

## PROJECT REPORT

### DESCRIPTION

#### Background of Problem Statement :

NYC 311's mission is to provide the public with quick and easy access to all New York City government services and information while offering the best customer service. Each day, NYC311 receives thousands of requests related to several hundred types of non-emergency services, including noise complaints, plumbing issues, and illegally parked cars. These requests are received by NYC311 and forwarded to the relevant agencies such as the police, buildings, or transportation. The agency responds to the request, addresses it, and then closes it.

#### Problem Objective :

Perform a service request data analysis of New York City 311 calls. You will focus on the data wrangling techniques to understand the pattern in the data and also visualize the major complaint types.

Domain: Customer Service

#### Analysis Tasks to be performed:

(Perform a service request data analysis of New York City 311 calls)

1. Import a 311 NYC service request.
2. Read or convert the columns 'Created Date' and Closed Date' to datetime datatype and create a new column 'Request\_Closing\_Time' as the time elapsed between request creation and request closing. (Hint: Explore the package/module datetime)
3. Provide major insights/patterns that you can offer in a visual format (graphs or tables); at least 4 major conclusions that you can come up with after generic data mining.
4. Order the complaint types based on the average 'Request\_Closing\_Time', grouping them for different locations.
5. Perform a statistical test for the following:

Please note: For the below statements you need to state the Null and Alternate and then provide a statistical test to accept or reject the Null Hypothesis along with the corresponding 'p-value'.

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

## Results: -

### Test Case-1: -

Here, we will be importing the data of the 311 NYC service request from csv file and validate the rows and columns that are generated from the csv file

The total rows are – 300698.

Columns are – 53.

Here, I have also took the data for first 5 rows.

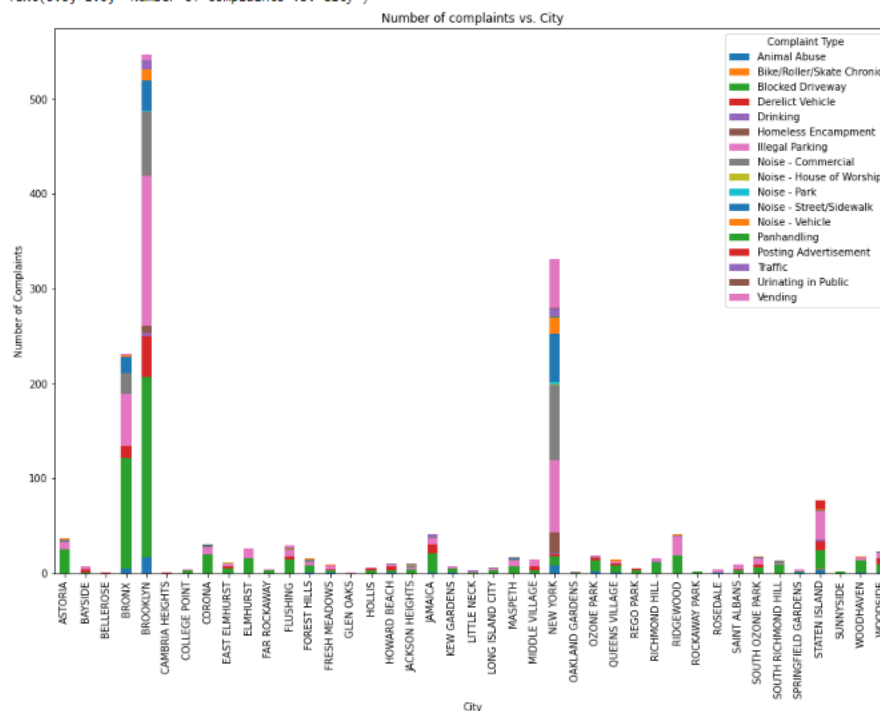
Here, we are also gathering data of the columns where all the categories are present.

### Test Case-2: -

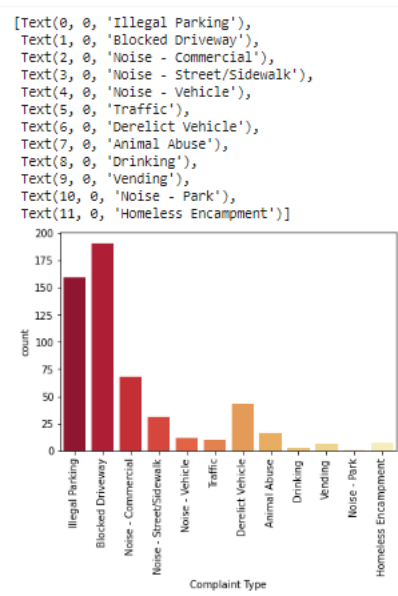
Here, I'm gathering information about the below.

Conclusion 1: City with maximum number of complaints was found to be Brooklyn with different types of complaints.

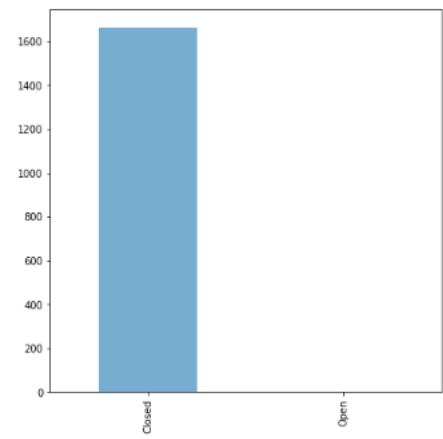
Text(0.5, 1.0, 'Number of complaints vs. City')



Conclusion 2: Since maximum number of complaints were filed in Brooklyn, Blocked Driveway was found to be the most frequent complaint lodged in the city of Brooklyn.

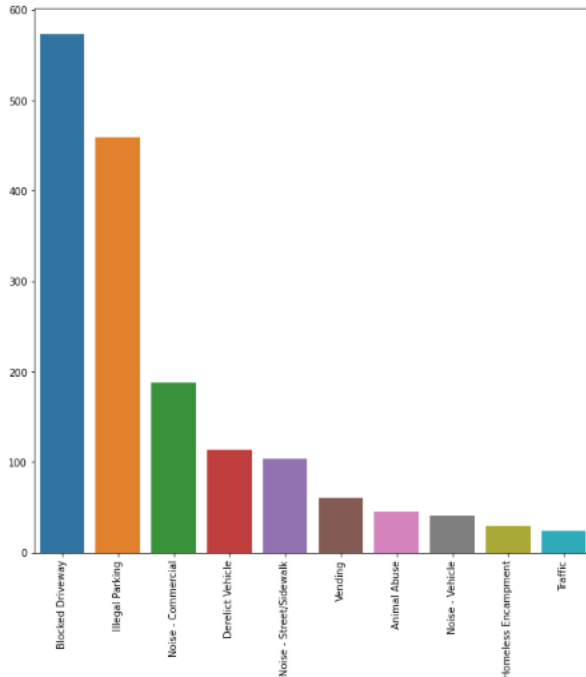


Conclusion 3 - Most of the complained are in closed status



Conclusion 4: Top 10 Complaint types and their count

```
[Text(0, 0, 'Blocked Driveway'),
Text(1, 0, 'Illegal Parking'),
Text(2, 0, 'Noise - Commercial'),
Text(3, 0, 'Derelict Vehicle'),
Text(4, 0, 'Noise - Street/Sidewalk'),
Text(5, 0, 'Vending'),
Text(6, 0, 'Animal Abuse'),
Text(7, 0, 'Noise - Vehicle'),
Text(8, 0, 'Homeless Encampment'),
Text(9, 0, 'Traffic')]
```



#### Test Case-4: -

Here I'm order the complaint types based on the average 'Request\_Closing\_Time', grouping them for different locations.

You can see 5 cities i.e., ASTORIA, BAYSIDE, BELLEROSE, BRONX, and BROOKLYN.

#### Test Case-5: -

Here, I'm checking if the average response time is different or not for the top 5 complaints from the "complaint\_type" data

And also I'm checking the "complaint\_type" and "location\_type" are related or independent

I have gathered the information of this which has data with **1667 rows × 2 columns** through **Chi Square**

