Deliverable 3
City of Revere

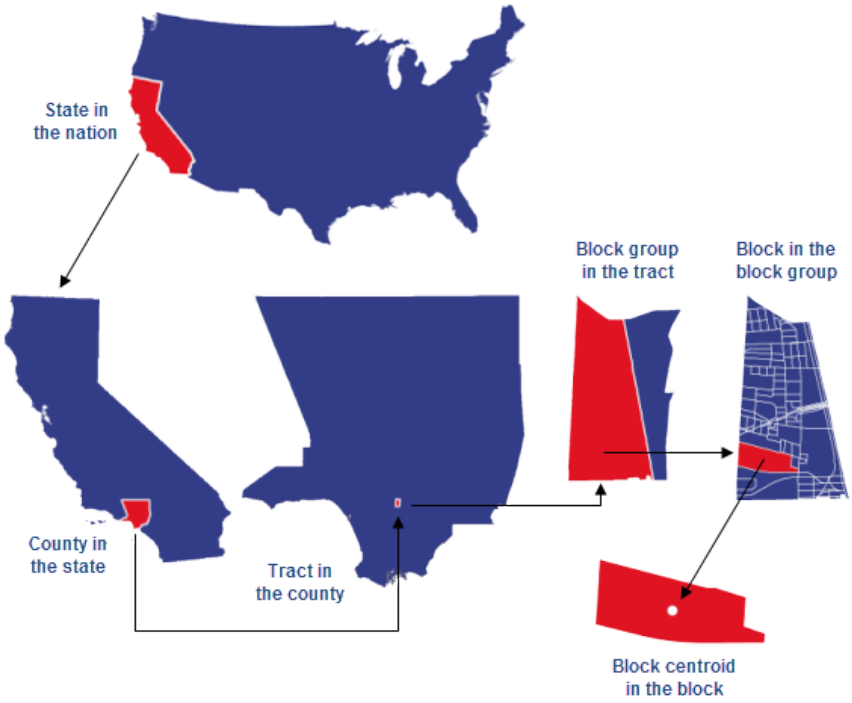
Group Members
(Name/email/github)
Client: Reuben Kantor rkantor@revere.org,
Spark Liaison: Greta Bruce gretab@bu.edu,
PM: Lingyang Jiang lingyanj@bu.edu,
Yuanming Chai jchai23@bu.edu **ychai23**
Mackenzie Knox mfknox@bu.edu **kenzieknox**,
Quinn Relyea qrelyea@bu.edu **Relishyeah**,
Giancarlo Sirio gsirio@bu.edu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

Key Question

There is no statistically significant difference between under reporting and over reporting of 311 issues by census block group with respect to a district having a “mostly non-hispanic white” (> 55%) or “mostly not non-hispanic white” residents. This holds for both COVID aid specific Requests and non COVID requests.

Background on Census Data Types

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. 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.



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

Data Collection and Cleaning

There are 6 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 two datasets from the 5-yearACS relating to tables B03002 relating to race and ethnicity, and table TABLE NAME relating to median income. All data has been preprocessed and used in some part of the analysis.

The Data from the census was already pared down enough to not need to be cleaned any further, but all of the datasets 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 that metric because it was the most segmented way we could analyze city services. For the parking ticket data, we removed parking meter and street sweeping tickets since these gave us fewer insights into potential inequalities. For the housing ticket data, the group member that was assigned the work did not do their part. For the 311 data several unnecessary columns were dropped, we refined the number of request types, and removed entries without a valid address. For the Department of Public Works, scheduled tasks such as emptying trash cans and cutting grass were removed.

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

	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 Assistance Data

	Under	Over	Total
< .55% white	16 14.5 0.155	5 6.5 0.346	21
>.55%	13	8 6.5	21

white	14.5 0.155	0.346	
Total	29	13	42

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

Non Covid Related Data

	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

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.

Visualization

1. **Parking Ticket Data:**
   First up is the parking ticket data, which will need to be refined and made into a more easily consumable visualization.

Figure 1. Color key for all parking infractions

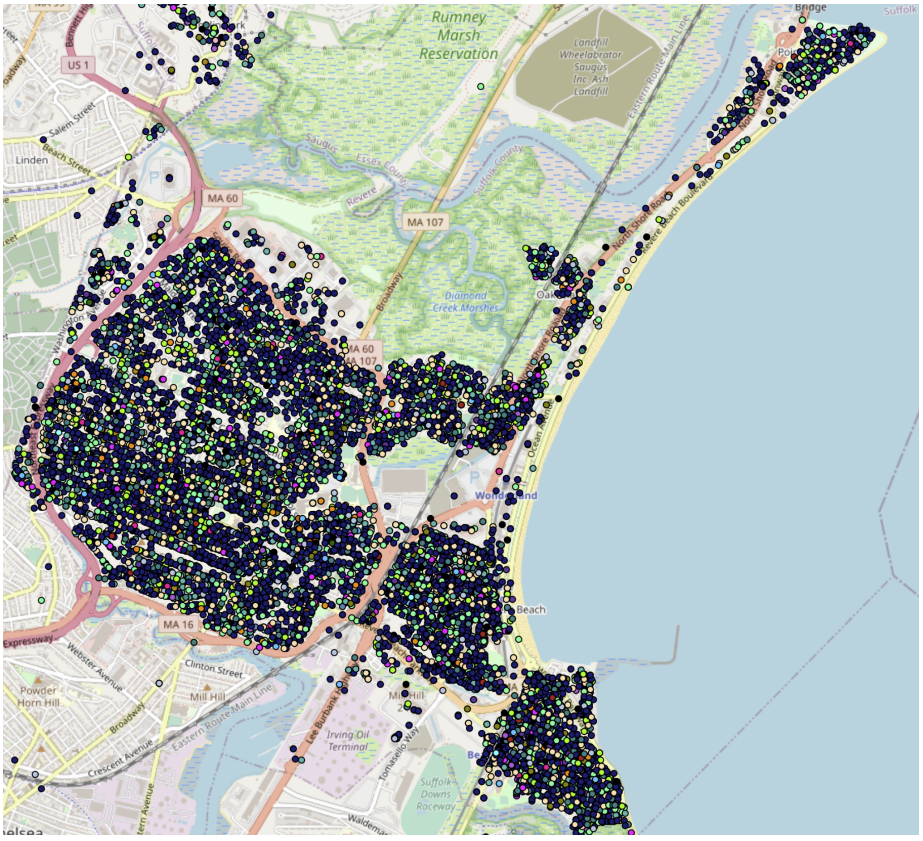


Figure 2. Map of parking infractions not coded parking meter (1, 2)

Code	Violation	Count
23	Parking to obstruct street sweeping	9015
33	Parking within permit parking (no sticker)	1846
35	Parking without valid certificate of inspection	1742
17	Parking all night (8PM to 8AM) in non commercial parking	1102
15	Parking within a posted area	794
12	Parking on a rotary or traffic island	682
34	Parking without valid registration or missing registration	558
1	Parking meter violation	463

Table 1. Top 8 parking infractions

There is no obvious pattern to the parking data in Figure 2, most likely due to the vast majority of parking infractions being coded as obstructing street sweeping (Table 1).

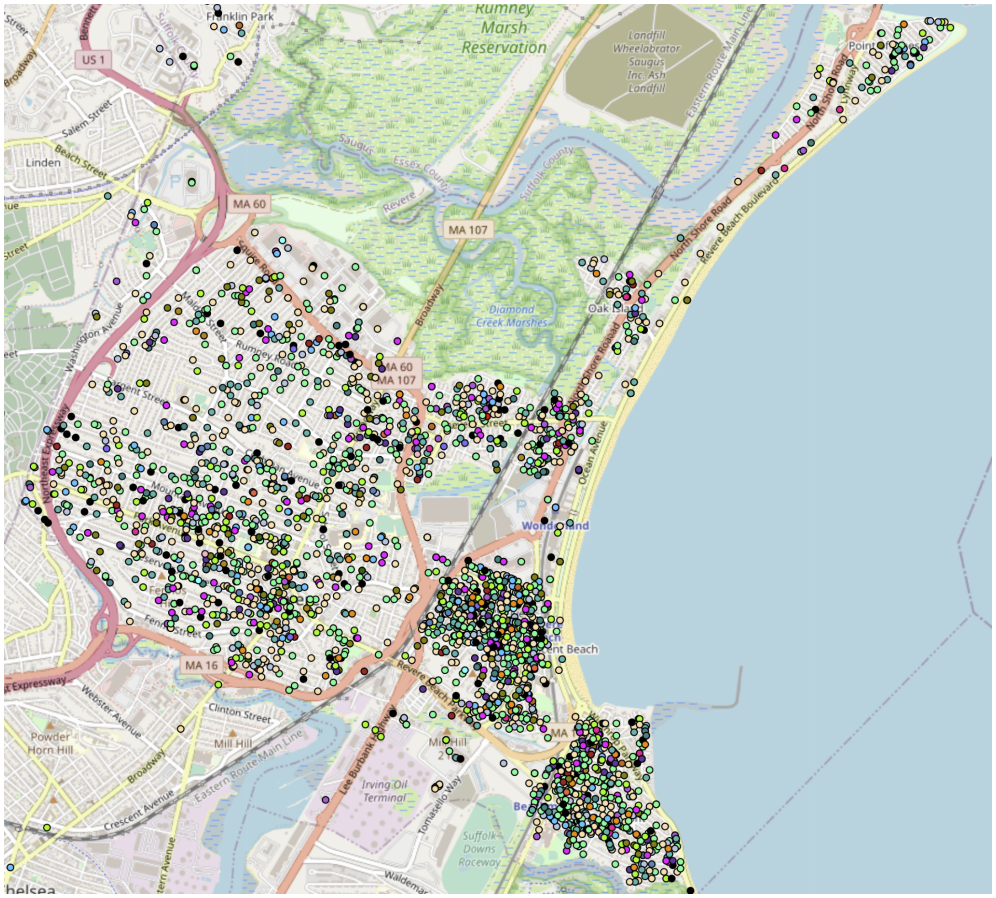


Figure 3. Map of parking fractions not coded parking meter (1, 2) or street sweeping (23)

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 3 illustrates the distribution of overall parking infractions. There is an obviously heavy bias towards tracts 1707.02, 1707.01, and 1708 in the general parking infraction data.

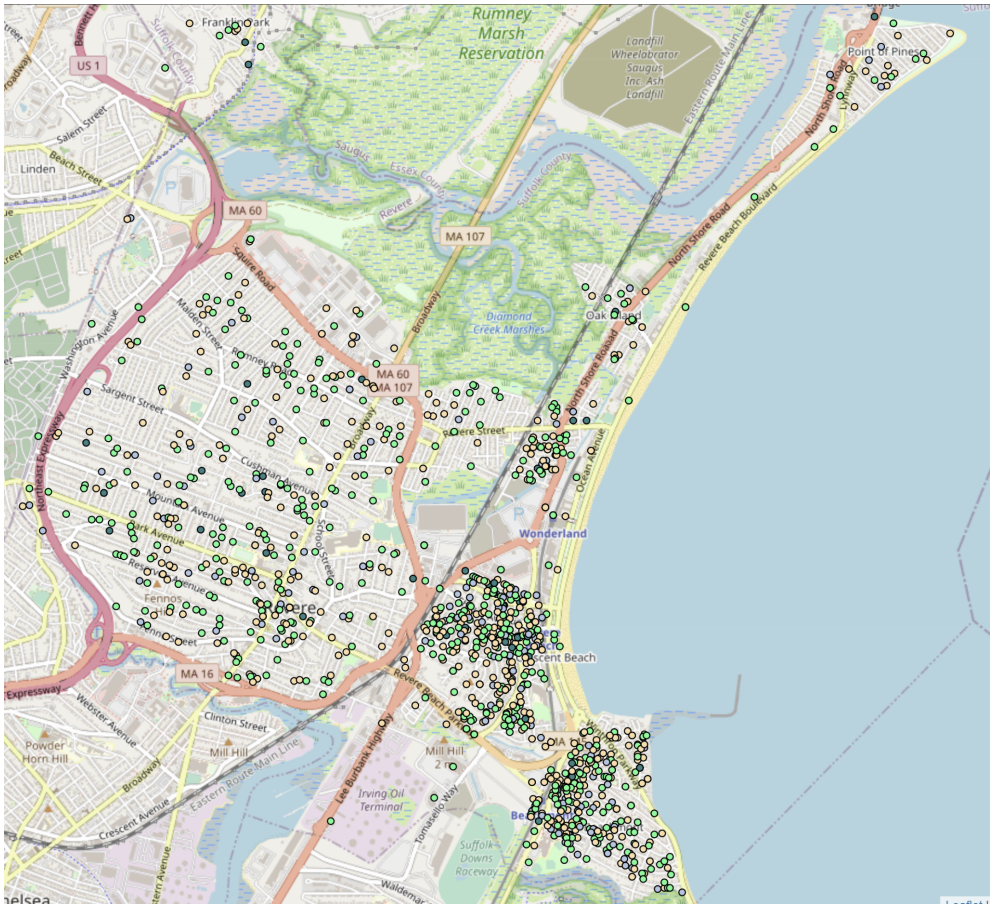


Figure 4. Map of parking infractions with codes for parking overnight in non commercial parking (17), parking without permit (33), parking without a valid registration (34), or parking without a valid certificate inspection (35).

Figure 4 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 (understandably) in households with tight budgets. High density areas are 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.

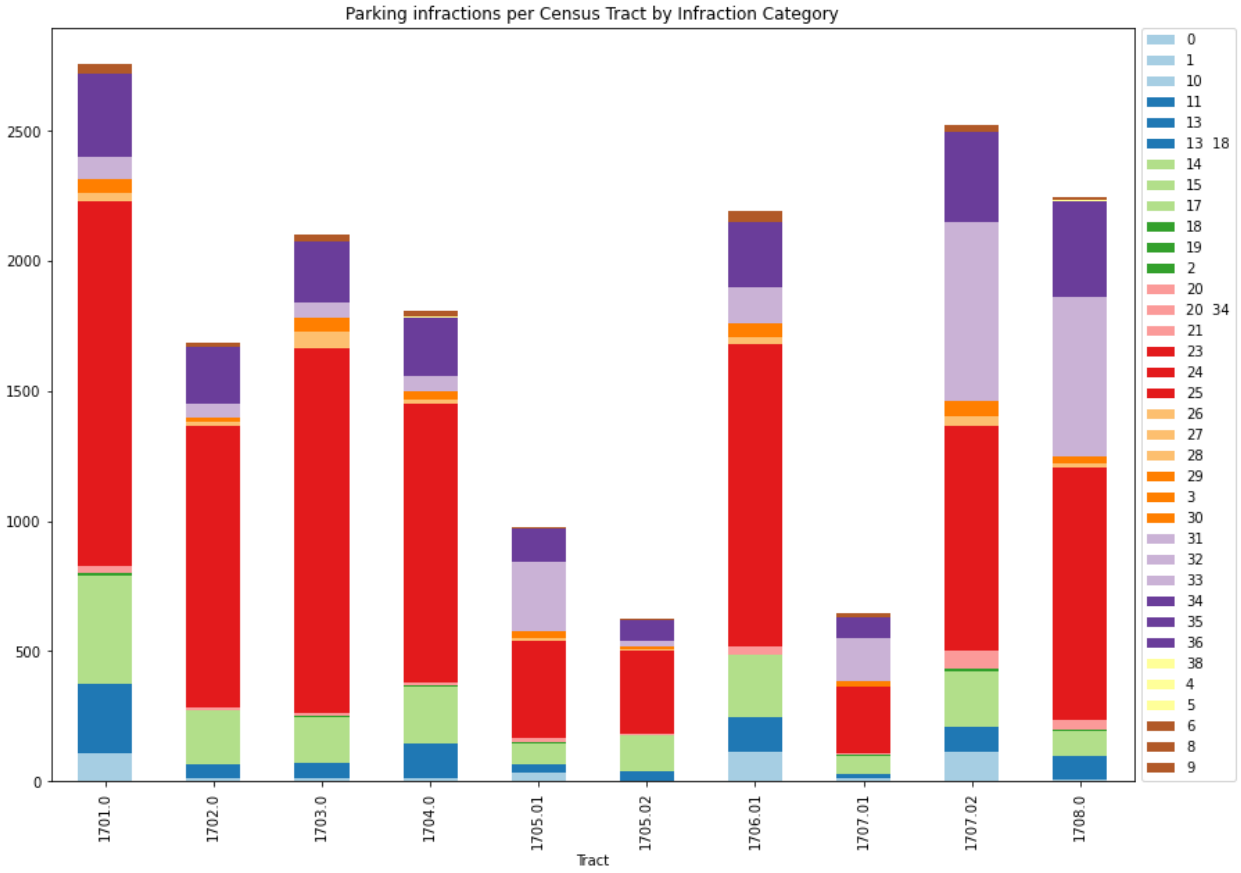
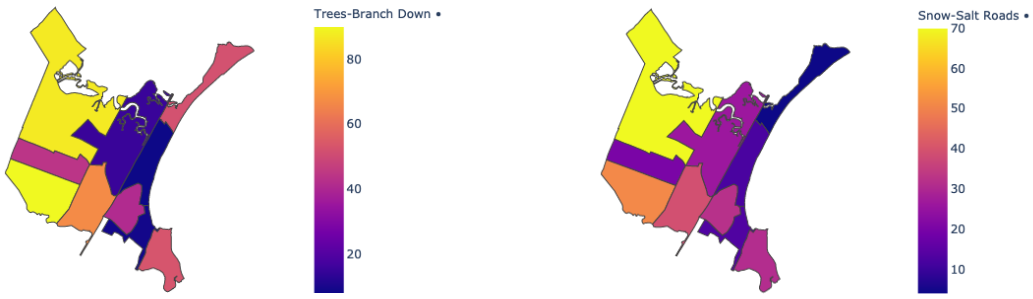


Figure 5. Stacked bar plot of parking infractions by tract, colored by infraction



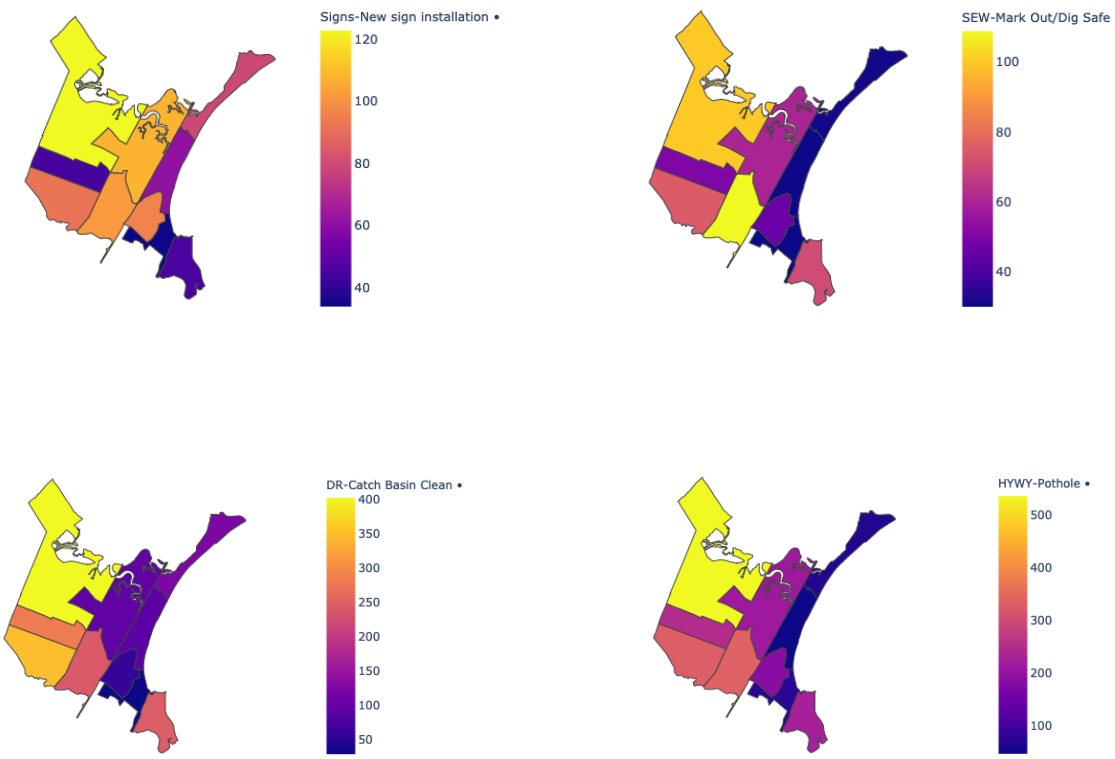


Figure 6. Choropleth maps of various DPW Services Completed. For a web visualization, it is highly recommended to visit: https://xav.ie/choropleths/. The counts are the number of services completed in a given tract. For example, the top left tract of "HWY-Pothole" was completed 500 times.

2. **Permit data**
   Next is the Permit data, which will need to be refined and made into a more easily consumable visualization.

   Figure 1: Number of permits issued from 2018-2021 so far

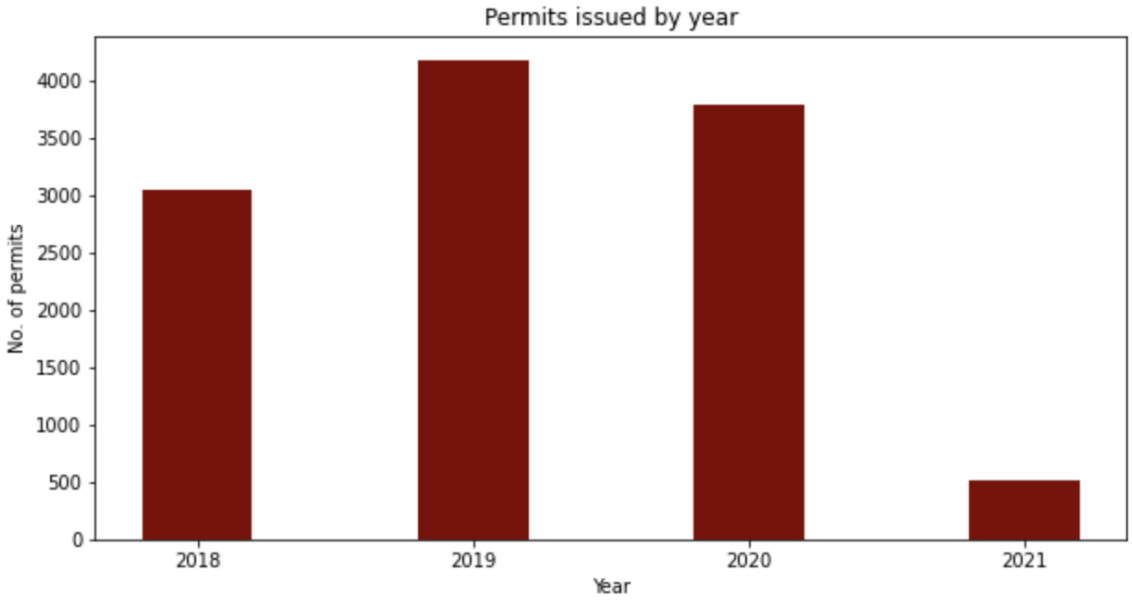


Figure1 shows the number of permits issued from 2018-2021 so far. There seems to be a peak number of permits issued in 2019 and a decrease in year 2020.

Figure 2: Top ten types of permits issued

['Building Permit', 'Electrical Permit', 'Plumbing Permit', 'Gas Permit', 'Fire Alarm System Permit', 'Certificate of Occupancy', 'Sheet Metal Permit', 'Fire Permit', 'Mechanical Permit', 'Fire System Alarm Permit']

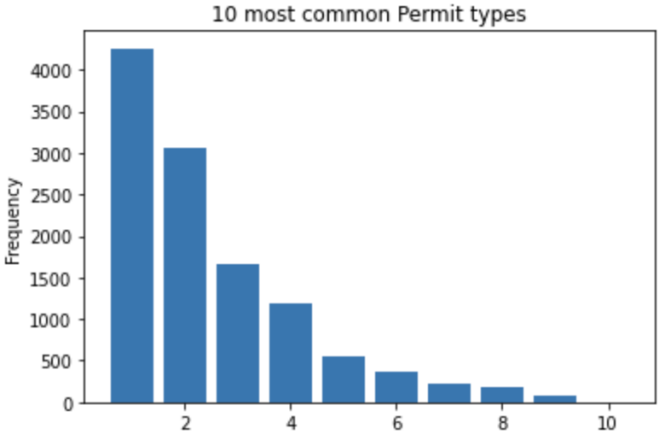
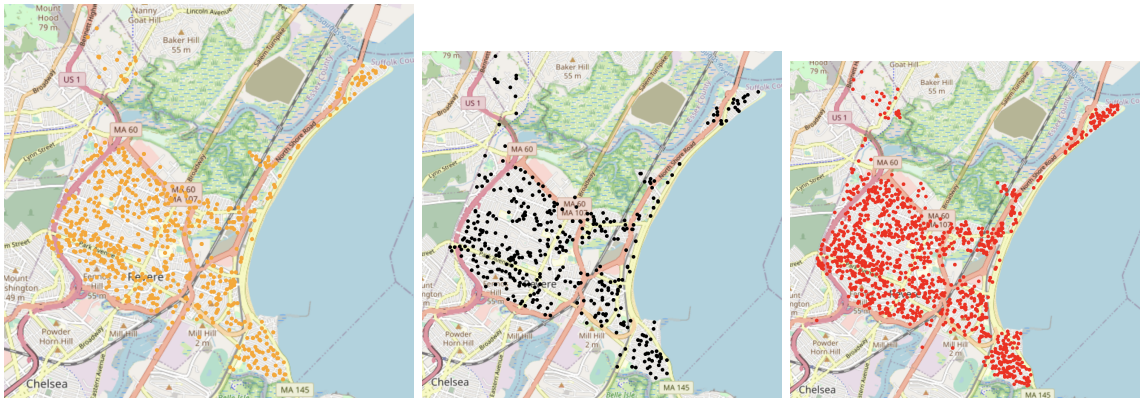


Figure 2 extracted the top 10 most popular permits types issued through 2018-2021 so far. And the top 3 permits are related to 'Building Permit', 'Electrical Permit' and 'Plumbing Permit'.

Figure 3: Geographic distribution of Top 3 permits in year 2020



Electricity Permits Plumbing Permits Building Permits

Figure 3 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 4: Permit Fees by year

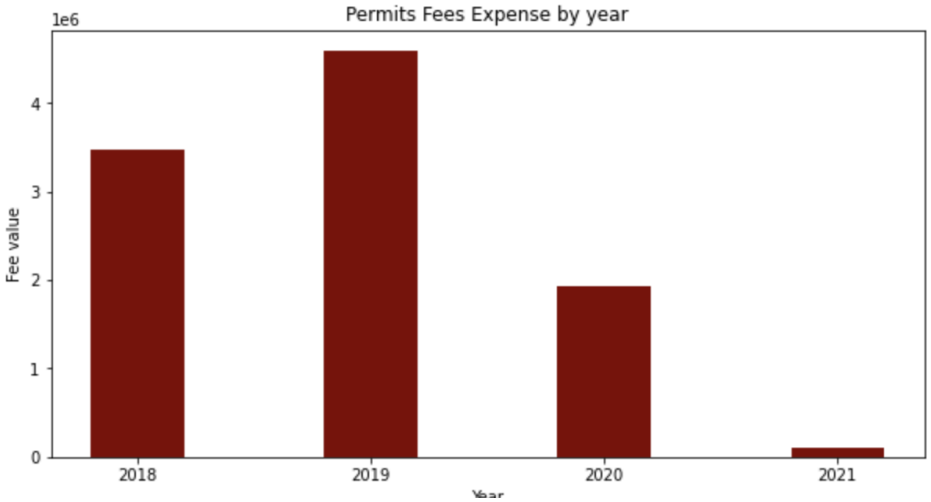


Figure 4 shows the amount of money spent on permit fees from 2018-2021. An interesting observation if comparing the bar graph for the number of requests is that in 2020, although the number of permits is more than 2018, the total expense fell compared to 2018.

<u>Conclusion from Permit Data:</u>

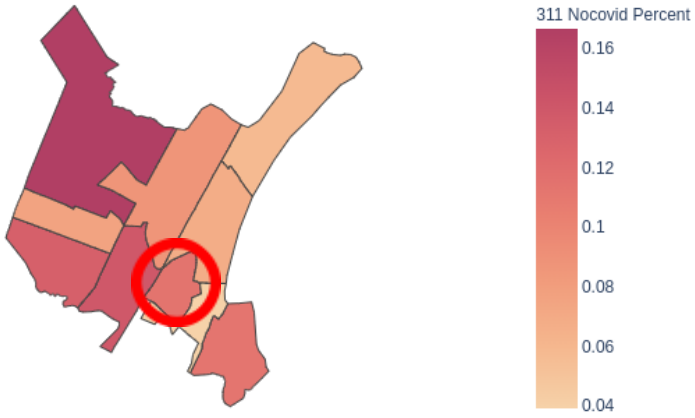
From the analysis above dealing with expense, types of permits, number of permits yearly and geographic distribution, we can draw a few conclusions:

- 2019 is the highest number of permits dealt with the most fees
- 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.
- 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.

3. 311 Data

Figure 1



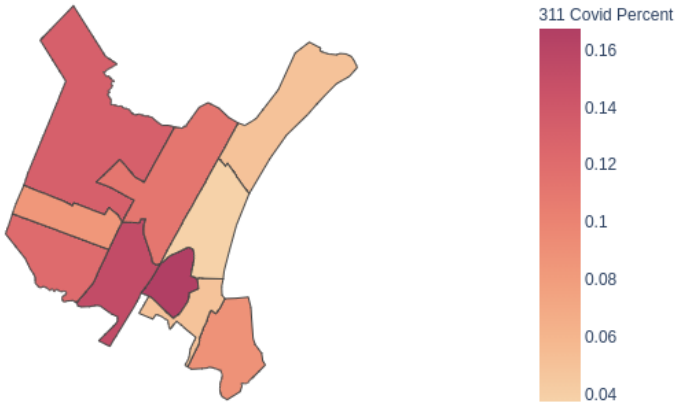


Figure 2

Figure 1&2: Here we see the density of 311 calls for Covid and Non Covid related requests.

One insight to gain here is the density of 311 Calls for the circled tract, 1707.02, which has one of the largest Hispanic and Latino populations at 50.71%. The Percent of all the covid related calls is much higher in this tract when compared to all types of requests. We know that the pandemic has affected marginalized communities in greater numbers, and while we cannot directly prove this is the fact here in Revere since calls are not mapped to a certain demographic group ,there is a notable correlation in the numbers.

4. Housing Violations 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. The next two categories revolve around trash, which could potentially be grouped together based on the client's discretion.

Figure 1.

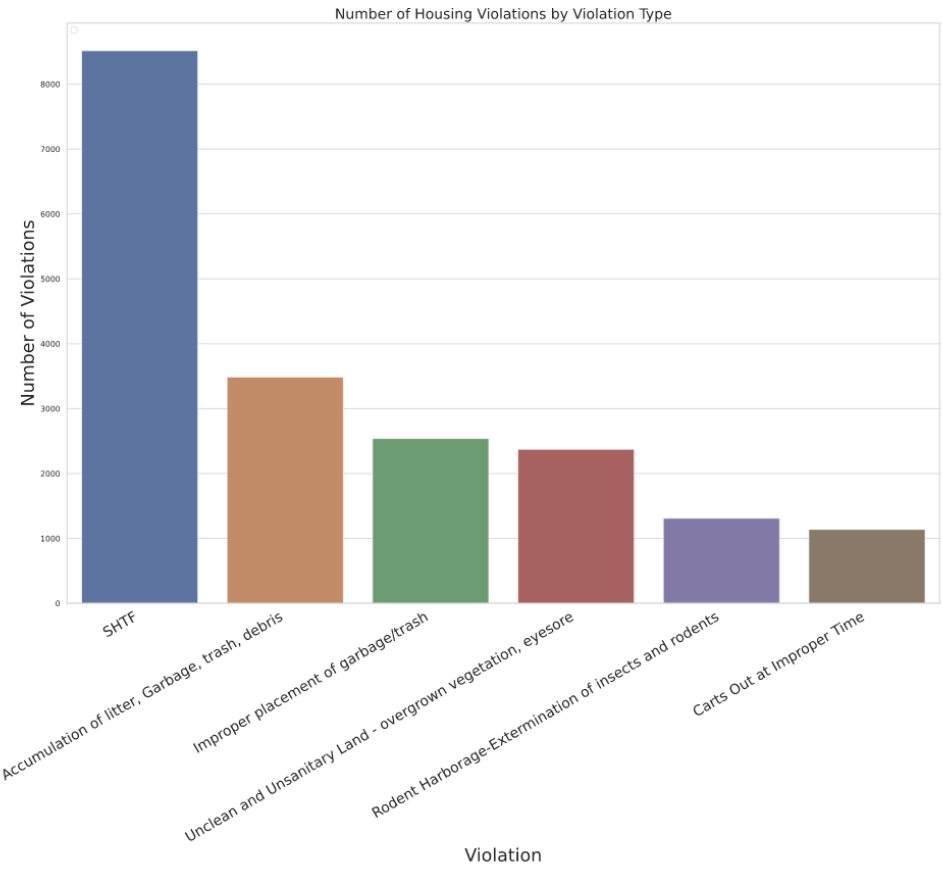
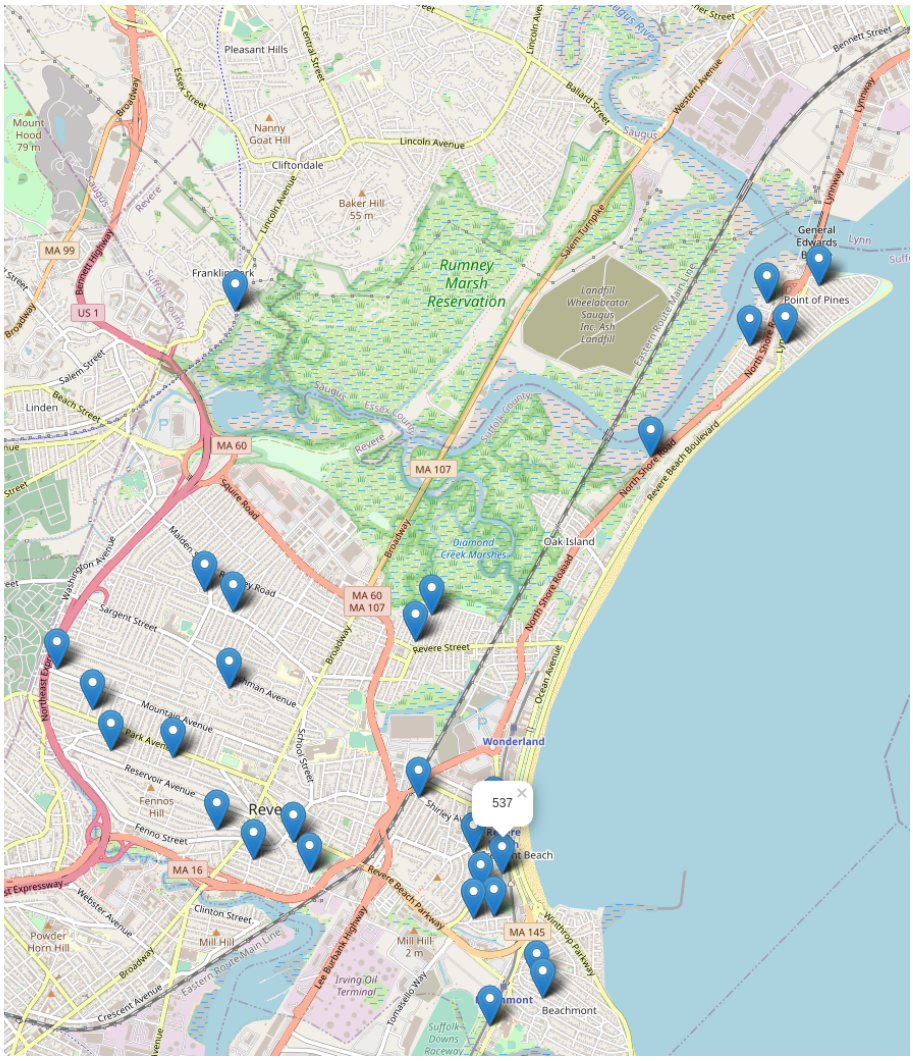


Figure 2.

Locations of Top 30 Violations



From this map we can see that the Top 30 housing violations are pretty spread but the address with the most violations is located near a middle school. The fact that there is a school district near that address might be a possible explanation but we can confirm with the client at a further data. In the next steps we look to geocode the rest of the addresses of each housing violation, map them to their corresponding tract and block group, and create some visualizations with relation to time as specified by the client.

Project Scope

Correlate demographic location data from 2018 onward from the census with city services including parking data, DPW data, and 311 Request Data.

Limitations

Median income is not listed in small enough segments for analysis on it to be useful in the American Community Survey.

Due to incomplete or inaccurate location tagging for some data points, our model might be less effective. The lack of tagged race and ethnicity data on 311 requests makes it difficult to make clear assumptions about the way race and ethnicity affects request behavior. Within districts, we are unaware of how individuals are behaving.

Potential Risks

Some of the statistics used in our reports are calculated with code. While the team puts their best effort in debugging and making sure that numbers being displayed are accurate, potential bugs in code could lead to different inaccurate numbers being reported. In addition, the demographic data being used is from the American Community Survey which provides estimates with margins of errors, and not exact figures.

Draft of Final Report

The most significant obstacle in this project is the fact that the data of city services/violations/tickets that was collected does not contain any demographics. Through the mappings of data to census block groups we are able to perform some assumptions of the underlying distribution but this might not be enough to provide any statistical significance to prove that there are underlying disparities in the city services.

After completing Deliverable 3, we know that through the use of data visualizations we can provide some interesting statistics and proportions of the data based on a specific location (Tract/Block Group). Through the use of statistical tests such as Chi-Squared Goodness-of-Fit, we can make assumptions of proportionality for some of the data sets to show that a set of block groups does not hold to our assumptions. Lastly, we can use data and time that is collected in most data sets to show more visualization of how the data changes over time.

Works Cited

"United States Census Geography." *United States Census Geography-Related Concepts | Documentation*, learn.arcgis.com/en/related-concepts/united-states-census-geography.htm.

All data should have been collected. All project questions should have been reviewed, answered, and submitted in a written document outlining findings as a PR. You will also be asked to submit the associated data and a README explaining what each label/feature in your dataset represents. Your team should meet with the client before this deliverable.

Checklist

1. All data is collected CHECK
2. Refine the preliminary analysis of the data performed in PD1&2 CHECK
3. Answer another key question CHECK
4. Attempt to answer overarching project question CHECK
5. Create a draft of your final report
6. Refine project scope and list of limitations with data and potential risks of achieving project goal CHECK
7. Submit a PR with the above report and modifications to original proposal