SCOPE

Dataset: Restaurant Violation
Document about dataset
Story Flow
Presentation
2nd copy of the spreadsheet

Finalized Visualizations !!

One Pager

Presentation Time: 15-20 mins

Presentation

- Include background on our dataset, how we cleaned the dataset what we did and how the data changed after we cleaned it

Problems we are trying to resolve / address

- Where the DOH should spend most of their time based on the highest violations in certain areas?
- Connie's suggestion (from 9/25/25)
    - Violation codes and incorporate zip codes, think abt geography (boro)
    - See if certain zipcodes are more populated like Time Square/Chinatown (commercial use) - maybe more dense areas are correlated to pest problems
        - Maybe compare these commercial areas to residential areas
    - Time series: has any violation codes changed over the years (ex. Increase in codes related to pest may have increased/decreased and ask why and what happened during those time frames to add context)
    - Inspections paused from March 2020 to July 2021

Project Name Ideas

-

Potential Stakeholders:

- NYC Department of Health
- ~~NYC Health and Hygiene~~
- ~~City Council / Community Boards (To encourage policy change in neighborhoods)~~
- ~~Public Health / Hospitals (illness outbreak correlations + seasonal foodborne illnesses)~~

VIOLATION CATEGORY	VIOLATION CODE	COUNTA of CRITICAL FLAG
+ Worker Hygiene Total		2190
+ Sanitization Total		7627
+ Pest Problem Total		11656
+ Food Protection Total		6183
+ Food Handling/Temperature Total		9818
+ Facility/Plumbing Total		1348
+ Certification/Training Total		2169
Grand Total		**40991**

Violation C..	Violation Category	Violation Description	
10F	Facility/Equipment	Non-food contact surface improperly constructed. Unacceptable material used. Non-food contact surface or equipment improperly maintained and/or not properly sealed, raised, spaced or movable to allow accessibility for cleaning on all sides, above and underneath the unit.	10,858
08A	Pest Problem	Facility not vermin proof. Harborage or conditions conducive to attracting vermin to the premises and/or allowing vermin to exist.	8,707
04L	Pest Problem	Evidence of mice or live mice present in facility's food and/or non-food areas.	5,691
06D	Sanitization	Food contact surface not properly washed, rinsed and sanitized after each use and following any activity when contamination may have occurred.	5,374
06C	Food Protection	Food not protected from potential source of contamination during storage, preparation, transportation, display or service.	4,461
02G	Food Handling/Temperature	Cold food item held above 41Â° F (smoked fish and reduced oxygen packaged foods above 38 Â°F) except during necessary preparation.	4,366
10B	Facility/Plumbing	Plumbing not properly installed or maintained; anti-siphonage or backflow prevention device not provided where required; equipment or floor not properly drained; sewage disposal system in disrepair or not functioning properly.	4,022
02B	Food Handling/Temperature	Hot food item not held at or above 140Â° F.	4,004
04N	Pest Problem	Filth flies or food/refuse/sewage-associated (FRSA) flies present in facilityâ€™s food and/or non-food areas. Filth flies include house flies, little house flies, blow flies, bottle flies and flesh flies. Food/refuse/sewage-associated flies include fruit flies, drain flies and Phorid flies.	3,364
04A	Certification/Training	Food Protection Certificate not held by supervisor of food operations.	2,169

Main Idea

Why are these zones critical?

-Most pests and restaurants are not maintaining food temperature then expand into the borough briefly and why are these zip codes getting people sick?

-Start up heat codes/dip deeper into critical/pest violations are the pertinent and dive deeper in why it's happening

-Kevin (Closure date versus inspection date)

Andy Choi

Kevin's seasonality can work with Andy's line graph.

- Maybe compare two categories and whether it's coordinated-Pest and sanitization or temp or food spoilage. Look deeper into violation connected with food spoilage

Kevin Li
Heatmap of Critical Counts in Zip Codes- "Hot Zones"
Heatmap of Violations that are Pest management

- Kevin suggested making around 4 heat maps
- Andy suggests narrow into borough and then zip codes

Is there a correlation with seasonality?

- Can incorporate a historical line graph of number violations. Might have to exclude 3/2020-7/2021 or make a note that no inspections were done there

hot zones may correlate with seasonality: line graph --> compare the number of restaurants
cuisines with food types, correlation, raw fish, street food
-can talk about the number of restaurants and density
scatter plot of average inspection score and
Grade distribution across boroughs?
-Which borough is most likely to get an "A rating"? Histogram

Jenny Wu
What is the DOH doing in regards to these violation codes?
-Brooklyn has the highest number of closures despite not having the most violations. Which violation codes (both non-critical or critical) are popular in these boroughs?
-Combine it with Kevin's heat map
-Deeper dive into action. Are actions mostly directly targeted at critical violations or non critical violations? What category is the most popularly targeted?
Top Violation Category

- Kevin's heat map. Count of pest problem in each zip code

Chris Bobadilla

- Everyone made a similar timeline focused on a rat problem
- % Restaurants with Grade
    - Difficult because most restaurants without grade (Kevin)
    - Find the best chart for this visualization (Ashley)
    - Can supplement the grade/cuisine into the story

Ashley Kwong
Top Cuisines with restaurant violations
-Won't be included because it's not relevant to the route we're going

Tableau Links:

- Chris: Restaurant Violation Charts

- Andy: Restaurant Violation Chart 1

Restaurant Violation Chart 2
Restaurant Chart
Critical Violation

- Jenny: Restaurant Violation Charts
  Restaurant Violation Pt 2

- Kevin: Restaurant Violation Charts
  Critical Violations Heatmap

- Ashley: Restaurant Violation Charts

TIMELINE

1st Meeting Notes - 9/2/25

Next Steps

- Look through the data and brainstorm what metrics you can get from the data set
  - Types of cuisine, different boroughs, coordination between different violations and other data etc
- Clean the data before putting it into Tableau

Ideas

- Create a heat map

2nd Meeting Notes - 9/17/25

Project Steps

- Look through the data
- Clean the data
- Determine what charts we want to create
- Build a report

Project Workflow

- 9/23/25
  - Progress Report
    - Progress on Data Cleaning
- 9/25/25
  - Meeting 6:30pm - 9:30pm
    - Data Cleaned
    - Begin Tableau Vizzes
    - ~~Begin PowerPoint Presentation~~
- 9/28/25
  - Meeting 4pm - 7pm

        - Reviewed everyone's vizzes and chose ones that best fit our story
        - Created project story flow
- 9/29/25
    - Progress Report
        - Worked on finalizing vizzes, story flow, and color scheme
    - Next steps
        - Reclean dataset (violation codes, violation category)
- 9/30/25
    - Meet during COOP time
    - Revisited dataset for dirty data
- 10/1/25
    - 8:30pm - 11:20pm lol
    - Finalized vizzes !!
- 10/2/25
    - Meeting agenda
        - Finalize vizzes with recleaned dataset
        - Assign slides for presentation
        - Try to get presentation done asap
- ~~10/2/25~~
    - ~~Checkpoint~~
        - ~~Finalize Presentation~~
        - ~~Rehearsals~~
- 10/5/25
    - Meeting 6pm - 8pm
        - Presentation should be complete
        - Dry Run presentation
        - Work on 1 pager
- 10/6/25
    - Touch base to ensure everything is ready for presentation tomorrow
- 10/7/25
    - Presentation Day!

NOTES

Project Plan

- Stakeholders:
    - (idea) Governance body that wants to start a restaurant stipend program and wants to know how or where to allocate funds
- Goals:

Joey recommends

- Broadening our stakeholder
    - Restaurants union
    - NYC health and hygiene
- Research why restaurants are having so many violations
- Research how restaurants are graded

Notes based on convo with Joey:

1. Restaurants by Borough

- NYC's infrastructure varies by borough, and older or more crowded areas (like Manhattan) may face higher risks of pest access (rats, cockroaches).
- Question to explore: Which boroughs consistently see higher rates of pest-related violations?
- Stakeholder value: Business owners and policymakers could use this to prioritize preventive measures (e.g., extra pest control budget in boroughs with higher risk).

2. Time-Series Analysis

- Inspections happen repeatedly across years. You could look at trends by month, quarter, or year.
- Example: Do rodent violations spike in the summer months?
- Stakeholder value: Restaurants could plan seasonal interventions (e.g., increase cleaning and pest monitoring in warmer months when infestations peak).

3. Geographic (Spatial) Analysis

- Use latitude/longitude or ZIP codes to map where violations cluster.
- Example: Are certain neighborhoods or community boards "hotspots" for critical violations?
- Stakeholder value: City agencies or business groups could target inspections or allocate grants to high-risk areas.

4. Business Impact / Actionable Insight

- Insights should go beyond just describing violations — they should suggest what businesses or stakeholders can do.
- Example scenario: A restaurant stipend program
    - If analysis shows that restaurants in Manhattan are disproportionately affected by pest violations, stakeholders could allocate stipends specifically for pest-control services.
    - Similarly, if kitchens in outer boroughs struggle more with refrigeration violations, stipends could cover equipment upgrades.

Data Cleaning Documentation

Jenny Wu

- Fixed boroughs labeled as 0 and the spelling for some places (starbucks, dunkin, popeyes, subway, chipotle)
    - 1 row (borough=0)
- Continued Filling Missing Zip Codes
    - 700 Missing Blank Zip Codes

Kevin Li

- Remove White Spaces
- Formatted Dates
- Removed Duplicate Rows
    - 5139 Duplicated Rows Found and Removed
- "No violation recorded at this time", add violation code "00X"

Ashley Kwong

- Narrowed years from 2020 - 2023
- Reduced Number of rows from 319,00 → 81,508
    - (Removed Irrelevant Rows and Columns)
- Removed Phone Number, NTA, BIN, BBL

Andy Choi

- Fixed the Spelling for McDonalds
- Filled in missing building numbers
    - 63 missing values
- No Violations were recorded with empty violation code and empty violation category and description
    - Violation Code: 00X
    - Violation Category: No Violations
    - With empty description

1. Heat Map
2. Explain noncritical violations versus critical
3.