



**VIT<sup>®</sup>**  
**Vellore Institute of Technology**  
(Deemed to be University under section 3 of UGC Act, 1956)

**School of Computer Science and Engineering**

VIT Chennai

Vandalur - Kelambakkam Road, Chennai - 600 127

**Final Review Report**

**Programme: Integrated M. tech in CSE with specialization in Business Analytics**

**Course: BIG DATA FRAMEWORKS**

**Slot: CSE3120**

**Faculty: G2**

**Component: J**

**Title: Chicago Crime data analysis**

**Team Member(s):**

NOEL BENNY                      20MIA1020

ANTONY GEORGE MATHEW   20MIA1022

DIYA AL DIN                      20MIA1048

## **ABSTRACT**

Crimes are an unavoidable aspect of life in this planet. We learn about them every day, and some of us have even participated in one or more of them. Being cautious and increasing safety is no longer a straightforward guideline. To combat this issue more effectively, we need to make use of contemporary technology and data science methodologies. The police department has amassed a vast amount of records and documentation over the years, which can be used as a significant source of information for data analytics jobs. Applying analytical tasks to these data provides us with useful knowledge that we can use to improve societal safety and reduce crime.

In order to better understand the security situation in this city, we analyse the Chicago Crime dataset (between the years 2001 and 2017), one of the most comprehensive open source data sets available. Our project's findings indicate that only 28.33 percent of the reported criminals were really apprehended. As compared to a number of crimes from 2008 to 2010, the number of crimes reported in subsequent years had significantly fallen. Which demonstrates that Chicago's security has improved throughout the years. In this investigation, we also examined the connection between various crime categories and the setting in which they occurred.

The findings revealed that the top four offences in Chicago, theft, battery, and criminal property damage, accounted for 65.7% of all crimes committed there. Additionally, our findings indicate that these crimes typically occur in common locations like sidewalks, streets, apartments, and homes. We also examined the statistics of three particular crimes in our research: theft, homicide, and sexual harassment. Our findings show that there has been a consistent decrease in homicides between the years of 2001 and 2016, with an average of 482 homicides per year. But the number of homicides has sharply increased in 2016. Similar to this, there has been an upsurge in theft and sexual harassment in 2016.

## **Problem Definition**

We came up with a few questions to help us get perspective on the security conditions in Chicago, and we addressed them during our data analysis study.

The questions are listed below:

1. How have certain offences' occurrence rates changed over time in Chicago?
2. How has the relationship between arrests and crimes altered over time in Chicago?
3. Do the offences that are being committed have any patterns?
4. What types of crimes are perpetrated most frequently?
5. Where are these crimes most frequently committed?
6. Are there specific areas with a high crime rate for specific offences (such as sexual offence)?
7. How have various crimes in Chicago changed in frequency over time, such as homicide?

We used the four primary KDD data mining pipeline steps—data preprocessing, data pre-processing, analysis, and post-processing—to provide answers to these questions .

## 2 Pre-Processing Data: Data Extraction

### 2.1 Data Investigation

Our data was obtained from the Kaggle website. This dataset "reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to 2017," according to the information given with this data collection. Data is taken from the CLEAR (Citizen Law Enforcement Analysis and Reporting) system of the Chicago Police Department. Addresses are only displayed at the block level and individual locations are not revealed to preserve the anonymity of crime victims [1]. In general, the data includes details like the date and time the crime occurred, the street where it happened, the sort of offense, a description of the place, whether an arrest was made, and the precise location.

#### 2.1.1 Data size:

There were 24 columns and 7,939,202 records in the data .

#### 2.1.2 a glimpse of the data

The data may be viewed at this address, but in order to give you a fast overview of the data set's structure, we've also included a screenshot of its first few records here.

Chicago\_Crimes\_2001\_to\_2004.csv (116.42 MB)

20 of 23 columns

	#	# ID	A Case N...	📅 Date	A Block	# IUCR	A Primary...	A Descrip...	A Locatio...	✓ Arrest	✓ Domes...
1	879	4786321	HM399414	01/01/2004 12:01:00 AM	082XX S COLES AVE	0840	THEFT	FINANCIAL ID THEFT: OVER \$300	RESIDENCE	False	False
2	2544	4676906	HM278933	03/01/2003 12:00:00 AM	004XX W 42ND PL	2825	OTHER OFFENSE	HARASSMENT BY TELEPHONE	RESIDENCE	False	True
3	2919	4789749	HM402220	06/20/2004 11:00:00 AM	025XX N KIMBALL AVE	1752	OFFENSE INVOLVING CHILDREN	AGG CRIM SEX ABUSE FAM MEMBER	RESIDENCE	False	False
4	2927	4789765	HM402058	12/30/2004 08:00:00 PM	045XX W MONTANA ST	0840	THEFT	FINANCIAL ID THEFT: OVER \$300	OTHER	False	False
5	3302	4677901	HM275615	05/01/2003 01:00:00 AM	111XX S NORMAL AVE	0841	THEFT	FINANCIAL ID THEFT:\$300 & UNDER	RESIDENCE	False	False
6	3633	4838048	HM440266	08/01/2004 12:01:00	012XX S HARDING	0841	THEFT	FINANCIAL ID	APARTMENT	False	False

20 of 23 columns											
Id...	Arrest	Domes...	# Beat	# District	# Ward	# Comm...	# FBI Code	# X Coord...	# Y Coord...	# Year	Update...
NCE	False	False	424	4.0	7.0	46.0	06			2004	08/17/2015 03:03:40 PM
NCE	False	True	935	9.0	11.0	61.0	26	1173974.0	1876757.0	2003	04/15/2016 08:55:02 AM
NCE	False	False	1413	14.0	35.0	22.0	20			2004	08/17/2015 03:03:40 PM
	False	False	2521	25.0	31.0	20.0	06			2004	08/17/2015 03:03:40 PM
NCE	False	False	2233	22.0	34.0	49.0	06	1174948.0	1831051.0	2003	04/15/2016 08:55:02 AM
IENT	False	False	1011	10.0	24.0	29.0	06			2004	08/17/2015 03:03:40 PM

## Data Extraction

The Stanford Visualization Group's Data Wrangler [2], Python pandas [3], OpenRefine [4], or Microsoft SQL Management Studio [5] are just a few examples of the numerous tools available for data preprocessing. These are excellent tools that can save hours. Additionally, their functionalities overlap. But because setting up and utilising OpenRefine is so simple and convenient, we used this tool to understand our data. Using OpenRefine, we discover that cleaning is one of the most crucial data pre-processing steps that the Chicago Crime dataset needs in order to meet our project's objectives.

### Our data must be pure by:

deleting redundant rows

removing null/NA values, missing values, etc. from the dataset

removing all dataset features that are not pertinent to our data analysis (etc. X Coordinate, Y Coordinate, Latitude Longitude).

We used SQL Management Studio to perform these preparation procedures on our dataset in the following order: First, we had to delete a lot of incorrect data from each entry. For instance, 70,627 records were filtered from a file containing 1,923,865 data because they did not match the column property. After that, in order to have cleaner data, we had to identify and eliminate any incorrect data from each column.

