Project 1 - 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
```

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

```
C:\Users\Vinosh\AppData\Local\Temp\ipykernel_8196\1424208058.py:1: DtypeWarning:
Columns (48,49) have mixed types. Specify dtype option on import or set
low_memory=False.
  dataset = pd.read_csv('311_Service_Requests_from_2010_to_Present.csv')
```

```
[3]: dataset
```

```
[3]:         Unique Key              Created Date              Closed Date Agency  \
     0        32310363  12/31/2015 11:59:45 PM  01/01/2016 12:55:15 AM   NYPD
     1        32309934  12/31/2015 11:59:44 PM  01/01/2016 01:26:57 AM   NYPD
     2        32309159  12/31/2015 11:59:29 PM  01/01/2016 04:51:03 AM   NYPD
     3        32305098  12/31/2015 11:57:46 PM  01/01/2016 07:43:13 AM   NYPD
     4        32306529  12/31/2015 11:56:58 PM  01/01/2016 03:24:42 AM   NYPD
     ...           ...                     ...                     ...    ...
     364553   29609918  01/01/2015 12:04:44 AM  01/01/2015 10:22:31 AM   NYPD
     364554   29608392  01/01/2015 12:04:28 AM  01/01/2015 02:25:02 AM   NYPD
     364555   29607589  01/01/2015 12:01:30 AM  01/01/2015 12:20:33 AM   NYPD
     364556   29610889  01/01/2015 12:01:29 AM  01/01/2015 02:42:22 AM   NYPD
     364557   29611816  01/01/2015 12:00:50 AM  01/01/2015 02:47:50 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
     ...                                  ...                       ...
     364553   New York City Police Department           Illegal Parking
     364554   New York City Police Department           Noise - Vehicle
     364555   New York City Police Department  Noise - Street/Sidewalk
```

```
364556  New York City Police Department        Blocked Driveway
364557  New York City Police Department        Blocked Driveway

                          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
…                                  …               …             …
364553              Blocked Hydrant   Street/Sidewalk       11421.0
364554              Car/Truck Horn    Street/Sidewalk       10468.0
364555              Loud Music/Party  Street/Sidewalk       10031.0
364556              No Access         Street/Sidewalk       10466.0
364557              No Access         Street/Sidewalk       11420.0

              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
…                            …  … …                 …
364553              84-25 85 ROAD  …                NaN
364554      2555 SEDGWICK AVENUE  …                NaN
364555        508 WEST 139 STREET  …                NaN
364556        931 EAST 226 STREET  …                NaN
364557         123-19 135 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
…                                    …         …                      …
364553                             NaN       NaN                    NaN
364554                             NaN       NaN                    NaN
364555                             NaN       NaN                    NaN
364556                             NaN       NaN                    NaN
364557                             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
...                 ...             ...                   ...        ...
364553            NaN             NaN                   NaN   40.695145
364554            NaN             NaN                   NaN   40.867830
364555            NaN             NaN                   NaN   40.821647
364556            NaN             NaN                   NaN   40.886361
364557            NaN             NaN                   NaN   40.674212

        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)
...            ...                                        ...
364553 -73.860949    (40.69514470265117, -73.86094888534394)
364554 -73.907178    (40.86782963689454, -73.90717786644662)
364555 -73.950873   (40.821646626438095, -73.95087342885292)
364556 -73.853290    (40.88636077906953, -73.85329048666742)
364557 -73.803585   (40.674211762243935, -73.80358548685278)

[364558 rows x 53 columns]
```

1 1. Understand the dataset:

1. Identify the shape of the dataset
2. Identify variables with null values

1.0.1 1) 1) Identify the shape of the dataset

```
[4]: dataset.shape
```

```
[4]: (364558, 53)
```

1.0.2 1) 2) Identify variables with null values

```
[5]: dataset.isnull().sum()
```

```
[5]: Unique Key              0
     Created Date            0
     Closed Date          2381
     Agency                  0
     Agency Name             0
     Complaint Type          0
     Descriptor           6501
     Location Type         133
```

```
Incident Zip                          2998
Incident Address                     51699
Street Name                          51699
Cross Street 1                       57188
Cross Street 2                       57805
Intersection Street 1               313438
Intersection Street 2               314046
Address Type                          3252
City                                  2997
Landmark                            364183
Facility Type                         2389
Status                                   0
Due Date                                 3
Resolution Description                   0
Resolution Action Updated Date        2402
Community Board                          0
Borough                                  0
X Coordinate (State Plane)            4030
Y Coordinate (State Plane)            4030
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        364558
Vehicle Type                        364558
Taxi Company Borough                364558
Taxi Pick Up Location               364558
Bridge Highway Name                 364261
Bridge Highway Direction            364261
Road Ramp                           364296
Bridge Highway Segment              364296
Garage Lot Name                     364558
Ferry Direction                     364557
Ferry Terminal Name                 364556
Latitude                              4030
Longitude                             4030
Location                              4030
dtype: int64
```

```
[6]: dataset.columns
```

```
[6]: Index(['Unique Key', 'Created Date', 'Closed Date', 'Agency', 'Agency Name',
            'Complaint Type', 'Descriptor', 'Location Type', 'Incident Zip',
            'Incident Address', 'Street Name', 'Cross Street 1', 'Cross Street 2',
            'Intersection Street 1', 'Intersection Street 2', 'Address Type',
            'City', 'Landmark', 'Facility Type', 'Status', 'Due Date',
            'Resolution Description', 'Resolution Action Updated Date',
            'Community Board', 'Borough', 'X Coordinate (State Plane)',
            'Y Coordinate (State Plane)', 'Park Facility Name', 'Park Borough',
            'School Name', 'School Number', 'School Region', 'School Code',
            'School Phone Number', 'School Address', 'School City', 'School State',
            'School Zip', 'School Not Found', 'School or Citywide Complaint',
            'Vehicle Type', 'Taxi Company Borough', 'Taxi Pick Up Location',
            'Bridge Highway Name', 'Bridge Highway Direction', 'Road Ramp',
            'Bridge Highway Segment', 'Garage Lot Name', 'Ferry Direction',
            'Ferry Terminal Name', 'Latitude', 'Longitude', 'Location'],
           dtype='object')
```

2 2. Perform basic data exploratory analysis:

1. Utilize missing value treatment

2. Analyze the date column and remove the entries if it has an incorrect timeline

3. Draw a frequency plot for city-wise complaints

4. Draw scatter and hexbin plots for complaint concentration across Brooklyn

2.0.1 2) 1) Utilize missing value treatment

```
[7]: service_dataset = dataset.drop(['School or Citywide Complaint','Vehicle␣
      ↪Type','Taxi Company Borough','Taxi Pick Up Location','Bridge Highway␣
      ↪Name','Bridge Highway Direction','Road Ramp','Bridge Highway␣
      ↪Segment','Garage Lot Name','Ferry Direction','Ferry Terminal␣
      ↪Name','Intersection Street 1','Intersection Street 2','Landmark','Park␣
      ↪Facility Name','School Name', 'School Number', 'School Region', 'School␣
      ↪Code','School Phone Number', 'School Address', 'School City', 'School␣
      ↪State','School Zip','School Not Found'],axis=1)
```

```
[8]: service_dataset.shape
```

```
[8]: (364558, 28)
```

```
[9]: service_dataset.isnull().sum()
```

```
[9]: Unique Key                          0
     Created Date                        0
     Closed Date                      2381
```

```
Agency                            0
Agency Name                       0
Complaint Type                    0
Descriptor                     6501
Location Type                   133
Incident Zip                   2998
Incident Address              51699
Street Name                   51699
Cross Street 1                57188
Cross Street 2                57805
Address Type                   3252
City                           2997
Facility Type                  2389
Status                            0
Due Date                          3
Resolution Description            0
Resolution Action Updated Date 2402
Community Board                   0
Borough                           0
X Coordinate (State Plane)     4030
Y Coordinate (State Plane)     4030
Park Borough                      0
Latitude                       4030
Longitude                      4030
Location                       4030
dtype: int64
```

[10]: `service_dataset`

[10]:
```
          Unique Key              Created Date              Closed Date Agency  \
0           32310363  12/31/2015 11:59:45 PM  01/01/2016 12:55:15 AM    NYPD
1           32309934  12/31/2015 11:59:44 PM  01/01/2016 01:26:57 AM    NYPD
2           32309159  12/31/2015 11:59:29 PM  01/01/2016 04:51:03 AM    NYPD
3           32305098  12/31/2015 11:57:46 PM  01/01/2016 07:43:13 AM    NYPD
4           32306529  12/31/2015 11:56:58 PM  01/01/2016 03:24:42 AM    NYPD
…                …                        …                        …     …
364553      29609918  01/01/2015 12:04:44 AM  01/01/2015 10:22:31 AM    NYPD
364554      29608392  01/01/2015 12:04:28 AM  01/01/2015 02:25:02 AM    NYPD
364555      29607589  01/01/2015 12:01:30 AM  01/01/2015 12:20:33 AM    NYPD
364556      29610889  01/01/2015 12:01:29 AM  01/01/2015 02:42:22 AM    NYPD
364557      29611816  01/01/2015 12:00:50 AM  01/01/2015 02:47:50 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
...                                 ...                          ...
364553  New York City Police Department              Illegal Parking
364554  New York City Police Department              Noise - Vehicle
364555  New York City Police Department  Noise - Street/Sidewalk
364556  New York City Police Department              Blocked Driveway
364557  New York City Police Department              Blocked Driveway

                          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
...                               ...              ...            ...
364553              Blocked Hydrant  Street/Sidewalk       11421.0
364554                Car/Truck Horn  Street/Sidewalk       10468.0
364555              Loud Music/Party  Street/Sidewalk       10031.0
364556                     No Access  Street/Sidewalk       10466.0
364557                     No Access  Street/Sidewalk       11420.0

            Incident Address  ...  \
0            71 VERMILYEA AVENUE  ...
1               27-07 23 AVENUE  ...
2          2897 VALENTINE AVENUE  ...
3           2940 BAISLEY AVENUE  ...
4                 87-14 57 ROAD  ...
...                        ...  ...
364553             84-25 85 ROAD  ...
364554   2555 SEDGWICK AVENUE  ...
364555     508 WEST 139 STREET  ...
364556     931 EAST 226 STREET  ...
364557       123-19 135 STREET  ...

                            Resolution Description  \
0       The Police Department responded and upon arriv…
1       The Police Department responded to the complai…
2       The Police Department responded and upon arriv…
3       The Police Department responded to the complai…
4       The Police Department responded and upon arriv…
...                                              ...
364553  The Police Department responded to the complai…
364554  The Police Department responded to the complai…
364555  The Police Department responded to the complai…
364556  The Police Department responded and upon arriv…
364557  The Police Department responded to the complai…
```

```
        Resolution Action Updated Date Community Board     Borough  \
0               01/01/2016 12:55:15 AM    12 MANHATTAN    MANHATTAN
1               01/01/2016 01:26:57 AM       01 QUEENS       QUEENS
2               01/01/2016 04:51:03 AM        07 BRONX        BRONX
3               01/01/2016 07:43:13 AM        10 BRONX        BRONX
4               01/01/2016 03:24:42 AM       04 QUEENS       QUEENS
...                                ...             ...          ...
364553          01/01/2015 10:22:31 AM       09 QUEENS       QUEENS
364554          01/01/2015 02:25:02 AM        07 BRONX        BRONX
364555          01/01/2015 12:20:33 AM    09 MANHATTAN    MANHATTAN
364556          01/01/2015 02:42:22 AM        12 BRONX        BRONX
364557          01/01/2015 02:47:50 AM       10 QUEENS       QUEENS

        X Coordinate (State Plane) Y Coordinate (State Plane) Park Borough  \
0                       1005409.0                   254678.0    MANHATTAN
1                       1007766.0                   221986.0       QUEENS
2                       1015081.0                   256380.0        BRONX
3                       1031740.0                   243899.0        BRONX
4                       1019123.0                   206375.0       QUEENS
...                           ...                        ...          ...
364553                  1022809.0                   192567.0       QUEENS
364554                  1009923.0                   255465.0        BRONX
364555                   997847.0                   238629.0    MANHATTAN
364556                  1024816.0                   262237.0        BRONX
364557                  1038733.0                   184971.0       QUEENS

         Latitude  Longitude                                      Location
0       40.865682 -73.923501   (40.86568153633767, -73.92350095571744)
1       40.775945 -73.915094   (40.775945312321085, -73.91509393898605)
2       40.870325 -73.888525   (40.870324522111424, -73.88852464418646)
3       40.835994 -73.828379   (40.83599404683083, -73.82837939584206)
4       40.733060 -73.874170   (40.733059618956815, -73.87416975810375)
...           ...        ...                                        ...
364553  40.695145 -73.860949   (40.69514470265117, -73.86094888534394)
364554  40.867830 -73.907178   (40.86782963689454, -73.90717786644662)
364555  40.821647 -73.950873   (40.821646626438095, -73.95087342885292)
364556  40.886361 -73.853290   (40.88636077906953, -73.85329048666742)
364557  40.674212 -73.803585   (40.674211762243935, -73.80358548685278)

[364558 rows x 28 columns]
```

```
[11]: service_dataset.dropna(inplace=True)
```

```
[12]: service_dataset.isnull().sum()
```

```
[12]: 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 Borough                      0
Latitude                          0
Longitude                         0
Location                          0
dtype: int64
```

2.0.2 2) 2) Analyze the date column and remove the entries if it has an incorrect timeline

```
[13]: service_dataset['Created Date']
```

```
[13]: 0          12/31/2015 11:59:45 PM
      1          12/31/2015 11:59:44 PM
      2          12/31/2015 11:59:29 PM
      3          12/31/2015 11:57:46 PM
      4          12/31/2015 11:56:58 PM
                          …
      364553     01/01/2015 12:04:44 AM
      364554     01/01/2015 12:04:28 AM
      364555     01/01/2015 12:01:30 AM
      364556     01/01/2015 12:01:29 AM
      364557     01/01/2015 12:00:50 AM
      Name: Created Date, Length: 300932, dtype: object
```

```
[14]: service_dataset['Created Date'] = pd.to_datetime(service_dataset['Created␣
      ↪Date'])
      service_dataset['Closed Date'] = pd.to_datetime(service_dataset['Closed Date'])
```

```
[15]: service_dataset.loc[service_dataset['Created Date']>=service_dataset['Closed␣
      ↪Date']].shape
```

```
[15]: (0, 28)
```

```
[16]: service_dataset = service_dataset[service_dataset['Closed Date'].notnull()]
```

2.0.3 2) 3) Draw a frequency plot for city-wise complaints

```
[17]: plt.figure(figsize=(10,10))
      service_dataset['City'].value_counts().plot(kind='barh',color='b')
      plt.xlabel('Count')
      plt.ylabel('City')
      plt.title('City-wise Complaints')
      plt.show()
```
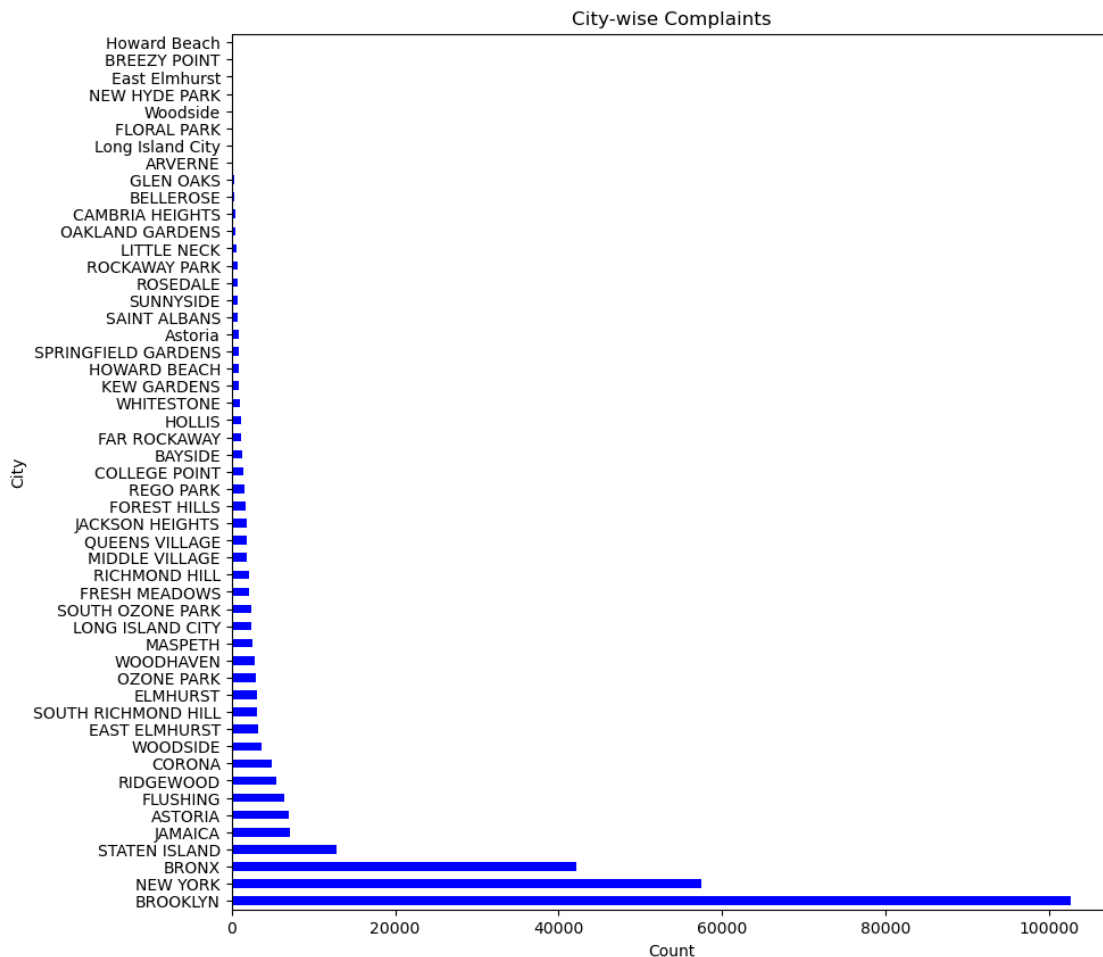
```
[18]: plt.figure(figsize=(10,10))
      sns.countplot(y=service_dataset['City'])
      plt.title('City-wise Complaints')
      plt.show()
```


City-wise Complaints
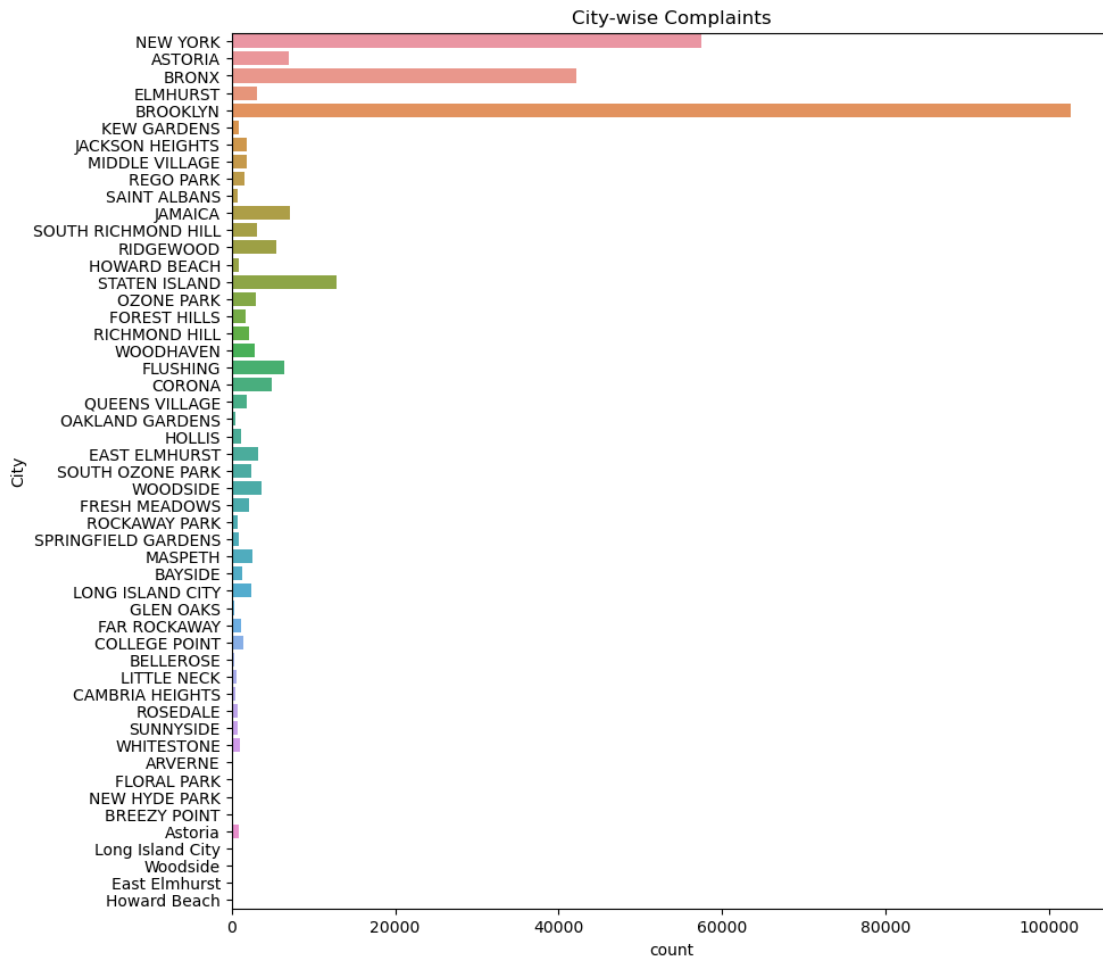
2.0.4 2) 4) Draw scatter and hexbin plots for complaint concentration across Brooklyn

```
[19]: def to_title(city):
          try:
              city = city.title()
              return city
          except:
              return np.nan
```

```
service_dataset['City'] = service_dataset['City'].apply(to_title)
service_dataset['City'].value_counts()
```

[19]:
```
Brooklyn                102568
New York                 57393
Bronx                    42117
Staten Island            12873
Astoria                   7927
Jamaica                   7080
Flushing                  6498
Ridgewood                 5456
Corona                    4946
Woodside                  3824
East Elmhurst             3213
South Richmond Hill       3156
Elmhurst                  3089
Ozone Park                3025
Woodhaven                 2819
Long Island City          2611
Maspeth                   2514
South Ozone Park          2346
Fresh Meadows             2128
Richmond Hill             2107
Middle Village            1909
Queens Village            1837
Jackson Heights           1810
Forest Hills              1668
Rego Park                 1537
College Point             1379
Bayside                   1254
Far Rockaway              1107
Hollis                    1099
Whitestone                1072
Kew Gardens                914
Howard Beach               907
Springfield Gardens        904
Saint Albans               806
Sunnyside                  783
Rosedale                   781
Rockaway Park              675
Little Neck                609
Oakland Gardens            512
Cambria Heights            500
Bellerose                  390
Glen Oaks                  275
Arverne                    208
Floral Park                168
```

```
New Hyde Park           109
Breezy Point             29
Name: City, dtype: int64
```

[20]: `service_dataset.groupby(['City','Complaint Type']).size()`

```
[20]: City        Complaint Type
      Arverne     Animal Abuse            45
                  Blocked Driveway        49
                  Derelict Vehicle        25
                  Disorderly Youth         1
                  Graffiti                 1
                                          …
      Woodside    Noise - Park            11
                  Noise - Street/Sidewalk 236
                  Noise - Vehicle         108
                  Traffic                  12
                  Vending                   7
      Length: 567, dtype: int64
```

[21]: `brooklyn = service_dataset.loc[service_dataset['City']=='Brooklyn']`

[22]:
```
brooklyn[['Longitude', 'Latitude']].plot(kind = 'hexbin', x='Longitude',
 ↪y='Latitude', gridsize=40,
    colormap = 'jet', mincnt=1, title = 'Complaints Concentration in Brooklyn',
 ↪figsize = (10, 8));
```

Complaints Concentration in Brooklyn
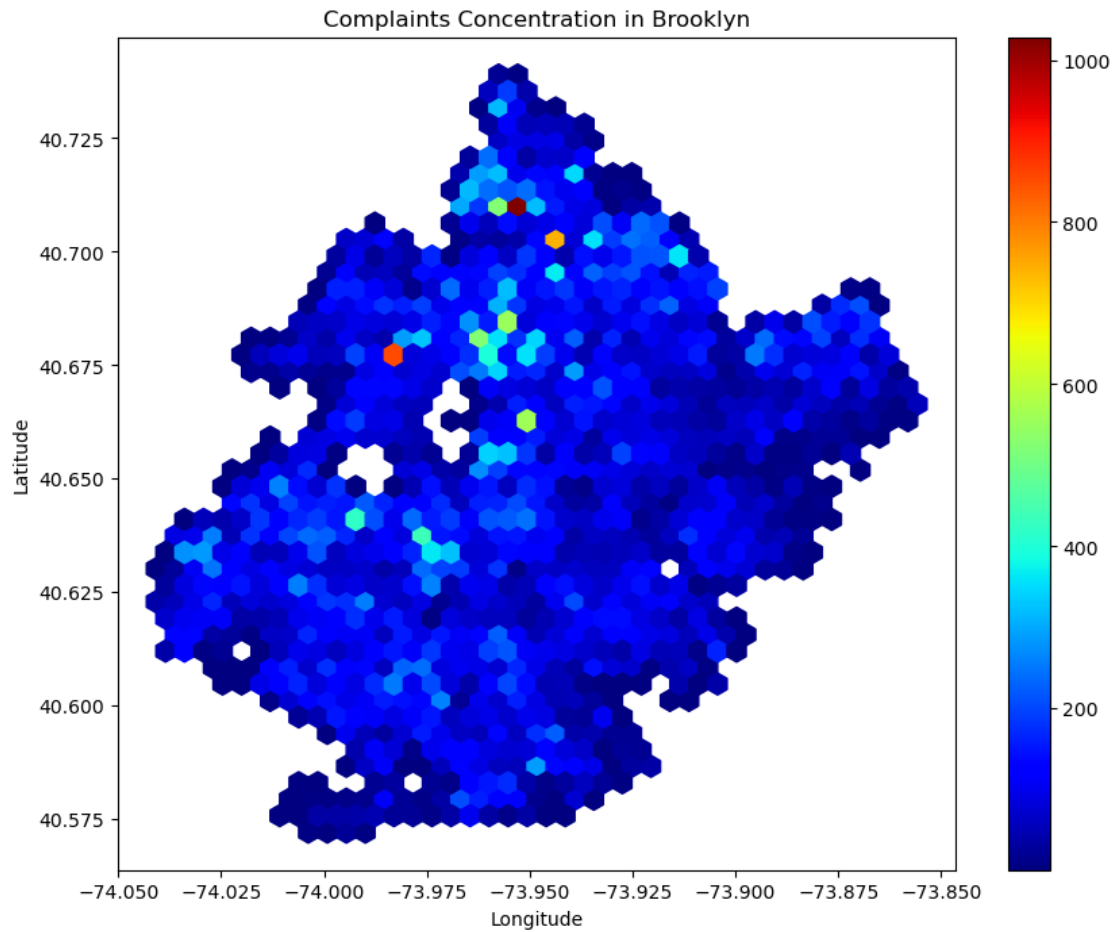
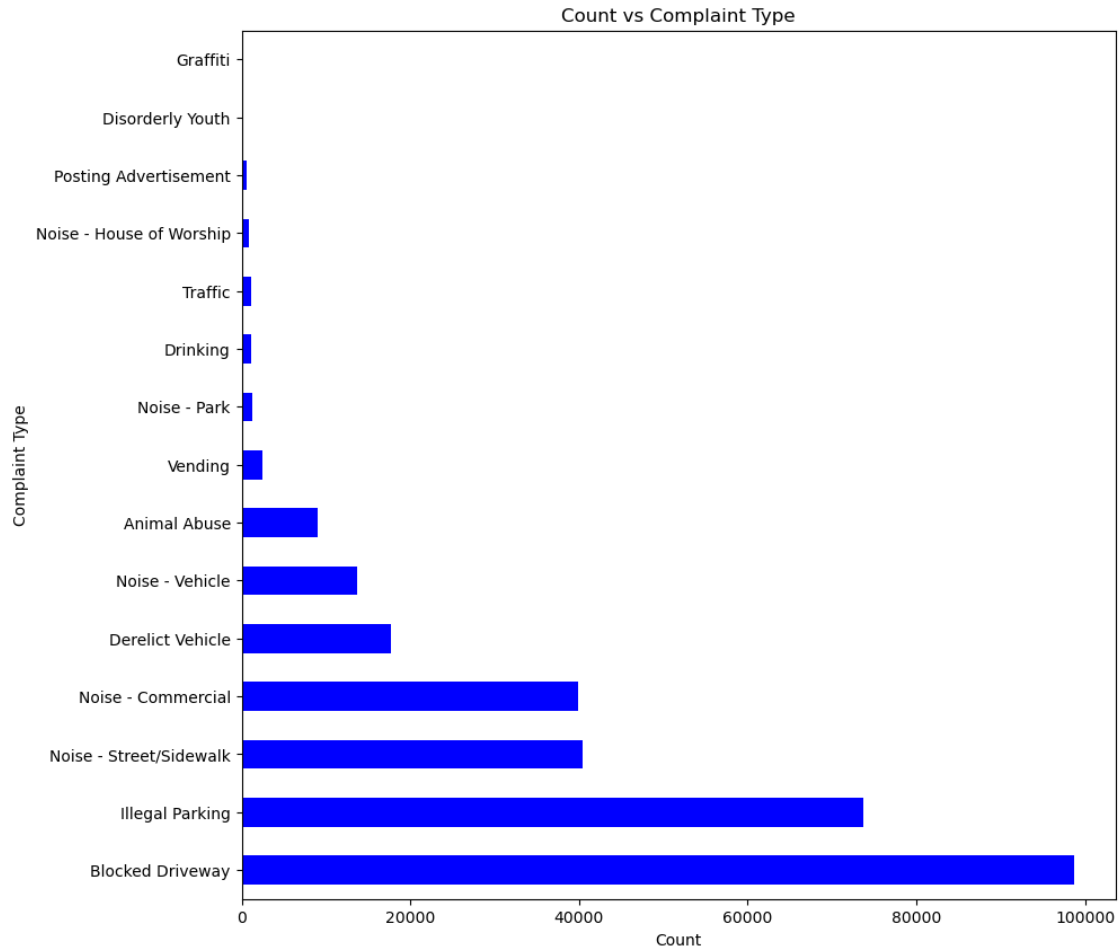# 3 3. Find major types of complaints:

1. Plot a bar graph of count vs. complaint types

2. Find the top 10 types of complaints

3. Display the types of complaints in each city in a separate dataset

3.0.1 3) 1) Plot a bar graph of count vs. complaint types

```
[23]: plt.figure(figsize=(10,10))
      sns.countplot(y=service_dataset['Complaint Type'])
      plt.title('Count vs. Complaint types')
      plt.show()
```

Count vs. Complaint types
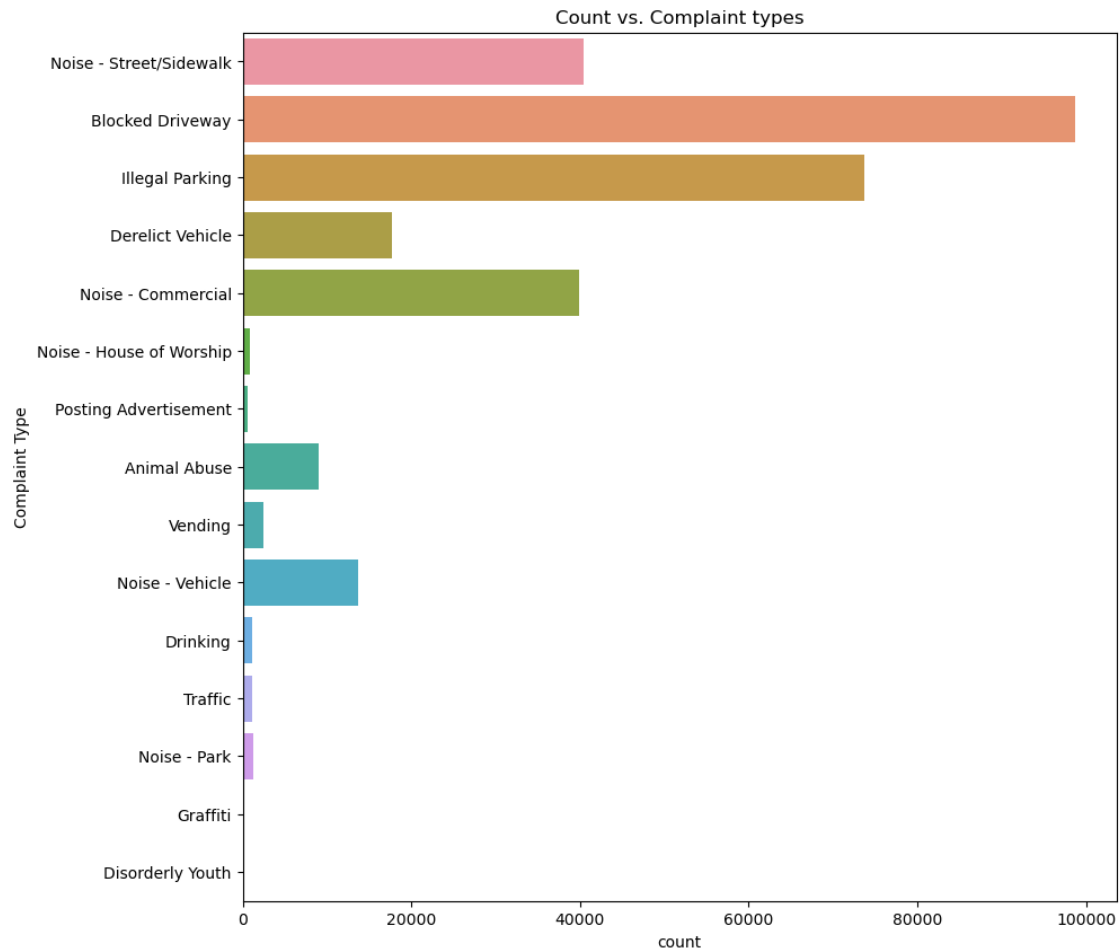
```
[24]: plt.figure(figsize=(10,10))
      service_dataset['Complaint Type'].value_counts().plot(kind='barh',color='b')
      plt.xlabel('Count')
      plt.ylabel('Complaint Type')
      plt.title('Count vs Complaint Type')
      plt.show()
```

Count vs Complaint Type

3.0.2 3) 2) Find the top 10 types of complaints

```
[25]: dfTop10=service_dataset.groupby('Complaint Type').size().
      ↪sort_values(ascending=False)[:10]
```

```
[26]: dfTop10
```

```
[26]: Complaint Type
      Blocked Driveway          98668
      Illegal Parking           73719
      Noise - Street/Sidewalk   40444
      Noise - Commercial        39919
      Derelict Vehicle          17706
      Noise - Vehicle           13763
      Animal Abuse               8968
      Vending                    2451
      Noise - Park               1289
```

```
Drinking                                  1124
dtype: int64
```

3.0.3 3) 3) Display the types of complaints in each city in a separate dataset

```
[27]: service_dataset.columns
```

```
[27]: Index(['Unique Key', 'Created Date', 'Closed Date', 'Agency', 'Agency Name',
             'Complaint Type', 'Descriptor', 'Location Type', 'Incident Zip',
             'Incident Address', 'Street Name', 'Cross Street 1', 'Cross Street 2',
             'Address Type', 'City', 'Facility Type', 'Status', 'Due Date',
             'Resolution Description', 'Resolution Action Updated Date',
             'Community Board', 'Borough', 'X Coordinate (State Plane)',
             'Y Coordinate (State Plane)', 'Park Borough', 'Latitude', 'Longitude',
             'Location'],
            dtype='object')
```

```
[29]: dfcomplaint = service_dataset.drop(['Unique Key', 'Created Date', 'Closed␣
      ↪Date', 'Agency', 'Agency Name',
             'Descriptor', 'Location Type', 'Incident Zip',
             'Incident Address', 'Street Name', 'Cross Street 1', 'Cross Street 2',
             'Address Type', 'Facility Type', 'Status', 'Due Date',
             'Resolution Description', 'Resolution Action Updated Date',
             'Community Board', 'Borough', 'X Coordinate (State Plane)',
             'Y Coordinate (State Plane)', 'Park Borough', 'Latitude', 'Longitude',
             'Location'],axis=1)
```

```
[30]: dfcomplaint
```

```
[30]:               Complaint Type              City
      0        Noise - Street/Sidewalk      New York
      1               Blocked Driveway       Astoria
      2               Blocked Driveway         Bronx
      3                Illegal Parking         Bronx
      4                Illegal Parking      Elmhurst
      …                            …             …
      364553           Illegal Parking     Woodhaven
      364554             Noise - Vehicle        Bronx
      364555  Noise - Street/Sidewalk      New York
      364556           Blocked Driveway         Bronx
      364557           Blocked Driveway  South Ozone Park

      [300932 rows x 2 columns]
```

```
[31]: dfcomplaint.to_csv('Types of Complaints')
```

4 4. Visualize the major types of complaints in each city

```
[34]: service_dataset.groupby(['City','Complaint Type']).size()
```

```
[34]: City       Complaint Type
      Arverne    Animal Abuse              45
                 Blocked Driveway          49
                 Derelict Vehicle          25
                 Disorderly Youth           1
                 Graffiti                   1
                                          ...
      Woodside   Noise - Park              11
                 Noise - Street/Sidewalk  236
                 Noise - Vehicle          108
                 Traffic                   12
                 Vending                    7
      Length: 567, dtype: int64
```

```
[38]: # pivot table
      crosstab=pd.
       ↪crosstab(index=service_dataset['City'],columns=service_dataset['Complaint␣
       ↪Type'])
      crosstab
```

```
[38]: Complaint Type    Animal Abuse  Blocked Driveway  Derelict Vehicle  \
      City
      Arverne                     45                49                25
      Astoria                    149              3531               359
      Bayside                     48               506               181
      Bellerose                   13               137                82
      Breezy Point                 2                 3                 3
      Bronx                     1742             16789              1960
      Brooklyn                  2845             35891              5342
      Cambria Heights             11               177               110
      College Point               31               587               198
      Corona                      98              3555                60
      East Elmhurst               79              1890               110
      Elmhurst                    41              1964                70
      Far Rockaway               101               377               151
      Floral Park                  7                33                57
      Flushing                   162              3558               441
      Forest Hills                61               850                57
      Fresh Meadows               60               664               288
      Glen Oaks                    3                46                52
      Hollis                      35               433               148
      Howard Beach                48               211               133
      Jackson Heights             45               688                35
```

Jamaica	263	3550	925
Kew Gardens	25	421	14
Little Neck	18	170	50
Long Island City	33	1064	185
Maspeth	53	983	399
Middle Village	30	655	308
New Hyde Park	1	74	13
New York	1355	2525	561
Oakland Gardens	26	171	84
Ozone Park	63	1652	416
Queens Village	77	769	383
Rego Park	31	760	73
Richmond Hill	49	1087	175
Ridgewood	138	2075	409
Rockaway Park	30	75	12
Rosedale	42	266	206
Saint Albans	34	306	197
South Ozone Park	70	1183	367
South Richmond Hill	35	1923	299
Springfield Gardens	36	324	221
Staten Island	733	2795	1774
Sunnyside	36	272	13
Whitestone	39	276	210
Woodhaven	50	1339	331
Woodside	75	2014	219

Complaint Type	Disorderly Youth	Drinking	Graffiti	Illegal Parking \
City				
Arverne	1	0	1	47
Astoria	4	33	4	1373
Bayside	1	1	3	453
Bellerose	2	1	0	98
Breezy Point	0	1	0	14
Bronx	50	168	13	7867
Brooklyn	60	235	54	28309
Cambria Heights	0	0	0	71
College Point	0	1	2	376
Corona	3	20	4	638
East Elmhurst	1	7	2	901
Elmhurst	1	11	1	661
Far Rockaway	1	3	0	252
Floral Park	1	1	0	63
Flushing	1	29	6	1801
Forest Hills	1	1	3	424
Fresh Meadows	0	2	0	1039
Glen Oaks	0	0	0	63
Hollis	0	3	0	142

Howard Beach	0	4	0	282
Jackson Heights	0	7	1	185
Jamaica	6	29	3	1238
Kew Gardens	0	1	0	228
Little Neck	2	1	0	276
Long Island City	1	6	3	794
Maspeth	0	5	1	874
Middle Village	0	2	0	838
New Hyde Park	0	0	0	15
New York	43	267	17	11235
Oakland Gardens	0	2	0	207
Ozone Park	3	10	0	625
Queens Village	0	3	0	460
Rego Park	0	2	1	501
Richmond Hill	0	8	1	390
Ridgewood	2	7	3	1796
Rockaway Park	1	23	0	267
Rosedale	0	2	2	186
Saint Albans	0	3	0	139
South Ozone Park	0	13	2	482
South Richmond Hill	2	21	0	507
Springfield Gardens	0	4	0	223
Staten Island	19	166	6	5096
Sunnyside	1	6	1	129
Whitestone	0	3	1	484
Woodhaven	0	4	0	761
Woodside	1	8	3	909

Complaint Type	Noise - Commercial	Noise - House of Worship \
City		
Arverne	2	13
Astoria	1788	17
Bayside	35	2
Bellerose	37	1
Breezy Point	4	0
Bronx	2611	61
Brooklyn	12769	257
Cambria Heights	13	2
College Point	24	2
Corona	259	1
East Elmhurst	37	24
Elmhurst	72	4
Far Rockaway	53	1
Floral Park	2	0
Flushing	204	2
Forest Hills	137	1
Fresh Meadows	20	0

Glen Oaks	78	0
Hollis	51	211
Howard Beach	209	0
Jackson Heights	597	2
Jamaica	485	11
Kew Gardens	198	1
Little Neck	69	0
Long Island City	252	0
Maspeth	51	1
Middle Village	11	0
New Hyde Park	4	0
New York	17195	162
Oakland Gardens	1	0
Ozone Park	109	4
Queens Village	43	0
Rego Park	70	0
Richmond Hill	246	0
Ridgewood	455	2
Rockaway Park	42	0
Rosedale	22	1
Saint Albans	29	0
South Ozone Park	64	4
South Richmond Hill	208	3
Springfield Gardens	31	0
Staten Island	669	10
Sunnyside	224	0
Whitestone	14	0
Woodhaven	206	1
Woodside	219	2

Complaint Type	Noise - Park	Noise - Street/Sidewalk	Noise - Vehicle \
City			
Arverne	0	18	6
Astoria	27	443	148
Bayside	1	5	16
Bellerose	0	9	7
Breezy Point	0	1	1
Bronx	214	7584	2679
Brooklyn	454	11200	4491
Cambria Heights	0	20	94
College Point	0	21	135
Corona	8	163	85
East Elmhurst	4	78	71
Elmhurst	3	182	55
Far Rockaway	7	97	56
Floral Park	0	3	1
Flushing	17	143	90

Forest Hills	6	72	39
Fresh Meadows	3	36	14
Glen Oaks	21	4	1
Hollis	2	28	42
Howard Beach	0	11	6
Jackson Heights	4	133	55
Jamaica	6	258	215
Kew Gardens	0	9	13
Little Neck	1	8	6
Long Island City	12	130	98
Maspeth	1	110	22
Middle Village	1	26	33
New Hyde Park	0	0	2
New York	425	17544	4142
Oakland Gardens	3	13	4
Ozone Park	4	79	54
Queens Village	1	46	44
Rego Park	18	41	34
Richmond Hill	0	84	48
Ridgewood	4	358	186
Rockaway Park	2	192	25
Rosedale	4	17	18
Saint Albans	0	52	37
South Ozone Park	1	69	76
South Richmond Hill	0	62	77
Springfield Gardens	0	29	32
Staten Island	20	704	292
Sunnyside	2	49	32
Whitestone	1	19	15
Woodhaven	1	58	58
Woodside	11	236	108

Complaint Type	Posting Advertisement	Traffic	Vending
City			
Arverne	0	0	1
Astoria	2	17	32
Bayside	0	2	0
Bellerose	0	3	0
Breezy Point	0	0	0
Bronx	13	85	281
Brooklyn	49	313	299
Cambria Heights	0	2	0
College Point	0	1	1
Corona	0	4	48
East Elmhurst	1	4	4
Elmhurst	1	4	19
Far Rockaway	0	4	4

```
Floral Park            0      0       0
Flushing               1     16      27
Forest Hills           4     10       2
Fresh Meadows          0      2       0
Glen Oaks              0      1       6
Hollis                 0      4       0
Howard Beach           0      2       1
Jackson Heights        1      2      55
Jamaica                8     75       8
Kew Gardens            1      2       1
Little Neck            1      7       0
Long Island City       2      8      23
Maspeth                0     10       4
Middle Village         0      5       0
New Hyde Park          0      0       0
New York              40    331    1551
Oakland Gardens        0      1       0
Ozone Park             1      4       1
Queens Village         1     10       0
Rego Park              0      4       2
Richmond Hill          1      4      14
Ridgewood              1     15       5
Rockaway Park          0      4       2
Rosedale               0      6       9
Saint Albans           0      8       1
South Ozone Park       0     14       1
South Richmond Hill    0      6      13
Springfield Gardens    0      3       1
Staten Island        515     56      18
Sunnyside              2     10       6
Whitestone             0     10       0
Woodhaven              0      6       4
Woodside               0     12       7
```

[39]: `crosstab.plot(kind='bar',figsize=(12,8),stacked=True,colormap='Paired')`

[39]: `<AxesSubplot:xlabel='City'>`
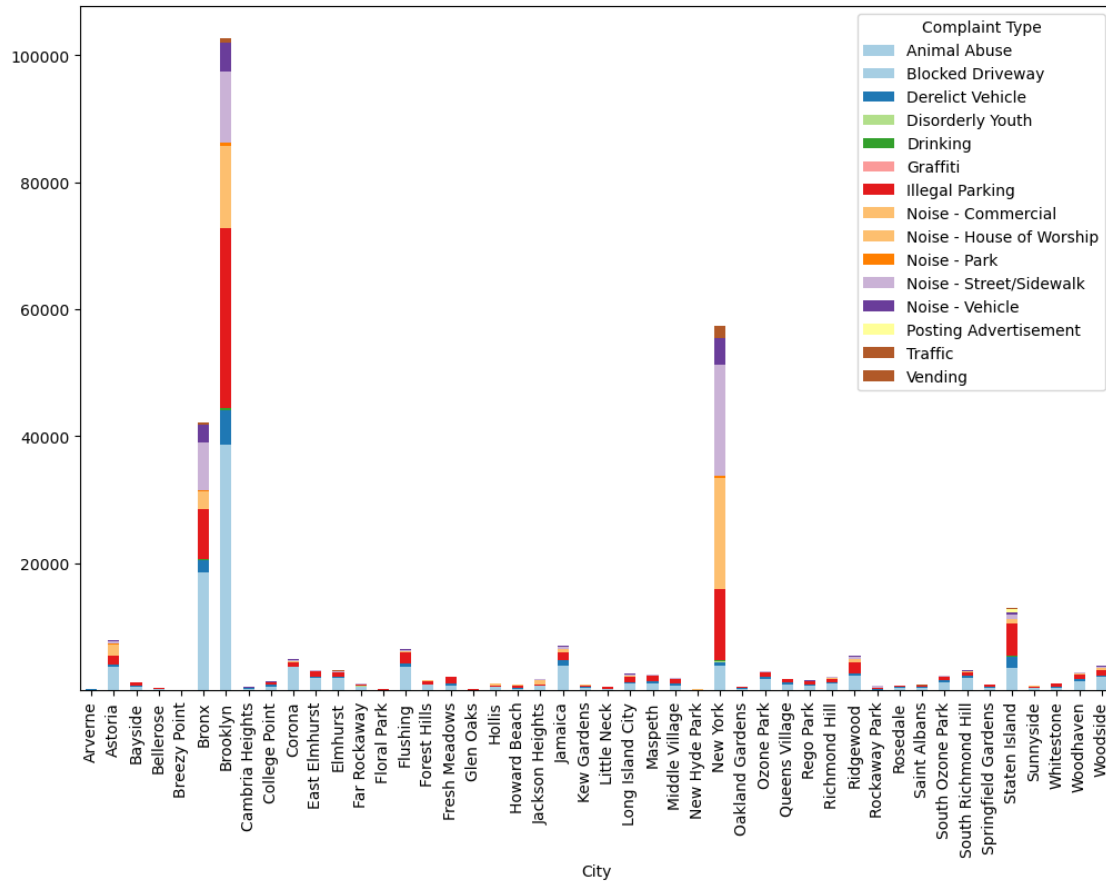
5 5. Check if the average response time across various types of complaints

```
[53]: service_dataset["Request_Closing_Time"]=(service_dataset["Closed␣
      ↪Date"]-service_dataset["Created Date"])

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

      service_dataset["Request_Closing_Time"]=Request_Closing_Time
```

```
[56]: pd.DataFrame(service_dataset.groupby("Location Type")["Request_Closing_Time"].
      ↪mean()).sort_values("Request_Closing_Time")
```

```
[56]:                          Request_Closing_Time
      Location Type
```

```
Highway                         118.261111
Club/Bar/Restaurant             183.121128
Store/Commercial                188.681364
Subway Station                  192.109524
House of Worship                195.050892
Park/Playground                 210.367193
Commercial                      229.288000
Street/Sidewalk                 263.398204
Residential Building            275.251621
House and Store                 293.174961
Residential Building/House      304.822417
Parking Lot                     309.087500
Vacant Lot                      405.950714
Roadway Tunnel                  838.352778
```

6   6. Identify significant variables by performing a statistical analysis using p-values and chi-square values

```python
[68]: from scipy import stats
      from scipy.stats import chi2_contingency

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

```python
[57]: service_dataset.shape
```

```
[57]: (300932, 29)
```

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

```python
[60]: 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']
Address Type            :  ['ADDRESS' 'BLOCKFACE' 'INTERSECTION']
Facility Type            :  ['Precinct']
Status             :  ['Closed']
```

```python
[61]: new_df.shape
```

```
[61]: (300932, 29)
```

```
[62]: 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)
```

```
[64]: new_df.head()
```

```
[64]:          Created Date         Closed Date Agency  \
      0 2015-12-31 23:59:45 2016-01-01 00:55:15   NYPD
      1 2015-12-31 23:59:44 2016-01-01 01:26:57   NYPD
      2 2015-12-31 23:59:29 2016-01-01 04:51:03   NYPD
      3 2015-12-31 23:57:46 2016-01-01 07:43:13   NYPD
      4 2015-12-31 23:56:58 2016-01-01 03:24:42   NYPD

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

         Incident Zip Address Type      City Facility Type  Status    Borough  \
      0       10034.0      ADDRESS  New York      Precinct  Closed  MANHATTAN
      1       11105.0      ADDRESS   Astoria      Precinct  Closed     QUEENS
      2       10458.0      ADDRESS     Bronx      Precinct  Closed      BRONX
      3       10461.0      ADDRESS     Bronx      Precinct  Closed      BRONX
      4       11373.0      ADDRESS  Elmhurst     Precinct  Closed     QUEENS

         Request_Closing_Time
      0             55.500000
      1             87.216667
      2            291.566667
      3            465.450000
      4            207.733333
```
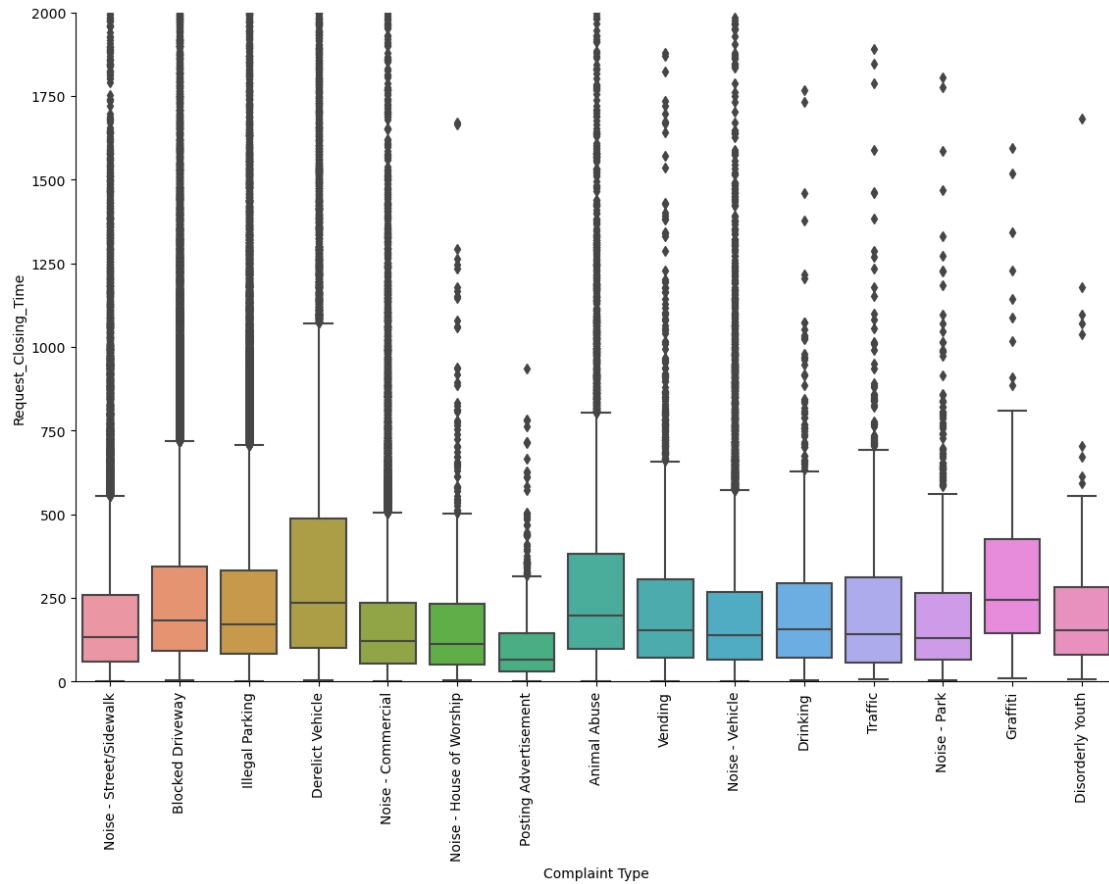
```
[65]: g=sns.catplot(x="Complaint␣
      ↪Type",y="Request_Closing_Time",kind="box",data=new_df)
      g.fig.set_figheight(8)
      g.fig.set_figwidth(15)
      plt.xticks(rotation=90)
      plt.ylim((0,2000))
```

```
[65]: (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

```
[66]: 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()
```

```
[66]:    Request_Closing_Time              Complaint
      0             55.500000  Noise - Street/Sidewalk
      1             87.216667          Blocked Driveway
      2            291.566667          Blocked Driveway
      3            465.450000           Illegal Parking
      4            207.733333           Illegal Parking
```

```
[69]: lm=ols("Request_Closing_Time~Complaint",data=anova_df).fit()
      table=sm.stats.anova_lm(lm)
      table
```

```
[69]:                df        sum_sq        mean_sq           F  PR(>F)
      Complaint    14.0  8.501602e+08  6.072573e+07  525.961595     0.0
      Residual  300917.0  3.474285e+10  1.154566e+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

```
[70]: 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)
```

```
[71]: data_crosstab = pd.crosstab( chi_sq["Location Type"],chi_sq["Complaint Type"])
```

```
[72]: 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

```
[ ]:
```