Deliverable 2
City of Revere

Group Members

(Name/email/github)
Client: Reuben Kantor rkantor@revere.org,
Spark Liaison: Greta Bruce gretab@bu.edu,
PM: Lingyang Jiang lingyanj@bu.edu,
Yuanming (John) Chai jchai23@bu.edu, **ychai23**,
Mackenzie Knox mfknox@bu.edu **kenzieknox**,
Quinn Relyea qrelyea@bu.edu **Relishyeah**,
Giancarlo Sirio gsirio@bu.edu **gian-s,**
Xavier Ruiz xruiz@bu.edu **xav-ie**

More data should have been collected to perform a more thorough analysis of the data and attempt to answer one additional question relevant to your project proposal which you will submit as a pull request.

Collect and pre-process a secondary batch of data

The DPW data contains lots of data similar to the data used in the paper that used the Kansas City 311 data. We were able to clean the data by removing invalid addresses and addresses that could not be geocoded. The data itself is very dense and we will need to make a choropleth map of it to better visualize it (using census block groups). We believe with this data we can now perform a similar analysis to the aforementioned paper.

Refine the preliminary analysis of the data performed in PD1

The 311 data was further refined and the Request types were made into 28 distinct categories from a total of ~180 thanks to help from Reuben. In order to get the demographics for the people who are making 311 calls we used Census data to map each 311 call data point to a specific Block Group. The American Community Survey is what we used to get up to data Census estimates for the race and ethnicity for the city. The smallest denomination which data is publicly available for is the block group, which allows for the most flexibility in statistical modeling. We have not incorporated the race and ethnicity data yet, choosing to investigate where the highest population density occurs before characterizing the demographic data.

From the 311 data, we want to further visualize the distribution of the complaints geographically and find out the most dense area of certain requests. The reasoning for this is that we hope the graphs will help us examine which specific area needs which type of improvements most. Thus, we picked the top 3 requests -- Covid-19, potholes, and overgrowth/unsanitary conditions and plotted these requests on the entire revere city map. Following are our results and analysis.

We have begun to preprocess parking data, and found that much of it was parking meter violations, which skewed the data towards places of business. We will examine the data more and see if we can find data related to tickets outside of normal homes, we will certainly consider analyzing those implications.

