

# City of Revere Final Report

## Team Members

(Name/email/github)

Client: Reuben Kantor rkantor@revere.org,

Spark Liaison: Greta Bruce gretab@bu.edu,

PM: Lingyang Jiang [lingyanj@bu.edu](mailto:lingyanj@bu.edu),

Yuanming 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**

## Executive Summary

### Key Takeaways

1. With regards to requests to 311, there are limited conclusions that we can make based off of the race and ethnicity of the supplicant because requests are not individually tagged. We were able to identify block groups that over or under reported incidents to 311, but we did not find evidence of a city wide discrepancy with regards to the proportion of Non Hispanic or Latino White residents in a block group.
2. An alternative analysis to overcome this problem addressed about, we focus on the top 3 complaints from the 311 data to obtain some local insights. From the analysis of top 3 complaints (covid, pothole and overgrowth), we find out that Tract 1707.02, 1706.01 and 1708 are the areas with the most concentrated complaints for the chosen categories. After examining the past income data, we discovered that tract 1707.02 has a low income rate compared to other tracts. Tract 1703. 01 and 1701 have the most population but have less dense requests than 1707.02, 1706.01 and 1708. Tract 1703. 01 and 1701. 01 have the top 2 rank the lowest not hispanic or latino population and top 2 on the white population. A suggestion for the planning in the future could be focusing on improving the city services in tract 1707.02 relating to covid, pothole and overpopulation. An insight to offer from examining the demographic data is that among the **top most dense** complaint tracts, majority hispanic tracts have more dense complaints compared to those that are mainly white.
3. Parking tickets not tied to fixed events like street sweeping or parking meters tend to be clustered in neighborhoods bordering the beach or near the MBTA stations. Without data about the drivers, it is unclear what proportion of the tickets in this area are from residents versus out-of-towners.
4. From the analysis of the permits data, we can draw some main conclusions: The City of Revere received most permits relating to Building, Electrical and Plumbing Types of

permits in general. Geographically speaking, Tract 1708 is the area with the most building permit and Tract 1702 in general has the least amount of permits in total in 2020. A suggestion that can be given to the city council based on the analysis are summarized as focusing more support or service on the building infrastructure support, especially in Tract 1708.

5. From the DPW data, we were able to find that Tract 1708 also has the most services among all the tracts. This is true for almost any category of service including potholes, fallen trees, and snow on roads that need to be salted.

### **Limitations and Proposed Solutions**

1. The largest problem with the project is the lack of segmented data on race and ethnicity through the American Community Survey. When the 2020 census data is publicly released, there will be a greater opportunity for specificity and actionable results from the currently curated data.
2. Without local knowledge coming from residents, students are limited in where to look for discrepancies. Obviously complicated by the pandemic, students should start off the semester with a trip to Revere to understand the neighborhood dynamics
3. Thanks to professor Dr. Lorena Estrada-Martínez of UMass Boston, we were able to understand ways in which our analysis could be improved. Unfortunately, this conversation came too late in the project cycle to be actionable. In future projects, meetings with experts should come earlier in the project's timeline and potentially occur more often to ensure that analysis is on the right track.
4. We also had trouble trying to identify blocks as over- or under-reporting without having a ground truth to how many requests should be coming in. In another study, the researchers were able to use satellite data and determine street quality for each census block and then look at how many pothole fixing requests there were in that block. Because the researchers had a ground truth of the true condition of that area, they were able to calculate how many requests were expected and then categorize the data into under- and over-reporting.

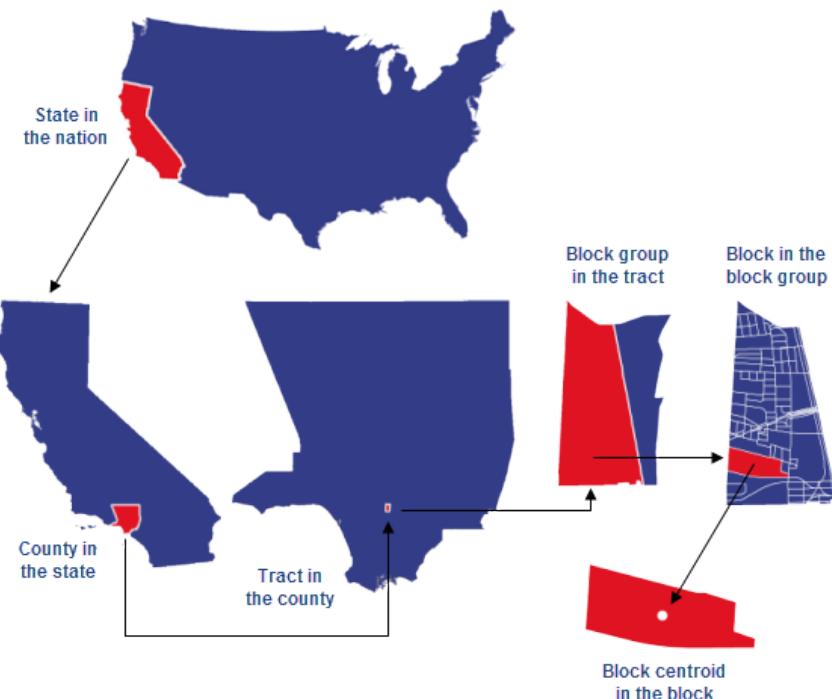
### **Motivation & Thesis**

The purpose of this study was to establish if there were demographic inequalities in the distribution of city services in Revere, Massachusetts. From previous studies on the demographics of Revere, it has become clear that the makeup of the city is rapidly shifting, with a large dichotomy between racial, economic, and age groups within the city's population. Using information about both requests for service and actual services completed, as well as violation data, we hope to gain insight into any patterns of inequity that may be present.

## Initial Studies and Data Set Introduction

In order to correlate the various datasets curated for this report, the data was grouped based on Census tract, with tract being added post-collection based on longitude and latitude if the data set did not already specify Tract information.

Within Suffolk County, Census data is broken up into several divisions. Under each County, there are a number of tracts, which break down further into block groups, and then individual Census blocks (see Figure 1 for visualization). We did not use information from the 2020 Census since it is not currently available for public use, but improvements to the analysis can be made at that time. Instead, we used the 2014-2019 5-year American Community Survey (ACS), which is a survey used to track changes over time in the years between censuses. Most of the ACS data is only available at the block group level.



**Figure 1.** Visualization of regional breakdowns from state through Census block.

<https://learn.arcgis.com/en/related-concepts/united-states-census-geography.htm>

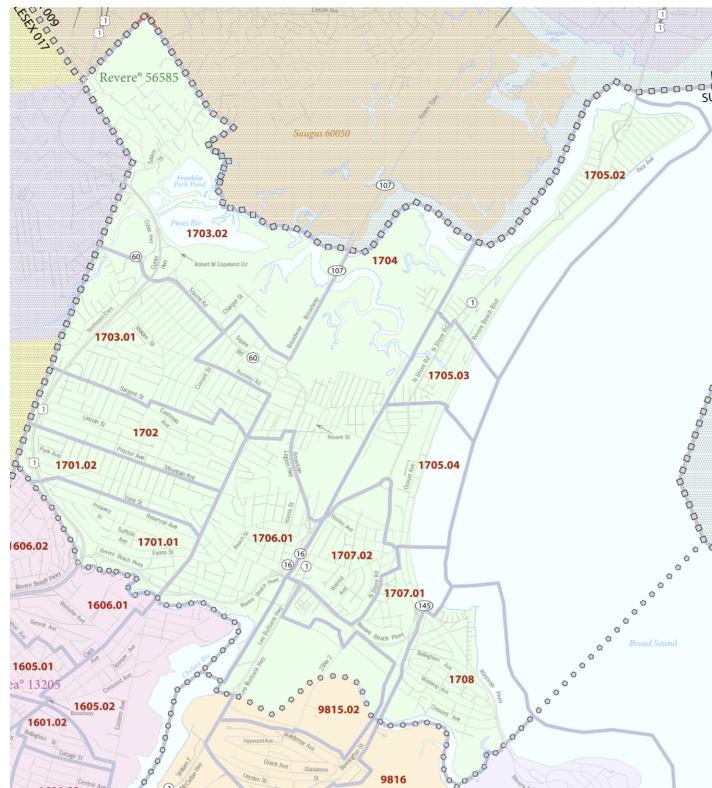
There are **5 datasets** used in this analysis: 4 from the city of Revere (Department of Public Works Jobs, 311 Helpline Requests, Parking Ticket data, and Housing Ticket data), and a dataset from the 5-year ACS relating to table B03002 relating to race and ethnicity. All data has been preprocessed and used in some part of the analysis.

The data from the Census was already parsed down enough to not need to be cleaned any further, but all of the datasets collected from the city of Revere were cleaned to improve the quality of the analysis. For all datasets, we used shape files from the Census to translate the coordinates of

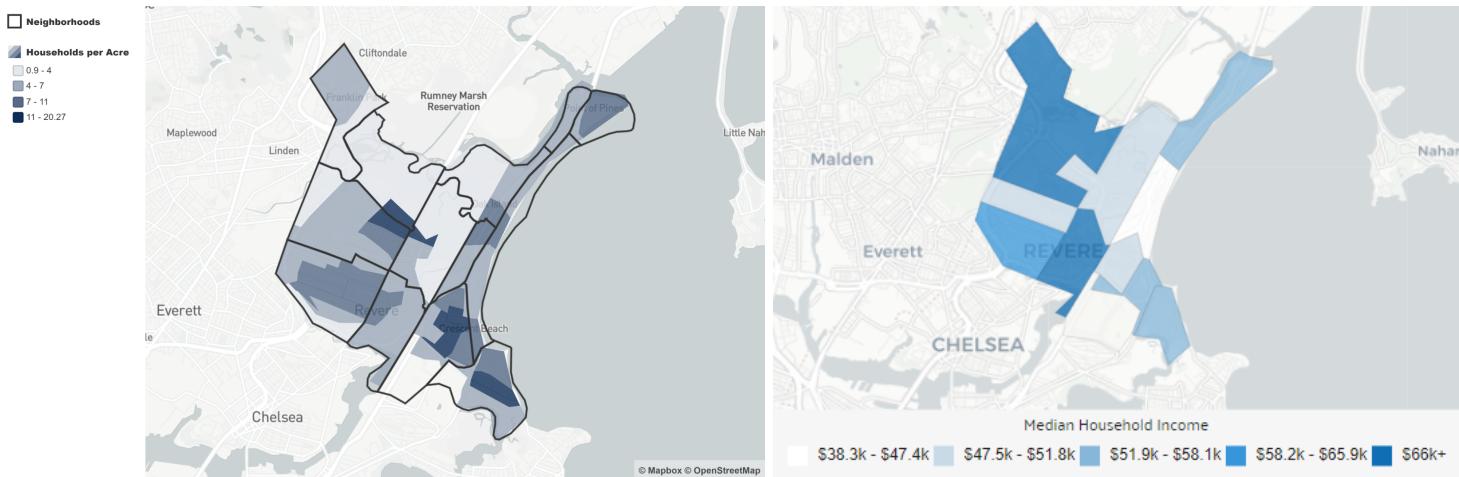
the incident to the corresponding Census block group. We decided to use this level of distinction as it was the smallest level of classification we could use without adding confusion and losing potential significance. For the parking ticket data, we removed parking meter and street sweeping tickets since these gave us fewer insights into potential inequalities and are on fixed routes or locations, and so provide less variable information. For the 311 data, several unnecessary columns were dropped, request type categories were refined, and entries without valid addresses were removed (as they could not be then assigned to a Census tract or block group). For the Department of Public Works, scheduled tasks such as emptying trash cans and cutting grass were removed for similar reasons as removal of parking ticket data.

## Data Analysis

### Visualization



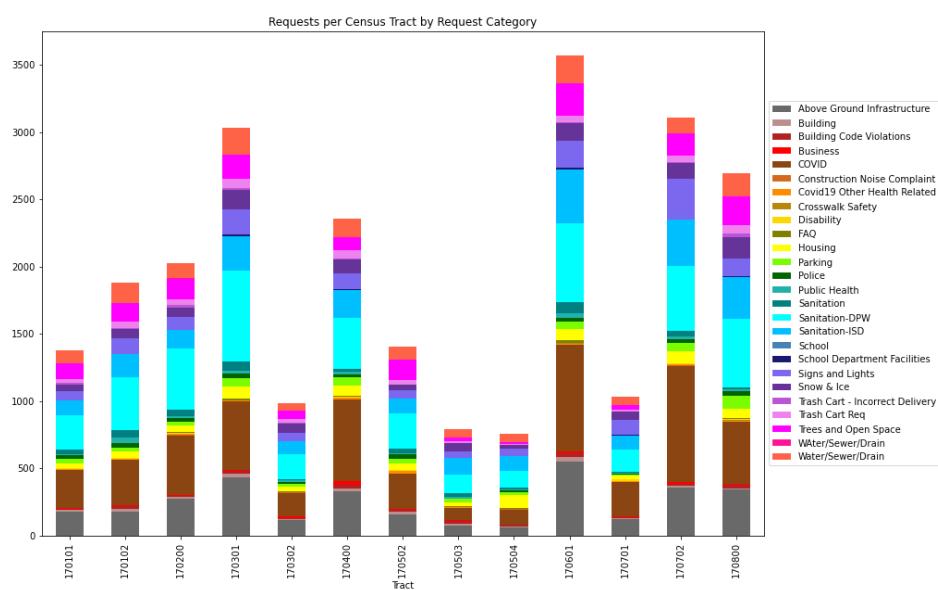
**Figure 2.** Map of census blocks in Revere, MA



