

Food Inspection Forecasting

Project Background



Food Inspections

- 36 inspectors (“sanitarians”)
- 13,000 food establishments

2015: Chicago Department of Public Health (CDPH) launches a data-driven inspection model.



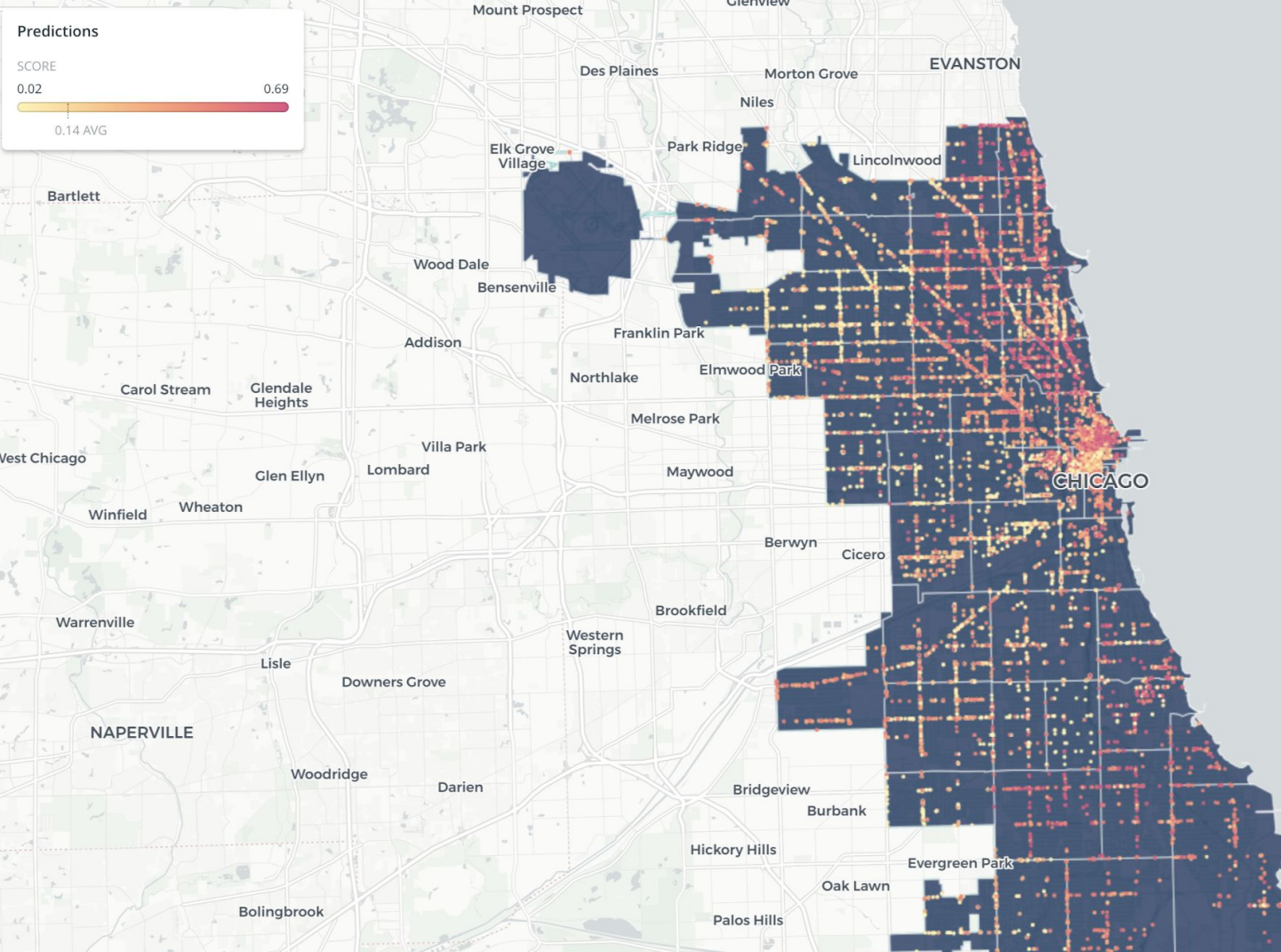
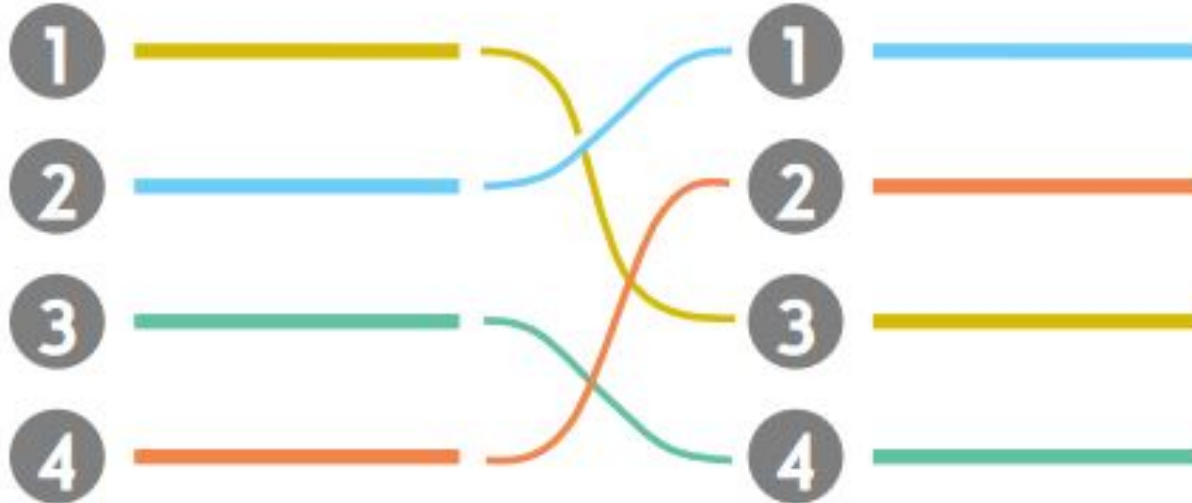


Image: City of Chicago, Carto

Typical order of inspections



Data-driven order of inspections



Forecasting restaurants with critical violations in Chicago

Tom Schenk Jr. (City of Chicago), Gene Leynes (City of Chicago), and Aakash Solanki (City of Chicago), Stephen Collins (Allstate Insurance), Gavin Smart (Allstate Insurance), Ben Albright, and David Crippin (Allstate Insurance)

May 15, 2015

The Chicago Department of Public Health (CDPH) inspects more than 15,000 restaurants with fewer than three dozen inspectors over the course of the year. This paper describes a predictive model designed to identify the presence of a critical violation in a particular food establishment. The goal of this model is to prioritize inspections by likelihood in order to identify the riskiest restaurants earlier, thereby reducing the length of exposure of risky restaurants to patrons. Critical violations were identified approximately 7.44 days earlier over a 60 day period compared to current operations in the out-of-sample test.

1 Introduction

In 2014 the [Chicago Department of Public Health](#) inspected performed over 20,000 inspections at nearly 13,000 food establishments across Chicago with fewer than three dozen inspectors. The majority of these food inspections were routine inspections that don't uncover serious problems, but some of these inspections uncovered issues that affect the health and safety of the patrons who visit these establishments. Traditionally, prioritizing these inspections is a largely manual task that relies on a combination of administrative processes and personal expertise.

The model set forth in this paper can help with the prioritization of scheduled, saving time and money as well as making the city's food safer. The model utilizes several data sources and through advanced modeling techniques the model provides additional insight into an establishment's current actual risk based on real-time data.

This paper is organized as follows: Section 1 provides an introduction and background to describe the current process and scope of the problem. Section 2 describes data that has been collected by the research team for this project and how that data was combined. Section 3 describes the model. Section 4 describes the model evaluation. Section 5 contains details of the model results from the experiment. Finally, the Summary

Progress

- Read Chicago paper
- Contacted CDPH to interview staff
 - Currently not available
 - Connected with paper authors, former Chief Data Officer
- Reproduced model and results (in Python)
- Set up reproducible data pipeline



Research Questions (Draft)

- How might we make risk scores more meaningful?
- What level of transparency does the public deserve?



TASK



Inspection Types

- Primary Types
 - Canvass
 - License
 - Complaint
- Very high pass rate for re-inspections

Inspection_Type	2014 Count
Canvass	12,229
Canvass Re-Inspection	2,284
License	2,418
License Re-Inspection	708
Complaint	1,653
Short Form Complaint	644
Complaint Re-Inspection	689
Other	172

Risk Levels

Inspection Rates

- 1. Twice a year
- 2. Once a year
- 3. Once every two years

Risk Category	Count of Licenses
Risk 1 (High)	8,510
Risk 2 (Medium)	3,128
Risk 3 (Low)	1,264
TOTAL	12,902



Violation Codes

CRITICAL
V1 - V14

SERIOUS
V15 - V29

MINOR
V30 - V44, V70

- Critical Violations: Associated with greatest risk of outbreak
- Even one critical → **FAILURE**
- Too many serious and minors → **FAILURE**

Top Critical Violations (Training Data)

Code	Count	Violation
V3	1292	All hot food shall be stored at a temperature of 140°F or higher. All cold food shall be stored at a temperature of 40°F or less.
V2	600	All food establishments that prepare, sell, or store hot food shall have adequate hot food storage facilities. All food establishments that display, prepare, or store potentially hazardous food shall have adequate refrigerated food storage facilities.
V8	409	Equipment and utensils should get proper exposure to the sanitizing solution during the rinse cycle. Bactericidal treatment shall consist of exposure of all dish and utensil surfaces to a rinse of clean water at a temperature of not less than 180°F.



Top Serious Violations (Training Data)

Code	Count	Violation
V18	2074	All necessary control measures shall be used to effectively minimize or eliminate the presence of rodents, roaches, and other vermin and insects on the premises of all food establishments, in food-transporting vehicles, and in vending machines.
V21	1260	A certified food service manager must be present in all establishments at which potentially hazardous food is prepared or served.
V29	706	A separate and distinct offense shall be deemed to have been committed for each Minor violation that is not corrected upon re-inspection by the health authority.



Bryan 1987

Risks of Practices, Procedures and Processes that Lead to Outbreaks of Foodborne Diseases

TABLE 1. *Factors that contributed to the occurrence of 766 outbreaks of foodborne disease, United States, 1977-1982.*

Contributory factor	Number	Percent ^a
Improper cooling	313	40.9
Holding at room/outside ambient temperature	162	21.1
Large/deep containers in refrigerator	151	19.7
Lapse of 12 or more hours between preparing and eating	193	25.2
Contaminated raw food/ingredient	175	22.8
Obtaining food from unsafe source	138	18.0
Colonized person handled implicated food	116	15.1
Inadequate cooking/canning/heat processing	109	14.2
Inadequate reheating	70	9.1
Improper hot holding	70	9.1
Cross contamination	29	3.8
Improper cleaning of equipment/utensils	27	3.8
Toxic containers/pipelines	23	3.0
Use of leftovers ^b	19	2.5
Intentional additives	19	2.5
Improper fermentation	12	1.6
Incidental additives	10	1.3
Mistaken for edible varieties	9	1.2
Inadequate/improper thawing	5	0.7
Inadequate acidification	5	0.7
Contaminated water	3	0.4
Slow/inadequate drying	2	0.3
Post processing contamination	2	0.3
Soaking time too short	1	0.1
Flies	1	0.1

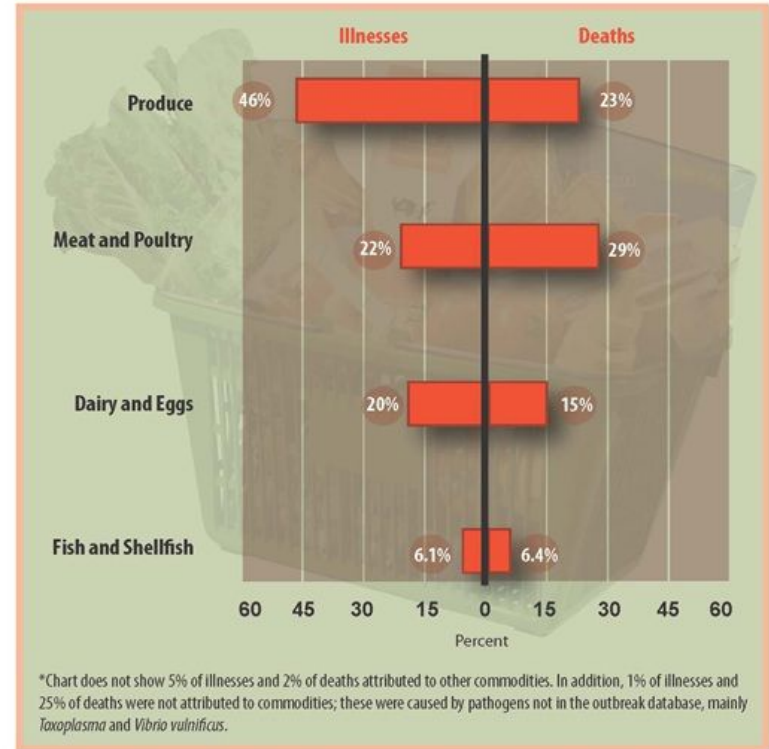
^aPercentage exceeds 100 because multiple factors contribute to single outbreaks.

^bAlso lapse of 12 or more hours.

Painter et al. 2013

Attribution of Foodborne Illnesses, Hospitalizations, and Deaths to Food Commodities by using Outbreak Data, United States, 1998–2008

Figure 1. Contribution of different food categories to estimated domestically-acquired illnesses and deaths, 1998–2008*



Source: Painter JA, Hoekstra RM, Ayers T, Tauxe RV, Braden CR, Angulo FJ, Griffin PM. Attribution of foodborne illnesses, hospitalizations, and deaths to food commodities by using outbreak data, United States, 1998–2008. *Emerg Infect Dis* [Internet]. 2013 Mar [date cited]. <http://dx.doi.org/10.3201/eid1903.111866>

DATA



Study Design

- Only CDPH senior management aware of project
- No date overlap between training and pilot set
- Simulate model performance on pilot data



Dataset

Training Data

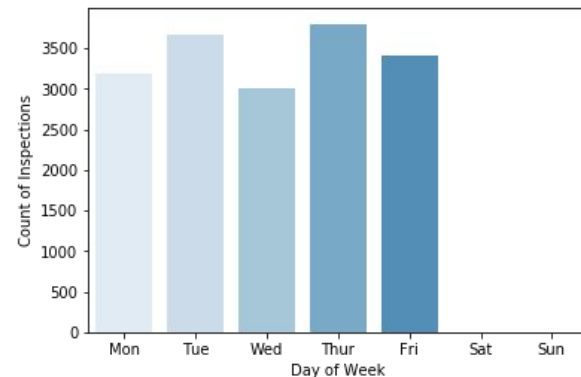
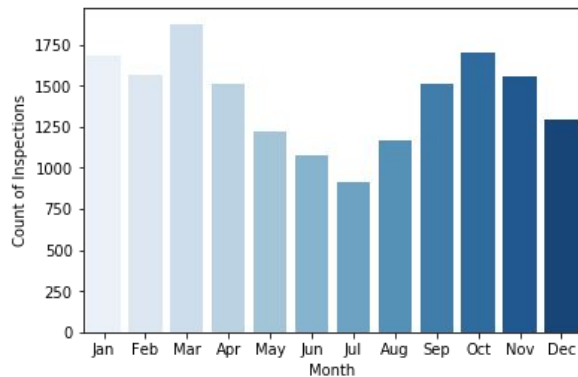
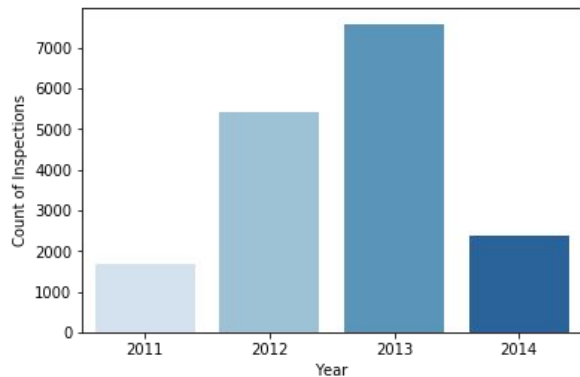
- $N = 17,075$
- $P(\text{Critical}) = 14.1\%$
- Sept 2011 - Apr 2014

Pilot Data

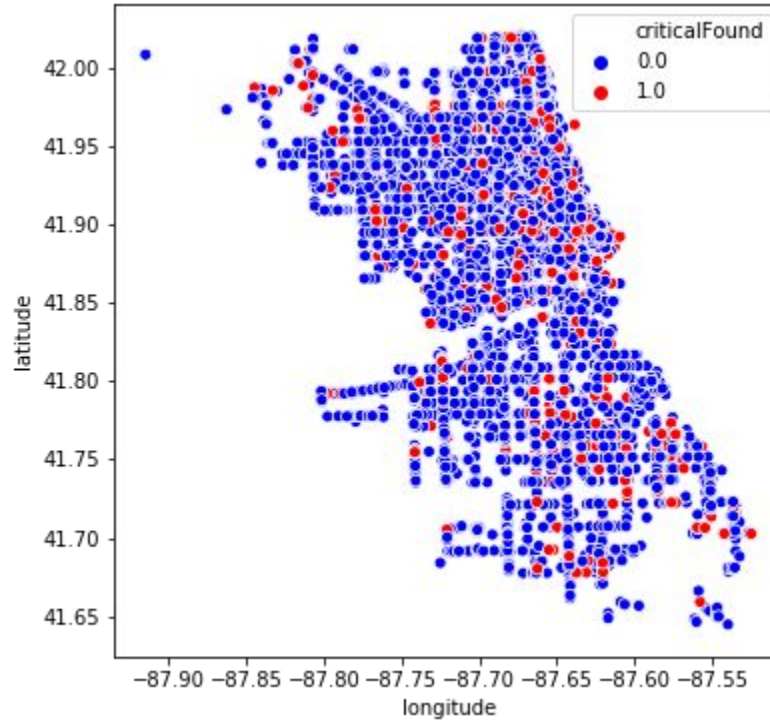
- $N = 1,637$
- $P(\text{Critical}) = 15.8\%$
- Sept 2014 - Oct 2014



Temporal Distribution (Training Data)



Spatial Distribution (Training Data)



Multiple Failed Inspections

dba_name	Failures
SOUTH BRANCH	4
KINCADE'S	3
CHOP SUEY HUANG	3
PITA KABAB, INC.	3
ANDIAMO'S O'HARE, LLC	3
FURAMA RESTAURANT INC	3
FRANKS 'N' DAWGS	3
DEMERA ETHIOPIAN RESTAURANT	3
ILLINOIS SPORTSERVICE, INC.	3
PRIMO PIZZA INC	3
ARAMARK@ LOYOLA SIMPSON RESIDENTIAL	3
GSB CAFE	3
MANDARIN KITCHEN	3
LEVY RESTAURANTS @ WRIGLEY FIELD	3
HON KEE REST INC	3

Violation Code Matrix

	1	2	3	4	5	6	7	8	9	10	...	36	37	38	39	40	41	42	43	44	70
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	1	0	0	0	1	0	0	1	0	0
3	0	0	0	0	0	0	0	0	0	0	...	1	0	0	0	0	1	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	1	0	0	0	1	0	0	0	0	0



Ten Predictors

1. B: Past Serious Violation
2. B: Past Critical Violation
3. C: Time Since Last Inspection
4. C: Age at Inspection (-)
5. B: Alcohol License
6. B: Tobacco License
7. C: Daily High Temperature
8. C: Intensity of Local Burglaries
9. C: Intensity of Local Sanitation Complaints
10. C: Intensity of Local Garbage Cart Requests (-)

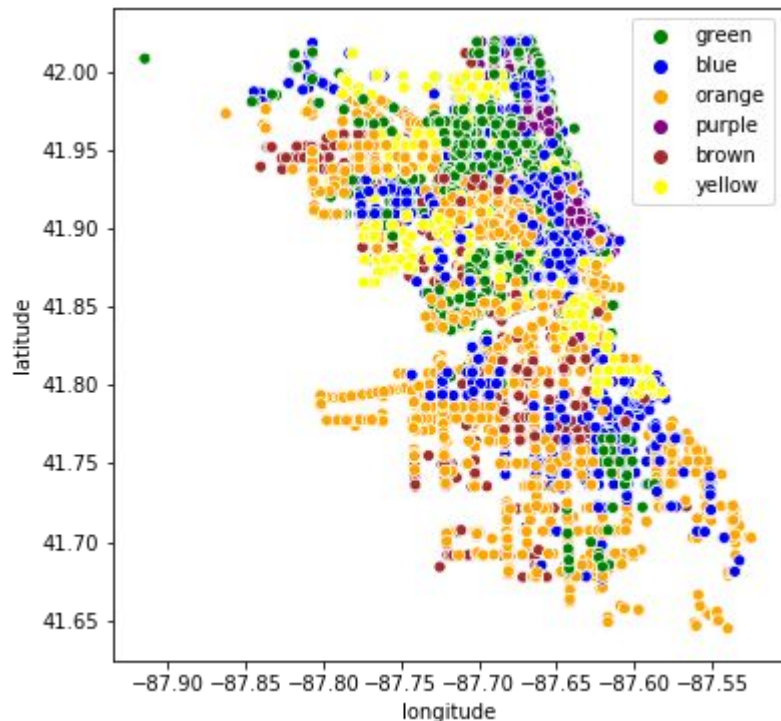


Coefficients

- Logistic Regression
 - Target: Critical Violation?
 - L1 regularization for inspector coefficients
 - No regularization for other coefficients

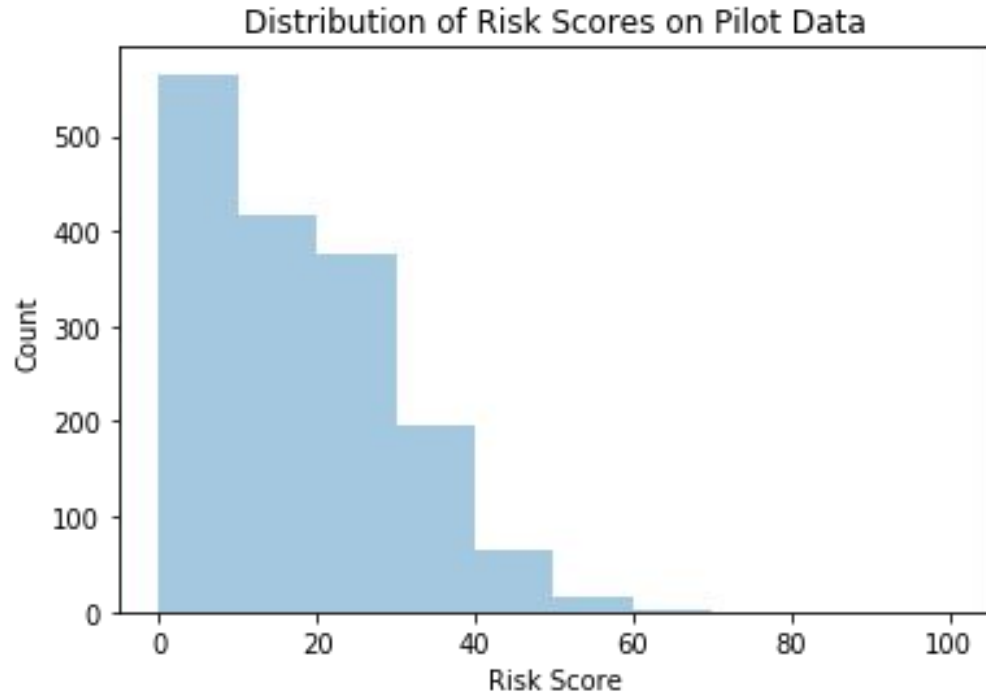
Inspector_blue	0.950
Inspector_brown	-1.306
Inspector_green	-0.244
Inspector_orange	0.202
Inspector_purple	1.555
Inspector_yellow	-0.697
pastSerious	0.302
pastCritical	0.427
timeSinceLast	0.097
ageAtInspection	-0.164
consumption_on_premises_incidental_activity	0.411
tobacco_retail_over_counter	0.171
temperatureMax	0.005
heat_burglary	0.002
heat_sanitation	0.002
heat_garbage	-0.004

Inspector Clusters



	Coefficient	count	hit_rate
Inspector_purple	1.555	1174	0.406
Inspector_blue	0.950	2897	0.265
Inspector_orange	0.202	3769	0.136
Inspector_green	-0.244	4595	0.095
Inspector_yellow	-0.697	2762	0.058
Inspector_brown	-1.306	1878	0.024

Risk Scores



Facility Types

“We filtered the inspections to only include Retail Food Establishments, which excluded establishments such as schools and hospitals. Their inspection schedules follow a different planning process, and we also believe that these establishments have different risk characteristics that are not generalizable across the entire population.” (Schenk Jr. et al. 2015)

	Count
Restaurant	13760
Grocery Store	2349
Bakery	335
School	138
Catering	132
Hospital	46
Long Term Care	15
BANQUET HALL	15
Liquor	13
STADIUM	10
Shelter	10
GAS STATION	10
Wholesale	7
CAFETERIA	6
GROCERY/RESTAURANT	6

EVALUATION



Assumptions

- All food establishments will be inspected
- Inspection capacity and schedule will not change, only order
- Critical violations are time invariant



Metrics

- Average change in time to identifying critical violations
- Standard deviation of change in time to identifying critical violations
- Proportion of critical violations found in first half



Chicago Model Performance

Mean Time to Critical Violation **7.44** days

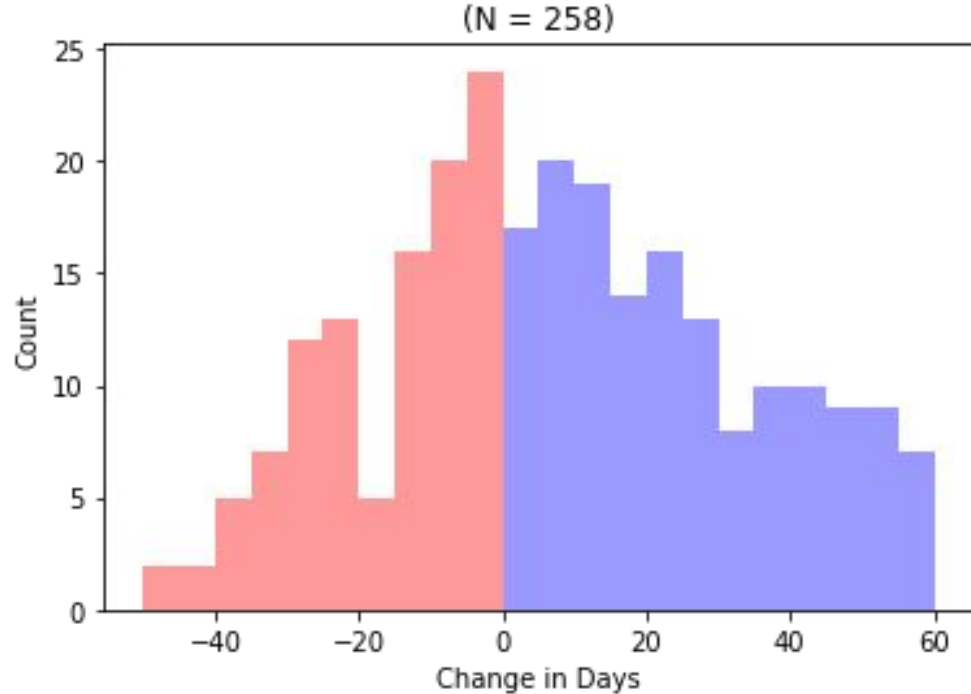
Stdev Time to Critical Violation **25.20** days

First Half Critical Proportion **0.55** of all critical violations

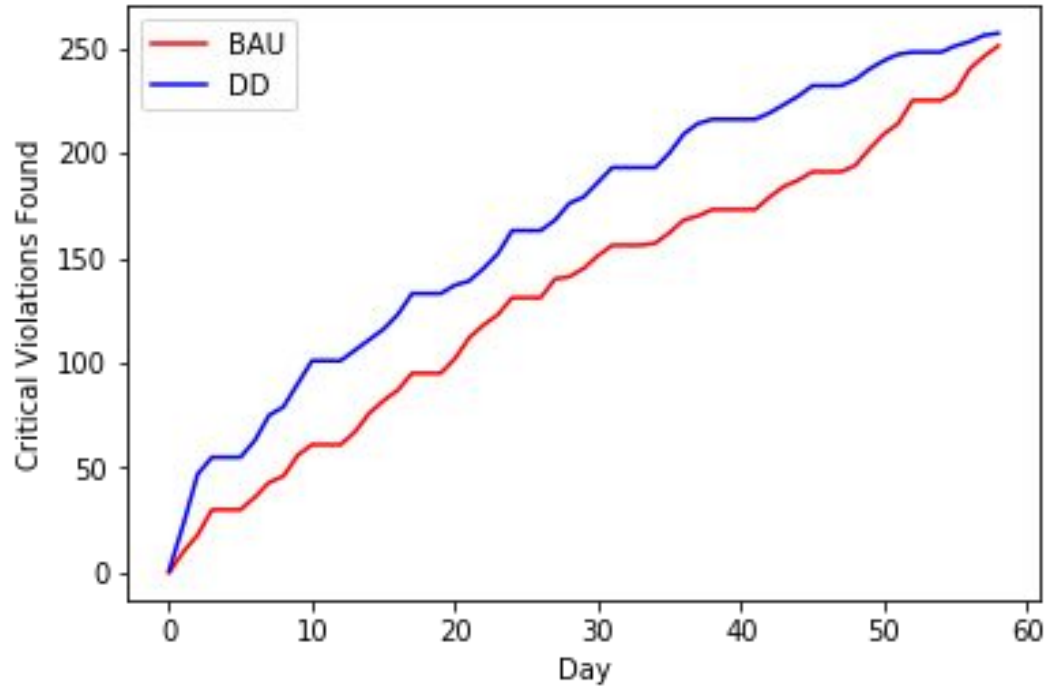


Time to Critical Violation

Mean 7.44
Stdev 25.20
Min -50.00
25% -9.00
50% 6.00
75% 25.00
Max 58.00



Cumulative Violations Found



Confusion Matrix (Pilot Data)

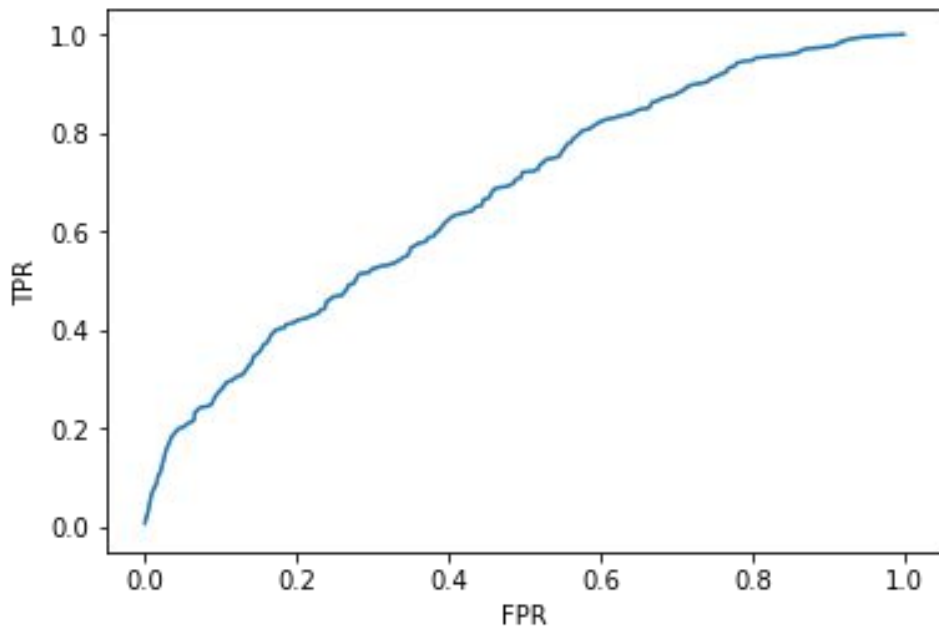
	Predicted Critical	Predicted None
Actual Critical	12	246
Actual None	7	1372

- F1-Score = 0.087
- Precision = 0.632
- Recall = 0.047

- Maximum F1-Score = 0.346
 - Threshold = 0.282



Area Under ROC Curve = 0.672



Production Model vs Pilot

- Director of Food Protection uses model results to inform choices
- Coefficients for specific inspectors
- DOIT changes the predictors in production

