PROCESS BOOK

NYC FOOD SAFETY INSPECTION REVIEW

CS 573 DATA VISUALIZATION FINAL PROJECT

CUT FAMELIA, SHAOWEI GONG, SHILE ZHAO, ZHAOJUN YANG

OVERVIEW

Food Safety Awareness

New York City (NYC) is one of the most diverse cities in the world, where many people come from many different backgrounds and cultures. Due to this fact, NYC offers many options of basic needs for the people living in the city. One of the needs is restaurants that provide different types of food based on the diversity such as American, Chinese, African, Mediterranean, and so forth.

Along with the increasing number of those thematic restaurants and the scaling up population, the food quality of the restaurants has become a serious concern for the government and the people. To address this issue, the government under the Department of Health and Mental Hygiene (DOHMH) has conducted official inspections towards all the restaurants within the city. According to NYC Health Report on Restaurant Grading in NYC at 18 Months, in July 2010 the Health Department began requiring restaurants in all five boroughs to post letter grades summarizing their sanitary inspection scores to help achieve three goals: to inform the public about a restaurant's inspection results in a simple, accessible way; to improve sanitary conditions and food safety practices in restaurants; and to reduce illnesses associated with dining out.

In fact, since January 2011, when the inspections began, the inspectors have found a big number of food safety violations in many of the restaurants. Based on the NYC Health report, New Yorkers eat food at restaurants almost a billion times every year. While most of them do not get sick, foodborne bacteria, viruses and other contaminants cause millions of cases of illness each year. It is estimated that more than 6,000 New Yorkers are hospitalized and 20,000 visit emergency rooms each year because of foodborne illnesses. Each year, NYC receives approximately 2,700 complaints about restaurant-acquired foodborne

illnesses and another 3,000 complaints about restaurant hygiene. Based on a study conducted by Baruch College Survey Research in 2012, 88% of New Yorkers take the grades into account when making dining decisions. Besides, 70% concern about getting sick from eating out in restaurants, delis and coffee shops, with 38% are very concerned.

Food Safety Grading System

According to the NYC Government Health report, the inspection program uses letter grades: A, B or C, with dual inspections to help restaurants improve their food safety practices before they are graded. If a restaurant does not get an A on its first inspection (initial inspection), the Health Department does not issue a grade but conducts a second inspection (re-inspection) around a month later and issues a grade based on the re-inspection. This series of inspections is called a cycle. Each health code violation results in points; total points are a restaurant's inspection score. The following is the score ranges for each letter grade:

A = 0 – 13 points (excellent food safety practices)

B = 14 – 27 points

C = 28 points or more

Restaurants that earn B or C on their initial inspection begin a new cycle of inspections sooner. Those earning A on their initial inspection start a new cycle in around 12 months, while those getting B or C begin a new cycle in approximately 6 months and 4 months, respectively.

MOTIVATION

Before letter grades, restaurants were motivated to practice food safety by their own desire to maintain healthful conditions and by the threat of fines for violations found at the time of inspection. From the data the inspectors have collected, in fact, not every restaurant has done improvement based on the inspection findings and recommendations. This fact is indicated by the lower grades given to a number of restaurants after the last inspection. To evaluate the impact of the inspections, it is important for DOHMH to see the trend of the grades given to each restaurant after every inspection was done, whether it is going up or going down, or fluctuating. Thus, DOHMH can easily do some analysis to get some insights for future inspections.

Furthermore, we believe that it is important for New Yorkers, especially those who like to eat outside and care about food safety to explore and learn the inspection results of each restaurant and do some analysis in an interactive way to help them in dining decision to maintain a healthy lifestyle. Also, the visualization will be useful for food safety supervisors and analysts to support their evidence-based works.

QUESTIONS

In general, this data visualization project aims to explore the NYC restaurant food safety inspection dataset to give new insights to New Yorkers to help them in food safety quality improvement in NYC. To achieve this goal, we figured out three important questions that need to be addressed:

1. How much the inspections have affected the restaurant food safety practices.
2. How the food safety grade ranking and summary of every restaurant type looks like.
3. How every restaurant food safety grade performance looks like based on different periods of time and how they differ among each other.

DATA

Source

We use a DOHMH New York City Restaurant Inspection Results dataset retrieved from NYC Open Data website (https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j). This website provides a dataset on restaurant inspections, violations, grades, and adjudication information of 439,301 records.

Structure

The structure of the data is described below:

- CAMIS: This is a unique identifier for the entity (restaurant)
- DBA:This field represents the name (doing business as) of the entity (restaurant)
- BORO:Borough in which the entity (restaurant) is located.
- BUILDING:This field represents the building number for the entity (restaurant)
- STREET:This field represents the street name at which the entity (restaurant) is located.
- ZIPCODE:Zip code as per the address of the entity (restaurant)
- PHONE :Phone Number
- CUISINE DESCRIPTION: This field describes the entity (restaurant) cuisine.
- INSPECTION DATE: This field represents the date of inspection
- ACTION: This field represents the actions that is associated with each restaurant inspection.

- VIOLATION CODE:This field represents the violation codes that is associated with each restaurant inspection.
- VIOLATION DESCRIPTION:This field is the description that corresponds to the violation codes
- CRITICAL FLAG:This indicates if Violation is critical or not.
- SCORE:Total Score for a particular inspection. If there was adjudication a judge may reduce the total points for the inspection and this field will have the update amount.
- GRADE:• N = Not Yet Graded • A = Grade A • B = Grade B • C = Grade C • Z = Grade Pending • P= Grade Pending issued on re-opening following an initial inspection that resulted in a closure
- GRADE DATE:The date when the current grade was issued to the entity (restaurant)
- RECORD DATE: The date when the extract was run to produce this data set
- INSPECTION TYPE:The type of inspection. A combination of the program and inspection type.

Attributes

The followings are the details of the data attributes along with each number of the value type:

CAMIS = 25,969

DBA = 20,506

BORO = 6

BUILDING = 7,228

STREET = 3,309

ZIPCODE = 230

CUISINE DESCRIPTION = 84

INSPECTION DATE = 1,335

ACTION = 6

VIOLATION CODE = 97

VIOLATION DESCRIPTION = 95

CRITICAL FLAG = 3

GRADE DATE = 1,256

INSPECTION TYPE = 34

Pre-Processing

The dataset takes 160 MB in memory, which is too much to fit into a web page. Besides, the dataset contains a lot of information, so we extract and re-organize the data based on what we need to address each question. Therefore, we do the following data pre-processing work:

1. Transforming some parts of the data into JSON data type to make it easier to visualize the grade summary

2. Building an index JSON for line chart

3. Compressing index JSON by using dictionary

4. Removing the restaurant which has only one record

5. Removing duplicated records

EXPLORATORY DATA ANALYSIS

To get the whole picture of what the dataset actually tells us, we firstly looked at the CSV file of the dataset and used Panda to do the calculation to see the data distribution. Here, we found that the dataset has much information that could be useful for analysis to get new insights. We then tried to figure out some important questions we wanted to answer based on the dataset and we found many interesting questions that can be addressed, which may lead to NYC food safety improvement in the future. For instance, the dataset tells the grade of each restaurant so that we can find out which restaurants have better food safety quality. This will help New Yorkers in making food decision. Besides, the dataset has "VIOLATION TYPE" column that specifies what kind of food safety rules that a restaurant breaks when being inspected. From this data, we can explore what kind of violations mostly done so that the Health Department can analyze why those violations occurred so that they can figure out what kind of actions they need to take towards the restaurants to reduce the violation rate.

However, for this final project with a limited amount of time, we would not have enough time to explore and visualize all the important data to answer all the interesting questions. Therefore, we finally decided to focus on the restaurants and the inspection performance, i.e. the grade distribution of the restaurants and the inspection impact on food safety improvement in the restaurants, which we break down into 3 questions as mentioned in "Question" section.

After collecting the questions, we deployed Panda to see the data validity and to get some calculation on the data values so that we could see how much workload we would need to do to make an interactive visualization to address the questions. We finally found that it is reasonable to implement our visualization design to address all the five questions in our list with the given length of time and the validity of the data. However, since the number of

restaurant is very huge: 20506, which is too many to be put in one visualization, we decided to shorten the number by putting the restaurants into a number of categories based on the restaurant type (cuisine description). We use this number in the restaurant population and grade summary charts only.

DESIGN EVOLUTION

Initial Design and Issues

After doing the exploratory data analysis and finding five questions to be addressed, we decided to create 5 visualizations, as described below:

Horizontal Bar Chart of Restaurant Count

Horizontal Bar Chart of A Grade Accumulation Percentage

Grouped Bar Chart of All Grade Summary

Tree Chart of Restaurant Names and Line Chart of Restaurant Score Trends

DNA Chart for Grade History

Revised Design

After considering efficiency factor and some suggestion from Prof. Lane Harrison during the proposal feedback meeting, we finally made some changes. The tree chart was transformed into a set of drop-down-selections / searching-bars which end up with a corresponding line chart of the score trend. The problem of using Tree to provide selections of restaurant is that first, there are too many children for each node of the Tree, so it is inconvenient for our users to select the restaurant that they are interested in. Also, using Tree, our user have to select their options hierarchically, which limit the flexibility of our app. However, with drop-down menu/searching-bar, we can user more flexible data structure under the hood, easier for our users to select out the restaurant that they are interested in.

According to the class feedback, the users are not willing to navigate the restaurant list especially when the drop-down is too long. According to this

observation, we added the searching bar to let user to specify the filters they want instead of looking for them in the long list and being lost.

From "tree chart" to "bar selection"

In addition, after the prototype presentation, based on some feedback from Prof. Harrison, we added a sparkline of score trend of each restaurant that pops up in the selection list for user convenience in looking at the trend comparison in brief before they start to explore the line chart trend one by one. With the help of the sparkline, our users can directly get the score trend of each restaurant that they are potentially interested in, and get a rough impression about restaurants, which they want to know more detailed information, they can click the restaurant that they are interested in, and in the left side, there is a line chart which will display the score trend with precise value and the date. In addition, we moved all the selection bars from the left-hand side to the top of the line chart to fill space gap so that we could minimize empty space size, and we added a searching bar below each selection bar to provide users an option to do faster selections.

Last version of bar selection and line chart

IMPLEMENTATION

Visualization #1 - Horizontal Bar Chart of Restaurant Count

The purpose of this horizontal bar chart is to give users an interactive overview of the distribution of each restaurant type in each borough. Each color represents each different restaurant type. When users mouse over any of the bar, the bar's color will be highlighted and tooltip showing the name of restaurant as well as the count of the restaurant will come up.

Visualization #2 - Horizontal Bar Chart of "A" Grade Summary

This chart is intended to show the percentage of the accumulative A grade of each restaurant type in NYC. The y-axis specifies each restaurant type and the x-axis demonstrates the percentage. If users mouse over a bar, the bar will be highlighted and a tooltip showing the restaurant type and the percentage of grade A will come up. We display the restaurant type and the percentage on the tooltip to make it easier for users to identify which bar belongs to which restaurant type in case for some people the x and y-axis look a little dense.

Visualization #3 - *Grouped Vertical Bar Chart of All Grade Summary*

We made the grouped vertical bar charts to show users the distribution of all letter grades achieved by each restaurant type in each borough. So, when a user clicks on a green bar from the horizontal bar chart, a grouped vertical bar chart displaying the grade summary of the corresponding restaurant type will pop up.

Visualization #4 - *Line Chart and sparklines of Restaurant Score Trends*

The bar selections are to provide users with filtering feature to choose any specific restaurant they want to explore. There are two ways that u user can do to do the filtering:

1. A user can pick a value from one or more selection bars. So it is not necessary to pick a value from each bar, it depends on needs. In general, the more number of selection bars a user uses, the less number of restaurants will come up.
2. A user can type a value they want to get in one or more search bars. If the value typed is not in the selection bar lists, then the default/current restaurant name(s) and line chart will not change and the user need to use the bar selection instead.

The "Restaurant List" box will display the restaurant name(s) that are picked by a user based on the filtering result, along with the sparkline of each restaurant next to it. The sparkline shows the score trend of every chosen restaurant from time to time. At the same time, the line chart will automatically come up, showing the default/current score trend of the first restaurant on the list with the details of score numbers on y-axis and inspection dates on x-axis. So, if a

user clicks on one of the displayed restaurant names, then the corresponding line chart will be shown on the right.

To make it quicker for users to see if the score performance of a restaurant is "better" or "worse", we added an arrow on the top-left of the line chart area. If the arrow turns green, it means that the last score is less than the first score, indicating that the restaurant food safety quality is getting better. In contrast, if the arrow turns red, it means that the last score is higher than the first score, so the restaurant food safety quality is getting worse. As described in the grading system section, the score represents the violence rate of food safety practices. Therefore, the more score a restaurant gets, the worse letter grade it will be given.

Visualization #5 - *DNA Chart of Grade History*

Once a user gets the restaurant(s) on the box based on the filtering result, the user can drag out any of them to the "Drag the Restaurant Here" DNA chart box below the line chart. Users can always drag as many restaurants as they want to compare on the grade records in each inspection interval time. Each bar in the DNA chart shows the time length between one particular inspection and the

following inspection, so the size of each is adjusted based on the time length. If users mouse over a bar, the bar will be highlighted and a tooltip showing the information details (restaurant name, grade, time length) on that particular length of time will pop up.

EVALUATION

Analysis and Insights

From the visualizations, we found some interesting insights based on our analysis:

1. The Nuts/Confectionary is the best restaurant type in terms of food safety quality, while French restaurants have the worst food safety practices.

2. As Manhattan has most of the different restaurant types, meaning that the demand of different types of cuisine is higher than those in the other boroughs. This may suggest that Manhattan is the most diverse borough/region in NYC.

3. There is a high variance in the scores trends among the restaurants since many of them have fluctuative score trends in the time length between the first inspection and the last one. Also, many of the restaurants have higher scores of food safety violation after the last inspections, but at the same time, there are also many restaurants having lower scores at the end. This means that the inspections have not been very effective to make all the restaurants improve their food safety practices. Therefore, it is important for the Health Department to evaluate their current inspection implementation and also to find out why many restaurants have not made better progress as many others have to promote better food safety quality.

Future Work

There is still a catalogue of data related to violation types we haven't touched yet. So in the future work, we are thinking, maybe we can connect what we have currently to the violation types and dates. It may be useful for the Health Department to do some analysis on which types of violation that they need to pay more attention to decrease food safety violation rate in the future.

REFERENCE

DOHMH New York City Restaurant Inspection Results dataset: https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j

NYC Health Report on Restaurant Grading in NYC at 18 Months: http://www1.nyc.gov/assets/doh/downloads/pdf/rii/restaurant-grading-18-month-report.pdf

D3 Gallery: https://github.com/d3/d3/wiki/Gallery

Stack Overflow: http://stackoverflow.com/