In one of the greatest cities in the world, NYC, we have an abundant amount of options to choose from when it comes to where we eat, whether we are looking for fine dining, fast food, or something in between. The cuisine choices seem just as endless. Is there a good way to help choose a restaurant to eat it or a general location? Although not an easy task, one thing we can all agree on is that nobody want to eat in a restaurant ridden with health violations. In our project, we will explore and analyze NYC’s restaurant inspection results from 2013-2017 of the five boroughs.

We are interested in answering:

1. Which cuisines have the least and most violations along with their associated score.
2. What parts of NYC have the least and most violations?
3. We hypothesize that location is highly associated with inspection grade and so we will be searching for patterns between these variables.
4. We are also interested in seeing how inspection grade and score changes over time (month and years) based upon cuisine and location.
5. A description of violations is also provided in the data and we would like to better understand the common causes of those violations based upon the description.
6. Is there a pattern/trend in violations/inspection score/grade based upon restaurant chains?

We found this data set by exploring NYC open data sets. Our love for food and health made this a great option. One of our team members who recently moved to NYC is highly selective about food she eats due to her rare health condition and hence, having the kinds of questions mentioned above answered can prove to be a great asset to both her and others suffering from health conditions or who may just be picky eaters! The data can be found and downloaded here: <https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/xx67-kt59>

To download, click the export button on the right -> Download as -> choose your format.

Team and Task allocation:

Team members:

1. Jonathan Galsurkar
2. Lakshya Garg

|  |  |  |
| --- | --- | --- |
| S. No | Task | Assignee |
| **1** | Cleaning up the data | Both |
| **2** | identifying missing data-including understanding what missing values mean | Both |
| **3** | Geocoding data | Both |
| **4** | Questions 1, 2, 5 | Jonathan Galsurkar |
| **5** | Questions 3, 4, 6 | Lakshya Garg |
| **6** | Pushing the data analysis to GITHUB website | Both |

We will have to do further research and analysis to see how scores and grade are related. There does not seem to be a clear pattern. At first, it appears that high scores