Data Mining and Data Warehousing

Course #: MASY1-GC 3510

Professor Joseph Ng

Fall 2019

New York Police Department Crime and Complaints Analysis

Submitted on Thursday, October 24th, 2019

Submitted by:

Name: Group 4 – Sarah Kim, Jiaoyu Guo, Sijia Fang, Ziyue Wang

Email: ek3203@nyu.edu, jg6254@nyu.edu, sf3254@nyu.edu, zw2013@nyu.edu

Table of Contents

Introduction 1

SQL Data Model 4

Location of Crime and Complaints 5

Time of Crime and Complaints 14

Description of Crime and Complaints 23

Conclusion and Recommendations 28

Appendix 30

Introduction

Background

The New York City Police Department (NYPD) is a municipal police department in the United States, NYPD is responsible for policing in a city of 8.5 million population, "by performing a wide variety of public safety, law enforcement, traffic management, counterterror, and emergency response roles." The NYPD records crime, traffic, and employee data and makes most of the statistics transparent to the public. NYPD publishes analyses of crime trends, traffic data regularly on their official website as well.

Problem Statement

Even though there are plenty of datasets on crime rates, there is not a single crime map or analysis on neighborhoods in New York City. Therefore, people in NYC are not aware of their environment on how safe or comfortable it is. For example, tourists often book cheap hotels in areas that are not so safe without knowing well. Additionally, parents may need to know whether it is the right decision for them to move into certain areas, seeking primary education and a safe place for their children.

Lastly, NYPD is not able to efficiently view the complaints and crime reports simultaneously as there are two separate datasets, therefore hindering them to dispatch police officers inefficiently.

Goal

The objective of this database project is to combine NYPD's crime database and NYPD's complaint database, extracting information and provide a report to the public. For example, after combining the database, we can easily find out which neighborhoods have the most arrests and complaints, also, we can identify if arrest rates are somehow related to complaints.

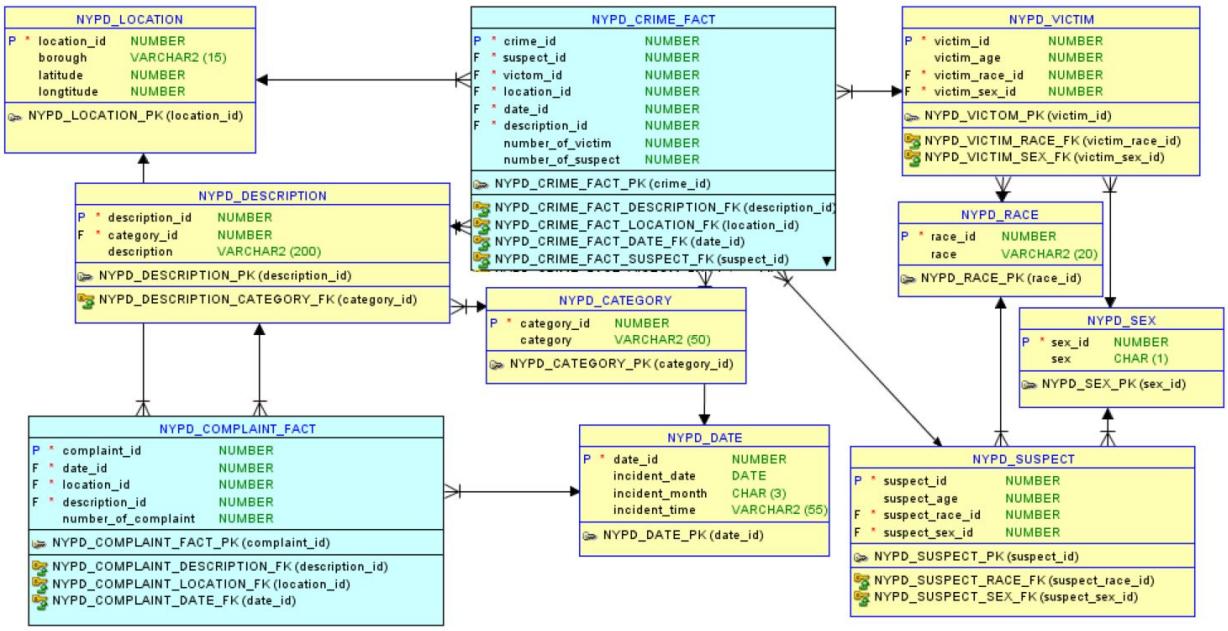
Though there are already database reports and publications of crime statistics and traffic data, I still believe its database can be upgraded to help to scale down criminal rates and ameliorating the quality of life. NYPD will be able to

deploy the right number of officers in the right places at the right time. For example, the current NYPD's crime database contains items such as murder, rape, robbery, shooting, and more. Those are good for basic search, however, if individuals want to advance the search to find the most well-guarded and placid neighborhood, more elements should be added on.

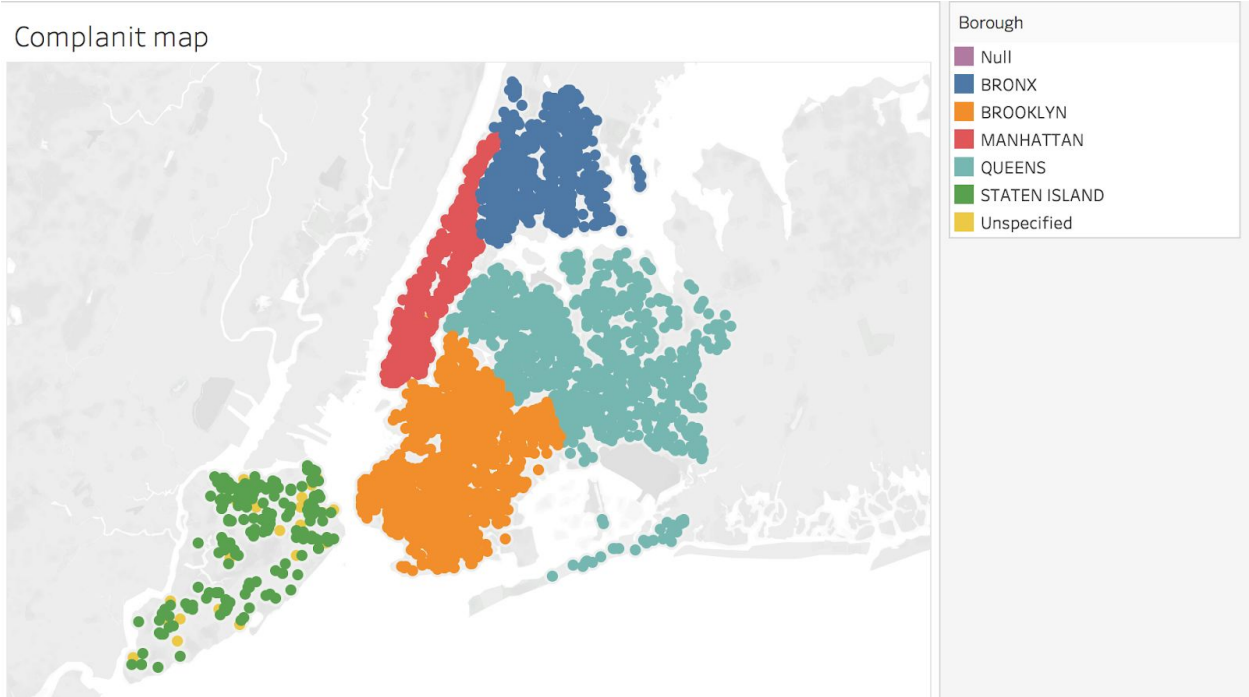
We will merge each dataset into a unified database by cleaning and standardizing the data by using Oracle SQL Developer. Furthermore, we will improve the database by optimizing queries performance, creating optimal indexes and removing data redundancy. By creating an efficient unified database and a luminous report, we can actively help the residents and newcomers who live in New York City looking for their gratified residence.

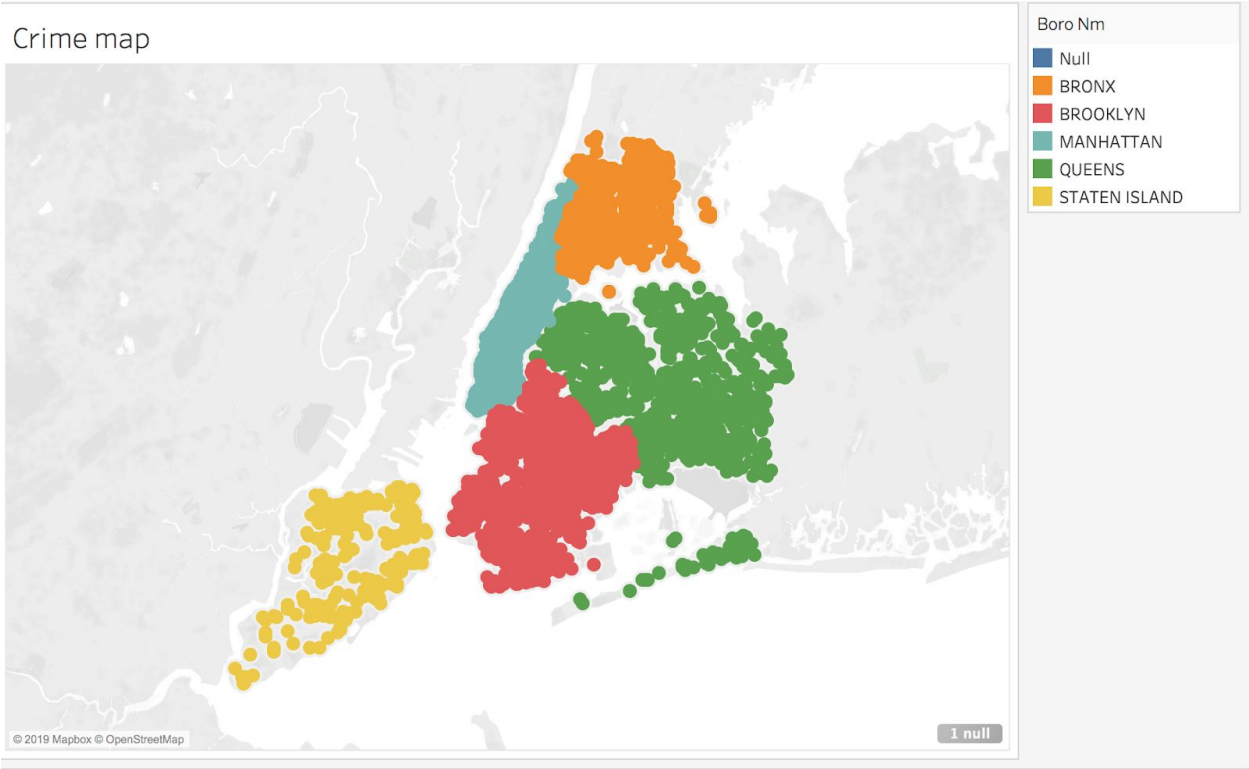
Finally, we need to visualize our datasets and analysis by making a crime and complaint map, giving people a clear understanding of how the environment is like in certain areas.

