



UNIVERSITY  
*of*  
GREENWICH

**ADETUNJI DAVID OGUNDALU**

**Student No: 000963223**

**MSC Big Data and Business Intelligence**

**Final Project**

# Chicago Crime Analysis for Public Safety

**Supervisor:** Mohammad Majid Al-Rifaie

**Submission Date:** September, 2012

**Word count:** 11,674

# Abstract

Crime is the general scenario of nowadays world which is harmful to mankind for sustainable livings and living peacefully. To stay in the world peacefully, the crime scenario should be reduced and try to make it minimized so that to make the environment full of peace. The Crime records maintained by the respective national police department. Like the other places, Chicago is one of those which have a handy crime record and the record is maintained by the police department and make it global so that the historical, as well as the updated records, can be found for the analysis of the crime. This paper deals with an in-depth analysis of the Crime scenario of the city of Chicago. The objective is to analyse of the data provided by the Chicago Police Department so that the perfect crime scenario can be obtained through it. Additionally, with the application of Machine learning, the Classification and Prediction of the Crime is done in the paper to clarify the insight of the crime data. All the variables declared in the data are analysed thoroughly to make the model more compact and to visualize the crime behaviour more perfectly with respect to the other analysis already exists.

---

*Table of Content*

---

<b>List of Figures .....</b>	
<b>List of Tables .....</b>	
<b>Abbreviations .....</b>	
<b>Chapter-1: Introduction .....</b>	<b>7</b>
1.1 Crimes in Chicago .....	8
1.2 Research Objectives .....	10
1.3 Research Questions .....	11
1.4 Impact on Society .....	11
<b>Chapter-2: Literature Review .....</b>	<b>12</b>
2.1 Historical Overview .....	13
2.2 Remedies Taken .....	14
<b>Chapter-3: Methodology .....</b>	<b>20</b>
3.1 Types of Data .....	21
3.2 Data Collection .....	23
3.3 Analysis Methods .....	24
3.4 Machine Learning Algorithm .....	24
<b>Chapter-4: Data Analysis .....</b>	
4.1 Data Overview .....	26
4.2 Data Variables .....	26
4.3 Programming Language .....	28
4.4 Application of Machine Learning .....	29
4.5 Feature Extraction .....	29
4.6 Analysis of Data .....	30
4.7 Result .....	41
<b>Chapter-5: Conclusion .....</b>	<b>42</b>

---

*List of Figures*

---

**Fig-1: Crime Statistics in 2019**

**Fig-2: Crime Statistics comparison of last 5 years**

**Fig-3: Percentage Change in Crimes**

**Fig-4: Percentage Change in Kidnapping Crime**

**Fig-5: Percentage Change in Kidnapping Murder**

**Fig-6: Percentage Change in Robbery Crime**

**Fig-7: Percentage Change in Sexual Assault Crime**

**Fig-8: Percentage Change in Theft Crime**

**Fig-9: Percentage Change in Violence Crime**

**Fig-10: Correlation of Crime Data**

**Fig-11: Overall Crime**

**Fig-12: Crime Analysis for Recent Year(2019)**

**Fig-13: Crime Analysis for 2015**

**Fig-14: Crime Analysis for 2016**

**Fig-15: Crime Analysis for 2017**

**Fig-16: Crime Analysis for 2018**

**Fig-17: Crime Analysis for Arson**

**Fig-18: Crime Analysis for Assault**

**Fig-18: Crime Analysis for Battery**

**Fig-20: Crime Analysis for Burglery**

**Fig-21: Crime Analysis for Violence**

**Fig-22: Yearly top five crime**

**Fig-23: Monthly top five crime**

**Fig-24: Crime Scenario of the year 2017**

**Fig-25: Crime Scenario of the year 2018**

**Fig-26: Crime Scenario of the year 2019**

**Fig-27: Arrest rate of the year 2019**

**Fig-28: Arrest rate of the year 2018**

**Fig-29: Arrest rate of the year 2017**

**Fig-30 Arrest rate of the year 2016**

**Fig-31: Arrest rate of the year 2015**

**Fig-27: Arrest rate of the year 2019**

**Fig-28: Arrest rate of the year 2018**

**Fig-29: Arrest rate of the year 2017**

**Fig-30 Arrest rate of the year 2016**

**Fig-31: Arrest rate of the year 2015**

**Fig-32: Crime map for 2015-2019**

---

#### *List of Tables*

---

**Table-1: Comparison of Crime in Chicago in last 5 years**

**Table-2: Percentage Change in Crime in last 5 years**

**Table-3: Correlation data**

**Table-4: Descriptive Analysis(2001-2004)**

**Table-5: Descriptive Analysis(2005-2007)**

**Table-6: Descriptive Analysis(2008-2011)**

**Table-7: Descriptive Analysis(2012-2019)**

---

#### *Abbreviations*

---

## Chapter-1: Introduction

A city as well the nation will stay in peace in the crime rate will be lower enough. Basically, the national police bureau controls the crimes and keep records of the same in a discrete manner. The reason behind the record holding is to keep a track of the crime for the future analysis and prediction so that, the crime can be controlled in an analytical technique. There are different ways to make the analysis of the crime from the record. But with the application of Machine Learning, the analysis can be made easily and efficiently for the predefined algorithm of the specified objectives (Peixin Zhao, 2014).

For the keeping of the crime data, different parameters word. As per the convention, the records are kept in ta structured data so that the future finding will be made easy. The Chicago has different crimes laws to fight against different crimes but the actual scenario is that those law didn't come into the scenario in a single day. With a day by day, month by month and yearly analysis, the Police and Law department of Chicago have decided to deploy such laws. So, that is a long and analytical process(H. Chen, 2004). There are different types of crimes that can be seen in Chicago. In most cases, the Race is the matter of violence and for which huge number of people were made the victims. The black race is still underestimated and most of the violence are made against them(John J. DiIulio, 1996). This is not the new fact but it is continuing from decades. Another major issue of the crime is the Murder which ca be called as the resultant issue of the black race and the race from the other religion who are not supported by the white caste of Chicago. Not only in Chicago, the types of crime are happening in most of the cities of United States with the same issue(Silvia Martorano Raimundo, 2018). The actual cause behind these all is the bifurcation of caste system and thus they make the caste by whitelisted and backward. With this get done, the backward castes are mostly attacked and they get victimized(J. Fagan, 2007).

On the other hand, as a result of the influence of such crime by some internal support, the other crimes like Physical assault, threat, theft, dacoity, homicide, gambling and even the sex trafficking are made executed and most of the time the general peoples are get affected by those type of crimes. So, the Chicago police department were active to protest those crimes in most of the cases with the application of different laws that were imposed earlier (Ludwig, 2007).

### 1.1 Crimes in Chicago

As mentioned in the previous section, there are different types of crimes have been and still are taking place. Those crimes are making the city unhealthy and full of fear in some specific time and areas and thus the general people generally afraid to live a peaceful like(Silvia Martorano Raimundo, 2018). With the implementation of different crimes laws, some of the crimes can be minimized but in a bread view, still the crimes rates are

higher is a categorical view of the crime types(Manski, 2000). The types of crimes in Chicago are briefed below:

#### 1.1.1 Assault

The attempt of assault, may be physically or mentally are the common phenomena in Chicago though the violent crime is decreased in New York and plenty of different cities like Chicago, however there are two distinct reasons to stay the champagne corked. Some of the important are robbery, murder, rape, and assault stay at historic high alerts as the streets of Manhattan, like those of Houston, city of Philadelphia, Detroit, Los Angeles and Chicago stay a lot of less safe these days than within the earlier 1950s and 1960s(Wilson, 1997). Worse, although policing and jail policies matter, nothing affects crime rates quite the number of young males within the population and by the year 2010, there'll be regarding 4.5 million additional males age seventeen or below than there have been in 1990: almost 8 % more whites and about 26 percent more blacks. Since around 6 % of young males prove to be career criminals, in keeping with the historical information, this increase can place associate calculable 270,000 additional young people on the streets in about 1990's, reverting back at USA in waves over succeeding 20 years(R. J. Sampson, 1997). Long and Continuous studies show that every succeeding generation of young male criminals commits regarding 3 times the maximum amount serious crime because the one before it: the occasional fatal scrap of street gangs of 1950s has given path to the frequent drive-by shootings of Nineties gangs. In the same pace of crimes, another important crime that happens frequently is the Sexual Assault for which the not only the women but also males are harassed (Mayer, 1990).

Page | 8

#### 1.1.2 Homicide

Another type of crime that is frequently observed is the homicide that is to murder a person by other. In most of the time in the crime time line, the homicide has been taken place as it is the type of crime that the expected criminal will not execute rather, they use a third person to execute it(S. Fazel, 2007). One of the critical issues for the homicide crime is the application of sleeper cell who generally works as per order. Most of the time this type of criminals cannot be properly identified and so the criminals get space to do the crime in a continuous fashion (Håkansson, 2011).

#### 1.1.3 Gambling

One of the serious issues of Chicago is the gambling. It is the game in terms of money and so there is the high possibility of having the problem with the players engaged which can be forwarded towards kidnapping and murder. From several decades, this tradition of gambling is going on in Chicago and in most cases, it cannot be prevented as for the political backbone. The place, where the gambling is done is basically a place where the criminals born(Tonry, 2019).

#### 1.1.4 Trace pass & Damage

In Chicago, one of the important and mentionable crime is the trace passing to other's property and damage it. In most of the cases it was observed that, while



executing the crime, another crimes scene may be taken place such as the robbery, theft and may be murder. Is may be treated as a type of homicide if that will be done by any third party(Hofer, 2019). The original criminal always takes the safe side for the crime scenario to make them masked.

### 1.1.5 Kidnapping

One of the widely executed crime in Chicago is the kidnapping. For the advantages of the criminal in terms of money, they do execute the crime and the victim may be male or female. In most of the cases the victims are the young girl who are kidnapped and exploited and hence may be trafficked which has the high possibility after the kidnapping of young girls (S. F. Greenfield, Substance Abuse in Women,, 2010). A huge number of records are found for the same in Chicago Police department and hence it is the critical and common scenario all over the world. Chicago police is now active to demolish such crime to b happened by using the proper crime laws and by getting the shooting orders.

## 1.2 Research Objectives

In this research paper, the in-depth analysis of the crime data and report will be done. The primary objective is to obtain the actual and official data for the crime from Chicago Police department. As the data download is not enough for the data analysis, so the data insight will be pulled up so that the objective that is incorporated inside the data can be understood. Then the data analysis with the application of Machine Learning will be done(John J. DiIulio, 1996).

Basically, there are different attributes in the report data. As the machine learning algorithm will be applied for the analysis of the data from the identification of crime and to make some prediction on it with respect to the variable defined, the data pre-processing is essentially required to minimize the error in analysis(Tonry, 2019). So, to get the perfect result, the data attributes to take wisely. So, on the way to achieve the result, first the data attributes to be taken into consideration. The important data variables in the report file are(Wilson, 1997):

- Case Number
- Block
- IUCR
- Primary Type
- Description
- Location Description
- Arrest
- Domestic
- Beat
- District
- Ward
- Community
- Area
- FBI Code
- X Coordinate Y Coordinate
- Year

While the analysis, the attributes to be taken carefully by correlating those so that the relationship can be found distinctly. This choice is because of the reason that, if some attributes will be chosen for analysis but if it will have no effect of the data and hence the analysis, then the analysis will be let into the imperfection which will cause the impractical result. The above attributes are important for the analysis (Mayer, 1990). Case number is the identification number allotted for each case is booked with which the case or crime can be tracked by location and type. IUCR is the special code through which the crime type can be identified and it is provided by the police department. Additionally, the Primary type crime attribute will help to check the category of crime that is being executed. The location description will describe in which location the crime was executed and it can be tracked by the coordinate attribute values (S. F. Greenfield, Substance Abuse in Women, 2010). In that case, the crime location can be visualized by map system. Now, to check the police activity, the three attributes are there namely, arrest, domestic and beat. The arrest attribute describes the number of criminals being arrested after launching the case and the number of beats with the police is described by the beat attribute. And all the crimes happened can be categorised by the district attribute. The year wise data is available in the year attribute. So, these are the analysis parameters of the project (S. Fazel, 2007).

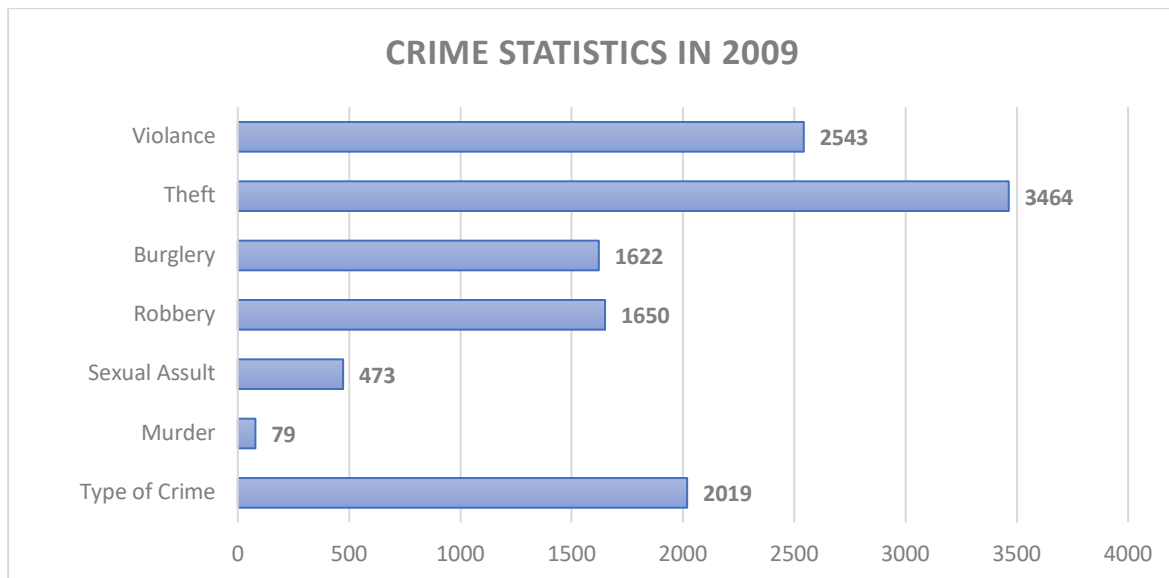
The typical objective of the research is chosen to determine the high number of crimes area wise and through which the prediction will be done for the determination of future crime. As the categorical and the continuous data are available, so, the prediction will be done by means of classifying the data and analysing the subset of the data which corresponds to the year. That is, by creating the data-cluster and with the application of classification and regression algorithm, the entire crime data will be analysed to get a real scenario of crime that has been happened in Chicago (Wilson, 1997).

For years of analysis from the different news and anti-crime agencies, the survey for the crime for the last 5 years shows the percentage of crime happened with percentage comparison for the same to understand the crime that have been taken place to the remedies taken (Chicago Police, 2019). The remedies can be observed if the percentage change of the year comparison to a particular crime type. If percentage change is positive, it will reflect that the crime is increasing in Chicago and if the percentage is negative, it shows that the crime is decreasing in Chicago. The survey data is shown in the following table, Table-1.

**Table-1: Comparison of Crime in Chicago in last 5 years**

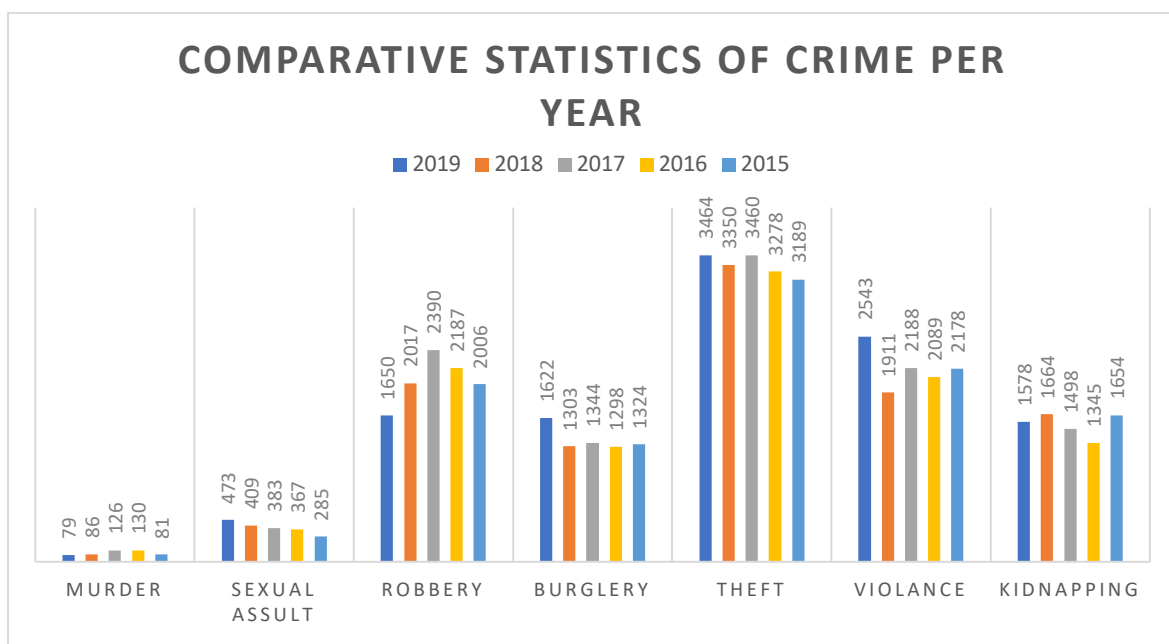
<b>Comparison of Crime Records in Chicago</b>									
<b>Type of Crime</b>	<b>2019</b>	<b>2018</b>	<b>2017</b>	<b>2016</b>	<b>2015</b>	<b>% Change (2018 - 2019)</b>	<b>% Change (2017 - 2018)</b>	<b>% Change (2016 - 2017)</b>	<b>% Change (2015 - 2016)</b>
<b>Murder</b>	79	86	126	130	81	-8.14	-31.75	-3.08	60.49
<b>Sexual Assault</b>	473	409	383	367	285	15.65	6.79	4.36	28.77
<b>Robbery</b>	1650	2017	2390	2187	2006	-18.20	-15.61	9.28	9.02
<b>Burglary</b>	1622	1303	1344	1298	1324	24.48	-3.05	3.54	-1.96
<b>Theft</b>	3464	3350	3460	3278	3189	3.40	-3.18	5.55	2.79
<b>Violence</b>	2543	1911	2188	2089	2178	33.07	-12.66	4.74	-4.09
<b>Kidnapping</b>	1578	1664	1498	1345	1654	-5.16	11.08	11.37	-18.68

From Table-1, it can be found that, some of the crimes has the negative change over the recent years and some other has the positive change. So, as per supervision of the police and the anti-crime bureau, some of the crimes are under control and some are still processing. The overall scenario of the crime statistics in the recent 5 years are shown below:



**Fig-1: Crime Statistics in 2019**

The statistics of last 5 years is as follows:



**Fig-2: Crime Statistics comparison of last 5 years**

With that comparison, it can be found that, somehow, the theft sustains the most compared to the other types of crime.

### 1.3 Research Questions

In this research, the analysis will be done with the application of machine learning algorithm. While starting the analysis, the research question will be taken into consideration on the basis of which the analysis and prediction will be performed. Research question is essential in research as it depicts the clear objective of the research orientation. So, in this paper, the research questions are chosen to clarify the research (H. Chen, 2004). As the objective of the research is to analyze the crime data year wise and thereby the prediction will be made, so the research questions are as follows:

***Research Question-1:***

Which area of Chicago is mostly affected by the crime?

***Research Question-2:***

What is the effect of crime in Chicago and how to make the remedy of it?

The solution to these questions is described in the analysis section where the answer to the research questions can be found through analysis and with the application of machine learning algorithm.

### 1.4 Impact on Society

Crime is the serious issue of the society which harms the society and left a negative impact on the public living. The action against the crimes can be taken by means of good analysis. The objective of the analysis is the finding out the density of crimes in different places in Chicago so that it can be clearly understood the most crime prone area and thus the protection should be given there to prevent the probable crimes (H. Chen, 2004). If the target will be successful, the rate of crime will surely to be decreased notably. This requires a thorough analysis which are done in this paper in the analysis section with the power of machine learning classification and regression algorithm through which the crime prone zone can be predicted and detected and so it will be easier to detect and be protected to those zones to attain the public safety.

## Chapter-2: Literature Review

It is being a long history from when the crimes are committed in Chicago. It was early 1800s when the detection of crime was taken place in the record with a motive to keep the crimes in track and to eliminate or to minimize that to the level best. But in the earlier age, there was no electronic device or even calculator was not there so to compute the statistics of the criminal issues. As the days are going on the technology was incorporated with the social life and the criminal jurisdiction(Inqilab Shahbazov, 2019). In the later 1900 era, with the invocation of computer and the editing software, the records were being kept in a software document file so that it can be accessed globally by using the internet. Moreover, with the implication of the software programming language, the calculations were made easy and everything were started to be done and computed using programming(John J. DiIulio, 1996). Machine learning is not a new study. It is the accumulation of different algorithm. The research on the algorithm were started from about 1870s but then the fewer algorithms were in use, In the middle 1900s, the research on the algorithm went faster and then different new algorithms were come into the field of statistics. After the application of computer and software language, the practical implementation was found(Schröder, 2019).

In the modern technological era, the algorithm which were capable of computing the statics, are in use to implement the machine learning model. Now a days, as the demand is still increasing for data science and machine learning, different language was being innovated and one of such language is Python. Previously, different analysis was made by machine learning for the crime data analysis using python. The analysis of crime is one of the hot-cake in the present market. The analysis of the crime scenario for Chicago is the most famous of all the crime data available(S. Fazel, 2007). This is because of the fact that, Chicago is one of the most cultured city or different issue and one of those is the criminal activity. Another reason is that, the availability of data in open source facility so that anyone can download and access the data for analysis. The third reason is that the structure of the data. Chicago crime data is one of suitably structured so that the analysis will be made easier and the records found in the dataset is mostly discrete and the number of observations in the dataset is large enough and for this reason, the discrete analysis is possible(ChicagoPolice, 2019).

## 2.1 Historical Overview

The historical perspective of Chicago is more important to understand the background of the crime in Chicago. For many years the Chicago was cultured for different facts like the movies, songs, in house play like casino and also famous for the Bar and Restaurants. These are the common acts where the problem scan be generated and this were most probably the reason behind the generation of crime because of high paid persons. Another critical issue of Chicago is the sex trafficking. As the number of casino and the palaces are there in Chicago, the sex racket is inevitable there and for this the crime scenario like sex trafficking and sex

crimes are also cause a massive part of the crime scenario in Chicago(Klaus Bachmann, 2019).

Chicago's criminal name long preceded gangster and also the brewage wars. Born within the same years because the ballyhoo artist penny press and Americans' new fears of a lord less social class, town appeared even to its earliest observers a hotbed of crime and immorality(Håkansson, 2011). The stories of rape, murder, arson, theft, and different mayhem stuffed the frontier town's several newspapers. In its 1840 grievance that “the business of stealing horsemeat,” has been “reduced to an everyday system,” the apsis echoed the perception of unnumbered Chicagoans. Constant year, the city's initial hanging—with a pair of 500 reportedly in attendance who confirmed the notice and readability. By the top of the decennium, observers each among town and on the far side often noted the existence of a diagnosable criminal underworld. Within the words of the Democrat, it had been “getting to be a disreputable undeniable fact that robbers, pickpockets, thimble riggers are utterly reception in our town”. The visibility of vice increased the city's criminal name. At mid-century, Chicago reportedly had additional gambling institutions than the larger town of metropolis and more per capita than the big apple. Vice initial targeted in a region on the Chicago watercourse referred to as “the Patches,” places, because the apsis places it, of “the most beastly sensual fact and darkest crimes” (Klaus Bachmann, 2019).

So wicked was the city's name that several saw the hearth of 1871 as divine retribution against a modern thrust and metropolis. A newspaper reported about the situation, “The city,” as “is troubled with a horde of thieves, burglars and cut-throats, out to plunder, and who won't hesitate to burn, pillage and even murder”(Wilson, 1997). The fire could not provide the redemption, and Chicago's name darkened within the late 1800s. The labours of violent labour got disputed, specially the crisis of Haymarket which added to the scenario of the lawlessness. By the Nineties the disreputable dike vice district attracted criticism—and visitors—from round the world. Chicago “makes a additional surprisingly open show of evil than the other town proverbial to Pine Tree State,” a visitant from London exclaimed. “Other places hide their blackness out of sight; Chicago treasures it within the heart of the business quarter and provides it a veneer” (Tonry, 2019).

The killing of vice leader Jim Colosimo in 1920, the primary year of national Prohibition, signal a brand new innovate Chicago violence. The bloody brewage wars of 1924–1930 created gangster noted and also the town substitutable with the new development of gangsterism. Chicago's infamy grew in an exceedingly series of violent episodes: the 1924 shooting of gang leader Dion O'Banion in his side flower look, the 1926 machine-gunning of kike Weiss on the steps of Holy Name Cathedral, the 1929 St. Valentine's Day Massacre of seven men in an exceedingly Clark Street garage(Tonry, 2019). Wide reported within the national and international press, these incidents were the topic of widespread modern books and plays. Even additional vital, *Underworld* (1927), *Public Enemy* (1931), *Alphonse Capone* (1932), and innumerable lesser films broadcast to eager audience's dramatic tales of Chicago crime scenario. In concert journalist place it in 1930, “In all the body of water and also the lands bordering on that there's most likely no name that additional quickly calls up thoughts of crime, violence and wickedness than will that of Chicago” (Mathew Zaia, 2019).

Yet even throughout the twenties and thirties, Chicago's levels of violence and vice were ne'er particularly high. Instead the city's name was a matter of story and image. For Chicago—in its

booming growth, unrestrained energy, and typically explosive conflicts—symbolized for several the guarantees and perils of America's urban future. In its “excessiveness”, One author on crime explained, Chicago “is like different yank cities—only additional thus” (Vincenzo Ruggiero, 2019).

## 2.2 Remedies Taken

Page | 16

For year long, with the accumulation of data and the reports launched, the Chicago police department as taken the serial steps against different crimes. As per the decision and analysis of the available data it was found different crimes ware been taken massively and thus those were considered as most sensitive to Chicago and thus serious and organised steps were been taken against it.

It was seen that, from the year of 2006 and within the range of 2008 the violence were been taken place massively and the culture were still continuing. It placed an massive effect on the general life of the people and thus it was the great concern for the police and anti-crime bureau of Chicago. Not only the violence, from 1990s, the rise of big gangs creates massive nuisance in the general life in the Chicago. The criminals were with their guns and thus no one dared to protest against it(J. Fagan, 2007). In view of eliminating those crime, the strategy of Chicago police department changed their strategy to handle the situation. Previously, the police were acting as the group activity but that was a failure to the police department as for the communication gap of the police department of different places(H. Chen, 2004). Thus, the implementation was done for the hotspot of the police which helps the information sharing of the crime report. The technology helps to share the instant information about the crime happening in different places with the application of machine learning which act as the backbone of the technology. This was implemented because it was found that about 70-80% crimes in Chicago deals with the Violence which are closely related to gang crime. So, the measures were taken carefully the police department of the way to eliminate or minimize the percentage of crime happened(Guangxing Zhu, 2019).

The other step was taken by the Chicago Police is to focus on repeat offenders. Dispiriting although it's to modify a similar individuals time and once more, the goal of targeting the habitual gun carriers and imposing stricter penalties on those condemned can hopefully send the message(Peixin Zhao, 2014).The policy isn't while not its detractors, however in an exceedingly climate wherever one thing should be happened, it a minimum of are a few things. Not amazingly police morale hit a coffee in an exceedingly culture of mistrust. Following the excessive shooting of a black juvenile person by a politician, the ripples flowed through both communities; anger, outrage and sorrow within the black community, firings and a possible foreign policy as an integrated part of the Chicago police(ChicagoPolice, 2019). Steps to boost trust and community relations are being adopted by the Chicago police on one aspect and leaders of the community on the opposite. Each acknowledge there's an extended profit to go. The police conjointly restructured the authority policing the police. This may currently be active body that has the large strength than the interior board. Another try is to provide officers on the scene choices, with Tasers for instance. Lastly, the commission intends to feature to the force. Not amazingly the quantity of police on the road has diminished because the danger inflated. It was a decent begin. The voters of Chicago are attempting to require back their streets, schools, churches, and stores and create them places of safety and luxury yet again. Within the wake of such staggering amounts of bloodshed, the



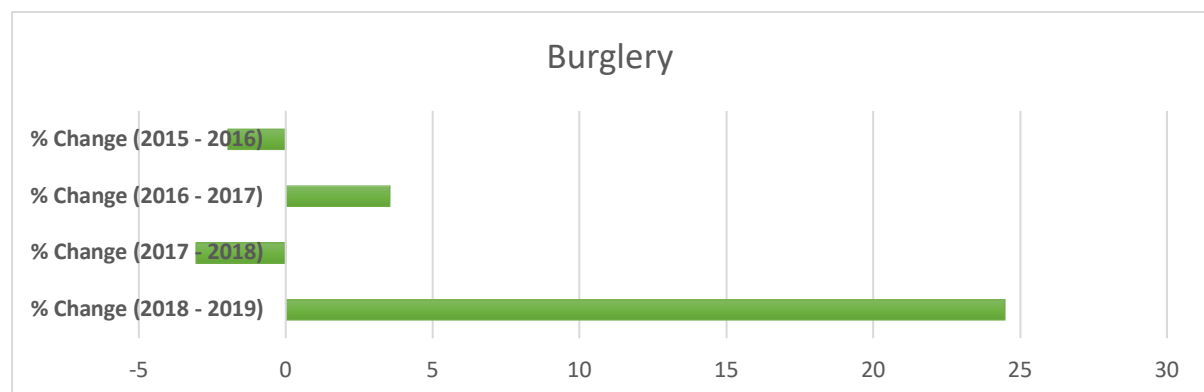
individuals are uninterested in living in concern, and are rallying along to assist enact changes to Illinois gun laws. The problems encompassing gun reform in our country are as advanced because the those who wield them(Schröder, 2019). The report doesn't apprehend that they feel anyone extremely is aware of the correct answer to the matter. It was apprehended that it has to be undertaken to unravel it. Youngsters ought to never need to be afraid to measure and learn. Once the truth is that they aren't safe in their colleges and community parks as a result of they could get hit with wide bullet, the time has come back to prevent this violence. There are multitudinous ways that to enact the changes we would like to visualize in our society. Due to the rights they have as voters to enact that modification, they have an obligation to our future to try to thus(S. Fazel, 2007).

To eliminate or to minimize the occurrence of the, the steps are taken by the police department and the outcome can be reflected in the following table(ChicagoPolice, 2019).

**Table-2: Percentage Change in Crime in last 5 years**

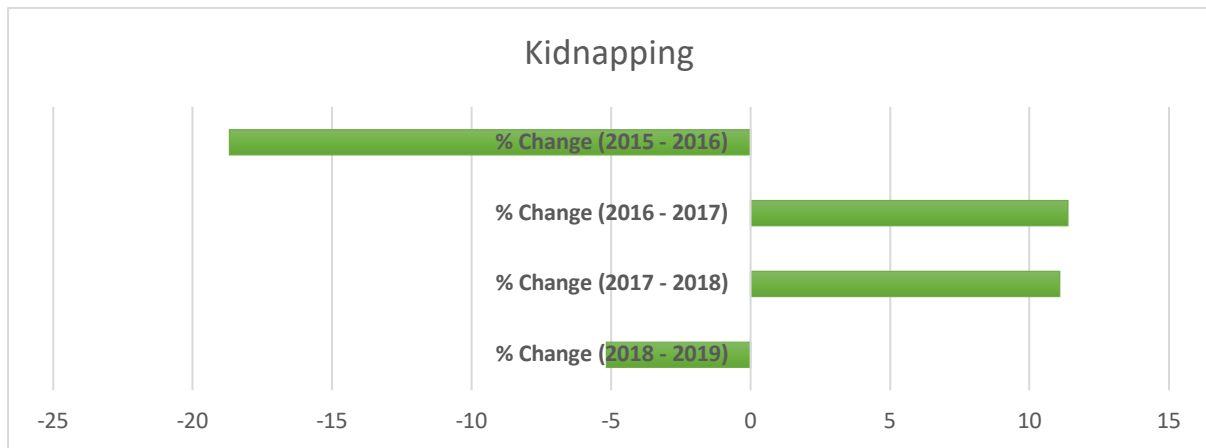
Type of Crime	% Change (2018 - 2019)	% Change (2017 - 2018)	% Change (2016 - 2017)	% Change (2015 - 2016)
Burglary	24.48	-3.05	3.54	-1.96
Kidnapping	-5.16	11.08	11.37	-18.68
Murder	-8.13	-31.74	-3.07	60.49
Robbery	-18.19	-15.60	9.28	9.02
Sexual Assault	15.64	6.78	4.35	28.77
Theft	3.40	-3.17	5.55	2.79
Violence	33.07	-12.65	4.73	-4.08

In this analytical report, the percentage change in the specific crime type is shown. The percentage change of the crimes for 2018 – 2019 is as follows:



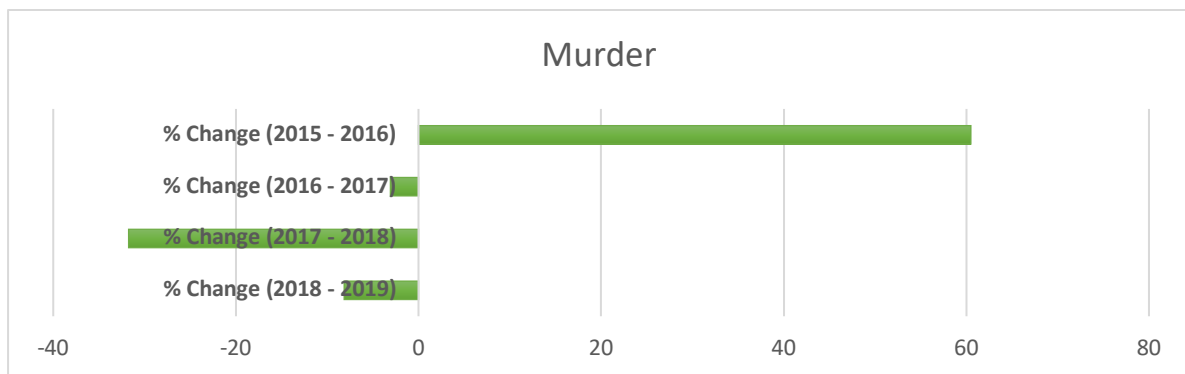
**Fig-3: Percentage Change in Burglary Crime**

The highest percentage change is found for the Burglary in 2018-2019. The change is positive that means, the action was taken against it and so the crime is somehow minimized one the year. The percentage change for Kidnapping is shown below:



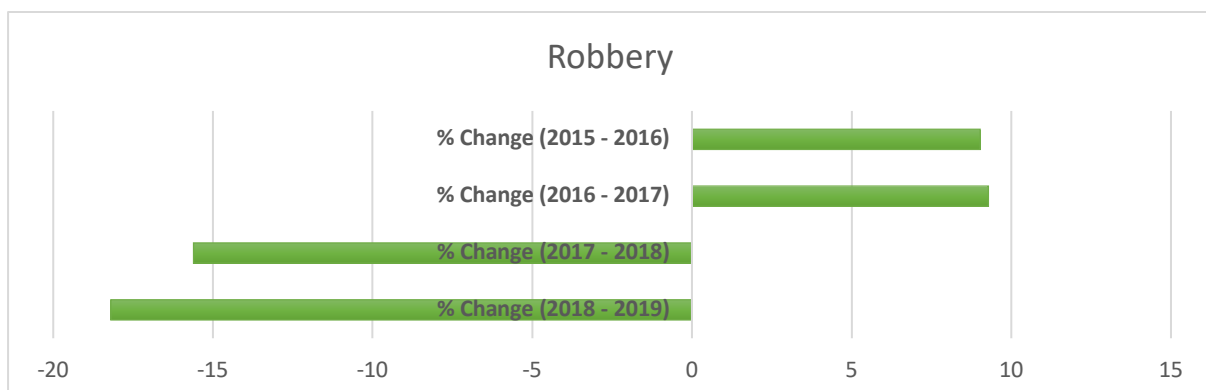
**Fig-4: Percentage Change in Kidnapping Crime**

From the analysis, it can be said that, no special steps were taken for the kidnapping and so the percentage of change is still in negative in 2018-2019 and the steps need to be taken by the police. The percentage of change of murder is shown below:



