homework vi

Sidhartha Amperayani, Moinuddin Memon, Amritha Menon

August 2, 2022

Contents

INTRODUCTION

Data helps a person make informed decisions. Data analysis is a crucial and effective step in decision-making and predicting trends and behavior. Visual Analytics makes studying data a lot simpler and more accurate. Visual Analytics fascilitates enterprises / individuals understand the market better to improve their system in a a targeted fasion. With the use of visualizations, non-tehnical stakeholders can understand the results better.

The exploratory data analysis of the NYC 311 Service Request dataset helps in analyzing the performance of different agencies in NYC when it comes to handling non-emergent requests. Through this, we can seek answers to various questions as to how the complaints are distributed, which complaint is most prevalent, which Borough makes most complaints and a lot more. These answers will help us in identifying different correlations between different aspects of the data. NYPD Complaints dataset from 2010 to 2015 is also used along with the 311 dataset to take a deeper dive to examine how NYPD handles its complaints. Through merging this data one can understand better how about the trends in crime over several boroughs. NYPD is a department that helps the city maintain law and order. Through this project we try to understand what crimes are being commited in the city and how well are they being handled. While working on this project, we understood the functionality of seeral agencies their day to day service request natures.

Two datasets are being used here. The NYC 311 Service Request dataset consists of service requests made from the years 2010 to 2019. However, we use service requests between the years 2010 and 2015 for data analysis. Services provided by NYPD, DOT, DPR, DCA, HRA and several other agencies in all the 5 Boroughs of the city, Bronx, Brooklyn, Manhattan, Queens and Statten Islands. The dataset contains 41 columns which has attributes like date of complaint creation and resolution, type of complaint, agency handling the complaint, address, and status of the case. Only the relevant attributes among 41 were selected for analysis.

This dataset was obtained from https://nycopendata.socrata.com/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9.

To help us better understand the the way a NYPD works, we import the NYPD complaints data-set which represents all the crime reported to NYPD. Through this dataset, we cn find the nature of crimes and when they were commited in the city. We can also learn how fast the complaints are resolved in different boroughs of the city. We sampled this from NYC Open Data, which can be found here: https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i

The dataset's attributes describes details related to the type of crime (valid felony, misdemeanor, and violation crimes) that has been reported to the New York City Police Department from 2006 to 2019. Since the 311 Service Request data contains data between the years 2010 and 2015, we have filtered crimes reported between the years 2010 and 2015 from NYPD Complaints dataset. This dataset includes date, location of the crime, victims ad suspects age group and other relevant features pertaining to the crime. Combining these datasets helped us analyse the complaints better.

PARAMETERS

All the parameters required for this file are initialized here like if the file name, sample number of rows, if a full run or test run must be made and also the key for the Google map API.

| CreatedDate	ClosedDate	Agency	AgencyName	ComplaintType	Descriptor	IncidentZip	StreetName	AddressType	Status	Borough	XCoordinate(StatePlane
04/14/2015 02:14:40 AM	04/14/2015 03:03:22 AM	NYPD	New York City Police Department	Vending	In Prohibited Area	10465	EAST TREMONT AVENUE	ADDRESS	Closed	BRONX	103375
04/14/2015 02:10:12 AM		NYPD	New York City Police Department	Blocked Driveway	No Access	11234	RYDER STREET	ADDRESS	Open	BROOKLYN	100154
04/14/2015 02:03:01 AM		NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	11204		INTERSECTION	Open	BROOKLYN	98467
04/14/2015 02:02:40 AM		NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Talking	11211	METROPOLITAN AVENUE	ADDRESS	Assigned	BROOKLYN	99647
04/14/2015 02:00:04 AM	04/14/2015 02:47:33 AM	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Talking	10025		INTERSECTION	Closed	MANHATTAN	99426
04/14/2015 01:52:15 AM	04/14/2015 02:11:10 AM	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Talking	11205		INTERSECTION	Closed	BROOKLYN	99400

INITIALIZATION

The packages to be used like tidyverse, ggplot, dplyr etc. are installed. The program runs according to the user's setting of **testRun** variable. All the irrelevant columns are removed and the duplicate rows are also removed. In the **Borough** column the **Unspecified** value is changed to the respective borough name based on the given zip codes and our knowledge of NYC. The date values are formatted from date type to integers for easy usage.

[1] 872266

FINDINGS

311 DATASET

TOP 5 COMPLAINTS ACROSS NYC

Top 5 Complaint Distribution across NYC
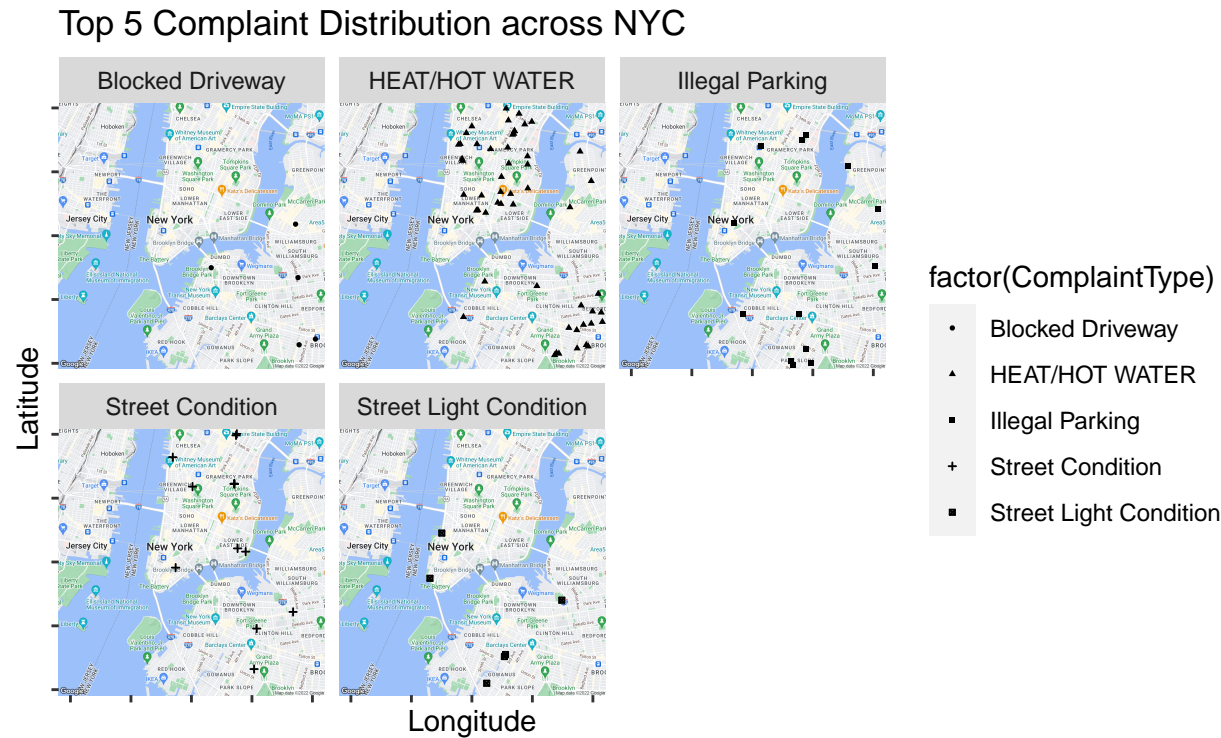
We can see that the top 5 complaints are Heat/Hot water, Water System, Blocked Driveway, Street Condition, and Street Light Conditions are the top 5 complaints in that order, based on their density distribution across the NYC map.

TOP 5 AGENCIES (BASED ON CLOSED CASES)

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

Top 5 Agencies in terms of Closed Cases
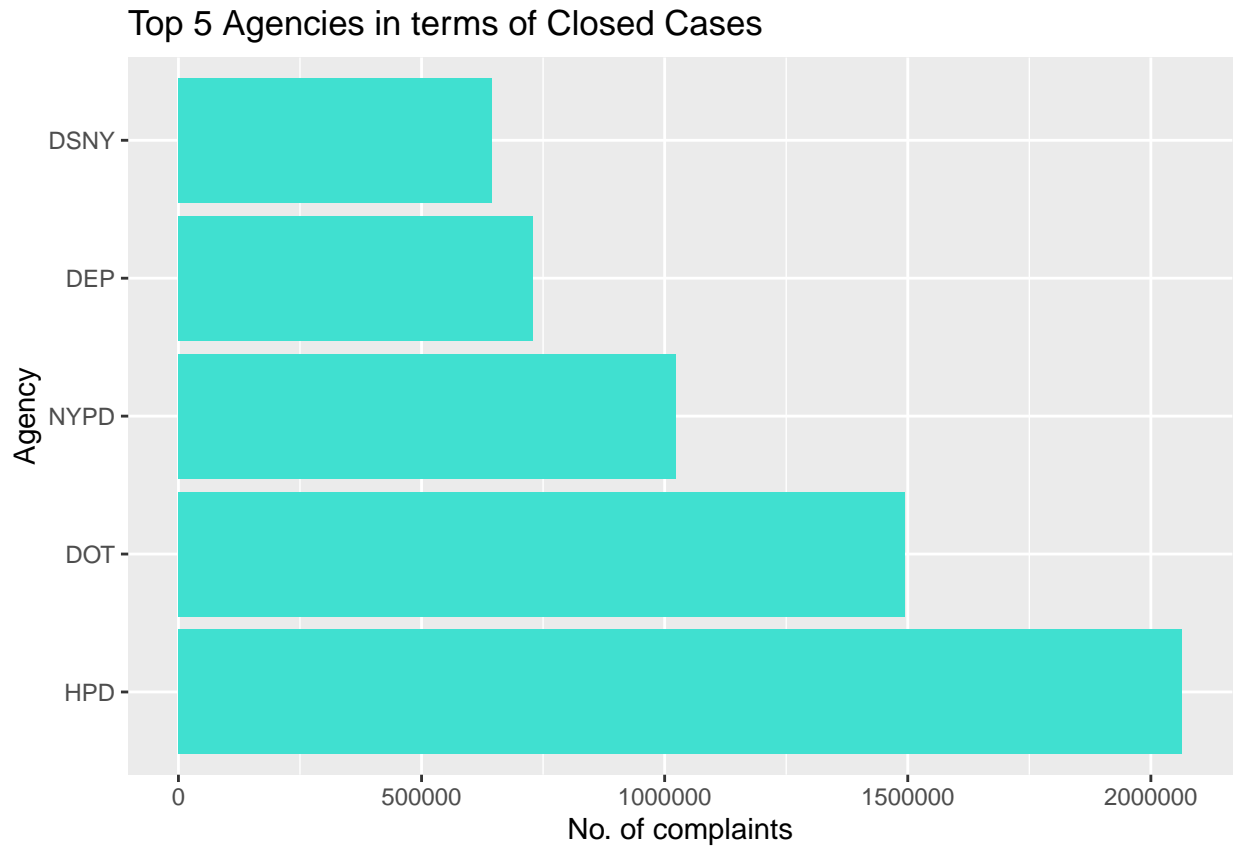
The HPD department, (Department of Housing, Preservation and Development) is the Agency which has closed the most cases, followed by DOT (Department of Transportation).

NYPD COMPLAINTS IN EACH BOROUGH

```
##
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':
##
##     discard

## The following object is masked from 'package:readr':
##
##     col_factor
```

NYPD Complaints in each Borough
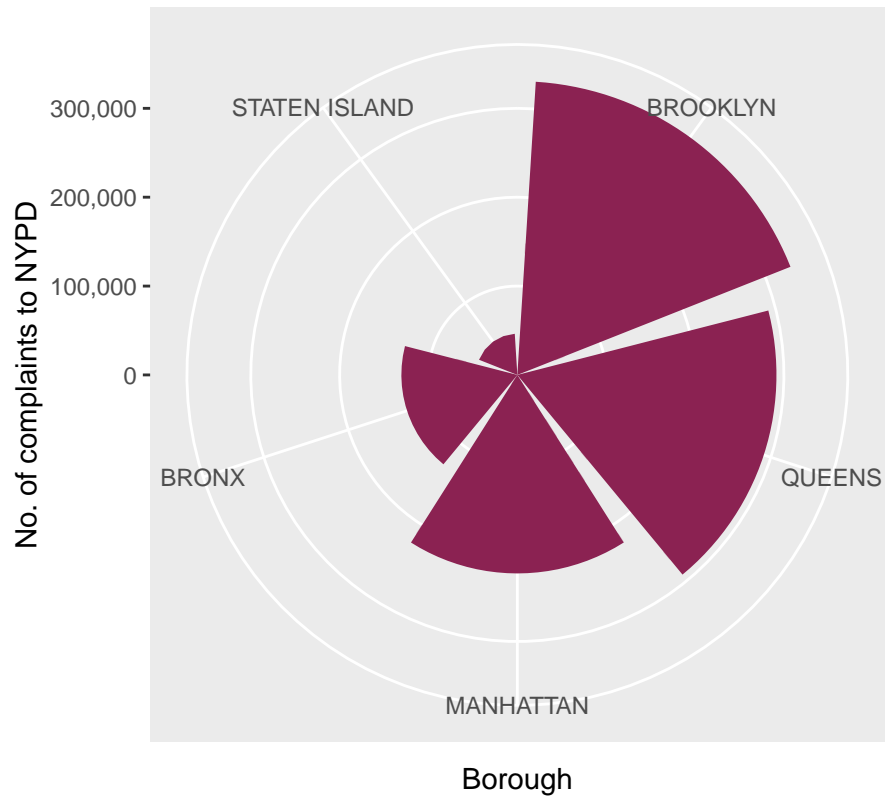
The Borough - Brooklyn has a lot of cases to which NYPD needs to respond.

# NOISE COMPLAINT DISTRIBUTION IN EACH BOROUGH

Distribution of Noise Complaints
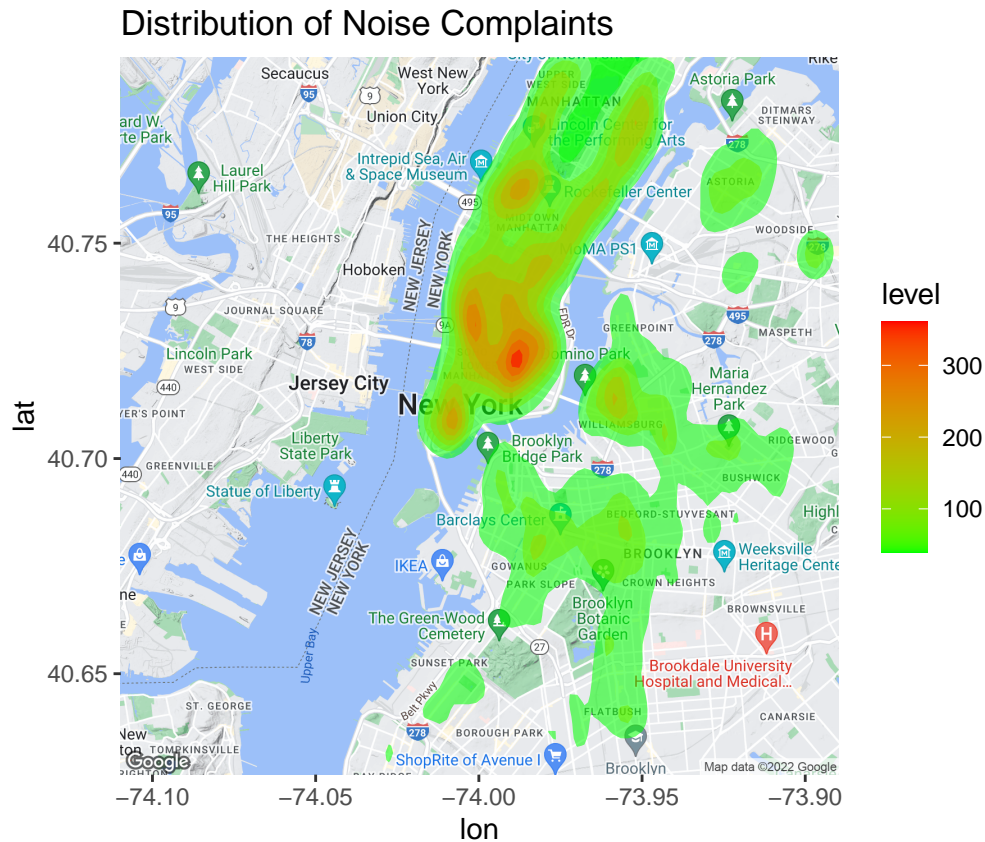
From the different types of noise complaints, we can see that helicopter noise, noise from house of worship, parks, streets and vehicles are on the lower side. But commercial noise and other noise complaints seem to be higher, especially in Manhattan. Staten Island and Bronx seem to have less number of noise complaints.

