Executive Summary

Introduction:

Ever imagined eating at a restaurant, enjoying your delicious food, only to see a rat run into the kitchen? Such has been the unfortunate case of several new Yorkers over the years. What makes the matter even more aggravating is when patrons fall sick because of eating foods likely contaminated by rats at restaurants where the patrons consumed the foods. Historically, rat sightings have been handled by the New York department of Health and Food Safety. This group of aspiring Data Scientists at Rutgers Data Science bootcamp 2024 set out to find out answers to

several questions regarding rats and restaurants in New York City. The goal of the group is to derive insights from publicly available data with an objective of providing crucial unbiased analysis and summary regarding the subject to the public. This will enable or promote good decision making on what restaurants to go to for food in New York.

Body

The student researchers ploughed through a plethora of datasets which are publicly available platform of Kaggle. We narrowed down to three most promising datasets for our project, namely **DOHMH New York City Restaurant Inspection Results** , **NYC Rat Sightings**. and a third dataset having latitudes and longitudes csv file.

According to a December 2021 BBC article titled "If you eat here, you're dining with rats", (Hooker, 2021) eaters have jumped off their feet in shock as they witness rats run around dining areas. Indeed, the author of the article has explained how eaters have lost their appetite quickly whenever they have sighted rats. The author argues in the article that now rats are "bigger and bolder, they jump at you. They're gymnastics, doing backflips" (Hooker, 2021).
To start this project off, the team came up with three Research Questions which guided the project analysis, namely: Do rat sightings lead to higher restaurant citations by the city of New York (NYC)? Is there any relationship between the number of rat sightings and the number of violations cited by NYC? Do high rat zones expect to have higher closures. To answer these questions, the team further developed more micro questions that led the analysis to discover the answers to the macro level research questions above. These included the following questions: 1. How many A, B, C restaurants were there in New York City from September 2010-September 2017? 1b. Is there an association between restaurant type and rating? 2. How many rat sightings were there in New York City from September 2010-September 2017? 3. Do certain Boroughs

have higher restaurant ratings? 3a. Do certain Boroughs have more restaurant violations? (Violations per restaurant by Borough) 4. Do certain Boroughs have more rat sightings? 4a. Do certain Zip Codes have more rat sightings? -created bins 5. Is there a relationship between rats and restaurant violations within zip code? 5a. Do zip codes with more rats have more violations? 5b. Do "high rat" zip codes have more restaurants closed by DOHMH when compared to "low rat" zip codes?

Literature Review:

According to New York Restaurant Grade Program, the New York City initial inspection does grade restaurants based on an A,B,C classification. The A grade are typically those that received less than 14 points, and are slated for next inspection in 5 – 7 months if initial score achieved was between 14 to 27 points. The B grade was awarded to restaurants that attained 14 to 27 points with a final grade posted until OATH Hearings Division inspection in 5 to 7 months. The C grade was awarded to restaurants with a score of 28 points with a scheduled follow up in 3 to 5 months.

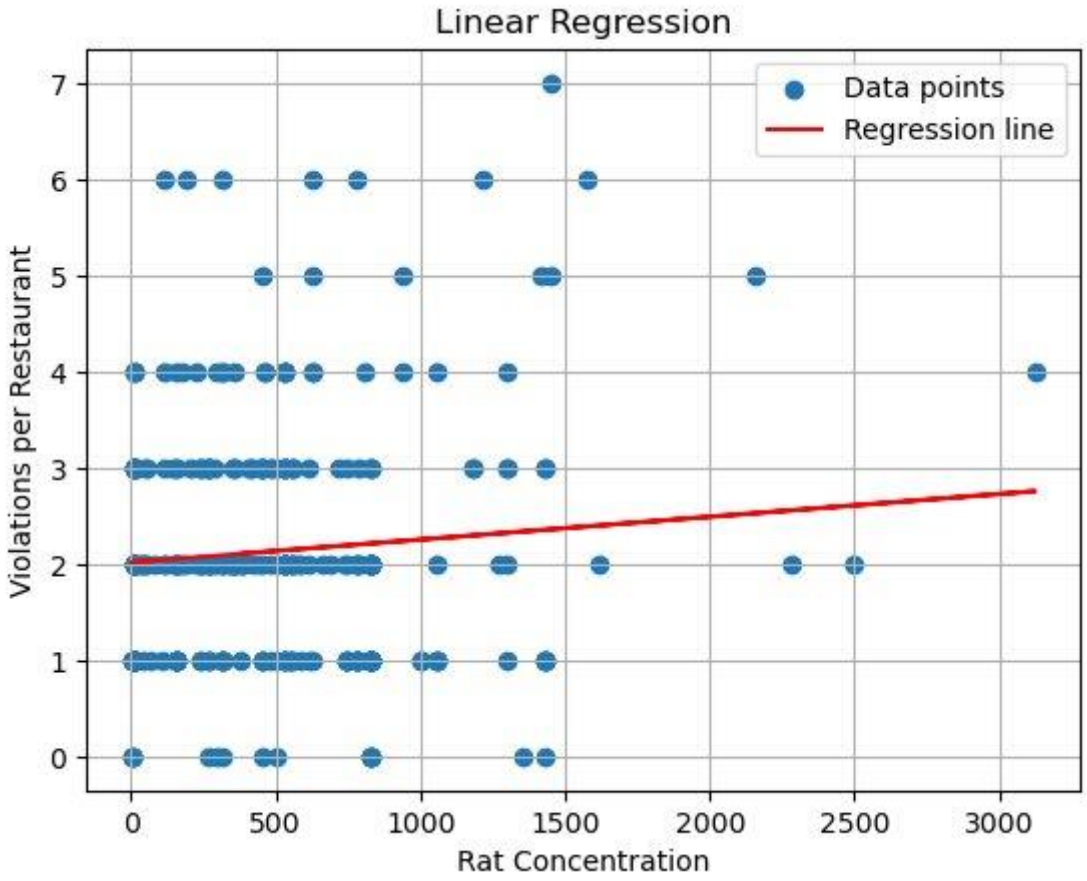
Criteria for choosing data:

The team wanted data with significant row entries or more than ten thousand rows and readily available data. Kaggle was chosen as a data source for this project because it is a robustly researched website frequently used by Data Scientists due to it's curated nature. This allows for quick analysis as well as use for the data in traditional data science workflows to produce insights.

Data Analysis and Findings:

The team cleaned the data using pandas and used the cleaned datasets to run various analysis using data analysis methods we have learned in the class thus far. Our research questions were pursued as follows:

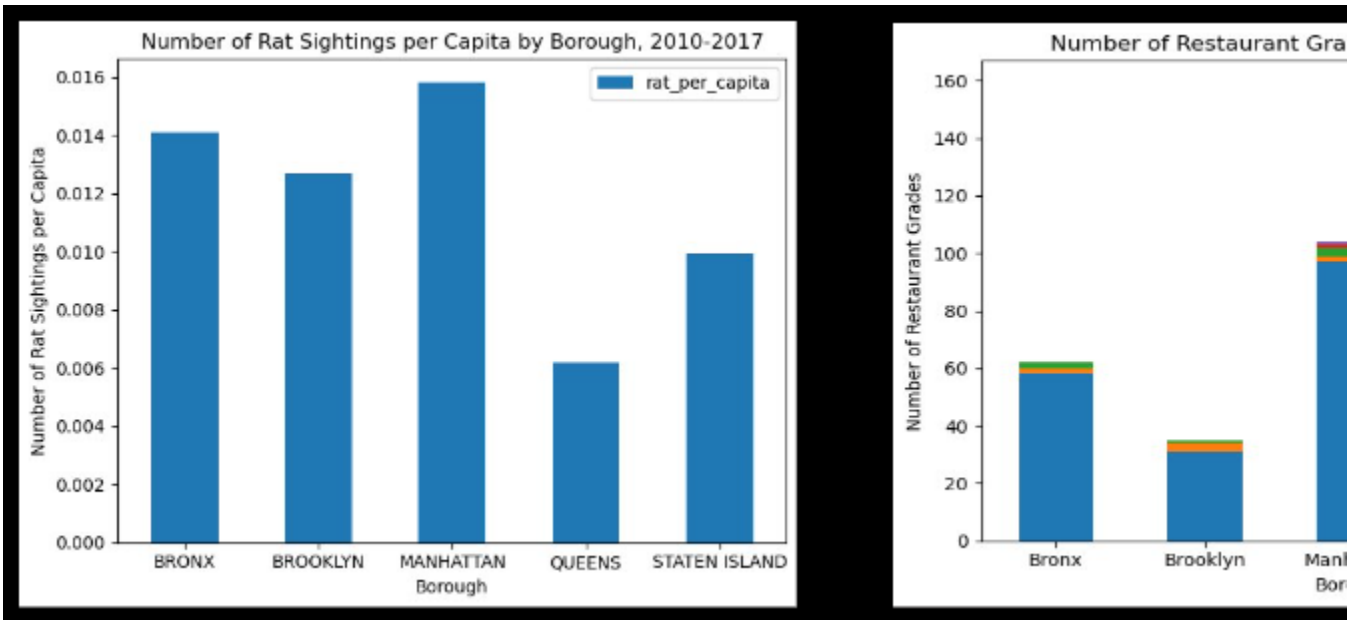
1. **Do rat sightings lead to higher restaurant citations by the city of New York (NYC)?** The team performed a Linear Regression analysis pitting Rat Concentration vs Violations per restaurant. As shown below in Index 1, With a p-value of 0.148, it was concluded that there was not enough statistical significance of the p-value to conclude a strong relationship between number of violations and rat concentration.



Index 1.

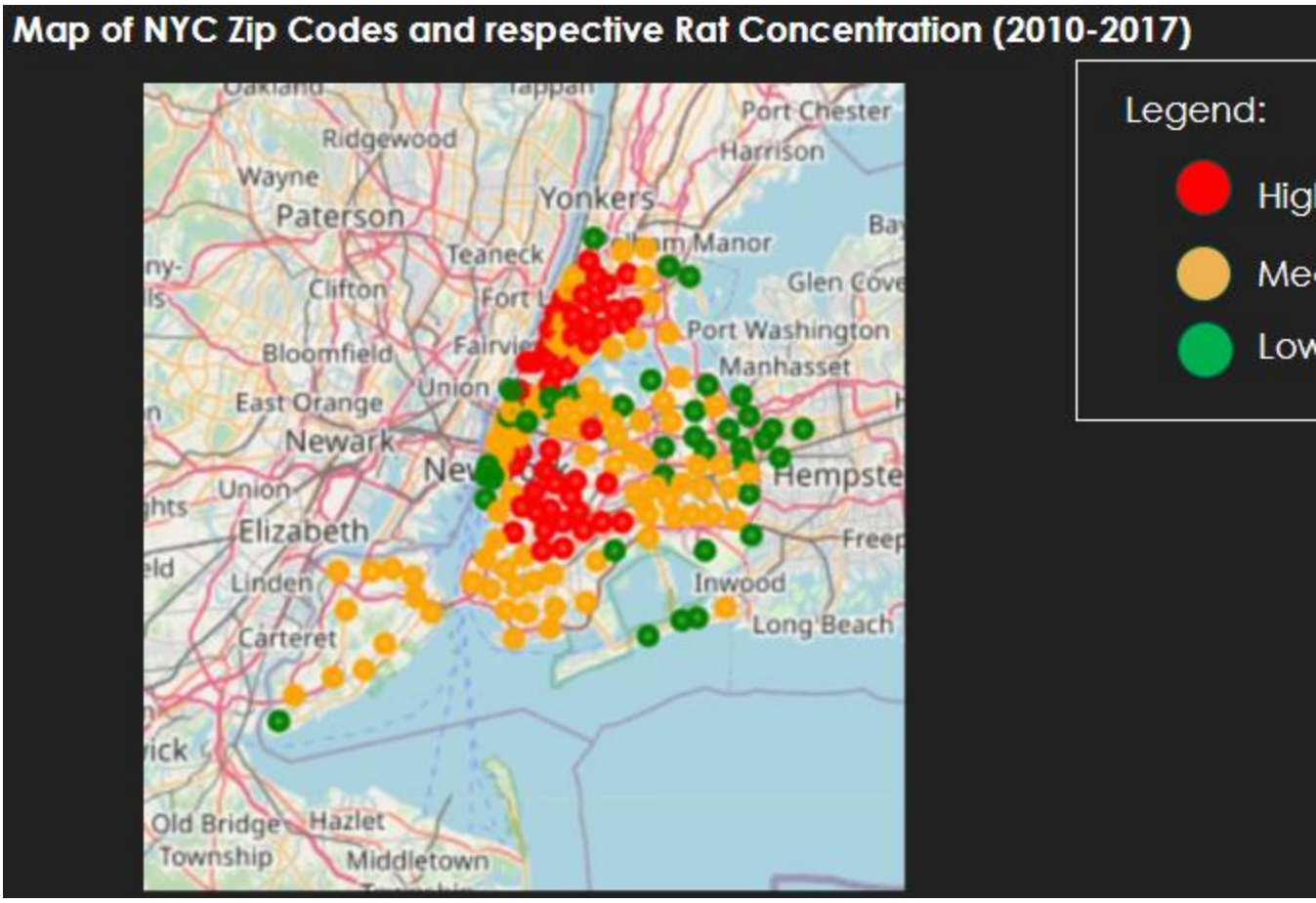
2. **Is there any relationship between the number of rat sightings and the number of violations cited by NYC?**

Based on the below histograms, there is no relationship between rat sightings and violations. For instance, Queens which has a record number of at least one violation is not among the boros with highest number of rat sightings



Index 2.

3. Do high rat zones expect to have higher closures. Based on the heat map shown below, High rat concentrations did not match zones with high closures of restaurants. This is good news for any investors that would like to open a restaurant in zones with high rat concentrations. The group pursued zip code mapping to identify visually areas with high rat, medium rat, and low rat sightings.



Index 3.

Statistical Tests that Validated above findings

Various tests were carried out on the two datasets for the period 2010 - 2017, including Stacked bar graphs with a Chi-Square test which evaluated the number of restaurants grades by Borough. The p-value was found to be 0.74402. This means that there is no statistically significant relationship between borough and restaurant rating. The group used a Pie Chart to visualize grades attributed to various cuisine types and restaurant ratings. With a p value of 0.15910, it was found that there was no statistically significant association between cuisine type and restaurant rating. Next, the team it's sight on evaluating the number of violations per restaurant by Borough during the same period. For this test, the team merged data frame with

division. For this test, the team found that p-value was to be 0.008639. This meant that there is a statistical significance relationship between violations category by borough. Higher number of at least one violation were found in Queens, New York. When the number of rat sightings were tested, Brooklyn had the highest number of sightings, followed by Manhattan and Bronx in that order. This did not exactly pan out as we had expected for Queens to have a high number of sightings.

Summary:

In spite of our initial bias that areas with high rat concentrations would likely have restaurants with high violation citations, its could not be validated statistically.

Citations:

Hooker, L. (2021) *'if you eat here, you're dining with rats', BBC News*. Available at: https://www.bbc.com/news/business-59624540 (Accessed: 17 June 2024).

Letter grading for restaurants (no date) *Letter Grading for Restaurants - NYC Health*. Available at: https://www.nyc.gov/site/doh/business/food-operators/letter-grading-for-restaurants.page#:~:text=Since%202010%2C%20New%20York%20City%20has%20required%20restaurants,can%20easily%20be%20seen%20by%20people%20passing%20by. (Accessed: 17 June 2024).

Retrieved from https://www.kaggle.com/datasets/joeleichter/us-zip-codes-with-lat-and-long

Retrieved from https://www.kaggle.com/datasets/new-york-city/nyc-rat-sightings

Retrieved from https://www.kaggle.com/datasets/shivanielakurthy/dohmh-new-york-city-restaurant-inspection-results