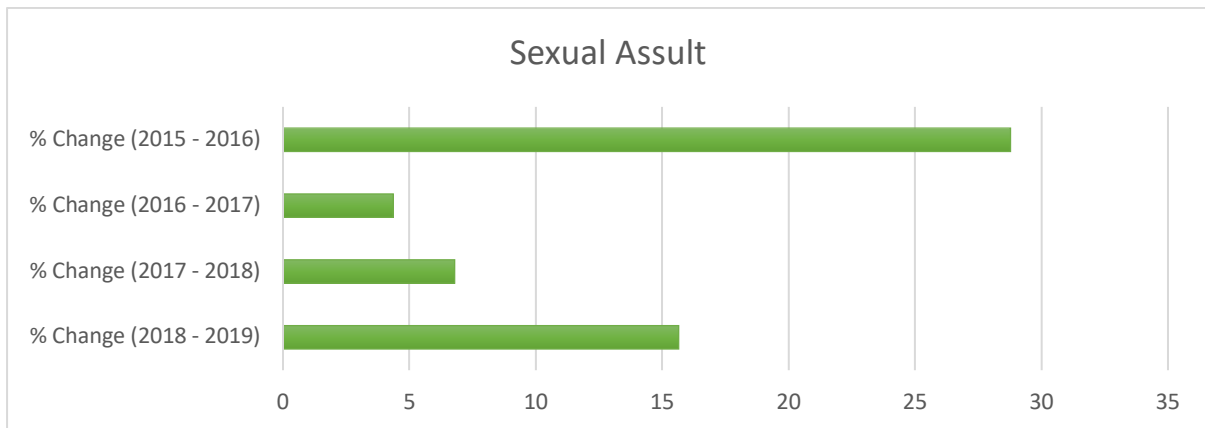
**Fig-5: Percentage Change in Murder Crime**

In this analytical report, it is clear that the necessary steps were taken against murder and so the significant change in percentage of that crime is being visualized. The percentage change in Robbery is shown below:



**Fig-6: Percentage Change in Robbery Crime**

Till now the analysis are shown, the percentage change is the highest for the case and that means the necessary steps were taken by the Chicago police. The percentage change for the sexual Assault is shown below:



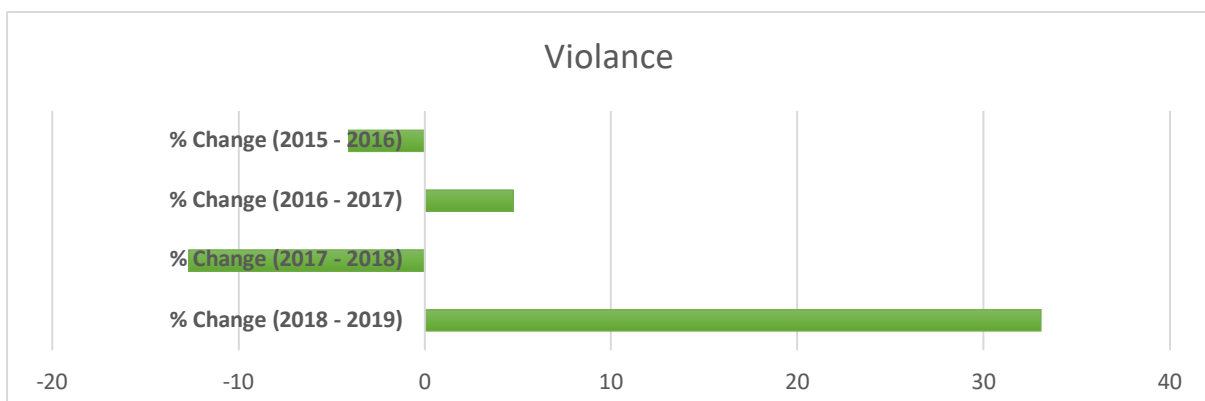
**Fig-7: Percentage Change in Sexual Assault Crime**

There were no specific steps were taken against the Sexual Assault and thus the percentage change is high positive value which indicate that this type of crime rate is increasing and if no step is taken it will be higher enough in future years. The percentage of change for theft crime is shown below:



**Fig-8: Percentage Change in Theft Crime**

In this case, still police department of Chicago is little bit inactive and need to take necessary steps in future to eliminate the crime. The percentage change of violence is shown below:



**Fig-9: Percentage Change in Violence Crime**

Like the former crime, no specific steps were taken against the violence whereas this type of crime is happening from many decades and the Chicago police should take steps to minimize the crime in Chicago.

So, in this paper, the analysis of the data will be done in view of the crime analysis and to focus the most cultured crime as depicted in the research question. The next section deals with the research methodology which will describe the data collection method for the analysis and the analysis technique with the implication of machine learning.

## Chapter-3: Methodology

This research paper will be meant for the analysis of the crime on the basis of the data. While doing the analysis, the data is required which is mentioned earlier section and that is the crime data. The procedure of the analysis will be determined after the successful fetching of the data from the repository. Basically, there are different sources of data which can be obtained in various way with different structures. The best structured data will be chosen to obtain the best result. That requires a survey for the search of data (Inqilab Shahbazov, 2019).

Practically, as the crime issue is focused on the Chicago and which is provided by the Chicago police itself, so the data can be collected from their respective repository. After collection of the data, the required analysis will be done. But before proceeding to the analysis, the type of data should be determined so that, it will be easier to find the attributes and the analysis can be performed accordingly (Mathew Zaia, 2019).

### 3.1 Types of Data

There are basically two types of data available on the basis of the source from where it can be obtained and they are namely, Primary data and the Secondary data. The main issue behind the categorization is to classify the types of the content and usage of the data. The types of data on the basis of the source and purpose are detailed below:

#### 3.1.1 Primary Data

Primary data is collected by the investigator itself or by his or her team for some specific purpose. This can also be referred as the survey data as it is being collected by means of survey. For example, if a student has to make a new thesis paper where new kind of data is required from the available sources in the world, this can be referred to as the Primary data. This data is basically a unique data and is to be used for the first time by the investigator who intends to do some new and unique project with new data (Marlene Matos, 2019). The advantage and disadvantages of the Primary Data are as follow:

##### 3.1.1.1 Advantages of Primary data:

1. The investigator usually collects the data from the survey for his or her project and that data is unique for his or her desire.

- T2. As the investigator is collecting the data by survey, the quality of the data will be good as that is collected by the selection of the parameter that the investigator have chosen.

3. The additional data, if required, can be collected on runtime of the data analysis by the investigator.

#### 3.1.1.2 Disadvantages of Primary data:

1. The investigator should collect the data with all possible attributes required, if anything is missing which he or she planned for, the analysis will be not done as per the desired.

Page | 22

2. It may require the incorporation of the funding agency who will provide the moral and economical support for the data collection. In return the agency may demand for the royalty.

3. The investigator should take care of the data quality in terms of the attribute collection and the number of observations.

4. In this data collection, no unnecessary data will be included and that means while collecting the data, the attributes should be considered and planned before the collection of data.

5. For the fulfilment of the data collection, no fake data should be included in the dataset otherwise it will provide the wrong result.

6. For this data collection, permission from different sources may be required from where the data can be collected. So, to collect the data properly, the permissible source should be found out.

#### 3.1.2 Secondary Data

Secondary data is actually collected, built and processed by some other for their own purpose by is utilized by the investigator or the researcher itself to fulfil his or her requirement. Actually, the secondary data is available in different websites and repository. With the objective of the research, the data can be selected (Inqilab Shahbazov, 2019). The secondary data is also collected by someone primarily to fulfil their work and after the execution, the data will be uploaded into some repository from where, with the similar objective, the data can be collected for some analysis. The advantage and disadvantages of the secondary data are shown below:

##### 3.1.2.1 Advantages of Secondary data:

1. As the data is already collected and processed primarily, so it is already a processed data and also an almost clean data which help to make the analysis faster and efficient.

2. As to collect the data, only the searching effort is required and the massive effort like the real collection of survey data is not required, so it is less expensive with compared to the primary data.

3. If the quality of the data is somehow deviated, the investigator will not take any responsibility for it as the analysis will go wrong for the data attributes and that is the passive advantage of the Secondary Data.

4. Secondary data of same topic will be available in the repository with different structure and different architecture and so the investigator can choose the most relevant data with respect to his or her research. So, this one is the added advantage of the research having Secondary Data.

### 3.1.2.2 Disadvantages of Secondary data:

1. There will be no control of the data and its structure as the investigator is not designed and structured it. So, as per the data is given, the analysis will be based upon this. So, the investigator cannot decide the analysis pattern and type as per their choice(Hofer, 2019).

2. The investigator has to depend on the quality of the data as they cannot change the structure of the data as the data is directly fetched from the repository link.

3. If, during the analysis, extra data is required, it will not be possible to be added with the existing data as the structure may be different.

4. Once the research model is designed using the Secondary data, that cannot be usually modified for the lack of flexibility of data availability. So, the research and its corresponding analysis will be remaining static.

Comparing these two data types, the Secondary data will be chosen for its two main advantages namely, the less expensive and can be collected with less effort. So, the research analysis will now be based on the Secondary data and hence first the data will be collected from the respective repository and the analysis will be made upon that data.

## 3.2 Data Collection

The Crime data will be collected from the Chicago police department repository where the data is uploaded periodically. The same data is also available in other websites but they are not the real time or dynamic data rather, those external mirror collects the data from the same Chicago police department repository and do their analysis. But those data can be outdated and hence the update data cannot be found most of the cases(John J. DiIulio, 1996). So, the Chicago police department repository is chosen for the betterment of the analysis and prediction.

For the crime data analysis, the quantitative data is required and thus the quantitative data is downloaded and collected from Chicago police department repository official website(ChicagoPolice, 2019).

### 3.2.1 Collection of secondary Data

Secondary data is particularly a kind of data that has been published in some newspapers or books or magazine or journals or any other on-line portals etc. there's an abundance of information offered in these sources concerning your analysis space in business studies, nearly notwithstanding the character of the analysis space. Therefore, application of acceptable set of criteria to pick secondary information to be utilized in the study plays a vital role in terms of skyrocketing the degree of analysis validity and dependableness(S. F. Greenfield, Substance Abuse in Women,, 2010). These criteria embody, however not restricted to this point of publication, credentials of the author, dependableness of the supply, quality of discussions, depth of analyses, the extent of contribution of the text to the event of the analysis space etc(Hofer, 2019).

#### 3.2.1.1 Quantitative Data Collection

Quantitative data assortment ways are based mostly in mathematical calculations in varied and typical formats. The techniques of quantitative data accumulation and analysis embody

the questionnaires who is with limited queries, ways of statistical implication like regression, correlation, mode, mean and median and all the others.

Quantitative ways are much cheaper to be used and that they are often applied among shorter length of your time compared to qualitative methods. Moreover, because of a high level of standardisation of quantitative ways, it's simple to create comparisons of findings.

### 3.2.1.2 Qualitative Data Collection

Qualitative analysis methods, on the contrary, don't involve numbers or mathematical calculations. Qualitative analysis is closely related to words, sounds, feeling, emotions, colours and different components that are non-quantifiable.

Qualitative studies aim to make sure larger level of depth of understanding and qualitative data assortment ways embody interviews, questionnaires with open-ended queries, focus teams, observation, game or role-playing, case studies etc.

In this project, the data is collected from the Chicago Police portal. As the data is designed by the Chicago police, and that same data is used in this analysis, so this can be considered as the Secondary data. If the data needs to be designed manually, a lot of time will be wasted and so this technique will be rejected in this case for smooth conduction of the analysis and additionally, as the data is readily available in the repository, so it is meaningless to design and make survey of the same data. So, in the project, the secondary data is used and the analysis will be made accordingly.

## 3.3 Analysis Methods

There are different types of analysis are available for the same dataset. While collecting the data, the variables are observed for the understanding the data insight which will help to decide the technique of analysis. The collected dataset has the record from 2001 to the present date and that means a huge record are available in the dataset. The general observation number for the dataset is about 43 lakhs. So, this is basically a huge and big data. The analysis method that will be applied in the research paper is described below:

### 3.3.1 Steps for Analysis

The steps for the analysis of the dataset is as follows(Tao Wang, 2019):

1. Obtain the dataset
2. Check the data profile
3. Check for missing data
4. If there is any missing data, the data cleaning operation will be performed.
5. To check the interrelation of the variables, correlation is done to understand the exploratory analysis which is essential for the crime prediction and analysis.
6. Classify the crime data depending upon attribute with:
  - i. One variable
  - ii. Two variables
  - iii. Multiple variables
7. Execute the crime result to show the location result with clusters.
8. Summarize the crime result and proceed to predict for the future crime.



These steps will be followed in the analysis section for the analysis of the crime data starting from the year 2001 to the present date(Qinglan Zhao, 2019).

### 3.4 Machine Learning Algorithm

To analyse and visualize the crime data, the machine learning approach is applied in this paper. The approach of the machine learning algorithm is appreciable in that case because with the use of the algorithm the analysis of ten crime data will be easier as the algorithm and the statistical interface are already incorporated in the library. As, Python is chosen as the language to analyse the data, it has the rich library where the algorithms are already incorporated and so the new implementation of the algorithm is not required and so the operation will be smoother and more efficient(Saruar Alam, 2019).

Page | 25

There are different algorithms are available in the machine learning domain. The algorithms are classified by the Classification and Regression. In this data analysis, the classification techniques will be used to classify the data and after the classification is done, the prediction based upon the data will be performed. As the data is already been collected, so the classification algorithm applied will be supervised. So, altogether, the analysis will be done using the Supervised Classification Algorithm on the structured data.

# Chapter-4: Data Analysis

The Crime data that is collected for the analysis is containing total 23 variables and a huge number of observations. Before proceeding to the analysis, the levels and the variables will be analysed to understand the types of the data that are available in the dataset. In the data overview section, the overview of the data and its levels are discussed (Klaus Bachmann, 2019). During the analysis, the identification of the variables is done so that the analysis can be done uniformly with high score value. After the variable selection is done, the feature will be extracted for creating different required subset of the data and to select the required variables for the analysis. Finally, the analysis of the data will be performed, based upon which the decision will be made for the crime happenings in Chicago (MadhuSudana Rao Nalluri, 2019).

## 4.1 Data Overview

In the collected dataset, there are different columns which are considered as the variables for the analysis. For the decision making of the crime data, the following variables are useful as those are providing the numerical result for crime data. The variables are:

- **Arrest:** This variable includes all the Boolean report whether True or False. It means, this variable implies whether the criminal was arrested or not. This variable makes a good impact on the dataset to show the activity of the Chicago police. If the arrest is shown as True, it means for that case, police was very much active which shows the efficiency of the working of the police.
- **Beat:** Beat is the variable which shows the arrested criminal was treated or not. Basically, in most of the cases, the criminals were not well treated by punishment. The number of punished criminal numbers are shown in the variable field.
- **District:** This is also a Boolean variable which shows the activity of the police with respect to the district of Chicago. The True result shows that the criminal of that observation was arrested and being jailed in that district.

- **Ward:** Similar to the district, the ward is the same type of variable which shows the arrest count in the particular ward. The count for this is shown the number of arrested in a particular ward.
- **Community Area:** The count for the community area variable shows how many crimes are treated from some community place.
- **FBI Code:** This variable shows the criminal code provided by the FBI itself. It is the record for the Chicago police as well as maintained by the FBI itself to count for the case launched.
- **Latitude&Longitude:** Latitude and Longitude are two different variables but works together to show the location of the crime happened.

## 4.2 Data Variables

The previous section has dealt with the useful variables that can be used in the analysis and by using those, the fruitful result can be obtained. Though for the analysis of multivariate, some of the columns are chosen so that the perfect analysis can be done. The variables are chosen as follows:

- Arrest
- Primary Types
- Latitude
- Longitude
- District

The descriptive analysis of the data is as follows:

**Table-4: Descriptive Analysis(2001-2004)**

	ID	Beat	District	Ward	Community Area	X Coordinate	Y Coordinate	Year	Latitude	Longitude
count	1048575	1048575	1048575	433808	433758	1027938	1027938	1048575	1027938	1027938
mean	1986269	1223.08552	11.36983	22.5314	37.056029	1164555	1886813	2001.615	41.84501	-87.6716
std	474034	709.638025	6.992544	14.00719	21.509026	16202.47	31502.53	0.627322	0.086645	0.058897
min	634	111	1	1	0	0	0	2001	36.61945	-91.6866
25%	1649161	624	6	10	23	1153339	1860657	2001	41.77307	-87.7123
50%	1977641	1121	10	22	32	1165936	1891354	2002	41.85751	-87.6663
75%	2304970	1821	17	34	56	1176264	1910220	2002	41.90957	-87.6288
max	1.1E+07	2535	25	50	77	1205119	1951622	2004	42.02291	-87.5245

**Table-5: Descriptive Analysis(2005-2007)**

	ID	Beat	District	Ward	Community Area	X Coordinate	Y Coordinate	Year	Latitude	Longitude
count	1048575	1048575	1048573	1048564	1048407	1045226	1045226	1048575	1045226	1045226
mean	5110249	1194.722	11.18865	22.39872	38.14755	1164619	1884989	2006.331	41.84	-87.6715

<b>std</b>	266547.4	702.2081	6.902997	13.73577	21.5674	16270.09	31397.09	0.479088	0.08635	0.059212
<b>min</b>	4673626	111	1	1	0	1092706	1813908	2005	41.6446	-87.9343
<b>25%</b>	4877854	623	6	10	23	1152967	1858632	2006	41.76745	-87.7138
<b>50%</b>	5099819	1034	10	21	34	1166006	1888774	2006	41.85031	-87.6661
<b>75%</b>	5337096	1731	16	33	58	1176401	1908441	2007	41.90446	-87.6282
<b>max</b>	5843197	2535	31	50	77	1205119	1951527	2007	42.02259	-87.5245

**Table-6: Descriptive Analysis(2008-2011)**

	ID	Beat	District	Ward	Community Area	X Coordinate	Y Coordinate	Year	Latitude	Longitude
<b>count</b>	1048575	1048575	1048534	1048557	1047996	1034337	1034337	1048575	1034337	1034337
<b>mean</b>	6852150	1207.172	11.3744	22.69286	37.95508	1164286	1885197	2008.814	41.84058	-87.6727
<b>std</b>	503664.3	706.3792	6.953051	13.71955	21.66959	18332.61	34407.06	0.770323	0.094732	0.065989
<b>min</b>	4785	111	1	1	0	0	0	2008	36.61945	-91.6866
<b>25%</b>	6439107	624	6	10	23	1152637	1858550	2008	41.76715	-87.7149
<b>50%</b>	6861209	1111	10	22	32	1165728	1889522	2009	41.85247	-87.6671
<b>75%</b>	7262896	1732	17	34	58	1176228	1909539	2009	41.90756	-87.6291
<b>max</b>	10549780	2535	31	50	77	1205119	1951511	2011	42.02255	-87.5245

**Table-7: Descriptive Analysis(2012-2019)**

	ID	Beat	District	Ward	Community Area	X Coordinate	Y Coordinate	Year	Latitude	Longitude
<b>count</b>	1048575	1048575	1048574	1048562	1048535	1039870	1039870	1048575	1039870	1039870
<b>mean</b>	9283843	1152.032	11.26615	22.81723	37.60436	1164362	1885286	2013.223	41.84082	-87.6724
<b>std</b>	522668.4	691.9712	6.90616	13.76245	21.45232	19255.95	35326.87	1.079359	0.097306	0.0691
<b>min</b>	20859	111	1	1	0	0	0	2012	36.61945	-91.6866
<b>25%</b>	8842990	613	6	10	23	1152482	1858631	2012	41.76743	-87.7155
<b>50%</b>	9264986	1023	10	23	32	1165955	1890867	2013	41.85619	-87.6663
<b>75%</b>	9705776	1711	17	34	57	1176364	1908680	2014	41.90521	-87.6281
<b>max</b>	10550397	2535	31	50	77	1205119	1951573	2016	42.02271	-87.5245

### 4.3 Programming Language

For the analysis, the Python programming language is chosen for the richness of the libraries. For the analysis, the prebuilt library like scikit learn and pandas are required for the data analytics and classification and by the classification, the future forecasting will be done and

hence the prediction is possible. Moreover, the python is the high-level language which can work faster because of it is the interpreted language. So, the analysis will be made faster and accurate. In this analysis, the map for the crime is also done so that to clarify the location of the crimes are happened and so the crimes location can be well identified and verified from the dataset. It will help the Chicago police service to identify the crime prone zone and to add more services if that particular zone is not well services and that can be determined by the arrest report. So, by the analysis, two separate issues can be solved, namely the crime location and the crime classification. For this operation, the following libraries will be used in the coding:

- import pandas as pd
- import numpy as np
- import matplotlib.pyplot as plt
- import seaborn as sns
- from sklearn.metrics import mean\_squared\_error, r2\_score
- from sklearn.linear\_model import LinearRegression
- import os
- import pandas\_profiling
- import folium
- import webbrowser
- from PIL import Image

## 4.4 Application of Machine Learning

The application of Machine learning is to analyse and classify the crime data but not limited to. In this paper, the data mining done primarily to fetch the data from the data set for the analysis. After the successful fetch of data, the, the required analysis such as the detection and identification of crime type that is the primary type and the values and thus the visualization, the number of arrest in year wise, month wise and weekly and the top crime types so that, the analysis and classification can be done and visualize to show the ultimate result for the analysis and to fulfil the requirement of the paper. So, after fetching of the data, the feature will be selected properly which will helps the analysis will be get done with the respective visualizations.

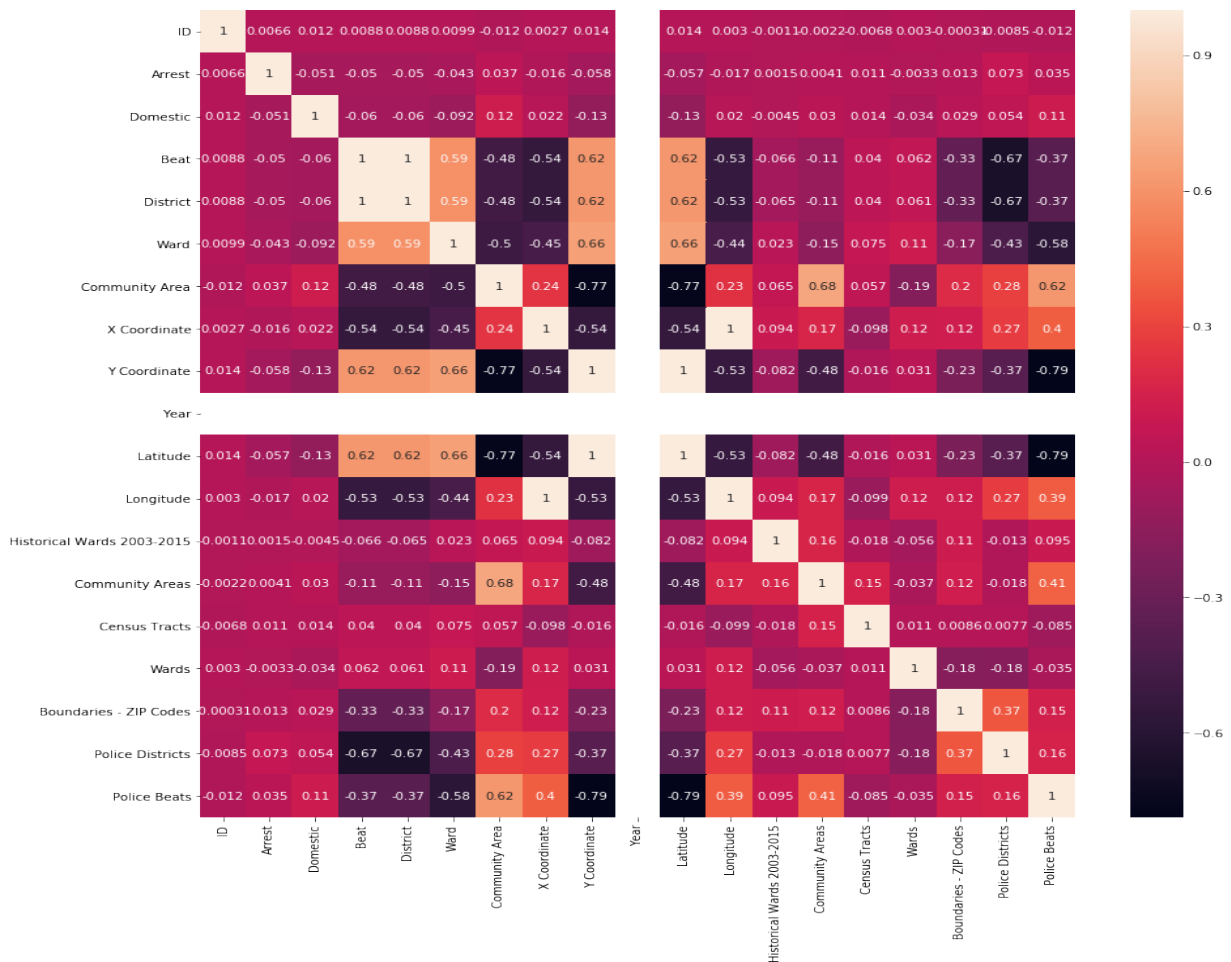
## 4.5 Feature Extraction

After choosing the variables, as described earlier, the feature selection is done. For example, the relationship between the variable will be computer with the application of statistics and machine learning. The analysis is performed using correlation as it signifies the property and relationship between two iterative variables. If the value for the correlation is near to +1, it means that if one variable is increased the other will also be increased. If the value is close to -1, it means, if one variable is increased, the other one will be decreased. If the value is near to 0, there will be no significant relationship between these two variables. The correlation can only be done if the iterative variable contains all the numerical data. The correlation among all the data is shown below:

**Table-3: Correlation data**

	ID	Arrest	Domestic	Beat	District	Ward	Community Area	X Coordinate	Y Coordinate
ID	1	0.00658	0.012413	0.0088	0.0088	0.00992	-0.0122	0.0027	0.0142
Arrest	0.0066	1	-0.05115	0.0498	-0.05	-0.0433	0.03668	-0.0163	-0.05761
Domestic	0.0124	-0.0512	1	-0.06	-0.06	-0.0915	0.11538	0.022	-0.12534
Beat	0.0088	-0.0498	-0.06002	1	0.9999	0.59346	-0.479	-0.5374	0.61643
District	0.0088	-0.0496	-0.06013	0.9999	1	0.59472	-0.4797	-0.5384	0.61766
Ward	0.0099	-0.0433	-0.09152	0.5935	0.5947	1	-0.503	-0.4475	0.65516
Community Area	-0.012	0.03668	0.115376	-0.479	-0.48	-0.503	1	0.2408	-0.77229
X Coordinate	0.0027	-0.0163	0.021973	0.5374	-0.538	-0.4475	0.24078	1	-0.53822
Y Coordinate	0.0142	-0.0576	-0.12534	0.6164	0.6177	0.65516	-0.7723	-0.5382	1

The visualization for the correlation is shown below:

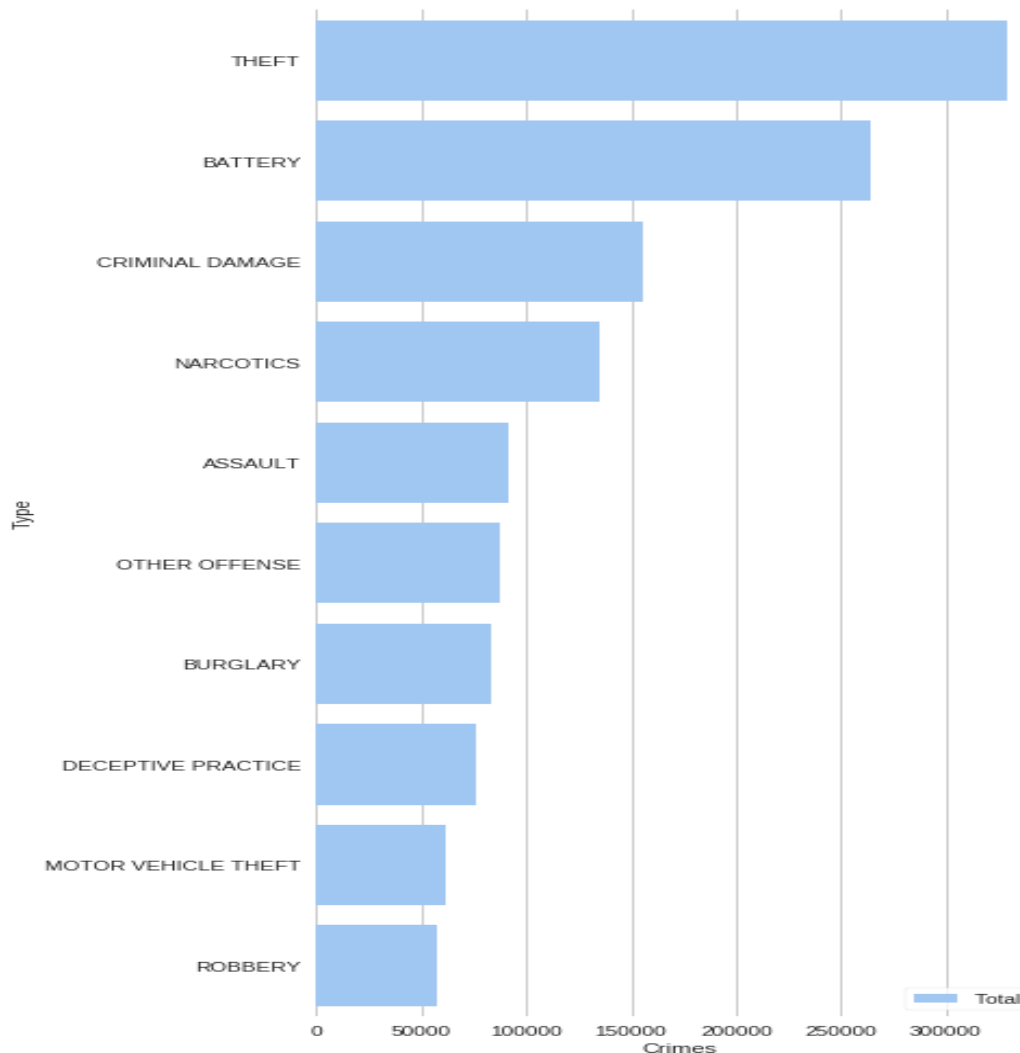


**Fig-10: Correlation of Crime Data**

So, for the analysis, the said variables are chosen and will be analysed in the next sections.

## 4.6 Analysis of Data

As the variables are already mentioned for the analysis, in this section the in-depth analysis and the discussion will be shown by drawing the analytical insight. The year wise analysis will be shown to check the relationship of progress of the crime of different types. The all over crime numbers can be visualized as follows:

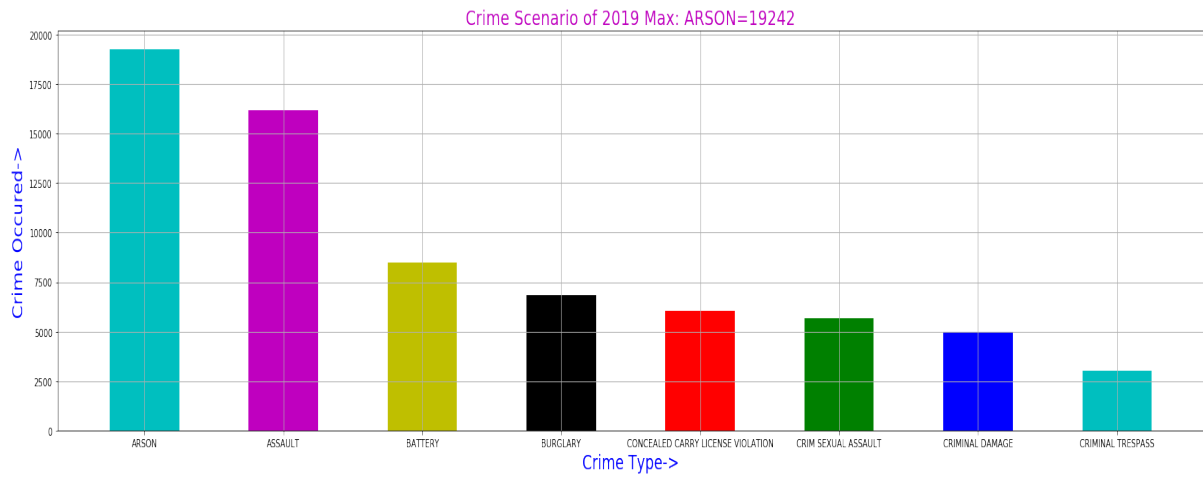


**Fig-11: Overall crime**

From Fig-1, it can be seen that the overall theft which includes the Arson crime and thus in the highest position of occurrence. The low value of the crime can be signified by Robbery. The year wise analysis is done. While analysing for the year 2019, the top eight crimes are found as:

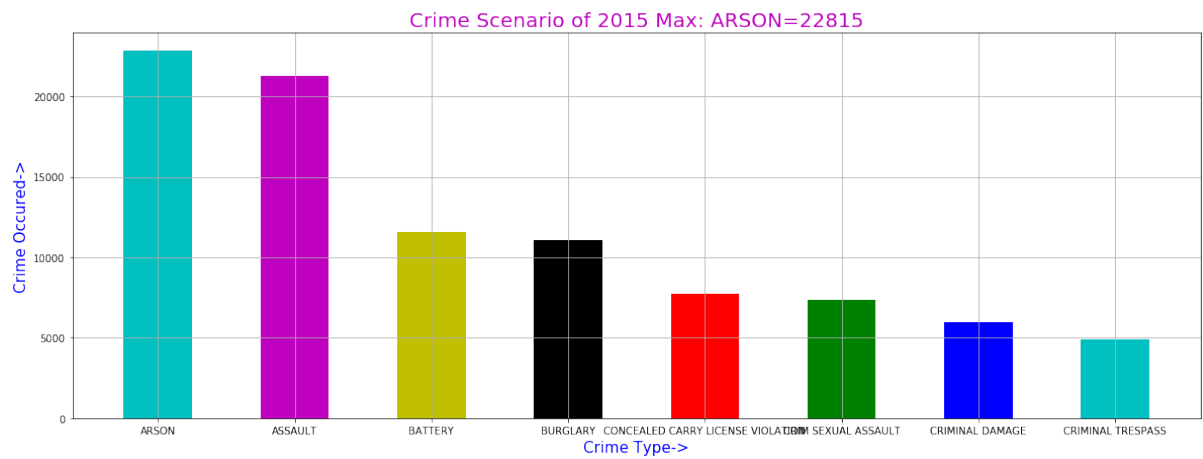
```
['ARSON' 'ASSAULT' 'BATTERY' 'BURGLARY'  
'CONCEALED CARRY LICENSE VIOLATION' 'CRIM SEXUAL ASSAULT'  
'CRIMINAL DAMAGE' 'CRIMINAL TRESPASS']  
[19242 16158 8467 6819 6017 5651 4974 3034]
```

The corresponding visualization is:

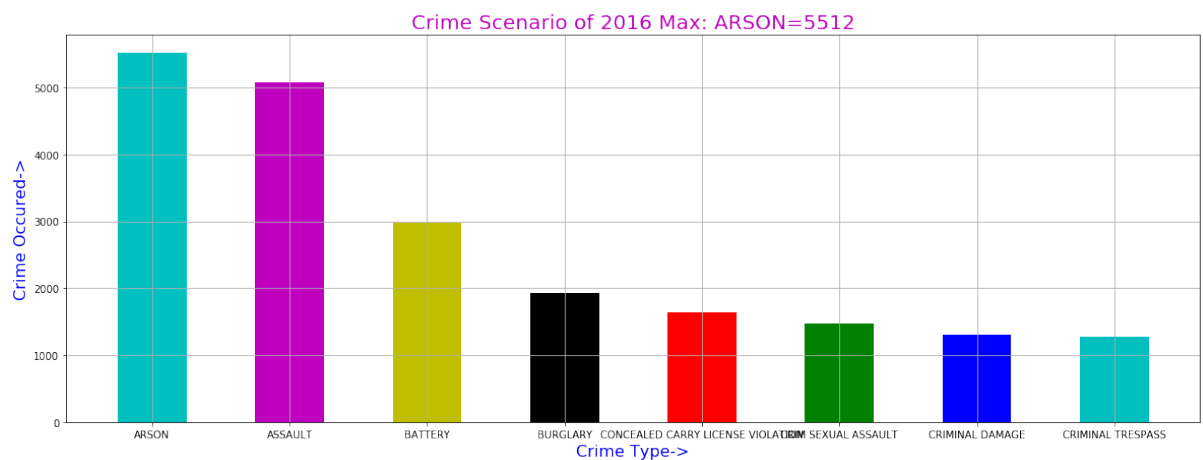


**Fig-12: Crime Analysis for Recent Year(2019)**

So, after the analysis, it is being visualized that the rate of crime Arson is higher in 2019. To clarify the fact, the same analysis will be done for 2018, 2017, 2016 & 2015 and the outcomes are shown below:

