

Optimizing Food Safety Inspections in Chicago

Enhancing the efficiency of inspections to improve overall public health in the city.

By Linda Fung (Data Engineering Final Project)





Executive Summary

In a city as bustling as Chicago, ensuring food safety standards across its diverse array of food establishments is paramount. However, the efficient allocation of resources for food safety inspections remains a challenge. This executive summary outlines a data-driven approach to optimize food safety inspections in Chicago by leveraging insights from food inspections datasets.



Research Objectives

Problem Statement

The primary problem to be addressed is the optimization of food safety inspections in Chicago to mitigate the risk of foodborne illnesses by:

- **Identifying higher-risk establishments and neighborhoods**
- **Prioritizing inspection resources to areas with the most need**

Data

The research will utilize two primary datasets:

- **City of Chicago Food Inspections Dataset**
- **City of Chicago Business Licenses Dataset**



Methodologies

Raw Data Cleaning & Preparation

OpenRefine

The initial dataset underwent a cleaning process aimed at eliminating redundant columns, filtering out rows with excessive missing data, and resolving inconsistencies and errors within the dataset to enhance its quality and reliability.



Data Modeling & Database Creation

MySQL

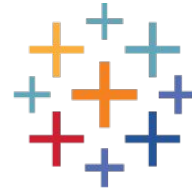
A data model was developed to organize the information, normalization was performed to 3NF, and structured data was imported into a relational database. Queries were then executed to address specific questions and extract relevant insights.



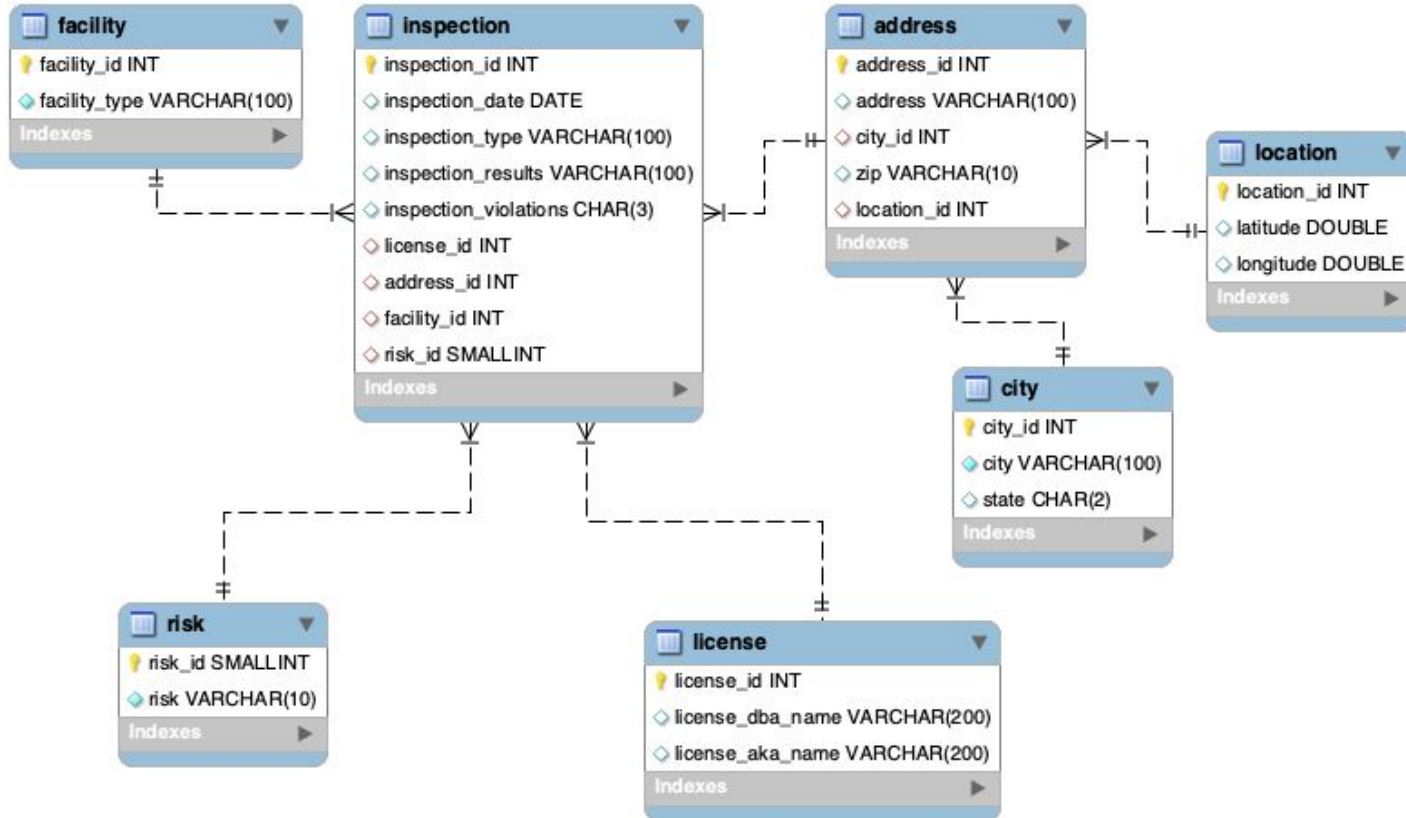
Data Visualization & Analytics

Tableau

An interactive dashboard with dynamic visuals was created to draw additional insights as well as support insights driven from the SQL queries.



EER Model



Data Analysis: SQL Queries





Overage Averages

What are the overall average risk, passing and failing rates?

- Higher risk establishments seem to be more common.
- Inspections seem to pass most of the time.

| avg_risk | avg_pass_rate | avg_fail_rate |
|----------|---------------|---------------|
| 84.0705 | 68.0609 | 19.7183 |

```
SELECT
    AVG(CASE
        WHEN risk_id = 1 THEN 100
        WHEN risk_id = 2 THEN 50
        WHEN risk_id = 3 THEN 0
    END) AS avg_risk,
    AVG(CASE WHEN inspection_results LIKE 'PASS%'
        THEN 1 ELSE 0 END) * 100
    AS avg_pass_rate,
    AVG(CASE WHEN inspection_results = 'FAIL'
        THEN 1 ELSE 0 END) * 100
    AS avg_fail_rate
FROM
    inspection;
```



Risk and Failure Rate

Are higher risk restaurants more prone to failing? Is there a correlation between risk and fail rate?

- There doesn't seem to be a strong correlation between risk and failing rate.
- Surprisingly, low risk establishments have a slightly higher fail rate.

| risk | total_inspectio... | failed_inspectio... | failure_rate |
|--------|--------------------|---------------------|--------------|
| HIGH | 196230 | 38032 | 19.3813 |
| MEDIUM | 48283 | 9329 | 19.3215 |
| LOW | 17614 | 4326 | 24.5600 |

```
SELECT
  r.risk,
  COUNT(*) AS total_inspections,
  SUM(CASE WHEN i.inspection_results = 'FAIL'
    THEN 1 ELSE 0 END) AS failed_inspections,
  SUM(CASE WHEN i.inspection_results = 'FAIL'
    THEN 1 ELSE 0 END) / COUNT(*) * 100 AS failure_rate
FROM
  risk r
JOIN
  inspection i ON r.risk_id = i.risk_id
GROUP BY risk;
```


Facility Types and Failure Rate

Are certain facility types more prone to failing?
Also, how high is their respective risk? Looking at facility types with a decent amount of inspections (200+):

- Wholesale, liquor, and tavern facility types have the highest failure rates.
- The facility types with failure rates higher than the average also tend to be high risk.

| facility_type | total_inspectio... | failed_inspectio... | failure_rate | risk_scale |
|-------------------------|--------------------|---------------------|--------------|------------|
| WHOLESALE | 510 | 169 | 33.1373 | 35.4902 |
| LIQUOR | 1102 | 365 | 33.1216 | 4.6279 |
| TAVERN | 414 | 128 | 30.9179 | 4.3478 |
| MOBILE FOOD DISPENSER | 782 | 216 | 27.6215 | 8.1202 |
| GAS STATION | 222 | 60 | 27.0270 | 17.3423 |
| MOBILE FOOD PREPARER | 908 | 234 | 25.7709 | 53.3040 |
| LONG TERM CARE | 2000 | 471 | 23.5500 | 99.9000 |
| GROCERY STORE | 32808 | 7536 | 22.9700 | 48.3602 |
| DAYCARE COMBO | 679 | 148 | 21.7968 | 99.8527 |
| SCHOOL | 16687 | 3624 | 21.7175 | 94.8703 |
| BAKERY | 3849 | 829 | 21.5381 | 70.5508 |
| DAYCARE UNDER 2 YEARS | 303 | 65 | 21.4521 | 100.0000 |
| DAYCARE 2 6 YEARS | 3124 | 616 | 19.7183 | 99.2318 |
| DAYCARE ABOVE AND UN... | 3597 | 683 | 18.9880 | 99.8471 |

```
SELECT
    f.facility_type,
    COUNT(*) AS total_inspections,
    SUM(CASE WHEN i.inspection_results = 'FAIL'
        THEN 1 ELSE 0 END) AS failed_inspections,
    SUM(CASE WHEN i.inspection_results = 'FAIL'
        THEN 1 ELSE 0 END) / COUNT(*) * 100 AS failure_rate,
    AVG(CASE
        WHEN r.risk_id = 1 THEN 100
        WHEN r.risk_id = 2 THEN 50
        WHEN r.risk_id = 3 THEN 0
    END) AS risk_scale
FROM
    facility f
JOIN
    inspection i
    ON f.facility_id = i.facility_id
JOIN
    risk r
    ON i.risk_id = r.risk_id
GROUP BY facility_type
HAVING total_inspections >= 200
ORDER BY failure_rate DESC;
```

Zip Codes and Failure Rate

Are certain zip codes more prone to failing? Also, how high is their respective risk? Looking at facility types with a decent amount of inspections (200+):

- The zip codes with failure rates higher than the average also tend to have very high risk.

| zip | total_inspecio... | failed_inspecio... | failure_rate | risk_scale |
|-------|-------------------|--------------------|--------------|------------|
| 60619 | 5217 | 1380 | 26.4520 | 73.9218 |
| 60628 | 4680 | 1197 | 25.5769 | 76.8376 |
| 60615 | 3002 | 753 | 25.0833 | 83.7941 |
| 60621 | 2263 | 561 | 24.7901 | 72.5144 |
| 60620 | 4988 | 1234 | 24.7394 | 75.3007 |
| 60637 | 2966 | 722 | 24.3425 | 79.1133 |
| 60636 | 3034 | 738 | 24.3243 | 70.4351 |
| 60644 | 3217 | 762 | 23.6867 | 72.6764 |
| 60623 | 6097 | 1439 | 23.6018 | 81.6303 |
| 60617 | 5179 | 1208 | 23.3250 | 83.2593 |
| 60639 | 7033 | 1621 | 23.0485 | 82.3262 |
| 60651 | 3696 | 848 | 22.9437 | 69.6023 |

```
SELECT
  a.zip,
  COUNT(*) AS total_inspections,
  SUM(CASE WHEN i.inspection_results = 'FAIL'
    THEN 1 ELSE 0 END) AS failed_inspections,
  SUM(CASE WHEN i.inspection_results = 'FAIL'
    THEN 1 ELSE 0 END) / COUNT(*) * 100 AS failure_rate,
  AVG(CASE
    WHEN r.risk_id = 1 THEN 100
    WHEN r.risk_id = 2 THEN 50
    WHEN r.risk_id = 3 THEN 0
  END) AS risk_scale
FROM
  address a
JOIN
  inspection i
  ON a.address_id = i.address_id
JOIN
  risk r
  ON i.risk_id = r.risk_id
GROUP BY zip
HAVING total_inspections >= 200
ORDER BY failure_rate DESC;
```



Inspections Over The Years

Are there any trends in the amount of inspections over the years? How about the pass rates?

- There was a slight increase in overall inspections between 2013 and 2017.
- Passing rates seem to be trending downward over the years.

| inspection_year | total_inspections | pass_count | pass_rate |
|-----------------|-------------------|------------|-----------|
| 2010 | 16777 | 12459 | 74.2624 |
| 2011 | 17689 | 12960 | 73.2659 |
| 2012 | 17188 | 12102 | 70.4096 |
| 2013 | 19562 | 13464 | 68.8273 |
| 2014 | 20615 | 14698 | 71.2976 |
| 2015 | 20338 | 14098 | 69.3185 |
| 2016 | 22437 | 15422 | 68.7347 |
| 2017 | 21261 | 14253 | 67.0382 |
| 2018 | 16979 | 11432 | 67.3302 |
| 2019 | 18786 | 12388 | 65.9427 |
| 2020 | 14851 | 9594 | 64.6017 |
| 2021 | 15677 | 10021 | 63.9217 |
| 2022 | 16685 | 10619 | 63.6440 |
| 2023 | 18010 | 11457 | 63.6147 |

```
SELECT
    YEAR(inspection_date) AS inspection_year,
    COUNT(*) AS total_inspections,
    SUM(CASE WHEN inspection_results LIKE 'PASS%'
        THEN 1 ELSE 0 END) AS pass_count,
    SUM(CASE WHEN inspection_results LIKE 'PASS%'
        THEN 1 ELSE 0 END) / COUNT(*) * 100 AS pass_rate
FROM
    inspection
WHERE YEAR(inspection_date) < 2024
GROUP BY inspection_year
ORDER BY inspection_year;
```

Inspections and Day of the Week

Which days of the week are busiest for inspections?
Also, what are their pass rates?

- Weekdays are busiest for inspections with Tuesday, Thursday, and Friday leading, in that order.
- Pass rates are higher toward the end of the week on Saturday, Friday, and Thursday.

| day_of_week | total_inspections | pass_count | pass_rate |
|-------------|-------------------|------------|-----------|
| Monday | 43246 | 29071 | 67.2224 |
| Tuesday | 59693 | 40217 | 67.3731 |
| Wednesday | 48602 | 33311 | 68.5383 |
| Thursday | 58817 | 40012 | 68.0280 |
| Friday | 51653 | 35710 | 69.1344 |
| Saturday | 104 | 77 | 74.0385 |
| Sunday | 12 | 8 | 66.6667 |

```
SELECT
    DAYNAME(inspection_date) AS day_of_week,
    COUNT(*) AS total_inspections,
    SUM(CASE WHEN inspection_results LIKE 'PASS%'
        THEN 1 ELSE 0 END) AS pass_count,
    SUM(CASE WHEN inspection_results LIKE 'PASS%'
        THEN 1 ELSE 0 END) / COUNT(*) * 100 AS pass_rate
FROM
    inspection
GROUP BY day_of_week
ORDER BY FIELD(day_of_week,
    'Monday', 'Tuesday', 'Wednesday', 'Thursday',
    'Friday', 'Saturday', 'Sunday');
```

Violations and Passing Rate

What are the number of businesses with violations by facility type and their passing rate?

- The establishments with the most violations are slightly above average in passing rates.
- Restaurants, grocery stores, schools tend to have the most violations.

| facility_type | count_with_violations | pass_rate |
|---------------------------------|-----------------------|-----------|
| RESTAURANT | 134474 | 75.7604 |
| GROCERY STORE | 23835 | 69.9308 |
| SCHOOL | 14295 | 74.8164 |
| CHILDRENS SERVICES FACILITY | 4277 | 75.9878 |
| BAKERY | 2909 | 71.8804 |
| DAYCARE ABOVE AND UNDER 2 YEARS | 2729 | 75.3390 |
| DAYCARE 2 6 YEARS | 2134 | 72.3993 |
| LONG TERM CARE | 1677 | 72.0930 |
| CATERING | 1072 | 77.9851 |
| LIQUOR | 622 | 48.0707 |
| MOBILE FOOD PREPARER | 576 | 59.8958 |
| GOLDEN DINER | 564 | 75.7092 |

```
SELECT
    f.facility_type,
    COUNT(*) AS count_with_violations,
    SUM(CASE WHEN inspection_results LIKE 'PASS%'
        THEN 1 ELSE 0 END) / COUNT(*) * 100 AS pass_rate
FROM
    facility f
JOIN
    inspection i
    ON f.facility_id = i.facility_id
WHERE i.inspection_violations = "YES"
GROUP BY facility_type
ORDER BY count_with_violations DESC;
```



Zip Codes and Violations

Which zip codes had the most violations?


- The zip codes with the most violations have passing rates that are slightly above average.

| zip | count_with_violations | pass_rate |
|-------|-----------------------|-----------|
| 60614 | 7095 | 73.2770 |
| 60647 | 6737 | 71.6936 |
| 60657 | 6403 | 75.0586 |
| 60618 | 5977 | 78.7519 |
| 60611 | 5893 | 80.8926 |
| 60622 | 5729 | 75.1091 |
| 60608 | 5607 | 71.3572 |
| 60625 | 5421 | 72.0347 |
| 60607 | 5412 | 75.6652 |
| 60639 | 5284 | 70.2687 |
| 60640 | 5258 | 72.5371 |

```
SELECT
    a.zip,
    COUNT(*) AS count_with_violations,
    SUM(CASE WHEN inspection_results LIKE 'PASS%'
        THEN 1 ELSE 0 END) / COUNT(*) * 100 AS pass_rate
FROM
    address a
JOIN
    inspection i
    ON a.address_id = i.address_id
WHERE inspection_violations = "YES"
GROUP BY a.zip
ORDER BY count_with_violations DESC;
```



Data Analysis: Data Visualization



Data Analysis & Visualization

- Passing rates have been trending downward over the years.
- The northeastern section of Chicago has the most establishments with violations.
- Fail rates are slightly higher in areas of southeastern Chicago.
- When analyzing levels of risk alone, there doesn't seem to be a correlation with pass or fail rates.
- However, when looking at facility types, the facilities with the lowest pass rates also tend to be high risk.

Average
Pass
Rate

68.06

Average
Fail
Rate

19.72

Average
Risk

84.07

Risk

- ☒ (All)
☐ HIGH
☒ LOW
☐ MEDIUM

Inspection Violations

- ☒ (All)
☒ Null
☒ NO
☒ YES

Zip

(All)

Inspection Type

(All)

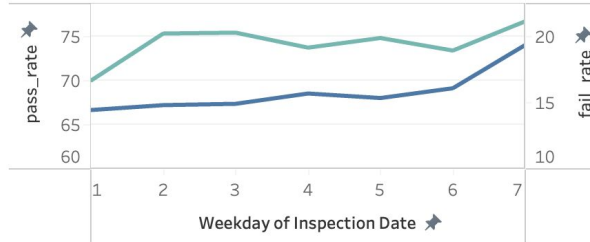
Inspection Results

(All)

Facility Type

(All)

Day of Week



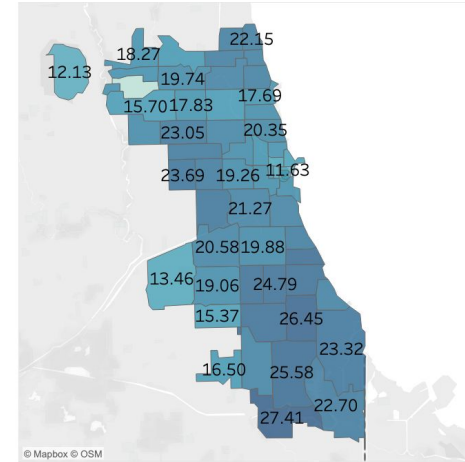
Facility Type

| Facility Type | Count of inspe.. | pass_rate | risk_scale |
|----------------------|------------------|-----------|------------|
| BANQUET HALL | 242 | 46 | 84 |
| LIQUOR | 1,102 | 53 | 5 |
| TAVERN | 414 | 53 | 4 |
| GAS STATION | 222 | 57 | 17 |
| WHOLESALE | 510 | 59 | 35 |
| MOBILE FOOD PREPA.. | 908 | 63 | 53 |
| DAYCARE COMBO | 679 | 63 | 100 |
| GROCERY STORE | 32,808 | 64 | 48 |
| MOBILE FOOD DISPEN.. | 782 | 65 | 8 |
| BAKERY | 3,849 | 65 | 71 |
| CATERING | 1,654 | 65 | 97 |
| DAYCARE 2-6 YEARS | 3,124 | 68 | 99 |

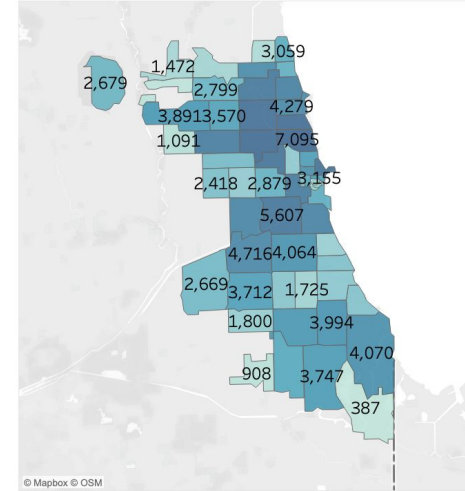
Year & Pass
Rate



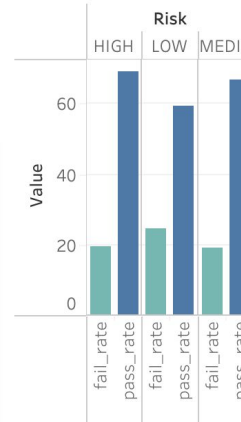
Zip Code and Fail Rate



Zip Code and Violations



Risk



Wrap Up: Conclusions





Recommendations

Corrective Measures

- **Geographic Focus:** Prioritize inspections and interventions in southeastern Chicago, which contains regions with failure rates higher than average. Implement preventive measures and educational campaigns tailored to the specific needs of these communities to improve food safety standards.
- **High-Risk Establishments:** Allocate more resources and attention to establishments classified as wholesale, liquor, and taverns, as they have shown to have the highest failure rates. Implement targeted inspections and enforcement measures to improve compliance in these establishments.
- **Violation Enforcement:** Restaurants, grocery stores, and schools tend to have the most violations. Focus efforts on the northeastern section of Chicago, where the most establishments with violations are located. Strengthen enforcement measures for establishments with repeated violations, including fines, penalties, and license suspensions for non-compliance.
- **Weekday Inspection Scheduling:** Given that weekdays are busiest for inspections and pass rates are higher toward the end of the week, consider adjusting inspection schedules to allocate more resources during peak inspection days. This may help improve efficiency and increase pass rates.

Scope for Improvement

- **Time-Based Analysis:** Investigate the reasons behind the slight increase in overall inspections between 2013 and 2017 and the downward trend in passing rates over the years. Review inspection protocols, training programs, and enforcement strategies to ensure effectiveness and address any underlying issues contributing to the decline in passing rates.



Lessons Learned

- **Continuous monitoring** and evaluation of food safety inspections are essential for identifying emerging risks, patterns, and addressing compliance gaps proactively.
- **Data-driven decision-making**, supported by in-depth analytics, data tools, and reporting capabilities, can provide valuable insights for improving food safety results.
- By **implementing** these **corrective measures** and applying lessons learned from past experiences, food, health, and safety agencies can work together to enhance food safety practices..

Thank you!

