

## ESILV - Python for data analysis

### Scoring method

Handover date: January 10 midnight

You are assigned a set of data to analyze, on the page of [attribution of datasets by student: https://docs.google.com/spreadsheets/d/1CHVkt2Bk7RBycCZypeMLjxcM6U2MynypbGFFWEfKf4 / edit? usp = sharing](https://docs.google.com/spreadsheets/d/1CHVkt2Bk7RBycCZypeMLjxcM6U2MynypbGFFWEfKf4/edit?usp=sharing) .

**You are asked to analyze the data set assigned to you and to do from there:**

1. A powerpoint explaining the ins and outs of the problem, your thoughts on the question asked, the different variables you created, how the problem is situated in the context of the study, etc: **25%**
2. A code in python:
  - at. Data-visualization (use matplotlib, seaborn, bokeh...) - show the link between the variables and the target: **25%**
  - b. Modeling - take scikit-learn try several algorithms, change the hyperparameters, do a search grid, compare the results of your models in graphs: **25%**
3. Transformation of the model into a Django API: **25%**

**An email + a github + the filling of a google drive sheet are expected:**

To answer you will need to:

1. Upload your work to your github account:
  - at. A Readme summarizing the task to be performed and your conclusions
  - b. The pdf version of your powerpoint
  - vs. The code in jupyter notebook format
  - d. The django or Flask API
2. Put the link to your github on the following page:  
[https://docs.google.com/spreadsheets/d/1CHVkt2Bk7RBycCZypeMLjxcM6U2MynypbGFFWEfKf4 / edit? usp = sharing](https://docs.google.com/spreadsheets/d/1CHVkt2Bk7RBycCZypeMLjxcM6U2MynypbGFFWEfKf4/edit?usp=sharing)
3. Send an email to [romain.jouin@gmail.com](mailto:romain.jouin@gmail.com) + the teacher you had in class, with:
  - at. Topic : " **ESILV - Python for data analysis - assignment 2021** "
  - b. Contents:
    - Last name First Name
    - Github link
  - vs. Attachment
    - Ppt (the same as on the github)

The professors remain at your disposal for any possible questions, and also try to contact the researchers responsible for the data sets: there are often their emails in the descriptions, and they will no doubt be happy to see that their work is reused, and we will see if they answer.

The main job of the data scientist is to understand data, and daring to ask questions is an essential quality in this profession.