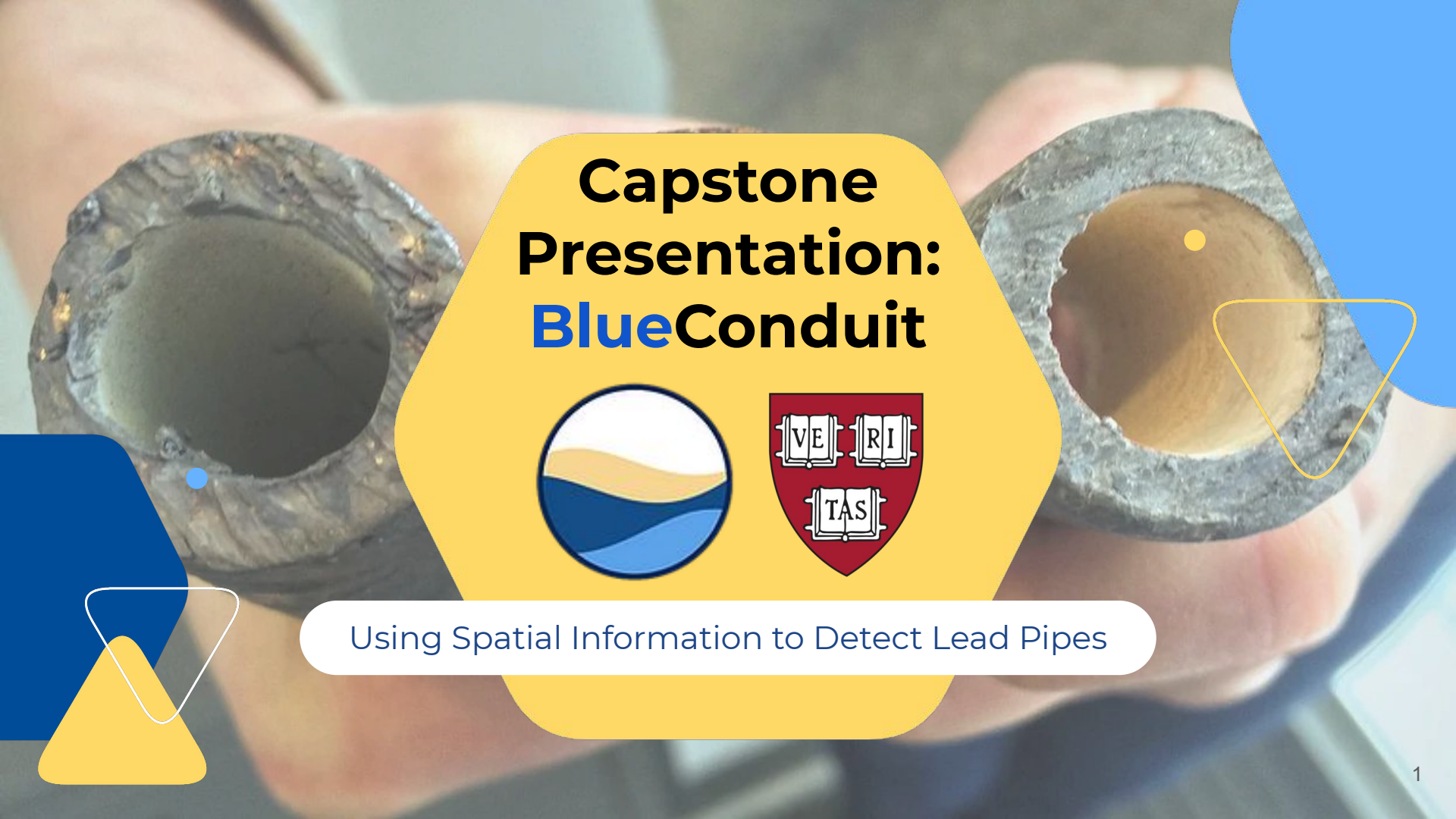
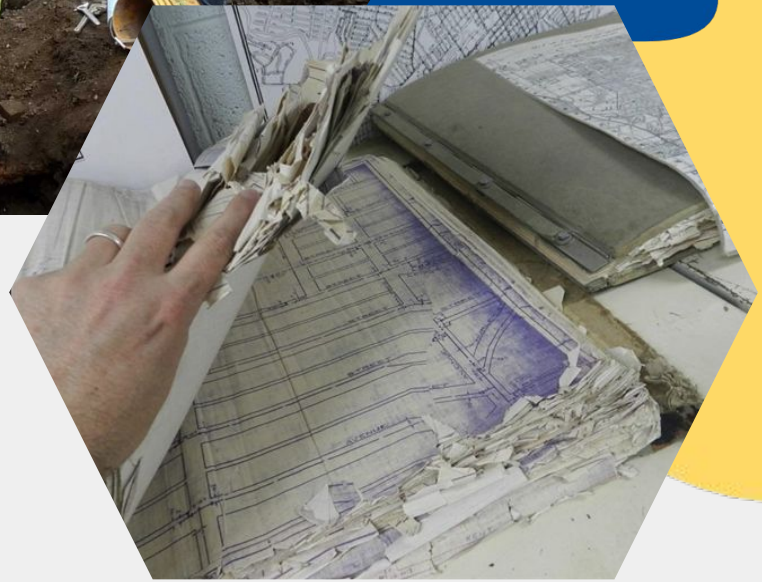


# Capstone Presentation: BlueConduit

Using Spatial Information to Detect Lead Pipes

# The Problem: Lead Pipes

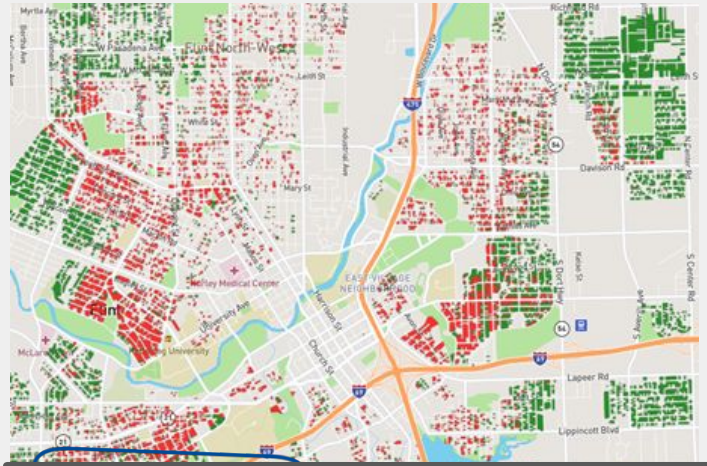
- When ingested, lead is poisonous to humans
  - Children are particularly vulnerable
- Lead water pipes were widely used in U.S.
  - Banned by EPA in 1986
  - Pre-existing lead pipes still prevalent



# Why is lead hard to find?

- City records are unreliable
- Digging is expensive

# BlueConduit's Innovation



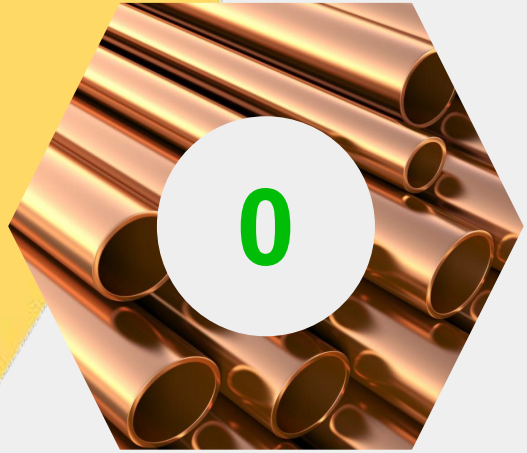

- Collected detailed data on homes in Flint, working with city and residents

| pid int64 | Property Zip Code float64 | Owner Type object | Owner State object | Homestead object | Homestead Percent float64 | HomeSEV int64 | Land Value int64 |
|---|---|---|---|---|---|---|---|
| 4012482018 | 48503 | Private | MI | Yes | 100 | 18400 | 932 |
| 4013226009 | 48503 | Private | MI | Yes | 100 | 11800 | 420 |
| 4012476011 | 48503 | Private | FL | No | 0 | 0 | 602 |
| 4012481022 | 48503 | Private | MI | Yes | 50 | 4550 | 781 |
| 4013226025 | 48503 | Private | MI | Yes | 100 | 12800 | 510 |

# BlueConduit's Innovation

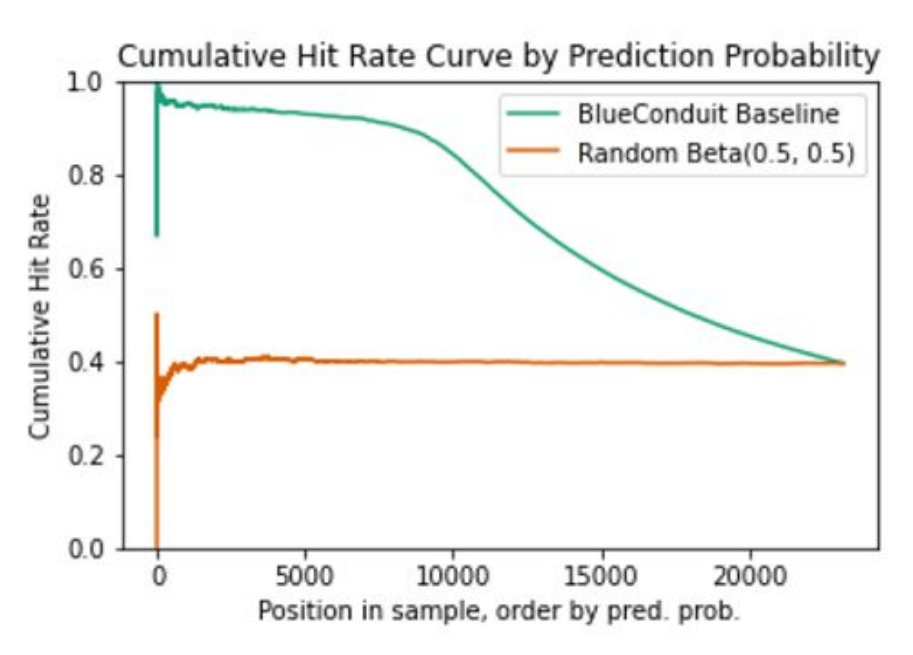
Used machine learning to predict copper/lead.

- Flint's initial digging:
  - 15% Hit Rate
- BlueConduit's digging:
  - 81% Hit Rate




# BlueConduit's Model

*XGBoost*

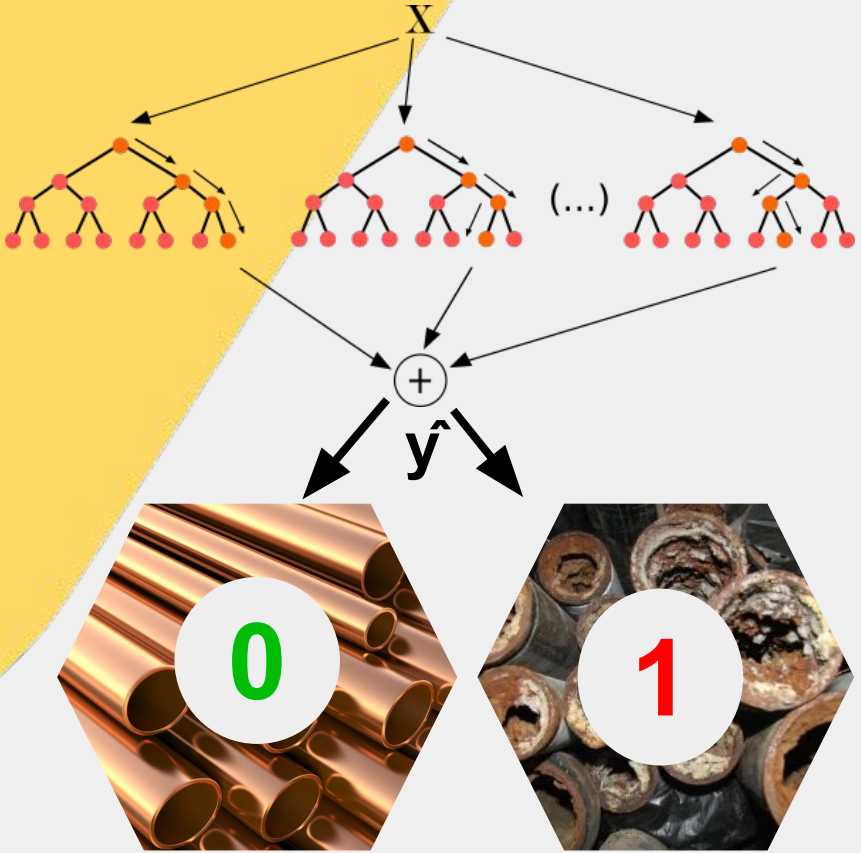



## Evaluation:

- Hit Rate Curve
- City Savings

# BlueConduit's Model

## XGBoost

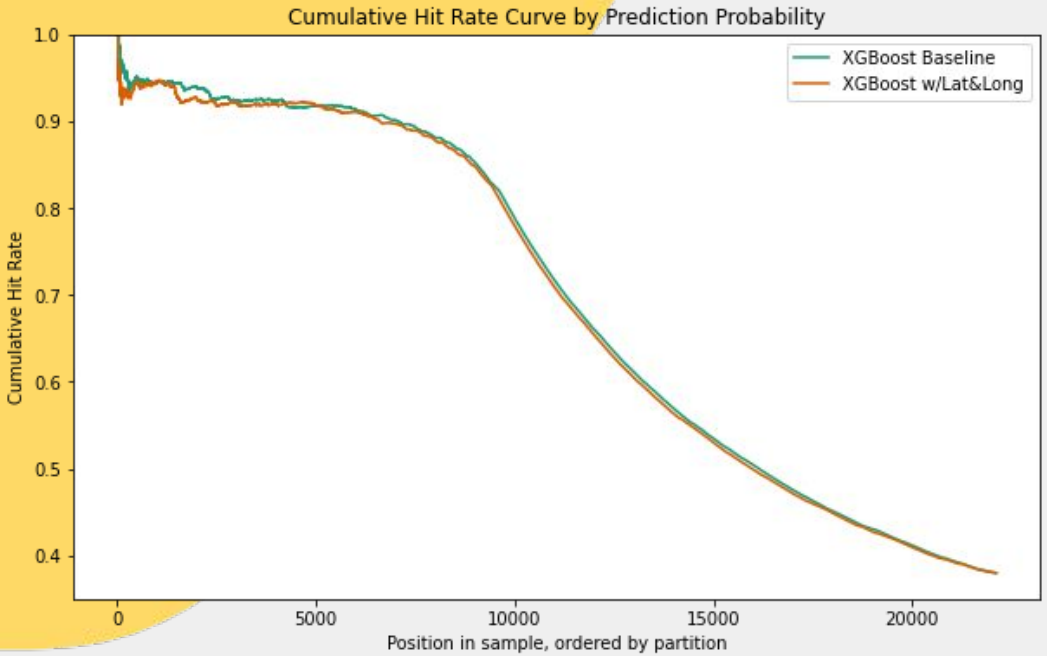



Important Features

- Age of home
- City record of copper
- Hydrant type
- Home value

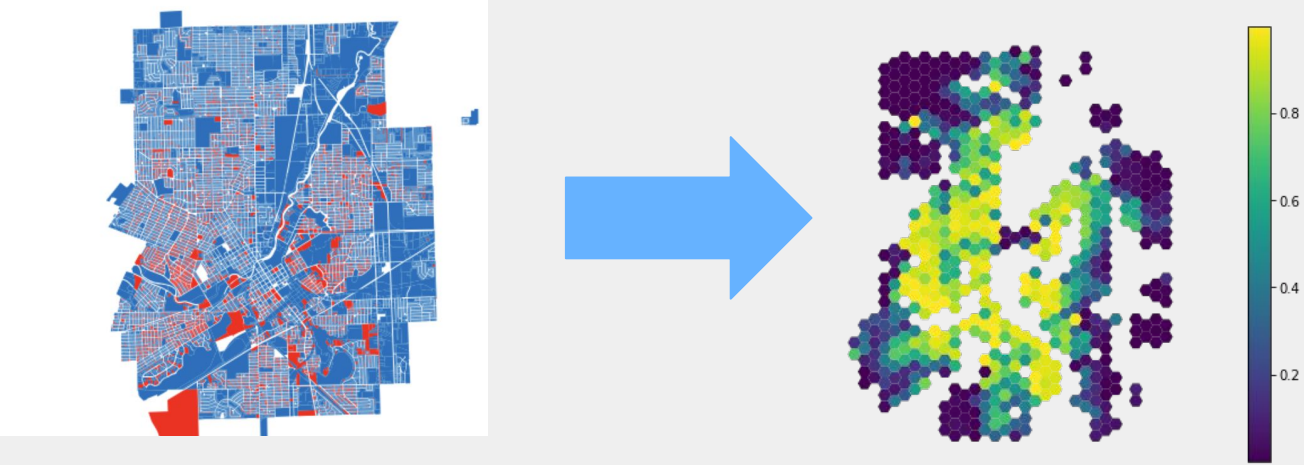
# Our Project: Problem

- BlueConduit's model **does not use** spatial info.
- Adding lat / longitude features does not help their model



# Our Project: Motivation

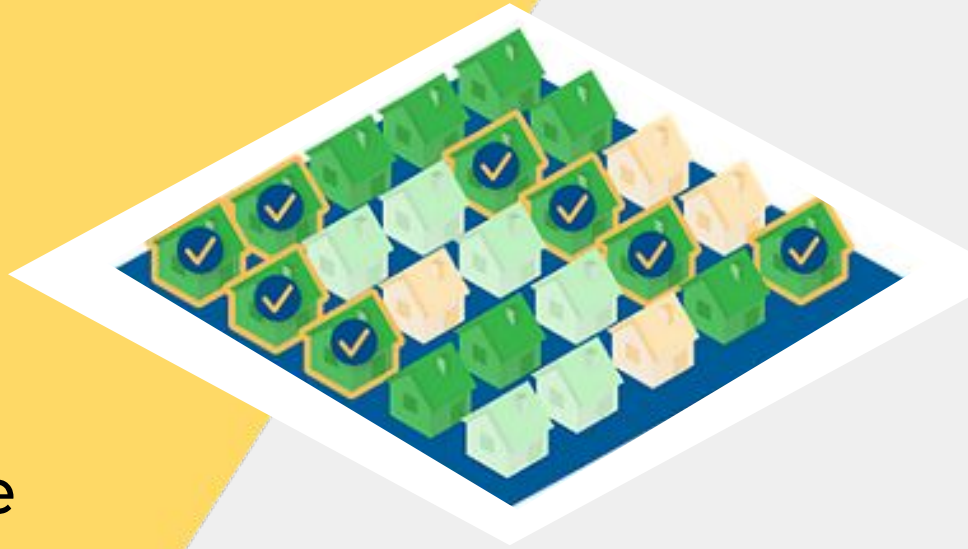
- Lead prevalence varies greatly by neighborhood.
- Home locations encode info about their construction, development, and materials.
- Neighborhood info should help model in *some way*.



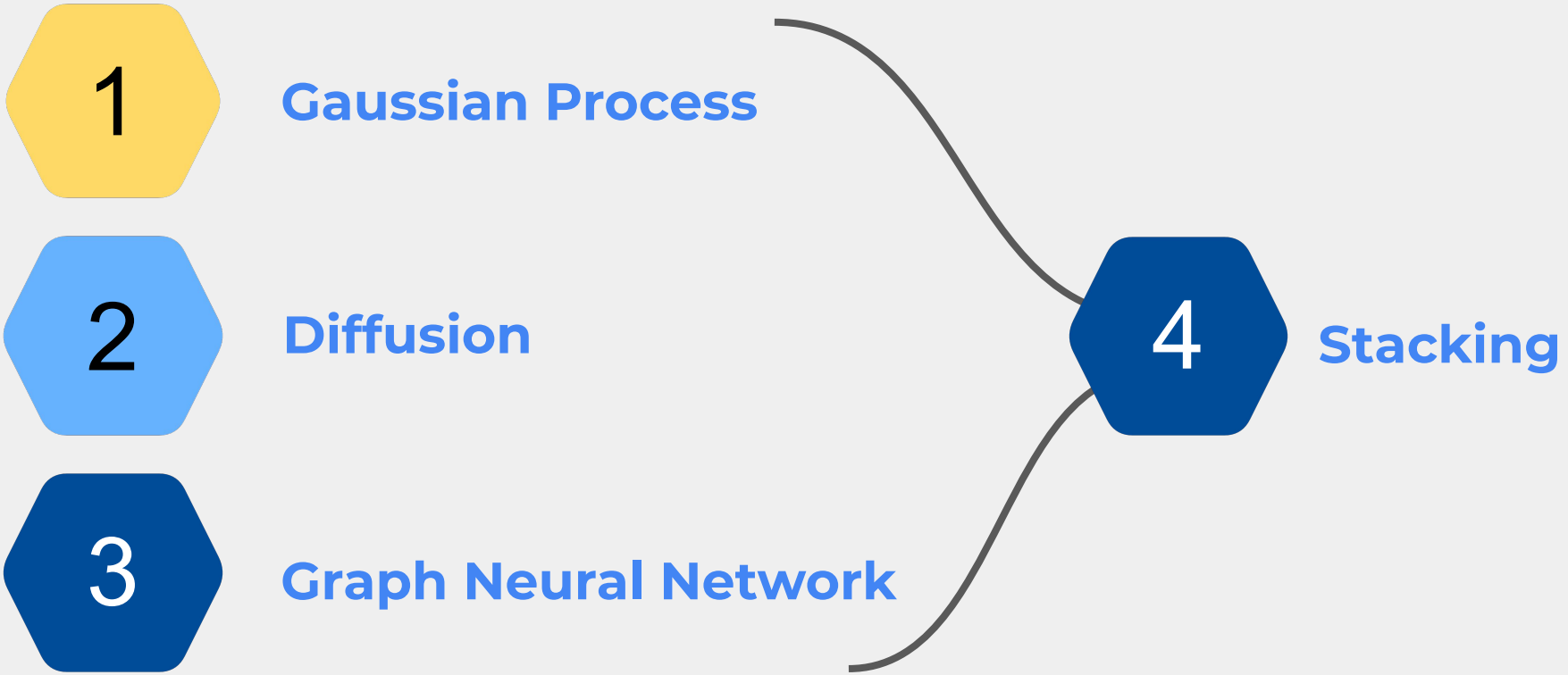
# Our Task

Explore whether using spatial information **can improve** BlueConduit's model.

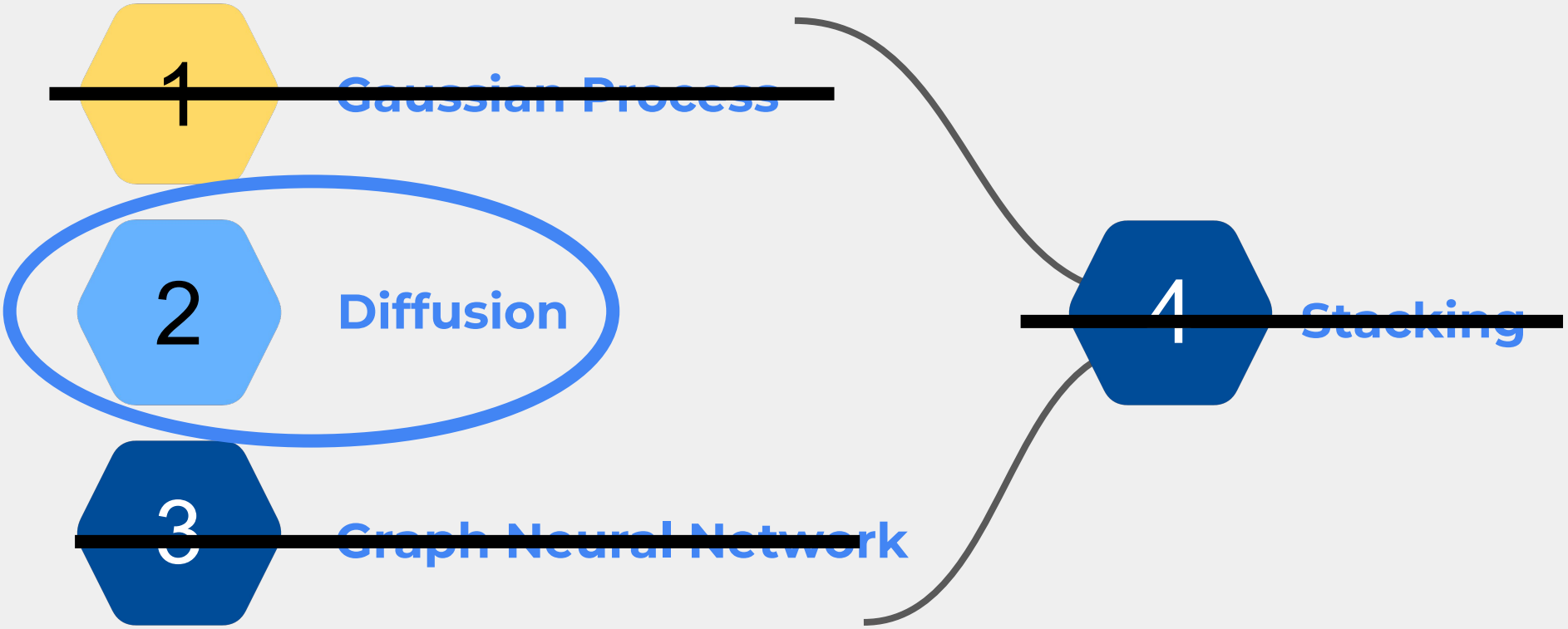
- Evaluate using hit rate curve and city savings for Flint dataset



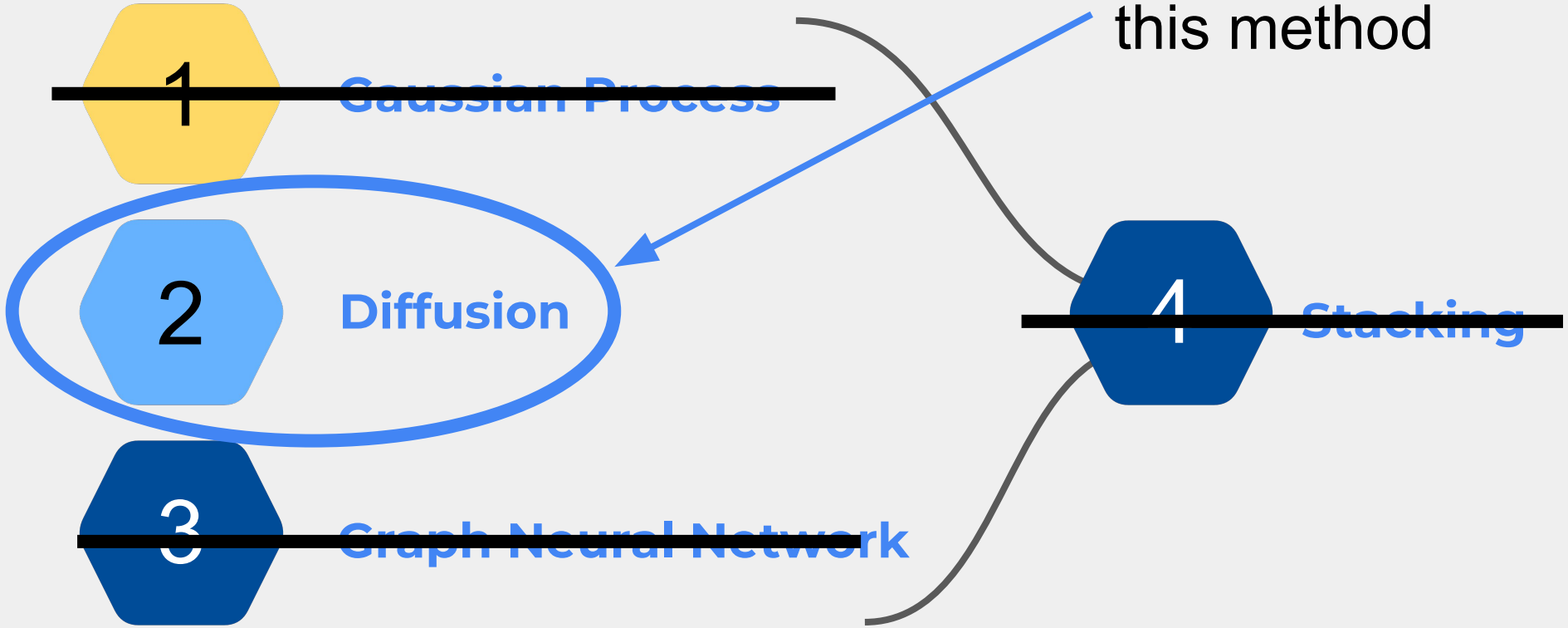
# What we tried



# What worked

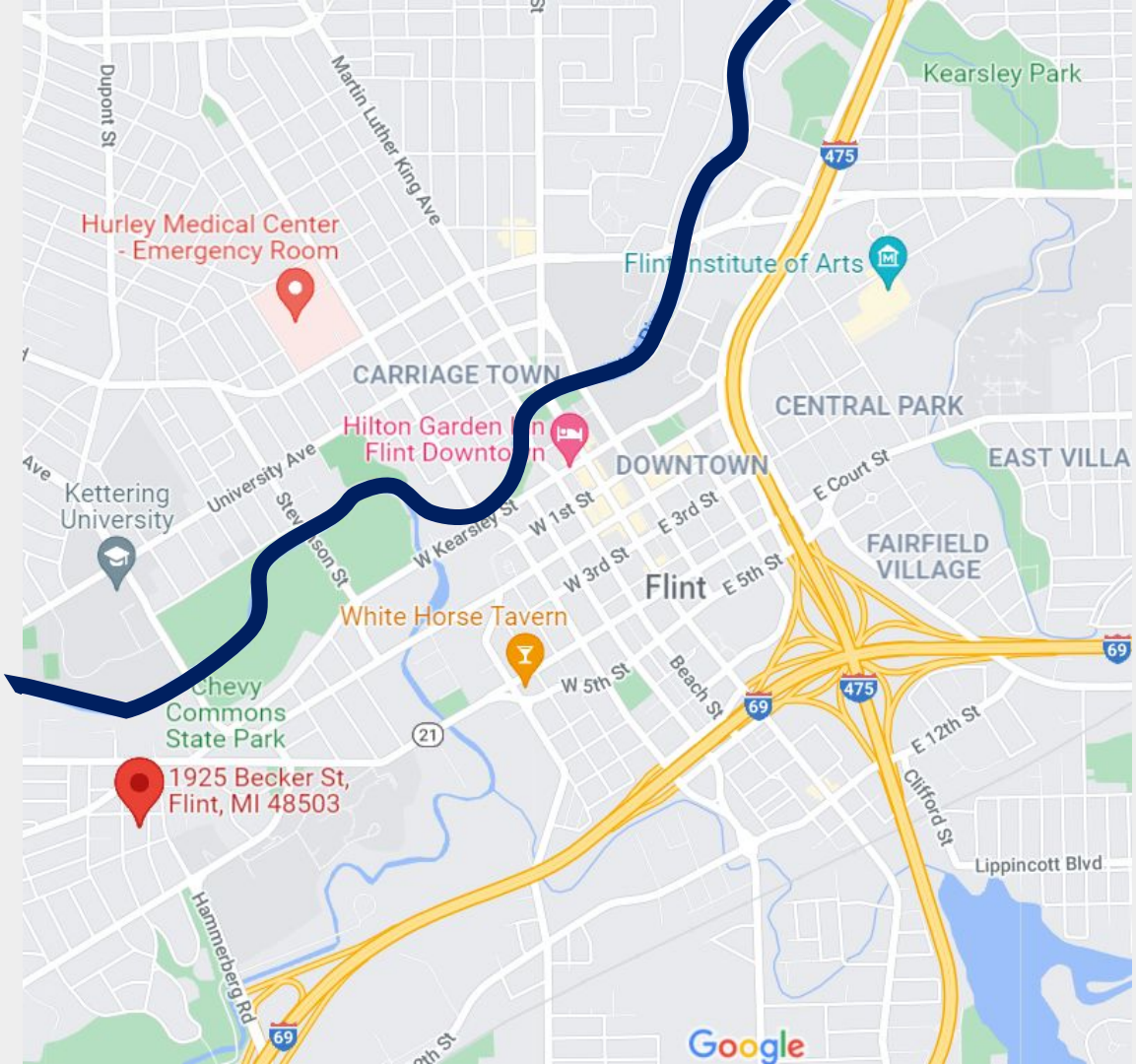

1
Gaussian Process
2
Diffusion
3
Graph Neural Network
4
Stacking

# What worked

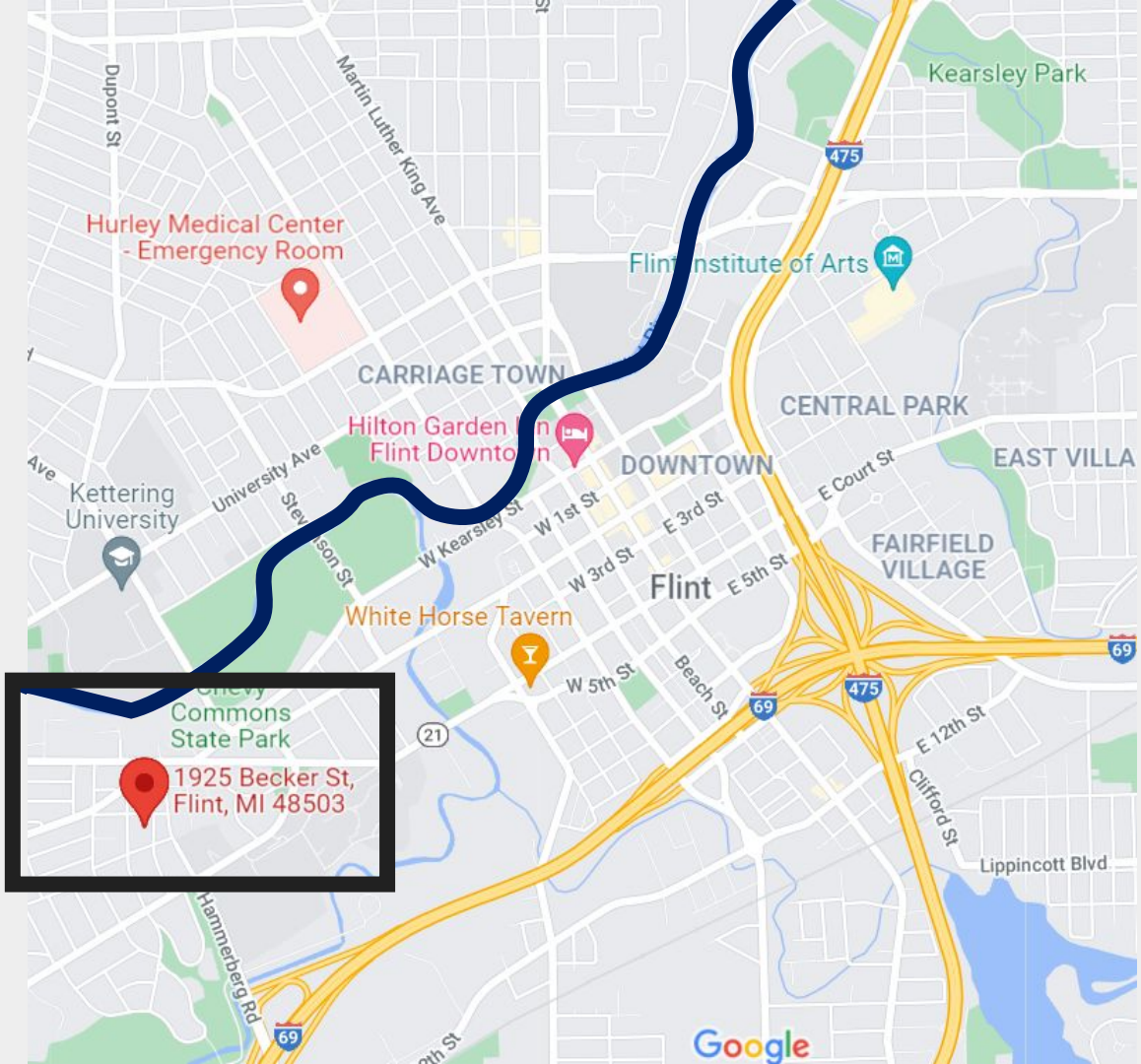

Let’s explore this method
1
Gaussian Process
2
Diffusion
3
Graph Neural Network
4
Stacking

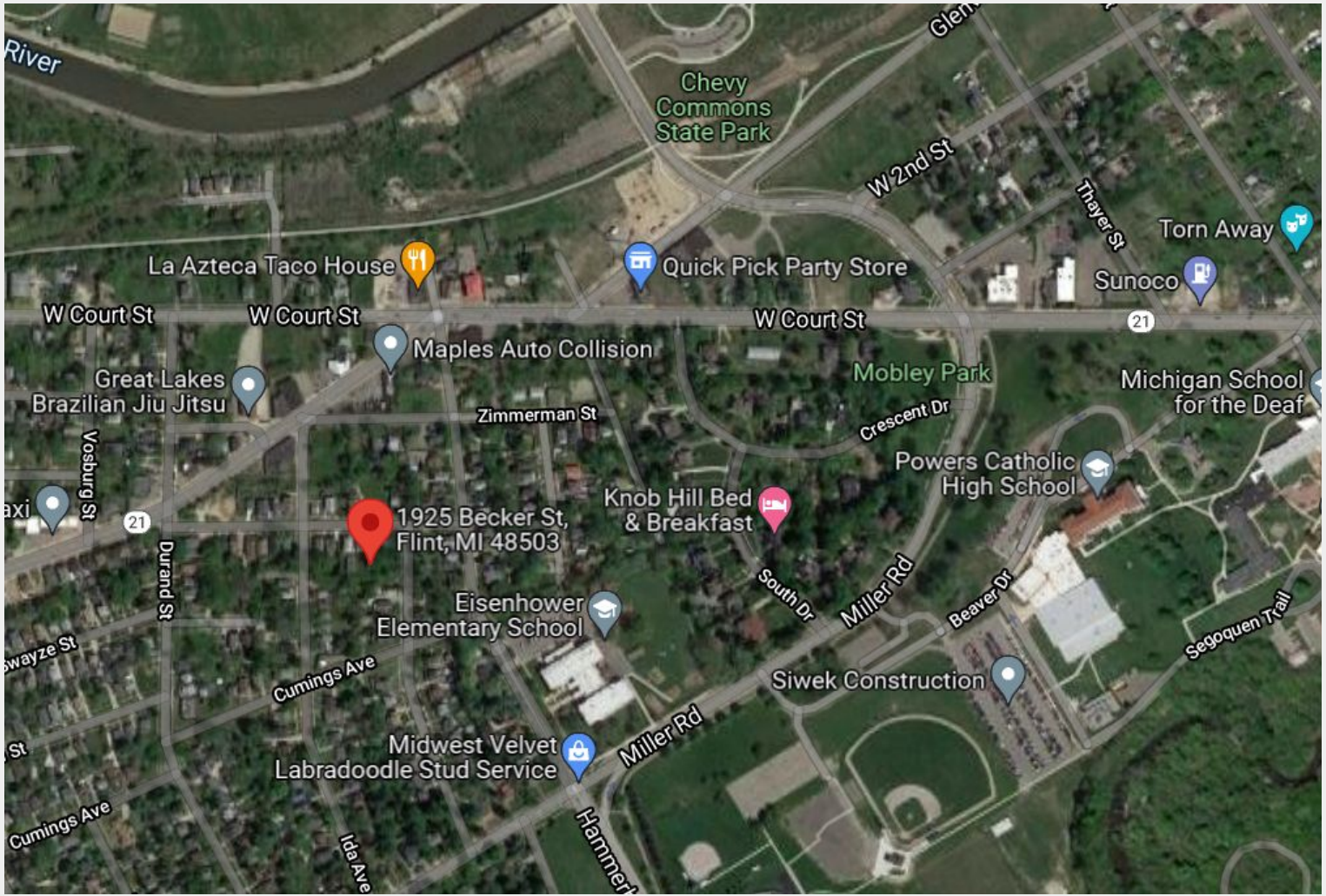
# Case Study:
# 1925 Becker St.

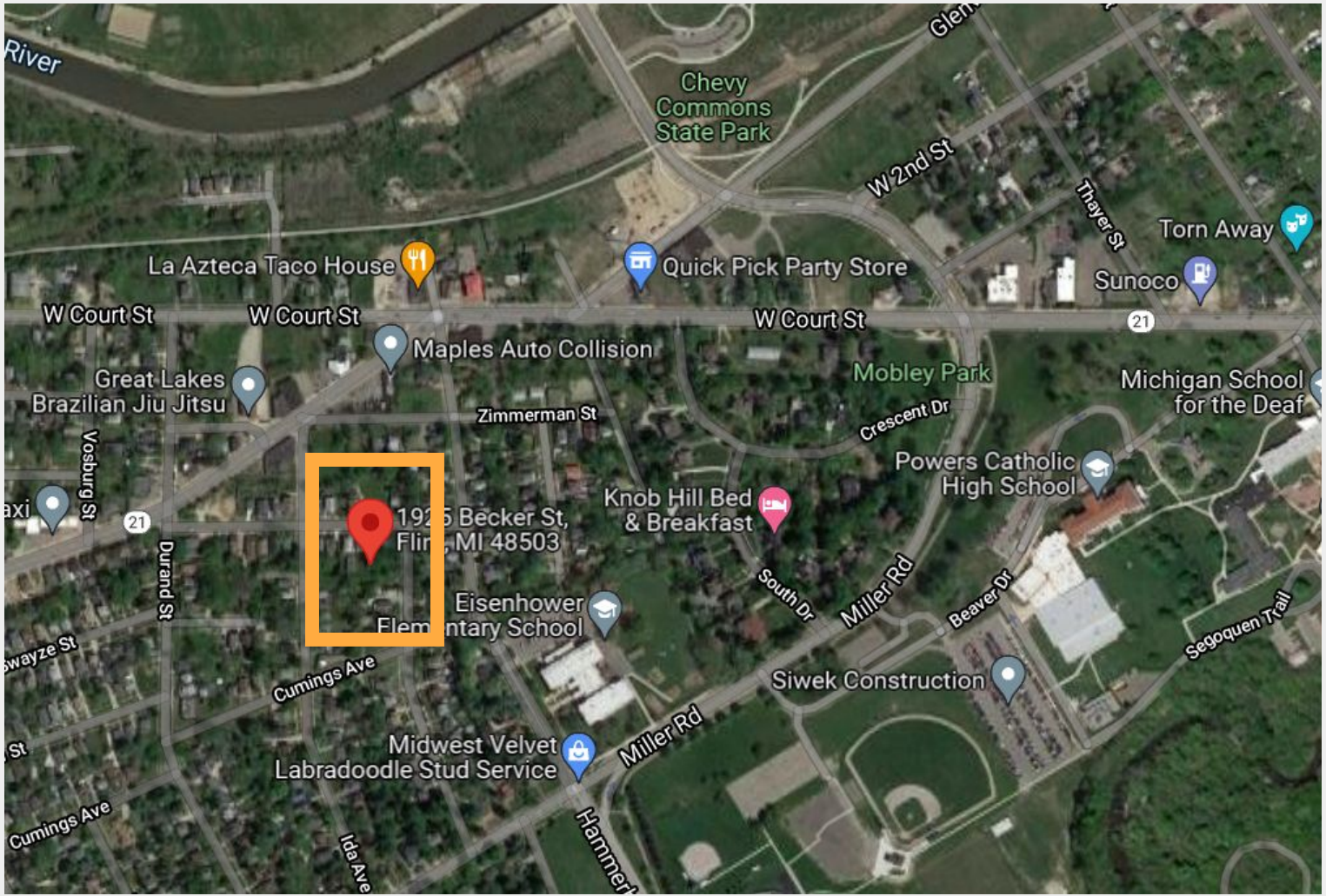
# Flint River

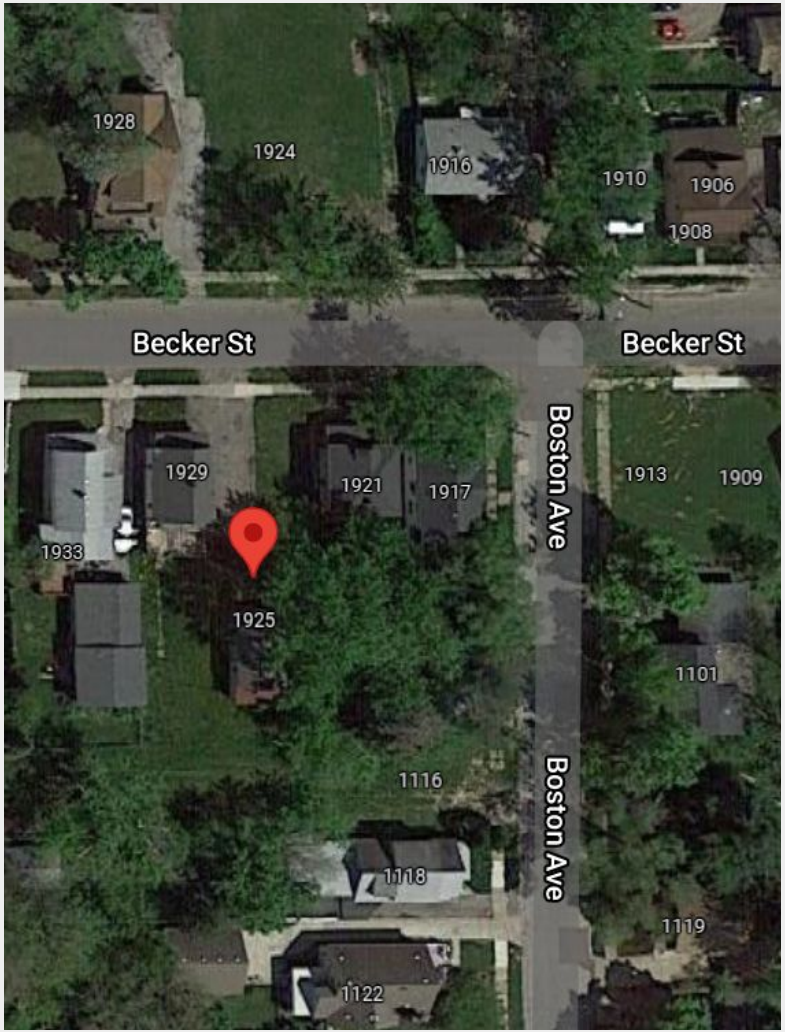

Kearsley Park
Dupont St
Martin Luther King Ave
475
Hurley Medical Center
- Emergency Room
Flint Institute of Arts
CARRIAGE TOWN
CENTRAL PARK
Hilton Garden Inn
Flint Downtown
DOWNTOWN
EAST VILLA
E Court St
University Ave
Kettering
University
W Kearsley St
W 1st St
E 3rd St
W 3rd St
Flint
E 5th St
FAIRFIELD
VILLAGE
White Horse Tavern
W 5th St
Beach St
69
475
Chevy
Commons
State Park
21
1925 Becker St,
Flint, MI 48503
E 12th St
Clifford St
Lippincott Blvd
Hammerberg Rd
69
Google

# Flint River

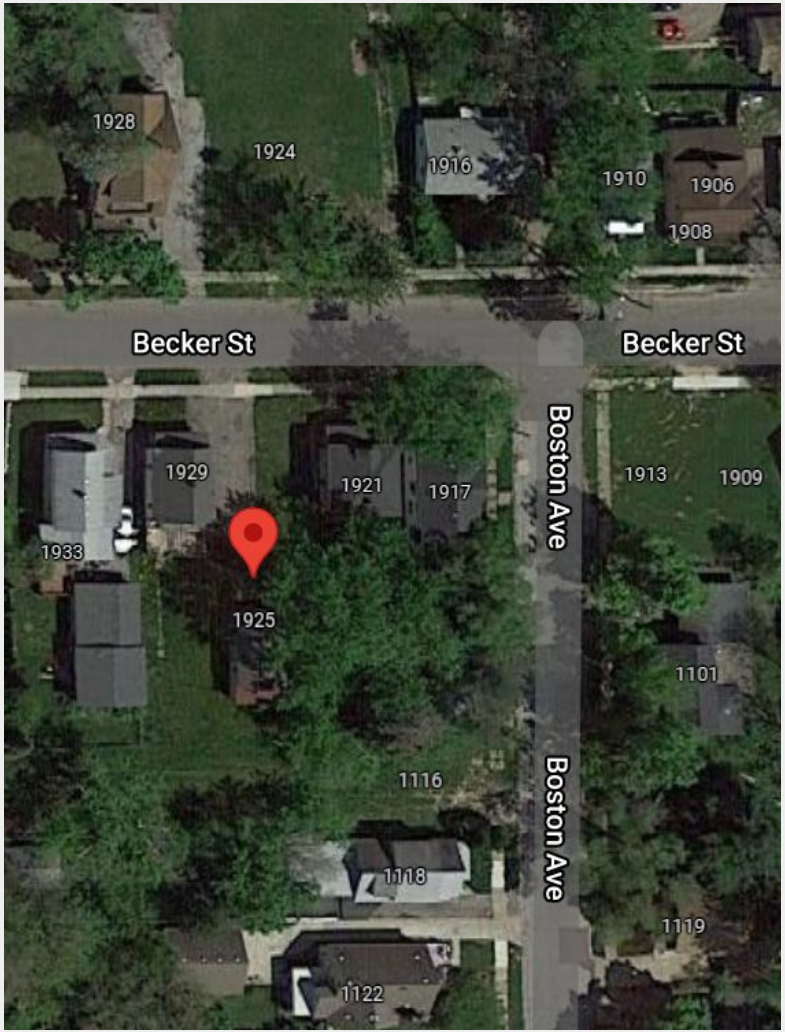

Kearsley Park
Dupont St
Martin Luther King Ave
475
Hurley Medical Center
- Emergency Room
Flint Institute of Arts
CARRIAGE TOWN
CENTRAL PARK
Hilton Garden Inn
Flint Downtown
DOWNTOWN
EAST VILLA
E Court St
University Ave
Kettering
University
W Kearsley St
W 1st St
E 3rd St
W 3rd St
Flint
E 5th St
FAIRFIELD
VILLAGE
White Horse Tavern
W 5th St
Beach St
69
475
E 12th St
Clifford St
Lippincott Blvd
Commons
State Park
1925 Becker St,
Flint, MI 48503
21
Hammerberg Rd
69
Google


River
Chevy Commons State Park
W 2nd St
Thayer St
Torn Away
Sunoco
La Azteca Taco House
Quick Pick Party Store
W Court St
W Court St
W Court St
21
Maples Auto Collision
Great Lakes Brazilian Jiu Jitsu
Mobley Park
Michigan School for the Deaf
Zimmerman St
Crescent Dr
Vosburg St
Powers Catholic High School
Knob Hill Bed & Breakfast
1925 Becker St, Flint, MI 48503
21
Durand St
South Dr
Miller Rd
Beaver Dr
Segoquen Trail
Eisenhower Elementary School
Cumings Ave
Siwek Construction
Midwest Velvet Labradoodle Stud Service
Miller Rd
Cumings Ave
Ida Ave

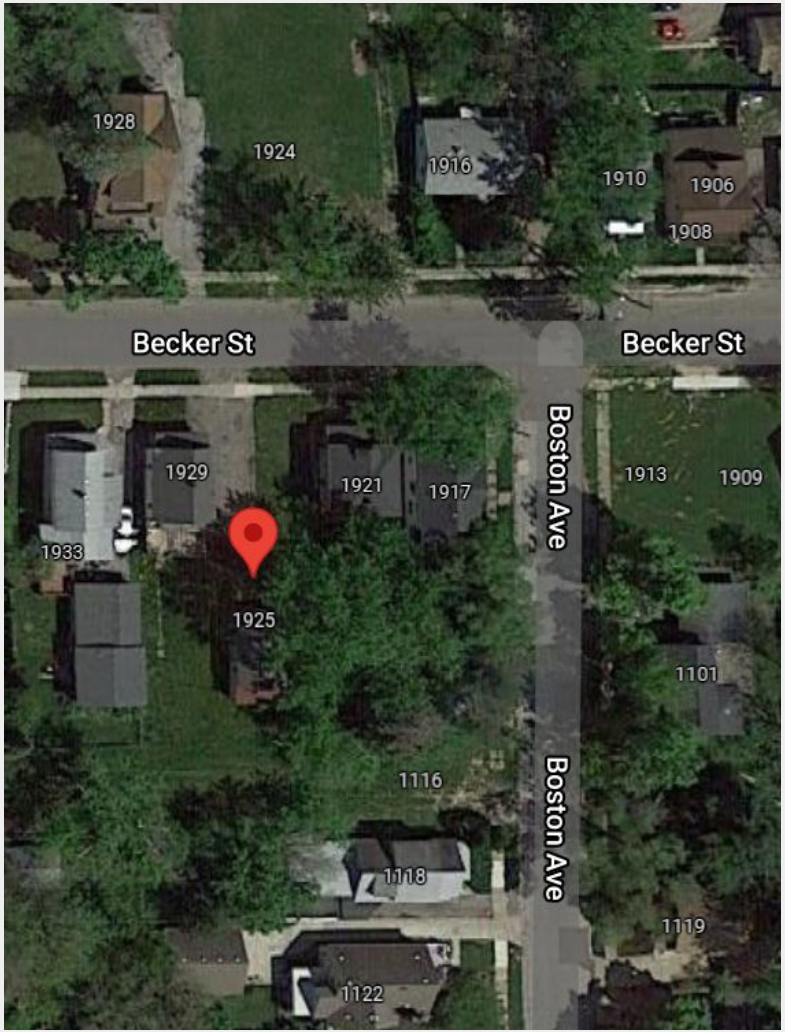

River
Chevy Commons State Park
W 2nd St
Thayer St
Torn Away
Sunoco
La Azteca Taco House
Quick Pick Party Store
W Court St
W Court St
W Court St
21
Maples Auto Collision
Great Lakes Brazilian Jiu Jitsu
Mobley Park
Michigan School for the Deaf
Zimmerman St
Crescent Dr
Vosburg St
Powers Catholic High School
Knob Hill Bed & Breakfast
21
Becker St, MI 48503
Durand St
South Dr
Miller Rd
Beaver Dr
Segoquen Trail
Eisenhower Elementary School
Cumings Ave
Siwek Construction
Midwest Velvet Labradoodle Stud Service
Miller Rd
Cumings Ave
Ida Ave



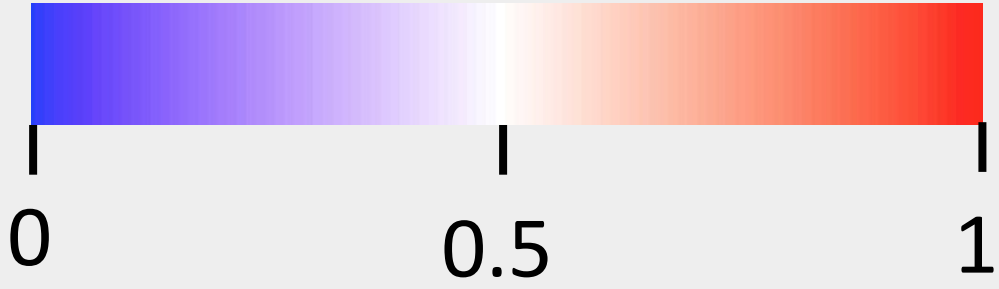

# 1925 Becker St.

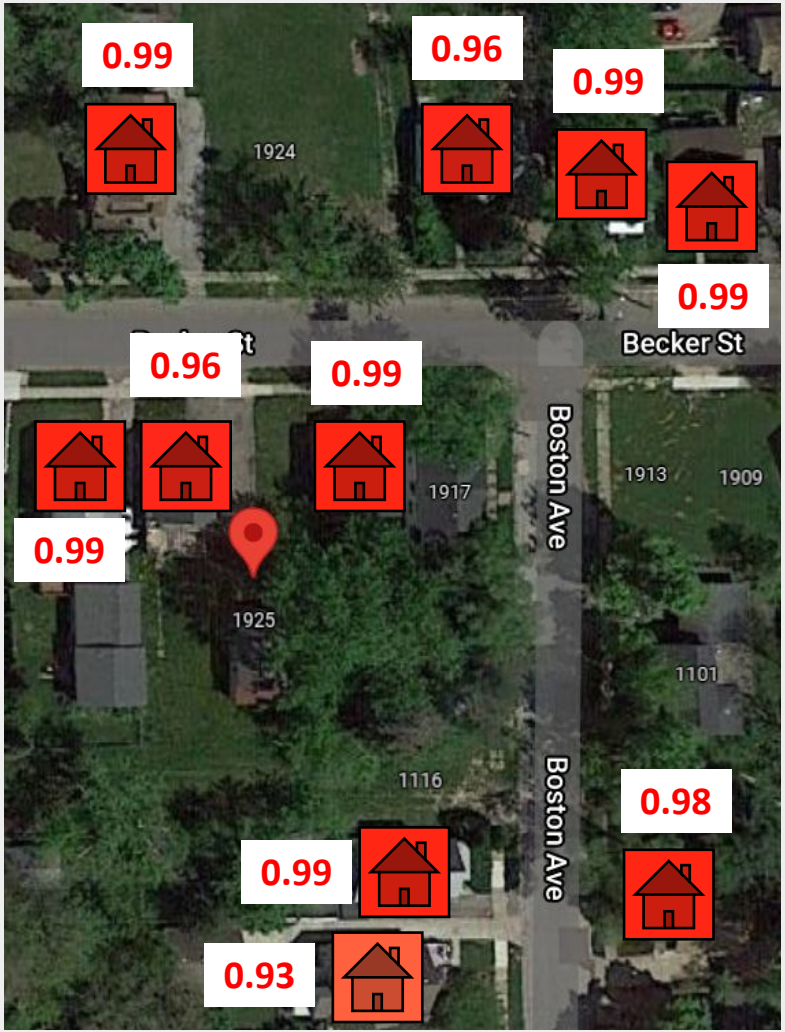



# All these homes have lead pipes

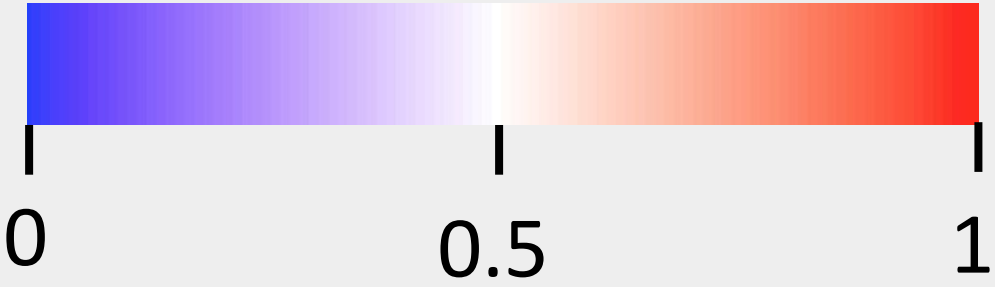


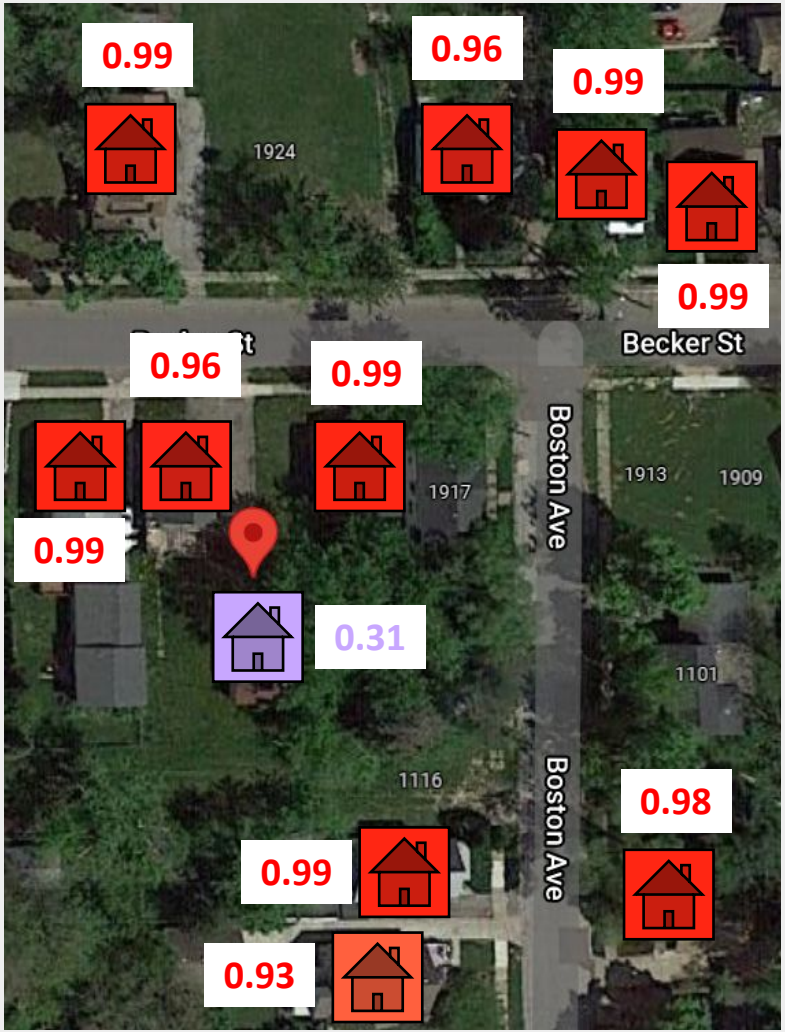

# BlueConduit's Model: Predicted Lead Probabilities



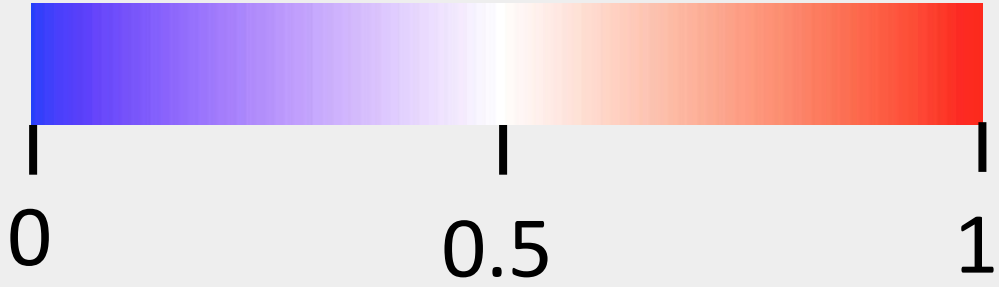


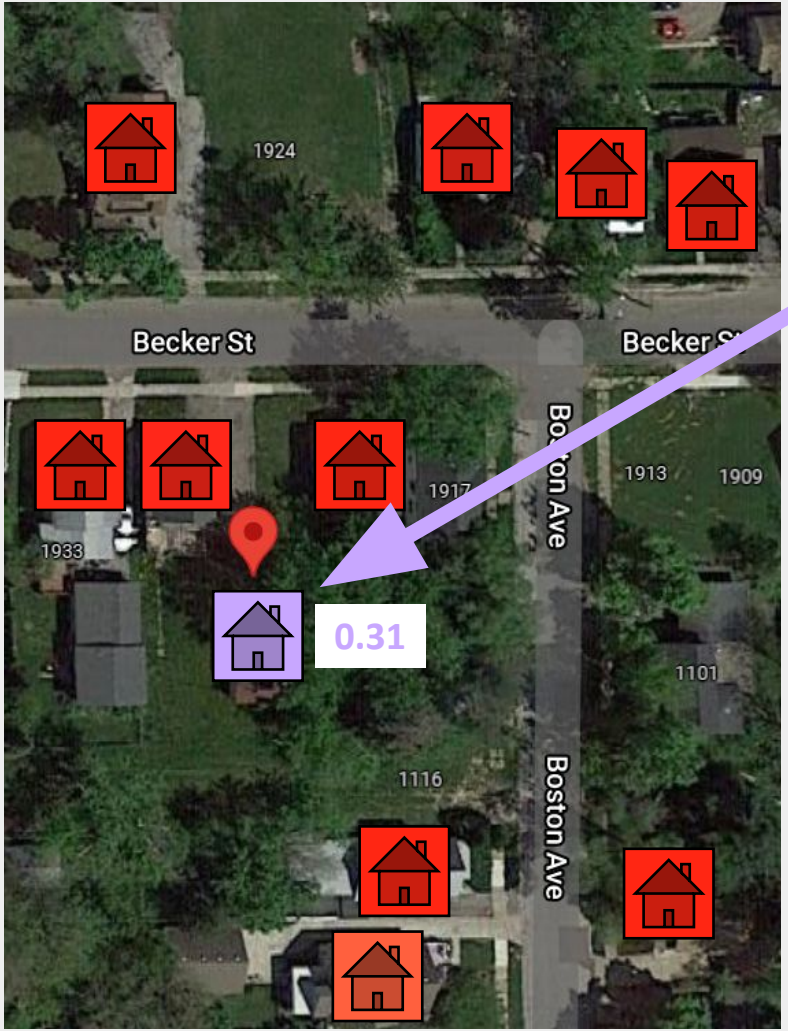

# BlueConduit's Model: Predicted Lead Probabilities



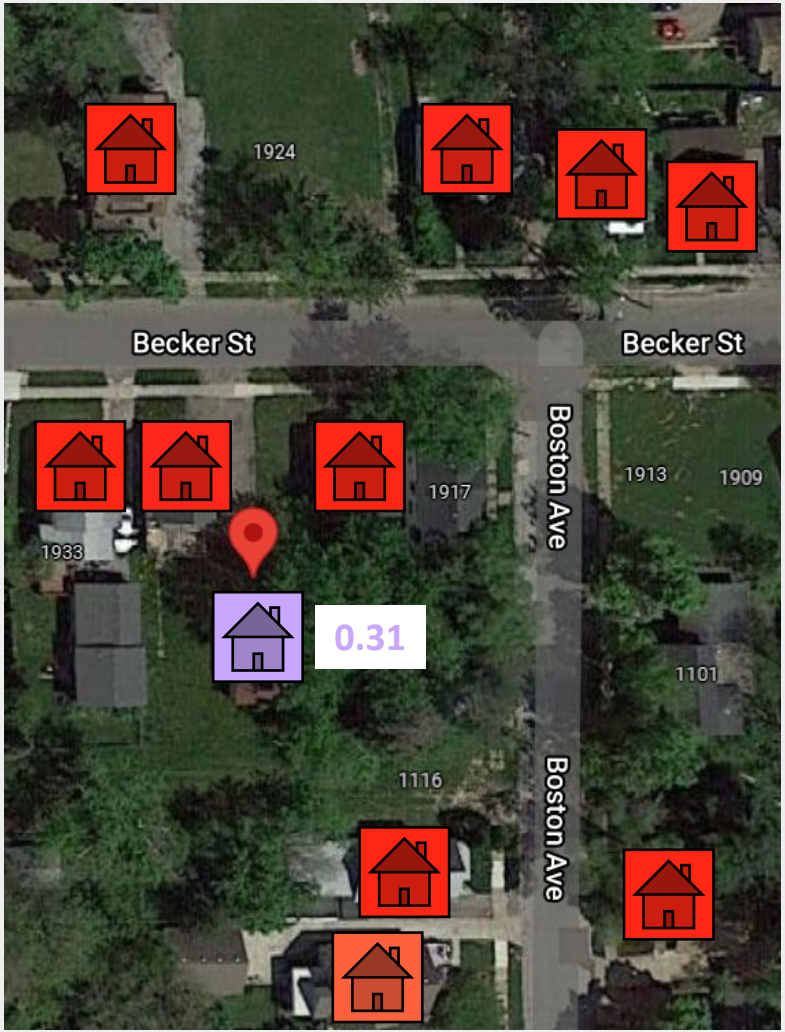



# BlueConduit's Model: Predicted Lead Probabilities





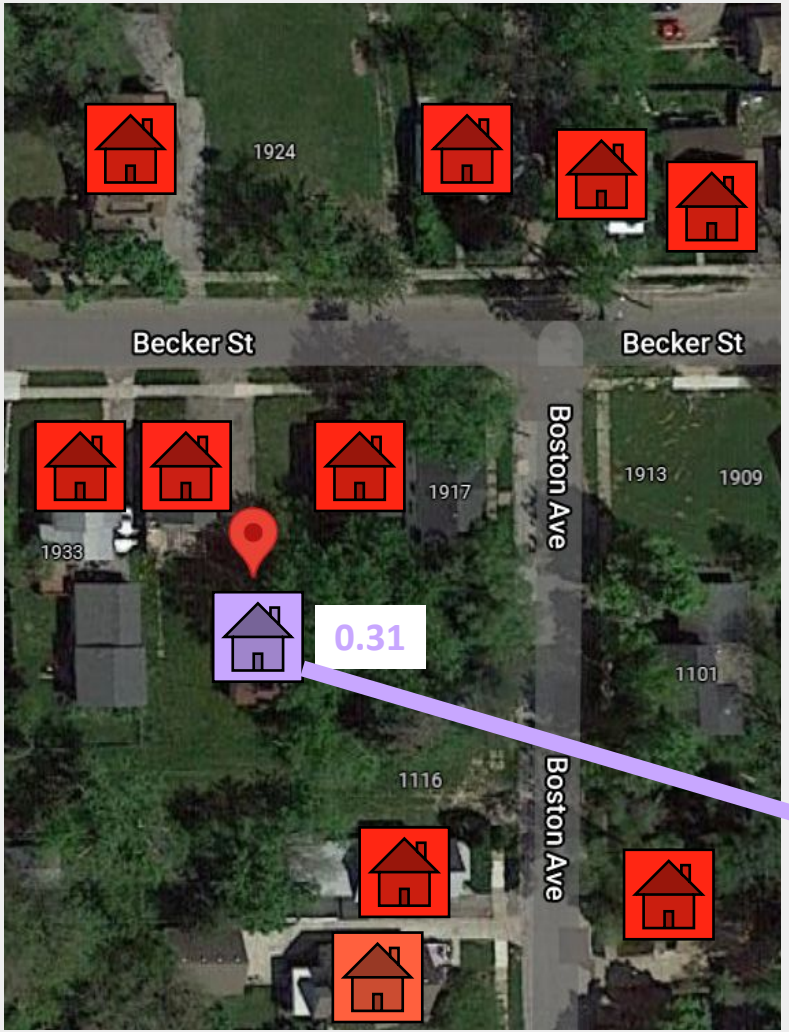

- City records inaccurately coded home as having **copper pipes.**
- Normally, copper pipe records are accurate.
- Fooled the model.




- First Dig

# Dig Queue

- Last Dig

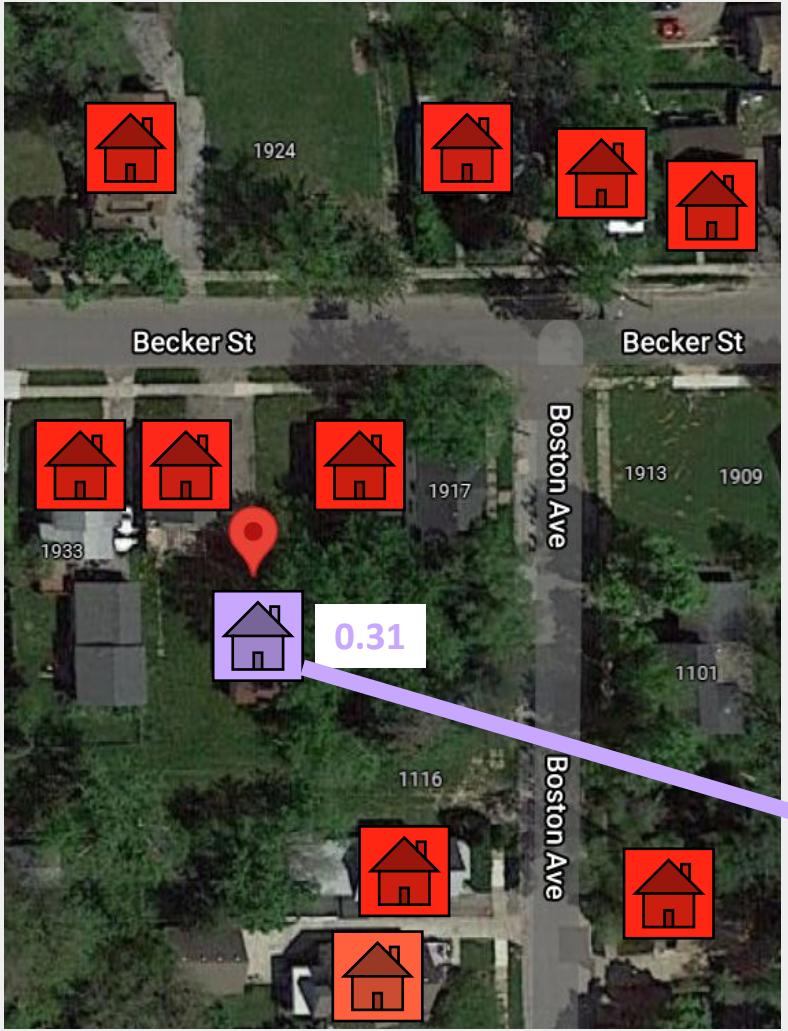



- First Dig

# Dig Queue

## Place in queue:

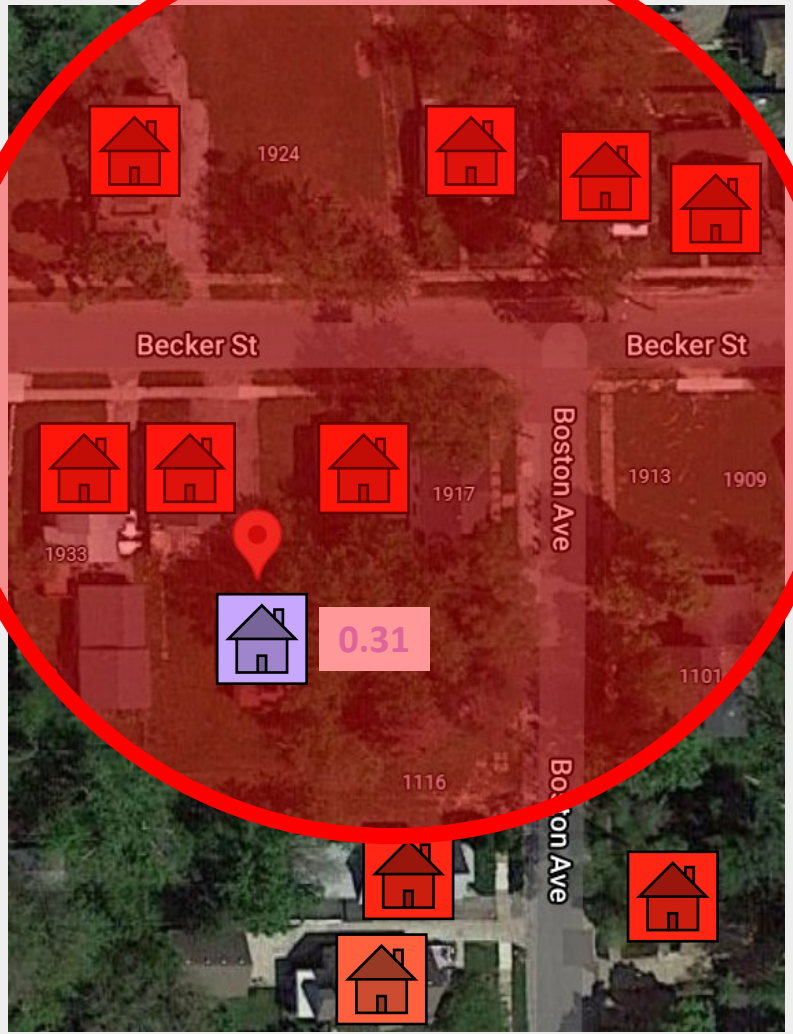
## 9,136



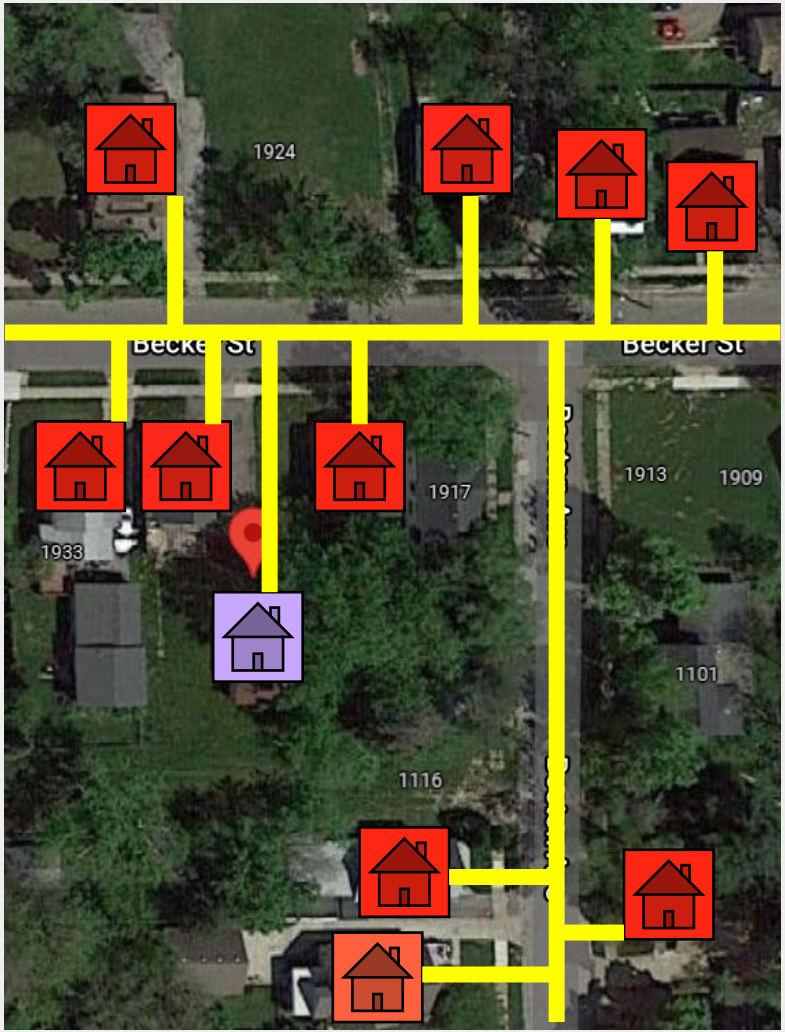

- With limited dig budget, may never get to 1925 Becker.
- Waste funds on digging non-lead homes first.

Place in queue:

<u>9,136</u>



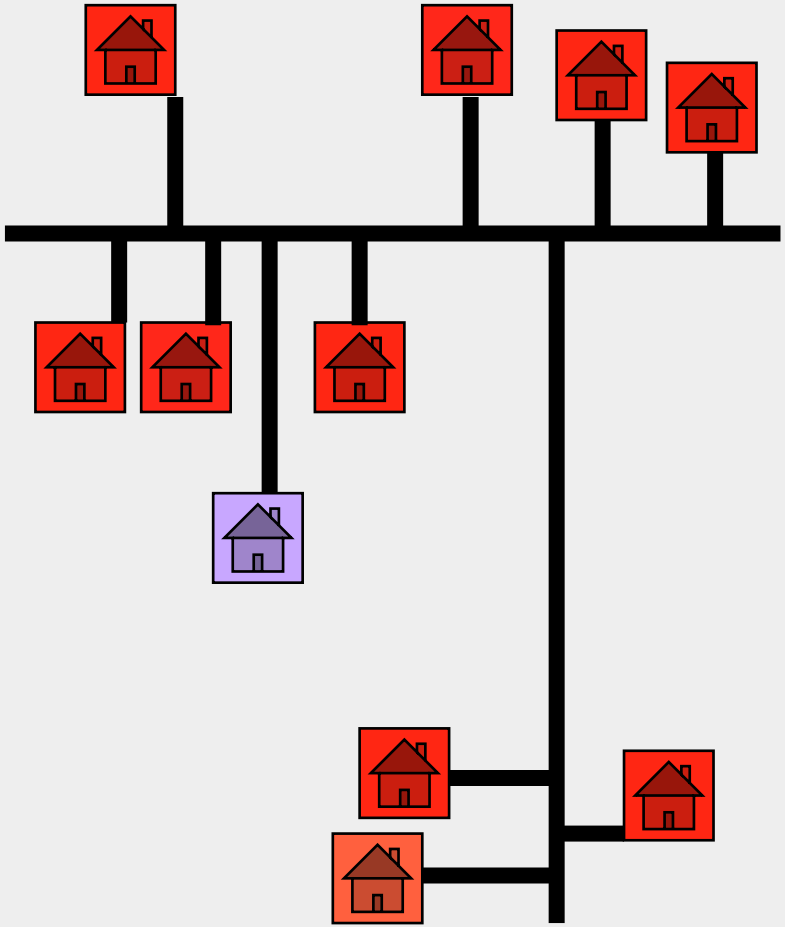

- More efficient for dig crews to dig up all homes in a single area (rather than travel around).
- We're missing a really efficient dig opportunity




# Diffusion

1. Find road distances

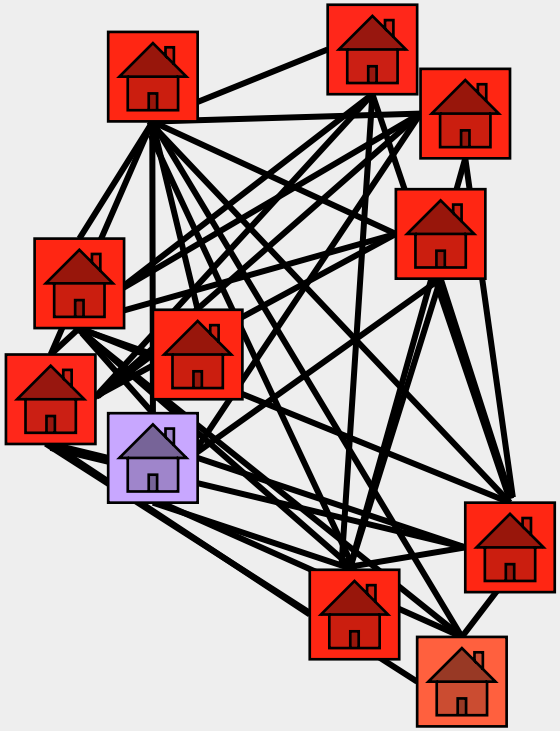
(pipes follow roads, so road distance encodes infrastructure distance)



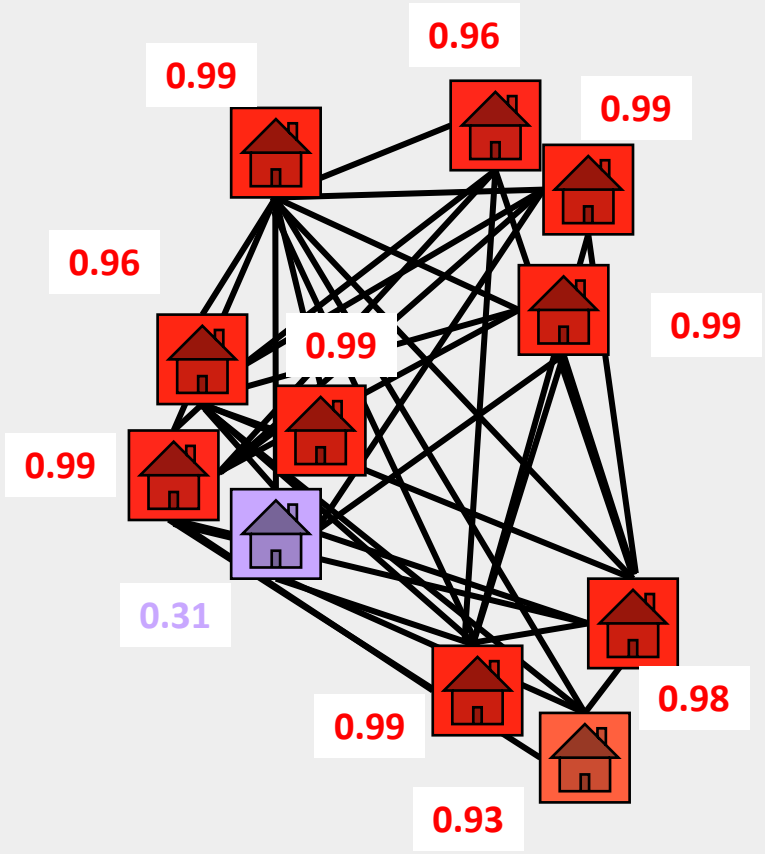
# Diffusion

2. Create a graph

# Diffusion

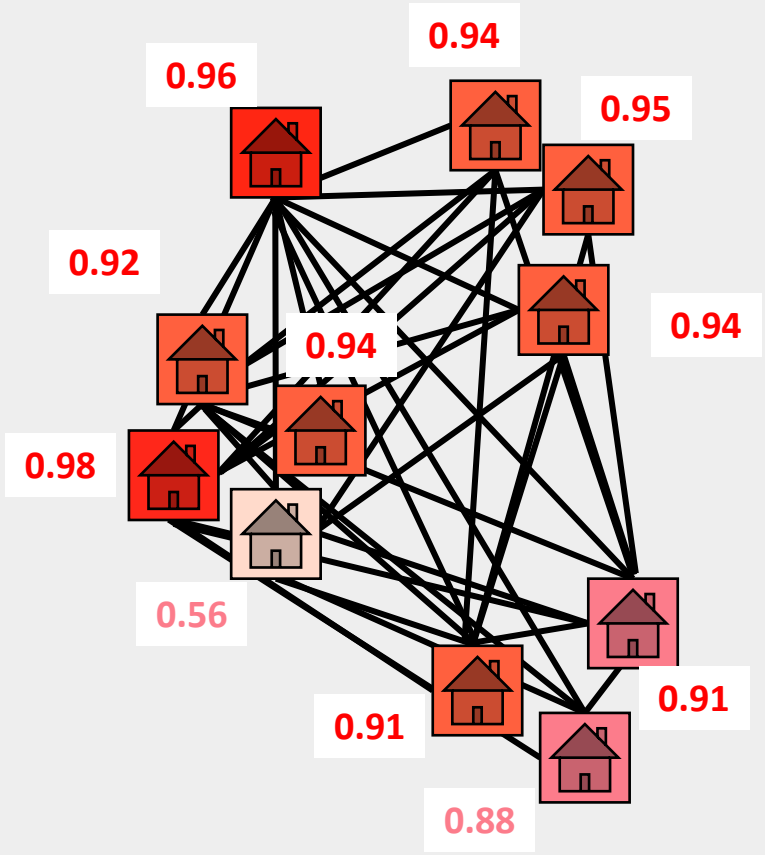


2. Create a graph




# Diffusion

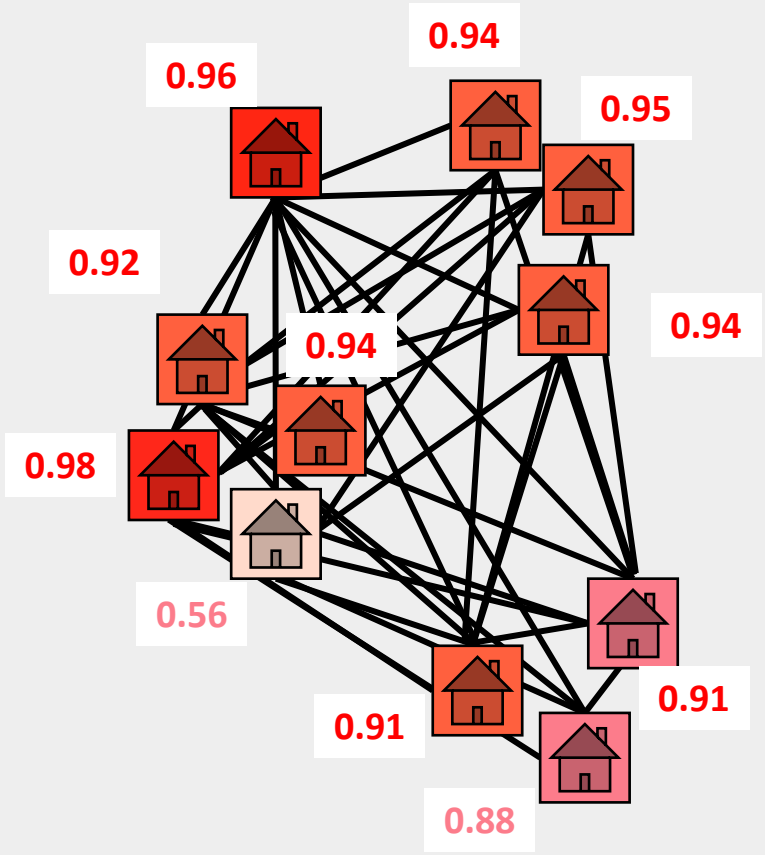
3. Take the baseline prediction probabilities and…

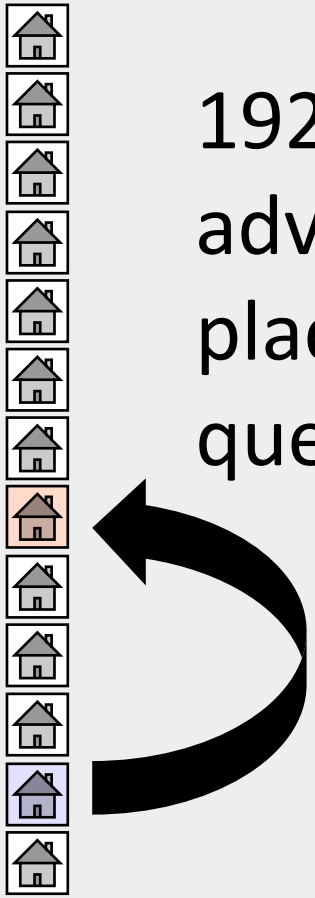



# Diffusion

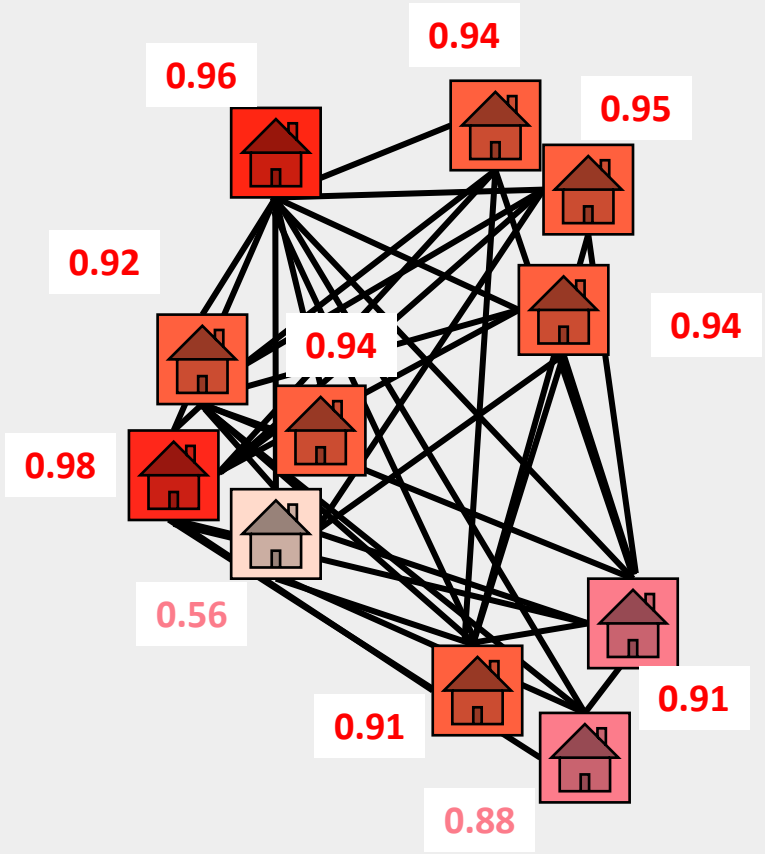
3. Take the baseline prediction probabilities and…

**Diffuse them**



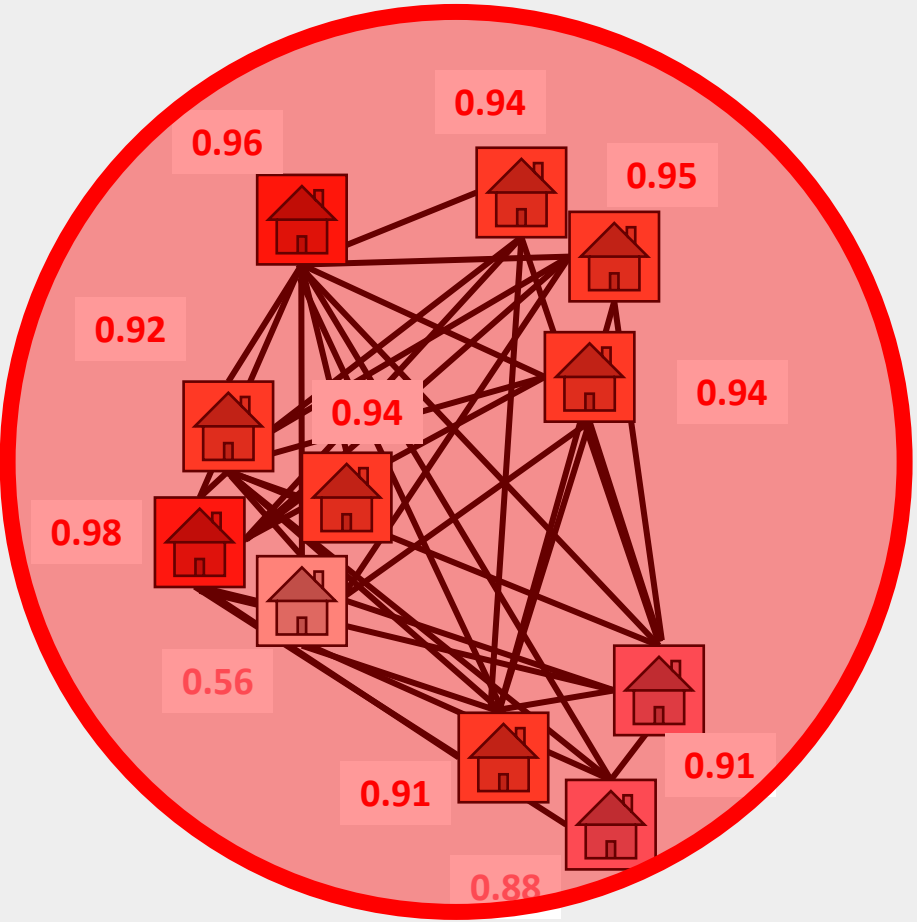
1925 Becker advanced 502 places in the dig queue!




