PYTHON PROJECT REPORT

(EAS_503 Aug - Dec, 2017)

Chicago Food Inspection

Submitted by

Group 3

Name	UB#
Alekya Kumar	50249052
Monalisa Mishra	50250673
Trupti Jadav	50249177

Abstract

Analysis of Food Inspection based on the data obtained from Chicago Data Portal. The Food 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 analysis was done using Python Pandas Library, Matplotlib Library and SQL. We were able to study about each of the predictors or the variables in the data set. The data set that we dealt with is mostly of textual type and involved a set of categorical variables. Studying the variables gave us an idea as to how each of them gives out a meaning for the data points. We were able to come up with relationships among the various input variables and the same were visualized.

Table of 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 foodborne 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
- Steps were taken to remove duplicate inspection reports. 180 duplicate records were removed.
- Violation column contained textual data. For analysis purpose, the data was parsed and disintegrated to two different columns, containing violation codes and their respective counts. And violation descriptions were maintained in a separate csv file.

3. Analysis & Results

3.1. Histogram of Results

We plotted a histogram for Results to see how the inspections were treated and recorded as. 'Results' is a categorical variable and it has only 3 values. Majority of the inspections were recorded as 'Pass', few others were recorded as 'Fail' and others were 'Pass with condition'. This indicates that the most establishments were adhered to the regulations.
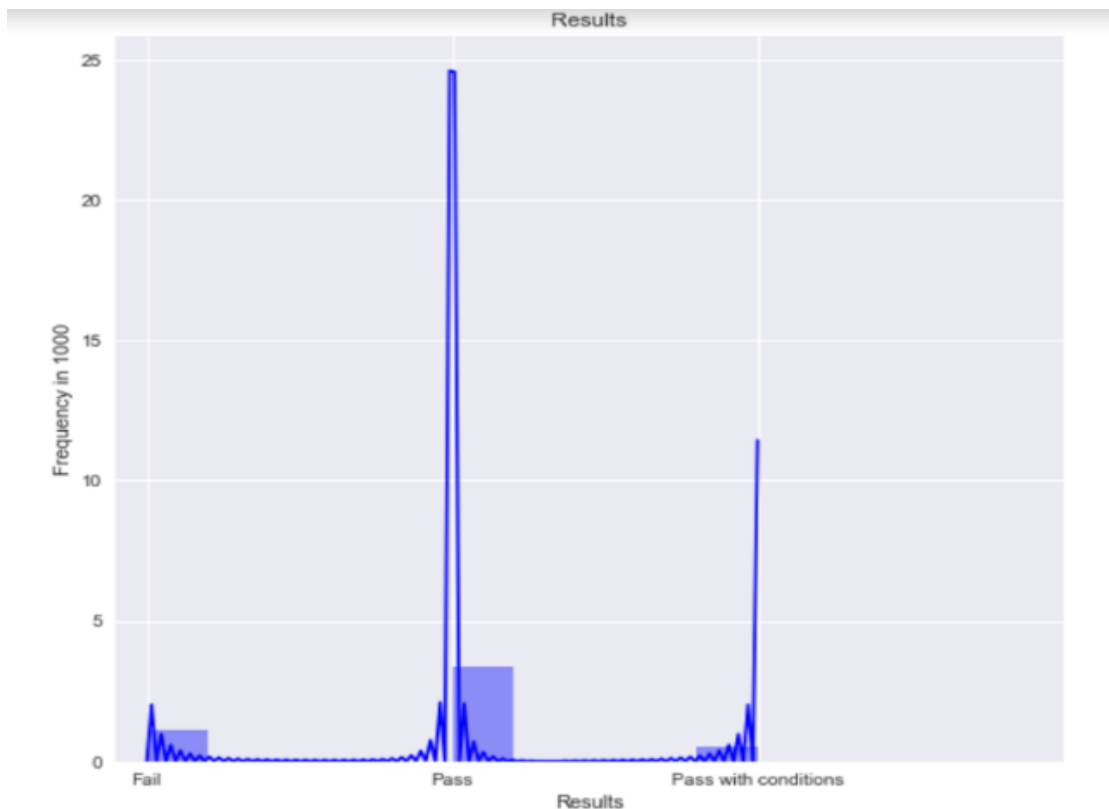

Fig1. Histogram of Results

3.2.   Time Series Plot for Inspection Type and Facility Type

The facility types had a lot of duplicates which include the misspelling of establishment names. The duplicates were corrected, and the data was presented. Using the pivot table, the number of inspections each facility type underwent in the entire span of January,2010 to September,2017 was calculated. The count was plotted with each line indicating the different facility types.

To get a better idea about what the graph depicts, we went ahead and retrieved the count for each facility type and grouped only the top ten among them.
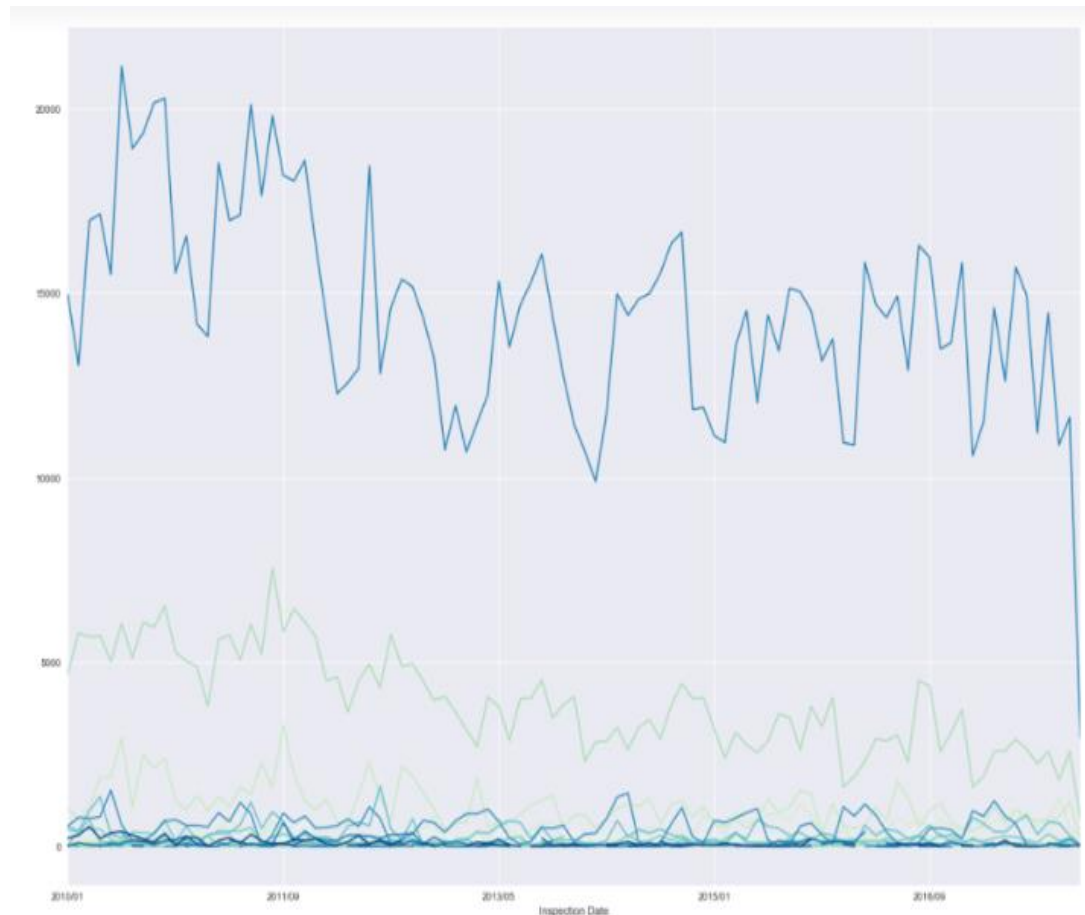

Fig2. Time Series Plot for Inspection Type and Facility Type

3.3. Facility Type against No. of Inspections

We analyzed 'Facility Type' who have undergone 500 and above inspections. From the table, we can see that 'Restaurant' has highest record of inspections done followed by 'Grocery/Retail Store', 'Private/Public School'.

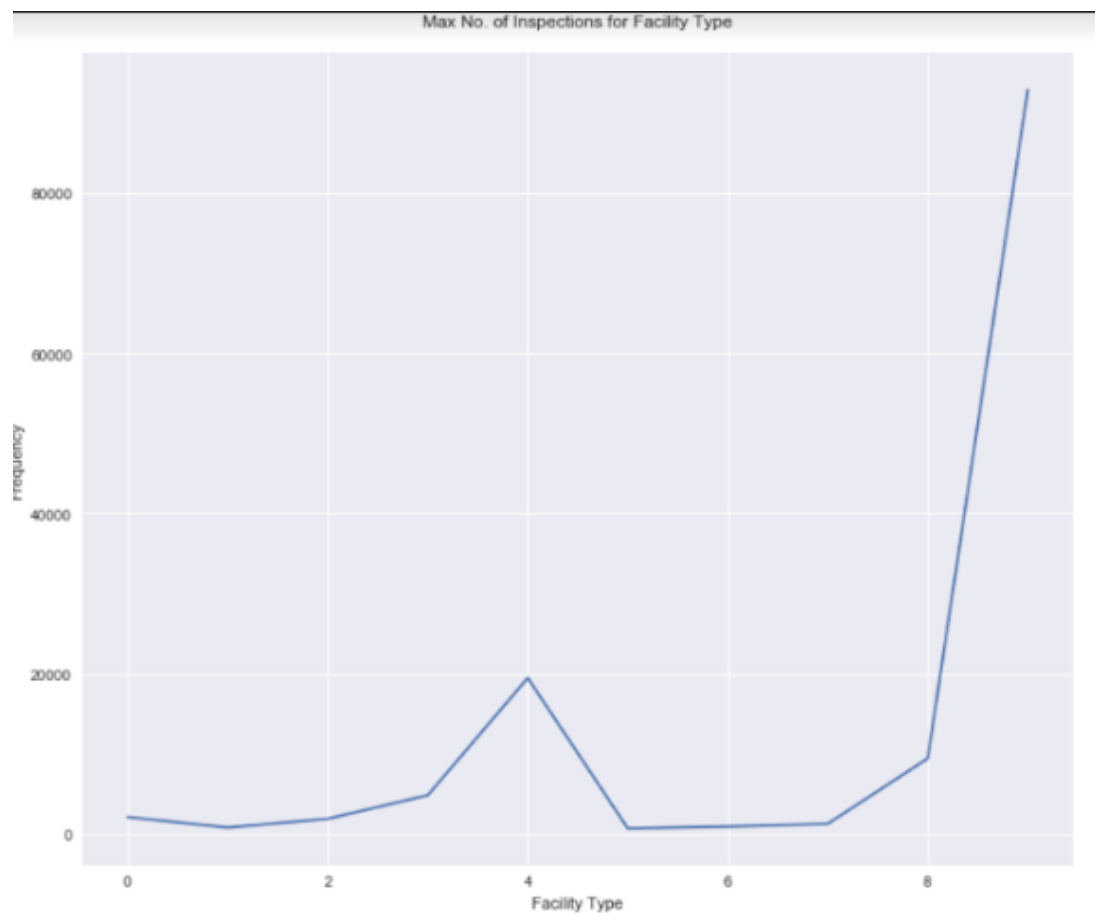|	facility_type	no_of_inspections
0	BAKERY/RESTAURANT	2087
1	CATERING SERVICE	787
2	CHILDRENS FACILITY	1876
3	DAY CARE	4818
4	GROCERY/RETAIL STORE	19474
5	LIQUOR STORE	683
6	LONG TERM CARE FACILITY	915
7	MOBILE FOOD TRUCK/DISPENSER	1254
8	PRIVATE/PUBLIC SCHOOL	9431
9	RESTAURANT	92946

Fig3. Facility Type and No. of Inspections

3.4. Result and No. of Violation Analysis

The base map of city of Chicago shows the areas where the establishments have more than 10 number of violations, given after inspection. There are three kinds pf results - Pass, Fail and Pass with conditions. As we can see in the map, "Pass with conditions" prevails the most in the upper part of the city. The result indicates that each establishment was awarded a violation in every inspection to be corrected before the next recurring inspection.
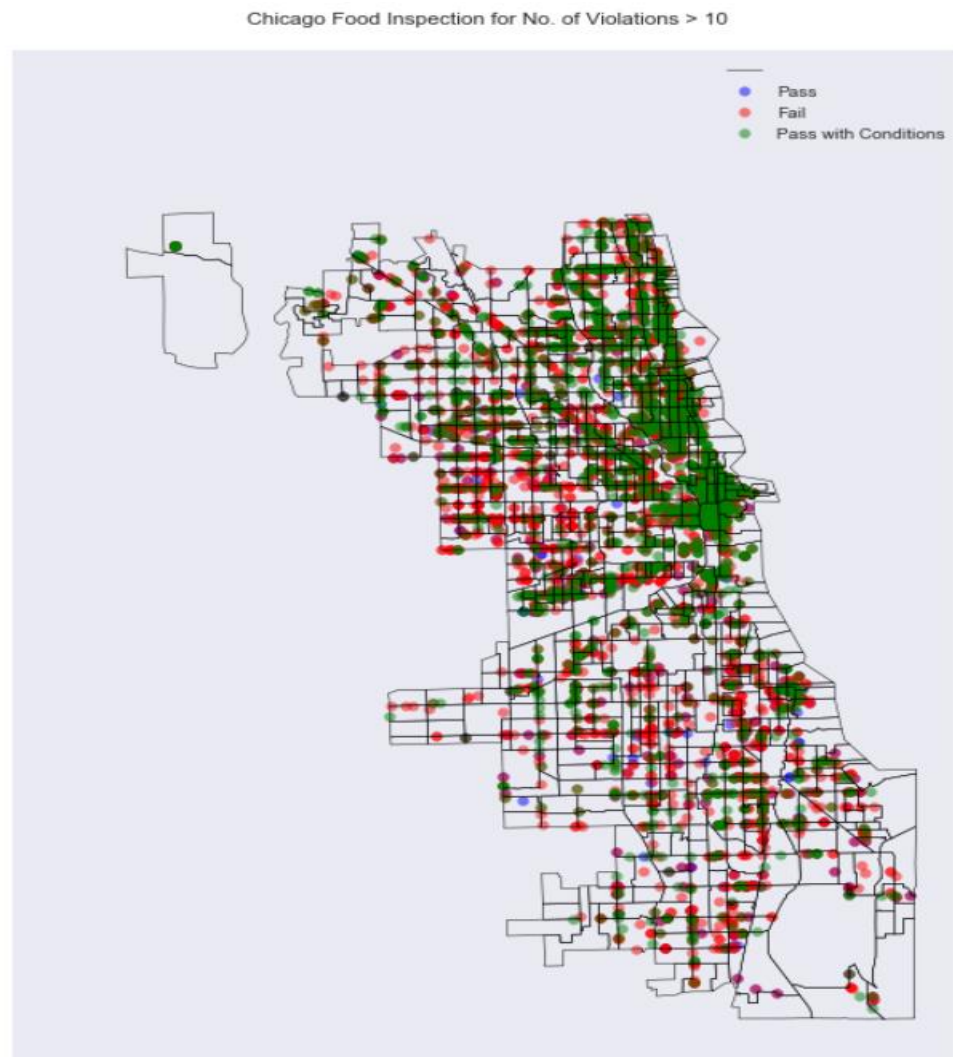

Fig4. Result and No. of Violation Analysis

3.5. Inspection Types and its respective counts

The next plot shows the various types of inspections each establishment undergoes and the corresponding count of the inspection.

Inspection types which occurred more than 100 times were taken into account. As we can see from the table, Canvass inspection type has the maximum count. Canvass is the type of inspection that is done in restaurants/food outlets to check whether the roofing and the storage area are properly maintained to keep the food in required optimal temperature.

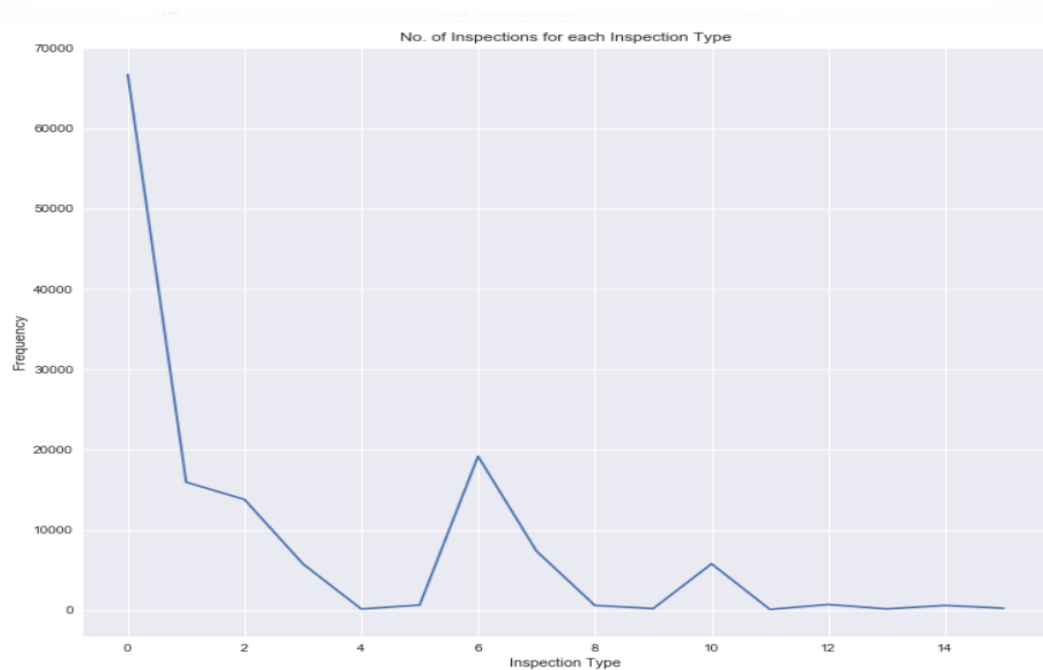|	Inspection_Type	no_of_inspections
0	CANVASS	66712
1	CANVASS RE-INSPECTION	15949
2	COMPLAINT	13790
3	COMPLAINT RE-INSPECTION	5778
4	COMPLAINT-FIRE	154
5	CONSULTATION	650
6	LICENSE	19171
7	LICENSE RE-INSPECTION	7379
8	LICENSE-TASK FORCE	603
9	RECENT INSPECTION	207
10	SHORT FORM COMPLAINT	5783
11	SHORT FORM FIRE-COMPLAINT	104
12	SUSPECTED FOOD POISONING	717
13	SUSPECTED FOOD POISONING RE-IN	165
14	TAG REMOVAL	601
15	TASK FORCE LIQUOR 1475	253

Fig5. Inspection Types and its respective counts

3.6. Risk Analysis in different zip codes

'Risk' according to their intensity are displayed on the Chicago map. This shows us the area wise distribution of the 'Risk'. We cannot tell much about the risk from the map as they are evenly scattered across the city. Predominantly, Risk 1 and Risk 2 are observed more. We can also see the combination of risks at many places (indicated by the brown points on the map). Also there are many crowded points observed in Main City of Chicago (Eastern region) indicating most inspection zone area.
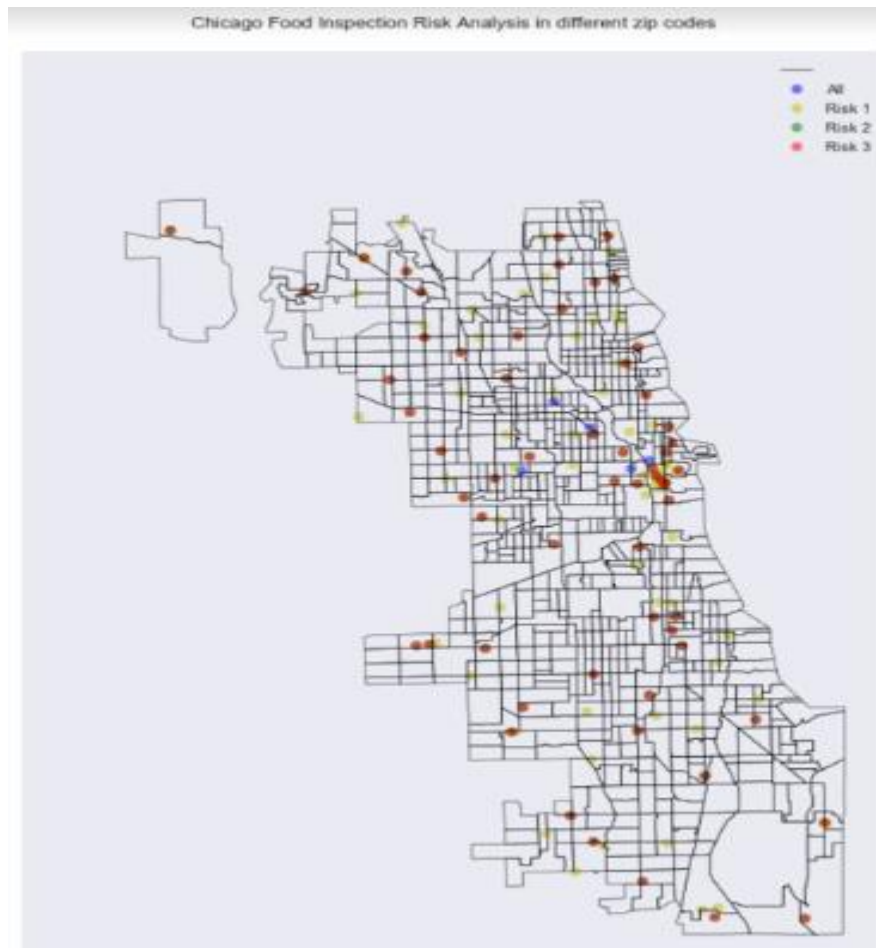

Fig6. Risk Analysis in different zip codes

3.7. Business Unit Branches that have maximum number of inspections

Business units with branches with most number of inspections were extracted from the data set. This graph depicts the business unit branch units which underwent the maximum number of inspections. The number of inspections is directly proportional to the risk associated with each business unit. Surprisingly, the Admiral café has the largest number of inspections in the city of Chicago.
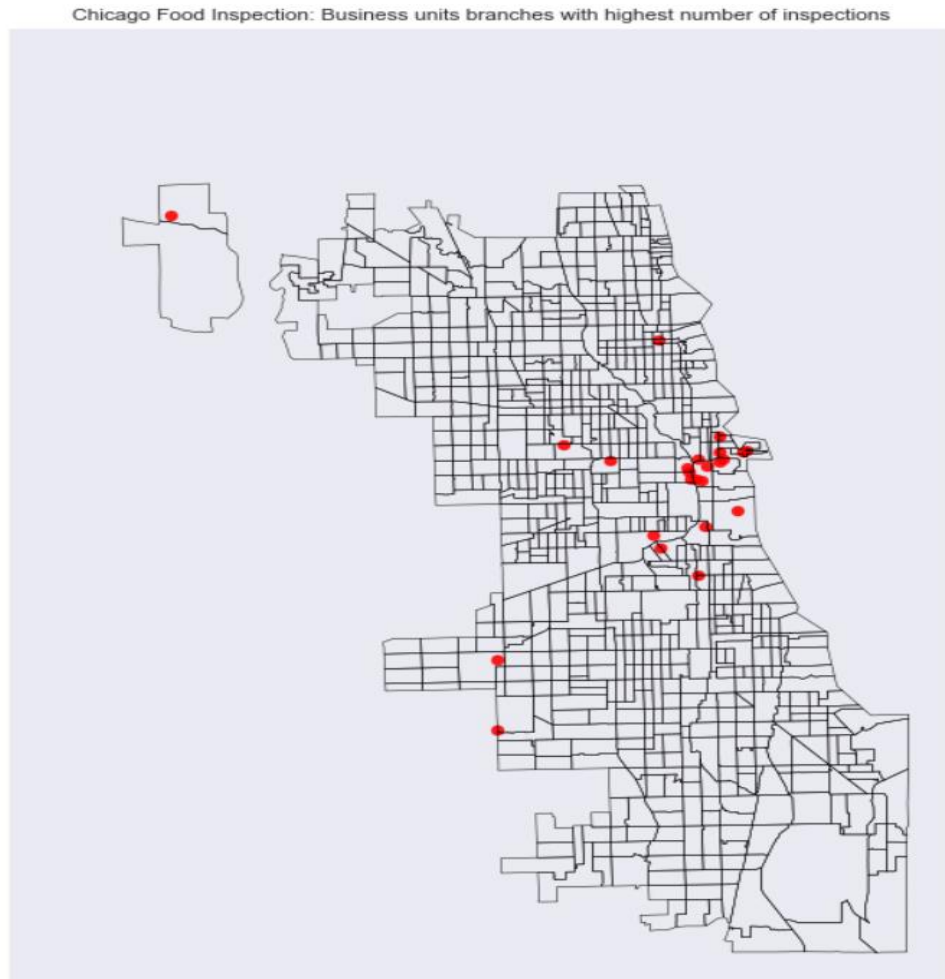

Fig7. Business Unit Branches having maximum number of inspections

The plot displayed above gave us the insight about the business unit that is inspected most with respect to branch. Displayed below, is the table that shows the inspection count for each Business Chain. This table gives the real data helping us conclude which establishment undergoes more number of inspections, cumulatively.

	No_of_Inspections	DBA_Name
0	2177	SUBWAY
1	1135	DUNKIN DONUTS
2	586	MCDONALD'S
3	378	7-ELEVEN
4	349	MCDONALDS
5	343	CHIPOTLE MEXICAN GRILL
6	222	POTBELLY SANDWICH WORKS LLC
7	200	POTBELLY SANDWICH WORKS
8	197	AU BON PAIN
9	196	CORNER BAKERY CAFE
10	179	SUBWAY SANDWICHES
11	171	WHOLE FOODS MARKET
12	170	DUNKIN DONUTS/BASKIN ROBBINS
13	159	DOMINO'S PIZZA
14	153	FRESHII
15	146	HAROLD'S CHICKEN SHACK
16	145	JIMMY JOHNS
17	137	SEE THRU CHINESE KITCHEN
18	134	KFC
19	115	MC DONALD'S
20	113	SHARKS FISH & CHICKEN
21	110	STARBUCKS
22	107	J & J FISH
23	106	PIZZA HUT
24	105	DUNKIN DONUTS / BASKIN ROBBINS
25	101	LAS ISLAS MARIAS
26	98	PAPA JOHN'S PIZZA
27	98	LOU MALNATI'S PIZZERIA
28	98	JIMMY JOHN'S

3.8. What type of ratings do establishments get?

Since we do not have a parameter that gives the ratings to the establishments, we have taken Risk factor for this analytical question. As we can see, establishments related to food have the maximum risk – Risk1 or Risk2. Banquets, Restaurants, Bakeries have the maximum risk. As indicated earlier, the restaurants with high risk will have undergone large number of repeated inspections.

	Facility_Type	Risk
0	(GAS STATION)	Risk 3 (Low)
1	1584-DAY CARE ABOVE 2 YEARS	Risk 1 (High)
2	A-NOT-FOR-PROFIT CHEF TRAINING PROGRAM	Risk 1 (High)
3	ADULT FAMILY CARE CENTER	Risk 1 (High)
4	AIRPORT LOUNGE	Risk 3 (Low)
5	ANIMAL SHELTER CAFE PERMIT	Risk 3 (Low)
6	ART GALLERY	Risk 3 (Low)
7	ASSISSTED LIVING	Risk 1 (High)
8	BAKERY/RESTAURANT	Risk 2 (Medium)
9	BANQUET FACILITY	Risk 1 (High)
10	BAR	Risk 3 (Low)
11	BAR/GRILL	Risk 2 (Medium)
12	BEFORE AND AFTER SCHOOL PROGRAM	Risk 2 (Medium)
13	BEVERAGE/SILVERWARE WAREHOUSE	Risk 3 (Low)
14	BLOCKBUSTER VIDEO	Risk 3 (Low)
15	BOWLING LANES/BANQUETS	Risk 1 (High)
16	BOYS AND GIRLS CLUB	Risk 2 (Medium)
17	BREWERY	Risk 3 (Low)
18	BREWPUB	Risk 1 (High)
19	BUTCHER SHOP	Risk 2 (Medium)

3.9. Which Restaurant chains fail inspections most?

There are failures common in Restaurants or food outlets. Analysis was made on the restaurants to check which restaurant faces the most number of failures since failures are more common in areas where food storage comes into picture. From the data, it's clear that Subway, Dunkin Donuts and McDonald's faced most no. of inspections.

	DBA_Name	Count_of_Fail_as_Result
0	SUBWAY	289
1	DUNKIN DONUTS	183
2	MCDONALD'S	127

3.10. Where are failures most common?

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/Retail, Private/Public Schools.

	Facility_Type	Count_of_Fail_as_Result
0	RESTAURANT	19513
1	GROCERY/RETAIL STORE	5261
2	PRIVATE/PUBLIC SCHOOL	1814
3	DAY CARE	1025
4	BAKERY/RESTAURANT	493
5	MOBILE FOOD TRUCK/DISPENSER	397
6	CHILDRENS FACILITY	353
7	nan	286
8	LIQUOR STORE	275
9	LONG TERM CARE FACILITY	185

4. 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, but also other parts of the country. 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.

5. Future Research Directions

The ratings for the business units were not given as part of the data set. If the rating for each of the establishments is give, then we can proceed with modelling the variables to predict the ratings.