Half Yearly Exam (2023-24)
CLASS XI SC
SUB: DATA SCIENCE(844)

Time allowed -2 hours   **Max Marks- 50**

==

General Instructions:
- Please check this question paper contains 23 questions.
- The paper is divided into 5 Sections- A, B, C, D and E.
- Section A, consists of 10 questions (1 to 10). Each question carries 1 Mark.
- Section B, consists of 5 questions (11 to 15). Each question carries 2 Marks.
- Section C, consists of 4 questions (16 to 19). Each question carries 3 Marks.
- Section D, consists of 2 questions (20 to 21). Each question carries 4 Marks.
- Section E, consists of 2 questions (22 to 23). Each question carries 5 Marks.
- All programming questions are to be answered using R Language only

Q No	Section A (Write outputs from 1 to 5)	10 Marks
1	class(TRUE) [1] "logical"	1
2	seq(1:10) [1] 1 2 3 4 5 6 7 8 9 10	1
3	rep(5,7)	1
4	x <- c(10,20,30) y <- c(40,50,60) >x+y [1] 50 70 90	1
5	a <- seq(1:9) matrix(a,nrow=3,byrow=TRUE) [,1] [,2] [,3] [1,] 1 2 3 [2,] 4 5 6 [3,] 7 8 9	1
6	Write R command to check the current working directory >getwd() [1] "C:/Users/user/Documents"	1
7	Write R command to change the working directory to the folder **r01** in **C drive** >setwd('c:/r01') >getwd() [1] "c:/r01"	1
8	Fact is same as Story. State **True/False**	1
9	Write command in R to check for packages installed. >installed.packages() Package LibPath Version Priority base "base" "C:/Program Files/R/R-3.2.2/library" "3.2.2" "base" boot "boot" "C:/Program Files/R/R-3.2.2/library" "1.3-17" "recommended" class "class" "C:/Program Files/R/R-3.2.2/library" "7.3-13" "recommended"	1
10	What is the advantage of the command **library(datasets)?** Loads the package datasets into memory	1
	Section B	10 Marks
11	A list in R is declared as spclist<- list('Darjeeling','Kalimpong','Mirik')	2

	Write command to replace "Kalimpong' with 'Kurseong' >spclist[2] <- 'Kurseong'	
12	Name any four data objects in R vectors, matrices, arrays, data frames, tables and lists	2
13	Name any four basic vectors of R. logical, integer, double, complex, character and raw	2
14	If x and y are two vectors of different length,findx+y x <- c(1:10) y <- c(11,12) [1] 12 14 14 16 16 18 18 20 20 22	2

15	> mat
	   [,1] [,2] [,3] [,4]
	[1,] 1 4 7 10
	[2,] 2 5 8 11
	[3,] 3 6 9 12
	Give R command to access only the 3rd row of the matrix
	> mat
	   [,1] [,2] [,3] [,4]
	[1,] 1 4 7 10
	[2,] 2 5 8 11
	[3,] 3 6 9 12
	>mat[3,]
	[1] 3 6 9 12
		2

Section C		12Marks
16	A list is declared as mylist<- list(c("Blue","Orange"),matrix(c(10,230,30,60),nrow=2), c(FALSE, TRUE)) And if elements of the list are named as below names(mylist) <- c("Colors","explMatrix","logicals") Write command to replace the number 60 with 80 in the matrix of mylist mylist$explMatrix[2,2] <- 80 >mylist$explMatrix [,1] [,2] [1,] 10 30 [2,] 230 80	3
17	Explain importance of incorporating ethics in Data Science Data Science Ethics is an important topic of discussion in today's world. Organizations and companies using data and implementing data science must follow a set of ethics while dealing with data. When used ethically, data may help you make better decisions and make a difference in the world.	3
18	The syntax used to create a line chart in R is:**plot (v, type, xlab, ylab, main, col)** Explain the various parameters used in the function.	3

The basic syntax to create a line chart in R is −

```
plot(v,type,col,xlab,ylab)
```

Following is the description of the parameters used −

- **v** is a vector containing the numeric values.
- **type** takes the value "p" to draw only the points, "l" to draw only the lines and "o" to draw both points and lines.
- **xlab** is the label for x axis.
- **ylab** is the label for y axis.
- **main** is the Title of the chart.
- **col** is used to give colors to both the points and lines.

19	Create a DataFrame in R using 3 vectors of same length. >empno<- c(101,102,103,104,105) >empname<- c('abc','def','ghi','jkl','mno') >projn<- c('p01','p02','p03','p04','p05') >empl<- data.frame(empno,empname,projn) >empl empnoempnameprojn 1 101 abc p01 2 102 def p02 3 103 ghi p03 4 104 jkl p04 5 105 mno p05	3
	Section D	08 Marks
20		4

	QTR1	QTR2	QTR3	QTR4
SBI	NA	NA	NA	NA
HDFC	NA	NA	NA	NA
L and T	NA	NA	NA	NA
ONGC	NA	NA	NA	NA

Create the above matrix in R with row and column names. Also update the QTR4 value of HDFC to 22000

```
> s=matrix(nrow=4,ncol=4)
>rownames(s) <-c('SBI','HDFC','L and T','ONGC')
>colnames(s)<-c('QTR1','QTR2','QTR3','QTR4')
> s
     QTR1 QTR2 QTR3 QTR4
SBI    NA   NANANA
HDFC   NA   NANANA
L and T NA   NANANA
ONGC   NA   NANANA

> s[2,4] <- 22000
> s
     QTR1 QTR2 QTR3  QTR4
SBI    NA   NANANA
```

	HDFC NA NANA 22000 L and T NA NANANA ONGC NA NANANA >	

21	A CSV file 'tt.csv' as shown below is stored in the working directory, give command to read data from the file. Also store the age of each person in a vector **age** by reading only the Age column from the file.
	  Name Age
	1 a 10
	2 s 20
	3 c 10
	4 r 20
	5 t 10
	6 y 10
	
	> ss=read.csv('tt.csv')
	> ss
	  Name Age
	1 a 10
	2 s 20
	3 c 10
	4 r 20
	5 t 10
	6 y 10
	> age<-ss$Age
	> age
	[1] 10 20 10 20 10 10
		4

	Section E	10 Marks

22	Write R command to create a data frame **marks** to store the **names**Chakachak,Jhakajhak,Takatak and Fatafat and marks obtained in data science **marksinDS**as 39 42 40 46. Display the content of the data frame in descending order of **marksinDS**.
	> names<-c('Chakachak','Jhakajhak','Takatak','Fatafat')
	>marksinDS<-c(39,42,40,46)
	> marks<-data.frame(names,marksinDS)
	> marks
	    names marksinDS
	1 Chakachak 39
	2 Jhakajhak 42
	3 Takatak 40
	4 Fatafat 46
	>
	
	> marks[order(-marks$marksinDS),]
	    names marksinDS
	4 Fatafat 46
	2 Jhakajhak 42
	3 Takatak 40
	1 Chakachak 39
	>
		5

23	Find the subset of the above data frame and display only those rows where marks is greater than 40. > subset(marks,marksinDS>40) names marksinDS 2 Jhakajhak 42	5

| | 4 Fatafat 46 > | |