311 Exploratory Analysis

With concluding recommendations for the NYC leadership, especially the Department of Information Technology and Telecommunications (DoITT)

October, 2017

"DoITT will redesign the 311 architecture to make the system more robust, further streamline customer service, enhance service request processing, and increase access to 311 data." ~DoITT Strategic Plan 2015-2017

Overview

This study explores the frequency and timing of 311 complaints, complaints per City department, the prevalence of noise complaints, and City response times. It uses citywide data to compare stats from zip code 11201 (the location of NYU CUSP) and zip code 11216 (where I lived while in graduate school) with the citywide averages. It also brings in demographic information and presents a model that can be used to compare income, race, and other factors with 311 usage rates and response times to determine if 311 implementation is equitable.

The analysis conducted here pertains primarily to these DoITT Strategic Plan Goals:

Goal 1: Enhance and Improve Services

Goal 2: Expand Strategic Role with Agencies

Goal 3: Invest in Human Capital

Goal 6: Facilitate Greater Access to Technology for All New Yorkers

Methods & data

METHODS

This entire project was done in Python plus one visualization using Carto. I created a variety of functions, which output a combination of descriptive statistics and plots: response_stats(df), plot_complaints_per_agency(df), complaints_per_day(df), day_of_week(df), and group_by_complaint_type(df). Creating functions allowed me to easily run the same analyses on any data inputs. Carto was used to show a spatio-temporal distribution.

I cleaned the data a few ways. Rows with zip code values with a number of characters not equal to 5 were removed (but only for analysis that required grouping by zip code). Complaints with a 'closed date' before the 'created date' were removed for the response time analysis. In some cases I removed outliers that were more than 3 standard deviations above the mean, so I could assess typical 311 behavior (these data should not be removed in all cases).

DATA

I used 311 complaints for March 1, 2016 to February 28, 2017 so I would have a single, complete year. I gathered 311 complaints from this period for the entire city and compared them with those for zip code 11201.

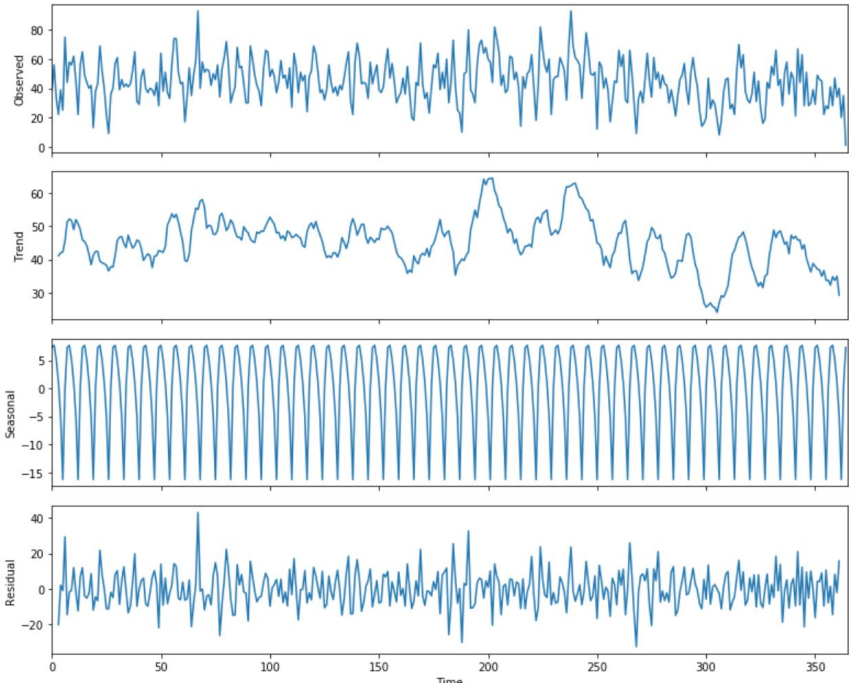
I also included demographics data. This poor quality data set was for demonstration purposes only.

Describing 311 calls in zip code 11201

There are an average of **44.85 complaints per day**. Weekdays are above average and weekends are below average. The graphic at right shows the **weekly seasonality** of the complaints (third chart), as well as a general downward trend over the last few months (second chart).

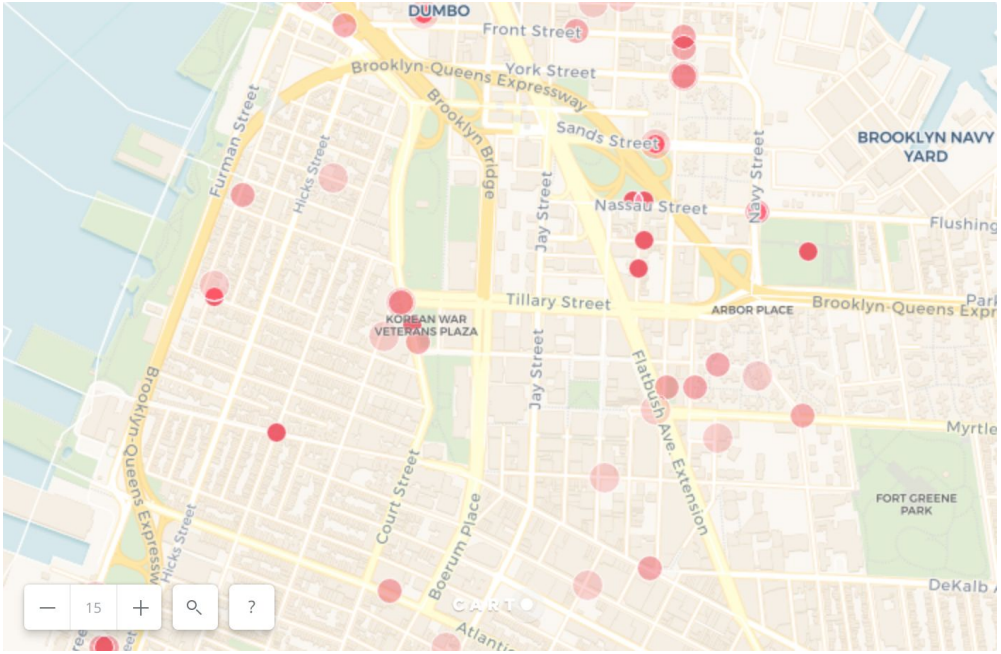
The **median response time (time from complaint creation to closing) is 30 hours** (more on this later).

Below, zip code 11201 is compared with my zip code, 11216 (parts of Bed-Stuy and Crown Heights). I'm surprised to find more complaints per capita in the poorer, more diverse neighborhood (I had heard that the wealthy used 311 more). Note: response time below is average days.



	zip	income	pop	whitepop	perc_white	median_age	res	biz	perc_biz	area	complaints	complaints_per_capita	response_time	pop_density
0	11216	47107	49071	7847	0.159911	34.0	21243	1338	0.059253	0.935	22094	0.450246	9.1617	52482.352941
1	11201	101150	61165	35835	0.585874	35.7	31207	3113	0.090705	1.421	16371	0.267653	8.9170	43043.631246

Noise is the top complaint in zip code 11201



	Number_of_Calls
Noise	1753
Illegal Parking	1472
Noise - Residential	1121
Street Condition	982
Broken Muni Meter	707
Street Light Condition	691
Noise - Street/Sidewalk	489
HEAT/HOT WATER	433
Taxi Complaint	431
Noise - Commercial	399
Water System	363
Blocked Driveway	339
Traffic Signal Condition	335
General Construction/Plumbing	324
Maintenance or Facility	300

Click HERE to see an animation of all noise-related complaints.

Comparing 11201 w/ the average NYC zip code

Zip Code	Median Response Time (hours)	Calls Per Day	Weekly Cycle
11201*	30:06:44.5	44.85	More complaints on weekdays
11216**	36:13:00	60.53	More even
NYC Average	26:32:00	between 13.90 and 21.62	More complaints on weekdays

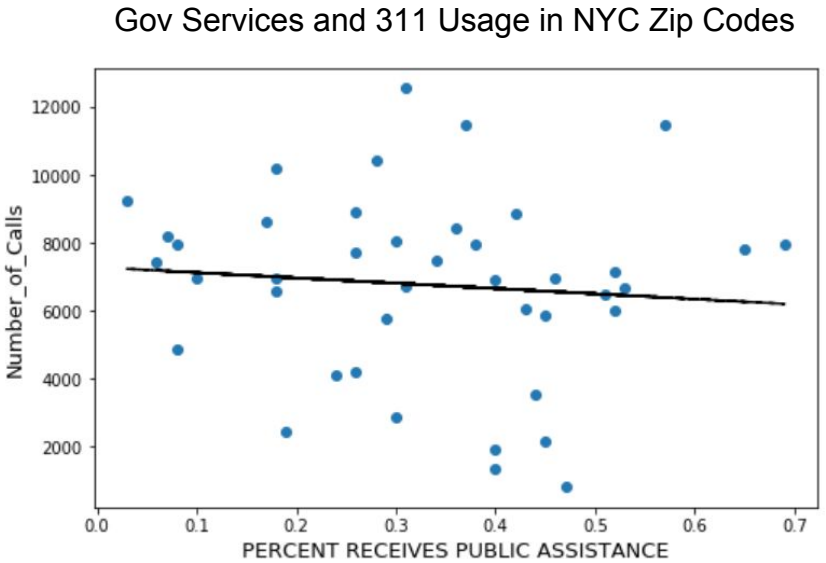
* Downtown Brooklyn
** My neighborhood

Who uses 311? Do all get equal service?

I wrote code to compare 311 usage per zip code with demographics per zip code. The model is built to show the correlation between a given social factor (i.e. income, % white, or age) and the number of calls per capita.

This method can also be used to assess the relationship between demographics and response times, to see if all zip codes are being given equal service.

NOTE: The demographic data that NYC shares at the zip code level reflects a survey of only a few people per zip code. This is too small a sample, so my results should not be used. The method, however, could be used on a better dataset to produce useful insights. Sufficient demographics for zip codes could be gotten by doing a weighted aggregation where the spatial area is used to weight each value.



Dear DOT: what's up with street lights?

In zip code 11201 (downtown Brooklyn) there were 171 rows where the closed date was before the created date, meaning one or both of these dates was entered wrong. Most of these impossible timestamps are from DOT and are of the type "Street Light Condition". **We need to go talk to the people who record these complaints.**

There also are several DOHMH "Rodent" complaints that are similarly entered wrong. Let's talk to them, too.

Next steps for this analysis

Acquire or create data for income and race per zip code. Compare with 311 response times.

Compare income, race, gender, and age data with 311 complaints per capita to answer the question - who complains? Consider how the population reflected in 311 complaints affects delivery of city services. In areas with lower 311 usage per capita, consider if increasing **mobile broadband access** could help boost usage (this would also serve Strategic Plan Goal 6).

When comparing 311 usage per capita between zip codes, maybe **normalize for population density**, which probably has a great effect on issues like transportation.

Methods improvement: normalize complaints per zip code by population, since not all zip codes have the same number of residents.

Recommendations for the Mayor's Office

Do everything possible to **make 311's userbase representative of New York**. A study in Houston found that different racial groups had different views about city planning (parks specifically, in that case).

Illegal parking was the second biggest complaint in zip code 11201. Please increase enforcement, especially in bike lanes, to increase the disincentive to park illegally.

Monitor the **median response time**. Try to decrease it over time. Also look at the distribution of response times based on location and complaint type to keep the customer service side of 311 equitable and responsive.

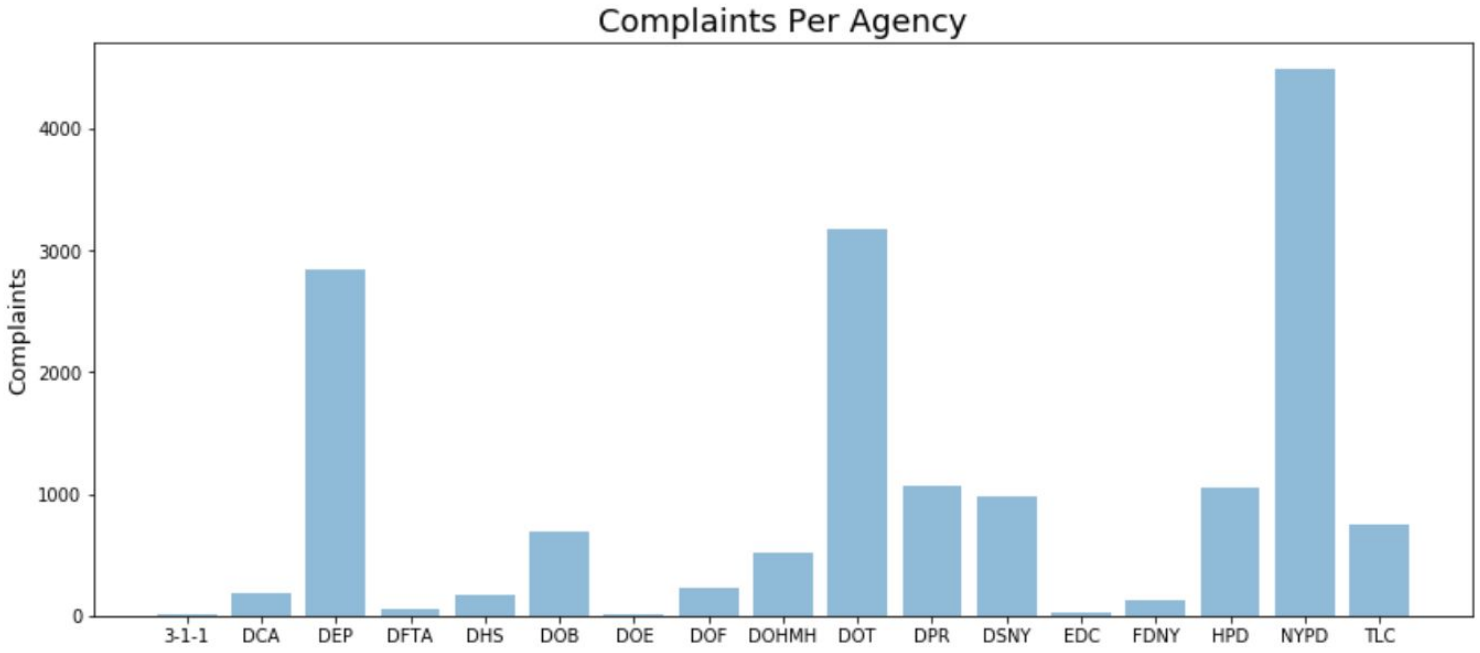
Recommendation for DoITT leadership

I had a conversation this week with Dan Ault, Chief Innovation Officer for Cary, NC. He said that a city's goal should be to "optimize human ingenuity and push decision-making to the edges, not just the top".

Recommendation: HACKATHONS

Maybe twice a year. Create several teams, each consisting of two DoITT folks and two from another agency. They should have different roles/skillsets and none of them should be close colleagues. Have them use 311 data to try to solve a problem for that agency. Stress that they should be educating each other along the way; one teaches about analytics, one about relevant policy, etc. This develops human capital (Strategic Plan Goal 3), builds relationships between employees at DoITT and other agencies (Goal 2), and may lead to improved services (Goal 1). When appropriate, a team could include someone from outside government as well (say, a resident, software developer, or transportation consultant).

Appendix 1: Complaints Per Agency in 11201



THANKS FOR READING!