

Crime Scene: NYC 2015-2018

by Sergei Suslin
April 11, 2020

Introduction

In ordinary language, a crime is an unlawful act punishable by a state or other authority. The term crime does not, in modern criminal law, have any simple and universally accepted definition. The most popular view is that crime is a category created by law; in other words, something is a crime if declared as such by the relevant and applicable law. The state has the power to restrict one's liberty for committing a crime. In modern societies, there are procedures to which investigations and trials must adhere. Law-enforcement authority which investigates the crimes is the police. The police are a constituted body of persons empowered by a state to enforce the law, to ensure the safety, health and possessions of citizens, and to prevent crime and civil disorder. Police accepts the complaints of crimes, these complaints will be a subject of this survey. The goal of this survey is to explore the crime statistics of NY City, to compare the neighborhoods by their safety and to select the most dangerous and most safe place in NY.

Estimating the safety of the neighborhood may be valuable for real-estate agencies, business owners and of course for the tourists or people who looking for property in NY.

Data Acquisition and Cleaning

Data Sources

The set of crime report headers is provided by the New York City Police Department, the Department of Information Technology and Telecommunications and the Mayor's Office of Data Analytics. Main source of data is NYPD historic dataset. It includes all valid felony, misdemeanor, and violation crimes reported to the New York City Police Department (NYPD) from 2006 to the end of 2018. Today this dataset contains the records for more than 6.8 millions of accidents.

Other source of data will be Foursquare.com. It will provide the information about business activity in the neighborhoods under survey.

Dataset will be processed using Python/pandas and presented on Python/Folium maps.

Data cleaning

This dataset contains a large quantity of data which will be excess for this work. After the brief reflection the shape of dataset was chosen as: the date of report, the coordinates of the accident, and the short description of accident. The rest of the data was omitted. The data before year 2015 was also omitted as irrelevant to our days. There was insignificant number of broken records, e.g. without the date of report and/or coordinates, these records were also discarded.

This sequence of operations gave reduced the initial data to about 1.91 million of simplified records.

Initial offense descriptions had 67 unique categories. These categories were reduced to 7 main categories by the subject of crime. This simplified categories include:

Theft: the crimes that assumes the silent stealing of property;

Violent crimes: the crimes against person by use of force, weapons and threats;

Acts of Disorder: the violations against public and administrative activities;

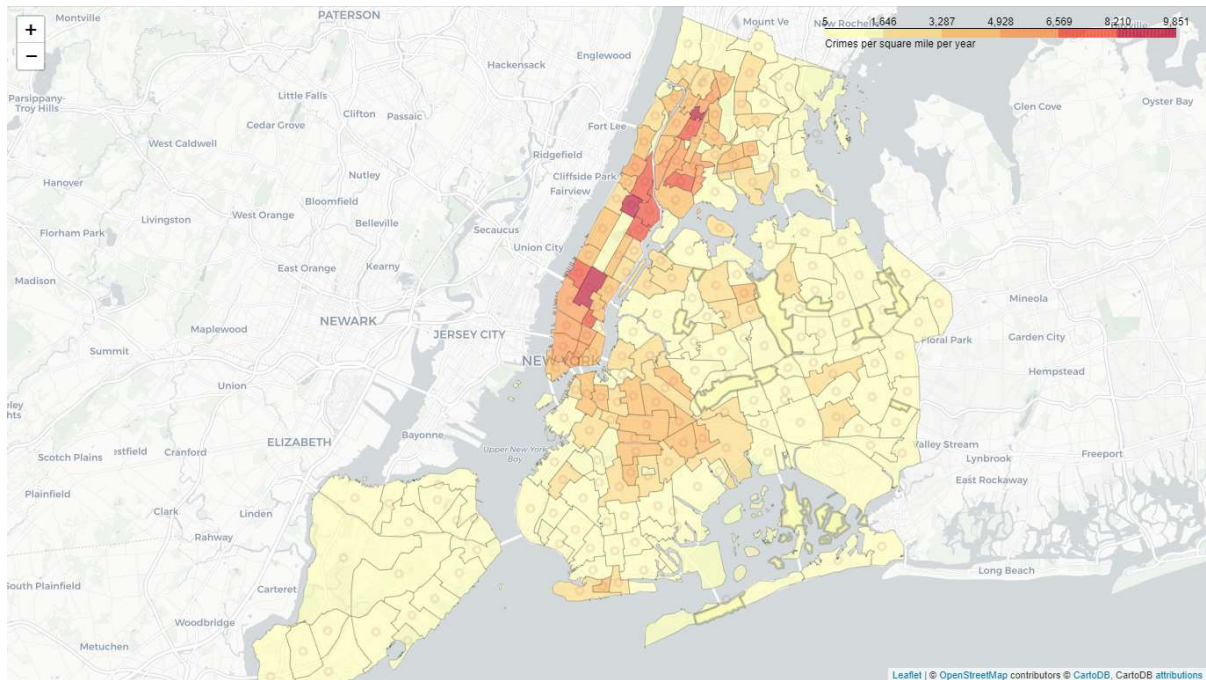
Drug-related crimes: posession of drugs, being under drugs, etc;

Sex-related: prostitution and and other non-violent;

Vehicle-related violations: traffic violations, accidents, intoxicated driving;

Miscellaneous: gambling, agriculture violations(?), fortune telling(seriously?), and others hard to classify.

