Ramakrishna Mission Vivekananda Educational and Research Institute
PO Belur Math, Howrah, West Bengal 711 202
School of Mathematical Sciences
Department of Computer Science

MSc Big Data Analytics : Batch 2023-25, Semester 1, Mid Term
Programming for Data Science
Instructor: Dr. Sudeep Mallick

*S*tudent Name (in block letters): Date: 20 Sep 2023
*S*tudent Roll No: *M*ax Marks: 60
*S*ignature: Time: 2 hrs

Please note the following:

- **Question 1 is COMPULSORY. Attempt ANY TWO other questions. All questions carry equal marks.**
- You are allowed to refer to your class notes, material, etc.
- You are **NOT** allowed use of the Internet during the examination unless instructed otherwise by the invigilator.
- Any indication of plagiarism or infraction would result in either cancellation of the exam or a particular question as the case may be.
- You are STRONGLY advised not to copy code from class examples, you may refer to them however. Any indication otherwise would result in either cancellation of the attempt or suitable deduction of marks as deemed appropriate during evaluation.

1. Attempt the following (COMPULSORY - *all parts carry equal marks*): [2x10 = 20]
   (a) Create an tuple of 5 literal elements - 2 numbers, 1 string and 2 boolean.
   (b) Convert the tuple created in part (a) into a list.
   (c) Add a new element to the list from part (b).
   (d) Create a new list of 2 literal elements and add the contents from this new list to the list from the part (c) such that the updated list is now made up of 7 literal elements
   (e) Given a string - 'examination' extract the substring 'nation' using two ways - using positive indices and negative indices.
   (f) Using the list created in part (d) select the alternative elements and create a new set from these selected elements.
   (g) Create a for loop to multiply the elements of a list with any number so that the original list is updated.
   (h) Create a while loop to multiply the elements of a list with any number so that the original list is updated.
   (i) Create a dictionary so that the keys are first letter of fruit names and the values are the corresponding fruit names.
   (j) Given a list of integers find the statistical range (difference of the largest number and smallest number) of the numbers in the list using some of the built-in functions.

2. Using the provided text file bda23_25_midterm.txt do the following: [8+8+4]
   (a) Find the count of all the vowels (a,e,i,o,u).
   (b) Find the count of all the words (anything separated by space is a word).
   (c) Arrange the words in descending order of counts.

3. Do the following: [8+12]

(a) Create a function that receives a tuple and a number as two parameters. The function returns a tuple with the contents of the tuple it received as one of the parameters multiplied by the number it received as the other parameter.

(b) The built-in chr() function in python converts an ASCII code into the corresponding ASCII character. Numbers frm 65 to 90 generate from A to Z. Do the following:

i. Use this function to generate a random collection of 1000 alphabets.

ii. Categorise the alphabet collection into 3 categories a1, c1 and c2 and put them into an appropriate dictionary based data structure. All the vowels A,E,I,O,U go to a1, C, S, X, Z, J go to c1 and rest go to c2.

iii. Print a report where on each line the frequency count is shown as stars beside the category labels (see sample report below):

```
a1 ***************
c1 **********************
c2 ************************************************************
```

4. Do the following: [8+12]

(a) Convert the following python datastructure A into datastructure B given below programmatically:

```
A = {'one': [5,2,7,1], 'two': ('z','q',3,6), 'three': {8,9,'m','n'}}
B = [['one','two','three'], [5,'z',8],[2,'q',9],[7,3,'m'],[1,6,'n']]
```

(b) Write a function that accepts two parameters (a list of sorted numbers and a value to be inserted into the list). The function inserts the second parameter in the correct location in the sorted list. Demonstrate the correctness of the function for all the following test cases using a sorted list having values 2, 3, 4, 56, 100, 102 in that order :

```
test case 1: the value to be inserted is 1
test case 2: the value to be inserted is 2
test case 3: the value to be inserted is 5
test case 4: the value to be inserted is 10
test case 5: the value to be inserted is 101
test case 6: the value to be inserted is 102
test case 7: the value to be inserted is 103
```