# Headline

### Preview

Every year, approximately 48 million Americans get sick from food poisoning with 128,000 of them being hospitalized, and 3,000 dying from food born illnesses, making food poisoning a real public health threat. As Americans consume about a third of their calories from food not prepared at home, and spend almost half of their food dollars on restaurant food, restaurant food safety is increasingly important. In New York City (NYC) alone, its residents eat out nearly 1 billions times a year, or about every person once every 3 days. And even though most people who eat out do not get sick, Jones and Angulo (2006) discovered a correlation between the consumption of food prepared outside the home and an increase in the rise of sporadic food borne diseases.

To keep this threat in check, the NYC Department of Health and Mental Hygiene (DOHMH) is charged with the responsibility of inspecting restaurants (including but not limited to bars, coffee shops, employee or university cafeterias, nightclubs, fixed site food stands, and bakeries). The manner in which the DOHMH does its inspections to measure the restaurant’s sanitary condition is by using a scoring system where the lower the score the better the establishment is at adhering to food safety.

Rodents are a prominent feature of the urban setting. Rodents are also detrimental to society as they are a carrier of food borne diseases such as Salmonella. The more rats in an establishment, the higher the chance of contamination of food, leading to a higher rate of food-born illnesses related to rats. To help contain this threat, The NYC Department of Health and Mental Hygiene (DOHMH) does inspections of reported rodent sightings across the 5 boroughs.

This analysis looks for a correlation between income, number of rodent sightings, and restaurant health score. It asks whether there exists a correlation between the mean income in a zip code and the number of reported rodent sightings in a zip code versus the mean restaurant health score for all restaurants in a zip code and in NYC. In terms of the correlation, this study did find evidence for a correlation between number of reported rodent sightings and restaurant health score in a zip code in NYC. There are multiple limitations to this study including that zip codes are not an ideal measure for defining an area. Further analysis is required before being able to state a definitive answer.

### Background

#### Rodents

Rats have been reported as being prevalent in all areas, especially in densely populated urban areas such as NYC (2). In fact, NYC has been reported by Sullivan (2008) to be one of the cities with the largest rat populations in the United States (27). While several types of rats can be found in NYC, the most common rat found in NYC is the Rattus norvegicus, commonly known as the Norway rat, a burrowing species commonly found in subway and sewer systems (Clinton, 1969). These rats are known carriers of diseases, such as Salmonella enterica (22, 23, 24, 25, 26). These diseases are transmitted by foods, causing issues for those dining at establishments with rat infestations. One of the most notable diseases that rats can transmit is Salmonella, a bacteria that can spread through contaminated water (1, 39). This is a major health concern since Salmonella represents one of the most common foodborne illnesses, with an average incidence rate of 12 cases per 100,000 a year in New York State (1, 41). Infection in humans is often through ingestion of foods contaminated with animal feces, such as rat droppings, or cross-contaminated by other sources. Salmonella isn’t just a theoretical concern but rather a serious matter since there have been an estimated 100,000 annual antimicrobial-resistant infections caused by Salmonella in the United States (2, 43). Therefore Salmonella is an issue that needs to be minimized in the restaurant setting, and given that rats are a significant transmitter of Salmonella it is important to know if there is a correlation between the number of rodents and restaurant health scores.

The NYC DOHMH maintains a dataset on rat inspections and intervention visits in NYC. This program is based on the idea that rodents rarely occupy a single property but rather they occupy entire blocks or larger areas. The way a rodent inspection occurs is first a person has to observe a rodent, or rodent activity in a specific location. Then they have to file a complaint, whether online through the 311 portal or over the telephone. At this point the complaint is routed to the DOHMH who within approximately two weeks of receiving the complaint inspects the property, unless the property was recently inspected (in this case it is considered a duplicate complaint) (8). When performing an inspection, an official DOHMH inspector searches for signs of rats or conditions that rats favor on the exterior of the property. Signs of rat activity include fresh gnaw marks, burrows (earthen or structural), tracks, fresh droppings, active rodent pathways, and live or dead rodents. Conditions that rodents favor include those that provide protection or shelter, such as clutter or overgrown vegetation and conditions that provide feed for rodents, such as exposed or uncontained garbage (8, 40). If the inspector does not find such signs or conditions, the property passes the inspection and as a result the 311 complaint is closed. If however the inspector does find signs of rats or conditions that rats favor during the inspection, then the property fails the inspection (8). The DOHMH does provide incentives for responsible parties to pass inspections, such as fines ranging from $300 to $2000 that the Office of Administrative Trials and Hearings (OATH) administers (8). The DOHMH collects this dataset to inform community members of where there are likely to be more rodents so that they can take informed action based on this data to treat their rodent issues.

#### Restaurants

Restaurants are a major part of NYC with there being over 8,500 establishments in 2019 (44). Given the frequency that New Yorkers eat out in addition to the health problems presented by eating out, it is of utmost importance that there are sanitary standards for the restaurants. Therefore there are federal guidelines that all jurisdictions have to implement in order to ensure a certain baseline health safety standard.

One of the core functions of local health authorities is regular inspection of restaurants for food safety, which is guided by the United States Food and Drug Administration (FDA) Food Code. Even though the sanitation codes in all states are modeled after the FDA code, each state has its own implementation methods. In NYC, the DOHMH is charged with the responsibility of inspecting restaurants, bars, coffee shops, employee or university cafeterias, nightclubs, fixed site food stands, and bakeries (hereafter, restaurants). The manner in which the DOHMH does its inspections to measure the restaurant’s compliance with the NYC Health code is by using a scoring system (where the lower the score the better the establishment is at adhering to food safety) which is updated regularly in order to maintain consistency with both the New York State Sanitary Code as well as the FDA Food Code. The DOHMH conducts approximately 24,000 restaurant inspections a year via unannounced visits to monitor compliance with both NYC and New York State food safety regulations and assigns a score based on how well it follows these standards. Official DOHMH inspectors check for “food handling, food temperature, personal hygiene, facility and equipment maintenance and vermin control. Each violation earns 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” (DOHMH). Before July 2010, these results were only posted online. However, this was not enough for two main reasons. The first is that those who are the most vulnerable to foodborne illnesses, especially the elderly, are not likely to search online before entering a restaurant. The second reason is that only having the scores available online shifts information from ‘point-of-purchase’ (where the consumer may enter a restaurant), thereby limiting informed choice and a lacking transparency in the system’s key leverage point. Therefore in July 2010 the DOHMH required restaurants to publicly display the letter grade in a visible window location that reflects the most recent health inspection conducted by the DOHMH. The letter grade is a function of the restaurant score. Scores in the range of 0-13 map to an A, scores in the range of 14-27 map to a B, and scores greater than or equal to 28 map to a C. The idea behind this policy change is that the consumer has easier access to inspection results, at the point of making a decision as to whether or not to eat at the location, encouraging restaurants to better comply with food safety rules to have a higher rating and thus hopefully attract more patronage. And this policy change appeared to have a significant impact since 88% of NYC respondents factored grades into their dining decisions, and Salmonella cases decreased by 5.3% in NYC versus the rest of New York state from 2011 to 2015 (right after the adoptions of the new policy). The DOHMH collects this dataset to grade restaurants based on their health standards, which is intended for use by the average consumer to know, in part, the amount of rodent infestation and infestation occurrence in the specific establishment, knowledge which previous to this new program was unavailable to the average consumer.

Restaurants play an important role in the health of New Yorkers because of the frequency that they eat out and the health problems that come along with eating out. NYC therefore conducts inspections of restaurants in order to ensure that they maintain certain sanitary standards. In July 2010 NYC mandated that there be a public facing, easy to read letter grade to both inform diners about the sanitary condition of the restaurant and to incentive restaurants to perform better on these inspections, thereby improving the sanitary conditions of the establishment.

### Data

#### Restaurant

This study uses the NYC DOHMH New York City Restaurant Inspection Results dataset which contains every current violation citation from all program inspections conducted up to three years prior to the most recent inspection for restaurants in an active status. As of March 11, 2023 there were 208,225 inspections. Of these, 39,291 occurred in 2019 with all other inspections being removed from the dataset. The reason for this decision is that 2019 was the last full pre Covid-19 era year and this study wants to avoid any Covid-19 fallout from the study. Of the 39,291 inspections, 722 did not have a valid zip code and were removed from the sample leaving 38,569 inspections. Of the 38,569 inspections, 1,197 did not have a score associated with it. This can be for a variety of reasons, such as a data entry error or the score is still pending. These 1,197 inspections were removed from the sample leaving 37,399 inspections which is more than 95% of the total number of restaurant inspections that occurred in 2019. These inspections were then grouped by their zip code with the average score of each restaurant taken by zip code, leaving a dataset with 206 observations having a zip code to average restaurant health score in that zip code association.

#### Rodent

This study uses the NYC DOHMH Rodent Inspection dataset which contains information on both rat inspections and intervention visits in NYC. It is important to note about this dataset that if a location does not appear in the dataset does not indicate a lack of rodents, rather that it has not been inspected. Related to this is that an area with an increased number of active rodent sightings does not necessarily have a higher number of rodents, but rather that area has had more inspections. As of March 12, 2023 there were 2,327,279 reported rodent sightings in the dataset of which 248,357 occurred in 2019. Of these 248,357 reported rodent sightings, 1,673 had an invalid zip code and were removed from the sample, leaving 246,684 reported rodent sightings. Of these 246,684 reported rodent sightings, only 33,652 resulted in “Rat Activity.” This is not surprising as there are a variety of reasons for an inspection, including follow up, as well a tendency for people to enjoy complaining about anything to get attention. We therefore only included these reported rodent sightings that resulted in “Rat Activity” into our sample frame. These observations were then grouped by zip code resulting in 168 observations. It is worth noting that even though a zip code may not be included in this dataset, we can easily infer that it had zero reported rodent sightings resulting in “Rat Activity” for otherwise it would have been included in this dataset.

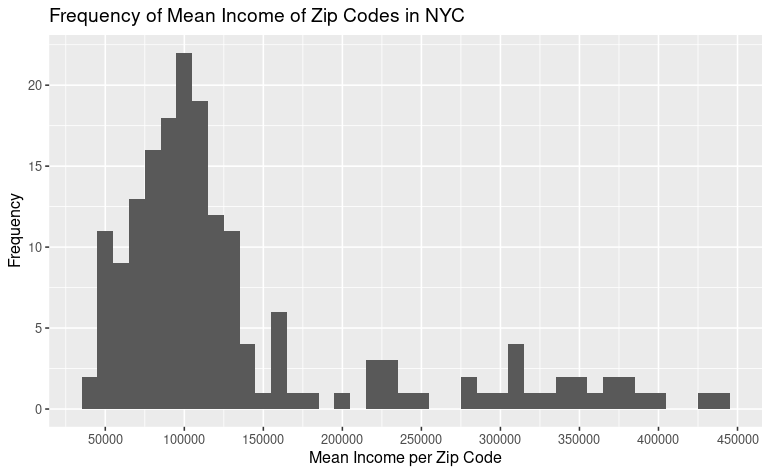
#### Income Data

The income data came from the United States Census Bureau and was accessed on February 24, 2023 and includes all zip codes for NYC where there is a non-zero population. Of the 223 NYC zip codes, 33 did not have a valid average income and were removed from the dataset leaving 190 observations. It is worth noting that there exists zip codes where there are restaurant(s) but have a zero population (such as JFK International Airport) as well as zip codes with a non-zero population but don’t have any restaurants. This dataset was then merged with the restaurant dataset resulting in a average income per zip code to average restaurant inspection score by zip code association.

### Descriptive Statistics

#### Income Dataset

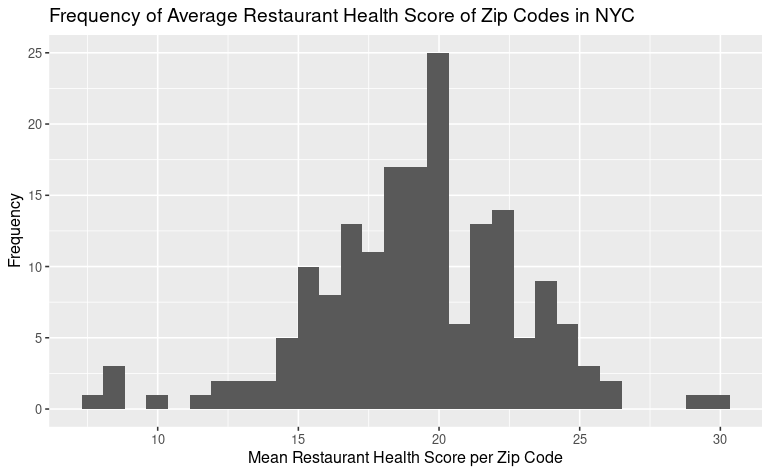
What is the distribution of mean incomes of zip codes in NYC? The mean income of zip codes in NYC is $135,672 with a standard deviation of $91,564 and a median of $103,872. Additionally the maximum average income in a zip code is $473,210 and the minimum average income in a zip code is $44,362. This distribution is illustrated below:

Figure 1: Frequency of Man Income of Zip Codes in NYC in 2019

We can see that the average income per zip code is right skewed, which matches our intuition; there are a handful of zip codes that require it’s residents to have exorbitant incomes with very few that allow for an income below the poverty line.

#### Restaurant Dataset

What is the distribution of average restaurant health score by zip code in NYC? The mean score is 19.24 points with a standard deviation of 3.67 points and a median of 19.4 points. Additionally, the lowest (best) average restaurant health score for a zip code was 7.5 points and the highest (worst) average restaurant health score for a zip code was 29.77 points. This distribution is illustrated below:

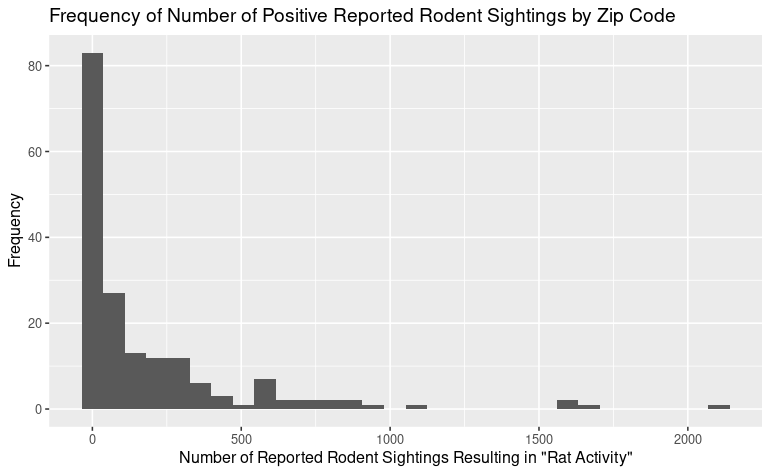
Figure 2: Frequency of Mean Restaurant Health Scoresof Zip Codes in NYC in 2019

We can see that the average restaurant health score is close to normally distributed which definitely surprises the author. We would have expected that there would have been more zip codes that have an average of A level (0-13 points) restaurants given how important restaurant health safety is.

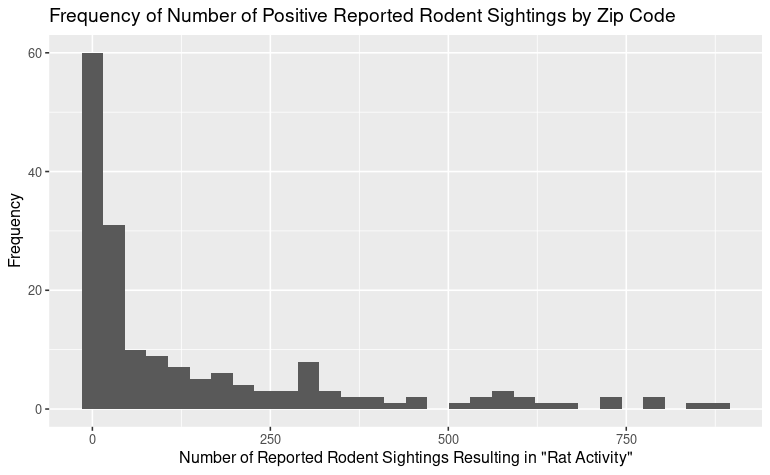
#### Rodent Dataset

What is the distribution of the total number of reported rodent sightings by zip code in NYC?

The mean number of reported rodent sightings was 187.90 with a median of 44.5 and a standard deviation of 322.53. Additionally the smallest number of reported rodent sightings in a zip code was 0 and the most was 2,104. This distribution is illustrated below:

Figure 3: Frequency of Number of Reported Rodent Sightings Resulting in “Rat Activity” by Zip Code

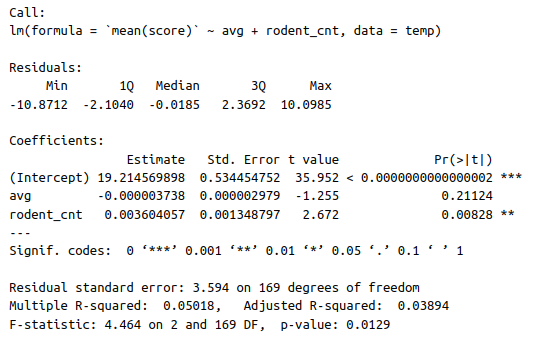
Looking at this distribution, we can clearly see that the data is right skewed and contains a handful of far out outliers. Tukey's fences, defined as [Q1-k(Q3-Q1), Q3+(Q3-Q1)] and k=3, we removed 6 zip codes from this dataset, leaving 172 zip codes. The distribution of the new dataset has the following distribution:

Figure 4: Frequency of Number of Reported Rodent Sightings Resulting in “Rat Activity” by Zip Code After Removing Outliers

The new rodent dataset is still right skewed, but much less so with a mean of 142.77 with a standard deviation of 203.87 and a median of 42.

### Model

The objective of this study is to determine whether or not there exists a correlation between average income per zip code and number of reported rodent sightings resulting in “Rat Activity” versus average restaurant health score by zip code in NYC. Given that the three metrics that will be used to measure these attributes are all continuous scales, we will use a scatter plot and a Pearson’s r to test this question.

Figure 5: Results of Linear regression of average income by zip code and number of positive reported rodent sightings in a zip code vs. average restaurant health score by Zip Code

This study did find evidence for a correlation between the average income of a zip code and number of positive reported rodent sightings versus average restaurant health score in that zip code in NYC with a p-value of 0.0129, a probability below our 0.05 cutoff. In other words, given our assumption that there is no correlation between the average income of a zip code and average restaurant health score in that zip code in NYC, we found evidence to the contrary and therefore reject our assumption. Therefore this study finds that knowing the average income of a zip code and number of positive reported rodent sightings does provide insight into the average health score of restaurants in that zip code. However, given that the coefficients are near 0, there is no meaningful signifcance to our findings. In other words, even though statistically there is evidence that a correlation exists, in practice there is no meaningful correlation.

### Conclusions

This analysis sought to determine whether there exists a correlation between the average income of a zip code and number of positive reported rodent sightings in a zip code against average restaurant health score in that zip code in NYC. The data did provide evidence that such a correlation exists, albeit not meaningful. However, there are several limitations to the analysis. One of the of limitations is with regards to the restaurant dataset. A limitations associated with this dataset is that the health score is based on a single metric that encompasses all aspects of health safety. Health safety from rodents is only one of a multitude of attributes that are considered when giving a restaurant a health score. Another limitation of this dataset is that the health score is given by a single individual and calculated based on a single visit. There are many variables and aspects that may or may not arise when the inspector is present that can greatly impact the health score.

One of the limitations is that the rodent dataset is based on self reported data. This is a limitation since even if a zip code doesn’t have any reported rodent sightings, it does not necessarily mean that there are no rodents in that zip code. Similarly, zip codes with a higher number of reported sightings do not necessarily correlate with a higher rodent population, but rather just a larger number of reported sightings. Another limitation to this study is that it focused solely on reported rodent sightings, rather than looking for other causes that might lead to a worse health score.

Additionally, the economic status of an area can be indicative of a restaurant's cleanliness because wealthier areas tend to have higher expectations for cleanliness and hygiene. This is because people in these areas generally have higher disposable incomes and are more likely to dine out regularly. As a result, restaurants in these areas need to maintain high levels of cleanliness to meet the expectations of their customers and stay in business. However, this analysis indicates that the economic status of an area is not a reliable indicator of a restaurant's cleanliness, at least in NYC. There are many clean and well-maintained restaurants in lower-income areas, and there may be dirty and unsanitary restaurants in wealthier areas as well. Ultimately, the cleanliness of a restaurant depends on the specific management and practices of that establishment.

The question of this analysis was, does there exists a correlation between the average income of a zip code and average restaurant health score in that zip code in NYC? The average income of a zip code can be indicative of the general health of the area, as higher earners can afford to maintain a healthier living environment. Knowing where these area are can help identify which areas have cleaner and more sanitary restaurants, especially given how much New Yorkers eat outside their homes. And given how much New Yorkers eat outside their homes, NYC conducts restaurant health inspections to enforce a baseline sanitary environment that is vital to the general health of the city. The hypothesis of whether the average income of a zip code is correlated with the average restaurant health score in that zip code was tested by employing a simple linear regression. The null hypothesis was accepted, indicating that no correlation was found. This is surprising given anecdotal evidence of the dirtiness of lower income areas.