Chicago’s Golden Future

Gold Team
Osama Ahmad, Taj Atassi, Mohamed Saeed, Abdullah Kidwai, Maxwell Dausch

Repository:

https://github.com/orgs/edgeslab/teams/gold-team-rules

Name	NetID	GitHub Handle
Osama Ahmad	oahmad7	oahmad7
Maxwell Dausch	mdausc2	mdausch
Taj Atassi	tatass2	tajatassi
Mohamed Saeed	msaeed7	MohamedSaeed99
Abdullah Kidwai	akidwa2	akidwhy

Problem: Objectives

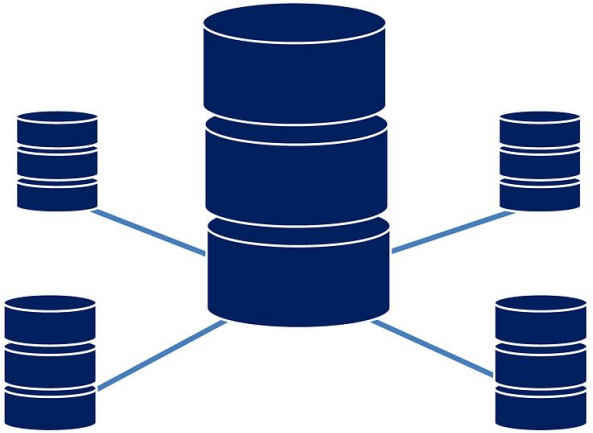
- Correlation between Chicago’s historical crime data and permit data
- Find investment opportunities when examined with Chicago’s government owned land
- Leads to return on investment when made at low prices

Problem: Expected Outcome

- New buildings and facilities -> investment increasing
- Decreasing crime rate, increasing investment rate = good future investment
- Increasing crime rate = poor investment

Data: Our Various Data Sets

- Crime Data
- Permit Data
- Land Data

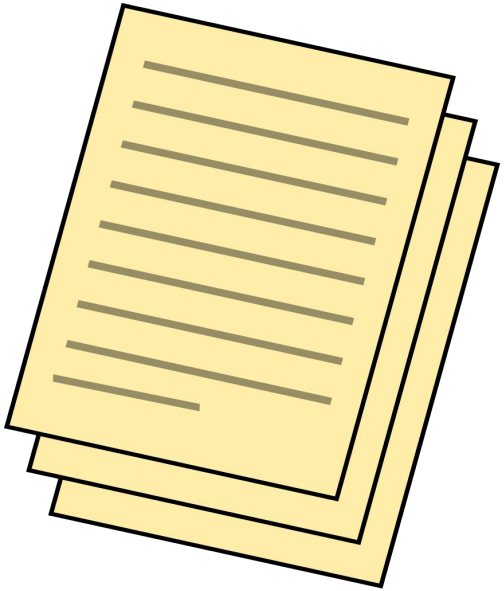


Data: Crime Dataset

- 6,805,668 rows
- 2001 - Present
- Location
- Crime Type
- Date of Crime
- Aids in finding:
  - Crime rate changes
  - Concentration of crime

Data: Permit Dataset

- 564,828 rows
- 2006 - Present
- Location
- Dates of permit application
- Aids in finding:
  - Types of renovations
  - Where investments are made



Data: Land Dataset

- 15,932 rows
- Location
- Zoning Information
- Square Footage
- Aids in finding:
    - Government owned land for sale
    - Areas for cheap long term investments

Data: Building the Databases

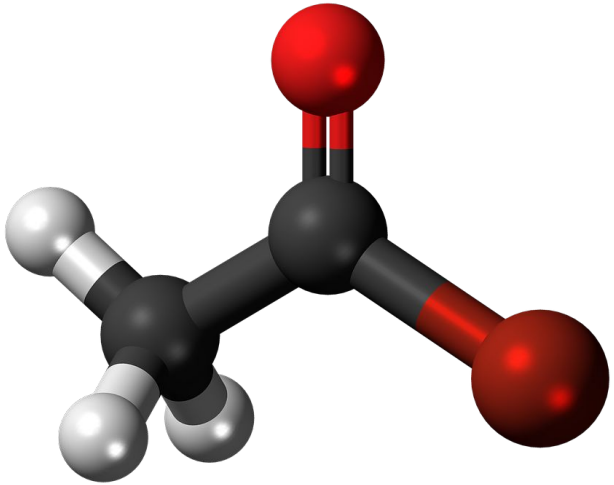
- Process a CSV into a SQLite3
    - Needed to bypass low memory errors
- Filter out the columns we don’t need

Data: Cleaning the Data

- Removed duplicated data
- Removed useless data (ex: 20 columns of contractor information)
- Removed dates outside scope of project
- Removed empty columns and null values

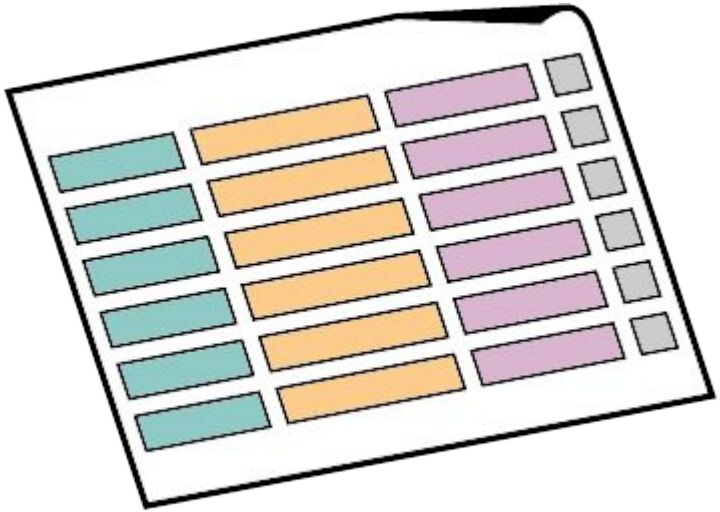
Exploratory Data Analysis: Structure

- Data is not very rectangular
- Some entries are missing
- No nested data (easy to manipulate)



Exploratory Data Analysis: Granularity

- Data is very fine
- Which Specific Buildings are available
- Many details about crimes



Exploratory Data Analysis: Scope

- Data is mostly complete
- Some entries are blank
- Only needed to focus 2006 - 2019
    - Machine Learning: 2018 only

Exploratory Data Analysis: Temporal

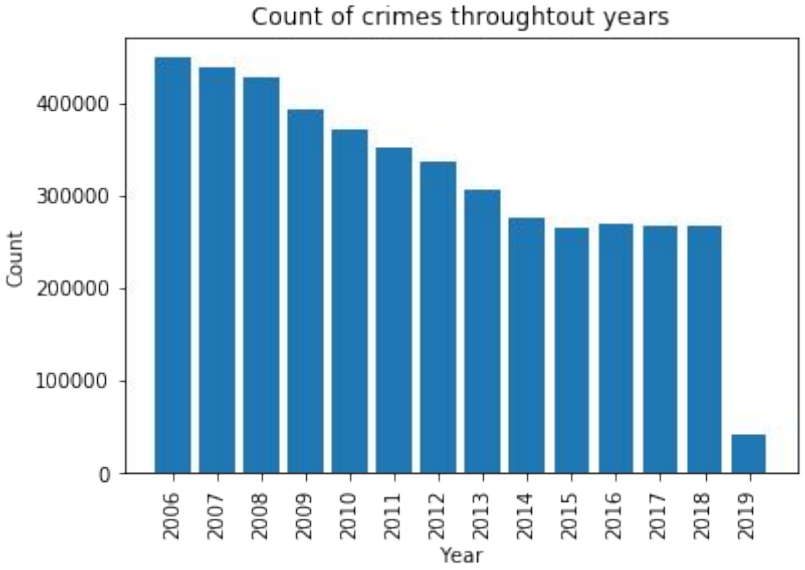
- Before cleaning -> different ranges for each database
- After cleaning -> data truncated
  - Crime Data: 2001-2019 -> 2006-2019
  - Permit Data: 2006-2019 -> 2006-2019
  - Land Data: Up-to-Date

Exploratory Data Analysis: Faithfulness

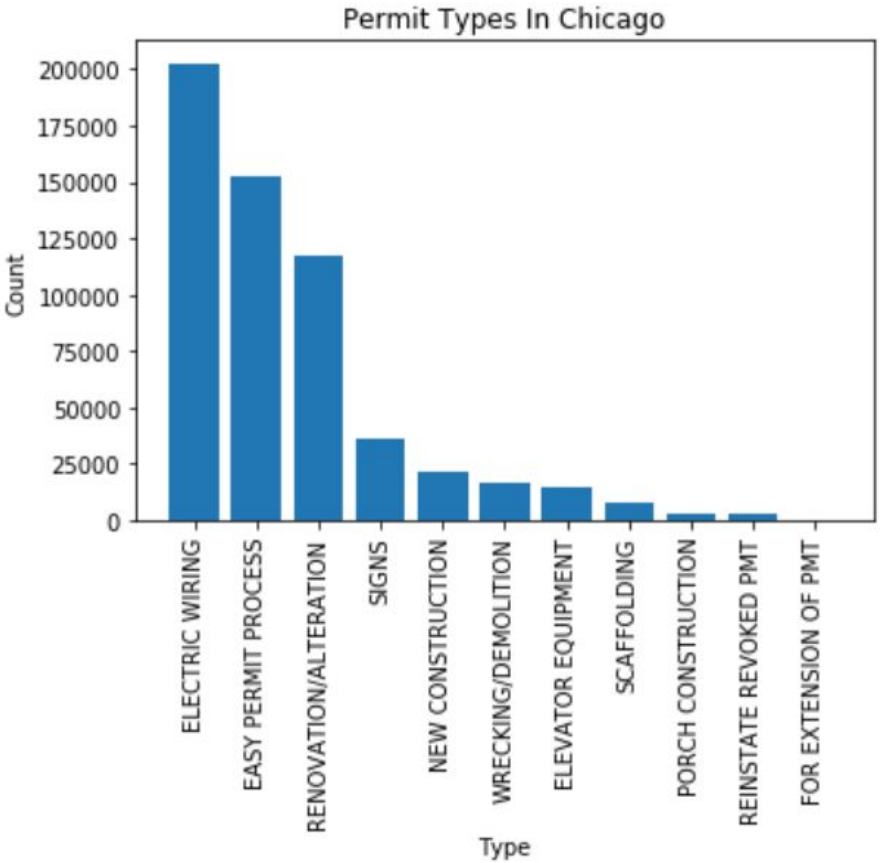
- Data taken directly from City of Chicago
- Can be Assumed to be trustworthy

Exploratory Data Analysis: Other Observations

- After cleaning data, dates range from 2006-2019
- fetchall() functions return a list of tuples always
- Need to extract proper items and place them into lists

