Project 3

Focus: Originally my entire spreadsheet has records for every restaurant of every cuisine in New York City but I decided to focus my attention on Chinese restaurants, specifically their health ratings and how they can be improved if necessary. I further decided to focus a bit on Chinatown, Manhattan since I feel it should be representative of Chinese food as I spent my childhood taking trips their on the weekend.

Steps Taken

Excel

* The initial data sheet had restaurants of all cuisines for all of New York City but I shaved it down beforehand to Chinese cuisine only as the focus for my project, and to make it load faster in excel
* Very little text cleaning was needed to be done besides a few characters in the wrong encoding
* Attempted to standardize streets but four variations per streets across the entirety of NYC proved futile, zipcode served the same purpose
* Removed several columns that served no purpose for the analysis (building, phone number, cuisine, record date) but left them in the raw just in case
* Although there appears to be large amounts of null values in grade scoring, after some research I found that blank values meant that the inspection was failed and thus a grade was not recorded.
  + After failing the first inspection, a restaurant is given a second chance at random date at least one month later to make up for it.
  + I made a new column for these grade reinspections and gave grades for the initial inspections that were ungraded as to better have an idea of the distributions of grades
  + Grade Dates are ignored because they were left blank on purpose by inspectors
* An inspection year column was created to make filtering things easier
* There was a peculiar method for health inspectors to record multiple violations by listing the same inspection record multiple times with a different violation, thus it was necessary to have access to SQL’s ability to distinctly aggregate data
* After cleaning my sheet and many struggles, I successfully imported my table into SQL

SQL

* In SQL my first goal was to count up the letter grades and convert them into percentages so I could figure out their ratios. I started with figuring out how many restaurants were in each borough and to figure out exactly how many new restaurants were built in 2016.
* By figuring out the distinct amount of grades received by restaurants, I could make a much more correct geomaps in Tableau
  + The first one is the average scores received in each zipcode
  + The second one is the percentage of restaurants that were able to receive A’s that year in each zipcode
* Found the distribution of A, B and C by borough each year
* After feeling satisfied with my grade aggregating, I moved onto violations
* I found the average amount of violations per inspection by restaurants, and then the average amount of violations per inspections of restaurants in Chinatown
* I then split them into critical violations and vermin related violations

Tableau

* I imported many of my queries straight to SQL with the intent of aggregating them there instead of in SQL, thus leaving some of my queries incomplete
* Successfully constructed a pair of geomaps that show the average scores and ratios of As to Bs and Cs
* Made a skinny bar graph to show the growth of restaurants in each borough by the year and another bar graph that shows the changing ratios of grades by boro each year
* Made a histogram of violations in Chinatown but can’t figure out how to close the bar gaps

Initial Findings and Thoughts

* Despite the rapid growth of Chinese Restaurants, health grades on a whole have actually and the amount of restaurants that receive As have increased 10% across the board by 2016
* Most violations are quite minor and deal with sanitation or food storage
* Chinatown Manhattan has seen very little improvement however
  + Chinatown’s stagnation can be attributed to several groups of restaurants that score extremely poorly year after year but somehow stay open despite critical violations such as live cockroaches and rodents. Bribery?
  + Chinatown also has the most violations for cockroaches and rodents, would recommend hiring exterminators or buying industrial sized packs of glue traps
* Would recommend you go to Chinatown, Flushing for the time being which has seen the same positive growth of other Chinese restaurants.
* Outlook mostly positive, will continue to observe as the second half of the 2016 reports come in