311 NYC Analysis Report

Table of Contents

NYC 311 and the problem stated

Complaints in New York

I want to investigate how to improve citizens lives in New York analyzing the information provided by the public organizations, like NYC 311.

I will download and study complaints information and join it with other relevant information available to be able to detect and (eventually) understand how NYC can improve public service.

Basic steps are:

- Descriptive statistics and visualizations
- Inferential statistics
- Predictive analytics / machine learning

What is NYC 311?

Based on the information of the official [web page](#):

311's mission is to provide the public with quick, easy access to all New York City government services and information while maintaining the highest possible level of customer service.

311 allows other government bodies to focus on their core missions and manage their workload efficiently. 311 also helps City agencies improve their service delivery by providing accurate and consistent data tracking and analysis of all service requests.

311 provides access to New York City government services through eight platforms: Call Center, Social Media, Mobile App, Text, Video Relay Service and TTY/text telephone.

Data used

NYC 311 data

New York City provides information freely available about the non-emergency calls and [related information](#).

From this information I planned to get enough insight to know where the complaints are more filled, which of them is the most frequent and basically see with a map how is it spread on the city.

Another important data source is PLUTO:

Extensive land use and geographic data at the tax lot level in comma–separated values (CSV) file format. The PLUTO files contain more than seventy fields derived from data maintained by city agencies.
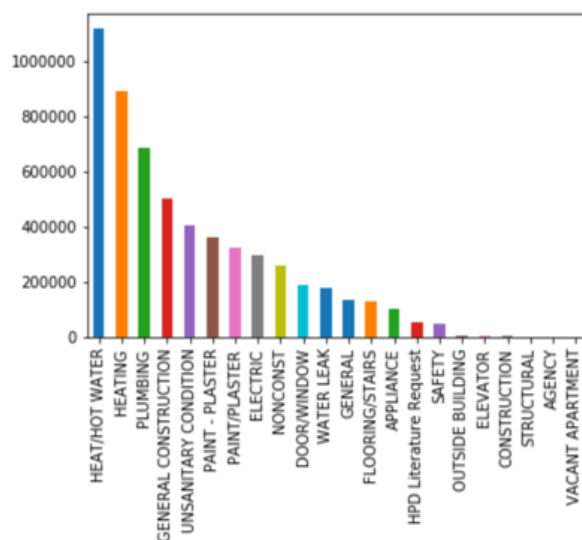
Process

Which is the most frequent complaint

This is somewhat easy, I had to just count them. Straightforward, isn't it?

Yes, it is that easy. Obviously HEAT/HOT WATER it is the most frequent, but then it comes close HEATING and PLUMBING which are related, at least functionally.

But, as the main purpose of the analysis is focusing on the most frequent, I will focus on HEAT/HOT WATER.

Having said that, any other study is very interesting… no doubt about that.

What about neighbors?

If I group by neighbor and then count them I can see which is the most problematic one.

Bronx and Brooklyn are our best customers for the following analysis. With a more detailed analysis it comes very clearly that, summing all the first incidents in the list, Bronx ends up being the winner by far.

Getting rid of the streets makes it easier:

borough	incident_zip	unique_key
BROOKLYN	11226.0	172197
BRONX	10467.0	140831
	10458.0	137696
	10453.0	130783
	10468.0	118377
	10452.0	118258
	10457.0	114950
	10456.0	106858
MANHATTAN	10031.0	100193
BROOKLYN	11225.0	97846

| borough	incident_zip	street_name	unique_key
BROOKLYN	11230.0	OCEAN AVENUE	18476
	11226.0	LINDEN BOULEVARD	18198
BRONX	10472.0	BOYNTON AVENUE	16878
MANHATTAN	10031.0	BROADWAY	15434
	10040.0	ARDEN STREET	15069
BROOKLYN	11226.0	OCEAN AVENUE	14798
BRONX	10451.0	GRAND CONCOURSE	13390
	10468.0	MORRIS AVENUE	12973
BROOKLYN	11212.0	ROCKAWAY PARKWAY	12690
BRONX	10468.0	GRAND CONCOURSE	12469
QUEENS	11373.0	ELMHURST AVENUE	12439
BRONX	10458.0	VALENTINE AVENUE	12412
	10453.0	DAVIDSON AVENUE	12279
	10452.0	GRAND CONCOURSE	11954
	10453.0	WALTON AVENUE	11938
	10456.0	SHERIDAN AVENUE	11907
	10458.0	GRAND CONCOURSE	11349
	10467.0	DECATUR AVENUE	11302
	10463.0	BAILEY AVENUE	11268
	10468.0	CRESTON AVENUE	10772

