EAS 503

Project Report
on
FOOD INSPECTION

Submitted To:

Professor: Varun Chandola

Submitted By:

Group 20
Mohan Krishna Gaddam (UB 50290269)
Jai Kishore (UB 50290269)
Kartik Bapna (UB 50291058)

Abstract

Chicago Food Inspection data was analyzed based on the data obtained from Chicago Data Portal, Protection Division of the Chicago Department of Public Health (CDPH) is committed to maintaining the safety of food bought, sold, or prepared for public consumption in Chicago by carrying out science-based inspections of all retail food establishments. These inspections promote public health in areas of food safety and sanitation and prevent the occurrence of food-borne illness.

Our observation from the data was carried out majorly from three phases:
1) Exploratory Data Analysis
2) Data Cleaning
3) Basic Text Mining

The interpretation of the result were carried out in Python using Pandas Library, Matplotlib Library and SQL, The data is comprised of both numerical and categorical data. Through data visualization we were able to draw significant inferences that can be used to improve food safety and avoid food illness.

Contents

1. Introduction

There are around 15,000 business establishments serving food and drink across the City of Chicago that are subject to sanitation inspections by the Department of Public Health at least once in a year. Factors such as worker's hygiene, correct holding and storing temperatures for foods and making sure the food comes from authorized sources with expiry dates maintained, etc. are all critical for ensuring public health. Each year a restaurant is subject to annual inspections to ensure continued compliance with City ordinances and regulations and to reduce public's exposure to any kind of food borne disease. In addition to recurring inspections, restaurants may also be inspected in response to a complaint. Some of these recurring inspections, such as the inspection by the Buildings Department, will be scheduled, while others will not

2. Data

2.1. Description

This data information was obtained from inspections of restaurants and other food and drink establishments in Chicago from January 1, 2010 to the present. Details of the significant variables:

- DBA: 'Doing business as.' This is legal name of the establishment.

- Facility Type: Each establishment is described by one of the following: bakery, banquet hall, candy store, caterer, coffee shop, day care center (for ages less than 2), day care center (for ages 2 – 6), day care center (for ages less than 2 and 2 – 6 combined), gas station, Golden Diner, grocery store, hospital, long term care center (nursing home), liquor store, mobile food dispenser, restaurant, school, shelter, tavern, social club, wholesaler, or Wrigley Field Rooftop.

- Risk: Each establishment is categorized as to its risk of adversely affecting the public's health, with 1 being the highest and 3 the lowest. The frequency of inspection is tied to this risk, with risk 1 facilities inspected most frequently and risk 3 least frequently.

- Street address and zip code of facility: Complete address where the facility is located.

- Inspection date: This is the date the inspection occurred.

- Inspection type: One of the following types:

  - canvass, the most common type of inspection performed at a frequency relative to the risk of the establishment;

  - consultation, when the inspection is done at the request of the owner prior to the opening of the establishment;

  - complaint, when the inspection is done in response to a complaint against the establishment; license, when the inspection is done as a requirement for the establishment to receive its license to operate;

  - suspect food poisoning, when the inspection is done in response to one or more persons claiming to have gotten ill as a result of eating at the establishment (a specific type of complaint-based inspection);

  - task-force inspection, when an inspection of a bar or tavern is done.

  - Re-inspections can occur for most types of these inspections and are indicated as such.

- Results: An inspection can pass, pass with conditions or fail. Establishments receiving a 'pass' were found to have no critical or serious violations (violation number 1-14 and 15- 29, respectively). Establishments receiving a 'pass with conditions' were found to have critical or serious violations, but these were corrected during the inspection. Establishments receiving a 'fail' were found to have critical or serious violations that were not correctable during the inspection.

- Violations: An establishment can receive one or more of 45 distinct violations (violation numbers 1-44 and 70). For each violation number listed for a given establishment, the requirement the establishment must meet in order for it to NOT receive a violation is noted, followed by a specific description of the findings that caused the violation to be issued.

