Lab Activity 17-12-2019 - Simulation of lab exam

Find the clusters in the included dataset.

The solution must be produced as a Python Notebook. The notebook must include appropriate comments and must produce:

1. the boxplots of the attributes and a comment on remarkable situations, if any (**2pt**)
2. a pairplot of the data (see Seaborn pairplot) and a comment on remarkable situations, if any (**2pt**)
3. a clustering schema using a method of your choice exploring a range of parameter values (**5pt**)
4. the plot of the global *inertia (SSD)* and *silhouette index* for the parameter values you examine (**4pt**)
5. the optimal parameters of your choice (**4pt**)
6. a pairplot of the data using as *hue* the cluster assignment with the optimal parameter (**3pt**)
7. a plot of the silhouette index for the data points, grouped according to the clusters (**4pt**)
8. A sorted list of the discovered clusters for decreasing sizes (**7pt**)

The python cells must be preceded by appropriate comments. Useless cells and pieces of code will be penalised

Naming style of variables must be uniform and in English.

Bad indentation and messy code will be penalised.

The notebook must contain in the first cell the first name and last name of the student.

The notebook name must be lastname_firstname.ipynb.

The solution must directly access the data in the same folder of the notebook.

The candidate can freely access the manuals available online in:

scikit-learn.org
docs.scipy.org
pandas.pydata.org
matplotlib.org
seaborn.pydata.org
The candidate can freely access the teaching materials available in the course website, including examples of python notebooks.

The notebook must be uploaded in **both** *original* and *compressed* form, as two separate files.