	Description	Location Description	Arrest	Domestic	District	Year
	PRO EMP HANDS NO/MIN INJURY	"SCHOOL	PUBLIC	BUILDING"	NULL	NULL
	FALSE FIRE ALARM	"SCHOOL	PUBLIC	BUILDING"	NULL	NULL
	SIMPLE	"SCHOOL	PUBLIC	BUILDING"	NULL	NULL
	EMBEZZLEMENT	"SCHOOL	PRIVATE	GROUNDS"	NULL	1167102
	PRO EMP HANDS NO/MIN INJURY	"SCHOOL	PUBLIC	BUILDING"	NULL	NULL
EN	AGG CRIM SEX ABUSE FAM MEMBER	"SCHOOL	PUBLIC	GROUNDS"	NULL	NULL
	AGG CRIMINAL SEXUAL ABUSE	"SCHOOL	PUBLIC	BUILDING"	NULL	NULL
	"TRUCK	BUS	MOTOR HOME"	VACANT LOT/LAND	NULL	NULL
	FINANCIAL ID THEFT: OVER \$300	"SCHOOL	PUBLIC	GROUNDS"	NULL	NULL
	"THEFT BY LESSEE	MOTOR VEH"	AIRPORT/AIRCRAFT	False	813	NULL
	OVER \$500	"SCHOOL	PUBLIC	BUILDING"	NULL	NULL
	OVER \$500	"SCHOOL	PRIVATE	BUILDING"	NULL	NULL
	TO PROPERTY	"SCHOOL	PRIVATE	BUILDING"	NULL	NULL
	"TRUCK	BUS	MOTOR HOME"	STREET	NULL	NULL
	"THEFT BY LESSEE	MOTOR VEH"	OTHER	False	1651	NULL
	"THEFT BY LESSEE	MOTOR VEH"	OTHER	False	1622	NULL
	"TRUCK	BUS	MOTOR HOME"	STREET	NULL	NULL
	"THEFT BY LESSEE	MOTOR VEH"	STREET	False	1651	NULL
	FINANCIAL IDENTITY THEFT OVER \$ 300	"SCHOOL	PRIVATE	GROUNDS"	NULL	NULL

We still need to remove all null values after that because we only wanted to concentrate on the true data. In order to find any null values in any column, we execute another query.

	Primary Type	Location Description	Arrest	Domestic	District	Year
VE	DECEPTIVE PRACTICE	NULL	0	0	15	2015
	DECEPTIVE PRACTICE	NULL	0	0	2	2014
	DECEPTIVE PRACTICE	NULL	0	0	5	2015
E	DECEPTIVE PRACTICE	NULL	0	0	25	2016
	DECEPTIVE PRACTICE	NULL	0	0	18	2016
AVE	DECEPTIVE PRACTICE	NULL	0	0	17	2016
	DECEPTIVE PRACTICE	NULL	0	0	12	2016
	THEFT	NULL	0	0	18	2012
	THEFT	NULL	0	0	22	2012
N ST	THEFT	NULL	0	0	19	2012
R AVE	THEFT	NULL	0	0	19	2012
	THEFT	NULL	0	0	8	2012
	THEFT	NULL	0	0	9	2012
VE	THEFT	NULL	0	0	19	2012
	THEFT	NULL	0	0	19	2012
	THEFT	NULL	0	0	3	2012
VE	THEFT	NULL	0	0	1	2012
R AVE	THEFT	NULL	0	0	2	2012
/E	DECEPTIVE PRACTICE	NULL	0	0	1	2016

We export the data into a csv file after making sure it is clean so that Spark SQL can use it. By cleaning the data, the size was reduced from approximately 5 GB to 700 MB.

### **3.1 Analysis deciding on technology (Spark SQL and Hadoop)**

Every data analytic project must be implemented using the technologies that are most appropriate for that particular project type because the technology will have an impact on the effectiveness, speed, and cost of the process as well as the project's completion. Take a machine learning project, for instance, that teaches Convolutional Neural Networks to classify photos. Because of its extensive machine learning libraries, data scientists frequently employ the Python programming language, which runs on cloud GPU computers, for projects like this one.

We chose to implement our project using Spark SQL and the Hadoop distributed file system after taking this into account as well as the scope and characteristics of our project.

We deal with a vast data set in our project, thus it was obvious that we needed to adopt a big data technology. It was difficult to choose between MapReduce and Spark, though. Each of these two frameworks has unique qualities and advantages. In conclusion, they both contribute to the effectiveness of big data processing. In actuality, the technique to processing is where they differ most from one another: Spark can do it in memory, whereas Hadoop MapReduce must read from and write to a disc. The processing speed varies greatly as a result.

Overall, we picked Spark as our primary framework since it met the demands of our project and because in-memory processing using Spark is significantly faster and more effective nowadays.

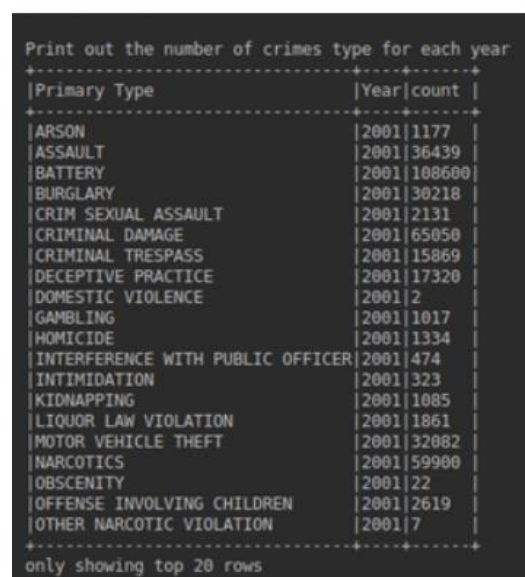
Also, because the goal of our project is to apply queries to the dataset in order to get the needed information, Spark SQL was utilised instead of Spark RDDs to create the queries that we applied to the dataset. We used Spark SQL because, on the one hand, it makes it easier to work with structured data using DataFrame and DataSet abstractions, and, on the other hand, it enables us to run standard SQL queries directly on top of Spark to get the answers we need.

### **3.2 Evaluation Section and outcome**

As previously mentioned, we use spark SQL to query the data set in order to provide answers to each of the questions we identified as the project's objectives . Continue reading for an explanation of how we responded to each of those questions, including the Spark SQL queries we used. We also provide screenshots of the results of each query, which serve as the final responses to each question.

#### **How have certain crimes in Chicago evolved in frequency throughout time?**

From this inquiry, we can see if each sort of crime has been declining or rising over the course of the year. The outcome is displayed in the following : Spark SQL.



```
Print out the number of crimes type for each year
+-----+-----+-----+
|Primary Type|Year|count|
+-----+-----+-----+
|ARSON|2001|1177|
|ASSAULT|2001|36439|
|BATTERY|2001|108600|
|BURGLARY|2001|30218|
|CRIM SEXUAL ASSAULT|2001|2131|
|CRIMINAL DAMAGE|2001|65050|
|CRIMINAL TRESPASS|2001|15869|
|DECEPTIVE PRACTICE|2001|17320|
|DOMESTIC VIOLENCE|2001|2|
|GAMBLING|2001|1017|
|HOMICIDE|2001|1334|
|INTERFERENCE WITH PUBLIC OFFICER|2001|474|
|INTIMIDATION|2001|323|
|KIDNAPPING|2001|1085|
|LIQUOR LAW VIOLATION|2001|1861|
|MOTOR VEHICLE THEFT|2001|32082|
|NARCOTICS|2001|59900|
|OBSCENITY|2001|22|
|OFFENSE INVOLVING CHILDREN|2001|2619|
|OTHER NARCOTIC VIOLATION|2001|7|
+-----+-----+-----+
only showing top 20 rows
```

### How has the amount of arrests in Chicago that are in line with the offences varied over time?

We can determine how many crimes were solved and how the crime occurred in Chicago between 2001 and 2017 by adding up the total number of crimes and arrests each year. We may obtain the annual number of crimes and arrests using the following SQL queries.

Spark SQL:

Crimes each Year:

```
CombineData.groupBy("Year").count().orderBy("Year").show(20,false);
```

Arrest each Year:

```
filterData.where("Arrest == 'True']").groupBy("Year").count().orderBy("Year").show(false);
```

Total Crime each year

Year	count
2001	568515
2002	490875
2003	475911
2004	388195
2005	455804
2006	794668
2007	621842
2008	851964
2009	783888
2010	700675
2011	351810
2012	335328
2013	306636
2014	274345
2015	262738
2016	264679
2017	11329

Print out number of arrest for each year

Year	count
2001	167888
2002	142803
2003	141499
2004	116679
2005	141159
2006	243779
2007	190129
2008	219196
2009	221052
2010	191503
2011	96224
2012	90511
2013	86218
2014	79055
2015	69396
2016	50385
2017	1902



### Are there any patterns in the crimes that are being committed?

In order to determine the type of crime that Chicago experiences the most frequently, we needed to look at the trending crimes between 2001 and 2017.

Primary Type	count
THEFT	1639777
BATTERY	1442702
CRIMINAL DAMAGE	922984
NARCOTICS	885426
OTHER OFFENSE	491917
ASSAULT	481650
BURGLARY	470948
MOTOR VEHICLE THEFT	370537
ROBBERY	300439
DECEPTIVE PRACTICE	279682
CRIMINAL TRESPASS	229365
PROSTITUTION	86399
WEAPONS VIOLATION	77424
PUBLIC PEACE VIOLATION	58547
OFFENSE INVOLVING CHILDREN	51441
CRIM SEXUAL ASSAULT	29865
SEX OFFENSE	28703
GAMBLING	18806
LIQUOR LAW VIOLATION	17513
INTERFERENCE WITH PUBLIC OFFICER	15710
ARSON	13097
HOMICIDE	9051
KIDNAPPING	7756
INTIMIDATION	4636
STALKING	3734
OBSCENITY	496
PUBLIC INDECENCY	163
OTHER NARCOTIC VIOLATION	144
NON-CRIMINAL	97
CONCEALED CARRY LICENSE VIOLATION	90
NON - CRIMINAL	38
RITUALISM	31
HUMAN TRAFFICKING	28
NON-CRIMINAL (SUBJECT SPECIFIED)	4
DOMESTIC VIOLENCE	2

### Which crimes are perpetrated the most frequently?

To be more specific, we need to figure out how many of each sort of crime occurred between 2001 and 2017 in order to determine what crime was committed most frequently throughout that time. We used the Spark SQL command below to discover the outcome:

Spark SQL

```
CombineData.groupBy("Primary Type").count().orderBy(desc("count")).show(40, false);
```



Here it displays the output of this operation, and it can be observed that throughout those years, for instance, 163977 thefts took place.

Primary Type	count
THEFT	1639777
BATTERY	1442702
CRIMINAL DAMAGE	922984
NARCOTICS	885426
OTHER OFFENSE	491917
ASSAULT	481650
BURGLARY	470948
MOTOR VEHICLE THEFT	370537
ROBBERY	300439
DECEPTIVE PRACTICE	279682
CRIMINAL TRESPASS	229365
PROSTITUTION	86399
WEAPONS VIOLATION	77424
PUBLIC PEACE VIOLATION	58547
OFFENSE INVOLVING CHILDREN	51441
CRIM SEXUAL ASSAULT	29865
SEX OFFENSE	28703
GAMBLING	18806
LIQUOR LAW VIOLATION	17513
INTERFERENCE WITH PUBLIC OFFICER	15710
ARSON	13097
HOMICIDE	9051
KIDNAPPING	7756
INTIMIDATION	4636
STALKING	3734
OBSCENITY	496
PUBLIC INDECENCY	163
OTHER NARCOTIC VIOLATION	144
NON-CRIMINAL	97
CONCEALED CARRY LICENSE VIOLATION	90
NON - CRIMINAL	38
RITUALISM	31
HUMAN TRAFFICKING	28
NON-CRIMINAL (SUBJECT SPECIFIED)	4
DOMESTIC VIOLENCE	2

### Where locations are these crimes most frequently committed?

The answer revealed that the majority of crimes in Chicago occur on the streets. This question allows us to determine where crimes occur most frequently in Chicago. The results will be shown and locally written to a csv file by the Spark SQL statements that follow.

Print out the frequent location that crime occur

Location Description	count
STREET	2101820
RESIDENCE	1341736
SIDWALK	815579
APARTMENT	812506
OTHER	294279
PARKING LOT/GARAGE(NON.RESID.)	225452
ALLEY	180151
SCHOOL, PUBLIC, BUILDING	173750
RESIDENCE-GARAGE	158550
RESIDENCE PORCH/HALLWAY	138492
SMALL RETAIL STORE	134003
VEHICLE NON-COMMERCIAL	128309
RESTAURANT	116290
GROCERY FOOD STORE	100812
DEPARTMENT STORE	95201
GAS STATION	82394
RESIDENTIAL YARD (FRONT/BACK)	75440
CHA PARKING LOT/GROUNDS	65123
PARK PROPERTY	61322
COMMERCIAL / BUSINESS OFFICE	58200
CTA PLATFORM	44710
CHA APARTMENT	41550
BAR OR TAVERN	41169
DRUG STORE	35671
SCHOOL, PUBLIC, GROUNDS	34772
BANK	32561
HOTEL/MOTEL	31475
CHA HALLWAY/STAIRWELL/ELEVATOR	30135
VACANT LOT/LAND	28033
TAVERN/LIQUOR STORE	26721
CTA TRAIN	25855
CTA BUS	25158
DRIVEWAY - RESIDENTIAL	23519
HOSPITAL BUILDING/GROUNDS	22801
AIRPORT/AIRCRAFT	21873
POLICE FACILITY/VEH PARKING LOT	20242
CHURCH/SYNAGOGUE/PLACE OF WORSHIP	17651
GOVERNMENT BUILDING/PROPERTY	16764

### Are there specific areas with a high crime rate for specific offences, such as sexual harassment?

From the response to the question, we may tell what kinds of crimes have occurred where and what kinds of crimes occur there most frequently . The results were written into a csv file using Spark SQL, which displayed the quantity of each category of crime for each region.

