No. of Printed Pages : **8**

A10.1-R5-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.

PART ONE

(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 supplied with the question paper, following instructions therein. (1x10)

1.1 What will be the value of m after execution of the following line of code ?

>>> m = [[x, x + 1, x + 2] for x in range(0, 3)]

(A) [[0], [1], [2]]

(B) [[0, 1, 2], [1, 2, 3], [2, 3, 4]]

(C) [[0, 1, 2], [1, 2, 3], [2, 3, 4], [3,4,5]]

(D) NameError: name 'x' is not defined

1.2 What will be the output of the following code :

```
capitalized_list = [word.capitalize() for word in ['new', 'best', 'super']]
print (capitalized_list)
```

(A) ['New', 'Best', 'Super']

(B) ['new', 'best', 'super']

(C) ['NEW', 'BEST', 'SUPER']

(D) Name capitalized_list not defined

1.3 What is the purpose of the pack geometry manager in Tkinter ?

(A) To organize widgets in a grid layout

(B) To organize widgets with absolute positioning

(C) To organize widgets in a horizontal or vertical stack

(D) To bind functions to mouse-click events

1.4 How can you load data into NumPy from various formats, such as text files or CSVs ?

(A) Using the load() function

(B) Using the read_csv() function

(C) Using the loadtxt() function

(D) Using the import_data() function

1.5 How can you filter rows in a Pandas DataFrame based on a condition ?

(A) Using the filter() method

(B) Using the select() function

(C) Using boolean indexing

(D) Using the exclude() function

1.6 What does the term "probability distribution" refer to in statistics ?

(A) The likelihood of an event occurring

(B) The spread of data points

(C) The arrangement of data in ascending order

(D) The range of a dataset

1.7 What is the main purpose of collecting data in the Data Science Process ?

(A) Identifying patterns in data

(B) Framing the problem statement

(C) Extracting knowledge and insights

(D) Creating visualizations

1.8 Which type of Machine Learning algorithm is suitable for problems where the output is a sequence, such as predicting the next word in a sentence ?

(A) Supervised Learning

(B) Unsupervised Learning

(C) Reinforcement Learning

(D) Recurrent Neural Networks (RNN)

1.9 How can you access a specific column in a Pandas DataFrame ?

(A) Using dot notation (e.g., df.column_name)

(B) Using square brackets (e.g., df['column_name'])

(C) Using both dot notation and square brackets interchangeably

(D) Using the access_column() function

1.10 Which of the following is NOT a valid NumPy array type ?

(A) int

(B) float

(C) str

(D) complex

2. Each statement below is either TRUE or FALSE. Choose the most appropriate one and enter your choice in the "OMR" answer sheet supplied with the question paper, following instructions therein. (1x10)

2.1 The concept of "duck typing" in Python implies that the type of an object is determined at runtime based on its behavior rather than its explicit declaration.

2.2 Universal functions in Numpy operate only on one-dimensional arrays.

2.3 In Pandas, using the 'apply' function with a lambda expression is the most efficient way to perform element-wise operations on large DataFrames.

2.4 The 'map' function in Python is always more efficient than a list comprehension when it comes to applying a function element-wise to a list.

2.5 Immutable variables in Python do not allow their values to be changed after assignment.

2.6 In Tkinter, the 'grid' geometry manager is always preferable over 'pack' for organizing layouts and widgets.

2.7 Pandas Nan objects are always treated as equal during comparison operations.

2.8 In statistical analysis, the median is always a more robust measure of central tendency than the mean.

2.9 Broadcasting in NumPy allows for efficient element-wise operations on arrays with different shapes without the need for explicit looping.

2.10 In a linear regression analysis using Python, a correlation coefficient close to 1 implies a strong linear relationship between the variables, guaranteeing predictive accuracy.

3. Match words and phrases in column X with the closest related meaning / word(s) / phrase(s) in column Y. Enter your selection in the "OMR" answer sheet supplied with the question paper, following instructions therein. (1x10)

	X		Y
3.1	method to filter data based on specific conditions	**A**	bind()
3.2	method for quick data overview in Dataframes	**B**	Tkinter
3.3	Cleaning and preprocessing data	**C**	slicing
3.4	Functions that operate on entire arrays without the need for explicit looping	**D**	munging
3.5	method for mouse click event in Tkinter	**E**	describe()
3.6	Graphical User Interface	**F**	statsmodels
3.7	Linear regression analysis	**G**	matplotlib
3.8	print ('Ghaziabad'[3:-2])	**H**	query()
3.9	Z-score	**I**	outliers
3.10	Heatmap	**J**	ufunc
		K	0.5
		L	aggregation
		M	Unsupervised Learning

4. Each statement below has a blank space to fit one of the word(s) or phrase(s) in the list below. Choose the most appropriate option, enter your choice in the "OMR" answer sheet supplied with the question paper, following instructions therein. (1x10)

A	Vectorized	**B**	Context diagram	**C**	tkinter library
D	Scatter plot	**E**	Dispersion	**F**	Boosting
G	Association	**H**	Slicing, Selecting	**I**	Label-based, Position-based
J	Git push	**K**	<class 'float'>	**L**	Inline
M	Supervised learning				

4.1 A _____ is used as the first step in developing a detailed DFD of a system.

4.2 _____ will be the output of code: print(type(1/2)).

4.3 _____ create a GUI component for displaying multiple-lines of text.

4.4 _____ shows either when one continuous variable affects another or when both continuous variables are independent.

4.5 _____ command is used to upload local repository content to a remote repository.

4.6 Lambdas in Pandas can be used for applying _____ functions to manipulate data in a concise and efficient manner.

4.7 NumPy's universal functions (ufuncs) enable _____ operations on entire arrays, optimizing performance and readability.

4.8 Data indexing and selection in Pandas allow for selecting specific columns or rows using _____ or _____ methods.

4.9 Correlation measures the _____ between two variables, indicating the strength and direction of their linear relationship.

4.10 _____ ensemble learning technique that combines the predictions of multiple weak learners to create a strong learner.

PART TWO
Answer any Four Questions

5. How does the mouse and keyboard events are programmed in 'tkinter' ? Explain with the help of specific example to use each one.

(a) Write the difference between 'lambda' and 'def' with a suitable example.

(b) We have the following code with unknown function f():

```
for x in f(5) :
   print x,
```

The output looks like this :

```
0 1 8 27 64
```

Write a function f() so that we can have the output as above. **(9+3+3)**

6. (a) A file 'world_temp.txt' dataset of size 14,830 contains the data of different cities of the world. The headers are: 'Date', 'Time', 'Latitude', 'Longitude', 'city' and 'temp'. Using pandas package, write python program to find the city wise mean and variance of the temperature. Use matplotlib module to draw a graph of mean temperature of the cities.

(b) Write a Python program to print the Fibonacci series 0, 1, 1, 2, 3, 5, 8..... **(10+5)**

7. (a) What is Data Science ? Describe Exploratory Data Analysis.

(b) What various techniques are used in statistics to analyze the data ? Explain any two with examples. **(7+8)**

8. (a) (i) Write NumPy codes to :

Generate the following 10* 10 matrix

```
array([[ 0.01,  0.02,  0.03,  0.04,  0.05,  0.06,  0.07,  0.08,  0.09,  0.1 ],
[ 0.11,  0.12,  0.13,  0.14,  0.15,  0.16,  0.17,  0.18,  0.19,  0.2 ],
[ 0.21,  0.22,  0.23,  0.24,  0.25,  0.26,  0.27,  0.28,  0.29,  0.3 ],
[ 0.31,  0.32,  0.33,  0.34,  0.35,  0.36,  0.37,  0.38,  0.39,  0.4 ],
[ 0.41,  0.42,  0.43,  0.44,  0.45,  0.46,  0.47,  0.48,  0.49,  0.5 ],
[ 0.51,  0.52,  0.53,  0.54,  0.55,  0.56,  0.57,  0.58,  0.59,  0.6 ],
[ 0.61,  0.62,  0.63,  0.64,  0.65,  0.66,  0.67,  0.68,  0.69,  0.7 ],
[ 0.71,  0.72,  0.73,  0.74,  0.75,  0.76,  0.77,  0.78,  0.79,  0.8 ],
[ 0.81,  0.82,  0.83,  0.84,  0.85,  0.86,  0.87,  0.88,  0.89,  0.9 ],
[ 0.91,  0.92,  0.93,  0.94,  0.95,  0.96,  0.97,  0.98,  0.99,  1.  ]])
```

(ii) Get standard deviation of all values in matrix of Q8. (i).

(iii) Get the sum of all the columns in the matrix of Q8. (i).

(b) What will be the output of following code ?

```
import numPy as np
    x =  np.arange(2, 11).reshape(3,3)
print(x)
```

(c) Explain distribution plots in detail. **(6+3+6)**

9. Briefly explain the following (Any three) :

(a) Data structures in Pandas

(b) Correlation and Regression

(c) Difference between Global and nonlocal variables

(d) Applications of Machine Learning **(5+5+5)**

- o 0 o -

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK