311 Service calls

Authors: Brandon ascencio, Michael Campos, Joey Tran, Garrett Tice, Kyle Alfonso

Department of Information Systems, California State University Los Angeles

CIS 4560-01 Introduction to big data

[Bascenc8@calstatela.edu](mailto:Bascenc8@calstatela.edu),- [Mcampo61@calstatela.edu](mailto:Mcampo61@calstatela.edu), [Kalfons@calstatela.edu](mailto:Kalfons@calstatela.edu), [jtran66@calstatela.edu](mailto:jtran66@calstatela.edu), [kalfons@calstatela.edu](mailto:kalfons@calstatela.edu)

**Abstract:** This paper investigates 311 Service Requests from 2010-2020. A 311 Service Request is a non-emergency phone number that people can call in many cities to find information about services, make complaints, or report problems like graffiti or road damage. We would be using methods learned in class to present the insights, data analysis, and visualization of the dataset. The purpose of our analysis of 311 service calls is to survey the different types of non-emergency calls within the New York City Area between 2010 and 2020. We want to discover patterns of the most and least frequent types of calls in different boroughs based on the most common complaints. Analyze the number of cases each agency responded to, and use the information in the reallocation of budgets. Lastly, to determine the most and least common service calls based on the responding agency and the area of new york city (I.E., park borough manhattan).

**1. Introduction**

The big apple, the concrete jungle, the capital of the world. If you have ever heard of a description about a city, most likely whoever said it was describing New York City. It is a city that is jammed packed and bustling with life. Many people reference it in songs, paintings, movies, and stories due to its fast-paced environment and glamorous hardships. Another saying goes that if you can make it in New York then you can make it anywhere. With the allure of making a living in the metropolis of New York, this study will try and discover where to move to, and which park borough is the safest. Using our data set we will sift and clean all the entries pertaining to this matter. The study will break down 311 calls in New York From 2010 to the present to try and find out which areas had the most calls and most disturbances.

2. Related Work

Thinking about 311 service calls in New York City and their usefulness in improving citywide problems it can be better understood looking into related works, that explain methods of analyzing the data. The basis of our research focuses on most common 311 service calls but understanding how 311 service calls system works and what is being done with data gathered is paramount to knowing how it can affect the progress made in NYC.

311 service calls relate to very meticulous geolocation, since each call is based on zip code and by neighborhood (I.E. Brooklyn) but locations of service calls become can be more precise. According to a publication by New York University there are lots of descriptors in the category Location. We can locate about 91.25% of requests with latitude and longitude from the data set. About 91.77% of the requests have recorded zip codes. Population in an area is one predictive feature to the number of 311 requests in that area (Zha & Veloso, 2014).

A geo-correlation coefficient helps define relationships based on type of service calls and location showing a relationship between the variables in this study.

Population density in NYC is significant and managing a system for 311 service calls is an aspect of working with large data when there are more than 7.5 million reports per year via app-based technology or by phone. A report by NY Edu states that Twenty percent of all calls and website visits result in creation of 311 Service Requests (Kontokosta, Hong, & Korsberg, 2017).This doesn’t include other means like numerous social media sites for citizens to report 311 requests on.

Further study into how 311 service data affects NYC is analyzed by NYU. The results of methods used to define relationships is significant by using methods and findings from the data to draw conclusions. Data are being used to predict problems and inform city decisions as a result the growing interest in the application of machine learning and data science to city management [8]. For example, widely available 311 data provide an opportunity to develop spatial-temporal proxy measures of neighborhood socioeconomic characteristics [8]. Specific complaint data, coupled with other city records, have also been used to predict the emergence of unsafe or unhealthy conditions, including rodent infestations and illegally converted buildings (Kontokosta, Hong, & Korsberg, 2017).

Being able to predict possible problem areas of cities and using past methodology helps a city operate more efficiently when responding to 311 service calls. Understanding the use of 311 data and the usefulness is described by John Jay College The scale of annual variation suggests that calls not only reflect service requests by residents, but are also impacted by other factors which may systematically increase call volume, including enhanced reporting efforts by public employees. For example, New York City developed a Street Conditions Observation Unit Team (SCOUT) Program in 2007, through which city agency inspectors travel the city and call 311 to report issues related to quality of life (Mulligan, Cuevas, Grimsley, Chauhan, & Bond, 2019).There is significance between relationships that officials and public must maintain to increase efficiency of NYC. Implementing initiatives to make changes by looking at past data and making further decisions clearly shows positive gains as a result on this study.

Our work focuses on most common 311 calls while analyzing a much broader view, findings can have impact to make change and allocate resources most efficiently in terms of response. The purpose of related works is using methodology to predict based on past events and how to make future decisions based on the data.

3. Specifications

The data set contains 311 calls or service requests from 2010 to the present time. NYC OpenData database is publicly open to anyone; because of this, we were able to access all of the data set fully. Due to the data set frequently being updated, the file size may differ from when it is downloaded. We downloaded the data set on November 29, 2020; because of this, our file size came out to be 12.3 GB. When compressing the file, it came out to be 2.18. The data set covers from January 01, 2010 - November 29, 2020. In the data set, there are 24.6 million rows and 41 columns. Below, Table 1 Data Specification shows the file size of it being unzipped and compressed.

Table 1 Data Specifications

|  |  |
| --- | --- |
| Data Set Unziped | Size ( 12.4 GB ) |
| Data Set Compressed | Size ( 2.18 GB ) |

Below is a table (Table 2) that shows the specification of the Orcale Cluster we used.

Table 2 H/W Specifications

|  |  |
| --- | --- |
| Cluster Version | 20.3.3-20 |
| Number of nodes | 3 |
| OCPU’s | 12 |
| CPU Speed | 2.20 GHz |
| Memory | 180 GB |
| Storgae | 957 GB |

4.Implementation of Flow Chart

We first started to download our dataset from data.cityofnewyork.us. NYC ope-n data is free public data published by New York City agencies and other companies. Then secondly, we uploaded the CSV files into Amazon Web Services. We uploaded it into Amazon Web Services because we wanted to transfer the data set into the HDFS easily. We also wanted to have the same data set when working on it because NYC open data regularly was updating the data set. Thirdly once uploading the CSV files into HDFS, we used Hive to clean and create tables using our data set. Fourthly once we had tables created, we started to create queries to analyze data using Tableau. Finally, in Tableau, we created visualization like bar charts, a geographic map, and a correlation matrix.

Diagram

Description automatically generated

Figure 1- Implementation of flowchart

5. Data Cleaning

Our data set was rather large, it came in at about 12.4 gigabytes of data. To narrow down what to use and to make sure our information was as relevant and precise, had to clean our data quite a bit. We had quite a large set of data to begin with. This being said, we had a goal in mind on what we wanted to achieve and focus on so we laser- focused in on complaint types and location. When running some queries in the data set to find things such as, most common complaint and areas with the most complaints, we found some strange returns come up. Some of the complaint types would have abnormalities and display data as “\_!@#(&&&)” which wasn’t relevant to use. Lucky enough, these outliers with symbols and odd characters only seemed to be entered once in the data set. To get around this problem we ran a query to select all the data that had more than one occurrence and that seemed to weed out all the irrelevant data.

6. Analysis and Visualization

After we cleaned out data in Hive and prepared the tables that we were going to use for our analysis, we downloaded the necessary files from the HDFS. From here we used Excel and Tableau to perform further analysis and create data visualizations. We used these visualizations to showcase the most common types of complaints, the areas that received the most complaints, and the times of day were complaints were received more often.

**6.1 3D Map in Excel**

Our first visualization (Figure 2) is a 3D Geo-Temporal Map we created using Excel. The map features a tour of the data, which is essentially showcases the map being populated with residential noise complaints as time goes on. The tour starts on March 19, 2010 and ends November 21,2020. It covers over ten and a half years of data, filtered to show the number of residential noise complaints that were received during this time. The first layer is a heatmap that is a count of the noise complaints received. The scale used blue to denote the lower number of call and red to show higher numbers. The Heat map allows us to see the areas that are receiving the most noise complaints and look closer at the bars(Layer 2) that are coming from them.

The second and third layer of the map are bars that represent the time/hour that received the most noise complaint calls. The orange bars represent calls made during the morning hours(12AM-11AM), and the green bars represent calls made in the afternoon(12PM-11PM). The height of the bars is determined by the number of calls received during the morning or afternoon.

With this map, we are able to see that in the northwestern area/region of new york seems to receive significantly more noise complaint calls in the afternoon hours, as displayed by the green bars. We were also able to see that the western and southeastern part of newyork recieves the most calls during the morning hours of the day, showcased by the orange bars.

Map

Description automatically generated

Figure 2- Count of Residential Noise Complaints-

**6.2 Vissualisations in Tableau**

As we were querying our data set, one of the columns that we chose to focus in on, was the complaint types. The complaint type acts as a category to help separate the different calls. Using tableau, we were able to visualize our analysis of the number of calls each complaint type received in the form of a bar chart (Figure 3). Residential noise complaints were by far the complaint type that received the highest total number of calls received with just under 2.2 million calls. The second highest was heat/hot water with about 1.4 million calls. Not only does the bar chart make it clear that residential noise complaints were the biggest cause for 311 service calls, but also illustrates the gap between it and the other complaint types.

Chart, histogram

Description automatically generated

Figure 3 - Bar chart of complaint types

With the knowledge and visualization of the most common complaint types, we began to look into the agencies that were actually handing these complaints. In the Part-to-Whole visualization below (Figure 4), we are able to get a good sense of what complaint type each agency is responsible for and the responsible for and how prevalent they are within the dataset itself. The chart uses squares of various sized to show how proportionate they are to the dataset. It also clusters some of these squares together based on the agency. For example, when looking below (Figure 4) you can see the cluster of squares in the top left of the graph. The cluster represents the complaint types that Ney York City Police Department (NYCPD) respond to and uses the sizes of the squares within the cluster to show how proportional each complaint type is to the total number of complaints they responded to. Along with the size of the squares, the graph also uses a color scale to help with the visualization. The lighter the color the lower the number of 311 calls is, and darker colors mean a higher number.

Chart, treemap chart

Description automatically generated

Figure 4 Complaint types by Agency

Using this visualization above (Figure 4), we can see than the NYCP takes up about a quarter of the graph, as does the Department of Housing Preservation and Development. One conclusion we summarized from this graph, was that, together, these two agencies can account for half of the total 311 services calls in the entire data set.

In the next visualization below (Figure 6), we made a heat map of New York that displayed the number of services calls that were received in each ZIP code. It used a red to green color scale. Going from dark green to indicate the zip codes with the fewest amount of service called placed, to Dark red to show the ZIP codes that received the most calls. This visualization made it easier to see which areas of New York were receiving the most calls.

The Brooklyn borough was the darkest red area on the map, indicating that it was the area that received the most calls with a total of 389,890 service calls placed within that area. We the map we can also see that the area that more red than green, seem to be more toward the center of the city areas.

**Map

Description automatically generated**

Figure 5 – 311 service calls by ZIP

The next visualization that we decided to make after seeing the results of Figure 3, was a bar chart of the cities that were receiving the largest amount of noise complaints. Residential noise complaints were the most common complaint type received in the data set, and we wanted to track down where the majority of them were coming from.

Looking at the graph we can immediately see a large disparity between the top three city results and the rest of the cities. As we saw in the previous figure (Figure 5), Brooklyn was the city that received the most calls. With the following graph, we can see that the Bronx follows closely behind, with New York city falling behind more so. After New York, there is a very steep drop off. Going from around 475 thousand call in New York, to under 75 thousand in Staten Island.

Brooklyn, Bronx, and Ney York were the cities that accounted for the majority of the residential noise complaints that were received.

**Graphical user interface, application, table, Excel

Description automatically generated**

Figure 6 – Residential Noise complaints by City

**7. Conclusion**

With our analysis we were able to conclude the following summary:

1. Northwestern areas/regions received more noise complaints during the afternoon hours
2. Western and south eastern received the most calls during the morning hours
3. Residential noise complaints were the most reported with 2.2 million calls
4. NYCPD and the Department of Housing Preservation and Development account for half the calls in the entire data set (~13 million)
5. Areas toward the center had higher numbers of calls placed
6. Brooklyn, Bronx, and New York received the most noise complaints

In closing, our studies found that New York’s highest complaint type was related to noise, residential areas have the highest complaints in their park borough, and out of all the park boroughs Brooklyn and the Bronx had the highest numbers of residents thus making them the leaders in complaints. Its no surprise that the more people in one area the more noise complaints there would be. Bronx and Brooklyn are one of the most population dense places having just about 4 million in population as of 2020. These are some of the things we thought one should consider when moving to the Empire State of New York. That being said, New York would not be New York if it wasn’t as robust and noisy as all the stories tell.

### References

[1] Kontokosta, C., Hong, B., & Korsberg, K. (2017). Equity in 311 Reporting: Understanding Socio-Spatial Differentials in the Propensity to Complain.

[2] Mulligan, K., Cuevas, C., Grimsley, E., Chauhan, P., & Bond, E. (2019). Understanding New York City's 311 Data

[3] Zha, Y., & Veloso, M. (2014). Profiling and Prediction of Non-Emergency Calls in New York City . *AAAI*