Identifying and Analysing Patterns in NYC Restaurant Inspection Results

Tom Wang • Capstone 1 • Galvanize DSI

About Me

Data Scientist, studied Economics, worked in E-commerce strategy

Family has been in the restaurant industry for 30 years

NYC Restaurant Industry

In 2019 B.C. (Before COVID), the restaurant industry in New York City consisted of 23,650 establishments and 317,800 jobs. **Both these numbers were all-time highs!** Further, from 2009-2019, the growth of the restaurant industry was **double** the overall growth rate of businesses in the city. [1]

The Department of Health and Mental Hygiene (DOHMH) conducts inspections of every one of these establishments on a regular cycle.

The Dataset

NYC OpenData - The DOHMH has a dataset consisting of every violation cited by a health inspector for a restaurant over the last 8 years.

- 410k observations
- Only restaurants in an active status
- This project will only look at 2017, 2018, and 2019 data

Income Data - PyPi project “uszipcode”

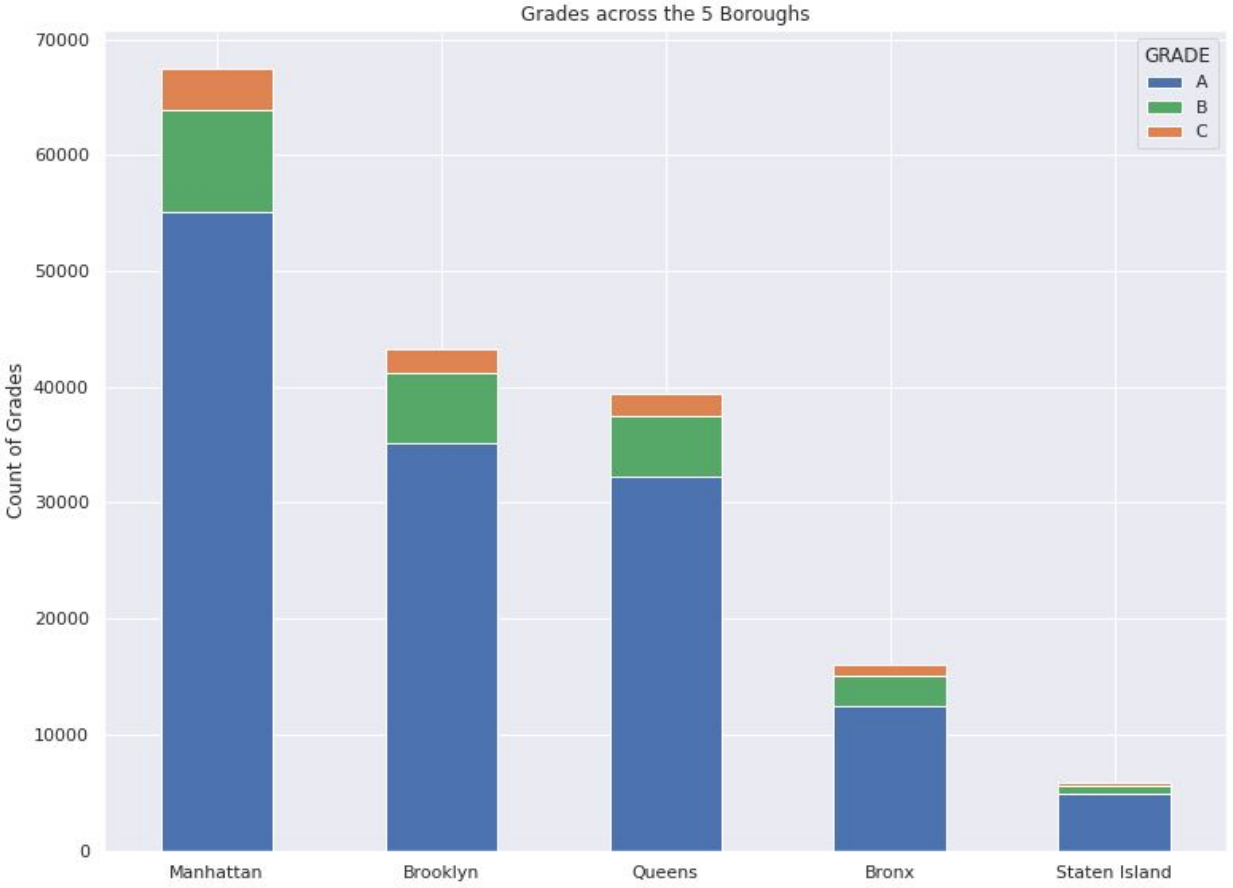
How Inspections Work

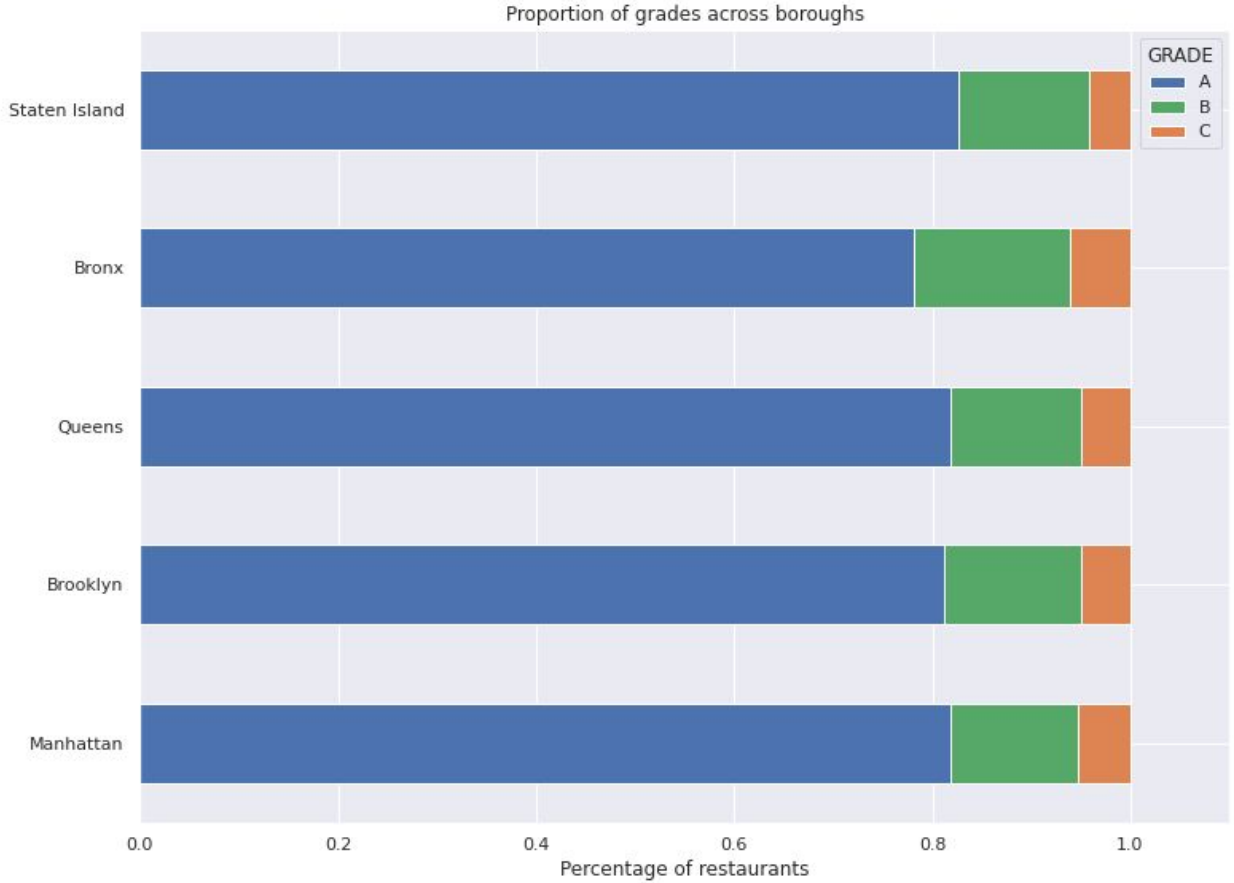
- "A" grade: 0 to 13 points for sanitary violations
- "B" grade: 14 to 27 points for sanitary violations
- "C" grade: 28 or more points for sanitary violations

