Reg. No															

B.Tech/M.Tech(Integrated) DEGREE EXAMINATION, NOVEMBER 2023

Third Semester

21CSS202T - FUNDAMENTALS OF DATA SCIENCE

(For the candidates admitted during the academic year 2022-2023 onwards)

Note:

i. **Part - A** should be answered in OMR sheet within first 40 minutes and OMR sheet should be handed over to hall invigilator at the end of 40th minute.

ii. **Part - B** and **Part - C** should be answered in answer booklet.

Time: 3 Hours **Max. Marks: 75**

PART - A (20 × 1 = 20 Marks)

Answer **all** Questions

		Marks	BL	CO
1.	hsplit() is equivalent to split with _____ and vsplit() is equivalent to split with_____. (A) axis=1, axis=0 (B) axis=0, axis=1 (C) axis=-1, axis=0 (D) axis=0, axis=-1	1	1	1
2.	To search an array, numpy offers which function? (A) search() (B) find() (C) where() (D) seek()	1	1	1
3.	Normal Distribution is applied for _____ (A) Continuous Random Distribution (B) Discrete Random Variable (C) Irregular Random Variable (D) Uncertain Random Variable	1	1	1
4.	Given two Numpy arrays, array1 = np.array([2, 4, 6, 8]) and array2 = np.array([1, 3, 5, 7]), what is the result of the following operation? Result = array1 + array2 (A) [3, 7, 11, 15] (B) [1, 3, 5, 7, 2, 4, 6, 8] (C) [2, 4, 6, 8, 1, 3, 5, 7] (D) [2, 4, 6, 8] + [1, 3, 5, 7]	1	2	1
5.	When a statement doesn't make any sense and is not meaningful, then we can say it is a: (A) compile time error (B) semantic error (C) logical error (D) run time error	1	2	2
6.	Which of the following is incorrect about classes. (A) A class is a user-defined blueprint or prototype from which objects are created. (B) classes are created using the keyword class (C) class is an instance of an object (D) attributes are the variables that belong to class.	1	2	2
7.	JSON stands for _____ (A) JavaScript Output Notation (B) Java Structured Object Network (C) JavaScript Object Network (D) JavaScript Object Notation	1	1	2
8.	What does "parsing" in JSON refer to? (A) Transforming JSON into a string (B) Extracting data from a JSON string and converting it into a usable format (C) Encrypting JSON data for security (D) Generating random data in a JSON format	1	1	2

9.	Which one of the key features on Django related to security? (A) Storing plain text passwords in cookies (B) Storing hashed passwords in cookies (C) Using client-side frameworks for security (D) Storing sensitive data in XML format	1	2	3
10.	What type of websites can be built using Django? (A) Only e-commerce websites (B) Only social networking sites (C) Almost any type of website (D) Only blogs and personal websites	1	2	3
11.	In the context of the client-server model, what is the role of a server? (A) Servers request data from clients. (B) Servers share their resources with clients. (C) Servers provide information or access to services to clients. (D) Servers only accept data packets from clients.	1	2	3
12.	What is the localhost url for Django? (A) localhost:3000 (B) http://127.0.0.1:3306/ (C) http://127.0.0.1:1/ (D) http://127.0.0.1:8000/	1	1	3
13.	Which of the following is not true about dataframe? (A) A dataframe object can be created by passing dictionaries. (B) A dataframe is size immutable. (C) A dataframe index can be string. (D) A column of dataframe can have data of different types.	1	2	4
14.	Which of the following is correct statement? (A) inplace argument of rename function is set to False, than original dataframe is changed with new index/columns. (B) del statement of dataframe can be used to delete rows. (C) When a dataframe object is created, all the columns are sorted automatically. (D) While specifying your own index sequence in DataFrame() function, Python doesn't care about length of index.	1	2	4
15.	Which of the following is/are not a iterative function for dataframe? (i) iterrows() (ii) iteritems() (iii) itercolumns() (A) (i) only (B) (i) and (ii) (C) (ii) only (D) (iii) only	1	1	4
16.	Which of the following statement for creating dataframe is valid? (A) df = pd.dataframe(dict1) (B) df = pd.Dataframe(dict1) (C) df = pd.dataFrame(dict1) (D) df = pd.DataFrame(dict1)	1	1	4
17.	The graphical representation of a frequency distribution is called (A) Bar chart (B) Line chart (C) Histogram (D) Pie Chart	1	1	5
18.	What is the difference between a bar chart and a histogram? (A) A histogram does not show the entire range of scores in a distribution (B) Bar charts are circular, whereas histograms are square (C) There are no gaps between the bars on a histogram (D) Bar charts represents numbers, whereas histograms represent percentages	1	2	5

19. What is the function of a contingency table, in the context of bivariate analysis? 1 1 5

(A) It shows the results you would expect to find by chance

(B) It summarises the frequencies of two variables so that they can be compared

(C) It lists the different levels of p value for tests of significance

(D) It compares the results you might get from various statistical tests

20. When might it be appropriate to conduct a multivariate analysis test? 1 2 5

(i) If the relationship between two variables might be spurious
(ii) If there could be an intervening variable
(iii) If a third variable might be moderating the relationship

(A) (i) only

(B) (ii) only

(C) (i) and (ii)

(D) (i), (ii) and (iii)

PART - B (5 × 8 = 40 Marks) Marks BL CO

Answer **all** Questions

21. (a) Imagine you are tasked with analyzing the performance of a mobile app for a ride-sharing company. Describe in detail the step-by-step process of Data Science that you would follow to derive valuable insights from the data. Incorporate appropriate examples to illustrate each stage of this process. 8 4 1

(OR)

(b) Describe in detail about Exponential Distribution and Binomial Distribution.

22. (a) Develop a program and discuss how JSON and Python Dictionaries Work Together. 8 3 2

(OR)

(b) Construct Regular Expression and show the basic function with examples.

23. (a) Sketch the Web development basic architecture and discuss the Django web framework. 8 4 3

(OR)

(b) Justify the need of URL mapping and URL Configuration with example.

24. (a) Implement a Pandas program to check whether alpha numeric values present in a given column of a DataFrame. 8 4 4

(OR)

(b) Implement a Pandas program to add, subtract, multiple and divide two Pandas Series.

25. (a) Implement a Python code snippet to plot a line chart with multiple lines. 8 4 5

(OR)

(b) Implement a multivariate function in Python for a random data x, y and z and show by scatter plot

PART - C (1 × 15 = 15 Marks) Marks BL CO

Answer **any 1** Questions

26. Analyze the methodology employed in examining the time intervals between consecutive customer inquiries within a customer service call center. Evaluate the potential impact of the distribution patterns of these time intervals on gauging the operational efficiency of the call center and the dynamics of inquiry resolution. Construct concrete scenarios or instances from the call center environment to illustrate the practical significance of these analyses. 15 4 1

27. Develop a program to create a form which receive data from the client with field's first name, last name, roll number and password and have a validator function for password. 15 4 3

* * * * *