TOTAL NUMBER OF COMPLAINTS, MONTH WISE

Total Complaints By Month
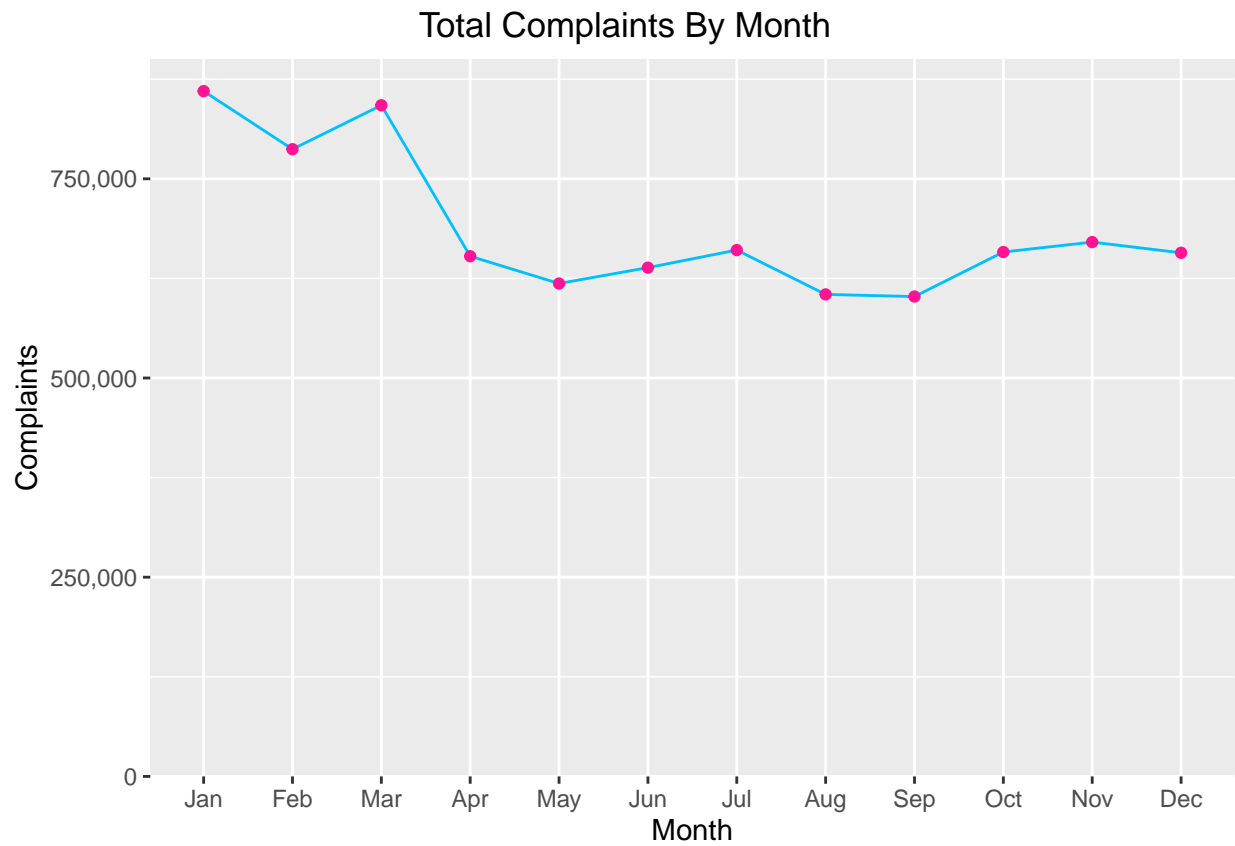
From the line graph, it is seen that the total complaints are higher during the month of January.

PENDING CASES IN EACH AGENCY

Top Total pending cases by agency
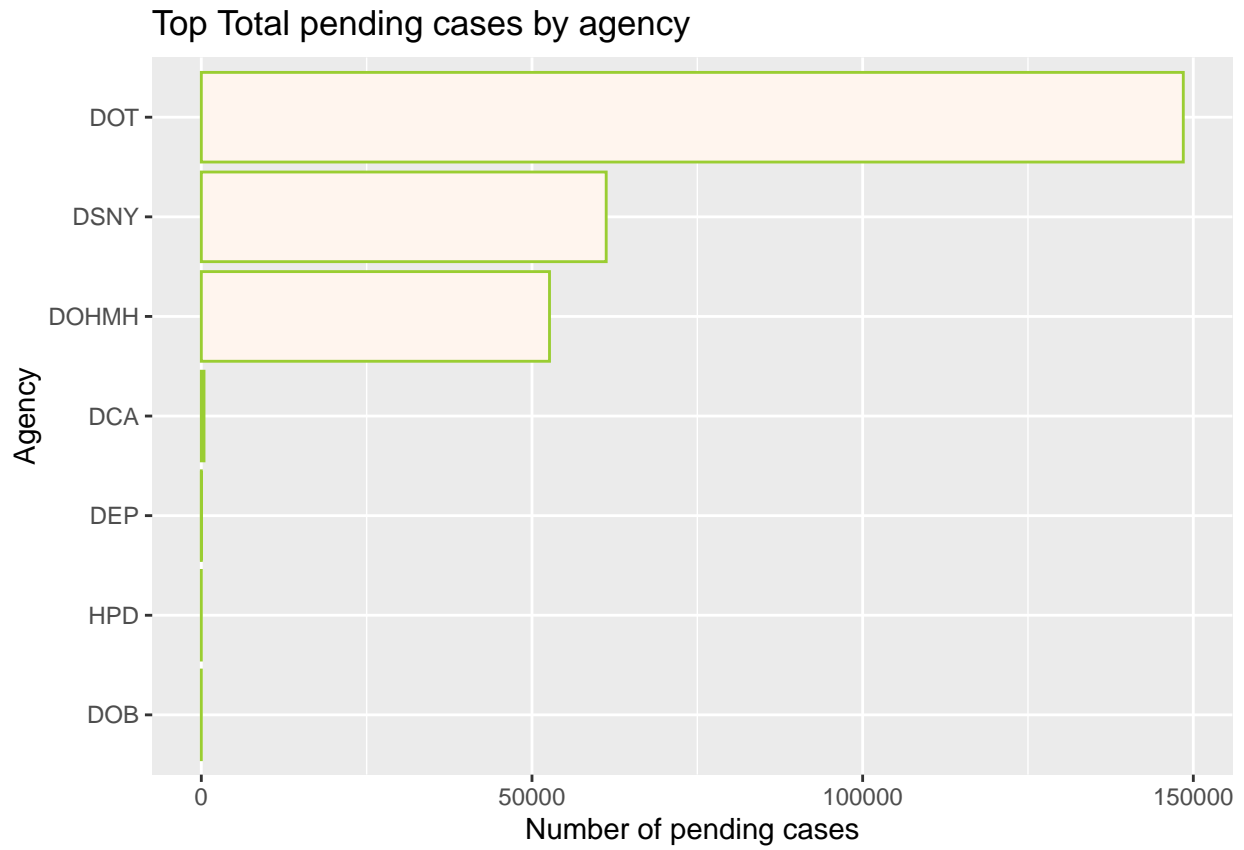
The agency with the most pending cases is DOT (Department of Transportation).

AVERAGE RESOLUTION TIME, AGENCY WISE

Average resolution time in hours Agency wise
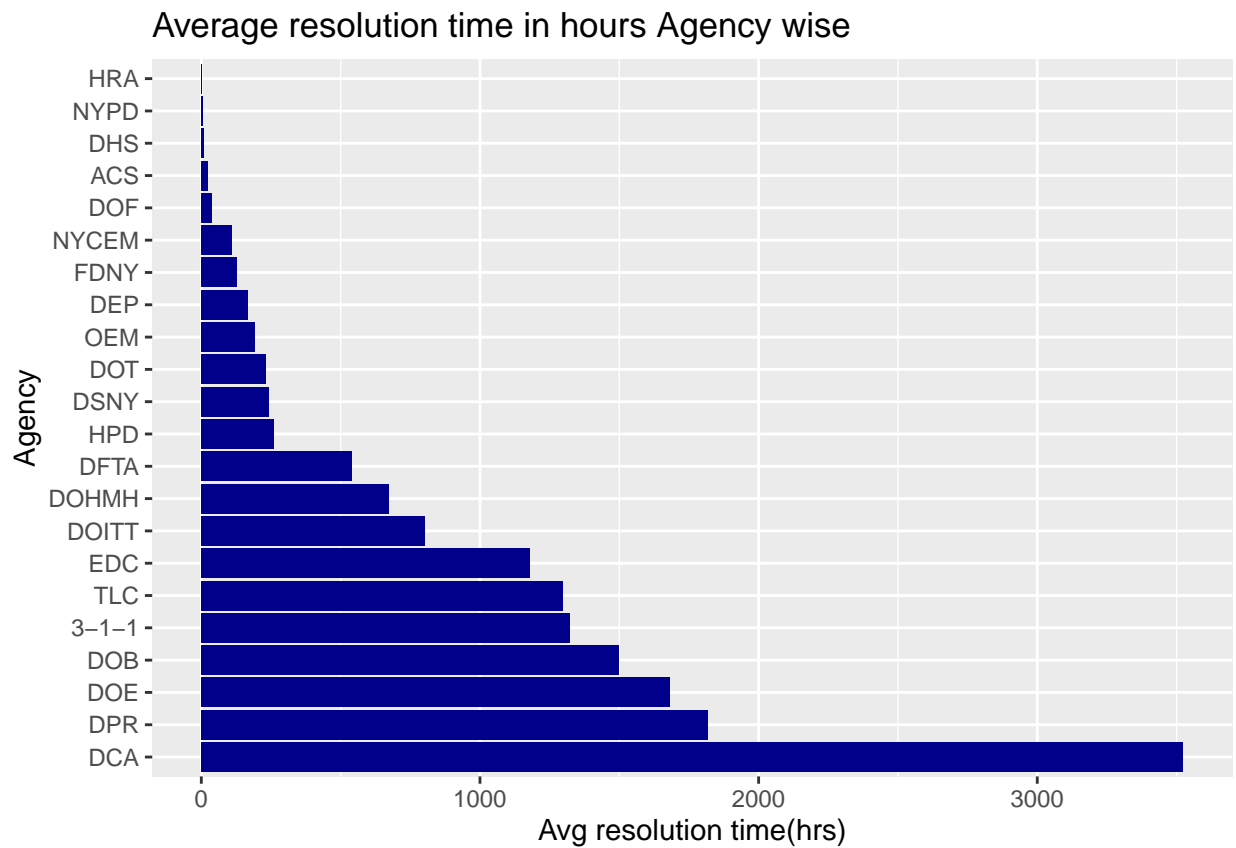
From the average resolution time (in hours) calculated for each agency, it is seen that the Agency DCA (Department of Consumer Affairs) has the highest resolution time.

