Data Science with Python Project

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

uploading the data file

```
data=pd.read_csv('311_Service_Requests_from_2010_to_Present.csv')
```

```
/usr/local/lib/python3.7/dist-packages/IPython/core/interactiveshell.py:3326: DtypeWarning: Columns (48,49) have mixed types.Sp
  exec(code_obj, self.user_global_ns, self.user_ns)
```

```
data.info()
```

```
Data columns (total 53 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   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  object
 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: float64(10), int64(1), object(42)
```

```
memory usage: 147.4+ MB
```

```
#Identify the shape of the dataset(Qus-1)
data.shape
```

```
(364558, 53)
```

```
#Identify variables with null values(Qus-2)
data.isnull().any()
```

```
Unique Key                      False
Created Date                    False
Closed Date                      True
Agency                          False
Agency Name                     False
Complaint Type                  False
Descriptor                       True
Location Type                    True
Incident Zip                     True
Incident Address                 True
Street Name                      True
Cross Street 1                   True
Cross Street 2                   True
Intersection Street 1            True
Intersection Street 2            True
Address Type                     True
City                             True
Landmark                         True
Facility Type                    True
Status                          False
Due Date                         True
Resolution Description          False
Resolution Action Updated Date   True
Community Board                 False
Borough                         False
X Coordinate (State Plane)       True
Y Coordinate (State Plane)       True
Park Facility Name              False
Park Borough                    False
```

```
School Name                     False
School Number                   False
School Region                    True
School Code                      True
School Phone Number             False
School Address                  False
School City                     False
School State                    False
School Zip                       True
School Not Found                False
School or Citywide Complaint     True
Vehicle Type                     True
Taxi Company Borough             True
Taxi Pick Up Location            True
Bridge Highway Name              True
Bridge Highway Direction         True
Road Ramp                        True
Bridge Highway Segment           True
Garage Lot Name                  True
Ferry Direction                  True
Ferry Terminal Name              True
Latitude                         True
Longitude                        True
Location                         True
dtype: bool
```

```
#Utilize missing value treatment
#finfing the percentage of missing values
round(data.isnull().sum()/len(data.index),1)*100
```

```
Unique Key                      0.0
Created Date                    0.0
Closed Date                     0.0
Agency                          0.0
Agency Name                     0.0
Complaint Type                  0.0
Descriptor                      0.0
Location Type                   0.0
Incident Zip                    0.0
```

```
Incident Address                  10.0
Street Name                       10.0
Cross Street 1                    20.0
Cross Street 2                    20.0
Intersection Street 1             90.0
Intersection Street 2             90.0
Address Type                       0.0
City                               0.0
Landmark                         100.0
Facility Type                      0.0
Status                             0.0
Due Date                           0.0
Resolution Description             0.0
Resolution Action Updated Date     0.0
Community Board                    0.0
Borough                            0.0
X Coordinate (State Plane)         0.0
Y Coordinate (State Plane)         0.0
Park Facility Name                 0.0
Park Borough                       0.0
School Name                        0.0
School Number                      0.0
School Region                      0.0
School Code                        0.0
School Phone Number                0.0
School Address                     0.0
School City                        0.0
School State                       0.0
School Zip                         0.0
School Not Found                   0.0
School or Citywide Complaint     100.0
Vehicle Type                     100.0
Taxi Company Borough             100.0
Taxi Pick Up Location            100.0
Bridge Highway Name              100.0
Bridge Highway Direction         100.0
Road Ramp                        100.0
Bridge Highway Segment           100.0
Garage Lot Name                  100.0
Ferry Direction                  100.0
Ferry Terminal Name              100.0
```

```
Latitude                        0.0
Longitude                       0.0
Location                        0.0
dtype: float64
```

```
#removing columns with more than 50% missing values
missing_columns=data.columns[100*(data.isnull().sum()/len(data.index))>50]
print(missing_columns)
```

```
Index(['Intersection Street 1', 'Intersection Street 2', 'Landmark',
       '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'],
      dtype='object')
```

```
data_1=data.drop(missing_columns,axis=1)
print(data_1)
```

```
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
```

```
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]
```

```
#Finding the percentage of missing the values in each column after removing 50%
round(data_1.isnull().sum()/len(data_1.index),1)*100
```

```
Unique Key                        0.0
Created Date                      0.0
Closed Date                       0.0
Agency                            0.0
Agency Name                       0.0
Complaint Type                    0.0
Descriptor                        0.0
Location Type                     0.0
Incident Zip                      0.0
Incident Address                 10.0
Street Name                      10.0
Cross Street 1                   20.0
Cross Street 2                   20.0
Address Type                      0.0
City                              0.0
Facility Type                     0.0
Status                            0.0
Due Date                          0.0
Resolution Description            0.0
Resolution Action Updated Date    0.0
Community Board                   0.0
Borough                           0.0
X Coordinate (State Plane)        0.0
Y Coordinate (State Plane)        0.0
Park Facility Name                0.0
Park Borough                      0.0
School Name                       0.0
School Number                     0.0
School Region                     0.0
School Code                       0.0
School Phone Number               0.0
School Address                    0.0
School City                       0.0
School State                      0.0
School Zip                        0.0
School Not Found                  0.0
Latitude                          0.0
Longitude                         0.0
```

```
Location                        0.0
dtype: float64
```

Analyze the date column and remove the entries if it is incorrect timeline

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

```
data['Request_Closing_Time']=data['Closed Date'].values-data['Created Date'].values
data['Request_Closing_Time_mins']=data['Request_Closing_Time']/np.timedelta64(1,'m')
```

```
#Verifying Created date column is having any null Values
data['Created Date'].isnull().sum()
```

```
0
```

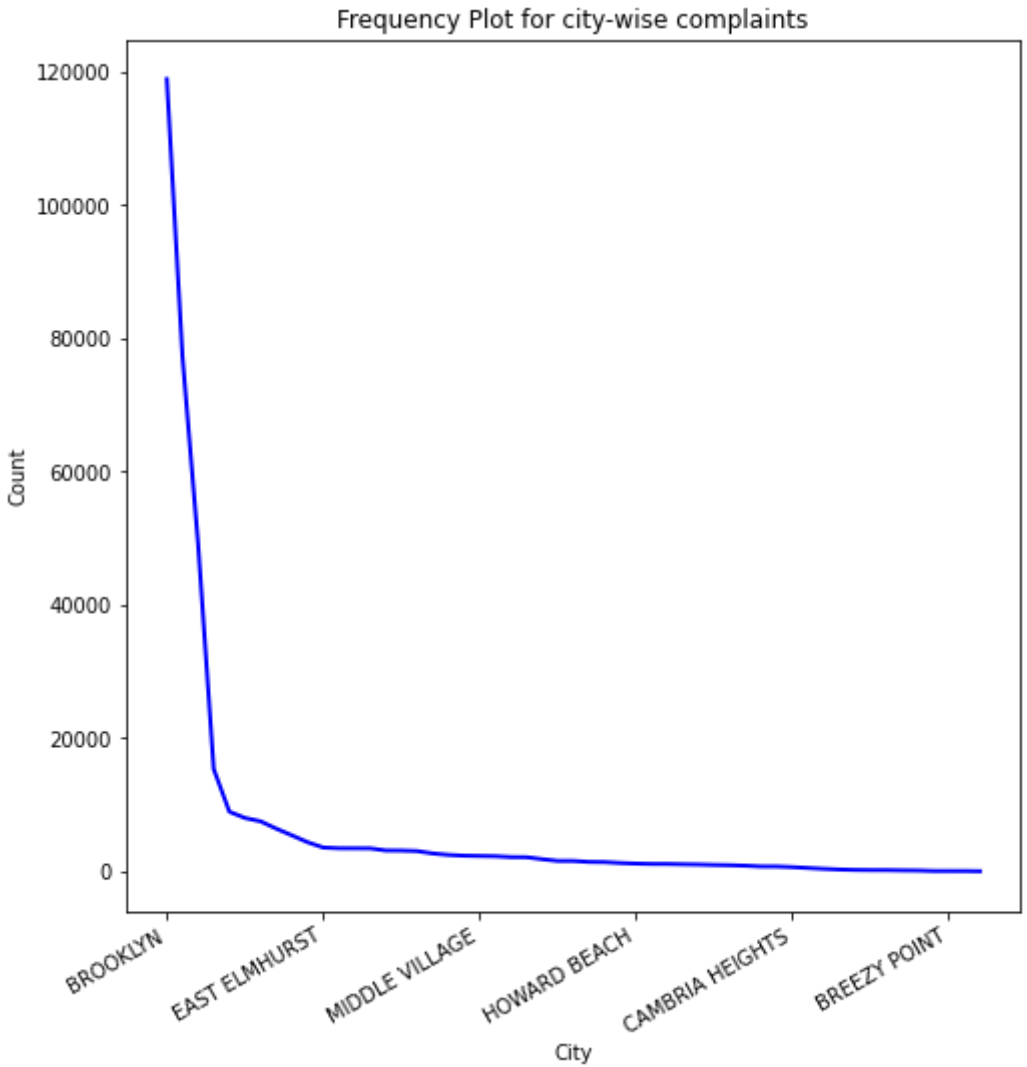
```
#Verifying Closed date column is having any null values
data['Closed Date'].isnull().sum()
```

```
2381
```

```
#Dropping all the null value column and verfying
data.dropna(subset=['Closed Date'],inplace=True)
data['Closed Date'].isnull().sum()
```