SQL Data Model



Location of Crime and Complaints

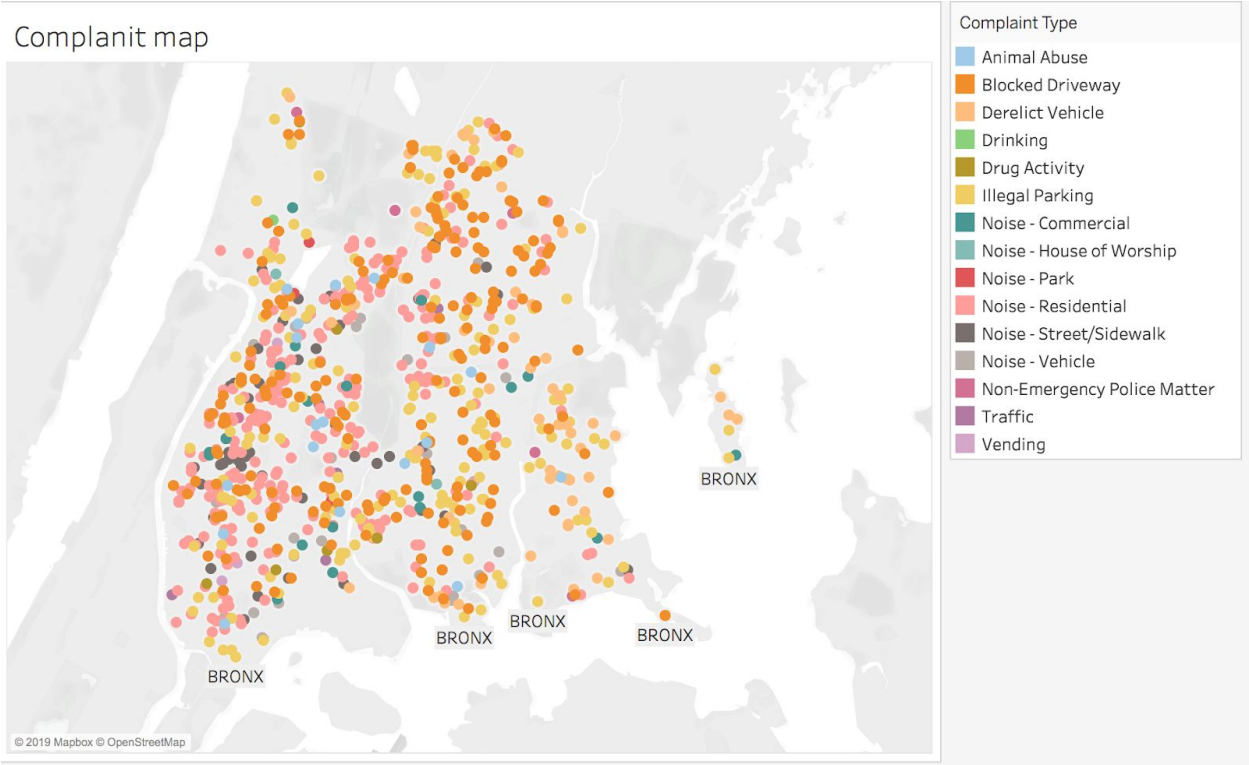


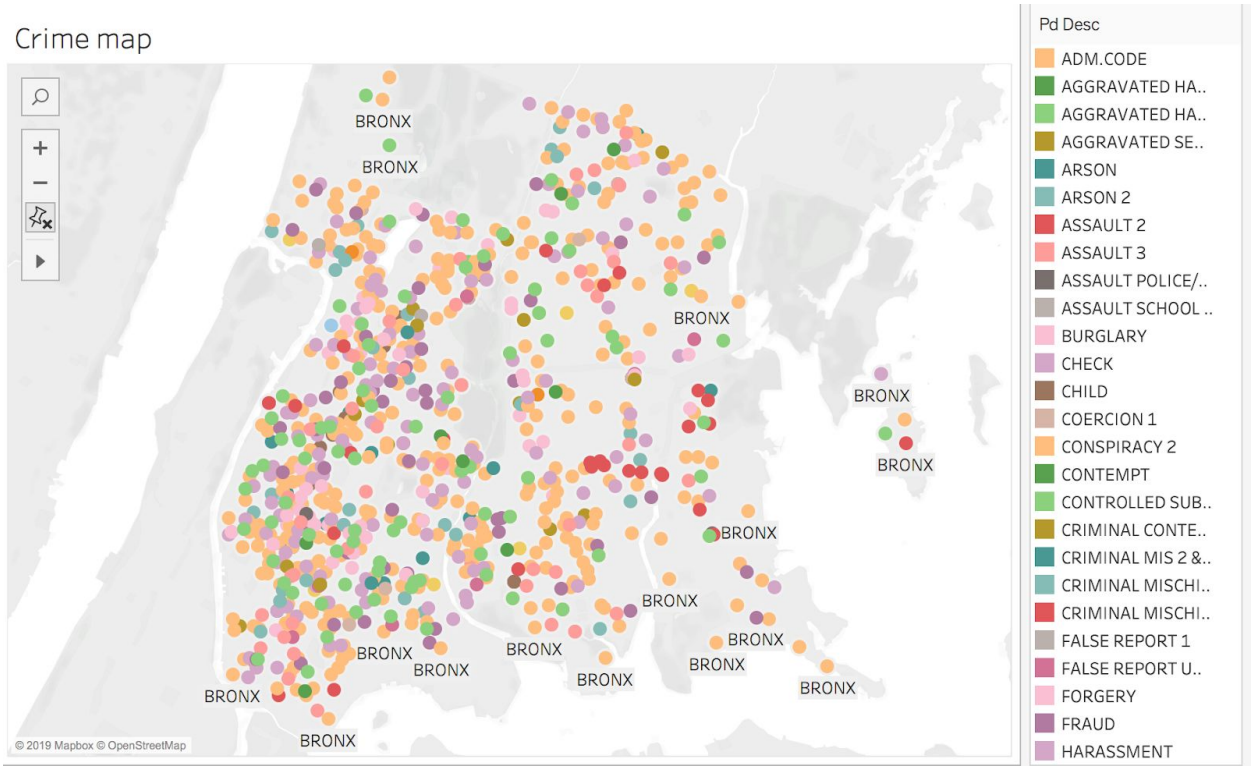


New York City encompasses five boroughs: The Bronx, Brooklyn, Manhattan, Queens, and Staten Island. Those two maps display complaints and crime distribution in each borough of New York City. According to the map, we can see

that Brooklyn and Queens get more complaint reports and more crimes than other boroughs while Staten Island seems to have the least complaints and crimes.

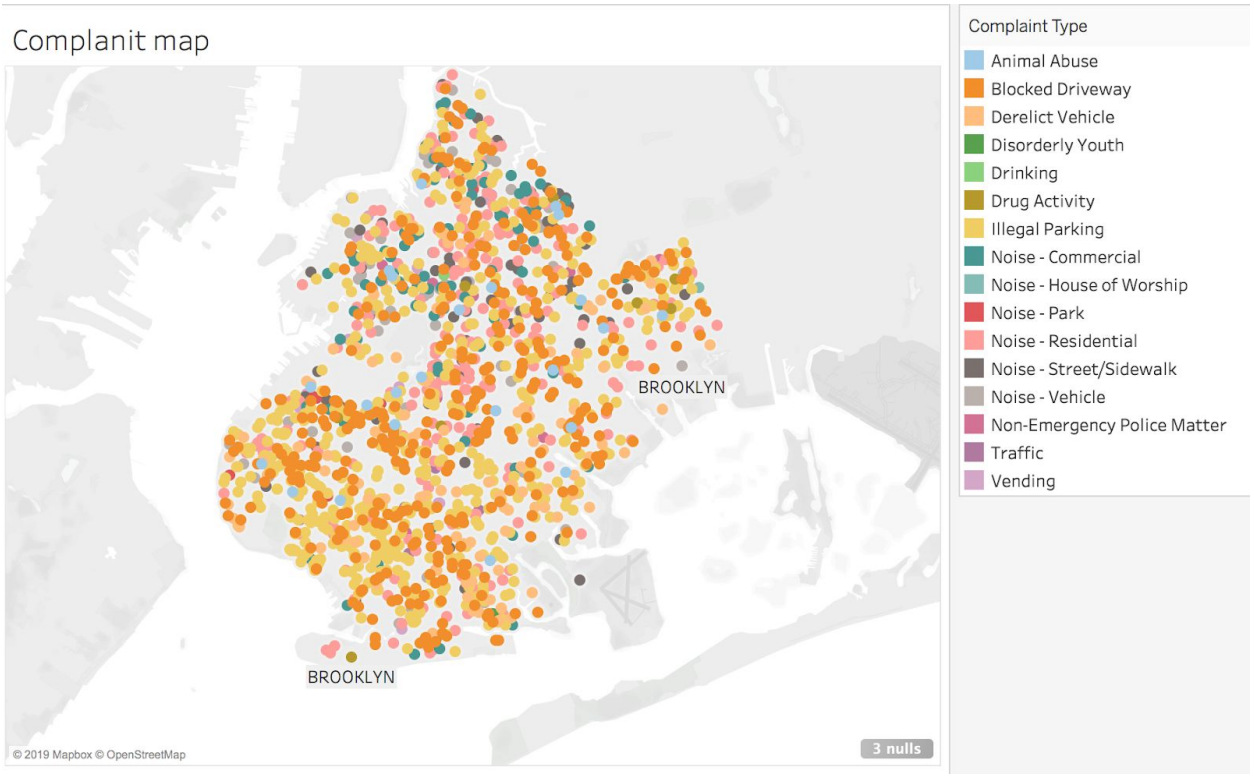
Bronx

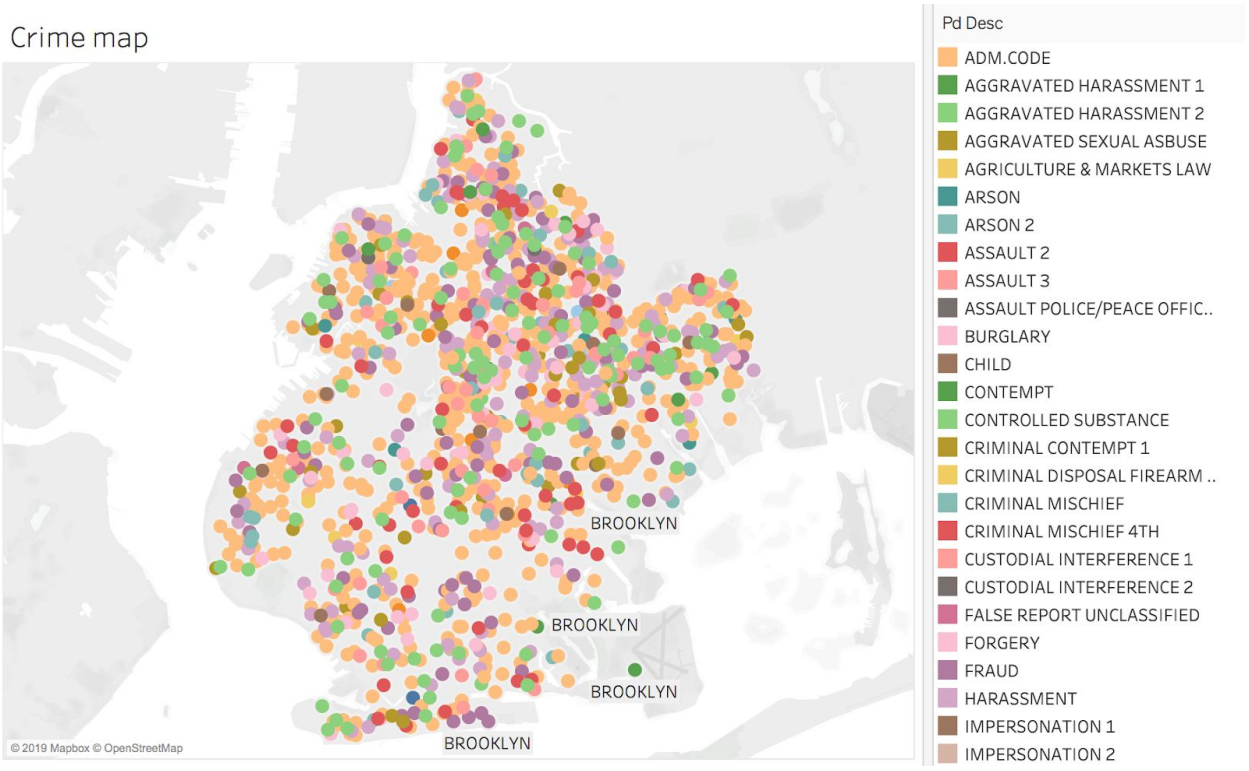




From the complaint-crime distribution map, we can see the residential noise, blocked driveway, illegal parking, are most frequently being complained. The Bronx has the most drug activity complaints compared with other four boroughs. Larceny is the most frequently-happen crime in the Bronx, besides that, harassment, fraud, tampering are responsible for the majority of the crimes. What is surprising to mention is that the Bronx is the only borough that has someone who assaults the school safety agent.

Brooklyn

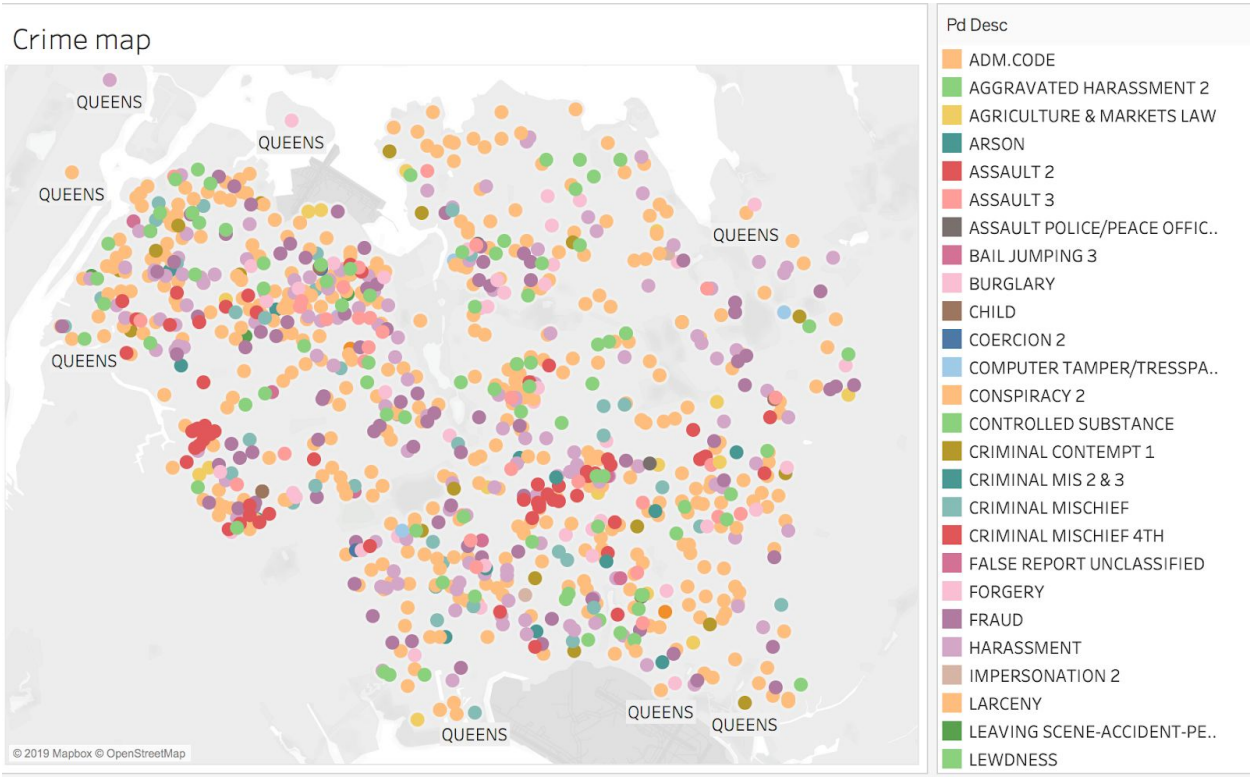


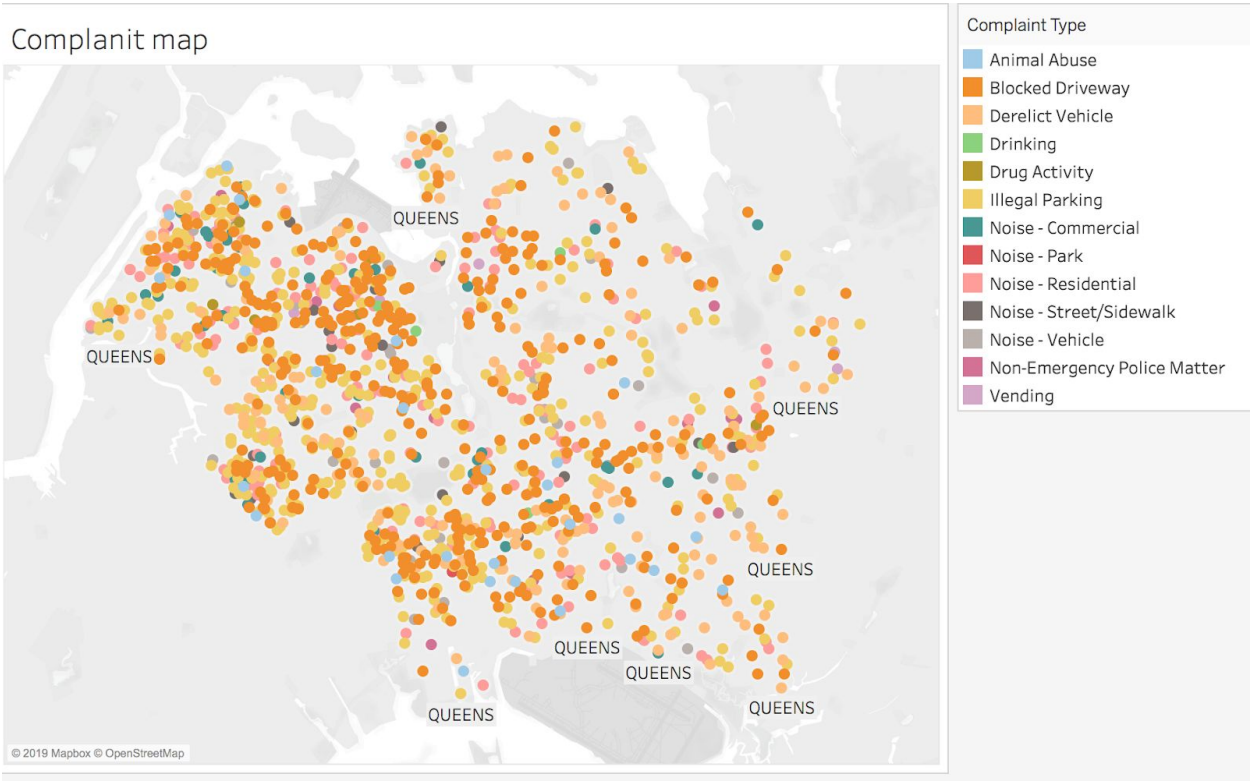


From the complaint distribution map above, we can see that Brookly has the most illegal parking complaint among five boroughs, which are almost 2.5 times more than that in Bronx and Manhattan, and 8 times more than that in Staten Island. Other

than the blocked driveway and illegal parking, noise is also a major complain reseason. People in this neighborhood were complaining a lot about the noise from the residential, street, commercial business, vehicle, or even the park. The complaint about commercial noise and street noise are significantly more than that in Bronx, Queens, and Staten Island, which is almost in the same level as that in Manhattan. In terms of the crime distribution map, larceny is the most frequent crime in Brooklyn, which is almost three times the amount of second frequent crime, harassment.

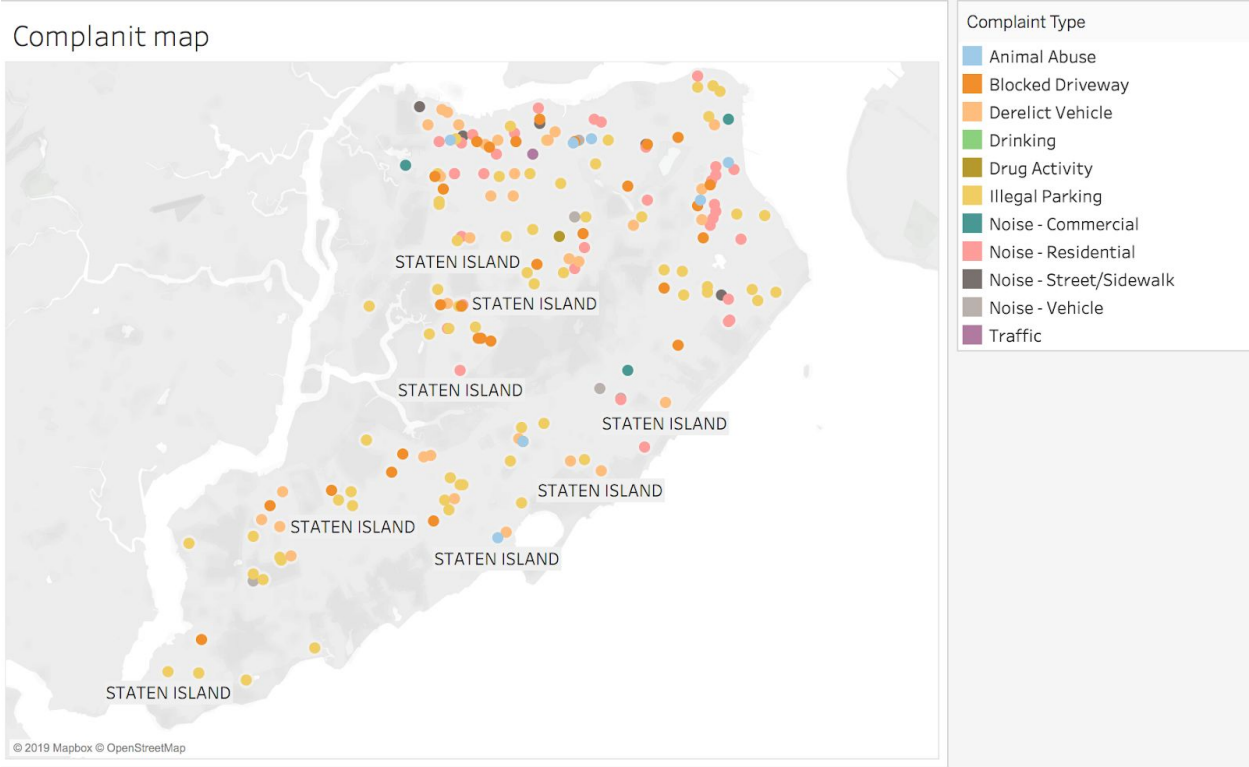
Queens

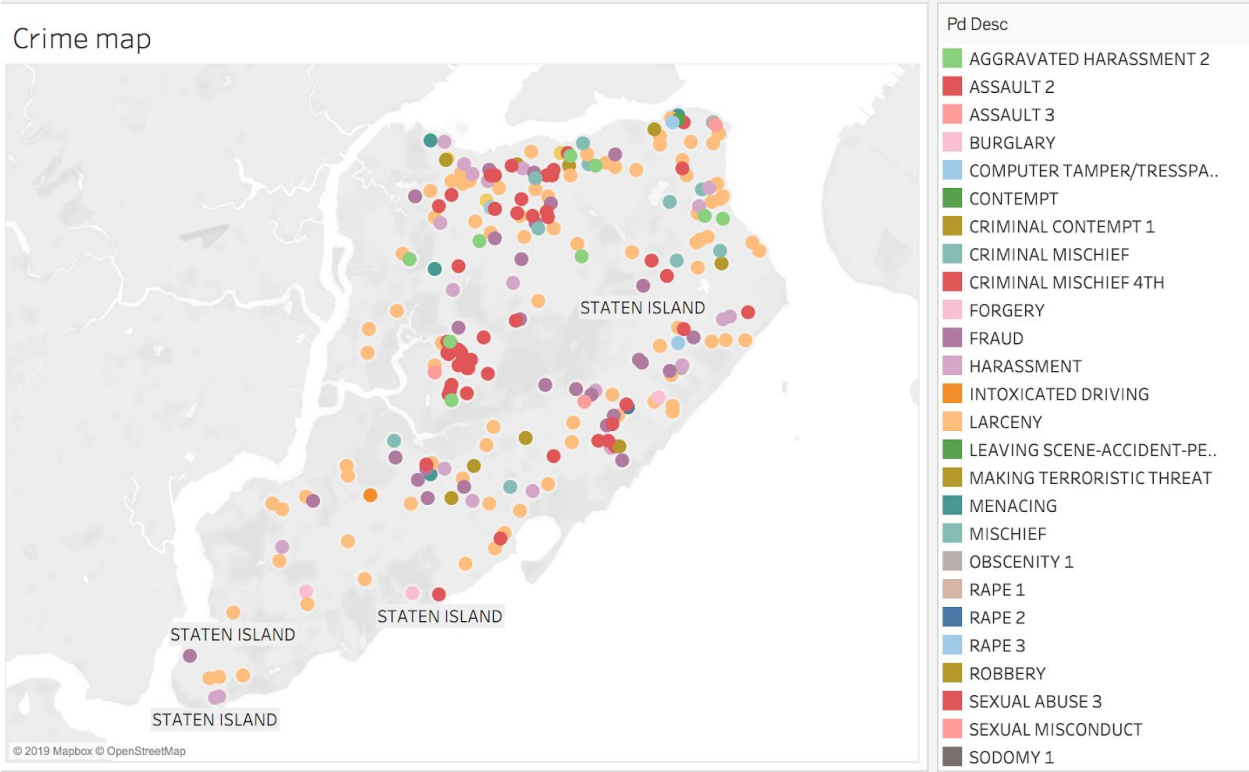




From the complaint distribution map, derelict vehicle, blocked driveway, illegal parking, residential noise are major complaining issues in Queens as well. As the fourth-most densely populated county in the United States, it doesn't have traffic complaints at all in the whole year. Queens is the most ethnically diverse urban area in the world, approximately 48 percent of them foreign-born. In New York City, Queens Is the only area has someone committed the abandoning animal crime.

