SRM VALLIAMMAI ENGINEERING COLLEGE

SRM Nagar , Kattankulathur – 603 203

DEPARTMENT OF
ARTIFICIAL INTELLINGENCE AND DATA SCIENCE

QUESTION BANK

V SEMESTER
1922502 – R PROGRAMMING IN DATA SCIENCE

Regulation – 2019
Academic Year 2022 – 2023 (ODD)

Prepared by

Mrs. R. Deepa, Assistant Professor / AI&DS

SRM VALLIAMMAI ENGINEERING COLLEGE

SRM Nagar, Kattankulathur – 603203.

DEPARTMENT OF
ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

QUESTION BANK

SUBJECT : 1922502 - R PROGRAMMING IN DATASCIENCE
SEM/YEAR : V / III

UNIT – I

Introduction to R: R Software, R packages, Data Types in R: Scalars, Vectors, Matrices, Data frames, Lists, Variables and Logical Operations. R Matrix Create, Print, Column, Slice, Factors in R, Categorical and Continuous Variables.

PART- A (2 Marks)

Q. No	Questions	BT Level	Competence
1	Define R programming.	Remembering	BTL-1
2	Differentiate between Scalars, vector, list, Matrix and Data frame.	Understanding	BTL-2
3	List out any five features of R.	Remembering	BTL-1
4	Differentiate between R and Python in terms of functionality.	Understanding	BTL-2
5	What are the applications of R?	Understanding	BTL-2
6	Why do we use the command-install.packages(file.choose(),repos=NULL)?	Analyzing	BTL-4
7	Summarize some packages in R, which can be used for data imputation?	Evaluating	BTL-5
8	How to get the name of the current working directory in R?	Applying	BTL-3

9	Write a R program to take input from the user (name and age) and display the values. Also print the version of R installation	Analyzing	BTL-4
10	What are the different values that can be assigned to a numeric datatype in R?	Analyzing	BTL-4
11	What are the different data types in R?	Analyzing	BTL-4
12	Explain RStudio.	Evaluating	BTL-5
13	Compare R with other technologies.	Understanding	BTL-2
14	Write a R program to create three vectors numeric data, character data and logical data. Display the content of the vectors and their type.	Evaluating	BTL-5
15	Define Merging and accessing list elements.	Remembering	BTL-1
16	Demonstrate the simple 3X3 matrix.	Applying	BTL-3
17	How do you access the elements in the 2^{nd} column and 4^{th} row of a matrix?	Applying	BTL-3
18	Define slice a matrix.	Remembering	BTL-1
19	How to create a Matrix?	Applying	BTL-3
20	Write R program to create a blank matrix.	Creating	BTL-6
21	Write a program to add two matrices.	Creating	BTL-6
22	List out the operations on Matrices.	Remembering	BTL-1
23	Define order of a Matrix.	Remembering	BTL-1
24	Difference between Nominal and ordinal categorical variable.	Understanding	BTL-2

PART – B (13 Marks)

1	i.Summarize the advantages and disadvantages of R? ii.Explain scalar and vector with an example.	**(6)** **(7)**	Remembering	BTL-1
2	i.Write a R program to get the first 10 Fibonacci numbers. ii.Generate the following: a. Access the element at 3^{rd} column and 1^{st} row in a matrix. b. Access only the second row	**(6)** **(7)**	Remembering	BTL-1

