Godavari Foundation's

Godavari College of Engineering, Jalgaon

Department of Computer

Continuous Assessment I/II

Introduction to Data Science with R

Date:-_____

Name of Student:- _____

Class:-_____ **PRN No:-_____**

Title: -

_____

_____

Aim: -

_____

_____

_____

_____

Software Requirement: _____

Hardware Requirement:- _____

Theory:-

Data Frames

A data frame is a list of vectors which are of equal length. A matrix contains only one type of data, while a data frame accepts different data types (numeric, character, factor, etc.).

Create a data frame

Create first data set by combining four variables of same length.

```
# Create a, b, c, d variables
```

```
a <- c(10,20,30,40)
b <- c('book', 'pen', 'textbook', 'pencil_case')
c <- c(TRUE,FALSE,TRUE,FALSE)
d <- c(2.5, 8, 10, 7)
# Join the variables to create a data frame
df <- data.frame(a,b,c,d)
df
```

Output:

```
##  a          b      c     d
## 1  1       book  TRUE   2.5
## 2  2        pen  TRUE   8.0
## 3  3    textbook  TRUE  10.0
## 4  4 pencil_case FALSE   7.0
```

Change the column name with the function names().

```
# Name the data frame
names(df) <- c('ID', 'items', 'store', 'price')
df
```

Output:

```
##   ID       items store price
## 1 10        book  TRUE   2.5
## 2 20         pen FALSE   8.0
## 3 30    textbook  TRUE  10.0
## 4 40 pencil_case FALSE   7.0
```

```
# Print the structure
str(df)
```

Output:

```
## 'data.frame':    4 obs. of  4 variables:
##  $ ID   : num  10 20 30 40
##  $ items: Factor w/ 4 levels "book","pen","pencil_case",..: 1 2 4 3
##  $ store: logi  TRUE FALSE TRUE FALSE
##  $ price: num  2.5 8 10 7
```

```
## Select row 1 in column 2
df[1,2]
```

```
## [1] book
## Levels: book pen pencil_case textbook
```

```
## Select Rows 1 to 2
df[1:2,]
```

Output:

```
##   ID items store price
## 1 10  book  TRUE   2.5
## 2 20   pen FALSE   8.0
```

```
## Select Columns 1
df[,1]
```

Output:

```
## [1] 10 20 30 40
```

```
## Select Rows 1 to 3 and columns 3 to 4
df[1:3, 3:4]
```

Output:

```
##   store price
## 1  TRUE   2.5
## 2 FALSE   8.0
## 3  TRUE  10.0
```

Select the columns with their names.

```
# Slice with columns name
df[, c('ID', 'store')]
```

Output:

```
##   ID store
## 1 10  TRUE
## 2 20 FALSE
## 3 30  TRUE
## 4 40 FALSE
```

Append a Column to Data Frame

Use the symbol $ to append a new variable.

```
# Create a new vector
quantity <- c(10, 35, 40, 5)

# Add `quantity` to the `df` data frame
```

```
df$quantity <- quantity
df
```

Output:

```
##   ID       items store price quantity
## 1 10        book  TRUE   2.5       10
## 2 20         pen FALSE   8.0       35
## 3 30    textbook  TRUE  10.0       40
## 4 40 pencil_case FALSE   7.0        5
```

Select a column of a data frame

```
# Select the column ID
df$ID
```

Output:

```
## [1] 1 2 3 4
```

Subset a data frame

Return only the items with price above 10

```
# Select price above 5
subset(df, subset = price > 5)
```

Output:

```
ID       items store price
2 20         pen FALSE     8
3 30    textbook  TRUE    10
4 40 pencil_case FALSE     7
```

Sort a Data Frame by Multiple Columns in R

First generate a data frame to manipulate.

```
# Generate data frame
dataframe <- data.frame(
  x = c("apple", "orange", "banana", "strawberry"),
  y = c("a", "d", "b", "c"),
  z = c(4:1))

# Print data frame
dataframe
```

```
        x y z
```

1 apple a 4
2 orange d 3
3 banana b 2
4 strawbery c 1

The Order Function

Use order() to simply sort a vector of five randomly ordered numbers

```
# Create unordered vector
vector = c(2, 5, 1, 3, 4)

# Print vector
vector

# Sort in ascending order
vector[order(vector)]
```

Output
[1] 2 5 1 3 4
[1] 1 2 3 4 5

Sorting a Data Frame by Vector Name

Sort by the vector z by adding the following code to the script

```
# Sort by vector name [z]
dataframe[with(dataframe, order(z)),]
```

Output

x y z
4 strawberry c 1
3 banana b 2
2 orange d 3
1 apple a 4

Sorting by Column Index

Sort based on the numeric index of a column in the data frame, rather than the specific name.
dataframe[order(dataframe[,1]),]

Output

```
          x y z
1      apple a 4
3     banana b 2
2     orange d 3
4 strawberry c 1
```

Sorting by Multiple Columns

```
# Sort by vector name [z] then [x]
dataframe[with(dataframe, order(z, x)),]
```