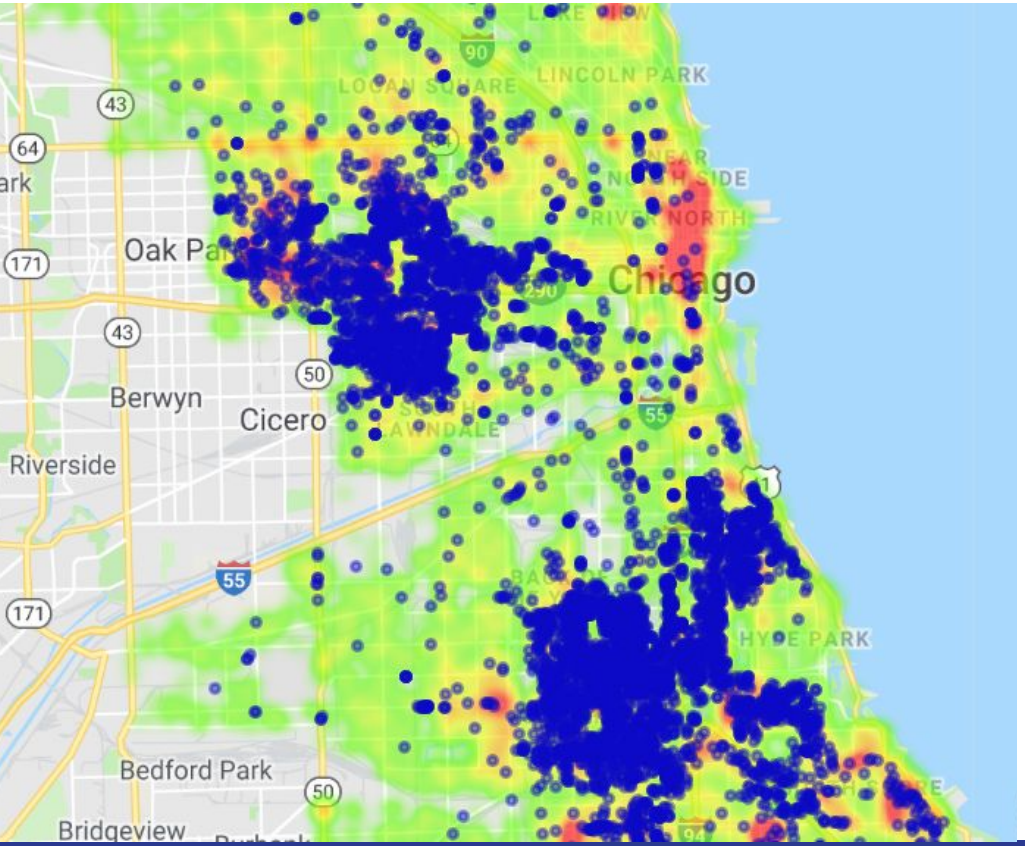


- Number of crimes from 2006 to present
- Decrease in overall crime
- Crime rate in past four years has stagnated



- Permit types granted in Chicago
- Many permits are tied to renovation
    - Can be seen as high investment

Heatmap of Chicago Crimes and City Land



- The heat map corresponds to the overall crime concentration
- Each blue dot is a piece of government owned land

60633

- ELECTRIC WIRING (64.2 %)
- EASY PERMIT PROCESS (24.0 %)
- RENOVATION/ALTERATION (6.3 %)
- WRECKING/DEMOLITION (3.0 %)
- NEW CONSTRUCTION (1.1 %)
- SIGNS (0.7 %)
- ELEVATOR EQUIPMENT (0.7 %)

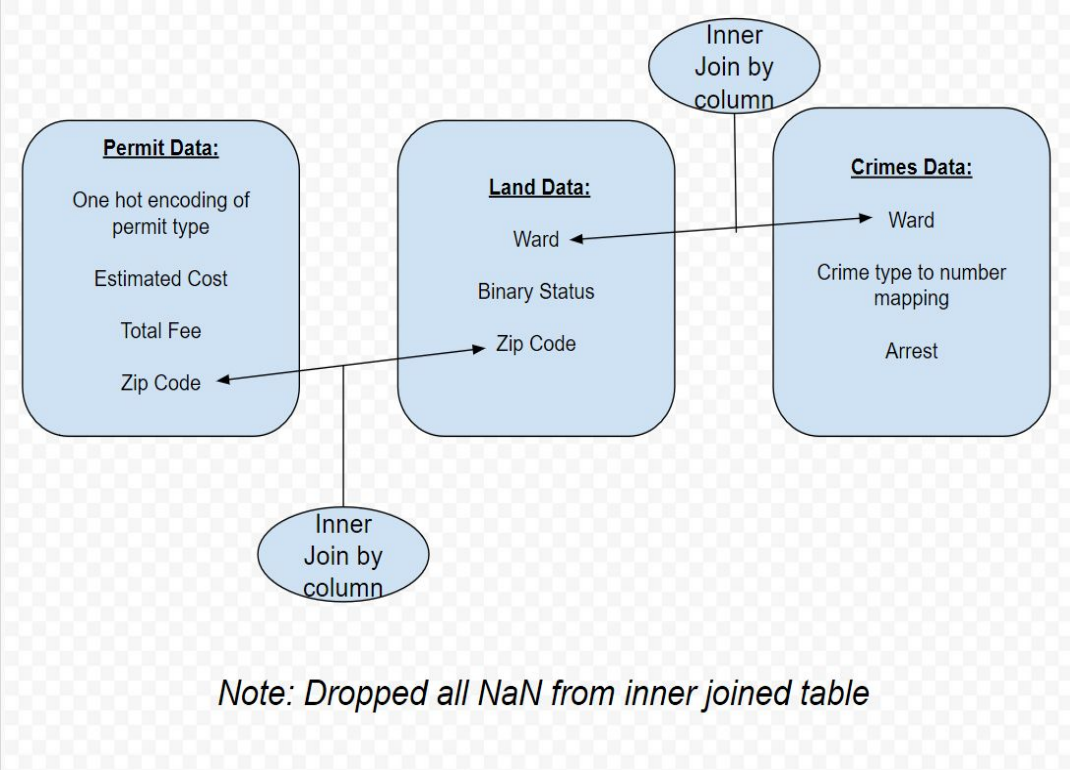
60607

- ELECTRIC WIRING (30.4 %)
- RENOVATION/ALTERATION (23.2 %)
- EASY PERMIT PROCESS (21.3 %)
- SIGNS (11.5 %)
- ELEVATOR EQUIPMENT (5.5 %)
- NEW CONSTRUCTION (5.0 %)
- WRECKING/DEMOLITION (2.1 %)
- SCAFFOLDING (0.9 %)
- REINSTATE REVOKED PMT (0.2 %)