	c. Access the element at 2nd column and 4th row in a matrix			
3	Write a R program to find the maximum and the minimum value of a given vector. Explain the functions with syntax.	**(13)**	Applying	BTL-3
4	Write a R program to find elements which are present in two given data frames	**(13)**	Applying	BTL-3
5	Write a R program to create a data frame using two given vectors and display the duplicated elements and unique rows of the data frame. Explain with a syntax.	**(13)**	Understanding	BTL-2
6	i. Illustrate the usage of all logical operator in R. ii. Explain the use of length () and mean() function.	**(6)** **(7)**	Understanding	BTL-2
7	Elaborate the statistical and programming features of R.	**(13)**	Understanding	BTL-2
8	i. Write a R program to add a new item g4 = "Python" to a given list. g1=1:10,g2=”R Program”,g3=”HTML”. ii. Explain Data frame operations.	**(6)** **(7)**	Evaluating	BTL-5
9	Write a R program to add 3 to each element of the first vector. Print the original and new vector.	**(13)**	Analyzing	BTL-4
10	Write a R program to reverse the order of given vector.	**(13)**	Analyzing	BTL-4
11	Check whether the value of the element of a given vector greater than 10 or not. Return TRUE or FALSE.	**(13)**	Understanding	BTL-2
12	Write a R program to create an ordered factor from data consisting of the names of months.	**(13)**	Analyzing	BTL-4
13	Write a R program to create a correlation matrix from a data frame of same datatype. Explain the functions with syntax.	**(13)**	Applying	BTL-3
14	Create the following:		Creating	BTL-6

	a. Create a matrix taking a given vector of numbers as input. Display the matrix. b. To access the element at 3rd column and 2nd row, only the 3rd row and 4th column of a given matrix.	**(7)** **(6)**		
15	List out the properties of the following: a. Matrix subtraction b. Matrix Division c. Matrix addition d. Matrix multiplication	**(4)** **(4)** **(3)** **(2)**	Remembering	BTL-1
16	i. What is Factor in R and its function? ii. Distinguish two types of variables with an example.	**(6)** **(7)**	Remembering	BTL-1
17	Explain categorical variables with an example.	**(13)**	Evaluating	BTL-5
	PART – C (15 Marks)			
1	i. Explain main features to write R code that runs faster. ii. Difference between package and library. With examples	**(8)** **(7)**	Evaluating	BTL-5
2	a. i.Let's create the following vectors: u <- 4 v <- 8 Use the elementary arithmetic operators +, -, *, /, and ^ to: • add u and v • subtract v from u • multiply u by v • divide u by v • raise u to the power of v ii. Write a R program to create a vector and find the length and the dimension of the vector.	**(7)** **(8)**	Creating	BTL-6

	b. i. Suppose u and v are not scalars, but vectors with multiple elements: u <- c(4, 5, 6) v <- c(1, 2, 3) Without using R, write down what you expect as the result of the same operations as in the previous exercise: • add u and v • subtract v from u • multiply u by v • divide u by v • raise u to the power of v ii. Create a Vector using : Seq() function iii. Write R program to find Sum, Mean and Product of a vector, ignore elements like NA or NaN.			
3	a. Write a R program to get all prime numbers up to a given number b. Write a R program to count the number of NA values in a data frame column. c. Create the following a. Creating a list b. Naming list elements c. Check whether a item exist or not	**(5)** **(5)** **(5)**	Evaluating	BTL-5
4	Perform the following operation in data frame: a. Write a R program to add a new column in a given data frame. b. Write a R program to add new row(s) to an existing data frame. c. Write a R program to drop column(s) by name from a given data frame. d. Write a R program to drop row(s) by number from a given data frame.	**(4)** **(4)** **(4)** **(4)**	Evaluating	BTL-5

	e. Write a R program to create inner, outer, left, right join(merge) from given two data frames	**(4)**		
5	Create the following: a. Create factor variables b. Create ordered factor variables c. Adding and dropping levels in factor variable	**(5)** **(5)** **(5)**	Creating	BTL-6

UNIT – II R DATA STRUCTURES

Scalars -Vectors Matrices - List - Data Frames-Factors -Packages - Data Reshaping –Data management with repeats, sorting, ordering and lists - Vector indexing, factors, Data management with strings, display and formatting.

PART – A (2-Marks)

Q. No	Questions	BT Level	Competence
1	Define R vector?	Understanding	BTL-2
2	Define R lists?	Understanding	BTL-2
3	To reverse the order of given vector using R vector.	Remembering	BTL-1
4	Difference between data frame and a matrix in R?	Analyzing	BTL-4
5	List out the various forms of reshaping data in a data frame.	Remembering	BTL-1
6	Examine why R- data reshaping is important?	Applying	BTL-3
7	Define Transpose of a matrix.	Understanding	BTL-2
8	Define melt() and cast() function.	Understanding	BTL-2
9	Define Tidyr package.	Understanding	BTL-2
10	Identify the use of sort () function.	Remembering	BTL-1
11	List out the various sorting mechanisms.	Remembering	BTL-1
12	Explain factor variable?	Evaluating	BTL-5
13	Explain the recycling of elements in an R vector? Give an example.	Evaluating	BTL-5
14	Differentiate vector index and Negative index.	Analyzing	BTL-4
15	Analyze what is meant by out-of-range index?	Analyzing	BTL-4

