**311 Service Requests Analysis in New York City**

Jane Liang ([bliang12@calstatela.edu](mailto:bliang12@calstatela.edu))

Weiqiang He ([whe8@calstatela.edu](mailto:whe8@calstatela.edu))

Yiqing Qian ([yqian6@calstatela.edu](mailto:yqian6@calstatela.edu))

Department of Information Systems, California State University Los Angeles

**Abstract:** This paper will analyze 311 service requests in New York City from 2010 to October 31, 2019. 311 is a non-emergency phone number that people can call in many cities to find information about services, make complaints, or report problems. This dataset is mainly categorized by the unique key of the case, created date, closed date, responding City Government agency, complaint type, location type, incident location, case status, and the channel type of complaint. After going through all the analyzation process, we are able to find out the most common channel type to file complaints, the most and least incident prone area, leading reasons, and also the time range to close the case.

**1. Introduction**

311 is a 24/7 non-emergency phone number. NYC311 was first launched in 2003. Its mission is to provide the public with quick, easy access to all New York City (NYC) government services and information; help Agencies improve service efficiency; provide insight to improve city government through accurate, consistent measurement and analysis of service deliver.

Analysis of 311 calls can be of great use for wide variety of purposes, ranging from a rich understanding of the status of a city to the efficiency of the government services in addressing such calls.

**2. Related Work**

Similar works have been done using the 311 data of NYC. For example, in Zha’s paper, separate-and-combine approach, observation and understanding of the individual patterns in subsets of data were used to extract more effective features, in order to get predictive model. Another example is in Minkoff’s paper, regression models were applied to analyze 311 data to understand the distribution of problem within NYC. Thus, our paper is different from those two previous works, since we use Hadoop Cluster in AWS and visualization tools to analyze our data.

**3. Background**

NYC has made its service-request 311 call data available through its NYC Open Data. Each day, NYC 311 receives thousands of requests related to several hundred types of non-emergency services. These requests are then forwarded to relevant agencies, such as Police and so on. The agency will respond to the request until the case is closed.

**4. Workflow**

First, we downloaded this dataset of 311 service requests analysis in New York City from a trusted source. The data set contains the detailed information of the past 10 years’ 311 service request, which includes Unique Key, Created Date, Closed date, Agency, Complaint Type, Incident Zip, Latitude, Longitude and so on. We use SCP command to copy dataset to the Cluster, then we move the dataset into hadoop distributed file system (HDFS). The data is analyzed using AWS Cloud and HIVEQL and the results files are stored in the cloud. When we need to do visualization , we download the results files and open them in Excel format. Using Tableau and Excel is to perform data visualization.

By analyzing the results, we could know that the year of highest number of cases, average days to close one case, the most popular channel type for service request, the most complaint types, etc.

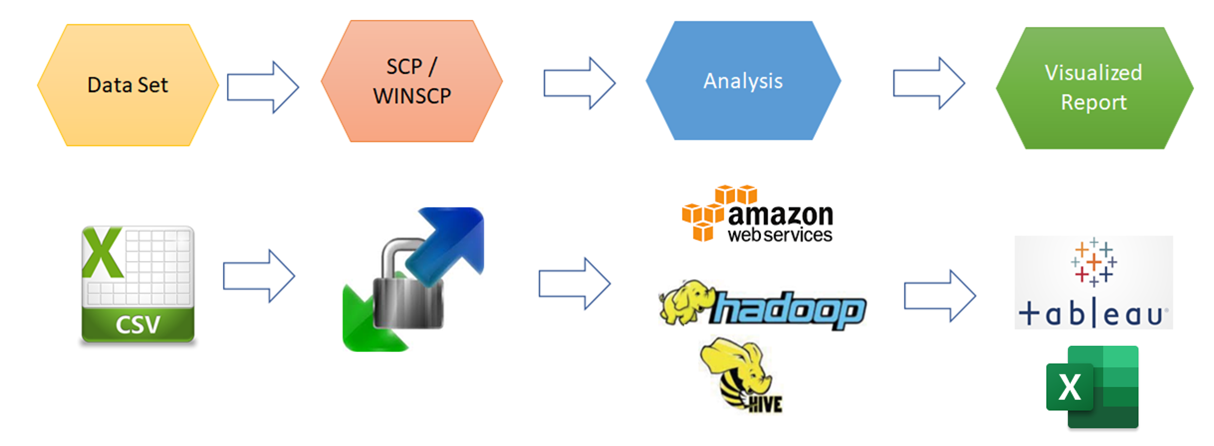


Figure 1. Workflow

**4. 1 Data Analysis and Representation**

In this Project, we performed visualization to have a better understanding of the dataset. We tried to get answers based on the following questions:

