### +

#### **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

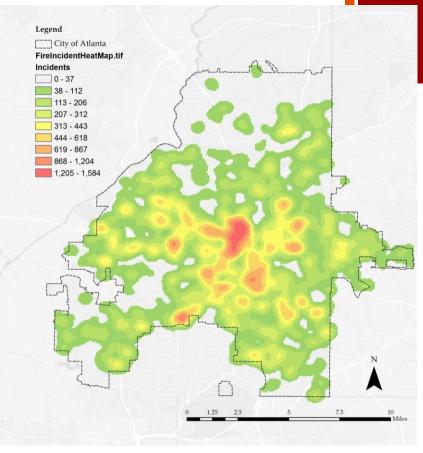




### +Problem

- 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

- <u>List of new properties:</u> from external business and property databases
- <u>Prioritized list:</u> 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	Data	Source			
	Fire Incident	E51, 1 1982			
- 0	Fire Inspection Permits	Atlanta Fire Department			
■ 6+ sources	Liquor License				
■ 2+ GB	Parcel Data	AS URGEN			
■ ~200,000	Atlanta Business Licenses	City of Atlanta			
Records	SCI Report				
	Neighborhood Planning Unit	Atlanta Regional Commission 🖊 🍋			
	Demographic Data	United States"			
	Socio-economic Data	U.S. Census Bureau			
	CoStar Property Report	CoStar Group, Inc			
	Business Location Data	Google APIs Google			

