Project 2 Proposal

Francisco Gutierrez, Cynthia Hu, David Harding

Analysis of Chicago Crime Data

Data Set :

https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2

Given the size of the dataset (over 6.2 million records covering fifteen years) we are considering various options to narrow the focus of our analysis. Potentially we will consider a more limited time period (e.g. 5 years), and or eliminate some of the less serious incidents (e.g. theft below a certain threshold), or exclude Domestic incidences to focus more on street crime or homicide (which has 'just' 8099 records).

Note in the proposal below that specific columns in the database are Capitalized.

Part 1: Executive summary of the project and conclusions or insights from the data

- Summary of the main conclusions
- Brief overview of the dataset and some of the challenges and constraints to the analysis encountered during the Project.

Part 2: Annual snapshot: the overall picture of crime in Chicago

Taking two twelve month time periods from the beginning and end of the dataset, we will give a snapshot of:

- A breakdown of the number of crimes by Primary Type (two pie charts or stacked bars)
- The change in frequency of each type (bar chart of % change between the two years)
- Arrest ratio: a chart showing the Arrest ratio for each Primary Type, for the two time periods
- Overview of location type: for each of the two annual periods, what is the overall distribution of Location Description

Part 3: Time Patterns of Crime: how is crime distributed throughout the day / week / month and year?

Are there times when all types of crime are high or low, are certain types of crime correlated by time, or do Primary Types have unique and different time characteristics?

- Analyze the dataset by Primary Type and Date to show their incidence over different time periods (daily, weekly, monthly) to discern if there are certain times when overall crime rates are abnormally high or low, and whether some crimes (e.g. theft and homicide) have similar time patterns.

- Select a couple of different time periods to see if patterns have changed over time (e.g. are more thefts or assaults occurring during the day?)

Part 4: Location Patterns of Crime: Are certain neighborhoods, Community Areas or Locations crime hotspots?

- Map the incidence of certain Primary types by location (using X,Y coordinates)
- Show the Location type broken down by Primary type
  - Maybe homicides are in buildings, while narcotics crimes occur on the street?
  - What crimes are most prevalent around schools?

Part 5: In depth analysis of specific issues: Given the overall trends of crime, and the patterns of time and location of the Primary Types, we select a couple of specific questions to answer in more depth. (The examples given may change after the initial data analysis)

- Is there a combination of time and location that appears to trigger or facilitate crime? e.g. car theft during the day at school?
- Abandoned Buildings at Night: Narcotics or sexual assault?
- Are some crimes predicted by the occurrence of others (taking different time lags), e.g. weapons violations and homicides?
- Arrest ratios:
  - Is there any correlation between either the absolute arrest ratio or trend in arrest ratio, and the trend in crime?
  - Are crimes with a low arrest ratio growing faster than crimes with a high arrest ratio?
  - If an arrest ratio has risen significantly for a Primary Type (or Primary Type within a Community Area) has there been a noticeable change in the level of that crime?

Proposal Appendix:

Initial Dataset Analysis

Fields	Sample Value	Questions Related
ID	10211392	
Case Number	HY397957	Count incidents
Date	8/24/2015 15:00	Part 3, 5
Block	075XX N PAULINA ST	Part 4
IUCR	890	
Primary Type	THEFT	All Parts
Description	FROM BUILDING	Maybe Part 5
Location Description	RESTAURANT	Part 4, 5
Arrest	FALSE	Part 2, 5
Domestic	FALSE	
Beat	2422	Maybe Part 5
District	24	Maybe Part 5
Ward	49	Maybe Part 5
Community Area	1	Part 4
FBI Code	6	
X Coordinate	1163690	Part 4
Y Coordinate	1950023	Part 4
Year	2015	Part 2, 3, 5
Updated On	8/31/2015 15:43	
Latitude	42.01849825	Maybe Part 5
Longitude	-87.67298323	Maybe Part 5
Location	(42.018498254, -87.672983233)	Maybe Part 4