Data Science Workshop 2016/17: Project Guidelines

Project phases:

1. Explore the datasets of the <u>World Data Bank</u>.
2. Choose a goal, related to one (or more) of the <u>Millennium Development Goals</u>.
3. Formulate a "data science" hypothesis, according to your goal.
4. Build a data model (preprocessing):
   1. Cleaning / missing data handling
   2. Feature selection and/or formulation
5. Implement/Choose the correct ML algorithms and methods over the data model to output results.
6. Evaluate your model statistically and use your results to improve the feature selection and parameter tuning.
7. Present your workflow and results in a user friendly Jupyter Notebook (or an *approved-by-staff* alternative).

Project administrations:

1. Teams will use Trello to divide and manage their tasks and track them, and be ready to present it if asked to.
2. Teams will use GIT for collaborative development, and be ready to present it if asked to

Submission:

1. A link to the Github repository that includes:
   1. *all* of your code
   2. Your data
   3. The Jupyter Notebook, containing all phases of the project - from cleaning to evaluation.
2. A documentation file (8 pages maximum, 1.5space, font size 12, PDF Only) describing your work.
   - All your choices (algorithms, parameters, features, cleaning ,evaluation) should be explained.
   - If you use external source - cite it.
   - Should include your final insights