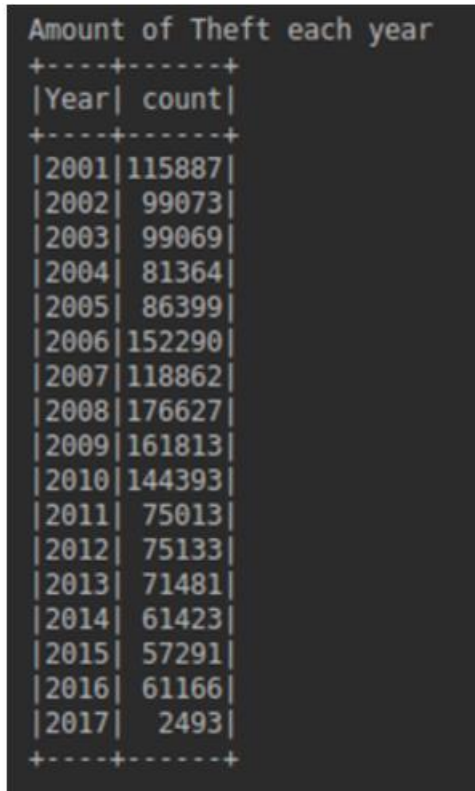
Calculate the number of crime x base on location

Primary Type	Location Description	count
ARSON	ABANDONED BUILDING	251
ARSON	AIRPORT EXTERIOR - NON-SECURE AREA	1
ARSON	ALLEY	430
ARSON	ANIMAL HOSPITAL	3
ARSON	APARTMENT	1031
ARSON	APPLIANCE STORE	1
ARSON	ATM (AUTOMATIC TELLER MACHINE)	1
ARSON	BANK	4
ARSON	BAR OR TAVERN	41
ARSON	BARBERSHOP	19
ARSON	BOAT/WATERCRAFT	2
ARSON	CAR WASH	6
ARSON	CHA APARTMENT	71
ARSON	CHA HALLWAY/STAIRWELL/ELEVATOR	27
ARSON	CHA PARKING LOT/GROUNDS	13
ARSON	CHURCH/SYNAGOGUE/PLACE OF WORSHIP	61
ARSON	CLEANING STORE	3
ARSON	COLLEGE/UNIVERSITY GROUNDS	8
ARSON	COLLEGE/UNIVERSITY RESIDENCE HALL	2
ARSON	COMMERCIAL / BUSINESS OFFICE	65

only showing top 20 rows

### How has the frequency of specific crimes in Chicago evolved over time (homicide, etc.)?

Given that theft is the most common crime in Chicago (see Figure 14) and that homicide involves the killing of individuals, we chose these two crimes to view in order to determine whether they have increased or decreased over time (see Figure 15). Also, we selected criminal sexual assault and sex offences because we wanted to check how Chicago's security and safety for women was (see Figure 16). The entire Spark SQL result set was then written to a csv file.



A terminal window displaying a table titled 'Amount of Theft each year'. The table has two columns: 'Year' and 'count'. The data shows a general downward trend in theft counts over the years, starting at 115,887 in 2001 and ending at 2,493 in 2017. The table is enclosed in a box with a dark background and light-colored text.

Year	count
2001	115887
2002	99073
2003	99069
2004	81364
2005	86399
2006	152290
2007	118862
2008	176627
2009	161813
2010	144393
2011	75013
2012	75133
2013	71481
2014	61423
2015	57291
2016	61166
2017	2493

Amount of Homicide each year

Year	count
2001	1334
2002	780
2003	604
2004	454
2005	453
2006	481
2007	448
2008	513
2009	460
2010	438
2011	437
2012	503
2013	422
2014	424
2015	497
2016	772
2017	31

Amount of Criminal Sexual Assault + Sex Offense each year

Primary Type	Year	count
CRIM SEXUAL ASSAULT	2001	2131
CRIM SEXUAL ASSAULT	2002	1896
CRIM SEXUAL ASSAULT	2003	1621
CRIM SEXUAL ASSAULT	2004	1430
CRIM SEXUAL ASSAULT	2005	1574
CRIM SEXUAL ASSAULT	2006	2545
CRIM SEXUAL ASSAULT	2007	2130
CRIM SEXUAL ASSAULT	2008	3014
CRIM SEXUAL ASSAULT	2009	2786
CRIM SEXUAL ASSAULT	2010	2439
CRIM SEXUAL ASSAULT	2011	1476
CRIM SEXUAL ASSAULT	2012	1406
CRIM SEXUAL ASSAULT	2013	1264
CRIM SEXUAL ASSAULT	2014	1305
CRIM SEXUAL ASSAULT	2015	1339
CRIM SEXUAL ASSAULT	2016	1445
CRIM SEXUAL ASSAULT	2017	64
SEX OFFENSE	2001	2616
SEX OFFENSE	2002	2219
SEX OFFENSE	2003	2106

