

Symposium - Spring 2020

Processing Big Data for Analytics Applications

Project Name: An Analysis of crime and complaints in NYC

Team Name: The Complainers

Team:

- Daniel Richardson
- Michael Losev

An Analysis of crime and complaints

Abstract:

Our original goal was to identify locations in New York City where crimes are reported by the NYPD in order to identify areas with unsafe crime levels and also to find areas where residents are reporting complaints via 311. In this way, we hoped to find trends between the level of crime of a location and the respective complaints brought up by its inhabitants. Defining level of crime is a nebulous task, and for our project we did so by weighting each felony crime committed by the average number of prison years for that crime.

Platform(s) where the application runs: **NYU Dumbo cluster**

An Analysis of crime and complaints

Motivation

Who are the users of this application?

The government of New York City.

Who will benefit from this application?

The NYC government benefits from this application by seeing one window into the effectiveness of both their police force and the 311 system. In this way, all those hoping to conduct analyses of the quality of NYC will benefit.

Why is this application important?

The application pinpoints areas of reform that are needed in different parts of the city and help highlight certain complaints that tend to rise with higher levels of crime.

An Analysis of crime and complaints

Goodness

What steps were taken to assess the ‘goodness’ of the analytic part of the application?

We used the average prison sentence multiplied by the amount of each respective crime in each respective area to assign a crime level to each location (precinct and latitude/longitude). In order to verify our method, we researched how dangerous some of our more extreme values for precincts actually are according to news articles and further NYPD data. Our results were overwhelmingly confirmed to be accurate.

An Analysis of crime and complaints

All Data Sources

Name: NYPD Complaint Data Current (Year To Date)

Description: Data about the crimes which take place in and around NYC, such as description of the crime, location where it took place, precinct, borough, etc.

Size of data: 150MB

Link to data: <https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Current-Year-To-Date-/5uac-w243>

Name: 311 Service Requests for 2019

Description: Data about the complaints which took place in and around NYC, such as description, location, borough, lat, long, date, etc. and how they were reported to 311.

Size of data: 1.27GB

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

An Analysis of crime and complaints

Data Sample: 311 Service Requests for 2019

unique_key	created_date	agency	agency_name	complaint_type	descriptor	borough	open_data_channel_type	latitude	longitude
41306282	01/01/2019 01:30:33 AM	NYPD	New York City Police Department	Noise - Commercial	Loud Music/Party	QUEENS	PHONE	40.70923327111128	-73.90689374176596
41306288	01/01/2019 12:21:04 AM	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	BROOKLYN	MOBILE	40.62374037922147	-73.91594318600238
41306289	01/01/2019 01:46:40 AM	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	BROOKLYN	ONLINE	40.63468819437153	-74.03191829007669
41306294	01/01/2019 12:15:26 AM	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	QUEENS	MOBILE	40.71366733255496	-73.88686040680959
41306296	01/01/2019 12:32:00 AM	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Talking	BROOKLYN	MOBILE	40.60558657019448	-73.95575378891145
41306359	01/01/2019 12:24:12 AM	NYPD	New York City Police Department	Illegal Fireworks	N/A	BROOKLYN	PHONE	40.63525769703462	-73.99098180830418
41306363	01/01/2019 01:05:16 AM	NYPD	New York City Police Department	Noise - Commercial	Loud Music/Party	QUEENS	PHONE	40.7614385543964	-73.76994132155568
41306377	01/01/2019 12:41:00 AM	NYPD	New York City Police Department	Noise - Residential	Banging/Pounding	BRONX	PHONE	40.8535846063531	-73.89437250001652
41306379	01/01/2019 12:50:09 AM	NYPD	New York City Police Department	Noise - Commercial	Loud Music/Party	QUEENS	ONLINE	40.70614246718544	-73.83176504853816
41306388	01/01/2019 01:03:18 AM	NYPD	New York City Police Department	Noise - Residential	Banging/Pounding	MANHATTAN	MOBILE	40.81558871186911	-73.95694003863639
41306389	01/01/2019 01:18:25 AM	NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	BROOKLYN	ONLINE	40.6785373240126	-73.97297834095316
41306394	01/01/2019 12:33:27 AM	NYPD	New York City Police Department	Noise - Commercial	Loud Music/Party	MANHATTAN	MOBILE	40.78649155707408	-73.97432120472271
41306408	01/01/2019 01:50:51 AM	NYPD	New York City Police Department	Noise - Residential	Banging/Pounding	QUEENS	PHONE	40.73088388676759	-73.85397195881608
41306409	01/01/2019 01:03:18 AM	NYPD	New York City Police Department	Noise - Residential	Banging/Pounding	BRONX	MOBILE	40.81609201946636	-73.88593482914274
41306410	01/01/2019 01:57:33 AM	NYPD	New York City Police Department	Noise - Residential	Banging/Pounding	MANHATTAN	ONLINE	40.79131256205503	-73.95237648622604
41306436	01/01/2019 02:11:12 AM	NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	BROOKLYN	PHONE	40.61495754058157	-73.9969420230285
41306437	01/01/2019 01:36:58 AM	NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	BROOKLYN	ONLINE	40.68508898050023	-73.98881168177468
41306438	01/01/2019 01:30:06 AM	NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	BRONX	MOBILE	40.81609201946636	-73.88593482914274
41306439	01/01/2019 01:44:59 AM	NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	MANHATTAN	ONLINE	40.79447505250334	-73.94165508212575
41306440	01/01/2019 01:47:35 AM	NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	MANHATTAN	ONLINE	40.70535931457295	-74.00860207956921

