

Identify the shape of the dataset

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

```
(364558, 53)
```

Identify variables with null values

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

| | |
|--------------------------------|--------|
| 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 |

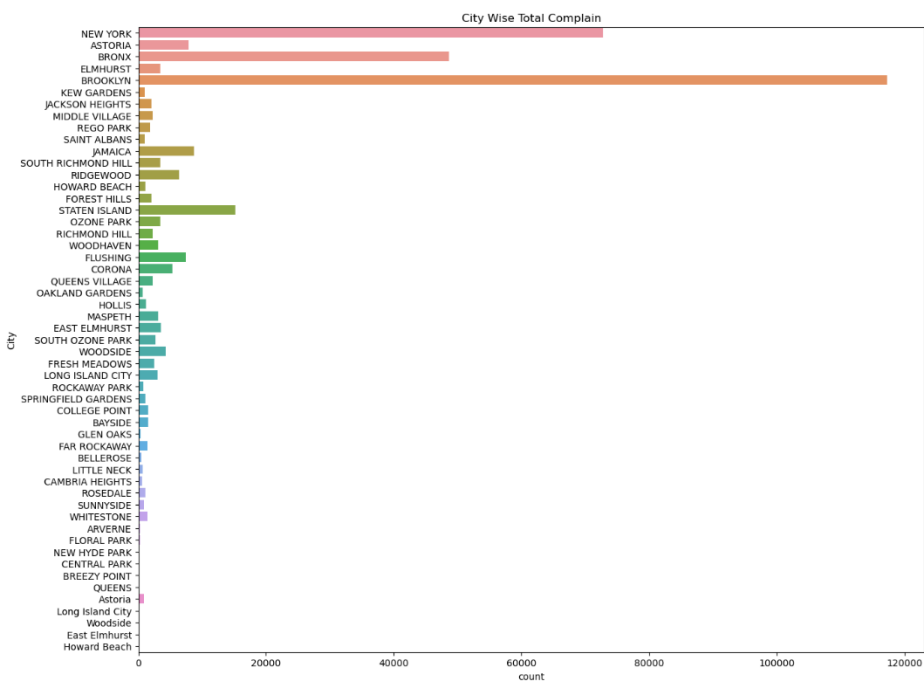
Utilize missing value treatment

```
df.dropna(inplace=True)
df.isnull().sum()/len(df)*100
```

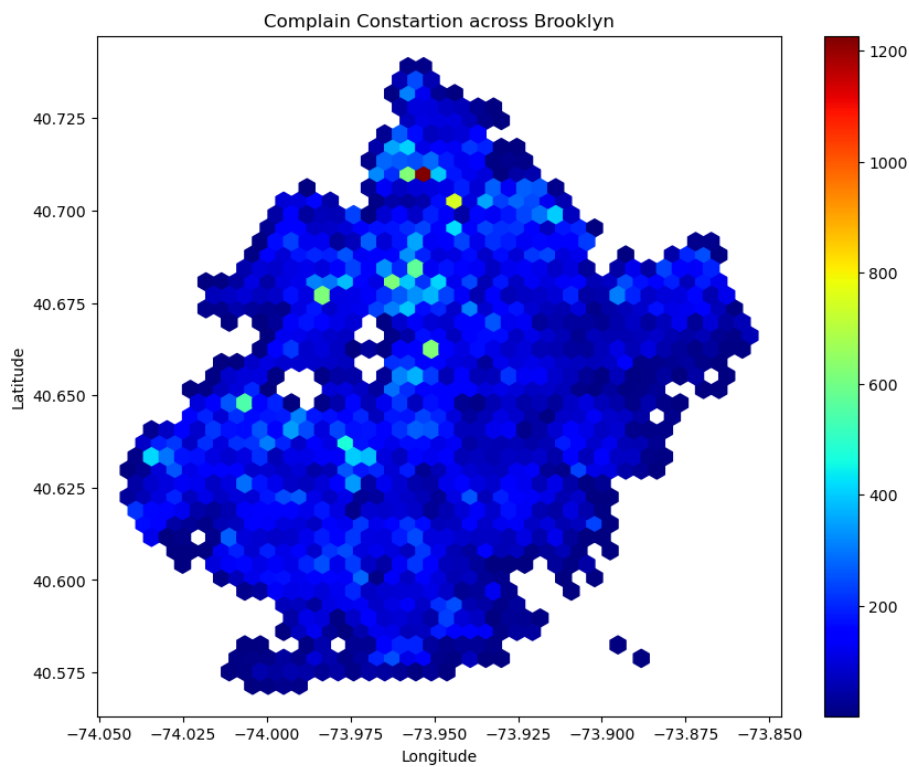
| | |
|------------------------|-----|
| Unique Key | 0.0 |
| Created Date | 0.0 |
| Closed Date | 0.0 |
| Agency | 0.0 |
| Complaint Type | 0.0 |
| Descriptor | 0.0 |
| Location Type | 0.0 |
| Incident Zip | 0.0 |
| City | 0.0 |
| Status | 0.0 |
| Resolution Description | 0.0 |
| Borough | 0.0 |
| Latitude | 0.0 |
| Longitude | 0.0 |
| Location | 0.0 |