16	Explain the use of length () function?	Evaluating	BTL-5
17	Point out the attributes of a factor.	Analyzing	BTL-4
18	Show how to count the number of NA values in a data frame column	Remembering	BTL-1
19	Convert a matrix to a 1-dimensional array using Rcode.	Applying	BTL-3
20	How will you read a .csv file in R language?	Applying	BTL-3
21	Convert a given pH levels of soil to an ordered factor using R Code.	Applying	BTL-3
22	Write a R program to get the structure of a given data frame	Creating	BTL-6
23	Write a R program to get the length of the first two vectors of a given list. g1=1:10,g2="R Program",g3="HTML".	Creating	BTL-6
24	Show the R code to Add 10 to each element of the first vector in a given list. g1=1:10,g2="R Program",g3="HTML".	Remembering	BTL-1

PART- B (13 Marks)

1	Explain in detail about data frame with example R code.	**(13)**	Remembering	BTL-1
2	Demonstrate an R code to find the factorial of a number (use recursion)	**(13)**	Applying	BTL-3
3	Explain list data structure and its operations with examples.	**(13)**	Remembering	BTL-1
4	Create a simple data frame from 3 vectors. Order the entire data frame by the first column.	**(13)**	Understanding	BTL-2
5	i.Explain about how to create a list in R with an example? ii.Explain how to access list element? iii. Explain how to operate on lists in R?	**(7)** **(6)**	Remembering	BTL-1
6	Convert the following multi-line operations to a single expression. Check that both approaches give the same result. Part a: w<- u + v w <- w / 2	**(13)**	Evaluating	BTL-5

	w <- w + u Part b: w1 <- u^3 w2 <- u - v w <- w1 / w2			
7	i. Define Factor. ii. How to create a factor and how to access components of a factor? iii. How to modify a factor?	**(4)** **(5)** **(4)**	Understanding	BTL-2
8	Explain with an example in changing the orders of levels?	**(13)**	Remembering	BTL-1
9	Summarize the functions to join columns and rows in a data frame.	**(13)**	Understanding	BTL-2
10	Show how to code R program to create a Data frames which contain details of 5 employees and display the details	**(13)**	Applying	BTL-3
11	Write a R code to check the following: a. Check available R packages b. Get the list of all packages installed c. Install new package d. Install package manually	**(4)** **(4)** **(3)** **(2)**	Creating	BTL-6
12	Write a R program to create a matrix taking a given vector of numbers as input and define the column and row names. Display the matrix.	**(13)**	Evaluating	BTL-5
13	Illustrate a R code using the following functions: seq() , paste(), print(), format(), mode(), order()	**(13)**	Analyze	BTL-4
14	Discuss any three commonly used packages with an example.	**(13)**	Understanding	BTL-2
15	Sketch out some popular repositories for R package.	**(13)**	Applying	BTL-3
16	Illustrate R program to create the system's idea of the current date with and without time. Explain with a syntax.	**(13)**	Analyzing	BTL-4

