Amal Nazeem

Bostonography Spring 2018

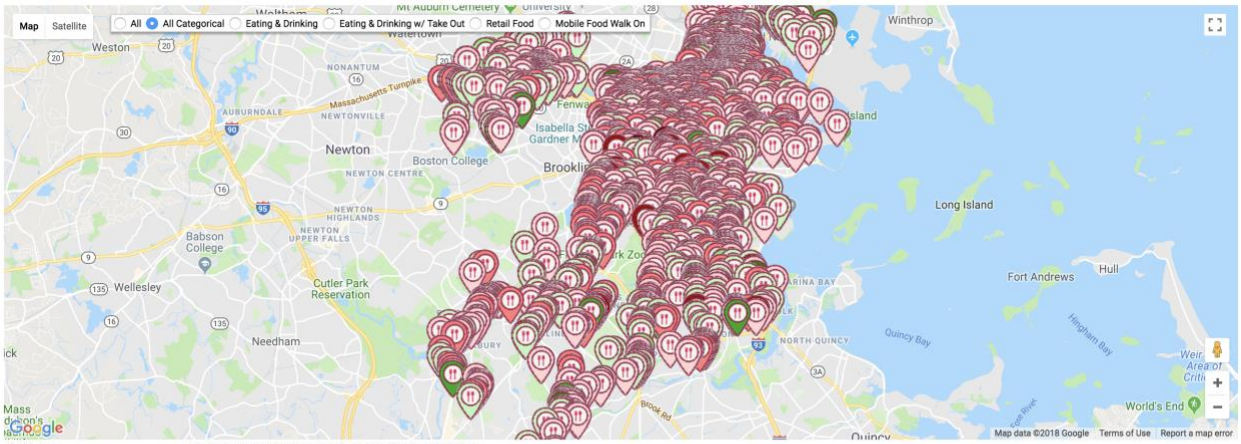
Final Project Write-Up

Food Viola: Food Violations in Boston

Introduction:

In order to access this tool visit http://foodviola.tomatotoaster.com/ and this write-up is available at http://foodviola.tomatotoaster.com/WriteUp.pdf. This tool was created for my Bostonography project where we had to analyze Boston data or create a tool which could give us some greater insight about the city we live in. I decided to create a tool that makes use of Boston's Food Establishment Inspections Database.

What is the Product:



Category	Amount	Average Score	Lowest Score	Highest Score
All	4191	14	-176	692
Eating & Drinking	1728	0	-176	692
Eating & Drinking w/ Take Out	1520	40	-135	692
Retail Food	943	-4	-63	265

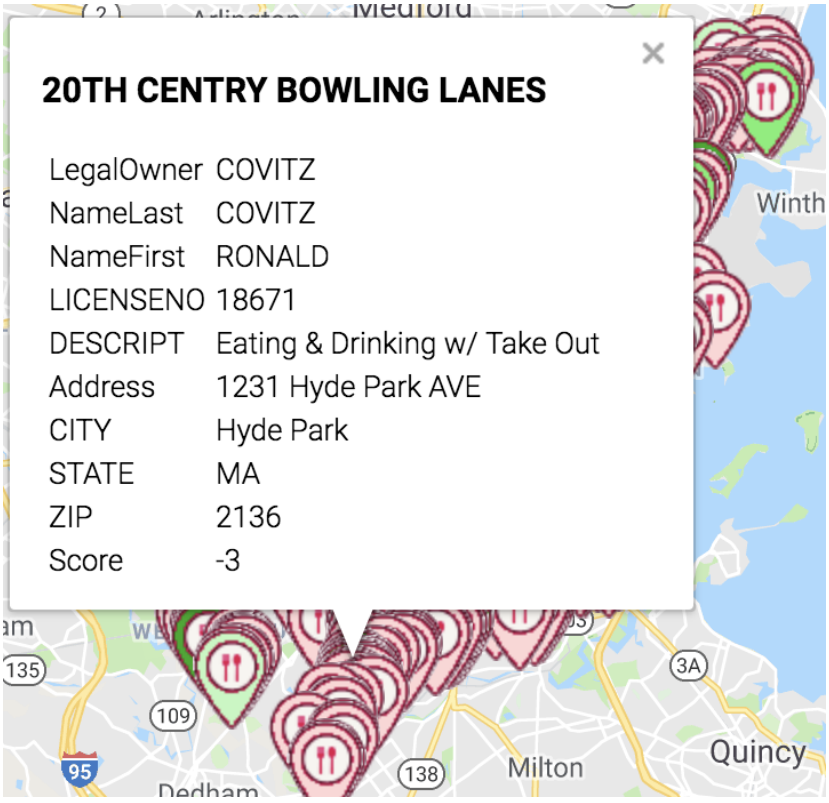
Link to Repo

What I've tried to create is a map based representation of the food inspection situation in Boston Massachusetts. The city of Boston leaves a public data set of all the food inspection violations that restaurants or retail stores in Boston have violated. Originally this was meant more to be more of a heat map of all the violations in Boston and which places have the worst

