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

```
In [3]:  ## Understand the dataset

         # Import the dataset

         dataset=pd.read_csv('D:\Simplilearn\Assessments\Data Science with Python\Customer Servic

         #Visualize the dataset

         dataset.head()
```

Out[3]:

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

5 rows × 53 columns

```
In [4]:  #Print the columns of the dataframe

         dataset.columns
```

Out[4]: 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'],
               dtype='object')

```
In [5]:  #Identify the shape of the dataset

         dataset.shape
```

Out[5]: (364558, 53)

```
In [6]:  #Identify the variables with null values

         dataset.isna()
```

Out[6]:

	Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Br High N
0	False	False	False	False	False	False	False	False	False	False	...	
1	False	False	False	False	False	False	False	False	False	False	...	
2	False	False	False	False	False	False	False	False	False	False	...	
3	False	False	False	False	False	False	False	False	False	False	...	
4	False	False	False	False	False	False	False	False	False	False	...	
...	...	...	...	...	...	...	...	...	...	...	...	
364553	False	False	False	False	False	False	False	False	False	False	...	
364554	False	False	False	False	False	False	False	False	False	False	...	
364555	False	False	False	False	False	False	False	False	False	False	...	
364556	False	False	False	False	False	False	False	False	False	False	...	
364557	False	False	False	False	False	False	False	False	False	False	...	

364558 rows × 53 columns

```
In [7]:  dataset.isna().sum()
```

Out[7]:
```
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
```
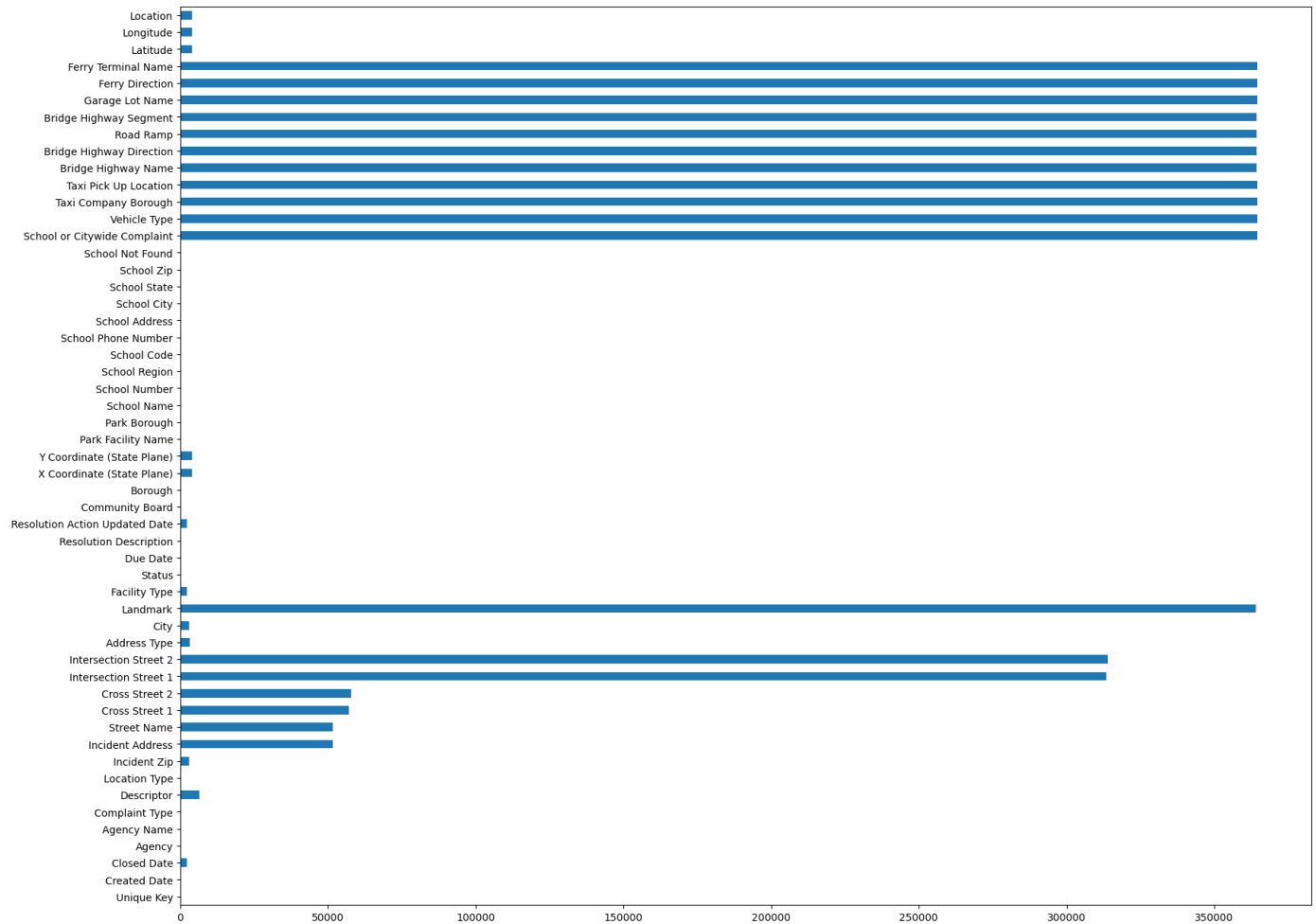
```
Y Coordinate (State Plane)              4030
Park Facility Name                         0
Park Borough                               0
School Name                                0
School Number                              0
School Region                              1
School Code                                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 [8]:
```python
## Perform basic data exploratory analysis
# Draw a frequency plot to show null values in each column of the dataframe

df_null=dataset.isna().sum()
plt.figure(figsize=(20,16))
df_null.plot(kind="barh")
plt.show()
```

