Mining The 21st Century's Gold From NYC311

Ilham Seladji

June 22nd, 2020

I. Introduction

New York City is the most populous city in the United States, with an estimated population of more than 8 million, distributed over about 302.6 square miles (784 km²). This large population has led its government to establish one of the largest and most effective call center services in the world, **NYC311**, to address various problems that citizens might encounter daily in New York City.

NYC311 was launched in March 2003 and it fields tens of thousands of calls daily, offering information about more than 3,600 topics: *school closings, recycling rules, homeless shelters, park events, pothole repairs… etc.* The service is available 24 hours a day, 7 days a week, 365 days a year, and it has translators on call to handle some 180 different languages.

NYC311 has a clear mission: providing the public with quick and easy access to all New York City government services and information while offering the best customer service possible.

Figure 1: NYC311 Logo

II. Context

The 21st century shifted from the industrial revolution to a new economic domination of information technology. Consequently, cities are continuously adapting their government services to the digital era to cope with their problems.

One example is to give the opportunity to every citizen to make their voice heard by reporting daily problems in the easiest, fastest and most convenient way, through a platform, and make such data open for public use to encourage data-driven problem solving and decision making.

In this report, we will highlight the importance of this digitization and how publicizing digital information can help data scientists to solve some of the hardest problems in the world and make decisions that would otherwise take years of inspection and expertise.

III. Data Exploration

We will come across two different data sets in this report:

1. The NYC311 data set: This is the main data set which gives all information about complaints reported in New York City from 2003 to mid-April 2015. It initially had 9,124,937 complaints characterized by 52 features.

2. The Population data set: This gives the populations of each Borough per year, from 2010 to 2014.

IV. Data Preprocessing

In order to have neat visualizations and do an accurate analysis with the best complexity possible, we have to clean our data sets following the below steps:

1. Drop all irrelevant features from the NYC311 data set and keep the most interesting ones, which are: The *Creation* and *Closure Dates* of a complaint, the *Agency* handling it, its *Type*, *Descriptor/Subtype*, *Status* and *Borough*, and the *Latitude* and *Longitude* of the location in which it happened.
2. Remove duplicates which might falsify results and increase the spatial and temporal complexities of our script.
3. Population values of each Borough in 2010 are coming from two different sources, we will keep one value for each Borough, which is provided by the same source from which population values have been collected in the remaining years.
4. Convert Dates and Times which are given in plain text into a format which parses Dates into Dates and Times, called **POSIXct**.
5. Add a new feature to the NYC311 data set, named **Duration**, which contains the time required to process each complaint.
6. We will ultimately join the NYC311 with the Population data set. But before, we have to rename the **Geography** feature to **Borough** in the Population data set to match its name in the NYC311 data set.
7. Finally, put Borough names in upper case in the Population data set to match their format in the NYC311 data set.

What does a Complaint record look like?

In order to make the above words clearer, a sample of each data set is given below:

1. NYC311:

Table 1: Table continues below

Created.Date	Closed.Date	Agency	Complaint.Type
2015-04-14 02:14:40	2015-04-14 03:03:22	NYPD	Vending
2015-04-14 02:00:04	2015-04-14 02:47:33	NYPD	Noise - Street/Sidewalk
2015-04-14 01:52:15	2015-04-14 02:11:10	NYPD	Noise - Street/Sidewalk

Created.Date	Closed.Date	Agency	Complaint.Type
2015-04-14 01:45:31	2015-04-14 04:03:14	NYPD	Derelict Vehicle
2015-04-14 01:44:15	2015-04-14 02:21:43	NYPD	Noise - Street/Sidewalk
2015-04-14 01:43:17	2015-04-14 02:21:41	NYPD	Illegal Parking

Descriptor	Status	Borough	Latitude	Longitude	Duration	Year
In Prohibited Area	Closed	BRONX	40.83	-73.82	2922 secs	2015
Loud Talking	Closed	MANHATTAN	40.8	-73.96	2849 secs	2015
Loud Talking	Closed	BROOKLYN	40.69	-73.96	1135 secs	2015
With License Plate	Closed	QUEENS	40.71	-73.9	8263 secs	2015
Loud Music/Party	Closed	BROOKLYN	40.67	-73.95	2248 secs	2015
Blocked Hydrant	Closed	QUEENS	40.7	-73.89	2304 secs	2015

2. Population:

| Borough	Year	Population
BRONX	2014	1431538
BROOKLYN	2014	2602680
MANHATTAN	2014	1630946
QUEENS	2014	2300148
STATEN ISLAND	2014	472075
BRONX	2013	1422340
BROOKLYN	2013	2588501
MANHATTAN	2013	1627806
QUEENS	2013	2288194
STATEN ISLAND	2013	471898

V. Exploration

First, let's explore the NYC311 data set on its own:

## 1. What are NYC311's complaints trends?

Complaints Trends
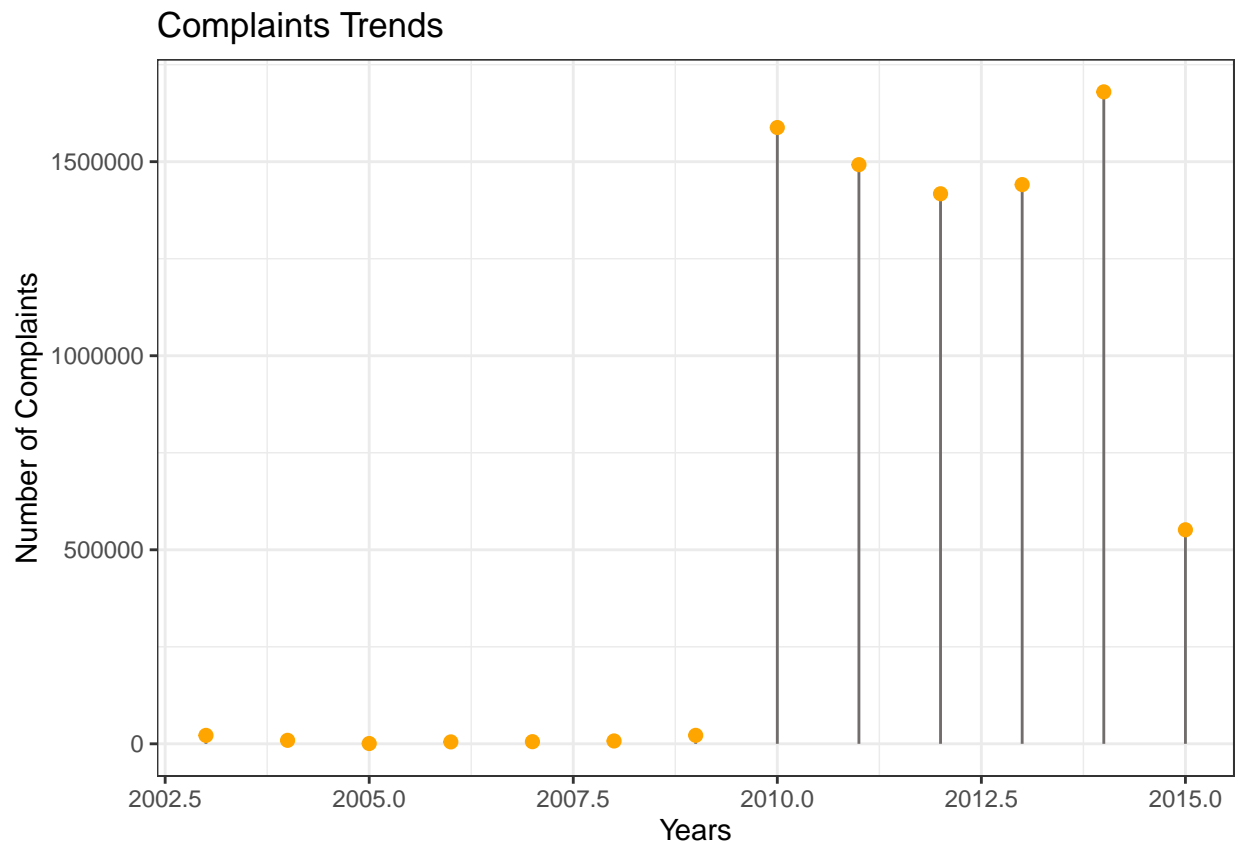
- Complaints were very rare between 2003 and 2009, but they increased abruptly in 2010. The obvious reason is that NYC311 was not fully implemented until 2010. In fact, it was not until March 3rd, 2010 that the Federal Chief Information Officer of the United States announced the creation of a uniform Open 311 API that would allow for greater standardization of modern 311 systems across jurisdictions. From then, many cities started to use it to build mobile apps for the 311 service.
- The available data shows that the peak was reached in 2014. We cannot say that the number of complaints dropped drastically in 2015, since we only have complaints through mid-April 2015.
- In order to do an unbiased and more focused analysis, the NYC311 will be restricted to the most important period with the highest number of complaints (i.e. 2010 to 2014) for the rest of the analysis.

2. In which period of the year do complaints increase?

Complaints By Month
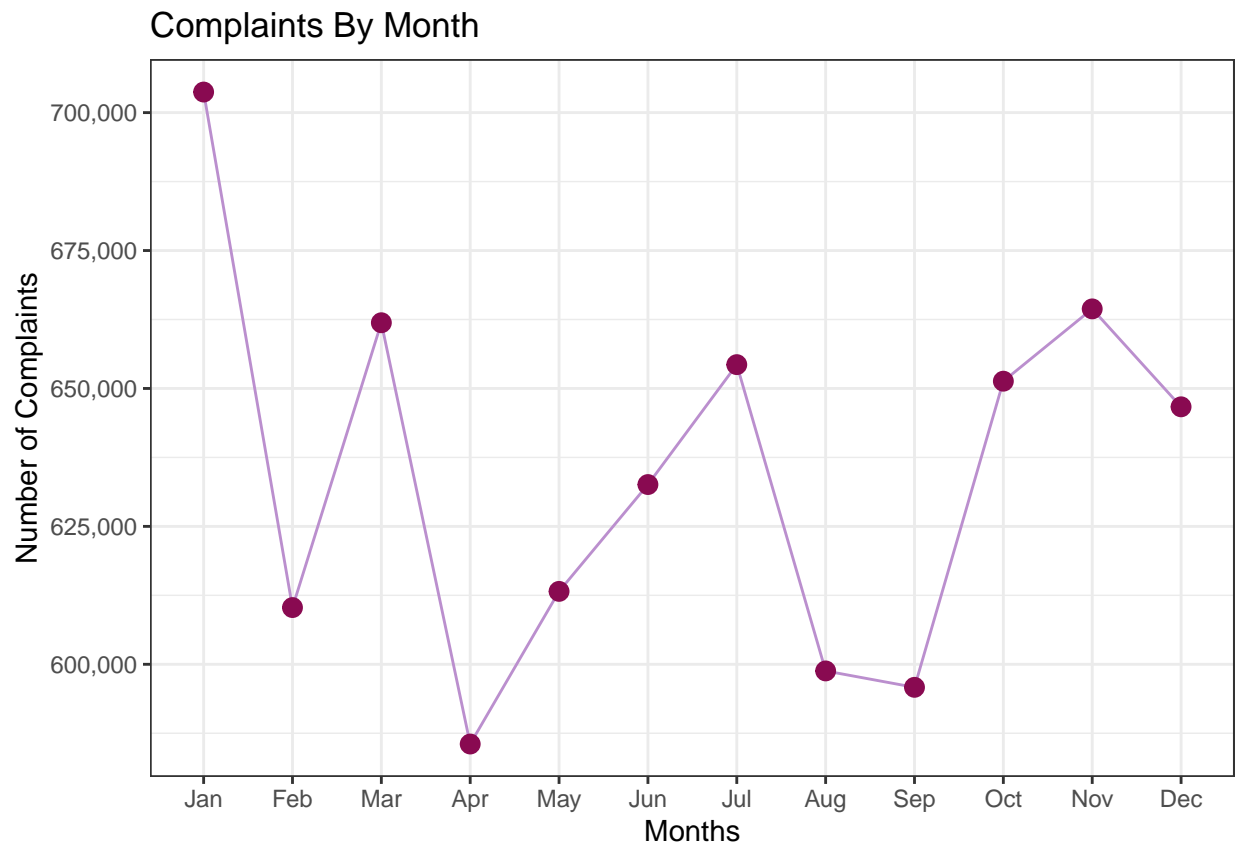
More complaints are received in **January** than in any other month of the year. But what are the most common types of complaints which are reported in January?