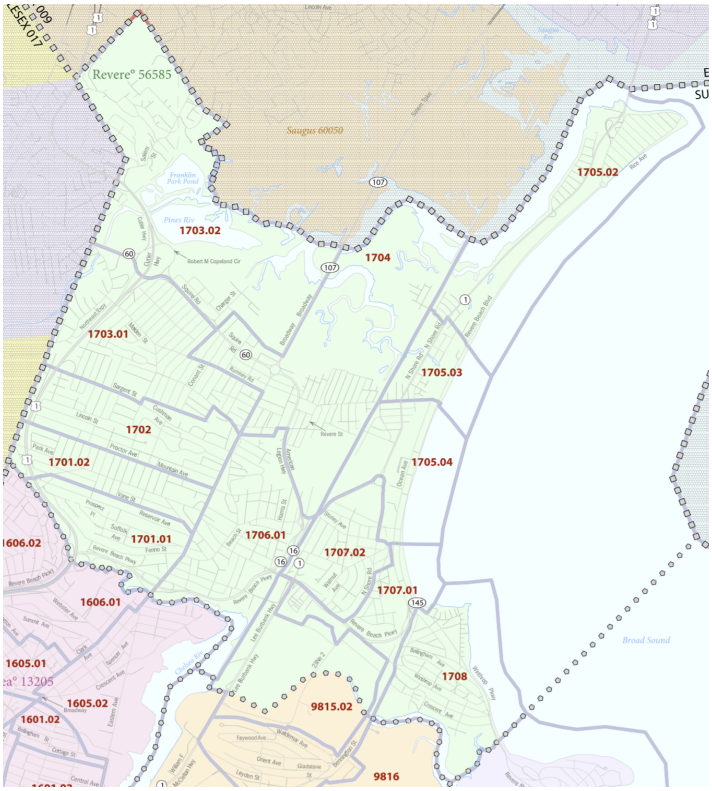


Figure 1: Map of census tracts

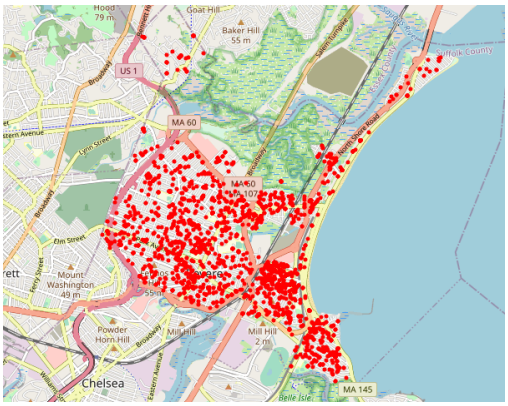


Figure 2: Covid related complaints distribution

Although the density seems to be universal throughout the city, specifically in areas such as Tract 1707.02, 1708 has the most dense complaints of covid-related issues and Tract 1705.02 has the most sparse complaints raised in the area. Thus, we can somehow conclude that Tract 1707.02 is the area that needs the most improvement on the social services relating to covid.

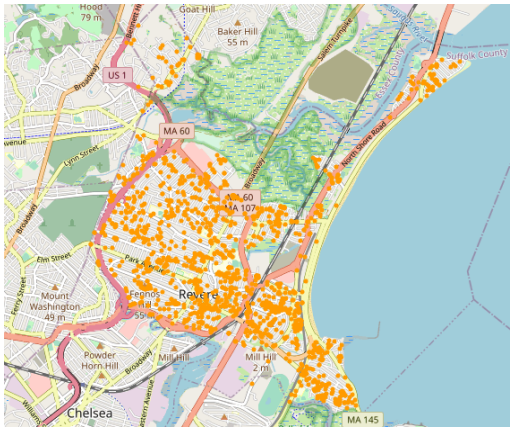


Figure 3: Pothole complaints distribution

Similar to the distribution of Figure 2, we can examine that 1706.01 and 1707.02 are the most dense areas and most requests for potholes are located in these two tracts.

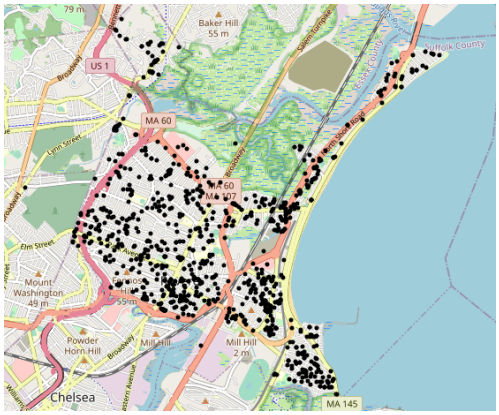


Figure 4: Overgrowth/unsanitary conditions complaints distribution

Again, similar to the distribution of Figure 2, we can examine that 1706.01 and 1707.02 and 1708 are the most dense areas and most requests for overgrowth are located in these two tracts.

Ranking	Tract	Population
1	Census Tract 1703, Suffolk, MA	9417.0
2	Census Tract 1701, Suffolk, MA	7935.0
3	Census Tract 1708, Suffolk, MA	6481.0
4	Census Tract 1707.02, Suffolk, MA	5798.0
5	Census Tract 1706.01, Suffolk, MA	5530.0
6	Census Tract 1704, Suffolk, MA	5041.0
7	Census Tract 1702, Suffolk, MA	4844.0

8	Census Tract 1705.01, Suffolk, MA	4117.0
9	Census Tract 1705.02, Suffolk, MA	2430.0
10	Census Tract 1707.01, Suffolk, MA	2099.0

