Problem Set 1: Part 1

MACS 30200

Leoson Hoay

1 How to access the data

The data is stored on the Chicago Data Portal and accessed via this link. The site provides some basic visualization, data manipulation, and viewing tools, while the entire dataset can be exported in various formats such as csv and JSON using the export button. According to the metadata provided, the dataset is curated by State of Illinois Licensed Environmental Health Practitioners (LEHPs) and owned by the Chicago Department of Public Health.

2 Key papers that use this dataset

I: One key paper that makes use of the dataset is the *2016 Chicago Department of Public Health Food Establishment Inspection Audit*, conducted by the Office of Inspector General of the City of Chicago. The audit was conducted the objectives to determine whether the Food Protection Division "conducted routine inspections of food establishments as often as required by the Departments rules and regulations incorporating State law, conducted inspections triggered by complaints and reinspections of known violations in a timely manner, and accurately reported the results of inspections and reinspections through the Citys Data Portal." The food inspections dataset was used alongside a number of other datasets in order to determine whether inspections fulfilled certain required standards such as frequency of inspections and timeliness of follow-ups.

Office of Inspector General, City of Chicago. (2016). Chicago Department of Public Health Food Establishment Inspection Audit. Chicago. Retrieved from http://chicagoinspectorgeneral.org/wp-content/uploads/2016/11/Audit-of-CDPH -Food-Establishment-Inspections.pdf

II: Another key paper that makes use of the dataset is *Fudging the Nudge: Information Disclosure and Restaurant Grading* by Daniel E Ho. The paper studies health inspections data from ten jurisdictions to examine the effectiveness of the restaurant letter grading system in alleviating actual manifestations of foodborne illness over time. The intriguing finding of the paper is that the grading system does not perform its function well, and possibly even causes grade-related disputes that directs resources away from addressing more serious health-related hazards in food establishments.

Ho, D. E. (2012). Fudging the Nudge: Information Disclosure and Restaurant Grading. The Yale Law Journal, 122(3), 574688.

III: One more paper that makes use of the dataset is *Co-creating an Open Government Data Driven Public Service: The Case of Chicagos Food Inspection Forecasting Model by Mcbride et. al.* The paper uses the dataset as a case study to examine the feasibility and potential of building predictive models from OGD (Open Government Data).

McBride, K., Aavik, G., Kalvet, T., & Krimmer, R. (2018). Co-creating an Open Government Data Driven Public Service: The Case of Chicagos Food Inspection Forecasting Model. The 51st Hawaii International Conference on System Sciences (HICSS). https://doi.org/10.24251

3 Descriptive characteristics

Table 1: Descriptives

| Variables	No. of Categories*	Max Type*	Max*	Min*
Risk	4	Risk 1 (High)	117586	22
Inspection Type	108	Canvass	88972	1
Results	7	Pass	98082	61
Facility Type	453	Restaurant	110828	1
DBA Name	25549	SUBWAY	2323	1
AKA Name	24401	SUBWAY	2800	1
Zip	100	60614	6384	1
No Violations	33679**	-	-	-

* Values included only for applicable variables
** Frequency of entries with no violations

The above table shows descriptives for 8 key variables - Risk, Inspection Type, Inspection Results, Facility Type, DBA Name, AKA Name, Zip, and Number of entries with no violations. For each of these variables, the number of categories, the category with the maximum frequency and its count, and the minimum frequency observed is displayed - if applicable. We can see that there are 33679 inspections that resulted in no violations being observed, and that the most frequently occurring inspection result is a 'Pass'.

4 Key Visualizations
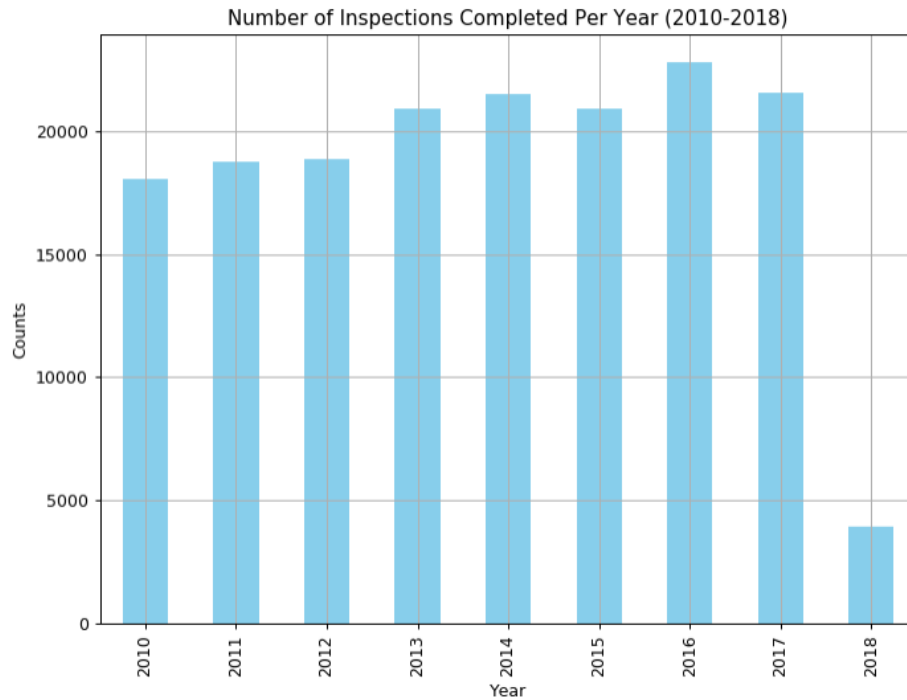
The above figure shows the frequencies of inspections for each year between 2010-2018. We can see that 2018 has the lowest inspection count (3926), which is most likely due to the fact that the year has not yet ended. For the other years that precede it, there is a slight upward trend in the total number of inspections completed per year (highest in 2016, with a count of 22816) - although it is unclear whether the direct reason for this is an increase in the number of restaurants in the Chicago market, or an increase in range, scope, and frequency of the inspections conducted by the Food Protection Division of the Chicago Department of Public Health (CDPH), or both. In any case, there is some variability in the number of inspections conducted each year.
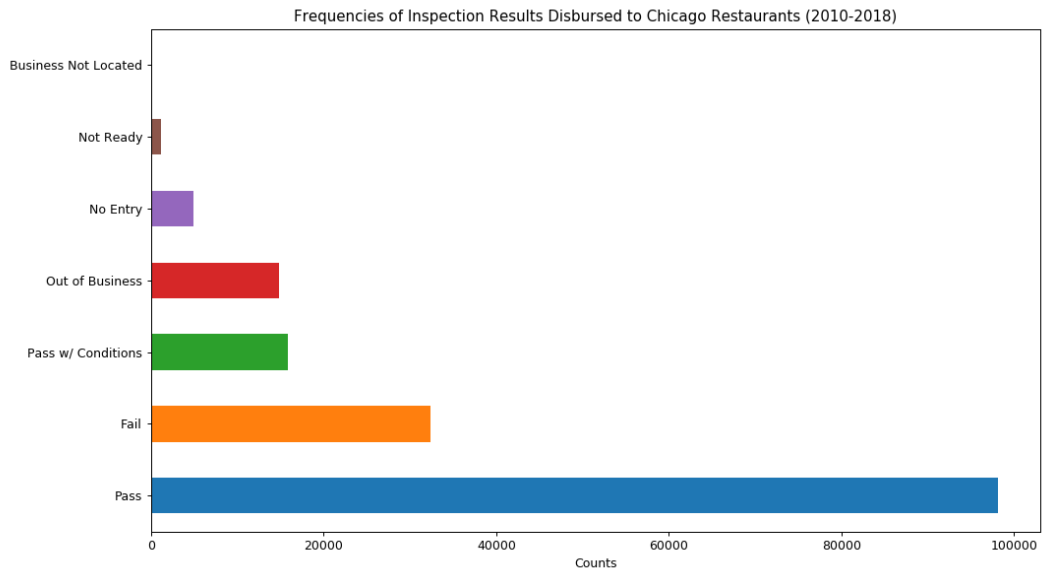
Frequencies of Inspection Results Disbursed to Chicago Restaurants (2010-2018)

The above figure shows the frequencies of inspections for each year between 2010-2018. We can see that 'Pass' is the most frequently awarded result (with a count of 98082 out of 167412) across the years. A simple calculation yields the fact that a 'Pass' has been awarded approximately 58.6% of the time.

5 Conditional Descriptives by Inspection Result Type

Table 2: Variable Maximum Categories and their Occurences

Result Type	Risk*	Inspection Type*	Facility*	DBA Name*	AKA Name*	Zip*	No Violations**	Year*
Pass	Risk 1 (High) / 71428	Canvass / 45754	Restaurant / 65374	SUBWAY / 1491	SUBWAY / 1796	60611 / 3360	10048	2016 / 12928
Pass w/ Conditions	Risk 1 (High) / 10244	Canvass / 9729	Restaurant / 12325	SUBWAY / 392	SUBWAY / 469	60614 / 873	121	2016 / 2633
Fail	Risk 1 (High) / 22389	Canvass / 15056	Restaurant / 20751	SUBWAY / 287	SUBWAY / 363	60614 / 1281	2903	2010 / 4504
Not Ready	Risk 1 (High) / 578	License / 1038	Restaurant / 716	PALETERIA / 17	LA MEXICANA / 17	60622 / 44	1099	2017 / 423
No Entry	Risk 1 (High) / 3899	Canvass / 3726	Restaurant / 4107	SHERNIQUES / 11	MARTINI CLUB / 16	60614 / 341	4605	2017 / 1160
Not Located	Risk 1 (High) / 27	Canvass / 29	Restaurant / 22	DUNKIN DONUTS / 3	HALE FRAN-CISCAN / 2	60608 / 6	61	2011 / 16
Out of Business	Risk 1 (High) / 7221	Canvass / 14637	Restaurant / 7533	SUBWAY / 142	SUBWAY / 159	60647 / 629	14842	2013 / 3171

* Maximum Category Type/Occurences
** Occurences

The above table displays conditional descriptive counts for 8 variables in the dataset based on the Inspection Result variable. It is interesting to note that the highest occurences of Out of Business facilities in the dataset was in 2013, and also that certain facilities that recorded no violations still failed the inspections.