Description

Our dataset is a relational table of every reported instance of crime in Chicago from 1/1/2001 to 1/17/2018 provided by the Chicago Police Department. With this information we intend on mining correlations between the datetime, location, and types of crime. For example, based on the concentration of drug-related crime we can predict where drug dealers might be located.

Datasets

Chicago Police Department CLEAR (Citizen Law Enforcement Analysis and Reporting) Reported Incidents of Crime from 2001 Present.

Found at:

https://catalog.data.gov/dataset/crimes-2001-to-present-398a4

(On Friedrich Amouzou's Machine)

Questions to answer

- How do types and amount of crime change throughout the city?
- Which neighborhoods can we classify as generally unsafe?
- Can we predict future specific crime trends in certain neighborhoods?

Data Preparation Work

- Data Cleaning
  - Eliminating incomplete reports / handling null data
- Data Preprocessing
  - Organize data by type
  - FBI codes are stored as floats
  - Date/time standardization
  - Reducing the dataset/ removing unnecessary attributes
- Data Integration
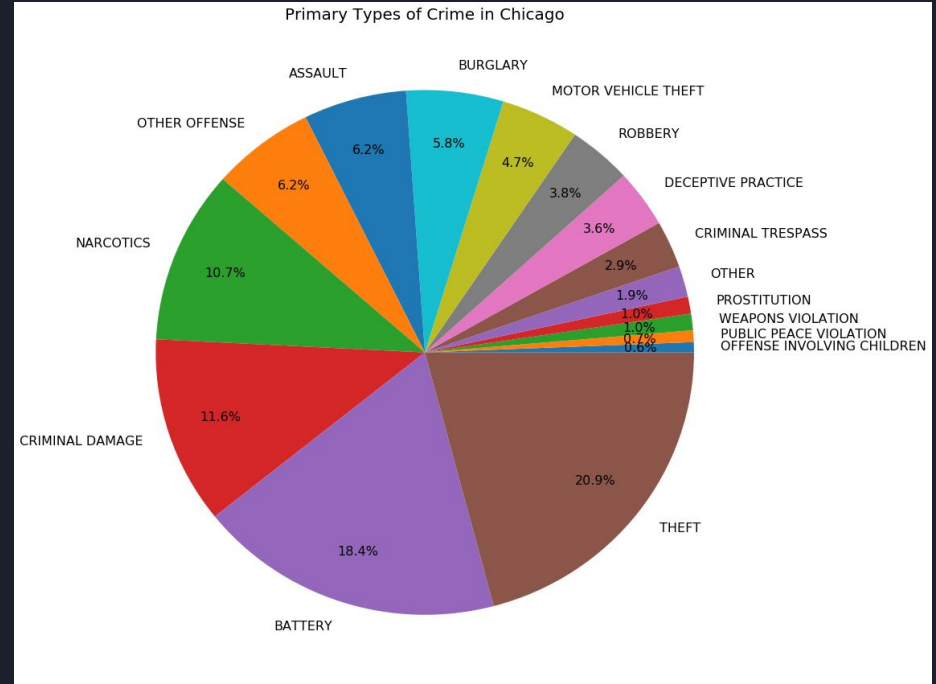 - Data visualization
  - Trend analysis
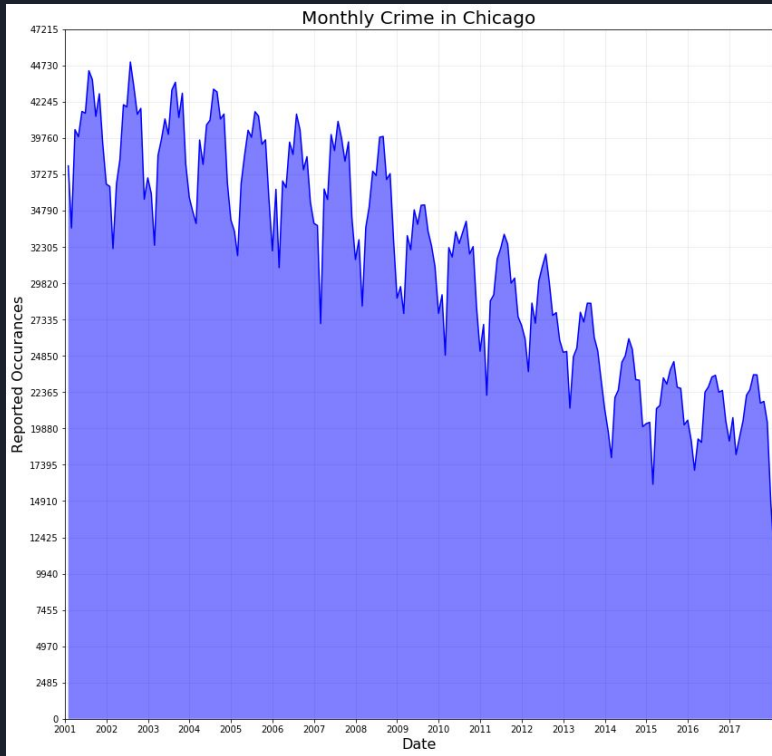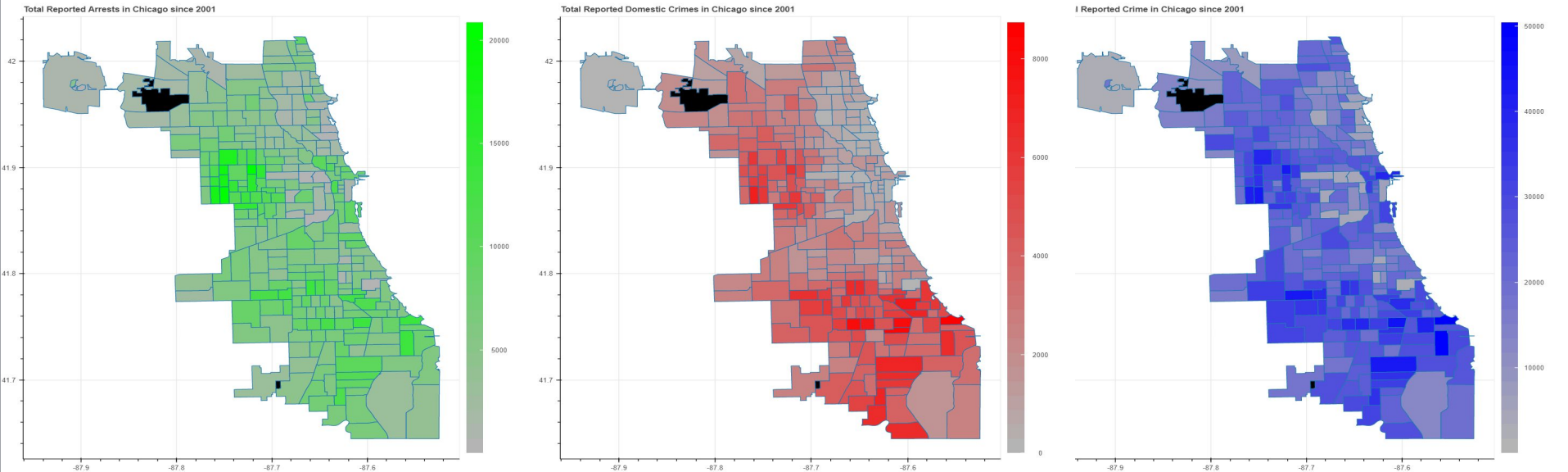
Tools

