**Lab 1.1 Data Exploration**

Data Exploration is always a crucial work to do at the initial stage, before any further analysis. Via data exploration, you can better understand your data, and thus execute some essential data pre-processing to clean your input data. As for figures, you can help others quickly seize the information about your data by proper data visualization.

In the first part of lab 1, you would be asked to explore the data from the following dataset from UCI: <https://archive.ics.uci.edu/ml/datasets/glass+identification>

, which is a dataset about ***classifying different glass types*** by *refractive index (折射率)* and several *metal oxide contents (金屬氧化物含量)*.

The following questions are about data exploration of this dataset. Please answer them in the form of a **short report** (**2 pages at most**). You are allowed to write the report in either **English** or **Chinese**. More detailed submission rules will be given at the end of this description.

**Questions (totally 50%)**

**Q1.** How many observations are in this dataset? How many attributes (including the target variable: types of glasses) are in this dataset? (2%)

Also, show the first 10 observations in this dataset, but please **remove the first id column and add the title row before showing.** (5%)

**Q2.** Is there any missing value in any column? (3%)

**Q3.** Among all 8 kinds of metal oxide, which of them has the highest mean? Which of them has the smallest standard deviation? (5%)

**Q4.** Which kind of plot is more suitable for drawing the target variable (glass types), histogram or bar chart? Briefly explain your thoughts and draw it out. Also, what do you see from the plot? (15%)

**Q5.** Draw out the **distributions plots**, **boxplots** for each variables, as well as plots that can demonstrate the **pairwise correlation** between each 2 variables. Briefly explain what do you see in these plots. (20%).

**Submission Rules**

**Due day:** 23:59:59 on **March 30, 2020**. **Late submission is not allowed.**

**Files to hand in:** (1) A short report (**2 pages at most**) in the format of:

**STUDENTID.pdf** (ex: 310554004.pdf).

(2) Data Exploration Code written in **Python**.