

Projects

This semester, Penn Data Science Group (PDSG) will be hosting a data science project competition. Any member of PDSG can participate. To become a member, simply fill out the signup form on penndsg.com/join and choose “Yes” when it asks if you’re interested in working on a team project. Projects must be done in teams of 2-4 people. Any project that aims to answer some question or otherwise extract value from a dataset (or multiple datasets) is valid.

The primary form of communication regarding team projects will be via the [PDSG Piazza account](#). Every PDSG member participating in team projects will be added to the account. Here are some of the main uses for the Piazza account:

1. **Finding team members.** There is “Search for Teammates” widget on Piazza. You can mark how many team members you’re still looking for, any pertinent information about yourself, like a type of project you want to work on, and what type of team members you’re looking for.
2. **Asking for help with your project.** You can post as yourself or anonymously to ask questions either to everyone or just to the “Instructors”, who in this case are the PDSG Board members.
3. **Providing feedback.** You can post feedback as yourself or anonymously to let us (the board members) know if you have suggestions, or if there are problems or complications with your team. We will respond to these as soon as we can.

We will be offering \$50 Amazon gift cards to each member (excluding board members) of winning teams. We guarantee at least 2 winning teams, provided there are at least 2 eligible teams, and we may have more depending on funding. Here are the eligibility requirements:

1. Teams must have 2, 3, or 4 members
2. Your team must post project idea and group members on Piazza by February 9
3. Your project must not be a research project already being done for an advisor.
4. Your project must not be a project that also counts for a class project.
5. Your team must submit a short report to be put up on the PDSG website by April 6.
6. Your team must present the results of your project in front of a panel of judges around April 20.
7. Your team must be chosen as one of the top projects presented to the panel of judges.

The criteria for judging the projects will be:

1. Novelty: Is this a new analysis, or something mostly from a tutorial?
2. Validity: Was the analysis meaningful and done correctly?
3. Clarity: Was the project presented clearly and effectively?
4. Impact: How does/could this project contribute to society?

If you don’t know what kind of project you want to do, a great place to start is a Kaggle competition. Kaggle competitions are great for starting out with machine learning and data science because they provide clean data sets and a clear objective. Each competition lets you

submit multiple predictions, and your prediction is evaluated immediately so you can easily compare performance of different models. Each competition also has a forum for questions and answers associated with it, as well as “kernels”, which are scripts that other users have uploaded that you can run and tweak right there on the website. We picked 5 completed Kaggle competitions that we feel are good places to start with interesting datasets.

1. [Titanic: Machine Learning from Disaster](#) -- Predict Titanic passenger survival based on data about passengers
2. [Digit Recognizer](#) -- Classify handwritten digits from famous MNIST data
3. [Shelter Animal Outcomes](#) -- Predict Austin Animal Center animal outcomes
4. [TalkingData Mobile User Demographics](#) -- Predict demographics characteristics from mobile phone data
5. [Predicting Facebook Check-ins](#) -- Predict where people like to check in on Facebook based on fictitious data similar to data Facebook employees use

Raffle Prizes

To encourage a broad range of participation, we will offer raffle tickets for participation, and raffle off prizes at the end of the semester. The prizes will be \$5 gift certificates to Amazon or Starbucks or other similarly priced items. To do so, we will draw raffle tickets one at a time, without replacement, giving a prize to whoever owns each raffle ticket, until either all prizes are gone or all raffle tickets are gone. Here’s how to earn raffle tickets:

1. Submitting a project report to go on the website earns 1 raffle ticket.
2. Giving a student talk at a PDSG event, whether it’s a presentation about your team project, an internship experience, research experience, or other side project experience, earns 1 raffle ticket. For team project presentations, each team member who presents earns 1 raffle ticket.
3. Introducing the PDSG board, via email or in person, to a data professional who ends up giving a career talk for PDSG, earns one raffle ticket to the student who referred the data professional.
4. Leading a workshop earns 2 raffle tickets.