Reg. No																

B.Tech DEGREE EXAMINATION, MAY 2024

Fifth Semester

18ECE372J - PYTHON FOR DATA SCIENCES

(For the candidates admitted during the academic year 2018-2019 to 2021-2022)

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: 100**

PART - A (20 × 1 = 20 Marks)

Answer **all** Questions

		Marks	BL	CO
1.	What will be the output of the following Python function? len(["good",1, 3, 5]) (A) 7 (B) 4 (C) 3 (D) error will be displayed	1	2	1
2.	What will be the output of the following Python function? min(max(True,0,-3), 2, 5) (A) 0 (B) -3 (C) True (D) 2	1	2	1
3.	--------- is a series of character (A) variable (B) string (C) word (D) function	1	1	1
4.	what will be the output of following code? str= "Umbrella" str[5]= "-" (A) Umbr_ell (B) Umbre_la (C) Type error (D) Umbre_lla	1	1	1
5.	Write the output of following code a= {1:'One',2:'Two',3:'Three'} print(a[2]+a[1]) (A) 21 (B) error (C) 3 (D) TwoOne	1	2	2
6.	To read the next line of the file from a file object infile, _____ to be used (A) infile.read(2) (B) infile.read() (C) infile.readline() (D) infile.readlines()	1	1	2
7.	Python uses special objects called -------- to manage errors that arise during a program's execution (A) Refactoring (B) Exception (C) Inheritance (D) Encapsulation	1	1	2
8.	languages = {'python', 'ruby', 'c'} (A) set (B) dictionary (C) list (D) tuple	1	1	2
9.	Which of the following is use to represent a data frame as numpy array? (A) shape (B) dtypes (C) values (D) size	1	1	3

10.	----------method to grab all the rows where the time portion of the DatetimeIndex is between two times (A) between_time( ) (B) mid_time( ) (C) time_between( ) (D) time_interval( )	1	1	3
11.	random sampling with replacement is called ---------- (A) resampling (B) stratified random sampling (C) bootstrapping (D) simple random sampling	1	1	3
12.	The -------- method of the *pandas* library converts categorical columns to numeric columns. (A) get_dummies (B) get_num (C) get_col (D) get_cat	1	1	3
13.	---------- plots to present the relationship between three numeric variables in two dimensions (A) Bar (B) Scatter (C) Histogram (D) Contour	1	1	4
14.	----------is a Python package that supports fast, flexible, and expressive data structures, as well as computing functions for data analysis (A) pandas (B) json (C) sklearn (D) struct	1	1	4
15.	scipy.special has special functions for -------- (A) plots (B) matrices (C) vectors (D) computational physics	1	1	4
16.	---------- denotes the functions of Fast Fourier transform. (A) scipy.fft (B) scipy_fft (C) scipy.fftpack (D) scipy.fillfft	1	1	4
17.	Which one of the following is not a naive Bayes model (A) Gaussian (B) Bernoulli (C) Mutinomial (D) Poisson	1	1	5
18.	------ does not build the model from the training data (A) KNN (B) Linear regression (C) Logistic regression (D) Json	1	1	5
19.	In regression model, partitioning until the point where the non-linear interactions are non-existent is called ----- partition (A) Complete (B) recursive (C) repetition (D) Full	1	1	6
20.	Ordinary Least squares (OLS) method provides ---------- (A) Linear regression (B) K nearest neighbours (C) Matlab plots (D) Arrays	1	5	6

PART - B (5 × 4 = 20 Marks)

Answer **any 5** Questions

		Marks	BL	CO
21.	Write a program to find numbers divisible by 3 from a given list using Continue statement (use for loop)	4	3	1
22.	Write a python program to get age from the user. Print the age is valid if it is greater than or equal to 18 and raise a value error otherwise.	4	3	2
23.	Write a python program to obtain a pie plot for male and female population in percentage.	4	3	3
24.	Compare Matplotlib and seaborn packages	4	3	4
25.	Write a python program to obtain pairplot using seaborn for Iris data set	4	3	5

26.	Write a python program to generate squares of first 10 natural numbers using for loop in list comprehension	4	1	1
27.	What is lambda in python? Illustrate it with a simple python code	4	2	4

PART - C (5 × 12 = 60 Marks)

Answer **all** Questions

		Marks	BL	CO
28.	(a) i) Write a program creating a tuple with 10 elements. Slice a tuple with given indices, first element to specified index value, specified index value to last element, first to last element, negative index values, increment between the elements, jump every 3 items, negative increments. (6 marks) ii) To create a string "BEAUTIFUL" and print all the letters in it using while loop. Include the test condition to check character is "T" with Continue, Break and Pass statements and write its output. (6 marks)	12	3	1
	(OR)			
	(b) i) Write a program to create a list of positive and negative numbers. Create a new tuple that has only negative numbers from the list and add an item (tuple) with and without specific index location in an existing tuple. (6 marks) ii) Write a python program using range function and while condition statement to obtain the following output 1 1 2 1 2 3 (6 marks)			
29.	(a) i) Create an user defined function with variable length argument using python to find minimum number (use for loop). (6 marks) ii) Write a program that prompts for the user's age. Use json.dump() to store this number in a file. Write a separate program that reads in this value and prints the message, "I know your age! It's _____." (6 marks)	12	2	2
	(OR)			
	(b) i) Write a while loop that prompts users for their name. When they enter their name, print a greeting to the screen and add a line recording their visit in a file called *visitor_book.txt*. Make sure each entry appears on a new line in the file. (6 marks) ii) What is exception in python? Write a python program to get a mark in percentage. Print the mark if it is less than or equal to 100 and raise a value error otherwise (6 marks)			
30.	(a) i) Create a panda data frame based on the dictionary of lists with persons name, age and income (5 items). Perform filtering the data 1) after a certain age 2) income greater than a specified value (6 marks) ii) Create a panda data frame based on the dictionary with two keys A and B and its values using range (9), range (1,10) respectively. Transform the data frame by increasing 5 times of existing value (6 marks)	12	3	3
	(OR)			
	(b) i) Create a data frame with players team name, position and age. Aggregate the data frame into group by one column and get Mean, Min, and Max values by group and group by multiple columns (6 marks) ii) Create a 3x3 matrix array from random input using arange and numpy and concatenate both row and column wise with a shuffled matrix of same size (6 marks)			
31.	(a) i) Write the python coding to create plot inside plot (separately) of not equal sizes using add axes and save it (6 marks) ii) Plot histogram bar with labels and title using Matlab library in python for income vs number of people (Use bins with interval of 15)) (6 marks)	12	1	4
	(OR)			
	(b) What are autocorrelation and lag plots? Write python code to display these plots for random data generated with NumPy.			

32. (a) Write a simple visualization code using sklearn with Iris data sheet with following specifications
standard scaler preprocessing, PCA decomposition,
convert column 2 to from cm to inches
convert column 3 to from cm to metres and scatter plot it with labels and title. 12 3 5

(OR)

(b) Construct a data frame using linear regression in python with X, (generate 100 normally distributed random numbers with mean 1.5 and standard deviation 2.5). For predicted value(Y),we assume an intercept of 2 and a slope of .2 Also calculate the values of α and β using the preceding data and observe efficacy of the model. For the actual value, residual term (res) a random variable distributed normally with mean 0 and a standard deviation of 0.5

* * * * *