

CHICAGO CRIMES DATA ANALYSIS

Team members: Ajinkya Pawale (ajpawale@iu.edu) , Anuhya Sankranti (ansankra@iu.edu) , Sharan Sumbad (ssumbad@iu.edu) , Shivani Vogiral (svogiral@iu.edu)

Executive Summary:

Almost every country is experiencing a tremendous rise in crimes. There is a pressing need to discover crime patterns and study various types of criminal activity. Security agencies across the globe are working hard to reduce these crimes; nevertheless, the volume of crime data is continually growing, making it challenging to manage such a large amount of data and keep track of crimes that occur across many geographies and over time periods. As a result, having a criminal information system that can process massive amounts of data in a short amount of time is critical. Through this project we worked on the Chicago crimes dataset to analyze different factors effecting the increase in crime rate in Chicago.

There are so many records and documentation in the police department that have been gathered during the years, which can be used as a valuable source of data for the data analytics tasks. Applying analytical task to these data bring us valuable information that can be used to increase the safety of our society and lower the crime rate.

In this project we are trying to analyze the proportion of arrests made for a particular crime after it was reported to the Chicago Police Department. There are various variables in the dataset such as crime type, location, arrest made etc. Our goal is to predict whether the arrest was made or not for a particular crime depending on the other variables in the dataset like crime type, location, month, day, etc.

For the year 2019, based on our analysis only 19 % of the crimes that were reported, were arrested. Out of all the crime types ‘theft’ was the highest committed crime with 20.85 % of the all the crimes reported, out of which only 5% were marked arrested. So, if you are planning a trip to Chicago anytime soon beware of theft and even if so, you end up in such a situation there is very thin changes that your case will be solved. To our surprise we noticed that assault has higher chances of being solved when compared to theft. Assault is a major crime type, consisting 18% of the total crimes reported.

After our analysis we found that, there is significant amount of interaction between the type of crime and location with respect to the crime leading to an arrest. Domestic crime variable i.e., whether the crime was domestic or not, doesn’t contribute much individually towards the analysis of the crime leading to an arrest. The variation between type of the crime and location of the crime with factors like day of the week and month is almost constant hence, we can say that they don’t contribute much towards the arrest.

In Chicago, the kind of crime and the location of the crime are crucial variables in determining whether a case is considered solved. They can be further analyzed by looking at the interaction between location and crime type on a model and the interaction of domestic crime variable with the crime type.

We have taken the data for 2019 for our use case and would like to work on data from other years and see how the interpretations match to them as compared to our assumptions for 2019. Also, we can implement a similar model for major cities across U.S and see how the trend varies for each one of them.