

Reg. No.

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

B.E. / B.TECH. DEGREE EXAMINATIONS, MAY 2024

Second Semester

CS22201 – PYTHON FOR DATA SCIENCE*(Computer Science and Engineering)***(Regulation 2022)****TIME: 3 HOURS****MAX. MARKS: 60**

COURSE OUTCOMES	STATEMENT	RBT LEVEL
CO 1	Students will be able to understand the mathematical foundation for data science.	2
CO 2	Students will be able to solve computational problems in python.	3
CO 3	Students will be able to handle Python arrays using NumPy package.	3
CO 4	Students will be able to manipulate data using Pandas.	3
CO 5	Students will be able to understand the pattern of data by graphical displays using Matplotlib.	2

PART- A (10 x 2 = 20 Marks)

(Answer all Questions)

	CO	RBT LEVEL
1. Define Descriptive Statistics.	1	2
2. Create the Stem and Leaf plot for the given data values: 6, 8, 10, 23, 35, 43, 44, 45, 48 Also find the mean and mode of the plot.	1	2
3. Differentiate package from a module in Python.	2	2
4. Define Recursion. Write a recursive function to find GCD of two numbers.	2	2
5. How will you concatenate two NumPy arrays of different dimensions?	3	2
6. Illustrate fancy indexing in NumPy with an example.	3	2
7. Differentiate join and merge function in Pandas.	4	2
8. How to plot error bar in Python using Matplotlib?	5	2
9. How do you manage null values or missing data in Seaborn?	5	2

10. How to plot a histogram in Python using Matplotlib?

5 2

PART- B (3 x 10 = 30 Marks)

11. (a) Describe in detail about Hypothesis testing and types of testing methods.

Marks (10) CO 1 RBT LEVEL 2

(OR)

(b) (i) The marks scored by 3 students in Maths and Computer Science are given

(5) 1 2

below. Find the Covariance matrix for the same.

Student id	Maths (X)	Computer Science (Y)
1	92	80
2	60	30
3	100	70

(ii) It is estimated that 50% of emails are spam emails. Some software has been used to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

(5) 1 2

12. (a) (i) Write a Python code to count number of characters, words and lines in file.

(5) 2 3

(ii) There are 7 floors in a flat and only 2 lifts. Initially Lift A is at the ground floor and Lift B at the top floor. Whenever someone calls the lift from Nth floor, the lift closest to that floor comes to pick him up. If both the lifts are at equidistant from the Nth floor, then the lift from the lower floor comes up. Write a program to find which lift will serve the user request for the given floor number.

(5) 2 3

(OR)

(b) With suitable examples, Explain Python dictionary in detail discussing its operations for the creation, insertion, deletion, accessing an item and printing all (keys, values and items) and how will you set the default value if the key is not found in the dictionary.

(10) 2 3

13. (a) Create a NumPy's Structured array of employees, which has different fields like empID, empDept, empDesignation and empSalary. Sort the array according to name and display the employee array such that empSalary is greater than Rs.25000/-.

(10) 3 3

(OR)

- (b) Write a program to convert a string element to uppercase, lowercase, and capitalize the first letter, title-case, and swapcase of a given NumPy array. **(10)** **3** **3**

PART- C (1 x 10 = 10 Marks)

(Q.No.14 is compulsory)

		Marks	CO	RBT LEVEL
14.	Investigate the effective technique of Data Manipulation with Pandas for the following: (i) Data Selection Dataframe (ii) Handling Missing Data	(10)	4	4
