March 5 Gro Hackathon - Problem Definition

The Data Source

The source data we’ll be working with is the NASS dataset from the USDA:

* Main Site: <http://www.nass.usda.gov/>
* Online Data Browser: <http://quickstats.nass.usda.gov/>
* Development API: <http://quickstats.nass.usda.gov/api/>
* FTP Download: [ftp.nass.usda.gov/quickstats/](http://ftp.nass.usda.gov/quickstats/)

.

The Hack

Areas of focus in the dataset for this hack

* Sector to use: Crops
* Lowest region level: County
* Data coverage: The app should be able to harvest all the data in the Crops sector from any given start date to any given end date

How to structure your app

Your project must have at least one file named ***harvest.py*** which will be the main entry point and should be able to be called from the command line with the following options:

1. **--start\_date**=%Y-%m-%d
2. **--end\_date**=%Y-%m-%d
3. **--database\_name**=
4. **--database\_host**=
5. **--database\_user**=
6. **--database\_pass**=

The code must retrieve the data within the provided range from the website and store it in a **PostgreSQL** database table called **fact\_data** with the following columns from the source data:

1. DOMAIN\_DESC
2. COMMODITY\_DESC
3. STATISTICCAT\_DESC
4. AGG\_LEVEL\_DESC
5. COUNTRY\_NAME
6. STATE\_NAME
7. COUNTY\_NAME
8. UNIT\_DESC
9. VALUE
10. YEAR

As a **bonus**, your app can also create a second table called **stats** which should contain some statistics about the data stored in the **fact\_data** table.

How to work

You will be required to put your code under version control, use either Github or Bitbucket whichever you prefer.

Each individual will submit their project as an individual, but this doesn’t mean you can’t ask for assistance from us or the rest of the hackers.

Deadline is Sunday evening, but try and maximize your time here today to learn stuff you didn’t already know and have fun.

Once done, you should submit a link to your repo via the Google form we’ll provide.

How to win

The winner will be evaluated based on:

1. Correctness - The app should fetch all the correct data and populate the database
2. How fast their app retrieved the data from the source and populated the fact\_data table in their database
3. The quality of their code
4. How much fun they had :-)