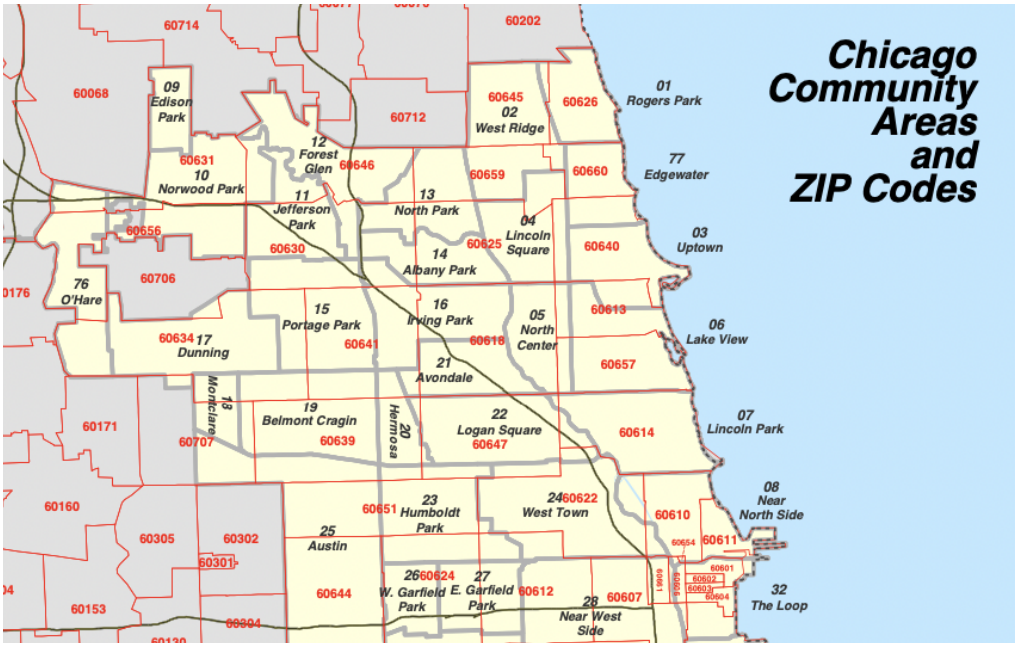
Jacob Liu
7/13/2022
Discovering Data Science

Failed Food

I've always loved food. I liked how interesting different foods from different countries and cultures were. So I decided to look into restaurant related data for my final project. After searching for how I could incorporate food into data science, I found a data set on restaurants and food inspections.[1] How many restaurants in Chicago fail their food inspections, and what general area are these restaurants in?

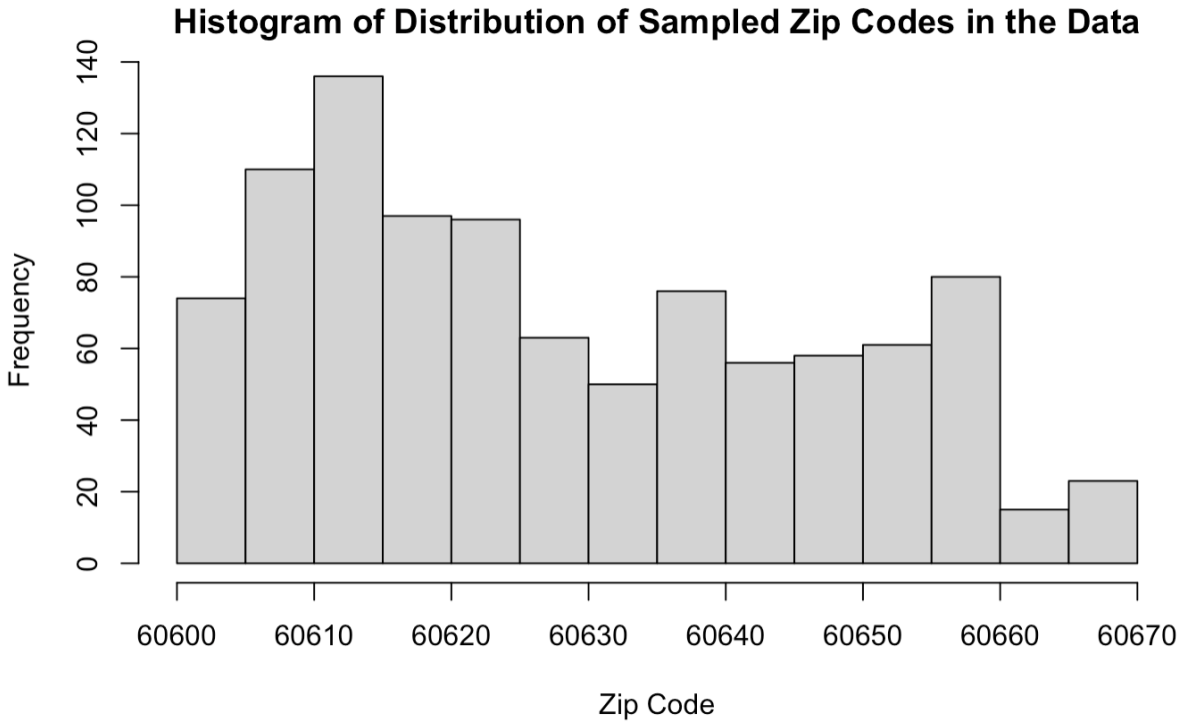
I found out that my data had a lot of garbage in it, including data from facilities that were not even restaurants, and data that was not from Chicago, even though the title of the data set said "Food Inspections in Chicago." However, after cleaning up the data, I decided to use the zip codes as a means of telling which neighborhoods these restaurants are in. Unfortunately, on this map of zip codes in Chicago, I could not see a



[1] https://www.kaggle.com/datasets/chicago/chicago-food-inspections

form of order or organization on how the zip codes are distributed. So when I was trying to describe which neighborhoods these restaurants were in, I would have two restaurants from neighboring zip codes, like 60644 - 60646, but the two would be on opposite sides of Chicago. These were the two big problems that I ran into while analyzing.

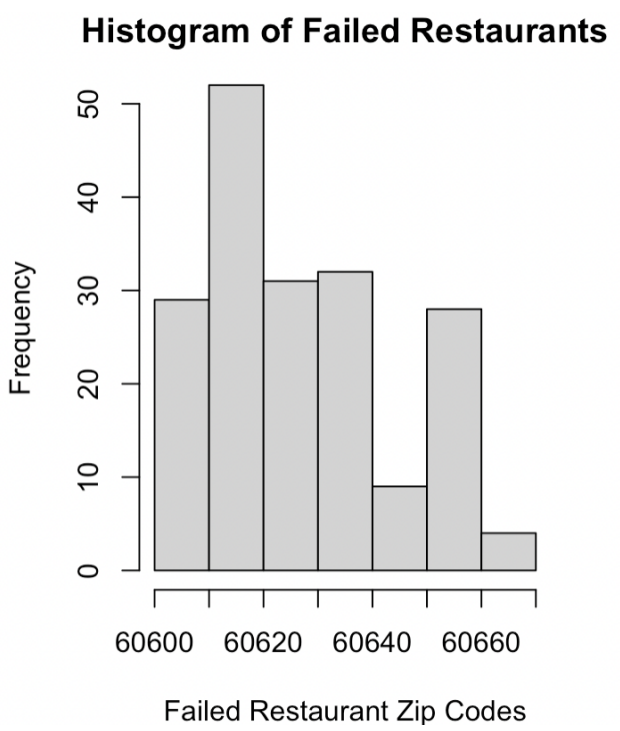
The results of my analysis entailed that a lot of failed restaurant inspections happened in the north side. That is a very general area since neighboring zip codes are not close to each other location-wise. So there was not really a general trend in a specific area of Chicago, however certain neighborhoods such as Lake View and Near West Side had many more restaurant data entries than other areas. These areas -



based on demographics by the government[2] - have a median annual income of around

[2] https://www.cmap.illinois.gov/documents/10180/126764/Near+West+Side.pdf
https://www.cmap.illinois.gov/documents/10180/126764/Lake+View.pdf

ninety thousand dollars. So the most sampled areas in this data set were not extremely well-off, but not terribly poor either. However, this led to the distribution of fails to be much higher in those areas because of the amount of restaurants surveyed being much higher. So I sampled a uniform number of data points from neighboring zip codes to get a proportion of failed restaurants in that area. I discovered that the zip codes that had the highest number of fails came from 60620 - 60630 and 60640 - 60650. Again, these zip codes do not cover a specific area in Chicago, but generally are on the north side.



The average number of fails in each zip code interval seemed to be around 25. The average percent failure for every region was 17.7% using my uniform number of sampled restaurants. However, when compared to using the entire data set, which had many more data points from Lake View and Near West Side, the percentage rose to 18.6%. Using the entire dataset, I found that the total number of failed restaurants in the dataset was 24,414. Using a sample from each zip code region, I was able to find that the general area where most restaurants failed was the north side, with a failure rate of approximately 25.3%.