# 6

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

**Current Inspections** 







**Current Inspections** 

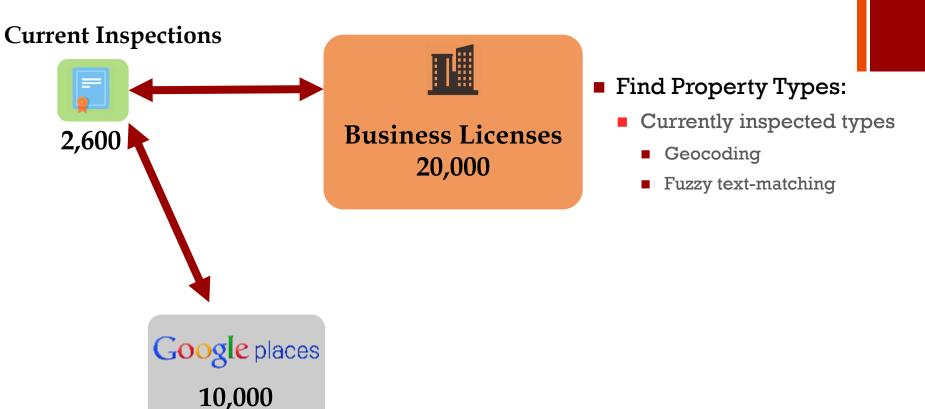




#### Find Property Types:

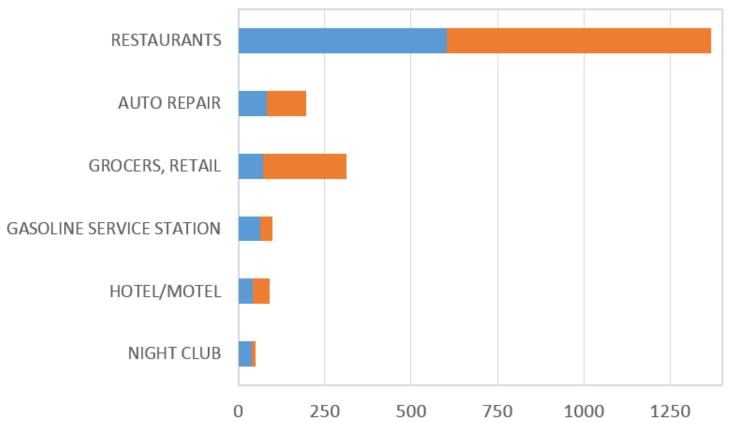
Currently inspected types

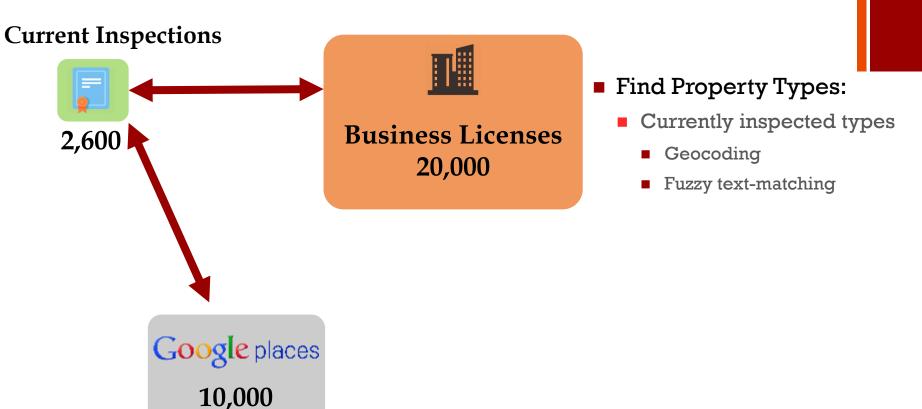




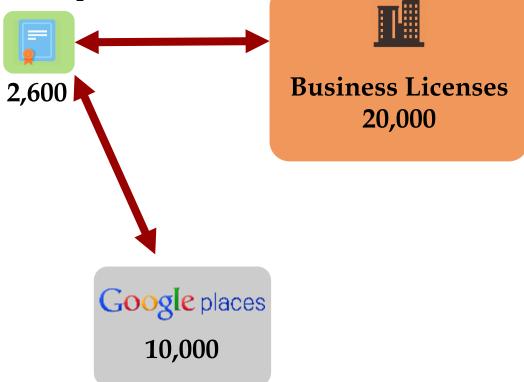
Current Inspections







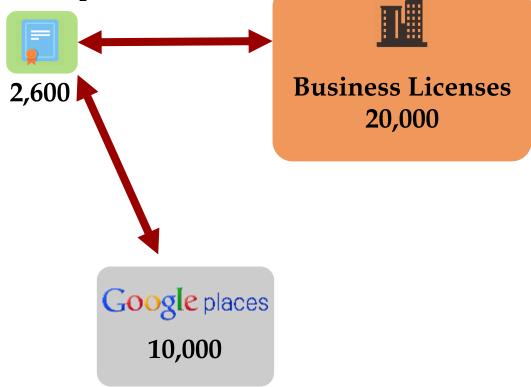
#### **Current Inspections**



#### • Find Property Types:

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

#### **Current Inspections**

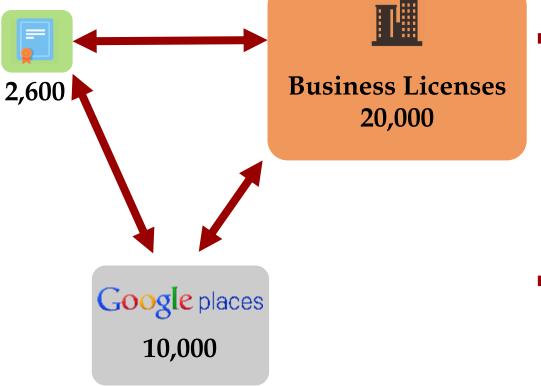


#### Find Property Types:

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

#### Generate unique property list

#### **Current 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

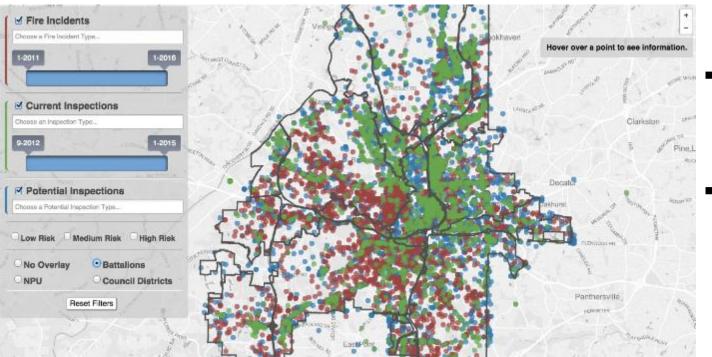
#### New potential Inspections: 6,500

- 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

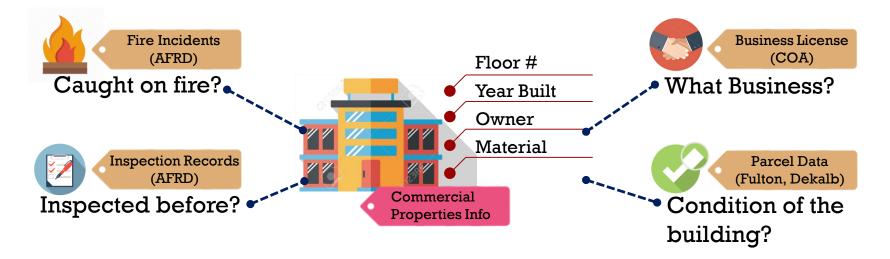
# +Interactive Inspection Map



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

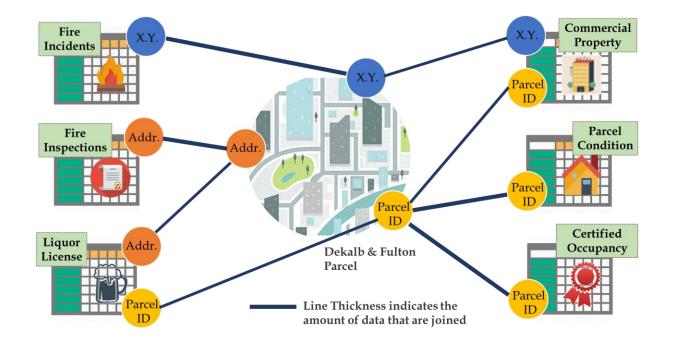
# How do we ensure the properties at greatest risk of fire are being inspected?

#### 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

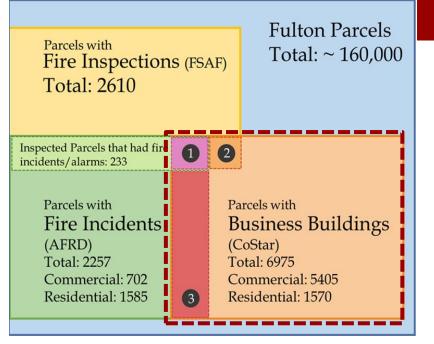
#### Example of linked dataset

Property ID	Address	Floor	Year Built	Material	Renovation year	Owner	Land Use	Lot Condition	Structure Condition			Inspection	Previous Fire
41815	Address l	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)			t		rcel Data (Fulton, Dekalb)	SCI I (City Atlar	y of	US Census Data	Created by us	Fire Inc and Insp			

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	<b>number of units</b> , 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

0

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

Predicted

# + Predictive Model Performance

Actual

1

0

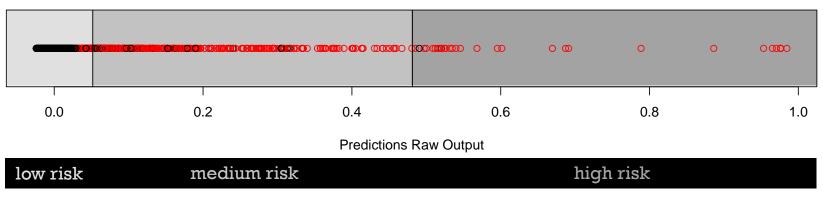
- Used data from 2011-2015
- Averaged results of 10-fold cross
   validation:
- Average accuracy: 0.78
- Average AUC: 0.73

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

Predicted

### + Applying Predictive Model to Potential Fire Inspections

had fireno fire





Fire Risk Rating (jittered)

### + Applying Predictive Model to Potential Fire Inspections

_2	A	B	C	D	E	F	H	1
1	name	address	phone	occup_type	b_sic_desc	google_type	fire_risk_rating	risk_category
2	CORPORATION CONTRACTOR	LABOR DEPENDENCES MALL MER PROMINED	100010-0010-0000	NIGHTCLUB	RESTAURANTS	NA	1	low risk
3	(Tringen Lines of the	CONTRACTOR AND A REPORT OF A REPORT OF	1000-1200-1200	RESTAURANT	RESTAURANTS	RESTAURANT	NA	NA
4	1-10-171-00084-00-10-10-10-10-10-10-10-10-10-10-10-10-	- MARE - JART 18 - CT - MARA	1000-000-000	RESTAURANT	RESTAURANTS	RESTAURANT	1	low risk
5	Miled Pasterias Property	1000 (1000) 1000 (10) 1110000 (00) - 000		RESTAURANT	RESTAURANTS	BAR	1	low risk
6	1	10000011100000000000000000000000000000	10101-0101	RESTAURANT	RESTAURANTS	NA	1	low risk
7	the sheet in the way wanted	101-10100-00-001	10001003-0278	RESTAURANT	RESTAURANTS	BAR	1	low risk
8	Conversion of the Content of the	LOUGH CALLERY TOLLING LOUP AND	1000-000-000	NIGHTCLUB	RESTAURANTS	NIGHT_CLUB	1	low risk
9	Line research week'r rule.	LOUGH CONTRACTOR AND AND AND AND AND	INTERIORE OF THE	RESTAURANT	RESTAURANTS	BAR	1	low risk
10	WARDING AND TO BE AND THE OWNER	HARDER HARD COLUMN TO A	10081-010-01-01	NIGHTCLUB	RESTAURANTS	BAR	2	medium risk
11	1919-10-1 (\$10.00 179-1.00)	LANSE (WILLS FRIME) OF BEI OFT	1000 1000 1000	NIGHTCLUB	NA	NA	NA	NA
12	Maria and Constant Palaster	ALC: NO THE OWNER OF SMALL	1000.0117.000	NIGHTCLUB	RESTAURANTS	RESTAURANT	1	low risk
13	11.1.0007 100.00. 100.0011000	MALEN AVAILATING ME	10700 1220-0270	RESTAURANT	RESTAURANTS	RESTAURANT	1	low risk
14	inter antipolitik	INCOMPANY OF AN INSTEED	-	RESTAURANT	RESTAURANTS	RESTAURANT	1	low risk
15	maria-	citable - refractable - anteriorate - and - in	76 34 444	RESTAURANT	RESTAURANTS	RESTAURANT	NA	NA
16	COMPANY CONTRACTOR OF CONTRACTOR	Louise - and a superior in the same		NIGHTCLUB	RESTAURANTS	BAR	1	low risk
17	THE LANGERSON	INCOMPANY OF ANYTHIN		RESTAURANT	RESTAURANTS	BAR	NA	NA
18	106-2012-2-017-108001-10-	HEPPARTARIAN, INCH.		NIGHTCLUB	RESTAURANTS	RESTAURANT	1	low risk
19	CARTE INFERNMENTER	LINE WORLDFIND OF ME AATTS O	1000 000 000	RESTAURANT	RESTAURANTS	CAFE	2	medium risk
20	10101-01-0-00	100-10710-00-001	1000-000-000	RESTAURANT	RESTAURANTS	RESTAURANT	NA	NA
21	WINDOW WIN COLOURS DECK	TO ANY TOP THAT THE PARTY NAME	Married Street	RESTAURANT	RESTAURANTS	RESTAURANT	NA	NA
22	Have - manage	1.000 1.000A.000 FM 127 100A	state in the last	RESTAURANT	RESTAURANTS	BAR	NA	NA
23	Construction of the construction of the second	THE OWNER AND INCOME.	HERE SHEET HAVE	RESTAURANT	RESTAURANTS	NA	1	low risk
24	(Aud a line)	1000 - 100000 TWI. 100/	10101-0020-0027	NIGHTCLUB	RESTAURANTS	NIGHT_CLUB	NA	NA

### + Applying Predictive Model to Potential Fire Inspections

					fire_risl	k_rating	risk_category	
	A	В	C	1		1	low risk	1
1	name	address	phone	occup		-		k_category
2	CONTRACTOR OF THE PROPERTY OF	LABOR DEPARTELL MARKED AND PROMITIONS		NIGH	NA		NA	w risk
3	- PT-RATE - AND - PT-	10-000 - Water ( # 190.01 - Wa	1000-0100-0100	REST				Α.
4		1000 - 100 10 - 01 - 0100	1000-000-000	REST		1	low risk	w risk
5	Married Teleforoide ( This State	TANK STATE THAT THE TRANSPORT OF A	-	REST/		-	IOW HISK	w risk
6	1-0.0001	10000001010000000000000000000000000000	10101-0101	REST		1	low risk	w risk
7	the control for single while interesting	100 - 100 VB-100 - 88	1000-000-0716	REST		-	iow hisk	w risk
8	Characteristic ( the Characteristic)	London - Annihi Talanihi Law - Anh	-	NIGH		1	low risk	w risk
9	- 2042 ( Malenatics painted ( Male)		INTERIORE STREET	REST				w risk
10	100011210000/1119100010111000000	Designed, respectively respectively wear	1000 1000 1111	NIGH		1	low risk	edium risk
11	101000.00100000000000000000000000000000	14.000 (TELLA CONTINUE) (T. 188) (T. 1)	-	NIGH				Ą
12	Maria M. Conter State.	store the water or state		NIGH		1	low risk	w risk
13	- 1.800 T 100/08. 100/08.	MARCHE ARTIST OF MAR	are the set	REST				w risk
14	1 TALE - A PARTING VE	HEIMAN IN THE LATER OF	-	REST/		1	low risk	w risk
15	(market)	interior - and the read of the second s	176 - 18 Mar	REST/				Α
16	COMPANY CONTRACT CONTRACTOR	- and - mail available - she had		NIGH		2	medium risk	w risk
17	-1000 -1.1.000 -000 page 10-	DECOMPTON OF CAMPUS OF		REST/				Ą
18	100.002.0021000110	14/11/14/0 Part 1995/86. 1091-101		NIGH	NA		NA	w risk
19	CONTRACTOR OF STREET,	A REAL PROPERTY OF AN ANTEL OF	-	REST/				edium risk
20	1010010-0-100	100	-	REST/		1	low risk	4
21	PERSONAL PROPERTY AND INCOME.	in application of the second second second		REST/		4		4
22	their r statute - set all 1 Made	- SHE LANDA LOS IN LCC. (MAR)	server to be the server	REST/		1	low risk	4
23	Laurent to artification and and	1777-1217-98-1287 18447-91-1888	Here in the	REST		4	lavv nlale	w risk
24	dauge of	- Allow - Table Barry Mr. Mark	1000-0030-007	NIGH		1	low risk	4

## + Applying Predictive Model to Potential Fire Inspections H I

					IIIe_IISK_Iatilig	TISK_Category	
	A	В	C	1	Q	high risk	1
1	name	address	phone	occup	0	Ingittisk	k_category
2	CONTRACTOR OF THE OWNER.	LABOR CONVERSIONAL METRICAL	10000-0010-0000	NIGH	8	high risk	w risk
3	THE PARTY OF THE P	CONTRACT PROVIDE A PERSON AND	10001030-1220	REST/		-	Δ.
4		1000-10010-00-000	1000 1000 1000	REST/	8	high risk	w risk
5	Would Indexed Transport	TANK STATE THE CONTRACTOR AND THE		REST/		-	w risk
6	1-01004-010	10000001100000000000000000000000000000	1010-1010-1011	REST/	8	high risk	w risk
7	the sheet to the "the transfer	100 - 100 - 100 - 100 - 100 -	-	REST/		-	w risk
8	Approximation and the construction	Longent - depender Traditional Land - depen-	-	NIGH	8	high risk	w risk
9	- 204 Halanders merer frage		NUMBER OF TAXABLE PARTY.	REST		-	w risk
10	100011210000701101000101-00001200	100.00110. (PROJ. 1000) (PERSONNEL 100.0	1000 1000 1111	NIGH	8	high risk	edium risk
11	191910-00 - 00123-0119-1-001	LANSE CTUDE CONTRACTOR OF THE OWNER OF THE	-	NIGH		-	Ą
12	Maria M. California Maria	store the balls of the		NIGH	8	high risk	w risk
13	- LOUT TRAD. TRADITION	10012-001-ARVER-101-001-001	are the set	REST		-	w risk
14	inter species to	THE LEWIS CO., MICH.		REST/	8	high risk	w risk
15	(market)	CANADA - CANTON AND ADDRESS OF THE CANE -	76 34 444	REST		_	Ą
16 17	CONSTRUCTION OF THE OWNER	1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.		NIGH	8	high risk	w risk
	-1081-1.1.0000000000	THE LOOP OF CALLSREE IN		REST		-	Ą
18	100-1002-1007-100001-10	1017 100 DOM 100 00. 100 101	100010000000000000000000000000000000000	NIGH	8	high risk	w risk
19	CONTRACTOR OF STREET	HARD PROPERTY AND IN THE CAPTURE OF	-	REST		-	edium risk
20	Hartman E. The	100	1000 1000 1000	REST	8	high risk	Α
21	WYNER WILL BELLEVILLE	The Application of the state of	-	REST	•	high sigh	Ą
22	the state of the second st	- sales - Advide, con the commonly	search in the later	REST	8	high risk	4
23	Laurent 11 - Archest - Ertrett	THE LOCATE IN CONTRACTOR	HERE INSTRUCTOR	RESTA	0	high rick	w risk
24	Badd He	salar - the start of start	10.01	NIGH	8	high risk	4

# + 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!)

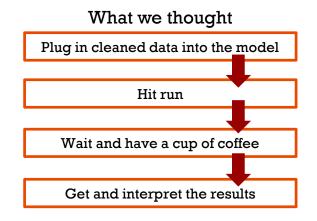
## + Practitioner's Guide

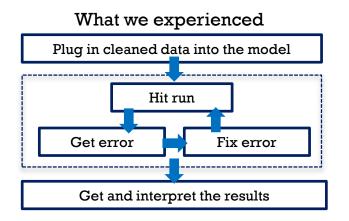
- Data are DIRTY
  - Formatting Issues
    - Address ------ Martin Luther King Boulevard vs. M. L. K. blvd
    - Parcel ID ----- 17-31000-xxxxxx vs. 17 310 0 xxxxxxx
    - Null Values ----- Empty, "", NAN, -1, 99, 9999, Null.....
  - Resolution Issues
    - Building vs. Parcel vs. Block vs. Census Tract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



