INDUSTRY RESEARCH PROJECT

CHICAGO CRIME ANALYTICS

Author: Syafiq Azman

Table of Contents

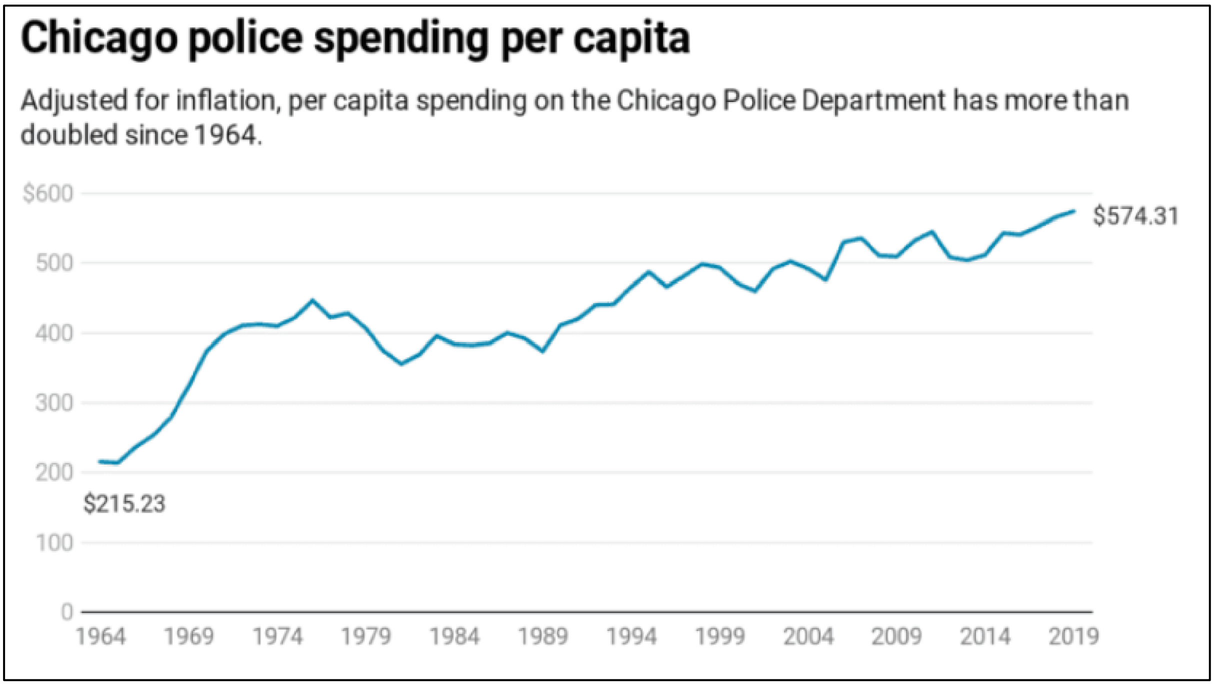
1. Executive Summary

The city of Chicago is known for its notoriously high crime, even in the country where crimes seem to be growing rampant. Despite massive investments, new legislation and harder laws from local and central governments, crimes, in general, are spreading like wildfires in the United States and Chicago especially. The research project aims to tackle the main business problem of reducing the crime rate in the city of Chicago using data analytics. The methodologies used include both descriptive and predictive analytics. The main tools used in the research include Exploratory software, Excel and Power BI. This paper mainly addresses how the Chicago Police Department (CPD) can utilise advanced analytics to reduce crime and protect its community. Based on the analytics insights gathered, CPD is recommended to increase foot patrol in crime hotspot areas, promote smart street campaigns and distribute resources effectively for high-profile crimes reported.

2. Industry Problem

In 2016, Time magazine released a headline titled, "Chicago is Responsible for Almost Half of the Increase in U.S Homicides". The article highlights the rise in crimes across the United States, especially Chicago as a leading contributor to murders and killings, with 500 murders in 2016 alone when it was published (Sanburn 2016). In 2021, almost 6 years apart, Chicago Sun-Times published an article titled "Downtown shootings up 220%, biggest spike in city: People are fed up". Crime rise is also evident in the increase in car thefts and sexual assaults, among other crimes growing rampant (Schuba et al. 2021). Many more articles, magazines, newspapers and media outlets report the same issue of what seems to be inevitable crimes that keep repeating over the years. In 2019, only 21% of arrests were made on the total of 260,899 crimes reported and less than 12% for 203,530 crimes reported in 2021, despite a generous budget allocation provided by the local government of $US 1.6 billion in 2020 and a steady increase in police spending per capita, referring to Figure 1. However, crimes continue to grow, and the number of arrests drops (Daniel 2021; Ballesteros 2020). The Chicago Police Department has the means and resources to make the city safer for its people, but every year, promises fail, precious lives are lost, and communities are scared (Hampton 2021; Molina 2022).

Figure 1: Chicago Police spending per Capita from 1964 to 2019



Source: Ballestros 2020

The business problem this research will focus on is "How can we predict crime arrests with high accuracy and reduce the crime rate in Chicago?" The data will be taken from the Chicago Data Portal, which provides data and information regarding the crimes reported with crime arrests, indicating how many successful and unsuccessful arrests were made over the years. The research project and business question are crucial to paving the way for governments and companies to use data analytics to create effective policing that can reduce crime and make communities safer. The data captured will provide insights on the type of frequent crimes, the locations and frequency of each crime, and most importantly, the number of criminal arrests for each type of crime. With the available data, the research project can deduct the type of crimes that require more attention and develop specific policies to tackle them using visualisation techniques. Then, machine learning can help model which crimes are more likely to result in arrest based on given parameters, so it is reliable to future data and used as actionable knowledge moving forward.

The data parameters were limited to only a few variables such as locations, type of crimes, arrest rates, date and time. Ideally, the project requires more data variables or parameters as independent variables, such as the criminal's behavioural data, the preferred mode of transportation or type of vehicle, weapons used, gang affiliation and many more. Thus, a better understanding of criminal patterns is achievable, making any recommendations far more powerful and accurate (Gray Matter 2022).

