Tutorial 1

Data Analysis and Visualization
using
Pandas

BT3017

Due date: 24th Jan 2021 (Monday) 1200hr (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_Tut1

- *Put your python file and also the results in this folder.*

- *Zip your folder. Name your zip file using the following convention:*

  StudentNumber_yourName_Tut1.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_Tut1.zip

- *Submit the zip file in the "Tutorial-1 Submit Here" folder in Luminus.*

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

Download a dataset containing football results from the following link:

Use the dataset results.csv that you downloaded from the above link for this tutorial.

Write python codes to answer the following questions.

(a) Use pandas to print the column names in the dataset.

(b) Use describe() of pandas to print a description of the numeric columns.

(c) Use pandas to create a new data frame containing only matches with both teams coming from 2 of the 4 countries: Argentina, Brazil, England and Germany.

(d) Use pandas to create a new data frame called **df6** containing only matches played between England and Germany.

   (i)   Do a scatterplot of the data contained in **df6**. Use 'home_score' and 'away_score' as x and y respectively.   Use hue to represent 'home_team'.

   (ii)   What is the maximum number of goals scored by one team in all the England vs Germany matches?

   (iii)   Print a heatmap of home vs away scores for England vs Germany matches.

   (iv)   Write a code segment to compute England's home win rate, away win rate, and total win rate against Germany.