**Problem**

There are many New York City restaurants and health inspectors that evaluate the restaurants a certain number of times each year. “Trained to prevent the spread of foodborne illnesses, inspectors look to ensure restaurants comply with guidelines to help prevent the spread of coronavirus, as well” (National Restaurant Association, 2020). The National Restaurant Association mentions that there is a list to make sure a restaurant is in order such as have all the paperwork the inspector needs ready, do not commit obvious and common violations, follow social distancing guidelines, take temperatures, and remember the operation’s food safety and sanitation is very important (National Restaurant Association, 2020).

Most customers that order food from restaurants do not keep up on how restaurants perform during health inspections. It seems like people hope for the best but when a restaurant comes out in the news for major health violations, customers tend to be a little shocked. Another article discusses the best ways to make sure companies do well on inspections. For instance, paying the employees well, provide training on the specific aspects of each job, self-inspections, cleaning and maintenance schedules, and how often are the health inspections (Menu Cover Depot, 2021).

**Hypothesis**

There will be a lot of New York City restaurants that will do well but there will be some that fail inspections. Various words in violations will potentially include food, temperature, undercooked, uncleanliness, and illness. I am not too familiar with the New York City boroughs and it is tough to tell which one will have the most inspections.

**Data Used**

The dataset I used is the “DOHMH New York City Restaurant Inspection Results”. This data goes from 2011 through 2021. There are 419,155 records with 26 fields.

The following fields appear in the dataset:

* CAMIS – Unique restaurant identifier
* DBA – Restaurant that is doing business as
* BORO – New York City Burrough
* Building – building number
* Street – Street the restaurant is on
* ZIPCode – ZIP code the restaurant is in
* Phone – Restaurant phone number
* Cuisine Description – What types of food the restaurant serves
* Inspection Date – date of inspection
* Action – What action, if any, was taken against the restaurant
* Violation Code – Unique identifier for the violation
* Violation Description – Descriptions of the violation
* Critical Flag – How critical is the violation
* Score – How the restaurant performed on the inspection
* Grade – Letter grade that was given to the restaurant
* Grade Date – Date the letter grade was assigned
* Record Date – Date when the extract was run to produce the dataset
* Inspection Type - Type of inspection that was performed
* Latitude – Location of the restaurant
* Longitude – Location of the restaurant
* Community Board – Unknown
* Council District – Unknown
* Census Tract – Unknown
* BIN – Unknown
* BBL – Unknown
* NTA – Unknown

**Methods**

I used various methods to evaluate the New York City restaurant inspection information. I used R programming and Python. I cleaned the data by removing fields that were not needed, summarizing the data, updating the data types, and removing null values.

Another method I used was word networks. Word networks are used to identify words that were used in violation descriptions. The bigger the word, the more times the word was used in all the violations that were assigned to restaurants. The words “Contact” and “Food” were used the most with several words in pink that were just behind those words.

Text

Description automatically generated

I summarized the “A”, “B”, and “C” grades for each borough in 2020. I found the highest amount with the A grade was in Manhattan. This could be due to there being a lot more restaurants in this borough than the other ones. Manhattan also have the highest in the B and C grades. Staten Island had the least number of grades.

Chart, bar chart

Description automatically generated

Lastly, I used the KDE plot to show that most restaurants have few health inspection violations. The lower the score, the more violations that were encountered.

Chart, histogram

Description automatically generated

**Conclusion**

Overall, there is a lot of data that can be evaluated. I found that Manhattan had the most A, B, and C grades. This could be due to there are more restaurants in this borough than any other New York City borough. The lower the Score and Grade, the higher amount of violations there will be for a specific restaurant. The most common word in the violation descriptions is “food” followed by “contact”, and “surface”.

**Ways to Improve**

There are several ways I can improve analyzing the New York City food inspections dataset. I could evaluate whether certain Inspectors gave more violations than other. Also, I could create a map of where all the restaurants appear in the New York City area. I could also evaluate whether certain areas had higher violations than others and how many restaurants improved during the re-inspections.

**Appendix**

Questions to answer in Presentation:

* What restaurants have the worst ratings/scores?
* What locations in New York City have the worst/best scores?
* What violations occur the most?
* What categories are the best/worst for violations?
* What Inspector employees were more favorable to give good/bad scores?
* What is the average inspection score?
* What days were inspections mostly completed on?

**References**

* NYC OpenData. 2021. DOHMH New York City Restaurant Inspection Results. Retrieved from <https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j>
* Centers for Disease Control and Prevention. 2021. Coronavirus Disease 2019 (COVID-19). <https://restaurant.org/articles/news/inspections-may-look-different-in-the-wake-covid19>
* <https://www.menucoverdepot.com/resource-center/articles/restaurant-health-inspections/>