1. What is the trend of the total cases numbers per year?
2. How many days to close one case in each year? What is the trend?
3. Which channel type do people use most when they report cases?
4. What are the most common complaint types in New York City?
5. Which city has the most number of cases?
6. Which city and compliant type has more than 150,000 during the past ten years?Number of Cases per Year?
7. On July 4th, 2019, where had the highest number of cases?

After running these related queries, the following screenshots of the results could answer our questions.

1) Number of Cases per year

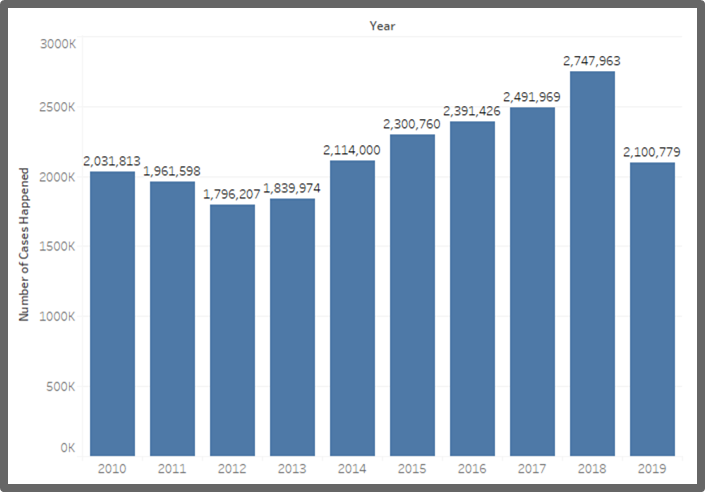


Figure 2. Caption of the Figure

From the bar Chart, we can know how many cases happened in each year. The number of cases is in a trend of going up by year. Because this dataset collected data only until October 31,2019, the collection of total cases happened in 2019 is incompleted. Year 2018 had the highest amount of cases, which is 2,747,963; Year 2012 had the lowest amount of cases, which is 1,796,207.

2) Average days taken to close one case for each year:

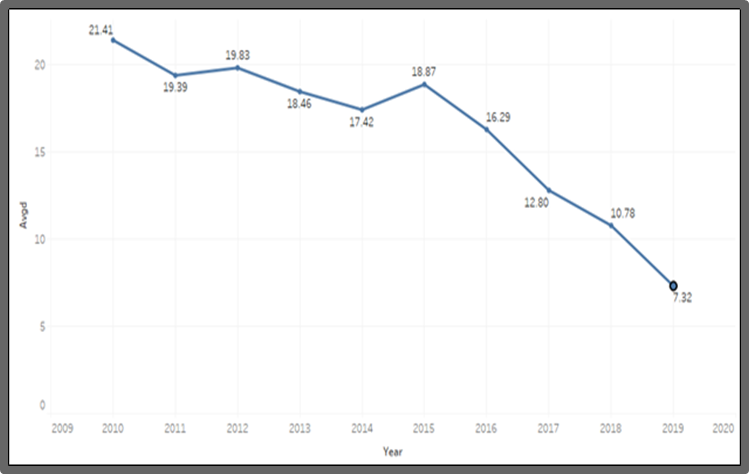


Figure 3. Average Days Taken to Close the case

This line chart shows that the trend for average days taken to close each case is decreasing yearly. So the longest time to close a case is 31.41 days, which happened in 2010; the shortest time is 7.32 days which happened in 2019. It could prove that the service provided by service agency are becoming more and more effective.

