32.	a. Write a Python program to load the breast cancer dataset available in sci-kit learn and perform logistic regression to classify tumors as either malignant or benign.	12	4	5
	(OR)			
	b. Develop a program that uses sci-kit learn to build a decision tree 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.			

* * * * *

Reg. No | | | | | | | | | | | | | | | |

B.Tech. DEGREE EXAMINATION, MAY 2023

Fifth Semester

18AIO353J - PYTHON FOR DATA ANALYTICS

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

Time: 3 Hours **Max. Marks: 100**

Part - A (20 × 1 Marks = 20 Marks)
Answer **All** Questions

		Marks	BL	CO

1.	Which of the following is the output of the following Python code?
	x = 5
	y = 2
	print(x ** y // x)
	(A) 2 (B) 1
	(C) 12 (D) 25
		1	1	1

2.	Which of the following is NOT a Python data type? (A) Integer (B) Float (C) Boolean (D) Complex	1	2	1
3.	Which of the following is a mutable data structure in Python? (A) Tuple (B) Set (C) String (D) Dictionary	1	1	1
4.	Which of the following functions can be used to sort a list in Python in descending order? (A) sorted() (B) sort(reverse=True) (C) sort() (D) reverse()	1	1	1
5.	Which of the following is the correct syntax to install the numpy library in Python? (A) pip install numpy (B) apt-get install numpy (C) easy_install numpy (D) conda install numpy	1	2	2
6.	What is the difference between a list and a numpy array in Python? (A) Numpy arrays are mutable, while lists are immutable. (B) Numpy arrays can hold elements of different data types, while lists can only hold elements of the same data type. (C) Numpy arrays have fixed sizes and shapes, while lists can grow and shrink dynamically. (D) Numpy arrays are one-dimensional, while lists can be multidimensional.	1	2	2
7.	What is the correct syntax to create a 2D numpy array with 3 rows and 4 columns, filled with zeros? (A) np.array(3, 4) (B) np.zeros((3, 4)) (C) np.ones((3, 4)) (D) np.full((3, 4), 0)	1	1	2
8.	Which of the following numpy functions can be used to find the maximum value in a 1D or 2D array? np.max() (A) np.sum() (B) np.average() (C) np.min() (D) np.max()	1	1	2

9.	When removing duplicates in a pandas DataFrame using the drop_duplicates() function, which of the following parameters can be used to specify the columns to consider when identifying duplicates? (A) keep (B) subset (C) inplace (D) ignore_index	1	1	3
10.	Which of the following is a valid syntax for dropping duplicates in a pandas DataFrame and keeping the last occurrence of each duplicate row? (A) df.drop_duplicates(subset='column_name', keep='last') (B) df.drop_duplicates(keep='last', subset='column_name') (C) df.drop_duplicates('column_name', keep='last') (D) df.drop_duplicates('column_name', last=True)	1	4	3
11.	Which of the following is a valid syntax for grouping data in a Pandas DataFrame based on a single column and then aggregating the results using the sum function? (A) df.groupby('column_name').agg(sum) (B) df.group('column_name').sum() (C) df.groupby('column_name').sum() (D) df.sum('column_name').groupby()	1	4	3
12.	When using the "Group by" function in Pandas, which of the following methods can be used to aggregate the data? (A) agg() (B) transform() (C) filter() (D) All of the above	1	2	3
13.	What is Matplotlib used for? (A) Data analysis (B) Machine learning (C) Data visualization (D) Data cleaning	1	1	4
14.	Which of the following commands can be used to save a plot created using Matplotlib as an image file? (A) savefig() (B) plot() (C) show() (D) xlabel()	1	2	4
15.	Which of the following types of data is best represented using a line chart? (A) Categorical data (B) Continuous data (C) Nominal data (D) Ordinal data	1	1	3
16.	Which of the following types of data is best represented using a pie chart? (A) Categorical data (B) Continuous data (C) Nominal data (D) Ordinal data	1	2	4
17.	Which of the following is not a module in sci-kit learn? (A) preprocessing (B) clustering (C) regression (D) classification	1	2	5
18.	What is the purpose of train_test_split() in sci-kit learn? (A) To split data into training and testing sets (B) To preprocess the data (C) To perform feature selection (D) To perform dimensionality reduction	1	2	5
19.	Which of the following is not a type of linear regression? (A) Simple linear regression (B) Multiple linear regression (C) Polynomial regression (D) Logistic regression	1	1	4
20.	Which of the following is an example of a supervised learning application? (A) Image classification (B) Sentiment analysis (C) Fraud detection (D) All of the above	1	3	5

Part - B (5 × 4 Marks = 20 Marks)

Answer **any 5** Questions

		Marks	BL	CO
21.	List down the differences between a list and a tuple in Python. Write one example for each.	4	2	1
22.	What is the numpy library in Python? With a sample program, list out some of the benefits of using numpy.	4	2	2
23.	Write a Python program that reads data from a CSV file called "sales_data.csv" using Pandas and displays the first 10 rows of the data.	4	3	3
24.	Write a Python program that loads an XML file called "books.xml" using Pandas and displays the book titles and authors in a table.	4	3	4
25.	Describe the steps to install the matplotlib library in Python and how to verify if it's installed correctly?	4	1	4
26.	Write a Python code snippet to create a bar chart that shows the number of pets owned by five different households. Label the X-axis as "Household" and the Y-axis as "Number of Pets."	4	3	4
27.	Discuss the applications of supervised learning in real-world scenarios.	4	2	4

Part - C (5 × 12 Marks = 60 Marks)

Answer **All** Questions

		Marks	BL	CO
28.	a. (i) Explain the difference between structured and unstructured data. Provide examples of each type and discuss their characteristics. (ii) Explain the difference between quantitative and qualitative data. Provide examples of each type and discuss their characteristics **(OR)** b. (i) Write a Python program that accepts a string from the user and counts the frequency of each character in the string using dictionaries. (ii) Write a Python program to find the maximum and minimum values in a given list using the built-in functions of Python.	12	4	1
29.	Give an example of how to use NumPy to perform a simple statistical analysis on a dataset. Describe the steps involved in the analysis and the NumPy functions that can be used for this purpose. **(OR)** b. (i) Write a Python code snippet to create a numpy array of shape (3,3) with random integers between 1 to 10. Use the numpy function 'linalg.det' to compute the determinant of the array. (ii) Write a Python code snippet to create a 2-dimensional numpy array with dimensions (3,4) and then reshape it into a 1-dimensional array. Also, print the shape of both the arrays before and after reshaping.	12	4	2
30.	a. Explain the various methods in Pandas to read data from CSV, XML, text, and HTML files. Provide examples to demonstrate the implementation of each method. Discuss the advantages of using Pandas for data reading and manipulation. **(OR)** b. Discuss the importance of data visualization in data analysis and how Pandas can be used to create visual representations of data. Provide examples to demonstrate the creation of various types of plots using Pandas and Matplotlib. Explain the advantages of using Pandas for data visualization.	12	1	3
31.	a. Explain the importance of data preprocessing in data visualization. Provide an example where data preprocessing is required before visualizing the data using Matplotlib. **(OR)** b. (i) How do you create a simple line chart using Matplotlib's Pyplot module? Provide an example code snippet. (ii) What is the architecture of Matplotlib? Explain the role of the Figure, Axes, and Axis objects in the Matplotlib architecture.	12	3	4