Francine Leech

Ideas for CIS550 Project

* Some sort of food related website or app
* Yelp dataset – millions of reviews

<https://www.yelp.com/dataset_challenge/dataset>

<https://www.yelp.com/dataset_challenge>

<https://www.yelp.com/html/pdf/Dataset_Challenge_Academic_Dataset_Agreement.pdf>

<https://www.yelp.com/dataset_challenge/dataset>

* Food/world cup

<https://github.com/fivethirtyeight/data/tree/master/food-world-cup>

* Food nutrition dataset

<https://www.google.com/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=food+related+datasets>

<https://public.tableau.com/en-us/s/search/all/food>

<https://public.tableau.com/en-us/s/gallery/global-and-regional-food-consumption-trends>

1. **Yelp Dataset:**

By adding a diverse set of cities, we want participants to compare and contrast what makes a particular city different.

**Seasonal Trends:**What about seasonal effects: Are HVAC contractors being reviewed mainly during winter, and manicure salons over the summer? Are there more reviews for sports bars on major game days and if so, could you predict that?

**Cultural Trends:**By adding a diverse set of cities, we want participants to compare and contrast what makes a particular city different. For example, are people in international cities less concerned about driving to a business, indicated by their lack of mention about parking? What cuisines do Yelpers rave about in these different countries? Do Americans tend to eat out late compared to those in Germany or the U.K.? In which countries are Yelpers sticklers for service quality? In international cities such as Montreal, are French speakers reviewing places differently than English speakers?

1. **Zika Vrius Epidemic:**

<https://www.kaggle.com/cdc/zika-virus-epidemic>

<https://github.com/cdcepi/zika> - updated Zika

What are the trends we are seeing in Zika?

* Over time
* Location (rural, urban)
* How are number of cases similar/diff across countries?
* Where will countries have more cases this summer?

Temperature data

* Can we predict cases with temperature data?
* Hot all year round
* Climate change – where could zika travel
* Where is it predicted to get warm enough to mosquito

How does the Zika epidemic compare to other mosquito borne viruses?

* <http://www.mosquito.org/mosquito-borne-diseases>
* Number of cases
* Duration of epidemic
* <https://github.com/cl65610/west_nile>

Some mosquito data

<https://data.world/zika-virus-data/mosquito-data>

<https://data.world/zika-virus-data/zika-travel-notices> travel notices - somehow collect data from CDC

Who is in the group

1. Develop an initial idea, and determine the technologies you wish to standardize on as a group. Amazon Web Services should be used for hosting your database and deploying your application.

Technologies:

MySQL

Node.js

JavaScript – make actual website

Python to put data in usable format

Amazon Web Services should be used for hosting your database and deploying your application

1. Provide 6-10 questions (in English) that someone might want to ask about the domain of your intended application. (You will be permitted to revise these questions later if needed.)

What is the current state of Zika?

What is the estimate spread of Zika?

I want to go to X country, is Zika there and what is the travel alert?

How many predicted cases will be in the future in Country X or Region X?

What is the predicted spread of Zika in the future?

How do Zika cases compare in rural and urban areas?

How do Zika cases compare in coastal vs non coastal cases? – Heat maps of different

How do cases compare between countries?

1. Setup Subversion/Git to share source code and starter data files. See http://www.seas.upenn.edu/cets/answers/subversion.html for details, and be sure that whoever sets it up grants access to everyone in the group. You should also add your assigned TA and Professor Naik to it so we can see what you are doing.
2. One group member should then upload a PDF document via Canvas, stating who is in the group, what your initial idea is, and what technologies you plan to use. The document should also include a timeline for the different milestones of your project, and a preliminary division of responsibilities.

Timeline:

2/21/17

12:30

* Meet in 2 weeks after playing with the data

Meet in

**Milestone 2** – 3/16/17

* NoSQL

Questions:

Data dump for data

D3.js

* Widely used
* Maps/highlighting maps
* NY times – open source see how it’s implemented
* On world map – Olympics –
* Wikipedia – query data live
* How it spread
* Cases per month

NoSQL – store as documents – MONGODB

Meanstack to communicate bw 2 databases

Which is going to to be with NOSQL vs tabular

* Links to videos, pictures – don’t need a table

User interface vs how to table should be modeled

Connecting the relational with nonrelational

* Parse out label of data

Design

How is going to be meaningful

Overlaying on top of map

Select vs pop up

Demographics of disease in specific country

Speed of spread

Mosquito demographic

<https://d3-geomap.github.io/>