3) Channel type by percentage:

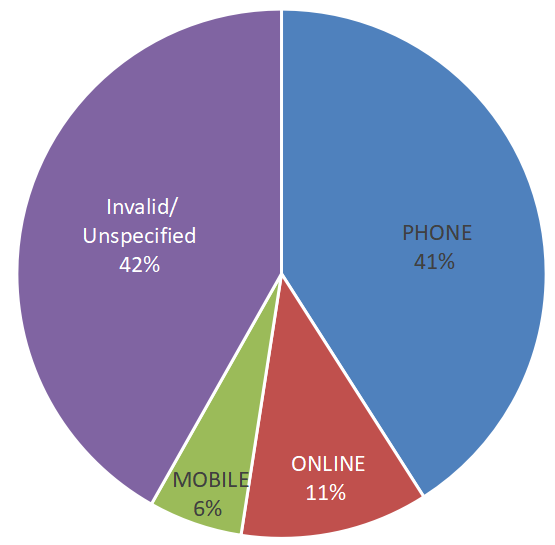


Figure 4. Caption of the Figure

Through this pie chart, we analyze that a majority of people usually use their phones (call center) to request 311 service.

4) Number of cases for the top 10 complaint types:

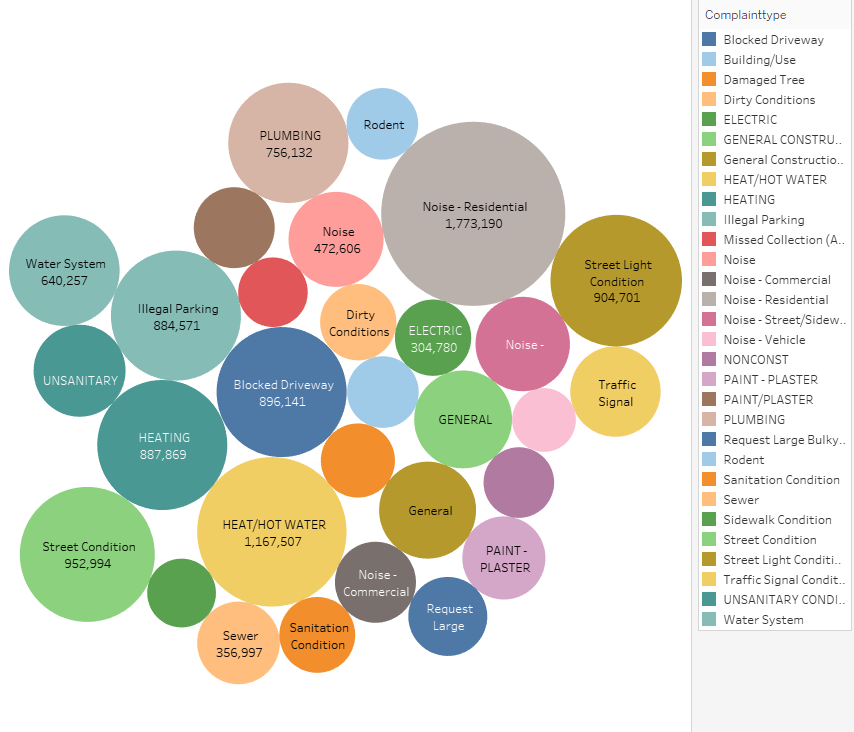


Figure 5. Top 30 by Complaint Type

This packed Bubble chart shows that the most common complaint types are :Noise-Residential, Heat/ hot water, Street condition, Heating, Blocked driveway, and Street light condition. The bubble size is bigger, which means that the number of cases for one complaint type is higher. In this case, the noise-residential complaint type has the highest number of cases in the past 10 years, which is 1,773,190.

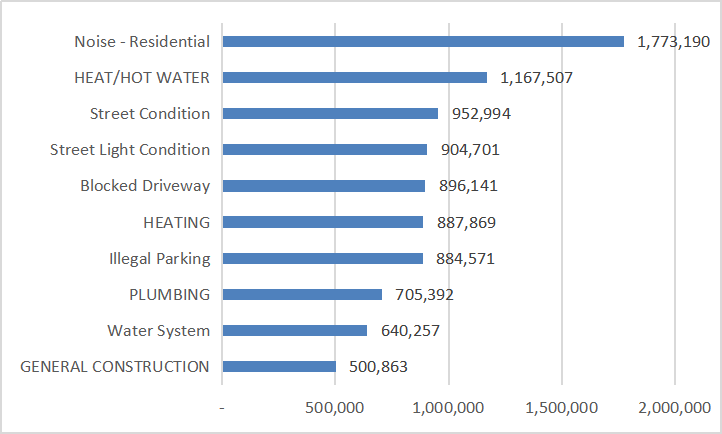


Figure 6. Top 10 by Complaint Type

In this bart chart, the noise-residential complaint type has the highest number of cases in the past 10 years, which is 1,773,190. The cases number for top 10 complaint types accounts for 43% of total case number(21.8m).

