City Exploration #3: Urban Hike

Lynsey Lawrence

Northeastern University

PPUA 5262: Big Data for Cities

Matthew MacNabb

April 13, 2024

City Exploration #3: Urban Hike

Background

This city walk was based on the "Restaurants" dataset explored in previous city exploration assignments, which examined different variables relating to inspections for those corresponding food service businesses. This dataset showcases data from all of the restaurants in Boston. Specifically, it examines each business's location. It is helpful for readers to understand the inspection data for individual locations, considering that each franchise or business is different, even if it is under the same brand, company, or corporation.

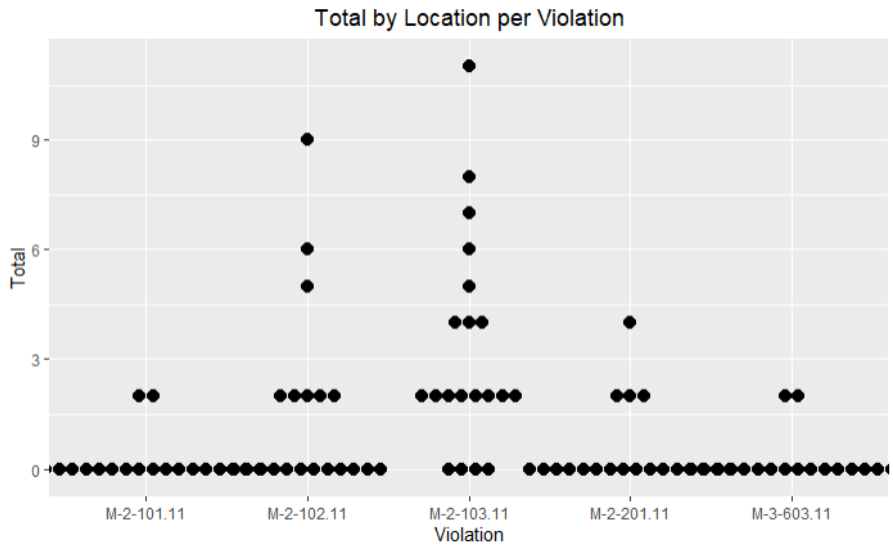
The dataset consists of 52 total variables used to examine 475,827 total inspections; this is not the total number of businesses. The total number of datapoints is not surprising, considering Boston is a large city, and the dataset includes all businesses selling food-related items. The focus of previous assignments was the statistics surrounding McDonald's locations across the city. Previous assignments found 27 individual McDonald's locations across the city, with 4,038 total inspections. One of the issues with the dataset was that each inspection was a single string for each location. Considering the large dataset and total number of inspections for McDonald's, it should be easier for readers to see the total number of inspections per location.

Violation Code

In previous assignments, it was discovered that the managerial violation with the highest composition was *M-2-103.11* (PIC Performing Duties). This violation has to do with employees' performance and how adequately they do their required responsibilities and tasks. Considering that performance is a large component to the business, it makes sense that inspectors would focus on that violation over others, like *M-2-102.11* (PIC Knowledge), however this was the second highest. The purpose of examining the specific codes in the previous assignment was to get a

better understanding of the descriptive statistics of a portion of the dataset. There were 85 types of Non-Critical violations, 23 Critical violations, and 35 Foodborne Critical violations. The analysis was simplified to only examine the five managerial violations: M-2-101.11, M-2-102.11, M-2-103.11, M-2-201.11/.14 .15, and M-3-603.11. The findings of the assignment can be seen in Figure 1 below.

Figure 1



Goal of Walk

Those findings served as the foundation for this City Walk assignment. Now that there was a basis for what to look for, it was important to try to understand why the composition was higher for certain locations over others. Limiting the visual scope to the M-2-103.11 violation in the figure, there was a datapoint that had a total value of 11. Each datapoint represents a single location. The total number of violations per location can be seen in Figure 2 below.

Figure 2
Total M-2-103.11 per Location

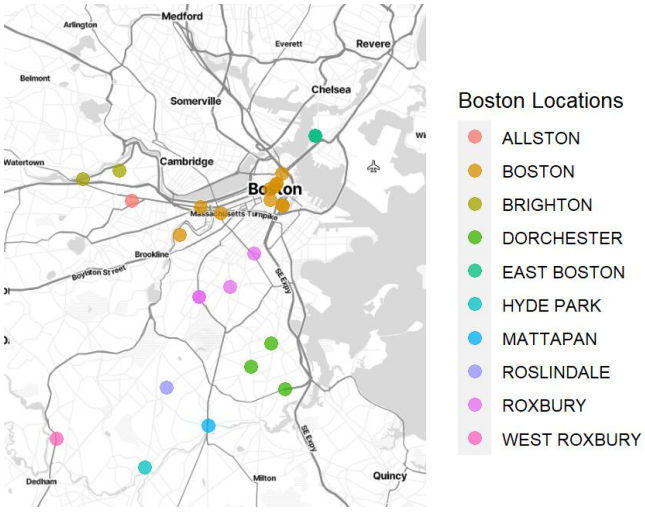
	Address	Total
1	3060 WASHINGTON ST	11
2	870 MASSACHUSETTS AVE	8
3	718 AMERICAN LEGION HWY	7
4	540 COMMONWEALTH AVE	6
5	640 ATLANTIC AV	5
6	3064 WASHINGTON ST	4
7	1650 VFW PKWY	4
8	1223 COMMONWEALTH AV	4
9	178 BORDER ST	2
10	1 UNION ST	2
11	327 WASHINGTON ST	2
12	702 WASHINGTON ST	2
13	340 LONGWOOD AV	2
14	607 WASHINGTON ST	2
15	511 GALLIVAN BLVD	2
16	1750 SOLDIERS FIELD RD	2
17	174 BORDER ST	0
18	315 WASHINGTON ST	0
19	417 WASHINGTON ST	0
20	146 TREMONT ST	0

As seen in Figure 2, 3060 Washington St had the highest total with 11 violations. These findings are important because that serves as a starting point for the city walk. It provides a location to observe or a reference point for the area to study.

The goal of the city walk was to recognize and examine any important factors affecting the location's likelihood of having this performance-based violation. Considering the issues that arose in previous assignments with analyzing factors that explain the data, it was important to understand that the factors examined were merely for observations, not an explanation. It was hard to draw strong conclusions or explanations on this subject. Previously, it was discussed that it is difficult to explain why employees performed the way that they did, especially with demographic information like race, gender, socioeconomic status, etcetera. As mentioned before, there were 4,038 inspections across 27 locations. Each neighborhood in Boston is different, whether it is based on socioeconomics, politics, religion, or racial/ethnic background. For this assignment, the observations only considered what was in the surrounding area of each location.

In this case, the focus was on the area around 3060 Washington St, which is in the neighborhood of Roxbury, as seen in Figure 3.

Figure 3
Locations by City



As seen in Figures 3 above, the distribution amongst the cities was not even. Boston has the most locations with 11, followed by Roxbury with four. In contrast to the prior City Exploration, this assignment focused more on poverty and median rent and how they could possibly influence the results found for 3060 Washington St.

Observations

3060 Washington St

An interesting fact about the findings above was that this location was observed in the first city exploration. Here were the findings of the original observation:

"The first observation occurred at the leading location, 3060 Washington St in Roxbury. Interestingly, there was a McDonald's location in the area, but it was not at the listed address. The McDonald's location next to the bank [at 3060 Washington St] was listed under 3064 Washington St. This means that the dataset has a major error with the addresses. The

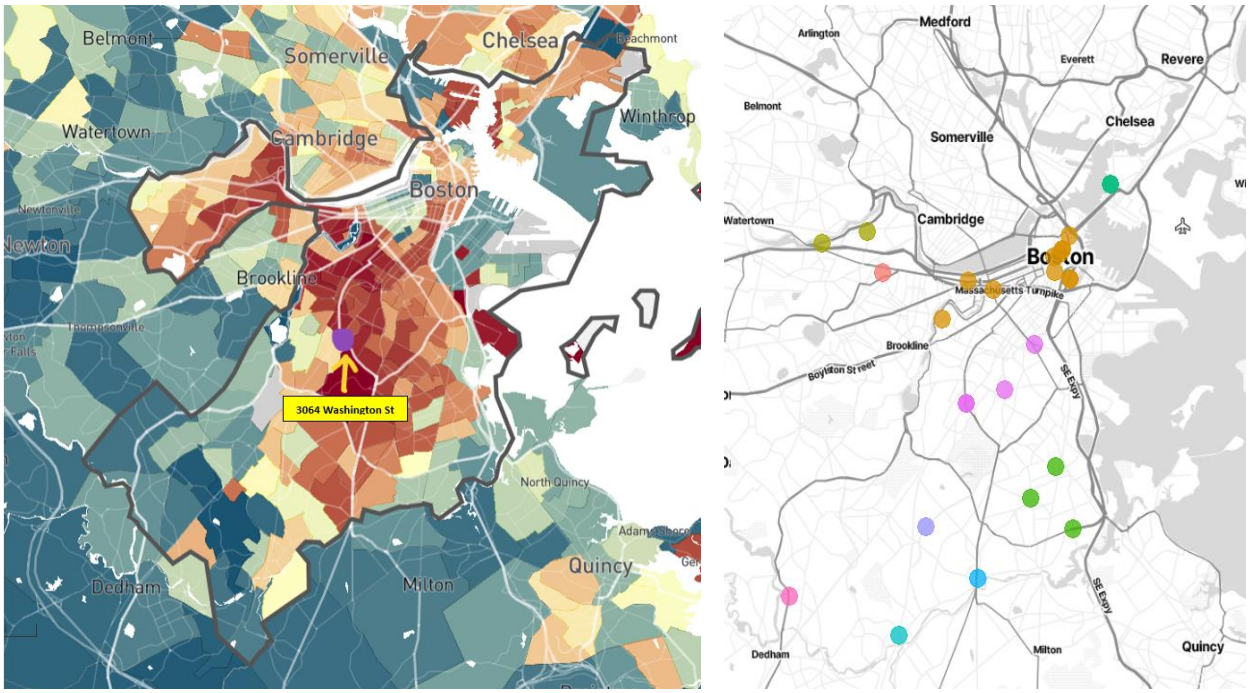
McDonald's location with the most inspections was listed under an incorrect address. Given the addresses, it can be assumed that the address would be changed to 3064, and the total from that listed address would be combined with the total for the 3060 address. […] This address is the actual location that the data should have represented throughout the dataset. Upon initial inspection, the sign outside the location says "Express," which is uncommon with McDonald's locations […] Inside the restaurant, it was clear what an "express" location meant. The location did not include a dining area like other locations. This was meant for customers to purchase their food to-go; an accompanying drive-thru was in the back of the building. Notably, it shared the same building as the Santander Bank, an H&R Block, and a Metro by T-Mobile."

Figure 4
3064 Washington St, Roxbury

Poverty Rate

Using data from the United States Census Bureau, the Opportunity Atlas provides data about locations across the United States on various topics. In this case, it would be interesting to see if poverty was a considerable factor. See Figure 5 below to see the composition of the poverty rate in Boston from 2012 to 2016 (United States Census Bureau, 2016).

Figure 5
Poverty Rate in Boston (2012-2016)



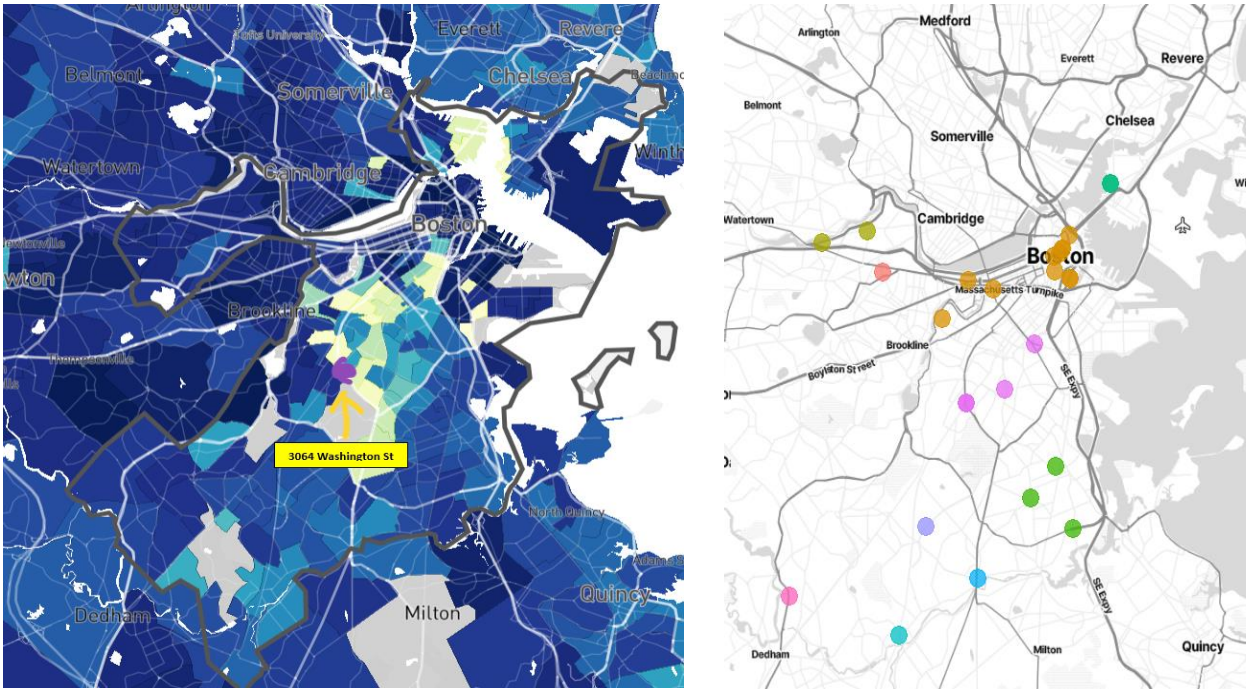
In the map above, the color *RED* corresponds to a high poverty rate. The color *YELLOW* corresponds to a medium poverty rate, and the color *GREEN* corresponds to a low poverty rate. The evolution/devolution is on a gradient scale. 3064 Washington St is in a high poverty area; however, it appears like a majority of Boston neighborhoods are in a high poverty area, especially the McDonald's locations. Considering this composition, it is highly unlikely that poverty played(s) a significant role in explaining why certain locations had more of the M-2-301.11 violation over others.

Median Rent

In addition to considering poverty rate, median rent was considered. Despite a majority of Boston falling in the high poverty status, it would be interesting to see how rent, or the total amount, had an effect. Rent in Boston, historically, had been known to be higher in certain locations than others. In this case, it would be interesting to see if that observation was true or if

it could really make a difference. See Figure 6 below to see the composition of the median rent in Boston from 2012 to 2016 (United States Census Bureau, 2016).

Figure 6
Median Rent in Boston (2012-2016)



In the map above, the color *BLUE* corresponds to a high median rent. The color *TEAL/GREEN* corresponds to a medium median rent, and the color *YELLOW* corresponds to a low median rent. The evolution/devolution is on a gradient scale. 3064 Washington St is in-between a high and medium rent area. The area itself was considered to be a low median rent area. The same was true for the other Roxbury locations. When comparing the map from the previous assignment to the figure, a majority of the total locations are in high median rent areas. This was interesting considering that the previous observation found that a majority of the neighborhoods were considered to have high poverty. Despite there being high poverty throughout the neighborhoods, the rent is medium to high in most neighborhoods.

Considering the data in Figure 1 and Figure 2, compared to 3060 Washington St, there were four locations with a total of zero: 174 Border St, 315 Washington St, 417 Washington St, and 146 Tremont St. Like 3060 Washington St, 174 Border St and 146 Tremont St fell within the same type of area composition for median rent. 315 and 417 Washington St were in high median rent areas. In this case, it would be difficult to draw any conclusions regarding the high to low rent areas. For example, since 3060 Washington St was in a low rent area, theoretically, one could conclude that this could increase the likelihood of receiving a performance violation. This could be due to stereotypes or assumptions regarding those who live in impoverished areas. However, this conclusion cannot be proven considering the locations with zero performance violations were split between high and low areas. This further proved that it was difficult to make accurate or supported observations or conclusions about what could cause the findings above.

Conclusion

The importance of considering these observations for this dataset was to examine external factors and the possible effects that they may have on the data. A large issue with this dataset was that it was difficult to find correlations or variables that could explain one another. To get a better understanding of the data, it was crucial to research and observe demographic information that was not present in the original dataset. In this case, it was poverty rates and median rent totals. Observing those two factors was interesting since there was a difference in their comparisons with the data. Considering that a majority of the neighborhoods were in high poverty areas, the rent data should not be too different, but when focusing on the individual locations, there were slight differences. Understanding the relationship between those two factors together is crucial because they are different but also similar at the same time.

In terms of the relevance of this work and dataset observations over the semester, this assignment did more to connect the data to actual city data compared to prior assignments. The work here, over the course of the semester, has slowly grown to be more connected to the community. Despite the lack of concrete conclusions, this work has made it possible to acknowledge or recognize key factors and variables that could explain variables that typically are not observed. Other datasets may involve demographic variables, like race, age, gender, but this dataset was limited to variables related to food violations and statuses. Working to connect those variables to city data was challenging, but it was possible. It was important to show that datasets like this could be connected to unconventional factors and that urban informatics can be analyzed at basic and advanced levels by anyone despite experience. The analysis from the first observation to this one should be evident of this perspective change.

References

United States Census Bureau. (2016). The Opportunity Atlas. *Opportunity Atlas*. https://www.opportunityatlas.org/