

Data Visualization of Crimes Database Chicago.

Sanchit Agarwala

December 30,2020

1.Introduction

The idea of this study is to reflect the reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent seven days. If the data set is viewed it can be concluded that the data set is quite complex to ground zero the information regarding the crimes happening in a specific are or location of Chicago, which are the crimes happening more frequently compared to other ones in a specific area, where there is a need to increase the number of police commandos, etc.

2.Dataset Description

This dataset reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent seven days. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. In order to protect the privacy of crime victims, addresses are shown at the block level only and specific locations are not identified. Should you have questions about this dataset, you may contact the Research & Development Division of the Chicago Police Department at PSITAdministration@ChicagoPolice.org.

Disclaimer: These crimes may be based upon preliminary information supplied to the Police Department by the reporting parties that have not been verified. The preliminary crime classifications may be changed at a later date based upon additional investigation and there is always the possibility of mechanical or human error. Therefore, the Chicago Police Department does not guarantee (either expressed or implied) the accuracy, completeness,

timeliness, or correct sequencing of the information and the information should not be used for comparison purposes over time. The Chicago Police Department will not be responsible for any error or omission, or for the use of, or the results obtained from the use of this information. All data visualizations on maps should be considered approximate and attempts to derive specific addresses are strictly prohibited. The Chicago Police Department is not responsible for the content of any off-site pages that are referenced by or that reference this web page other than an official City of Chicago or Chicago Police Department web page.

The user specifically acknowledges that the Chicago Police Department is not responsible for any defamatory, offensive, misleading, or illegal conduct of other users, links, or third parties and that the risk of injury from the foregoing rests entirely with the user. The unauthorized use of the words "Chicago Police Department," "Chicago Police," or any colourable imitation of these words or the unauthorized use of the Chicago Police Department logo is unlawful. This web page does not, in any way, authorize such use. Data are updated daily. To access a list of Chicago Police Department - Illinois Uniform Crime Reporting (IUCR) codes, go to <http://data.cityofchicago.org/Public-Safety/Chicago-Police-Department-Illinois-Uniform-Crime-R/c7ck-438e>

To provide the people the necessary information, I'll be combining Chicago Crime Data's 2001 to present data that contains Case Number, ID, Date District, etc.

3.Methodology

Following steps will be taken into account to reduce the complexity and make it clearer:

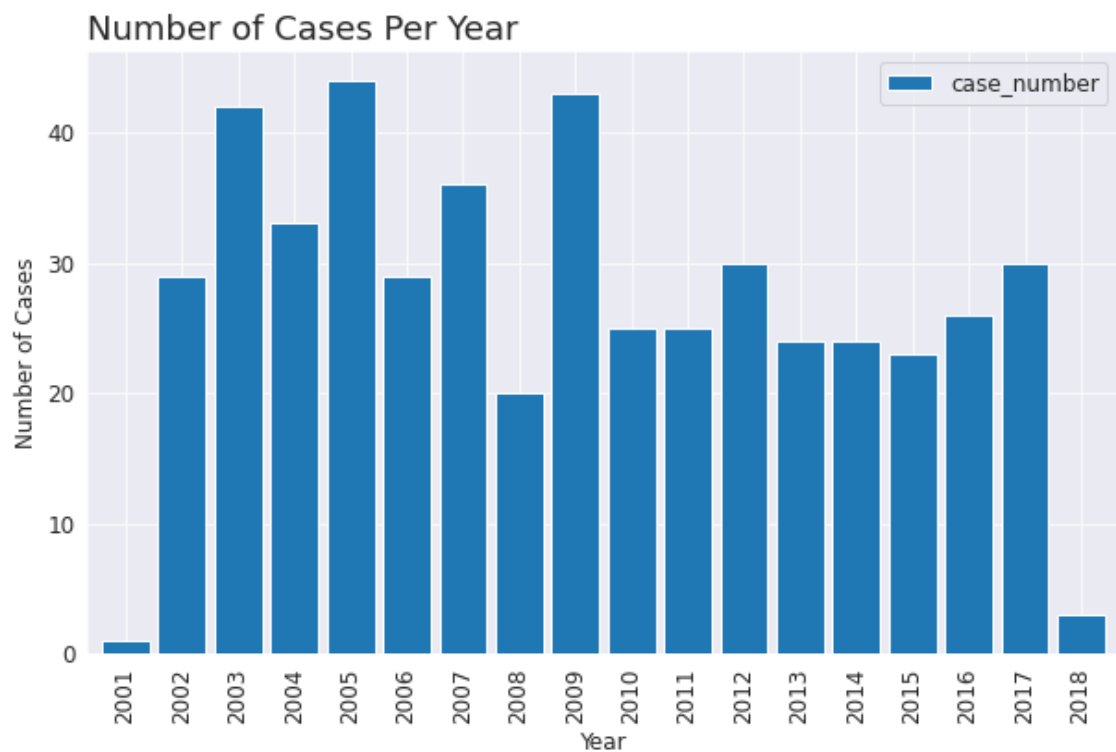
1. Importing the needed libraries.
2. Loading the dataset.
3. Displaying the details of the database.
4. Clean up the data and prepare for visualization.
5. Data Visualization of the number of crimes.

