CS 507
311 Requests in District 7
Project Deliverable 1

Project goal:

To understand how district 7 is being served vs. other districts in Boston in regards to 311 service requests, to obtain a better picture of how city resources are being allocated. To do so, we have a source of public information and are exploring the analysis up to five years in the past to map the evolution of these requests.

Preliminary analysis:

In our preliminary analysis of the data, we first found the number of requests from D7 per year and per month. From the last 5 years, from 2017 to 2021, requests averaged between 28,000 and 32,000 requests per year. The summer months averaged more requests than the others, with the most occurring during August and September with more than 14,000 requests. We then looked at the total requests over the last 5 years from each district and the districts' per capita average of requests. D7 had the third highest number of total requests, following D1 and D2. The per capita averages of requests followed a similar pattern and showed a similar graph.

Lastly, we analyzed the average time from start to close for requests in D7 and requests in all districts outside of D7 in the last 5 years. To do this, we filtered out the requests that were never closed and found that D7 had about 10.66% and all other districts had about 11.83% of their requests not being closed. Of the requests that were closed, D7 averaged 11.5 days and all other districts averaged 11.17 days to close their requests. Due to outliers in extremely long start to close times for certain requests, we believe that with further data analysis by filtering out such outliers, we may be able to find numbers that show a bigger difference in average start to close time.

Questions answered.

In the analysis above, we were able to answer some of the initial questions that our client inquired about. Initially a general idea of how D7 is compared to other districts. Additionally, the stationarity of the requests clustered in a greater proportion during specific months

raises some interesting hypotheses of why those months are particularly high, which will eventually help us answer the question of where the requests are coming from.

Scope and limitations re-evaluation.

One of the aspects which interests the client a lot, is to have a way to identify requests within specific regions, some of which may or may not coincide with the variables classification of the initial data sets. Some alternatives have been discussed, such as using the zip code, addresses, and neighborhood, or a combination of them to obtain the desired classification intended.

Tasks detail of the team members:

- Brian:
    - Average time from start to close for entire dataset
    - Average time from start to close for D7
- Kevin:
    - Analysis of frequency of 311 requests in D7 by time frame, monthly and yearly.
- Ryan:
    - Count of requests in D7 compared to the rest
    - Gather population of each district (for later normalization)
    - Requests/population for D7 vs non-D7
    - Generated role ideas for discussion on splitting work
- Samantha:
    - Filter 311 dataset and produce csv file with district 7 data and csv file for all districts except district 7
    - Download and start to understand and filter census data for demographic information
    - Create presentation for meeting with client