Principles and Applications of Data Science
Homework #2

Due: June 16, 2021

This assignment is to practice how to create a grid of bar plots with tools, for example MatPlotLib, using the data of number of passengers at Taipei MRT stations. We will provide the data file (.csv) and the table for mapping station name to station ID. In the data file, there are five **attributes**, including

- date

- hour

- in: Entrance (station)

- out: Exit (station)

- num: Number of Passengers

With the data file and table, your homework is done according to the following **requirements**:

- Add up the number of passengers at each station.

- Show the top 10 numbers of passengers with bar chart.

- Produce two figures, one for the number of passengers entering the station and the other for exiting the station, for each year (2017 and 2018)

    - Each figure includes 12 bar plots showing the statistics per month.
    - Thus, there are totally 4 figures including 48 bar plots.

- Use Station ID as the tick label in the bar plots.

The provided files of this homework will be put on the **i-school(Plus)** (`https://istudy.ntut.edu.tw/lear` platform of school for you to access.

Homework Submission

- Please compress and upload your homework file named as `HW2_MRT_studentID.zip`(or `.rar`) to **i-school(Plus)** (`https://istudy.ntut.edu.tw/learn/index.php`). The submitted file includes:

    - The source code (.ipynb).
    - 4 Figures names as 2017_enter.png (.jpg or other format)

- The **deadline** is the **midnight of June 16**, 2021 and **Late work** is not acceptable.

- Honest Policy: We encourage students to discuss their work with the peer. However, each student should write the program or the problem solutions on her/his own. Those who copy others work will get 0 on the homework grade.