Day 13: Project Proposals and Continuing to Learn PfDA

Raymond Yee

March 5, 2013 (`http://bit.ly/wwod1313`)

Agenda

- Keep on truckin' with Python/Pandas
- Projects

Event tomorrow

Software is Rebooting Journalism: Data Mining and Visualization in the Public Interest

Wednesday, March 6, 2013 – 306 Soda Hall (HP Auditorium) 4:00 - 5:00 pm / 3:30 - Refreshments will be served

Jeff Larson News Applications developer, ProPublica

ABSTRACT:

ProPublica is a non profit news outlet in New York City that publishes investigative journalism. They also have a team of seven News Application developers who create software that tells stories and finds meaningful stories in data. Their News Applications include databases such as Dollars for Doctors, which lists pharmaceutical company payments to doctors, and the Message

Homework solutions Posted / Getting Help

I've uploaded the answers to Day 10 homework, which was due on Friday:

- Day_10_A_fixed_width_parsing_completed.ipynb
- Day_10_freebase_cursor_completed.ipynb

BTW, you can follow commits on github for the working-open-data repository.

I'm backlogged on grading homework. I've posted homework solutions for all the problem sets.

I count on you as graduate students to contact me when you need help. Please don't hesitate to talk to me, come to my office hours, and post questions to Piazza.

Continuing Learning Pandas

We've taken a first pass at the major pieces of `PfDA` but we now need to get conversant by practice – on the examples in `PfDA` and increasingly on the data sets in the projects.

Good lectures I can commend on YouTube:

- Introduction to NumPy and Matplotlib
- Advanced Matplotlib

Working with Data Sets in PfDA

`https://github.com/pydata/pydata-book`

For example:

- calculating the most common domain names in usagov_bitly_data2012-03-16-1331923249.txt
- verifying that the timestamps represented in the file do actually fall on 2012-03-16

Think of questions you have for the 3 data sets in Chapter 2, and we'll come back to seeing whether/how to answer them:
`http://bit.ly/wwod13dataq`

Project Proposals due this Friday

Due by Friday, March 8 at noon: a project proposal that will serve our learning goals and a solid effort at answering questions listed in Project Proposal Template.

I hope that your project is a topic you find fascinating, because it will help with your learning!!

Answer the following questions in
`http://bit.ly/wwod13projideas`

- What's the title for your project?
- Who are your team members (2-4 members)?
- What are some open-ended questions of your project?
- What are some concrete immediate first/next steps in your projects?
- What open data sets do you plan to work with? (Show that the data set(s) you plan to work with are indeed open.)
- What are possible ties to Wikipedia (or any Wikimedia related projects)?
- What challenges do you anticipate running into?

Work on Projects Today

- Making sure everyone has identified a project.
- Helping every project identify a a specific data set to start working on and a specific next step.
- Having project teams talk to each other

Homework

Project Proposals