3. Data Processing and Management

The data is sourced from the City of Chicago Data Portal, an official website of all statistics and data-related information for the city of Chicago. The dataset can be found under the public safety tab in the data catalogue, sorted by each year as early as 2001 (Chicago Data Portal 2022). The dataset contains 22 columns, as shown in Table 1 below, and 1,048,526 rows which are crucial to understanding the crime patterns over the years and addressing the business problem. For instance, the Chicago Police can use the insight from the data to predict future crimes with low arrest rates and develop effective strategies to combat each specific crime as it occurs in real-time. Figure 1 also contains the description and data types for each variable.

After data was acquired, the data source was verified for credibility and consistency. Then, the dataset was cleansed to identify and remove missing values, blank cells, date and time inconsistencies and duplicate data. Data cleansing is an integral process as it helps confirm the data validity, accuracy, reliability, completeness and consistency to generate accurate reporting (Mittal 2021). The final data count for analysis was at 976,087 after incomplete or inconsistent data were removed.

Table 1: Dataset Column Name, Description and Type

Columns in this Dataset

Column Name	Description	Type
ID	Unique identifier for the record.	Number #
Case Number	The Chicago Police Department RD Number (Records Divisi...	Plain Text T
Date	Date when the incident occurred. this is sometimes a best ...	Date & Time
Block	The partially redacted address where the incident occurred...	Plain Text T
IUCR	The Illinois Unifrom Crime Reporting code. This is directly li...	Plain Text T
Primary Type	The primary description of the IUCR code.	Plain Text T
Description	The secondary description of the IUCR code, a subcategory...	Plain Text T
Location Description	Description of the location where the incident occurred.	Plain Text T
Arrest	Indicates whether an arrest was made.	Checkbox ✔
Domestic	Indicates whether the incident was domestic-related as de...	Checkbox ✔
Beat	Indicates the beat where the incident occurred. A beat is th...	Plain Text T
District	Indicates the police district where the incident occurred. Se...	Plain Text T
Ward	The ward (City Council district) where the incident occurred...	Number #
Community Area	Indicates the community area where the incident occurred....	Plain Text T
FBI Code	Indicates the crime classification as outlined in the FBI's Na...	Plain Text T
X Coordinate	The x coordinate of the location where the incident occurre...	Number #
Y Coordinate	The y coordinate of the location where the incident occurre...	Number #
Year	Year the incident occurred.	Number #
Updated On	Date and time the record was last updated.	Date & Time
Latitude	The latitude of the location where the incident occurred. T...	Number #
Longitude	The longitude of the location where the incident occurred. ...	Number #
Location	The location where the incident occurred in a format that a...	Location

Source: Chicago Data Portal 2022

The research project mainly uses descriptive and predictive analytics techniques to address and tackle the business problem. For descriptive analytics, dashboards compiling charts, figures, and tables can be used to visualise the historical and present information. For example, a bar chart with years and the highest number of crimes reported can provide a deeper understanding of the city's most frequent type of crimes each year. While predictive analytics, using machine learning techniques, the police can identify the type of crimes reported that will likely result in an arrest and what resources can be better distributed to ensure high success in arrest rates. Both descriptive and predictive analytics techniques will be applicable and critical to solving the business problem.

4. Data Analytics Methodology

The research project uses three main data analytics methodologies, exploratory, descriptive and predictive analytics. First, exploratory analytics looks at the overall statistical observations and summarises the main attributes of the dataset. Excel and Exploratory software are used to achieve this task by looking at essential data characteristics: median, minimum, maximum, range, standard deviation, unique values, blank cells, distributions and outliers. Figures 2 and 3 demonstrate how the exploratory task was achieved. Exploratory allows preliminary observations and hypotheses to be made on the dataset concerning the business problem (Masters in Data Science 2022).

Figure 2: Exploratory Analytics using Exploratory Software for Location, Arrest, Domestic and Beat Variables

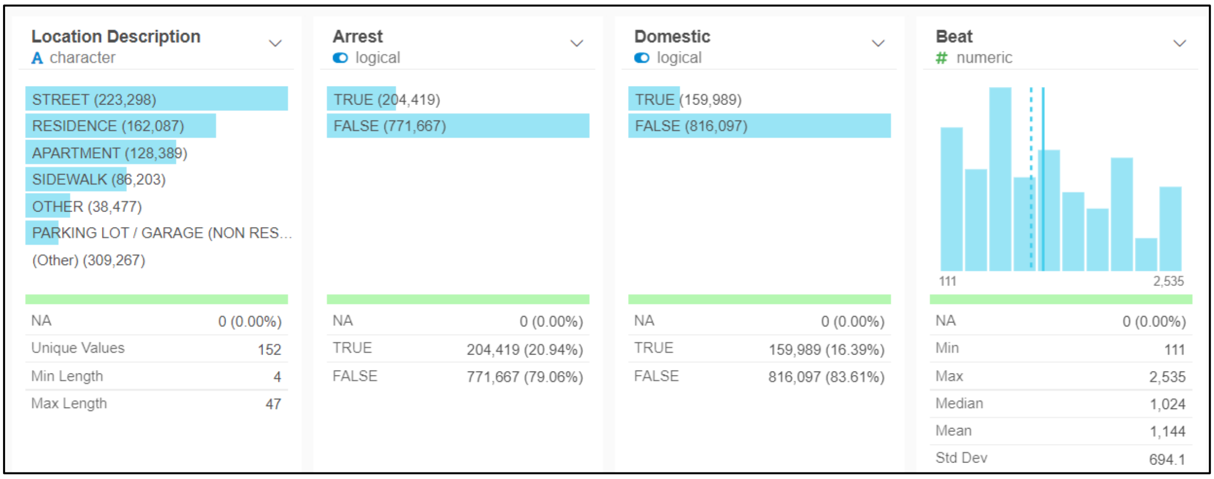
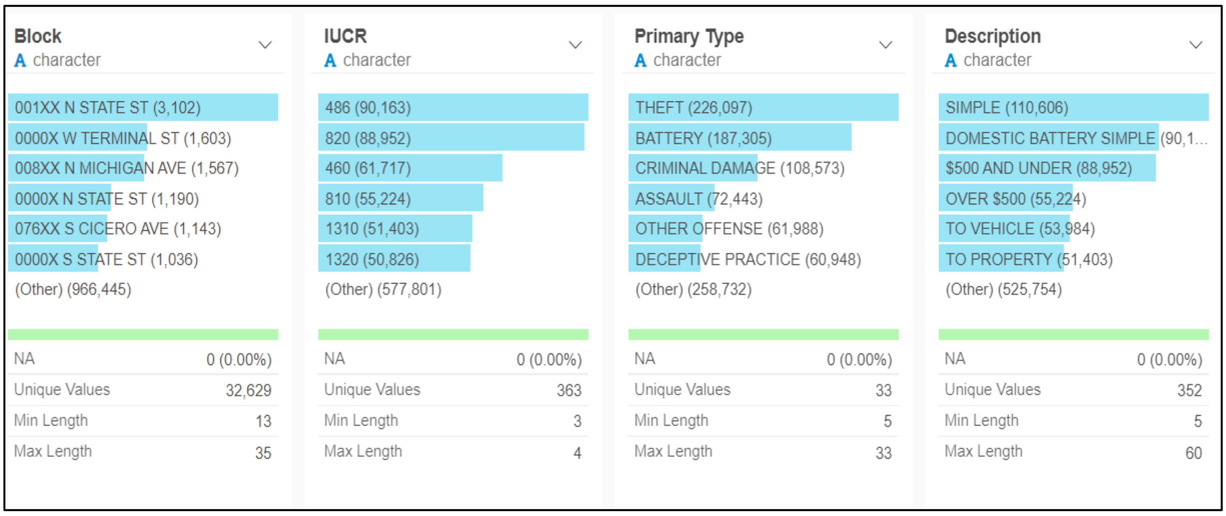


