Predicting Fire Risk in Atlanta

Data Science for Social Good – Atlanta Fire Rescue Department

Team: Xiang Cheng, Oliver Haimson, Michael Madaio, Wenwen Zhang

Advisors: Dr. Polo Chau, Dr. Bistra Dilkina

Partner: Atlanta Fire Rescue Department
Dr. Matt Hinds-Aldrich (AFRD)

+ Data Science for Social Good & Atlanta Fire Rescue Department

Team Members:

- Oliver Haimson | UC Irvine | ohaimson@uci.edu
- Michael Madaio | Georgia Tech | mmadaio@gatech.edu
- Xiang Cheng | Emory University | xcheng7@emory.edu
- Wenwen Zhang | Georgia Tech | wzhang300@gatech.edu

Partner:

- Atlanta Fire Rescue Department (AFRD)
- Dr. Matt Hinds-Aldrich (AFRD) | mhinds-aldrich@atlantaga.gov

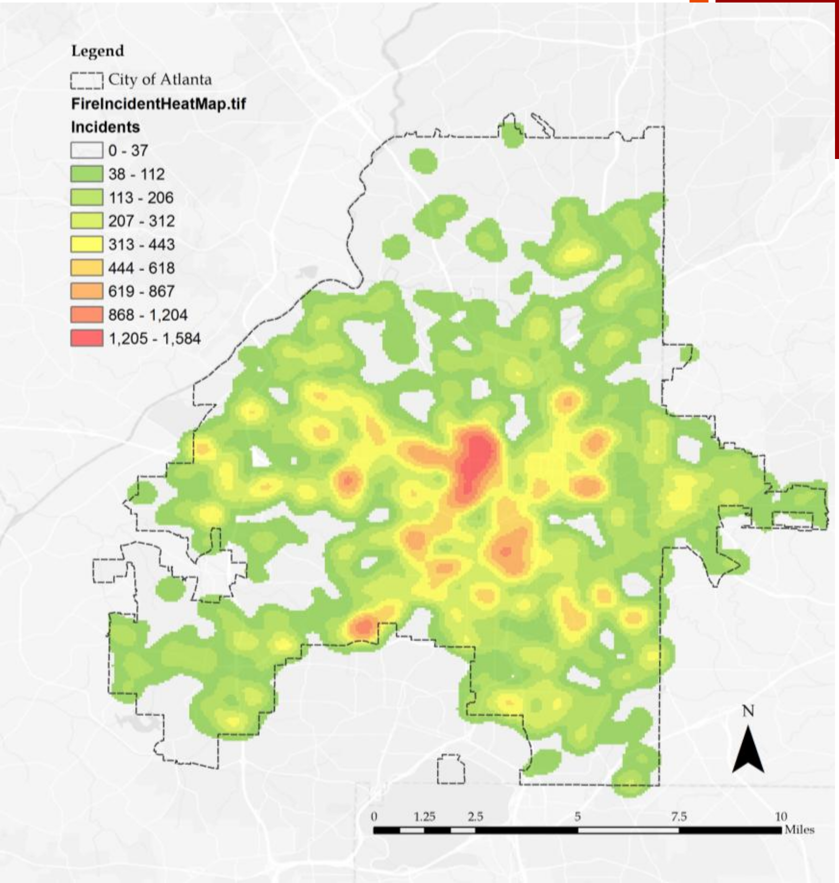
Mentors:

- Dr. Polo Chau | Georgia Tech | polo@gatech.edu
- Dr. Bistra Dilkina | Georgia Tech | bdilkina@cc.gatech.edu

+Problem

- Hundreds of fires occur in Atlanta every year
- 2,600 properties are inspected per year
- How do we help AFRD find new commercial properties that need inspection?
- How do we ensure the properties at greatest risk of fire are being inspected?

Fire incidents heat map (2011-present)



Goal 1: Find new properties to inspect

- List of new properties: from external business and property databases
- Prioritized list: using risk scores from the model
- Interactive map to view inspected properties, fire incidents, and potential inspections in Atlanta

Goal 2: Prioritize inspections

- Integrated database of buildings with the most complete property information
- Make a predictive model to generate risk score for properties

+ Data

- 6+ sources
- 2+ GB
- ~200,000 Records

Data	Source
Fire Incident	Atlanta Fire Department
Fire Inspection Permits	
Liquor License	
Parcel Data	City of Atlanta
Atlanta Business Licenses	
SCI Report	
Neighborhood Planning Unit	Atlanta Regional Commission
Demographic Data	U.S. Census Bureau
Socio-economic Data	
CoStar Property Report	CoStar Group, Inc
Business Location Data	Google APIs

How do we help AFRD find new properties that need inspection?

