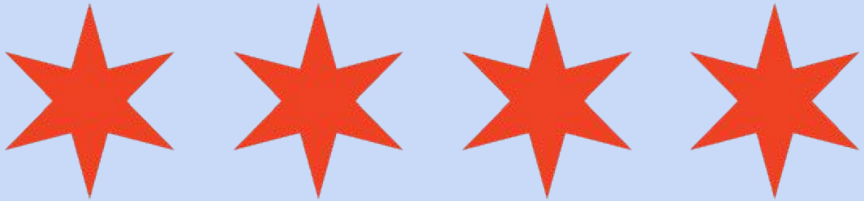


Final Project

Use Chicago Department of Public Health Food Inspection Data to Predict Restaurant Closures

The problem

Problem statement

Can we provide lenders and investors a prediction of restaurant closures based on on historical food inspection data?

Data Set

- ➔ Includes 10 years of historical food inspection data
- ➔ Parsed down to only include restaurants for the purpose of this analysis
- ➔ Updated by Department of Public Health on a weekly basis

Metrics

- ➔ **Risk** : Low, Medium, High risk to public health
- ➔ **Results:** Inspection results (Pass, Fail, Out of Business, etc)
- ➔ **Violations:** Violation # & Description
- ➔ **Risk Number:** (Numeric Risk Mapping)
- ➔ **Results Number:** (Numeric Result Mapping)
- ➔ **Number of Violations:** Total number of violations for any given restaurant

Approach & Process

Initial Approach

Predict restaurant closings based on number of violations

Hypothesis

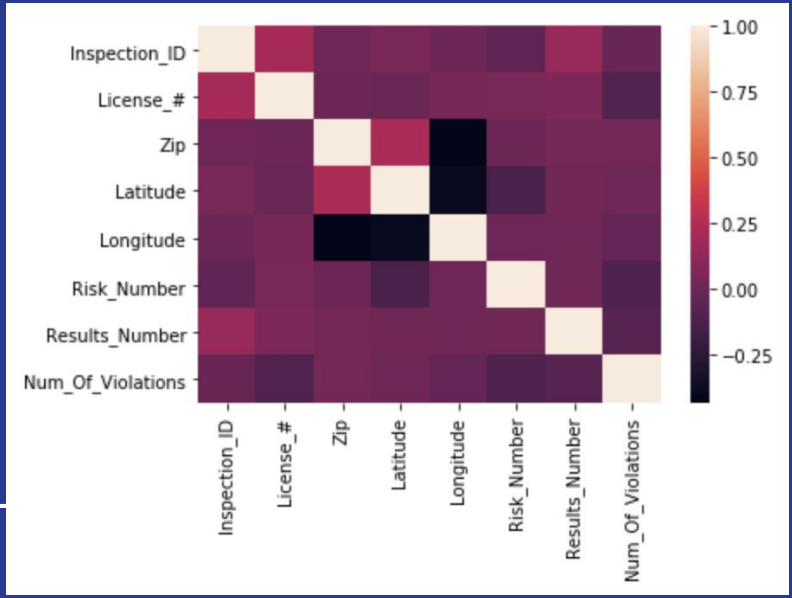
Restaurants with more violations are likely to go out of business

EDA

- ➔ Parse data to Restaurants only
- ➔ Manually calculate number of violations per restaurant
- ➔ Determine if there is a correlation between # of violations and result of inspection

Result

- ➔ **# of Violations & Result Type have no correlation.**
- ➔ **Move forward with text analysis of violations**



New Approach

Use a classification model to predict the probability of a restaurant closure

Models Considered

Logistic Regression
&
Multinomial Naive Bayes

Choosing A Model

Logistic Regression

- ➔ Process: Used number of violations & risk number to find the probability of a restaurant going out of business
- ➔ Accuracy of Model: 66%
- ➔ Null Accuracy: 57%
- ➔ Only provides the probability of a business closing

```
array([[0.31360041, 0.68639959],
       [0.40334058, 0.59665942],
       [0.43615511, 0.56384489],
       [0.28534025, 0.71465975],
       [0.6095206 , 0.3904794 ],
       [0.34331491, 0.65668509],
       [0.39357702, 0.60642298],
       [0.48986995, 0.51013005],
       [0.48986995, 0.51013005],
       [0.28534025, 0.71465975]])
```

Multinomial Naive Bayes

- ➔ Process: Used natural language processing on the data's violation field to find the probability & prediction of a business closing
- ➔ Accuracy of Model: 90%
- ➔ Null Accuracy: 88%
- ➔ Calculated probability to predict and label if a business will close down in the future

Impact of Model

Who can use this model?

Use Cases

1) **Investors:** Should I invest in this restaurant based on prediction?

2) **Bank Loan Lenders**: Should I lend to this restaurant owner based on their inspection history?

-

Next Steps

How can we make a stronger prediction?

Recommendations

- ➔ Use more features, such as location, in combination NLP to predict a restaurant closure.
- ➔ Take a look at merging restaurant review data to predict restaurant closure.
- ➔ Predict the avg. number of years before a restaurant closure to give investors and lenders more confidence in their decision

-