

# Guidelines for the project report

- Be sure the following information is present
  - Context
    - What is the dataset about? How has it been used?
    - What is the goal (or the goals) you want to pursue with your analysis
      - You should prevent your report from being a sequence of “I applied this and this is the result. Then, I applied this other algorithm, and this is the result, etc.”
      - Before the application of any algorithm, explain what you want to do
    - Characterize and understand your dataset
      - ... via data exploration
      - ... but do not report everything you found: only report what you think is useful or interesting to observe.
      - Explain the meaning of the columns and the values
    - Explain your data-science protocol clearly
      - What pre-processing you did and why?
      - How do you get your training and test data? Which operations do you perform on the former and on the latter?
    - Analyse the results
      - Do not just show “here are the results”.
        - Tell if you expected them to be like they are. Or if they surprise you? Try to get some explanations of why you get these results. Why performing some modification to the algorithms’ hyper-parameters gave you better results, etc.
      - Link the results
        - The report becomes more interesting if you include your analysis path. For instance, you might say that you first applied a certain method, but then you observed a certain thing, which gave you the idea of doing some additional pre-processing or changing some hyper-parameter, etc.

## Be critic with your project

- You are not trying to sell a project. So, try to openly find the limits of your project
- Try to critically assess whether the methods explored would be feasible in real situations

## Add only important information

- Avoid too much writing.
- Clean the information before putting it in the report.
  - For instance, avoid to copy and paste the output of a Python notebook as it is. Only take the informative part. As another example, if you have categories encoded in

some integers numbers, prefer to show the name of the categories instead of numbers.

- If not needed, avoid using too many decimal numbers. For instance, instead of writing 5.7598766%, just 5.8% would be enough in most cases! For the sake of clarity, you may use colours (e.g., green, yellow, orange, red to indicate whether the numbers you show are good or bad)
- Be sure that the pictures or graphs you add are readable (fontsize is large enough) and of good resolution