Group 5 Project Report

Part 1: Introduction

Law enforcement and policing are vital components of any society, working to maintain order, uphold the law, and ensure public safety. The primary objective of law enforcement and policing is to prevent crime and protect citizens from harm. Data analytics has become an increasingly important tool for law enforcement and policing, allowing these organizations to better understand crime patterns and trends, allocate resources more efficiently, and prevent and solve crimes. The use of data analytics has revolutionized the way law enforcement operates, enabling police officers and law enforcement officials to target criminal activity and identify potential crime hotspots more effectively. This project we aim to demonstrate how data analytics can help answer various questions which will benefit law enforcement agency, policy makers and general public to better strategies their decisions.

Part 2: Data Source Description

Source of Dataset:

We have used multiple datasets in our analysis with our primary data source as NYC Open Data. Below is the list of datasets used in our analysis.
NYPD Arrest Data -> Each record represents an arrest effected in NYC by the NYPD and includes information about the type of crime, the location and time of enforcement. In addition, information related to suspect demographics is also included.
NYPD Complaint Data -> This dataset includes all valid felony, misdemeanor, and violation crimes reported to the New York City Police Department (NYPD).
NYPD Shooting Incident Data -> Each record represents a shooting incident in NYC and includes information about the event, the location and time of occurrence. In addition, information related to suspect and victim demographics is also included.

Characteristics of Dataset:

NYPD Arrest Data - the data contains 5.5 million observations and 19 features.
Strengths: a) Location information for arrest. b) Precinct information. c) Demographic information of perpetrator.

Weakness: a) No victim information. b) 0.0478 % of missing data

NYPD Complaint Data -> The data contains 8.36 million observations and 35 features.
Strengths: a)Time of occurrence of reported event. b)Precinct and location where the event occurred. c)Description of offence. d) Indicator if the crime was successful. e)Suspect and Perpetrator demographic information.

Weakness: 1. 20.73% of missing data

NYPD Shooting Incident Data -> The data contains 27.3 K observations and 21 features.

Strengths: a)Location of shooting. b)Occurrence date and time. c)Perpetrator and Victim demographic information.
Weakness: 1. 8.83 % of missing data.

Data Cleaning Steps:

1. Omitting columns with more than 70-75% of NA values.
2. Imputing columns with less than 5% of NA values with approximate values.
3. Removing the observations where year value is out of year range specified at source.
4. Removing observation where age value is negative or more than 100 and is not falling in the specified age group.

Part 3: Data Visualization and Storytelling

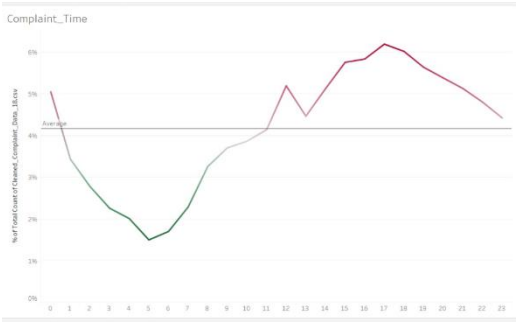
The Crimes Dominant in an Area and How They Span Across Different Time categories:

Area information and changes in crime rates can provide valuable insights into crime trends and patterns. Analyzing weekly, monthly, and yearly changes in crime rates can help to identify trends over time and highlight areas of concern that may require targeted interventions or resource allocation.

To start the study we look at the types of crime being conducted at different locations in our data, using bar plot by plotting having area as our main category, crime category as our sub category and count of crimes conducted. We observe that Queens and Brooklyn has the majority of crimes being committed with the two top categories as public order offence and Violent crimes.

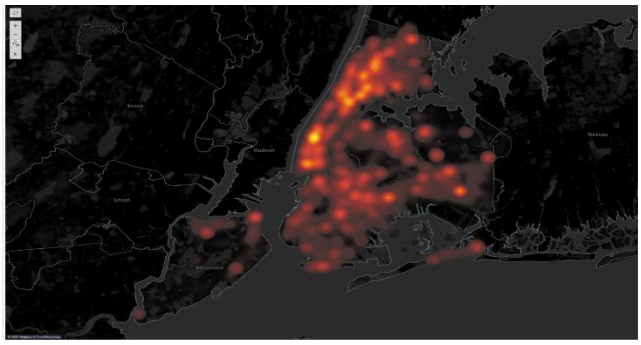
We then investigate how the crime rate has changed over years, to do so we plot years on x axis and count on y axis. We observe that there has been a declining trend in our arrest data but we cannot conclusively say the same for our complaint data. According to our analysis of complaint data, between 2018 and 2021, the year 2020 saw the most crimes committed, and complaints filed.

We further analyze the complaint data as it is a better indicator of crime being committed, and observe how the crime changes across various months. Monthly changes in crime rates can help to identify longer-term trends in crime activity. For example, if crime rates increase consistently over several months, it may suggest a systemic issue that requires more extensive intervention or policy changes.

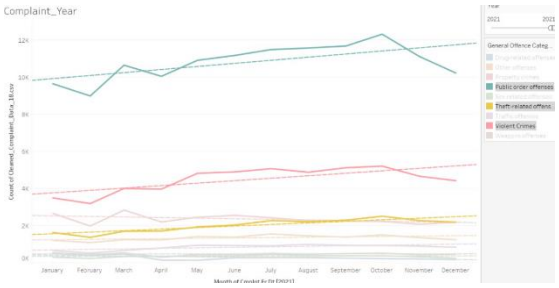


we achieve this by plotting months and count of complaints in which the crime was completed and observe that the highest number of crimes are perpetrated from May to November in terms of months.

We also analyze how the crime changes over the week and hour of days and observe that Wednesday through Saturday during the workweek have seen the most crime activity and The highest number of crimes was reported between 12 p.m. and 11 a.m.



To analyze the geographical crime hotspots, we plot a geo-plot and construct a density map. By analyzing crime data on a map, law enforcement agencies can identify areas with high concentrations of criminal activity, such as theft, assault, or drug-related crimes. This information can then be used to allocate resources to those areas, such as increased patrols or targeted investigations, in order to reduce crime and increase public safety. We observe that Manhattan and Bronx are the primary hotspots for committing crimes.



To predict the crime trends and tell what crimes will be increasing in the following year we plot trend lines for each crime category for the year of 2021 and observe that public order offences, theft related offences and violent crimes are more likely to increase.

For preventive measures to control crime in an area and how can law enforcement use the analysis to allocate resources more effectively we suggest the following:

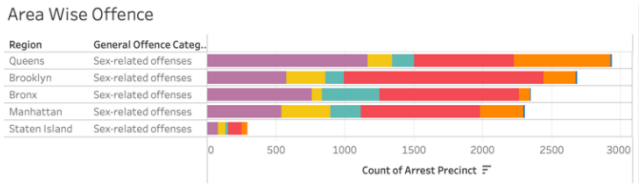
1. Police forces can use the above observations to pinpoint the crime hotspots using the geo-plot, allocate resources in different time of day and week according the above observations such as crime is most active in the evening and hence more resources can be allocated for that time.
2. Increase police presence: Law enforcement can increase patrols in high-crime areas to deter criminal activity and respond quickly to incidents.
3. Targeted enforcement: Law enforcement can focus on specific types of crime or areas where crime is more prevalent, such as drug trafficking or gang activity.
4. Crime prevention programs: These can include neighborhood watch programs, youth programs, and crime prevention education.

The Use of Demographic Data to Investigate Criminal Activity

Demographic data can be a valuable tool for police forces in a variety of ways, from identifying areas of high crime rates to developing more effective policing strategies that consider the unique needs and circumstances of different communities. Demographic data can be used to inform community policing efforts. By understanding the unique needs and circumstances of different communities, police forces can develop more effective strategies for building trust and working collaboratively with community members to prevent crime. This can include initiatives such as cultural sensitivity training for officers, outreach programs to engage with at-risk youth, and partnerships with community organizations to provide social services and support to those in need.

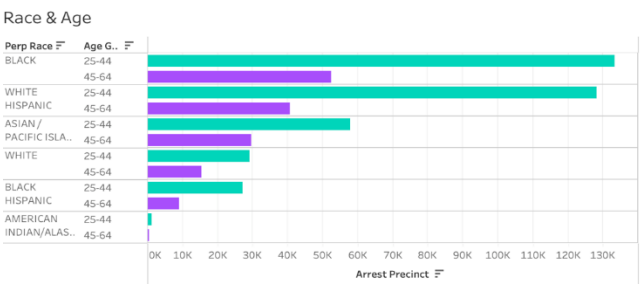
We investigate the demographics by providing information on both **perpetrators and victims,** and to make our analysis more detailed we select the sex-related offences as our focal point.

Perpetrators –>



We start by looking at the areas where majority of the sex related offences occurred and which race was responsible for majority of the occurrence of this offence. We plot a bar graph with region and counts of occurrence and color the bars by race. We can observe that most of the crime are being conducted in Queens followed by Brooklyn and the crime is being committed by Black's the most followed by White Hispanics.

We further look into the age group committing the most sex-related offences for the above two races, we plot the race and age group on our vertical axis and the count on our horizontal axis to obtain the graph. We observe that the age group of '25-44' commits the most crime followed by '45-64' in this crime category across all the races. We further look into the sex of the criminals by region to find if males are more involved in the crime category of females for a particular region, we accomplish this using a simple bar graph by region with bar color as sex and observe that males are the majority. We also construct a geo-plot to have a view of which age group, and race is responsible for the crime and if we can observe anything if we plot it on the map.



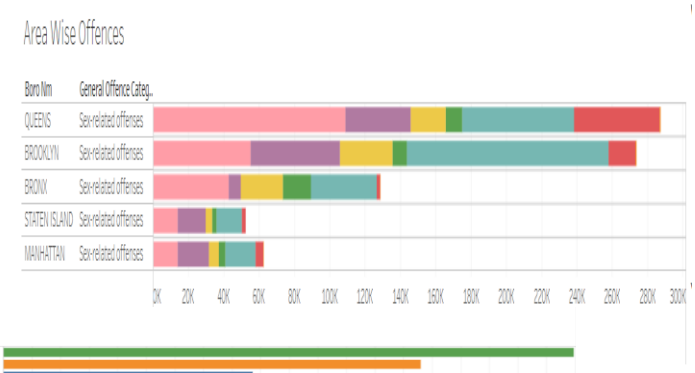
With the above observations we can construct a simple model to predict the crime category for an observation, such as if an observation is Black-male or White Hispanic-male in the age group of 25-44 years, and is operating in the Queens then we can conclude that it is very highly likely that the crime conducted by this observation is in the category of sex-related offences.

Measures which should be implemented for the prevention of such crimes for a demographic group could include:

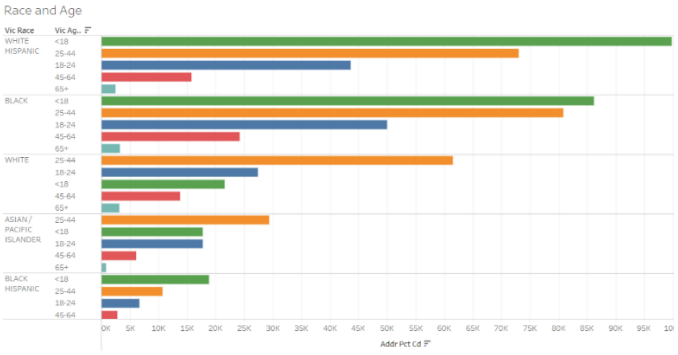
1. Targeted community policing, education and training programs focused on healthy relationships and consent, mental health support services, economic empowerment initiatives, and restorative justice programs could be implemented.
2. Improved information sharing by state so that general public is aware of the following observations and can take appropriate steps to educate their relevant communities.
3. Better monitoring of individuals which have the demographic characteristics which falls under a certain type of crime.

Victim –>

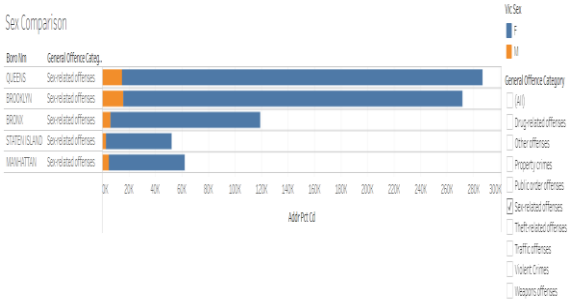
We start looking at the areas where majority of the sex related offences occurred and which race was the most targeted for this offence category. We plot a bar graph with region & counts of occurrence and color the bars by race. We can observe that most of the crime are being conducted in Queens followed by Brooklyn and the ethnic group that is most targeted is the White Hispanic ethnic group.



We further look into the age group being the most targeted in case of sex related offences, we plot the race and age group on our vertical axis and the count on our horizontal axis to obtain the graph. We observe that the age group of 'Less than 18' is the most targeted followed by age group '25-44' in the sex related crime category across all the races.



We further look into the sex of the victims by region to find females are more targeted in case of sex related offences or males for a particular region, we accomplish this by using a simple bar graph by region with bar color a sex and observe that females are the most targeted in case of sex related offences.



We also plot a geo plot to have a view of which age group, and race is being targeted for a crime and if we can observe anything if we plot it on the map.

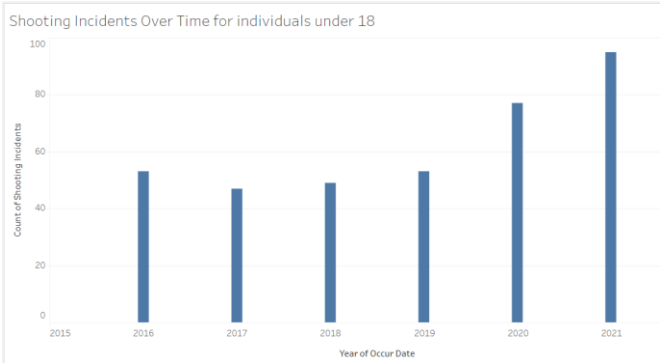
With the above observations we construct a simple model to predict the crime category for an observation, such as if an observation is White Hispanic race, queens region, and under 18 age group, public order offense is the most likely criminal activity that the females of this demographic group would be affected by.

Measures:

1. To reduce the prevalence of sex-related crimes among White Hispanic individuals under 18 in Queens, targeted interventions such as increased law enforcement presence, community outreach programs, and education campaigns on consent and healthy relationships could be implemented.
2. Additionally, interventions specifically tailored to address the needs and experiences of female victims, such as providing access to support services and resources, could also be effective.

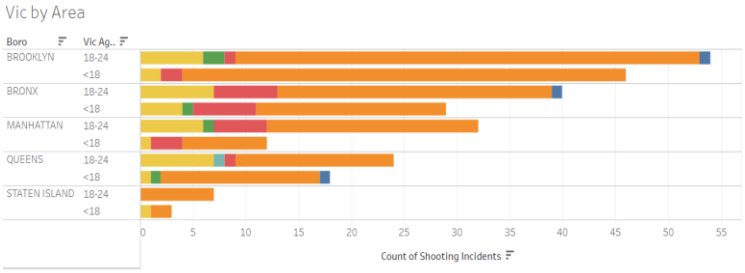
The Impact of Criminal Justice Policies

We consider the “Raise the Age” (RTA) policy implemented in New York. The "Raise the Age" legislation increased the minimum age of criminal responsibility from 16 to 18 years old. As a result, individuals who are 16 and 17 years old and accused of misdemeanors are now directed to family court, where they may be adjudicated as juvenile delinquents. RTA allows for an AO’s case to continue in adult court under “extraordinary circumstances exist that should prevent the transfer of the action to family court”. But these " extraordinary circumstances” are not clearly defined allowing individuals who are 16 and 17 to be pardoned for crimes which are considerably violent. RTA changed the age of criminal responsibility to 17 as of October 1, 2018 and 18, as of October 1, 2019.



We analyzed various types of crime activity for the age group of “Less than 18” using relevant filters in tableau, with the year range of 2016-2021 and found that the only area where there was a significant increase in the criminal activity was the shooting incidents.

