311 customer service

April 11, 2023

1 NYC Service Requests

1.0.1 You've been asked to perform data analysis of service request (311) calls from New York City. You've also been asked to utilize data wrangling techniques to understand the pattern in the data and visualize the major types of complaints.

[1]:
```
# import dependencies
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
from matplotlib.pyplot import figure
import scipy.stats as stats
```

[2]:
```
df=pd.read_csv("311_Service_Requests_from_2010_to_Present.csv",␣
↪low_memory=False)
```

[3]:
```
# first 5 rows to visualize
df.head()
```

[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

                        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

                    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

        Incident Address  … Bridge Highway Name Bridge Highway Direction  \
0    71 VERMILYEA AVENUE  …                  NaN                       NaN
1        27-07 23 AVENUE  …                  NaN                       NaN
2  2897 VALENTINE AVENUE  …                  NaN                       NaN
3    2940 BAISLEY AVENUE  …                  NaN                       NaN
4          87-14 57 ROAD  …                  NaN                       NaN

  Road Ramp Bridge Highway Segment Garage Lot Name Ferry Direction  \
0       NaN                     NaN             NaN             NaN
1       NaN                     NaN             NaN             NaN
2       NaN                     NaN             NaN             NaN
3       NaN                     NaN             NaN             NaN
4       NaN                     NaN             NaN             NaN

  Ferry Terminal Name   Latitude  Longitude  \
0                 NaN  40.865682 -73.923501
1                 NaN  40.775945 -73.915094
2                 NaN  40.870325 -73.888525
3                 NaN  40.835994 -73.828379
4                 NaN  40.733060 -73.874170

                                Location
0   (40.86568153633767, -73.92350095571744)
1  (40.775945312321085, -73.91509393898605)
2  (40.870324522111424, -73.88852464418646)
3   (40.83599404683083, -73.82837939584206)
4  (40.733059618956815, -73.87416975810375)

[5 rows x 53 columns]
```

[4]:
```
# reload in the data and change date columns to datetime
df=pd.read_csv("311_Service_Requests_from_2010_to_Present.csv", parse_dates =␣
↪['Created Date','Closed Date','Due Date'], infer_datetime_format = True,␣
↪low_memory=False)
```

[5]:
```
# checking new date columns
df.head()
```

[5]:
```
   Unique Key        Created Date         Closed Date Agency  \
0    32310363 2015-12-31 23:59:45 2016-01-01 00:55:15   NYPD
1    32309934 2015-12-31 23:59:44 2016-01-01 01:26:57   NYPD
2    32309159 2015-12-31 23:59:29 2016-01-01 04:51:03   NYPD
```

```
3    32305098 2015-12-31 23:57:46 2016-01-01 07:43:13   NYPD
4    32306529 2015-12-31 23:56:58 2016-01-01 03:24:42   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

                    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

        Incident Address  … Bridge Highway Name Bridge Highway Direction  \
0    71 VERMILYEA AVENUE  …                 NaN                       NaN
1        27-07 23 AVENUE  …                 NaN                       NaN
2  2897 VALENTINE AVENUE  …                 NaN                       NaN
3    2940 BAISLEY AVENUE  …                 NaN                       NaN
4          87-14 57 ROAD  …                 NaN                       NaN

  Road Ramp Bridge Highway Segment Garage Lot Name Ferry Direction  \
0       NaN                    NaN             NaN             NaN
1       NaN                    NaN             NaN             NaN
2       NaN                    NaN             NaN             NaN
3       NaN                    NaN             NaN             NaN
4       NaN                    NaN             NaN             NaN

  Ferry Terminal Name   Latitude  Longitude  \
0                 NaN  40.865682 -73.923501
1                 NaN  40.775945 -73.915094
2                 NaN  40.870325 -73.888525
3                 NaN  40.835994 -73.828379
4                 NaN  40.733060 -73.874170

                                Location
0   (40.86568153633767, -73.92350095571744)
1  (40.775945312321085, -73.91509393898605)
2  (40.870324522111424, -73.88852464418646)
3   (40.83599404683083, -73.82837939584206)
4  (40.733059618956815, -73.87416975810375)

[5 rows x 53 columns]
```

[6]:
```
#info and column names
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 364558 entries, 0 to 364557
Data columns (total 53 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   Unique Key                      364558 non-null  int64
 1   Created Date                    364558 non-null  datetime64[ns]
 2   Closed Date                     362177 non-null  datetime64[ns]
 3   Agency                          364558 non-null  object
 4   Agency Name                     364558 non-null  object
 5   Complaint Type                  364558 non-null  object
 6   Descriptor                      358057 non-null  object
 7   Location Type                   364425 non-null  object
 8   Incident Zip                    361560 non-null  float64
 9   Incident Address                312859 non-null  object
 10  Street Name                     312859 non-null  object
 11  Cross Street 1                  307370 non-null  object
 12  Cross Street 2                  306753 non-null  object
 13  Intersection Street 1           51120 non-null   object
 14  Intersection Street 2           50512 non-null   object
 15  Address Type                    361306 non-null  object
 16  City                            361561 non-null  object
 17  Landmark                        375 non-null     object
 18  Facility Type                   362169 non-null  object
 19  Status                          364558 non-null  object
 20  Due Date                        364555 non-null  datetime64[ns]
 21  Resolution Description          364558 non-null  object
 22  Resolution Action Updated Date  362156 non-null  object
 23  Community Board                 364558 non-null  object
 24  Borough                         364558 non-null  object
 25  X Coordinate (State Plane)      360528 non-null  float64
 26  Y Coordinate (State Plane)      360528 non-null  float64
 27  Park Facility Name              364558 non-null  object
 28  Park Borough                    364558 non-null  object
 29  School Name                     364558 non-null  object
 30  School Number                   364558 non-null  object
 31  School Region                   364557 non-null  object
 32  School Code                     364557 non-null  object
 33  School Phone Number             364558 non-null  object
 34  School Address                  364558 non-null  object
 35  School City                     364558 non-null  object
 36  School State                    364558 non-null  object
 37  School Zip                      364557 non-null  object
 38  School Not Found                364558 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             297 non-null     object
 44  Bridge Highway Direction        297 non-null     object
 45  Road Ramp                       262 non-null     object
 46  Bridge Highway Segment          262 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                        360528 non-null  float64
 51  Longitude                       360528 non-null  float64
 52  Location                        360528 non-null  object
dtypes: datetime64[ns](3), float64(10), int64(1), object(39)
memory usage: 147.4+ MB
```

[7]:
```
# size/shape of data
df.shape
```

[7]:
```
(364558, 53)
```

1.1 Missing value treatment

[8]:
```
# checking missing/null values
df.isnull().sum().sort_values(ascending=False)
```

[8]:
```
Taxi Company Borough             364558
Taxi Pick Up Location            364558
School or Citywide Complaint     364558
Garage Lot Name                  364558
Vehicle Type                     364558
Ferry Direction                  364557
Ferry Terminal Name              364556
Bridge Highway Segment           364296
Road Ramp                        364296
Bridge Highway Name              364261
Bridge Highway Direction         364261
Landmark                         364183
Intersection Street 2            314046
Intersection Street 1            313438
Cross Street 2                    57805
Cross Street 1                    57188
Incident Address                  51699
Street Name                       51699
Descriptor                         6501
```

```
Location                          4030
Longitude                         4030
X Coordinate (State Plane)        4030
Y Coordinate (State Plane)        4030
Latitude                          4030
Address Type                      3252
Incident Zip                      2998
City                              2997
Resolution Action Updated Date    2402
Facility Type                     2389
Closed Date                       2381
Location Type                      133
Due Date                             3
School Code                          1
School Region                        1
School Zip                           1
School Name                          0
Complaint Type                       0
Agency Name                          0
Agency                               0
Created Date                         0
School Not Found                     0
School Phone Number                  0
School State                         0
School City                          0
School Address                       0
Park Borough                         0
Status                               0
Resolution Description               0
Community Board                      0
Borough                              0
School Number                        0
Park Facility Name                   0
Unique Key                           0
dtype: int64
```

[9]:
```
# % of missing values
df.isnull().sum().sort_values(ascending=False)*100/364558
```

[9]:
```
Taxi Company Borough            100.000000
Taxi Pick Up Location           100.000000
School or Citywide Complaint    100.000000
Garage Lot Name                 100.000000
Vehicle Type                    100.000000
Ferry Direction                  99.999726
Ferry Terminal Name              99.999451
Bridge Highway Segment           99.928132
```

```
Road Ramp                         99.928132
Bridge Highway Name               99.918531
Bridge Highway Direction          99.918531
Landmark                          99.897136
Intersection Street 2             86.144317
Intersection Street 1             85.977540
Cross Street 2                    15.856187
Cross Street 1                    15.686941
Incident Address                  14.181283
Street Name                       14.181283
Descriptor                         1.783255
Location                           1.105448
Longitude                          1.105448
X Coordinate (State Plane)         1.105448
Y Coordinate (State Plane)         1.105448
Latitude                           1.105448
Address Type                       0.892039
Incident Zip                       0.822366
City                               0.822091
Resolution Action Updated Date     0.658880
Facility Type                      0.655314
Closed Date                        0.653120
Location Type                      0.036483
Due Date                           0.000823
School Code                        0.000274
School Region                      0.000274
School Zip                         0.000274
School Name                        0.000000
Complaint Type                     0.000000
Agency Name                        0.000000
Agency                             0.000000
Created Date                       0.000000
School Not Found                   0.000000
School Phone Number                0.000000
School State                       0.000000
School City                        0.000000
School Address                     0.000000
Park Borough                       0.000000
Status                             0.000000
Resolution Description             0.000000
Community Board                    0.000000
Borough                            0.000000
School Number                      0.000000
Park Facility Name                 0.000000
Unique Key                         0.000000
dtype: float64
```

[10]:
```
df2=pd.DataFrame(df.columns.to_list()).set_index(0)
df2
```

[10]:
```
Empty DataFrame
Columns: []
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]
```

[11]:
```
df2=df2[df.isnull().sum()*100/364558 < 80 ].reset_index()
df2
```

[11]:

	0
0	Unique Key
1	Created Date
2	Closed Date
3	Agency
4	Agency Name
5	Complaint Type
6	Descriptor
7	Location Type
8	Incident Zip
9	Incident Address
10	Street Name
11	Cross Street 1
12	Cross Street 2
13	Address Type
14	City
15	Facility Type
16	Status
17	Due Date
18	Resolution Description
19	Resolution Action Updated Date
20	Community Board
21	Borough
22	X Coordinate (State Plane)
23	Y Coordinate (State Plane)
24	Park Facility Name

```
25                    Park Borough
26                     School Name
27                   School Number
28                   School Region
29                     School Code
30             School Phone Number
31                  School Address
32                     School City
33                   School State
34                     School Zip
35                School Not Found
36                        Latitude
37                       Longitude
38                        Location
```

[12]:
```
# school has a lot of unspecified values
df3=df[df2[0].to_list()]
df3
```

[12]:
```
        Unique Key        Created Date         Closed Date Agency  \
0         32310363 2015-12-31 23:59:45 2016-01-01 00:55:15   NYPD
1         32309934 2015-12-31 23:59:44 2016-01-01 01:26:57   NYPD
2         32309159 2015-12-31 23:59:29 2016-01-01 04:51:03   NYPD
3         32305098 2015-12-31 23:57:46 2016-01-01 07:43:13   NYPD
4         32306529 2015-12-31 23:56:58 2016-01-01 03:24:42   NYPD
…              …                   …                   …    …
364553    29609918 2015-01-01 00:04:44 2015-01-01 10:22:31   NYPD
364554    29608392 2015-01-01 00:04:28 2015-01-01 02:25:02   NYPD
364555    29607589 2015-01-01 00:01:30 2015-01-01 00:20:33   NYPD
364556    29610889 2015-01-01 00:01:29 2015-01-01 02:42:22   NYPD
364557    29611816 2015-01-01 00:00:50 2015-01-01 02:47:50   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  …  School Code School Phone Number  \
0          71 VERMILYEA AVENUE  …  Unspecified         Unspecified
1              27-07 23 AVENUE  …  Unspecified         Unspecified
2        2897 VALENTINE AVENUE  …  Unspecified         Unspecified
3          2940 BAISLEY AVENUE  …  Unspecified         Unspecified
4                87-14 57 ROAD  …  Unspecified         Unspecified
…                          …  …          …                   …
364553           84-25 85 ROAD  …  Unspecified         Unspecified
364554    2555 SEDGWICK AVENUE  …  Unspecified         Unspecified
364555     508 WEST 139 STREET  …  Unspecified         Unspecified
364556     931 EAST 226 STREET  …  Unspecified         Unspecified
364557       123-19 135 STREET  …  Unspecified         Unspecified

        School Address  School City School State   School Zip School Not Found  \
0          Unspecified  Unspecified  Unspecified  Unspecified                N
1          Unspecified  Unspecified  Unspecified  Unspecified                N
2          Unspecified  Unspecified  Unspecified  Unspecified                N
3          Unspecified  Unspecified  Unspecified  Unspecified                N
4          Unspecified  Unspecified  Unspecified  Unspecified                N
…                    …            …            …            …                …
364553     Unspecified  Unspecified  Unspecified  Unspecified                N
364554     Unspecified  Unspecified  Unspecified  Unspecified                N
364555     Unspecified  Unspecified  Unspecified  Unspecified                N
364556     Unspecified  Unspecified  Unspecified  Unspecified                N
364557     Unspecified  Unspecified  Unspecified  Unspecified                N

          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 39 columns]
```

[13]:
```
#confirming that df3 has dropped all the categories with more than 80% missing values
df3.isnull().sum().sort_values(ascending=False)*100/364558
```

[13]:
```
Cross Street 2                    15.856187
Cross Street 1                    15.686941
Incident Address                  14.181283
Street Name                       14.181283
Descriptor                         1.783255
Location                           1.105448
Y Coordinate (State Plane)         1.105448
Latitude                           1.105448
Longitude                          1.105448
X Coordinate (State Plane)         1.105448
Address Type                       0.892039
Incident Zip                       0.822366
City                               0.822091
Resolution Action Updated Date     0.658880
Facility Type                      0.655314
Closed Date                        0.653120
Location Type                      0.036483
Due Date                           0.000823
School Region                      0.000274
School Code                        0.000274
School Zip                         0.000274
Park Facility Name                 0.000000
School State                       0.000000
Created Date                       0.000000
Agency                             0.000000
Agency Name                        0.000000
Complaint Type                     0.000000
School Not Found                   0.000000
School Phone Number                0.000000
School City                        0.000000
School Address                     0.000000
Park Borough                       0.000000
Status                             0.000000
Resolution Description             0.000000
School Number                      0.000000
Community Board                    0.000000
Borough                            0.000000
School Name                        0.000000
```

```
Unique Key                          0.000000
dtype: float64
```

[14]:
```
df3.shape
```

[14]:
```
(364558, 39)
```

[15]:
```
#columns went from 53 to 39 after removing nulls
df3.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 364558 entries, 0 to 364557
Data columns (total 39 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   Unique Key                      364558 non-null  int64
 1   Created Date                    364558 non-null  datetime64[ns]
 2   Closed Date                     362177 non-null  datetime64[ns]
 3   Agency                          364558 non-null  object
 4   Agency Name                     364558 non-null  object
 5   Complaint Type                  364558 non-null  object
 6   Descriptor                      358057 non-null  object
 7   Location Type                   364425 non-null  object
 8   Incident Zip                    361560 non-null  float64
 9   Incident Address                312859 non-null  object
 10  Street Name                     312859 non-null  object
 11  Cross Street 1                  307370 non-null  object
 12  Cross Street 2                  306753 non-null  object
 13  Address Type                    361306 non-null  object
 14  City                            361561 non-null  object
 15  Facility Type                   362169 non-null  object
 16  Status                          364558 non-null  object
 17  Due Date                        364555 non-null  datetime64[ns]
 18  Resolution Description          364558 non-null  object
 19  Resolution Action Updated Date  362156 non-null  object
 20  Community Board                 364558 non-null  object
 21  Borough                         364558 non-null  object
 22  X Coordinate (State Plane)      360528 non-null  float64
 23  Y Coordinate (State Plane)      360528 non-null  float64
 24  Park Facility Name              364558 non-null  object
 25  Park Borough                    364558 non-null  object
 26  School Name                     364558 non-null  object
 27  School Number                   364558 non-null  object
 28  School Region                   364557 non-null  object
 29  School Code                     364557 non-null  object
 30  School Phone Number             364558 non-null  object
 31  School Address                  364558 non-null  object
```

```
 32  School City                     364558 non-null  object
 33  School State                    364558 non-null  object
 34  School Zip                      364557 non-null  object
 35  School Not Found                364558 non-null  object
 36  Latitude                        360528 non-null  float64
 37  Longitude                       360528 non-null  float64
 38  Location                        360528 non-null  object
dtypes: datetime64[ns](3), float64(5), int64(1), object(30)
memory usage: 108.5+ MB
```

[16]:
```
# all of school code are unspecified
df3['School Code'].value_counts()
```

[16]:
```
Unspecified    364557
Name: School Code, dtype: int64
```

[17]:
```
# all but one of school name are unspecified
df3['School Name'].value_counts()
```

[17]:
```
Unspecified                          364557
Alley Pond Park - Nature Center           1
Name: School Name, dtype: int64
```

[18]:
```
df3[df3['School Name'] != 'Unspecified']
```

[18]:
```
         Unique Key        Created Date         Closed Date Agency  \
283132    30427220 2015-04-18 09:44:55 2015-05-02 10:35:29   NYPD

