****

**Data Visualization – Summer’23**

**Exploring the Urban dynamics through datasets made of Hate crime, Arrest, Shooting, Eviction and Collision of New York City.**

***Prepared By:***

***Fuad Bin Saif***

***U41393121***

[***fuadbinsaif@usf.edu***](mailto:fuadbinsaif@usf.edu)

**Introduction:**

Understanding complicated societal dynamics is more important than ever in a time when technology is developing quickly, and data is becoming more readily available. Using data on arrests, shootings, evictions, car collisions and hate crimes, in this data visualization project I tried to shed light on the many facets of our society. This initiative intends to give policymakers, scholars, and the public alike a comprehensive view on these important topics and enable educated decision-making using data visualization.

We all are aware of the enormous influence these topics have on people and communities as we work on this project. The societal institutions, economic circumstances, and cultural dynamics of New York city are all intertwined with arrests, shootings and hate crimes; they are not isolated incidents. The goal of this project is to reveal these links and the underlying narratives that influence how we see these phenomena.

I will analyze patterns, find trends, and highlight correlations within and across various datasets through thoughtful visualizations to identify potential intersections between data on arrests, shootings, evictions, car collisions and hate crimes. To further provide a contextualized view of these social issues over time and across New York city, the project has geographical visuals.

This project’s aim is to allow significant insights as well as accessible data presentation by displaying it in the form of interactive dashboards, maps, and graphs. This encourages users to pose queries, investigate hypotheses, and come to their own conclusions. This initiative acts as a forum for discussion and investigation, fostering a deeper comprehension of the difficulties facing our society and the possible routes leading to constructive change.

This project will explore some of the research questions that I have already developed. The visuals that reveal facts, tell tales will answer these research questions. In addition to the questions there is room for more exploration using this data.

**Research Questions:**

1. Where were the arrests made in terms of location in NY city and did any race have precedence over other races?
2. What was the age range of the arrests that were made? What was the number in terms of gender?
3. What was the scenario of arrest made and shooting in NYC?
4. For vehicle collisions, what factors contributed towards collision of different types of vehicles? Analize the visuals.
5. What trend do we see in arrest, shooting, hate crime and collision?
6. What does a borough wise comparative analysis of arrest, shoot, collision, eviction and hate crime look like? Which borough had the highest percentage of the variables? What can we infer from the visuals?

**Methodology:**

All my data was collected from [data.gov](https://data.gov/) and [CityPopulation](https://www.citypopulation.de/%20). This website link data from different other websites like NY city website or NYPD website. Since one place had a collection of data, I did not mention the other sources separately. The datasets that I have included are:

**Arrest:** This data contains various information like arrest location, date, precinct, borough etc. about arrests made in in New York city from the year 2006.

**Shooting:** This dataset contains all shooting that took place in NY city from the year 2006 to 2022. This dataset contains columns like borough, occur date, victims’ race, victims’ sex etc.

**Hate Crime**: This is the dataset about hate crime committed in NY city from the year 2019 to 2023. This contains various information about hate crime which includes borough, Motive description, offense description etc.

**Eviction**: This data is about the evictions made in NY city from 2017 to 2023. This dataset also has various information about the eviction.

**Collision**: This is the car collision data of NY city from the year 2012 to 2023. This dataset contains information about the collision vehicle and factors influencing the crash.

Please refer to the TWBX file for details about the attributes of each table.

**Data Processing:**

The datasets that I have used are very large in size due to the volume and the details of information they contain. I have decided to use such large datasets because this means that we can use this project to create more visualization than that is used to answer the given research questions.

But before combining the dataset I had to do some data cleaning. There was room for more furnishing of data but since this project has a different focus, I did not spend much time on this part. I have used R studio to do data cleaning. The code is as follows:

getwd()

setwd("C:/Study Materials/USF Courses/Summer 23/Data Visualization/Project related Data/NY City Data/Final Datasets")

library(readxl)

library(readr)

library(dplyr)

#Following is the processing of the data related to Hatecrime and shooting.

NYPDshooting<- read\_csv("NYPD\_Shooting\_Incident\_Data\_\_Historic\_.csv")

unique(NYPDshooting$BORO)

NYPDhatecrimes<- read\_csv("NYPD\_Hate\_Crimes.csv")

unique(NYPDshooting$BORO)

unique(NYPDhatecrimes$'Patrol Borough Name')

colnames(NYPDhatecrimes)[c(6)] <- c("BORO")

colnames(NYPDhatecrimes)

colnames(NYPDhatecrimes) <- trimws(colnames(NYPDhatecrimes))

unique(NYPDhatecrimes$`BORO`)

print(NYPDhatecrimes$`BORO`)

NYPDhatecrimes <- NYPDhatecrimes %>%

mutate(BORO = case\_when(

BORO == 'PATROL BORO BRONX' ~ 'BRONX',

BORO == 'PATROL BORO BKLYN SOUTH' ~ 'BROOKLYN',

BORO == 'PATROL BORO BKLYN NORTH' ~ 'BROOKLYN',

BORO == 'PATROL BORO MAN SOUTH' ~ 'MANHATTAN',

BORO == 'PATROL BORO MAN NORTH' ~ 'MANHATTAN',

BORO == 'PATROL BORO QUEENS NORTH' ~ 'QUEENS',

BORO == 'PATROL BORO QUEENS SOUTH' ~ 'QUEENS',

BORO == 'PATROL BORO STATEN ISLAND' ~ 'STATEN ISLAND',

TRUE ~ BORO

))

print(nrow(NYPDhatecrimes))

write.csv(NYPDhatecrimes, "Hatecrime.csv", row.names = FALSE)

write.csv(NYPDshooting, "Shooting.csv", row.names = FALSE)

#the above codes can change the column names and column values

#Following is the processing of the data related to Collision.

Collisions<- read\_csv("Motor\_Vehicle\_Collisions\_-\_Crashes - Copy.csv")

colnames(Collisions)[c(3)] <- c("BORO")

Collision\_f <- Collisions %>%

filter(!is.na(BORO))

nrow(Collisions)

write.csv(Collision\_f, "Collision.csv", row.names = FALSE)

#Following is the processing of the data related to Restaurants.

Restaurant<- read\_csv("DOHMH\_New\_York\_City\_Restaurant\_Inspection\_Results.csv")

Restaurant\_f1 <- Restaurant\_f %>%

filter(BORO != '0')

Restaurant\_f <- Restaurant %>%

filter(!is.na(ACTION))

rm(Restaurant\_f2)

unique(Restaurant\_f1$`BORO`)

Restaurant\_f1 <- Restaurant\_f1 %>%

filter(!is.na(ZIPCODE))

write.csv(Restaurant\_f1, "Restaurant.csv", row.names = FALSE)

#Following is the processing of the data related to arrests.

Arrests<- read\_csv("NYPD\_Arrests\_Data\_\_Historic\_ - Copy.csv")

colnames(Arrests)[c(9)] <- c("BORO")

unique(Arrests$`BORO`)

Arrests <- Arrests %>%

filter(!is.na(BORO))

Arrests <- Arrests %>%

mutate(BORO = case\_when(

BORO == 'B' ~ 'BRONX',

BORO == 'K' ~ 'BROOKLYN',

BORO == 'M' ~ 'MANHATTAN',

BORO == 'Q' ~ 'QUEENS',

BORO == 'S' ~ 'STATEN ISLAND',

TRUE ~ BORO

))

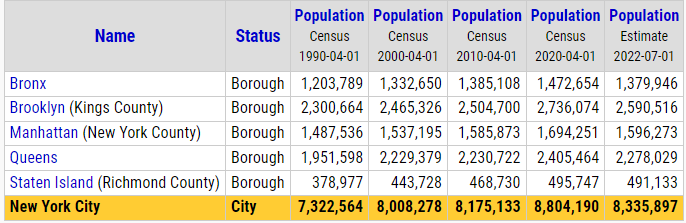
write.csv(Arrests, "Arrest.csv", row.names = FALSE)

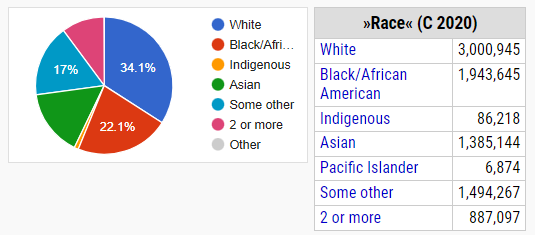
The primary purpose of the above code was to set the proper borough name for each datasets since it is the borough name that will connect all the dataset together. Also, I removed null values from borough so that the dataset contains less noise.

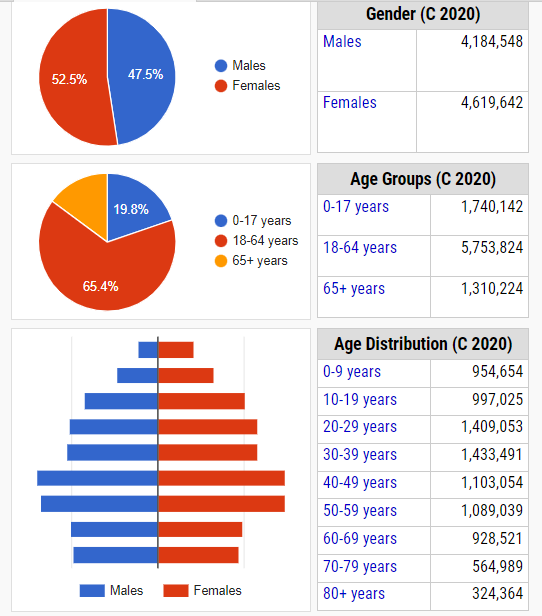
**Analysis:**

**Population scenarios:**

Before starting to answer the research questions let is look at the population of different boroughs of NY city:





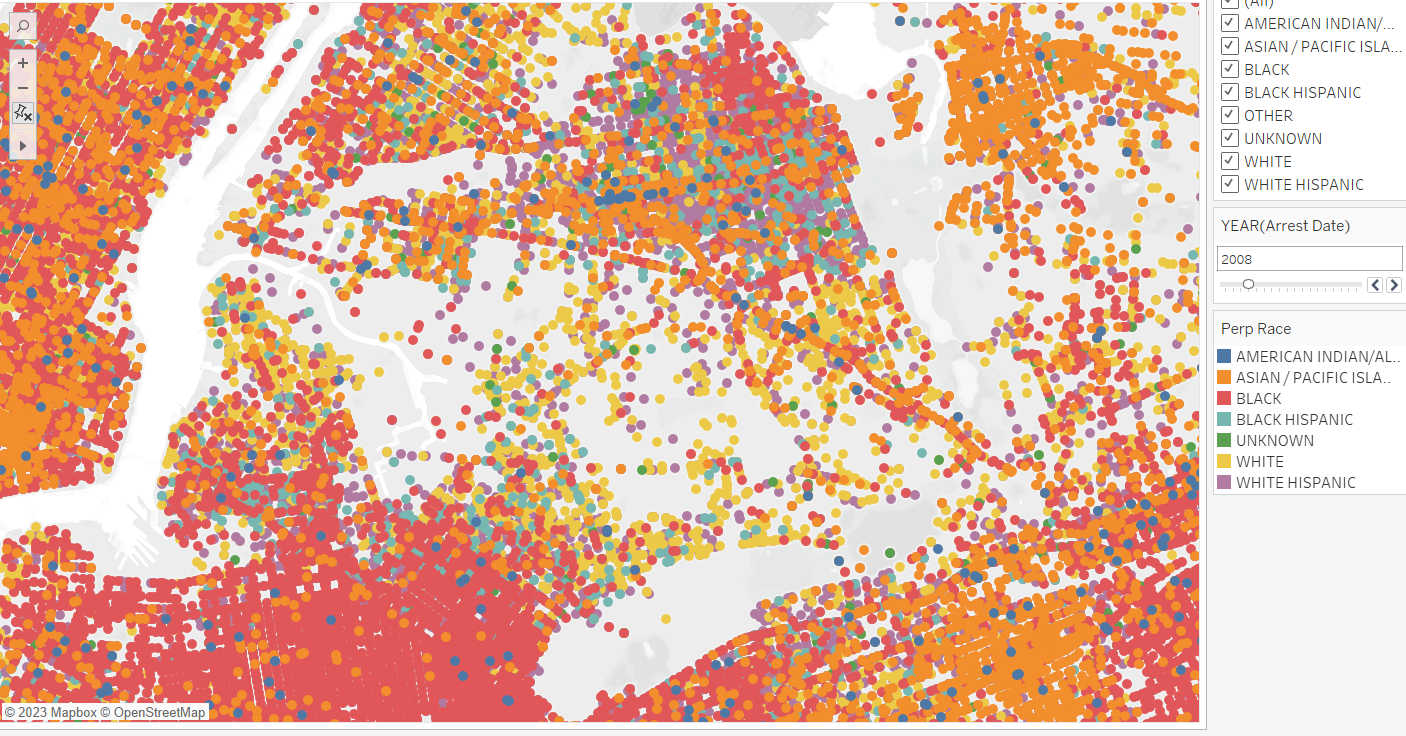


Since some of the visuals will be related to the number of incidents that happened, it is imperative to take the population into consideration as well. For example, we will see Staten Island has ranked lowest in all variables. But if we look at the population of this borough, we will see that it has the lowest population in comparison to all other boroughs. Since we can dig deep while addressing the issue of urban dynamics for the sake of preparing a report that is not too difficult, I have not linked any data related to population in borough other than the above-mentioned table. The above-mentioned table gives us a rough idea if population could be a big factor in deciding the number of instances in some variables. But I have included the population related question in the additional research questions. A normalization of the data in respect to the number of populations is important for getting clearer pictures. The next focus of this project should be to normalize the results by adding a population related dataset.

**1) Where were the arrests made in terms of location in NY city and did any race have precedence over other races?**

A map of the world

Description automatically generatedAs clearly shown on the map, the locations where any arrests were made from the year 2006. The above visual gives us the freedom to select each year separately and to dig deeper by selecting each race separately. We can zoom into the Map to see better where the arrests were made. The following image makes it clear how the arrests are plotted on the map.

To answer the second part of the question we can use the above shown visual. But that will not be a good decision since it will violate the basic theories of a good visual, that is a visual should be easily discerning. So, I have included a second visualization. This shows that Brooklyn had the highest arrest rate and Black was the highest perpetrators race. White Hispanic was high in number in Bronx, which was 427,386, which is 7.77% of the total arrests made. Black in B

A group of colorful circles with black text

Description automatically generated

**2. What was the age range of the arrests that were made? What was the number in terms of gender?**

The arrests made in New York city were divided into four groups. They were 18-24, 25-44, 45-64 and 65+. 25-44 age range had the highest number of apprehensions, and they were mostly males. As a matter of fact, in all age categories males had a significant higher number of arrests in comparison to females.

A graph of numbers and a number of people

Description automatically generated with medium confidence

We can also see a significantly higher number in the first two age categories. We can further investigate into these two age groups specially amongst to male groups. If we can find the top reasons for this spike, then we can significantly mitigate the arrest numbers in the city. Increased number of arrests also can translate into increase in crime. So digging deeper into this question could unravel many questions that will address crime problems of New York city.

**3. What was the scenario of arrests and shooting in NYC? Is there any relationship between shooting and arrest in NY city?**

Arrests and shooting showed an interesting trend where they were plotted in a time series. If we look at the arrest scenario from the year 2006, we will see that it has not increased much. We see an approximate rise of 30 thousand in the arrest in the year 2022.

A graph with lines and numbers

Description automatically generated

The good news is the number of shooting incidents decrease from 2006 to 2022. We can say in the long run that the shooting and the arrest are moving in the opposite direction. But if we do a year-by-year analysis we will see that every time arrest increase shooting decreased. From the year 2007 to 2010 they show a wave pattern and moving in the opposite direction every year. But after 2011 it shooting incidents kept on decreasing gradually and went as low as 958, which was almost half of what it was in 2006. We can investigate these few years and check what could have influenced the reduced number of shootings in these years. Such information could help us reduce unfortunate events like shooting in future or in other cities as well. During 2011 to 2020 the number of arrests kept a wavy trend. The year 2015 saw the highest number of arrests which was 5.49 million. Apart from this there were no significant changes in the arrest trend. We could try dig down borough wise or quarter wise to see if we could find a pattern. It is important to note that in the year 2020 the shooting incidents drastically increase almost to its double. This increased in the next year as well. So, the proper authority should check if there were any anomaly that broke the trend of decreased shooting. It just did not break the trend, it increased drastically.

**4. For vehicle collisions, what factors contributed towards collision of different types of vehicles?** **Analize the visuals.**

Our datasets classified all vehicles that crashed into five categories and provided the collision contribution factor for each type of vehicles.

A screen shot of a computer

Description automatically generated

As we can see driver’s inattention is the primary reason for crash of vehicle type one. Queens had the highest collision which is 79 thousand. The second highest reason was failing to yield right of way and Queens also tops this category by having 31 thousand collisions. Queens also tops the third highest reason of collision, which is backing unsafely. Although in this category Brooklyn comes very close to queens, they had 18,401 and 17,662 collisions respectively. We can say that Staten Island is the safest if we compare the first three reason of collision as it stayed in the lowest rank for all three.

If we investigate vehicle type 2, we will see that for this type of vehicle, the highest collision was in Manhattan, 20 thousand approx., and the main reason was Distraction. And Staten Island is also ranked the lowest in this category. Manhattan had 11,383 collisions attributed to other vehicular reasons. As usual Staten Island had the lowest collision in this category as well.

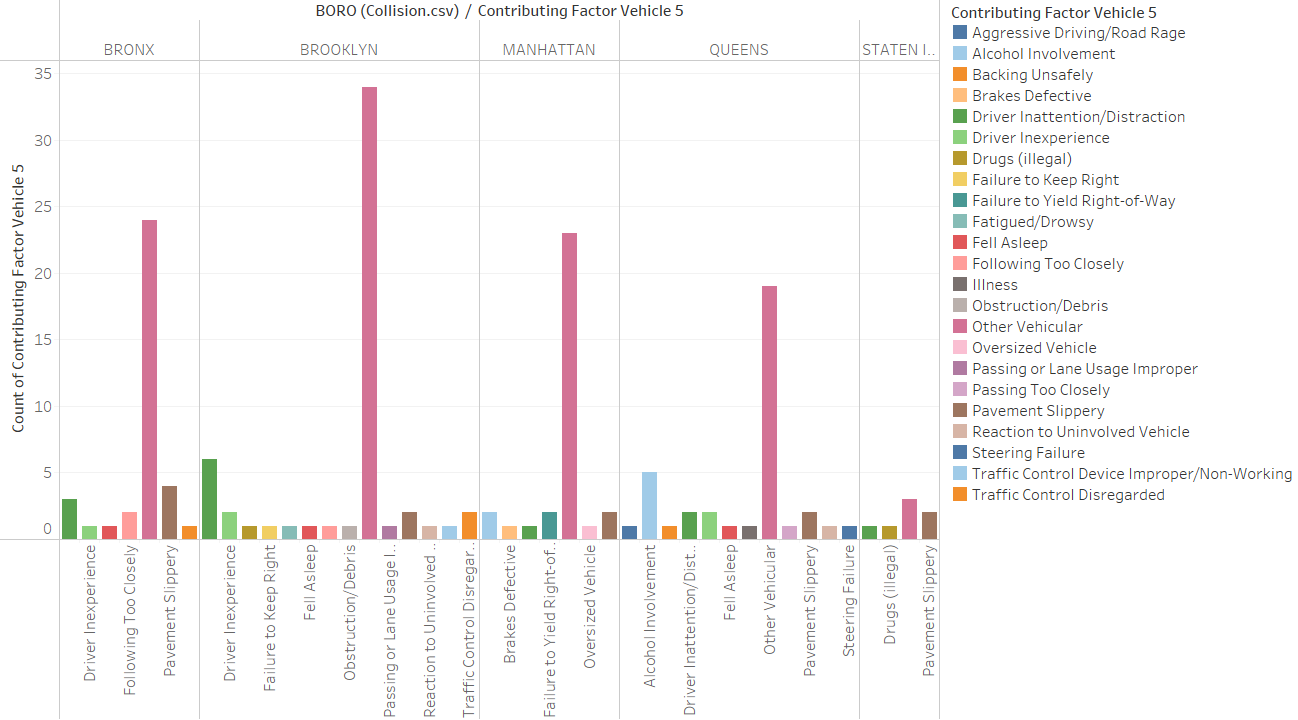
A screen shot of a computer

Description automatically generated

I have created similar visuals for the other two types of vehicles as well.

A graph with text and numbers

Description automatically generated with medium confidence



If we look into the last two types of vehicles, we will see that Brooklyn had the highest collision and the primary reason for both of them was “Other Vehicular”. Let us look into the first three kinds of vehicles in an interactive dashboard as these three types of vehicles contributed mostly to the collisions.

A screenshot of a graph

Description automatically generated

We can say that Queens and Manhattan had the highest collisions and Staten Island was the safest. Drivers’ inattentiveness and other vehicular were the two main reasons for collisions. We need to see that vehicle type one had a significantly higher collision rate in Queens. It was almost 7 times higher than Staten Island in case of Vehicle type one. Although the collision number was close to Brooklyn and Manhattan as well. Brooklyn had 75 thousand and Manhattan had 65 thousand collisions. We can also see that other reasons contributed a small percentage in comparison to the primary reason. So, we can say the if we can reduce the collisions in Queens, Brooklyn and Manhattan due to driver’s inattentiveness then the total number of collisions can be reduced significantly. Additionally, if we dig deeper into “Other Vehicular” factor then we can also address the collision issue further.

**5. What trend do we see in arrest, shooting, hate crime?**

Let us look into a Dashboard that I prepared which contains the time series for Hatecrime, Shooting, Attest and Evictions. As we can see that Hatecrime increased in 2022 to 667 in the

A graph of different types of lines

Description automatically generated with medium confidence

year 2022. But we can see a drastic fall in the year 2023. But we must keep in mind that the year 2023 is not complete so it is showing result for first two quarters only. In case of shooting, we can see a reduction in the year 2019 and spiked in the year 2020 to 2,011. Arrest showed a wave pattern, and it did not see a significant increase, but it did increase in comparison to the year 2006. The evictions also decreased in the year 2021 to 268. It is a radicle number since the eviction in 2017 was 22,522. So, if we investigate the datasets from the year 2018 to 2022, evictions and shooting showed a downward trend while arrest and hate crime showed upward trend.

**6. What does a borough wise comparative analysis of arrest, shoot, collision, eviction and hate crime look like? Which borough had the highest percentage of the variables? What can we infer from the visuals?**

A graph of different colored bars

Description automatically generated with medium confidence

The first look at this visual we will see that Brooklyn has high columns for all variables. Brooklyn contributes 34.5% of arrests whereas the highest contribution came from Manhattan which is 35.4%, a little bit higher than Brooklyn. Next if we look into number of car collisions we can see that Brooklyn contributes 31.72% to the total number of arrest. This is the highest and Queens comes second in this contribution. If we scrutinize the shooting data, we will see that Brooklyn contributes 29.8% which is the second highest. Bronx ranks top in this category with 32.2% contribution. Although, Bronx does not contribute much towards other variables except evictions. Hate crime is topped by Manhattan with 35.4% and Brooklyn contributes 34.5% in this category as well. And in evictions, as mentioned earlier, Bronx is the top contributor with 32.1% and Brooklyn contributed 29.2%, which is the second highest.

