Deliverable 1
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**

Data batch collections:

311data.csv
Description: a file containing information (location, district…) of all social services in the City of Revere.

Preprocessing Data

We looked at non-null sections of location data to plot on a map, chose a preliminary set of features to analyze in the 311 dataset, and started looking at which labels in the Request Type feature were the most relevant.

Preliminary analysis:

Demographic distributions using k-means clustering

1. We collected the data from the client (311 data), a file containing information (location, district…) of all social services in the City of Revere.
2. We decided to start by choosing the columns of data that will be helpful for our preliminary analysis in order to perform clustering and counting the distributions of social services. Specifically we chose:
    a. cols = ['Address', 'Comments', 'Create Date', 'Department', 'Department ID', 'District', 'Master Request', 'Request ID', 'Request Type']
3. When analyzing, we found several rows without a specific address, for the sake of accurately defining the locations of social service, we decide to get rid of the rows that have incomplete address info (Fig 3)
    a. There were also several rows with inaccurate latitude/longitude. If addresses are available, the coordinates will need to be updated, otherwise the row will need to be removed.
4. The most popular request by far looks to be COVID Food Assistance, making up 4154 requests over the entire 26856 requests from 7/17/2017, but more work

needs to be done reclassifying Request type, so this could be less pronounced. (Fig 1)

5. Looking at the full years for where the database has entries (2018-2020) the counts of the years is 5451, 6925, and 11069respectively (Fig 2). The graph in Figure 4 denotes the enormity of city resources that were directed towards the pandemic, which raises questions about the locations of COVID relief requests.

Key Question and Answer:

Question: What are the most important datasets, and what data could be very useful if the group had it?
Answer: The 311 dataset seems very useful for comparing the location of data points with the sort of request that was made there. We are able to use the requests to retrieve information of the most popular request and also the distribution of the requests by district. The data included detailed information and we chose several columns that are useful for our research and analysis in the graphs below. More work has to be done to whittle down the 180+ distinct categories in the request, likely following the example set by Kansas City.

Project Scope

Correlate demographic location data from 2018 onward from the Census with locations of where people are requesting help from the city and where the public works requests are coming from.

New Limitations & Risks:

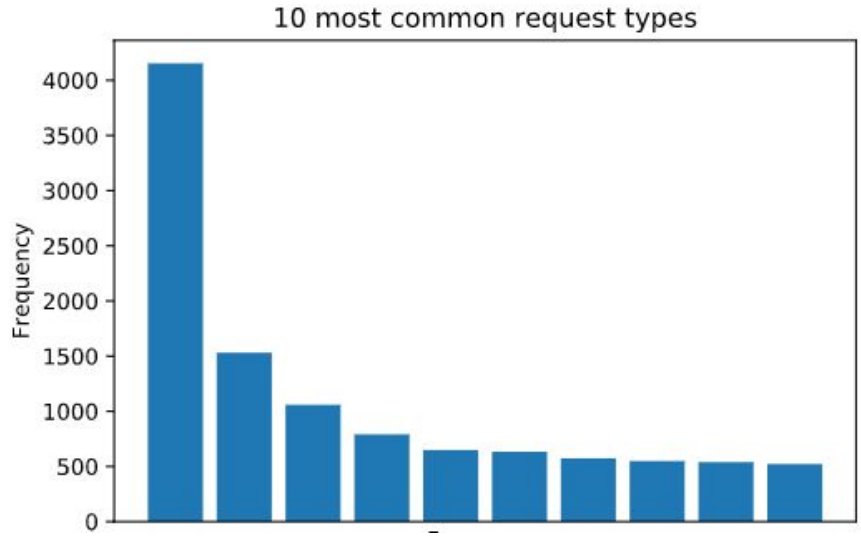
Through the discussions between group mates and with the client, there is no change to the product scope, but this could change when we present our initial findings. Potential risks to achieving the project goal include the implications of so many city resources going towards COVID rather than their normal duties, but more discussion about inner workings of the city government should be addressed in the next meeting.

Questions for the next meeting:

1. District ID or District?
    a. Do we know how Census blocks line up against Revere's Districts?
2. If we can get the building and land values? Property info? (price and owner useful for avg)?
3. A list of distinct social services locations?
4. What changes in the role of the city have you noticed in the last year?

Visual Analysis

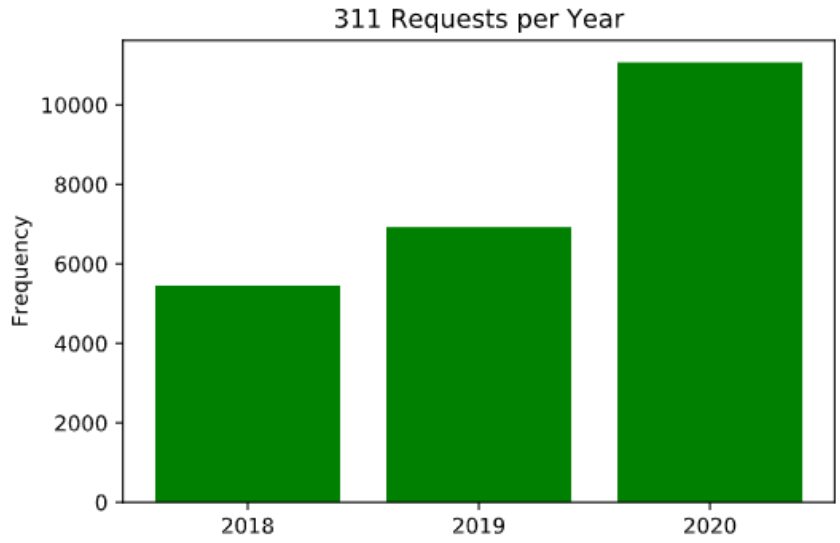
Fig 1 - Common Requests Types in 311 data



Requests Categories: ['Covid19 Food Assistance', 'Pothole', 'Overgrowth/Unsanitary Conditions', 'Sidewalk repair', 'Private Property: Unsanitary or improper trash disposal', 'Request for New/Replacement Recycling Bin', 'Debris/Litter Pickup', 'Dead animal', 'Tree maintenance request', 'Sign Missing/Damaged']

Discovery: The top three request categories are Covid19 Food Assistance, Pothole and Overgrowth/Unsanitary Conditions.

Fig 2 - Number of Requests over the past 3 years



The graph indicates a gradually increasing number of requests over the past 3 years.

Fig 3 - Data plotted by latitude/longitude, colored by District (311data.csv)

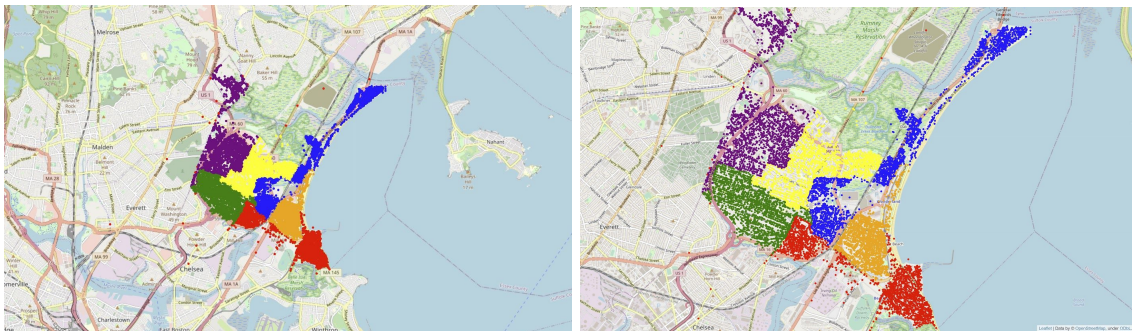
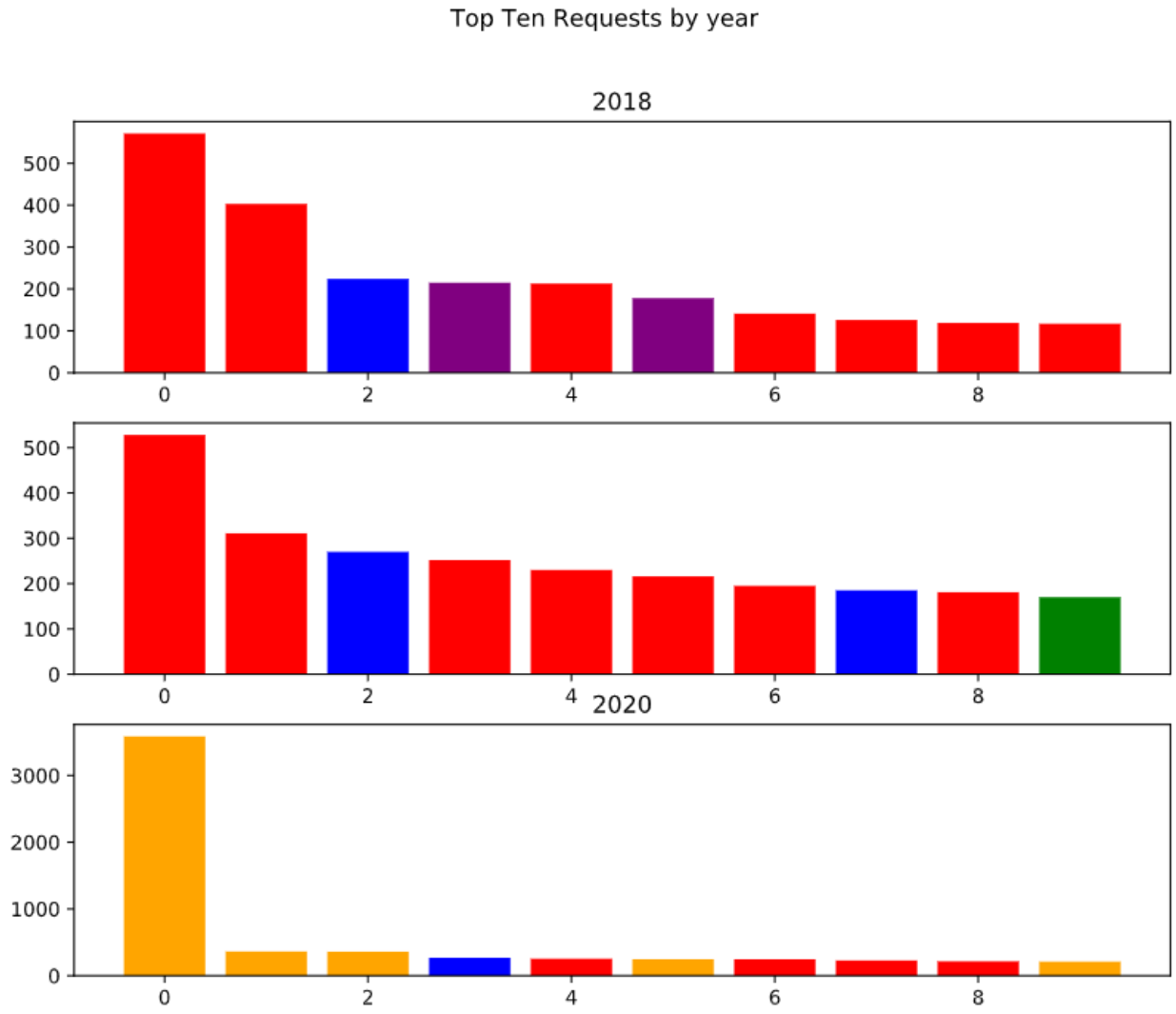


Fig 4 - Notice each graph is on a different scale.



Key: Orange: Covid Related,
Blue: Request for Item,
Red: Request for city work,
Purple: Pest Related

Counts of Tasks in graph 4, by year: {

2018: [('Pothole', 571), ('Overgrowth/Unsanitary Conditions', 403), ('Request for New/Replacement Recycling Bin', 224), ('Rodents Private Property', 215), ('Sidewalk repair', 213), ('Dead animal', 178), ('Sign Missing/Damaged', 141), ('Private Property: Unsanitary or improper trash disposal', 126), ('City Overgrowth/Unsanitary Conditions', 119), ('Tree maintenance request', 117)],

2019: [('Pothole', 528), ('Overgrowth/Unsanitary Conditions', 311), ('Request for New/Replacement Recycling Bin', 270), ('Private Property: Unsanitary or improper trash disposal', 252), ('Salt request', 230), ('Sidewalk repair', 216), ('Debris/Litter Pickup', 195), ('Trash Cart - DPW Assigned', 185), ('Sign Missing/Damaged', 181), ('Illegally parked vehicle', 170)],

2020: [('Covid19 Food Assistance', 3585), ('Covid19 Other Assistance', 362), ('Covid19 Case Related', 359), ('Request Recycling Sticker', 268), ('Debris/Litter Pickup', 254), ('Covid19 Business Related', 247), ('Pothole', 246), ('Overgrowth/Unsanitary Conditions', 228), ('Plow request', 215), ('Tree maintenance request', 210)]

}