**R LAB – 06**

**Task – 01 :**

**Aim :** using the apply , aggregate , data.table functionalites manipulate the any one random data..

**Program :**

# apply function

theMatrix = matrix(1:9, nrow = 3)

theMatrix

# 1 means sum of rows and 2 means sum of columns

apply(theMatrix, 1, sum)

apply(theMatrix, 2, sum)

theMatrix[2,3] <- NA

apply(theMatrix, 1, sum)

apply(theMatrix, 1, sum, na.rm = TRUE)

rowSums(theMatrix, na.rm = TRUE)

colSums(theMatrix, na.rm = TRUE)

# lapply & sapply functions

theList <- list(A = matrix(1:9,3), B = 1:5, C = matrix(1:4,2), D = 2)

theList

lapply(theList, sum) # returns list

sapply(theList, sum) # returns data frame

theNames <- c("Tyson", "Rizwanullah", "Dragon Emperor", "A.Charan")

lapply(theNames, nchar)

sapply(theNames, nchar)

# mapply function

f3 <- function(x,y)

{

NROW(x) + NCOL(y)

}

flist = list(A = matrix(1:16,4), B = matrix(1:16,2), C = 1:5)

flist

slist = list(A = matrix(1:16,4), B = matrix(1:16,2), C = 15:1)

slist

mapply(identical, flist, slist)

mapply(f3, flist, slist)

# using aggregate function

eid = c(2501 : 2509)

eid

ename = c("Rizwan", "Ajay", "Mounav", "Charan", "Srinivas", "Vamsi", "Deepak", "Abhi", "Pavan")

ename

desig = c("sales", "accounts", "manager", "sales", "sales", "accounts", "accounts", "manager", "sales")

desig

dept\_id = c(10,10,10,10,20,20,20,30,30)

dept\_id

salary = c(23000, 35000, 40000, 80000, 230000, 98000, 50000, 85000, 130000)

salary

employee = data.frame(eid, ename, desig, dept\_id, salary)

employee

aggregate(salary~dept\_id, employee, mean)

aggregate(salary~dept\_id, employee, max)

aggregate(salary~dept\_id+desig, employee, mean)

aggregate(salary~dept\_id+desig, employee, min)

#data()

install.packages("plyr")

require("plyr")

#data(package = "plyr")

# using plyr package

employee1 = data.frame(eid, ename, desig, dept\_id, salary)

employee1

employee1$eid[employee1$salary < 36000] <- 0

employee1

any(is.na(employee1$eid))

# using data table

install.packages("data.table")

require("data.table")

theDT = data.table(eid, ename, desig, dept\_id, salary)

theDT

class(theDT$ename)

class(employee1$ename)

theDT[1:2,]

theDT[theDT$eid >= 2504]

theDT[, list(eid,desig)]

theDT[, ename]

theDT[, list(ename)]

# if yu use column name as character name then yu should use with attribute

theDT[, "ename", with = FALSE]

theDT[, c("ename","eid"), with = FALSE]

# get all the tables

tables()

# setting the key in the data table

setkey(theDT, salary)

theDT

key(theDT)

tables()

theDT[theDT$salary > 50000]

setkey(theDT, desig, dept\_id, salary)

tables()

theDT

# aggregate on data tables

theDT[, mean(salary), by = dept\_id]

theDT[, list(price = mean(salary)), by = dept\_id]

theDT[, mean(salary), by = list(dept\_id,desig)]

theDT[, list(price = mean(salary), desig = mean(eid)), by = dept\_id]

#data()

**Output :**

> theMatrix = matrix(1:9, nrow = 3)

> theMatrix

[,1] [,2] [,3]

[1,] 1 4 7

[2,] 2 5 8

[3,] 3 6 9

> apply(theMatrix, 1, sum)

[1] 12 15 18

> apply(theMatrix, 2, sum)

[1] 6 15 24

> theMatrix[2,3] <- NA

> apply(theMatrix, 1, sum)

[1] 12 NA 18

> apply(theMatrix, 1, sum, na.rm = TRUE)

[1] 12 7 18

> rowSums(theMatrix, na.rm = TRUE)

[1] 12 7 18

> colSums(theMatrix, na.rm = TRUE)

[1] 6 15 16

> theList <- list(A = matrix(1:9,3), B = 1:5, C = matrix(1:4,2), D = 2)

> theList

$A

[,1] [,2] [,3]

[1,] 1 4 7

[2,] 2 5 8

[3,] 3 6 9

$B

[1] 1 2 3 4 5

$C

[,1] [,2]

[1,] 1 3

[2,] 2 4

$D

[1] 2

> lapply(theList, sum) # returns list

$A

[1] 45

$B

[1] 15

$C

[1] 10

$D

[1] 2

> sapply(theList, sum) # returns data frame

A B C D

45 15 10 2

> theNames <- c("Tyson", "Rizwanullah", "Dragon Emperor", "A.Charan")

> lapply(theNames, nchar)

[[1]]

[1] 5

[[2]]

[1] 11

[[3]]

[1] 14

[[4]]

[1] 8

> sapply(theNames, nchar)

Tyson Rizwanullah Dragon Emperor A.Charan

5 11 14 8

> f3 <- function(x,y)

+ {

+ NROW(x) + NCOL(y)

+ }

> flist = list(A = matrix(1:16,4), B = matrix(1:16,2), C = 1:5)

> flist

$A

[,1] [,2] [,3] [,4]

[1,] 1 5 9 13

[2,] 2 6 10 14

[3,] 3 7 11 15

[4,] 4 8 12 16

$B

[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]

[1,] 1 3 5 7 9 11 13 15

[2,] 2 4 6 8 10 12 14 16

$C

[1] 1 2 3 4 5

> slist = list(A = matrix(1:16,4), B = matrix(1:16,2), C = 15:1)

> slist

$A

[,1] [,2] [,3] [,4]

[1,] 1 5 9 13

[2,] 2 6 10 14

[3,] 3 7 11 15

[4,] 4 8 12 16

$B

[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]

[1,] 1 3 5 7 9 11 13 15

[2,] 2 4 6 8 10 12 14 16

$C

[1] 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

> mapply(identical, flist, slist)

A B C

TRUE TRUE FALSE

> mapply(f3, flist, slist)

A B C

8 10 6

> eid = c(2501 : 2509)

> eid

[1] 2501 2502 2503 2504 2505 2506 2507 2508 2509

> ename = c("Rizwan", "Ajay", "Mounav", "Charan", "Srinivas", "Vamsi", "Deepak", "Abhi", "Pavan")

> ename

[1] "Rizwan" "Ajay" "Mounav" "Charan" "Srinivas" "Vamsi" "Deepak" "Abhi" "Pavan"

> desig = c("sales", "accounts", "manager", "sales", "sales", "accounts", "accounts", "manager", "sales")

> desig

[1] "sales" "accounts" "manager" "sales" "sales" "accounts" "accounts" "manager" "sales"

> dept\_id = c(10,10,10,10,20,20,20,30,30)

> dept\_id

[1] 10 10 10 10 20 20 20 30 30

> salary = c(23000, 35000, 40000, 80000, 230000, 98000, 50000, 85000, 130000)

> salary

[1] 23000 35000 40000 80000 230000 98000 50000 85000 130000

> employee = data.frame(eid, ename, desig, dept\_id, salary)

> employee

eid ename desig dept\_id salary

1 2501 Rizwan sales 10 23000

2 2502 Ajay accounts 10 35000

3 2503 Mounav manager 10 40000

4 2504 Charan sales 10 80000

5 2505 Srinivas sales 20 230000

6 2506 Vamsi accounts 20 98000

7 2507 Deepak accounts 20 50000

8 2508 Abhi manager 30 85000

9 2509 Pavan sales 30 130000

> aggregate(salary~dept\_id, employee, mean)

dept\_id salary

1 10 44500

2 20 126000

3 30 107500

> aggregate(salary~dept\_id, employee, max)

dept\_id salary

1 10 80000

2 20 230000

3 30 130000

> aggregate(salary~dept\_id+desig, employee, mean)

dept\_id desig salary

1 10 accounts 35000

2 20 accounts 74000

3 10 manager 40000

4 30 manager 85000

5 10 sales 51500

6 20 sales 230000

7 30 sales 130000

> aggregate(salary~dept\_id+desig, employee, min)

dept\_id desig salary

1 10 accounts 35000

2 20 accounts 50000

3 10 manager 40000

4 30 manager 85000

5 10 sales 23000

6 20 sales 230000

7 30 sales 130000

> employee1 = data.frame(eid, ename, desig, dept\_id, salary)

> employee1

eid ename desig dept\_id salary

1 2501 Rizwan sales 10 23000

2 2502 Ajay accounts 10 35000

3 2503 Mounav manager 10 40000

4 2504 Charan sales 10 80000

5 2505 Srinivas sales 20 230000

6 2506 Vamsi accounts 20 98000

7 2507 Deepak accounts 20 50000

8 2508 Abhi manager 30 85000

9 2509 Pavan sales 30 130000

> employee1$eid[employee1$salary < 36000] <- 0

> employee1

eid ename desig dept\_id salary

1 0 Rizwan sales 10 23000

2 0 Ajay accounts 10 35000

3 2503 Mounav manager 10 40000

4 2504 Charan sales 10 80000

5 2505 Srinivas sales 20 230000

6 2506 Vamsi accounts 20 98000

7 2507 Deepak accounts 20 50000

8 2508 Abhi manager 30 85000

9 2509 Pavan sales 30 130000

> any(is.na(employee1$eid))

[1] FALSE

> theDT = data.table(eid, ename, desig, dept\_id, salary)

