CHICAGO FOOD INSPECTIONS 2010-2025

By Michael ashley

CLEANING

- Renamed all column names to lowercase with underscores for consistency.

- Converted "inspection_date" from object to datetime format.

- Converted "license_#" and "zip" to strings

- Filled missing "aka_name" values using "dba_name"

- Imputed "risk" values based on dominant risk per "facility_type"; ramining nulls were filled as "Unknown Risk"

- Cleaned and standardized "city" and filled missing values using "zip" when available

- Filled missing "state" values based on known city (Chicago)

- Dropped one row with missing "inspection_type", "results", and "violations" (incomplete record)

- New column added "has_violations", a Boolean column indicating whether a violation was recorded during the inspection

QUESTIONS

1. Which areas in Chicago may need health and safety reform?

    - Can we visualize heat zones on the map of Chicago?

2. Are there any patterns in inspection results or violations by facility types?

    - Do certain facility types consistently perform better or worse?

3. How did facilities adapt to the 2018 change in inspection definitions?

    - Did establishments maintain consistent inspection results before or after the rule change?

4. Are complaint-based inspections more likely to result in failing inspection?

    - Does inspection types correlate with failure rate?

5. Are specific business chains and franchises more likely to fail inspection?

    - using "dba_name" values, can we identify patterns amongst large brands?

IMPORTANT NOTES
(FOR QUESTION 1)

- Date ranges from 2021-2025

- Minimum inspections per zip was 20 to reduce noise.

- Inspection results limited to "Pass", "Fail", and "Pass with Conditions"

- Zip codes 60827, 60617, and 60628 had the highest inspection fail rates in the city

- Only zip code 60827 had a higher fail rate than pass rate

QUESTION 1

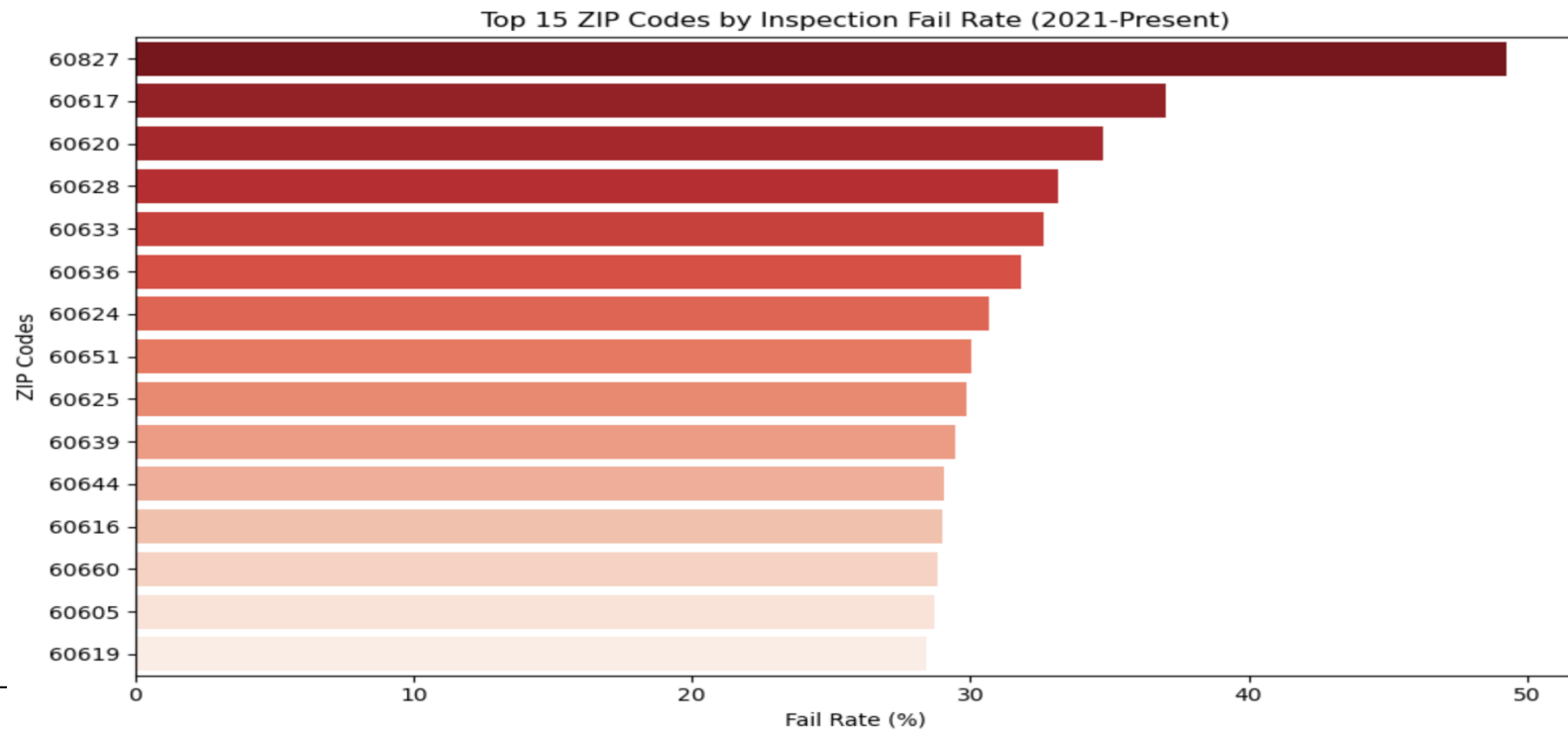WHICH AREAS IN CHICAGO MAY NEED HEALTH AND SAFETY REFORM? (2021-2025)

