

Final Project Proposal

S670

Group Members:

- Shivani Vogiral (svogiral@iu.edu)
- Ajinkya Pawale (ajpawale@iu.edu)
- Sharanbasav Sumbad (ssumbad@iu.edu)
- Anuhya Sankranti (ansankra@iu.edu)

Project DataSet:

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 2019 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.

Data Description:

The data is available at

<https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2>

For the project we are considering only crime data from 2019 to present,

The data was queried from the above mentioned site using bigquery API out of the total 22 columns and 637729 rows (observations) from the year 2019- present we are only using 15 column for data analysis they are as follows:

- **Date** : Date when the incident occurred. This is sometimes a best estimate.
- **ID**: Unique identifier for a record
- **Primary_type** :The primary description of the IUCR code.
- **Description**: The secondary description of the IUCR code, a subcategory of the primary description.
- **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.
- **Year**: Year the incident occurred.
- **Updated_on**:Date and time the record was last updated.
- **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.
- **X_coordinate**:The x coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.

- **Y_coordinate:** The y coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.
- **Beat:** Indicates the beat where the incident occurred. A beat is the smallest police geographic area – each beat has a dedicated police beat car. Three to five beats make up a police sector, and three sectors make up a police district. The Chicago Police Department has 22 police districts.
- **Community_area :** Indicates the community area where the incident occurred. Chicago has 77 community areas.
- **Location:** The location where the incident occurred in a format that allows for creation of maps and other geographic operations on this data portal. This location is shifted from the actual location for partial redaction but falls on the same block.
- **Block :** The partially redacted address where the incident occurred, placing it on the same block as the actual address.
- **District:** Indicates the police district where the incident occurred.

Research Questions:

The research's major Question would be: **How many days would it typically take for a Crime's Report to be marked as arrested ?** Based on the Current data there is no straight way answer for now , but with some analysis and preprocessing with available variables we try to answer the Question. Along the way we also analyse

- **Monthly Crime Trend over the years:** During the exploratory analysis we will also observe the monthly trend in crime for the respective years. Is there a certain increase in type of crime (primary type) during any specific time of the year? Average reported crimes on a monthly basis ?
- **Domestic crime trend:** How have the reported domestic crimes been marked arrested/not arrested compared to the ones which or not considered domestic (since domestic variable is available in the dataset why not analyse it), How many days does it typically take for a domestic crime to be marked as arrested.
- **Location trend of the crimes:** Does location have any influence in the crime committed or type of crime committed, based on the police district/beat/ward or block/community area: how does the crime vary ? Can location of a crime variable such as (x_coordinate,y_coordinate , location) be used as a predictor to predict if an arrest will be made or not ?
- **Conclusion:** Based on all the analysis is there an approximate number of days that we can say that for a reported crime to be marked as arrested keeping in mind the type of crime. A plausible cause and effect story in the monthly trend of the crime committed in the windy city of Chicago.

Note: since the data was available from 2001 to present due to computational restrictions we will be working with the more recent time frame i.e 2019-2021

