# Boulder County Restaurant Health Inspection Scores v.s.

**Google Ratings**

## Summary of Study

This study seeks to answer the question:

*Is there a correlation between restaurant health inspection scores and customer ratings?*

Using a publicly available dataset from [Boulder County Public Health](https://www.bouldercounty.org/families/food/restaurant-inspection-data/), as well as Google Places API, we created a database of regular restaurant inspections in Boulder County prior to 2020. We plan to use logistic regression to determine if there is a correlation between health inspection score and Google rating.

## Example Dashboard Layout Explore the Dataset

*Select Restaurant* ⇩ Avg. Inspection Score Avg. Google Rating

List of violations Example 1

Example 2

Example 3

Example 4

## Prediction Machine

*Enter Inspection Score* Predicted Avg. Google Rating

*Enter Google Rating* Predicted Avg. Inspection Score

\*We plan on using either Heroku or building the application using HTML/Flask hosted on one of our personal URLs.

# Exploration of Violations

### Critical Violations

Additionally, we will explore if there is a correlation between individual violations and Google rating. This will tell us if there are particular violations that seem to have an outsized effect on the guest experience.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 01a | Approved Source | 15 | 04a | Manual | 10 |
| 01b | Wholesome, free of spoilage | 15 | 04b | Mechanical | 10 |
| 01c | Cross-contamination | 20 | 04c | In-place | 10 |
| 01d | HACCP plan | 15 |  |  |  |
| 01e | Date Marking | 5 | **05** | **Water, Sewage, Plumbing Systems** |  |
| 01f | Consumer Advisory | 5 | 05a | Safe water source | 15 |
|  |  |  | 05b | Hot and cold water under pressure | 10 |
| **02** | **Personnel** |  | 05c | Backflow | 10 |
| 02a | Personnel with infections restricted | 20 | 05d | Sewage disposal | 15 |
| 02b | Wounds properly covered | 20 |  |  |  |
| 02c | Hands washed as needed | 20 | **06** | **Handwashing and Toilet Facilities** |  |
| 02d | Hygenic practices | 5 | 06a | Adequate number, location, design | 15 |
| 02e | Smoking, eating, drinking | 10 | 06b | Accessible | 10 |
| 02f | Training needed | 15 | 06c | Soap and drying devices | 5 |
| 02g | No bare hand contact | 20 |  |  |  |
|  |  |  | **07** | **Pest Control** |  |
| **03** | **Food Temperature Control** |  | 07a | Evidence of pests | 10 |
| 03a | Rapidly cool to 41 F or lower | 20 | 07b | Pesticide application | 15 |
| 03b | Rapidly reheat to 165 F or greater | 20 | 07c | Animals prohibited | 5 |
| 03c | Hot hold at 135 F or greater | 15 |  |  |  |
| 03d | Required cooking temperatures | 20 | **08** | **Poisonous or Toxic Items** |  |
| 03e | Cold hold at 41 F or lower | 15 | 08a | Properly stored | 10 |
| 03f | Food thermometer (probe type) | 5 | 08b | Properly labeled | 5 |
| 03g | Adequate equipment to food maintain | 20 | 08c | Toxic items properly used | 15 |
|  | temperatures |  |  |  |  |

We have chosen only critical criteria that are key to restaurant hygiene and food safety, rather than administrative and clerical criteria.

## Most Common Violations

Example 1

Example 2

Example 3

Example 4

## Most Impactful Violations

Example 1

Example 2

Example 3

Example 4

**01 Food Source Points**

**04 Sanitization Rinse Points**

# Data Exploration & Analysis

## Data Exploration:

* Original dataset: 471,209 rows
* Filter criteria:
  + Only restaurants, bars, or food trucks
  + Before January 1st, 2020
  + Only routine inspections
* After filtering: 220,888 rows
* Number of facilities: 895
* Number of inspection scores: 2687

## Analysis:

To do our analysis we will be taking the average score for each facility and compare it with the facility’s Google rating using logistic regression.

For the analysis of violations, we will filter the data down to only the violations with which we are interested, generate a list of all relevant violations for a given facility, and then compare them to the facility’s Google rating.

This analysis will use Random Forest learning to determine feature importance.

# Prediction Machine with Violations

**Select violations:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **01** | **Food Source** | **Points** | **04** | **Sanitization Rinse** | **Points** |
| * 01a | Approved Source | 15 | * 04a | Manual | 10 |
| * 01b | Wholesome, free of spoilage | 15 | * 04b | Mechanical | 10 |
| * 01c | Cross-contamination | 20 | * 04c | In-place | 10 |
| * 01d | HACCP plan | 15 |  |  |  |
| * 01e | Date Marking | 5 | **05** | **Water, Sewage, Plumbing Systems** |  |
| * 01f | Consumer Advisory | 5 | * 05a | Safe water source | 15 |
|  |  |  | * 05b | Hot and cold water under pressure | 10 |
| **02** | **Personnel** |  | * 05c | Backflow | 10 |
| * 02a | Personnel with infections restricted | 20 | * 05d | Sewage disposal | 15 |
| * 02b | Wounds properly covered | 20 |  |  |  |
| * 02c | Hands washed as needed | 20 | **06** | **Handwashing and Toilet Facilities** |  |
| * 02d | Hygenic practices | 5 | * 06a | Adequate number, location, design | 15 |
| * 02e | Smoking, eating, drinking | 10 | * 06b | Accessible | 10 |
| * 02f | Training needed | 15 | * 06c | Soap and drying devices | 5 |
| * 02g | No bare hand contact | 20 |  |  |  |
|  |  |  | **07** | **Pest Control** |  |
| **03** | **Food Temperature Control** |  | * 07a | Evidence of pests | 10 |
| * 03a | Rapidly cool to 41 F or lower | 20 | * 07b | Pesticide application | 15 |
| * 03b | Rapidly reheat to 165 F or greater | 20 | * 07c | Animals prohibited | 5 |
| * 03c | Hot hold at 135 F or greater | 15 |  |  |  |
| * 03d | Required cooking temperatures | 20 | **08** | **Poisonous or Toxic Items** |  |
| * 03e | Cold hold at 41 F or lower | 15 | * 08a | Properly stored | 10 |
| * 03f | Food thermometer (probe type) | 5 | * 08b | Properly labeled | 5 |
| * 03g | Adequate equipment to food maintain | 20 | * 08c | Toxic items properly used | 15 |
|  | temperatures |  |  |  |  |

## Predicted Google Rating:

★★★★★

## Inspection Rating:

### Excellent

Inspection Ratings Key: Excellent 0-19

Good 20-39

Fair 40-69

Marginal 70-99

Unacceptable >100

