

# DS-GA 1019 Advanced Python for Data Science Projects

## 1 Goals of the projects

Students within a group will decide on a problem that they would like to work on. During “project proposal presentations” (Lecture 7 on March 6th), students will present the topic and goals of their project to the instructor/class. The instructor will then schedule and meet with groups whose proposals are either overly or insufficiently complex. The final project presentations will take place during the last lecture, where students will present the problem they worked on as well as the obtained results.

The proposal and final project will be graded based on:

- Demonstrated understanding of advanced Python concepts
- Demonstrated understanding of the final project problem
- Quality of documentation: **one report of maximum 4 pages**; for final project only
- Quality of the presentation: clarity of the exposition, slides quality, and duration (the groups should respect the time slot, 5 minutes with extra 1-2 minutes for questions).

## 2 Examples of project topics

The topics of the projects should be related to data science in some way.

- Fast numerical optimization tools/solvers for your favorite mathematical models and tools for data science;
- Social Network Analysis;
- Twitter analysis;
- Dynamical Processes: Covid modeling, market adoption, etc.;
- Efficient or large-scale learning of some neural networks architectures;
- Processing and/or statistics on high-dimensional data: e.g. fast or memory efficient algorithm on: images, audio, biological data, or other datasets/data streams;
- NYC Taxi fare prediction: <https://www.kaggle.com/c/new-york-city-taxi-fare-prediction/data>
- Stock Portfolio Optimization;
- Efficient forecasting over large datasets or streams;

- Stochastic models for predictions of: road/train-system delays; product adoption; spread of epidemics;
- Improvement of some of your previous machine learning projects by using the tools from this course;

### **3 Student groups**