R FOUNDATION PRACTICE QUESTIONS (Not for Submission)

Ground Floor, SCO 382
(Next to Iffco Chowk Metro station)
Sector 29,
Gurgaon, Haryana

First Floor, Bldg-41,
14th Main Road
Sector 7, HSR Layout,
Bengaluru, Karnataka

1. Use R code to get a description of the version of R and its attached packages used in the current session?
2. Create an object that has the list of all the installed packages in your system.
3. Create an object with the following data types and check their data types by using the class function after creation
   a. integer
   b. numeric
   c. character
   d. logical
   e. complex
4. Create the following objects:
   object_1 <- “my_text”
   object_2 <- “123.456”
   object_3 <- 123.456
   object_4 <- 0
   object_4 <- FALSE

Perform type casting on each of the object. (Convert each of the objects into different available data types and analyze the results).

5. List all the objects in the current session.
6. Convert the data “1/December/1998” into a Date type.
7. Find the number of years from the Independence of India till present.
8. Convert the following text into a Date format:
   “On the 16th of December the hostilities between the two countries ended and a new nation was formed in the winter of 1971” (Hint: Use lubridate)
9. Convert the string “December-31-2018” to a date format and extract the month. (Hint: use lubridate)
10. Create a vector with 5 numbers and find the data type of it.
11. Create a vector with 5 numbers and 1 text and find the data type of it.
12. Create a vector with number from 1 to 1000 with a step of 10. (Hint: use seq())
13. Create a vector V1 having numbers from 11 to 20. (Hint: use n:m)
14. Extract the 5th, 7th and 9th element of the vector v1 and save it in an object V2.
15. Save the first 9 elements of the vector V1 in a separate object naming ‘V3’.
16. Create a vector “Airport.Codes” with the following information:
    DEL, BOM, PUN, MAA, DEL, MAA, PUN, DEL, MAA, PUN, PUN, DEL, BOM
17. PUN is the incorrect airport code and it should be PNQ. Replace the values in the vector Airport.Codes.

18. Create vector “x” with the values {4, 4, 5, 6, 7, 2, 9}.
    a. Calculate the number of observation (n), mean, sum, max, min, variance of the vector ‘x”.
    b. Also print the 3rd element, odd positions elements and elements from 2 to 6 positions.
19. Create a vector ‘Age’ with values from 1 to 99.
20. Create a vector ‘Young.Old’ by extracting values from 1 to 18 and 80 to 99 from Age Vector.
21. Find the number of elements in the vector Young.Old.
22. Find the sum, mean, median, min and max age in the vector Young.Old.
23. Create 6x4 matrix (6 rows and 4 columns) using 1 to 24 numbers
24. Create a vector ‘for.m’ which has 30 numeric values. Use this vector to create a matrix my_matrix with 10 rows.
25. Extract the 1st, 5th and 9th row of the 1st and 2nd column from the matrix my_matrix.
26. Create data frame with the below vectors
    StoreID - (111, 208, 113, 408)
    Tenure - (25, 34, 28, 52)
    StoreType - ("Type1", "Type2", "Type1", "Type1")
    status - ("Poor", "Improved", "Excellent", "Poor")
27. Print the data in different programs.
    a. only storeId, tenure
    b. only storetype and status
    c. only tenure
28. Create a data.frame ‘Employee.Info’ by combining vectors. The table should have 4 columns with 5 rows. The columns should be – ‘Employee.ID’ , ‘Employee. Name’, ‘Employee.Salary’ , ‘Employee.Department’. (Provide arbitrary values for the rows).
29. Create different factors using the below vectors and label the values and order levels for factors “outcome” and “status”.
    a. ethnicity - White", "African amrican", "White", "Asian"
    b. status - Poor", "Improved", "Excellent", "Poor"
    c. outcome - c(1, 3, 2, 4, 3, 1, 1) – labels- “Poor", "Average", "Good", "Excellent"
30. Create list called “mylist” with the title "My First List" and with the below objects called “ages” with below h, j and k. Also print the different combinations of objects.
    h –numeric vector with the values 25, 26, 18, 39
    j – matrix with 5 rows and 2 columns with the values 1 to 10
    k – character vector with the values "one", "two", "three"

Basic Exploratory Data Analysis

31. Import the stores.csv dataset and save it as df1.
32. Apply below functions on stores data frame and observe the outputs.
    class(df1)
    names(df1)
    length(df1)
    dim(df1)
    str(df1)
    head(df1)
    tail(df1)
    fix(df1)
    summary(df1)
    Hmisc::describe(df1)
33. Import the second sheet of the excel file CarData.xlsx and save it as Car.Info
34. Save the object df1 as an .RData file in the hard drive (df1.RData).
35. Save all the objects created so far in a .RData File. Name it as My_Backup.
36. Delete df1 and Car.Info from the global environment.
37. Find the names of all the current objects in the global environment and save this information in an object 'Current_objects_in_GE'.
38. Delete all the objects available in the global environment.
39. Import the df1.RData file from the hard drive.
40. Find the summary details of df1 (Hint: use summary() or Hmisc::describe()).

Structure Based Data Manipulation

41. Find details regarding the structure of df1 (Hint: use str())
42. Extract the column Staff_Cnt from df1 and save it in an object 'Staff_Count'.
43. Add 10 to every value of the vector 'Staff_Count' and name it as 'Updated_ Staff_Count'.
44. Add this vector 'Updated_ Staff_Count' as a column in the data.frame df1.
45. Create an object 'Store_Details' by extracting the following columns from df1 – StoreCode, StoreName, StoreType, Location, OwnStore.
46. Find the class of Store_Details.
47. Change the data type of the column StoreName from factor to character.
48. Rename the column 'AcqCostPercust' to 'ACPC' and 'ProfitPercust' to 'PPC'.
49. Create a column 'MarketingSpend' which has a constant value of 200. Use this column to create another column 'UpdatedOperatingCost' which is OperatingCost + MarketingSpend.
50. Delete the column MarketingSpend.
51. In df1, the UpdatedOperatingCost should be next to the column OperatingCost.
52. In df1, Create new variable store_class as follows.
    If total sales<120 then "Low Perform store"
    If total sales>=120 and total sales<240 then "Average Perform store"
    If total sales>240 then "High Perform store"

Content Based Data Manipulation

53. Find the names of the top 3 SuperMarket stores of Delhi with the highest TotalSales.
54. Find the information such as the StoreName, TotalSales, OperatingCost and Staff_Cnt of the stores where the store is owned and there is an online presence.
55. Check if there are any duplicate rows in df1
56. Check if there are any duplicate rows in df1 on the basis of StoreCode.
57. Create a new dataset that has removed duplicates on the basis of StoreName and StoreCode.
58. How to find missing values in the df1 dataset and recode missing values with 0? Delete the missing values (if there are any) from the data set?
59. Sort the df1 data as follows
    a. Create new dataset (newstores) with sorted data by “Storetype”
    b. Create new dataset (newstores) with sorted data by location(ascending) and totalsales in descending order.
60. Create subsets of data from stores data as following conditions.
    a. Only Columns 5,7, 8,9
    b. Excluding columns 5, 7, 8,9
    c. Selecting first 10 observations
    d. Observations with Storetype= Apparel and totalsales>100
    e. Columns (storecode, storename, location, totalsales) with totalsales between 100 & 300.
    f. Include all columns from Storecode to Basketsize with storetype=Electronics and totalsale>100