Course code	Course Title	L	T	P	C
BCSE207L	Programming for Data Science	2	0	0	2
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. To provide necessary knowledge on data manipulation and to perform analysis on the practical problems using a programming approach.
2. To generate report and visualize the results in graphical form using programming tools.
3. To learn and implement R programs for data science.

Course Outcome

Upon completion of the course, the student will be able to

1. Engrave and use R language to solve problems.
2. Design a suitable form for analysis from real-time data.
3. Formulate insights from the data through statistical inferences.
4. Evaluate and visualize the results, analyze the performance of the models.

Module:1	**Functions in R**	**2 hours**

Programming with R- Running R Code - Including Comments - Defining Variables, Functions -Built-in R Functions - Loading Functions - Writing Functions - Using Conditional Statements.

Module:2	**Vectors and Lists**	**3 hours**

Vector - Vectorized Operations - Vector Indices - Vector Filtering - Modifying Vectors, Lists - Creating Lists - Accessing List Elements - Modifying Lists- Applying Functions to Lists with lapply().

Module:3	**Data Wrangling**	**4 hours**

Understanding Data - The Data Generation Process - Finding Data - Types of Data - Interpreting Data - Using Data to Answer Questions - Data Frames - Working with Data Frames -Working with CSV Data.

Module:4	**Manipulating Data with dplyr and tidyr**	**5 hours**

Data Manipulation - Core dplyr Functions- Performing Sequential Operations -Analyzing Data Frames by Group - Joining Data Frames Together - dplyr in Action: Analyzing Flight Data- Reshaping Data with tidyr -From Columns to Rows: gather() - From Rows to Columns: spread() - tidyr in Action: Exploring Educational Statistics.

Module:5	**Accessing Databases and Web APIs**	**5 hours**

An Overview of Relational Databases -A Taste of SQL-Accessing a Database from R - Accessing Web APIs -RESTful Requests -Accessing Web APIs from R -Processing JSON Data -APIs in Action: Finding Cuban Food in Seattle.

Module:6	**Data Visualization**	**6 hours**

Designing Data Visualizations - The Purpose of Visualization - Selecting Visual Layouts - Choosing Effective Graphical Encodings - Expressive Data Displays - Enhancing Aesthetics - Creating Visualizations with ggplot2- A Grammar of Graphics - Basic Plotting with ggplot2 - Complex Layouts and Customization - Building Maps- ggplot2 in Action: A case study.

Module:7	**Interactive Visualization in R**	**3 hours**

The Plotly Package - The Rbokeh Package - The Leaflet Package - Interactive Visualization in Action: Exploring Changes to the City of Seattle.

Module:8	**Contemporary Issues**	**2 hours**
	Total Lecture hours:	**30 hours**

Text Book(s)

1. Michael Freeman and Joel Ross, Programming Skills for Data Science: Start Writing

	Code to Wrangle, Analyze, and Visualize Data with R, Addison-Wesley, 2018.		
Reference Books			
1.	Benjamin S. Baumer, Daniel T. Kaplan and Nicholas J. Horton, Modern Data Science with R, Chapman and Hall/CRC, 2021.		
2.	John Mount and Nina Zumel, Practical Data Science with R, 2nd edition, Wiley, 2019.		
Mode of Evaluation : Continuous Assessment Tests, Quizzes, Assignment, Final Assessment Test			
Recommended by Board of Studies	12-05-2022		
Approved by Academic Council	No. 66	Date	16-06-2022