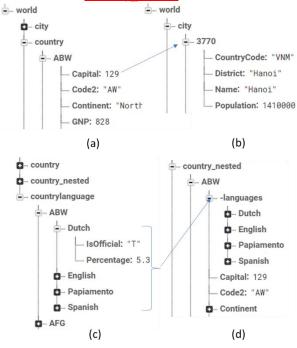
DSCI 551 - HW1 (Fall 2020)

100 points, Due September 20, Sunday (11:59pm)

In this homework, you are provided with 3 tables exported from a MySQL database (the world sample database). They are in JSON format. Consider the following 3 queries.

- A. Find names of countries and their capital cities for all countries in "North America". e.g., United States, Washington (DC)
- B. Find names of countries and their official languages for all countries in "North America". e.g., United States, English
- C. Find, for each continent, the average life expectancy of all countries in the continent which has at least 5 countries whose GNP is more than 10000.
- 1. [15 points] For each query above, write a Python script that uses Pandas DataFrames to answer the query. Name your scripts pandas-A.py, pandas-B.py, etc.
- 2. [30 points] Now write a Python script, without using Pandas, to implement each of the queries above. Name your scripts python-A.py, etc.
- 3. [15 points] Write a Python script "load-world.py" that loads the data in the JSON files to your Firebase database, under a node "world". It should look like the following (see parts: a, b, and c). Also create a "country nested" node to embed the languages of the country (see part: d).



- 4. [30 points] Write a Python script to answer each of the above queries using the data in Firebase. You can use the package "requests".
 - For question B, give two solutions: one using the "country_nested" node; the other without.

- Name your scripts: firebase-A.py, firebase-B-nested.py, firebase-B.py, and firebase-C.py.
- 5. [10 points] Analyze the performance of scripts in question 4, using two criteria: the number of requests sent to firebase; and the size of data downloaded from Firebase.

Submission & Execution Format:

Please use <u>Python 3</u> and include all scripts for 5 questions in a zip folder <u>Firstname_Lastname.zip</u> containing following files. All of the results for Q1-Q4 should be printed out directly in terminal. <u>DO</u> **NOT** save results to text file for Q1-Q4.

Q1: pandas-A.py, pandas-B.py, pandas-C.py.

Execution Format: python pandas-A.py country.json city.json countrylanguage.json

Similar for pandas-B.py and pandas-C.py.

(Please include all three json files as arguments to all your scripts in Q1)

Q2: python-A.py, python-B.py, python-C.py.

Execution Format: python python-A.py country.json city.json countrylanguage.json

Similar for python-B.py, python-C.py.

(Please include all three json files as arguments to all your scripts in Q2)

Q3: load-world.py

Execution format: python load-world.py country.json city.json countrylanguage.json

Q4: firebase-A.py, firebase-B-nested.py, firebase-B.py, firebase-C.py

Execution format: python firebase-A.py "North America"

Similar for firebase-B-nested.py, firebase-B.py

Execution format: python firebase-C.py

Q5: please include all of your analysis in one text file

Wrong submission format or execution format will result in the deduction