Analyzing COVID-19's Impact On Facility Violations

Group members: Wasay Ahmed, Sahil Chauhan, Arthur Floriano, Kiyoon Jang, Miguel Rodriguez

Problem

Hypothesis:

The COVID-19 pandemic has reverberated widely, leading to shifts in facility violations.

Our study aims to investigate the pandemic's impact on grocery stores and restaurants,

expecting increased violations in the former and decreased violations in the latter during

the pandemic when compared to before it. Furthermore, we anticipate that

violation types will vary based on the facility type and their association with nearby areas.

Established pandemic dates as the following:

Pre-Pandemic Dates: 10/01/16 - 12/15/19
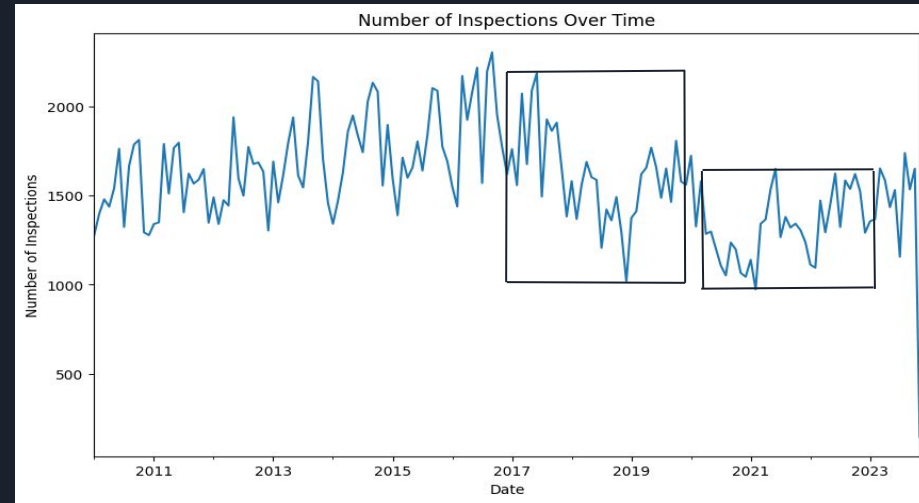
Pandemic Dates: 03/01/20 - 05/11/23

Data: Why restaurants and grocery stores?
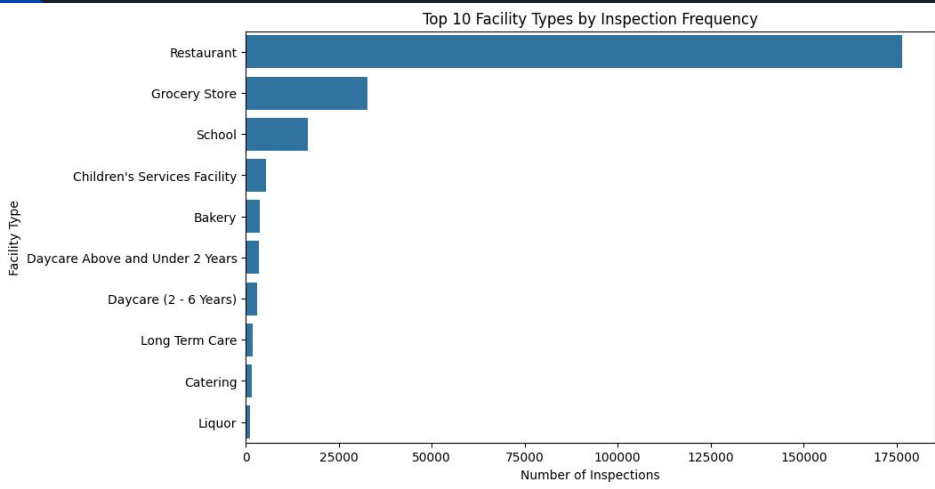
Top 10 Facility Types by Inspection Frequency

Pre-Pandemic Dates: 10/01/16 - 12/15/19
Pandemic Dates: 03/01/20 - 05/11/23

Number of Inspections Over Time

Why Care:
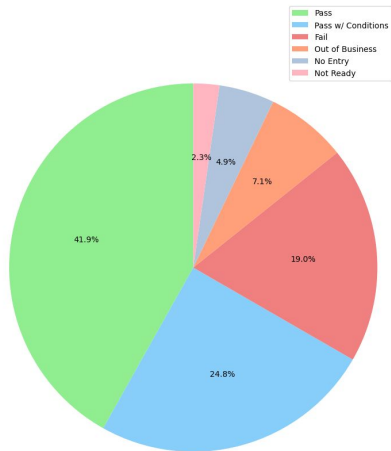- Significant dip in inspections between 2020 and 2022

