

```
In [1]: import pandas as pd
import numpy as np
import math
import json
import matplotlib.pyplot as plt

%matplotlib inline
# read in the csv files
df = pd.read_csv('Desktop/311_Service_Requests_from_2010_to_Present.csv')
```

C:\Users\farah\Anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3057: DtypeWarning: Columns (48,49) have mixed types. Specify dtype option on import or set low_memo ry=False.
interactivity=interactivity, compiler=compiler, result=result)

```
In [2]: df.head()
```

Out[2]:

	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	01-01-16 0:55	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	Street/Sidewalk	100-00000-00000
1	32309934	12/31/2015 11:59:44 PM	01-01-16 1:26	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	111-00000-00000
2	32309159	12/31/2015 11:59:29 PM	01-01-16 4:51	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	104-00000-00000
3	32305098	12/31/2015 11:57:46 PM	01-01-16 7:43	NYPD	New York City Police Department	Illegal Parking	Commercial Overnight Parking	Street/Sidewalk	104-00000-00000
4	32306529	12/31/2015 11:56:58 PM	01-01-16 3:24	NYPD	New York City Police Department	Illegal Parking	Blocked Sidewalk	Street/Sidewalk	113-00000-00000

5 rows × 53 columns

```
In [3]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300698 entries, 0 to 300697
Data columns (total 53 columns):
Unique Key                300698 non-null int64
Created Date              300698 non-null object
Closed Date               298534 non-null object
Agency                    300698 non-null object
Agency Name               300698 non-null object
Complaint Type            300698 non-null object
Descriptor                294784 non-null object
Location Type             300567 non-null object
Incident Zip              298083 non-null float64
Incident Address           256288 non-null object
Street Name                256288 non-null object
```

```

Cross Street 1           251419 non-null object
Cross Street 2           250919 non-null object
Intersection Street 1    43858 non-null object
Intersection Street 2    43362 non-null object
Address Type              297883 non-null object
City                      298084 non-null object
Landmark                  349 non-null object
Facility Type             298527 non-null object
Status                     300698 non-null object
Due Date                  300695 non-null object
Resolution Description    300698 non-null object
Resolution Action Updated Date 298511 non-null object
Community Board            300698 non-null object
Borough                   300698 non-null object
X Coordinate (State Plane) 297158 non-null float64
Y Coordinate (State Plane) 297158 non-null float64
Park Facility Name        300698 non-null object
Park Borough               300698 non-null object
School Name                300698 non-null object
School Number               300698 non-null object
School Region               300697 non-null object
School Code                 300697 non-null object
School Phone Number         300698 non-null object
School Address              300698 non-null object
School City                 300698 non-null object
School State                300698 non-null object
School Zip                  300697 non-null object
School Not Found            300698 non-null object
School or Citywide Complaint 0 non-null float64
Vehicle Type                0 non-null float64
Taxi Company Borough       0 non-null float64
Taxi Pick Up Location      0 non-null float64
Bridge Highway Name         243 non-null object
Bridge Highway Direction    243 non-null object
Road Ramp                   213 non-null object
Bridge Highway Segment      213 non-null object
Garage Lot Name              0 non-null float64
Ferry Direction              1 non-null object
Ferry Terminal Name          2 non-null object
Latitude                     297158 non-null float64
Longitude                    297158 non-null float64
Location                     297158 non-null object
dtypes: float64(10), int64(1), object(42)
memory usage: 121.6+ MB

```

```
In [9]: #convert column to datetime
df['Created Date'] = pd.to_datetime(df['Created Date'])
df['Closed Date'] = pd.to_datetime(df['Closed Date'])
```

```
In [10]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300698 entries, 0 to 300697
Data columns (total 53 columns):
Unique Key                  300698 non-null int64
Created Date                300698 non-null datetime64[ns]
Closed Date                 298534 non-null datetime64[ns]
Agency                      300698 non-null object
Agency Name                 300698 non-null object
Complaint Type              300698 non-null object
Descriptor                  294784 non-null object
Location Type               300567 non-null object
Incident Zip                298083 non-null float64
Incident Address             256288 non-null object
```

