INTRODUCTION TO DATA ANALYSIS WITH PYTHON AND R
(CSEN 3135)

Time Allotted : 3 hrs   **Full Marks : 70**

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and
<u>any 5 (five)</u> from Group B to E, taking <u>at least one</u> from each group.

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

Group – A
(Multiple Choice Type Questions)

1. Choose the correct alternative for the following: **10 × 1 = 10**

    (i)   In a Python code, there is a statement x = {}. Then, x is a
            (a) set   (b) tuple
            (c) dictionary   (d) none of these.

    (ii)   Which one of the following is a mutable data type?
            (a) String   (b) Integer   (c) Tuple   (d) Dictionary.

    (iii)   What will be displayed when the following piece of code is executed in R?
            c(13, 23, 33) %in% c(3, 13, 23, 53, 73)
            (a) TRUE TRUE FASLE   (b) c(13, 23)
            (c) FALSE TRUE TRUE FALSE FALSE   (d) c(3, 13, 23, 33, 53, 73).

    (iv)   What will be displayed when the following piece of code is executed in Pandas?
            import pandas as pd
            obj = pd.Series(['b', 'a', 'a', 'c', 'd']*4)
            print(obj.describe().freq)
            (a) 2   (b) 4   (c) 8   (d) 16.

    (v)   What do we use to define a block of code in Python language?
            (a) Key   (b) Brackets   (c) Indentation   (d) None of these.

    (vi)   What will be displayed, when the statement as.logical(1) is executed in R?
            (a) TRUE   (b) FALSE   (c) "1"   (d) ASCII code of 1.

    (vii)   Which one of the following has the same precedence level?
            (a) Division, Power, Multiplication, Addition and Subtraction
            (b) Division and Multiplication
            (c) Subtraction and Division
            (d) Power and Division.

(viii)   import re
txt = "The rain in Spain"
x = re.split("\s", txt, 1)
print(x)
What would be the output of the above code?
(a) ['The', 'rain in Spain'] (b) ['The']
(c) ['The', 'rain', 'in', 'Spain'] (d) None of the above.

(ix)   Which of the following can be considered as object attribute in R?
(a) Dimensions (b) Class
(c) Length (d) All of the above.

(x)   Point out the correct statement with respect to R
(a)   Empty vectors can be created with the vector() function.
(b)   A sequence is represented as a vector but can contain objects of different classes.
(c)   "raw" objects are commonly used directly in data analysis.
(d)   The value NaN represents undefined value.

Group – B

2.   (a)   What are the characteristics of big data? Name four components of big data ecosystem.

(b)   What do you understand by equality and identity of mutable values? Give examples.

(c)   Write a Python Function with name 'reverse_lst' which takes a list (in_lst) of numbers as its argument and returns another list (out_lst), which is basically in_lst in reversed order. Do not use any in-built function.

(2 + 2) + (2 + 2) + 4 = 12

3.   (a)   Write a Python Function (without using any in-built function) with name 'power' to compute x^n, which takes two arguments as follows:
(i)   a number (x)
(ii) an integer (n), which can be negative, zero or positive, with a default value of 10.

(b)   Write a Python code to read every line from a text file named "in-data.txt" and copy alternate lines (starting from the first line of 'in-data.txt') into a file "out-data.txt".

(c)   The number of students (ns) is 41 and their average age (avage) is 19.985. It is required to display these values in the following format (5 spaces to be reserved for displaying the value of ns and total 6 spaces to be reserved for displaying the value of avage, including the decimal point and 2 places after decimal point). What should be the print statement?
"No. of students: 41, Average age: 19.99"

6 + 4 + 2 = 12

Group – C

4. (a) Write a Python Program to count the number of characters in a string using dictionaries. Display the keys and their values in alphabetical Order.

    (b) Write a Python Program to count the number of times an item appears in the list.

    6 + 6 = 12

5. (a) Write Python code, using (i) List comprehension method and (ii) Functional programming approach, to eliminate from a list, all the multiples of its first element.

    (b) 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.

    (c) Write a Python code, to initialize a 3 × 3 matrix with some arbitrary values. Then using list comprehension, filter the matrix so that only the cells divisible by 4 in rows that sum to 15 or less, will be remaining in the result matrix.

    (2 + 2) + (2 + 2) + 4 = 12

Group – D

6. (a) Using NumPy, create a 2 X 3 array containing real numbers and a 1-D array of 3 numeric elements. Show step-by-step in details, how Broadcasting rules are applied to add the above two arrays.

    (b) Mention two important points why NumPy arrays are faster than 'for' looping over lists in Python.

    (c) Using NumPy, create two 1-D arrays (x and y), each containing 5 numeric elements. Also create a Boolean array (barray) containing 5 elements. It is required to generate an array (z) taking the value from x, if the corresponding element in barray is True, otherwise taking the value from y.
    (i)   Write a list comprehension to do this
    (ii)   Mention two problems of this method
    (iii) How the above problems can be overcome in NumPy?

    6 + 2 + (1 + 2 + 1) = 12

7. (a) Write Python code snippet to compute the End Time of an Opera, while Start Time and Duration are given.

    (b) Given a point(x, y), write a Python program to find whether it lies in the First, Second, Third or Fourth Quadrant of x - y Plane.

    6 + 6 = 12

Group – E

8. (a) Explain the difference between sample(1:6, size = 2) and sample(1:6, size = 2, replace = TRUE).

(b)   Explain with an example what is 'factor' in R.

(c)   Write down two rules for coercion of data types in R with examples.

(d)   For the table given below, what is the R statement to create a data frame with name 'popul'

	City	Population	Capital
1	Bengaluru	1.2327	TRUE
2	Kolkata	1.4850	TRUE
3	Pune	0.6629	FALSE
4	Surat	0.7185	FALSE

(e)   Write a function in R which takes an integer vector and returns the mean of the elements of the vector, without using any in-built function. Use 'repeat' loop in the code.

2 + 2 + (2 + 2) + 1+ 3 = 12

9.   (a)   What are the types of vectors in R programming?

(b)   What are the features of data frame?

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

4 + 4 + 4 = 12

Department & Section	Submission link:
CSE A + B + C	https://classroom.google.com/c/MjE1MDg4MTk5ODgy/a/Mjc0NDUw MDQ5Njc4/details