```
In [9]:   # Remove the records whose closed Date values are null

          dataset.dropna(subset=['Closed Date'], inplace=True)
          dataset.isna().sum()
```

Out[9]:
```
Unique Key                              0
Created Date                            0
Closed Date                             0
Agency                                  0
Agency Name                             0
Complaint Type                          0
Descriptor                           6496
Location Type                         130
Incident Zip                          675
Incident Address                    51686
Street Name                         51686
Cross Street 1                      55331
Cross Street 2                      55464
Intersection Street 1              311549
Intersection Street 2              311673
Address Type                          929
City                                  674
Landmark                           361802
Facility Type                          18
Status                                  0
Due Date                                1
Resolution Description                  0
Resolution Action Updated Date         39
Community Board                         0
Borough                                 0
X Coordinate (State Plane)           1707
Y Coordinate (State Plane)           1707
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    362177
                Vehicle Type                    362177
                Taxi Company Borough            362177
                Taxi Pick Up Location           362177
                Bridge Highway Name             361880
                Bridge Highway Direction        361880
                Road Ramp                       361915
                Bridge Highway Segment          361915
                Garage Lot Name                 362177
                Ferry Direction                 362177
                Ferry Terminal Name             362177
                Latitude                          1707
                Longitude                         1707
                Location                          1707
                dtype: int64
```

In [10]: `dataset.shape`

Out[10]: `(362177, 53)`

In [11]:
```python
##Check the missing value percentage of all rows in a dataframe, Since in the previous s

missing_percentage=(dataset.isna().sum(axis=0)/dataset.shape[0])*100
missing_percentage
```

Out[11]:
```
                Unique Key                        0.000000
                Created Date                      0.000000
                Closed Date                       0.000000
                Agency                            0.000000
                Agency Name                       0.000000
                Complaint Type                    0.000000
                Descriptor                        1.793598
                Location Type                     0.035894
                Incident Zip                      0.186373
                Incident Address                 14.270923
                Street Name                      14.270923
                Cross Street 1                   15.277337
                Cross Street 2                   15.314059
                Intersection Street 1            86.021200
                Intersection Street 2            86.055437
                Address Type                      0.256504
                City                              0.186097
                Landmark                         99.896459
                Facility Type                     0.004970
                Status                            0.000000
                Due Date                          0.000276
                Resolution Description            0.000000
                Resolution Action Updated Date    0.010768
                Community Board                   0.000000
                Borough                           0.000000
                X Coordinate (State Plane)        0.471317
                Y Coordinate (State Plane)        0.471317
                Park Facility Name                0.000000
                Park Borough                      0.000000
                School Name                       0.000000
```

```
School Number                              0.000000
School Region                              0.000276
School Code                                0.000276
School Phone Number                        0.000000
School Address                             0.000000
School City                                0.000000
School State                               0.000000
School Zip                                 0.000276
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                       99.917996
Bridge Highway Direction                  99.917996
Road Ramp                                 99.927660
Bridge Highway Segment                    99.927660
Garage Lot Name                          100.000000
Ferry Direction                          100.000000
Ferry Terminal Name                      100.000000
Latitude                                   0.471317
Longitude                                  0.471317
Location                                   0.471317
dtype: float64
```

In [12]: `missing_percentage.mean()`

Out[12]: 27.087349816083353

In [13]:
```python
## Since the percentage of missing vlaues in the dataset is less than 30% of the industr
## Hence we can drop the columns with null values having more than 80% from the dataset

remove_cols=pd.DataFrame(dataset.columns.to_list()).set_index(0)
remove_cols=remove_cols[dataset.isna().sum()/dataset.shape[0]*100 < 80].reset_index()
remove_cols
```

Out[13]:

	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

In [14]:
```python
dataset=dataset[remove_cols[0].to_list()]
dataset
```

Out[14]:

	Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incid
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	1003
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	1110
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	1045
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	1046
4	32306529	12/31/2015 11:56:58	01/01/2016 03:24:42	NYPD	New York City Police	Illegal Parking	Blocked Sidewalk	Street/Sidewalk	1137

		PM	AM		Department				
...	...	...	...	...	...	...	...	...	...
364553	29609918	01/01/2015 12:04:44 AM	01/01/2015 10:22:31 AM	NYPD	New York City Police Department	Illegal Parking	Blocked Hydrant	Street/Sidewalk	1142
364554	29608392	01/01/2015 12:04:28 AM	01/01/2015 02:25:02 AM	NYPD	New York City Police Department	Noise - Vehicle	Car/Truck Horn	Street/Sidewalk	1046
364555	29607589	01/01/2015 12:01:30 AM	01/01/2015 12:20:33 AM	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	Street/Sidewalk	1003
364556	29610889	01/01/2015 12:01:29 AM	01/01/2015 02:42:22 AM	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	1046
364557	29611816	01/01/2015 12:00:50 AM	01/01/2015 02:47:50 AM	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	1142

362177 rows × 39 columns

In [15]:
```python
## The School related columns contains most of entries as unspecified
dataset['School Name'].value_counts()
```

Out[15]:
```
Unspecified                       362176
Alley Pond Park - Nature Center        1
Name: School Name, dtype: int64
```

In [17]:
```python
dataset['School Code'].value_counts()
```

Out[17]:
```
Unspecified    362176
Name: School Code, dtype: int64
```

In [18]:
```python
## Lets remove all school related columns as most of the entries are Unspecified
dataset=dataset.drop(dataset.filter(regex='School').columns,axis=1)
dataset
```

Out[18]:

	Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incid
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	1003
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	1110
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	1045
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	1046
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	1137

...	...	...	...	...	...	...	...	...	...
364553	29609918	01/01/2015 12:04:44 AM	01/01/2015 10:22:31 AM	NYPD	New York City Police Department	Illegal Parking	Blocked Hydrant	Street/Sidewalk	1142
364554	29608392	01/01/2015 12:04:28 AM	01/01/2015 02:25:02 AM	NYPD	New York City Police Department	Noise - Vehicle	Car/Truck Horn	Street/Sidewalk	1046
364555	29607589	01/01/2015 12:01:30 AM	01/01/2015 12:20:33 AM	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	Street/Sidewalk	1003
364556	29610889	01/01/2015 12:01:29 AM	01/01/2015 02:42:22 AM	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	1046
364557	29611816	01/01/2015 12:00:50 AM	01/01/2015 02:47:50 AM	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	1142

362177 rows × 29 columns

In [19]:
```python
## Checking similar entries in Borough, Park Borough and community Board
dataset['Borough'].value_counts()
```

Out[19]:
```
BROOKLYN          118851
QUEENS            100754
MANHATTAN          77439
BRONX              49164
STATEN ISLAND      15334
Unspecified          635
Name: Borough, dtype: int64
```

In [20]:
```python
dataset['Park Borough'].value_counts()
```

Out[20]:
```
BROOKLYN          118851
QUEENS            100754
MANHATTAN          77439
BRONX              49164
STATEN ISLAND      15334
Unspecified          635
Name: Park Borough, dtype: int64
```

In [21]:
```python
dataset['Community Board'].value_counts()
```

Out[21]:
```
12 MANHATTAN             14134
01 BROOKLYN              12802
05 QUEENS                11821
01 QUEENS                11637
09 QUEENS                10027
                          ...
26 BRONX                    11
80 QUEENS                   10
56 BROOKLYN                  9
Unspecified QUEENS           2
Unspecified STATEN ISLAND    2
Name: Community Board, Length: 75, dtype: int64
```

In [22]:
```python
## Since Community board, Borough and Park Borough has similar entries, let's remove Com
dataset=dataset.drop(['Community Board'], axis=1)
dataset=dataset.drop(['Park Borough'], axis=1)
dataset.head()
```

Out[22]:

Unique	**Created**	**Closed**	**Agency**	**Agency**	**Complaint**	**Descriptor**	**Location Type**	**Incident**

	Key	Date	Date	Name	Type			Zip
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
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
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
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
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

5 rows × 27 columns

```
In [23]:   ## Park Facility Name also has unspecified entries, check the column
           dataset['Park Facility Name'].value_counts()
```

```
Out[23]:   Unspecified                      362176
           Alley Pond Park - Nature Center       1
           Name: Park Facility Name, dtype: int64
```

```
In [24]:   dataset=dataset.drop(['Park Facility Name'],axis=1)
           dataset.head()
```

Out[24]:

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

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

5 rows × 26 columns

In [25]: `dataset.info()`

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 362177 entries, 0 to 364557
Data columns (total 26 columns):
 #   Column                         Non-Null Count   Dtype
---  ------                         --------------   -----
 0   Unique Key                     362177 non-null  int64
 1   Created Date                   362177 non-null  object
 2   Closed Date                    362177 non-null  object
 3   Agency                         362177 non-null  object
 4   Agency Name                    362177 non-null  object
 5   Complaint Type                 362177 non-null  object
 6   Descriptor                     355681 non-null  object
 7   Location Type                  362047 non-null  object
 8   Incident Zip                   361502 non-null  float64
 9   Incident Address               310491 non-null  object
 10  Street Name                    310491 non-null  object
 11  Cross Street 1                 306846 non-null  object
 12  Cross Street 2                 306713 non-null  object
 13  Address Type                   361248 non-null  object
 14  City                           361503 non-null  object
 15  Facility Type                  362159 non-null  object
 16  Status                         362177 non-null  object
 17  Due Date                       362176 non-null  object
 18  Resolution Description         362177 non-null  object
 19  Resolution Action Updated Date 362138 non-null  object
 20  Borough                        362177 non-null  object
 21  X Coordinate (State Plane)     360470 non-null  float64
 22  Y Coordinate (State Plane)     360470 non-null  float64
 23  Latitude                       360470 non-null  float64
 24  Longitude                      360470 non-null  float64
 25  Location                       360470 non-null  object
dtypes: float64(5), int64(1), object(20)
memory usage: 74.6+ MB
```

In [26]:
```python
## Created and Closed date are in object type. Hence, need to change it to datetime type

dataset["Created Date"]=pd.to_datetime(dataset["Created Date"])
dataset["Closed Date"]=pd.to_datetime(dataset["Closed Date"])
dataset['Due Date']=pd.to_datetime(dataset['Due Date'])
dataset.head()
```

Out[26]:

	Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incid Add

	Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	
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	VERMII AVE
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	27-0 AVE
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	VALEN AVE
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	BAI AVE
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	87-1 R(

5 rows × 26 columns

In [27]:
```python
## Lets check the info
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 362177 entries, 0 to 364557
Data columns (total 26 columns):
 #   Column                         Non-Null Count   Dtype
---  ------                         --------------   -----
 0   Unique Key                     362177 non-null  int64
 1   Created Date                   362177 non-null  datetime64[ns]
 2   Closed Date                    362177 non-null  datetime64[ns]
 3   Agency                         362177 non-null  object
 4   Agency Name                    362177 non-null  object
 5   Complaint Type                 362177 non-null  object
 6   Descriptor                     355681 non-null  object
 7   Location Type                  362047 non-null  object
 8   Incident Zip                   361502 non-null  float64
 9   Incident Address               310491 non-null  object
 10  Street Name                    310491 non-null  object
 11  Cross Street 1                 306846 non-null  object
 12  Cross Street 2                 306713 non-null  object
 13  Address Type                   361248 non-null  object
 14  City                           361503 non-null  object
 15  Facility Type                  362159 non-null  object
 16  Status                         362177 non-null  object
 17  Due Date                       362176 non-null  datetime64[ns]
 18  Resolution Description         362177 non-null  object
 19  Resolution Action Updated Date 362138 non-null  object
 20  Borough                        362177 non-null  object
 21  X Coordinate (State Plane)     360470 non-null  float64
 22  Y Coordinate (State Plane)     360470 non-null  float64
 23  Latitude                       360470 non-null  float64
 24  Longitude                      360470 non-null  float64
```

```
 25  Location                              360470 non-null  object
dtypes: datetime64[ns](3), float64(5), int64(1), object(17)
memory usage: 74.6+ MB
```

In [28]:
```python
## Calculate the time elapsed in closed and created date for Response and closure

dataset["Elapsed_Time"]=dataset['Closed Date'] - dataset['Created Date']
Elapsed_Time=[]
for x in dataset["Closed Date"]-dataset["Created Date"]:
    close=x.total_seconds()
    Elapsed_Time.append(close)

dataset["Elapsed_Time"]=Elapsed_Time
```

In [29]:
```python
## Print the column of Elapsed_Time from the dataset to check if it is converted into se
dataset.head()
```

Out[29]:

	Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Inci Add
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	VERMII AVE
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	27-0 AVE
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	VALEN AVE
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	BAIS AVE
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	87-1 R(

5 rows × 27 columns

In [30]:
```python
## View the descriptive statistics of newly created column i.e. Elapsed_Time
dataset_new=dataset
pd.options.display.float_format = "{:.2f}".format
dataset_new["Elapsed_Time"].describe()
```

Out[30]:
```
count   362177.00
mean     15113.30
std      21102.55
min         61.00
25%       4533.00
50%       9616.00
```

```
75%         18878.00
max       2134342.00
Name: Elapsed_Time, dtype: float64
```

In [31]:
```python
## Check the number of null values in the Complaint_Type and City columns
dataset_new.columns
```

Out[31]:
```
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',
       'Address Type', 'City', 'Facility Type', 'Status', 'Due Date',
       'Resolution Description', 'Resolution Action Updated Date', 'Borough',
       'X Coordinate (State Plane)', 'Y Coordinate (State Plane)', 'Latitude',
       'Longitude', 'Location', 'Elapsed_Time'],
      dtype='object')
```

In [32]:
```python
dataset_new['Complaint Type'].isna().sum(),dataset_new['City'].isna().sum()
```

Out[32]:
```
(0, 674)
```

In [33]:
```python
## City column has 674 null values
## Remove null values from City column
dataset_new.dropna(subset=['City'], inplace=True)
```

In [34]:
```python
dataset_new['City'].isna().sum()
```
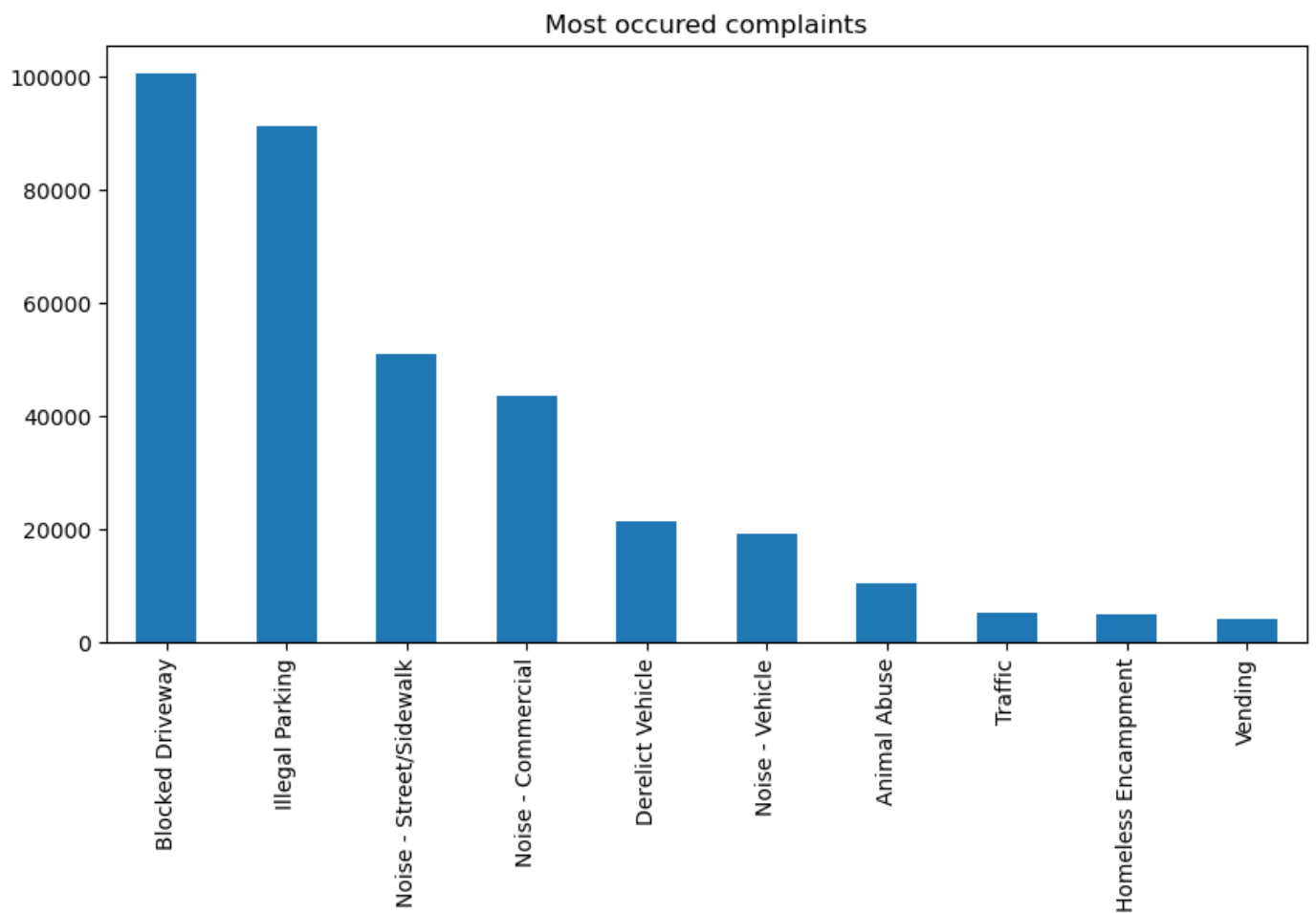
Out[34]:
```
0
```

In [35]:
```python
dataset_new['City'].unique()
```

Out[35]:
```
array(['NEW YORK', 'ASTORIA', 'BRONX', 'ELMHURST', 'BROOKLYN',
       'KEW GARDENS', 'JACKSON HEIGHTS', 'MIDDLE VILLAGE', 'REGO PARK',
       'SAINT ALBANS', 'JAMAICA', 'SOUTH RICHMOND HILL', 'RIDGEWOOD',
       'HOWARD BEACH', 'FOREST HILLS', 'STATEN ISLAND', 'OZONE PARK',
       'RICHMOND HILL', 'WOODHAVEN', 'FLUSHING', 'CORONA',
       'QUEENS VILLAGE', 'OAKLAND GARDENS', 'HOLLIS', 'MASPETH',
       'EAST ELMHURST', 'SOUTH OZONE PARK', 'WOODSIDE', 'FRESH MEADOWS',
       'LONG ISLAND CITY', 'ROCKAWAY PARK', 'SPRINGFIELD GARDENS',
       'COLLEGE POINT', 'BAYSIDE', 'GLEN OAKS', 'FAR ROCKAWAY',
       'BELLEROSE', 'LITTLE NECK', 'CAMBRIA HEIGHTS', 'ROSEDALE',
       'SUNNYSIDE', 'WHITESTONE', 'ARVERNE', 'FLORAL PARK',
       'NEW HYDE PARK', 'CENTRAL PARK', 'BREEZY POINT', 'QUEENS',
       'Astoria', 'Long Island City', 'Woodside', 'East Elmhurst',
       'Howard Beach'], dtype=object)
```

In [75]:
```python
(dataset_new['Complaint Type'].value_counts()).head(10).plot(kind='bar',figsize=(10,5),t
```