Street Name	256288	non-null object
Cross Street 1	251419	non-null object
Cross Street 2	250919	non-null object
Intersection Street 1	43858	non-null object
Intersection Street 2	43362	non-null object
Address Type	297883	non-null object
City	298084	non-null object
Landmark	349	non-null object
Facility Type	298527	non-null object
Status	300698	non-null object
Due Date	300695	non-null object
Resolution Description	300698	non-null object
Resolution Action Updated Date	298511	non-null object
Community Board	300698	non-null object
Borough	300698	non-null object
X Coordinate (State Plane)	297158	non-null float64
Y Coordinate (State Plane)	297158	non-null float64
Park Facility Name	300698	non-null object
Park Borough	300698	non-null object
School Name	300698	non-null object
School Number	300698	non-null object
School Region	300697	non-null object
School Code	300697	non-null object
School Phone Number	300698	non-null object
School Address	300698	non-null object
School City	300698	non-null object
School State	300698	non-null object
School Zip	300697	non-null object
School Not Found	300698	non-null object
School or Citywide Complaint	0	non-null float64
Vehicle Type	0	non-null float64
Taxi Company Borough	0	non-null float64
Taxi Pick Up Location	0	non-null float64
Bridge Highway Name	243	non-null object
Bridge Highway Direction	243	non-null object
Road Ramp	213	non-null object
Bridge Highway Segment	213	non-null object
Garage Lot Name	0	non-null float64
Ferry Direction	1	non-null object
Ferry Terminal Name	2	non-null object
Latitude	297158	non-null float64
Longitude	297158	non-null float64
Location	297158	non-null object

dtypes: datetime64[ns](2), float64(10), int64(1), object(40)
memory usage: 121.6+ MB

```
In [11]: #drop useless columns which are include huge rate of Null values
df = df.drop(['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'], axis=1)
```

```
In [12]: df.describe()
```

	Unique Key	Incident Zip	X Coordinate (State Plane)	Y Coordinate (State Plane)	Latitude	Longitude
count	3.006980e+05	298083.000000	2.971580e+05	297158.000000	297158.000000	297158.000000
mean	3.130054e+07	10848.888645	1.004854e+06	203754.534416	40.725885	-73.925630
std	5.738547e+05	583.182081	2.175338e+04	29880.183529	0.082012	0.078454
min	3.027948e+07	83.000000	9.133570e+05	121219.000000	40.499135	-74.254937

	Unique Key	Incident Zip	X Coordinate (State Plane)	Y Coordinate (State Plane)	Latitude	Longitude
25%	3.080118e+07	10310.000000	9.919752e+05	183343.000000	40.669796	-73.972142
50%	3.130436e+07	11208.000000	1.003158e+06	201110.500000	40.718661	-73.931781
75%	3.178446e+07	11238.000000	1.018372e+06	224125.250000	40.781840	-73.876805
max	3.231065e+07	11697.000000	1.067173e+06	271876.000000	40.912869	-73.700760

In [13]:

```
#Display all columns
pd.set_option('display.max_columns', None)
df.head()
```

Out[13]:

	Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incide Z
0	32310363	2015-12-31 23:59:45	2016-01-01 00:55:00	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	Street/Sidewalk	10034
1	32309934	2015-12-31 23:59:44	2016-01-01 01:26:00	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	11105
2	32309159	2015-12-31 23:59:29	2016-01-01 04:51:00	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	10458
3	32305098	2015-12-31 23:57:46	2016-01-01 07:43:00	NYPD	New York City Police Department	Illegal Parking	Commercial Overnight Parking	Street/Sidewalk	10461
4	32306529	2015-12-31 23:56:58	2016-01-01 03:24:00	NYPD	New York City Police Department	Illegal Parking	Blocked Sidewalk	Street/Sidewalk	11373

In [14]: df1=df.copy()

In [15]: df['new_date'] = [d.date() for d in df['Closed Date']]

In [139...]: df1['year_Request_Closing_Time'] = df1['Closed Date'].dt.year

In [140...]: df1['month_Request_Closing_Time'] = df1['Closed Date'].dt.month

In [14]: df1.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300698 entries, 0 to 300697
Data columns (total 46 columns):
Unique_Key           300698 non-null int64
Created Date         300698 non-null datetime64[ns]
Closed Date          298534 non-null datetime64[ns]
Agency               300698 non-null object
Agency_Name          300698 non-null object
Complaint_Type      300698 non-null object
Descriptor           294784 non-null object
Location_Type        300567 non-null object
Incident_Zip         298083 non-null float64
Incident_Address     256288 non-null object
Street Name          256288 non-null object
Cross Street 1       251419 non-null object
Cross Street 2       250919 non-null object
Intersection Street 1 43858 non-null object
Intersection Street 2 43362 non-null object
Address Type         297883 non-null object
City                 298084 non-null object
Landmark              349 non-null object
Facility Type         298527 non-null object
Status                300698 non-null object
Due Date              300695 non-null object
Resolution Description 300698 non-null object
Resolution Action Updated Date 298511 non-null object
Community Board       300698 non-null object
Borough               300698 non-null object
X Coordinate (State Plane) 297158 non-null float64
Y Coordinate (State Plane) 297158 non-null float64
Park Facility Name    300698 non-null object
Park Borough           300698 non-null object
School Name            300698 non-null object
School Number          300698 non-null object
School Region          300697 non-null object
School Code             300697 non-null object
School Phone Number    300698 non-null object
School Address          300698 non-null object
School City              300698 non-null object
School State             300698 non-null object
School Zip                300697 non-null object
School Not Found        300698 non-null object
Latitude                  297158 non-null float64
Longitude                  297158 non-null float64
Location                  297158 non-null object
hour_Request_Closing_Time 298534 non-null float64
minute_Request_Closing_Time 298534 non-null float64
year_Request_Closing_Time 298534 non-null float64
month_Request_Closing_Time 298534 non-null float64
dtypes: datetime64[ns](2), float64(9), int64(1), object(34)
memory usage: 105.5+ MB
```

In [16]: #df1['Dates'] = pd.to_datetime(df1['Closed Date']).dt.date
#df1['Request_Closing_Time'] = pd.to_datetime(df1['Closed Date']).dt.time

In [17]: df1['hour_Request_Closing_Time'] = df1['Closed Date'].dt.hour

In [18]: df1['minute_Request_Closing_Time'] = df1['Closed Date'].dt.minute

In [20]: #drop useless columns which are include huge rate of Null values

```
#df1 = df1.drop(['second_Request_Closing_Time'], axis=1)
```

In [21]: `df1.head()`

Out[21]:

	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 00:55:00	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	Street/Sidewalk	10034
1	32309934	2015-12-31 23:59:44	2016-01-01 01:26:00	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	11105
2	32309159	2015-12-31 23:59:29	2016-01-01 04:51:00	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	10458
3	32305098	2015-12-31 23:57:46	2016-01-01 07:43:00	NYPD	New York City Police Department	Illegal Parking	Commercial Overnight Parking	Street/Sidewalk	10461
4	32306529	2015-12-31 23:56:58	2016-01-01 03:24:00	NYPD	New York City Police Department	Illegal Parking	Blocked Sidewalk	Street/Sidewalk	11373

In [22]: `df1.rename(columns={'Unique Key': 'Unique_Key', 'Agency Name': 'Agency_Name', 'Complain`

In [23]: `df1.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300698 entries, 0 to 300697
Data columns (total 44 columns):
Unique_Key          300698 non-null int64
Created Date        300698 non-null datetime64[ns]
Closed Date         298534 non-null datetime64[ns]
Agency              300698 non-null object
Agency_Name          300698 non-null object
Complaint_Type      300698 non-null object
Descriptor          294784 non-null object
Location_Type       300567 non-null object
Incident_Zip         298083 non-null float64
Incident_Address    256288 non-null object
Street_Name          256288 non-null object
Cross_Street_1       251419 non-null object
Cross_Street_2       250919 non-null object
Intersection_Street_1 43858 non-null object
```

```

Intersection Street 2           43362 non-null object
Address Type                   297883 non-null object
City                           298084 non-null object
Landmark                        349 non-null object
Facility Type                  298527 non-null object
Status                          300698 non-null object
Due Date                        300695 non-null object
Resolution Description          300698 non-null object
Resolution Action Updated Date 298511 non-null object
Community Board                 300698 non-null object
Borough                         300698 non-null object
X Coordinate (State Plane)    297158 non-null float64
Y Coordinate (State Plane)    297158 non-null float64
Park Facility Name             300698 non-null object
Park Borough                    300698 non-null object
School Name                     300698 non-null object
School Number                   300698 non-null object
School Region                   300697 non-null object
School Code                      300697 non-null object
School Phone Number             300698 non-null object
School Address                  300698 non-null object
School City                      300698 non-null object
School State                     300698 non-null object
School Zip                       300697 non-null object
School Not Found                300698 non-null object
Latitude                         297158 non-null float64
Longitude                        297158 non-null float64
Location                         297158 non-null object
hour_Request_Closing_Time       298534 non-null float64
minute_Request_Closing_Time     298534 non-null float64
dtypes: datetime64[ns](2), float64(7), int64(1), object(34)
memory usage: 100.9+ MB

```

In [25]: `df1['Agency_Name'].unique()`

Out[25]: `array(['New York City Police Department', 'NYPD', 'Internal Affairs Bureau'], dtype=object)`

In [26]: `df1['Descriptor'].unique()`

Out[26]: `array(['Loud Music/Party', 'No Access', 'Commercial Overnight Parking', 'Blocked Sidewalk', 'Posted Parking Sign Violation', 'Blocked Hydrant', 'With License Plate', 'Partial Access', 'Unauthorized Bus Layover', 'Double Parked Blocking Vehicle', 'Double Parked Blocking Traffic', 'Vehicle', 'Loud Talking', 'Banging/Pounding', 'Car/Truck Music', 'Tortured', 'In Prohibited Area', 'Congestion/Gridlock', 'Neglected', 'Car/Truck Horn', 'In Public', 'Other (complaint details)', nan, 'No Shelter', 'Truck Route Violation', 'Unlicensed', 'Overnight Commercial Storage', 'Engine Idling', 'After Hours - Licensed Est', 'Detached Trailer', 'Underage - Licensed Est', 'Chronic Stoplight Violation', 'Loud Television', 'Chained', 'Building', 'In Car', 'Police Report Requested', 'Chronic Speeding', 'Playing in Unsuitable Place', 'Drag Racing', 'Police Report Not Requested', 'Nuisance/Truant', 'Homeless Issue', 'Language Access Complaint', 'Disruptive Passenger', 'Animal Waste'], dtype=object)`

In [170...]: `df1['Complaint_Type'].unique()`

Out[170...]: `array(['Noise - Street/Sidewalk', 'Blocked Driveway', 'Illegal Parking', 'Derelict Vehicle', 'Noise - Commercial', 'Noise - House of Worship', 'Posting Advertisement',`

```
'Noise - Vehicle', 'Animal Abuse', 'Vending', 'Traffic',
'Drinking', 'Bike/Roller/Skate Chronic', 'Panhandling',
'Noise - Park', 'Homeless Encampment', 'Urinating in Public',
'Graffiti', 'Disorderly Youth', 'Illegal Fireworks',
'Ferry Complaint', 'Agency Issues', 'Squeegee', 'Animal in a Park'],
dtype=object)
```

In [157]: df1.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300698 entries, 0 to 300697
Data columns (total 46 columns):
Unique_Key                      300698 non-null int64
Created Date                     300698 non-null datetime64[ns]
Closed Date                      298534 non-null datetime64[ns]
Agency                           300698 non-null object
Agency_Name                      300698 non-null object
Complaint_Type                  300698 non-null object
Descriptor                        294784 non-null object
Location_Type                    300567 non-null object
Incident Zip                     298083 non-null float64
Incident Address                 256288 non-null object
Street Name                      256288 non-null object
Cross Street 1                   251419 non-null object
Cross Street 2                   250919 non-null object
Intersection Street 1            43858 non-null object
Intersection Street 2            43362 non-null object
Address Type                     297883 non-null object
City                             298084 non-null object
Landmark                          349 non-null object
Facility Type                    298527 non-null object
Status                            300698 non-null object
Due Date                          300695 non-null object
Resolution Description           300698 non-null object
Resolution Action Updated Date  298511 non-null object
Community Board                  300698 non-null object
Borough                           300698 non-null object
X Coordinate (State Plane)     297158 non-null float64
Y Coordinate (State Plane)     297158 non-null float64
Park Facility Name              300698 non-null object
Park Borough                      300698 non-null object
School Name                       300698 non-null object
School Number                     300698 non-null object
School Region                     300697 non-null object
School Code                       300697 non-null object
School Phone Number              300698 non-null object
School Address                    300698 non-null object
School City                        300698 non-null object
School State                      300698 non-null object
School Zip                         300697 non-null object
School Not Found                 300698 non-null object
Latitude                          297158 non-null float64
Longitude                         297158 non-null float64
Location                           297158 non-null object
hour_Request_Closing_Time        298534 non-null float64
minute_Request_Closing_Time      298534 non-null float64
year_Request_Closing_Time        298534 non-null float64
month_Request_Closing_Time       298534 non-null float64
dtypes: datetime64[ns](2), float64(9), int64(1), object(34)
memory usage: 105.5+ MB
```

In [27]: df1['Unique_Key'].nunique()

Out[27]: 300698

```
In [188... P_MIN =df1.minute_Request_Closing_Time.mean()  
P_MIN
```

```
Out[188... 29.10856049897164
```

```
In [161... p_hour = df1.hour_Request_Closing_Time.mean()  
p_hour
```

```
Out[161... 12.146576269369653
```

```
In [165... p_month = df1.month_Request_Closing_Time.mean()  
p_month
```

```
Out[165... 7.93286526827765
```

```
In [180... p_year = df1.year_Request_Closing_Time.mean()
```

```
In [169... n_Descriptor = df1.query('Descriptor == "Loud Music/Party"').shape[0]  
n_Descriptor
```

```
Out[169... 61430
```

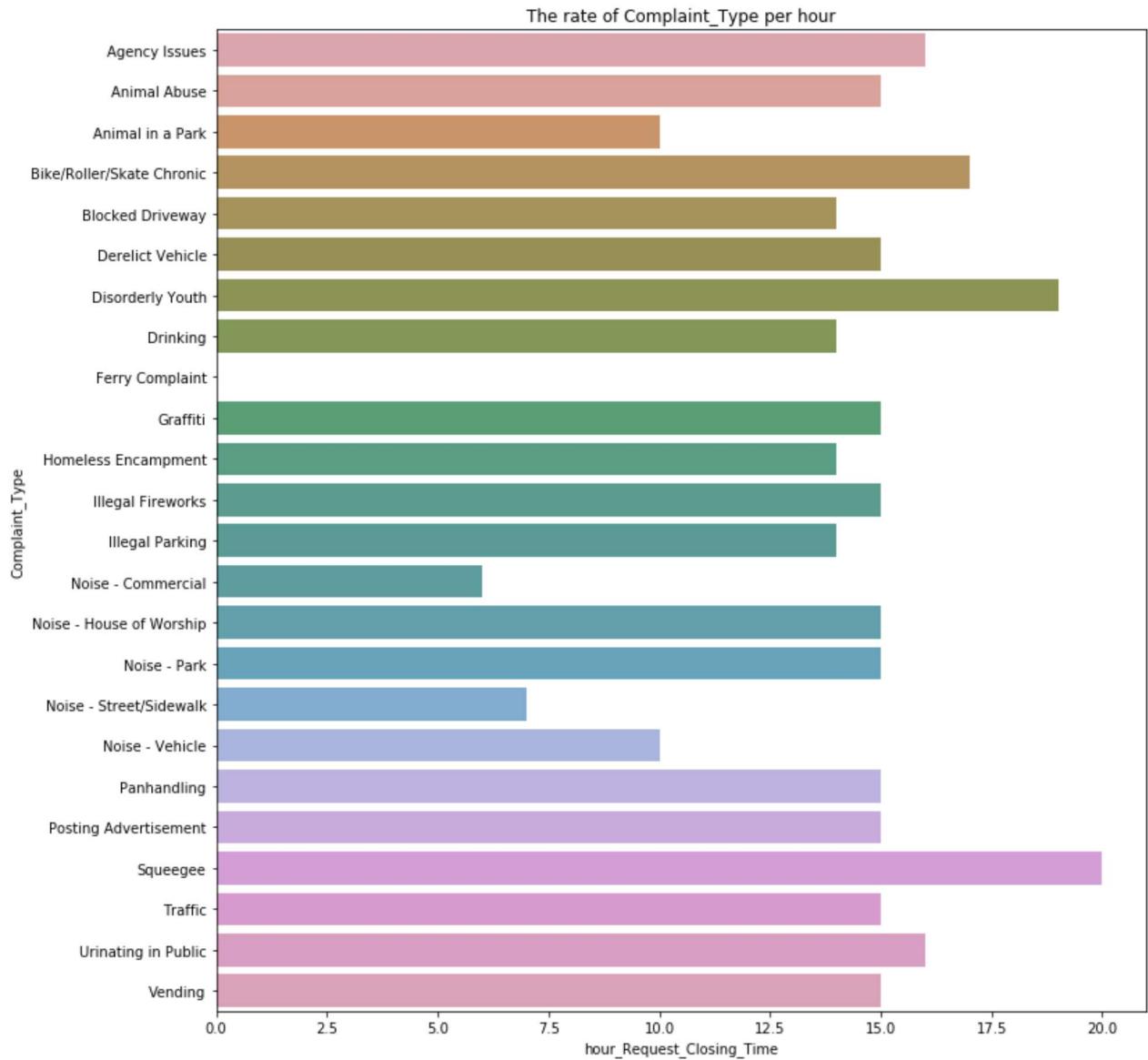
```
In [172... n_Complaint_Type = df1.query('Complaint_Type== "Noise - Street/Sidewalk"').shape[0]  
n_Complaint_Type
```

