CSE 7212c_CUTe_R and Python

(Max. Duration: 4 hours)   **(Max Marks : 50)**

INSTRUCTIONS:

- You can use R/Python lab activity sheets for help with commands/ syntax.
- Directly work on R-Studio/Jupyter notebook and save the files.
  Naming convention:
  - **- FirstName_LastName_CUTe01_7212c.Rmd**
  - **- FirstName_LastName_CUTe01_7212c.ipynb**
- Appropriate comments are mandatory for every problem.
- Submit only one Rmd and one .ipynb code in a .zip file in the grader tool.
  - **Naming convention: CUTe_CSE_7212c _FirstName_LastName**

PART A - R (20 marks):

1. The dataset "go_tracks.csv" contains various information collected by a 'goTracks App' based on GPS information. The details of the dataset is provided in the text file "Data_Description.txt".
   a. Import the 'go_tracks.'csv' file into R. (1 mark)
   b. Look at the structure and summary of the datasets. Observe the data and report your findings as comments in your R file. (2 marks)
   c. Convert the attributes into appropriate data types. (2 marks)
   d. Impute the na values, "?" should be treated as na. (1 mark).
   e. Write a user-defined function which takes in a numeric attribute and returns the range of that attribute. (Do not use any inbuilt functions like range, max, min or apply functions etc).
      Apply the function you made and find the range of the following attributes in this dataset - 'speed', 'time' and 'distance'. (5 marks)
   f. Use any of the apply functions in R and find the mean of the attributes - 'speed' and 'distance'. (1 mark)

Inspire...Educate...Transform.

g. Use a 'for loop' to loop through the values in the attribute 'distance' and find the maximum distance travelled. You should **not** use the inbuilt function for this. (2 marks)

h. Access the rows 1-10 for the columns - 'speed', 'time', 'distance'. Save the result into a new df called GPS_df. Is it a numeric or a character df? Find the sum of the columns of the newly created df and report your results. (2 marks)

i. Calculate the Variance for all the numeric columns in GPS_df. (1 mark)

j. Standardize the GPS_df columns. (1 mark)

k. Find the average speed when the traffic rating given is 'bad' <u>and</u> the weather is 'raining'. (2 marks)

PART B - Python (30 marks):

2. Answer the following questions :

a. Print all the integers till 1000 which are not a multiple of 8 but are divisible by 4, also print the count of such occurrences. (2 marks)

b. Create a list of 20 integers and write code using **only** lambda , filter, map functions: (10 marks)

    i. Print the sum of the list of squared values

    ii. Print the odd values

    iii. Print the even values

    iv. Print the numbers that are divisible by 5 but are not divisible by 10

    v. Print the numbers that are not prime

c. Write a function to print "Palindrome" if the input string is a palindrome, else "Not a Palindrome".   (3 marks)
Eg: If input is "Rotator" the function should return "Palindrome"
If input is "Insofe" the function should return "Not a Palindrome"

d. Write a function that takes a string as an input and returns the sum of the numbers corresponding to each alphabets position from a to z. (Hint: can use a dictionary, handle cases of the input string)   (5 Marks)

Inspire...Educate...Transform.

CSE 7212c_CUTe_R and Python

(Max. Duration: 4 hours)   **(Max Marks : 50)**

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

Ex: Regards ---> 18 + 5 + 7 + 1 + 18 + 4 + 19 = 72

3. Read the 'bank.csv' file as DataFrame with Pandas and answer/solve the following questions.

   a. Reading in the data, print the Summary, Head and Tail of the data (2 Marks)

   c. What is the most common occupation in the 'job' column? (2 Marks)

   d. What is the mean 'balance' for people with marital status as married and education as tertiary?   (2 Marks)

   e. What is the mean 'age' for different 'education' categories? (1 Mark)

   f. Drop the following columns from the dataframe: job, day, month, pdays, previous, poutcome and store the resulting dataframe in a new variable. (1 Mark)

   g. Recode loan variables with 2 levels as 0 & 1 using a function. Return 0 for level 'no' and return 1 for level 'yes'. (2 Marks)

Inspire...Educate...Transform.