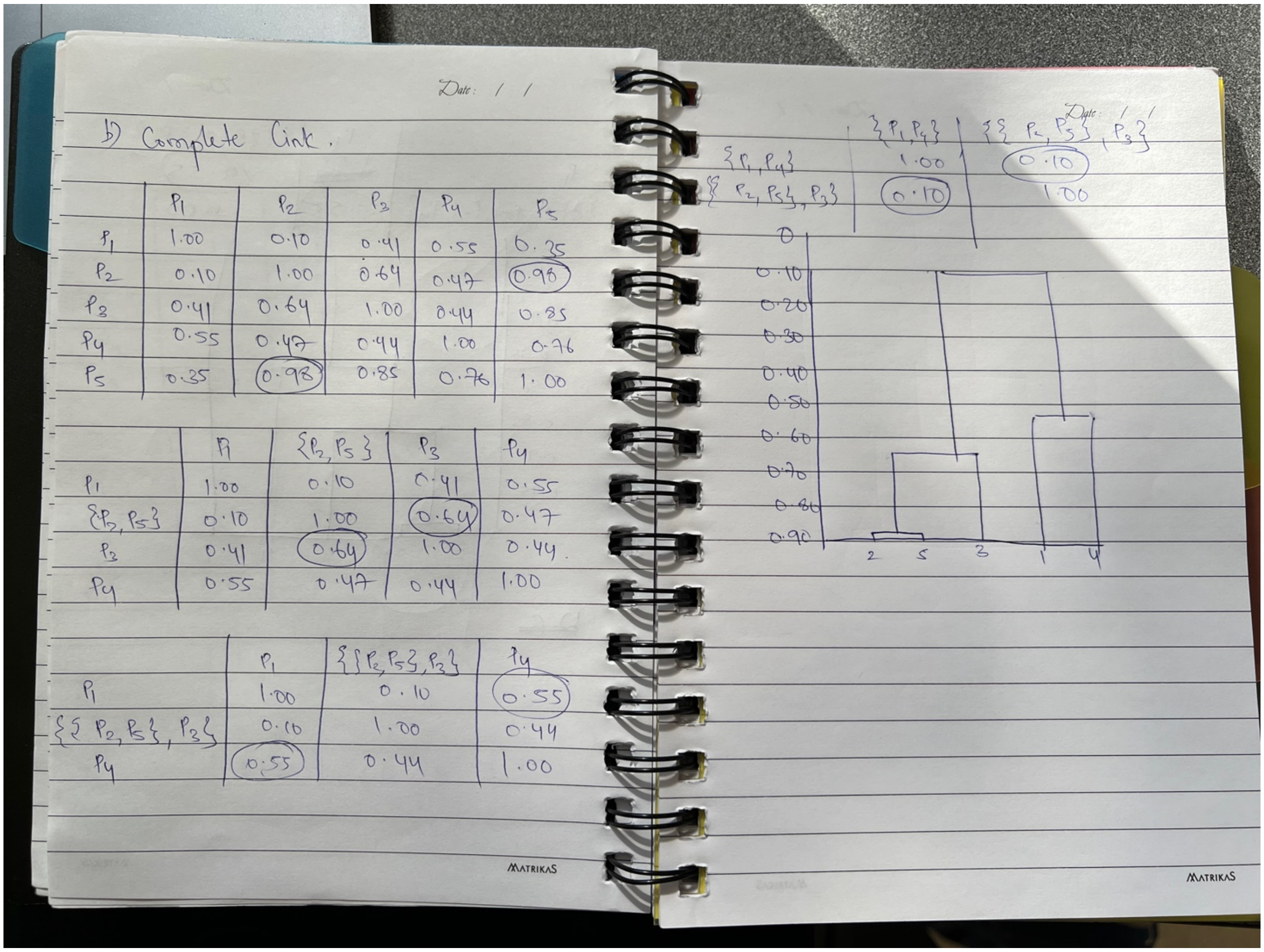
	P_1	$\{P_2, P_5\}$	P_3	P_4
P_1	1.00	0.35	0.41	0.55
$\{P_2, P_5\}$	0.35	1.00	0.85	0.76
P_3	0.41	0.85	1.00	0.44
P_4	0.55	0.76	0.44	1.00

	P_1	$\{\{P_2, P_5\}, P_3\}$	P_4
P_1	1.00	0.41	0.55
$\{\{P_2, P_5\}, P_3\}$	0.41	1.00	0.76
P_4	0.55	0.76	1.00

	P_1	$\{\{\{P_2, P_5\} P_3\} P_4\}$
P_1	1.00 ~~0.55~~	0.55
$\{\{\{P_2, P_5\}, P_3\}, P_4\}$	0.55	1.00

~~b) Co~~

b.



Date: / /

b) Complete Link.

	P1	P2	P3	P4	P5
P1	1.00	0.10	0.41	0.55	0.35
P2	0.10	1.00	0.64	0.47	0.98
P3	0.41	0.64	1.00	0.44	0.85
P4	0.55	0.47	0.44	1.00	0.76
P5	0.35	0.98	0.85	0.76	1.00

	P1	{P2, P5}	P3	P4
P1	1.00	0.10	0.41	0.55
{P2, P5}	0.10	1.00	0.64	0.47
P3	0.41	0.64	1.00	0.44.
P4	0.55	0.47	0.44	1.00

	P1	{{P2, P5}, P3}	P4
P1	1.00	0.10	0.55
{{P2, P5}, P3}	0.10	1.00	0.44
P4	0.55	0.44	1.00

Date: / /

	{P1, P4}	{{P2, P5}, P3}
{P1, P4}	1.00	0.10
{{P2, P5}, P3}	0.10	1.00

0
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
2
5
3
1
4