Introduction

The dataset consists of reported crimes in the City of Chicago for the year 2022. This dataset reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred. Data is extracted from the Chicago’s CLEAR (Citizen Law Enforcement Analysis and Reporting) system.

Objective

The objective was to get meaningful insights like how the crime is distributed in the city, how many arrests are made, how many domestic crimes were committed. These insights will determine the public safety of the city and city sections, school violence, police response etc. These visualizations and analysis are approximate and do not represent the exact crime analysis or police response

Methodology

Most of the plots depended on showing the number of criminal cases based on location, crime type, wards etc., since the case number is unique alphanumeric a Seaborn *countplot* seemed better approach to count the number of cases for each criterion

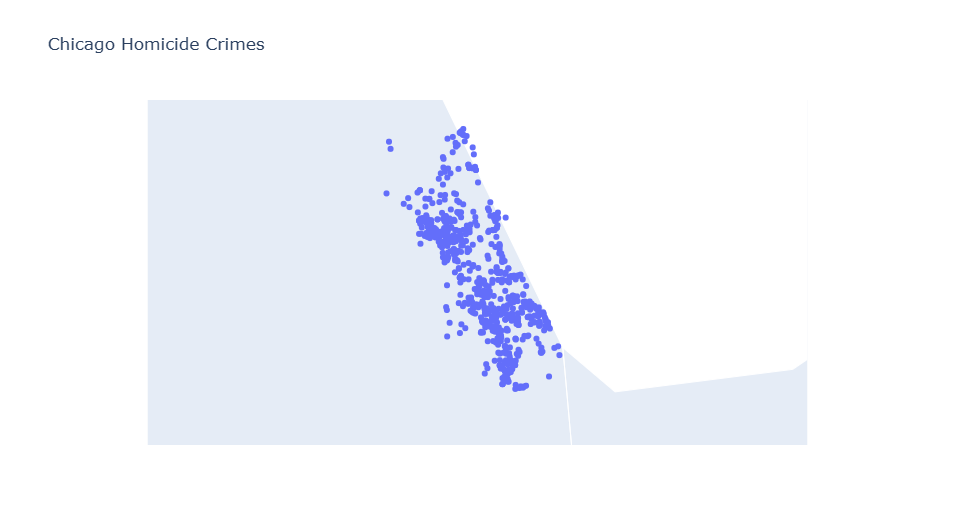
Pie charts were used to show percentage of crimes committed at Bank and percentage of crimes by public and private school. The pie chart gives a better breakdown when you want to see the % breakdown and derive conclusion about most dominant factor

Geographic maps were chosen to show the cases reported as Latitude and Longitude was available in columns. Visualization on geographic maps shows the spread of the cases and easier classification of locations. This gives the audience a quick view of the affected city locations rather than just counts by areas

This dataset is publicly available from City of Chicago data portal. The city regularly updates the data and available for download and analyze.

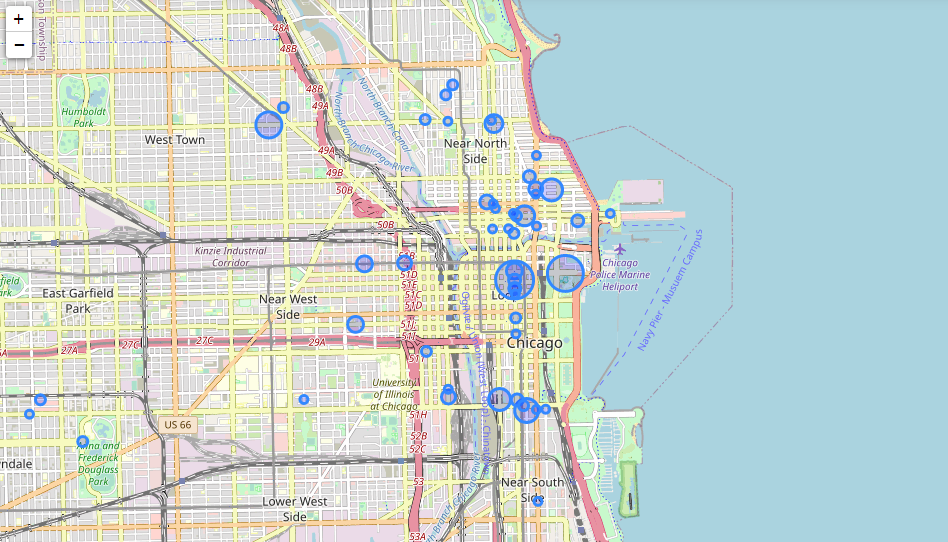
Homicide by geographic locations:

