In [2]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:
```python
df = pd.read_csv(r"C:\Users\HP\Desktop\project\311_Service_Requests_from_2010_to_Present.csv")
```

```
C:\ANACONDA\lib\site-packages\IPython\core\interactiveshell.py:3165: DtypeWarning: Columns (48,49) have mixed types.Specify dtype option on import or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

In [3]:
```python
df.head()
```

Out[3]:

	Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Bridge Highway Name	Bridge Highway Direction	Road Ramp	Bridge Highway Segment	Garage Lot Name	Ferry Direction	Ferry Terminal Name	Latitude	Longitude	
0	32310363	12/31/2015 11:59:45 PM	01/01/2016 12:55:15 AM	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	Street/Sidewalk	10034.0	71 VERMILYEA AVENUE	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	40.865682	-73.923501	(40.86568 -73.92350
1	32309934	12/31/2015 11:59:44 PM	01/01/2016 01:26:57 AM	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	11105.0	27-07 23 AVENUE	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	40.775945	-73.915094	(40.77594 -73.91509
2	32309159	12/31/2015 11:59:29 PM	01/01/2016 04:51:03 AM	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	10458.0	2897 VALENTINE AVENUE	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	40.870325	-73.888525	(40.87032 -73.88852
3	32305098	12/31/2015 11:57:46 PM	01/01/2016 07:43:13 AM	NYPD	New York City Police Department	Illegal Parking	Commercial Overnight Parking	Street/Sidewalk	10461.0	2940 BAISLEY AVENUE	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	40.835994	-73.828379	(40.83599 -73.82837
4	32306529	12/31/2015 11:56:58 PM	01/01/2016 03:24:42 AM	NYPD	New York City Police Department	Illegal Parking	Blocked Sidewalk	Street/Sidewalk	11373.0	87-14 57 ROAD	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	40.733060	-73.874170	(40.73305 -73.87416

5 rows × 53 columns

In [5]:
```python
df['Created Date'] = pd.to_datetime(df['Created Date'])
df['Created Date']
```

Out[5]:
```
0        2015-12-31 23:59:45
1        2015-12-31 23:59:44
2        2015-12-31 23:59:29
3        2015-12-31 23:57:46
4        2015-12-31 23:56:58
                 ...
364553   2015-01-01 00:04:44
364554   2015-01-01 00:04:28
364555   2015-01-01 00:01:30
364556   2015-01-01 00:01:29
364557   2015-01-01 00:00:50
Name: Created Date, Length: 364558, dtype: datetime64[ns]
```

In [6]:
```python
df2=df[['Created Date']]
df2
```

Out[6]:

	Created Date
0	2015-12-31 23:59:45
1	2015-12-31 23:59:44
2	2015-12-31 23:59:29
3	2015-12-31 23:57:46
4	2015-12-31 23:56:58
...	...
364553	2015-01-01 00:04:44
364554	2015-01-01 00:04:28
364555	2015-01-01 00:01:30
364556	2015-01-01 00:01:29
364557	2015-01-01 00:00:50

364558 rows × 1 columns

In [9]:
```python
df2['year'] = df2['Created Date'].dt.year
df2['month'] = df2['Created Date'].dt.month
df2['day'] = df2['Created Date'].dt.day
df2['hour'] = df2['Created Date'].dt.hour
df2['minute'] = df2['Created Date'].dt.minute
df2['second'] = df2['Created Date'].dt.second
df2
```

```
<ipython-input-9-49bc007256e3>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df2['year'] = df2['Created Date'].dt.year
<ipython-input-9-49bc007256e3>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df2['month'] = df2['Created Date'].dt.month
<ipython-input-9-49bc007256e3>:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df2['day'] = df2['Created Date'].dt.day
<ipython-input-9-49bc007256e3>:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df2['hour'] = df2['Created Date'].dt.hour
<ipython-input-9-49bc007256e3>:5: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df2['minute'] = df2['Created Date'].dt.minute
<ipython-input-9-49bc007256e3>:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df2['second'] = df2['Created Date'].dt.second
```

Out[9]:

	Created Date	year	month	day	hour	minute	second
0	2015-12-31 23:59:45	2015	12	31	23	59	45
1	2015-12-31 23:59:44	2015	12	31	23	59	44
2	2015-12-31 23:59:29	2015	12	31	23	59	29
3	2015-12-31 23:57:46	2015	12	31	23	57	46
4	2015-12-31 23:56:58	2015	12	31	23	56	58
...	...	...	...	...	...	...	...
