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 output for the following code ?

```
a = list(range(10))
del a[::2]
for i in a:
    i+=1
print(a)
```

(A) [2, 4, 6, 8, 10]

(B) [0, 2, 4, 6, 8]

(C) [9, 7, 5, 3, 1]

(D) [1, 3, 5, 7, 9]

1.2 What is the output when following code is executed ?

```
>>>str1="helloworld"
>>>str1[::-1]
```

(A) dlrowolleh

(B) hello

(C) world

(D) helloworld

1.3 What will following code segment print ?

```
a = True
b = False
c = False
if not a or b:
    print 1
elif not a or not b and c:
    print 2
elif not a or b or not b and a:
    print 3
else:
    print 4
```

(A) 1

(B) 2

(C) 3

(D) 4

1.4 What will be the output of the following ?

```
>>>t = (1, 2)
>>>2 * t
```

(A) (1, 2, 1, 2)

(B) [1, 2, 1, 2]

(C) (1, 1, 2, 2)

(D) [1, 1, 2, 2]

1.5 Which of the following is one of the key data science skills ?

(A) Data Visualization

(B) Machine Learning

(C) Statistics

(D) All of the mentioned

1.6 What does the eye( ) function in the NumPy package return ?

(A) A diagonal matrix

(B) An identity matrix

(C) A null matrix

(D) A symmetric matrix with only 1s and 0s

1.7 Which of the following is the correct syntax for pandas series ?

(A) `pandas_Series( data, index, dtype, copy)`

(B) `pandas.Series( data, index, dtype)`

(C) `pandas.Series( data, index, dtype, copy)`

(D) `pandas_Series( data, index, dtype)`

1.8 Which plot displays distribution of data based on five-number summary ?

(A) scatter plot

(B) box plot

(C) line plot

(D) chart plot

1.9 `config()` in Tkinter is used for _____.

(A) destroy the widget

(B) place the widget

(C) change property of the widget

(D) configure the widget

1.10 What is the correct way to draw a line in tkinter ?

(A) `line()`

(B) `canvas.create_line()`

(C) `create_line(canvas)`

(D) `None`

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 pass statement is helpful when you have created a code block but it is no longer required.

2.2 By default numpy.sort() does sorting in descending order.

2.3 NumPy package is capable to do fast operations on arrays.

2.4 Using Pandas library, a series can be created using various inputs like Array, dictionary, scalar value.

2.5 Two DataFrames cannot be merged on Multiple Keys.

2.6 If data is an ndarray, index must be the same length as data.

2.7 When you create a dataframe object, it never gets its row numbers and column labels automatically.

2.8 The pie() function can plot multiple data series.

2.9 Linear Regression is a supervised machine learning algorithm.

2.10 fg in tkinter widget stands for forgap.

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 supplied with the question paper, following instructions therein.**

(1x10)

No.	X		Y
3.1	Command to sort list1 in descending order	A	`fit()`
3.2	Function of `numpy.extract()` is to _____	B	`list1.sort(reverse=1)`
3.3	_____ function of Pandas library is to apply some aggregation to one or more column.	C	`dtype`
3.4	_____ identifies the datatype of a given NumPy array.	D	text
3.5	A panel is a _____ container of data.	E	`pyplot`
3.6	Function to compute variance of data along specified axis of an array.	F	3D
3.7	Function invoked with LinearRegression() to get OLS model	G	`std()` function of Pandas
3.8	Widget used for user entry data is	H	`dataFrame.aggregate()`
3.9	Function to Calculate the standard deviation of the given set of numbers, DataFrame, column, and rows.	I	Return the elements of an array that satisfy some condition.
3.10	Name the module of matplotlib library required for plotting of graph.	J	matplot
		K	type
		L	`numpy.var()`
		M	2D

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	Y ticks	B	GUI	C	np.all()
D	row-wise(i.e. on axis=1)	E	resize, reshape	F	Y values
G	loc()	H	arange	I	Three
J	database	K	input()	L	stdev()
M	reshape, resize				

4.1 _____ function is used to read a string.

4.2 We can assign a multiline string to a variable by using _____ quotes.

4.3 To create sequences of numbers, NumPy provides a function _____ analogous to range that returns arrays instead of lists.

4.4 In 2D array, by default, sorting is done _____.

4.5 The _____ function returns its argument with a modified shape, whereas the _____ method modifies the array itself.

4.6 The _____ function returns True when all the elements of ndarray passed to the first parameter are True and returns False otherwise.

4.7 To access subset of a dataframe we can use _____ method.

4.8 Values that are displayed on y-axis are called as _____.

4.9 Statistics module in Python provides a function known as _____, which can be used to calculate the standard deviation.

4.10 Tkinter library in python provides the _____.

PART TWO

(Answer any FOUR questions)

5. (a) What is data analytics ? Mention the steps involved in data analytics ? What are its benefits ?

   (b) Differentiate supervised and unsupervised machine learning algorithms. Mention names of two algorithms for each of these.

   (c) What do you mean by exploratory data analysis ? What types of graphs are created in this analysis ?

   (5+5+5)

6. (a) Write a program that rotates the element of a list so that the element at the first index moves to the second index, the element in the second index moves to the third index, etc., and the element in the last index moves to the first index.

   (b) Write a program that inputs the index of an element of a string to be removed and then print the rest of the string.

   (c) Write a program that extracts all tuples having K digit elements from a list of tuples. The list of tuples should be hard coded in the program.

   (5+5+5)

7. (a) What is NumPy in Python ? What are its applications ? Why is NumPy Array good compared to Python Lists ?

   (b) Write a NumPy program to reverse an array (first element becomes last).

   (c) How is data visualization done in Python ? What do you mean by scatter plot, bar graph and histogram.

   (5+5+5)

8. Using given dataset (only few records are shown below for reference), write code for following tasks to be done.

company	num-of-cylinders	horsepower	average-mileage	price
alfa-romero	four	111	21	13495
audi	five	110	19	15250
audi	five	110	19	18920
bmw	four	101	23	16430
bmw	four	101	23	16925
bmw	six	121	21	20970
bmw	six	182	16	30760
chevrolet	four	70	38	6295
chevrolet	four	70	38	6575
honda	four	101	24	12945
honda	four	100	25	10345
isuzu	four	78	24	6785
jaguar	six	176	15	35550
jaguar	twelve	262	13	36000
mazda	four	68	30	5195
mazda	four	68	31	6095

(i)   Count total cars per company
(ii) Find each company's Highest price car
(iii) Find the average mileage of each car making company
(iv) Sort all cars by decreasing order of Price column
(v)   Print All Toyota Cars details

(5x3)

9. (a) What is Probability Distribution ? What are the different types of Probability Distribution used in Data Science ?
   (b) What are the Measures of Central Tendency ? Name all python functions for the same. Write small code to explain three of these functions.
   (c) What are widgets in Tkinter ? Explain any five widgets.

(5+5+5)

- o O o -

SPACE FOR ROUGH WORK