Figure 3: Exploratory Analytics using Exploratory Software for Block, IUCR, Primary Type and Description



Next is using descriptive analytics to gain a deeper data intelligence and information on the crime statistics to answer the business problem. This methodology uses charts and figures through Power BI dashboards to visualise data in a meaningful way to identify key insights. From Figure 4, the heatmap visualisation allows a better understanding of which areas where crimes are happening the most down to their exact location based on the reports made. Figure 5 displays the type of crimes that occur the most from 2001 to 2020.

Figure 4: Number of Crimes Reported (ID) by Location (Latitude and Longitude)

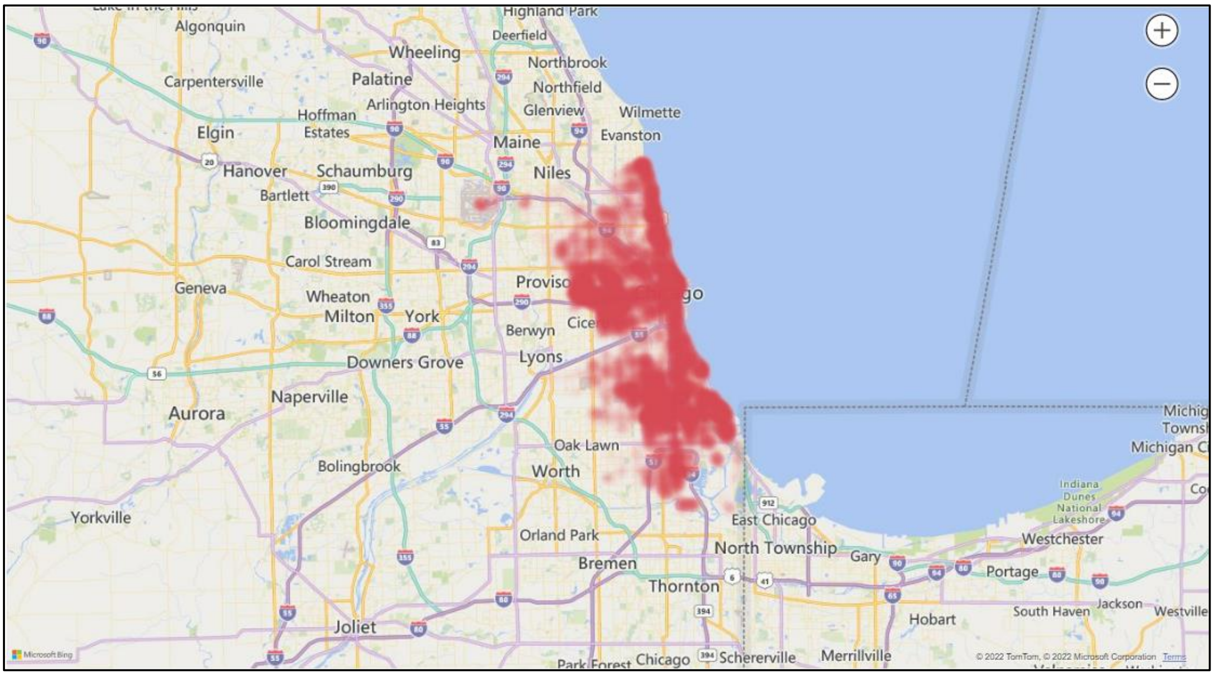
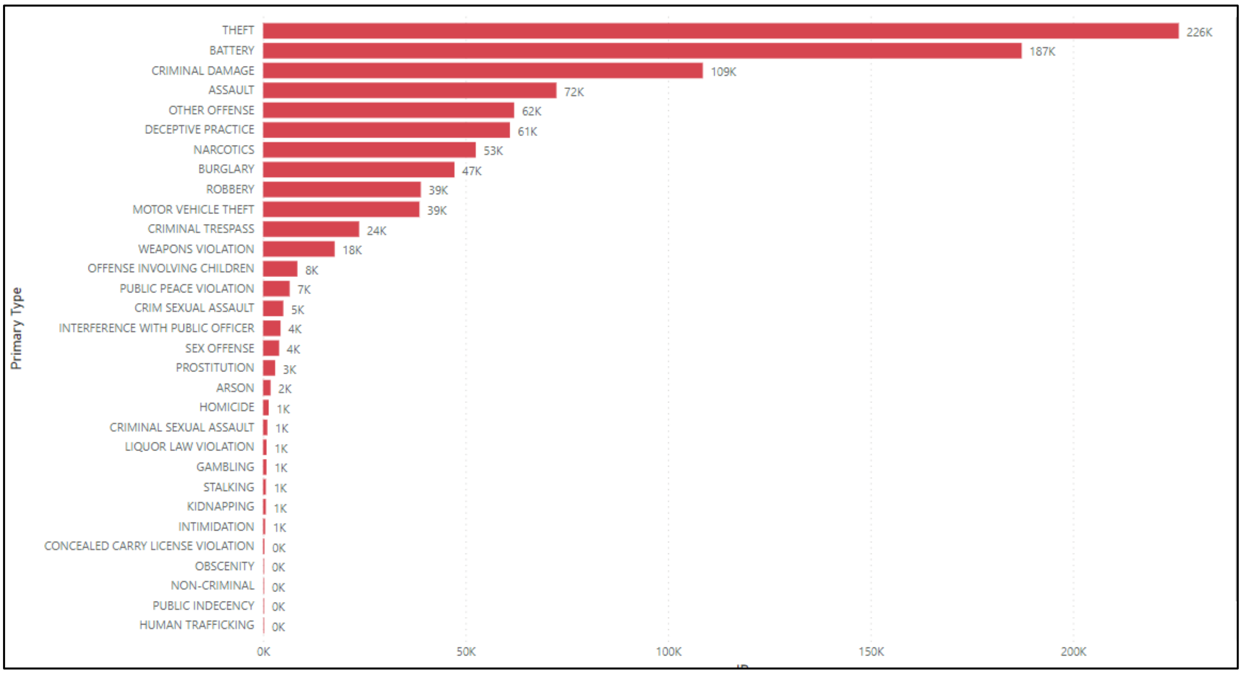


Figure 5: Number of Crime reported (ID) by Primary Type

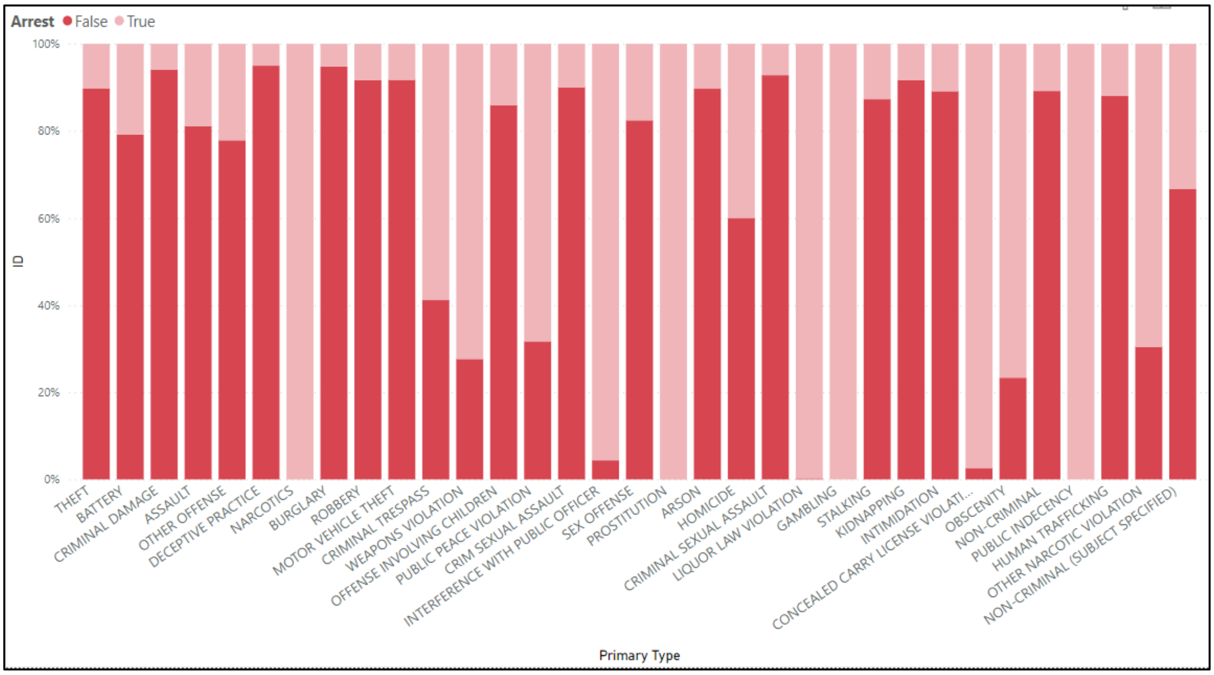


The third and final methodology is predictive analytics using Exploratory software, where predictive models are generated using machine learning techniques, logistic regression and random forest. The predictive modelling was created by looking at the top 5 categories for the primary type: theft, battery, criminal damage, assault and for all, referring to the whole dataset. The target variable is the arrest, while the independent variables are all the other variables but were run two different times; first with all variables, and second with only variables with high importance. The goal of predictive analytics is to find the most effective machine learning models to predict arrests accurately, given the set of variables and replicate the same techniques over new datasets in the future in real-time.

5. Visualisation and Evaluation of Results

First, with descriptive analytics, Figure 6 visualises the arrest rate by the primary type (crime category). The figure shows that the number of false or unsuccessful arrests is high for most crimes, especially for the top 4 categories: theft, battery, criminal damage and assault. The CPD only perform well in narcotics, interference with a public officer, prostitution, gambling, liquor law violation, public indecency, concealed carry case violation and a weapons violation, with only as high as 27% for false arrests. However, the CPD performs poorly in other crime categories, with around 60 to 95% of false arrests recorded. The crime categories with a high number of crimes reported as well as having high false arrests must be put on high alert whenever the crime is reported in real-time.

Figure 6: Total Successful vs Unsuccessful Arrests by Primary Type



Next, Figure 7 reports the changes in the number of crimes reported from 2001 to 2020. As shown, there was a sharp increase for all primary types in 2016, which convincingly aligns with the news report and findings covered earlier in Section 2. Figures 8 and 9 visualise the insight obtained from looking at the contributions of the increase in theft from 2015 to 2016. Figure 8 shows the increase of crimes reported by location, with streets contributing the most, followed by residence and parking lot. Figure 9 shows which subset of theft crimes contribute the most, with theft of $500 over and under in the top 2 followed by building and retail theft.