60631

- ELECTRIC WIRING (39.5 %)
- EASY PERMIT PROCESS (27.9 %)
- RENOVATION/ALTERATION (19.7 %)
- SIGNS (5.1 %)
- NEW CONSTRUCTION (3.0 %)
- WRECKING/DEMOLITION (1.8 %)
- ELEVATOR EQUIPMENT (1.4 %)
- REINSTATE REVOKED PMT (0.9 %)
- SCAFFOLDING (0.7 %)

Machine Learning Data



Goal:

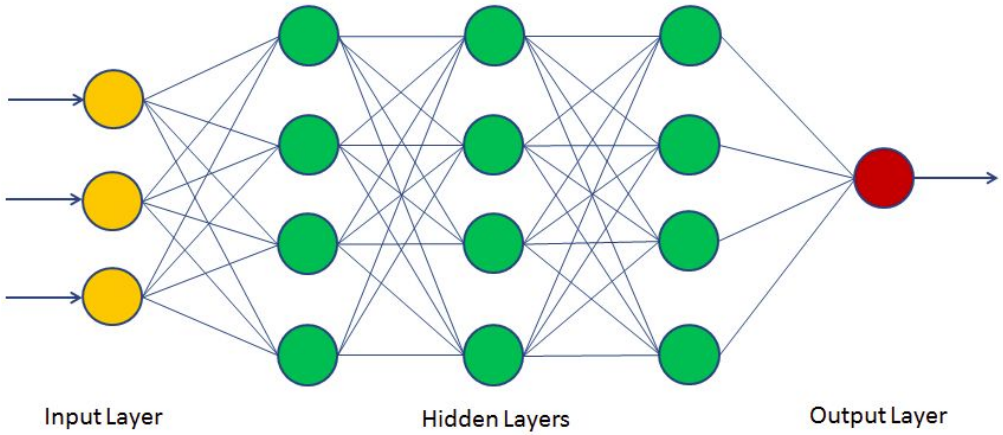
Predict Chicago's City

Owned Land status (Sold,

Not Sold)

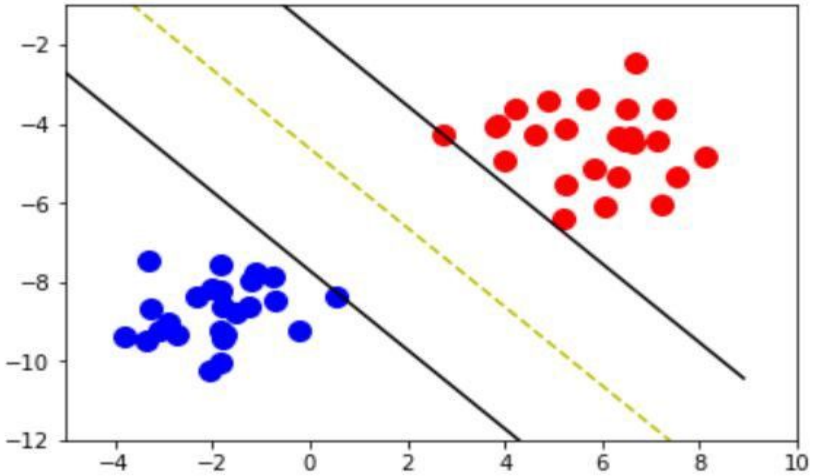
Machine Learning: Neural Network

- Used 10 fold cross validation.
- Max_iter: 10000, Activation: relu, solver: adam, epsilon:0.001, alpha: 0
- Found 1 layer 10 nodes to be optimal



Machine Learning: Support Vector Machine

- Used bagging to speed process
- Less error than the Neural Network



Takeaways:

- High overlap between high crime areas and government owned land
- Much more rehabilitation than building new facilities
- Crime rate is decreasing year over year
- Small amount of prime candidates for investment are available

Next Steps:

- Use scraping sites (Zillow or Realtor) to find accurate representation of housing data
- More Credit in Google API
- Find a way to lower error of Neural Network
- Run ML Models on complete Dataset
- Predict for each neighborhood instead of overall