

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)







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

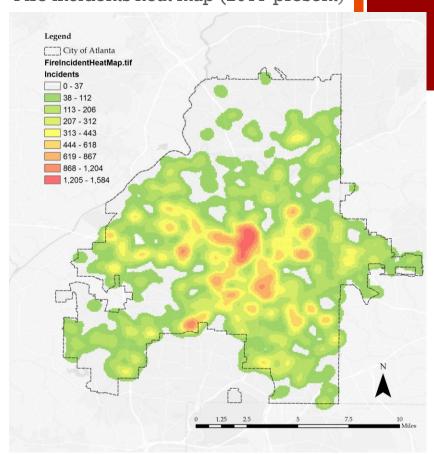






- 2,600 properties are inspected per year
- How do we help AFRD find new commercial properties that need inspection?
- How do we ensure the properties at greatest risk of fire are being inspected?

Fire incidents heat map (2011-present)





Goal 1: Find new properties to inspect

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

Goal 2: Prioritize inspections

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

+Data

■ 6+ sources

■ 2+ GB

■ ~200,000 Records Data

Fire Incident

Fire Inspection Permits

Liquor License

Atlanta Fire Department

Atlanta Regional Commission

nt

EST, 1892

Parcel Data

Atlanta Business Licenses

SCI Report

City of Atlanta

Source

TANTA GA

Neighborhood Planning Unit

Demographic Data

Census
Bureau

Socio-economic Data

U.S. Census Bureau

CoStar Property Report CoStar Group, Inc



Business Location Data Google APIs Google



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

Current Inspections



2,600



Business Licenses 20,000

Google places
10,000

Current Inspections



2,600

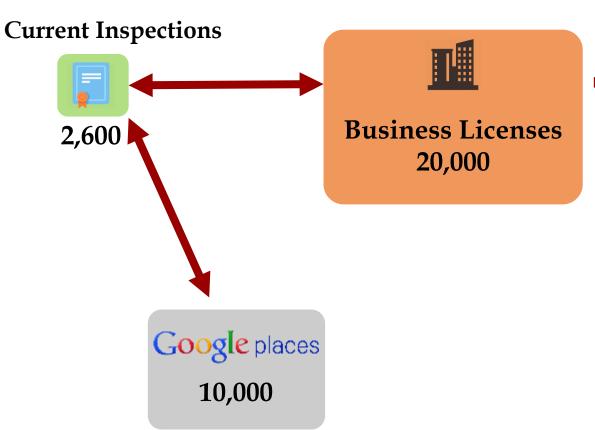


Business Licenses 20,000



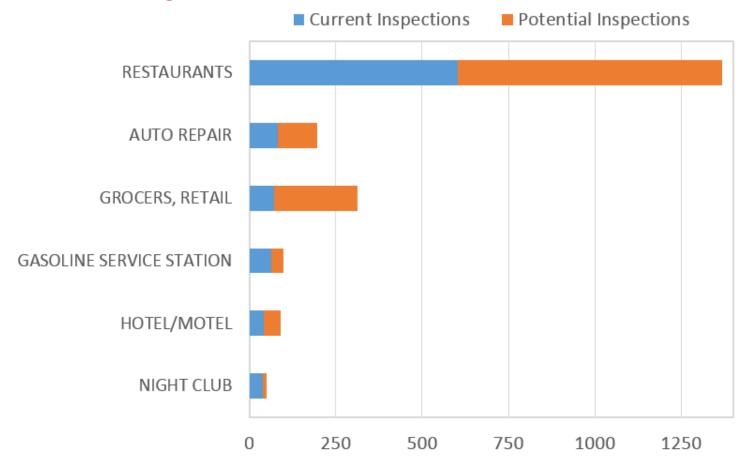
Currently inspected types

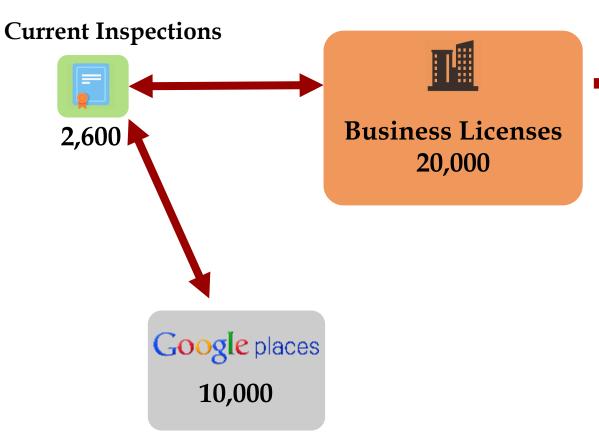
Google places
10,000



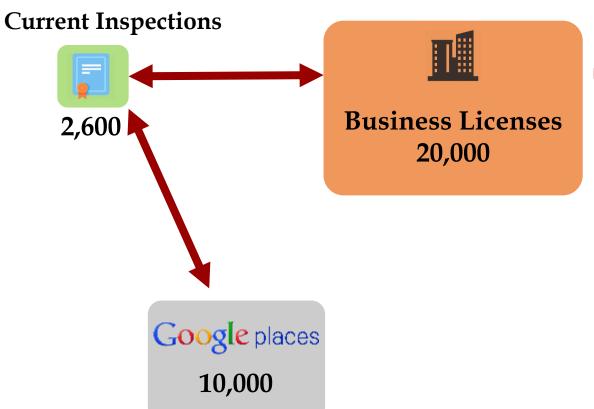
■ Find Property Types:

- Currently inspected types
 - Geocoding
 - Fuzzy text-matching



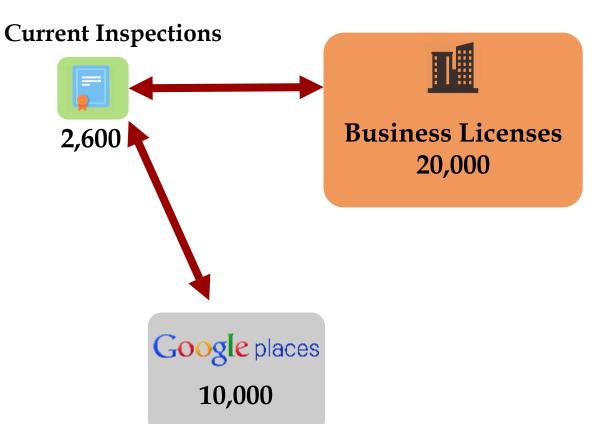


- Find Property Types:
 - Currently inspected types
 - Geocoding
 - Fuzzy text-matching

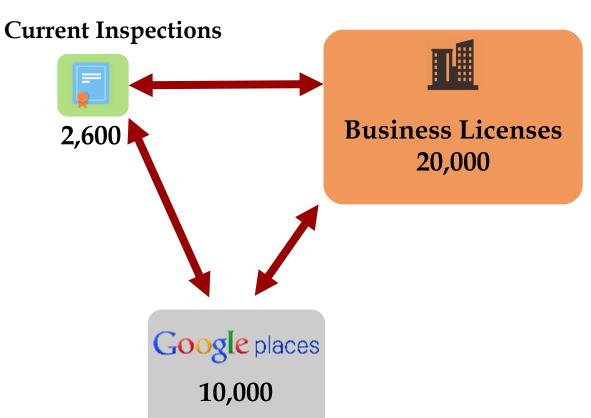


■ Find Property Types:

- Currently inspected types
 - Geocoding
 - Fuzzy text-matching
- Text-mining of the Fire Code of Ordinances
- Fire inspectors focus group



- 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



- 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

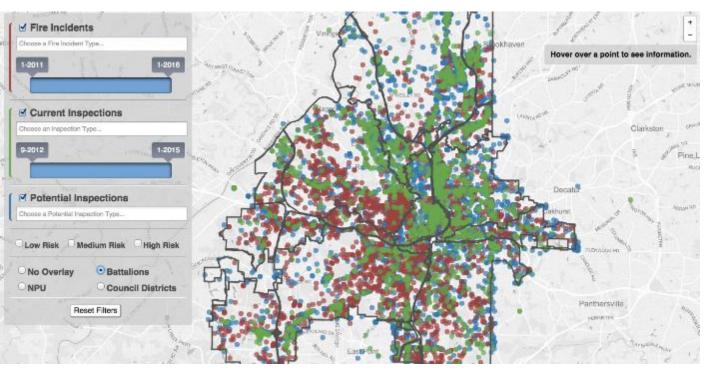
+ Inspection List

- 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

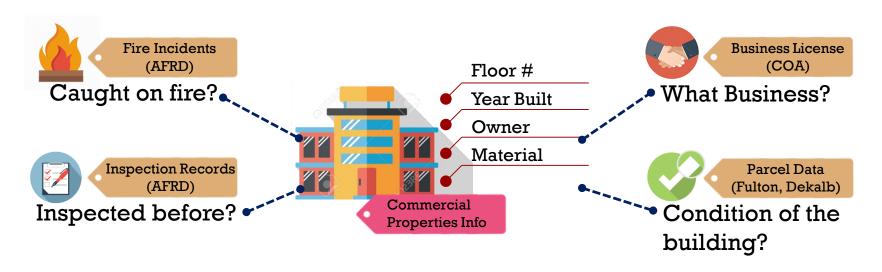


- Made with D3, Leaflet, and Mapbox
- Displays the current inspections, potential inspections, and fire incidents



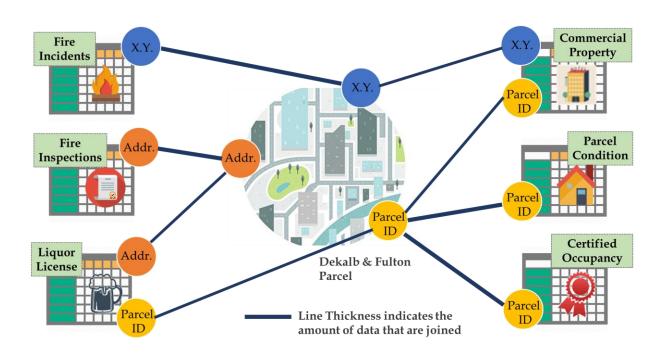
How do we ensure the properties at greatest risk of fire are being inspected?

■ Data from various sources



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

Joining data from different sources



Approach:

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

■ 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)		Inspection	Previous Fire
41815	Address 1	20	1929	Masonry	2006	xxl	Office	Good	Fair	1291.3	0.7	0	0
7381715	Address 2	11	1972	Wood Frame	-	xx2	Garden Apartment	Poor	Deteriorat ed	107.3	445.3	1	7

Commercial Property Dataset (Costar)

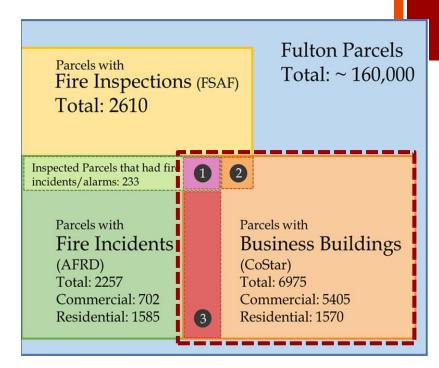
Parcel Data (Fulton, Dekalb) SCI Data (City of Atlanta)

US Census Data Created by us

Fire Incidents and Inspections

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

- 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

Actual

- 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

false negatives (had fire; predicted no fire)

> n = 38 0.2669

true negatives (no fire; predicted no fire)

> n = 1577 0.7678

true positives (had fire; predicted fire)

n = 104 0.7331

false positives (no fire; predicted fire)

> n = 468 0.2322

0 Predicted

+ Predictive Model Performance

Actual

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

false negatives (had fire; predicted no fire)

> n = 16 0.3244

true negatives (no fire; predicted no fire)

> n = 6090.7914

true positives (had fire; predicted fire)

