Project Proposal
MSAI 339
Prof. Jennie Rogers
10/08/18

Jack Richard
Quincia Hu
Ikhlas Attarwala

Introduction

This proposal will identify specific queries and relations we are interested in learning about throughout the duration of this project. Our group will be augmenting the C.P.D.P. dataset with the Chicago Crime dataset provided by Chicago Data Portal. We will be paying close attention to the relationships between crimes committed per district and complaints filed against these districts' officers. Ultimately we would like to provide a heuristic that identifies significance in an allegation for projected chance of disciplinary action based on the officer's history. Extrapolating from this, we hope to also learn of correlations between the frequency of crimes in an area based on number of complaints, what crimes happen most often in high complaint-ridden areas, and maybe classify complaints on whether they are more or less emotionally-driven. This goal is heavily dependent on the ability to affirm a connection between crimes and complaints. We believe the following list of questions will help us answer this growing concern.

Description and Summarization

1. What is the salary range of officers with more than a single complaint? We're curious to see if the distribution of salary has any effect on predicting the type or number of complaints an officer may receive.
2. What is the distribution of complaints over the various police districts?
3. What officer ranks receive the most number of complaints?
4. How many different types of allegations are there against police officers?

Data Integration

1. Are there a significant number of complaints with near-equivalent entries of date and location in the crime database? This may help to computationally infer a set of coordinated crimes by an offender.
2. Given connections between crimes and complaints in a location at a given time, what classifications of crimes lead to a greater number of filed complaints?
3. Is there a stronger likelihood that a complaint will be filed in areas that historically show more domestic crimes? Here we are again interested in emotion as a factor of filing complaints.
4. What is the ratio of complaints to crimes per police district?

Workflow Analytics

1. Is there a specific district that consistently sees a higher percentage of complaints and crimes? If we are able to isolate this district, we may be able to focus further into issues on the individual-scale.

2. Is there a relationship between the area type and clusters of officers? ie. In a given area, is there a single officer that has racked up a plethora of complaints, or is there no significant relation in the area type?

3. What is the correlation coefficient of the relationship between police officer salary and the number of complaints received? What about the relationship between police officer rank and number of complaints received?

4. What is an officer's track record of complaints over a period of time, and which officers have sudden and a large number of complaints from an otherwise "quieter" behavior.

Machine Learning

1. How reasonable is a complaint, given the history of the officer and the area it was filed in? The reason for this prediction is that complaints can often be emotionally-driven, more so than objectively.

2. For a given district, can we predict the number of complaints during an upcoming time period? If a crime type's rate dramatically changes over a specific time period, it would be helpful for local residents to know in advance.

Modeling With Neural Networks

Our classifier will accept a document from the C.P.D.P and identify key components such as witness information, type of complaint, type of complainant, location of event, and whether the incident was domestic. Given this information, it will classify the document on a normalized scale of 0 to 1 to determine how likely the allegation will lead to disciplinary action.

Visualization

1. Historically what areas have the highest rate of crime, and what areas exhibit the most complaints?

2. How many unique officers have received complaints in each district or area of interest?

3. What factors have a higher probability of leading to complaints?

Database Schema

The Chicago Crime Dataset documents each crime recorded since 2001. The schema and clusters derived from related attributes are described below.

Table

Crime	
ID	integer
Case Number	integer
Date	datetime
Block	text
IUCR	integer
Primary Type	string
Description	text
Location Description	text
Arrest	boolean
Domestic	boolean
Beat	integer
District	integer
Ward	integer
Community Area	integer
FBI Code	string
X Coordinate	integer
Y Coordinate	integer
Updated On	datetime
Latitude	float
Longitude	float
Location	string
Add field	

Clustered Attributes

There are a number of attributes than can be removed due them not being applicable or being redundant. Ultimately, clusters can be defined surrounding the related attributes. The clusters can be seen below.

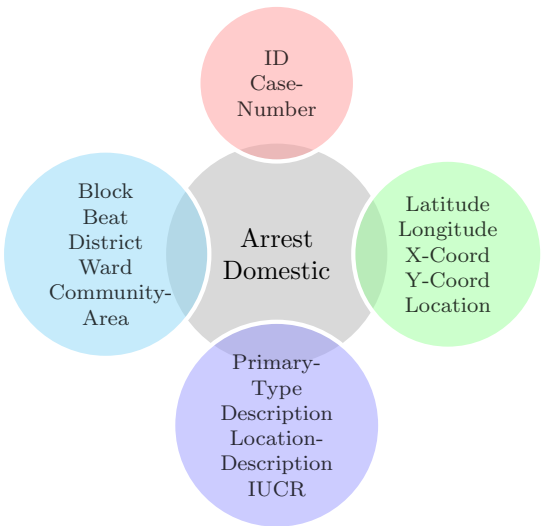

Table Samples

Below is the result of querying the first five rows of the Chicago Crime Data set. Due to the dataset being a single table with 21 attributes, the sample is split up for documentation

Attributes 1-7

ID	Case Number	Date	Block	IUCR	Primary Type	Description
634	G000705	01/01/2001 10:40:00 AM	024XX W MONROE ST	0110	HOMICIDE	FIRST DEGREE MURDER
635	F803142	01/01/2001 03:10:00 PM	023XX N MILWAUKEE AVE	0110	HOMICIDE	FIRST DEGREE MURDER
636	G008548	01/04/2001 10:30:00 PM	035XX W 12 PL	0110	HOMICIDE	FIRST DEGREE MURDER
637	G005960	01/06/2001 10:35:00 AM	014XX N HARDING ST	0110	HOMICIDE	FIRST DEGREE MURDER
638	G008793	01/05/2001 04:22:00 PM	016XX W 80 ST	0110	HOMICIDE	FIRST DEGREE MURDER
639	G011288	01/06/2001 08:54:00 AM	017XX S ASHLAND AVE	0110	HOMICIDE	FIRST DEGREE MURDER
640	G011894	01/06/2001 01:30:00 PM	117XX S MORGAN ST	0110	HOMICIDE	FIRST DEGREE MURDER
641	G013038	01/06/2001 11:38:00 PM	058XX W FULLERTON AVE	0110	HOMICIDE	FIRST DEGREE MURDER
642	G013102	01/07/2001 12:48:00 AM	007XX E 79 ST	0110	HOMICIDE	FIRST DEGREE MURDER
643	D362053	01/04/2001 10:05:00 PM	002XX E 46 ST	0110	HOMICIDE	FIRST DEGREE MURDER
644	G011117	01/09/2001 01:30:00 PM	087XX S BURLEY AVE	0110	HOMICIDE	FIRST DEGREE MURDER

Attributes 8-18

Location Description	Arres	Domestic	Beat	District	Ward	Comm	FBI Coc	X Coordinat	Y Coordinat	Year
CHA STAIRWELL			1125	011	2	28	01A	1159921	1899604	2001
AUTO			1414	014	35	22	01A	1156620	1915710	2001
BASEMENT	✓		1021	010			01A	1153021	1894062	2001
STREET	✓		2535	025			01A	0	0	2001
STREET			0611	006			01A	1166645	1851658	2001
AUTO			1222	012	25	31	01A	1165996	1891724	2001
ALLEY	✓		0524	005	34	53	01A	1171768	1827002	2001
STREET	✓		2515	025	37	19	01A	1137239	1915361	2001
STREET			0624	006			01A	1182418	1852784	2001
ABANDONED BUILDING			0221	002			01A	1178501	1874572	2001
STREET	✓		0424	004	7	46	01A	1199187	1847605	2001

Attributes 19-21

Latitude	Longitude	Location
41.880224549	-87.688248952	(41.880224549°, -87.688248952°)
41.924488265	-87.699933231	(41.924488265°, -87.699933231°)
41.865156224	-87.71373199	(41.865156224°, -87.71373199°)
36.619446395	-91.686565684	(36.619446395°, -91.686565684°)
41.748514115	-87.664928601	(41.748514115°, -87.664928601°)
41.858473695	-87.666167161	(41.858473695°, -87.666167161°)
41.680743671	-87.646876459	(41.680743671°, -87.646876459°)
41.923901331	-87.771156586	(41.923901331°, -87.771156586°)
41.751253211	-87.607096334	(41.751253211°, -87.607096334°)
41.81113151	-87.620788483	(41.81113151°, -87.620788483°)
41.736636821	-87.545821131	(41.736636821°, -87.545821131°)