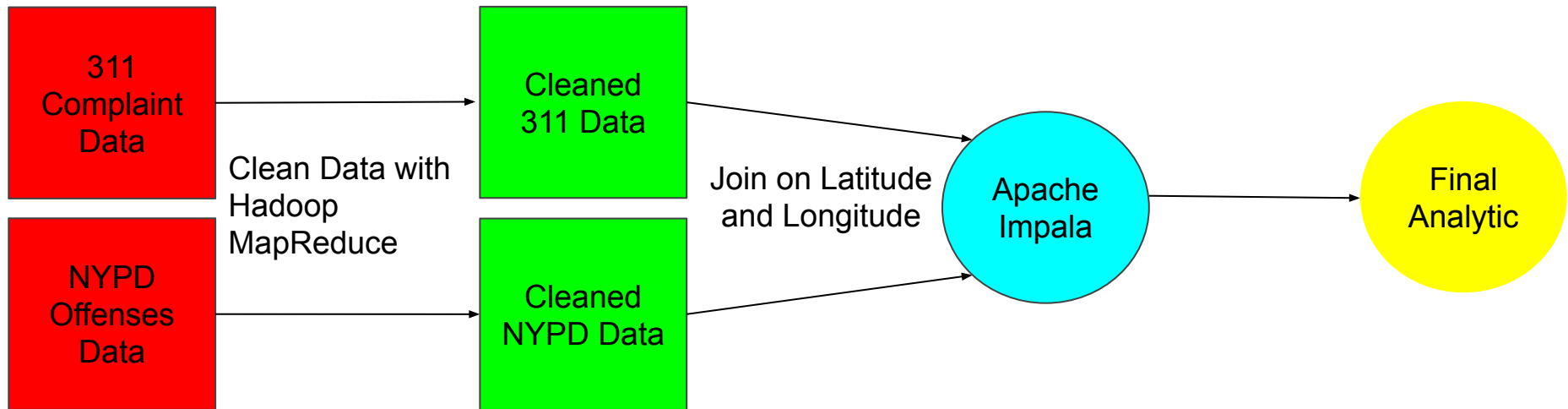
An Analysis of crime and complaints

Data Sample: NYPD Offenses Data Current (Year To Date)

cmplt_num	addr_pct_cd	borough	juris_desc	law_cat_cd	ofns_desc	reported_date	latitude	longitude
100000426	75	BROOKLYN	N.Y. POLICE DEPT	MISDEMEANOR	PETIT LARCENY	10/17/2019	40.67499474100003	-73.87481428499996
100000941	75	BROOKLYN	N.Y. POLICE DEPT	VIOLATION	HARRASSMENT 2	11/05/2019	40.66843408700004	-73.87423541899994
100001492	13	MANHATTAN	N.Y. POLICE DEPT	MISDEMEANOR	ASSAULT 3 & RELATED OFFENSES	05/20/2019	40.74041591300004	-73.97597511999999
100002195	109	QUEENS	N.Y. POLICE DEPT	MISDEMEANOR	PETIT LARCENY	11/10/2019	40.78775134300002	-73.83687595999999
100004615	25	MANHATTAN	N.Y. POLICE DEPT	MISDEMEANOR	PETIT LARCENY	01/02/2019	40.79829308000006	-73.94186988299998
100008946	112	QUEENS	N.Y. POLICE DEPT	MISDEMEANOR	UNAUTHORIZED USE OF A VEHICLE	01/09/2019	40.70901607900004	-73.85376158199995
100009823	104	QUEENS	N.Y. POLICE DEPT	VIOLATION	HARRASSMENT 2	12/26/2019	40.70565413500003	-73.90040078599996
100012751	69	BROOKLYN	N.Y. POLICE DEPT	MISDEMEANOR	CRIMINAL MISCHIEF & RELATED OF	02/26/2019	40.64281427600001	-73.89324821699995
100016271	44	BRONX	N.Y. POLICE DEPT	MISDEMEANOR	CRIMINAL MISCHIEF & RELATED OF	07/17/2019	40.84098498200007	-73.91600985799995
100017074	68	BROOKLYN	N.Y. POLICE DEPT	FELONY	RAPE	05/05/2019	40.63902167200007	-74.022463524
100019576	121	STATEN ISLAND	N.Y. POLICE DEPT	FELONY	DANGEROUS WEAPONS	05/16/2019	40.62568490900002	-74.16034600499995
100019578	121	STATEN ISLAND	N.Y. POLICE DEPT	MISDEMEANOR	PETIT LARCENY	12/07/2019	40.63677780500007	-74.12884605299998
100020028	42	BRONX	N.Y. POLICE DEPT	MISDEMEANOR	PETIT LARCENY	12/11/2019	40.82931452500002	-73.89289480599997
100022582	47	BRONX	N.Y. POLICE DEPT	MISDEMEANOR	ASSAULT 3 & RELATED OFFENSES	04/09/2019	40.87203753300003	-73.83784793999997
100023083	46	BRONX	N.Y. POLICE DEPT	FELONY	GRAND LARCENY	10/21/2019	40.85878942700003	-73.89837477699997
100030496	44	BRONX	N.Y. POLICE DEPT	FELONY	FELONY ASSAULT	06/04/2019	40.83421942500007	-73.91253114899997
100030509	18	MANHATTAN	N.Y. TRANSIT POLICE	FELONY	GRAND LARCENY	03/10/2019	40.76502407900006	-73.98483594999993
100036138	75	BROOKLYN	N.Y. HOUSING POLICE	MISDEMEANOR	ASSAULT 3 & RELATED OFFENSES	05/06/2019	40.65690638000007	-73.88509115599999
100040859	50	BRONX	N.Y. HOUSING POLICE	FELONY	GRAND LARCENY	07/02/2019	40.87513794200004	-73.90614578699996
100043252	52	BRONX	N.Y. POLICE DEPT	VIOLATION	HARRASSMENT 2	02/18/2019	40.86652284100006	-73.89404245199995

