

LEON MARCO A. DEVELA

COM232

GIVEN:

ADVANCED MACHINE LEARNING
EXERCISE 2

CUSTOMER	Avg Bet per Visit (USD)	Visits per Month
A	10	1
B	20	2
C	55	6
D	70	8
E	60	10
F	35	4

① Compute all pairwise Euclidean distances between customers

$$d(A, B) = \sqrt{(10-20)^2 + (1-2)^2}$$

$$d(A, B) = 10.05$$

$$d(A, C) = \sqrt{(10-55)^2 + (1-6)^2}$$

$$d(A, C) = 45.28$$

$$d(A, D) = \sqrt{(10-70)^2 + (1-8)^2}$$

$$d(A, D) = 60.41$$

$$d(A, E) = \sqrt{(10-60)^2 + (1-10)^2}$$

$$d(A, E) = 50.80$$

$$d(A, F) = \sqrt{(10-35)^2 + (1-4)^2}$$

$$d(A, F) = 25.18$$

$$d(B, C) = \sqrt{(20-55)^2 + (2-6)^2}$$

$$d(B, C) = 35.23$$

$$d(B, D) = \sqrt{(20-70)^2 + (2-8)^2}$$

$$d(B, D) = 50.36$$

$$d(B, E) = \sqrt{(20-60)^2 + (2-10)^2}$$

$$d(B, E) = 40.79$$

$$d(B, F) = \sqrt{(20-35)^2 + (2-4)^2}$$

$$d(B, F) = 15.13$$

$$d(D, E) = \sqrt{(70-60)^2 + (8-10)^2}$$

$$d(D, E) = 10.20$$

$$d(D, F) = \sqrt{(70-35)^2 + (8-4)^2}$$

$$d(D, F) = 35.23$$

$$d(E, F) = \sqrt{(60-35)^2 + (10-4)^2}$$

$$d(E, F) = 25.71$$

② Create a matrix that will contain all the distances

	A	B	C	D	E	F
A	0	10.05	45.28	60.41	50.80	25.18
B	10.05	0	35.23	50.36	40.79	15.13
C	45.28	35.23	0	15.13	6.40	20.10
D	60.41	50.36	15.13	0	10.20	35.23
E	50.80	40.79	6.40	10.20	0	25.71
F	25.18	15.13	20.10	35.23	25.71	0

③ Using single linkage, determine the merging order step by step

C - E

A - B

AB - F

CE - D

ABF - CED

④ Record the distance at which each merge occurs

C - E = 6.40

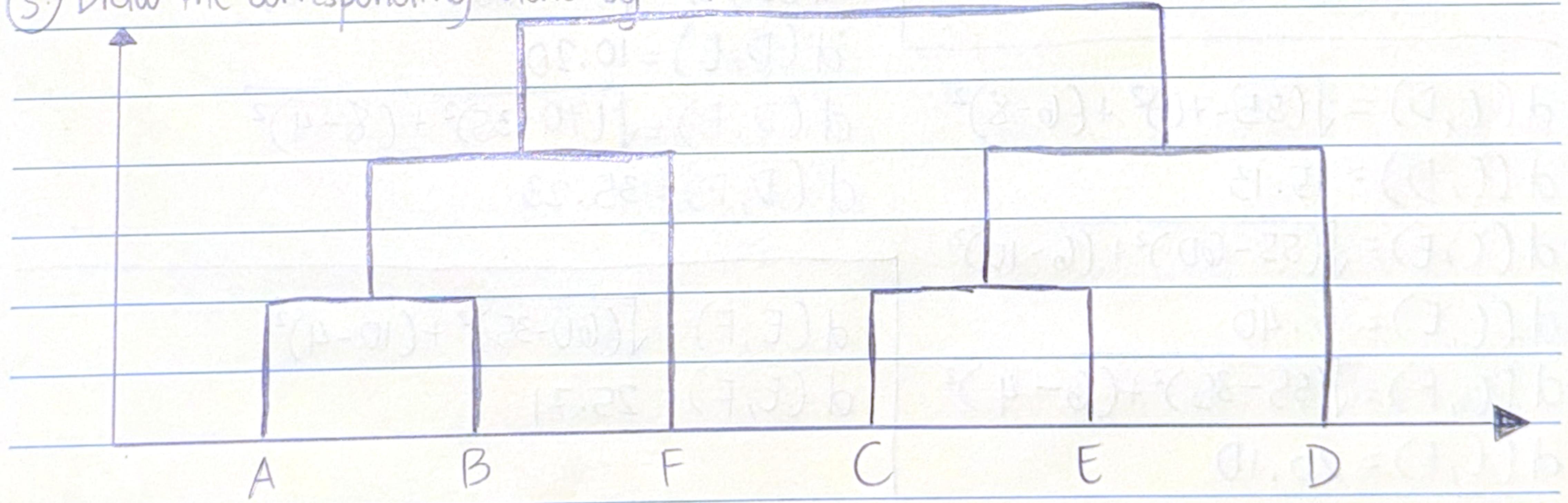
A - B = 10.05

AB - F = 15.13

CE - D = 15.13

ABF - CED = 20.10

⑤ Draw the corresponding dendrogram



⑥ Interpret the clusters. What customer behavior does each cluster represent?

The clusters ABF and CED makes the most sense, meaning by dividing the dataset into two clusters, we can interpret some insights. The first cluster, ABF, are the casual casino goer because they visit less and spend less. The CED cluster are the addicted one because they visit the casino more often and bet higher.