**Figure 3.** A) Housing density in Revere, B) Median household income in 2018 by Census Tract

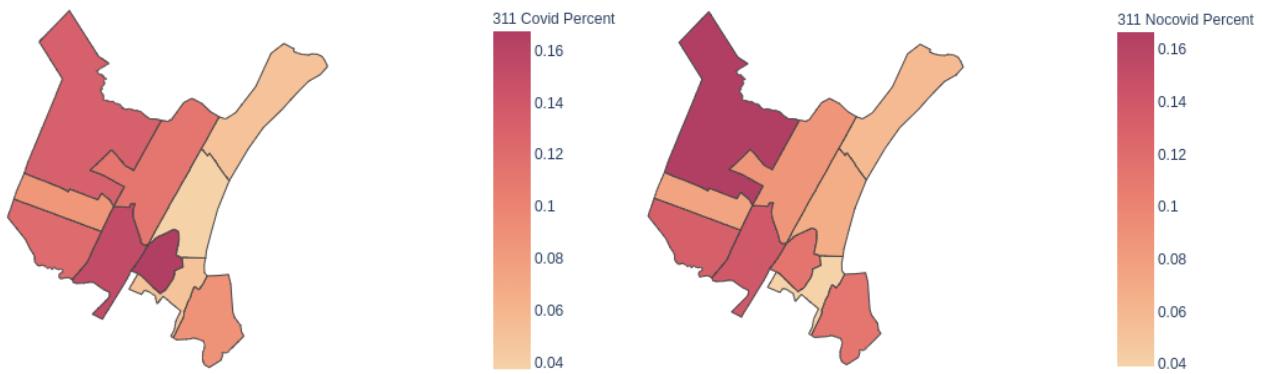
<https://www.revere.org/next-stop-revere/data>, <https://datausa.io/profile/geo/revere-ma>

### 311 Complaint Data



**Figure 4.** Total 311 complaint requests by Census Tract, colored by request category

Figure 4 depicts the overall distribution of 311 request calls placed for each Census tract. The difference in total number of requests between tracts is not a clear indication of bias between tracts due to differences in housing density and total area of the tracts, but is a good indication of what may be important to continue to investigate. Sanitation and COVID-19 related calls were 2 of the most frequent complaint types, as were potholes.

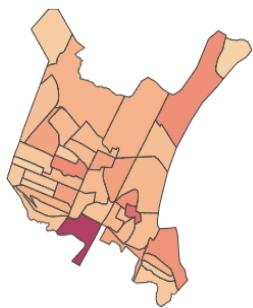


**Figure 5.** Density of 311 complaint calls, separated by COVID-19 and non-COVID-19 related request

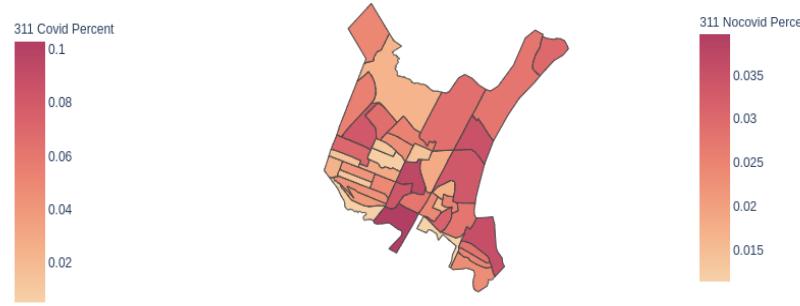
Figure 5 illustrates the density of 311 calls for both COVID-19 and non-COVID-19 related requests. Tract 1702.02 is shown to have the highest density of COVID-19 related complaints compared to the other tracts. This tract also has the highest proportion of Latino residents (at 50.71%) and is one of the lower-income tracts in Revere (median income was less than \$50,000 per year in 2018; datausa.io). The COVID-19 pandemic has had the greatest effect on marginalized communities, both economically and health-wise, so the difference in COVID-19 related requests for this tract is not out of place. Interesting to note is tract 1705.01, directly North of 1702.02, has a much lower density of COVID-19 related requests despite having the lowest median income in 2018 (\$45,000; datausa.io).

In order to get a more accurate look at the different various Tracts we see above, we decided to map the 311 Calls by Census Block Groups.

Percent of Covid related 311 Calls by Census Block Group



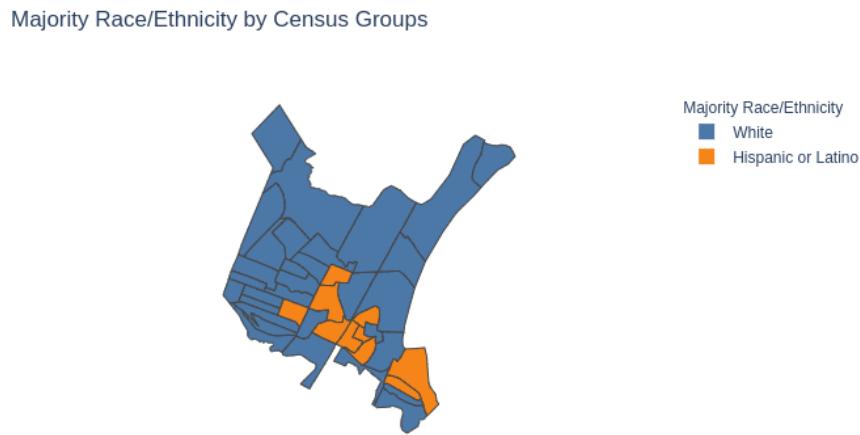
Percent of Non-Covid 311 Calls by Census Block Group



