



---

FOOD INSPECTION IN ILLINOIS

---

Data Analysis Using Python



[CIS – 5810-01 HealthCare Information Systems](#)

DECEMBER 7, 2017

VRUNDA SHAH  
Prof. 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

## 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).

## **B] Data Cleaning**

### **Complete Code Screenshot**

The screenshot shows the Spyder Python IDE interface. The main window displays a Python script named Data\_Cleaning.py. The code is as follows:

```
1 # -*- coding: utf-8 -*-
2 """
3 Created on Sat Nov 25 22:00:54 2017
4
5 @author: Vrunda B Shah
6 """
7
8 import pandas as pd
9 import numpy as np
10 data1 = pd.read_csv('Food_Inspections.csv')
11 data1.replace('Risk 2 (Medium)', '2')
12 # dropping rows with blank values
13 rem_missing = data1.dropna()
14 rem_missing[rem_missing.notnull()]
15 # Splitting Inspection Date to access only year data from it - String Functions
16 rem_missing[['Month','Date','Year']] = rem_missing['Inspection Date'].str.split('/', expand = True)
17 rem_missing[['Num', 'ViolationDescription']] = rem_missing['Violations'].str.split('.', 1,expand = True)
18 # Adding Inspection Type into new variable to save it in text file for future use
19 rem_missing1 = rem_missing[['Inspection Type']]
20 # Dropping of columns
21 rem_missing = rem_missing.drop(['DBA Name', 'AKA Name', 'Inspection Date', 'Month', 'Date'], axis = 1)
22
23 print (rem_missing.head())
24 rem_missing.to_csv('file_clean.csv', index = False)
25
26 np.savetxt('a.txt', rem_missing1, fmt='%s', delimiter = '\t', newline = '\n')
27 np.savetxt('b.txt', rem_missing1, fmt='%s', delimiter = ',', newline = '\n')
```

The right side of the interface includes a 'Usage' help panel and a 'Python console' window. The taskbar at the bottom shows various open applications like File Explorer, Task View, and a browser.

## 1) Removal of NA and Blank values

### Code

```
rem_missing = data1.dropna()
```

```
rem_missing[rem_missing.notnull()]
```

### Screenshot

**Food\_Inspections (1) - Excel**

Inspect:	DBA Na	AKA Na	License	Facility	Risk	Address	City	State	Zip	Inspect:	Inspection Type	Results	Violatic	Latitude	Longitude	Location
2099021	DS TEQUIL DS TEQUIL	2542467	Restauran	Risk 1	(Hig 3350-3352 CHICAGO IL	60657	#####	IL	60657	#####	Sort A to Z	Fail	18. NO EVI	41.94332	-87.6495	(41.94331835640129, -87.649514755563)
3099016	DS TEQUIL DS TEQUIL	2542468	Restauran	Risk 1	(Hig 3350-3352 CHICAGO IL	60657	#####	IL	60657	#####	Sort Z to A	Pass	18. NO	41.84879	-87.7053	(41.84879049508066, -87.705310981556)
4099006	LA V BAKE LA V BAKE	2550054	Bakery	Risk 2	(Me 2338 S KEE CHICAGO IL	60623	#####	IL	60623	#####	Sort by Color	Pass	31. CLEAN	41.86484	-87.6568	(41.86485873399433, -87.6568659479)
5098949	GILCHRIST GILCHRIST	2514038	Daycare Al Risk	Risk 2	(Me 1104 S RAI CHICAGO IL	60608	#####	IL	60608	#####	Clear Filter From "Inspection Type"	Fail	38.	41.86906	-87.626	(41.8690625105567, -87.62601598275)
6098967	FRANNIE'S COLUMBI	251098	Restauran	Risk 2	(Me 1104 S WA CHICAGO IL	60605	#####	IL	60605	#####	Filter by Color	Fail	18. NO	41.74754	-87.605	(41.74753844847344, -87.605049857308)
7098958	PARADISE PARADISE	2517736	Grocery St Risk 2	(Me 8101 S CO CHICAGO IL	60619	#####	IL	60619	#####	Text Filters	Pass	38. VENTI	41.80767	-87.7306	(41.80767249318478, -87.730609902682)	
8098931	EL MUELLIEL MUELLI	1484896	Restauran	Risk 1	(Hig 4253 W 47 CHICAGO IL	60632	#####	IL	60632	#####	Search	Fail	35.	41.99765	-87.6883	(41.99764733751051, -87.68826118291)
9098927	SAFA MAR SAFA MAR	2283942	Grocery St Risk 2	(Me 2319 W DI CHICAGO IL	60659	#####	IL	60659	#####	Check Box	Pass	14. PREVIC	41.99768	-87.6989	(41.997682115657, -87.698913591067)	
10098956	EGGSXPERIENCE	2560093	Restauran	Risk 1	(Hig 1230 W TA CHICAGO IL	60607	#####	IL	60607	#####	Task Force	Fail	12. HAND	41.88199	-87.625	(41.88199089398375, -87.62503959142)
11098954	POTBELLY POTBELLY	2534897	Restauran	Risk 1	(Hig 756 E 1111 CHICAGO IL	60628	#####	IL	60628	#####	Task Force Liquor	Pass	16. FOOD	41.86951	-87.6559	(41.86950816613224, -87.65591982305)
12098955	Daallo Me Daallo Me	2013707	Grocery St Risk 2	(Me 2742 W De CHICAGO IL	60659	#####	IL	60659	#####	Task Force Night	Pass	14. PREVIC	41.99768	-87.6989	(41.997682115657, -87.698913591067)	
13098953	POKEWOR POKEWOR	2560710	Restauran	Risk 1	(Hig 79 E MAD CHICAGO IL	60602	#####	IL	60602	#####	Taste of Chicago	Fail	12. HAND	41.88199	-87.625	(41.88199089398375, -87.62503959142)
14098945	A10 A10	2241029	Restauran	Risk 1	(Hig 1460-1466 CHICAGO IL	60615	#####	IL	60615	#####	Two People Ate and Got Sick	Pass	16. FOOD	41.79965	-87.5894	(41.7996512604028, -87.589432855582)
15098937	MELLOW MELLOW	1542	Restauran	Risk 1	(Hig 1508 E 531 CHICAGO IL	60615	#####	IL	60615	#####	(Blanks)	Not Applicable	10. NO EVI	41.79965	-87.5887	(41.79965205200755, -87.58875171085)
16098940	TAQUERIA TAQUERIA	1823377	Restauran	Risk 1	(Hig 3452 W 55 CHICAGO IL	60629	#####	IL	60629	#####	Two People Ate and Got Sick	Pass	33. FOOD	41.79965	-87.5887	(41.79965205200755, -87.58875171085)
17098934	KUSANYA KUSANYA	2294418	Restauran	Risk 1	(Hig 825 W 691 CHICAGO IL	60621	#####	IL	60621	#####	Two People Ate and Got Sick	Pass	35. WALLS	41.76871	-87.6456	(41.76871169711559, -87.645577780655)
18098926	DON DE VILON DE VI	2172653	Grocery St Risk 2	(Me 3543 S LA CHICAGO IL	60625	#####	IL	60625	#####	Two People Ate and Got Sick	Pass	Out of Business	41.96825	-87.7176	(41.96824854706044, -87.71763240512)	
19098505	EAT & DRI EAT & DRI	2069092	Restauran	Risk 1	(Hig 4649 N DA CHICAGO IL	60625	#####	IL	60625	#####	Two People Ate and Got Sick	Pass	33. FOOD	41.80041	-87.5884	(41.800409523947955, -87.58836368635)
20098929	NATIVE FC NATIVE FC	2431737	Restauran	Risk 1	(Hig 1518 E HA CHICAGO IL	60615	#####	IL	60615	#####	Two People Ate and Got Sick	Pass	38. VENTI	41.80041	-87.5884	(41.800409523947955, -87.58836368635)
21098922	CO-OPTIM CO-OPTIM	2542098	BAR	Risk 3	(Loj 550 W WA CHICAGO IL	60661	#####	IL	60661	#####	Two People Ate and Got Sick	Fail	38. VENTI	41.88329	-87.6417	(41.88328561813559, -87.641746760772)
22098923	CAFETERIA Y NEVERU	2560709	Restauran	Risk 2	(Me 4204-4216 CHICAGO IL	60639	#####	IL	60639	#####	Two People Ate and Got Sick	Pass	41.90993	-87.7316	(41.909932470575885, -87.73155721632)	
23098912	MC KINNE MC KINNE	2511356	Restauran	Risk 1	(Hig 200 W 103 CHICAGO IL	60628	#####	IL	60628	#####	Two People Ate and Got Sick	Pass	40. REFRIG	41.70721	-87.6284	(41.707210358472786, -87.62835881151)
24098912	MC KINNE MC KINNE	2215689	Daycare	L Risk 1	(Hig 5733-5743 CHICAGO IL	60651	#####	IL	60651	#####	Two People Ate and Got Sick	Pass	38. VENTI	41.90198	-87.7694	(41.9019797953018, -87.769353282123)
25098906	GIORDANI GIORDANI	35633	Restauran	Risk 1	(Hig 5311 S BLU CHICAGO IL	60615	#####	IL	60615	#####	Two People Ate and Got Sick	Pass	32. FOOD	41.779916	-87.5903	(41.7799155754303854, -87.59028080505)
26098902	ZACATACC ZACATACC	50551	Restauran	Risk 1	(Hig 5295 S PU1 CHICAGO IL	60629	#####	IL	60629	#####	Two People Ate and Got Sick	Fail	33. FOOD	41.78505	-87.7228	(41.78505749032044, -87.72279048656)
27098889	LA ESCUELA ESCUEL	2247067	Daycare	A Risk 1	(Hig 3001-3005 CHICAGO IL	60647	#####	IL	60647	#####	Two People Ate and Got Sick	Pass	24. DISH W	41.93195	-87.7026	(41.93194853574754, -87.70258032777)
28098900	THE LEARN THE LEARN	2464806	Children's	Risk 1	(Hig 5331 N LIN CHICAGO IL	60625	#####	IL	60625	#####	Two People Ate and Got Sick	Fail	2. FACILIT	41.97865	-87.6925	(41.978648374745106, -87.6924770324)
29098904	A & J FOOL A & J FOOL	69143	Grocery St Risk 1	(Hig 342 W 751 CHICAGO IL	60621	#####	IL	60621	#####	Two People Ate and Got Sick	Pass w/ Cc 21.	*	41.7582	-87.6338	(41.758199702466, -87.633757062717)	
30098944	SYMPHON SYMPHON	2447171	Long Term Care	Risk 1	(Hig 4437 S CIC CHICAGO IL	60632	#####	IL	60632	#####	Two People Ate and Got Sick	Fail	18. NO EVI	41.81199	-87.7431	(41.811989599987115, -87.74312824835)

## After Cleaning

**file\_clean - Excel**

Inspection ID	License #	Facility Type	Risk	Address	City	State	Zip	Inspection Type	Results	Violations
2099016	2542468	Restaurant	1	3350-3352 N HALSTED ST	CHICAGO	IL	60657	Lic	Fail	18. NO EVI
2099006	2550054	Bakery	1	2338 S KEDZIE AVE	CHICAGO	IL	60623	Lic	Pass	18. NO EVI
2098969	2314038	Daycare Above and Under 2 Years	1	1312 S RACINE AVE	CHICAGO	IL	60608	Lic	Pass	31. CLEAN
2098967	2551098	Restaurant	1	2100 S WABASH AVE	CHICAGO	IL	60615	Lic	Fail	18. NO EVI
2098958	2517736	Grocery Store	1	8101 S COTTAGE GROVE AVE	CHICAGO	IL	60619	Co	Pass	38. VENTI
2098931	1488496	Restaurant	1	14253 W 47TH ST	CHICAGO	IL	60632	Co	Pass	38. VENTI
2098927	2283942	Grocery Store	1	2319 W DEVON AVE	CHICAGO	IL	60659	Co	Pass	35. WALLS
2098955	2013707	Grocery Store	1	2742 W Devon AVE	CHICAGO	IL	60659	Co	Pass	14. PREVIC
2098953	2560710	Restaurant	1	179 E MADISON ST	CHICAGO	IL	60602	Lic	Pass	33. FOOD A
2098937	1542	Restaurant	1	1508 E 53RD ST	CHICAGO	IL	60615	Sh	Pass	32. FOOD A
2098940	1823377	Restaurant	1	13452 W 59TH ST	CHICAGO	IL	60629	Ca	Pass	35. WALLS
2098934	2294418	Restaurant	1	1825 W 69TH ST	CHICAGO	IL	60621	Ca	Pass	33. FOOD A
2098929	2431737	Restaurant	1	1518 E HARPER CT	CHICAGO	IL	60615	Ca	Pass	24. DISH W
2098916	2511356	Restaurant	1	1200 W 103RD ST	CHICAGO	IL	60628	Co	Pass	40. REFRIG
2098912	2215689	Daycare (Under 2 Years)	1	5733-5743 N DIVISION ST	CHICAGO	IL	60651	Lic	Pass	38. VENTI
2098906	35633	Restaurant	1	5311 S BLACKSTONE AVE	CHICAGO	IL	60615	Ca	Pass	32. FOOD A
2098902	50551	Restaurant	1	5925 S PULASKI RD	CHICAGO	IL	60629	Ca	Pass	33. FOOD A
2098889	2247067	Daycare Above and Under 2 Years	1	3001-3009 W DIVERSEY AVE	CHICAGO	IL	60647	Lic	Pass	21. * CERTI
2098900	2464806	Children's Services Facility	1	1533 N LINCOLN AVE	CHICAGO	IL	60625	Co	Pass	18. NO EVI
2098994	69143	Grocery Store	1	1342 W 75TH ST	CHICAGO	IL	60632	Co	Pass	38. VENTI
2098892	2470228	Children's Services Facility	1	15467 S UNIVERSITY AVE	CHICAGO	IL	60615	Co	Pass	34. FLOORS
2098893	2215869	Daycare (Under 2 Years)	1	1657 S WELLS ST	CHICAGO	IL	60607	Lic	Pass	11. ADEQU
2098768	2528979	Restaurant	1	300 E CERMACK RD	CHICAGO	IL	60616	Lic	Pass	34. FLOORS
2098883	2142129	Grocery Store	1	13629 W 63RD ST	CHICAGO	IL	60629	Co	Pass	33. FOOD A
2098882	15790	Hospital	1	326 W 64TH ST	CHICAGO	IL	60621	Co	Pass	33. FOOD A
2098877	1768391	Long Term Care	1	10450 S MICHIGAN AVE	CHICAGO	IL	60628	Co	Pass	11. ADEQU
2098875	2560708	Restaurant	1	2633 S ASHLAND AVE	CHICAGO	IL	60636	Lic	Fail	11. ADEQU

## 2) Replacing Values

## Code

```
data1.replace('Risk 1 (High)', '1')
```

```
data1.replace('Risk 2 (Medium)', '2')
```

```
data1.replace('Risk 3 (Low)', '3')
```

## Screenshot

Inspect	DBA Name	AKA Name	License	Facility	Risk	Address	City	State	Zip	Inspect	Inspection Type	Results	Violat	Latitude	Longitude	Location
2	2099021 DS TEQUIL DS T				Sort A to Z	3350-3352 CHICAGO IL				60657 #####	License	Fail	41.94332	-87.6495	(41.94331835640129,-87.649514)	
3	2099016 DS TEQUIL DS T				Sort Z to A	3350-3352 CHICAGO IL				60657 #####	License	Fail	18. NO EV	41.94332	-87.6495	(41.94331835640129,-87.649514)
4	2099006 LA V BAKE (LA V				Sort by Color	2338 S KEE CHICAGO IL				60623 #####	License Re-Inspection	Pass	18. NO	41.84879	-87.7053	(41.84879049508066,-87.705310)
5	2098969 GILCHRIST GILC					1312 S RAI CHICAGO IL				60608 #####	License	Pass	31. CLEAN	41.85486	-87.6568	(41.86485873399433,-87.656765)
6	2098967 FRANNIE'S COL					1104 S WA CHICAGO IL				60605 #####	License	Fail	38.	41.86906	-87.626	(41.86906251655677,-87.626015)
7	2098958 PARADISE					1010 S WA CHICAGO IL				60619 #####	Complaint	Fail	18. NO	41.74754	-87.605	(41.74753844847344,-87.60498)
8	2098931 EL MUELLER M					4252 W 7 CHICAGO IL				60632 #####	Canvass	Pass	38. VENTI	41.80767	-87.7306	(41.80767249318478,-87.730609)
9	2098927 SAFA MAR JAF					2319 W DE CHICAGO IL				60659 #####	Canvass	Fail	35.	41.99765	-87.6883	(41.99764733751051,-87.688261)
10	2098956 EGGSPIERIENCE					1230 W 7 CHICAGO IL				60607 #####	License Re-Inspection	Pass	16. FOOD	41.86951	-87.6579	(41.8695086613224,-87.657919)
11	2098954 POTBELLY POT					756 E 111 CHICAGO IL				60628 #####	License	Pass	41.6929	-87.6027	(41.69297593328423,-87.602664)	
12	2098955 Daallo M. Daal					2742 W DE CHICAGO IL				60659 #####	Canvass Re-Inspection	Pass	14. PREVIC	41.99768	-87.6980	(41.997682115657,-87.69801315)
13	2098953 POKEWOR POK					79 E MADJ CHICAGO IL				60602 #####	License	Fail	12. HAND	41.88199	-87.625	(41.88199083993875,-87.62503)
14	2098945 A10 A10					1460-146 CHICAGO IL				60615 #####	Canvass	No Entry	41.79965	-87.5384	(41.7996512604028,-87.5894032)	
15	2098937 MELLOW MEL					1508 E 5TH CHICAGO IL				60615 #####	Short Form Complaint	Pass	33. FOOD	41.79966	-87.5887	(41.799656205200755,-87.58873)
16	2098940 TAQUERIA AQ					3452 W 8 CHICAGO IL				60629 #####	Canvass	Pass	32. FOOD	41.78625	-87.7104	(41.78624677993347,-87.710431)
17	2098934 KUSANYA KUS					825 W 6TH CHICAGO IL				60621 #####	Canvass	Pass	35. WALLS	41.76871	-87.645577	(41.76871169711559,-87.645577)
18	2098926 DON DE VIDA DON					3543 W 1 CHICAGO IL				60625 #####	Canvass	Out of Business	41.9682	-87.7176	(41.96824854706049,-87.717563)	
19	2098905 EAT & DRINK AT					4649 N DE CHICAGO IL				60625 #####	Canvass	No Entry	41.96658	-87.679	(41.966582946351025,-87.678975)	
20	2098929 NATIVE FC NAT					1518 E FA CHICAGO IL				60615 #####	Canvass	Pass	33. FOOD	41.80041	-87.5384	(41.800409523947955,-87.588363)
21	2098922 CO-OPTIM CO-OPTIM	2542098_BAR	Risk 3 (Low)			550 W VA CHICAGO IL				60661 #####	License Re-Inspection	Fail	38. VENTI	41.88329	-87.6417	(41.88328561813559,-87.641746)
22	2098923 CAFETERIA Y NEVERU	2560709	Restauran Risk 2 (Medium)			4204-4216 CHICAGO IL				60639 #####	License	Pass	41.90993	-87.73135	(41.90992547057885,-87.73155)	
23	2098916 JIMMY'S E	2511356	Restauran Risk 1 (High)			200 W 103 CHICAGO IL				60628 #####	Complaint	Pass	40. REFRIG	41.70721	-87.6284	(41.707210358472786,-87.628358)
24	2098912 MC KINNE MC KINNE	2215689	Daycare (L Risk 1 (High)			5733-5743 CHICAGO IL				60651 #####	License	Pass	38. VENTI	41.90198	-87.76694	(41.90197977973018,-87.769353)
25	2098906 GIORDANO GIORDANO	35633	Restauran Risk 1 (High)			5311 S BLU CHICAGO IL				60615 #####	Canvass	Pass	32. FOOD	41.779916	-87.5903	(41.779915754303854,-87.590280)
26	2098904 ZACATACC ZACATACC	50551	Restauran Risk 1 (High)			5925 S PU CHICAGO IL				60629 #####	Canvass	Fail	33. FOOD	41.7850	-87.7227	(41.7850649032044,-87.722790)
27	2098886 LA ESCUELA ESCUE	2247067	Daycare Al Risk 1 (High)			3001-3005 CHICAGO IL				60647 #####	License Re-Inspection	Pass	24. DISH W	41.93195	-87.7026	(41.93194853574754,-87.702580)
28	2098900 THE LEARN THE LEARN	2464806	Children's Risk 1 (High)			5331 N LIP CHICAGO IL				60625 #####	License Re-Inspection	Fail	2. FACILIT	41.97865	-87.6925	(41.978648374745106,-87.69247)
29	2098904 A & J FOOL A & J FOOL	69143	Grocery St Risk 1 (High)			342 W 75T CHICAGO IL				60621 #####	Canvass	Pass w/ Cc 21. *	41.7582	-87.6338	(41.758197027466,-87.63375706)	
30	2098894 SYMPHON SYMPHON	2447171	Long Term Risk 1 (High)			4437 S CIC CHICAGO IL				60632 #####	Canvass	Fail	18. NO EV	41.81199	-87.74312	(41.81198599987115,-87.74312)

## After Cleaning

Inspection ID	License #	Facility Type	Risk	Address	City	State	Zip	Inspection Type	Results	Violations
2099016	2542468	Restaurant	Sort Smallest to Largest	3350-3352 N HALSTED ST	CHICAGO	IL	60657	License	Fail	18. NO EVI
2099006	2550054	Bakery	Sort Largest to Smallest	2338 S KEDZIE AVE	CHICAGO	IL	60623	License Re-Inspection	Pass	18. NO EVI
2098969	2314038	Daycare A	Sort by Color	1312 S RACINE AVE	CHICAGO	IL	60608	License	Pass	31. CLEAN I
2098967	2551098	Restaurant	Clear Filter From "Risk"	1104 S WABASH AVE	CHICAGO	IL	60605	License	Fail	38. VENTIL
2098958	2517736	Grocery St	Filter by Color	8101 S COTTAGE GROVE AVE	CHICAGO	IL	60619	Complaint	Fail	18. NO EVI
2098931	1488496	Restaurant	Number Filters	4253 W 47TH ST	CHICAGO	IL	60632	Canvass	Pass	38. VENTIL
2098927	2283942	Grocery St	Search	2319 W DEVOU AVE	CHICAGO	IL	60659	Canvass	Fail	35. WALLS,
2098955	2013707	Grocery St	OK	2742 W Devon AVE	CHICAGO	IL	60659	Canvass Re-Inspection	Pass	14. PREVIO
2098953	2560710	Restaurant	Cancel	79 E MADISON ST	CHICAGO	IL	60602	License	Fail	12. HAND V
2098937	1542	Restaurant		1508 E 53RD ST	CHICAGO	IL	60615	Short Form Complaint	Pass	33. FOOD A
2098940	1823377	Restaurant		3452 W 59TH ST	CHICAGO	IL	60629	Canvass	Pass	32. FOOD A
2098934	2294418	Restaurant		825 W 69TH ST	CHICAGO	IL	60621	Canvass	Pass	35. WALLS,
2098929	2431737	Restaurant		1518 E HARPER CT	CHICAGO	IL	60615	Canvass	Pass	33. FOOD A
2098922	2542098	BAR		550 W WASHINGTON BLVD	CHICAGO	IL	60661	License Re-Inspection	Fail	38. VENTIL
2098916	2511356	Restaurant		200 W 103RD ST	CHICAGO	IL	60628	Complaint	Pass	40. REFRIG
2098912	2215689	Daycare (U		5733-5743 W DIVISION ST	CHICAGO	IL	60651	License	Pass	38. VENTIL
2098906	35633	Restaurant		5311 S BLACKSTONE AVE	CHICAGO	IL	60615	Canvass	Pass	32. FOOD A
2098902	50551	Restaurant		5925 S PULASKI RD	CHICAGO	IL	60629	Canvass	Fail	33. FOOD A
2098889	2247067	Daycare A		3001-3009 W PIVERSEY AVE	CHICAGO	IL	60647	License Re-Inspection	Pass	24. DISH W.
2098900	2646406	Children's Services Facility		1 5331 N LINCOLN AVE	CHICAGO	IL	60625	License Re-Inspection	Fail	2. FACILITIE
22	69143	Grocery Store		1 5416 W 75TH ST	CHICAGO	IL	60621	Canvass	Pass with conditions	21. * CERTIF
23	2447171	Long Term Care		1 4437 S CICERO AVE	CHICAGO	IL	60632	Canvass	Fail	18. NO EVI
24	2470228	Children's Services Facility		1 5467 S UNIVERSITY AVE	CHICAGO	IL	60615	Canvass Re-Inspection	Pass	38. VENTIL
25	2098893	2215869 Daycare (Under 2 Years)		1 657 S WELLS ST	CHICAGO	IL	60607	License	Pass	34. FLOORS
26	2098768	2528979 Restaurant		3 200 E CERMAK RD	CHICAGO	IL	60616	License Re-Inspection	Pass	11. ADEQU
27	2098883	2142129 Grocery Store		1 3629 W 63RD ST	CHICAGO	IL	60629	Canvass	Pass	34. FLOORS
28	2098882	15790 Hospital		1 326 W 64TH ST	CHICAGO	IL	60621	Canvass	Pass	33. FOOD A
29	2098877	1768391 Long Term Care		1 10450 S MICHIGAN AVE	CHICAGO	IL	60628	Canvass	Pass	33. FOOD A
30	2098875	2560785 Restaurant		2 6336 S ASHLAND AVE	CHICAGO	IL	60636	License	Fail	11. ADEQU

### 3) Removal of Unwanted Columns

#### Code

```
rem_missing = rem_missing.drop(['DBA Name', 'AKA Name', 'Inspection Date', 'Month', 'Date'],
axis = 1)
```

#### Screenshot:

Food\_Inspections (1) - Excel

	A	B	C	D	
1	Inspection ID	DBA Name	AKA Name	License #	Facility Type
2	2099021	DS TEQUILA CO.	DS TEQUILA CO.	2542467	Restaurant
3	2099016	DS TEQUILA CO.	DS TEQUILA CO.	2542468	Restaurant
4	2099006	LA V BAKERY, INC	LA V BAKERY, INC	2550054	Bakery
5	2098969	GILCHRIST-MARCHMAN CHILD DEVELOPMENT CENTER	GILCHRIST-MARCHMAN CHILD DEVELOPMENT CENTER	2314038	Daycare Above and
6	2098967	FRANNIE'S CAFE	COLUMBIA COLLEGE / FRANNIE'S CAFE	2551098	Restaurant
7	2098958	PARADISE GROCERY INC.	PARADISE GROCERY INC.	2517736	Grocery Store
8	2098931	EL MUELLE	EL MUELLE	1488496	Restaurant
9	2098927	SAFA MARKET	SAFA MARKET	2283942	Grocery Store
10	2098956	EGGSPERIENCE	POTBELLY SANDWICH SHOP	2560093	Restaurant
11	2098954	POTBELLY SANDWICH SHOP	DAALO MEAT & GROCERY	2534897	Restaurant
12	2098955	Daalo Meat & Grocery, Inc	POKEWORKS	2013707	Grocery Store
13	2098953	POKEWORKS CHICAGO	A10	2560710	Restaurant
14	2098945	A10	MELLOW YELLOW	2241029	Restaurant
15	2098937	MELLOW YELLOW	TAQUERIA LA HACIENDA, INC.	1542	Restaurant
16	2098940	TAQUERIA LA HACIENDA, INC.	KUSANYA INC	1823377	Restaurant
17	2098934	KUSANYA INC	DON DE VIDA	2294418	Restaurant
18	2098926	DON DE VIDA	EAT & DRINK	2172653	Grocery Store
19	2098905	EAT & DRINK, INC	NATIVE FOODS	2069902	Restaurant
20	2098929	NATIVE FOODS CAFE	CO-OPTIM WEST LOOP, LLC.	2431737	Restaurant
21	2098922	CO-OPTIM WEST LOOP, LLC.	JIMMY'S BEST	2542098	BAR
22	2098923	CAFETERIA Y NEVERA JIMENEZ	MC KINNEY'S EARLY LEARNING ACADEMY	2560709	Restaurant
23	2098916	JIMMY'S BEST	GIORDANO'S PIZZA	2511356	Restaurant
24	2098912	MC KINNEY'S EARLY LEARNING ACADEMY	ZACATACOS	2215689	Daycare (Under 2 Years)
25	2098906	GIORDANO'S PIZZA	LA ESCUELA NUMERO DOS	35633	Restaurant
26	2098902	ZACATACOS	THE LEARNING BLOCK	50551	Restaurant
27	2098889	LA ESCUELA NUMERO DOS	A & J FOOD MARKET INCORPORATED	2247067	Daycare Above and
28	2098900	THE LEARNING BLOCK	SYMPHONY MIDWAY	2464806	Children's Services Facility
29	2098904	A & J FOOD MARKET INCORPORATED		69143	Grocery Store
30	2098894	SYMPHONY MIDWAY		2447171	Long Term Care

## After Cleaning

file\_clean - Excel

	B	C	D	E	F	G	H	I	J		
1	Inspection ID	License #	Facility Type	Risk	Address	City	State	Zip	Inspection Type	Results	Violations
2	2099016	2542468	Restaurant	1	3350-3352 N HALSTED ST	CHICAGO	IL	60657	License Re-Inspection	Fail	18. NO EVI
3	2099006	2550054	Bakery	2	2338 S KEDZIE AVE	CHICAGO	IL	60623	License Re-Inspection	Pass	18. NO EVI
4	2098969	2314038	Daycare Above and Under 2 Years	1	1312 S RACINE AVE	CHICAGO	IL	60608	License	Pass	31. CLEAN I
5	2098967	2551098	Restaurant	2	1104 S WABASH AVE	CHICAGO	IL	60605	License	Fail	38. VENTIL
6	2098958	2517736	Grocery Store	2	8101 S COTTAGE GROVE AVE	CHICAGO	IL	60619	Complaint	Fail	18. NO EVI
7	2098931	1488496	Restaurant	1	4253 W 47TH ST	CHICAGO	IL	60632	Canvas	Pass	38. VENTIL
8	2098927	2283942	Grocery Store	2	2319 W DEVON AVE	CHICAGO	IL	60659	Canvas	Fail	35. WALLS,
9	2098955	2013707	Grocery Store	2	2742 W Devon AVE	CHICAGO	IL	60659	Canvas Re-Inspection	Pass	14. PREVID
10	2098953	2560710	Restaurant	1	79 E MADISON ST	CHICAGO	IL	60602	License	Fail	12. HAND V
11	2098937	1542	Restaurant	1	1508 E 53RD ST	CHICAGO	IL	60615	Short Form Complaint	Pass	33. FOOD A
12	2098940	1823377	Restaurant	1	3452 W 59TH ST	CHICAGO	IL	60629	Canvas	Pass	32. FOOD A
13	2098934	2294418	Restaurant	1	825 W 69TH ST	CHICAGO	IL	60621	Canvas	Pass	35. WALLS,
14	2098929	2431737	Restaurant	1	1518 E HARPER CT	CHICAGO	IL	60615	Canvas	Pass	33. FOOD A
15	2098922	2542098	BAR	3	550 W WASHINGTON BLVD	CHICAGO	IL	60661	License Re-Inspection	Fail	38. VENTIL
16	2098916	2511356	Restaurant	1	200 W 103RD ST	CHICAGO	IL	60628	Complaint	Pass	40. REFRIG
17	2098912	2215689	Daycare (Under 2 Years)	1	5733-5743 W DIVISION ST	CHICAGO	IL	60651	License	Pass	38. VENTIL
18	2098906	35633	Restaurant	1	5311 S BLACKSTONE AVE	CHICAGO	IL	60615	Canvas	Pass	32. FOOD A
19	2098902	50551	Restaurant	1	5925 S PULASKI RD	CHICAGO	IL	60629	Canvas	Fail	33. FOOD A
20	2098889	2247067	Daycare Above and Under 2 Years	1	3001-3009 W DIVERSEY AVE	CHICAGO	IL	60647	License Re-Inspection	Pass	24. DISH W.
21	2098900	2464806	Children's Services Facility	1	5331 N LINCOLN AVE	CHICAGO	IL	60625	License Re-Inspection	Fail	2. FACILITIE
22	2098904	69143	Grocery Store	1	342 W 75TH ST	CHICAGO	IL	60621	Canvas	Pass with conditions	21. * CERTI
23	2098894	2447171	Long Term Care	1	4437 S CICERO AVE	CHICAGO	IL	60632	Canvas	Fail	18. NO EVI
24	2098892	2470228	Children's Services Facility	1	5467 S UNIVERSITY AVE	CHICAGO	IL	60615	Canvas Re-Inspection	Pass	38. VENTIL
25	2098893	2215869	Daycare (Under 2 Years)	1	657 S WELLS ST	CHICAGO	IL	60607	License	Pass	34. FLOORS
26	2098768	2528979	Restaurant	3	200 E CERMACK RD	CHICAGO	IL	60616	License Re-Inspection	Pass	11. ADEQU
27	2098883	2142129	Grocery Store	1	3629 W 63RD ST	CHICAGO	IL	60629	Canvas	Pass	34. FLOORS
28	2098882	15790	Hospital	1	326 W 64TH ST	CHICAGO	IL	60621	Canvas	Pass	33. FOOD A
29	2098877	1768391	Long Term Care	1	10450 S MICHIGAN AVE	CHICAGO	IL	60628	Canvas	Pass	33. FOOD A
30	2098875	2560785	Restaurant	2	6336 S ASHLAND AVE	CHICAGO	IL	60636	License	Fail	11. ADEQU

#### **4) Splitting of Column**

Splitting violation number and violation description for calculation how many time same violation occurred from 2010 – 2017 by counting unique violation no.

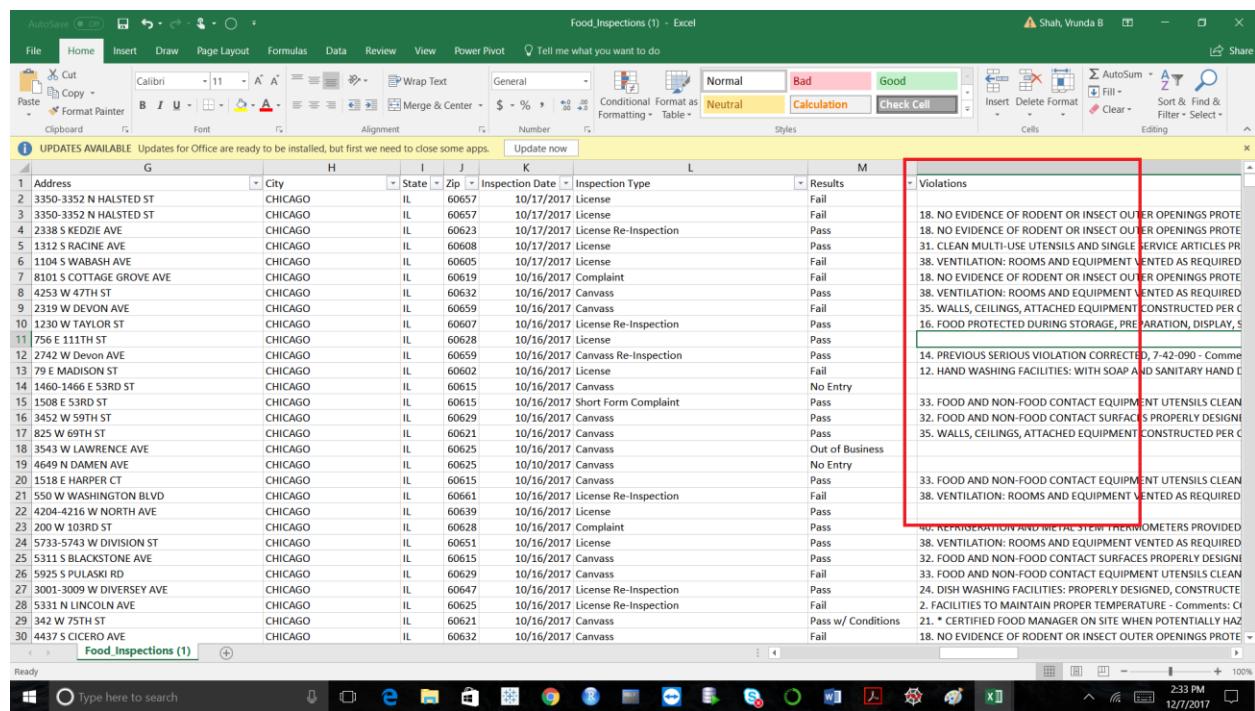
We will be using Violation Description to identify top 10 violations that took place from 2010- 2017. We have used *string function SPLIT()*

#### **TECHNIQUE USED – STRING FUNCTIONS**

#### **Code**

```
rem_missing[['Num', 'ViolationDescription']] = rem_missing['Violations'].str.split('.', 1,expand = True)
```

#### **Screenshot**



G	H	I	J	K	L	M
Address	City	State	Zip	Inspection Date	Inspection Type	Results
2 3350-3352 N HALSTED ST	CHICAGO	IL	60657	10/17/2017	License	Fail
3 3350-3352 N HALSTED ST	CHICAGO	IL	60657	10/17/2017	License	Fail
4 2338 S KEDZIE AVE	CHICAGO	IL	60623	10/17/2017	License Re-Inspection	Pass
5 1312 S RACINE AVE	CHICAGO	IL	60608	10/17/2017	License	Pass
6 1104 S WABASH AVE	CHICAGO	IL	60605	10/17/2017	License	Fail
7 8101 S COTTAGE GROVE AVE	CHICAGO	IL	60619	10/16/2017	Complaint	Fail
8 4253 W 47TH ST	CHICAGO	IL	60632	10/16/2017	Canvass	Pass
9 2319 W DEVON AVE	CHICAGO	IL	60659	10/16/2017	Canvass	Fail
10 1230 W TAYLOR ST	CHICAGO	IL	60607	10/16/2017	License Re-Inspection	Pass
11 756 E 111TH ST	CHICAGO	IL	60628	10/16/2017	License	Pass
12 2742 W Devon AVE	CHICAGO	IL	60659	10/16/2017	Canvass Re-Inspection	Pass
13 79 E MADISON ST	CHICAGO	IL	60602	10/16/2017	License	Fail
14 1460-1466 E 53RD ST	CHICAGO	IL	60615	10/16/2017	Canvass	No Entry
15 1508 E 53RD ST	CHICAGO	IL	60615	10/16/2017	Short Form Complaint	Pass
16 3452 W 59TH ST	CHICAGO	IL	60629	10/16/2017	Canvass	Pass
17 825 W 69TH ST	CHICAGO	IL	60621	10/16/2017	Canvass	Pass
18 3543 W LAWRENCE AVE	CHICAGO	IL	60625	10/16/2017	Canvass	Out of Business
19 4649 N DAMEN AVE	CHICAGO	IL	60625	10/10/2017	Canvass	No Entry
20 1518 E HARPER CT	CHICAGO	IL	60615	10/16/2017	Canvass	Pass
21 550 W WASHINGTON BLVD	CHICAGO	IL	60661	10/16/2017	License Re-Inspection	Fail
22 4204-4216 W NORTH AVE	CHICAGO	IL	60639	10/16/2017	License	Pass
23 200 W 103RD ST	CHICAGO	IL	60628	10/16/2017	Complaint	Pass
24 5733-5743 W DIVISION ST	CHICAGO	IL	60651	10/16/2017	License	Pass
25 5311 S BLACKSTONE AVE	CHICAGO	IL	60615	10/16/2017	Canvass	Pass
26 5925 S PULASKI RD	CHICAGO	IL	60629	10/16/2017	Canvass	Fail
27 3001-3009 W DIVERSEY AVE	CHICAGO	IL	60647	10/16/2017	License Re-Inspection	Pass
28 5331 N LINCOLN AVE	CHICAGO	IL	60625	10/16/2017	License Re-Inspection	Fail
29 342 W 75TH ST	CHICAGO	IL	60621	10/16/2017	Canvass	Pass w/ Conditions
30 4437 S CICERO AVE	CHICAGO	IL	60632	10/16/2017	Canvass	Fail

	P	ViolationDescription	Num
1			18
2		IVITY IN BASEMENT IN CORNER BEHIND STEAMER. NOTED APPROX. 35 SMALL FLIES ON WALLS AND CEILIN	18 NO EVIDENCE OF RODENT OR INSECT OUTER OPEN
3		41.94331838 -87.64951476 (41.94331835640129, -87.64951475556387)	18 NO EVIDENCE OF RODENT OR INSECT OUTER OPEN
4		41.8487905 -87.70531098 (41.84879049508066, -87.70531098155864)	201
5		JUNTS. MUST STORE INVERTED.   32. FOOD AND NON-FOOD CONTACT SURFACES PROPERLY DESIGNED, C	31. CLEAN MULTI-USE UTENSILS AND SINGLE SERVICE
6		41.8645873 -87.65767659 (41.8645873399433, -87.65676659479097)	38. VENTILATION: ROOMS AND EQUIPMENT VENTED A
7		e does not work. Must repair and maintain.	18 NO EVIDENCE OF RODENT OR INSECT OUTER OPEN
8		41.86906252 -87.7260159 (41.86906251655677, -87.726015982796)	38. VENTILATION: ROOMS AND EQUIPMENT VENTED A
9		JOT RODENT PROOF AT THE BOTTOM. FOUND A 1/4 INCH GAP AT THE BOTTOM OF THE DOOR; MUST	18 NO EVIDENCE OF RODENT OR INSECT OUTER OPEN
10		41.7473845 -87.60498573 (41.74738448447344, -87.6049857308018)	38. VENTILATION: ROOMS AND EQUIPMENT VENTED A
11		IS, ATTACHED EQUIPMENT CONSTRUCTED PER CODE: GOOD REPAIR, SURFACES CLEAN AND DUST-LESS CL	35. WALLS, CEILINGS, ATTACHED EQUIPMENT CONST
12		8  S THROUGH OUT THE PREMISES, CRACK WALL BY HAND WASH SINK, WALL BEHIND THE MEAT PREP	14. PREVIOUS SERIOUS VIOLATION CORRECTED, 7-4-14
13		9 'ENTIN DEVICE. #41,   32. FOOD AND NON-FOOD CONTACT SURFACES PROPERLY DESIGNED, CONSTRUC	12. HAND WASHING FACILITIES: WITH SOAP AND SABU
14		10 INKS NOR IN THE WASHROOM. INSTRUCTED TO PROVIDE PAPER TOWELS AND MAINTAIN. CRITICAL VIOLA	33. FOOD AND NON-FOOD CONTACT EQUIPMENT UTE
15		11. PREP TABLES IN FIRST FLOOR AND BASEMENT PREP AREAS, SHELVES IN BASEMENT DRY STORAGE AREAS,	32. FOOD AND NON-FOOD CONTACT SURFACES PROPE
16		12 MOVE THE RUST/REPLACE.   33. FOOD AND NON-FOOD CONTACT EQUIPMENT UTENSILS CLEAN, FREE OF	35. WALLS, CEILINGS, ATTACHED EQUIPMENT CONST
17		13 THE CEILING OF THE WASHROOM; CLEAN VENTILATION FAN TO REMOVE DUST OBSERVED.   38. VENTILAT	33. FOOD AND NON-FOOD CONTACT EQUIPMENT UTE
18		14 CODE, CLEANED, GOOD REPAIR, COVING INSTALLED, DUST-LESS CLEANING METHODS USED - Comments: C	38. VENTILATION: ROOMS AND EQUIPMENT VENTED A
19		15 tion device.   40. REFRIGERATION AND METAL STEM THERMOMETERS PROVIDED AND CONSPICUOUS - Co	40. REFRIGERATION AND METAL STEM THERMOMETE
20		16 ATION: ROOMS AND EQUIPMENT VENTED AS REQUIRED: PLUMBING: INSTALLED AND MAINTAINED - Com	38. VENTILATION: ROOMS AND EQUIPMENT VENTED A
21		17 S ROOM HAND SINKS.	32. FOOD AND NON-FOOD CONTACT SURFACES PROPE
22		18 LLS, CEILINGS, ATTACHED EQUIPMENT CONSTRUCTED PER CODE: GOOD REPAIR, SURFACES CLEAN AND D	33. FOOD AND NON-FOOD CONTACT EQUIPMENT UTE
23		19 ITIZ.   29. PREVIOUS MINOR VIOLATION(S) CORRECTED 7-42-090 - Comments: OBSERVED PREVIOUS VIOL	24. DISH WASHING FACILITIES: PROPERLY DESIGNED C
24		20 RAINING PROPERLY.   18. NO EVIDENCE OF RODENT OR INSECT OUTER OPENINGS PROTECTED/RODENT I	2. FACILITIES TO MAINTAIN PROPER TEMPERATURE
25		21 NED - Comments: CORRECTED (EXPOSED HAND SINK HAS BEEN INSTALLED WITH HAND SOAP HAND DRINK	21. * CERTIFIED FOOD MANAGER ON SITE WHEN PO
26		22. MANAGER STATES THAT HE TOOK THE CLASS TO RENEW HIS CITY OF CHICAGO FOOD SANITATION	18. NO EVIDENCE OF RODENT OR INSECT OUTER OPEN
27		23 ON WALLS, DOOR FRAME AND CEILING IN PREP AREA. INSTD TO ELIMINATE FLIES, CLEAN AND SANITIZE	38. VENTILATION: ROOMS AND EQUIPMENT VENTED A
28		24.22.	34. FLOORS: CONSTRUCTED PER CODE, CLEANED, GOO
29		25. CEILINGS, ATTACHED EQUIPMENT CONSTRUCTED PER CODE: GOOD REPAIR, SURFACES CLEAN AND DUST	11. ADEQUATE NUMBER, CONVENIENT, ACCESSIBLE, D
30		26. ED TO REMOVE.   35. WALLS, CEILINGS, ATTACHED EQUIPMENT CONSTRUCTED PER CODE: GOOD REPAIR,	34. FLOORS: CONSTRUCTED PER CODE, CLEANED, GOO
		27. CE MACHINE HAS SLIGHT BUILD-UP.MUST CLEAN/MAINTAIN.   35. WALLS, CEILINGS, ATTACHED EQUIPM	33. FOOD AND NON-FOOD CONTACT EQUIPMENT UTE
		28. EASE.	11. ADEQUATE NUMBER, CONVENIENT, ACCESSIBLE, D
		29. CE MACHINE HAS SLIGHT BUILD-UP.MUST CLEAN/MAINTAIN.   35. WALLS, CEILINGS, ATTACHED EQUIPM	33. FOOD AND NON-FOOD CONTACT EQUIPMENT UTE
		30. EXPOSED HANDSINK IN SAME AREA. VIOLATION 7-38-030 CRITICAL.   38. VENTILATION: ROOMS AP	11. ADEQUATE NUMBER, CONVENIENT, ACCESSIBLE, D

## 5) Refining Data via File System

```
rem_missing1 = rem_missing['Inspection Type']
```

```
# Dropping of columns
```

```
rem_missing = rem_missing.drop(['DBA Name', 'AKA Name', 'Inspection Date', 'Month', 'Date'],
```

```
axis = 1)
```

```
print (rem_missing.head())
```

```
rem_missing.to_csv('file_clean.csv', index = False)
```

```
np.savetxt('a.txt', rem_missing1, fmt='%5s', delimiter = ',', newline = '\n')
```

```
fhand = open('b.txt')
```

```
count = 0
```

```
d = dict()

for line in fhand:

    line = line.rstrip() # strip off additional line space

    line1 = line.split(",")

    for temp in line1:

        print (temp)

        if temp not in d:

            d[temp] = 1

        else:

            d[temp] = d[temp] + 1

print(d)
```

```
lst = list(d.keys())

print(lst)

lst.sort()

for key in lst:

    print(key, d[key])
```

```
import pandas as pd

dict1 = d

df = pd.DataFrame(data=dict1, index=[0])
```

```

df = (df.T)

print (df)

df.to_csv('dict1.csv')

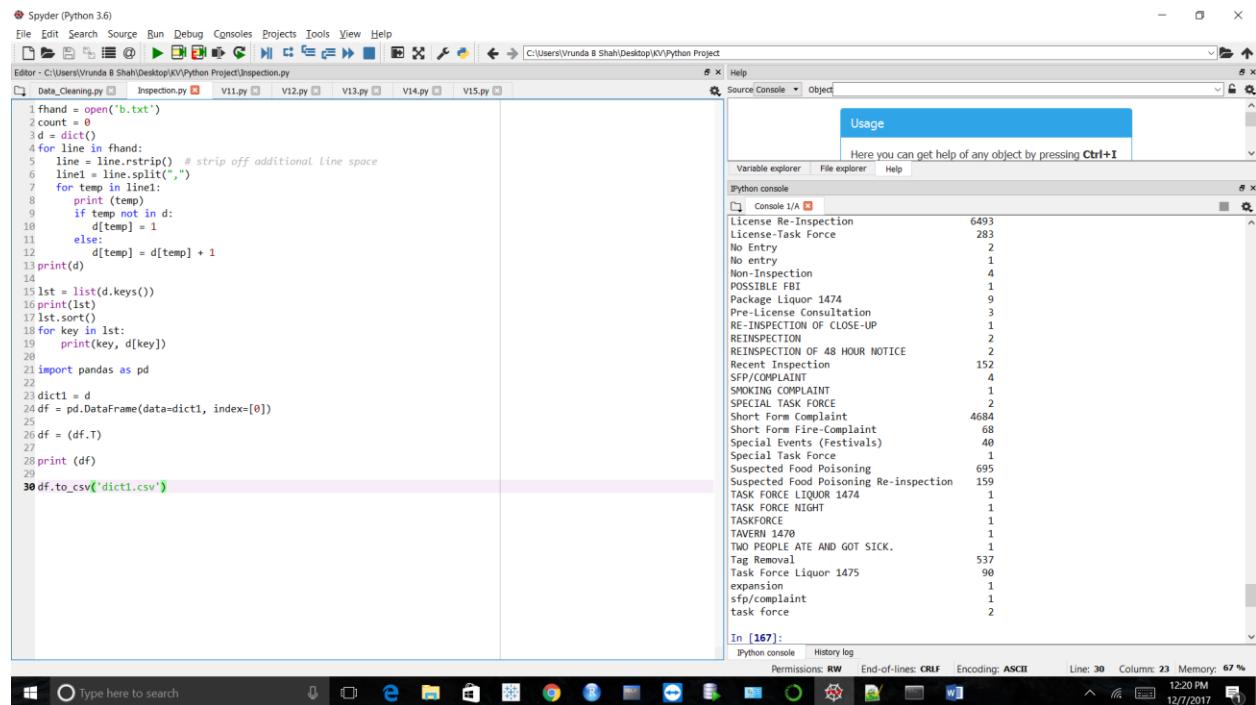
```

## SUMMARY

We have used *list*, *dictionary* to get number of times particular ‘Inspection Type’ was requested from year 2010 – 2017.

It will help us to understand which Inspection type was requested maximum and least during the tenure of 7 years. We will be further saving end result in different excel sheet(*Usage of file system*), which will be used later in visualizing top 10 violations.

## TECHNIQUE USED – FILES, LIST, DICTIONARY



The screenshot shows the Spyder Python 3.6 IDE interface. The code editor contains a script named Data\_Cleaning.py with the following content:

```

1 fhand = open('b.txt')
2 count = 0
3 d = dict()
4 for line in fhand:
5     line = line.rstrip() # strip off additional Line space
6     line1 = line.split(',')
7     for temp in line1:
8         print (temp)
9         if temp not in d:
10             d[temp] = 1
11         else:
12             d[temp] = d[temp] + 1
13 print(d)
14
15 lst = list(d.keys())
16 print(lst)
17 lst.sort()
18 for key in lst:
19     print(key, d[key])
20
21 import pandas as pd
22
23 dict1 = d
24 df = pd.DataFrame(data=dict1, index=[0])
25
26 df = (df.T)
27
28 print (df)
29
30 df.to_csv('dict1.csv')

```

The IPython console shows the output of the code, which is a dictionary of inspection types and their counts. A 'Usage' window is also visible, providing help for the object 'd'.

Inspection Type	Count
LICENSE RE-INSPECTION	6493
LICENSE-TASK FORCE	283
No Entry	2
No entry	1
Non-Inspection	4
POSSIBLE FBI	1
PACKAGE LIQUOR 1474	9
PRE-LICENSE CONSULTATION	3
RE-INSPECTION OF CLOSE-UP	1
REINSPECTION	2
REINSPECTION OF 48 HOUR NOTICE	2
Recent Inspection	152
SFP/COMPLAINT	4
SMOKING COMPLAINT	1
TASK FORCE	2
Short Form Complaint	4684
Short Form Fire-Complaint	68
Special Events (Festivals)	40
Special Task Force	1
Suspected Food Poisoning	695
Suspected Food Poisoning Re-inspection	159
TASK FORCE LIQUOR 1474	1
TASK FORCE NIGHT	1
TASKFORCE	1
TAVERN 1470	1
TWO PEOPLE ATE AND GOT SICK.	1
Tag Removal	537
Task Force Liquor 1475	90
expansion	1
sfp/complaint	1
task force	2

```
Command Prompt
Illegal Operation 1
KIDS CAFE 1
LICENSE DAYCARE 1586 1
LICENSE RENEWAL FOR DAYCARE 2
LICENSE RENEWAL INSPECTION FOR DAYCARE 1
LIQUOR CATERING 1
License 11961
License Re-Inspection 6493
License-Task Force 283
No Entry 2
No entry 1
Non-Inspection 4
POSSIBLE FBI 1
Package Liquor 1474 9
Pre-License Consultation 3
RE-INSPECTION OF CLOSE-UP 1
REINSPECTION 2
REINSPECTION OF 48 HOUR NOTICE 2
Recent Inspection 152
SFP/COMPLAINT 4
SMOKING COMPLAINT 1
SPECIAL TASK FORCE 2
Short Form Complaint 4684
Short Form Fire-Complaint 68
Special Events (Festivals) 40
Special Task Force 1
Suspected Food Poisoning 695
Suspected Food Poisoning Re-inspection 159
TASK FORCE LIQUOR 1474 1
TASK FORCE NIGHT 1
TASKFORCE 1
TAVERN 1478 1
TWO PEOPLE ATE AND GOT SICK. 1
Tag Removal 587
Task Force Liquor 1475 90
expansion 1
sfp/complaint 1
task force 2

C:\Users\Vrunuda B Shah\Desktop\KV\Python Project>
```

## C) STATISTICS SUMMARY and FUNCTION

### Code

```
import pandas as pd  
df = pd.read_csv('file_clean.csv', index_col=0)
```

```
def maxrisk(l):
```

```
    #count = 0
```

```
    d = dict()
```

```
    for data in l:
```

```
        if data not in d:
```

```
            d[data] = 1
```

```
        else:
```

```
            d[data] = d[data] + 1
```

```
    print(d)
```

```
lst = list(d.keys())
```

```
print(lst)
```

```
lst.sort()
```

```
for key in lst:
```

```
    print(key, d[key])
```

```
return maxrisk

num1 = maxrisk(df['Risk'])

#print('Total Summary of risk',num1 )

mean_risk = df['Risk'].mean()

print(mean_risk)

print(df['Risk'].describe())
```

## **SUMMARY**

### **TECHNIQUE USED – Function, Dictionary and describe()**

We have used ***dictionary, function, and describe()*** and We have calculated average summary for Risk column, which has risk data associated with Inspection Type. Risk is categorized by 1 – High, 2 – Medium and 3 – Low. We have removed summary of Risk from year 2010 – 2017.

Using ***function*** we have also identified, how many Inspections had Risk 1 - High, Risk 2 – Medium, Risk 3 – Low.

### **Raw Output of Function**

{1.0: 91032, 2.0: 24864, 3.0: 7472}

[1.0, 2.0, 3.0]

1.0 - 91032

2.0 - 24864

3.0 – 7472

From the raw output of function, we have found majority of inspection requested was for violations which had risk category 2.

### **Raw Statistics Output**

count - 123368.000000

mean - 1.322677

std - 0.582832

min - 1.000000

25% - 1.000000

50% - 1.000000

75% - 2.000000

Max - 3.000000

### **SCREENSHOT**

Spyder (Python 3.6)

```

1 import pandas as pd
2 df = pd.read_csv('file_clean.csv', index_col=0)
3 #creating function maxval.
4
5 def maxrisk(l):
6     #count = 0
7     d = dict()
8     for data in l:
9         if data not in d:
10             d[data] = 1
11         else:
12             d[data] = d[data] + 1
13     print(d)
14
15 lst = list(d.keys())
16 print(lst)
17 lst.sort()
18 for key in lst:
19     print(key, d[key])
20
21 num1 = maxrisk(df['Risk'])
22 #print('Total Summary of risk',num1 )
23
24
25

```

In [193]:

```

In [194]: runfile('C:/Users/Vrunda B Shah/Desktop/KV/Python Project/V15.py', wdir='C:/Users/Vrunda B Shah/Desktop/KV/Python Project')
(1.0: 91032, 2.0: 24864, 3.0: 7472)
[1.0, 2.0, 3.0]
1.0 91032
2.0 24864
3.0 7472

```

In [193]:

IPython console History log Permissions: RW End-of-lines: CRLF Encoding: UTF-8 Line: 25 Column: 1 Memory: 75 % 12:55 PM 12/7/2017

Spyder (Python 3.6)

```

1 import pandas as pd
2 df = pd.read_csv('file_clean.csv', index_col=0)
3 #creating function maxval.
4
5 def maxrisk(l):
6     #count = 0
7     d = dict()
8     for data in l:
9         if data not in d:
10             d[data] = 1
11         else:
12             d[data] = d[data] + 1
13     print(d)
14
15 lst = list(d.keys())
16 print(lst)
17 lst.sort()
18 for key in lst:
19     print(key, d[key])
20
21 num1 = maxrisk(df['Risk'])
22 #print('Total Summary of risk',num1 )
23
24 mean_risk = df['Risk'].mean()
25 print(mean_risk)
26
27 print(df['Risk'].describe())
28
29

```

In [201]:

```

In [201]: runfile('C:/Users/Vrunda B Shah/Desktop/KV/Python Project/V15.py', wdir='C:/Users/Vrunda B Shah/Desktop/KV/Python Project')
(1.0: 91032, 2.0: 24864, 3.0: 7472)
[1.0, 2.0, 3.0]
1.0 91032
2.0 24864
3.0 7472
1.3226768692043118
count    123360.000000
mean      1.332677
std       0.582832
min      1.000000
25%     1.000000
50%     1.000000
75%     2.000000
max      3.000000
Name: Risk, dtype: float64

```

In [202]:

IPython console History log Permissions: RW End-of-lines: CRLF Encoding: UTF-8 Line: 29 Column: 1 Memory: 72 % 10:06 PM 12/7/2017

## D) ANALYSIS AND VISUALISATIONS

### Question1 :

Identify Top 10 'Inspection Type' requested during tenure of 2010 – 2017?

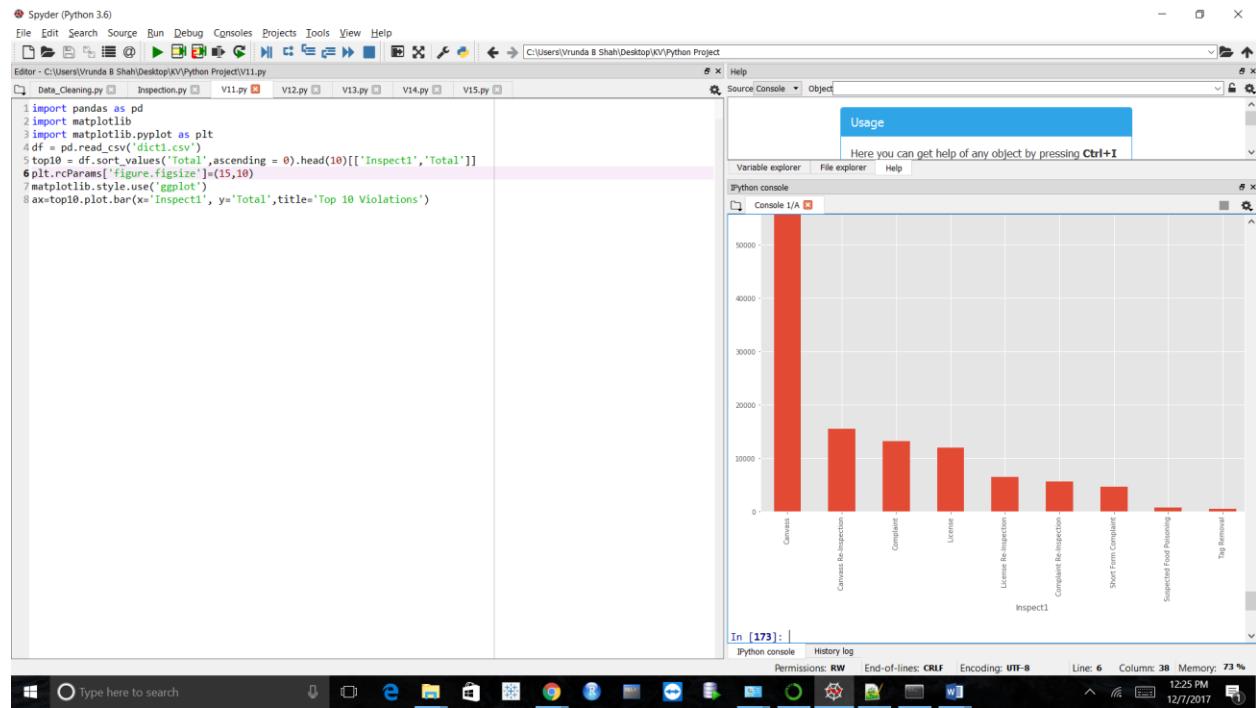
## Analysis

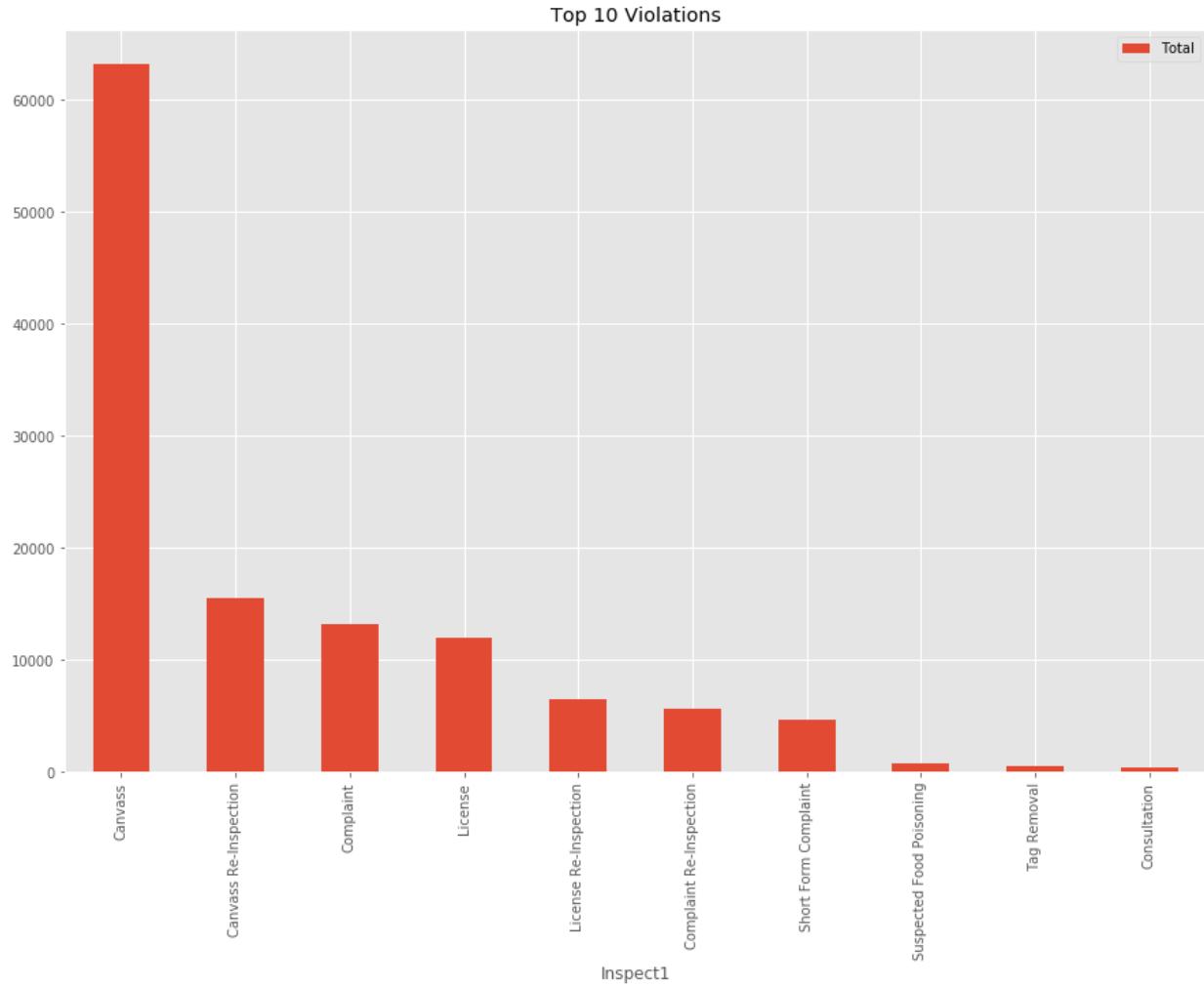
We have found Canvass, Canvass – Reinspection, Complaint, License, License Re -Inspection, Complaint Re-Inspection, Tag removal, Short for complaint, Suspected Food Poisoning were the Top 10 Inspections requested. We have used Bar Chart.

Canvass Inspection was highly requested and Tag Removal was least requested. We also found License Inspection and Re-Inspection was also frequently Inspected.

## TECHNIQUE USED – MATPLOT, PANDAS, DATAFRAME

## VISUALISATION





## Question 2

Find out frequency distribution of risk associated with Inspection Type ?

## Analysis

Majority of the inspection requested had Risk associated in the range of Risk 1 and the least associated risk range is falls under Risk 3.

We have used Histogram

## TECHNIQUE USED – MATPLOT, PYPLOT

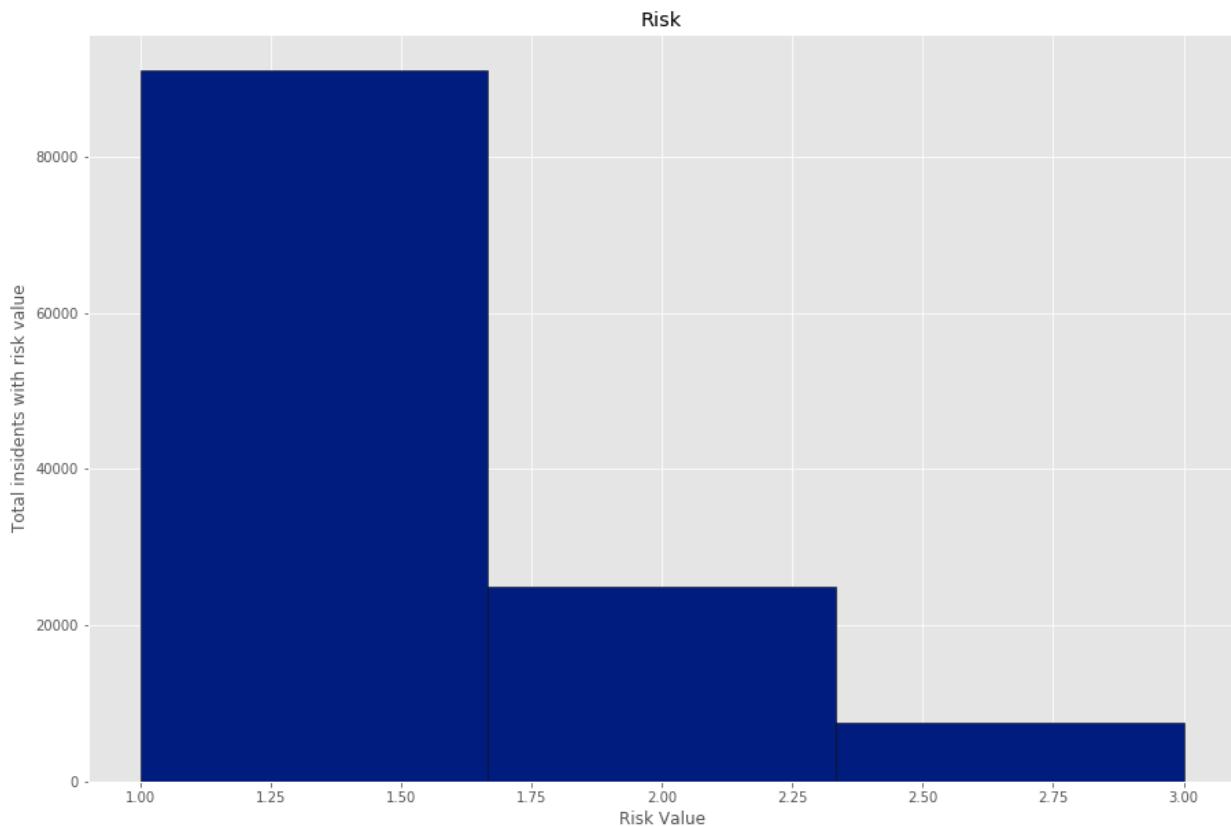
## VISUALISATION

The screenshot shows the Spyder Python IDE interface. The code editor contains the following Python script:

```
1 import pandas as pd
2 import seaborn as sns
3 import matplotlib
4 import matplotlib.pyplot as plt
5 df = pd.read_csv('file_clean.csv', index_col=0)
6 matplotlib.style.use('seaborn-dark-palette') # plot style
7 df.hist(column='Risk', edgecolor='black', bins = 3) #color and column
8 plt.xlabel('Risk Value') #axis label
9 plt.ylabel('Total incidents with risk value') #axis label
10 plt.suptitle('Frequency Distribution of Average Risk for Food Inspection from 2010 - 2017', ha='center')
11
```

The IPython console displays a histogram titled "Risk" with the subtitle "Frequency Distribution of Average Risk for Food Inspection from 2010 - 2017". The x-axis is labeled "Risk Value" and ranges from 1.00 to 3.00. The y-axis is labeled "Total incidents with risk value" and ranges from 0 to 80,000. The histogram has three major bars: one from 1.00 to 1.75 (~85,000), one from 1.75 to 2.50 (~25,000), and one from 2.50 to 3.00 (~10,000). The plot area has a light gray background with white grid lines.

Frequency Distribution of Average Risk for Food Inspection from 2010 - 2017



### **Question 3**

Find how many percent of Inspection type has result Pass, Pass with conditions and Fail?

### **Analysis**

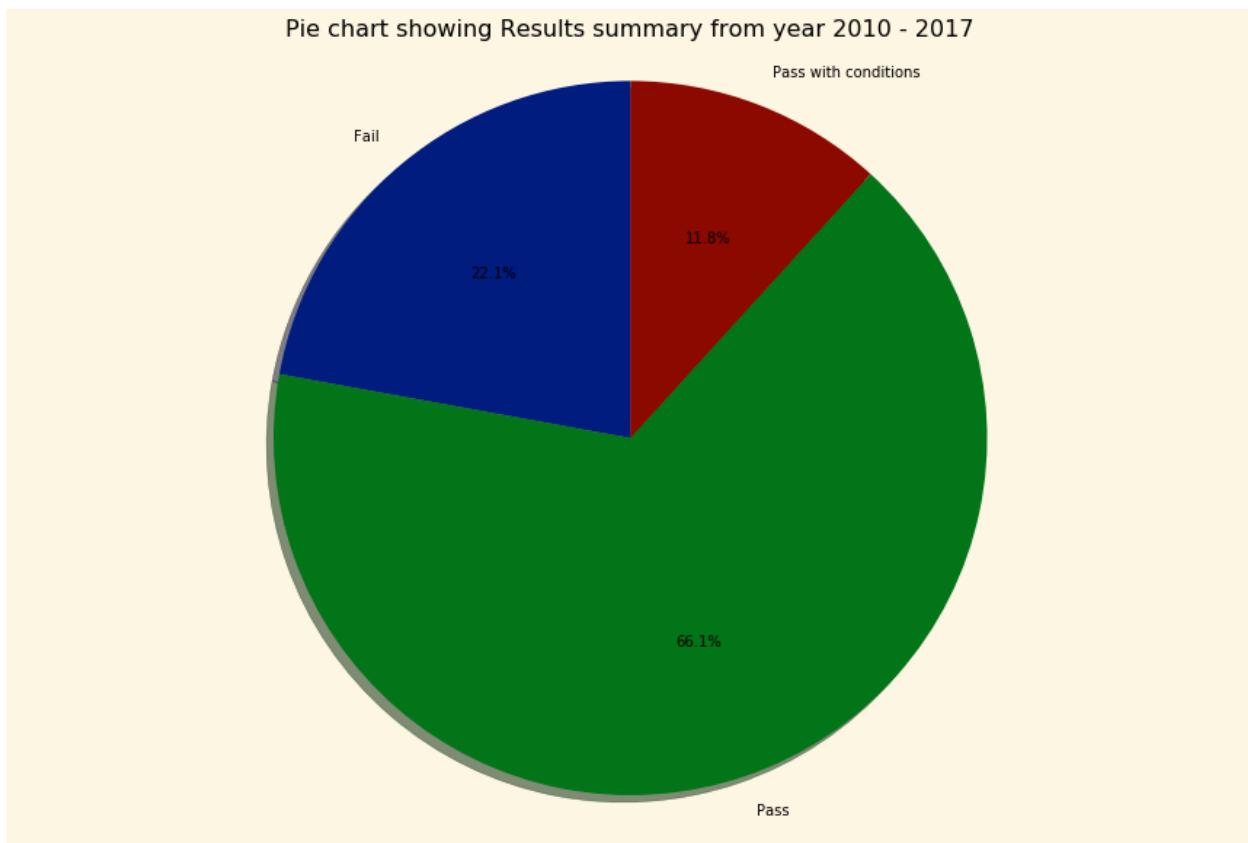
We have identified majority of the Inspection that took place in tenure of 2010 – 2017 has Pass status (66.1 %), 22.1% has outcome Pass with conditions and 11.8% of Inspection results falls under Fail category. We have used Pie chart.

**Pass - 66.1**

**Pass With Conditions - 22.1**

**Fail - 11.8**

**TECHNIQUE USED: List, pandas, numpy**



Spyder (Python 3.6)

File Edit Search Source Run Debug Consoles Projects Tools View Help

Editor - C:\Users\Yrunda B Shah\Desktop\KV\Python Project\V14.py

```

1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 df = pd.read_csv('file_clean.csv',index_col=0)
5 res_count=df.groupby(['Results']).size().reset_index(name='count')
6 print(res_count)
7 nparray_count=res_count['count'].values
8 list_count=np.array(nparray_count).tolist()
9 nparray_reses_count['Results'].values
10 list_reses=np.array(nparray_res).tolist()
11 plt.pie(list_count,labels=list_reses,autopct='%1.1f%%', shadow=True, startangle=90)
12 plt.axis('equal')
13 plt.title('Pie chart showing Results summary from year 2010 - 2017')
14 plt.show()
15

```

Usage  
Here you can get help of any object by pressing **Ctrl+I**

Variable explorer File explorer Help

IPython console

Console 1/A

	Results count
0	Fail 27307
1	Pass 81551
2	Pass with conditions 14510

Pie chart showing Results summary from year 2010 - 2017

Category	Percentage
Pass	66.1%
Fail	22.1%
Pass with conditions	11.8%

Type here to search

## [E\] Code](#)

### ➤ Data Cleaning

```
import pandas as pd

import numpy as np

data1 = pd.read_csv('Food_Inspections.csv')

data1.replace('Risk 2 (Medium)', '2')

# Removal of NA and null Values

rem_missing = data1.dropna()

rem_missing[rem_missing.notnull()]

# Splitting Inspection Date to access only year data from it - String Functions

rem_missing[['Month','Date','Year']] = rem_missing['Inspection Date'].str.split('/', expand = True)

rem_missing[['Num', 'ViolationDescription']] = rem_missing['Violations'].str.split('.', 1,expand = True)

# Adding Inspection Type into new variable to save it in text file for future use

rem_missing1 = rem_missing['Inspection Type']

# Dropping of columns

rem_missing = rem_missing.drop(['DBA Name', 'AKA Name', 'Inspection Date', 'Month', 'Date'], axis = 1)

print (rem_missing.head())

rem_missing.to_csv('file_clean.csv', index = False)
```

```
np.savetxt('a.txt', rem_missing1, fmt='%5s', delimiter = ',', newline = '\n')

np.savetxt('b.txt', rem_missing1, fmt='%5s', delimiter = ',', newline = ',')
```

➤ **Data Refinement Using File System**

```
fhand = open('b.txt')

count = 0

d = dict()

for line in fhand:

    line = line.rstrip() # strip off additional line space

    line1 = line.split(",")

    for temp in line1:

        print (temp)

        if temp not in d:

            d[temp] = 1

        else:
```

```
            d[temp] = d[temp] + 1

print(d)
```

```
lst = list(d.keys())

print(lst)

lst.sort()
```

```
for key in lst:
```

```
    print(key, d[key])
```

```
import pandas as pd
```

```
dict1 = d
```

```
df = pd.DataFrame(data=dict1, index=[0])
```

```
df = (df.T)
```

```
print (df)
```

```
df.to_csv('dict1.csv')
```

➤ **Function**

```
import pandas as pd
```

```
df = pd.read_csv('file_clean.csv', index_col=0)
```

```
#creating function maxval
```

```
def maxrisk(l):
```

```
    #count = 0
```

```
    d = dict()
```

```
    for data in l:
```

```
        if data not in d:
```

```
            d[data] = 1
```

```
else:  
    d[data] = d[data] + 1  
  
print(d)  
  
  
lst = list(d.keys())  
  
print(lst)  
  
lst.sort()  
  
for key in lst:  
    print(key, d[key])  
  
return maxrisk  
  
num1 = maxrisk(df['Risk'])
```

➤ **Function + Summary**

```
import pandas as pd  
  
df = pd.read_csv('file_clean.csv', index_col=0)
```

```
def maxrisk(l):  
    #count = 0  
  
    d = dict()  
  
    for data in l:  
        if data not in d:
```

```
d[data] = 1

else:
    d[data] = d[data] + 1

print(d)

lst = list(d.keys())
print(lst)
lst.sort()

for key in lst:
    print(key, d[key])

return maxrisk

num1 = maxrisk(df['Risk'])

#print('Total Summary of risk',num1 )

mean_risk = df['Risk'].mean()

print(mean_risk)

print(df['Risk'].describe())

➤ Visualisation1

import pandas as pd

import matplotlib
```

```
import matplotlib.pyplot as plt  
  
df = pd.read_csv('dict1.csv')  
  
top10 = df.sort_values('Total', ascending = 0).head(10)[['Inspect1','Total']]  
  
plt.rcParams['figure.figsize']=(15,10)  
  
matplotlib.style.use('ggplot')  
  
ax=top10.plot.bar(x='Inspect1', y='Total',title='Top 10 Inspections')
```

### ➤ Visualisation 2

```
import pandas as pd  
  
#import pylab as plt  
  
import matplotlib  
  
import matplotlib.pyplot as plt  
  
df =pd.read_csv('file_clean.csv',index_col=0)  
  
matplotlib.style.use('seaborn-dark-palette') # plot style  
  
df.hist(column='Risk',edgecolor='black', bins = 3) #color and column  
  
plt.xlabel('Risk Value') #xaxis label  
  
plt.ylabel('Total incidents with risk value') #yaxis label  
  
plt.suptitle('Frequency Distribution of Average Risk for Food Inspection from 2010 - 2017',  
ha='center')
```

### ➤ Visualisation 3

```
import pandas as pd
```

```
import numpy as np

import matplotlib.pyplot as plt

df=pd.read_csv('file_clean.csv',index_col=0)

res_count=df.groupby(['Results']).size().reset_index(name='count')

print(res_count)

nparray_count=res_count['count'].values

list_count=np.array(nparray_count).tolist()

nparray_res=res_count['Results'].values

list_res=np.array(nparray_res).tolist()

plt.pie(list_count,labels=list_res,autopct='%1.1f%%', shadow=True, startangle=90)

plt.axis('equal')

plt.title('Pie chart showing Results summary from year 2010 - 2017')

plt.show()
```