6. Create a folium map with a different colour per crime
7. Data Preparation for Modelling
8. K Nearest Neighbour (KNN)

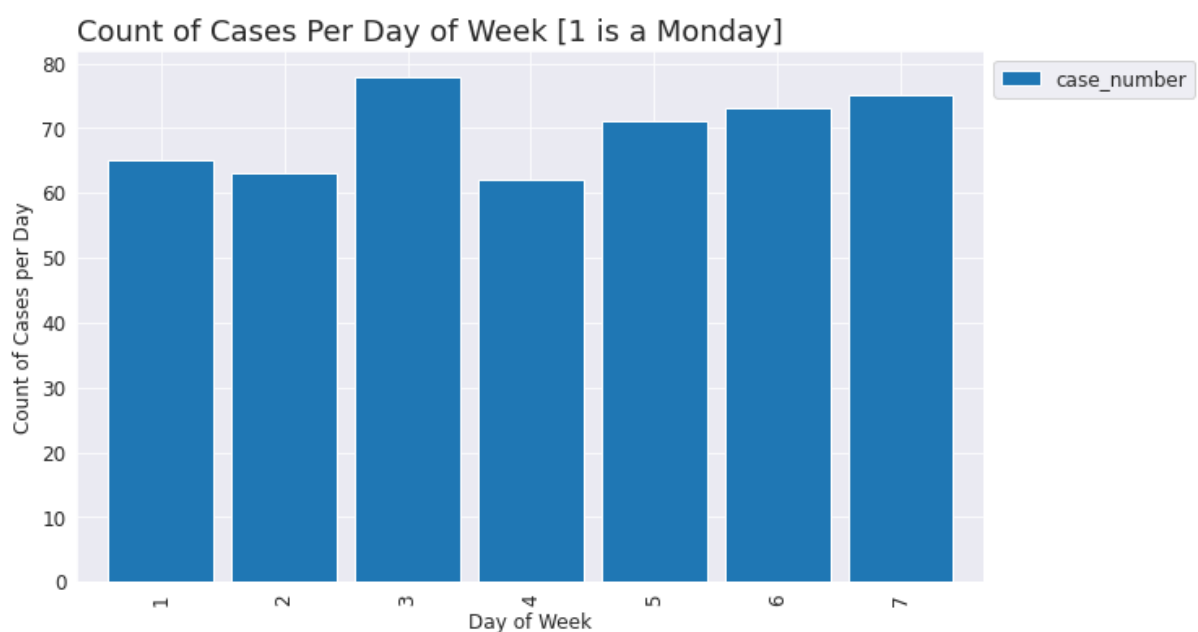
4.Results

Data Visualization:

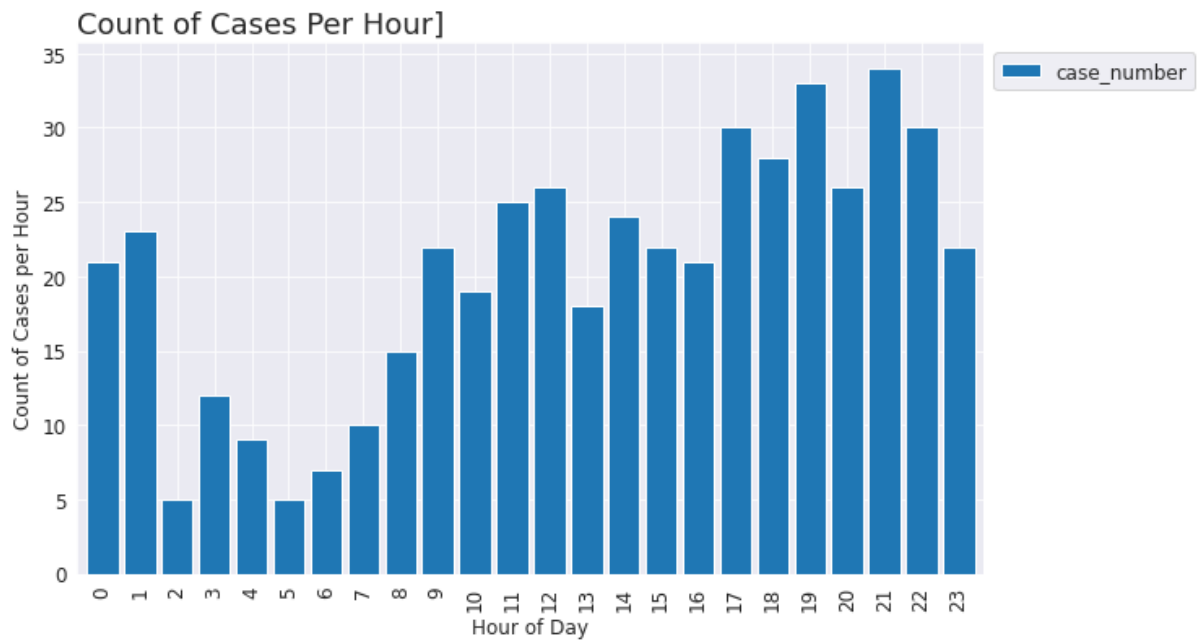
1. Number of Crimes per Year



2. Number of crimes occurring on each day

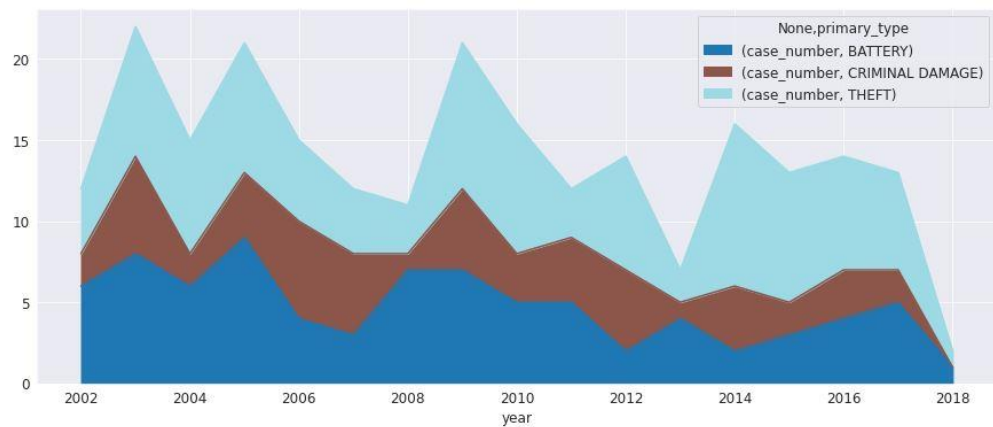


3. Number of crimes occurring in each hour

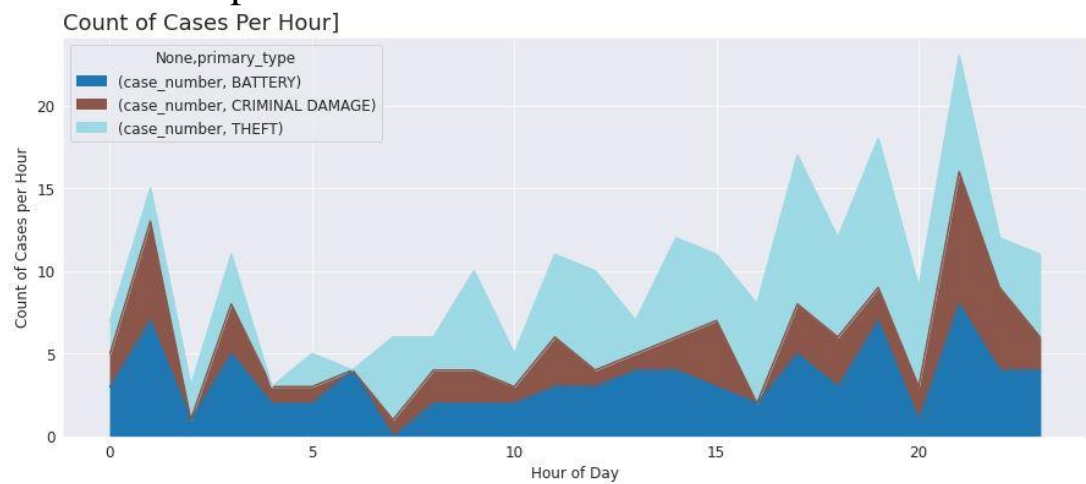


4. Primary type per year

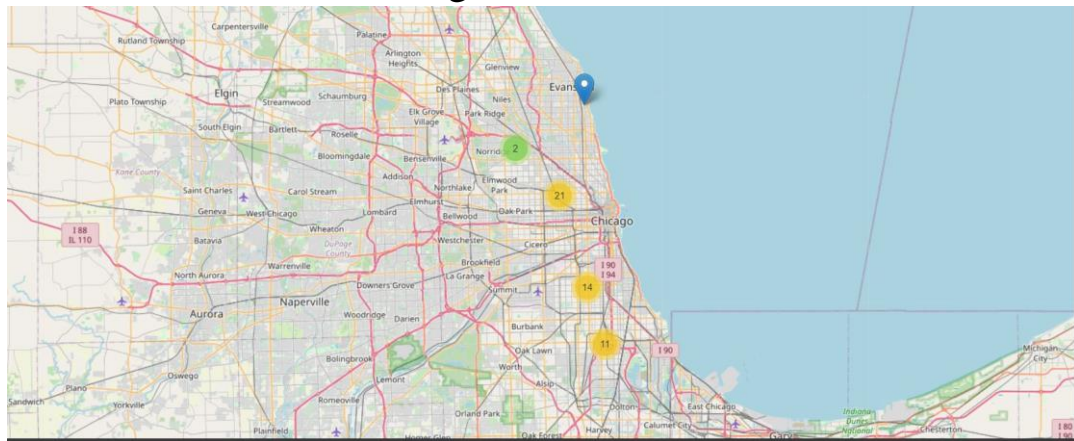
Out[19]: <matplotlib.axes._subplots.AxesSubplot at 0x7fd4250ecbd0>