**Figure 6.** Density of 311 Calls by Census Block Group

A few things to note is that the map with the Percent of 311 Calls related to Covid is affected by a Census Tract 1706.01, Block Group 2, whose values do not seem to change regardless of the

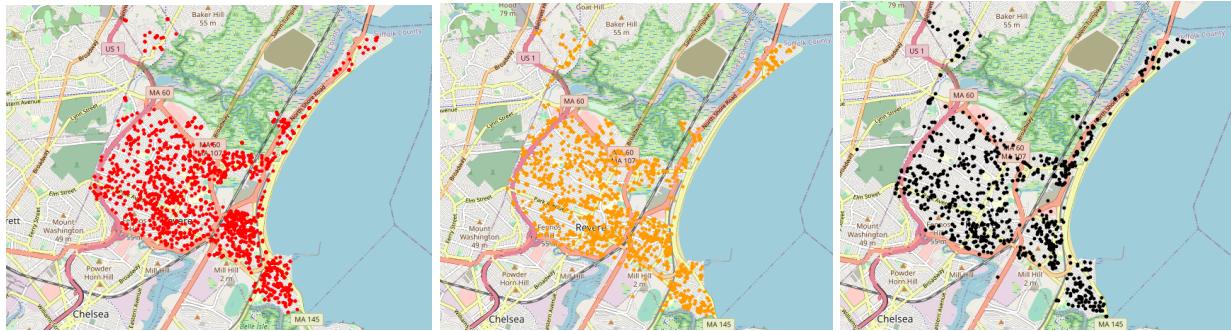
map. In order to get a better look at the Block Groups, we decided to get the Block Groups with the dominant Race/Ethnicity in for its demographics. We can see below that there are about 10 Block Groups that are majority Hispanic/Latino.



**Figure 7.** 10 Block Groups are Majority Hispanic/Latino

Tract	Block Group	Percent Non Covid 311	Percent Covid 311	Differences
Census Tract 1708	Block Group 1	3.58%	4.27%	0.69%
Census Tract 1707.02	Block Group 4	1.73%	4.27%	2.54%
Census Tract 1701	Block Group 7	2.43%	3.54%	1.11%
Census Tract 1704	Block Group 3	1.40%	3.33%	1.93%
Census Tract 1708	Block Group 2	2.84%	2.49%	-0.35%
Census Tract 1707.02	Block Group 3	2.41%	2.15%	-0.26%
Census Tract 1707.02	Block Group 2	3.16%	2.01%	-1.15%
Census Tract 1706.01	Block Group 1	3.73%	1.95%	-1.78%
Census Tract 1707.02	Block Group 1	1.55%	1.61%	0.06%
Census Tract	Block Group	2.78%	1.34%	-1.44%

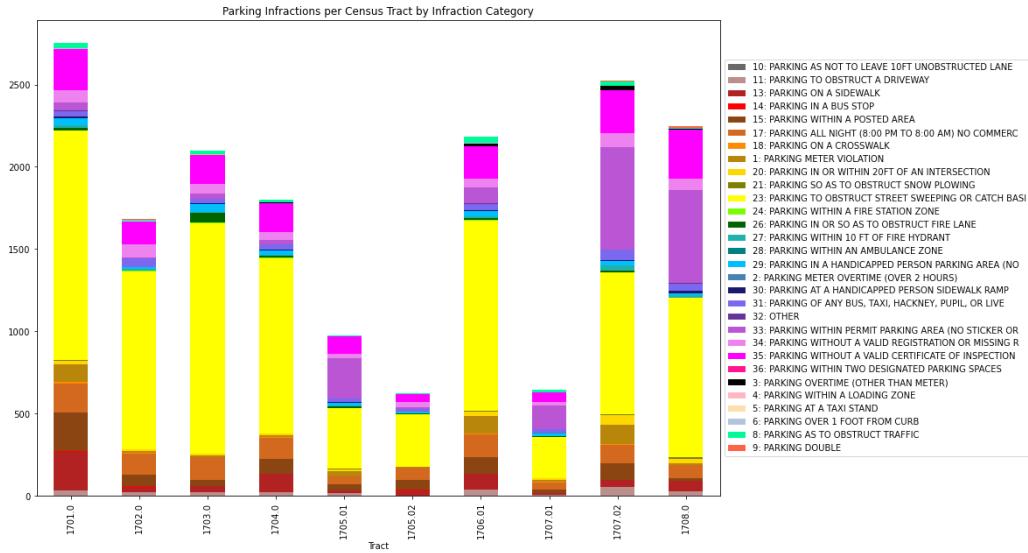
**Figure 8.** Table contains percentages of majority Hispanic/Latino Block Groups and their corresponding densities. Here we see that the difference in densities of the 10 Block Groups that are a majority Hispanic/Latino. Half of the Block Groups have an increase in density of 311 calls. From the map and the table it seems that the density of 311 Calls related to Covid increased from the usual 311 Calls, but the trend was not unique to a certain demographic.



**Figure 9.** Top 3 Type complaint requests by Census Tract (COVID-19, Pothole, Overgrowth)

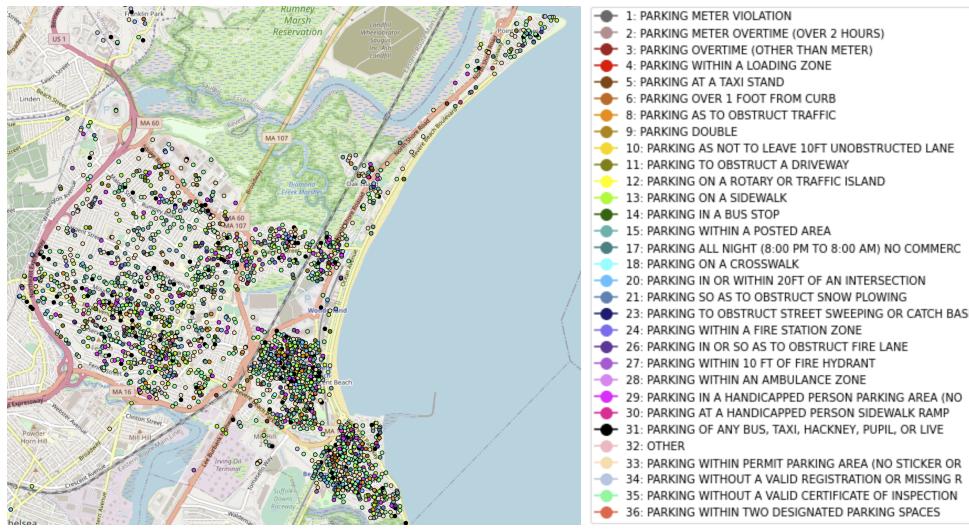
From the analysis of top 3 complaints, we find out that Tract 1707.02, 1706.01 and 1708 are the areas with the most concentrated complaints for the chosen categories. It is possible as well that the majority of the population is located in these areas (since they have the top number of complaints) and the amount of improvements that need to be done should be focusing on these 3 tracts. We find out that there is an unequal distribution of complaint density throughout the city. There are specifically three areas with the most concentrated complaints for the chosen categories. After examining the past income data, we discovered that tract 1707.02 has a low income rate compared to other tracts. Tract 1703. 01 and 1701 have the most population but have less dense requests than 1707.02, 1706.01 and 1708. Tract 1703. 01 and 1701. 01 have the top 2 rank the lowest not hispanic or latino population and top 2 on the white population.

## Parking Data



**Figure 10.** Total parking infractions per Census Tract, colored by infraction code

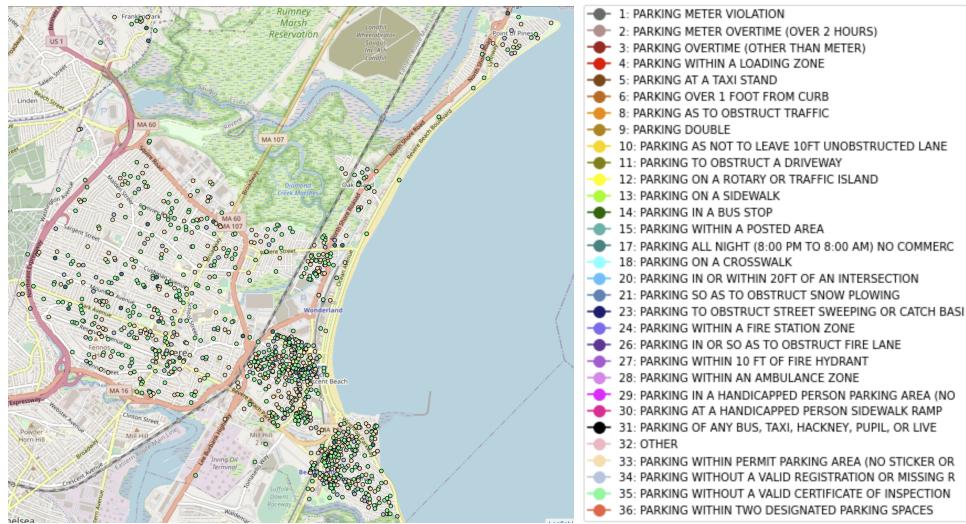
Figure 7 illustrates the number of parking tickets distributed per Census tract, as well as the breakdown of parking violation type. The vast majority of parking infractions are due to obstruction of street sweeping, which for the purposes of our analysis and visualization was removed due to being a “fixed” infraction as it is a scheduled event. Similarly to Figure 4, the overall distribution of parking tickets is not a clear indicator of bias towards any particular tract due to differences in commercial zoning, total area of the tract, and population density.



**Figure 11.** Map of parking fractions not coded parking meter (1, 2) or street sweeping (23)

With removal of street sweeping related parking infractions and parking meters (which are located mainly in areas zoned “general business” (commercial) or “central business” (residential and commercial), Figure 7 illustrates the distribution of overall parking infractions (with the previously mentioned infractions removed). There is an obviously heavy bias towards tracts 1707.02, 1707.01, and 1708 in the general parking infraction data. Density of parking infractions

clustered in beach areas could also be due to tourists and beach-goers, and not necessarily over policing of residents of these neighborhoods, however this would be difficult to determine without registration information for the cars given the ticket.

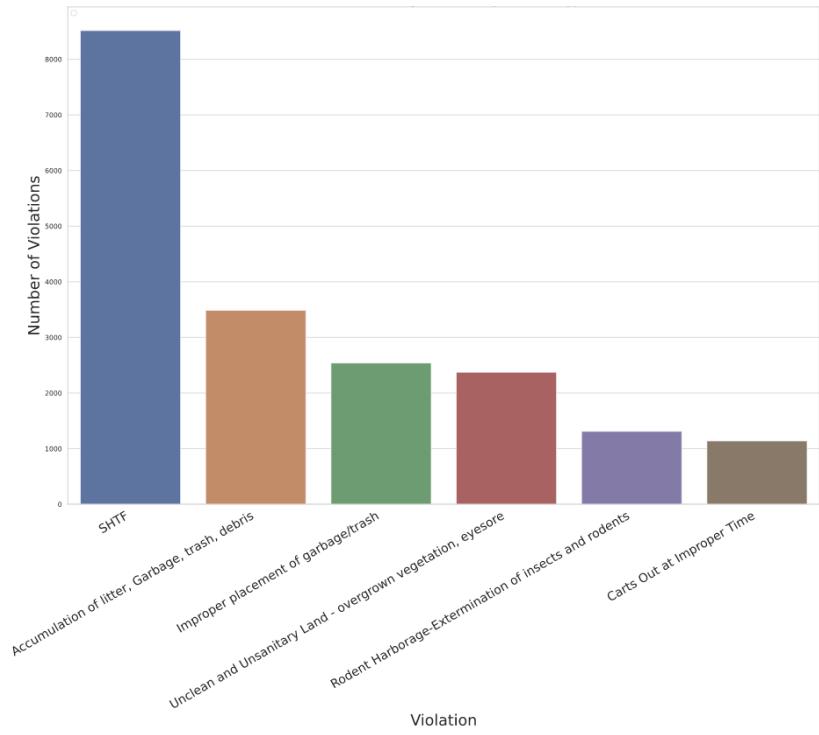


**Figure 12.** Map of parking infractions with codes for parking overnight in non commercial parking (17), parking without permit (33), parking without a valid registration (34), or parking without a valid certificate inspection (35).

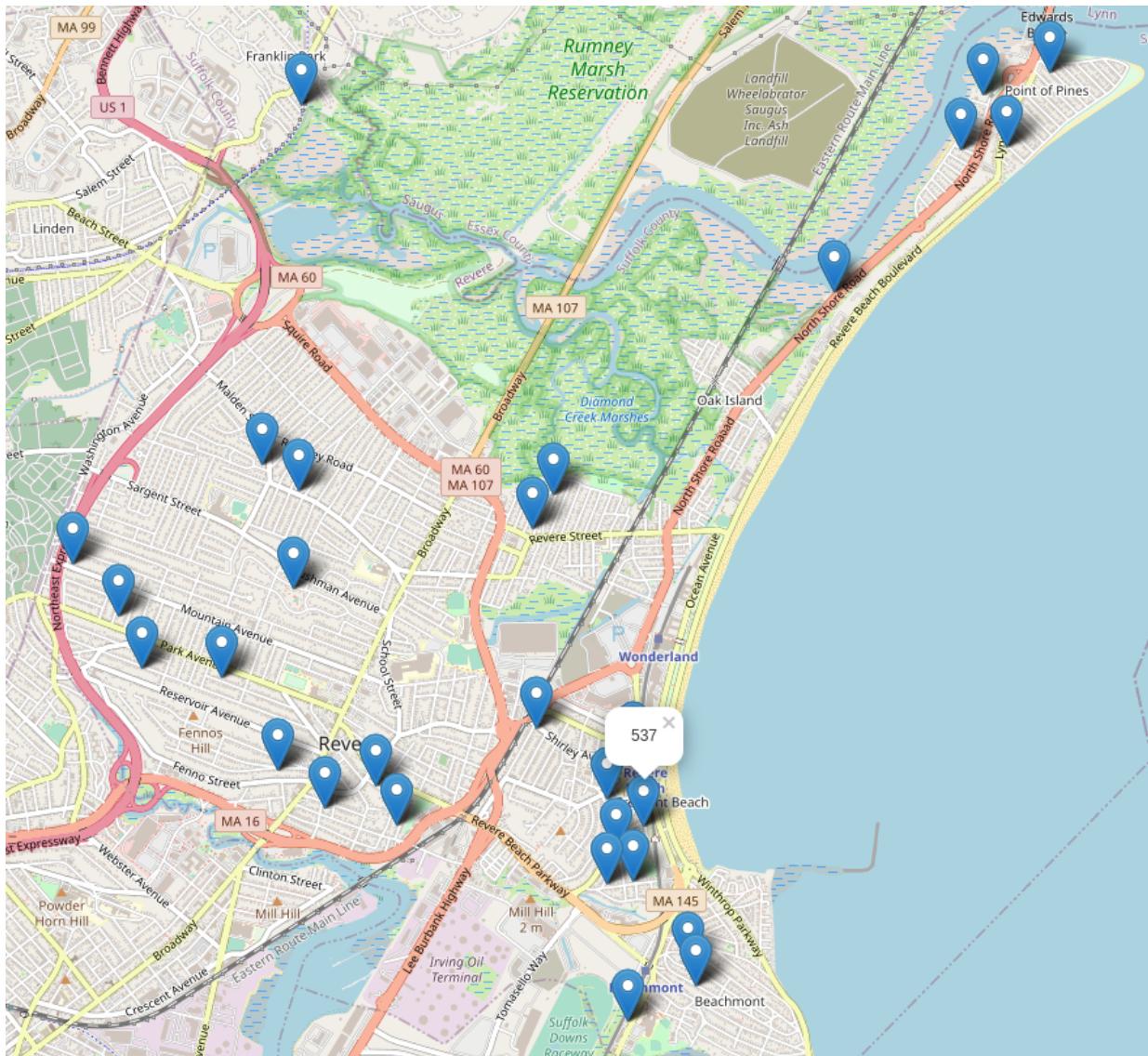
Figure 9 illustrates parking infractions hypothesized to be most likely to be experienced by low income car owners, as permits, registration, inspections, and/or private parking spaces may not be a priority, or even an option, in households with tight budgets. High density areas are also less likely to have parking spaces available to tenants without additional fees, or at all. Census tracts 1707.02, 1707.01, and 1708 appear to have the most parking violations within the designated categories. Permit parking violations may also be due to tourists and beach-goers parking in street parking meant for residents. However, as mentioned previously, this would be difficult to determine without registration information for the cars given the ticket.

### *Housing Violation Data*

For housing we began a preliminary analysis and processing of the data. From this barplot we see that SHTF or Safe Housing Task Force has the most number of violations. This could be due to it potentially being a violation which encompasses many other violations.



**Figure 13.** Number of housing violations by type

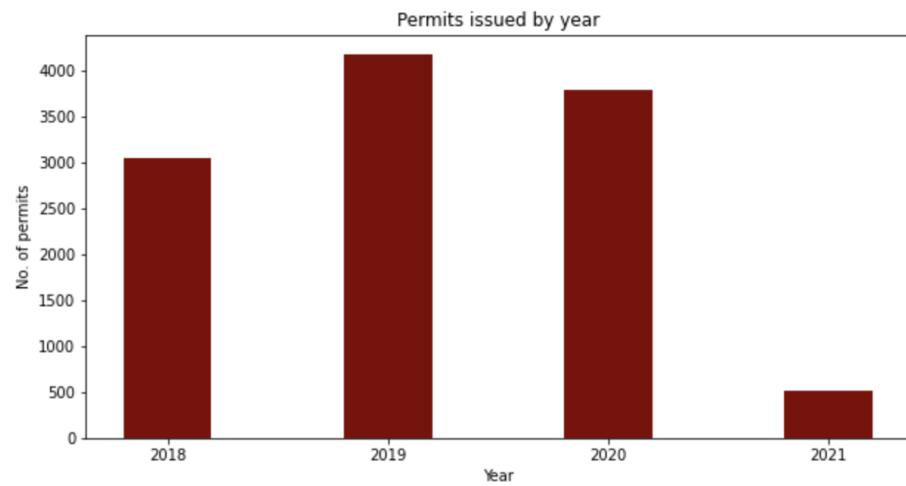


**Figure 14.** Locations of top 30 housing violations

From this map we can see that the Top 30 housing violations are well spread out but the address with the most violations is located near a middle school which could have an influence on the number of violations. That cluster of violations is also near Census Tract/Block groups with a greater population of people who identify as Hispanic/Latino, which is an avenue open for exploration in future work on this project. In addition, future work that can be done is completing the geocoding of all the addresses and finding an efficient way to display the data on a map. It would be interesting to see the top violation by Tract/Block Group, or even some more interesting groupings such as school district or clusters of housing density.

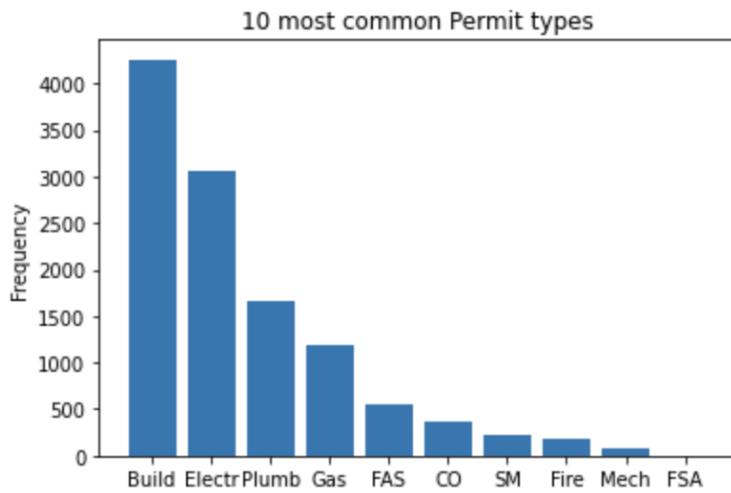
### *Permits Data*

Along with the Housing Violations, another city service we sought to analyze was permits granted for building projects in residential and commercial buildings. Our analysis consisted of showing the most common permit types, counts for each year, and some maps to give a visual representation of where permits were issued. A good starting point for future work to be done is to correlate the demographic data we collected to specific Census Tracts/Block Groups and further analyze the differences in permits issued by year.



**Figure 15.** Number of permits issued from 2018-2021 so far

Figure 15 shows the number of permits issued from 2018-2021 so far. There seems to be a peak number of permits issued in 2019 and a decrease in year 2020.



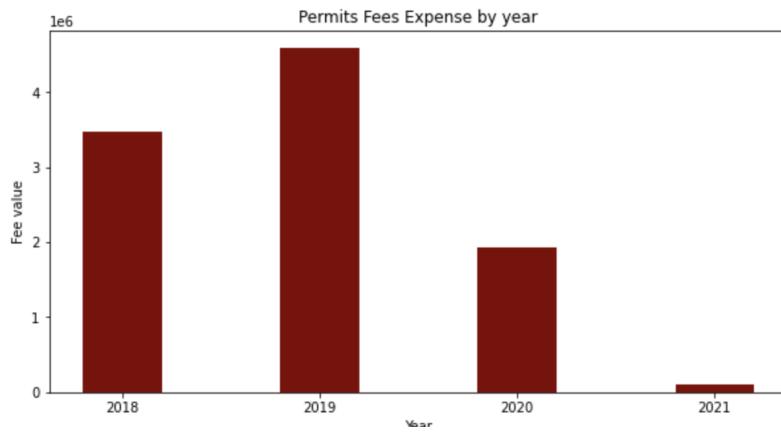
**Figure 16.** Top ten types of permits issued

Figure 16 extracted the top 10 most popular permits types issued through 2018-2021 so far. And the top 3 permits are related to 'Building Permit', 'Electrical Permit' and 'Plumbing Permit'.



**Figure 17.** Geographic distribution of Top 3 permits in year 2020

Figure 17 consists of multiple map plots that shows the density distribution of the top three complaints in 2020. We choose year 2020 to investigate the most recent permit status. Analyzed from these graphs, we can see a relative low density in Tract 1702 for electricity and plumbing permits. Tract 1708 has the most dense permits issued relating to building.

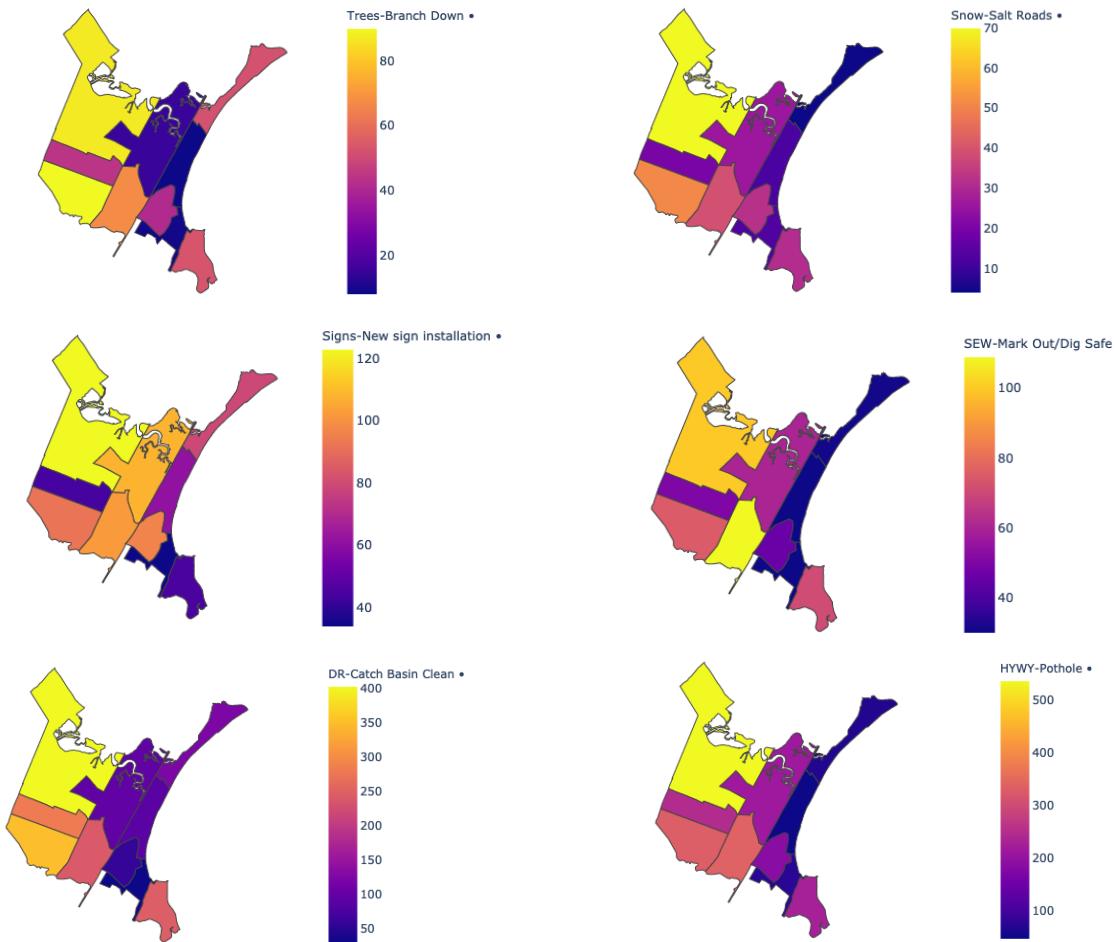


**Figure 18.** Permit Fees by year

Figure 18 shows the amount of money spent on permit fees from 2018-2021. An interesting observation if comparing the bar graph for the number of requests is that in 2020, although the number of permits is more than 2018, the total expense fell compared to 2018.