+ Finding potential inspections

Current Inspections

2,600

Business Licenses
20,000

Google places

10,000

+ Finding potential inspections

Current Inspections

2,600

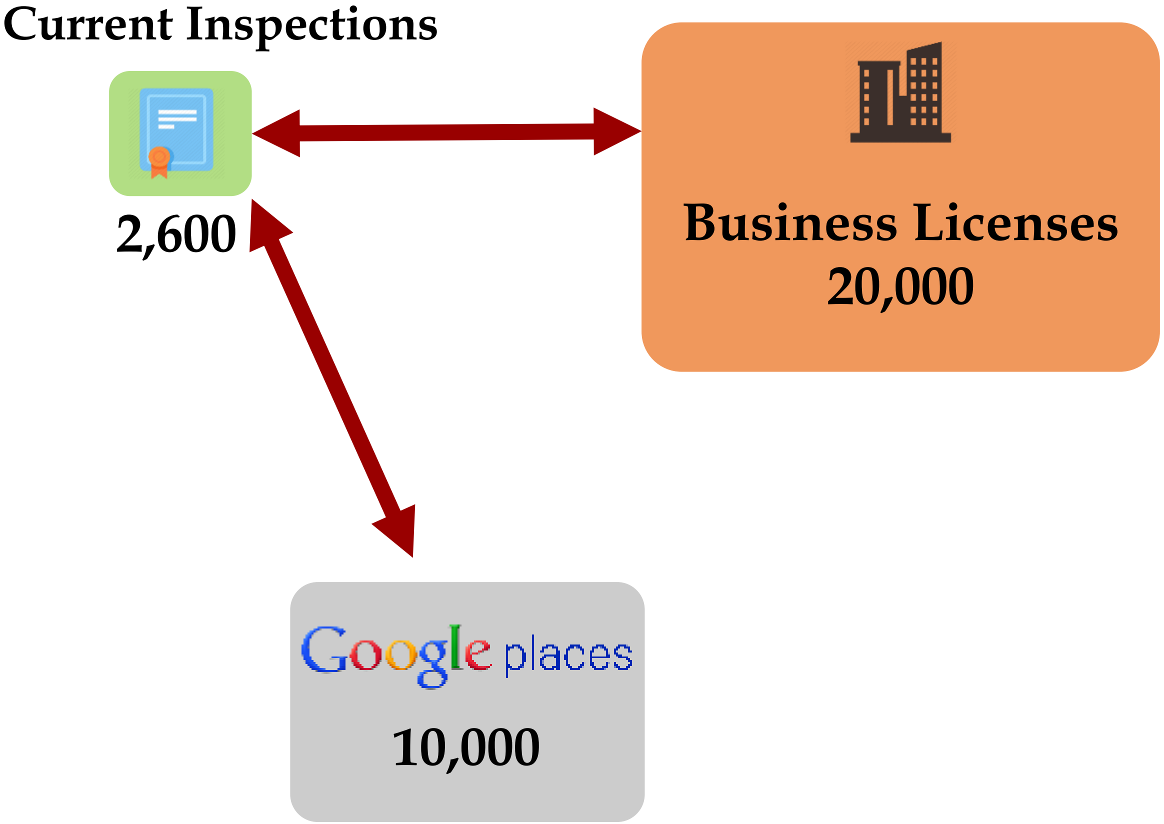
Business Licenses
20,000

- Find Property Types:
  - Currently inspected types

Google places

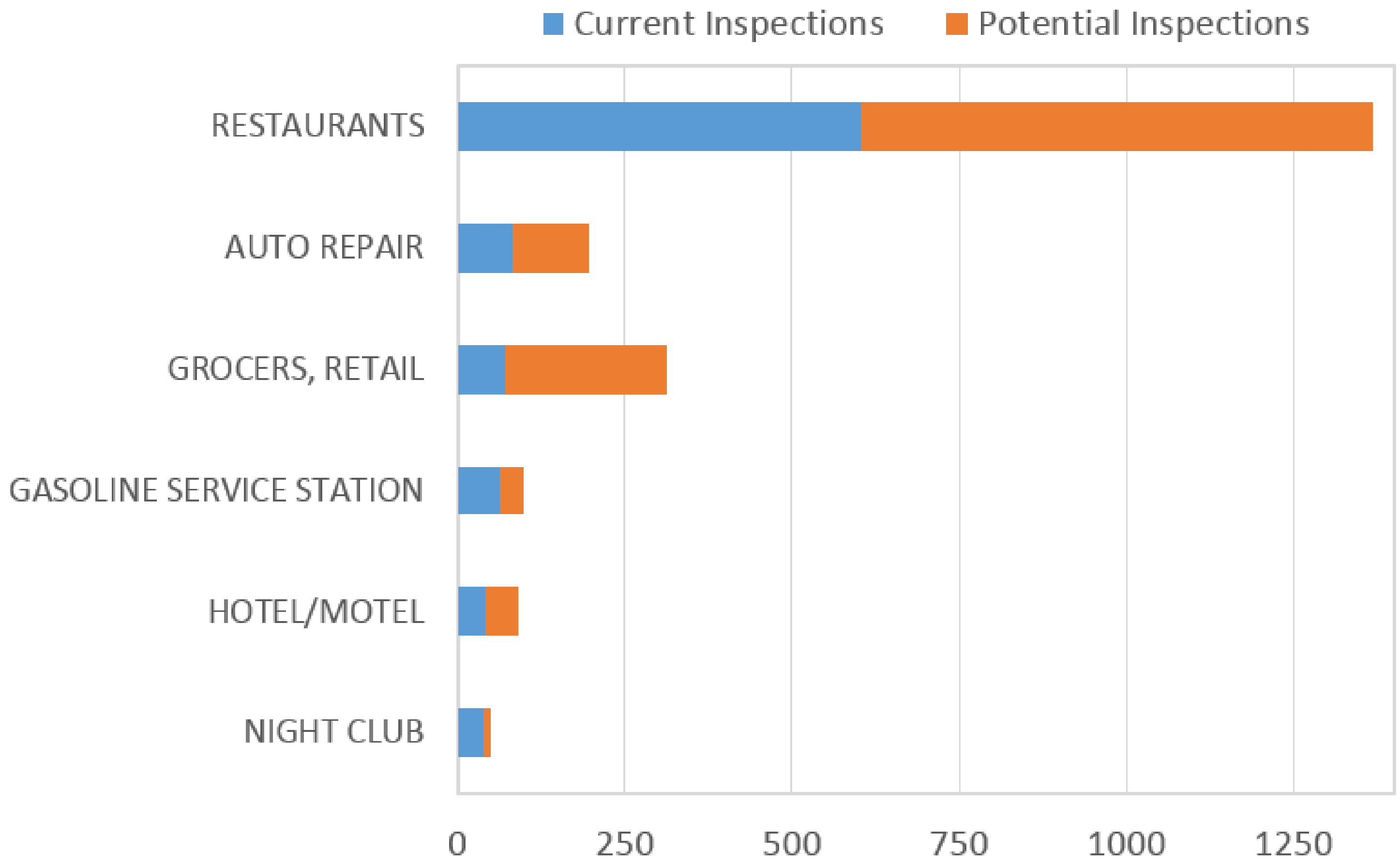
10,000

+ Finding potential inspections

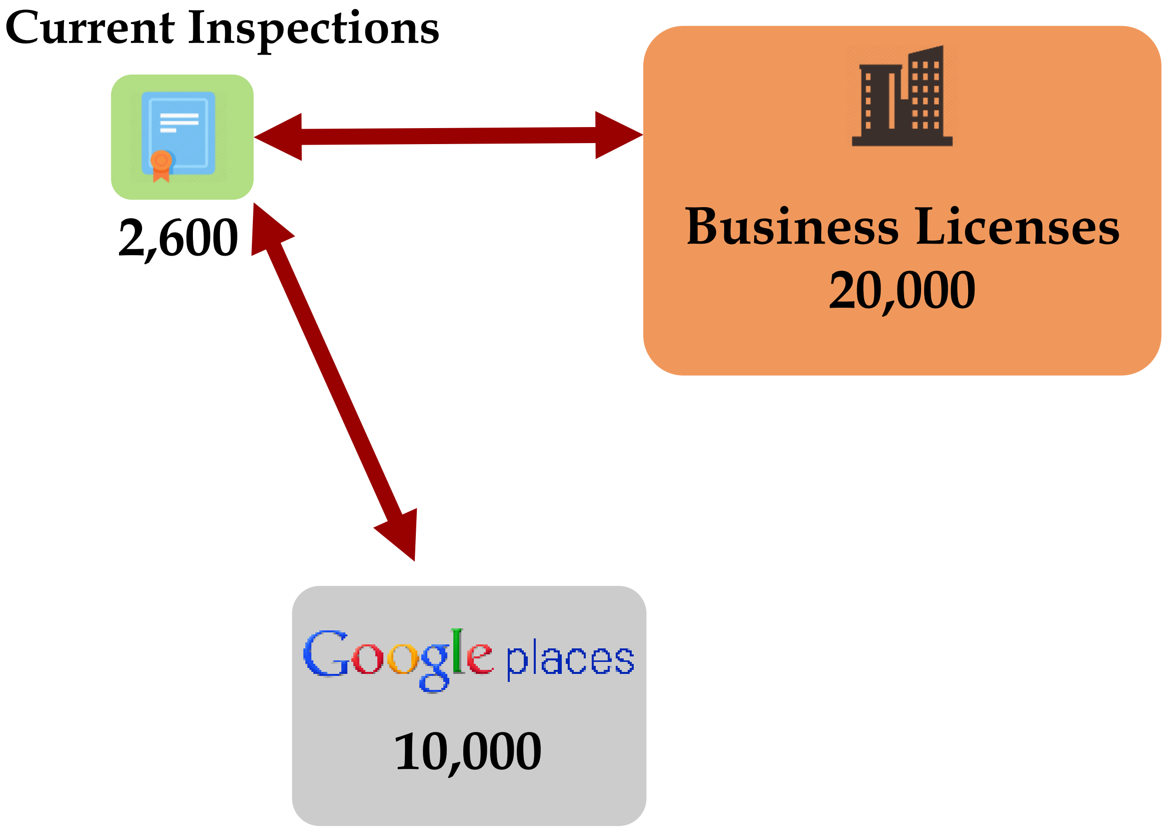


- Find Property Types:
  - Currently inspected types
    - Geocoding
    - Fuzzy text-matching

+ Finding potential inspections

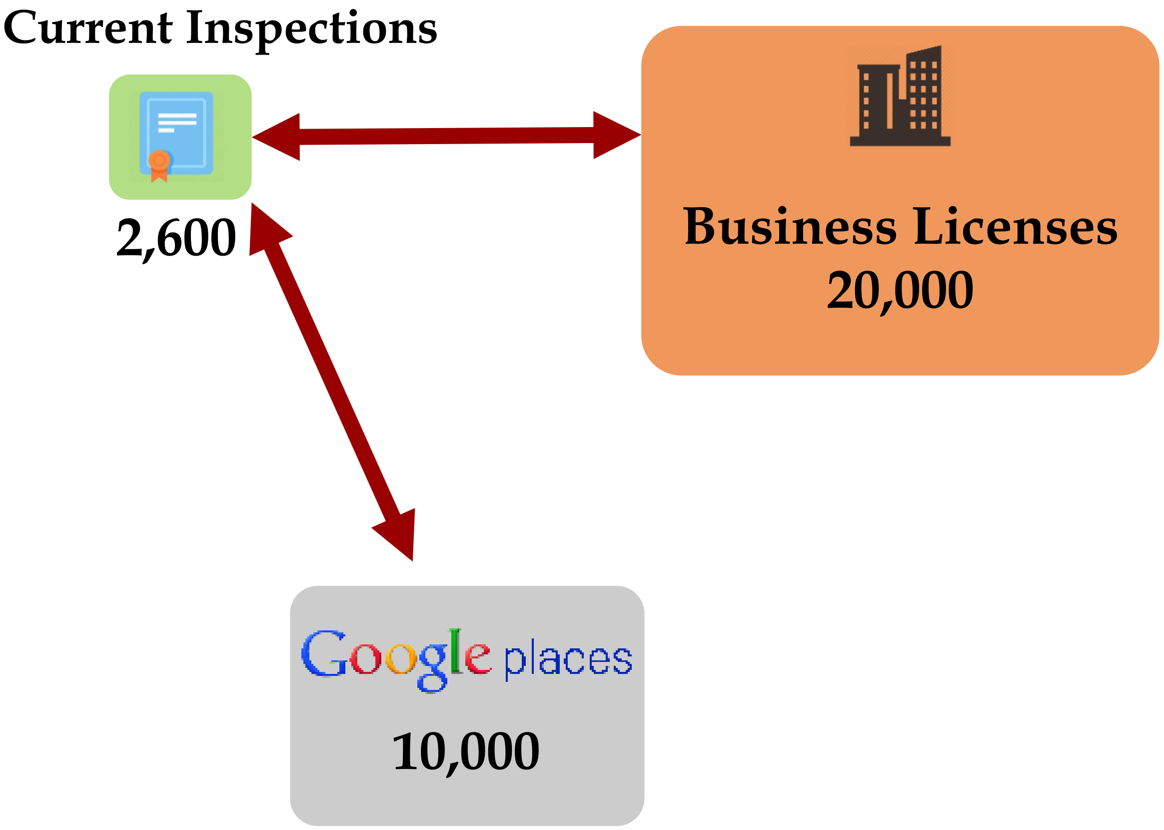


+ Finding potential inspections



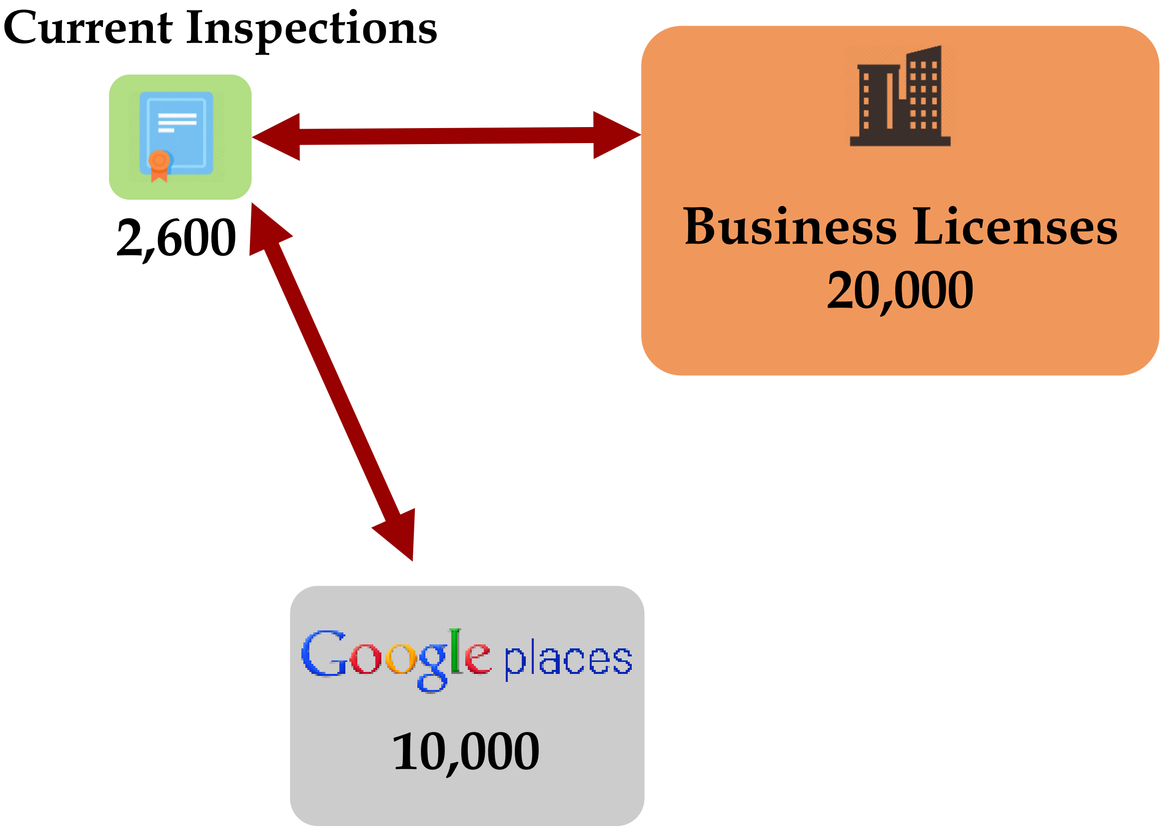
- Find Property Types:
  - Currently inspected types
    - Geocoding
    - Fuzzy text-matching

+ Finding potential inspections



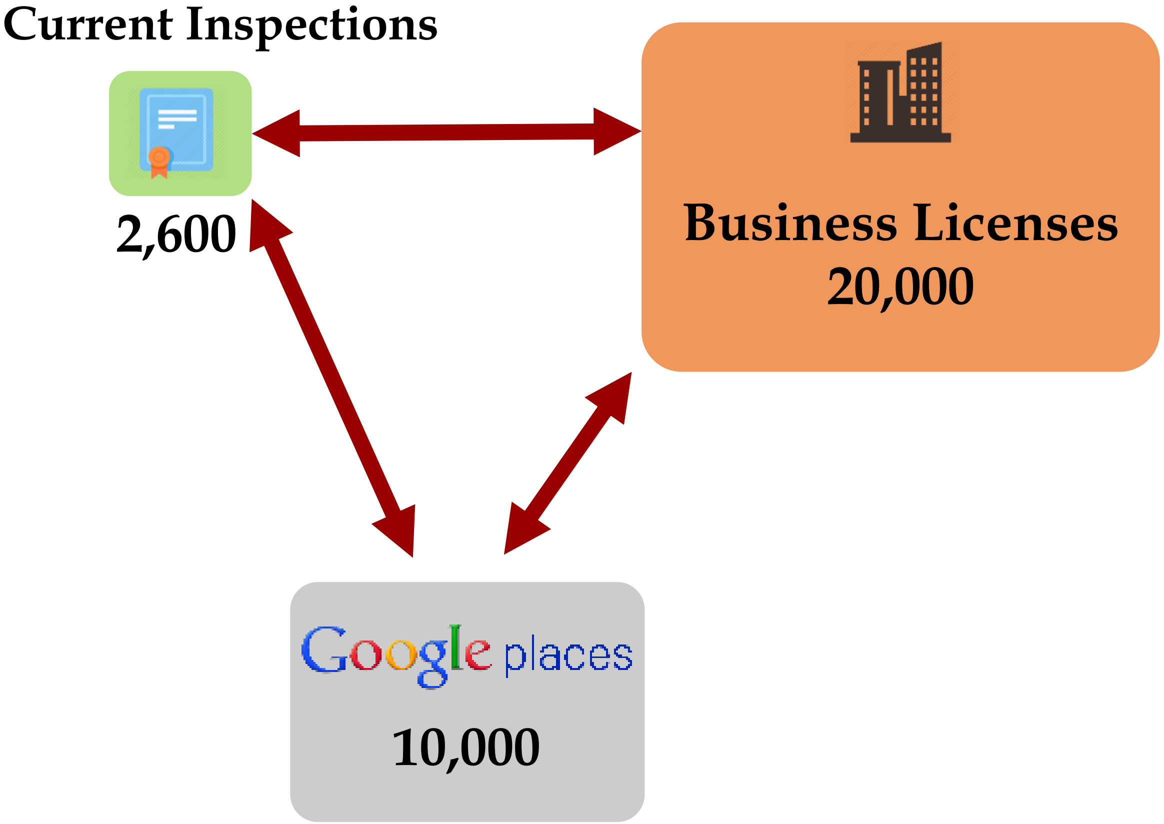
- Find Property Types:
  - Currently inspected types
    - Geocoding
    - Fuzzy text-matching
  - Text-mining of the *Fire Code of Ordinances*
  - Fire inspectors focus group

+ Finding potential inspections



- Find Property Types:
  - Currently inspected types
    - Geocoding
    - Fuzzy text-matching
  - Text-mining of the *Fire Code of Ordinances*
  - Fire inspectors focus group
- Generate unique property list

+ Finding potential inspections

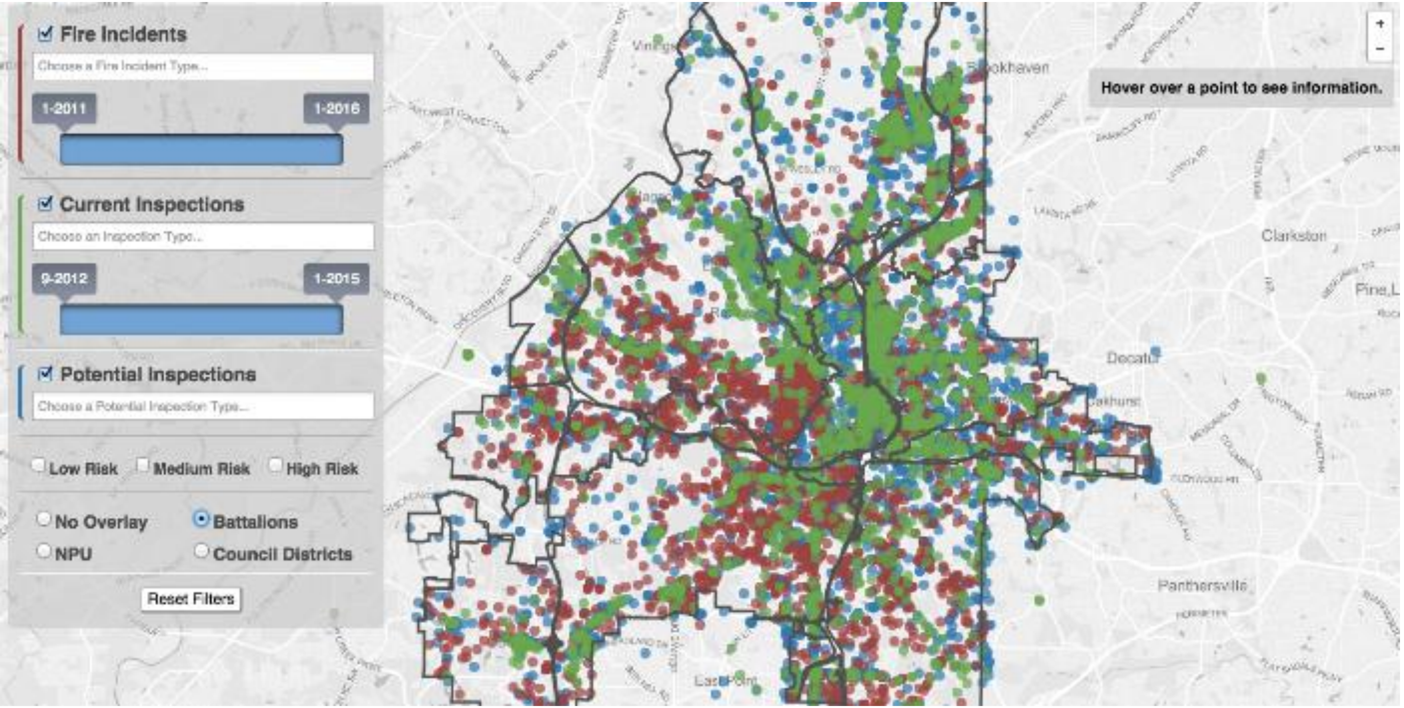


- Find Property Types:
  - Currently inspected types
    - Geocoding
    - Fuzzy text-matching
  - Text-mining of the *Fire Code of Ordinances*
  - Fire inspectors focus group
- Generate unique property list

+ Inspection List

- List of ~9,000 properties
  - Current Inspections: 2,600
  - **New potential Inspections: 6,500**
    - Business Licenses: 2,000
    - Google Places: 3,000
    - Liquor Licenses: 400
    - Pre K: 1,000
    - Child Car: 100
- Information:
  - Name, address, phone, type
  - Business ID, Google ID, Liquor License ID
  - Risk scores

+Interactive Inspection Map

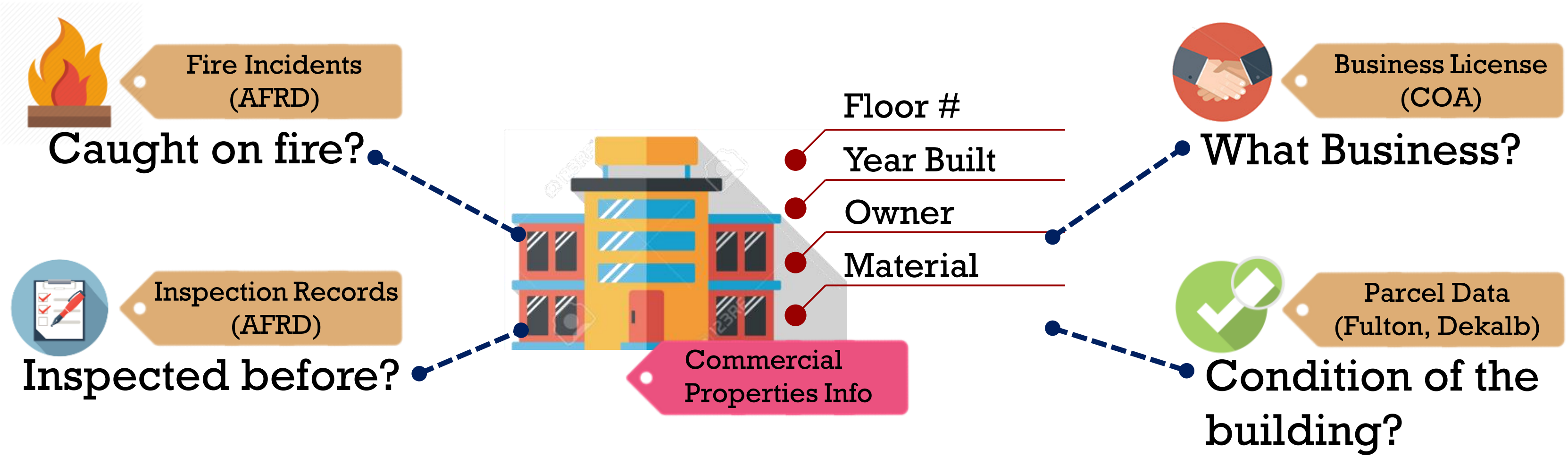


- Made with D3, Leaflet, and Mapbox
- Displays the current inspections, potential inspections, and fire incidents

How do we ensure the properties at greatest risk of fire are being inspected?

+ Fire Risk Predictive Model (Goal 2)

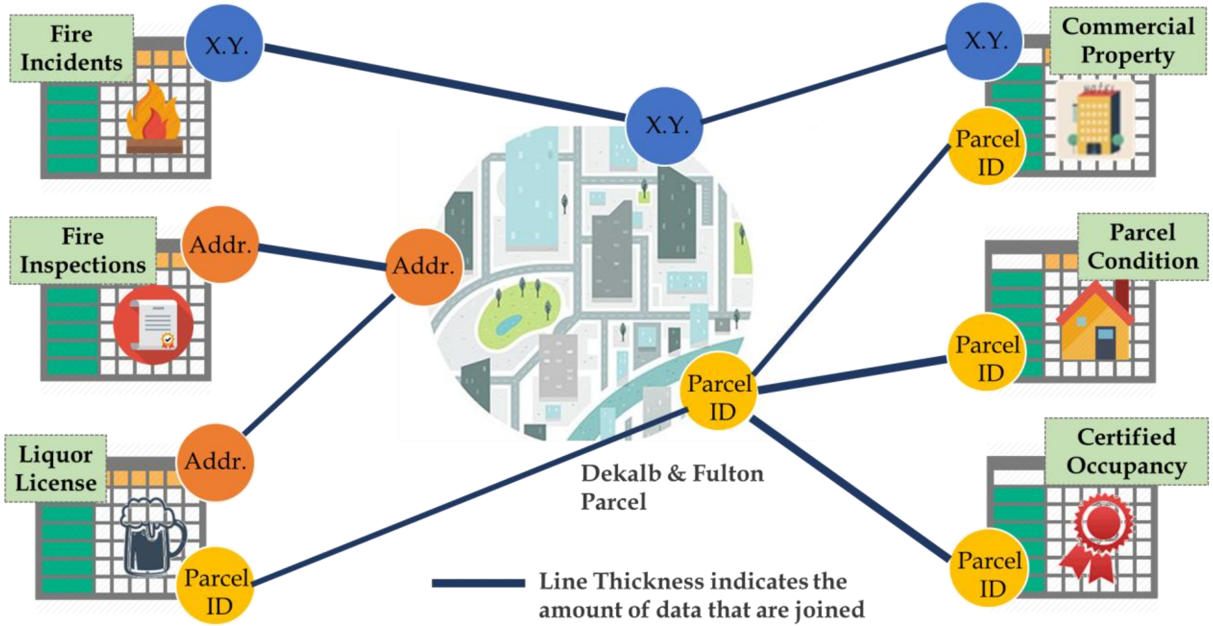
- Data from various sources



How do we CONNECT data from various sources together, so that they can talk to each other?

+ Fire Risk Predictive Model (Goal 2)

- Joining data from different sources



Approach:

- Geographic Information System (GIS)
- Google Geocoding API
- USPS mail address validation API

+ Fire Risk Predictive Model (Goal 2)

- Example of linked dataset

Property ID	Address	Floor	Year Built	Material	Renovation year	Owner	Land Use	Lot Condition	Structure Condition	Employment Density (per Sq Mi)	Owner Distance (Mile)	Inspection	Previous Fire
41815	Address 1	20	1929	Masonry	2006	xx1	Office	Good	Fair	1291.3	0.7	0	0
7381715	Address 2	11	1972	Wood Frame	-	xx2	Garden Apartment	Poor	Deteriorated	107.3	445.3	1	7

Commercial Property Dataset (Costar)

Parcel Data (Fulton, Dekalb)

SCI Data (City of Atlanta)

US Census Data

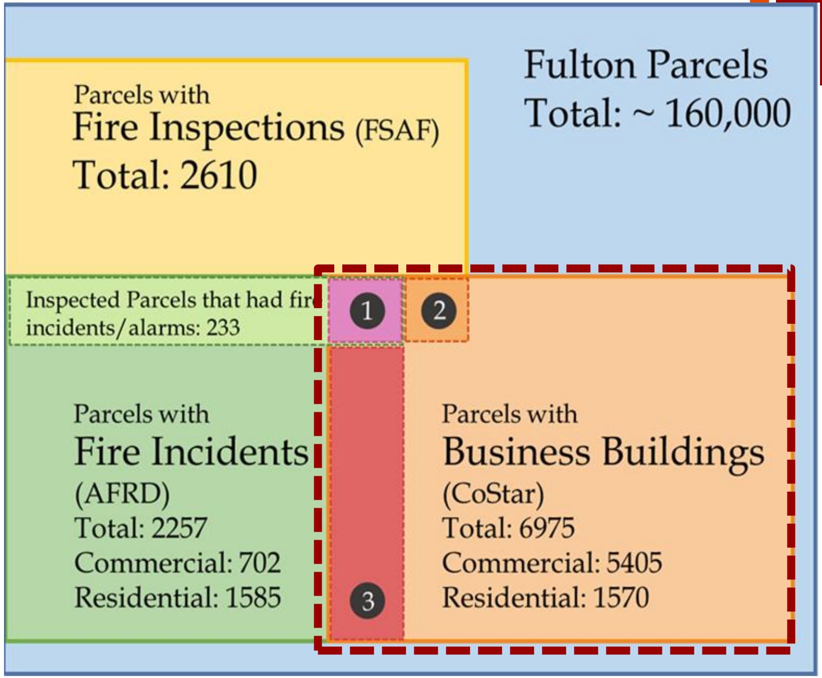
Created by us

Fire Incidents and Inspections

Final Table: **252** Variables describing different aspects of property

+ Fire Risk Predictive Model (Goal 2)

- Approaches
  - Machine Learning
    - SVM Model
    - 58 independent variables
    - Fire as binary dependent variable



1. Business Buildings with Inspections AND Fire Incidents
2. Business Buildings with Inspections
3. Business Buildings with Fire Incidents

+ Predictive Factors

Location	NPU (Neighborhood Planning Unit), zip code, submarket, neighborhood, tax district
Land / property use	property/business type, land use codes, zoning
Financial	tax value, appraisal value
Time-based	year built, year renovated
Condition	lot condition, structure condition, sidewalks
Occupancy	vacancy, units available, percent leased
Size	land area, building square feet
Building	number of units, style, stories, structure, construction materials, sprinklers, last sale date
Owner	owner or property management company, owner's distance from Atlanta
Demographics of location (based on traffic analysis zone)	density, land use diversity, intersection features, crime density, racial makeup
Inspection	whether or not the parcel had been inspected by AFRD

+ Predictive Factors

Location	**NPU (Neighborhood Planning Unit), zip code, submarket, neighborhood**, tax district
Land / property use	**property/business type, land use codes**, zoning
Financial	tax value, appraisal value
Time-based	year built, year renovated
Condition	lot condition, structure condition, sidewalks
Occupancy	vacancy, units available, percent leased
Size	**land area**, building square feet
Building	**number of units**, style, stories, structure, construction materials, sprinklers, last sale date
Owner	owner or property management company, owner's distance from Atlanta
Demographics of location (based on traffic analysis zone)	density, land use diversity, intersection features, crime density, racial makeup
Inspection	**whether or not the parcel had been inspected by AFRD**

+ Predictive Model Performance

- Used data from 2011 – 2014 to predict fires from 2014 – 2015
- Averaged results of 10 bootstrapped samples:
- Average accuracy: 0.77
- Average AUC: 0.75

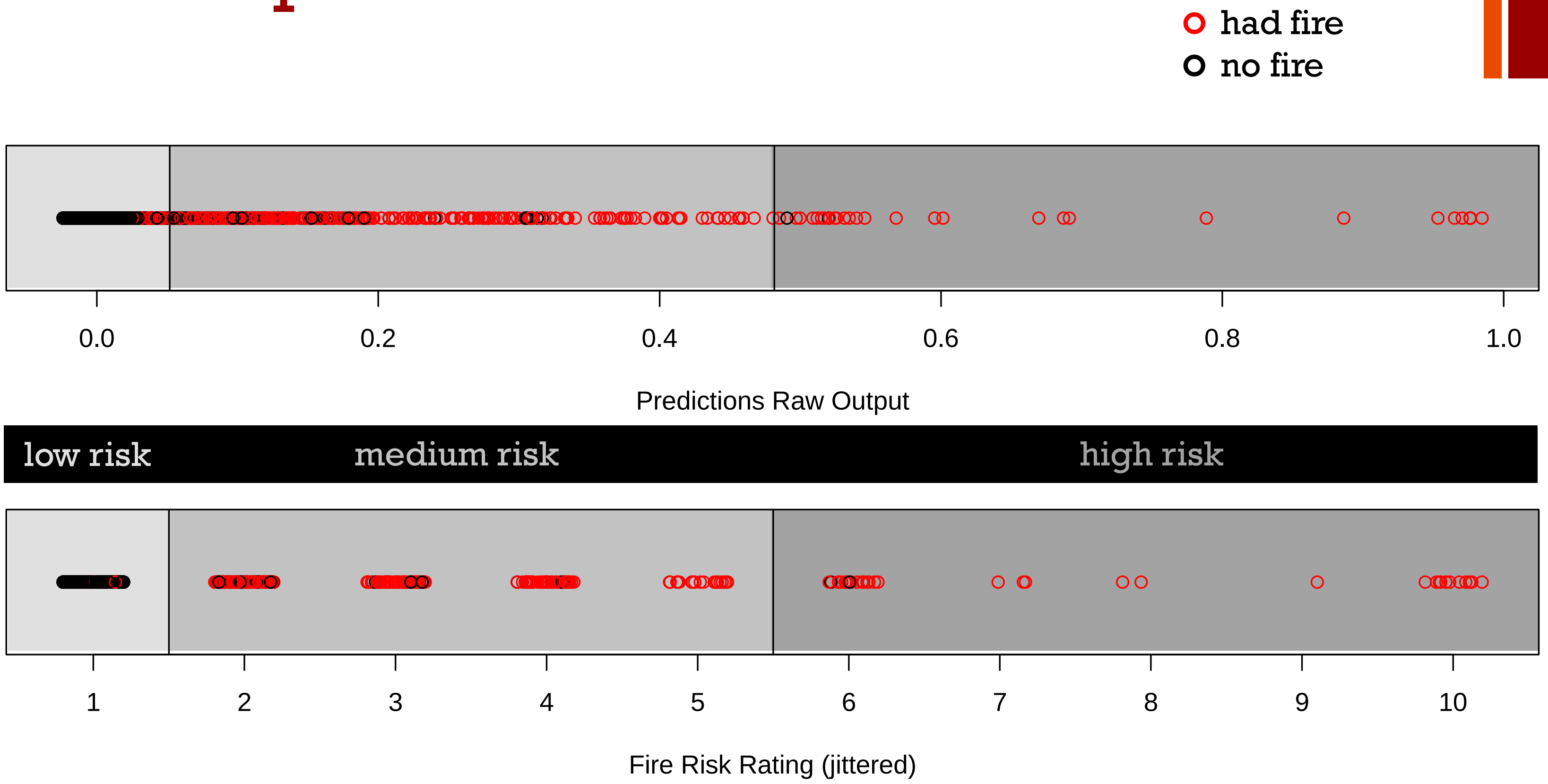
Actual \ Predicted	0	1
1	false negatives (had fire; predicted no fire) n = 38 0.2669	true positives (had fire; predicted fire) n = 104 0.7331
0	true negatives (no fire; predicted no fire) n = 1577 0.7678	false positives (no fire; predicted fire) n = 468 0.2322