TOP 5 ZIP CODES WITH THE HIGHEST INSPECTION FAIL RATE

results	Fail	Pass	Pass w/ Conditions	total	fail_rate	pass_rate
zip						
60827	34	28	7	69	49.275362	40.579710
60617	532	729	175	1436	37.047354	50.766017
60620	449	635	208	1292	34.752322	49.148607
60628	366	587	151	1104	33.152174	53.170290
60633	47	78	19	144	32.638889	54.166667

FAILURE RATE BY ZIP CODE BAR CHART

Top 15 ZIP Codes by Inspection Fail Rate (2021-Present)

TOP 5 ZIP CODES WITH THE HIGHEST INSPECTION PASS RATE

results	Fail	Pass	Pass w/ Conditions	total	fail_rate	pass_rate
zip						
60638	161	712	107	980	16.428571	72.653061
60601	86	668	176	930	9.247312	71.827957
60637	112	493	85	690	16.231884	71.449275
60615	117	525	101	743	15.746972	70.659489
60602	56	276	59	391	14.322251	70.588235

PASSING RATES BY ZIP CODE BAR CHART

Top 15 ZIP Codes by Inspection Pass Rate (2021–Present)
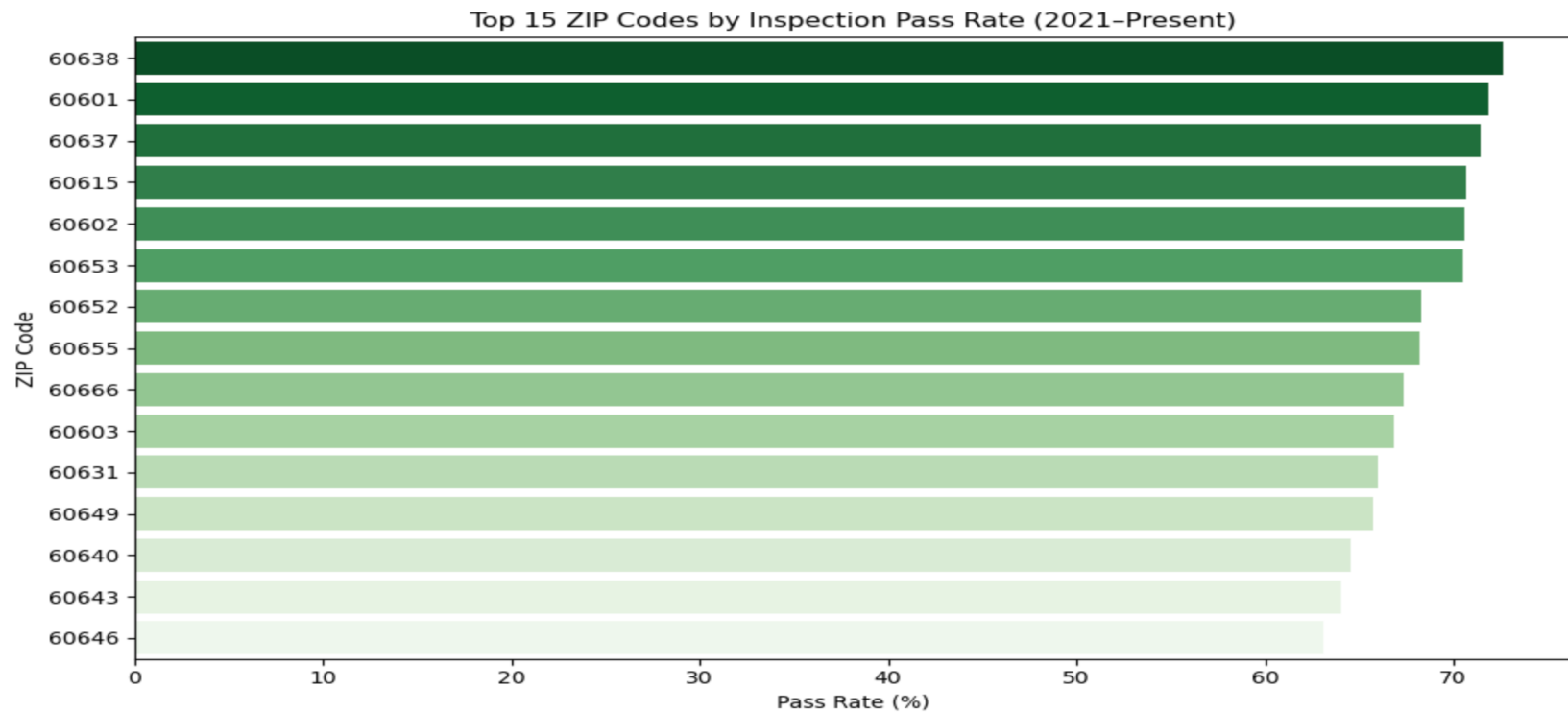
CHOROPLETH MAP

- Let's head on over to my jupyter notebook so we can see the heat map I have plotted over the map of Chicago!

QUESTION 2

ARE THERE ANY PATTERNS IN INSPECTION RESULTS OR VIOLATIONS BY FACILITY TYPES?

- "Unknown" facility types were excluded.

- A Large portion of these businesses were reported out of business and not ready

- Keeping them would distort fail rates and misinterpret active businesses.

- Results were filtered to reflect from 2021-Present

results	
Not Ready	34.0
Out of Business	31.2
Pass	14.3
Fail	9.0
No Entry	7.6
Pass w/ Conditions	3.1
Business Not Located	0.8

dba_name	
WHALE FISH & CHICKEN	9
SOULE' CHICAGO CORP	8
ORANGE LINE	7
LUCKY STRIKE DOWNTOWN CHICAGO	7
TACOS VELAZQUEZ	6
THE SLICE SHOP	5
JOE & THE JUICE NEW YORK LLC	4
THE SECOND CITY	4
DABBLE	4
FOOD ROCKET	4
HERBAN PRODUCE	4
AVENUE GROCERY MART	3

TOP 20 FACILITY TYPES BY INSPECTION FAIL RATES 2021-PRESENT

Top 20 Facility Types by Inspection Fail Rates (%) - Post 2020
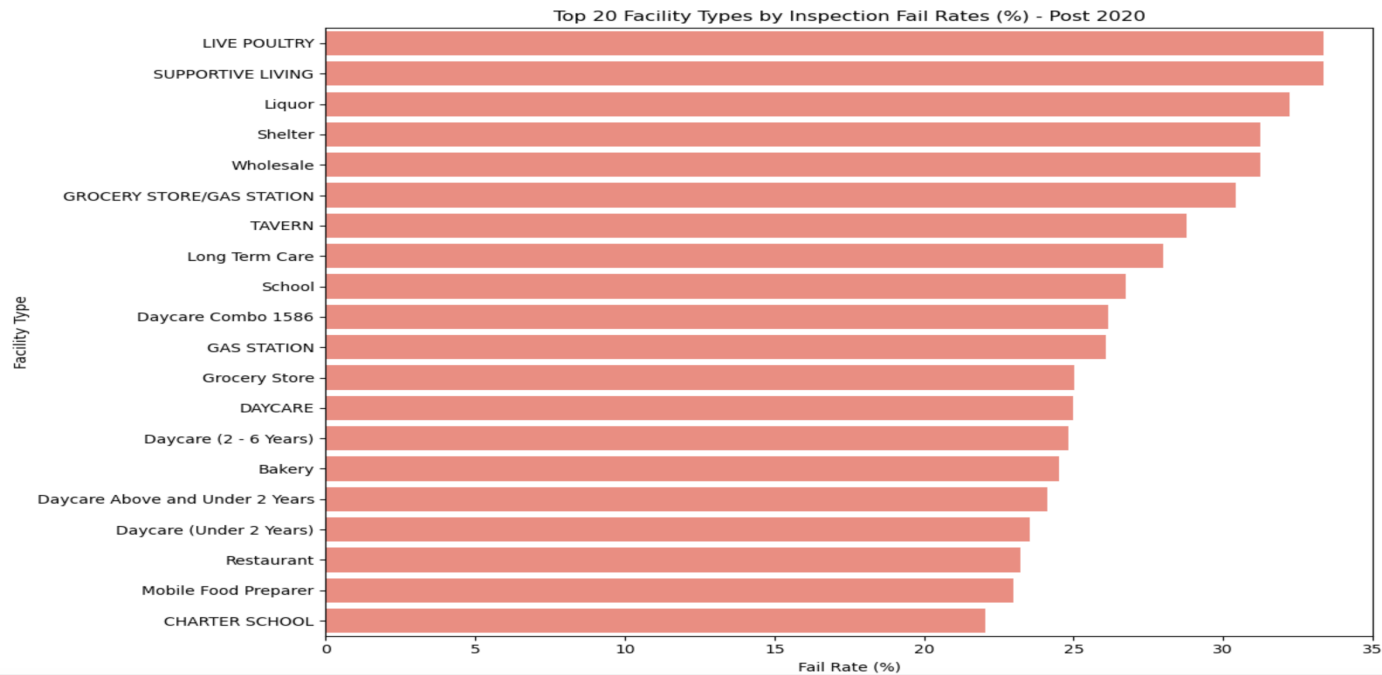
#3 HOW DID FACILITIES ADAPT TO THE 2018 CHANGE IN INSPECTION DEFINITIONS?

	index	facility_type	pre_fail_rate	post_fail_rate	fail_rate_change
0	32	Unknown	60.069444	31.451613	-28.617832
1	31	Mobile Prepared Food Vendor	31.481481	14.516129	-16.965352
2	30	Mobile Food Dispenser	32.313830	18.018018	-14.295812
3	29	ROOFTOP	13.888889	0.000000	-13.888889
4	28	Mobile Frozen Desserts Vendor	13.333333	0.000000	-13.333333
5	27	Mobile Food Preparer	33.702882	24.007937	-9.694946
6	26	Wholesale	37.130802	28.571429	-8.559373
7	25	Liquor	40.540541	33.248082	-7.292459
8	24	TAVERN	40.000000	33.082707	-6.917293
9	23	CHARTER SCHOOL	32.352941	26.315789	-6.037152
10	22	Daycare Combo 1586	27.171492	23.148148	-4.023344
11	21	BANQUET	25.714286	21.875000	-3.839286
12	20	Special Event	19.318182	15.833333	-3.484848
13	19	Daycare (Under 2 Years)	25.698324	22.727273	-2.971051
14	18	PRIVATE SCHOOL	14.285714	11.363636	-2.922078

	facility_type	pre_fail_rate	post_fail_rate	fail_rate_change
0	Shelter	11.627907	26.530612	14.902705
1	DAYCARE	7.500000	21.739130	14.239130
2	LIVE POULTRY	27.450980	36.170213	8.719232
3	School	20.041793	25.734285	5.692492
4	Long Term Care	22.232916	27.507163	5.274247
5	Daycare Above and Under 2 Years	18.809777	22.845336	4.035560
6	Bakery	23.777479	26.346434	2.568955
7	Shared Kitchen User (Long Term)	2.688172	4.784689	2.096517
8	Hospital	15.186916	16.908213	1.721297
9	Restaurant	21.077100	22.668837	1.591737
10	Shared Kitchen	19.178082	20.000000	0.821918
11	Children's Services Facility	18.408551	18.152524	-0.256027
12	Grocery Store	26.679510	25.722708	-0.956802
13	Golden Diner	19.144144	17.611940	-1.532204
14	Daycare (2 – 6 Years)	23.026647	21.474774	-1.551873

FAILURE RATE CHANGE PRE AND POST 2018

Change in Inspection Failure Rates pre and post 2018
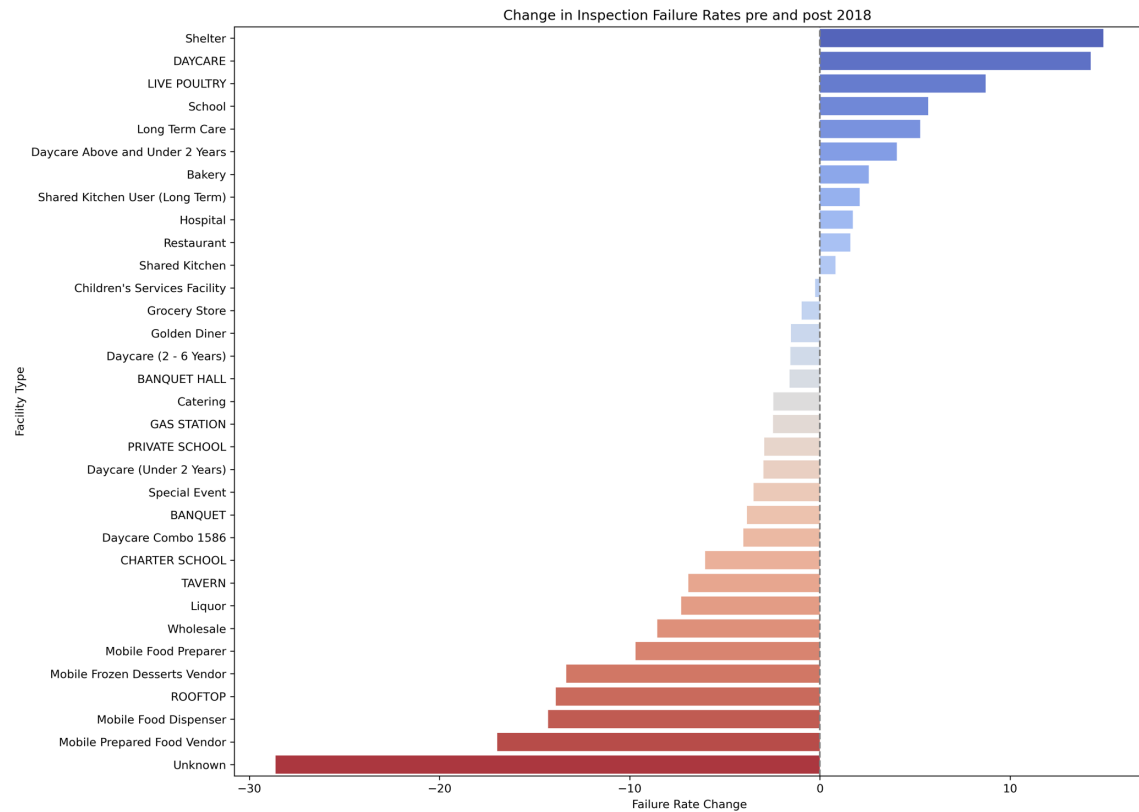
```
dba_name
WHALE FISH & CHICKEN
SOULE' CHICAGO CORP
LUCKY STRIKE DOWNTOWN CHICAGO
ORANGE LINE
THE SECOND CITY
MAN-O-SALWA
TACOS VELAZQUEZ
BOTY'S ICE CREAM ROLL INC..
TOKYO LUNCH BOXES & CATERING
BELEN
THE SLICE SHOP
JOE & THE JUICE NEW YORK LLC
FOOD ROCKET
HERBAN PRODUCE
DABBLE
Name: count, dtype: int64
```

THE UNKNOWNS

- "Unknowns" appear to be restaurants

- Without confirmed metadata I can not say for certain

- Did not reclassify to avoid a potential bias

QUESTION 4
ARE COMPLAINT-BASED INSPECTIONS MORE LIKELY TO RESULT IN FAILING INSPECTION?

results	Fail	Pass	Pass w/ Conditions	total	fail_rate
is_complaint					
False	44732	124205	34831	203768	21.952416
True	11540	25450	8850	45840	25.174520

- Ran a Chi-Square Test to assess independence between complaints and inspection outcomes

- Test adjusts for unequal group sizes

- Ensures difference in fail rates is statistically significant and not random

- Chosen because both variables are categorical
    - Variable 1: Whether the inspection was complaint-based or not
    - Variable 2: Inspection Result

FAILURE RATES BY COMPLAINTS AND NON COMPLAINTS CHART

Failure Rates by Complaints and Non-Complaints
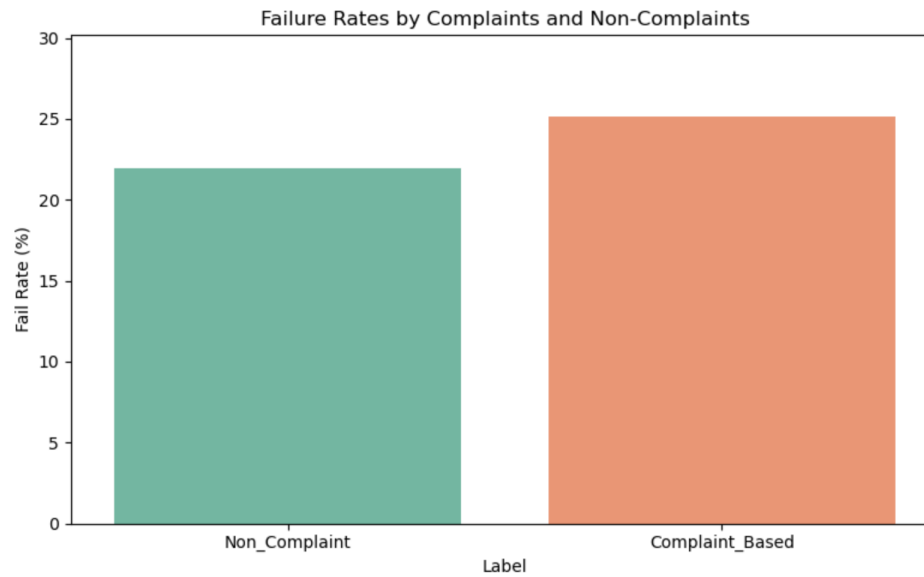
CHI-SQUARE TEST

- **Null Hypothesis**: There is no association between complaint-based inspections and inspection outcomes. The likelihood of failing an inspection is the same whether the inspection was triggered by a complaint.

- **Alternative Hypothesis**: There is an association between complaint-based inspections and outcomes. The likelihood of failing an inspection is increased when driven by a complaint.

- **Goal**: To determine if inspection outcome (fail vs. non-fail) is statistically dependent on whether a complaint triggered the inspection

- **Chi-Square Statistic**: 222.31(Measures how much the observed values deviate from expected values. A high number indicates a strong difference between complaint and non-complaint inspection outcomes)

- **P-Value**: < 0.0001 — statistically significant

- **Interpretation**: Strong evidence to reject the null hypothesis

- **Conclusion**: Complaint-based inspections are more likely to result in failure — the difference is not random

QUESTION 5

ARE SPECIFIC BUSINESS CHAINS AND FRANCHISES MORE LIKELY TO FAIL INSPECTION?

- I explored whether some businesses consistently fail but found that most entries appear to reflect individual locations.

- Without a reliable chain franchise, I can not draw confident conclusions about about business level trends

- Popeye's appears 2 times within the top 50 but appears to represent 2 different locations and not the franchise.

Top Chains by Inspection Fail Rate
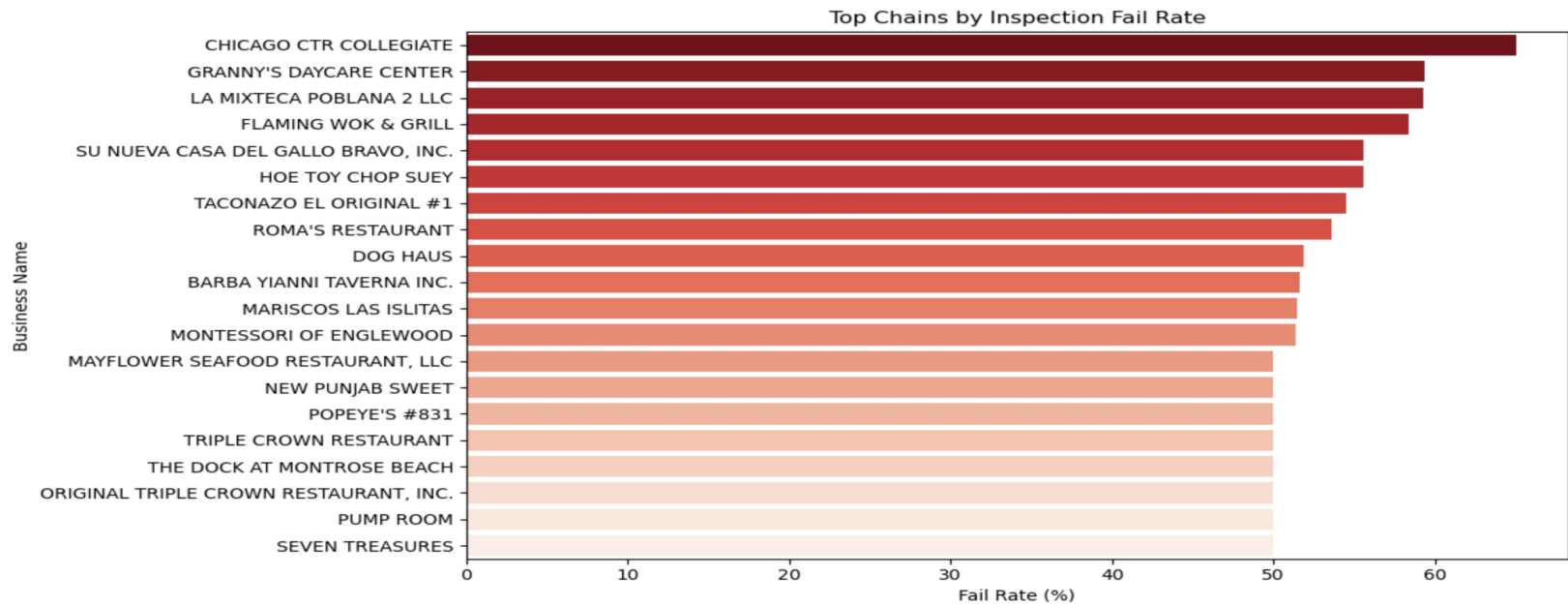
TOP 20 CHAINS BY INSPECTION FAIL RATE
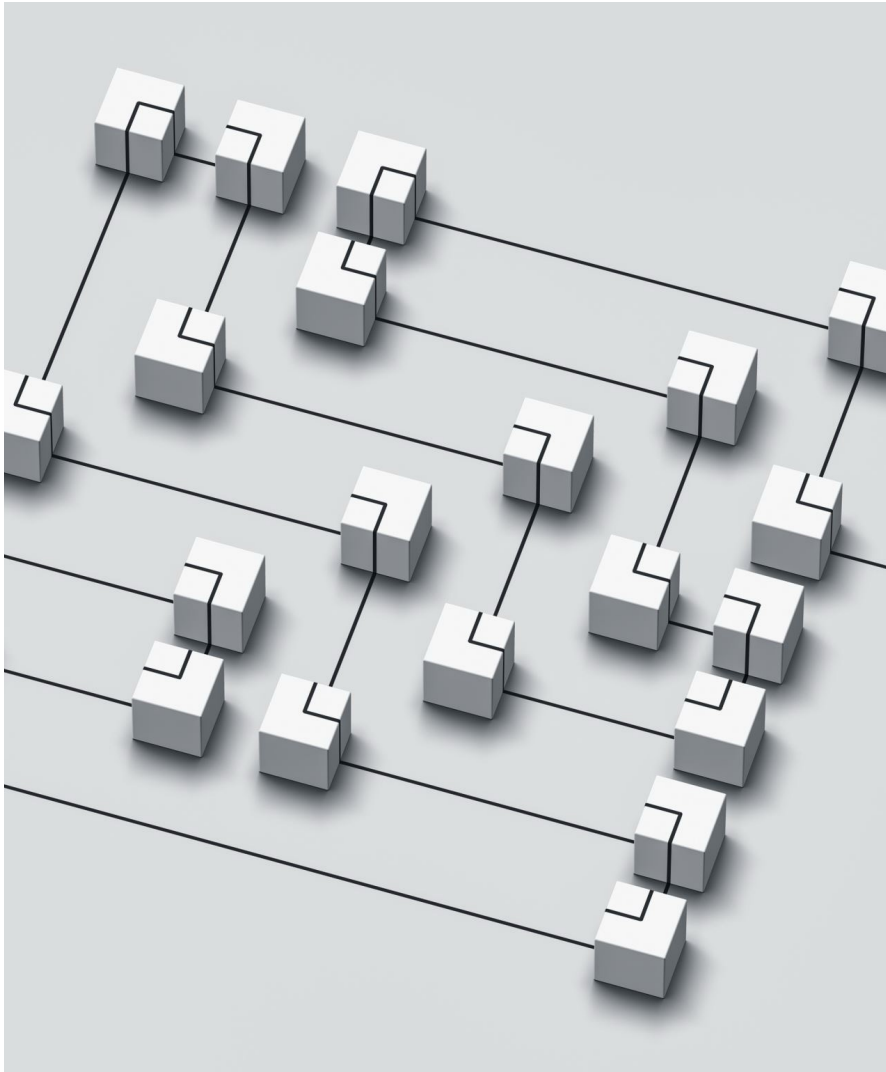
LIMITATIONS

- Facility Type accuracy: Some establishments had "Unknown" facility types, limiting my ability to categorize them. While many appear to be restaurants, assumptions were avoided preserve integrity.

- Chain Identification: The "dba_name" column represents individual locations. Without a standardized chain ID, franchise level patterns could not be confidently assessed.

- Geographic Boundaries: Choropleth map was based on zip codes, not neighborhoods.

- Complaint Bias: Complaint driven inspections are more likely to fail, but causation can not be confirmed.

- Small Sample Sizes: Facility types or ZIP codes with few inspections (under 20) were excluded to avoid unreliable fail rate calculations.

- Data Entry Errors: Possibility of typos, inconsistencies, or outdated records such as "out-of-business" listings still present.

FINAL THOUGHTS

- Facilities serving vulnerable groups of people such has schools, shelters, day cares, and medical care showed some of the highest increases in inspection fail rates after 2018, highlighting a need for stronger oversight.

- Complaint based inspections are statistically more likely to result in inspection failure as confirmed by the Chi-Squared Test that accounts for the difference in group sizes, which means the comparison is adjusted for how many inspections occurred in each group and is unlikely due to random chance.

- Most inspections conducted do not stem from complaints, and most facilities pass but outliers exist and often clutter in certain zip codes and facility types.

- Zip codes 60827, 60617, and 60628 displayed the highest inspection fail rates. However, only zip code 60827 had a higher failure rating than passing rate.