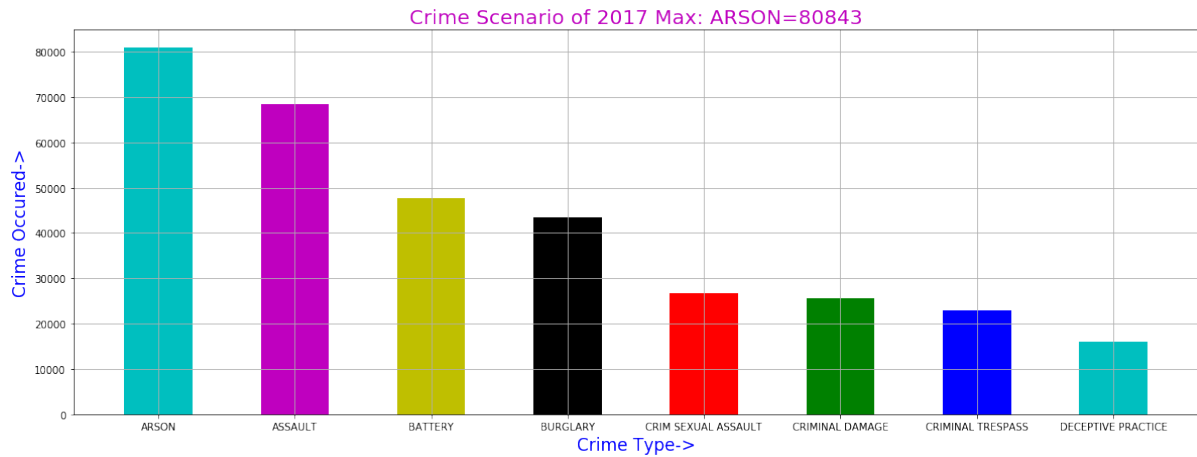


**Fig-13: Crime Analysis for 2015**

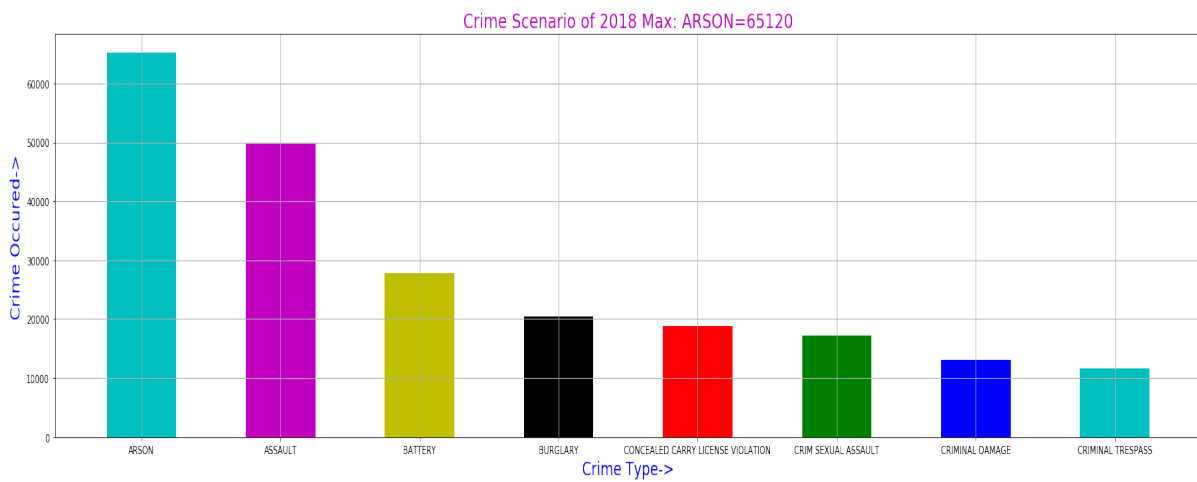


**Fig-14: Crime Analysis for 2016**





**Fig-15: Crime Analysis for 2017**



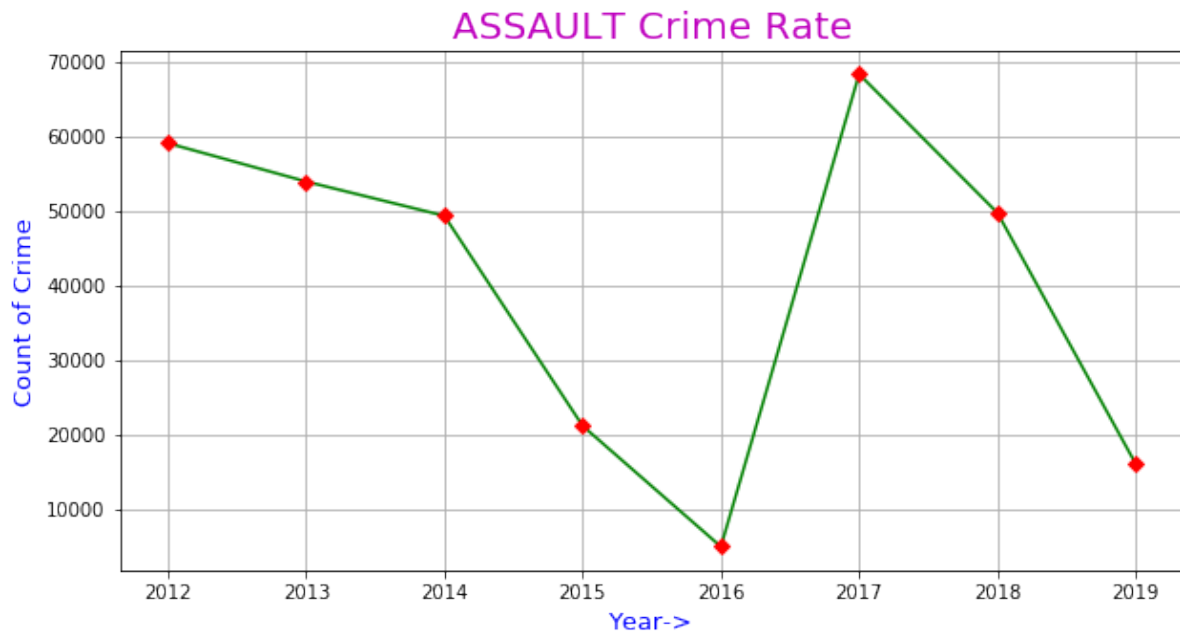
**Fig-16: Crime Analysis for 2018**

For the analysis from 2015 to 2019, it can be seen that, the Arson takes the highest place for every year and this the analysis for the Arson for past few years are required and hence done. The analysis for Arson is shown below:



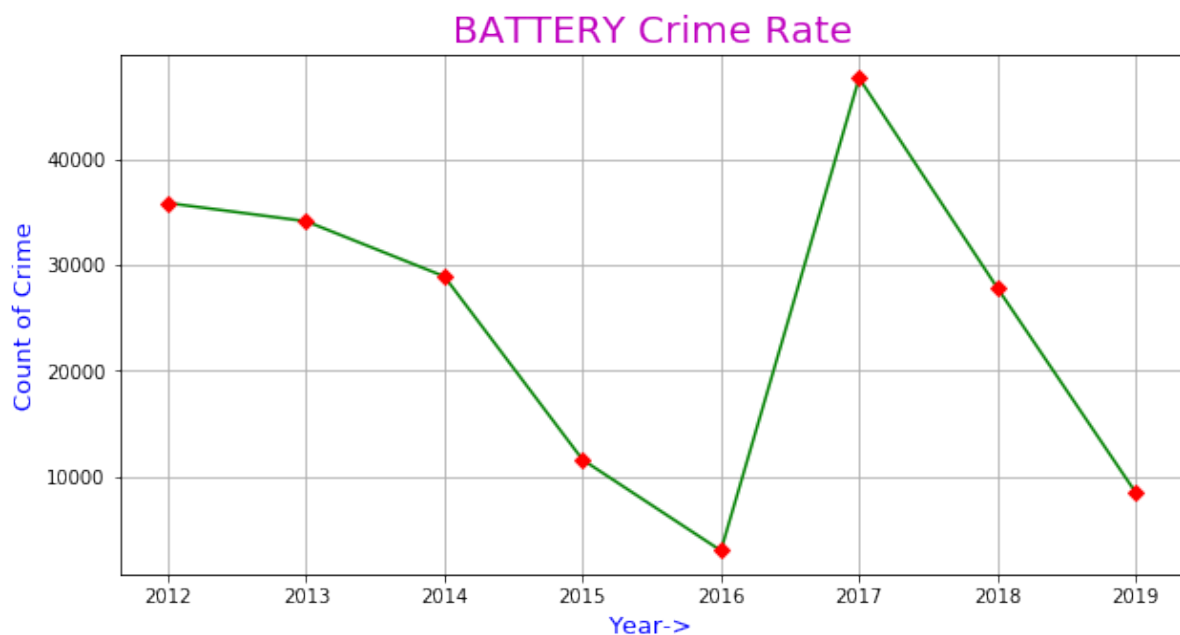
**Fig-17: Crime Analysis for Arson**

Though the Arson takes the highest place in every year analysis, that's why it is considered to be the most sensitive crime among all. The Chicago police, for this reason, have been taken the necessary steps and thus the rate of crime is in the downward while the years approaches. On the other hand, if we will look into the Assault crime, it also shows the same type of result through which it can be said that, Chicago police also active for this as well.



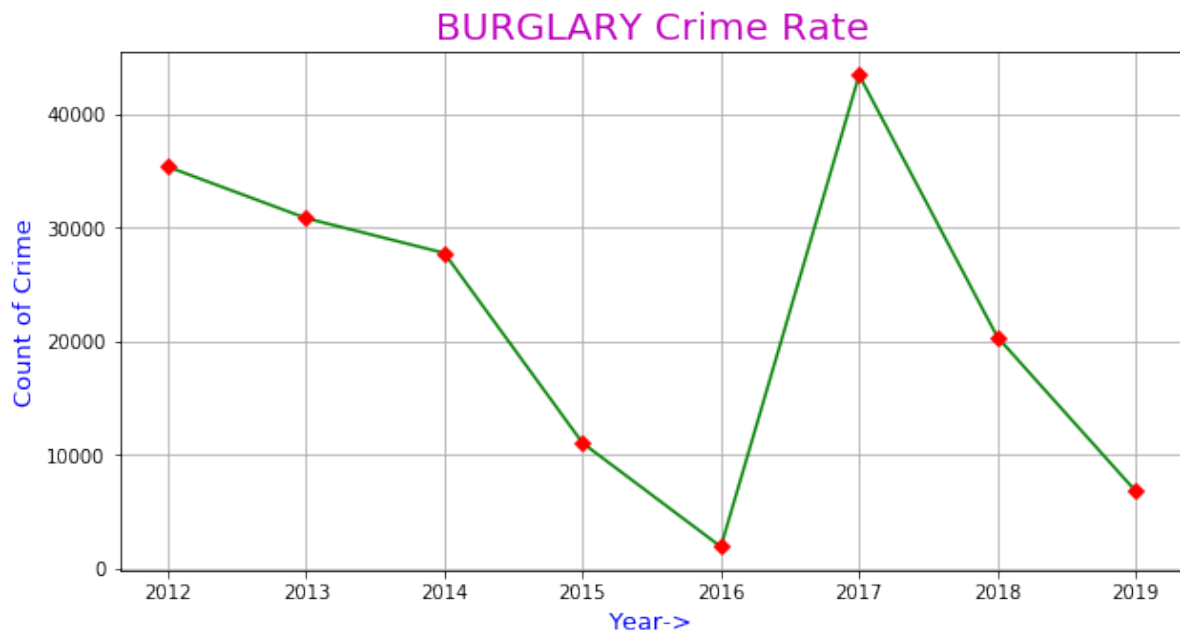
**Fig-18: Crime Analysis for Assault**

While doing the analysis, the crime like Battery is also well known in Chicago and it is also been so sensitive so that the police have taken steps as per required.

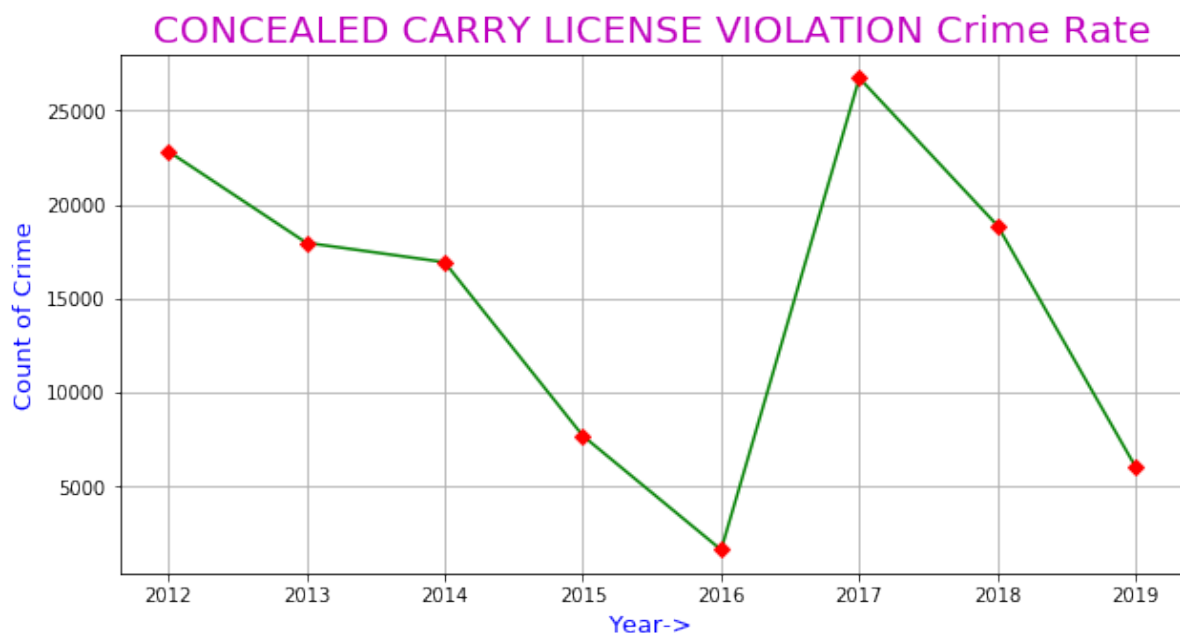


**Fig-19: Crime Analysis for Battery**

On the top 5 crime in Chicago, the other two are the Burglary and violence due to the local imbalance of law and thus the criminals are finding their way to do so. But, the analysis is done, it can be found the volume of such crime are still in decreasing mode and so the police have worked efficiently in view of the minimization of the crime.

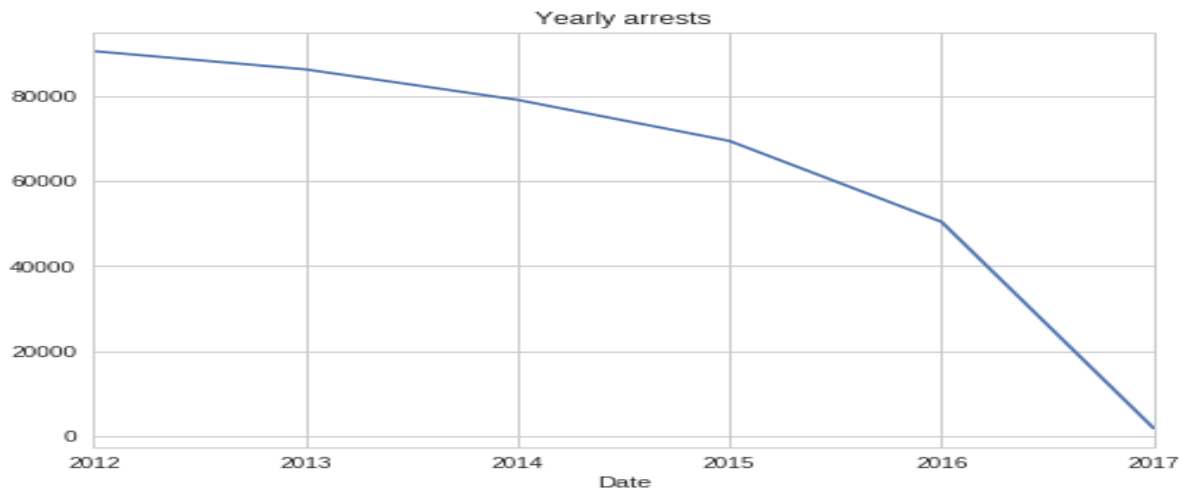


**Fig-20: Crime Analysis for Burglary**



**Fig-21: Crime Analysis for Violence**

So, in this analysis, the partial crime rate can be found. For more compact result, the combined analysis is made to visualize the top five crimes and their yearly and monthly progress of happening. The yearly count of the top five crime is shown below:



**Fig-22: Yearly top five crime**

If the analysis can be observed, it can be found that the yearly arrest is decreasing while increasing in crime. From, the Fig-17 to Fig-21, the crime rate have reach to its extent in the year of 2017 whereas, the arrest rate is lowest in the same year. As the arrest is related po all the types of crimes that have been taken place in Chicago still now, the rate of protection is not good and so the criminals have chosen their way to execute crimes openly.