Results



This map demonstrates the density of crime in New York City in 2015-2018.

The most criminal areas are:

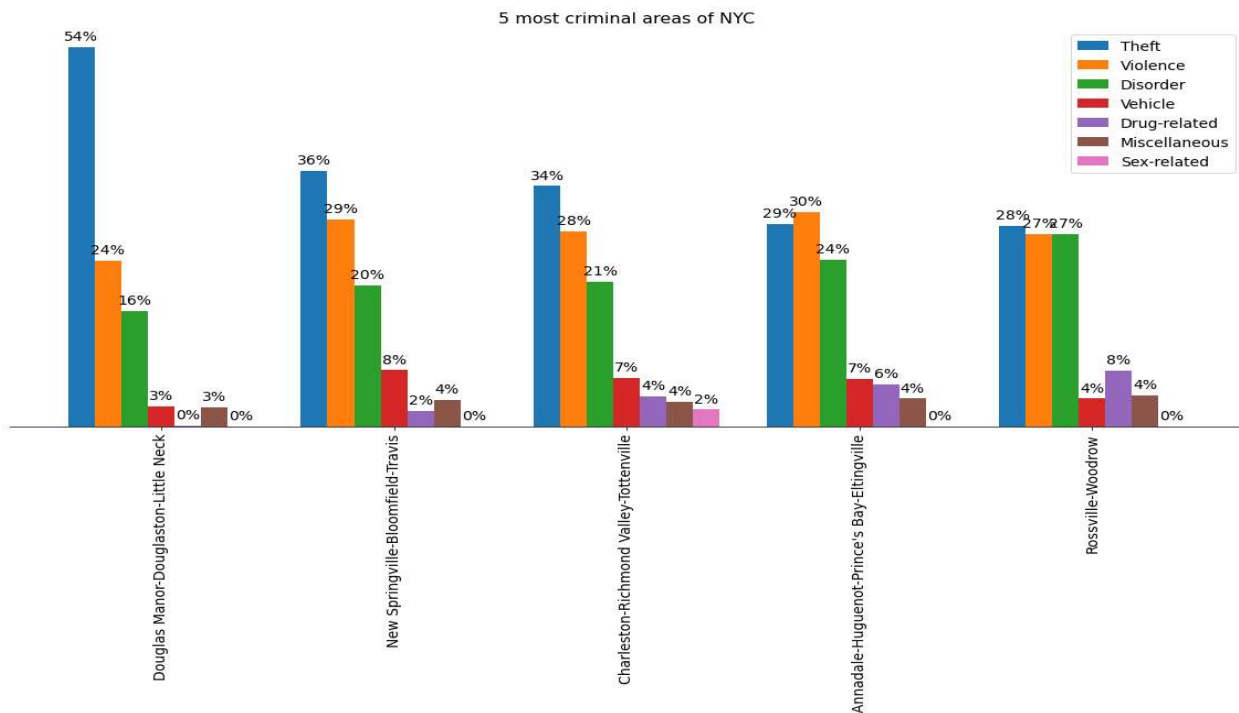
BORO	NTA	Total number of crimes	Crime density
Manhattan	Midtown-Midtown South	48021	9851.50
Manhattan	Central Harlem South	21395	9177.46
Bronx	Fordham South	8838	8676.77
Bronx	Melrose South-Mott Haven North	22065	7912.67
Manhattan	Gramercy	9040	7438.87

The safest places in the NYC:

BORO	NTA	Total number of crimes	Crime density
Staten Island	New Springville-Bloomfield-Travis	6037	114.08
Staten Island	Annadale-Huguenot-Prince's Bay-Eltingville	2825	124.04
Staten Island	Charleston-Richmond Valley-Tottenville	3971	168.83
Staten Island	Rossville-Woodrow	1772	168.94
Queens	Douglas Manor-Douglaston-Little Neck	1934	175.14

Midtown of Manhattan looks like the lair of The Evil of NY City. What are the most advanced criminal skills of all these guys?