Out[75]:
```
<AxesSubplot:title={'center':'Most occured complaints'}>
```

Most occured complaints

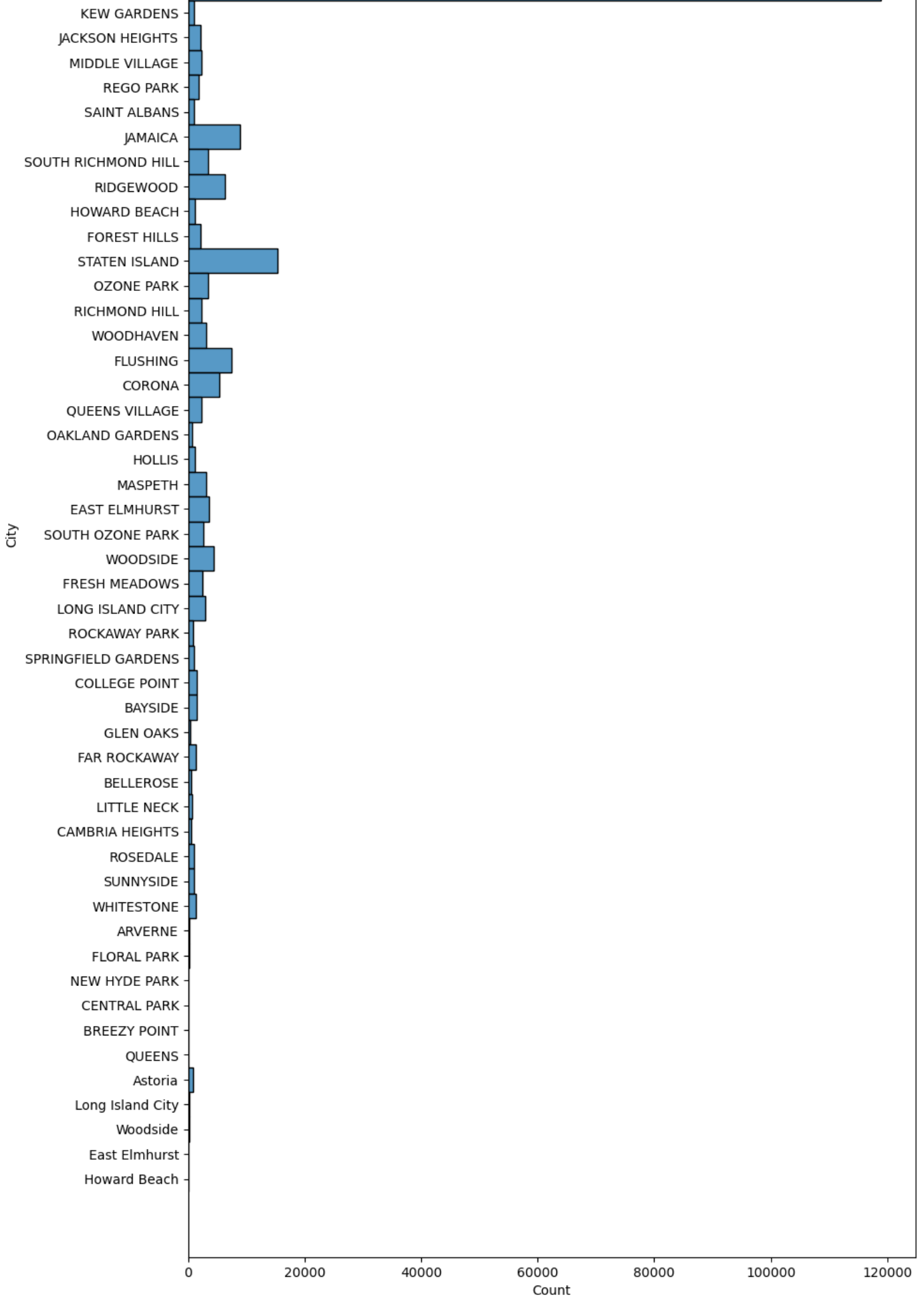## Conclusion 1:

Top 10 Complaints and The above graph shows that the top2 most occured complaints are "Blocked Driveway" and "illegal parking"

In [37]:
```python
## Lets see the complaints based on cities through Pie chart

dataset_new['City'].value_counts().plot(kind='pie',autopct='%1.2f%%',startangle=40,figsi
```
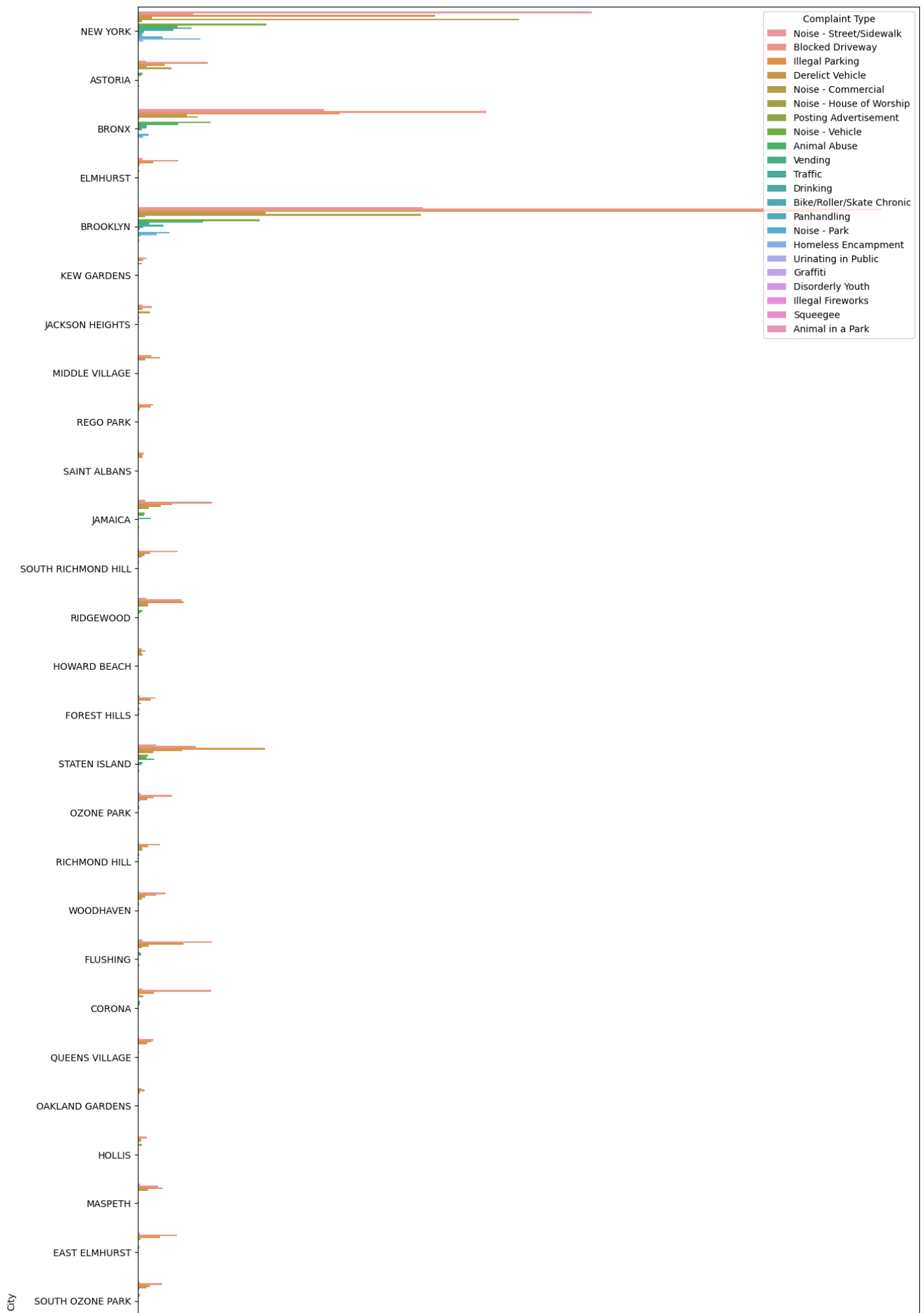
Out[37]: <AxesSubplot:ylabel='City'>