An Analysis of crime and complaints

Design Diagram



An Analysis of crime and complaints

Code Challenge 1

```
Query: select grand_lar.addr_pct_cd, c_grand_lar + isnull(c_arson,0) + isnull(c_hom_neg,0) + isnull(c_burg,0) + isnull(c_child,0) + isnull(c_crim_mis,0) + isnull(c_drugs,0) + isnull(c_weapons,0) + isnull(c_endan,0) + isnull(c_assault,0) + isnull(c_sex_crimes,0) + isnull(c_forgery,0) + isnull(c_gambling,0) + isnull(c_grand_vehicle,0) + isnull(c_hom_vehicle,0) + isnull(c_intox,0) + isnull(c_kid,0) + isnull(c_kid_rel,0) + isnull(c_kid_rel2,0) + isnull(c_misc,0) + isnull(c_murder,0) + isnull(c_unc1,0) + isnull(c_other,0) + isnull(c_oss,0) + isnull(c_pros,0) + isnull(c_rape,0) + isnull(c_rob,0) + isnull(c_s_crimes,0) + isnull(c_theft,0) as sum from (select addr_pct_cd, ofns_desc, count(*) * 2 as c_grand_lar from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "GRAND LARCENY") as grand_lar left join (select addr_pct_cd, ofns_desc, count(*) * 22.5 as c_arson from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "ARSON") as arson on grand_lar.addr_pct_cd = arson.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 9.25 as c_hom_neg from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "\"HOMICIDE-NEGLIGENT") as hom_neg on hom_neg.addr_pct_cd = grand_lar.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 15 as c_burg from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "BURGLARY") as burg on grand_lar.addr_pct_cd = burg.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) as c_child from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "CHILD ABANDONMENT/NON SUPPORT") as child on grand_lar.addr_pct_cd = child.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 2 as c_crim_mis from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "CRIMINAL MISCHIEF & RELATED OF") and law_cat_cd = "FELONY") as crim_mis on grand_lar.addr_pct_cd = crim_mis.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) as c_drugs from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "DANGEROUS DRUGS" and law_cat_cd = "FELONY") as drugs on grand_lar.addr_pct_cd = drugs.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 15 as c_weapons from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "DANGEROUS WEAPONS" and law_cat_cd = "FELONY") as weapons on grand_lar.addr_pct_cd = weapons.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) as c_endan from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "ENDAN WELFARE INCOMP" and law_cat_cd = "FELONY") as endan on grand_lar.addr_pct_cd = endan.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 15 as c_assault from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "FELONY ASSAULT" and law_cat_cd = "FELONY") as assault on grand_lar.addr_pct_cd = assault.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 15 as c_sex_crimes from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "FELONY SEX CRIMES" and law_cat_cd = "FELONY") as sex_crimes on grand_lar.addr_pct_cd = sex_crimes.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) as c_forgery from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "FORGERY" and law_cat_cd = "FELONY") as forgery on grand_lar.addr_pct_cd = forgery.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) as c_gambling from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "GAMBLING" and law_cat_cd = "FELONY") as gambling on grand_lar.addr_pct_cd = gambling.