Final Project Guidelines

"If you torture the data long enough, it will confess" – Ronald Coase

The goal of the project is to assess your ability to use python in the development of a data science project. Since this is an open topic assignment, it is also expected that the specific steps taken in the elaboration of the project will be different from group to group.

As a Data Scientist you will often be challenged to bring new useful insights to problems or questions that are relevant to the organization you are working for. More often than not you will resort to data to obtain such insights. Tasks that you might find yourself doing include: filtering, cleaning and merging data from multiple sources; classify information; clustering; build regression models; train machine learning algorithms; identify non intuitive correlations between features in the dataset; test scientifically your findings; design meaningful visualizations; and more importantly construct a narrative to report your findings.

Use this project as an opportunity to show and advertise your creativity in developing relevant data science projects. You will be able to carry the final project with you as part of your portfolio, it can be a great opportunity to show your next prospective employee your skills, or to approach your prospective thesis supervisor.

Check the document "Online Resources.docx" for links to examples of data science projects developed by undergrad students in data science courses.

Important Dates and Deliveries

Date: 8th of October at 23:59
Delivery: Project proposal.
Download and fill the proposal template that is available in Moodle.
This will help us to understand which groups might need help, and which groups would prefer to work in pre-defined projects. Please make one submission per group, and groups of at least three members.

Date: December
Deliveries:

1. A report (in Word or PDF formats) with a maximum of 3500 words plus References, Figures and abstract;
2. A Jupyter notebook (ipynb) with all the code used in the elaboration of the project, with the appropriated comments and documentation;
3. A representative sample of the dataset that can be used for reproducibility;
4. The report should include the following elements:
    a. Title;
    b. List of Authors;
    c. Group Number;
    d. Introduction;
    e. Data Description;
    f. Results and Discussion;
    g. Conclusions;
    h. References;
    i. Statement detailing the contributions of each member.

Date: December/Januart
Oral Presentation: Prepare a 10 minute presentation plus 5 minutes for discussion. All members of the group should be prepared to explain a different section of the project.
Date: Third Week of December (to be defined)