COMPLAINTS BY THE HOUR
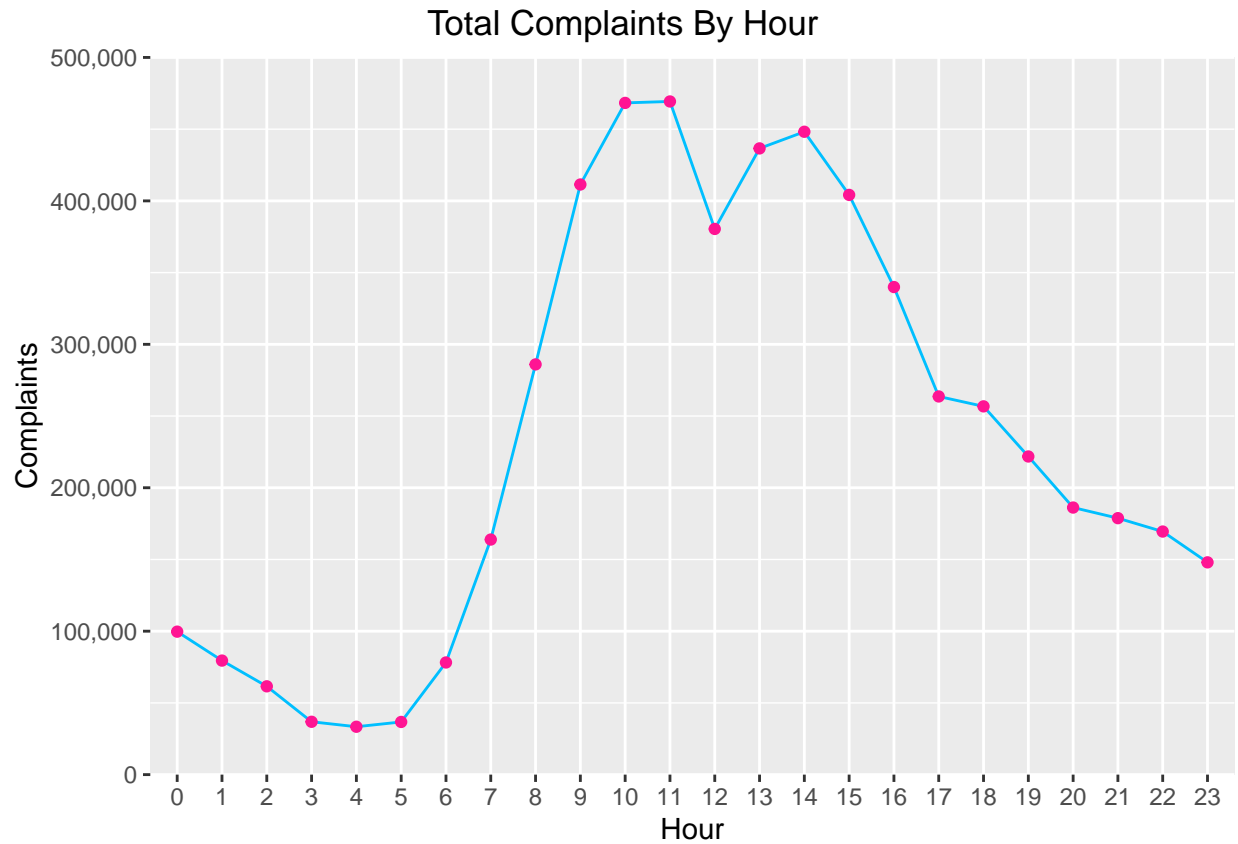
We can find in the pattern of the number of complaints reported against the hours in a day that the peak time of the reporting is between 9-12. We also removed the auto-generated complaint times which were set to 0 hour by default by checking against the minute value, If it was also 0 then it was by default and can be removed.

TOP COMPLAINT WORDS

Top words used in Descriptor Field by Borough
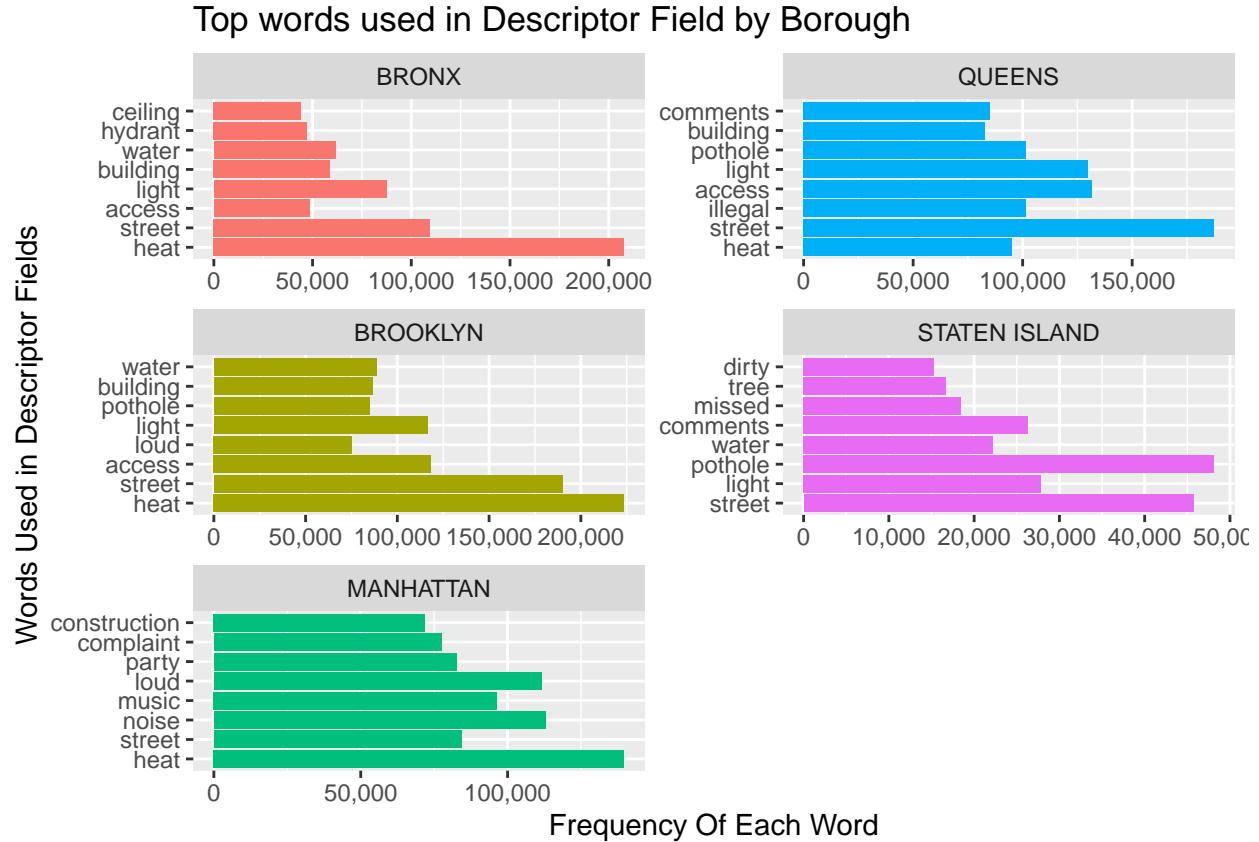
We follow the word counts in the descriptor field of the data-set to get insights about how the complaints are reported. As we can see, "heat" is a keyword found the most in the descriptor field across the Boroughs Bronx, Brooklyn,Manhattan while "street" is found most in Staten Islands and Queens complaints. These insights give an idea about the context in which the complaints are reported and are very insightful.