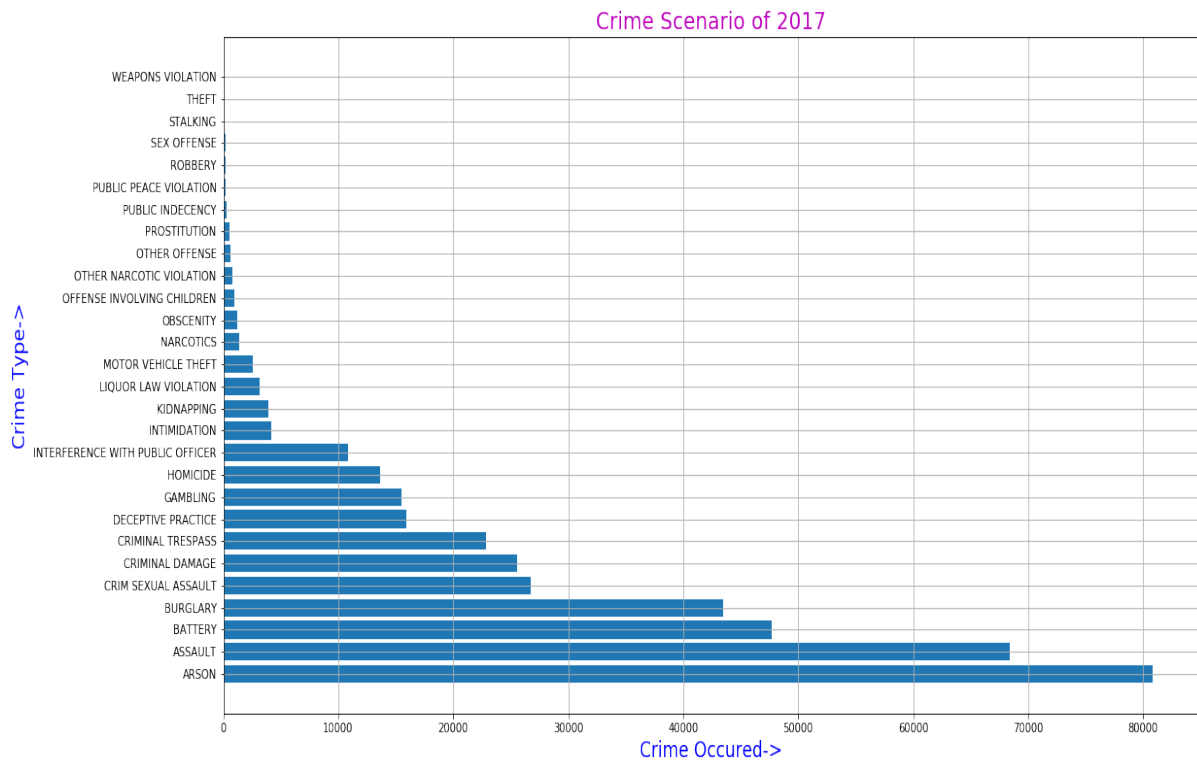
**Fig-23: Monthly top five crime**

The same picture can be observer in the monthly view. So, it is required to view the crime scenario for the year 2017. The crime counts are shown below:

ARSON	615
ASSAULT	22834
BATTERY	68369
BURGLARY	26720
CRIM SEXUAL ASSAULT	1381
CRIMINAL DAMAGE	47663
CRIMINAL TRESPASS	10834
DECEPTIVE PRACTICE	13648
GAMBLING	990
HOMICIDE	231

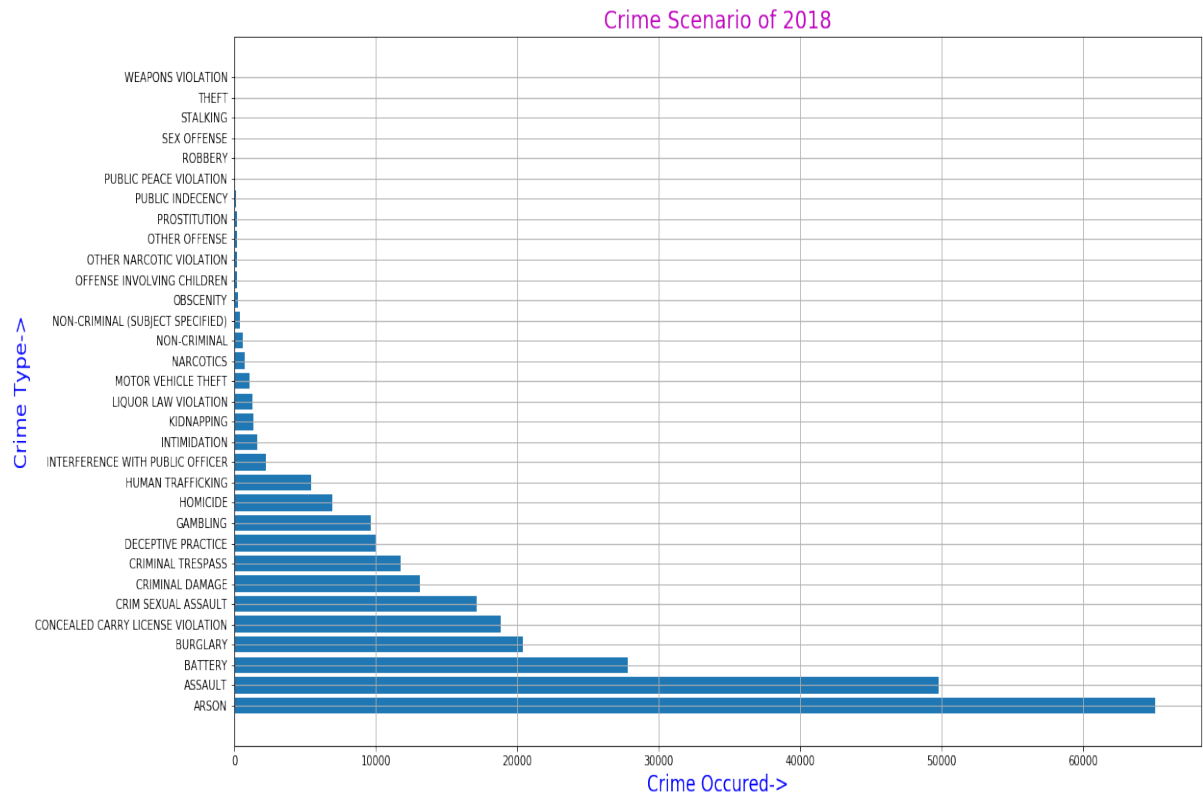
INTERFERENCE WITH PUBLIC OFFICER	572
INTIMIDATION	231
KIDNAPPING	293
LIQUOR LAW VIOLATION	745
MOTOR VEHICLE THEFT	15460
NARCOTICS	43452
OBSCENITY	21
OFFENSE INVOLVING CHILDREN	2523
OTHER NARCOTIC VIOLATION	7
OTHER OFFENSE	25569
PROSTITUTION	3939
PUBLIC INDECENCY	10
PUBLIC PEACE VIOLATION	3138
ROBBERY	15952
SEX OFFENSE	1233
STALKING	167
THEFT	80843
WEAPONS VIOLATION	4157

The overall crime scenario for the year 2017 is as follows:



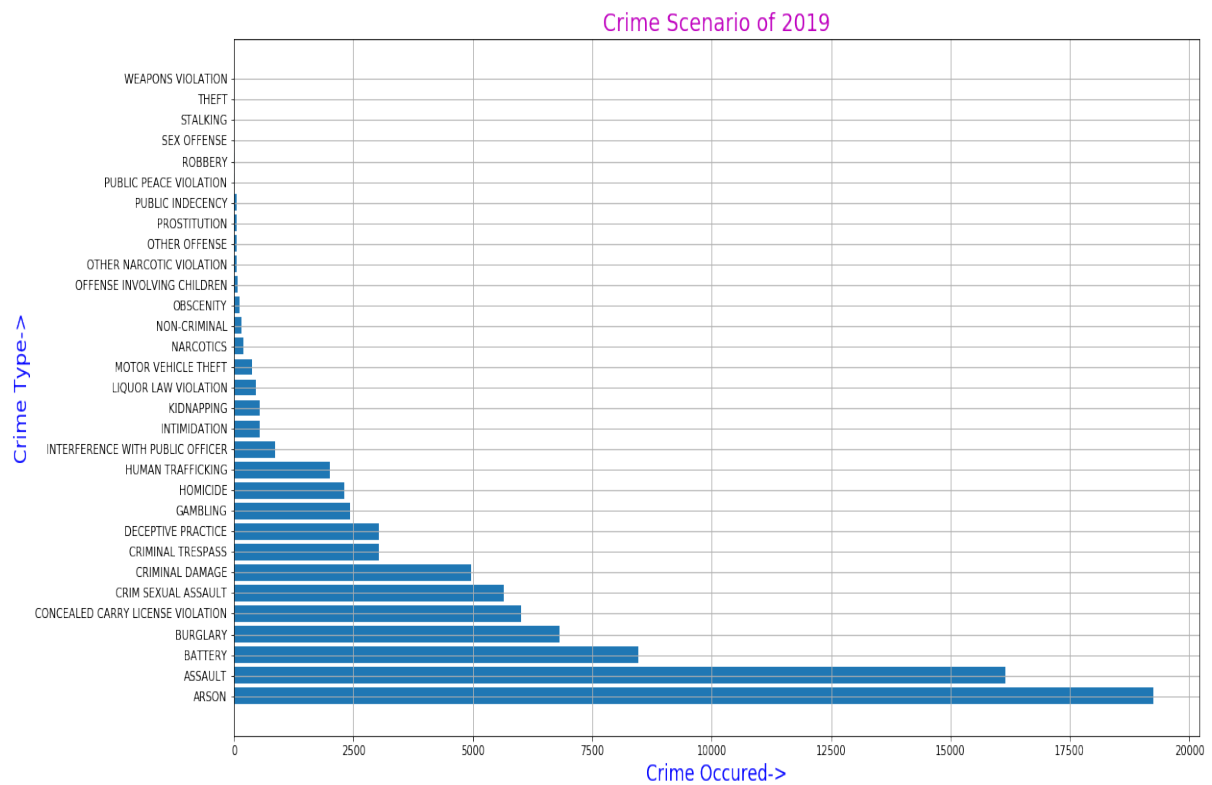
**Fig-24: Crime Scenario of the year 2017**

On the other hand, the crime scenario for the year 2018 is as follows:



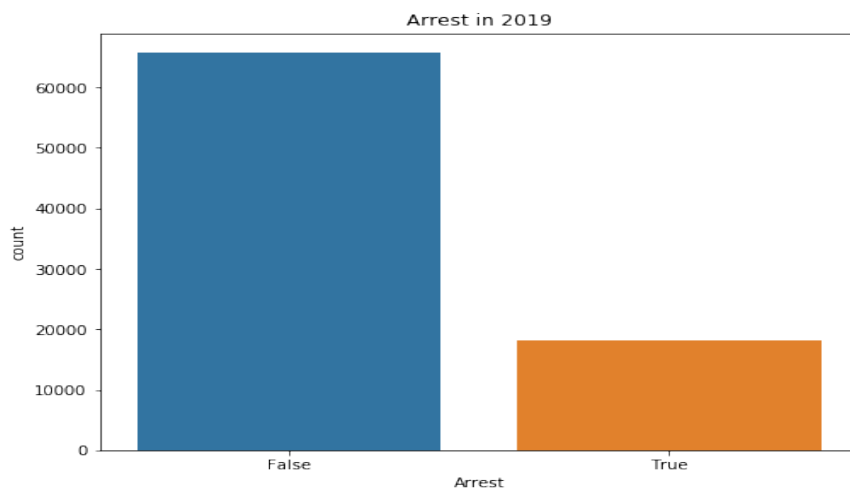
**Fig-25: Crime Scenario of the year 2018**

The analysis result for the year 2019 is shown below:

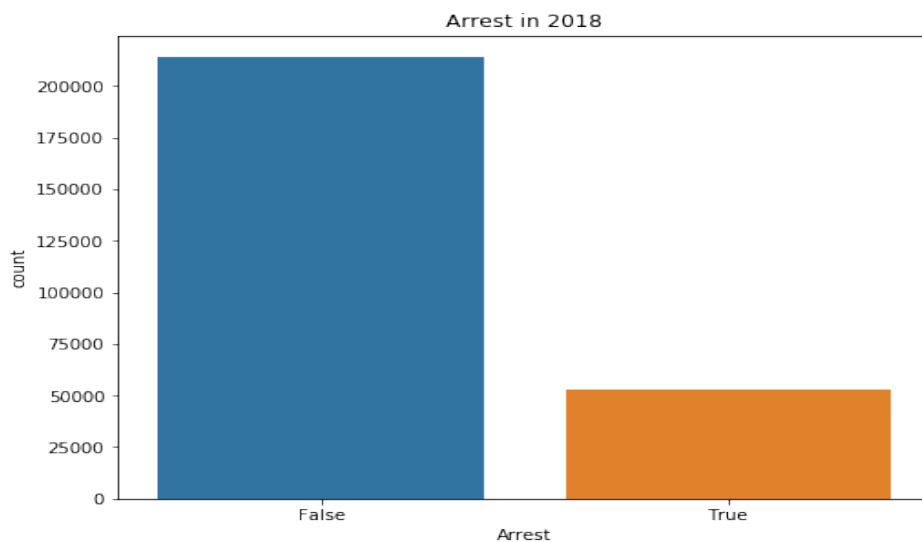


**Fig-26: Crime Scenario of the year 2019**

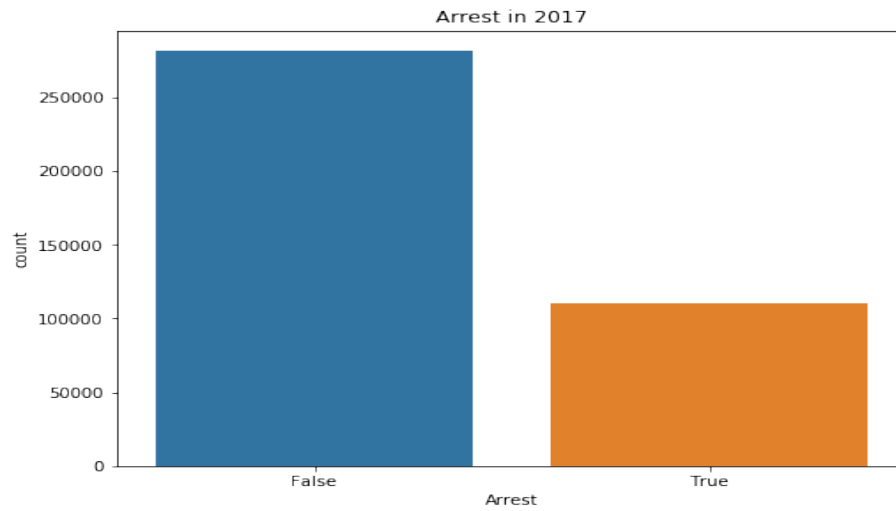
After analysing the data related to the number of crimes happened and with the reference of year, the emphasis is given on the arrest rate of last five years. This is required analysis as because this will determine whether the crime will be increased in future or not. The arrest attribute is depicted as the True or False in the data which means it has two binary value, True stands for the criminal is arrested and False stands for the criminal is not arrested. So, if the analysis is down, it will be clear enough to understand that actions that have been taken by the police. The Arrest rate for last five years are shown below:



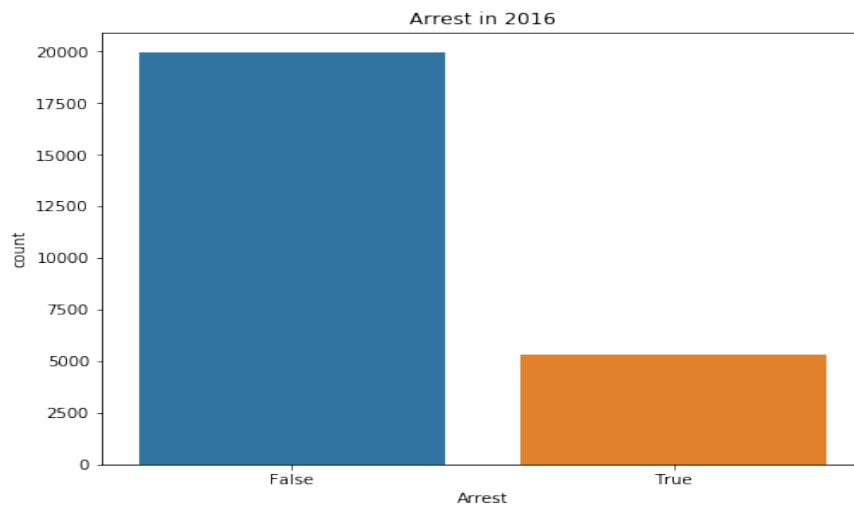
**Fig-27: Arrest rate of the year 2019**



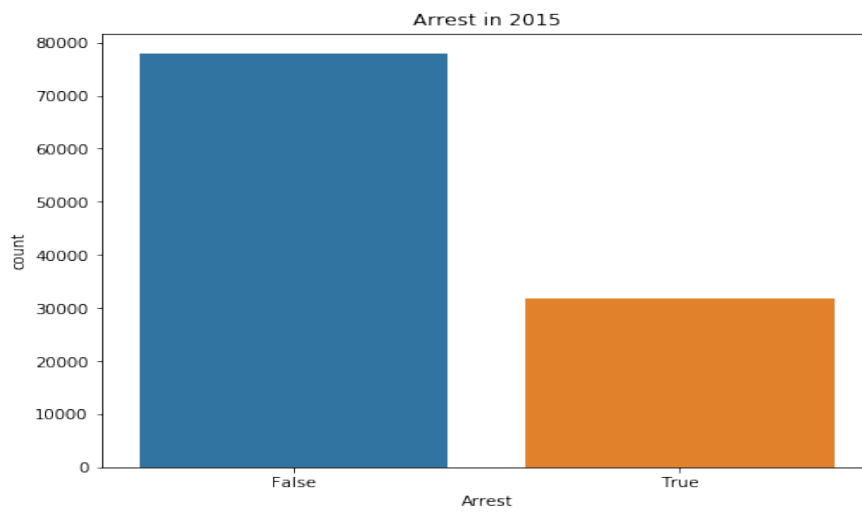
**Fig-28: Arrest rate of the year 2018**



**Fig-29: Arrest rate of the year 2017**



**Fig-30Arrest rate of the year 2016**



**Fig-31: Arrest rate of the year 2015**

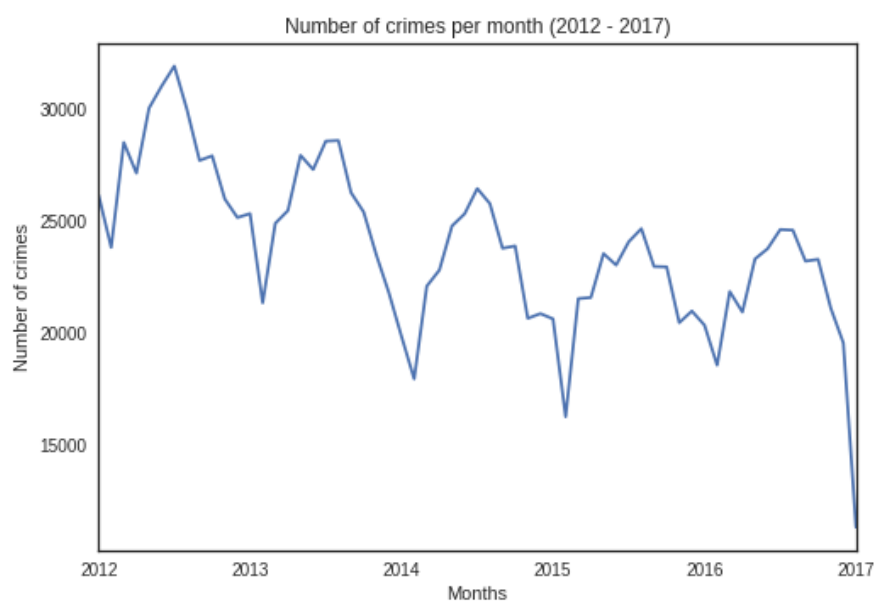


As seen from the earlier analysis, the results are matched for the arrest rate in Chicago. Though the crime rate is higher in Chicago, the arrest rate is still very poor. For the last five-year analysis, it is found that the case was launched, the IUCR and FBI code is given though the arrest was not made for most of the cases.

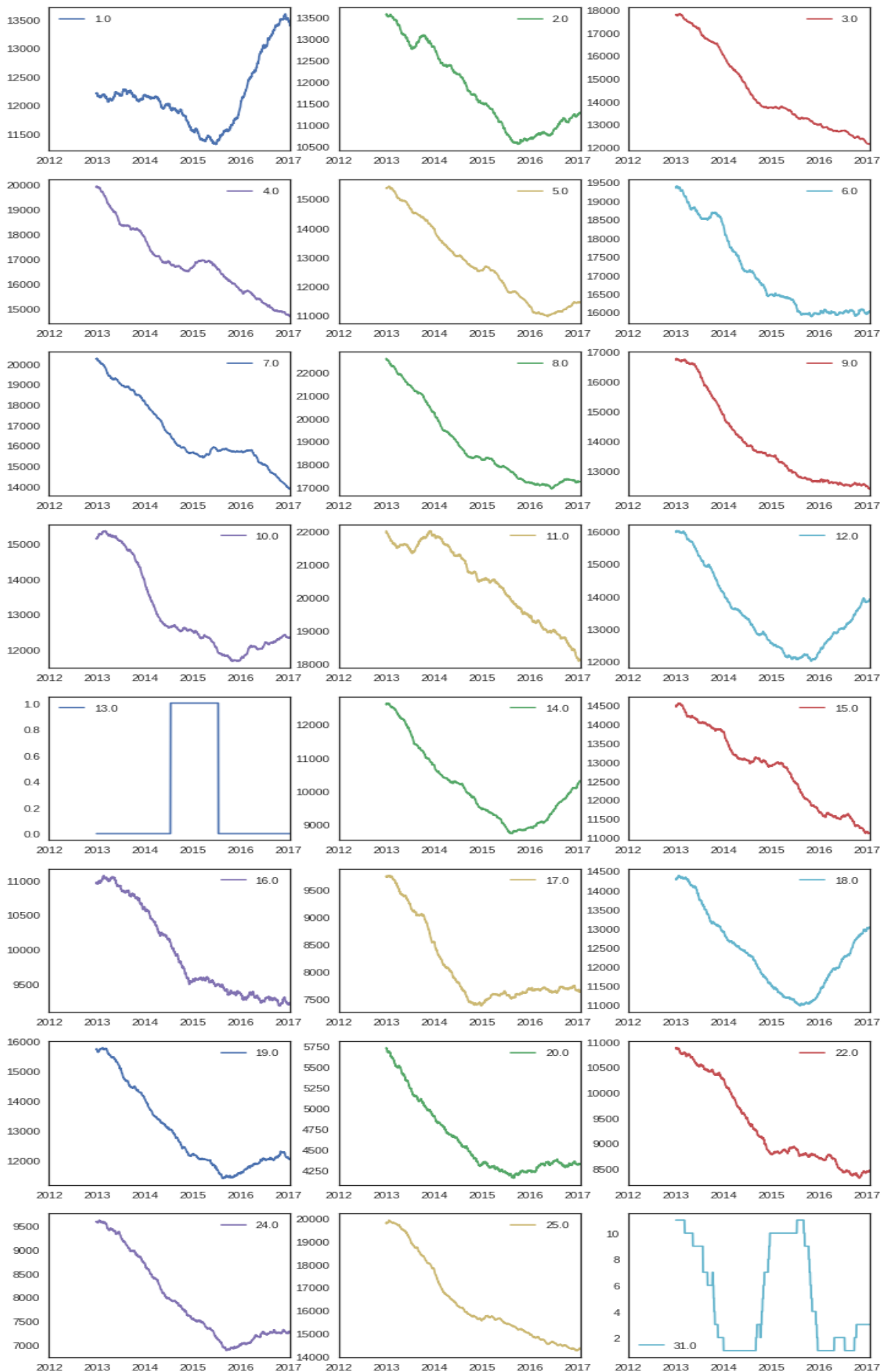
## 4.8 Evaluation of Crime using

### 4.8.1 K-Means Clustering

K-Means clustering is the unsupervised algorithm in Machine learning which is classify the data with respect to chosen parameters. In this case the crime scenario to be presented using the clustering method to get insight of the data which is selected here for the representation of crime. Hence, primarily the crime scenario to be presented in year wise and month wise as follows:

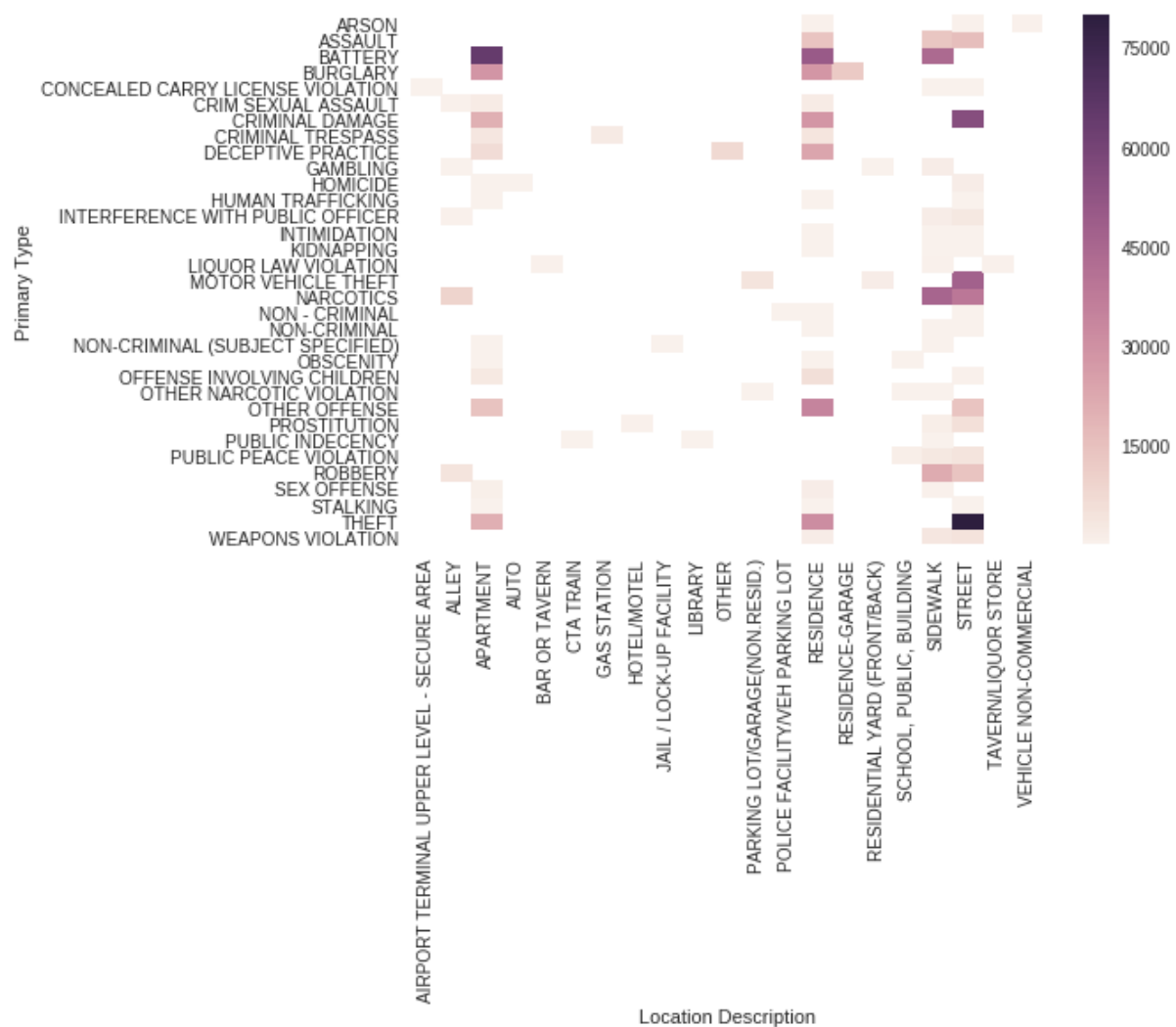


In this analysis, the crime occurrence is shown for every month in a year and it can be seen that the occurrence of crime is being reduced with respect to year. The district wise crime is shown as follows which depicts the crime occurrence in each district.



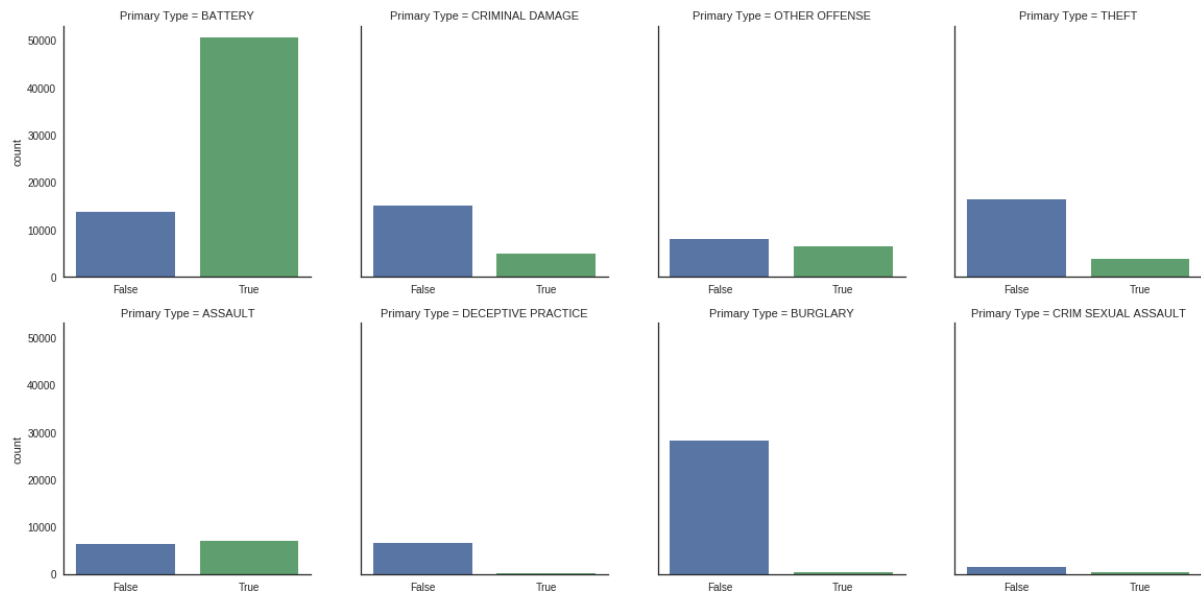
From the district wise analysis of the crime report, it can be seen that, the crime in the districts is being reduced except the district code 1, 12, 14 & 18. So, this indicate the good sign of the taking care of the crime scenario against crime happenings.

Next if we look into the crime scenario on the basis of the Primary Type and the location using the heatmap, the following result is obtained which depicts the fact that the happening of crime in a particular location.

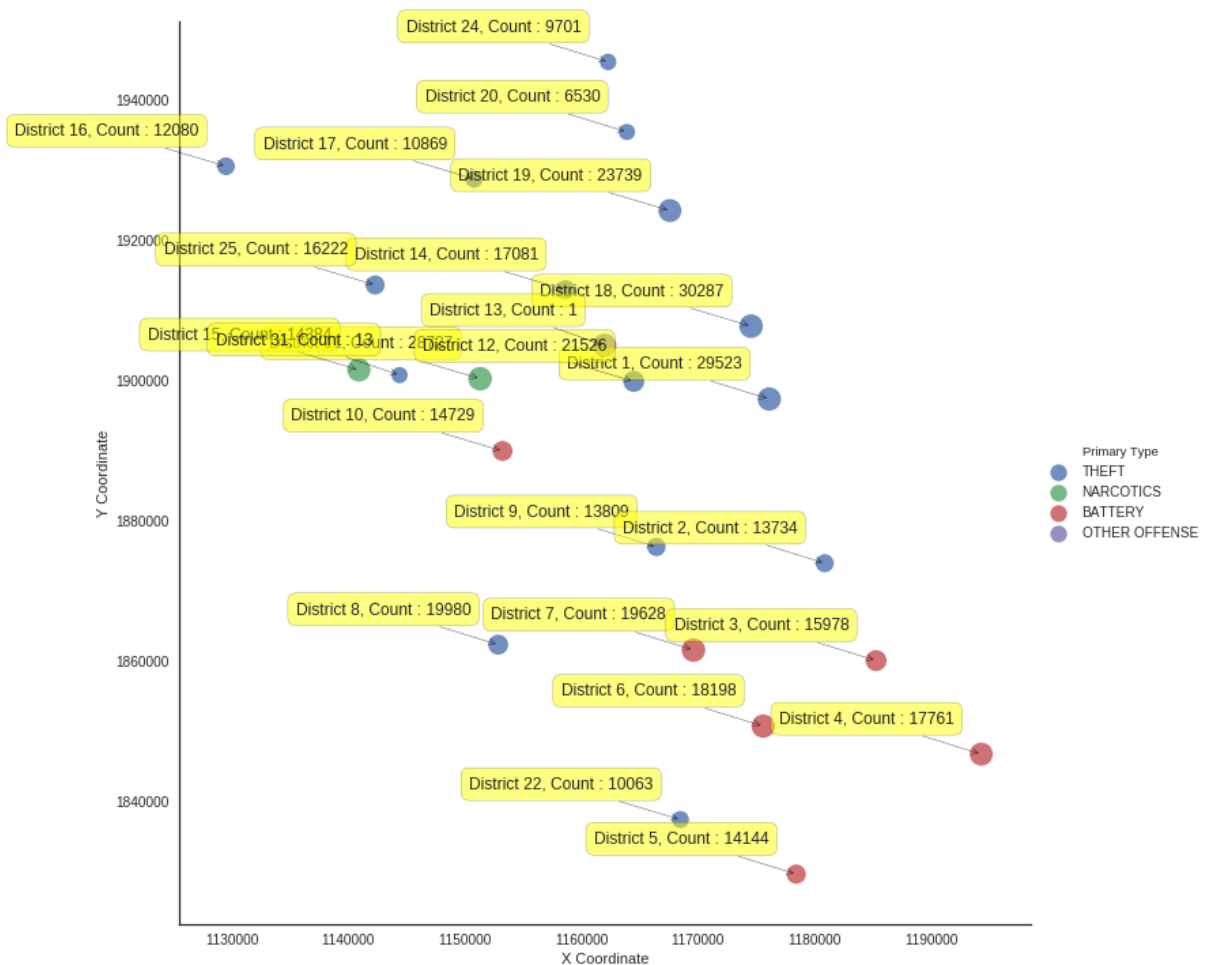


It can be seen that, for an instance, the crime type BATTERY is mostly happened at the locations of apartment. The dense color shows the high occurrence of crime in that particular location and hence it is the type of unsupervised classification and this will be more clarified in the next simulation on behalf of the crime detection.

