Day 11: Project Proposals and Pandas!

Raymond Yee

February 25, 2013 (`http://bit.ly/wwod1311`)

Agenda

Our priorities:

- Make progress on project proposals
- prepare for visit from Harrison Dekker and Tim Dennis of the UC Data Lab, who will help us learn how to use the data resources of UC Berkeley.

Visit from UC Data Lab folks

Add your questions and data sets to data sets you are looking and and any questions you have for Harrison and Tim

Projects and Pandas

I want everyone to identify a **specific** data set tied to a project idea today on Project Ideas/ Proposals for WwOD 2013:

- Your project is listed under the Project Ideas section
- You link to the project in the Who's Interested in What? section under your name.
- By specific data, I mean something you can meaningfully use what we have learned to produce a pandas `DataFrame` and answer a useful (even if simple and preliminary) question. List the data source in the project idea section.

Reminder: Project Requirements:

- 2-4 people.
- Must involve in a non-trivial manner some open data – though you are welcome to use closed data also
- Deadline for project proposal/team formation is Friday, March 8, 2013 at noon

Refining project proposals

To the questions I've already posed about projects:

- ▶ What open data sets might you be working with?
- ▶ Possible ties to Wikipedia (or any Wikimedia related projects)?
- ▶ What challenges do you anticipate running into?
- ▶ What skills will a team need to develop to solve this problem?

I'll add two more to consider:

- ▶ What are some open ended questions of your project?
- ▶ What are some immediate next steps in your projects?

If you are looking a project to work on. . . .

I have some project ideas:

- Bach!
- Books!
- Berkeley!
- Brian's Bailiwick! (i.e., CourtListener)
- Bisecting Big Brother! (org chart for the US Federal Goverment)
- Building Blocks!

Homework

We will work on this in class today.

data questions

- ► Add your questions and data sets to data sets you are looking and and any questions you have for Harrison and Tim

project selection and refinement

I want everyone to identify a **specific** data set tied to a project idea today on Project Ideas/ Proposals for WwOD 2013:

- ► Your project is listed under the Project Ideas section
- ► You link to the project in the Who's Interested in What? section under your name.
- ► By specific data, I mean something you can meaningfully use what we have learned to produce a pandas `DataFrame` and answer a useful (even if simple and preliminary) question. List the data source in the project idea section.