Exploring Crime Patterns in Boston to reduce the level of crime with use of Big Data Analytics

Kieran Colhoun

Faculty of Science and Engineering

University of Wolverhampton

Wolverhampton, United Kingdom

K.Colhoun@wlv.ac.uk

Loucas Xiourouppa

Faculty of Science and Engineering

University of Wolverhampton

Wolverhampton, United Kingdom

L.Xiourouppa@wlv.ac.uk

Peer Review - Contributions to this Report:

**Abstract**- Loucas Xiourouppa & Kieran Colhoun

**Background of Study**- Kieran Colhoun & Loucas Xiourouppa

**Related Work**- Loucas Xiourouppa & Kieran Colhoun

**Methodology**- Kieran Colhoun & Loucas Xiourouppa

**Result and Discussion**- Loucas Xiourouppa & Kieran Colhoun

**Conclusion**- Kieran Colhoun & Loucas Xiourouppa

**Code Appendix (Part B)**- Loucas Xiourouppa & Kieran Colhoun

**Table of Contents**

[Abstract 2](#_Toc133492310)

[Background of study 2](#_Toc133492311)

[Related work 3](#_Toc133492312)

[How our work differs 3](#_Toc133492313)

[Methodology 3](#_Toc133492314)

[Block Diagram 3](#_Toc133492315)

[Result and Discussion 4](#_Toc133492316)

[Experimental setup 4](#_Toc133492317)

[Discussion of the findings 4](#_Toc133492318)

[Analysis of the findings 5](#_Toc133492319)

[Conclusion 7](#_Toc133492320)

[References 8](#_Toc133492321)

[Code Appendix (Part B) 8](#_Toc133492322)

# Abstract

Boston City is the capital of the commonwealth of Massachusetts. The estimated total population is around 617,594 and is the tenth-largest metropolitan area in the country [1]. Boston also contributes substantially towards the population density within Massachusetts with 13,841 people per square mile and that makes Massachusetts the 3rd largest densely populated state in the US [2].

Managing crime in densely populated areas is an issue tackled around the globe and it can leave local authorities facing an endless number of issues such as appropriate resource allocation, physical police presence, amount of policing staff as well as more specific issues with being under-utilised to face more violent crimes. Just like with other dense cities, Boston also struggles to manage high rime rates and the control of narcotic issues. Effort has been made to face the problem, but it has not been substantial enough and with the full proper utilisation of the Boston Police Force, there can be significant improvements to the crime rate.

The main aim of this report is to provide suggestions that will contribute towards the improvement of Boston’s Crime rate. To achieve this goal we will complete the following objectives: explore Boston’s crime patterns, analyse and discuss the findings of the patterns gathered through big data and present our findings clearly through the use of various diagrams. The social impact of this will lead to consistent lower crime rates year on year as the methods are more refined and consequently improve the quality of life within Boston City as well as the economic and environmental factors. A brief overview of areas we will be investigating include policy strategies, economic conditions, demographic shifts and changes to law.

# Background of study

Within this report the aim is to use big data analytics to explore crime patterns in Boston to try and reduce crime. The report utilises data from the previous years to analyse crime across different districts that Boston has. Through identifying the trends in data, we aim to target Boston’s districts with tailored strategies to reduce the level of crime and improve the safety and overall quality of life in Boston. [3]

The US has historically had an issue with crime and the management of narcotics during the ‘80s & ‘90s [4], this prompted a country wide crackdown. Unfortunately, like many other states and cities Boston has not improved as well as expected. Crime is a major issue in Boston and it is important to understand the patterns in the crime to develop an effective strategy to  reduce it. Research has shown that while non-violent crimes such as robbery has declined *(-2.2%)* the violent crimes such as homicide *(+2.5%)* and assault *(+7.8%)* has seen an increase year on year [5] and Boston’s crime rate is deemed *“considerably higher than the national average across all communities in America”* by Neighbourhood Scout, which is originally sourced from the Federal Bureau of Investigation (FBI) [6]. The use of big data analytics can help provide an insight into the crime patterns that might not be obvious to the naked eye and allow a targeted and effective approach to reducing crime. By analysing the data, the project aims to identify  the most common crimes, the areas with the highest crime and any other factors that may impact the level of crime in Boston. [7]

The aim of this paper is to use the power of big data to explore crime patterns in Boston and develop a strategy to reduce crime in Boston. The aims are:

* Analyse crime from the data set we have and identify most common crimes.
* Identify the areas with highest crime rates and see if there are other factors at play that may increase the level of crime in Boston.
* Develop a targeted strategy to reduce crime in Boston based on the insights we get from the data analysis.

