University of Kerala

Discipline	Mathematics				
Course Code	UK4DSEMAT202				
Course Title	Data Analysis using Python				
Type of Course	DSE				
Semester	IV				
Academic Level	200-299				
Course Details	Credit	Lecture per week	Tutorial per week	Practical per week	Total Hours per week
	4	3		2	5
Pre-requisites	Basics of python, Descriptive Statistics, Statistical plotting, Testing of hypothesis, simple linear regression and correlation				
Course Summary	This course in Data Visualization and Statistical Analysis using Python equips students with the skills to visualize data through various plots and graphs, obtain statistical measures, conduct hypothesis tests, and perform correlation and regression analysis, all within the Python programming environment.				

Detailed Syllabus

Module	Unit	Contents	Hrs
I		**Introduction to Python**	**9**
	1	Installing and setting up Python, Basic syntax, variables, and expressions, Strings, lists, dictionaries, tuples, sets, Control structures: conditionals and loops, Functions and modules.	
		Chapter 2 of Text [1]	
II		**Data Visualization**	**9**
	2	Introduction to matplotlib, Bar charts, line charts, histograms, Scatter plots and correlation visualization, Simple statistical visualization: box plots, distribution plots.	
		Chapter 2 of Text [1]	
III		**Essential Mathematics for Data Science**	**9**

21

Module	Unit	Contents	Hrs
	3	Linear algebra essentials: vectors, matrices, operations, Statistics: mean, variance, correlation, Probability basics.	
		Chapter 4, 5, 6 of Text [1]	
IV		**Hypothesis Testing and Gradient Descent**	**9**
	4	What is inference?, Hypotheses testing and Statistical significance, p-values, A/B testing, Resampling methods (e.g., bootstrap), Confidence intervals, Practical inference examples, Gradient descent	
		Chapter 7 and 8 of Text [1]	
V		**Suggestions for the teacher designed module**	**9**
	5	Simple linear regression, Multiple linear regression, Additional examples and problems related to topics from chapters 1–7.	

Practical Session (30 hours)

All the topics (including those in the suggestions for the teacher designed module) can be used for practical sessions.

Problems for the practical examination

1. Basic Python Structures: Write a Python program to read a list of numbers from the user and compute: sum, average, maximum, minimum Use only built-in Python data structures (list, loop, function).

2. Write a Python program to compute mean, median, mode, and variance of a given list of numbers.

3. Implement vector addition, dot product, and magnitude computations.

4. Simulate and Visualize Random Variables Generate 500 random values from: Uniform distribution, Normal distribution, Plot histograms and compare their shapes using matplotlib.

5. Create a program to visualize a histogram for randomly generated data.

6. Write a function to compute correlation between two lists.

7. Generate a scatter plot with labelled axes and title.

8. Visualize a bar chart of categorical data.

9. Optimization using gradient descent.

10. Simple linear regression using gradient descent.

A record should be maintained with atleast 7 problems from the above. Each problem in the record must have a description of the problem, algorithm (step by step procedure), commands used, input given and output obtained accordingly. For the ESE, from the list of above problems, the student should be able to answer two selected (from the 7 available in the record) by the examiner.

Textbooks

1. Joel Grus, Data Science from Scratch: First Principles with Python,second edition, O'Reilly, 2019.

References

1. Huy Hoang Nguyen, Paul N Adams, Stuart J Miller, Building Statistical Models in Python, Packt Publishing, 2023.

2. Samir Madhavan, Mastering Python for Data Science, Packt Publishing, 2015

3. Wes McKinney, Python for data analysis, third edition, O'Reilly, 2022.

4. Allen B. Downey, Think Stats: Probability and Statistics for Programmers; 2nd Edition, O'Reilly Media, 2014

5. Wes McKinney, Python for Data Analysis, 2nd Edition; O'Reilly Media, 2017.

6. William Mendenhall, Robert J. Beaver, and Barbara M. Beaver ; Introduction to Probability and Statistics, 15th Edition, Cengage Learning, 2016.

23

Course Outcomes

CO No.	Upon completion of the course the graduate will be able to	PO/PSO	Cognitive Level	Knowledge Category	Lecture(L) Tutorial (T)	Practical (P)
CO 1	Construct appropriate data visualizations to explore and interpret given datasets using Python.	PSO1, PSO2, PSO4, PO1, PO3	U	L	C	P
CO 3	Compute and interpret basic statistical and linear algebraic measures relevant to data analysis problems.	PSO3, PO3	Ap, E	L	P	P
CO 4	Perform statistical inference using hypothesis testing techniques.	PSO1, PSO2, PSO5, PO6	E, Ap, An	L	P	P
CO 5	Implement optimization and modelling techniques such as gradient descent and regression by integrating Python programming.	PSO1, PSO2, PSO5, PO6	E, Ap, An	L	P	P

(R-Remember, U-Understand, Ap-Apply, An-Analyse, E-Evaluate, C-Create)
(F-Factual, C-Conceptual, P-Procedural, M-Metacognitive)

Mapping of CO with PSOs and POs

	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	-	3	-	3	-	-	2	-	-	-	-	-	-	-
CO2	3	3	-	2	-	-	2	-	2	-	-	-	-	-
CO3	-	-	3	-	-	-	-	-	2	-	-	-	-	-
CO4	3	3	-	-	2	-	-	-	-	-	-	2	-	-
CO5	2	3	-	2	2	-	-	-	-	-	-	2	-	-

(- -Nill, 1-Slightly/Low, 2-Moderate/Medium, 3-Substantial/High)

24

Assessment Rubrics

- Quiz/Assignment/Discussion/Seminar

- Midterm Exam

- Programming Assignments

- Final Exam

Mapping of COs to Assessment Rubrics

	Internal Examination	Assignment	Project Evaluation	End Semester Exam
CO1	✓	✓		✓
CO2	✓	✓	✓	✓
CO3	✓	✓	✓	✓
CO4	✓	✓	✓	✓
CO5	✓	✓	✓	✓

25