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 2 as c_grand_vehicle from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "GRAND LARCENY OF MOTOR VEHICLE" and law_cat_cd = "FELONY") as grand_vehicle on grand_lar.addr_pct_cd = grand_vehicle.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 9.25 as c_hom_vehicle from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "HOMICIDE-NEGLIGENT-VEHICLE" and law_cat_cd = "FELONY") as hom_vehicle on grand_lar.addr_pct_cd = hom_vehicle.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) as c_intox from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "INTOXICATED/IMPAIRED DRIVING" and law_cat_cd = "FELONY") as intoxic on grand_lar.addr_pct_cd = intoxic.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 22.5 as c_kid from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "KIDNAPPING" and law_cat_cd = "FELONY") as kid on grand_lar.addr_pct_cd = kid.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 15 as c_kid_rel from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "KIDNAPPING & RELATED OFFENSES" and law_cat_cd = "FELONY") as kid_rel on grand_lar.addr_pct_cd = kid_rel.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 15 as c_kid_rel2 from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "KIDNAPPING AND RELATED OFFENSES" and law_cat_cd = "FELONY") as kid_rel2 on grand_lar.addr_pct_cd = kid_rel2.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 8.3 as c_misc from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "MISCELLANEOUS PENAL LAW" and law_cat_cd = "FELONY") as misc on grand_lar.addr_pct_cd = misc.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 22.5 as c_murder from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "MURDER & NON-NEGL. MANSLAUGHTER" and law_cat_cd = "FELONY") as murder on grand_lar.addr_pct_cd = murder.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 8.3 as c_unc1 from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "NYS LAWS-UNCLASSIFIED FELONY" and law_cat_cd = "FELONY") as uncl on grand_lar.addr_pct_cd = uncl.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 8.3 as c_other from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "OTHER STATE LAWS (NON PENAL LA" and law_cat_cd = "FELONY") as other on grand_lar.addr_pct_cd = other.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 2 as c_oss from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "POSSESSION OF STOLEN PROPERTY" and law_cat_cd = "FELONY") as oss on grand_lar.addr_pct_cd = oss.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 2 as c_pros from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "PROSTITUTION & RELATED OFFENSES" and law_cat_cd = "FELONY") as pros on grand_lar.addr_pct_cd = pros.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 15 as c_rape from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "RAPE" and law_cat_cd = "FELONY") as rape on grand_lar.addr_pct_cd = rape.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 15 as c_rob from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "ROBBERY" and law_cat_cd = "FELONY") as rob on grand_lar.addr_pct_cd = rob.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 15 as c_s_crimes from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "SEX CRIMES" and law_cat_cd = "FELONY") as s_crimes on grand_lar.addr_pct_cd = s_crimes.addr_pct_cd left join (select addr_pct_cd, ofns_desc, count(*) * 2 as c_theft from data_nypd group by addr_pct_cd, law_cat_cd, ofns_desc having ofns_desc = "THEFT FRAUD" and law_cat_cd = "FELONY") as theft on grand_lar.addr_pct_cd = theft.addr_pct_cd order by sum
```

