CRIME REPORTING IN THE STATE OF NEW YORK

Submitted to: E.A.S. (UB)

December 2021

Group No. 17

University at Buffalo
Graduate School of Education

Abstract

This report aims to visualize the crimes in the state of New York over a period of 30 years (1990 – 2020). We have captured the data from the New York state government website. The link for the csv file of the data - https://data.ny.gov/Public-Safety/Index-Crimes-by-County-and-Agency-Beginning-1990/ca8h-8gjq. A pipeline from EDA to visualization was carried out and was the m focus of this paper. We have visualized the crimes reported in the entire state by using a heatmap and shown a graph of how the crimes have decreased/increased over time from 1990 – 2020.

Keywords

Python, New York Crime, Visualization, SQL, Geo-Visualization, Jupyter Notebook

1. Introduction

Crime mapping has been used in crime analysis at least since 1900. The New York City police department uses it in many cases. The representation was merely consisting of traditional crime map with pins representing the location of the crimes and some of its characteristics. Now-a-days, crime mapping, and analysis upgraded from using pins to computer software such as GIS technology which has a significant influence on crime analysis. The demographic characteristics of the crime's locations and the data of the crimes collected by the police department regarding the features of cityscapes and landscapes are extensive and valuable data. It can be visualized on a map to analyze where, how and why crime occurs. The ability to combine all these data is implemented using GIS. The primary objective of crimes analysis and visualization provides information that can make the police department more efficient in carrying out its mission; to prevent and suppress crime. The main contribution of our paper is to find the most occurrence types of crimes and where, when and why the crimes occur. Furthermore, precautions could be suggested to the police department to prevent and suppress crimes.

2. Data Overview

```
df.head()
```

✓ 0.2s Python

	County	Agency	Year	Index Total	Violent Total	Murder	Rape	Robbery	Aggravated Assault	Property Total	Burglary	Larceny	Motor Vehicle Theft
0	Albany	Albany City PD	2020	3547.0	875.0	18.0	61.0	164.0	632.0	2672.0	430.0	1958.0	284.0
1	Albany	Albany County Park PD	2020	2.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	2.0	0.0
2	Albany	Albany County Sheriff	2020	127.0	12.0	0.0	4.0	0.0	8.0	115.0	11.0	96.0	8.0
3	Albany	Albany County State Police	2020	103.0	26.0	0.0	18.0	2.0	6.0	77.0	4.0	71.0	2.0
4	Albany	Altamont Vg PD	2020	6.0	2.0	0.0	0.0	0.0	2.0	4.0	0.0	4.0	0.0

Fig. 2 Cleaned Data overview

Fig. 2 shows us the cleaned data that we have made ready to use for further analysis and processing. The data consists of County's from NY with their respective agencies and the year column ranges from 1990 – 2020. The Index Total is the summation of Violent Total and Property Total.

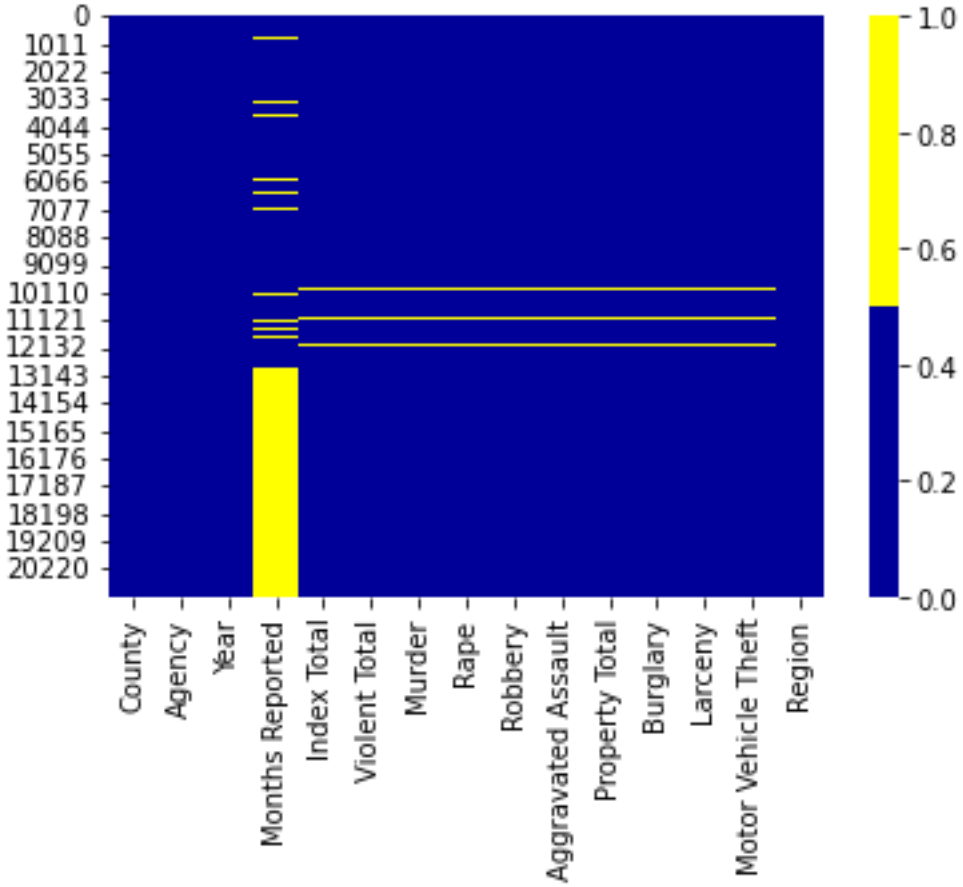

Fig. 2.1 Non-normalized data

Fig. 2.1 shows the uncleaned data (raw data off the web). We further had to do imputations and pre-processing to get the data in a form which we could work on and make visualizations of the same. The yellow parts in fig. 2.1 represents missing values in the dataset, whereas the blue part represents relevant piece of information in the dataset.

3. Schema

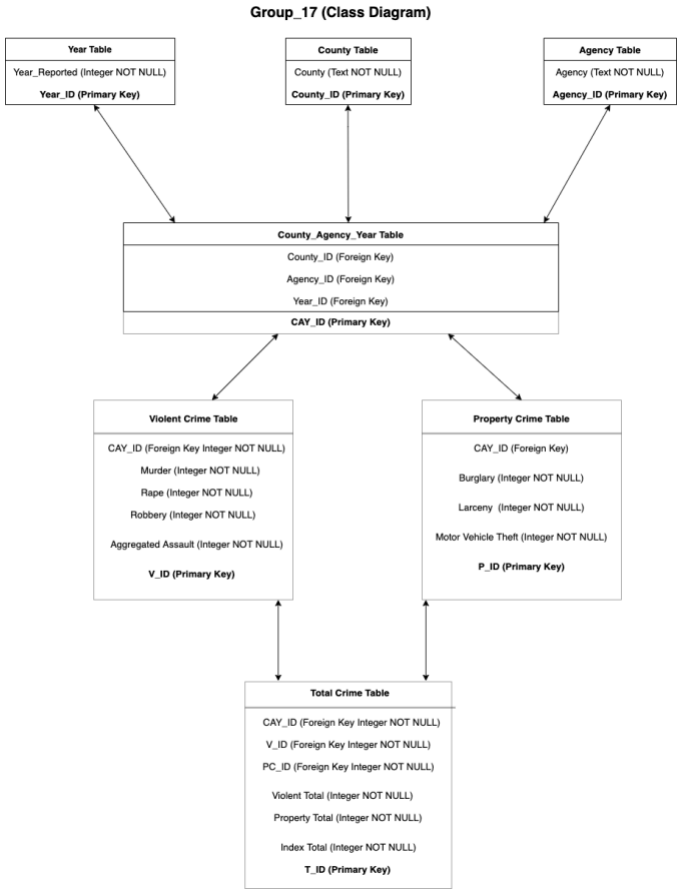

Fig. 3 Database Schema

The above image shows a pictorial representation of how we have structured the database. The County Agency Year Table in the middle of the diagram is the table that stores all the primary keys from various tables and acts as a connector to other tables. This table is not visualized outside. It is merely for the purpose of joining tables and is hidden in the data abstraction layer.

4. Results and Discussions

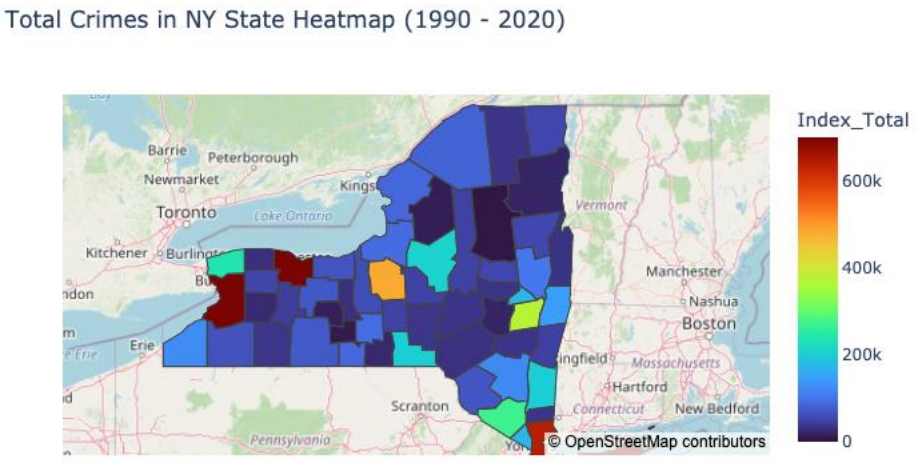

Fig. 4.1 Crime in NY Heatmap

Fig 4.1, the subset sample of the data represents a general insight about the distribution of crimes in the state of New York.

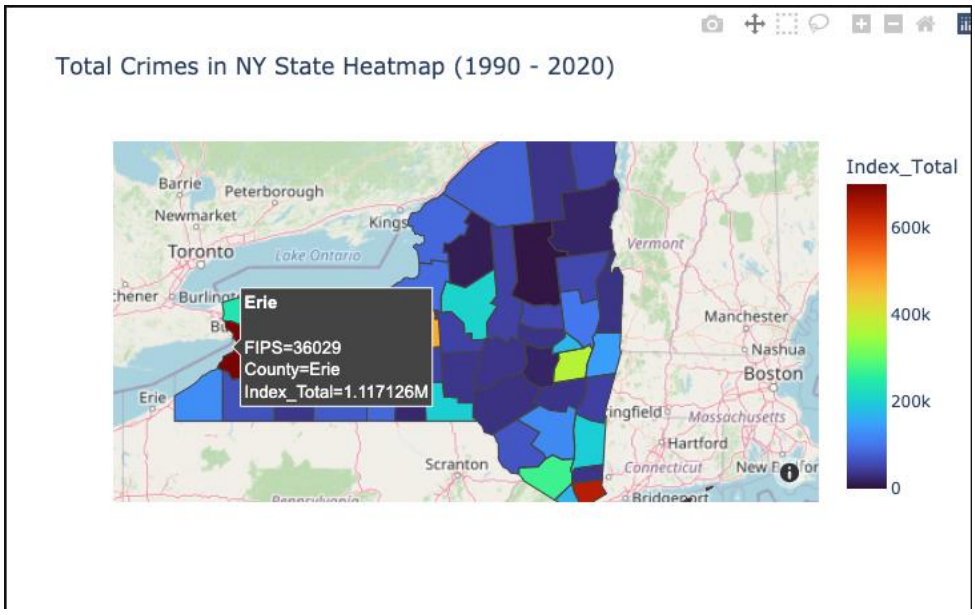

Fig. 4.2 Erie County Crime Visualization

Further dwelling into visualizations, when we hover the cursor over the heatmap, each county's Crime total is seen as it can be seen in Fig. 4.2. The number for Erie County in crime total is 1.17million from 1990-2020. The contributing factors to these numbers could be many like unemployment, recession, etc.

Fig. 4.3 Violent Crime vs Property Crime Comparison (1990-2020)

Fig. 4.3 shows a pictorial visualization how the violent crime and property crime has decreased in the state of New York from 1990 – 2020.

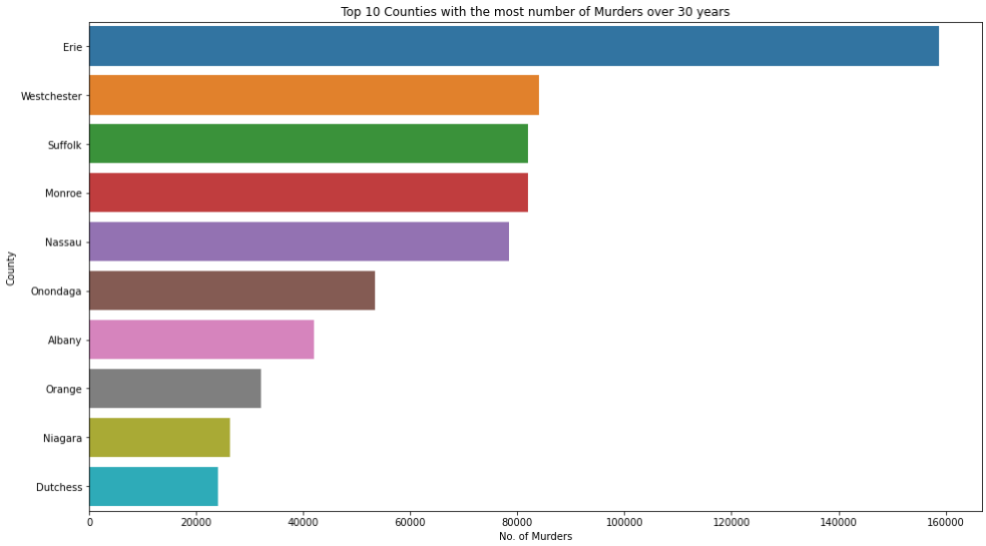

Fig. 4.4 Counties with most murder (Violent Crime)

We have then decided to visualize the top 10 county's contributing the most to Violent Crime. As seen in Fig. 4.3, the topmost crime in violent crime is Murder. Consequently, we visualized top county's contributing to the latter. We then see that Erie County is the highest in the Murder category of Violent Crime from 1990 – 2020.
Furthermore, we had dived more into it by dissecting the agency's that reported the greatest number of murder cases in Erie County as seen in Fig. 4.5.

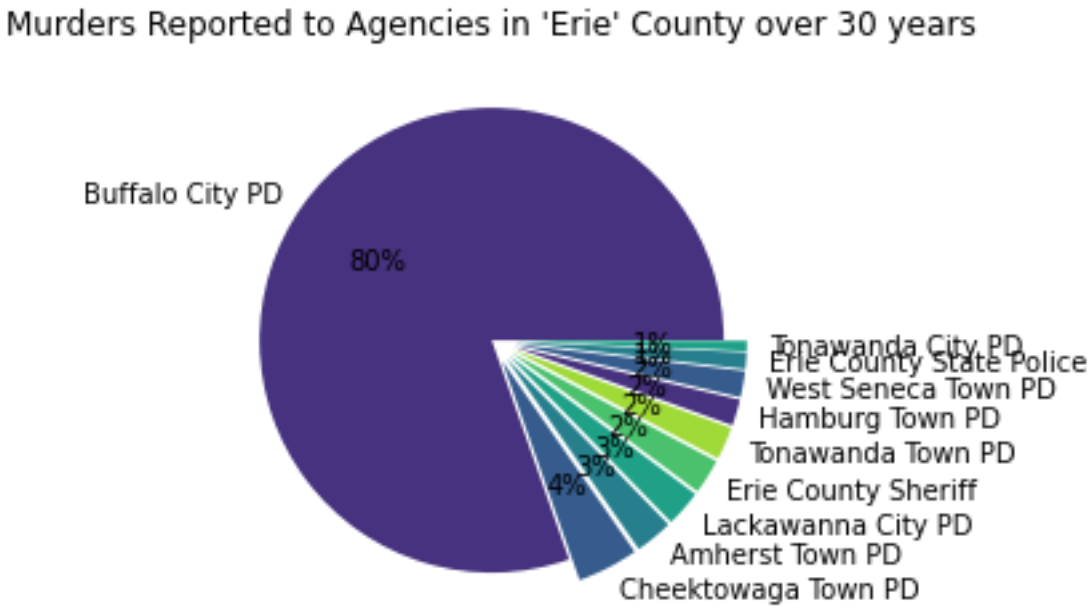

Fig. 4.5 Agency in Erie with the most Murder reported Cases

Similarly, we computed the analysis for Property Crime and as seen in Fig. 4.3 (RHS), Larceny was the most reported over the thirty years from 1990 – 2020. We then handled it in the similar way with first computing the top 10 counties in NY state that reported Larceny and got result as Suffolk County. (Fig. 4.6)

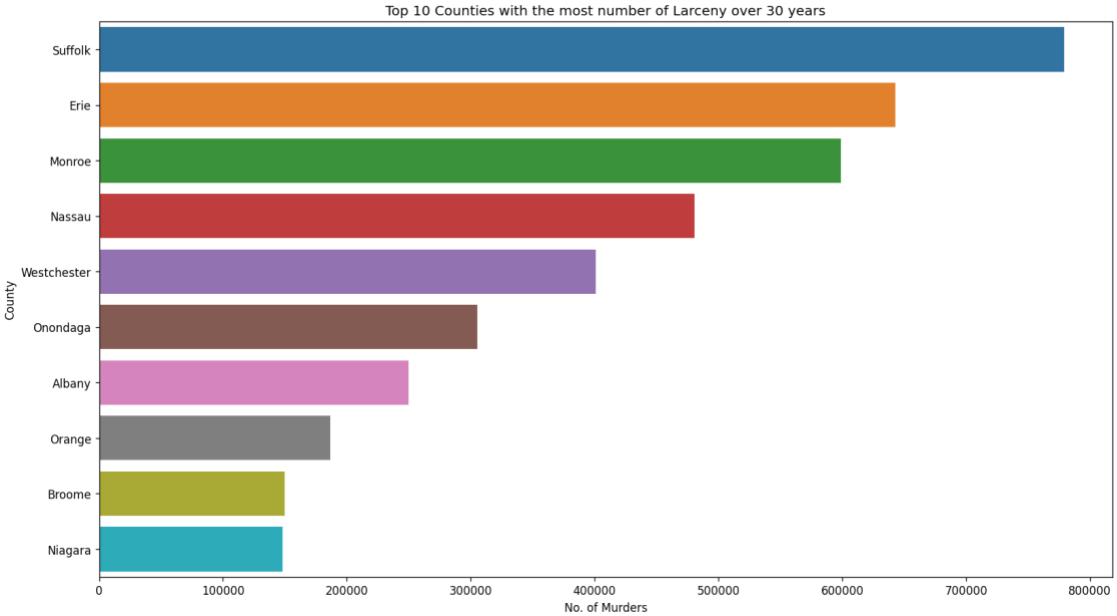

Fig. 4.6 County's with the most Larceny reported (Property Crime)

As a result, we then used SQL queries after data encapsulation into required tables and got the agency's that reported the most Larceny cases from 1990 – 2020 in Suffolk County.

