Tutorial 2

Data Analysis and Visualization
using
Pandas and Seaborn

BT3017

Due date: 31st Jan 2022 (Monday) 1200 hrs (i.e. noon)

Semester 2, AY21/22, School of Computing, National University of Singapore

IMPORTANT:

For this tutorial, you are supposed to submit your project file to LUMINUS.

Instruction for submission:

- *Create a folder using the following naming convention:*

  StudentNumber_yourName_Tut2

- *Follow the submission instructions at the end of this document. Zip your folder. Name your zip file using the following convention:*

  StudentNumber_yourName_Tut2.zip

  For example, if your student number is A1234567B, and your name is Chow Yuen Fatt, for this tutorial, your file name should be A1234567B_ChowYuenFatt_Tut2.zip

- *Submit the zip file in the "Tutorial-2 Submit Here" folder in Luminus.*

Note: you should not need to pay for the website recommended.

Download a dataset containing 120 years of Olympic history from the following link (note: you should not need to pay for downloading this dataset):

https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results

You should see two files: *athlete_events.csv* and *noc_regions.csv*

Use the above datasets that you downloaded from the above link for this tutorial.

Examine the datasets. Answer the following questions:

(a) What are the top three favourite sports of yours?

(b) What are the top three questions you want to ask regarding the data related to at least one of these three sports?

(c) Using Pandas and Seaborn, analyze and visualize the data to help answer your three questions raised in Step (b) above.

(d) Submission instructions:

i. Consolidate your answers to (a), (b), (c) above in a PDF file.
ii. Put all your codes in a Python file that is runnable on its own. You can also submit in Jupyter format.
iii. Put the datafiles in the same directory as your code.