dtype: float64

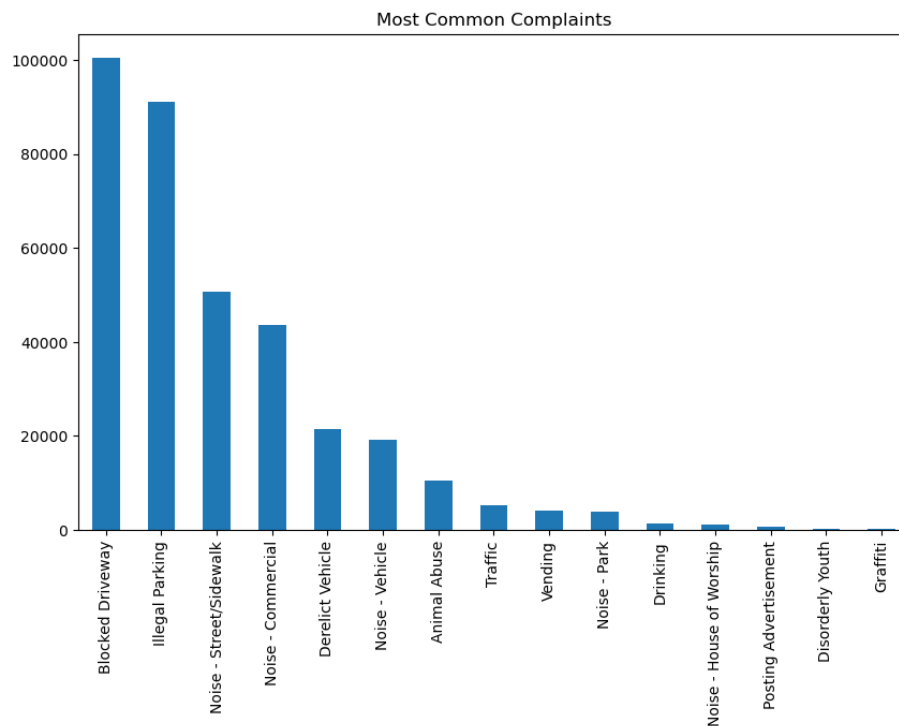
Draw a frequency plot for city-wise complaints



Draw scatter and hexbin plots for complaint concentration across Brooklyn



Plot a bar graph of count vs. complaint types



Find the top 10 types of complaints

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

```
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
```

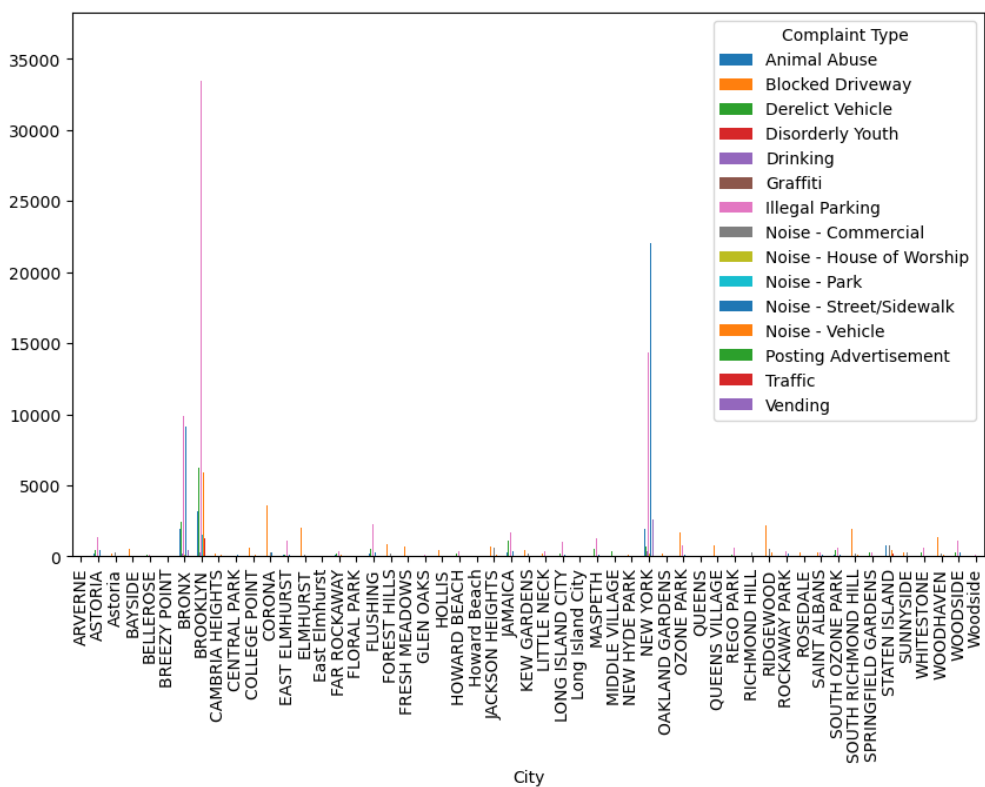
Display the types of complaints in each city in a separate dataset

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

df2

| | Complaint Type | Animal Abuse | Blocked Driveway | Derelict Vehicle | Disorderly Youth | Drinking | Graffiti | Illegal Parking | Noise - Commercial | Noise - House of Worship | Noise - Park | Noise - Street/Sidewalk | Noise - Vehicle | Posting Advertisement | Traffic |
|------|----------------|--------------|------------------|------------------|------------------|----------|----------|-----------------|--------------------|--------------------------|--------------|-------------------------|-----------------|-----------------------|---------|
| City | | | | | | | | | | | | | | | |
| | ARVERNE | 46.0 | 50.0 | 32.0 | 2.0 | 1.0 | 1.0 | 62.0 | 2.0 | 14.0 | 2.0 | 29.0 | 9.0 | 0.0 | 1 |
| | ASTORIA | 170.0 | 3436.0 | 426.0 | 5.0 | 43.0 | 4.0 | 1337.0 | 1640.0 | 21.0 | 64.0 | 408.0 | 236.0 | 3.0 | 60 |
| | Astoria | 0.0 | 159.0 | 14.0 | 0.0 | 0.0 | 0.0 | 277.0 | 310.0 | 0.0 | 0.0 | 145.0 | 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 | 17.0 | 24.0 | 0.0 | 9 |
| | BELLEROSE | 15.0 | 138.0 | 120.0 | 2.0 | 1.0 | 0.0 | 131.0 | 38.0 | 1.0 | 1.0 | 13.0 | 11.0 | 1.0 | 9 |
| | BREEZY POINT | 2.0 | 3.0 | 3.0 | 0.0 | 1.0 | 0.0 | 16.0 | 4.0 | 0.0 | 0.0 | 1.0 | 1.0 | 0.0 | 0 |
| | BRONX | 1966.0 | 17048.0 | 2398.0 | 66.0 | 205.0 | 15.0 | 9853.0 | 2941.0 | 90.0 | 523.0 | 9118.0 | 3544.0 | 17.0 | 425 |
| | BROOKLYN | 3185.0 | 36414.0 | 6242.0 | 79.0 | 291.0 | 60.0 | 33446.0 | 13847.0 | 387.0 | 1557.0 | 13943.0 | 5932.0 | 58.0 | 1253 |

Visualize the major types of complaints in each city



Check if the average response time across various types of complaints

