1/8/2022

**Crime Data Analysis**

CSCI-322

Salma Khaled 18102479

Samar Ibrahim 18102357

Haidy Tarek 18102411

Ahmed Khaled 18101381

Saif Abdelaziz 18101353

Mohamed Wageeh 18102283

Table of Contents

[**Introduction** 2](#_Toc92735927)

[What we want to observe 2](#_Toc92735928)

[**Data Description** 3](#_Toc92735929)

[**Results** 4](#_Toc92735930)

[**Conclusion** 8](#_Toc92735931)

# **Introduction**

There is a rapid increase in crime in almost every country. There is a strong need to identify crime patterns and analyze different areas of crime. Security agencies all over countries are working hard to reduce these crimes, however, the size of crime information is increasing rapidly, and it becomes difficult to manage such a huge amount of data and to keep records of crimes that are geographically widespread and at different times period. Thus. it is very necessary to have a crime information system that can process a large amount of data in a short period. Data mining using clustering, classification, association mining is one of the most effective ways to explore, analyze, detect patterns and predict future crimes in a huge amount of data.

## What we want to observe

* Top crimes in Chicago in the year 2018?
* Most occurring top 5 crimes in Chicago in 2018
* Modify the date column to a pythonic format
* Convert the date-time column to a known format
* Create a new column month
* The probability of arrest of Chicago
* Distribution of arrests across the months
* Crime vs time
* What are the most unsafe hours?
* Is your house safe from burglary during the day?
* Visualize a crime pattern for24 hours?
* Crime vs location
* Most common Occurrences per district
* Which region is given crime concentrated in?
* Visualizing Narcotics (to get an understanding of the most prevalent of drugs in the area
* Visualizing with heatmaps
* Using Supervised machine learning to predict crime hotspots

# **Data Description**

In this case study, we analyze the crime trend in the city of Chicago in the period 2001-2018 which was uploaded from Kaggle https://www.kaggle.com/chicago/chicago-crime. The dataset contains 22 columns and has 3374977 cases. Every record identifies criminal activity using various features. These attributes include:

* **ID**: Unique identifier for a record.
* **Case Number**: Chicago Police Division Record Number.
* **Date**: Date of the incident.
* **Block**: The abbreviated address for the criminal activity.
* **IUCR**: Internal Uniform Chicago Crime Reporting Code.
* **Primary Type**: Type of crime
* **Description**: Little more details about the criminal activity
* **Location Description**: Location where the crime occurred.
* **Arrest**: Boolean indicating whether the arrest was made.
* **Domestic**: Indicates whether the incident was domestic-related.
* **Beat**: Indicates the Beat where the incident occurred.
* **District**: Indicates Police District where the incident occurred.
* **Ward**: The City Council District where the incident occurred.
* **Community Area**: The Community where the incident occurred.
* **FBI Code**: Indicates the crime classification code.
* **X Coordinate**: The X coordinate of the location of the incident.
* **Y Coordinate**: The Y coordinate of the location of the incident
* **Year**: Year the incident occurred.
* **Updated On**: The date of the last update of the record.
* **Latitude**: Lattitude of the location of the incident
* **Longitude**: Longitude of the location of the incident.
* **Location**: Lattitude and longitude saved as a tuple.

# **Results**

Chart, bar chart

Description automatically generated

Dealing with missing values by dropping all rows that have at least one missing value.

Chart, bar chart

Description automatically generated

Here is the most occurring criminal in Chicago. The plot shows that theft is the most occurring value and a high count of crimes indicates the presence of a physically violent community.

Chart, line chart

Description automatically generated

The plot shows that theft is rising during summer. All types of crime were at their lowest in February. Assault and deceptive practice are smoothies that don’t have ups and downs.

Chart, bar chart

Description automatically generated

79.47 of the crimes saw no arrests of all reported crimes. The plot shows that the total number of crimes with no arrest is greater than crimes with arrest.

Chart, bar chart

Description automatically generatedChart, bar chart, histogram

Description automatically generated

The months of May, July, and August record the greatest number of crimes in the city. So, crime rises in summer.

Chart, bar chart, histogram

Description automatically generated

Noon is the peak hour of the rate of crimes in the city.

Chart

Description automatically generated

Visualization of NARCOTICS shows that POSS: HEROIN(WHITE) is the most prevalent type of drug in the city that counts 276.

A picture containing graphical user interface

Description automatically generated

Plotting for District vs Month in Gang crimes

# **Conclusion**

We were able to uncover answers to various queries we had about crimes in Chicago through the examination of crime data. We were able to attain our goals by following a systematic method. We discovered answers to queries such as which crimes were committed the most in 2018, the most dangerous hours, the most dangerous areas, the most common gang-related activities in Chicago, and so on. We also discovered the stunning reality that in 80% of the incidents, no arrests were made. However, because the information gathered was based on first-hand testimonies, we may be more positive and say that 20% of the crimes committed resulted in arrests within the first 24 hours of the incident being reported!