New York City Crime Analysis

Author : Mo. Helmi Boudhraa

SUMMARY

Title 1
Summary 2
Figures 3
Introduction 4
Background 4
Problem 4
Interest 4
Data acquisition and cleaning 5
Data sources 5
Data cleaning 6
Features selection 7
Exploratory Data Analysis 8
Crime Distribution across the Boroughs 8
Crime Distribution across the Boroughs (Incidents / Million) 9
Offense Levels Distribution across the Boroughs 10
Offenses by Suspect Race 11
Offenses by Age Group 11
Offenses by Sex 12
10 most common crimes in NYC 13
Most common crimes by Borough 14
Offense / Borough Standard Residual Table 14
Correlation Study Demographic factors / Incidents count 15
Visualization of Incidents locations 16
Visualization of Incidents locations Clusters 17
Visualization of safety degree of venues 17
Incidents By Venue Category 18
Most common Offenses / Venue Category 19
Offenses (By Venue Category) 20
Offenses (By Venue) 20
Results and Discussions 22
Conclusion 24

FIGURES

Dropped Features (Table) ……. 7
Kept Features (Table) …… 7
Crime Distribution across the Boroughs (Bar Chart) …… 8
Crime Distribution across the Boroughs (Incidents / Million) (Bar Chart) .. 9
Offense Levels Distribution across the Boroughs (Bar Chart) ……. 10
Offenses by Suspect Race (Bar Chart) …… 11
Offenses by Age Group (Bar Chart) ……. 12
Offenses by Sex (Bar Chart) …….. 12
10 most common crimes in NYC (Bar Chart) …….. 13
Most common crimes in by Borough (Bar Chart) …… 14
Offense / Borough Standard Residual Table (Residual Table) …….. 14
Demographic factors/Incidents count (Correlation Matrix) ……. 15
Visualization of Incidents locations (Map) ……. 16
Visualization of Incidents locations Clusters (Map) ……. 17
Visualization of safety degree of venues (Map) …… 18
Incidents By Venue Category (Bar Chart) ……. 18
Most common Offenses / Venue Category (Stacked Bar Chart) …… 19
Offenses (By Venue Category) (Correlation Matrix) ……... 20
Offenses (By Venue) (Correlation Matrix) ……..... 21

1. Introduction

1.1 Background

New York is one of the biggest cities in the world having a large population density and relatively high crime rate making the job of law enforcement agents extremely difficult.
One way to help analyze and predict crimes in New York is by exploring various data about the city, a process made easier by the NYC Open Data Project that makes the wealth of public data generated by various New York City agencies and other City organizations available for public use.

1.2 Problem

Crime analysis based on traditional socio-demographic data is of limited value because it fails to capture the complexity and dynamicity of human activity in cities. With the rise of ubiquitous computing, there is the opportunity to improve crime analysis (and prediction) with crowdsourced data (such as Foursquare data) that make for better proxies of human activity.

1.3 Interest

- Law Enforcement agencies are really interested / encourage any crime analyzing / prediction made by the public which will be helpful in their war against crime in the city.
- We will be doing some Venues - Crimes analysis so people who want to visit “safe” venues (we will discuss the criteria of safety later on) will be interested.

2. Data acquisition and cleaning

2.1 Data sources

- **New York City official Incidents Data : NYPD Complaint Data Current (Year To Date)**
  This Data contains a lot of interesting information including the coordinates (longitude, latitude) of each incident which we can exploit using Foursquare API.

- **New York City Population Data : NYC Population by Borough**
  This Data will be helpful in calculating the Incidents / Population / Borough

- **New York City Census Data : NYC Census Data**
  This Data will allow us to explore The (Demographics - Crime) relation.
  We will explore this data and extend the results with data acquired from Foursquare API.

- **Foursquare API Data : Foursquare API**
  We will be using The **Explore** feature of Foursquare API to fetch Nearby Venues for each Incidents and analyze the resulting data.

2.2 Data Cleaning :

NYC official Incidents data had a lot of missing values, and some random incidents that didn't happen in 2020.

The missing values were in both Suspect_Race and Borough features, all the Nans were replaced with 'Unknown'.

All the Incidents that happened before 2020 were dropped (nearly 5% of the dataset).

Census and Population data were cleaned and merged in the same table, we also added a Population feature to the incidents dataset to be able to calculate the incidents / million / borough rate.

Two Dataframes were extracted from incidents coordinates 'crimes_by_venue' and 'crimes_by_venue_category', I queried Foursquare for the top 20 venues within 100 meters radius from each incidents and constructed two dataframes one for the crimes / venue and one for the crimes / venue category.

Overall, the data wasn't very messy and did not require any special Missing Values Engineering.

2.2 Features selection :

NYC official Incidents data contained a lot of columns (35) most of the columns weren't necessary for our project so we dropped all the unneeded features, the Table Below summarize the features selection process:

Features	Reason for Dropping
ADDR_PCT_CD, CMPLNT_NUM, CMPLNT_FR_DT, CMPLNT_FR_TM, CRM_ATPT_CPTD_CD, JURIS_DESC , KY_CD, LOC_OF_OCCUR_DESC, PATROL_BORO, PD_CD, SUSP_SEX, SUSP_AGE_GROUP, PREM_TYP_DESC, RPT_DT, X_COORD_CD, Y_COORD_CD.	Irrelevant to our Project.
HADEVELOPT, HOUSING_PSA, JURISDICTION_CODE, PARKS_NM, STATION_NAME, TRANSIT_DISTRICT.	High number of missing values (> 50%)

Table1. Dropped Features.

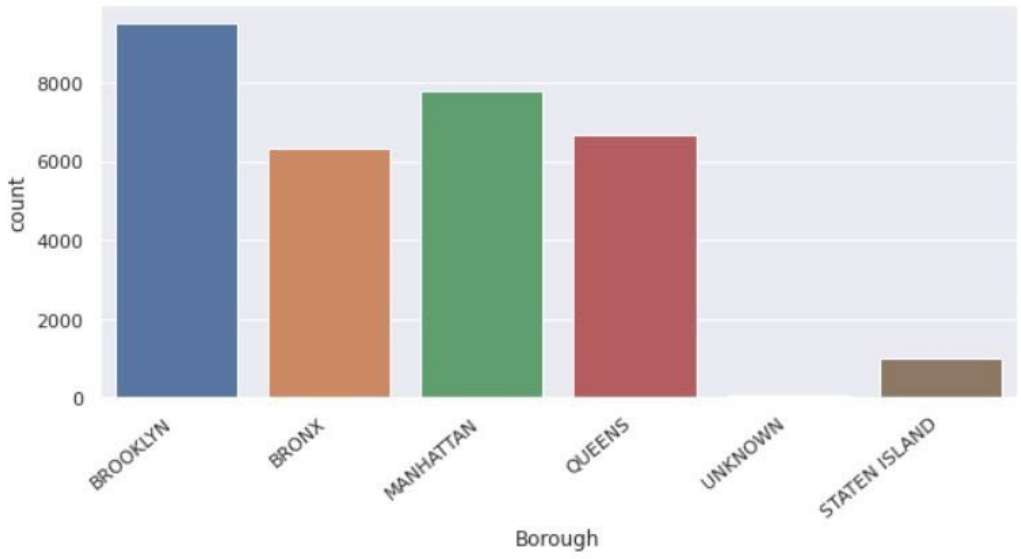
Kept Feature	Description
BORO_NM	The name of the borough in which the incident occurred
LAW_CAT_CD	Level of offense: felony, misdemeanor, violation
OFNS_DESC	Description of offense
VIC_AGE_GROUP	Victim's age group
SUSP_RACE	Suspect's race
VIC_SEX	Victim's Sex
Latitude, Longitude	Latitude / Longitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)

Table2. Kept Features.

3. Exploratory Data Analysis

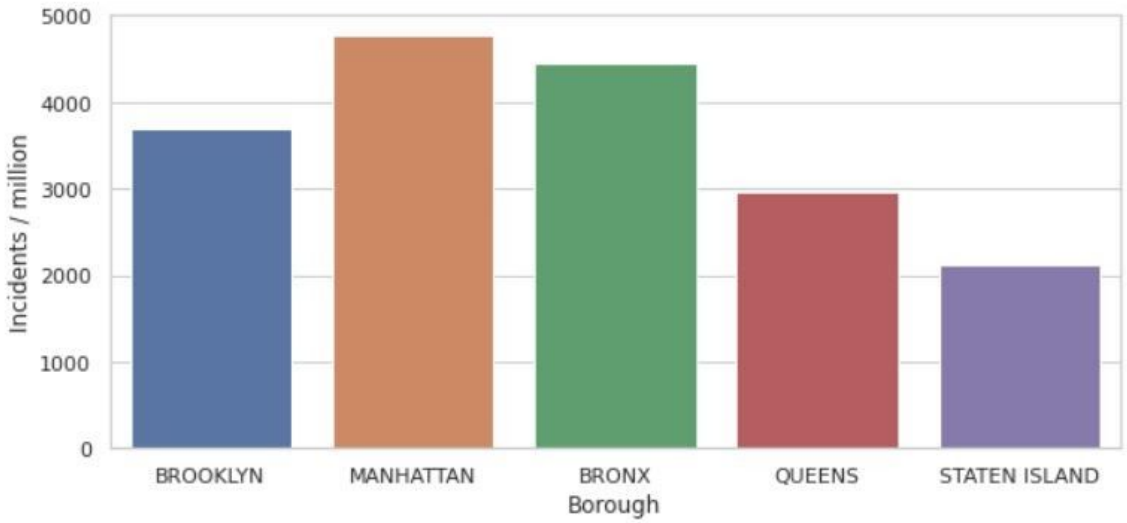
3.1 Crime Distribution across the Boroughs:

We will plot the crime count / borough to get an understanding of the crime distribution across NYC Boroughs.



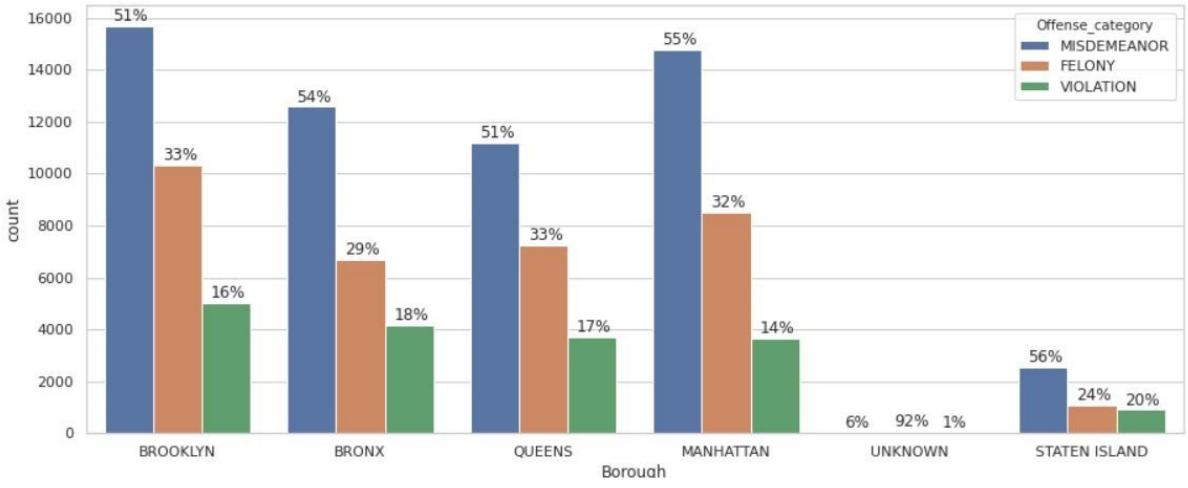
3.2 Crime Distribution across the Boroughs (Incidents / Million) :

Taking into consideration the difference in the population count, it will be more significant if we plot the Incidents / Million.



3.3 Offense Levels Distribution across the Boroughs:

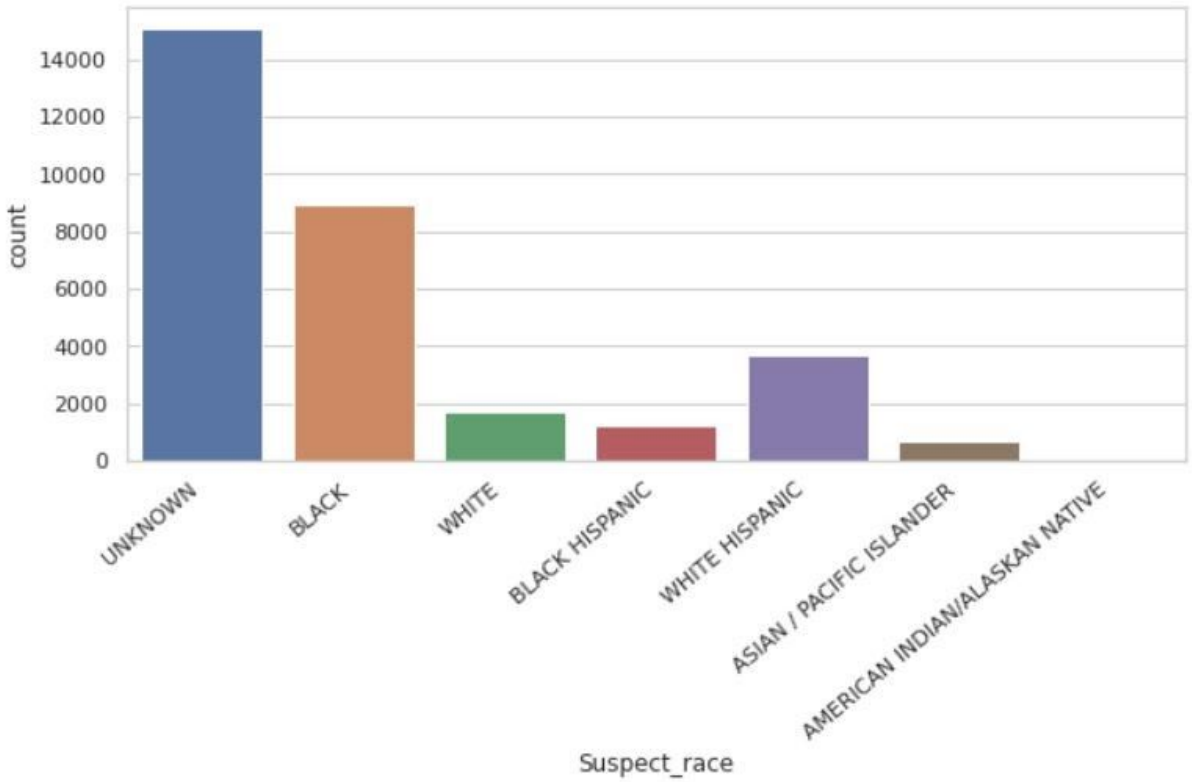
Exploring the distribution of Offense Categories across the borough is a good way to see if some offenses are more common in some boroughs.



- Staten Island has a substantially lower percentage of Felony than the other 4 boroughs (which have about 31%). This could mean that Staten Island is a much more peaceful area, with not only a lower total crime number but also less serious ones. (This also means that the Violation level of crime in Staten Island has a much higher percentage than all the other areas)

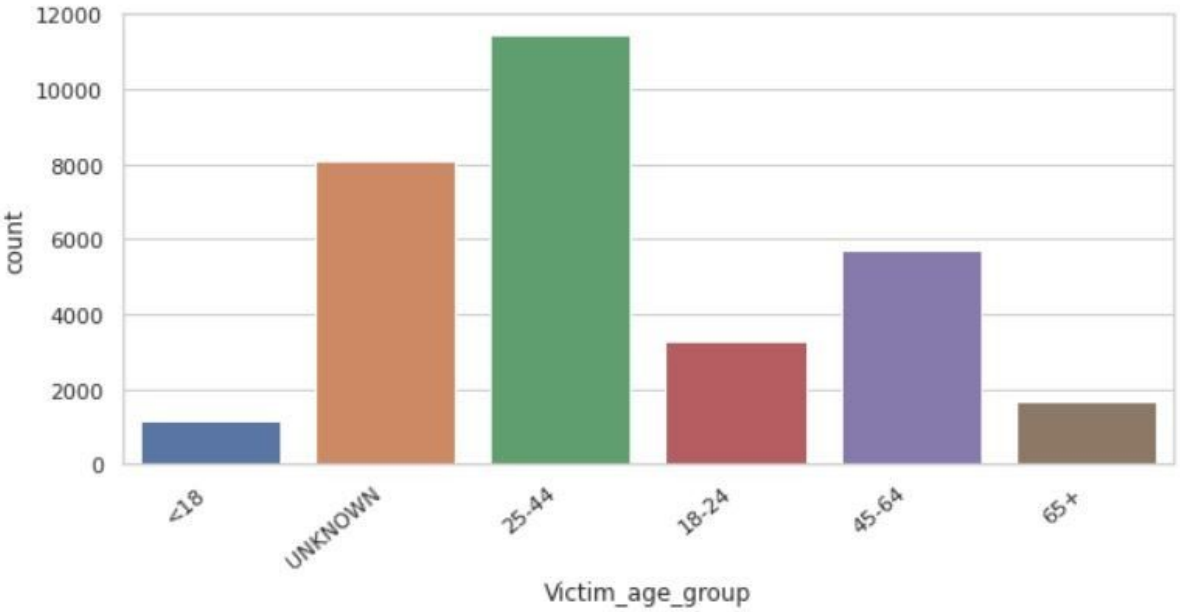
- Bronx also has a smaller percentage of Felony types of crime than the other 3 popular boroughs. This could mean that most of the crime that happens in Bronx is not pressing. This insight can affect the common belief of neighborhood security of Bronx.

3.4 Offenses by Suspect Race :



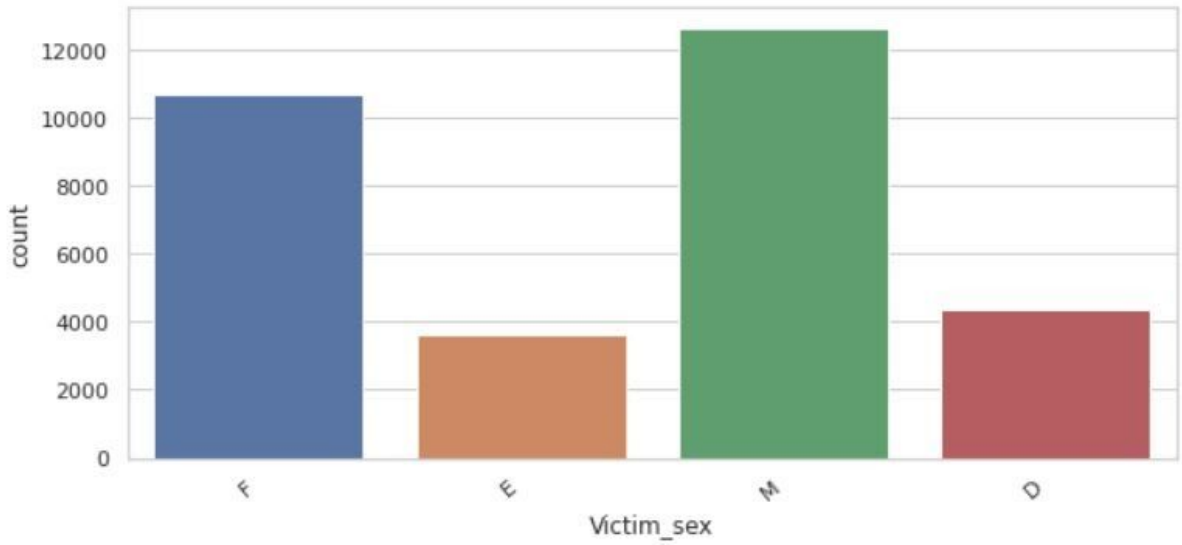
3.5 Offenses by Age Group :

One way to understand the crime dynamics in NYC is to inspect the age groups of the victims



- Nearly a Third of victims (~11000 / ~31000) are between 25 and 44 years old.

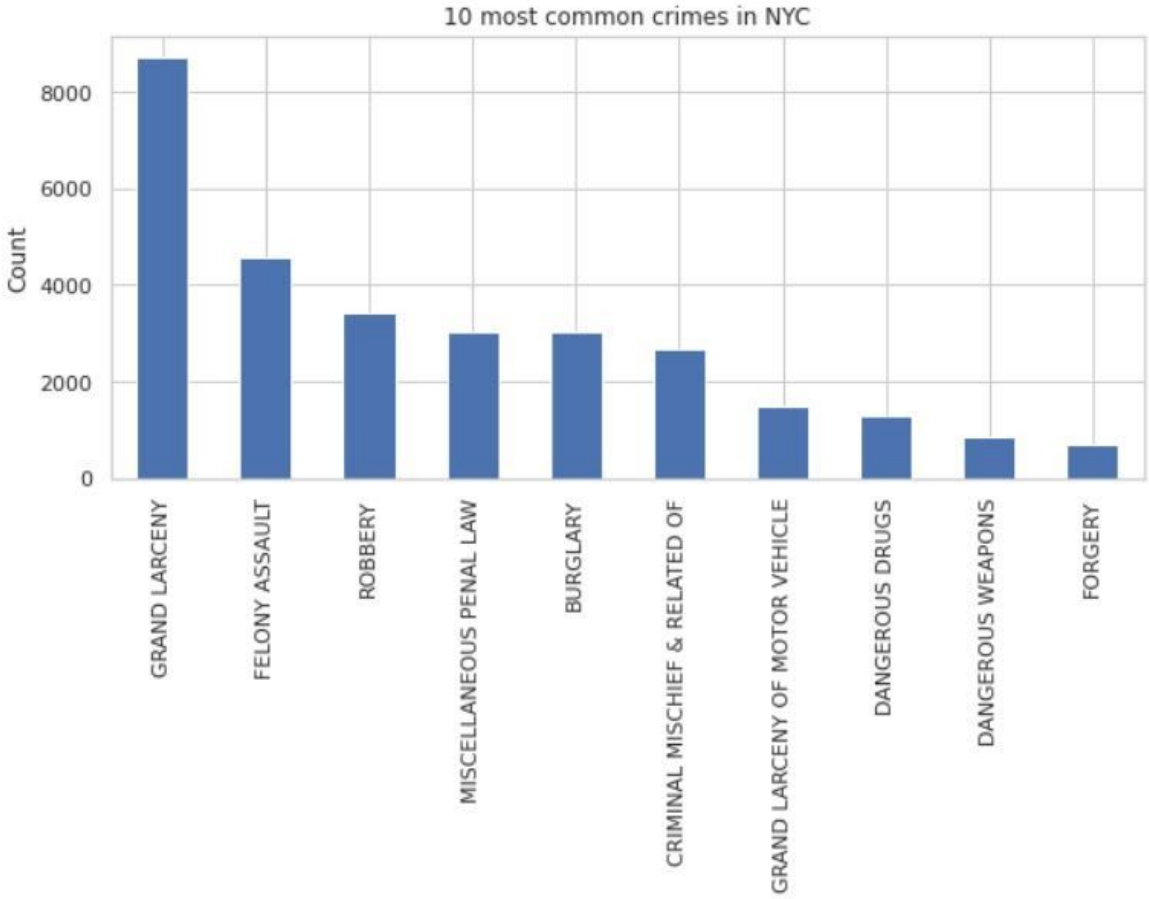
3.6 Offenses by Sex :



(In the Victim_sex feature 'E' and 'D' refers to material objects)

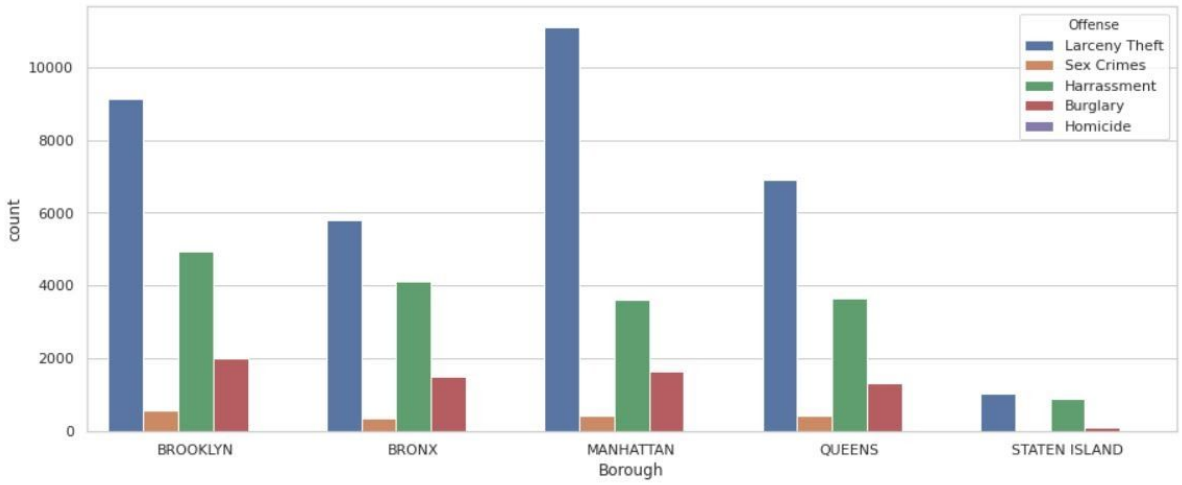
- Victim's sex count doesn't imply much, Male Victims are slightly higher than Females.

3.7 10 most common crimes in NYC :

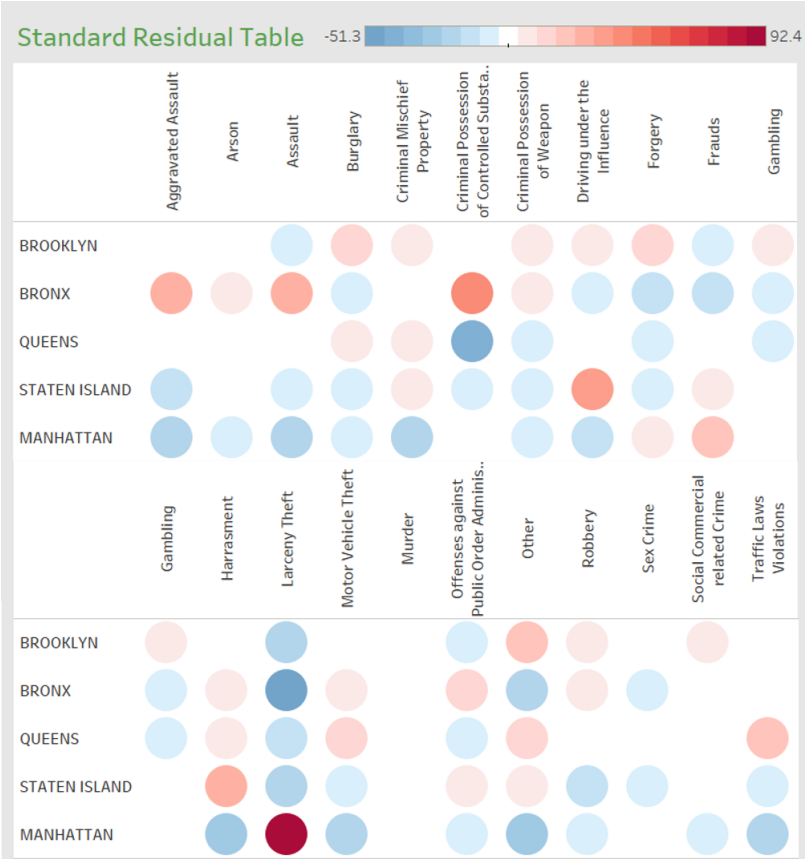


- Grand larceny (which is the theft of another's property – including money – over a certain value) is the by far the most frequent incident.

3.8 Most common crimes by Borough :



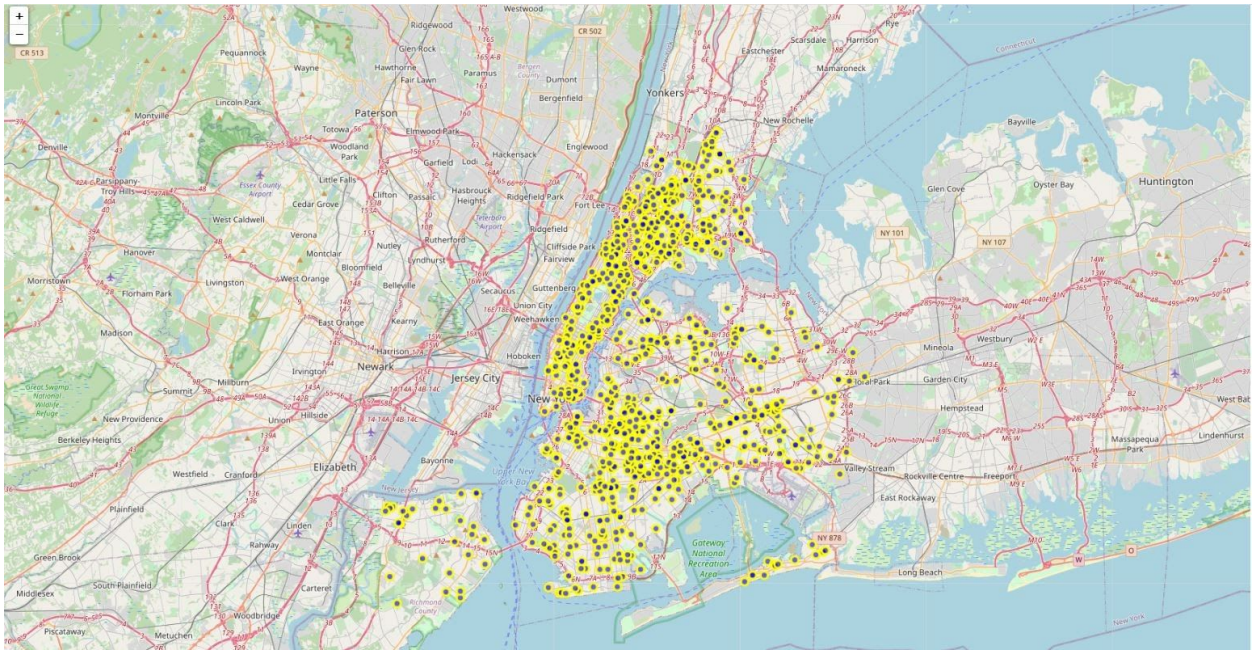
3.9 Offense / Borough Standard Residual Table (Plot made in R and Took from Kaggle) :



3.10 Correlation Study between Demographic factors and Incidents count :

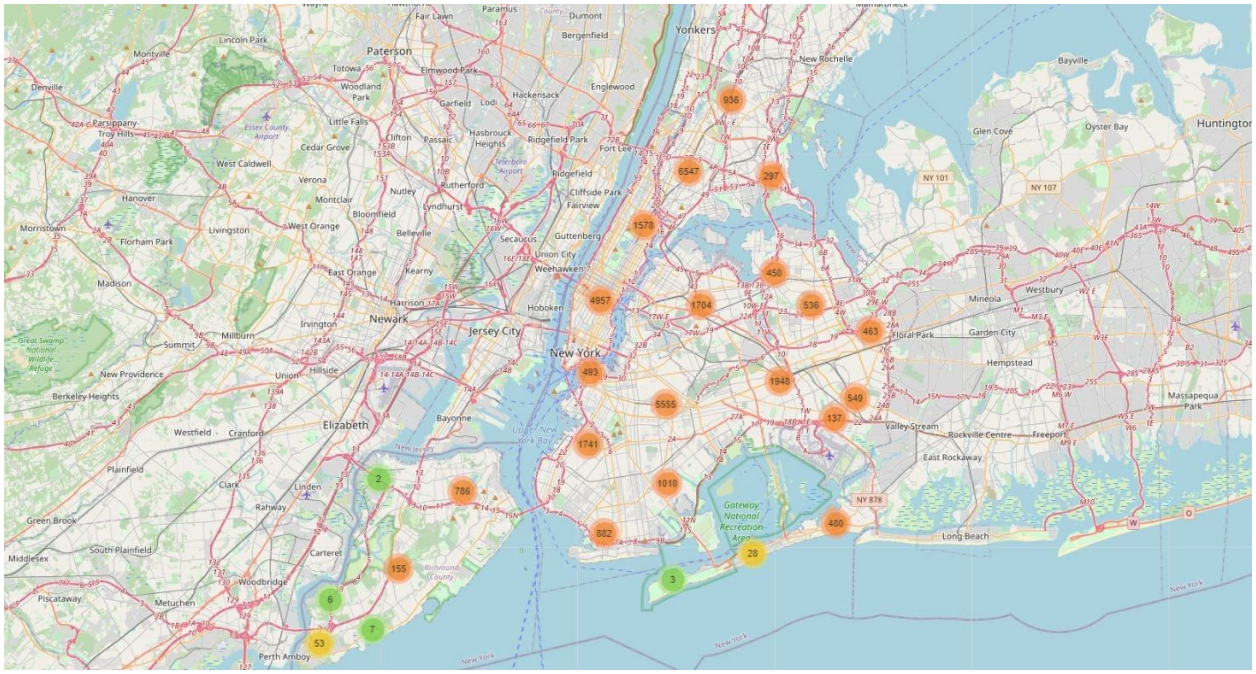
- Positive Correlation between Incidents_count and the Percentage of Poverty, the poorer the borough the higher the crime
- Negative Correlation between Incidents_count and Median Household Income.

3.11 Visualization of Incidents locations:



- **Confirming the previous plots, we can observe incidents clusters in both Manhattan and Brooklyn**

3.12 Visualization of Incidents locations Clusters :

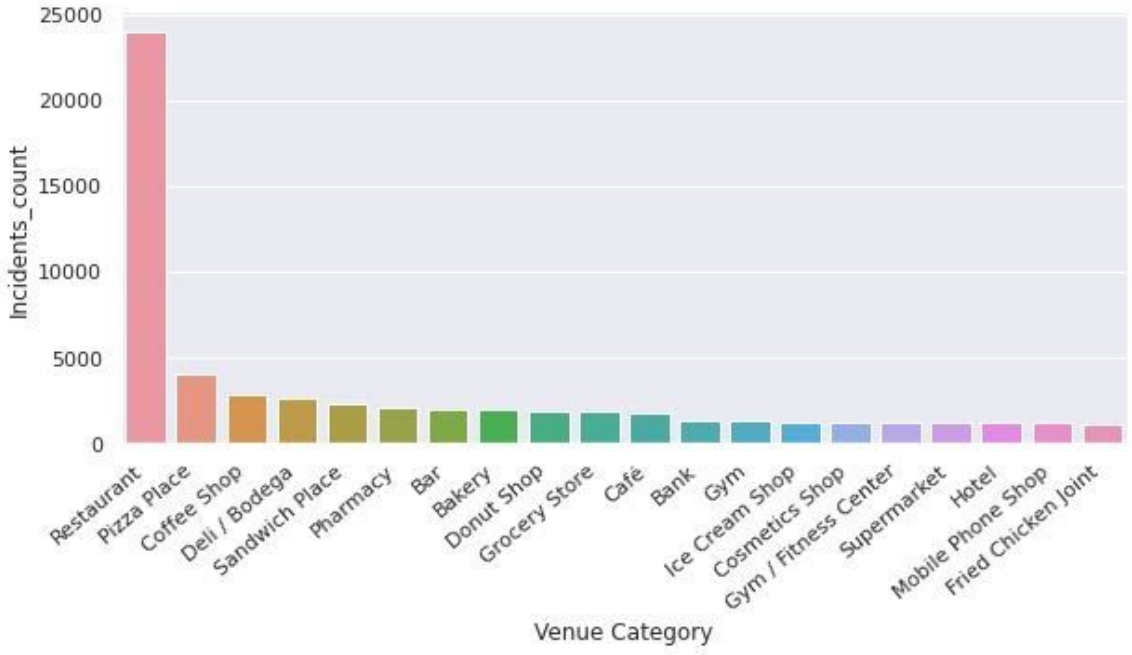


3.13 Visualization of safety degree of venues :

The safety degree is based on the number of incidents that happened in 100 meters radius from the venue:

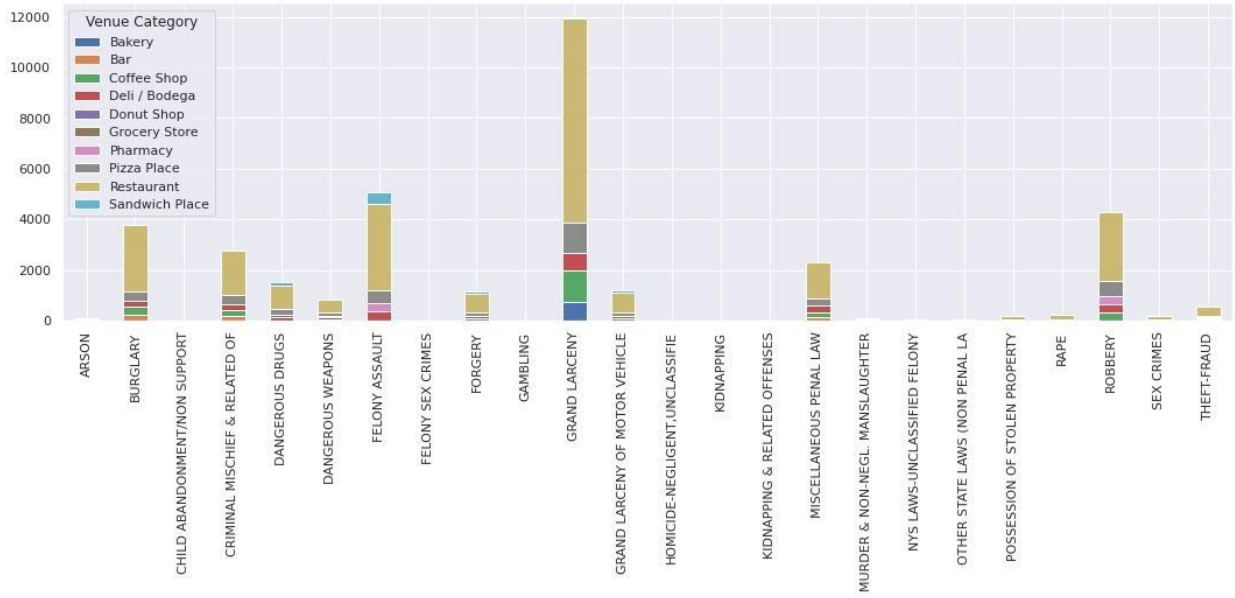
- Venues number of nearby incidents < 80 are considered safe. (Green Pins)
- Venues number of nearby incidents >= 80 and < 150 are considered partially safe. (Orange Pins)
- Venues number of nearby incidents > 150 are considered dangerous. (Red Pins)

3.14 Incidents By Venue Category:



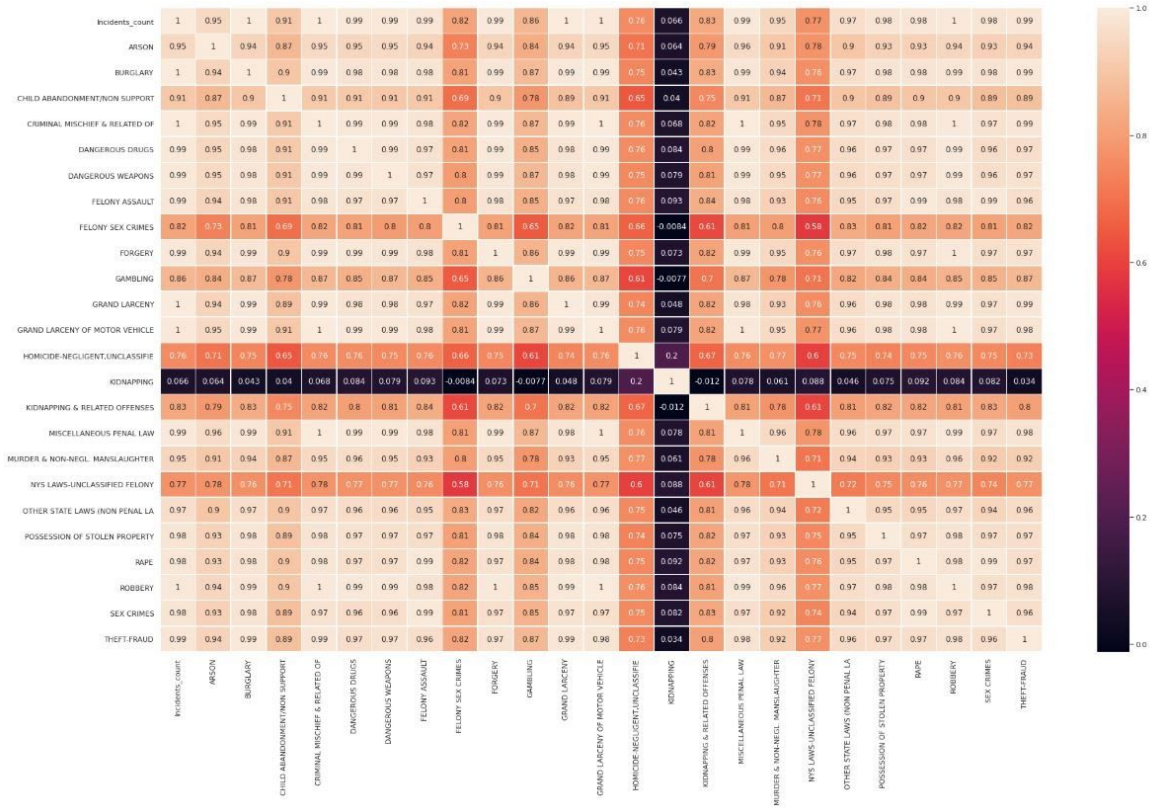
- Restaurants have the highest Incidents count (within 100 meters) because, probably, because they are everywhere across the city !

