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:-

What is a Matrix?

In R, a matrix is a collection of elements of the same data type (numeric, character, or logical), that is arranged into a fixed number of rows and columns.
In R, matrix can be constructed with the matrix function, for example

```
matrix(1:9, byrow=TRUE, nrow=3)
```

The first argument, is the collection of elements that R will arrange into the rows and columns of the matrix.`1:9` which constructs the vector c(1,2,3,4,5,6,7,8,9)

The argument `byrow` indicates that the matrix is filled by the rows. If the vector to be filled by the columns, just place `bycol=TRUE` or `byrow=FALSE`.

The third argument `nrow` indicates that the matrix should have three rows.

1. Create a matrix with 3 rows containing the numbers 1 upto 9 in the editor
   matrix(1:9, byrow=TRUE, nrow=3)

Analyzing matrices

Analyze the box office numbers of the Star Wars franchisee.

In the editor, three vectors are defined, each representing the box office numbers from the first three Star Wars movies. The first element of each vector indicates the US box office revenue, the second element of each vector refers to the Non-US box office

1. Construct a matrix with one row for each movie (thus with 3 rows), a column for the US box office revenue, and a second column for the non-US box office revenue. Name the matrix `star.wars.matrix`.

```
# Box office Star Wars: In Millions (!)
# First element: US, Second element: Non-US
new.hope <- c( 460.998007, 314.4);
empire.strikes <- c(290.475067, 247.900000);
return.jedi <- c(309.306177,165.8)

# Add code below to construct the matrix
star.wars.matrix <- matrix( c(new.hope,empire.strikes,return.jedi),
                            nrow=3, byrow=TRUE)
# Show
star.wars.matrix
```

Naming a matrix

Naming a matrix helps to read the data, it is also useful to select certain elements from the matrix later.
Similar to vectors, Add names for the rows and the columns of a matrix with:
rownames(my.matrix) <- row.names.vectors
colnames(my.matrix) <- col.names.vectors

1. Give the columns of `star.wars.matrix` the names `"US"` and `"non-US"`.
2. Give the rows of the matrix `star.wars.matrix` the names of the three movies. The movie names are: "A new hope", "The empire strikes back" and "Return of the Jedi".

```
# Box office Star Wars: In Millions (!)
# First element: US, Second element: Non-US
new.hope <- c( 460.998007, 314.4);
empire.strikes <- c(290.475067, 247.900000);
return.jedi <- c(309.306177,165.8)

# Construct matrix:
star.wars.matrix <- matrix(c(new.hope,empire.strikes,return.jedi), nrow=3,byrow=TRUE)
colnames(star.wars.matrix) <- c("US","non-US")
rownames(star.wars.matrix) <- c("A new hope","The empire strikes back","Return of the
Jedi")

# Print the matrix to the console:
star.wars.matrix
```

Calculating the worlwide box office

To calculate the total box office revenue for the three Star Wars movies, take the sum of the US revenue column and the non-US revenue column.
In R, the function `rowSums()` calculates the totals for each row of a matrix.

1. Calculate the worldwide box office figures for the three movies and put these in the vector named `worldwide.vector`.

```
# Box office Star Wars: In Millions (!)
# Construct matrix:
box.office.all <- c(461, 314.4,290.5, 247.9,309.3,165.8)
movie.names    <- c("A new hope","The empire strikes back","Return of the Jedi")
col.titles     <-  c("US","non-US")
star.wars.matrix       <- matrix(box.office.all, nrow=3,
byrow=TRUE,dimnames=list(movie.names,col.titles))
worldwide.vector <- rowSums(star.wars.matrix)

# Show
worldwide.vector
```

Adding a column for the Worlwide box office (2)

If a column or multiple columns are added to a matrix. A good way to do this is `cbind()`, which merges matrices and/or vectors together by column.

1. Add `worldwide.vector` as a new column to the `star.wars matrix` and assign to `all.wars.matrix`. Use the `cbind()` function.

```
# Box office Star Wars: In Millions (!)
# Construct matrix:
box.office.all <- c(461, 314.4,290.5, 247.9,309.3,165.8)
movie.names    <- c("A new hope","The empire strikes back","Return of the Jedi")
col.titles     <-  c("US","non-US")
star.wars.matrix      <- matrix(box.office.all,
nrow=3,byrow=TRUE,dimnames=list(movie.names,col.titles))

# Print the matrix to the console:
worldwide.vector <-  rowSums(star.wars.matrix);
worldwide.vector

# Print worldwide revenue per movie
# Bind the new variable total.per.movie as a column to star.wars
all.wars.matrix <- cbind( star.wars.matrix, worldwide.vector )

# Show
all.wars.matrix
```

Adding a column for the Worlwide box office

1. Assign to `all.wars.matrix` a new matrix with `star.wars.matrix` in the first three rows and `star.wars.matrix2` in the next three rows.

```
# Box office Star Wars: In Millions (!)
star.wars.matrix    # Matrix containing first trilogy box office
star.wars.matrix2 # Matrix containing second trilogy box office

# Combine the both Star Wars trilogies in one matrix
all.wars.matrix <- rbind(star.wars.matrix, star.wars.matrix2)

# Show
all.wars.matrix

# Construct matrix:
```

```
box.office.all <- c(461, 314.4,290.5, 247.9,309.3,165.8)
movie.names    <- c("A new hope","The empire strikes back","Return of the Jedi")
col.titles     <-  c("US","non-US")
star.wars.matrix       <- matrix(box.office.all, nrow=3, byrow=TRUE,
                         dimnames=list(movie.names,col.titles))

# Construct matrix2:
box.office.all2 <- c( 474.5, 552.5, 310.7, 338.7, 380.3, 468.5 )
movie.names2    <- c( "The Phantom Menace", "Attack of the Clones", "Revenge of the
Sith")

star.wars.matrix2 <- matrix(box.office.all2, nrow=3, byrow=TRUE,
                         dimnames=list(movie.names2,col.titles))
```

Adding a column for the Worlwide box office

1. Calculate the total revenue for the US and the non-US region and `assign total.revenue.vector`. Use the `colSums()` function.

Use the `colSums()` function with `star.wars.matrix` as argument to find the total box office per region.

```
# Print box office Star Wars: In Millions (!) for 2 trilogies
all.wars.matrix
total.revenue.vector <- colSums( all.wars.matrix )

# Construct matrix:
box.office.all <- c(461, 314.4,290.5, 247.9,309.3,165.8)
movie.names    <- c("A new hope","The empire strikes back","Return of the Jedi")
col.titles     <-  c("US","non-US")
star.wars.matrix       <- matrix(box.office.all, nrow=3, byrow=TRUE,
                         dimnames=list(movie.names,col.titles))

# Construct matrix2:
box.office.all2 <- c( 474.5, 552.5, 310.7, 338.7, 380.3, 468.5 )
movie.names2    <- c( "The Phantom Menace", "Attack of the Clones", "Revenge of the
Sith")
star.wars.matrix2 <- matrix(box.office.all2, nrow=3, byrow=TRUE,
                         dimnames=list(movie.names2,col.titles))
```

```
# Box office Star Wars: In Millions (!)
star.wars.matrix    # Matrix containing first trilogy box office
star.wars.matrix2 # Matrix containing second trilogy box office

# Combine the both Star Wars trilogies in one matrix
all.wars.matrix <- rbind( star.wars.matrix, star.wars.matrix2 )
```

Selection of matrix elements

Use the square brackets `[ ]` to select one or multiple elements from a matrix. Whereas vectors have 1 dimension, matrices have 2 dimensions, therefore use a comma to separate what to select from the rows and what from the columns. For example:
- `my.matrix[1,2]` selects from the first row the second element
- `my.matrix[1:3,2:4]` selects rows 1,2,3 and columns 2,3,4.

If all elements of a row or column are to be selected, no number is needed before or after the comma:
- `my.matrix[,1]` selects all elements of the first column
- `my.matrix[1,]` selects all elements of the first row.

1. Calculate the average Non-US revenue per movie. Assign this to the `non.us.all variable'
Use the function `mean()` to compute the average.

```
# Box office Star Wars: In Millions (!)
# Construct matrix:
box.office.all <- c(461, 314.4,290.5, 247.9,309.3,165.8)
movie.names    <- c("A new hope","The empire strikes back","Return of the Jedi")
col.titles     <-  c("US","non-US")
star.wars.matrix  <- matrix(box.office.all, nrow=3, byrow=TRUE,
                      dimnames=list(movie.names,col.titles))
star.wars.matrix
non.us.all  <- mean( star.wars.matrix[,2] )
non.us.some <- mean( star.wars.matrix[1:2,2] )

# Print to console both averages:
non.us.all
non.us.some
```

A little arithmetic with matrices

The standard operators like `+`, `-`, `/`, `*`, etc. work in an element-wise way on matrices in R.
For example: `2*my.matrix` multiplies each element of `my.matrix` by two.

Assume that the price of a ticket was 5 dollars. Box office numbers divided by the ticket price gives you the number of visitors.

1. Assign to `visitors` the matrix with the estimated number of Non-US and US visitors for the three movies.

```
# Box office Star Wars: In Millions (!)
# Construct matrix:
box.office.all <- c(461, 314.4,290.5, 247.9,309.3,165.8)
movie.names    <- c("A new hope","The empire strikes back","Return of the Jedi")
col.titles     <- c("US","non-US")
star.wars.matrix      <- matrix(box.office.all,
nrow=3,byrow=TRUE,dimnames=list(movie.names,col.titles))
visitors <- star.wars.matrix/5;

# Show me (also in millions!) the
Visitors
```

Source Code:-

```
#Construct a matrix
new.hope <- c( 460.998007, 314.4);
empire.strikes <- c(290.475067, 247.900000);
return.jedi <- c(309.306177,165.8)
star.wars.matrix <- matrix( c(new.hope,empire.strikes,return.jedi),
                    nrow=3, byrow=TRUE)

# Show
star.wars.matrix

## Naming a matrix
colnames(star.wars.matrix) <- c("US","non-US")
rownames(star.wars.matrix) <- c("A new hope","The empire strikes back","Return of the Jedi")

# Print the matrix to the console:
star.wars.matrix
```

```
## Calculating the worlwide box office
worldwide.vector <- rowSums(star.wars.matrix)

# Show
worldwide.vector

## Adding a column for the Worlwide box office (2)
all.wars.matrix <- cbind( star.wars.matrix, worldwide.vector )

# Show
all.wars.matrix

## Adding a column for the Worlwide box office
box.office.all2 <- c( 474.5, 552.5, 310.7, 338.7, 380.3, 468.5 )
star.wars.matrix2 <- matrix(box.office.all2, nrow=3, byrow=TRUE)

all.wars.matrix <- rbind(star.wars.matrix, star.wars.matrix2)

# Show me the
all.wars.matrix

## Selection of matrix elements

all.wars.matrix[1,2]
all.wars.matrix[1:3,1:2]
all.wars.matrix[,1]
all.wars.matrix[1,]

#Calculate the average
non.us.all  <- mean( star.wars.matrix[,2] )
non.us.some <- mean( star.wars.matrix[1:2,2] )

# Print to console both averages:
non.us.all
non.us.some

#Arithmetic
visitors <- star.wars.matrix/5;

# Show
visitors
```

Output:-

```
> #Construct a matrix
> new.hope <- c( 460.998007, 314.4);
> empire.strikes <- c(290.475067, 247.900000);
> return.jedi <- c(309.306177,165.8)
> star.wars.matrix <- matrix( c(new.hope,empire.strikes,return.jedi),
+                             nrow=3, byrow=TRUE)

> # Show
> star.wars.matrix
        [,1]  [,2]
[1,] 460.9980 314.4
[2,] 290.4751 247.9
[3,] 309.3062 165.8

> ## Naming a matrix
> colnames(star.wars.matrix) <- c("US","non-US")
> rownames(star.wars.matrix) <- c("A new hope","The empire strikes
back","Return of the Jedi")
>
> # Print the matrix to the console:
> star.wars.matrix
                          US non-US
A new hope              460.9980  314.4
The empire strikes back 290.4751  247.9
Return of the Jedi      309.3062  165.8
>
> ## Calculating the worlwide box office
> worldwide.vector <- rowSums(star.wars.matrix)
>
> # Show
> worldwide.vector
           A new hope The empire strikes back      Return of the Jedi
             775.3980               538.3751               475.1062
>
> ## Adding a column for the Worlwide box office (2)
> all.wars.matrix <- cbind( star.wars.matrix, worldwide.vector )
>
> # Show
> all.wars.matrix
```

```
                          US non-US worldwide.vector
A new hope              460.9980  314.4         775.3980
The empire strikes back 290.4751  247.9         538.3751
Return of the Jedi      309.3062  165.8         475.1062
>
> ## Adding a column for the Worlwide box office
> box.office.all2 <- c( 474.5, 552.5, 310.7, 338.7, 380.3, 468.5 )
> star.wars.matrix2 <- matrix(box.office.all2, nrow=3, byrow=TRUE)
>
> all.wars.matrix <- rbind(star.wars.matrix, star.wars.matrix2)
>
> # Show
> all.wars.matrix
                          US non-US
A new hope              460.9980  314.4
The empire strikes back 290.4751  247.9
Return of the Jedi      309.3062  165.8
                        474.5000  552.5
                        310.7000  338.7
                        380.3000  468.5
>
> ## Selection of matrix elements
>
> all.wars.matrix[1,2]
[1] 314.4
> all.wars.matrix[1:3,1:2]
                          US non-US
A new hope              460.9980  314.4
The empire strikes back 290.4751  247.9
Return of the Jedi      309.3062  165.8
> all.wars.matrix[,1]
          A new hope The empire strikes back      Return of the Jedi
            460.9980              290.4751              309.3062

            474.5000              310.7000              380.3000
> all.wars.matrix[1,]
    US  non-US
460.998 314.400
>
```

```
> #Calculate the average
> non.us.all  <- mean( star.wars.matrix[,2] )
> non.us.some <- mean( star.wars.matrix[1:2,2] )
>
> # Print to console both averages:
> non.us.all
[1] 242.7
> non.us.some
[1] 281.15
>
> #Arithmetic
> visitors <- star.wars.matrix/5;
>
> # Show
> visitors
                          US non-US
A new hope              92.19960  62.88
The empire strikes back 58.09501  49.58
Return of the Jedi      61.86124  33.16
```

Conclusion:-

_____

_____