
Gun Violence

Team 4

Final Report

**Snigdha Reddy Pulim, Patrick Wright, Tarek Mourad,
Vaishnavi Vadlamudi**

Year: spring 2023

Email: snigdhap@bu.edu, pmwright@bu.edu, tkmourad@bu.edu, vaishv@bu.edu

Page Numbers: 28

Introduction

Project Goal and Overview

Gun violence has become a major issue in many communities, and Boston's District 4 is no exception. Councilor Worrell has expressed concern about the increasing number of firearms possessed by young people in the district and the impact of gun violence on the community. To address these concerns, a data science project has been undertaken to understand the drivers of gun violence in District 4 and the rest of the city, as well as the comparative volume and geographic distribution of gun violence in the district and beyond.

Motivation

The goal of this project is to understand the drivers of gun violence in Boston's District 4 and compare it to the rest of the city. The project aims to determine the rate of gun violence in the district, identify patterns in terms of location and type of violence, and analyze existing programs meant to curb violence. The results of this analysis will inform policies that can be beneficial for improving the district.

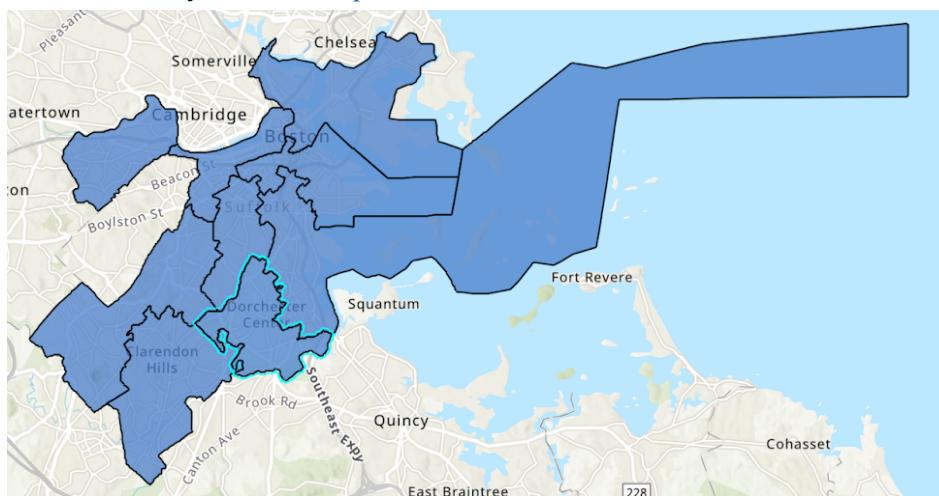
Background

Gun violence has been on the rise among the youth in Boston's District 4, which is a cause of concern for Councilor Worrell. In the project, we use various data sets and sources, such as police records of violence/firearm activity, discipline records from schools, and police presence to analyze the problem. We also gather additional data sets related to schools, environmental/community factors, and interactions to gain a better understanding of the issue.

Base Analysis

Data:

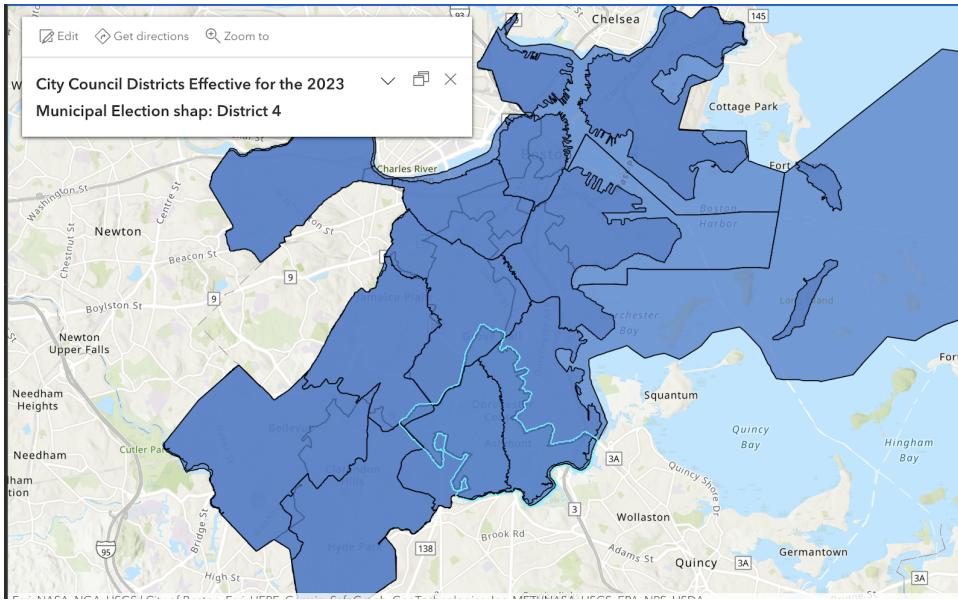
1. Districts for City Council [Shapefile](#)



2. Police Records of Violence/Firearm Activity [Shots Fired](#) and [Shootings - Datasets - Analyze Boston](#)

3. Police Districts:

- a. [Boston City Police Districts shapefile](#) (2 ii)
- b. Overlapped the above shape file with the city council shape file

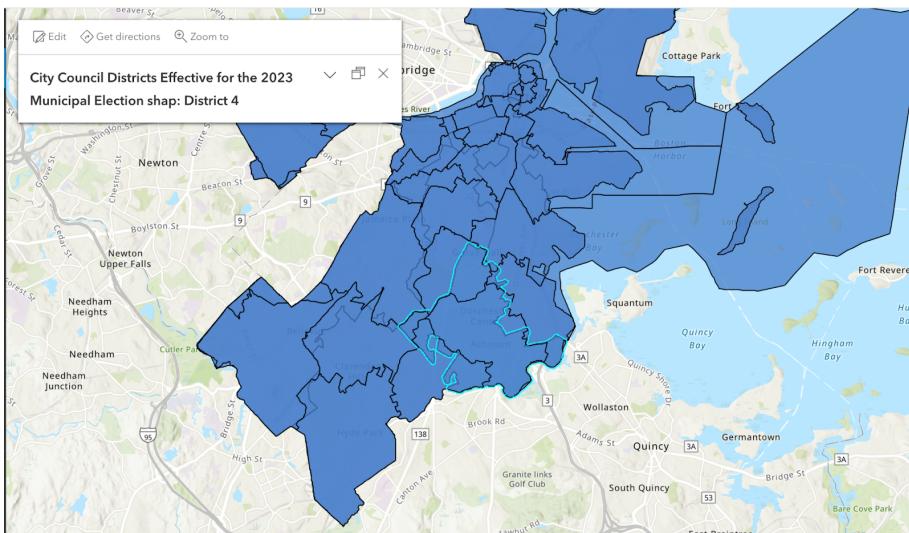


- c. We used the above Police Districts (B3, C11, E13, B2, E5) which overlap with city council district 4 to represent D4.

4. [Crime Incident Reports - Datasets - Analyze Boston](#)

5. Police Presence: [BPD Field Interrogation and Observation \(FIO\) - Datasets - Analyze Boston](#)

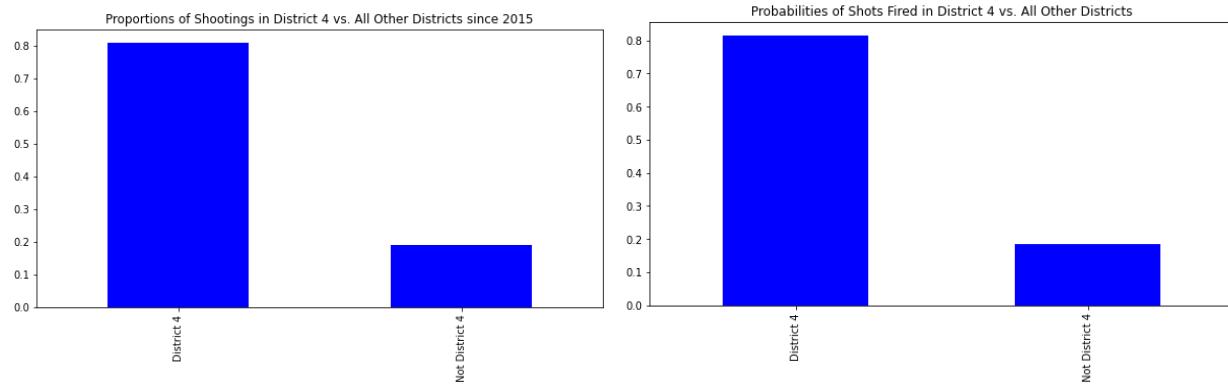
6. Overlapped the above with the city council shape file



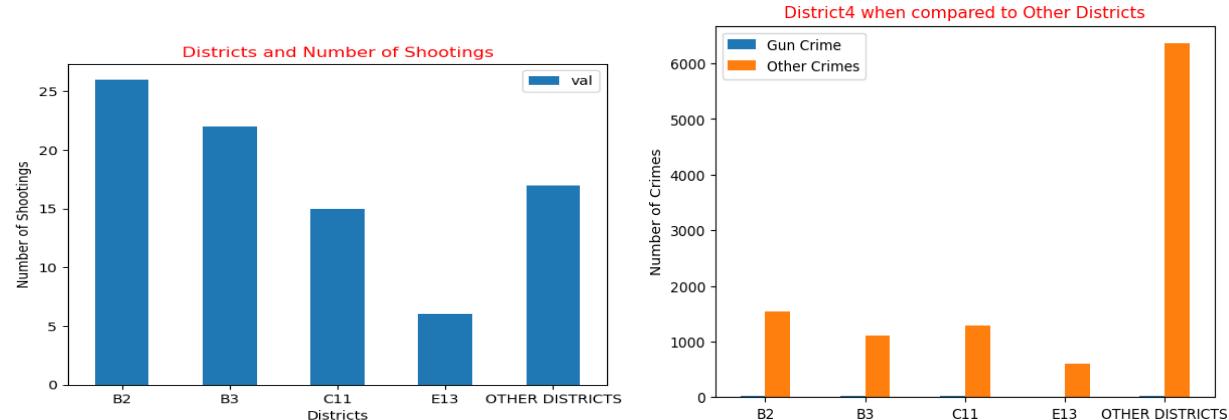
- a. The overlapping zip codes are
 - i. 02124 (most of the district)
 - ii. 02131
 - iii. 02126
 - iv. 02121
 - v. 02122

Key Questions

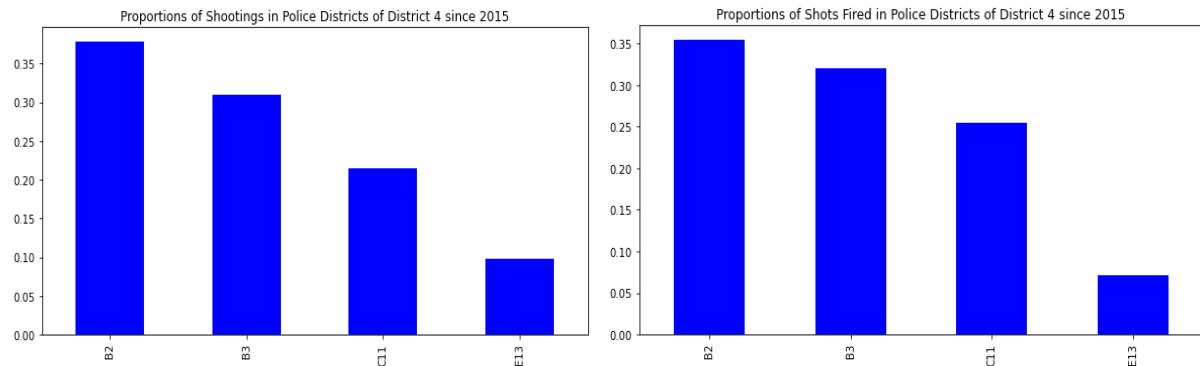
- What is the rate of gun violence in District 4? How does this compare to the rest of the city?



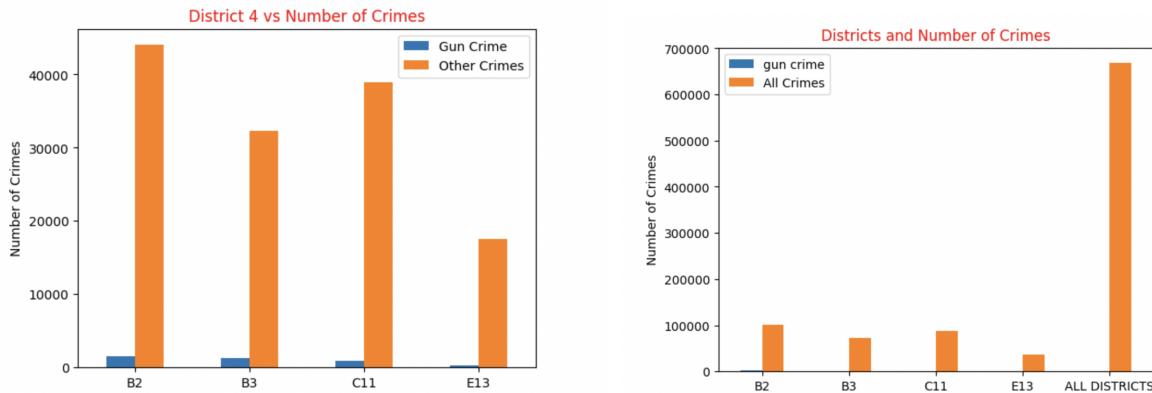
As shown in the bar plots “Propotions of Shootings in District 4 vs All other Districts since 2015” and “Probabilities of Shots Fired in District 4 vs All Other Districts”, it is evident that the rate of gun violence in District 4 is higher than the rest of the city.



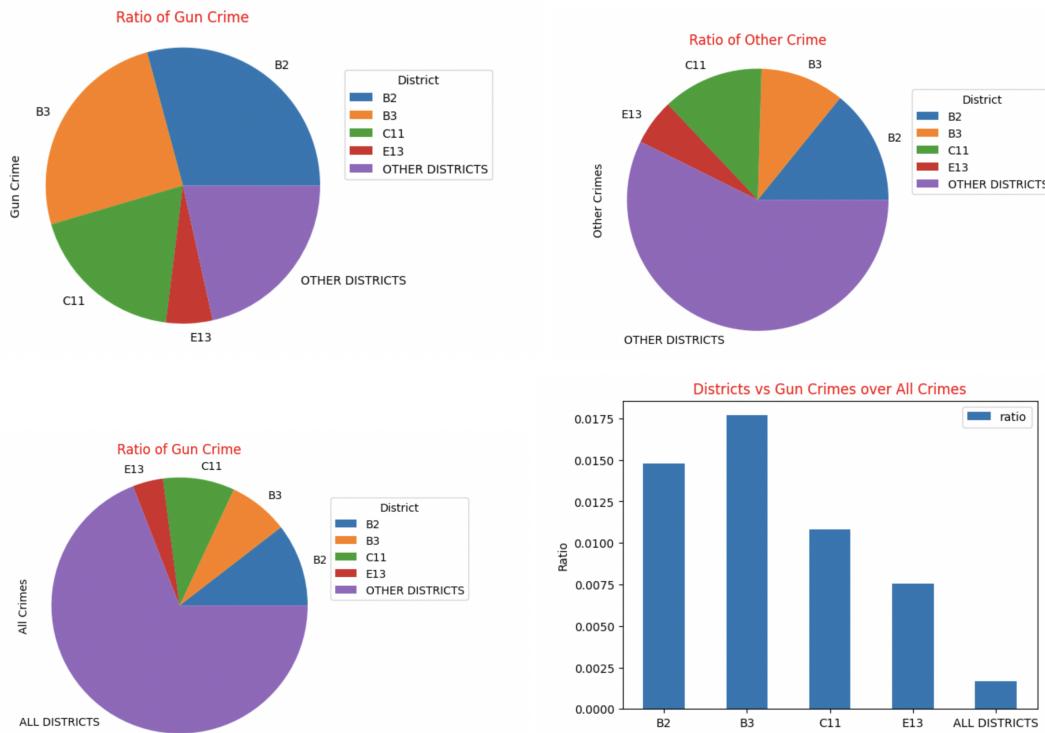
Here, with “Districts and Number of shootings” and “Distrit4 when compared to Other Districts” it can be seen that within District 4, Police District B2 (Roxbury) has the most number of crimes. This is followed by B3 (Mattapan) and C11 (Dorchester).



“Proportions of Shootings in Police Districts of District 4 since 2015” and “Proportions of Shots Fired in Police Districts of District 4 since 2015” also follow the same pattern as above.



The above graph illustrates District 4 Gun crimes when compared to the other crimes. The below graph illustrates District 4 crimes when compared to the rest of the city.



It appears that the numbers are decreasing in a consistent trend over the past few months.

However, we can calculate the year-on-year growth rate to identify any patterns.

Year-on-year growth rates:

$$\text{From 2015 to 2016: } ((302-196)/196)*100 = 54.08\%$$

$$\text{From 2016 to 2017: } ((352-302)/302)*100 = 16.56\%$$

$$\text{From 2017 to 2018: } ((318-352)/352)*100 = -9.66\%$$

$$\text{From 2018 to 2019: } ((647-318)/318)*100 = 103.77\%$$

$$\text{From 2019 to 2020: } ((889-647)/647)*100 = 37.39\%$$

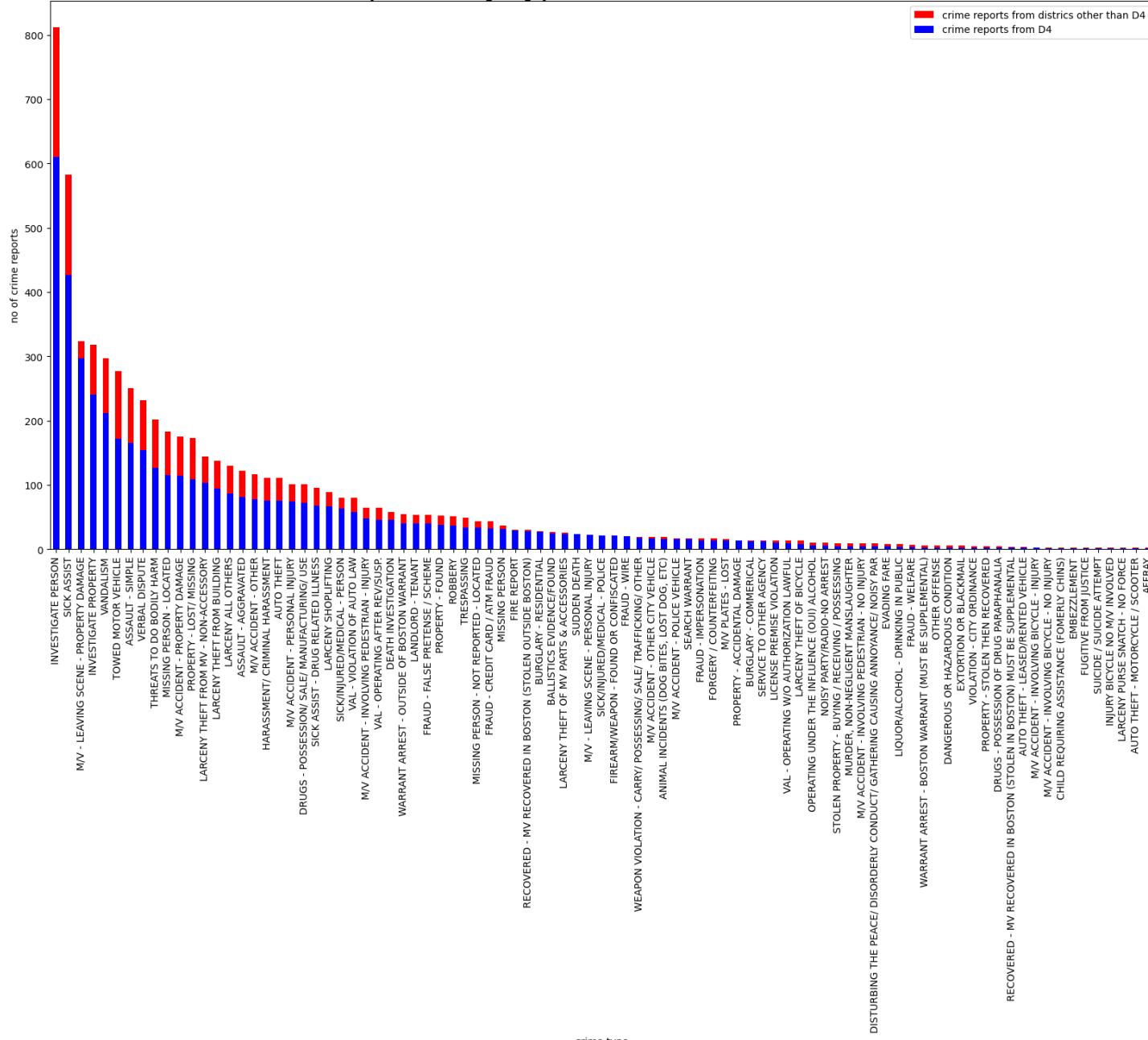
$$\text{From 2020 to 2021: } ((713-889)/889)*100 = -19.80\%$$

From 2021 to 2022: $((529-713)/713)*100 = -25.80\%$

Based on the above calculation, we can see that the year-on-year growth rate has been decreasing over the past few years. This suggests that the trend has been slowing down, and the numbers are not growing as quickly as they have in the past. However, as noted earlier, the latest observation for February 2023 is lower than the previous year's observation, suggesting a decrease in the trend.

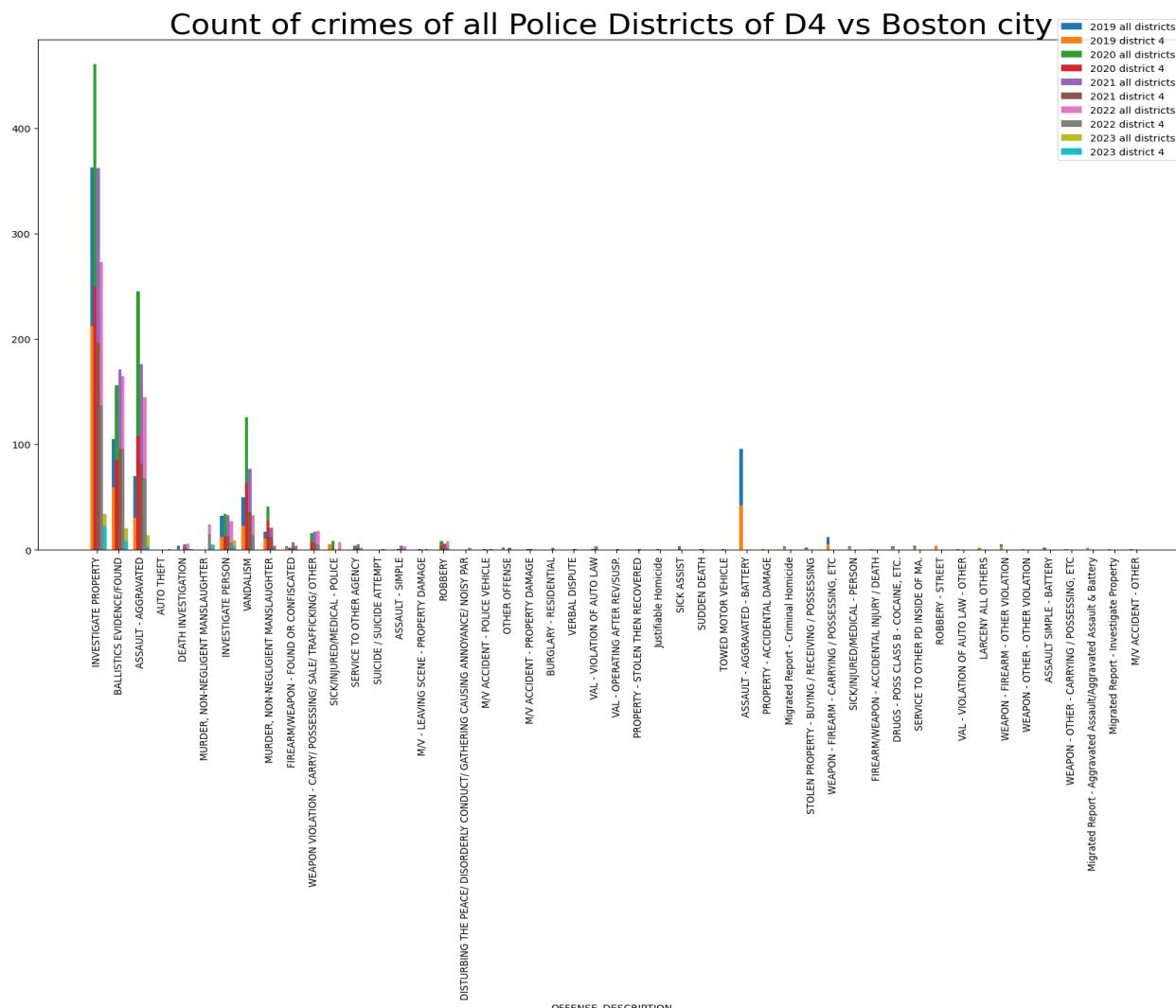
- Are there patterns of violence in terms of location in District 4? How does this compare to the rest of the city?

no of crime reports (by type of crime) outside D4 and in D4



This graph shows the “no of crime reports (by type of crime) outside D4 and in D4”. The pattern of crimes in terms of the popularity of the crime remained the same both in District 4 and the rest of Boston City. The most popular crimes are:

1. Investigate Person
2. Sict Assist
3. M/V - Leaving Scene - Property Damage
4. Investigate Property
5. Vandalism
6. Towed Motor Vehicle
7. Assault - Simple
8. Verbal Dispute
9. Threats to do Bodily Harm
10. Missing Person - Located

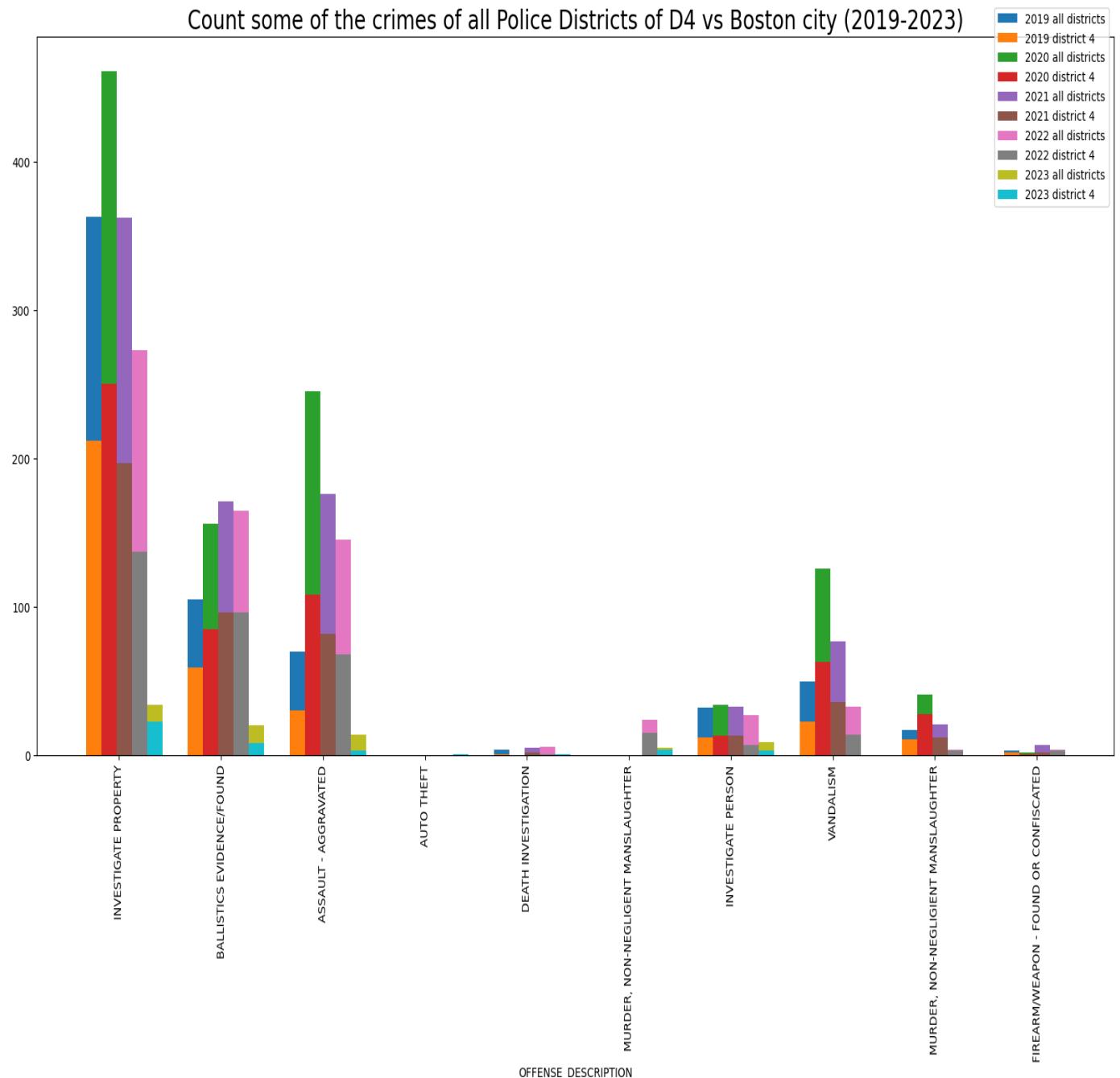


Based on the “Count of crimes of all Police Districts of D4 vs Boston city” graph, the most frequent crimes involving shooting are:

- Investigate Property

- Ballistics evidence / Found
- Assault - Aggravated
- Vandalism
- Investigate Person
- Murder, Non-Negligent Manslaughter

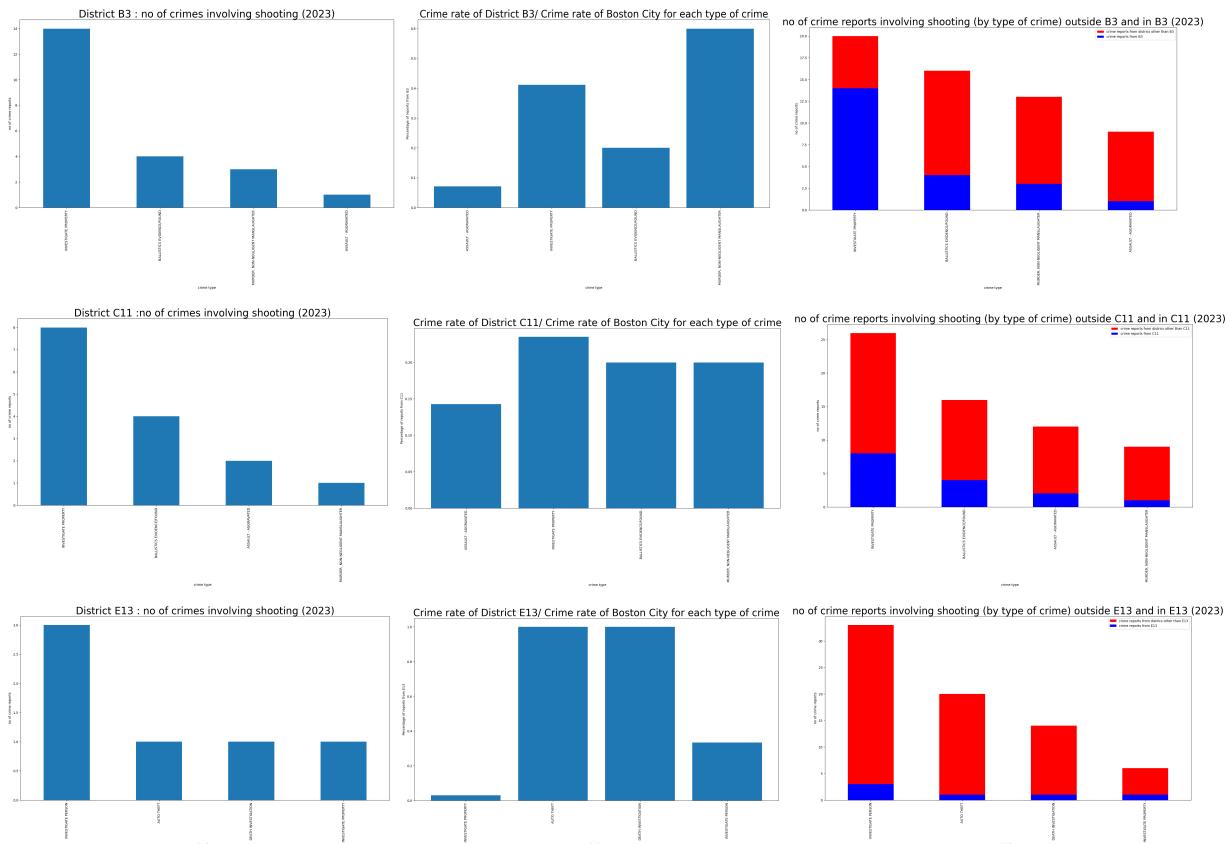
Count some of the crimes of all Police Districts of D4 vs Boston city (2019-2023)



From the “Count of some Crimes of all Police Districts of D4 vs Boston City (2019-2023)” data, it is evident that:

1. Both D4 and the rest of the city experienced a significant decrease in crime rate starting in 2020. However, for Ballistics evidence, this decrease began in 2021.

2. 2020 witnessed a significant increase in most types of crimes, leading to a peak for all crimes. This could be a possible response to the lockdown measures.



As seen in the above police districts' crime (involving shooting) data, E13's addition to the crime rate of D4 is very small.

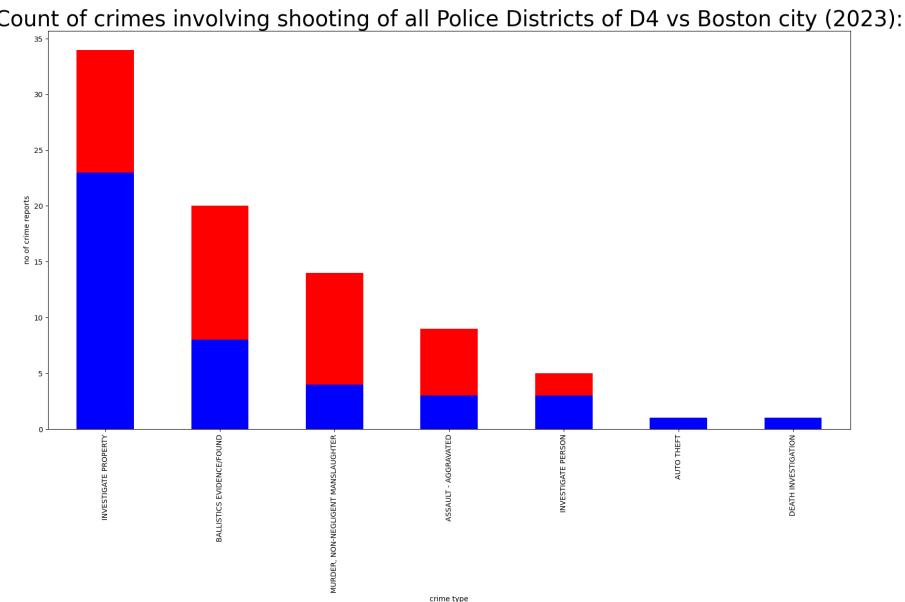
For 2023 data:

The top crimes in B3 and C11 are:

1. Investigate Property
2. Ballistics evidence / Found
3. Murder, Non-Negligent Manslaughter
4. Assault - Aggravated

For E13:

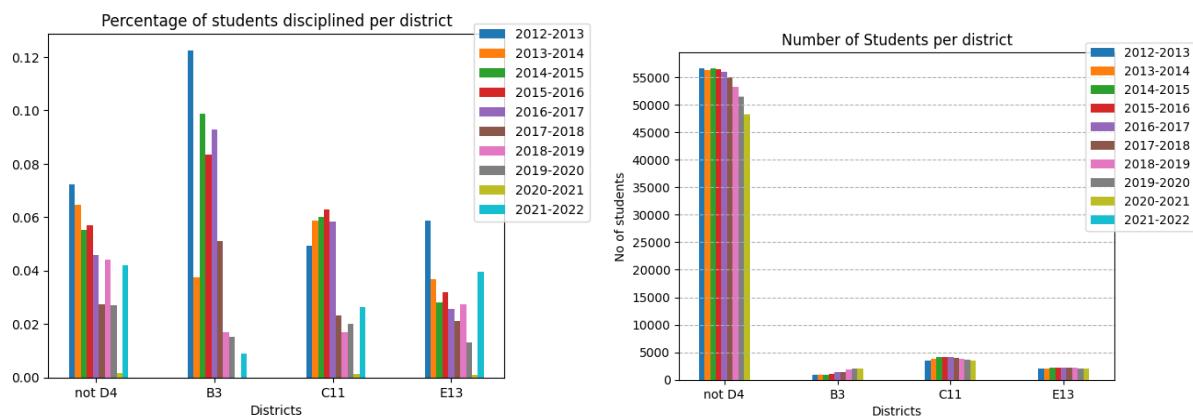
1. Investigate Person
2. Auto Theft
3. Death Investigation
4. Investigate Property

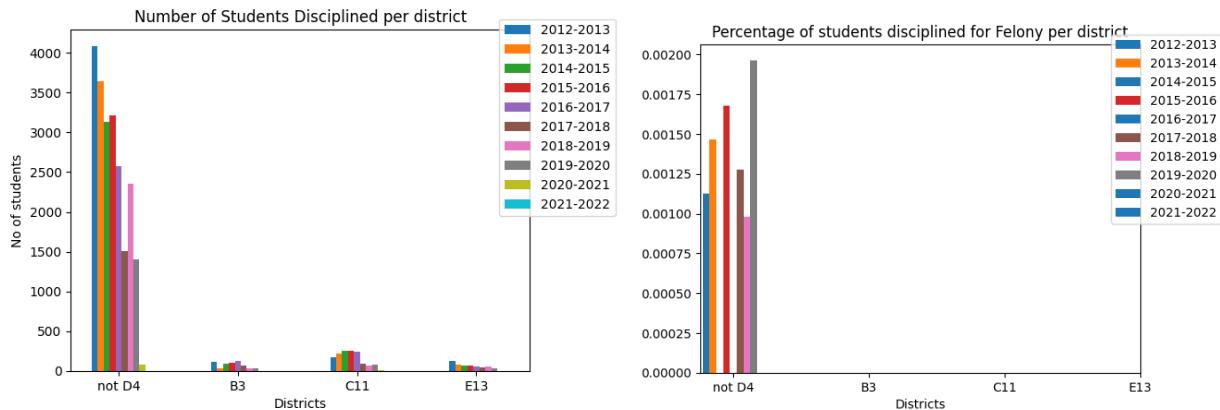


This graph shows the “Count of crimes involving shootings of all Police Districts of D4 vs Boston city (2023)”. The pattern of crimes in terms of the popularity of the crime remained the same both in District 4 and the rest of Boston City. The most popular crimes are

1. Investigate Property
 2. Ballistics evidence / Found,
 3. Murder, Non-Negligent Manslaughter
 4. Assault - Aggravated,
 5. Investigate Person
 6. Auto Theft
 7. Death Investigation

3. How do the patterns of student discipline in District 4 compare to those in Boston City as a whole?





Based on the "Percentage of Students Disciplined" metric:

- B3 had the highest percentage of students being disciplined until the 2017-2018 academic year.
- However, the percentage has decreased significantly since 2018-2019 and has continued to consistently decrease.
- C11 has a higher percentage of students being disciplined.
- Nonetheless, C11 has maintained a lower percentage since 2017-2018.

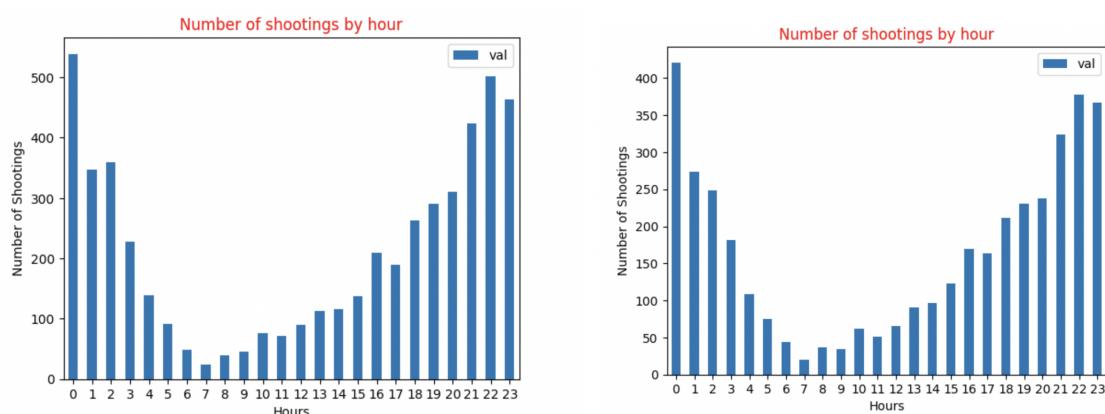
Based on the "Number of Students per District" metric:

- Non-district4 schools have shown a slight decrease in the number of students each year since 2014-2015.
- B3 has experienced an increase or maintained a similar student count over the years.
- E13 has maintained its student count while C11 has followed the pattern of non-district4 schools.

From the "Number of Students Disciplined per District" metric, it can be observed that there has been a decrease in the number of students disciplined across District 4 since the 2016-2017 academic year.

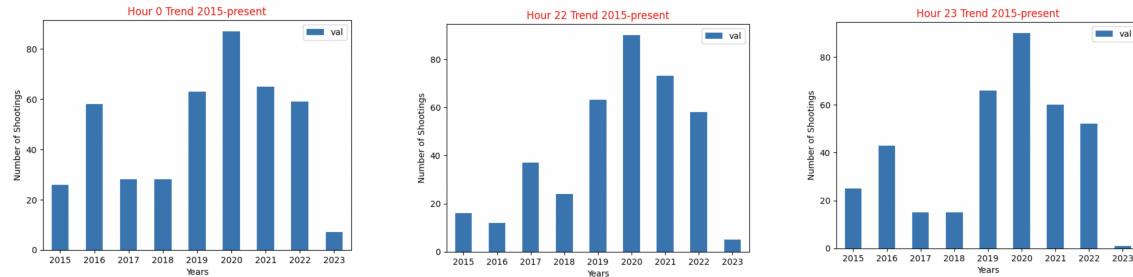
Based on the "Percentage of Students Disciplined for Felony per District" metric, there have been no students disciplined for felony convictions or complaints in District 4 between 2012 and 2022.

4. Within District 4, at which hour do most shootings occur within our dataset which contains information about all shootings reported from 2015 to about the present day?



The First graph is for the number of shootings per hour in all the districts throughout the last 7 years. The Second graph illustrates the number of shootings per hour in District 4 throughout the past 7 years.

The number of shootings ranges from 26 to 87. The mean is 49.25. The standard deviation is 23.38. We can say that the data is moderately spread out with a slightly skewed distribution towards the lower end.



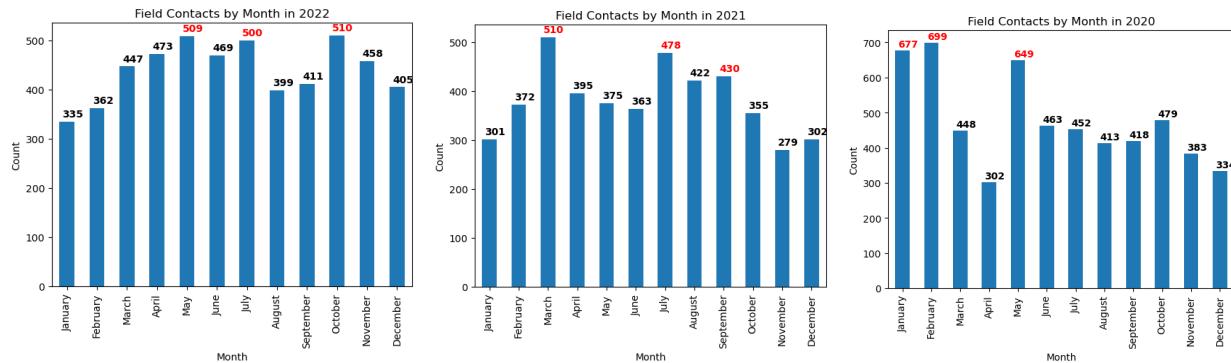
The number of shootings in Hour 22 ranges from 12 to 90. The mean is 47.125, and the standard deviation is 27.362. The data is moderately spread out, with a slightly skewed distribution towards the higher end.

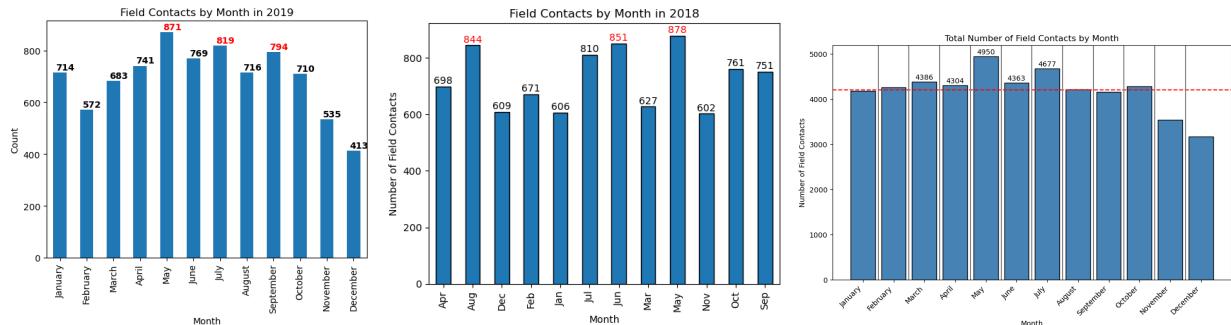
It's worth noting that while the mean for Hour 22 is close to the mean for Hour 0, the range and standard deviation are slightly larger. This suggests that there is more variability in the number of shootings for Hour 22, and the distribution is slightly more spread out than it was for Hour 0.

The number of shootings in Hour 23 ranges from 15 to 90. The mean is 47.5, and the standard deviation is 25.267. The data is moderately spread out, with a slightly skewed distribution towards the higher end.

Compared to Hour 0 and Hour 22, the range of Hour 23 is similar to Hour 22, but the standard deviation is closer to Hour 0. This suggests that there is a moderate amount of variability in the number of shootings for Hour 23, and the distribution is slightly more spread out than it was for Hour 0.

5. What are the patterns and trends in the distribution of field contact counts over time (both annually and monthly)?



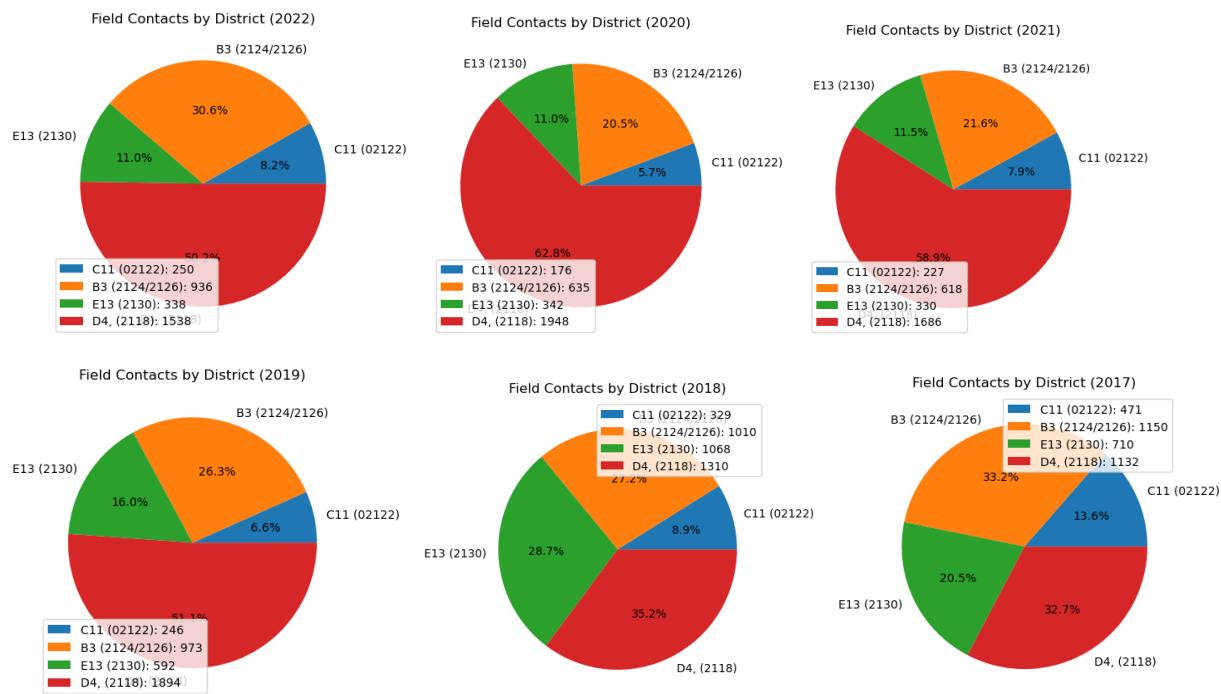


The first 5 charts above represent the count of field contacts for each individual year from 2018 until 2022. There are some irregularities in the data that represent the year in which covid appeared. In general, the months with the highest counts of field contacts are in the spring/summertime.

The last chart represents a bar chart of the sum of the counts of field contacts for each month from 2016 until 2022. It has the average count for all months as well as the labeled count of the top 5 months with the most counts of field contacts.

The top 5 counts for all months revolve around the summer.

The average count for all months is around 4200.



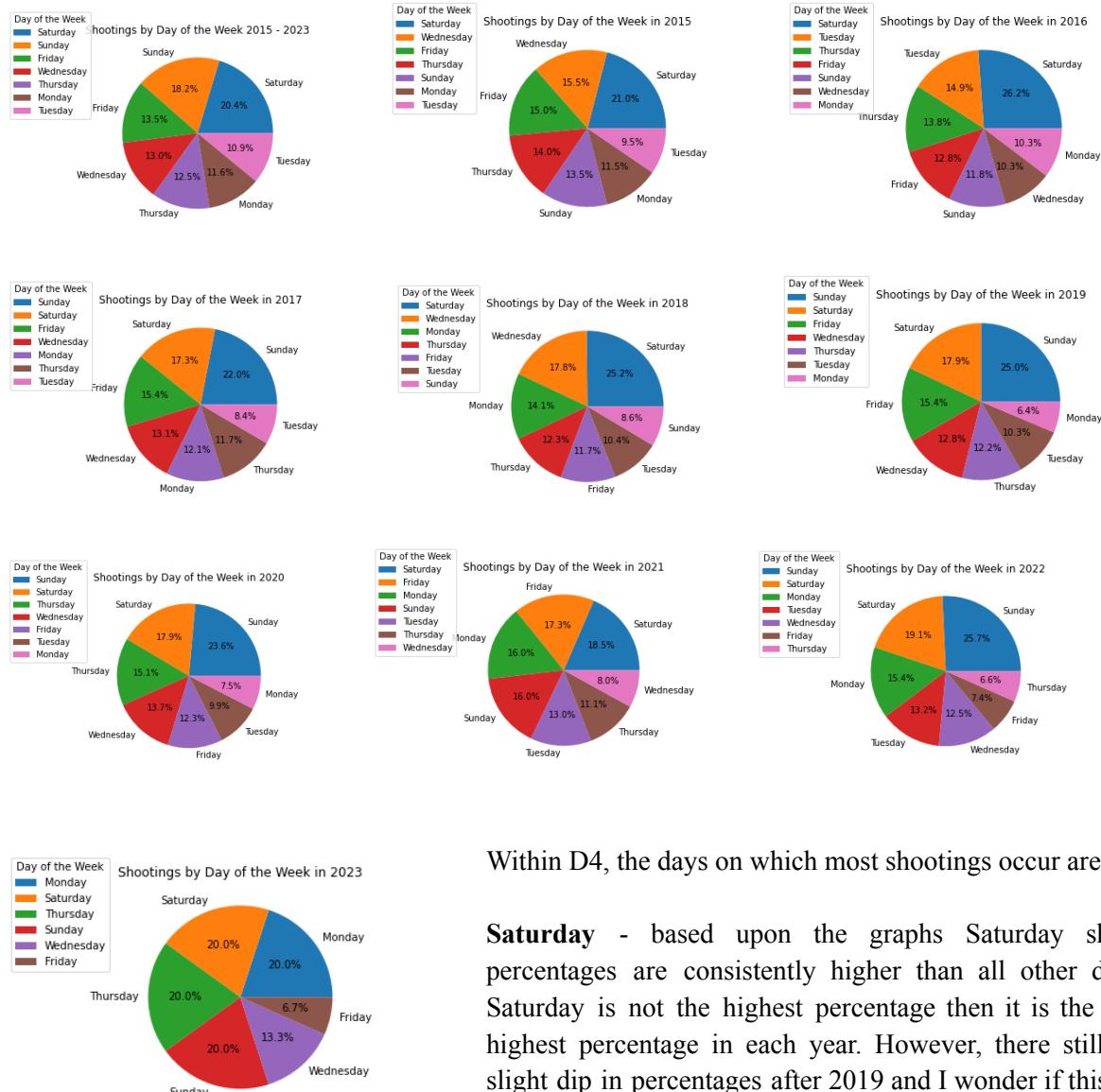
The 6 pie charts above represent the percentage of field contacts within districts E13, B3, C11, and D4.

District D4 has a much higher number of field contacts compared to other districts

District D4's field contacts have increased by almost 100% since 2017.

District E13 field contacts have decreased from around 20% to 11%.

6. Within District 4, on which days do most shootings occur within our dataset which contains information about all shootings reported from 2015 to about the present day?



Within D4, the days on which most shootings occur are:

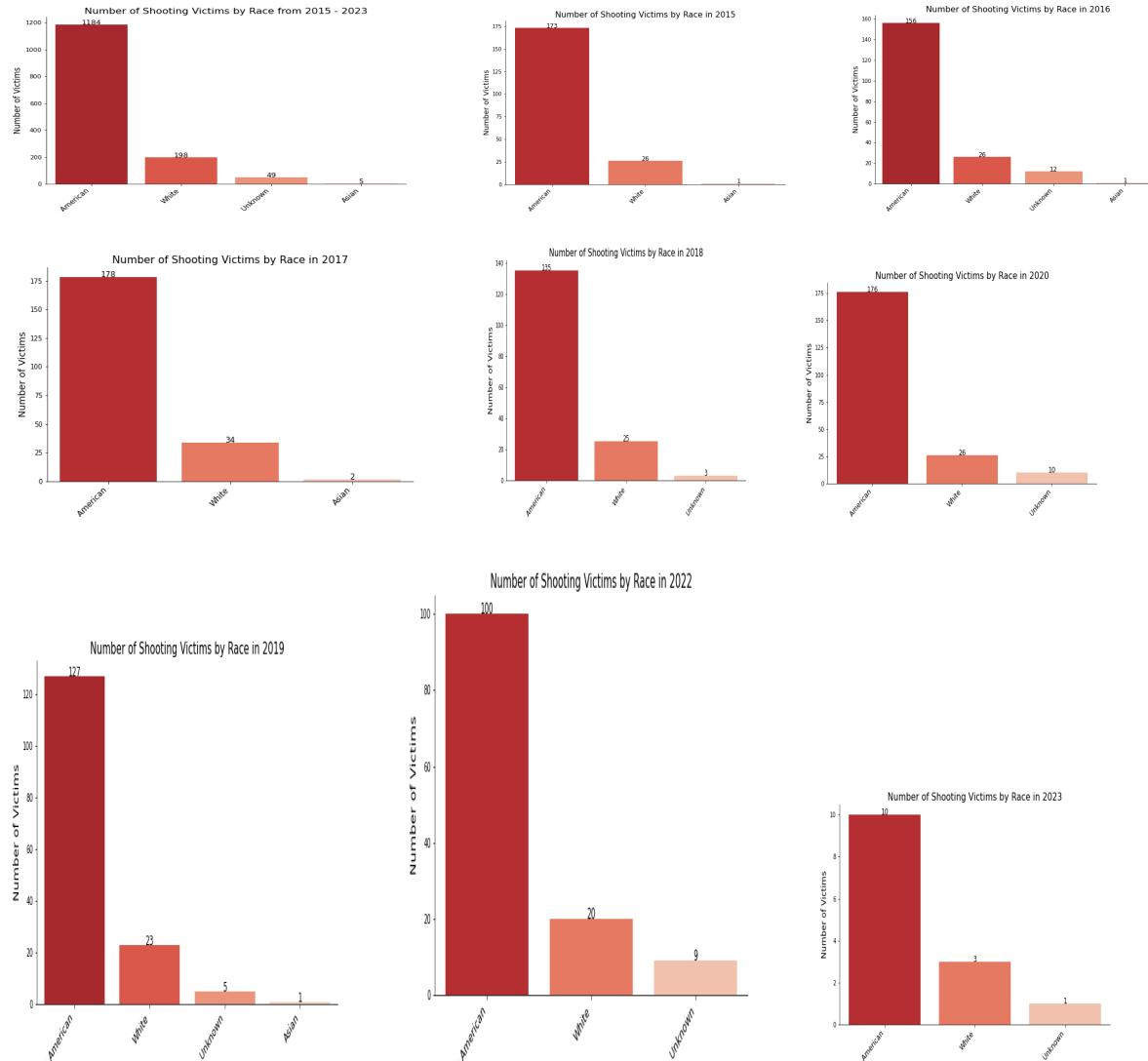
Saturday - based upon the graphs Saturday shooting percentages are consistently higher than all other days, if Saturday is not the highest percentage then it is the second highest percentage in each year. However, there still was a slight dip in percentages after 2019 and I wonder if this is due to more police presence on Saturdays in D4.

Sunday - Sunday was the second highest percentage within our dataset. In 2017, Sunday had the highest percentage out of all the days, which was unique to the first 4 years of the dataset because, in every year within that 4 year span, Saturday had the highest percentage over Sunday by a decent margin. But after 2018, there was a spike in shootings that occurred on Sundays within D4 excluding one of the years. Excluding 2023 due to limited data, Sunday had the highest percentage of shootings in 3 out of 4 years from 2019 - 2022. I wonder if this change has to do with the number of patrols or police presence within D4 or if there is another reason for this change in the data.

Friday - It seems like the rest of the days of the week all jumped around a little bit from year to year, but Friday was the third highest percentage within this span right next to Wednesday and Thursday. Monday

and Tuesday not falling too far behind them either. Most of the shootings happen from Friday to Sunday: a little over 50% of shootings have happened from Friday to Sunday in our 9 years (2015 - 2023) of data covering shootings within D4. It might be best to increase the patrols within D4 during these days of the week.

7. Also looking at all the shootings that have taken place just within District 4, what is the number of shootings victims by race per year from 2015 - 2023?

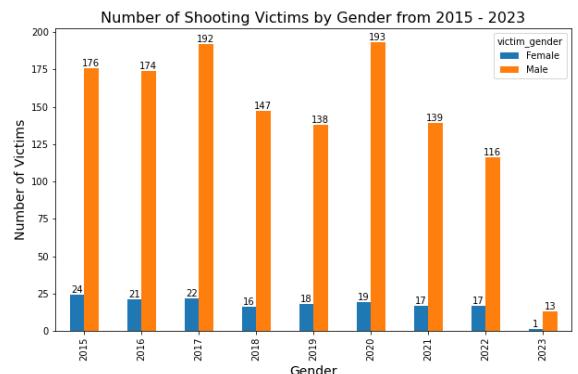


Here we are using the same modified dataset as before that focuses on shootings that have occurred within D4 from 2015 - 2023. I am focusing on data that has more to do with the victims themselves

This data above shows the race of the victims within this time period, Black/African American victims are by far the highest compared to any other race.

- Due to D4 being a predominantly black area this data makes sense, but hopefully, there is a solution we can find to drastically decrease the number of black shootings within D4 as well as decrease the number of shootings within D4 as a whole.

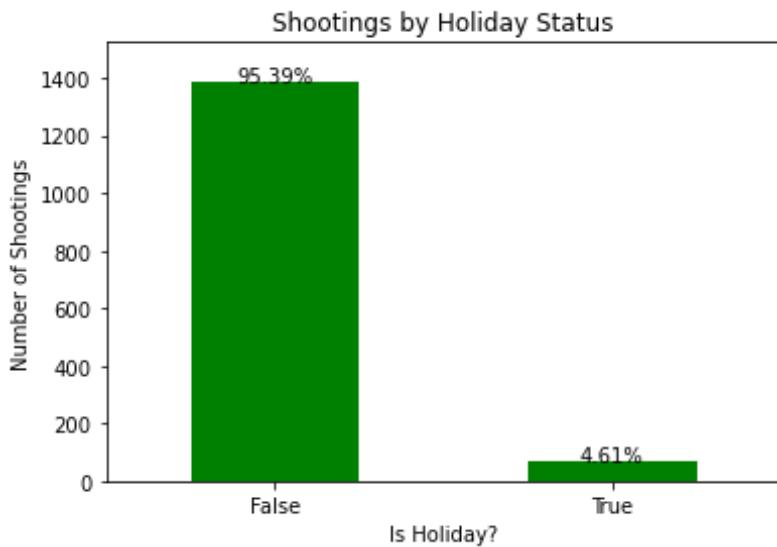
- The number of black victims is nearly 6x that of the next highest race/ethnicity of victims within D4.
8. What is the number of shooting victims by gender per year from 2015 - 2023?



This again reflects the victim data for D4 shootings.

- This data as well comes as no surprise due to the fact that most of the shootings that happen within D4 are gang-related and most gang members, so I've been told, are male rather than female.
- The number of male victims is over 8x that of the number of female victims within D4.
- We believe to limit and disrupt gang activity within D4 would hopefully lead to fewer shootings within the district.

