Title of the Course	Statistics using R and Python

Credits	T	P	E	C
	3	1	0	4
Course Type (Theory/Practical/Integrated)	Integrated			
Course Category	Core Discipline			
Pre-Requisite	1. Basic statistics understanding – foundational topics like mean, variance, probability, hypothesis testing 2. Basic programming skills – knowledge of any programming language is sufficient			
Learning Objectives	1. Understand and apply Python and R programming for statistical data analysis. 2. Perform data manipulation and preprocessing using Pandas and dplyr libraries. 3. Visualize data effectively using statistical plots in Python and R. 4. Apply probability distributions and hypothesis testing techniques to real-world datasets. 5. Analyze data using regression, ANOVA, and multivariate statistical methods.			

Course Outcomes & Bloom's Level

CO Code	Course Outcome Statement	Bloom Level
CO1	Explain fundamental concepts of Python, R, and data handling techniques for statistical analysis.	L2
CO2	Apply data manipulation, cleaning, and transformation techniques using Pandas and dplyr libraries.	L3
CO3	Analyze datasets using visualization techniques and descriptive statistical measures.	L4
CO4	Evaluate statistical inference methods including hypothesis testing and probability distributions.	L5
CO5	Develop models using regression, ANOVA, and multivariate analysis techniques.	L3
CO6	Design and implement data-driven solutions using Python and R for real-world case studies.	L6

Course Elements

Course Element	Coverage Level
Skill Development	Yes

	Entrepreneurship	Partial
	Employability	Yes
	Professional Ethics	Partial
	Gender	No
	Human Values	Partial
	Environment & Sustainability	No
SDG (Goals)	SDG 4: Quality Education	
Total Hours of Pedagogy	45 hours Theory (15 hours Self-paced content + 30 hours lecture) 30 hours Practical	

Module#	Content	Pedagogy
M-1	**Introduction to Python and Pandas** Introduction to Python, Python installation, variables and operators, reading data from keyboard, Data types - numbers, lists, dictionary; Decision making Loops - for, while; Loop control statements - break, continue, pass; Introduction to Pandas - Pandas Installation, data structures - series, data frame; Reading from csv files, loc(), iloc() function, descriptive statistics, inserting columns into DataFrame, deleting columns from DataFrame, concatenating DataFrame, writing back to csv files, reading from excel files; Merging or joining DataFrame; Group by and aggregate functions; Connecting Python to MySQL	Self-paced content, Lecture, Modular Assignment
M-2	**Introduction to R Programming** Introduction to R programming, R programming installation and running, Variables, Operators, Data Types, Decision making, Loops, Functions, Reading from csv files, writing back to csv files, reading from excel files, Packages - installation, loading; Introduction to dplyr Library, dplyr functions, arrange, join, select, rename, filter, summarize, Pipe operator, group_by, do, join, if, across; Connecting R to MySQL	Self-paced content, Lecture, Modular Assignment
M-3	**Data Visualization and Elementary Statistics** Data Visualization - bar charts, grouped bar charts, stacked bar charts, histogram, line charts, pie charts, box plots, scatter plots, density plots, strip charts, QQ plots - implementation in both Python and ; Probability distributions - binomial, Poisson, normal, exponential, chi-squared, student's t; F distribution and Uniform distribution; Z-Test, one sample Z-Test, two sample Z-Test, F-Test, Student's t-Test - implementation of tests in Python and R	Self-paced content, Lecture, Modular Assignment
M-4	**Tests of Hypothesis and Non-parametric Tests** Tests of Hypothesis - population mean with known variance, population with unknown variance, population proportion, goodness of fit, chi-square test; Implementation of tests in Python and R; Non-parametric tests - one-sample test, two-sample test - two-sample sign test with binomial distribution, two-sample sign test with normal approximation; Wilcoxon signed-rank test, Mann-Whitney-Wilcoxon	Self-paced content, Lecture, Modular Assignment

	test, Kolmogorov-Smirnov test, K-Samples test - Kruskal-Wallis H test - implementation of tests in Python and R	
M-5	**ANOVA and Multivariate Analysis** Analysis of variance - one-way ANOVA, Two-way ANOVA; ANOVA - implementation in Python and R; Multivariate Analysis - correlation, covariance, forecasting - simple moving average, weighted moving average, simple exponential smoothing implementation in Python and R; Regression - simple linear regression, multiple linear regression, polynomial regression, logistic regression, poisson regression	Self-paced content, Lecture, Modular Assignment
M-6	**Advanced Graphs, Case studies and Discussion** Advanced graphs - lattice graphs, ggplot2 graphs, probability graphs, mosaic graphs, correlograms; implementation in Python and R; Case studies and discussion	Self-paced content, Lecture, Modular Assignment

List of DIY Modular Assignments

1. Data Exploration and Transformation Using Python and Pandas
2. Analytical Workflows in R: Tidy Data from Spreadsheets to MySQL
3. Practical Data Science: Real-World Dataset Analysis Using Python and R
4. Applied Data Analysis Using Python and R
5. Comparative Group Analysis Using ANOVA in Python and R
6. Case-Based Analysis with Advanced Graphing in Python and R
7. Cross-Platform Probability & Hypothesis Testing Analysis
8. Applied Statistical Testing in Python and R

Modular Assignment Mapping

S. No.	***DIY Assignment Title***	***Mapped CO(s)***	***Bloom Level***
1	Data Exploration and Transformation using Python and Pandas	CO2	L3
2	Analytical Workflow in R – Tidy Data from Spreadsheets to MySQL	CO2, CO6	L3
3	Practical Data Science – Real World Dataset Analysis using Python and R	CO3, CO6	L4
4	Applied Data Analysis using Python and R	CO3	L4
5	Comparative Group Analysis using ANOVA in Python and R	CO4, CO5	L5
6	Case-based Analysis with Advanced Graphing in Python and R	CO3, CO6	L5
7	Cross-platform Probability and Hypothesis Testing Analysis	CO4	L5
8	Applied Statistical Testing in Python and R	CO4, CO5	L5

Reference Books	1. Think Stats: Exploratory Data Analysis in Python, Allen B. Downey, 2nd Edition, O'Reilly Media, 2014. 2. Python for Data Analysis, Wes McKinney, 2nd Edition, O'Reilly Media, 2017.

	3. An Introduction to Statistical Learning, Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, 2nd Edition, Springer, 2021. 4. The R Book, Michael J. Crawley, 2nd Edition, Wiley, 2012. 5. R for Data Science, Hadley Wickham & Garrett Grolemund, 1st Edition, O'Reilly Media, 2017. 6. Practical Statistics for Data Scientists, Peter Bruce & Andrew Bruce, 2nd Edition, O'Reilly Media, 2020.

Course Articulation Matrix

CO \ PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	2	1	1	2	–	–	–	–	1	–	2
CO2	2	3	2	2	3	–	–	–	–	1	–	2
CO3	2	3	2	3	3	–	–	–	1	2	–	2
CO4	2	3	1	3	2	–	–	1	–	1	–	2
CO5	3	3	3	2	3	–	–	–	1	1	1	2
CO6	2	3	3	3	3	1	–	1	2	2	2	3