



# DATA MINING APPLICATIONS IN CRIME STATISTICS

Julia Chen

# DESCRIPTION

The dataset provides information about crimes that occurred in the City of Chicago from 2001 to September 2023. Features include the date of incident, primary type of crime and description, and location of the incident.

Questions to address:

- What types of crime are the most frequent in each of Chicago's wards?
- Are there any significant frequent patterns of crime for time of year, type of crime, and location of the event?
- Where is each type of crime (theft, assault, homicide, etc.) most likely to occur?

# PRIOR WORK

S. Yadav, M. Timbadia, A. Yadav, R. Vishwakarma and N. Yadav, "Crime pattern detection, analysis & prediction," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2017, pp. 225-230, doi: 10.1109/ICECA.2017.8203676.

H. Chen, W. Chung, J. J. Xu, G. Wang, Y. Qin and M. Chau, "Crime data mining: a general framework and some examples," in Computer, vol. 37, no. 4, pp. 50-56, April 2004, doi: 10.1109/MC.2004.1297301.

# DATASET

This project will use one dataset which includes approximately 7.91M rows and 22 features.

The dataset was downloaded from [Data.gov](https://data.gov) and was published by the City of Chicago.

The dataset has been downloaded and is stored on my personal computer.

# PROPOSED WORK

## Data Cleaning

- Change each feature to correct DataFrame type.
- For NaN feature values of interest, remove rows if possible. Most of the features of interest have few NaN values.

## Data Transformations

- For time of year question, add an additional feature which will be of one of four values (0,1,2,3) which signify winter, spring, summer, and fall.

# LIST OF TOOLS

## Python

- Pandas
- NumPy
- SciPy
- Sklearn
- Mlxtend

## Data Visualization

- Matplotlib

# EVALUATION

- Do our results have a strong association with a significant correlation between the itemsets?
- Are our results interesting in that they are easily understood, valid on new or test data, potentially useful, and give new insights into crime patterns in the city of Chicago?