MAKING THE CUT

Predicting Inspection Results for Chicago Based Restaurants

Shannon McDonnell

Agenda

Part 1: Introduction

- ❖ Prediction Question
- ❖ Objectives
- ❖ Relevance
- ❖ Background Information

Part 2: Implementation and Results

- ❖ Data
- ❖ Feature Engineering
- ❖ Classification Modeling
- ❖ Evaluation Metrics
- ❖ Results and Applications
- ❖ Next Steps

Will a restaurant pass or fail inspection?

Objectives:

- ★ Predict if a certain restaurant will pass or fail inspection in Chicago using ML classification modeling
- ★ Determine what parameters and/or features are influential to target
- ★ Keep an eye out for potential improvements we can make to improve our modeling in the future

The overall intent is to better predict the outcome of a restaurant passing or failing inspection, which could then help improve the quality of restaurants throughout the Chicago area, and help prevent the spread of foodborne diseases.

Relevance

Given COVID-19 this is a particularly relevant issue for restaurants right now.

The measures originally put in place for inspection results have dramatically changed.

It will be interesting to see what happens next...

Note: This dataset is pre-covid

Why does all of this matter?

72%

HIGH risk of FAILING!

Food Protection Services

WHO: Health Inspections and Food Protection Services

WHAT: A subsector of the Public Health Department

WHY: Ultimate goal is to prevent the spread of foodborne diseases

***HOW:** Enter our modeling features!*

Background Information:

Where does this data come from?

- ❑ This data is derived from inspections of restaurant and other food establishments in the city of Chicago
- ❑ Collected from January 1, 2010 - present
- ❑ Roughly 215,000 data points
- ❑ Collected and performed by the Chicago Department of Public Health's Food Program
- ❑ Standardized procedure is performed and the results of the inspections are inputted into a database and evaluated by LEHP

*For more information about Food Inspections, go to https://www.cityofchicago.org/city/en/depts/cdph/provdrs/healthy_restaurants/svcs/food-protection-services.html.

Data Introduction

Description of unclear feature names:

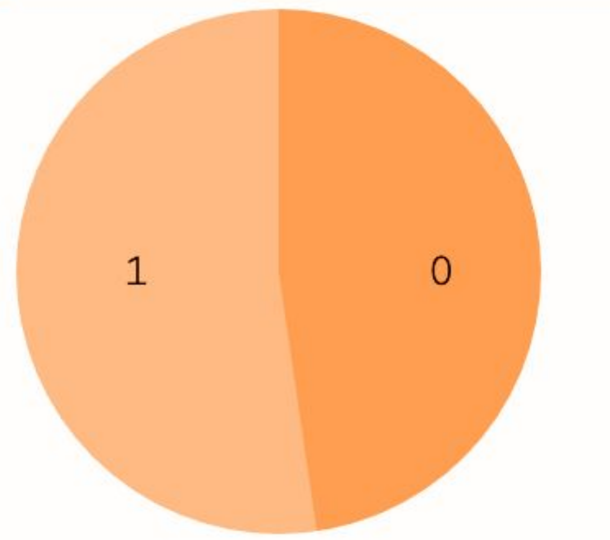
- ***licence #*** is referring to a unique establishment, in this instance a particular restaurant*
- ***Risk*** is the level of risk that adversely affects the public's health
  - 1 (lowest) - 3 (highest)

*Restaurant** is a term being used to capture a variety of facilities under inspection, it is not a feature name

  - Includes restaurants, bakeries, coffee shops, schools, shelters, taverns…

DATA

Binary Classification



TARGET = Results

- PASS (1)
- FAIL (0)
- Visibly well-balanced

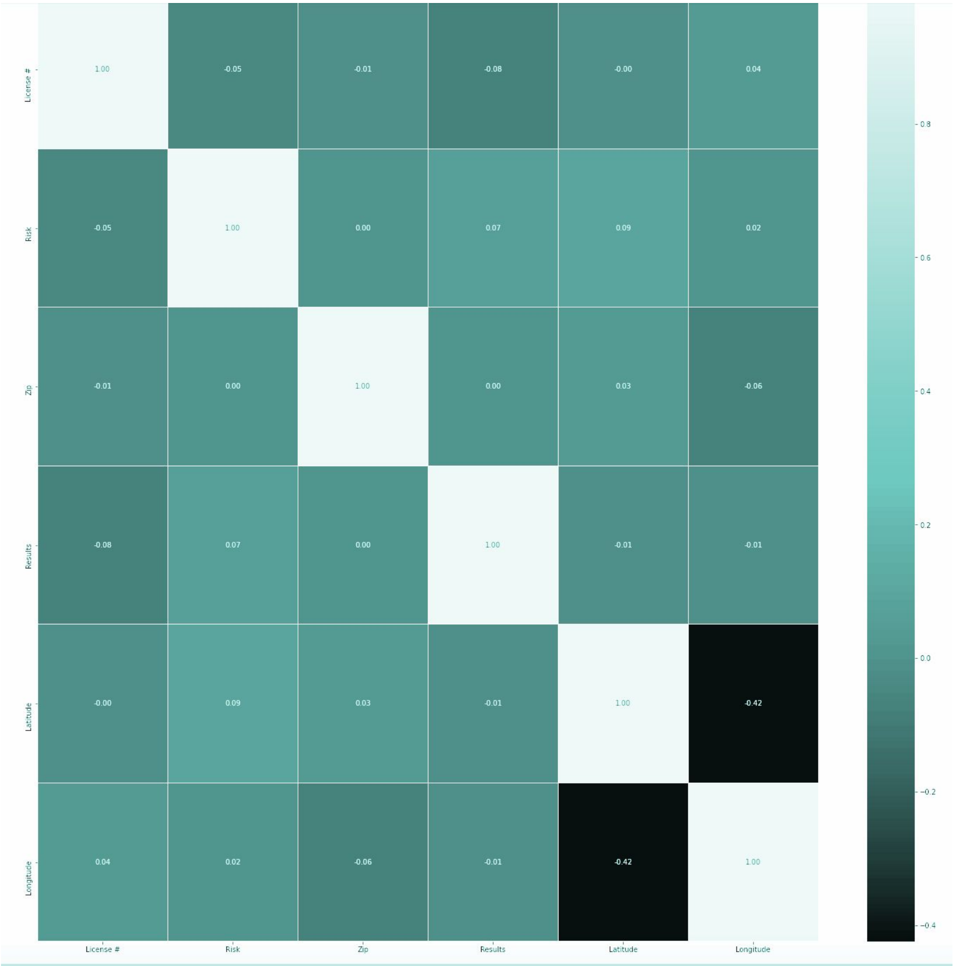
FEATURES:

- License #
- Inspection Type
- Latitude
- Longitude
- Risk

Heat Map

Feature Correlation

- License #
- Risk
- Zip
- Results
- Latitude
- Longitude



Zip Code

- The differences in percentages (of pass/fail) across different zip codes can be huge and varied

Risk

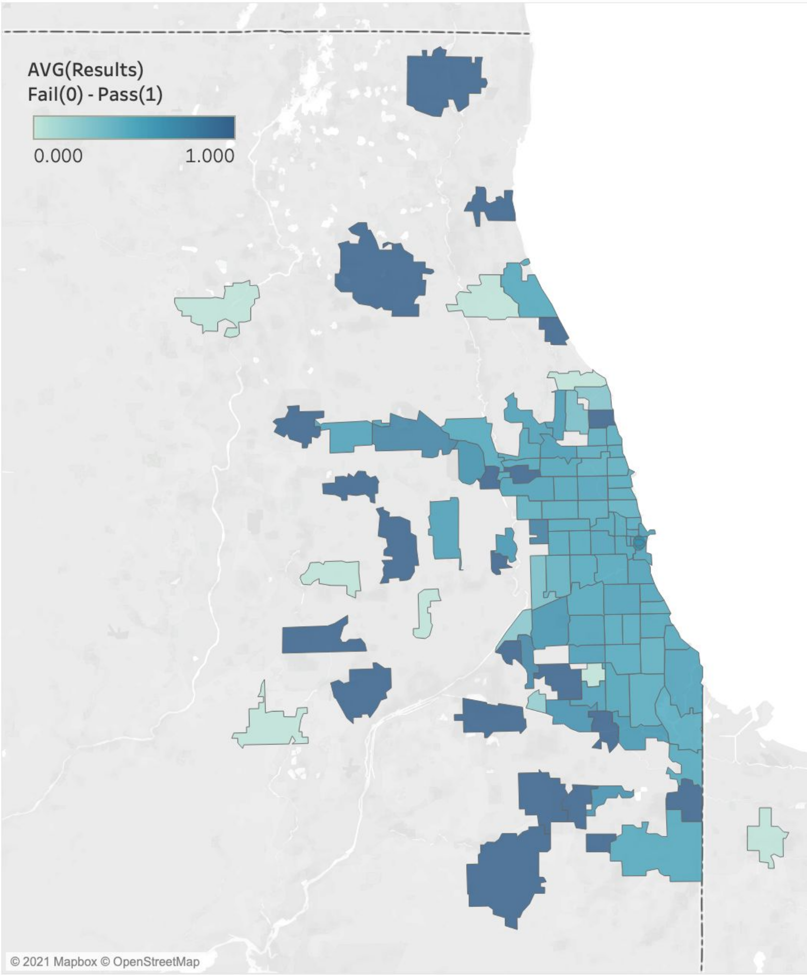
- The risk levels are inconsistent and deceiving if looked at alone

Latitude and Longitude

- Perhaps the best visualization of the features

- Based on the average results of the target, it appears restaurants that are inland are more likely to pass inspection

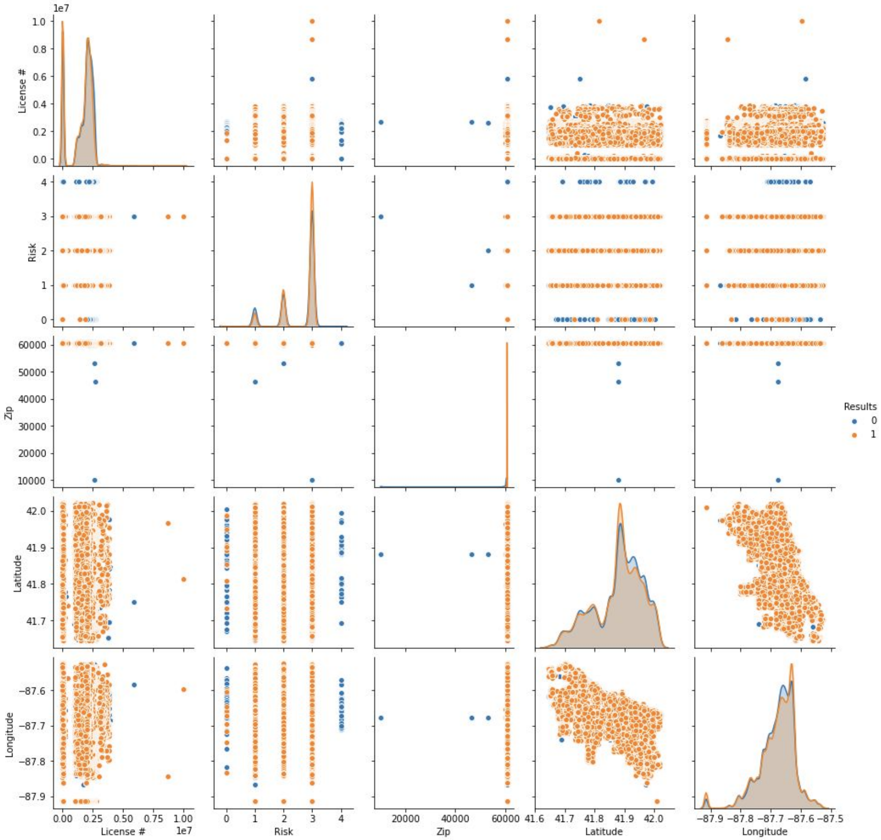
Average Results Based on Location



Feature Distribution

Pair Plot

- Latitude
- Longitude
- Zip
- Risk



Classification Modeling

- LOGISTIC REGRESSION
  - okay, at first glance
- KNN
  - very expensive
- NAIVE BAYES
  - deceiving
- RANDOM FOREST
- DECISION TREES

Evaluation Metrics

All models evaluated based on:

- Confusion matrices
- Accuracy
- Recall
- Precision
- F1 scores

Decision Metric:

F1 score

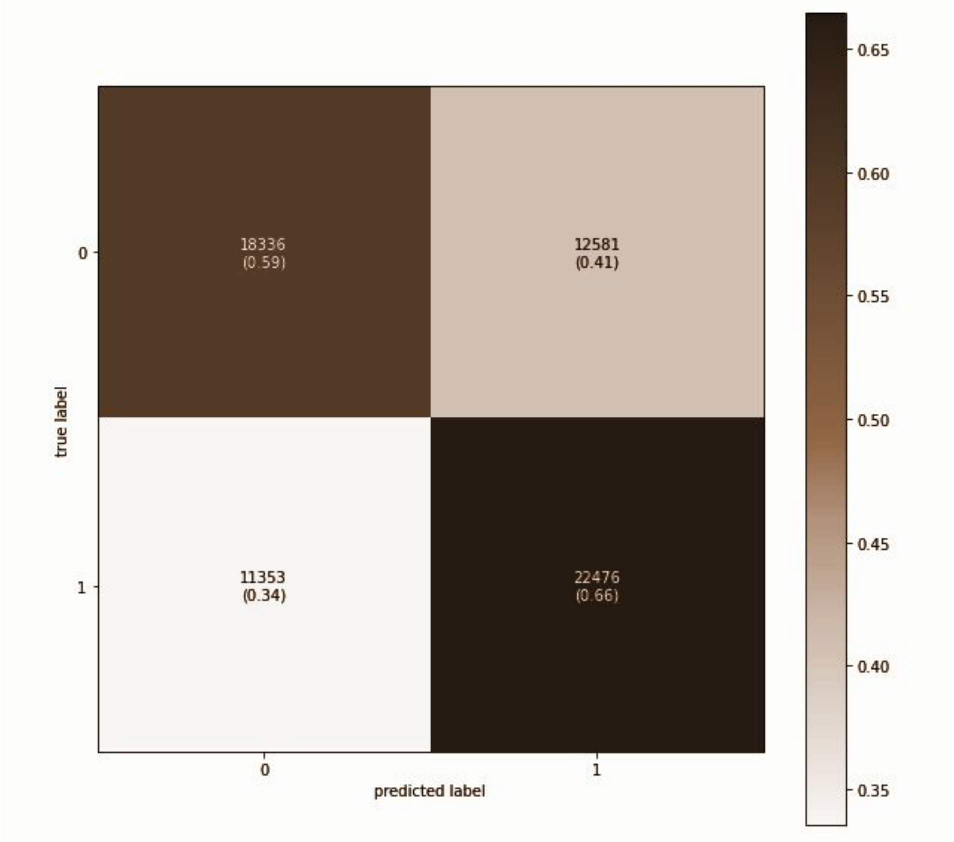
Evaluations Across All Models

	Accuracy	Recall	Precision	F1 Score
Logistic Regression	.62	.58	.66	.62
KNN	.63	.66	.64	.65
Naive Bayes	.55	.99	.52	.67
Random Forest	.62	.64	.63	.64
Decision Tree	.60	.56	.63	.60

Results

Best Model: KNN

✓ F1 Score = 65%



Practical Application

- ➔ KNN Classification Modeling can be used to predict restaurant inspection results *before* they happen
  - ◆ This will keep businesses up and running
  - ◆ Help keep people safe
  - ◆ Save Public Health Resources
- ➔ This modeling can help restaurant owners better place their businesses based on geographical features where they are more likely to pass inspection
- ➔ This is good for the restaurant owners, the Department of Public Health and Food Protection Services, and consumers...like you!

Next Steps...

1. Combine zip code data with median household incomes
2. Examine how much geographical location matters when it comes to income levels in poorer vs. richer neighborhoods
   a. Potential troubles: borders are arbitrary and hard to define
   b. Examine North vs. South sides of Chicago
3. Need more specific examples
   a. Look at types of restaurants specifically
   b. Look at cuisines
4. Look at more publically available datasets in the Chicago Public Portal
   a. Other features examined in other projects include weather, nearby burglaries, tobacco/liquor licenses, sanitation complaints, length of time establishment has been open

*Adding depth to data

Additional Reading

- “Food Inspection Forecasting: Optimizing Inspections with Analytics”

  (GitHub Write-up on similar topic)

  Food Inspection Forecasting - City of Chicago

- Food Protection Services and COVID-19

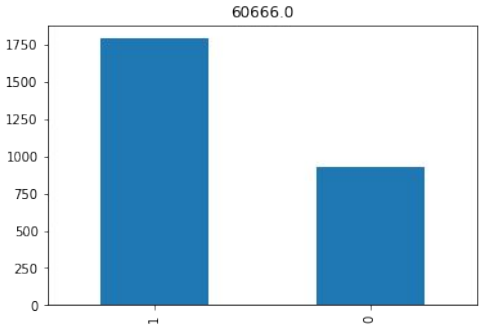
  Food Protection Services

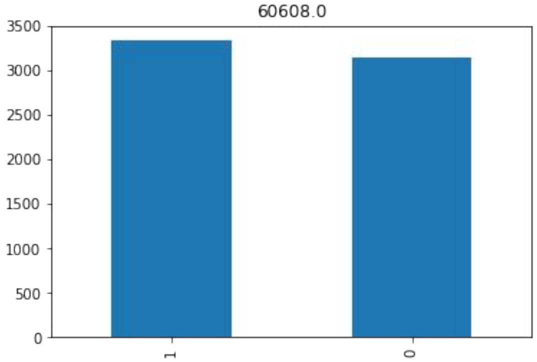
Questions?

Thank you!

Appendix

Comparing 2 Zip Codes





Comparing 2 Risk Levels