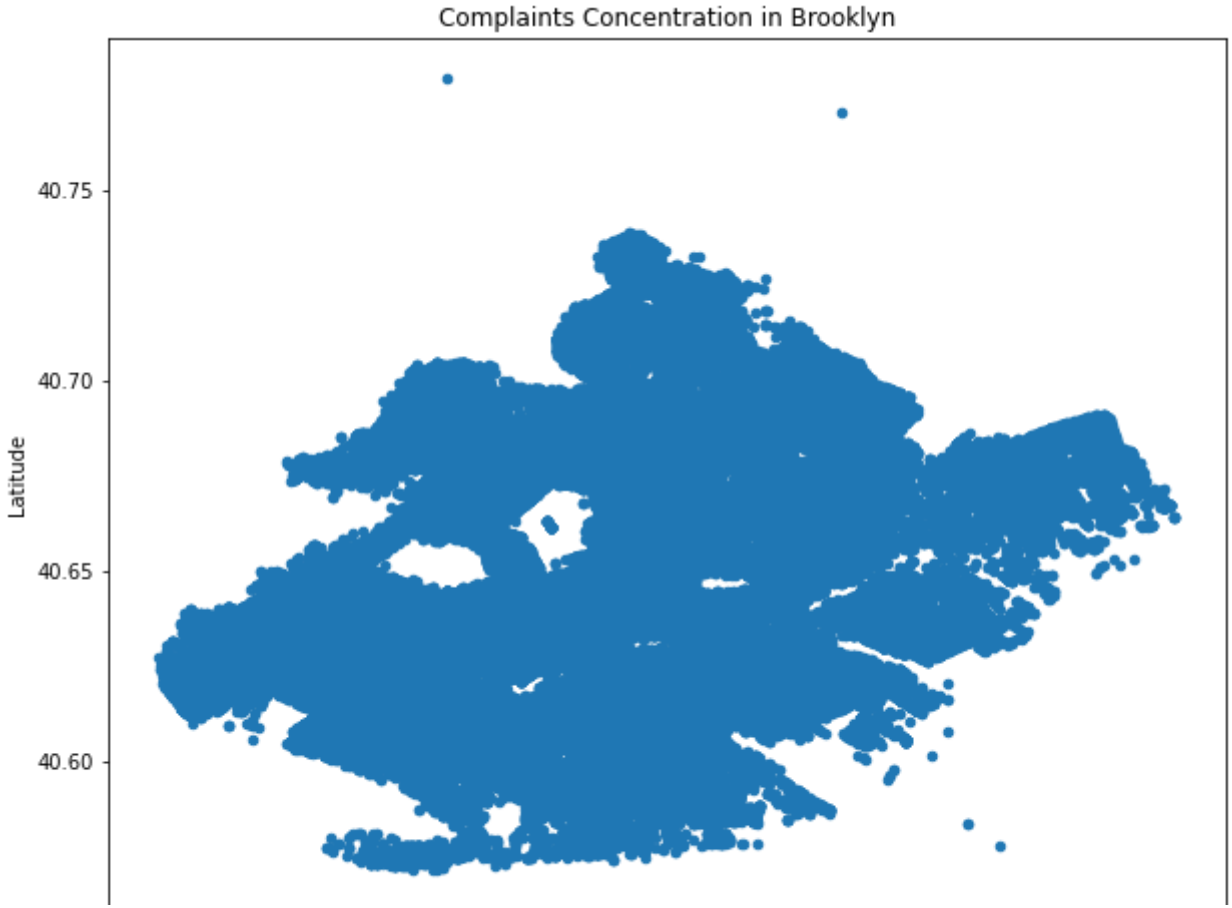
```
0
```

```
#Draw a frequency plot for city-wise complaints
fig=data['City'].value_counts().plot(xlabel='City',ylabel='Count', title='Frequency Plot for city-wise complaints',linewidth=2,marke
plt.setp(fig.get_xticklabels(), rotation=30, horizontalalignment='right')
plt.show()
```



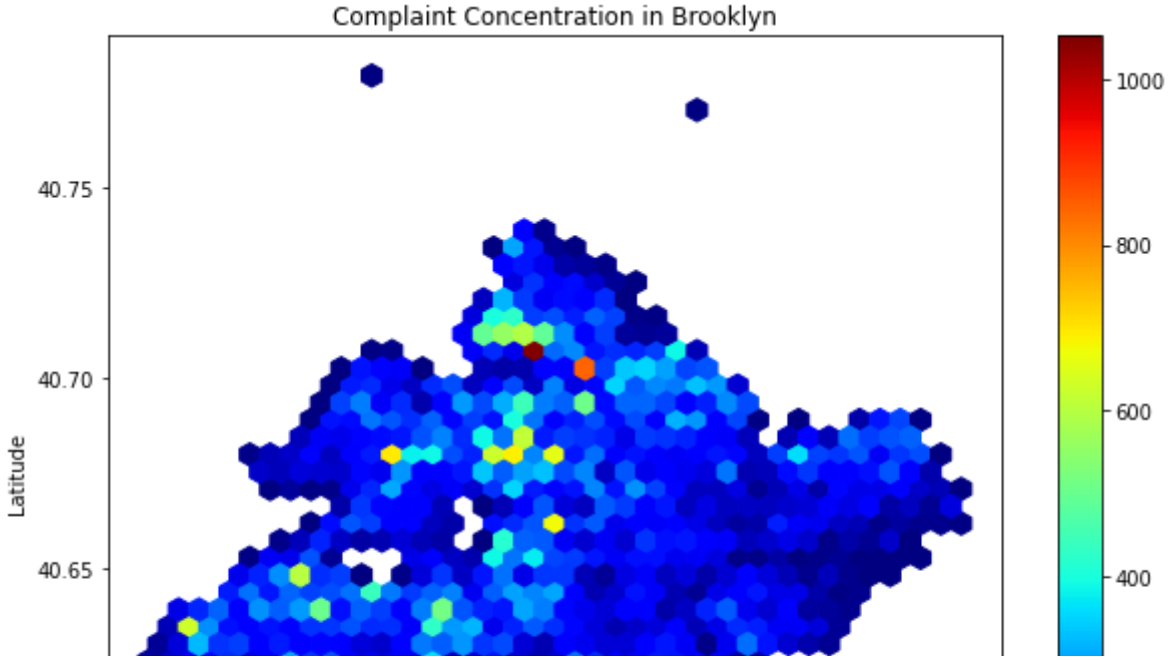

```
#Draw scatter and hexbin plots for complaint concentration across Brooklyn
data_bkn = data.loc[data['City'] == 'BROOKLYN']
data_bkn[['Longitude', 'Latitude']].plot(kind = 'scatter', x='Longitude', y='Latitude', title = 'Complaints Concentration in Brooklyn
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f4398b5a310>
```




```
data_bkn[['Longitude', 'Latitude']].plot(kind = 'hexbin', x='Longitude', y='Latitude', gridsize=40,
    colormap = 'jet', mincnt=1, title = 'Complaint Concentration in Brooklyn', figsize = (10, 8))
```

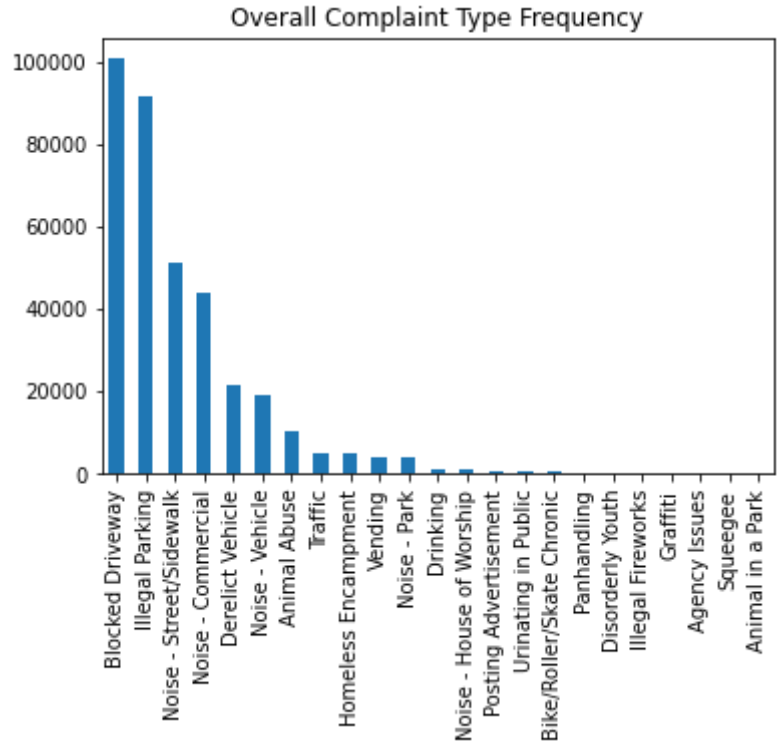
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f76b4575590>
```




```
#Major Complaint Type
data[['Complaint Type','City']]
```

	Complaint Type	City
0	Noise - Street/Sidewalk	NEW YORK
1	Blocked Driveway	ASTORIA
2	Blocked Driveway	BRONX

```
#Plot a bar graph of count vs. complaint types
data['Complaint Type'].value_counts().plot(kind = 'bar', title = 'Overall Complaint Type Frequency');
```




```
#Find the top 10 types of complaints
data['Complaint Type'].value_counts().head(10)
```

```
Blocked Driveway           100624
Illegal Parking             91716
```

```
Noise - Street/Sidewalk     51139
Noise - Commercial          43751
Derelict Vehicle            21518
Noise - Vehicle             19301
Animal Abuse                10530
Traffic                      5196
Homeless Encampment          4879
Vending                      4185
Name: Complaint Type, dtype: int64
```

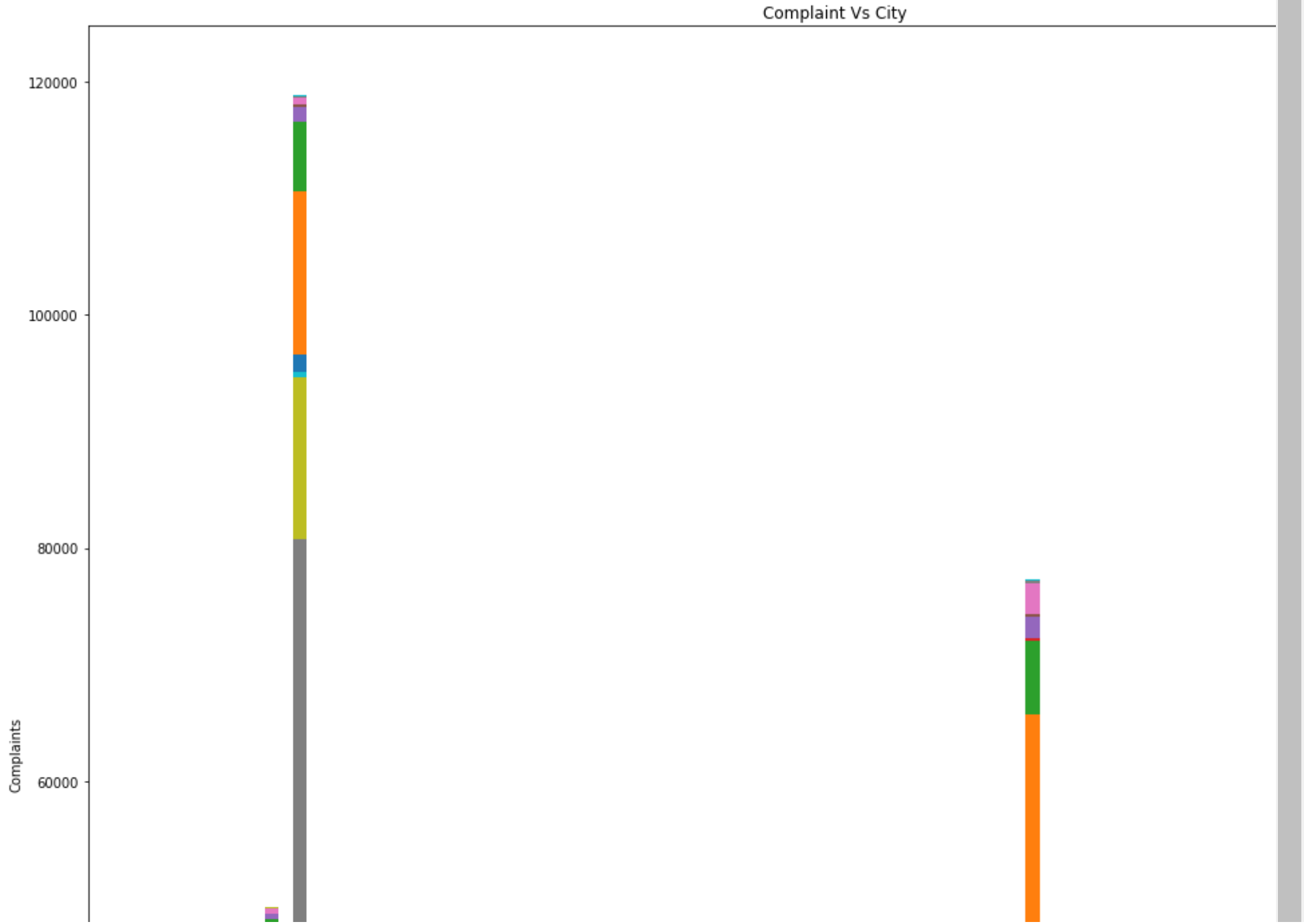
```
#Display the types of complaints in each city in a separate dataset
data_2=data.groupby(['City','Complaint Type']).size().unstack().fillna(0)
print(data_2)
```

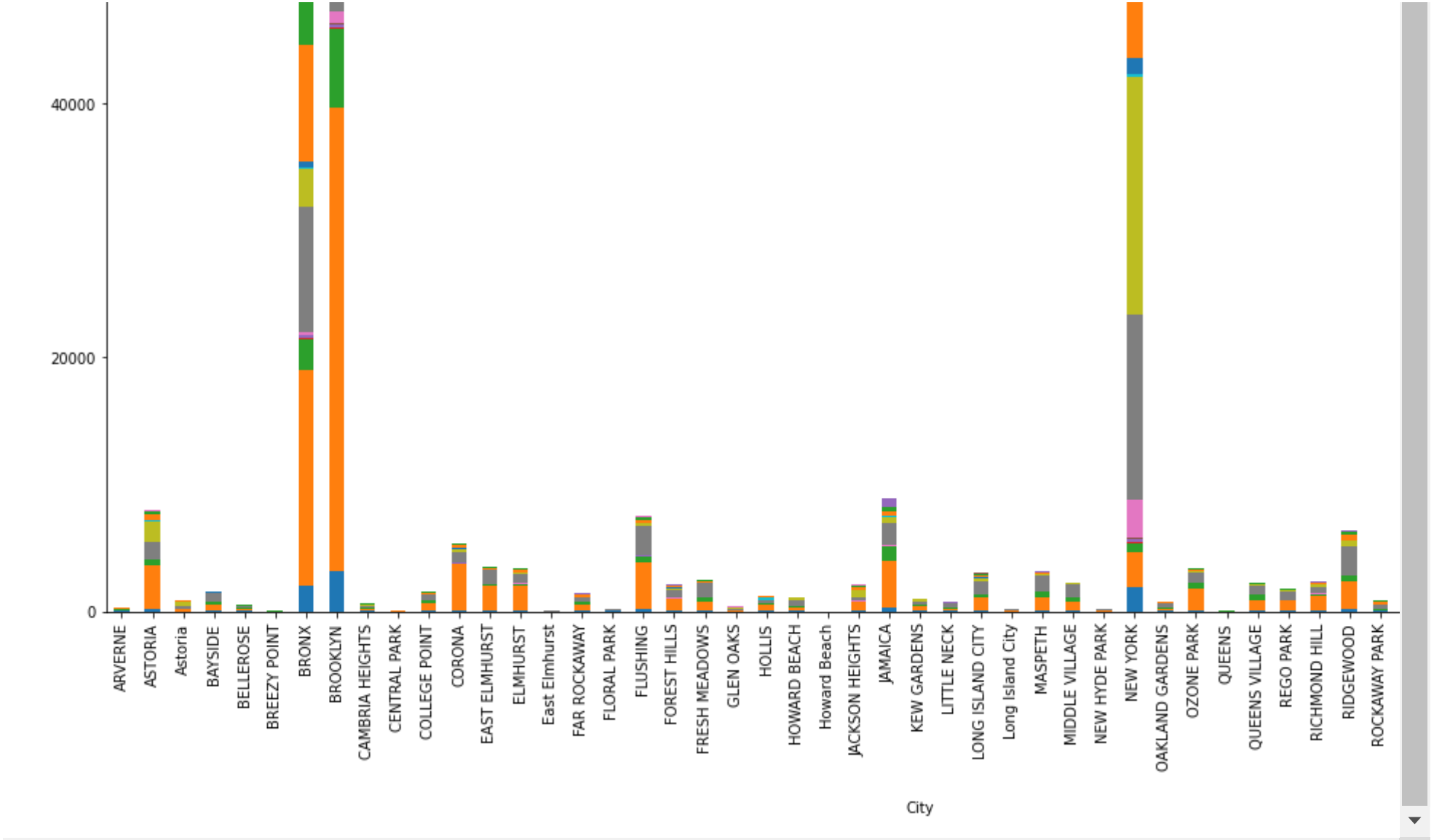
```
MIDDLE VILLAGE               0.0     14.0                0.0      0.0
NEW HYDE PARK                0.0      0.0                0.0      0.0
NEW YORK                   206.0   1769.0              264.0   2638.0
OAKLAND GARDENS              0.0      6.0                0.0      2.0
OZONE PARK                   7.0     21.0                4.0      1.0
QUEENS                       0.0      2.0                1.0      0.0
QUEENS VILLAGE               1.0     27.0                5.0      2.0
REGO PARK                    0.0     16.0                1.0      3.0
RICHMOND HILL                0.0      8.0                5.0     15.0
RIDGEWOOD                    0.0     50.0                9.0      9.0
ROCKAWAY PARK                0.0      7.0                1.0      2.0
ROSEDALE                     0.0     25.0                0.0     19.0
SAINT ALBANS                 0.0     14.0                1.0      2.0
SOUTH OZONE PARK             0.0     36.0                2.0      5.0
SOUTH RICHMOND HILL          0.0     12.0                1.0     24.0
SPRINGFIELD GARDENS          2.0     12.0                3.0      1.0
STATEN ISLAND               13.0    229.0               19.0     25.0
SUNNYSIDE                    0.0     17.0                2.0     15.0
WHITESTONE                   0.0     32.0                0.0      1.0
WOODHAVEN                    1.0      7.0                2.0      6.0
WOODSIDE                     0.0     45.0                8.0     15.0
Woodside                     0.0      0.0                0.0      0.0

Complaint Type       Bike/Roller/Skate Chronic  Illegal Fireworks  \
City
ARVERNE                                    0.0                0.0
ASTORIA                                   16.0                4.0
```

Astoria	0.0	0.0
BAYSIDE	0.0	0.0
BELLEROSE	1.0	1.0
BREEZY POINT	0.0	0.0
BRONX	22.0	24.0
BROOKLYN	124.0	61.0
CAMBRIA HEIGHTS	0.0	1.0
CENTRAL PARK	0.0	0.0
COLLEGE POINT	0.0	0.0
CORONA	0.0	0.0
EAST ELMHURST	1.0	0.0
ELMHURST	2.0	1.0
East Elmhurst	0.0	0.0
FAR ROCKAWAY	0.0	0.0
FLORAL PARK	0.0	0.0
FLUSHING	3.0	2.0
FOREST HILLS	6.0	1.0
FRESH MEADOWS	0.0	0.0
GLEN OAKS	0.0	0.0
HOLLIS	0.0	0.0
HOWARD BEACH	1.0	4.0
Howard Beach	0.0	0.0
JACKSON HEIGHTS	2.0	1.0
JAMAICA	3.0	4.0
KEW GARDENS	0.0	0.0
LITTLE NECK	0.0	0.0
LONG ISLAND CITY	3.0	0.0
Long Island City	0.0	0.0
MASPETH	1.0	1.0
MIDDLE VILLAGE	1.0	0.0
NEW HYDE PARK	0.0	0.0
NEW YORK	254.0	38.0

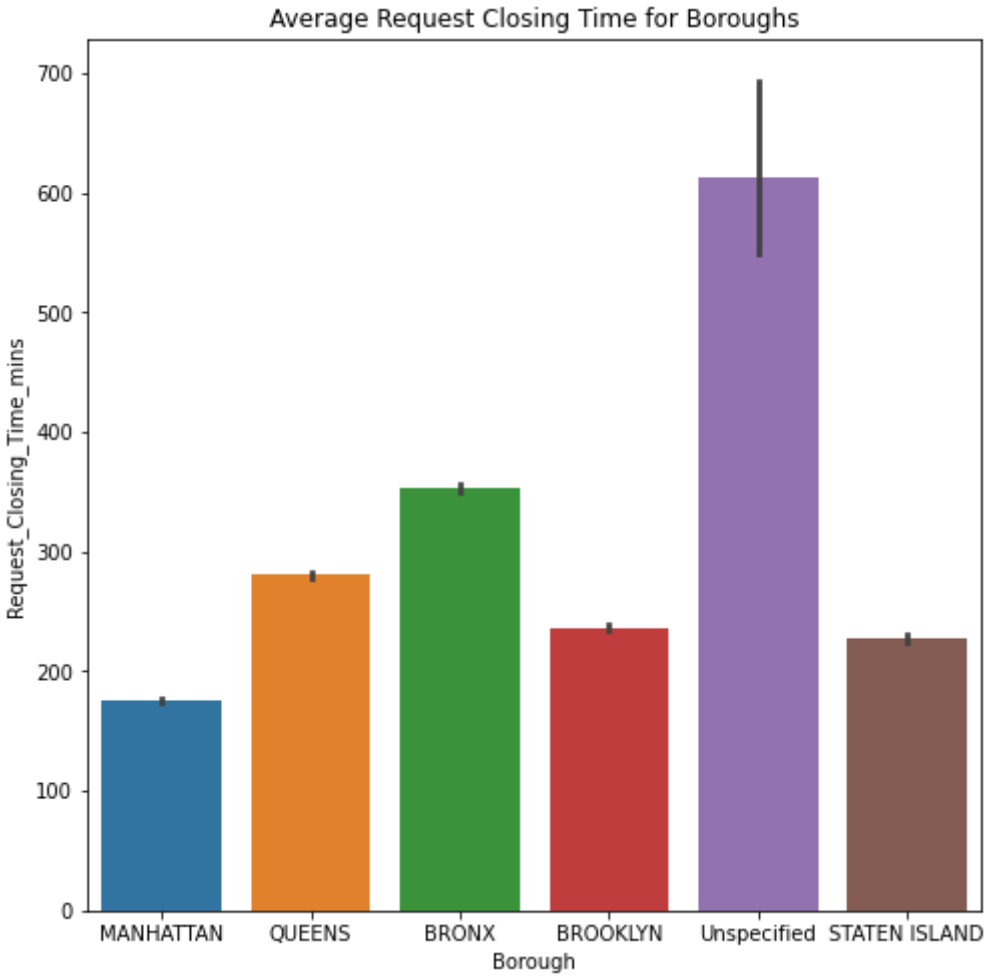
```
#Visualize the major types of complaints in each city
data_2.plot.bar(figsize=(20,20), stacked=True)
plt.xlabel('City')
plt.ylabel('Complaints')
plt.title('Complaint Vs City')
```

```
Text(0.5, 1.0, 'Complaint Vs City')
```






```
#Check if the average response time across various types of complaints
plt.figure(figsize=(8,8))
sns.barplot(x='Borough', y='Request_Closing_Time_mins', data=data)
plt.title('Average Request Closing Time for Boroughs')
plt.show()
```


Average Request Closing Time for Boroughs
Request_Closing_Time_mins
0
100
200
300
400
500
600
700
MANHATTAN
QUEENS
BRONX
BROOKLYN
Unspecified
STATEN ISLAND
Borough

Colab paid products - Cancel contracts here

✓ 7s completed at 4:39 PM