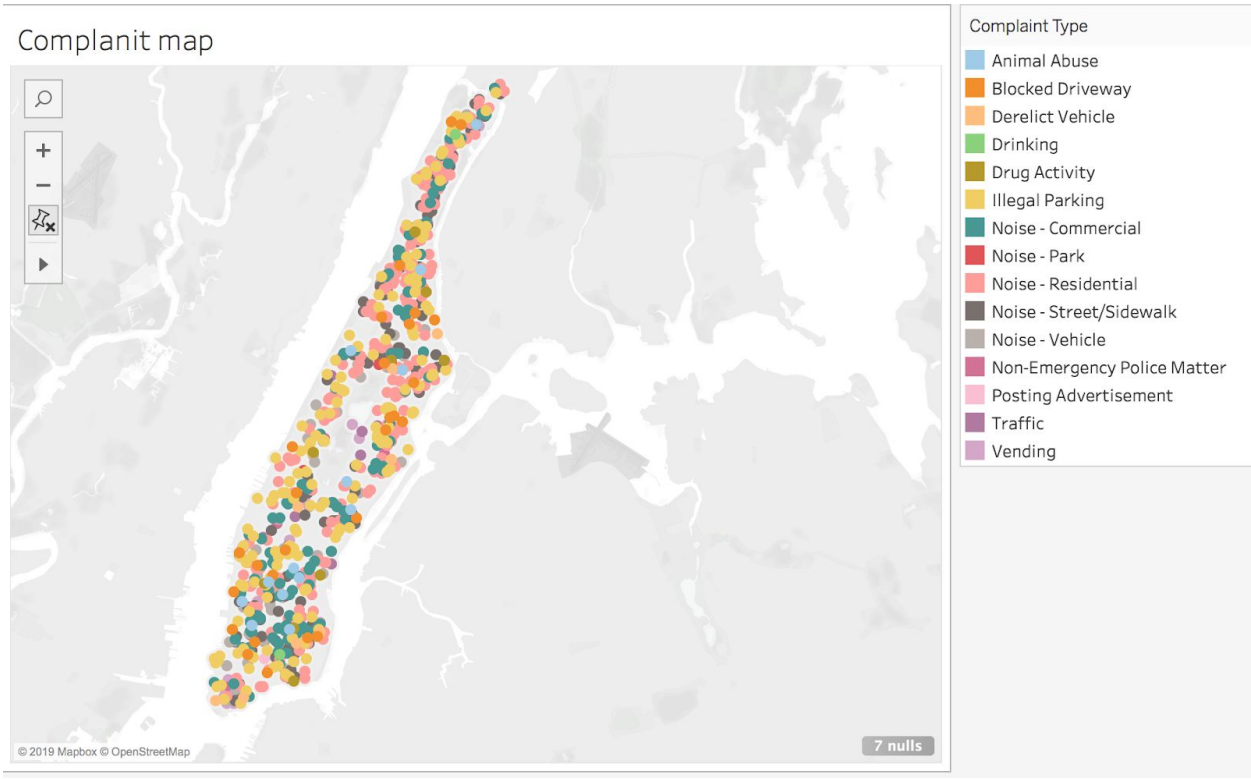
Staten Island

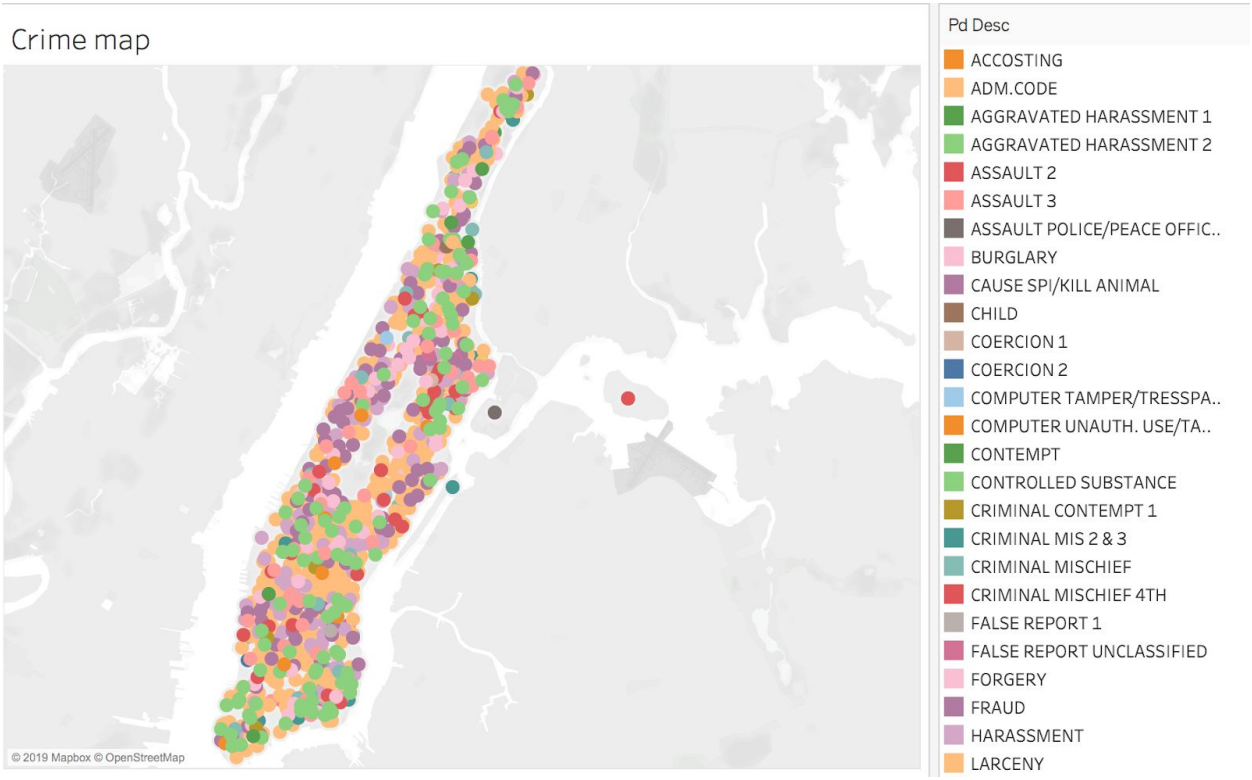




Staten Island has the least complaints and crimes compared with other boroughs. It only has one traffic complaint, 1 drug activity and 1 drinking complaint in the whole year while other places have several. It only has two sexual misconduct crimes that happened in terms of all sexual crimes. It also doesn't have any weapon-related crimes. Staten Island seems to be a relatively safe and peaceful area for residents.

Manhattan



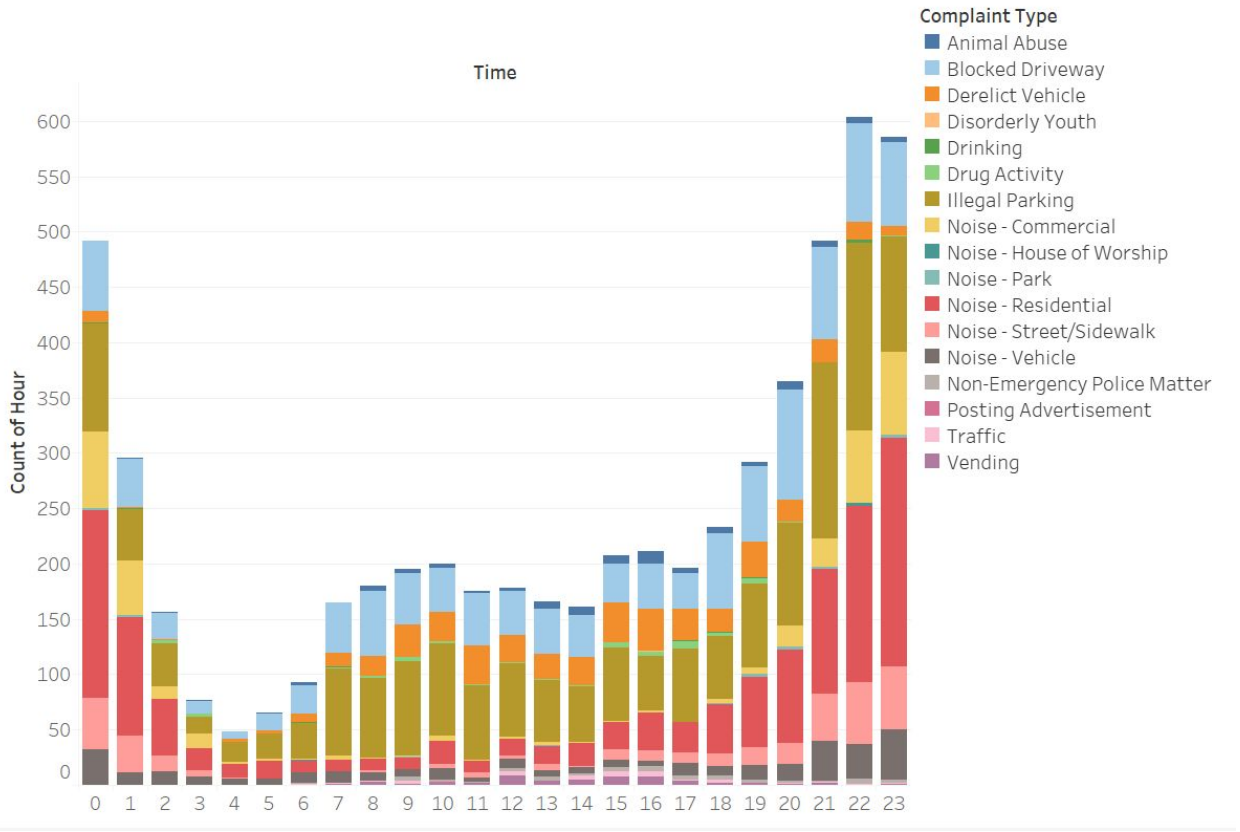


Manhattan is the most densely populated borough in New York City. Manhattan is the city's economic and administrative center. That explains why Manhattan is the only place has the posting advertisement compliant and accosting crime. Different from other boroughs, in Manhattan, noise is a major concern, among the top five complaining reasons, noise complaint account for four of them, including residential noise, commercial business noise, street noise, and vehicle noise. Manhattan has the most complaints about vending and traffic in New York City as well. In other boroughs, blocked driveway seems to be a major issue, however, Manhattan only got 28 complaints about blocked driveway in the year, which are tremendously lower than that in Brooklyn and Queens at 395 complaints and 445 complaints respectively. Manhattan had the least derelict vehicle complaints as well, only 8 complaints about the derelict vehicle, which is even lower than that in Staten Island at 32.

Time of Crime and Complaints

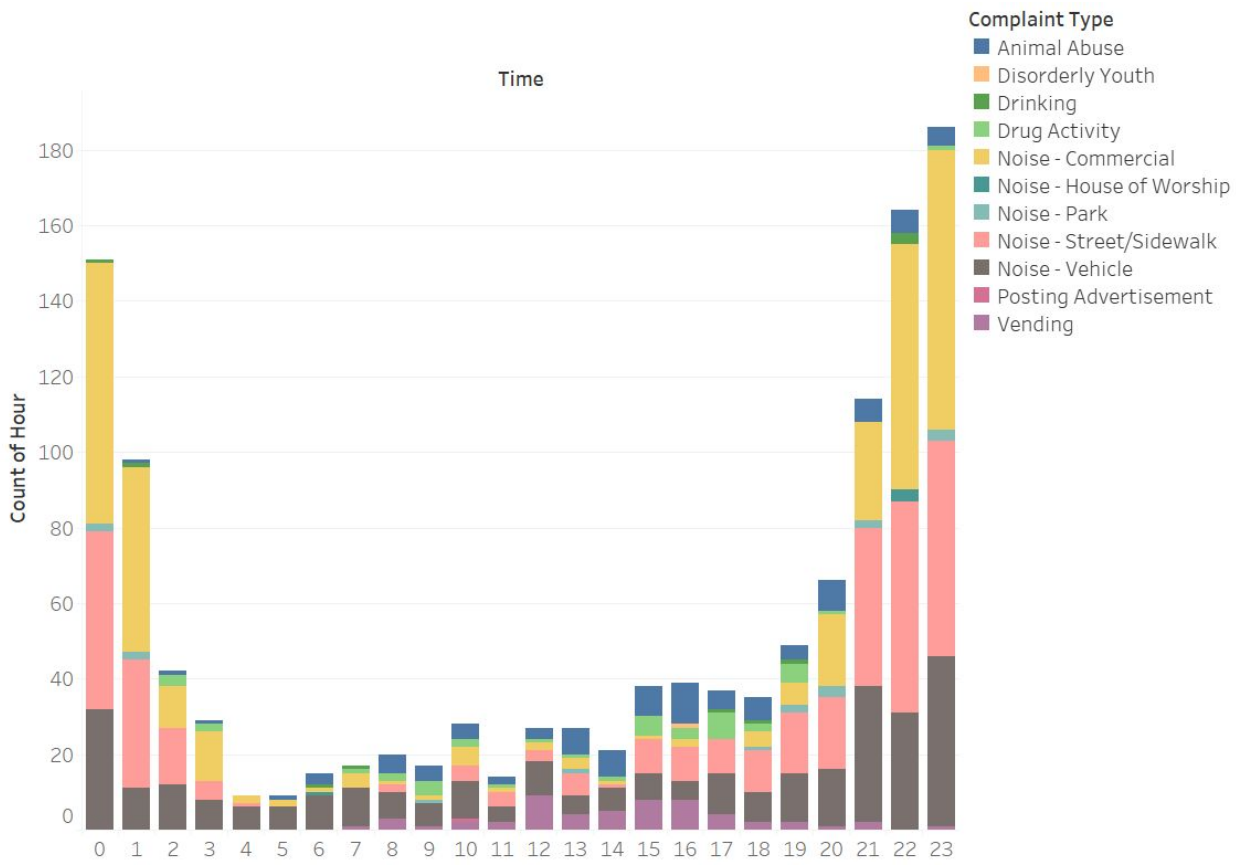
Time of Complaints

This is all complaints filed in New York City according to its occurrence time. The horizontal axis shows the time, from 0 am (12 a

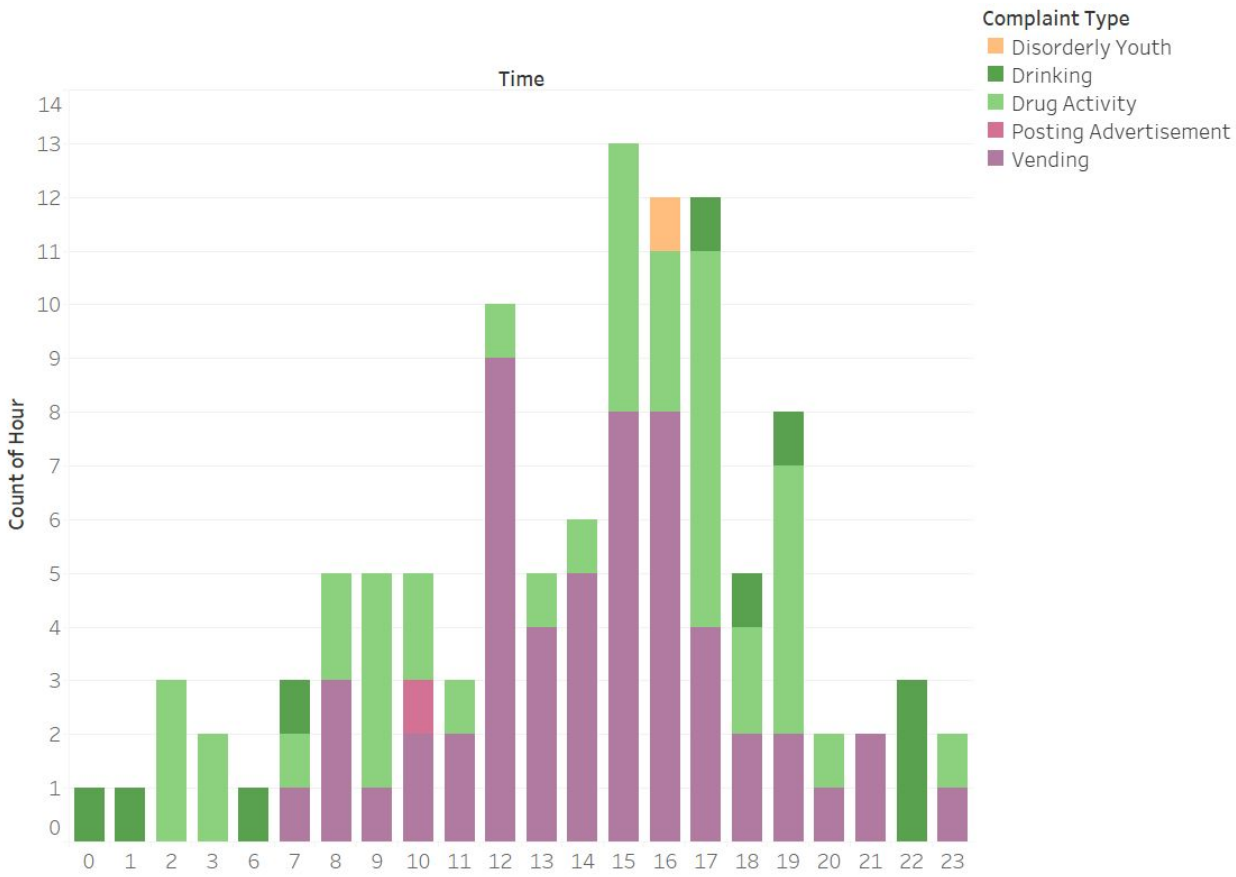


m) to 23 pm (11 pm). As you can see from the pink-colored bars, in the very early morning and late-night, there are higher complaints filed for residential noise. However, as expected, during the daytime, there are significantly fewer complaints filed for residential noise. Additionally, blocked driveway in a light-blue color and illegal parking in a beige color shows there are traffic-complaints throughout the day. This shows that NYPD can dispatch more officers around neighborhoods with a high number of residential noise complaints from 9 pm to 1 am for a better and safer neighborhood. Also, NYPD may dispatch more traffic officers to reduce the illegal parking and blocked driveway.

For the better analysis in safety, complaints related to traffic and residential noise are excluded in the below analysis.

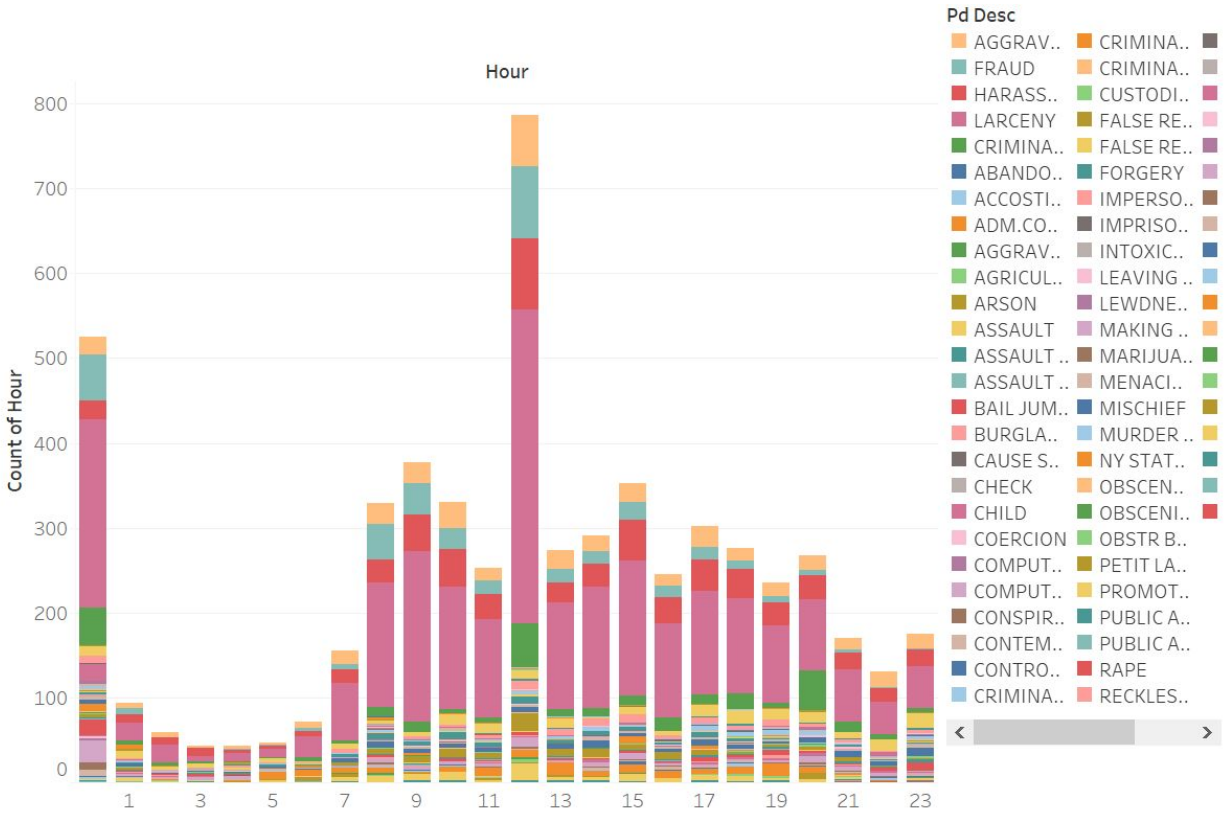


It is still shown that the majority of complaints are from noise, especially street/sidewalk noise and vehicle noise at late night and early morning. From this, it is clear that NYPD must dispatch more officers to further control the noise, especially from the evening to the early morning. Additionally, it is demonstrated there are a fair number of complaints in animal abuse, from the afternoon until late at night. It is very important for NYPD to work with animal welfare organizations to improve this condition by educating people about the Animal Welfare Act (AWA).

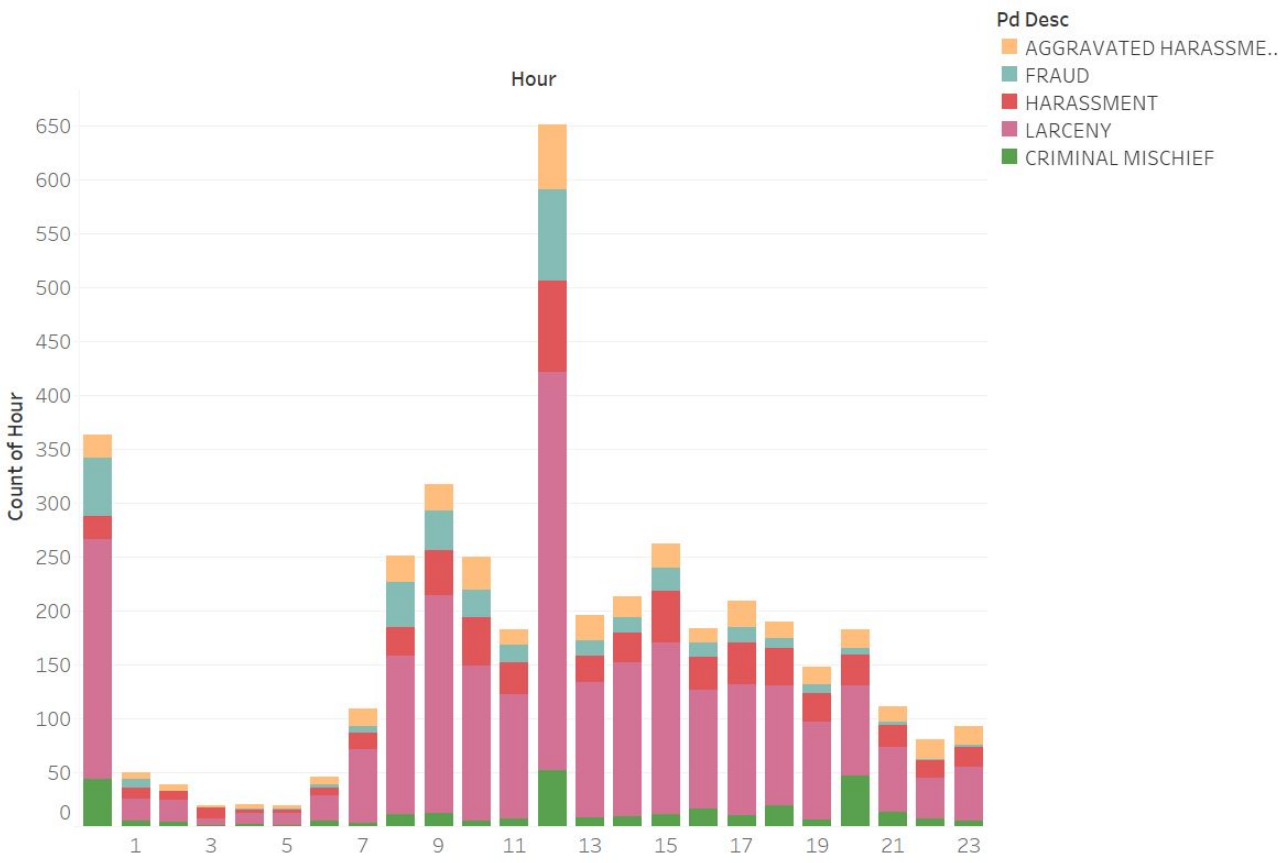


Lastly, the above bar graph shows complaints filed excluding noise, traffics, and animal abuse. Although there is not enough data to determine the time with the related complaints and crimes, we can assume what time certain complaints usually filed around. During the day time, most complaints filed were vending. Vending happens from 7 am until very late at night, and it peaks during the afternoon. NYPD can dispatch more officers to target from 12 pm to 4 pm for vending activities. Lastly, drug activities also happen throughout the day. As drug activities do not have a certain time that happens the most, NYPD should consider to watch out for drug activities throughout the day.

Time of Crimes



The above bar chart is crimes reported according to its reported time. From the bar chart, it can be seen that 12 pm has the highest number of crimes. It is interesting that crimes happened during the daytime rather than late at night or very early in the morning. Therefore, NYPD should dispatch more officers during the time to secure people's safety. It is shown that from 1 am to 7 pm, there are fewer crimes than the other times, but it is possible that NYPD already has positioned more officers to prevent crimes during these hours. The below bar chart is created with the top 5 crimes that happened the most as this bar chart is hard to know when and what kind of crimes were reported.

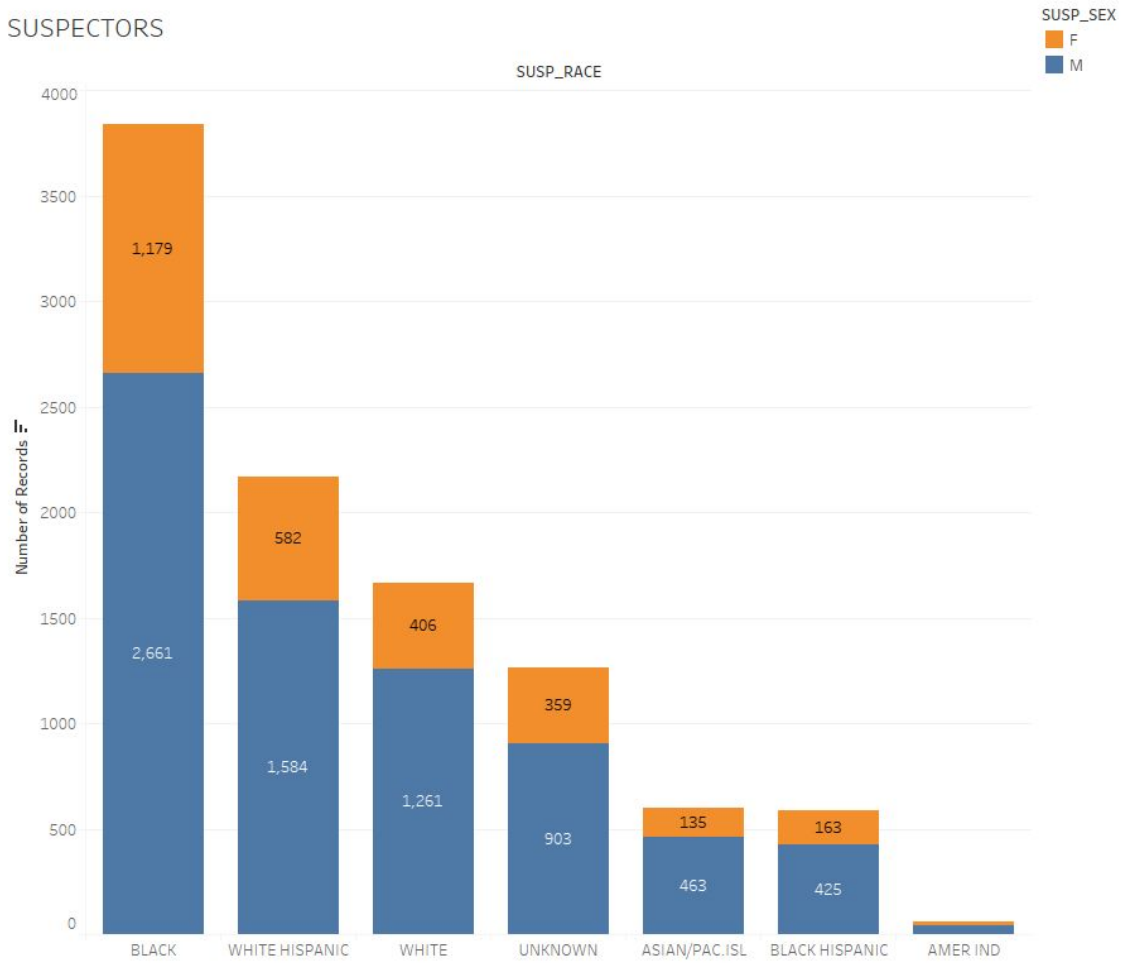


In this bar chart, it is shown that larceny happens the most at 12 pm, but also fairly high from 7 am until 12 am. It can be assumed that larceny probably includes shoplifting as many stores open from 9 am until 5 pm. As it is demonstrated that there is less larceny from 1 am to 7 am, NYPD is recommended to deploy more officers during the daytime according to its location where larceny happens the most. However, what must be considered is that it is possible that victims reported crimes not right away when the crime happened, but later after the crimes.

Finally, these two bar charts of complaints and crimes indicate that the complaints are mostly filed during the late night until early morning while the crimes are mostly reported during the day.

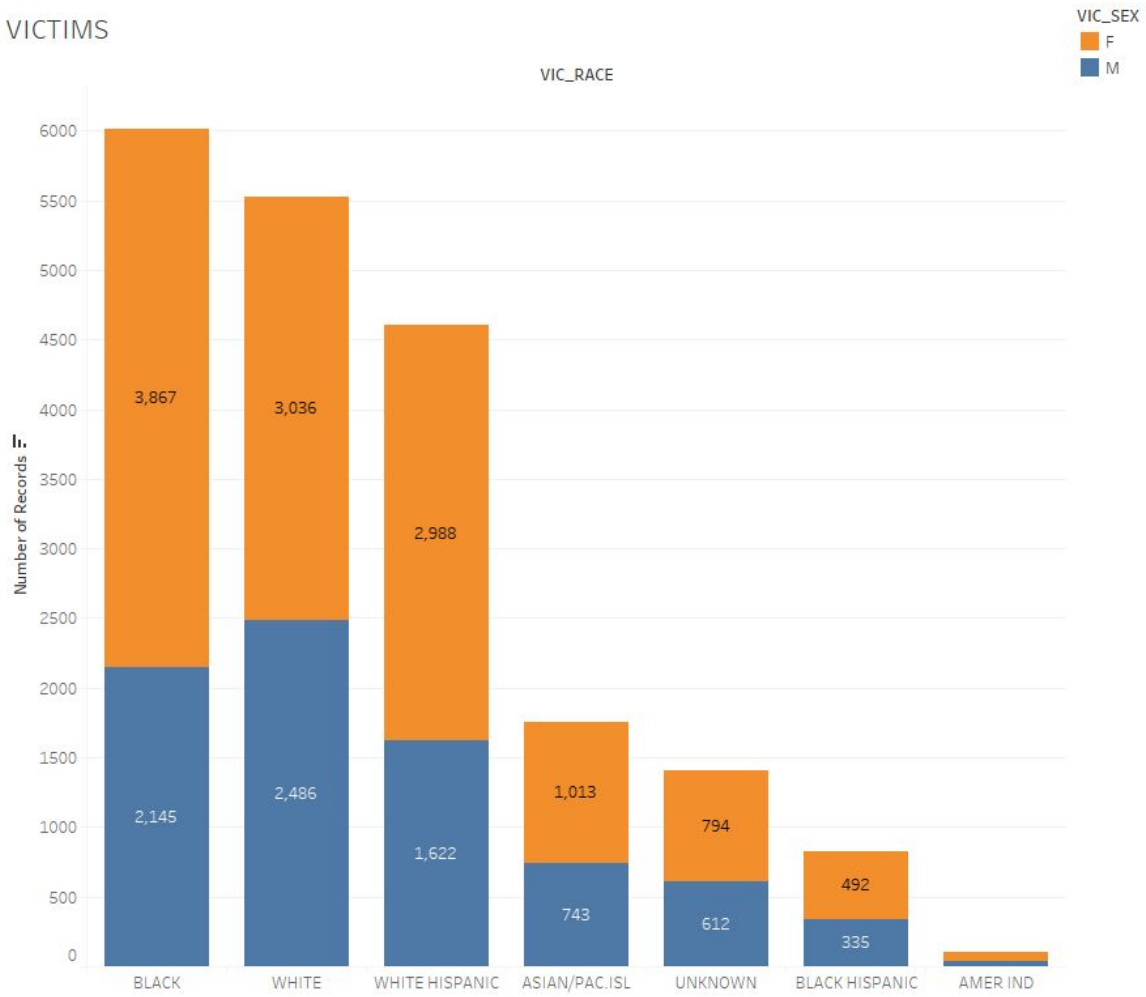
Race and Sex of Susceptor and Victims

Race and Sex of Susceptor



Based on the bar chart above, we can clearly see the distribution of race and gender among suspects. The number of black suspects is significantly higher than that of other races. The number of black male suspects is pretty high and it can exceed the total number of male and female suspects of other races. In addition to racial differences, we can also see that the number of male suspects is much larger than that of female suspects, which is characteristic of suspects of all races. Of the same race, the number of male suspects is about three times larger than the number of female suspects.

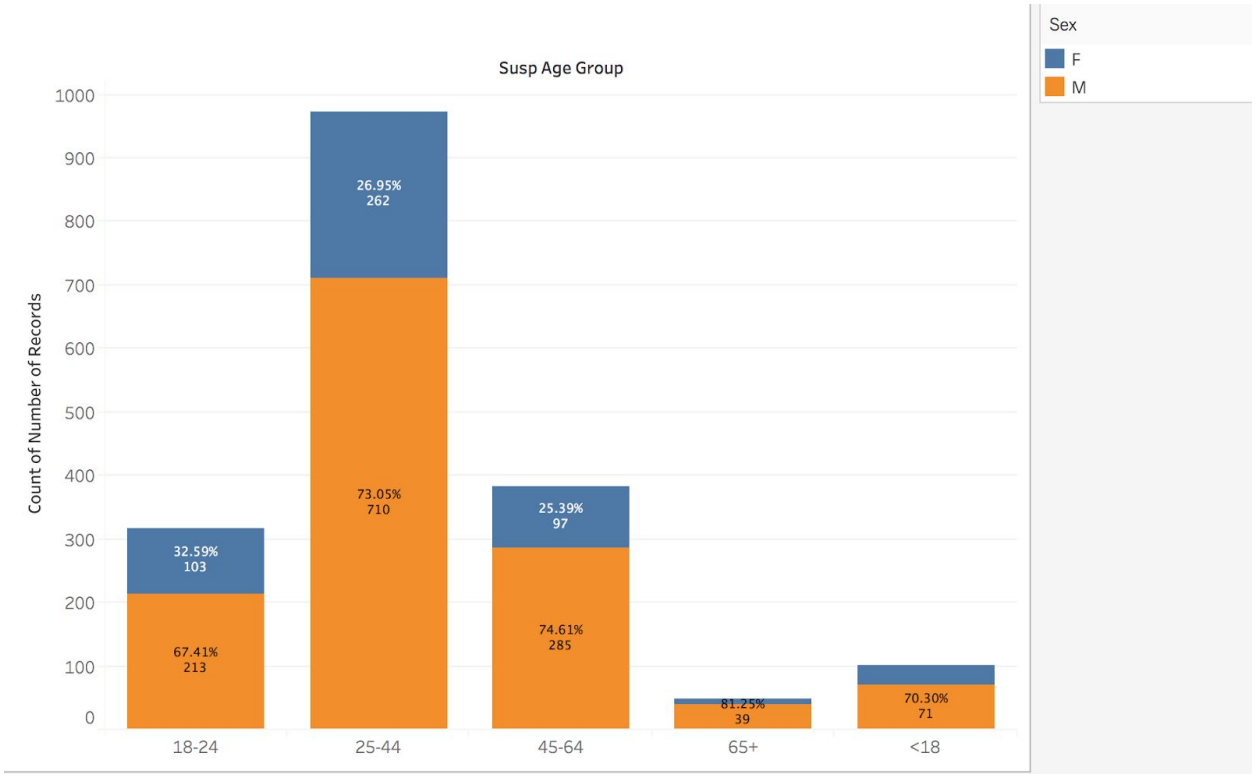
Race and Sex of Victims



In this bar chart, we are able to find that blacks have the highest number of victims, followed by whites and white Hispanics. The number of victims of these three races is much higher than the number of victims of other races. The number of female victims is much higher than the number of male victims, with black women having the highest number of victims. Although the total number of victims of white male and female is lower than the total number of victims of white, the number of victims of white men is higher than that of white male victims.

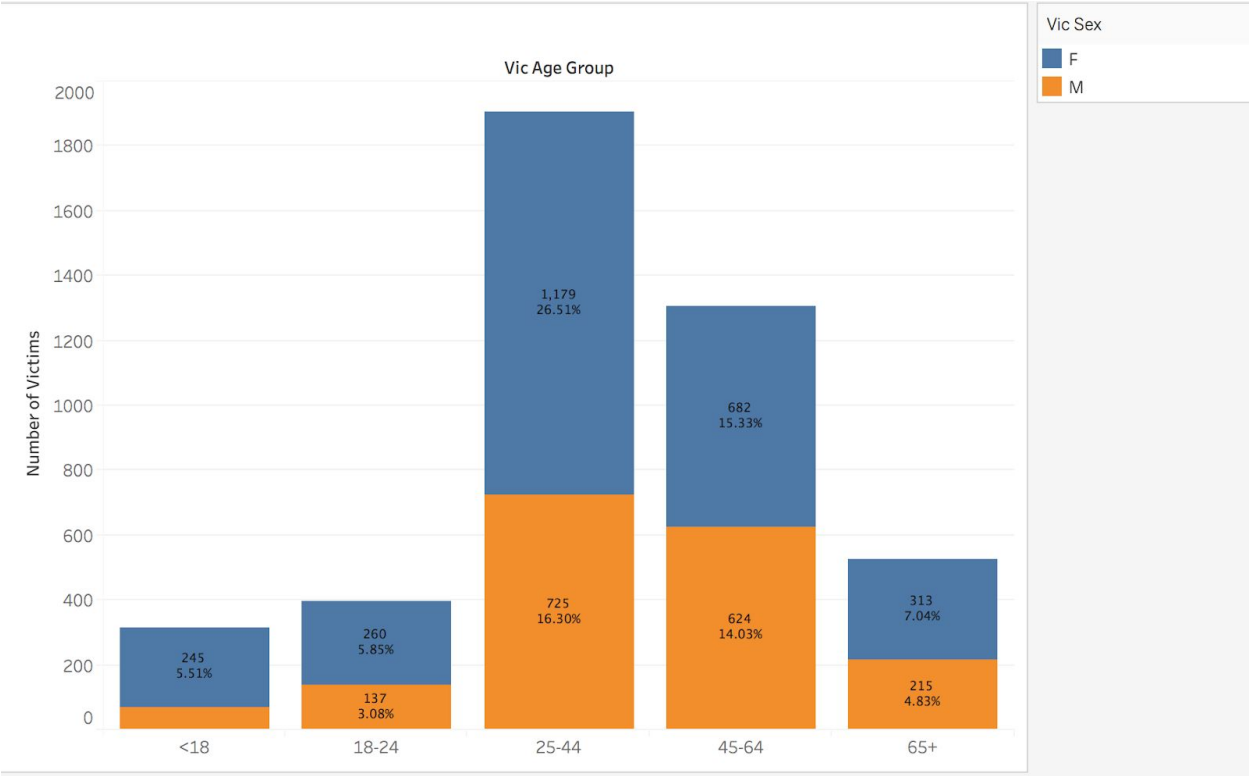
Age of Suspects and Victims

Age of susceptors



The graph above shows the number of suspects in different age groups. First, the suspects' age are most frequently at 25-44, while mile suspects account for 73.05% at 710 while female suspects account for 26.95% at 262, a lot higher than that of the age group of 65+ and age group of 18-. As for group 16-24 and group 45-64, the male suspects between 18-24 of 213 people are lower than the male suspects between 45-64 of 285 people, while the population of female suspects between 18-24 at 103 is slightly higher than the population of female suspects between 45-64 at 97. What is worth mentioning is the number of suspects between 25-44 is more than 20 times the suspect's group who are older than 65.

Age of victims

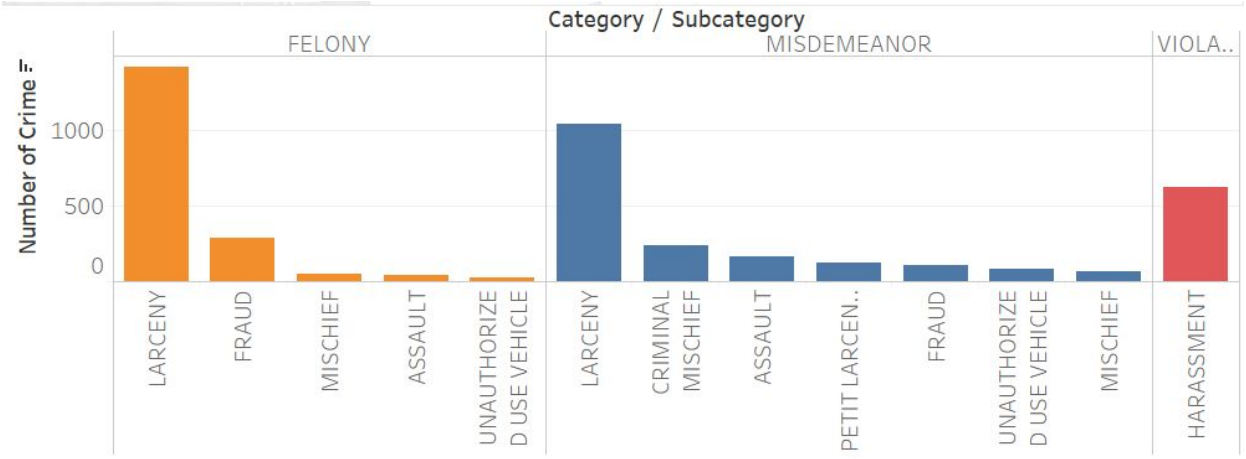


The bar chart below shows the number of suspects and victims in different age groups. It is clear that female victims are significantly more than male victims in each age group. Of the 5 age groups, females from ages 25 to 44 are the most easily targeted group, there were 1179 female victims in the whole year. By contrast, a male under 18 seems to be the least targeted group.