BROOKLYN

32.88%

NEW YORK

21.38%

13.60%

4.24%

2.47%
2.21%
2.07%
1.77%
1.49%
1.21%
0.98%
0.95%
0.95%
0.95%
0.86%
0.86%
0.84%
0.74%
0.68%
0.65%

City

BRONX

STATEN ISLAND

JAMAICA

ASTORIA

FLUSHING

RIDGEWOOD

CORONA

WOODSIDE

EAST ELMHURST

OZONE PARK

ELMHURST

SOUTH RICHMOND HILL

MASPETH

WOODHAVEN

LONG ISLAND CITY

SOUTH OZONE PARK

FRESH MEADOWS

RICHMOND HILL

MIDDLE VILLAGE

QUEENS VILLAGE

FOREST HILLS

JACKSON HEIGHTS

REGO PARK

BAYSIDE

COLLEGE POINT

FAR ROCKAWAY

WHITESTONE

HOWARD BEACH

GARDENS

KEW GARDENS

Conclusion 2:

Brooklyn has the most number of complaints

Conclusion 3:

Frequency plot for city-wise complaints

```
In [37]:  ## Frequency plot for City-wise complaints
          plt.figure(figsize=(10,20))
          sns.histplot(data=dataset_new,y='City')
```

Out[37]: <AxesSubplot:xlabel='Count', ylabel='City'>
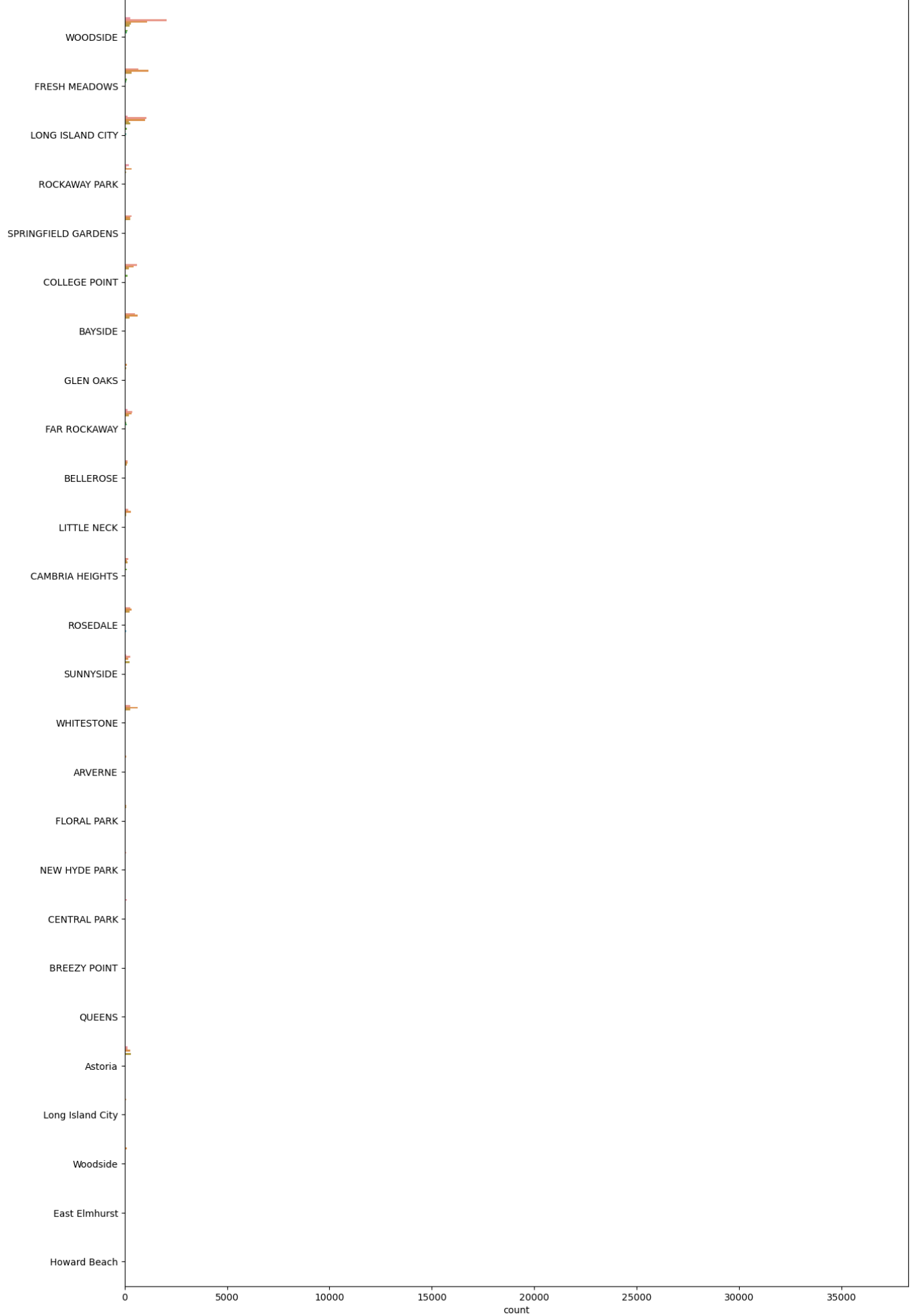

NEW YORK

ASTORIA

BRONX

ELMHURST

BROOKLYN

`## Major types of complaints in each city`

```
plt.figure(figsize=(15,50))
sns.countplot(y='City', hue='Complaint Type', data=dataset_new)
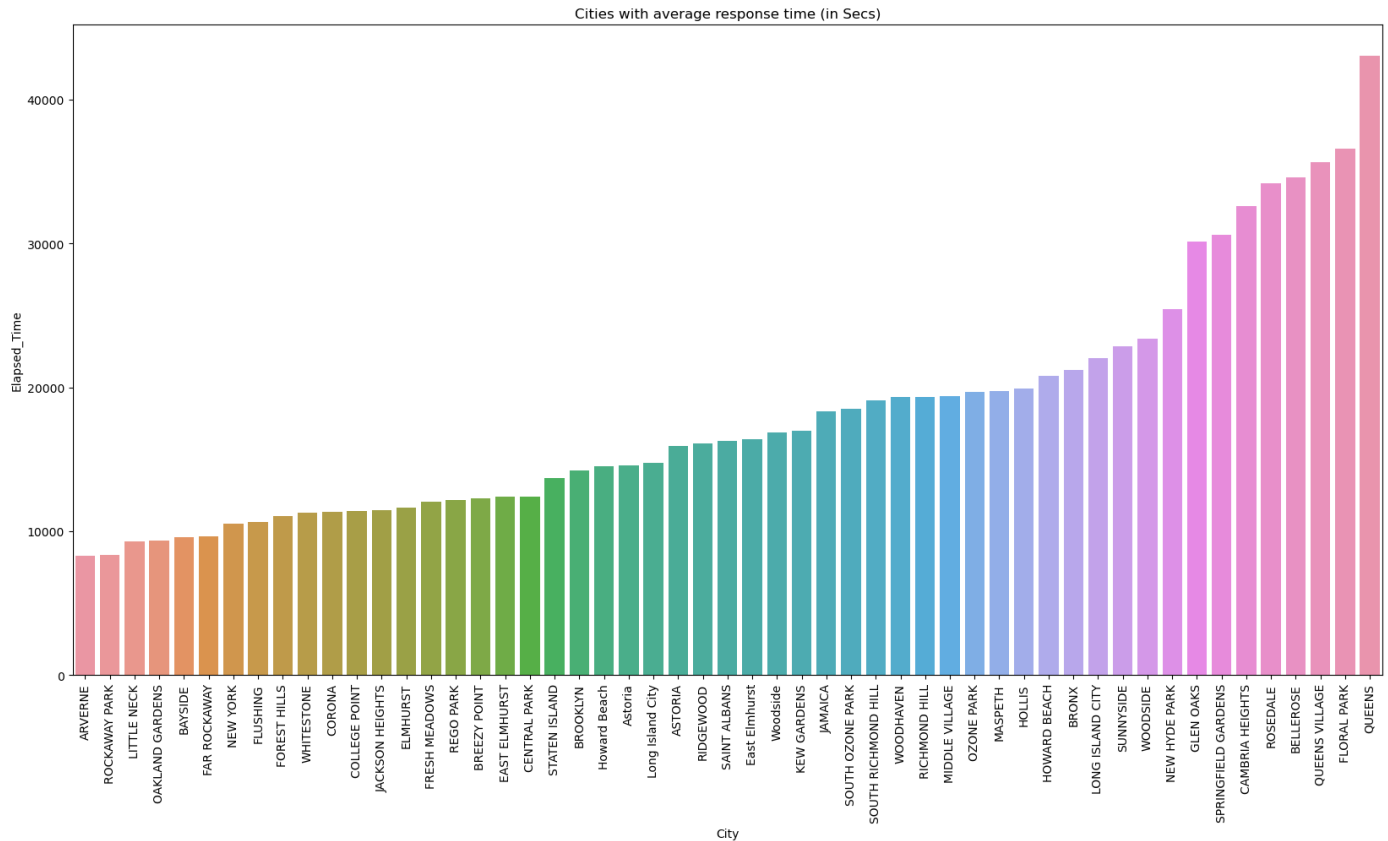```

Out[38]:  `<AxesSubplot:xlabel='count', ylabel='City'>`

y-axis labels (top to bottom):
WOODSIDE
FRESH MEADOWS
LONG ISLAND CITY
ROCKAWAY PARK
SPRINGFIELD GARDENS
COLLEGE POINT
BAYSIDE
GLEN OAKS
FAR ROCKAWAY
BELLEROSE
LITTLE NECK
CAMBRIA HEIGHTS
ROSEDALE
SUNNYSIDE
WHITESTONE
ARVERNE
FLORAL PARK
NEW HYDE PARK
CENTRAL PARK
BREEZY POINT
QUEENS
Astoria
Long Island City
Woodside
East Elmhurst
Howard Beach

x-axis: count
0   5000   10000   15000   20000   25000   30000   35000

Conclusion 4:

Majot types of complaints in each city. Newyork has maximum Noise related complaints whereas Brooklyn has
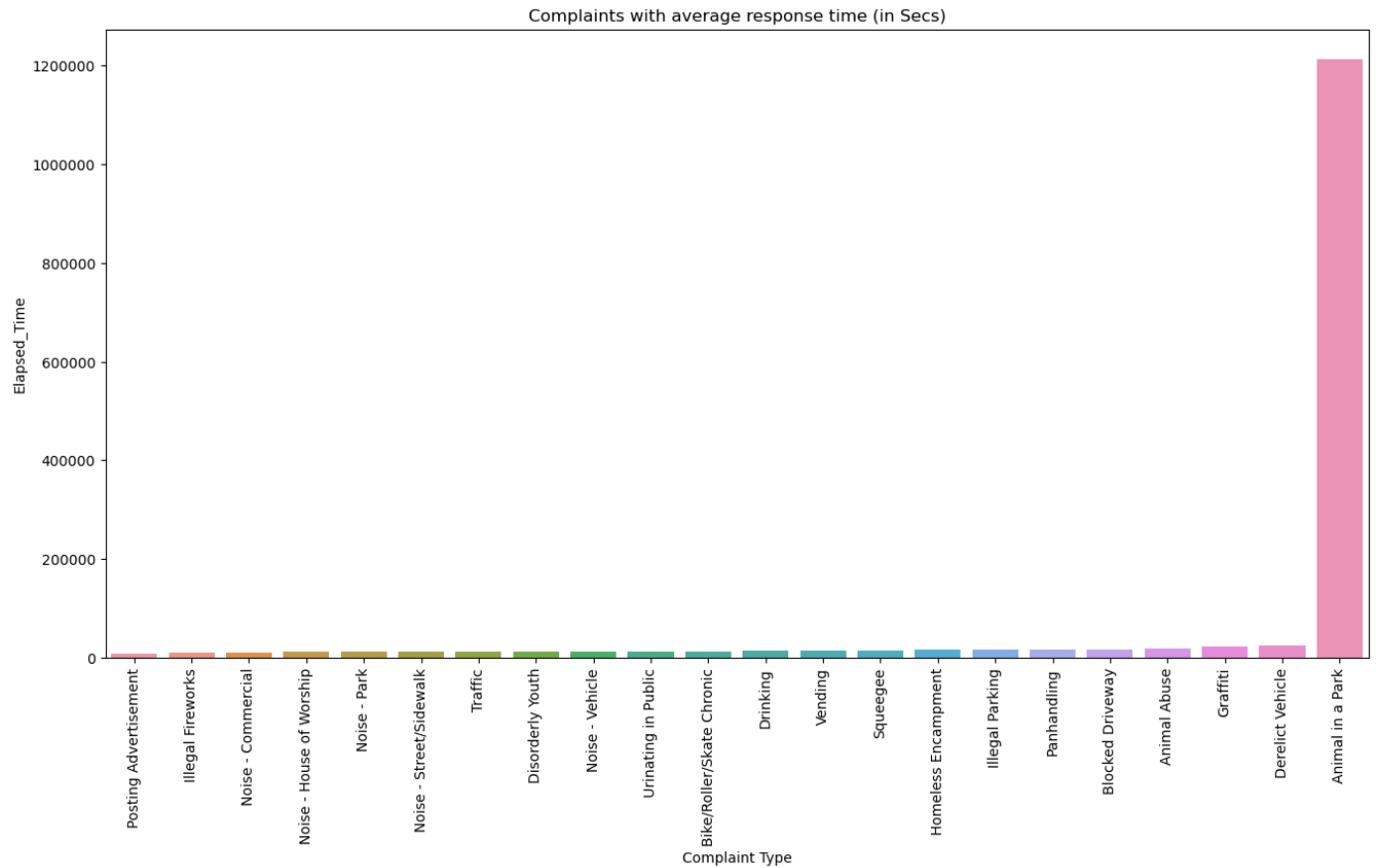maximum Blocked Driveway and Illegal parking complaints

In [39]:
```python
df_city_Elapsed=dataset_new[['City','Elapsed_Time']]
df_city_Elapsed_Avg=df_city_Elapsed.groupby('City')['Elapsed_Time'].mean().to_frame()
df_city_Elapsed_Avg=df_city_Elapsed_Avg.sort_values('Elapsed_Time')
df_city_Elapsed_Avg['City']=df_city_Elapsed_Avg.index
plt.figure(figsize=(20,10))
sns.barplot(x='City',y='Elapsed_Time',data=df_city_Elapsed_Avg)
plt.title("Cities with average response time (in Secs)")
plt.xticks(rotation=90)
plt.show()
```


Conclusion 5:

The fastest response time is in Arverne and Rockway Park while the least response time is in
Floral Park and Queens for the complaints

In [40]:
```python
df_Complaint_Elapsed=dataset_new[['Complaint Type','Elapsed_Time']]
df_Complaint_Elapsed_Avg=df_Complaint_Elapsed.groupby('Complaint Type')['Elapsed_Time'].
df_Complaint_Elapsed_Avg=df_Complaint_Elapsed_Avg.sort_values('Elapsed_Time')
df_Complaint_Elapsed_Avg['Complaint Type']=df_Complaint_Elapsed_Avg.index
plt.figure(figsize=(16,8))
plt.ticklabel_format(style='plain')
sns.barplot(x='Complaint Type',y='Elapsed_Time',data=df_Complaint_Elapsed_Avg)
plt.title("Complaints with average response time (in Secs)")
plt.xticks(rotation=90)
plt.show()
```

Complaints with average response time (in Secs)

Conclusion 6:

Posting Advertisement has the minimum response time and Animal in a park has maximum response time for complaint types

```
In [74]: df_cce=dataset_new[['City','Complaint Type','Elapsed_Time']].copy()
         cce=pd.DataFrame({'Count':df_cce.groupby(['City','Complaint Type']).size()})
         cce.head(60)
```

Out[74]:

City	Complaint Type	Count
ARVERNE	Animal Abuse	46
	Blocked Driveway	50
	Derelict Vehicle	32
	Disorderly Youth	2
	Drinking	1
	Graffiti	1
	Homeless Encampment	4
	Illegal Parking	62
	Noise - Commercial	2
	Noise - House of Worship	14
	Noise - Park	2
	Noise - Street/Sidewalk	29

	Noise - Vehicle	10
	Panhandling	1
	Traffic	1
	Urinating in Public	1
	Vending	1
ASTORIA	Animal Abuse	170
	Bike/Roller/Skate Chronic	16
	Blocked Driveway	3436
	Derelict Vehicle	426
	Disorderly Youth	5
	Drinking	43
	Graffiti	4
	Homeless Encampment	32
	Illegal Fireworks	4
	Illegal Parking	1340
	Noise - Commercial	1653
	Noise - House of Worship	21
	Noise - Park	64
	Noise - Street/Sidewalk	409
	Noise - Vehicle	236
	Panhandling	2
	Posting Advertisement	3
	Traffic	60
	Urinating in Public	10
	Vending	57
Astoria	Blocked Driveway	159
	Derelict Vehicle	14
	Illegal Parking	277
	Noise - Commercial	310
	Noise - Street/Sidewalk	145
BAYSIDE	Animal Abuse	53
	Blocked Driveway	514
	Derelict Vehicle	231
	Disorderly Youth	2
	Drinking	1
	Graffiti	3
	Homeless Encampment	2

	Illegal Parking	638
	Noise - Commercial	47
	Noise - House of Worship	3
	Noise - Park	4
	Noise - Street/Sidewalk	17
	Noise - Vehicle	24
	Traffic	9
	Vending	2
BELLEROSE	**Animal Abuse**	15
	Bike/Roller/Skate Chronic	1
	Blocked Driveway	138

Conclusion 7:

Types of Complaints in each city in a separate dataset called "cce"