Analyzing 311 Requests - Team 3

Thiru Satya Surya Mahaveer Bonagiri (2023)
mahaveer@bu.edu

Mary Choe (2025)
marychoe@bu.edu

Christian DeAsis (2023)
cdeasis@bu.edu

Aryaan Upadhyay (2023)
uaryaan@bu.edu

April 28, 2023

1 Introduction

1.1 Project Description

At-Large City Councillor Julia Mejia seeks to understand the city's response to 311 service requests and if they are resolving them in an equitable manner. We used 311 data to understand which communities feel empowered in Boston to demand services and how the city responds to empowered residents across the city.

1.2 Importance

This project is important as it helps the councilor understand if the 311 call requests are addressed in an efficient manner. This is important as for smooth and effective functioning it is important that calls are addressed in a timely and appropriate manner. This also allows us to understand which communities feel empowered to use the 311 call service.

1.3 Methodology

To understand the discrepancy of the unresolved cases to the resolved cases, we decided to analyze the datasets in 3 parts. The first part included us analyzing the unresolved cases, while the second involved us analyzing the resolved cases. The third part will include us analyzing the communities around which the unresolved cases exist. We decided to do this because this would help us analyze and breakdown the different parts of this project in an efficient and thorough manner. The datasets we used for the first, second, and third part were the census data for 311 requests for Boston from 2020 - 2024, census data for 311 requests for Boston from 2020 - 2024, and the social vulnerability index.

To navigate the first part, we decided to do an analysis of the unresolved 311 cases by neighborhood/zip code. We did this as we wanted a general consensus of the ratio of the unresolved cases per neighborhood. This helped provide a general overview of where most of the requested unresolved cases lied. To further build on this part of the project, we decided to granualize the data by analyzing the zip codes through different blocks. This involved appending block IDs to each row in the dataset, which was accomplished through a combination of geocode API utilization and extensive data preprocessing and merging. Consequently, we re-examined the unresolved requests by block, which yielded insights that differed from those obtained through zip code analysis, ultimately resulting in enhanced clarity.

To navigate the second part, we decided to do an analysis on the resolved 311 cases. We followed the basic template that we had developed from part 1 as this was the most beneficial plan we had developed. Moreover, it helped us understand the question on a micro and macro level.

To navigate the third part we decided to do an analysis of people with limited English proficiency (LEP) using the social vulnerability index dataset. We wanted to find out how people with LEP's ability to use the 311 call service was affected. We checked it was negatively affected and as expected the data reflected that people with LEP tend to not use the 311 call service as much as those without this language barrier. We analyzed the data to see just how much this specific social vulnerability affected their ability to use the call service.

2 Key Questions

1. Conduct analysis of Unresolved 311 requests
2. What Service requests are most common for the city overall and by Geography
3. Different Unresolved request types
4. Ratios of Resolved vs Unresolved requests
5. rate of closure for different types of service requests by Geography

3 Overview

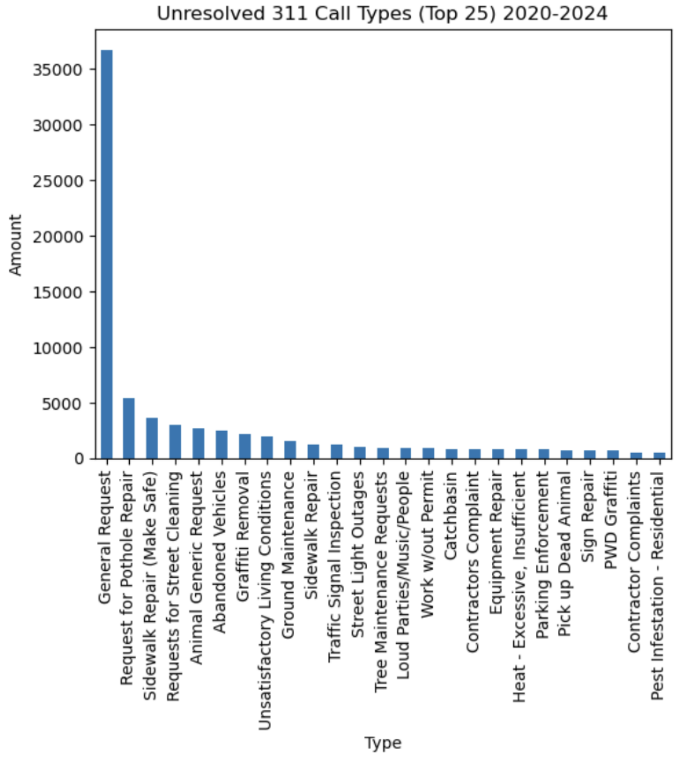
We extracted all of the unresolved 311 service requests and conducted an overview analysis of the responses. Then we went and added the block IDs to our main dataset to analyze the data according to different geographical divisions. We have graphs below showing and describing which service requests are the most common and how the communities are empowered based on the different service requests types. Then, we looked at the rate of closure for different types of service requests across different geography. It was a bit difficult to understand which requests were actually closed and which were not so we add the reason of closure for each request. We were able to find this from a different dataset and added it to our own. Below, we go into depth of our answers by showing all of the graphs we made and explaining them in detail and how they answer the key questions.

4 Analysis

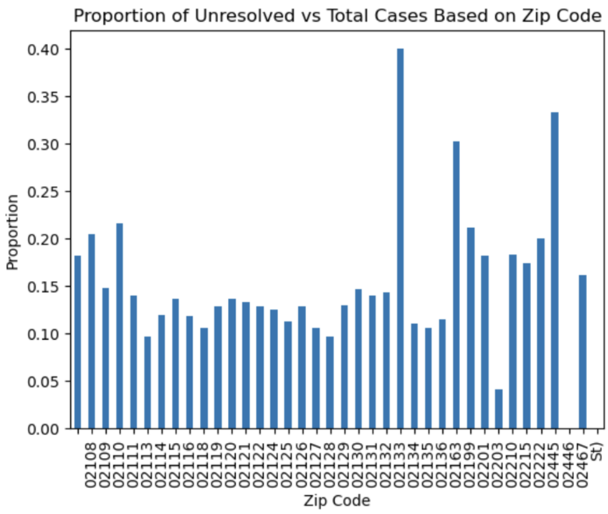
As the questions asked us to analyze the 311 calls for the time period of 2011 - 2024 we decided to use the different datasets over this time period. However, we ran into issues with this process due to the amount of information we had. Our computers would often crash and fail to perform the analysis on the datasets due to sheer volume of records. In order to combat this we decided to perform our analysis on the time period from 2020-2024. This proved beneficial as it was the latest set of records with updated values. Moreover, it would account for spillages. By that we mean that old possibly open cases will either be accounted for within this dataset by either being closed or being open. The following are the components mentioned above.

4.1 Analysis by Zipcode

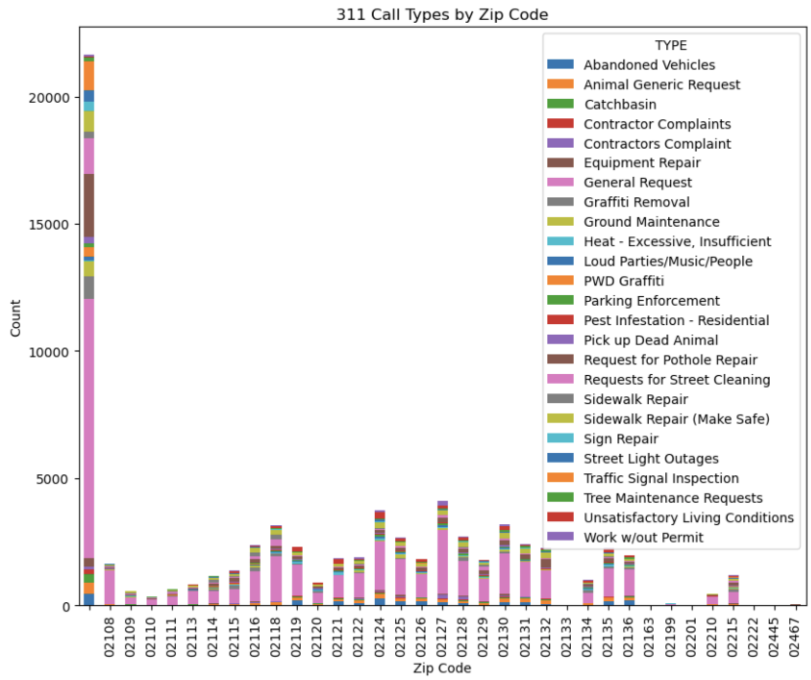
To understand the occurrences of certain request types for unresolved cases, we decided to see the 25 most requested types. This was important as it helped us understand why certain requests are more frequent than others. As seen from the graph below, "general requests" are leading by a significant margin. This makes us hypothesize that the associated authorities should work with the 311 request-solving department for reducing these general requests.