17	Illustrate R program to create two 2x3 matrix and add, subtract, multiply and divide the matrixes.	**(13)**	Analyzing	BTL-4
	PART – C(15 Marks)			
1	Create the vectors: (a) (1, 2, 3, . . . , 19, 20) (b) (20, 19, . . . , 2, 1) (c) (1, 2, 3, . . . , 19, 20, 19, 18, . . . , 2, 1) (d) (4, 6, 3) and assign it to the name tmp. For parts (e), (f) and (g) look at the help for the function rep. (e) (4, 6, 3, 4, 6, 3, . . . , 4, 6, 3) where there are 10 occurrences of 4. (f) (4, 6, 3, 4, 6, 3, . . . , 4, 6, 3, 4) where there are 11 occurrences of 4, 10 occurrences of 6 and 10 occurrences of 3. (g) (4, 4, . . . , 4, 6, 6, . . . , 6, 3, 3, . . . , 3) where there are 10 occurrences of 4, 20 occurrences of 6 and 30 occurrences of 3.	**(2)** **(2)** **(2)** **(2)** **(2)** **(2)** **(3)**	Creating	BTL-6
2	i. Explain operations on vectors. ii. Write R program to check a given number is Even or Odd.	**(7)** **(6)**	Evaluating	BTL-5
3	Explain R function for differentiation and integration with an example?	**(15)**	Evaluating	BTL-5
4	i.Consider two vectors u and v: u <- c(8, 9, 10) v <- c(1, 2, 3) Create a new vector w in a single line of code: w <- (2 * u + v) / 10 or carry out each operation on a separate line: w <- 2 * u	**(7)**	Creating	BTL-6

	w <- w + v w <- w / 10 ii.Convert the following expressions to separate operations, and check that both approaches give the same result: w <- (u + 0.5 * v) ^ 2 w <- (u + 2) * (u - 5) + v w <- (u + 2) / ((u - 5) * v)	**(6)**		
5	i.Create a simple data frame from 3 vectors. Order the entire data frame by the first column. ii.Create a data frame from a matrix of your choice, change the row names so every row says id_i (where i is the row number) and change the column names to variable_i (where i is the column number). I.e., for column 1 it will say variable_1, and for row 2 will say id_2 and so on.	**(5)** **(10)**	Creating	BTL-6

UNIT – III -DATA PREPARATIONS

R Data Frame: Create, Append, Select, Subset. R sort a data Frame using Order (), R Dplyr: Data manipulation and Cleaning, Merge Data Frames in R: Full and Partial Match, Functions in R programming.

PART-A (2 Marks)

Q. No	Questions	BT Level	Competence
1	List out the characteristics of a data frame.	Remembering	BTL-1
2	Define the structure of a data frame using str () function.	Applying	BTL-3
3	What is the use of nrow () function?	Understanding	BTL-2
4	What is the use of subset() function?	Understanding	BTL-2
5	Write R code to select a column of a data frame.	Analyzing	BTL-4
6	List down the methods to sort a data frame.	Remembering	BTL-1

7	What is the use of order() function.	Applying	BTL-3
8	What is dplyr() function?	Applying	BTL-3
9	List out the performance of R -dplyr package.	Remembering	BTL-1
10	How to install and load dplyr package?	Understanding	BTL-2
11	What is the use of rename() and filter() function?	Analyzing	BTL-4
12	What is the use of with () and by () functions in R?	Analyzing	BTL-4
12	What is the use of summarise () function with synax?	Analyzing	BTL-4
13	List out the common symptoms of messy data.	Remember	BTL-1
14	Write a R code to remove empty rows and columns.	Creating	BTL-6
15	How to handle missing value in R?	Understanding	BTL-2
16	Write down the syntax of grep() function and it use?	Creating	BTL-6
17	List out the function components.	Remember	BTL-1
18	What is the use of return () function?	Understanding	BTL-2
19	List out some built-in functions.	Remembering	BTL-1
20	Compare the types of functions in R programming?	Evaluating	BTL-5
21	Explain Argument matching.	Evaluating	BTL-5
22	Explain Lazy evaluation.	Evaluating	BTL-5
23	Analyze on how library () and require () functions are used.	Analyzing	BTL-4
24	List out the characteristics of a data frame.	Applying	BTL-3

PART-B (13 Marks)

1	Show how to extract data from data frame. Explain with an example.	**(13)**	Applying	BTL-3
2	Explain how to append rows to R data frame with an example?	**(13)**	Evaluating	BTL-5
3	Analyse R program to create dataframe with 2 columns and order based on particular columns in decreasing order. Displayed the Sorted dataframe based on subjects in decreasing order, displayed the Sorted dataframe based on rollno in decreasing order	**(13)**	Analyzing	BTL-4
4	List out the dplyr function and its equivalent SQL with an example.	**(13)**	Remembering	BTL-1

5	List out the functions to select variables based on their names.	**(13)**	Remembering	BTL-1
6	List out the purpose of data cleaning in R with an example.	**(13)**	Remembering	BTL-1
7	Explain about Data Manipulation with dplyr package	**(13)**	Evaluating	BTL-5
8	With the dataset swiss, create a data frame of only the rows 1, 2, 3, 10, 11, 12 and 13, and only the variables Examination, Education and Infant.Mortality. a) The infant mortality of Sarine is wrong, it should be a NA, change it. b) Create a row that will be the total sum of the column, name it Total. c) Create a new variable that will be the proportion of Examination (Examination / Total)	**(5)** **(5)** **(3)**	Creating	BTL-6
9	Show how to clean the column names of a data frame using R Programming with an example.	**(13)**	Applying	BTL-3
10	Explain the following a. rbind() to merge two R data frames b. cbind() to merge two R data frames c. merge()	**(5)** **(5)** **(3)**	Understanding	BTL-2
11	Analyze R code for the following: a. Find partial match in a specific column b. Find several partial matches	**(7)** **(6)**	Analyzing	BTL-4
12	Write the syntax for writing functions in R with a sample program.	**(13)**	Analyzing	BTL-4
13	Write the R code for the following a. Calling a function with default arguments b. Calling a function with arguments c. Calling a function without arguments	**(5)** **(4)** **(4)**	Applying	BTL-3
14	Summarize the features of R function?	**(13)**	Understanding	BTL-2

	Explain the following: a. Full Match b. Partial match			
15	Summarize the functions which helps in importing data from other applications in R. With an example.	**(13)**	Understanding	BTL-2
16	List out the commonly used functions in dplyr package.	**(13)**	Remembering	BTL-1
17	i.Explain the following example for writing a function a. Throwing a die ii.Write a short note on: a.Data Manipulation b.Data Cleaning	**(13)**	Understanding	BTL-2
	PART-C (15 Marks)			
1	Develop the R code for the following: a. Subset data frame by selecting columns b. Subset data frame by excluding columns **c.** Subset data frame by selecting rows	**(15)**	Creating	BTL-6
2	Create the dataframe data <- data.frame(x1 = 1:6, x2 = c(1, 2, 2, 3, 1, 2), x3 = c("F", "B", "C", "E", "A", "D")) Use the following functions a. Arrange function b. Filter function c. Mutate function d. Pull function e. Rename function f. Sample)n function g. Select function	**(5)** **(1)** **(1)** **(1)** **(2)** **(2)** **(2)** **(1)**	Creating	BTL-6
3	Explain with a Sample (Dummy) Data in R and perform data manipulation with R.	**(15)**	Evaluating	BTL-5

4	Explain in detail about math function in R with an example each?	**(15)**	Evaluating	BTL-5
5	i.Create a function that will return the sum of 2 integers. ii.Create a function that given a vector will print by screen the mean and the standard deviation, it will Optionally also print the median.	**(7)** **(6)**	Creating	BTL-6

UNIT – IV – DATA FRAMES

Data frames, import of external data in various le formats, statistical functions, compilation of data - Graphics and plots, statistical functions for central tendency, variation, skewness and kurtosis, handling of bivariate data through graphics, correlations, programming and illustration with examples

PART-A (2 Marks)

Q. No	Questions	BT Level	Competence
1	Discuss data frames.	Evaluating	BTL-5
2	Show how to access columns from a data frame?	Applying	BTL-3
3	Classify rbind () and cbind () function.	Analyzing	BTL-4
4	How to find the number of columns in a data frame with an example?	Creating	BTL-6
5	Name the function to check if a variable is a data frame or not.	Analyzing	BTL-4
6	List out few basic statistic functions.	Remembering	BTL-1
7	Write a formula to normalize a variable.	Applying	BTL-3
8	How to draw an empty R plot?	Understanding	BTL-2
9	How to set the axis labels and title of the R plots?	Understanding	BTL-2
10	How to save a plot as an image on disc?	Understanding	BTL-2
11	Define plot () function.	Remembering	BTL-1
12	Define Skewness.	Remembering	BTL-1
13	Define Kurtosis.	Remembering	BTL-1
14	Define Visualizing.	Remembering	BTL-1
15	Define Bivariate analysis	Remembering	BTL-1

