COMP9321 Assignment One: Welcome to Barcelona

Version:1.0
Weight: 10 Marks

Due: **23:59:59 9th March, 2019**

Pre-requirements

- You should able to write some workable python code fluently ($\leq$ 999 lines)
- Know how to use basic python data structure and library like matplotlib,math,csv....
- Python version is **3.7.2** which is the version installed on CSE machine. Packages are not limited in this assignment, as long as it is pre-installed in standard CSE environment.
- You can always try your codes on CSE login servers[1]

Overview

In this assignment, you will be given some 'real experience' on your week 1-2 lecture content. You are asked to make some CRUD(create,read,update and delete) operations to the data-set which we provided. After finish this assignment you should have some idea about how to do data ingestion,cleansing and manipulation on a small-size data-set(less than 5MB). In this assignment, we will work on the data which from Barcelona.

Tasks

Question 0.
Download the zip file a1.zip which contain a python code template a1.py and a few csv files: Do **not** change the file name.(0 mark)

Question 1.
(1 mark)
Barcelona is second largest city on Spain, accidents are quite normal in a city of such scale. In this question, you are require to read the accident data("accidents_2017.csv") correctly and print table head with first 10 lines of data.

Question 2.
(2 mark)
You may noticed during the first question that some fields in this file is "unknown". In this question you need to remove all lines with "unknown" fields, and save to "result_q2.csv".

[1]See https://taggi.cse.unsw.edu.au/FAQ/Logging_In_With_SSH/#CSE_Login_Servers

Question 3.
(3 mark)
Statistics of accidents can be useful, and in this question, you are asked to produce and print a table of total numbers of accidents in different district("District Name" in the dataset) with names, descending ordered.

Question 4.
(4 marks)
It is also interesting to view different data together. "air_stations_Nov2017.csv" contains air quality station information while "air_quality_Nov2017.csv" is the air quality logs. Firstly, print the air station names with its district names, in json format; Secondly, print the first 10 records that the air quality is **NOT** "Good"; Finally, save the **accident** data when the air quality is **NOT** "Good", into 'result_q3.csv', in the same format of the original "accidents_2017.csv".

Question 5.
(Bonus 3 marks, total marks for "Assignment 1" is up to 10)
Plot a Heat Map to show the total accident data on the provided map("Map.png"), where the coordinates start from UTM 31N 686563 5048578 to 31N 717652 5074225. The ploted map should be saved as "plot.png" with the same size of the original map.

Submission

Submit you python source code through give commend on CSE machine(or VLAB/SSH):

```
give cs9321 assn1 a1.py
```

Note: If your code fails to run, you will get 0 on this assignment. Please make sure your python version is 3.7.2, and it works properly on the CSE machine. We don't accept any reason like:' It works on my own computer, but I don't know why it doesn't work on CSE'.

Plagiarism

Plagiarism will result in 0 marks in Term 1, 2019 for this course.

Finally, and also very importantly, have fun!
COMP9321 administration team.