A graph of different colored bars

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A graph of different colored bars

Description automatically generated

A graph of different colored bars

Description automatically generated with medium confidence

A graph of different colored lines

Description automatically generated

Brooklyn has the highest contribution in collisions but has very high contribution in other categories as well. If more arrests are made in this borough, it can mean that more crime is also committed in this borough. Additionally, we can see that this borough also ranks very high in hate crimes as well. So, this means that the crime data in this borough might reveal more valuable information if we dig deeper. We will also see that this area has higher evictions in comparison. This might suggest that the people in this area are either poorer or have a tendency of not paying rent. We will need more detailed data to find out the reason and get a clearer picture. Shooting is also a concern in this borough as 29% of the shooting happened in this borough. Except for Bronx, other boroughs contribute almost half of what Brooklyn contributes. So, this indicated that safety and security might be a concern over here. As there is high number of shootings taking place over here, we could also explore the work of police in this borough.

Staten Island is doing excellent, it has the lowest contribution in all the variables taken into consideration for this report. Only 3.7% arrest, 4.1% in vehicle collisions, 2.8% in shooting, 3.74% in hate crime, 3.5% in eviction. These are excellent stats if compared to any other borough in New York city. But this raises a deeper concern. What could possibly going on in New York city where one borough is so much drastically better than others. The city could compare them and find out what is going right in Staten Islam. Also, we should keep in mind that statistics can be misleading sometimes, so we must make sure that we are comparing apples with apples.

**Conclusion:**

The aim of the project was to delve into the complex dynamics of urban life in the city of New York through the analysis of various data related to hate crime, arrests, collisions, evictions and shootings. Several visuals like interactive dashboard, maps and graphs help us analyze the data better, as a result we could get a comprehensive view of the crucial topics, deeper understanding and informed decision making. There was a contrasting pattern between the arrests and shooting over the years. Shooting gradually increased but we could see a fluctuation in the arrest data, especially during the year 2020 and 2021. If we further investigate the specific years and compare them with other years or types of analysis, we will be able to find the potential triggers. This investigation could help us prevent crimes and formulate preventive strategies. Analysis into the age and gender trends revealed that 25-44 had the highest arrest and males consistently outnumbered women in all age categories. This demographic information can help us prepare a more targeted strategy for each age group for crime reduction. Visualization of collisions data by vehicle type and major contributing factors highlight areas of concern. We can see that the driver’s inattention was the major cause of vehicle crashes for all categories. Brooklyn contributed highest towards collisions. Making efforts to increase driver attention could significantly reduce the number of collisions. And Brooklyn can be the location to start with since they have the highest collisions. A borough wise analysis of all the variables allows us to see how each variable performed in each borough. Brooklyn has made major contributions in all categories. This shows that there is room for work in Brooklyn, open doors for a deeper analysis which will reveal new information to make the urban dynamics better. Data on hate crimes showed a fluctuating pattern, with an odd spike in 2022. In 2021, evictions significantly decreased. These patterns might be a result of evolving social dynamics, legislative changes, or other outside forces that demand additional research. While this project attempted to offer insightful information, it is important to understand possible data constraints. A multi-dimensional approach beyond quantitative research is also necessary due to the complexity of societal dynamics. This data visualization project tried to demonstrate if and how New York City's urban processes are interrelated. This project provides a scope for educated dialogues, evidence-based policymaking, and deeper studies into the challenges and opportunities the city faces by revealing trends, patterns, and discrepancies within the data. Initiatives like these can help create a society that is better informed and more empowered as technology develops and data becomes more readily available.

**Additional research questions:**

1. What was the crime scenario like in areas where arrests had the highest density?
2. Could there be an economic factor that contributed to arrests in NY city?
3. If we normalize the dataset, then could we expect to see a different trend?
4. If hate crimes were plotted against evictions will there be any co relation?