Chicago Health Score: The integration of Chicago Food Inspections and Yelp Reviews
Jose Luis Rodriguez Orjuela
COMP300

Progress Report

Chicago Health Score: The integration of Chicago Food Inspections and Yelp Reviews

Through the analysis of the City of Chicago Food Inspection and Yelp reviews data a business opportunity emerge by providing accurate information about the quality of the service as well as insights about the conditions on how food providers located in a particular zip code do business in the City of Chicago.

The first challenge encountered in this research was in how to sort and analyze the data to gain meaningful information To overcome this obstacle it was necessary to focus the research in just business that the city of Chicago consider Restaurants then in a thorough manner eliminate rows with empty cell and wrong data that could give us problems later. At this stage the data is clean and organize ready for some text mining to create the correlation with the Yelp reviews data.

Also a classification by zip code was performed to better identify where the restaurants with

highest risk were located, here is the map:
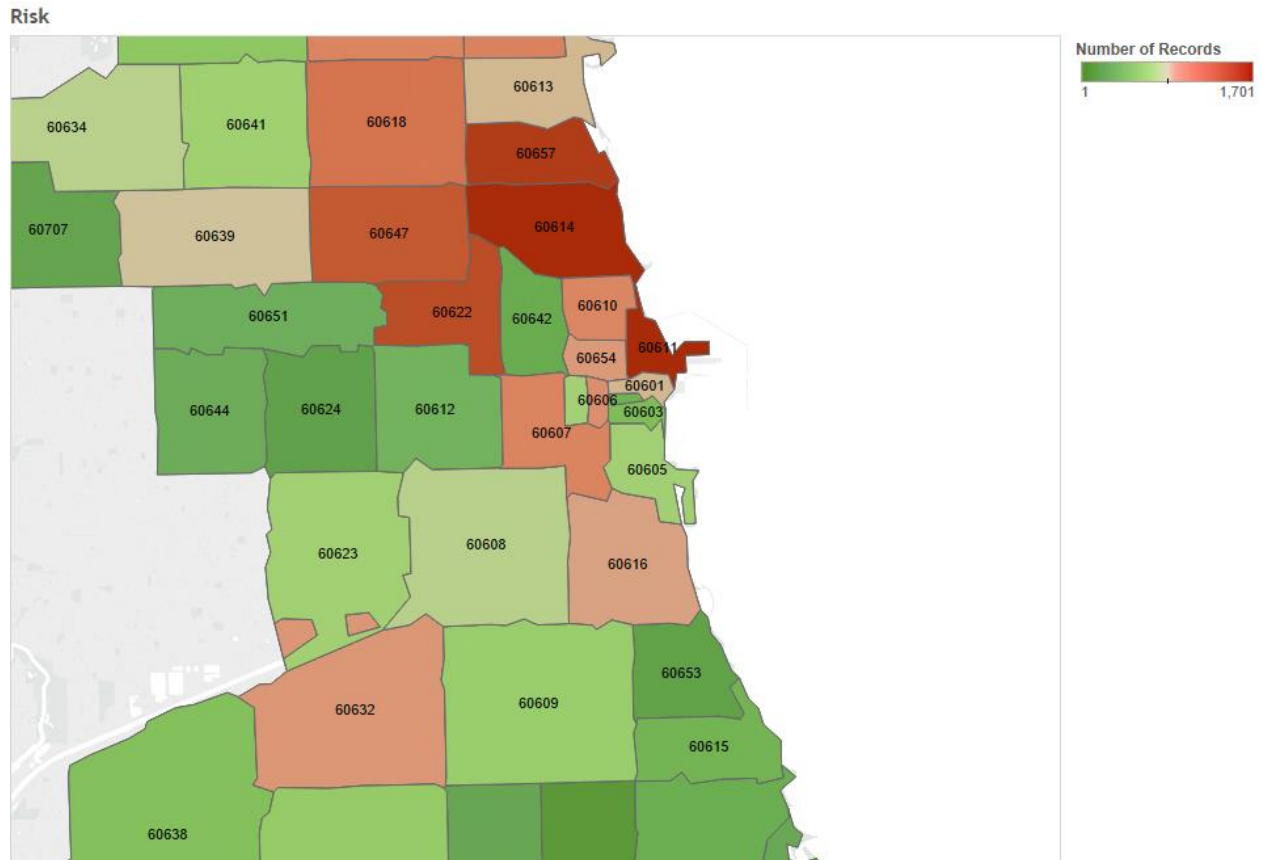
The next step for this research is to perform a text mining on a sample set of restaurants and get

the data for this restaurant on Yelp to create a correlation between the food inspection comments

and the yelp reviews.

Chicago Health Score: The integration of Chicago Food Inspections and Yelp Reviews
Jose Luis Rodriguez Orjuela
COMP300

Description of the Data

Inspection type

- **Canvass:** Random check of the business.
- **Consultation:** When a business is going to open for the first time.
- **Complain:** When a complain about the business is filed.
- **License:** When a business is obtaining a particular license.
- **Suspect food poisoning:** When a business is suspected of making a person ill.
- **Task-force inspection:** Bar or tavern inspection.
- **Re-inspection:** After one of the previous inspections is performed.

Results

- **Pass:** When no critical violations are found.
- **Pass with conditions:** When a critical violation is found but its corrected during the inspection.
- **Fail:** A critical violation that couldnt be fix during the inspection.
- **Others:** Business not located, out of business, no entry.

Risk (To affect public health)

- **Risk 1:** High
- **Risk 2:** Medium
- **Risk 3:** Low

Yelp Data

Yelp API returns a json-object per line. Every object contains a 'type' field, where we can determine if is a business, a user, or a review.

Business Objects

Basic information about local businesses. We can determine the 'business_id' of restaurants in the Chicago Food Inspection data to fetch more information about a particular business. The fields are as follows:

Chicago Health Score: The integration of Chicago Food Inspections and Yelp Reviews
Jose Luis Rodriguez Orjuela
COMP300

```
{
  'type': 'business',
  'business_id': (a unique identifier for this business),
  'name': (the full business name),
  'neighborhoods': (a list of neighborhood names, might be empty),
  'full_address': (localized address),
  'city': (city),
  'state': (state),
  'latitude': (latitude),
  'longitude': (longitude),
  'stars': (star rating, rounded to half-stars),
  'review_count': (review count),
  'photo_url': (photo url),
  'categories': [(localized category names)]
  'open': (is the business still open for business?),
  'schools': (nearby universities),
  'url': (yelp url)
}
```

Review Objects

Contain the review text, the star rating, and information on votes Yelp users have cast on the

review. The business_id is crucial in this part to associate this review with others of the same

business.

```
{
  'type': 'review',
  'business_id': (the identifier of the reviewed business),
  'user_id': (the identifier of the authoring user),
  'stars': (star rating, integer 1-5),
  'text': (review text),
  'date': (date, formatted like '2011-04-19'),
  'votes': {
    'useful': (count of useful votes),
    'funny': (count of funny votes),
    'cool': (count of cool votes)
  }
}
```

User Objects

Contain aggregate information about a single user across all of Yelp (including businesses and

reviews). For further research the User Object can determine the weight of the review, as users

that are more active in Yelp will be different to those that just create an account to make a

complain about a particular restaurant.

```
{
  'type': 'user',
  'user_id': (unique user identifier),
  'name': (first name, last initial, like 'Matt J.'),
  'review_count': (review count),
  'average_stars': (floating point average, like 4.31),
  'votes': {
    'useful': (count of useful votes across all reviews),
    'funny': (count of funny votes across all reviews),
    'cool': (count of cool votes across all reviews)
  }
}
```