From the analysis above dealing with expense, types of permits, number of permits yearly and geographic distribution, we can draw the following conclusions: Year 2019 is the highest number of permits dealt with the most fees. The City of Revere received most permits relating to Building, Electrical and Plumbing Types of permits in general. In 2020, although the amount of permits dealt increased than 2019, the permit fees severely decreased. Geographically speaking, Tract 1708 is the area with the most building permit and Tract 1702 in general has the least amount of permits in total in 2020.

## Department of Public Works Data



**Figure 19.** Choropleth maps of various DPW Services Completed. For a web visualization, it is highly recommended to visit: <https://xav.ie/choropleths/>. The counts are the number of services completed in a given tract. For example, the top left tract of “HWY-Pothole” was completed 500 times.

## Statistical Analysis

Methodology: All requests were tied to their respective Census Block group, which is how the observed number of Request types was tabulated. The Expected value is calculated by multiplying the total number of requests by the proportion of the population that live in a given block group. A group was tagged as over reporting if the observed number of requests is greater than the expected number of requests, and vice versa for the under reporting block groups. The 55 percent of the population being non-hispanic white since it divided the block groups nicely in half.

A note on simpson's paradox: When all the 311 requests are looked at together, there is the potential that the subgroups have opposite effects, skewing the result. Thus, what follows is a chi

square test for goodness of fit first on all 311 requests, then on the data broken up into COVID related requests and non COVID related requests.

Each entry in the table includes the observed value, the expected value and then the chi-squared residual.

For the following Pearson Chi-Squared goodness of fit tests, the hypotheses are as follows.

H0: The number of under reporting and over reporting is the same for mostly non-hispanic white block groups and not mostly non-hispanic white block groups.

H1: H0 is False

**All Data**

<u>A</u>	Under	Over	Total
<.55% white	10 8.5 0.2647	11 12.5 0.18	21
>.55% white	7 8.5 0.2647	14 12.5 0.18	21
Total	17	25	42

Chi-Squared Value : 0.889, df:1, p-value = 0.3456

**COVID-19 Assistance Data**

<u>B</u>	Under	Over	Total
<.55% white	16 14.5 0.155	5 6.5 0.346	21
>.55% white	13 14.5 0.155	8 6.5 0.346	21
Total	29	13	42

Chi-Squared Value : 1.002, df:1, p-value = 0.3167

**Non-COVIDp19 Related Data**

<u>C</u>	Under	Over	Total

< .55% white	10 9 0.111	11 12 0.083	21
>.55% white	8 9 0.111	13 12 0.083	21
Total	18	24	42

Chi-Squared Value : 0.38888, df:1, p-value =0.5329

**Figure 20.** A) Analysis of over and under estimation of all data, B) Analysis of over and under estimation of COVID-19 assistance data, C) Analysis of over and under estimation of non-COVID-19 related data

In each of the three datasets, there is no statistical significance as the p-value is larger than our alpha value of 0.05, so we fail to reject the null hypothesis at the 95% significance level.

## Limitation and Future Directions

The greatest limitation of this research is that the implications about race can only go down to the block group level, due to limitations with ACS data. Additionally median income is not listed in small enough segments for analysis on it to be useful in the American Community Survey. When the 2020 census becomes available, more detailed race and ethnicity data will be available, and there will be insights into racial disparity by block instead of block group.

Due to the ongoing pandemic, limited resources, and lack of information about Revere, there is only so much that the team could analyze with regards to implications available only to residents and administrators of Revere. For furniture investigations, the semester should start with a tour of the various neighborhoods of Revere so students can get their bearings.

The block groups do not necessarily align along neighborhood boundaries, so by using administrative boundaries the insights we are able to make with regards to community are limited. This is something that is harder to fix, and is implicit with using Census Data. If future analysis was to be done on neighborhoods, census blocks could be grouped together to better reflect true neighborhood boundaries.

Through conversations with Dr. Lorena Estrada-Martínez of UMass Boston, rather than selecting an arbitrary divide for feature analysis, we would have gotten better results from using latent class analysis. Unfortunately, there are currently no Python modules with LCA support, so that would require becoming comfortable with either R or other statistical software. For future

semesters, a class or two on R would give students more ability to be confident with quantitative methods.

## Conclusion

Much of the exploration into the relationship between race/ethnicity and city resources is warped by the lack of ability to segment the data into smaller parcels, like those that will be available when the 2020 Census data is released. The use of 311 request metrics as a proxy for access to city resources is imperfect; contacting 311 is not the same as getting the problem fixed, and it is difficult to look into response times in different neighborhoods. Those limitations put aside, this investigation, while not able to point out any glaring disparities, is part of an ongoing conversation in the City of Revere and cities across the US looking to use data to improve city resources. A general recommendation the team can make is to inform more citizens in underprivileged neighborhoods about making 311 requests.

Parking tickets also may not be a good indicator of inadequacies or bias. The main clusters of tickets are in tracts with lower median incomes and are proportionally less-white, but they are also located closer to the beaches and touristy areas. Are these residents truly facing over-policing with regards to parking tickets, or are there more tickets due to tourists and commercial activity? Without more data on who received the tickets, it is very difficult to draw conclusions regarding any inequities.

Even though our data has not accurately determined an overall problem, talking to residents of Revere reveals a different picture than doing analysis by administrative block. Neighborhoods in real cities are not defined by the lines drawn in paper ones. Census tracts, voting blocks, and other administrative labels do not accurately reflect the makeup of neighborhoods within cities, and may even be causing issues in our analysis of Revere's city services. Communities of various groups may be split between administrative regions, giving the illusion of better application of city services or a lower density of said group when lumped in with bits of other neighborhoods.

## **Additional Citations and References**

<https://www.revere.org/next-stop-revere/data>

<https://datausa.io/profile/geo/revere-ma>

<https://datausa.io/profile/geo/revere-ma/#:~:text=Median%20household%20income%20in%20Revere%2C%20MA%20is%20%2455%2C020>

## **Code:**

<https://github.com/BU-Spark/CS506Spring2021Repository/tree/master/City%20of%20Revere%20Service%20Data%20Project>