+ Predictive Model Performance

- Used data from 2011-2015
- Averaged results of 10-fold cross validation:
- Average accuracy: 0.78
- Average AUC: 0.73

Actual \ Predicted	0	1
1	false negatives (had fire; predicted no fire) n = 16 0.3244	true positives (had fire; predicted fire) n = 39 0.6756
0	true negatives (no fire; predicted no fire) n = 609 0.7914	false positives (no fire; predicted fire) n = 158 0.2086

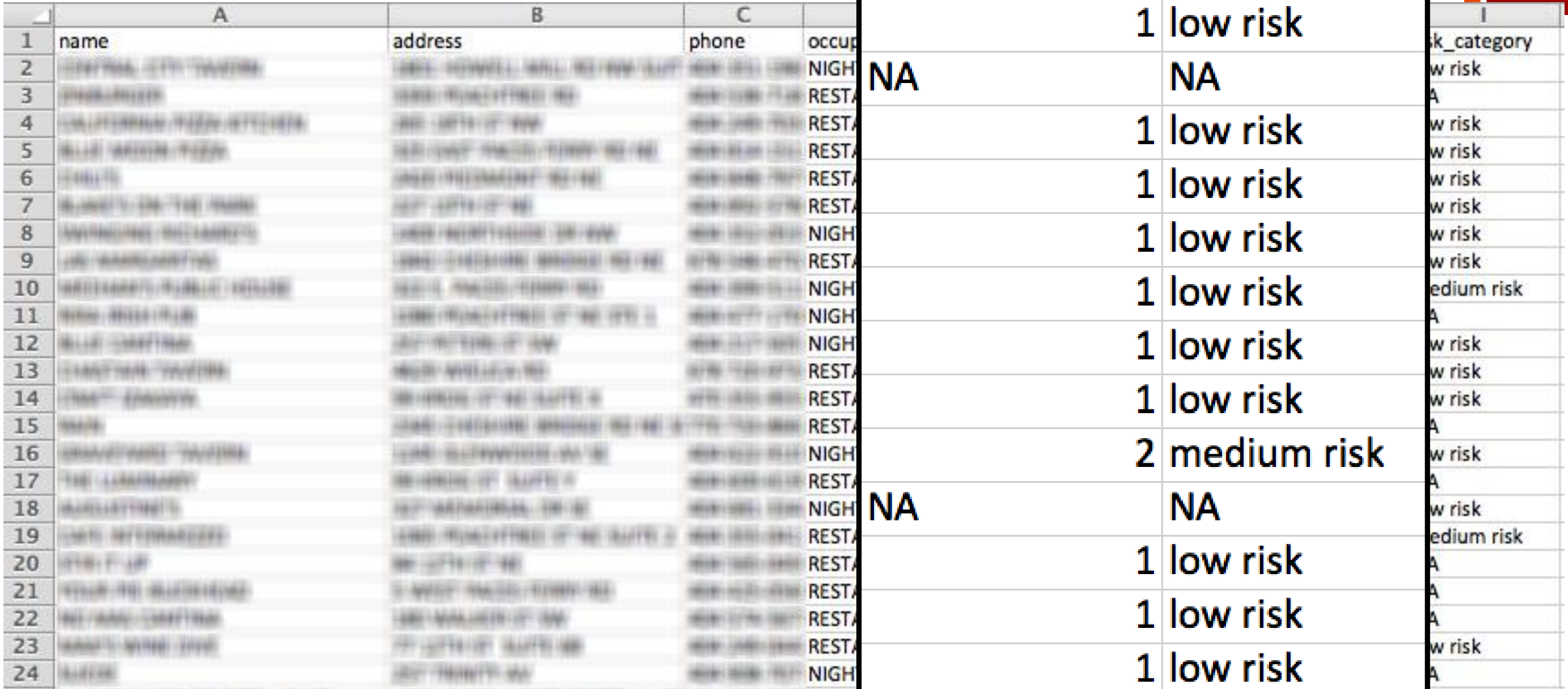
+ Applying Predictive Model to Potential Fire Inspections



+ Applying Predictive Model to Potential Fire Inspections

	A	B	C	D	E	F	H	I
1	name	address	phone	occup_type	b_sic_desc	google_type	fire_risk_rating	risk_category
2	[illegible]	[illegible]	[illegible]	NIGHTCLUB	RESTAURANTS	NA	1	low risk
3	[illegible]	[illegible]	[illegible]	RESTAURANT	RESTAURANTS	RESTAURANT	NA	NA
4	[illegible]	[illegible]	[illegible]	RESTAURANT	RESTAURANTS	RESTAURANT	1	low risk
5	[illegible]	[illegible]	[illegible]	RESTAURANT	RESTAURANTS	BAR	1	low risk
6	[illegible]	[illegible]	[illegible]	RESTAURANT	RESTAURANTS	NA	1	low risk
7	[illegible]	[illegible]	[illegible]	RESTAURANT	RESTAURANTS	BAR	1	low risk
8	[illegible]	[illegible]	[illegible]	NIGHTCLUB	RESTAURANTS	NIGHT_CLUB	1	low risk
9	[illegible]	[illegible]	[illegible]	RESTAURANT	RESTAURANTS	BAR	1	low risk
10	[illegible]	[illegible]	[illegible]	NIGHTCLUB	RESTAURANTS	BAR	2	medium risk
11	[illegible]	[illegible]	[illegible]	NIGHTCLUB	NA	NA	NA	NA
12	[illegible]	[illegible]	[illegible]	NIGHTCLUB	RESTAURANTS	RESTAURANT	1	low risk
13	[illegible]	[illegible]	[illegible]	RESTAURANT	RESTAURANTS	RESTAURANT	1	low risk
14	[illegible]	[illegible]	[illegible]	RESTAURANT	RESTAURANTS	RESTAURANT	1	low risk
15	[illegible]	[illegible]	[illegible]	RESTAURANT	RESTAURANTS	RESTAURANT	NA	NA
16	[illegible]	[illegible]	[illegible]	NIGHTCLUB	RESTAURANTS	BAR	1	low risk
17	[illegible]	[illegible]	[illegible]	RESTAURANT	RESTAURANTS	BAR	NA	NA
18	[illegible]	[illegible]	[illegible]	NIGHTCLUB	RESTAURANTS	RESTAURANT	1	low risk
19	[illegible]	[illegible]	[illegible]	RESTAURANT	RESTAURANTS	CAFE	2	medium risk
20	[illegible]	[illegible]	[illegible]	RESTAURANT	RESTAURANTS	RESTAURANT	NA	NA
21	[illegible]	[illegible]	[illegible]	RESTAURANT	RESTAURANTS	RESTAURANT	NA	NA
22	[illegible]	[illegible]	[illegible]	RESTAURANT	RESTAURANTS	BAR	NA	NA
23	[illegible]	[illegible]	[illegible]	RESTAURANT	RESTAURANTS	NA	1	low risk
24	[illegible]	[illegible]	[illegible]	NIGHTCLUB	RESTAURANTS	NIGHT_CLUB	NA	NA

+ Applying Predictive Model to Potential Fire Inspections



H	I
fire_risk_rating	risk_category
1	low risk
NA	NA
1	low risk
1	low risk
1	low risk
1	low risk
1	low risk
1	low risk
2	medium risk
NA	NA
1	low risk
1	low risk
1	low risk

+ Applying Predictive Model to Potential Fire Inspections

	A	B	C		I
1	name	address	phone	occup	k_category
2	[illegible]	[illegible]	[illegible]	NIGH	w risk
3	[illegible]	[illegible]	[illegible]	RESTA	A
4	[illegible]	[illegible]	[illegible]	RESTA	w risk
5	[illegible]	[illegible]	[illegible]	RESTA	w risk
6	[illegible]	[illegible]	[illegible]	RESTA	w risk
7	[illegible]	[illegible]	[illegible]	RESTA	w risk
8	[illegible]	[illegible]	[illegible]	NIGH	w risk
9	[illegible]	[illegible]	[illegible]	RESTA	w risk
10	[illegible]	[illegible]	[illegible]	NIGH	edium risk
11	[illegible]	[illegible]	[illegible]	NIGH	A
12	[illegible]	[illegible]	[illegible]	NIGH	w risk
13	[illegible]	[illegible]	[illegible]	RESTA	w risk
14	[illegible]	[illegible]	[illegible]	RESTA	w risk
15	[illegible]	[illegible]	[illegible]	RESTA	A
16	[illegible]	[illegible]	[illegible]	NIGH	w risk
17	[illegible]	[illegible]	[illegible]	RESTA	A
18	[illegible]	[illegible]	[illegible]	NIGH	w risk
19	[illegible]	[illegible]	[illegible]	RESTA	edium risk
20	[illegible]	[illegible]	[illegible]	RESTA	A
21	[illegible]	[illegible]	[illegible]	RESTA	A
22	[illegible]	[illegible]	[illegible]	RESTA	A
23	[illegible]	[illegible]	[illegible]	RESTA	w risk
24	[illegible]	[illegible]	[illegible]	NIGH	A

H	I
fire_risk_rating	risk_category
8	high risk
8	high risk
8	high risk
8	high risk
8	high risk
8	high risk
8	high risk
8	high risk
8	high risk
8	high risk
8	high risk
8	high risk
8	high risk

+ Summary of Deliverables

- Predictive model to generate fire risk score
- Integrated database of building information
- Prioritized list of properties to inspect
  - Currently Inspected (2,600)
  - Potential Inspections (5,300)
- Interactive map to view fires, inspections, and potential inspections

+ Practitioner's Guide

- Data Availability
- API daily query limits
  - Google Geocoding API – 1500 per key
  - Zillow API – 1000 per key
  - Walk score API – 5000 per key (approximately a week to get an active key!)

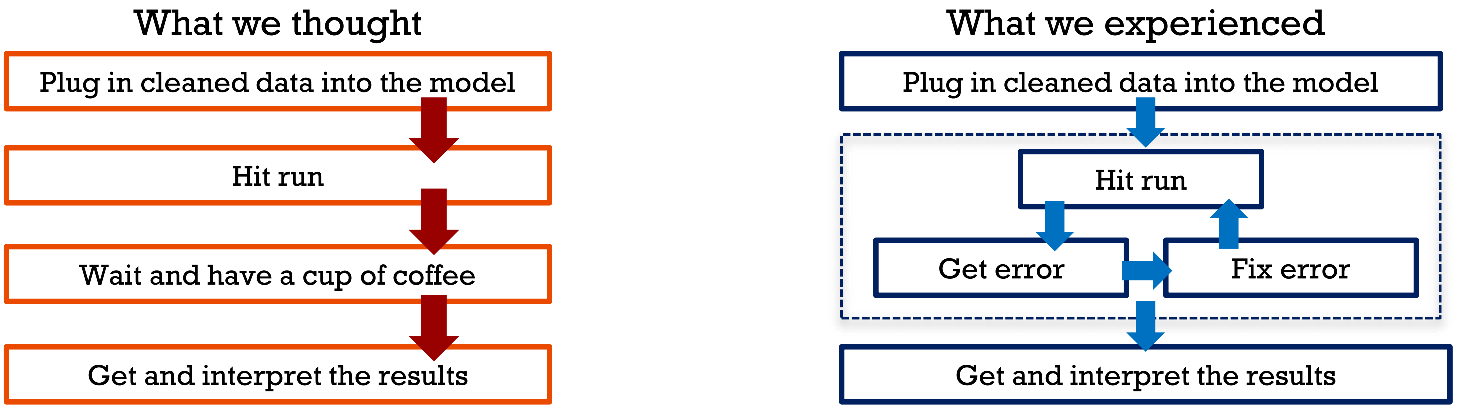
+ Practitioner's Guide

- Data are DIRTY
  - Formatting Issues
    - Address ---------- Martin Luther King Boulevard **vs.** M. L. K. blvd
    - Parcel ID ---------- 17-31000-xxxxxxx **vs.** 17 310 0 xxxxxxx
    - Null Values ------ Empty, " ", NAN, -1, 99, 9999, Null……
  - Resolution Issues
    - Building vs. Parcel vs. Block vs. Census Tract Level

ONE MONTH OF CLEARNING AND JOINING!

+ Practitioner's Guide

- Model Development
  - Understand your data: what to include in the model?
  - Model Error Fixing



+ Thank you!

Data Science for Social Good – Atlanta Fire Rescue Department

Team Members:

- Oliver Haimson | UC Irvine | ohaimson@uci.edu
- Michael Madaio | Georgia Tech | mmadaio@gatech.edu
- Xiang Cheng | Emory University | xcheng7@emory.edu
- Wenwen Zhang | Georgia Tech | wzhang300@gatech.edu

Partner:

- Atlanta Fire Rescue Department (AFRD)
- Dr. Matt Hinds-Aldrich (AFRD) | mhinds-aldrich@atlantaga.gov

Mentors:

- Dr. Polo Chau | Georgia Tech | polo@gatech.edu
- Dr. Bistra Dilkina | Georgia Tech | bdilkina@cc.gatech.edu