                            Agency Name    Complaint Type    Descriptor  \
283132  New York City Police Department  Animal in a Park  Animal Waste

       Location Type  Incident Zip Incident Address  … School Code  \
283132          Park           NaN              NaN  …         NaN

       School Phone Number                                School Address  \
283132          7182176034  Grand Central Parkway, near the soccer field

       School City School State School Zip School Not Found Latitude  \
283132      QUEENS           NY        NaN                N      NaN

       Longitude Location
283132       NaN      NaN

[1 rows x 39 columns]
```

[19]:
```
# remove all school related columns if they have 'unspecified'
```

```
removed_school_list = [col for col in df3.columns.to_list() if col.
 ↪find('School')<0]
removed_school_list
```

[19]:
```
['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 Facility Name',
 'Park Borough',
 'Latitude',
 'Longitude',
 'Location']
```

[20]:
```
# made new dataset without any 'School' related columns
pd.set_option('display.max_columns', None)
df4=df3[removed_school_list]
df4.head(10)
```

[20]:

	Unique Key	Created Date	Closed Date	Agency	\
0	32310363	2015-12-31 23:59:45	2016-01-01 00:55:15	NYPD	
1	32309934	2015-12-31 23:59:44	2016-01-01 01:26:57	NYPD	
2	32309159	2015-12-31 23:59:29	2016-01-01 04:51:03	NYPD	
3	32305098	2015-12-31 23:57:46	2016-01-01 07:43:13	NYPD	
4	32306529	2015-12-31 23:56:58	2016-01-01 03:24:42	NYPD	
5	32306554	2015-12-31 23:56:30	2016-01-01 01:50:11	NYPD	
6	32306559	2015-12-31 23:55:32	2016-01-01 01:53:54	NYPD	

```
7    32307009 2015-12-31 23:54:05 2016-01-01 01:42:54   NYPD
8    32308581 2015-12-31 23:53:58 2016-01-01 08:27:32   NYPD
9    32308391 2015-12-31 23:53:58 2016-01-01 01:17:40   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
5  New York City Police Department          Illegal Parking
6  New York City Police Department          Illegal Parking
7  New York City Police Department         Blocked Driveway
8  New York City Police Department          Illegal Parking
9  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
5  Posted Parking Sign Violation  Street/Sidewalk       11215.0
6               Blocked Hydrant  Street/Sidewalk       10032.0
7                     No Access  Street/Sidewalk       10457.0
8  Posted Parking Sign Violation  Street/Sidewalk       11415.0
9                     No Access  Street/Sidewalk       11219.0

          Incident Address         Street Name     Cross Street 1  \
0      71 VERMILYEA AVENUE    VERMILYEA AVENUE     ACADEMY STREET
1          27-07 23 AVENUE           23 AVENUE          27 STREET
2    2897 VALENTINE AVENUE    VALENTINE AVENUE    EAST 198 STREET
3      2940 BAISLEY AVENUE      BAISLEY AVENUE      EDISON AVENUE
4            87-14 57 ROAD             57 ROAD     SEABURY STREET
5            260 21 STREET           21 STREET           5 AVENUE
6      524 WEST 169 STREET     WEST 169 STREET   AMSTERDAM AVENUE
7      501 EAST 171 STREET     EAST 171 STREET  WASHINGTON AVENUE
8  83-44 LEFFERTS BOULEVARD  LEFFERTS BOULEVARD               BEND
9           1408 66 STREET           66 STREET          14 AVENUE

      Cross Street 2 Address Type         City Facility Type  Status  \
0    WEST 204 STREET      ADDRESS     NEW YORK      Precinct  Closed
1          28 STREET      ADDRESS      ASTORIA      Precinct  Closed
2    EAST 199 STREET      ADDRESS        BRONX      Precinct  Closed
3          B STREET       ADDRESS        BRONX      Precinct  Closed
4      HOFFMAN DRIVE      ADDRESS     ELMHURST      Precinct  Closed
5          6 AVENUE       ADDRESS     BROOKLYN      Precinct  Closed
```

```
6      AUDUBON AVENUE      ADDRESS     NEW YORK      Precinct  Closed
7            3 AVENUE      ADDRESS        BRONX      Precinct  Closed
8               BEND       ADDRESS  KEW GARDENS      Precinct  Closed
9  NEW UTRECHT AVENUE      ADDRESS     BROOKLYN      Precinct  Closed

             Due Date                                Resolution Description  \
0 2016-01-01 07:59:45  The Police Department responded and upon arriv…
1 2016-01-01 07:59:44  The Police Department responded to the complai…
2 2016-01-01 07:59:29  The Police Department responded and upon arriv…
3 2016-01-01 07:57:46  The Police Department responded to the complai…
4 2016-01-01 07:56:58  The Police Department responded and upon arriv…
5 2016-01-01 07:56:30  The Police Department responded and upon arriv…
6 2016-01-01 07:55:32  The Police Department issued a summons in resp…
7 2016-01-01 07:54:05  The Police Department responded to the complai…
8 2016-01-01 07:53:58  The Police Department responded to the complai…
9 2016-01-01 07:53:58  The Police Department responded and upon arriv…

  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
5         01/01/2016 01:50:11 AM     07 BROOKLYN   BROOKLYN
6         01/01/2016 01:53:54 AM    12 MANHATTAN  MANHATTAN
7         01/01/2016 01:42:54 AM        03 BRONX      BRONX
8         01/01/2016 08:27:32 AM       09 QUEENS     QUEENS
9         01/01/2016 01:17:40 AM     11 BROOKLYN   BROOKLYN

   X Coordinate (State Plane)  Y Coordinate (State Plane) Park Facility Name  \
0                   1005409.0                    254678.0        Unspecified
1                   1007766.0                    221986.0        Unspecified
2                   1015081.0                    256380.0        Unspecified
3                   1031740.0                    243899.0        Unspecified
4                   1019123.0                    206375.0        Unspecified
5                    986312.0                    180032.0        Unspecified
6                   1001578.0                    245627.0        Unspecified
7                   1011117.0                    244417.0        Unspecified
8                   1030662.0                    196163.0        Unspecified
9                    984378.0                    166541.0        Unspecified

  Park Borough   Latitude  Longitude                                  Location
0    MANHATTAN  40.865682 -73.923501   (40.86568153633767, -73.92350095571744)
1       QUEENS  40.775945 -73.915094  (40.775945312321085, -73.91509393898605)
2        BRONX  40.870325 -73.888525  (40.870324522111424, -73.88852464418646)
3        BRONX  40.835994 -73.828379   (40.83599404683083, -73.82837939584206)
4       QUEENS  40.733060 -73.874170  (40.733059618956815, -73.87416975810375)
```

```
5     BROOKLYN  40.660823 -73.992568   (40.66082272389114, -73.99256786342693)
6    MANHATTAN  40.840848 -73.937375   (40.840847591440415, -73.9373750864581)
7        BRONX  40.837503 -73.902905   (40.83750262540012, -73.90290517326568)
8       QUEENS  40.704977 -73.832605   (40.704977164399935, -73.8326047502584)
9     BROOKLYN  40.623793 -73.999539  (40.623793065806524, -73.99953890121567)
```

[21]:
```
# remaining 29 columns
df4.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 364558 entries, 0 to 364557
Data columns (total 29 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   Unique Key                      364558 non-null  int64
 1   Created Date                    364558 non-null  datetime64[ns]
 2   Closed Date                     362177 non-null  datetime64[ns]
 3   Agency                          364558 non-null  object
 4   Agency Name                     364558 non-null  object
 5   Complaint Type                  364558 non-null  object
 6   Descriptor                      358057 non-null  object
 7   Location Type                   364425 non-null  object
 8   Incident Zip                    361560 non-null  float64
 9   Incident Address                312859 non-null  object
 10  Street Name                     312859 non-null  object
 11  Cross Street 1                  307370 non-null  object
 12  Cross Street 2                  306753 non-null  object
 13  Address Type                    361306 non-null  object
 14  City                            361561 non-null  object
 15  Facility Type                   362169 non-null  object
 16  Status                          364558 non-null  object
 17  Due Date                        364555 non-null  datetime64[ns]
 18  Resolution Description          364558 non-null  object
 19  Resolution Action Updated Date  362156 non-null  object
 20  Community Board                 364558 non-null  object
 21  Borough                         364558 non-null  object
 22  X Coordinate (State Plane)      360528 non-null  float64
 23  Y Coordinate (State Plane)      360528 non-null  float64
 24  Park Facility Name              364558 non-null  object
 25  Park Borough                    364558 non-null  object
 26  Latitude                        360528 non-null  float64
 27  Longitude                       360528 non-null  float64
 28  Location                        360528 non-null  object
dtypes: datetime64[ns](3), float64(5), int64(1), object(20)
memory usage: 80.7+ MB
```

[22]:
```
# Borough and community board have similar values so drop 'community board'
df5 = df4.drop(columns = ['Community Board'])
df5
```

[22]:

	Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	Street Name	Cross Street 1
0	32310363	2015-12-31 23:59:45	2016-01-01 00:55:15	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	Street/Sidewalk	10034.0	71 VERMILYEA AVENUE	VERMILYEA AVENUE	ACADEMY STREET
1	32309934	2015-12-31 23:59:44	2016-01-01 01:26:57	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	11105.0	27-07 23 AVENUE	23 AVENUE	27 STREET
2	32309159	2015-12-31 23:59:29	2016-01-01 04:51:03	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	10458.0	2897 VALENTINE AVENUE	VALENTINE AVENUE	EAST 198 STREET
3	32305098	2015-12-31 23:57:46	2016-01-01 07:43:13	NYPD	New York City Police Department	Illegal Parking	Commercial Overnight Parking	Street/Sidewalk	10461.0			
4	32306529	2015-12-31 23:56:58	2016-01-01 03:24:42	NYPD	New York City Police Department	Illegal Parking	Blocked Sidewalk	Street/Sidewalk	11373.0			
…	…	…	…	…	…	…	…	…	…			
364553	29609918	2015-01-01 00:04:44	2015-01-01 10:22:31	NYPD	New York City Police Department	Illegal Parking	Blocked Hydrant	Street/Sidewalk	11421.0			
364554	29608392	2015-01-01 00:04:28	2015-01-01 02:25:02	NYPD	New York City Police Department	Noise - Vehicle	Car/Truck Horn	Street/Sidewalk	10468.0			
364555	29607589	2015-01-01 00:01:30	2015-01-01 00:20:33	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	Street/Sidewalk	10031.0			
364556	29610889	2015-01-01 00:01:29	2015-01-01 02:42:22	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	10466.0			
364557	29611816	2015-01-01 00:00:50	2015-01-01 02:47:50	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	11420.0			

```
3         2940 BAISLEY AVENUE    BAISLEY AVENUE       EDISON AVENUE
4               87-14 57 ROAD           57 ROAD      SEABURY STREET
…                         …                …                 …
364553          84-25 85 ROAD           85 ROAD      FOREST PARKWAY
364554   2555 SEDGWICK AVENUE   SEDGWICK AVENUE       BAILEY AVENUE
364555    508 WEST 139 STREET   WEST 139 STREET    AMSTERDAM AVENUE
364556    931 EAST 226 STREET   EAST 226 STREET    BRONXWOOD AVENUE
364557      123-19 135 STREET        135 STREET  ROCKAWAY BOULEVARD

         Cross Street 2 Address Type             City Facility Type  Status  \
0       WEST 204 STREET      ADDRESS          NEW YORK      Precinct  Closed
1             28 STREET      ADDRESS           ASTORIA      Precinct  Closed
2       EAST 199 STREET      ADDRESS             BRONX      Precinct  Closed
3             B STREET       ADDRESS             BRONX      Precinct  Closed
4         HOFFMAN DRIVE      ADDRESS          ELMHURST      Precinct  Closed
…                   …           …                …           …     …
364553        85 STREET      ADDRESS         WOODHAVEN      Precinct  Closed
364554             BEND      ADDRESS             BRONX      Precinct  Closed
364555   HAMILTON PLACE      ADDRESS          NEW YORK      Precinct  Closed
364556  PAULDING AVENUE      ADDRESS             BRONX      Precinct  Closed
364557    SUTTER AVENUE      ADDRESS  SOUTH OZONE PARK      Precinct  Closed

                  Due Date                                Resolution Description  \
0       2016-01-01 07:59:45  The Police Department responded and upon arriv…
1       2016-01-01 07:59:44  The Police Department responded to the complai…
2       2016-01-01 07:59:29  The Police Department responded and upon arriv…
3       2016-01-01 07:57:46  The Police Department responded to the complai…
4       2016-01-01 07:56:58  The Police Department responded and upon arriv…
…                       …                                                 …
364553 2015-01-01 08:04:44  The Police Department responded to the complai…
364554 2015-01-01 08:04:28  The Police Department responded to the complai…
364555 2015-01-01 08:01:30  The Police Department responded to the complai…
364556 2015-01-01 08:01:29  The Police Department responded and upon arriv…
364557 2015-01-01 08:00:50  The Police Department responded to the complai…

        Resolution Action Updated Date    Borough  X Coordinate (State Plane)  \
0              01/01/2016 12:55:15 AM  MANHATTAN                   1005409.0
1              01/01/2016 01:26:57 AM     QUEENS                   1007766.0
2              01/01/2016 04:51:03 AM      BRONX                   1015081.0
3              01/01/2016 07:43:13 AM      BRONX                   1031740.0
4              01/01/2016 03:24:42 AM     QUEENS                   1019123.0
…                                  …          …                          …
364553         01/01/2015 10:22:31 AM     QUEENS                   1022809.0
364554         01/01/2015 02:25:02 AM      BRONX                   1009923.0
364555         01/01/2015 12:20:33 AM  MANHATTAN                    997847.0
364556         01/01/2015 02:42:22 AM      BRONX                   1024816.0
364557         01/01/2015 02:47:50 AM     QUEENS                   1038733.0
```

```
        Y Coordinate (State Plane) Park Facility Name Park Borough   Latitude  \
0                         254678.0        Unspecified    MANHATTAN  40.865682
1                         221986.0        Unspecified       QUEENS  40.775945
2                         256380.0        Unspecified        BRONX  40.870325
3                         243899.0        Unspecified        BRONX  40.835994
4                         206375.0        Unspecified       QUEENS  40.733060
…                               …                  …            …         …
364553                    192567.0        Unspecified       QUEENS  40.695145
364554                    255465.0        Unspecified        BRONX  40.867830
364555                    238629.0        Unspecified    MANHATTAN  40.821647
364556                    262237.0        Unspecified        BRONX  40.886361
364557                    184971.0        Unspecified       QUEENS  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 28 columns]
```

[23]:
```
# creating new column as time elapsed between request creation and closing (to␣
 ↪show how fast the response time is)
df5['Closing Time'] = df5['Closed Date'] - df5['Created Date'] #calculating␣
 ↪time taken
```

[24]:
```
df5.head()
```

[24]:
```
   Unique Key        Created Date         Closed Date Agency  \
0    32310363 2015-12-31 23:59:45 2016-01-01 00:55:15   NYPD
1    32309934 2015-12-31 23:59:44 2016-01-01 01:26:57   NYPD
2    32309159 2015-12-31 23:59:29 2016-01-01 04:51:03   NYPD
3    32305098 2015-12-31 23:57:46 2016-01-01 07:43:13   NYPD
4    32306529 2015-12-31 23:56:58 2016-01-01 03:24:42   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

                    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

        Incident Address       Street Name   Cross Street 1   Cross Street 2  \
0    71 VERMILYEA AVENUE  VERMILYEA AVENUE   ACADEMY STREET  WEST 204 STREET
1        27-07 23 AVENUE         23 AVENUE        27 STREET        28 STREET
2  2897 VALENTINE AVENUE  VALENTINE AVENUE  EAST 198 STREET  EAST 199 STREET
3    2940 BAISLEY AVENUE    BAISLEY AVENUE    EDISON AVENUE         B STREET
4          87-14 57 ROAD           57 ROAD   SEABURY STREET    HOFFMAN DRIVE

  Address Type      City Facility Type  Status            Due Date  \
0      ADDRESS  NEW YORK      Precinct  Closed 2016-01-01 07:59:45
1      ADDRESS   ASTORIA      Precinct  Closed 2016-01-01 07:59:44
2      ADDRESS     BRONX      Precinct  Closed 2016-01-01 07:59:29
3      ADDRESS     BRONX      Precinct  Closed 2016-01-01 07:57:46
4      ADDRESS  ELMHURST      Precinct  Closed 2016-01-01 07:56:58

                             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…

  Resolution Action Updated Date    Borough  X Coordinate (State Plane)  \
0         01/01/2016 12:55:15 AM  MANHATTAN                   1005409.0
1         01/01/2016 01:26:57 AM     QUEENS                   1007766.0
2         01/01/2016 04:51:03 AM      BRONX                   1015081.0
3         01/01/2016 07:43:13 AM      BRONX                   1031740.0
4         01/01/2016 03:24:42 AM     QUEENS                   1019123.0

   Y Coordinate (State Plane) Park Facility Name Park Borough   Latitude  \
0                    254678.0        Unspecified    MANHATTAN  40.865682
1                    221986.0        Unspecified       QUEENS  40.775945
2                    256380.0        Unspecified        BRONX  40.870325
3                    243899.0        Unspecified        BRONX  40.835994
4                    206375.0        Unspecified       QUEENS  40.733060

   Longitude                                  Location    Closing Time
0 -73.923501   (40.86568153633767, -73.92350095571744) 0 days 00:55:30
```

```
 1 -73.915094  (40.775945312321085, -73.91509393898605) 0 days 01:27:13
 2 -73.888525  (40.870324522111424, -73.88852464418646) 0 days 04:51:34
 3 -73.828379   (40.83599404683083, -73.82837939584206) 0 days 07:45:27
 4 -73.874170  (40.733059618956815, -73.87416975810375) 0 days 03:27:44
```

[25]:
```
df5['Time Bin'] = pd.cut(df5['Closing Time'], 100)#
```

[26]:
```
df5.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 364558 entries, 0 to 364557
Data columns (total 30 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   Unique Key                      364558 non-null  int64
 1   Created Date                    364558 non-null  datetime64[ns]
 2   Closed Date                     362177 non-null  datetime64[ns]
 3   Agency                          364558 non-null  object
 4   Agency Name                     364558 non-null  object
 5   Complaint Type                  364558 non-null  object
 6   Descriptor                      358057 non-null  object
 7   Location Type                   364425 non-null  object
 8   Incident Zip                    361560 non-null  float64
 9   Incident Address                312859 non-null  object
 10  Street Name                     312859 non-null  object
 11  Cross Street 1                  307370 non-null  object
 12  Cross Street 2                  306753 non-null  object
 13  Address Type                    361306 non-null  object
 14  City                            361561 non-null  object
 15  Facility Type                   362169 non-null  object
 16  Status                          364558 non-null  object
 17  Due Date                        364555 non-null  datetime64[ns]
 18  Resolution Description          364558 non-null  object
 19  Resolution Action Updated Date  362156 non-null  object
 20  Borough                         364558 non-null  object
 21  X Coordinate (State Plane)      360528 non-null  float64
 22  Y Coordinate (State Plane)      360528 non-null  float64
 23  Park Facility Name              364558 non-null  object
 24  Park Borough                    364558 non-null  object
 25  Latitude                        360528 non-null  float64
 26  Longitude                       360528 non-null  float64
 27  Location                        360528 non-null  object
 28  Closing Time                    362177 non-null  timedelta64[ns]
 29  Time Bin                        362177 non-null  category
dtypes: category(1), datetime64[ns](3), float64(5), int64(1), object(19),
timedelta64[ns](1)
memory usage: 81.0+ MB
```

[27]:
```
df5['Closing Time Sec'] = df5['Closing Time'].apply(lambda x:x.seconds)
```

[28]:
```
df5['Closing Time Sec']
```

[28]:
```
0          3330.0
1          5233.0
2         17494.0
3         27927.0
4         12464.0
           ...
364553    37067.0
364554     8434.0
364555     1143.0
364556     9653.0
364557    10020.0
Name: Closing Time Sec, Length: 364558, dtype: float64
```

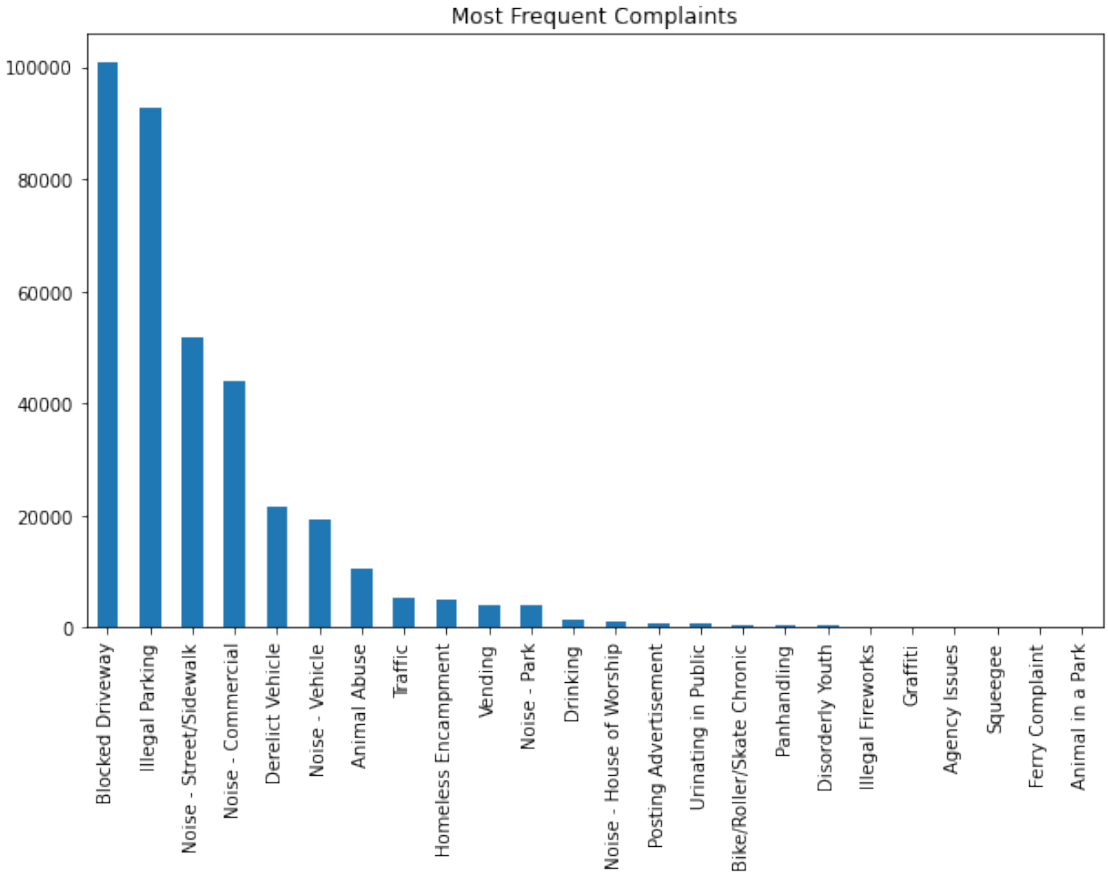
[29]:
```
df5.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 364558 entries, 0 to 364557
Data columns (total 31 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   Unique Key                      364558 non-null  int64
 1   Created Date                    364558 non-null  datetime64[ns]
 2   Closed Date                     362177 non-null  datetime64[ns]
 3   Agency                          364558 non-null  object
 4   Agency Name                     364558 non-null  object
 5   Complaint Type                  364558 non-null  object
 6   Descriptor                      358057 non-null  object
 7   Location Type                   364425 non-null  object
 8   Incident Zip                    361560 non-null  float64
 9   Incident Address                312859 non-null  object
 10  Street Name                     312859 non-null  object
 11  Cross Street 1                  307370 non-null  object
 12  Cross Street 2                  306753 non-null  object
 13  Address Type                    361306 non-null  object
 14  City                            361561 non-null  object
 15  Facility Type                   362169 non-null  object
 16  Status                          364558 non-null  object
 17  Due Date                        364555 non-null  datetime64[ns]
 18  Resolution Description          364558 non-null  object
 19  Resolution Action Updated Date  362156 non-null  object
 20  Borough                         364558 non-null  object
 21  X Coordinate (State Plane)      360528 non-null  float64
 22  Y Coordinate (State Plane)      360528 non-null  float64
```

```
 23  Park Facility Name              364558 non-null  object
 24  Park Borough                    364558 non-null  object
 25  Latitude                        360528 non-null  float64
 26  Longitude                       360528 non-null  float64
 27  Location                        360528 non-null  object
 28  Closing Time                    362177 non-null  timedelta64[ns]
 29  Time Bin                        362177 non-null  category
 30  Closing Time Sec                362177 non-null  float64
dtypes: category(1), datetime64[ns](3), float64(6), int64(1), object(19),
timedelta64[ns](1)
memory usage: 83.8+ MB
```

1.2 Finding major types of complaints

[30]:
```
# bar graph of count vs. complaint types
df5['Complaint Type'].value_counts().plot(kind='bar',figsize=(10,6),title='Most␣
 ↪Frequent Complaints')
plt.show()
```



[31]:
```
# showing top 10 most amount of complaints with blocked driveway and illegal␣
 ↪parking as the most by a lot
df5['Complaint Type'].value_counts().head(10)
```

[31]:
```
Blocked Driveway           100881
Illegal Parking             92679
Noise - Street/Sidewalk     51692
Noise - Commercial          44109
Derelict Vehicle            21661
Noise - Vehicle             19352
Animal Abuse                10541
Traffic                      5198
Homeless Encampment          4879
Vending                      4192
Name: Complaint Type, dtype: int64
```

[32]:
```
# Showing how many complaints per Borough but some are unspecified
df5['Borough'].value_counts()
```

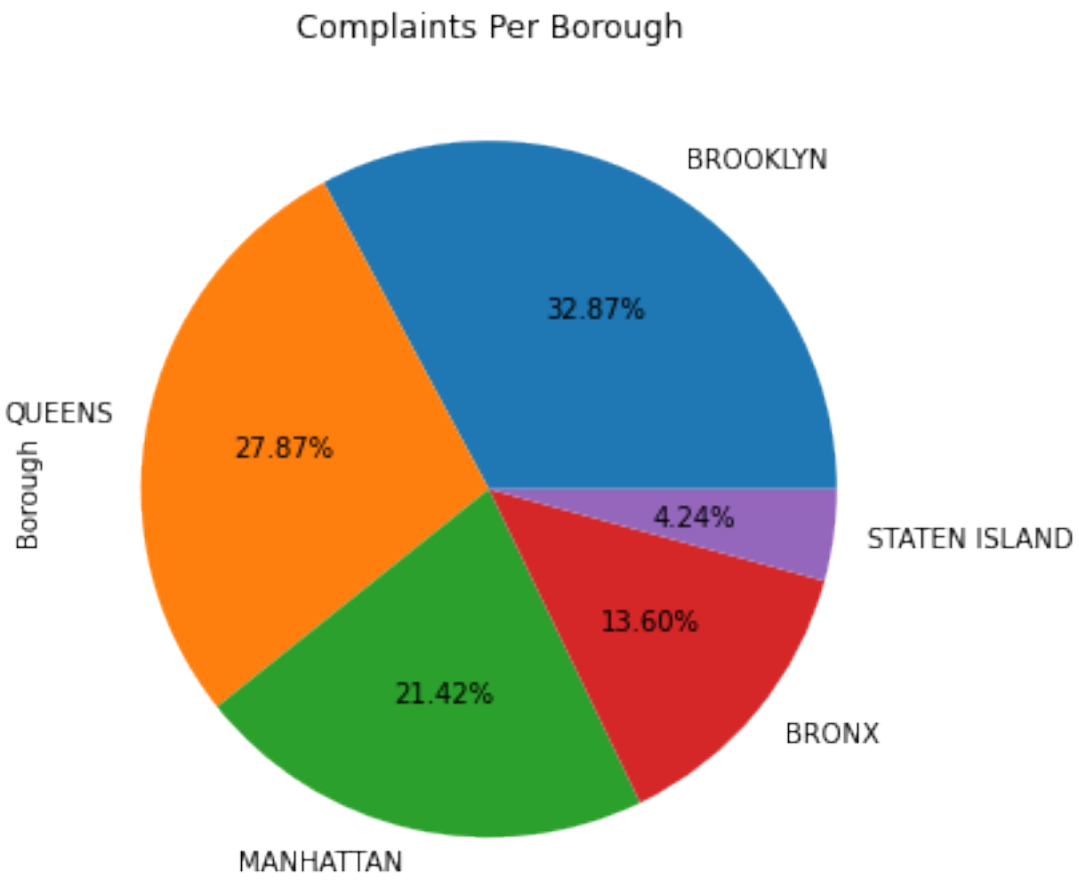
[32]:
```
BROOKLYN         118864
QUEENS           100766
MANHATTAN         77462
BRONX             49169
STATEN ISLAND     15339
Unspecified        2958
Name: Borough, dtype: int64
```

[33]:
```
# removing unspecified complaints
df5.replace('Unspecified',np.NaN,inplace=True)
df5=df5.dropna(how='all')
```

[34]:
```
# confirming unspecified are removed
df5['Borough'].value_counts()
```

[34]:
```
BROOKLYN         118864
QUEENS           100766
MANHATTAN         77462
BRONX             49169
STATEN ISLAND     15339
Name: Borough, dtype: int64
```

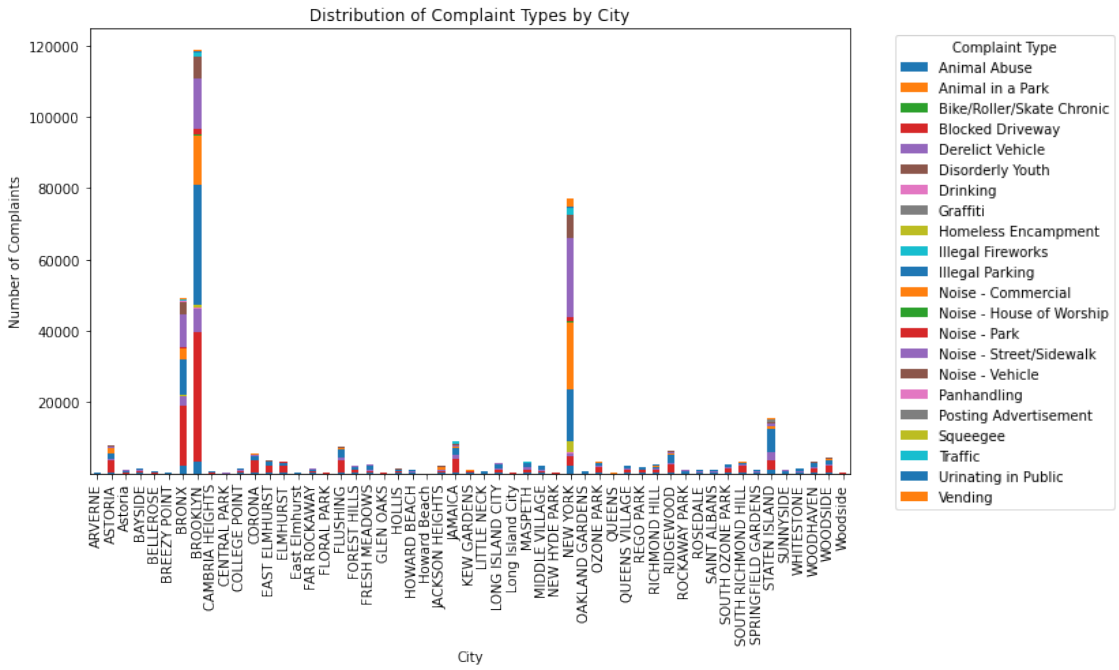
[35]:
```
# pie plot showing Borough wise percentage of complaints and that Brooklyn␣
 ↪followed by Queens have the most
df5['Borough'].value_counts().plot(kind='pie',autopct='%.
 ↪2f%%',figsize=(10,6),title='Complaints Per Borough')
plt.show()
```



1.3 Visualize the major types of complaints in each city

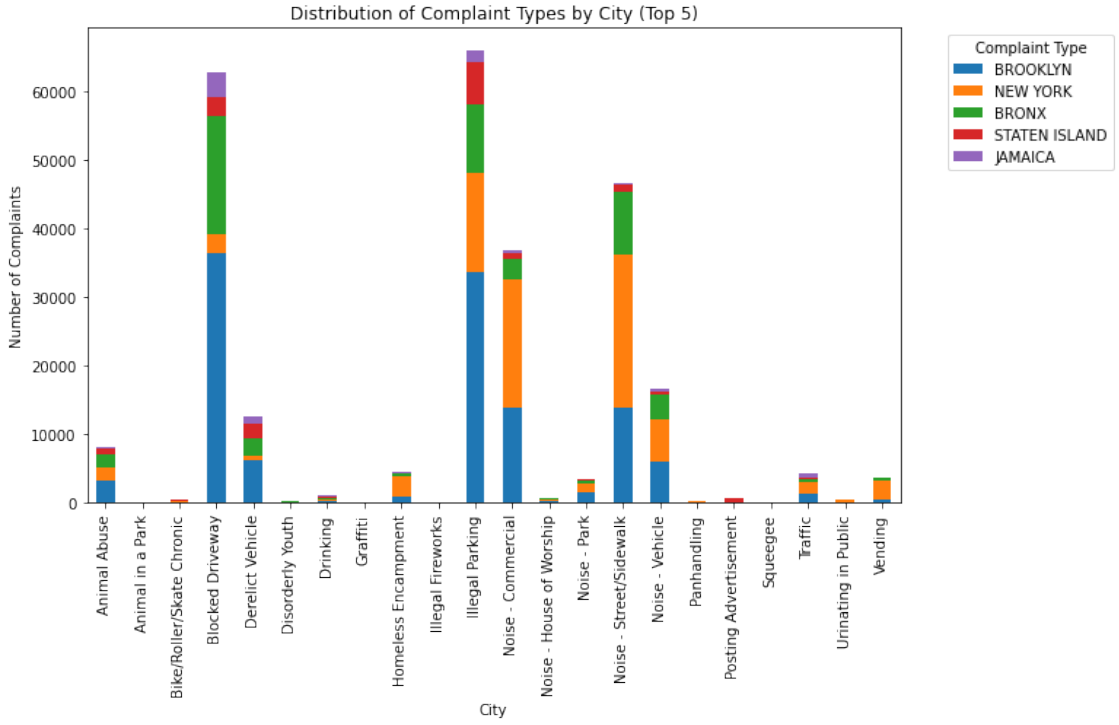
[36]:
```
# Group the data by city and complaint type, and count the number of complaints
grouped_data = df5.groupby(['City', 'Complaint Type']).size().
 ↪reset_index(name='count')

# Pivot the data to create a matrix with cities as rows, complaint types as␣
 ↪columns, and complaint counts as values
pivoted_data = grouped_data.pivot(index='City', columns='Complaint Type',␣
 ↪values='count')

# Create a bar plot with stacked bars for each city, showing the distribution␣
 ↪of complaint types
pivoted_data.plot(kind='bar', stacked=True, figsize=(10, 6))

# Add labels and titles to the plot
```

```
plt.xlabel('City')
plt.ylabel('Number of Complaints')
plt.title('Distribution of Complaint Types by City')
plt.legend(title='Complaint Type', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.show()
```

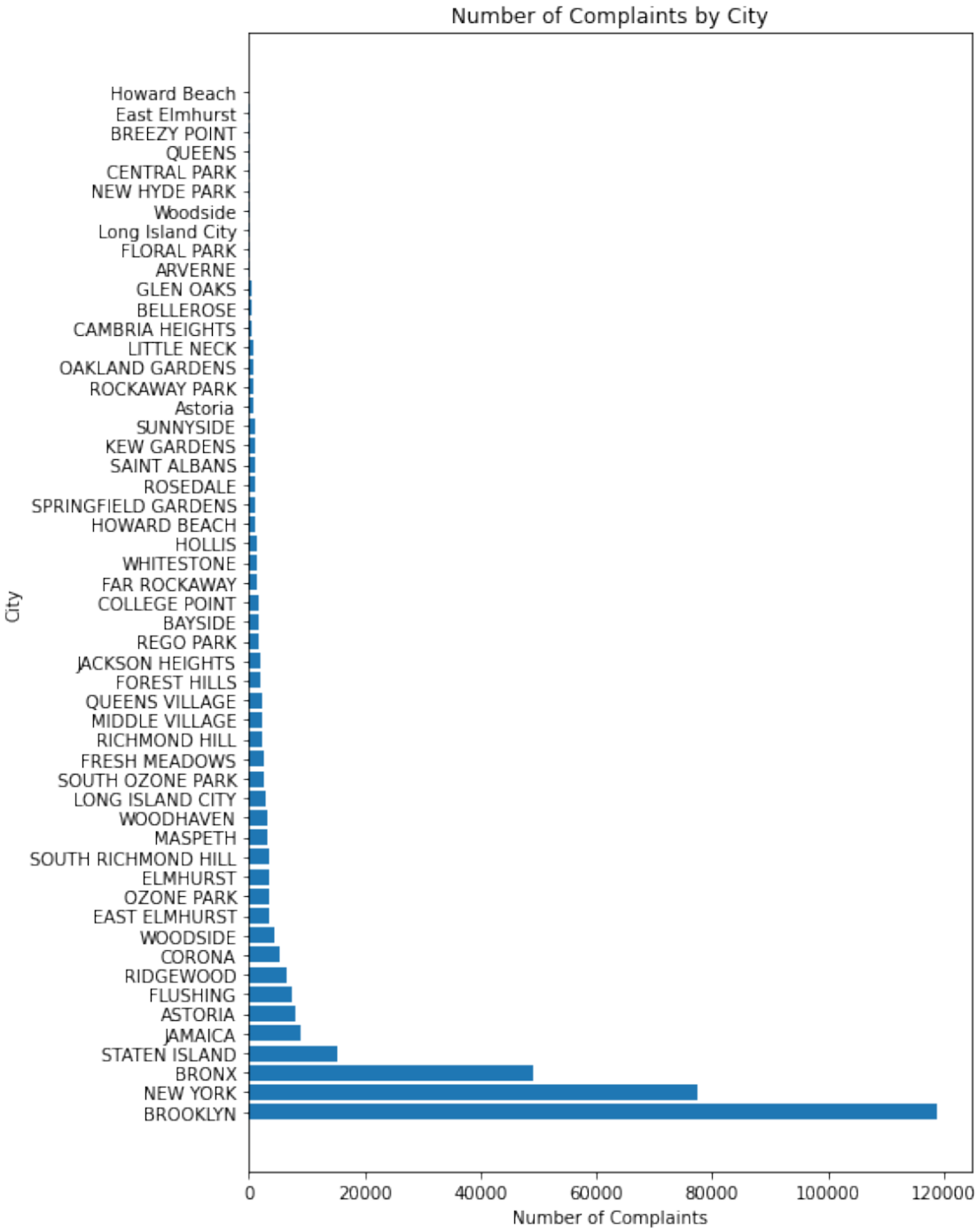


[37]:
```
# Calculate the total number of complaints for each city
city_totals = pivoted_data.sum(axis=1)

# sort the cities by this value and pick the top 5 (the first 5 elements of the␣
 ↪index, from the beginning up to (but not including) the element at index 5)
sorted_cities = city_totals.sort_values(ascending=False).index[:5]

# Filter the pivoted data to show only the top 5 cities and their most common␣
 ↪complaint types
pd1 = pivoted_data.loc[sorted_cities]
pd2 = pd1.loc[pd1.sum(axis=1).sort_values(ascending=False).index[:5]].T

# Create a bar plot with stacked bars for each city, showing the distribution␣
 ↪of complaint types
pd2.plot(kind='bar', stacked=True, figsize=(10, 6))

# Add labels and titles to the plot
plt.xlabel('City')
plt.ylabel('Number of Complaints')
```

```
plt.title('Distribution of Complaint Types by City (Top 5)')
plt.legend(title='Complaint Type', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.show()
```

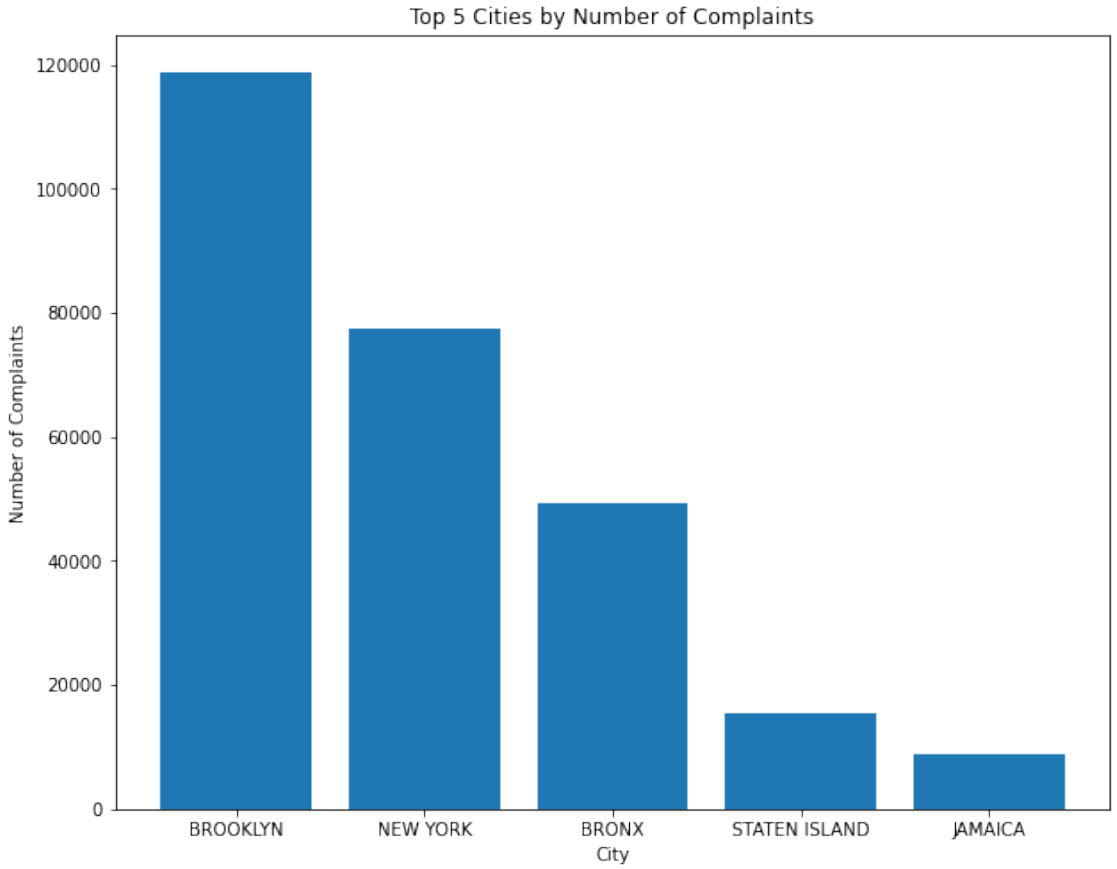


[38]:
```
# Get the value counts for each city
city_counts = df['City'].value_counts()

# Create a horizontal bar chart
plt.figure(figsize=(8, 10))
plt.barh(city_counts.index, city_counts.values)

# Add labels and titles to the plot
plt.xlabel('Number of Complaints')
plt.ylabel('City')
plt.title('Number of Complaints by City')

# Adjust the padding between subplots to prevent overlapping labels
plt.tight_layout()
plt.show()
```



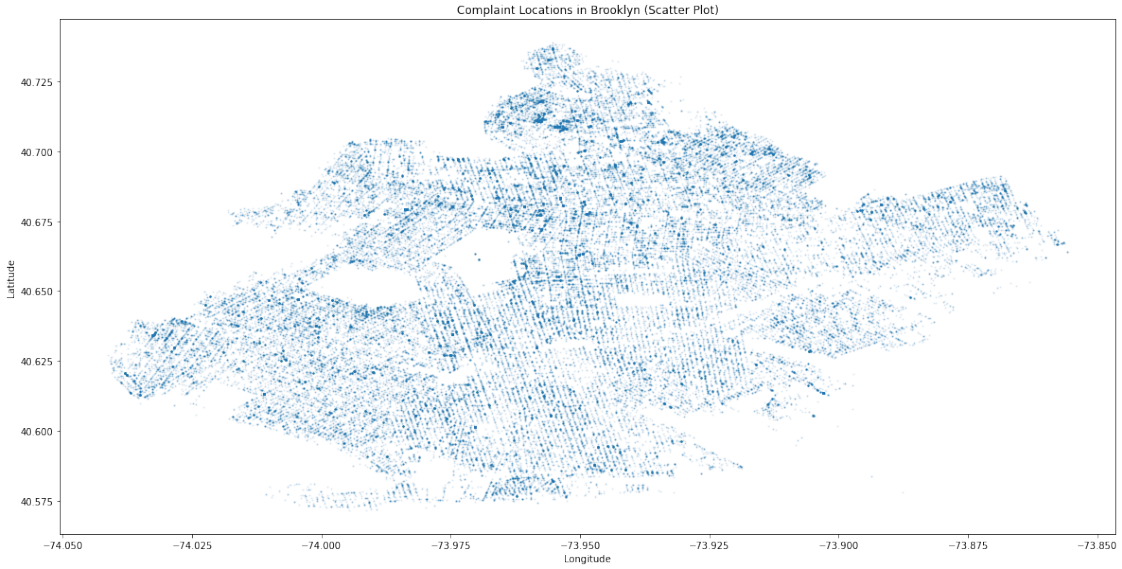
[39]:
```
# get the frequency of complaints by city
complaint_freq = df5['City'].value_counts()

# create a bar plot of the complaint frequencies
fig, ax = plt.subplots(figsize=(10, 8))
```

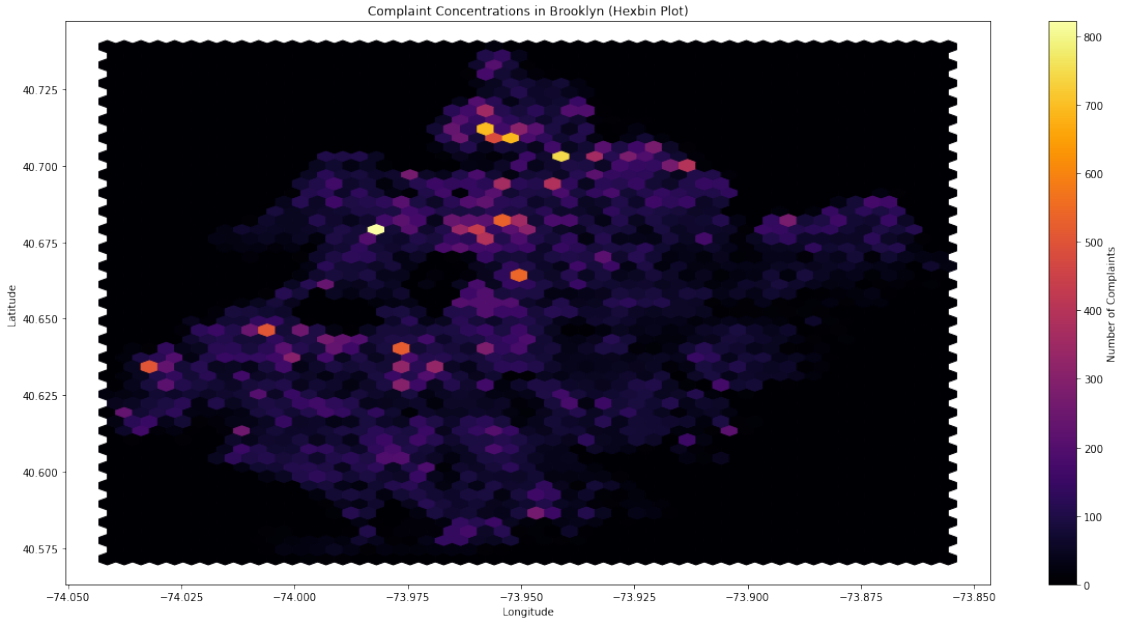
```
# the x-axis values are top 5 complaint types (extracted from the index using .
 ↪index) and y-axis values are frequency of each complaint type (extracted␣
 ↪from the complaint_freq series)
ax.bar(complaint_freq.index[:5], complaint_freq[:5])

# Add labels and titles to the plot
ax.set_xlabel('City')
ax.set_ylabel('Number of Complaints')
ax.set_title('Top 5 Cities by Number of Complaints')
plt.show()
```



[40]:
```
# select only the rows of df5 where the borough is 'BROOKLYN'
brooklyn_data = df5[df5['Borough'] == 'BROOKLYN']

# scatter plot of complaint locations in Brooklyn
fig, ax = plt.subplots(figsize=(20,10))
ax.scatter(brooklyn_data['Longitude'], brooklyn_data['Latitude'], s=1, alpha=0.
 ↪1)
```

```
# Add labels and titles to the plot
ax.set_xlabel('Longitude')
ax.set_ylabel('Latitude')
ax.set_title('Complaint Locations in Brooklyn (Scatter Plot)')
plt.show()

# Create a hexbin plot of complaint concentrations in Brooklyn
fig, ax = plt.subplots(figsize=(20,10))
hb = ax.hexbin(brooklyn_data['Longitude'], brooklyn_data['Latitude'],␣
↪gridsize=50, cmap='inferno')

# Add a colorbar and labels and titles to the plot
cb = plt.colorbar(hb, ax=ax)
cb.set_label('Number of Complaints')
ax.set_xlabel('Longitude')
ax.set_ylabel('Latitude')
ax.set_title('Complaint Concentrations in Brooklyn (Hexbin Plot)')
plt.show()
```





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

[41]:
```
# create contingency table of the variables we want to test
cont_table = pd.crosstab(df5['Complaint Type'], df5['Borough'])
```

[42]:
```
# chi-square test of independence to calculate the p-value and chi-square value
chi2, p, dof, expected = stats.chi2_contingency(cont_table)
```

[43]:
```
# print out the p-value and chi-square value
print("p-value:", p)
print("chi-square:", chi2)
```

```
p-value: 0.0
chi-square: 93431.06106800103
```

-Null Hypothesis: H0 - there is NO significant relation between type of complaint and location

-Alternate Hypothesis: Ha - there IS significant relation between type of complaint and location

1.5 Results

1.5.1 p-value 0.00 < 0.05

1.5.2 -REJECT null hypothesis

1.5.3 -complaint type and location ARE related

[ ]: