311 Service Requests Explanatory Analysis

Data and Methods

311 serves the public and handles all the requests for NYC government agencies and non-emergency services. I obtained the 311 Service Requests data from NYC OpenData. The dataset contains the service requests since 2010 with detailed information on each request. **My primary goal is to explore the processing time of agencies to investigate the potential inefficiency, misuse, and unfairness of dealing with people's requests.**

The entire project was done in Python with visualization in matplotlib, seaborn, and folium. I cleaned the data in a few ways. I selected 11 useful columns out of total 41 columns of the dataset. Rows with zip codes that don't belong to NYC were removed. I calculated the processing time of each request from its 'created date' and 'closed date,' and I removed the entry errors of the date. Since all the selected columns will be used, I also removed rows with any null values.

Investigation on 2019 DOHMH 311 Data

Because the whole dataset was huge, I only selected 2019 service requests to begin my research. Firstly, I calculated the average processing time of each agency shown below to check how quickly they can deal with the requests. **We can see DOHMH (Department of Health and Mental Hygiene) had the second-highest average processing time around 40 days. Since DOHMH is critical to the NYC public's life, it is worthy of further investigation of its performance.**
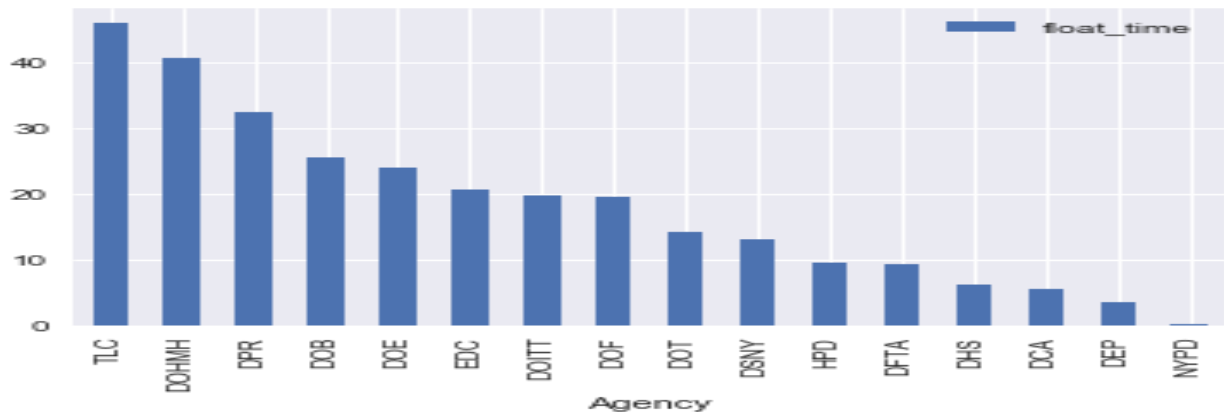
I was interested in the complaint/request types for DOHMH. As a result, "Rodent" complaints dominate with 58.5% of the total complaints.
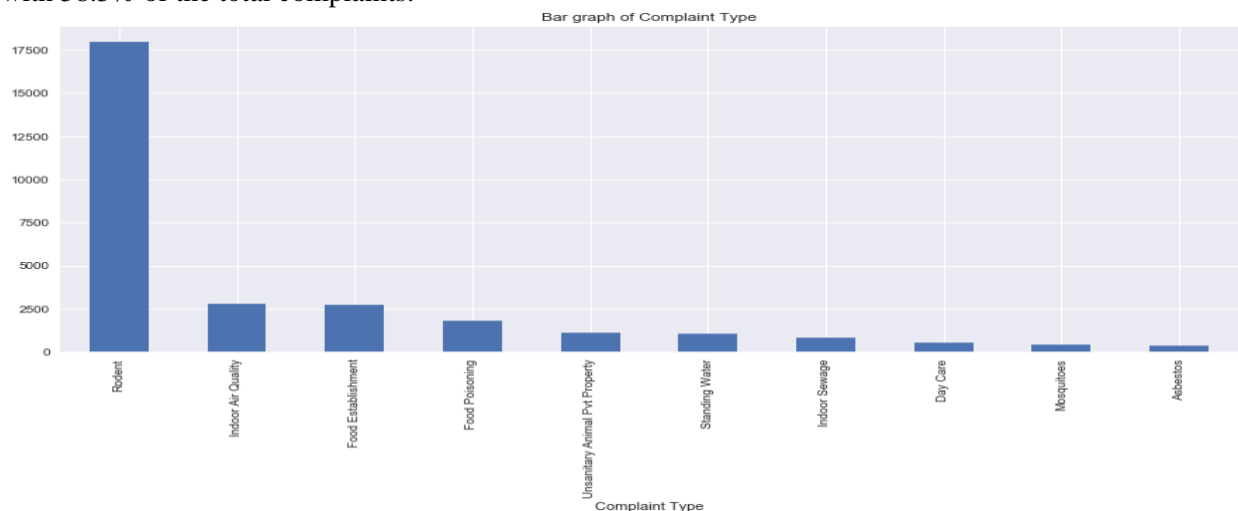
Since the processing time of every compliant type might be different by nature and "Rodent" complaints are the most, I decided to focus on the 2019 "Rodent" cases to avoid changing factors.

Before probing into details of the processing time, I investigated the resolution rate of "Rodent" complaints. The dataset has a column called "Resolution Description", which described in detail how the agency resolved requests. I categorized the various description into main summaries so that we can have a clear look at how many cases have been solved. According to the statistics, most cases have been closed still under review status. In other words,

under review wait	15179
violations found	1000
duplicate complaint	715
no violation found	679
no access to the property	197
ambiguous	173
cannot complete inspection	2

DOHMH only recorded action for 15.4% incidents (all the incidents besides under review status), no matter whether they have resolved the incidents or not. Understandably, an agency cannot take action for every complaint, **but a 15.4% action rate still seems too low, and it shows DOHMH potential ineffectiveness of allocating its resources to resolve public complaints.**

Then, I grouped the incidents by 2019 months and calculated the monthly number of cases and mean processing time as below. We can tell from the table that the number of complaints increased with the approaching of summer, while the processing time also increased significantly.

yyyymm	count	mean
201901	1998.0	6.277139
201902	1862.0	5.306095
201903	2389.0	4.720261
201904	2766.0	15.482170
201905	3235.0	115.807194
201906	3108.0	87.659187
201907	1.0	138.118437
201908	3.0	100.973588
201909	24.0	72.344057
201910	169.0	26.510755
201911	989.0	8.432228
201912	1401.0	4.616989

For example, compared with January, the number of complaints increased by around 50% in May, but the mean processing time increased by 18 times. **This situation indicates the inefficiency of the DOHMH's capability in dealing with increasing incidents.**

Nevertheless, it is suspicious that there were only single digit number of cases in July and August. So, I suspect that there might be some serious data entry errors or some significant changes within the DOHMH system. In order to investigate further, I extracted the 2016 – 2019 311 service requests for DOHMH to re-check my findings.
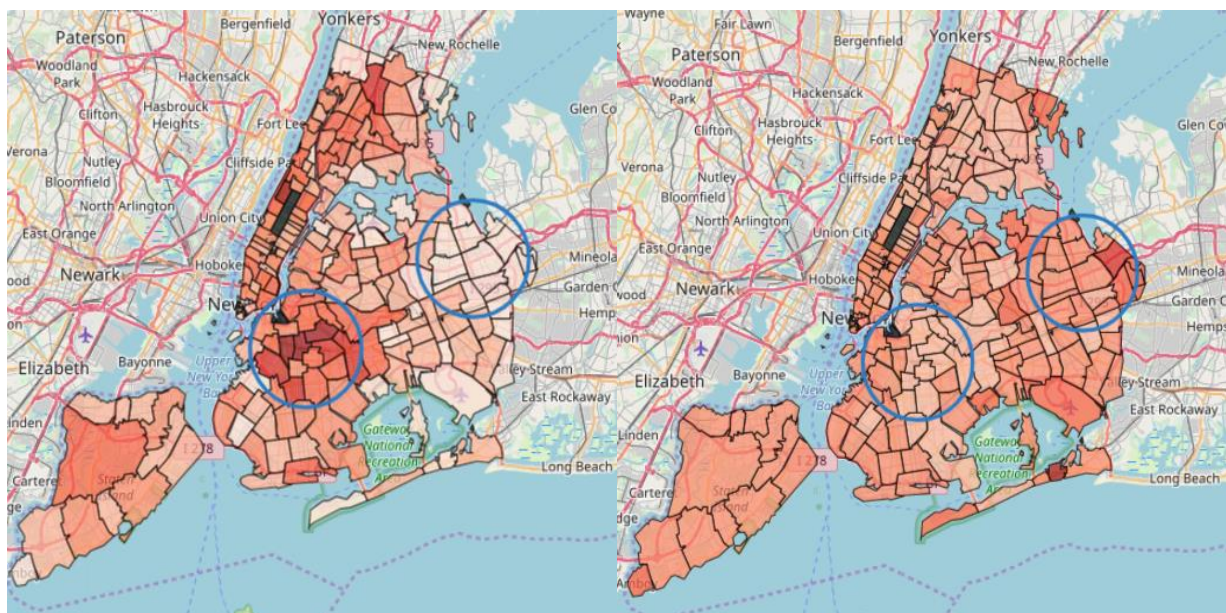
Investigation on 2016-2019 DOHMH 311 Data by Time
As before, I still focused on the "Rodent" complaints from 2016-2019 DOHMH data. The "Rodent" complaints again took more than half, with 56.5% of total incidents. I plotted the time series data in terms of the number of cases and mean processing time on monthly basis. We can see the number of incidents followed the periodic manner (high in summer and low in winter) in the past four years, but the trend abnormally dropped in 2019 summer. In addition, the processing days have been stable before 2019, but suddenly jumped up after 2019. **The abnormal 2019 data validates the assumption of possible data entry errors from DOHMH or any systematic changes within the agency that resulted in its inefficiency. We can consult with DOHMH to check what exactly happened.**
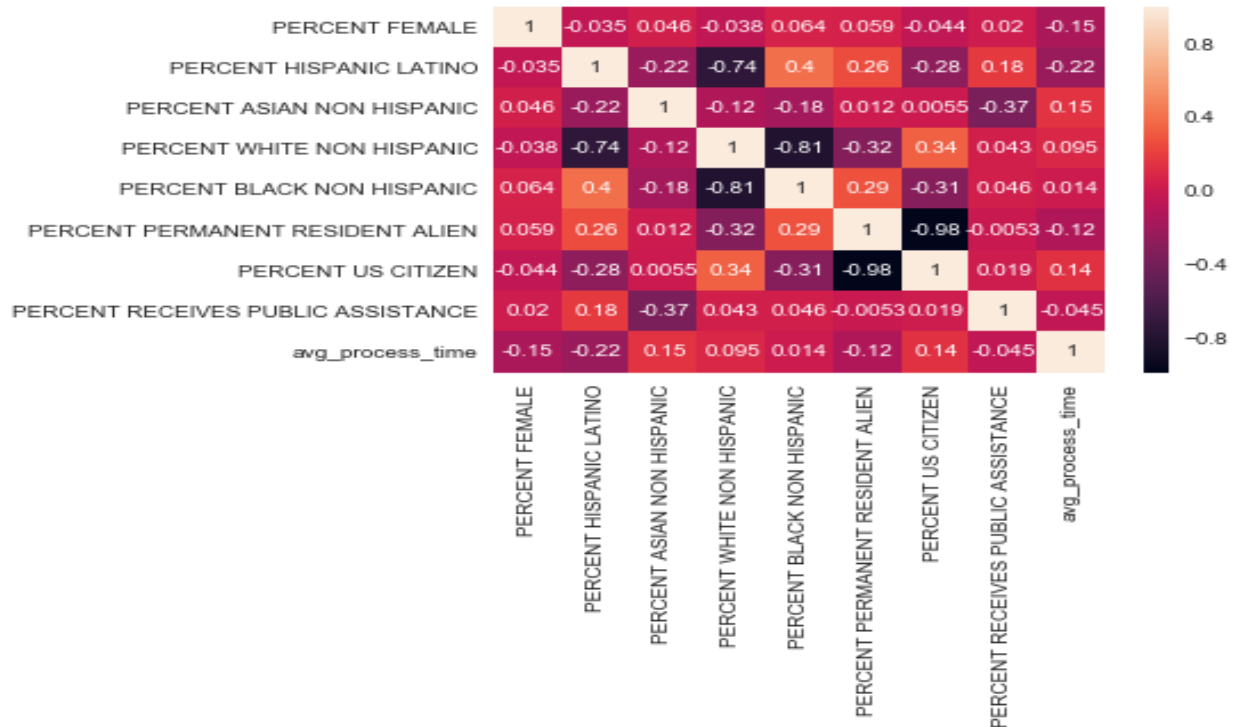
incidents of Rodent Compaints on monthly basis

processing days of Rodent Compaints on monthly basis
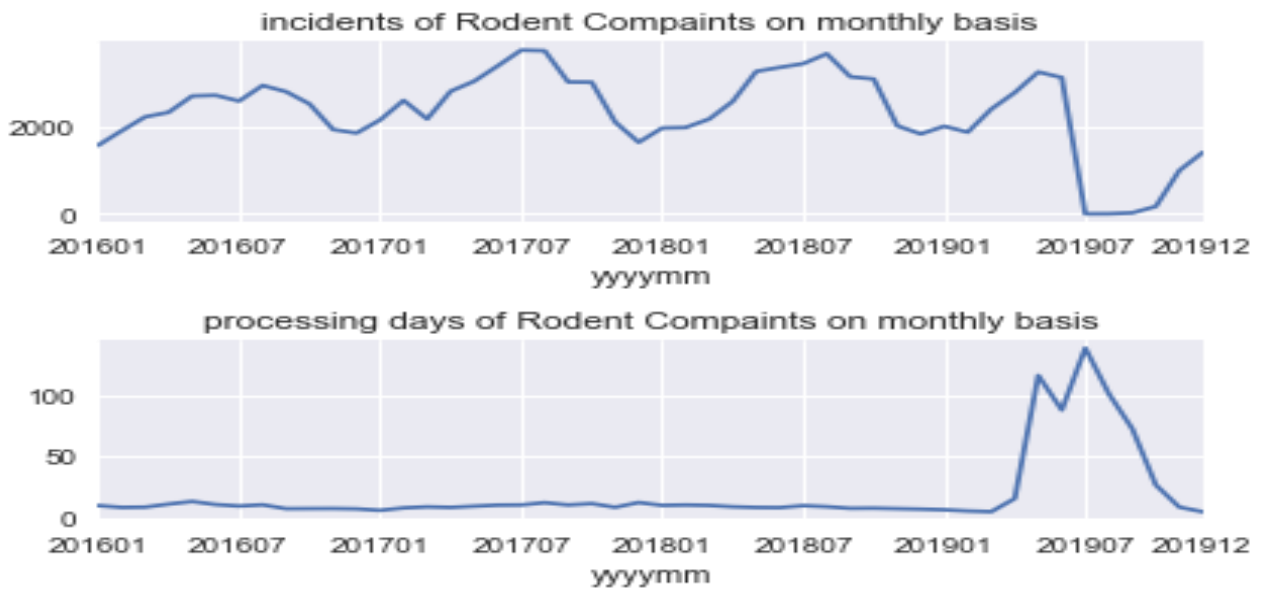
Investigation on 2016-2019 DOHMH 311 Data by Zip Code

The 311 dataset also contains the zip code for every complaint. So, I could exploit this feature to study if all zip codes were given equal 311 services from DOHMH. Continued on the 2016-2019 DOHMH "Rodent" complaints data, I constructed the heat maps of total the number of cases (the picture on the left) and average processing time (the picture on the right) by zip code region. The darker the color and the larger the measuring units of that area. The two pictures display the regional inequality of the service received. For example, the blue circle on the Brooklyn borough shows a large number of incidents with short processing days, while the blue circle on the Queens borough indicates a small number of incidents with long processing days. **It is good for us to see the region with a lot of complaints received more attention and care, but we should not neglect the areas that report fewer cases. Thus, we should notice the regional inequality of services provided by DOHMH.**

Investigation on 2016-2019 DOHMH 311 Data and Demographic Statistics

In terms of equality and fairness of the services, we can dig deeper by studying the relationships between demographics and processing time per zip code. I found the [Demographic Statistics By Zip Code](#) on NYC OpenData. The dataset recorded the voluntarily participated NYC residents' information such as their gender, ethnicity, public assistance, and etc., by zip code. However, the data quality is not good since there were only, on average, 17.66 participants for each zip code. Nevertheless, we can still merge this dataset with the 2016-2019 DOHMH 311 Service Requests dataset to have an initial investigation on the relationship between demographic statistics and the processing time. We can infer the correlations of variables from the correlation matrix below. For instance, the processing time is negatively correlated with the population percent of Hispanic Latino and positively correlated with the population percent of Asians. My current results should not be used immediately since the poor data quality. If we have a better dataset in the future, we can use this method to produce useful insights to investigate whether the agency has potentially discriminated against certain groups.

	PERCENT FEMALE	PERCENT HISPANIC LATINO	PERCENT ASIAN NON HISPANIC	PERCENT WHITE NON HISPANIC	PERCENT BLACK NON HISPANIC	PERCENT PERMANENT RESIDENT ALIEN	PERCENT US CITIZEN	PERCENT RECEIVES PUBLIC ASSISTANCE	avg_process_time
PERCENT FEMALE	1	-0.035	0.046	-0.038	0.064	0.059	-0.044	0.02	-0.15
PERCENT HISPANIC LATINO	-0.035	1	-0.22	-0.74	0.4	0.26	-0.28	0.18	-0.22
PERCENT ASIAN NON HISPANIC	0.046	-0.22	1	-0.12	-0.18	0.012	0.0055	-0.37	0.15
PERCENT WHITE NON HISPANIC	-0.038	-0.74	-0.12	1	-0.81	-0.32	0.34	0.043	0.095
PERCENT BLACK NON HISPANIC	0.064	0.4	-0.18	-0.81	1	0.29	-0.31	0.046	0.014
PERCENT PERMANENT RESIDENT ALIEN	0.059	0.26	0.012	-0.32	0.29	1	-0.98	-0.0053	-0.12
PERCENT US CITIZEN	-0.044	-0.28	0.0055	0.34	-0.31	-0.98	1	0.019	0.14
PERCENT RECEIVES PUBLIC ASSISTANCE	0.02	0.18	-0.37	0.043	0.046	-0.0053	0.019	1	-0.045
avg_process_time	-0.15	-0.22	0.15	0.095	0.014	-0.12	0.14	-0.045	1

Conclusion

Through my analysis of DOHMH 311 Data, we can uncover some insights about work errors, inefficiency, and inequality of DOHMH dealing with the public's requests. We should talk with DOHMH to further understand the situation and collaborate with them to solve potential problems. Moreover, the analysis methods can also be applied to any other agency to start the oversight and investigation process in a data-driven manner.