To build off the above graph we decided to find the proportion of unresolved cases by zip code. This proved crucial as we wanted to understand how each zip code contributes to the total number of unresolved cases. As seen from the graph below we realize that zip code 02132 has the highest proportion of unsolved cases. We dug deeper to understand why there were sudden high spikes in the proportion of unresolved cases. We found that within the dataset that we used, there were several locations that were at intersections and not definitive locations. These locations were more prone to 311 requests. These locations were included in zip codes such as 02132.

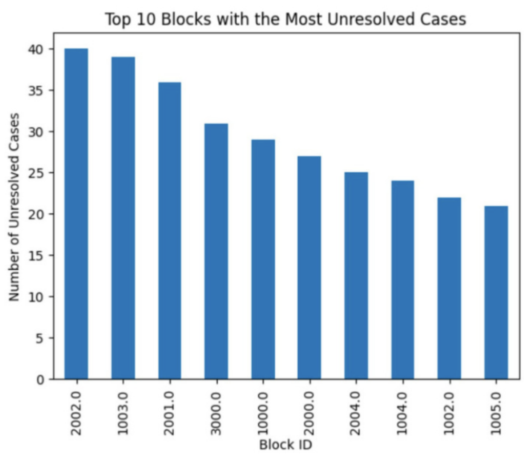


Having seen why certain locations/zip codes were more prone to having unresolved cases, we wanted to understand the breakdown of these calls based on zip code. We did so because we tried to find how location impacts the type of call being requested. From the below graph we found that our above finding is not consistent with the number of calls based on zip code. While 02132 does not have the highest count of requests it has the highest proportion of unresolved cases. This means that more attention needs to be prepared to this zip code. This theme can also be applied to different zip codes to understand where more attention needs to be paid.



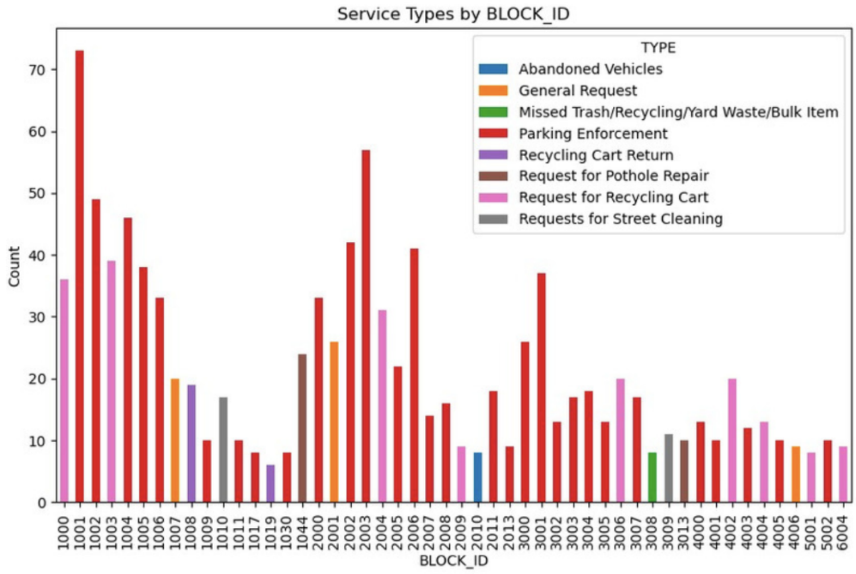
Additionally, in the graph above we see that the first bar is astronomically high - this is because certain locations did not have any zip codes mentioned.