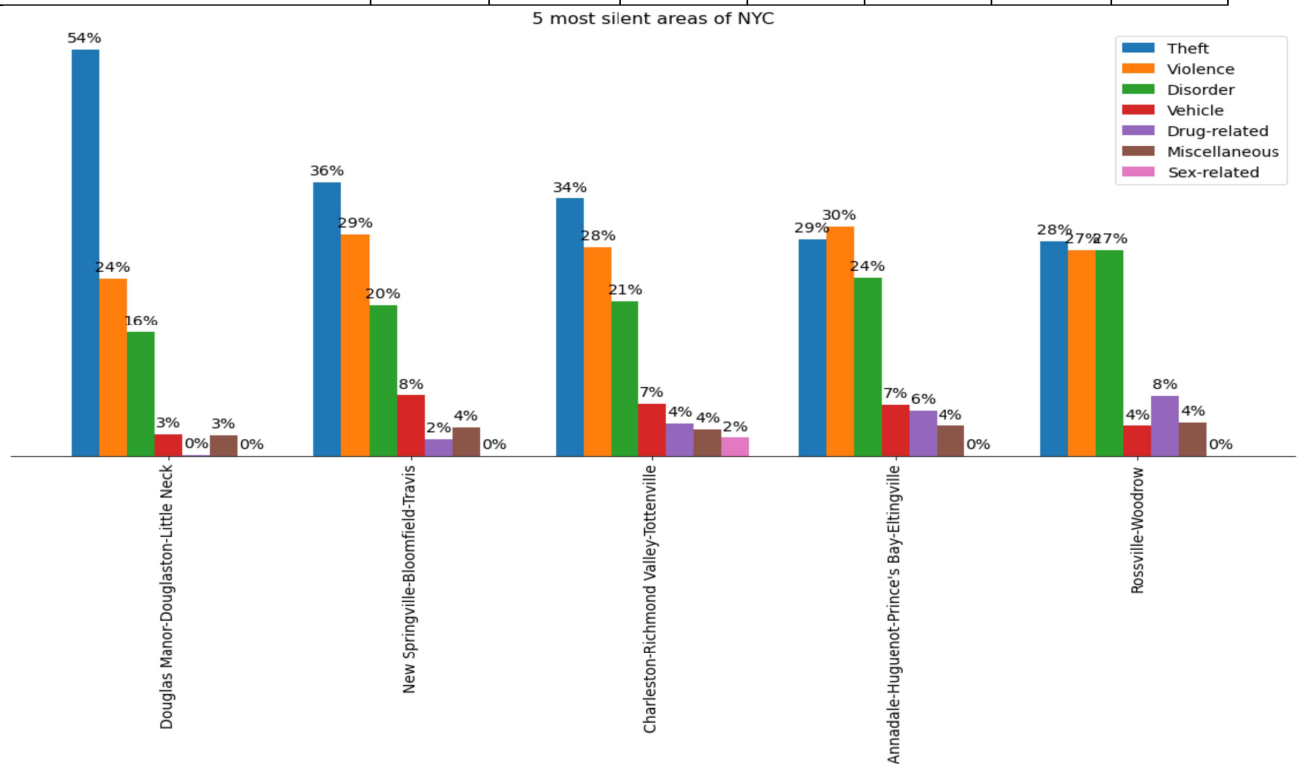
NTA	Theft	Violence	Disorder	Drug-related	Misc	Sex-related	Vehicle
Midtown-Midtown South	68.46	18.21	8.26	2.02	1.84	0.02	1.15
Melrose South - Mott Haven North	30.26	37.09	15.32	10.41	5.33	0.00	1.32
Central Harlem South	37.48	30.97	17.16	8.03	3.85	1.24	1.18
Gramercy	47.81	23.13	12.18	1.39	6.40	7.00	2.07
Fordham South	27.78	43.37	14.82	8.44	4.40	0.01	1.04



It revealed that more than 2/3 of Midtown crimes is the ordinary theft(remember, in this survey we assume the robbery is a violent crime and not a theft). Of course it doesn't look funny if somebody stole your cellphone and shoelaces, but this isn't an assault. At least you will be alive and in one piece, and your health insurance will stay untouched. Our 3rd top criminal area, Fordham South, have the similar crime rate per square mile, but the profile of Bronx guys is definitely more dangerous.

Staten Island looks like the safest place in the entire NY. Let's see what kind of crimes do they prefer to commit:

NTA	Theft	Violence	Disorder	Vehicle	Drug-related	Misc	Sex-related
Douglas Manor-Douglaston-Little Neck	53.88	23.63	16.44	2.90	0.21	2.79	0.00
New Springville-Bloomfield-Travis	36.31	29.44	20.03	8.05	2.25	3.78	0.00
Charleston-Richmond Valley-Tottenville	34.20	27.70	20.62	6.95	4.36	3.53	2.44
Annadale-Huguenot-Prince's Bay-Eltingville	28.78	30.48	23.75	6.80	6.05	4.00	0.00
Rossville-Woodrow	28.50	27.31	27.37	4.01	8.01	4.46	0.00



These thieves from Queens disrupted all the statistics! However it is clear the people of Staten Island are not the same as Manhattan guys. Traffic violations were almost invisible in Bronx, but here it is 7-8%! Does that mean all the Staten is driving like nuts? More likely these people don't committing so much of other crimes, so the vehicle violations became the significant part... of almost nothing. If the citizen of Silent Staten Island decides to go bad, he is more into make some disorder. If he knocks over the dumpster, will it be too much? And there are no sex crimes in Staten. Maybe there is no sex at all.

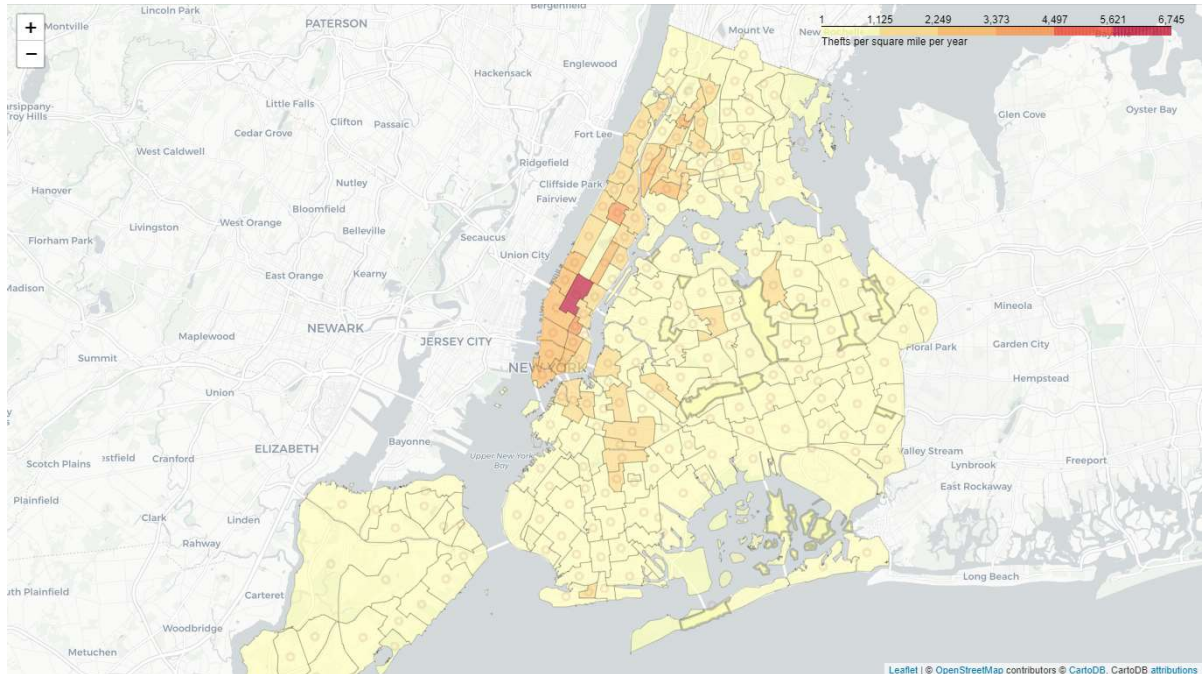
Let's check which crimes are the most common in the entire NYC.

CRIME_TYPE	Crime Count
Theft	684618
Violence	637387
Disorder	322852
Miscellaneous	98511
Drug-related	84602
Vehicle	56074
Sex-related	26861
Arson	3274

Now it's pretty clear that the 3 most common criminal cases are the Theft, the Violence, and the Disorder. Let's dive into this a little bit deeper.

Theft

The theft is the taking of another person's property or services without that person's permission

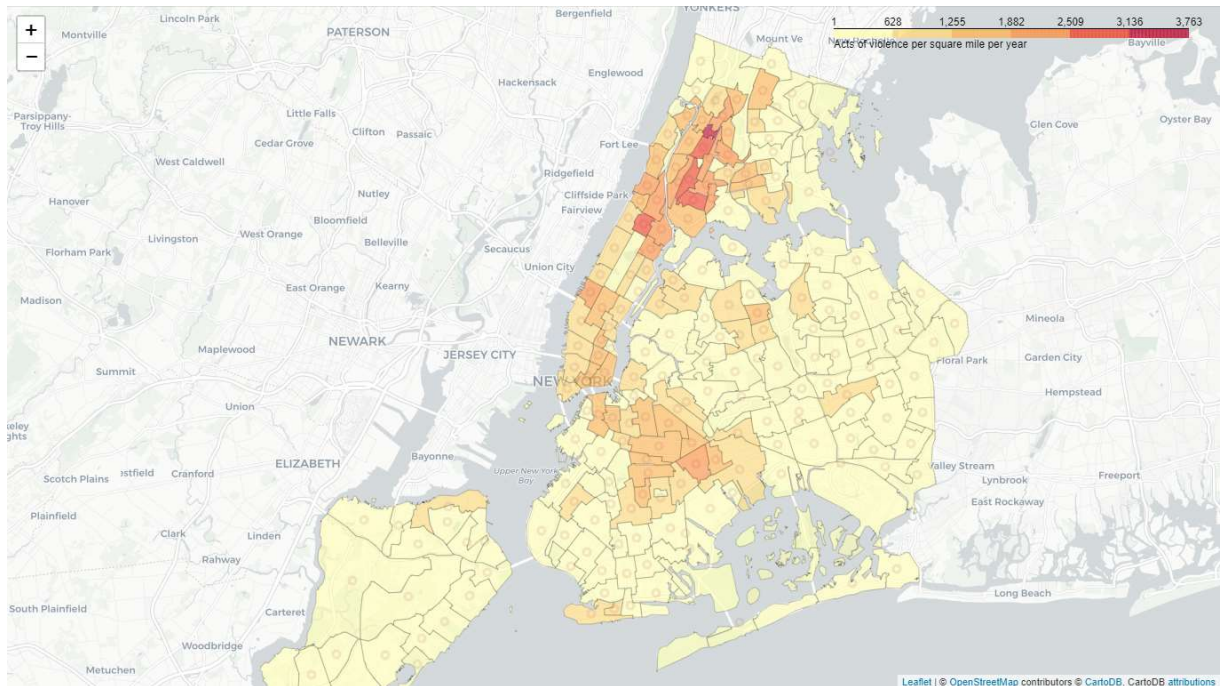


Yes. The Department of Stealh Disposal of Unnecessary Objects is headquartered in Midtown, Manhattan. As we already know. The theft is the taking of another person's property or services without that person's permission

The safest area for the property is Staten Island. 50-100 stealings per year.

Violence

A violent crime or crime of violence is a crime in which an offender or perpetrator uses or threatens to use force upon a victim. This entails both crimes in which the violent act is the objective, such as murder or rape, as well as crimes in which violence is used as a form coercion. Violent crimes may, or may not, be committed with weapons. Depending on the jurisdiction, violent crimes may vary from homicide to harassment.



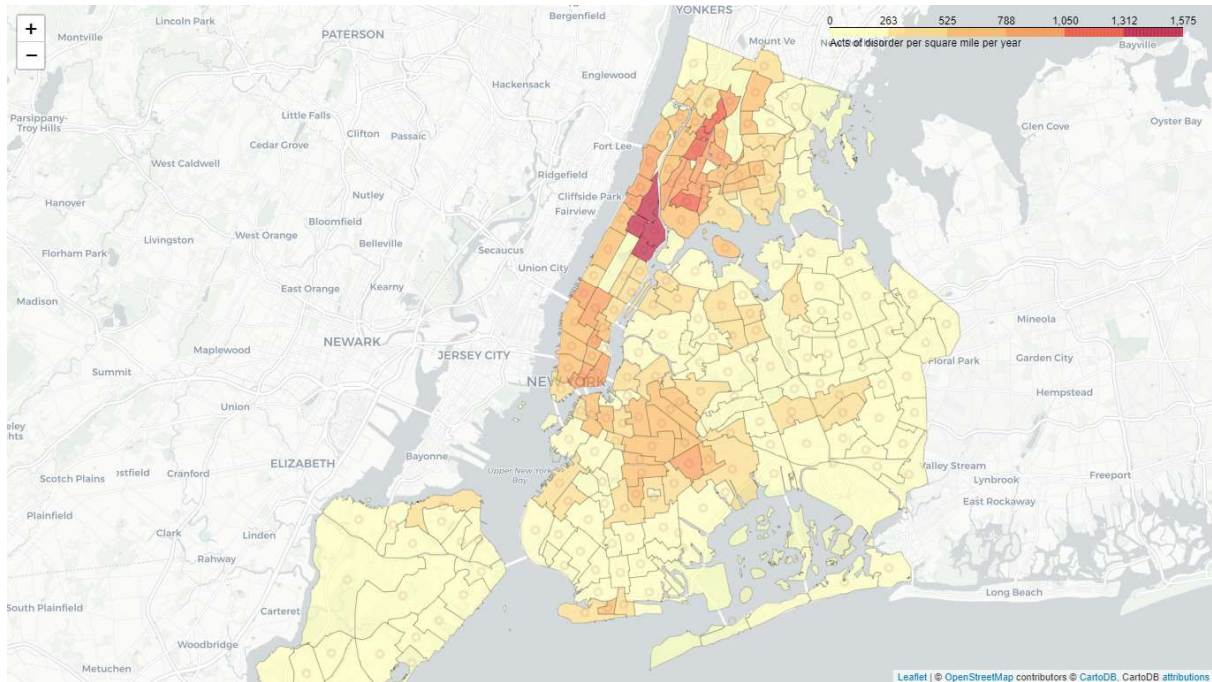
The leader in this nomination is Bronx. Fordham South is the NYC capital of everything related to violence. At least 10 acts of violence per day per square mile. You better don't wander in the heart of Bronx.

Central Harlem South is not very safe place too. That movie didn't lie.

Eastern part of Queens is the safest place in the NYC. But not as safe as Staten Island, of course.