Figure 7: Crimes Reported (ID) by Year and Primary Type

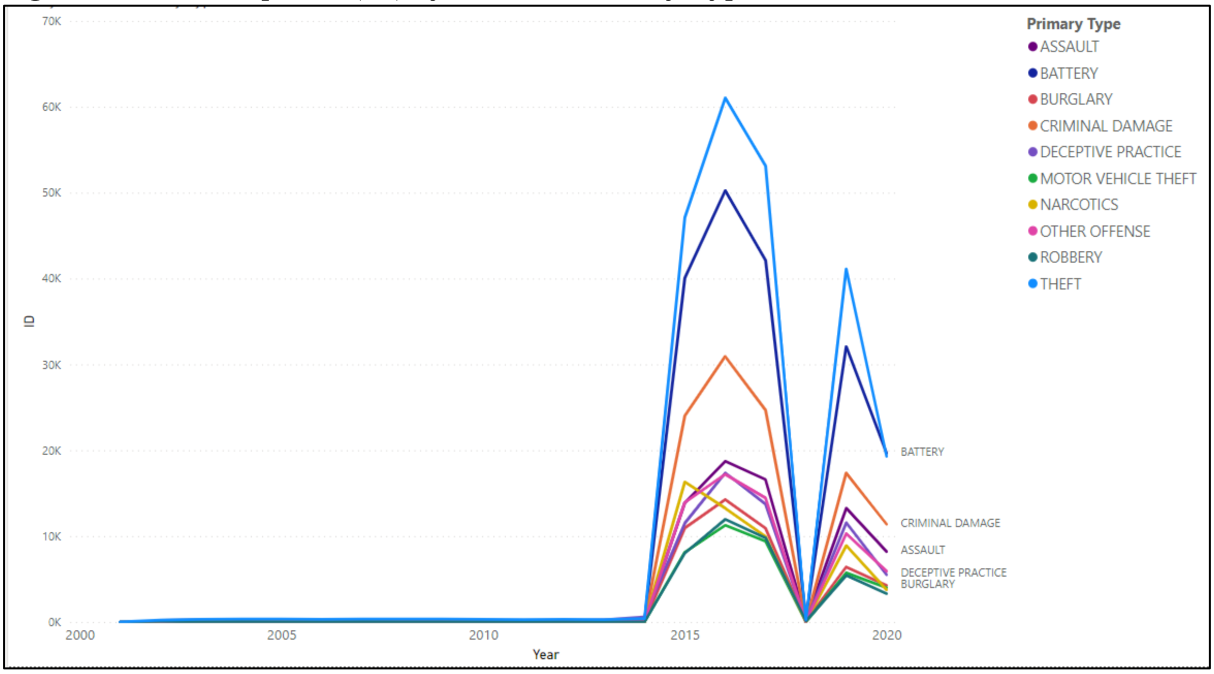


Figure 8: 2015 Vs 2016 Theft by Location Description

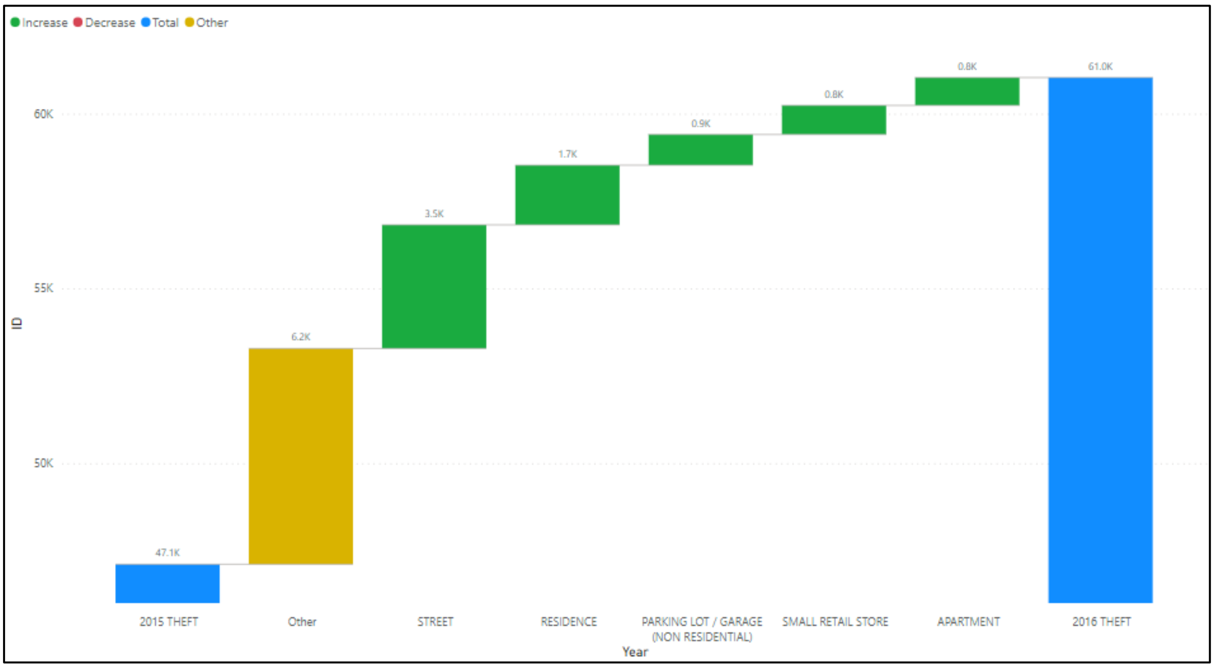
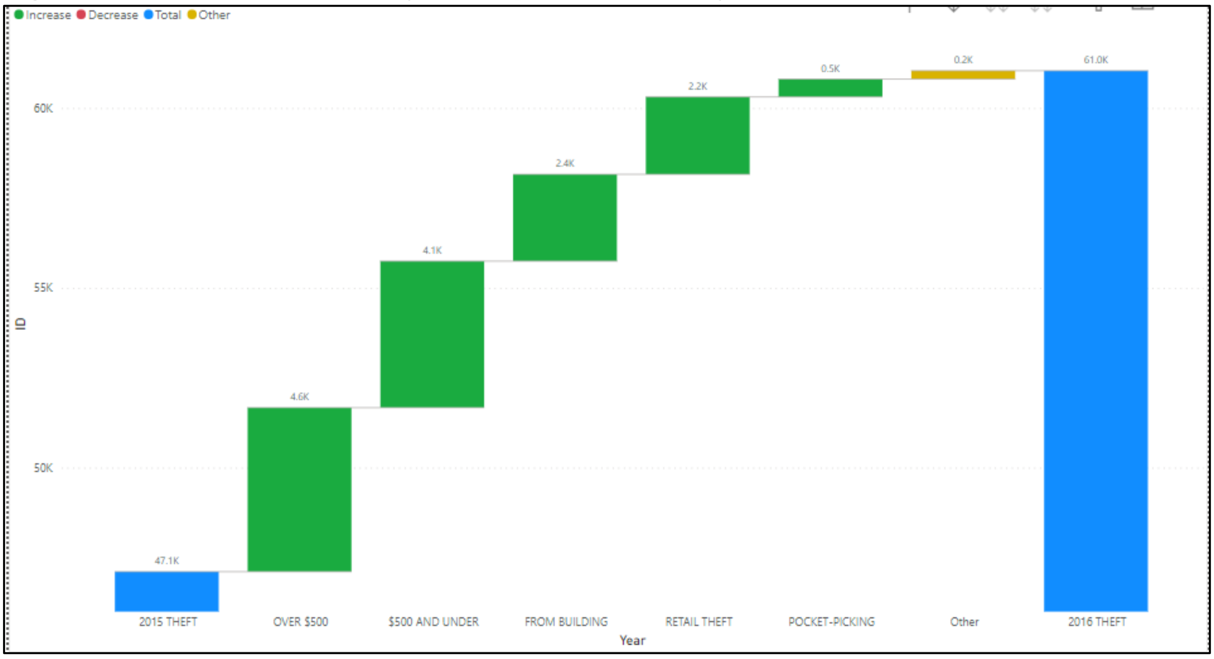


Figure 9: 2015 Vs 2016 Theft by Description (Secondary)



Second, with 5 separate categories, predictive analytics were modelled at 2 different times using all variables and only highly important variables. On top of that, each category was modelled using 2 different algorithms, logistic regression and random forest. All 20 models are then compared by looking at the performance and key indicators; ROC curve, prediction, prediction matrix, accuracy rate, misclassification rate and AUC (area under the curve).

Figures 10 and 11 show the logistic regression and random forest model for all crimes category, with random forest performing slightly better at 87.7% accuracy and lower false positive and false negative rates on the prediction matrix.

Figure 10: Logistic Regression Summary and Prediction Matrix for All Primary Type

Summary Prediction Importance Coefficients Coef. (Significant) Coef. Table Collinearity Pred. Matrix Prot

Data Type	AUC	F Score	Accuracy Rate	Misclassification Rate	Precision	
Training	0.86128400557	0.62287749524	0.86991428571	0.13008571429	0.80445014976	0.5
Test	0.85905777811	0.61902912621	0.8692	0.1308	0.7942202292	0.5

		Predicted	
Data Type	**Actual**	**TRUE**	**FALSE**
Test	TRUE	10.63%	10.33%
	FALSE	2.75%	76.29%
Training	TRUE	10.74%	10.40%
	FALSE	2.61%	76.25%

Figure 11: Random Forest Summary and Prediction Matrix for All Primary Type

Summary Prediction Importance Pred. Matrix Probability ROC Data

Data Type	AUC	F Score	Accuracy Rate	Misclassification Rate	Precision	
Training	0.93485606453	0.67589752094	0.88831428571	0.11168571429	0.86392539212	0.5
Test	0.85851015622	0.63443629879	0.877	0.123	0.81767109295	0.5

		Predicted	
Data Type	Actual	TRUE	FALSE
Test	TRUE	10.67%	9.92%
	FALSE	2.38%	77.03%
Training	TRUE	11.65%	9.33%
	FALSE	1.83%	77.19%

Figure 12 and 13 show the machine learning models for the primary type - theft, with random forest performing slightly better at a 92.3% accuracy rate and lower false-negative rate but the same false-positive rate.

Figure 12: Logistic Regression Summary and Prediction Matrix for Primary Type - Theft

Summary | Prediction | Importance | Coefficients | Coef. (Significant) | Coef. Table | Collinearity | Pred. Matrix | Proba

Data Type	AUC	F Score	Accuracy Rate	Misclassification Rate	Precision	
Training	0.88813799912	0.55879494655	0.92217142857	0.077828571429	0.69055244195	0.46
Test	0.88868064717	0.5418138987	0.9222614841	0.077738515901	0.67055393586	0.45

Summary | Prediction | Importance | Coefficient

		Predicted	
Data Type	**Actual**	**TRUE**	**FALSE**
Test	TRUE	4.60%	5.51%
	FALSE	2.26%	87.63%
Training	TRUE	4.93%	5.57%
	FALSE	2.21%	87.29%

Figure 13: Random Forest Summary and Prediction Matrix for Primary Type - Theft

Summary | Prediction | Importance | Pred. Matrix | Probability | ROC | Data

Data Type	AUC	F Score	Accuracy Rate	Misclassification Rate	Precision	
Training	0.97378581404	0.70422083133	0.94734285714	0.052657142857	0.85870841487	0.59
Test	0.870668032	0.5458135861	0.92332822188	0.076671778119	0.68078817734	0.45

		Predicted	
Data Type	**Actual**	**TRUE**	**FALSE**
Test	TRUE	4.61%	5.51%
	FALSE	2.16%	87.73%
Training	TRUE	6.27%	4.23%
	FALSE	1.03%	88.47%

Figures 14 and 15 show the primary type – battery models, with logistic regression performing slightly better at 80.4% on accuracy rate and lower false negatives by 0.66%. However, the false-positive rate was slightly higher than the random forest.

Figure 14: Logistic Regression Summary and Prediction Matrix for Primary Type – Battery

Summary Prediction Importance Coefficients Coef. (Significant) Coef. Table Collinearity Pred. Matrix Proba

Data Type	AUC	F Score	Accuracy Rate	Misclassification Rate	Precision	
Training	0.61994228827	0.11718252452	0.80197142857	0.19802857143	0.83941605839	0.06
Test	0.6202806141	0.11842500751	0.80446666667	0.19553333333	0.84913793103	0.06

		Predicted	
Data Type	Actual	TRUE	FALSE
Test	TRUE	1.31%	19.32%
	FALSE	0.23%	79.13%
Training	TRUE	1.31%	19.55%
	FALSE	0.25%	78.88%

Figure 15: Random Forest Summary and Prediction Matrix for Primary Type - Battery

Summary Prediction Importance Pred. Matrix Probability ROC Data

Data Type	AUC	F Score	Accuracy Rate	Misclassification Rate	Precision	
Training	0.93402275463	0.27188778492	0.82202857143	0.17797142857	0.92891373802	0.1
Test	0.59178421795	0.13086276781	0.79986666667	0.20013333333	0.6295264624	0.07

		Predicted	
Data Type	Actual	TRUE	FALSE
Test	TRUE	1.51%	19.13%
	FALSE	0.89%	78.48%
Training	TRUE	3.32%	17.54%
	FALSE	0.25%	78.88%

Figures 16 and 17 show the primary type – criminal damage models, with almost identical accuracy rates that only favour logistic regression marginally, and also identical for both false positive and negative rates at 5.81%

Figure 16: Logistic Regression Summary and Prediction Matrix for Primary Type – Criminal Damage

Summary Prediction Importance Coefficients Coef. (Significant) Coef. Table Collinearity Pred. Matrix Proba

Data Type	AUC	F Score	Accuracy Rate	Misclassification Rate	Precision	
Training	0.71372641714	0.0009842519685	0.942	0.058	0.25	0.0004
Test	0.72192047217		0.94193333333	0.058066666667		

		Predicted	
Data Type	**Actual**	**TRUE**	**FALSE**
Test	TRUE	0.00%	5.81%
	FALSE	0.00%	94.19%
Training	TRUE	0.00%	5.79%
	FALSE	0.01%	94.20%

Figure 17: Random Forest Summary and Prediction Matrix for Primary Type – Criminal Damage

Summary Prediction Importance Pred. Matrix Probability ROC Data

Data Type	AUC	F Score	Accuracy Rate	Misclassification Rate	Precision	
Training	0.96853215034	0.033931168202	0.94305714286	0.056942857143	1	(
Test	0.67674866303	0.0090909090909	0.94186666667	0.058133333333	0.44444444444	0.

		Predicted	
Data Type	**Actual**	**TRUE**	**FALSE**
Test	TRUE	0.03%	5.78%
	FALSE	0.03%	94.16%
Training	TRUE	0.10%	5.69%
	FALSE	0.00%	94.21%

Figures 18 and 19 show the primary type – assault models, where logistic regression performs better at 82.5% and lower false positives and negatives rates on average for logistic regression.

Figure 18: Logistic Regression Summary and Prediction Matrix for Primary Type – Assault

Summary Prediction Importance Coefficients Coef. (Significant) Coef. Table Collinearity Pred. Matrix Proba

Data Type	AUC	F Score	Accuracy Rate	Misclassification Rate	Precision	
Training	0.669242389	0.16427216048	0.82382857143	0.17617142857	0.79319371728	0.09
Test	0.66754878619	0.16603295311	0.82453333333	0.17546666667	0.8036809816	0.

		Predicted	
Data Type	**Actual**	**TRUE**	**FALSE**
Test	TRUE	1.75%	17.12%
	FALSE	0.43%	80.71%
Training	TRUE	1.73%	17.17%
	FALSE	0.45%	80.65%

Figure 19: Random Forest Summary and Prediction Matrix for Primary Type – Assault

Summary Prediction Importance Pred. Matrix Probability ROC Data

Data Type	AUC	F Score	Accuracy Rate	Misclassification Rate	Precision	
Training	0.9338428441	0.34056720099	0.8472	0.1528	0.92312834225	0.20
Test	0.64393255442	0.1862745098	0.82293333333	0.17706666667	0.70046082949	0.1

		Predicted	
Data Type	**Actual**	**TRUE**	**FALSE**
Test	TRUE	2.03%	16.84%
	FALSE	0.87%	80.27%
Training	TRUE	3.95%	14.95%
	FALSE	0.33%	80.77%

Both machine learning models, logistic regression and random forest, perform very well with close marginal of success between each other. Thus, it is advisable to deploy both models using predictive analytics. However, ideally, more machine learning models should be tested with the data, such as XGBoost, decision tree, linear regression, clustering, naïve Bayes, support vector machines and neural networks (Fumo 2017).

6. Recommendations

Based on the descriptive and predictive analytics outputs and insights, several recommendations can be made for the CPD and the city of Chicago to serve better the public's interest in ensuring that their safety and wellbeing are protected. The recommendations will be relevant to the research project objective to reduce crime in the city by predicting which crimes will result in arrest once the primary type is reported. First, using the insights gained in descriptive analytics based on Figures 4 and 5, the CPD can pinpoint the crime hotspot, filter by its primary and secondary type, and the most likely location for the crime to happen. Now, CPD can mobilise resources to prevent crimes from ever happening in the first place by placing patrol units in crime hotspots. For instance, Figures 8 and 9 show that the street was the highest jump in theft crime from 2015 to 2016. Moreover, the secondary description also helps each department in CPD focus more on a specific type of crime to reduce the growing rampant crimes across the city. For example, theft involving a monetary value of $500 under and above is two of the biggest contribution to theft crimes from 2015 to 2016. The recommended strategy is to increase police foot patrolling on streets, especially in the hotspots of crime and look out for possible offenders that act suspicious such as following and stalking. Besides that, CPD can create an awareness campaign on the street smart to promote good habits while out in the streets, such as travelling in groups, carrying small cash and avoiding dangerous areas when out late at night (Columbia University 2022; Link 2018).

Then, combining the insights gathered in the descriptive section, now predictive analytics can target specific widespread primary type crimes such as theft, battery, criminal damage and assault. Referring to all of the models generated from Figures 10 to 19, the CPD can utilise the machine learning models for future crime data. For instance, in the future and in real-time, any of the assault crimes reported, the CPD can quickly understand that it is a high crime rate while having a low arrest rate. As a strategy, any of the high-profile primary types must be given immediate attention to increase the number of arrests while simultaneously decreasing the crime rate. A study also shows that over 25,000 criminal offenders examined, around 77% are likely to get re arrested within 5 years, which is one indicator that offenders are very likely to commit future crimes again. Hence, by responding to high crime rate crimes with extensive resources to ensure arrests, future crime rates will drop significantly as offenders are unable to commit further crimes in the community (The Marshall Project 2022). Furthermore, based on the variables that are used to predict the arrest rates, the CPD can also use the same variables when similar crimes are reported under the same circumstance. This will allow the CPD to know in advance whether the specific crime reported will result in an arrest or not. Then resources can be better distributed to increase arrest rates, such as an increase in staffing and patrol vehicles.

Nevertheless, the methodologies, techniques and data have limitations that must be addressed. First, the data collected did not contain enough variables to predict arrest rates accurately. Ideally, more data on each crime's specificity will help generate advanced machine learning models that understand more inputs. As mentioned earlier, more machine learning algorithms should also be tested, such as XGBoost and neural networks techniques. This allows the CPD to understand how different algorithms work on different data types and choose one or a combination of techniques when addressing a business problem (Fumo 2017).

In the future, as the project and data grow exponentially, most data analytics methodologies are recommended for change to fit better with the existing software and tools required to produce reliable insights. For example, when dealing with big data, the CPD must invest heavily in system requirements such as the cloud systems and business intelligence tools, talent resourcing to attract data analytics professionals, data governance compliance and procedures as well as training toward data-driven culture within CPD (Kumar 2021; Gartner 2022).

7. Data Ethics and Security

Data analytics projects are no stranger to data ethics, privacy and security issues, and this research project has taken the necessary steps to follow strict industry best practices. All the collected data contain no personal information and are completely anonymous at every stage of the project, even from the beginning when the data source is already anonymised. The data was downloaded from open-source government data, where all legal compliance procedures have been met. The data visualisations created in descriptive and exploratory analytics are transparent, as the visuals represent and are displayed as the data suggests. The visuals are not presented in a way that is biased and supports only one side of the argument or influences the audience. The visuals are presented to tackle the main business problem of utilising data analytics to generate machine learning models to predict arrests. Strict guidelines are also followed for predictive analytics, with only the actual data source being used in the model and no other data being manipulated.

However, many ethical, privacy and security concerns need to be considered moving forward as the analytics will get more advanced with a more considerable volume of data. For example, more personal details for both the offender and victim can be collected to obtain a more accurate analysis of the crime patterns. The data collected then will have privacy concerns from all parties regarding how they would like their data to be used, even if it is for a good cause (Panelli 2018). The city or police department also needs to improve data security, and as data are transitioning into the cloud environment, cyber hacks and cybercrime will pose a significant threat. For a better and more comprehensive guideline, the CPD can follow the general data protection regulation (GDPR), which can help organisations comply with strict data ethical procedures (GDPR 2022).

8. References

Ballesteros, C 2020. "Chicago Has Nearly Tripled Per-Capita Police Spending Since 1964, Data Shows" Block Club Chicago, retrieved 10 June 2022, https://blockclubchicago.org/2020/06/10/chicago-has-nearly-tripled-per-capita-police-spending-since-1964-data-shows/

Chicago Data Portal 2022. "Crimes – 2020" Chicago Data Portal, City of Chicago, retrieved 12 June 2022, https://data.cityofchicago.org/Public-Safety/Crimes-2020/qzdf-xmn8

Columbia University 2022. "Crime Prevention Tips", Columbia University, retrieved 19 June 2022, https://publicsafety.columbia.edu/crimepreventiontips

Daniel, E 2021. "Chicago Crime Statistics By Year" Chicago Reporter, retrieved 10 June 2022, https://www.chicagoreporter.net/chicago-crime-statistics-by-year/

Fumo, D 2017. "Types of Machine Learning Algorithms You Should Know", Medium, Towards Data Science, retrieved 15 June 2022, https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861#:~:text=There%20so%20m%20e%20variations%20of%20how%20to,Unsupervised%20Learning%203%20Semi-supervised%20Learning%204%20Reinforcement%20Learning

Gartner 2022. "5 Steps to Maximise Data Center Efficiency", Gartner, Gartner Inc, retrieved 22 June 2022, https://www.gartner.com/smarterwithgartner/5-steps-to-maximize-data-center-efficiency

GDPR 2022. "GDPR Summary", GDPR Summary, GDPR, retrieved 17 June 2022, https://www.gdprsummary.com/gdpr-summary/

Gray Matter 2022. "Using Big Data to Predict Crime Patterns" Gray Matter, retrieved 10 June 2022, https://graymattersystems.com/big-data-crime-patterns/

Hampton, D, J 2021. "Violent crime is rising in Chicago. Business leaders are concerned companies will leave.", Yahoo News, Yahoo, retrieved 10 June 2022, https://news.yahoo.com/chicago-business-leaders-rising-crime-083015880.html

Kumar, R 2021. "Data Strategy: 5 Key Elements & Best Practices" Hevo Data, Hevo Data Inc, retrieved 21 June 2022, https://hevodata.com/learn/data-strategy/

Link 2018. "Street smart safety tips to avoid being a target for criminals", Link, University of California, retrieved 18 June 2022, https://link.ucop.edu/2018/07/09/street-smart-safety-tips-to-avoid-being-a-target-for-criminals/

Masters in Data Science 2022. "What Is Exploratory Data Analysis?", Masters in Data Science, 2U Inc, retrieved 14 June 2022, https://www.mastersindatascience.org/learning/what-is-exploratory-data-analysis/#:~:text=Simply%20defined%2C%20exploratory%20data%20analysis%20%28EDA%20for%20short%29,data%20analysts%20to%20approach%20a%20database%20without%20assumptions.

Mittal, A 2021. "Data Cleansing: Why It's Important", Dataversity, retrieved 12 June 2022, https://www.dataversity.net/data-cleansing-why-its-important/

Molina, T 2022. "Multiple people were shot in downtown Chicago neighborhoods over holiday weekend as crime keeps rising", CBS, CBS Broadcating Inc, retreievd 10 June 2022, https://www.cbsnews.com/chicago/news/multiple-people-were-shot-in-downtown-chicago-neighborhoods-memorial-day-weekend-crime-statistics-keep-rising/

Panelli, J 2018. "Viterbi Conversations in Ethics", Viterbi Conversations in Ethics, USSC Viterbi, retrieved 18 June 2022, https://vce.usc.edu/volume-2-issue-2/ethics-of-predictive-policing/

Sanburn, J 2016. "Chicago Is Responsible for Almost Half of the Increase in U.S. Homicides", Time, retrieved 10 June 2022, https://time.com/4497814/chicago-murder-rate-u-s-crime/

Schuba, T, Struett, D, Grimm, A, Main, F and Boyle, A 2021. "Downtown shootings up 220%, biggest spike in city: 'People are fed up", Chicago Sun-Times, retrieved 10 June 2022, h What Is Exploratory Data Analysis? What Is Exploratory Data Analysis?ttps://chicago.suntimes.com/2021/10/29/22751518/downtown-chicago-loop-shootings-murder-river-north

The Marshall Project 2022. **"**Seven Things to Know About Repeat Offenders**"** The Marshall Project, retrieved 19 June 2022, https://www.themarshallproject.org/2016/03/09/seven-things-to-know-about-repeat-offenders