

Department of Computer Science & Engineering

QUESTION BANK FOR VII SEMESTER (Term: Aug-Dec 2017)

Data Analytics Laboratory (CSL717)

I.A. Marks : 50

Credits: 0:1:1

Exam Hours: 03

Exam Marks: 50

Develop the following programs using R

1. **Store all dept (9 depts) students names,USN,Dept names, 5 subject marks, grades and SGPA in csv file using ms excel.**
 - a. Extract each dept students names separately.
 - b. Extract S grade scores in all subjects in each dept separately.
 - c. Extract students who have scored at least S grades in any 2 subjects
 - d. Extract students who have scored above 9 SGPA in each dept
 - e. Store students marks numerically, transform into grades and store in new data frame
 - f. Check whether students grades are identical or not in each subject
2. **Store all dept faculty names with designation, number of papers published, training programs attended and salary details.**
 - a. Extract each dept faculty details separately
 - b. Extract Professors of each dept separately
 - c. Extract people who earn more than 1.5 lakh in each dept where their designation is prof, associate or assistant
 - d. Find out the cost of professors in each dept .(sum up their salary to get cost of them)
 - e. Find the individual cost of each dept faculty and average cost of faculty in dept
 - f. Find out Which dept has highest average cost of faculty and which dept has lowest cost of faculty
3. **Store all dept (9 depts) students names,USN,Dept names, 5 subject marks , grades and SGPA in csv file using ms excel.**
 - a. Store students marks numerically, transform into grades and store in new data frame
 - b. Check whether students grades are identical or not in each subject
 - c. Extract students' marks in each subject separately. If the student has scored greater than 80 map it as "good", if it is between (80 and 60) map it as "moderate", if it is between (40 and 60) map it as "need improvement", else map it as "poor".
 - d. Consider dataset given in (a) , map the S,A, grades as " GOOD"; map ,B,C grades as "average", D,E grades as "below average"; 'F' grade as "poor".
 - e. Transform dept names to numerical data.
 - f. Using factor() and mapvalues() convert dept names to numerical data.
4. **Store all dept (9 depts) students names,USN,Dept names, 5 subject marks , grades and SGPA in csv file using ms excel.**
 - a. Create table from student data with USN and names only.
 - b. Display the type of each column.
 - c. Write a function to provide recommendations of students about their performance in the exams like "Good", "Moderate", "Need improvement", etc.,
 - d. Write a function to perform statistical analysis of students data.

HOD, Dept. of CSE

- e. Use supply to perform (c) and (d)
- 5. Store all dept faculty names with designation, number of papers published, training programs attended and salary details.**
 - a. Store faculty salary numerically, transform into factor.
(eg. 50000 to 75000 as 1, 75000 to 100000 as 2 , and so on)
 - b. Check whether faculty paper publication count and number of training program attended are same or not.
 - c. Extract paper published count separately. If the count is greater than 15 map it as “Excellent”, if it is (10-15)map it as “good”, if it is (5 to 10) map it as “ moderate”), if it is (1-5) map it as “need to improve”, else “poor, start your research”.
 - d. Consider the solution of (c). Map “Excellent”, “good” as “ Good performers”;Map “ Moderate” and “need to improve” as “ Ok, Keep it up”. Else “ map to “ You may be fired!”
 - e. Transform designations to numerical data
- 6. Store all dept faculty names with designation, , number of papers published, training programs attended and salary details.**
 - a. Using factor() and mapvalues() convert designations to numerical data.
 - b. Create table for faculty names and designations.
 - c. Write a function to provide recommendation about faculty performance based on the number of papers published and number of training programs attended. Like Good, Moderate, Need improvement .
 - d. Write a function to perform statistical analysis of faculty data to identify faculty performance of depts.
 - e. Use supply to perform all the above functions.
- 7. Store all dept (9 depts) students names,USN,Dept names, 5 subject marks, grades and SGPA in csv file using ms excel.**
 - a. Plot , in each dept, how many students have scored above 9 SGPA
 - b. Create subset of students , who have scored S grade in any subject and failed in any subject.
 - c. Find out average SGPA of each dept students and find out average score of each subject for each dept.
 - d. Extract 10 toppers of each dept.
 - e. Sort students details of each dept separately.
 - f. Search for a particular student name in the data set, and retrieve his/her details.
- 8. Store all dept faculty names with designation, number of papers published, training programs attended and salary details.**
 - a. Plot, in each dept, how many faculty are earning more than 1 lakh.
 - b. Create subset of faculty, who have published more than 10 papers and their designation is Associate professor.
 - c. Find out average papers published by each dept, designation wise and Find out Average training programs attended by faculty each dept wise.
 - d. Extract top 3 performers among faculty dept wise(who have more papers published and more training programmes attended)
 - e. Sort faculty details ,dept wise separately.
 - f. Search for faculty name in the data set and retrieve his/her details.