364553	2015-01-01 00:04:44	2015	1	1	0	4	44
364554	2015-01-01 00:04:28	2015	1	1	0	4	28
364555	2015-01-01 00:01:30	2015	1	1	0	1	30
364556	2015-01-01 00:01:29	2015	1	1	0	1	29
364557	2015-01-01 00:00:50	2015	1	1	0	0	50

364558 rows × 7 columns

In [23]:
```python
print(df2['year'] != 2015)
```

```
0         False
1         False
2         False
3         False
4         False
          ...
364553    False
364554    False
364555    False
364556    False
364557    False
Name: year, Length: 364558, dtype: bool
```

In [28]:
```python
yf= (df2['year'] != 2015)
df[yf]
```

Out[28]:

Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Ferry Terminal Name	Latitude	Longitude	Location	year	month	day	hour	minute	second

0 rows × 59 columns

In [29]:
```python
yf= (df2['month'] > 12)
df[yf]
```

Out[29]:

Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Ferry Terminal Name	Latitude	Longitude	Location	year	month	day	hour	minute	second

0 rows × 59 columns

In [30]:
```python
yf= (df2['day'] > 31)
df[yf]
```

Out[30]:

Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Ferry Terminal Name	Latitude	Longitude	Location	year	month	day	hour	minute	second

0 rows × 59 columns

In [31]:
```python
yf= (df2['hour'] > 23)
df[yf]
```

Out[31]:

Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Ferry Terminal Name	Latitude	Longitude	Location	year	month	day	hour	minute	second

0 rows × 59 columns

In [32]:
```python
yf= (df2['minute'] > 60)
df[yf]
```

Out[32]:

Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Ferry Terminal Name	Latitude	Longitude	Location	year	month	day	hour	minute	second

0 rows × 59 columns

In [37]:
```python
yf= (df2['second'] > 60)
df[yf]
```

Out[37]:

Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Ferry Terminal Name	Latitude	Longitude	Location	year	month	day	hour	minute	second

0 rows × 59 columns

In [38]:
```python
df['Closed Date'] = pd.to_datetime(df['Closed Date'])
df['Closed Date']
```

Out[38]:
```
0        2016-01-01 00:55:15
1        2016-01-01 01:26:57
2        2016-01-01 04:51:03
3        2016-01-01 07:43:13
4        2016-01-01 03:24:42
                 ...
364553   2015-01-01 10:22:31
364554   2015-01-01 02:25:02
364555   2015-01-01 00:20:33
364556   2015-01-01 02:42:22
364557   2015-01-01 02:47:50
Name: Closed Date, Length: 364558, dtype: datetime64[ns]
```

In [57]:
```python
df3=df[['Closed Date']]
df3
```

Out[57]:

	Closed Date
0	2016-01-01 00:55:15
1	2016-01-01 01:26:57
2	2016-01-01 04:51:03
3	2016-01-01 07:43:13
4	2016-01-01 03:24:42
...	...
364553	2015-01-01 10:22:31
364554	2015-01-01 02:25:02
364555	2015-01-01 00:20:33
364556	2015-01-01 02:42:22
364557	2015-01-01 02:47:50

364558 rows × 1 columns

In [58]:
```python
df3['year'] = df3['Closed Date'].dt.year
df3['month'] = df3['Closed Date'].dt.month
df3['day'] = df3['Closed Date'].dt.day
df3['hour'] = df3['Closed Date'].dt.hour
df3['minute'] = df3['Closed Date'].dt.minute
df3['second'] = df3['Closed Date'].dt.second
df3
```

```
<ipython-input-58-37b376a16e11>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df3['year'] = df3['Closed Date'].dt.year
<ipython-input-58-37b376a16e11>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df3['month'] = df3['Closed Date'].dt.month
<ipython-input-58-37b376a16e11>:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df3['day'] = df3['Closed Date'].dt.day
<ipython-input-58-37b376a16e11>:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df3['hour'] = df3['Closed Date'].dt.hour
<ipython-input-58-37b376a16e11>:5: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df3['minute'] = df3['Closed Date'].dt.minute
<ipython-input-58-37b376a16e11>:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df3['second'] = df3['Closed Date'].dt.second
```

Out[58]:

	Closed Date	year	month	day	hour	minute	second
0	2016-01-01 00:55:15	2016.0	1.0	1.0	0.0	55.0	15.0
1	2016-01-01 01:26:57	2016.0	1.0	1.0	1.0	26.0	57.0
2	2016-01-01 04:51:03	2016.0	1.0	1.0	4.0	51.0	3.0
3	2016-01-01 07:43:13	2016.0	1.0	1.0	7.0	43.0	13.0
4	2016-01-01 03:24:42	2016.0	1.0	1.0	3.0	24.0	42.0
...	...	...	...	...	...	...	...
364553	2015-01-01 10:22:31	2015.0	1.0	1.0	10.0	22.0	31.0
364554	2015-01-01 02:25:02	2015.0	1.0	1.0	2.0	25.0	2.0
364555	2015-01-01 00:20:33	2015.0	1.0	1.0	0.0	20.0	33.0
364556	2015-01-01 02:42:22	2015.0	1.0	1.0	2.0	42.0	22.0
364557	2015-01-01 02:47:50	2015.0	1.0	1.0	2.0	47.0	50.0

364558 rows × 7 columns

In [61]:
```python
xf= (df3['year'] < 2015)
df[xf]
```

Out[61]:

Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Ferry Terminal Name	Latitude	Longitude	Location	year	month	day	hour	minute	second

0 rows × 59 columns

In [62]:
```python
xf= (df3['month'] > 12)
df[xf]
```

Out[62]:

Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Ferry Terminal Name	Latitude	Longitude	Location	year	month	day	hour	minute	second

0 rows × 59 columns

In [63]:
```python
xf= (df3['day'] > 31)
df[xf]
```

Out[63]:

Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Ferry Terminal Name	Latitude	Longitude	Location	year	month	day	hour	minute	second

0 rows × 59 columns

In [64]:
```python
xf= (df3['hour'] > 23)
df[xf]
```

Out[64]:

Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Ferry Terminal Name	Latitude	Longitude	Location	year	month	day	hour	minute	second

0 rows × 59 columns

In [65]:
```python
xf= (df3['minute'] > 60)
df[xf]
```

Out[65]:

Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Ferry Terminal Name	Latitude	Longitude	Location	year	month	day	hour	minute	second

0 rows × 59 columns

In [67]:
```python
xf= (df3['second'] > 60)
df[xf]
```

Out[67]:

Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Ferry Terminal Name	Latitude	Longitude	Location	year	month	day	hour	minute	second

