

**CSE 240 Data Science with R**

**STUDENT WORK BOOK**

|  |  |  |
| --- | --- | --- |
| **Name** | **:** | M SIVANT |
| **Unique ID** | **:** | E00119004 |
| **Year** | **:** | II |
| **Quarter** | **:** | Q6 |
| **Department** | **:** | B.Tech CSE (CyS & IoT or AI &ML) |
| **Faculty Name** | **:** | Prof.B.Nirmala or Prof.N.Chiranjeevi |
| **Academic Year** | **:** | 2020-2021 |

**Date: 02-11-2020**

**Questions:**

1.Consider 2 vectors c(9,10,11,12) and c(13,14,15,16). Create a 4 by 2 matrix from these two vectors

2.Write an R program to take input from the user (user ID and Group/Branch) and display the values

3. Create a data frame Write a R program to create a data frame from four given vectors. a name b. Subject C. Score d. Rank

4.Write a R program to get the statistical summary and nature of the data of a given data frame. (use 3rd Question dataframe)

5. Write a R program to extract specific column from a data frame using column name

6. Write a R program to extract first two rows from a given data frame

**Program:**

# question 1

cat("Creation of Matrix:\n\n")

a <- c(9,10,11,12)

b <- c(13,14,15,16)

matrix\_4\_by\_2 <- matrix(data = c(a,b), nrow = 4, ncol =2)

print(matrix\_4\_by\_2)

# question 2

cat("\nGetting input from the user:\n\n")

id = readline("Enter userID : ")

batch = readline("Enter batch : ")

cat(id, batch,'\n')

# question 3

cat("\nCreating a dataframe:\n\n")

name = c("Siva","King","Star")

subject = c("AI & ML","Python","IOT")

score = c(19,20,18)

rank = c(2,1,3)

df = data.frame(name, subject, score, rank)

print(df)

# question 4

cat("\nDisplaying summary:\n\n")

print(summary(df))

# question 5

cat("\nDisplaying name column of the dataframe:\n\n")

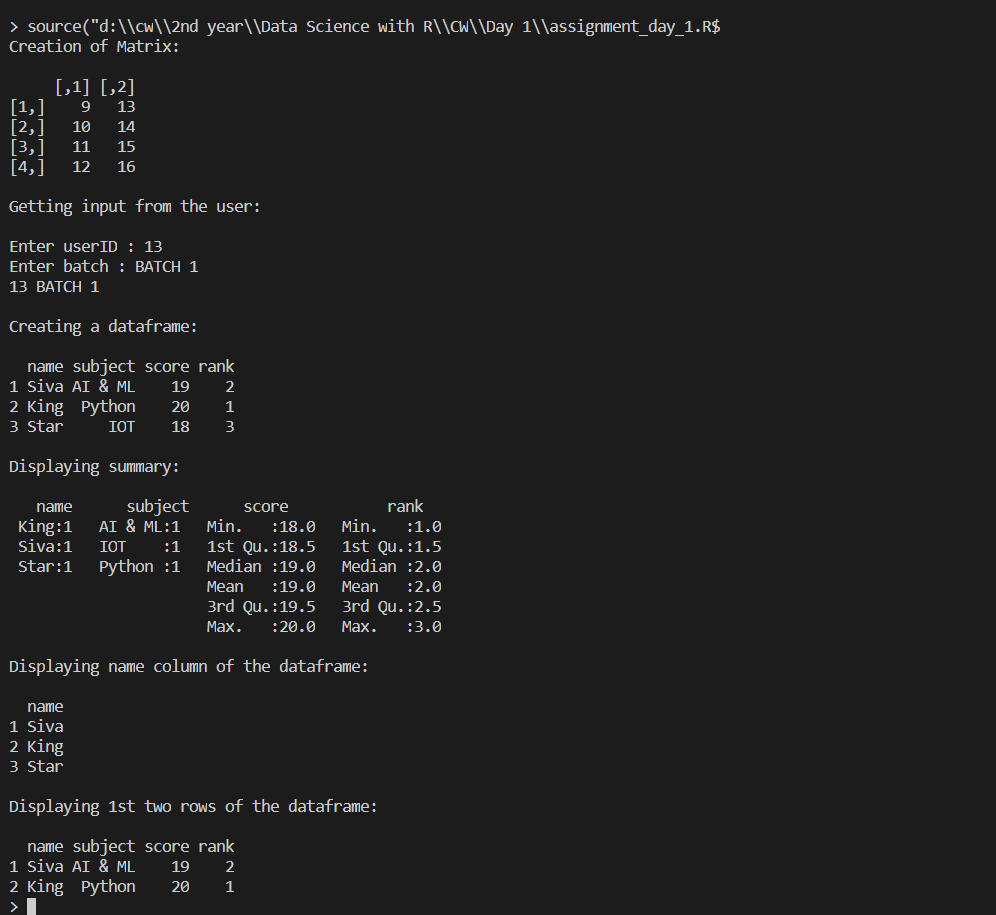
print(df['name'])

# question 6

cat("\nDisplaying 1st two rows of the dataframe:\n\n")

print(df[1:2,])

**Output:**



**Explanation: Concept or Program**

1. Vector:

Vector is a collection of elements

**Syntax** = c (val1, val2, val3, …, valn)

1. Matrix:

Matrices are the R objects in which the elements are arranged in a two-dimensional rectangular layout.

**Syntax** = matrix(data, nrow, ncol, byrow, dimnames)

1. DataFrame:

A data frame is a table or a two-dimensional array-like structure in which each column contains values of one variable and each row contains one set of values from each column.

**Syntax** = data.frame(column1, column2, …, columnn)

1. Summary:

The statistical summary and nature of the data can be obtained by applying summary() function.

**Syntax** = df.summary()

1. as.numeric(val):

as.numeric() is used to convert variable val to numberic type

1. print():

print() is used to display a message or value stored in a variable

1. cat():

cat () is used to concatenate 2 or more messages or values stored in a variable and finally display the values at the console

1. slicing:
   1. df['name']: display specific column in a data frame
   2. df[1:2,]: display first two columns in a data frame

**Date:**

**Q.NO : Question**

**Program:**

**Output:**

**Explanation:**