1925 Becker advanced <u>502</u> places in the dig queue!

500 digs

→ **$1.5 million**




Dig crew digs <u>all</u> homes in the same area. Saves more **$$$** by reducing travel time/costs.

It’s not just 1925 Becker…

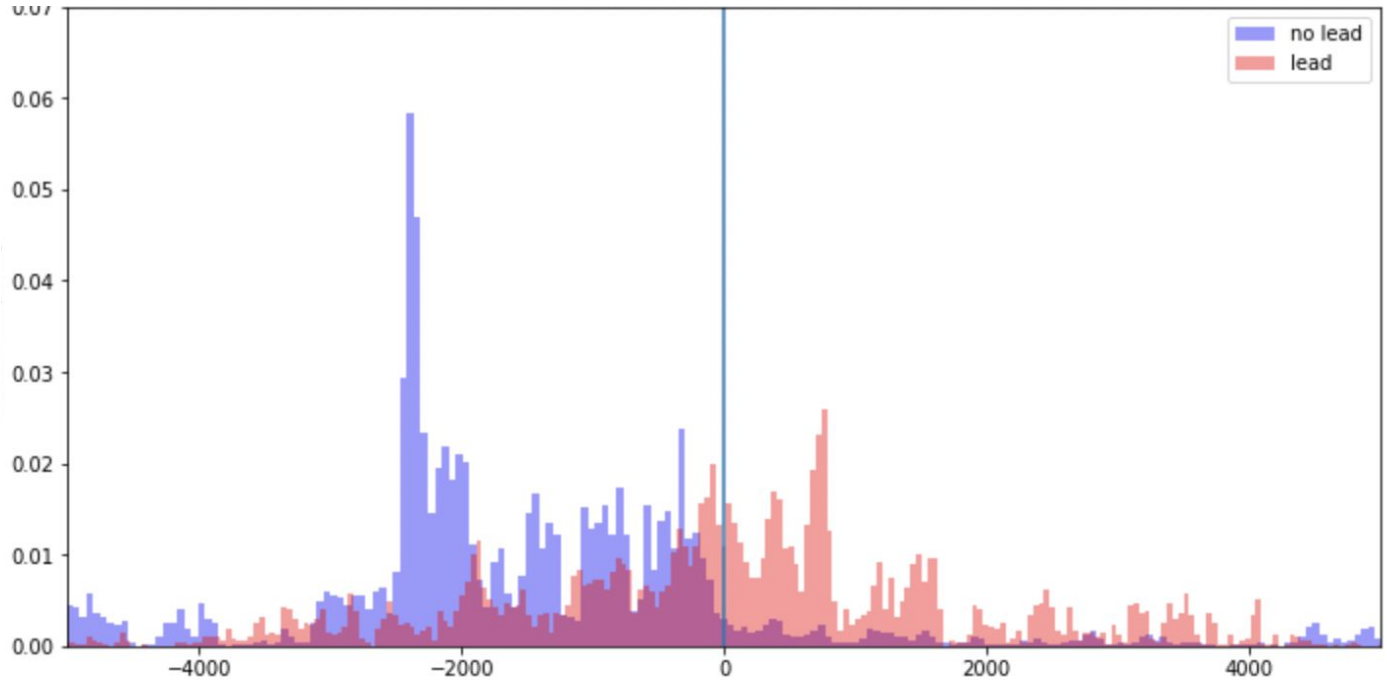
# Because of diffusion, on average …

- Lead homes **climbed 327 positions** in the dig queue.
- Non-lead homes **fell 195 positions** in the dig queue.

# Because of diffusion …

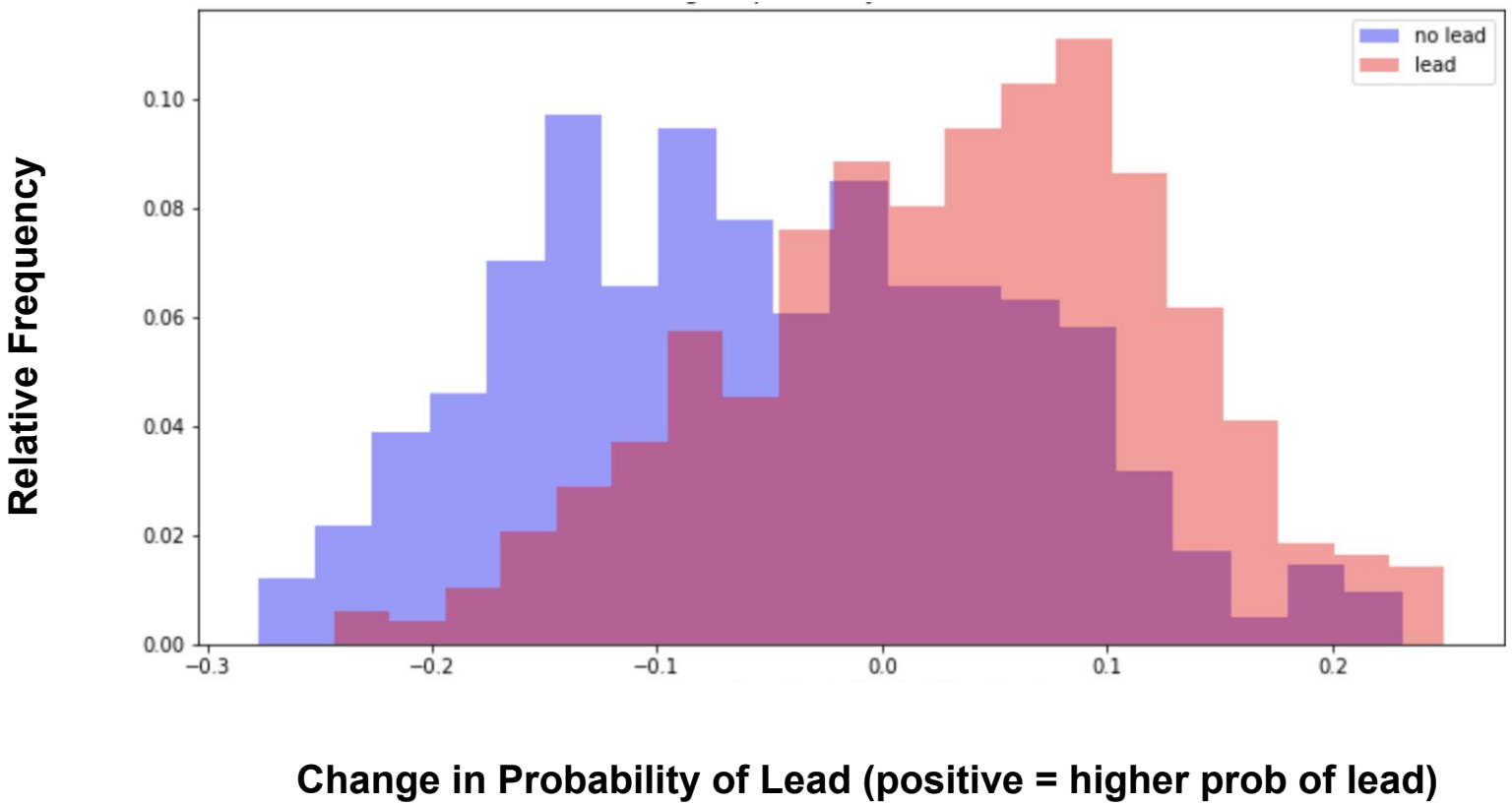
- **53%** of lead homes **climbed** in the dig queue.
- **79%** of non-lead homes **fell** in the dig queue.

# Changes in Dig Order after Diffusion



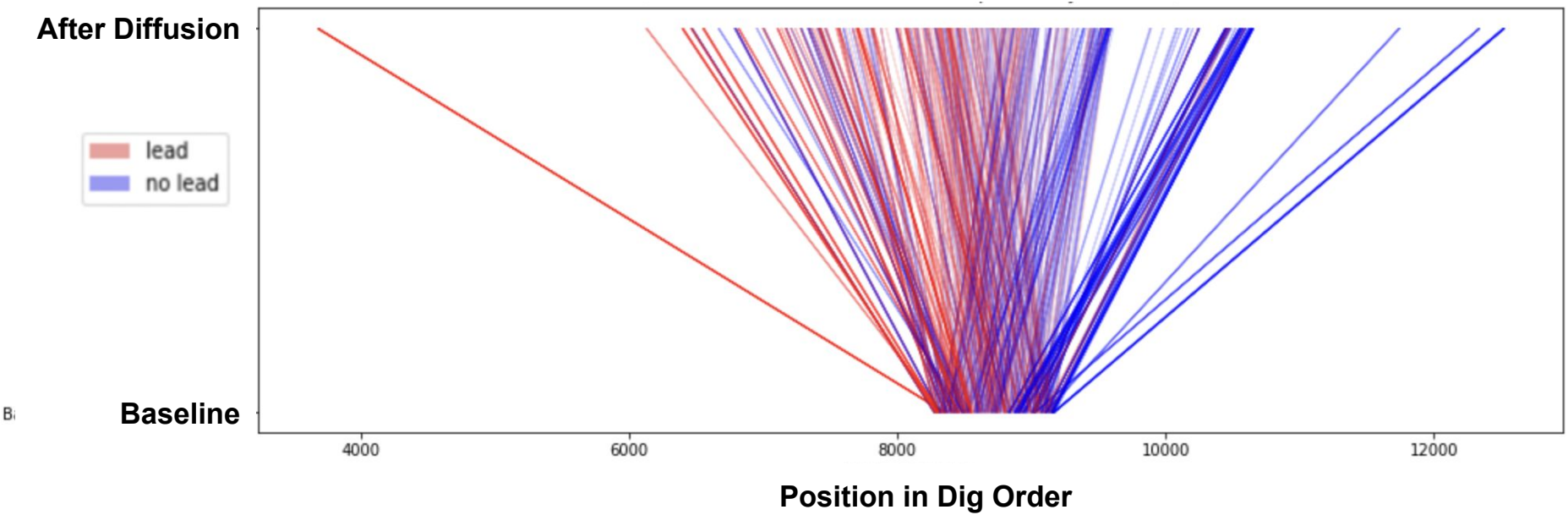
# Change in Lead Probability Among Uncertain Homes

(baseline probability 30%-70%)

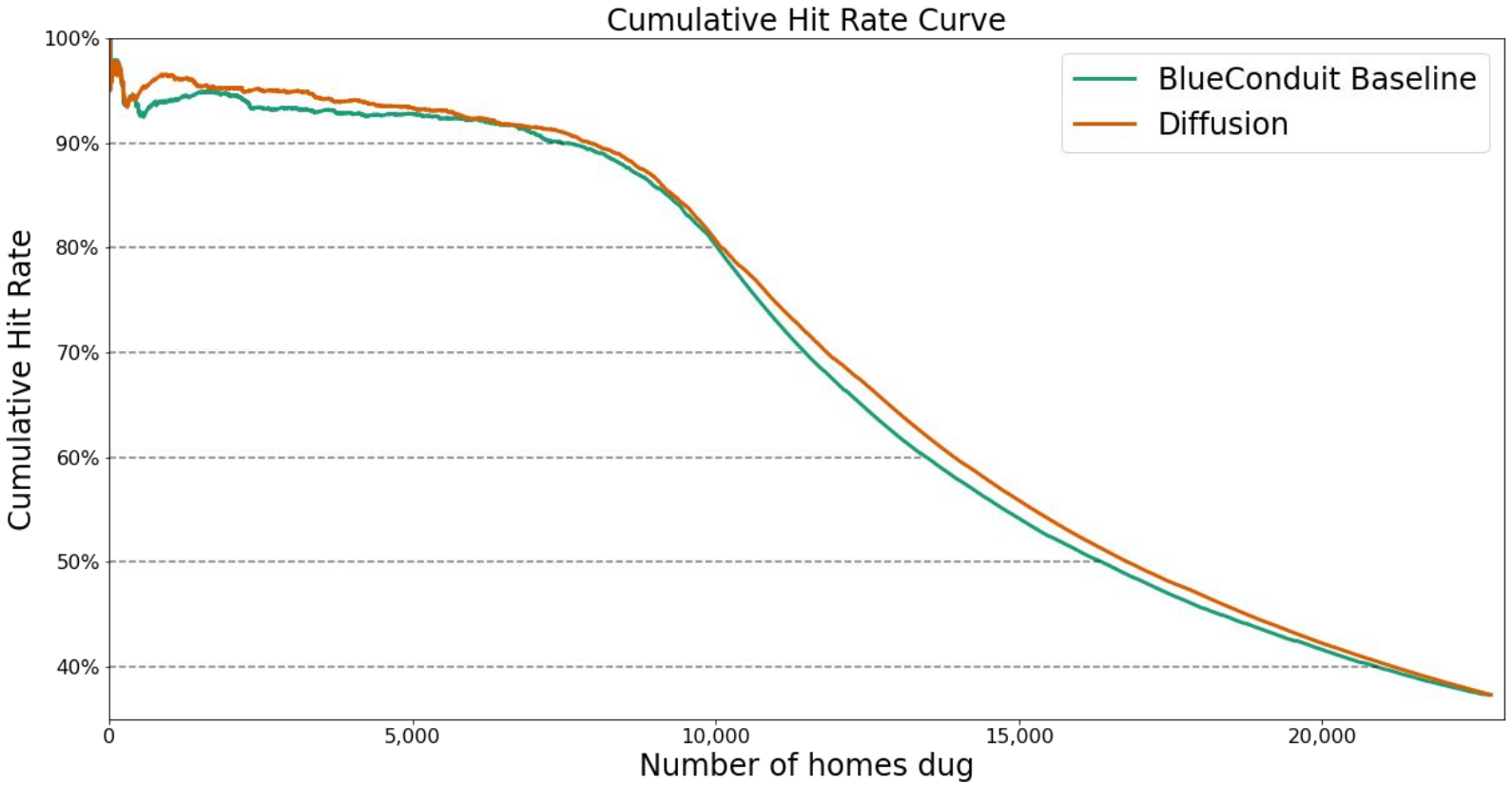


# Change in Dig Order Among Uncertain Homes

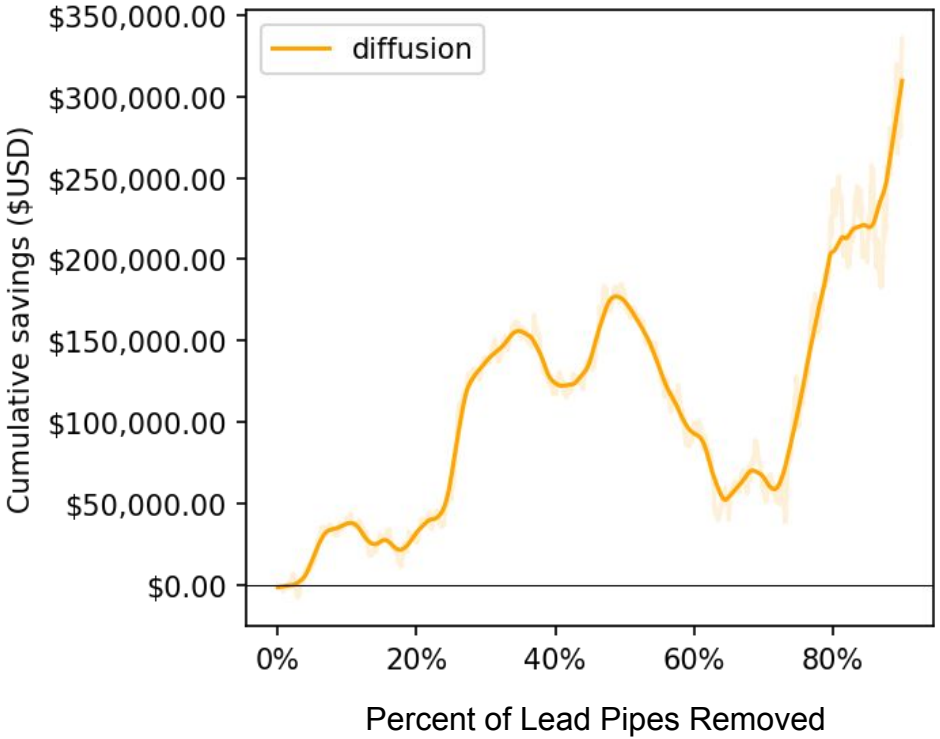
(baseline probability 30%-70%)

