Python for Data Analysis

ESILV S1 2023–2024

[Project]

Please find your assigned dataset on the following link:
https://docs.google.com/spreadsheets/d/1kaQUR8rA9Kbk93ZPM5Blid9pjfM6BZts6I6Odhr8Efc/edit?usp=sharing

To analyze the dataset, you should answer the following questions:

1. A PowerPoint explaining the ins and outs of the problem, your thoughts on the asked question, the different variables you created, how the problem fits in the context of the study, etc: 25%.
2. A code in python:
    a) Data pre-processing: encoding, normalization, imputation… 20%.
    b) Data visualization (use matplotlib, seaborn, bokeh ...): show the link between the variables and the target: 20%.
    c) Modeling: use the scikit-learn library to try several algorithms, change the hyper parameters, do a grid search, compare the results of your models using graphics: 20%
3. Transformation of the model into an API of your choice (Django or flask): 5%.
4. Bonus : 10%

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

1. Put your work on your GitHub account:
    a) A readme summarizing the task to be accomplished and your conclusions.
    b) The pdf version of your PowerPoint.
    c) 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/1kaQUR8rA9Kbk93ZPM5Blid9pjfM6BZts6I6Odhr8Efc/edit?usp=sharing
3. Send an email to your teacher that you had in class, with:
    a) Subject: "ESILV - Python for data analysis - project 2023/2024".
    b) Content:
        ▪ Last name / first name
        ▪ GitHub link
    c) attached file: Ppt (the same as on GitHub).

The professors remain at your disposal for any question. If needed, try to contact the responsible of the dataset (via email in the dataset description).