Choosing the Data:
- Restaurants and grocery stores are most frequently inspected

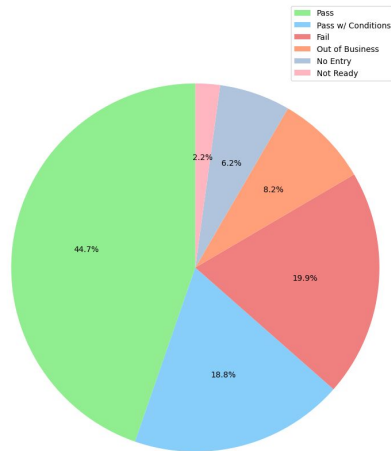Graphs are findings from exploratory data analysis

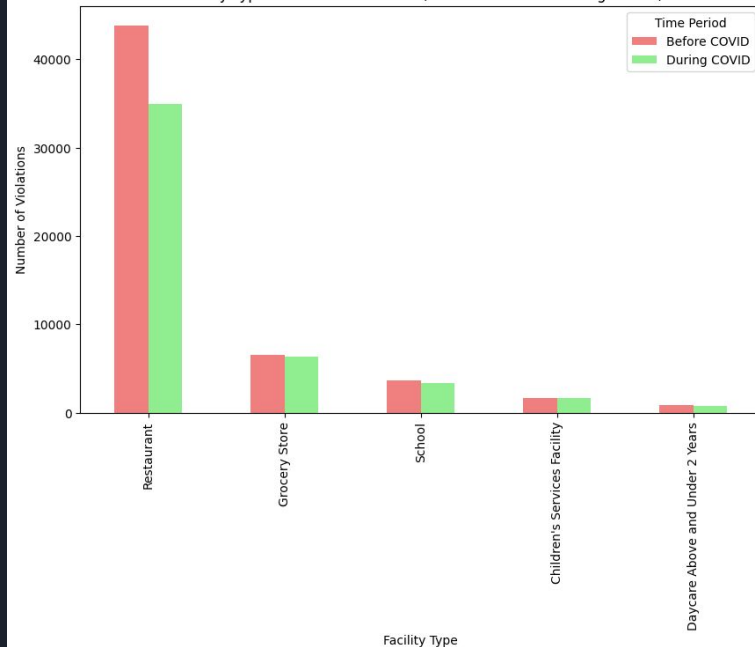# Solution/Findings: Pre-Pandemic vs. Pandemic Violations

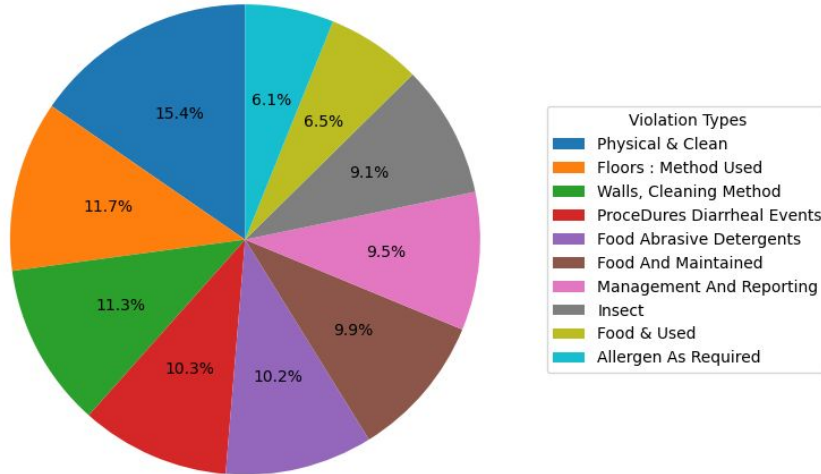*Total violations before pandemic vs pandemic*

Top 5 Facility Types with most violations before pandemic vs pandemic

Pre-Pandemic Dates: 10/01/16 - 12/15/19
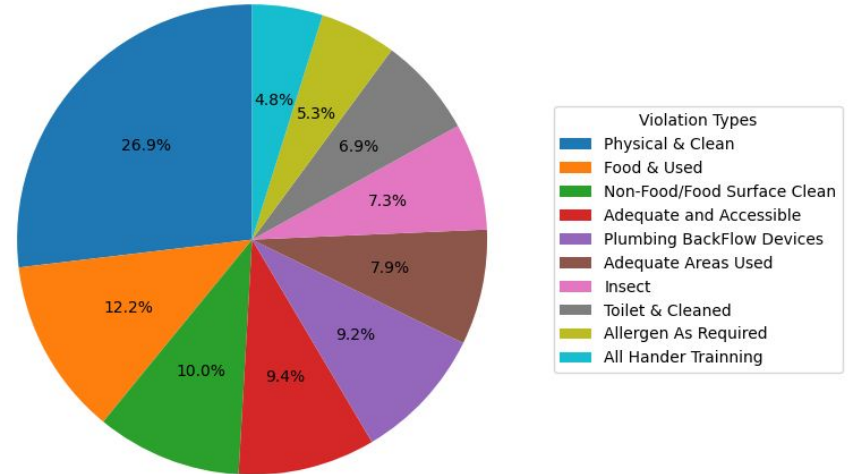*Pandemic Dates: 03/01/20 - 05/11/23*

Solution/Findings: Common Violation Types
Pre-Pandemic vs. Pandemic

Top 10 Pre-Pandemic Violations

Violation Types
- Physical & Clean
- Floors : Method Used
- Walls, Cleaning Method
- ProceDures Diarrheal Events
- Food Abrasive Detergents
- Food And Maintained
- Management And Reporting
- Insect
- Food & Used
- Allergen As Required

15.4%, 6.1%, 6.5%, 9.1%, 9.5%, 9.9%, 10.2%, 10.3%, 11.3%, 11.7%

Top 10 During-Pandemic Violations

Violation Types
- Physical & Clean
- Food & Used
- Non-Food/Food Surface Clean
- Adequate and Accessible
- Plumbing BackFlow Devices
- Adequate Areas Used
- Insect
- Toilet & Cleaned
- Allergen As Required
- All Hander Trainning

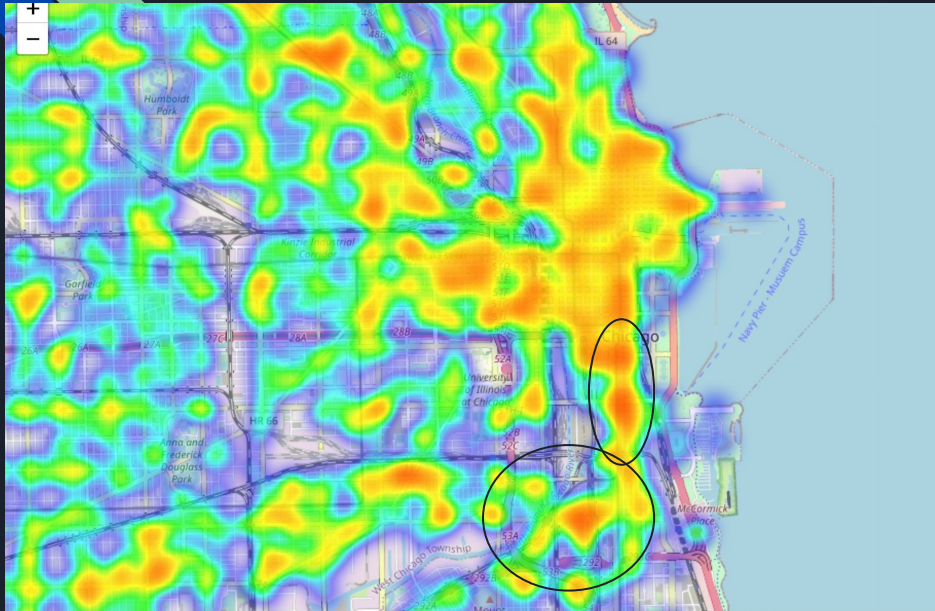26.9%, 4.8%, 5.3%, 6.9%, 7.3%, 7.9%, 9.2%, 9.4%, 10.0%, 12.2%

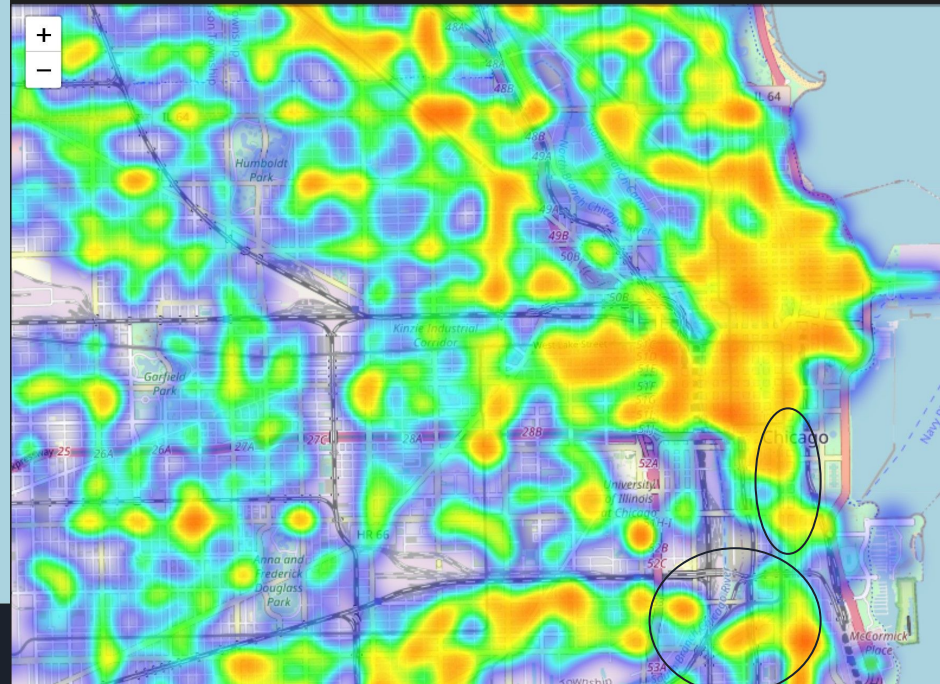*Top 10 most common violation types before pandemic vs pandemic*

Pre-Pandemic Dates: 10/01/16 - 12/15/19
Pandemic Dates: 03/01/20 - 05/11/23

Solution/Findings: Pre-Pandemic vs. Pandemic Violation Location Analysis

Pre-pandemic: Higher concentrations of violations

Pandemic: Lower Concentrations of violations

Pre-Pandemic Dates: 10/01/16 - 12/15/19
Pandemic Dates: 03/01/20 - 05/11/23

Solution/Findings: Machine Learning Analysis

Histogram Classifier

Accuracy: 0.54

Classification Report:

	precision	recall	f1-score	support
Business Not Located	0.04	0.14	0.06	22
Fail	0.53	0.01	0.01	10086
No Entry	0.97	0.23	0.37	2102
Not Ready	0.49	0.05	0.08	640
Out of Business	0.78	0.22	0.35	4547
Pass	0.53	0.99	0.69	27119
Pass w/ Conditions	0.18	0.00	0.00	8047
accuracy			0.54	52563
macro avg	0.50	0.23	0.22	52563
weighted avg	0.52	0.54	0.41	52563

Linear Regression Model

Accuracy: 0.53

Classification Report:

	precision	recall	f1-score	support
Business Not Located	1.00	0.00	0.00	22
Fail	1.00	0.00	0.00	10086
No Entry	1.00	0.00	0.00	2102
Not Ready	1.00	0.00	0.00	640
Out of Business	0.62	0.19	0.29	4547
Pass	0.52	0.99	0.68	27119
Pass w/ Conditions	1.00	0.00	0.00	8047
accuracy			0.53	52563
macro avg	0.88	0.17	0.14	52563
weighted avg	0.72	0.53	0.38	52563

Did Not Work:
- No baseline established in findings
- Accuracies too low
- Must reduce data file for higher accurate results -> Speed up training data

Different Attempts:
- Different models took too long: must decrease file size

Future Attempts -> Improvements:
- Attempt new models with decreased file size: SVM, etc.

Results

Hypothesis:

The COVID-19 pandemic has reverberated widely, leading to shifts in facility violations.

Our study aims to investigate the pandemic's impact on grocery stores and restaurants,

expecting increased violations in the former and decreased violations in the latter during

the pandemic when compared to before it. Furthermore, we anticipate that

violation types will vary based on the facility type and their association with nearby areas.

WHY SOLUTION WORKS:

- Data analysis supporting why restaurants/grocery stores need specific focus
- Inspections over time denote dip in inspections during pandemic
- Various visuals highlight decrease in violations during pandemic
- Heat map showcases violation concentration based on surrounding areas

WHAT DID NOT WORK + IMPROVEMENTS

- ML analysis accuracies are too low
  - Adopt new analysis to reflect higher accuracies
  - Minimize the data set and adjust for overfitting
  - Implement MLC baseline
- Alternative models
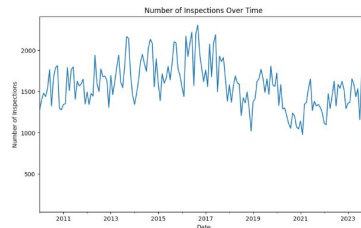 - SVM, Neural Networks, Decision Tree

Next Steps