Day – 4 Programs

1 Input:

@relation dataset

@attribute customerid{1,2,3,4,5}

@attribute gender{male,female}

@attribute age{19,21,20,23,31}

@attribute annualincome{15,16,17}

@attribute spendingscore{39,81,6,77,40}

@data

1 male 19 15 39

2 male 21 15 81

3 female 20 16 6

4 female 23 16 77

5 female 31 17 40

2 Input:

@relation dataset

@attribute id{1,2,3,4,5,6}

@attribute gender{male,female}

@attribute age{28,25,26,30,29}

@attribute salary{15,16,17,20}

@attribute credit{39,27,42,40,64,72}

@data

1 male 28 15 39

2 male 25 15 27

3 female 26 16 42

4 female 25 16 40

5 female 30 17 64

6 male 29 20 72

4 Input:

@relation dataset

@attribute person{gopu,babu,baby,gopal,krishna,jai,dev,malini,hema,anu}

@attribute vegeterian{yes,no}

@data

gopu yes

babu yes

baby yes

gopal no

krishna yes

jai no

dev no

malini yes

hema yes

anu yes

5 Input:

@relation dataset

@attribute x{4,1,5,7,10,2,50,25,90,36}

@attribute y{12,5,13,19,31,7,153,72,275,110}

@data

4,12

1,5

5,13

7,19

10,31

2,7

50,153

25,72

90,275

36,110

R programming:

x\_values <- c(4, 1, 5, 7, 10, 2, 50, 25, 90, 36)

y\_values <- c(12, 5, 13, 19, 31, 7, 153, 72, 275, 110)plot(x\_values, y\_values, main="Scatter Plot of Mobile Phones Sold",

xlab="Number of Mobile Phones Sold", ylab="Money",

pch=16, col="blue")

grid()

6 Input:

@relation dataset

@attribute transid{1,2,3,4,5}

@attribute bread{true,false}

@attribute cheese{true,false}

@attribute egg{true,false}

@attribute juice{true,false}

@attribute milk{true,false}

@attribute yogurt{true,false}

@data

1,true,true,true,true,false,false

2,true,true,false,true,false,false

3,true,false,false,false,true,true

4,true,false,false,true,true,false

5,false,true,false,true,true,false

8 Input:

marks <- c(55, 60, 71, 63, 55, 65, 50, 55, 58, 59, 61, 63, 65, 67, 71, 72, 75)

bins\_a <- cut(marks, breaks = 3, labels = c("Low", "Medium", "High"))

bins\_b <- cut(marks, breaks = seq(min(marks), max(marks), length.out = 4), labels = c("Low", "Medium", "High"))

k <- 3

clusters <- kmeans(matrix(marks), centers = k)

bins\_c <- cut(clusters$centers[clusters$cluster], breaks = 3, labels = c("Low", "Medium", "High"))

par(mfrow = c(1, 3))

hist(marks, main = "Equal-frequency (equi-depth) partitioning", col = "skyblue", breaks = 3)

hist(marks, main = "Equal-width partitioning", col = "lightgreen", breaks = seq(min(marks), max(marks), length.out = 4))

hist(marks, main = "Clustering", col = "lightpink", breaks = 3)

10 Input:

@relation dataset

@attribute id{1,2,3,4,5,6,7,8,9}

@attribute sony{true,false}

@attribute bpl{true,false}

@attribute lg{true,false}

@attribute samsung{true,false}

@attribute onida{true,false}

@data

1,true,true,true,false,false

2,false,true,false,true,false

3,false,true,false,false,true

4,true,true,false,true,false

5,true,false,false,false,true

6,false,true,false,false,true

7,true,false,false,false,true

8,true,true,true,false,true

9,true,true,false,false,true

11 Input:

strike\_rates <- c(100, 70, 60, 90, 90)

min\_max\_normalization <- function(x) {

(x - min(x)) / (max(x) - min(x))

}

normalized\_min\_max <- min\_max\_normalization(strike\_rates)

z\_score\_normalization <- function(x) {

(x - mean(x)) / sd(x)

}

normalized\_z\_score <- z\_score\_normalization(strike\_rates)

mad\_normalization <- function(x) {

(x - mean(x)) / mad(x)

}

normalized\_mad <- mad\_normalization(strike\_rates)

decimal\_scaling\_normalization <- function(x) {

x / 10^(ceiling(log10(max(x))))

}

normalized\_decimal\_scaling <- decimal\_scaling\_normalization(strike\_rates)

cat("Original Data:", strike\_rates, "\n\n")

cat("(a) Min-Max Normalization:", normalized\_min\_max, "\n")

cat("(b) Z-Score Normalization:", normalized\_z\_score, "\n")

cat("(c) Z-Score Normalization (MAD):", normalized\_mad, "\n")

cat("(d) Normalization by Decimal Scaling:", normalized\_decimal\_scaling, "\n")

output:

> cat("(a) Min-Max Normalization:", normalized\_min\_max, "\n")

(a) Min-Max Normalization: 1 0.25 0 0.75 0.75

> cat("(b) Z-Score Normalization:", normalized\_z\_score, "\n")

(b) Z-Score Normalization: 1.095445 -0.7302967 -1.338877 0.4868645 0.4868645

> cat("(c) Z-Score Normalization (MAD):", normalized\_mad, "\n")

(c) Z-Score Normalization (MAD): 1.214083 -0.8093889 -1.48388 0.5395926 0.5395926

> cat("(d) Normalization by Decimal Scaling:", normalized\_decimal\_scaling, "\n")

(d) Normalization by Decimal Scaling: 1 0.7 0.6 0.9 0.9

12 Input:

avg\_speed <- c(78, 81, 82, 74, 83, 82, 77, 80, 70)

total\_time <- c(39, 37, 36, 42, 35, 36, 40, 38, 46)

sd\_avg\_speed <- sd(avg\_speed)

sd\_total\_time <- sd(total\_time)

var\_avg\_speed <- var(avg\_speed)

var\_total\_time <- var(total\_time)

cat("Standard Deviation of AvgSpeed:", sd\_avg\_speed, "\n")

cat("Standard Deviation of TotalTime:", sd\_total\_time, "\n\n")

cat("Variance of AvgSpeed:", var\_avg\_speed, "\n")

cat("Variance of TotalTime:", var\_total\_time, "\n")

13 Input:

@relation dataset

@attribute M{true,false}

@attribute O{true,false}

@attribute N{true,false}

@attribute K{true,false}

@attribute E{true,false}

@attribute Y{true,false}

@attribute D{true,false}

@attribute A{true,false}

@attribute U{true,false}

@attribute C{true,false}

@attribute I{true,false}

@data

true true true true true true false false false false false false

false true true true true true true false false false false false

true false false true true false false false true false false false

true false false true false true false false false true true false

false true false true true false false false false false true true