

EXERCISE 3 Answer Key

Dataset (Customer, Avg Bet, Visits/month):

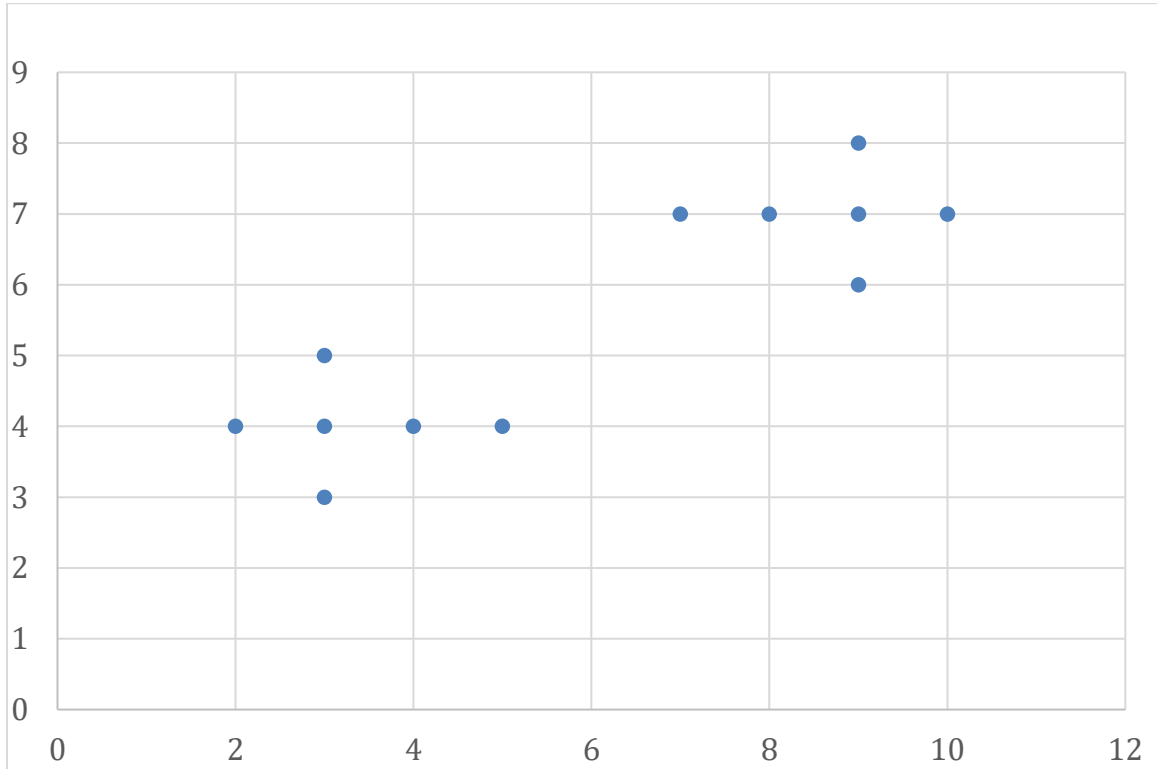
Customer	Avg Bet per Visit (USD)	Visits per Month
A	3	4
B	2	4
C	3	3
D	3	5
E	9	7
F	8	7
G	9	6
H	9	8
I	4	4
J	10	7
T	5	4
U	7	7

DBSCAN Hyperparameters:

eps = 1

min_samples = 2

1. Create a scatter plot



2. Compute all pairwise Euclidean distances between customers.
3. Create a full distance matrix.

	A (3,4)	B (2,4)	C (3,3)	D (3,5)	E (9,7)	F (8,7)	G (9,6)	H (9,8)	I (4,4)	J (10,7)	T (5,4)	U (7,7)
A(3,4)	0	1	1	1	6.7	5.8	6.3	7.2	1	7.6	2	5
B(2,4)	1	0	1.4	1.4	7.6	6.7	7.2	8	2	8.5	3	5.8
C(3,3)	1	1.4	0	2	7.2	6.4	6.7	7.8	1.4	8	2.2	5.6
D(3,5)	1	1.4	2	0	6.3	5.3	6	6.7	1.4	7.2	2.2	4.4
E(9,7)	6.7	7.6	7.2	6.3	0	1	1	1	5.8	1	5	2
F(8,7)	5.8	6.7	6.4	5.3	1	0	1.4	1.4	5	2	4.2	1
G(9,6)	6.3	7.2	6.7	6	1	1.4	0	2	5.3	1.4	4.4	2.2
H(9,8)	7.2	8	7.8	6.7	1	1.4	2	0	6.4	1.4	5.6	2.2
I(4,4)	1	2	1.4	1.4	5.8	5	5.3	6.4	0	6.7	1	4.2
J(10,7)	7.6	8.5	8	7.2	1	2	1.4	1.4	6.7	0	5.8	3
T(5,4)	2	3	2.2	2.2	5	4.2	4.4	5.6	1	5.8	0	3.6
U(7,7)	5	5.8	5.6	4.4	2	1	2.2	2.2	4.2	3	3.6	0

4. For each point A-U, list all of its neighbours based on **eps**.
 5. Create a table showing each point and the **number of neighbors** in its neighborhood.
 6. For each point A-U, determine if it is a core point, non-core point and noise based on **min_samples**.
 7. Simulate DBSCAN
 - a. Select the a core point and begin a new cluster
 - b. Expand the cluster using its neighbours
 - c. Add newly discovered core points to the expansion queue.
 - d. Continue until no more points can be reached
 - e. Repeat until all points are assigned or labeled noise.
 8. Create a table showing each point and its cluster.
- Starting from core point A, start green cluster and join all its neighbors B,C,D,I
 - B,C,D are non-core points. The green cluster cannot extend further
 - E is an core point, start blue cluster and join its neighbors F,G,H,J in the blue cluster
 - F is a core point and is in the blue cluster, extend blue cluster to its neighbor U
 - G,H are non core points. Blue cluster cannot extend further.
 - I is a core point in the green cluster. We extend the green cluster to its neighbor T
 - J,T,U are non core points. We cannot extend their clusters further.

Customer	<i>Neighbours</i>	<i>Number of neighbours</i>	Point Type	Cluster
A	B,C,D,I	4	Core	GREEN
B	A	1	Non-core	GREEN
C	A	1	Non-core	GREEN
D	A	1	Non-core	GREEN
E	F,G,H,J	4	Core	BLUE
F	E,U	1	Core	BLUE
G	E	1	Non-core	BLUE

H	E	1	Non-core	BLUE
I	A,T	2	Core	GREEN
J	E	1	Non-core	BLUE
T	I	1	Non-core	GREEN
U	F	1	Non-core	BLUE