In [2]:

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

In [3]:

```
df=pd.read_csv("/Users/david/Downloads/Customer_Service_Requests_Analysis_Dataset/Customer_Servic
```

In [4]:

```
#Understand the dataset:
```

In [5]:

```
#Identify the shape of the dataset
df.shape
```

Out[5]:

```
(364558, 53)
```

In [6]:

```
#Identify variables with null values
df.isnull().sum()
```

Out[6]:

```
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
```

In [7]:

```
#Perform basic data exploratory analysis:
```

In [8]:

```
#Utilize missing value treatment

df.isnull().sum()/len(df)*100
```

Out[8]:

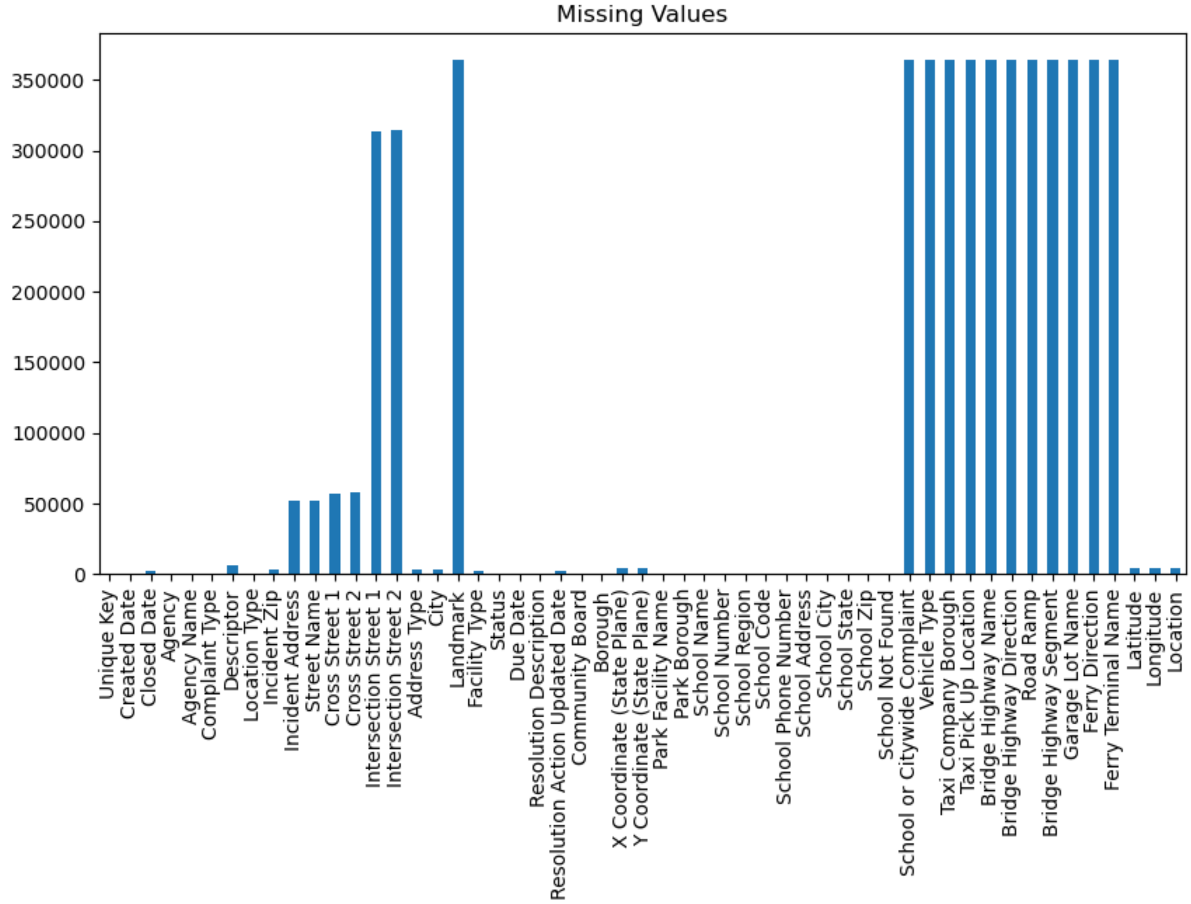
```
Unique Key                          0.000000
Created Date                        0.000000
Closed Date                         0.653120
Agency                              0.000000
Agency Name                         0.000000
Complaint Type                      0.000000
Descriptor                          1.783255
Location Type                       0.036483
Incident Zip                        0.822366
Incident Address                   14.181283
Street Name                        14.181283
Cross Street 1                     15.686941
Cross Street 2                     15.856187
Intersection Street 1              85.977540
Intersection Street 2              86.144317
Address Type                        0.892039
City                                0.822091
Landmark                           99.897136
Facility Type                       0.655314
Status                              0.000000
Due Date                            0.000823
Resolution Description              0.000000
Resolution Action Updated Date      0.658880
Community Board                     0.000000
Borough                             0.000000
X Coordinate (State Plane)          1.105448
Y Coordinate (State Plane)          1.105448
Park Facility Name                  0.000000
Park Borough                        0.000000
School Name                         0.000000
School Number                       0.000000
School Region                       0.000274
School Code                         0.000274
School Phone Number                 0.000000
School Address                      0.000000
School City                         0.000000
School State                        0.000000
School Zip                          0.000274
School Not Found                    0.000000
School or Citywide Complaint      100.000000
Vehicle Type                      100.000000
Taxi Company Borough              100.000000
Taxi Pick Up Location             100.000000
Bridge Highway Name                99.918531
Bridge Highway Direction           99.918531
Road Ramp                          99.928132
Bridge Highway Segment             99.928132
Garage Lot Name                   100.000000
Ferry Direction                    99.999726
Ferry Terminal Name                99.999451
Latitude                            1.105448
Longitude                           1.105448
Location                            1.105448
dtype: float64
```

In [9]:

```
df.isnull().sum().plot(kind="bar",figsize=(10,5),title=("Missing Values"))
```

Out[9]:

```
<Axes: title={'center': 'Missing Values'}>
```



In [10]:

```
un_col=['School or Citywide Complaint','Vehicle Type','Taxi Company Borough','Taxi Pick Up Locati
       'Bridge Highway Name','Bridge Highway Direction','Road Ramp','Bridge Highway Segment',
       'Garage Lot Name','Ferry Direction','Ferry Terminal Name','Landmark','Intersection Street
       'Intersection Street 1']
```

In [11]:

```
df.drop(un_col,axis=1,inplace=True)
```

In [12]:

```
df.isnull().sum()/len(df)*100
```

Out[12]:

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

In [13]:

```
df=df[['Unique Key','Created Date','Closed Date','Agency',
      'Complaint Type','Descriptor','Location Type','Incident Zip','City','Status',
      'Resolution Description','Borough','Latitude','Longitude','Location']]
```

In [14]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 364558 entries, 0 to 364557
Data columns (total 15 columns):
 #   Column                  Non-Null Count   Dtype  
---  ------                  --------------   -----  
 0   Unique Key              364558 non-null  int64  
 1   Created Date            364558 non-null  object 
 2   Closed Date             362177 non-null  object 
 3   Agency                  364558 non-null  object 
 4   Complaint Type          364558 non-null  object 
 5   Descriptor              358057 non-null  object 
 6   Location Type           364425 non-null  object 
 7   Incident Zip            361560 non-null  float64
 8   City                    361561 non-null  object 
 9   Status                  364558 non-null  object 
 10  Resolution Description  364558 non-null  object 
 11  Borough                 364558 non-null  object 
 12  Latitude                360528 non-null  float64
 13  Longitude               360528 non-null  float64
 14  Location                360528 non-null  object 
dtypes: float64(3), int64(1), object(11)
memory usage: 41.7+ MB
```

In [15]:

```
df.isnull().sum()
```

Out[15]:

```
Unique Key                   0
Created Date                 0
Closed Date               2381
Agency                       0
Complaint Type               0
Descriptor                6501
Location Type              133
Incident Zip              2998
City                      2997
Status                       0
Resolution Description       0
Borough                      0
Latitude                  4030
Longitude                 4030
Location                  4030
dtype: int64
```

In [16]:

```
df.isnull().sum()/len(df)*100
```

Out[16]:

```
Unique Key                0.000000
Created Date              0.000000
Closed Date               0.653120
Agency                    0.000000
Complaint Type            0.000000
Descriptor                1.783255
Location Type             0.036483
Incident Zip              0.822366
City                      0.822091
Status                    0.000000
Resolution Description    0.000000
Borough                   0.000000
Latitude                  1.105448
Longitude                 1.105448
Location                  1.105448
dtype: float64
```

In [17]:

```
df.dropna(inplace=True)
```

In [18]:

```
df.isnull().sum()/len(df)*100
```

Out[18]:

```
Unique Key                0.0
Created Date              0.0
Closed Date               0.0
Agency                    0.0
Complaint Type            0.0
Descriptor                0.0
Location Type             0.0
Incident Zip              0.0
City                      0.0
Status                    0.0
Resolution Description    0.0
Borough                   0.0
Latitude                  0.0
Longitude                 0.0
Location                  0.0
dtype: float64
```

In [19]:

```
df.head()
```

Out[19]:

	Unique Key	Created Date	Closed Date	Agency	Complaint Type	Descriptor	Location Type	Incident Zip	
0	32310363	12/31/2015 11:59:45 PM	01/01/2016 12:55:15 AM	NYPD	Noise - Street/Sidewalk	Loud Music/Party	Street/Sidewalk	10034.0	NEW
1	32309934	12/31/2015 11:59:44 PM	01/01/2016 01:26:57 AM	NYPD	Blocked Driveway	No Access	Street/Sidewalk	11105.0	AS
2	32309159	12/31/2015 11:59:29 PM	01/01/2016 04:51:03 AM	NYPD	Blocked Driveway	No Access	Street/Sidewalk	10458.0	B
3	32305098	12/31/2015 11:57:46 PM	01/01/2016 07:43:13 AM	NYPD	Illegal Parking	Commercial Overnight Parking	Street/Sidewalk	10461.0	B
4	32306529	12/31/2015 11:56:58 PM	01/01/2016 03:24:42 AM	NYPD	Illegal Parking	Blocked Sidewalk	Street/Sidewalk	11373.0	ELMH

In [21]:

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

In [37]:

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

In [38]:

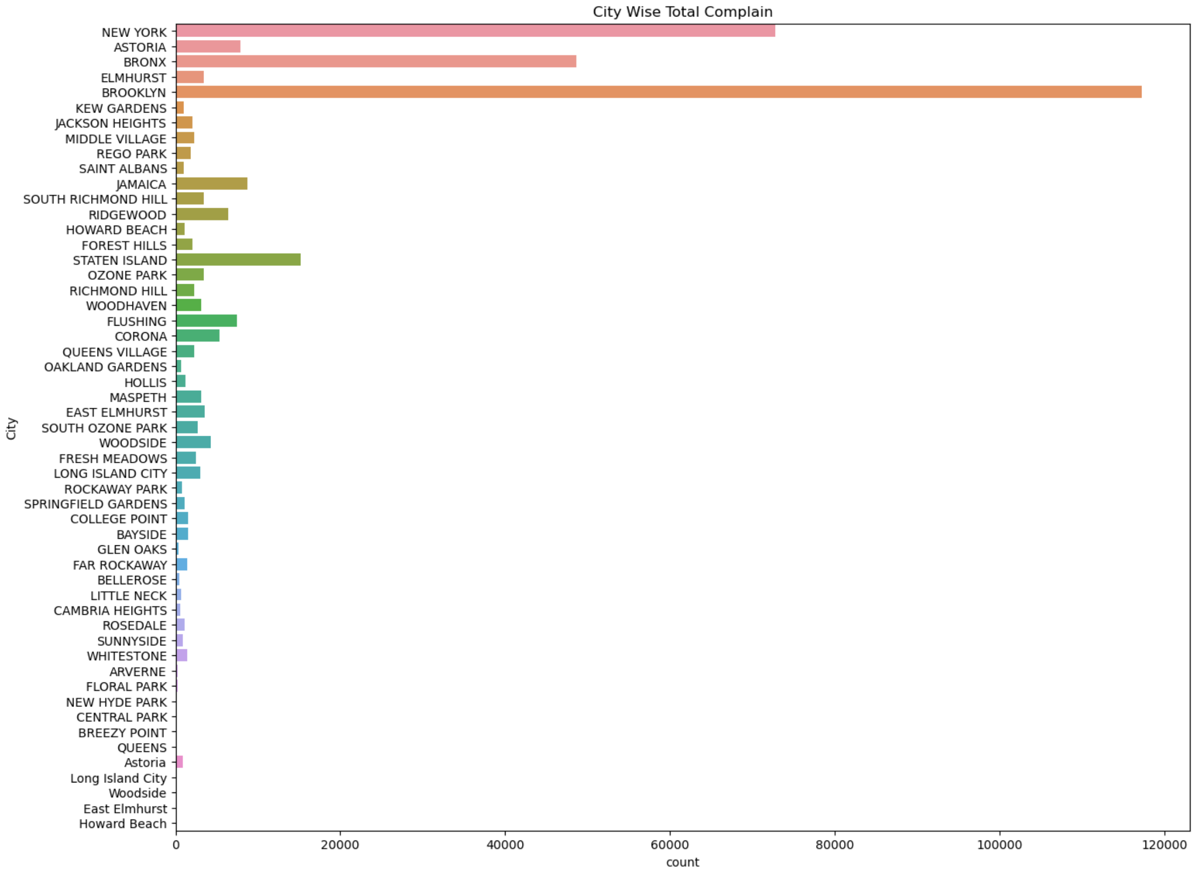
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 353891 entries, 0 to 364557
Data columns (total 15 columns):
 #   Column                  Non-Null Count   Dtype
---  ------                  --------------   -----
 0   Unique Key              353891 non-null  int64
 1   Created Date            353891 non-null  datetime64[ns]
 2   Closed Date             353891 non-null  datetime64[ns]
 3   Agency                  353891 non-null  object
 4   Complaint Type          353891 non-null  object
 5   Descriptor              353891 non-null  object
 6   Location Type           353891 non-null  object
 7   Incident Zip            353891 non-null  float64
 8   City                    353891 non-null  object
 9   Status                  353891 non-null  object
 10  Resolution Description  353891 non-null  object
 11  Borough                 353891 non-null  object
 12  Latitude                353891 non-null  float64
 13  Longitude               353891 non-null  float64
 14  Location                353891 non-null  object
dtypes: datetime64[ns](2), float64(3), int64(1), object(9)
memory usage: 43.2+ MB
```

In [59]:

```
#Draw a frequency plot for city-wise complaints

plt.figure(figsize=(15,12))
sns.countplot(data=df,y='City')
plt.title("City Wise Total Complain")
plt.show()
```



In [40]:

```
#Draw scatter and hexbin plots for complaint concentration across Brooklyn
```

In [41]:

```
df['Borough'].unique()
```

Out[41]:

```
array(['MANHATTAN', 'QUEENS', 'BRONX', 'BROOKLYN', 'STATEN ISLAND',
       'Unspecified'], dtype=object)
```

In [23]:

```
df_brooklyn=df[df['Borough']=='BROOKLYN']
```

In [43]:

```
df_brooklyn
```

Out[43]:

	Unique Key	Created Date	Closed Date	Agency	Complaint Type	Descriptor	Location Type	Incident Zip	
5	32306554	2015-12-31 23:56:30	2016-01-01 01:50:11	NYPD	Illegal Parking	Posted Parking Sign Violation	Street/Sidewalk	11215.0	BR(
9	32308391	2015-12-31 23:53:58	2016-01-01 01:17:40	NYPD	Blocked Driveway	No Access	Street/Sidewalk	11219.0	BR(
13	32305074	2015-12-31 23:47:58	2016-01-01 08:18:47	NYPD	Illegal Parking	Posted Parking Sign Violation	Street/Sidewalk	11208.0	BR(
17	32310273	2015-12-31 23:44:52	2016-01-01 00:36:10	NYPD	Noise - Commercial	Loud Music/Party	Club/Bar/Restaurant	11217.0	BR(
18	32306617	2015-12-31 23:40:59	2016-01-01 02:37:28	NYPD	Noise - Commercial	Loud Music/Party	Club/Bar/Restaurant	11234.0	BR(
...	...	...	...	...	...	...	...	...	
364539	29608505	2015-01-01 00:23:55	2015-01-01 02:58:38	NYPD	Blocked Driveway	No Access	Street/Sidewalk	11201.0	BR(
364541	29612697	2015-01-01 00:19:22	2015-01-01 02:41:10	NYPD	Blocked Driveway	No Access	Street/Sidewalk	11211.0	BR(
364544	29613295	2015-01-01 00:17:48	2015-01-01 03:24:48	NYPD	Noise - Commercial	Loud Music/Party	Store/Commercial	11217.0	BR(
364545	29613456	2015-01-01 00:17:47	2015-01-01 00:51:13	NYPD	Blocked Driveway	No Access	Street/Sidewalk	11237.0	BR(
364546	29613402	2015-01-01 00:15:45	2015-01-01 02:04:54	NYPD	Blocked Driveway	No Access	Street/Sidewalk	11218.0	BR(

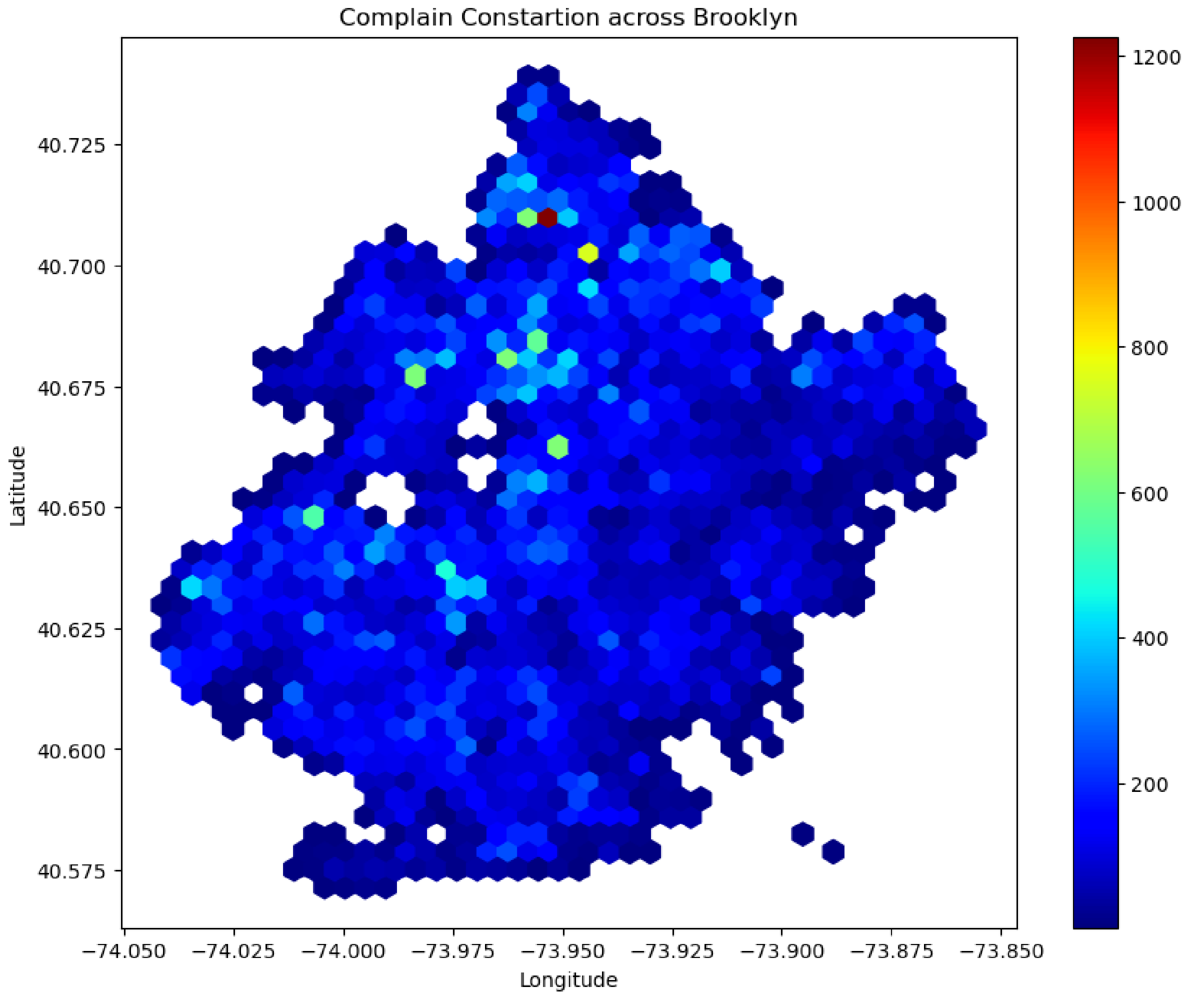
117270 rows × 15 columns

In [24]:

```
df_brooklyn[['Latitude','Longitude']].plot(kind='hexbin',x='Longitude',y='Latitude',
                                           title="Complain Constartion across Brooklyn",
                                           figsize=(10,8),
                                           gridsize=40,
                                           colormap='jet',
                                           mincnt=1)
```

Out[24]:

```
<Axes: title={'center': 'Complain Constartion across Brooklyn'}, xlabel='Longitud
e', ylabel='Latitude'>
```



In [ ]:

```

```

In [46]:

```
#Plot a bar graph of count vs. complaint types
```

In [47]:

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

Out[47]:

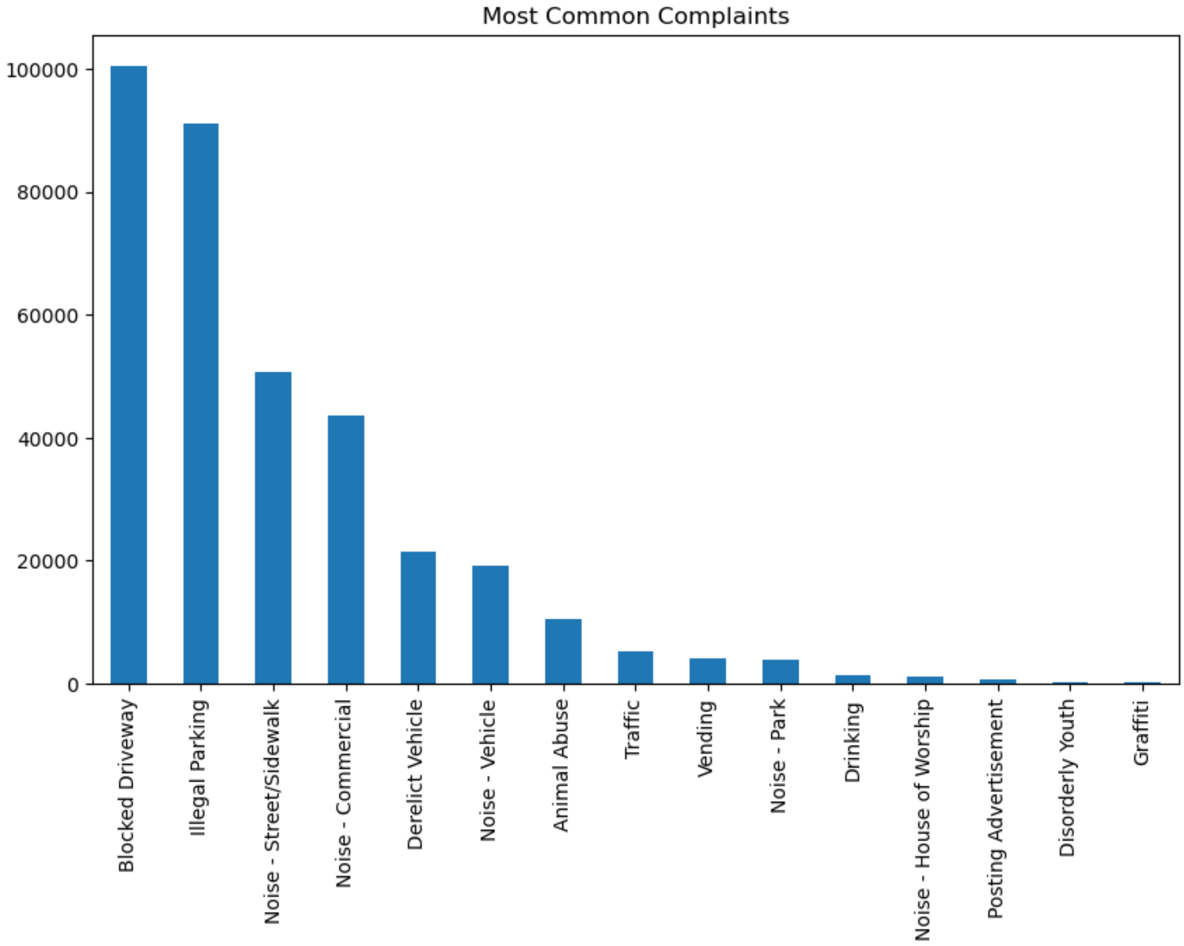
```
Blocked Driveway            100455
Illegal Parking              91057
Noise - Street/Sidewalk      50791
Noise - Commercial           43623
Derelict Vehicle             21419
Noise - Vehicle              19122
Animal Abuse                 10500
Traffic                       5161
Vending                       4162
Noise - Park                  3994
Drinking                      1399
Noise - House of Worship      1059
Posting Advertisement          678
Disorderly Youth               314
Graffiti                       157
Name: Complaint Type, dtype: int64
```

In [48]:

```
df['Complaint Type'].value_counts().plot(kind="bar",figsize=(10,6),title="Most Common Complaints"
```

Out[48]:

```
<Axes: title={'center': 'Most Common Complaints'}>
```



In [49]:

```
#Find the top 10 types of complaints
```

In [50]:

```
df['Complaint Type'].value_counts().sort_values(ascending=False)[:10]
```

Out[50]:

```
Blocked Driveway           100455
Illegal Parking             91057
Noise - Street/Sidewalk     50791
Noise - Commercial          43623
Derelict Vehicle            21419
Noise - Vehicle             19122
Animal Abuse                10500
Traffic                      5161
Vending                      4162
Noise - Park                 3994
Name: Complaint Type, dtype: int64
```

In [51]:

```
#Display the types of complaints in each city in a separate dataset

df2=df.groupby(['City','Complaint Type']).size().unstack().fillna(0)
```

In [52]:

```
df2
```

Out[52]:

Complaint Type	Animal Abuse	Blocked Driveway	Derelict Vehicle	Disorderly Youth	Drinking	Graffiti	Illegal Parking	Noise - Commercial	Noise - House of Worship	Noise - Park	Stre
City											
ARVERNE	46.0	50.0	32.0	2.0	1.0	1.0	62.0	2.0	14.0	2.0	
ASTORIA	170.0	3436.0	426.0	5.0	43.0	4.0	1337.0	1640.0	21.0	64.0	
Astoria	0.0	159.0	14.0	0.0	0.0	0.0	277.0	310.0	0.0	0.0	
BAYSIDE	53.0	513.0	231.0	2.0	1.0	3.0	635.0	47.0	3.0	3.0	
BELLEROSE	15.0	138.0	120.0	2.0	1.0	0.0	131.0	38.0	1.0	1.0	
BREEZY POINT	2.0	3.0	3.0	0.0	1.0	0.0	16.0	4.0	0.0	0.0	
BRONX	1966.0	17048.0	2308.0	66.0	205.0	15.0	9853.0	2941.0	90.0	523.0	

In [ ]:

```
```

In [30]:

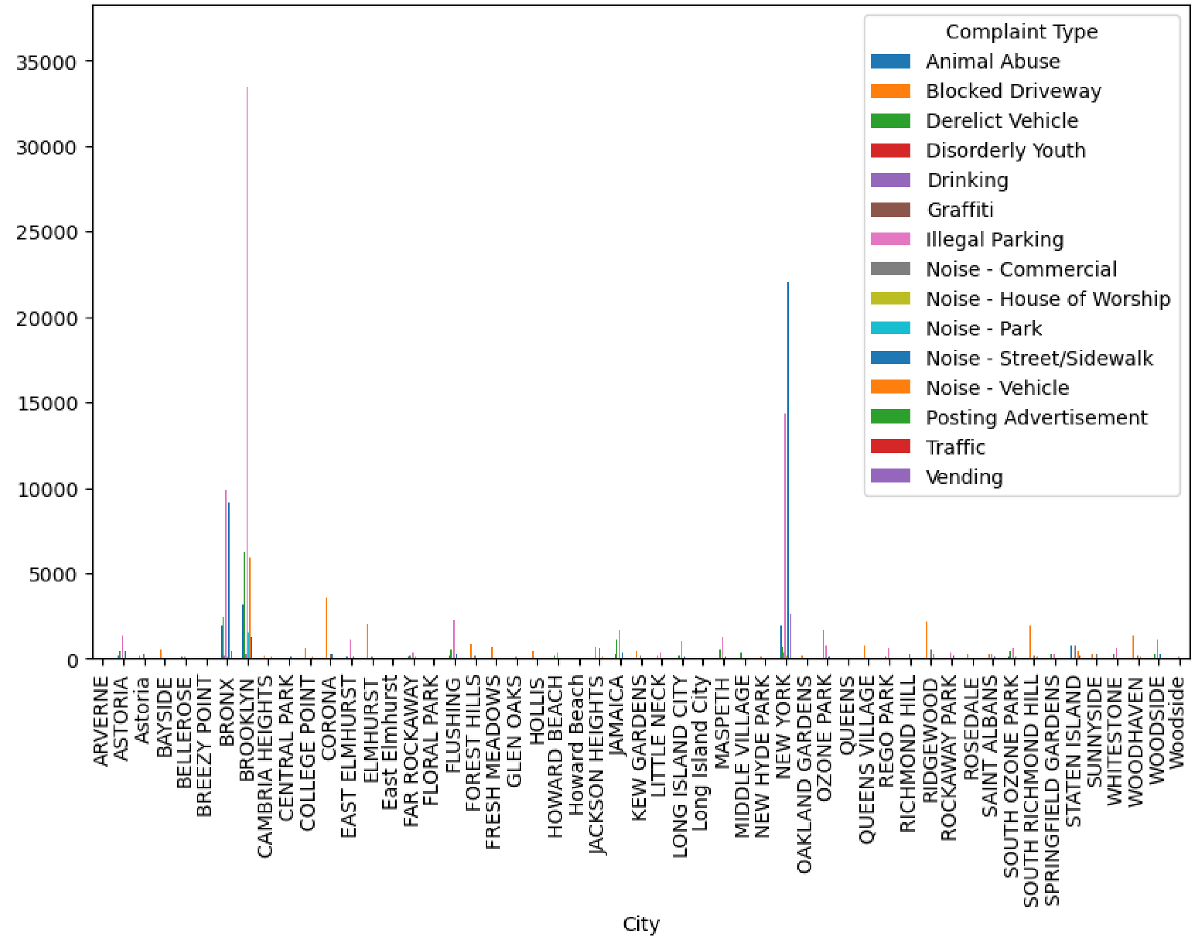
```
# Visualize the major types of complaints in each city
```

In [92]:

```
df.groupby(['City','Complaint Type']).size().unstack().fillna(0).plot(kind='bar',figsize=(10,6))
```

Out[92]:

```
<Axes: xlabel='City'>
```



In [ ]:

```

```

In [ ]:

```

```

In [54]:

```
#Check if the average response time across various types of complaints
```

In [33]:

```
df['Resolution_Time']=(df['Closed Date']-df['Created Date']).dt.days
```

In [56]:

```
(df['Closed Date']-df['Created Date']).dt.days
```

Out[56]:

```
0         0
1         0
2         0
3         0
4         0
         ..
364553    0
364554    0
364555    0
364556    0
364557    0
Length: 353891, dtype: int64
```

In [57]:

```
df['Resolution_Time'].value_counts()
```

Out[57]:

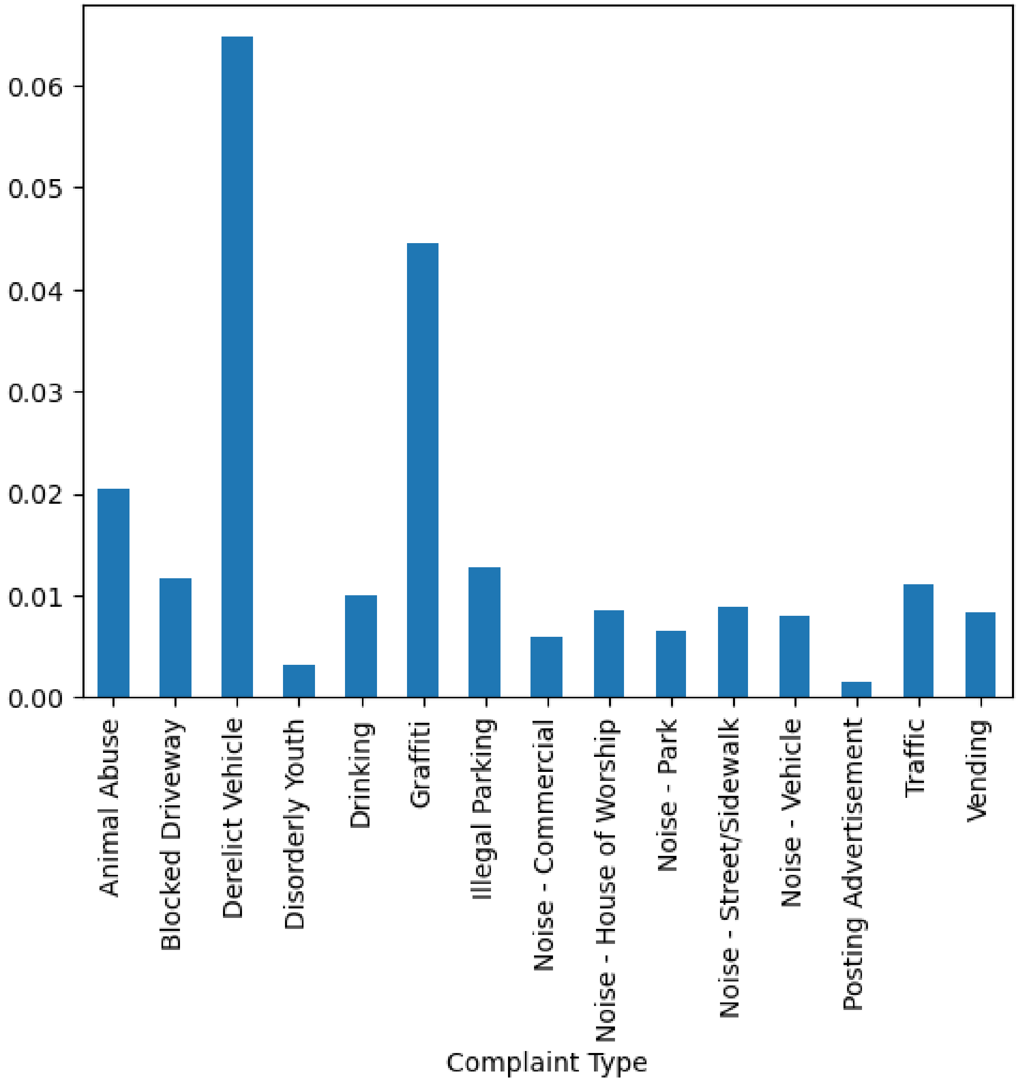
```
0     350061
1       3172
2        445
3        126
5         34
4         30
6         13
8          3
24         3
9          2
21         1
7          1
Name: Resolution_Time, dtype: int64
```

In [58]:

```
df.groupby('Complaint Type')['Resolution_Time'].mean().plot(kind='bar')
```

Out[58]:

```
<Axes: xlabel='Complaint Type'>
```



In [ ]:

In [ ]:

In [ ]: