A10.1-R5.1-DATA SCIENCE USING PYTHON

DURATION : 03 Hours   **MAXIMUM MARKS : 100**

OMR Sheet No. :

Roll No. :   **Answer Sheet No. :**

Name of Candidate : _____ ; **Signature of Candidate :** _____

INSTRUCTIONS FOR CANDIDATES :

- Carefully read the instructions given on Question Paper, OMR Sheet and Answer Sheet.

- Question Paper is in English language. Candidate has to answer in English language only.

- There are **TWO PARTS** in this Module/Paper. **PART ONE** contains **FOUR** questions and **PART TWO** contains **FIVE** questions.

- **PART ONE** is Objective type and carries **40** Marks. **PART TWO** is Subjective type and carries **60** Marks.

- **PART ONE** is to be answered in the **OMR ANSWER SHEET** only, supplied with the question paper, as per the instructions contained therein. **PART ONE** is **NOT** to be answered in the answer book for **PART TWO**.

- Maximum time allotted for **PART ONE** is **ONE HOUR**. Answer book for **PART TWO** will be supplied at the table when the Answer Sheet for **PART ONE** is returned. However, Candidates who complete **PART ONE** earlier than one hour, can collect the answer book for **PART TWO** immediately after handing over the Answer Sheet for **PART ONE** to the Invigilator.

- **Candidate cannot leave the examination hall/room without signing on the attendance sheet and handing over his/her Answer Sheet to the invigilator. Failing in doing so, will amount to disqualification of Candidate in this Module/Paper.**

- After receiving the instruction to open the booklet and before answering the questions, the candidate should ensure that the Question Booklet is complete in all respects.

DO NOT OPEN THE QUESTION BOOKLET UNTIL YOU ARE TOLD TO DO SO.

(Answer all the questions; each question carries ONE mark)

1. **Each question below gives a multiple choice of answers. Choose the most appropriate one and enter in the "OMR" answer sheet attached to the question paper, following instructions therein.**

(1x10)

1.1 What will be the output of the below Python code ?

str1="security"

 print(str1[2:8])

(A) curity

(B) secure

(C) ecurit

(D) none of above

1.2 Which of the following is a step performed by data scientist after acquiring data ?

(A) Data preprocessing

(B) Data Cleaning

(C) Data Replication

(D) All of above

1.3 What will be the output of following code ?

import numpy as np

a = np.array([8, 13, 7, 4, 13, 16])

per = np.percentile(a, 50)

print(per)

(A) 11.5

(B) 8

(C) 10.5

(D) None of above

1.4 What will be output of following code ?

import pandas as pd

list1 = [[15,20], [55,60], [75,80]]

df = pd.DataFrame(list1)

print(df)

(A)   15 20

   0   55 60

   1   75 80

(B)   0   1

   0 15   20

   1 55   60

   2 75   80

(C)   0   1

   15   20

   55   60

   75   80

(D)   0   1   2

   0 15 55 75

   1 20 60 80

1.5 What will be the median of the 7 consecutive number 12,7,9,21,14,23,25 ?

(A) 21

(B) 9

(C) 14

(D) 12

1.6 What will be the arithmetic mean of the data : 5,2,7,4,10 ?

(A) 5.8

(B) 14

(C) 6

(D) 5.6

SPACE FOR ROUGH WORK

1.7 Which of the following chart plots by default using plot() functions ?

(A) Histogram

(B) Pie chart

(C) Bar chart

(D) Line chart

1.8 In Tkinter widget pack() function is used for_____.

(A) To perform a task by the widget

(B) To destroy the widget

(C) To pack the widget on a screen

(D) To define the size of widget

1.9 Machine Learning is a subset of____.

(A) Artificial intelligence

(B) Deep learning

(C) Data learning

(D) All of above

1.10 We can use _____ to access the value of data frame using raw labels.

(A) iloc

(B) at

(C) iat

(D) loc

2. **Each statement below is either TRUE or FALSE. Choose the most appropriate one and ENTER in the "OMR" sheet attached to the question paper, following instructions therein.**

(1x10)

2.1 In python, a dictionary does not allow duplicates.

2.2 Data cleaning does not remove inconsistent data.

2.3 Changing the size of an ndarray will create a new array and delete the original.

2.4 Ndarray is also known as the axis array.

2.5 Missing data can be represented as NaN in floating point as well as non-floating point data.

2.6 A type II error (false-negative) occurs if an investigator rejects a null hypothesis that is actually true in the population.

2.7 We can display horizontal box plot by using plt.boxplot(data, vert=false) statement.

2.8 It is possible to draw a circle directly in Tkinter canvas.

2.9 Screen inside another screen is possible by creating frames.

2.10 Decision tree algorithm is used in classification as well as regression problem.

3. **Match words and phrases in column X with the closest related meaning / words(s) / phrase(s) in column Y. Enter your selection in the "OMR" answer sheet attached to the question paper, following instructions therein.**

(1x10)

No.	X		Y
3.1	Count()	**A**	Pad the string with specified number
3.2	String Center()	**B**	It is used to change property of the widget.
3.3	Data frame	**C**	Returns a panda series containing the counts of unique values.
3.4	value_counts()	**D**	It is used to give statistical summary.
3.5	ndarray.shape	**E**	Returns the number of times an object appears in a list.
3.6	Continuous random variable	**F**	It is used to identify data by its color patterns.
3.7	Legend	**G**	Defines the dimensions of array.
3.8	Config()	**H**	It is structural representation of data.
3.9	Box plot	**I**	Unsupervised Learning
3.10	Clustering	**J**	It is a variable that can define at an exact value.
		K	It is a variable that can define over a range of values.
		L	It helps to access rows and column in dataset.
		M	Supervised Learning

4. **Each statement below has a blank space to fit one of the word(s) or phrase(s) in the list below. Enter your choice in the "OMR" answer sheet attached to the question paper, following instructions therein.**

(1x10)

A	Vertical	**B**	reshape(array,shape)	**C**	df.head()
D	random.shuffle	**E**	Numerical python	**F**	Destroy()
G	Import matplotlib.pyplot as plt	**H**	Poisson's	**I**	Training
J	GUI	**K**	Import matplotlib as plt	**L**	Numerical pyplot
M	Reshape(value)				

4.1 _____ function is used to shuffle the list().

4.2 A column is a _____ representation of data.

4.3 NumPY stands for _____.

4.4 Syntax of reshape() function is _____

4.5 _____ will display first five rows of dataframe.

4.6 $E(X) = \lambda$ is used for _____ distribution.

4.7 We can load matplotlib library using _____ to plot a graph.

4.8 _____ function is used to delete any widget from the screen.

4.9 Machine learning models build using _____ dataset.

4.10 Tkinter tool in python is used to provide _____.

PART TWO

(Answer any FOUR questions)

5. (a) Explain following terms :
   (i) Data Science
   (ii) Data Analytics

(b) Define Exploratory data analysis(EDA). List and explain the tasks performed in EDA.

(c) Explain following numpy operations with example.
   (i) max/min
   (ii) ndim
   (iii) itemsize
   (iv) dtype
   (v) addition.

(5+5+5)

6. (a) What is the use of slicing function in python ? Write an example of slicing to fetch slice from given string: "InformationSecurity" by using slice(0,10,3) and slice(-1,0,-3).

(b) Explain the following operations on panda's data frame with example.
   (i) tail()
   (ii) fillna()
   (iii) dropna().

(c) How to read data from the text files with pandas ? Summarize with syntax.

(3+6+6)

7. (a) Differentiate List and Tuple in Python.

(b) What are the stages for data science pipeline ? Summarize each stage.

(c) A coin that is fair in nature is tossed n number of times. The probability of the occurrence of a head six times is the same as the probability that a head comes 8 times, then find the value of n.

(5+6+4)

8. (a) Define Line plot. Write a python code to Plot the following data using a line plot :

Days	1	2	3	4	5	6	7
Ticket sold	2200	2500	3100	2300	1900	2700	1000

(b) Differentiate numpy and pandas in Python.

(c) Which are the methods used to load text files into numpy array ? Explain concept of fancy indexing in numpy with example.

(3+6+6)

9. (a) Differentiate correlation and regression.

(b) List out Tkinter widget. Explain any two with example.

(c) What is machine learning ? Briefly explain any two applications of machine learning.

(5+5+5)

- o O o -

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK