



Analysis of Chicago Crime Trends



Amanda Killeen, Isaac Hammes, Graham Dominick,
Daniel Kingsley



Goals

We intend to explore Chicago crime data for trends and correlations in attributes including crime type, locations, and timeframes.

Possible Questions:

- Is there a correlation between type of crime and location description or district?
- Do arrests occur more often in different locations for the same crime?
- What are the most common crimes for different areas?
- Are there periods when there were waves of certain types of crimes?
- Trends in crime types?
- Can an accurate model be generated for “arrest made” based on other information about a crime?
(Logistic Regression or KNN classification)

Motivation:

This type of analysis could be useful in discovering crime waves, trends, clusters and other useful knowledge that could be used to plan or take action by law enforcement entities.

Prior Work

- Chicago Crime Data Analysis:
 - <https://www.kaggle.com/djonafegnem/chicago-crime-data-analysis>
- Medium Article:
 - <https://medium.com/@stafa002/my-notes-on-chicago-crime-data-analysis-ed66915dbb20>
- Number of reported crimes by each attribute visualizations:
 - <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present-Dashboard/5cd6-ry5g>
- Crime Type Prediction
 - <https://github.com/Tbhangale/Chicago-Crime-Analysis>
 - <https://pdfs.semanticscholar.org/a4ba/0cd981ac32fa7f6f8a489f525edf3dae99e2.pdf>

Dataset

Chicago crimes 2001-Present:

<https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present-Dashboard/5cd6-ry5g>

Contains around 7,000,000 entries of reported crimes.

Key attributes:

- Type/Description
- Area (Address/Neighborhood)
- Location (e.g. Apartment, Bar, etc.)
- Date
- Arrest

Proposed Work

- **Data cleaning:**
 - Remove unneeded columns
 - Clean string data (ensure uniformity between descriptions, locations, etc.)
- **Preprocessing:**
 - Bin by location
 - Filter by crime type / arrest
- **Analysis:**
 - Frequent Patterns
 - Classification (Arrest Made? Expected Crime type?)
 - Prediction (“How many robberies in Summer 2020?”)

Tools

- Python
- Jupyter Notebook
- Sklearn (KNN)
- Python statsmodels.api (Linear/Logistic Regression)
- Numpy
- Pandas
- Tableau

Evaluation

- Ultimately depends on questions we decide to ask
- Use methods that are familiar to us
- Accuracy, error rate, and confusion matrices for classification tasks
- Mean squared error for inference/prediction tasks
- Holdout method: train and test sections of data
- For models:
 - Generation time
 - Prediction time