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 respect.

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 Which of the following is/are one of the key data science skills ?

(A) Statistics

(B) Machine learning

(C) Data Visualization

(D) All of the above

1.2 How would you join the two arrays ?

Note : Numpy library has been imported as np

(A) resulting_set = train_set.append (test_set)

(B) resulting_set = np.concatenate ([train_set, test_set])

(C) resulting_set = np.vstack([train_set, test_set])

(D) None of these

1.3 How would you import a decision tree classifier in sklearn ?

(A) from sklearn.decision_tree import DecisionTreeClassifier

(B) from sklearn.ensemble import DecisionTreeClassifier

(C) from sklearn.tree import DecisionTreeClassifier

(D) None of these

1.4 Suppose you are defining a tuple given below :
tup = (1, 2, 3, 4, 5)

Now, you want to update the value of this tuple at 2nd index to 10. Which of the following option will you choose ?

(A) tup(2) = 10

(B) tup[2] = 10

(C) tup{2} = 10

(D) none of these

1.5 Which of these is user defined data type of Pythan ?

(A) Tuple

(B) List

(C) Dictionaries

(D) Class

1.6 Which of the following statements is used to create an empty set ?

(A) { }

(B) set( )

(C) [ ]

(D) ( )

SPACE FOR ROUGH WORK

1.7 What is the output of the following piece of code when executed in the python shell ?

a={1,2,3}

a.intersection_update({2,3,4,5})

a

(A) {2,3}

(B) Error, duplicate item present in list

(C) Error, no method called intersection_update for set data type

(D) {1,4,5}

1.8 Which is not a plot in Matplotlib ?

(A) Scatter Plot

(B) Line Plot

(C) Square Plot

(D) Histogram

1.9 Which module in the python standard library parses options received from the command line ?

(A) getopt

(B) os

(C) getarg

(D) main

1.10 What is the type of sys.argv ?

(A) set

(B) list

(C) tuple

(D) array

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 Pandas Python library is used for Data Munging.

2.2 In Python, you can define a variable without specifying its data type.

2.3 A comment in Python language can start with # sign.

2.4 Turtle.lift() can lift the pen of in turtle.

2.5 Linear Regression is a supervised machine learning algorithm

2.6 Dictionaries in Python are immutable.

2.7 Overfitting is more likely when you have huge amount of data to train.

2.8 Linear regression is not sensitive to outliers.

2.9 Copy.write an object in Python.

2.10 Tuple is the type of sys.argv.

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)

	Column X		Column Y
3.1	Tkinter	A.	Seaborn
3.2	Bootstrap	B.	NumPy, SciPy
3.3	pylab	C.	list
3.4	Range() returns	D.	statistical programming method
3.5	self	E.	supervised machine learning algorithm
3.6	bias	F.	Machine Learning Library
3.7	Linear regression	G.	Sampling Method
3.8	decision tree	H.	Graphical User Interface
3.9	Library that uses matplotlib underneath	I.	an instance
3.10	SciKit-Learn	J.	Capture rigidity of the model.
		K.	object
		L.	unsupervised machine learning algorithm
		M.	Global variable

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	getopt	B	packet	C	window = Tk()	D	len(s)
E	Tuples	F	pandas	G	window.mainloop()	H	Tkinter
I	Matplotlib	J	Slicing	K	Use Message	L	byte array
M	Data Cleaning						

4.1 _____ is the python library used for plotting.

4.2 _____ Python library you would prefer to use for Data Munging.

4.3 _____ module in the Python standard library parses options received from the command line.

4.4 _____ create a window.

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

4.6 s = {1, 2, 4, 3}, _____ returns 4.

4.7 _____ are immutable.

4.8 _____ create an event loop.

4.9 _____ is a computationally fast way to access a range of items from sequence types.

4.10 _____ is the de-facto standard GUI package of Python.

PART -TWO

(Answer any FOUR questions)

5. Diamonds :

   This dataset contains the prices and other attributes of many diamonds.

   Content

Column Name	Description
price	price in RS (3260 - 18,823)
carat	weight of the diamond (0.2 - 5.01)
cut	quality of the cut (Fair, Good, Very Good, Premium, Ideal)
color	diamond colour, from J (worst) to D (best)
clarity	a measurement of how clear the diamond is (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best))
x	length in mm (0 - 10.74)
y	width in mm (0 - 58.9)
z	depth in mm (0 - 31.8)
depth	total depth percentage = z / mean(x, y) = $2 * z / (x + y)$ (43 - 79)
table	width of top of diamond relative to widest point (43 - 95)

   (a) Write a Pandas program to remove the second column of the diamonds Dataframe

   (b) Write a Pandas program to rename two of the columns of the diamonds Dataframe.

   (c) What is the main difference between a Pandas series and a single-column DataFrame in Python ?

   (5 + 5 + 5)

6. (a) What are two main branches of Statistics ?

(b) What is Data Science and what is the relationship between Data science and Statistics ?

(c) What is the difference between Join and Merge in Pandas ?

(4 + 4 + 7)

7. (a) Write a Python program to draw a line using given axis values taken from a text file, with suitable label in the x axis, y axis and a title.

(b) Write a Python program to create multiple plots.

(7 + 8)

8. (a) Write a NumPy program to append values to the end of an array.

(b) Write a NumPy program to create a 2d array with 1 on the border of array and 0 inside the array.

(c) Write a NumPy program compare two arrays using numpy.

(5 + 5 + 5)

9. (a) Explain different widget classes built into Tkinter

(b) Write a Python GUI program to create three single line text-box to accept a value from the user using Tkinter module.

(c) Differentiate Supervised from Unsupervised machine learning

(6 + 5 + 4)

- o O o -

SPACE FOR ROUGH WORK