Homework 7

PS 398 - Computational Frameworks for Social Science

This homework is designed to get you started using databases to store your data.

Problem Description

In this homework you are going to refactor your previous two assignments such that they store information into a database. Specifically:

- **HW5** should store all of the info that you previously outputted to a table called "scrapes". You should add an additional column with a reference to a *Source* model that has the name and URL of the source that you crawled.

- **HW6** should store the twitter network that you crawled into a local DB, i.e. the user information in a *Users* model. You should have a separate table with a relation to Users that is called "crawls" which specifies a unique crawl and is associated with the attempt to crawl the network (as this information changes over time and we want snapshots). You should at a minimum store the starting user_id for you crawl and the time you initiated it.

Spend some time before you start to sketch out what the schema of your tables will look like. As you refactor, think about how you might have written your code originally to more easily facilitate swapping out the data storage layer.

One nice thing about schemas is that they are a specification for how the data should look. So long as your code conforms to the schema then all of homework assignments could theoretically be storing their data in some shared data repository.