Department of Computer Science & Engineering (Data Science)

Lesson Plan &Work-done Diary for AY: 2023-24, ODD Semester

Course with Code: Data Analytics with R - BDS306C				Faculty: Dr Anitha D B			Semester & Section: 3	
Class No.	**Date planned (DD/MM)**	**Topics to be covered**	**TLP Planned**	**Class No.**	**Date of Conduction (DD/MM)**	**Topics Covered**	**TLP Executed**	**Remarks if any deviation**
MODULE-1								
1.		Basics of R Introducing R, Initiating R, Packages in R	Chalk & Talk PPT					
2.		Environments and Functions	Chalk & Talk PPT					
3.		Basic Data Types in R	Chalk & Talk PPT					
4.		Flow Controls	Chalk & Talk PPT					
5.		Loops,	Chalk & Talk PPT					
6.		Vectors	Chalk & Talk PPT					
MODULE-2								
7		Matrices	Chalk & Talk PPT, Kiel µvision4					
8		Arrays	Chalk & Talk PPT, Kiel µvision4					
9		Lists	Chalk & Talk PPT, Kiel µvision4					
10		Data Frames	Chalk & Talk PPT, Kiel µvision4					
11		Factors, Strings	Chalk & Talk PPT, Kiel µvision4					
12		Dates and Times	Chalk & Talk PPT.					

Course with Code: Data Analytics with R - BDS306C				Faculty: Dr Anitha D B			Semester & Section: 3	
Class No.	**Date planned (DD/MM)**	**Topics to be covered**	**TLP Planned**	**Class No.**	**Date of Conduction (DD/MM)**	**Topics Covered**	**TLP Executed**	**Remarks if any deviation**
MODULE-3								
13		Data Preparation : Datasets	Chalk & Talk PPT, Kiel µvision4					
14		Importing and Exporting files	Chalk & Talk PPT, Kiel µvision4					
15		Accessing Databases	Chalk & Talk PPT, Kiel µvision4					
16		Data Cleaning	Chalk & Talk PPT, Kiel µvision4					
17		Transformation	Chalk & Talk PPT, Kiel µvision4					
18		More Datasets	Chalk & Talk PPT, Kiel µvision4					
MODULE-4								
19		Graphics using R Exploratory Data Analysis	Chalk & Talk PPT					
20		Main Graphical Packages	Chalk & Talk PPT					
21		Pie Charts, Scatter Plots	Chalk & Talk PPT					
22		Line Plots, Histograms	Chalk & Talk PPT					
23		Box Plots, Bar Plots	Chalk & Talk PPT					
24		Other Graphical packages	Chalk & Talk PPT					

Course with Code: Data Analytics with R - BDS306C				Faculty: Dr Anitha D B			Semester & Section: 3	
Class No.	**Date planned (DD/MM)**	**Topics to be covered**	**TLP Planned**	**Class No.**	**Date of Conduction (DD/MM)**	**Topics Covered**	**TLP Executed**	**Remarks if any deviation**
MODULE-5								
25		Statistical Analysis using R Basic Statistical Measures	Chalk & Talk PPT					
26		Normal distribution	Chalk & Talk PPT					
27		Binomial distribution	Chalk & Talk PPT					
28		Correlation Analysis	Chalk & Talk PPT					
29		Regression Analysis	Chalk & Talk PPT					
30		Linear Regression Analysis of Variance	Chalk & Talk PPT					

Course with Code: Data Analytics with R - BDS306C				Faculty: Dr Anitha D B			Semester & Section: 3	
Class No.	**Date planned (DD/MM)**	**Topics to be covered**	**TLP Planned**	**Class No.**	**Date of Conduction (DD/MM)**	**Topics Covered**	**TLP Executed**	**Remarks if any deviation**
PRACTICAL COMPONENT								
1.		Introduction to the tool	R Studio					
2.		Demonstrate the steps for installation of R and R Studio.	R Studio					
3.		Perform the following: a) Assign different type of values to variables and display the type of variable. Assign different types such as Double, Integer, Logical, Complex and Character and understand the difference between each data type. b) Demonstrate Arithmetic and Logical Operations with simple examples.	R Studio					
4.		Perform the following: a) Demonstrate generation of sequences and creation of vectors. b) Demonstrate Creation of Matrices c) Demonstrate the Creation of Matrices from Vectors using Binding Function. e) Demonstrate element extraction from vectors, matrices and arrays.	R Studio					
5.		Assess the Financial Statement of an Organization being supplied with 2 vectors of data: Monthly Revenue and Monthly Expenses for the Financial Year. You can create your own sample data vector for this experiment) Calculate the financial metrics:	R Studio					
6.		Develop a program to create two 3 X 3 matrices A and B and perform the following operations a) Transpose of the matrix b) addition c) subtraction d) multiplication	R Studio					

7.		Develop a program to find the factorial of given number using recursive function calls.	R Studio					
8.		Develop an R Program using functions to find all the prime numbers up to a specified number by the method of Sieve of Eratosthenes.	R Studio					
9.		The built-in data set mammals contain data on body weight versus brain weight. Develop R commands to: a) Find the Pearson and Spearman correlation coefficients. Are they similar? b) Plot the data using the plot command. c) Plot the logarithm (log) of each variable and see if that makes a difference.	R Studio					
10.		Develop R program to create a Data Frame with the given details and do the following operations a) Subset the Data frame and display the details of only those items whose price is greater than or equal to 350. b) Subset the Data frame and display only the items where the category is either “Office Supplies” or “Desktop Supplies” c) Create another Data Frame called “item-details” with three different fields itemCode, ItemQtyonHand and ItemReorderLvl and merge the two frames	R Studio					
11.		Let us use the built-in dataset air quality which has Daily air quality measurements in New York, May to September 1973. Develop R program to generate histogram by using appropriate arguments for the following statements. a) Assigning names, using the air quality data set. b) Change colors of the Histogram c) Remove Axis and Add labels to Histogram d) Change Axis limits of a Histogram e) Add Density curve to the histogram	R Studio					

12.		Design a data frame in R for storing about 20 employee details. Create a CSV file named “input.csv” that defines all the required information about the employee such as id, name, salary, start_date, dept. Import into R and do the following analysis. a) Find the total number rows & columns b) Find the maximum salary c) Retrieve the details of the employee with maximum salary d) Retrieve all the employees working in the IT Department. e) Retrieve the employees in the IT Department whose salary is greater than 20000 and write these details into another file “output.csv”	R Studio					
13.		Using the built in dataset mtcars which is a popular dataset consisting of the design and fuel consumption patterns of 32 different automobiles. The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models). Format A data frame with 32 observations on 11 variables : [1] mpg Miles/(US) gallon, [2] cyl Number of cylinders [3] disp Displacement (cu.in.), [4] hp Gross horsepower [5] drat Rear axle ratio,[6] wt Weight (lb/1000) [7] qsec 1/4 mile time, [8] vs V/S, [9] am Transmission (0 = automatic, 1 = manual), [10] gear Number of forward gears, [11] carb Number of carburetors Develop R program, to solve the following: a) What is the total number of observations and variables in the dataset? b) Find the car with the largest hp and the least hp using suitable functions c) Plot histogram / density for each variable and determine whether continuous variables are normally distributed or not.	R Studio					

		If not, what is their skewness? d) What is the average difference of gross horse power(hp) between automobiles with 3 and 4 number of cylinders(cyl)? Also determine the difference in their standard deviations. e) Which pair of variables has the highest Pearson correlation?						
14.		Demonstrate the progression of salary with years of experience using a suitable data set (You can create your own dataset). Plot the graph visualizing the best fit line on the plot of the given data points. Plot a curve of Actual Values vs. Predicted values to show their correlation and performance of the model. Interpret the meaning of the slope and y-intercept of the line with respect to the given data. Implement using lm function. Save the graphs and coefficients in files. Attach the predicted values of salaries as a new column to the original data set and save the data as a new CSV file.	R Studio					