ANALYSIS OF BUFFALO CRIME DATASET

EAS 503

Submitted By : Group 22

December 2022

Jeevalkant Dandona (50485395)

Maitrai Sharma (50476049)

Mancy Saxena (50477980)

Nihal Mishra (50480980)

Varun Muppalla (50485121)

University at Buffalo
UB | Graduate School of Education

Abstract

This project aims to analyze the crime patterns in the city of Buffalo, New York. The dataset was collected from Open Data Buffalo, (Buffalo's official open data program that makes public information publicly available). This dataset is updated daily and offers a preliminary look at crime reports in the City of Buffalo. We extracted the data using Socrata Open Data API. The consumed data was then stored in a csv file, from which the data was ingested in database using SQLite. EDA was then performed on the data to find underlying patterns, and visualized using various Python libraries. Finally, predictions were made to categorize if a certain type of incident is a Larceny/Theft or not.

Keywords :
Buffalo Crime Analysis, EDA, Python, Visualization, Matplotlib, SQLite, sodapy

1. Introduction

Crime dataset of any city include where, when and how an incident occurred, which aids in problem solving by developing patterns and trends, identifying investigative leads/suspects, and clearing cases. In the recent months, unfortunately, Buffalo's safety has been in question due to several incidents that have made their way in the global news.

Analyzing the crime patterns in the city will assist in police operations such as crime investigation, patrolling activities, and crime prevention.

2. Objective

By analyzing the crime dataset, we aimed to answer the following questions:

a. Analyze crime patterns over months and identify the month with the highest number of criminal activities recorded
b. Analyze most frequently occurring incidents(Larceny/Murder/Sexual Assault)
c. Analyze the most unsafe neighborhoods in Buffalo
d. Analyze if crimes happen more over weekends than weekday

3. Implementation

We used the public dataset provided daily by the Buffalo Police Department. It offers a preliminary look at crime reports in the City of Buffalo. The dataset consists of 287K rows with 29 columns, each with a crime incident **(Crime Incidents | OpenData Buffalo)**
The data was consumed via Socrata Open Data API. The data was then stored in a csv file from which the data was ingested in the database.

The data was normalized and stored in 3 tables namely: INCIDENT TYPE, INCIDENT INFO and INCIDENT LOCATION

University at Buffalo
Graduate School of Education
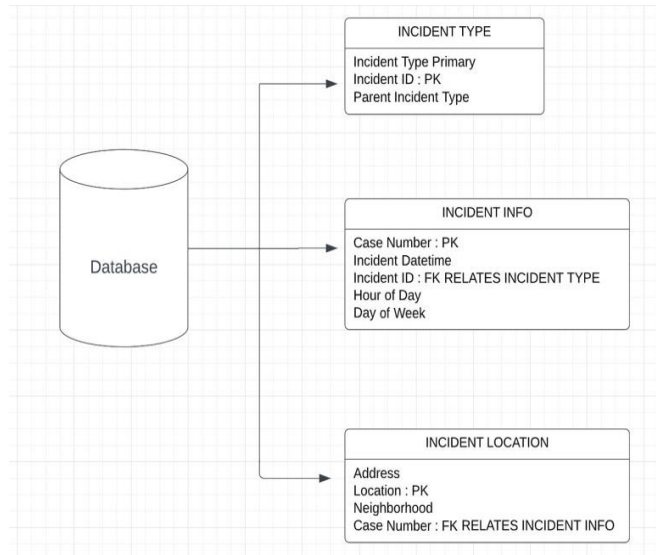
We queried the tables and stored the result in dataframes to find the following patterns:


Crime Distribution by Neighborhood
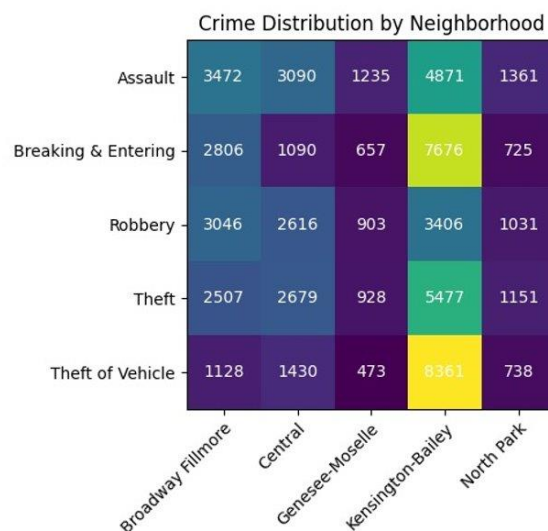
Through this heatmap we see that, the most common crime is Theft of Vehicle and Breaking and Entering which occurs majorly in Kensington-Bailey neighborhood.
Genesee-Moselle neighborhood has the least number of Theft of Vehicle crimes.

Crime Type in Buffalo
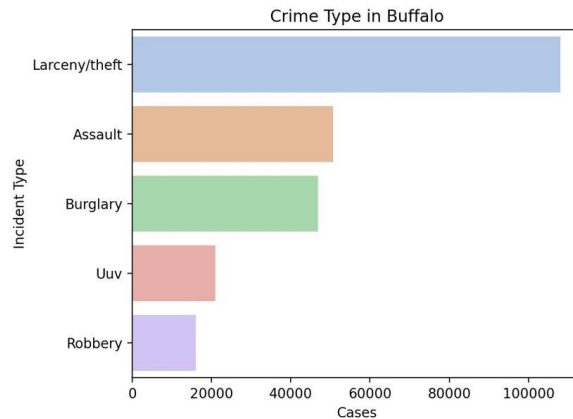
The bar graph shows the most common crimes that occur in Buffalo.
Larceny/Theft crimes constitute more than half the crimes reported, followed by Assault and Burglary.
The least occurring crime is SODOMY.

Most Dangerous Neighborhoods in Buffalo
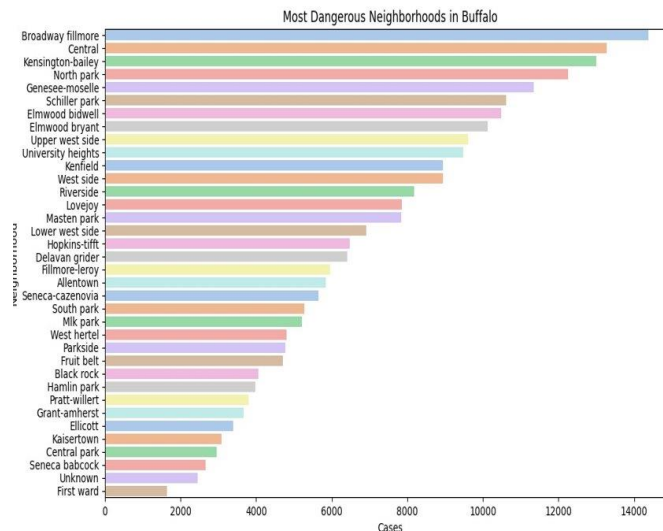
The bar graph depicts the frequency of crime in various Buffalo neighborhoods. Broadway fillmore is the most unsafe neighborhood with over 14,000 cases in total while First ward is the safest with less than 2000 cases.

Violent Crimes by Season
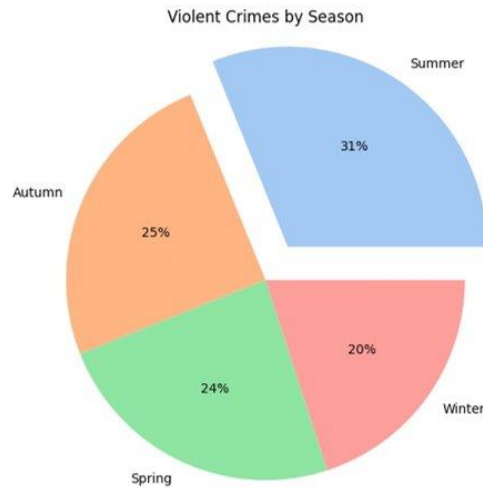
This pie chart shows the distribution of violent crimes (Assault, Sexual Abuse, Murder, Manslaughter, Rape , SODOMY) happening across seasons.

We see that the crimes are more common in Summer with 31% and least common in Winters with 20%.
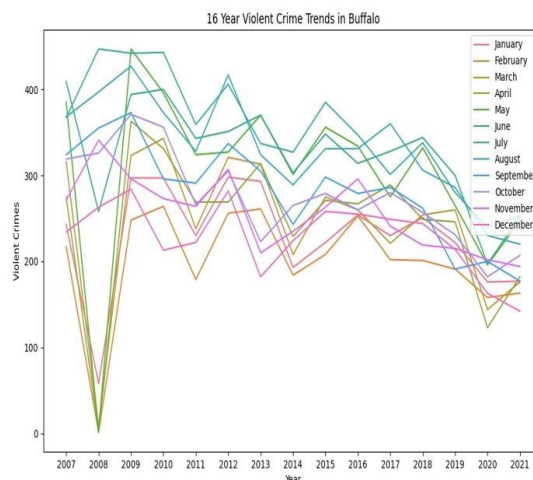
The line graph shows crime rate in every month for the last 14 years.

There is a sudden drop in crimes in 2008 and then 2020. We see that the summer months (June, July, August, September) have more reported crimes than the rest of the year.

Since the most frequently occurring incident is larceny/theft (~124k) we want to label if the incident is/is not larceny/theft based on Address, Neighborhood, Date. We achieved ~65% accuracy with XG Boost.

Future Work :

Using this dataset, and a time series model, we can predict when a crime can happen based on location, hour of the day, neighborhood, date etc. This will help the police identify potential unsafe areas and prevent crime.

University at Buffalo
Graduate School of Education

References :

1. https://www.kaggle.com/code/zikazika/analysis-of-world-crime
2. https://github.com/xmunoz/sodapy/blob/master/README.md
3. https://www.kaggle.com/code/ashkanranjbar/crime-analysis-in-the-city-of-chicago

University at Buffalo
Graduate School of Education