## Delivering Faster Results with Food Inspection Forecasting

Chicago's Analytics-Driven Plan to Prevent Foodborne Illness

By Sean Thornton, May 19 2015

<http://datasmart.ash.harvard.edu/news/article/delivering-faster-results-with-food-inspection-forecasting-631>

All of Chicago’s 15,000 restaurants are required to adhere to food safety regulations to avoid foodborne illness. To enforce these regulations, the City of Chicago employs 30- 35 food inspectors. With so many restaurants and so few inspectors, there are not enough resources to fully police all restaurants at all times.

The traditional method of choosing who to inspect finds about 15% of inspections result in a critical violation – usually the food temperature is in the danger zone, resulting in dramatically increased odds of foodborne illness. The City of Chicago wondered whether data and analytics could help them improve their efficiency of choosing who to inspect.

Chicago is particularly well-positioned to try a data-driven approach to reduce foodborne illness. Chicago was recently awarded $1M from the Bloomberg Philanthropies Mayor’s Challenge, in which they proposed to aggregate and analyze data to make better and faster decisions. This led to the creation of the SmartData Platform, a place to access well-structured and up-to-date data from many sources.

With well-structured data available, the city also wanted to secure a collaboration for the project. With the help of Civic Consulting Alliance, they were able to partner with a team from Allstate Insurance, a Chicago-based company with strong data science talent.

The team, including Allstate Insurance and the City of Chicago, was able to predict which establishments are at highest risk of violating food safety regulations. They found that the factors most associated with a critical food safety violations were: a history of violations; having a tobacco or alcohol consumption license; length of time since the last inspection; location in the city; nearby garbage and sanitation complaints; nearby burglaries; and 3-day average high temperature.

Being able to predict the critical violations allows inspectors to prioritize those establishments and to catch violations earlier, keeping the city safer. In a trial month, the standard inspection approach was able to catch 141 violations; using the data-driven approach, the team estimates that the inspectors would have been able to catch 178 violations, a 26% improvement.

The City of Chicago has released its data and code for this project with the hopes that other cities can implement similar procedures or help them improve their own. However, most other cities would struggle to aggregate the data from many sources since they do not have a program similar to the SmartData Platform.

## Food Inspection Forecasting

Optimizing Inspections with Analytics

<https://chicago.github.io/food-inspections-evaluation/>

The Chicago Department of Public Health collected almost 100,000 sanitation inspection reports and gathered data from other city departments based on 311 reports concerning pest sightings, business characteristics, and other information. The project resulted in restaurants with critical violations being found on average 7 days sooner during a two month evaluation.

Critical violations, typically pertaining to improper temperature control for food, facilitate the spread of food-borne diseases and their detection is the primary goal of food inspections. Only one critical violation is sufficient to categorize a restaurant as “failing” the inspection. Roughly 15% of inspections result in such a violation, and the longer an establishment operates with unsafe conditions the greater the risk posed to the general public. This provided the impetus to partner with the Department of Innovation and Technology and Allstate Insurance to combine datasets in an effort to expedite detection of critical violations.

One dozen variables were determined to have significant relationships to the likelihood of an establishment failing an inspection. These include prior violations, above average temperatures for previous days, nearby sanitation complaints, nearby burglaries, possession of a tobacco or alcohol consumption license, and which inspector is assigned.

Every variable used in the model is available on Chicago’s open data portal that exists to improve transparency for city government, allow developers to make apps to aid residents, and to increase commercial activity for small and large businesses. The portal aided the city’s dissemination of a large volume of data and proved to be an effective tool for allowing collaborative research.

When testing the model, a probability of failure was assigned to each restaurant. The researchers then estimated the number of violations that would have been caught if the order of inspections was rearranged to prioritize visiting the higher risk institutions first. Using this data-driven workflow, nearly 70% of critical violations would have been found in the first month compared to 55% using the standard workflow. This translates to 37 establishments being discovered sooner rather than later, potentially saving patrons from falling ill.

The forecasting model is available as an open-source project. All coding was done in R, which is also open source and a widely-known language for statistical analysis. Others are encouraged to reproduce and improve upon the project, and all necessary materials are available on GitHub.