The contributions of this project include:

* The use of big data analytics to find patterns and trends that might not be obvious.
* Development of targeted strategies to reduce crime based on the insights we gained.
* Identify specific areas and crime types that require more attention from law/ Government.
* Potential approach to be applied in other cities to improve public safety.

The report is organised into several sections to clearly show parts of the report. It will be structured as follows:

* Introduction/Abstract
* Background of study
* Related work
* Results and discussion
* Conclusion
* References

# Related work

There's already been a number of studies that use the power of big data in other cities like New York for example. The studies focused on using big data analytic techniques to find crime patterns and trends to develop targeted strategies to reduce the level of crime. These studies show that the approach of using big data to create a crime predictability through analysing crime patterns within big data is a sustainable method in reducing crime.

A study conducted by Wang et al. (2018) in New York City, big data techniques were used to find where there was a hotspot for crime. The study was a predicted study which means they discovered that, using big data analytics, they could improve the accuracy of crime prediction which could potentially reduce crime rate in high crime areas. [8]

Similarly, a study by Yang et al. (2019) was studied, however this time it was in London with its big problem of knife crime. Once again big data analytics were used to find a data driven approach to improve “crime prevention” and potentially reduce crime rate. [9]

Another study was conducted by Song et al. (2020) In Shanghai, China to analyse crime patterns and trends to develop an application which is a “predictive model” for crime hotspots. The study came to find out that the application created was effective at identifying what's called a High-risk area and was used to guide targeted preventions. [10]

## How our work differs

While these strategic studies were used to decrease crime in other cities, our work differs in different ways. First we pay attention to Boston's crime patterns which, with each city, will have its unique characteristic challenges. Secondly we will use a combination of statistics and visualisation techniques to try and identify the crime pattern which will provide a better understanding in the data too. Finally we will try to come up with a strategy to expectantly reduce crime, help local law enforcement and organisations in Boston by implementing interventions based on the insights we gain from data analysis.

# Methodology

The data set used in this report was provided by the Boston Police Department’s Crime Incident Reports database via the site Kaggle.com. This dataset holds records of information on criminal incidents reported to the police during the years of 2015 – 2022. The data was analysed using descriptive statistics such as mean, median, and standard deviation, as well as graphical representations such as charts and maps. The limitations of the data include underreporting of some crimes, incomplete or inaccurate data, and potential biases in police reporting and enforcement.

## Block Diagram

Diagram

Description automatically generated

**Step 1** : To begin an investigation we must first acquire an appropriate dataset that would be suitable for analysis in regard to our area of study within this report. In this case we are investigating Boston’s crime rate so we sourced a dataset via Kaggle that is titled “Boston Crime Dataset 2022”.

**Step 2** : We must now examine the data set and consider which parts of the data is suitable to our investigation in order to start the cleaning process. Firstly, we removed any incomplete rows that were within the dataset so that all remaining data was thorough and reliable. This is done so when the data is used during the study, there is no incomplete outputs or anomaly results which can influence the outcome of the analysis.

**Step 3** : Just like with the previous step with the rows, we must now apply the cleaning process to any columns within the data. The dataset contained a column defined as *“OFFENCE CODE”.* These codes were not valuable to the study as they categorised crimes by police code and there was already another clearer crime identifier column called *“OFFENCE DESCRIPTION”* which better explained crime types and would provide more readable results to the public through analysis. Another column we decided to remove was the *“SHOOTING”* column as it gave too many inconsistencies when providing data, additionally any shootings were already categorised within the offence description column previously mentioned.

**Step 4** :  Once we have completed any bigger modifications to the dataset we must now begin refining the cleaning process. We used a function to remove any rows that contained data that was dated before 2018, the purpose of this is to achieve more relevant and accurate results within our findings of any crime patterns.

**Step 5** : Now that the csv file has been cleaned we can begin to analyse the dataset by using python libraries such as Pandas and Matplotlib to provide us with insightful information on Boston’s crime rate and patterns of crime. The process was to utilise Pandas library to clean the data and then use Matplotlib to analyse the data through bar and pie charts as well as tools such as Google My Maps to geographically present our findings to visualise the dataset so we can discover patterns of crime and begin to answer questions of why certain crimes were more dominant in different districts of Boston.

**Step 6** : Finally, we must now examine the data that is cleaned, observe the charts and graphs we have created and now report on any patterns we have discovered in order to aid our report. Now we can begin to make actions towards providing assistance and suggestions to the Boston Police Department through the data provided and present it to them in a professional and understandable manner. We can provide an insight into which districts are affected by which type of crime and the approaches they could take.

# Result and Discussion

## Experimental setup

To ensure we gain the correct data for analysis we must apply the correct methods or else the findings will be inaccurate. Acting on the findings and presenting it appropriately is the most important step in big data analytics because if the study is not understandable then it's not able to communicate any clear points to the intended audience.

## Discussion of the findings

Now the data has been refined and cleansed we can begin to discuss and present the findings that have been discovered within the dataset. Through analysis we can reveal some insights that could be crucial to the leading factor that causes crime within Boston and its districts

The next step is to visualise the data, this step is crucial as it is the foundation of presenting our findings in a clear and concise manner. The more we can visualise the better our assistance can be for the Boston Police Department.

Chart, pie chart

Description automatically generatedThe following is a pie chart listing the Boston crime rate percentages for each crime reported within the dataset from the years 2018-2022. As shown in Figure 1, the *‘investigative person’* and *‘verbal disputes’* crimes are the two leading crimes. With investigative person having 6.7% of the total crimes reported and verbal disputes having 6.4%.

Fig. 1.

In Figure 2, we can observe that the district B2, which is located within the centre of Boston City has the highest contribution to crime in Boston with a total of 15.8%. The second and third highest contributors to Boston’s total crime is district C11 with 13.7% and district D4 with 12.2%. These districts hold significantly higher crime reports than districts A7 (4.1%) and district E5 (4.6%). This graph highlights the key districts the Boston Police Department should turn their attention to.

Chart, pie chart

Description automatically generated

Fig. 2.

In Figure 3, we can see that throughout the year the reported crime *‘investigative person’* is the most prominent through the year with it being the highest crime for 6 total months of the year. Additionally we can identify that during the summer months of Boston (months 5-7) the crime is noticeably higher than the rest of the year, especially compared to the winter months (months 10-12).

Chart, bar chart

Description automatically generated

Fig. 3.

In Figure 4 we can see a clear growth of crime and a clear pattern of crime. Through 2018-01 to 2019-01 we can see a clear spike in crime within the summer months similar to Figure 3. Once again, through 2019-01 to 2020-01 there is the same pattern of an increase of crime in the earlier months which hits a peak in July/August then a decline in crime as it heads towards the winter months. With the highest spike reaching higher crime levels year on year (8300 incidents to 8500 incidents).

Chart, line chart

Description automatically generated

Fig. 4.

# Analysis of the findings

Now we have observed the presented data and discussed the findings such as the district specific crime percentages, the crimes most dominant through Boston and the months that show elevated levels of crime, we can now dive deeper into the statistics and start to present methods and suggestions to the Boston Police Department to aid their enforcement. Figure 5 is a map that illustrates the geographical mapping of Boston’s 12 total districts and will be referred to within analysis.

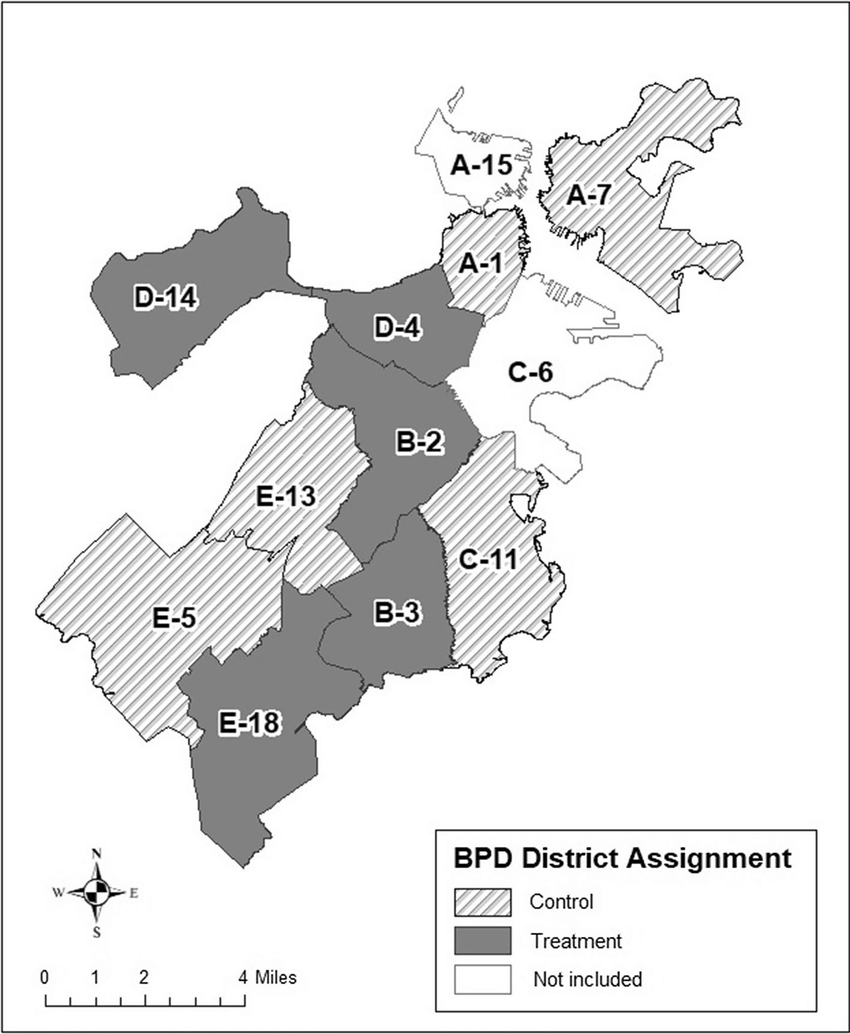


Fig.5.

Fig. 1: Within Figure 1, we identified verbal disputes as the highest reported crime. Although the context of the verbal disputes isn’t included actions can still be taken, stronger legislations can be governed to create a deterrence to antisocial behaviour. This is then down to the Boston Police Department to enforce the updated laws and to sanction individuals through tougher fines which can then be directed back into the police to create additional funding and therefore enable higher productivity. Repeat offenders should either receive a higher fine or be sentenced to create stronger police presence. As well as this, officers should also make a passive effort to increase good will within the community, this can build sustainable relationships with residents and can deescalate disputes and is historically shown to make a difference to crime. [11]

Fig. 2: Within Figure 2 we observed that the highest rates of crime were found within districts B2, C11 and D4. These 3 districts have noticeably higher cases of crime as apposed to the lower districts such as A7 and E5. If we now use Figure 5 we can associate higher levels of crime with the city centre and lower levels of crime with the outskirts of the city that are more rural. Statistically, city centres will have higher crime rates due to the population density but also because there are more opportunities for criminals to operate and utilise transport links to remain virtually anonymous [12]. Now we have identified which crimes appear the most and where the crimes are appearing the most, we can now evidently say that the attention of Boston Police Department must be focused on the city centre. The force needs to delegate sufficient resources such as officers to either patrol the streets more regularly or to have more officers based in central so the police presence is more dominant, this consequently results in a deterrence in crime, this information can play a crucial role for law enforcement agencies to get a better understanding of the hotspots of crime.

Fig. 3: Within Figure 3 we discovered that the *‘investigative persons’* crime was the most prominent through the year. We also seen that there is a spike of this particular crime during the summer months, this is a typical correlation that can be observed when investigating crimes within the summer [13]. Whilst we know the police department cannot control the weather, we can definitely show that this is a typical pattern of a spike in crime they should learn to expect and should prepare for. I would advise that the Boston police once again have more regular patrols to create a stronger police presence. Moreover the governors of Boston could introduce a city-wide curfew within summer months to keep the exposure of potential crime to a minimum. This in-turn keeps people off the streets and will have a substantial impact on the level of crime experienced within the months of June-August. Additionally, as *‘investigative persons’* is the highest crime reported, this means that the police still need to carry out further investigations into the crime as it is still a vague crime description. Boston police should aim to attend crime scenes at a quicker pace to gain a timely result, and if not they should create incentives for civilians to cooperate with the police in order to get the crimes resolved.

Fig. 4: Similarly to Figure 3, Figure 4 shows a similar pattern of crime where during the summer months there is an increase of crime and this pattern is repeated year on year from 2018 through to 2020 and it shows a clear increase each year. Interestingly, the graph shows that the crime decreased significantly during the start of 2020. This is due to the global COVID-19 pandemic; however this supports the previous suggestion observed from Figure 3 that potential curfews could provide a decrease in crime. Curfews are utilised based upon Opportunity Theory - curfews reduce the opportunities that crimes can be committed in and is deemed to have a very strong impact on crime [14].

Finally, the following figures indicate a clear pattern of crime over the course of 3 years (2018-2020). The graphs illustrate the crimes committed (x-axis) against the number of crimes per crime (y-axis). There is a graph for each year to show the direct correlation. Fig. 6 (2018), Fig. 7 (2019) and Fig. 8 (2020).

Chart, bar chart

Description automatically generated

Fig. 6.

Chart, bar chart

Description automatically generated

Fig. 7.

Chart, bar chart, histogram

Description automatically generated

Fig. 8.

# Conclusion

The analysis of the crime reports completed within this report indicate clear patterns of crime and highlight the antisocial behaviours within Boston City. Furthermore, this study identifies the effectiveness of big data to investigate city-wide crime issues as well as the attention it can drive towards the specific areas the crime is most prominent and the patterns that are observed within crime data. These results should be used by the Boston Police Department to create tailored strategies to tackling crime in order to overtime build up community reassurance in the police to create social sustainability. [15]

This analysis provides clear evidence of crime patterns generated through the Boston Crime Dataset 2022 and allows predictions to made in order to arrange preventative measures. In particular we highlight through illustrations the key districts that are suffering with crime rates. This report should be studied and developed further going forward to create a sustainable police force within Boston city.

# References

|  |  |
| --- | --- |
| [1] | Boston Municipal Research Bureau, “About Boston,” 2019. [Online]. Available: https://www.bmrb.org/about-boston/#:~:text=The%20city%20is%20the%20anchor,metropolitan%20area%20in%20the%20country.. |
| [2] | World Popuation Review, “Boston, Massachusetts Population 2023,” 2023. [Online]. Available: https://worldpopulationreview.com/us-cities/boston-ma-population. [Accessed 1 January 2023]. |
| [3] | V. S. H. T. K. B. a. J. S. A. J. Park, “Discovering Crime Trends and Patterns Using Three-Dimensional Visual Analysis,” *Proceedings of the 2019 IEEE 10th Annual Informtation Technology, Electronics and Mobile Communication Conference (IEMCON),* no. 10, pp. 545-549, 2019. |
| [4] | J. Robison, “Decades of Drug Use: The '80s and '90s,” Gallup, 9 July 2002. [Online]. Available: https://news.gallup.com/poll/6352/decades-drug-use-80s-90s.aspx. |
| [5] | F. McColgan, “Boston’s overall crime rate is down 1.5% in 2022, but fatal shootings rose by 8 over 2021,” Boston Herald, 3 January 2023. [Online]. Available: https://www.bostonherald.com/2023/01/03/bostons-overall-crime-rate-is-down-1-5-in-2022-but-fatal-shootings-rose-by-8-over-2021/. [Accessed 6 January 2023]. |
| [6] | Neighbourhood Scout, “Boston, MA Crime Rates,” Neighbourhood Scout, October 2022. [Online]. Available: https://www.neighborhoodscout.com/ma/boston/crime. |
| [7] | e. a. A. Almoqbil, “The correlation between sustance abuse and crime in the United States,” *Proceedings of the 2020 IEEE International Conference on Big Data,* pp. 5625-5626, 2020. |
| [8] | X. L. a. L. W. Y. Wang, “Predicting crime hotspot with big data analytics: A case study in New York City,,” *Comput Environ. Urban Syst. ,* vol. 70, pp. 97-107, 2018. |
| [9] | Y. L. X. L. Y. Z. a. Y. W. Z. Yang, “A framework for spatiotemporal crime pattern analysis using big data analytics: A study of London,” *PLoS One,* vol. 14, no. 7, 2019. |
| [10] | Q. Z. T. C. J. C. H. L. a. H. Z. X. Song, “Crime Hotspot Prediction with a fusion model based on multi-source data,” *ISPRS Int. J. Geo-Inf,* vol. 9, no. 2, p. 99, 2020. |
| [11] | College of Policing, “What works in policing to reduce crime?,” 1 July 2021. [Online]. Available: https://www.college.police.uk/research/what-works-policing-reduce-crime#:~:text=As%20well%20as%20preventing%20crime,make%20a%20difference%20to%20crime.. |
| [12] | M. Oliveira, “More crime in cities? On the scaling laws of crime and the inadequacy of per capita rankings—a cross-country study,” Crime Science, USA, 2021. |
| [13] | I. Ahmed, “Why US gun violence spikes in warm weather,” 4 June 2022. [Online]. Available: https://phys.org/news/2022-06-gun-violence-spikes-weather.html. |
| [14] | College of Policing, “Youth Curfews,” 7 May 2019. [Online]. Available: https://www.college.police.uk/research/crime-reduction-toolkit/youth-curfews#:~:text=Opportunity%20theory%20%E2%80%93%20curfews%20reduce%20the,a%20deterrent%20to%20criminal%20behaviour.. |
| [15] | S. C. a. V. T. R. A. V. Kumar, “Crime Data Analysis usinf Big Data Analytics and Visualizastions using Tableau,” *Int. Conf. Electron Commun. Aerospace Technology ICECA,* no. 6, pp. 627-632, 2022. |
| [16] |  |
| [17] | [Online]. |

# Code Appendix (Part B)

**Description of what the code does**

Code:

**Complete code that loads the data from the boston\_crimerate.csv file, removes the rows with missing values, and saves the cleaned data back to the same file:**

*import pandas as pd*

*# Loads the data from csv excel file*

*df = pd.read\_csv("boston\_crimerate.csv")*

*# removes rows with any missing values*

*df.dropna(inplace=True)*

*# this line saves the file for any changes just made*

*df.to\_csv("boston\_crimerate.csv", index=False)*

**This code deleted all rows that have missing values and saved the file, went from 50 000 rows to just 1665.**

**Have to have pandas library to work, also excel file has to be in csv format which it is.**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**THIS CODE REMOVED THE 3 COLUMNS THAT WE DIDN'T WANT. THE UPDATED EXCEL SHEET IS CALLED boston\_crimerate WHICH IS UPLOADED IN DATASET FOLDER.**

import pandas as pd

# Loads data from file

df = pd.read\_csv("boston\_crimerate.csv")

# Removes three columns

df.drop(columns=["INCIDENT\_NUMBER", "OFFENSE\_CODE", "SHOOTING"], inplace=True)

# Save the file again

df.to\_csv("boston\_crimerate.csv", index=False)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**THIS CODE BELLOW DELETED COLUMNS UCR\_PART AND OFFENCE\_CODE\_GROUP, updated version called boston\_crimerate\_modified**

import pandas as pd

# Loads the file

df = pd.read\_csv("boston\_crimerate.csv")

# Drops columns UCR\_PART and OFFENSE\_CODE\_GROUP

df = df.drop(columns=["UCR\_PART", "OFFENSE\_CODE\_GROUP"])

# Saves the file with new changes

df.to\_csv("boston\_crimerate\_modified.csv", index=False)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Delete rows that date 2017, 2016, 2015.**

import pandas as pd

df = pd.read\_csv("boston\_crimerate\_modified.csv")

df = df[~df["YEAR"].isin([2015, 2016, 2017])]

df.to\_csv("boston\_crimerate\_modified.csv", index=False)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Delete rows with missing information. File went from 246,706 rows to 225,372.**

import pandas as pd

# reads the CSV file

df = pd.read\_csv('boston\_crimerate\_modified.csv')

# removes rows with any missing values

df = df.dropna()

