Sarita Himthani

Zenil Patel

Jhanvi Munim

Professor Sridevi Ayloo

Data Science- 1 IS680

29 November 2024

**Unveiling Trends in Restaurant Health Inspections (2016-2019)**

Foodborne infections, which impact millions of people every year, can result from inappropriate food handling, preparation, or storage, making food safety an essential public health issue. Regular inspections by health officials guarantee adherence to safety requirements, and restaurants are essential in upholding these standards. These inspections assist reduce risks and safeguard consumers by assessing food safety procedures, allocating ratings, and highlighting infractions.

With an emphasis on inspection trends, infraction patterns, and compliance rates across neighborhoods, this research examines historical restaurant inspection data from 2016 to 2019. This study looks at more than 50,000 records in an effort to find useful information that will help important stakeholders. The results can be used by public health authorities to create focused interventions and distribute resources efficiently. While customers receive useful information to help them make educated dining decisions, restaurant operators may pinpoint typical compliance issues to enhance operations.

This analysis is significant because it can identify patterns, evaluate hazards, and guarantee accountability for food safety. Given the increasing intricacy of urban food systems and the amount of inspections, data-driven insights are crucial to promoting a more dependable and safe food service sector, which will eventually enhance customer confidence and public health.

**Data Origin:**

Records of past restaurant inspections from 2016 to 2019 make up the dataset. It was most likely taken by a local health department or other organization in charge of making sure food safety regulations are followed. A thorough examination of food safety trends is made possible by the data, which includes information about companies, inspection findings, infractions, and related dangers.

**Important variables to investigate:**

The dataset has twenty-two variables. Crucial ones for this investigation consist of:

**Inspection Score:** A numerical value that represents adherence to food safety regulations.

**Inspection Type:** Types include follow-ups, new ownership inspections, and routine inspections.

**Description of Violation:** Describes particular infractions found during inspections.

**Risk Category:** A classification of infractions according to their level of seriousness, such as low, moderate, or high risk.

For geographical analysis, geographic variables include supervisor districts, police districts, and neighborhoods.

Business names, addresses, and contact information are extra variables that are helpful for context.

**Procedure for Cleaning Data:**

To fix inconsistent and missing entries, the dataset needed to be cleaned. Among the steps were:

Discarding records that lack inspection scores since analysis depends on them.

Dates in the inspection\_date column are parsed and standardized.

Addressing conflicting or missing geographic data for analysis at the neighborhood level.

To guarantee correctness, duplicate entries are eliminated.

**Descriptive Statistics:**

There are more than 50,000 inspection records in the dataset. A reasonable level of compliance is indicated by the average score of 85.6, which is valid for 73% of these records. With 60% of inspections being "Routine - Unscheduled," this is the most common type of inspection. Violations are noted in about 25% of records, with "Unclean or degraded floors, walls, or ceilings" being one of the most frequent. Fifteen percent of all infractions are high-risk, indicating areas that require urgent attention. Inspection densities vary by neighborhood, according to data, with certain districts having noticeably greater non-compliance rates.

**Exploratory Data Analysis:**

This section presents an exploratory analysis of food safety inspection data to uncover violation patterns, risk categorization, and inspection performance. Four key questions guide the analysis, each answered through a focused visualization.

1. **What are the most common violations observed in the data?**

To understand the primary areas of concern in food safety, we examined the top 10 most frequently reported violations. Using a horizontal bar chart, we visualized the counts of these violations. The chart revealed that issues such as improper food handling, inadequate hygiene practices, and equipment sanitation deficiencies were the most prevalent.

This analysis helps stakeholders identify recurring problems and prioritize intervention efforts. For example, targeted training programs for food handlers could address improper food handling violations, while stricter equipment cleaning protocols could mitigate sanitation-related issues.

1. **How are violations distributed across different risk categories?**

To assess the severity of reported violations, we categorized them into high, medium, and low-risk categories and visualized the distribution using a bar chart. High-risk violations— such as those involving contamination, foodborne illness hazards, or unsafe temperature control— accounted for the majority of cases. Medium-risk violations included infractions that could lead to potential safety concerns if left unaddressed, while low-risk violations were less likely to pose immediate health risks.

This analysis underscores the critical need to focus resources and inspection efforts on high-risk categories, as they have the greatest potential impact on public health.

1. **How are inspection scores distributed geographically?**

We created an interactive map to explore the spatial distribution of inspection scores. Each business location was marked on the map, with markers color-coded to represent inspection performance: green for scores of 90+ (excellent), orange for scores between 70 and 89 (moderate), and red for scores below 70 (needs improvement).

The map revealed distinct clusters of underperforming businesses in certain areas, suggesting potential regional challenges such as insufficient compliance enforcement or resource limitations. This insight allows regulators to allocate resources strategically, focusing on regions with consistently low scores.

1. **Where are businesses located geographically?**

A scatterplot of business latitude and longitude provided a comprehensive view of the geographical spread of businesses included in the dataset. While not explicitly tied to violations or scores, this visualization offered foundational insights into the spatial coverage of inspections. By overlaying this information with data on inspection outcomes or violations, future analysis could identify whether specific locations or business clusters are more prone to violations, enabling a more targeted approach to inspections and compliance monitoring.

**Summary:**

The analysis of San Francisco’s restaurant inspection data from 2016 to 2019 reveals significant insights into food safety and compliance trends. Key patterns include compliance behaviors, public health risks, and geographical clusters of violations and inspection scores. High-risk violations, which directly impact the transmission of foodborne illnesses, require targeted attention. Meanwhile, moderate and low-risk violations present opportunities for improvement. The findings emphasize the need strategic resource, allocation, business education, and robust inspection protocols to ensure food safety.

Although the dataset is no longer updated due to the discontinuation of the LIVES standard, it remains a valuable historical resource for informing policies and interventions. For current data, stakeholders should consult the Health Department’s official resources.

**Policy recommendations:**

Policy recommendations based on the analysis include targeted actions to address key issues. Educational campaigns and stricter regulations should be developed to address the most common violations. High-risk businesses should be prioritized through monitoring and assistance programs. For areas with lower inspection scores, targeted outreach and support should be introduced. Lastly, inspection resources should be strategically deployed in regions with higher business densities to ensure efficient coverage and food safety compliance.

**References:**

1. <https://www.cdc.gov>
2. <https://www.fda.gov>
3. "Restaurant Scores in SF." *DataSF*, City and County of San Francisco. Accessed December 1, 2024. <https://data.sfgov.org/Health-and-Social-Services/Restaurant-Scores-in-SF/6ud7-8ksb>.
4. "[Historical] Restaurant Inspection Scores (2016-2019)." *DataSF*, City and County of San Francisco. Accessed December 1, 2024. <https://data.sfgov.org/Health-and-Social-Services/-Historical-Restaurant-Inspection-Scores-2016-2019/pyih-qa8i>.
5. "Open Restaurant Inspection Data and Public Health." *Public Health Reports*. Accessed December 1, 2024. (Search result discussing the utility and insights from restaurant inspection datasets, especially for trends and public health risks).