

Programming for Data Analytic

SOFT8032

First Assessment

October 2021

1 Open Book Examination

You are allowed to use any materials that are associated with this module. This includes your lecture notes, lab exercises and any other materials that are shared in Canvas with you.

You are NOT allowed to use email or any other communication tools (e.g., whatsapp facebook etc) during the examination. You may only use email in case you want to communicate with your lecturer.

A python file is provided for you in Canvas. Please download the python file and complete the file based on the following tasks requirement. You will then need to submit the completed python file without anything else.

Input files are assumed to be in the same folder as the working python file.

Please type your name and ID in the designated area in the template file.

Once you finish the tasks, rename the template.py file into S12345678.py where 12345678 is your student ID and submit it via Canvas.

1.1 Tasks

1. Complete the function called *LineDetector* in the template file as follows: The function has no input argument. The function should open the text file (**CountOfMonteCristo.txt**) and return an immutable collection of lines from the file with the following conditions:
 - (a) Each line must either start with " or ' (i.e., straight double quotes or straight single quote).
 - (b) Each line must either end with " or ' or . (i.e., straight double quotes or straight single quote or dot).
2. Complete the function called *IntegerDetector* in the template file as follows: The function has an input argument (a string). The function does not return anything. The function should detect all the sub-strings from the input string that can be converted into an integer number.
 - Note1: A sub-string of a string is not necessarily separated by spaces. For instance "11" is a sub-string of "11am". "11" is also a sub-string of "Number 11 is odd".

- Note2: A sub-string can be converted into a negative integer if it starts with - For instance "-5" is sub-string of "to-5".
- Note3: "123" and "55" are two sub-strings of "123.55" that can be converted into integer numbers.
- Note4: You should not convert the sub-strings into integers, only print them as string.

The following is a list of strings with their sub-strings that can be converted into integer:

"farshad 2020to2021 toosi -67-7" → [2020, 2021, -67, -7]

"Number -5to -1 is freezing1982" → [5, -1, 1982]

"198 is 2 less then 200." → [198, 2, 200]

"1.50 is 2 less then 3.5." → [1, 50, 2, 3, 5]

3. Complete the function called *CasualUsersAMPM* in the template file as follows:

Use **bikeSharing.csv** file and write a program that will compare the average number of casual users before midday (AM) and after midday (PM). The information about hours (i.e., before and after midday) can be found in the 5th column.

The function does not return anything. It only prints the average number of casual users for AM (before midday) and the average number of casual users for PM (after midday).

See Figure 1 for information about the columns.

1. instant: record index
2. season : season (1:springer, 2:summer, 3:fall, 4:winter)
3. yr : year (0: 2011, 1:2012)
4. mnth : month (1 to 12)
5. hr : hour (0 to 23)
6. holiday: whether day is holiday or not (extracted from [Web Link])
7. weekday : day of the week
8. workingday : if day is neither weekend nor holiday is 1, otherwise is 0.
9. + weathersit :
 - i. 1: Clear, Few clouds, Partly cloudy, Partly cloudy
 - ii. 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
 - iii. 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
 - iv. 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
10. temp : Normalized temperature in Celsius. The values are divided to 41 (max)
11. atemp: Normalized feeling temperature in Celsius. The values are divided to 50 (max)
12. hum: Normalized humidity. The values are divided to 100 (max)
13. windspeed: Normalized wind speed. The values are divided to 67 (max)
14. casual: count of casual users
15. registered: count of registered users
16. cnt: count of total rental bikes including both casual and registered

Figure 1: Information