> n = 39 0.6756

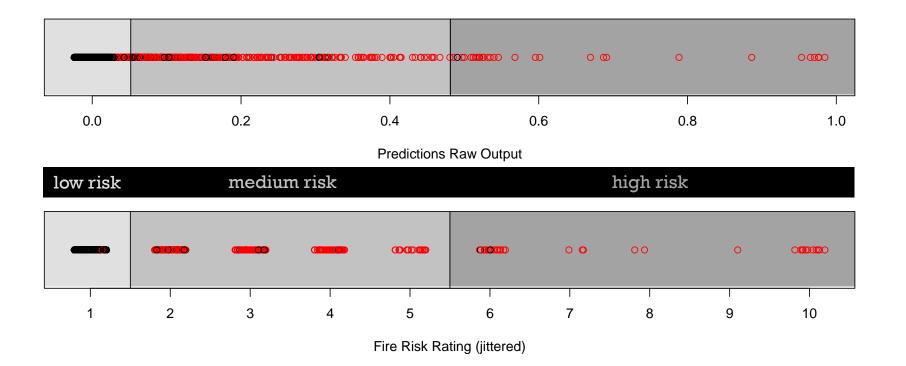
false positives (no fire; predicted fire)

> n = 158 0.2086

0 Predicted

+ Applying Predictive Model to Potential Fire Inspections

had fireno fire



+ Applying Predictive Model to Potential Fire Inspections

	A	В	C	D	E	F	Н	1
1	name	address	phone	occup_type	b_sic_desc	google_type	fire_risk_rating	risk_category
2	CONTROL CONTROL	TABLE DESWELL WILL METRIM SLIT	1600-1610-1600	NIGHTCLUB	RESTAURANTS	NA	1	low risk
3	PERSONAL PROPERTY.	SERVICE PROPERTY AND ADDRESS OF THE PARTY AND	Maria Comp. 15 (4)	RESTAURANT	RESTAURANTS	RESTAURANT	NA	NA
4	CONTRACTOR AND THE PROPERTY OF	JACK SETTLEF THE	969-199-793	RESTAURANT	RESTAURANTS	RESTAURANT	1	low risk
5	Miles Minister Trappy	MATERIAL PROPERTY OF THE	Miles Man 1516	RESTAURANT	RESTAURANTS	BAR	1	low risk
6	10000000	CONTROL STREET STREET STREET	NEW YORK THE	RESTAURANT	RESTAURANTS	NA	1	low risk
7	Shared to the TEE MARKET	AND AND DOM: NO	機能 排放 安张	RESTAURANT	RESTAURANTS	BAR	1	low risk
8	Approximation of the Contraction	THE RESERVE THE SAME THE PROPERTY.	Mar 1857-9816	NIGHTCLUB	RESTAURANTS	NIGHT_CLUB	1	low risk
9	John Valences paint / 1964	THE PERSON NAMED IN COLUMN	NUMBER OF STREET	RESTAURANT	RESTAURANTS	BAR	1	low risk
10	MATERIAL PLANT DESIGN	PRINCIPAL PRINCI	168 (88 153)	NIGHTCLUB	RESTAURANTS	BAR	2	medium risk
11	STREET STREET STREET	HARRISTON, OFFICE OF THE OFFI	MEN STYLET	NIGHTCLUB	NA	NA	NA	NA
12	Miller State Man.	ACTOR TORS OF SAME	100	NIGHTCLUB	RESTAURANTS	RESTAURANT	1	low risk
13	THREE TREE TRUITME	ORGANI ANNI LIE W. THE	879-750-875	RESTAURANT	RESTAURANTS	RESTAURANT	1	low risk
14	chan ganerie	MICHELL OF THE BUTTO O	of the later many	RESTAURANT	RESTAURANTS	RESTAURANT	1	low risk
15	Seasole	conditional matter as a	76 70 88	RESTAURANT	RESTAURANTS	RESTAURANT	NA	NA
16	STREAM THERE THE STREET	STATE OF THE PARTY OF THE PARTY OF	MENTAL STATE	NIGHTCLUB	RESTAURANTS	BAR	1	low risk
17	THE LANGUAGE	DESIGNATION OF SAMPLE OF	1681 1681 1635	RESTAURANT	RESTAURANTS	BAR	NA	NA
18	MARKARTHET IS	THE PARTACIONAL DECISION	1600 1000 1000	NIGHTCLUB	RESTAURANTS	RESTAURANT	1	low risk
19	CART METERARDISE	SHOUNDSTINGS OF HE AUTO &	1604 (810-984)	RESTAURANT	RESTAURANTS	CAFE	2	medium risk
20	STEEL TO LET	SECURITY SEC	66 (60) 340	RESTAURANT	RESTAURANTS	RESTAURANT	NA	NA
21	THE PERSON NAMED IN	S AND THE SECTION SEC	Mile Hill Hill	RESTAURANT	RESTAURANTS	RESTAURANT	NA	NA
22	HALL / BORNE - CARRY HARL	HER WAS DEED IT THE	Mile line less	RESTAURANT	RESTAURANTS	BAR	NA	NA
23	MARKET TO SECTION STORY	PERSONAL PROPERTY AND PERSONS ASSESSMENT	Mile Ministra	RESTAURANT	RESTAURANTS	NA	1	low risk
24	those on	SECTION TO MA	NUMBER OF	NIGHTCLUB	RESTAURANTS	NIGHT_CLUB	NA	NA

+ Applying Predictive Model to Potential

Fire Inspections

		•			fire_r	risk_r	ating	risk_category	
	A	В	C				1	low risk	1
1	name	address	phone	occup				IOW IISK	k_category
2	CONTROL CTT TAVETOR	LABOR TREMATEL MALL THE FROM TALLY	1000 (01) 1300	NIGH	NA			NA	w risk
3	SURFACE DESCRIPTION OF THE PERSON OF T	ARREST PROPERTY HEALTH HEALTH	Mile time / Las	RESTA	ראורו			140	A
4	1300,073 MMARC TAPPE (877) 1878.	ARE SETTLED THAT	1616 (1617 1917)	RESTA			1	low risk	w risk
5	MALES MARKET PROPERTY.	DESCRIPTION OF THE OWNER, THE PARTY HERE	Military Harry State	RESTA				TOW TISK	w risk
6	10000	AND DESCRIPTION OF THE PARTY.	NUMBER OF	RESTA			1	low risk	w risk
7	SEASON TO THE PERSON	1877 - 1877 B 127 - MG	機能 排放 安徽	RESTA				10 W TISK	w risk
8	ANTHORN BUT THE LABOUR TO	AND THE PERSON NAMED IN COLUMN	Mile ING-HU	NIGH			1	low risk	w risk
9	July 1 Marketti pater 1 Mg.	THE PERSON NAMED IN	NAME OF TAXABLE PARTY.	RESTA					w risk
8 9 10	WATERWAY CHARLE DESIGN	THE R. P. LEWIS CO., LANSING MICH.	168 (88 153	NIGH			1	low risk	edium risk
11	WHITE STREET WAR	SHIRL PROPERTY OF THE STATE OF	MEN SETT STE	NIGH					A
11 12	MillER COMPT MAIN	ACTIVITIES OF AM	NUMBER OF STREET	NIGH			1	low risk	w risk
13	1 186 T 188 TYPE 1984	MODE AND AND ASSAULT	新物 马拉 田市	RESTA					w risk
14	I THE SHOWING	MICHIGAN OF THE PARTY OF	afti Michigan	RESTA			1	low risk	w risk
15 16	SMA-No.	COMPLETED AND DESIGNATION OF THE R	76 70 88	RESTA					A
16	STRANG MINE TANGERS	HART BUT WATER ON THE	MERCHANIST MAN	NIGH			2	medium risk	w risk
17	THE LANGERSHIP	SECOND OF BUTTON	100010000100000	RESTA					A
17 18	MARKANITHE V	THE PARTMENT OF RE	MERCHANIC TRA	NIGH	NA			NA	w risk
19	COSTS SECTIONAL COSTS	SAME PROSPERED OF SECULIFIED	1696 (83) (84)	RESTA					edium risk
19 20	INTER TO ART	SECURITY SEC	100 100 100	RESTA			1	low risk	A
21	PERSONAL PROPERTY AND ADDRESS OF THE PARTY AND	S MIST THESE TURNS HE	Marie States	RESTA					A
22	1865 / 600/G - 1-6080 THINKS	HER WAS DESCRIPTION.	Miles living later	RESTA			1	low risk	A
23	MARKET TO SPECIAL PROPERTY.	PET SEPTEMBER GALFFEI MER	169-149-140	RESTA					w risk
24	Budient:	SECTIONAL TO MAY	168 168 167	NIGH			1	low risk	4

+ Applying Predictive Model to Potential

Fire Inspections

H

Inspections

					fire_risk_rating	risk_category	
	A	В	C		g	high risk	1
1	name	address	phone	occup	0	HIGHTISK	k_category
3	CONTROL CTT TAYOUR	THE REPORT LAND, WE FROM THAT	1000-1001-100	NIGH	8	high risk	w risk
3	PROPERTY.	CONTROL PROPERTY HAVE A	168 108 758	RESTA	U	mgm msk	A
4	1300 PT 1000 Pt 100 Pt 100 PT 100 PT	SHE SHITE OF THE	1696-1690-1933	RESTA	8	high risk	w risk
5	Mild Minte Tippe	MATCHET PRODUCTIONS OF THE	Miles Miles 124	RESTA			w risk
6	100010	CONTRACTOR SECURE	NEW YEAR THE	RESTA	8	high risk	w risk
7	SEASON TO THE PERSON	107 1071 17 16	MEN MILE STR	RESTA		_	w risk
9	MANAGEMENT THE LABOUR.	THE PROPERTY SHARE SWITTER	MR 100-300	NIGH	8	high risk	w risk
9	JOSE : MARKET MARKET PARKET	THE THE PERSON NAMED IN	NAME AND PARTY	RESTA		_	w risk
10	MARIE AND TO PERSON TO THE OWNER.	SECULAR PROPERTY AND ADDRESS OF THE PARTY AND	100 000 000	NIGH	8	high risk	edium risk
11	Martine - Martin 179 Life	AND PROPERTY OF	MEN NETT LYTS	NIGH		_	A
12 13	Michael College Made	AND THE RESIDENCE	MAN THE REAL PROPERTY.	NIGH	8	high risk	w risk
13	THE TRUE TRUE TO STATE OF THE	MICH. WHEN IN W.	前海 "持"	RESTA		_	w risk
14 15 16	The factorie	MERCHANT OF THE BAPTE OF	975 Mil-80	RESTA	8	high risk	w risk
15	Table 181	CONTRACTOR SERVICE SERVICE	76-70-88	RESTA		_	A
16	CONTRACTOR CONTRACTOR	THE RESERVE OF BUILDING	MENTER MAI	NIGH	8	high risk	w risk
17 18	THE LUMBERSHIP	SECURIOR OF SLATTER W	100001000010001000000000000000000000000	RESTA		_	A
18	PROPERTY NAME OF	THE PARTMANNAL THE RE	100100-000	NIGH	8	high risk	w risk
19 20	CARTE OFFERBRUSES	AND TRACTION OF BUILDING	MIN (83) (84)	RESTA		_	edium risk
20	STATE OF STA	ME ATTEM ME	160 160 160	RESTA	8	high risk	A
21 22 23	PERSONAL PROPERTY AND ADDRESS OF THE PERSON NAMED AND ADDRESS	S WAS THE STATE OF THE	用作业的由	RESTA	•	hi-h wi-le	A
22	FREE / HUNGE - LANGE THINKS	SECONDALISED TO UN	Mile Info Inc	RESTA		high risk	A
	ANADOL - MATERIAL STORE	THE SHIP BUT BUT SHIPE SHIP	16th 16th 16th	RESTA	0	himb viale	w risk
24	Bud HE	SECTION TO SEC	50K HB: 107	NIGH	8	high risk	A

+ 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!)

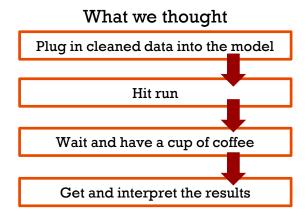
+ Practitioner's Guide

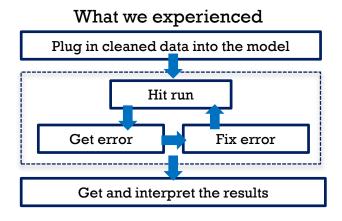
- Data are DIRTY
 - Formatting Issues
 - Address ----- Martin Luther King Boulevard vs. M. L. K. blvd
 - Parcel ID ----- 17-31000-xxxxxxx vs. 17 310 0 xxxxxxx
 - Null Values ----- Empty, "", NAN, -1, 99, 9999, Null......
 - Resolution Issues
 - Building vs. Parcel vs. Block vs. Census Tract Level

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







