



# PREDICTING INSPECTION RESULTS AT FOOD ESTABLISHMENTS

*Prepared by Geovanna Meier*

## THE ISSUE AND WHY IT MATTERS

Health departments across the United States conduct unannounced inspections at food establishments to ensure they meet the food and safety requirements to protect public's health.

### **Why it matters:**

It's vital to prioritize inspections at restaurants that are most likely to have food and safety violations to avoid foodborne outbreaks and maximize the effectiveness of inspections.

## APPROACH

Use publicly available data on restaurant inspections and Yelp's ratings and business data to predict non-compliant results.

## OPEN DATA

### **Restaurant Inspection results:**

Publicly available data obtained from the Southern Nevada Health District with details on food establishment inspections.

Source: <https://opendata.lasvegasnevada.gov/API-Integration-/Restaurant-Inspections/q8ye-5kwk>

### **Business and Reviews Yelp Academic Datasets:**

Source: <https://www.yelp.com/dataset>

## STAKEHOLDERS

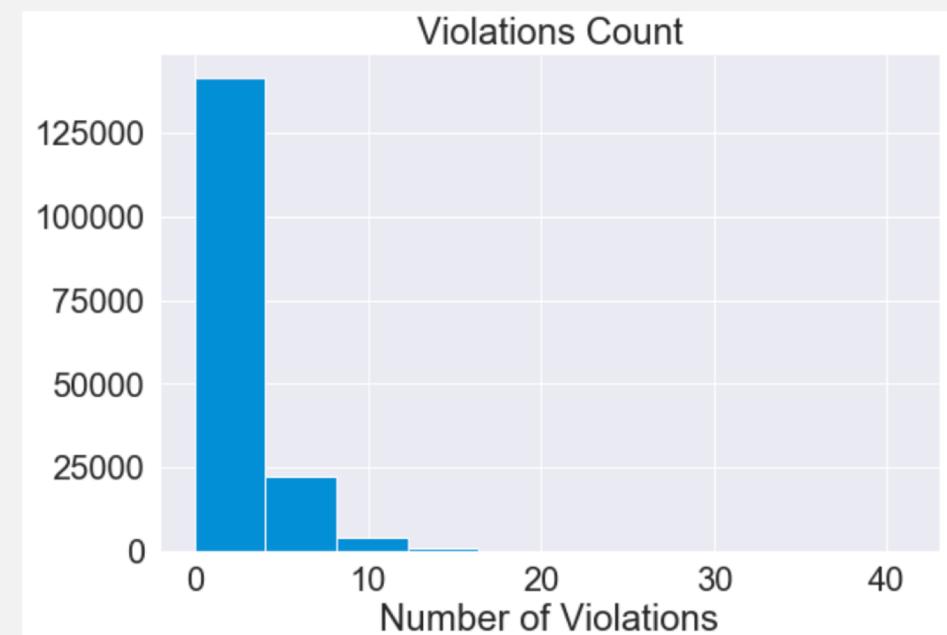
The results can provide actionable insights to local officials to prioritize inspections and scheduled visits to food establishments that are most likely to have critical violations or non-compliance outcome.

# INSPECTION RESULTS ANALYSIS

## THE STATE OF INSPECTIONS

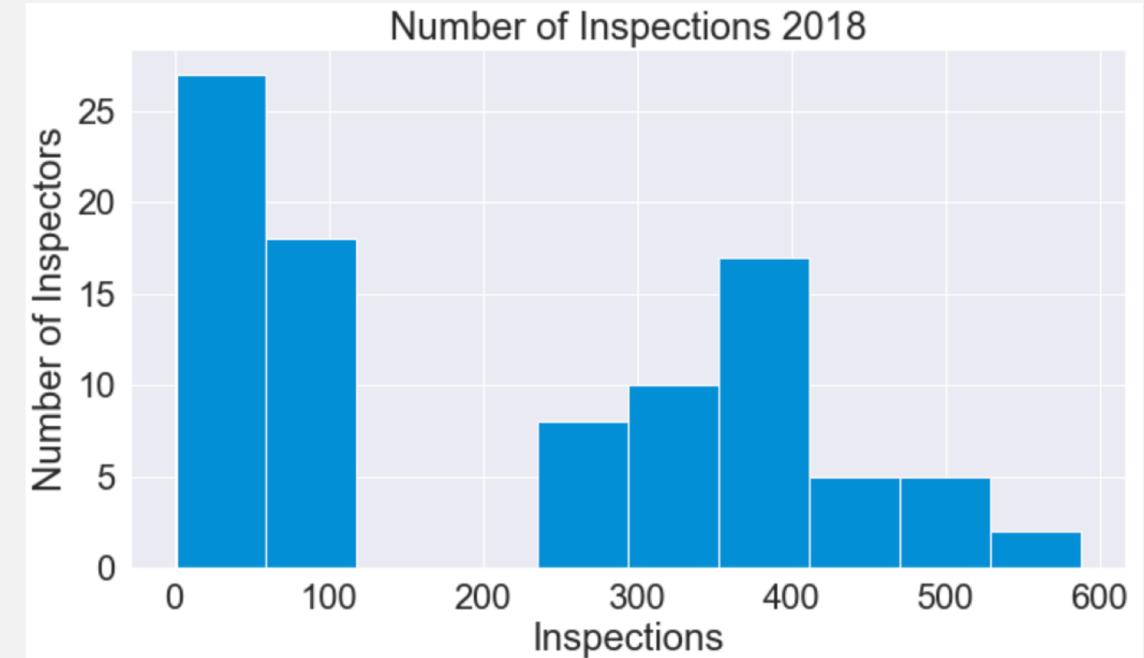
Overall compliance rate among food establishments is high, with 88% of inspections resulting on passing grades.

In fact, most inspections have less than 10 violations.

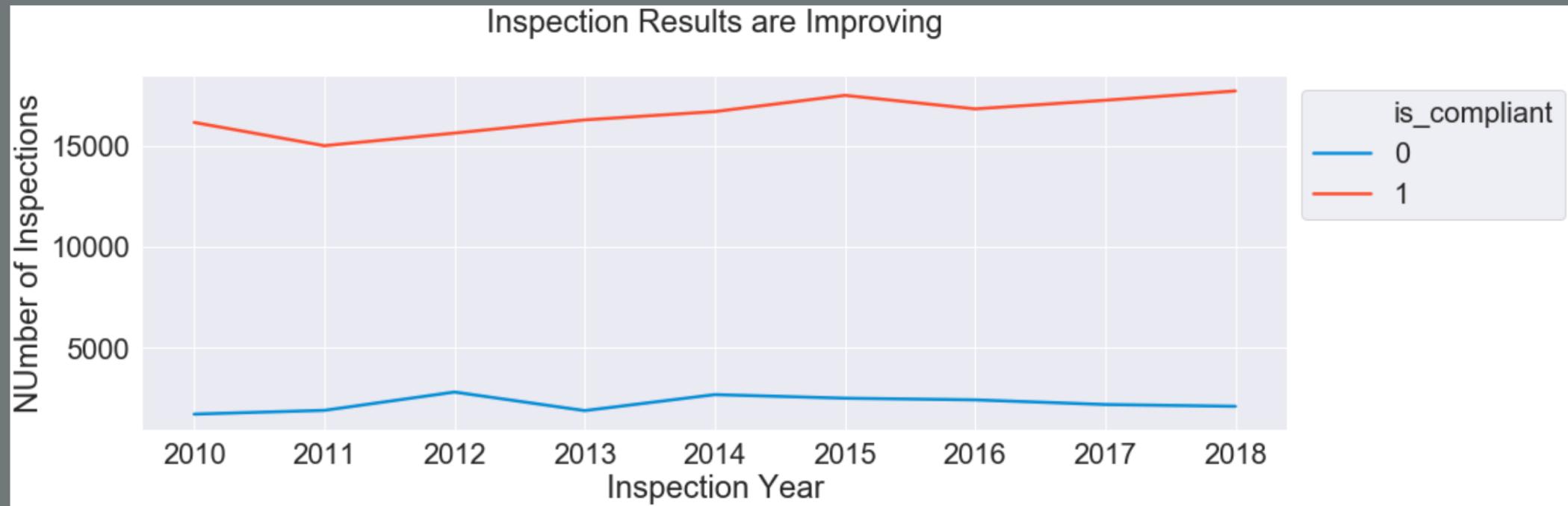


## INSPECTIONS AND INSPECTORS

In 2018, a small number of inspectors performed a large number of visits.

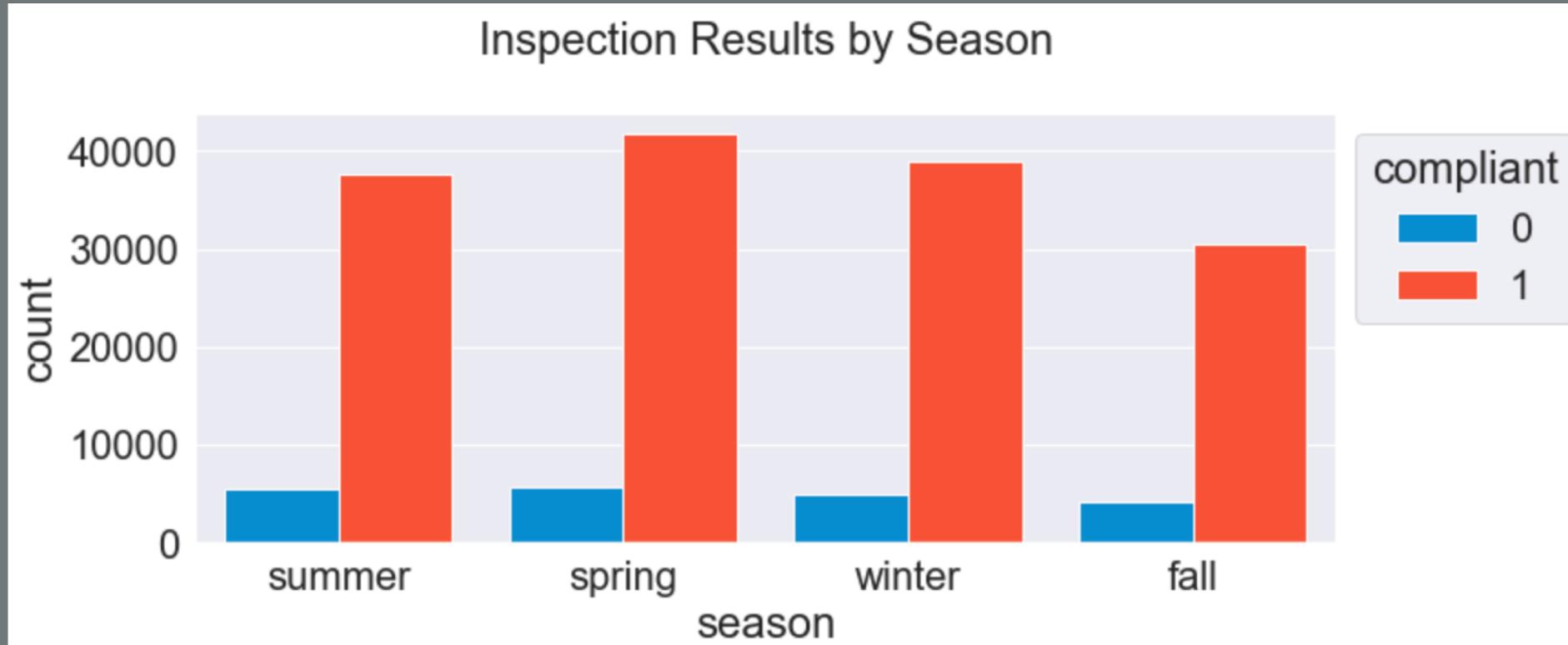


# COMPLIANCE OVERTIME IS IMPROVING



Since 2011 food establishments have obtained more passing grades, with only a slight decline between 2015 and 2016 when grades were consolidated and inspectors started using A grades for compliance.

# SEASONS AND INSPECTION RESULTS



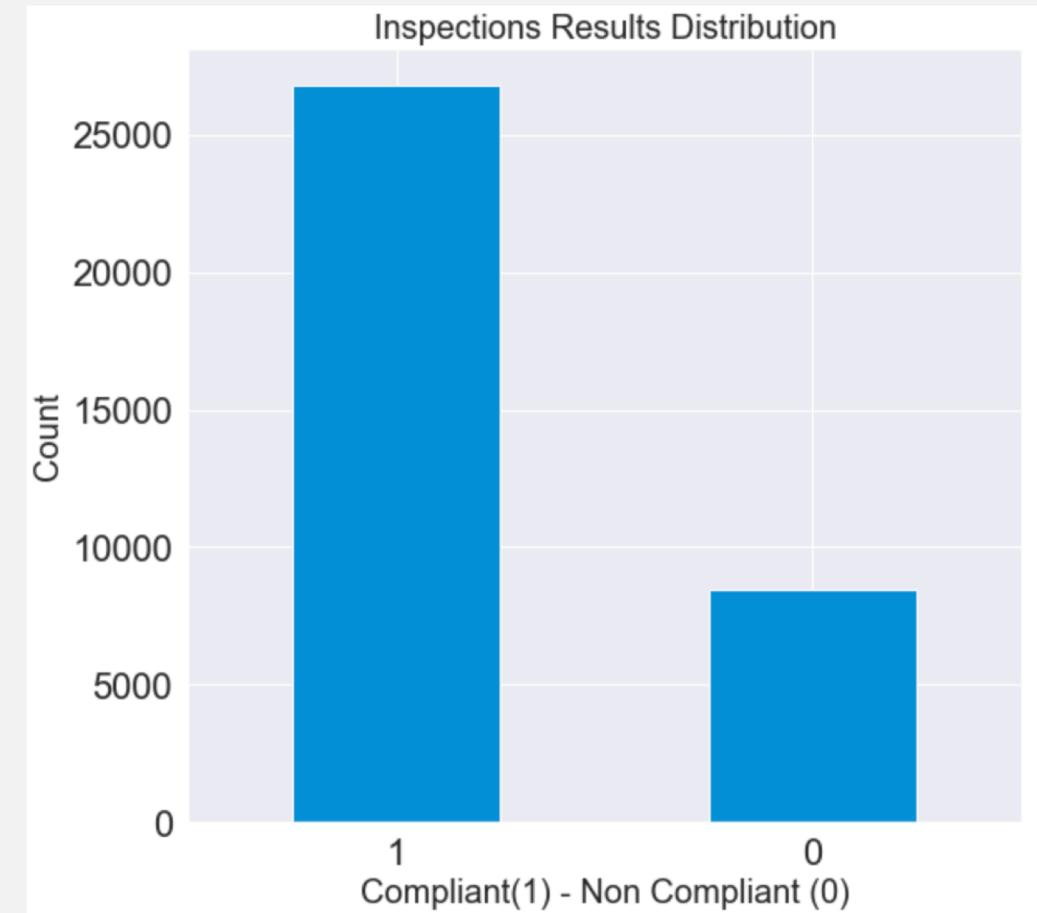
Based on the data, the lowest compliance rate happens during the Summer with 87.4% compared to winter when compliance is 88.7 %. Since the Summer months tend to be especially busy, being able to predict critical health violations, can minimize the risks of foodborne illness outbreaks.

# **PREDICTIVE ANALYTICS INSPECTIONS RESULTS AND YELP DATA**

## MERGED DATA

The distribution of compliance rate was maintained after the inspections and Yelp data was merged.

The proportion of passing grades is 76%, which is a fairly balanced distribution for the predictive models.



## PREDICTIVE MODELS

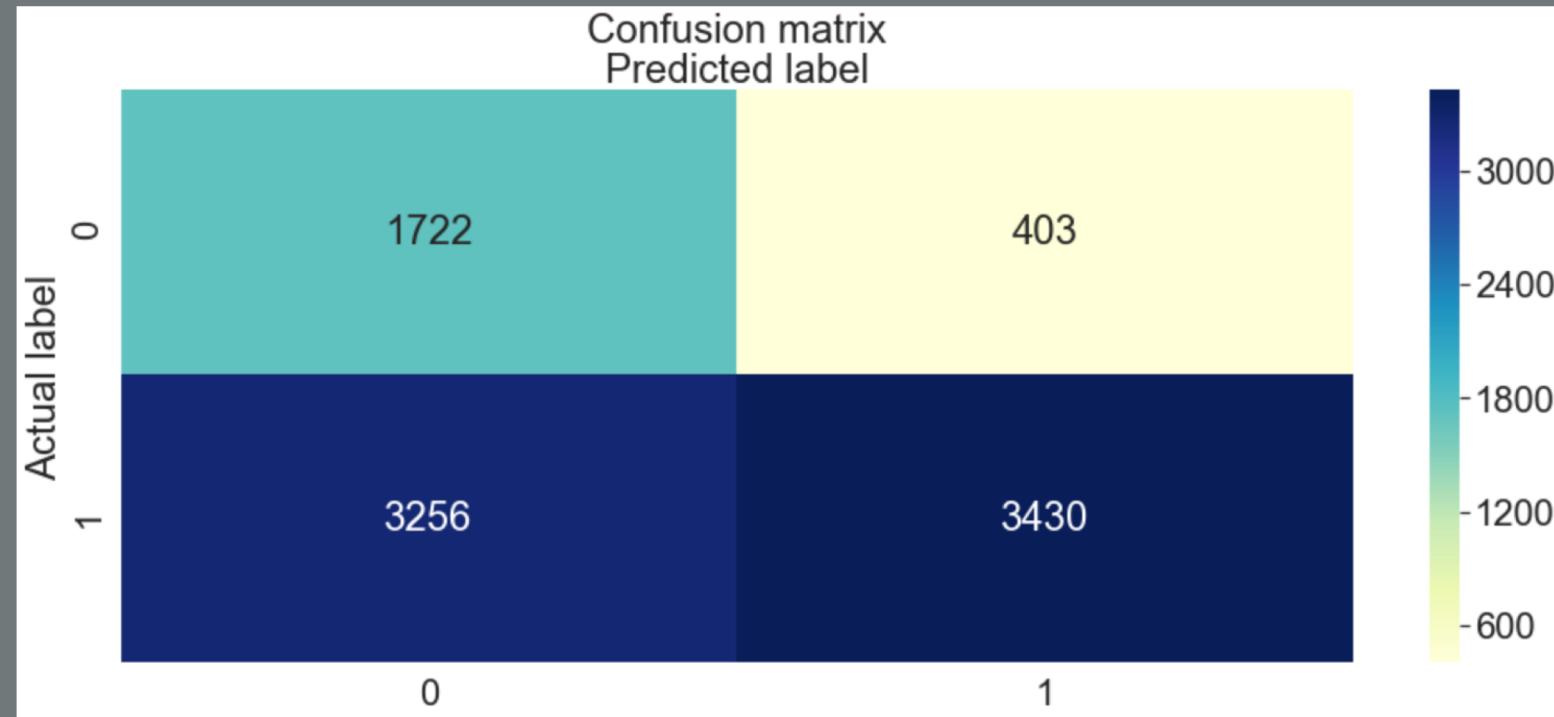
- The final dataset had 35,242 records and 92 attributes.
- Four models with two different algorithms were trained with slightly different predictors to compare the predictive power of each model.
- Logistic regression and Random Forest

# LOGISTIC REGRESSION I

	precision	recall	f1-score	support
0	0.00	0.00	0.00	2125
1	0.76	0.76	0.86	6686
micro avg	0.76	0.76	0.76	8811
macro avg	0.38	0.5	0.43	8811
weighted avg	0.58	0.76	0.65	8811

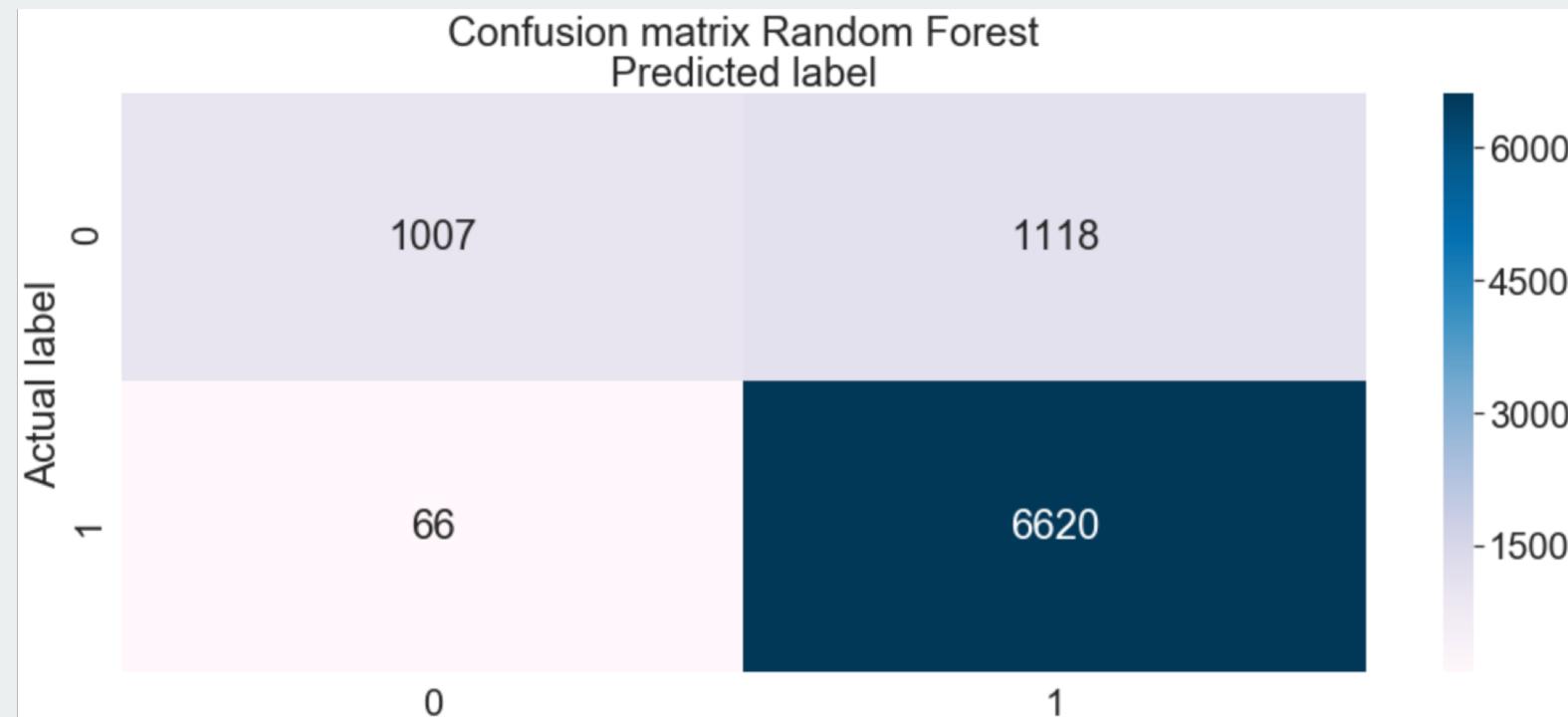
The first model failed to identify any non-compliance results.

## LOGISTIC REGRESSION 2



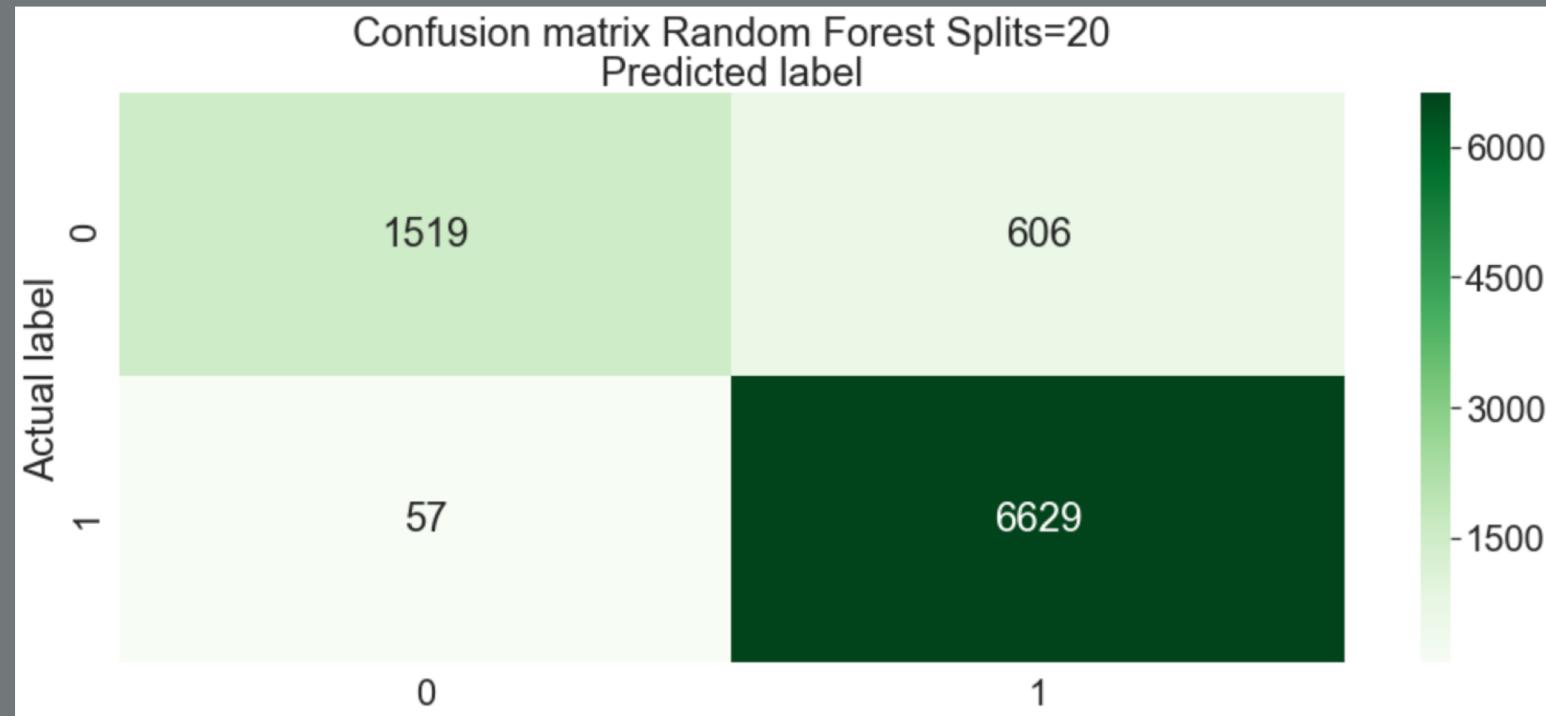
Using a 'balanced' class weight, the model identified 81% but it did not perform well with passing grades, identifying only 51%. Overall performance was only 58%.

# RANDOM FOREST MIN. SAMPLE SPLIT = 50



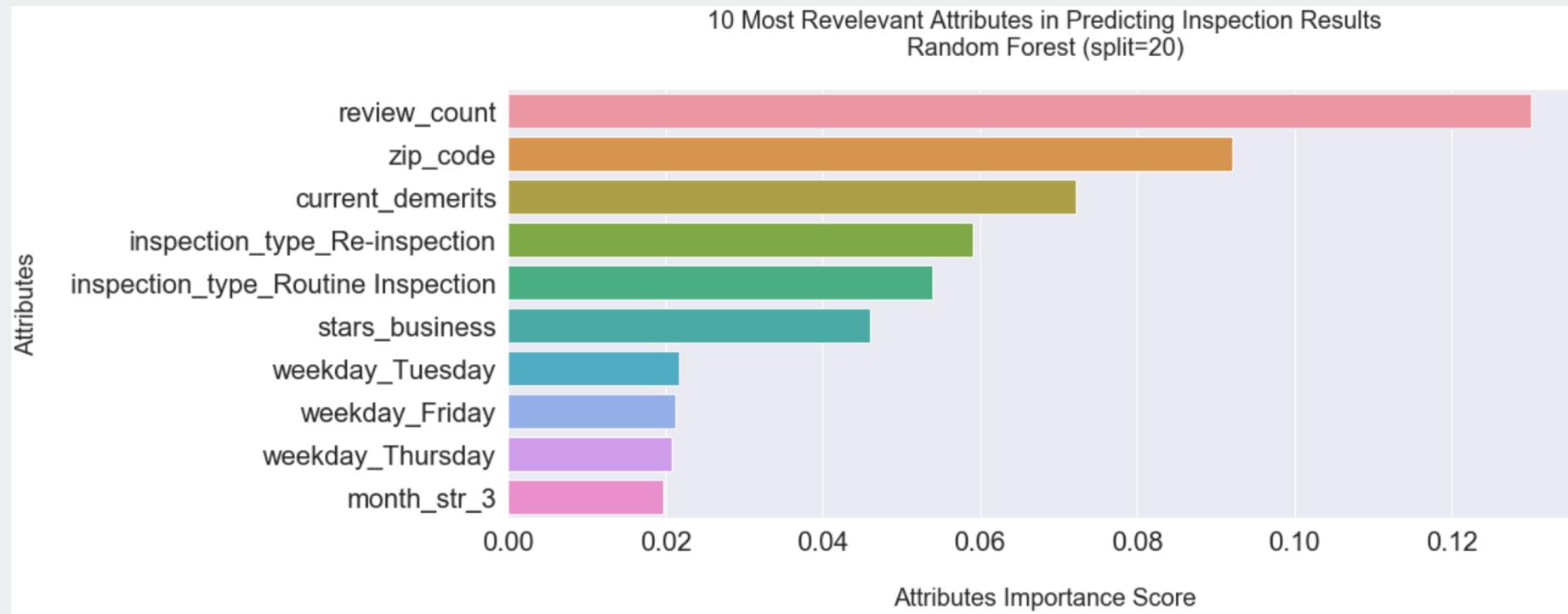
The accuracy of the model on unseen data was 86.65%. However, the model was able to identify only 47% of non-compliance results.

## BEST MODEL: RANDOM FOREST MIN. SAMPLE SPLIT = 20



The predictive accuracy of the model was 92.5% on unseen data. It accurately predicted 72% of the actual non-compliance results and 99% of compliance results.

# BEST MODEL: RANDOM FOREST



The most relevant predictors are related to the number of reviews, location, and inspection details, such as the type, day, current violation points, and less to the type of cuisine or specific rating given by users on a given day.

## BEST MODEL: MOST SIGNIFICANT PREDICTORS

- **Why Review counts?**
  - Better Grades brings more Customers and more reviews.
  - Popularity comes from standing out. Good or bad.
- **Why Zip code?**
  - Segmentation. Areas with predominant fine vs. casual dining restaurants. Generally, upscale restaurants would maintain a high level of health and safety standards.

## BEST MODEL: MOST SIGNIFICANT PREDICTORS

- **Why Current demerits?**
  - Restaurants that had over 40 points in the last inspection might not immediately go down to 0 points or vice versa.
- **Why Type of inspection** (re-inspection or routine inspection):
  - Although the compliance rate should increase in the first re-inspection, establishments with a significant number of critical violations may not pass a first re-inspection.
- **Number of stars the business receives:**
  - Indicates what customers see and experience.

## BEST MODEL: MOST SIGNIFICANT PREDICTORS

- **Why Tuesday, Friday, and Thursday ?**
  - Establishments may be closed on Monday, thus, this is the first to clean kitchens, serving stations, and bathrooms after a busy weekend.
  - Monday is the first available day for inspection.
  - Thursday and Friday could be produce delivery days, hence areas and surfaces may still be unclean when inspectors arrive.
- **Inspections in March (month\_str\_3):**
  - This explanatory variable is not as clear as others on the correlation of inspections in March and inspection results.

## CONCLUSIONS

The previous analysis used publicly available data on inspection results and social media data on business attributes and customer ratings, to predict food and safety violations, and improve inspection visits by prioritizing establishments that are likely to be non-compliant.