Disorder

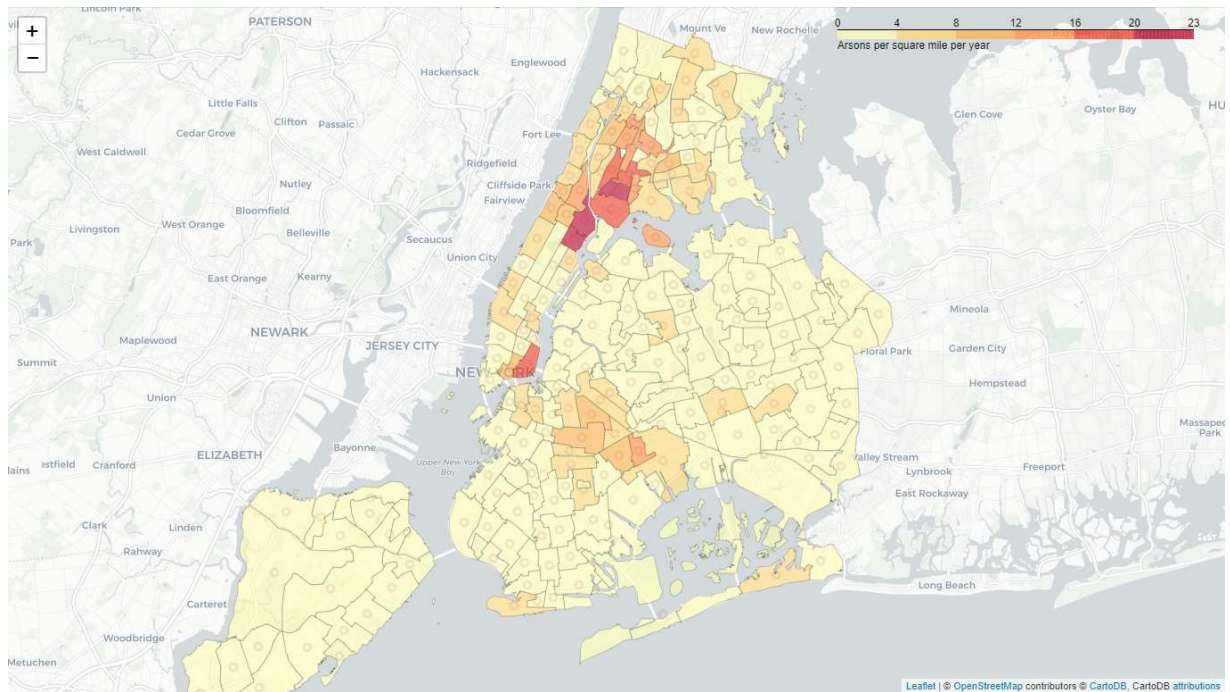
This survey will use the word "disorder" for the acts against public order, for the criminal mischief, for the offenses against public administration, etc.



Harlem is the Stronghold of NYC Disorder. There is nothing else to say. The guys from Bronx are not very into disorder. There is no reason to stick to misbehavior if you are able to be really bad. And they are, we already figured it out.

Our leader of safety, The Staten Island, produces about 50 cases of disorder per square mile per year. After all of these previous maps, it looks doubtful. Good people of Staten Island can not make such a mess. Maybe a cat on a tree is the definition of disorder on Staten Island. Anyway, the data is highly-likely overstated.

Bonus: Arson!



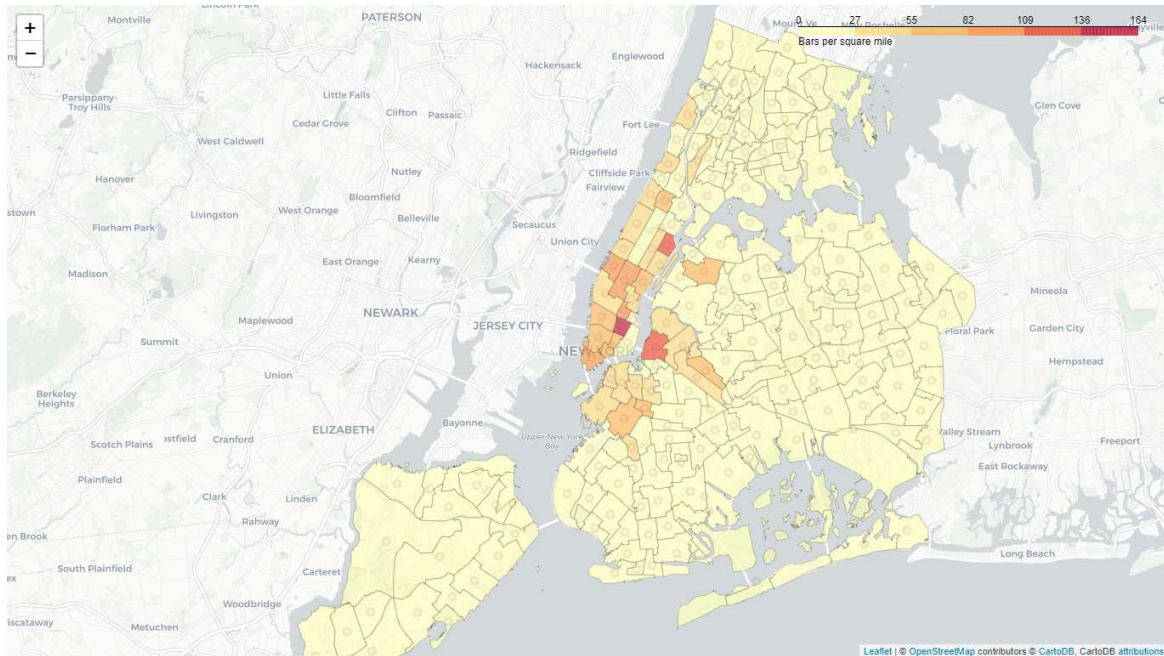
The guys from East Harlem are in competition with their colleagues from Bronx. Twenty fires per year, that's the spirit!

But look at that, who's hiding on the south of Manhattan? Lower East Side, looks interesting. Their total rate of crime is one of the lowest all over the Manhattan, however these shy boys definitely have some matches in their pockets. 17.8 cases per square mile, go Lower East Side!

Staten Island demonstrates the total lack of motivation as always. What's going on at Staten Island? The people of this boro are unable to make even one tiny little fire per year. Are they alive in there at all?

