**Lab 1.2 Data Analysis**

Now that you have completed data exploration, you can proceed to execute the following data analysis. For the **classification** problems, there are some widely used *machine learning* algorithms, including logistic regression, decision tree, random forest, SVM, KNN and Naïve Bayesian Classifier. Some easy vanilla neural networks might do the trick as well. When your analysis is done, you will also need to show the results. A common way is to calculate and draw the confusion matrix, which can be further used for deriving other important metrics such as accuracy, sensitivity(recall), precision and F score.

**Questions (totally 50%, bonus 10%)**

**Q1.** Split the dataset into training dataset and testing dataset, with explanatory variables and target variables separated. Show the shapes of the 4 subsets. (5%)

**Q2.** Try at least 3 different algorithms for classifying the glass types. Briefly describe what you did. (30%)

(Note: Freely choose, not limited to the methods mentioned above)

**Q3.** Draw the confusion matrix and calculate the accuracy, precision, recall and F score on the **testing** data, for the method with the **best performance** you used in question 2 above. (15%)

**Q4. (Bonus)** What else did you do to try making the predictions better? (10%) *It's totally fine that these tricks didn't improve the performances of your models. Just list out what techniques you tried. I might give you some bonus points based on these efforts.*

**Submission Rules**

Please answer the above questions in the form of a **short report** (**4 pages at most**). You are allowed to write the report in either **English** or **Chinese**

**Violation of these rules will lead to -5 points‼**

**Files to hand in:** (1) A short report in the format of: **STUDENTID.pdf**.

(2) Data Exploration Code written in **Python** and **name it as your student ID number as well**. That means: **STUDENTID.py** or **STUDENTID.ipynb**

**You don't need to zip your file.**

**Due day:** 23:59:59 on **May 1, 2022**. **Late submission is not allowed.**