

Sl. No.

A10.1-R5 : DATA SCIENCE USING PYTHON

अवधि : 03 घंटे

DURATION : 03 Hours

अधिकतम अंक : 100

MAXIMUM MARKS : 100

ओएमआर शीट सं. :

OMR Sheet No. :

रोल नं. :

Roll No. :

उत्तर-पुस्तिका सं. :

Answer Sheet No. :

परीक्षार्थी का नाम :

Name of Candidate :

परीक्षार्थी के हस्ताक्षर :

Signature of Candidate :

परीक्षार्थियों के लिए निर्देश :**Instructions for Candidate :**

कृपया प्रश्न-पुस्तिका, ओएमआर शीट एवं उत्तर-पुस्तिका में दिये गए निर्देशों को ध्यानपूर्वक पढ़ें।	Carefully read the instructions given on Question Paper, OMR Sheet and Answer Sheet.
प्रश्न-पुस्तिका की भाषा अंग्रेजी है। परीक्षार्थी केवल अंग्रेजी भाषा में ही उत्तर दे सकता है।	Question Paper is in English language. Candidate can answer in English language only.
इस मॉड्यूल/पेपर के दो भाग हैं। भाग एक में चार प्रश्न और भाग दो में पाँच प्रश्न हैं।	There are TWO PARTS in this Module/Paper. PART ONE contains FOUR questions and PART TWO contains FIVE questions.
भाग एक "वैकल्पिक" प्रकार का है जिसके कुल अंक 40 हैं तथा भाग दो "व्यक्तिपरक" प्रकार का है और इसके कुल अंक 60 हैं।	PART ONE is Objective type and carries 40 Marks. PART TWO is Subjective type and carries 60 Marks.
भाग एक के उत्तर, ओएमआर उत्तर-पुस्तिका पर ही दिये जाने हैं। भाग दो की उत्तर-पुस्तिका में भाग एक के उत्तर नहीं दिये जाने चाहिए।	PART ONE is to be answered in the OMR ANSWER SHEET only. PART ONE is NOT to be answered in the answer book for PART TWO.
भाग एक के लिए अधिकतम समय सीमा एक घण्टा निर्धारित की गई है। भाग दो की उत्तर-पुस्तिका, भाग एक की उत्तर-पुस्तिका जमा कराने के पश्चात् दी जाएगी। तथापि, निर्धारित एक घंटे से पहले भाग एक पूरा करने वाले परीक्षार्थी भाग एक की उत्तर-पुस्तिका निरीक्षक को सौंपने के तुरंत बाद, भाग दो की उत्तर-पुस्तिका ले सकते हैं।	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.
परीक्षार्थी, उपस्थिति-पत्रिका पर हस्ताक्षर किए बिना और अपनी उत्तर-पुस्तिका, निरीक्षक को सौंपे बिना, परीक्षा हॉल/कमरा नहीं छोड़ सकते हैं। ऐसा नहीं करने पर, परीक्षार्थी को इस मॉड्यूल/पेपर में अयोग्य घोषित कर दिया जाएगा।	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.
प्रश्न-पुस्तिका को खोलने के निर्देश मिलने के पश्चात् एवं उत्तर लिखना आरम्भ करने से पहले उम्मीदवार जाँच कर यह सुनिश्चित कर लें कि प्रश्न-पुस्तिका प्रत्येक दृष्टि से संपूर्ण है।	After receiving the instruction to open the booklet and before starting to answer the questions, the candidate should ensure that the Question Booklet is complete in all respect.

जब तक आपसे कहा न जाए, तब तक प्रश्न-पुस्तिका न खोलें।

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

PART ONE

(Answer all the questions)

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 these is user defined data type of Python ?

(A) List
(B) Tuples
(C) Dictionaries
(D) Class

- 1.2 Which of these is **not** a part of Data Science Process ?

(A) Data Collection
(B) Data Duplication
(C) Data Cleaning
(D) Data Visualization

- 1.3 Output of

`a = np.arange(20)`

`a[-8:17:1]`

will be :

(A) [12 13 14 15 16]
(B) [11 12 13 14 15]
(C) [11 12 13 14 15 16]
(D) [12 13 14 15 16 17]

- 1.4 What function can be used to create an identity matrix in NumPy ?

(A) `Numpy.matrix(3,3,1)`
(B) `Numpy.eye(3)`
(C) `Numpy.id_matrix(3)`
(D) `Numpy.mat(3,3,type=identity)`

- 1.5 If `a = [1,2,3,4,5]` and `b = [6,7,8,9]` and the task is to create a list which has all the elements of `a` and `b` in one dimension like `a = [1,2,3,4,5,6,7,8,9]` then the function to be used is :

(A) `a.append(b)`
(B) `a.concatenated(b)`
(C) `a.extend(b)`
(D) None of the above

- 1.6 Which of the following is **not** a valid pandas read method ?

(A) `Read_csv()`
(B) `Read_xml()`
(C) `Read_excel()`
(D) `Read_json()`

- 1.7 What is Regression used for ?