The dataset was filtered and new dataset created for total of case numbers grouped by Longitude and Latitude. Plotly geographic map was used to display the locations with markers



Crimes by geographic locations

New dataset created for total of case numbers grouped by Longitude and Latitude. Folium library geographic map was used to display the locations with markers

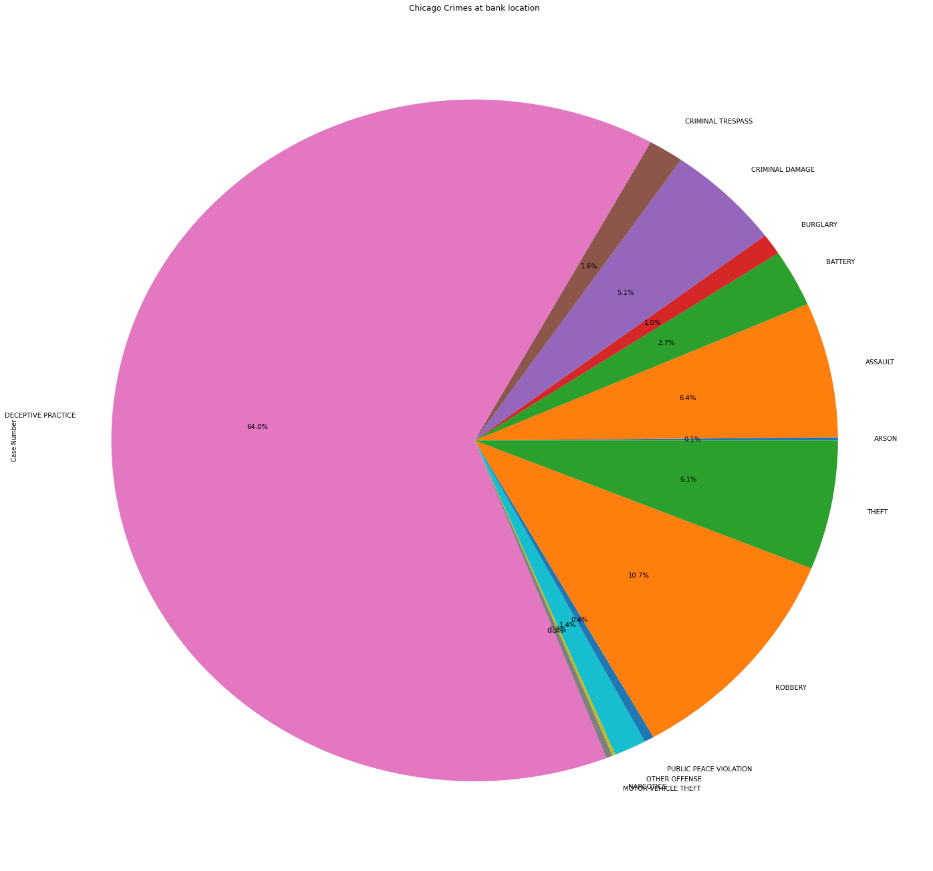


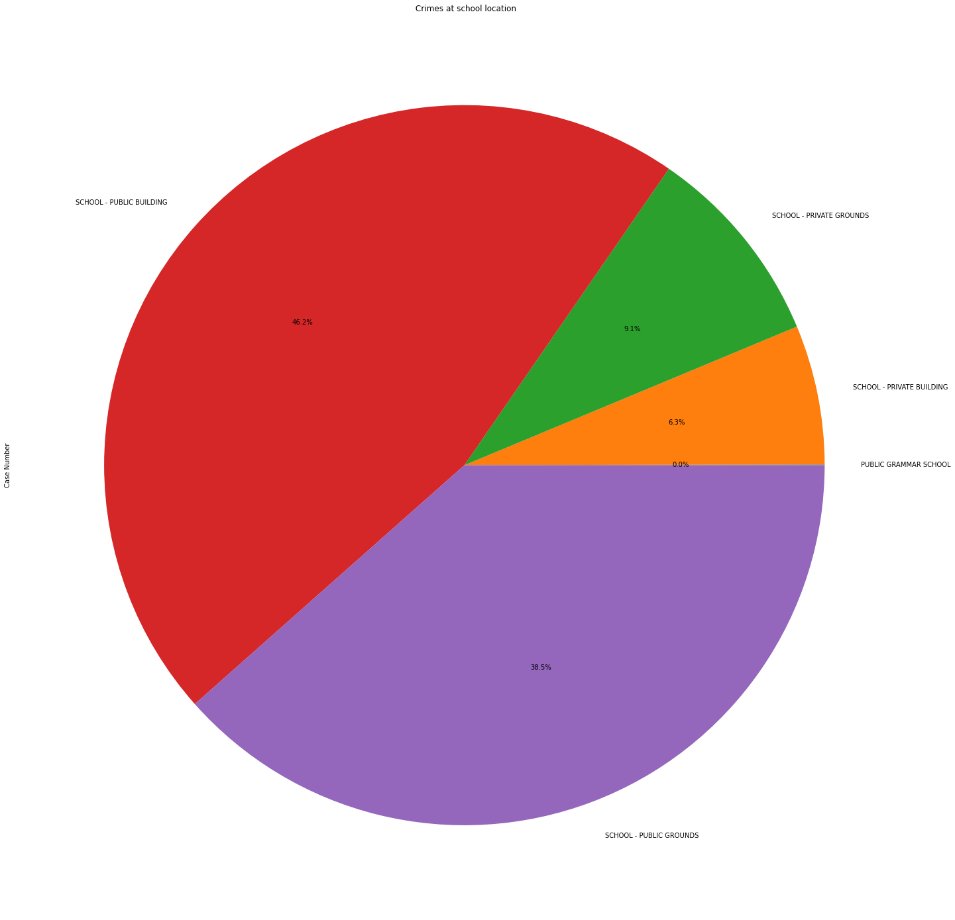
Below Bar plots were created for visualization in Python/R

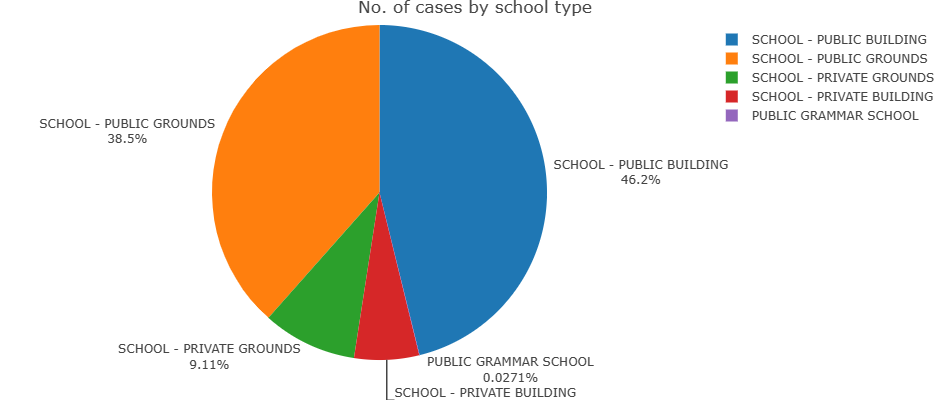
* Bar plot showing Total No. of cases vs Crime type (Python)
* Bar plot showing Total No. of cases by crime type and arrests for each crime type(Python)
* Bar plot showing Total No. of cases vs Ward(Python)
* Bar plot showing Total No. cases vs Domestic(Python)
* Bar plot showing Toal No. of cases vs Arrest made(R)
* Bar plot showing Total No. of Domestic cases vs Arrest made (R)
* Bar plot to show Total No. of crimes by each month. A new column was created by stripping month out of Date column in the dataset(Python)
* Bar plot to show Total No. of crimes by Community area(Python)

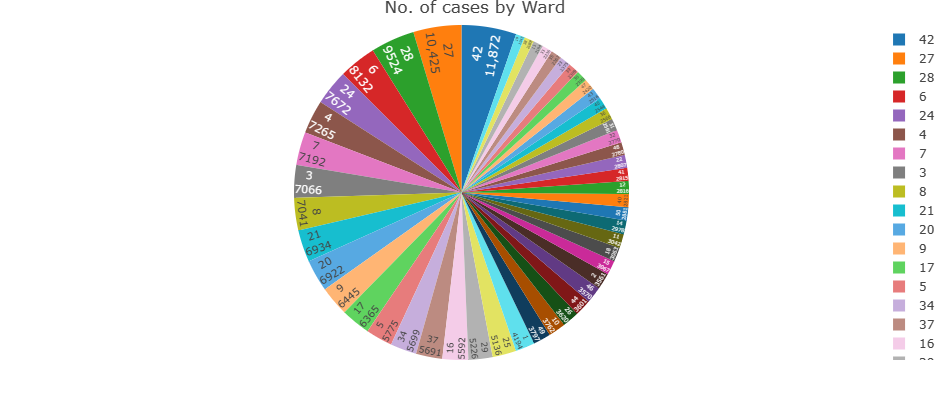
Below Pie charts were created for visualization in Python/R

* Pie chart showing type of Crimes percentage committed at *Bank* location. The dataset was filtered and new dataset created for total of case numbers grouped by crime type(Python/R)
* Pie chart showing percentage of crimes at *School* locations. The dataset was filtered and new dataset created for total of case numbers grouped by Location(Python/R)
* Pie chart showing Total No. of crimes vs Ward(R)









Data

Some of the key fields of the dataset used for visualization

|  |  |
| --- | --- |
| Column Name | Description |
| ID | Unique identifier for the record. |
| Case Number | The Chicago Police Department RD Number (Records Division Number), which is unique to the incident. |
| Date | Date when the incident occurred. this is sometimes a best estimate. |
| Primary Type | The primary description of the IUCR code. |
| Location Description | Description of the location where the incident occurred. |
| Arrest | Indicates whether an arrest was made. |
| Domestic | Indicates whether the incident was domestic-related as defined by the Illinois Domestic Violence Act. |
| Ward | The ward (City Council district) where the incident occurred |
| Latitude | The latitude of the location where the incident occurred |
| Longitude | The longitude of the location where the incident occurred |

* Any rows with Null(NA) values was deleted using *dropna* in Python and *na.omit* in R.
* Slicing the dataset to *groupby* by Latitude and Longitude. Create a new dataset and rename the column for count of cases
* Slicing the dataset to filter where crime type is Homicide, to use it for Plotly mapping
* Slicing of data was done to create new dataset when the crime location is BANK and *groupby* crime type
* Slicing of data was done to create new dataset when the crime location contains \*School\* and *groupby* crime type.

Conclusion

About **64**% of the crimes committed at the Bank was of nature **Deceptive Practice** whereas **robbery, theft** etc. accounted for about **25%** only. This means the crimes were of financial nature contrary to thinking that crimes reported at bank are usually robbery or theft

Around **84%** of the crimes reported at school were for **Public** schools with only 15% of school crimes were at **Private** schools. The number of public schools are higher meaning the total number of cases will be greater than smaller number private schools (Based on the fact the Chicago Public District is 3rd largest in the country). However, the wide difference leans towards that public school crimes are higher than private

The number of crimes is the most in the month of **October** and higher in the months of July, August and September. The crimes are lowest in **December** and lower in January and February. This can be attributed to the weather in Chicago with extreme winters and milder summers

Only ~ **16**% of the reported Domestic cases result in an arrest

From the geographic map visualization of the crimes, Homicide crime was most concentrated towards the South and crimes in general were more in the Downton area

It would like to include more colored palettes for some of the bar plots instead of simple colors. The categorization on x-axis for the bar plot was overlapping as the categorization name were longer. I increased the plot size to accommodate but was expecting better solution to include longer names

Sources

cityofchicago.org. (2023, Jan 2) Crimes - 2022. data.cityofchicago.org. https://data.cityofchicago.org/Public-Safety/Crimes-2022/9hwr-2zxp

cityofchicago.org. (2023, Jan 2) Crimes - 2001 to Present. data.cityofchicago.org. https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2