Analysis on Shooting Incidents, Arrests & Court Summons in New York City

1st Shashank Sanjay Tomar
MSc in Data Analytics
National College of Ireland
x19213280@student.ncirl.ie

2nd Sumit Singh
MSc in Data Analytics
National College of Ireland
x20135769@student.ncirl.ie

3rd Vani Mirg
MSc in Data Analytics
National College of Ireland
x19211538@student.ncirl.ie

***Abstract*—The paper aims to assess the Law-and-Order situation of one of the biggest Metropolitans in the world, New York. The NYPD police department maintains a database of all the crimes in the city based on several factors. We consider the data for the shootings that have happened in the city, arrests made for the crimes and summons issued by the courts to maintain order in the city. We use Python 3.0 as the base technology for analyzing data and use two different databases for storing that data. MongoDB is an unstructured database that will store the records fetched from the NYC open data database. From MongoDB, we fetch the data and analyze it to make various inferences about the shootings, arrests, and summons in the city. We ensure to make individual inferences from each dataset and present different use cases in the form of visualizations and exhibit the deductions using various graphs, plots, and animations. We have also created several maps to highlight the key areas where the crimes have taken place. We then use MySQL as a structured database to store the data from our individual inferences and then combine all the information to make inferences as a group about the city. We highlight key factors based on the boroughs with maximum crimes, along with the race and gender of the offenders.**

***Index Terms*—New York City (NYC), New York Police Department (NYPD), Shootings, Arrests, Summons, Python, MongoDB, MySQL**

I. INTRODUCTION

Study in the field of crime analysis has been used to reduce crime and improve public safety for many years. The New York Police Department is the country's largest municipal force policing an 8.5million-person city. The number of cases brought under Borough, Precinct Crime, Robbery, Assaults, Shooting Incidents, Arrests, Court Summons and other categories has skyrocketed in recent years. According to the crime reports collected by NYPD, the majority of the offences are hate crime offences and illegal activities. Researchers have also looked into the impact of factors like age, race, gender, alcohol and other factors on crime rates. With the emergence of Big Data, public data access and e-governance, the development of visualization for policing and managing public-safety measures has increased at an almost exponential rate.

New York is the most populated city in the United States of America and as per the urban scaling theory, it is no secret that population and crimes such as shootings and murders have got a super-linear association amongst themselves. The recent social movements revolving around topics like shootings, racial discrimination etc. across the United States of America, inspired analysis of the shootings that occurred in the city of New York in the last decade (2010-2019). The rise of such movements across the nation mainly focused on raising awareness amongst the citizens, governments, and administration of the states. The analysis revolved around affirming whether shooting incidents decreased overtime or not and identifying a pattern amongst the aspects of shooting like the borough (the place where the shooting occurred), race, gender of the victim and the perpetrator, timings of the shootings etc. The aforementioned inferences will eventually be amalgamated with the patterns and inferences made from the Court Summons and Arrests made in New York City to present a relatively colossal overview of the safety and felonies committed across the city.

The "Big Apple", formally known as New York City is one of the biggest metropolitans in the world. New York is known to have a large and diverse population. However, with a large population, the crime rate in the city is also quite high, In this paper, we will access the law-and-order situation of the city by analysing the number of arrests made by the New York Police Department (NYPD) over 10 years. We analyse the numbers and severity of the crimes committed over the period and assess the demographic impact of the population on the crimes committed. We make inferences based on different age group, gender, the race of the citizens, while also diving deep into the types of crime being committed in the city and investigate the safest areas in the city.

Between 2010 and 2015, the New York City Police Department (NYPD) issued 1,839,414 summonses offences such as public urination and possession of small amounts of marijuana. There are several legitimate reasons to issue such summonses, most notably to address community concerns and police offences in question. Further, maintaining order is a goal in itself. New York City is a safer city today than it was in years past. In the last decade, felony rates continued to decline and remain and historic lows. What factors contributed to this safer city is a worthy inquiry because identifying what works will help the Department become more strategic and more efficient. It is also necessary to determine which factors are responsible for the court summons that have occurred in the last decade to determine whether factors such as gender, age, race, and boroughs affect issuing summons or not.

This paper aims to examine the criminal offences occurring in New York City using data rather than testimony. Seeing the population, safety measures can only be taken for the public are justified by doing visualizations on it to obtain insights into the trend. To reach conclusions, shooting accidents, convictions, and court summons cases are carefully investigated. The NYPD Department investigated whether there is a statistically relevant association between court warrants, convictions, and shooting events with this Study. Both the essence of data processing and interpretation has been thoroughly explored.

II. Literature Survey

A. Dataset 1: NYPD Shooting Incidents

In the paper [1] , the authors (John L. Worrall, Stephen A. Bishopp, Scott C. Zinser, Andrew P. Wheeler, Scott W. Phillips) analysed the presence of any bias while making a shooting decision, based on the race of the victim. The author concluded that a particular race was one third as likely as any other victim race amongst the total shootings that occurred in the southern states of the United States of America. This paper became an inspiration to analyse the racial composition of the victims and perpetrators.

In the paper [2], the author (Adam Lankford) conducted a study on the mass shootings that occurred between 2006-2014 to try and find any correlation between the race composition of the perpetrators committing mass murder and any other criminals. The author concluded that the racial makeup of the perpetrators was almost similar to the homicides of any other type and it may be largely explained by societal causes like racial disparities and discrimination. It provided us with an insight into the racial disparity amongst the criminals.

In the paper [3], the authors (Jason R. Silva, Joel A. Capellan, Margaret A. Schmuhl, Colleen E. Mills) conduct an extensive quantitative analysis of the shootings that occurred in the USA between 1966 and 2018, which were target towards a particular gender. The author also highlighted that when compared with all other gender-based crimes, mass shooting had substantial discrepancies. This paper motivated us to take a deeper look into the gender-composition amongst the victims.

In a news article [4] published by the renowned journalism website 'Gothamist LLC', the author (Sydney Pereira) highlighted the recent rise in the number of shootings that occurred in the city of New York. The author also showcased a timeline of the number of shooting incidents from the year 2006 up until 2020. It also suggested that the trend was declining from the early 1990s till 2019 but has risen again in the year 2020.

B. Dataset 2: NYPD Arrests Data (Historic)

In paper [5], the authors Edward M. Shepard and Paul R. Blackley highlight the key influence of Drug Enforcement on the arrests and crime rate in the state of New York. They evaluate 16 regression models across a sample data of 62 New York State counties arrest rates. They consider key factors like hard drug sales, hard drug possession and Marijuana sales. Based on these factors they associated multiple arrests that resulted due to possession of drugs in separate crimes like assault, robbery, burglary and larceny.

In paper [6], the authors Andrew Golub and Henry H. Brownstein, analyze the sample data from the Arrestee Drug Abuse Monitoring Program (ADAP) program of the US government to evaluate the influence of drugs on arrestees of different age groups across various states in the country. In the research it was concluded that most arrests in the manhattan region were made for the consumption, possession and distribution of hard drugs like crack, heroin, methamphetamine and cocaine and these drugs were mostly consumed by the adult population, while the younger arrestees were more prone use consume marijuana.

In paper [7], the authors Kim Jaeok and Shawn D. Bushway study the relationship between the age of and arrestees and the severity of the crime. The researchers create coherent specific curve models and estimate the rate of criminal involvements as a function of age. Hierarchical mixed linear modelling is done across different levels and it was deduced that the level of crime was lower in younger cohorts (¡18) as compared to the older cohorts. It was also noticed that the level of crimes increased as the average age of cohorts increased. The major crimes amongst younger cohorts were the destruction of property and petty thefts, while the older arrestees resorted to Major thefts, larceny, Assault/Attacks, or drugs.

In paper [8], researchers assess the National longitudinal study of adolescent to adult health (ADD) data of the arrestees using a multi-group longitudinal panel model. They observe that amongst different ethnic groups, black young adults had the highest arrest rate, almost 7 times that of white young adults. They evaluate key factors that play a quantitative role leading to arrests in different groups, the most dominant being the neighbourhood disadvantages, exposure to violence, parent-child bond, and family income. It was discovered that even though the black adults have a stronger family bond, and are less accustomed to alcohol consumption and drugs, the arrests rate is quite high to their white counterparts, which indicates disparity and racial bias.

In paper [9], Authors Nadia Campaniello and Evelina Gavrilova calculate the probability of arrests based on the gender of the arrestee using the data from the U.S. National Incident-Based Reporting System. They consider a period of 20 years, 1995 to 2015. They consider a wide range of criminal activities such as illegal earning, larceny, stolen property, theft, embezzlement amongst others. The conclusion was that even though the number of crimes for females was less than their counterparts, with the mean being 20.12 for females over 38.99 for males; the mean probability of arrests was relatively the same with 0.35 for females and 0.36 for the males.

C. Dataset 3: NYPD Court Summons (Historic)

In paper [10], the researchers examine court appearance data for summonses before and after implementation of the Criminal Justice Reform Act (CJRA) in New York City. They have concluded that males and adults 35-65 years old were less likely to respond to their summons compared to adolescents (16-17 years old). It was also found that summonses for littering and public consumption of alcohol were less likely to appear than those summonses for public urination and parks offences. Study shows that neighbourhood characteristics are strong predictors of the likelihood of court summons.

In paper [11], the author talks about the distinct procedures for summons: 1) Criminal summons and 2) the desk appearance ticket (DAT). Criminal summonses are issued for a number of lower-level offences that do not require fingerprinting and arrests. Whereas, a DAT is issued for a specific set of eligible misdemeanours (for a class E felony); recipients of DAT are arrested and fingerprinted and then issued a ticket indicating the date and location for their future arraignment in court. (NYCLA, 2011)

In the paper [12], the author examines the court nonappearance rate in Colorado which is over 20% for misdemeanours and traffic offences while in New Orleans, there is a nonappearance rate of 52% among traffic misdemeanour and felony hearings. Researchers have also found that court appearance rates are lower for less serious offensives relative to felonies.

In paper [13], empirical research has been done on the factors that impact the likelihood of court appearance, as multiple studies have found that a longer case processing time and indicators of more extensive criminal history increase the likelihood of failure to appear. A study by Siddiqi (2009) in New York City found that the odds of pretrial misconduct (which included failure to appear along with rearresting for a violent offence) were lower for individuals who were older or White, while Black, Hispanic, and younger individuals had a higher likelihood of pretrial misconduct.

III. Methodology

A. Data Acquisition

1) Dataset 1: NYPD Shooting Incidents:

- To initiate the analysis on the shooting incidents that occurred in New York City, the data was to be fetched from the official publishing authority (New York Police Department (NYPD)). There were multiple ways of extracting the data from the official owner of the dataset (NYC OpenData) like:
  - API: NYC OpenData lets you request an API published by them along with an APP Token, attained after creating an account with them.
  - CSV, XML: NYC OpenData also provides the same dataset in popular formats such as CSV XML.
  - JSON, GEOJSON: NYC OpenData also provides the same dataset in JSON as well as GEOJSON formats.
- The option of getting the dataset in the format of GEOJSON was a lucrative option as it would have allowed us to plot the points of shooting incidents on a map easily, using the geospatial data.
- However, API was chosen to be the final choice of data consumption since it was relatively customizable and provided us with a host of options like filtering the data based on features like the 'OCCUR_DATE' (Occurrence Date) of the shooting incident as well as setting a limit of the number of records we fetch from the API and We opted for utilising the longitude and latitude columns instead, at a later stage to display the points of shooting incidents onto a map. The API call was made by utilising the 'requests' package and firing a 'get' request while passing the endpoint of the API and the relevant parameters to receive the data. NYC OpenData handled authentication by needing an APP Token in the request call, which was provided to a user after creating an account on the portal and registering the project as an application (DAP_Project_NCI_Jan_21) that will access the API.
- Once, the data was attained, it went through a host of pre-processing, transformations and Databases like MongoDB and MySQL. A lot of libraries like Matplotlib, Plotly, Basemap, Seaborn, Squarify were used to create interactive plots and animations showcasing the identified inferences.

2) Dataset 2: NYPD Arrests Data (Historic):

- The data is available with the open data library [] of the city of new york in various formats such as CSV, KML and Shapefile, but since the volume is data is too large and we wish to analyze the data between 2011 and 2019, we will consume the API which is available in the JSON format.
- To access the API, we will register with the data library of New York City and create an API ID along with an access token.
- To import data into Python, we will initiate an HTTP call using the get() function [] from the request's library available with Python.
- Dealing with a large dataset of 5.2 million records over a period of 14 years, we need to ensure uniformity in

such a way that data is available for the entire intended time period, we will extract a more compatible sample size for the analysis by running a loop starting January 2011 for each week till January 2019 and will extract 20 records from each week using timedelta[], a function of the DateTime library. This way we can ensure to have varying data from the period 2011 to 2019.

- Then we take the processed data and store it in the MySQL database using MySQL.Connector library.

3) Dataset 3: NYPD Court Summons (Historic):

- The foremost and the most important task was to acquire the summons data from the NYPD website. There are various methods of extracting the data such as using API, CSV, XML or JSON format.
- Knowing that fetching all the data 5 million was not only time consuming but not useful, therefore only date from the year 2011 till 2020 was retrieved using API.
- API can be generated by registering with the data library of New York which will generate an APP token used for accessing the OpenData. The data fetched is the JSON format. Packages like requests, JSON was installed prior to fetching to make an HTTP call and a get request was made.
- After the data was successfully stored, the records were stored in MongoDB; records were fetched from MongoDB by making a connection and lastly, data was stored in a data frame for pre-processing, transformation and visualisation.
- Several packages namely Squarify, pygal, bokeh, Folium and many more are used to plot charts using interactive widgets and dashboards.

