**Chicago Food Inspections Dataset**

Department of Information Systems, California State University Los Angeles

Marco Aramburo, Henry Flores, Antonio Ponce, Michael Mijos

CIS 3200: Data Analytics

**Abstract:**

Our work was about the Chicago Food Inspections between the years of 2010 and 2019. These inspections take into account food markets, restaurants, and every established food distributor in the entire city of Chicago, IL. All inspections were completed through the Chicago Department of Public Health’s Food Protection Program. We chose this topic because it related to the socio-economic status and health of the city of Chicago and the food distributors within it. How these food distributors are affecting the health of millions of people is important to know as it shows how well the city is taking care of the health of its population and making sure the food is being properly maintained and distributed.

This dataset gives detailed data on the type of violation of violators, the risk of violation depending on the type of food that is being handled, and whether the food distributors passed or not. We believe it is important for food to be properly handled and maintained in the city and the city creating a healthy environment for the people. This data is useful as it can be applied to other cities helping bring attention to the importance of food being properly handled.

**Introduction:**

**Why we chose this topic and its importance:**

As briefly discussed in the Abstract, we chose this topic because we believed it was served as valuable data to be transformed into information. This information can be used to help other cities, anywhere, to discover the importance of keeping a healthy and stable food distributing environment for their citizens. This data also gives a good representation of what kind of food is most present in Chicago, showing the favorite type of food and what is most enjoyed. On top of that, showing the economic status of the population in Chicago. If there are a lot of fast food restaurants, sit down restaurants, and if there are a lot of markets or schools.

This topic is also very important as it shows the risks that these facilities face for being a simple food distributor. Especially if the type of food they are handling is sensitive. The importance of this topic also relates to the results of these inspections and how many food distributors have made violations or make bad scores. Choosing this topic has also made us research more into it and gave us a more in-depth analysis of the importance of inspecting in general. There are very big gaps of health hazards between a score of A and a score of B. B scores already contain rodents and such.

The importance of this subject comes from the prevention of any diseases or viruses being spread out into the public. This is surely something that must be controlled and maintained authority over. This is why inspections are so important when it comes to these food distributors, making sure the health of the people buying it is considered.

There are many variables when it comes to food maintenance and keeping it in the proper conditions to make sure it does not go bad. Expiration dates must be considered by the food distributors and make sure to throw out anything that may be expired or else that may lead to people complaining if they decide to cook it and sell in case they try to get away with not throwing out expired food and wasting money.

Overall, the importance of this topic is making sure that society as a whole maintains a healthy and happy demeanor when it comes to eating out. It is important to make sure that all the food being distributed is edible and do not carry any infections bacteria. Food is very sensitive and that is why it must be treated with the proper care. City inspectors make sure of that.

**Background of our work:**

The background of our work is given through the Chicago Department of Public Health’s Food Inspection Program. They created this dataset and update it often. It is open to the public to download and analyze. For this reason, there are many other people or institutions who have taken this data and done some analysis of their own which we will bring up and show later in the paper. The city of Chicago stated, “We work to ensure that all food preparation in every restaurant, grocery store and food truck across the city is done in a way that meets the health code” when they decided to release this inspection data and become more aware of the food distributors and the health of the population in the city.

As there are so many ways of visualizing and analyzing data, we decided to use the tools we learned in this course and use Microsoft Azure ML. Through this machine learning platform, we plan to break down the dataset, clean up missing values, and create visualizations for the areas that were inspected, the risks many of these restaurants face, and how many of the these food distributors violated the food safety practice laws.

**Related Work:**

The first only paper we came across that analyzed the same data was the city of Chicago who worked in conjunction with the county. Both gathered each of their own restaurant inspections conducted in their respective area. In total they collected over 100,000 sanitation inspections. Both of the departments collaborated together in order to achieve the same goal to uncover any foodborne illness before they reach and infect the public.

The city and the county made the data accessible to every to analyze and review on their own within two portals. The first being the Chicago Data Portal where you can access several data sets from the city. Another way they shared their datasets was through GitHub where they provided their code and also all data, reports, etc.

One of the main differences in their study came down to using a different tool than what we used. They used R studio which is a free open source code software but similar to our project we didn’t use an expensive software licenses to run and visualize our data. The ability to run it on an open source code software is being able to share the code with everyone in order to build of it.

In their analysis of the data sets they were aiming to predict whether or not the retail food establishment will end up getting a critical violation during an inspection. In their analysis they were able to identify and distinguish three levels of violations the first being critical violations which basically meant if the issues would pose immediate health hazards, the second serious violations issues that represent potential health hazards, and lastly would be minor violations which do not pose any immediate threat to the community.

They were able to achieve these three criteria was through several predicting variables that provided them with the likelihood of a restaurant receive a violation. The criteria included: prior history of critical violation, possession of a tobacco and/or incidental alcohol consumption license, length of time establishment has been operating, length of time since last inspection, location of establishment, nearby garbage and sanitation complaints, nearby burglaries, and lastly three-day average high temperature.

Their use of the data was to be able to visit those restaurants who had the potential to have repeated critical violations in the course month first before they would see others who had a lower chance. This would avoid the chance of any person getting sick from the restaurants malpractice and further reducing the amount of critical violation they would catch on a monthly basis.

Our data analysis didn’t differentiate between which violations would be the most critical to least critical but many focused on ensuring we were able to map and take account of all violations within the city and county.

In our study we wanted to achieve a similar method where we were able to duplicate their study and provide a visualization where we could be able to distinguish which restaurants would have repeating critical violations. In our study we wanted to use AzureML studio to be able to get to the same conclusion as the city and county of Chicago.

**Background/Existing Work:**

Our data was retrieved from the City of Chicago Open Data on kaggle.com. It consists of daily updated datasets from inspections performed by staff from the Chicago Health Department of Public Health’s Food Protection Program. The city of Chicago has an open data platform and its information is updated according to the amount of data brought in.

The Food Protection Services is a business that prioritizes prevention of the spread of food-borne diseases. They do this by recalling food, inspecting food businesses, and responding to complaints. Chicago is home to 16,000 food establishment including restaurants, grocery stores, bakeries, lunchrooms, food vendors, and any other operations that store, prepare, package, serve, and provide food for human consumption.

Our dataset consists of:

* Inspection ID
* Restaurant's name
* Restaurants’ “AKA” name
* License number
* Facility type
* Risk level
* Address, City, State, and ZIP
* Inspection Date
* Inspection Type

The data recorded was based off of inspections taken place between 2010 and 2019. In February 2019, however, there were major changes to the Chicago Food Code. The major changes of the food code were related to its Terms, Required Written Documents, Food Safety Training, Food Safety, Inspection Report, Variances, and Equipment Compliance. The enforcement of these new changes would take into effect on February 1st, 2019, but no citation will be issued for any violations. However, establishments are required to address these violations within 90 days.

The inspections were rated by their level of risk, with Risk 1 being a High-Risk establishment and Risk 3 being a Low Risk establishment.

* Risk Level 1:

“This risk level is assigned to a permit that allows limited preparation steps of potentially hazardous foods, such as hot dogs, and includes sectioning of melons, heating of individually pre-packaged ready-to-eat foods for immediate service without opening of the package, and preparation of espresso and/or blended drinks. It also includes cold holding of commercially pre-packaged ready-to-eat foods, such as sandwiches, without opening of the package. Risk 1 does not include hot holding of food. Mobile cart operations with espresso are included in this risk category as are mobile trucks with frozen foods or meat. **These types of operations are inspected once a year.**”

* Risk Level 2:

“This risk level is assigned to a permit that allows food processing steps such as receiving, storing, preparing, cold holding, and serving potentially hazardous foods. It does not include hot holding of food. It includes limited preparation steps, such as baking bread, frying donuts, and grilling or toasting sandwiches for immediate service. Examples of this type of operation include on-site baking with raw shell eggs as an ingredient, making smoothies with fresh fruit and/or fresh vegetables, opening ready to eat prepackaged foods for heating or service, cooking waffle cones or cake mixes. Grocery stores with pre-packaged raw meat, poultry, or seafood are also included. **These operations receive one routine inspection and one educational visit each year with the exception of schools which receive two routine inspections.**”

* Risk Level 3:

“This risk level is assigned to a permit that allows operations with complex food preparation steps, including thawing, cutting, cooking, cooling, cold holding, reheating, hot holding, and serving of potentially hazardous foods. It includes all operations that provide cooking or hot holding of foods, including meat and seafood markets and mobile trucks. **These operations receive two routine inspections and one educational visit each year.**”

There are four main inspection types in the City of Chicago, and these are shown within our dataset. The four main inspection types are:

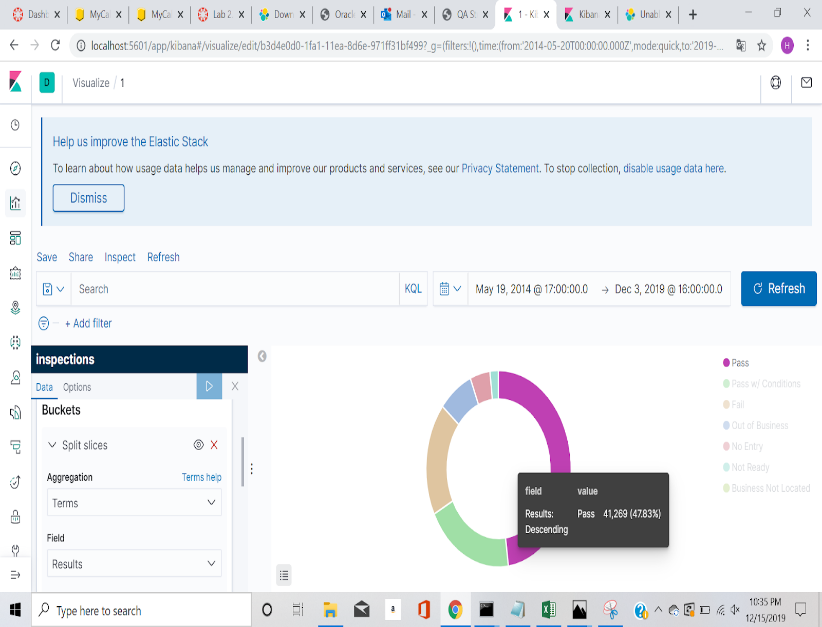
* Permits: inspection to achieve a permit
* Annual/Quarterly: inspection to maintain status of an establishment
* Business License: required inspection to obtain new licensing
* Complaints: inspection in response to a complaint

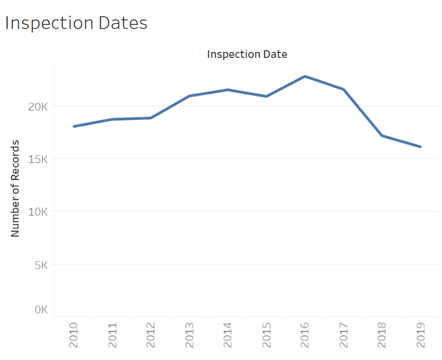
According to our dataset, the results of these inspections can be one of seven. Fail, No Entry, Not Ready, Out of Business, Pass, Pass with Conditions, or Business Not Located.

**Our Work:**

Data Size: 222Mb

Hardware Specs: Intel Core i5 6 core processor 3.2 Ghz, 16gb RAM, 500gb SSD storage.

****

****

**Conclusion:**

In this experiment we inspected a database that was about food health inspections. It specifically targeted inspections in Chicago and showed a plethora of information about each inspection such as but not limited to reason for failing, what health code violations they had, what establishment it was (Restaurant, Grocery store, School, and Children’s Daycare). We believe our work with this project is important because health is something that everyone universally cares about. Food is a huge part of that, so understanding how food inspections work should be something that everyone should be aware about.

We learned that restaurants have a risklevel attached to them. Risk 1 is considered a high-risk establishment and deals with relatively noncomplex goods like hot dog stands and fruit stands. Risk 2 is a step up in complexity and is a category designed for places like bakery and grocery stores with prepackaged raw meat. Risk 3 is reserved for the most complex establishments. These restaurants have complex food preparation and are held to a higher standard due to the probability of food-borne illness occurring if health code standards are not met.

By analyzing this data set we uncovered that only 48.76% of the establishments were able to pass an inspection on the first try. Another 19.94% of the establishments were able to pass with conditions, meaning that their passing grade was contingent on them updating their equipment/standards up to code. 18.55% failed the inspection, 7.7% of the establishments went out of business before another inspection could be performed. No entry was responsible for 4.46% of the institutions and 1.89% of the restaurants were simply not ready for the inspection. These results were surprising since only 68.7% of the institutions had standards that were up to code. Considering the disparity between what constitutes a “A” grade and a “B” grade shows how much slack these inspections give. Even a “C” is technically passing but is not recommended due to the fact that it is barely fit for human consumption. Considering these facts, it's shocking to see that only 68.7% of places passed.

**References:**

**City of Chicago."Health Inspections". *Chicago.Gov*, 2010,** [**https://www.chicago.gov/city/en/depts/cdph/provdrs/healthy\_restaurants.html. Accessed 15 Dec 2019**](https://www.chicago.gov/city/en/depts/cdph/provdrs/healthy_restaurants.html.%20Accessed%2015%20Dec%202019)**.**

**KingCounty. “Risk levels and permit classifications”.**

**KingCoutny.gov, 2019,**

**https://www.kingcounty.gov/depts/health/environmental-health/food-safety/food-business-permit/risk-levels-permit-classifications.aspx**

**Kannan, Vinesh, et al. “[PDF] Hindsight Analysis of the Chicago Food Inspection Forecasting Model: Semantic Scholar.” *Undefined*, 1 Jan. 1970,** [**www.semanticscholar.org/paper/Hindsight-Analysis-of-the-Chicago-Food-Inspection-Kannan-Shapiro/be88e4783329a7dd5298c797ce7b946e93271359**](http://www.semanticscholar.org/paper/Hindsight-Analysis-of-the-Chicago-Food-Inspection-Kannan-Shapiro/be88e4783329a7dd5298c797ce7b946e93271359)**.**

[**https://studio.azureml.net**](https://studio.azureml.net)

[**https://www.kaggle.com/chicago/chicago-food-inspections**](https://www.kaggle.com/chicago/chicago-food-inspections)

[**https://github.com/Ensaria/3200-Project**](https://github.com/Ensaria/3200-Project)