Figure 5: Population ranking by Tract

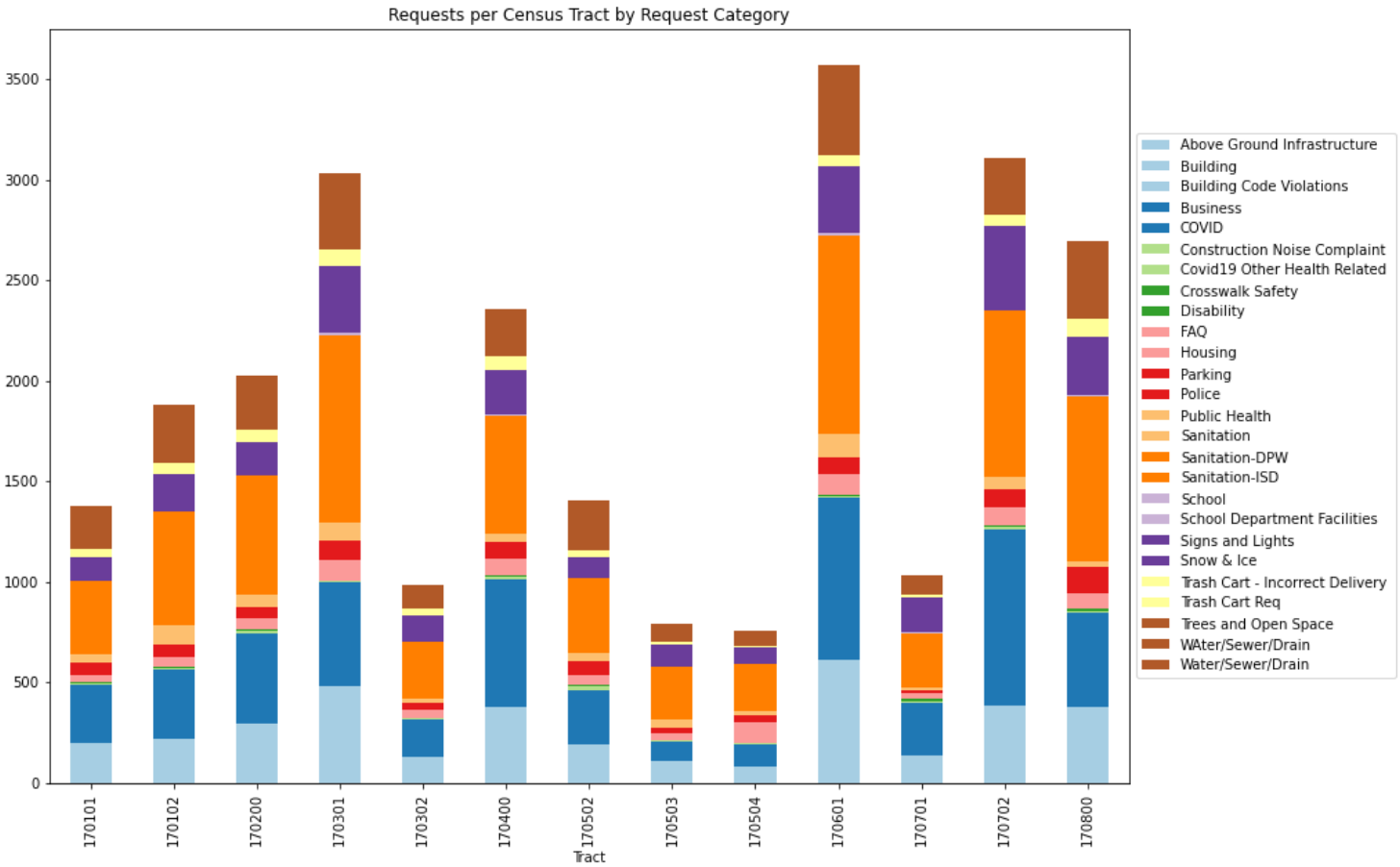


Figure 6: Total requests by tract, by request category

The above plot confirms the analyses of the previous figures, showing that tracts 1706.01 (Figure 5, rank 5th), 1707.02 (Figure 5, rank 4th), and 1708 (Figure 5, rank 3rd) have the most requests. In addition to the previously noted tracts, trat 1703.01 (Figure 5, rank 1st) has a significant number of requests, but this was not evident from the 3 top request categories (and was especially sparse in the overgrowth/unsanitary conditions request category).

