

Principles and Applications of Data Science

Homework #1

Due: Mar 27, 2019

This assignment is to practice how to use Jupyter/IPython notebook 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 project will be put on the *ischool* (<http://www.ischool.ntut.edu.tw/>) 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 your homework project named as HW1-Abalone-dataset-SID.ipynb to *ischool* platform (<http://www.ischool.ntut.edu.tw/>).
- The **deadline** is the **midnight of Mar 27**, 2019 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.