

**INTRODUCTION TO DATA ANALYSIS WITH PYTHON AND R
(CSEN 3142)**

Time Allotted : 2½ hrs

Full Marks : 60

Figures out of the right margin indicate full marks.

*Candidates are required to answer Group A and
any 4 (four) from Group B to E, taking one from each group.*

Candidates are required to give answer in their own words as far as practicable.

Group – A

1. Answer any twelve:

12 × 1 = 12

Choose the correct alternative for the following

- (i) An interpreter
(a) Scans the entire program and translates it as a whole into machine code
(b) Generates intermediate object code which further requires linking, hence requires more memory
(c) Makes debugging more difficult
(d) Translates a program one statement at a time.
- (ii) Low level programming languages
(a) Are closer to natural languages
(b) Are closer to machine code
(c) Are taught at junior levels in school
(d) None of the above.
- (iii) range(0,n) produces
(a) A sequence that is of type list
(b) A sequence that is not of type list
(c) Neither a sequence nor a list
(d) A list or a sequence depending on the where the function is called.
- (iv) When a child class modifies or replaces the behaviour inherited from the parent class, this is called
(a) Overriding
(b) Overloading
(c) Encapsulation
(d) Abstraction.
- (v) Which function is used to find the amount of rows and columns in an array in R?
(a) dim() (b) nchar() (c) length() (d) dim_len()
- (vi) Which of the following is true for a vector in R?
(a) It is a homogeneous one-dimensional data structure
(b) It is a heterogeneous one-dimensional data structure
(c) It is a homogeneous two-dimensional data structure
(d) It is a heterogeneous two-dimensional data structure.

- (vii) Default value of an argument for a function must be provided
 - (a) At the time of defining the function
 - (b) At the time of calling the function
 - (c) Either at the time of defining or calling the function
 - (d) At a time that depends on how the function is defined
- (viii) If a is given list, print([x*x for x in a if x > 2 and x < 5]) is an example of
 - (a) Reduction
 - (b) List comprehension
 - (c) Concatenation
 - (d) Inheritance
- (ix) __init__ function is termed as a
 - (a) Constructor
 - (b) Destructor
 - (c) Getter
 - (d) Setter
- (x) Rprof() function is a built-in tool that enables which of the following?
 - (a) Write professional level R code
 - (b) Offer on-demand help to programmers
 - (c) Determine where a program spends most of its execution time
 - (d) None of the above.

Fill in the blanks with the correct word

- (xi) The block which will always be executed whether an exception is encountered or not in is _____.
- (xii) The programming model in which instructions have the effect of making changes to memory is known as _____.
- (xiii) Deriving a child class from more than one parent classes is called _____.
- (xiv) The function pow(x,y,z) is correctly evaluated by the expression _____.
- (xv) R is a dynamic programming language, which means R automatically _____ the code as it is run.

Group - B

- 2. (a) Describe different components of big data ecosystems. [[CO1](Remember/LOCQ)]
 - (b) Write a program that randomly generates a number. Raise a user-defined exception if the number is below 0.5. [[CO1](Apply/IOCQ)]
 - (c) Explain the benefits of inheritance. [[CO1](Analysis/IOCQ)]
- 5 + 5 + 2 = 12**
- 3. (a) Discuss Data Science Process life cycle with a suitable diagram. [[CO1](Understand/LOCQ)]
 - (b) Write a python program using function that accepts two positive numbers n and m, where m<=n. The function returns numbers between 1 and n, which are divisible by m. [[CO1](Apply/IOCQ)]
 - (c) Define facets of data? How can they help in data visualization? [[CO1](Understand/LOCQ)]
- 4 + 4 + (2 + 2) = 12**

Group - C

4. (a) Write a python program to write the even numbers from a specified list to abc.txt file. The entire list should be passed through a function where only the even numbers will be inserted into the file abc.txt. [[CO2](Apply/IOCQ)]
- (b) Write a python program to find the minimum window in a given string which will contain all the characters of another given string. For example, if string1 is "PRWSOERIUSFK" and string2 is "OSU", then the minimum window is "OERIUS". [[CO6](Evaluate/HOCQ)]
- (c) Give examples of mutable and immutable data structures in python. [[CO2](Remember/LOCQ)]
4 + 6 + 2 = 12
5. (a) Write a Python code, using (i) List comprehension method and (ii) Functional programming approach to find a list, z, containing the elements found in both the lists x and y. [[CO2](Analyse/IOCQ)]
- (b) Write a python program to count the number of characters in a string using dictionaries. Display the keys and their values in alphabetical order. [[CO2](Apply/IOCQ)]
- (c) Discuss the ways to write Pythonic code. [[CO2](Remember/LOCQ)]
6 + 4 + 2 = 12

Group - D

6. (a) Compare and contrast array versus matrix in NumPy. [[CO4](Remember/LOCQ)]
- (b) Explain Boolean indexing and fancy indexing in NumPy with examples. [[CO4](Remember/LOCQ)]
- (c) What are universal functions in NumPy? Name three of their characteristics. [[CO4](Remember/LOCQ)]
4 + 4 + 4 = 12
7. (a) Discuss the applications of Index objects and re-indexing. [[CO3](Understand/LOCQ)]
- (b) Write a program using NumPy to capitalize the first letter, lowercase, uppercase, swapcase, and title-case of all the elements of a given array. [[CO3](Apply/IOCQ)]
- (c) Explain the mechanism of handling missing information. [[CO3](Analysis/LOCQ)]
4 + 6 + 2 = 12

Group - E

8. (a) What are the rules for coercion in R? [[CO5](Remember/LOCQ)]
- (b) In R, what are atomic vectors and how many types of atomic vectors are there? [[CO5](Remember/LOCQ)]
- (c) Describe the environment system in R. What is the active environment? [[CO5](Remember/LOCQ)]
4 + 4 + 4 = 12
9. (a) Explain different functions for probability distributions in R. [[CO4](Analysis/LOCQ)]

- (b) Consider the following table. Write a R program to create a data frame with name "Information". Check the data types of the data frame. Print the data frame with assigning row numbers.

	ID	Name	DOB
1	10	Sai	1990-10-02
2	11	Ram	1981-03-24
3	12	Deepika	1987-06-14
4	13	Dipesh	1985-08-16

[[CO6](Evaluate/IOCQ)]

- (c) Write a R program to convert a given matrix to a 1 dimensional array.

[[CO1](Understand/IOCQ)]

4 + 4 + 4 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	53.12	40.62	6.26

Course Outcome (CO):

After the completion of the course students will be able to

- CSEN 3142.1.** Learn and understand the basics of the Python Programming Language.
- CSEN 3142.2.** Learn about basic Python data structures.
- CSEN 3142.3.** Learn about the NumPy and pandas libraries in Python.
- CSEN 3142.4.** Learn and understand the basics of the R Programming Language.
- CSEN 3142.5.** Learn about R data structures.
- CSEN 3142.6.** Learn how to apply Python and R in building solutions to basic data analysis problems.

**LOCQ: Lower Order Cognitive Question; IOCQ: Intermediate Order Cognitive Question; HOCQ: Higher Order Cognitive Question.*