Group 3: Natalia Russu, Justin Brander and Harvey Zhang

NYC Restaurants - YELP - and Inspections Data - Analysis

1) Introduction and the Business Problem:

New York City is a metropola known for having some of the best dishes in the world. Here you can find people of diverse nationalities and backgrounds, therefore everyone has specific food preferences and NYC is offering plenty of food options for everyone. You can basically find any type of cuisine you wish for, offered in small/ family type establishments or even street cars, or up to Michelin type restaurants. So, NYC offers plenty of opportunity for having a successful and flourishing business in the hospitality industry as long as their owners follow the legal and food safety requirements.

Meanwhile, a critical part of a restaurant's success is to have and drive customers into the business, to keep the loyal clientele happy and constantly acquiring new potential clients, while also competing with the market. Due to the evolution of technology, people now can make you famous or vice versa with a few posts over social media platforms or other online tools, like various reviewing apps. Yelp, is an online directory that connects people and businesses, the latter are listed there and can be found by customers who respectively have the possibility of posting reviews on businesses. The reviews, given that they may be negative, positive or neutral, perhaps, can impact the businesses accordingly. A business can be rated from 1 to 5, and also people get to write their opinions and even share pictures. A higher rating is better, it will likely motivate more people to choose a business with a higher rating as opposed to a lower rated one.

The project's proposed **problem** is to analyze and determine whether there is a correlation between the ratings of NYC's restaurants and inspection grade the restaurants were given by the city. A sanitary grade of A is what restaurants want, while a "B" or "C" will indicate that there are serious violations that could potentially be harmful for one's health. To be noted that these grades must be always posted in the front window, so it's clearly visible and easily accessible to the public.

So, the **goal** of our analysis is to dive into NYC restaurant Yelp data and inspection data in order to determine if any aspects of a restaurant, including the reviews on Yelp, can be influenced by the inspection grade; to see if there is any correlation at all between the reviews and the inspection grade assigned by the NYC authority .

This analysis could serve those who are interested in opening a restaurant or simply in joining the restaurant world in NYC, as it can help them identify certain aspects (i.e. cuisine type, location, food price etc.) that can influence the Yelp ratings/reviews and inspection grades, which in turn may impact the potential success or failure of a restaurant.

2) Business Impact & Persona

For this project we will consider the budget aspect and the clients' desires/expectations.

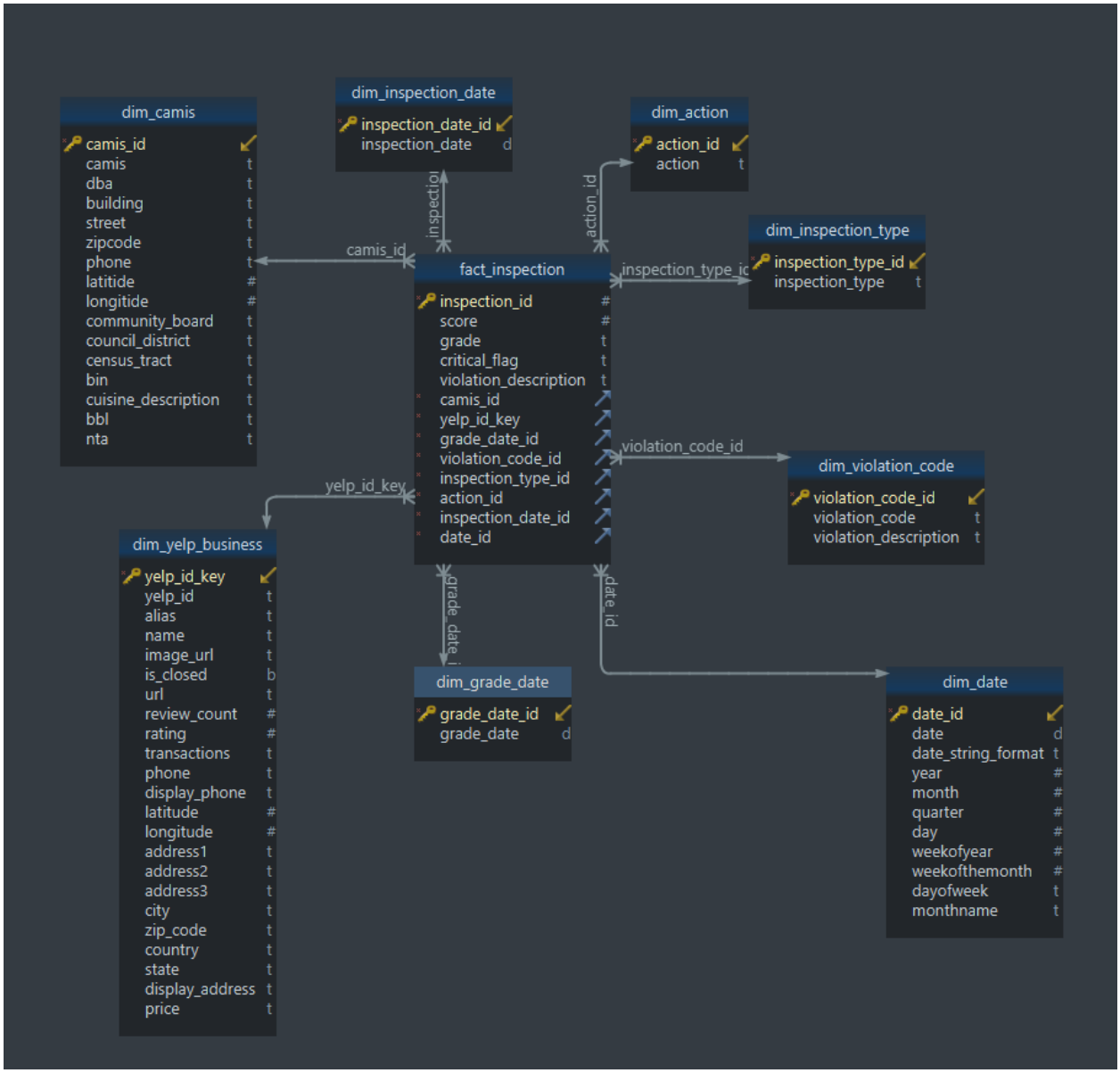
The ideal customers for consuming this data would be NYC residents looking for a consolidated view of Yelp reviews and inspection data to decide where to get food, also potential restaurant owners/ managers.

3) Data

Can be accessed through our Github:

https://github.com/NataliaRussuu/CIS4400spring2023Group3

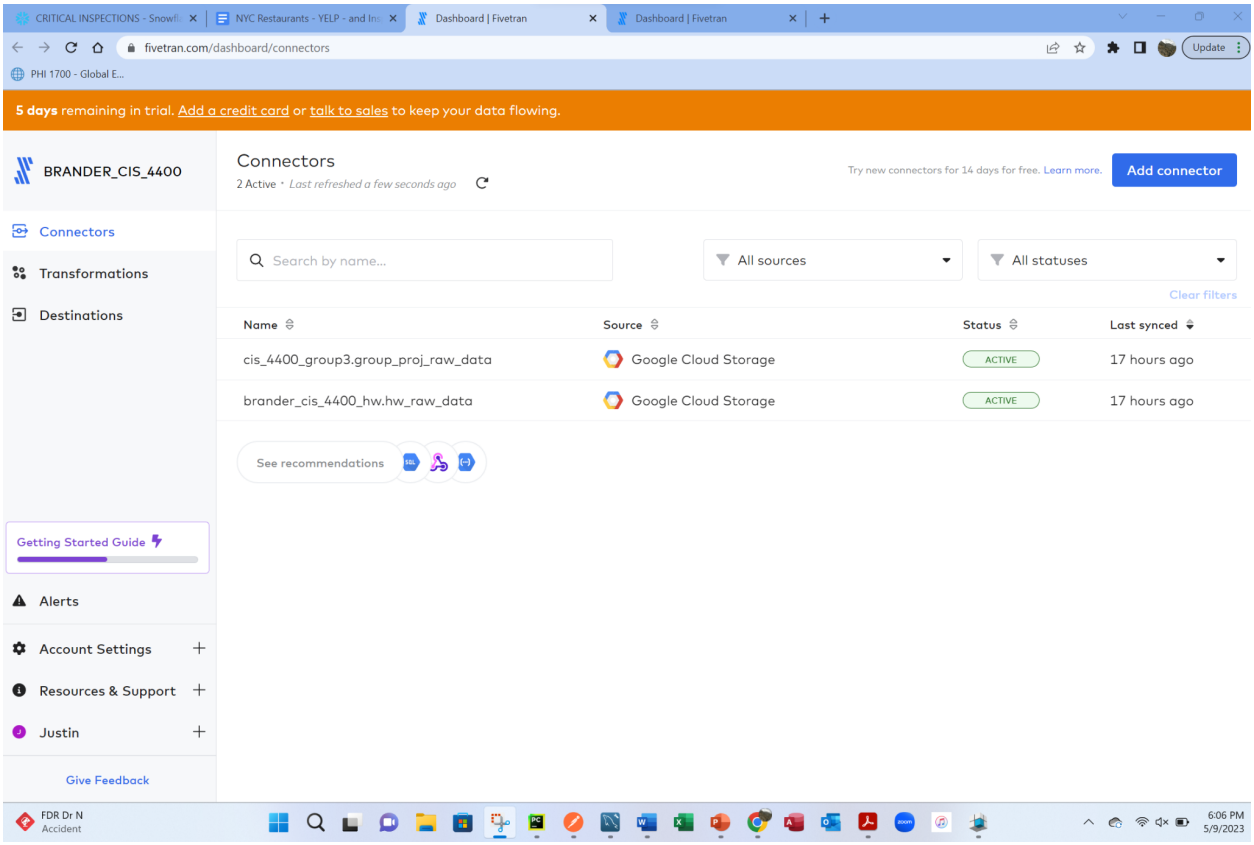
Dimensional model:



Our data comes from the Yelp API and the NYC Open Data Source (.csv)

4) Methods

Fivetran - ELT , https://fivetran.com/dashboard/connectors



Snowflake - https://app.snowflake.com/us-east4.gcp/re42164/w2SImtkMWRIX#query

5) Data Tools

Data Storage: we used **google cloud** to store data;

Data Processing: through a python script, fivetran;

Data Orchestration: fivetran/ Snowflake;

6) Interface

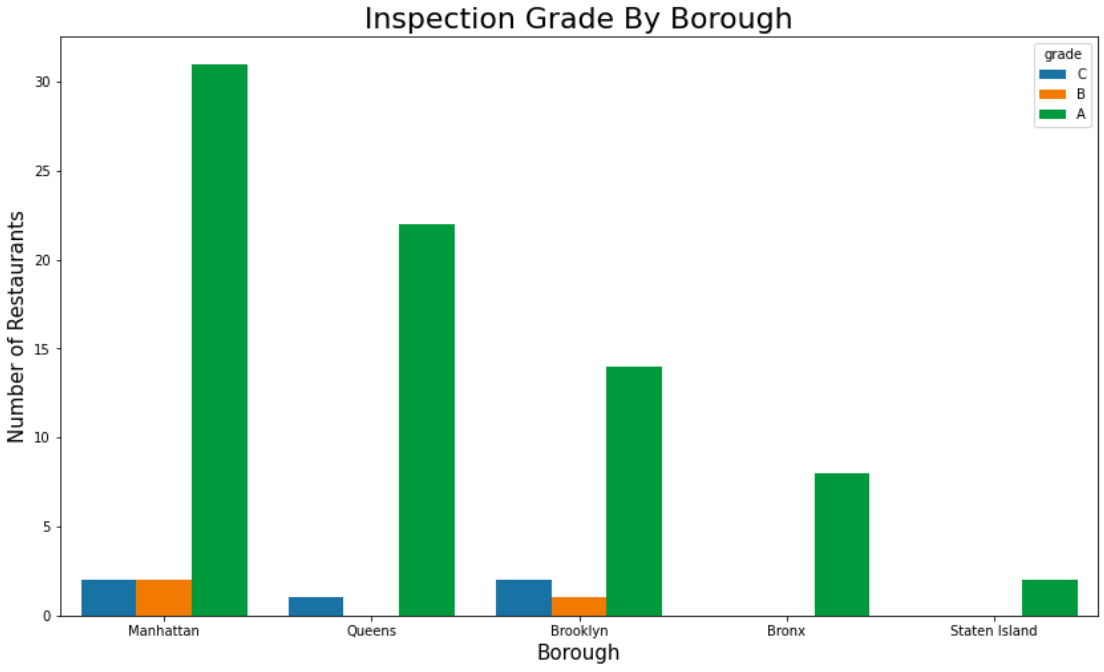
We used a few tools to visualize the data, including Public Tableau, the links can be accessed below.

Data Visualization

Fig. 1

```
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(14,8))
df_insp_filtered_new=df_insp_filtered_new.sort_values("grade", ascending=False)
sns.countplot(data=df_insp_filtered_new, x='boro',hue='grade')
#df_insp_filtered_new=df_insp_filtered_new.sort_values("A")
plt.title('Inspection Grade By Borough', fontsize=22)
plt.xlabel('Borough', fontsize=15)
plt.ylabel('Number of Restaurants', fontsize=15);
```



Manhattan has the most A, followed by Queens that has 2nd most A, then Brooklyn follows. It makes sense since our data set contains more entries in Manhattan, then Queens and then Brooklyn (see the EDA step in Github, Data Analysis).

Although most of the restaurants have not been yet flagged into any category, again we see that Manhattan has been flagged critical the most(a bit over 50), followed by Brooklyn and then Queens. Staten Island is on the lower tail - has none “critical” applications.

Fig. 3 (Created in Jupyter Notebook)

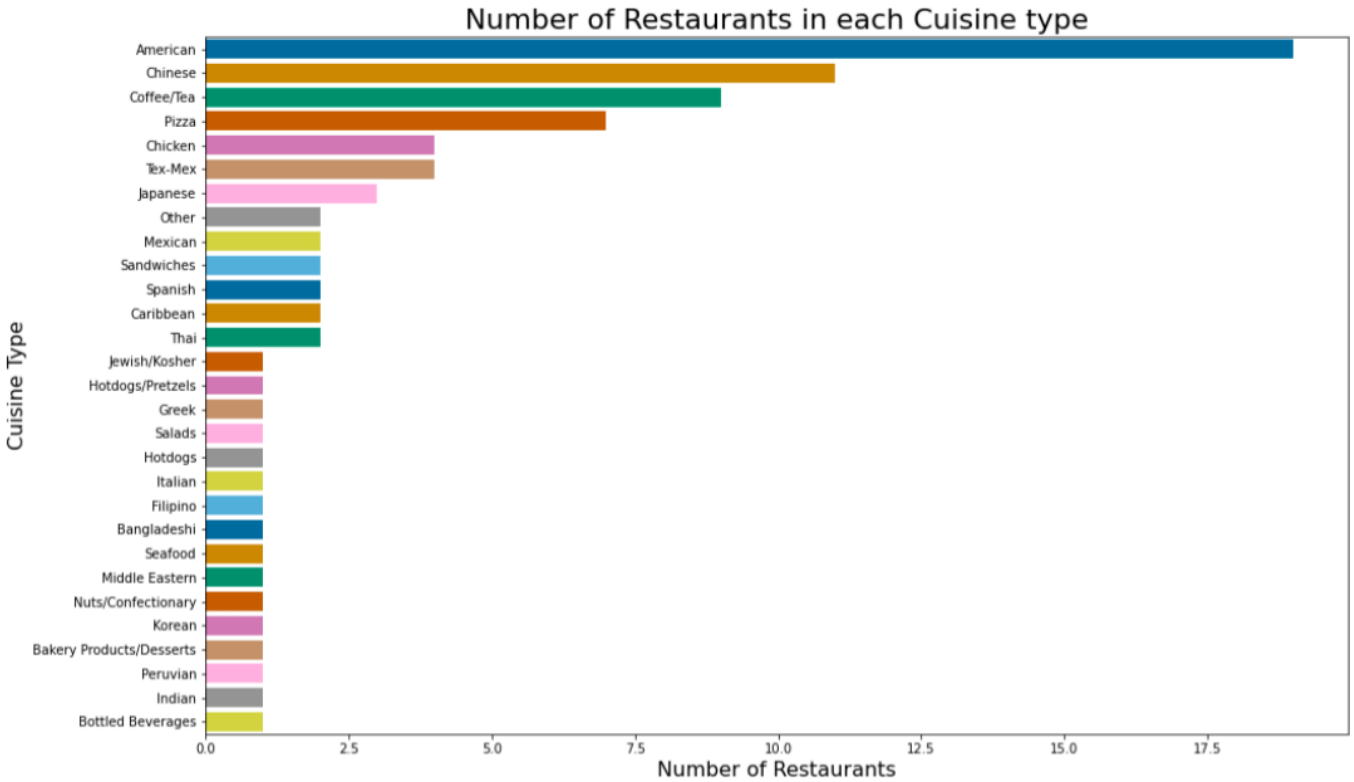
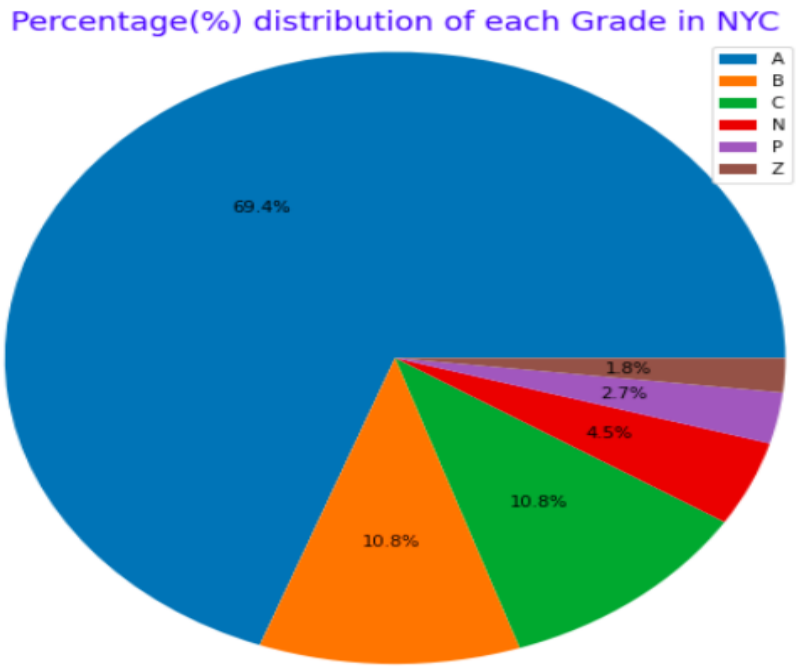


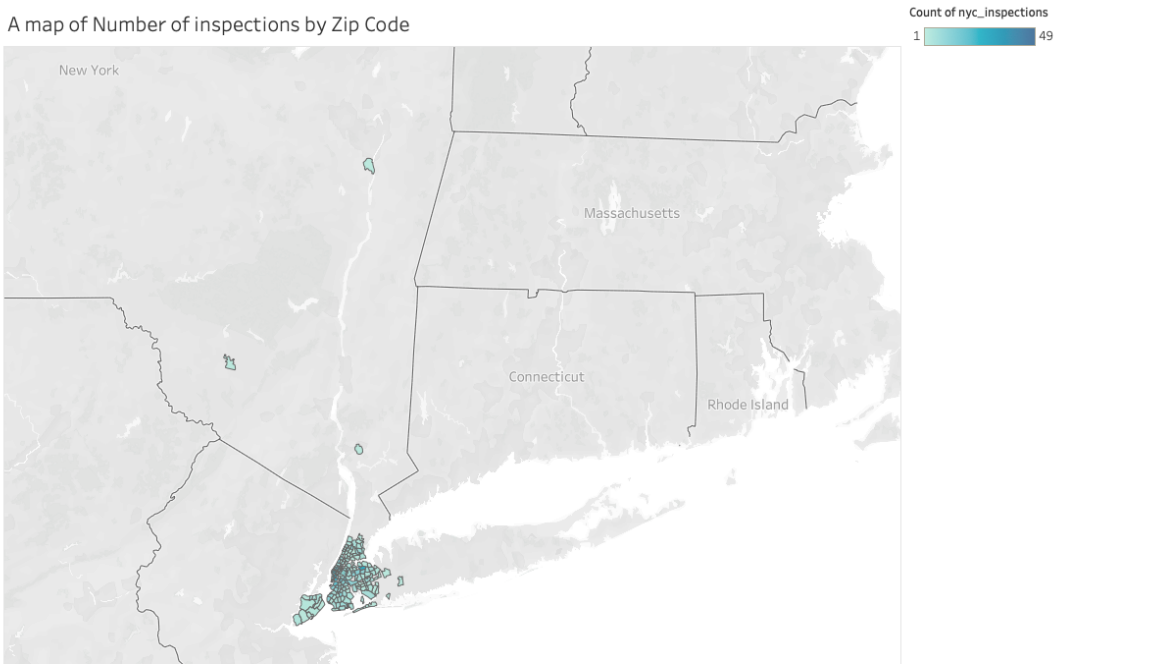
Fig. 3 depicts the distribution of cuisine kinds across NYC. American, no wonder, is the most popular one, followed by Chinese and Pizza establishments, NYC is indeed known for its pizza. On the lower end is Peruvian and Indian food.

Fig. 4 (Created in Jupyter Notebook)



Above is the distribution of the grades given by city's authorities to restaurants, as we see, there is a good pattern: most of the grades are inspiring: Grade A dominates with 69.4%, B and C share the same stake: 10.8%, though no restaurants want C or B, it will scare the customers, there are no grades given to restaurants below C.
An A grade is a score of 0-13, B grade is a score of 14-27, and C grade is 28+.

Fig. 5
Link to Tableau Public:
https://public.tableau.com/app/profile/natalia4559/viz/MapofnumberofNYCrestaurantsinspectionsbyZipCode/Dashboard2



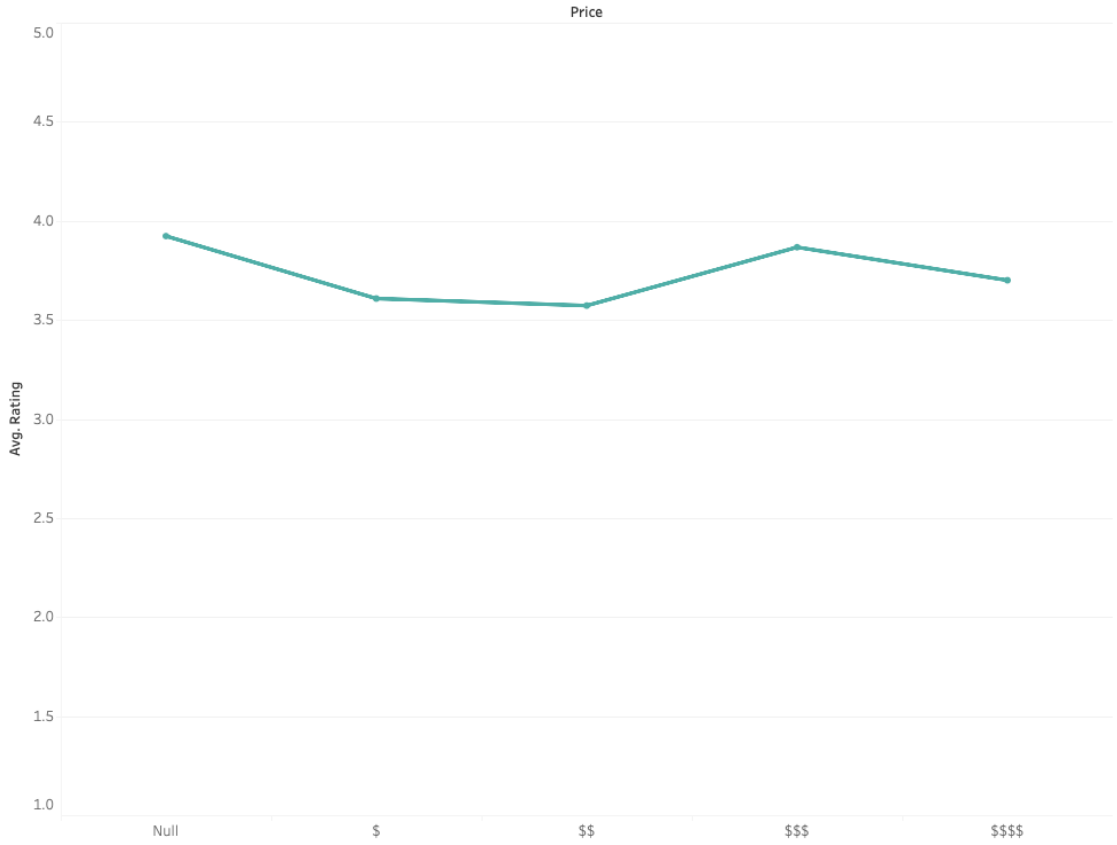
This map, which can be accessed on Tableau Public shows the distribution of all inspections done in NYC by Zip Code, again it appears that most are in Manhattan.

Fig. 6
https://public.tableau.com/app/profile/natalia4559/viz/NumberofNYCrestaurantsinspectionsbyBorough1000entries/Dashboard1?publish=yes

This graph shows an easier way of visualizing the number of inspections.

Fig. 7
https://public.tableau.com/app/profile/natalia4559/viz/AlineplotshowingtheaverageratingacrossNYCrestaurantsandtheirpricelevel/Dashboard3?publish=yes

A line plot showing the average rating across NYC restaurants and their price level



Based on the line graph above we will say that a price level of “$$$” is the one that generates or is correlated with a higher Yelp rating, while a price of “$$” shows the lowest rating, but again for “$$$$” the rating is lower - so, there is not a direct correlation between price and ratings.

Conclusions:

Is there any relationship between the inspection grades and the Yelp reviews a restaurant gets?
Our analysis shows that there is basically no correlation, so a specific grade will not influence or will influence very little the reviews.
Factors that influence the number of reviews may be considered the location, cuisine type or the price range.

Recommendations:
For the interested personas, we recommend to focus on the restaurants with a price range around 3 dollar sign, towards 4. Looks like these types of restaurants receive better ratings. Manhattan looks a strategic location, given the higher number of customers and more business opportunities.
We recommend keeping the number of violations to minimum, or to avoid them at all. Also, to try to keep an A grade, that looks to bring more reviews, thus more visibility for the restaurants, so more traffic.