DSE I2700
Project Sketches
Team Members: Stephanie Luk, Kuan Chen, Jiffar Abakoyas, Andrii Luchko

Visualizing NYC Restaurant Inspection Data

Introduction

If you ever traveled on the streets of New York City, chances are that you have noticed big letter grades (of mostly an 'A' or 'B') framed on the glass window of a restaurant or cafe's store-front. These letter grades are the results of unannounced inspections of restaurants conducted by the Health Department at least once a year. Inspectors check for compliance in food handling, food temperature, personal hygiene and vermin control. For the start of this project, we created sketches of visualizations that could show the overall trend and insight to certain relationships of the restaurants in the city based on the results of these inspections.

What is the data?

The data we are analyzing is the result of the restaurant inspections. It is the New York City's Department of Health and Mental Hygiene's DOHMH New York City Restaurant Inspection Results. The data can be found here: Restaurant Inspection Results.

Each violation of a regulation gets a certain number of points. At the end of the inspection, the inspector totals the points and this number is the restaurant's inspection score — **the lower the score, the better the grade**. Restaurants with a score between 0 and 13 points earn an A, those with 14 to 27 points receive a B and those with 28 or more a C. Inspection results are posted on the Health Department's website. More about how a restaurant is graded can be found here. Since July 2010, the Health Department has required restaurants to post letter grades showing their most recent inspection results.

Dimensions and features of the data

The dataset contains 395,000 rows, where each row represents one record of an inspection. There are 26 columns each of which represents a feature of an inspection. Some of the important features which we will use for our analysis include:

- unique identifier, name, borough, zip code, cuisine description, inspection date, violation code, violation description, critical flag, score, grade, inspection type, latitude, longitude

Why is this important?

According to the DOH, there are 24,000 restaurants in New York City, giving restaurant goers plenty of options. People looking for a place to eat out can find the grade of the most recent inspection posted on the window of the restaurant, but also they can look at their particular restaurant on the database and look at all the details of its inspection history before even visiting the establishment, helping them to decide where to safely eat.

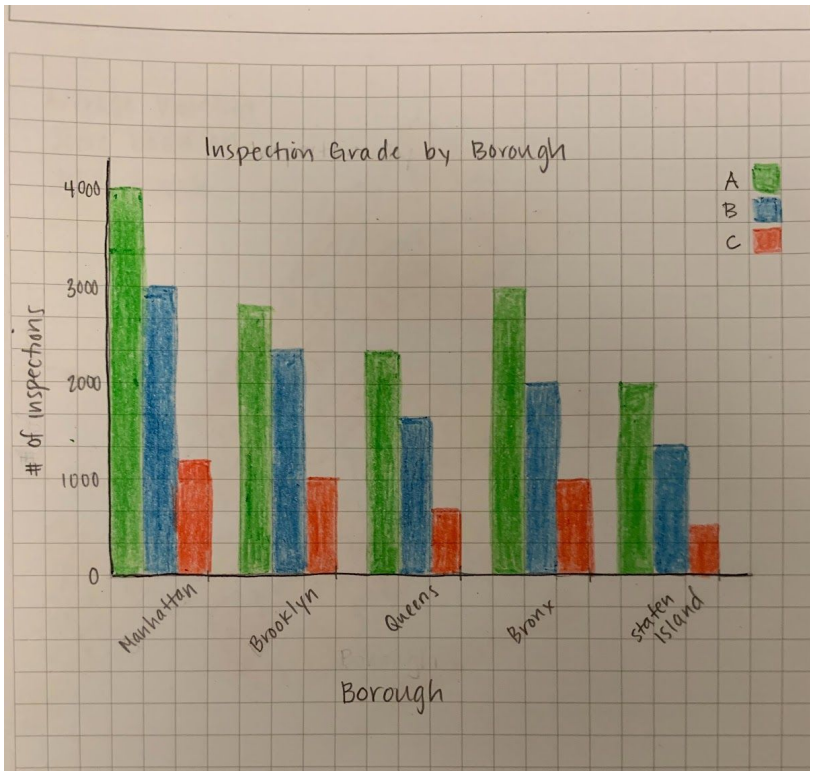
Beyond the power the database gives individuals the freedom of informed choice, this dataset can also be used by policymakers as well as concerned citizens to improve the overall conditions of the restaurant industry in NYC by discovering insights and trends that can help bring change at a larger scale. In order to do that, it is possible to look at the overall trend of restaurant inspections at a city level and visualize different features and see how they are changing over time. That is what we set out to do in this project.

Research questions?

- Do establishments near subway entrances have a higher chance of certain violations (ie. rats)?
- How do grades differ across cuisine types?
- How do grades/scores differ across zip codes?
- Are certain regions (areas) more prone to certain violations?
- Are there any violations that more frequently receive citations by region? (Knowing this can help the city focus its prevention and inspection).
- Is it possible to predict an establishment's next inspection grade?

Why the chosen visualizations is an appropriate ones to answer those questions?

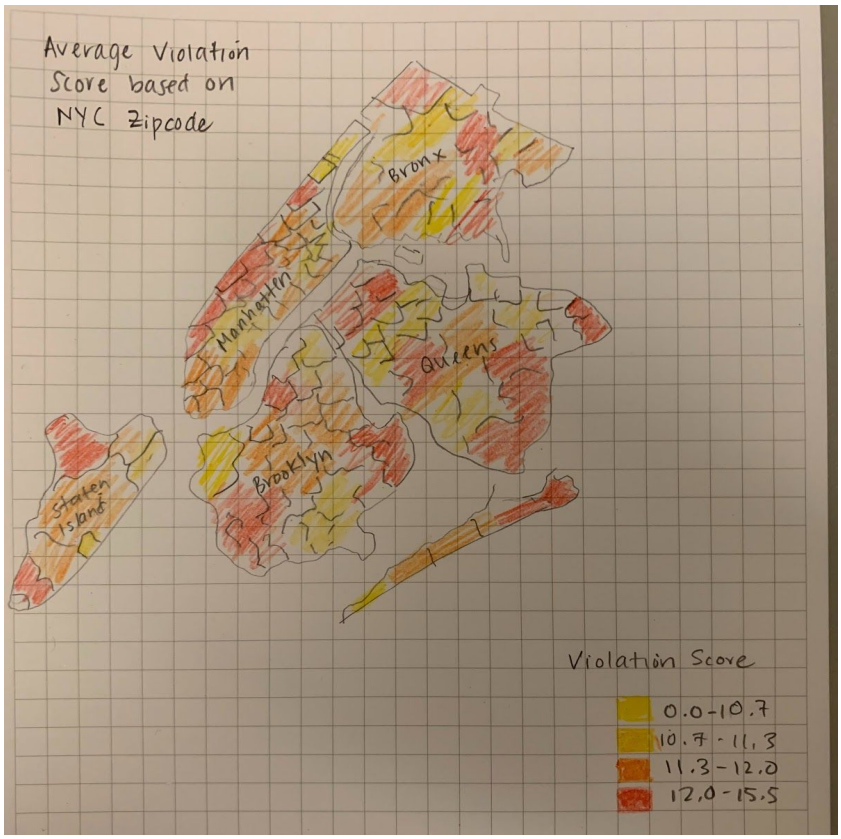
Stephanie



Inspection Grade by Borough (Bar chart)

Each establishment is given a Grade after their inspection. One way to show how many grades (A's, B's, and C's) were given to establishments by each borough, is by a bar chart. Each

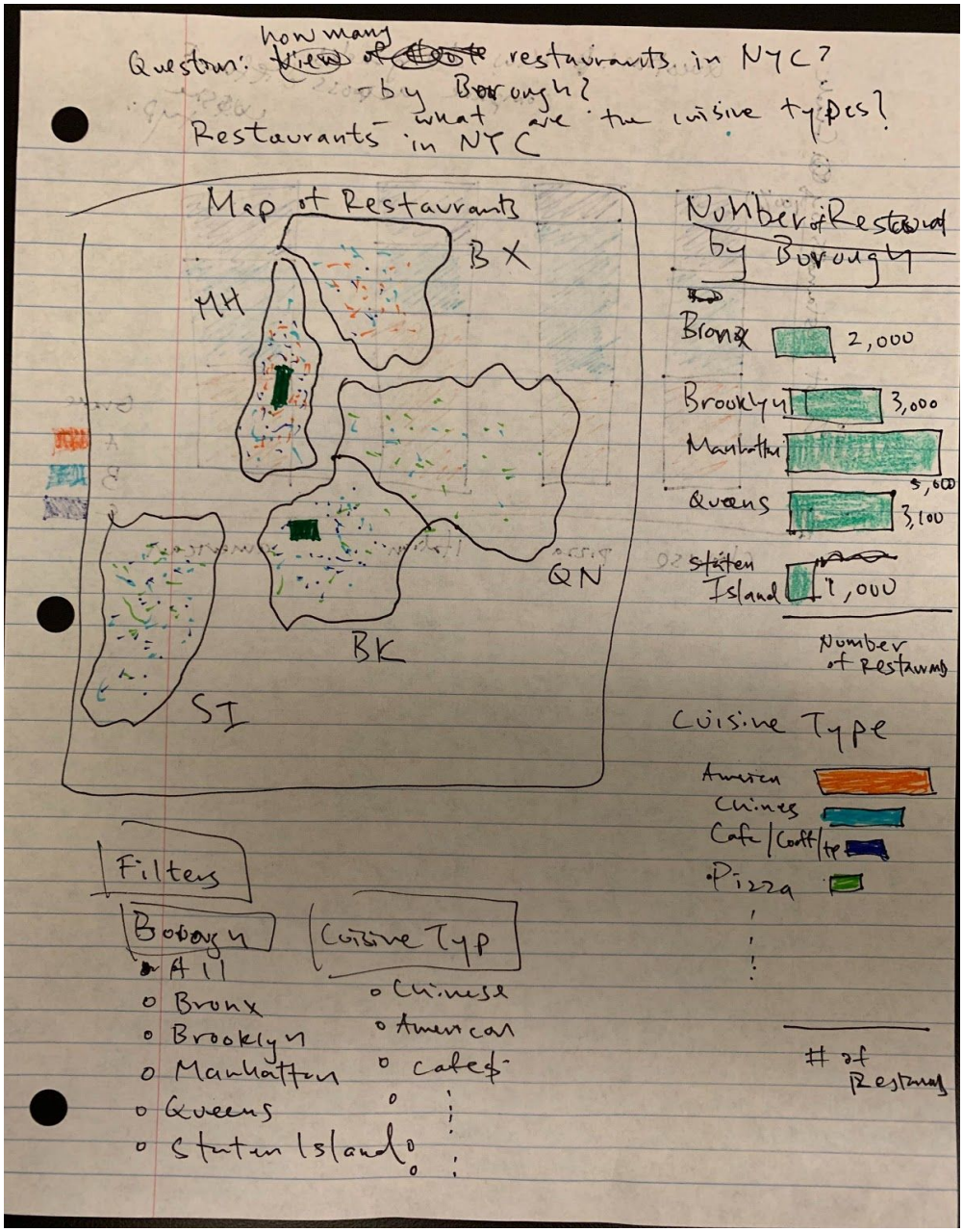
borough has three columns, one for each grade (A, B, and C), and a different color to represent each grade. By showing how many inspections in each borough were given for each grade, it gives an overview of how the distribution of grades between each borough. This visualization shows potential relationships between the location of an establishment and their respective grades. It also shows which borough has the most or least number of establishments, and which grade is the most frequent in each borough. This type of graph could also be done by percentage instead of number of inspections along the y-axis.



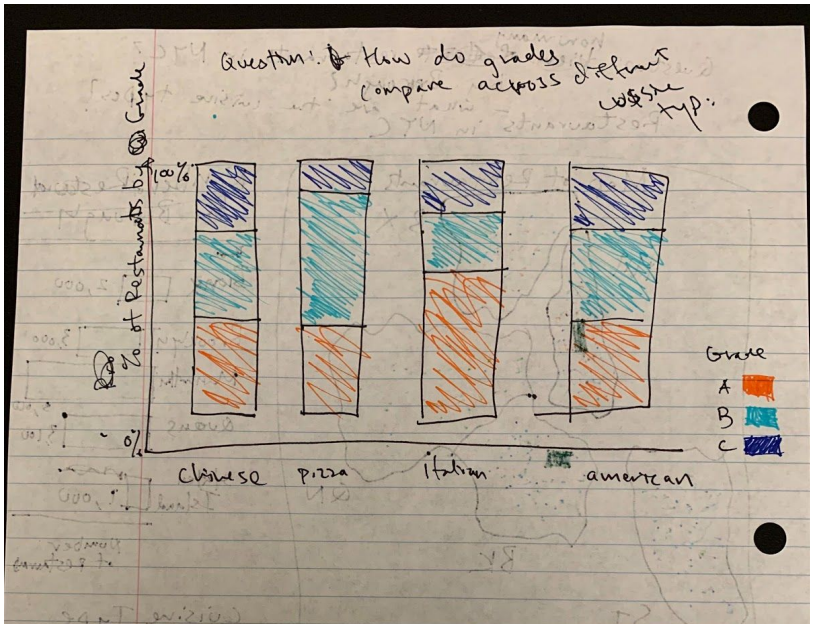
Average Violation Score Based On NYC Zipcode (NYC Heatmap)

It would be interesting to use a heatmap over a map of NYC to show the average violation score based on NYC zip codes. It gives a better idea of the spread of the establishment violation score throughout NYC. It could show how certain areas show higher violation scores, how certain areas show lower violation scores, or if certain areas have a mixture of both high and low violation scores.

Kuan



Geographic locations of restaurants by borough and cuisine type — Dashboard:
It would be useful to have a dashboard mapping out all the restaurants in the five boroughs. The end user would be able to filter the restaurants on the map by borough and cuisine type. A hover over tooltip would show the details of the restaurant — including the name, address, and current sanitation grade of the restaurant — when the cursor is pointed at a restaurant (point) on the map. To the right of the interactive map, counts of the number of restaurants by borough and cuisine type would be shown in bar charts in descending order.



***Visualizing sanitation grades across different cuisine types**:*
I am interested in how sanitation grades vary across different cuisine type. Since the number of restaurants for certain cuisines — such as pizzerias, chinese fast food restaurants, American cuisines — greatly dominate in counts than others, using a 100% stacked chart to compare the relative frequency of sanitation grades (e.g., percentage of pizzerias that have grade A, B, C, totaling 100%) across different cuisines could give us some insight into this question.

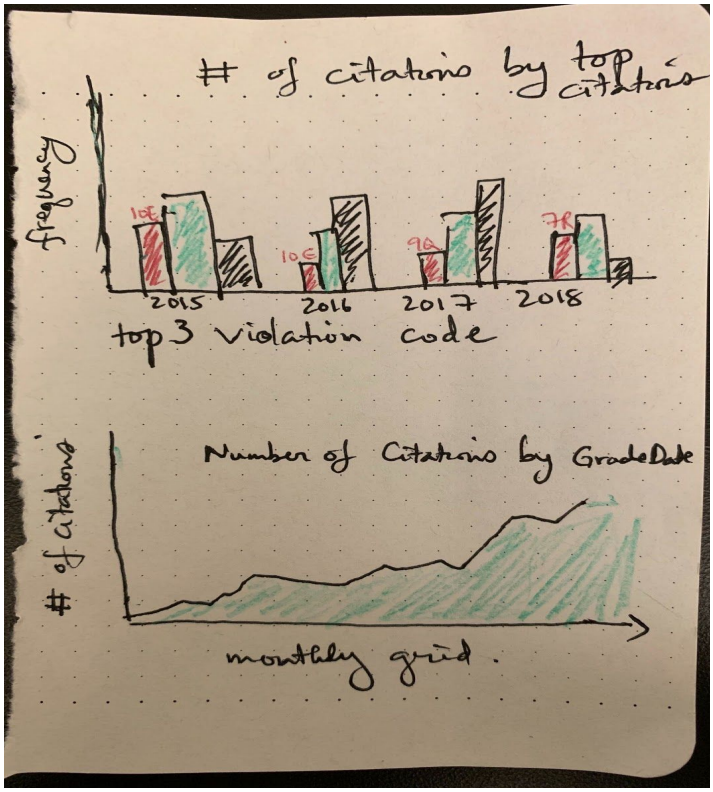
Jiffar

1.Proposed studying the most frequently occuring violation codes. This can help understand if certain types of violations occur more commonly. Knowing this can help city officials implement policies that focus on addressing those issues. This will address the root cause and reduce the infraction in the future. The city can implement preventive efforts to educate restaurants on better health practices using the results of this visualization.

2. Also proposed visualizing the number of negative inspections over time. This will help understand if the city has been issuing more negative results or less over the past three years (taking the opening of new restaurants into consideration). For example, do we have the same ratio of total negative citations to total restaurants over the years? This will show us if restaurants have become safer or not over time? It will also serve as a measure of the effectiveness of the inspection program.

Please see next page for visualization sketches.

Please see the below sketches to show visualization for the above two points:

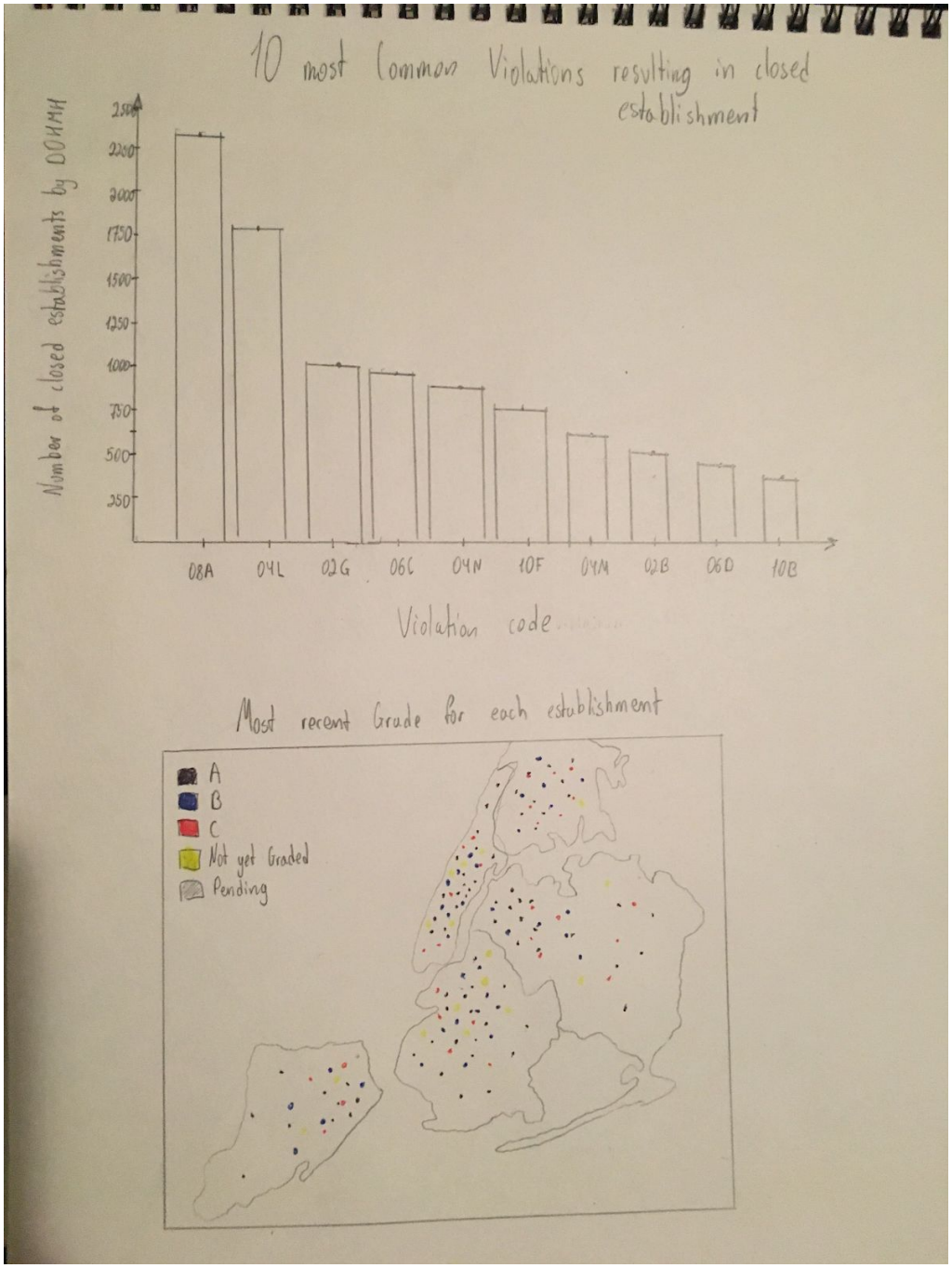


Title: Top 3 violation types by year.
Method: Clustered Bar Chart.
X-axis: Bar charts grouped by year
Y-axis: Count for the violation

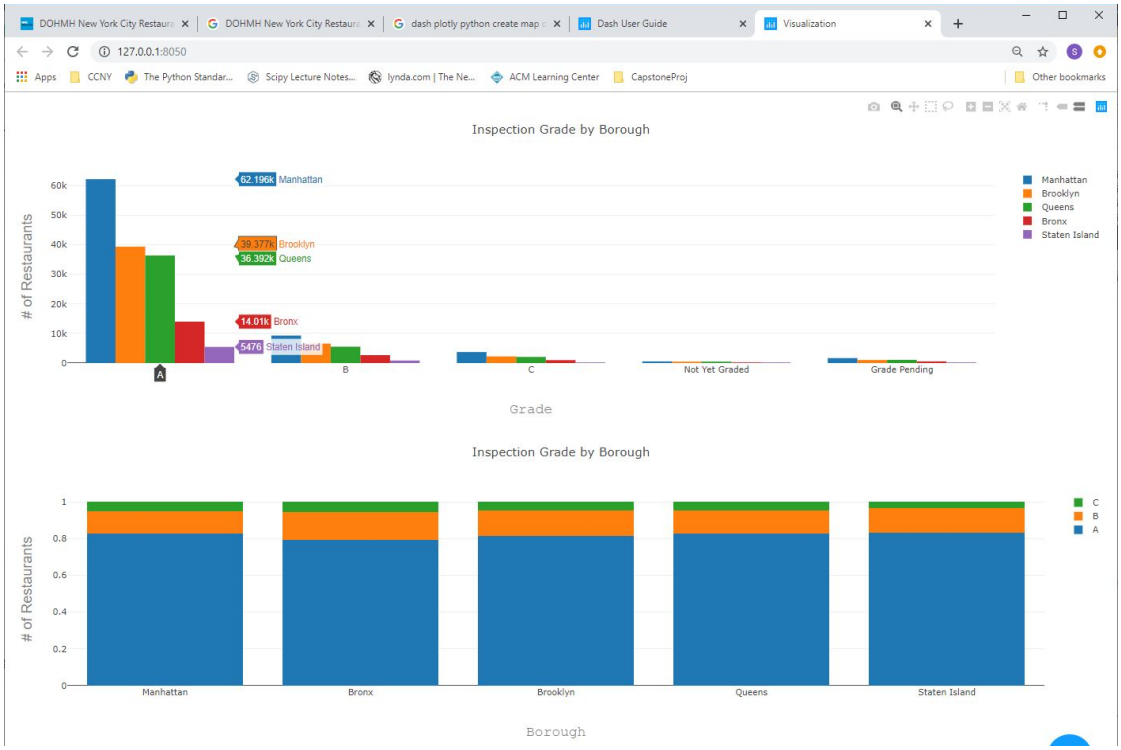
==========

Title: Number of inspections with grades of C
Method: Time-series plot
X-axis: Time in months.
Y-axis: Count of inspection

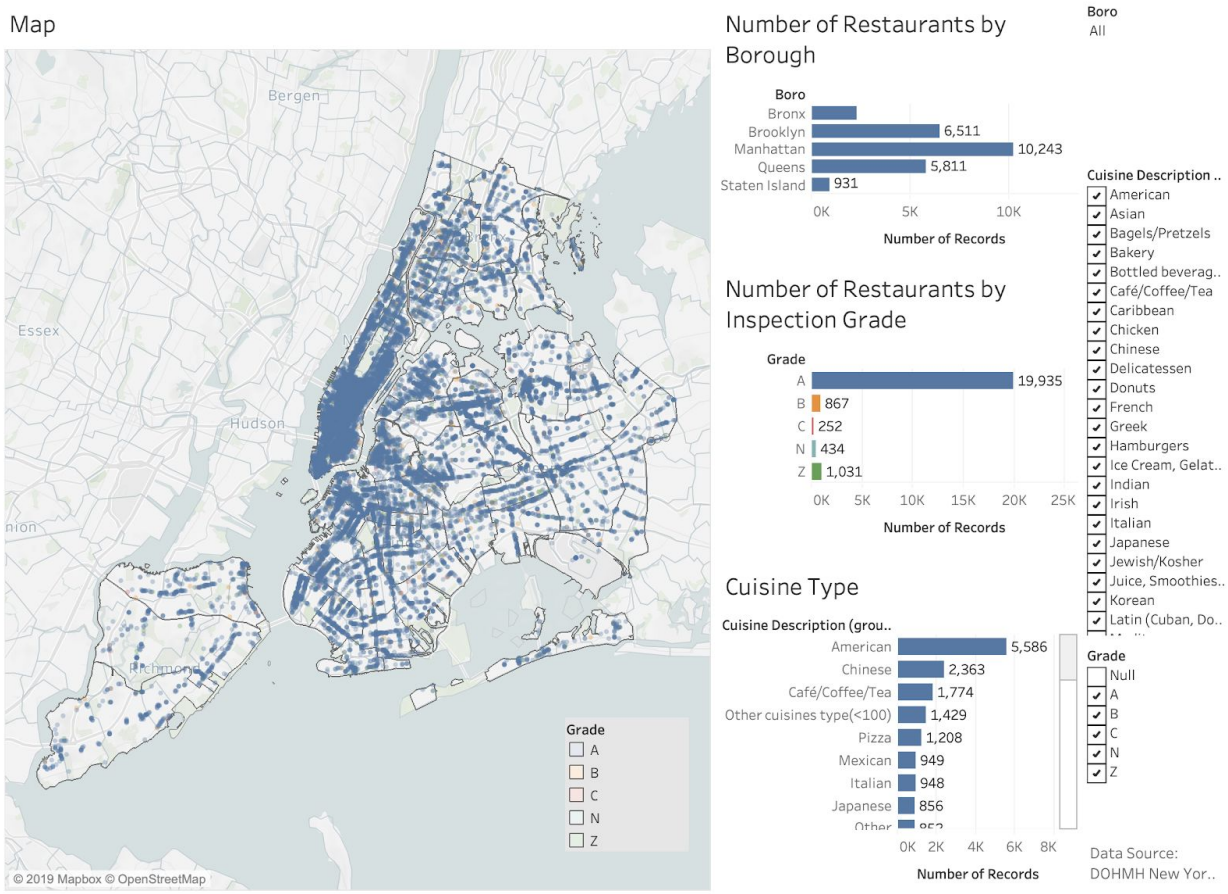
Andrii


10 most Common Violations resulting in closed establishment
Number of closed establishments by 00HMH
2500
2250
2000
1750
1500
1250
1000
750
500
250
08A
04L
02G
06C
04N
10F
04M
02B
06D
10B
Violation code
Most recent Grade for each establishment
A
B
C
Not yet Graded
Pending

Screenshot of bar charts on Dash app



Bar charts showing Inspection Grade by Borough. The top one is a multiple bar chart: the x-axis represents the boroughs and each column represents the boroughs. When you hover over the columns, it shows the labels and values of each of the columns. The bottom bar chart is a stacked bar chart: the x-axis is the borough, and each color in the column represents the grades A, B, and C.



Our next step is to have an interactive visualization like this:
This visualization has a map of NYC and points representing all the restaurants. You can filter the data by borough, inspection grade, cuisine type, and cuisine description. As you click and select the attributes you want, the map will zoom in or identify the results accordingly.