4.2 Analysis by Block

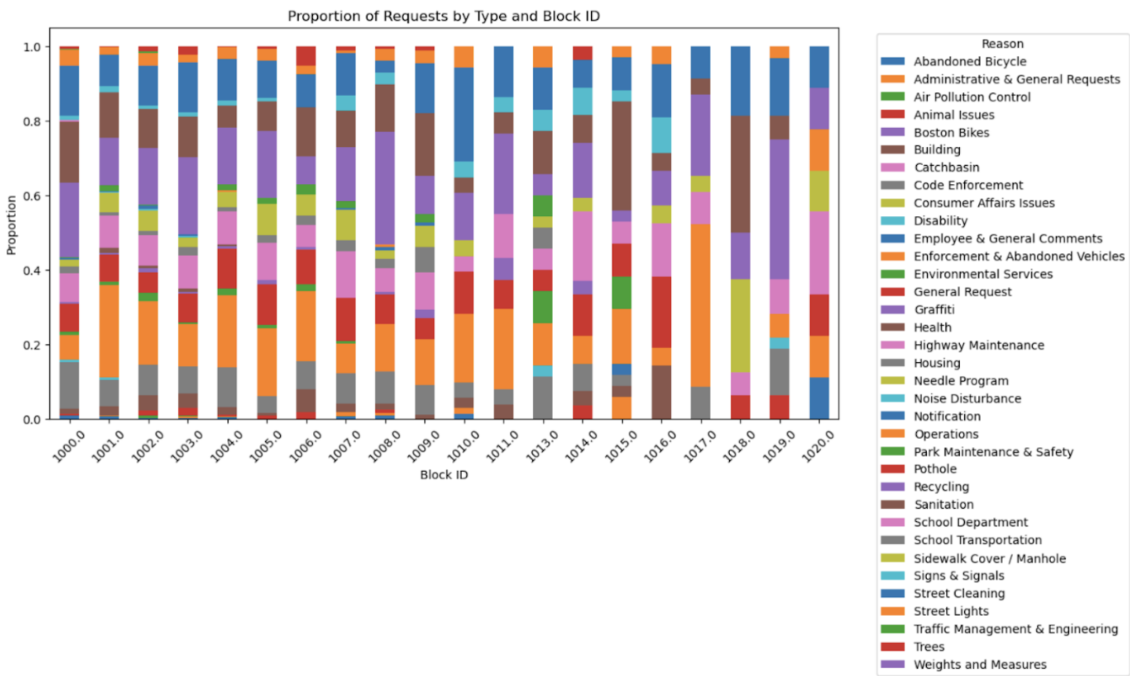


For the above graph, we take a closer look at unresolved cases and blocks. In the graph to the left, we can see the top 10 blocks that have the highest number of unresolved cases. This is important because it leads to questions as to why certain blocks seem to have a higher number of unresolved cases in relation to others. Perhaps to take a deeper look into this, we can look into proportions, such as the proportion of unresolved vs total cases in these 10 blocks, which can indicate whether or not these numbers are statistically significant or not. For example, if a block has a high raw number of unresolved cases but has a low proportion of unresolved vs total cases, this can be seen as somewhat acceptable, whereas the opposite would point to larger problems. As the project moves along, one of our goals is to continue to see why certain blocks or areas see higher overall traffic and a higher overall amount of unresolved cases.

In the graph above, we can see the denominations of service types, or 311 request types, by block ID. From here, we can see that the most common seems to be parking enforcement, with the second being requests for recycling carts. From this, we can draw some simple conclusions that parking