2.2. Steps of Data Cleaning

- Unnecessary columns with respect to this analysis were removed.
- Data of facility type and inspection type were not maintained uniformly. So, steps were taken to maintain a uniformity in data in Python and SQL
- Attributes like Location were having 668 missing values so removed them
- Removed values in Results columns which were not relevant and keep fields like pass, fail or pass with conditions
- Steps were taken to remove duplicate inspection reports.
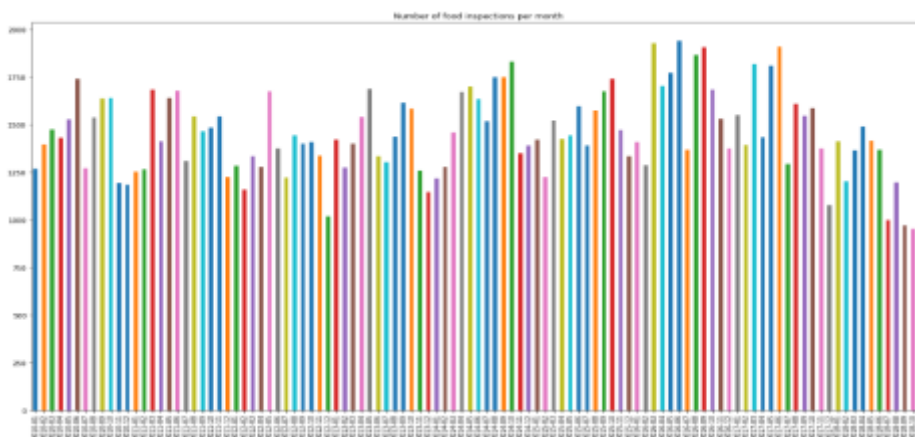- For easy interpretation of the data columns, column name were changed to upper case

3. Visualization

3.1 Histogram plot for results
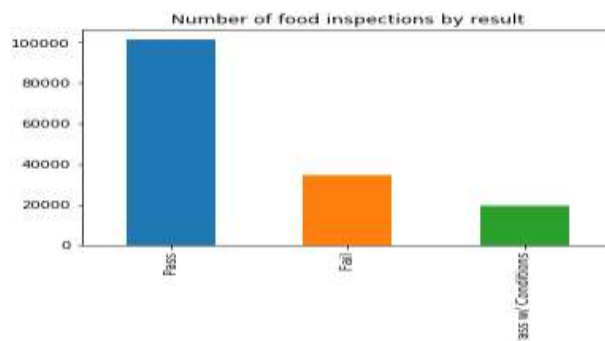
We can observe that majority of the food inspection result passed the inspection audit and approx to 18% failed

3.2 Histogram plot for food inspection per month

The above plot shows the number of inspections per month from 2010 to 2018 and the maximum number of inspection was during June 2016

3.3 Histogram plot for each risk type

Number of food inspections by risk
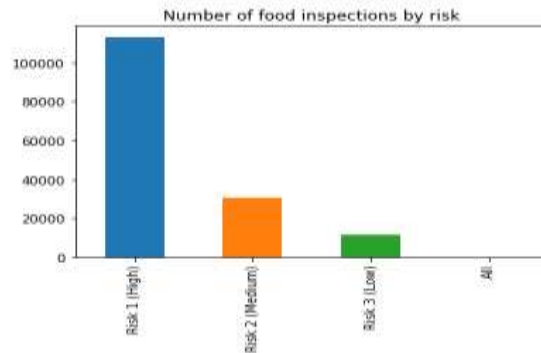
Here each establishment is categorized as to its risk of adversely affecting the public's health, with 1 being the highest and 3 the lowest. The frequency of inspection is tied to this risk and we can infer from the plot that majority of the food inspection were considered as High

3.4 Plot the number of inspections by zipcode

Number of food inspections by Zip
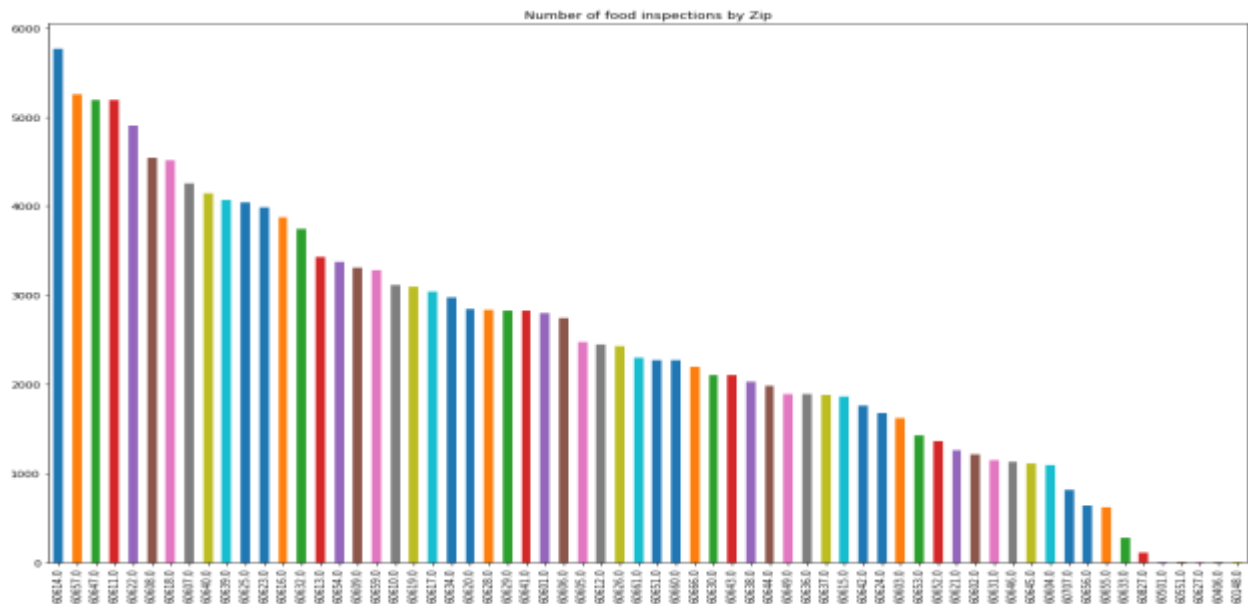
Zipcode 606140 had the maximum number of inspection

3.5 Plot with respect to facility type
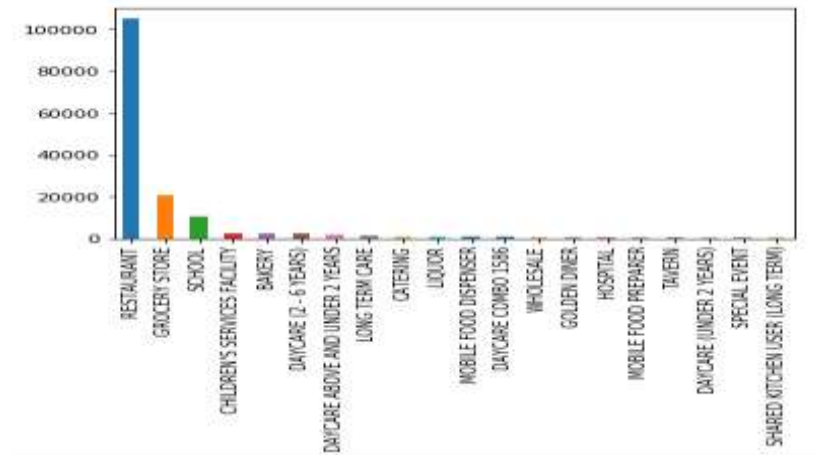
We can observe that majority of the inspection were in restaurant and least were in shared kitchen user

3.6 Plot of addresses with most inspection
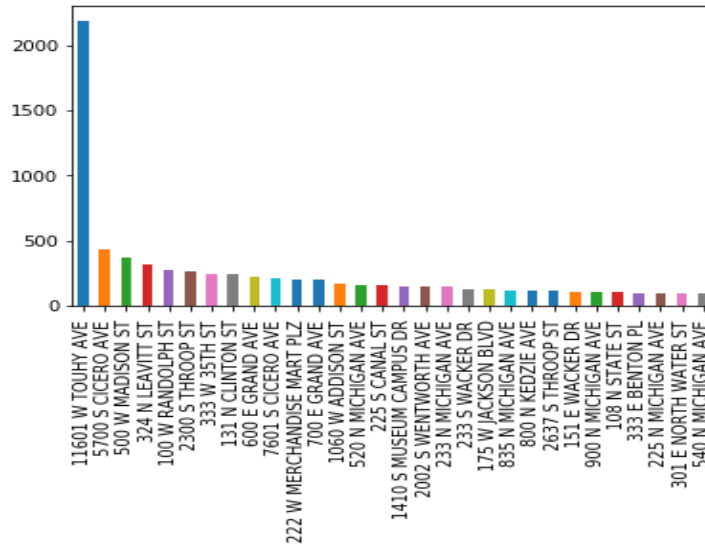
Among all the address listed Touhy Avenue had the max inspections

3.7 Plot for type of Inspection
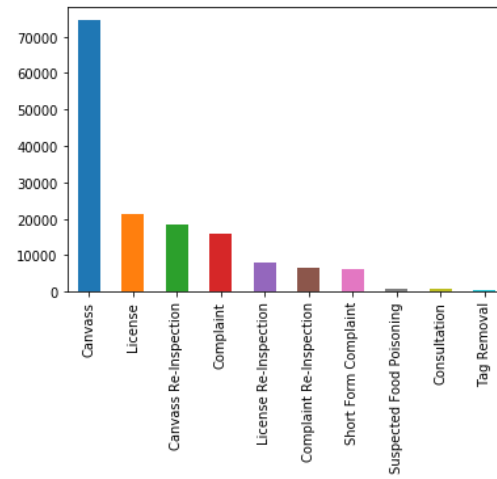
Canvass was the most significant type of inspection

3.8 Plot with respect to type of food chain

Subway chain was inspected the most

3.9 Chicago census plot where most of the inspection failed

4.0 Using SQl some of the results which were found

4.1 Number of failed inspections per facility type:

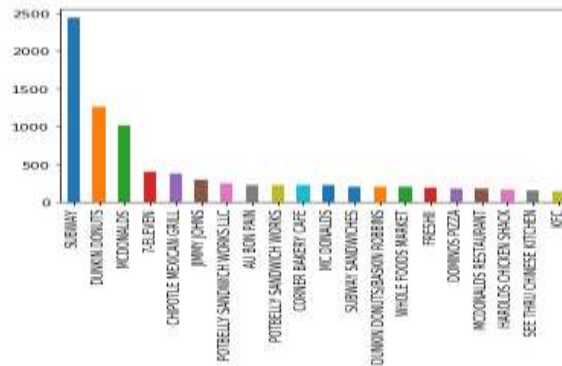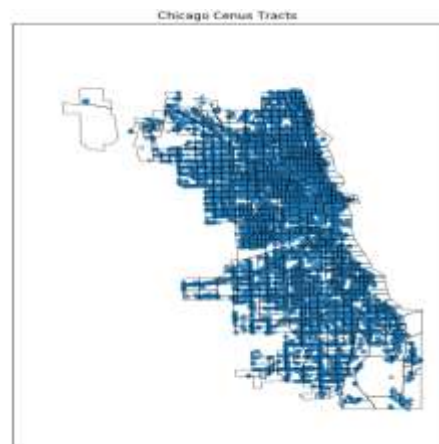|	Facility_Type	Total_Failed_Inspections
0	Restaurant	22087
1	Grocery Store	5549
2	School	2158
3	Bakery	557
4	Daycare (2 - 6 Years)	495

From the data, we can see that majority of inspections were done for facility type Restaurant. Then we checked the Results of these frequently inspected facility type. So we came up with analysis where we showed the all the Facility types which have maximum failed results. On doing that we see that Restaurant has highest number of failed records followed by Grocery,Schools,Bakery and Daycare.

4.2 How much number of inspections varied from 2016 to 2017?

	Decrease_in_Inspections_from_2016_to_2017
0	1234

We can see there was a decrease in inspection from 2016 to 2017

4.3 How much number of failed inspections varied from 2016 to 2017?

	Decrease_in_failedinspections_from_2016_to_2017
0	179

We can see there was a decrease of 179 failed inspection

4.4 Which place had maximum number of failed inspections ?

	DBA_Name	Address	Location	Num_of_Failed_Inspections
0	ILLINOIS SPORTSERVICE INC.	333 W 35TH ST	(41.83078366228312, -87.6352957830455)	31
1	Orly's/Jalapeno	1660 E 55TH ST	(41.79633443394228, -87.58427242325949)	16
2	JIMMY G'S	307 S KEDZIE AVE	(41.87707753773826, -87.70586030035169)	15
3	JOE'S BARBQUE, INC	4900 W MADISON ST	(41.880571145964275, -87.74795724892677)	14
4	IYANZE	4623-4627 N BROADWAY	(41.96606299325794, -87.65773412452839)	14
5	Food 4 Less Midwest #662	7030 S ASHLAND AVE	(41.76579077043167, -87.66396688645497)	13
6	Congress Pizza	2033 N Milwaukee AVE	(41.91826595568366, -87.68952838043421)	13
7	McDONALD'S	70 E GARFIELD BLVD	(41.79479402426613, -87.62291510355041)	13
8	BLUE CORAL CONTEMPORARY SUSHI AND LOUNGE	1265 N MILWAUKEE AVE	(41.90485662426828, -87.66859020422658)	13
9	LATIN AMERICAN REST & LOUNGE	2743 W DIVISION ST	(41.90282132997605, -87.69614329943664)	13

Illinois Sport Service has the most failed inspection

4.5 What is the percentage which didn`t pass the inspections that were carried out because of complaints?

	Percentage
0	55.15657

55.15 percent were not able to pass the inspection which were carried out due to complaints

4.6 What is the percentage failed the inspections that were carried out because of "Suspected Food Poisoning"?

	Percentage
0	51.41026

51.41 percent failed because of the suspect that there is food poisoning

4.7 What is the percentage which didn`t pass the inspections that were carried out because of "Complaint Re-Inspection"?

	Percentage
0	20.28622

20.28 percent failed post Re Inspection

4.8 What is the percentage which didn`t pass the inspections that were carried out because of "Consultation"?

	Percentage
0	37.31343

37.31 percent failed because of Consultation

4.9 Which Restaurants chains have most failed inspections?

	DBA_Name	No_of_failed_inspections
0	SUBWAY	338
1	DUNKIN DONUTS	205
2	MCDONALD'S	140

Subway chain failed maximum number of inspections and the same can be observed from the below plot.


Restaurant chains with most failed inspections
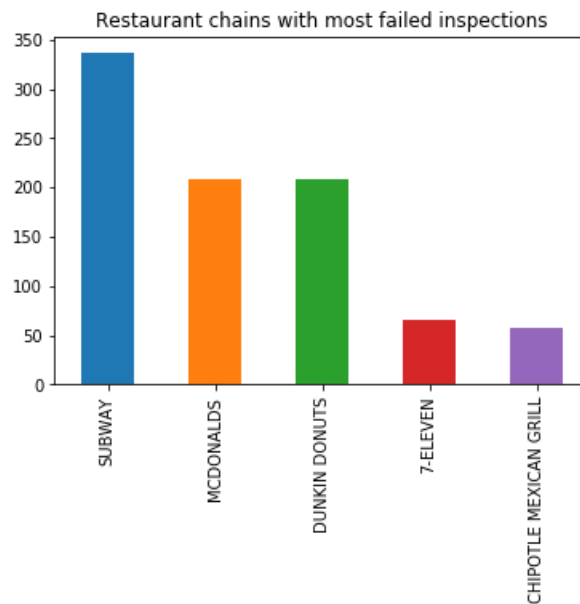
4.10 Which subway restaurants failed most inspections?

	License	Address	Zip	No_of_failed_inspections
0	2262950	2008 N HALSTED ST	60614	6
1	2032865	1212 S MICHIGAN AVE	60605	6
2	1679112	8711 S ASHLAND AVE	60620	6
3	25182	5040 S Cottage Grove AVE	60615	6
4	2129629	2811 S KEDZIE AVE	60623	6

We can see from the above table that all these subway outlets failed maximum time during inspection

4.11 Which individual restaurants failed most inspections?

|	DBA_Name	Address	Zip	No_of_failed_inspections
0	Orly's/Jalapeno	1660 E 55TH ST	60615	16
1	JIMMY G'S	307 S KEDZIE AVE	60612	15
2	IYANZE	4623-4627 N BROADWAY	60640	14
3	JOE'S BARBQUE, INC	4900 W MADISON ST	60644	14
4	LATIN AMERICAN REST & LOUNGE	2743 W DIVISION ST	60622	13

Orly's Jalapeno failed maximum type under the individual restaurant category and can also be observed from the plot below
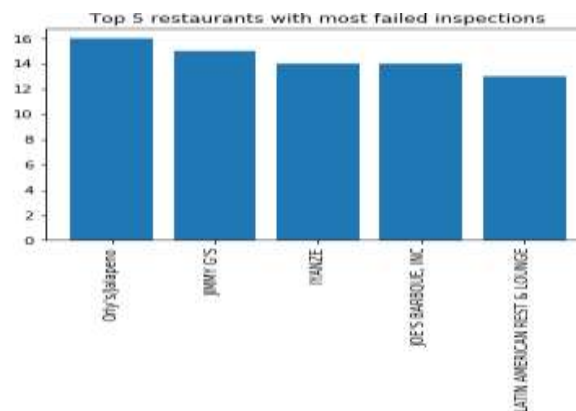
4.12 Addresses with most number of failed inspections ?

	Address	No_of_failed_inspections
0	11601 W TOUHY AVE	302
1	324 N LEAVITT ST	93
2	2300 S THROOP ST	78
3	500 W MADISON ST	73
4	5700 S CICERO AVE	60

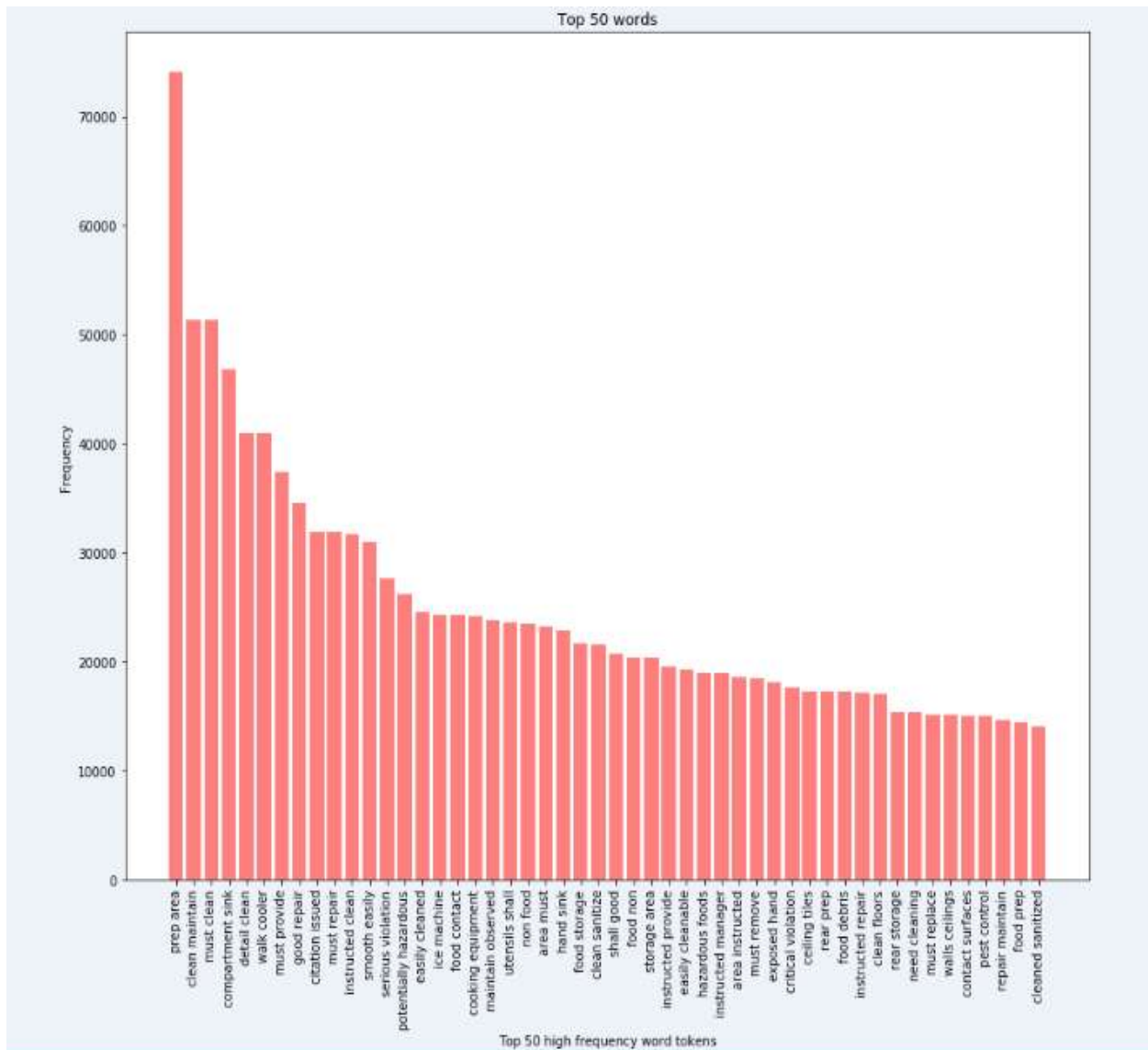We can see that Touhy Avenue is the address where maximum number of food inspection failed.

Top 5 Addresses with most failed inspections
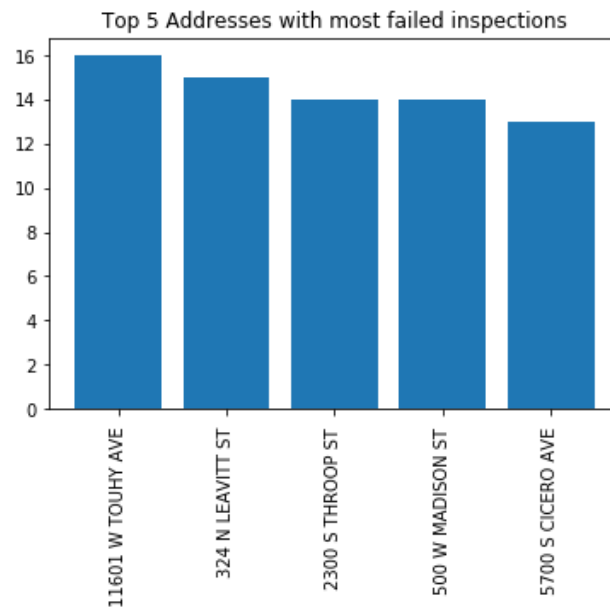
5. Text Mining

5.1 Top 50 highest frequency word tokens that were used in the violations column

Top 50 words
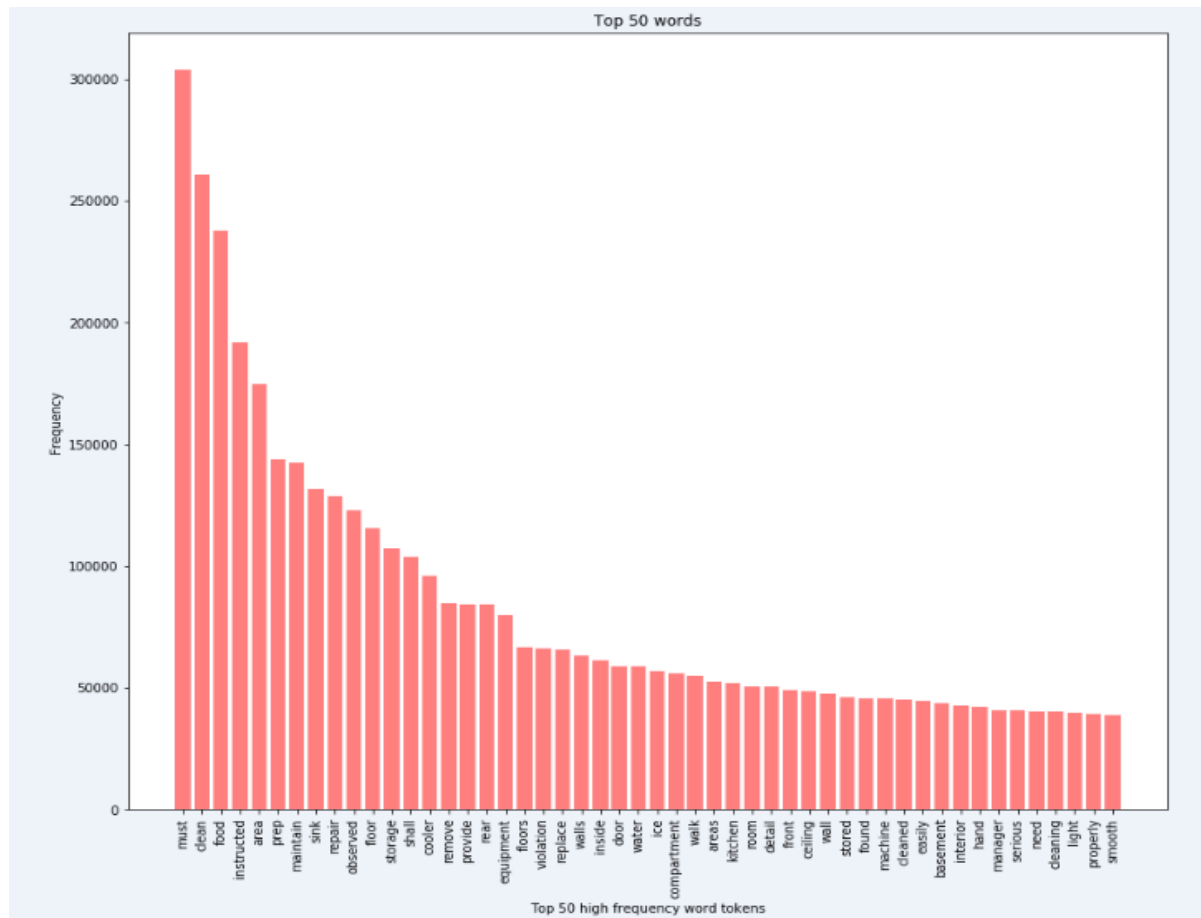
5.2 50 highest bi-grams used in the text

Top 50 words

5.3 Word Cloud for the entire corpus

6.0 Conclusion

The analysis made on the data set gave us a fair idea as to how food inspection is carried out not only in the city of Chicago, The violations provided good insight as to how the restaurants are being classified according to the number of the violation they are penalized with. On top of that, we inferred about the establishments/business units that are being inspected frequently and the risk involved with each of them. When factoring all these together, the research provided the vision as to which establishments should be inspected first.

7.0 Future Research

Based on the score obtained after Sentiment Analysis on the text in the violations column, we will be able to predict the outcome of the inspection i.e, whether the inspection failed/passed.