16	Show Z-test and t-test explain with an equation?	Applying	BTL-3
17	Show the purpose of using ANOVA test?	Applying	BTL-3
18	Write the syntax of Covariance and Correlation.	Creating	BTL-6
19	Discuss about variance.	Evaluating	BTL-5
20	Discuss about standard deviation.	Evaluating	BTL-5
21	Explain histogram.	Evaluating	BTL-5
22	Explain Time series analysis.	Analyzing	BTL-4
23	How R can be used for predictive analysis?	Understanding	BTL-2
24	How would you measure correlation in R?	Understanding	BTL-2

PART-B (13 Marks)

1	Summarize the operations that can be performed on a Data frame.	**(13)**	Evaluating	BTL-5
2	Demonstrate with syntax how to select the subset of the data frame.	**(13)**	Applying	BTL-3
3	How to access components of a Data Frame?	**(13)**	Understanding	BTL-2
4	Illustrate how to import data in R programming.	**(13)**	Applying	BTL-3
5	List out the various methods that one can export data to a text file with a syntax.	**(13)**	Remembering	BTL-1
6	How to create two different x and y-axes? Explain with an example.	**(13)**	Understanding	BTL-2
7	How to add or change the R plot's legend? Write a syntax with an example.	**(13)**	Understanding	BTL-2
8	How to adjust the size of points in an R plot? Write a syntax with an example.	**(13)**	Understanding	BTL-2
9	Illustrate the bivariate analysis of two categorical variables.	**(13)**	Applying	BTL-3
10	Point out the function which is used for the conversion of covariance to correlation in R. Explain the function with syntax.	**(13)**	Analyzing	BTL-4
11	List out the methods for calculating the correlation with an example.	**(13)**	Remembering	BTL-1
12	Elaborate variance for regression model with an example program.	**(13)**	Analyzing	BTL-4

13	Analyze the difference between covariance and correlation.	**(13)**	Analyzing	BTL-4
14	i. List out few applications of covariance. ii.Briefly explain about statistical functions for central tendency.	**(7)** **(6)**	Remembering	BTL-1
15	i.List out few applications of correlations. ii.How to handle the bivariate data through graphics?	**(7)** **(6)**	Remembering	BTL-1
16	i.Discuss about plot () function. ii.Create the scatterplot for the relation between weight and miles per gallon.	**(7)** **(6)**	Evaluating	BTL-5
17	Create the following for line chart: a. Simple line graph in R code with plot function b. Saving line graph in the PNG file. **c.** Create multiple lines in the line chart and add a legend to line graph	**(5)** **(4)** **(4)**	Creating	BTL-6
	PART – C (15 Marks)			
1	Explain with an example a. how to create a data frame b. To add the new variables to data frame. **c.** How to modify a data frame in R?	**(5)** **(5)** **(5)**	Evaluating	BTL-5
2	Write a code to demonstrate various charts using tree datasets for the following a. Histogram b. Scatter plot c. Box plot d. Line chart	**(4)** **(4)** **(4)** **(3)**	Creating	BTL-6
3	Show the inferences about skewness and kurtosis of a population given below: **Frequency distribution of litter size in rats, n-815** Litter Size: 1 2 3 4 5 6 7 8 9 10 11 12 Freq-ency: 7 38 58 116 125 126 121 107 56 37 25 4	**(15)**	Creating	BTL-6

4	Illustrate with the following example to covert the covariance value to correlation. Pass two vectors a and b such that they obey all the terms of a square matrix. Further, using cov2cor() function, we achieve a corresponding correlation matrix for every pair of the data values.	**(15)**	Evaluating	BTL-5
5	Create histogram using hist() function for the built-in dataset airquality which has " Daily air quality measurements in New York ".	**(15)**	Creating	BTL-6

UNIT – V – INTERFACING

R – CSV Files – Excel File – Binary Files – XML files – Web Data – Database – Regression – Decision Tree – Random Forest, R Random Forest, Generalized Linear Model in R with example, K- means Clustering in R with example