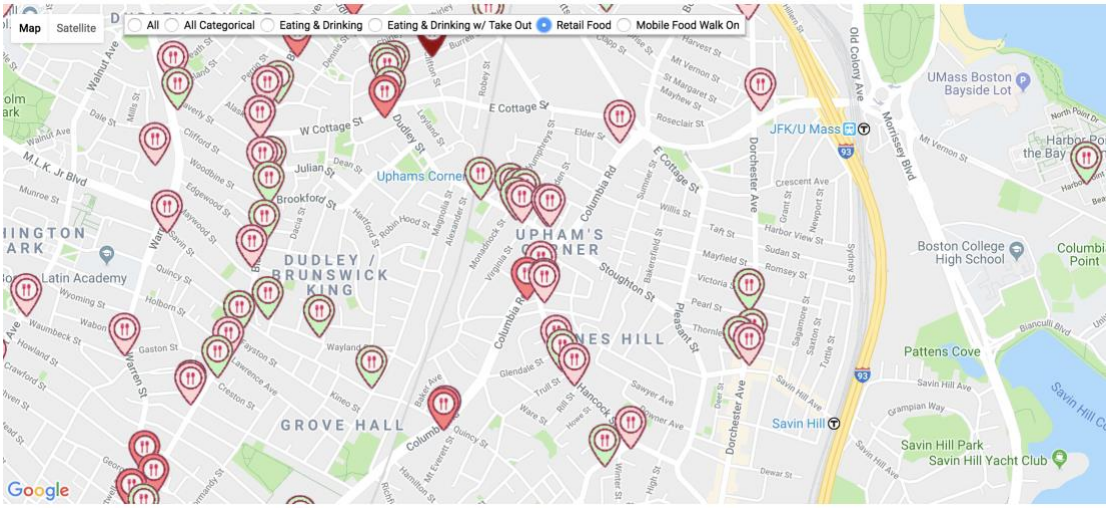
violations. However, that seemed very unfair to food establishments that are trying to proactively fix the violations they have been called out for. I decided that it would be wiser to reward those food establishments and score them against establishments that leave their violations unfixed. The scoring system is based on how many fixed violations a restaurant/retail store has minus the ones they have left open. This tool's ultimate goal is to showcase proactive restaurants and punish failing restaurants.

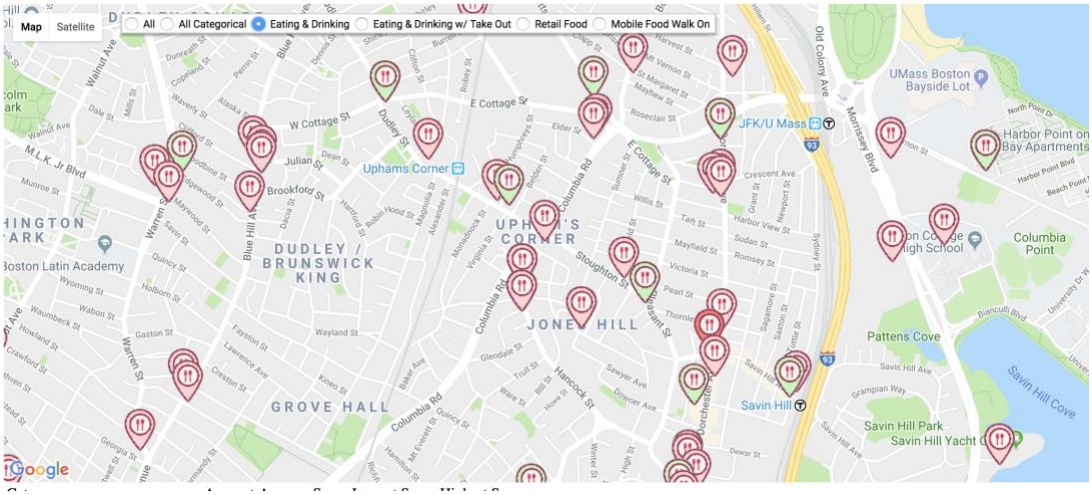
How to use:

Clicking on any of the points shows you details about the establishment. Most of the data comes directly from the Boston database but Score is computed live. (Scores can be negative)



The radio buttons help you filter by each category





The first two radio buttons show all the points. The difference is that the first one categorizes and colors the points as compared to all food establishments, whereas the second one more fairly assesses them based on their respective categories, but shows all of them.

