NYC Health vs Yelp: Analysis to Compare NYC Restaurant Rating with Diners' Yelp Reviews

Sam Falk <sjf374>, Mei Guan <yg833>, Jerome Louison <jl9489>
All in Wednesday class

Introduction:

Implemented in July 2010, New York City's Health Rating system requires restaurants to post letter grades of the respective establishment's sanitary inspection results. This policy reflects the mission of the Health Department to protect public health by assuring that restaurants practice safe food handling, preventing foodborne illnesses. The New York State Department of Health reported that "61% of contributing factors identified in foodborne illness outbreaks" can be traced to one of 4 key unsafe food practices such as: a) contaminated ingredient (21%), b) a food handler who was infected with a pathogen (19%), c) naturally occuring toxins (12%), and d) inadequate cooking (9%). This method of restaurant health ratings has become a standard for monitoring urban public health and similar grading systems can be found in cities like Boston, and San Francisco.

Problem Statement:

However, for a city as large as New York City with 32,000+ restaurants to book on OpenTable year round and the NYC Health Department reports inspecting about 24,000 restaurants annually--it leads one to doubt the accuracy of the rating system. These metrics imply that our city government may not have the capacity to inspect all New York City restaurants at least once a year. Additionally, an analysis from IQuantNY in 2014 highlighted an unusual distribution of restaurant health ratings with many restaurants receiving an "A" rating by a narrow margin of 1-2 points.

If inspections are infrequent and biased towards an "A" rating but diners could eat out year-round, how else could the city government monitor safe food handling practices? Can user generated data on social media be used to predict restaurant health violations? Our project aims to use user generated reviews on Yelp.com to supplement the health inspections ratings data with the goal of developing a model to help identify restaurants with high risk for unsafe food practices, which may lead to outbreaks of foodborne illness.

Data & Methods:

We will utilize NYC Open Data portal to identify restaurants with low health department ratings. Those ratings will be supplemented with Yelp review inputs to predict health violations, which will then be applied to high scoring restaurants to recommend restaurants for inspections. The Yelp reviews will be processed through Natural Language Processing (NLP) methods to create variables for sentiment analysis and key words related to abdominal distress and foodborne illnesses. Examples of these include "feeling sick", "hair in food", "coughing" and "undercooked"

to name a few. We will aim to build and train a number of different models and compare across for the best performance (accuracy and precision). Then we will apply this model to restaurants who may have received higher health ratings previously and may not be inspected anytime soon. Our goal is a model to help the NYC DOH prioritize their inspections, given their limited resources (i.e. number of inspectors), and to improve monitoring of potential outbreaks of foodborne illnesses.

About the Team:

Our team is comprised of all part-time graduate students at NYU's Center for Urban Science + Progress, taking the Wednesday PM session of Applied Data Science. For this project, we will each tackle different components preparing the raw data into analysis ready datasets (i.e. cleaning and aggregating). We will also each take part in creating models and will compare the performance of our models before selecting the final model for the project. In terms of project set-up, write-up and research, we will all take equal parts in that work. Work product submissions for the project will come from Mei Guan (yg833) on behalf of the team.

References:

New York State Department of Health. Understanding Contributing Factors. Jul 2019. https://www.health.ny.gov/publications/6665.pdf (accessed 19 Oct 2019)

NYC Health. How we score and grade. Mar 2012. https://www1.nyc.gov/assets/doh/downloads/pdf/rii/how-we-score-grade.pdf (accessed 19 Oct 2019)

NYC Open Data. DOHMH New York City Restaurant Inspection Results. 14 Oct. 2019. Department of Health and Mental Hygiene (DOHMH). https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j (accessed 19 Oct 2019)

San Francisco Department of Health. Food Safety Program: Restaurant Safety Scores. https://www.sfdph.org/dph/EH/Food/Score/ (accessed 19 Oct 2019)

Schomberg JP, Haimson OL, Hayes GR, Anton-Culver H (2016) Supplementing Public Health Inspection via Social Media. PLOS ONE 11(3): e0152117. https://doi.org/10.1371/journal.pone.0152117 (accessed 19 Oct 2019)

Rocheleau, Matt. Chart: Food safety grades for every restaurant in Boston. 24 Aug 2016. https://www.bostonglobe.com/metro/2016/08/15/chart-food-safety-grades-for-every-restaurant-boston/A3adU1opTucixN6cVMhmpL/story.html (accessed 19 Oct 2019)

Wellington, Ben. "I Quant NY." I Quant NY, Tumblr, 17 Feb. 2014, https://iquantny.tumblr.com/post/76928412519/think-nyc-restaurant-grading-is-flawed-heres (accessed 19 Oct 2019)