Project 2 - Customer Service Requests Analysis

December 19, 2022

[1]:

```
# import the library
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

1 1. Import a 311 NYC service request.

[10]:

```
df = pd.read_csv('NYC311_Service_Requests_from_2010_to_Present.csv')
```

```
C:\Users\Vinosh\AppData\Local\Temp\ipykernel_9808\2749672019.py:1: DtypeWarning:
Columns (48,49) have mixed types. Specify dtype option on import or set
low_memory=False.
  df = pd.read_csv('NYC311_Service_Requests_from_2010_to_Present.csv')
```

[11]:

```
df
```

[11]:

```
        Unique Key            Created Date             Closed Date Agency  \
0         32310363  12/31/2015 11:59:45 PM           01-01-16 0:55   NYPD
1         32309934  12/31/2015 11:59:44 PM           01-01-16 1:26   NYPD
2         32309159  12/31/2015 11:59:29 PM           01-01-16 4:51   NYPD
3         32305098  12/31/2015 11:57:46 PM           01-01-16 7:43   NYPD
4         32306529  12/31/2015 11:56:58 PM           01-01-16 3:24   NYPD
…              …                       …                       …     …
300693    30281872  03/29/2015 12:33:41 AM                     NaN   NYPD
300694    30281230  03/29/2015 12:33:28 AM  03/29/2015 02:33:59 AM   NYPD
300695    30283424  03/29/2015 12:33:03 AM  03/29/2015 03:40:20 AM   NYPD
300696    30280004  03/29/2015 12:33:02 AM  03/29/2015 04:38:35 AM   NYPD
300697    30281825  03/29/2015 12:33:01 AM  03/29/2015 04:41:50 AM   NYPD

                            Agency Name           Complaint Type  \
0       New York City Police Department  Noise - Street/Sidewalk
1       New York City Police Department         Blocked Driveway
2       New York City Police Department         Blocked Driveway
3       New York City Police Department          Illegal Parking
4       New York City Police Department          Illegal Parking
…                                     …                       …
```

```
300693  New York City Police Department       Noise - Commercial
300694  New York City Police Department         Blocked Driveway
300695  New York City Police Department       Noise - Commercial
300696  New York City Police Department       Noise - Commercial
300697  New York City Police Department       Noise - Commercial

                          Descriptor        Location Type  Incident Zip  \
0                   Loud Music/Party      Street/Sidewalk       10034.0
1                          No Access      Street/Sidewalk       11105.0
2                          No Access      Street/Sidewalk       10458.0
3       Commercial Overnight Parking      Street/Sidewalk       10461.0
4                   Blocked Sidewalk      Street/Sidewalk       11373.0
…                              …                  …           …
300693              Loud Music/Party  Club/Bar/Restaurant           NaN
300694                Partial Access      Street/Sidewalk       11418.0
300695              Loud Music/Party  Club/Bar/Restaurant       11206.0
300696              Loud Music/Party  Club/Bar/Restaurant       10461.0
300697              Loud Music/Party     Store/Commercial       10036.0

                Incident Address  … Bridge Highway Name  \
0            71 VERMILYEA AVENUE  …                 NaN
1                27-07 23 AVENUE  …                 NaN
2          2897 VALENTINE AVENUE  …                 NaN
3            2940 BAISLEY AVENUE  …                 NaN
4                  87-14 57 ROAD  …                 NaN
…                          …  …                 …
300693           CRESCENT AVENUE  …                 NaN
300694          100-17 87 AVENUE  …                 NaN
300695         162 THROOP AVENUE  …                 NaN
300696  3151 EAST TREMONT AVENUE  …                 NaN
300697        251 WEST 48 STREET  …                 NaN

       Bridge Highway Direction Road Ramp Bridge Highway Segment  \
0                           NaN       NaN                    NaN
1                           NaN       NaN                    NaN
2                           NaN       NaN                    NaN
3                           NaN       NaN                    NaN
4                           NaN       NaN                    NaN
…                         …       …                    …
300693                      NaN       NaN                    NaN
300694                      NaN       NaN                    NaN
300695                      NaN       NaN                    NaN
300696                      NaN       NaN                    NaN
300697                      NaN       NaN                    NaN

       Garage Lot Name Ferry Direction Ferry Terminal Name   Latitude  \
0                  NaN             NaN                 NaN  40.865682
```

```
1                  NaN             NaN                 NaN  40.775945
2                  NaN             NaN                 NaN  40.870325
3                  NaN             NaN                 NaN  40.835994
4                  NaN             NaN                 NaN  40.733060
…                …             …                 …        …
300693             NaN             NaN                 NaN        NaN
300694             NaN             NaN                 NaN  40.694077
300695             NaN             NaN                 NaN  40.699590
300696             NaN             NaN                 NaN  40.837708
300697             NaN             NaN                 NaN  40.760583

        Longitude                                  Location
0      -73.923501   (40.86568153633767, -73.92350095571744)
1      -73.915094  (40.775945312321085, -73.91509393898605)
2      -73.888525  (40.870324522111424, -73.88852464418646)
3      -73.828379   (40.83599404683083, -73.82837939584206)
4      -73.874170  (40.733059618956815, -73.87416975810375)
…           …                                       …
300693        NaN                                       NaN
300694 -73.846087    (40.69407728322387, -73.8460866160573)
300695 -73.944234   (40.69959035300927, -73.94423377144169)
300696 -73.834587    (40.8377075854206, -73.83458731019586)
300697 -73.985922   (40.76058322950115, -73.98592204392392)

[300698 rows x 53 columns]
```

[12]:
```
df.describe()
```

[12]:
```
         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
```

[13]:

```
df.shape
```

[13]:

```
(300698, 53)
```

2 2. Read or convert the columns 'Created Date' and Closed Date' to datetime datatype and create a new column 'Request_Closing_Time' as the time elapsed between request creation and request closing.

[41]:

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

[39]:

```
df['Request_Closing_Time']=(df['Closed Date']-df['Created Date'])

Request_Closing_Time=[]
for x in (df['Closed Date']-df['Created Date']):
    close=x.total_seconds()/60
    Request_Closing_Time.append(close)

df['Request_Closing_Time']=Request_Closing_Time
```

[40]:

```
df['Request_Closing_Time'].head()
```

[40]:

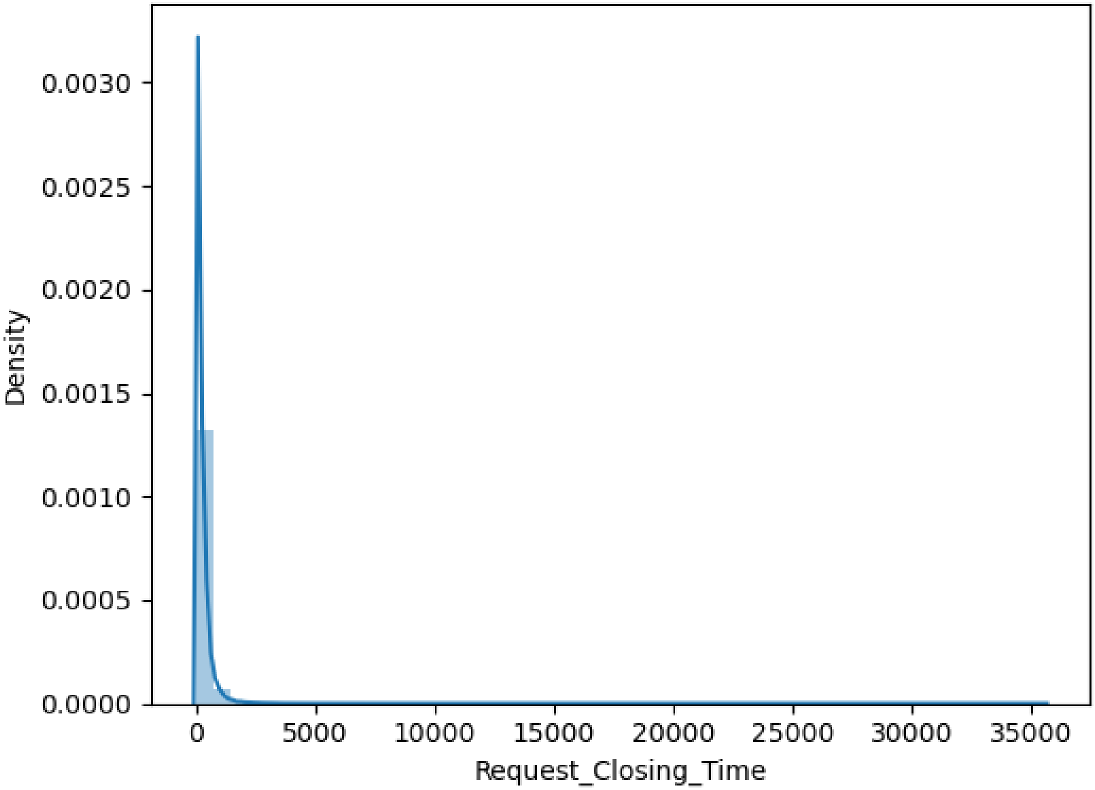
```
0     55.250000
1     86.266667
2    291.516667
3    465.233333
4    207.033333
Name: Request_Closing_Time, dtype: float64
```

3 3. Provide major insights/patterns that you can offer in a visual format (graphs or tables); at least 4 major conclusions that you can come up with after generic data mining

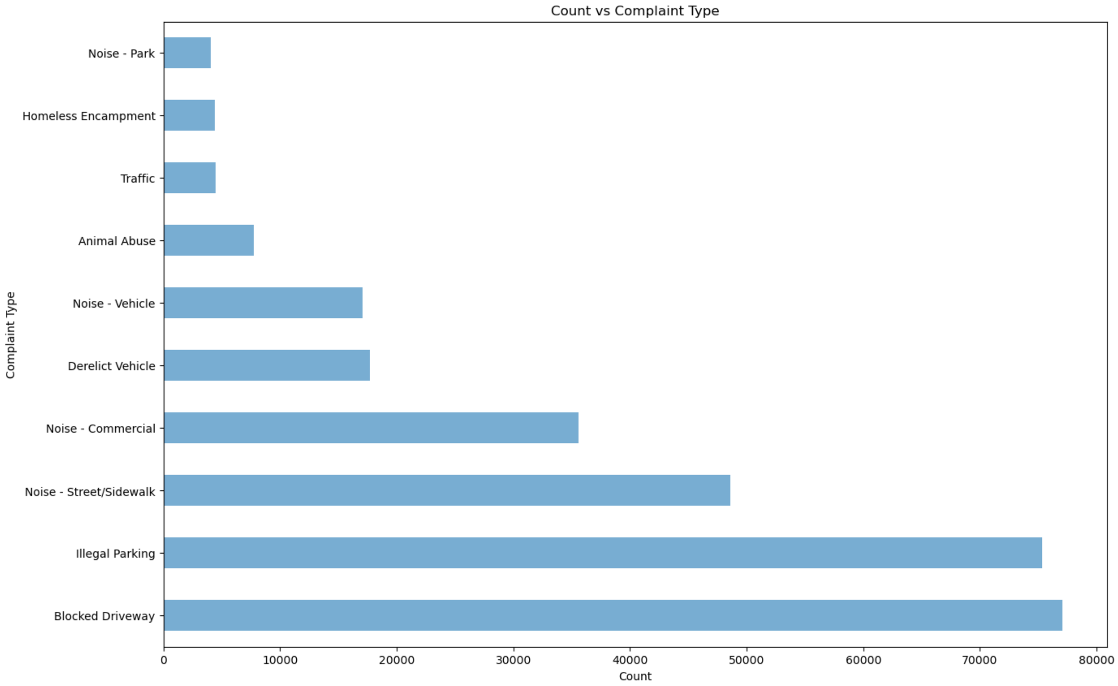
[23]:

```
sns.distplot(df['Request_Closing_Time'])
plt.show()
```

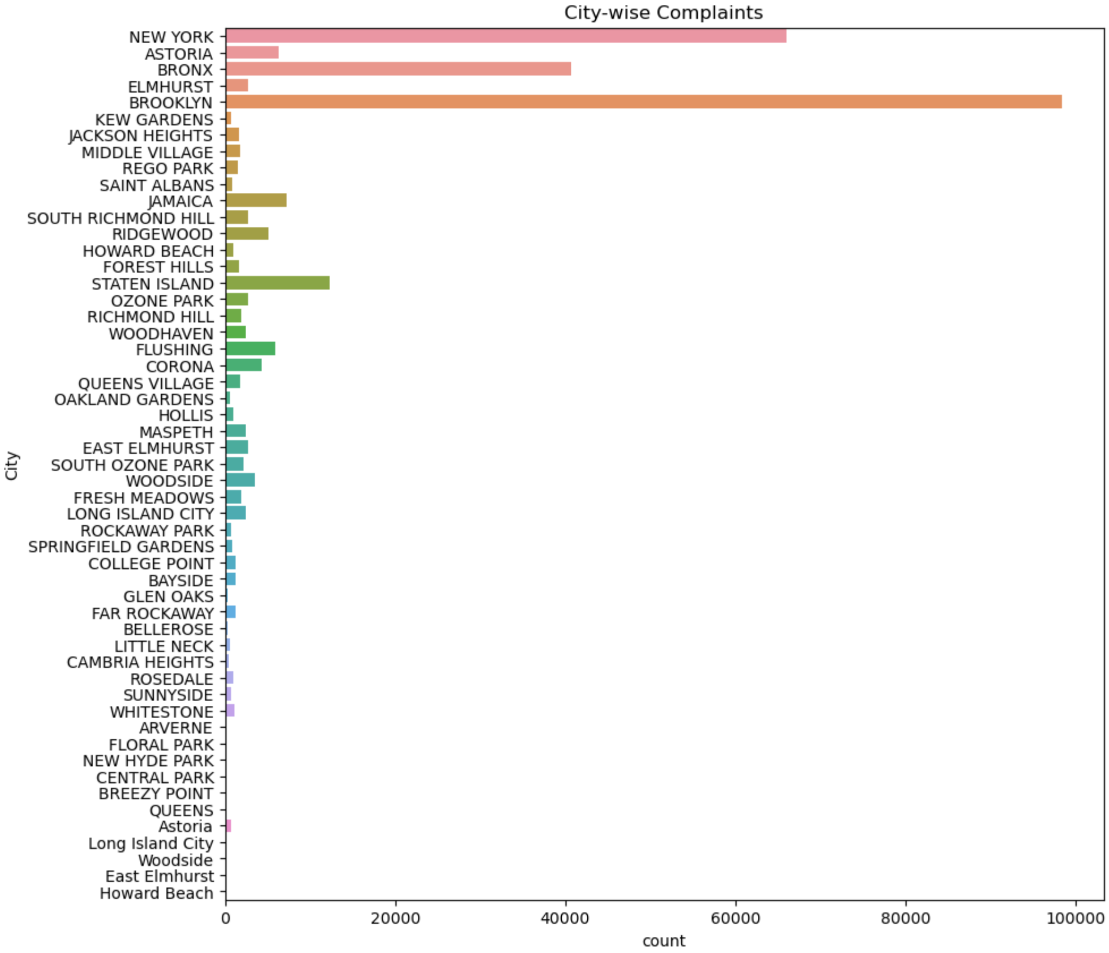
```
C:\Users\Vinosh\anaconda3\lib\site-packages\seaborn\distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
  warnings.warn(msg, FutureWarning)
```



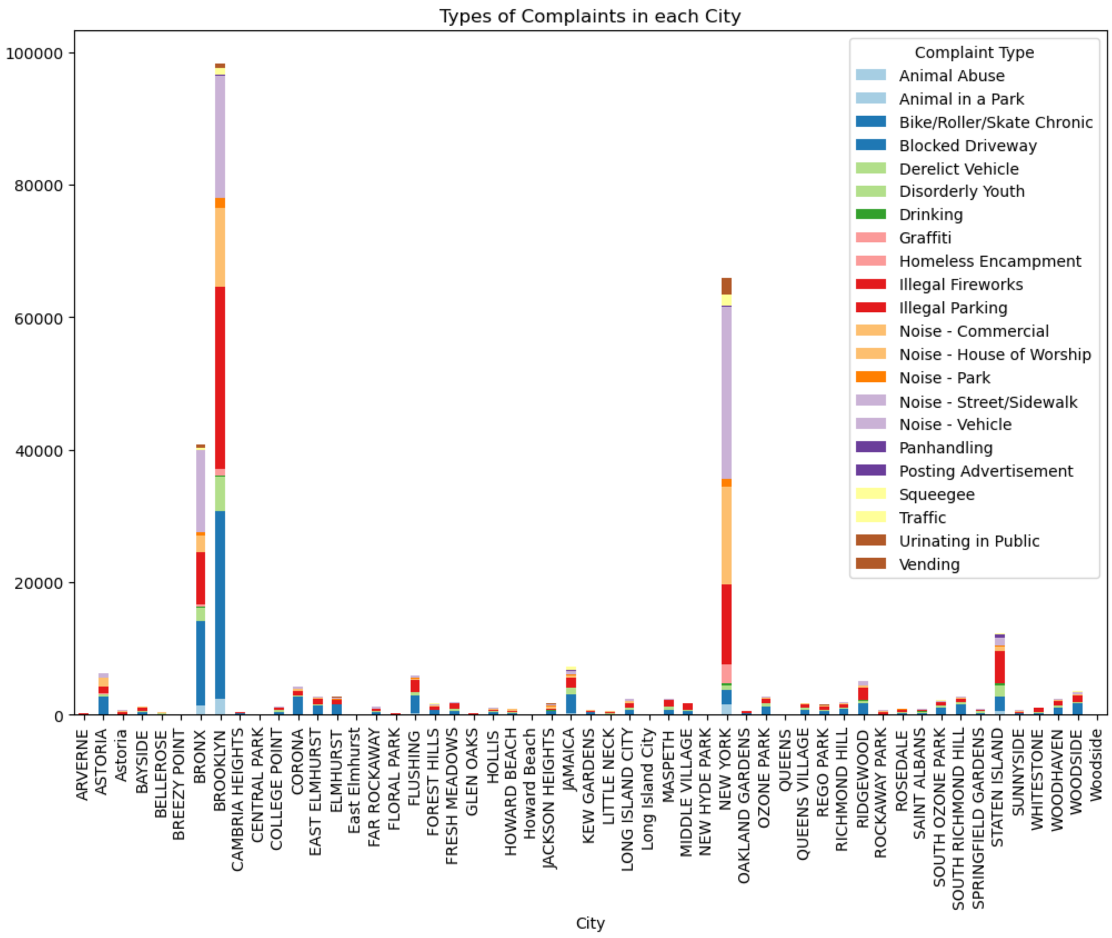
[28]:
```
df['Complaint Type'].value_counts()[:10].plot(kind='barh',alpha=0.
 ↪6,figsize=(15,10))
plt.xlabel('Count')
plt.ylabel('Complaint Type')
plt.title('Count vs Complaint Type')
plt.show()
```



[27]:
```
plt.figure(figsize=(10,10))
sns.countplot(y=df['City'])
plt.title('City-wise Complaints')
plt.show()
```



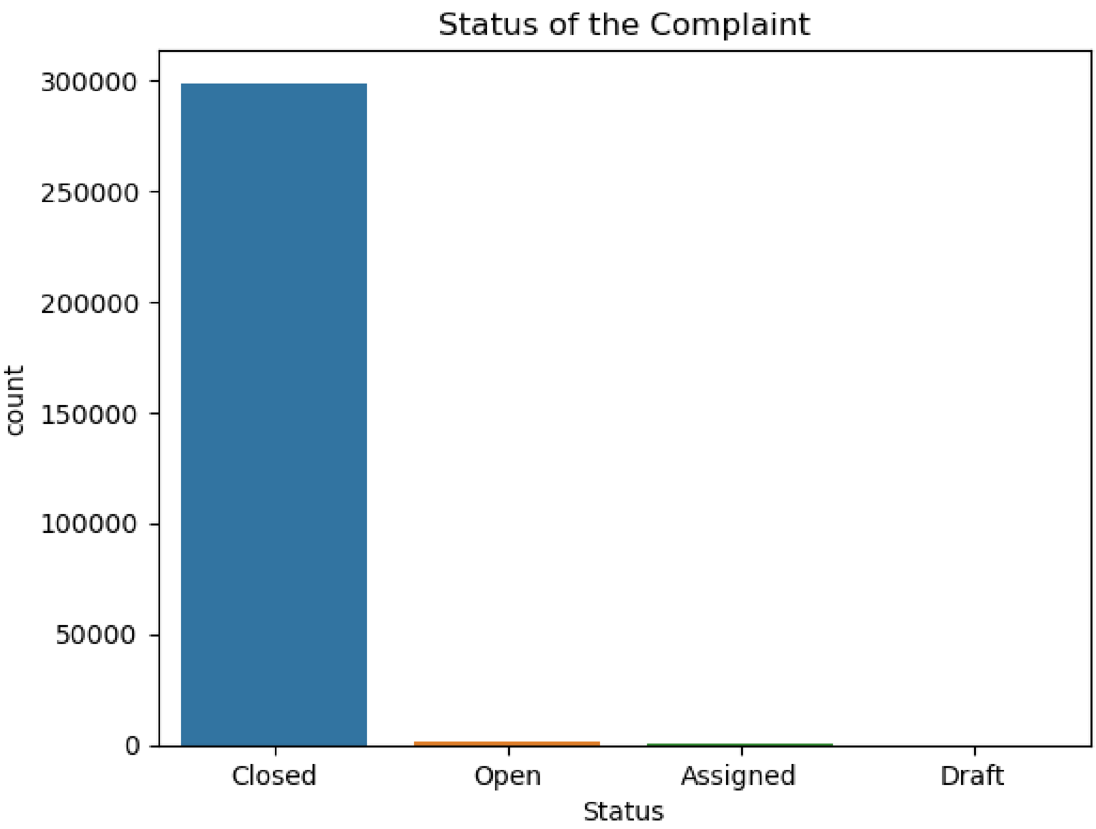
```
[31]: crosstab=pd.crosstab(index=df['City'],columns=df['Complaint Type'])
      crosstab.plot(kind='bar',figsize=(12,8),stacked=True,colormap='Paired')
      plt.title('Types of Complaints in each City')
      plt.show()
```



[34]:

```
sns.countplot(df['Status'])
plt.title('Status of the Complaint')
plt.show()
```

```
C:\Users\Vinosh\anaconda3\lib\site-packages\seaborn\_decorators.py:36:
FutureWarning: Pass the following variable as a keyword arg: x. From version
0.12, the only valid positional argument will be `data`, and passing other
arguments without an explicit keyword will result in an error or
misinterpretation.
  warnings.warn(
```



4 4. Order the complaint types based on the average 'Request_Closing_Time', grouping them for different locations

```
[36]: df['Location Type'].unique()
```

```
[36]: array(['Street/Sidewalk', 'Club/Bar/Restaurant', 'Store/Commercial',
             'House of Worship', 'Residential Building/House',
             'Residential Building', 'Park/Playground', 'Vacant Lot',
             'House and Store', 'Highway', 'Commercial', 'Roadway Tunnel',
             'Subway Station', 'Parking Lot', 'Bridge', 'Terminal', nan,
             'Ferry', 'Park'], dtype=object)
```

```
[37]: pd.DataFrame(df.groupby('Location Type')['Request_Closing_Time'].mean()).
       ↪sort_values('Request_Closing_Time')
```

```
[37]:                            Request_Closing_Time
      Location Type
      Subway Station                       142.250980
      Club/Bar/Restaurant                  186.074330
```

```
House of Worship                      191.833279
Store/Commercial                      198.089073
Park/Playground                       207.137129
Highway                               223.424221
Bridge                                229.158333
Roadway Tunnel                        266.525714
Street/Sidewalk                       268.515306
Residential Building                  289.089941
House and Store                       300.795699
Residential Building/House            309.505679
Parking Lot                           320.130342
Commercial                            320.566129
Vacant Lot                            448.435498
Park                                20210.083333
Ferry                                        NaN
Terminal                                     NaN
```

5 5. Perform a statistical test for the following:

Please note: For the below statements you need to state the Null and Alternate and then provide a statistical test to accept or reject the Null Hypothesis along with the corresponding 'p-value'.

Whether the average response time across complaint types is similar or not (overall), Are the type of complaint or service requested and location related?

[53]:

```
from scipy import stats
from scipy.stats import chi2_contingency

import statsmodels.api as sm
from statsmodels.formula.api import ols
```

[42]:

```
new_df=df.loc[:,(df.isnull().sum()/df.shape[0]*100)<=50]
```

[43]:

```
print('Old DataFrame Shape :',df.shape)
print('New DataFrame Shape : ',new_df.shape)
```

```
Old DataFrame Shape : (300698, 54)
New DataFrame Shape :  (300698, 40)
```

[47]:

```
rem=[]
for x in new_df.columns.tolist():
    if new_df[x].nunique()<=3:
        print(x+ ' '*10+' : ',new_df[x].unique())
        rem.append(x)
```

```
Agency           :  ['NYPD']
Agency Name           :  ['New York City Police Department' 'NYPD' 'Internal
Affairs Bureau']
Facility Type             :  ['Precinct' nan]
```

```
Park Facility Name           :  ['Unspecified' 'Alley Pond Park - Nature
Center']
School Name           :  ['Unspecified' 'Alley Pond Park - Nature Center']
School Number           :  ['Unspecified' 'Q001']
School Region           :  ['Unspecified' nan]
School Code           :  ['Unspecified' nan]
School Phone Number           :  ['Unspecified' '7182176034']
School Address           :  ['Unspecified' 'Grand Central Parkway, near the
soccer field']
School City           :  ['Unspecified' 'QUEENS']
School State           :  ['Unspecified' 'NY']
School Zip           :  ['Unspecified' nan]
School Not Found           :  ['N']
```

[48]:

```
new_df.drop(rem,axis=1,inplace=True)
```

```
C:\Users\Vinosh\AppData\Local\Temp\ipykernel_9808\2260080562.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  new_df.drop(rem,axis=1,inplace=True)
```

[49]:

```
new_df.shape
```

[49]:

```
(300698, 26)
```

[50]:

```
rem1=["Unique Key","Incident Address","Descriptor","Street Name","Cross Street␣
 ↪1","Cross Street 2","Due Date","Resolution Description","Resolution Action␣
 ↪Updated Date","Community Board","X Coordinate (State Plane)","Y Coordinate␣
 ↪(State Plane)","Park Borough","Latitude","Longitude","Location"]

new_df.drop(rem1,axis=1,inplace=True)
```

```
C:\Users\Vinosh\AppData\Local\Temp\ipykernel_9808\1160912026.py:3:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  new_df.drop(rem1,axis=1,inplace=True)
```

[51]:

```
new_df.head()
```

[51]:

```
         Created Date         Closed Date           Complaint Type  \
0 2015-12-31 23:59:45 2016-01-01 00:55:00  Noise - Street/Sidewalk
1 2015-12-31 23:59:44 2016-01-01 01:26:00         Blocked Driveway
2 2015-12-31 23:59:29 2016-01-01 04:51:00         Blocked Driveway
```

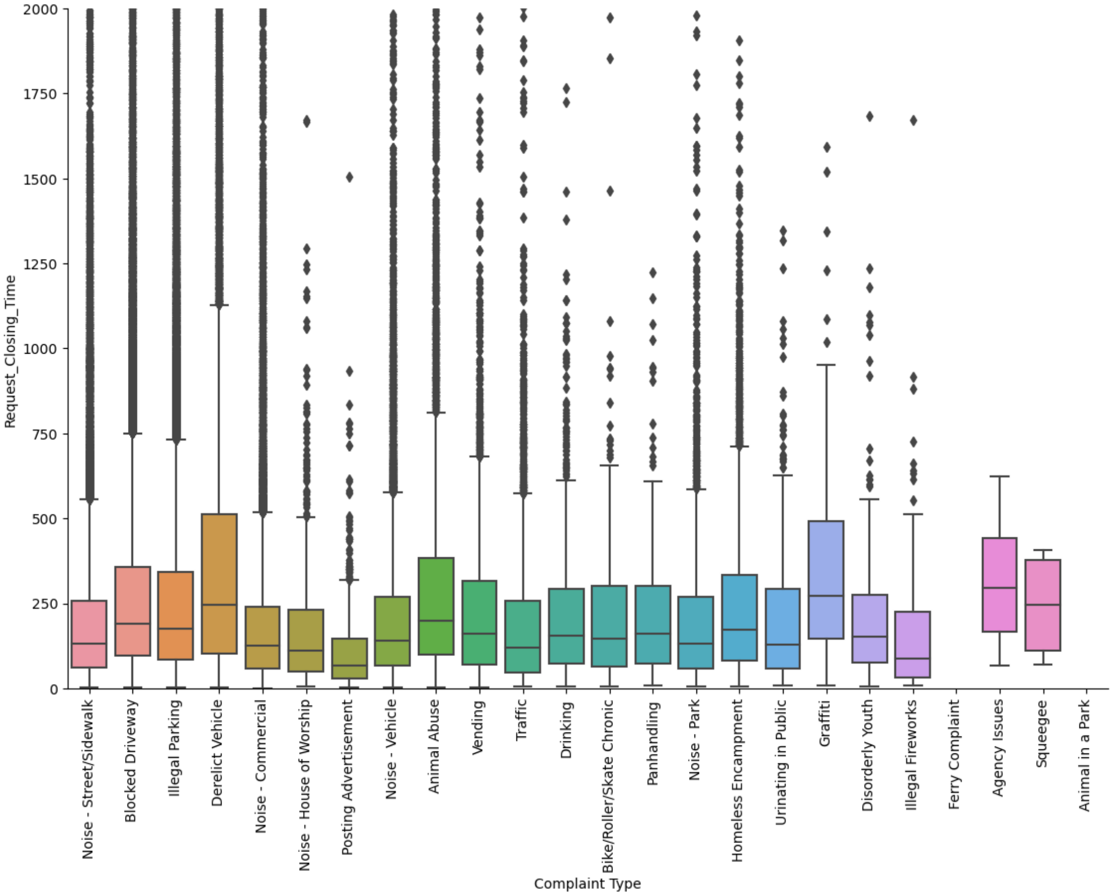
```
3 2015-12-31 23:57:46 2016-01-01 07:43:00          Illegal Parking
4 2015-12-31 23:56:58 2016-01-01 03:24:00          Illegal Parking

     Location Type  Incident Zip Address Type      City  Status    Borough  \
0  Street/Sidewalk       10034.0      ADDRESS  NEW YORK  Closed  MANHATTAN
1  Street/Sidewalk       11105.0      ADDRESS   ASTORIA  Closed     QUEENS
2  Street/Sidewalk       10458.0      ADDRESS     BRONX  Closed      BRONX
3  Street/Sidewalk       10461.0      ADDRESS     BRONX  Closed      BRONX
4  Street/Sidewalk       11373.0      ADDRESS  ELMHURST  Closed     QUEENS

   Request_Closing_Time
0             55.250000
1             86.266667
2            291.516667
3            465.233333
4            207.033333
```

[52]:

```
g=sns.catplot(x='Complaint␣
 ↪Type',y='Request_Closing_Time',kind='box',data=new_df)
g.fig.set_figheight(8)
g.fig.set_figwidth(15)
plt.xticks(rotation=90)
plt.ylim((0,2000))
```

[52]:

```
(0.0, 2000.0)
```



H0: there is no significant different in mean of Request__Closing__Time for different Complaint

H1:there is signficant different in mean of Request__Closing__Time for different Complaint

[55]:
```
anova_df=pd.DataFrame()
anova_df['Request_Closing_Time']=new_df['Request_Closing_Time']
anova_df['Complaint']=new_df['Complaint Type']

anova_df.dropna(inplace=True)
anova_df.head()
```

[55]:
```
   Request_Closing_Time                Complaint
0             55.250000  Noise - Street/Sidewalk
1             86.266667         Blocked Driveway
2            291.516667         Blocked Driveway
3            465.233333          Illegal Parking
4            207.033333          Illegal Parking
```

[57]:

```
lm=ols('Request_Closing_Time~Complaint',data=anova_df).fit()
table=sm.stats.anova_lm(lm)
table
```

[57]:

```
                  df        sum_sq       mean_sq           F  PR(>F)
Complaint       22.0  1.455049e+09  6.613860e+07  514.177089     0.0
Residual   298511.0  3.839747e+10  1.286300e+05         NaN     NaN
```

Since p value for the Complaint is less that 0.01 thus we accept alternate hypothesis

H0:Complaint Type and Location Type are independent

H1:Complaint Type and Location Type are related

[60]:

```
chi_sq=pd.DataFrame()
chi_sq['Location Type']=new_df['Location Type']
chi_sq['Complaint Type']=new_df['Complaint Type']

chi_sq.dropna(inplace=True)
```

[61]:

```
data_crosstab = pd.crosstab( chi_sq['Location Type'],chi_sq['Complaint Type'])
```

[62]:

```
stat, p, dof, expected = chi2_contingency(data_crosstab)

alpha = 0.05
if p <= alpha:
    print('Dependent (reject H0)')
else:
    print('Independent (H0 holds true)')
```

```
Dependent (reject H0)
```

Since p value for the chi square test is less than 0.05(LOS) we can conclude that Complaint Type is dependent on Location Type

[ ]:

```

```