

fishpercolator / wardmill

Watch 1 Star 0 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Pulse Graphs



Branch: master wardmill / README.md

Find file Copy path

 pedantic-git Add a basic README 492c4b3 on 14 Apr

1 contributor

80 lines (57 sloc) 3.2 KB

Raw Blame History   

Ward Mill

Ward Mill is a web app that partitions datasets on Leeds Data Mill into their wards.

Demo: <http://wardmill.herokuapp.com/>

Problem

There's a lot of data on Leeds Data Mill but currently it's only partitioned by dataset. If you want to answer a question like "get me all the data relating to mortality in Adel & Wharfedale" you'd need to first find all the mortality data and then split them up so you only had the Adel & Wharfedale data.

Proof of concept

The proof of concept, developed in 3 hours during an Innovation Lab, is available [here](#) and represents 4 real datasets.

- If you visit the [wards](#) page, you can see a list of all wards. Clicking on one of these wards will show you the four datasets, but only those rows that related to the ward in question.
- If you visit the [datasets](#) page you can see each dataset broken down by ward. For large datasets, different colours are used for different numbers of rows (an example of a free visualization that can apply to all datasets).

The data is partitioned according to a heuristic that can be found in the [DataRow](#) model, which boils down to:

- If a column name starts with 'ward', it's assumed to be the name of a ward.
- If a column name looks like 'post code'/'post_code'/'postcode' it's assumed to be a postcode. Ward names are determined from postcode using data from Ordnance Survey.

Running it yourself

The app is functional in its current state, and can be used for any dataset on Leeds Data Mill, as long as the datasets in question meet the following criteria:

1. The dataset must be in CSV format and the data in question must be the first resource listed in the dataset.
2. The dataset must have column names that match the heuristic above.

The easiest way to run this is on [Heroku](#). Note that the database has more than 10,000 rows so you'll need to add credit card details to pay for a Hobby Basic database (\$9 per month).

- `git checkout https://github.com/fishpercolator/wardmill.git`
- `cd wardmill`
- `heroku apps:create <your app name>`
- `git push heroku master`

- `heroku addons:create heroku-postgresql:hobby-basic`
- `heroku pg:promote <new db name from previous command>`
- `heroku run rake db:migrate`
- `heroku run rake wards:create`

Now you can import data from Leeds Data Mill by going to the Rails console (`heroku run rails console`) and typing, for example:

```
Dataset.load_from_ldm('blue-plaques-of-leeds')
```

Future

Obviously this is 3-hour proof of concept and lots more could be done with it. Taking it forward to a full prototype it could do the following:

1. Support CSVs that are not on Leeds Data Mill (e.g. Leeds Observatory)
2. Allow users to import datasets into it
3. Present the data in more useful ways (e.g. as per-ward CSV files that can be downloaded)
4. Have more visualizations like the colour example
5. Support geographic data for identifying ward (this is a hard problem)
6. Display the wards as a geographic map instead of in alphabetical order
7. Support other kinds of partitions as well as ward (e.g. demographics, date)

