



## Objective

This time we want to support your CP2 submission! We prepared a tutorial to help you to understand how you can pre-process your database using two different tools: the Pentaho Data Integration (Kettle) and the Pandas library (Python). Follow the tutorial and then clean a new dataset to reinforce your learning!

## Requirements

Look at the list of videos published in the Lab Classes page in Moodle. You must watch the *Pentaho Data Integration Tutorial* and the *Pandas Python Library Tutorial*. Then, you must choose one of those tools and install either the pandas library in your Python environment or the Pentaho Data Integration CE 9.2 version (<https://sourceforge.net/projects/pentaho/>). We have included a guide in the Lab Classes page to help you troubleshoot any issues you face. There are already a lot of datasets in the Lab Classes page. You can train your skills with those.

## Task

Each class shift has their own exercise. Check the list below to understand which exercise you must do.

- **VIL03:** Clean the Income by Country dataset. Produce a dataset with the average GDP per capita per country identified by its ISO Code.
- **VIL04:** Clean the WHO Suicide Statistics dataset. Produce a dataset with the average suicides\_no per country identified by its ISO Code.
- **VIL05:** Clean the Wine Reviews dataset. Produce a dataset with the average price per country of origin identified by its ISO Code.
- **VIL06:** Clean the Income by Country dataset. Produce a dataset with the average GNI per capita per country identified by its ISO Code.
- **VIL07:** Clean the WHO Suicide Statistics dataset. Produce a dataset with the average population per country identified by its ISO Code.
- **VIL08:** Clean the Wine Reviews dataset. Produce a dataset with the average points per country of origin identified by its ISO Code.

## Tasks to perform during the lab

The group will present the task in the laboratory to the professor. Each student in the group needs to attend the class to be granted the grade. **The professor will also ask each group member to complete an individual task.** These individual tasks will require each student to tweak the code/project

presented to implement Task II. Keep in mind that the individual evaluation depends on the quality of the answers.