PART –A(2 Marks)

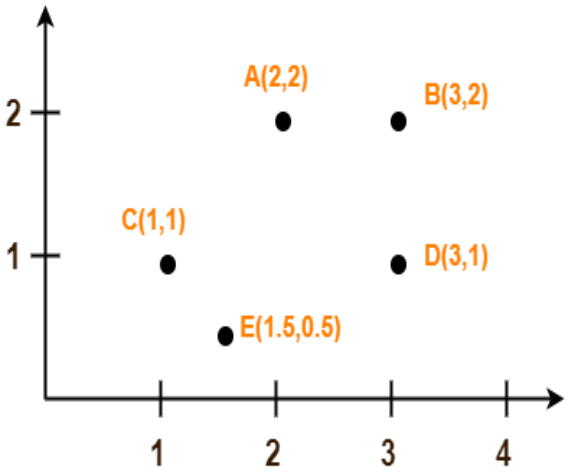
Q. No	Questions	BT Level	Competence
1	Show how to delete the content from files?	Applying	BTL-3
2	Distinguish between binary and text files.	Understanding	BTL-2
3	Discuss some binary file properties	Understanding	BTL-2
4	Discuss about some of the packages in R which are used to scrap data from the web.	Understanding	BTL-2
5	What do you mean by normal distribution?	Analyzing	BTL-4
6	What is the difference between Correlation and Regression?	Analyzing	BTL-4
7	Show how do you identify outliers?	Applying	BTL-3
8	Give an example scenario where a multiple linear regression model is necessary.	Creating	BTL-6
9	Demonstrate some of the Evaluation Metrics for regression model.	Applying	BTL-3
10	What does Intercept means?	Remembering	BTL-1
11	Difference between Mean Absolute Error (MAE) vs Mean Squared Error (MSE)?	Understanding	BTL-2

12	Why do you need to prune the decision tree?	Remembering	BTL-1
13	Define Tree Boosting?	Remembering	BTL-1
14	How is a Random Forest related to Decision trees?	Applying	BTL-3
15	Define Entropy.	Remembering	BTL-1
16	What is Out-of-Bag error?	Remembering	BTL-1
17	What do you mean by Bagging?	Remembering	BTL-1
18	Can random forest algorithm be used both for continuous and categorical target variables?	Evaluating	BTL-5
19	Write Generalized Linear Model (GLM) function.	Evaluating	BTL-5
20	How to create generalized linear model in R?	Understanding	BTL-2
21	Write out the generalized linear model in R?	Evaluating	BTL-5
22	What is the main difference between k-Means and k-Nearest Neighbours?	Analyzing	BTL-4
23	What is the difference between the Manhattan Distance and Euclidean Distance in Clustering?	Analyzing	BTL-4
24	Write about how to pre-process the data for k-Means?	Creating	BTL-6

PART-B (13 Marks)

1	Illustrate with an example how to read a particular file from the working directory.	**(13)**	Applying	BTL-3
2	Explain merge () function and its use with an example.	**(13)**	Evaluating	BTL-5
3	List out text file properties with examples	**(13)**	Remembering	BTL-1
4	List out the basic assumptions of linear regression.	**(13)**	Remembering	BTL-1
5	i. How would you detect over fitting in linear models and how to avoid it? ii. Identify the problem of over fitting and under fitting.	**(7)** **(6)**	Understanding	BTL-2
6	How is the Error calculated in a Linear Regression model?	**(13)**	Understanding	BTL-2
7	i.How does the CART algorithm produce Regression Trees? ii.Brief the following	**(7)**	Understanding	BTL-2

	a. XML Files b. Binary Files c. CSV Files	**(2)** **(2)** **(2)**		
8	Compare Linear Regression and Decision trees.	**(13)**	Analyzing	BTL-4
9	i.Explain the structure of decision tree. ii.Explain about Generalized Linear Model in R with example.	**(7)** **(6)**	Evaluating	BTL-5
10	Compare how the random forest give output for classification and regression problems.	**(13)**	Analyzing	BTL-4
11	Summarize advantages of using Random Forest.	**(13)**	Understanding	BTL-2
12	Analyze how does Random Forest handle missing values?	**(13)**	Analyzing	BTL-4
13	Describe how it is possible to perform Unsupervised learning with random forest?	**(13)**	Remembering	BTL-1
14	Use the adult data set to illustrate Logistic regression. The "adult" is a great dataset for the classification task. The objective is to predict whether the annual income in dollar of an individual will exceed 50.000. The dataset contains 46,033 observations and ten features: • age: age of the individual. Numeric • education: Educational level of the individual. Factor. • marital.status: Marital status of the individual. Factor i.e. Never-married, Married-civ-spouse, … • gender: Gender of the individual. Factor, i.e. Male or Female	**(13)**	Applying	BTL-3

	• income: Target variable. Income above or below 50K. Factor i.e. >50K, <=50K amongst others			
15	Describe some cases where k-means clustering fails to give good results?	**(13)**	Remembering	BTL-1
16	Use k-Means Algorithm to create two clusters:  	**(13)**	Creating	BTL-6
17	a. What are some stopping criteria for k-Means Clustering? b. How does the k-means algorithm works?	**(7)** **(6)**	Applying	BTL-3
PART – C (15 Marks)				
1	Problem Statement: Consider the R inbuilt data "mtcars". a. First we create a csv file from it and convert it to a binary file and store it as a OS file. b. Next we read this binary file created into R.	**(10)** **(5)**	Evaluating	BTL-5
2	Problem Statement: Let's assume we want to play badminton on a particular day — say Saturday — how will you decide whether to play or not. Let's say you go out and check if it's hot or cold, check the speed of the wind and humidity, how the weather is,	**(15)**	Evaluating	BTL-5

<table>
<tr><td></td><td>i.e. is it sunny, cloudy, or rainy. Take all these factors into account to decide if you want to play
<table><tr><th>Day</th><th>Weather</th><th>Temperature</th><th>Humidity</th><th>Wind</th></tr><tr><td>1</td><td>Sunny</td><td>Hot</td><td>High</td><td>Weak</td></tr><tr><td>2</td><td>Cloudy</td><td>Hot</td><td>High</td><td>Weak</td></tr><tr><td>3</td><td>Sunny</td><td>Mild</td><td>Normal</td><td>Strong</td></tr><tr><td>4</td><td>Cloudy</td><td>Mild</td><td>High</td><td>Strong</td></tr><tr><td>5</td><td>Rainy</td><td>Mild</td><td>High</td><td>Strong</td></tr><tr><td>6</td><td>Rainy</td><td>Cool</td><td>Normal</td><td>Strong</td></tr><tr><td>7</td><td>Rainy</td><td>Mild</td><td>High</td><td>Weak</td></tr><tr><td>8</td><td>Sunny</td><td>Hot</td><td>High</td><td>Strong</td></tr><tr><td>9</td><td>Cloudy</td><td>Hot</td><td>Normal</td><td>Weak</td></tr><tr><td>10</td><td>Rainy</td><td>Mild</td><td>High</td><td>Strong</td></tr></table>or not. So, calculate all these factors for the last ten days and form a lookup table like the one below.</td><td></td><td></td><td></td></tr>
<tr><td>3</td><td>Problem Statement:
To build a Random Forest model that can study the characteristics of an individual who was on the Titanic and predict the likelihood that they would have survived.</td><td>(15)</td><td>Creating</td><td>BTL-6</td></tr>
<tr><td>4</td><td>Write in detail about k-Means Algorithm with an example.</td><td>(15)</td><td>Creating</td><td>BTL-6</td></tr>
<tr><td>5</td><td>Cluster the following eight points (with (x, y) representing locations) into three clusters:

A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)

Initial cluster centres are: A1(2, 10), A4(5, 8) and A7(1, 2).

The distance function between two points a = (x1, y1) and b = (x2, y2) is defined as-

$P(a, b) = |x2 – x1| + |y2 – y1|$</td><td>(15)</td><td>Evaluating</td><td>BTL-5</td></tr>
</table>

	Use K-Means Algorithm to find the three cluster centres after the second iteration.			