

Analysis of Chicago Crime

Zahra Sinaei

May 2020

Goals

- ▶ We have collected the data from the data set [Chicago Crimes 2001-present](#).
- ▶ The data has 7.11M rows and 22 Columns. In the following we only analyzed 1M rows.

This project has two goals:

1. Study of so called [outdoor crimes](#) per community area and determining the [theft danger zones](#) with respect to each community area.
2. Study of the performance of police department in each district.

Project 1: Analysis of outdoor crime per community area

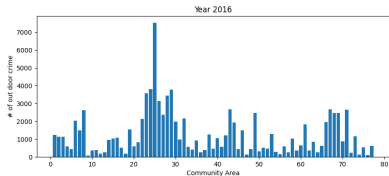
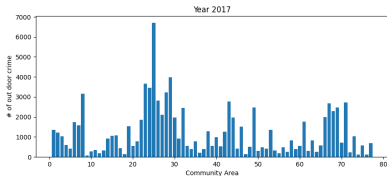
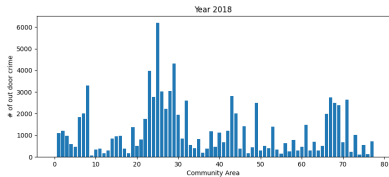
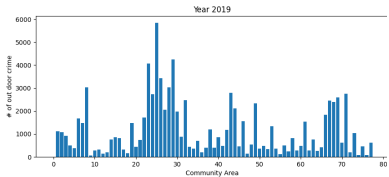
Definition

We call a crime an *Outdoor Crime* if its Location Description is one of the following:

[sidewalk, street, public, alley, CTA station, ATM, construction site, gas station, CTA platform, lakefront/waterfront/riverbank, CTA bus stop, CTA subway station]

Project 1: Analysis of outdoor crime per community area

- ▶ The city of Chicago has 77 community areas.
- ▶ The purpose of this project is to determine the community areas which are safer for walking and determining what are the theft safest zones in the outdoor community areas.
- ▶ Our preliminary study is for the number of outdoor crimes per community area in each of the years 2016–2019.
- ▶ Figure 1 shows the bar chart of outdoor crimes per community area in 2016–2019.
- ▶ Further study: we can also determine the number of outdoor crimes in each month, in weekdays, and weekends and analyze the corresponding data.



Project 1: Analysis of outdoor crime per community area

- ▶ We use the K-mean clustering algorithm to create clusters for location distribution of outdoor crimes with primary type theft in year 2019. We would like to find the optimal number of clusters in each community area.
- ▶ Figure 2 shows a scatter plot of theft type outdoor crime location distribution in community area 30.

Scatter plot for location distribution of community area 30 with 10 clusters

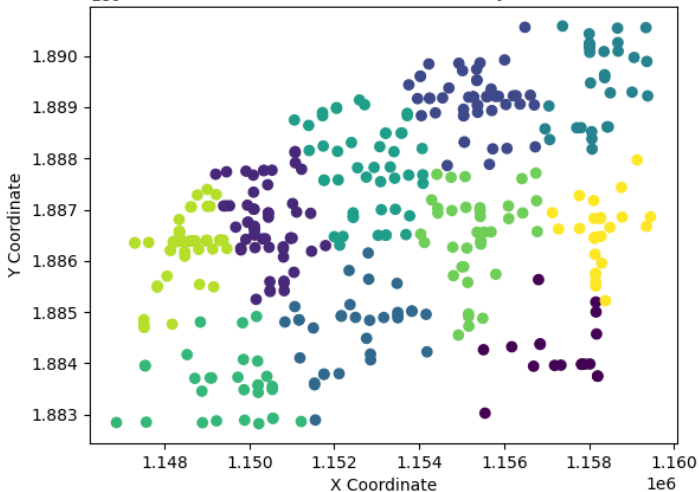


Figure: Figure 2

Project 2: Analysis of police performance

- ▶ City of Chicago has 25 police districts.
- ▶ First we determine the percentage of arrests per the total number of crimes in each district and determine the districts with highest such percentage.
- ▶ Figure 3 shows a plot of average arrests (that is, the number of arrests divided by the total number of crimes) per district in years 2016–2019.

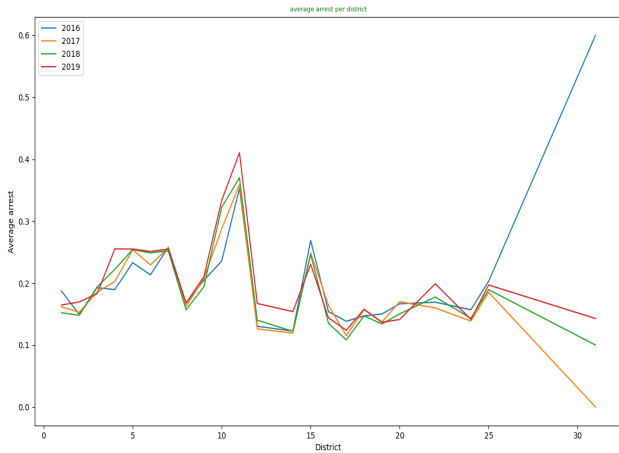


Figure: Figure 3

Project 2: Analysis of police performance

- ▶ Figure 4 shows the scatter plot of location distribution of crimes in year 2019.
- ▶ Our goal is to use the KNN algorithm with optimal number of neighborhoods to predict an arrest given the crime location.

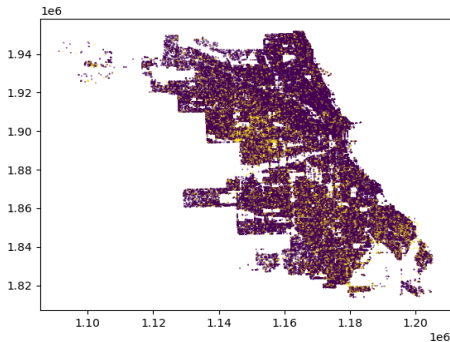


Figure: Figure 4