

```
[1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline

In [2]: pwd

Out[2]: 'C:\\Users\\HP'

In [4]: df = pd.read_csv(r"C:\Users\HP\Desktop\project\311_Service_Requests_from_2010_to_Present.csv")

In [5]: df.head()

Out[5]:
```

	Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Bridge Highway Name	Bridge Highway Direction	Road Ramp	Bridge Highway Segment	Garage Lot Name	Ferry Direction	Ferry Terminal Name	Latitude	Longitude
0	32310363	12/31/2015 11:59:45 PM	01/01/2016 12:55:15 AM	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	Street/Sidewalk	10034.0	71 VERMILYEA AVENUE	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	40.865682	-73.923501
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	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	40.775945	-73.915094
2	32309159	12/31/2015 11:59:29 PM	01/01/2016 04:51:03 AM	NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	10458.0	VALENTINE AVENUE	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	40.870325	-73.888525
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	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	40.835994	-73.828379
4	32306529	12/31/2015 11:56:58 PM	01/01/2016 03:24:42 AM	NYPD	New York City Police Department	Illegal Parking	Blocked Sidewalk	Street/Sidewalk	11373.0	87-14 57 ROAD	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	40.733060	-73.874170

5 rows × 53 columns

```
In [6]: df.shape

Out[6]: (364558, 53)

In [7]: df.isnull().any()

Out[7]: Unique Key          False
Created Date          False
Closed Date           False
Agency               False
Agency Name          False
Complaint Type        False
Descriptor             True
Location Type          True
Incident Zip          True
Incident Address      True
Street Name           True
Cross Street 1         True
Cross Street 2         True
Intersection Street 1  True
Intersection Street 2  True
Address Type           True
City                   True
Landmark               True
Facility Type          True
Status                 False
Due Date               True
Resolution Description  False
Resolution Action Updated Date True
Community Board        False
Borough                False
X Coordinate (State Plane) True
Y Coordinate (State Plane) True
Park Facility Name     False
Park Borough           False
School Name            False
School Number          False
School Region          True
School Code            True
School Phone Number   False
School Address         False
School City            False
School State           False
School Zip             True
School Not Found       False
School or Citywide Complaint True
Vehicle Type           True
Taxi Company Borough  True
Taxi Pick Up Location  True
Bridge Highway Name    True
Bridge Highway Direction True
Road Ramp              True
Bridge Highway Segment True
Garage Lot Name        True
Ferry Direction        True
Ferry Terminal Name    True
Latitude               True
Longitude              True
Location               True
dtype: bool

In [8]: df.isna().any()

Out[8]: Unique Key          False
Created Date          False
Closed Date           True
Agency               False
Agency Name          False
Complaint Type        False
Descriptor             True
Location Type          True
Incident Zip          True
Incident Address      True
Street Name           True
Cross Street 1         True
Cross Street 2         True
Intersection Street 1  True
Intersection Street 2  True
Address Type           True
City                   True
Landmark               True
Facility Type          True
Status                 False
Due Date               True
Resolution Description  False
Resolution Action Updated Date True
Community Board        False
Borough                False
X Coordinate (State Plane) True
Y Coordinate (State Plane) True
Park Facility Name     False
Park Borough           False
School Name            False
School Number          False
School Region          True
School Code            True
School Phone Number   False
School Address         False
School City            False
School State           False
School Zip             True
School Not Found       False
School or Citywide Complaint True
Vehicle Type           True
Taxi Company Borough  True
Taxi Pick Up Location  True
Bridge Highway Name    True
Bridge Highway Direction True
Road Ramp              True
Bridge Highway Segment True
Garage Lot Name        True
Ferry Direction        True
Ferry Terminal Name    True
Latitude               True
Longitude              True
Location               True
dtype: bool

In [9]: column_nan = [col for col in df.columns if df[col].isna().any() ]

In [10]: print(column_nan)

['Closed Date', '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', 'Due Date', 'Resolution Action Updated Date', 'X Coordinate (State Plane)', 'Y Coordinate (State Plane)', 'School Region', 'School Code', 'School Zip', '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']

In [11]: df.isna().sum()

Out[11]: Unique Key          0
Created Date          0
Closed Date          2381
Agency              0
Agency Name         0
Complaint Type       0
Descriptor           6581
Location Type        133
Incident Zip         2998
Incident Address     51699
Street Name          51699
Cross Street 1       57188
Cross Street 2       57805
Intersection Street 1 313438
Intersection Street 2 314946
Address Type         3252
City                 2997
Landmark             364183
Facility Type        2389
Status               0
Due Date             3
Resolution Description 0
Resolution Action Updated Date 2482
Community Board      0
Borough              0
X Coordinate (State Plane) 4030
Y Coordinate (State Plane) 4030
Park Facility Name   0
Park Borough         0
School Name          0
School Number        0
School Region        1
School Code          1
School Phone Number  0
School Address       0
School City          0
School State         0
School Zip           1
School Not Found     0
School or Citywide Complaint 364558
Vehicle Type         364558
Taxi Company Borough 364558
Taxi Pick Up Location 364558
Bridge Highway Name  364261
Bridge Highway Direction 364261
Road Ramp            364296
Bridge Highway Segment 364296
Garage Lot Name      364558
Ferry Direction      364557
Ferry Terminal Name  364556
Latitude             4030
Longitude            4030
Location             4030
dtype: int64

In [12]: df.isna().sum().sum()

Out[12]: 5262284

In [13]: df.isna().mean()

Out[13]: Unique Key          0.000000
Created Date          0.000000
Closed Date          0.006531
Agency              0.000000
Agency Name         0.000000
Complaint Type       0.000000
Descriptor           0.017833
Location Type        0.000365
Incident Zip         0.008224
Incident Address     0.141813
Street Name          0.141813
Cross Street 1       0.156869
Cross Street 2       0.158562
Intersection Street 1 0.859775
Intersection Street 2 0.861443
Address Type         0.008920
City                 0.008221
Landmark             0.998971
Facility Type        0.006553
Status               0.000000
Due Date             0.000008
Resolution Description 0.000000
Resolution Action Updated Date 0.006589
Community Board      0.000000
Borough              0.000000
X Coordinate (State Plane) 0.011054
Y Coordinate (State Plane) 0.011054
Park Facility Name   0.000000
Park Borough         0.000000
School Name          0.000000
School Number        0.000000
School Region        0.000003
School Code          0.000003
School Phone Number  0.000000
School Address       0.000000
School City          0.000000
School State         0.000000
School Zip           0.000003
School Not Found     0.000000
School or Citywide Complaint 1.000000
Vehicle Type         1.000000
Taxi Company Borough 1.000000
Taxi Pick Up Location 1.000000
Bridge Highway Name  0.999185
Bridge Highway Direction 0.999185
Road Ramp            0.999281
Bridge Highway Segment 0.999281
Garage Lot Name      1.000000
Ferry Direction      0.999997
Ferry Terminal Name  0.999995
Latitude             0.011054
Longitude            0.011054
Location             0.011054
dtype: float64

In [24]: df1=df.fillna(value=5)

In [25]: df1.head()

Out[25]:
```

	Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Bridge Highway Name	Bridge Highway Direction	Road Ramp	Bridge Highway Segment	Garage Lot Name	Ferry Direction	Ferry Terminal Name	Latitude	Longitude
0	32310363	12/31/2015 11:59:45 PM	12/55:15 AM	NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	Street/Sidewalk	10034.0	71 VERMILYEA AVENUE	...	5	5	5	5	5.0	5	5	40.865682	-73.923501
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	...	5	5	5	5	5.0	5	5	40.775945	-73.915094</