An Analysis of crime and complaints

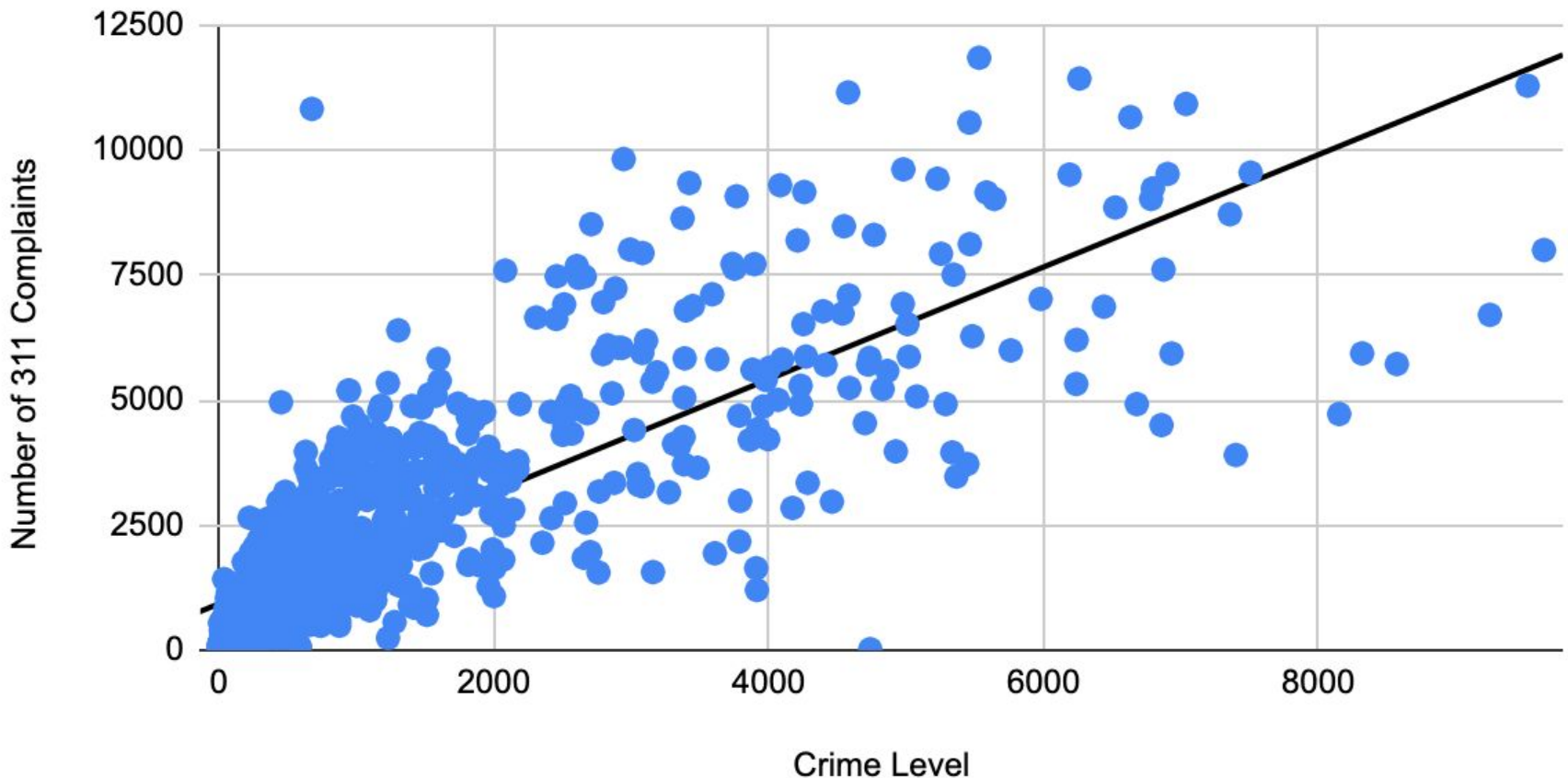
Insights

1. As the level crime in a location increased, so did the number of 311 complaints.
2. Some 311 complaint types did not follow the above trend, with Blocked Driveways and Large Item Pickup occurring the most. This means that in high crime areas, residents are less likely to contact 311 in regards to city services in favor of resolving complaints like noise or graffiti.