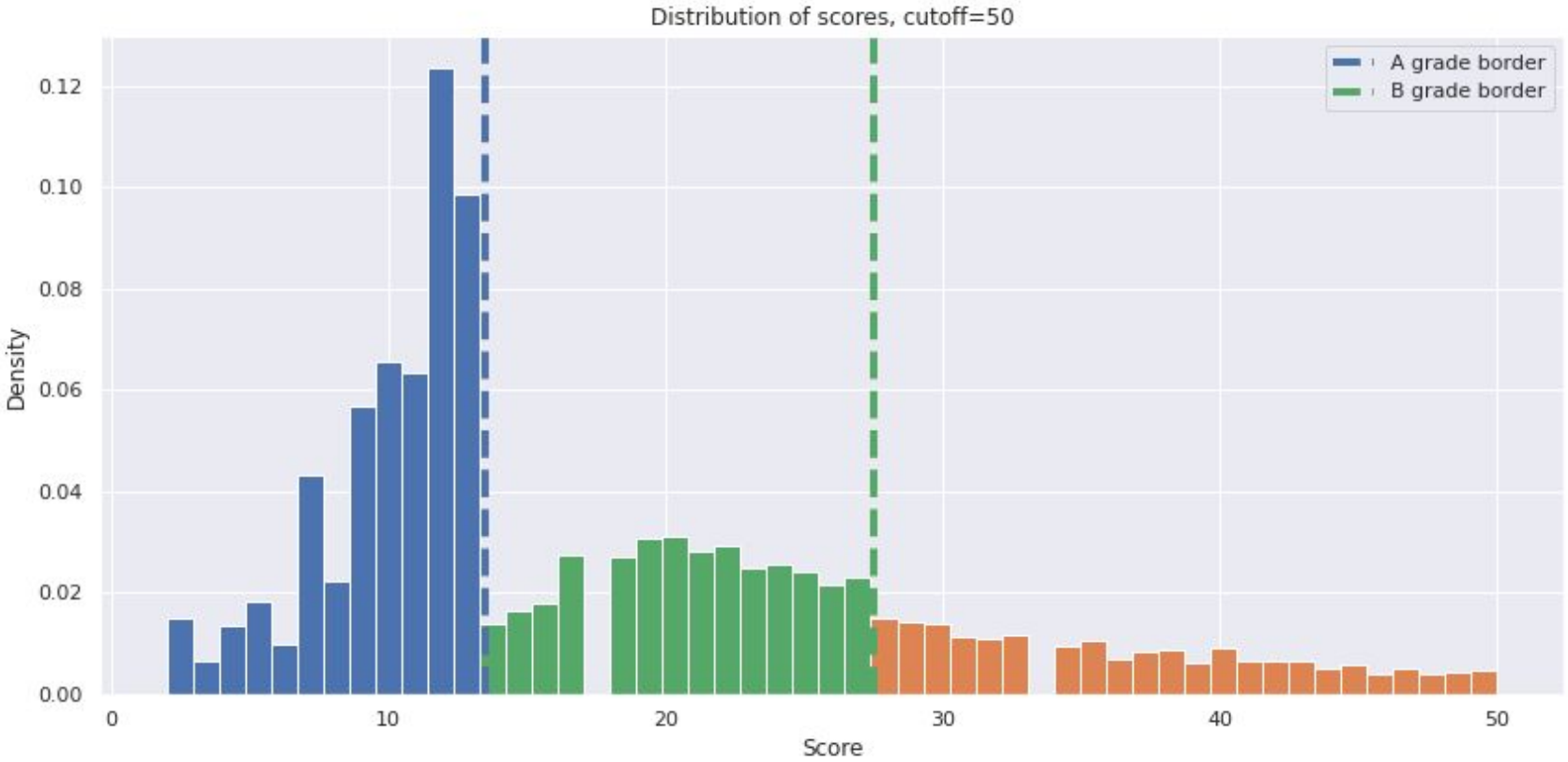
Violations fall into three categories:

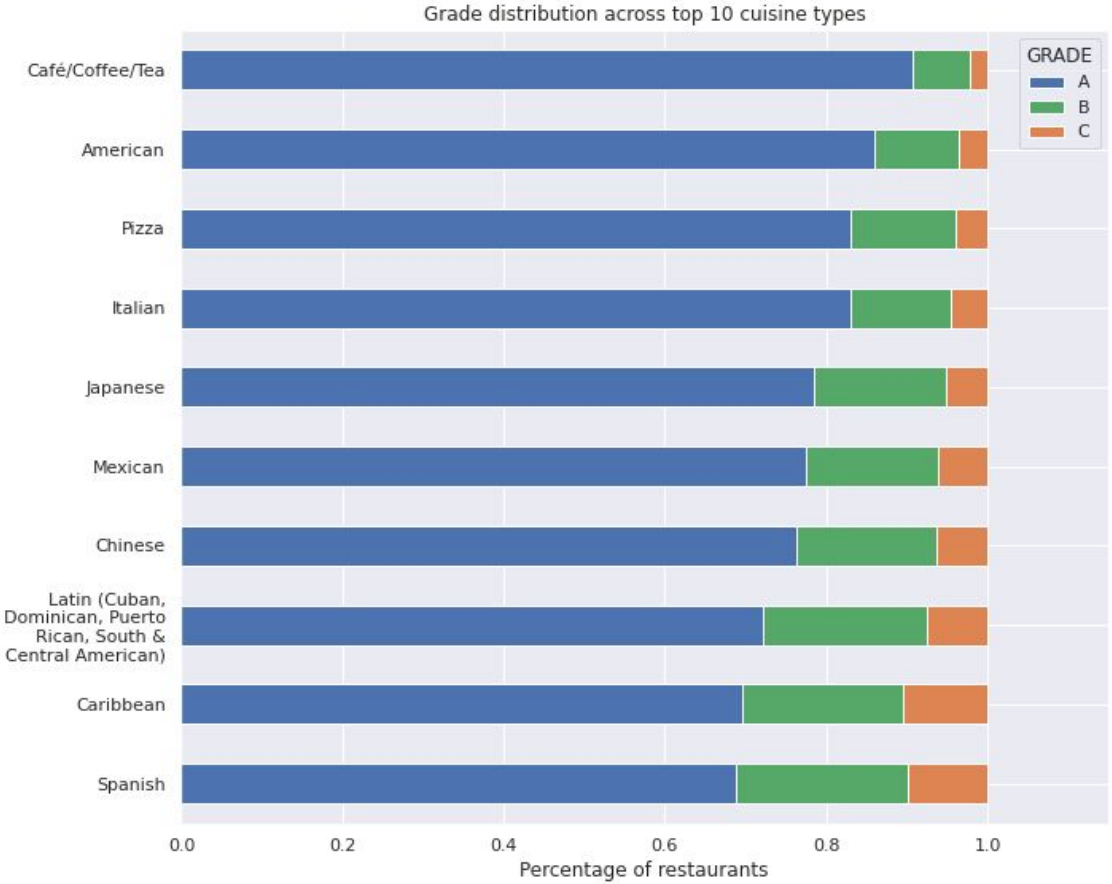
- Public health hazard: min. 7 points
- Critical violation: min. 5 points
- General violation: min. 2 points

Additionally, A-graded restaurants typically will not be inspected for another 12 months, while B's and C's can receive re-inspections up to 3 months later.

