

UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

DATA SCIENCE LAB ON SMART CITIES

FINAL ESSAY

Crime Data Analysis

(In reference to the regions in Boston located in Massachusetts, United States)

Authors:

Induni Sandapiumi Nawarathna Pitiyage - 906451 - i.pitiyage@campus.unimib.it

Sara Campolattano - 906453 - s.campolattano@campus.unimib.it

12 September 2024



Abstract

This report aims to present a crime analysis based on the criminal data reported from 2015 to 2018 in Boston (located in Massachusetts, United States), focusing on the measures to mitigate criminal activities in urban areas. By examining various crime indicators such as crime types, temporal and geographic dimensions, the report provides a detail understanding of criminal patterns that exist in that urban areas. In particular, the crime type indicators explored focus on crimes against property. Temporal indicators reflect how crime patterns fluctuates based on the time related factors, including the day of the week and time of the day. On the other hand, geographic indicators include the identification of crime hotspots, proximity to key locations and street level crime analysis. By mapping out areas with high concentration of criminal activity and examining the relationship between the crime locations to urban facilities such as schools, parks and transportation hubs, the report provides actionable insights for targeted law enforcement interventions.

Contents

1. Introduction.....	4
1.1 Challenges and Ethical Implications.....	5
1.2 Indicators.....	6
2. Data Analysis.....	8
2.1 Dataset.....	8
2.2 Data Pre-processing.....	8
2.3 Descriptive Statistics.....	8
2.4 Spatial Analysis.....	10
Conclusions.....	17
References.....	18

1. Introduction

Modern urban security planning and the governance are greatly impacted by two major phenomena: the rapid urbanization of the city and the advancement of the information communication technology. The concept of smart cities utilizes advanced technology and innovative solutions to enhance the quality of life among the residents, improve efficiency of urban services, and promote sustainable development. The primary objective of smart cities is to maximise resource management, reduce negative environmental effects, enhance public services (such as healthcare, education, transportation and security) and promote economic development.

Among the different fields, security management in smart cities holds a relevant position due to its features and functionalities. Improving one's own security and elevating the general sense of security within the city are both necessary for a safe city. Security management against crimes in a city involves the strategic planning, coordination and implementation measures to mitigate criminal activities ensuring the safety and protection of the citizens, infrastructure and public spaces. Here are some key considerations that should be considered to prevent, detect and respond to criminal activities.

1. Surveillance systems: surveillance systems in smart cities interrogate advanced technologies like artificial intelligent and IoT, which are able to extract structured information such as licensed plate numbers, vehicle colours/models and facial IDs. These technologies aid in preventing crime by law enforcement giving them access to timely warnings and actionable intelligence. Smart cities can improve public safety by identifying high-risk locations and periods by examining the patterns in the collected data.
2. Data driven policy: data driven policies in smart cities rely on huge amount of data and advanced analytics to inform decision making processes related to crime prevention and law enforcement. By analysing crime statistics, trends and urban infrastructural data, policymakers are able to derive targeted interventions and allocate resources more effectively. By relying on these datasets, policymakers can ensure the policies are made to be evidence-based, flexible and suited to the particular requirements of the community.
3. Community engagement: in the context of crime and smart cities, community engagement through social media and digital platforms helps to promote the contact between public citizens and law enforcement. Residents can report events, discuss safety issues and get real time update from the authorities. This kind of community engagement foster trust, enhance public awareness and promote cooperative efforts to lower the crime.

4. Technological innovation: implementation of new technological innovations and deployment of advanced tools in smart cities are aimed at improving public safety and prevention of crime. This includes innovations such as predictive policy algorithms, drone for aerial surveillance, intelligent street lighting that adjust based on the activity levels.
5. Law enforcement strategies: to increase the effectiveness and efficiency, law enforcement initiatives in smart cities incorporate technology into conventional policing and techniques. This includes the use of mobile apps for public outreach, deployment of body cameras to maintain the transparency and data analytics for crime prediction and crime prevention. These strategies help law enforcement to be responsive, proactive and accountable, which also helps to lower the crime rates and fosters the public trust.
6. Urban design: urban design in smart cities plays a crucial role in crime prevention by fostering environments that deter criminal activity. This involves creating easy monitoring public spaces that promote natural surveillance. To make sure of the citizens safety, urban planning leverage on data on crime hotspots to guide the layout of the streets, parks and buildings. Smartly planned urban cities can drastically cut down on crime by promoting accessible safer community.

Security management in smart cities requires a multifaced approach to address various types of crime. Leveraging on advanced technologies and data driven strategies can help mitigating the criminal activities. Although various types of crimes afflict the world, in this study we focused on crimes against property.

To effectively implement the above discussed measures, smart cities need to establish a robust infrastructural data extraction, analysis and the real-time communication among various parties including law enforcement, government agencies, private sector parties and the community. Continuous monitoring and adaptation to security measures are essential in mitigating criminal activities in connected environment.

1.1 Challenges and Ethical Implications

The management of crime against property in cities involves a wide range of challenges and ethical considerations. Here are some challenges that are typically encountered in managing relevant security measures to mitigate criminal activities.

- Integration of data from several sources, including social media, IoT devices and CCTV cameras are essential to ensure cities' safety. However, it might be difficult to ensure the compatibility across many systems and data formats. Errors in data integration also leads to inaccurate crime detection and response.

- The extensive use of data collection, and surveillance technology might also lead to concerns about data privacy and security. Unauthorized access or misuse of personal data can violate citizen's privacy rights and jeopardize people's trust in smart city initiatives.
- Challenges in allocation of scarce resources such as police officers and surveillance equipment in areas with the highest risks of property crimes. Misallocation of resources can also lead to under-protection of some areas.
- Bias in the training data inherited by AI and machine learning algorithms employed for crime prevention might lead to unfair policy actions that disproportionately target particular groups, which consequences may cause discrimination.
- With the development of advanced technological equipment, cities interconnectedness also increases and eventually become more vulnerable to cyberattacks that could disrupt services and compromise security measures.

Ethical implications that arise along with the challenges that should be considered are:

- Policies and practices must be placed to ensure informed consent from the residents.
- All the citizens must be treated and benefit equally from smart city initiatives.
- Mechanism placed in the society must provide transparency and hold accountability for the decisions to be made from automated services.
- Ethical guidelines must be implemented to govern the use and sharing of data related to criminal activities.

Maintaining the rights and the liberties of residents in these cities while utilising the technology to increase the security against crimes must be carefully balanced in order to overcome these ethical dilemmas and challenges.

1.2 Indicators

Cities are growing very quickly through many technological achievements along with the criminal active areas. As previously mentioned, the main objective of this study is to

analyze areas interested by crime that occurred in 2015-2018 in Boston. To achieve this, trends and patterns of criminal activities were inspected, following indicators.

1. Crime type Indicators

- **Crime Category:** this analysis focuses on specific types of crime such as burglary, larceny, vandalism, arson or theft. By understanding the prevalence and the distribution of different crime categories, authorities are able to allocate their resources based in their severity to the public.

2. Temporal Indicators

- **Day of the week:** this analysis explores how crime patterns fluctuate based on the day of the week. For instance, certain types of crimes may be prevalent on weekends, while others may be prevalent on weekdays. By understanding these locations, law enforcement can better allocate their resources.
- **Time of the day:** crimes often follow specific time patterns. This indicator helps to breakdown crime occurrences into periods such as morning, afternoon and night. It helps in analysing peak hours for specific crimes, aiding targeting high risk crimes.
- **Seasonal pattern:** this indicator helps understand whether occurred crimes vary in seasons or months of the year. For instance, crime against property such as burglaries might occur in summer when people are on vacation.

3. Geographic Indicators

- **Crime Hotspots:** this involves identifying areas with high concentration of criminal activities. These could be districts, neighbourhoods or local areas likes crosswalks. By identifying these hotspot's, law enforcement can concentrate their efforts on the most impacted areas.
- **Proximity to key locations:** analysing crime in relation to key locations such as schools, parks and transportation hubs provides insights into how the presence of these facilities increases the crimes rate. For instance, crimes might be more frequent near crowded transit like stations and parks.
- **Street level crime analysis:** this indicator helps in identifying specific street locations among districts with high crime rates. By pinpointing these micro-locations, law enforcement can deploy target patrol or community outreach programmes.

2. Data Analysis

2.1 Dataset

The Dataset for the study is obtained from the Kaggle platform. Crime incident reports included in this dataset are provided by the Boston Police Department (BPD) and contain initial details around the surrounding of an incident to which BDP officers respond. The dataset encompasses 327,820 individual records on 17 attributes. Various criminal activities are recorded from June 14, 2015 to September 3, 2018.

2.2 Dataset Pre-processing

Crimes can be classified into various categories based on their nature, severity and impact. Since the original dataset is encompassed with various types of crimes, only data related to crimes against property were extracted.

The extracted dataset is composed with 67,131 crime records reported from June 15, 2015, to September 3, 2018, including variables such as incident number, offense code, offense code group, offense description, district, reporting area, shooting, occurred on date, year, month, day of the week, hour, UCR part, latitude, longitude and location.

The dataset presents with missing values for several attributes. Specifically, the attributes with missing values along with the percentage of missing records are: shooting (99.9%), district (0.3%), reporting area (3.2%), street (1.6%), latitude (3.3%) and longitude (3.3%). This dataset also encompassed several misinformation related to crime location. It was therefore decided to remove the “Shooting” attribute from the dataset, given the high missing values percentage. Missing records contained in attributes “latitude” and “longitude” and “district” were also removed, along with those crime records that contain misinformation related to location.

Feature engineering was performed by creating a “date” and “time” from the original attribute “occurred on date” to capture trends of the criminal activities. Moreover, using “longitude” and “latitude”, data proximity distances such as the distance from the place in which the criminal activity has occurred to amenities (schools and parks) and public transport (stations and stop positions) were engineered to gain insights about the neighbourhood of the crime.

Following pre-processing, the final dataset comprises of 64,643 crime records on 24 attributes.

2.3 Descriptive Statistics

Understanding the distribution of crime type is essential for identifying which categories of crime are most prevalent in a given area. *Figure 1* displays the crime count for each year (from 2015 to 2018) that have been reported in Boston.

As it can be seen in *Figure 1*, the most frequent type of crime against property that can be observed across several years is larceny, followed by vandalism and larceny from motor vehicles. It can also be seen that the highest crime count for all the crime types was recorded in 2016 and 2017, while the crime count observed 2015 is comparatively low.

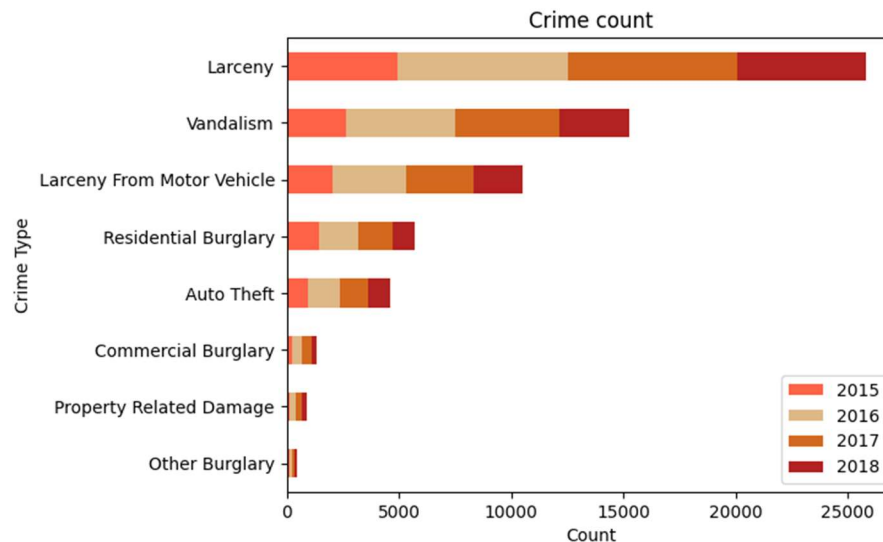


Figure 1: Crime counts by category

Crimes do not seem to be uniformly distributed across time, and certain days may witness higher crime than in others. *Figure 2* illustrates the crime count by the day of the week that observed in years 2015 to 2018.

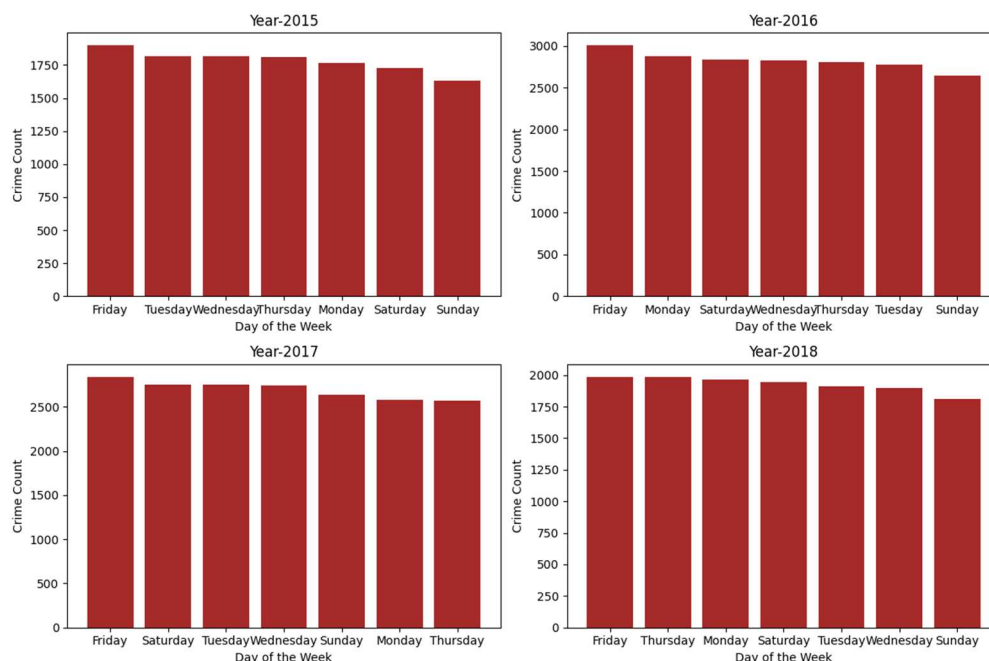


Figure 2: Crime count by the day of week

As depicted in *Figure 2*, in 2015, 2016 and 2018 the least crime occurrence day is Sunday while only for 2017 it is Thursday. As for the highest crime occurrence day, for all the observed years is Friday.

Criminal activity often fluctuates through the day, with certain hours being more prone to incidents than others. *Figure 3* shows the distribution of the crimes by hour of the day.

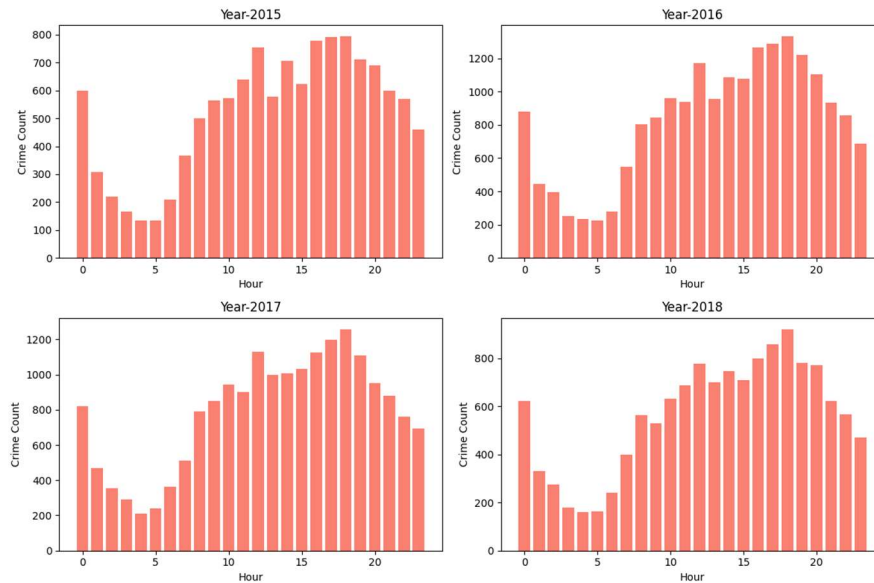
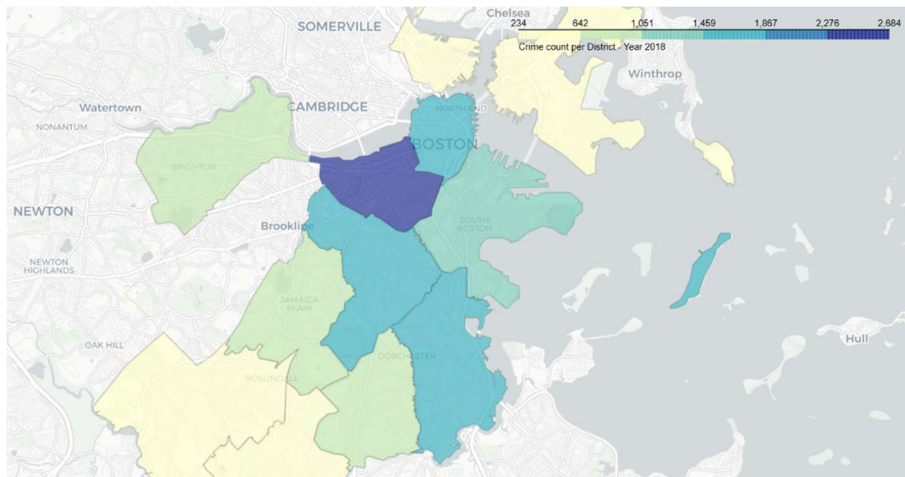


Figure3: Crime count by the hour of the day

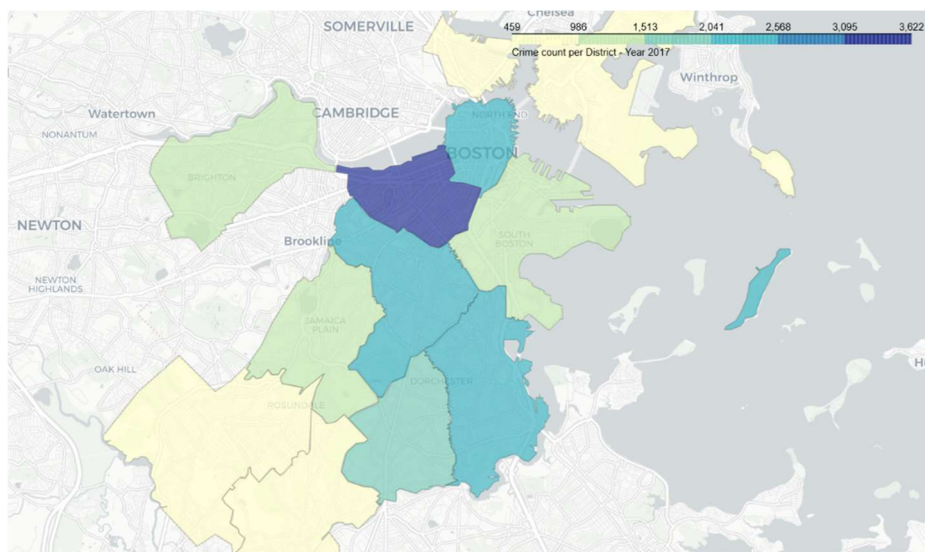
Hourly patterns shown in *Figure 4*, display similar patterns for all the years. This analysis reveals the peak crime hours are mid-day and evening hours from 14:00 to 16:00 while the least crime hours for all the years are around 1:00 to 7:00. These hourly patterns emphasize that crimes have often likely happened during the day or working hours.

2.4 Spatial Analysis

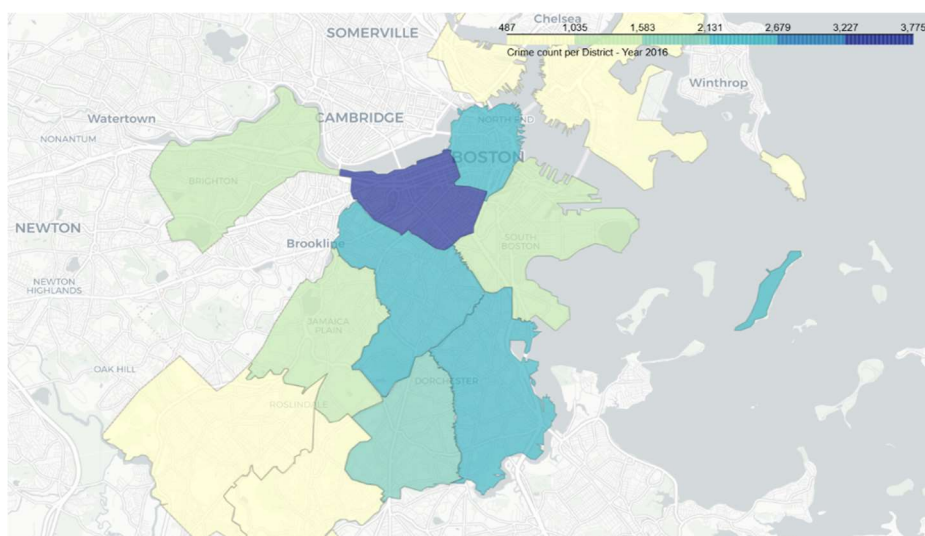
In the following, the geographic distribution of crime, with each police districts shaded according to the frequency of the crimes that occurred respective to each year is displayed. These crime maps show that the crime concentration is not shifted throughout the periods (2015 to 2018). However, the crime frequency over the periods have been changed. The highest number of crimes for all the districts have occurred in 2016 while the least number of crimes for all the districts have occurred in 2015.



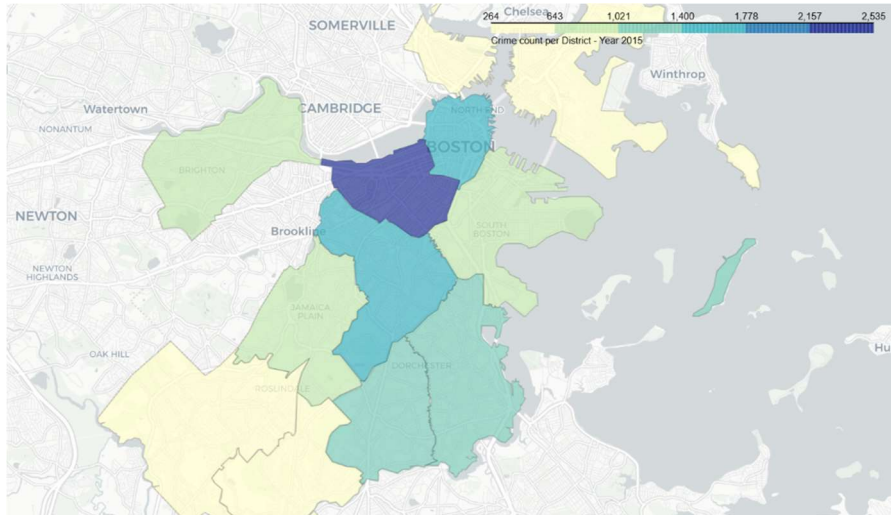
(a) Crime counts per District – Year 2018



(d) Crime counts per District – Year 2017



(d) Crime counts per District – Year 2016



(e) Crime counts per District – Year 2017

Figure 4: Choropleth maps for Crime count per Police District (2015-2018).

To further understand the distribution of crime with respect to different type of crimes under crimes against property for the year of 2018 following plots have been incorporated.

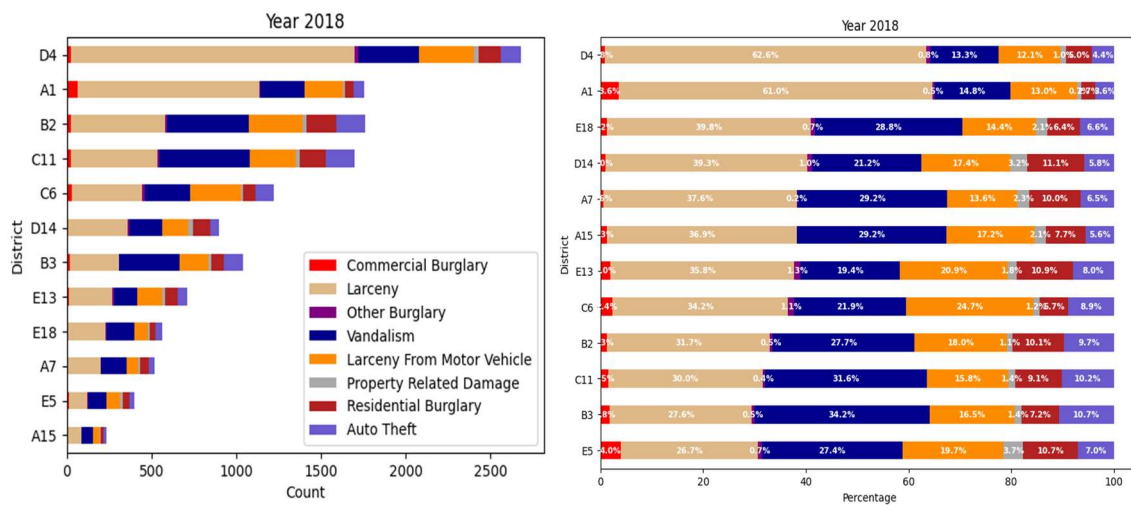
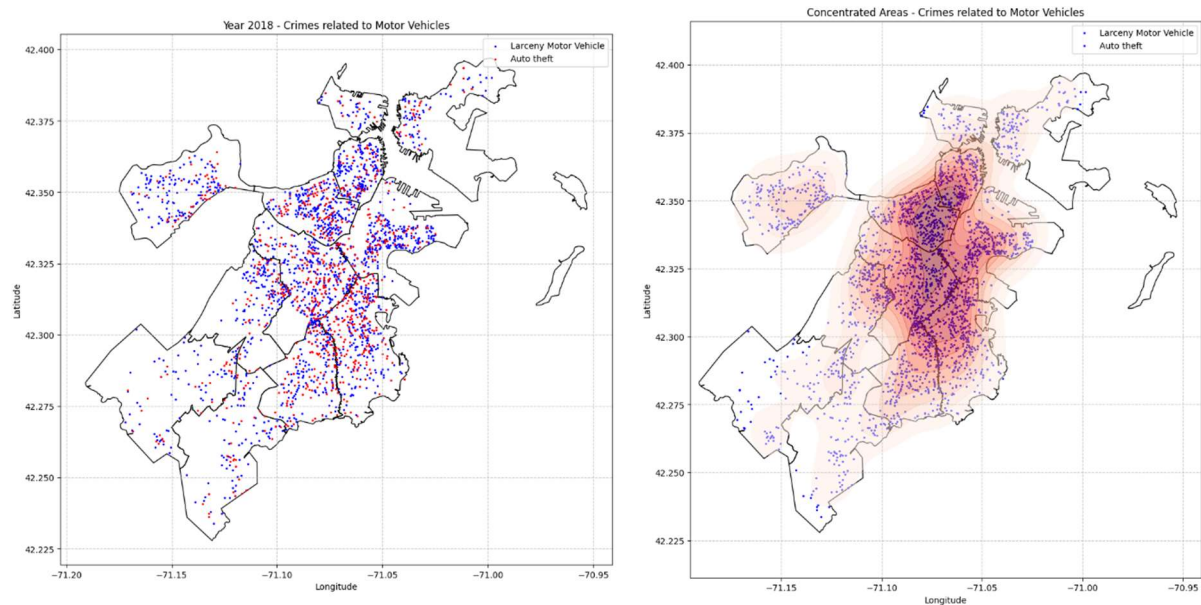


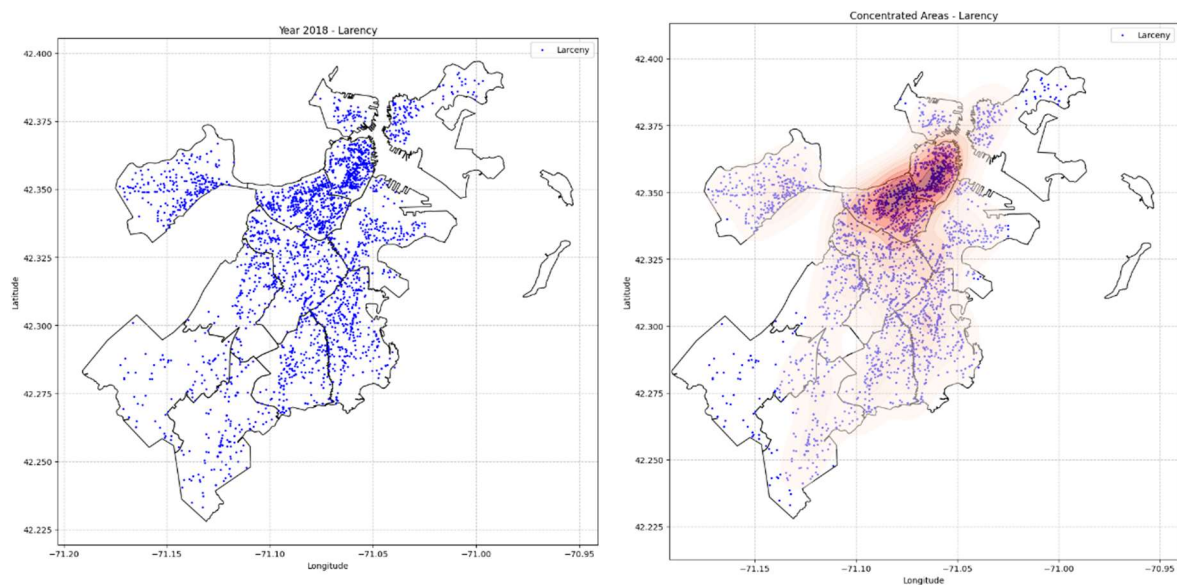
Figure 5: Crimes occurred in Police Districts for year 2018.

