C S 488/508 Introduction to Data Mining
Homework 0: Python practice

1 Objective

This homework is to let you practice your Python programming knowledge. This is an *individual* homework.

2 Questions

Write python program to do the following.

(1) (10 points) Read in the **titanic_train.csv** dataset using functions in Pandas package and print the data in the first two rows. The dataset can be downloaded from Canvas data folder.

(2) (10 points) Print the names and data types of all the columns.

(3) (10 points) Calculate and print the number of rows and columns that this dataset contains. Note that this file's first row has all the column names. When you calculate the number of rows, you should exclude the column-name row.

(4) (10 points) Calculate and print the distinct values of the last column.

(5) (10 points) Calculate and print the distinct values of the "Survived" column.

(6) (10 points) Calculate and print the number of rows when the value of the 2nd column is 1.

(7) (10 points) Calculate and print the number of rows that do not have missing values in the 6th column.

(8) (10 points) Calculate the average value of the 6th column ("Age" column). The rows with missing values should be discarded in the calculation.

(9) (10 points) Calculate the average value of the 6th column ("Age" column) when the 2nd column has value 1. The rows with missing values should be discarded in the calculation.

(10) (10 points) Draw a scatter plot with the data of the 6th column ("Age" column) and the 10th column ("Fare" column). The y axis represents the 10th column and x axis represents the 6th column. Show the points in different colors and shapes when the values of the 2nd column ("Survived" column) are different. The rows with missing values should be discarded.

General requirements

- Put proper comments in your code.
- Try to **avoid using loops**.
- The lines of code you need to write to answer each question are generally 2-5 lines. Do not make the questions unnecessarily complicated.
- Put the code for all these questions to one file. Properly organize the code to make grading easy.

3 Submission instructions

A zipped file `hw-lastname.zip` consisting of all the code.

4 Grading criteria

(1) The score allocation has been put beside the questions.

(2) Please make sure that you test your code **thoroughly**.

(3) FIVE points will be deducted if files are not submitted in the required format.