DATA ANALYTICS – 4027

LAB-4

Name: Hari Krishna P

Reg No: 19BCE7675

DATE: 28/10/21

Contents:

- **Statistical Functions**
- **Data frames**
- **Join Operations**

Submitted to:

Prof . Hari Seetha

Ex.4

1.Check all the mathematical functions and show the results of each function

```
> x=1.2
>
> sqrt(x)
[1] 1.095445
>
> round(x,digits = 2)
[1] 1.2
>
> round(x,digits = 10)
[1] 1.2
>
> x=1.2
> round(x,digits = 2)
[1] 1.2
>
> round(x,digits = 10)
[1] 1.2
>
> round(x,digits = 1)
[1] 1.2
>
> x=1.2222
> round(x,digits = 1)
[1] 1.2
>
> signif(x,digits = 2)
[1] 1.2
>
```

```
> cos(x)
[1] 0.3415789
>
> sin(x)
[1] 0.9398531
>
> tan(x)
[1] 2.751496
>
> log(x)
[1] 0.2006525
>
> log10(x)
[1] 0.08714228
>
> exp(x)
[1] 3.394648
>
```

```
> factorial(x)
[1] 1.115353
>
```

2.Implement charcater functions and show the results

```
> a="Hari is good"
> toupper(a)
[1] "HARI IS GOOD"
>
> tolower(a)
[1] "hari is good"
>
> diff(x,lag=1)
numeric(0)
>
> x=10
> diff(x,lag=1)
numeric(0)
>
>
> b = "& A Legend"
> is.character(a)
[1] TRUE
>
> as.character(a)
[1] "Hari is good"
>
> strsplit(a,split = 1)
[[1]]
[1] "Hari is good"

>
> paste(a,b)
[1] "Hari is good & A Legend"
>
```

3.Create a vector of numbers and work with all statistical functions and report the results

```
> x<-c(10,6,55,100,2,1)
> mean(x)
[1] 29
>
> median(x)
[1] 8
>
> mode(x)
[1] "numeric"
>
> var(x)
[1] 1624
>
> sum(x)
[1] 174
>
> sd(x)
[1] 40.29888
>
> scale(x)
            [,1]
[1,] -0.4714771
[2,] -0.5707354
[3,]  0.6451792
[4,]  1.7618354
[5,] -0.6699938
[6,] -0.6948083
attr(,"scaled:center")
[1] 29
attr(,"scaled:scale")
[1] 40.29888
>
```

4.Create a dataframe Emp to store (Empid,empname,age,sal).Sort the data in the descending order of age

```
> emp <- data.frame(Empid=c(20,10,30,50,2),empname=c("Arvinth","Hari","Ram","Nadella","Mark"),age=c(20,21,22,25,23),sal=c("10k","80k","50k","80k","20k"))
> emp
  Empid empname age sal
1    20 Arvinth  20 10k
2    10    Hari  21 80k
3    30     Ram  22 50k
4    50 Nadella  25 80k
5     2    Mark  23 20k
>
```

```
> attach(emp)
>
> New_emp <- emp[order(age),]
> New_emp
  Empid empname age sal
1    20 Arvinth  20 10k
2    10    Hari  21 80k
3    30     Ram  22 50k
5     2    Mark  23 20k
4    50 Nadella  25 80k
>
```

Change is placed at 4th and 5th row

5.Sort the data in Emp in the descending order of sal and ascending order of name

```
> New_emp<-emp[order(sal,decreasing=TRUE,na.last=TRUE),]
> New_emp
  Empid empname age sal
2    10     Hari  21 80k
4    50 Nadella  25 80k
3    30      Ram  22 50k
5     2     Mark  23 20k
1    20 Arvinth  20 10k
> New_emp<-emp[order(empname),]
> New_emp
  Empid empname age sal
1    20 Arvinth  20 10k
2    10     Hari  21 80k
5     2     Mark  23 20k
4    50 Nadella  25 80k
3    30      Ram  22 50k
>
```

6.Add a column DeptNo to Emp.

```
> emp1=emp['DeptNo'] <- NA
> emp
  Empid empname age sal DeptNo
1    20 Arvinth  20 10k     NA
2    10     Hari  21 80k     NA
3    30      Ram  22 50k     NA
4    50 Nadella  25 80k     NA
5     2     Mark  23 20k     NA
>
```

7. Create a data frame Dept(DeptNo ,DeptName) and Projects(DeptNo,PNo,Pname)

```
> Dept<-data.frame(DeptNo=c(1,2,3,4,5),DeptName=c("CSE","ECE","MECH","M.TECH","BBA"))
> Dept
  DeptNo DeptName
1      1      CSE
2      2      ECE
3      3     MECH
4      4   M.TECH
5      5      BBA
> Projects<-data.frame(DeptNo=c(1,2,3,4,5),Pno=c(1001,2001,3001,4001,5001),Pname=c("Data Mining","Circuit System","Automater Controlled Car","Home Automated S
ystem","Management Analysis"))
> Projects
  DeptNo  Pno                  Pname
1      1 1001            Data Mining
2      2 2001         Circuit System
3      3 3001 Automater Controlled Car
4      4 4001    Home Automated System
5      5 5001      Management Analysis
>
```

8. Perform Inner Join and Cross Join using Emp and Dept

Inner Join:

```
> df= merge(x=emp,y=Dept,by="DeptNo")
> df
[1] DeptNo   Empid    empname  age      sal      DeptName
<0 rows> (or 0-length row.names)
```

As employee and dept table doesn't have any common column's

Cross Join:

```
> df= merge(x=Projects,y=Dept,by=NULL)
> df
   DeptNo.x  Pno                    Pname DeptNo.y DeptName
1         1 1001             Data Mining        1      CSE
2         2 2001          Circuit System        1      CSE
3         3 3001 Automater Controlled Car        1      CSE
4         4 4001    Home Automated System        1      CSE
5         5 5001      Management Analysis        1      CSE
6         1 1001             Data Mining        2      ECE
7         2 2001          Circuit System        2      ECE
8         3 3001 Automater Controlled Car        2      ECE
9         4 4001    Home Automated System        2      ECE
10        5 5001      Management Analysis        2      ECE
11        1 1001             Data Mining        3     MECH
12        2 2001          Circuit System        3     MECH
13        3 3001 Automater Controlled Car        3     MECH
14        4 4001    Home Automated System        3     MECH
15        5 5001      Management Analysis        3     MECH
16        1 1001             Data Mining        4   M.TECH
17        2 2001          Circuit System        4   M.TECH
18        3 3001 Automater Controlled Car        4   M.TECH
19        4 4001    Home Automated System        4   M.TECH
20        5 5001      Management Analysis        4   M.TECH
21        1 1001             Data Mining        5      BBA
22        2 2001          Circuit System        5      BBA
23        3 3001 Automater Controlled Car        5      BBA
24        4 4001    Home Automated System        5      BBA
25        5 5001      Management Analysis        5      BBA
>
```

9.Perform Left Join,Right Join,Outer Join using Emp and Project

RIGHT JOIN:

```
> df= merge(x=emp,y=Dept,by="DeptNo",all.y= TRUE)
> df
  DeptNo Empid empname age  sal DeptName
1      1    NA     <NA>  NA <NA>      CSE
2      2    NA     <NA>  NA <NA>      ECE
3      3    NA     <NA>  NA <NA>     MECH
4      4    NA     <NA>  NA <NA>   M.TECH
5      5    NA     <NA>  NA <NA>      BBA
> 
```

OUTER JOIN:

```
> df= merge(x=emp,y=Dept,by="DeptNo",all= TRUE)
> df
   DeptNo Empid empname age  sal DeptName
1       1    NA     <NA>  NA <NA>      CSE
2       2    NA     <NA>  NA <NA>      ECE
3       3    NA     <NA>  NA <NA>     MECH
4       4    NA     <NA>  NA <NA>   M.TECH
5       5    NA     <NA>  NA <NA>      BBA
6      NA    20 Arvinth  20  10k      <NA>
7      NA    10     Hari  21  80k      <NA>
8      NA    30      Ram  22  50k      <NA>
9      NA    50 Nadella  25  80k      <NA>
10     NA     2     Mark  23  20k      <NA>
> 
```

10.Rename Column DeptNo to DNo in Emp.

```
> names(Dept)[1]<- "DNo"
> Dept
  DNo DeptName
1   1      CSE
2   2      ECE
3   3     MECH
4   4   M.TECH
5   5      BBA
> 
```

11. Add anew Emp (101,“Viswa”, NA,10000)

```
> emp1<-data.frame(Empid=c(101),empname=c("Viswa"),age=c("NA"),sal=c("10000"))
> New_Emp<-rbind(emp,emp1)
> New_Emp
  Empid empname age   sal
1    20 Arvinth  20   10k
2    10    Hari  21   80k
3    30     Ram  22   50k
4    50 Nadella  25   80k
5     2    Mark  23   20k
6   101   Viswa  NA 10000
> 
```

12.Replace missing value of age with mean,median and a value

```
> New_Emp
  Empid empname age   sal
1    20 Arvinth  20   10k
2    10    Hari  21   80k
3    30     Ram  22   50k
4    50 Nadella  25   80k
5     2    Mark  23   20k
6   101   Viswa  NA 10000
> 
```

```
x <- New_Emp$age[is.na(New_Emp$age)] <- mean(New_Emp$age,na.rm = TRUE)
x
[1] 20

y <- FinalNewEmp$age[is.na(FinalNewEmp$age)] <- median(FinalNewEmp$age,na.rm = TRUE)
>y
[1] 20
```