

CS 109A/AC 209A/STAT 121A Data Science: Project Guidelines

Harvard University

Fall 2016

Instructors: W. Pan, P. Protopapas, K. Rader

Overview:

The course project is your opportunity to explore a topic in depth; you get to choose from a list of topics provided for you from broad fields from to sports analytics. You will work in groups of 2 to 4 students. If you are looking for project partners, we suggest you post on Piazza.

The final products of the project will be a website and a poster to be presented at one of two poster sessions.

Requirements, Milestones, and Deadlines:

1. Milestone #1 due **Tues, Sep 20**: Project Selection. Form a team of 2 to 4 students and select your top 3 to 5 choices, in rank order with 1 being top choice, for a project from the provided list (found here: <https://canvas.harvard.edu/courses/12656/pages/projects>).
2. Milestone #2 due **Tues, Oct 19** Literature Review. As a team, submit a 1-page (or less) summary of each of two readings related to your selected project (up to 2 pages total).
Finalize your vision of the project with your TF and write an updated project summary (included in the 2 pages)..
3. Milestone #3 due **Tues, Nov 1**: Data Exploration. Collect the relevant data and submit a document with 10-12 visuals and 1 to 2 page write-up of key findings. Some summary measures should be submitted as well; be smart to only include what helps you explain what has been noticed.
4. Milestone #4 due **Tues, Nov 15**: Baseline Model. Fit a specific model provided in your project topic's guidelines, provide the results from Python along with a roughly half page write-up of what you notice. Visuals are helpful here too.
5. Milestone #5 also due **Tues, Nov 15**: Proposal of Future Work. Provide a roughly 1 page outline of what approaches, models, etc. you would like to use for the final project results.
6. Deliverable #1 due either **Thurs, Dec 8 or Fri Dec 9**: Poster presentation: your group will present your project with a poster at one of 2 poster sessions (required poster size: 36" x 48" either using a tri-fold board (like from Staples) or *professionally* printed). The poster should have large enough font to be read from at least 5 feet away. More details to come.
7. Deliverable #2 due **Wed, Dec 14**: Website: design a website to communicate your final results and models used. You are welcome to update your approach from your poster based on comments received. More details to come.

General Advice:

- Each project topic will have a TF assigned to it (possibly 2). This will be your contact person for questions...preferably through use of TF-led office hours.
- There will be at most 1 class after Thanksgiving and no problem sets due after the second midterm. So the entirety of the class will be focused on the project at that point.
- A picture is worth a thousand words: often times a wisely chosen visual helps illustrate your results better than any text ever could.
- If you like, you can think of organizing your project like a scientific paper, which generally include sections for intro and motivation, methods, results, and discussion. This organization is not required.
- The content, approach, and interpretation is much more important than bells and whistles in both the poster and the website.

Grading:

Each member of your group will receive the same grade on the following components:

- **1/4 of grade** - Milestones along the way.
- **1/4 of grade** - Poster and presentation during poster session. The poster will be worth 2/3 of this portion and the presentation/discussion during the poster session will be worth 1/3 of this portion.
- **1/2 of grade** - Completed website. Content and communication of results are more important than design.