9. What percentage of crimes are occurring on holidays?



A significantly small percentage (4.61%) of crimes are occurring on holidays.

To be noted, the percentage of holidays in the time period considered is 3.26%.

Extension Analysis

Reasons for Choosing Extension Idea

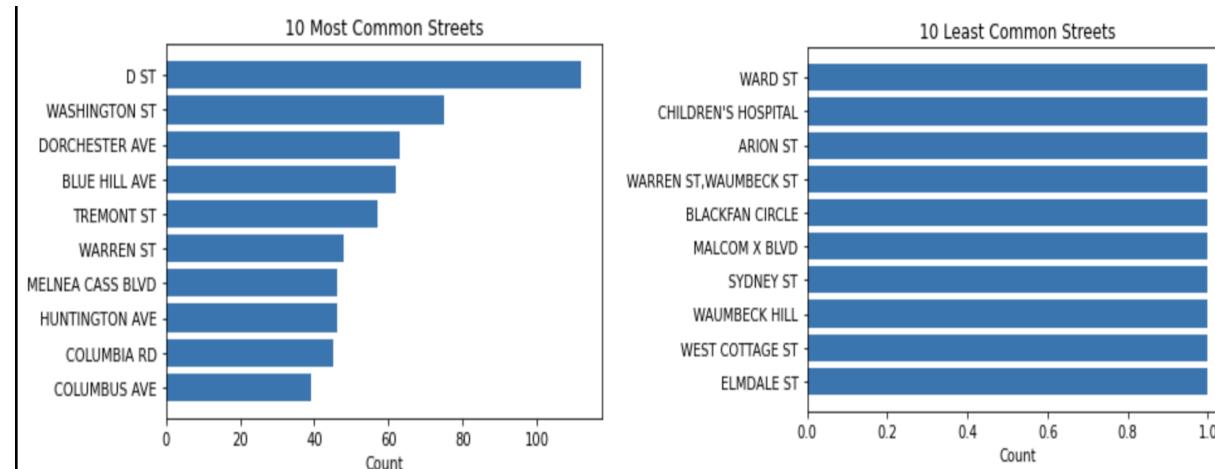
To further enhance our project's insights and contribute to urban planning and community development recommendations, we aim to expand the scope of our project by incorporating Environmental/Community Factors. Specifically, we will focus on factors such as green space, tree canopy, community programming, interactions, and pedestrian/mobility data.

Overview of Team's Extension Idea

The inclusion of Environmental/Community Factors is essential as it acknowledges the significant role that the natural environment and community programming play in shaping the quality of life for residents. Access to green spaces promotes physical activity, mental health, and social cohesion within a community. Tree canopy data is essential for understanding the distribution and health of trees in urban areas, which can have numerous environmental and health benefits. Community programs can provide essential resources and services for residents, particularly in underserved areas. Analyzing interactions between different factors, such as routes from schools to public transit stops, can shed light on mobility patterns and potential barriers to accessing public transportation. Pedestrian and mobility data can provide insights into the safety and accessibility of walking and mobility options, which are essential for promoting active transportation and reducing reliance on cars. Our team plans to utilize data sets and sources such as green space and tree canopy data, community programming, pedestrian/mobility data, and more, to provide a more comprehensive analysis of the community's environmental health and well-being, as well as its impact on mobility and access to public transit. Through data analysis, visualizations, community engagement, and policy recommendations, the project can contribute to informed decision-making for urban planning and community development, promoting sustainability, equity, and well-being in the community.

Key Questions and Answers

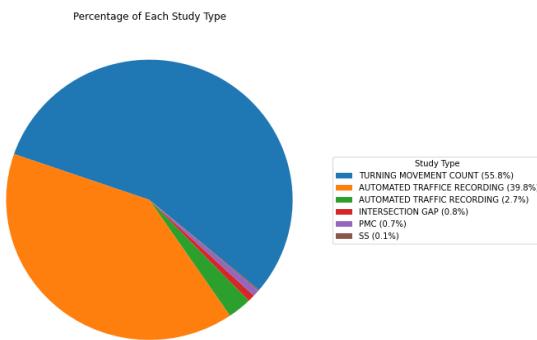
1. Is there any correlation between the popularity of a street and the number of crimes on the street?



We had slight trouble with the data

We cut the data to focus solely on D4, however, after some research we found out that some of the streets were not found in D4.

Keeping those ideas in mind, the graphs above show the top 10 most and least common streets that we found in the Traffic Dataset.



The pie chart above shows the percentage of each study type in traffic data. The chart displays the most common study types and their corresponding percentage of the total count. The 'Turning Movement Count' study type is the most common type of study found within the dataset. These TMC counts are used for intersection analysis: traffic operation analysis, intersection design, and transportation planning applications. From my understanding, these types in the traffic data refer to normal roads and intersections, which makes me wonder if shootings within District 4 are more likely to occur on streets where there are fewer bus stops and T stops due to

there likely being fewer people. Again, I would need to gather more data in order to make a more accurate conclusion, but this is my thought process for the data at hand at the moment.

```

Study types for COLUMBUS AVE:
TURNING MOVEMENT COUNT      37
SS                            2
Name: Study Type, dtype: int64
Study types for COLUMBIA RD:
TURNING MOVEMENT COUNT      45
Name: Study Type, dtype: int64
Study types for HUNTINGTON AVE:
TURNING MOVEMENT COUNT      44
PMC                           2
Name: Study Type, dtype: int64
Study types for MELNEA CASS BLVD:
TURNING MOVEMENT COUNT      46
Name: Study Type, dtype: int64
Study types for WARREN ST:
TURNING MOVEMENT COUNT      48
Name: Study Type, dtype: int64
Study types for TREMONT ST:
TURNING MOVEMENT COUNT      57
Name: Study Type, dtype: int64
Study types for BLUE HILL AVE:
TURNING MOVEMENT COUNT      60
INTERSECTION GAP             2
Name: Study Type, dtype: int64
Study types for DORCHESTER AVE:
TURNING MOVEMENT COUNT      61
INTERSECTION GAP             2
Name: Study Type, dtype: int64
Study types for WASHINGTON ST:
TURNING MOVEMENT COUNT      75
Name: Study Type, dtype: int64
Study types for D ST:
TURNING MOVEMENT COUNT      106
INTERSECTION GAP             4
PMC                           2
Name: Study Type, dtype: int64

```

Here is some more data that I collected to further show that the 10 most common streets found within the data are of type 'TMC'.

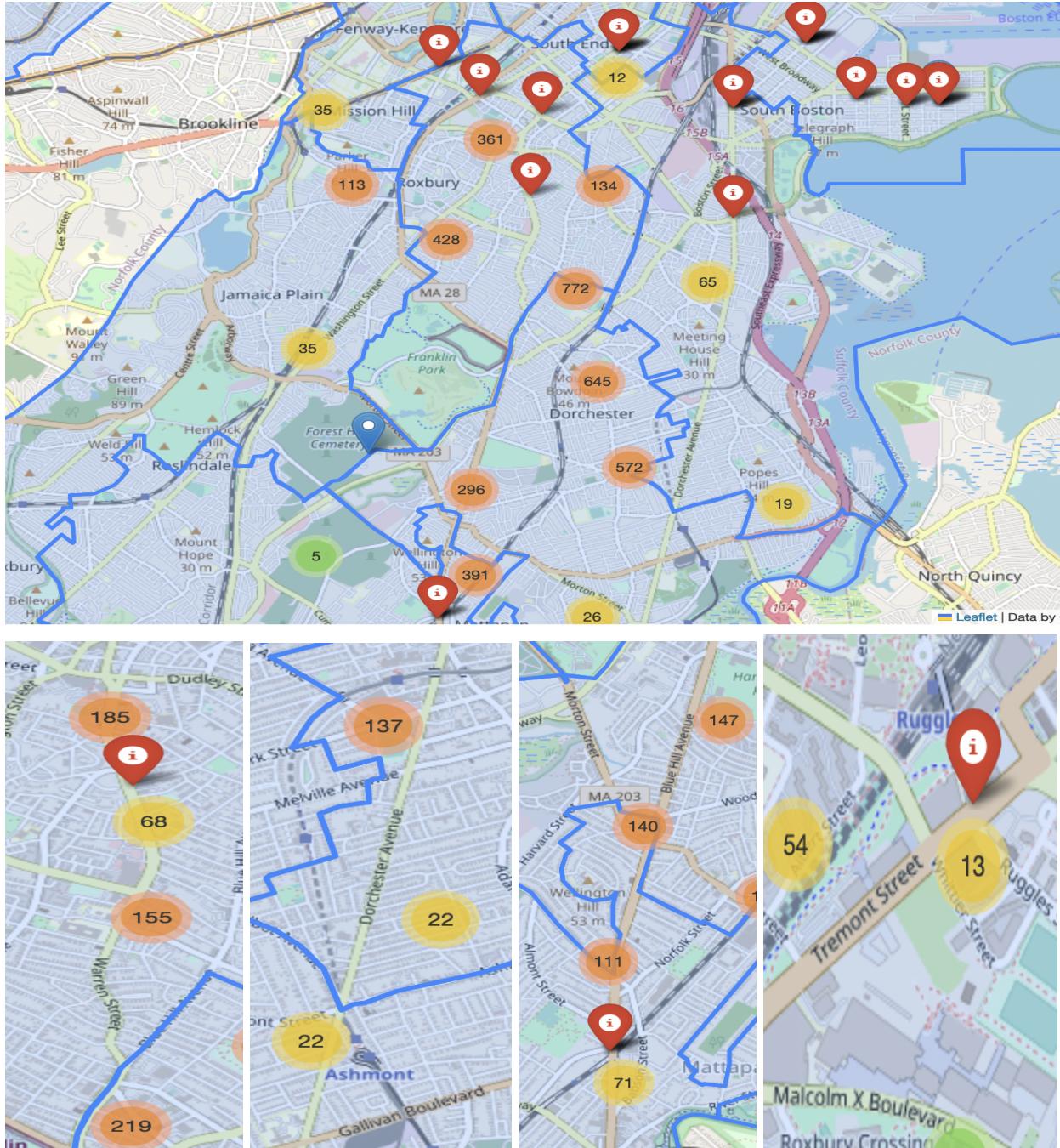
With this information, I also looked at some of the information that Vaishnavi found by using her map. I compared what I found vs. what she found in order to reach a hypothetical solution. From here, I can take this hypothetical solution and continue to look into the information that could make the conclusion more accurate.

As mentioned above, I believe that this data shows us that shootings seem more likely to occur in places with fewer things going on. (All aspects: people, transportation, etc.



These findings are significant and have informed policies to increase police presence in these areas from 10 pm to 12 am, where gun crime has been observed to be the highest. This information is presented in the Hourly Analysis section of the Deliverable 2 Report. The increased police presence is part of a broader effort to reduce the incidence of gun violence in the affected neighborhoods.

This was the basis for the next part of the data visualization.



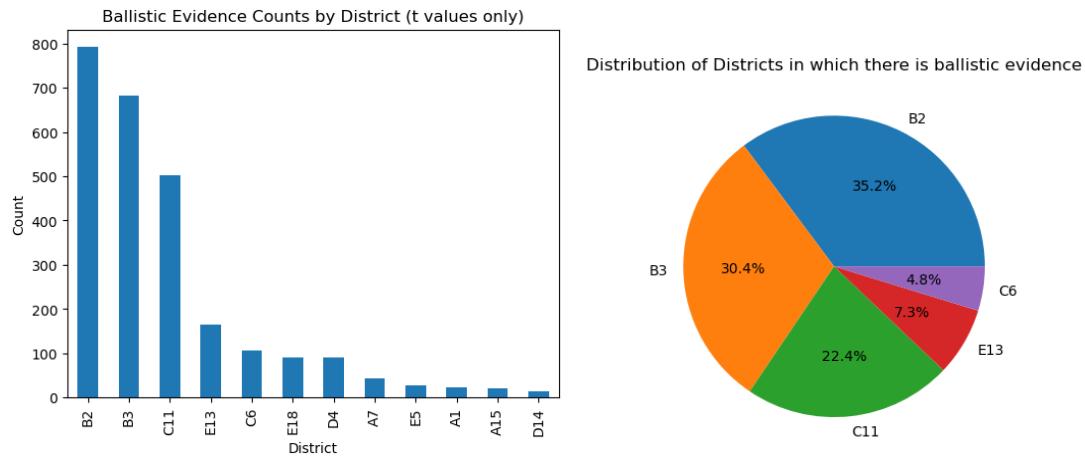
Furthermore, the analysis of the gun crime reports and the identification of the high-risk areas have led to additional visualizations of the data. For instance, we have created heat maps to provide a more comprehensive view of the prevalence of gun violence across the city. By visualizing the hotspots of gun violence, we can identify areas where intervention is most urgently needed.

In addition to the heat maps, we have also looked at the relationship between public transportation and gun violence.

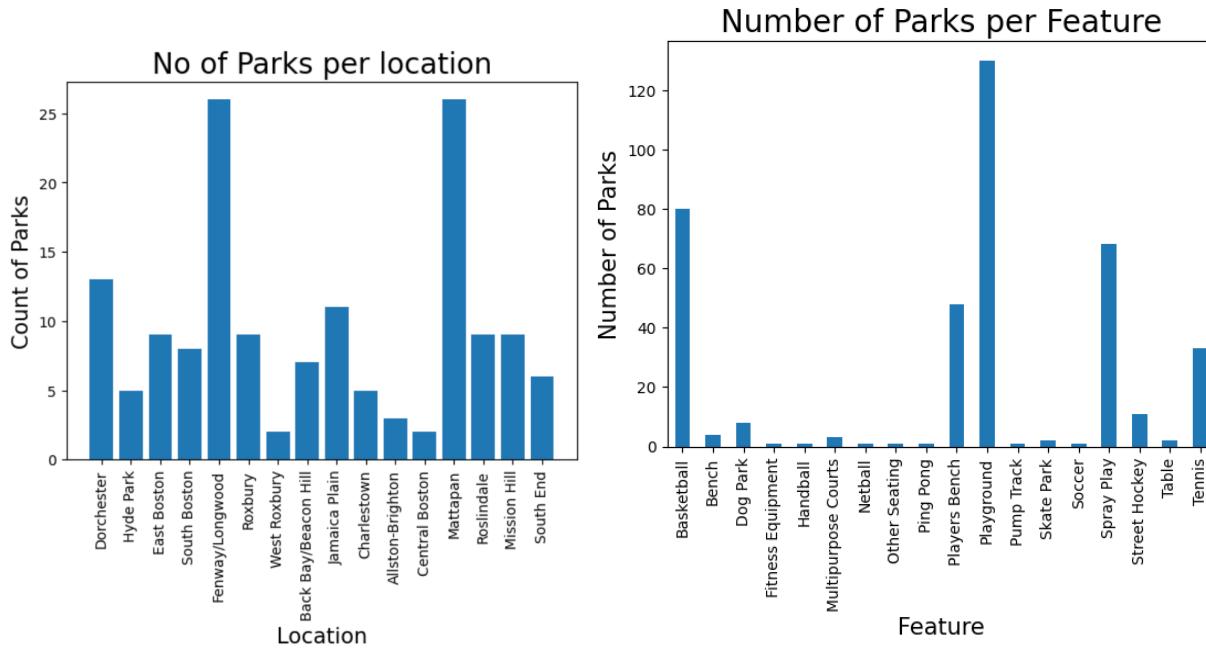
Our analysis has shown that the streets with the highest number of public transportation stops are also the areas with the highest crime incidents. This finding is significant because it suggests that public transportation may be a contributing factor to gun violence in these areas.

Overall, the data visualizations have provided valuable insights into the prevalence of gun violence in Boston and have informed policies to reduce the incidence of crime in the affected neighborhoods. The next step in our analysis is to look at other factors that may be contributing to gun violence in these areas, such as socioeconomic factors and access to mental health services.

2. What are the Ballistic evidence trends?



3. How does the presence of parks affect the crime rate?



Top 5 neighbourhoods by count of Parks in the neighbourhood:

- Fenway/Longwood
- Mattapan
- Dorchester
- Jamaica Plain
- Roslindale; Mission Hill

Top 4 neighbourhoods by count of Parks in the neighbourhood in District 4:

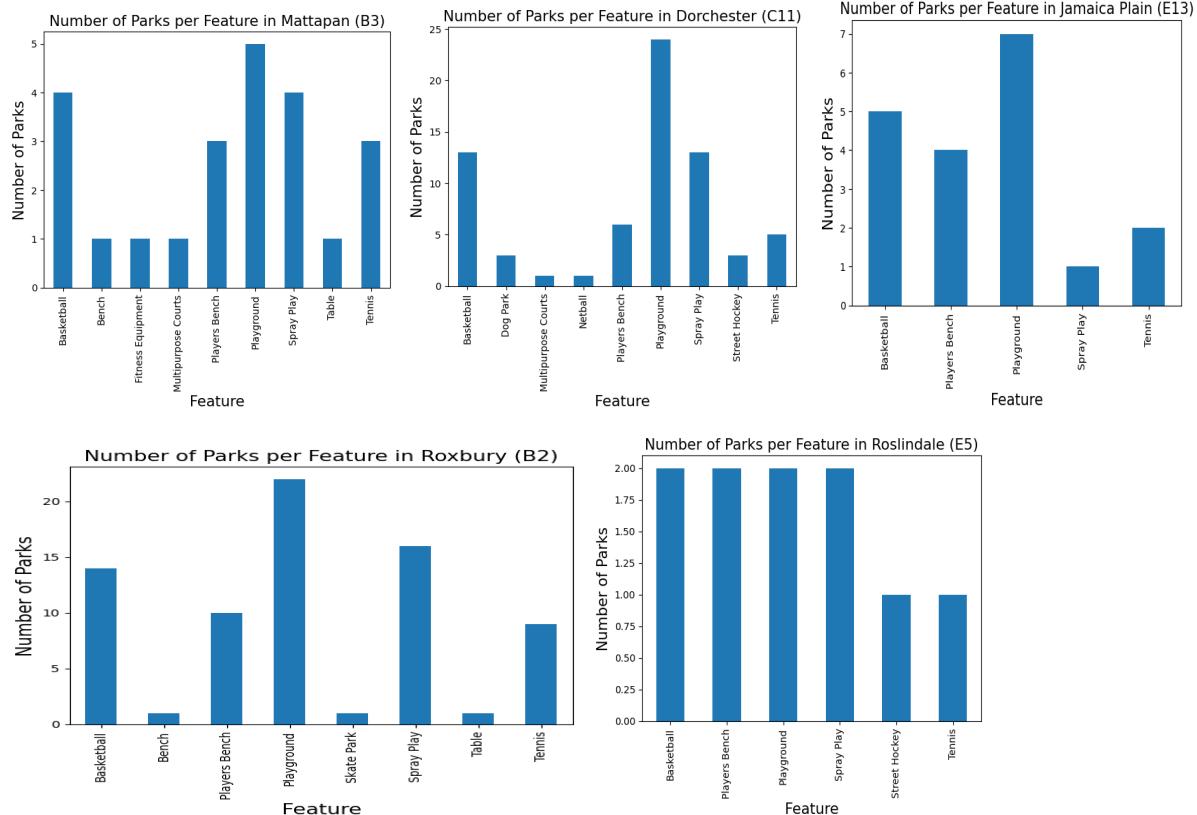
- Mattapan
- Dorchester
- Jamaica Plain
- Roslindale

Despite the high crime rate in District 4, it is interesting to note that the number of parks is also high, particularly in the top 4 neighborhoods listed above. This finding challenges the assumption that more parks necessarily lead to a reduction in crime.

From the plot, we can see that the top 5 features with the most parks are Playground, Basketball, Spray Play, Players Bench, and Tennis. This information can be useful for city planners and park management to understand the popular features among park visitors and make informed decisions on park upgrades and improvements.

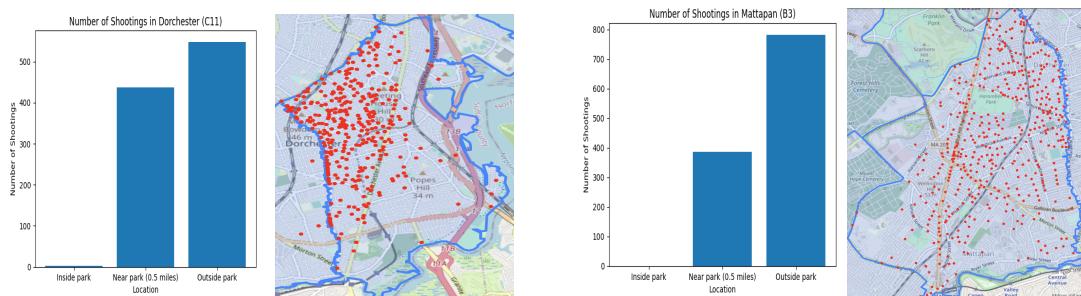
Additionally, we can see that there are some features with very few parks, such as Boathouse and Ice Skating, which may indicate opportunities for further development in those areas.

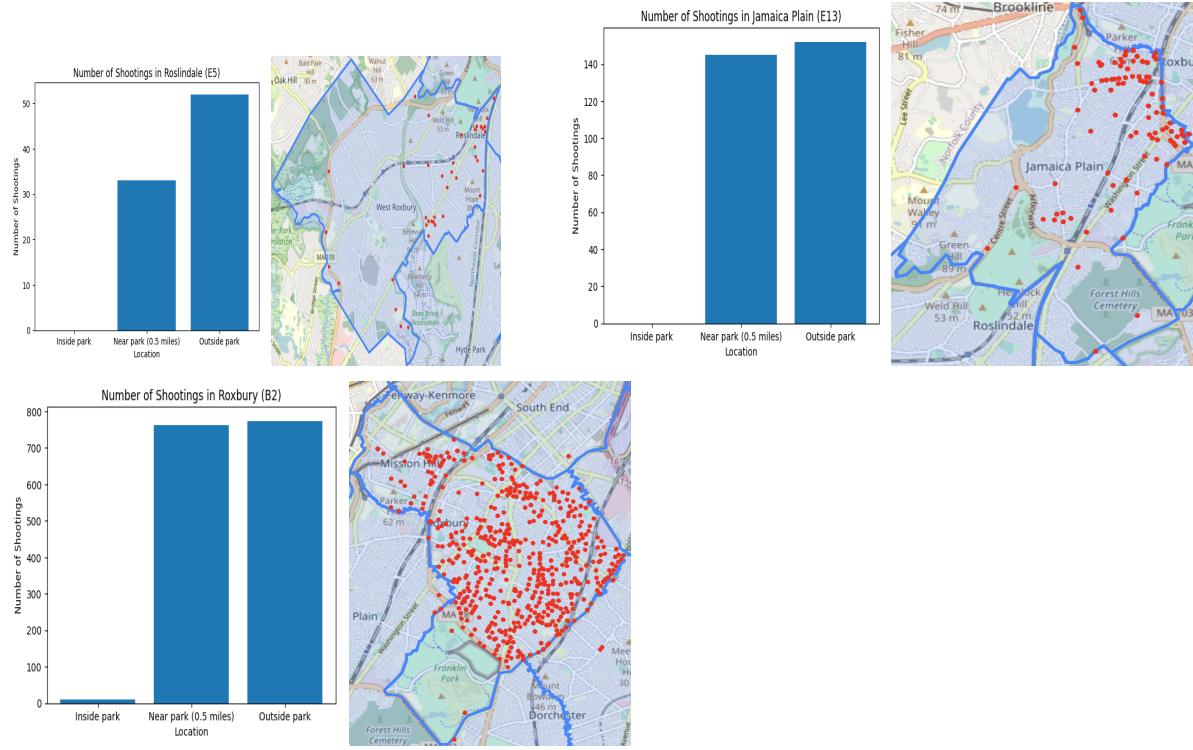
Overall, this analysis provides insights into the distribution of park features and can aid in decision-making for park management and planning.



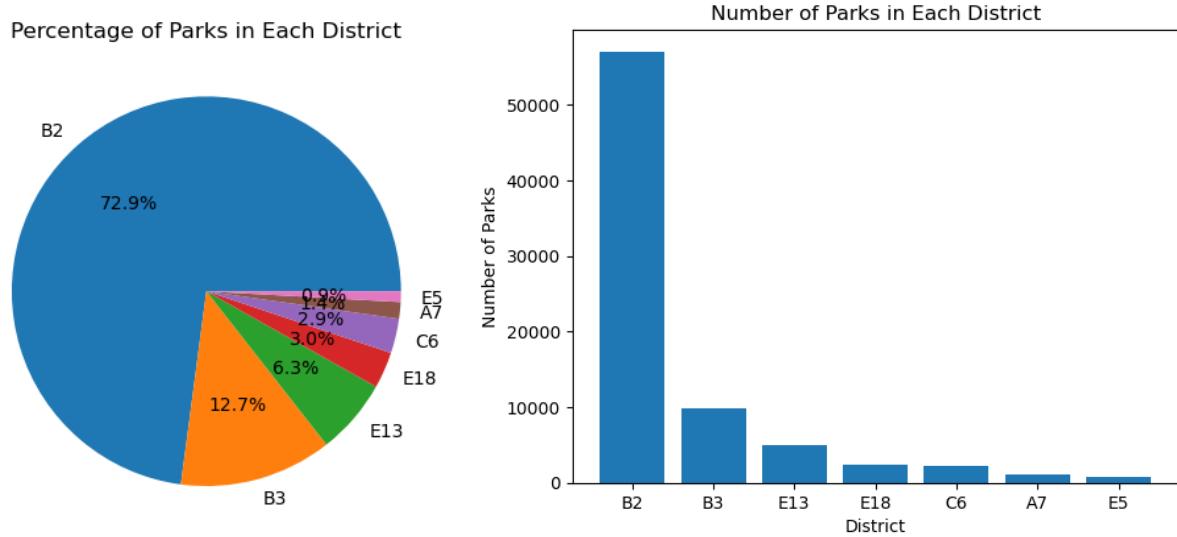
The plots reveal that different features are more prevalent in some locations than others. For example, in Dorchester and Roxbury, playgrounds are the most common feature in parks, while in Roslindale, basketball courts and baseball fields are more common. In Jamaica Plain, trails and paths are the most common feature in parks, and in Mattapan, spray pools and water play areas are more common.

Overall, the code provides useful insights into the distribution of park features in different areas of Boston, which can be helpful for city planners, community organizations, and residents who are interested in improving and maintaining their local parks.





The analysis reveals that, for the most part, there are few incidents of shootings occurring inside parks. More concerning is that a high percentage of shootings occurred within 0.5 miles of a park, although this rate varies across the districts.



Districts B2, B3, and E13 have the highest number of parks. Keep in mind district B2 contains Franklin Park, which is one of the biggest parks in Boston. District B2 has the highest count of parks and ballistic evidence and 4 out of the 5 districts with the highest counts of ballistic evidence are within the top 5 districts with the highest counts of parks.

To conclude, there is a correlation between the number of parks and the number of shootings. The higher the number of parks within a district, the higher the number of shootings within that district.

Contribution of Team Members

1. What is the rate of gun violence in District 4? How does this compare to the rest of the city?
 - Patrick Wright
 - Vaishnavi Vadlamudi
 - Tarek Mourad
2. Are there patterns of violence in terms of location in District 4? How does this compare to the rest of the city?
 - Snigdha Reddy Pulim
3. How do the patterns of student discipline in District 4 compare to those in Boston City as a whole?
 - Snigdha reddy pulim
4. Within District 4, at which hour do most shootings occur within our dataset which contains information about all shootings reported from 2015 to about the present day?
 - Vaishnavi Vadlamudi
5. What are the patterns and trends in the distribution of field contact counts over time (both annually and monthly)?
 - Tarek Mourad
6. Within District 4, on which days do most shootings occur within our dataset which contains information about all shootings reported from 2015 to about the present day?
 - Patrick Wright
7. Also looking at all the shootings that have taken place just within District 4, what is the number of shootings victims by race per year from 2015 - 2023?
 - Patrick Wright
8. What percentage of crimes are occurring on holidays?
 - Patrick Wright
9. Is there any correlation between the popularity of a street and the number of crimes on the street?
 - Patrick Wright
 - Vaishnavi Vadlamudi
10. What are the Ballistic evidence trends?
 - Tarek Mourad
11. How does the presence of parks affect the crime rate?
 - Snigdha Reddy Pulim
 - Tarek Mourad

Conclusion

This project aimed to analyze the drivers of gun violence in Boston's District 4, considering multiple datasets and community/environmental factors. The findings suggest that the location and type of violence can be identified by analyzing police records, discipline records from schools, and community/environmental factors. The incorporation of environmental/community factors, such as green space, tree canopy, community programming, interactions, and pedestrian/mobility data, provided a more

comprehensive analysis of the community's environmental health and well-being and its impact on mobility and access to public transit.

Furthermore, the correlation between the number of parks and the number of shootings indicates that the higher the number of parks within a district, the higher the number of shootings within that district. It highlights the importance of designing public spaces to promote safety and well-being.

Analyzing the trends of reported gun crime, we observed that the year-on-year growth rate has been decreasing over the past few years. Though there are some irregularities due to the COVID-19 pandemic, the data suggests that the trend has been slowing down, and the numbers are not growing as quickly as they have in the past.

District D4 had a significantly higher number of field contacts compared to other districts, which increased by almost 100% since 2017. District E13 had a decrease in field contacts from around 20% to 11%. The months with the highest counts of field contacts were in the spring/summertime. The findings can inform policies that would involve higher police presence in the areas where the gun crime has been observed the highest.

Finally, the analysis of community/environmental factors and trends in reported gun crime emphasizes the importance of promoting sustainability, equity, and well-being in the community. The average counts for all months being around 4200, it is crucial to continue monitoring the situation and implementing measures to ensure public safety.

Appendix:

Code Documentation:

Repository Navigation:

Here is the list of files and folders in the repository along with a brief description of what each file or folder contains:

Deliverables:

Deliverable 1:

- crime_patterns_per_PD_snigdha.ipynb: This Jupyter notebook provides analysis on the number of crimes in Boston city and per Police district in city council district 4.
- CrimeIncidentReport_Vaishnavi.ipynb: This Jupyter notebook presents analysis on crimes involving shootings in district 4, rest of the districts, and Boston city, along with comparisons.
- ReadingShapeFiles_Vaishnavi.ipynb: This Jupyter notebook showcases different shape files of Boston city and city council districts.
- gun_violence_PatW.ipynb: This Jupyter notebook utilizes the Shootings dashboard, which contains information on shooting incidents where a victim was struck by a bullet, to visualize the proportion of shootings within district 4 compared to the rest of the city.
- Field_Contacts_Data.ipynb: This Jupyter notebook analyzes the Top Zip Codes within Boston City based on the number of field interrogations, using the BPD FIELD INTERROGATION AND OBSERVATION data. Additionally, for more detailed information on deliverable 1, you can refer to the Deliverable 1.pdf file. This PDF file provides insights, analysis, and visualizations related to the crime patterns, shootings, shape files, gun violence, and field contacts data in Boston city and city council district 4.

Deliverable 2:

code folder, along with a brief description of what each file contains:

- CrimeIncidentReport_AllYears_Vaishnavi.ipynb: This Jupyter notebook continues from "CrimeIncidentReport_Vaishnavi.ipynb" in deliverable 1 and includes analysis on the number of shootings by hour and its trend for the last 10 years.
- Field_Contacts_Data.ipynb: This Jupyter notebook continues from "Field_Contacts_Data.ipynb" in deliverable 1 and provides analysis on the trend in field contacts by months for the last 5 years.
- gun_violence_PatW.ipynb: This Jupyter notebook continues from "gun_violence_PatW.ipynb" in deliverable 1 and identifies patterns based on the days of shootings, race of the victim, and gender of the victim for the last 7 years.
- shooting_crime_patterns.ipynb: This Jupyter notebook presents analysis on the number of crimes containing shootings in Boston city and per Police district in city council district 4, focusing on data from 2019 onwards (last 5 years).
- student_discipline.ipynb: This Jupyter notebook provides analysis on Student Discipline Data for the last decade, with a focus on district 4. Additionally, for more detailed information on deliverable 2, you can refer to the Deliverable 2.pdf file in the repository. This PDF file contains insights, visualizations, and analysis related to the code files listed above. In addition to the aforementioned, our project is committed to broadening its scope by incorporating Environmental/Community Factors. We will specifically focus on factors such as Green Space, Tree Canopy, Community Programming, Interactions, and Pedestrian/Mobility data. By incorporating these additional factors, our project aims to provide a more comprehensive analysis of the community's environmental health and well-being, as well as its impact on mobility and access to public transit. This expansion will enhance our project's insights and contribute to urban planning and community development recommendations.

Deliverable 3

code folder, along with a brief description of what each file contains:

- count_on_borders.ipynb: This Python code imports GeoPandas and Shapely libraries to load a shapefile of police districts and crime data in CSV format, convert the latitude and longitude columns into Shapely Point objects, loop through each police district and crime point to check if the crime point is on the police district border, and count the number of crimes on the police district borders.
- CrimeIncidentReport_AllYears_Vaishnavi.ipynb: This code is a Python script that imports the Pandas and Matplotlib libraries to read and plot data from a CSV file containing crime incident reports. The script filters the data to show the number of shootings and other crimes in different police districts, and creates a bar plot to visualize the results. It also calculates the year-on-year growth rate of reported gun crime for each year from 2015 to 2022, and prints the results to the console. Additionally, the script filters the data for two specific police districts (B2 and B3) and shows the number of shootings and other crimes in each district. Overall, this script provides a basic analysis of gun crime trends and the distribution of crime in different police districts.
- gv2.ipynb: The Python code analyzes two datasets from the city of Boston: the "Shots Fired" dataset and the "Boston Park Assets" dataset. The code aims to find a correlation between the number of parks in an area and the counts of ballistic evidence within that area. The code uses pandas for data filtering and merging, and Matplotlib to create a pie chart showing the distribution of districts with ballistic evidence. The resulting merged dataframe contains information about each park, its location, and the corresponding zip code group.
- parks.ipynb: The file describes an analysis of crime rates in different neighborhoods (police districts) in Boston and their correlation with parks and recreational areas. The analysis includes several code snippets in Python, which explore different aspects of the relationship between parks and crime, such as the number of parks per location, the number of parks per feature in Boston, and the number of shootings inside, near, and outside parks in different police districts. The analysis provides useful insights for city planners and law enforcement agencies to make informed decisions about park management and public safety.

- `ReadingShapeFiles.ipynb`: The code consists of several Python modules used for data analysis and visualization of crime incidents in the city of Boston. It utilizes various libraries such as geopandas, pandas, matplotlib, and folium to read and manipulate shapefiles, CSV files, and data frames, and to create maps and plots. The data is filtered to show active districts and shootings, and the locations are plotted on a map using a marker cluster. The code also adds additional markers to the map to indicate specific locations of interest. Overall, the code is designed to provide insights into crime patterns and locations in the city of Boston.
- `TransportationGV.ipynb`: This code snippet imports the necessary libraries such as numpy, matplotlib.pyplot, and pandas. It reads a CSV file named `BTDTrafficData.csv` into a pandas dataframe and then drops certain rows based on the 'Neighborhood Code' column. It also adds a new column 'Street Count' to the dataframe which counts the number of streets in each cell of the 'Associated Streets' column. It then extracts all unique street names from the 'Associated Streets' column and counts their occurrences in the dataframe, displaying the 10 most common and 10 least common streets in two separate bar graphs. Finally, it counts the occurrences of each study type in the 'Study Type' column and displays the result in a pie chart. ##### Deliverable 4 In Deliverable 4, the team made several updates to the codebase and added a new file. Here are the details: code folder, along with a brief description of what each file contains:
Updated Files
 - `gun_violence_PatW.ipynb`: This file remained mostly unchanged from Deliverable 3, but a new graph was added to depict "Shootings by Holiday Status". This graph displays the percentage of shootings that occurred during holidays, which is 3.26%.
 - `TransportationGV.ipynb`: The code in this file is similar to the file in Deliverable 3 with the same name. However, some updates were made to the "Percentage of Each Study Type" graph. Specifically, the graph now displays only two study types: "TURNING MOVEMENT COUNT" and "AUTOMATED TRAFFIC RECORDING".
 - `ReadingShapeFiles.ipynb`: This file has been updated by adding more maps to show the "Crime incident reports along the busy streets of district 4" and the "Crime incident reports in the busy streets." The addition of these maps provides a more detailed view of crime incidents in specific areas of the district, which can be helpful in identifying patterns or trends. New File
 - `district_maps_with_crime_location.ipynb`: This file allows the user to run the code to get police district maps with crime locations marked on them. The data used for this map is from 2015 to 2023, providing a comprehensive view of crime incidents in the district. This addition can be particularly helpful in identifying crime hotspots or trends over time. Overall, these changes improve the quality of the project and provide a more comprehensive view of the data.

Dataset Navigation Instruction:

All the data is in the data folder under the Sprint23-team-4 folder

Results Reproduction Instruction

To get plots for updated data:

1. Replace the particular data file in the data folder with the updated file (use the same name).
2. Go to the appropriate files in the deliverables folder.
3. Install the required applications to run the jupyter notebook file
4. Run the jupyter notebook file