Source Code:-

```
## Create a data frame
a <- c(10,20,30,40)
b <- c('book', 'pen', 'textbook', 'pencil_case')
c <- c(TRUE,FALSE,TRUE,FALSE)
d <- c(2.5, 8, 10, 7)
# Join the variables to create a data frame
df <- data.frame(a,b,c,d)

## Show df
df

## Change the column name
names(df) <- c('ID', 'items', 'store', 'price')

## Show df
df

# Print the structure
str(df)

## Select row 1 in column 2
df[1,2]

## Select Rows 1 to 2
df[1:2,]

## Select Columns 1
df[,1]

## Select Rows 1 to 3 and columns 3 to 4
df[1:3, 3:4]
```

```
## Select the columns with their names
df[, c('ID', 'store')]

## Append a Column to Data Frame
# Create a new vector
quantity <- c(10, 35, 40, 5)

# Add `quantity` to the `df` data frame
df$quantity <- quantity

## Show df
df

## Select a column of a data frame
df$ID

# Select price above 5
subset(df, subset = price > 5)

# Create unordered vector
vector = c(2, 5, 1, 3, 4)

# Print vector
vector

# Sort in ascending order
vector[order(vector)]

# Generate data frame
x = c("apple", "orange", "banana", "strawberry")
y = c("a", "d", "b", "c")
z = c(4:1)
dataframe <- data.frame(x,y,z)

# Print data frame
dataframe


# Sort by vector name [z]
dataframe[with(dataframe, order(z)),]

# Sorting by Column Index
dataframe[order( dataframe[,1] ),]

# Sorting by Multiple Columns
dataframe[with(dataframe, order(z, x)),]
```

Output:-

```
> ## Create a data frame
> a <- c(10,20,30,40)
> b <- c('book', 'pen', 'textbook', 'pencil_case')
> c <- c(TRUE,FALSE,TRUE,FALSE)
> d <- c(2.5, 8, 10, 7)
> # Join the variables to create a data frame
> df <- data.frame(a,b,c,d)
>
> ## Show df
> df
   a           b     c    d
1 10        book  TRUE  2.5
2 20         pen FALSE  8.0
3 30    textbook  TRUE 10.0
4 40 pencil_case FALSE  7.0
>
> ## Change the column name
> names(df) <- c('ID', 'items', 'store', 'price')
>
> ## Show df
> df
  ID       items store price
1 10        book  TRUE   2.5
2 20         pen FALSE   8.0
3 30    textbook  TRUE  10.0
4 40 pencil_case FALSE   7.0
>
> # Print the structure
> str(df)
'data.frame':   4 obs. of  4 variables:
 $ ID   : num  10 20 30 40
 $ items: Factor w/ 4 levels "book","pen","pencil_case",..: 1 2 4 3
 $ store: logi  TRUE FALSE TRUE FALSE
 $ price: num  2.5 8 10 7
```

```
>
> ## Select row 1 in column 2
> df[1,2]
[1] book
Levels: book pen pencil_case textbook
>
> ## Select Rows 1 to 2
> df[1:2,]
  ID items store price
1 10  book  TRUE   2.5
2 20   pen FALSE   8.0
>
> ## Select Columns 1
> df[,1]
[1] 10 20 30 40
>
> ## Select Rows 1 to 3 and columns 3 to 4
> df[1:3, 3:4]
  store price
1  TRUE   2.5
2 FALSE   8.0
3  TRUE  10.0
>
> ## Select the columns with their names
> df[, c('ID', 'store')]
  ID store
1 10  TRUE
2 20 FALSE
3 30  TRUE
4 40 FALSE
>
> ## Append a Column to Data Frame
> # Create a new vector
> quantity <- c(10, 35, 40, 5)
>
> # Add `quantity` to the `df` data frame
> df$quantity <- quantity
>
> ## Show df
```

```
> df
  ID       items store price quantity
1 10        book  TRUE   2.5       10
2 20         pen FALSE   8.0       35
3 30    textbook  TRUE  10.0       40
4 40 pencil_case FALSE   7.0        5
>
> ## Select a column of a data frame
> df$ID
[1] 10 20 30 40
>
> # Select price above 5
> subset(df, subset = price > 5)
  ID       items store price quantity
2 20         pen FALSE     8       35
3 30    textbook  TRUE    10       40
4 40 pencil_case FALSE     7        5
>
> # Create unordered vector
> vector = c(2, 5, 1, 3, 4)
>
> # Print vector
> vector
[1] 2 5 1 3 4
>
> # Sort in ascending order
> vector[order(vector)]
[1] 1 2 3 4 5
>
> # Generate data frame
> x = c("apple", "orange", "banana", "strawberry")
> y = c("a", "d", "b", "c")
> z = c(4:1)
> dataframe <- data.frame(x,y,z)
>
> # Print data frame
> dataframe
          x y z
1     apple a 4
```

```
2     orange d 3
3     banana b 2
4 strawberry c 1
>
>
> # Sort by vector name [z]
> dataframe[with(dataframe, order(z)),]
          x y z
4 strawberry c 1
3     banana b 2
2     orange d 3
1      apple a 4
>
> # Sorting by Column Index
> dataframe[order( dataframe[,1] ),]
          x y z
1      apple a 4
3     banana b 2
2     orange d 3
4 strawberry c 1
>
> # Sorting by Multiple Columns
> dataframe[with(dataframe, order(z, x)),]
          x y z
4 strawberry c 1
3     banana b 2
2     orange d 3
1      apple a 4
```

Conclusion:-

_____

_____