- What about a good old visualization?
- Yes, please. Let me see it plotted
- Here you go:
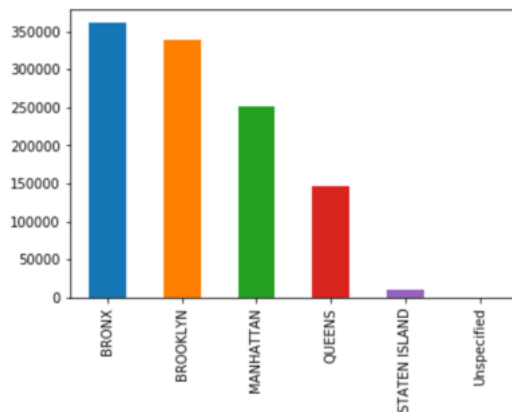
But even better than that... Is it there a way to graphically see how all the complaints are distributed in neighbors and be able to navigate zooming in an out?
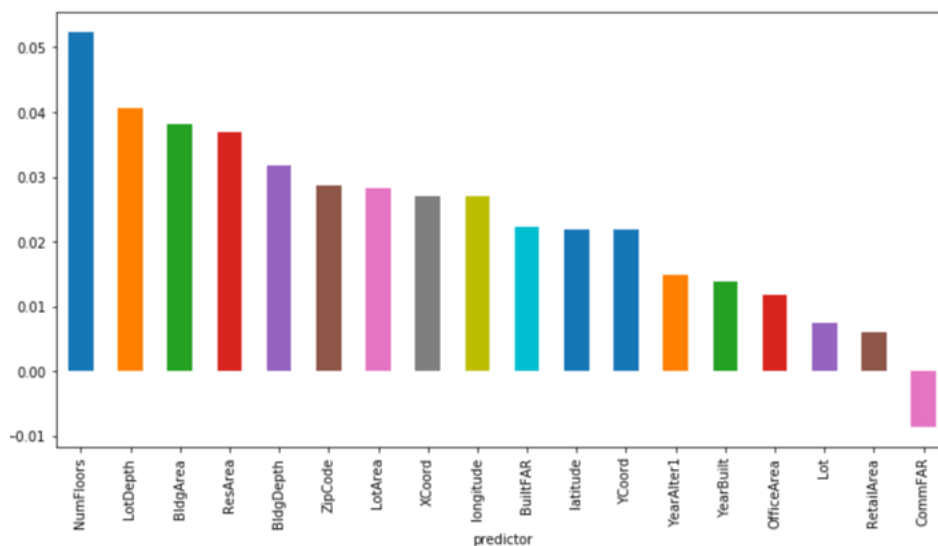
Well, yes.

It is using Folium, a great library to map things in Python.
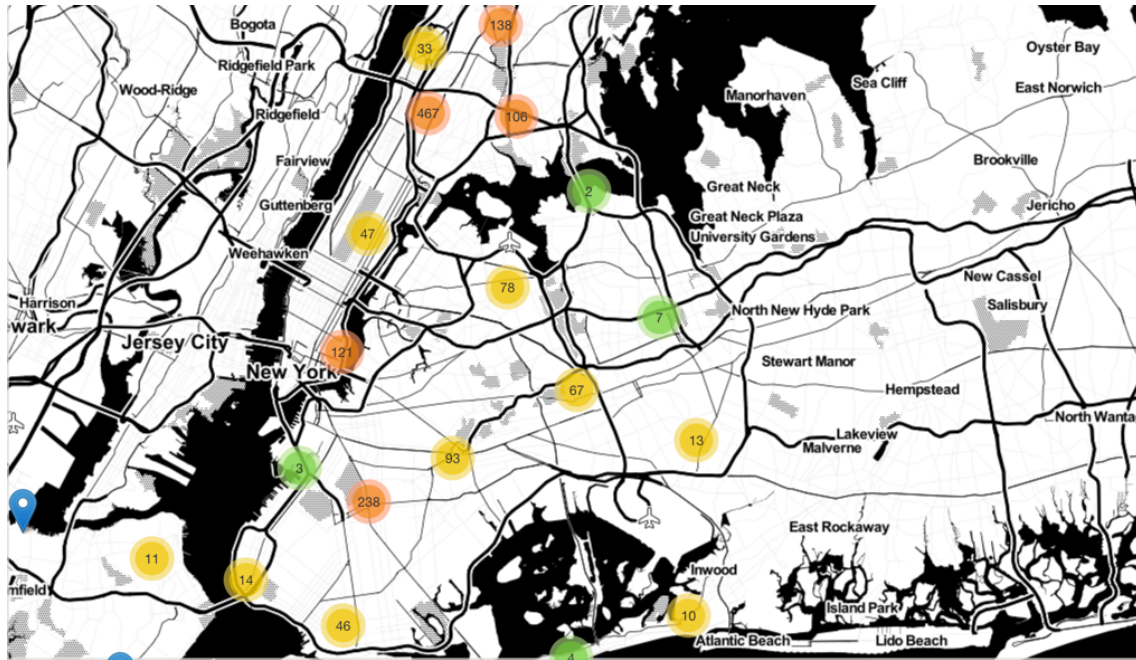
More about predictions

Calculate predictors

Getting information from the estates where the complaints were filled, we can see if there is some variable that can predict if the inhabitants are prone to have some kind of complaint. Being more specific, the one we detected as more frequent.

This means that the number of floors in the estate it is important to predict if the living being in there is going to call 311 to fill a complaint about HEAT/HOT WATER.

Does it make sense? Absolutely. But even if it does not make sense it is maths, so we should at least take care about it.

Prediction

By the end of the journey we have to finish in a peaceful port, dwelling in a knowledgeable future and being able to foresee what is in front of us. And this means being able to infer from our actual knowledge what is in front of us.

Let him who have the means deduce the foreseeable future reckon the prognosis and plunge into the stormy seas of ignorance to prevent us from falling into the illiterate behavior.

And this means xgboost.

Having said that, the result is very interesting: I was able to detect up to 93% of the complaints based on the characteristics of the estates.