0 rows × 59 columns

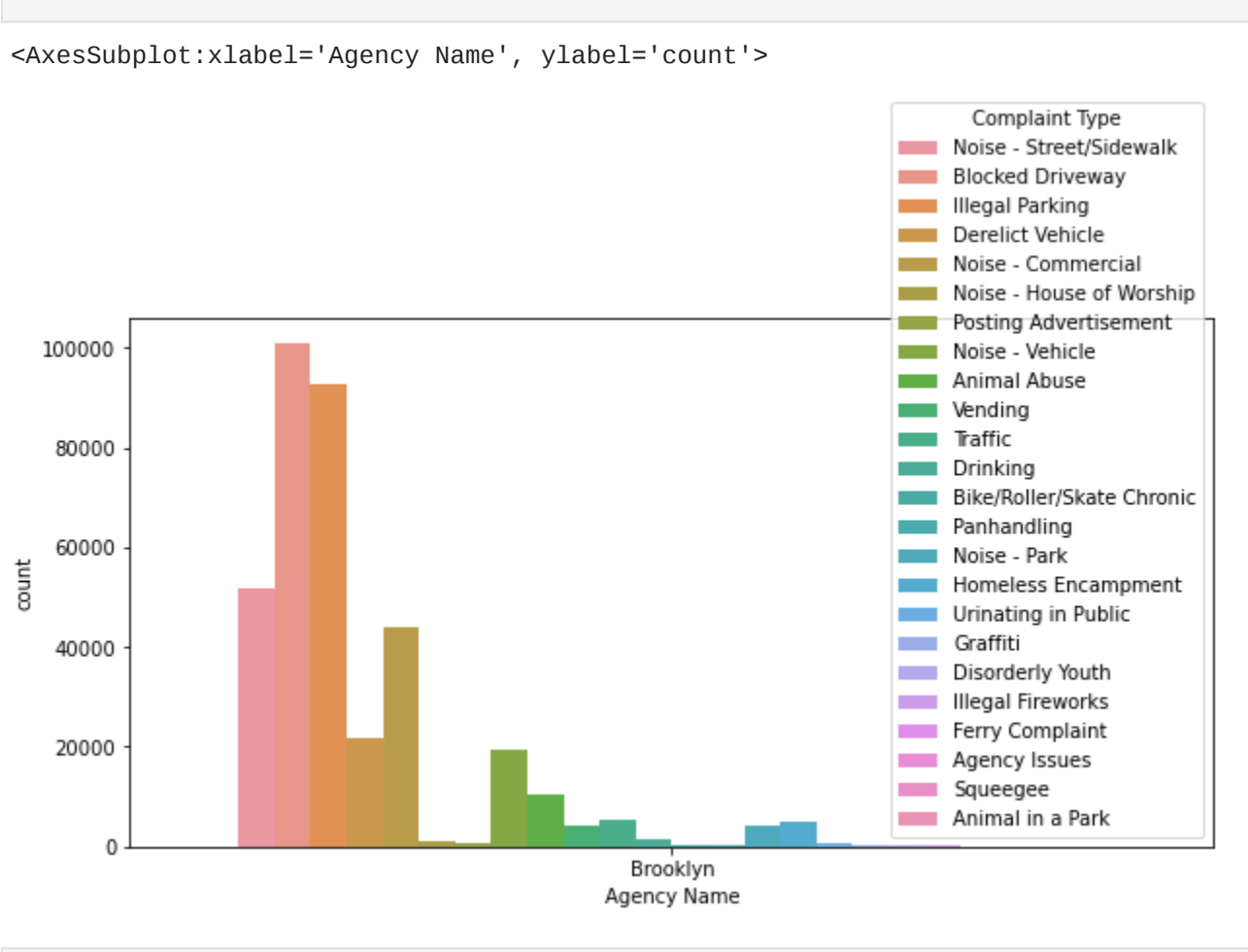
In [86]:
```python
import seaborn as sns
```

In [84]:
```python
df['Agency Name']
```

Out[84]:
```
0         Brooklyn
1         Brooklyn
2         Brooklyn
3         Brooklyn
4         Brooklyn
            ...
364553    Brooklyn
364554    Brooklyn
364555    Brooklyn
364556    Brooklyn
364557    Brooklyn
Name: Agency Name, Length: 364558, dtype: object
```

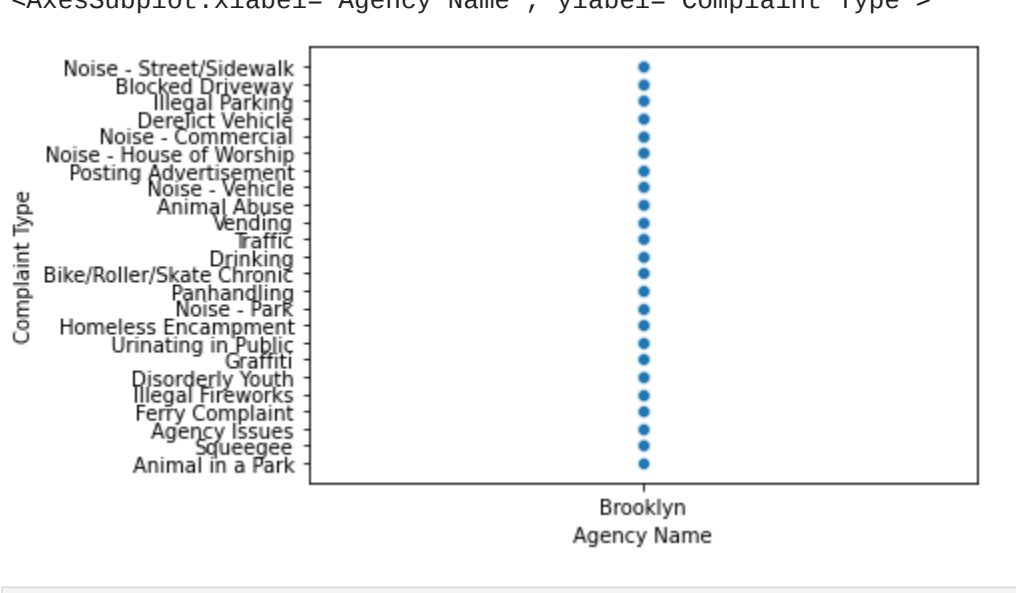
In [94]:
```python
plt.figure(figsize=(10,5))
sns.countplot(x="Agency Name" ,hue= "Complaint Type", data=df)
```

Out[94]: <AxesSubplot:xlabel='Agency Name', ylabel='count'>



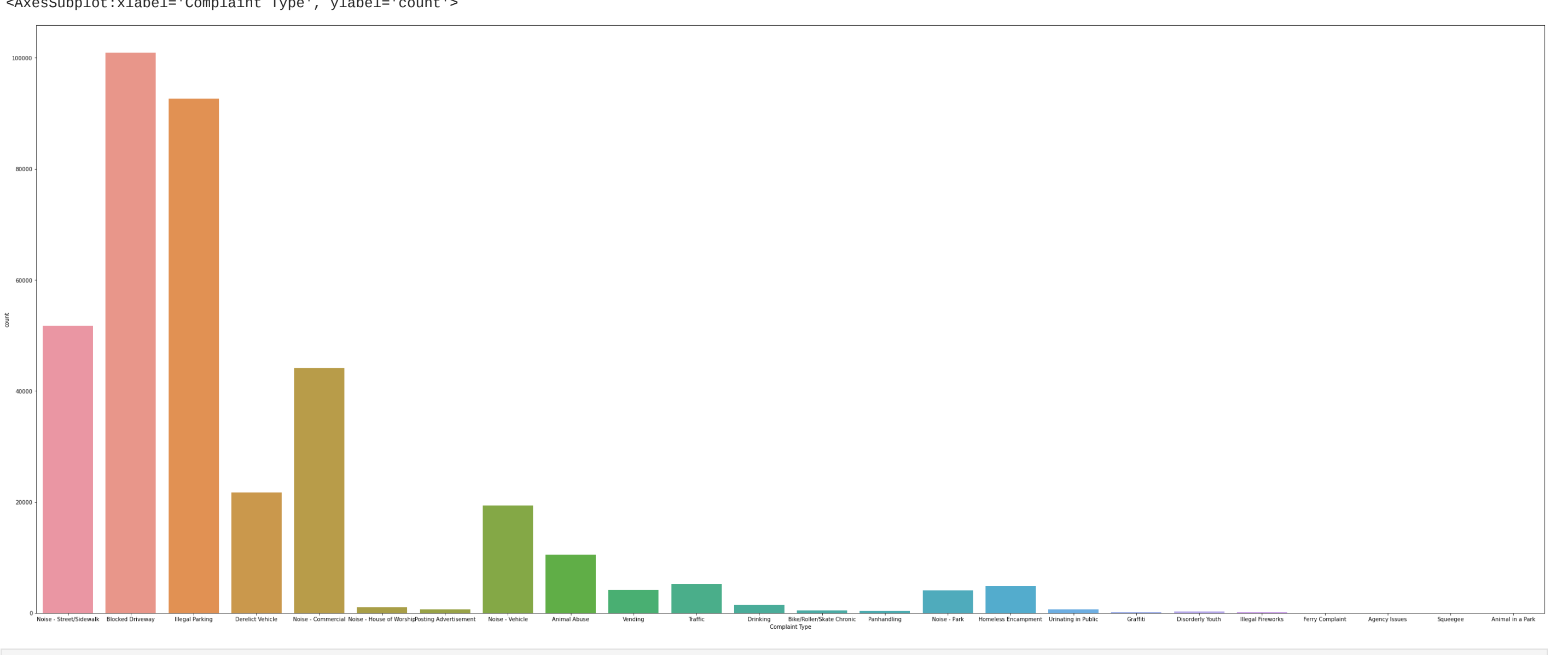
In [97]:
```python
sns.scatterplot(data=df, x="Agency Name", y= "Complaint Type")
```

Out[97]: <AxesSubplot:xlabel='Agency Name', ylabel='Complaint Type'>



In [106]:
```python
plt.figure(figsize=(50,20))
sns.countplot(x= "Complaint Type", data=df)
```

Out[106]: <AxesSubplot:xlabel='Complaint Type', ylabel='count'>



In [108]:
```python
top_complaints= df['Complaint Type'].value_counts()
top_complaints
```

Out[108]:
```
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
Noise - Park                   4109
Drinking                       1409
Noise - House of Worship       1070
Posting Advertisement           681
Urinating in Public             641
Bike/Roller/Skate Chronic       478
Panhandling                     327
Disorderly Youth                315
Illegal Fireworks               172
Graffiti                        157
Agency Issues                     8
Squeegee                          4
Ferry Complaint                   2
Animal in a Park                  1
Name: Complaint Type, dtype: int64
```

In [109]:
```python
top_complaints.head(10)
```

Out[109]:
```
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
```

In [122]:
```python
x=np.array(top_complaints)
x
```

Out[122]:
```
array([100881,  92679,  51692,  44109,  21661,  19352,  10541,   5198,
         4879,   4192,   4109,   1409,   1070,    681,    641,    478,
          327,    315,    172,    157,      8,      4,      2,      1],
      dtype=int64)
```

In [123]:
```python
agen= df['Agency Name'].value_counts()
agen
```

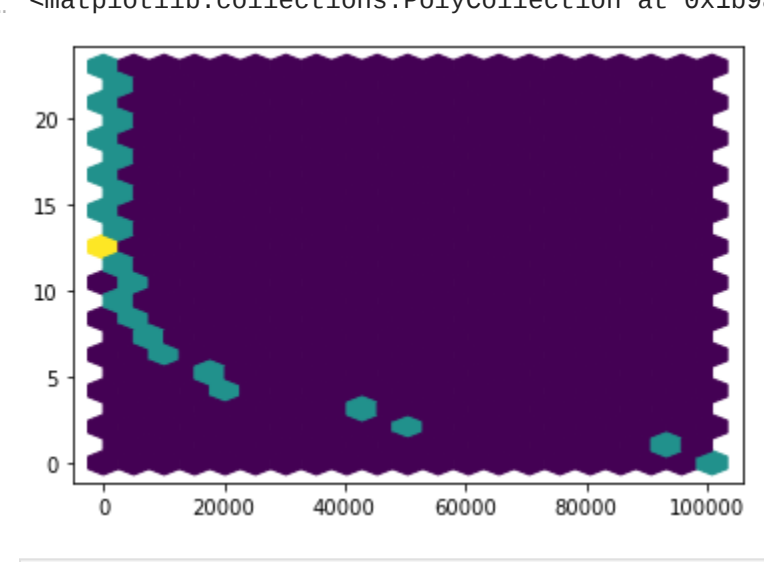
Out[123]:
```
Brooklyn    364558
Name: Agency Name, dtype: int64
```

In [127]:
```python
y=np.arange(0,24)
y
```

Out[127]:
```
array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
       17, 18, 19, 20, 21, 22, 23])
```

In [128]:
```python
plt.hexbin(x,y,gridsize=20)
```

Out[128]: <matplotlib.collections.PolyCollection at 0x1b9a2eaae80>



In [129]:
```python
df[['Agency Name','Complaint Type']]
```

Out[129]:

	Agency Name	Complaint Type
0	Brooklyn	Noise - Street/Sidewalk
1	Brooklyn	Blocked Driveway
2	Brooklyn	Blocked Driveway
3	Brooklyn	Illegal Parking
4	Brooklyn	Illegal Parking
...	...	...
364553	Brooklyn	Illegal Parking
364554	Brooklyn	Noise - Vehicle
364555	Brooklyn	Noise - Street/Sidewalk
364556	Brooklyn	Blocked Driveway
364557	Brooklyn	Blocked Driveway

364558 rows × 2 columns

In [146]:
```python
top_complaints
```

Out[146]:
```
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
Noise - Park                   4109
Drinking                       1409
Noise - House of Worship       1070
Posting Advertisement           681
Urinating in Public             641
Bike/Roller/Skate Chronic       478
Panhandling                     327
Disorderly Youth                315
Illegal Fireworks               172
Graffiti                        157
Agency Issues                     8
Squeegee                          4
Ferry Complaint                   2
Animal in a Park                  1
Name: Complaint Type, dtype: int64
```

In [148]:
```python
x=np.array(top_complaints)
x
```

Out[148]:
```
array([100881,  92679,  51692,  44109,  21661,  19352,  10541,   5198,
         4879,   4192,   4109,   1409,   1070,    681,    641,    478,
          327,    315,    172,    157,      8,      4,      2,      1],
      dtype=int64)
```

In [156]:
```python
p=df['Complaint Type']
p
```

Out[156]:
```
0         Noise - Street/Sidewalk
1                Blocked Driveway
2                Blocked Driveway
3                 Illegal Parking
4                 Illegal Parking
                   ...
364553            Illegal Parking
364554            Noise - Vehicle
364555    Noise - Street/Sidewalk
364556           Blocked Driveway
364557           Blocked Driveway
Name: Complaint Type, Length: 364558, dtype: object
```

In [157]:
```python
z=np.array(p)
z
```

Out[157]:
```
array(['Noise - Street/Sidewalk', 'Blocked Driveway', 'Blocked Driveway',
       ..., 'Noise - Street/Sidewalk', 'Blocked Driveway',
       'Blocked Driveway'], dtype=object)
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: