Chicago Crime: A Predictive Analysis

Katie Mason
4/2017

Agenda

- Background on Chicago Crime
- Objectives and Success Criteria
- Dataset Overview
- Data Preparation
- Modeling & Evaluation
- Deployment
- Conclusion

Background

- Highest levels of violent crime in 20 years
- Homicide rate increased over 50%
- Gun shooting victims increased by 46%
- Potential causes: distrust of police force, low police morale, low arrest rates, gang violence and influx of firearms
- Violent crime: homicide, rape and sexual assault, robbery, aggravated assault and aggravated battery

Objectives

- When violent crimes occur using classification
- Where violent crimes occur using classification
- Profiles of crimes committed using clustering

Success Criteria

- Predict days and times violent crimes occur
- Identify places with highest crime rates
- Determine profiles of crimes

Chicago Crime Dataset

- 1/19/15 - 2/6/15, 3 weeks of data
- Target variable: Violent Crime - IUCR Codes, Primary Type
- Variables of interest:
  - District
  - DateTime
  - Arrest
  - Domestic

Data Preparation

DateTime: Derive Weekday and Time Intervals

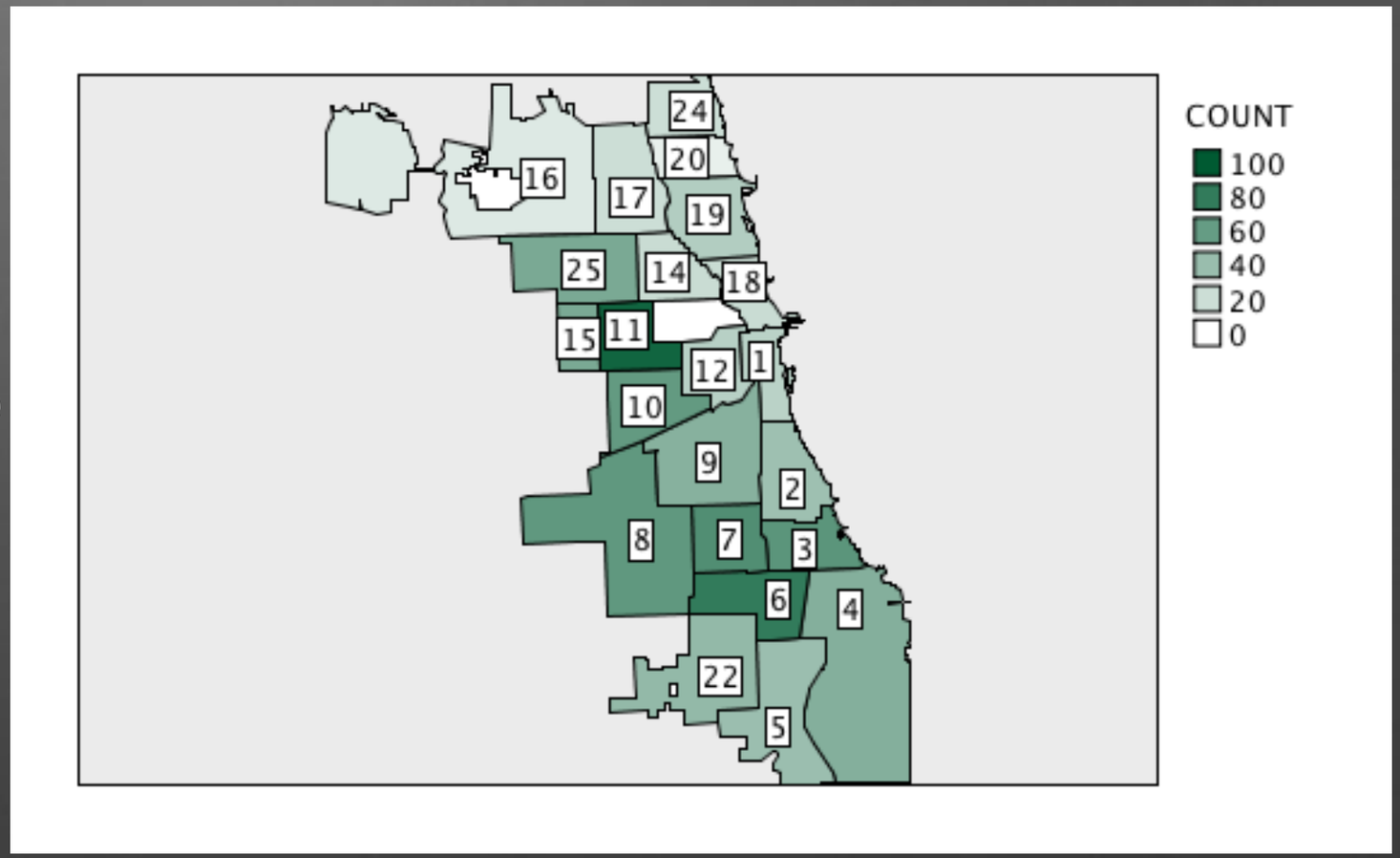
District: Missing data and formatting

Location Description: Reclassification for consolidation and missing values

IUCR: Classify violent from non-violent crimes

Distribution of Violent Crime

- An initial look at the distribution of violent crime reveals it is much more concentrated to the west and south of the city
- Districts 6 &11 have highest rates
- Districts 16 & 20 have the lowest rates



Feature Selection

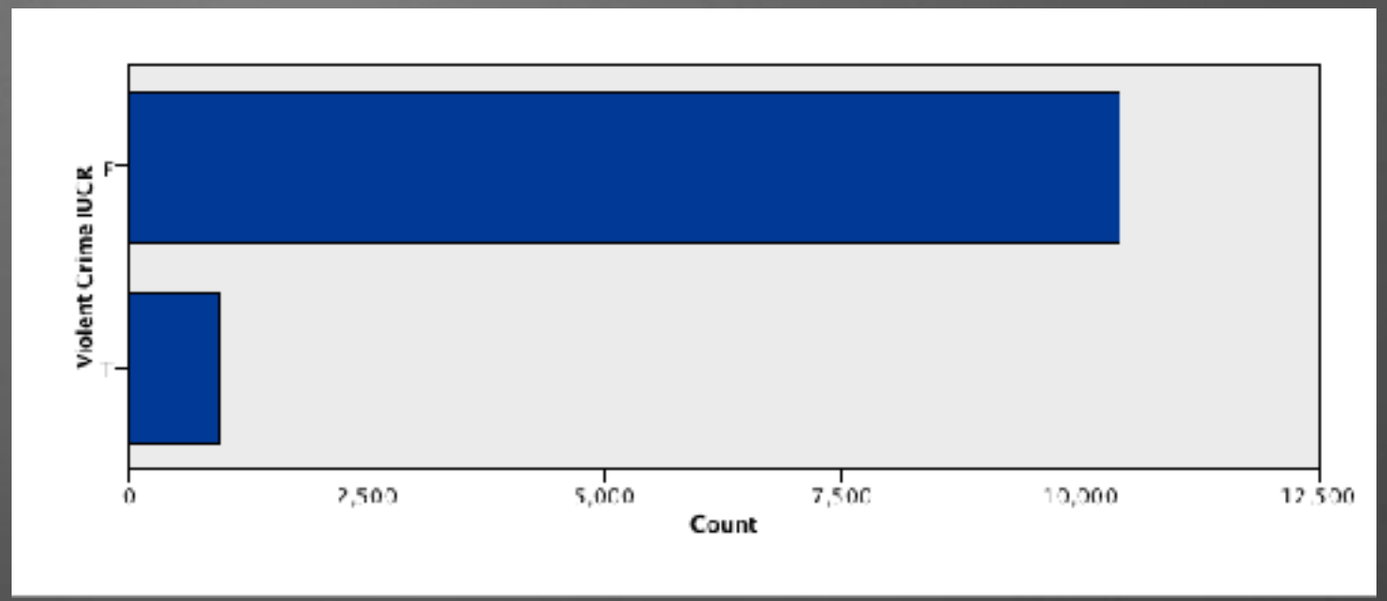
- Important fields:
  - Location Description
  - District
  - Arrest
  - Time Intervals
- Unimportant fields:
  - Weekday
  - Domestic

Anomaly Detection

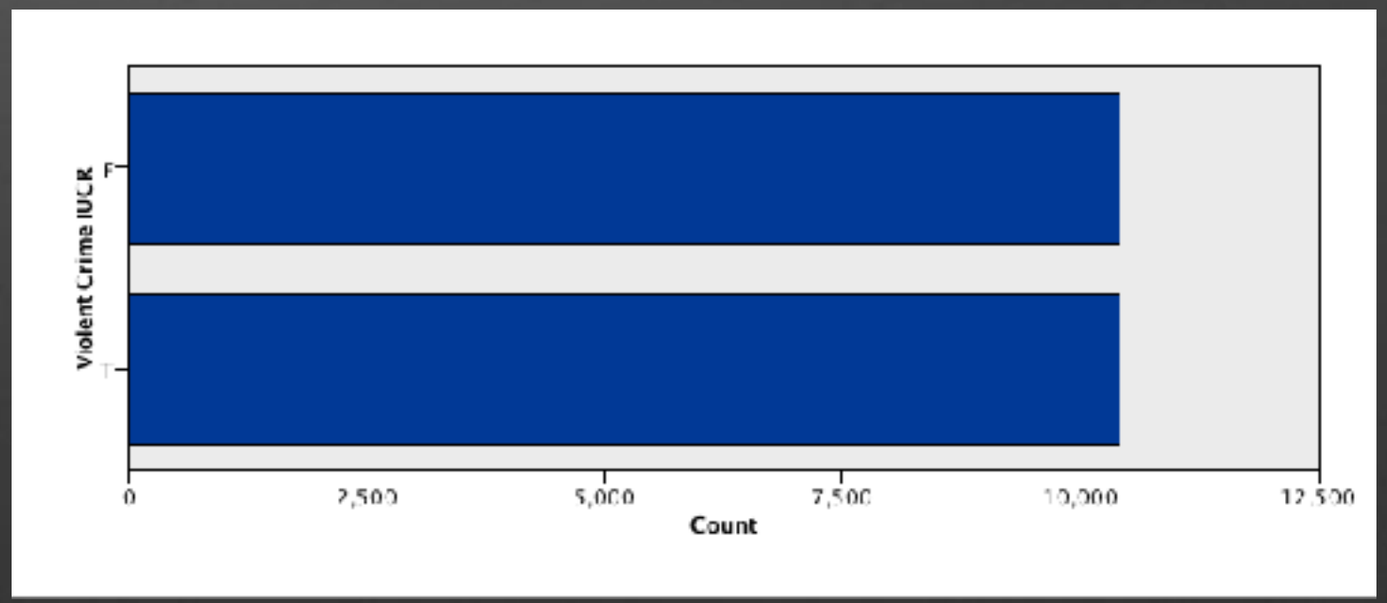
- 4 Peer Groups
- 20 Anomalous records
- Attributed mostly to missing Location Description values
- Excluding blank data reduced anomalous records to 9

Modeling: Boosting

- Target variable unbalanced
- Only 8% True

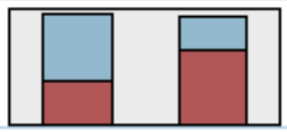
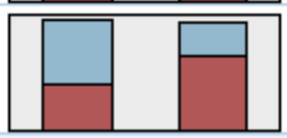


- After boosting
- 50% True



Modeling: Auto-Classifier

- Auto-classifier on decision tree models

Graph	Model	Build Time (mins)	Overall Accuracy (%)	No. Fields Used
	CHAID 1	< 1	65.997	4
	C&R Tree 1	< 1	64.614	5
	C5 1	< 1	64.033	5

- Inputs: Location Description, Arrest, Time Interval, District, & Anomaly

- Results:
  - CHAID: highest accuracy & true positives, minimal overfitting
  - C5.0: overfitted to the training data
  - C&R Tree: lower accuracy and overfit to training data

Evaluation: CHAID

- Important predictors: Arrest, District, & Location Description
- Still some overfitting:
  - Training accuracy: 71%
    Testing Accuracy: 65%
- Violent Crime Prediction:
  - Districts 22 & 3 at Apartment, Building, Private Space, or Vehicle
  - Chicago Housing Authority spaces and public spaces, except Districts 12 & 16
  - Districts 17, 18, 25 in late morning and early evening hours
  - Chicago Transit Authority spaces
  - Healthcare facilities and parking lots were not predicted to have violent crimes
  - Residences and stores in Districts 10 & 6
  - Districts 19, 2, and 7 in late night and early morning hours

Modeling: Clustering

- 3 Inputs: Arrest, Time Intervals, & Location Description
- All inputs ranked equally important predictors
- 4 Clusters with only fair cluster quality (0.4):
  - Arrests: 28%
  - Night Crimes: 40%
  - Midafternoon Thefts: 13%
  - Early afternoon Apartment Crimes: 18%

Evaluation: Clustering

- Clusters Characteristics:
  - Arrests: All arrests; 37% Narcotics, 11% Theft
  - Night Crimes: Crimes between 6 pm - noon, 12% violent
  - Midafternoon thefts: Crimes between 3-6 pm, 28% Thefts
  - Early afternoon apartment crimes: 51% at Apartments, 58% between noon and 3pm, 35% battery & assault

- Clusters variable only benefitted training accuracy
- Overfitting to training set by 10%

Deployment

- Increase police presence Chicago Housing Authority & Transit spaces
- Increase staffing on evening and night shifts
- Increase police presence around apartments in afternoon and keep in mind higher violence rates
- 18% of Thefts in Districts 1 & 18, target afternoon and early evening
- Drug-free campaign initiative to address narcotics

Conclusion

- Predictions for time and locations
- Weekday was not useful predictor
- 13% of violent crimes result in arrest
- Allows optimal cost-effective approach to reducing crime
- Limitations of the study
- Drug-free initiative

Appendix A - Violent Crimes

- Violent crimes were classified using IUCR codes and according to the definition by the City of Chicago, as specified at the Chicago ClearMap Crime Summary website: http://gis.chicagopolice.org/clearmap_crime_sums/crime_types.html#N03

- The following codes were classified as Violent Crimes:

- 01A Homicide
  0110, 0130

- 02 Criminal sexual assault
  0261, 0262, 0263, 0264, 0265, 0266, 0271, 0272, 0273, 0274, 0275, 0281, 0291, 1753, 1754

- 03 Robbery
  0312, 0313, 031A, 031B, 0320, 0325, 0326, 0330, 0331, 0334, 0337, 033A, 033B, 0340

- 04A Aggravated Assault
  051A, 051B, 0520, 0530, 0550, 0551, 0552, 0553, 0555, 0556, 0557, 0558

- 04B Aggravated Battery
  041A, 041B, 0420, 0430, 0450, 0451, 0452, 0453, 0461, 0462, 0479, 0480, 0481, 0482, 0483, 0485, 0488, 0489, 0490, 0491, 0492, 0493, 0495, 0496, 0497, 0498, 0510

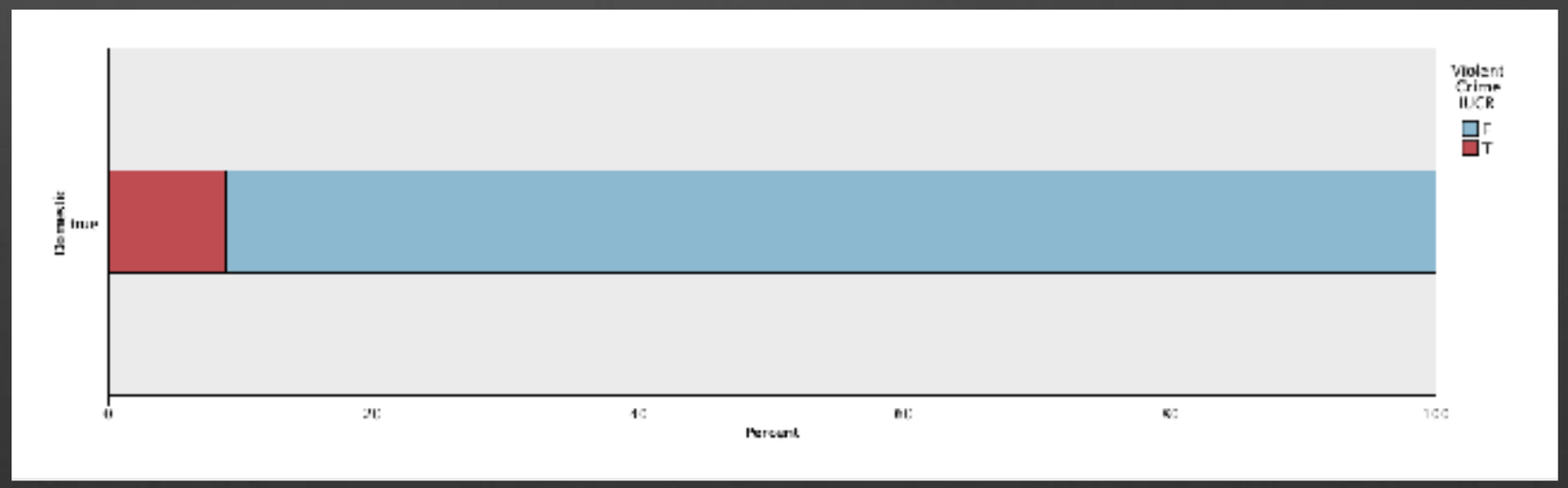
- All other crimes were classified as non-violent.

Appendix B - Domestic Violence

- Although domestic violence is included as a variable in the dataset, its definition is based on the Illinois Domestic Violence Act and did not completely coincide with the definition of a violent crime as determined by the IUCR codes (4,5). Only 9% of crimes classified as domestic violence met the criteria of violent crime, as defined by IUCR. Due to the delicate nature of domestic violence and its special impact on families and communities, it deserves a full assessment which is outside the scope of this analysis.

- **Violent Crimes**
  Count = 147
  Percent = 8.914

- Non-violent Crimes
  Count = 1502
  Percent = 91.086



References

1. Gorner, J. (2017, March 03). Few answers as Chicago hit with worst violence in nearly 20 years. Retrieved April 23, 2017, from http://www.chicagotribune.com/news/local/breaking/ct-chicago-violence-2016-met-20161229-story.html

2. Sanburn, J., & Johnson, D. (2017, January 30). Violent Crime Is On the Rise in U.S. Cities. Retrieved April 23, 2017, from http://time.com/4651122/homicides-increase-cities-2016/

3. Chicago Police GIS. Retrieved April 23, 2017, from http://gis.chicagopolice.org/clearmap_crime_sums/crime_types.html#N04B

4. Illinois State Agency Links. (n.d.). Retrieved April 23, 2017, from http://www.isp.state.il.us/crime/domesticviol.cfm

5. Chicago Data Portal. Retrieved April 23, 2017, from https://data.cityofchicago.org/view/5cd6-ry5g