> theDT

eid ename desig dept\_id salary

1: 2501 Rizwan sales 10 23000

2: 2502 Ajay accounts 10 35000

3: 2503 Mounav manager 10 40000

4: 2504 Charan sales 10 80000

5: 2505 Srinivas sales 20 230000

6: 2506 Vamsi accounts 20 98000

7: 2507 Deepak accounts 20 50000

8: 2508 Abhi manager 30 85000

9: 2509 Pavan sales 30 130000

> class(theDT$ename)

[1] "character"

> class(employee1$ename)

[1] "character"

> theDT[1:2,]

eid ename desig dept\_id salary

1: 2501 Rizwan sales 10 23000

2: 2502 Ajay accounts 10 35000

> theDT[theDT$eid >= 2504]

eid ename desig dept\_id salary

1: 2504 Charan sales 10 80000

2: 2505 Srinivas sales 20 230000

3: 2506 Vamsi accounts 20 98000

4: 2507 Deepak accounts 20 50000

5: 2508 Abhi manager 30 85000

6: 2509 Pavan sales 30 130000

> theDT[, list(eid,desig)]

eid desig

1: 2501 sales

2: 2502 accounts

3: 2503 manager

4: 2504 sales

5: 2505 sales

6: 2506 accounts

7: 2507 accounts

8: 2508 manager

9: 2509 sales

> theDT[, ename]

[1] "Rizwan" "Ajay" "Mounav" "Charan" "Srinivas" "Vamsi" "Deepak" "Abhi" "Pavan"

> theDT[, list(ename)]

ename

1: Rizwan

2: Ajay

3: Mounav

4: Charan

5: Srinivas

6: Vamsi

7: Deepak

8: Abhi

9: Pavan

> theDT[, "ename", with = FALSE]

ename

1: Rizwan

2: Ajay

3: Mounav

4: Charan

5: Srinivas

6: Vamsi

7: Deepak

8: Abhi

9: Pavan

> theDT[, c("ename","eid"), with = FALSE]

ename eid

1: Rizwan 2501

2: Ajay 2502

3: Mounav 2503

4: Charan 2504

5: Srinivas 2505

6: Vamsi 2506

7: Deepak 2507

8: Abhi 2508

9: Pavan 2509

> tables()

NAME NROW NCOL MB COLS KEY

1: theDT 9 5 0 eid,ename,desig,dept\_id,salary

Total: 0MB

> setkey(theDT, salary)

> theDT

eid ename desig dept\_id salary

1: 2501 Rizwan sales 10 23000

2: 2502 Ajay accounts 10 35000

3: 2503 Mounav manager 10 40000

4: 2507 Deepak accounts 20 50000

5: 2504 Charan sales 10 80000

6: 2508 Abhi manager 30 85000

7: 2506 Vamsi accounts 20 98000

8: 2509 Pavan sales 30 130000

9: 2505 Srinivas sales 20 230000

> key(theDT)

[1] "salary"

> tables()

NAME NROW NCOL MB COLS KEY

1: theDT 9 5 0 eid,ename,desig,dept\_id,salary salary

Total: 0MB

> theDT[theDT$salary > 50000]

eid ename desig dept\_id salary

1: 2504 Charan sales 10 80000

2: 2508 Abhi manager 30 85000

3: 2506 Vamsi accounts 20 98000

4: 2509 Pavan sales 30 130000

5: 2505 Srinivas sales 20 230000

> setkey(theDT, desig, dept\_id, salary)

> tables()

NAME NROW NCOL MB COLS KEY

1: theDT 9 5 0 eid,ename,desig,dept\_id,salary desig,dept\_id,salary

Total: 0MB

> theDT

eid ename desig dept\_id salary

1: 2502 Ajay accounts 10 35000

2: 2507 Deepak accounts 20 50000

3: 2506 Vamsi accounts 20 98000

4: 2503 Mounav manager 10 40000

5: 2508 Abhi manager 30 85000

6: 2501 Rizwan sales 10 23000

7: 2504 Charan sales 10 80000

8: 2505 Srinivas sales 20 230000

9: 2509 Pavan sales 30 130000

> theDT[, mean(salary), by = dept\_id]

dept\_id V1

1: 10 44500

2: 20 126000

3: 30 107500

> theDT[, list(price = mean(salary)), by = dept\_id]

dept\_id price

1: 10 44500

2: 20 126000

3: 30 107500

>

> theDT[, mean(salary), by = list(dept\_id,desig)]

dept\_id desig V1

1: 10 accounts 35000

2: 20 accounts 74000

3: 10 manager 40000

4: 30 manager 85000

5: 10 sales 51500

6: 20 sales 230000

7: 30 sales 130000

> theDT[, list(price = mean(salary), desig = mean(eid)), by = dept\_id]

dept\_id price desig

1: 10 44500 2502.5

2: 20 126000 2506.0

3: 30 107500 2508.5

**Result :** Successfully Completed the Aim