# saves file with  cleaned data back

df.to\_csv('boston\_crimerate\_modified.csv', index=False)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**This code deletes all years but the one chosen and deletes all rows except 10 000 to map geographical.**

import pandas as pd

# Reads the CSV file

df = pd.read\_csv('boston\_crimerate\_modified\_newest\_version.csv')

# Keeps only rows where the year is 2018

df = df[df['YEAR'] == 2018]

# Keeps random 10,000 rows

df = df.head(10000)

# Saves  modified CSV file

df.to\_csv('boston\_crimerate\_modified\_2018.csv', index=False)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**This code  sums up  in python which district has the most crimes which the result was this.**

import pandas as pd

# load data

df = pd.read\_csv('boston\_crimerate\_modified\_newest\_version.csv')

# uses value\_counts to count number of crimes in each district

district\_counts = df['DISTRICT'].value\_counts()

# displays the result

print(district\_counts)

**Outcome:**

Graphical user interface

Description automatically generated with medium confidence

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**This code analysed the street occurrences which produced this result:**

import pandas as pd

# loads the dataset

df = pd.read\_csv('boston\_crimerate\_modified\_newest\_version.csv')

# use the value\_counts method to sum every crime number in every street.

street\_counts = df['STREET'].value\_counts()

# displays the result

print(street\_counts)

**Outcome:**

Text

Description automatically generated with medium confidence

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**This counts all offence descriptions and totals it up to each crime committed.**

import pandas as pd

# load data

df = pd.read\_csv('boston\_crimerate\_modified\_newest\_version.csv')

# use the value\_counts method to count the number of occurrences of each offence

offense\_counts = df['OFFENSE\_DESCRIPTION'].value\_counts()

# display the result

print(offense\_counts)

**Outcome:**

Graphical user interface, text

Description automatically generated

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**This code uses matplotlib to create a bar chart**

import pandas as pd

import matplotlib.pyplot as plt

# Load data from file

df = pd.read\_csv('boston\_crimerate\_modified\_newest\_version.csv')

# Create plot

plt.figure(figsize=(10, 6))

df['OFFENSE\_DESCRIPTION'].value\_counts().plot(kind='bar', fontsize=5)

# Adds labels and titles for chart to make sense

plt.xlabel('Offense Description', fontsize=12)

plt.ylabel('Count', fontsize=12)

plt.title('Crime Distribution in Boston', fontsize=14)

# Shows chart

plt.show()

**Outcome:**

Chart, bar chart

Description automatically generated

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Pie chart that shows crime rate of each crime in percentage.**

import pandas as pd

import matplotlib.pyplot as plt

# Load data from file

df = pd.read\_csv('boston\_crimerate\_modified\_newest\_version.csv')

# Get values for column

counts = df['OFFENSE\_DESCRIPTION'].value\_counts()

# Creates pie chart

fig, ax = plt.subplots(figsize=(12, 12))

ax.pie(counts.values, labels=counts.index, startangle=90, autopct='%1.1f%%', textprops={'fontsize': 8})

# Adds a title

ax.set\_title('Boston Crime Rate by Offense Description', fontsize=12)

# Moves the legend to the left

ax.legend(loc='center left', bbox\_to\_anchor=(-0.7, 0.5), fontsize=8)

# Show the plot

plt.show()

**Outcome:**

Chart, pie chart

Description automatically generated

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Pie chart for districts**

import pandas as pd

import matplotlib.pyplot as plt

# Loads data from file

df = pd.read\_csv('boston\_crimerate\_modified\_newest\_version.csv')

counts = df['DISTRICT'].value\_counts()

# Create a pie chart  for every day of the week

fig, ax = plt.subplots(figsize=(10, 10))

ax.pie(counts.values, labels=counts.index, startangle=90, autopct='%1.1f%%', textprops={'fontsize': 8})

ax.set\_title('Crime reports in each district', fontsize=14)

ax.legend(loc='center left', bbox\_to\_anchor=(-0.4, 0.5), fontsize=10)

plt.show()

**Outcome:**

Chart, pie chart

Description automatically generated

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**This code gives a linear graph and a correlation matrix that shows the crimes in all of of the months in the dataset.**

import pandas as pd

import matplotlib.pyplot as plt

# Load the data file

df = pd.read\_csv('boston\_crimerate\_modified\_newest\_version.csv')

# Creates column 'DATE' combining YEAR and MONTH columns

df['DATE'] = pd.to\_datetime(df[['YEAR', 'MONTH']].assign(DAY=1))

# Groups the data by month and count the number of crimes per month

monthly\_crime\_count = df.groupby('DATE')['OFFENSE\_DESCRIPTION'].count()

# Plot the crime count by month

plt.plot(monthly\_crime\_count)

plt.title('Monthly Crime Count')

plt.xlabel('Date')

plt.ylabel('Number of Crimes')

plt.show()

subset\_df = df[['YEAR', 'MONTH', 'OFFENSE\_DESCRIPTION']]

corr\_matrix = subset\_df.corr()

fig, ax = plt.subplots(figsize=(10, 8))

im = ax.imshow(corr\_matrix, cmap='coolwarm')

for i in range(corr\_matrix.shape[0]):

    for j in range(corr\_matrix.shape[1]):

        ax.text(j, i, '{:.2f}'.format(corr\_matrix.iloc[i, j]), ha='center', va='center', fontsize=8)

ax.set\_xticks(range(corr\_matrix.shape[0]))

ax.set\_xticklabels(corr\_matrix.columns)

ax.set\_yticks(range(corr\_matrix.shape[1]))

ax.set\_yticklabels(corr\_matrix.columns)

plt.colorbar(im)

plt.title('Correlation Matrix')

plt.show()

**Outcome:**

Chart, line chart

Description automatically generated

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**This code gives us top 20 crimes committed  for each year (2018)**

import pandas as pd

import matplotlib.pyplot as plt

# Load data from Excel file

df = pd.read\_csv('boston\_crimerate\_modified\_newest\_version.csv')

# Filter data for 2018 only

df = df[df['YEAR'] == 2018]

# Get the top 20 crime types

top\_crimes = df['OFFENSE\_DESCRIPTION'].value\_counts().head(20)

# Create a bar chart of the top 20 crime types

plt.figure(figsize=(10,8))

plt.bar(top\_crimes.index, top\_crimes.values)

plt.xticks(rotation=90)

plt.xlabel('Crime Type')

plt.ylabel('Number of Crimes')

plt.title('Top 20 Crime Types in Boston (2018)')

plt.show()

**Outcome:**

Chart, bar chart

Description automatically generated

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**This code gives us top 20 crimes committed  for each year (2019)**

import pandas as pd

import matplotlib.pyplot as plt

# Load data from Excel file

df = pd.read\_csv('boston\_crimerate\_modified\_newest\_version.csv')

# Filter data for 2019 only

df = df[df['YEAR'] == 2019]

# Get the top 20 crime types

top\_crimes = df['OFFENSE\_DESCRIPTION'].value\_counts().head(20)

# Create a bar chart of the top 20 crime types

plt.figure(figsize=(10,8))

plt.bar(top\_crimes.index, top\_crimes.values)

plt.xticks(rotation=90)

plt.xlabel('Crime Type')

plt.ylabel('Number of Crimes')

plt.title('Top 20 Crime Types in Boston (2019)')

plt.show()

**Outcome:**

Chart, bar chart

Description automatically generated

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**This code gives us top 20 crimes committed  for each year (2020).**

import pandas as pd

import matplotlib.pyplot as plt

# Load data from Excel file

df = pd.read\_csv('boston\_crimerate\_modified\_newest\_version.csv')

# Filter data for 2020 only

df = df[df['YEAR'] == 2020]

# Get the top 20 crime types

top\_crimes = df['OFFENSE\_DESCRIPTION'].value\_counts().head(20)

# Create a bar chart of the top 20 crime types

plt.figure(figsize=(10,8))

plt.bar(top\_crimes.index, top\_crimes.values)

plt.xticks(rotation=90)

plt.xlabel('Crime Type')

plt.ylabel('Number of Crimes')

plt.title('Top 20 Crime Types in Boston (2020)')

plt.show()

**Outcome:**

Chart, bar chart, histogram

Description automatically generated

**Block phase diagram created in PowerPoint:**

Diagram

Description automatically generated

**Tableau - used to see dataset geographically:**

Map

Description automatically generated