

# Data Analytics

## Academic Year 2023-24

### Course Assignment N. 12: Wine Dataset

Prof. Fabio Crestani

TAs: Navdeep Singh Bedi, Lili Lu

For this assignment you will work in small groups to carry out simple tasks of data analysis given a specific dataset. The goal of this assignment is to use Python and complementary libraries on a given dataset to *explore* and *analyze* the given data and *draw conclusions*.

#### Description

The [Wine Dataset for Clustering](#) contains various attributes such as alcohol, malic acid, ash, alcalinity of ash, etc. The attributes are found in each of the three types of wines. Your goal is to cluster the wine based on their various attributes.

Your tasks are to:

- Explore and describe the data (*i.e.*, standard descriptive statistics, visualize the variables with different graphs, draw distributions and histograms of variables, are there outliers? Any interesting observation? Any correlations? Etc.)
- Pre-process the data (*i.e.*, handle and fill unknowns if there are any, etc.)
- Use at least two different clustering algorithms and compare them. What is the most optimal number of clusters? (Supposed you do not know it.)
- Evaluate and compare the accuracy of the different models.

#### Submission procedure and evaluation

You should produce a report of your work and its evaluation along with the source code. It will be a concise explanation of how you tackled the different tasks, the reasons of your choices, successive conclusions, graphs you produced, results of the decisions and their accuracy, *etc.*

Use Jupyter Notebook to produce results of the commands in a single .ipynb file. For more information check: <https://jupyter.org/documentation>

The report (max 5 pages) and the code of the project need to be submitted via iCorsi.

Please, upload all the required items in a single file and name it following the structure: **no\_Project.[zip|tar.gz|7z]**. For instance, 05\_projectname.tar.gz