3.15 Stacked Bar Chart of the Most common Offenses / Venue Category:

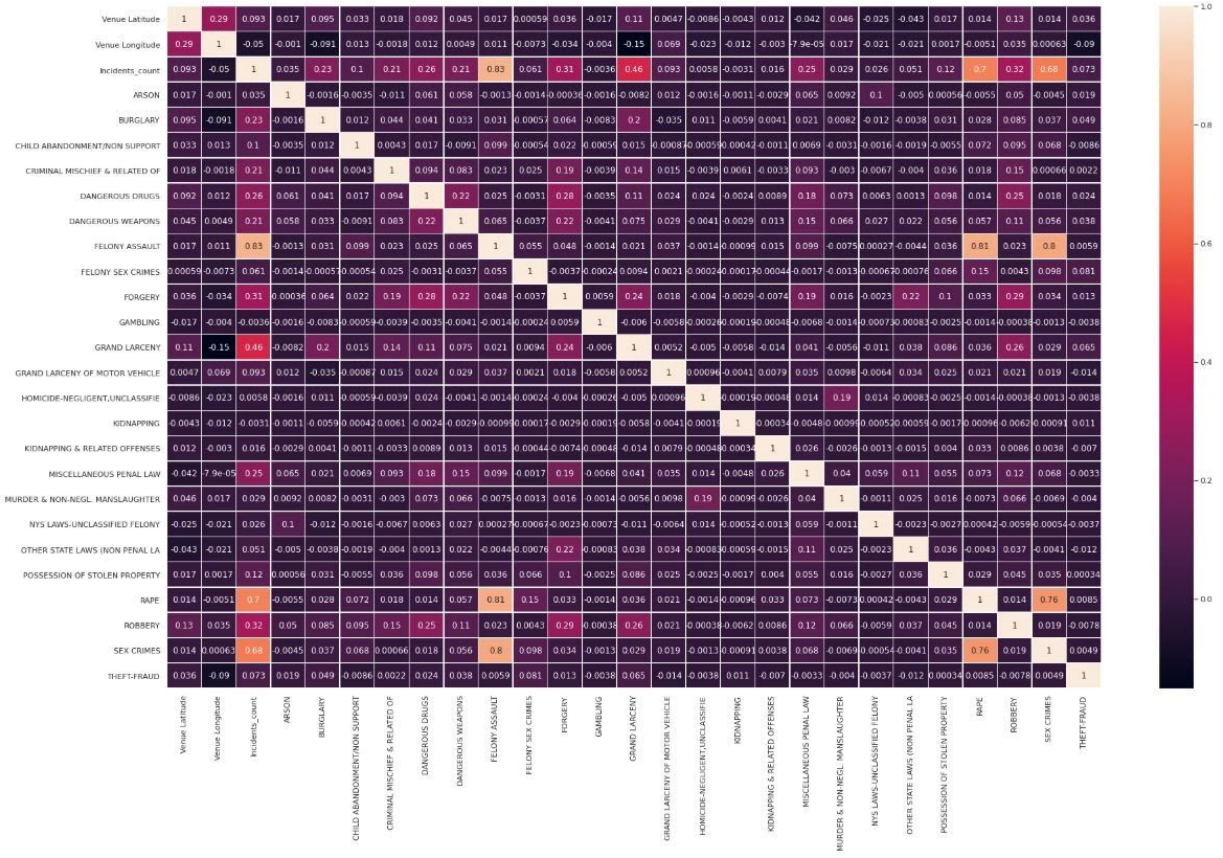


- We can observe that grave incidents such as (Sex crimes, rape, homicides, kidnapping) often happen in isolated places with no venues around.

3.16 Offenses Correlation Matrix (By Venue Category):



3.17 Offenses Correlation Matrix (By Venue):



- We notice a correlation between Felony Assaults and Sex Crimes / Rape.
- We also notice a correlation between Sex Crimes and Rape, that maybe explained that the venues that are more isolated have more sex crimes / rape incidents in their surroindings than other venues
- There are Several other Correlations that are not very strong between multiple crimes.

3. Results and Discussions

- **Misdemeanor** is the most popular level of crime in each borough, and it consists of a similar percentage in each group (about 51% to 56%; Manhattan has the highest percentage of 55%).

- **Felony** The second popular.

- **Violation** The third popular.

- **Staten Island** has a substantially lower percentage of Felony than the other 4 boroughs (which have about 31%).

- This could mean that **Staten Island** is a much more peaceful area, with not only a lower total crime number but also less serious ones.

- Based on the plotted maps above, most of New York City Venues are **safe** (in term of number of Incidents that happened within a 100 meters radius) except a few Venues that were near a big number of incidents, this doesn't imply that the problem is with the venue itself but it maybe be due to its bad location.

- We observed a strong correlation between demographic data and crime rates (The higher the income the lower the incidents rate)

- Being the most popular place in terms of incidents, Brooklyn has a lower frequency of Fraud and High-Value theft. This is also true in Bronx, Queens and Staten Island.

- The data from Bronx and Brooklyn shows high frequencies of a few public-concerning crime categories such as weapons, drugs, rubbery and serious assault. This could be one of the factors that some people have conflicting opinions about the neighborhood safety of these places.

- Manhattan's popular crime types, except sex crime, are mostly not physical crimes (violent crime) but the crimes involve personal or business property. This could be because of the high proportion of business or commercial districts in this borough.

- Queens has a distinctly high frequency of crime in regards to vehicle and traffic law. This is a little different from close boroughs such as Manhattan and Brooklyn. It could be that Queens' residents own more vehicles to commute, or Queens has more vehicle-related industrial areas than other districts.

4. Conclusion

- From the exploratory analysis and statistical approaches, I can say that different boroughs do have significant differences in terms of crime level and crime type with each other.

- Most of the venues in NYC are safe, except some that have a very high of incidents happening in their surroundings.

- Confirming the obvious, poverty is the **essence** of crime !

- Common assumptions on the safety of some Boroughs (Bronx) are, surprisingly, not validated by the data.

- Due to the extreme diversity (entropy) of NYC and the big spread of venues all over the city, we can't strongly relate some types of crimes to some specific venues.