- Python (including utilization of libraries like pandas, numpy, scipy, etc.)
- Jupyter Notebooks (for visualization)
- R (if necessary)

Summary of Results

Monthly Crime in Chicago

Primary Types of Crime in Chicago

Summary of Results

Total Reported Arrests in Chicago since 2001

Total Reported Domestic Crimes in Chicago since 2001

I Reported Crime in Chicago since 2001

Predictions

	Logistic Regression	Logistic +PCA	Logistic +LDA
Accuracy	0.8900	0.8886	0.8906
Sensitivity	0.5822	0.5850	0.5952
False Positive Rate	0.0261	0.0286	0.0289
Specificity	0.9739	0.9714	0.9711

Predictions (cont.)

	Linear SVM + PCA	Linear SVM + LDA	Radial SVM + PCA	Radial SVM + LDA
Accuracy	0.8661	0.8916	0.8915	0.8920
Sensitivity	0.5426	0.5705	0.5700	0.5686
False Positive Rate	0.0329	0.0209	0.0207	0.0199
Specificity	0.9761	0.9791	0.9793	0.9801

Applications of results

Current:

Visualizations can be used to form a general idea of crime and arrests for individual police beats. Clearly show views of the relative violence throughout the city.

Future:

Police beats can use data to redraw beat districts to better apply police to areas that increase in crime over time. Our predictions will show crime trends that we expect which will help to form a fully comprehensive view of crime.