

Course Code: CSE1006	Course Title: Foundations for Data Analytics	TPC	2	2	3
Version No.	1.1				
Course Pre-requisites/ Co-requisites	None				
Anti-requisites (if any).	None				
Objectives:	1. To introduce concepts related to the data analytics 2. To perform data analytics using R and Python				
CO's Mapping with PO's and PEO's					
Course Outcomes	Course Outcome Statement	PO's / PEO's			
CO1	Understand the importance of data analytics	PO1, PO2, PO3/PEO1			
CO2	Manipulate data using R	PO1, PO2, PO3/PEO1, PEO2, PEO3, PE04			
CO3	Perform exploratory data analysis using R	PO1, PO2, PO3/PEO1, PEO2, PEO3, PE04			
CO4	Analyze data using python	PO1, PO2, PO3/PEO1, PEO2, PEO3, PE04			
Total Hours of Instructions: 30					
Module No. 1	Introduction to Data Analytics				4 Hours
Data-Information- - characteristics of data - data munging-scraping-sampling-cleaning- importance of data analytics -Success Stories					
Module No. 2	Introduction to R				6 Hours
Introduction to R, R installation, Basic operations in R using command line, use of IDE R Studio, ‘R help’ feature in R, Data types and function, Variables in R, Scalars, Vectors, Matrices, List, Data frames, functions in R, Factors					
Module No. 3	Data manipulation				5 Hours
Data sorting, Find and remove duplicates record, Cleaning data, Recording data, Merging data					
Module No. 4	Data Analysis				5 Hours
Data Import: Reading Data, Writing Data in R, data cleaning and summarizing with dplyr package, Exploratory Data Analysis: Box plot, Histogram, Pie graph, Line chart,Barplot,Scatter Plot					
Module No. 5	Using Python for Data Science				5 Hours
Overview of Python,Introduction to NumPy, NumPy standard data types, the basics of NumPy Arrays: NumPy Array Attributes, Array Indexing: Accessing Single Elements, Array Slicing: Accessing Subarrays, Reshaping of Arrays, Array Concatenation and Splitting, Aggregations,Computations on Arrays, NumPy’s Structured arrays					
Module No. 6	Introduction to Pandas				5 Hours
Introducing Pandas Objects, Data Indexing and Selection, Operating on Data in Pandas, Handling missing data, Hierarchical Indexing, Vectorized String Operations, Visualization with Matplotlib					
Text Books					
1. Wes McKinny, “Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython” O’Reilly Media, 2017.					
2. Nina Zumel, Jim Porzak, John Mount “Practical Data Science with R”, Publisher: Dreamtech, 2014.					
References					
1. Jeffrey S. Saltz and Jeffrey M. Stanton, “An Introduction to Data Science” Sage Publication, 2018.					
2. Mark Lutz, “Learning Python”, O’Reilly Media, 5th Edition 2017.					

3. John D. Kelleher and Brendan Tierney, “Data Science” MIT Press, 2018. 4. Sinan Ozdemir, “Principles of Data Science” 2018	
Lab Exercises	
1. Basic Operations In R 2. Data Types and Functions In R 3. Data Sorting 4. Finding and Removing Duplicate Records 5. Data Cleaning 6. Data Recording 7. Data Merging 8. Reading and Writing Data In R 9. Data Cleaning and Summarizing with dplyr Package 10.Exploratory Data Analysis 11. Basic Operations on Numpy 12.Computations on Arrays 13.Numpy’s Structured Arrays 14.Introducing Pandas Objects 15.Data Indexing and Selection 16.Operating on Data in Pandas 17.Handling Missing Data 18.Hierarchical Indexing 19.Vectorized String Operations 20.Visualization with Matplotlib	
Course Type	Embedded Theory and Lab (ETL)
Mode of Evaluation	Theory – 75%
	Continuous Assessment Test-1 15
	Continuous Assessment Test-2 15
	Digital Assignments/Quizzes (Min) 30
	Final Assessment Test 40
	Lab – 25%
Recommended by the Board of Studies on	14th BoS, 11.05.2024
Date of Approval by the Academic Council	12th Academic Council, 25.05.2024