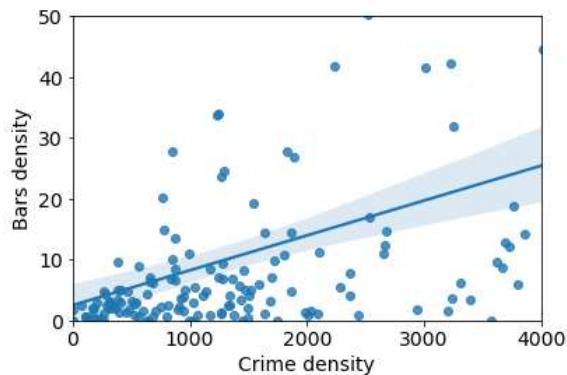
Foursquare Bar

Data of NY bars were provided by Foursquare.com and downloaded by the standalone script `fours.py` due to problems with debugging. This script is uploaded to GitHub along with this notebook.



Now it's absolutely clear why Staten Island have no crime. That's because they have no bars!

The next idea looks pretty stupid, but why not? Let's correlate!



It looks weird. There is no sufficient correlation, but picture is enough to illustrate the general dependency: the more the bars - the more the crimes. It is well known that the correlation don't means causation, that's why the "Bars density" is placed at y-axis. The bars don't causes crimes. The crimes causes the bars! If go more serious, this example was pretty unsuccessful.

This part of this work was mandatory. The author is not into public places anyway.

Clustering and criminal profiling of the NYC neighborhoods

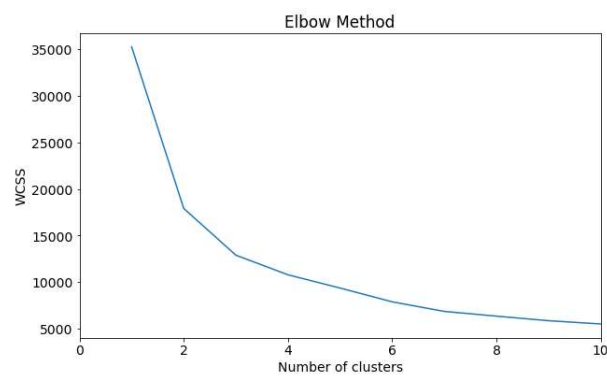
The idea is to find the similarities of the neighborhoods using their criminal profile.

After all the statistics was calculated, the dataset contains the densities of the different crime activities for each neighborhood. It is possible to segment the records using K-Means algorithm.

Densities for each neighborhood were recalculated into proportions for each type of criminal activity. The dataset looks like that:

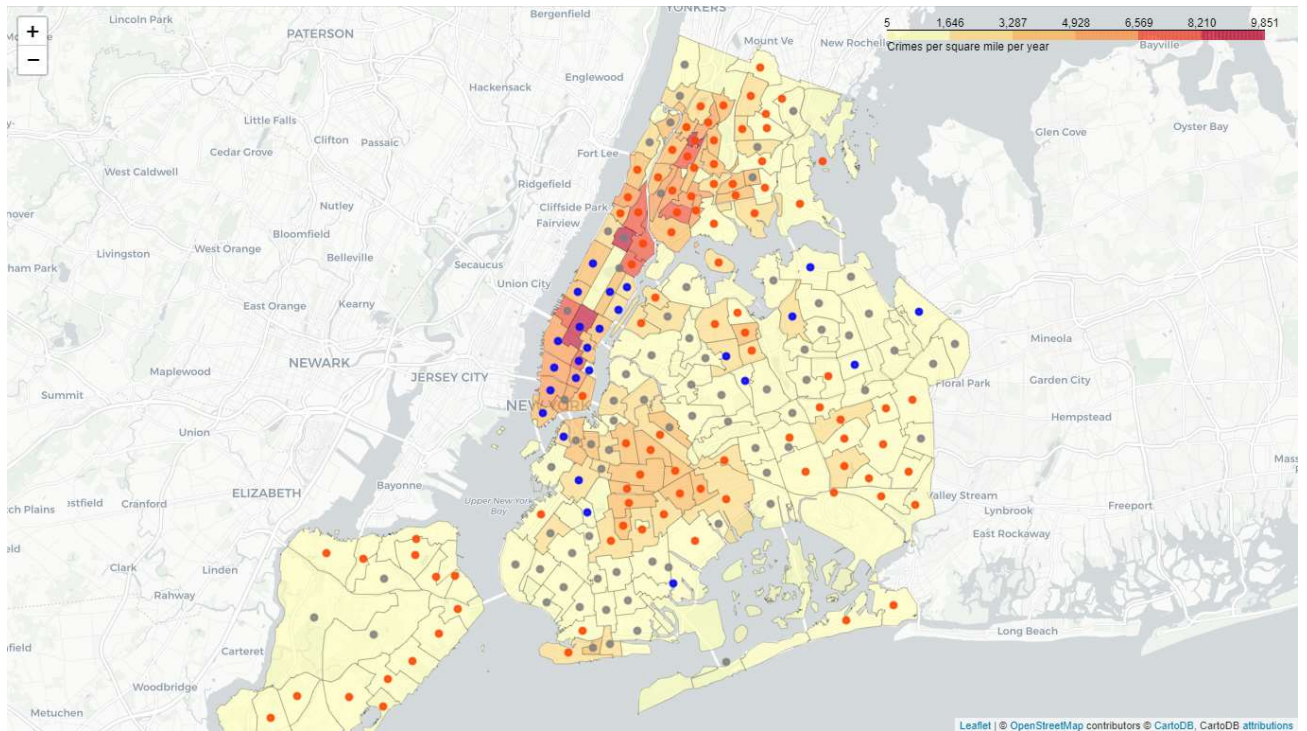
	Total	Thefts	Violence	Disorder	Arsons	Sex Crimes	Vehicle Violations	Drug Violations	Misc
0	1.000000	0.398908	0.272286	0.160703	0.001334	0.040509	0.052395	0.041601	0.032262
1	1.000000	0.373156	0.364728	0.179098	0.000843	0.000000	0.033291	0.007375	0.041509
2	1.000000	0.307081	0.386278	0.161703	0.001544	0.000221	0.063755	0.030664	0.048754
...
195	1.000000	0.280884	0.333094	0.215054	0.001673	0.021505	0.026284	0.070012	0.051493

The Elbow Method was used to find the optimal number of clusters:



It looks like the data can be segmented in 2 or 3 clusters. 3 was selected as the number of clusters.

The result of segmentation by K-Means method is the following map:



As you can see, the segmentation to 3 clusters was successful enough.

Let's describe the clusters of criminal activity in NYC.

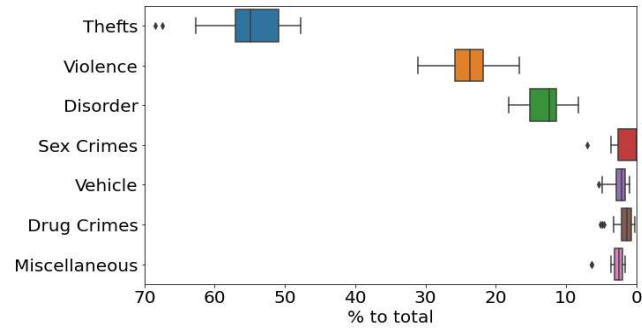
Blue dots belongs to cluster 0

Orange dots are in cluster 1

Gray dots are in cluster 2.

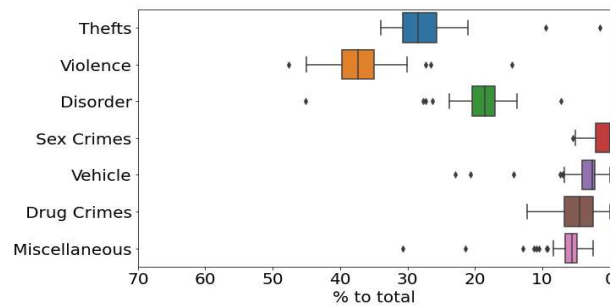
Let's explore what all this clusters are.

Cluster 0: Areas of the Thieves



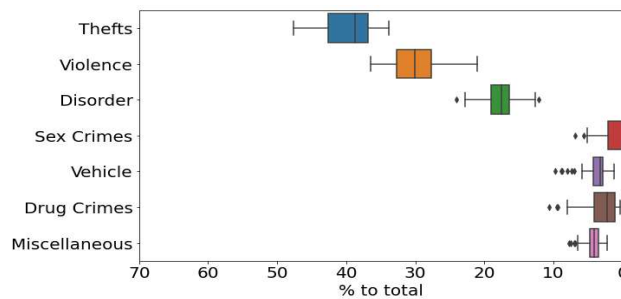
This guys definitely have a specific preference what to do. They are not very into disorder or assaults. Manhattan have the highest concentration of this kind of criminals.

Cluster 1: Areas of the Thugs



This kind of criminals prefers to perform the acts of violence in contact with the victim. Assaults, robbery, even murders. Bronx is their stronghold.

Cluster 2: There is the Regular guys



This group performs a regular set of crimes. Somewhat of violence, somewhat of stealing, this, that. But they don't stick to a specific kind of crime.

Conclusion:

The conclusion is pretty simple: if you looking for a house and want to live in safe and silent place, choose Staten Island. If you get bored you always can make a tour. The adventure is always waiting for you in NY City. The Bronx is ready to give you a good fight, and after that you can participate in some Harlem activities. But hold your wallet when going home through Manhattan. If you need some fire - look at Lower East Side. Everyone there have some matches for you. As for Brookyn and Queens, it's safer to live north of the Queens-Brooklyn border.

And remember: Staten Island looks safe and peaceful, but don't piss them off. They are more like the guys from Bronx deep inside.

References:

NYPD Complaint Data Historic: <https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i>

Neighborhood Tabulation Areas: <https://data.cityofnewyork.us/City-Government/Neighborhood-Tabulation-Areas-NTA-/cpf4-rkhq>

Pandas User Guide: https://pandas.pydata.org/docs/user_guide/

Folium documentation: <https://python-visualization.github.io/folium/>

Disclaimer:

The author is not a criminal and never been.

The author is not a law enforcement officer and never been.

The author is not an expert in data science and never been.

The author have never been in NY.

The goal of this survey is to play a little bit with some data-science tools which the author accidentally found.

The author waives any responsibility of any losses or revenues occurred as the result of proper or improper use of this "survey".