Description of Crime and Complaints

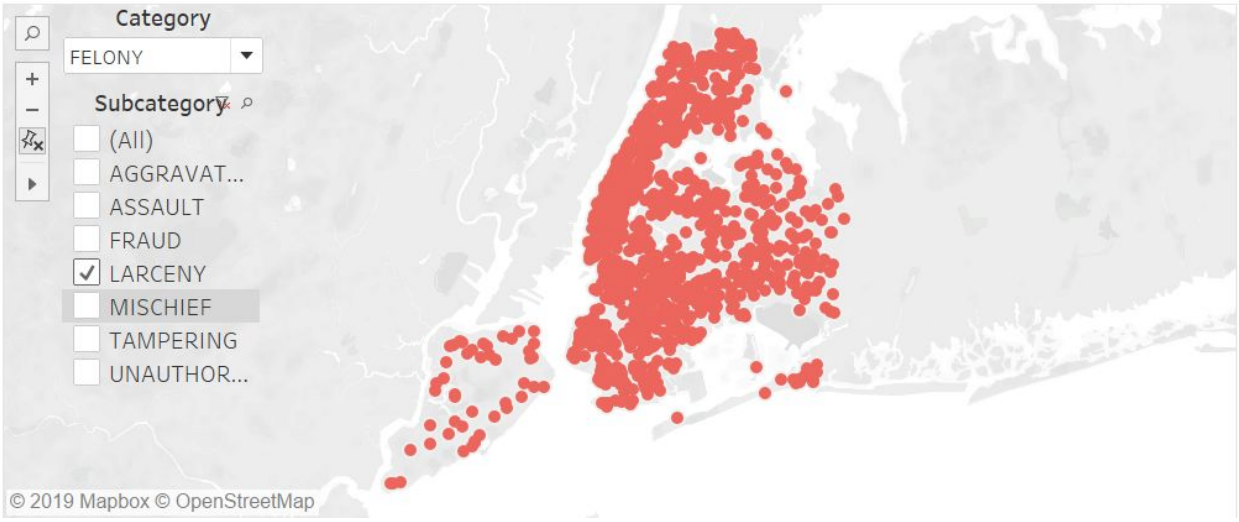
Crime Analysis

Crime Analysis focuses on the safety of living areas with different types of crime and their primary locations. There are three categories of crime: felony, misdemeanor, and violation.

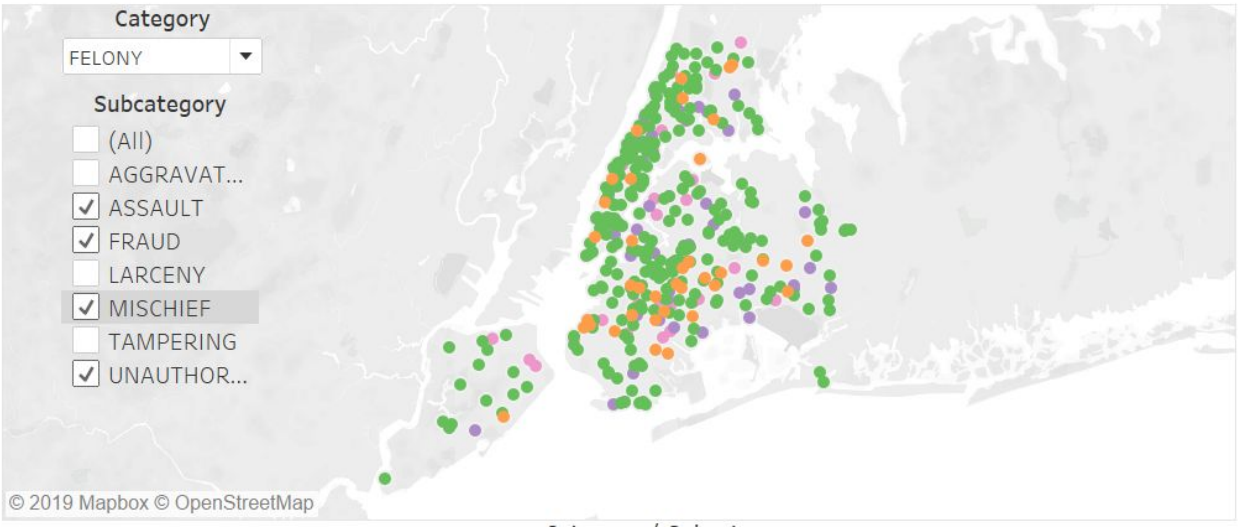


- Felony

Felony, the serious crime which leads to punishment of prison, has six main subcategories: larceny, fraud, tampering, mischief, assault, and unauthorized-use vehicles. Larceny, which is stealing, accounts for the most significant part of the crime, which happened 1,416 times in 2018. From the density of dot-map, we can know that serious stealing often occurs almost everywhere in Manhattan, lower Bronx, and East Brooklyn.

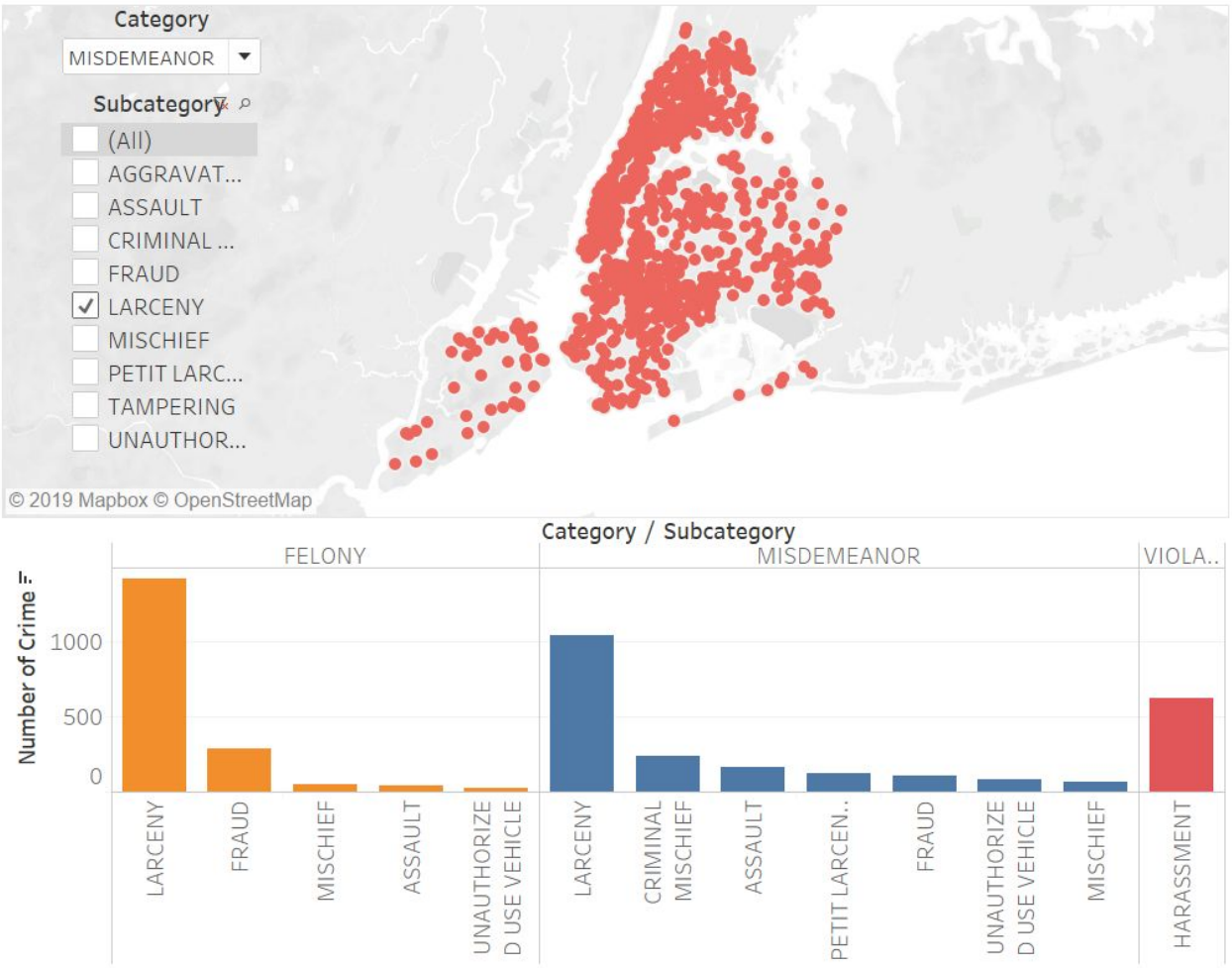


Other crimes happen in a similar trait. Manhattan always has the most significant number of crimes because of its population density.

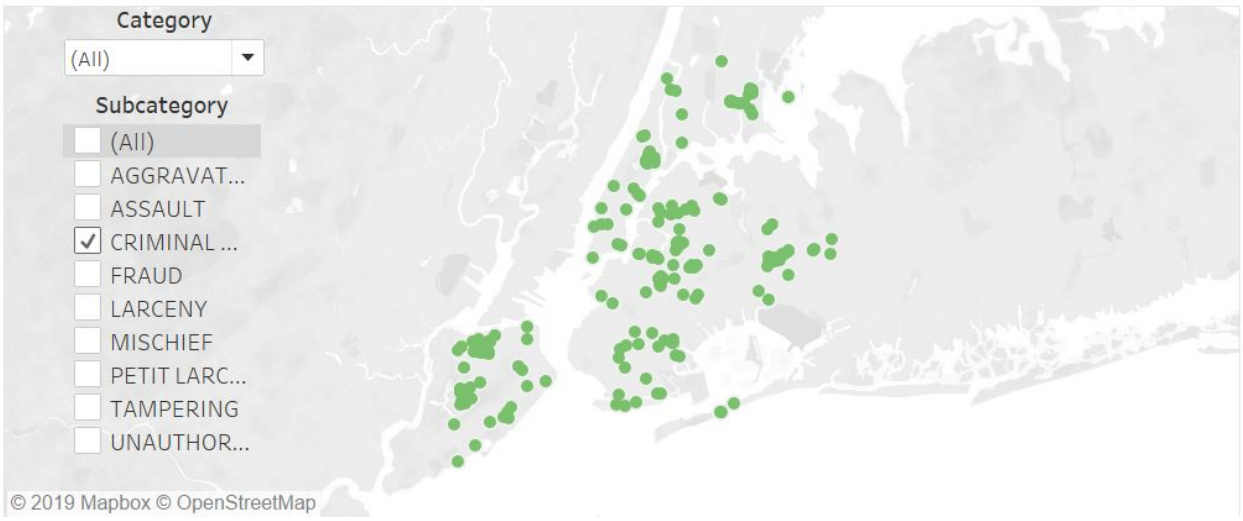


- Misdemeanor

Misdemeanor refers to a less severe type of crime. It shares almost the same subcategories with a felony but with less serious damage. Larceny still accounts for the largest number of crimes, with 1,041 records in one year, which happened a lot last year, primarily in Manhattan, the Bronx, and northeast Brooklyn.



While the largest number of a misdemeanor crime is still stealing, the second one changes to criminal mischief, with records of 239 for the past year.

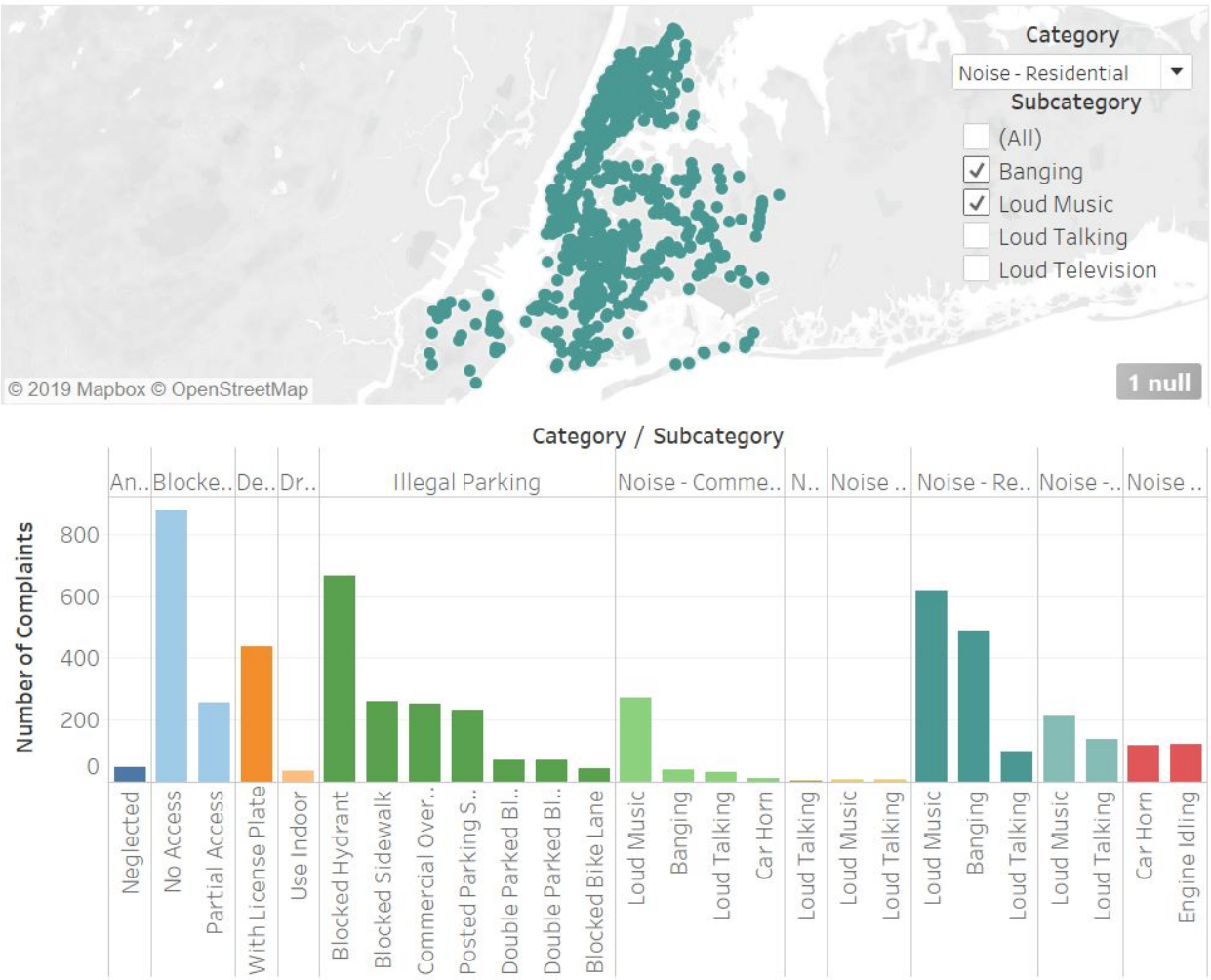


- Violation

The primary subcategory of violation is harassment, referring to behaviors that annoy or upset someone. It happened 620 times in the past year, especially in mid-Manhattan, lower Bronx, and northeast Brooklyn.

Complaint analysis

Complaint analysis focuses on the comfort of living areas with different types of crime and their primary locations.



When talking about comfort and safety, the priority thought maybe peace and quietness. The complaint about loud music and banging, which is the sound of shooting, in-residence happen 618 and 488 times, respectively, in 2018, mainly located in the Bronx, lower Manhattan, and northeast Brooklyn.

Conclusion and Recommendations

Location

- **Bronx**

NYPD should hire people to remove junk cars or ask the owner to move unused cars to prevent illegal parking or blocked driveway. Also, the highest number of crimes in the Bronx is larceny, and the following is harassment. NYPD should focus on these two crimes especially in the Bronx, and dispatch police officers accordingly.

- **Brooklyn**

Brooklyn has the highest number of complaints among the New York City boroughs with approximately over 1,800 filed complaints in 2018. Especially, similar to the Bronx, illegal parking and the blocked driveway are the most common complaints filed in Brooklyn. Additionally, residential noise is the most problematic along with commercial noise and street/sidewalk noise. Therefore, NYPD should dispatch more traffic officers to strengthen traffic law. NYPD can install noise meters in areas with a higher number of noise complaints to monitor the noise level.

For the crimes in Brooklyn, larceny is the most frequently-happened crime, next to larceny, harassment is the second most happened crime in Brooklyn. Therefore, NYPD should focus on these two crimes by deploying more police officers, and also NYPD should put more effort into educating people on how to stop and the serious consequences of committing such crimes.

- **Queens**

The major issues bothering the residents in Queens are illegal parking, blocked driveway, and derelict vehicle. People also complained about the noise from residential a lot. In order to relieve this problem, NYPD should dispatch more traffic officers to control the illegal parking and blocked driveway. For the derelict vehicle, NYPD can hire junk car removal services to further reduce the problem.

The situation in Queens is similar to Brooklyn, larceny is the main crime reported in Queens, followed by harassment. Therefore, NYPD in Queens should also dispatch more police officers and educate people to reduce the crime rate.

- **Manhattan**

In Manhattan, instead of the blocked driveway and derelict vehicle, noise complaint seems to be the most serious problem. In that case, NYPD should reduce the workforce in traffic and increase in noise control. In terms of the crimes that happened in Manhattan, the police should put more attention in reducing larceny, harassment, and fraud since those crimes happened more frequently.

- **Staten Island**

It seems Staten Island is the most peaceful and safe place among all boroughs in New York City. It has the least number of Larceny, with only 96 records in 2018. Also, illegal parking and noise are not a huge problem in the place. Therefore, NYPD can dispatch more officers to other places to reduce costs.

Time of Complaints and Crimes

Most of the complaints were filed from 7 am to 12 am. The graph shows the least complaints were filed in the early morning from 3 am to 4 am. Also, most of the crimes were reported from 7 am to 12 am. The graph also shows that from 1 am to 7 am, there were fewer crimes reported. Therefore, NYPD is recommended to place more officers during these hours to reduce the complaints and crimes.

Race and Sex of Susceptors and Victims

From the chart, it is shown that there is a much higher number of male susceptors than female susceptors. Additionally, for the victims, it is demonstrated that black female is more susceptible to crimes than other race. Next to black females, white female and white Hispanic females are also susceptible to crimes.

Age of suspects and victims

The age of susceptors is the highest in 25-44 years old. And the age of victims the highest in 25-44 years old, and the second highest in 45-65 years old.

Appendix

```
--1. Insert data into NYPD_SEX table.
create sequence sex_seq;
drop sequence sex_seq;

insert into NYPD_SEX(sex_id, sex)
select sex_seq.nextval, sex
    from (
            select distinct susp_sex sex
                from nypd_stage_crime
            union
            select distinct vic_sex sex
                from nypd_stage_crime
            )
    where sex is not null;

select count(*) from NYPD_SEX;
```

```
--2. Create NYPD_RACE table.
create sequence race_seq;

insert into NYPD_RACE(race_id, race)
select race_seq.nextval, race
    from (
            select distinct susp_race race
                from nypd_stage_crime
            union
            select distinct vic_race race
                from nypd_stage_crime
            )
    where race is not null;

select count(*) from NYPD_RACE;
```

```sql
--3. Create NYPD_CATEGORY table.
CREATE SEQUENCE category_seq;
--DROP SEQUENCE category_seq;

INSERT INTO nypd_category (
    category_id,
    category
)
    SELECT
        category_seq.NEXTVAL,
        category
    FROM
        (
            SELECT DISTINCT
                complaint_type category
            FROM
                nypd_stage_complaint
            UNION
            SELECT DISTINCT
                law_cat_cd category
            FROM
                nypd_stage_crime
        )
    WHERE category IS NOT NULL;

select count(*) from nypd_category;

--Create NYPD_LOCATION table.
CREATE SEQUENCE location_seq;
--DROP SEQUENCE location_seq;

INSERT INTO nypd_location (
    location_id,
    borough,
    latitude,
    longtitude

)
SELECT
    location_seq.NEXTVAL,
```

```sql
SELECT
    location_seq.NEXTVAL,
    borough,
    latitude,
    longitude
FROM
    (
        SELECT DISTINCT borough borough, latitude latitude, longitude longitude
        FROM nypd_stage_complaint
        UNION
        SELECT DISTINCT boro_nm borough, latitude latitude, longitude longitude
        FROM nypd_stage_crime
    );

select count(*) from nypd_location
```

```sql
--4.Create NYPD_LOCATION table.
CREATE SEQUENCE location_seq;
--DROP SEQUENCE location_seq;

INSERT INTO nypd_location (
    location_id,
    borough,
    latitude,
    longtitude

)
SELECT
    location_seq.NEXTVAL,
    borough,
    latitude,
    longitude
FROM
    (
        SELECT DISTINCT borough borough, latitude latitude, longitude longitude
        FROM nypd_stage_complaint
        UNION
        SELECT DISTINCT boro_nm borough, latitude latitude, longitude longitude
        FROM nypd_stage_crime
    );

select count(*) from nypd_location
```

```sql
--5. Create NYPD_Date table.
CREATE SEQUENCE date_seq;
--DROP SEQUENCE date_seq;



ALTER TABLE NYPD_STAGE_COMPLAINT rename column TIME to CMPLNT_FR_TM;
INSERT INTO nypd_date (
    date_id,
    incident_date,
    incident_month,
    incident_time

)
SELECT
    date_seq.NEXTVAL,
    in_date,
    in_month,
    in_time
FROM
    (
        SELECT DISTINCT to_date(cmplnt_fr_dt,'DD-Mon-YYYY') in_date, to_char(cmplnt_fr_dt, 'Mon')
        FROM nypd_stage_complaint
        UNION
        SELECT DISTINCT to_date(cmplnt_fr_dt,'DD-Mon-YYYY') in_date, to_char(cmplnt_fr_dt, 'Mon')
        FROM nypd_stage_crime
    );

select count(*) from nypd_date;
```

```
--6.Create NYPD_DESCRIPTION table.
create sequence des_seq;
--drop sequence des_seq;

insert into nypd_description(description_id, category_id, description)
select des_seq.nextval, category_id, description
    from (
            select distinct nc.category_id category_id, nsc.pd_desc description
                from nypd_stage_crime nsc
                left join nypd_category nc on nc.category = nsc.law_cat_cd
            union
            select distinct nc.category_id category_id, nsc.descriptor description
                from nypd_stage_complaint nsc
                left join nypd_category nc on nc.category = nsc.complaint_type
            );

select count(*) from nypd_description;

--7.Create NYPD_SUSPECT table.
create sequence sus_seq;
--drop sequence sus_seq;


insert into nypd_suspect(suspect_id, suspect_age, suspect_race_id, suspect_sex_id)
select sus_seq.nextval, suspect_age, suspect_race_id, suspect_sex_id
    from (
            select susp_age_group suspect_age, nr.race_id suspect_race_id, ns.sex_id suspect_sex_id
                from nypd_stage_crime nsc
                left join nypd_race nr on nr.race = nsc.susp_race
                left join nypd_sex ns on ns.sex = nsc.susp_sex
                where nr.race_id is not null
            );

select count(*) from nypd_suspect;

--8.Create NYPD_VICTIM table.
create sequence vic_seq;
--drop sequence vic_seq;


insert into nypd_victim(victim_id, victim_age, victim_race_id, victim_sex_id)
select vic_seq.nextval, victim_age, victim_race_id, victim_sex_id
    from (
            select vic_age_group victim_age, nr.race_id victim_race_id, ns.sex_id victim_sex_id
                from nypd_stage_crime nsc
                left join nypd_race nr on nr.race = nsc.susp_race
                left join nypd_sex ns on ns.sex = nsc.susp_sex
                where nr.race_id is not null
            );

select count(*) from nypd_victim;
```

```
--9.Create NYPD_COMPLAINT_FACT table.
create sequence complaint_seq;
--drop sequence complaint_seq;

insert into nypd_complaint_fact(complaint_id, date_id, location_id, description_id, number_of_comp
select complaint_seq.nextval, date_id, location_id, description_id, 1
    from (
            select distinct nsc.key, nd.date_id, location_id, description_id
                from nypd_stage_complaint nsc
                left join nypd_date nd on nd.incident_date||nd.incident_time = nsc.cmplnt_fr_dt||r
                left join nypd_location nl on nl.latitude||nl.longtitude = nsc.latitude||nsc.longi
                left join nypd_description nde on nde.description = nsc.descriptor
            );

select count(*) from nypd_complaint_fact;

select count(*) from
(
    select distinct nsc.key, nd.date_id, location_id
        from nypd_stage_complaint nsc
        left join nypd_date nd on nd.incident_date||nd.incident_time = nsc.cmplnt_fr_dt||nsc.cmplr
        left join nypd_location nl on nl.latitude||nl.longtitude = nsc.latitude||nsc.longitude
--        left join nypd_description nde on nde.description = nsc.descriptor
);

create view nypd_complaint_loc_view as
select nl.location_id
    from nypd_stage_complaint nsc
    left join nypd_location nl on nl.latitude||nl.longtitude = nsc.latitude||nsc.longitude;

--Check the number of complaints.
select complaint_id, num_of_inc
    from
    (
        select complaint_id, count(complaint_id) num_of_inc
            from nypd_complaint_fact
            group by complaint_id
    )
    where num_of_inc > 1;
```

```
--10. Create NYPD_COMPLAINT_FACT table.
select count(*) from
(
    select nd.date_id, nl.location_id, nc.category_id
        from nypd_stage_crime nsc
        left join nypd_date nd on nd.incident_date||nd.incident_time = nsc.cmplnt_fr_dt||nsc.cmplr
        left join nypd_location nl on nl.latitude||nl.longtitude = nsc.latitude||nsc.longitude
        left join nypd_description nc on nc.description = nsc.pd_desc
);
```