Analysis Recap

From the analysis of top 3 complaints, we find out that Tract 1707.02, 1706.01 and 1708 are the areas with the most concentrated complaints for the chosen categories. There are a few potential causes for the density of complaints. The first is that the people who live within these tracts are more likely to report the issues than people located in the other tracts. Another possibility is that these tracts see less upkeep with things like potholes and

overgrowth/unsanitary conditions without prompting compared to the rest of Revere. It is possible as well that the majority of the population is located in these areas (since they have the top number of complaints) and the amount of improvements that need to be done should be focusing on these 3 tracts.

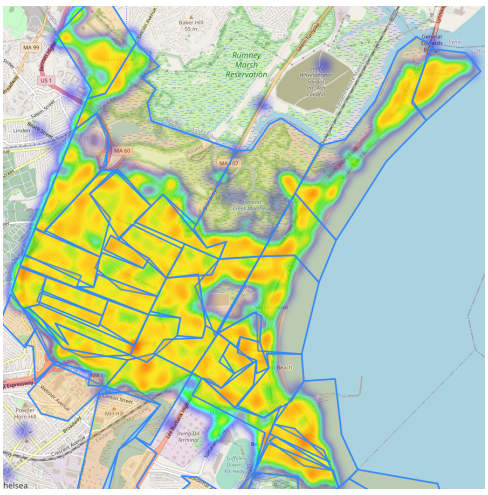
A future question we intend to address regarding the census data and Covid-19 request density is whether or not there is a correlation between request density (of both Covid-19 and non-Covid-19 requests) and average income, property values, or other census variables. Correlation between request density and housing density is another factor that may yield some insight into 311 requests. We will be investigating these relationships more as we proceed with our investigation.

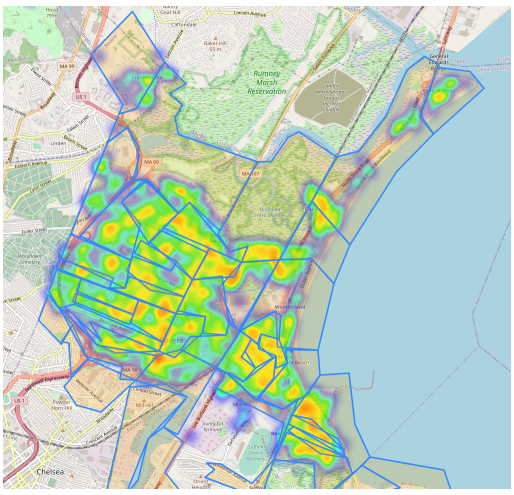
Answer another key question

Question: What region by block group is making the most requests?

Below we have a table with the five Census Tracts with the most 311 requests. As we can see Tract 1703 has the most with 3999 requests. This should be expected as it is the most populated of the Census tracts.

Census Tract	Number of Requests	Total Population	Requests per 100	Hispanic or Latino per100
1703	3999	9417	42.5	22.8
1706.01	3566	5530	64.5	42.1
1701	3263	7935	41.1	37.9
1707.02	3137	5798	54.1	50.7
1708	2709	6481	41.8	42.9





7.1 - Non-COVID related requests **7.2 -** All COVID related requests

Another question we are looking to answer is how 311 requests have changed pre and post COVID-19? From the heat maps we can see that the amount of COVID-19 related requests is spread out evenly. For future deliverables, we will work on improving the heatmap so we can get better distinctions for the amount of requests.

At this time, we do not yet have the statistical methodology to see if there is a disparity by race, but there is other research we are looking into to ensure that the methodology is proper and will examine any statistical significance that does exist.

Limitations and Risk Analysis

Limitations :

- Without actual demographic data for each individual request, it's hard to be super certain about the disparity based on race.
- Based on the research done concerning Kansas City's 311 data, we were able to find out that they also did not use demographic data to classify locations as under- or over-reporting but instead used pothole data and later analyzed the demographics.

Risk analysis:

- Currently the limitations we mentioned before may cause some uncertainty on the percentage of the results. However, based on the information we obtained, we are able to get an overview of the data and analyze the race disparity.

Project Scope

Correlate demographic location data from 2018 onward from the census with where people are requesting help from the city and where the public works funds are going.