In [14]:

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import os

# 1.Import a 311 NYC service request.
os.getcwd()
df=pd.read_csv(r'C:\Users\Shubham\Downloads\311_Service_Requests_from_2010_to_Present.csv')

print(df.head())
```

```
C:\Users\Shubham\Anaconda3\lib\site-packages\IPython\core\interactiveshell.p
y:3049: DtypeWarning: Columns (48,49) have mixed types. Specify dtype option
on import or set low_memory=False.
  interactivity=interactivity, compiler=compiler, result=result)
```

```
   Unique Key            Created Date    Closed Date Agency  \
0    32310363  12/31/2015 11:59:45 PM  01-01-16 0:55   NYPD
1    32309934  12/31/2015 11:59:44 PM  01-01-16 1:26   NYPD
2    32309159  12/31/2015 11:59:29 PM  01-01-16 4:51   NYPD
3    32305098  12/31/2015 11:57:46 PM  01-01-16 7:43   NYPD
4    32306529  12/31/2015 11:56:58 PM  01-01-16 3:24   NYPD

                      Agency Name           Complaint Type  \
0  New York City Police Department  Noise - Street/Sidewalk
1  New York City Police Department         Blocked Driveway
2  New York City Police Department         Blocked Driveway
3  New York City Police Department          Illegal Parking
4  New York City Police Department          Illegal Parking

                   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

        Incident Address  ... Bridge Highway Name Bridge Highway Direction
\
0    71 VERMILYEA AVENUE  ...                NaN                     NaN
1        27-07 23 AVENUE  ...                NaN                     NaN
2  2897 VALENTINE AVENUE  ...                NaN                     NaN
3    2940 BAISLEY AVENUE  ...                NaN                     NaN
4          87-14 57 ROAD  ...                NaN                     NaN

  Road Ramp Bridge Highway Segment Garage Lot Name Ferry Direction  \
0       NaN                   NaN            NaN             NaN
1       NaN                   NaN            NaN             NaN
2       NaN                   NaN            NaN             NaN
3       NaN                   NaN            NaN             NaN
4       NaN                   NaN            NaN             NaN

  Ferry Terminal Name   Latitude  Longitude  \
0                 NaN  40.865682 -73.923501
1                 NaN  40.775945 -73.915094
2                 NaN  40.870325 -73.888525
3                 NaN  40.835994 -73.828379
4                 NaN  40.733060 -73.874170
```

```
                                  Location
0   (40.86568153633767, -73.92350095571744)
1  (40.775945312321085, -73.91509393898605)
2  (40.870324522111424, -73.88852464418646)
3   (40.83599404683083, -73.82837939584206)
4  (40.733059618956815, -73.87416975810375)

[5 rows x 53 columns]
```

In [15]:

```
print(df.shape)
```

```
(300698, 53)
```

In [16]:

```
print(df.isnull().sum())
```

```
Unique Key                          0
Created Date                        0
Closed Date                      2164
Agency                              0
Agency Name                         0
Complaint Type                      0
Descriptor                       5914
Location Type                     131
Incident Zip                     2615
Incident Address                44410
Street Name                     44410
Cross Street 1                  49279
Cross Street 2                  49779
Intersection Street 1          256840
Intersection Street 2          257336
Address Type                     2815
City                             2614
Landmark                       300349
Facility Type                    2171
Status                              0
Due Date                            3
Resolution Description              0
Resolution Action Updated Date   2187
Community Board                     0
Borough                             0
X Coordinate (State Plane)       3540
Y Coordinate (State Plane)       3540
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   300698
Vehicle Type                   300698
Taxi Company Borough           300698
Taxi Pick Up Location          300698
Bridge Highway Name            300455
Bridge Highway Direction       300455
Road Ramp                      300485
Bridge Highway Segment         300485
Garage Lot Name                300698
Ferry Direction                300697
Ferry Terminal Name            300696
Latitude                         3540
Longitude                        3540
Location                         3540
dtype: int64
```

In [17]:

```
print(df[df['Closed Date'].isnull()])
```

```
...           ...                      ...         ...    ...
295003    30329648          04-04-15 20:49         NaN   NYPD
295549    30328285           04-04-15 2:06         NaN   NYPD
295636    30322486           04-04-15 0:25         NaN   NYPD
295639    30323888           04-04-15 0:24         NaN   NYPD
295840    30326689          04-03-15 21:51         NaN   NYPD
296033    30327613          04-03-15 17:11         NaN   NYPD
296570    30316718          04-02-15 22:50         NaN   NYPD
296848    30317612          04-02-15 18:01         NaN   NYPD
297038    30317663          04-02-15 13:14         NaN   NYPD
297358    30310830           04-02-15 0:53         NaN   NYPD
297372    30307739           04-02-15 0:20         NaN   NYPD
297464    30307698          04-01-15 22:43         NaN   NYPD
297497    30308674          04-01-15 22:07         NaN   NYPD
297738    30310517          04-01-15 17:14         NaN   NYPD
297743    30307738          04-01-15 17:00         NaN   NYPD
297851    30311643          04-01-15 14:19         NaN   NYPD
298344    30305570  03/31/2015 09:39:30 PM         NaN   NYPD
298375    30298328  03/31/2015 08:58:10 PM         NaN   NYPD
298388    30299132  03/31/2015 08:39:36 PM         NaN   NYPD
```

In [18]:

```
print(df.dtypes)
```

```
Unique Key                        int64
Created Date                     object
Closed Date                      object
Agency                           object
Agency Name                      object
Complaint Type                   object
Descriptor                       object
Location Type                    object
Incident Zip                    float64
Incident Address                 object
Street Name                      object
Cross Street 1                   object
Cross Street 2                   object
Intersection Street 1            object
Intersection Street 2            object
Address Type                     object
City                             object
Landmark                         object
Facility Type                    object
Status                           object
Due Date                         object
Resolution Description           object
Resolution Action Updated Date   object
Community Board                  object
Borough                          object
X Coordinate (State Plane)      float64
Y Coordinate (State Plane)      float64
Park Facility Name               object
Park Borough                     object
School Name                      object
School Number                    object
School Region                    object
School Code                      object
School Phone Number              object
School Address                   object
School City                      object
School State                     object
School Zip                       object
School Not Found                 object
School or Citywide Complaint    float64
Vehicle Type                    float64
Taxi Company Borough            float64
Taxi Pick Up Location           float64
Bridge Highway Name              object
Bridge Highway Direction         object
Road Ramp                        object
Bridge Highway Segment           object
Garage Lot Name                 float64
Ferry Direction                  object
Ferry Terminal Name              object
Latitude                        float64
Longitude                       float64
Location                         object
dtype: object
```

In [19]:

```
# 2. Read or convert the columns ‘Created Date’ and Closed Date’ to datetime datatype and c

import datetime as dt
import datetime, time

df['Created Date'] = pd.to_datetime(df['Created Date'])
print(df['Created Date'].dtype)
```

```
datetime64[ns]
```

In [20]:

```
df['Closed Date'] = pd.to_datetime(df['Closed Date'])
print(df['Closed Date'].dtype)
```

```
datetime64[ns]
```

In [21]:

```
df['Request_Closing_Time'] = df['Closed Date'] - df['Created Date']
df['Request_Closing_Time'].head()
```

Out[21]:

```
0   00:55:15
1   01:26:16
2   04:51:31
3   07:45:14
4   03:27:02
Name: Request_Closing_Time, dtype: timedelta64[ns]
```

In [22]:

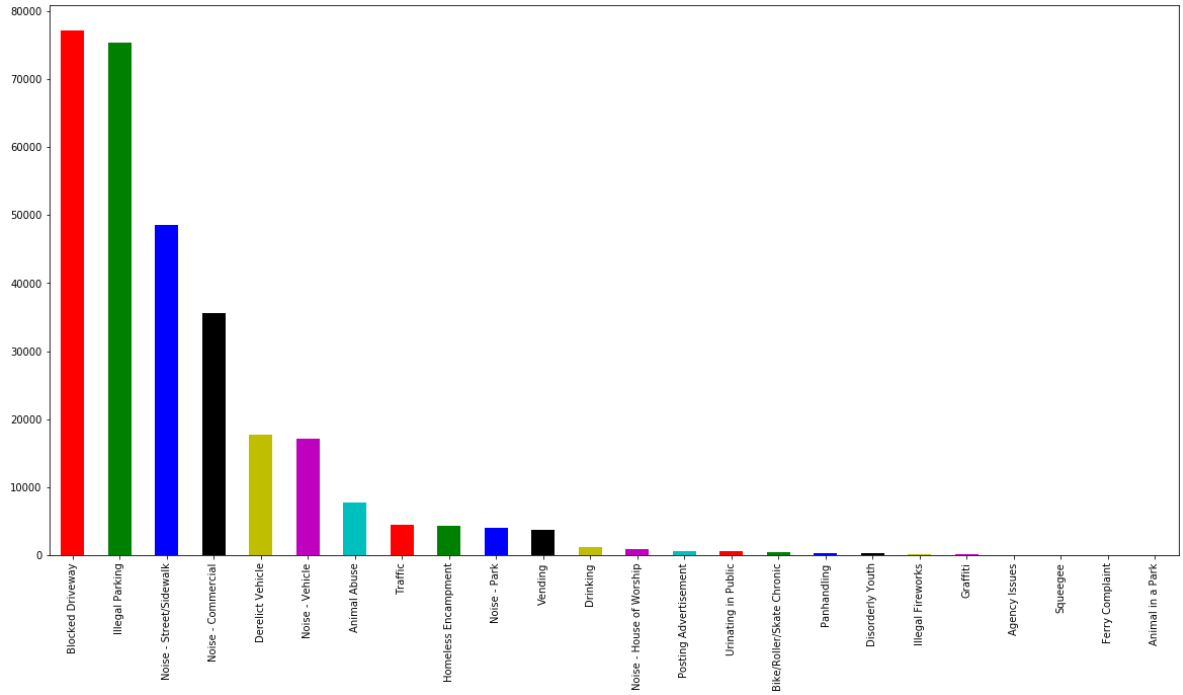
```
# 3. Provide major insights/patterns that you can offer in a visual format (graphs or table
# Conclusion 1: Frequency of each complain type

print(df['Complaint Type'].value_counts())
```

```
Blocked Driveway            77044
Illegal Parking             75361
Noise - Street/Sidewalk     48612
Noise - Commercial          35577
Derelict Vehicle            17718
Noise - Vehicle             17083
Animal Abuse                 7778
Traffic                      4498
Homeless Encampment          4416
Noise - Park                 4042
Vending                      3802
Drinking                     1280
Noise - House of Worship      931
Posting Advertisement         650
Urinating in Public           592
Bike/Roller/Skate Chronic     427
Panhandling                   307
Disorderly Youth              286
Illegal Fireworks             168
Graffiti                      113
Agency Issues                   6
Squeegee                        4
Ferry Complaint                 2
Animal in a Park                1
Name: Complaint Type, dtype: int64
```

In [23]:

```
df['Complaint Type'].value_counts().plot(kind="bar", color=list('rgbkymc'), figsize=(20,10)
plt.show()
```



In [24]:

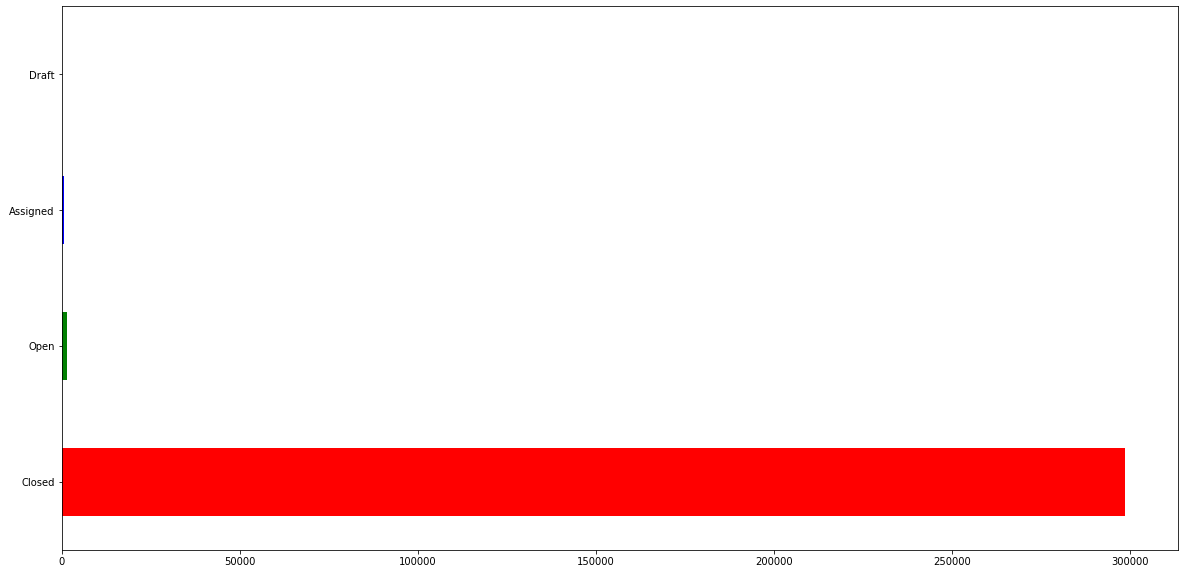
```
# Conclusion 2: Status of Complaints

print(df['Status'].value_counts())
```

```
Closed      298471
Open          1439
Assigned       786
Draft            2
Name: Status, dtype: int64
```

In [30]:

```
df['Status'].value_counts().plot(kind="barh", color=list('rgbkymc'), figsize=(30,10))
plt.show()
```



In [31]:

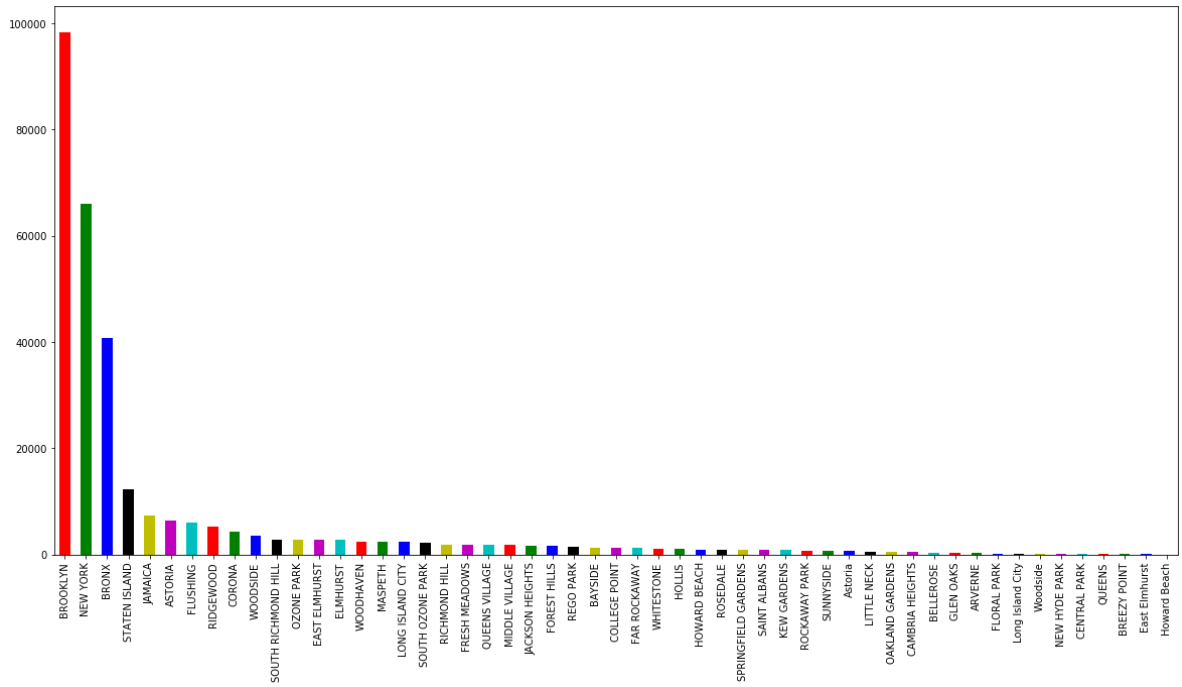
```
# Conclusion 3: Frequency of complaints from different cities

print(df['City'].value_counts())
```

```
BROOKLYN               98307
NEW YORK               65994
BRONX                  40702
STATEN ISLAND          12343
JAMAICA                 7296
ASTORIA                 6330
FLUSHING                5971
RIDGEWOOD               5163
CORONA                  4295
WOODSIDE                3544
SOUTH RICHMOND HILL     2774
OZONE PARK              2755
EAST ELMHURST           2734
ELMHURST                2673
WOODHAVEN               2464
MASPETH                 2462
LONG ISLAND CITY        2437
SOUTH OZONE PARK        2173
RICHMOND HILL           1904
FRESH MEADOWS           1899
QUEENS VILLAGE          1814
MIDDLE VILLAGE          1765
JACKSON HEIGHTS         1689
FOREST HILLS            1688
REGO PARK               1486
BAYSIDE                 1221
COLLEGE POINT           1220
FAR ROCKAWAY            1179
WHITESTONE              1098
HOLLIS                  1012
HOWARD BEACH             931
ROSEDALE                 922
SPRINGFIELD GARDENS      883
SAINT ALBANS             834
KEW GARDENS              771
ROCKAWAY PARK            745
SUNNYSIDE                723
Astoria                  717
LITTLE NECK              559
OAKLAND GARDENS          551
CAMBRIA HEIGHTS          477
BELLEROSE                375
GLEN OAKS                306
ARVERNE                  220
FLORAL PARK              152
Long Island City         134
Woodside                 120
NEW HYDE PARK             98
CENTRAL PARK              97
QUEENS                    32
BREEZY POINT              30
East Elmhurst             14
Howard Beach               1
Name: City, dtype: int64
```

In [32]:

```
df['City'].value_counts().plot(kind="bar", color=list('rgbkymc'), figsize=(20,10))
plt.show()
```



In [33]:

```
# Conclusion 4:

def toHour(timeDel):
    days = timeDel.days
    hours = round(timeDel.seconds/3600, 2)
    result = (days * 24) + hours
    return result


df['Request_Closing_In_Hour'] = df['Request_Closing_Time'].apply(toHour)
print(df['Request_Closing_In_Hour'].head())
```

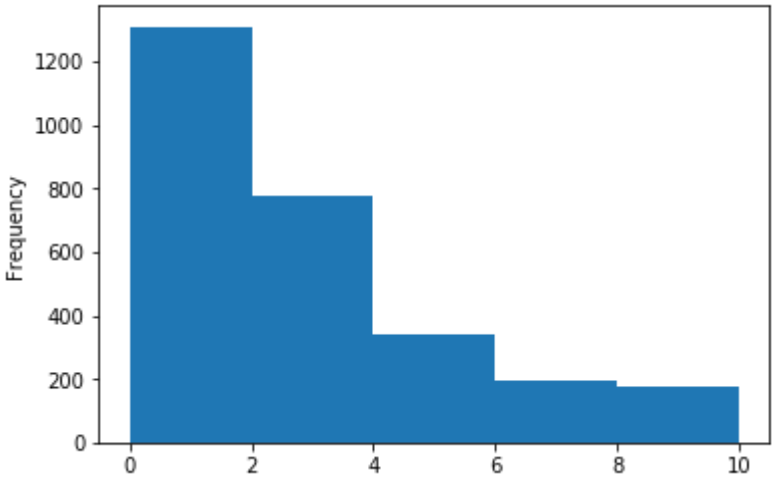
```
0    0.92
1    1.44
2    4.86
3    7.75
4    3.45
Name: Request_Closing_In_Hour, dtype: float64
```

In [34]:

```
print(df['Request_Closing_In_Hour'].mean())

df['Request_Closing_In_Hour'].value_counts().plot(kind='hist',bins=[0,2,4,6,8,10],rwidth=1)
plt.show()
```

```
4.314398862441142
```



In [35]:

```
# Conclusion 5:

months = pd.Series({1: 'Jan', 2: 'Feb', 3: 'Mar', 4: 'Apr', 5: 'May', 6: 'Jun', 7: 'Jul', 8
print(months)
print(months[12])

def getMonth(Date):
    a = str(Date)
    date = datetime.datetime.strptime(a, "%Y-%m-%d %H:%M:%S")
    return months[date.month]
```

```
1     Jan
2     Feb
3     Mar
4     Apr
5     May
6     Jun
7     Jul
8     Aug
9     Sep
10    Oct
11    Nov
12    Dec
dtype: object
Dec
```

In [36]:

```
df['Created_Month'] = df['Created Date'].apply(getMonth)
df['Created_Month']
```

Out[36]:

```
0         Dec
1         Dec
2         Dec
3         Dec
4         Dec
5         Dec
6         Dec
7         Dec
8         Dec
9         Dec
10        Dec
11        Dec
12        Dec
13        Dec
14        Dec
15        Dec
16        Dec
17        Dec
18        Dec
19        Dec
20        Dec
21        Dec
22        Dec
23        Dec
24        Dec
25        Dec
26        Dec
27        Dec
28        Dec
29        Dec
         ...
300668    Mar
300669    Mar
300670    Mar
300671    Mar
300672    Mar
300673    Mar
300674    Mar
300675    Mar
300676    Mar
300677    Mar
300678    Mar
300679    Mar
300680    Mar
300681    Mar
300682    Mar
300683    Mar
300684    Mar
300685    Mar
300686    Mar
300687    Mar
300688    Mar
300689    Mar
300690    Mar
```

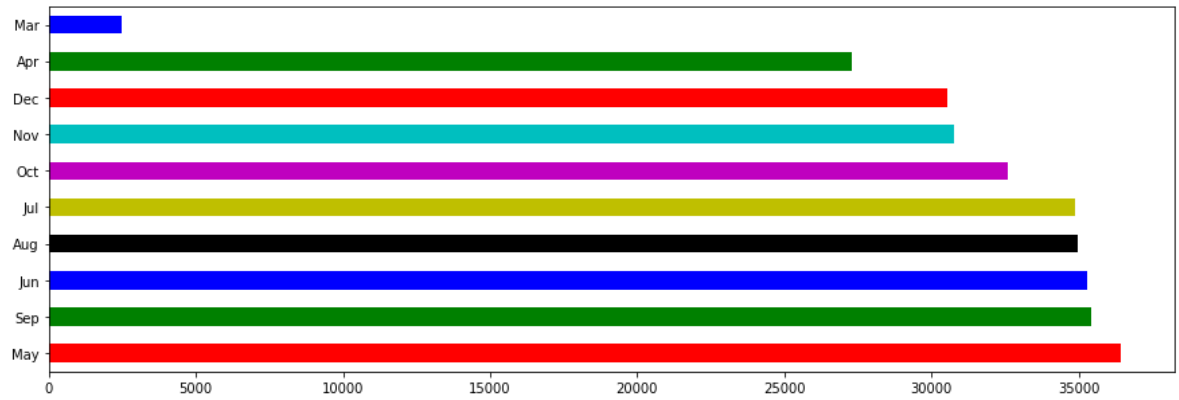
```
300691    Mar
300692    Mar
300693    Mar
300694    Mar
300695    Mar
300696    Mar
300697    Mar
Name: Created_Month, Length: 300698, dtype: object
```

In [37]:

```
print(df['Created_Month'].value_counts())

df['Created_Month'].value_counts().plot(kind="barh", color=list('rgbkymc'), figsize=(15,5))
plt.show()
```

```
May    36437
Sep    35427
Jun    35315
Aug    34956
Jul    34888
Oct    32605
Nov    30773
Dec    30521
Apr    27305
Mar     2471
Name: Created_Month, dtype: int64
```



In [38]:

```
# 4. Order the complaint types based on the average ‘Request_Closing_Time’, grouping them f

print(df['City'].isnull().sum())
```

```
2614
```

In [39]:

```
df['City'].fillna('NA', inplace=True)

print(df['City'].head())
```

```
0    NEW YORK
1     ASTORIA
2       BRONX
3       BRONX
4    ELMHURST
Name: City, dtype: object
```

In [40]:

```
grouped_df=df.groupby(['City', 'Complaint Type'])
```

In [41]:

```
RC_mean = grouped_df.mean()['Request_Closing_In_Hour']
print(RC_mean)
```

```
City       Complaint Type
ARVERNE    Animal Abuse                 2.153158
           Blocked Driveway             2.526000
           Derelict Vehicle             2.968889
           Disorderly Youth             3.595000
           Drinking                     0.240000
           Graffiti                     1.530000
           Homeless Encampment          1.812500
           Illegal Parking              2.316207
           Noise - Commercial           2.285000
           Noise - House of Worship     1.562727
           Noise - Park                 1.285000
           Noise - Street/Sidewalk      1.992759
           Noise - Vehicle              1.860000
           Panhandling                  1.030000
           Urinating in Public          0.690000
           Vending                      0.480000
ASTORIA    Animal Abuse                 5.000640
           Bike/Roller/Skate Chronic    1.740667
           Blocked Driveway             4.816134
           Derelict Vehicle             9.689117
           Disorderly Youth             2.903333
           Drinking                     4.722571
           Graffiti                    14.097500
           Homeless Encampment          4.918750
           Illegal Fireworks            2.772500
           Illegal Parking              4.833399
           Noise - Commercial           3.133039
           Noise - House of Worship     2.022632
           Noise - Park                 2.994754
           Noise - Street/Sidewalk      3.450829
                                         ...
WOODHAVEN  Noise - House of Worship     3.306667
           Noise - Park                 1.380000
           Noise - Street/Sidewalk      5.237907
           Noise - Vehicle              3.403784
           Traffic                      1.833333
           Urinating in Public          3.410000
           Vending                      2.841667
WOODSIDE   Animal Abuse                 8.439710
           Bike/Roller/Skate Chronic   12.150000
           Blocked Driveway             6.473280
           Derelict Vehicle             9.384089
           Disorderly Youth             1.220000
           Drinking                     5.481333
           Graffiti                     8.993333
           Homeless Encampment          6.717879
           Illegal Fireworks            2.470000
           Illegal Parking              7.245926
           Noise - Commercial           6.687990
           Noise - House of Worship     4.740000
           Noise - Park                 6.751842
           Noise - Street/Sidewalk      6.623760
           Noise - Vehicle              5.481714
           Traffic                      4.837436
           Urinating in Public          6.421250
```