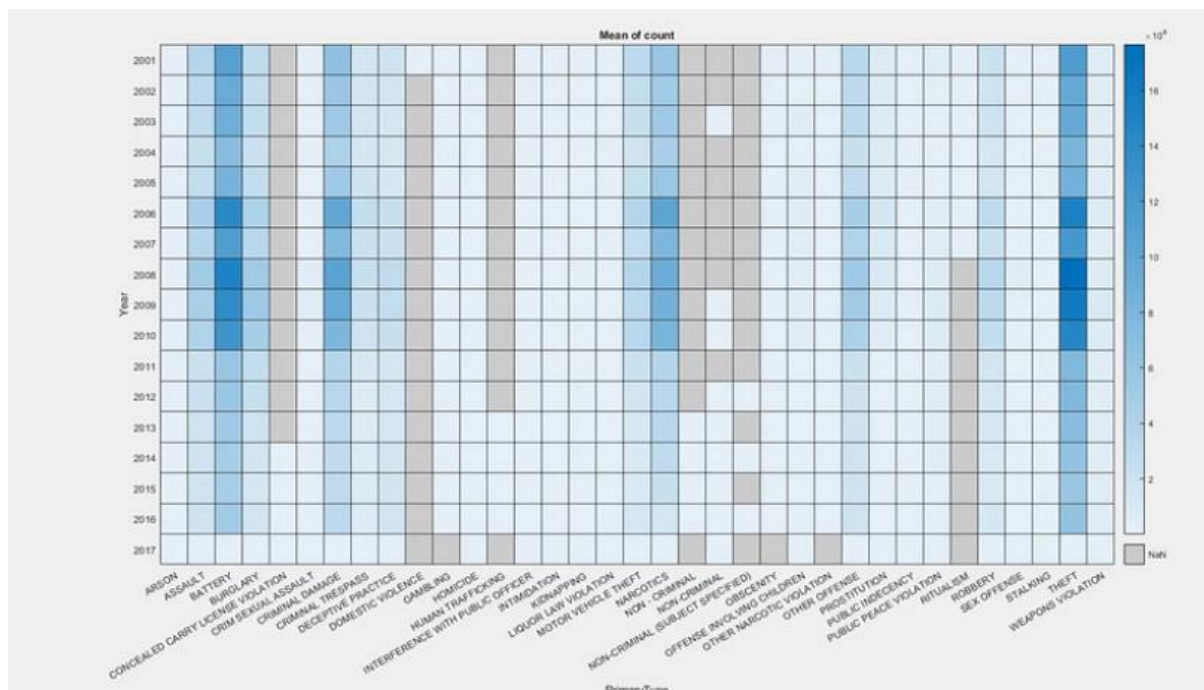
only showing top 20 rows

### Post-processing of Data: Visualization

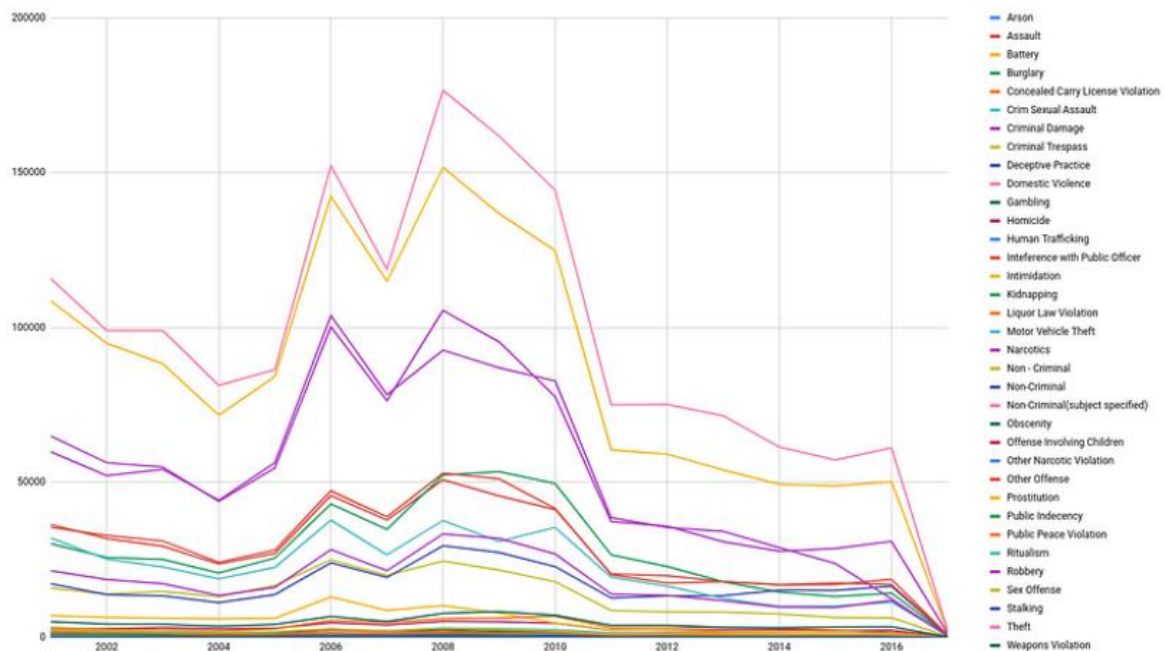
After we performed the analysis and obtained the results using Spark SQL , we imported the data into Google Sheet and produced charts to provide a better understanding of the resultant data through better visualisation. The following charts are just photos, however you can access an interactive website we made with these charts by clicking [here](#).

### How have certain crimes in Chicago evolved in frequency throughout time?

With the use of a stacked column chart, we can see the data and determine which years had the most and least crime. Looking at the graphic, we can see that there was a significant drop in crime in 2011 after having a high level of crime from 2006 to 2010. We link the rapid decrease in crime from 2007 to 2011 with the recovery from the recession.



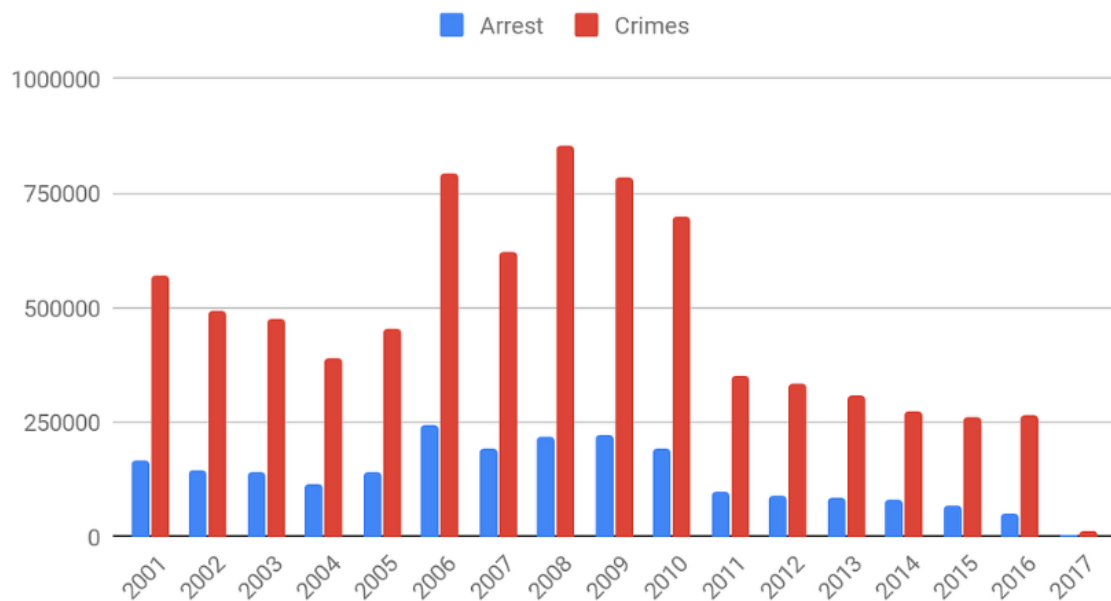
Rate of each Crime Per Year



### How has the amount of arrests in Chicago that are in line with the offences varied over time?

Since none of the arrests were more than half the crime rate for any given year, as seen in the bar chart, we can conclude that police capabilities were not very strong. (see Figure 19). This graph demonstrates that Chicago's security does not appear to be very good given the number of crimes that were not addressed by an arrest.

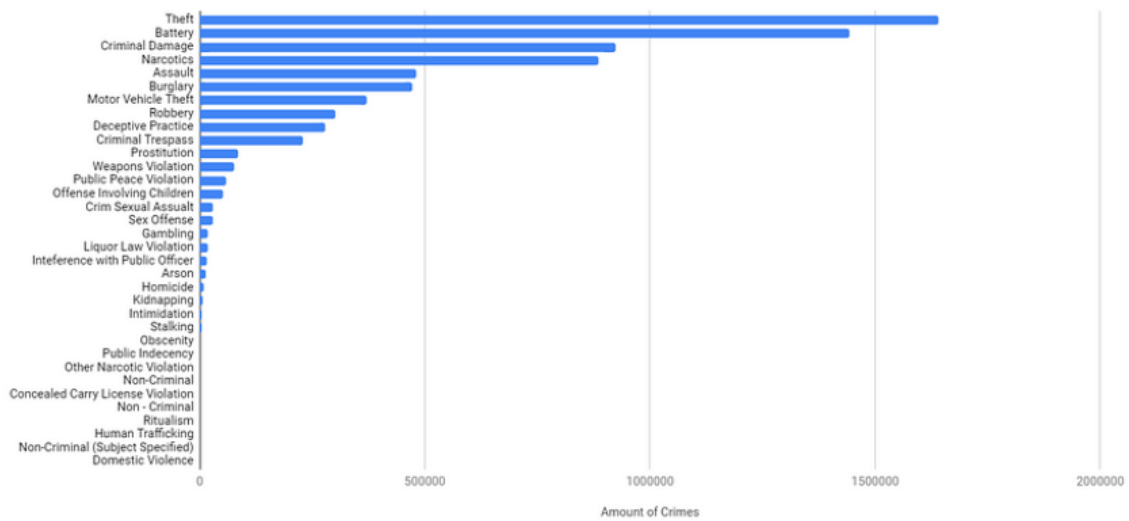
#### Arrest and Crimes



### Exist any patterns in the crimes that are being committed?

The table shows that theft and battery occur at quite a different rate from other crimes, so it is advisable to be on the lookout for theft and battery if you live in or visit Chicago.

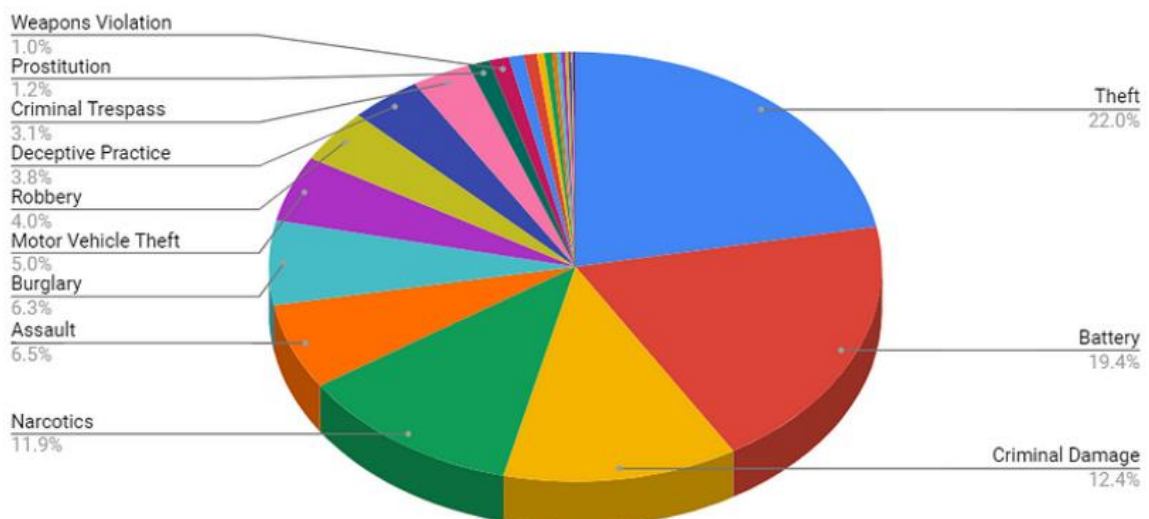
Amount of Crimes



### Which crimes are perpetrated the most frequently ?

As can be seen from the pie chart, theft is the most common crime, accounting for 22% of all crimes, perhaps due to the fact that it carries a low risk of detection and a high potential reward. The most common crimes in Chicago, which account for 65.7% of all crimes, are theft, battery, criminal damage, and drug possession, as seen in the graph.

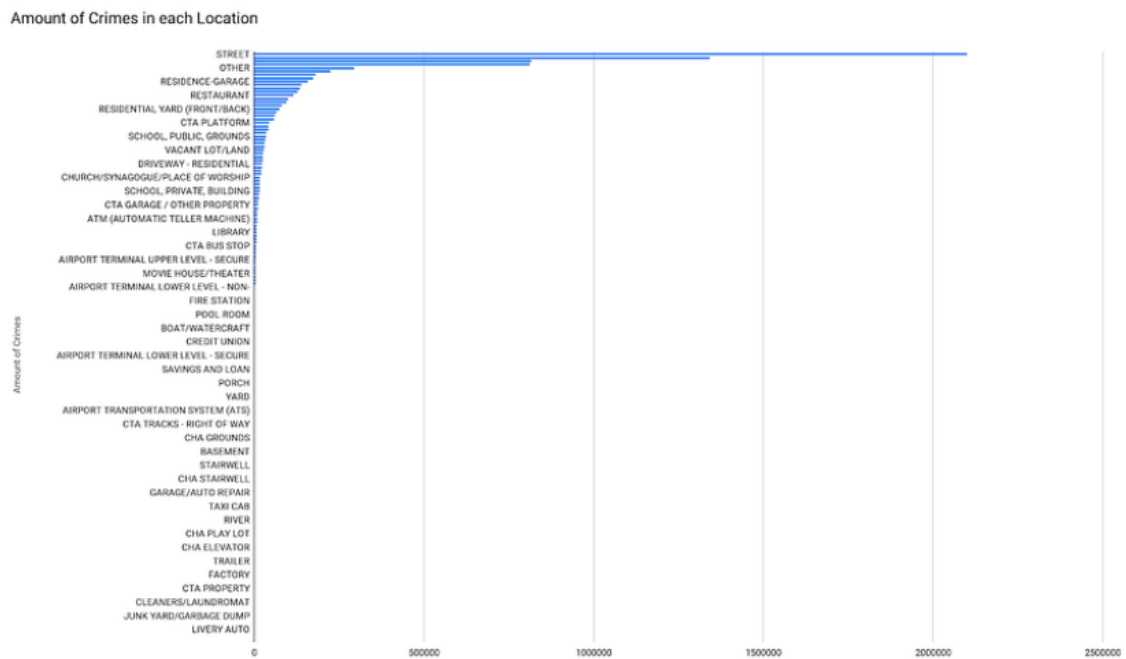
Amount of Crimes





### Where are these crimes most frequently committed?

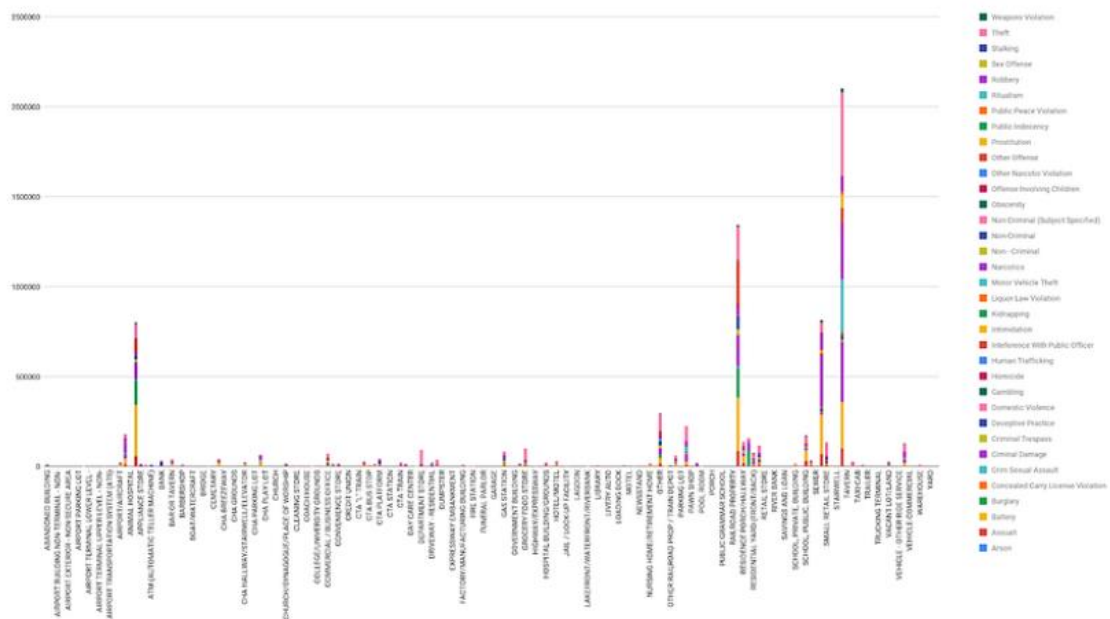
Inferring from the chart that there is a high rate of crime, we believe that this is because of the ease with which offenders may flee the streets because they are familiar with the region and because of the dense population that lines them.



### Do specific crimes, such as sexual harassment, have high crime areas?

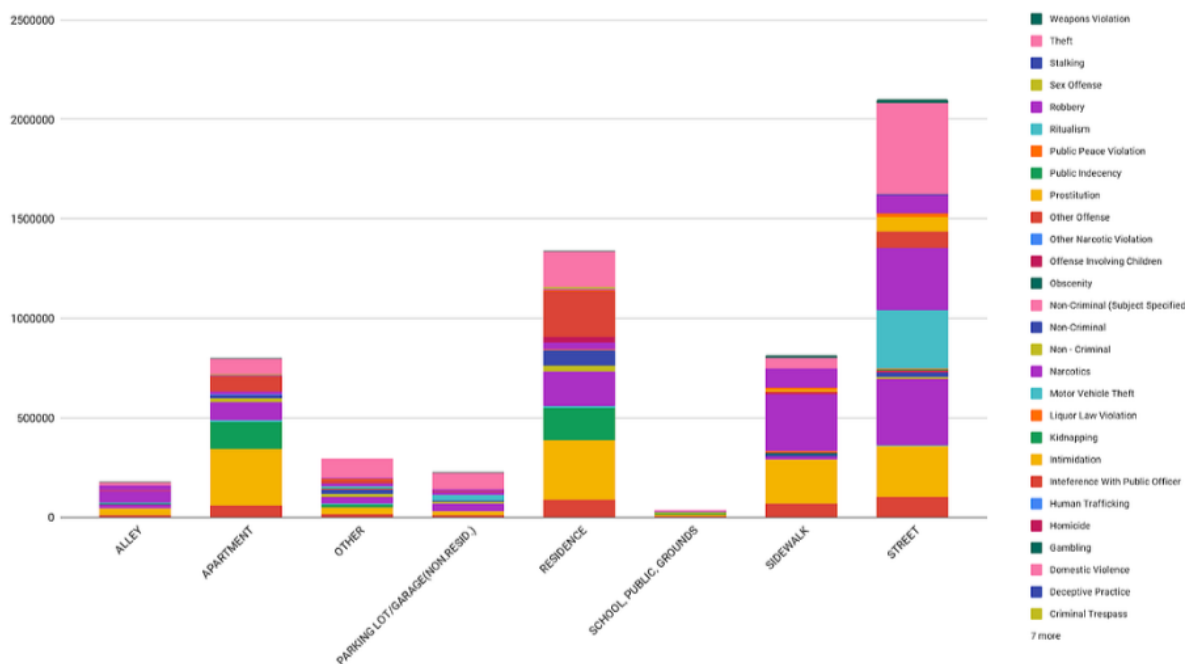
The most common crime in each location, according to the results, is theft, which indicates that Chicago's security situation is not very good. As a result, people should be cautious of having their possessions taken on the street.

Rate of Each Crime Per Location

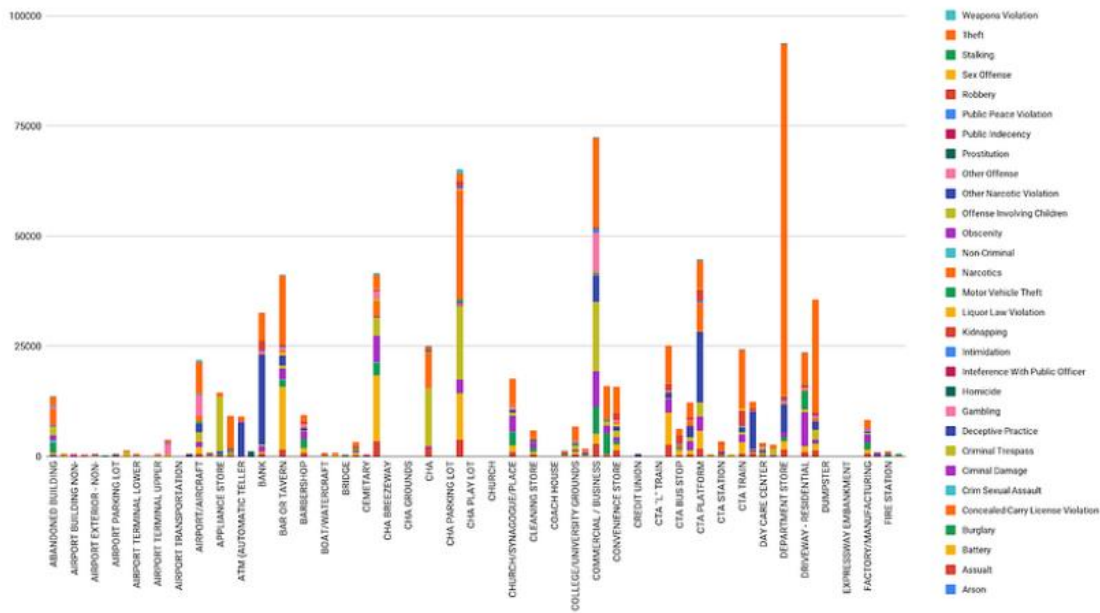


We wish to gain a deeper understanding of the prevalence of each crime in each area because the first chart, which shows the overall number of each crime type for each location, had too much information. In order to make the chart easier to view, we divided them according to the places where crimes occur most frequently.

Top Locations with each Crime

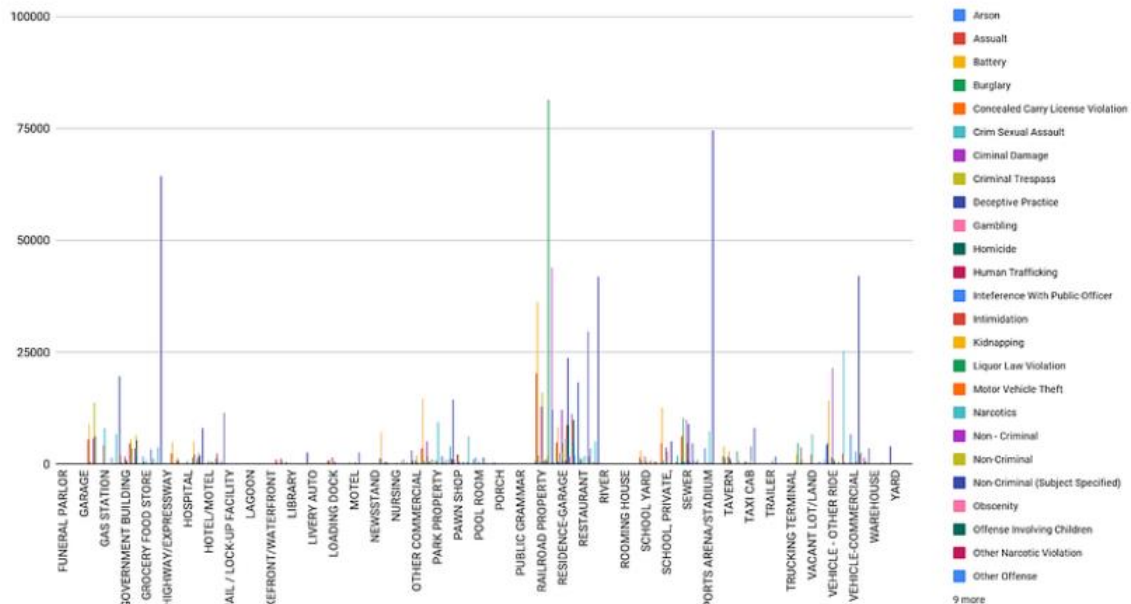


First Half of Each Crime Per Location



The majority of burglaries in Chicago occur in residential garages, therefore people there should use caution when getting off their automobiles there and should lock the garage when leaving the house, as seen in the second half of the crime by location figure.

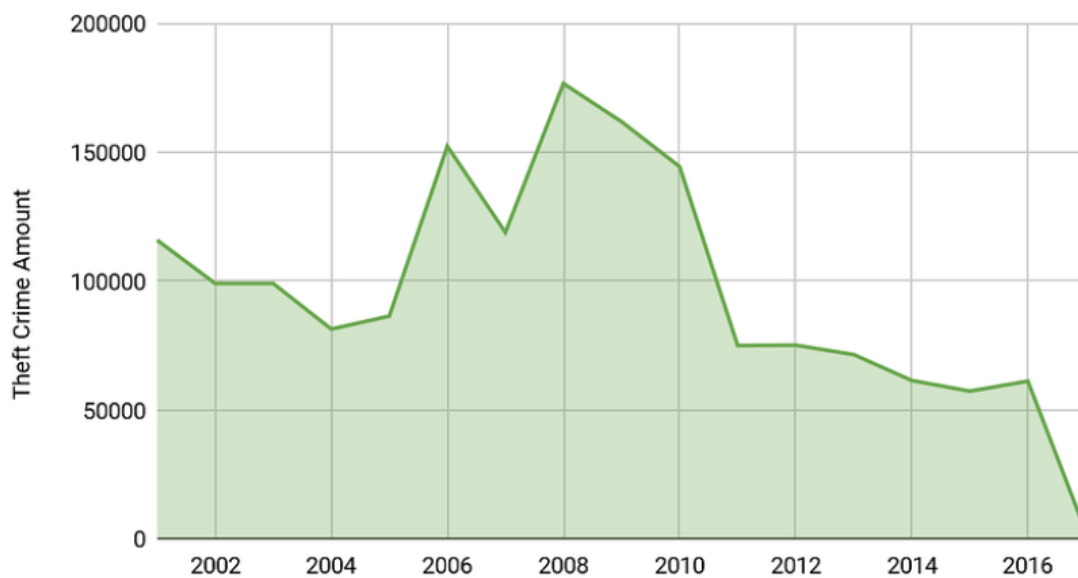
Second Half of Crime Per Location



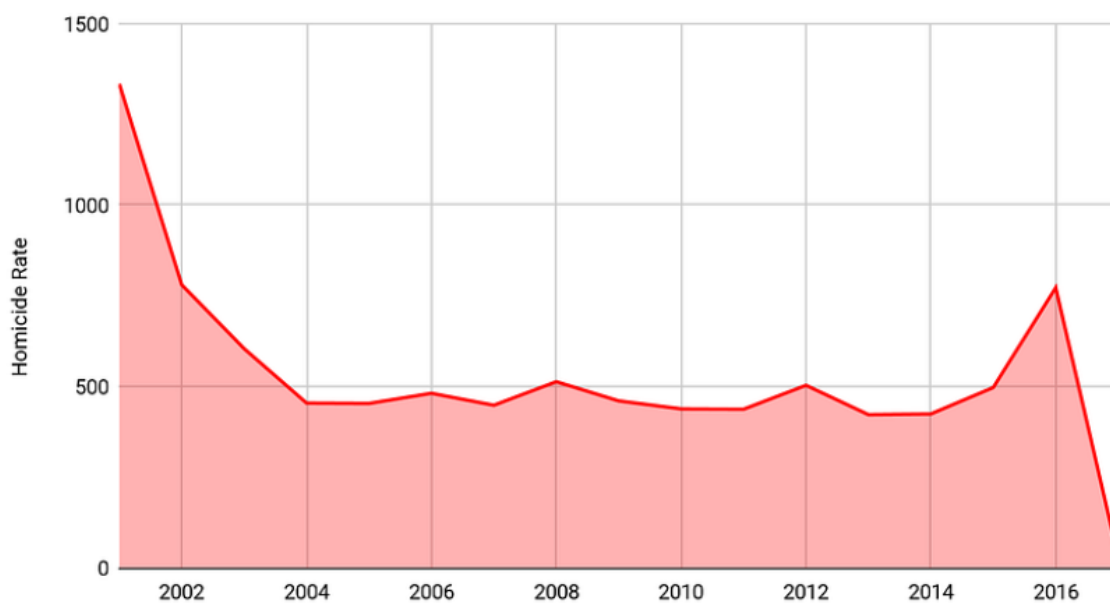
### How has Chicago's crime rate—for example, homicide—changed over time?

Because we believe it is tied to the great depression, which starts in December 2007 and lasts until 2009 and causes people to turn to stealing, we can notice a strong spike in theft from 2007 to 2010 on the theft bar column chart. The homicide rate saw a dramatic decline beginning in 2001, which we believe was caused by increased security in the wake of 9-11. Also, there was an increase in sexual offence offences in 2016, which is not good, especially for the safety of women.

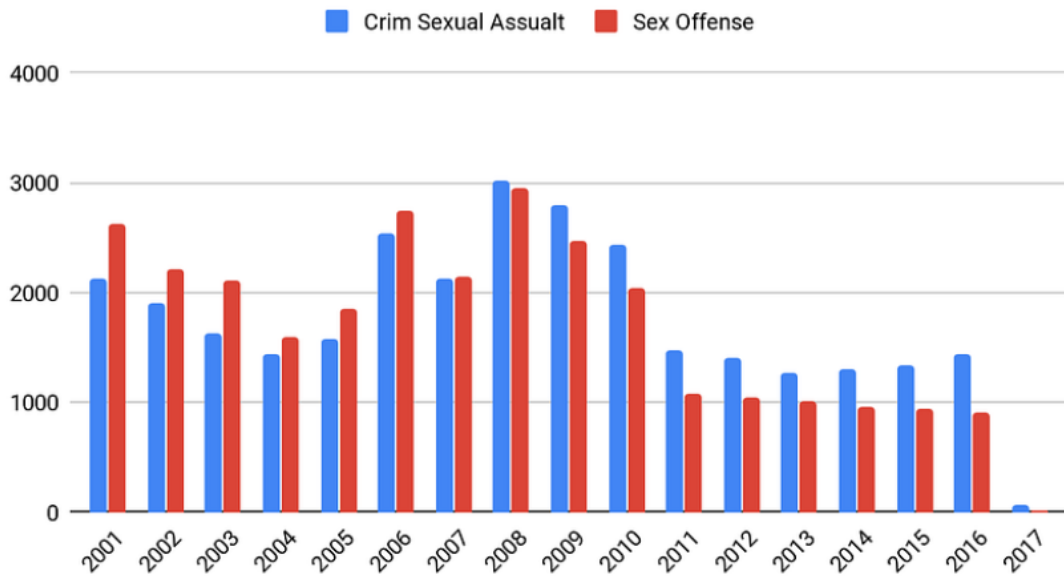
#### Theft Rate Each Year



#### Homicide Rate Each Year



## Criminal Sexual Assault and Sex Offense each Year



### Conclusion

We feel that our data analysis effort has provided us with a factual assessment on the level of security and crime in the city of Chicago. We may see the most common crimes and the most common places where crimes took place based on the analysis's findings and visualisation. According to these data, theft, battery, criminal property damage, and drug use accounted for 65.7% of all reported offences. Crimes are most frequently committed on streets, sidewalks, homes, and apartments because these are the places where most people are. For example, we especially looked at theft, homicide, and sexual offences to see how they have changed over time.

Even though there were many crimes recorded in Chicago every year, the arrest rate was only around 50%, which led us to assume that the city's police tactics for making arrests and conducting investigations were insufficient. If our data analytics can provide us with all of this information about the security situation in the city of Chicago, we believe a larger data analytics project will do the same and provide much more valuable data that can be used as a powerful source for making informed decisions that improve the security situation in our cities.