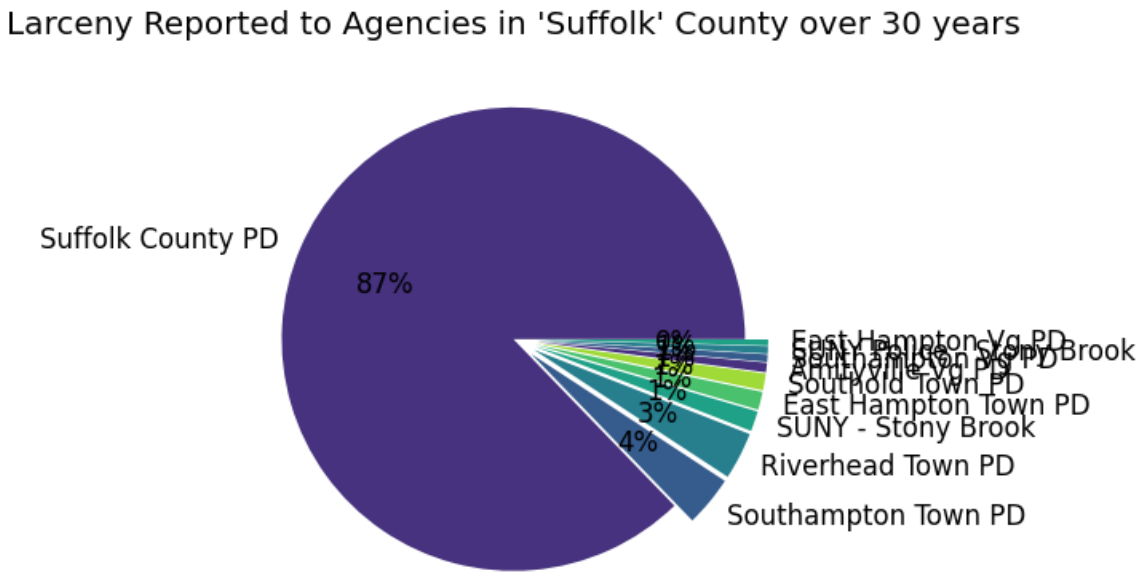

Fig. 4.7 Agency in Erie with the most Murder reported Cases

5. Conclusion and Future Work

The primary goal of crime visualization is to support the officers in decision making upon crimes. Therefore, our study focused on analysis and interacted visualization of crimes data with three types (crimes against the person, crimes against property, and crimes against society) from 1990-2020 in the state of New York, USA. To achieve that, we integrated SQLite and used matplotlib and other libraries in python for visualization. Accordingly, we got profound and actionable insights of 1) Erie County being the most contributing county to Violent Crime and 2) Suffolk County with the most Property Crime. Furthermore, we can incorporate the machine learning models to predict the likelihood of a crime in a particular county with predicting the time and day of a particular crime with the appropriate data. We are planning to build adaptive geo of crimes system to allow interactive geographical visualization of various data.

6. References

[1] National Institute of Justice, “Mapping and Analysis for Public Safety”.

[2] D. Vasiliauskas, G. Beconyt. “Cartography of crime: Portrait of metropolitan Vilnius.” Journal of Maps, 12:5, 1236-1241, 2016.

[3] S. Sim., Walker. W, J. Cook, R. Doyle, and L. Mathews,” Exploratory Spatial-Temporal Visualization of Hurricane Impacts on Crime Events in Miami, Florida,” 2011.

[4] Q.A. Kester. “Criminal Geographical Profiling: Using FCA for Visualization and Analysis of Crime Data.” ArXiv Prepr. ArXiv13100864, 2013.

[5] J. Hagenauer, M. Helbich, and M. Leitner. “Visualization of crime trajectories with self-organizing maps: a case study on evaluating the impact of hurricanes on spatiotemporal crime hotspots,” in Proceedings of the 25th conference of the International Cartographic Association, Paris, 2011.

[6] https://towardsdatascience.com/data-cleaning-in-python-the-ultimate-guide-2020-c63b88bf0a0d (Data Cleaning in Python).

[7] https://github.com/nikhilkumarsingh/choropleth-python-tutorial (Visualization in Python).