|  |
| --- |
| **Basic Data Structures** |
| 1. Mention the use cases of following functions in R :  c(),is.numeric(),as.character(),rm(),setwd(),getwd(),class(),str()  print(),,names(),dimnames(),matrix(),list(),data.frame() |
| 2. What is the difference between a matrix and data frame ? |
| 3. What is the concept of "coercion" and when does it happen in R ? |
| 4. What is the difference between "basic data types" and "data objects" in R ? |
| 5. What are the 4 main types of "data objects" in R ? |
| 6. Create a list of 2 vectors of different types (e.g. a character vector and a numeric vector) ? |
| 7. Create a vector with ("Male","Female","Male","Female","Male","Male","Female","Female") as input. |
| Convert it into a factor format. |
| 8. Create a dataframe containing 2 vectors: Vector 1: "hard Disk","Laptop Screen" and |
| Vector 2:"ID101","ID201". |
| 9.After creating the above data frame, access the element in the 2nd row and 1st column. |
| 10. Create two different 4\*3 matrix and fill the values column wise and row wise. ((Hint : Use of matrix(),byrow) |
| 11. What is the importance of the "byrow" parameter in the matrix() function ? |
| 12. What is the difference between "[]"(single square bracket) & "[[]]" (double square bracket) ? (Hint: Take the list example) |

|  |
| --- |
| **Data Manipulation (data to be used: campaign\_data.csv)** |
|  |
| 1. Perform null value treatment for the given dataset . (using na.strings while reading the dataset campaing\_data.csv) |
| 2. Create three subsets(Young,middle,old) of the original dataset based on age. "Young" - <30,"Middle" - 30-50, |
| "Old"- 50+ . |
| 3. List the number of users who are contacted over telephone for more than |
| 500 seconds (Hint: Use column duration) |
| 4. Identify the user with the highest contacted duration in |
| "Young" group. (Hint" Use max function in the Young data set) |
| 5. Identify the number of subscribed clients who have both the housing and |
| the personal loan (Hint: Use column "Housing" & "loan") |
|  |

|  |
| --- |
| **Table function and loop** |
| 1. Get the frequency distribution of different job types (Hint: Use column job, use table() function to get frequency) |
| 2. Create a vector of numbers 1:10 using a loop. |
| 3. Read ten numbers from the user and store them as a vector. |
| 4. Create a list of numbers 1:10 using a loop. |

|  |
| --- |
| **Aggregate function** |
| 1. Find the average age for each job type |
| 2. Find the average call duration of clients for each month and and education type |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **User Defined functions (dataset: campaign\_data.csv)** | | | | | | | | | | |
| 1. Create a function with user job type as input and the average age of that group as output | | | | | | | |  |  |  |
| Hint: The argument getting passed in the function can be any job type e.g. unemployed,services,blue-collar etc. | | | | | | | | | |  |
| And corresponding to the job type, the function should return the average age for that job type | | | | | | | | |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| **String manipulation** | | | | | | | | | | |
| 1. Concatenate column "Job" and "education" to get a new column called "Job\_Education". | | | | | | | |  |  |  |
| 2. What is the difference between gsub & chartr ? | | | | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| **Apply functions** | | | | | | | | | | |
| 1. Subset the Campaign\_dataset for numerical columns. (Hint: Use is.numeric() with sapply()) | | | | | | | |  |  |  |
| 2. Identify the count of unique values of each column of the Campaign\_datasets (Hint: Use unique() with sapply) |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| **Plots** | | | | | | | | | | |
| 1. Plot the number of customers for each education type. (Plot a barchart) | | | | | | |  |  |  |  |
| 2. Plot a boxplot of the duration of the client call . (Box plot of duration column) | | | | | | |  |  |  |  |
| 3. Plot the scatter plot for number of customers for each contact type during each month (Scatter plot between "Contact" and "month" column.) | | | | | | | | | | |
| 4. Draw a histogram for the "duration" column to understand the frequency distribution  of the same. | | | | | | | |  |  |  |