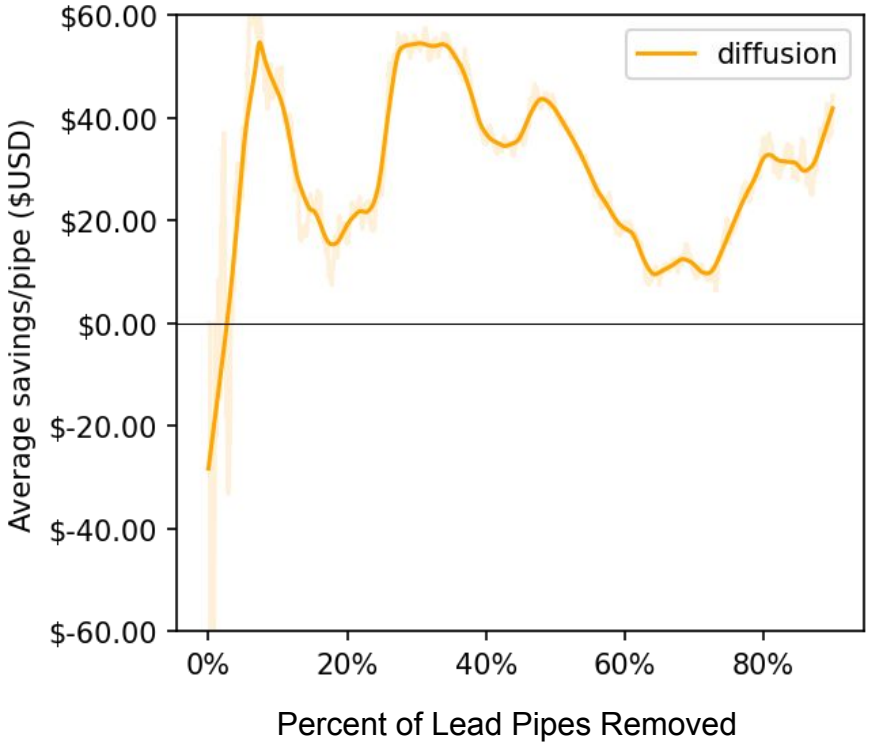


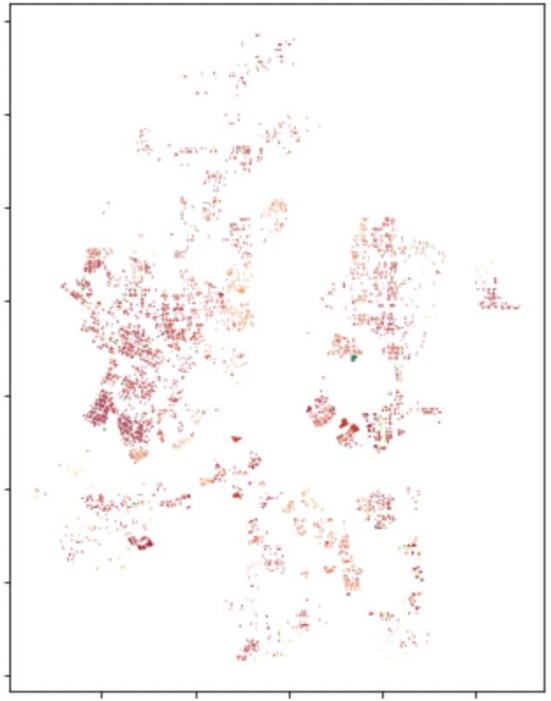
# Hit Rate Curve is Lifted



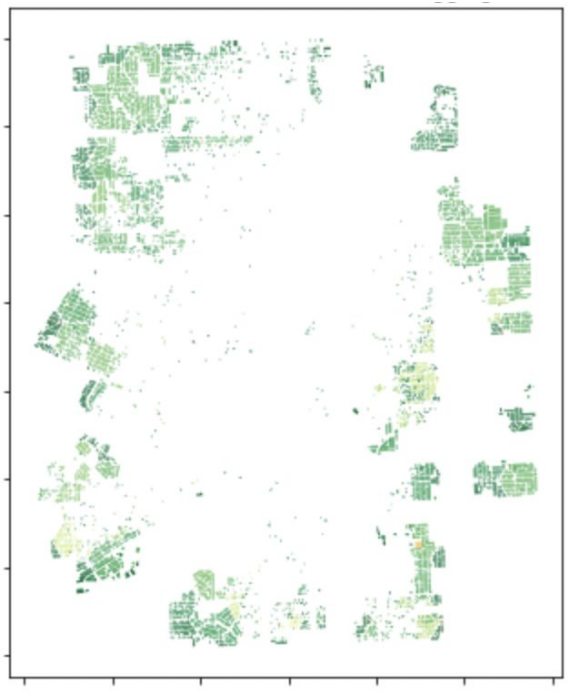
# Savings Over BlueConduit's Baseline





Homes with lead that rose in dig order



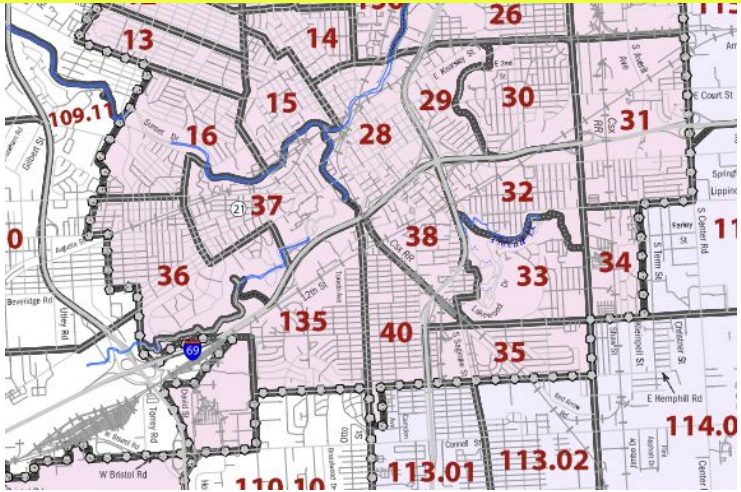
Homes without lead that fell in dig order

## What This Looks Like In Flint

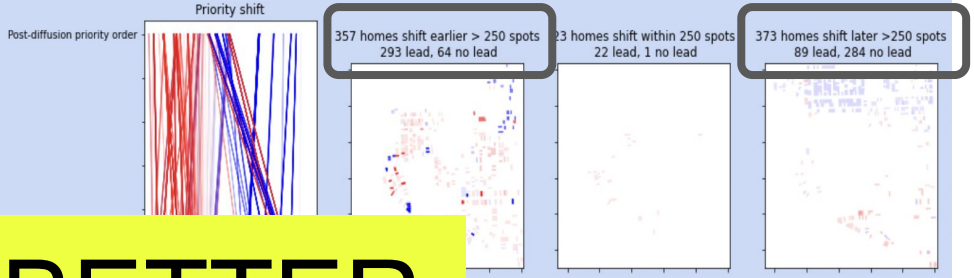
- Homes in city center (high lead density) rose in dig order
- Homes in suburbs (low lead density) fell in the dig order

Which census tracts are most impacted by diffusion?

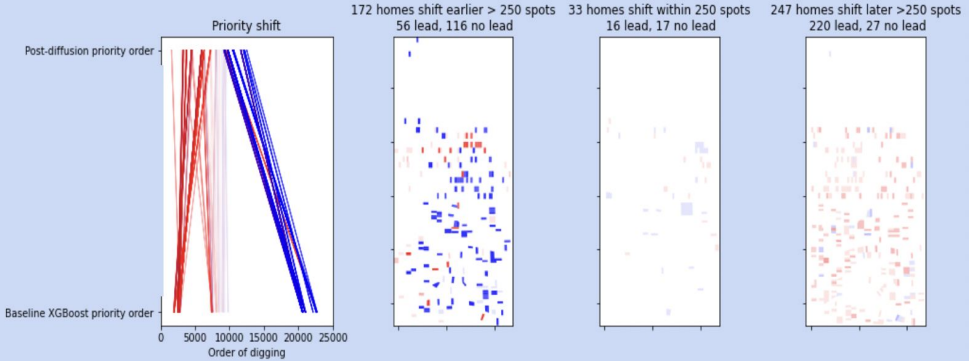

SAY/DEPICT THIS BETTER




Census tract 9




Census Tract 23

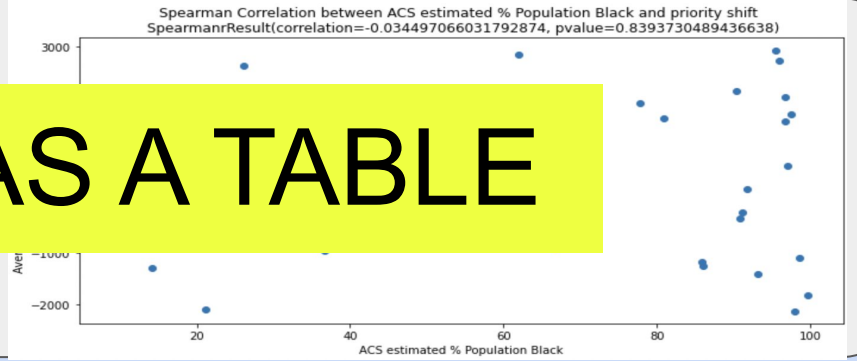


***% population Black:***

No cor

Very low confidence (p-value 0.84)

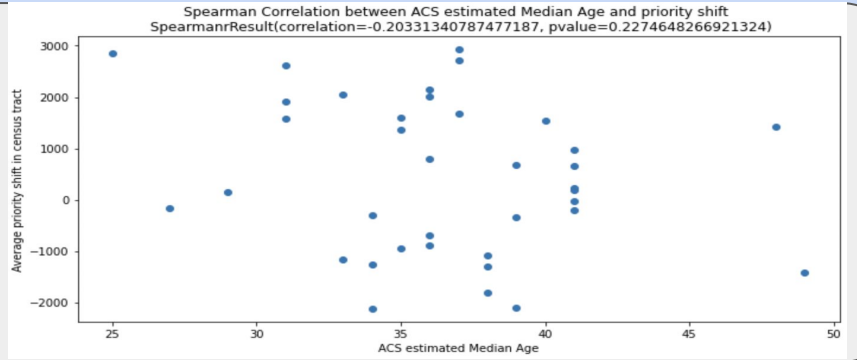
LAY THIS OUT AS A TABLE




***Median age:***

Negative correlation (Spearman coeff -0.2)

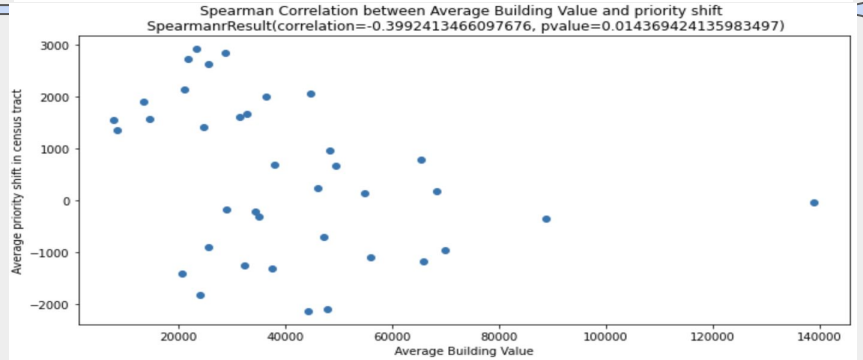
Low confidence (p-value 0.22)




***Residential Building Value:***

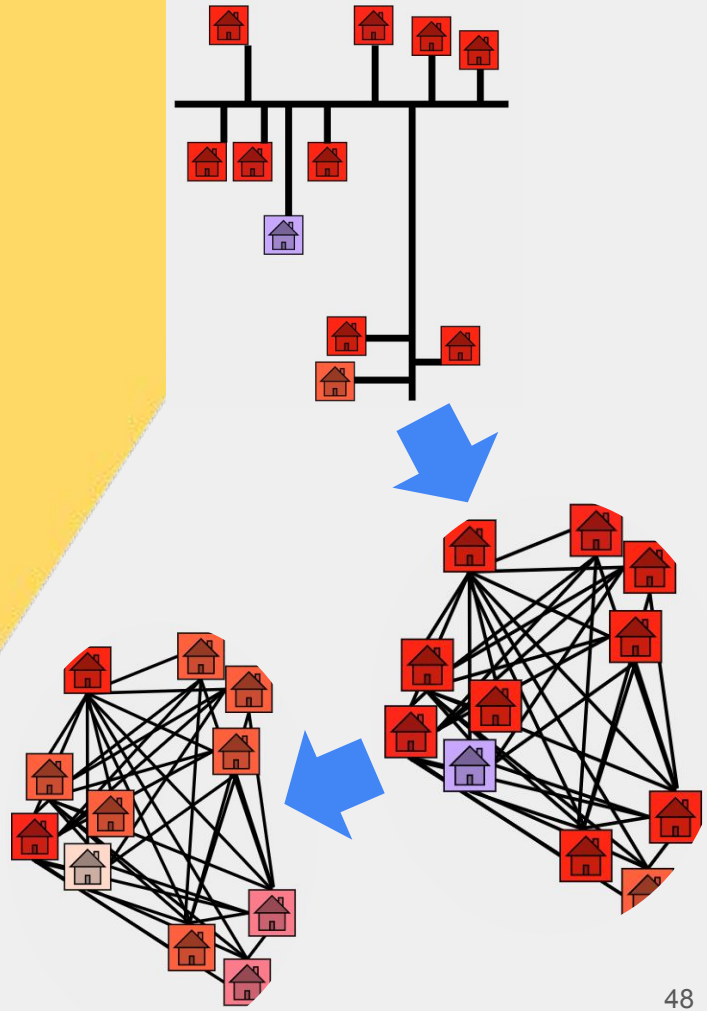
Negative correlation (Spearman coeff -0.4)

High confidence (p-value 0.01)



# Conclusions

- Spatial information **can improve** lead pipe predictions.
- Best performance when ...
  - Encode distance using roads
  - Allow for information sharing between neighbors



# Future Work

- Only use diffusion probabilities among initially uncertain homes.
- Test if results generalize to cities other than Flint.
- Explore other graph smoothing methods beyond diffusion.

# Thank you!


Kevin, Javiera, Max, Dash