C601

What factors drive restaurant inspection failures in NYC?

Data-driven strategies for improving food safety outcomes

Morgan, Luai, Briana

Agenda

- Background

- Objectives

- Data Visualizations

- Summary

- Recommendations

- Conclusion

coopcareers.org

New York is a cultural food capital showcasing a diverse body of restaurant types

New York City boasts
25,000+ restaurants
reflecting its rich cultural
and culinary diversity,
but that scale also creates
public health challenges.

Background

What is a restaurant inspection?
A restaurant inspection is an official evaluation conducted by local health department officials to ensure restaurants and food service establishments are following food safety regulations and health codes.

Why are restaurant inspections important?
Restaurant inspections are crucial for several key reasons:

- **Public health protection:** prevent foodborne illnesses that can cause serious health problems
- **Food safety enforcement:** ensure restaurants follow essential safety protocols
- **Early problem detection:** identify potential hazards from infestations to equipment failures
- **Consumer confidence:** transparency and peace of mind about the safety standards
- **Industry standards:** maintain consistent safety practices and accountability

Research Question & Objectives

What factors contribute to restaurant inspection failures in NYC, and what targeted interventions could improve food safety outcomes across different cuisines and neighborhoods?

| Objectives |
| --- |
| ● Identify geographic patterns in inspection failures |
| ● Analyze cuisine-specific violation trends |
| ● Determine most common violation types |
| ● Develop actionable, data-driven recommendations |

Data description
& headers Dtypes

Header	Dtype	represents
CAMIS	Text	The unique ID for each restaurant. 10-digit integer, static per restaurant permit
DBA	Text	The "Doing Business As" name of the restaurant. (Public business name)
BORO	Text	The borough where the restaurant is located (e.g., Manhattan, Brooklyn, Bronx, Queens, Staten Island). NOTE: There may be discrepancies between zip code and listed boro due to differences in an establishment's mailing address and physical location according to DOHMH FAQ
BUILDING	Text	The building number of the restaurant's address.
STREET	Text	The street name of the restaurant's address.
ZIPCODE	Text	The five-digit zip code.
PHONE	Text	The restaurant's phone number.
CUISINE DESCRIPTION	Text	The type of cuisine served at the restaurant.
INSPECTION DATE	Floating Timestamp	The date of the inspection. "1/1/1900 "= not yet had an inspection
ACTION	Text	The action taken as a result of the inspection (e.g., "Violations were cited," "No violations were recorded," "Closed", "Reopened").

Header	Dtype	represents
VIOLATION CODE	Text	A specific code assigned to each violation found.
VIOLATION DESCRIPTION	Text	A detailed description of the violation corresponding to the code.
CRITICAL FLAG	Text	Indicates if a violation is considered "Critical" or "Not Critical" or "Not Applicable".
SCORE	Number	The numeric score assigned to the restaurant's inspection. A lower score is better.
GRADE	Text	The letter grade assigned based on the score (N = Not Yet Graded, A = Grade A, B = Grade B, C = Grade C, Z = Grade Pending, P= Grade Pending issued on re-opening following an initial inspection that resulted in a closure).
GRADE DATE	Floating Timestamp	The date the grade was issued.
RECORD DATE	Floating Timestamp	The date the inspection record was entered into the system.
INSPECTION TYPE	Text	The type of inspection (e.g., "Cycle Inspection," "Re-inspection").
Latitude	Number	The geographic latitude of the restaurant.
Longitude	Number	The geographic longitude of the restaurant.
Community Board	Text	The community board district number.

Header	Dtype	represents
Council District	Text	The city council district number.
Census Tract	Text	The census tract identifier.
BIN	Text	The Building Identification Number.
BBL	Text	The Borough, Block and Lot number.
NTA	Text	The Neighborhood Tabulation Area.

Data cleaning procedure
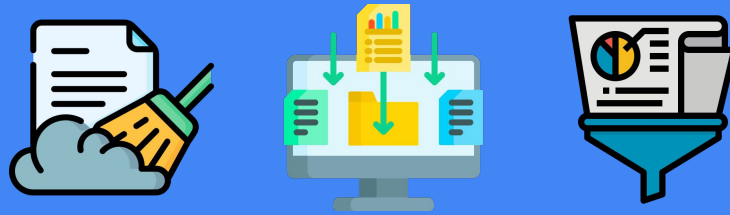
Data preparation steps

1. **Data Overview**
   - DOHMH Dataset included 319,806 rows and 26 columns from NYC Department of Health
   - Contained restaurant details, inspection results, and violation codes
2. **Data Structuring**
   - Split the large dataset into smaller, manageable batches for efficiency
   - Created a Primary Key to uniquely identify each restaurant and link records accurately
3. **Data Quality Checks**
   - Verified and standardized data types (dates, text, numbers)
   - Flagged and corrected invalid or missing entries, such as:
     - Placeholder dates → 1/1/1900
     - Missing coordinates → (0,0)
     - Invalid borough entries → BORO = 0
4. **Cleaning & Validation**
   - Removed duplicates and filled null values where appropriate
   - Ensured the final dataset was accurate, consistent, and ready for analysis in Excel and Tableau

https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j/about_data
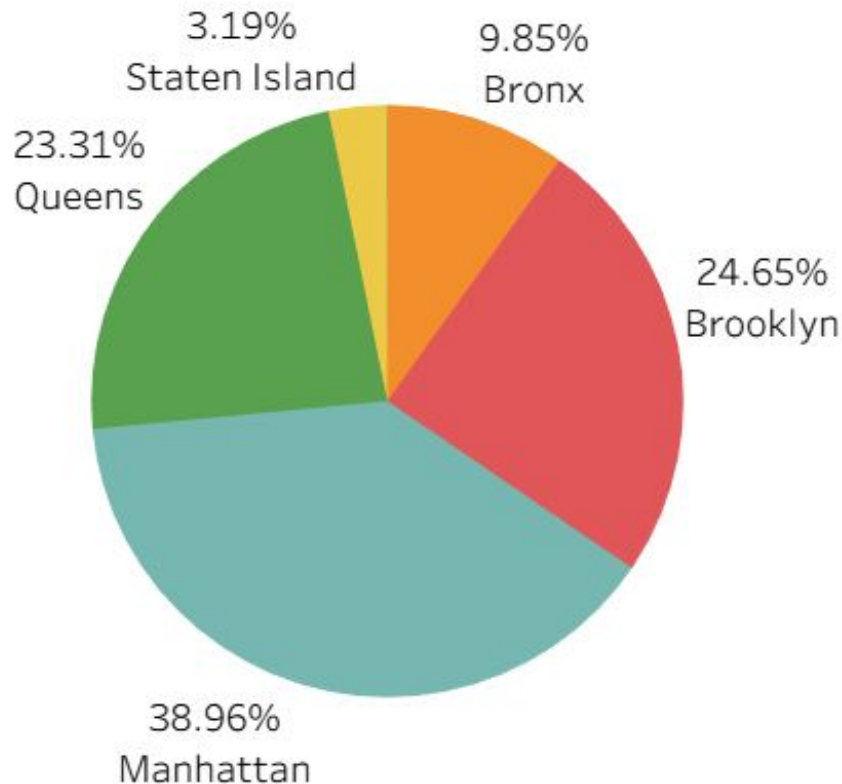
Key Findings

Finding #1 - Geography
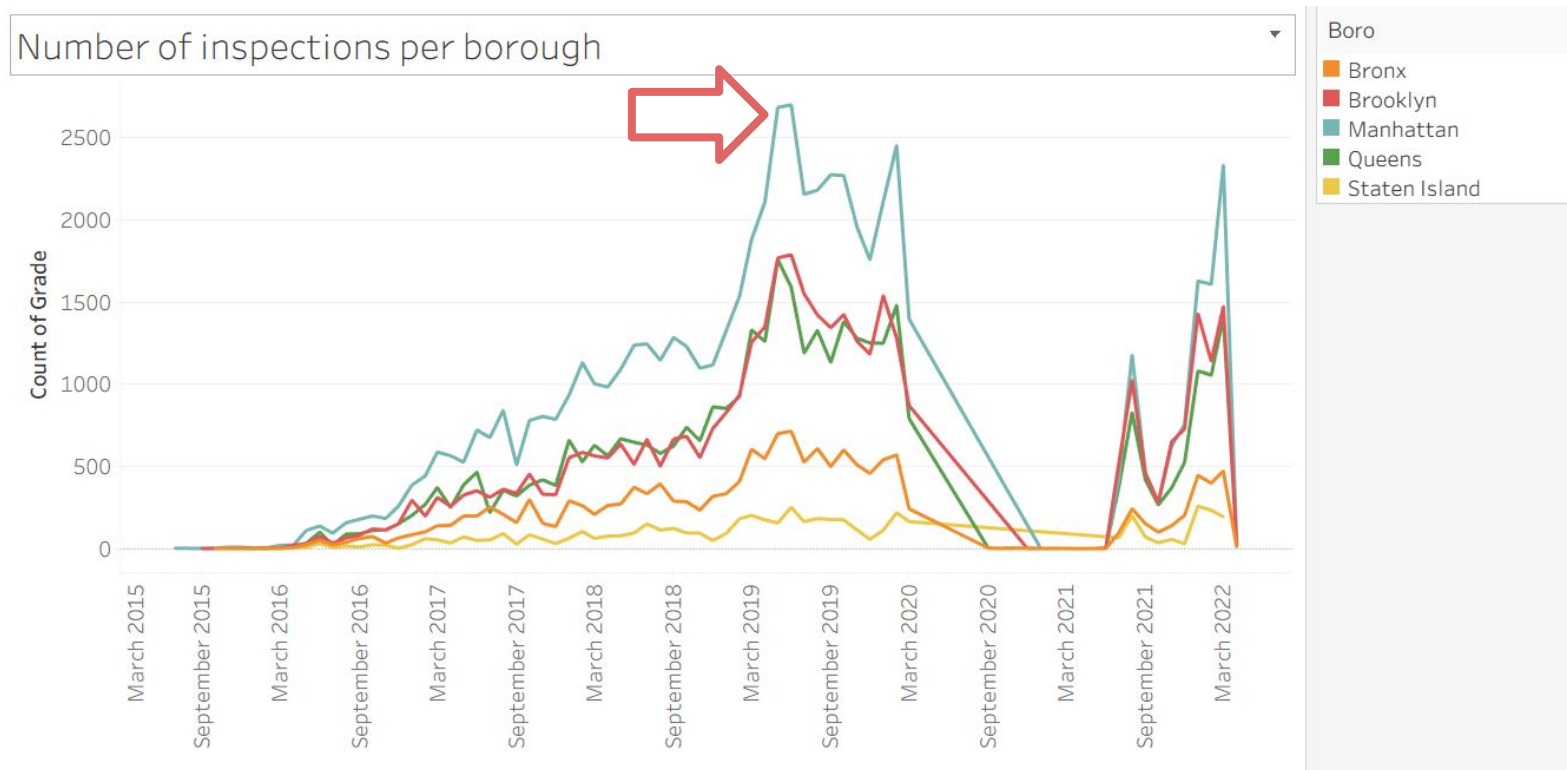
Manhattan dominates inspection failures

Potential reasons:

- High restaurant density in these areas
- Tourist-heavy locations may cause more frequent inspections
- Older building infrastructure in some Manhattan neighborhoods

Inspection Trends Over Time by Borough

Manhattan maintains highest failure rate consistently from 2015-2022, with brief COVID-19 disruption in 2020

https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j/about_data

Top 10 ZIP Codes Account for 18% of All Failures

Top 3 Hotspots:

1. 10003 (East Village) - 7,506 violations
2. 10019 (Midtown West) - 7,058 violations
3. 10036 (Times Square) - 6,603 violations

Insight:

- 8 of top 10 hotspots are in Manhattan's high-traffic tourist and commercial districts
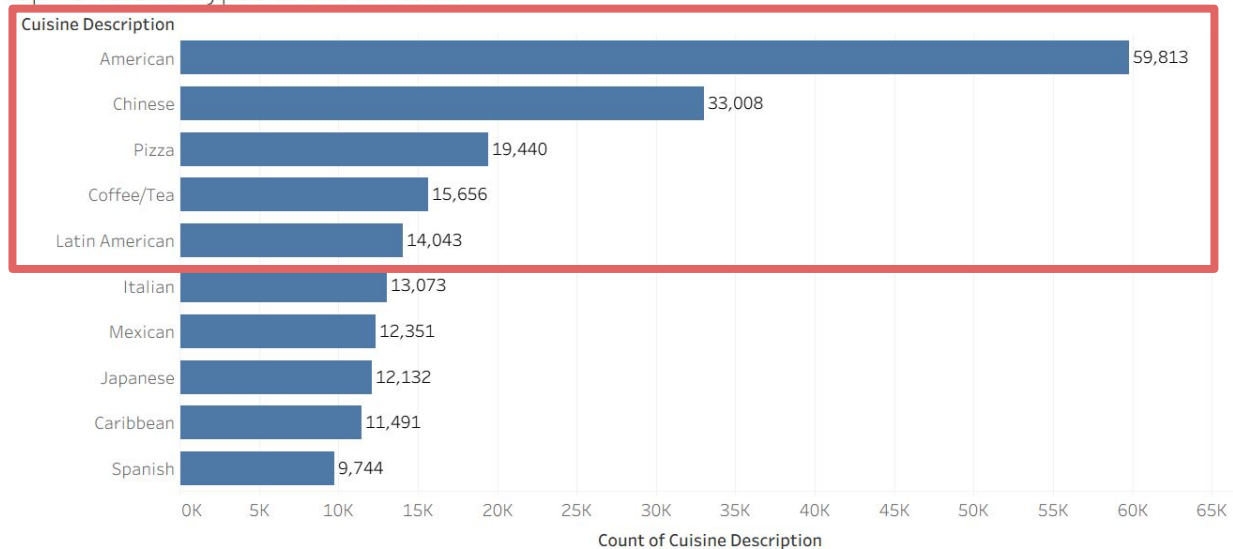
ZIP Code ⌄	Neighborhood ⌄	Borough ⌄
10003	East Village/Gramercy	Manhattan
10019	Midtown West/Hell's Kitchen	Manhattan
10036	Times Square/Theater District	Manhattan
10013	Tribeca/Soho	Manhattan
10002	Lower East Side/Chinatown	Manhattan
10001	Chelsea/Flatiron	Manhattan
10022	Midtown East	Manhattan
10016	Gramercy/Murray Hill	Manhattan
11354	Flushing	Queens
11220	Sunset Park	Brooklyn

Finding #2 - Cuisine Types

American, Chinese, Pizza lead failures

- **Key point:** Top 5 cuisines = 39% of failures

Top 10 cuisine types in the NY area

Cuisine Description	Count
American	59,813
Chinese	33,008
Pizza	19,440
Coffee/Tea	15,656
Latin American	14,043
Italian	13,073
Mexican	12,351
Japanese	12,132
Caribbean	11,491
Spanish	9,744

Count of Cuisine Description

Finding #3 - Severity

Over half of violations in each borough are critical

https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j/about_data
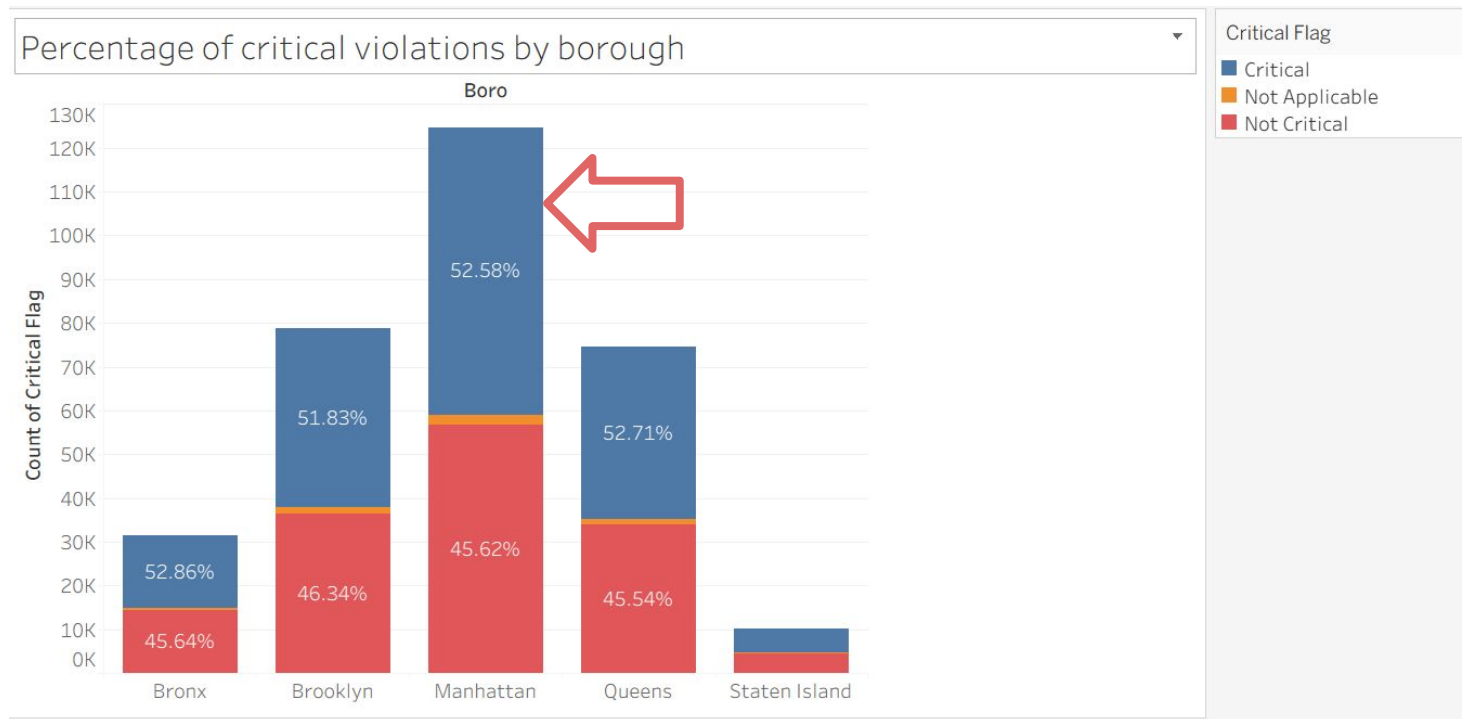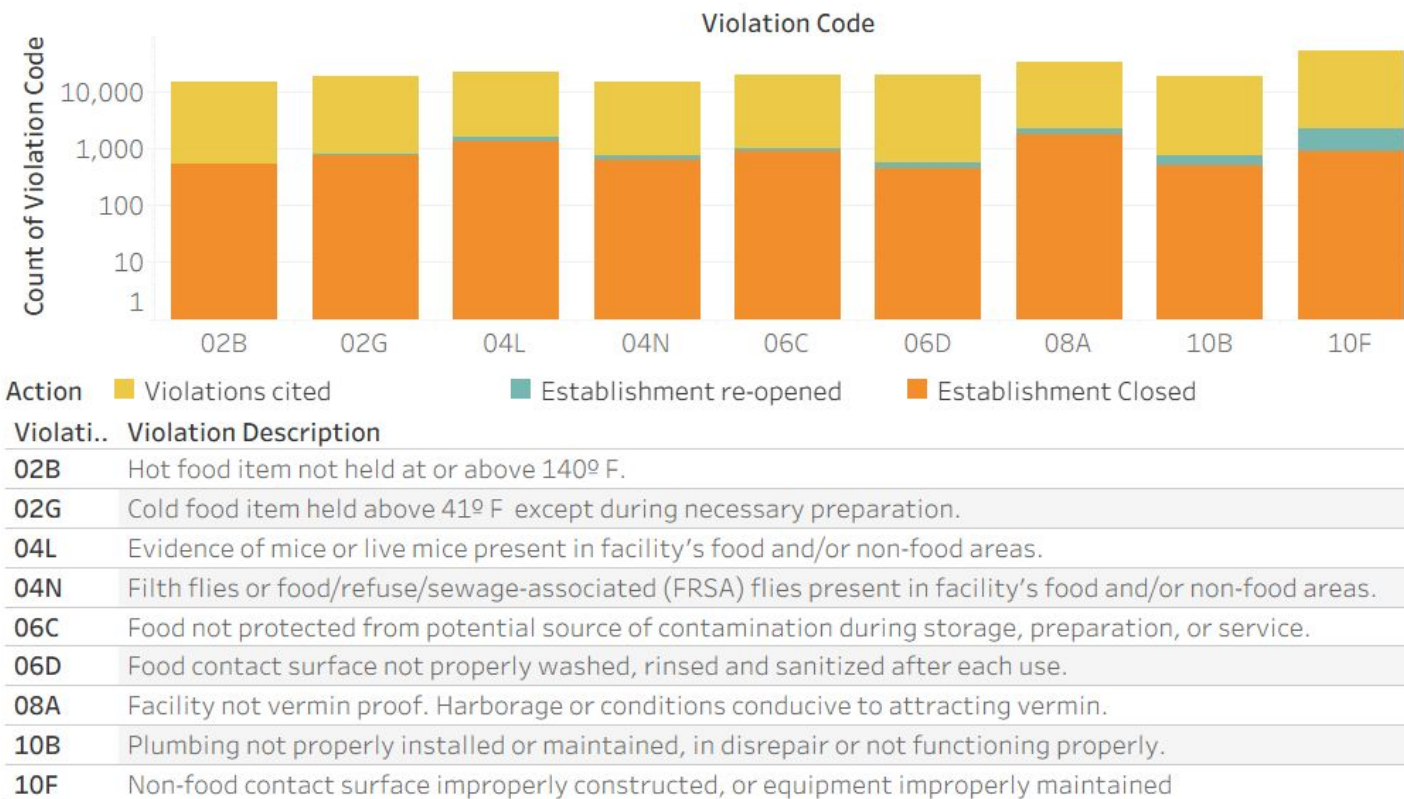
Finding #4 - Violation Types

Infrastructure & pests constitute the majority of restaurant violations.

Improper construction or maintenance 10F(17.5%),

Conditions conducive to attracting vermin 08A (10.6%),

Evidence of mice in facility 04L (7.0%)

Top ten violation codes and actions taken

Violation Code

Action	Violations cited	Establishment re-opened	Establishment Closed

Violati..	Violation Description
02B	Hot food item not held at or above 140º F.
02G	Cold food item held above 41º F except during necessary preparation.
04L	Evidence of mice or live mice present in facility's food and/or non-food areas.
04N	Filth flies or food/refuse/sewage-associated (FRSA) flies present in facility's food and/or non-food areas.
06C	Food not protected from potential source of contamination during storage, preparation, or service.
06D	Food contact surface not properly washed, rinsed and sanitized after each use.
08A	Facility not vermin proof. Harborage or conditions conducive to attracting vermin.
10B	Plumbing not properly installed or maintained, in disrepair or not functioning properly.
10F	Non-food contact surface improperly constructed, or equipment improperly maintained

Finding #4 - Violation Types

Despite Manhattan having the most restaurants and violations, it maintains the lowest closure rate of 3.21% as compared to boroughs of similar population density.
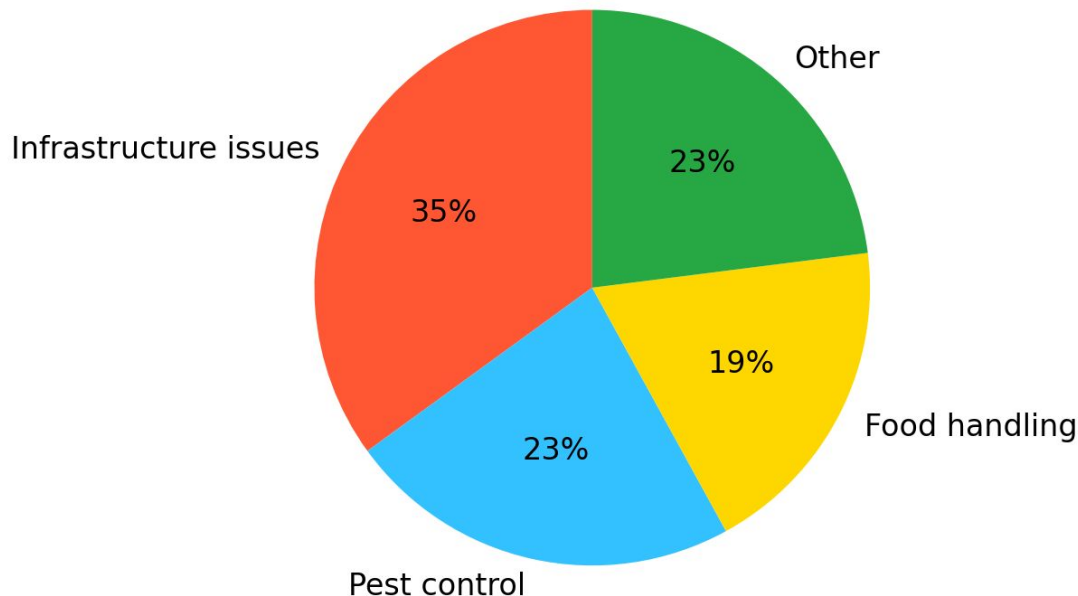

Distribution of actions taken accross boroughs
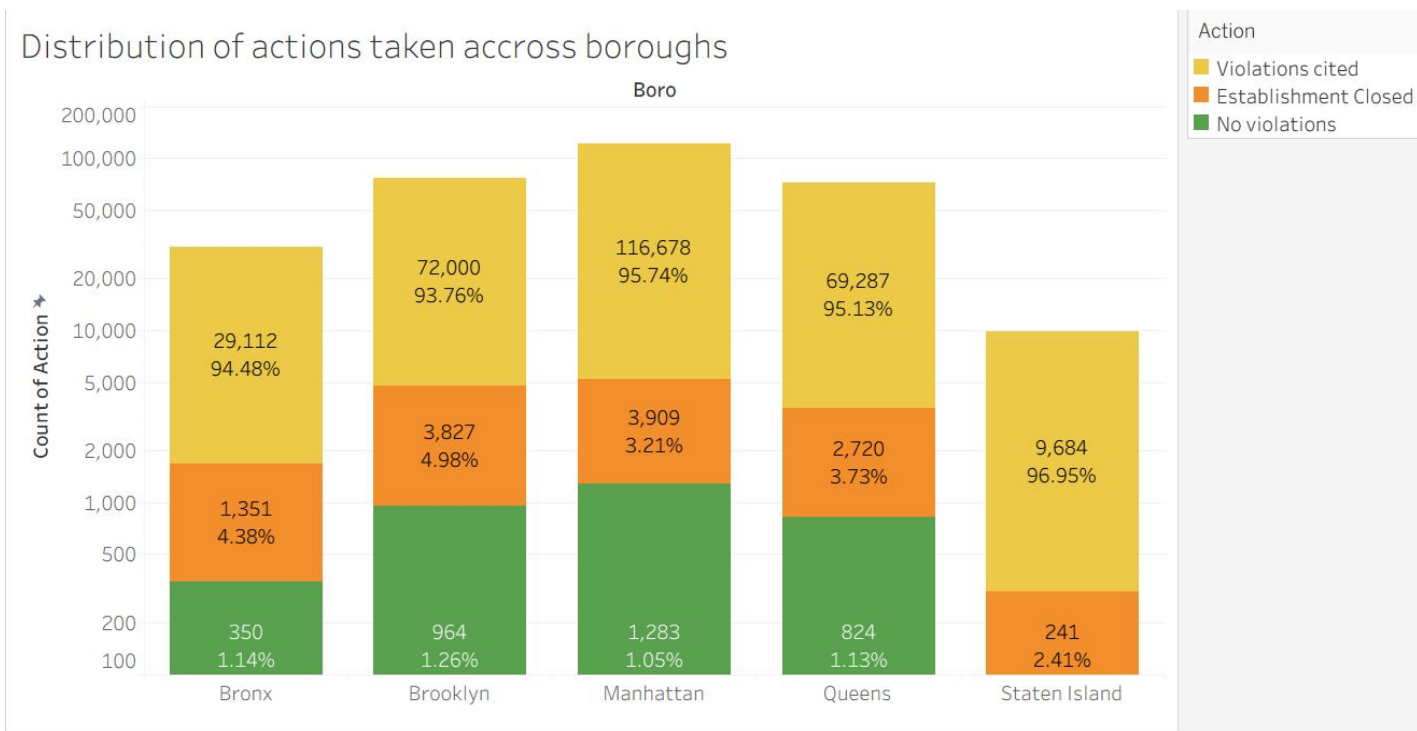
What this means

Problem Categories:

- Infrastructure issues: 35%
- Pest control: 23%
- Food handling: 19%
- Other: 23%

Key insight: Building conditions drive failures, not just food handling

Inspection Issues Breakdown

Infrastructure issues 35%
Other 23%
Food handling 19%
Pest control 23%

Recommendations

Recommendations

#1 Geographic Hotspot Program

What: 50% more inspections in top 10 violating ZIP codes

Impact: Concentrates enforcement resources where violations are most prevalent

#2 Restaurant Cleanliness Training

What: Require staff cleanliness training modules

Impact: Equips restaurant staff with cuisine-specific food safety protocols

#3 Addressing Critical Violations

What: 48-hour re-inspection period for critical violations

Impact: Ensures immediate correction of serious public health threats

#4 Infrastructure Improvement

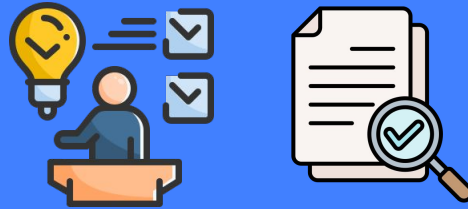**What:** Grants/loans for facility upgrades

Impact: Removes structural barriers preventing compliance in older buildings

#5 Pest Management Program

What: Subsidized monthly pest control

Impact: Prevents recurring infestations through professional intervention

Conclusion

Conclusion

Key Takeaways:
- Addressing infrastructure and location-specific issues, not just penalizing repeat violations, creates sustainable food safety improvement

Impact:
- Potential to prevent thousands of foodborne illness cases annually
- This approach shifts from punishment to support, helping restaurants succeed rather than just citing failures

Disclaimer - Inconsistent Health Score reporting

- The **NYC Department of Health and Mental Hygiene(DOHMH)** assigns letter grades using a clear scoring system.

- However, **scores are sometimes released without an updated grade**, or **old grades are kept despite new scores**.

- This leads to **inconsistent reporting** across restaurants.

- **Businesses are impacted** when third-party platforms like **Yelp, Uber Eats, and DoorDash** rely on outdated or mismatched data to promote restaurants.

Finding #5 - Inconsistent reporting

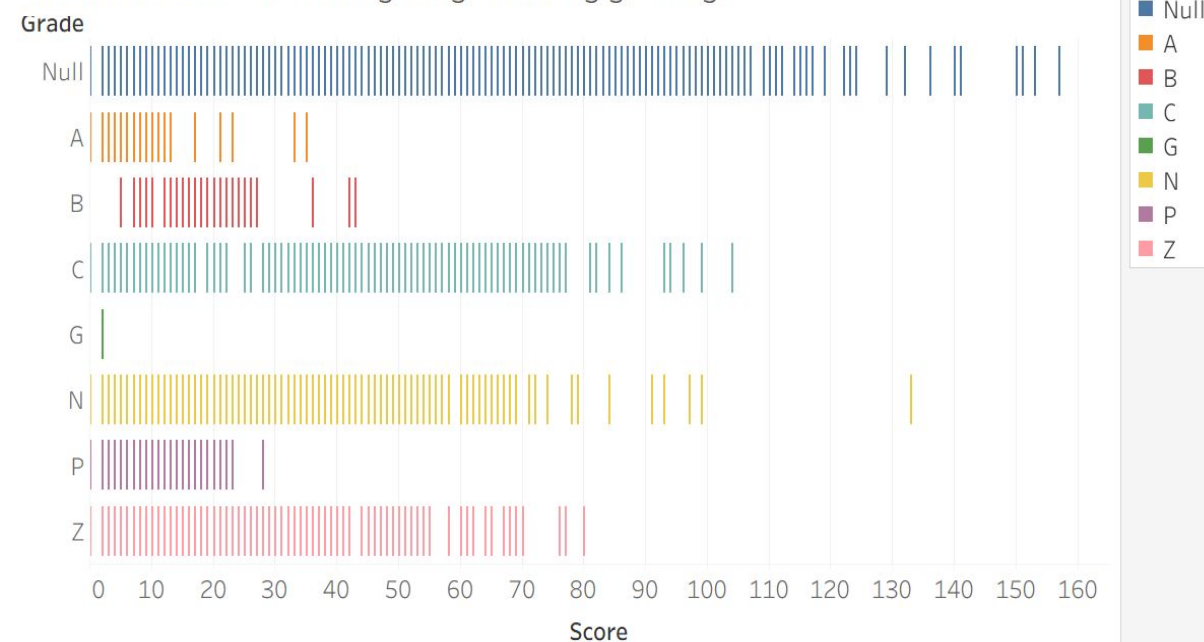The grades should represent scores of

A   0–13

B   14–27

C   28+

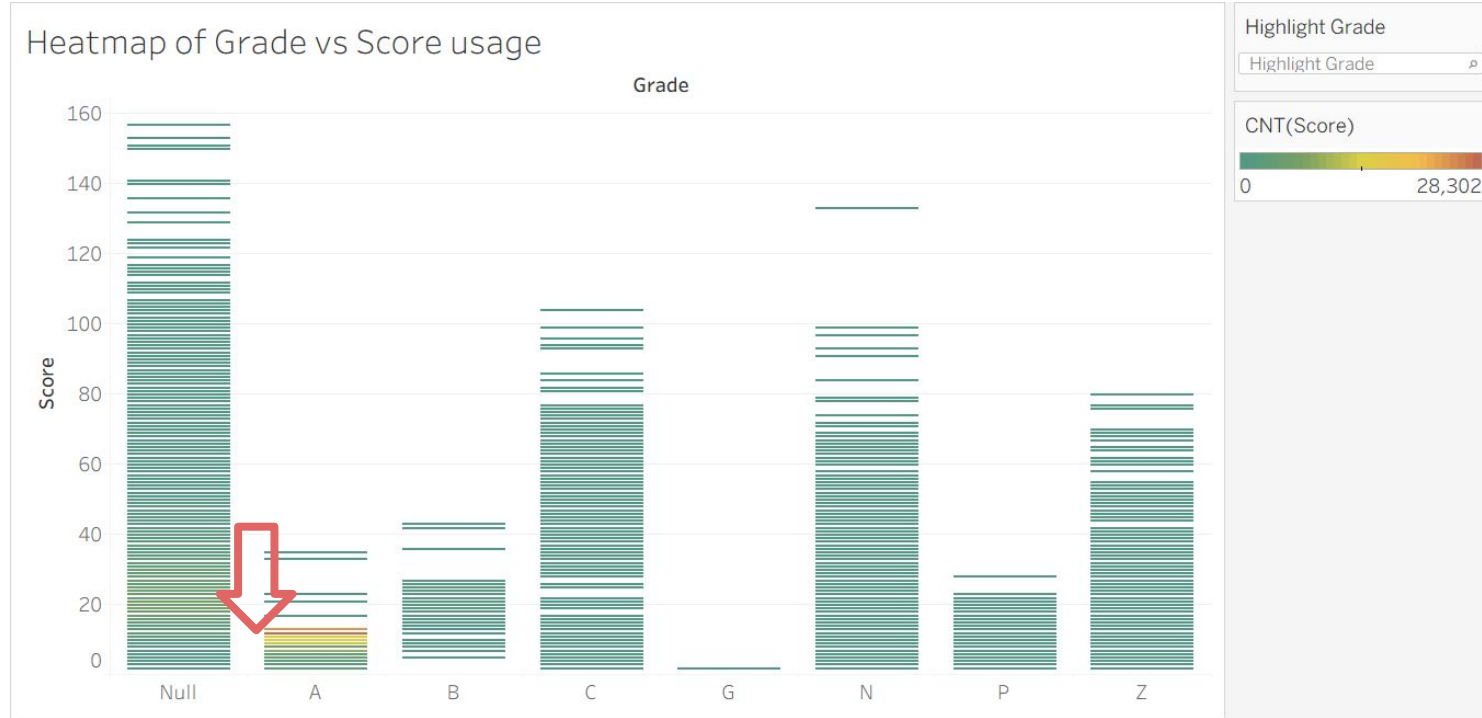And G,N,P,Z representing non finalized grading

However consistently the scores exist outside of their ranges.

Inconsistant use of scoring ranges during grading

Finding #5 - Inconsistent reporting

The vast majority of scores are given in the Grade of A, despite similar scores distributed across all grades

Heatmap of Grade vs Score usage

Resources

- https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j/about_data

Thank you!