First trimester courses

Master on Visual Tools to Empower Citizens

Universitat de Girona
2020-2021

Karina Gibert, Marc Comas, Matt Osborn, and Anton Bardera

Long Term practical work

Elaboration of a complete data science process to extract knowledge from a real data set

Final Presentation Guidelines

You will work in groups of 3 or 4 persons and apply the contents of the course to the reference dataset. Each group works independently.

A report will be delivered by Friday, January 8th, 2021 at 12:00 M.

A public presentation of each work will be done on Friday, January 8th, 2021 at 17:00 PM.

Common discussion of the different approaches followed by the working teams on the same dataset will be held as the final part of your training program in this course.

In the following, details on the different aspects of the final presentation are detailed:

- The presentation duration will be 40 minutes plus 20 minutes of questions and discussion
- You will have to describe the whole data analysis pipeline: statistical analysis, data engineering components, and visualization
- Slides are strongly recommended, although you can show some other kind of supports during the presentation (Python code and execution, Jupyter notebooks, Observable notebooks…)
- The presentation will be composed of the following parts:
    1. Introduction and problem description
    2. Data sources and preparation
    3. Statistical analysis and data mining methods
    4. Data engineering components (cloud services used, where data is being pulled from, where it is being uploaded to, etc.)
    5. Data and models visualization
    6. Conclusions
- At each one of the project steps, you have to specify the main challenges or difficulties encountered and how they were addressed
- Probably you won't have time to explain all the elements on the final report, so you will have summarize the work
- All the team members must participate on the presentation