Exam for Machine Learning Python Lab

Consider the file `income.csv`, explore the data, drop the columns that you consider useless for the classification and find the best classification scheme.
The solution must be produced as a Python Notebook, assuming that the dataset is in the same folder as the notebook.

The notebook must include appropriate comments and must operate as follows:

1. Load the data file and explore the data, showing size, data descriptions, data distributions with boxplot, pairplots **2pt**
2. Comment the exploration of step 1 pointing out if there are imbalanced distributions, outliers, missing values **2pt**
3. Drop the columns that are not relevant for the classification operation, if any, and explain why you do that.
   Deal with missing values, if any **4pt**
4. find the best classification scheme considering two classification methods, find the best hyperparameters using cross validation; the optimization must be focused on the f1_macro measure **4pt**
5. Show the performance measures and the confusion matrices for the best hyperparameters of each model **2pt**
6. Comment the results **2pt**

Quality of the code **4pt**

- Include appropriate comments with reference to the numbered requirements
- Useless cells, pieces of code and non-required output will be penalised
- Remove the code you use for testing and inspecting the variables during the development
- Naming style of variables must be uniform and in English
- Bad indentation and messy code will be penalised
- Non generalised solution, such as three sequential statements with the same kind of operation instead of a loop, will be penalised

Additional directions, the assignments not compliant with the rules below will not be considered:

- The notebook name must be `youremailusername.ipynb` in lowercase letters (underscore instead of dot inside the email username can also be accepted
  E.G. if your email is `mario.rossi45@studio.unibo.it`, the notebook filename will be `mario.rossi45.ipynb` (`mario_rossi45.ipynb` can also be accepted)
- The solution must directly access the data in the same folder of the notebook, the name of the file must be the same as the file provided. If the notebook is developed using *Google Colab*, the code must be able to work also out of the Google Colab environment without any change.
- Upload the notebook only to `http://eol.unibo.it` in the activity specified by the teacher, any other way of submitting the notebook will be ignored

Cooperative work will be heavily sanctioned
The candidate can freely access any kind of materials.