

This is the project to perform data analysis of service request (311) calls from New York City. You've also been asked to utilize data wrangling techniques to understand the pattern in the data and visualize the major types of complaints.

In [13]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
color = sns.color_palette()
```

In [14]:

```
pd.set_option('display.max_columns', None)
data = pd.read_csv('311_Service_Requests_from_2010_to_Present.csv')
```

C:\Users\aksha\AppData\Local\Temp\ipykernel_20512\3779929358.py:2: DtypeWarning: Columns (48,49) have mixed types. Specify dtype option on import or set low_memory=False.
 data = pd.read_csv('311_Service_Requests_from_2010_to_Present.csv')

In [15]:

```
data.head()
```

Out[15]:

	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
0	32310363	12/31/2015 11:59:44 PM	01/01/2016 12:55:15 AM	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	Street/Sidewalk	10034.0	71 VERMILYEA AVENUE	VERMILYEA AVENUE	ACADEMY STREET	WEST 204 STREET	NaN	NaN	ADDRESS	NEW YORK	NaN	Precinct
1	32309934	12/31/2015 11:59:44 PM	01/01/2016 01:26:57 AM	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	11105.0	27-07 23 AVENUE	23 AVENUE	27 STREET	28 STREET	NaN	NaN	ADDRESS	ASTORIA	NaN	Precinct
2	32309159	12/31/2015 11:56:58 PM	01/01/2016 04:51:03 AM	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	10458.0	2897 VALENTINE AVENUE	VALENTINE AVENUE	EAST 198 STREET	EAST 199 STREET	NaN	NaN	ADDRESS	BRONX	NaN	Precinct
3	32305098	12/31/2015 11:57:46 PM	01/01/2016 07:43:13 AM	NYPD	New York City Police Department	Illegal Parking	Commercial Overnight Parking	Street/Sidewalk	10461.0	2940 BAISLEY AVENUE	BAISLEY AVENUE	EDISON AVENUE	B STREET	NaN	NaN	ADDRESS	BRONX	NaN	Precinct
4	32306529	12/31/2015 11:59:44 PM	01/01/2016 03:24:42 AM	NYPD	New York City Police Department	Illegal Parking	Blocked Sidewalk	Street/Sidewalk	11373.0	87-14 57 ROAD	57 ROAD	SEABURY STREET	HOFFMAN DRIVE	NaN	NaN	ADDRESS	ELMHURST	NaN	Precinct

In [16]:

```
#missing Values
from datetime import date, time, datetime
data['Created Date'] = data['Created Date'].astype('datetime64[ns]')
data['Closed Date'] = data['Closed Date'].astype('datetime64[ns]')
data['time_diff'] = data['Closed Date'] - data['Created Date']
data['Request_Closing_Time'] = (data['time_diff'].astype('timedelta64[s']))./(60*60)
#data.drop("time_diff_hr",axis=1,inplace=True)
data.head(3)
#data type of datetime column is a datetime64[ns] object
#the [ns] means the nano second-based time format that specifies the precision of the Datetime object
```

Out[16]:

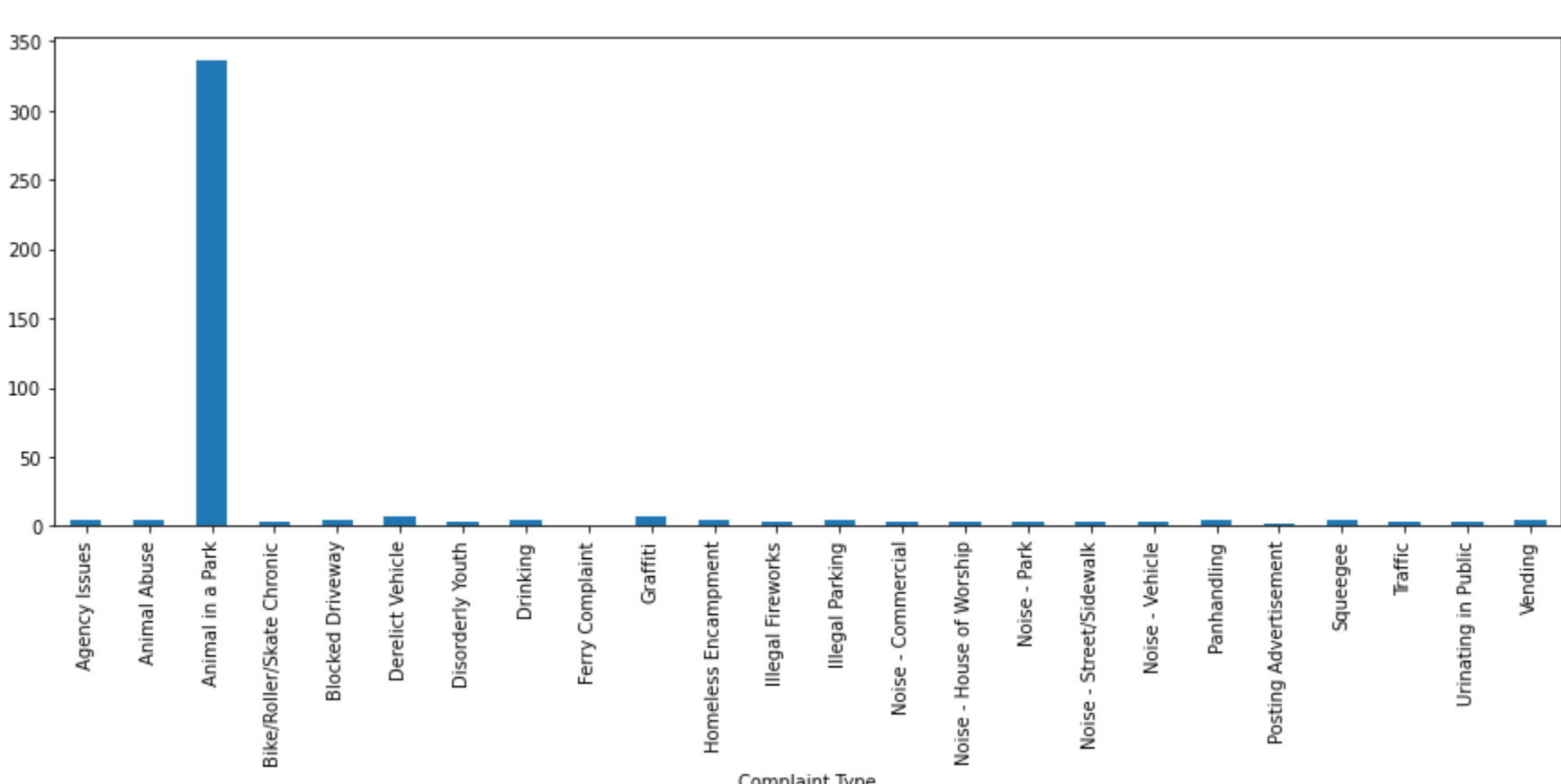
	Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	Street Name	Cross Street 1	Cross Street 2	Intersection Street 1	Intersection Street 2	Address Type	City	Landmark	Facility Type	Status
0	32310363	2015-12-31 23:59:44	2016-01-01 00:55:15	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	Street/Sidewalk	10034.0	71 VERMILYEA AVENUE	VERMILYEA AVENUE	ACADEMY STREET	WEST 204 STREET	NaN	NaN	ADDRESS	NEW YORK	NaN	Precinct	Closed
1	32309934	2015-12-31 23:59:44	2016-01-01 01:26:57	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	11105.0	27-07 23 AVENUE	23 AVENUE	27 STREET	28 STREET	NaN	NaN	ADDRESS	ASTORIA	NaN	Precinct	Closed
2	32309159	2015-12-31 23:59:29	2016-01-01 04:51:03	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	10458.0	2897 VALENTINE AVENUE	VALENTINE AVENUE	EAST 198 STREET	EAST 199 STREET	NaN	NaN	ADDRESS	BRONX	NaN	Precinct	Closed

In [17]:

```
import matplotlib.pyplot as plt
import seaborn as sns
plt.rcParams["figure.figsize"]=(15,5)
data.groupby("Complaint Type")["Request_Closing_Time"].mean().plot.bar()
# Animal in a park takes maximum average time close the tickets
```

Out[17]:

<AxesSubplot: xlabel='Complaint Type'>



In [18]:

```
import matplotlib.pyplot as plt
import seaborn as sns
plt.rcParams["figure.figsize"]=(15,5)
data.loc[data["Complaint Type"]!="Animal in a Park",].groupby("Complaint Type")["Request_Closing_Time"].mean().plot.bar()
#Animal in a Park seems like an outlier, hence we remove it then we notice that Derelict Vehicle and Graffiti takes the maximum time to close the tickets
```

Out[18]:

<AxesSubplot: xlabel='Complaint Type'>

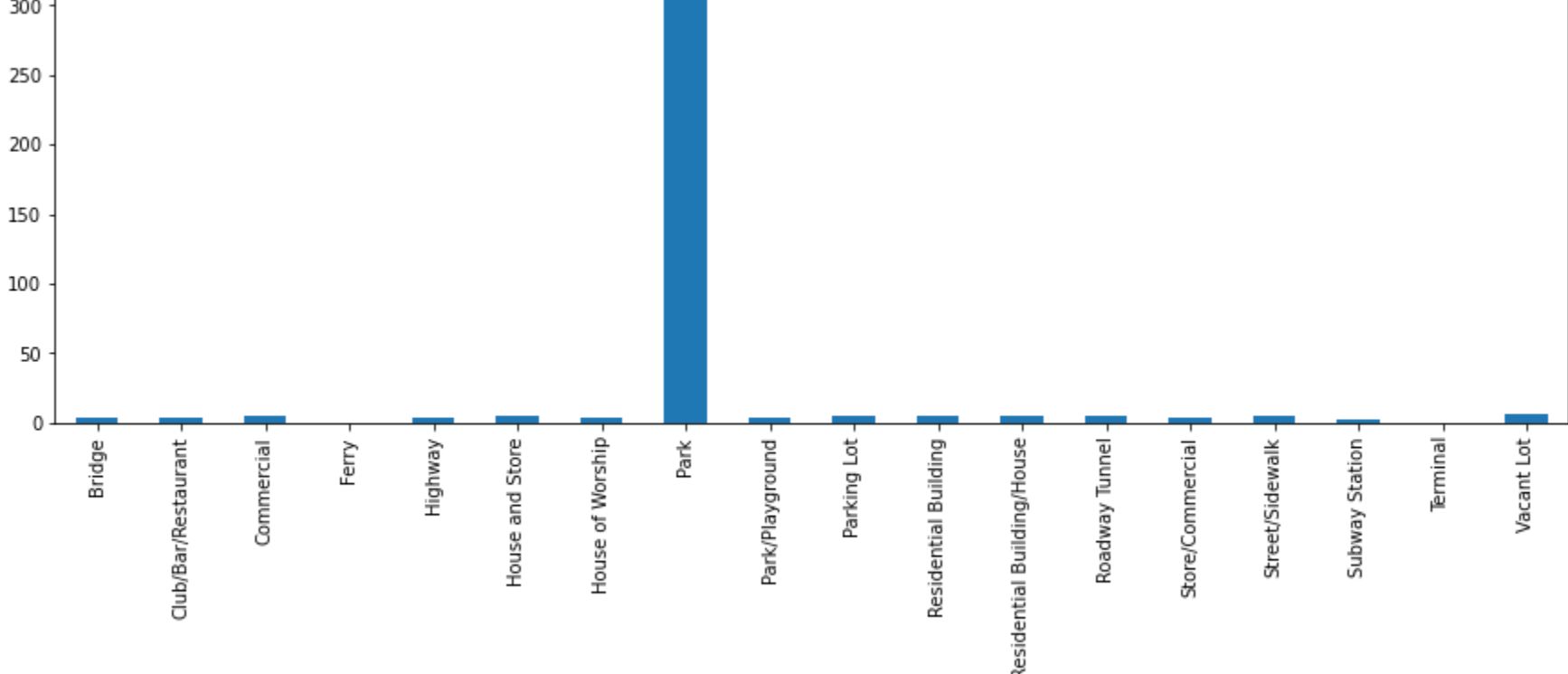


In [19]:

```
import matplotlib.pyplot as plt
import seaborn as sns
plt.rcParams["figure.figsize"]=(15,5)
data.groupby("Location Type")["Request_Closing_Time"].mean().plot.bar()
#Park takes the maximum average time to close the tickets
```

Out[19]:

<AxesSubplot: xlabel='Location Type'>



In [20]:

```
import matplotlib.pyplot as plt
import seaborn as sns
plt.rcParams["figure.figsize"]=(15,5)
data.loc[data["Location Type"]=="Park",].groupby("Location Type")["Request_Closing_Time"].mean().plot.bar()
#Park seems to be outlier, Hence if we remove it then we notice that vaccant lot takes the maximum time to close the tickets
```

Out[20]:

<AxesSubplot: xlabel='Location Type'>



In [21]:

```
data.head(3)
```

Out[21]:

	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
0	32310363	2015-12-31 23:59:45	2016-01-01 00:55:15	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	Street/Sidewalk	10034.0	71 VERMILYEA AVENUE	VERMILYEA AVENUE	ACADEMY STREET	WEST 204 STREET	NaN	NaN	ADDRESS	NEW YORK	NaN	Precinct	Closed
1	32309934	2015-12-31 23:59:44	2016-01-01 01:26:57	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	11105.0	27-07 23 AVENUE	23 AVENUE	27 STREET	28 STREET	NaN	NaN	ADDRESS	ASTORIA	NaN	Precinct	Closed
2	32309159	2015-12-31 23:59:29	2016-01-01 04:51:03	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	10458.0	2897 VALENTINE AVENUE	VALENTINE AVENUE	EAST 198 STREET	EAST 199 STREET	NaN	NaN	ADDRESS	BRONX	NaN	Precinct	Closed

In [22]:

```
#order the complaint types based on the average 'Request_Closing_Time',grouping time for different locations.
a = pd.DataFrame(data.groupby(["Location","Complaint Type"])["Request_Closing_Time"].mean())
a.sort_values("Request_Closing_Time")
```

Out[22]:

Request_Closing_Time		
Location	Complaint Type	
(40.72895633655987, -74.00074325193769)	Noise - Commercial	0.043611
(40.678429539269835, -73.98361397732425)	Noise - Commercial	0.043611
(40.76848580086362, -73.912352505723272)	Noise - Vehicle	0.045278
(40.69371028050496, -73.95499211670034)	Illegal Parking	0.046389
(40.76459666240956, -73.97372678473167)	Noise - Street/Sidewalk	0.046944
...
(40.73681034405816, -73.93570112335456)	Illegal Parking	NaN
(40.74285387870548, -73.97913408166315)	Blocked Driveway	NaN
(40.749256156669986, -74.00919964319513)	Noise - Street/Sidewalk	NaN
(40.7735541162767, -73.95172479233456)	Traffic	NaN
(40.80405676738835, -73.95343293482638)	Noise - Street/Sidewalk	NaN

178175 rows × 1 columns

(40.80405676738835, -73.95343293482638) Noise - Street/Sidewalk NaN

178175 rows × 1 columns

In [23]:

```
#whether the average response time across the Complaint type is similar or nor (overall)

import statsmodels.formula.api as sm
mod = sm.ols(formula = """Request_Closing_Time ~ Q('Complaint Type')""", data = data).fit()
mod.summary()
#Anova_hypothesis
```

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 1.66e-28. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

In [24]:

```
# Are the type of complaint or service requested and location related
CrosstabResult = pd.crosstab(index=data['Complaint Type'],columns = data['Location'])
from scipy.stats import chi2_contingency
ChiSqResult = chi2_contingency(CrosstabResult)
print('The P-Value of the ChiSq Test is',ChiSqResult[1])

#H0 : there is no relation
#H1 : there is a relation

#The P-Value is less than 0.05, hence we reject theNull Hypothesis and conclude that there is a relation between
#complaint and location

The P-Value of the ChiSq Test is 0.0
```

In []: