311 NYC service Request

October 5, 2020

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[ ]:
```
#TASK1
```

[2]:
```
df = pd.read_csv('311_Service_Requests_from_2010_to_Present.csv')
```

```
/usr/local/lib/python3.7/site-packages/IPython/core/interactiveshell.py:3063:
DtypeWarning: Columns (48,49) have mixed types.Specify dtype option on import or
set low_memory=False.
  interactivity=interactivity, compiler=compiler, result=result)
```

[3]:
```
df.shape
```

[3]:
```
(300698, 53)
```

[4]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300698 entries, 0 to 300697
Data columns (total 53 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   Unique Key                      300698 non-null  int64
 1   Created Date                    300698 non-null  object
 2   Closed Date                     298534 non-null  object
 3   Agency                          300698 non-null  object
 4   Agency Name                     300698 non-null  object
 5   Complaint Type                  300698 non-null  object
 6   Descriptor                      294784 non-null  object
 7   Location Type                   300567 non-null  object
 8   Incident Zip                    298083 non-null  float64
 9   Incident Address                256288 non-null  object
 10  Street Name                     256288 non-null  object
 11  Cross Street 1                  251419 non-null  object
 12  Cross Street 2                  250919 non-null  object
```

```
 13  Intersection Street 1           43858 non-null   object
 14  Intersection Street 2           43362 non-null   object
 15  Address Type                    297883 non-null  object
 16  City                            298084 non-null  object
 17  Landmark                        349 non-null     object
 18  Facility Type                   298527 non-null  object
 19  Status                          300698 non-null  object
 20  Due Date                        300695 non-null  object
 21  Resolution Description          300698 non-null  object
 22  Resolution Action Updated Date  298511 non-null  object
 23  Community Board                 300698 non-null  object
 24  Borough                         300698 non-null  object
 25  X Coordinate (State Plane)      297158 non-null  float64
 26  Y Coordinate (State Plane)      297158 non-null  float64
 27  Park Facility Name              300698 non-null  object
 28  Park Borough                    300698 non-null  object
 29  School Name                     300698 non-null  object
 30  School Number                   300698 non-null  object
 31  School Region                   300697 non-null  object
 32  School Code                     300697 non-null  object
 33  School Phone Number             300698 non-null  object
 34  School Address                  300698 non-null  object
 35  School City                     300698 non-null  object
 36  School State                    300698 non-null  object
 37  School Zip                      300697 non-null  object
 38  School Not Found                300698 non-null  object
 39  School or Citywide Complaint    0 non-null       float64
 40  Vehicle Type                    0 non-null       float64
 41  Taxi Company Borough            0 non-null       float64
 42  Taxi Pick Up Location           0 non-null       float64
 43  Bridge Highway Name             243 non-null     object
 44  Bridge Highway Direction        243 non-null     object
 45  Road Ramp                       213 non-null     object
 46  Bridge Highway Segment          213 non-null     object
 47  Garage Lot Name                 0 non-null       float64
 48  Ferry Direction                 1 non-null       object
 49  Ferry Terminal Name             2 non-null       object
 50  Latitude                        297158 non-null  float64
 51  Longitude                       297158 non-null  float64
 52  Location                        297158 non-null  object
dtypes: float64(10), int64(1), object(42)
memory usage: 121.6+ MB
```

[5]:
```
df.describe()
```

[5]:
```
          Unique Key   Incident Zip  X Coordinate (State Plane)  \
count  3.006980e+05  298083.000000                2.971580e+05
```

```
mean   3.130054e+07   10848.888645               1.004854e+06
std    5.738547e+05     583.182081               2.175338e+04
min    3.027948e+07      83.000000               9.133570e+05
25%    3.080118e+07   10310.000000               9.919752e+05
50%    3.130436e+07   11208.000000               1.003158e+06
75%    3.178446e+07   11238.000000               1.018372e+06
max    3.231065e+07   11697.000000               1.067173e+06

       Y Coordinate (State Plane)  School or Citywide Complaint  Vehicle Type  \
count                297158.000000                           0.0           0.0
mean                 203754.534416                           NaN           NaN
std                   29880.183529                           NaN           NaN
min                  121219.000000                           NaN           NaN
25%                  183343.000000                           NaN           NaN
50%                  201110.500000                           NaN           NaN
75%                  224125.250000                           NaN           NaN
max                  271876.000000                           NaN           NaN

       Taxi Company Borough  Taxi Pick Up Location  Garage Lot Name  \
count                   0.0                    0.0              0.0
mean                    NaN                    NaN              NaN
std                     NaN                    NaN              NaN
min                     NaN                    NaN              NaN
25%                     NaN                    NaN              NaN
50%                     NaN                    NaN              NaN
75%                     NaN                    NaN              NaN
max                     NaN                    NaN              NaN

            Latitude      Longitude
count  297158.000000  297158.000000
mean       40.725885     -73.925630
std         0.082012       0.078454
min        40.499135     -74.254937
25%        40.669796     -73.972142
50%        40.718661     -73.931781
75%        40.781840     -73.876805
max        40.912869     -73.700760
```

[6]:
```
df.isnull().sum()
```

[6]:
```
Unique Key                      0
Created Date                    0
Closed Date                  2164
Agency                          0
Agency Name                     0
Complaint Type                  0
Descriptor                   5914
```

```
Location Type                       131
Incident Zip                       2615
Incident Address                  44410
Street Name                       44410
Cross Street 1                    49279
Cross Street 2                    49779
Intersection Street 1            256840
Intersection Street 2            257336
Address Type                       2815
City                               2614
Landmark                         300349
Facility Type                      2171
Status                                0
Due Date                              3
Resolution Description                0
Resolution Action Updated Date     2187
Community Board                       0
Borough                               0
X Coordinate (State Plane)         3540
Y Coordinate (State Plane)         3540
Park Facility Name                    0
Park Borough                          0
School Name                           0
School Number                         0
School Region                         1
School Code                           1
School Phone Number                   0
School Address                        0
School City                           0
School State                          0
School Zip                            1
School Not Found                      0
School or Citywide Complaint     300698
Vehicle Type                     300698
Taxi Company Borough             300698
Taxi Pick Up Location            300698
Bridge Highway Name              300455
Bridge Highway Direction         300455
Road Ramp                        300485
Bridge Highway Segment           300485
Garage Lot Name                  300698
Ferry Direction                  300697
Ferry Terminal Name              300696
Latitude                           3540
Longitude                          3540
Location                           3540
dtype: int64
```

[7]:
```
for column in df.columns:
    if df[column].isnull().sum()>250000:
        df.drop([column],axis=1,inplace=True)
```

[8]:
```
df.shape
```

[8]:
```
(300698, 39)
```

[9]:
```
df.dropna(inplace=True)
```

[10]:
```
df.shape
```

[10]:
```
(245664, 39)
```

[11]:
```
df.isnull().sum()
```

[11]:
```
Unique Key                        0
Created Date                      0
Closed Date                       0
Agency                            0
Agency Name                       0
Complaint Type                    0
Descriptor                        0
Location Type                     0
Incident Zip                      0
Incident Address                  0
Street Name                       0
Cross Street 1                    0
Cross Street 2                    0
Address Type                      0
City                              0
Facility Type                     0
Status                            0
Due Date                          0
Resolution Description            0
Resolution Action Updated Date    0
Community Board                   0
Borough                           0
X Coordinate (State Plane)        0
Y Coordinate (State Plane)        0
Park Facility Name                0
Park Borough                      0
School Name                       0
School Number                     0
School Region                     0
School Code                       0
School Phone Number               0
```

```
School Address                    0
School City                       0
School State                      0
School Zip                        0
School Not Found                  0
Latitude                          0
Longitude                         0
Location                          0
dtype: int64
```

[12]:
```
df.duplicated().value_counts()
```

[12]:
```
False    245664
dtype: int64
```

[ ]:
```
#TASK2
```

[13]:
```
df['Created Date'] = pd.to_datetime(df['Created Date'])
df['Closed Date'] = pd.to_datetime(df['Closed Date'])
```

[14]:
```
df.dtypes
```

[14]:
```
Unique Key                           int64
Created Date                datetime64[ns]
Closed Date                 datetime64[ns]
Agency                              object
Agency Name                         object
Complaint Type                      object
Descriptor                          object
Location Type                       object
Incident Zip                       float64
Incident Address                    object
Street Name                         object
Cross Street 1                      object
Cross Street 2                      object
Address Type                        object
City                                object
Facility Type                       object
Status                              object
Due Date                            object
Resolution Description              object
Resolution Action Updated Date      object
Community Board                     object
Borough                             object
X Coordinate (State Plane)         float64
Y Coordinate (State Plane)         float64
Park Facility Name                  object
```

```
Park Borough                          object
School Name                           object
School Number                         object
School Region                         object
School Code                           object
School Phone Number                   object
School Address                        object
School City                           object
School State                          object
School Zip                            object
School Not Found                      object
Latitude                             float64
Longitude                            float64
Location                              object
dtype: object
```

[15]:
```
df['Request_Closing_Time'] = df['Closed Date']-df['Created Date']
```

[16]:
```
df.dtypes
```

[16]:
```
Unique Key                                int64
Created Date                     datetime64[ns]
Closed Date                      datetime64[ns]
Agency                                   object
Agency Name                              object
Complaint Type                           object
Descriptor                               object
Location Type                            object
Incident Zip                            float64
Incident Address                         object
Street Name                              object
Cross Street 1                           object
Cross Street 2                           object
Address Type                             object
City                                     object
Facility Type                            object
Status                                   object
Due Date                                 object
Resolution Description                   object
Resolution Action Updated Date           object
Community Board                          object
Borough                                  object
X Coordinate (State Plane)              float64
Y Coordinate (State Plane)              float64
Park Facility Name                       object
Park Borough                             object
School Name                              object
```

```
School Number                            object
School Region                            object
School Code                              object
School Phone Number                      object
School Address                           object
School City                              object
School State                             object
School Zip                               object
School Not Found                         object
Latitude                                float64
Longitude                               float64
Location                                 object
Request_Closing_Time           timedelta64[ns]
dtype: object
```

[ ]:
```
#TASK3
```

[17]:
```
df['Community Board'].value_counts().head(10)
```

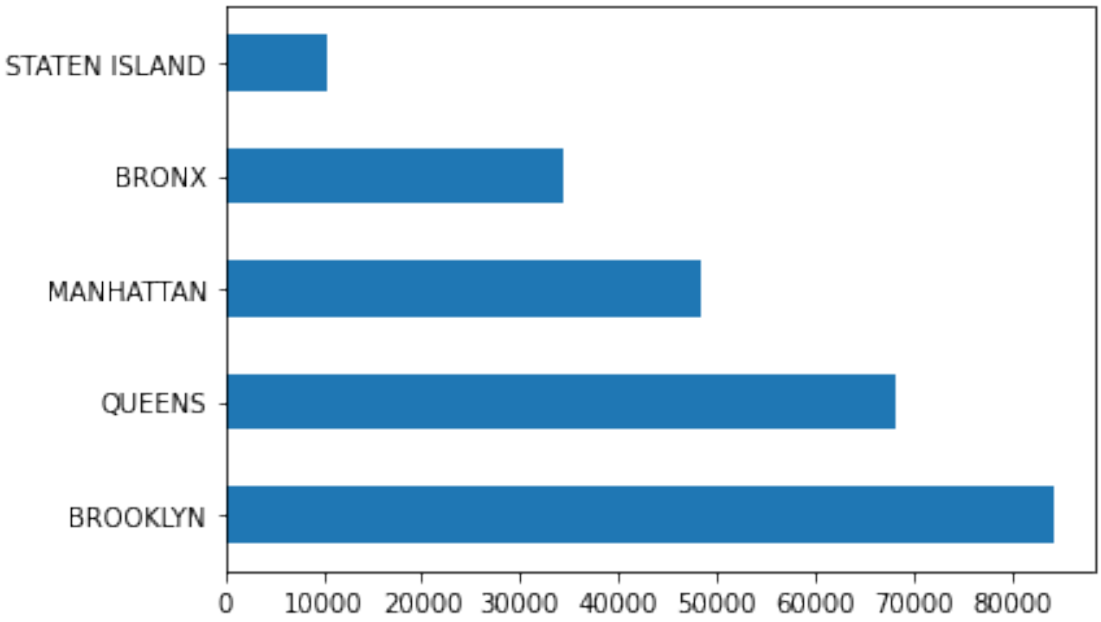
[17]:
```
12 MANHATTAN    10576
01 BROOKLYN      8676
05 QUEENS        7879
01 QUEENS        7828
09 QUEENS        7169
18 BROOKLYN      6671
12 BROOKLYN      6429
07 QUEENS        6370
03 BROOKLYN      6016
11 BROOKLYN      5735
Name: Community Board, dtype: int64
```

[18]:
```
df['Borough'].value_counts()
```

[18]:
```
BROOKLYN         84197
QUEENS           68237
MANHATTAN        48429
BRONX            34518
STATEN ISLAND    10283
Name: Borough, dtype: int64
```

[19]:
```
df['Borough'].value_counts().plot(kind='barh')
```

[19]:
```
<AxesSubplot:>
```



[20]:
```
df['City'].value_counts().head(10)
```

[20]:
```
BROOKLYN         84197
NEW YORK         48429
BRONX            34518
STATEN ISLAND    10283
JAMAICA           5706
ASTORIA           5517
FLUSHING          5170
RIDGEWOOD         4373
CORONA            3918
WOODSIDE          2973
Name: City, dtype: int64
```

[21]:
```
df['Descriptor'].value_counts().head(10)
```

[21]:
```
No Access                        55656
Loud Music/Party                 48829
Partial Access                   19635
Loud Talking                     17671
Posted Parking Sign Violation    17231
Blocked Hydrant                  14505
With License Plate               14477
Commercial Overnight Parking      9541
Blocked Sidewalk                  9132
Car/Truck Music                   8442
```

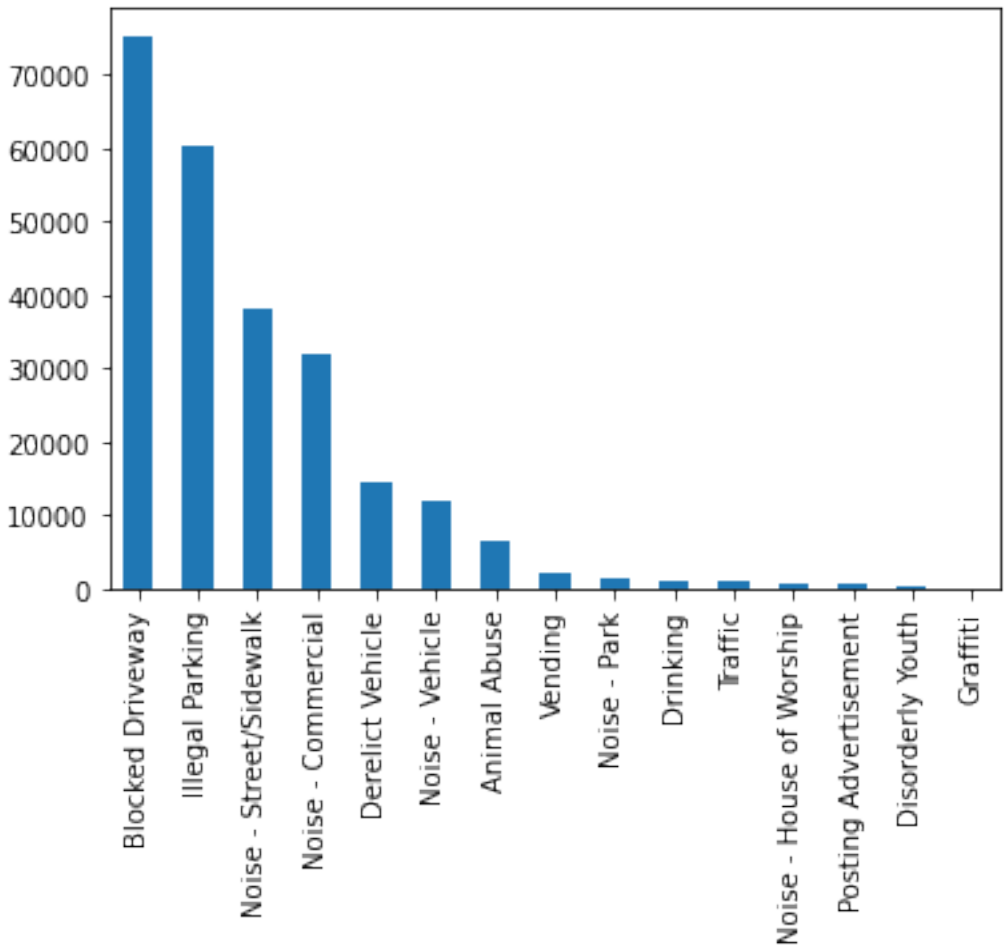
```
Name: Descriptor, dtype: int64
```

[22]:
```
df['Complaint Type'].value_counts()
```

[22]:
```
Blocked Driveway            75291
Illegal Parking             60145
Noise - Street/Sidewalk     38033
Noise - Commercial          32020
Derelict Vehicle            14477
Noise - Vehicle             12122
Animal Abuse                 6573
Vending                      2211
Noise - Park                 1270
Drinking                     1007
Traffic                       918
Noise - House of Worship      696
Posting Advertisement         621
Disorderly Youth              185
Graffiti                       95
Name: Complaint Type, dtype: int64
```

[26]:
```
df['Complaint Type'].value_counts().plot(kind='bar')
```

[26]:
```
<AxesSubplot:>
```



[31]:
```
df['Request_Closing_Time'] = df['Request_Closing_Time'].astype('timedelta64[s]')
```

[32]:
```
df['Request_Closing_Time']
```

[32]:
```
0          3315.0
1          5176.0
2         17491.0
3         27914.0
4         12422.0
             ...
300692     2309.0
300694     7231.0
300695    11237.0
300696    14733.0
300697    14929.0
Name: Request_Closing_Time, Length: 245664, dtype: float64
```

[ ]:
```
#TASK4
```

[33]:
```
grouped = df.groupby(['Location Type','Complaint Type']).mean()
grouped.head(5)
```

[33]:

Location Type	Complaint Type	Unique Key	Incident Zip	X Coordinate (State Plane)	Y Coordinate (State Plane)	Latitude	Longitude	Request_Closing_Time
Club/Bar/Restaurant	Drinking	3.130877e+07	10753.664557	1.004922e+06	205100.670886	40.729576	-73.925384	15942.683544
	Noise - Commercial	3.130618e+07	10667.385033	9.993469e+05	206492.532710	40.733421	-73.945510	11047.840790
Commercial	Animal Abuse	3.153133e+07	10362.310345	1.001490e+06	215999.931034	40.759513	-73.937776	14429.344828
Highway	Derelict Vehicle	3.131128e+07	10236.500000	1.002373e+06	230754.000000	40.800018	-73.934512	4616.500000
House and Store	Animal Abuse	3.136862e+07	11027.061728	1.011212e+06	193236.703704	40.696980	-73.902710	18747.876543

[34]:
```
grouped.iloc[:,-1]
```

[34]:
```
Location Type               Complaint Type
Club/Bar/Restaurant         Drinking                   15942.683544
                            Noise - Commercial         11047.840790
Commercial                  Animal Abuse               14429.344828
Highway                     Derelict Vehicle            4616.500000
House and Store             Animal Abuse               18747.876543
House of Worship            Noise - House of Worship   11893.221264
Park/Playground             Animal Abuse               10423.375000
```

```
                                Drinking                    12676.105263
                                Noise - Park                12597.240157
                                Vending                     12589.619048
Parking Lot                     Animal Abuse                22317.453125
                                Posting Advertisement        7616.714286
Residential Building            Animal Abuse                17656.799020
Residential Building/House      Animal Abuse                19536.332025
                                Disorderly Youth            14563.044118
                                Drinking                    12722.051095
                                Graffiti                    19841.870370
                                Posting Advertisement       12261.113208
                                Vending                     15624.861635
Roadway Tunnel                  Derelict Vehicle            59095.750000
Store/Commercial                Animal Abuse                15120.742788
                                Disorderly Youth             9968.142857
                                Drinking                    11586.704225
                                Graffiti                    20605.724138
                                Noise - Commercial          11447.456800
                                Posting Advertisement        9126.000000
                                Vending                     14561.705556
Street/Sidewalk                 Animal Abuse                20585.888035
                                Blocked Driveway            17063.280153
                                Derelict Vehicle            26022.454697
                                Disorderly Youth            12286.118182
                                Drinking                    13287.055046
                                Graffiti                    28185.666667
                                Illegal Parking             16128.593133
                                Noise - Street/Sidewalk     12551.489154
                                Noise - Vehicle             13316.820492
                                Posting Advertisement        6161.787770
                                Traffic                     14100.942266
                                Vending                     14152.691203
Subway Station                  Animal Abuse                12222.750000
Vacant Lot                      Derelict Vehicle            27153.736842
Name: Request_Closing_Time, dtype: float64
```

1 TASK5

Ho: Complaint Type and Location is not related Ha: Complaint Type and Location is related alpha = 0.05

[66]:
```
contingency_table = pd.crosstab(df['Complaint Type'],df['Location Type'])
```

[67]:
```
contingency_table
```

[67]:

Location Type / Complaint Type	Club/Bar/Restaurant	Commercial	Highway
Animal Abuse	0	29	0
Blocked Driveway	0	0	0
Derelict Vehicle	0	0	2
Disorderly Youth	0	0	0
Drinking	316	0	0
Graffiti	0	0	0
Illegal Parking	0	0	0
Noise - Commercial	15087	0	0
Noise - House of Worship	0	0	0
Noise - Park	0	0	0
Noise - Street/Sidewalk	0	0	0
Noise - Vehicle	0	0	0
Posting Advertisement	0	0	0
Traffic	0	0	0
Vending	0	0	0

Location Type / Complaint Type	House and Store	House of Worship	Park/Playground
Animal Abuse	81	0	24
Blocked Driveway	0	0	0
Derelict Vehicle	0	0	0
Disorderly Youth	0	0	0
Drinking	0	0	19
Graffiti	0	0	0
Illegal Parking	0	0	0
Noise - Commercial	0	0	0
Noise - House of Worship	0	696	0
Noise - Park	0	0	1270
Noise - Street/Sidewalk	0	0	0
Noise - Vehicle	0	0	0
Posting Advertisement	0	0	0
Traffic	0	0	0
Vending	0	0	21

Location Type / Complaint Type	Parking Lot	Residential Building
Animal Abuse	64	204
Blocked Driveway	0	0
Derelict Vehicle	0	0
Disorderly Youth	0	0
Drinking	0	0
Graffiti	0	0
Illegal Parking	0	0
Noise - Commercial	0	0
Noise - House of Worship	0	0

Noise - Park	0	0
Noise - Street/Sidewalk	0	0
Noise - Vehicle	0	0
Posting Advertisement	7	0
Traffic	0	0
Vending	0	0

Location Type	Residential Building/House	Roadway Tunnel	\
Complaint Type			
Animal Abuse	4840	0	
Blocked Driveway	0	0	
Derelict Vehicle	0	4	
Disorderly Youth	68	0	
Drinking	274	0	
Graffiti	54	0	
Illegal Parking	0	0	
Noise - Commercial	0	0	
Noise - House of Worship	0	0	
Noise - Park	0	0	
Noise - Street/Sidewalk	0	0	
Noise - Vehicle	0	0	
Posting Advertisement	53	0	
Traffic	0	0	
Vending	159	0	

Location Type	Store/Commercial	Street/Sidewalk	Subway Station	\
Complaint Type				
Animal Abuse	416	911	4	
Blocked Driveway	0	75291	0	
Derelict Vehicle	0	14414	0	
Disorderly Youth	7	110	0	
Drinking	71	327	0	
Graffiti	29	12	0	
Illegal Parking	0	60145	0	
Noise - Commercial	16933	0	0	
Noise - House of Worship	0	0	0	
Noise - Park	0	0	0	
Noise - Street/Sidewalk	0	38033	0	
Noise - Vehicle	0	12122	0	
Posting Advertisement	5	556	0	
Traffic	0	918	0	
Vending	360	1671	0	

Location Type	Vacant Lot
Complaint Type	
Animal Abuse	0
Blocked Driveway	0

```
Derelict Vehicle                 57
Disorderly Youth                  0
Drinking                          0
Graffiti                          0
Illegal Parking                   0
Noise - Commercial                0
Noise - House of Worship          0
Noise - Park                      0
Noise - Street/Sidewalk           0
Noise - Vehicle                   0
Posting Advertisement             0
Traffic                           0
Vending                           0
```

[68]:
```
Observed_Values = contingency_table.values
```

[70]:
```
from scipy import stats
```

[78]:
```
b =stats.chi2_contingency(contingency_table)
```

[72]:
```
Expected_Values = b[3]
```

[73]:
```
from scipy.stats import chi2
```

[74]:
```
p_value=1-chi2.cdf(x=b[0],df=1)
```

[79]:
```
p_value
```

[79]: 0.0

p_value is less than 0.05 so we reject the null hypothesis that Complaint Type and Location Type is not related.