

DATA ANALYTICS – 4027

LAB-4

Name: Hari Krishna P

Reg No: 19BCE7675

DATE: 30/10/21

Contents:

- **Data frame**
- **Functions**
- **NA Value's**

Submitted to:

Prof . Hari Seetha

Ex-5

1. If `X <- c(22,3,7,NA,NA,67)` what will be the output for the R statement `length(X)`

```
> x<-c(22,3,7,NA,NA,67)
> length(x)
[1] 6
> |
```

2. `X = c(NA,3,14,NA,33,17,NA,41)` write some R code that will remove all occurrences of NA in X.

```
> x <- c(NA,3,14,NA,33,17,NA,41)
> y <- na.omit(x)
> y
[1] 3 14 33 17 41
attr(,"na.action")
[1] 1 4 7
attr(,"class")
[1] "omit"
> |
```

3. If `Y = c(1,3,12,NA,33,7,NA,21)` what R statement will replace all occurrences of NA with 11?

```
> Y <- c(1,3,12,NA,33,7,NA,21)
> Y[is.na(Y)]<-11
> Y
[1] 1 3 12 11 33 7 11 21
> |
```

4. If `X = c(34,33,65,37,89,NA,43,NA,11,NA,23,NA)` then what will count the number of occurrences of NA in X?

```
> X <- c(34,33,65,37,89,NA,43,NA,11,NA,23,NA)
> sum(is.na(x))
[1] 3
> |
```

5. Create a dataframe using following features:

name = ('Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'),

score = (12.5, 9, 16.5, 12, 9, 20, 14.5, 13.5, 8, 19),

attempts = (1, NA, 2, NA, 2, NA, 1, NA, 2, 1),

qualify = ('yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes')

```
> df <- data.frame(name = c("Anastasia", "Dima", "Katherine", "James", "Emily", "Michael", "Matthew", "Laura", "Kevin", "Jonas"), score = c(12.5, 9, 16.5, 12, 9, 20, 14.5, 13.5, 8, 19), attempts = c(1, NA, 2, NA, 2, NA, 1, NA, 2, 1), qualify = c("yes", "no", "yes", "no", "no", "yes", "yes", "no", "no", "yes"))
> df
```

	name	score	attempts	qualify
1	Anastasia	12.5	1	yes
2	Dima	9.0	NA	no
3	Katherine	16.5	2	yes
4	James	12.0	NA	no
5	Emily	9.0	2	no
6	Michael	20.0	NA	yes
7	Matthew	14.5	1	yes
8	Laura	13.5	NA	no
9	Kevin	8.0	2	no
10	Jonas	19.0	1	yes

```
> |
```

6. m to replace NA values with 3 in a given data frame.

```
> df[is.na(df)] = 3
> df
```

	name	score	attempts	qualify
1	Anastasia	12.5	1	yes
2	Dima	9.0	3	no
3	Katherine	16.5	2	yes
4	James	12.0	3	no
5	Emily	9.0	2	no
6	Michael	20.0	3	yes
7	Matthew	14.5	1	yes
8	Laura	13.5	3	no
9	Kevin	8.0	2	no
10	Jonas	19.0	1	yes

```
> |
```

7. Write a R program to count the number of NA values in a data frame Write a R program column

```
> df
  name score attempts qualify
1 Anastasia 12.5      1    yes
2 Dima      9.0     NA     no
3 Katherine 16.5      2    yes
4 James    12.0     NA     no
5 Emily     9.0      2     no
6 Michael  20.0     NA     yes
7 Matthew  14.5      1    yes
8 Laura    13.5     NA     no
9 Kevin     8.0      2     no
10 Jonas   19.0      1    yes
>
> sum(is.na(df))
[1] 4
> |
```

7. Write a R program to change a column name of a given data frame.

```
> colnames(df)[colnames(df) == "qualify"] <- "Option"
> df
  name score attempts Option
1 Anastasia 12.5      1    yes
2 Dima      9.0     NA     no
3 Katherine 16.5      2    yes
4 James    12.0     NA     no
5 Emily     9.0      2     no
6 Michael  20.0     NA     yes
7 Matthew  14.5      1    yes
8 Laura    13.5     NA     no
9 Kevin     8.0      2     no
10 Jonas   19.0      1    yes
> |
```

8. Write a R program to change more than one column name of a given data frame

```
> colnames(df) <- c("Name", "S1", "A1", "O1")
> df
  Name    S1 A1 O1
1 Anastasia 12.5 1 yes
2      Dima  9.0 NA no
3 Katherine 16.5 2 yes
4      James 12.0 NA no
5      Emily  9.0 2 no
6   Michael 20.0 NA yes
7   Matthew 14.5 1 yes
8      Laura 13.5 NA no
9      Kevin  8.0 2 no
10     Jonas 19.0 1 yes
> |
```

9. Write a R program to display rows with missing values

```
> rowSums(is.na(df))
[1] 0 1 0 1 0 1 0 1 0 0
> |
```

```
> is.na(df$A1)
[1] FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE FALSE
```

Full:

```
> is.na(df)
  Name    S1    A1    O1
[1,] FALSE FALSE FALSE FALSE
[2,] FALSE FALSE  TRUE FALSE
[3,] FALSE FALSE FALSE FALSE
[4,] FALSE FALSE  TRUE FALSE
[5,] FALSE FALSE FALSE FALSE
[6,] FALSE FALSE  TRUE FALSE
[7,] FALSE FALSE FALSE FALSE
[8,] FALSE FALSE  TRUE FALSE
[9,] FALSE FALSE FALSE FALSE
[10,] FALSE FALSE FALSE FALSE
> |
```

10. Identify the location of missing values in the given data frame.

```
> which(is.na(df))
[1] 22 24 26 28
> |
```

11. Write a R program to create a data frame using two given vectors and display the duplicated elements rows of the said data frame.

a = (10,20,10,10,40,50,20,30)

b = (10,30,10,20,0,50,30,30)

Data Frame:

```
> df_1 <- data.frame( a = c(10,20,10,10,40,50,20,30), b = c(10,30,10,20,0,50,30,30))
> df_1
  a b
1 10 10
2 20 30
3 10 10
4 10 20
5 40  0
6 50 50
7 20 30
8 30 30
> |
```

Duplicated:

```
> duplicated(df_1)
[1] FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE
> |
```

12. Display unique rows of the above said data frame?

```
> unique(df_1)
  a b
1 10 10
2 20 30
4 10 20
5 40  0
6 50 50
8 30 30
> |
```