

# WINNING THE RAT RACE

WHERE ARE BOSTON'S RODENTS?





Sasha Lioutikova



City of Boston Team: Jeff Lambart, Hope Ruse, John Ulrich, Gina Belmonte, Kelly Mackey

### PROBLEM STATEMENT

The City currently mitigates rodent activity in two ways:

**REACTIVELY** (via 311 reports)











INTERVENTION

**PROACTIVELY** 

(via consistent activity)

**CONSISTENT ACTIVITY** 

INTERVENTION

Current operations are **BIASED** to locations that report more & might MISS RODENT ACTIVITY

### **DATA**

### SPATIAL DATA

# **RODENT DATA**



**ENVIRONMENT** 

Parks, sewers, restaurants, trash pickup, sidewalks

Condition, age,

**CENSUS DATA** 

Population density

material



**VIOLATIONS** 

311 REPORTS



**PROACTIVE BAITING** 

# **OBJECTIVE:**

**Equitably predict and explain** rodent activity across Boston

### **METHODOLOGY**

### **SAMPLING**

### **GOAL: Collect Ground Truth Data to Train and Validate**

1) Find most informative locations

**Select 1 Point per Block** 

600 candidate locations

**Max(Spatial Uniformity)** 

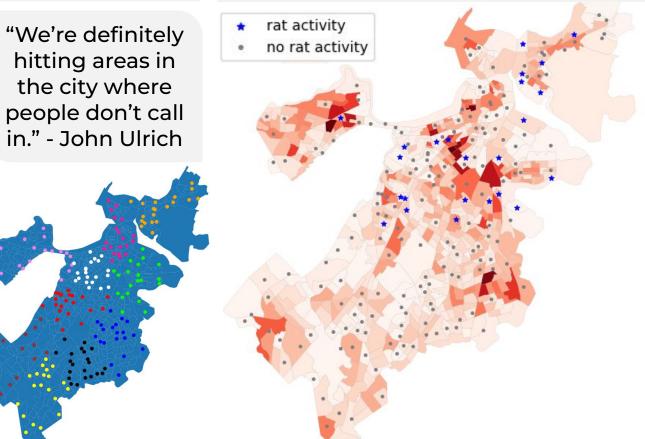
400 intermediate locations

**Max(Feature Uniformity)** 

200 final locations

Cluster

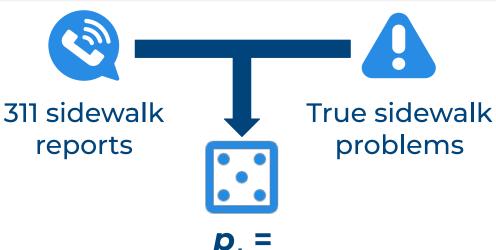




### **DEBIASING**

### **GOAL: Quantify and Offset Bias in 311 Complaints**

1) Estimate Reporting Bias per Tract



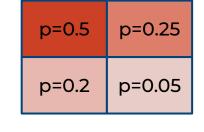
Proportion of sidewalk problems that are reported in tract t

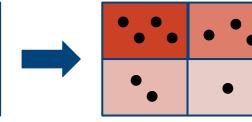
Why sidewalks?

- Have ground truth and 311 data
- Similar "inconvenience" as rodents

2) Pseudosample Negatives

Sample synthetic negatives ("pseudosample") from each tract, proportional to tract's reporting bias





Negatives have the same sampling bias as positives



Prevents model from learning bias

### GENERAL RODENT ACTIVITY MODEL

**GOAL:** Quantify **Drivers** of Rodent Activity

**GOAL:** 

**Predict** 

Actionable

Rodent

**Hotpots** 

### **FEATURES**

Spatial data

Census data

Building data

### **SUMMER 2024**

**LABELS** 

311 Complaints

Violations

**LOGISTIC REGRESSION** 

WITH INTERPRETABILITY

Proactive Inspections Sampled Absences

### **RESULTS & INTERPRETATION**

+ \$1 Average Building Value -54% odds of rodent activity

+ 1 Sewer Junctions Density +47% odds of rodent activity

**70%** 

0.76 **ACCURACY AUC** 

### Average **Building Value Distance** to Food **Establishments** Average **Building Age** Density of **Sewer Junctions Building Density Distance** to **Parks**

# **Bar Chart of Feature Importance**

### **ACTIONABLE RODENT ACTIVITY MODEL**

Spatial data

**FEATURES** 

Census data

Building data

Reporting bias metric

### TRAIN LABELS

**SUMMER 23 - SPRING 24** 

Violations

Absences

Proactive Inspections

Pseudosampled

### **TEST LABELS SUMMER 24**

Violations

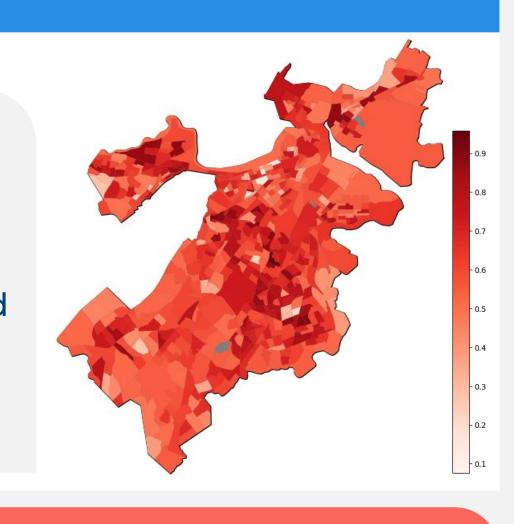
Proactive Inspections

Sampled Absences

## **RESULTS**

**72% PRECISION** 

72% of predicted locations have actionable activity



Informed by Reporting Bias from Debiasing methodology

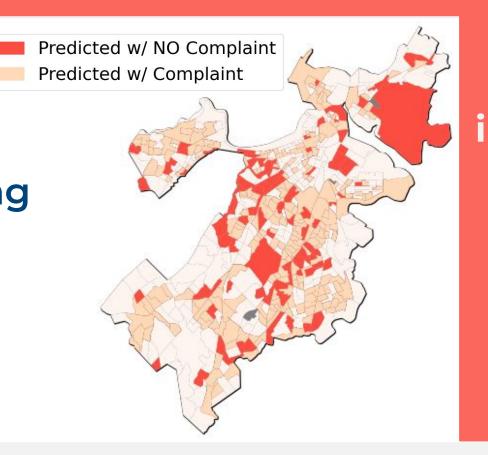


**RANDOM FOREST** WITH TIME SERIES CV

### **IMPACT**

Support equitable rodent mitigation

Identified 102 non-reporting areas with high risk of rodent activity



Increase inspection success rate

**72**% **PRECISION** 

140% increase from baseline of 30% from 311 complaints & baiting

**Facilitate** dialogue between departments



causes of rodent activity

Identified