B. Data Source & Data Description

1) Dataset 1: NYPD Shooting Incidents:

a) Data Sources:

- Historic data about the shootings that occurred in New York City is accessible through an API developed by the NYC OpenData. The dataset contains information about all the shootings that occurred in the city from 2006-2020. However, for the scope of this project, we wanted to restrict the focus to only the last decade and thus, an upper and lower limit for the 'OCCUR_DATE' was introduced and the shooting incidents between 1st Jan 2010 and 31st Dec 2019 were requested from the NYC OpenData API.
- API Endpoint and Documentation:
  - https://data.cityofnewyork.us/resource/833y-fsy8.json
  - https://dev.socrata.com/foundry/data.cityofnewyork.us/833y-fsy8

b) Data Description:

- A total of 13890 records are fetched. The Extracted data is in JSON format and each record in the dataset contains the following set of information:

Key Name	Description
INCIDENT_KEY	Randomly generated persistent ID for each arrest
OCCUR_DATE	Exact date of the shooting incident
OCCUR_TIME	Exact time of the shooting incident
BORO	Borough where the shooting incident occurred
PRECINCT	Precinct where the shooting incident occurred
JURISDICTION_CODE	Jurisdiction where the shooting incident occurred.
LOCATION_DESC	Location of the shooting incident
STATISTICAL_MURDER_FLAG	Shooting resulted in the victim's death which would be counted as a murder
PERP_AGE_GROUP	Perpetrator's age within a category
PERP_SEX	Perpetrator's sex description
PERP_RACE	Perpetrator's race description
VIC_AGE_GROUP	Victim's age within a category
VIC_SEX	Victim's sex description
VIC_RACE	Victim's race description
X_COORD_CD	Midblock X-coordinate
Y_COORD_CD	Midblock Y-coordinate
Latitude	Latitude coordinate
Longitude	Longitude coordinate
Lon_Lat	Longitude and Latitude Coordinates for mapping

Fig. 1. Data Description of Dataset 1

2) Dataset 2: NYPD Arrests Data (Historic):

a) Data Sources:

- The dataset is a breakdown of the arrests made by the New York police department (NYPD) in New York City between the year 2006 to 2020. Data is collected and reviewed every six months by The Office of Management Analysis and Planning (OMAP) and includes a total of 19 factors: ranging from nature, location and time of the arrest to the demographic of the arrestee. The data is available in the public domain for exploration of the nature of law enforcement in the city of New York by the NYPD.
- API Endpoint and Documentation:
  - https://data.cityofnewyork.us/resource/8h9b-rp9u.json
  - https://dev.socrata.com/foundry/data.cityofnewyork.us/8h9b-rp9u

b) Data Description:

- The dataset consists of historical data of the arrests made by the NYPD over the period of 15 years, from 2006 to 2021. It consists of 19 factors with key features like arrest date, offence description, arrest borough, age group, gender, and race of the perpetrator along with the longitude and latitude of the arrest. The shape of the data obtained is 8440 X 19.

Field Name Description	
COLUMN NAME	DESCRIPTION
ARREST_KEY	Random ID generated for each arrest
ARREST_DATE	Exact date of arrest of the event
PD_CD	Internal code used for incident classification(3 digit)
PD_DESC	Description about the PD code
KY_CD	A more granular Internal code than PD_CD(3 digit)
OFNS_DESC	Description if the offence committed
LAW_CODE	Law code charges corresponding to the NYS Penal Law
LAW_CAT_CD	Level of offense: felony, misdemeanour, violation
ARREST_BORO	Borough of arrest. B(Bronx), S(Staten Island), K(Brooklyn), M(Manhattan), Q(Queens)
ARREST_PRECINCT	Precinct where the arrest occurred
JURISDICTION_CODE	Jurisdiction responsible for arrest: 0(Patrol), 1(Transit) and 2(Housing)
AGE_GROUP	Perpetrator's age within a category
PERP_SEX	Perpetrator's sex description
PERP_RACE	Perpetrator's race description
X_COORD_CD	Midblock X-coordinate for New York State Plane
Y_COORD_CD	Midblock Y-coordinate for New York State Plane
Latitude	Latitude of the arrest location, coordinate for Global Coordinate System
Longitude	Longitude of the arrest location, coordinate for Global Coordinate System

Fig. 2. Data Description of Dataset 2

3) Dataset 3: NYPD Court Summons (Historic):

a) Data Sources:

- The dataset is about the historic court summons that has had happened between the timeframe of 2010-2020. The data has been made easily accessible by the New York Police Department which is available in the public domain. It contains information about the court summons and its factors like race, sex, borough, demographics, summons category type etc.
- API Endpoint and Documentation:
  - https://data.cityofnewyork.us/resource/sv2w-rv3k.json
  - https://dev.socrata.com/foundry/data.cityofnewyork.us/sv2w-rv3k

b) Data Description:

- The data extracted from the API was in JSON format. The structure of the data is depicted in Figure No. The size of the JSON file was 5.9 MB having 46,999 records

COLUMN NAME	DESCRIPTION
SUMMONS_KEY	Randomly generated persistent ID for each violation
SUMMONS_DATE	Exact date of violation for the reported event
OFFENSE_DESCRIPTION	Description of the violation committed
LAW_SECTION_NUMBER	NYS Penal Law and Local Law section number
LAW_DESCRIPTION	Description of the law dictionary
SUMMONS_CATEGORY_TYPE	General description of the violation category
AGE_GROUP	Perpetrator's age within a category
SEX	Perpetrator's sex description
RACE	Perpetrator's race description
JURISDICTION_CODE	Jurisdiction responsible for issued violation
BORO	New York city boroughs
PRECINCT_OF_OCCUR	Precinct where violation was issued
X_COORDINATE_CD	Midblock X-coordinate
Y_COORDINATE_CD	Midblock Y-coordinate
LATITUDE	Latitude coordinate
LONGITUDE	Longitude coordinate
LON_LAT	Latitude and Longitude for mapping

Fig. 3. Data Description of Dataset 3

C. Database Management

The data for the project is being consumed in a semi-structured format using the NYC Open Data API. The data is received in a JSON format. We will be using MongoDB to store the data as it is a schema-less database and does not require complicated join for linking different records, it also provides scalability and does not require mapping of each object with database objects. Once the data is stored in the database, we will fetch it for data cleaning and data pre-processing i.e., remove the null values, duplicate values, and transform each column like arrest date into a more computable format. Once the transformation is completed, we will save the data into a relational database like MySQL and created structured tables with rows and columns.

D. Process Flow Diagrams

The Process flow Diagrams for Dataset 1(NYPD Shooting Incidents), Dataset 2(NYPD Arrests Data (Historic)) and Dataset 3(NYPD Court Summons (Historic)) are:

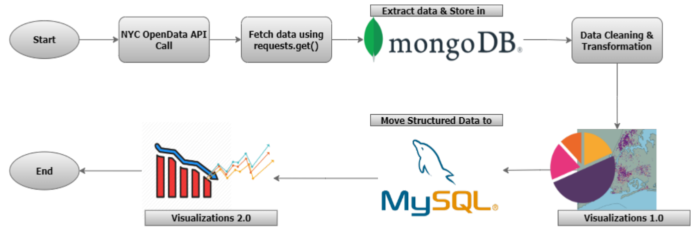

Fig. 4. Process Flow Diagram of Database 1

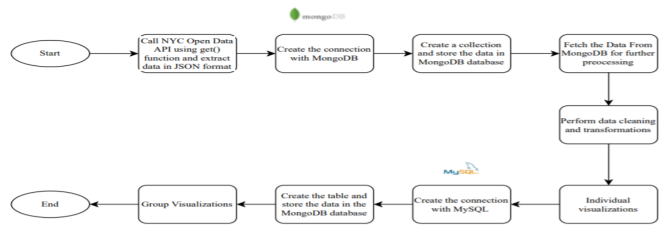

Fig. 5. Process Flow Diagram of Database 2

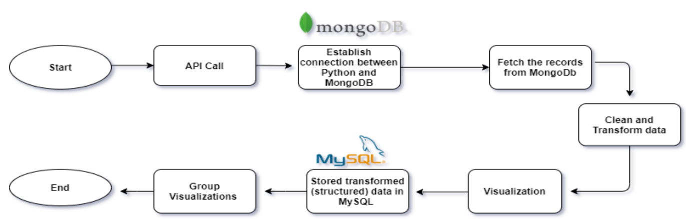

Fig. 6. Process Flow Diagram of Database 3

IV. Data Cleaning and Transformation

A. Dataset 1: NYPD Shooting Incidents

1) Data Preparation: The extracted dataset went through the following cleaning and transformation process:

- After pushing the extracted dataset to MongoDB, It was observed that there were few sets of records that had Null Values in features of Perpetrator (perp_age_group, perp_race) and victim (vic_age_group, vic_race). Such records were removed from our dataset as the Race and the Gender of the victim as well as the perpetrator are one of the topics of the focus of subsequent visualizations. The dataset was also cleaned with disregarding irrelevant like values 'U' in perp_sex and vic_sex
- It was also observed that Location_Desc also had a substantial number of Null Values and hence, the records could not be removed on this basis as this would have caused a loss of a substantial amount of data. Hence, the missing values were replaced by 'NA' for this column.

- Transformations for visualizations:
  - A column named 'day_night' was added to the dataset data frame after getting the data from MongoDB.This column was derived on 'Occur_Time' and hours 0600-1800 were considered as Day and 1800-0600 were considered to be Night. This was then used to plot a map of NYC showcasing the dispersion of shootings that occurred in Day/Night.
  - While plotting visualizations to showcase the year-wise count of shootings and race dispersion 'Year' was extracted from the 'Occur_Date' column.
  - A line chart on a monthly basis required the 'Month' to be derived from the 'Occur_Date' column
  - A clock plot showcasing the time-wise distribution of the shooting incidents was made after making an 'Hour' column from the 'Occur_Time'.
  - A bivariate encoding in the 'Statistical_Murder_Fag' was switched to Boolean values while plotting points of shooting incidents on the Map of NYC.
- The data was then moved to MySQL and typecasted into the appropriate Data Type for future use, through MySQL queries.

B. Dataset 2: NYPD Arrests Data (Historic)

1) Data Preparation:

- When fetching the data from the API, we will transform the arrest_date, a DateTime factor to a standard format(YYYY-MM-DDTHH:MM: SS) so it is easier to compute.
- After storing the data into MongoDB, we fetch it and remove columns that do not help our purposes like _id, pd_cd, and a point value factor "lon_lat".We remove the point value because we already have a separate column for latitude and longitude, which can be used to create a map and further processing.
- Then, we check the columns for null values along with duplicate rows. After treating both, the shape of the dataset becomes 8030 X 17.
- Then, we check each column based on their unique values to check if there are any unknown values in the data, we find that in some cases the race of the perpetrator is not known, we drop those rows, and the shape of the dataset becomes 7963 X 17.
- Then we take the processed data and store it in the MySQL database using MySQL.Connector library.

C. Dataset 3: NYPD Court Summons (Historic)

1) Data Preparation: The data after being fetched from MongoDB is not fully prepared for making visualisation. Therefore, the very basic first step is to know your dataset by cleaning and transforming it. The steps used for data preparation are as follows:

- The data that has been loaded in the dataframe has 46,999 rows and 19 features. The possibility of having numerous amounts of Null Values and missing values is quite high. The columns '_id' and 'lat_lon' were removed before cleaning because '_id' is an auto-generated string that is unique to every record which would not be used in the future 'lat_lon' is only the repetition of the coordinates for latitude and longitudes, so it was removed for reducing the redundancy.
- To check the number of the null values, the info method was used which returns the number of non-null columns in every column. It was seen that race, sex, summons_cateory_type, age_group, boro and law_description has a good number of null values. For more clarity, a dataframe with each column's name with the null values' percentage was formed which showed that in the race had approximately 24% of the values null, sex had 9.9% of the values null followed by summons_cateory_type having 4%.
- All the null values were removed using the dropna function. The resultant dataframe was reduced to having only 9677 rows.
- Even after removing the null values, the data was not perfect to move further; columns such as age, sex and race still had some unwanted values which were to be handled. The race column had some 'unknown' and 'other' which were inappropriate. Similarly, the age_group column had ages like 340, 963 which also did not make any sense and sex had an unknown 'U'. After handling all the missing values, the shape of the dataframe became 8684 X 18.
- A new column was derived to the dataframe named 'Year' from the column 'summons_category_field' for a better understanding of a year-wise trend of court summons.
- All this pre-processing and transformation was implemented using libraries such as NumPy and pandas which was an essential step in moving forward with the visualisations. As seen, the original dataset had 46,999 records and the final dataset after the transformation has 8,369 records shows the power of transformation that helps in reducing redundancy.

V. CHARTS, PLOTS ANIMATIONS

A. Dataset 1: NYPD Shooting Incidents

Various inferences were made from the data of the shooting incidents across the city of New York. The visualization libraries used to plot visualizations and animations are Matplotlib, Seaborn, Basemap, Squarify, Plotly :

- The Map of New York City clearly states that a large proportion of the shooting incidents occur during the night and the streets of NYC are much more perilous during the Nights when Compared to Day (Fig .7)
- The stacked Bar chart showcases that number of shootings has declined every year until 2019. However, the racial composition of the 'victim race' has not changed much and most numbers of victims in shooting incidents are still 'Blacks' (Fig .8)
- The Line chart of shooting incidents for each year showcases the numbers rise from April to July and then

drop from July to September, while the other months experience a somewhat similar number of shootings.(Fig .9)
- The Clock Plot represents a clock, with the number of shootings that occurred in each hour of the day, comprising of both Statistical Murders and Non-Statistical Murders. It can be inferred that the number of shootings is higher between 1600 HRS 0300 HRS, peaking at 2100 HRS (Fig.10)
- The Grouped Bar chart showcases that a large chunk of Perpetrators and victims both fall within the age groups of 18-24 and 25-44. (Fig.11)
- The histogram shows the number of shooting incidents in each Borough of New York City. It shows that Staten Island is the safest borough in NYC.(Fig.12)
- The animated NYC map showcases the number of statistical and Non-Statistical murders from 2010-2019.(Fig.13)
- The Treemap showcases the Gender divide between the perpetrators and victims.(Fig.14)

B. Dataset 2: NYPD Arrests Data (Historic)

Analyzing the data from the arrest's dataset of the NYPD, we were able to make multiple inferences. We were able to observe multiple patterns that emerge in the city based on the number of offences and types of crimes that are dominant in the city. We will create multiple maps, graphs, and animation to highlight several key relationships between the demography of the city and the crimes committed.

- A map of the New York city plotting the location of all the arrests that have taken place in the city over the period of 2011 to 2019. (Fig.15)
- A line graph highlighting the number of arrests based on the age group of the arrestees that have taken place over the period. (Fig.16)
- A horizontal bar graph showcasing the arrestees based on their gender. (Fig.17)
- An area graph comparing the number of arrests that taken place in each borough of New York City. (Fig.18)
- An animation comparing the number of arrests in each year and comparing that to the major crimes that the arrests were made for. (Fig.19)
- A scatter plot highlighting the frequency of individual crimes over 9 years. (Fig.20)
- A bar graph highlighting the number of arrests made based on the ethnic race of the perpetrator. (Fig.21)
- A bar graph highlighting the most frequent crimes committed based on the age group of the arrestee. (Fig.22)

C. Dataset 3: NYPD Court Summons (Historic)

Visualizations withdrawn from the summons dataset was useful to gain knowledge about the pattern of crime over the last decade. Visualizations and animations were made using libraries Folium, Bookeh, Plolty, Pygal, Matplotlib, Seaborn and many others. The use of animations made it easy to read the charts using interactive widgets.

- The Bar chart showcases the top 10 offences across all boroughs. It is found that Marijuana, consumption of alcohol and trespass were at the top. (Fig.23)
- A Catplot describes the relationship between year, race and boroughs. It was observed that every year, the highest number was summons were called in Brooklyn followed by Bronx and Queens. In every borough, 'blacks' were found to be in majority. (Fig.24)
- A Grouped Bar chart showcases that in every borough, the Males' ratio was more than 6 times of Females' (Fig.25)
- Using the Stacked Bar chart, it is seen that every borough has a majority race of blacks.(Fig.26)
- Histogram showcases that the mean age group of the people being called for court summons fall under the age group of 25-44.(Fig.27)
- A Map is built which showcases the area of the boroughs which were active over the last decade. (Fig.28)
- Dashboard showcases a pivot table that can be used to make all types of plots such as heatmap, bar chart, line chart. Using the same, a bar chart is built boro by year. (Fig.29)

D. Group Visualisation

All the three datasets were coagulated and a few common inferences were made:
- Staten Island is one the safest boroughs in NYC, considering the low numbers of Arrests, Shooting Incidents and Court Summons (Fig.30)
- Females account for roughly 50% when it comes to being arrested or summoned by the court. However, when it comes to being a perpetrator in a shooting incidents, the numbers are very low. (Fig.31)
- There is a probable inclination towards a few Races, when it comes to Arrests, Shooting Incidents and Court Summons (Fig.32)

VI. Conclusion

In this article, we attempted to extensively investigate the factors influencing crime rates, convictions, and juvenile court summons in New York City. This article provides an in-depth examination of the factors influencing crime prevention on various levels. Key data from these databases provide insights into the impact of ethnicity, ethnicity, age group, and location on violence. We collected data from the NYPD's shooting shootings, convictions, and criminal summons in an attempt to detect racial profiling, hate crime, and gender bias.

This project visualizes the facts using a dataset that was obtained via API and then uploaded to MongoDB and then to MySQL. The final derivations were derived from the common features of boroughs, gender, and race, which demonstrated a similarity between all three.

VII. ACKNOWLEDGEMENT

The authors of this study would like to express their heartfelt gratitude to the National College of Ireland (NCI) and Prof. (Dr.) Anthanasios Staikopoulos for their unwavering assistance, direction, and motivation, without which this project would have been difficult to complete. In these trying times, the writers did their best to contribute their analysis and interpretation in order to encourage the exchange of reliable and credible facts.

REFERENCES

[1] J. L. Worrall, S. A. Bishopp, S. C. Zinser, A. P. Wheeler, and S. W. Phillips, "Exploring bias in police shooting decisions with real shoot/don't shoot cases," *Crime & Delinquency*, vol. 64, no. 9, pp. 1171–1192, 2018.

[2] A. Lankford, "Race and mass murder in the united states: A social and behavioral analysis," *Current Sociology*, vol. 64, no. 3, pp. 470–490, 2016.

[3] J. R. Silva, J. A. Capellan, M. A. Schmuhl, and C. E. Mills, "Gender-based mass shootings: an examination of attacks motivated by grievances against women," *Violence against women*, p. 1077801220981154, 2021.

[4] S. Pereira, "Gothamist LLC," https://gothamist.com/news/after-painfully-violent-2020-nyc-shootings-continue-spike, 2021.

[5] E. M. Shepard and P. R. Blackley, "Drug enforcement and crime: Recent evidence from new york state," *Social Science Quarterly*, vol. 86, no. 2, pp. 323–342, 2005.

[6] A. Golub and H. H. Brownstein, "Drug generations in the 2000s: an analysis of arrestee data," *Journal of drug issues*, vol. 43, no. 3, pp. 335–356, 2013.

[7] J. Kim and S. D. Bushway, "Using longitudinal self-report data to study the age–crime relationship," *Journal of Quantitative Criminology*, vol. 34, no. 2, pp. 367–396, 2018.

[8] C. Schleiden, K. L. Soloski, K. Milstead, and A. Rhynehart, "Racial disparities in arrests: a race specific model explaining arrest rates across black and white young adults," *Child and adolescent social work journal*, vol. 37, no. 1, pp. 1–14, 2020.

[9] N. Campaniello and E. Gavrilova, "Uncovering the gender participation gap in crime," *European Economic Review*, vol. 109, pp. 289–304, 2018.

[10] C. Cuevas, E. Grimsley, P. Chauhan, and K. Mulligan, "Criminal and civil summons court appearance: Predictors of timely response to summonses for lower-level offenses in new york city," *Criminology, Crim. Just. L & Soc'y*, vol. 20, p. 1, 2019.

[11] H. R. Zettler and R. G. Morris, "An exploratory assessment of race and gender-specific predictors of failure to appear in court among defendants released via a pretrial services agency," *Criminal Justice Review*, vol. 40, no. 4, pp. 417–430, 2015.

[12] S. Q, "Predicting the likelihood of pretrial failure to appear and/or re-arrest for a violent offense among new york city defandants: An analysis of the 2001 dataset," *Criminal Justice*, vol. 46, pp. 1–45.

[13] D. B. Rottman and A. Tomkins, "Public trust and confidence in the courts: What public opinion surveys mean to judges," 1999.

VIII. APPENDIX: IMAGES

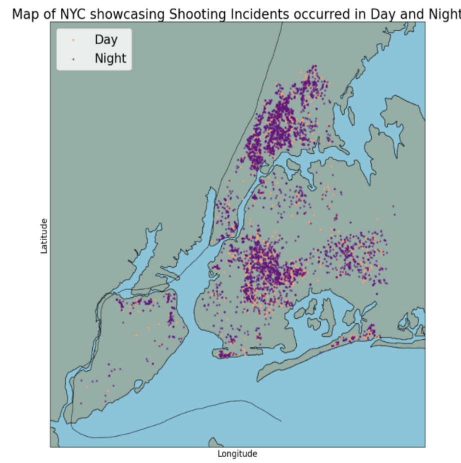

Fig. 7. Map Of NYC Showcasing Shooting Incidents in Day & Night

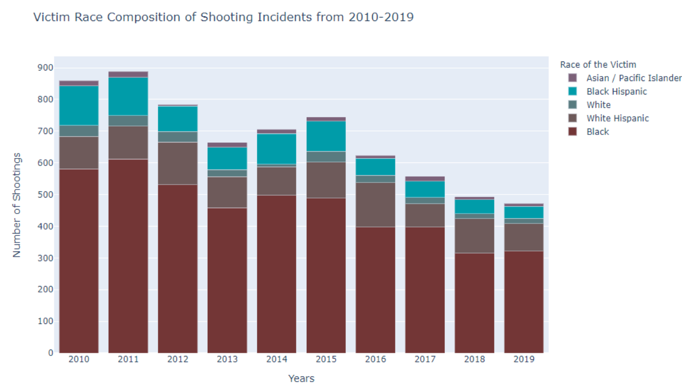

Fig. 8. Victim Race Composition

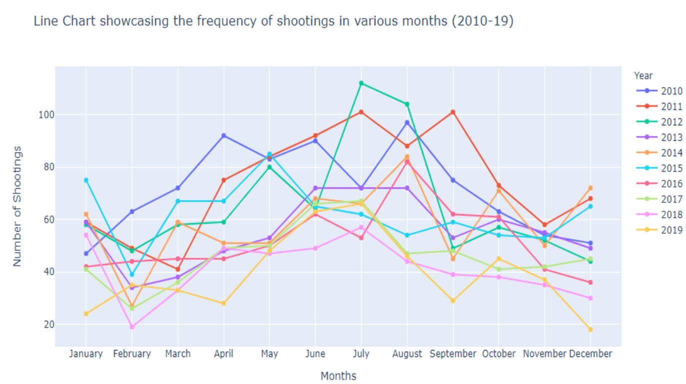

Fig. 9. Line Chart of Shootings in all months from 2010-2019

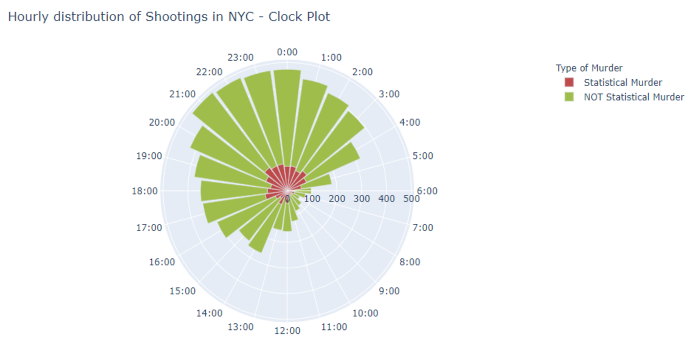

Fig. 10. Clock Plot showing timings of Shootings

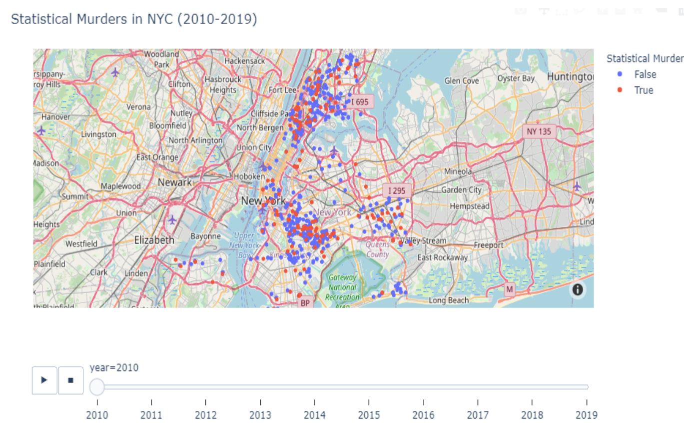

Fig. 13. Animated MAP of NYC, showcasing Statistical Murders (2010-19)

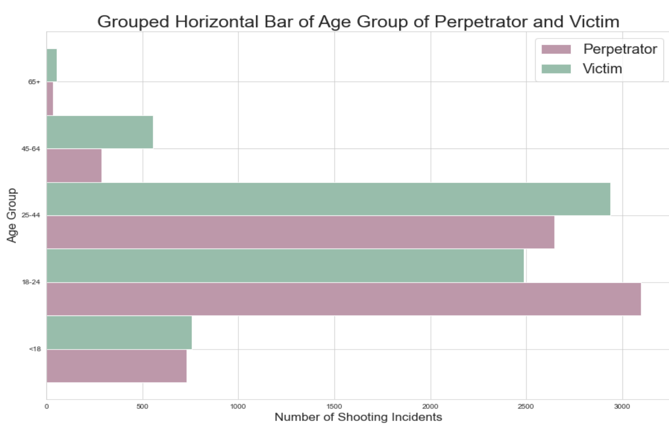

Fig. 11. Grouped Age Bar of Perpetrators and Victims

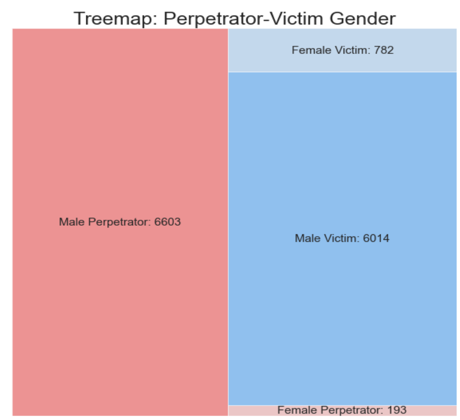

Fig. 14. Treemap of Perpetrator and Victim Gender

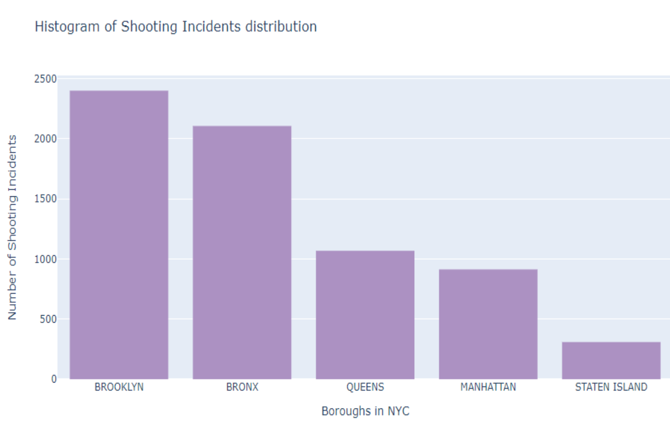

Fig. 12. Histogram of shootings in Boroughs

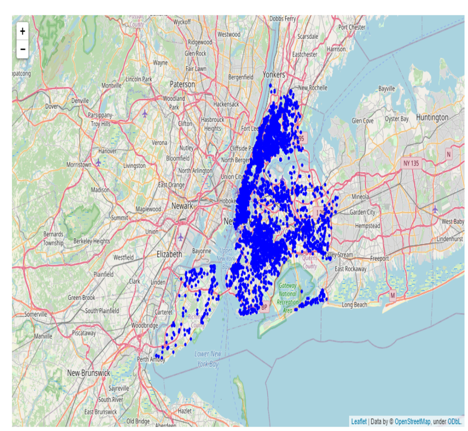

Fig. 15. NYC Map Arrests

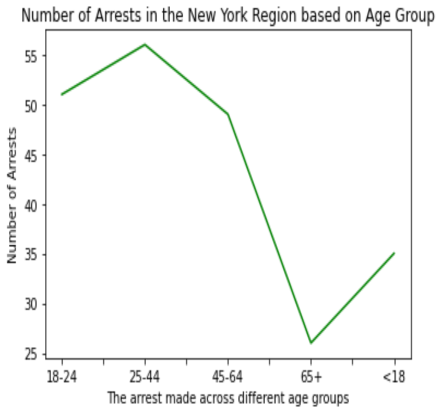

Fig. 16. Arrests Line Chart

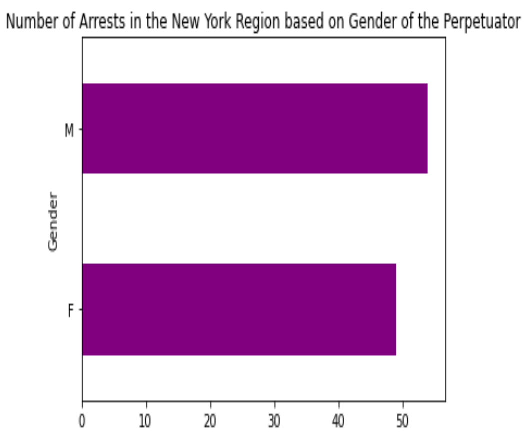

Fig. 17. Gender Bar Chart

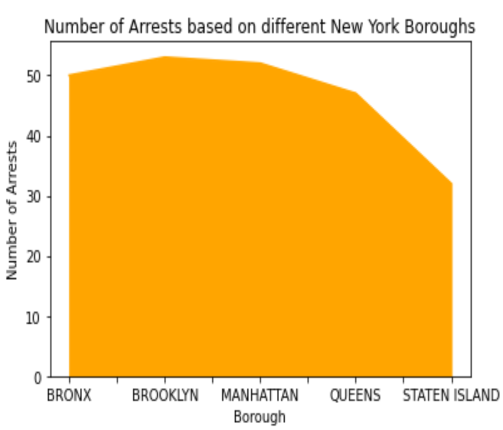

Fig. 18. Area Graph: Borough

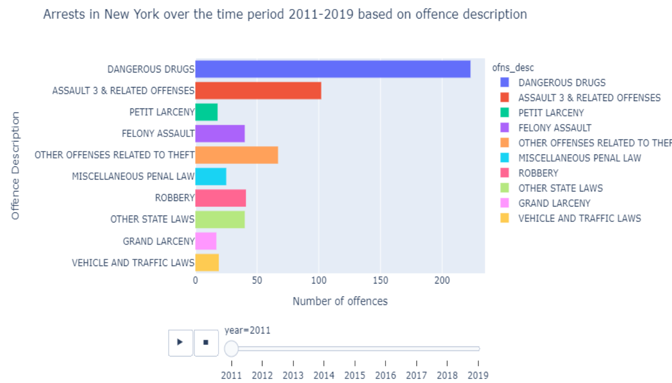

Fig. 19. Animation: Arrests each year

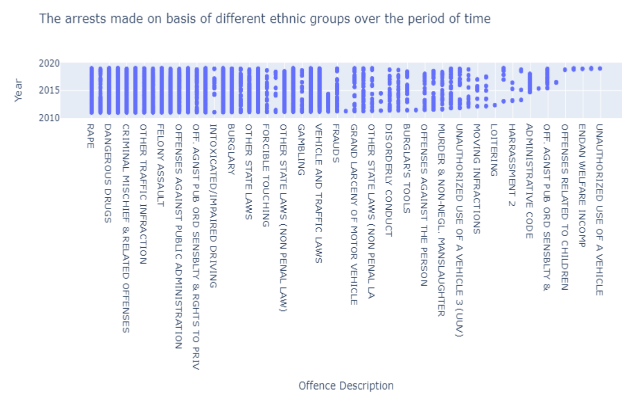

Fig. 20. Scatter Plot of Arrests

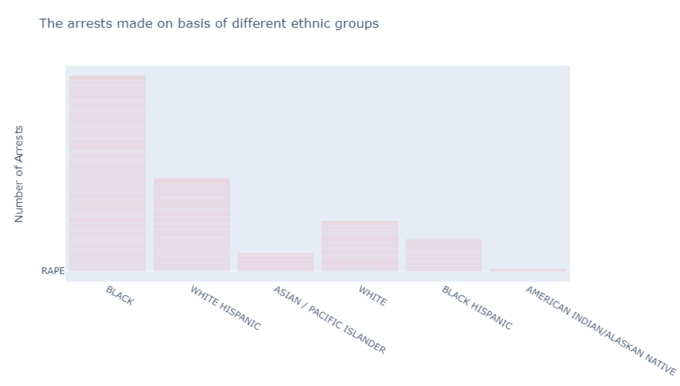

Fig. 21. Bar Chart of Ethnicity of the Perpetrator

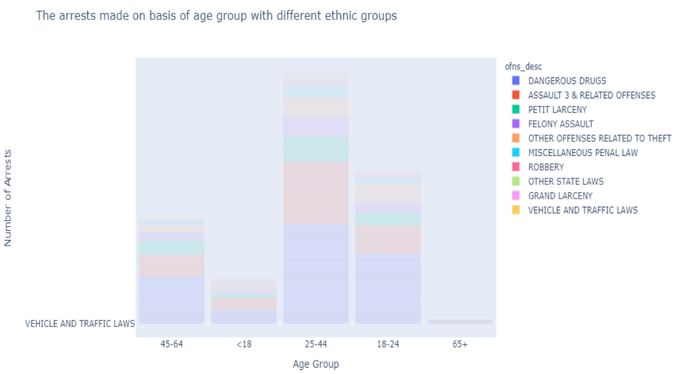

Fig. 22. Bar Chart of Age Group of the Perpetrator

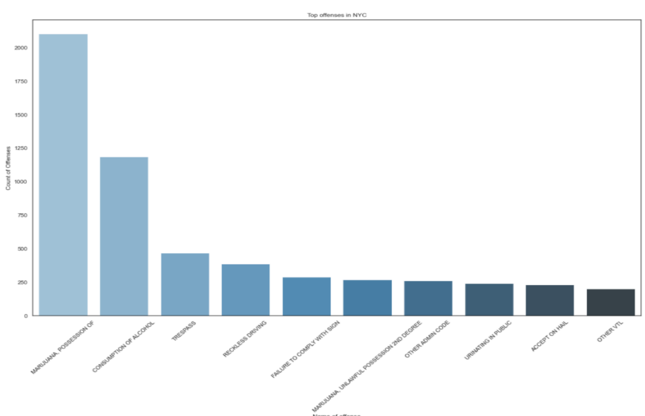

Fig. 23. Bar Chart of Top offences

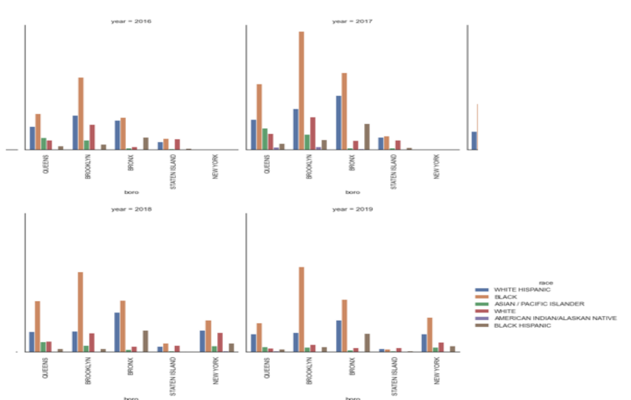

Fig. 24. Catplot

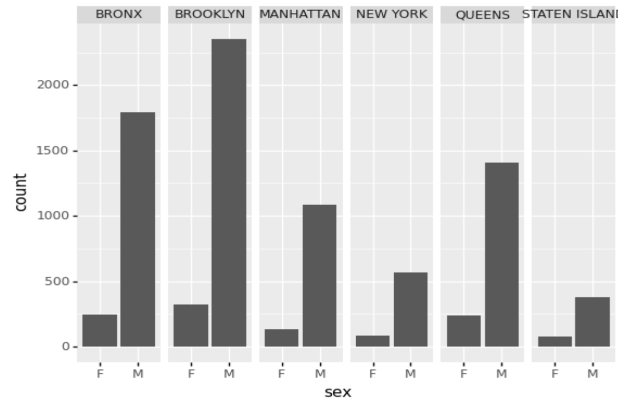

Fig. 25. Gender composition across Boroughs

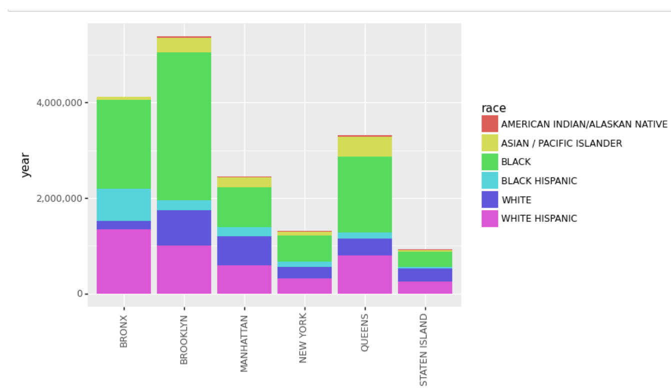

Fig. 26. Race composition across borough

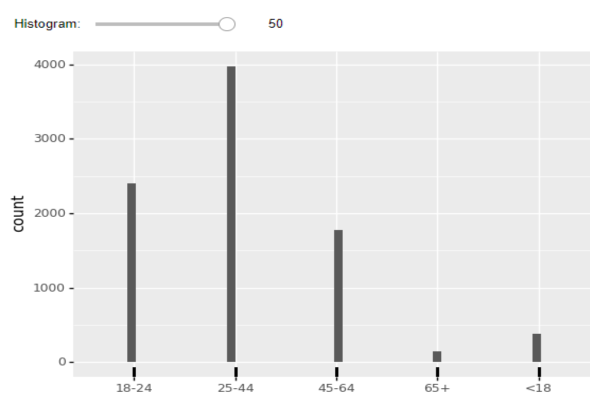

Fig. 27. Age Group of people called for summons

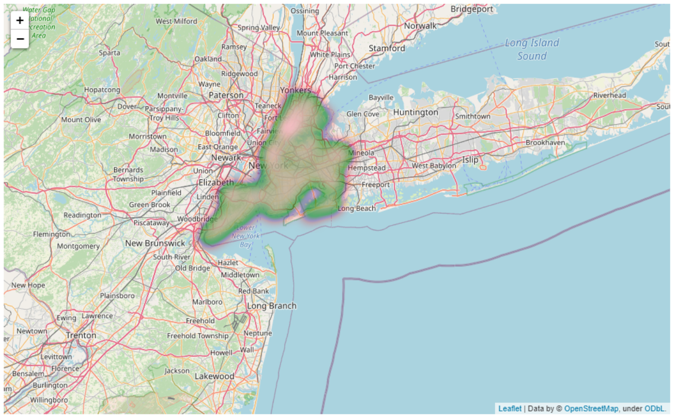

Fig. 28. Map of NYC

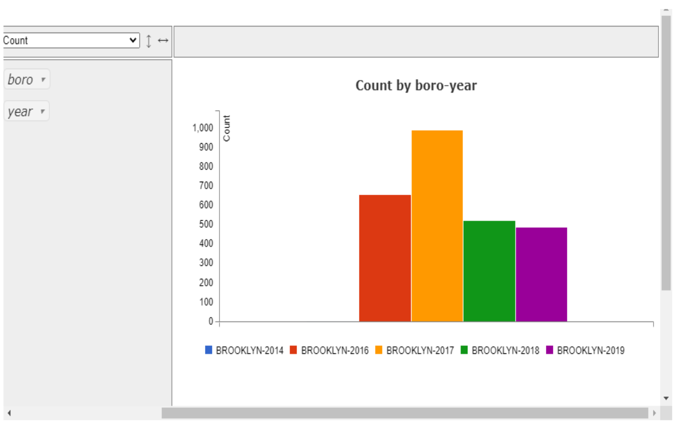

Fig. 29. Map of NYC

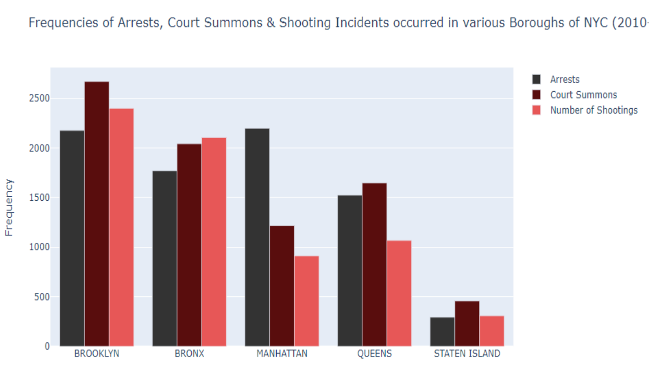

Fig. 30. Grouped Bar Chart of Arrests, Shooting Incidents and Court Summons

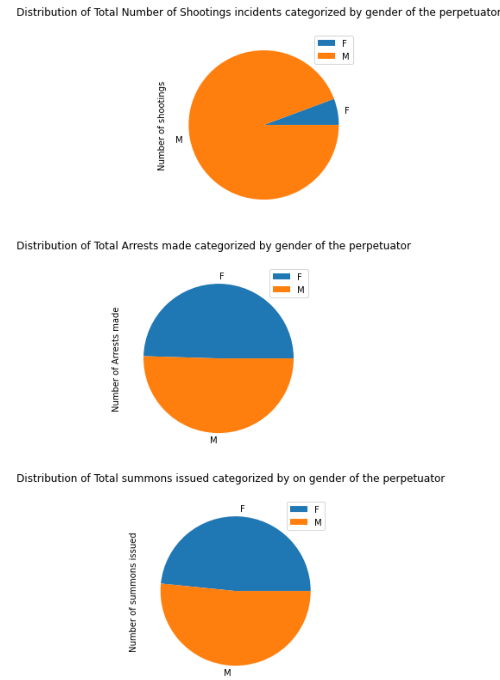

Fig. 31. Pie Chart of Gender composition amongst Arrests, Shooting Incidents and Court Summons

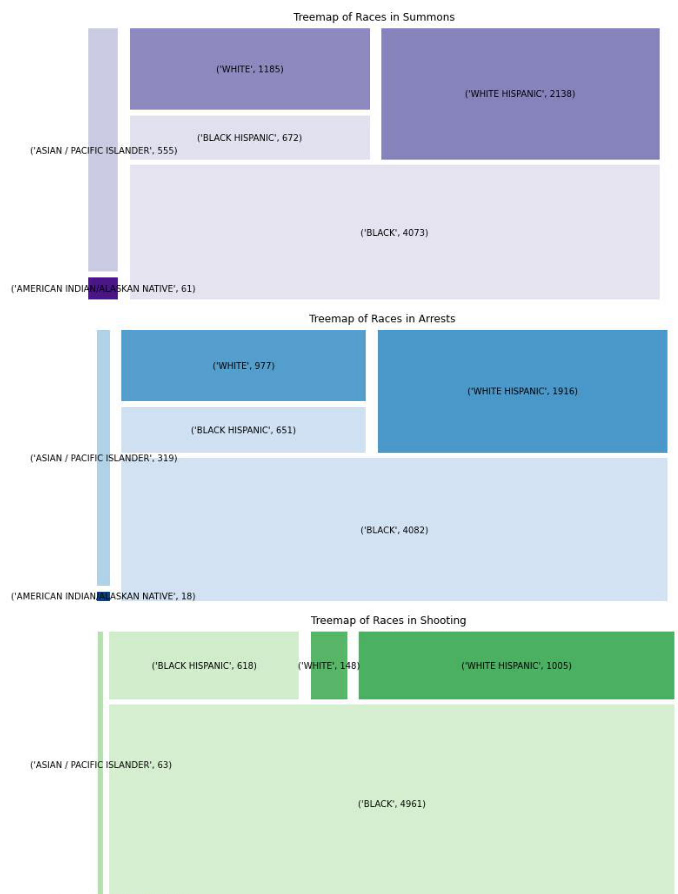

Fig. 32. Treemap of Race composition amongst Arrests, Shooting Incidents and Court Summons