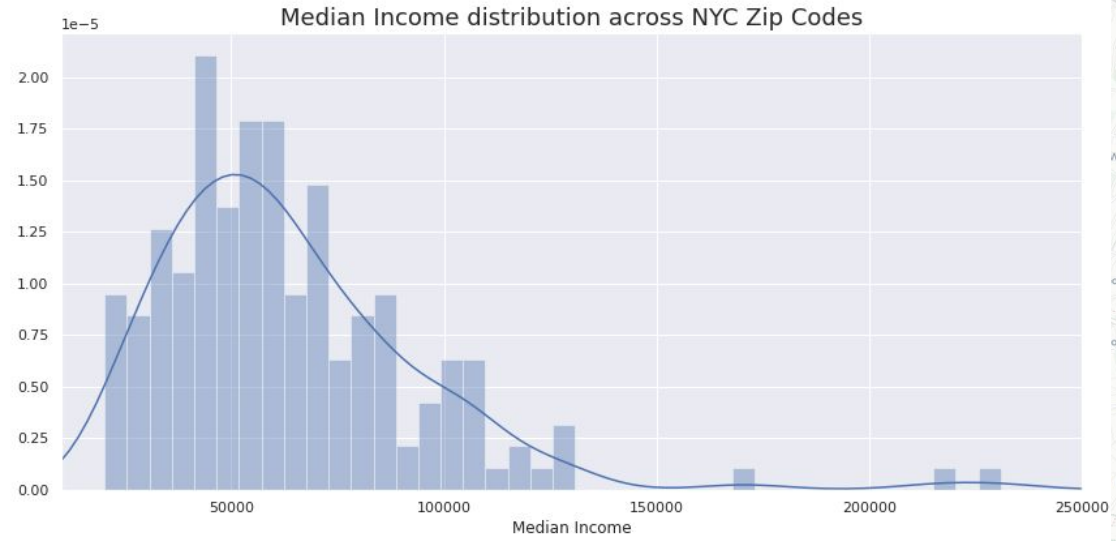

Grades across the 5 Boroughs
Count of Grades
70000
60000
50000
40000
30000
20000
10000
0
GRADE
A
B
C
Manhattan
Brooklyn
Queens
Bronx
Staten Island

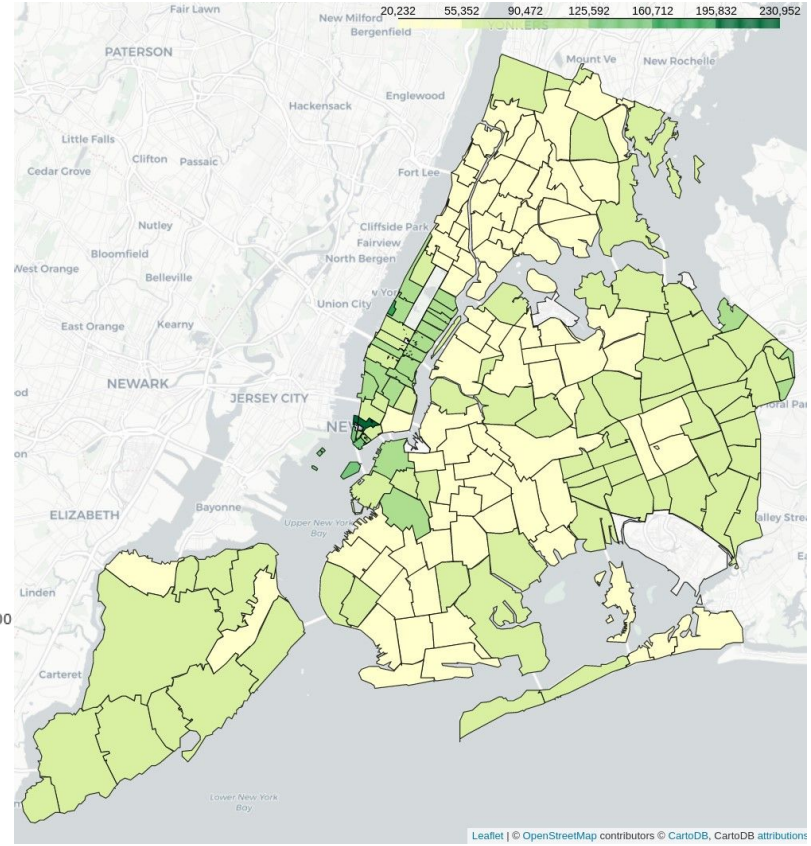

Proportion of grades across boroughs
GRADE
A
B
C
Staten Island
Bronx
Queens
Brooklyn
Manhattan
0.0
0.2
0.4
0.6
0.8
1.0
Percentage of restaurants


Distribution of scores, cutoff=50
A grade border
B grade border
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0
10
20
30
40
50
Score


Grade distribution across top 10 cuisine types
GRADE
A
B
C
Café/Coffee/Tea
American
Pizza
Italian
Japanese
Mexican
Chinese
Latin (Cuban, Dominican, Puerto Rican, South & Central American)
Caribbean
Spanish
0.0
0.2
0.4
0.6
0.8
1.0
Percentage of restaurants

Incorporating Median Income Data





Trim data down to unique restaurants, ~24K observations

Mann Whitney U-Test

H_0 : The given pairs of populations of incomes are equal.

H_A : The given pairs of populations of incomes are not equal.

$\alpha = .05$

Pairs to be tested: A-B, A-C, B-C

A-B p-value = 0.000

A-C p-value = 0.006

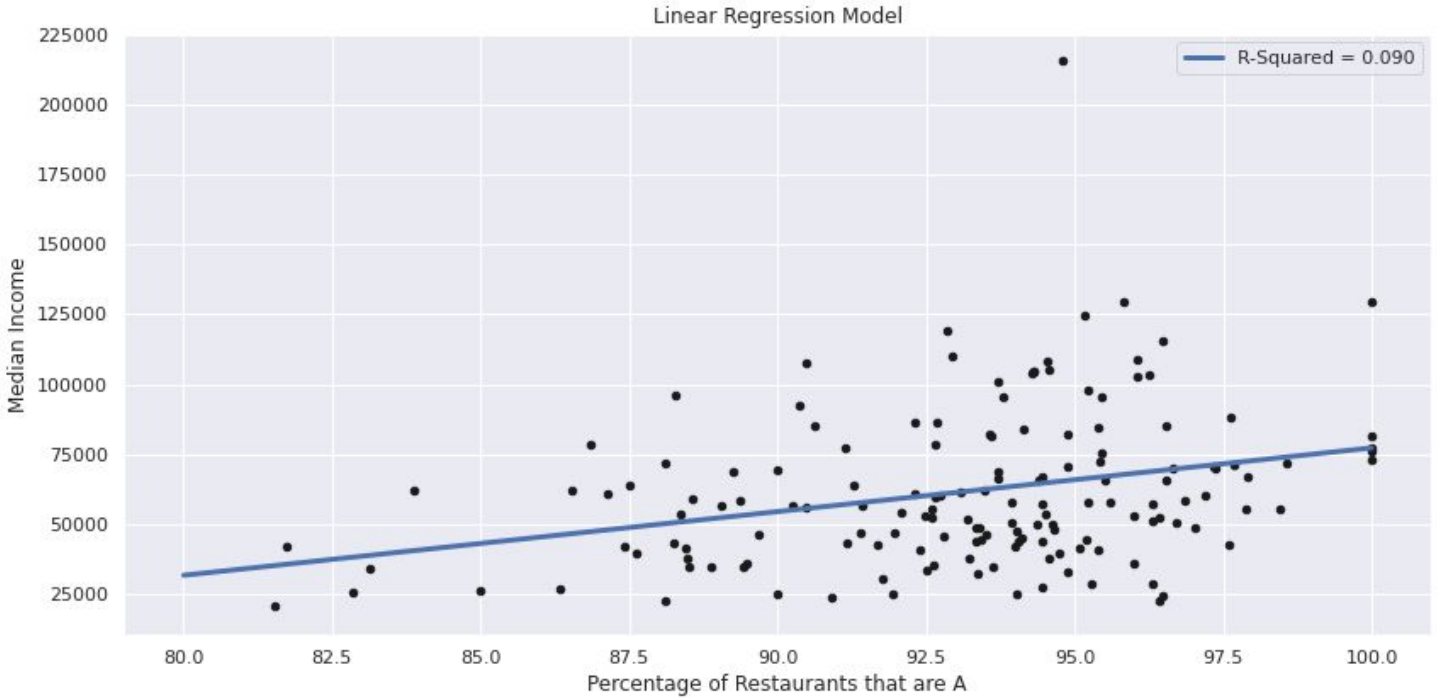
B-C p-value = 0.557

We can reject the null hypothesis that the two populations of incomes are equal for pairs A-B and A-C.

We fail to reject the null hypothesis for the third pair of populations, B-C.

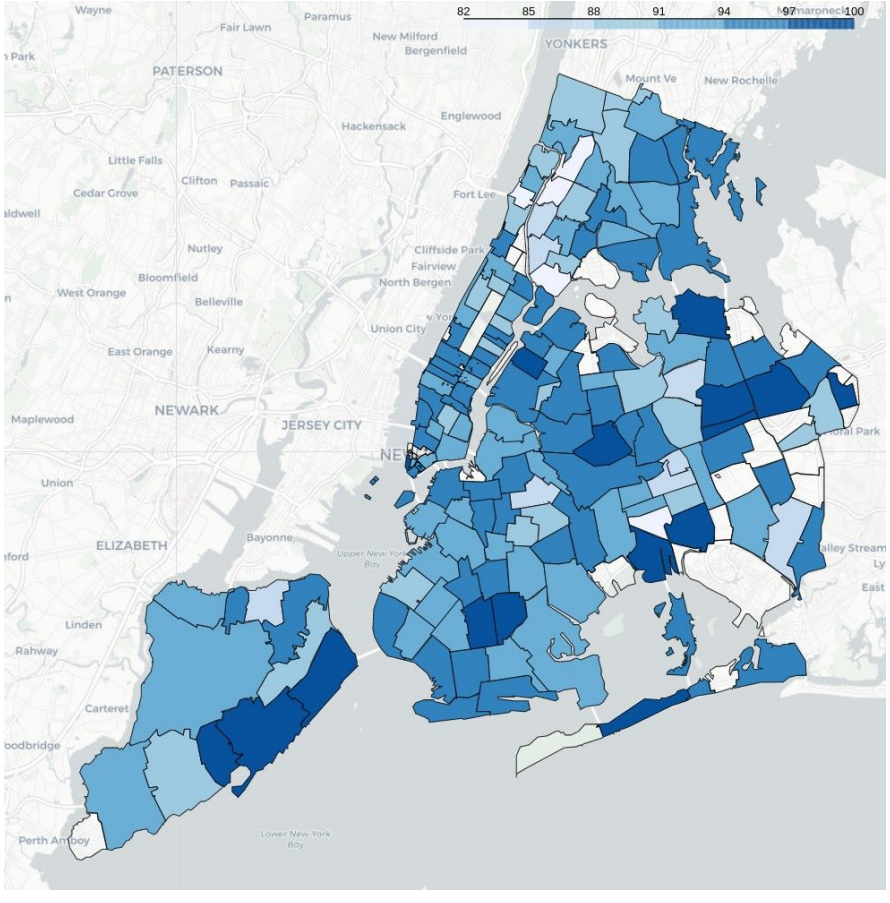
Grouping by Zip Code

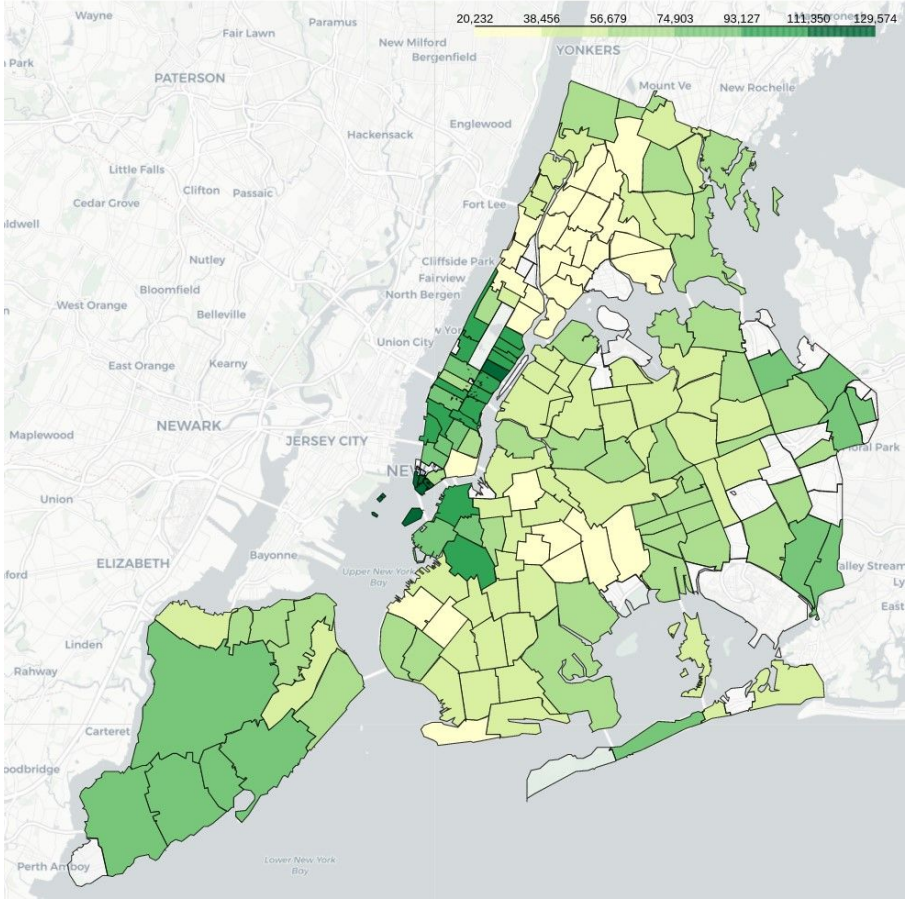
- Create a scatter plot of each zip code's Percentage of restaurants that are graded A, and the zip code's Median Income



Note: Zip Codes with total # of restaurants < 20 were removed as this caused some outlier data to appear

Choropleth map comparison - Left: %A, Right: Income





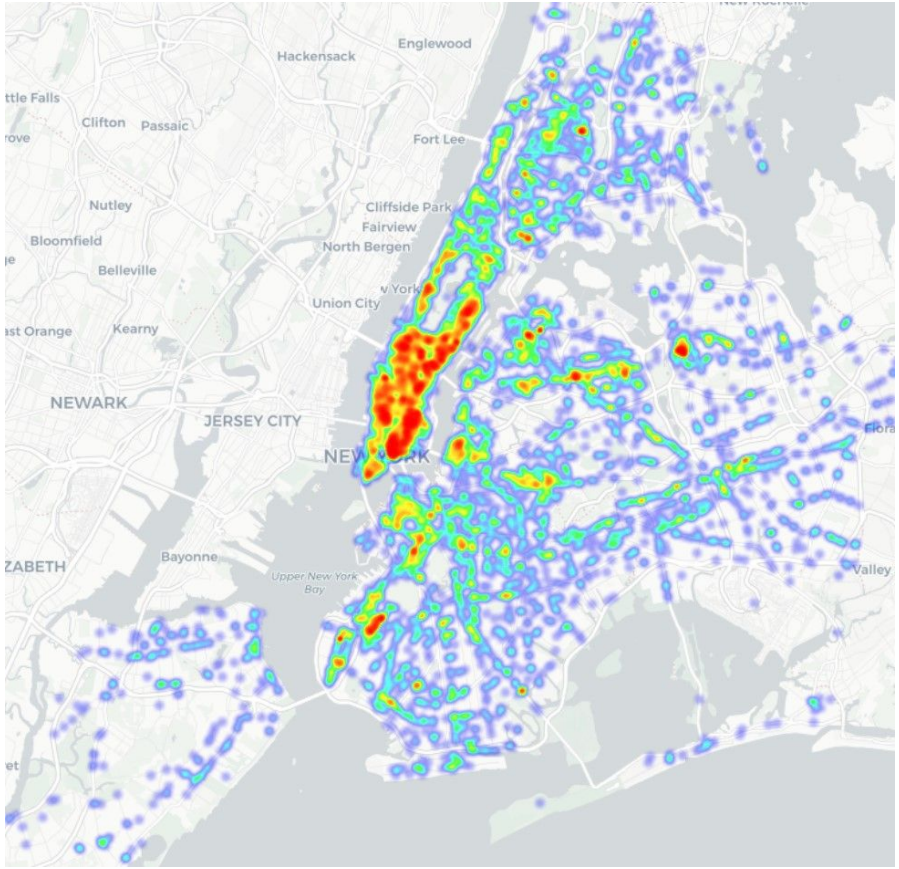
Looking at individual types of violations

We have breakdown on the exact nature of the violation that was cited. E.g.

- “Evidence of mice or live mice present in facility's food and/or non-food areas.”
- “Live roaches present in facility's food and/or non-food areas.”

Maybe certain types of violations are clustered together geographically?

Heat Maps - Left: Control group, Right: Vermin-related



Further Work

1. Obtain Yelp price point data.
2. Factor in ethnic demographics of zip codes. Look further into types of cuisines that make up a neighborhood.

Sources

[1] : https://www.osc.state.ny.us/files/reports/osdc/pdf/nyc-restaurant-industry-final.pdf

Images

- https://en.wikipedia.org/wiki/New_York_City_Department_of_Health_and_Mental_Hygiene
- https://medium.com/cusp-civic-analytics-urban-intelligence/what-can-we-learn-from-restaurant-grading-in-new-york-city-f31c5b079543

Questions, Comments, Criticism Welcome