Most Common Types of Complaints Reported in January
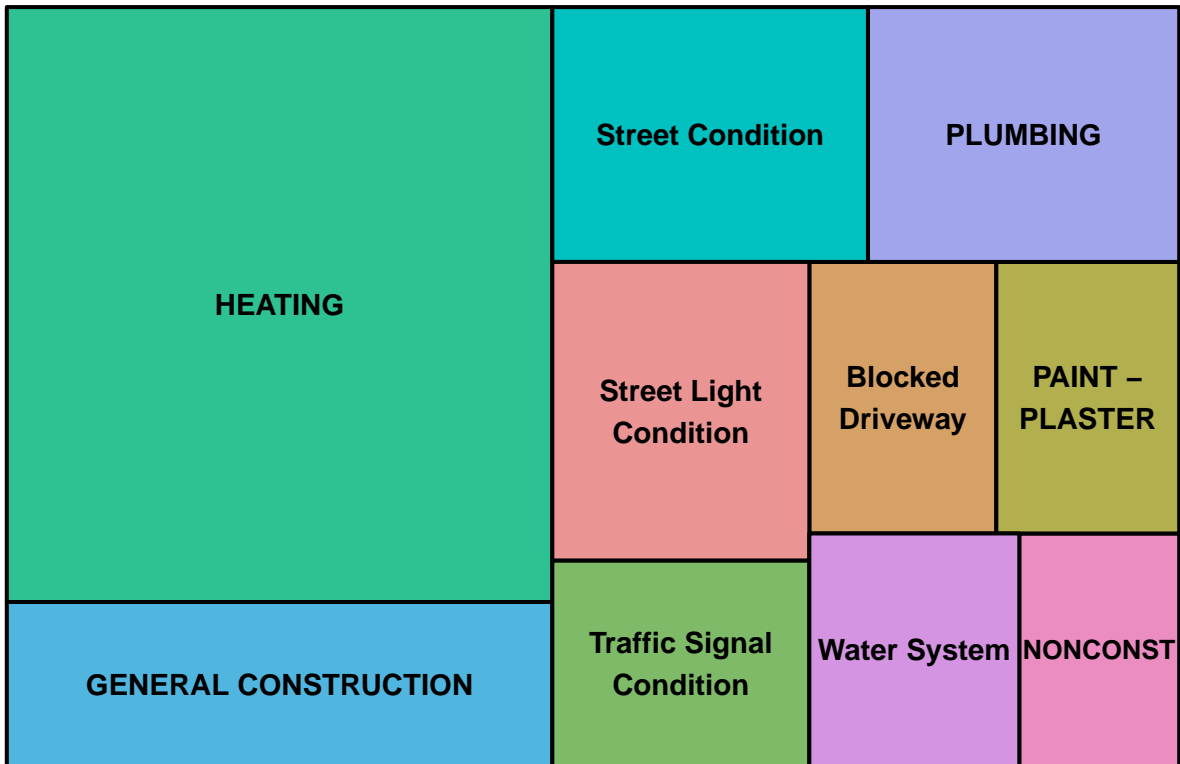
As expected, **Heating** complaints are the most common types of complaints in January. Indeed, January is the coldest month of the year in NYC with an average low temperature of 28°F and high of 39°F. Heating complaints shall increase during this month.
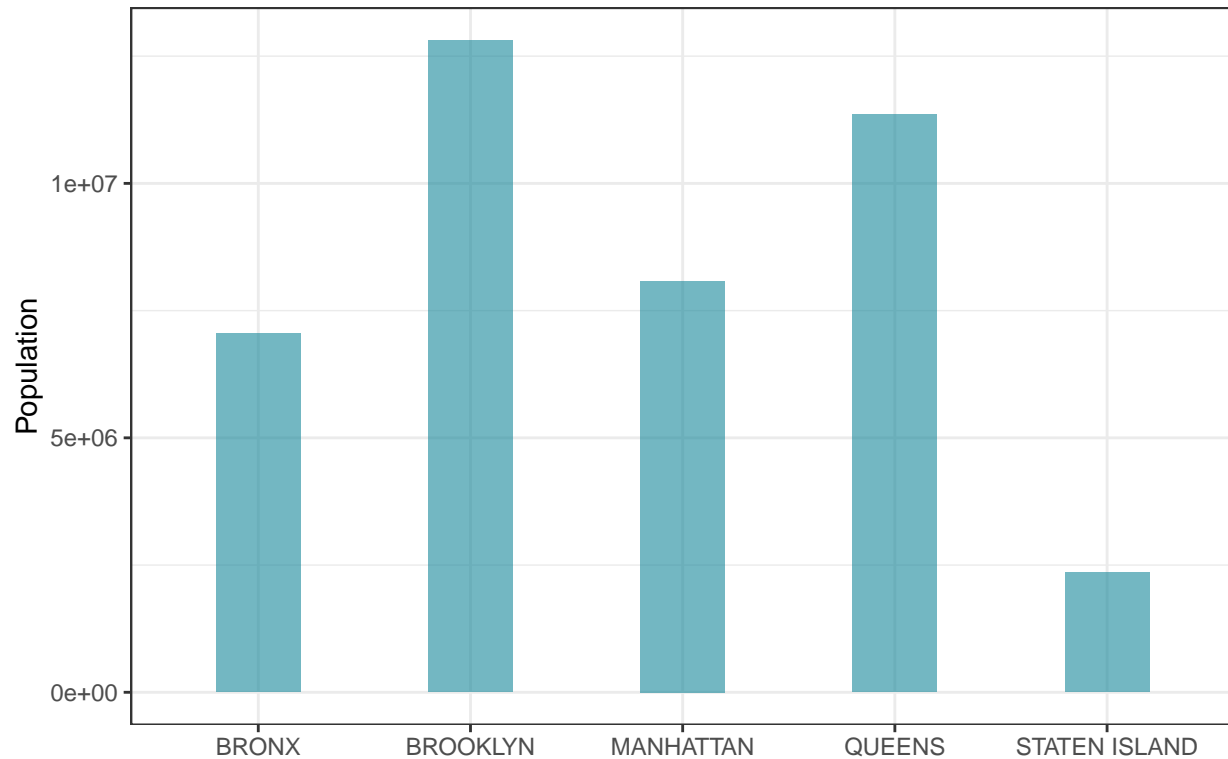
2. Which Boroughs have the highest number of complaints?
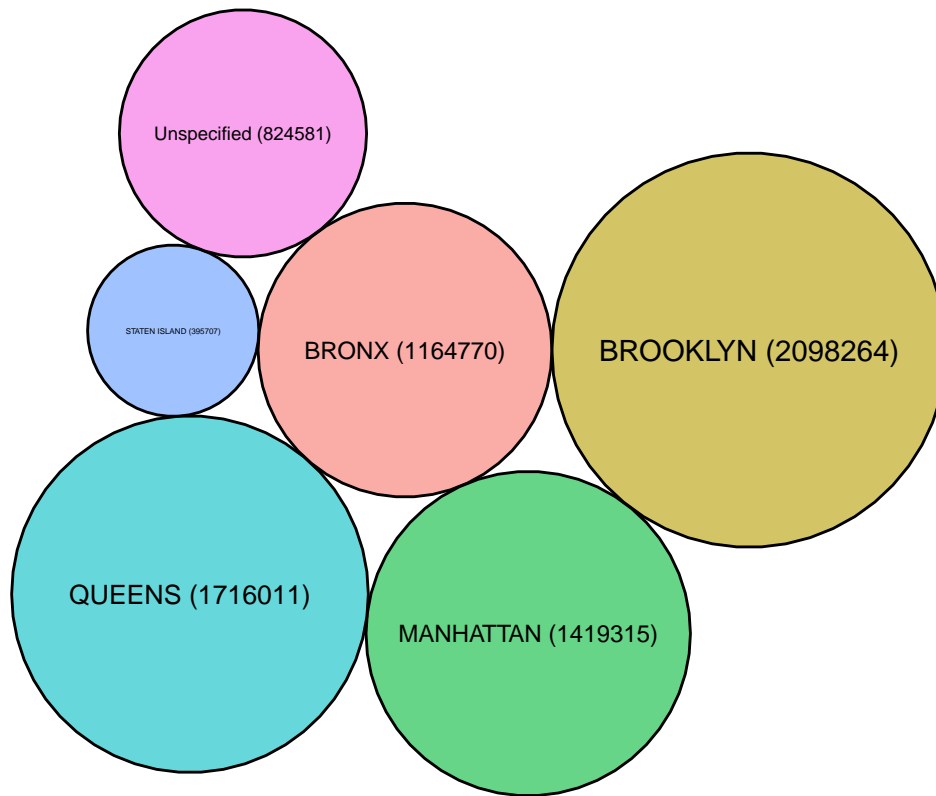
Brooklyn takes the lead when it comes to the number of complaints with 2,098,264 complaints, followed by **Queens** (1,716,011 complaints) and **Manhattan** (1,419,315 complaints).

Let us see if the population of a Borough can affect the number of complaints reported in it.

Overall Population Per Borough

As we can see, **Brooklyn**, **Queens** and **Manhattan** are the most populous Boroughs of NYC. This supports our previous finding. The higher the population, the higher the number of complaints. Generally speaking, big numbers tend to lead to more troubles.

Let's highlight this point a bit more. Suppose we have a 10-Flat residence in which a fire happened. If all 10 flats are occupied, we have twice more chances to receive complaints than if 5 flats only were occupied.

This explains why NYC311 should get prepared to work more with populous Boroughs.

But, this says that the most populous Boroughs are expected to have the highest number of complaints, and that is true. But can any random Borough have a biggest complaint rate (Number of complaints per capita) than the most populous Borough?

`summarise()` ungrouping output (override with `.groups` argument)

| Borough	Population	Complaints	Rate
BRONX	7050312	1164770	0.1652
BROOKLYN	12810756	2098264	0.1638
MANHATTAN	8079765	1419315	0.1757
QUEENS	11351391	1716011	0.1512
STATEN ISLAND	2355252	395707	0.168

Manhattan seems to have the highest complaint rate, followed by **Staten Island** and **Bronx**. **Brooklyn** and **Queens**, which are the most populous Boroughs with the highest number of complaints, happen to have a better complaint rate than other Boroughs.

A per capita study allows for a more correct analysis.

**3. Which Complaint Types are the most frequent within the data set?**

Most Frequent Complaints
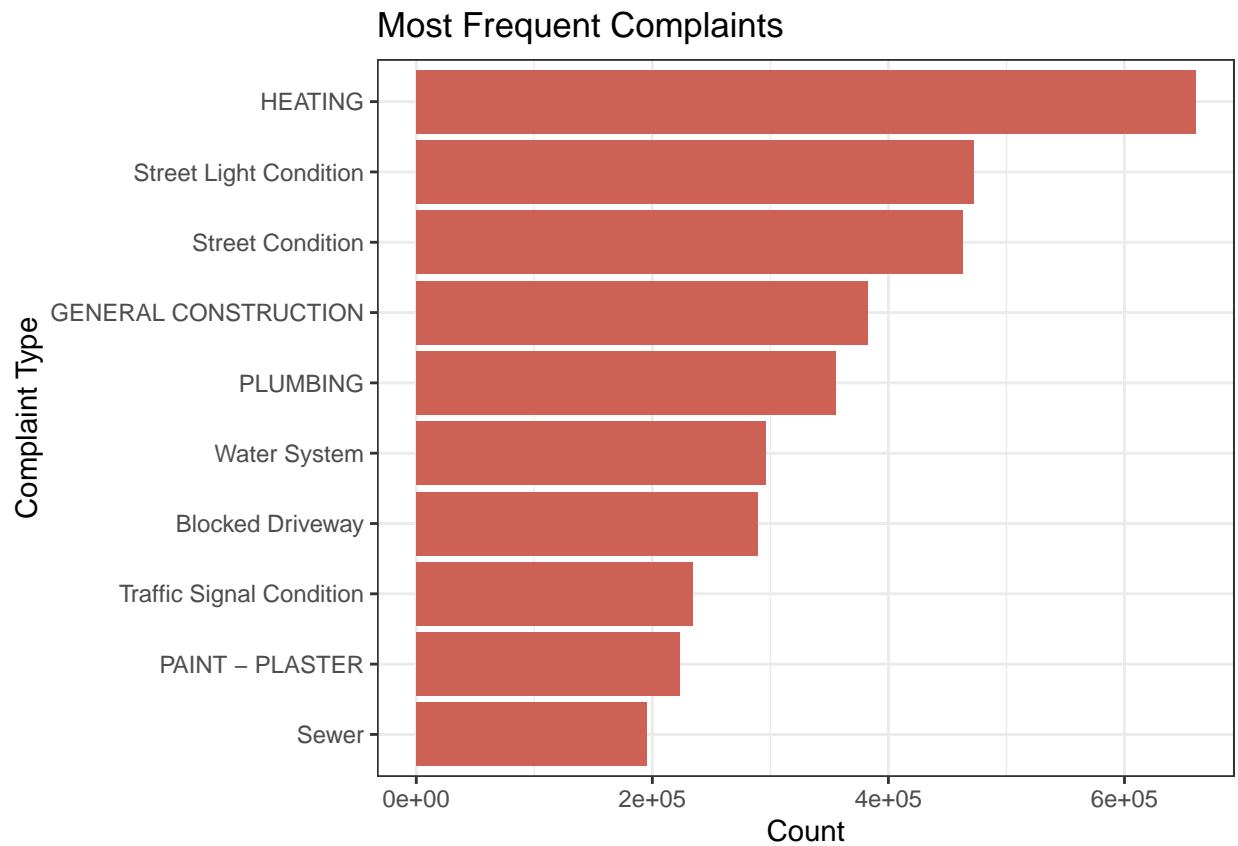
Heating, **Street Light Condition** and **Street Condition** are the three most common types of complaints. In which Boroughs are they frequent?
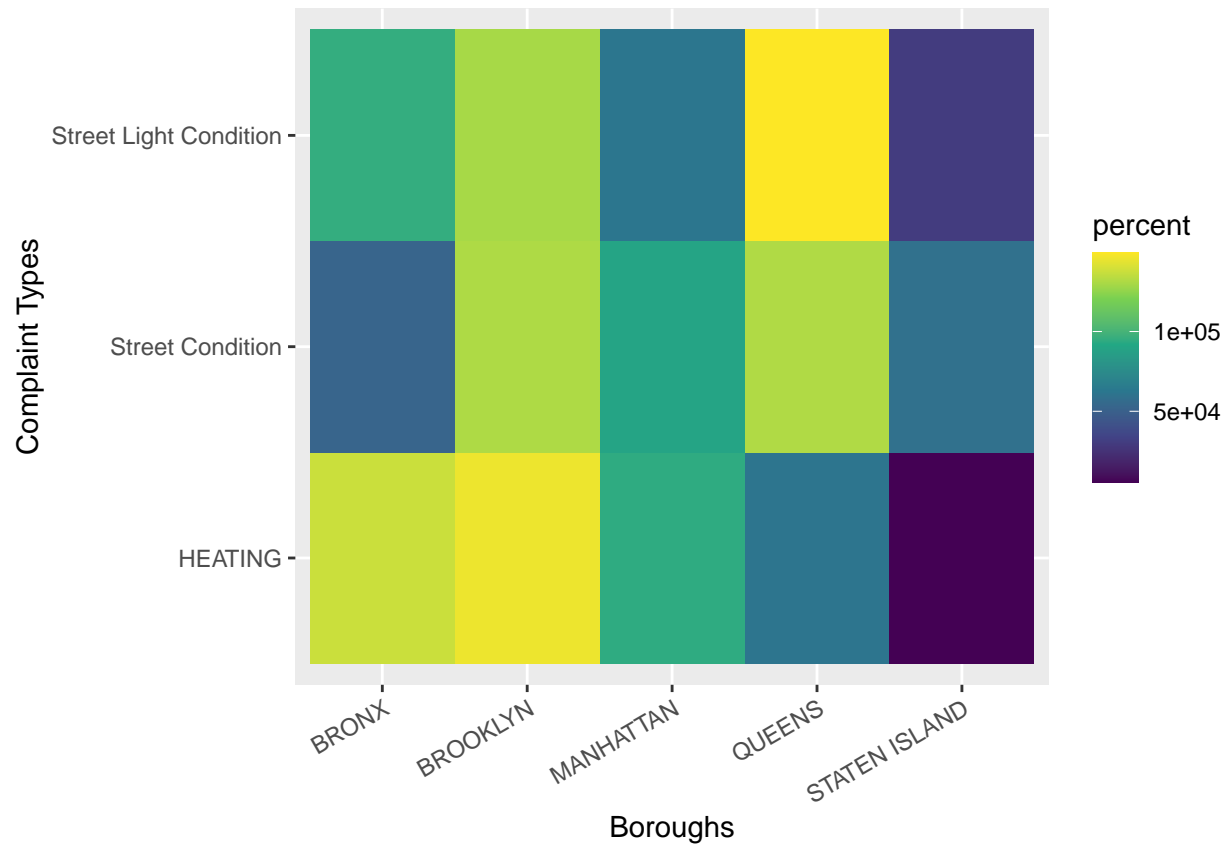
"Heating" complaints are more frequent in **Brooklyn** and *"Street Condition"* and *"Street Light Condition"* complaints are more frequent in **Queens**.

4. What are the Biggest Agencies (i.e. Agencies which have received more than 1000 complaints?)


Agencies with +1000 Complaints
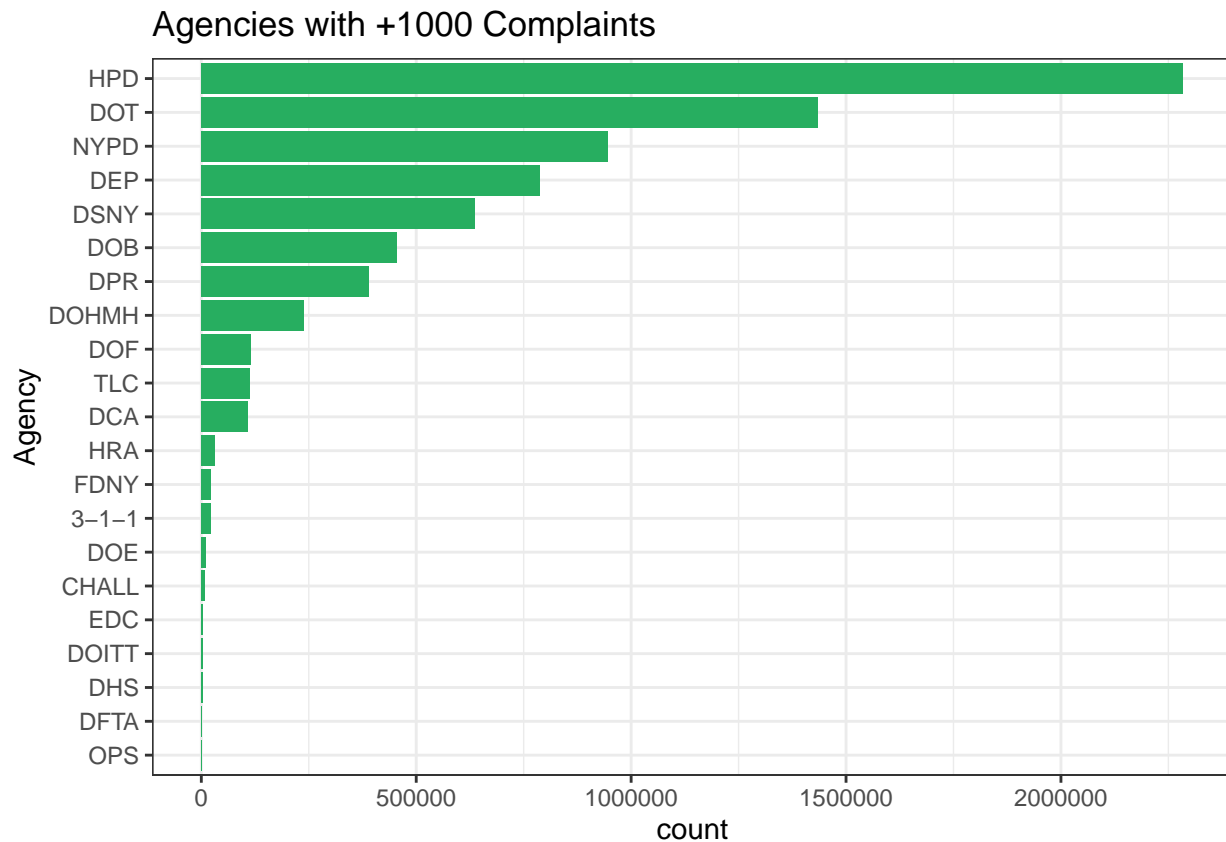
21 agencies have received more than 1000 complaints. *"New York City Department of Housing Preservation and Development (HPD)"* is the most targeted agency by complaints, followed by *"New York City Department of Transportation (DOT)"* and then *"New York City Police Department (NYPD)"* on the podium.

5. Is there a relationship between the most common types of complaints and the biggest agencies?

Let's see what kinds of complaints **HPD**, **DOT** and **NYPD** handle.

Most Common Complaint Types Handled By The 3 Giant
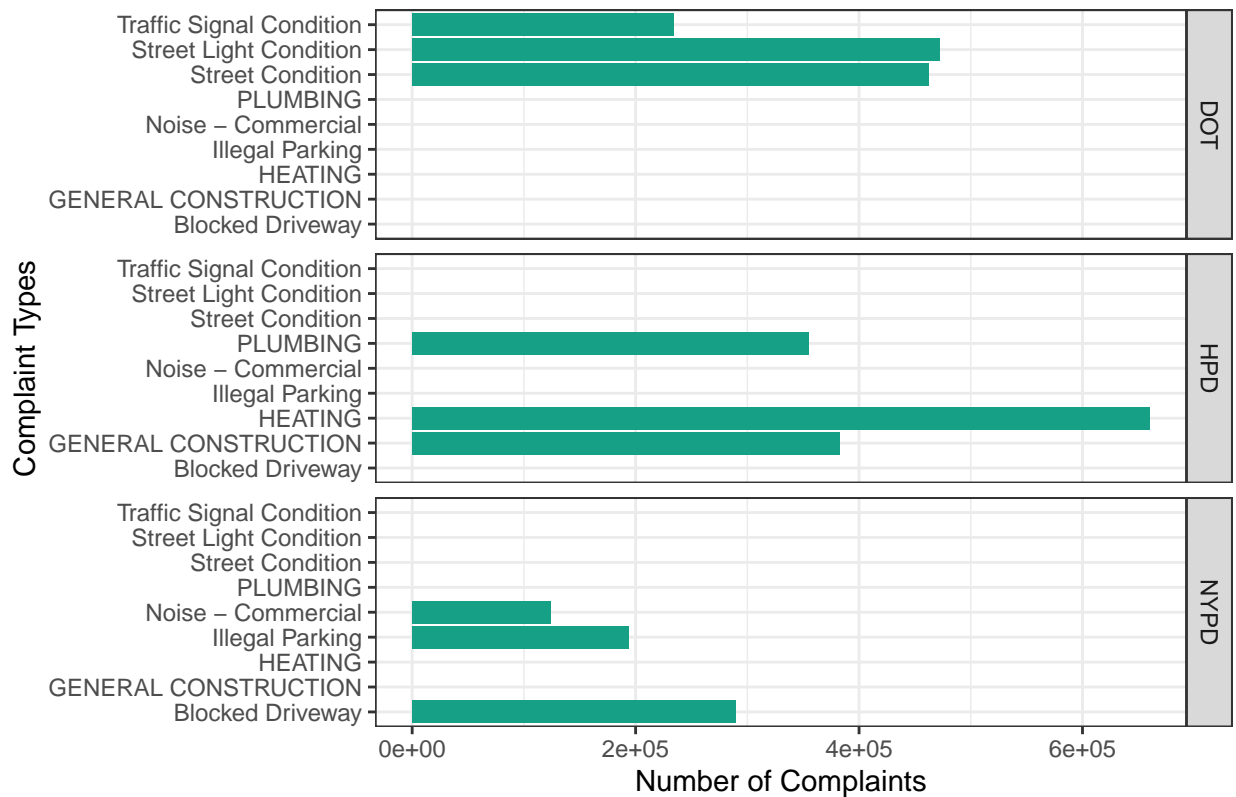
- **Heating** complaints, which are the most common types of complaints are handled by the biggest agency, **HPD**.
- **Street Light Condition** and **Street Condition** complaints, which are the second and third most common types of complaints are handled by the second biggest agency **DOT**.
- We can see that there is a relationship between agencies and complaint types. The most common complaint types are handled by the biggest agencies. That was expected. If you say that an agency receives a lot of complaints and at the same time there is a type of complaints which is more frequent than other types, then there should be a chance that the most frequent complaints are handled by the agency which deals with the biggest number of complaints.

6. How much time on average does it take the biggest agencies to process a complaint and close it?

A complaint gets closed after 15 days on average. This result does not reflect the whole performance of NYC311 since it takes different agencies and different complaint types into consideration.

Instead, let's have a look at the average processing time of each big agency separately:

Biggest Agencies' Processing Times
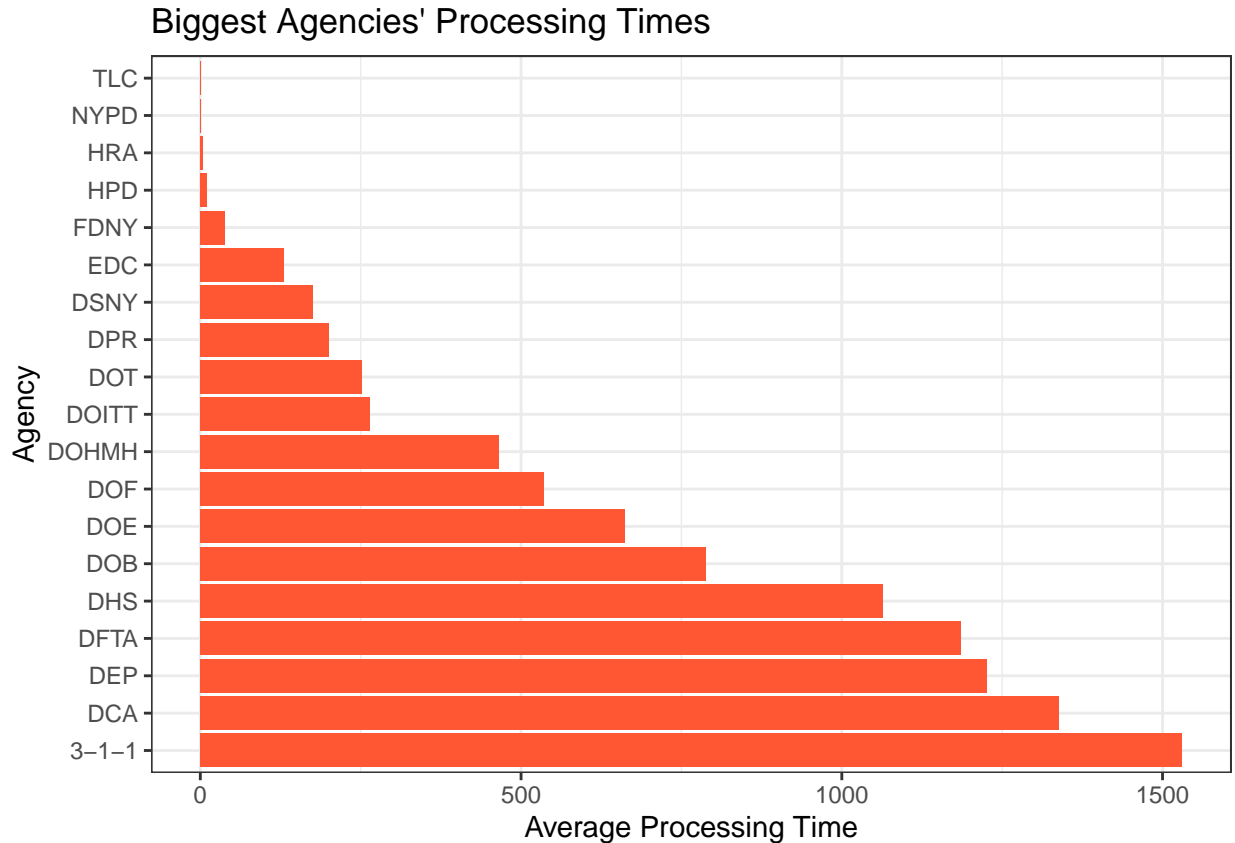
New York City Taxi and Limousine Commission (TLC), **New York City Police Department (NYPD)** and the **Human Resources Administration (HRA)** have the fastest services. Their complaints are solved almost immediately (after 45.8s, 2m12s and 3h40m respectively, on average). On another hand, **3-1-1**, **New York City Department of Consumer Affairs (DCA)** and **New York City Department of Environmental Protection (DEP)** have the slowest services amongst big agencies (it takes them approximately 64 days, 56 days and 51 days respectively on average to solve a complaint).

TLC and **NYPD** have easy and urgent cases respectively, thus, they ought to deliver very fast services. On another hand, **DCA** and **DEP** generally treat time-consuming and non-urgent cases.
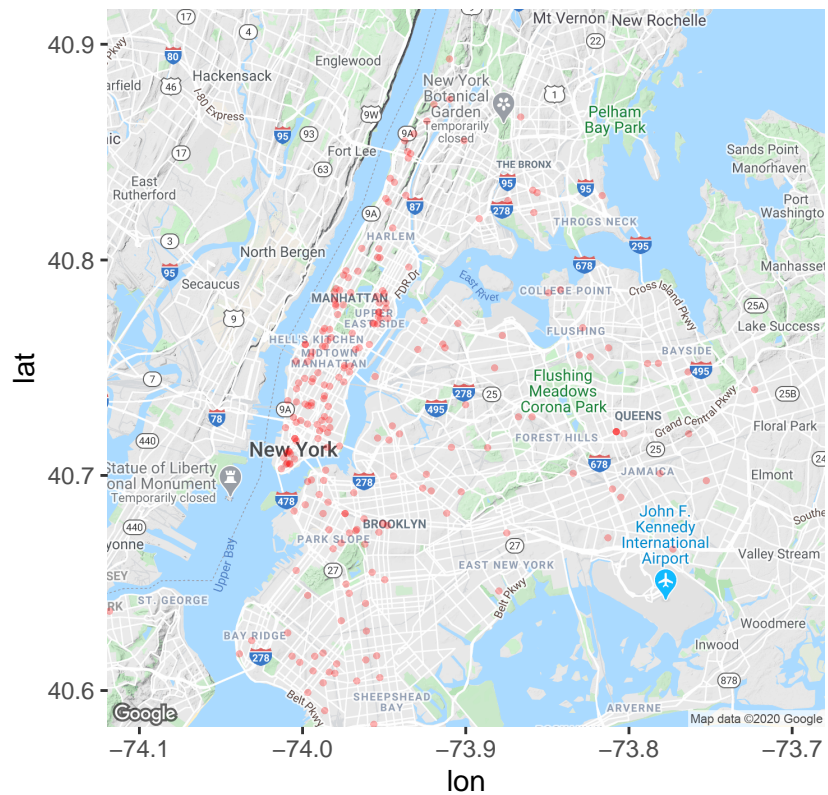
7. How are *"Noise"* complaints spread in NYC?

We are going to use a sample of 10,000 complaints in this part.

Noise Complaints by Borough

BROOKLYN
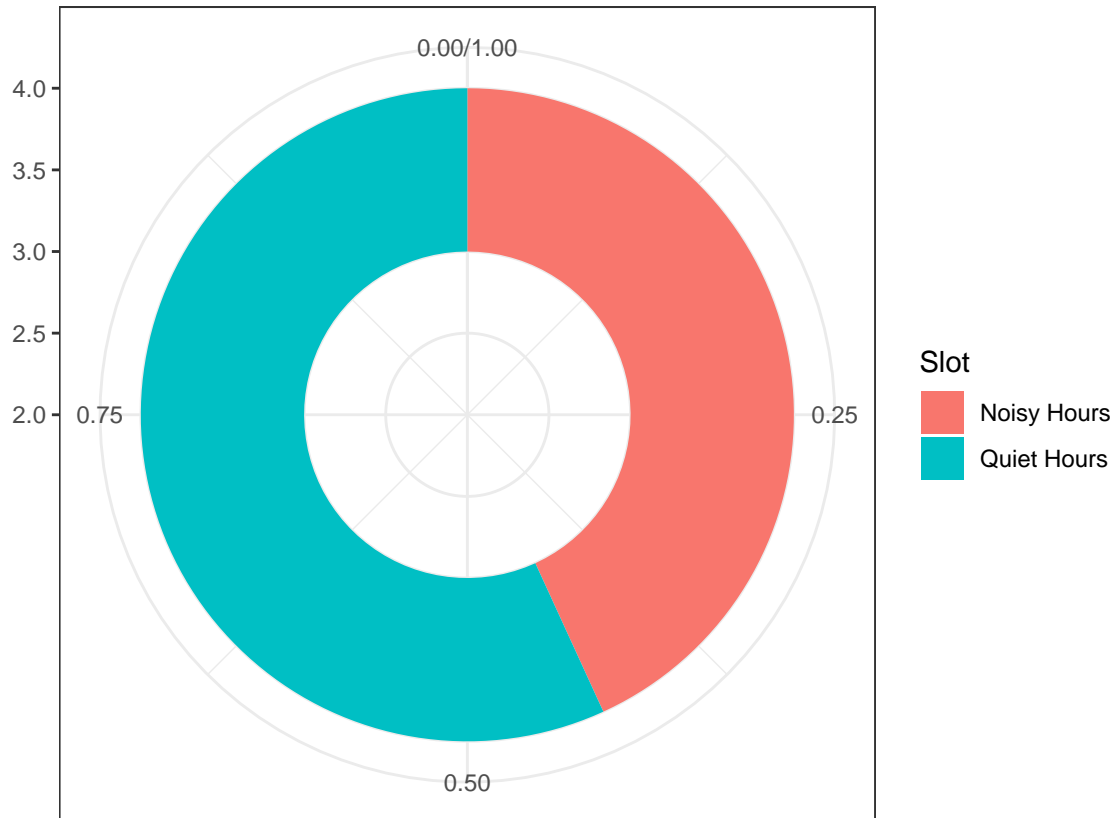
BRONX

STATEN ISLAND

QUEENS
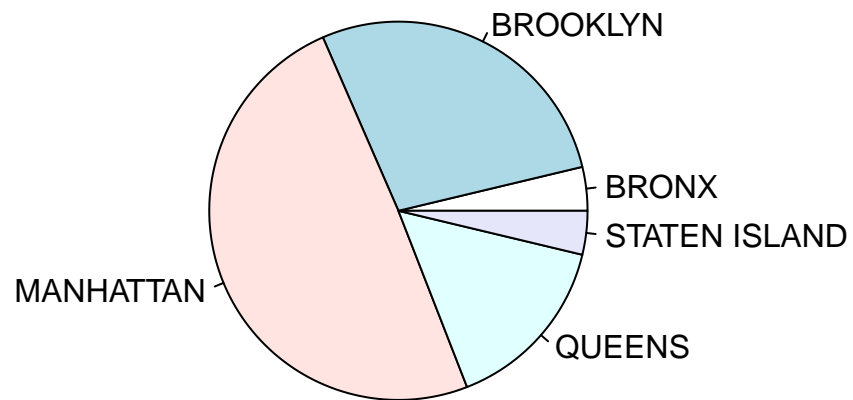
MANHATTAN

According to the above pie chart, **"Manhattan"** is the noisiest borough. The following map confirms our findings:
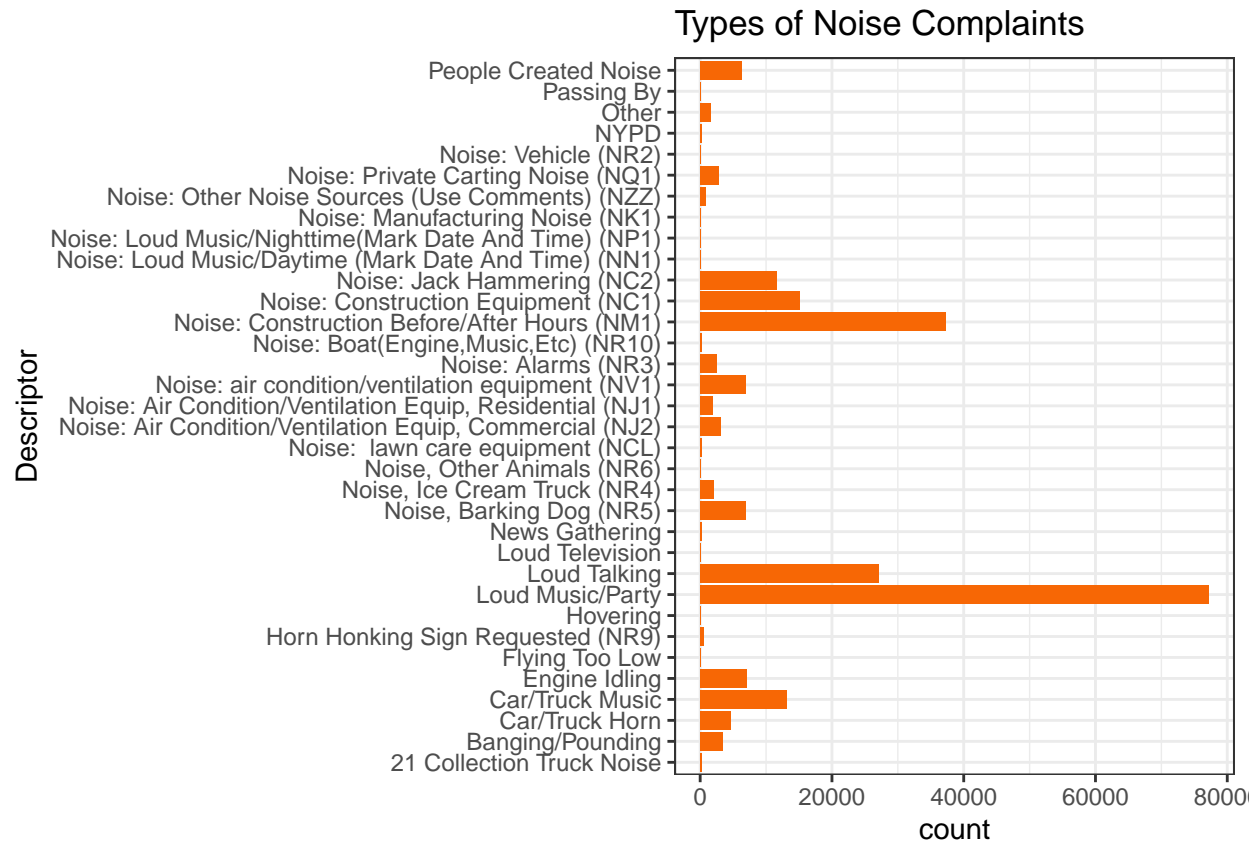
Map of Noise Complaints in NYC

As we can see, most **Noise** complaints are centered in Manhattan.

But, in which time of the day do Noise complaints increase in Manhattan?

Since the noise ordinance in NYC permits less noise between 10PM and 7AM, what we call *Quiet Hours*, then it is totally logical that noise complaints would increase during *Quiet Hours*. *Noisy Hours* receive less noise complaints, since noise is more tolerated during *Noisy Hours*, and this is confirmed in the above doughnut plot.

What is the main source of Manhattan's noise?

Types of Noise Complaints

The big majority of Manhattan's noise complaints concern *Loud Music/Party* noises, followed by *Construction Before/After Working Hours*. This supports our previous claim: *Non-Working hours* are part of *Quiet Hours*, that's why complaints increase during *Quiet Hours*.

8. Are complaint rates the same in weekdays as in weekends?

Complaints By Day
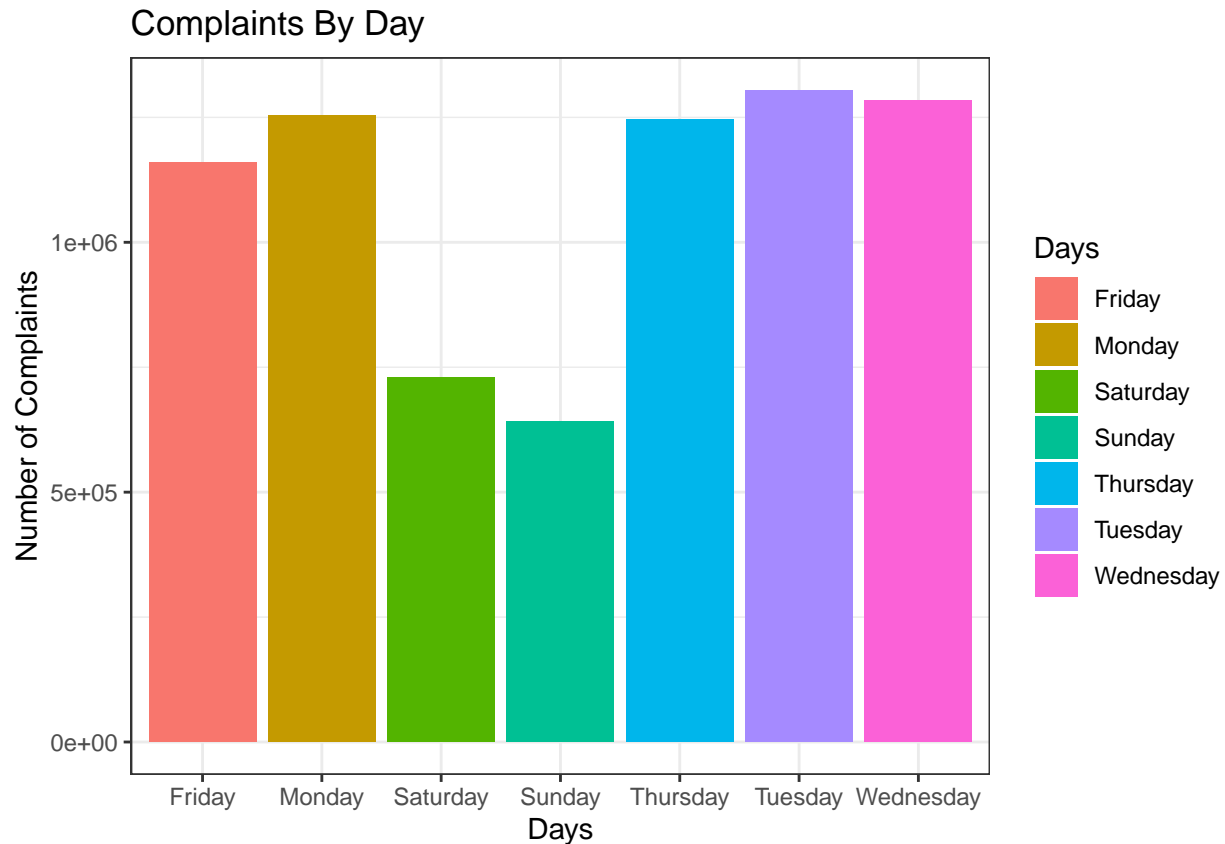
Complaints drop drastically during weekends (Saturday & Sunday). Indeed, some types of complaints such as **Transportation**, **School-related** and **Construction** complaints are not expected to be reported during weekends.

V. Conclusion

The NYC311 is a great source of knowledge. Complaints say a lot about the challenges facing Boroughs, the most well-performing agencies, the behavior of citizens in different months of the year and different days of the week and endless other insights that we can draw from different perspectives.

For instance, our analysis has led us to the following points:

1. More data leads to a better analysis. Indeed, we have analyzed complaints which happened between 2010 and 2014 when we discovered that, due to the partial implementation of the 311 service, some years were missing a lot of data, which might lead to false conclusions.
2. Some kinds of complaints reach the peak in some periods and then become less famous the rest of the time. These kinds of complaints, which are specific to a certain period, should be expected and prepared for to be treated in a timely manner.
3. The most populous areas receive the biggest number of complaints. However, they do not necessarily have the biggest complaint rate.
4. The biggest agencies in terms of the number of resolved complaints do not necessarily reflect their efficiency, as the average processing time of a complaint should be taken into consideration as well.

5. Some complaints require more time to be processed than others. The fact that an agency takes more time to solve its complaints than other agencies does not necessarily mean that it has a weaker performance.
6. We can think that the highest number of noise-related complaints are received during the noisiest hours of the day. In fact, it might be the opposite, since noise is supposed to be made during noisy hours, but it should be reported and complained about during quiet hours.
7. Noise complaints subtypes in a Borough can give an idea about whether a Borough is good to be lived in or not. For instance, quiet people may refuse to live in Manhattan.
8. Complaints increase during weekdays and decrease during weekends. More troubles are associated to work/school.

Things NYC311 should consider changing:

1. Complaints should be clustered according to their level of difficulty. Complex complaints need to be treated aside because they might affect numbers in terms of agency performance.
2. As mentioned in part 2, the same complaint can be reported by many people at once. NYC311 should detect duplicate complaints before saving them to ease the preprocessing task.
3. Tidying such a massive data set takes time and requires great processing resources. In order to ease it, NYC311 can at least try to avoid NAs by using several means such as reverse Geocoding to fill in missing Boroughs. Values need to be controlled before entering the system to avoid infelicities.
4. The database skeleton needs to be revised. Features which characterize 5% of the whole complaints need not stay in the database.

VII. Appendix

The following Data Dictionary defines each feature which has been used to mine information from the NYC311 data set in this project:

Field Name	Description	Type	Example
Created Date	Date Service Request was created	Date & Time	04/14/2015 02:14:40 AM
Closed Date	Date Service Request was closed by responding agency	Date & Time	04/14/2015 03:03:22 AM
Agency	Acronym of responding City Government Agency	Plain Text	NYPD
Complaint Type	This is the first level of a hierachy identifying the topic of the incident or condition. Complaint Type may have a corresponding Descriptor or may stand alone	Plain Text	Illegal Parking
Descriptor	This is associated to the Complaint Type, and provides further detail on the incident or condition. Descriptor values are dependent on the Complaint Type, and are not always required in Service Requests	Plain Text	Blocked Sidewalk
Status	Status of Service Request submitted	Plain Text	Closed
Borough	Provided by the submitter and confirmed by geovalidation. Boroughs can have five possible values: Bronx, Brooklyn, Manhattan, Queens, Staten Island	Plain Text	BROOKLYN
Latitude	Geo based Latitude of the incident location	Number	40.82573
Longitude	Geo based Longitude of the incident location	Number	-73.82111
Population	Population of the Borough where the incident happened, in the year it happened	Number	2,602,680