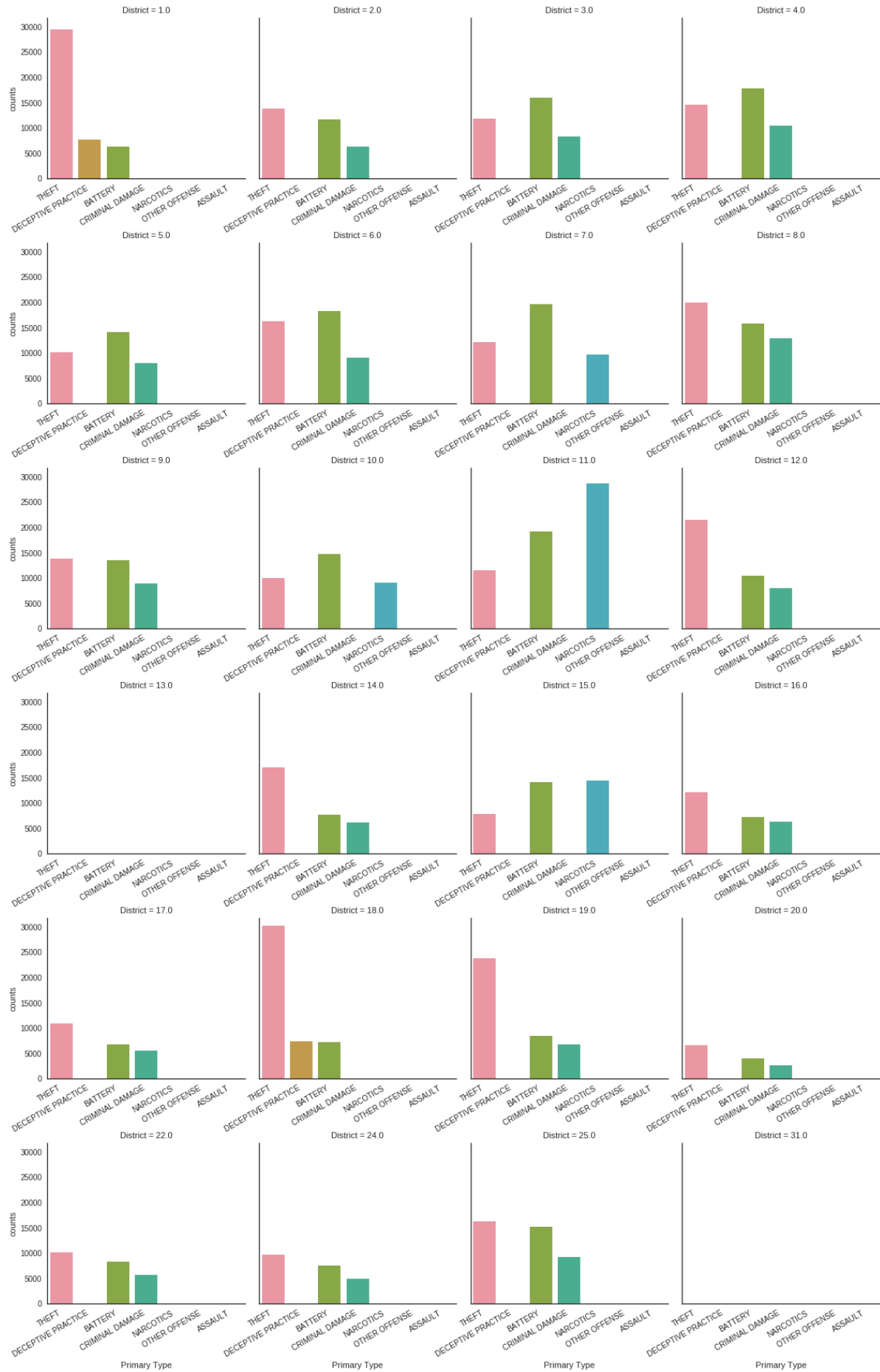
When the number of crimes is analysed whether the crime is happened for how many times, this will be signified here by using the bar chart for first and top 8 crime type. The analysis result is shown below:



It can be seen that BATTERY was happened for highest time and hence is indicated by True. Now, the analysis will be done for the crime occurrence with respect to the location by using the geographical latitude and longitude and the result is shown below:

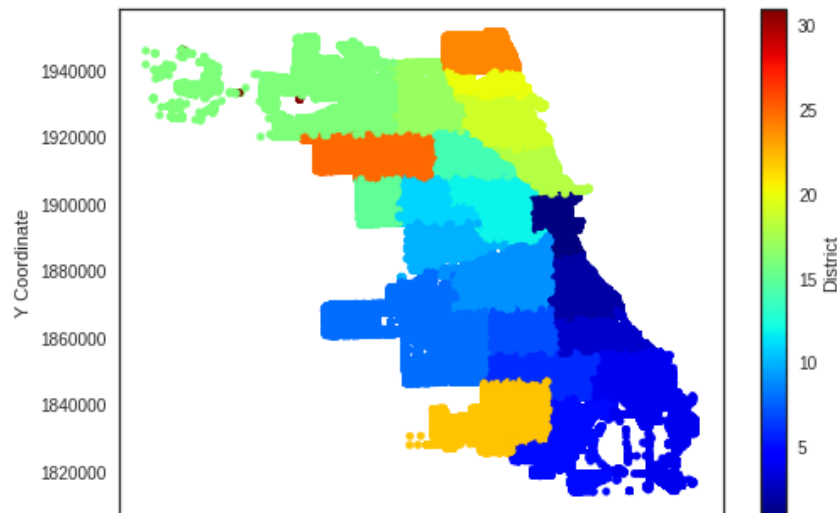


Next the clustering is applied for getting the outcome of the type of crime happened for each district with their respective values and hence is shown below:



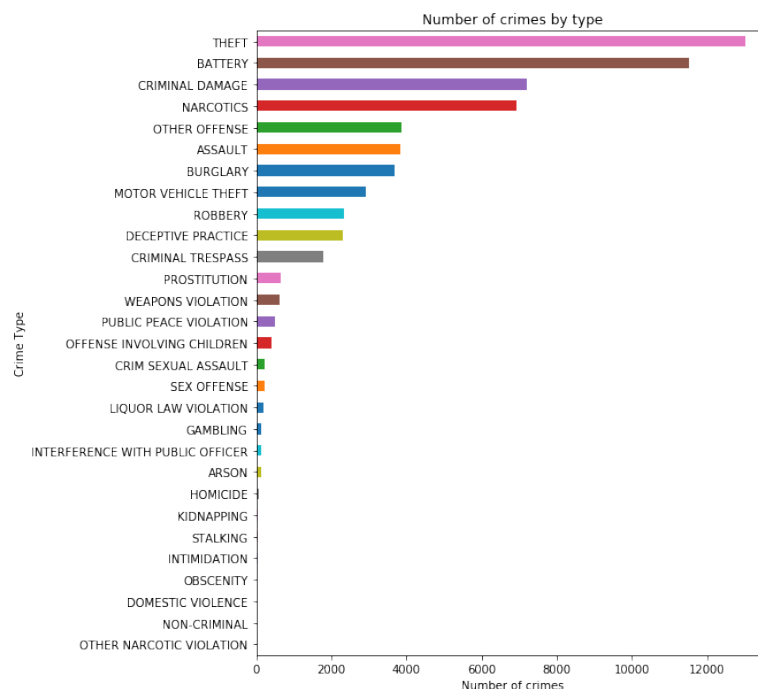
#### 4.8.1 Analysis using PCA and Elbow

In this section, the PCA with the implication of the Elbow method is used for the identification of the crime in different angles and to determine the number of clusters that can be drawn from the data of the crime in Chicago.



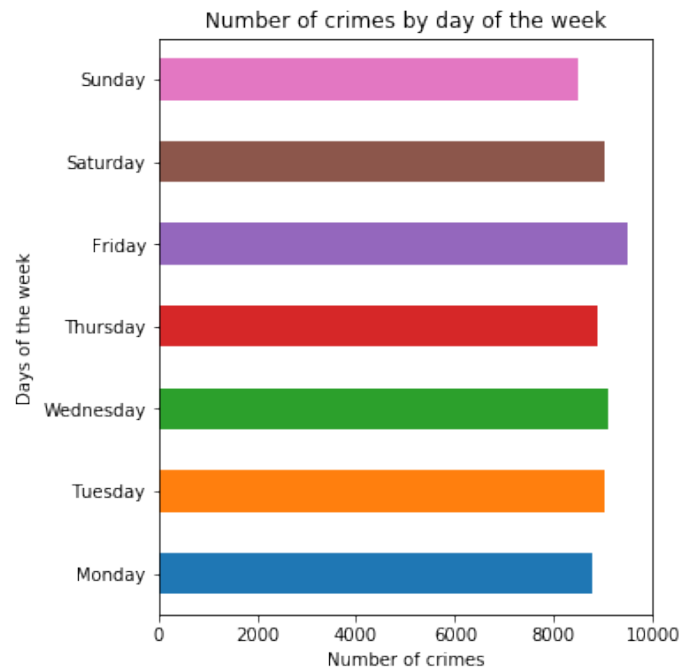
Primarily the crimes are clustered with the chosen dimension of the location and based on the location, the heatmap is applied to signify the sensitivity of the locations with respect to crime. The red is signified for high sensitive location for the crime and the deep blue is signified for the least sensitive location for the crime.

After that, using the unsupervised method, the number of happening of crimes is determined with respect to the crime type and the analysis result is shown below:

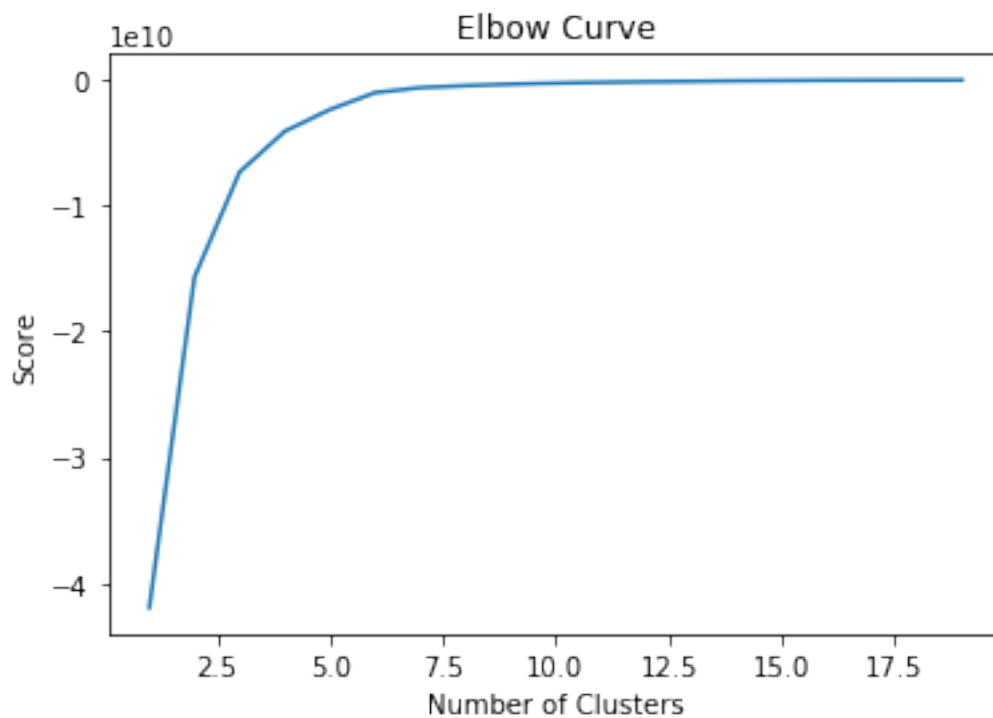


So, the highest occurrence is found for the crime type Theft in all perspectives.

After that the same procedure is applied to check out the crime happened with respect to the day of a week.

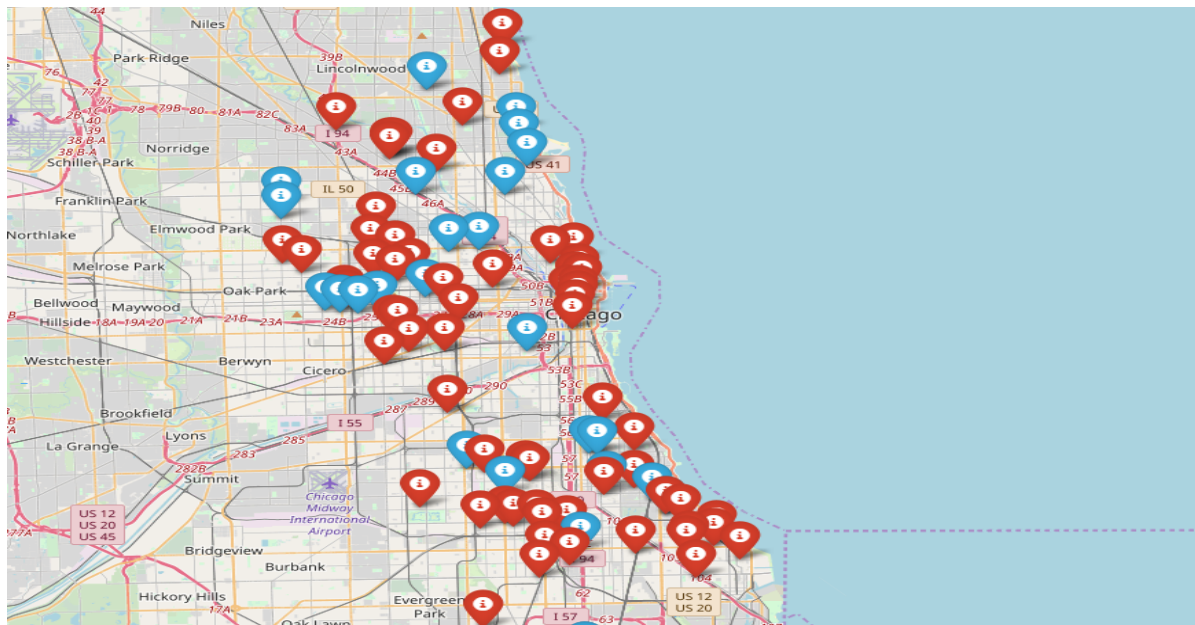


With that execution, now the elbow method is applied to find the number of clusters. The number of clusters can be found out using the k-means clustering also, but if the method or algorithm is applied, the accuracy will be increased and thus the number of cluster will be more perfect.



## 4.8 Result

The analysis made for the crime data for the attributes like Crime type or Primary type, Arrest conditions, rate of crime etc and their respective visualization are done. As basically, the crime data analysis is one kind of population analysis, which here in form of crime and the arrest and case launch rate, the bar charts are show so clarify the result as the magnitude of bar show more clear result of crime that other types of chart. S was found from the analysis; the crime rate is highest in the year 2017 in Chicago whereas the arrest rate is lower which means the Chicago police didn't solve the cases efficiently and thus the crime rate is still growing on. The location for the crime places is also analysed to show the crime density and the crime prone zone in Chicago. The Crime map is shown below:



**Fig-31: Crime map for 2015-2019**

In this output, the red location shows the fact that, the case was launched and the criminal was arrested but the blue location show that the case was launched but no action was taken means, the arrest was not made by the police.



## Chapter-5: Conclusion

In this paper, the data mining for the crime data is done properly so that the data can be imported and made ready for the analysis. After importing the crime data, the data was checked for inconsistency that is if the data contains any NULL value. So, after that checking is done, the NULL values are dropped so that it will not affect the entire analysis and observation

data. So, after the data pre-processing is done, the data was then ready for analysis. Then the required analysis was done over the data to classify the crimes and for the decision making whether the crime rate will increase in future or not. As the research question given in the Introductory section, the answer to the first question was found for the analysis that is Chicago is mostly affected by Arson and the analysis for the same was shown in the analysis section year wise and for every analysis, the result contains the high value for Arson. The remedy to get rid of such nuisances is that the Chicago police should increase the criminal arrest rate which is found very poor and for most of the case launched, all of those are not taken care of and thus no actions were taken for such cases. If such will go on, definitely, in the coming year, the crime rate will be increased and the arrest rate will be fallen down and so the Chicago will be mostly effected by different crimes which will hamper the general and healthy life in Chicago.

## References

- ChicagoPolice. (2019, July). *chicagopolice*. Retrieved from <https://home.chicagopolice.org/inside-the-cpd/statistical-reports/>.
- Guangxing Zhu, A. K. (2019). Early sexual initiation in Europe and its relationship with legislative change: A systematic review. *International Journal of Law, Crime and Justice*, 70-82.
- H. Chen, W. C. (2004). Crime data mining: a general framework and some examples. *Computer*.
- Håkansson, F. S. (2011). Associations between polysubstance use and psychiatric problems in a criminal justice population in Sweden. *Drug and Alcohol Dependence*, 5-11.
- Hofer, P. J. (2019). No AccessFederal Sentencing after Booker. *Crime and Justice*.
- Inqilab Shahbazov, E. M. (2019). Excessive use of pre-trial detention in Azerbaijan: Examination of the causes. *International Journal of Law, Crime and Justice*.
- J. Fagan, D. L. (2007). Social contagion of violence. *The Cambridge Handbook of Violent Behavior and Aggression*.
- John J. DiIulio, J. (1996). *My Black Crime Problem, and Ours*. City Journal.
- Klaus Bachmann, A. F. (2019). Accepting the political face of international criminal justice. *International Journal of Law, Crime and Justice*, 26-35.
- Ludwig, J. R. (2007). Is Crime Contagious? *The Journal of Law and Economics*.
- MadhuSudana Rao Nalluri, K. K. (2019). Hybrid Disease Diagnosis Using Multiobjective Optimization with Evolutionary Parameter Optimization. *Machine Learning Theory and Applications*.
- Manski, C. F. (2000). Economic analysis of social interactions,. *Journal of Economic Perspectives*, 115–136.
- Marlene Matos, H. G. (2019). Stalking victimization in Portugal: Prevalence, characteristics, and impact. *International Journal of Law, Crime and Justice*.
- Mathew Zaia, B. U. (2019). Investigating accountability of public police in the private employment realm. *International Journal of Law, Crime and Justice*, 36-46.
- Mayer, C. J. (1990). The Social Consequences of Growing Up. *National Academy of Sciences*.
- Peixin Zhao, M. D.-Q. (2014). Analyses of Crime Patterns in NIBRS Data Based on a Novel Graph Theory Clustering Method: Virginia as a Case Study. *The Scientific World Journal*.

- Qinglan Zhao, S. Y. (2019). A QR Code Secret Hiding Scheme against Contrast Analysis Attack for the Internet of Things. *Privacy and Security of Information Processing in Industrial Big Data and Internet of Things*.
- R. J. Sampson, S. W. (1997). Neighborhoods and violent crime: a multilevel study of collective efficacy. *Science*, 918-924.
- S. F. Greenfield, S. E. (2010). Substance Abuse in Women. *Psychiatric Clinics of North America*, 339-355.
- S. F. Greenfield, S. E. (2010). Substance Abuse in Women,. *Psychiatric Clinics of North America*, 339-355.
- S. Fazel, P. B. (2007). Substance abuse and dependence in prisoners: A systematic review,. *Addiction*, 181-191.
- Saruar Alam, G.-R. K.-I.-S. (2019). Twin SVM-Based Classification of Alzheimer’s Disease Using Complex Dual-Tree Wavelet Principal Coefficients and LDA. *Machine Learning Theory and Applications* .
- Schröder, T. (2019). Corporate crime, the lawmaker's options for corporate criminal laws and Luhmann's concept of “useful illegality”. *Go to journal home page - International Journal of Law, Crime and Justice*, 13-25.
- Silvia Martorano Raimundo, H. M. (2018). Contagious Criminal Career Models Showing Backward Bifurcations: Implications for Crime Control Policies. *Journal of Applied Mathematics*.
- Tao Wang, B. Y. (2019). An Approach Enabling Various Queries on Encrypted Industrial Data Stream. *Privacy and Security of Information Processing in Industrial Big Data and Internet of Things*.
- Tonry, M. (2019). Fifty Years of American Sentencing Reform: Nine Lessons. *Crime and Justice*, 1-34.
- Vincenzo Ruggiero, w. t. (2019). Front-line practitioners versus received theories of crime and terrorism. *International Journal of Law, Crime and Justice*, 59-69.
- Wilson, W. J. (1997). The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy. *The University of Chicago Press*.