NYPD COMPLAINT DATASET

NYPD CRIMINAL COMPLAINT DATA SET

CMPLNT_FR_DT	CMPLNT_FR_TM	OFNS_DESC	PD_DESC	LAW_CAT_CD	BORO_NM	SUSP_AGE_GROUP	SUSP_RACE	SUSP_SEX	Latitude	Longitude	VIC_AGE_GROUP
2013-10-18	20:34:00	MURDER & NON-NEGL. MANSLAUGHTER		FELONY		45-64	WHITE HISPANIC	M	40.67253	-73.89603	18-24
2014-03-20	12:40:00	MURDER & NON-NEGL. MANSLAUGHTER		FELONY		45-64	BLACK	M	40.90549	-73.84845	18-24
2014-04-26	21:42:00	MURDER & NON-NEGL. MANSLAUGHTER		FELONY					40.88216	-73.83191	18-24
2015-06-27	00:05:00	MURDER & NON-NEGL. MANSLAUGHTER		FELONY					40.79679	-73.93940	25-44
2013-06-09	21:00:00	MURDER & NON-NEGL. MANSLAUGHTER		FELONY					40.66668	-73.88156	25-44
2012-01-01	12:00:00	THEFT-FRAUD	FRAUD,UNCLASSIFIED-FELONY	FELONY	BROOKLYN				40.62003	-74.01311	<18

NYPD CRIMES COMPLAINT BY MONTH

Total NYPD Complaints By Month
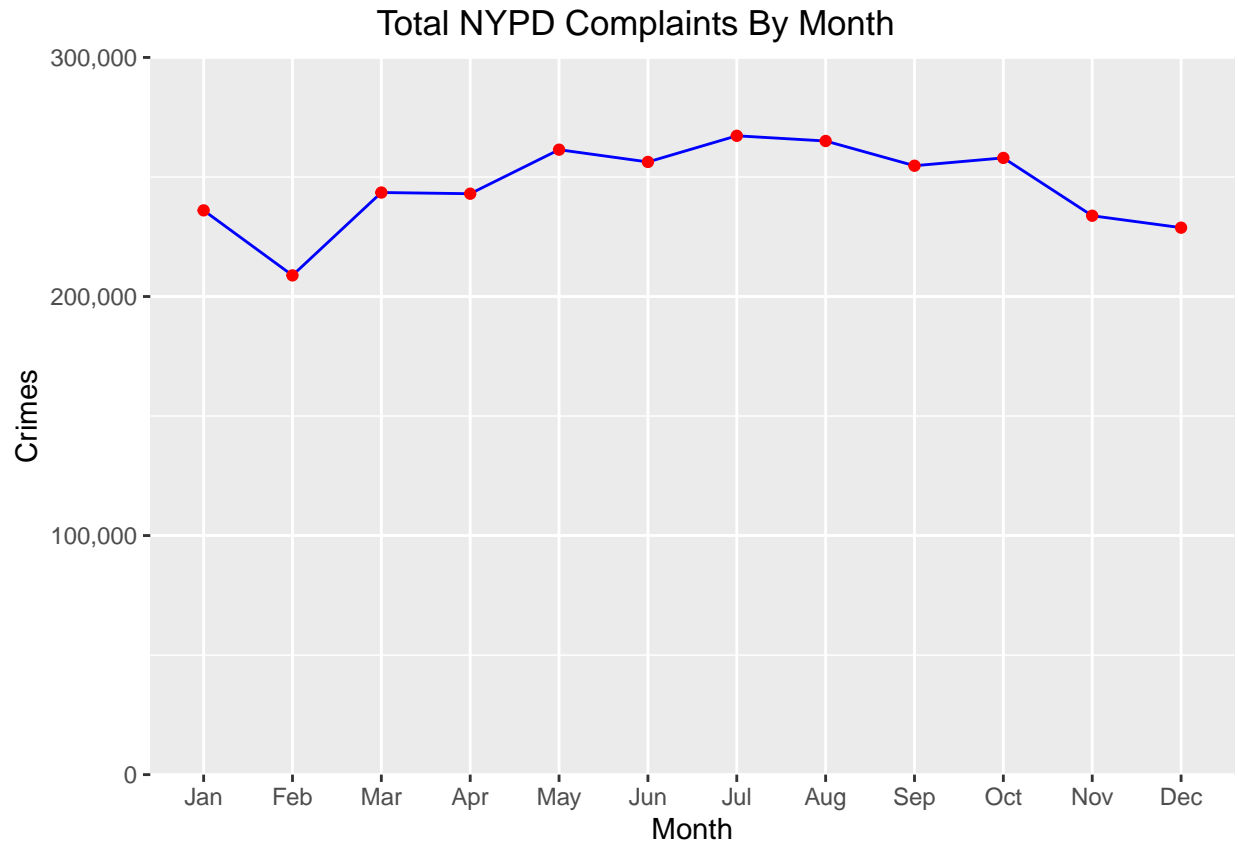
The graph tells us that the colder months in NYC (November, December, January, and February) the criminal complaints dip down but otherwise the crime rate remains high at a peak in NYC.

NYPD COMPLAINTS BY HOUR

Total Complaints By Hour
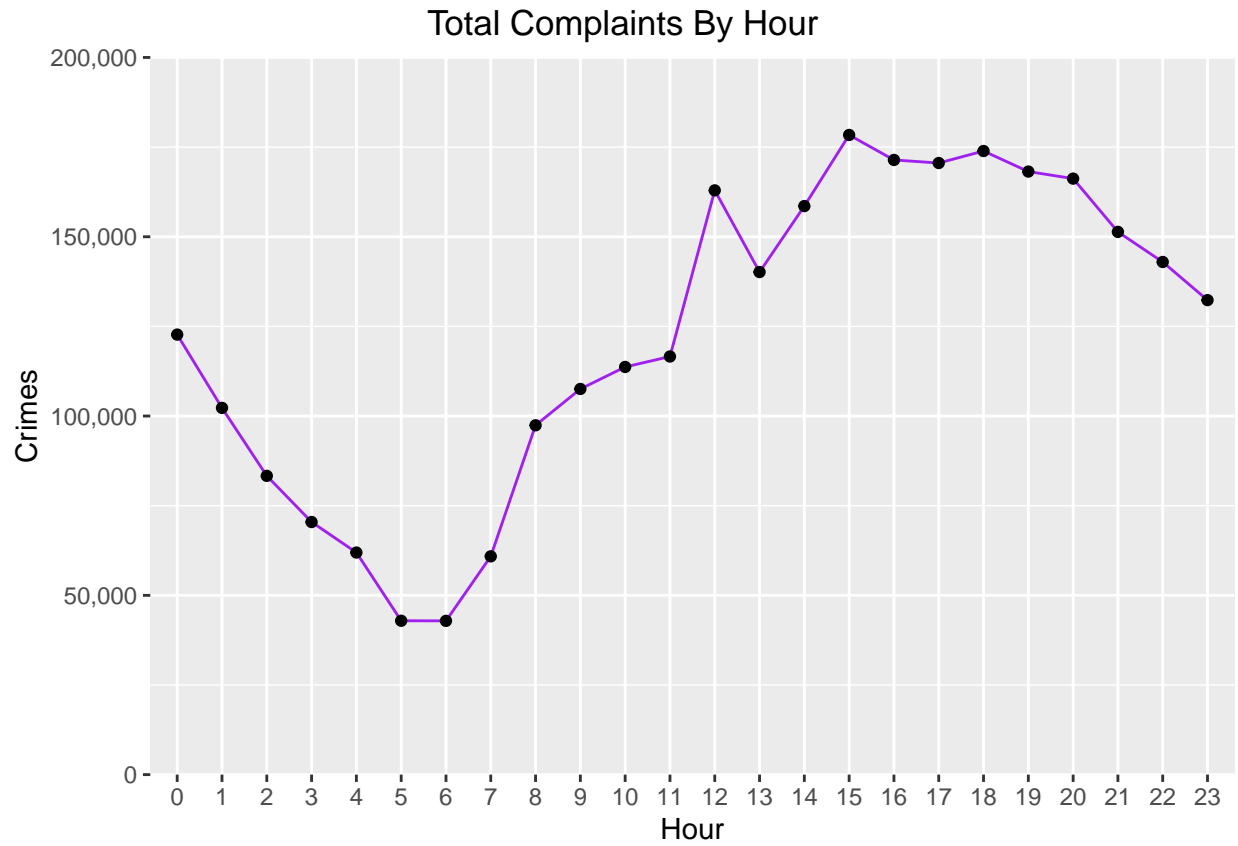
Here, all the complaints that were reported exactly at 12 is removed. The crime keeps rising from the morning and is at its peak during the evening hours.

NYPD CRIMES BY YEAR

Total NYPD Criminal Complaints By Year
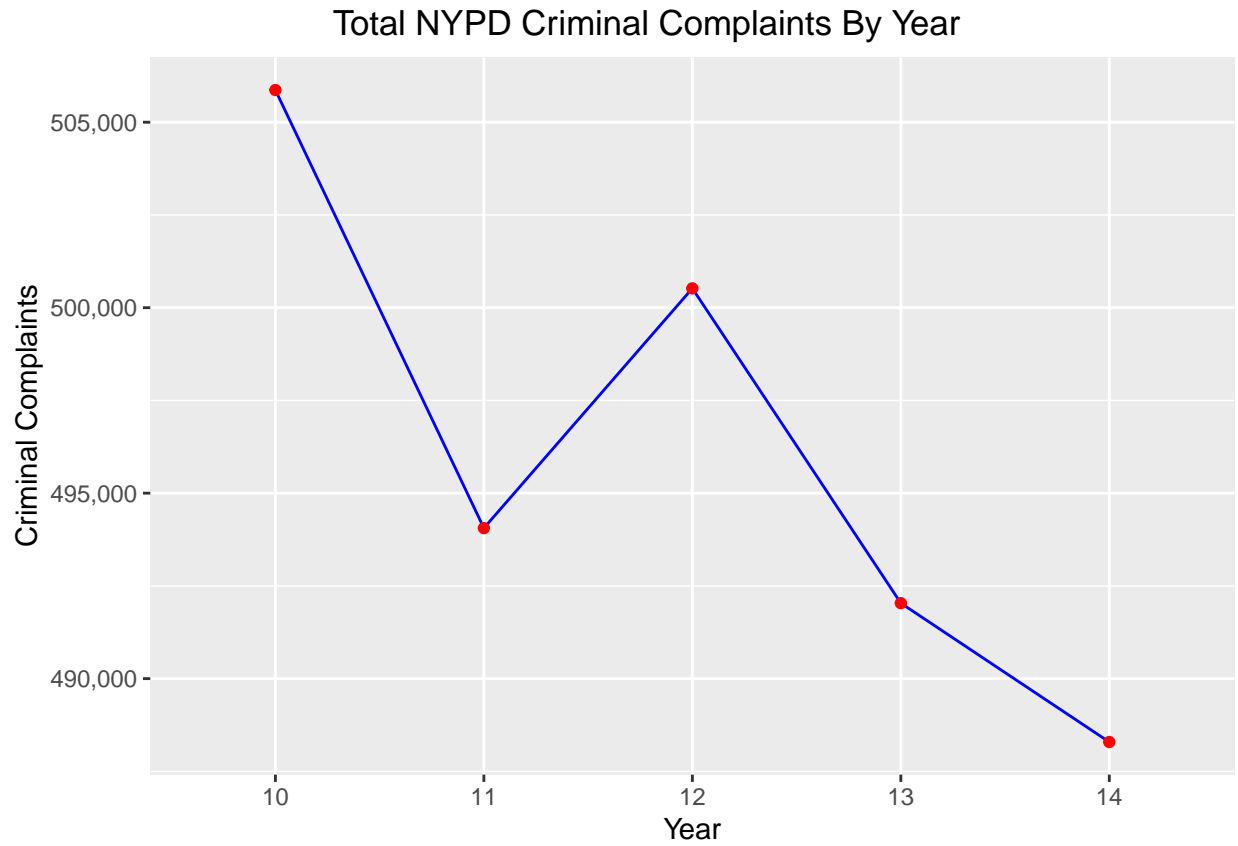
Over the years from 2010 to 2014 the crime count seems to be decreasing which is a good thing which means the NYPD is doing a good job.

TOP 10 CRIMES IN NYC

Top 10 Crimes
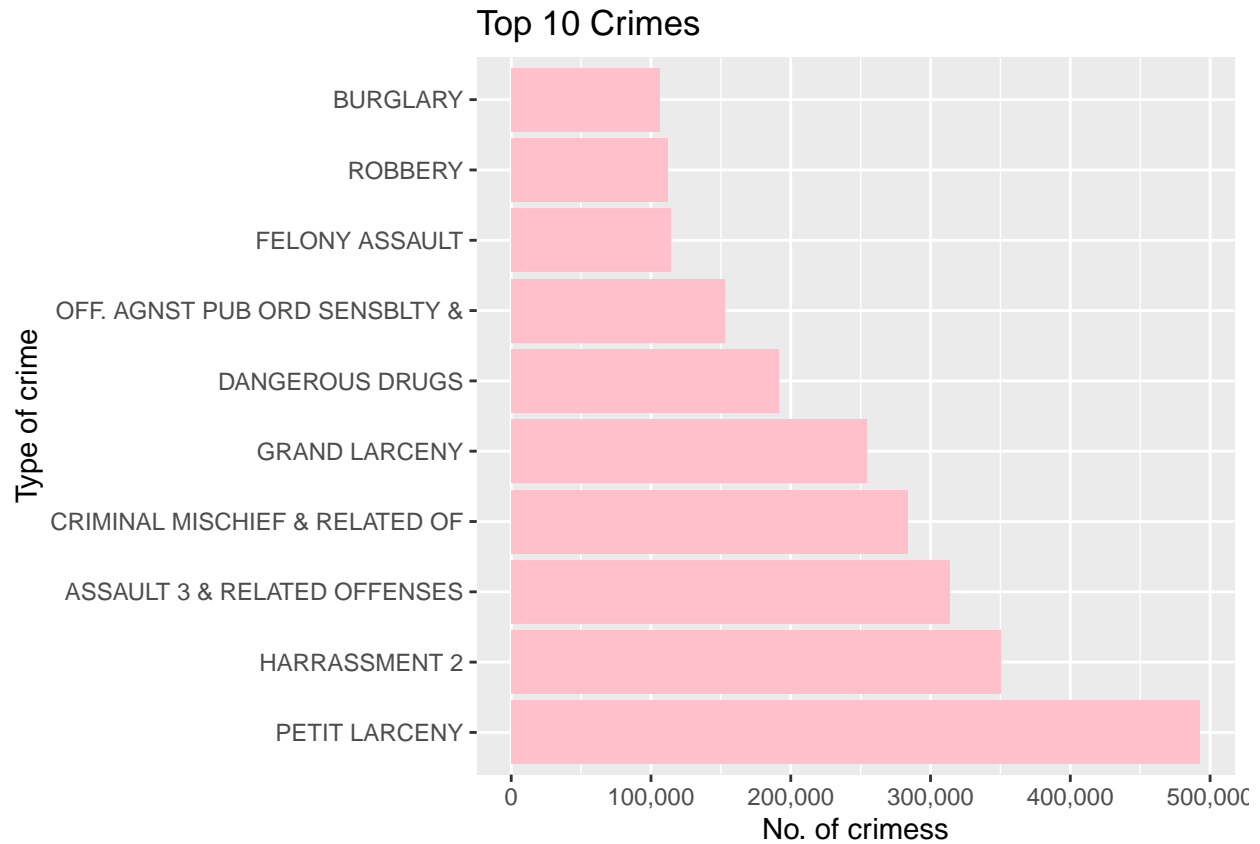
The number one crime in NYPD is petite larceny with felony assault being the lowest

CRIMINAL COMPLAINT TYPES ACROSS BOROUGH

```
## `summarise()` has grouped output by 'BORO_NM'. You can override using the
## `.groups` argument.
## Coordinate system already present. Adding new coordinate system, which will
## replace the existing one.
```

Criminal Complaints in each Borough
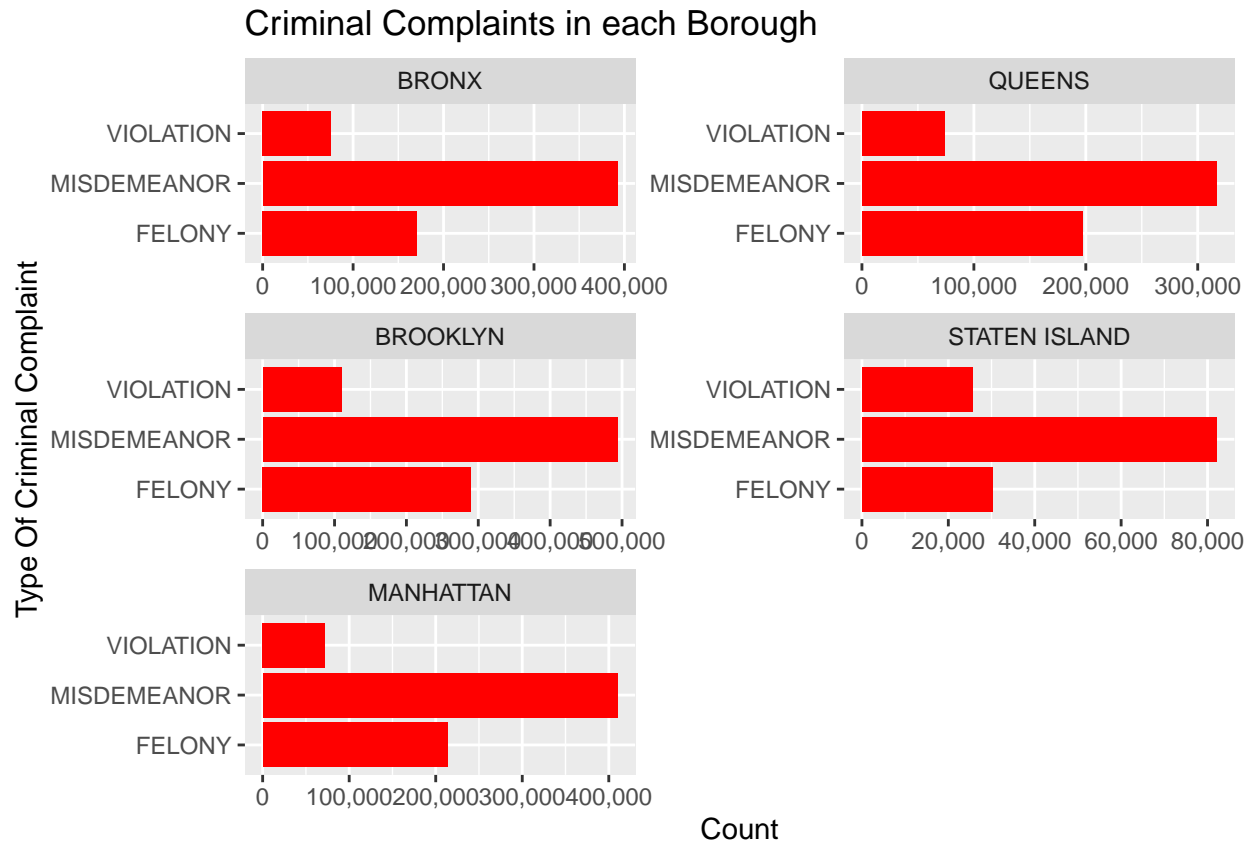
Brooklyn has the most number of criminal complaints reported, whereas Staten Island's criminal complaint counts are really small when compared to Brooklyn. Also, the Misdemeanor criminal complaint type is the highest among all the Boroughs.

MERGING BOTH DATASETS

Join on Borough, Year, Month, Day

```
## `summarise()` has grouped output by 'Borough', 'Created_year', 'Created_month'.
## You can override using the `.groups` argument.
## `summarise()` has grouped output by 'Borough', 'Year', 'Month'. You can
## override using the `.groups` argument.
## `summarise()` has grouped output by 'Borough'. You can override using the
## `.groups` argument.
## `summarise()` has grouped output by 'Month'. You can override using the
## `.groups` argument.
```

JOINED DATA SET

Borough	Year	Month	Day	Total_Complaints	Total_Offenses
BRONX	10	Jan	1	36	462
BRONX	10	Jan	2	32	264
BRONX	10	Jan	3	28	264
BRONX	10	Jan	4	37	237
BRONX	10	Jan	5	31	303
BRONX	10	Jan	6	26	321

Borough name, year, month and, day are the attributes that as a team we collectively decided to join the datasets upon. With the help of this join, it would enable us to make a variety of informative visualizations.

First, a snapshot of the agency wise complaints from the 311 dataset with the attributes Agency, Borough, Created_day, Created_month, Created_year, ComplaintType and filtered it by just taking NYPD complaints with years greater than or equal to 2010. Then the data was grouped based on the total complaints for each row. A subset of the NYPD complaint dataset was also taken with the attributes BORO_NM, YEAR, MONTH, DAY, LAW_CAT_CD. The column names of this dataset was changed to the name of corresponding column name of th 311 dataset. All the months in both the datasets were converted from numbers to the respective months. The two datasets were joined using left join by the Borough, Year and, Month. The total number of offense were calculated by the same Borough, Year and, Month and the joined dataset is grouped by the total number of offenses. So now joined_data contains both the dataset.

NO. OF COMPLAINTS VS CRIMES BOROUGH WISE

No. Of Complaints And Criminal Complaints by Borough

17
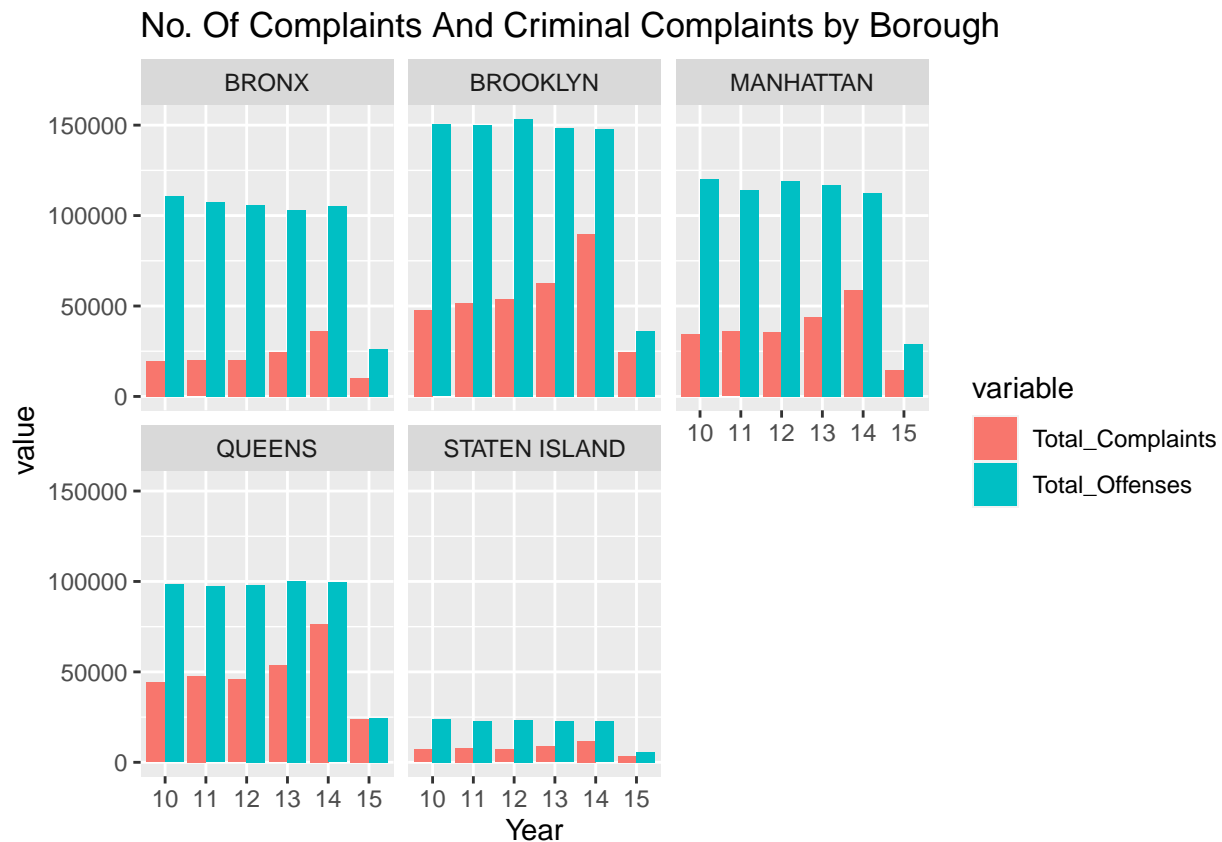
From the above scatter plot, it can be seen that the number of offenses and 311 requests are always low in Staten Island. In Brooklyn and Manhattan both keep rising with the number of offense count being higher than 311 request counts at most times. Also, one interesting observation is that the rate of Offenses in Queens stay the same even though the Complaints vary each year.

NO. OF COMPLAINTS VS CRIMES MONTH WISE

311 Requests vs Crimes in NYC Month wise
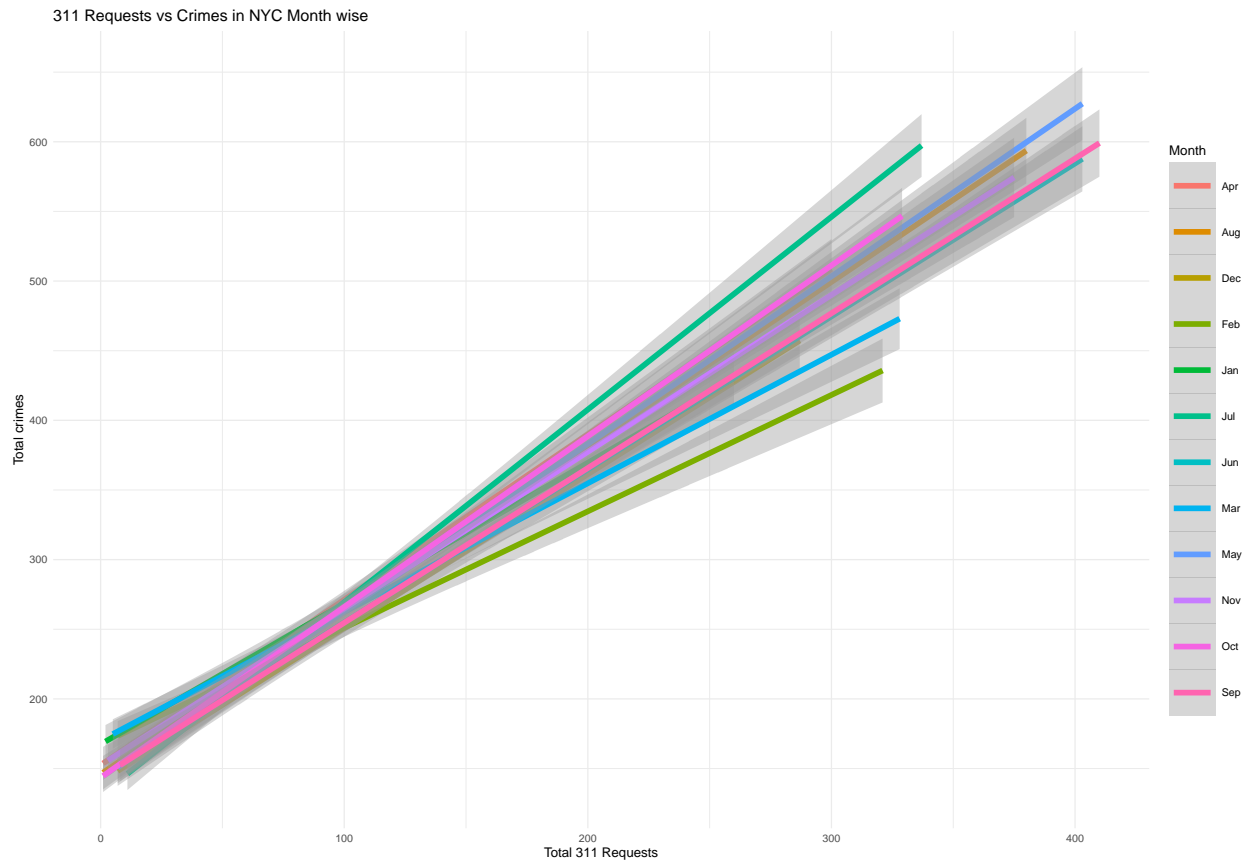
The smooth plot helps in understanding the relationship between the 311 requests and crimes in NYPD which happens to increase over the months but the number of offense is always more than the total number of 311 requests.

