Chicago Crime Analysis

Arjun Ravikumar, Tejal Shanbhag , Utkarsh Havle

1. INTRODUCTION

Crimes are something which happen every day in all parts of the world. Some crimes are organised crimes and others are unorganised. There have been multiple risk factors that have been the underlying reason for these crimes happening. Three major categories of crime risk factors: (1) biological factors; (2) socioeconomic factors; and (3) psychological factors, as a result of the individual value system of those involved in criminal activity. Chicago is a city which has a history of having one of the highest violent crime counts in the country. According to [3] Chicago has had 40,000 homicides from 1957 - 2020. For the first time since 1957 the homicides went below 500 for a year in 2019 in chicago. This might not seem like a large number but we need to consider that these numbers are of just a city and not a state in the US. From the beginning of the 20th century the Chicago Police department has been tracking the crimes happening in the city and making them accessible to the general public through their data website [2].

We propose to study these relationships between the socioeconomic factors of the residents of the area to the amount of crimes happening in that area. The dataset we are utilising for the same is of Hardship Index [4] and Crimes: 2001-present[1] from the City of Chicago's data portal.The Crimes: 2001-present dataset[1] consists of 7,279,930 rows of crimes between 2001 - February 18th 2021. This dataset[1] has 22 attributes which are id, case_number, date, block, iucr, primary_type, description, location_description, arrest, domestic, beat, district, ward, community_area, fbi_code, x _ coordinate, y_coordinate, year, updated_on, latitude, longitude and location. The HarshShip Index data [4] consists of 78 rows and has a selection of six socioeconomic indicators of public health significance and a "hardship index". This data[4] has 6 attributes which are hardship_index, community_area_name, percent _ households _below _ poverty, percent_aged_25_without_high_school_diploma, percent_aged_16 _ unemployed, percent_aged_under _18_or _ over _64 and per_capita_income_. In this project we intend to do data exploration where we try to visualise the data to analyse the trends in the data. We try to visualize the number of crimes against the year, day of the week, hour of the day, location. We also try to visualise the type of crimes and their count and the severe and not severe crimes against the hardship index and income. Finally we try to predict the severity of the crime using Naive Bayes Classifier.

The rest of the study would be conducted as follows: Section 2 provides the motivation for choosing this topic. Section 3 provides information about the design of the project. Section 4 describes the implementation and analysis of the methods and data respectively in detail. Section 5 will describe inferences from the analysis followed by the future work in the project.

2. REFERENCES

[1] Chicago Police Department. Crimes - 2001 to present. https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2. Accessed: 2021-02-19.

[2] City Of Chicago. Chicago data portal. "https://data.cityofchicago.org". Accessed: 2021-02-19.

[3] Kyle Bentle & Jonathon Berlin & Ryan Marx & Kori Rumore. 40,000 homicides: Retracing 63 years of murder in chicago. https://www.chicagotribune.com/news/breaking/ct-history-of-chicago-homicides-htmlstory.html. Accessed: 2021-02-20.

[4] U.S Census Bureau. hardship index: Census data - selected socioeconomic indicators in chicago, 2008 – 2012. https://data.cityofchicago.org/Health-Human-Services/hardship-index/792q-4jtu. Accessed: 2021-02-15.