As shown in Figure 5, most recorded crimes were reported from the district D4 followed by the districts A1, B2, and C11 exceeding 1500 crime records per year. Despite the bigger land areas in E18 and E5, recorded criminal activities are quite low. Most of the crimes are reported regard larceny and vandalism for all the districts while very few crimes are reported in relation to arson and property related damage.

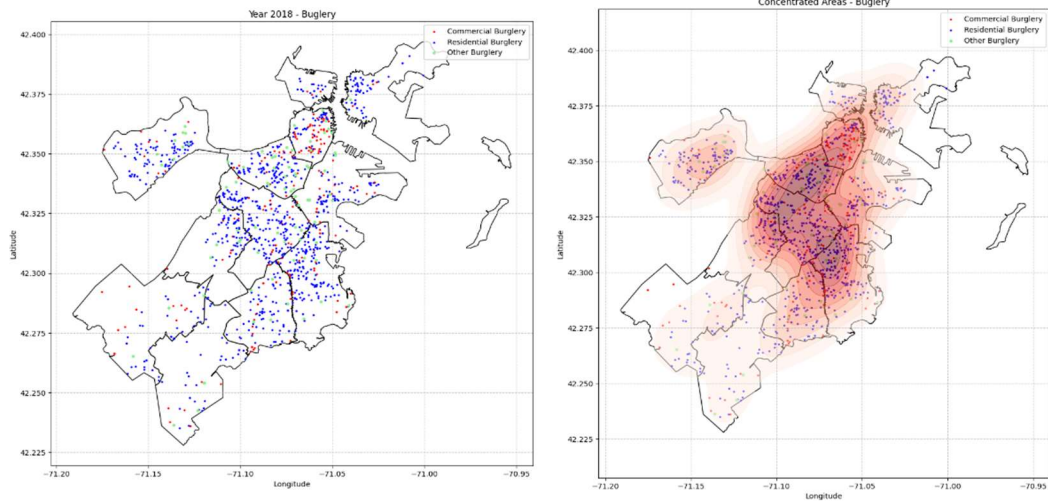
Moreover, to understand the exact location where criminal activities have occurred and to gain insights about whether these crimes occurred in the same place, following plots depicted in *Figure 6* were incorporated. Criminal activities related to motor vehicles for districts A1, D4, B2, C11, D14 and A7 have occurred close to each other, while for districts E18 and E5 they are quite dispersing with large land area.



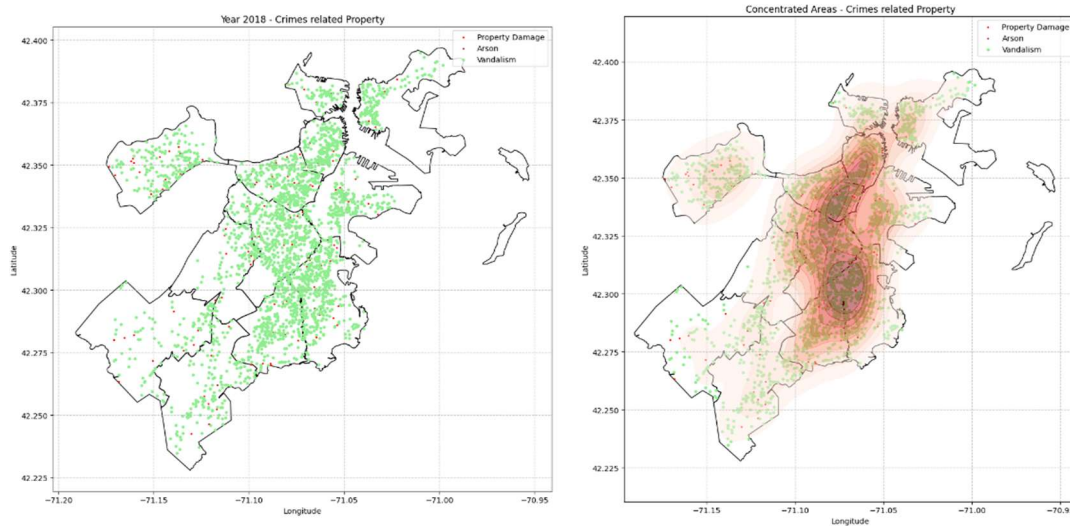
(a) Crime concentrated Areas (Motor Vehicles theft)- Year 2018



(b) Crime concentrated Areas (Larceny) – Year 2018



(c) Crime concentrated Areas (Burglary) – Year 2018



(d) Crime concentrated Areas (Crime related to Property) – Year 2018

Figure 6: Crime concentrated areas for crimes against property for year 2018.

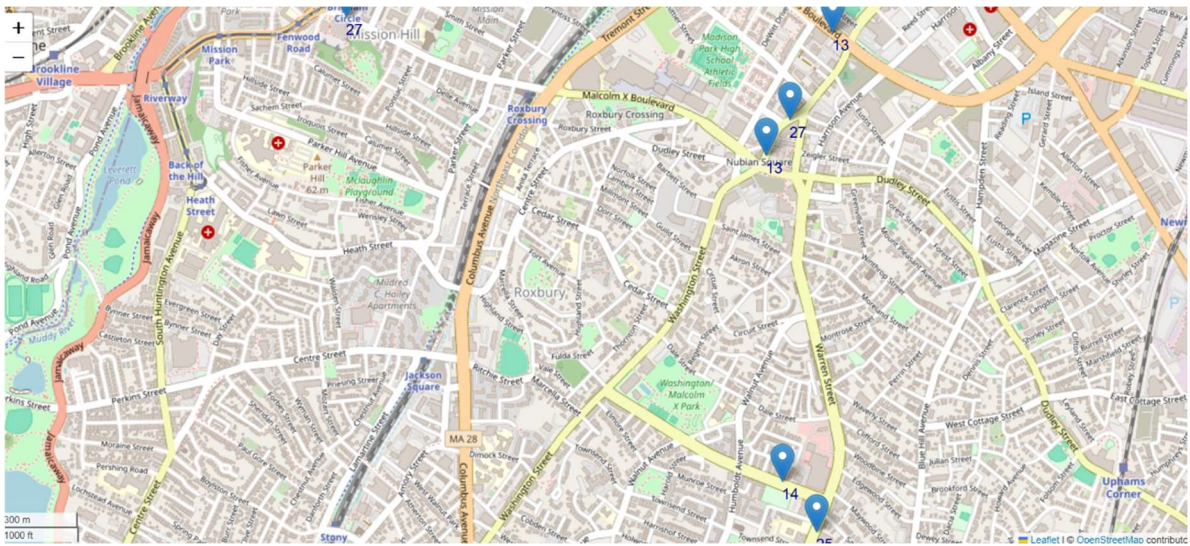
Similarly, crimes such as burglary and property related crimes also concentrated in same places in districts A1, D4, B2, C11, D14 and A7. However, larceny, except for districts A1 and D4, in all other regions it is quite disperse.

Since the number of crimes in districts A1, B2, C11 and D4 are relatively higher than other police districts in Boston, the following plots have been incorporated to further analyse criminal hotspots or repeated locations where these crimes occurred, specifically in 2018.

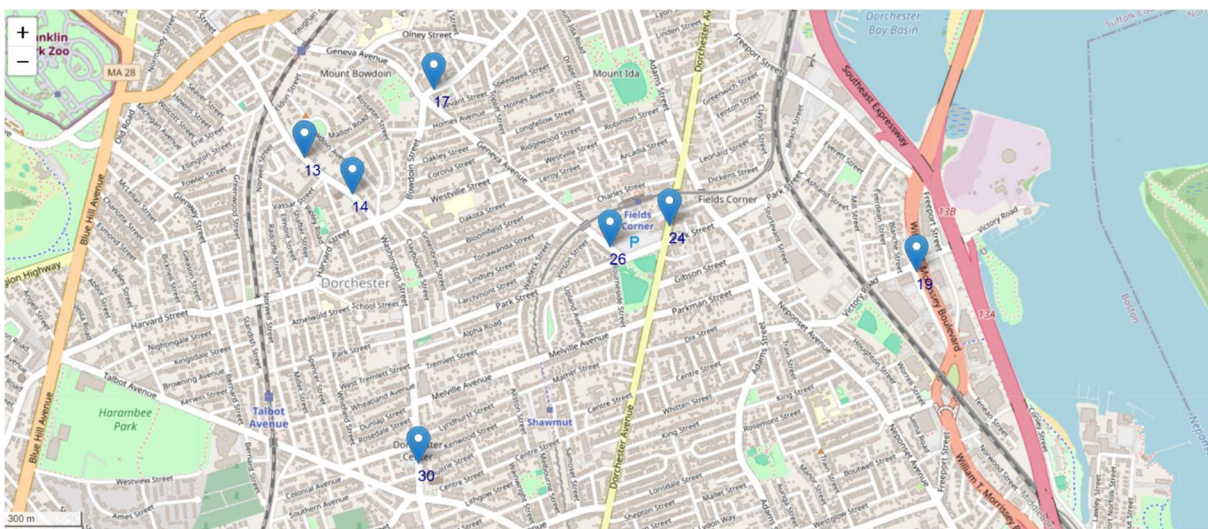
Figure 7 shows the top ten criminal hotspots for each district (A1, B2, C11 and D4) along with the number of crimes that occurred during 2018. This analysis helps identifying specific street locations among the districts with higher number of crimes. By pinpointing these micro-locations, law enforcement can deploy target patrol or community outreach programs to mitigate the criminal activities that, to some extent, prevail in that area.



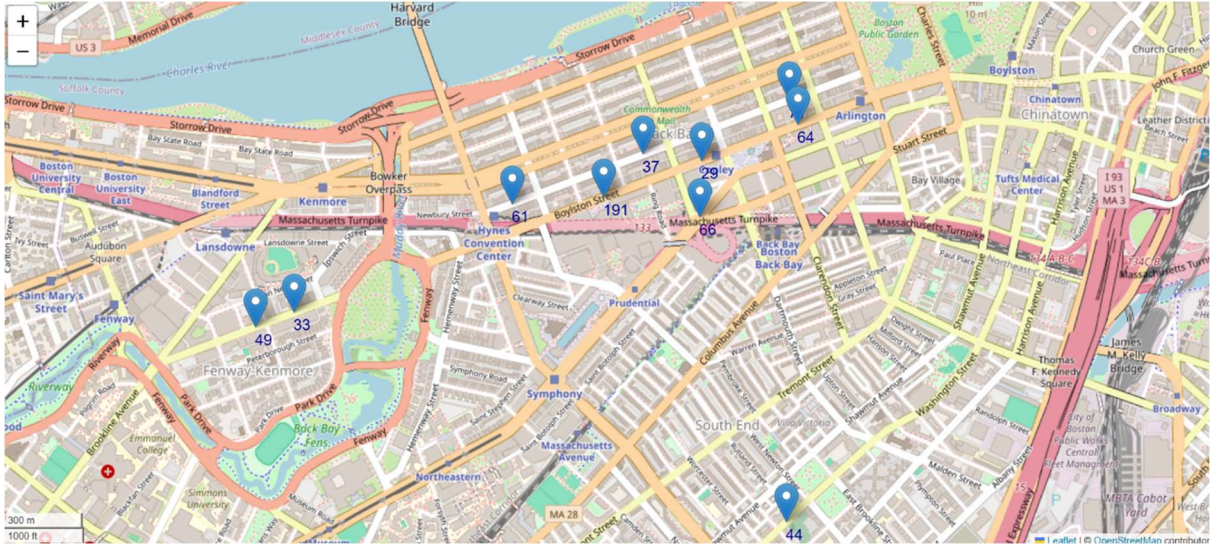
(a) A1 District



(b) B2 District



(c) C11 District



(d) D4 District

Figure 7: Crime Hotspots for A1, B2, C11 and D4 districts in 2018

It is interesting to observe that in district D4, which is the police district with the highest number of crimes reported, most of crimes are repeatedly reported from Boylston Street. Allocation of police officers and surveillance equipment to these areas might help to mitigate crime risks to certain extent.

Analysing crime in relation to key locations such as schools, parks and transportation hubs provides insights into how the presence of these facilities increases the crimes rate. Proximity to public places such as schools, parks and transportation hubs can significantly influence the crime patterns in urban areas.

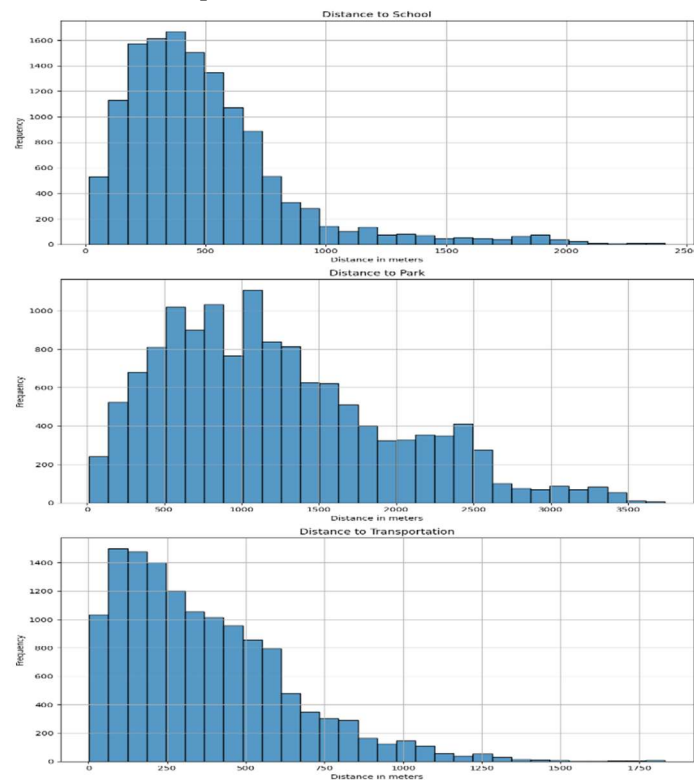


Figure 8: Proximity from schools, parks and transportation hubs to nearest crime location.

Figure 8 shows the proximity from the schools, parks and transportation hubs to the nearest crime locations of crimes that occurred in 2018. As depicted in *Figure 8*, the majority of the crime locations are quite near to these public spaces, which suggests that most of crime activities does not occur in isolated environments.

However, to overcome these problems, active community involvement, effective lightening, surveillance and visible police presence can help to mitigate crime risks in these areas. Understanding these proximity factors is crucial for urban planning and crime prevention strategies, helping to create a safer environment for all the residents.

Conclusions

The comprehensive analysis of crime data through crime type, temporal and geographic indicators has revealed several insights that can inform future policy-making and law enforcement strategies. Examining the crime types related to crimes against property revealed that most common type of crimes observed for all police districts in Boston are larceny, larceny from motor-vehicles and vandalism. Out of all the years (2015 -2018), 2017 and 2016 have seen the highest crime count for all the crime types while the crime count observed in 2015 is comparatively low. The analysis of crime pattern by the day of the week and the time of the day has shown that the crime days and the time periods are particularly prone to higher crime rates. For instance, for all the years, the highest number of crimes are reported on Fridays, except 2017, while during Sunday the least number of crimes are reported, most likely due to the presence of residents. Hourly pattern of the crimes for all the years showed a similar pattern revealing the peak crime hours are mid-day and evening hours from 14:00 to 16:00 while the least crime hours for all the years are around 1:00 to 7:00. These hourly patterns emphasize that crimes have often likely happened during the day or working hours. The geographic distribution of the crimes for all the police district in Boston revealed that crime concentration is not shifted throughout the periods (2015 to 2018). However, the crime frequency over the periods has changed. The highest number of crimes for all the districts has occurred in the 2016 while the least number of crimes for all the districts has occurred in the 2015. To further understand the crime concentrated areas and key crime hotspots, crime data that reported in year 2018 was taken into consideration. Through this analysis, it was possible to identify the main crime concentrated districts, which are A1, B2, C11 and D4. Out of these districts, D4 is the highest crime reported district. This analysis helped in identifying specific street locations among the districts with highest repeated number of crimes. To further gain insights into the neighbourhood around the crime location, proximity of the crimes to critical locations such as schools, public parks and transportation hubs were considered. The analysis revealed that the majority of the crime locations are quite near to these public spaces (less than 1000m from crime

location to public spaces), suggesting that most of the crimes against property have not occurred in isolated environments. The intersection of these crime type, temporal and geographic indicators provide a multidimensional understanding of the crime dynamics in the area, providing law enforcement and policy-makers with crucial information for future interventions.

References

Magdalena Tutak¹, Jaroslaw Brodny² “A Smart city is a safe city: Analysis of the State of Crime and safety in Polish cities”, *Applied Science and Humanities for Smart Cities*, published: 29 November 2023.

Kai Seidensticker¹, Katharina Schwarz² “Using Forecasting methods on Crime Data: The KALA Approach of the State Office for Criminal Investigation of North Rhine-Westphalia”, *The 8th International Conference on Time Series and Forecasting*, Published: 11 July 2022.

Juanjo Medina and Reka Solymosi,” chapter 7: Time matters”, *Crime Mapping and Spatial Data*, Published : 09 September 2022.

Dawei Wang ^a, Wei Ding^a, Henry Lo ^a, Melissa Morabito ^b, Ping Chen ^e, Josue Salazr, Tomas z Stepinski ^d “Understanding the spatial distribution of crime based on its related variables using geospatial discriminative patterns.” , *Computers, Environment and Urban Systems*, Volume 39, pages 93-106. Available online 9 April 2013.

Agnieszka Lisowska-Kierepka, “How to analyse spatial distribution of crime? Crime risk indicator in an attempt to design an original method of spatial crime analysis” , *Cities*, Volume 120, Available online 2 August 2021.

Giedrė Beconytė, Michael Govorov, Andrius Balčiūnas and Darius Vasiliauskas, “Spatial distribution of criminal events in Lithuania in 2015–2019”, *Journal of Maps Volume 17*, Published online: 07 Dec 2021

Data preparation and data analysis for the study:

<https://drive.google.com/drive/folders/1u6DJf56d3PM9FXuou8bUaxSo77dH2raX?usp=sharing>