CORRELATION BETWEEN 311 COMPLAINTS & CRIMINAL COMPLAINTS PER YEAR


Correlation of Joined Complaints and Crimes per Year

CONCLUSION

Through this assignment we were able to analyze the NYC 311 service request dataset and explore more about and find new information. We asked a set of questions and through the visualizations made, we were able to answer those questions.

Initially, we cleaned the data and pre-processed it and made it ready for analysis. Using the latitude and longitude columns from the dataset we marked the areas of the top 5 complaints across NYC. It must be noted that, for this particular analysis we used a sample of the dataset which has 400,000 rows. Amongst the top complaints, **Heat/Hot water** complaint is prevalent throughout NYC. Following Heat/Hot Water complaint is **Water System** compliant which is equally distributed throughout the city, although less dense than the former complaint. An interesting observation we made is that in the northern region of NYC (Upper Manhattan), although being known as one of the busiest locations in the world, there are hardly any complaints of **Blocked Driveways.**

We looked into the efficiency of top 5 agencies based on the closed cases. **The Department of Housing Preservation and Development** has closed the most cases. We also analyzed the pending cases of all the agencies and saw that the Agencies **DOT, DSNY,** and **DOHMH** have at least 50000 complaints, topping the list is the **Department of Transportation (DOT)** with registered pending complaints close to 150,000. While the remaining agencies have almost no pending cases but it could also mean the case could be pending but the Status could be **Assigned** or **Email sent**, etc.

Amongst all the five boroughs, most complaints were registered to **NYPD** were from **Brooklyn** and **Queens**. The lowest number of complaints originate from **Staten Island.** While studying the dataset, we understood that under the umbrella of noise complaints there are several varieties of noise complaints. To analyze in depth, we made a visualization which explains all the different noise complaints registered from different boroughs. According to the previous visualization, although Queens and Brooklyn have the most complaints registered to NYPD, Manhattan is where top 3 noise complaint types (Noise, Noise-Commercial, andNoise-Street/Sidewalk) are registered.

In NYC, the weather can get very cold during the first quarter of the year, this is the period where the most number of complaints are registered with January being the highest. This could be because of the high volume of complaints due to **Heat/Hot Water.**

We made an important visualization which talks about how efficiently all the agencies resolve their respective service requests. We calculated the average resolution time taken (in hours) for each agency. **The Department of Consumer Affairs** take the most time to resolve their cases and is not very efficient.

We were also able to localize the time slot between which most of the complaints are reported in a day. The time around noon between 9-12 a.m is where most of the complaints are reported which makes sense as these are normal working hours of any agency office. One more peculiar observation is that even after normal office hours there is a steady incoming of complaints for more than 2 hours after which it declines rapidly.

We were also able to categorize the popular descriptor words used while complaining across the Boroughs. We see a divison of Complaint words across the 2 groups of Boroughs viz Manhattan,Bronx,Brooklyn and Staten Island and Queens, which gives us a better idea of where to replicate which resolution strategy.

We explored the NYPD_Complaints_Data_Historic dataset. From this dataset we could get an understanding of the crimes happening in NYC and how NYPD is able to handle them. Over the years from 2010 to 2015, the number of crimes keep decreasing showing us that NYPD works very hard when it comes to handling both emergent and non emergent complaints. From the visualizations made, it is evident that Brooklyn has the more number of crimes happening which makes sense because there are a variety of people living there and is a very contrasting to Staten Island where the crime rate is low which could be because it is primarily a peaceful suburban area.

When looking at crimes by each hour, it can be seen that the crime rate keeps increase from morning to evening with the evening time being the prime crime time and it slowly decreases as night progresses.

Using dplyr package, we were able to perform a join between the nyc311 complaint and the nypd crime dataset. We were able to combine the data by projecting only the data between the year 2010 and 2015. Based on this , we could find that there were many columns like Boroughs, Year, Month, Time, etc. that we can use in joining the two datasets. We cleaned the new data in the same manner as we did for the nyc311 datasets in earlier assignments. We also created custom columns like the Year,month, day from the complaint time column of this dataset. As directed we created several short tables to signify the quality of the join on the datasets. So for a given Borough and time , we can point out how many complaints on the 311 helpline have been registered over the years and months,etc. and also report the crime statistics across the same demography.

Looking at the joined data, the relationship between the number of 311 requests and number of offenses were observed. The number of crimes are always higher than the number of 311 requests which is the scary part to live in NYC. Looking the relationship Borough wise, Brooklyn stands the highest at both emergent and non-emergent complaints which tells us that Brooklyn could be a place that one should be careful while visiting NYC.

Overall, we were able to understand better about how all the departments in NYC work in handling their complaints and it can be said that most departments are very efficient in handling cases especially NYPD. The NYPD handles both emergent and non-emergent cases in an efficient manner and tries to resolve them as quickly as possible.

APPENDICES

DATA DICTIONARY - 311 DATASET

There are 13 columns we find relevant. Here the data is described.

Relevant columns

AgencyName

The full names of the agencies and also the department which handles the service request in that agency is provided. Example : New York Police Department, Department of Transportation.

Agency

The acronyms of the different agencies which handles the service request are listed.

Complaint Type

This column identifies the topic of the incident or service request. Complaint Type could also have Descriptor which provides further details. Example : Hot Water problem, Street Condition, Noise complaint etc.

Descriptor

This column is dependent on Complaint Type. It provides further details about the complaint. Example : For a noise complaint, it says the cause of the nice like music, traffic etc.

Incident Zip

The column consists of the incident location's zip code. Example: 11205, 11378 etc.

Street Name

Consists of the street name of the given incident address. Example : University Avenue, 70th Street etc.

AddressType

Provides detail on the type of the Incident location Example : Address, Intersection etc.

Status

Provides the status of the service request. Example: Assigned, Closed etc.

Borough

Specifies the borough of the incident location. This also consists of many unspecified values. Example Bronx, Brooklyn, Queens etc

X Coordinate

Gives the X coordinate of the incident location. Example : 1033758

Y Coordinate

Gives the Y coordinate of the incident location. Example : 2401

Park Borough

Specifies the borough of the incident if it is in a Parks Department facility.

School Zip

Specifies the zip code of the incident if it is in a School zone.

Latitude

Specifies the latitude of the incident location. Example : 40.8273

Longitude

Specifies the longitude of the incident location. Example : -73.82111

CreatedDate

The Created Date specifies the date and time the incident was reported.

Example: 12/24/2015 2:55:20 AM

ClosedDate

The Closed Date specifies the date and time the incident report was closed.

Example: 12/24/2015 5:10:10 AM

NEW COLUMNS

Here, the Created date and Closed date is split into Month, day and it's hour.

New columns -> Created_month, Created_day, Created_hour, Closed_month, Closed_day and Closed_hours are added to the dataset

DATA DICTIONARY - NYPD CRIMES DATASET

DATE_TIME(Type: Date)

-Exact date time of Complaint

ADDR_PCT_CD

-The precinct in which the incident occurred

OFNS_DESC(Type: CHAR)

-Description of offense corresponding with key code

LAW_CAT_CD(TYPE: CHAR)

- Level of offense: felony, misdemeanor, violation

BORO_NM(TYPE:CHAR)

- The name of the borough in which the incident occurred

JURIS_DESC(TYPE:CHAR)

- Description of the jurisdiction code

Latitude(TYPE: FLOAT)

- Midblock Latitude coordinate for Global Coordinate System, WGS 1984, decimal degrees

Longitude(TYPE: FLOAT)

- Midblock Longitude coordinate for Global Coordinate System, WGS 1984, decimal degrees

SUSP_AGE_GROUP(TYPE: CHAR)

- Suspect's Age Group

VIC_AGE_GROUP(TYPE: CHAR)

- Victim's Age Group

Custom Columns

Day(Type: Number)

-Integer representing the Day, derived from Date_Time

Month(Type: Number)

-Integer representing the month, derived from Date_Time

Year(Type: Number)

-Integer representing the starting Year, derived from Date_Time