```
Out[172... 48612
```

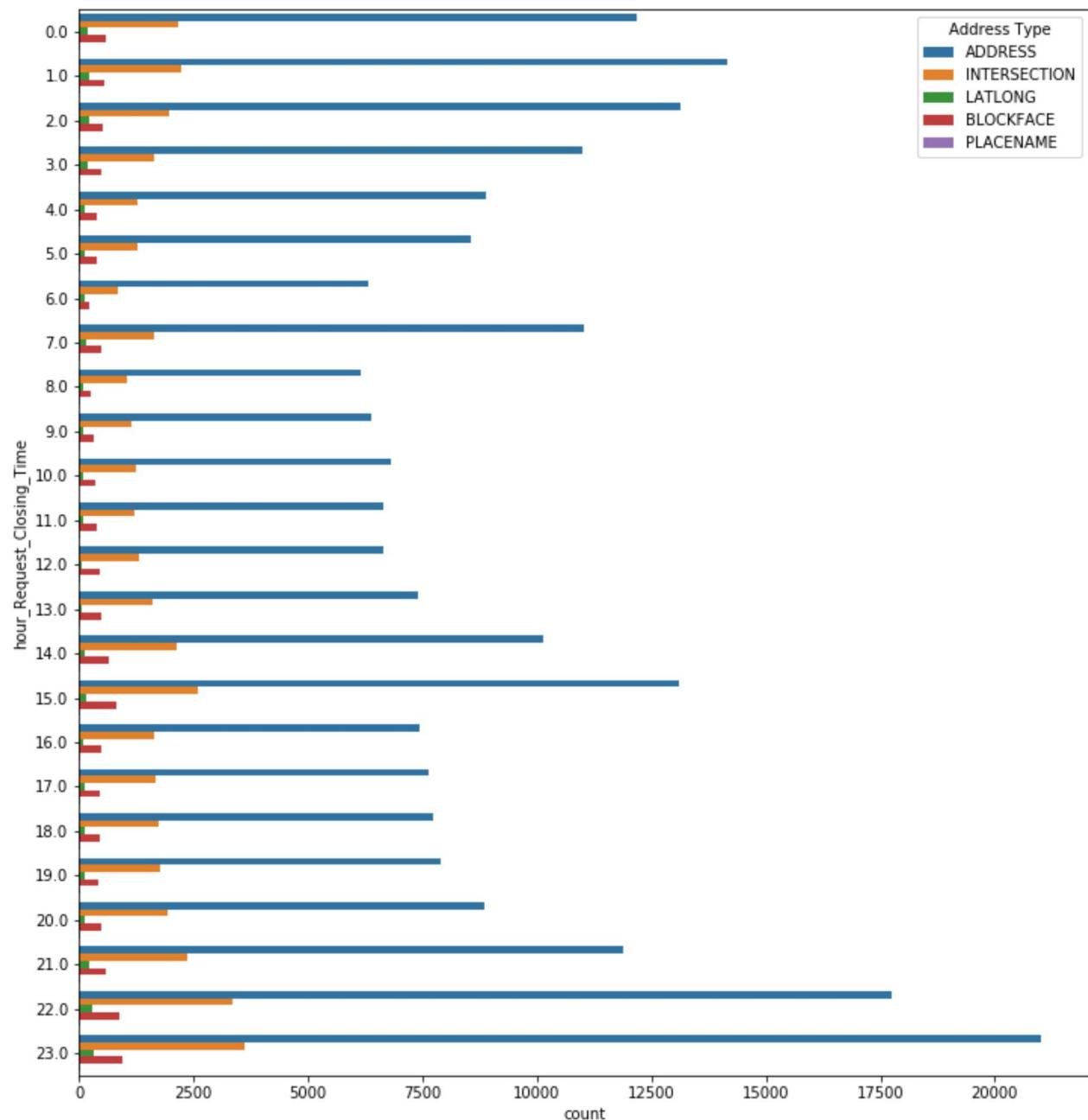
```
In [194... #n_Descriptor_Loud= np.random.choice([1,0], size=n_Descriptor, p=[p_year, (1-p_year)])
```

```
In [72]: result = df1.groupby(["Complaint_Type"])['hour_Request_Closing_Time'].aggregate(np.me
```

```
In [101... sns.barplot(y='Complaint_Type', x='hour_Request_Closing_Time' ,data=result, saturation=plt.rcParams['figure.figsize']=(12,13)
```



```
In [109]: sns.countplot(y ='hour_Request_Closing_Time', hue = "Address Type", data = df1)
plt.rcParams['figure.figsize']=(12,13)
```



```
In [154]: rate_Request_Closing_Time_in_month = df1.groupby('month_Request_Closing_Time').sum()
rate_Request_Closing_Time_in_month.sort_values(by ='hour_Request_Closing_Time', ascending=False)
```

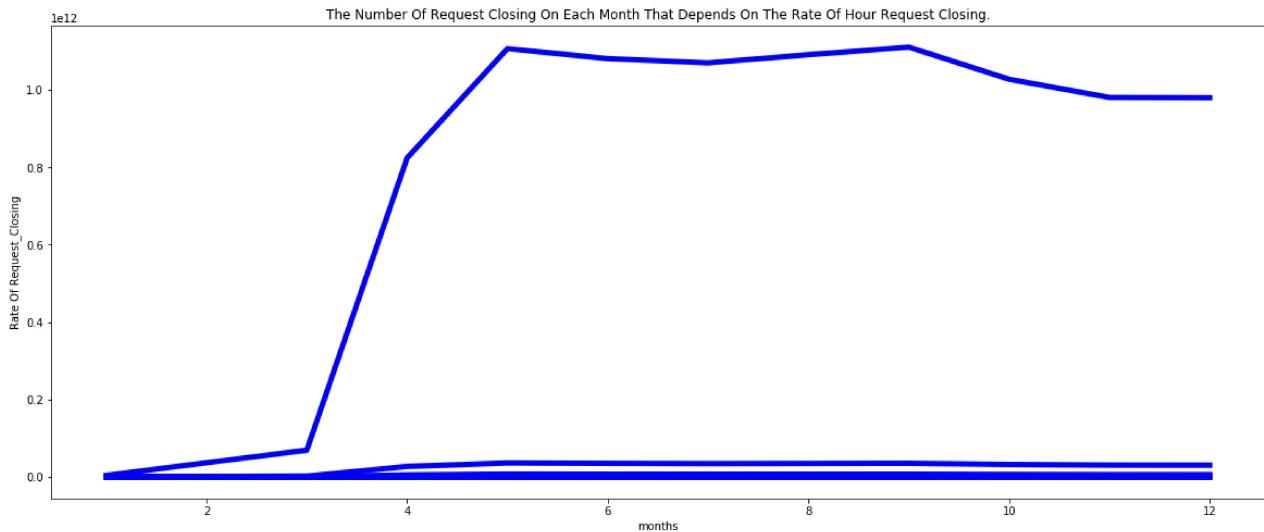
Out[154]:

	Unique_Key	Incident_Zip	X_Coordinate (State Plane)	Y_Coordinate (State Plane)	Latitude	Longitude
month_Request_Closing_Time						
0.0	4070795209	1340229.0	1.239863e+08	2.541449e+07	5.010250e+03	-75.200000
1.0	69250613118	24847405.0	2.285794e+09	4.584043e+08	9.255694e+04	-75.200000
2.0	824253541601	292971757.0	2.707925e+10	5.489842e+09	1.097961e+06	-75.200000
3.0	980506175616	332888576.0	3.067676e+10	6.159862e+09	1.243958e+06	-75.200000
4.0	979779789018	330642282.0	3.046233e+10	6.113049e+09	1.234832e+06	-75.200000
10.0	1027532317488	350804823.0	3.236849e+10	6.534947e+09	1.312347e+06	-75.200000
11.0						
12.0						
13.0						
14.0						
15.0						
16.0						
17.0						
18.0						
19.0						
20.0						
21.0						
22.0						
23.0						

	Unique_Key	Incident_Zip	X_Coordinate (State Plane)	Y_Coordinate (State Plane)	Latitude
month_Request_Closing_Time					
7.0	1069943952564	372267730.0	3.444315e+10	7.048681e+09	1.394733e+06
8.0	1090894417955	376238696.0	3.477119e+10	7.104118e+09	1.408622e+06
9.0	1110651869776	381468844.0	3.519358e+10	7.127433e+09	1.426721e+06
6.0	1080743419881	379545414.0	3.502140e+10	7.091760e+09	1.418911e+06
5.0	1106745753254	390258271.0	3.611848e+10	7.382142e+09	1.464092e+06

In [155...]

```
plt.figure(figsize=(20,8))
plt.plot(rate_Request_Closing_Time_in_month, color='blue', linewidth=5)
plt.xlabel('months')
plt.ylabel('Rate Of Request_Closing')
plt.title('The Number Of Request Closing On Each Month That Depends On The Rate Of Hour')
plt.show;
```

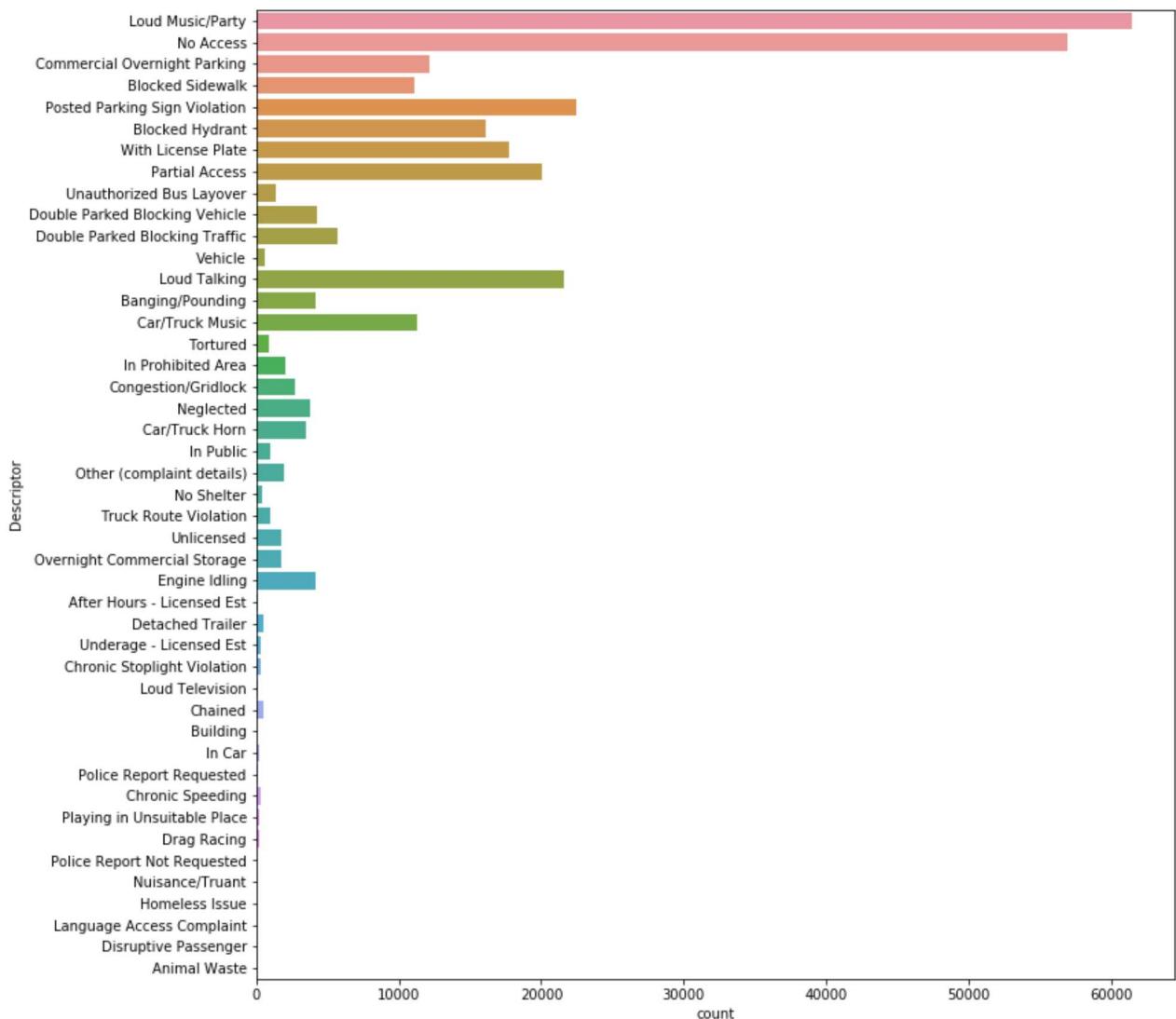


In [119...]

```
#sns.countplot(df1['Descriptor'])
# plt.title('Number of events In Transcripts')
# plt.rcParams['figure.figsize']=(12,13)

sns.countplot(y ='Descriptor', data = df1)
plt.rcParams['figure.figsize']=(12,13)

plt.show();
```



In []: