Reg. No															

B.Tech DEGREE EXAMINATION, NOVEMBER 2023

Fifth Semester

18AIO353J - PYTHON FOR DATA ANALYTICS

(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.	What statement below best describes why we do data analytics in business (A) Analytics improve our understanding of how the business works (B) We must show a return on the investment we make in data & analytical resources (C) We need specific insights to make business decisions (D) We have to calculate & report financial results to owners / shareholders	1	1	1
2.	Ages of students in a statistics class is (A) Nominal (B) Ordinal (C) Interval (D) Ratio	1	1	1
3.	If a function does not return a value then what value will be returned by python in a function? (A) int (B) void (C) bool (D) none	1	2	1
4.	Which of the following is a suitable structure to store the register number and names of the students as a single entry? (A) dictionary (B) list (C) tuple (D) class	1	2	1
5.	Which of the following is used to create an identity matrix in NumPy? (A) zeros() (B) eye() (C) identity() (D) one()	1	1	2

6.	What is the output of the following code?
	`import numpy as np`
	`a = np.array([1, 2, 3])`
	`b = np.array([4, 5, 6])`
	`c = np.stack((a, b))`
	`print(c)`
	(A) [[1, 2, 3], [4, 5, 6]] (B) [[1, 4], [2, 5], [3, 6]]
	(C) [1, 2, 3, 4, 5, 6] (D) Error
		1	4	2

7.	What is the output of the following code snippet?
	`import numpy as np`
	`a = np.array([2,3,4,5])`
	`b= np.arrange(4)`
	`print(a+b)`
	(A) [2 3 4 5] (B) [3 4 5 6]
	(C) [1 2 3 4] (D) [2 4 6 8]
		1	2	2

8. What will be the output of the following code snippet? 1 2 2

```
import numpy as np
a=np.eye(2)
b=np.matrix([1,2],[3,4])
c=a*b
```

(A) matrix([1,0],[0,1]) (B) matrix([1,2],[3,4])
(C) matrix([4,2],[3,1]) (D) Error

9. Which of the following is correct statement for creating a series to assign None to all series of 5 elements? 1 2 3

(A) s = pd.Series(none,5) (B) s = pd.Series(none*5)
(C) s = pd.Series(None,index= [1,2,3,4,5]) (D) s = pd.Series()

10. What will be output for the following code? 1 2 3

```
import pandas as pd
import numpy as np
s = pd.Series(np.random.randn(4))
print (s.ndim)
```

(A) 0 (B) 1
(C) 2 (D) 3

11. Sujeet working on a Data Analytics problem, found that missing values in a column and want to replace missing values with 500. What was the code snippet will do so. 1 2 3

(A) df.replacena(500) (B) df.fillna = 500
(C) df.fillna(500) (D) df.replacena = 500

12. Ram wants to add new employee details to the existing dataframe. Help him to write the code snippet 1 2 3

(A) emp.iloc[len(df)] = [142,'Raghu',2300] (B) emp.loc[-1] = [142,'Raghu',2300]
(C) emp.iloc[-1] = [142,'Raghu',2300] (D) emp.loc[len(df)] = [142,'Raghu',2300]

13. which function is used to save a matplotlib plot as as image file? 1 2 4

(A) saveimage() (B) saveimg()
(C) savefig() (D) exportimage()

14. Which of the following is correct to display histogram with well-defined edges 1 2 4

(A) df.plot(type='hist', edge='red') (B) df.plot(type='hist', edgecolor='red')
(C) df.plot(type='hist', line='red') (D) df.plot(type='hist', linecolor='red')

15. What will be the color of the last bar in the following code: 1 2 4

```
plt.bar(cities, population, color = ['r','g','b', 'm'])
```

(A) Blue (B) Green
(C) Magenta (D) Red

16. How do you set a title for the axis using matplotlib library? 1 1 4

(A) ax.set_title("Axis Title") (B) ax.title("Axis Title")
(C) ax.axis_title("Axis Title") (D) ax.setTitle("Axis Title")

17. _____ is a machine learning algorithm that provides class probability as output. 1 1 5

(A) SVM (B) Naïve Bayes
(C) K Nearest Neighbour (D) Logistic Regression

18. What is the purpose of regularization in machine learning? 1 1 5

(A) To increase the accuracy of the model (B) To speed up the training process
(C) To prevent overfitting and improve generalization (D) To reduce the number of features in a model

19. The problem of finding hidden structure in unlabeled data is called. 1 2 5

(A) supervised learning (B) unsupervised learning
(C) reinforcement learning (D) Transfer learning

20. Which of the following is not a supervised learning problem? 1 1 5

(A) Grouping related documents from an unannotated corpus.
(B) Predicting credit approval based on historical data.
(C) Predicting if a new image has cat or dog based on the historical data of other images of cats and dogs, where you are supplied the information about which image is cat or dog.
(D) Fingerprint recognition of a particular person used in biometric attendance from the fingerprint data of various other people and that particular person.

PART - B (5 × 4 = 20 Marks) **Marks BL CO**

Answer **any 5** Questions

21. Differentiate the list and dictionary data types of python by their characteristics along with example. 4 2 1

22. What do you mean by slicing operating in list of python? Write an example snippet to slice the list elements from position start to end and display the same. 4 3 2

23. Write a Numpy code to create a 3*3 identify matrix and replace all 0's with any random number from 5 to 20. 4 3 2

24. Write a python program that reads data from a CSV file called sales_data.csv using pandas and display the first 5, last 5 rows and summary of the data set. 4 1 3

25. Explain how will you identify and deal with the missing values in a data frame? 4 2 3

26. State the different types of machine learning algorithms with an example 4 2 5

27. Write a python code to create a bar chart to visualize the comparisons. Use proper legend, title, and axis labels. 4 3 4

PART - C (5 × 12 = 60 Marks) **Marks BL CO**

Answer **all** Questions

28. (a) i) Explain the difference between quantitative and qualitative data. Provide examples of each type and discuss their characteristics. (4 Marks) 12 4 1
ii) Emma is playing a new mobile game involving clouds numbered from 1 to n. There are two types of clouds, ordinary clouds and thunderclouds. The game ends if Emma jumps onto a thundercloud, but if she reaches the last cloud, she wins the game. Write a code to simulate the game. (8 Marks)

(OR)

(b) i) A jail has N prisoners, and each prisoner has a unique id number,S , ranging from 1 to N. There are M sweets that must be distributed to the prisoners. But wait—there's a catch—the very last sweet S is poisoned! Can you find and print the ID number of the last prisoner to receive a sweet so he can be warned? (8 Marks)
ii) Explain the difference between structured and unstructured data. Provide examples of each type and discuss their characteristics. (4 Marks)

29. (a) Using Numpy illustrate the following with code snippets: 12 3 2

1. Create a 6X4 integer array and Prints its attributes
2. Return array of items by taking the third column from all rows
3. Resize the array to new size of 3X8
4. Delete the second column from a reshaped array and insert the a new column in its place.

(OR)

(b) Using numpy, create two 1-D array with radom numbers from 10 to 100 and perform the following:

1. Remove from one array those items that exist in another
2. Get the positions where elements of two arrays match
3. Replace all odd numbers in arr with -1
4. Find the positions of elements in x where its value is more than its corresponding element in y.

30. (a) Use Pandas to perform the following operations on the given data label. 12 4 3

sample dictionary data and list labels:
exam-data ' {'name': ['Anastasia, ' Dima', 'Katherine', 'James', 'Emily',
'Michael', 'Matthew', 'Laura, 'Kevin', 'Jonas'.
'Score': [12.5, 9, 16.5, 15, 9, 20, 14.5, 18, 8, 19].
'attempts': {1,3,2,3,2,3,1,1,2,1],
'quality':['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']
labels = ['a','b','c','d','e','f','g','h','i','j']

1. Calculate the sum of the examination attempts by the students.
2. Calculate the mean score for each different student in dataframe.
3. Select the rows where number of attempts in the examination is less than 2 and score greater than 15.
4. Find the name of the student who made maximum number of attempts.

(OR)

(b) Create Sample dataset to store student information such as name, regno, age, height and weight with some missing value in age, height and weight column.
Illustrate with code to identify the columns with missing values.
Discuss the different techniques to counter the missing value with suitable code snippet.

31. (a) Use the below dataset: 12 3 4

month_number	facecream	facewash	toothpaste	bathingsoap	shampoo	moisturizer	total_units	total_profit
1	2500	1500	5200	9200	1200	1500	21100	211000
2	2630	1200	5100	6100	2100	1200	18330	183300
3	2140	1340	4550	9550	3550	1340	22470	224700
4	3400	1130	5870	8870	1870	1130	22270	222700
5	3600	1740	4560	7760	1560	1740	20960	209600
6	2760	1555	4890	7490	1890	1555	20140	201400
7	2980	1120	4780	8980	1780	1120	29550	295500
8	3700	1400	5860	9960	2860	1400	36140	361400
9	3540	1780	6100	8100	2100	1780	23400	234000
10	1990	1890	8300	10300	2300	1890	26670	266700
11	2340	2100	7300	13300	2400	2100	41280	412800
12	2900	1760	7400	14400	1800	1760	30020	300200

i) Read sales data of bathing soap of all months and show it using a bar chart. Save this plot as image file
ii) Calculate total sale data for last year for each product and show it using a Pie chart

(OR)

(b) Create a figure with three subplots, aligned vertically.

a = [89,87,97,67,64,88,99,46,76,57]
b = [77,69,89,48,92,100,75,89,74,66]

Create a bar graph for a and b on the first two, but at the last one draw both a and b, but put b on top of a! (Hint: use the bottom argument in the bar function)
Make sure the color for a and b are consistent in all the subplots!

32. (a) Develop a program that uses sci-kit learn to build a SVM classifier on the iris dataset to predict the species of a flower based on its sepal length, sepal width, petal length, and petal width. Print the classification report of the model. 12 4 5

(OR)

(b) i) What do you mean by unsupervised machine learning? Give example? (4)
ii) Illustrate the Supervised model, Linear Regression using a sample dataset? (8)

* * * * *