



---

## FOOD INSPECTION IN ILLINOIS

---

An academic research project using IBM WATSON ANALYTICS cloud Platform



CIS – 5810-01 Healthcare Information Systems

OCTOBER 22, 2017

VRUNDA SHAH  
Professor Shilpa Balan

## A] Data Set URL:

**Food Inspection:** This data is related to inspection of food quality via various government bodies in Illinois.

**URL:** <https://data.cityofchicago.org/Health-Human-Services/Food-Inspections/4ijn-s7e5>

**Hashtags Used:** #foodinspection, #foodquality, #foodhealth, #foodsafety

## B] Data Description

### Food Inspections:

This dataset is related to food inspections that is carried in all the cities of Illinois. Inspection is not only done on food that is served to customers but it is also done on staff, owner. This inspection is performed by Department of Public Health's Food Protection Program using a standardized parameter mentioned below in detail. The result submitted by public health's food protection is later reviewed and approved by state of Illinois Licensed Environmental Health Practitioner (LEHP).

### Fields in dataset:

- **DBA:** Doing Business As, legal name of facility type such as restaurant, clubs, shelter etc.
- **AKA:** ‘Also known as’, other name of facility type like real life example you have a name and also nickname; so nickname in above scenario is AKA.
- **License Number:** Each facility type has a unique number which is used as a reference legally. If we take real life example then passport no or car no. In the same way each person who does business legally has been given license number so government can track them wherever and whenever needed.

- **Type of facility:** In which genre a particular person's business falls into. In our dataset which is related to food inspection has several categories as follows, 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 (combo, 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, psalteria, school, shelter, tavern, social club, wholesaler, or Wrigley Field Rooftop.
- **Risk Category of facility:** It categorize risk associated for each facility type.

1 – High risk

2 – Medium risk

3 – Low risk

- **Street address, city, state, zip code of facility:** This provides complete address with location of the person who runs particular facility type.
- **Inspection Date:** Date on which inspection was scheduled.
- **Inspection Type:** There are several types of inspection that can be take place
  - i) **Canvass:** It is the most common type of Inspection type which is held to check the establishment of an owner
  - ii) **License:** This inspection is done to verify everything related to establishment and receive a license number to operate legally.

- iii) **Consultation:** This inspection is done with the due request of owner prior establishment of one of the facility type mentioned above.
- iv) **Complaint:** This inspection is done related to complaint against establishment.
- v) **Task-force:** This inspection is related to bar and tavern.
- vi) **Suspect food poisoning:** This inspection is done when customers files a complaint against establishment of falling ill after eating food at establishment location like restaurants.

**Note: By default, re-inspection of all the six categories mentioned above can be done.**

- **Results:** It is the final outcome after inspection has been completed by the authority. Final outcome is of 3 types:
  - i) Pass
  - ii) Pass with conditions
  - iii) Fail
- **Violations:** If an establishment fail to follow any rule or regulation then it is considered as violation for not following government norms. An establishment can receive one or more 45 distinct violations (1-44 and 70).

## C] Data Refinement

## **1) Data Cleaning in Excel:**

### **i) Spelling Mistake**

In the Inspection Type column ‘Canvass’ spelling is misspelled to ‘Canvas’. We have replaced all the misspelled Inspection type with the original Inspection type ‘Canvass Spelling’.

The screenshot shows an Excel spreadsheet titled "Food\_Inspections (2) - Excel". The data consists of columns: Facility Type, Risk, Address, City, State, Zip, Inspection Date, and Results. A dropdown menu is open over the "Inspection Type" column, specifically over the cell containing "Canvas". A filter dialog box is displayed, showing several options under the "Canvas" category, including "CANVASS FOR RIB FEST", "CANVASS RE INSPECTION OF CLOSE U", and "CANVASS Re-Inspection". The "OK" button is visible at the bottom right of the dialog.

Facility Type	Risk	Address	City	State	Zip	Inspection Date	Results	
1 License	Facility Type	Risk 2 (Medium)	4359 N PULASKI RD	CHICAGO	IL	60641	10/20/2013	Pass w/
2 2009262 Restaurant	Risk 1 (High)	58 E ONTARIO ST	CHICAGO	IL	60611	10/20/2013	Pass	
3 2536006 Restaurant	Risk 1 (High)	58 E ONTARIO ST	CHICAGO	IL	60611	10/20/2013	Pass	
4 2536005 Restaurant	Risk 1 (High)	7435 W TALCOTT AVE	CHICAGO	IL	60631	10/20/2013	Pass	
5 205217 Long Term Care	Risk 1 (High)	7020 W GRAND AVE	CHICAGO	IL	60707	10/20/2013	Pass	
6 2172667 Restaurant	Risk 1 (High)	3759 N DAMEN AVE	CHICAGO	IL	60618	10/20/2013	Pass	
7 1139359 Restaurant	Risk 2 (Medium)	4420 -04430 W MONTROSE AVE	CHICAGO	IL	60641	10/20/2013	Pass	
8 28863 Restaurant	Risk 3 (Low)	8201 S RACINE AVE	CHICAGO	IL	60620	10/20/2013	Pass	
9 2319946 Grocery Store	Risk 3 (Low)	10200 S CENTRAL PARK AVE	CHICAGO	IL	60655	10/17/2013	Pass	
10 2464642 Grocery Store	Risk 2 (Medium)	4010 N WESTERN AVE	CHICAGO	IL	60618	10/20/2013	Pass	
11 86763 Restaurant	Risk 2 (Medium)	2604 W LAWRENCE AVE	CHICAGO	IL	60625	10/20/2013	Pass	
12 1542518 Restaurant	Risk 1 (High)	7262 W PETERSON AVE	CHICAGO	IL	60631	10/20/2013	Pass	
13 2116608 Restaurant	Risk 1 (High)	5424 N Oketo (7400W)	CHICAGO	IL	60656	10/20/2013	Pass	
14 24771 School	Risk 1 (High)	3607 W FULLERTON AVE	CHICAGO	IL	60647	10/20/2013	Cor	
15 1717867 Grocery Store	Risk 1 (High)	3333 N KIMBALL AVE	CHICAGO	IL	60618	10/20/2013	Sho	
16 28024 Grocery Store	Risk 1 (High)	6400 S CICERO AVE	CHICAGO	IL	60638	10/20/2013	Lice	
17 2563501 Restaurant	Risk 1 (High)	1310 E 53RD ST	CHICAGO	IL	60615	10/20/2013	Lice	
18 2559750 Restaurant	Risk 1 (High)	305-309 N CICERO AVE	CHICAGO	IL	60644	10/18/2013	Lice	
19 2369187 Children's Services Facility	Risk 1 (High)	180 N MORGAN ST	CHICAGO	IL	60607	10/20/2013	Car	
20 2452240 Restaurant	Risk 3 (Low)	5215 S HARPER AVE	CHICAGO	IL	60616	10/20/2013	License	
21 2542554 Restaurant	Risk 2 (Medium)	5206 N WESTERN AVE	CHICAGO	IL	60625	10/20/2013	Canvass	
22 1445472 Grocery Store	Risk 1 (High)	2855-2859 N HARLEM AVE	CHICAGO	IL	60707	10/20/2013	Canvass	
23 1766467 Bakery	Risk 2 (Medium)	3401 W BELMONT AVE	CHICAGO	IL	60618	10/20/2017	Short Form Complaint	
24 1476422 Restaurant	Risk 1 (High)	5737 W IRVING PARK RD	CHICAGO	IL	60634	10/20/2017	Complaint Re-Inspection	
25 76170 Restaurant	Risk 1 (High)	5957 W BELMONT AVE	CHICAGO	IL	60634	10/20/2017	Canvass Re-Inspection	
26 1473964 Restaurant	Risk 1 (High)	1365-67 W ERIE ST	CHICAGO	IL	60642	10/19/2017	Canvass	
27 2437391 Restaurant	Risk 2 (Medium)	2110 N MILWAUKEE AVE	CHICAGO	IL	60647	10/19/2017	Canvass	
28 2483009 Grocery Store	Risk 3 (Low)	3206 N MAJOR AVE	CHICAGO	IL	60634	10/19/2017	Canvass	
29 2412054 Grocery Store	Risk 1 (High)	3457-3459 W LAWRENCE AVE	CHICAGO	IL	60625	10/19/2017	Out of E	
30 30963 Restaurant	Risk 2 (Medium)	5725 W BELMONT AVE	CHICAGO	IL	60634	10/19/2017	Pass	
31 2560754 Restaurant							Pass	

**ii) Misfielded Values with long description (Implementation of Correction of Data)**

In the Risk column risk values are mentioned in the long descriptive format. Originally it should

be 1 – 3.

1 – Represents High risk

2 – Represents Medium risk

3 – Represents Low risk

For the purpose of data cleaning I have replaced long descriptive value to numerical values as

mentioned in below screenshot.

Food Inspections (1)													
AKA Name	C	D	E	F	G	H	I	J	K	L	M	N	O
2015 SEE THRU CHINESE KITCHEN		License#	Facility Type	Risk	Address	City	Sta	Zip	Inspection Da	Inspection T			
SHRIMP'S FISH CHICKEN		2463164	Restaurant	1	409 E 79TH ST	CHICAGO	IL	60619	10/17/2017	Complaint			
BAR CARGO		2549497	Restaurant	2	3355 S COTTAGE GROVE AVE	CHICAGO	IL	60653	10/17/2017	License Re-Ir			
BEST BBQ II		2535322	Restaurant	3	405-607 N WELLS ST	CHICAGO	IL	60619	10/17/2017	License Re-Ir			
THE BEST LEARNING SCHOOL		2458404	Restaurant	4	1419 W 79TH ST	CHICAGO	IL	60620	10/17/2017	Canvass			
LA PETITE FOULE		1084600	Restaurant	5	1400 W 79TH ST	CHICAGO	IL	60620	10/17/2017	License Re-Ir			
EAT N RUN CHICKEN AND FISH		69466	Restaurant	6	1204 E 55TH ST	CHICAGO	IL	60619	10/17/2017	Canvass			
GEORGIA'S FOOD DEPOT		2442368	Restaurant	7	1235 S ASHLAND AVE	CHICAGO	IL	60620	10/17/2017	Canvass			
KIKUYA RESTAURANT		2419039	Restaurant	8	7525 HALSTED ST	CHICAGO	IL	60621	10/17/2017	Canvass			
JOLLY PUMPKIN PIZZERIA & BREWERY		2487634	Restaurant	9	1222 W 63RD ST	CHICAGO	IL	60621	10/17/2017	Canvass			
COLUMBIA COLLEGE / FRANNIES' CAFE		1546681	Restaurant	10	6001 E 55TH ST	CHICAGO	IL	60616	10/17/2017	Canvass			
SNAIL THAI CUISINE		2542553	Restaurant	11	2115 S HARPER AVE	CHICAGO	IL	60616	10/17/2017	License			
DS TECUILA CO.		2551098	Restaurant	12	1045 W WABASH AVE	CHICAGO	IL	60605	10/17/2017	License Re-Ir			
TOP LAND		1617795	Restaurant	13	849 E 55TH ST	CHICAGO	IL	60615	10/17/2017	Canvass			
THE BEECH OF LOVE AND FAITH		2126030	Children's Services Facility	14	3350-3352 N HALSTED ST	CHICAGO	IL	60657	10/17/2017	License			
CUTIE MIES' KITCHEN		1397000	Restaurant	15	4600 W ADDISON ST	CHICAGO	IL	60634	10/17/2017	License Re-Ir			
LAV & BAKERY, INC		2437784	Restaurant	16	117 W 79TH ST	CHICAGO	IL	60620	10/17/2017	Canvass			
JOLLY PUMPKIN PIZZERIA & BREWERY		2505054	Bakery	17	75 W 79TH ST	CHICAGO	IL	60620	10/17/2017	Canvass Re-Ir			
FOODA SWITCHING STATIONS		2542552	Restaurant	18	2038 S KEDOZIE AVE	CHICAGO	IL	60623	10/17/2017	License Re-Ir			
KENNEDY KING COLLEGE		2559668	Restaurant	19	2115 S HARPER AVE	CHICAGO	IL	60616	10/17/2017	License			
ST XAVIER COUGAR FUEL (SHANNON CENTER)		2419579	Restaurant	20	1400 S LA SALLE ST	CHICAGO	IL	60605	10/17/2017	License Re-Ir			
LITTLE KIDS VILLAGE LEARNING		2464645	Restaurant	21	543 S HALSTED ST	CHICAGO	IL	60621	10/17/2017	Canvass			
TBD		2418331	Daycare Above and Under 2 Years	22	8000 W 103RD ST	CHICAGO	IL	60655	10/17/2017	Canvass			
TBD		2551380	Restaurant	23	1732 S 63RD ST	CHICAGO	IL	60621	10/17/2017	License			
NICKY'S RESTAURANT		2551384	Restaurant	24	848N W HARLEM AVE	CHICAGO	IL	60634	10/17/2017	License			
GLORIA'S KOREAN CHILD DEVELOPMENT CENTER		49913	Restaurant	25	848N W HARLEM AVE	CHICAGO	IL	60634	10/17/2017	License			
CONTRIBUCIONES LA HACIENDA NO 1 INC		2314038	Daycare Above and Under 2 Years	26	7029 S WESTERN AVE	CHICAGO	IL	60620	10/17/2017	Canvass			
COLUMBIA COLLEGE / FRANNIES' CAFE		2424206	Restaurant	27	1312 S RACINE AVE	CHICAGO	IL	60608	10/16/2017	License			
HYATT PLACE CHICAGO SOUTH		2323869	Restaurant	28	6001 W 63RD ST	CHICAGO	IL	60605	10/16/2017	License			
LITTLE KIDS VILLAGE LEARNING		2315822	Daycare (2 - 5 Years)	29	2045 S WABASH AVE	CHICAGO	IL	60605	10/17/2017	License			
PARADISE GROCERY INC.		2517776	Grocery Store	30	1255 S HARPER AVE	CHICAGO	IL	60615	10/17/2017	Complaint			
POTBELLY SANDWICH SHOP		2560939	Restaurant	31	8556 W 71ST ST	CHICAGO	IL	60629	10/17/2017	License			
Daallo Meat & Grocery		2534897	Restaurant	32	8101 S COTTAGE GROVE AVE	CHICAGO	IL	60619	10/16/2017	Complaint			
POKEWORKS		2013707	Grocery Store	33	1230 W TAYLOR ST	CHICAGO	IL	60607	10/16/2017	License Re-Ir			
A10		2560710	Restaurant	34	1756 E 111TH ST	CHICAGO	IL	60628	10/16/2017	License			
10'S JAMAICAN TROPICAL CAFE		2241029	Restaurant	35	2742 W Devon AVE	CHICAGO	IL	60659	10/16/2017	Canvass Re-I			
		3451675	Restaurant	36	79 E MADISON ST	CHICAGO	IL	60602	10/16/2017	License			
				37	1460-1466 E 53RD ST	CHICAGO	IL	60615	10/16/2017	Canvass			
				38	1810 S ASHLAND AVE	CHICAGO	IL	60620	10/16/2017	Canvass			

## IBM Watson Cleaning

### iii) Data Group:

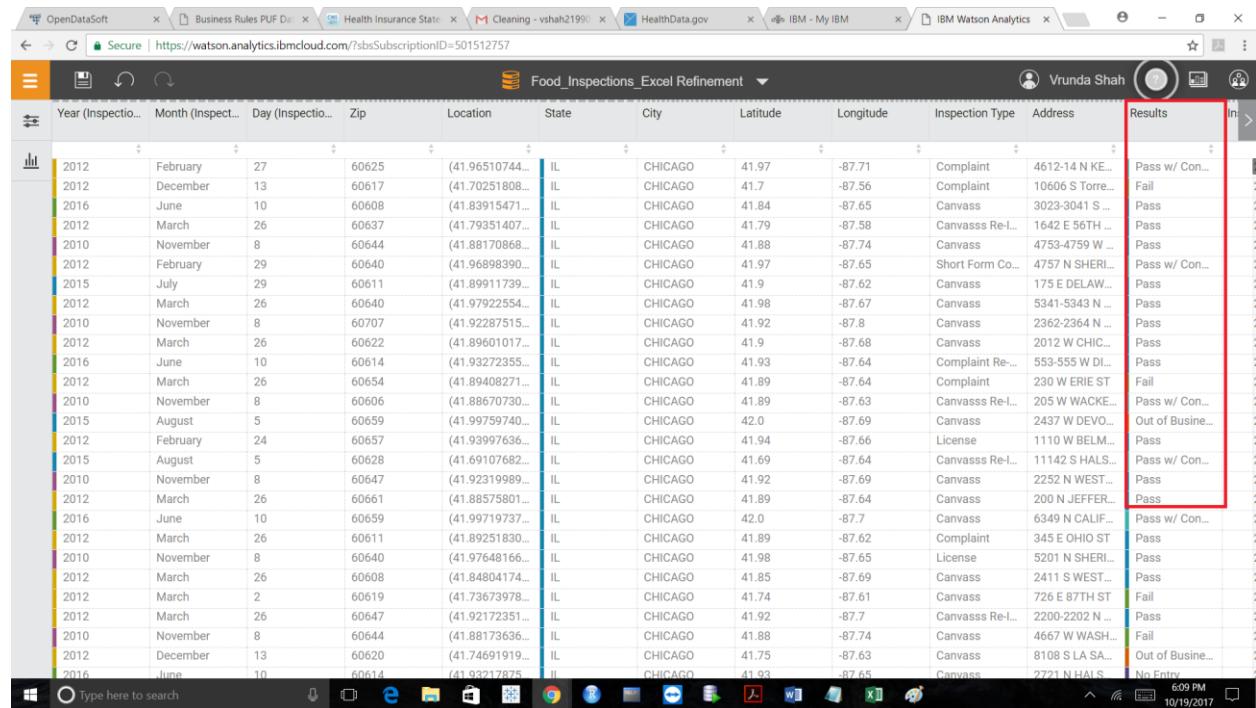
Results field should be classified into three values (Pass, Fail, Pass with condition). But in our original data it had several extra values such as '**No entry, Not Ready, Out of Business**' which falls under Fail status, so we have classified data in group like Pass, Fail and Pass with Condition.

#### Details:

Pass – Pass

Pass with condition - Pass with condition

Fail – Fail, No entry, Not Ready, Out of Business



Year (Inspectio...	Month (Inspectio...	Day (Inspectio...	Zip	Location	State	City	Latitude	Longitude	Inspection Type	Address	Results
2012	February	27	60625	(41.96510744...	IL	CHICAGO	41.97	-87.71	Complaint	4612-14 N KE...	Pass w/ Con...
2012	December	13	60617	(41.70251808...	IL	CHICAGO	41.7	-87.56	Complaint	10606 S Torre...	Fail
2016	June	10	60608	(41.83915471...	IL	CHICAGO	41.84	-87.65	Canvass	3023-3041 S ...	Pass
2012	March	26	60637	(41.79351407...	IL	CHICAGO	41.79	-87.58	Canvass Re-L...	1642 E 56TH ...	Pass
2010	November	8	60644	(41.88170868...	IL	CHICAGO	41.88	-87.74	Canvass	4753-4759 W ...	Pass
2012	February	29	60640	(41.96898390...	IL	CHICAGO	41.97	-87.65	Short Form Co...	4757 N SHERL...	Pass w/ Con...
2015	July	29	60611	(41.89911739...	IL	CHICAGO	41.9	-87.62	Canvass	175 E DELAW...	Pass
2012	March	26	60640	(41.97922554...	IL	CHICAGO	41.98	-87.67	Canvass	5341-5343 N ...	Pass
2010	November	8	60707	(41.92287515...	IL	CHICAGO	41.92	-87.8	Canvass	2362-2364 N ...	Pass
2012	March	26	60622	(41.89601017...	IL	CHICAGO	41.9	-87.68	Canvass	2012 W CHIC...	Pass
2016	June	10	60614	(41.93272355...	IL	CHICAGO	41.93	-87.64	Complaint Re...	553-555 W DI...	Pass
2012	March	26	60654	(41.89408271...	IL	CHICAGO	41.89	-87.64	Complaint	230 W ERIE ST	Fail
2010	November	8	60606	(41.88670730...	IL	CHICAGO	41.89	-87.63	Canvass Re-L...	205 W WAKEF...	Pass w/ Con...
2015	August	5	60659	(41.99759740...	IL	CHICAGO	42.0	-87.69	Canvass	2437 W DEVO...	Out of Busine...
2012	February	24	60657	(41.93997636...	IL	CHICAGO	41.94	-87.66	License	1110 W BELM...	Pass
2015	August	5	60628	(41.69107682...	IL	CHICAGO	41.69	-87.64	Canvass Re-L...	11142 S HALS...	Pass w/ Con...
2010	November	8	60647	(41.92319989...	IL	CHICAGO	41.92	-87.69	Canvass	2252 N WEST...	Pass
2012	March	26	60661	(41.88575801...	IL	CHICAGO	41.89	-87.64	Canvass	200 N JEFFER...	Pass
2016	June	10	60659	(41.99719737...	IL	CHICAGO	42.0	-87.7	Canvass	6349 N CALIF...	Pass w/ Con...
2012	March	26	60611	(41.89251830...	IL	CHICAGO	41.89	-87.62	Complaint	345 E OHIO ST	Pass
2010	November	8	60640	(41.97648166...	IL	CHICAGO	41.98	-87.65	License	5201 N SHERL...	Pass
2012	March	26	60608	(41.84804174...	IL	CHICAGO	41.85	-87.69	Canvass	2411 S WEST...	Pass
2012	March	2	60619	(41.73673978...	IL	CHICAGO	41.74	-87.61	Canvass	726 E 87TH ST	Fail
2012	March	26	60647	(41.92172351...	IL	CHICAGO	41.92	-87.7	Canvass Re-L...	2200-2202 N ...	Pass
2010	November	8	60644	(41.88173636...	IL	CHICAGO	41.88	-87.74	Canvass	4667 W WASH...	Fail
2012	December	13	60620	(41.74691919...	IL	CHICAGO	41.75	-87.63	Canvass	8108 S LA SA...	Out of Busine...
2016	June	10	60614	(41.93217875...	IL	CHICAGO	41.93	-87.65	Canvass	2721 N HALS...	No Entry

	Address	Results	Inspection Date	Inspection ID	License # Does not equal...	Violations	Facility Type	AKA Name	DBA Name	Risk	Risk_Avg	Result
	4612-14 N KE...	Pass w/ Con...	2012-02-27	670827	1869313	3. POTENTIAL...	Restaurant	MI LINDO PAL...	MI LINDO PAL...	1	1.39	Pass with co...
	10606 S Torre...	Fail	2012-12-13	1154755	1986474	18. NO EVIDE...	Restaurant	Gus's Pizza	Gus's Pizza	1	1.39	Fail
	3023-3041 S ...	Pass	2016-06-10	1932546	2093849	32. FOOD AND...	Grocery Store	CERMACK FRE...	CERMACK FRE...	1	1.39	Pass
Re-L...	1642 E 56TH ...	Pass	2012-03-26	1154216	1908007	24. DISH WAS...	Restaurant	PICCOLO MO...	PICCOLO MO...	1	1.39	Pass
n Co...	4753-4759 W ...	Pass	2010-11-08	456490	1493464	38. VENTILATI...	Daycare (2 - 6 ...	FIRST START ...	FIRST START ...	1	1.39	Pass
	4757 N SHERI...	Pass w/ Con...	2012-02-29	670984	2069675	21. * CERTIFI...	Restaurant	PAPA RAY'S PL...	PAPA RAY'S PL...	1	1.39	Pass with co...
	175 E DELAWA...	Pass	2015-07-29	1447917	1822782	36. LIGHTING...	Grocery Store	POTASH GOU...	POTASH GOU...	1	1.39	Pass
	5341-5343 N ...	Pass	2012-03-26	1106206	1245300	32. FOOD AND...	Restaurant	Calo Lounge	Calo Lounge, I...	1	1.39	Pass
	2362-2364 N ...	Pass	2010-11-08	456485	1843726	34. FLOORS: C...	Restaurant	BISTRO GRAND	BISTRO GRAND	1	1.39	Pass
	2012 W CHIC...	Pass	2012-03-26	680026	1477940	33. FOOD AND...	Restaurant	SUNRISE CAF...	SUNRISE CAF...	1	1.39	Pass
Re-L...	553-555 W DL...	Pass	2016-06-10	1932554	2391406	29. PREVIOUS...	Restaurant	JINYA RAMEN...	JINYA RAMEN...	1	1.39	Pass
	230 W ERIE ST	Fail	2012-03-26	1134209	2060219	34. FLOORS: C...	Restaurant	UNION SUSHI ...	UNION SUSHI ...	1	1.39	Fail
Re-L...	205 W WACKE...	Pass w/ Con...	2010-11-08	343486	2008992	21. * CERTIFI...	Restaurant	ALL STARS O...	ALL STARS O...	1	1.39	Pass with co...
	2437 W DEVO...	Out of Busine...	2015-08-05	1506619	2373988		Restaurant	DIVINE GARD...	DIVINE GARD...	1	1.39	Fail
Re-L...	1110 W BELM...	Pass	2012-02-24	670771	44304		Children's Serv...	THRESHOLDS	THRESHOLDS	1	1.39	Pass
	11142 S HALS...	Pass w/ Con...	2015-08-05	1561865	1822894	14. PREVIOUS...	Restaurant	BLUE SOUL RE...	BLUE SOUL RE...	1	1.39	Pass with co...
	2252 N WEST...	Pass	2010-11-08	456484	1820435	31. CLEAN M...	Restaurant	CAFE BOLERO	CAFE BOLERO	1	1.39	Pass
	200 N JEFFER...	Pass	2012-03-26	1184224	1823750	33. FOOD AND...	Restaurant	JIMMY JOHN'S	JIMMY JOHN'S	1	1.39	Pass
	6349 N CALIF...	Pass w/ Con...	2016-06-10	1932555	3953	41. PREMISES...	Restaurant	TEL AVIV RES...	TEL AVIV RES...	1	1.39	Pass with co...
	345 E OHIO ST	Pass	2012-03-26	1138212	1804575	33. FOOD AND...	Restaurant	D4 IRISH PUB	D4 IRISH PUB	1	1.39	Pass
	5201 N SHERI...	Pass	2010-11-08	456472	2065024		Restaurant	SUSHI	AFC SUSHI @ ...	1	1.39	Pass
	2411 S WEST...	Pass	2012-03-26	1182215	44026	31. CLEAN M...	Restaurant	TAQUERIA TA...	TAQUERIA TA...	1	1.39	Pass
	726 E 87TH ST	Fail	2012-03-02	671049	2049913	6. HANDS WA...	Restaurant	SUBWAY	SUBWAY	1	1.39	Fail
Re-L...	2200-2202 N ...	Pass	2012-03-26	1214282	1575555	18. NO EVIDE...	Restaurant	PANCHO III C...	PANCHO III C...	1	1.39	Pass
	4667 W WASH...	Fail	2010-11-08	456470	1902909		Restaurant	LAYMON'S	LAYMON'S	1	1.39	Fail
	8108 S LAS SA...	Out of Busine...	2012-12-13	1151022	1931		School	CECIL PARTEE...	TED LENART ...	1	1.39	Fail
	2721 N HALS...	No Entry	2016-06-10	1922552	32782		Restaurant	TRINITY BAR	TRINITY BAR	1	1.39	Fail

#### iv) Removal of Blank Values

There were several missing values in License # column. For the quality of data we have filtered out blank values as shown in below screenshots.

OpenDataSoft Business Rules PUF Data Health Insurance State Cleaning - vshah21999 HealthData.gov IBM - My IBM IBM Watson Analytics

Secure | https://watson.analytics.ibmcloud.com/?sbsSubscriptionID=50151279

Vrunda Shah

**Food\_Inspections\_Excel Refinement**

Address Results Inspection Date Inspection ID License # Violations Facility Type AKA Name DBA Name Risk Risk\_Avg Result

121 W HUBBA... Pass 2012-05-03 1134264 License # 33188 of 33188 OOD IN O... Restaurant RED VIOLET RED VIOLET 1 1.39 Pass

2643 N MILW... Pass w/ Con... 2014-06-26 1474095 CERTIFI... Restaurant HALF ITALIAN HALF ITALIAN 1 1.39 Pass

10740 S CALH... Pass 2012-05-03 1118295 LOORS: C... School BRIGHT ELEM... 1 1.39 Pass

7601 S CICER... Pass 2013-12-12 1151878 Set Condition OOD AND... Restaurant SMOKEHOUS... SMOKEHOUS... 1 1.39 Pass

Re-1010 N NOBL... Pass 2012-05-03 1202263 OEVIDE... School NOBLE STREET SODEXO AME... 1 1.39 Pass

936 W ADDIS... Pass w/ Con... 2014-06-26 1474092 OOD AND... Restaurant EL BURRITO EL BURRITO 1 1.39 Pass

4400-4404 N ... Fail 2012-05-03 1214645 Restaurant ANNA MARIA ... ANNA MARIA ... 1 1.39 Fail

4131 N SHERI... Fail 2013-09-04 1106818 DOD AND... ASSISTED LIVI... EDEN SUPPO... 1 1.39 Fail

4356 W ARMI... Pass 2012-05-03 1214613 ENTILATI... Restaurant KIM LONG RE... KIM LONG RE... 1 1.39 Pass

238 W 33RD ST Pass 2014-06-26 1474091 OOD AND... Restaurant TURTLES TURTLES 1 1.39 Pass

6731 N STRE... Fail 2012-05-03 1194271 OEVIDE... Restaurant CAFE TOUCHE CAFE TOUCHE 1 1.39 Fail

333 S State ST Fail 2013-12-12 1115181 ND'S WA... Restaurant ERIK'S DELI ERIK'S DELI 1 1.39 Fail

1541 W BRYN ... Pass 2012-05-03 1214609 Restaurant LITTLE BAD W... LITTLE BAD W... 1 1.39 Pass

1371 W CHIC... Pass 2014-06-26 1474090 LOORS: C... Restaurant BOW & STERN... BOW & STERN... 1 1.39 Pass

2964 N LINCO... Out of Busine... 2012-05-03 1198250 Restaurant TA TONG THA... TA TONG THA... 1 1.39 Fail

Re-5456 N HARL... Pass 2013-09-04 1360664 5852857 UTSIDE ... Restaurant WENDY'S WENDY'S #00... 1 1.39 Pass

2964 N LINCO... Out of Busine... 2012-05-03 1198249 Restaurant Rice Bistro Rice Bistro 1 1.39 Fail

1 N LA SALLE ... Pass 2014-06-26 1372609 DOD AND... Restaurant POTBELLY SA... POTBELLY SA... 1 1.39 Pass

2964 N LINCO... Pass 2012-05-03 1198248 ✓ 33389955 LOEVIDE... Restaurant RICE BISTRO RICE BISTRO 1 1.39 Pass

302 E ILLINOI... Pass 2013-12-12 1276012 ✓ (blank) DOD IN O... Restaurant BELLWETHER BELLWETHER 1 1.39 Pass

1401 S MICHL... Fail 2012-05-03 1026293 ✓ (blank) OEVIDE... Restaurant THE CHICAGO... THE CHICAGO... 1 1.39 Fail

1824 S WOOD... Pass 2014-06-26 1474088 Invert IALLS, CE... Golden Diner HARRISON PA... HARRISON PA... 1 1.39 Pass

1414 W 87TH ... Fail 2012-05-03 1214610 CILITIES... Daycare Abov... SWEET PEA A... SWEET PEA A... 1 1.39 Fail

Re-3326 W BELM... Pass 2013-09-04 1360665 Clear O SMOKL... Restaurant NEW BUFFET... NEW BUFFET... 1 1.39 Pass

3732 W 59TH ... Pass 2012-05-03 1227236 DOD AND... Restaurant BIG TONY'S PI... BIG TONY'S PI... 1 1.39 Pass

11142 S HALS... Pass w/ Con... 2014-06-26 1474093 1822894 32. FOOD AND... Restaurant BLUE SOUL RE... BLUE SOUL RE... 1 1.39 Pass

6124 S ASHLA... Pass 2012-05-03 1214604 2152117 36. LIGHTING... Restaurant KRISPY KRUN KRISPY KRUN 1 1.39 Pass

612 PM 10/19/2017

OpenDataSoft Business Rules PUF Data Health Insurance State Cleaning - vshah21999 HealthData.gov IBM - My IBM IBM Watson Analytics

Secure | https://watson.analytics.ibmcloud.com/?sbsSubscriptionID=50151279

Vrunda Shah

**Food\_Inspections\_Excel Refinement**

Address Results Inspection Date Inspection ID License # Does not equal... Violations Facility Type AKA Name DBA Name Risk Risk\_Avg Result

4612-14 N KE... Pass w/ Con... 2012-02-27 670827 License # Does not equal... TENTRAL... Restaurant MI LINDO PAL... MI LINDO PAL... 1 1.39 Pass

10606 S Torre... Fail 2012-12-13 1154755 Does not equal... OEVIDE... Restaurant Gusi's Pizza Gusi's Pizza 1 1.39 Fail

3023-3041 S ... Pass 2016-06-10 193254 Cancel Done DOD AND... Grocery Store CERMACK FRE... CERMACK FRE... 1 1.39 Pass

Re-1642 E 56TH... Pass 2012-03-26 1154211 Does not equal... ISHWAS... Restaurant PICCOLO MO... PICCOLO MO... 1 1.39 Pass

4753-4759 W ... Pass 2010-11-08 4564900 Does not equal... ENTITATL... Daycare (2 - 6 ... FIRST START... FIRST START... 1 1.39 Pass

n Co... 4757 N SHERI... Pass w/ Con... 2012-02-29 670984 (blank) CERTIFI... Restaurant PAPA RAY'S PI... PAPA RAY'S PI... 1 1.39 Pass

175 E DELAWA... Pass 2015-07-29 1447911 Does not equal... GHTING... Grocery Store POTASH GOU... POTASH GOU... 1 1.39 Pass

5341-5343 N ... Pass 2012-03-26 1106200 Add Condition LOORS: C... Restaurant Calo Lounge Calo Lounge, I... 1 1.39 Pass

2362-2364 N ... Pass 2010-11-08 4564857 Clear Condition DOD AND... Restaurant BISTRO GRAND BISTRO GRAND 1 1.39 Pass

2012 W CHIC... Pass 2012-03-26 680026 REVIEWS... Restaurant SUNRISE CAF... SUNRISE CAF... 1 1.39 Pass

Re-553-555 W DL... Pass 2016-06-10 1932555 Does not equal... 34. FLOORSC... Restaurant JINYA RAMEN JINYA RAMEN... 1 1.39 Pass

230 W ERIE ST Fail 2012-03-26 1134201 (blank) 2060219 34. FLOORSC... Restaurant UNION SUSHI ... UNION SUSHI ... 1 1.39 Fail

Re-205 W WACKE... Pass w/ Con... 2010-11-08 3434866 2008992 21.\* CERTIFI... Restaurant ALL STARS O... ALL STARS O... 1 1.39 Pass

2437 W DEVO... Out of Busine... 2015-08-05 1506611 2373988 1822894 14. PREVIOUSU... Restaurant DIVINE GARD... DIVINE GARD... 1 1.39 Fail

1110 W BELM... Pass 2012-02-24 6707711 44304 1820435 14. PREVIOUSU... Children's Serv... THRESHOLDS THRESHOLDS 1 1.39 Pass

Re-11142 S HALS... Pass w/ Con... 2015-08-05 1561866 1820435 31. CLEAN M... Restaurant BLUE SOUL RE... BLUE SOUL RE... 1 1.39 Pass

2252 N WEST... Pass 2010-11-08 4564844 1820435 31. CLEAN M... Restaurant CAFE BOLERO CAFE BOLERO 1 1.39 Pass

200 N JEFFER... Pass 2012-03-26 1184224 1823750 33. FOOD AND... Restaurant JIMMY JOHN'S JIMMY JOHN'S 1 1.39 Pass

6349 N CALIF... Pass w/ Con... 2016-06-10 1932555 3953 41. PREMISES... Restaurant TEL AVIV RES... TEL AVIV RES... 1 1.39 Pass

345 E OHIO ST Pass 2012-03-26 1138212 1804575 33. FOOD AND... Restaurant D4 IRISH PUB ... D4 IRISH PUB ... 1 1.39 Pass

5201 N SHERI... Pass 2010-11-08 4564727 2065024 Restaurant SUSHI AFC SUSHI @ ... 1 1.39 Pass

2411 S WEST... Pass 2012-03-26 1182215 44026 31. CLEAN M... Restaurant TAQUERIA TA... TAQUERIA TA... 1 1.39 Pass

726 E 87TH ST Fail 2012-03-02 671049 2049913 6. HANDS WA... Restaurant SUBWAY SUBWAY 1 1.39 Fail

Re-2200-2202 N ... Pass 2012-03-26 1214282 1575555 18. NO EVIDE... Restaurant PANCHO III C... PANCHO III C... 1 1.39 Pass

4667 W WASH... Fail 2010-11-08 4564707 1902909 Restaurant LAYMONS LAYMONS 1 1.39 Fail

8108 S LA SA... Out of Busine... 2012-12-13 1151022 1931 School CECIL PARTEE... TED LENART ... 1 1.39 Fail

2721 N HALS... No Entry 2016-06-10 1932552 33783 Restaurant TRINITY BAR TRINITY BAR 1 1.39 Fail

610 PM 10/19/2017

## 2) a) Data Grouping

To get more insight about the factor affecting ‘Result’ from year 2010 – 2017 , we have divided Month(Inspection Date) into 4 quarters, to get insight result about each quarter and further compare each quarters result with other to identify more factors which influences target Result.

1<sup>st</sup> quarter – January, February, March

2<sup>nd</sup> Quarter - April, May, June

3<sup>rd</sup> Quarter - July, August, September

4<sup>th</sup> Quarter - October, November, December

The screenshot shows a web-based data analysis tool. At the top, there are several tabs: OpenDataSoft, Business Rules PUF Da..., Health Insurance State..., Cleaning - vshah2199..., HealthData.gov, IBM - My IBM, and IBM Watson Analytics. The main area is titled "Food\_Inspections\_Excel Refinement". On the left, there's a sidebar with a tree view showing years from 2010 to 2016. The main table has columns: Year (Inspectio..., Month (Inspectio..., Day (Inspectio..., Zip, Location, State, City, Latitude, Longitude, Inspection Type, Address, Results, and In. A red box highlights the "Month (Inspection Date)" dropdown menu, which is open and shows the months from January to October. Below the dropdown are buttons for "Invert" and "Clear". The table contains data for Chicago locations with various inspection types like Canvass, License, and Complaint, and results like Pass and Fail.

The screenshot shows a Microsoft Excel-like interface for data analysis. A modal dialog box titled "Data group" is open, allowing the user to create hierarchical groups based on inspection dates. The dialog tree includes categories for "Month (Inspection Date)", "Quarterly Report", "3rd Quarter", and "4th Quarter". The main table below contains columns for Year, Month, Day, Zip, Location, State, City, Latitude, Longitude, Inspection Type, Address, Results, and Violations. The "Violations" column is highlighted with a red border.

Year (Inspection Date)	Month (Inspection Date)	Day (Inspection Date)	Zip	Location	State	City	Latitude	Longitude	Inspection Type	Address	Results	Violations
2015	July	30	60651	(41.894814, -87.666667)	IL	CHICAGO	41.89	-87.67	Canvass	5215 W CHIC...	Fail	
2013	September	12	60659	(41.98639, -87.666667)	IL	CHICAGO	41.99	-87.67	Canvass	2021 S MORG...	Pass	
2015	July	30	60657	(41.94102195, -87.666667)	IL	CHICAGO	41.94	-87.66	License	2456 W 47TH ...	Pass	
2016	June	17	60622	(41.90513175, -87.666667)	IL	CHICAGO	41.91	-87.67	License	2132-2134 W ...	Fail	
2015	July	30	60615	(41.79955945, -87.666667)	IL	CHICAGO	41.8	-87.6	Canvass	812 W VAN B...	Pass	
2013	September	12								6355 N MAPL...	Pass	
2015	July	30								2625 N HARL...	Pass	
2016	June	17								10951 S MICH...	No Entry	
2015	July	28								10400 S HALS...	Pass	
2010	May	20								2263 N LINCO...	Pass	
2015	July	30								1000 S LEAVI...	Pass	
2013	March	7								131 S DEARB...	Pass w/ Con...	
2015	July	30								5201 N SHERI...	Pass	
2013	September	12								3025 W CERM...	Pass	
2015	July	30								600 W CHICA...	Fail	
2016	June	17								1130 S WABA...	Pass w/ Con...	
2015	July	30								2925 W IRVIN...	No Entry	
2013	September	12								131 S CLINTO...	Pass	
2015	July	30								630 W HARRI...	Fail	
2016	June	17								540 N WELLS ...	Pass	
2014	April	1								6075 N MILWU...	Pass	
2013	September	12								7622 W BELM...	Pass	
2015	July	30								2462 W 47TH ...	Fail	
2013	March	7								5776-5778 N ...	Pass	
2015	July	30								3231-3233 N ...	Fail	
2013	September	12								1300 N ASHL...	Pass	
2014	April	1								1206 E 53RD ST	Pass	

## 2b) Hierarchy

In the below screenshot, we have created Hierarchy in which Result column(Top Level) -> Violations column (Bottom) to identify violations that can be taken into consideration during inspection and based on particular violation criticality we can classify end result of Inspection into 3 categories Pass, Pass with conditions and Fail.

Screenshot of the IBM Watson Analytics interface showing the 'Food\_Inspections\_Excel\_Refinement' dataset. A red box highlights the 'Violations Hierarchy' section, which displays a tree structure for violations. The hierarchy includes levels like 'Name', 'Violations Hierarchy', 'Levels', 'Result', and 'Violations'. Below the hierarchy, there are buttons for 'Add level', 'Delete hierarchy', and 'Cancel/Hierarchy/Done'.

Longitude	Inspection Type Does not contain..	Address Does not equal..	Results	Inspection Date	Inspection ID	License #	Violations Does not equal..	Facility Type Does not equal..	AKA Name Does not equal..	DBA Name Does not equal..	Risk	
-87.64	Complaint Re...	500 W MONR...	Pass	2015-10-20	1579168	1579365	29. PREVIOUS...	Restaurant	VENICE CAFE...	VENICE CAFE...		
-87.71	Complaint	6249 N MCCO...	Fail	2017-03-31	2009687	2510903	18. NO EVIDE...	Restaurant	NEW HONG K...	NEW HONG K...		
-87.61	Complaint	5758 S MARY...	Pass	2015-10-20	1386501	2374305	32. FOOD AN...	Restaurant	CAFE CENTRAL...	ARAMARK FO...		
-87.71	Canvas	3440 W MONT...	Pass	2016-06-01	1931928	2404664	38. VENTILATI...	Grocery Store	SAN FELIPE F...	SAN FELIPE F...		
-87.67	Complaint	7038 N CLAR...	Fail	2017-09-12	1955876	2293343	18. NO EVIDE...	Restaurant	A&T PANCAK...	A&T PANCAK...		
-87.81	Canvas	7148 N HARL...	Fail	2017-03-31	2009685	2179179	3. POTENTIAL...	Restaurant	TAQUERIA LO...	TAQUERIA LO...		
-87.67	License	4710 N RAVE...	Pass	2015-10-20	1579169	2398386	37. TOILET RO...	Restaurant	BAND OF BOH...	BAND OF BOH...		
-87.76	Canvas	5242 - 5244 W...	Pass	2016-06-01	1931914	2358956	35. WALLS, C...	Children's Serv...	RACHELS LEA...	RACHELS LEA...		
-87.7	Canvas	3139 W 63RD...	Fail	2017-09-12	2081925	26483	18. NO EVIDE...	Restaurant	TAQUERIA VIL...	TAQUERIA VIL...		
-87.65	Canvas	950 - 952 W AR...	Pass	2016-06-01	1931916	23581	1992138	34. FLOORS, C...	Restaurant	GIRL & THE G...	GIRL & THE G...	
-87.65	Canvas	7525 S Harvar...	Fail	2017-03-31	2009682	2032949	2204605	32. FOOD AN...	Long Term Care	ROSE ANGEL...	ROSE ANGEL...	
-87.7	Complaint Re...	6840 N SACR...	Pass	2017-03-31	2009682	1932480	31. CLEAN M...	Restaurant	ANTICA PIZZE...	ANTICA PIZZE...		
-87.6	Canvas	3440 W MONT...	Pass	2016-12-20	1978469	1943058	32. FOOD AN...	Grocery Store	ADAS MARKE...	ADAS MARKE...		
-87.64	Complaint Re...	7038 N CLAR...	Pass	2016-05-31	1931910	2334341	21. * CERTIFIE...	Restaurant	GEORGE'S CA...	GEORGE'S CA...		
-87.76	Canvas	4021 N MILW...	Pass w/ Con...	2016-05-31	1931910	2334341	18. NO EVIDE...	Restaurant	EXPLORER CA...	ARAMARK-FIE...		
-87.66	Canvas	7600 S ASHL...	Pass	2017-09-12	2081903	1933841	20. PREVIOUS...	Restaurant	MURKIN'S...	MURKIN'S...		
-87.66	Canvas	6920 S ASHL...	Pass w/ Con...	2017-03-31	2009681	2446619	11. ADEQUAT...	Restaurant	SUBWAY...	SUBWAY...		
-87.66	Canvas	9200 S ASHL...	Pass	2017-09-12	2081906	2308673	32. FOOD AN...	Restaurant	BUFFET CITY...	BUFFET CITY...		
-87.65	Canvas	1234 N HALS...	Pass	2016-05-31	1931908	2138602	32. FOOD AN...	Restaurant	JUST TURKEY...	JUST TURKEY...		
-87.64	Complaint	554 W ROOSE...	Fail	2017-09-12	2081896	1694688	2. FACILITIES...	Restaurant	EPPLES REST...	EPPLES REST...		
-87.72	Suspected Fo...	3545 W Fulton	Pass	2017-03-31	2009680	46471	38. VENTILATI...	School	AL RABY HIG...	AL RABY HIG...		

Screenshot of the IBM Watson Analytics interface showing the 'Food\_Inspections\_Excel\_Refinement' dataset. A red box highlights the 'Violations Hierarchy' section, which displays a tree structure for violations. The hierarchy includes levels like 'Name', 'Violations Hierarchy', 'Levels', 'Result', and 'Violations'. Below the hierarchy, there are buttons for 'Add level', 'Delete hierarchy', and 'Cancel/Hierarchy/Done'.

Longitude	Inspection Type Does not contain..	Address Does not equal..	Results	Inspection Date	Inspection ID	License #	Violations Does not equal..	Facility Type Does not equal..	AKA Name Does not equal..	DBA Name Does not equal..	Risk	Result	Quarterly Report	Violations Hier...
-87.64	Complaint Re...	500 W MONR...	Pass	2015-10-20	1579168	1579365	29. PREVIOUS...	Restaurant	VENICE CAFE...	VENICE CAFE...		Pass		Violations Hierarchy
-87.71	Complaint	6249 N MCCO...	Fail	2017-03-31	2009687	2510903	18. NO EVIDE...	Restaurant	NEW HONG K...	NEW HONG K...		Fail		
-87.61	Complaint	5758 S MARY...	Pass	2015-10-20	1386501	2374305	32. FOOD AN...	Restaurant	CAFE CENTRAL...	ARAMARK FO...		Pass		
-87.71	Canvas	3440 W MONT...	Pass	2016-06-01	1931928	2404664	38. VENTILATI...	Grocery Store	SAN FELIPE F...	SAN FELIPE F...		Pass		
-87.67	Complaint	7038 N CLAR...	Fail	2017-09-12	1955876	2293343	18. NO EVIDE...	Restaurant	A&T PANCAK...	A&T PANCAK...		Pass		
-87.81	Canvas	7148 N HARL...	Fail	2017-03-31	2009685	2179179	3. POTENTIAL...	Restaurant	TAQUERIA LO...	TAQUERIA LO...		Pass		
-87.67	License	4710 N RAVE...	Pass	2015-10-20	1579169	2398386	37. TOILET RO...	Restaurant	BAND OF BOH...	BAND OF BOH...		Pass		
-87.76	Canvas	5242 - 5244 W...	Pass	2016-06-01	1931914	2358956	35. WALLS, C...	Children's Serv...	RACHELS LEA...	RACHELS LEA...		Pass		
-87.7	Canvas	3139 W 63RD...	Fail	2017-09-12	2081925	26483	18. NO EVIDE...	Restaurant	TAQUERIA VIL...	TAQUERIA VIL...		Pass		
-87.65	Canvas	809-813 W RA...	Pass	2012-06-05	1184367	1992138	34. FLOORS, C...	Restaurant	GIRL & THE G...	GIRL & THE G...		Pass		
-87.78	Canvas	4200 N AUSTL...	Pass	2017-09-12	2081924	2204605	32. FOOD AN...	Long Term Care	ROSE ANGEL...	ROSE ANGEL...		Pass		
-87.67	Canvas	5663-65 N CL...	Pass	2016-06-01	1727319	1932480	31. CLEAN M...	Restaurant	ANTICA PIZZE...	ANTICA PIZZE...		Pass		
-87.8	Complaint	6165 N NORT...	Pass	2017-09-12	2081921	2354088	32. FOOD AN...	Grocery Store	ADAS MARKE...	ADAS MARKE...		Pass		
-87.76	Canvas	3012-3014 N...	Pass	2012-06-05	1271209	11587	30. FOOD IN O...	Grocery Store	LARAMIE BAK...	LARAMIE BAK...		Pass		
-87.73	Canvas	4000 W DIVER...	Fail	2017-09-12	2081920	2451427	29. PREVIOUS...	Grocery Store	CERMAR FRC...	CERMAR FRC...		Pass		
-87.63	Canvas	216 W JACKS...	Pass w/ Con...	2014-02-19	1286375	1579968	38. VENTILATI...	Restaurant	JIMMY JOHNS...	JIMMY JOHNS...		Pass		
-87.68	Canvas	6550 S Seeley...	Pass	2017-09-12	2081919	24801	18. NO EVIDE...	School	O'TOOLE ELE...	O'TOOLE ELE...		Pass		
-87.65	Canvas	6006 S Peoria...	Pass	2012-06-05	1158570	22181	33. FOOD AN...	School	Nicholson	Nicholson		Pass		
-87.6	Complaint Re...	1236 E 53RD ST	Pass	2017-09-12	2081910	2313667	18. NO EVIDE...	Restaurant	LEONAS REST...	LEONAS REST...		Pass		
-87.64	Canvas	931 N ORLEA...	Pass	2016-05-31	1931909	1959848	19. OUTSIDE ...	Restaurant	MOGADISHU...	MOGADISHU...		Pass		
-87.7	Canvas	7301 S KEDZL...	Pass	2017-08-22	2078790	7332	33. FOOD AN...	Restaurant	GEORGE'S CA...	GEORGE'S CA...		Pass		
-87.62	Canvas	1400 S LAKE...	Pass w/ Con...	2017-03-31	2009684	2284601	2. FACILITIES...	Restaurant	EXPLORER CA...	ARAMARK-FIE...		Pass		
-87.69	Canvas	2355 W ADDIS...	Pass	2017-09-12	2081909	255774	38. VENTILATI...	Restaurant	Popeye's #812...	Popeye's #812...		Pass		
-87.65	Canvas	950-952 W AR...	Pass	2016-05-31	1930332	11030	18. NO EVIDE...	Restaurant	KINCADES	KINCADES		Pass		
-87.63	Canvas	7525 S Harvar...	Fail	2015-10-20	1579166	23581	14. PREVIOUS...	School	HARVARD ELE...	HARVARD ELE...		Fail		
-87.77	Complaint Re...	6840 N SACR...	Pass	2017-03-31	2009682	2032949	8. SANITIZING...	Restaurant	PARK PLAZA...	PARK PLAZA...		Pass		
-87.59	Canvas	6718 S STON...	Pass	2016-12-20	1978469	1943058	32. FOOD AN...	Restaurant	SUBWAY...	SUBWAY...		Pass		
-87.75	Canvas	4021 N MILW...	Pass w/ Con...	2016-05-31	1931910	2334341	21. * CERTIFIE...	Restaurant	Jimmy John	Jimmy Johns		Pass		
-87.66	Canvas	7600 S ASHL...	Pass	2017-09-12	2081903	1933841	18. NO EVIDE...	Restaurant	MAXWELLS	MAXWELLS		Pass		
-87.66	Canvas	6920 S ASHL...	Pass w/ Con...	2017-03-31	2009681	2446619	11. ADEQUAT...	Restaurant	SUBWAY...	SUBWAY...		Pass		
-87.72	Short Form Co...	7010 S PULAS...	Pass	2017-09-12	2081906	2308673	32. FOOD AN...	Restaurant	BUFFET CITY...	BUFFET CITY...		Pass		
-87.65	Canvas	1234 N HALS...	Pass	2016-05-31	1931908	2138602	32. FOOD AN...	Restaurant	JUST TURKEY...	JUST TURKEY...		Pass		
-87.64	Complaint	554 W ROOSE...	Fail	2017-09-12	2081896	1694688	2. FACILITIES...	Restaurant	EPPLES REST...	EPPLES REST...		Pass		
-87.72	Suspected Fo...	3545 W Fulton	Pass	2017-03-31	2009680	46471	38. VENTILATI...	School	AL RABY HIG...	AL RABY HIG...		Pass		

## D] Data Quality

### Description:

**Quality of Data prior Refinement: 42%**

**Quality of Data Afte Refinement via Excel : 52%**

**Quality of Data After Refinement via IBM Watson : 62%**

The screenshot shows the IBM Watson Analytics interface. At the top, there are five main navigation cards: Explore, Predict, Assemble, Social Media (with a 'New!' badge), and Refine. Below this is a search bar and a sidebar with 'Shared' and 'Personal' sections. The main area displays a grid of datasets under the 'Personal' section. A red box highlights the third dataset from the left, which has a 'LOW QUALITY' score of 48. Other datasets shown include:

Dataset	Quality Score	Description
Food_Inspections_Excel Refinement	62 (MEDIUM)	DATA SET CSV
Food_Inspections_Excel	52 (MEDIUM)	DATA SET CSV
Food_Inspections_Prior Excel	48 (LOW)	DATA SET CSV
Public_Health_Statistics_Diabetes_hospitalizations_in_Chicago_2000_-	91 (HIGH)	DATA SET CSV
Medicaid Beneficiaries Inpatient Admissions Refinement	50 (LOW)	DATA SET Refinement

The screenshot shows the IBM Watson Analytics interface. At the top, there are five main navigation tabs: Explore, Predict, Assemble, Social Media, and Refine. Below these, a search bar and a sidebar with 'Shared' and 'Personal' sections are visible. The main content area displays a dashboard titled 'Data Quality after refinement in IBM Watson'. This dashboard lists five data sets with their quality scores and details:

Dataset	Quality Score	Description	Last Updated
Food_Inspections_Excel Refinement	(62) MEDIUM QUALITY	DATA SET Refinement	Oct 19, 2017
Food_Inspections_Excel	(52) MEDIUM QUALITY	DATA SET CSV	Oct 19, 2017
Food_Inspections_Prior Excel	(48) LOW QUALITY	DATA SET CSV	Oct 18, 2017
Public_Health_Statistics-Diabetes_hospitalizations_in_Chicago_2000-	(91) HIGH QUALITY	DATA SET CSV	Oct 18, 2017
Medicaid Beneficiaries Inpatient Admissions Refinement	(50) LOW QUALITY	DATA SET Refinement	Sep 16, 2017

Annotations in the image highlight the first two datasets in the list, with arrows pointing from the text 'Data Quality after refinement in IBM Watson' to each dataset's title.

## E] Data Exploration

### Question1

**What is the value of risk for each quarter from year 2010 -2017?**



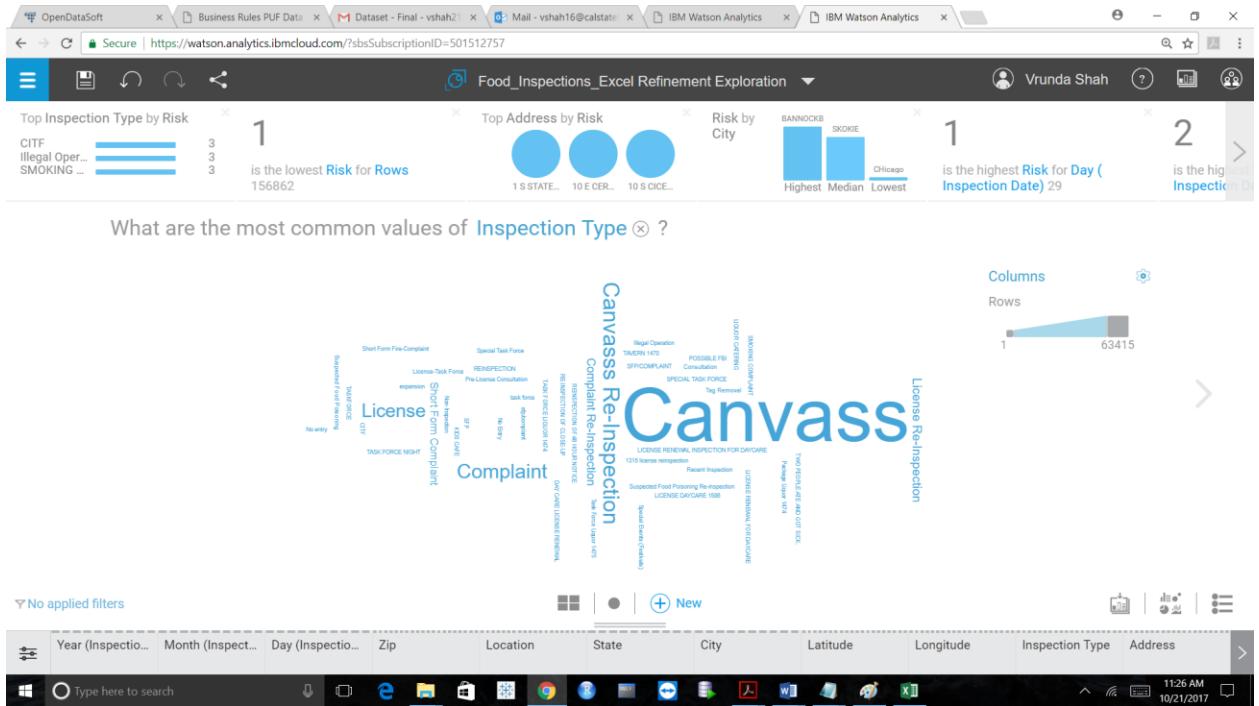
## Graph Type: Bar Chart

### Description

Risk 1 is considered as a High risk, Risk 2 – Medium and Risk 3 – Low. From the above visualization we can say during 1<sup>st</sup> quarter of 2016 average risk for state of illionis was 1.18 which is considered to be the highest among all for the tenure of 2010 – 2017 and the lowest risk during this tenure was 1.51 which is 4<sup>th</sup> quarter for year 2015.

### Question 2

**What are the most common values of Inspection Type?**



## Graph: Word Cloud

### Description:

We created above visualization to identify which type of Inspection is requested more in the state of Illinois. We found out Canvass and Canvass Inspection is requested more. We also can visualize from above word cloud that License, License Inspection and Complaint Re-inspection were requested moderately and task-force and suspect food poisoning inspection were requested by very few people.

### Question 3

What are the connection between Facility type and Inspection type?



### Description:

Above screenshot summarizes which kind of Inspection type is frequently used by particular facility type. Facility Type Shelter, Banquet, Wholesale and Bakery has strong connection to Inspection Type Complaint and Complaint Re-inspection. It is also evident from the visualization that Facility type gas station has strong connection with consultation inspection type; whereas Hospital, Long term care and rooftop facility type has strong connection with suspecting food poisoning inspection type. Facility type social club, coffee shop, liquor store, day care, candy shop, mobile and Paleteria has strong connection with inspection type Canvass, Canvass re-inspection, License and License re-inspection.

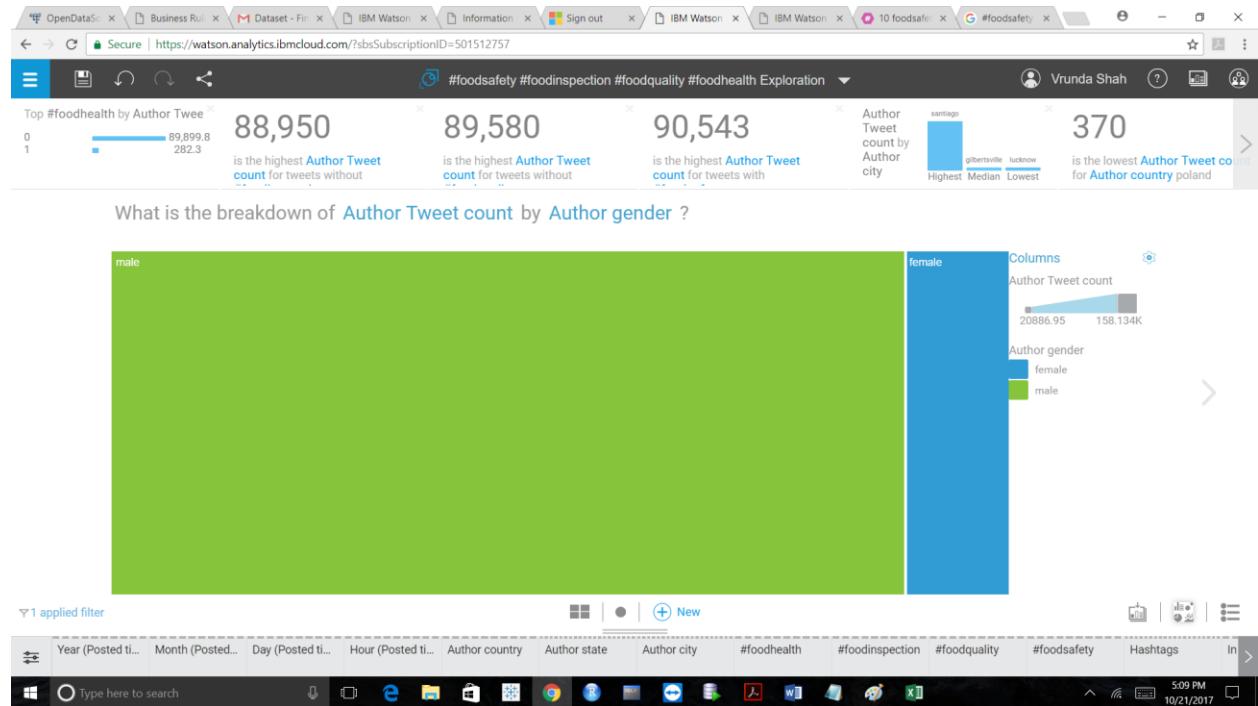
### Two Important points:

- 1) Inspection type Canvass, Canvass re-inspection, License is majorly used by all of the facility type

- 2) Task -force inspection type is only used by Long Term Care facility type.

## **Question 4**

**What is the breakdown of Author Tweet Count by Author gender?**



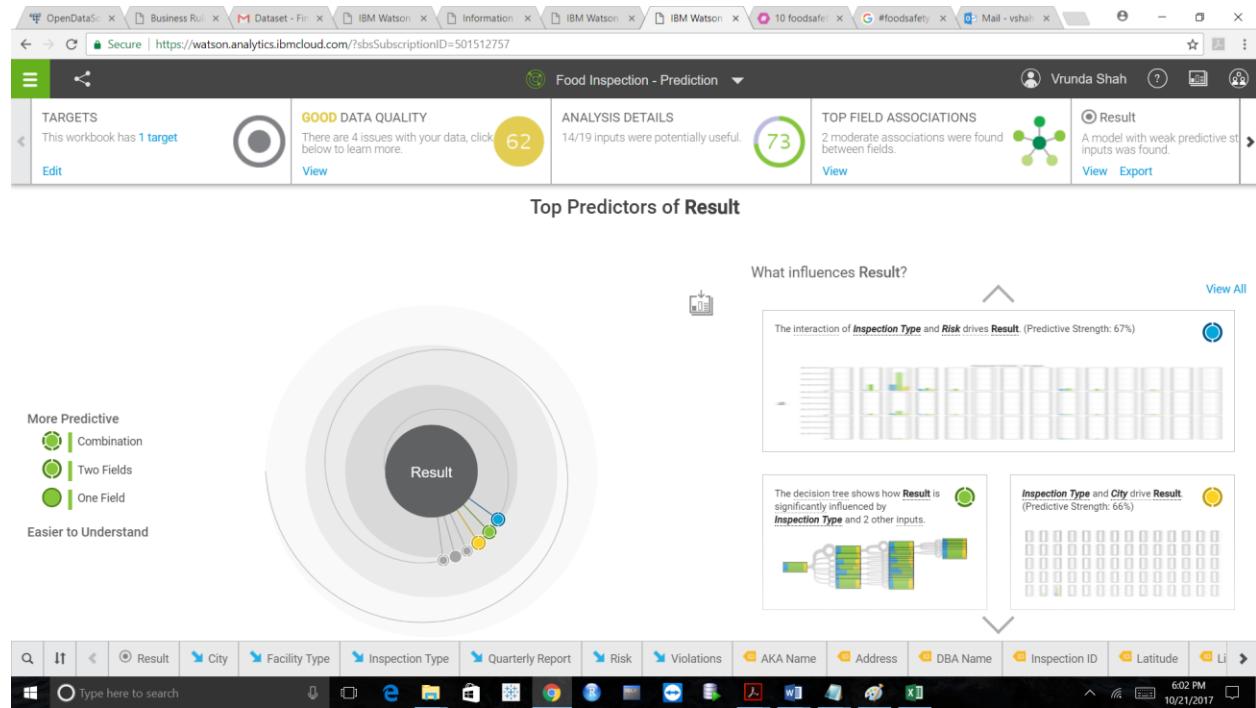
**Hashtag Used:** #foodsafety, #foodinspection, #foodquality, #foodhealth

**Graph Used:** Tree Map

**Description:** We have used hashtags and Twitter data to identify which gender of the population of globe have shared their views on food inspection or its quality or its safety from 1<sup>st</sup> January 2017 to 21<sup>st</sup> October 2017. We found that across the globe ‘Male’ people have tweeted more about food inspection related topic than female. In the above visualization 158133.82 males have tweeted and 20886.95 females have tweeted related to food inspection or quality in the tenure mentioned above.

## F] Prediction

### Prediction



We have made prediction for Result field which has outcomes (Pass, Pass with conditions, Fail) based on the input provided. In the above prediction visualization Target Value is Result and Input Values provided for the prediction of results are City, Facility Type, Inspection Date(Year), Inspection Type, Risk, Violations, Quarterly Report. Analysis Details ratio is 73% which stated predication model is of high quality.

Inspection Type and Risk Drives Result. Decision Tree shows **Result** is significantly influenced by Inspection Type, Quarterly Report and Risk value.

### Decision Table

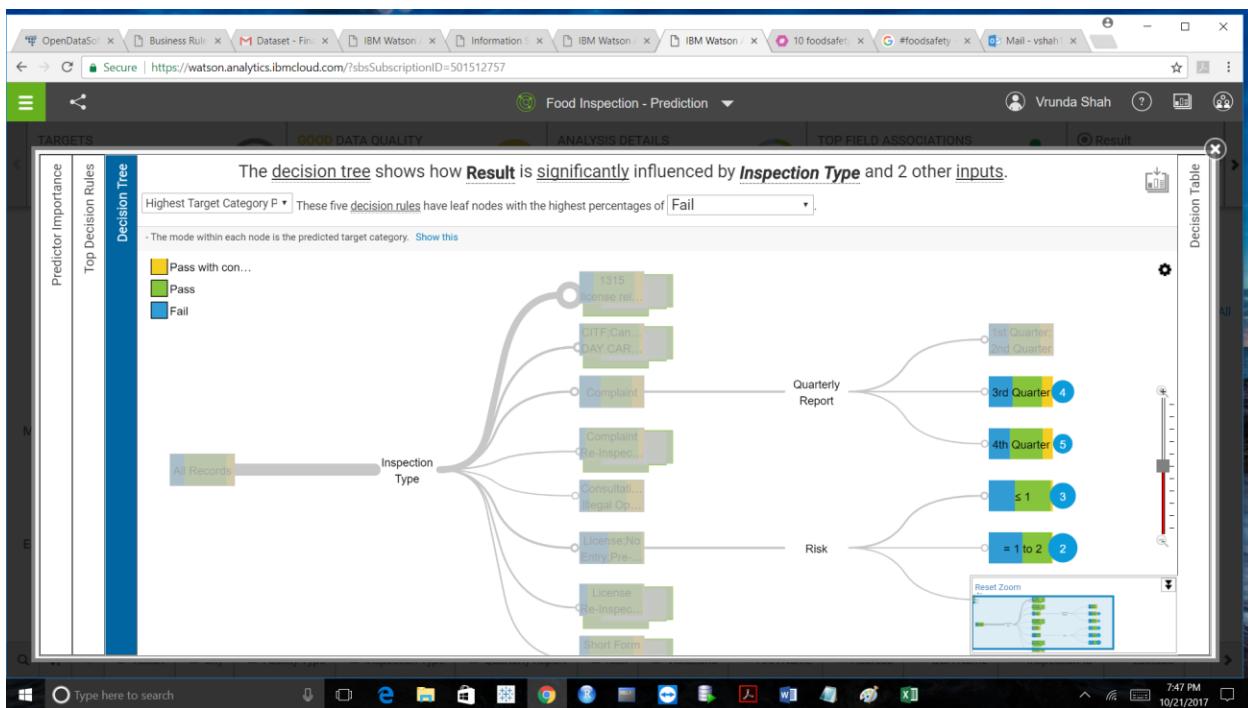
The screenshot shows the Watson Analytics interface with the title "Food Inspection - Prediction". The main area displays a "Decision Table" for a categorical target. The table has columns for Predictor Importance, Top Decision Rules, and Decision Tree. The rows represent different input parameters: Inspection Type, Risk, Quarterly Report, and Predicted Result. The table highlights a specific row and column with red borders, indicating a predicted result of "Pass" based on the conditions: (Inspection Type = Complaint Re-Inspection) AND (Risk ≤ 1). The status bar at the bottom shows the date and time as 10/21/2017 8:11 PM.

Predictor Importance	Top Decision Rules	Decision Tree	Decision Table							
			<b>Result</b> is a categorical target so a CHAID classification tree is used. Prediction Table The prediction table shows a different decision rule in each column. - <b>Result</b> is predicted by <b>Inspection Type</b> and 2 other inputs with 66% accuracy Statistical Details - In order to optimize the analysis, field transformations were performed. Statistical Details							
			<b>Input</b> Read the input values down each column to get the predicted Result							
Inspection Type	≤ 1	> 1	Complaint	Complaint	Complaint	Complaint Re-Inspection	Complaint Re-Inspection	License No En		
Risk	≤ 1	> 1				≤ 1	> 1	≤ 1		
Quarterly Report			1st Quarter, 2nd Quarter	3rd Quarter	4th Quarter					
Quarterly Report										
Predicted Result	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass		

IF (Inspection Type = Complaint Re-Inspection) AND (Risk ≤ 1) THEN predicted Result = Pass .

**Description:** CHAID Classification is used for decision table. It is a matrix which has various parameters and final outcome. For example let's take the example of highlighted to column, it is interpreted as If Inspection Type = Complaint Re-Inspection AND Risk <= 1 then Result is Pass.

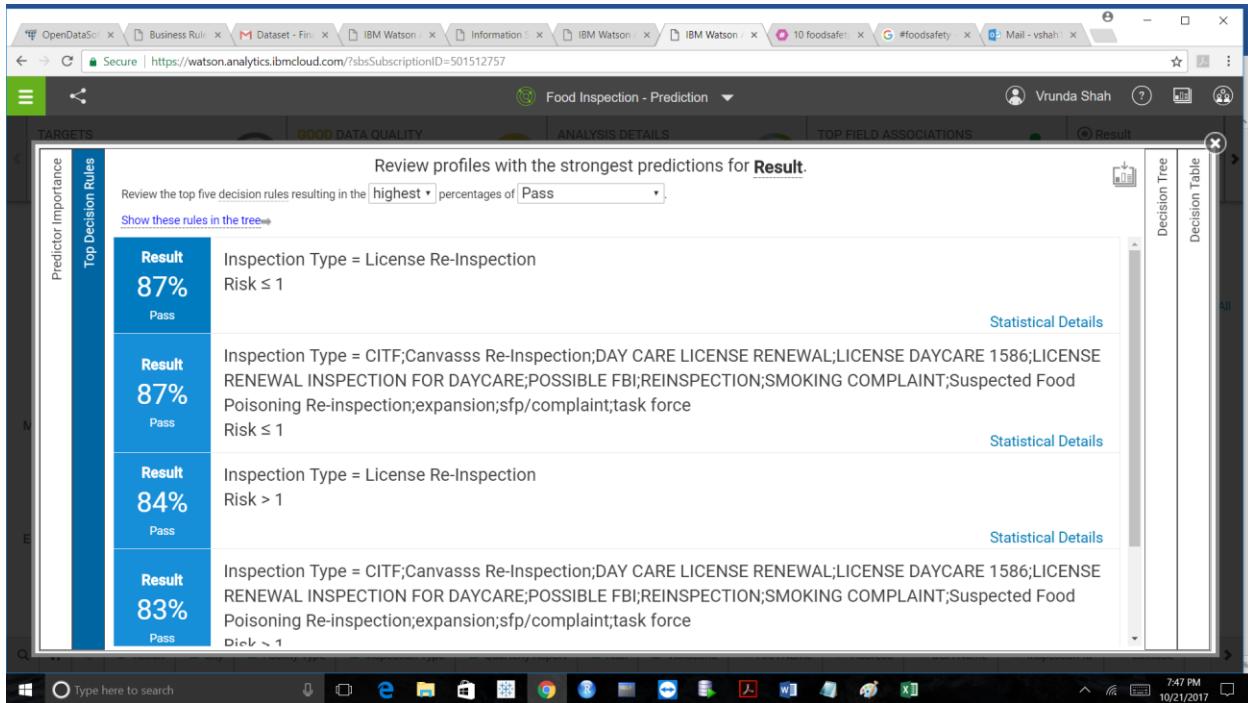
## **Decision Tree**



## Description

Above decision tree shows All Records are been classified differently based on Inspection Type. Further to predict result Inspection Type interact with Risk, which comprises risk value (1 – 3). It also uses Quarterly Report parameter to determine final result falls under which quarter after considering all parameters Inspection Type and Risk falls into which category of result(Pass, Fail, Pass with conditions).

## **Top Decision Rules**



### **Description:**

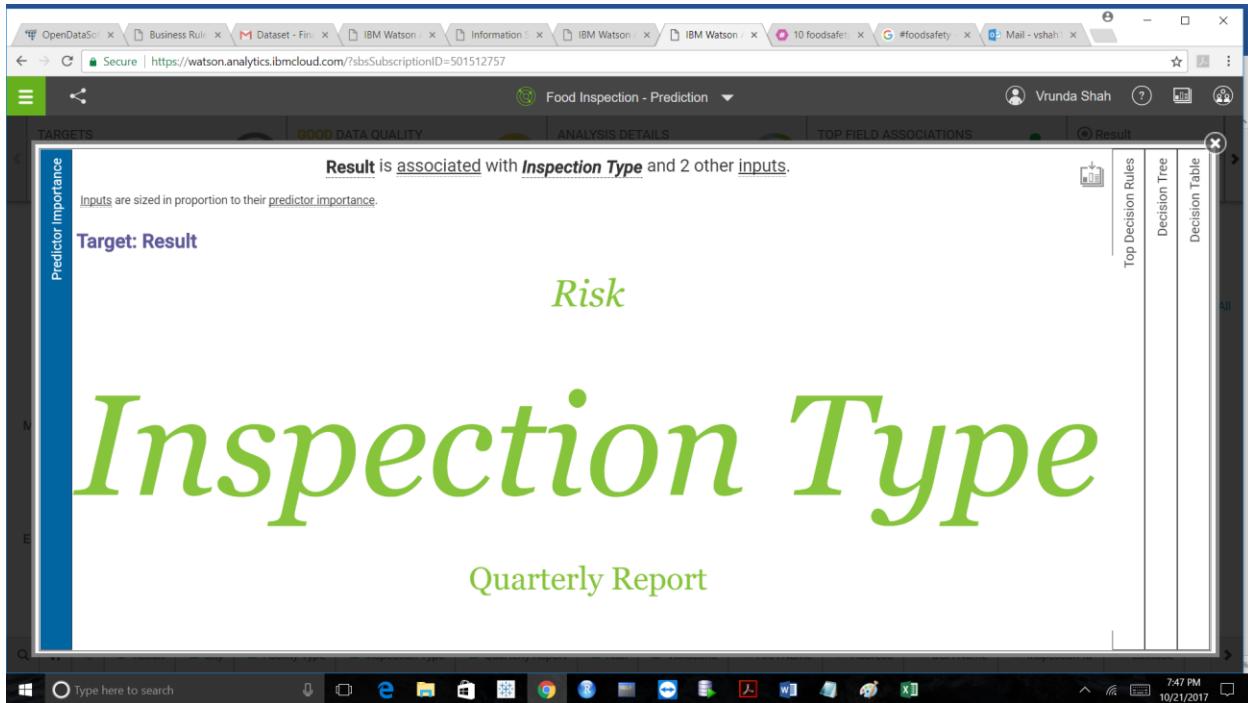
Top Decision Rules will act as a future reference that model will follow for the of future decision making. This rules were formed based on the input parameters + Data.

Above visualization shows Top Decision rules for Result (Pass, Pass with condition and Fail)

Below are the few top Decision Rules for Pass Result. It is interpreted as mentioned below

- i) If Inspection Type = License Re-Inspection and Risk  $\leq 1$  Then 87% result is Pass
- ii) If Inspection Type – CITF, Canvass Re – Inspection, License Inspection; Suspected food poisoning Re-Inspection; task force, Complaint and Risk  $\leq 1$  then 87% Result is Pass.
- iii) If Inspection Type = License Re-Inspection and Risk  $> 1$  then 84% times Result = Pass.

## Predictor Importance

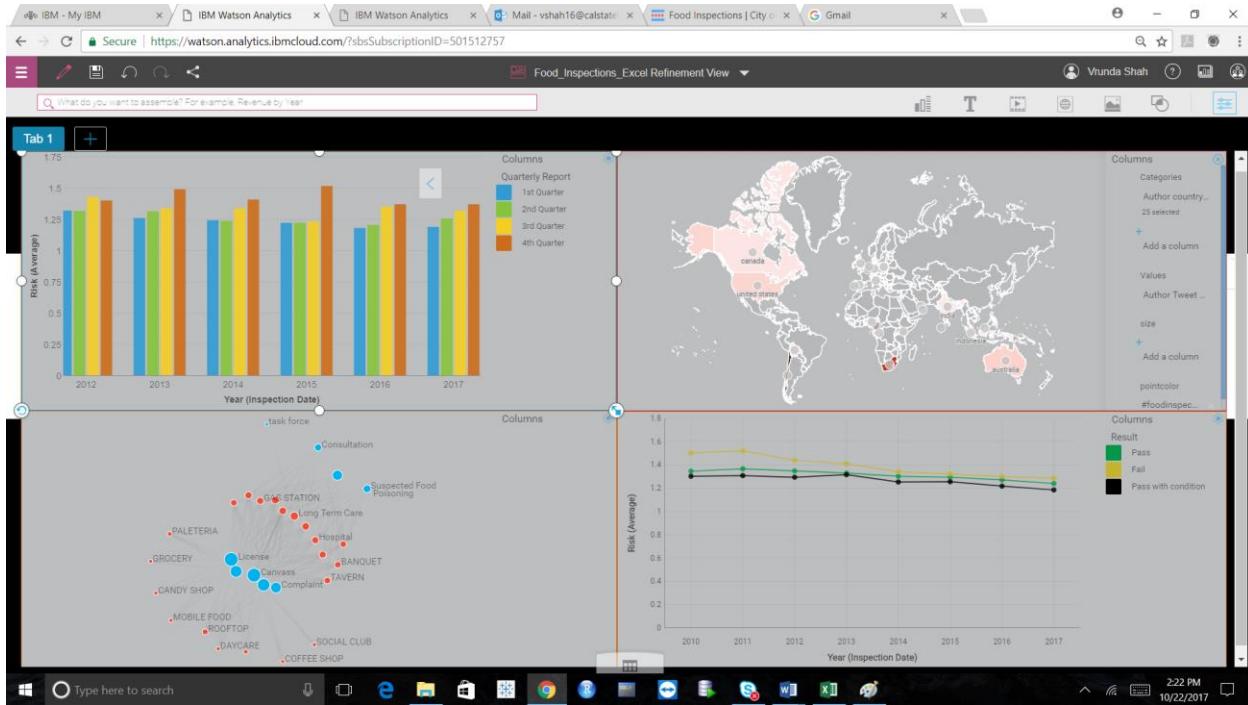


### Description:

The above visualization provides information which input values were considered important and drives our end result.

Result, our final output is based on 3 parameters Risk, Inspection type, quarterly Report.

## G] Dashboard



To summarize, the above dashboard displays all the important insights from our data analysis. In the first screenshot we are evaluation risk each quarter from year 2010 – 2017. In second screenshot, we can also include majority author tweets which are related to #foodinspection is from United States, Canada, India and Indonesia. In third screenshot, we are summarizing how Inspection type is related with Facility type. Basically we are trying to gain insight which kind of Inspection type is frequently used by particular facility type. We can clear identify two important points Inspection type Canvass, Canvass re-inspection, License is majorly used by all of the facility type. Task -force inspection type is only used by Long Term Care facility type. In the fourth screenshot, we have analyzed trend of the risk over tenure of 2010 – 2017 by Result (Pass, Fail, Pass with condition) using line chart. We have found that facility type who had average risk of 1.5 use to fail inspection in 2010 has reduced to risk 1.29. On the other hand there is mere change in average risk for the facility type who had cleared inspection with Pass or Pass with condition status.