```
            Vending                      7.302000
Woodside    Blocked Driveway             6.405455
            Derelict Vehicle             4.965000
            Illegal Parking              5.219500
            Noise - Commercial           2.390000
            Noise - Street/Sidewalk      3.410000
Name: Request_Closing_In_Hour, Length: 782, dtype: float64
```

In [42]:

```
print(RC_mean.isnull().sum())
```

```
4
```

In [43]:

```
grouped_df = df.groupby(['City','Complaint Type']).agg({'Request_Closing_In_Hour': 'mean'})
print(grouped_df)
```

```
                                    Request_Closing_In_Hour
City      Complaint Type
ARVERNE   Animal Abuse                             2.153158
          Blocked Driveway                         2.526000
          Derelict Vehicle                         2.968889
          Disorderly Youth                         3.595000
          Drinking                                 0.240000
          Graffiti                                 1.530000
          Homeless Encampment                      1.812500
          Illegal Parking                          2.316207
          Noise - Commercial                       2.285000
          Noise - House of Worship                 1.562727
          Noise - Park                             1.285000
          Noise - Street/Sidewalk                  1.992759
          Noise - Vehicle                          1.860000
          Panhandling                              1.030000
          Urinating in Public                      0.690000
          Vending                                  0.480000
ASTORIA   Animal Abuse                             5.000640
          Bike/Roller/Skate Chronic                1.740667
          Blocked Driveway                         4.816134
          Derelict Vehicle                         9.689117
          Disorderly Youth                         2.903333
          Drinking                                 4.722571
          Graffiti                                14.097500
          Homeless Encampment                      4.918750
          Illegal Fireworks                        2.772500
          Illegal Parking                          4.833399
          Noise - Commercial                       3.133039
          Noise - House of Worship                 2.022632
          Noise - Park                             2.994754
          Noise - Street/Sidewalk                  3.450829
...                                                     ...
WOODHAVEN Noise - House of Worship                 3.306667
          Noise - Park                             1.380000
          Noise - Street/Sidewalk                  5.237907
          Noise - Vehicle                          3.403784
          Traffic                                  1.833333
          Urinating in Public                      3.410000
          Vending                                  2.841667
WOODSIDE  Animal Abuse                             8.439710
          Bike/Roller/Skate Chronic               12.150000
          Blocked Driveway                         6.473280
          Derelict Vehicle                         9.384089
          Disorderly Youth                         1.220000
          Drinking                                 5.481333
          Graffiti                                 8.993333
          Homeless Encampment                      6.717879
          Illegal Fireworks                        2.470000
          Illegal Parking                          7.245926
          Noise - Commercial                       6.687990
          Noise - House of Worship                 4.740000
          Noise - Park                             6.751842
          Noise - Street/Sidewalk                  6.623760
          Noise - Vehicle                          5.481714
```

```
           Traffic                          4.837436
           Urinating in Public              6.421250
           Vending                          7.302000
Woodside   Blocked Driveway                 6.405455
           Derelict Vehicle                 4.965000
           Illegal Parking                  5.219500
           Noise - Commercial               2.390000
           Noise - Street/Sidewalk          3.410000

[782 rows x 1 columns]
```

In [44]:

```
print(grouped_df[grouped_df['Request_Closing_In_Hour'].isnull()])
```

```
                                Request_Closing_In_Hour
City Complaint Type
NA   Ferry Complaint                                NaN
     Noise - House of Worship                      NaN
     Panhandling                                   NaN
     Posting Advertisement                         NaN
```

In [45]:

```
grouped_df=grouped_df.dropna()

print(grouped_df.isnull().sum())
```

```
Request_Closing_In_Hour    0
dtype: int64
```

In [46]:

```
print(grouped_df)
```

```
                                    Request_Closing_In_Hour
City      Complaint Type
ARVERNE   Animal Abuse                              2.153158
          Blocked Driveway                          2.526000
          Derelict Vehicle                          2.968889
          Disorderly Youth                          3.595000
          Drinking                                  0.240000
          Graffiti                                  1.530000
          Homeless Encampment                       1.812500
          Illegal Parking                           2.316207
          Noise - Commercial                        2.285000
          Noise - House of Worship                  1.562727
          Noise - Park                              1.285000
          Noise - Street/Sidewalk                   1.992759
          Noise - Vehicle                           1.860000
          Panhandling                               1.030000
          Urinating in Public                       0.690000
          Vending                                   0.480000
ASTORIA   Animal Abuse                              5.000640
          Bike/Roller/Skate Chronic                 1.740667
          Blocked Driveway                          4.816134
          Derelict Vehicle                          9.689117
          Disorderly Youth                          2.903333
          Drinking                                  4.722571
          Graffiti                                 14.097500
          Homeless Encampment                       4.918750
          Illegal Fireworks                         2.772500
          Illegal Parking                           4.833399
          Noise - Commercial                        3.133039
          Noise - House of Worship                  2.022632
          Noise - Park                              2.994754
          Noise - Street/Sidewalk                   3.450829
...                                                      ...
WOODHAVEN Noise - House of Worship                  3.306667
          Noise - Park                              1.380000
          Noise - Street/Sidewalk                   5.237907
          Noise - Vehicle                           3.403784
          Traffic                                   1.833333
          Urinating in Public                       3.410000
          Vending                                   2.841667
WOODSIDE  Animal Abuse                              8.439710
          Bike/Roller/Skate Chronic                12.150000
          Blocked Driveway                          6.473280
          Derelict Vehicle                          9.384089
          Disorderly Youth                          1.220000
          Drinking                                  5.481333
          Graffiti                                  8.993333
          Homeless Encampment                       6.717879
          Illegal Fireworks                         2.470000
          Illegal Parking                           7.245926
          Noise - Commercial                        6.687990
          Noise - House of Worship                  4.740000
          Noise - Park                              6.751842
          Noise - Street/Sidewalk                   6.623760
          Noise - Vehicle                           5.481714
          Traffic                                   4.837436
          Urinating in Public                       6.421250
```

```
                Vending                          7.302000
Woodside        Blocked Driveway                 6.405455
                Derelict Vehicle                 4.965000
                Illegal Parking                  5.219500
                Noise - Commercial               2.390000
                Noise - Street/Sidewalk          3.410000

[778 rows x 1 columns]
```

In [48]:

```
sorted_group=grouped_df.sort_values(['City', 'Request_Closing_In_Hour'])
print(sorted_group)
```

```
                                  Request_Closing_In_Hour
City      Complaint Type
ARVERNE   Drinking                                0.240000
          Vending                                 0.480000
          Urinating in Public                     0.690000
          Panhandling                             1.030000
          Noise - Park                            1.285000
          Graffiti                                1.530000
          Noise - House of Worship                1.562727
          Homeless Encampment                     1.812500
          Noise - Vehicle                         1.860000
          Noise - Street/Sidewalk                 1.992759
          Animal Abuse                            2.153158
          Noise - Commercial                      2.285000
          Illegal Parking                         2.316207
          Blocked Driveway                        2.526000
          Derelict Vehicle                        2.968889
          Disorderly Youth                        3.595000
ASTORIA   Panhandling                             1.150000
          Bike/Roller/Skate Chronic               1.740667
          Noise - House of Worship                2.022632
          Illegal Fireworks                       2.772500
          Disorderly Youth                        2.903333
          Noise - Park                            2.994754
          Noise - Commercial                      3.133039
          Noise - Street/Sidewalk                 3.450829
          Noise - Vehicle                         3.509069
          Urinating in Public                     4.626667
          Drinking                                4.722571
          Blocked Driveway                        4.816134
          Illegal Parking                         4.833399
          Homeless Encampment                     4.918750
...                                                    ...
WOODHAVEN Noise - Commercial                      3.891543
          Animal Abuse                            4.967333
          Noise - Street/Sidewalk                 5.237907
          Blocked Driveway                        5.522871
          Illegal Parking                         5.729018
          Homeless Encampment                     7.106667
          Derelict Vehicle                        7.450162
WOODSIDE  Disorderly Youth                        1.220000
          Illegal Fireworks                       2.470000
          Noise - House of Worship                4.740000
          Traffic                                 4.837436
          Drinking                                5.481333
          Noise - Vehicle                         5.481714
          Urinating in Public                     6.421250
          Blocked Driveway                        6.473280
          Noise - Street/Sidewalk                 6.623760
          Noise - Commercial                      6.687990
          Homeless Encampment                     6.717879
          Noise - Park                            6.751842
          Illegal Parking                         7.245926
          Vending                                 7.302000
          Animal Abuse                            8.439710
          Graffiti                                8.993333
```

```
           Derelict Vehicle                 9.384089
           Bike/Roller/Skate Chronic       12.150000
Woodside   Noise - Commercial               2.390000
           Noise - Street/Sidewalk          3.410000
           Derelict Vehicle                 4.965000
           Illegal Parking                  5.219500
           Blocked Driveway                 6.405455

[778 rows x 1 columns]
```

In [49]:

```
#5. Perform a statistical test for the following:
#Please note: For the below statements you need to state the Null and Alternate and then pr

#Whether the average response time across complaint types is similar or not (overall)
#Are the type of complaint or service requested and location related?

import scipy.stats as stats
from math import sqrt

# Since we have to compare average of more than two variables. Therefore, we use ANOVA for

# Null hypothesis H0 : All Complain Types average response time mean is similar
# Alternate hypothesis H1 : All Complain Types average response time mean is Not similar

print(df['Complaint Type'].value_counts())
```

```
Blocked Driveway             77044
Illegal Parking              75361
Noise - Street/Sidewalk      48612
Noise - Commercial           35577
Derelict Vehicle             17718
Noise - Vehicle              17083
Animal Abuse                  7778
Traffic                       4498
Homeless Encampment           4416
Noise - Park                  4042
Vending                       3802
Drinking                      1280
Noise - House of Worship       931
Posting Advertisement          650
Urinating in Public            592
Bike/Roller/Skate Chronic      427
Panhandling                    307
Disorderly Youth               286
Illegal Fireworks              168
Graffiti                       113
Agency Issues                    6
Squeegee                         4
Ferry Complaint                  2
Animal in a Park                 1
Name: Complaint Type, dtype: int64
```

In [50]:

```
top_complaints_type = df['Complaint Type'].value_counts()[:5]
print(top_complaints_type)
```

```
Blocked Driveway           77044
Illegal Parking            75361
Noise - Street/Sidewalk    48612
Noise - Commercial         35577
Derelict Vehicle           17718
Name: Complaint Type, dtype: int64
```

In [51]:

```
top_complaints_type_names = top_complaints_type.index
print(top_complaints_type_names)
```

```
Index(['Blocked Driveway', 'Illegal Parking', 'Noise - Street/Sidewalk',
       'Noise - Commercial', 'Derelict Vehicle'],
      dtype='object')
```

In [52]:

```
sample = df.loc[df['Complaint Type'].isin(top_complaints_type_names), ['Complaint Type', 'R
print(sample.head())
```

```
            Complaint Type  Request_Closing_In_Hour
0  Noise - Street/Sidewalk                     0.92
1         Blocked Driveway                     1.44
2         Blocked Driveway                     4.86
3          Illegal Parking                     7.75
4          Illegal Parking                     3.45
```

In [67]:

```
print(sample.shape)

print(sample.isnull().sum())
```

```
(254312, 2)
Complaint Type                0
Request_Closing_In_Hour    2059
dtype: int64
```

In [68]:

```
sample.dropna(inplace=True)
print(sample.isnull().sum())
```

```
Complaint Type             0
Request_Closing_In_Hour    0
dtype: int64
```

In [70]:

```
set1 = sample[sample['Complaint Type'] == top_complaints_type_names[1]].Request_Closing_In_
print(set1.head())
```

```
3    7.75
4    3.45
5    1.89
6    1.96
8    8.55
Name: Request_Closing_In_Hour, dtype: float64
```

In [71]:

```
set2 = sample[sample['Complaint Type'] == top_complaints_type_names[2]].Request_Closing_In_
print(set2.head())
```

```
0     0.92
12    2.48
19    0.78
38    0.49
54    1.50
Name: Request_Closing_In_Hour, dtype: float64
```

In [72]:

```
set3 = sample[sample['Complaint Type'] == top_complaints_type_names[3]].Request_Closing_In_
print(set3.head())
```

```
17    0.85
18    2.93
22    1.26
29    2.50
30    1.99
Name: Request_Closing_In_Hour, dtype: float64
```

In [73]:

```
set4 = sample[sample['Complaint Type'] == top_complaints_type_names[4]].Request_Closing_In_
print(set4.head())
```

```
14     10.49
151     3.95
255     1.36
256     4.13
295     0.75
Name: Request_Closing_In_Hour, dtype: float64
```

In [74]:

```
set5 = sample[sample['Complaint Type'] == top_complaints_type_names[0]].Request_Closing_In_
print(set5.head())
```

```
1     1.44
2     4.86
7     1.80
9     1.38
10    7.80
Name: Request_Closing_In_Hour, dtype: float64
```

In [75]:

```
stats.f_oneway(set1, set2, set3, set4, set5)
```

Out[75]:

```
F_onewayResult(statistic=1799.5986832389517, pvalue=0.0)
```

In [62]:

```
# Since, pvalue<0.05 we reject null hypothesis.
# Therefore, All Complain Types average response time mean is Not similar

# Try ChiSquare Test for part2  -  Are the type of complaint or service requested and locat

# Null Hypothesis H0 : Complain Type and Location is not related
# Alternate Hypothesis H1 : Complain Type and Location is related

from scipy.stats import chi2_contingency
top_location = df['City'].value_counts()[:5]
print(top_location)
```

```
BROOKLYN         98307
NEW YORK         65994
BRONX            40702
STATEN ISLAND    12343
JAMAICA           7296
Name: City, dtype: int64
```

In [63]:

```
top_location_names = top_location.index
print(top_location_names)
```

```
Index(['BROOKLYN', 'NEW YORK', 'BRONX', 'STATEN ISLAND', 'JAMAICA'], dtype
='object')
```

In [64]:

```
sample2 = df.loc[(df['Complaint Type'].isin(top_complaints_type_names)) & (df['City'].isin(
print(sample2.head())
```

```
            Complaint Type      City
0  Noise - Street/Sidewalk  NEW YORK
2          Blocked Driveway     BRONX
3           Illegal Parking     BRONX
5           Illegal Parking  BROOKLYN
6           Illegal Parking  NEW YORK
```

In [65]:

```
C_table=pd.crosstab(sample2['Complaint Type'], sample2['City'])
print(C_table)
```

```
City                     BRONX  BROOKLYN  JAMAICA  NEW YORK  STATEN ISLAND
Complaint Type
Blocked Driveway         12755     28148     2818      2072           2142
Derelict Vehicle          1953      5181      954       537           1766
Illegal Parking           7859     27462     1421     12128           4886
Noise - Commercial        2434     11463      429     14550            678
Noise - Street/Sidewalk   8892     13356      339     20433            819
```

In [66]:

```
ch2, p, dof, tb1 = chi2_contingency(C_table)

print(ch2,p,dof)

# Since, p<0.05 we reject null hypothesis.
# Therefore, Complain Type and Location is related
```

```
40522.79928349593 0.0 16
```

In [78]:

```
if(p<0.05):
    print("""Reject Null Hypothesis.
             Complain Type and location is related.""")
else:
     print("""Reject Null Hypothesis.
             Complain Type and location is not related.""")
```

```
Reject Null Hypothesis.
             Complain Type and location is related.
```

In [ ]:

```

```