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 then 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.

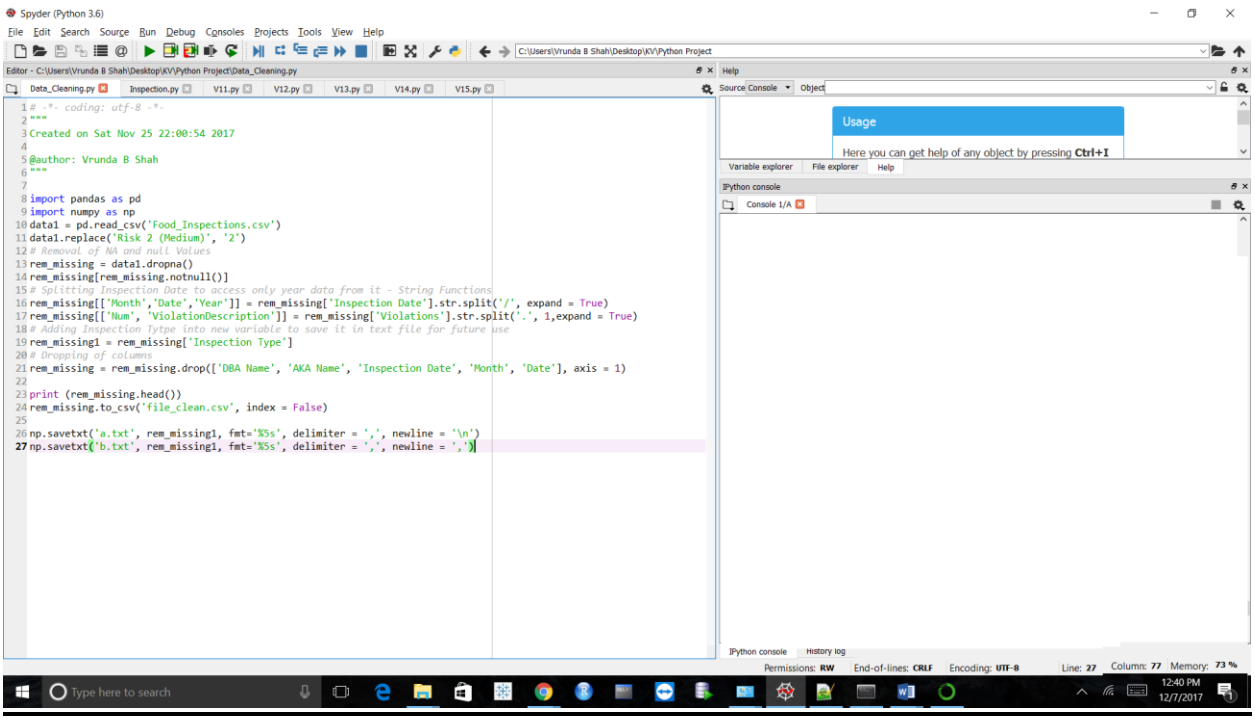
vi) **Suspect food poisoning:** This inspection is done when customers files a complaint against establishment of falling ill after eating food at establishment location like restaurants.

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



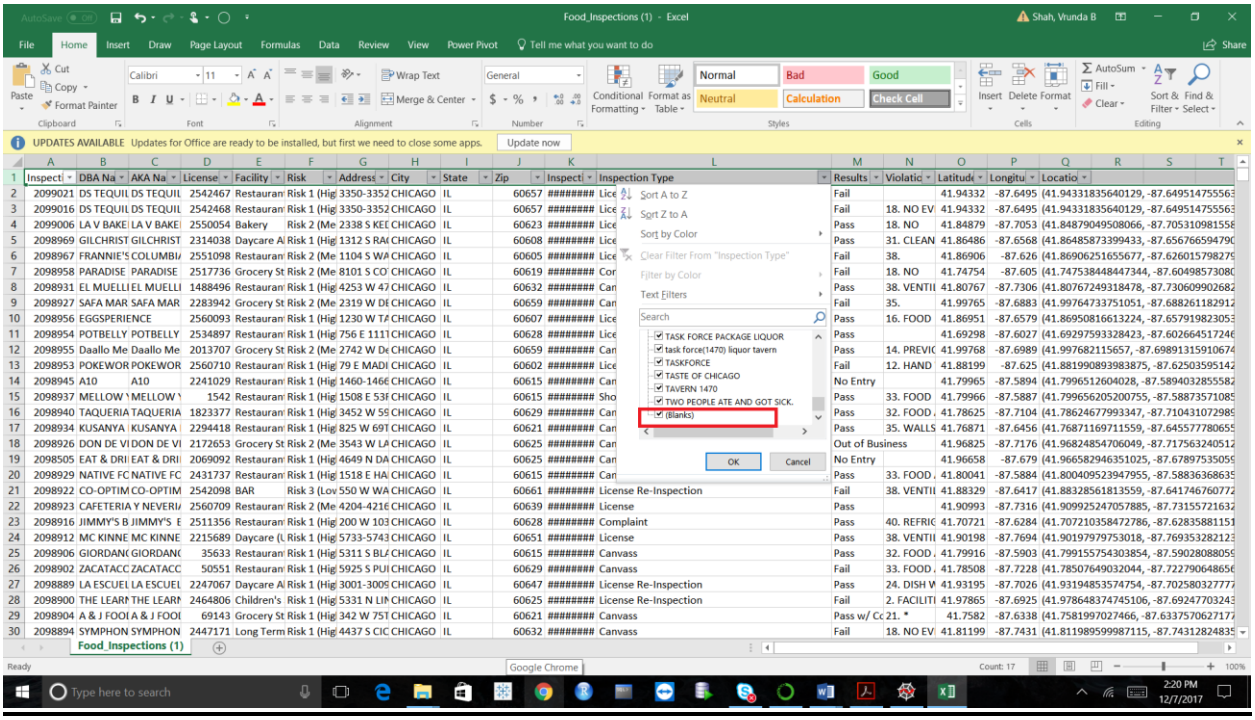
1) Removal of NA and Blank values

Code

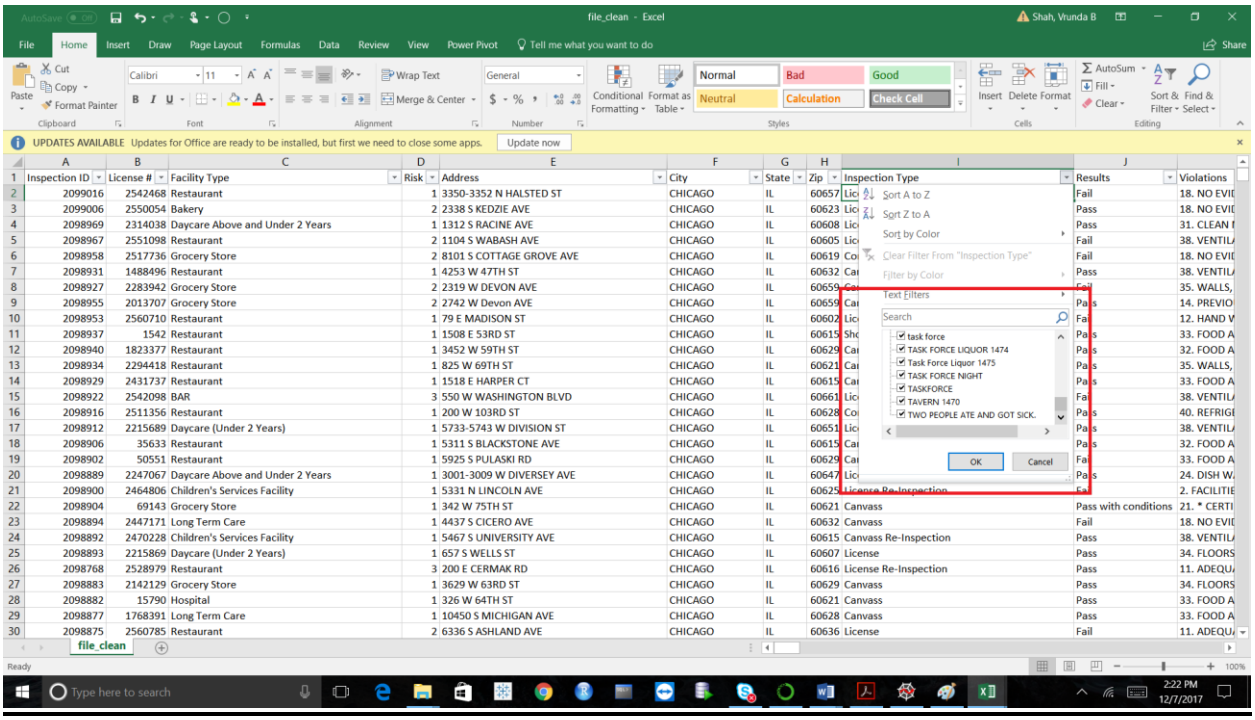
rem_missing = data1.dropna()

rem_missing[rem_missing.notnull()]

Screenshot



After Cleaning

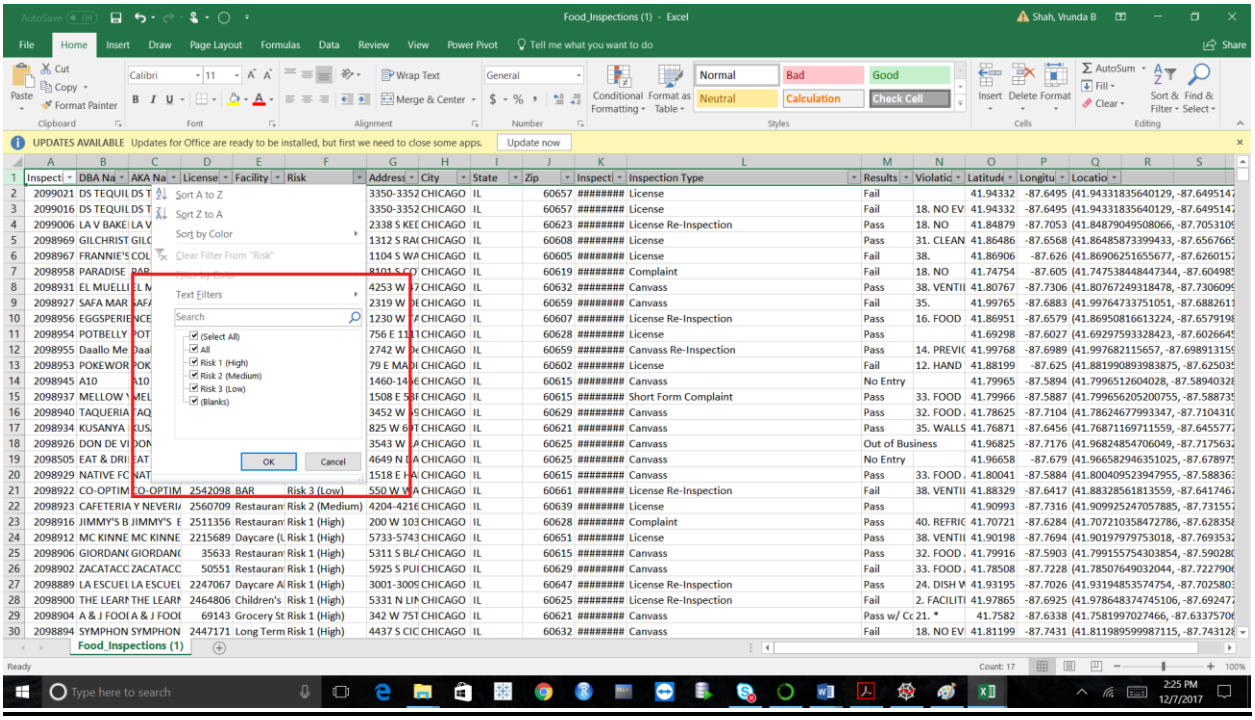


2) Replacing Values

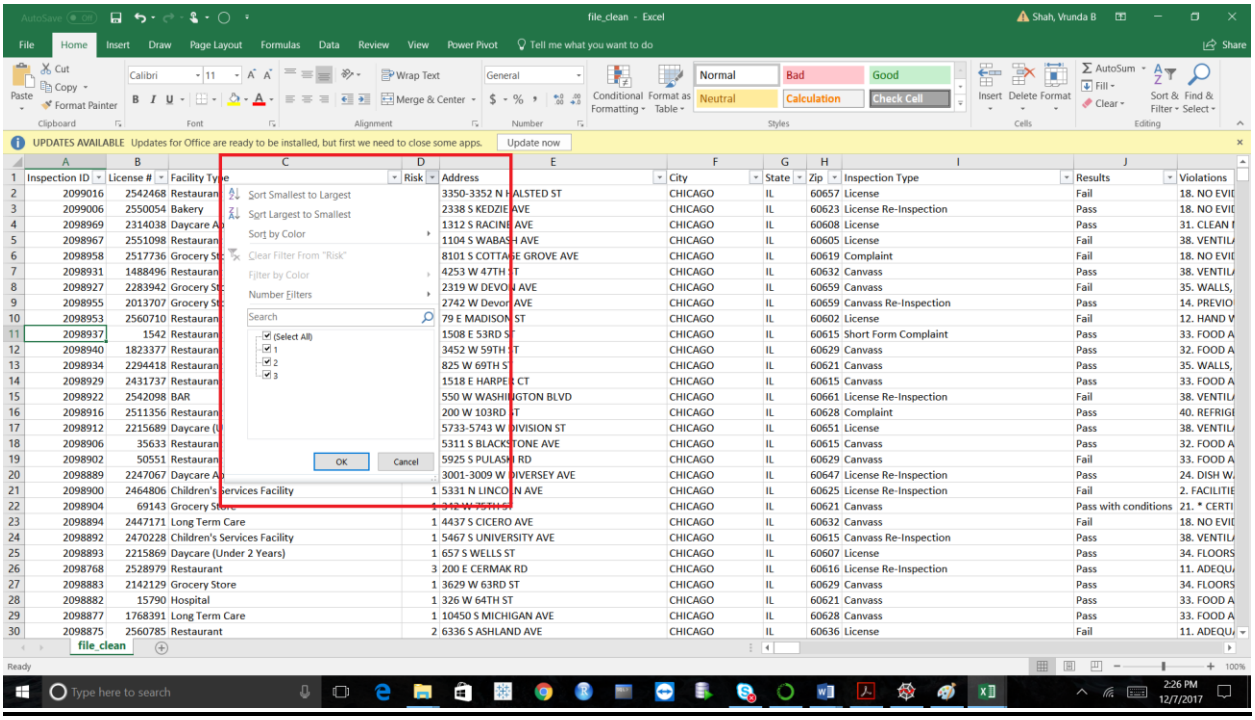
Code

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

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

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

Screenshot



After Cleaning

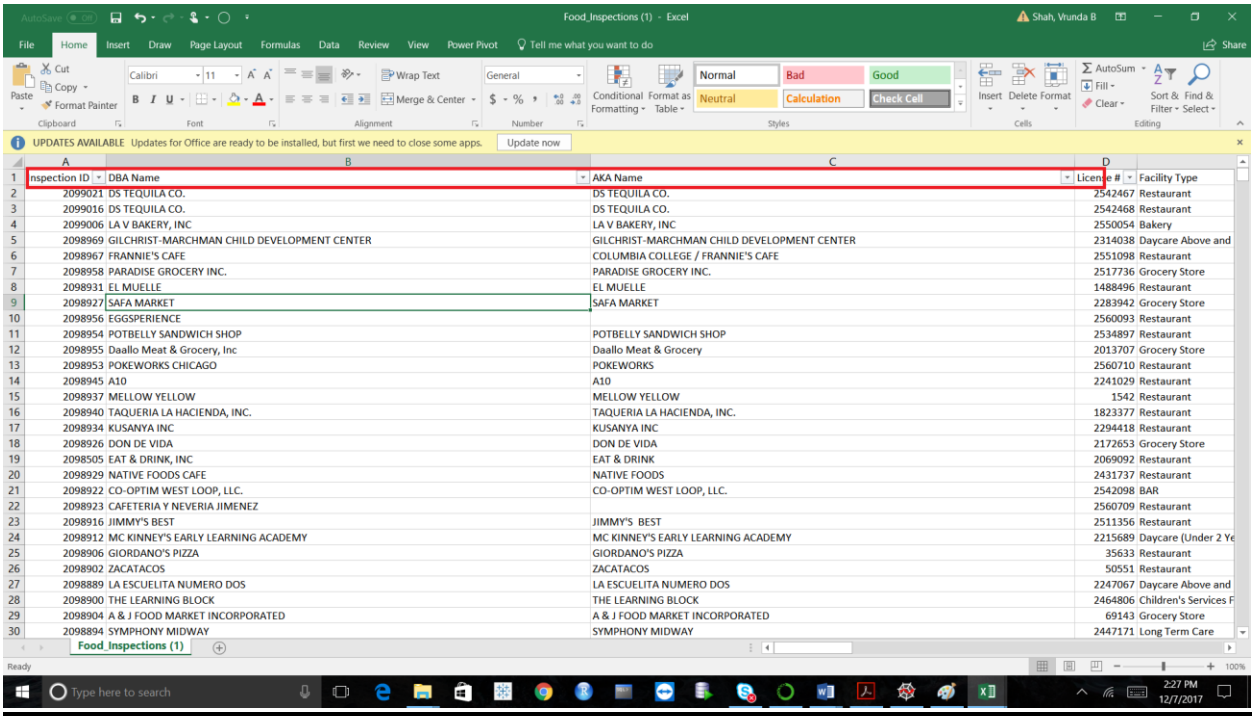


3) Removal of Unwanted Columns

Code

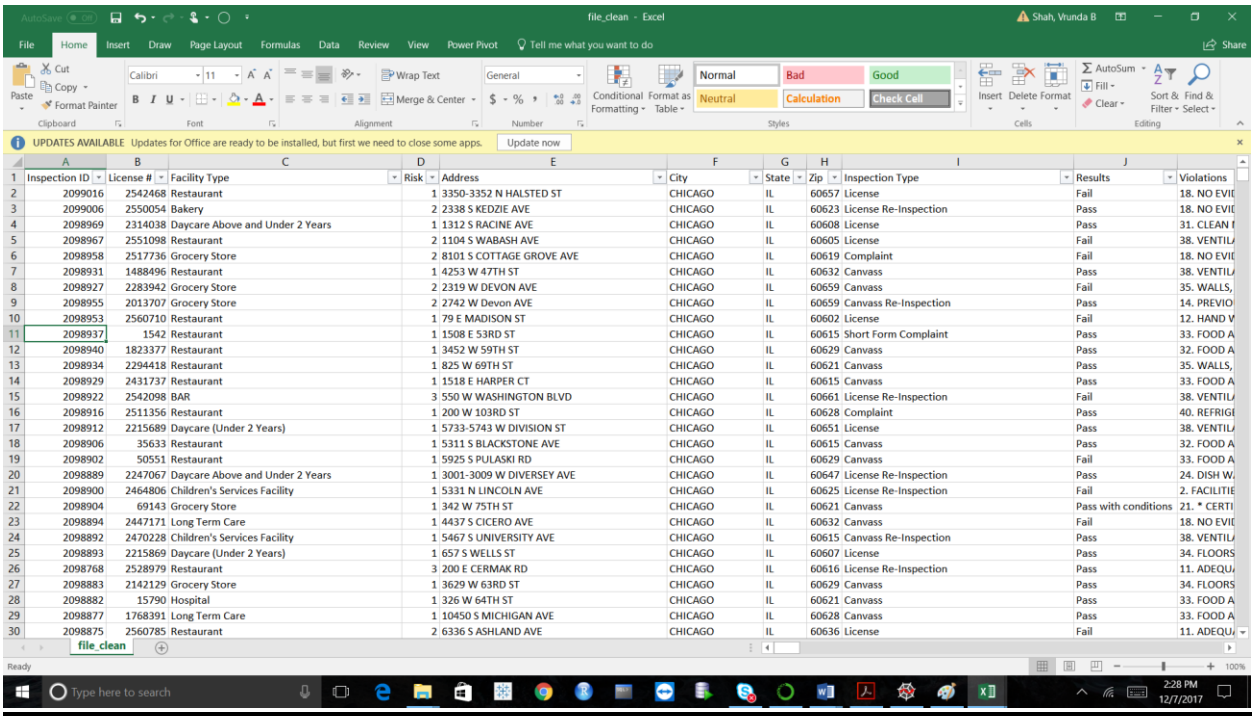
```
rem_missing = rem_missing.drop(['DBA Name', 'AKA Name', 'Inspection Date', 'Month', 'Date'],
axis = 1)
```

Screenshot:



Inspection ID	DBA Name	AKA Name	License #	Facility Type
2099021	DS TEQUILA CO.	DS TEQUILA CO.	2542467	Restaurant
2099016	DS TEQUILA CO.	DS TEQUILA CO.	2542468	Restaurant
2099006	LA V BAKERY, INC	LA V BAKERY, INC	2550054	Bakery
2098969	GILCHRIST-MARCHMAN CHILD DEVELOPMENT CENTER	GILCHRIST-MARCHMAN CHILD DEVELOPMENT CENTER	2314038	Daycare Above and
2098967	FRANNIE'S CAFE	COLUMBIA COLLEGE / FRANNIE'S CAFE	2551098	Restaurant
2098958	PARADISE GROCERY INC.	PARADISE GROCERY INC.	2517736	Grocery Store
2098931	EL MUELLE	EL MUELLE	1488496	Restaurant
2098927	SAFA MARKET	SAFA MARKET	2283942	Grocery Store
2098956	EGGSPERIENCE		2560093	Restaurant
2098954	POTBELLY SANDWICH SHOP	POTBELLY SANDWICH SHOP	2534897	Restaurant
2098955	Daallo Meat & Grocery, Inc	Daallo Meat & Grocery	2013707	Grocery Store
2098953	POKEWORKS CHICAGO	POKEWORKS	2560710	Restaurant
2098945	A10	A10	2241029	Restaurant
2098937	MELLOW YELLOW	MELLOW YELLOW	1542	Restaurant
2098940	TAQUERIA LA HACIENDA, INC.	TAQUERIA LA HACIENDA, INC.	1823377	Restaurant
2098934	KUSANYA INC	KUSANYA INC	2294418	Restaurant
2098926	DON DE VIDA	DON DE VIDA	2172653	Grocery Store
2098505	EAT & DRINK, INC	EAT & DRINK	2069092	Restaurant
2098929	NATIVE FOODS CAFE	NATIVE FOODS	2431737	Restaurant
2098922	CO-OPTIM WEST LOOP, LLC.	CO-OPTIM WEST LOOP, LLC.	2542098	BAR
2098923	CAFETERIA Y NEVERIA JIMENEZ		2560709	Restaurant
2098916	JIMMY'S BEST	JIMMY'S BEST	2511356	Restaurant
2098912	MC KINNEY'S EARLY LEARNING ACADEMY	MC KINNEY'S EARLY LEARNING ACADEMY	2215689	Daycare (Under 2 Ye
2098906	GIORDANO'S PIZZA	GIORDANO'S PIZZA	35633	Restaurant
2098902	ZACATACOS	ZACATACOS	50551	Restaurant
2098889	LA ESCUELITA NUMERO DOS	LA ESCUELITA NUMERO DOS	2247067	Daycare Above and
2098900	THE LEARNING BLOCK	THE LEARNING BLOCK	2464806	Children's Services F
2098904	A & J FOOD MARKET INCORPORATED	A & J FOOD MARKET INCORPORATED	69143	Grocery Store
2098894	SYMPHONY MIDWAY	SYMPHONY MIDWAY	2447171	Long Term Care

After Cleaning



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	License	Fail	18. NO EVI
2099006	2550054	Bakery	2	2338 S KEDZIE AVE	CHICAGO	IL	60623	License Re-Inspection	Pass	18. NO EVI
2098969	2314038	Daycare Above and Under 2 Years	1	1312 S RACINE AVE	CHICAGO	IL	60608	License	Pass	31. CLEAN
2098967	2551098	Restaurant	2	1104 S WABASH AVE	CHICAGO	IL	60605	License	Fail	38. VENTIL
2098958	2517736	Grocery Store	2	8101 S COTTAGE GROVE AVE	CHICAGO	IL	60619	Complaint	Fail	18. NO EVI
2098931	1488496	Restaurant	1	4253 W 47TH ST	CHICAGO	IL	60632	Canvass	Pass	38. VENTIL
2098927	2283942	Grocery Store	2	2319 W DEVON AVE	CHICAGO	IL	60659	Canvass	Fail	35. WALLS,
2098955	2013707	Grocery Store	2	2742 W Devon AVE	CHICAGO	IL	60659	Canvass Re-Inspection	Pass	14. PREVIO
2098953	2560710	Restaurant	1	79 E MADISON ST	CHICAGO	IL	60602	License	Fail	12. HAND W
2098937	1542	Restaurant	1	1508 E 53RD ST	CHICAGO	IL	60615	Short Form Complaint	Pass	33. FOOD A
2098940	1823377	Restaurant	1	3452 W 59TH ST	CHICAGO	IL	60629	Canvass	Pass	32. FOOD A
2098934	2294418	Restaurant	1	825 W 69TH ST	CHICAGO	IL	60621	Canvass	Pass	35. WALLS,
2098929	2431737	Restaurant	1	1518 E HARPER CT	CHICAGO	IL	60615	Canvass	Pass	33. FOOD A
2098922	2542098	BAR	3	550 W WASHINGTON BLVD	CHICAGO	IL	60661	License Re-Inspection	Fail	38. VENTIL
2098916	2511356	Restaurant	1	200 W 103RD ST	CHICAGO	IL	60628	Complaint	Pass	40. REFRIG
2098912	2215689	Daycare (Under 2 Years)	1	5733-5743 W DIVISION ST	CHICAGO	IL	60651	License	Pass	38. VENTIL
2098906	35633	Restaurant	1	5311 S BLACKSTONE AVE	CHICAGO	IL	60615	Canvass	Pass	32. FOOD A
2098902	50551	Restaurant	1	5925 S PULASKI RD	CHICAGO	IL	60629	Canvass	Fail	33. FOOD A
2098889	2247067	Daycare Above and Under 2 Years	1	3001-3009 W DIVERSEY AVE	CHICAGO	IL	60647	License Re-Inspection	Pass	24. DISH W
2098900	2464806	Children's Services Facility	1	5331 N LINCOLN AVE	CHICAGO	IL	60625	License Re-Inspection	Fail	2. FACILITIE
2098904	69143	Grocery Store	1	342 W 75TH ST	CHICAGO	IL	60621	Canvass	Pass with conditions	21. * CERTI
2098894	2447171	Long Term Care	1	4437 S CICERO AVE	CHICAGO	IL	60632	Canvass	Fail	18. NO EVI
2098892	2470228	Children's Services Facility	1	5467 S UNIVERSITY AVE	CHICAGO	IL	60615	Canvass Re-Inspection	Pass	38. VENTIL
2098893	2215869	Daycare (Under 2 Years)	1	657 S WELLS ST	CHICAGO	IL	60607	License	Pass	34. FLOORS
2098768	2528979	Restaurant	3	200 E CERMAK RD	CHICAGO	IL	60616	License Re-Inspection	Pass	11. ADEQUA
2098883	2142129	Grocery Store	1	3629 W 63RD ST	CHICAGO	IL	60629	Canvass	Pass	34. FLOORS
2098882	15790	Hospital	1	326 W 64TH ST	CHICAGO	IL	60621	Canvass	Pass	33. FOOD A
2098877	1768391	Long Term Care	1	10450 S MICHIGAN AVE	CHICAGO	IL	60628	Canvass	Pass	33. FOOD A
2098875	2560785	Restaurant	2	6336 S ASHLAND AVE	CHICAGO	IL	60636	License	Fail	11. ADEQUA

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.

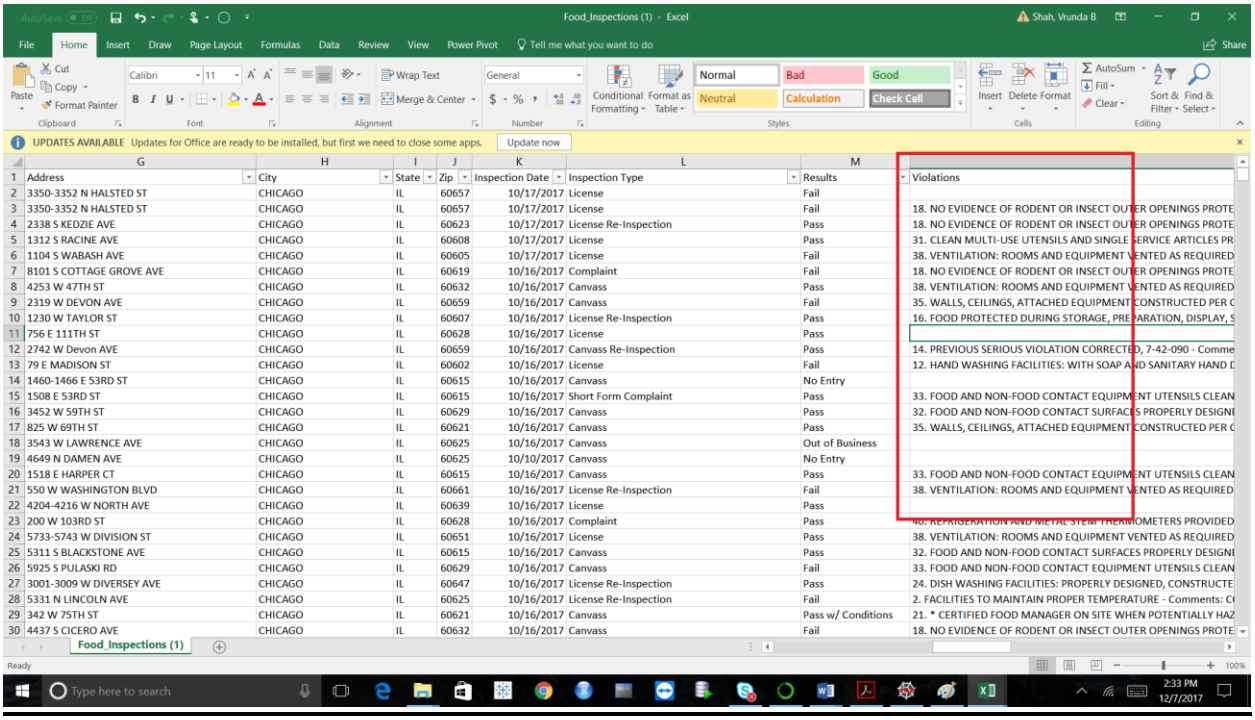
We will be using Violation Description to identify top 10 violations that took place from 2010- 2017. We have used ***string function SPLIT()***

TECHNIQUE USED – STRING FUNCTIONS

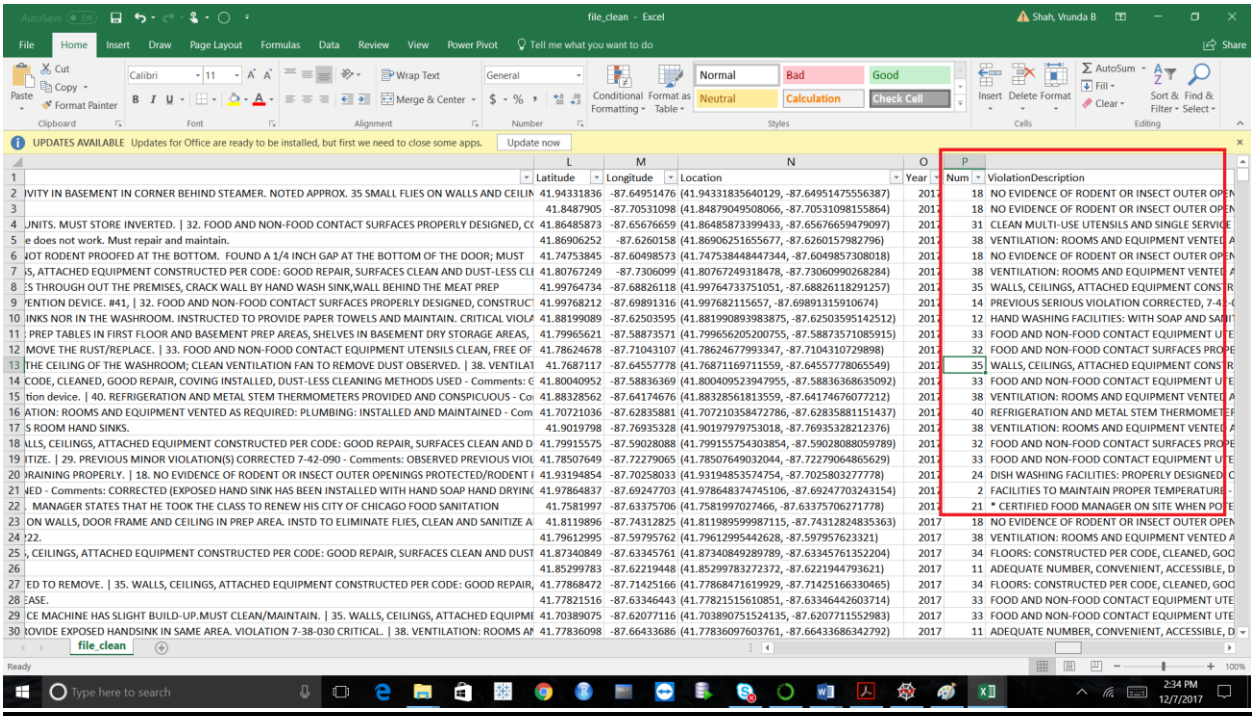
Code

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

Screenshot



	Address	City	State	Zip	Inspection Date	Inspection Type	Results	Violations
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	18. NO EVIDENCE OF RODENT OR INSECT OUTER OPENINGS PROTE
4	2338 S KEDZIE AVE	CHICAGO	IL	60623	10/17/2017	License Re-Inspection	Pass	18. NO EVIDENCE OF RODENT OR INSECT OUTER OPENINGS PROTE
5	1312 S RACINE AVE	CHICAGO	IL	60608	10/17/2017	License	Pass	31. CLEAN MULTI-USE UTENSILS AND SINGLE SERVICE ARTICLES PR
6	1104 S WABASH AVE	CHICAGO	IL	60605	10/17/2017	License	Fail	38. VENTILATION: ROOMS AND EQUIPMENT VENTED AS REQUIRED
7	8101 S COTTAGE GROVE AVE	CHICAGO	IL	60619	10/16/2017	Complaint	Fail	18. NO EVIDENCE OF RODENT OR INSECT OUTER OPENINGS PROTE
8	4253 W 47TH ST	CHICAGO	IL	60632	10/16/2017	Canvass	Pass	38. VENTILATION: ROOMS AND EQUIPMENT VENTED AS REQUIRED
9	2319 W DEVON AVE	CHICAGO	IL	60659	10/16/2017	Canvass	Fail	35. WALLS, CEILINGS, ATTACHED EQUIPMENT CONSTRUCTED PER C
10	1230 W TAYLOR ST	CHICAGO	IL	60607	10/16/2017	License Re-Inspection	Pass	16. FOOD PROTECTED DURING STORAGE, PREPARATION, DISPLAY, S
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	14. PREVIOUS SERIOUS VIOLATION CORRECTED, 7-42-090 - Comme
13	79 E MADISON ST	CHICAGO	IL	60602	10/16/2017	License	Fail	12. HAND WASHING FACILITIES: WITH SOAP AND SANITARY HAND D
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	33. FOOD AND NON-FOOD CONTACT EQUIPMENT UTENSILS CLEAN
16	3452 W 59TH ST	CHICAGO	IL	60629	10/16/2017	Canvass	Pass	32. FOOD AND NON-FOOD CONTACT SURFACES PROPERLY DESIGNE
17	825 W 69TH ST	CHICAGO	IL	60621	10/16/2017	Canvass	Pass	35. WALLS, CEILINGS, ATTACHED EQUIPMENT CONSTRUCTED PER C
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	33. FOOD AND NON-FOOD CONTACT EQUIPMENT UTENSILS CLEAN
21	550 W WASHINGTON BLVD	CHICAGO	IL	60661	10/16/2017	License Re-Inspection	Fail	38. VENTILATION: ROOMS AND EQUIPMENT VENTED AS REQUIRED
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	40. REFRIGERATION AND METAL STEM THERMOMETERS PROVIDED
24	5733-5743 W DIVISION ST	CHICAGO	IL	60651	10/16/2017	License	Pass	38. VENTILATION: ROOMS AND EQUIPMENT VENTED AS REQUIRED
25	5311 S BLACKSTONE AVE	CHICAGO	IL	60615	10/16/2017	Canvass	Pass	32. FOOD AND NON-FOOD CONTACT SURFACES PROPERLY DESIGNE
26	5925 S PULASKI RD	CHICAGO	IL	60629	10/16/2017	Canvass	Fail	33. FOOD AND NON-FOOD CONTACT EQUIPMENT UTENSILS CLEAN
27	3001-3009 W DIVERSEY AVE	CHICAGO	IL	60647	10/16/2017	License Re-Inspection	Pass	24. DISH WASHING FACILITIES: PROPERLY DESIGNED, CONSTRUCTE
28	5331 N LINCOLN AVE	CHICAGO	IL	60625	10/16/2017	License Re-Inspection	Fail	2. FACILITIES TO MAINTAIN PROPER TEMPERATURE - Comments: C
29	342 W 75TH ST	CHICAGO	IL	60621	10/16/2017	Canvass	Pass w/ Conditions	21. * CERTIFIED FOOD MANAGER ON SITE WHEN POTENTIALLY HAZ
30	4437 S CICERO AVE	CHICAGO	IL	60632	10/16/2017	Canvass	Fail	18. NO EVIDENCE OF RODENT OR INSECT OUTER OPENINGS PROTE



	L	M	N	O	P	
	Latitude	Longitude	Location	Year	Num	ViolationDescription
2	41.94331836	-87.64951476	(41.94331835640129, -87.64951475556387)	2017	18	NO EVIDENCE OF RODENT OR INSECT OUTER OPEN
3	41.8487905	-87.70531098	(41.84879049508066, -87.70531098155864)	2017	18	NO EVIDENCE OF RODENT OR INSECT OUTER OPEN
4	41.86485873	-87.65676659	(41.86485873399433, -87.65676659479097)	2017	31	CLEAN MULTI-USE UTENSILS AND SINGLE SERVICE
5	41.86906252	-87.6260158	(41.86906251655677, -87.6260157982796)	2017	38	VENTILATION: ROOMS AND EQUIPMENT VENTED
6	41.74753845	-87.60498573	(41.747538448447344, -87.6049857308018)	2017	18	NO EVIDENCE OF RODENT OR INSECT OUTER OPEN
7	41.80767249	-87.7306099	(41.80767249318478, -87.73060990268284)	2017	38	VENTILATION: ROOMS AND EQUIPMENT VENTED
8	41.99764734	-87.68826118	(41.99764733751051, -87.68826118291257)	2017	35	WALLS, CEILINGS, ATTACHED EQUIPMENT CONST
9	41.99768212	-87.69891316	(41.997682115657, -87.69891315910674)	2017	14	PREVIOUS SERIOUS VIOLATION CORRECTED, 7-4
10	41.88199089	-87.62503595	(41.88199089393875, -87.62503595142512)	2017	12	HAND WASHING FACILITIES: WITH SOAP AND SA
11	41.79965621	-87.58873571	(41.79965620520755, -87.58873571085915)	2017	33	FOOD AND NON-FOOD CONTACT EQUIPMENT UT
12	41.78624678	-87.71043107	(41.78624677993347, -87.7104310729898)	2017	32	FOOD AND NON-FOOD CONTACT SURFACES PROP
13	41.7687117	-87.64557778	(41.76871169711559, -87.64557778065549)	2017	35	WALLS, CEILINGS, ATTACHED EQUIPMENT CONS
14	41.80040952	-87.58836369	(41.80040952394955, -87.58836368635092)	2017	33	FOOD AND NON-FOOD CONTACT EQUIPMENT UT
15	41.88328562	-87.64174676	(41.88328561813559, -87.64174676077212)	2017	38	VENTILATION: ROOMS AND EQUIPMENT VENTED
16	41.70721036	-87.62835881	(41.70721035872786, -87.62835881151437)	2017	40	REFRIGERATION AND METAL STEM THERMOMETE
17	41.9019798	-87.76935328	(41.90197979753018, -87.76935328212376)	2017	38	VENTILATION: ROOMS AND EQUIPMENT VENTED
18	41.79915575	-87.59028088	(41.79915574303854, -87.59028088059789)	2017	32	FOOD AND NON-FOOD CONTACT SURFACES PROP
19	41.78507649	-87.72279065	(41.78507649032044, -87.72279064865629)	2017	33	FOOD AND NON-FOOD CONTACT EQUIPMENT UT
20	41.93194854	-87.70258033	(41.93194853574754, -87.70258033277778)	2017	24	DISH WASHING FACILITIES: PROPERLY DESIGNED
21	41.97864837	-87.69247703	(41.97864837474510, -87.69247703243154)	2017	2	FACILITIES TO MAINTAIN PROPER TEMPERATURE
22	41.7581997	-87.63375706	(41.7581997027466, -87.63375706271778)	2017	21	* CERTIFIED FOOD MANAGER ON SITE WHEN PO
23	41.8119896	-87.74312825	(41.81198959987115, -87.74312824835363)	2017	18	NO EVIDENCE OF RODENT OR INSECT OUTER OPEN
24	41.79612995	-87.59795762	(41.79612995442628, -87.59795762323321)	2017	38	VENTILATION: ROOMS AND EQUIPMENT VENTED
25	41.87340849	-87.63345761	(41.87340849289789, -87.63345761352204)	2017	34	FLOORS: CONSTRUCTED PER CODE, CLEANED, GOO
26	41.85299783	-87.62219448	(41.85299783272372, -87.62219447936621)	2017	11	ADEQUATE NUMBER, CONVENIENT, ACCESSIBLE, D
27	41.77868472	-87.71425166	(41.77868471619929, -87.71425166330465)	2017	34	FLOORS: CONSTRUCTED PER CODE, CLEANED, GOO
28	41.77821516	-87.63346443	(41.77821515610851, -87.63346442603714)	2017	33	FOOD AND NON-FOOD CONTACT EQUIPMENT UT
29	41.70389075	-87.62077116	(41.70389075124135, -87.6207711552983)	2017	33	FOOD AND NON-FOOD CONTACT EQUIPMENT UT
30	41.77836098	-87.66433686	(41.77836097603761, -87.66433686342792)	2017	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)

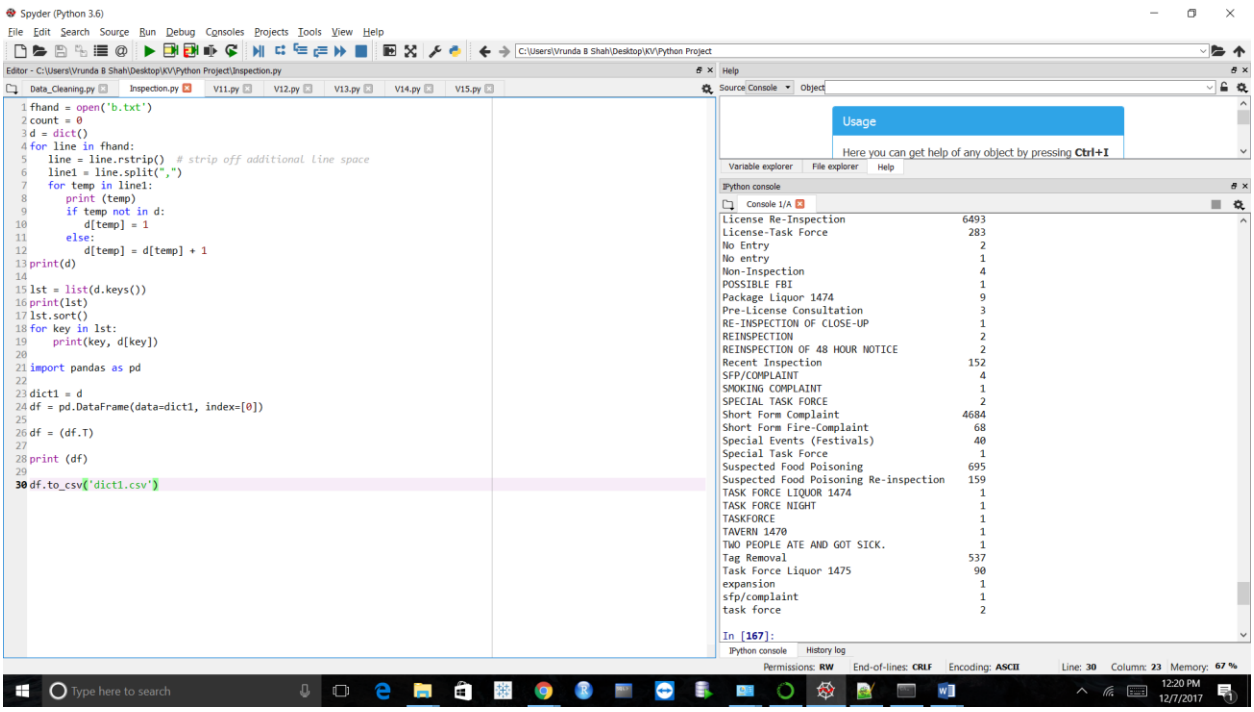
df.to_csv('dict1.csv')

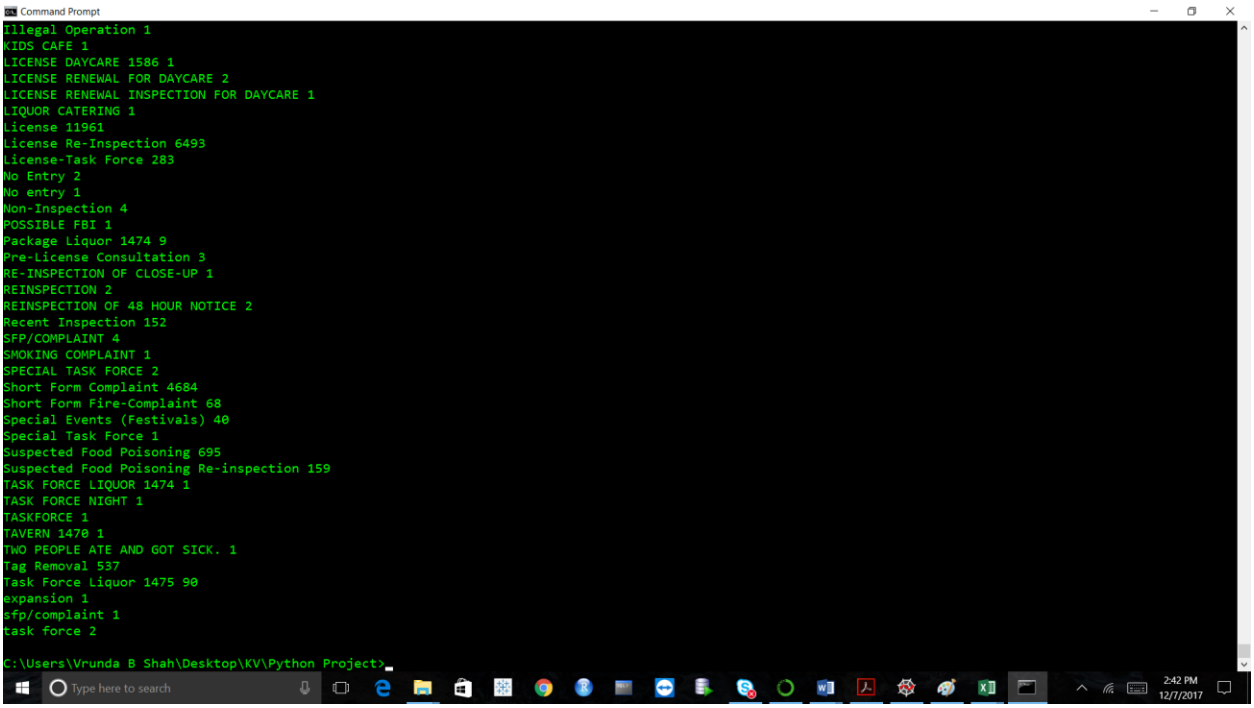
SUMMARY

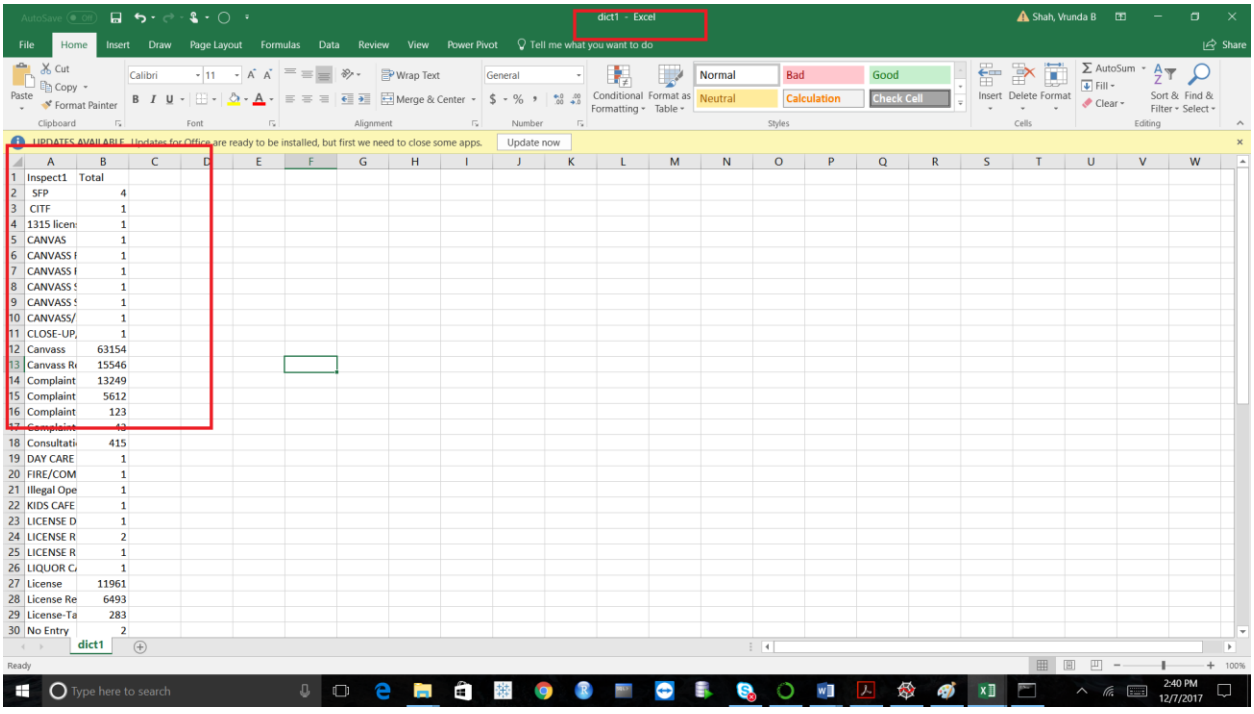
We have used ***list, dictionary*** to get number of times particular 'Inspection Type' was requested from year 2010 – 2017.

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




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 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
C:\Users\Vrunda B Shah\Desktop\KV\Python Project>
Type here to search
2:42 PM
12/7/2017


dict1 - Excel
Shah, Vrunda B
File Home Insert Draw Page Layout Formulas Data Review View Power Pivot Tell me what you want to do
Share
Paste Cut Copy Format Painter
Calibri 11
Wrap Text
Merge & Center
General
Conditional Formatting
Format as Table
Normal
Bad
Good
Neutral
Calculation
Check Cell
Insert Delete Format
AutoSum Fill Clear
Sort & Filter Find & Select
Clipboard Font Alignment Number Styles Cells Editing
UPDATES AVAILABLE Updates for Office are ready to be installed, but first we need to close some apps.
Update now
Inspect1 Total
SFP 4
CITF 1
1315 licen 1
CANVAS 1
CANVASS I 1
CANVASS I 1
CANVASS S 1
CANVASS S 1
CANVASS/ 1
CLOSE-UP/ 1
Canvass 63154
Canvass Re 15546
Complaint 13249
Complaint 5612
Complaint 123
Complaint 43
Consultati 415
DAY CARE 1
FIRE/COM 1
Illegal Ope 1
KIDS CAFE 1
LICENSE D 1
LICENSE R 2
LICENSE R 1
LIQUOR CA 1
License 11961
License Re 6493
License-Ta 283
No Entry 2
dict1
Ready
Type here to search
2:40 PM
12/7/2017

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 dercribe()*** 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

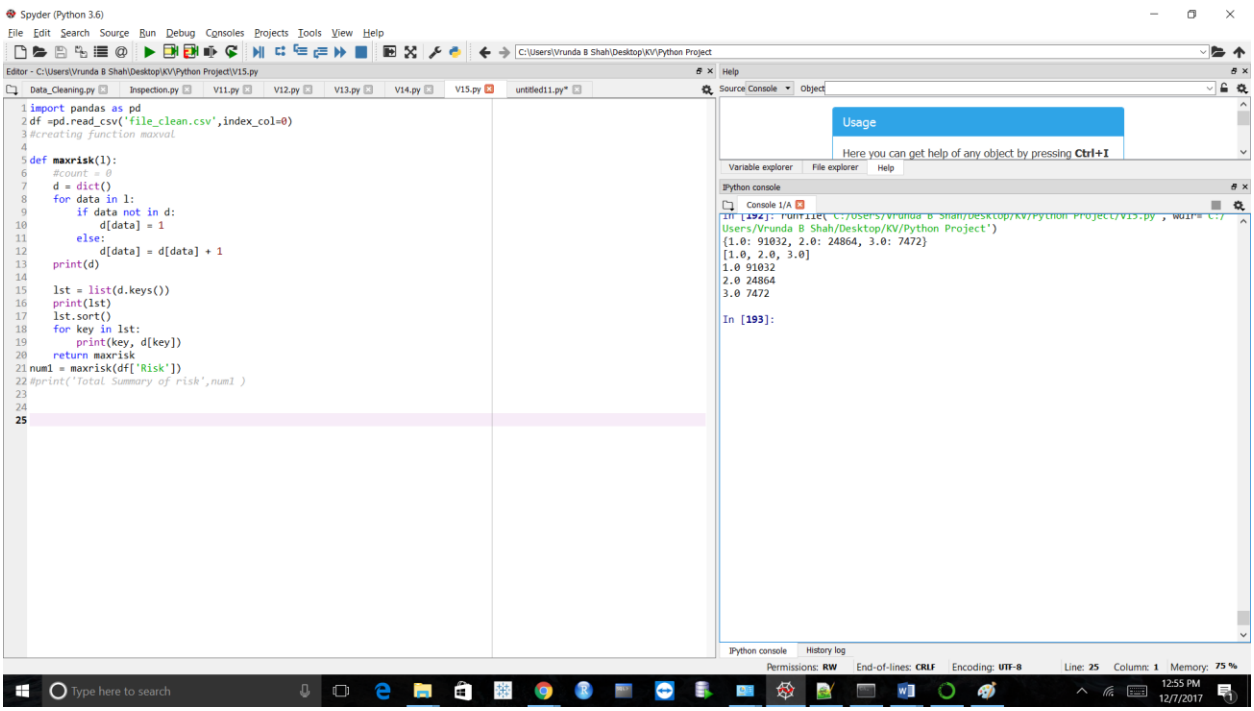
25% - 1.000000

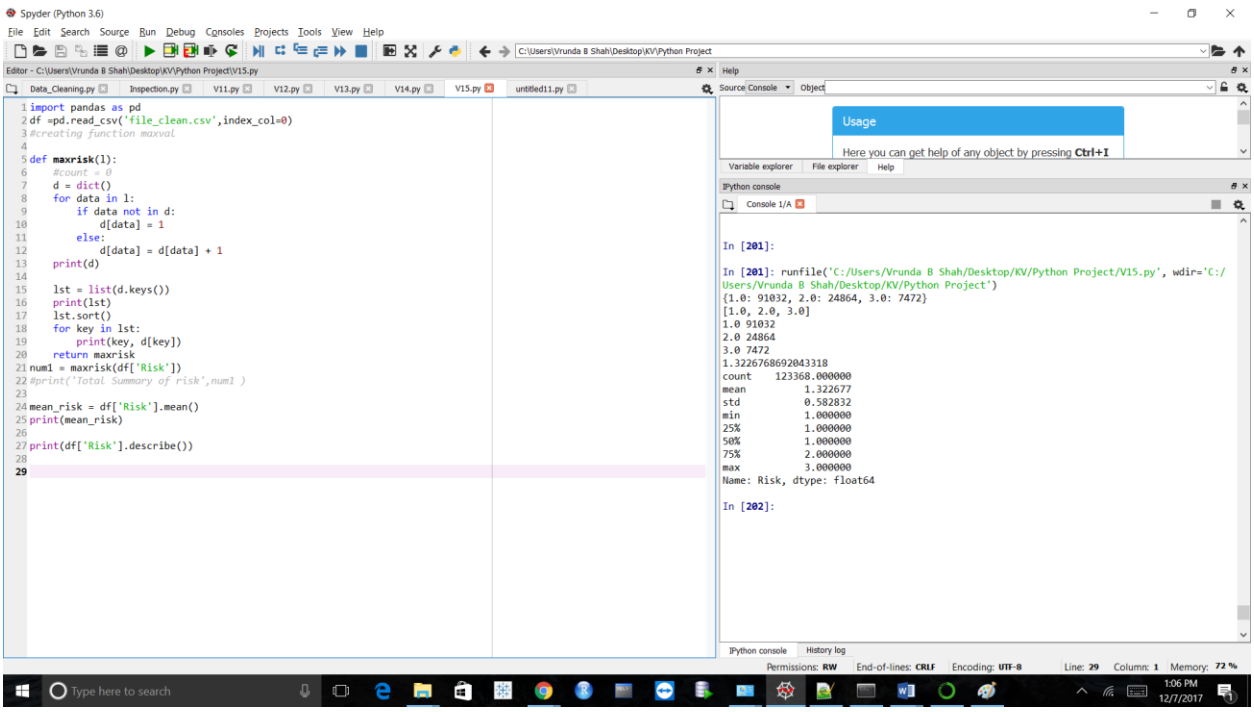
50% - 1.000000

75% - 2.000000

Max - 3.000000

SCREENSHOT





D] ANALYSIS AND VISUALISATIONS

Question1 :

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

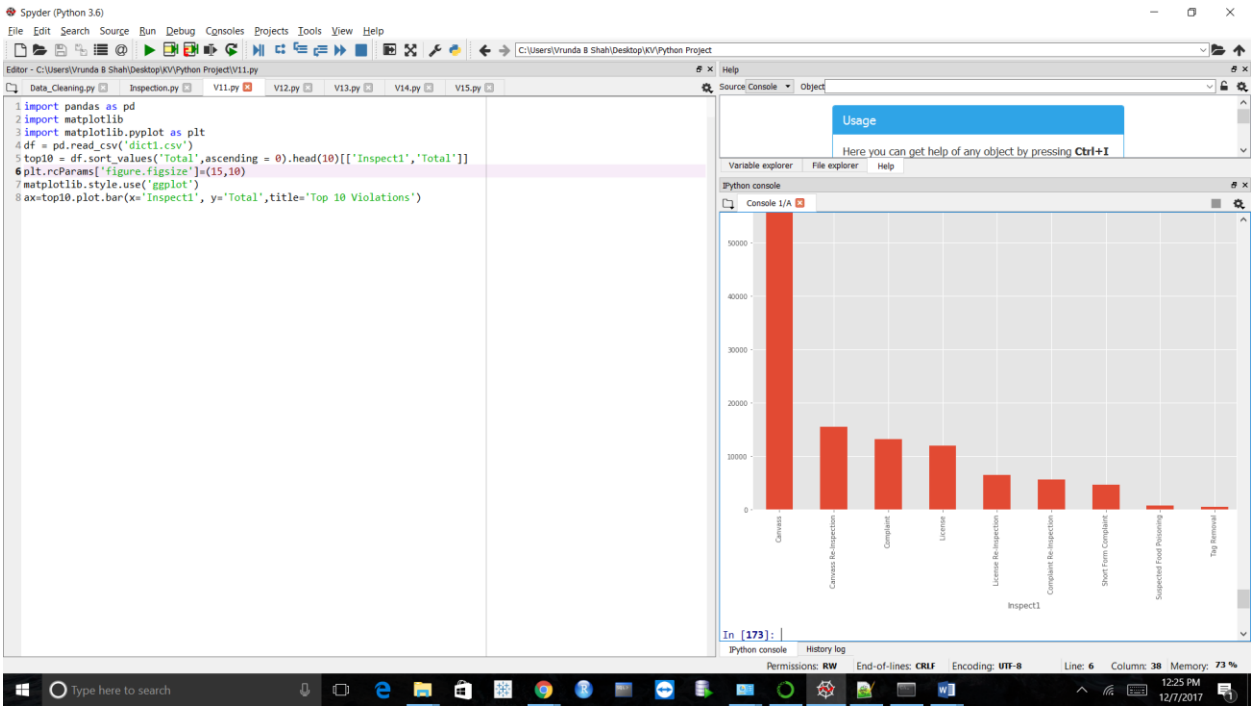
Analysis

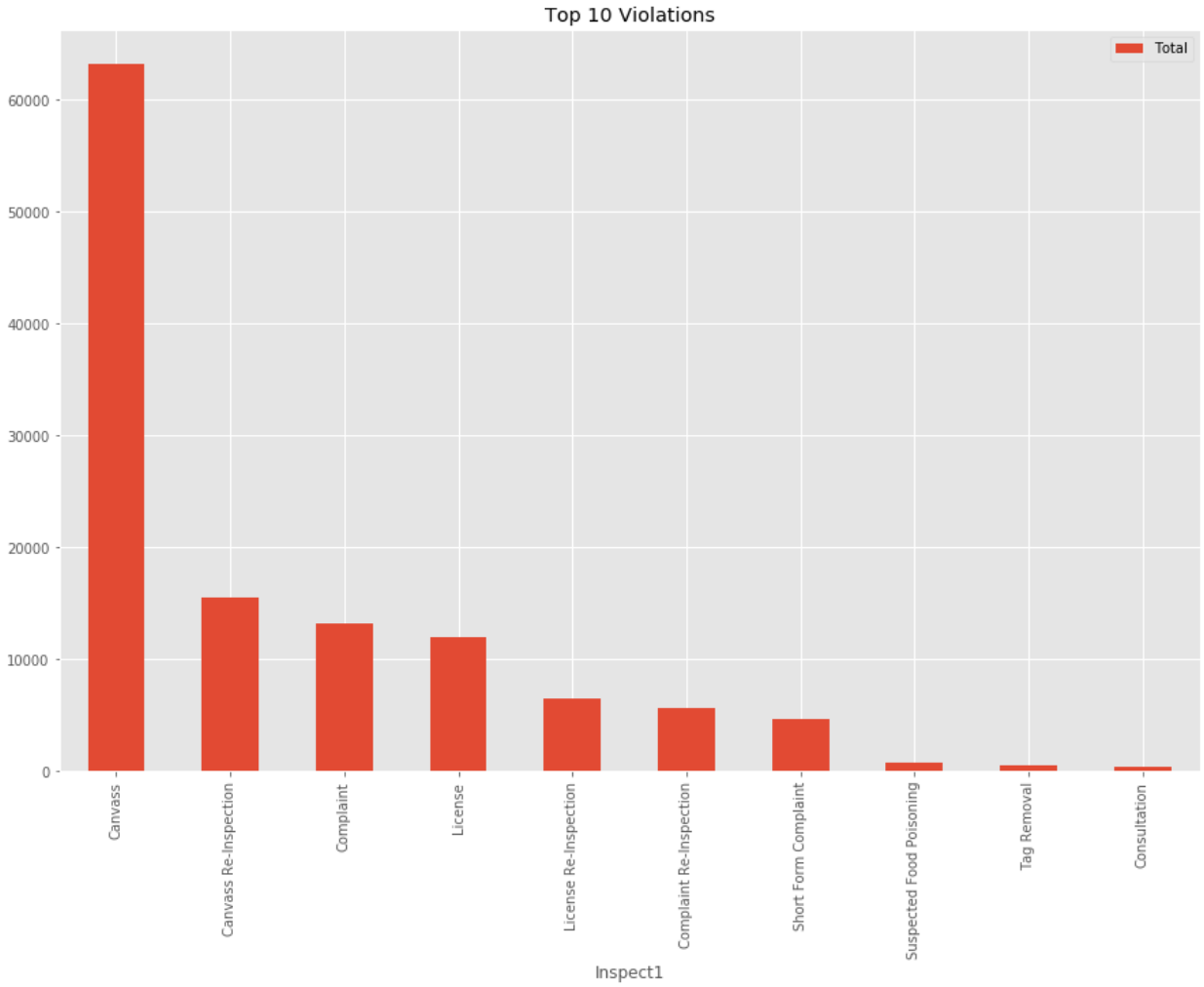
We have found Canvass, Canvass – Reinspection, Complaint, License, License Re -Inspection, Complaint Re-Inspection, Tag removal, Short for complaint, Suspected Food Poisoning were the Top 10 Inspections requested. We have used Bar Chart.

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 ?

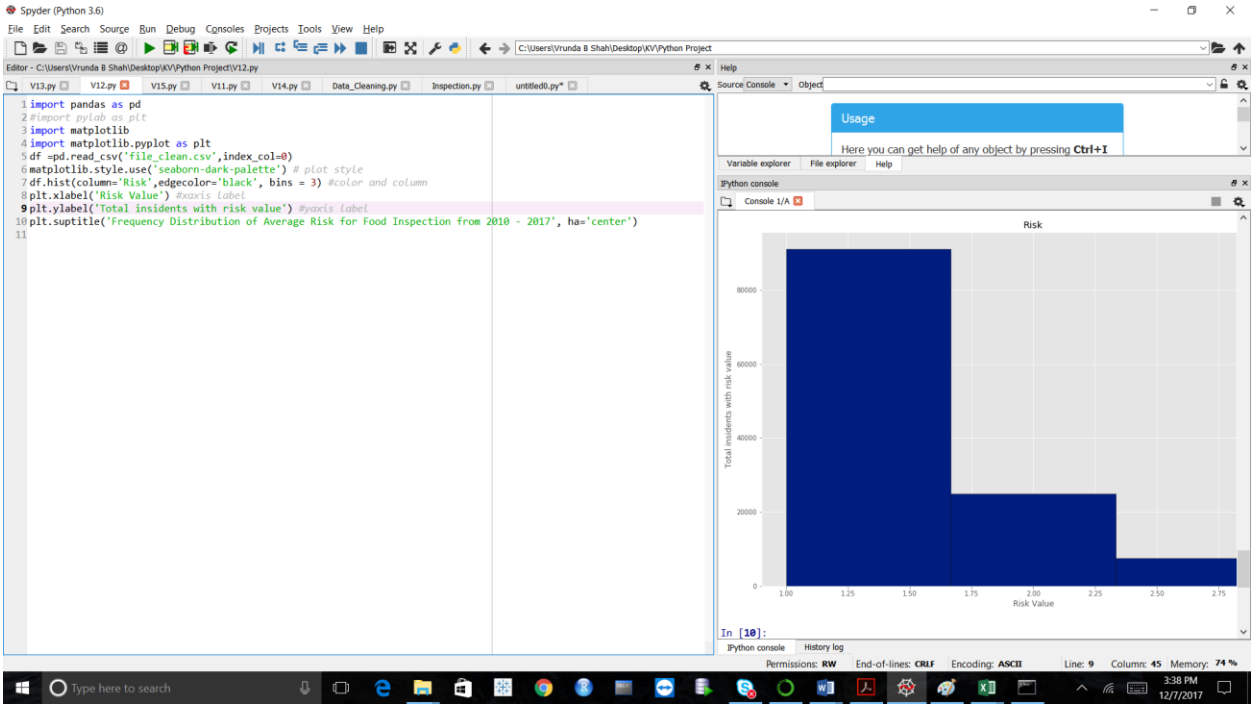
Analysis

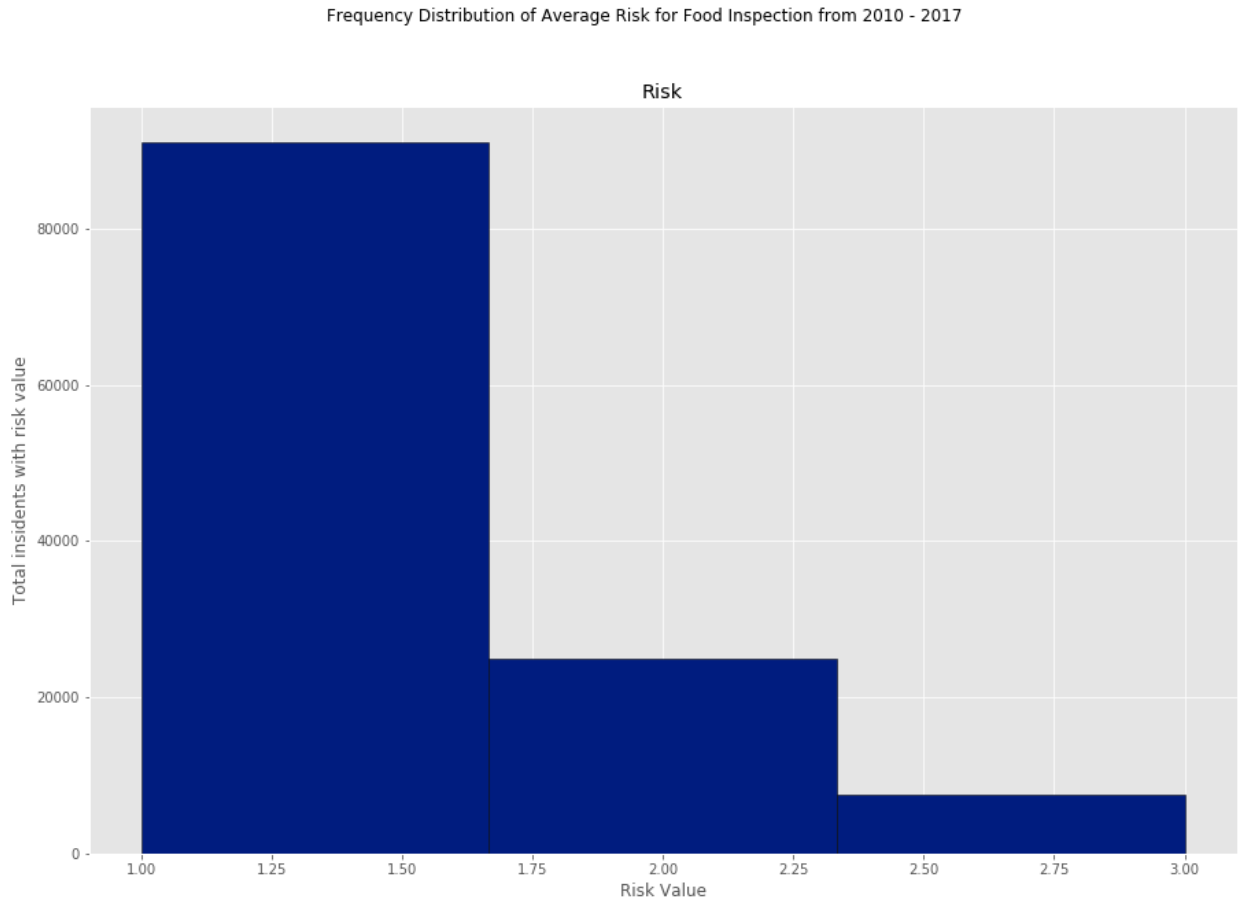
Majority of the inspection requested had Risk associated in the range of Risk 1 and the least associated risk range is falls under Risk 3.

We have used Histogram

TECHNIQUE USED – MATPLOT, PYPLOT

VISUALISATION





Question 3

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

Analysis

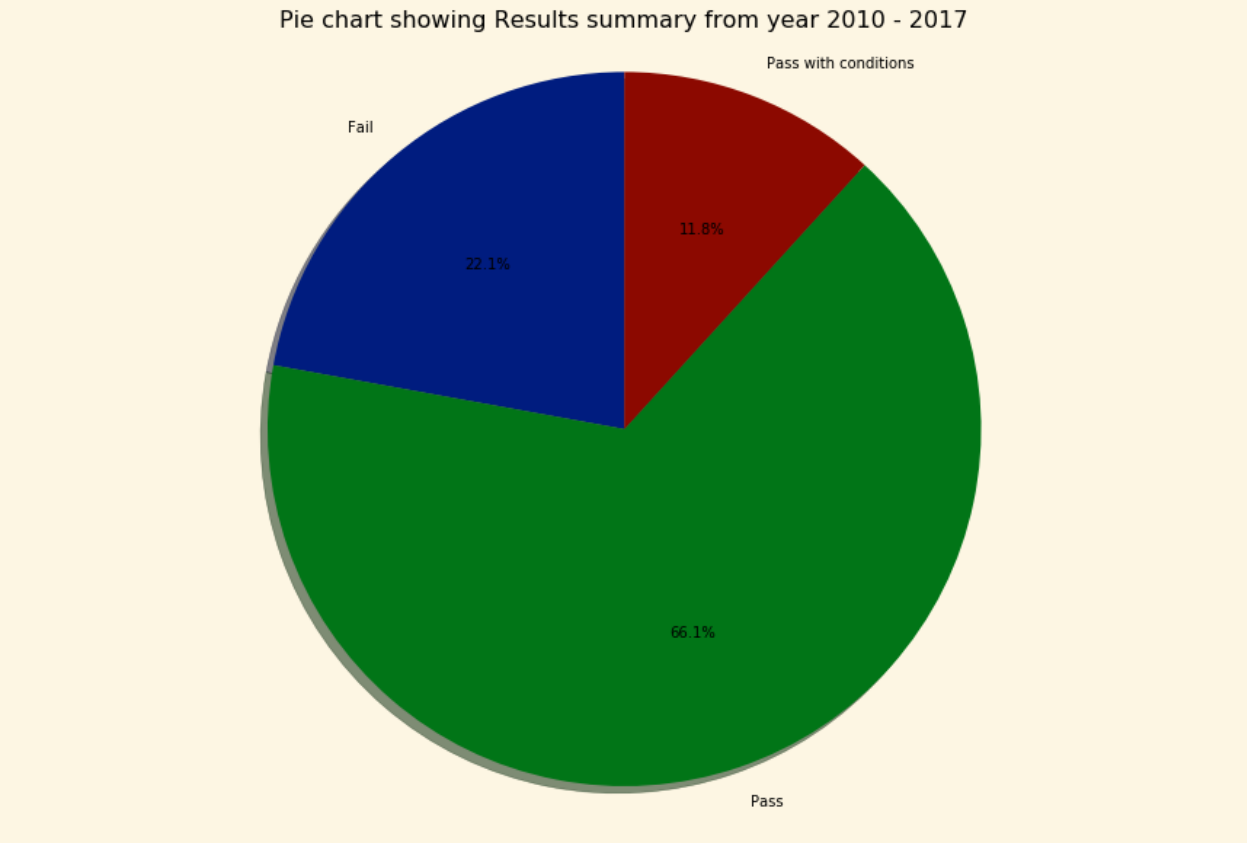
We have identified majority of the Inspection that took place in tenure of 2010 – 2017 has Pass status (66.1 %), 22.1% has outcome Pass with conditions and 11.8% of Inspection results falls under Fail category. We have used Pie chart.

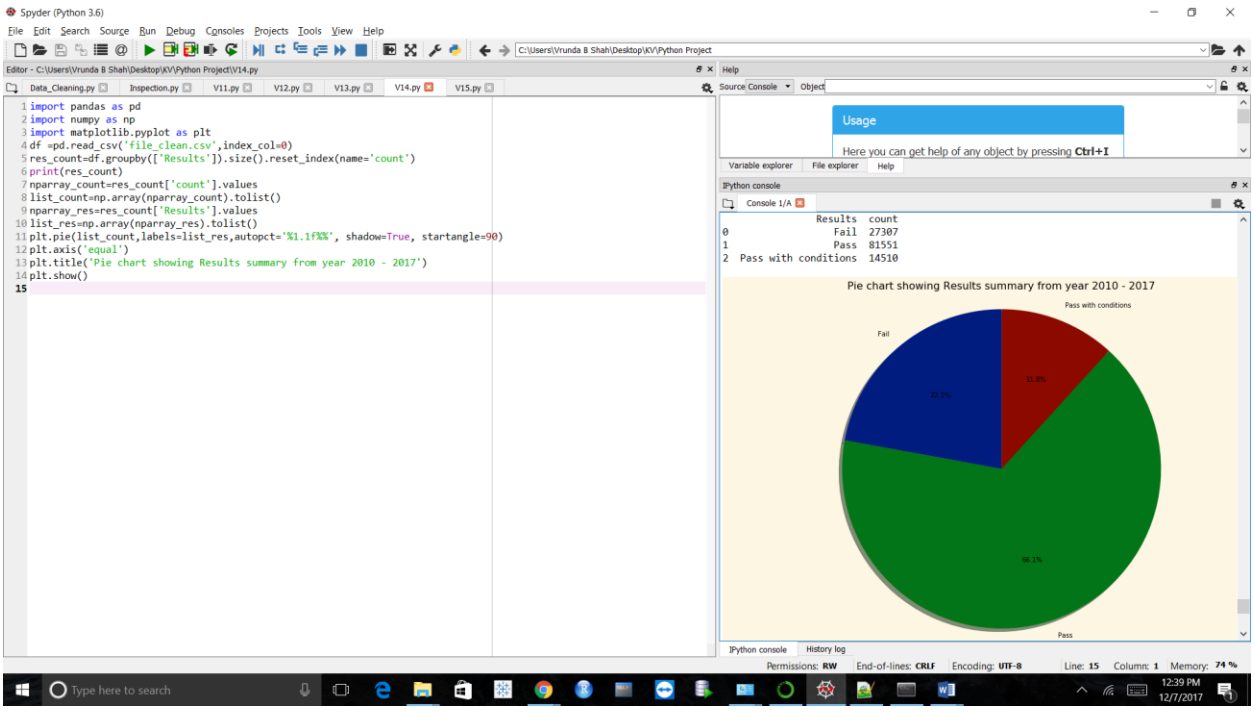
Pass - 66.1

Pass With Conditions - 22.1

Fail - 11.8

TECHNIQUE USED: List, pandas, numpy


Pie chart showing Results summary from year 2010 - 2017
Pass with conditions
Fail
11.8%
22.1%
66.1%
Pass


Spyder (Python 3.6)
File Edit Search Source Run Debug Consoles Projects Tools View Help
Editor - C:\Users\Vrunda B Shah\Desktop\KV\Python Project\V14.py
Data_Cleaning.py Inspection.py V11.py V12.py V13.py V14.py V15.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_res=res_count['Results'].values
10 list_res=np.array(nparray_res).tolist()
11 plt.pie(list_count,labels=list_res,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
Help
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
Pass with conditions
Fail
11.8%
22.1%
66.1%
Pass
IPython console History log
Permissions: RW End-of-lines: CRLF Encoding: UTF-8 Line: 15 Column: 1 Memory: 74 %
Type here to search
12:39 PM
12/7/2017

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 Tytpe 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 insidents 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()
```