customer-service-analysis

March 13, 2021

[1]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
```

[2]:
```
df_nyc_requests = pd.read_csv("311_Service_Requests_from_2010_to_Present.csv")
```

```
/home/satheesh/anaconda3/lib/python3.8/site-
packages/IPython/core/interactiveshell.py:3146: DtypeWarning: Columns (48,49)
have mixed types.Specify dtype option on import or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

[3]:
```
df_nyc_requests.columns
```

[3]:
```
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')
```

[4]:
```
df_nyc_requests.head(5)
```

[4]:

	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]
```

[5]:
```
df_nyc_requests.shape
```

[5]: (300698, 53)

[6]:
```
df_nyc_requests.isnull()
```

[6]:
```
        Unique Key  Created Date  Closed Date  Agency  Agency Name  \
0            False         False        False   False        False
1            False         False        False   False        False
2            False         False        False   False        False
3            False         False        False   False        False
4            False         False        False   False        False
…              …             …            …       …            …
300693       False         False         True   False        False
300694       False         False        False   False        False
300695       False         False        False   False        False
300696       False         False        False   False        False
300697       False         False        False   False        False

        Complaint Type  Descriptor  Location Type  Incident Zip  \
0                False       False          False         False
1                False       False          False         False
2                False       False          False         False
3                False       False          False         False
4                False       False          False         False
…                  …           …              …             …
300693           False       False          False          True
300694           False       False          False         False
300695           False       False          False         False
300696           False       False          False         False
300697           False       False          False         False

        Incident Address  …  Bridge Highway Name  Bridge Highway Direction  \
0                  False  …                 True                      True
1                  False  …                 True                      True
2                  False  …                 True                      True
3                  False  …                 True                      True
4                  False  …                 True                      True
…                    …    …                  …                         …
300693             False  …                 True                      True
300694             False  …                 True                      True
300695             False  …                 True                      True
300696             False  …                 True                      True
300697             False  …                 True                      True

        Road Ramp  Bridge Highway Segment  Garage Lot Name  Ferry Direction  \
0            True                    True             True             True
1            True                    True             True             True
2            True                    True             True             True
```

```
3            True                    True             True             True
4            True                    True             True             True
…             …                       …                …                …
300693       True                    True             True             True
300694       True                    True             True             True
300695       True                    True             True             True
300696       True                    True             True             True
300697       True                    True             True             True

        Ferry Terminal Name  Latitude  Longitude  Location
0                      True     False      False     False
1                      True     False      False     False
2                      True     False      False     False
3                      True     False      False     False
4                      True     False      False     False
…                       …        …          …         …
300693                 True      True       True      True
300694                 True     False      False     False
300695                 True     False      False     False
300696                 True     False      False     False
300697                 True     False      False     False

[300698 rows x 53 columns]
```

[7]:
```
#counting the null values that have been observed per column
df_nyc_requests.isnull().sum()
```

[7]:
```
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
```

0.1 Conclusion #1:

It is evident that the columsn "School or Citywide Complaint", "Vehicle Type", "Taxi Company Borough", "Taxi Pick Up Location", "Garage Lot Name" have no data which can be used for analysis. This is because all of their values are null

The other columns such as "Ferry Direction", "Ferry Terminal Name" have only 1 or 2 records

It's safe to assume that these columns can be dropped and we should be able to proceed with our data analysis

Analysis around Bridge, Landmark and Road ramp for meaningful data The data from the columns "Bridge Highway Name", "Bridge Highway Direction", "Bridge Highway segment",

"Landmark", "Road ramp" all have data that is less than 1% of data. So it woul be ideal to drop these columns

0.2 Dropping the columns described above

[8]:
```
df_nyc_requests.drop(['Landmark', '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='columns', inplace=True)
```

[9]:
```
#verifying the shape after dropping the columns
df_nyc_requests.shape
```

[9]:
```
(300698, 41)
```

[10]:
```
df_nyc_requests.columns
```

[10]:
```
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', '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',
       'Latitude', 'Longitude', 'Location'],
      dtype='object')
```

[11]:
```
#verifying the first data point after dropping columns
df_nyc_requests.iloc[0]
```

[11]:
```
Unique Key                                                      32310363
Created Date                                         12/31/2015 11:59:45
PM
Closed Date                                                  01-01-16
0:55
Agency                                                           NYPD
Agency Name                                  New York City Police
Department
Complaint Type                                          Noise -
```

Street/Sidewalk
Descriptor Loud
Music/Party
Location Type
Street/Sidewalk
Incident Zip
10034
Incident Address 71 VERMILYEA
AVENUE
Street Name VERMILYEA
AVENUE
Cross Street 1 ACADEMY
STREET
Cross Street 2 WEST 204
STREET
Intersection Street 1
NaN
Intersection Street 2
NaN
Address Type
ADDRESS
City NEW
YORK
Facility Type
Precinct
Status
Closed
Due Date 01-01-16
7:59
Resolution Description The Police Department responded and upon
arriv…
Resolution Action Updated Date 01-01-16
0:55
Community Board 12
MANHATTAN
Borough
MANHATTAN
X Coordinate (State Plane)
1.00541e+06
Y Coordinate (State Plane)
254678
Park Facility Name
Unspecified
Park Borough
MANHATTAN
School Name
Unspecified

```
School Number
Unspecified
School Region
Unspecified
School Code
Unspecified
School Phone Number
Unspecified
School Address
Unspecified
School City
Unspecified
School State
Unspecified
School Zip
Unspecified
School Not Found
N
Latitude
40.8657
Longitude
-73.9235
Location                                    (40.86568153633767,
-73.92350095571744)
Name: 0, dtype: object
```

[12]:
```
#checking if school name, number region etc is always unspecified
df_nyc_requests["School Name"].value_counts()
```

[12]:
```
Unspecified                          300697
Alley Pond Park - Nature Center           1
Name: School Name, dtype: int64
```

[13]:
```
df_nyc_requests["School Number"].value_counts()
```

[13]:
```
Unspecified    300697
Q001                1
Name: School Number, dtype: int64
```

[14]:
```
df_nyc_requests["School Region"].value_counts()
```

[14]:
```
Unspecified    300697
Name: School Region, dtype: int64
```

[15]:
```
df_nyc_requests["School Code"].value_counts()
```

[15]:
```
Unspecified    300697
Name: School Code, dtype: int64
```

[16]:
```
df_nyc_requests["School Phone Number"].value_counts()
```

[16]:
```
Unspecified    300697
7182176034          1
Name: School Phone Number, dtype: int64
```

[17]:
```
df_nyc_requests["School Address"].value_counts()
```

[17]:
```
Unspecified                                    300697
Grand Central Parkway, near the soccer field        1
Name: School Address, dtype: int64
```

[18]:
```
df_nyc_requests["School City"].value_counts()
```

[18]:
```
Unspecified    300697
QUEENS              1
Name: School City, dtype: int64
```

[19]:
```
df_nyc_requests["School State"].value_counts()
```

[19]:
```
Unspecified    300697
NY                  1
Name: School State, dtype: int64
```

[20]:
```
df_nyc_requests["School Zip"].value_counts()
```

[20]:
```
Unspecified    300697
Name: School Zip, dtype: int64
```

[21]:
```
#checking for the type of crime that is committed near/ at a school
school_complaints = df_nyc_requests[df_nyc_requests["School Name"]!
 ↪="Unspecified"]
school_complaints
```

[21]:
```
        Unique Key            Created Date     Closed Date Agency  \
283132    30427220  04/18/2015 09:44:55 AM  05-02-15 10:35   NYPD

                            Agency Name    Complaint Type    Descriptor  \
283132  New York City Police Department  Animal in a Park  Animal Waste

       Location Type  Incident Zip Incident Address  … School Code  \
283132          Park           NaN              NaN  …         NaN

       School Phone Number                                School Address  \
```

```
283132          7182176034  Grand Central Parkway, near the soccer field

       School City School State School Zip School Not Found Latitude  \
283132      QUEENS           NY        NaN                N      NaN

       Longitude Location
283132       NaN      NaN

[1 rows x 41 columns]
```

[22]:
```
school_complaints.iloc[0]
```

[22]:
```
Unique Key
30427220
Created Date                                                  04/18/2015 09:44:55
AM
Closed Date                                                          05-02-15
10:35
Agency
NYPD
Agency Name                                        New York City Police
Department
Complaint Type                                                 Animal in a
Park
Descriptor                                                        Animal
Waste
Location Type
Park
Incident Zip
NaN
Incident Address
NaN
Street Name
NaN
Cross Street 1
NaN
Cross Street 2
NaN
Intersection Street 1
NaN
Intersection Street 2
NaN
Address Type
NaN
City
QUEENS
Facility Type
```

NaN
Status
Closed
Due Date 05-02-15 9:44
Resolution Description The condition was determined to be an issue ap…
Resolution Action Updated Date 05-02-15 10:35
Community Board 0
Unspecified
Borough
Unspecified
X Coordinate (State Plane)
NaN
Y Coordinate (State Plane)
NaN
Park Facility Name Alley Pond Park - Nature Center
Park Borough
Unspecified
School Name Alley Pond Park - Nature Center
School Number
Q001
School Region
NaN
School Code
NaN
School Phone Number
7182176034
School Address Grand Central Parkway, near the soccer field
School City
QUEENS
School State
NY
School Zip
NaN
School Not Found
N
Latitude
NaN
Longitude
NaN
Location
NaN

```
Name: 283132, dtype: object
```

0.3 Conclusion #2

There is only 1 occurence of a complaint which is associated with a school in Queens. It's safe to assume that the New York Police department has been keeping all school areas well guarded/ protected

[23]:
```
#Examining the Unspecified for Park Facility Name
df_nyc_requests["Park Facility Name"].value_counts()
```

[23]:
```
Unspecified                          300697
Alley Pond Park - Nature Center           1
Name: Park Facility Name, dtype: int64
```

[24]:
```
#keeping this as a separate data set for understanding the nature of the␣
↪complaint
park_complaints = df_nyc_requests[df_nyc_requests["Park Facility Name"]!=␣
↪"Unspecified"]
```

[25]:
```
park_complaints.iloc[0]
```

[25]:
```
Unique Key
30427220
Created Date                                                  04/18/2015 09:44:55
AM
Closed Date                                                         05-02-15
10:35
Agency
NYPD
Agency Name                                          New York City Police
Department
Complaint Type                                                  Animal in a
Park
Descriptor                                                         Animal
Waste
Location Type
Park
Incident Zip
NaN
Incident Address
NaN
Street Name
NaN
Cross Street 1
NaN
Cross Street 2
```

NaN
Intersection Street 1
NaN
Intersection Street 2
NaN
Address Type
NaN
City
QUEENS
Facility Type
NaN
Status
Closed
Due Date 05-02-15
9:44
Resolution Description The condition was determined to be an issue
ap…
Resolution Action Updated Date 05-02-15
10:35
Community Board 0
Unspecified
Borough
Unspecified
X Coordinate (State Plane)
NaN
Y Coordinate (State Plane)
NaN
Park Facility Name Alley Pond Park - Nature
Center
Park Borough
Unspecified
School Name Alley Pond Park - Nature
Center
School Number
Q001
School Region
NaN
School Code
NaN
School Phone Number
7182176034
School Address Grand Central Parkway, near the soccer
field
School City
QUEENS
School State
NY

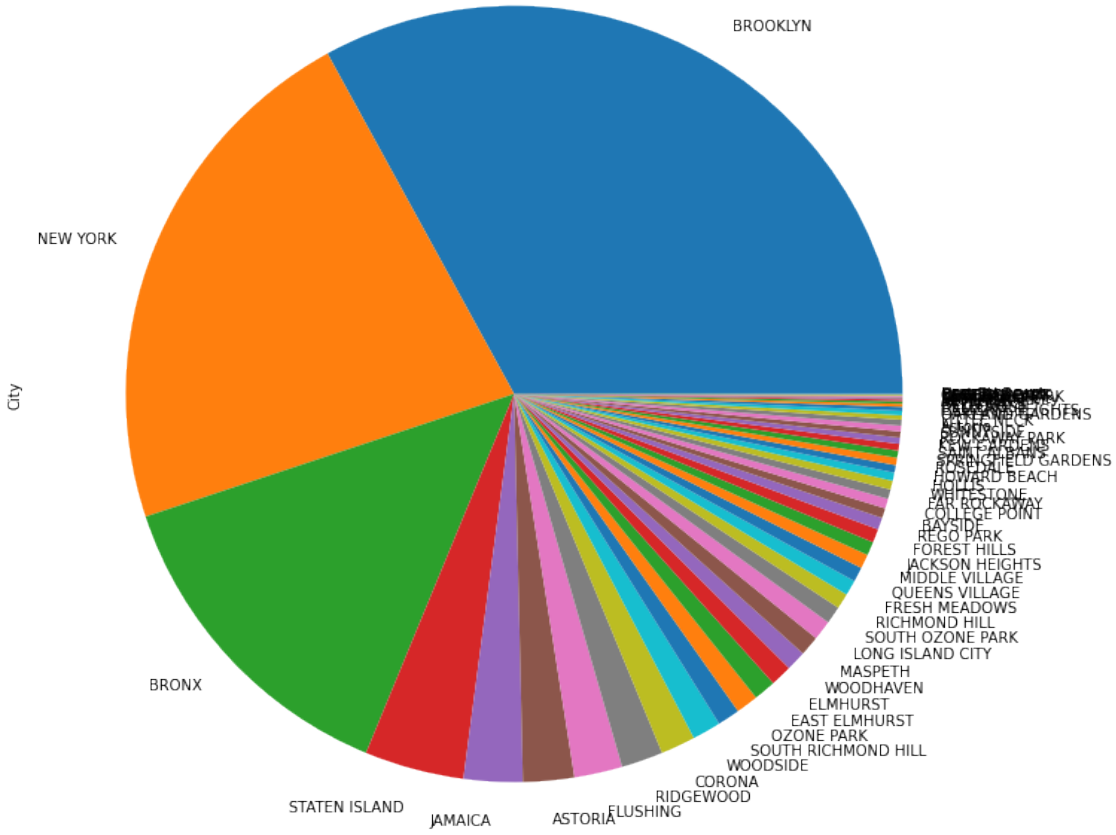
```
School Zip
NaN
School Not Found
N
Latitude
NaN
Longitude
NaN
Location
NaN
Name: 283132, dtype: object
```

The complaint from the school is of the park and there are no other part related complaints where the facily name is explicitly mentioned

0.4 Examining distribution of complaints

[26]:
```
#Distribution by city
df_nyc_requests["City"].value_counts().plot(kind='pie', figsize =(12,12))
```

[26]:
```
<AxesSubplot:ylabel='City'>
```



0.4.1 Conclusion #3

Brooklyn, New York, Bronx and Staten Island contribute to over 70% of the total complaints received by the New York Police Department. This could be owing to the larger amount of population in these areas

0.5 Examining complaints based on date

[27]:
```
#converting the available dates to date-time format
df_nyc_requests["Created Date"] = pd.to_datetime(df_nyc_requests["Created␣
↪Date"])
df_nyc_requests["Closed Date"] = pd.to_datetime(df_nyc_requests["Closed Date"])
df_nyc_requests["Due Date"] = pd.to_datetime(df_nyc_requests["Due Date"])
df_nyc_requests["Resolution Action Updated Date"] = pd.
↪to_datetime(df_nyc_requests["Resolution Action Updated Date"])
```

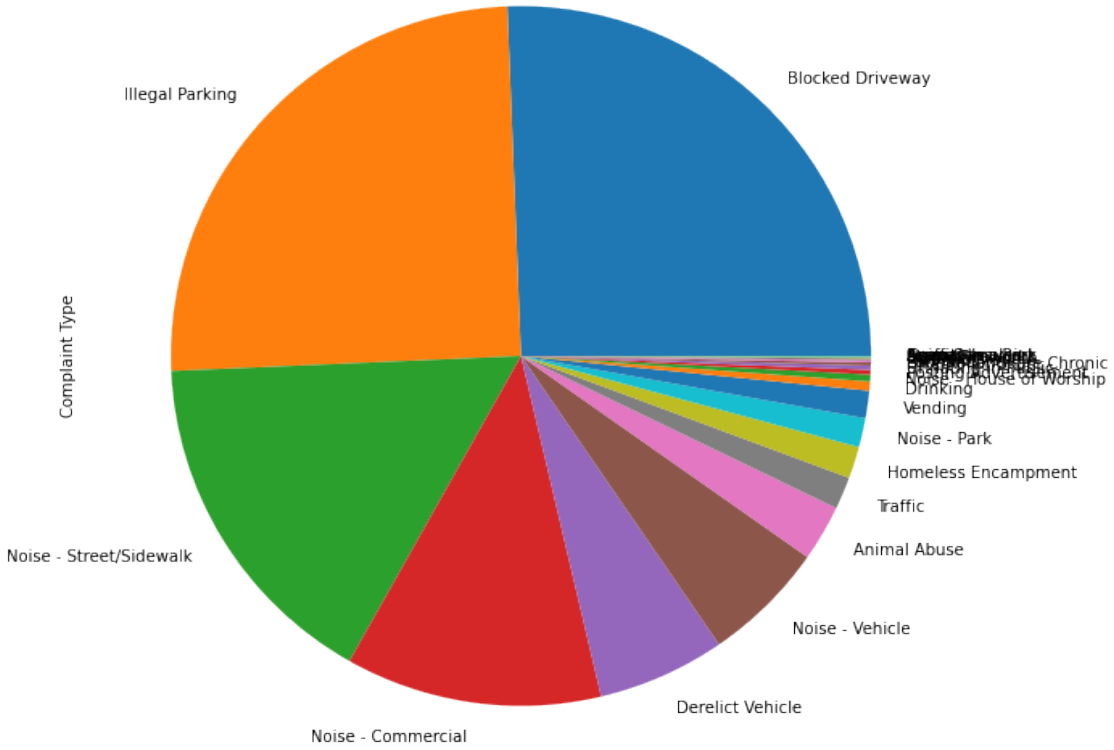
[28]:
```
df_nyc_requests.iloc[0]
```

[28]:
```
Unique Key                                                      32310363
Created Date                                                2015-12-31 23:59:45
Closed Date                                                 2016-01-01 00:55:00
Agency                                                             NYPD
Agency Name                                    New York City Police Department
Complaint Type                                                 Noise - Street/Sidewalk
Descriptor                                                        Loud Music/Party
Location Type                                                  Street/Sidewalk
Incident Zip                                                        10034
Incident Address                                            71 VERMILYEA AVENUE
Street Name                                                    VERMILYEA AVENUE
Cross Street 1                                                   ACADEMY STREET
Cross Street 2                                                  WEST 204 STREET
Intersection Street 1                                               NaN
Intersection Street 2                                               NaN
Address Type                                                    ADDRESS
City                                                                NEW YORK
Facility Type                                                      NaN
Precinct                                                            NaN
Status                                                           Closed
Due Date                                                    2016-01-01 07:59:00
Resolution Description          The Police Department responded and upon arriv…
Resolution Action Updated Date                              2016-01-01 00:55:00
Community Board                                                  12 12 ADULT
```

```
MANHATTAN
Borough
MANHATTAN
X Coordinate (State Plane)
1.00541e+06
Y Coordinate (State Plane)
254678
Park Facility Name
Unspecified
Park Borough
MANHATTAN
School Name
Unspecified
School Number
Unspecified
School Region
Unspecified
School Code
Unspecified
School Phone Number
Unspecified
School Address
Unspecified
School City
Unspecified
School State
Unspecified
School Zip
Unspecified
School Not Found
N
Latitude
40.8657
Longitude
-73.9235
Location                                    (40.86568153633767,
-73.92350095571744)
Name: 0, dtype: object
```

[29]:
```
#complaint types split up
df_nyc_requests["Complaint Type"].value_counts().plot(kind='pie',␣
↪figsize=(10,10))
```

[29]:
```
<AxesSubplot:ylabel='Complaint Type'>
```



0.6 Conclusion #4

Blocked Driveway and Illegal Parking forms over 50% of the total complaints to the NYPD which is an indication that parking is generally a huge problem.

The top 4 complaint areas "Blocked Driveway", "Illegal Parking", "Noise - Street/Sidewalk", "Noise - Commercial" form over 75% of the complaints which is an indication that Noise is the second largest contributor towards complaints to the NYPD

[30]:
```
#Creating a column to find out resolution time
df_nyc_requests["Expected Resolution Time"] = df_nyc_requests["Due Date"] -␣
↪df_nyc_requests["Created Date"]
df_nyc_requests["Actual Resolution Time"] = df_nyc_requests["Closed Date"] -␣
↪df_nyc_requests["Created Date"]
```

[31]:
```
df_nyc_requests.iloc[3]
```

[31]:
```
Unique Key                      32305098
Created Date                    2015-12-31
```

23:57:46
Closed Date 2016-01-01
07:43:00
Agency
NYPD
Agency Name New York City Police
Department
Complaint Type Illegal
Parking
Descriptor Commercial Overnight
Parking
Location Type
Street/Sidewalk
Incident Zip
10461
Incident Address 2940 BAISLEY
AVENUE
Street Name BAISLEY
AVENUE
Cross Street 1 EDISON
AVENUE
Cross Street 2 B
STREET
Intersection Street 1
NaN
Intersection Street 2
NaN
Address Type
ADDRESS
City
BRONX
Facility Type
Precinct
Status
Closed
Due Date 2016-01-01
07:57:00
Resolution Description The Police Department responded to the
complai…
Resolution Action Updated Date 2016-01-01
07:43:00
Community Board 10
BRONX
Borough
BRONX
X Coordinate (State Plane)
1.03174e+06

```
Y Coordinate (State Plane)
243899
Park Facility Name
Unspecified
Park Borough
BRONX
School Name
Unspecified
School Number
Unspecified
School Region
Unspecified
School Code
Unspecified
School Phone Number
Unspecified
School Address
Unspecified
School City
Unspecified
School State
Unspecified
School Zip
Unspecified
School Not Found
N
Latitude
40.836
Longitude
-73.8284
Location                                (40.83599404683083,
-73.82837939584206)
Expected Resolution Time                                    0 days
07:59:14
Actual Resolution Time                                      0 days
07:45:14
Name: 3, dtype: object
```

[32]:
```
#Creating an SLA breach column to understand if the SLA has been breached
df_nyc_requests["SLA Breach"] = df_nyc_requests["Actual Resolution␣
↪Time"]>df_nyc_requests["Expected Resolution Time"]
```

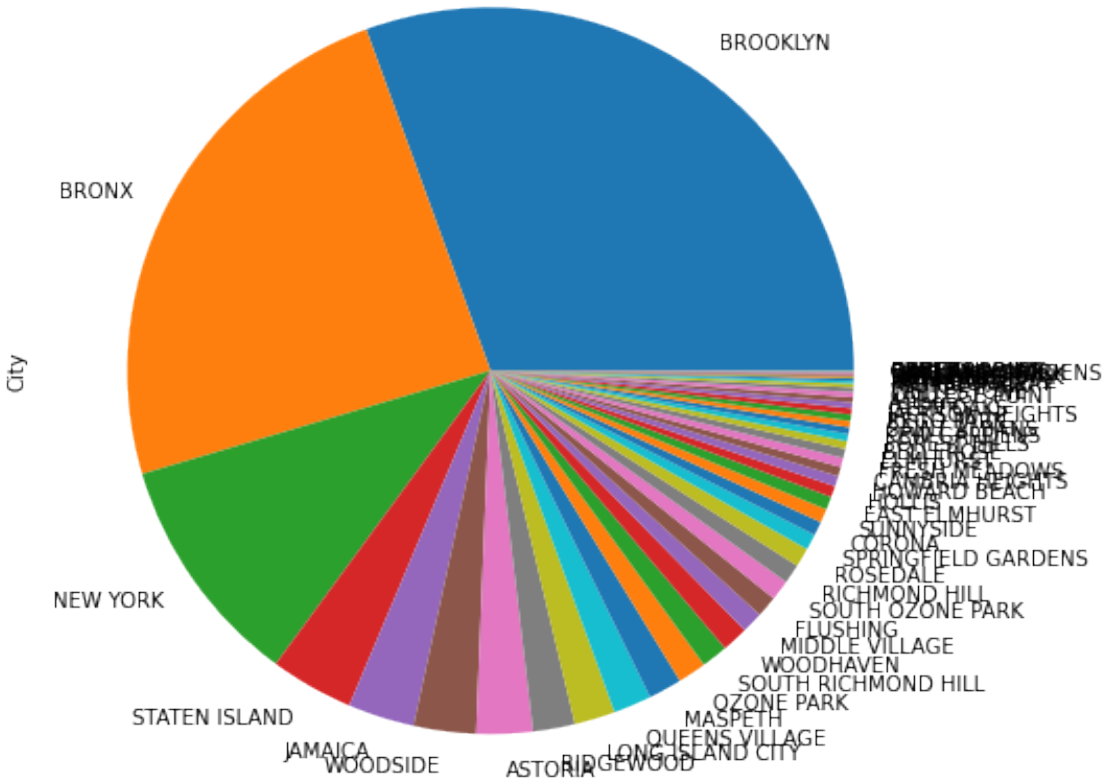
[33]:
```
df_nyc_requests["SLA Breach"].value_counts()
```

[33]:
```
False    262118
True      38580
Name: SLA Breach, dtype: int64
```

[34]:
```
df_sla_breach = df_nyc_requests[df_nyc_requests["SLA Breach"]==True]
```

[35]:
```
df_sla_breach["City"].value_counts().plot(kind='pie', figsize=(8,8))
```

[35]: <AxesSubplot:ylabel='City'>



0.7 Conclusion #5

Brooklyn, Bronx and New York form the top 60% of the SLA breaches where the resolution time that was expected has been exceeded

[36]:
```
df_nyc_requests.columns
```

[36]:
```
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', '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',
       'Latitude', 'Longitude', 'Location', 'Expected Resolution Time',
       'Actual Resolution Time', 'SLA Breach'],
      dtype='object')
```

[37]:
```
df_nyc_requests.iloc[12]
```

[37]:
```
Unique Key                                                    32306612
Created Date                                                  2015-12-31
23:48:03
Closed Date                                                   2016-01-01
02:17:00
Agency                                                            NYPD
Agency Name                                    New York City Police
Department
Complaint Type                                          Noise -
Street/Sidewalk
Descriptor                                                        Loud
Music/Party
Location Type
Street/Sidewalk
Incident Zip
10461
Incident Address                                              1701 PILGRIM
AVENUE
Street Name                                                       PILGRIM
AVENUE
Cross Street 1                                                    ROBERTS
AVENUE
Cross Street 2                                                WESTCHESTER
AVENUE
Intersection Street 1
NaN
Intersection Street 2
NaN
Address Type
ADDRESS
City
BRONX
Facility Type
Precinct
```

```
Status
Closed
Due Date                                                    2016-01-01
07:48:00
Resolution Description           The Police Department responded to the
complai…
Resolution Action Updated Date                              2016-01-01
02:18:00
Community Board                                                      10
BRONX
Borough
BRONX
X Coordinate (State Plane)
1.03029e+06
Y Coordinate (State Plane)
247376
Park Facility Name
Unspecified
Park Borough
BRONX
School Name
Unspecified
School Number
Unspecified
School Region
Unspecified
School Code
Unspecified
School Phone Number
Unspecified
School Address
Unspecified
School City
Unspecified
School State
Unspecified
School Zip
Unspecified
School Not Found
N
Latitude
40.8455
Longitude
-73.8336
Location                              (40.845545043640215,
-73.83358471831198)
Expected Resolution Time                                     0 days
```

```
07:59:57
Actual Resolution Time                                          0 days
02:28:57
SLA Breach
False
Name: 12, dtype: object
```

[38]:
```
df_nyc_requests["Facility Type"].value_counts()
```

[38]:
```
Precinct    298527
Name: Facility Type, dtype: int64
```

0.8 Resolution time and Location Analysis

[39]:
```
#creating a new dataframe with only location related information and resolution␣
 ↪times for complaints
df_nyc_location = df_nyc_requests[["Unique Key", "Agency Name", "Complaint␣
 ↪Type", "Descriptor", "Location Type", "Incident Zip", "Incident Address",␣
 ↪"Street Name", "Cross Street 1", "Cross Street 2", "Address Type" , "City",␣
 ↪"Status", "Actual Resolution Time"]]
```

[40]:
```
df_nyc_location.iloc[0]
```

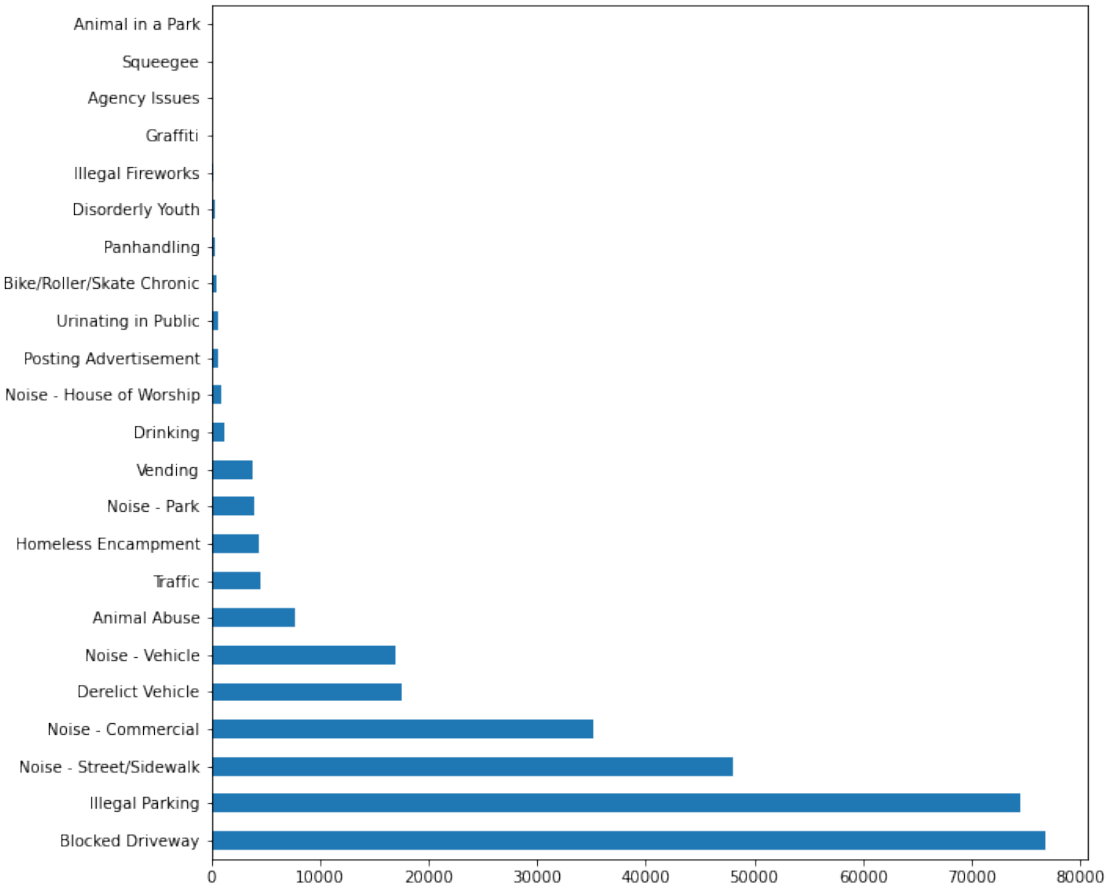
[40]:
```
Unique Key                                    32310363
Agency Name                New York City Police Department
Complaint Type                     Noise - Street/Sidewalk
Descriptor                               Loud Music/Party
Location Type                             Street/Sidewalk
Incident Zip                                        10034
Incident Address                    71 VERMILYEA AVENUE
Street Name                            VERMILYEA AVENUE
Cross Street 1                           ACADEMY STREET
Cross Street 2                          WEST 204 STREET
Address Type                                    ADDRESS
City                                           NEW YORK
Status                                           Closed
Actual Resolution Time                0 days 00:55:15
Name: 0, dtype: object
```

[41]:
```
#dropping not closed complaints for analysis purposes and examining the␣
 ↪dataframe shape
df_nyc_location = df_nyc_location[df_nyc_location["Status"] == "Closed"]
df_nyc_location.shape
```

[41]:
```
(298471, 14)
```

[42]:
```
#complaint distribution
df_nyc_location["Complaint Type"].value_counts().plot(kind = 'barh', figsize=␣
↪(10,10))
```

[42]: <AxesSubplot:>



[43]:
```
#viewing total compaints duration
df_nyc_location.groupby(by="Complaint Type")["Actual Resolution Time"].sum()
```

[43]:
```
Complaint Type
Agency Issues                   1 days 07:33:43
Animal Abuse                 1687 days 00:59:09
Animal in a Park               14 days 00:50:05
Bike/Roller/Skate Chronic      66 days 12:58:48
Blocked Driveway            15169 days 05:21:46
Derelict Vehicle             5395 days 05:35:38
Disorderly Youth               42 days 09:45:18
Drinking                      205 days 03:50:13
Graffiti                       33 days 16:05:29
```

```
Homeless Encampment          800 days 05:18:32
Illegal Fireworks             19 days 07:52:17
Illegal Parking            13973 days 04:08:46
Noise - Commercial          4621 days 17:23:24
Noise - House of Worship     123 days 14:34:27
Noise - Park                 571 days 03:34:58
Noise - Street/Sidewalk     6900 days 13:20:22
Noise - Vehicle             2547 days 00:00:34
Panhandling                   55 days 13:41:39
Posting Advertisement         53 days 06:28:08
Squeegee                       0 days 16:10:57
Traffic                      645 days 20:57:11
Urinating in Public           89 days 10:59:07
Vending                      634 days 09:54:17
Name: Actual Resolution Time, dtype: timedelta64[ns]
```

[44]:
```
df_nyc_location.groupby(by="Complaint Type")["Actual Resolution Time"].count()
```

[44]:
```
Complaint Type
Agency Issues                  6
Animal Abuse                7766
Animal in a Park               1
Bike/Roller/Skate Chronic    424
Blocked Driveway           76793
Derelict Vehicle           17585
Disorderly Youth             286
Drinking                    1275
Graffiti                     113
Homeless Encampment         4410
Illegal Fireworks            168
Illegal Parking            74515
Noise - Commercial         35245
Noise - House of Worship     929
Noise - Park                4021
Noise - Street/Sidewalk    48068
Noise - Vehicle            17032
Panhandling                  305
Posting Advertisement        647
Squeegee                       4
Traffic                     4493
Urinating in Public          592
Vending                     3793
Name: Actual Resolution Time, dtype: int64
```

[45]:
```
#Converting all the datetime into minutes for statistical analysis
```

```
df_nyc_location['Actual Resolution Time'] = (df_nyc_location['Actual Resolution␣
↪Time'].dt.seconds.div(60).astype(int) + df_nyc_location['Actual Resolution␣
↪Time'].dt.days.multiply(1440).astype(int))
```

[46]:
```
df_nyc_location["Location Type"].value_counts()
```

[46]:
```
Street/Sidewalk                247503
Store/Commercial                20183
Club/Bar/Restaurant             17227
Residential Building/House       6953
Park/Playground                  4751
House of Worship                  927
Residential Building              227
Highway                           214
Parking Lot                       117
House and Store                    93
Vacant Lot                         77
Commercial                         62
Roadway Tunnel                     35
Subway Station                     34
Bridge                              2
Park                                1
Name: Location Type, dtype: int64
```

[47]:
```
#Grouping NYC data by City
grouped_nyc = df_nyc_location.groupby(['City', 'Complaint Type'])
```

[48]:
```
grouped_nyc['Actual Resolution Time'].mean()
```

[48]:
```
City       Complaint Type
ARVERNE    Animal Abuse                128.894737
           Blocked Driveway            151.200000
           Derelict Vehicle            177.740741
           Disorderly Youth            215.000000
           Drinking                     14.000000
                                           ...
Woodside   Blocked Driveway            384.181818
           Derelict Vehicle            298.000000
           Illegal Parking             312.830000
           Noise - Commercial          143.000000
           Noise - Street/Sidewalk     204.400000
Name: Actual Resolution Time, Length: 764, dtype: float64
```

[49]:
```
grouped_nyc['Actual Resolution Time'].sum()
```

[49]:
```
City       Complaint Type
ARVERNE    Animal Abuse                 4898
```

```
            Blocked Driveway            5292
            Derelict Vehicle            4799
            Disorderly Youth             430
            Drinking                      14
                                         ...
Woodside    Blocked Driveway            4226
            Derelict Vehicle             596
            Illegal Parking            31283
            Noise - Commercial           286
            Noise - Street/Sidewalk     1022
Name: Actual Resolution Time, Length: 764, dtype: int64
```

[50]:
```
grouped_nyc['Actual Resolution Time'].count()
```

[50]:
```
City      Complaint Type
ARVERNE   Animal Abuse               38
          Blocked Driveway           35
          Derelict Vehicle           27
          Disorderly Youth            2
          Drinking                    1
                                     ...
Woodside  Blocked Driveway           11
          Derelict Vehicle            2
          Illegal Parking           100
          Noise - Commercial          2
          Noise - Street/Sidewalk     5
Name: Actual Resolution Time, Length: 764, dtype: int64
```

[51]:
```
city_mean_df = grouped_nyc['Actual Resolution Time'].mean()
```

[52]:
```
city_mean_df
```

[52]:
```
City      Complaint Type
ARVERNE   Animal Abuse              128.894737
          Blocked Driveway          151.200000
          Derelict Vehicle          177.740741
          Disorderly Youth          215.000000
          Drinking                   14.000000
                                       ...
Woodside  Blocked Driveway          384.181818
          Derelict Vehicle          298.000000
          Illegal Parking           312.830000
          Noise - Commercial        143.000000
          Noise - Street/Sidewalk   204.400000
Name: Actual Resolution Time, Length: 764, dtype: float64
```

1 Hypothesis Definition and Testing

2 Question 1: Whether the average response time across complaint types is similar or not (overall)

2.1 Hypothesis #1. Are complaints resolved in the same mean time across categories?

Null Hypothesis: Mean resolution time across different complaints handled by the NYC police department are same

Alternate Hypothesis: Mean resolution time across different complaints handled by the NYC police department are not the same

Preparing the dataset required for analysis

[67]:
```
#taking a subset including only the city name, complaint type, caseid and the␣
 ↪resolution time
df_nyc_analysis = df_nyc_location[['City', 'Complaint Type', 'Location Type',␣
 ↪'Actual Resolution Time']]
df_nyc_analysis.head(5)
```

[67]:

	City	Complaint Type	Location Type	Actual Resolution Time
0	NEW YORK	Noise - Street/Sidewalk	Street/Sidewalk	55
1	ASTORIA	Blocked Driveway	Street/Sidewalk	86
2	BRONX	Blocked Driveway	Street/Sidewalk	291
3	BRONX	Illegal Parking	Street/Sidewalk	465
4	ELMHURST	Illegal Parking	Street/Sidewalk	207

[85]:
```
#taking the required subset for the ANOVA test for complaint type and␣
 ↪resolution time
df_nyc_complaints = df_nyc_analysis[['Complaint Type', 'Actual Resolution␣
 ↪Time']]
df_nyc_complaints = df_nyc_complaints.rename(columns = {
    'Complaint Type': 'Complaint',
    'Actual Resolution Time': 'Time'
}, inplace = False)
```

[86]:
```
df_nyc_complaints.head(5)
```

[86]:

	Complaint	Time
0	Noise - Street/Sidewalk	55
1	Blocked Driveway	86
2	Blocked Driveway	291
3	Illegal Parking	465
4	Illegal Parking	207

[89]:
```
df_nyc_complaints.groupby('Complaint').mean()
```

[89]:

Complaint	Time
Agency Issues	315.333333
Animal Abuse	312.512490
Animal in a Park	20210.000000
Bike/Roller/Skate Chronic	225.693396
Blocked Driveway	284.142643
Derelict Vehicle	441.499119
Disorderly Youth	213.167832
Drinking	231.391373
Graffiti	428.752212
Homeless Encampment	261.000907
Illegal Fireworks	165.470238
Illegal Parking	269.724982
Noise - Commercial	188.524159
Noise - House of Worship	191.287406
Noise - Park	204.239741
Noise - Street/Sidewalk	206.425917
Noise - Vehicle	215.039220
Panhandling	262.072131
Posting Advertisement	118.262751
Squeegee	242.500000
Traffic	206.695526
Urinating in Public	217.302365
Vending	240.542578

On the outset, it seems that the mean time to solve different types of complaints is not the same. We will use the ANOVA test to analyse if this is the case

[78]:
```
import statsmodels.api as sm
from statsmodels.formula.api import ols
```

[87]:
```
lm = ols('Time~Complaint', data = df_nyc_complaints).fit()
table = sm.stats.anova_lm(lm)
print(table)
```

```
                df        sum_sq       mean_sq           F  PR(>F)
Complaint     22.0  1.454406e+09  6.610938e+07  513.960687     0.0
Residual  298448.0  3.838856e+10  1.286273e+05         NaN     NaN
```

The p value that is derived from the ANOVA test is very insignificant even with a confidence interval of 0.05 and so we reject the null hypothesis.

Hence the mean time to resolve across various types of complaints is not equal

[90]:
```
#examining if the animal in the park is the outlier and if removed the mean␣
 ↪time to resolve across other categories is the same
df_nyc_animals = df_nyc_complaints[df_nyc_complaints['Complaint']=='Animal in a␣
 ↪Park']
```

```
df_nyc_animals.value_counts()
```

[90]:
```
Complaint         Time
Animal in a Park  20210    1
dtype: int64
```

[92]:
```
#removing this datapoint alone to examine the dataset again using the ANOVA test
df_nyc_noanimals = df_nyc_complaints[df_nyc_complaints['Complaint']!='Animal in␣
↪a Park']
```

[94]:
```
df_nyc_noanimals.groupby('Complaint').mean()
```

[94]:

Complaint	Time
Agency Issues	315.333333
Animal Abuse	312.512490
Bike/Roller/Skate Chronic	225.693396
Blocked Driveway	284.142643
Derelict Vehicle	441.499119
Disorderly Youth	213.167832
Drinking	231.391373
Graffiti	428.752212
Homeless Encampment	261.000907
Illegal Fireworks	165.470238
Illegal Parking	269.724982
Noise - Commercial	188.524159
Noise - House of Worship	191.287406
Noise - Park	204.239741
Noise - Street/Sidewalk	206.425917
Noise - Vehicle	215.039220
Panhandling	262.072131
Posting Advertisement	118.262751
Squeegee	242.500000
Traffic	206.695526
Urinating in Public	217.302365
Vending	240.542578

[93]:
```
no_animals = ols('Time~Complaint', data = df_nyc_noanimals).fit()
table_noanimals = sm.stats.anova_lm(no_animals)
print(table_noanimals)
```

```
                df        sum_sq       mean_sq           F  PR(>F)
Complaint     21.0  1.056344e+09  5.030210e+07  391.068618     0.0
Residual  298448.0  3.838856e+10  1.286273e+05         NaN     NaN
```

Even after removing the possible outlier of 'Animal in a Park', the p value that is derived from the ANOVA test is very insignificant even with a confidence interval of 0.05 and so we reject the null hypothesis.

Hence the mean time to resolve across various types of complaints is not equal

2.2 ANOVA Conclusion (Hypothesis #1): The mean time to resolve various types of complaints is not similar

2.2.1 Examining the same across the prominent cities of Brooklyn and Bronx

2.2.2 Hypothesis #2

Null Hypothesis: Mean resolution time across different complaints across the city 'BROOKLYN' are same

Alternate Hypothesis: Mean resolution time across different complaints across the city 'BROOKLYN' are not the same

[106]:
```
#getting the necessary dataset
df_brooklyn = df_nyc_analysis[df_nyc_analysis['City']=='BROOKLYN']
df_brooklyn.head(5)
```

[106]:

	City	Complaint Type	Location Type	Actual Resolution Time
5	BROOKLYN	Illegal Parking	Street/Sidewalk	113
9	BROOKLYN	Blocked Driveway	Street/Sidewalk	83
13	BROOKLYN	Illegal Parking	Street/Sidewalk	510
17	BROOKLYN	Noise - Commercial	Club/Bar/Restaurant	51
18	BROOKLYN	Noise - Commercial	Club/Bar/Restaurant	176

[107]:
```
df_brooklyn = df_brooklyn.rename(columns = {
    'Complaint Type': 'Complaint',
    'Actual Resolution Time': 'Time'
}, inplace = False)
```

[108]:
```
df_brooklyn.drop(['City', 'Location Type'], axis = 1)
```

[108]:

	Complaint	Time
5	Illegal Parking	113
9	Blocked Driveway	83
13	Illegal Parking	510
17	Noise - Commercial	51
18	Noise - Commercial	176
…	…	…
300681	Noise - Street/Sidewalk	174
300682	Noise - Commercial	385
300683	Noise - Street/Sidewalk	175
300689	Noise - Street/Sidewalk	218
300695	Noise - Commercial	187

[98275 rows x 2 columns]

[100]:
```
df_brooklyn['Complaint'].value_counts()
```

[100]:
```
Blocked Driveway             28139
Illegal Parking              27454
Noise - Street/Sidewalk      13354
Noise - Commercial           11458
Derelict Vehicle              5179
Noise - Vehicle               5176
Animal Abuse                  2393
Noise - Park                  1554
Traffic                       1085
Homeless Encampment            855
Vending                        514
Noise - House of Worship       340
Drinking                       257
Urinating in Public            136
Bike/Roller/Skate Chronic      111
Disorderly Youth                72
Illegal Fireworks               61
Panhandling                     49
Posting Advertisement           45
Graffiti                        43
Name: Complaint, dtype: int64
```

[101]:
```
df_brooklyn.groupby('Complaint').mean()
```

[101]:

Complaint	Time
Animal Abuse	289.633932
Bike/Roller/Skate Chronic	299.972973
Blocked Driveway	264.310743
Derelict Vehicle	356.535238
Disorderly Youth	248.680556
Drinking	212.124514
Graffiti	494.232558
Homeless Encampment	281.561404
Illegal Fireworks	140.180328
Illegal Parking	255.982298
Noise - Commercial	178.859749
Noise - House of Worship	183.873529
Noise - Park	188.453668
Noise - Street/Sidewalk	197.460761
Noise - Vehicle	196.709428
Panhandling	258.673469
Posting Advertisement	201.422222
Traffic	186.470046
Urinating in Public	233.698529
Vending	271.225681

[102]:
```
#ANOVA test for brooklyn
brooklyn = ols('Time~Complaint', data = df_brooklyn).fit()
table_brooklyn = sm.stats.anova_lm(brooklyn)
print(table_brooklyn)
```

```
                df        sum_sq       mean_sq          F  PR(>F)
Complaint     19.0  1.903002e+08  1.001580e+07  86.69383     0.0
Residual   98255.0  1.135147e+10  1.155307e+05        NaN     NaN
```

Examining the brooklyn dataset, the p value that is derived from the ANOVA test is very insignificant even with a confidence interval of 0.05 and so we reject the null hypothesis.

Hence the mean time to resolve across various types of complaints in brooklyn is not equal

2.3 ANOVA Conclusion (Hypothesis #2): The mean time to resolve various types of complaints in brooklyn is not similar

2.3.1 Hypothesis #3:

Null Hypothesis: Mean resolution time across different complaints across the city 'BRONX' are same

Alternate Hypothesis: Mean resolution time across different complaints across the city 'BRONX' are not the same

[105]:
```
#getting the necessary dataset
df_bronx = df_nyc_analysis[df_nyc_analysis['City']=='BRONX']
df_bronx.head(5)
```

[105]:

	City	Complaint Type	Location Type	Actual Resolution Time
2	BRONX	Blocked Driveway	Street/Sidewalk	291
3	BRONX	Illegal Parking	Street/Sidewalk	465
7	BRONX	Blocked Driveway	Street/Sidewalk	107
11	BRONX	Blocked Driveway	Street/Sidewalk	667
12	BRONX	Noise - Street/Sidewalk	Street/Sidewalk	148

[109]:
```
df_bronx = df_bronx.rename(columns = {
    'Complaint Type': 'Complaint',
    'Actual Resolution Time': 'Time'
}, inplace = False)
```

[110]:
```
df_bronx.drop(['City', 'Location Type'], axis = 1)
```

[110]:

	Complaint	Time
2	Blocked Driveway	291
3	Illegal Parking	465
7	Blocked Driveway	107
11	Blocked Driveway	667
12	Noise - Street/Sidewalk	148
…	…	…

```
300643          Illegal Parking   176
300652        Blocked Driveway    105
300656        Blocked Driveway    450
300690          Illegal Parking   486
300696      Noise - Commercial    245

[40690 rows x 2 columns]
```

[111]:
```
df_bronx['Complaint'].value_counts()
```

[111]:
```
Blocked Driveway             12751
Noise - Street/Sidewalk       8890
Illegal Parking               7857
Noise - Vehicle               3395
Noise - Commercial            2433
Derelict Vehicle              1952
Animal Abuse                  1415
Noise - Park                   547
Vending                        379
Traffic                        355
Homeless Encampment            247
Drinking                       188
Noise - House of Worship        79
Disorderly Youth                63
Urinating in Public             51
Illegal Fireworks               24
Bike/Roller/Skate Chronic       20
Panhandling                     19
Posting Advertisement           16
Graffiti                         9
Name: Complaint, dtype: int64
```

[112]:
```
df_bronx.groupby('Complaint').mean()
```

[112]:
```
                                Time
Complaint
Animal Abuse                439.826855
Bike/Roller/Skate Chronic   207.300000
Blocked Driveway            375.418948
Derelict Vehicle            553.342725
Disorderly Youth            254.015873
Drinking                    347.297872
Graffiti                    533.666667
Homeless Encampment         446.222672
Illegal Fireworks           336.333333
Illegal Parking             394.493445
Noise - Commercial          281.542951
```

```
Noise - House of Worship   273.265823
Noise - Park               281.606947
Noise - Street/Sidewalk    313.278965
Noise - Vehicle            333.379676
Panhandling                852.684211
Posting Advertisement      213.250000
Traffic                    295.121127
Urinating in Public        323.196078
Vending                    409.245383
```

[113]:
```
#ANOVA test for bronx
bronx = ols('Time~Complaint', data = df_bronx).fit()
table_bronx = sm.stats.anova_lm(bronx)
print(table_bronx)
```

```
                df         sum_sq       mean_sq           F         PR(>F)
Complaint     19.0  1.446558e+08  7.613462e+06   36.908754  3.333669e-135
Residual   40670.0  8.389324e+09  2.062779e+05         NaN            NaN
```

Examining the bronx dataset, the p value that is derived from the ANOVA test is very insignificant even with a confidence interval of 0.05 and so we reject the null hypothesis.

Hence the mean time to resolve across various types of complaints in bronx is not equal

2.4 ANOVA Conclusion (Hypothesis #3): The mean time to resolve various types of complaints in bronx is not similar

3 _____

4 Answer 1: The mean time to resolve various types of complaints is not similar; it is also not similar when we examine it across the top cities as well

4.1 _____

5 Question 2: Are the type of complaint or service requested and location related?

[126]:
```
#taking subset of the nyc_complaints dataset required for analysis
df_nyc_services = df_nyc_location[['City','Complaint Type','Location Type']]
```

[127]:
```
df_nyc_services.head(5)
```

[127]:

	City	Complaint Type	Location Type
0	NEW YORK	Noise - Street/Sidewalk	Street/Sidewalk

```
1   ASTORIA          Blocked Driveway  Street/Sidewalk
2     BRONX          Blocked Driveway  Street/Sidewalk
3     BRONX           Illegal Parking  Street/Sidewalk
4  ELMHURST           Illegal Parking  Street/Sidewalk
```

[128]:
```
#renaming columns for ease of use
df_nyc_services = df_nyc_services.rename(columns = {
    'Complaint Type': 'Complaint',
    'Location Type': 'Location'
}, inplace = False)
```

[129]:
```
df_nyc_services.head(5)
```

[129]:
```
        City                Complaint         Location
0  NEW YORK  Noise - Street/Sidewalk  Street/Sidewalk
1   ASTORIA          Blocked Driveway  Street/Sidewalk
2     BRONX          Blocked Driveway  Street/Sidewalk
3     BRONX           Illegal Parking  Street/Sidewalk
4  ELMHURST           Illegal Parking  Street/Sidewalk
```

[130]:
```
#dropping the city column & using a label encoder for the complaints and␣
 ↪location
df_nyc_nocity = df_nyc_services.drop('City', axis=1)
```

[131]:
```
from sklearn.preprocessing import LabelEncoder
number = LabelEncoder()
df_nyc_nocity['Complaint'] = number.fit_transform(df_nyc_nocity['Complaint'].
 ↪astype('str'))
df_nyc_nocity['Location'] = number.fit_transform(df_nyc_nocity['Location'].
 ↪astype('str'))
```

[132]:
```
df_nyc_nocity.head(5)
```

[132]:
```
   Complaint  Location
0         15        13
1          4        13
2          4        13
3         11        13
4         11        13
```

5.0.1 Hypothesis #4

Null Hypothesis: There is no relationship between the complaint type and the location

Alternate Hypothesis: There exists some relationship between the complaint type and the location

[133]:
```
#spearmans rank correlation to check if the
from scipy.stats import spearmanr
```

[134]:
```
data1 = df_nyc_nocity['Complaint']
data2 = df_nyc_nocity['Location']
```

[137]:
```
#converting the pandas series to numpy array
data1.to_numpy()
data2.to_numpy()
```

[137]:
```
array([13, 13, 13, …,  1,  1, 12])
```

[138]:
```
stat, p = spearmanr(data1, data2)
```

[139]:
```
print('stat=%.3f, p=%.3f' % (stat, p))
```

```
stat=-0.143, p=0.000
```

[140]:
```
if p > 0.05:
    print('The two variables are probably independent')
else:
    print('The two variables are probably dependent')
```

```
The two variables are probably dependent
```

According to spearman's rank correlation, the two variables are probably interdependent

5.1 Spearman's rank correlation result:

There seems to exist some form of correlation which might be monotonic in nature between the complaints raised and the location

5.1.1 Performing a Chi-squared test to confirm

[141]:
```
from scipy.stats import chi2_contingency
```

[142]:
```
table = [data1,data2]
```

[143]:
```
stat_chi, p_chi, dof, expected = chi2_contingency(table)
print('stat=%.3f, p=%.3f' % (stat, p))
if p_chi > 0.05:
    print('Probably independent')
else:
    print('Probably dependent')
```

```
stat=-0.143, p=0.000
Probably dependent
```

[ ]:
```

```

There seems to exist some form of correlation between the complaints raised and the location

6 Answer 2: There is a some correlation between the complaints raised and the location. This is explained by the low value of p in the spearman's correlation