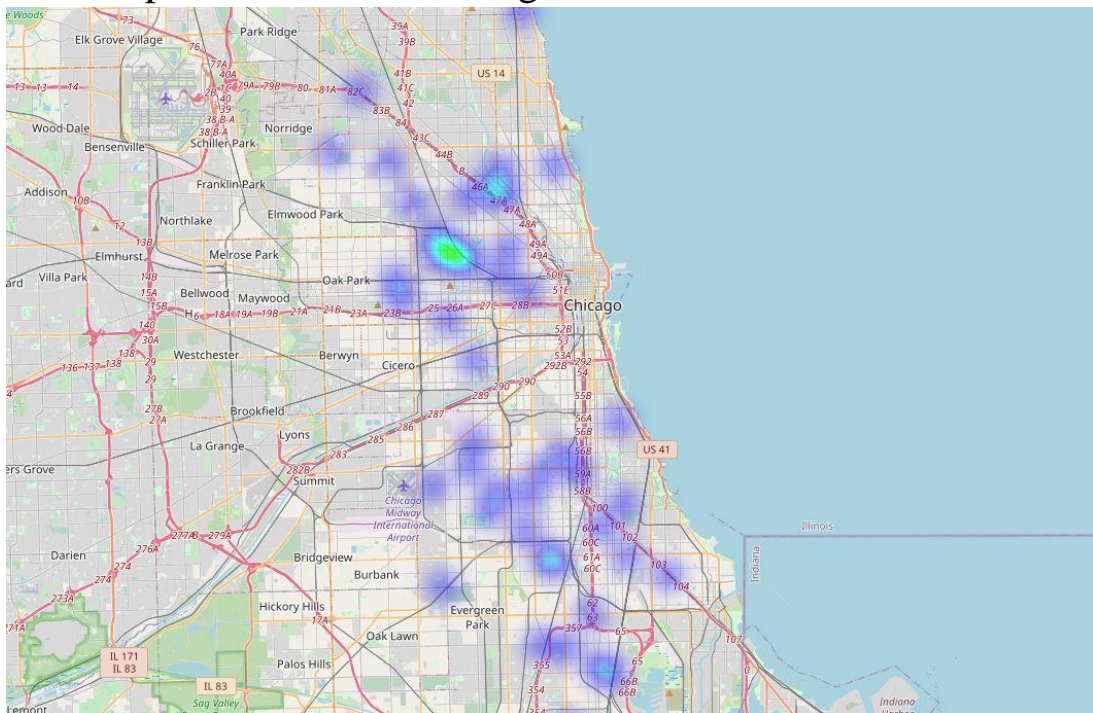
5. Count of Cases per Hour



6. Crimes in the month of August

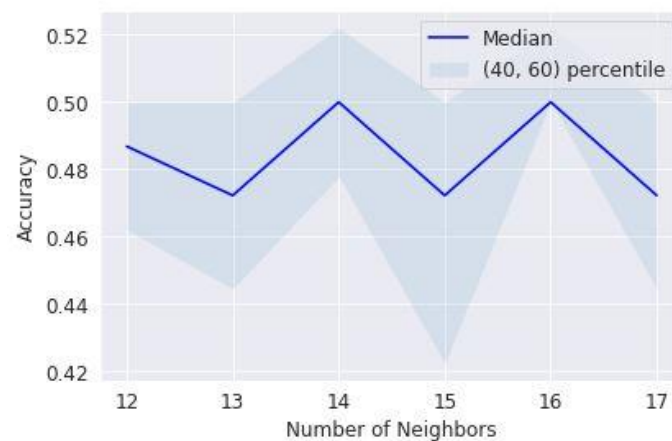


7. Heat Map for the month of August

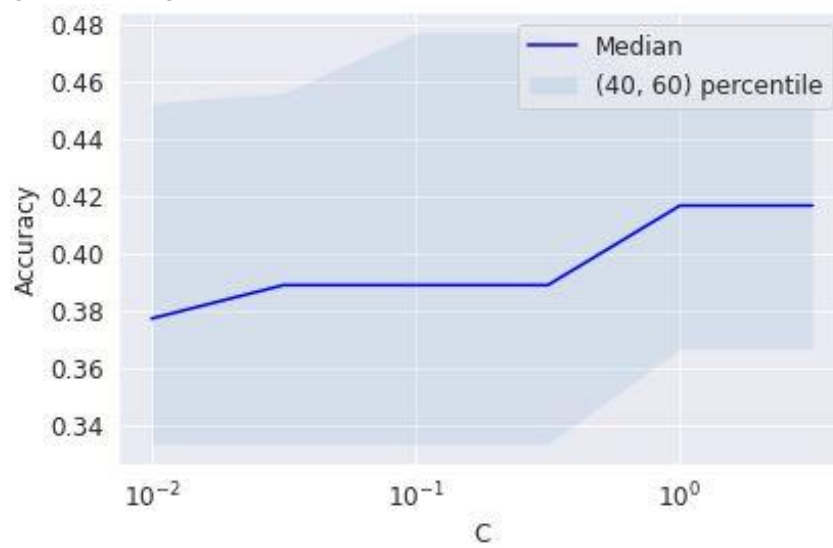


8. K Nearest Neighbour (KNN)

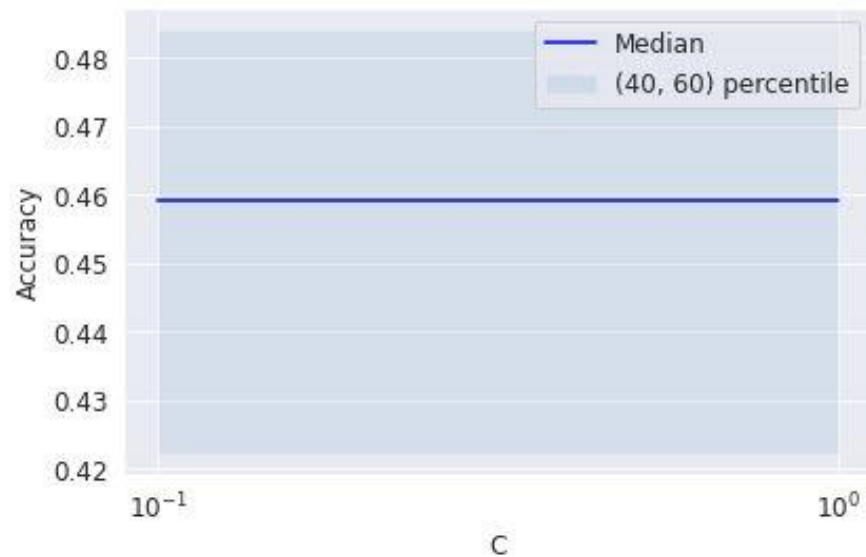
i) Decision Tree



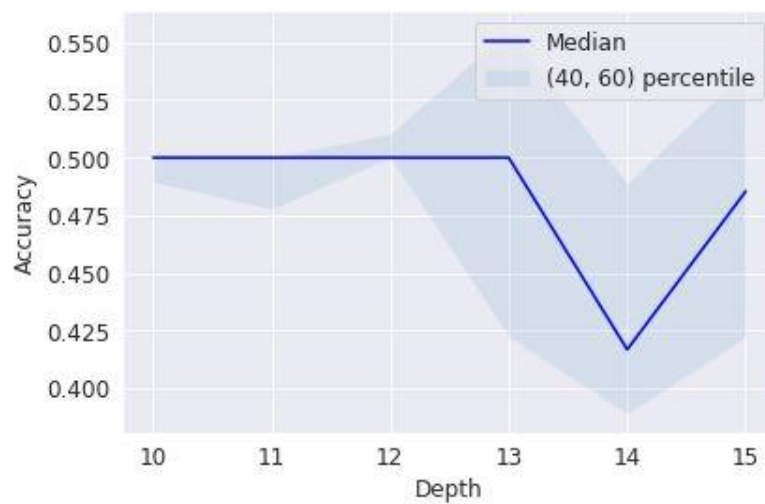
ii) Logistic Regression



iii) Naïve Bayes



iv) Using a Random Forest



5. Conclusion

In this study, I analysed the relationship between crimes improvement/decline and biographic data. I identified the variation in the demographics and spread of the crimes in the city of Chicago. I built both regression models and classification models to predict whether and how much crimes would improve/decline. These models can be very useful in helping the police department in a number of ways.