(A) To predict a data as a category
(B) To predict data as a choice
(C) To predict data as a continuous value
(D) All of the above

<p>1.8 Which of these is not a type of plot in Matplotlib ?</p> <p>(A) Scatter Plot</p> <p>(B) Histogram</p> <p>(C) Line Plot</p> <p>(D) Square Plot</p>	<p>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)</p> <p>2.1 Dictionaries in Python are immutable.</p> <p>2.2 The goal for all data collection is to capture quality evidence that allows analysis to lead to the formulation of convincing and credible answers to the questions that have been posed.</p> <p>2.3 The reshape function returns its argument with a modified shape, whereas the resize method modifies the NumPy array itself.</p>
<p>1.9 What is tkinter ?</p> <p>(A) Graphical User Interface</p> <p>(B) Guided User Interface</p> <p>(C) Python Inbuilt Variable</p> <p>(D) Function</p>	<p>2.4 A universal function (or ufunc for short) is a function that operates on numpy arrays in an element-by-element fashion, supporting array broadcasting, type casting, and several other standard features.</p> <p>2.5 An ndarray cannot be used to create a DataFrame in Pandas but a Pandas DataFrame can be converted into an ndarray.</p> <p>2.6 The standard marker for missing data in Pandas is NaN.</p> <p>2.7 Correlation is a statistical technique through which some variables can be used to minimize other variables' relation between each other.</p>
<p>1.10 Which of these is not a type of Machine Learning Algorithm ?</p> <p>(A) Supervised Learning</p> <p>(B) MultiCloud-Supervised Learning</p> <p>(C) Semi-Supervised Learning</p> <p>(D) Unsupervised Learning</p>	<p>2.8 Using Matplotlib you can give a title to your Plot/Graph but you cannot label the axes separately.</p> <p>2.9 To create a tkinter you first Import the module - tkinter then create the main window which acts like a container and then add any widgets you like to the main window.</p> <p>2.10 There are some tasks that cannot be programmed. Examples include tasks using unordered data. Machine Learning comes into the picture here.</p>

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

	X		Y
3.1	A sequence of immutable Python objects	A.	Array Slicing
3.2	The method used for dealing with missing values in Pandas Dataframe	B.	Scatter Plot
3.3	Standard elements in Tkinter like different kinds of buttons and menus.	C.	Tuple
3.4	To extract a part of array	D.	Array Aggregation
3.5	The value that appears most often in a set of data values.	E.	Data Munging
3.6	A diagram for showing either when one continuous variable affects another or when both continuous variables are independent.	F.	Line Plot
3.7	To summarize the values in an array or to obtain statistically significant value(s) from an array	G.	Fillna()
3.8	The process of transforming data with the intent of making it more appropriate for analytics	H.	head()
3.9	A Supervised Machine Learning Algorithm used for classification	I.	Data Visualization
3.10	The method used to display a select number of rows from the beginning of a Pandas Dataframe.	J.	Lambdas
		K.	Logistic Regression
		L.	Widget
		M.	Mode

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.	Series	B.	Root widget	C.	Apply
D.	Broadcasting	E.	Outliers	F.	(5, 2)
G.	Supervised	H.	Self awareness	I.	plt.show()
J.	Data Cleaning	K.	Unsupervised	L.	Unlabeled
M.	ndarray				

- 4.1 list.insert(2, 5) inserts element _____ at index _____ in the list.
- 4.2 _____ is the process of detecting and correcting (or removing) corrupt or inaccurate records from a dataset.
- 4.3 Numpy array class is called _____.
- 4.4 The term _____ describes how numpy treats arrays with different shapes during arithmetic operations.
- 4.5 _____ is a one-dimensional labeled array available in Pandas capable of holding any data type.
- 4.6 The _____ method is used to execute user defined functions for entire pandas dataframe.
- 4.7 Data points far from the trend line are called _____.
- 4.8 When Matplotlib Pyplot is imported as plt, a graph or a plot can be displayed using the function _____.
- 4.9 The method "root = Tkinter.Tk()" will create a _____.
- 4.10 kNN(k-Nearest Neighbours) is a _____ machine learning algorithm.

PART TWO

(Answer any FOUR questions)

5. (a) What is Data Science ? What skills are required to become a data scientist ?
- (b) Explain briefly about Python language. Which frameworks are used with Python ?
- (c) Write a Python program to sort an array in ascending order. (5+5+5)
6. (a) Enlist and briefly explain the most frequent used aggregation functions in Pandas.
- (b) Why are visualizations important ? Explain data visualization using Matplotlib.
- (c) Write down the steps to create a Python GUI-tkinter. (5+5+5)
7. (a) Write a Python program to find maximum element of each Row in a Matrix ?
- (b) What are the key features of Pandas library ?
- (c) Differentiate between lists and tuples. (7+3+5)
8. (a) What is a DataFrame in Pandas ? What are the different ways in which a DataFrame can be created in Pandas ? How to create an empty DataFrame in Pandas ?
- (b) Mention the applications of Machine Learning. (8+7)
9. Briefly explain the following (Any three) :
- (a) Data types in Python Language
- (b) Correlation and Regression
- (c) Indexing and Selecting Data with Pandas
- (d) Training data in Machine Learning (5+5+5)
- o o o -

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK