Principles and Applications of Data Science
Homework #1

Due: May 12, 2021

This assignment is to practice Jupyter/IPython notebook with simple descriptive statistics and run a toy example (Abalone-dataset). Please follow the steps below to have the project on notebook. The generic template (`HW1-Abalone-dataset.ipynb`) of this homework will be put on the **i-school(Plus)** (`https://istudy.ntut.edu.tw/learn/index.php`) platform of school.

Step 1
> Use Pandas (`https://pandas.pydata.org/`) to load Abalone Data Set (`https://archive.ics.uci.edu/ml/datasets/Abalone`) as the dataframe from UCI Machine Learning Repository.

Step 2
> Get 20 data items as sample randomly and show them.

Step 3
> Show 20 data items of Abalone-infant as sample randomly.

Step 4
> Sort the data by Whole weight (grams) and present the top 20 data items.

Step 5
> Show the simple statistical information.

> **Use matplotlib (`https://matplotlib.org/`) to show 2D images about data.

Step 6
> Plot the distribution of three types of Abalone using different colors where x-axis is the diameter and y-axis the whole weight.

Step 7
> Show the proportion of three types of Abalone using pie chart.

About submitting this homework

- Please upload the completed `.ipynb` file with the filename as `HW1_studentID.ipynb` to **i-school(Plus)** (`https://istudy.ntut.edu.tw/learn/index.php`).

- The **deadline** is the **midnight of May 12**, 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.