The Dirty Side of New York City Restaurants

By: Daniel McKay, Joseph Caltabiano

Final Project, CS480x

C-term, 2019

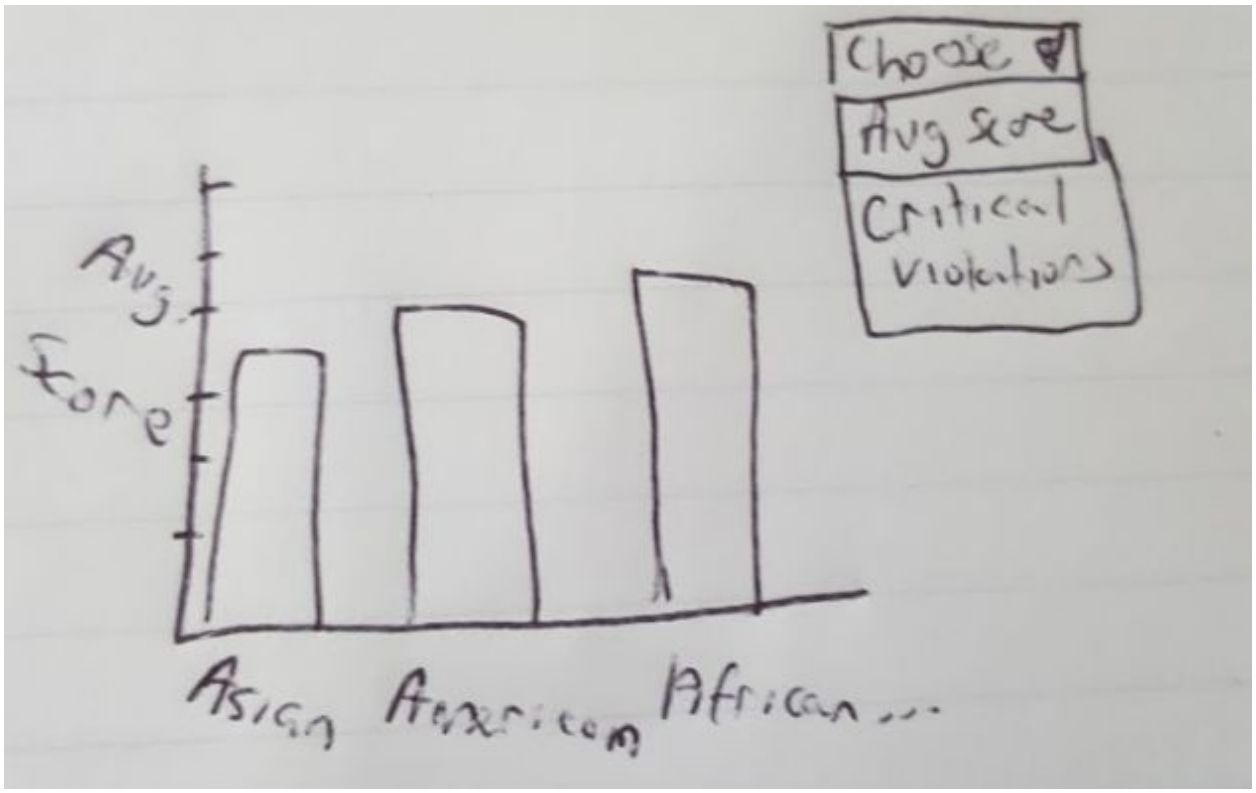
Overview and Motivation:
Our goal is to inform people about the cleanliness or lack of it in New York city restaurants. We want to first inform people on the type and number of restaurants per zip code, and then look into each type of restaurant's number of violations, and what specific violations they have committed, as well as looking at their health scores.
Related work:
There was a visualization done for this already, https://www.menglugao.com/blog/2017/12/7/new-york-city-restaurants-data-visualization
This visualization looked at cleanliness of restaurants, however we found it to not entirely be accurate, so we decided to do it our own way.

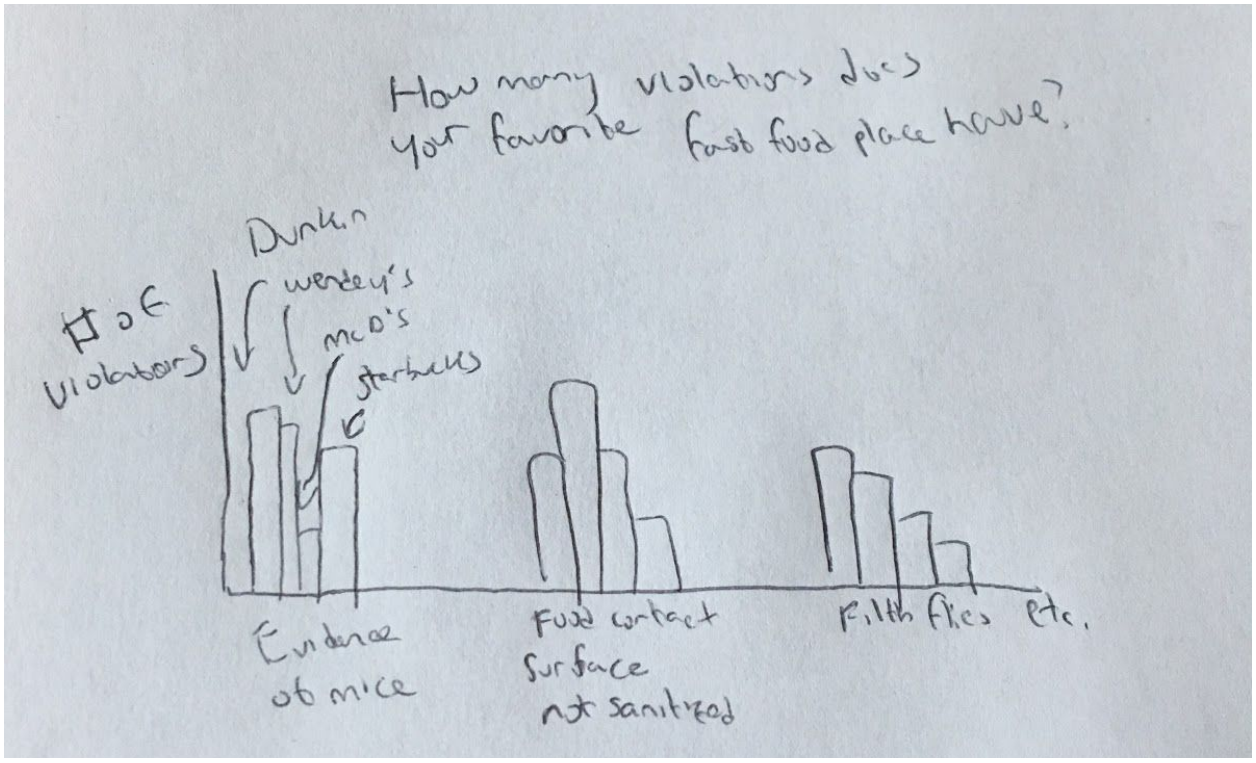
Questions:
We are trying to answer how clean New York City(NYC) restaurants really are. We looked at the visualization from the related works section, but we did not feel that it was an accurate representation. After doing some research, we found that if restaurants did not reach an 'A' grade on an initial inspection, it would be a no grade until the next inspection. This gives them time to prepare, but also messes with the system. That is why on the related visualization, we found that saying that almost every restaurant type had 70%-100% grade A ratings was dishonest. That is why we wanted to answer the question of how dirty NYC restaurants truly are, by looking at every inspection and not just looking at the grade.

Data:
We got our initial data from https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j which is NYC open data. This contains all inspections for every restaurant for the past 3 years in NYC. We had to do a lot of processing for this data. First, there were 92 categories of food initially. Many were repetitive, such as hot dogs, hot dogs/pretzels, so we condensed them down into 11 categories so the reader could easily look at all categories of food at once: indian, african, seafood, asian, latin, middle-eastern, european, comfort food, american, health food and bakery. Condensing this data, we used Ms Access to parse the data, and used this to get the average score for each category, the critical violations for each category, and for the amount of specific violation flags for 5 fast food restaurants: Dunkin, Subway, Domino's, McDonald's and Starbucks. We also used Ms Access to generate our map data, as the map data had the most common restaurant type per zip code, average score for each zip code, as well as the average score for each category for each zip code.

Exploratory Data Analysis:

We knew from the beginning that we wanted to use a choropleth map that the user could interact with somehow. We wanted to have supplementary charts of some sort that the user could click on and change the map. Here are some mock ups of these supplementary charts:





After making these charts, we were able to get some cool insights into our data. First, for the number of violations for each fast food place, the violations would obviously scale more with places that had many locations, and therefore would get more inspections. We found this to be misleading, and changed the data to be the % of inspections that a specific violation was found. This caused the data to even out more, and we were able to see who the real biggest offenders were for each violation.

Implementation:

For our final implementation, we decided to have the choropleth map take up the left hand side of the screen, and the two supplementary charts take up the right hand side of the screen, stacked on top of each other. We wanted everything to be in one screen so that the user doesn't have to scroll at while looking at the visualization. This way, the user can look on the right for something that looks interesting click on it, and get the corresponding visualization on the left hand side map.

Evaluation:

We learned some interesting things from the data. First, NYC restaurants are not as clean as we were lead to believe from the initial visualization that we looked at. Every single category of restaurant had at least 50% of their inspections result in a critical violation. That is worse than we imagined. There is also not a huge difference by category, the worst offender was indian food at 58% of their inspections being critical,

and the best being bakeries, with ~51% of their inspections being critical.

Looking at fast food restaurants, it is gross to see how bad some can be. McDonalds in NYC had 11% of their inspections result in them finding evidence of filth flies, and for 10% of inspections Subway and Dunkin they found that food contact surfaces were not properly washed. It was interesting to see how the 5 fast food places compared with each other, and gross to see just how many violations these places got.

We feel our visualization works well for portraying NYC restaurants' cleanliness. To improve, we would have added more interaction, such as adding another selection on the choropleth map. You can hover over each fast food chain's bar in the chart and it will tell you the boro that is the biggest offender, but I would have preferred to have the actual location with the most offenses, and then highlight the zip code on the map. Similarly, when hovering over the bar charts for the types of restaurants, it will tell you the boro with the best and worst average scores. I would have liked to add in the specific restaurant with the best and worst average score, to name and shame, and also name and praise specific places for their good/bad habits in cleanliness.