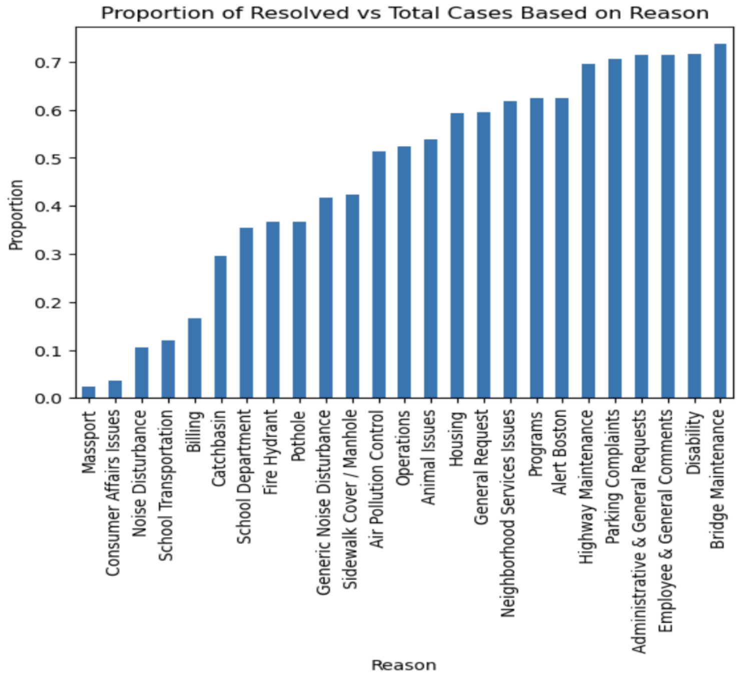
does seem to be a glaring and popular problem throughout the different blocks, as well as a need for sanitation and recycling. Being able to see these requests visually can be very helpful to see the overall problems that persist throughout the area, as well what problems have higher priority. It is also important to note that in this graph, we only included the most common type per block ID, and thus it is crucial to understand that these block IDs also have other various underlying types and issues not depicted on this graph.

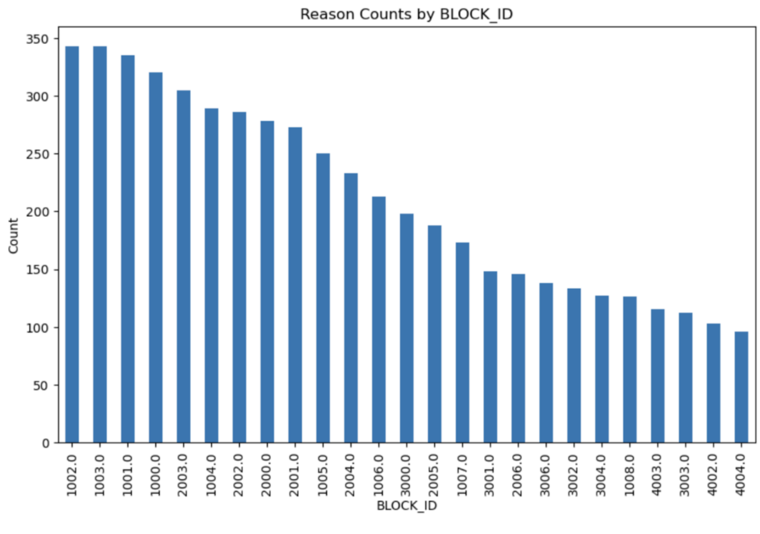


Upon observation, it can be noted that a greater number of requests have been raised for Block 1019 pertaining to Boston Bike issues. Furthermore, a higher incidence of health-related concerns has been reported for Block 1015, potentially indicative of suboptimal cleanliness and hygiene maintenance practices. Additionally, a significant volume of administrative requests for Block 1015 may suggest the presence of underlying issues requiring intervention from higher authorities. Therefore, the observed proportions may provide valuable insights into the challenges faced by the respective blocks.

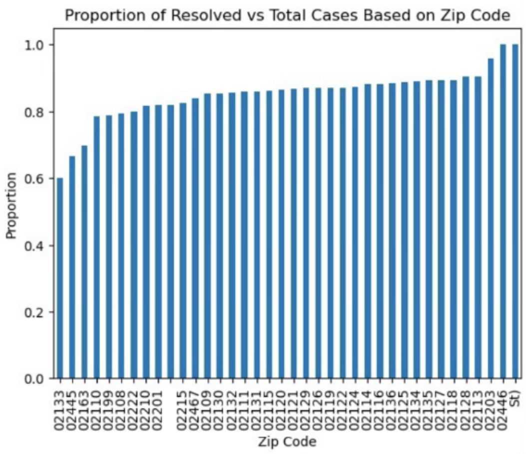
4.3 Analysis for Resolved Cases

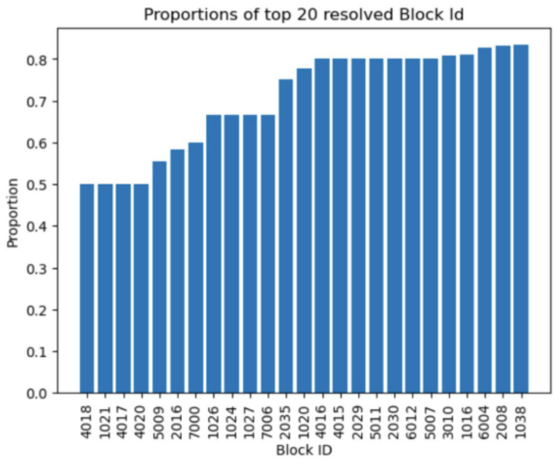
Having focused on the unresolved cases, we decided to move on to the resolved cases. We thought this was a great question to analyze because, like one of the above-mentioned graphs, it would help us understand/see if certain zip codes are being more focused on. However, before we could get to that, we also wanted to see if certain requests were more answered than others. This was a very important piece of analysis because it could indicate that certain request types were more catered to others. Using this approach we found that the highest proportion of resolved cases was bridge maintenance.





In the above graph, we take a further look at the total number of resolved cases vs total, based on the block ID. From this, we can see a clear declining trend, which indicates that certain blocks receive much more attention than others. For example, the block that sees the highest number of resolved cases has approximately more than 200 such instances than the block with the lowest number of resolved cases. Such a distinction is important because, as previously stated, it is quite evident that some areas have a better rate of resolution than others. This may be due to various factors such as their specific location or prominence or the amount of resources available at a certain area's disposal.





Above, we have two more graphs, both of which depict the proportion of resolved cases vs the total number of cases, with the left graph denoted by zip code and the right graph denoted by block. In the graph directly below, by zip code, we can see that overall, there seems to be around an 80% proportion of cases throughout the zip code areas that have resolved cases. There are also two cases in which the proportion is 100% and one that is about 90%. However, there are some areas that seem way lower, and by that I mean 70% or below, with the lowest proportion being 60%.

This information is important because while there are some cases of zip code areas having nearly 100% resolved to total cases, there are others that have glaring issues with resolution. Furthermore, having most of the areas having a proportion of around 80% suggests that one out of every five cases does not get resolved, which may be caused by various factors.

If we take a look at the right graph with block codes within the zip codes. Here, the data shows more variety, with the lowest proportion of resolved vs total cases being 50% and the highest being 80%.

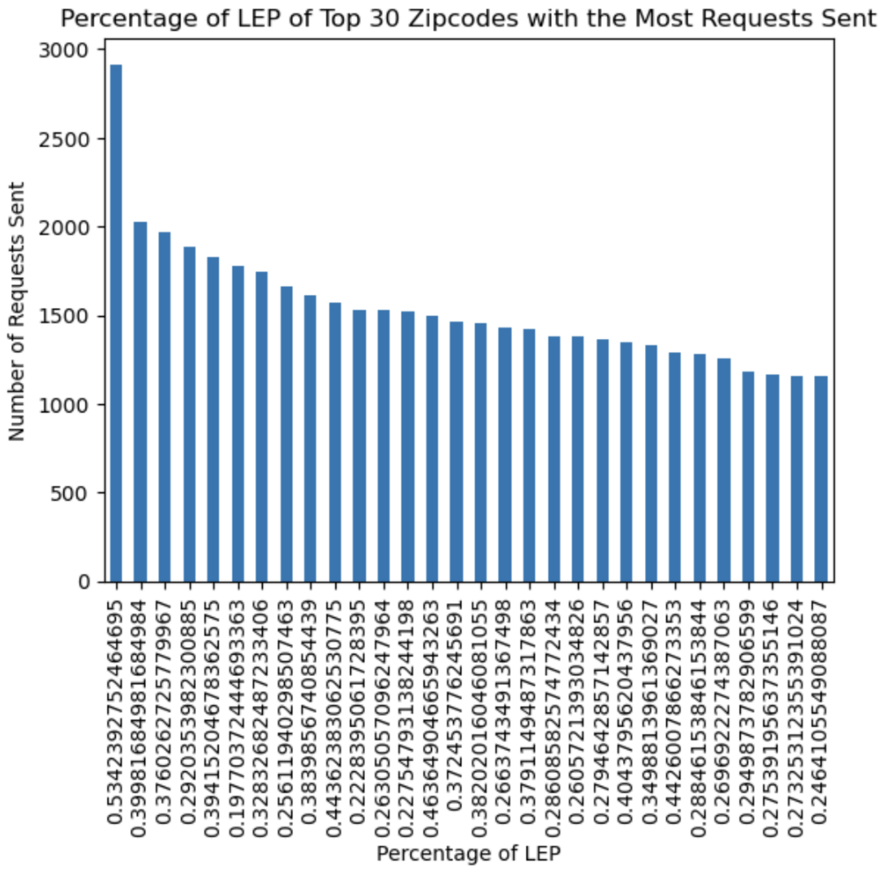
This graph is important because it works in accordance with the graph denoting the zip codes. It shows that, for example, while a zip code area can have a high proportion of resolved cases, the blocks within may have disparities, with some block areas within the zip code seeing higher proportions than others. It is also important to note that top 20 proportions of resolved cases have a minimum value of 50%, which indicates that other block areas

5 Extension Analysis

Our proposal is to analyze the data to better help those with limited English proficiency using the data from Analyze Boston's Social Vulnerability Index dataset. Without adequate English skills, residents can miss crucial information on how to prepare for hazards. Cultural practices for information sharing, for example, may focus on word-of-mouth communication. In a flood event, residents can also face challenges communicating with emergency response personnel. If residents are more socially isolated, they may be less likely to hear about upcoming events. Finally, immigrants, especially ones who are undocumented, may be reluctant to use government services out of fear of deportation or general distrust of the government or emergency personnel. The following segment provides a series of inquiries and offers corresponding responses to each.

1. Is there a concentration of people with limited English proficiency in certain geographical blocks?
2. Is it affecting the amount of requests that are made in proportion to other English-speaking blocks?
3. How does limited English proficiency actually impact the 311 calls?

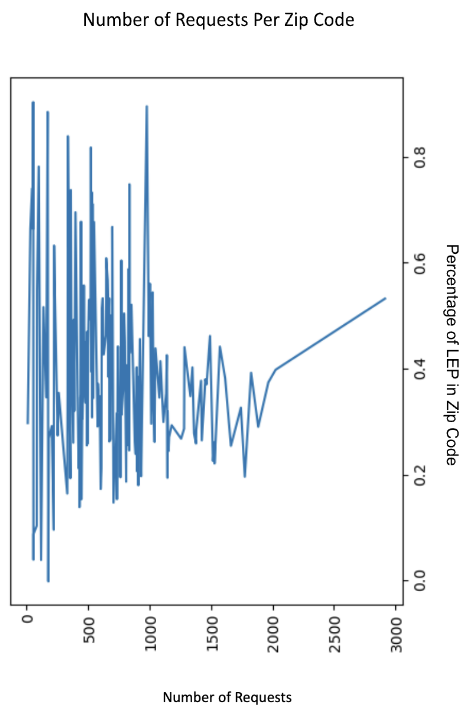
This graph shows the top 30 zip codes that have sent the most requests. As we are using information from the social vulnerability index, we are limited in the area divisions. We are only able to

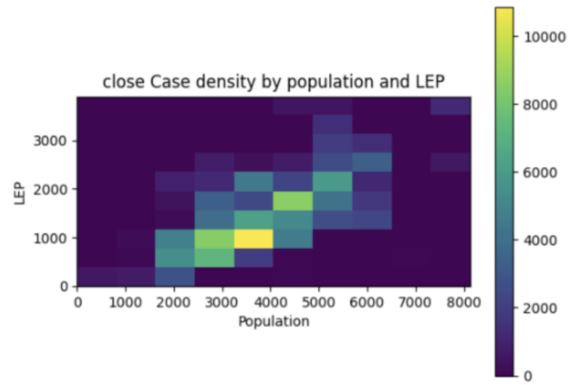


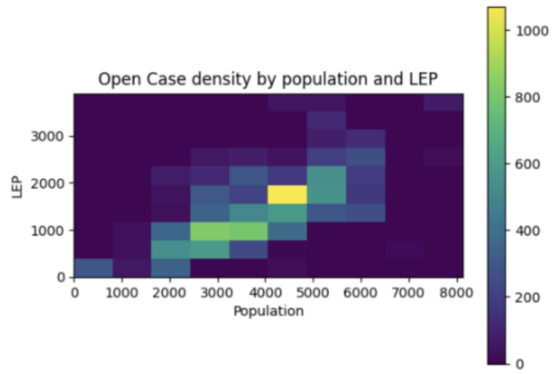
see the number of people with limited English proficiency (LEP) living in a certain zip code. We are not able to be more specific with the data such as that of the block level because the information available is too general. Based on what we were able to get, we found the number of requests made from each zip code. We were also able to see how many people in the zip code had LEP and analyze how many calls were made in zip codes with large numbers of people with LEP. From the number of LEP people in a zip code, we divided it by the total number of people living in the zip code area.

Above, we have a graph that shows how many 311 calls each zip code has made from 2020-2022. Instead of indicating which particular zip code made which amount of calls, we can see what percentage of people with limited English proficiency (LEP) reside in that zip code. While we do lose the information about which particular zip codes made that many requests, we already have graphs that look at the information. The reason for this graph is to analyze how people with limited English proficiency use the 311 call service if they do use it at all. Also, the zip codes themselves are simply area divisions.

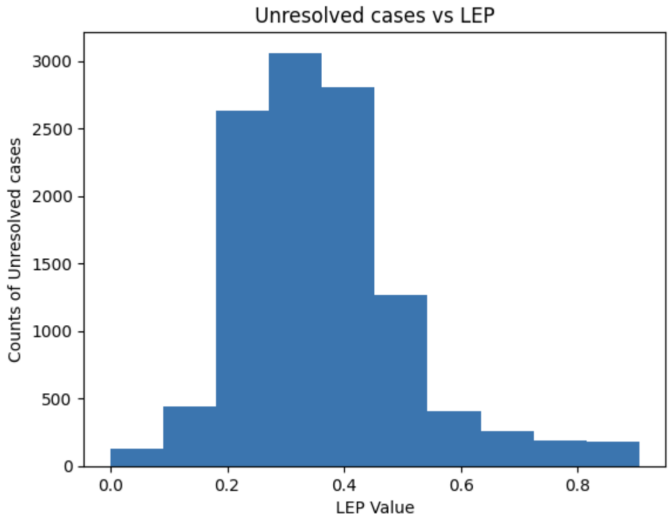
Based on the graph, we can see in general that the zip codes with a higher LEP percentage have fewer requests made. Specifically, we can see that all of the zip codes that made more than 1000 requests were all zip codes with an LEP percentage less than 60%. This trend is expected as the 311 service is a call service, and those with LEP people tend to be immigrants who would probably not know of this government service. It is also inherently more difficult for an LEP person to use the 311 call service as it is a call service.







The graphs presented above illustrate the relationship between Limited English Proficiency (LEP) and the number of open and closed cases in comparison to the population. The data reveals that individuals with limited proficiency in English tend to have a greater number of open cases, whereas those with a better command of the language tend to have a greater number of closed cases.



The graph above provide a detailed breakdown of the number of open cases based on the Limited English Proficiency (LEP) value. The data indicates that a significant number of unresolved cases, as denoted by the high value counts, correspond to LEP values of 0.2, 0.3, and 0.4. This observation suggests that individuals with lower English proficiency levels have a comparatively higher rate of unresolved requests.

6 Conclusion

In conclusion, those with LEP are making significantly less 311 calls than those who can speak English fluently. This may come from their ignorance of the existence of the 311 call service or how to operate it. This also may come simply from not being able to speak English and the 311 service is a call service that particularly requires speaking in English. It is also noticeable that in the geographical areas, in this case zip codes, that have larger amounts of people with LEP, there are less calls being made to the 311 service. Thus, the actual neighborhoods that have majority immigrants do make less calls.