Reg. No															

B.Tech DEGREE EXAMINATION, DECEMBER 2023

Fifth Semester

18ECE372J - PYTHON FOR DATA SCIENCES

(For the candidates admitted during the academic year 2020 - 2021 & 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.	Which of the following variable is correct? (A) 1face (B) _face (C) new face (D) face$	1	1	1
2.	Which of the following is used to define a block of code in Python language? (A) Key (B) Indentation (C) parantheses (D) flower Braces	1	1	1
3.	Which of the following declarations is incorrect in python language? (A) abcd = 5,123,678 (B) a,b,c,d= 5, 123, 678, 4 (C) a_b_c_d= 5,123,678,4 (D) a b c d= 5 123 678 4	1	2	1
4.	x = 'abcd' for i in range(len(x)): print(i) (A) error (B) 1 2 3 4 (C) a b c d (D) 0 1 2 3	1	2	1
5.	What will be the output of the following Python code snippet? z = {"car":35, "cycle":70} z["cycle"] (A) 35 (B) 70 (C) "car" (D) "cycle"	1	1	2
6.	---------- are enclosed in triple quotes (A) string (B) docstring (C) exceptions (D) numstring	1	1	2
7.	open() needs ---------- argument (A) no (B) 1 (C) 2 (D) 3	1	1	2
8.	When Python reads from a ------file, it interprets all contents in the file as a string (A) json (B) .py (C) .ipynb (D) text	1	1	2
9.	Data --------- is the process of preparing the data and getting it into a format that can be used for analysis. (A) collection (B) mining (C) wrangling (D) streamlining	1	1	3
10.	---------- method brings values forward. (A) 'fillf': (B) 'ffill': (C) 'fill_f': (D) 'f_fill":	1	1	3

11.	The seed gives a starting point for ---------- (A) performing calculation (B) printing the data frame (C) merging the cells (D) generation of pseudorandom numbers	1	1	3
12.	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
13.	In matlab, subplot(2,2,3) provides ---------- maximum subplots (A) 2 (B) 3 (C) 4 (D) 6	1	1	4
14.	--------- compute the frequency of unique values (A) counts() (B) value_counts() (C) true_counts() (D) unique()	1	1	4
15.	In MATLAB, horizontal bar plots can be obtained by ---------- (A) plt.hbar() (B) plt.barh() (C) hbar.plt() (D) barh.plt()	1	1	4
16.	freq='B' provides --------- (A) Business day frequency (B) Business month start frequency (C) Business year start frequency (D) Business hour frequency	1	1	4
17.	---------- sub module uses supervised learning (A) Cluster (B) Linear (C) Association (D) Naive Bayes	1	1	5
18.	Which one of the following is unsupervised learning? (A) Naive bayes (B) Linear regression (C) Cluster (D) Decision trees	1	1	5
19.	Iris datasheet contains a total of ---------- samples (A) 50 (B) 100 (C) 150 (D) 200	1	1	6
20.	---------- coefficients will be required to estimate in a simple linear regression model with one independent variable. (A) 1 (B) 2 (C) 3 (D) 4	1	1	6

PART - B (5 × 4 = 20 Marks)

Answer **any 5** Questions

		Marks	BL	CO
21.	What is List Comprehension? Write one suitable python code using it	4	1	1
22.	Create a function in python with GST (Goods Service Tax) as default argument to calculate the total bill amount inclusive of tax (round off to 2 digits)	4	3	2
23.	Write a python program to create a function with side of square as argument and to return multiple values such as diagonal, area and perimeter of it	4	3	3
24.	Plot histogram bar with labels and title using matlab library in python for age vs number of people (Use bins with interval of 10).	4	3	4
25.	Write a python program to display lag plot for random generation of 100 numbers	4	3	5
26.	Create a count plot using matlab and seaborn packages (use titanic dataset).Give suitable title and labels	4	2	4
27.	Write a python program to print "Hello World" five times using control statement and comment statement.	4	3	1

PART - C (5 × 12 = 60 Marks)

Answer **all** Questions

		Marks	BL	CO
28.	(a) i)Write a python program to calculate average of 'N' subject marks and print the total of marks, average and Grade based on the average value given below using nested if statements (8 marks)	12	3	1

Average	Grade
90-100	S
80-89	A
70-79	B
60-69	C
50-59	D
Less than 50	Fail

ii) Remove the blank space in the given string "welcome " and change into other cases using python (4 marks)

(OR)

(b) Write a python program to
i) create tuple with 10 elements and slice it 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) count the occurrence of a given character in a given string (6 marks)

29.	(a) i)Write a program to sort (ascending or descending) both key and values in the given dictionary (6 marks) original = {'c': [12, 2], 'm': [7, 6, 5], 'a': [11, 4], 'b':[3,9] } ii)Create a python function using required (positional) arguments to find HCF using Euclidean Algorithm (6 marks)	12	3	2

(OR)

(b) Write a python program to
i) calculate the sum of given numbers by creating a function with variable number of arguments (use for loop and formatting statement) (6 marks)
ii) 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. (6 marks)

30.	(a) Illustrate the operation of merging data frames with a python program	12	3	3

(OR)

(b) create 2 data frames from dictionaries
{'X': [1, 2, 3], 'y': [4, 5, 6]}) and {'Z': [1, 2, 3], 'W': [7, 8, 9]}
Perform the following operations on data frames
i) append
ii) append ignoring index
iii) concatenate along rows
iv) add keys and use.loc indexer

31.	(a) (i) What is a bootstrap plot? Illustrate it with a simple python code (6 marks) (ii) Write the python coding to create plot inside plot (separately) of not equal sizes and save it (6 marks)	12	3	4

(OR)

(b) i) Import Scipy to generate a random sample of 5,000 items and plot it using the mattplotlib.pyplot module for the following distributions (6 marks)
1) Binomial 2) Poisson
ii) Write a simple python code for swarm plot in seaborn(use dataset from "fmri") What is the advantage of it over strip plot in seaborn? (6 marks)

32. (a) Write a simple visualization code using sklearn with Iris data sheet with following specifications: 12 1 5
standard scaler preprocessing ,PCA decomposition,
convert column 1 to from cm to inches
convert column 2 to from cm to metres and scatter plot it with labels and title

(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 1 and a slope of 0.2 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 2 and a standard deviation of 0.4

* * * * *