

City Council Project: Counselor Mejia x City Services

Team 5:

Rithvik Doshi,
Haowei Li,
Ankur Ankit,
Amy Seedhom (Team Lead)

Table of Contents

- Background
- Project Goal
- Data Analysis
 - Census Dataset
 - Rental Assistance Dataset
 - Business Licenses Dataset
 - Capital Budget Dataset
- Challenges and Roadblocks
- Limitations and Assumptions
- Extension Project
- Future work

Background

The city of Boston has allocated money and resources to offer relief to the businesses and people of Boston. A budget serves as a plan for the city on how resources will be raised and allocated in order to provide quality services to residents. It is a commitment made by the City Council about how money collected will be spent in the upcoming year. The budget provides accountability to residents on how the city of Boston spends money. It also provides the City Council with the opportunity to examine how revenues and expenditures relate to long-term funding priorities and established City Council goals.

Budgets are often separated into a number of smaller sub-budgets or funds. The first separation often discussed is between operating and capital budgets. A Capital Budget supports the long-term assets of the city. It encompasses maintenance and repair of existing infrastructure, construction of new infrastructure, plus construction or purchase of new facilities and/or equipment.

In short, the budget of the city and where it is spent has a direct impact on the neighborhoods, and the well being of the city as a whole.

Project Goal

Knowing how impactful the budgets can be to the city of Boston and development of different neighborhoods, our goal is to analyze these allocations and determine whether the resources have been equitably distributed, so the city can look at our analysis and design a better plan for city development and people's well being. We believe that this project may have the potential to impact many residents and businesses in the city of Boston, and may assist to decrease inequality between social classes and ethnicities living in Boston. Therefore, we will discover how city funding has permeated and how business licenses and other resources have been distributed across Boston.

Data Analysis

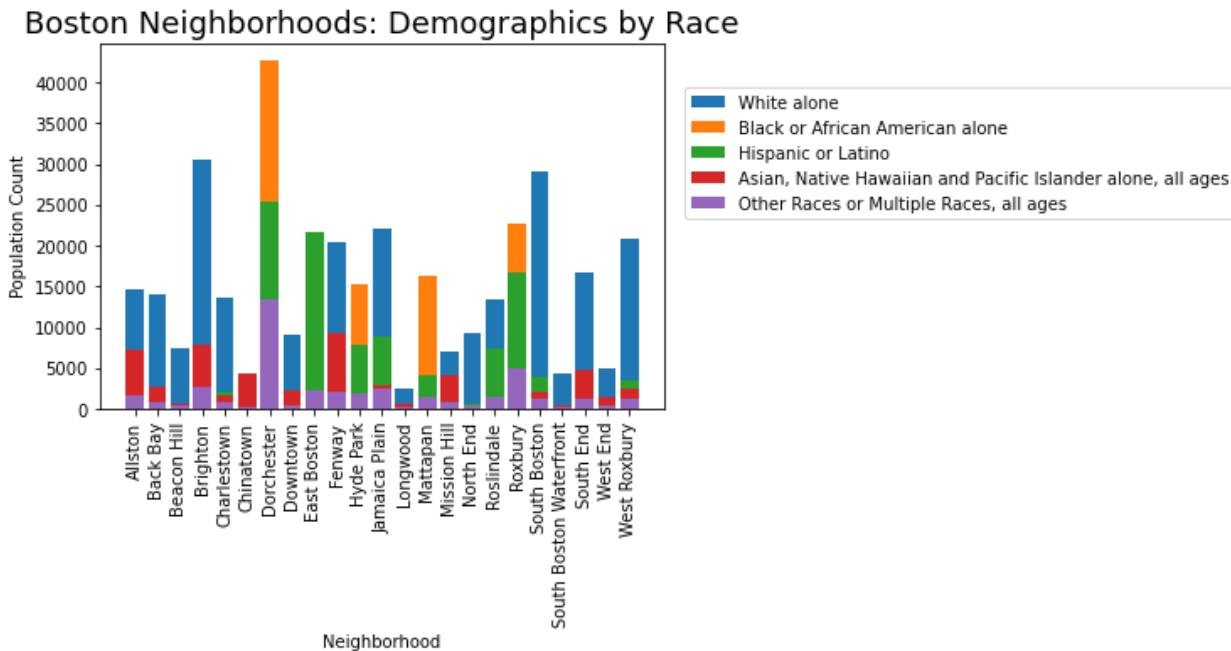
We worked with [census](#), [rental assistance](#), [business licenses](#), and [capital budget](#) datasets.

Census Dataset:

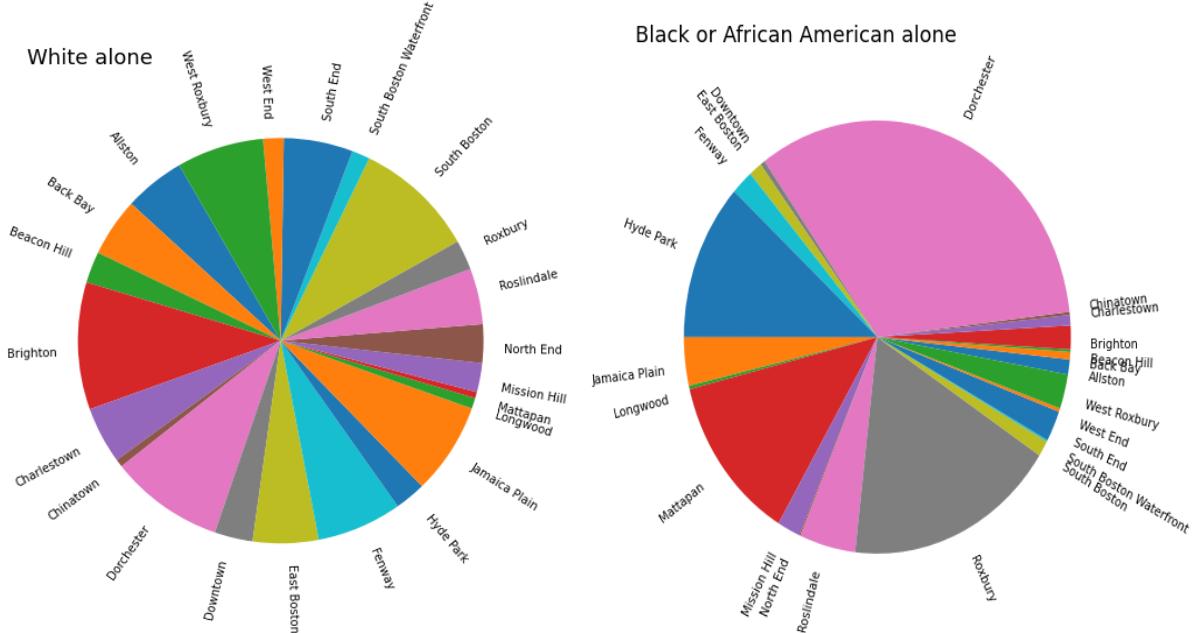
This included census data for the city of Boston, Boston neighborhoods, census tracts, block groups, and voting districts. In the 2020 Census, the U.S. Census Bureau divided Boston into 207 census tracts (~4,000 residents) made up of 581 smaller block groups. The Boston Planning and

Development Agency uses the 2020 block groups to approximate Boston neighborhoods.

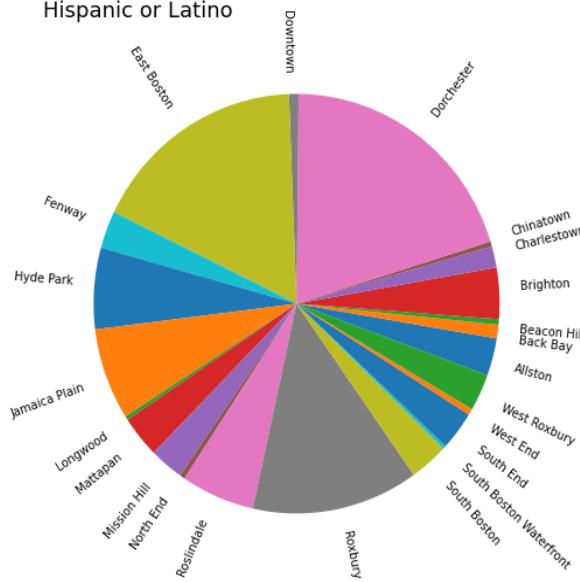
We first used these approximated 23 Boston neighborhoods to visualize the demographic compositions of each of them by race, as shown below.



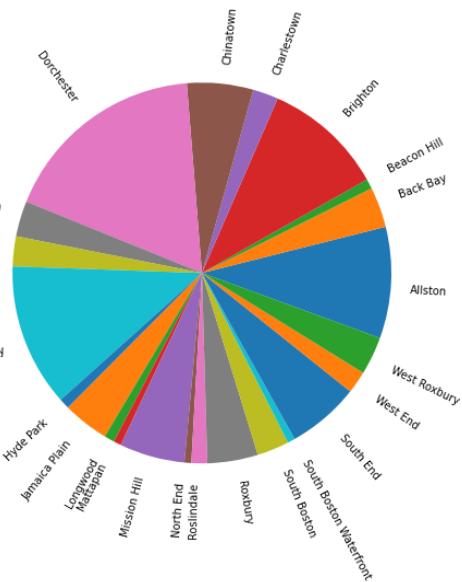
We then examined each race and determined in which neighborhoods they were most concentrated in throughout Boston, as shown below.



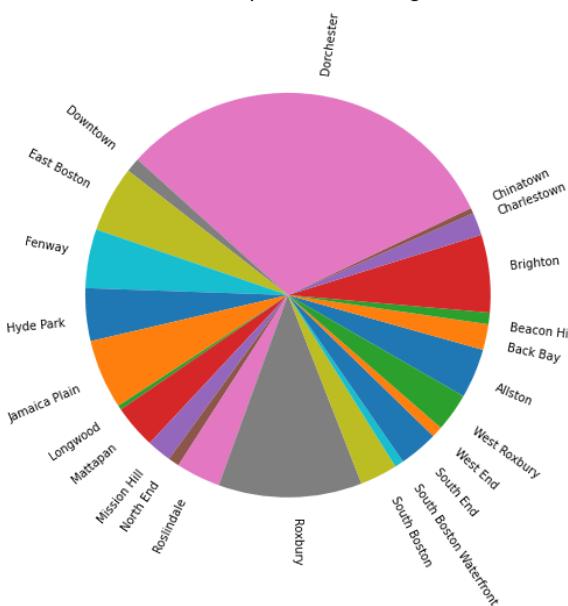
Hispanic or Latino



Asian, Native Hawaiian and Pacific Islander alone, all ages



Other Races or Multiple Races, all ages



With this information, we used this as the basis of our further analysis and conclusions to determine which are equity communities and whether or not resources were being well distributed to them.

Rental Assistance Dataset:

This data is maintained by the Mayor's Office for Housing/Office of Housing Stability. The data were collected from Oct 2020 to Sept 2021, with the interval of 6 months. The first dataset contains the information about RRF provided before the pandemic (Covid-19). The second dataset contains the information about RRF provided during the pandemic (Covid-19). Both the datasets contain information like Recipient Id (Street address + zip codes), Amount Requested, Amount Provided as RRF, RRF Award Type (Futures, Arrears, or both), Months of assistance provided.

The first plot provides the data about how much RRF was requested from different areas, the second plot provides the data about how much RRF was provided to different areas. The third plot provides the analysis between the two plots, which was done by subtracting the data points of RRF requested from RRF provided.

Analysis (Oct 2020 - Mar 2021)

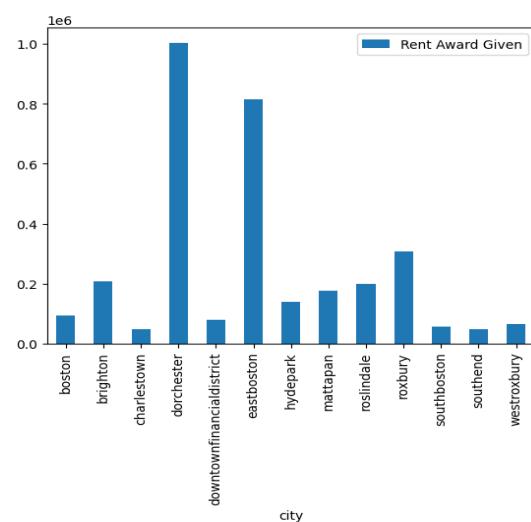
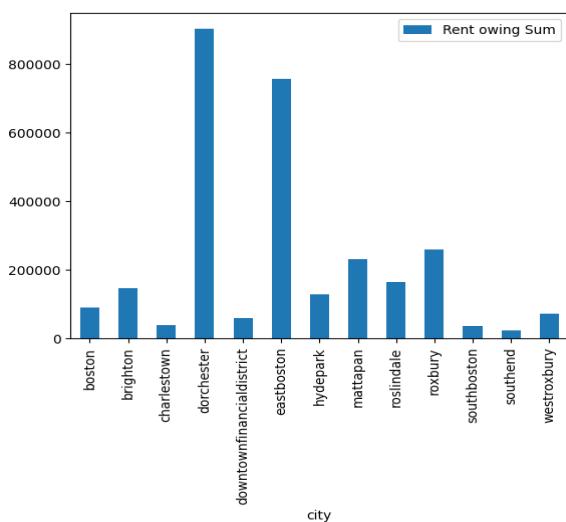
Total RRF Request : \$2.9 Million

Total RRF Given: \$3.25 Million

Most RRF is requested by the people of Dorchester (about \$900,000) and most RRF is given to Dorchester as well (\$1 Million), followed by East Boston.

Both the plots visually follow the same trend.

As per the third plot, all the areas except Mattapan and West Roxbury received excess funds compared to what was requested.



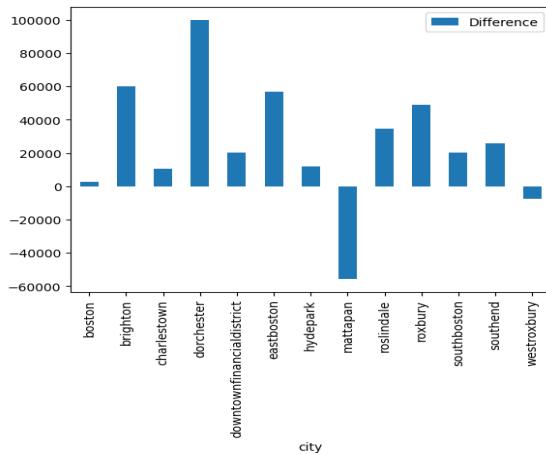


Fig RRF Plots for Oct 20 - Mar 21

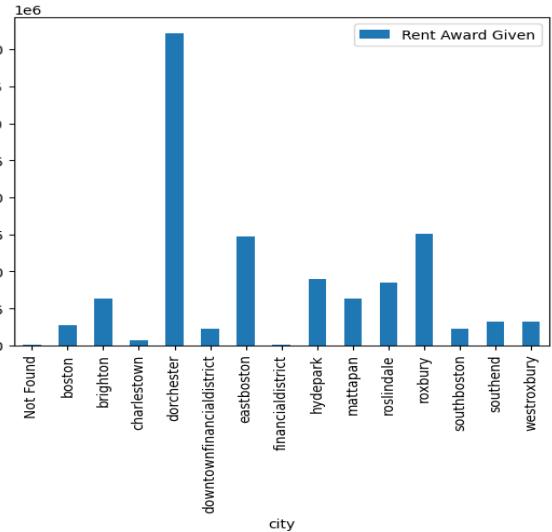
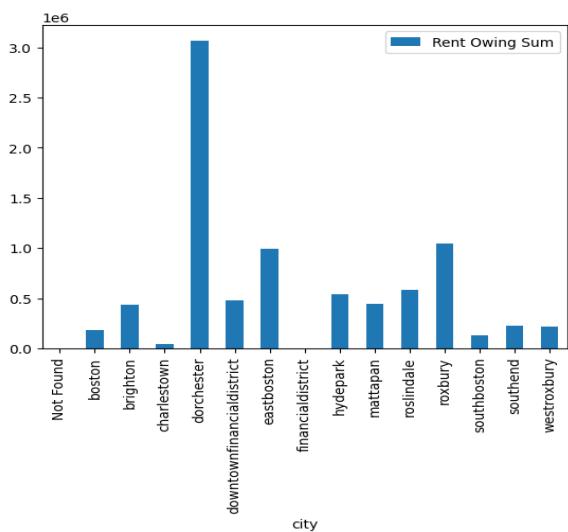
Analysis (Apr 2021 - Sept 2021)

Total RRF Requested: \$8.4 Million

Total RRF Given: \$11.68 Million

Most RRF is requested by the people of Dorchester (about \$3 Million) and most RRF is given to Dorchester as well (\$4.2 Million).

From the third plot, except for the Downtown/Financial District area, all other areas received excess funds compared to what was requested.



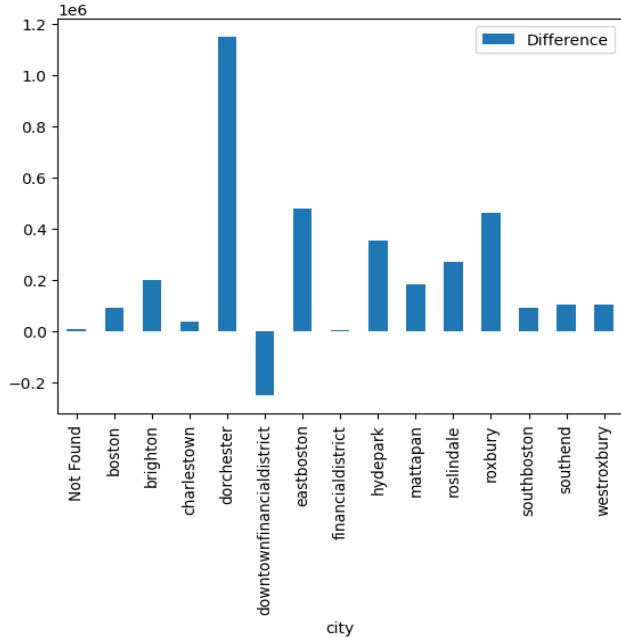
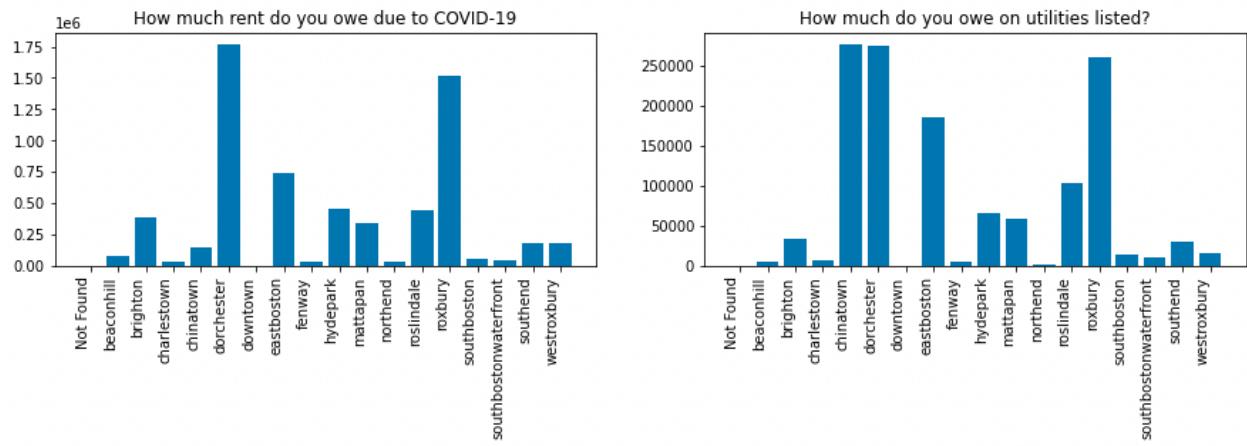


Fig. RRF Plots for Apr 21 - Sept 21

Finally, we decided to plot the responses to the two questions asked in the Apr 21-Sept 21 dataset, aggregated by neighborhood.

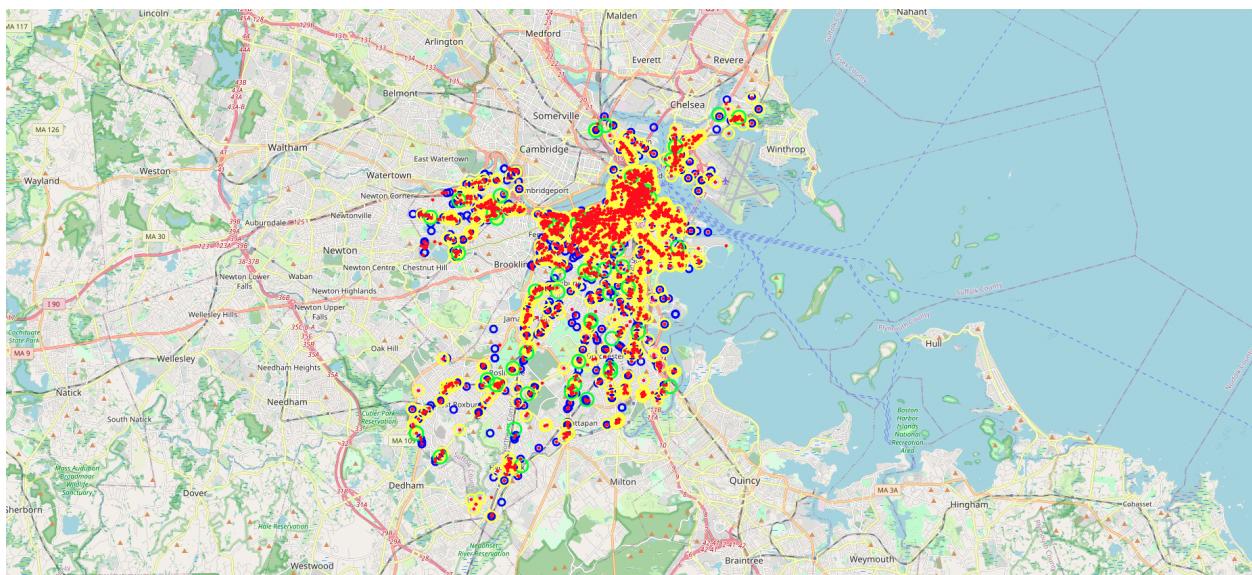


As you can see, neighborhoods like Dorchester and Roxbury seem to owe a lot more on rent and utilities due to the Covid-19 pandemic as opposed to other neighborhoods, suggesting that they may be greatly impacted by the effects of the pandemic. Other neighborhoods have nominally

high utility costs, such as Chinatown and East Boston, and this might be looked into to determine why they are so high compared to other locales.

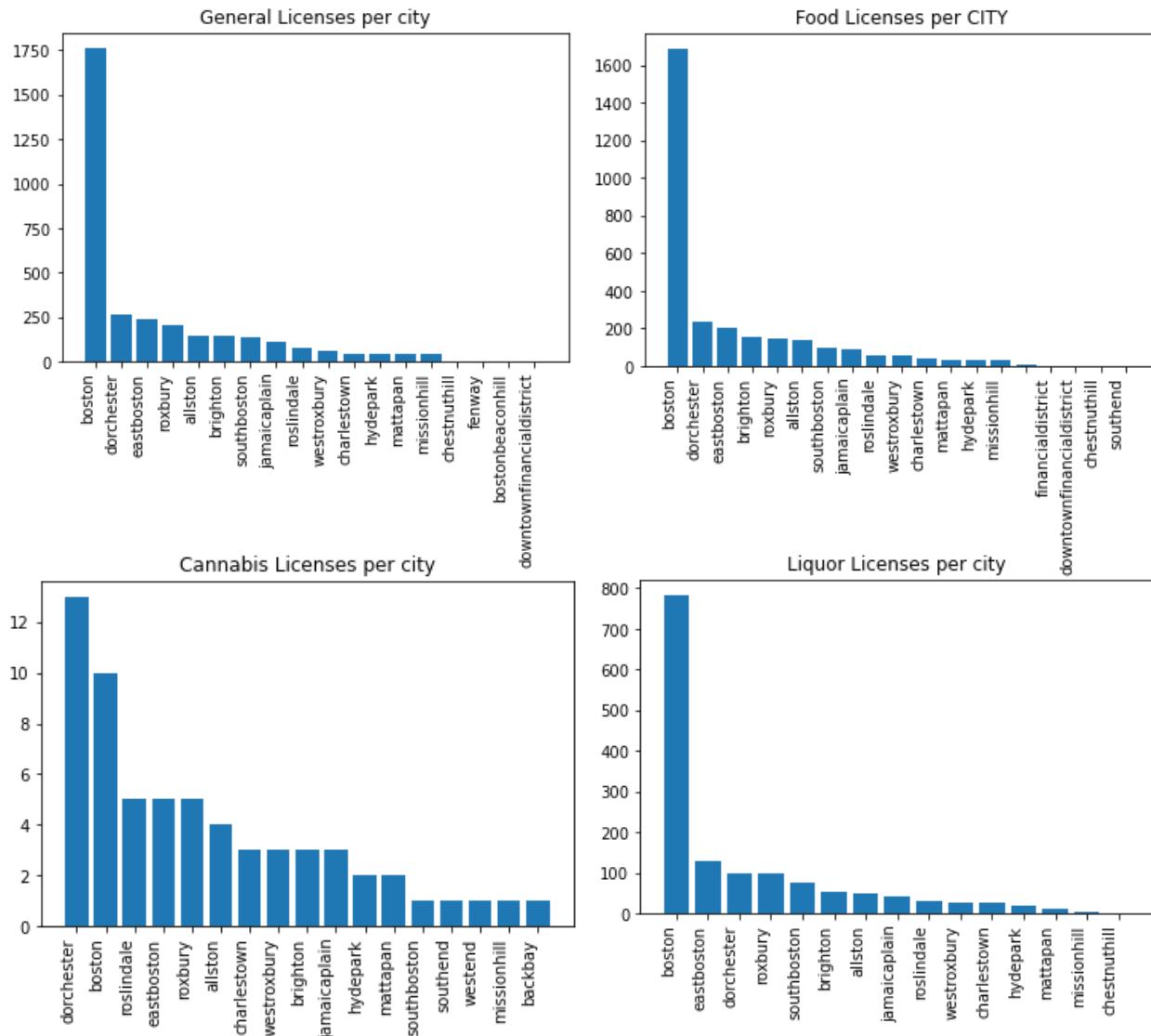
Business Licenses Dataset:

The Licensing Board issues a variety of license under MGL Chapter 138 and Chapter 140 including Billiards, Bowling, Clubs/Veterans' Groups, Common Victualler (Food Service), Alcohol Beverage, Innholder, Dormitories/Lodging houses, and Retail Package Stores. The licenses are divided into four main categories: food, alcohol, cannabis and general licenses. The following figure shows a map of the city of Boston with the location of these licenses.

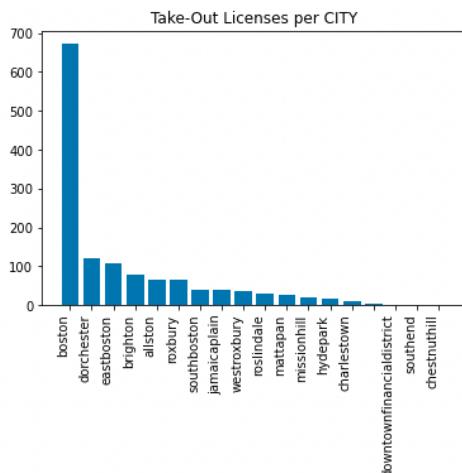


Red points indicate general licenses, blue indicates food, yellow indicates alcohol, and lime indicates cannabis. The general licenses overlap with many of the specific licenses. From the map, you can see that many of the licenses are concentrated around the city center of Boston and become more sparse the further away they are from the center.

As for the number of licenses per city, in each dataset Boston, Dorchester and East Boston have some of the highest amounts of licenses, while Hyde Park, Chestnut Hill and Mattapan have some of the lowest amounts of licenses. It's hard to say whether this gap is due to the difference in population, population density or difference in funding given the fact that in this portion of the analysis we didn't normalize for population.

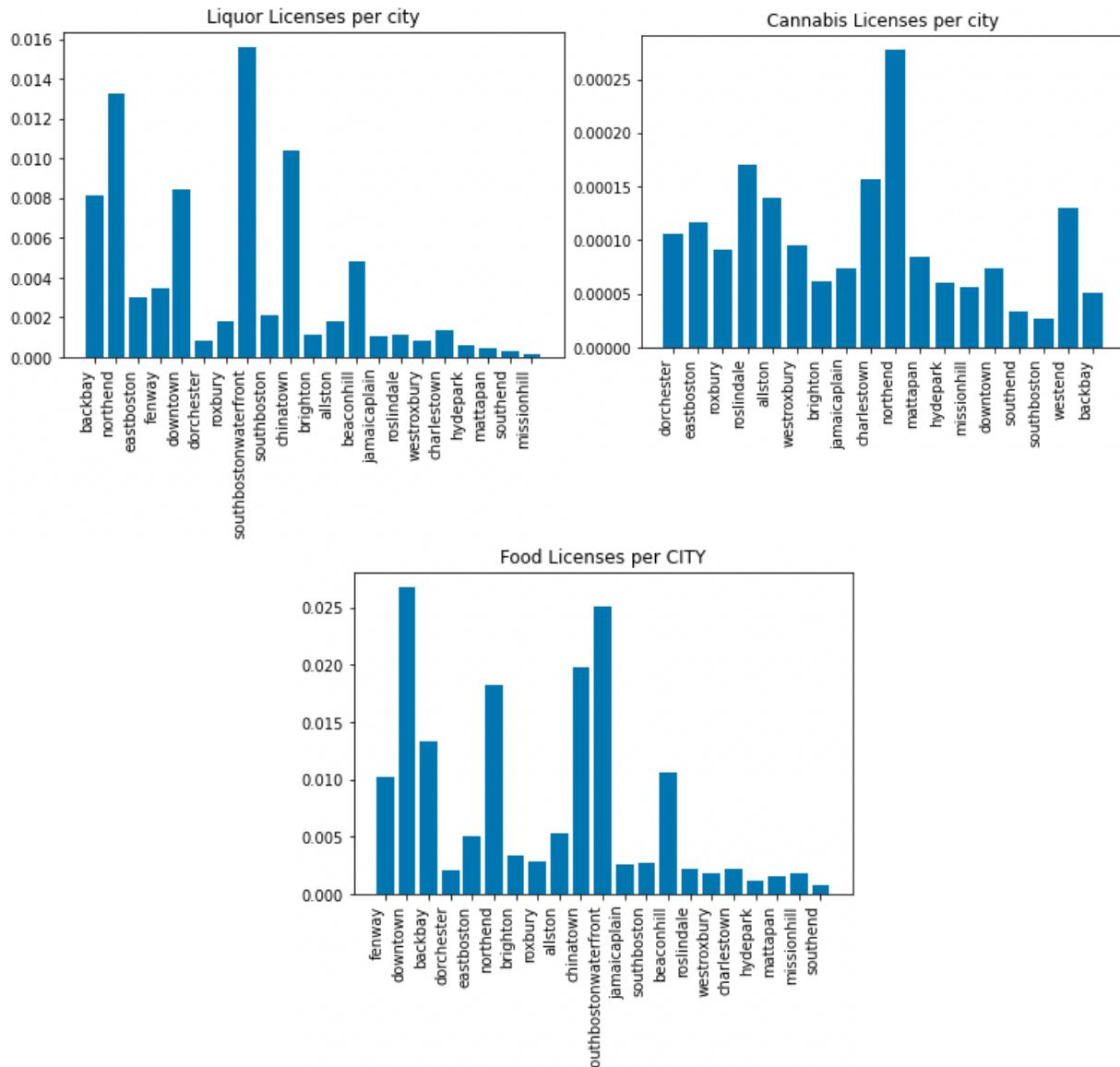


We also analyzed the amount of food licenses specific to take out, which is especially relevant given a global pandemic or any future public health emergency. Again, this is consistent with the above findings, suggesting on a surface level that those communities on the lower end may not be prepared for a situation where restaurants are closed for dine-in.



Additionally, we analyzed the relative time periods where licenses are being issued. Liquor licenses and general licenses have seldom been issued since 2014, whereas food licenses are increasingly being issued. As for license expiry dates, they mostly seem to be located about 1-2 years in the future, suggesting that businesses do not have a lot of time to update their licenses.

An issue we ran into at this point was the ambiguity of the city term ‘Boston’, as that could potentially mean any neighborhood located in the city of boston. As such, we planned to de-aggregate this term based on the specific zip code provided and matched that to recognized neighborhoods in boston. As such, we managed to incorporate the census dataset from before into our analysis to measure the number of licenses per capita present in each neighborhood.



From these charts, the relative distribution of businesses in each neighborhood begins to take shape. Neighborhoods such as the North End, South Boston Waterfront, Chinatown and Downtown have relatively high amounts of food and liquor licenses per capita, while Dorchester, Jamaica Plain and West Roxbury are places on the opposite spectrum. This suggests that there may be an imbalance in the availability of food and drink depending on the neighborhood in Boston. Further analysis could be done to determine how this compares to city finances and whether the number of business licenses is impacted by the amount of funding it receives.

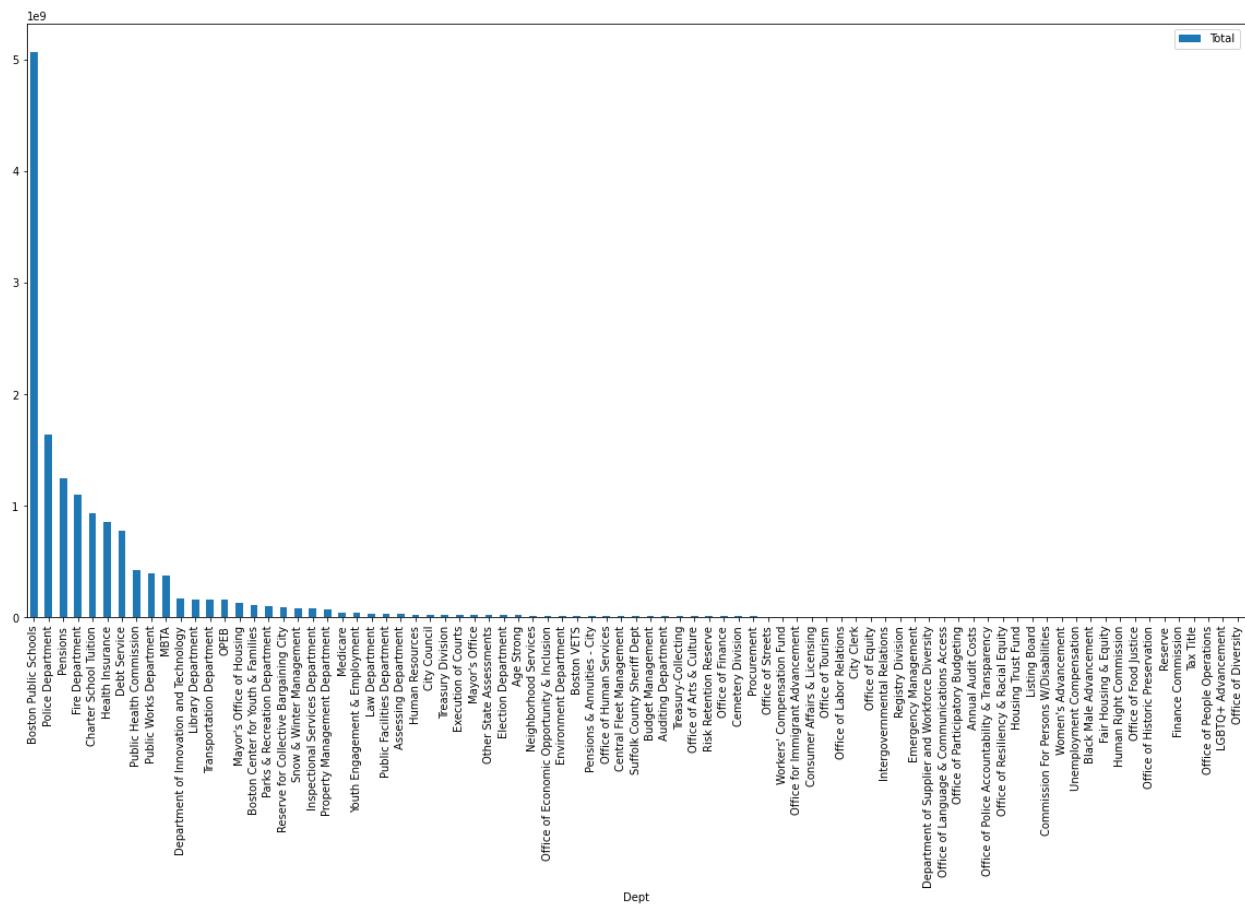
Capital Budget Dataset:

This data set is the FY23-27 Boston Capital plan. It is the City's plan to make critical investments in the City's infrastructure like, schools, roads, bridges, bike lanes, libraries, and more. This year's \$3.6B plan includes new infrastructure funds supporting affordable housing, revitalized parks, multi-modal transportation improvements, school improvements across the city.

After a series of calculations, we have calculated the total amount of the capital budget is 10,086,802,849 dollars. Below is a list of distribution indicating where the total budget goes:

Rank	Department	Total	percentage
0	Public Works Department	2585431733	0.256318
1	Boston Public Schools	2382281862	0.236178
2	Property Management Department	1084043619	0.107471
3	Parks and Recreation Department	1001583019	0.099296
4	Transportation Department	627863344	0.062246
5	Mayor's Office of Housing	427500000	0.042382
6	Environment Department	384824797	0.038151
7	Boston Public Library	325440000	0.032264
8	Fire Department	311263191	0.030858
9	Police Department	296820000	0.029427

10	Public Health Commission	269424000	0.026711
11	Boston Centers for Youth and Families	249450000	0.024730
12	Department of Innovation and Technology	70297284	0.006969
13	Office of Arts & Culture	54980000	0.005451
14	Boston Planning and Development Agency	15000000	0.001487
15	Office of New Urban Mechanics	600000	0.000059

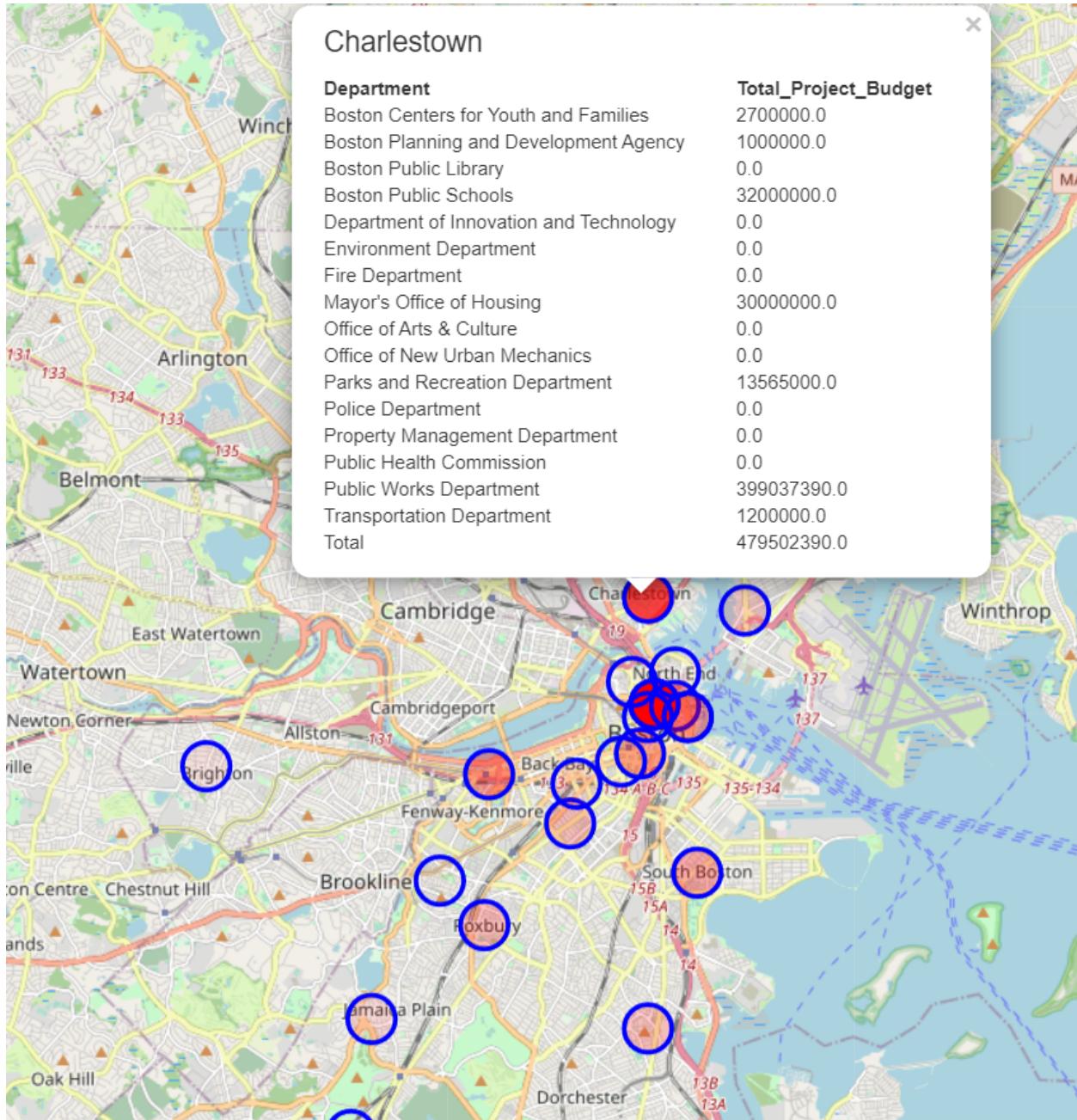


The table shows that the majority of the city budgets went to education and public facilities. However, just looking at the number won't help to determine the differences in equality in development for different neighborhoods. Therefore, we will divide them by neighborhood.

Neighborhood	Total_Project_Budget	percentage
Citywide	680174651	0.193289
Charlestown	479502390	0.136263
Various neighborhoods	341588600	0.097071
Fenway/Kenmore	239348258	0.068017
Chinatown	238872172	0.067882
Government Center/Faneuil Hall	209323273	0.059485
South Boston	197815000	0.056214
Roxbury	158723299	0.045105
Dorchester	139578681	0.039665
South End	139295034	0.039584
Harbor Islands	134396224	0.038192
Financial District/Downtown	128850000	0.036616
East Boston	101632890	0.028882
Jamaica Plain	73850000	0.020986
Back Bay	58293934	0.016566
Mattapan	48821850	0.013874
Allston/Brighton	46360000	0.013174
West Roxbury	29710000	0.008443
Beacon Hill	26100000	0.007417
North End	11975000	0.003403
Roslindale	8240000	0.002342
Mission Hill	7497212	0.002131
Hyde Park	6520000	0.001853
Various Neighborhoods	5960000	0.001694
West End	5000000	0.001421
Bay Village	1525000	0.000433

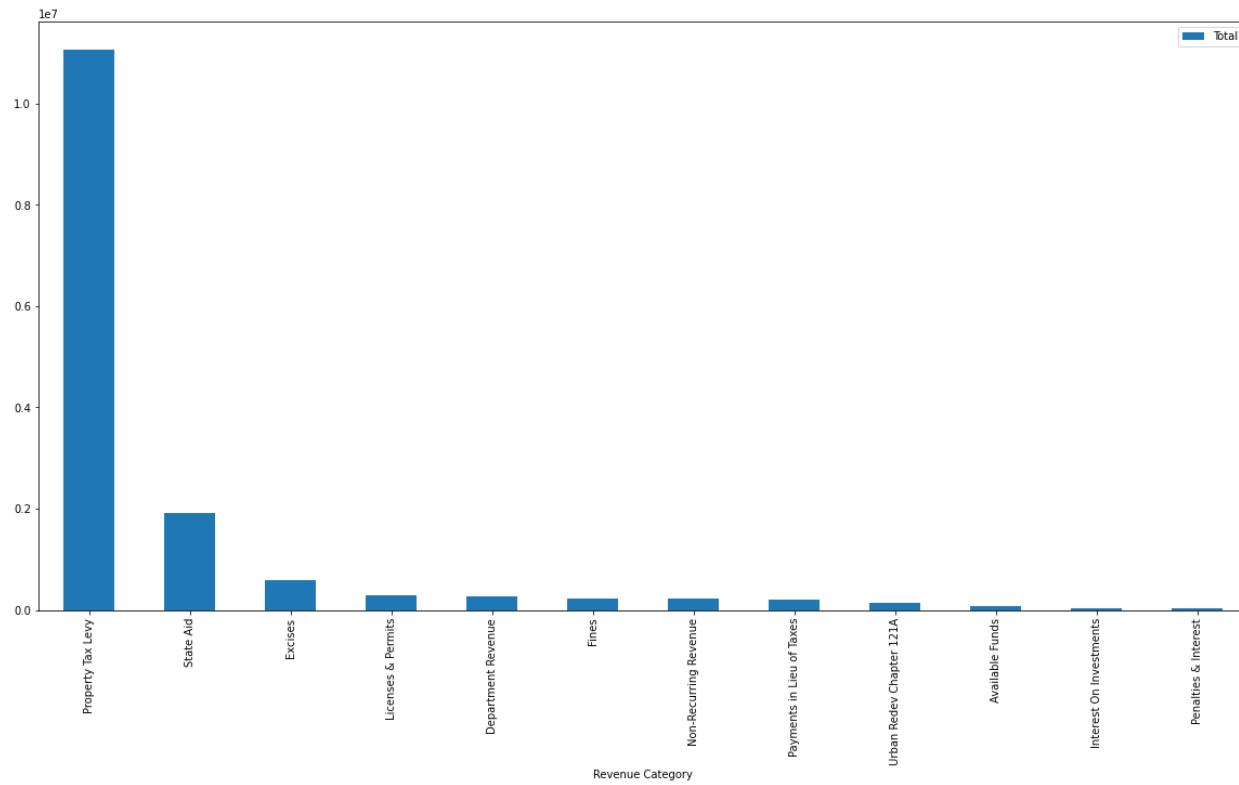
This shows most of development is happening in Charlestown, but it does not necessarily mean the city is biased toward them.

Below is a snapshot of an interactive [map](#) that we generated through the data, in which the darker the color of the circle, the more of the budget they received.



This indicates that the Charlestown area allocated most of its money into the Public Works Department.

Below is a chart indicating where the Capital Budget mainly comes from.



Rank	Revenue Category	Total	Percentage
0	Property Tax Levy	11065636.0k	0.734546
1	State Aid	1908654.0k	0.126698
2	Excises	593472.0k	0.039395
3	Licenses & Permits	284039.0k	0.018855
4	Department Revenue	257967.0k	0.017124
5	Fines	229280.0k	0.015220
6	Non-Recurring Revenue	215000.0k	0.014272
7	Payments in Lieu of Taxes	208757.0k	0.013857
8	Urban Redev Chapter 121A	142581.0k	0.009465

9	Available Funds	81900.0k	0.005437
10	Interest On Investments	38886.0k	0.002581
11	Penalties & Interest	38423.0k	0.002551

It is straightforward that a lot of money comes from property tax, and some neighborhoods are richer than others, therefore, it can explain why some places have more budget for their neighborhoods than the others. However, it would result in a bad circle, because the more investment they get, the more expensive their land and houses are, and the more future budget they will get, which is very unfair to the other neighborhoods that are underdeveloped.

Challenges and Roadblocks

A major roadblock that we faced was merging different datasets together. This was challenging given that we need to find a good feature upon which to merge the datasets. We considered a number of features upon which to merge the data, including zip codes and city names which were often the most common ones to appear. Oftentimes, datasets would have an inconsistency that needed to be smoothed out, such as the business licenses dataset containing the keyword ‘Boston’, or the location data and zip codes not matching up from time to time due to errors in the dataset. Converting from different data formats was key but time consuming, such as converting from a local coordinate projection to standard latitude and longitude. Furthermore, some datasets had different ways of classifying locations, making it even more challenging to interpret and merge. For example, in the extension project data that is explained below, there was different data for the locations of “Allston / Brighton” and “Allston” and “Brighton.”

Cleaning the datasets was also rather challenging. There were a number of miscellaneous types of invalid data and we dealt with them in a number of ways. Overall, we wanted to minimize the amount of data lost by simply getting rid of rows with any invalid data, so we applied different techniques such as mapping the data and filling NaN values with the mean value if it was numerical.

Finally, a challenge presented itself in the number of ways to analyze features. Given that we analyzed roughly 15 different data sources, each having a number of different features to analyze, there are many combinations of features that could be examined for results. There are so many possibilities when it comes to the potential analysis, and we have captured a small subset of the potential of the rest of the dataset.

Limitations and Assumptions

As for the business license data, one key challenge we faced was with interpreting the license issue and expiry data. It seemed like the majority of these licenses were either issued at the very beginning of the range of time, or set to expire at the very end of the range of time. This is rather ambiguous, as it could be interpreted that either these licenses begin/expire exactly on those days, or that the licenses could have begun/will expire outside the range of time but were aggregated to that date for some reason. In addition, we don't know enough about the license renewal process to know whether it is common for licenses to be issued at a specific time every x years rather than evenly distributed across time.

We are assuming that some neighborhoods have higher property tax, which means housing is more expensive, due to the fact that some people think certain neighborhoods are safer and have better influences than the other neighborhoods. Therefore, the popular neighborhoods would tend to have higher housing prices because more people want to buy houses from there, hence property tax is higher in these neighborhoods.

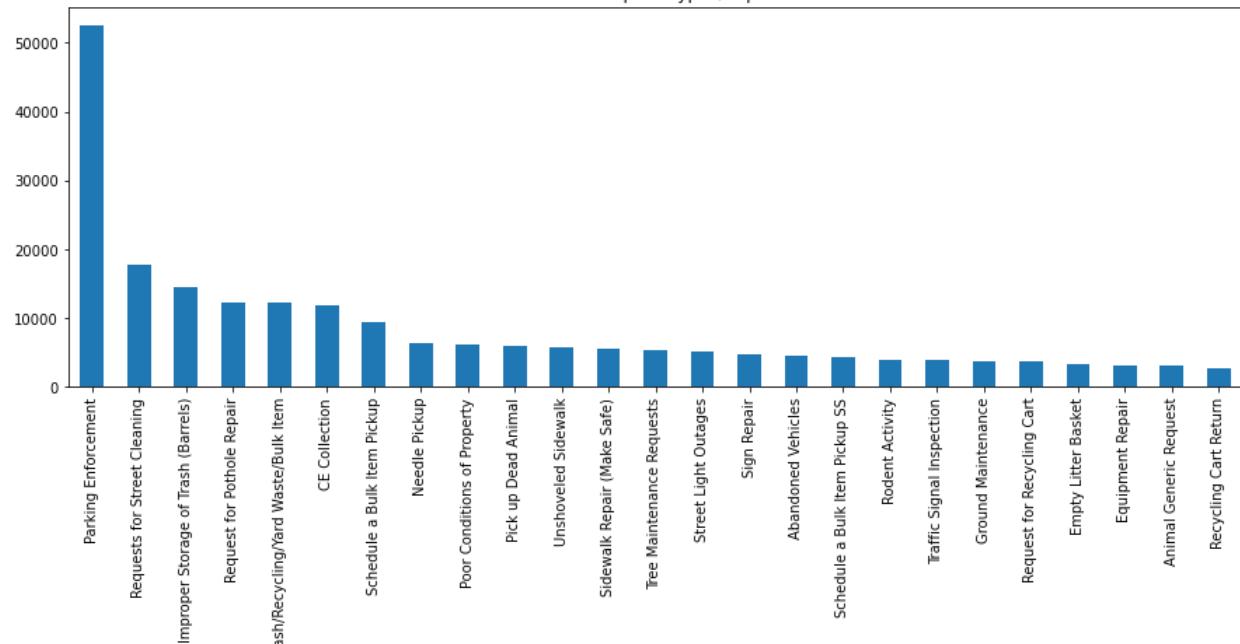
Extension Project

After completing analysis on the base project, we had further questions about how different neighborhoods of Boston had different issues that needed to be addressed by the city, which is why we decided to continue our work by analyzing 311 Service Requests. Boston 311 is a service center where residents can report non-emergent issues that are handled by different departments within the city, ranging from Public Works to Transportation and more. These service requests placed by residents can all be found publicly, and we worked specifically with the [2022](#) dataset.

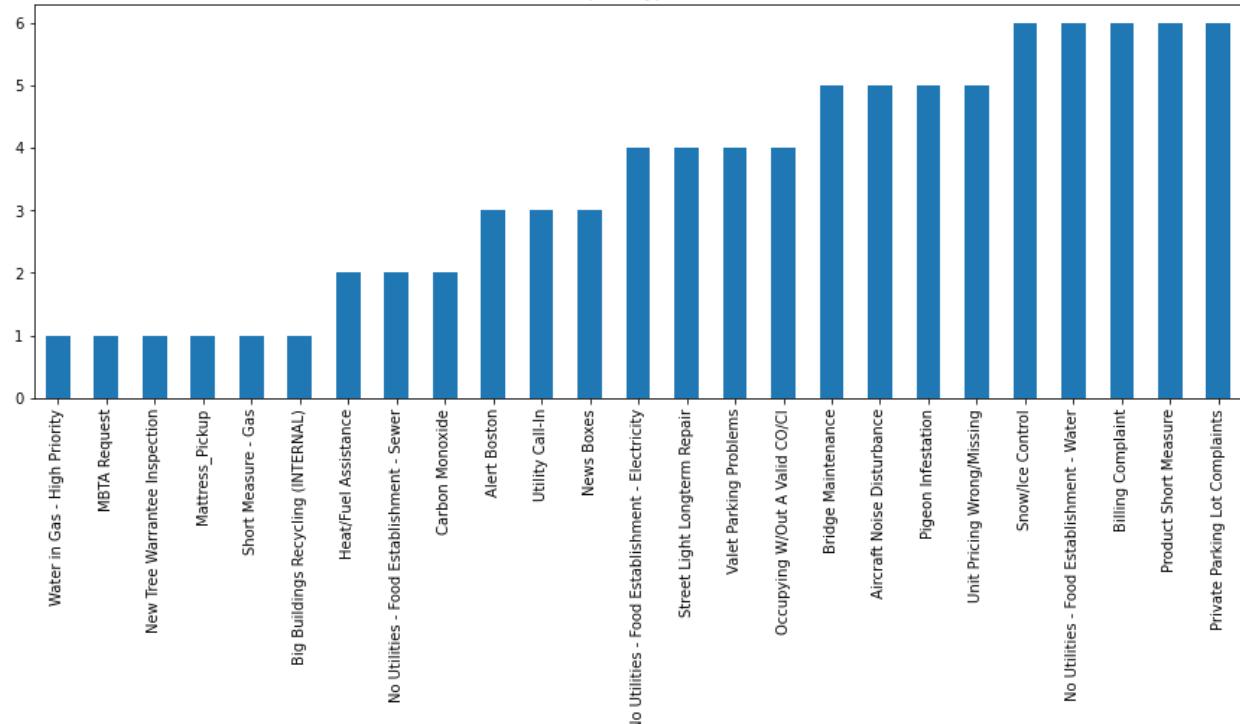
The dataset provided us with an entry for each and every dataset that was placed so far, and there were approximately 250,000 entries thus far. Each entry was recorded with longitude, latitude, the type of the request, source, closure status, and more. Our focus was placed on the type of requests.

There were 171 unique request types. These ranged from as many as 52,592 requests of that type, as with Parking Enforcement, to just 1 request of that type, as with MBTA Request for instance. Throughout all of the Boston neighborhoods, Parking Enforcement was by far the most requested service type, with Requests for Street Cleaning coming in second with less than half the number of requests.

Service Request Types, Top 25

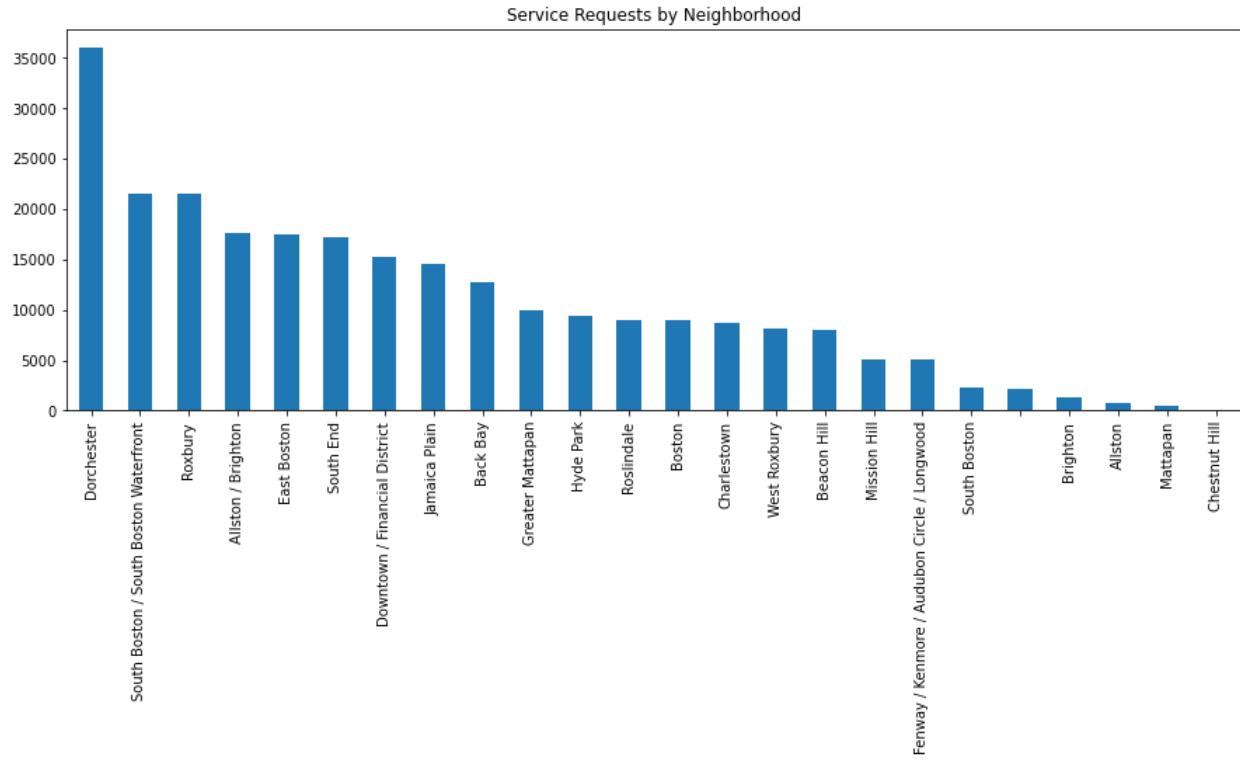


Service Request Types, Lowest 25

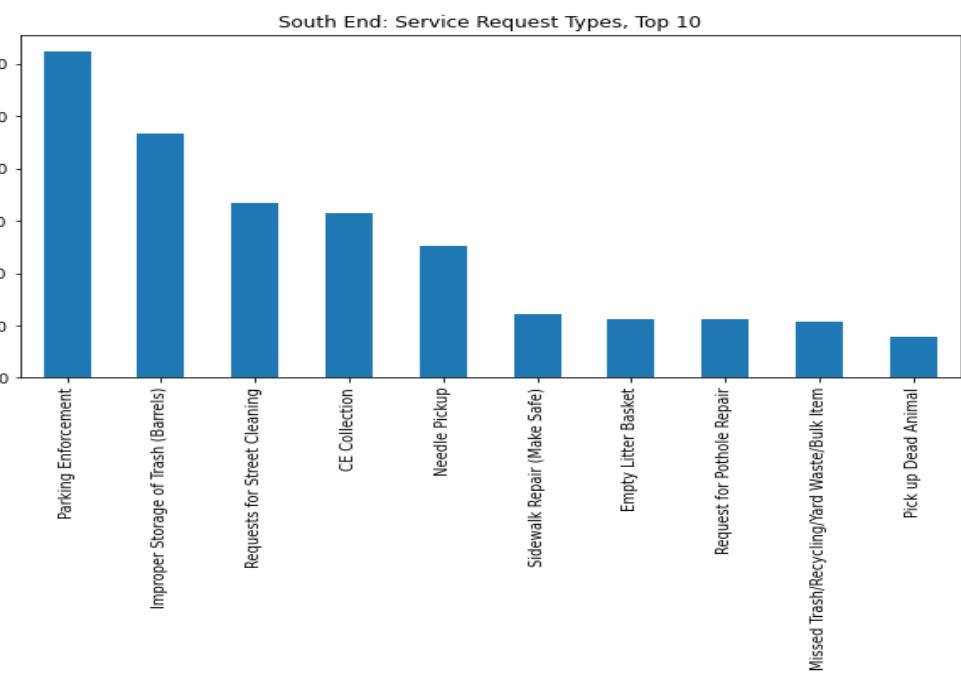
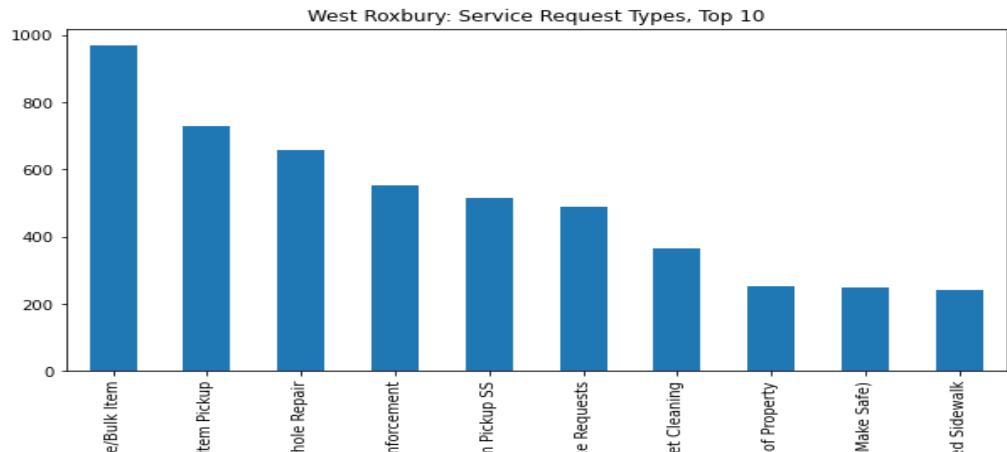


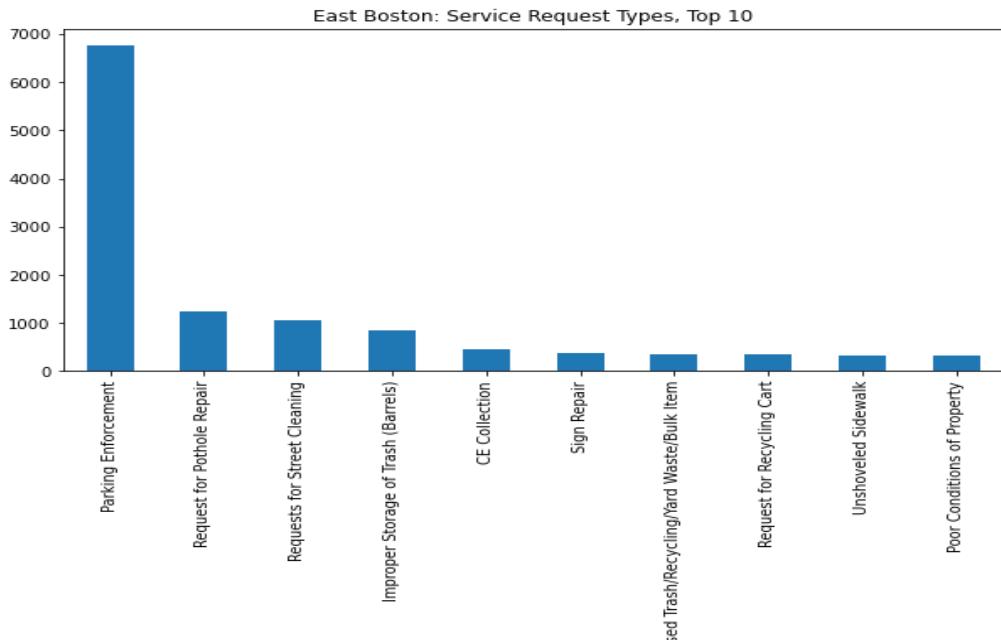
Out of all the Boston neighborhoods, Dorchester has unsurprisingly had the highest number of requests, with 36,038 requests, considering Dorchester's large population. Other neighborhoods,

like Allston and Mattapan, have had less than 1000 requests, and Chestnut Hill is the only neighborhood to have less than 100 requests thus far.

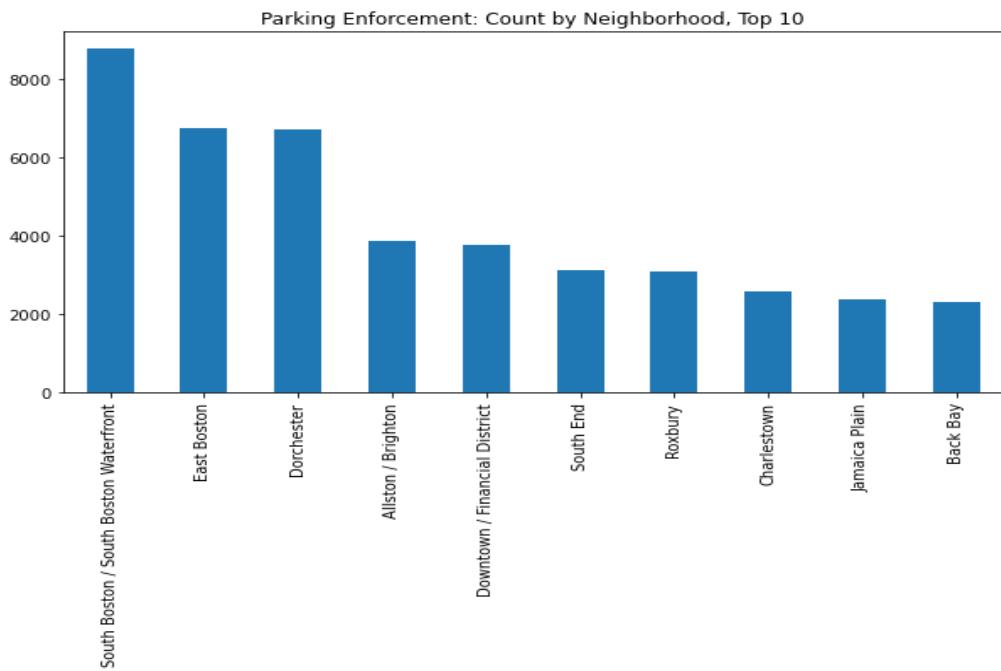


We then analyzed each neighborhood to determine their most popular types of service requests and see if we could discover any significant differences. All of the neighborhoods had Parking Enforcement as their most requested service type, except for West Roxbury, Beacon Hill, Mattapan, Greater Mattapan, and Hyde Park. However, the more interesting findings came from the next few most requested types. For example, only 6 neighborhoods, which are Downtown / Financial District, South End, Back Bay, Beacon Hill, Fenway / Kenmore / Audubon Circle / Longwood, and Boston had Empty Litter Basket within their top ten most requested service types. Below are three example charts for West Roxbury, South End, and East Boston, which each have different distributions of types and well varied most requested types.

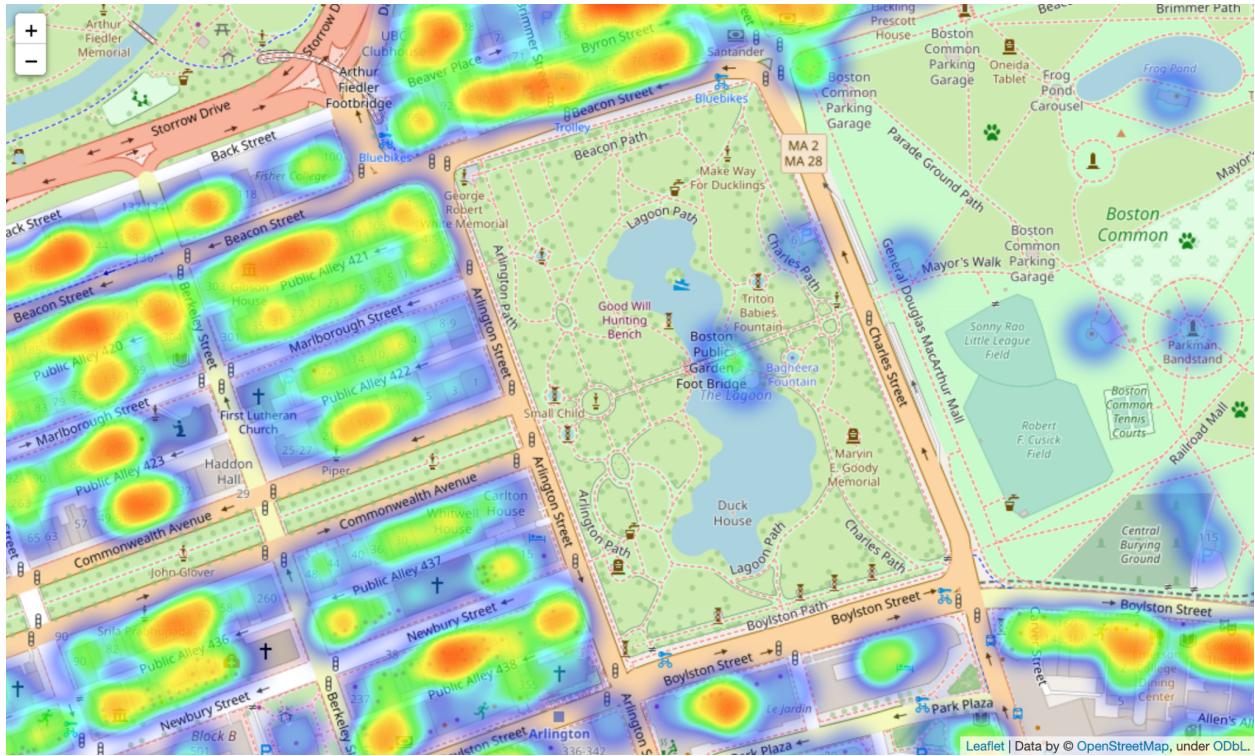




We also analyzed each of the top 10 most requested types of service to determine how they were distributed across the neighborhoods in Boston. Below is an example chart for Parking Enforcement, the most requested type of all, which shows that it was most requested in South Boston / South Boston Waterfront.



Finally, we created an interactive heatmap to visually display the locations of 311 requests across the city of Boston. Below is an example snapshot which shows the requests in and surrounding the Boston Public Garden.



Future Work

We want to be able to continue this work by answering the following important questions for each major class of datasets:

Business Licenses:

- How do business licenses compare by different names in different cities? Do certain businesses dominate certain cities/areas?
- How do opening and closing times compare across different cities? Does that impact the number of businesses in that area?
- How are businesses distributed by license category or type?
- Looking further into Cannabis dataset: map out different application statuses, and the split of facilities seeking Boston Equity Program and where they are respectively located, map out different types of Marijuana Licenses.

Rental Relief Funds:

- How are Rental Relief Funds distributed per capita per city?
- Is there a correlation between the above quantity and the capital budget or the amount of business licenses per capita in a given city?

Capital Budget:

- How does the money from different city departments diffuse into different communities and cities?

Additionally, in any future work that we might undertake, we have to be careful not to make any assumptions that might negatively affect people in their respective neighborhoods. To elaborate, we must avoid stereotyping or reinforcing any stereotypes about neighborhoods in Boston and make sure that neighborhoods are given equal chances of success given their predispositions.