5) Top 5 cities by total case numbers

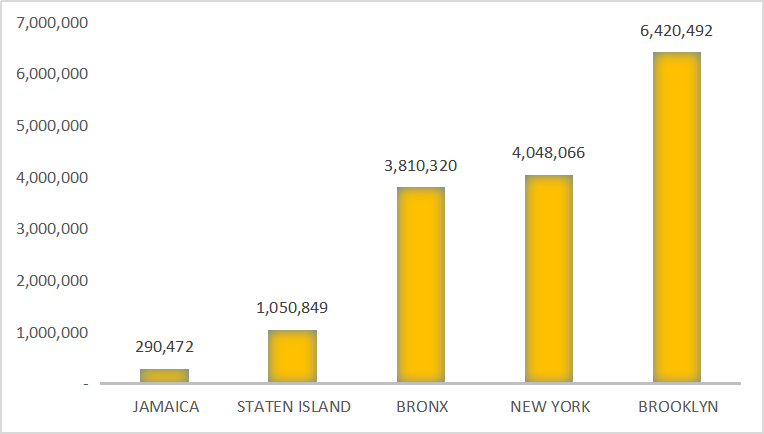


Figure 7. Top 5 Cities by Total Case Numbers

From this bart chart, it shows that Brooklyn, New York, and Bronx have the most significant complaint case numbers, which accounts for 66% of all complaints (21.8 m) in the past 10 years.

6) Number of cases in different borough and complaint type:

Condition: the total number of cases greater than 150,000 for major borough and complaint type.

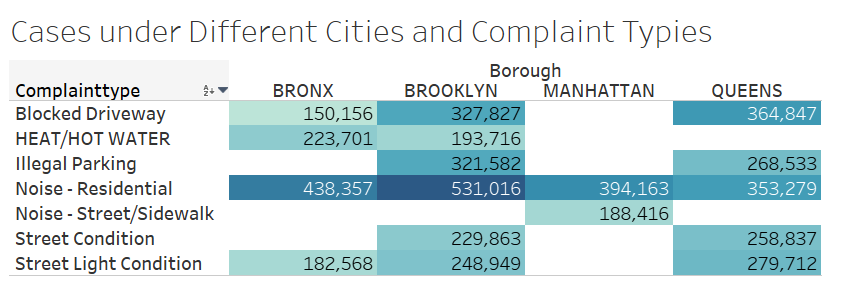


Figure 8. Number of Cases in Different Borough and Complaint Type

Through this table, we noticed that different boroughs have their different major complaint types. Brooklyn has the most cases in noise-residential complaint.

7) Complaint distribution on July 4th, 2019

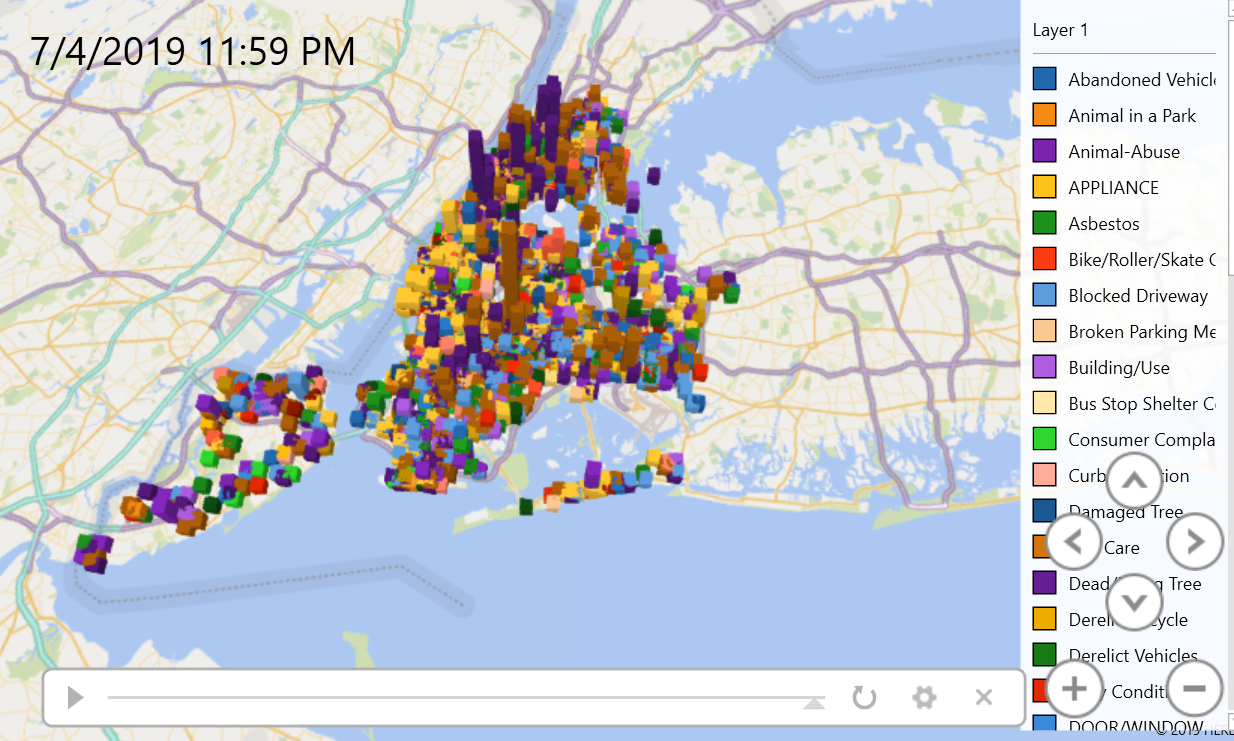


Figure 9. Complaint Distribution on July 4th, 2019

Through this geo-spatial map, we can see that the distribution of complaints happened in New York Area on the Independence Day of 2019 by time.

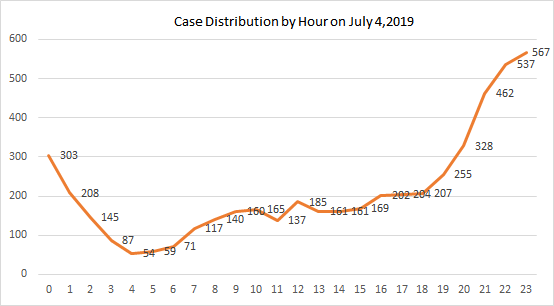


Figure 10. Case Distribution by Hour on July 4th, 2019

This is to visualize the case distribution by each hour throughout the day on July 4,2019. It shows that more cases were reported in night from 8:00PM to 1:00AM, and less cases were reported in earlier morning from 3:00AM to 7:00AM.

**5. Conclusion**

From the above analysis, we can conclude the following:

* Total number of cases is increasing year by year, with year of 2018 hit the most significant cases - 2,747,963.
* Agencies have improved their efficiency to decrease the average days to close each case.
* Making phone calls is the most common channel to request 311 service.
* Residential noise ranks #1 most common complaints.
* Most complaints happened in Brooklyn and New York City.

**6. Limitations**

The data in the dataset is not clear enough and our analysis might not reflect the real situation, for example:

* For case closed date: A lot of cases have invalid closed date (619,420 cases with empty closed date, 5986 cases with closed date in year 1900).
* For the complaint channel: 42% of cases were unspecified or invalid.

**7. References**

Dataset URL:

<https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9>.

GitHub Link: <https://github.com/aweihe/CIS5200-Group2-Project/tree/master>

Minkoff, Scott, “*NYC 311: A Tract-Level Analysis of Citizen-Government Contacting in New York City*,” <https://www.researchgate.net/publication/274708724_NYC_311_A_Tract-Level_Analysis_of_Citizen-Government_Contacting_in_New_York_City>

New York City Council

<https://council.nyc.gov/data/311-services/>

The Official Website of the City of New York

<https://portal.311.nyc.gov/>

Zha, Yilong, “*Profiling and Prediction of Non-Emergency Calls in New York City*,” <https://pdfs.semanticscholar.org/6591/0aa19450177650970f5c97bdcaa8420d9c93.pdf>