- 9. Store all dept faculty names with designation, number of papers published, training programs attended and salary details.**
- Change the column names of faculty data set.
 - Use `map values()`, `as.factor()` and `transform()` to change the designation column to have numerical values. 1- Prof, 2-Asso.Prof, 3-Asst.Prof; to change gender column 1-Male, 2-Female
 - Using `with()` and `tapply()`, calculate the mean training programs attended and no. of papers published in each department. Format it for markdown.
 - Using `with()` and `aggregate()`, calculate the mean training programs attended and no. of papers published in each department. Format it for markdown.
 - Check whether the mean value of no. of papers published in depts. are influenced by training programs attended and designation.
 - Do regression analysis using `aggregate()` for faculty performance.
- 10. Store all dept faculty names with designation, , number of papers published, training programs attended and salary details.**
- Create the table output for designation and papers published. Use `with()` and `table()`.
 - Find the odds of lower no. of paper published with respect to designation and no. of training programs attended
 - Is the designation affects the training programs attended? Check it with the data. Prove it.
 - Find the correlation of papers published and training programs attended using `with()`, `cor()`.
 - Find the correlation of training programs attended and designation using `with()`, `cor()`.
 - Using `by()` combine the operations of above questions (d) and (e) using function. And do the correlation analysis using `cor()` with in the function.
- 11. Store all dept faculty names with designation, number of papers published, training programs attended and salary details.**
- Plot average training programs attended against designations(only 3 designations) of the institution using `plot()`
 - Plot average papers published against designations of the institution using `plot()`.
 - Change the x axis, color and y-axis labels respectively. Add legends.
 - Plot the graphs of (a) and (b) using `with()` and `plot()`.
 - Draw scatter plot for above questions. Draw box plot for above question. Draw bar plots for above questions. Draw single variable plots for above questions.
 - Plot the prof, asst.prof and asso.prof average performance in different colors using `rep()`, `colorpalatte()` functions. Represent each designation average performance by different symbols.
- 12. Store all dept (9 depts) students names,USN,Dept names, 5 subject marks , grades and SGPA in csv file using ms excel.**
- Make some string entries in student marks. Make some numerical entries in names.
 - Using `gsub` remove character data in marks column. Using `gsub` remove numerical data in name column
 - Clean the name and marks column and put it in a new student data set variable using `transform` function
 - Use `table` to get summary of student data
 - Use `sapply` to perform the cleaning of data mentioned above
 - Use `lapply` to perform the cleaning of data mentioned above.
- 13. Store all dept (9 depts) students names,USN,Dept names, 5 subject marks , grades and SGPA in csv file using ms excel.**
- Display the summary using `summary()`
 - Include gender details for students as F/M in csv file.
 - Attach Mr or Ms. For each student, using paste command . Display all students details.
 - Define user defined functions to perform the above operations.

- e. Use while loop, for loop to access students marks and find the grades. Put this with in a Function
- f. Use apply, sapply, lapply and tapply to perform the above operation over all the columns of students data set. Use with() function to apply the above operation over students dataset. Use any() function to apply the above operation over students dataset.

14. Store all dept (9 depts) students names, USN, Dept names, 5 subject marks , grades and SGPA in csv file using ms excel.

- a. Change the column names of Student data set.
- b. Use map values() , as.factor() and transform ()
- c. To change the Grade column(S,A,B,etc) to have numerical values. 1- S, 2-A, 3-B,etc.;
- d. To change gender column 1-Male, 2-Female
- e. Using with() and tapply() , calculate the mean of marks in each subject scored by students in each department and mean of CGPA of students in each dept with respect to gender. Format it for markdown.
- f. Using with() and aggregate() , calculate the mean of marks in each subject scored by students in each department and with respect to the gender of students in each dept. Format it for markdown.

Note:

Conduction and Result	: 30 Marks (Write up: 10 Marks, Execution: 20 Marks; All subdivisions a to f , carries equal marks)
Viva	: 10 Marks
Mini Project Evaluation	: 10 Marks
For Change of question	: 5 Marks deduction