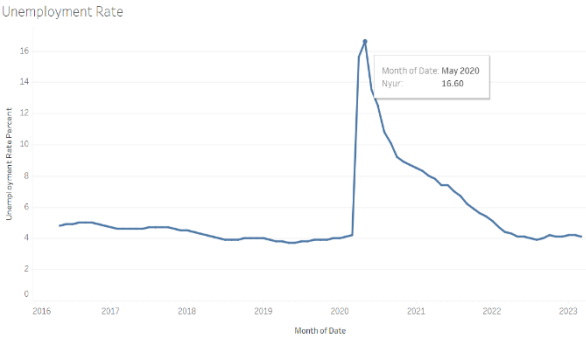
We observed an increase of 102% from 2017 to 2021 in the shootings performed by individuals in the above-mentioned age category. To further analyze the nature of our finding, we start by investigating what our perpetrators characteristics are and how shootings are distributed by area and what Race groups are the most active in conducting the shooting incidents. To understand this, we construct a bar graph plotting the count of shooting incidents by region conducted by the above age group, we use race as color in the graph to see which race is responsible for most of the shootings. We observe that Blacks followed by White Hispanics are the most active group in each region and Brooklyn followed by Bronx is the most effected regions where shootings are conducted.



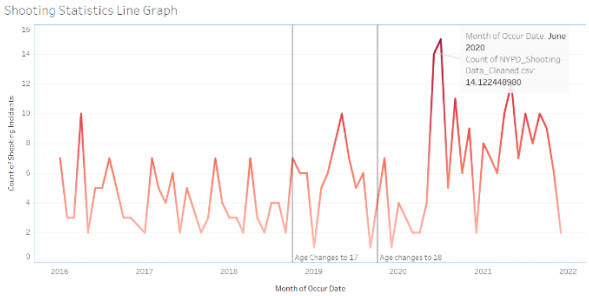
To further investigate the victims by plotting the victims of our shooting incidents across regions and age groups and color the bars by victim race. We observe that the age group of “18-24” is the most effected group followed by “Less than 18”. This is a rather interesting observation as the shootings are conducted by the age group of “Less than 18”. We can conclude that the victims of the targets of these shootings were usually individuals older than the shooter.

As our observations also include the COVID19 period we cannot conclusively say that the increase in the above crime is only due to RTA policy. To further investigate this matter we consider other factor such as unemployment rate which is directly related to the COVID19 affect, with the assumption that higher unemployment rates may lead to a more stressful environment for the adolescents in their home, unemployment can also have negative impacts on mental health and well-being, leading to increased, stress, depression, and other psychological issues. This could serve as a trigger for the criminal activity.

To investigate this we plot the unemployment rates observed form 2016 to 2023 using a line graph. We plot the Yearly data on our x-axis and unemployment rate in percent on our Y-axis. We observe that we see the peak of unemployment at May 2020 with a value of 16.60% at this point which is among the highest recorded for the state.



We plotted one more graph which Is a detailed version for shootings performed by the age group in question by year and months. We plot months and years on x-axis and the count of shooting incidents on y-axis and see if there is a spike around the unemployment spike observed on May 2020.



We can see the peak in this graph occurs at June 2020 and July 2020, which is a month after the unemployment peak observed. With this data we can conclude that the unemployment rate can be a factor used to explain the shooting incidents in our observation.

RTA contains loopholes which leads to such outcomes. We suggest the following to futureproof the RTA:

1. Youth-part judges must be allowed to access defendant’s full history
2. RTA must account for real world problems or scenarios which occur such as: gang assaults, display of what appears to be a firearm or deadly weapon

The study for this can serve as a reference point for other states implementing RTA. They can leverage the observations made in this study to further bulletproof the policy according to their state.

Part 4: Conclusion

Data visualization has become an increasingly important tool for criminal data analysis, allowing law enforcement agencies to gain deeper insights into patterns and trends in criminal activity. By presenting data in a visual format, such as graphs, charts, and maps, data visualization helps to highlight important relationships and trends that might otherwise be difficult to detect.

We were able to answer most of the questions in a very efficient manner by visualizing the data we were able to observe the trends over different periods of time using line plot, we were able to compare various factors such as region and race using a bar graph and we were able to see how the crime is distributed over the geographical map allowing us better insights into our data and easy location of criminal hot spots.

One of the key benefits of data visualization is that it allows for the quick and efficient analysis of large amounts of complex data. With traditional methods of data analysis, such as spreadsheets or text reports, it can be difficult to identify important patterns or trends in the data. However, data visualization tools can quickly and easily transform large amounts of raw data into meaningful visual representations, allowing analysts to identify trends and patterns with greater accuracy and efficiency.

Data-Source and Tableau-Workbook link