

**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
any 5 (five) from Group B to E, taking at least one 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) Which one of the following is a valid example of a set?
 (a) {{3:5}, {-7: 9}, {-11: 13}} (b) {{3, 5}, {-7, 9}, {-11, 13}}
 (c) {(3, 5), (-7, 9), (-11, 13)} (d) {[3, 5], [-7, 9], [-11, 13]}
- (ii) Which one of the following statements is true, when a function 'fun' is defined as: def fun(a, b=0):
 (a) The argument b always takes the value 0
 (b) If the function is called with the value of the second argument as 0, then only the function will be executed
 (c) If the second argument is provided at the time of calling the function, its value is assumed to be 0
 (d) If the second argument is not provided at the time of calling the function, its value is assumed to be 0.
- (iii) Which of the following statements is false?
 (a) Duplicates are automatically removed in a set
 (b) s= {} cannot be used to create an empty set
 (c) Set does not maintain the elements in a particular order
 (d) Values of only mutable types can be the elements of a Python set.
- (iv) 2 + 3 = 5 and "ab" + "cd" = "abcd" are examples of:
 (a) Addition (b) Operator overloading (c) Operator overriding (d) All of the above.
- (v) As per PEP-8 style guide, which of the following is the correct style for naming a class in Python?
 (a) Bank_Account (b) BankAccount
 (c) _BankAccount (d) bank_account.
- (vi) The _____ statement allows a program to come out of a loop even if the exit condition of the loop has not been satisfied.
 (a) in (b) except (c) stop (d) break
- (vii) Which of the statements is true about the following program?

```
import = int(input("Enter a number: "))
if import==2:
    print "Yes"
else:
    print "No"
```

 (a) There is nothing wrong with the program (b) The program will run very slowly
 (c) The program will throw a runtime error (d) There is a syntax error in the program .
- (viii) In a Python code, there is a statement x = {}. Then, x is a
 (a) Set (b) Tuple (c) Dictionary (d) List.
- (ix) 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.

(x) Consider the following code snippet written in Python:

```
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) ['The rain', 'in Spain'].

Group- B

2. (a) What will be the output when the following code is run? [(CO2)(Evaluate/HOCQ)]

```
def my_function(n):
    return_value = None
    if n == 0 or n == 1:
        return_value = False
    i=2
    while i < n**0.5:
        if n%i==0:
            return_value = False
            break
        i+=1
    return_value = True
    return return_value
```

```
print(my_function(37))
```

- (b) In Python, what are anonymous functions and how are they defined? [(CO2)(Remember/LOCQ)]
 (c) How does the special function `__str__` operate?. Explain with examples. [(CO3)(Understand/LOCQ)]
 (d) How is the special function `__add__` invoked? Explain with examples. [(CO3)(Understand/LOCQ)]

3 + 3 + 3 + 3 = 12

3. (a) In Python, what is the syntax for defining a child class which inherits from multiple parent classes? Explain with examples. [(CO3)(Remember/LOCQ)]

(b) In Python, which function is used to initialize a newly created instance of a class? Explain with examples. [(CO3)(Understand/LOCQ)]

(c) Why do we need files over and above standard input/output? [(CO2)(Understand/LOCQ)]

(d) In the syntax for a file handle for opening a file such as `fh = open("___", "___")`, what do the first and second arguments represent? [(CO2)(Remember/LOCQ)]

(e) Fill in the blanks in the following sentences: In Python, the code which can raise exception(s) is placed inside the _____ clause. The code that handles the exceptions is written in the _____ clause. [(CO2)(Understand/LOCQ)]

(f) Give an example of a positive use of exception handling to fulfil a functionality. [(CO2)(Analyze/IOCQ)]

2 + 2 + 1 + 2 + 2 + 3 = 12

Group - C

4. (a) Write a Python code, using List comprehension method, to eliminate from a list, all the multiples of its first element. [(CO3)(Analyse/IOCQ)]

(b) Write a Python code, using Functional programming approach to find a list (y), containing the cubes of the odd elements of another list x. [(CO3)(Understand/LOCQ)]

(c) Initialize a list with all integer numbers from 1 to 18 (both inclusive). Write a Python code, using anonymous function,

- (i) to display the squares of each of the elements of the list
 (ii) to display only the elements which are divisible by 5
 (iii) to display the sum of all the elements of the list.

[(CO3)(Analyze/IOCQ)]

(d) (i) How do you create an empty set in Python?

(ii) $s = \{(2,),(6,),(9,)\}$ Is this set a valid set? Justify your answer.

(iii) Let d be a dictionary. What does the statement $d.clear()$ do? [(CO1)(Remember, Understand/LOCQ)]

$2 + 2 + (1 + 2 + 1) + (1 + 2 + 1) = 12$

5. (a) If d is a Python dictionary, what do the following functions do? [(CO1)(Understand/LOCQ)]
 (i) $d.values()$ (ii) $d.items()$
- (b) What will be the output of the following code? [(CO1)(Evaluate/HOCQ)]
- ```
def some_function(p):
 dp = {}
 for order in p:
 if not order==0:
 dp[order-1] = order*p[order]
 return dp
print(some_function({0:-3,3:2,5:-1}))
```
- (c) What is a tuple? Explain with examples. [(CO1)(Understand/LOCQ)]
- (d) How do you define a one element tuple? [(CO1)(Remember/LOCQ)]
- (e) What are regular expressions and how are they useful? Explain with examples. [(CO3)(Understand/LOCQ)]
- $3 + 4 + 2 + 1 + 2 = 12$

### Group - D

6. (a) What is/are the central data structure(s) of the NumPy library? How do you define it/them? [(CO4)(Understand/LOCQ)]
- (b) What will be the output of the following code? [(CO4)(Analyze/IOCQ)]
- ```
arr = np.array([1, 7, 5, -4, 8, -3])
arr[1:6:2]
```
- (c) What is fancy indexing? Explain with examples. [(CO4)(Understand/LOCQ)]
- (d) What will be the output of the following code? [(CO4)(Analyze/IOCQ)]
- ```
m = np.matrix('2 1 8; 0 5 6; 7 9 3')
np.sort(m)
```
- $4 + 3 + 2 + 3 = 12$
7. (a) In the context of Python, what is "pandas" and why is it useful? [(CO4)(Understand/LOCQ)]
- (b) What are the two most important data structures in pandas? [(CO4)(Remember/LOCQ)]
- (c) What are index objects in pandas? [(CO4)(Understand/LOCQ)]
- (d) How does pandas handle missing data? [(CO4)(Analyze/IOCQ)]
- (e) What is/are the benefits of hierarchical indexing? [(CO4)(Understand/LOCQ)]
- $3 + 2 + 2 + 2 + 3 = 12$

### Group - E

8. (a) Write a function (name: roll) in R to display the sum of two independent random numbers between 1 and 6 (both inclusive). [(CO5)(Analyze/IOCQ)]
- (b) Explain the difference between  $sample(1:6, size = 2)$  and  $sample(1:6, size = 2, replace = TRUE)$  [(CO5)(Understand/LOCQ)]
- (c) Explain briefly with examples 4 different types of atomic vectors in R. [(CO5)(Remember/LOCQ)]
- (d) What will be displayed when the following piece of code is executed in R:  
 $gender \leftarrow factor(c('male', 'female', 'female', 'male'))$   
 $gender$   
 $typeof(gender)$   
 $unclass(gender)$  [(CO5)(Understand/LOCQ)]
- $3 + 2 + 4 + 3 = 12$
9. (a) What is a data frame in R? How is it used? [(CO5,CO6)(Understand/LOCQ)]
- (b) What will the following piece of code display?  
 $vec \leftarrow c(6, -1, 3, 8, 2, -7)$   
 $vec[2:4]$  [(CO5)(Analyze/IOCQ)]

- (c) How does R use the environment system?
- (d) How does a repeat loop work in R?

[(CO5)(Understand/LOCQ)]

[(CO5)(Understand/LOCQ)]

**3 + 3 + 3 + 3 = 12**

| Cognition Level         | LOCQ  | IOCQ  | HOCQ  |
|-------------------------|-------|-------|-------|
| Percentage distribution | 68.75 | 23.96 | 07.29 |

**Course Outcome (CO):**

After the completion of the course students will be able to

CSEN3135.1. Understand the basics of data science and of the Python Programming Language and its data structures viz. Lists, Tuples, Dictionaries and Sets

CSEN3135.2. Develop Python codes using iterations, recursion, function, input/output with files and using exception handling

CSEN3135.3. Learn how to manipulate strings, use regular expression, object-oriented features of Python and also how to write good and efficient codes in Python

CSEN3135.4. Apply NumPy library and develop codes using Pandas data structures (Series and Data Frames) and other features of Pandas

CSEN3135.5. Learn and understand the basics of the R Programming Language and its data structures to develop R codes using Selection, Modification, Logical sub-setting, conditionals, loops and other advanced features of R language and debug the code.

CSEN3135.6. 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.