<u>Datasets and Tools Leveraged</u>:

(**Boston Food Inspections**) https://data.boston.gov/dataset/food-establishment-inspections

This is the primary data set that I used. I originally had plans to incorporate Boston's Food Establishment license database as well, but I decided that it wouldn't help me much since the only additional piece of data it had in it was each food establishment's phone number on record. I was hoping to get information about restaurant size and cuisine type (e.g. Chinese, American, Fast Food), so I had also tried to exploring possibilities in cross-referencing the yelp database to get information about restaurants, but too many establishment's used names that wouldn't match up with an entry on yelp. In Limitations/Further Additions, I'll talk about what I can do in the future to get around this and find more ways to categorically compare these food establishments. What this data set did offer is a rough description of what kind of food establishments each entry in the data set were (e.g. Eating and Drinking w/Takeout, Takeout Only, Retail Store). As a result, I was able to compare stores that were in the same categories against each other.

(**Google Maps API**) https://developers.google.com/maps/documentation/javascript/

I had originally went back and forth on whether I would use Leaflet or Google Maps' api to make the map in this project. However, after exploring a lot of the documentation between mapbox (what leaflet uses to actually draw the map), I found that Google's services were more reliable, had better api key security, better documented, and aesthetically looked better. I also wanted to stick to one account for the api key because I was already using a Google Cloud Compute Engine that was tied to the FoodViola project. I ended up building the info window and some control options for the map thanks to the help of the Maps documentation.

(**Google Cloud Compute Engine**) https://cloud.google.com/compute/

I used this as a VM service to host the webapp because of pricing (300 dollars for free with every account and the server cost is roughly 4 dollars per month).

(**Mapbox csv2geojson**) http://mapbox.github.io/csv2geojson/

This is a mapbox service that I was using back when I tried to make this work with leaflet. This is what helps me turn the csv I compiled using R into a GeoJSON format that works with popular mapping services like Google Maps and Leaflet.

(**Templatic Map Icons**) https://templatic.com/newsblog/100-free-templatic-map-icons/

I needed to make icons that were color coded in order to represent levels between establishments (i.e. darker green means they have a higher score relative to their establishment).

How it works:

First I used R Code to squish all the violations into individual restaruants and counted their fixed violations and their unfixed violation.

```r
library(data.table)

fread("./data/mayorsfoodcourt.csv", stringsAsFactors=F) -> food_viola


library(magrittr)
library(plyr)


filtered <- dplyr::filter(food_viola, ViolStatus == 'Fail' & Location != '')
filteredp <- dplyr::filter(food_viola, ViolStatus == 'Pass' & Location != '')

df <- ddply(filtered,
            .(businessName, LegalOwner, NameLast, NameFirst, LICENSENO, DESCRIPT, Address, CITY, STATE, ZIP, Location),
            summarise,
            Severity = length(businessName))

pf <- ddply(filteredp,
            .(businessName),
            summarise,
            Positivity = length(businessName))

df <- merge(df,pf,by="businessName")
df <- df %>% tidyr::extract(Location, c("lat", "lng"), "\\(([^,]+), ([^)]+)\\)")

write.table(df, file = "./data/FoodViolationData.csv", sep=",",row.names=FALSE)

```

After that I use mapbox's csvtogeojson tool to convert that csv into a big GeoJSON file. In my maps code (found in map/index.html in my repo) I pull that GeoJSON data plug it into Google Maps api tool and do some front-end manipulation on the data so that I can color categorize them and display info windows.

Limitations/Future Additions:

All of the stores that were considered Takeout Only (Mobile Food Walk on) had no location data, so they were not included in the map. However, if one day the database has a "Mobile Food Walk On" store with a location database, it will be seen in the list and the table at the bottom will automatically incorporate that establishment.

I wasn't able to categorize establishments based on more unique characteristics because of the weird naming that some restaurants had, but I am trying to build a script that will go through the food establishment list, search the yelp api for it based on name and address, and then scrape information from the closest search term.

I've left the search bar out because it conflates my restaurant data with google map's data about places. Since it's difficult to make and not crucial for visualization purposes, I've decided to leave it out of the app.

Currently the "grading" of this app is best understood as the color differentiation between the food establishments. The icons that go r1 to r4 are for negative restaurants and g1 to g4 are for positive restaurants (i.e. those that fix more violations than they leave open). It would be more helpful if I showed a letter grade based on which icon category I gave each establishment. This way the consumer can just find a restaurant and see it's letter grade on the window (expect to see this change within the next few days).

Value:

The consumers of this app are Boston residents, tourists, food inspectors, and restaurants. My hope is that the residents and tourists can be more knowledgeable about the food inspection status of restaurants (which is currently abstracted away from consumers). With increased use, the food inspection committee might be motivated to fine tune their database and reporting of food inspections. That way it becomes more easily digestible (no pun intended) for their customers and my map/data analysis can be more nuanced. Another positive is that restaurants might become more competitive in fixing food inspection violations if the average consumer is able to better understand where violations are happening. This might seem unfair to restaurants at first, but I think the average consumer will win from this.