3. Calling the 311 number was the most-used method to report complaints, though it declined as crime level increased in favor of online and text reporting.

An Analysis of crime and complaints

Crime Level vs Total Number of 311 Complaints

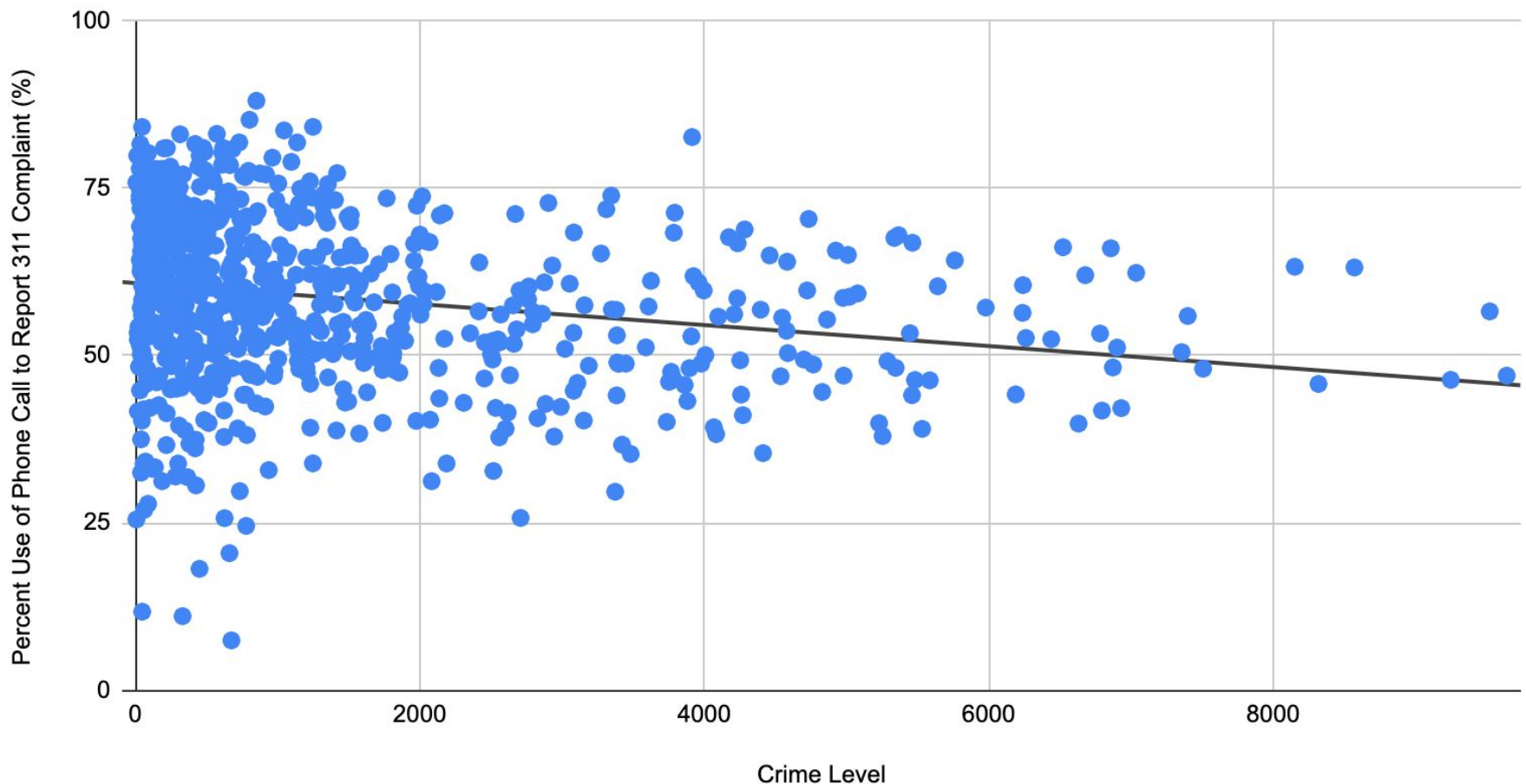
Each data point is a location of latitude and longitude rounded to 2 decimal places



An Analysis of crime and complaints

Crime Level vs Use of Phone Call in 311 Complaints

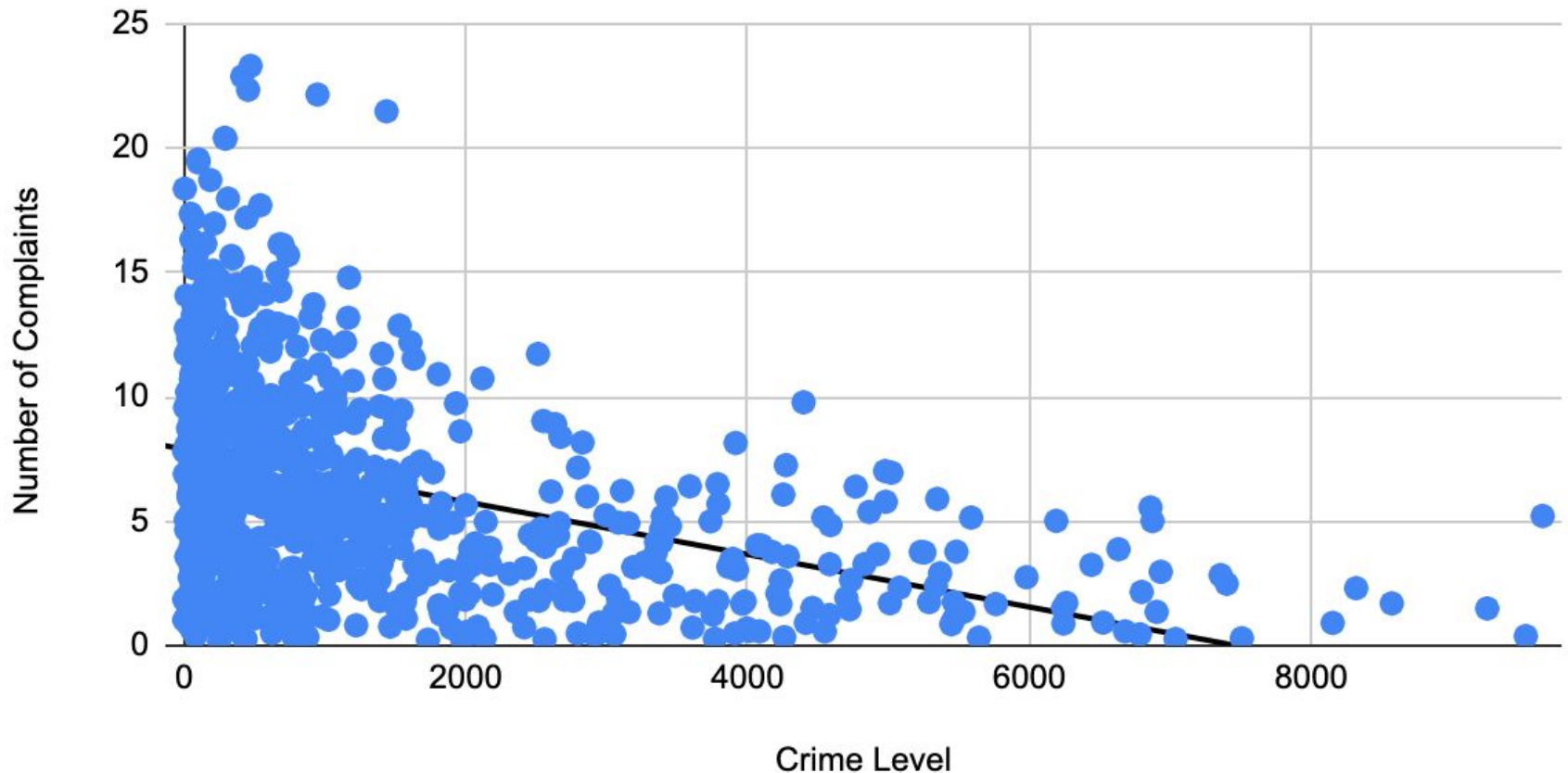
Each data point is a location of latitude and longitude rounded to 2 decimal places



An Analysis of crime and complaints

Crime Level vs Bulky Item Collection

Each data point is a location of latitude and longitude rounded to 2 decimal places



An Analysis of crime and complaints

Obstacles

1. One of the obstacles that we encountered in the beginning was that in cleaning the data some of the fields had extra commas in them which made it hard to get the exact fields that we wanted.
2. Another obstacle that we encountered was coming up with a method to determine crime level in a location. Neither of us are domain experts, so we had to use different sources to come up with our own formula of multiplying prison sentences by number of crimes.
3. We also struggled in the beginning of the project as this topic was not what we had initially intended to analyze. We planned to use an API to get weather data that could be joined with data about COVID-19, but the site ended up not being free. Our COVID dataset was also rather skeletal and would have difficult to use alongside another.

An Analysis of crime and complaints

Summary:

In summary we found this project engaging and a good use of the material that we learned in this class. We were able to string together complex queries to find interesting results, particularly in regards to crime and quality of life in New York City.

Some future changes that could improve the project if it were repeated:

- Working with a domain expert to get a more accurate way to determine crime level with the NYPD data.
- The inclusion of a precinct field in the 311 data would have made joining the sets much easier.
 - Using latitude and longitude relied on rounding and is less tangible than a precinct or borough.
- Further mathematical analysis of the data - neither of us have much of a background in statistics.

An Analysis of crime and complaints

Acknowledgements

- Thank you Professor McIntosh! You helped us a lot through the project and it was a great experience to develop our own analytic.

An Analysis of crime and complaints

References

- <https://opendata.cityofnewyork.us/> - NYC Open Data, where we got both of our data sets.
- http://ypdcrime.com/penal.law/offense_level.htm - The site whose numbers we used to determine the average prison sentences in New York State.
- <https://nypost.com/2019/01/23/nypd-announces-new-plan-to-combat-crime-in-dangerous-neighborhoods/> - One of the sources we used to make sure our determined “high crime” locations were accurate.

An Analysis of crime and complaints

Thank you!