311 Data Project

Romauli Butarbutar, Willey Chen, Inna Logunova, Danni Tian, Steven Tong, Daisy Wen, Zhengjia Xu

Problem statement

NYC 3-1-1 is a service channel that allows New Yorkers to access government services and to file complaints via multiple platforms, such as phone calling, texting and other social media (About NYC311).

The number of 3-1-1 calls made in New York City has increased every year since 2012 with millions of unique calls being made each year [1]. It is important to note that the wait time of the service has been increasing over the last few years going from 15 seconds in 2016 to 33 seconds in 2020. Our client, the New York City Municipal Government, wants to introduce automations calls to improve the efficiency of service requests.

Our client provided a huge amount of calling data. For each call, it includes a wealth of information including the reason for the call, caller location, time of day, and so on. However, it is not readily usable in an unstructured format. Since not all service calls can apply automation, we need to create a well structured database to help our client identify the suitable service requests to apply automation. With the new database, they can make more efficient decisions to allocate their resources by service channel type, location, and so on. New Yorkers could also benefit from a faster service response and better service experience.

Proposal

Our team aims to help our client, the NYC government, to relieve the stress of the call center, allocate public resources efficiently, and maintain quality of service by incorporating tools such as call automation. This could reduce the cost of hiring and training seasonal call center agents, and help the government to take more service calls with less customer wait time.

In order to help the authority to apply automation calls in the service request, we will focus on three dimensions. First, we want to understand if there are specific categories that have a high rate of problem solving in a short duration. Next would be using that information to decide on categories which are suitable to incorporate automation. For example, calls that could be resolved with a phone redirect to another department.Third,

we will look at different platforms and apply automation calls or forwarding for the most demanding ones.

Our dataset is provided by the city of New York from the NYC OpenData in Excel format (OpenData). The record contained information about each 3-1-1 service request in New York City. The original dataset contains information from the year of 2010 until present with 27.1 million records and 41 columns. We will focus on the pandemic years and only extract records starting from the year of 2020.

Team structure and Timeline

For our team contract, we broke off into groups, and separated it based on the tasks that were required to complete at each checkpoint. There were leaders for each task, and would alert the other group members in our team Slack and WhatsApp to review before posting.

The timeline along with responsibilities and challenges goes as follows:

- Checkpoint 1: Writing up the discussion post, and finalizing the project scenario.
    - Group members in charge: Aansh, Willey (complete by 11/18)
- Checkpoint 2: Developing a team contract, and Creating 15 tables in 3NF:
    - Group members in charge: Inna, Willey, Zhengjia
    - Completion date: (when it will be completed: 11/29)
    - Challenges:
        - Outlining the 15 tables
        - Categorizing 41 variables into 15 tables

- Checkpoint 3: Submitting the data schema for review: The database schema taking into account all integrity constraints, triggers, etc., including both the ER diagram and the SQL code.
    - Group members in charge: Aansh, Daisy, Zhengjia
    - Completion date: (when it will be completed: 11/29) (Links to an external site.).
    - Challenges:
        - Setting PK and FK, adding new PK

- Checkpoint 4: Submit the Data Plan for Review: Plan for transforming and entering the data to the database system. Include the Python/R scripts.

    - Group members in charge: Danni, Romauli,Steven
    - Completion date: (when it will be completed: 12/1)
    - Challenges:
        - Duplication of efforts
        - Making sure the code worked

- Checkpoint 5: Submit the Customer Interaction Plan for Review.
    - Group members in charge: Daisy, Danni,Romauli,Steven,Willey
    - Completion date: (when it will be completed: 12/4)
    - Challenges:
        - Filtering useful information from the large dataset (27.1M data)

- Paper :
    - Group members in charge: Everyone with a focus on what discussion they helped with
    - Completion date: (when it will be completed: 12/6).
- Presentation:
    - Group members in charge: Aansh, Daisy, Inna
    - Completion date: (when it will be completed: 12/6)

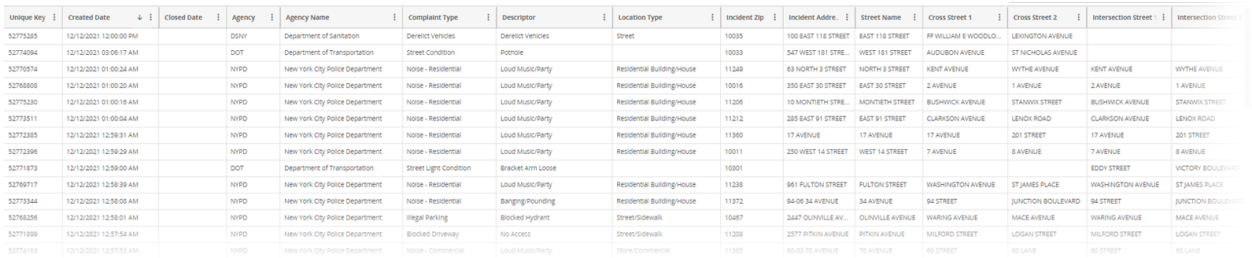
Database Schema

A description of the designed system, complete with logic, ER diagrams, and code listings. Present a sample of your data and provide links to the full dataset. Explain why you chose this dataset. Present your normalization plan and describe all steps in detail. **ER Diagrams must be attached as PDF, providing links to Lucidchart too. Provide SQL code for the normalized tables.**

The designed system of our group is when you enter the complaint ID, you can easily check the agency information, incident information, and the service information.

Here is the link of ER Diagrams:
https://drive.google.com/drive/folders/1mVYtYgjT4vpmP5apOGgWlHg52obycGfT
Here is the link of code listings:
https://drive.google.com/drive/folders/1mVYtYgjT4vpmP5apOGgWlHg52obycGfT
Our project full dataset:
https://data.cityofnewyork.us/api/views/erm2-nwe9/files/68b25fbb-9d30-486a-a571-7115f54911cd?download=true&filename=311_SR_Data_Dictionary_2018.xlsx

Here is one sample of our data:



Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Addre...	Street Name	Cross Street 1	Cross Street 2	Intersection Street ...	Intersection Street ...
52775265	12/12/2021 12:00:00 PM		DSNY	Department of Sanitation	Derelict Vehicles	Derelict Vehicles	Street	10035	100 EAST 118 STREET	EAST 118 STREET	FF WILLIAM E WOODLO...	LEXINGTON AVENUE		
52774094	12/12/2021 03:06:17 AM		DOT	Department of Transportation	Street Condition	Pothole		10033	547 WEST 181 STRE...	WEST 181 STREET	AUDUBON AVENUE	ST NICHOLAS AVENUE		
52770574	12/12/2021 01:00:24 AM		NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	Residential Building/House	11249	63 NORTH 3 STREET	NORTH 3 STREET	KENT AVENUE	WYTHE AVENUE	KENT AVENUE	WYTHE AVENUE
52768808	12/12/2021 01:00:20 AM		NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	Residential Building/House	10016	350 EAST 30 STREET	EAST 30 STREET	2 AVENUE	1 AVENUE	2 AVENUE	1 AVENUE
52775230	12/12/2021 01:00:16 AM		NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	Residential Building/House	11206	10 MONTIETH STRE...	MONTIETH STREET	BUSHWICK AVENUE	STANWIX STREET	BUSHWICK AVENUE	STANWIX STREET
52773511	12/12/2021 01:00:04 AM		NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	Residential Building/House	11212	285 EAST 91 STREET	EAST 91 STREET	CLARKSON AVENUE	LENOX ROAD	CLARKSON AVENUE	LENOX ROAD
52772365	12/12/2021 12:59:31 AM		NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	Residential Building/House	11360	17 AVENUE	17 AVENUE	17 AVENUE	201 STREET	17 AVENUE	201 STREET
52772396	12/12/2021 12:59:29 AM		NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	Residential Building/House	10011	250 WEST 14 STREET	WEST 14 STREET	7 AVENUE	8 AVENUE	7 AVENUE	8 AVENUE
52771873	12/12/2021 12:59:00 AM		DOT	Department of Transportation	Street Light Condition	Bracket Arm Loose		10301					EDDY STREET	VICTORY BOULEVARD
52769717	12/12/2021 12:58:39 AM		NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	Residential Building/House	11238	961 FULTON STREET	FULTON STREET	WASHINGTON AVENUE	ST JAMES PLACE	WASHINGTON AVENUE	ST JAMES PLACE
52773344	12/12/2021 12:58:08 AM		NYPD	New York City Police Department	Noise - Residential	Banging/Pounding	Residential Building/House	11372	94-06 34 AVENUE	34 AVENUE	94 STREET	JUNCTION BOULEVARD	94 STREET	JUNCTION BOULEVARD
52768256	12/12/2021 12:58:01 AM		NYPD	New York City Police Department	Illegal Parking	Blocked Hydrant	Street/Sidewalk	10467	2447 OLINVILLE AV...	OLINVILLE AVENUE	WARING AVENUE	MACE AVENUE	WARING AVENUE	MACE AVENUE
52771899	12/12/2021 12:57:54 AM		NYPD	New York City Police Department	Blocked Driveway	No Access	Street/Sidewalk	11208	2577 PITKIN AVENUE	PITKIN AVENUE	MILFORD STREET	LOGAN STREET	MILFORD STREET	LOGAN STREET
52774183	12/12/2021 12:57:53 AM		NYPD	New York City Police Department	Noise - Commercial	Loud Music/Party	Store/Commercial	11385	60-09 70 AVENUE	70 AVENUE	60 STREET	60 LANE	60 STREET	59 LANE

We choose this dataset because it is big enough and has well-labeled fields. Which include a dictionary of what each field is:
311_SR_Data_Dictionary_2018 (1).xlsx

In the 1NF, with 41 columns available, we will make sure each record in the dataset is unique, and categorize them into four big categories: **The service request, The responding agency, The complaint details, and The location of the request.**

After the 1NF, we can now be more specific on other details for the 2NF. We will create the **service request table** to record requester information, such as what service they are requesting, the channel they are connecting from, the start and the end date of the request. Second, we will create **the complaint table**. With hundreds of complaint types, we want to know what the most complaints are about. Third, we can create **the agency table** to collect agency information. Then we can measure each department's performance, and provide proper training. Fourth, as we understand the requester and responder, we will record location details in the **location table**. Thus, we can better allocate the government resources region by region.

Now, we can move on from the 2NF to the 3NF. Since there are tens of columns describing the location details of each incident, we can create **10 more tables** based on location, such as location by requester, by specific geo-coordination, by transportation, by community etc,. For the complaint details, it contains details about resolutions. We will take resolution details from the complaint category to **the resolution table** in order to check the progress of the service request. While each agency, or department, can receive different types of complaints, we will have an **agency _complaint table** to document the details. In this stage, the 15 table creations are completed.

Data insertion ipynb
APAN5310_PROJECT_Data_Transformation.ipynb

Our ten analytical procedures are as follows:

- Which departments receive the highest incident volume?
- What is the rate of call completions? Are there any differences between the departments?
- How fast are issues resolved per department from start to finish?
- What are the top complaint types?
- Are there areas/zones where call volume is particularly high?
- Are certain times of the day where the call volume spikes?
- Are there any open calls that took longer than 30 days to resolve? 60 days?
- How many calls are left unresolved per year?
- What is the volume of calls for each data type? How has that changed through the years?
- What is the most used open data channel type currently?
  https://github.com/dogedaze/sql_term_project

Analytics Applications

After putting it into the database, our customer interaction plan is broken into several levels:

Interactive dashboards can benefit analysts, their managers, and c-level executives. Dashboards make it easy to separate a large amount of data into something more specific to the end user's roles. It can also simplify analytics into something for anyone, no matter their programming experience, to understand the data. Especially as a majority of decision makers in a business only have basic skills in programming, and data analytics. Most BI softwares prefer hosting on cloud, since NYCOpenData already includes the data for public use, they can continue using their cloud based hosting. This is even more important with the amount of agencies within their jurisdiction. This way the data would be up to date when making reports.

On the analyst level, dashboards are great to make sure everything is running but they should be able to access certain parts of the schemas and views to generate reports of their own. Direct querying would also be helpful but would be needed. Views could be created for analysts and separated by department. The risks that come with direct querying would be users accidentally deleting data, or dropping tables. This will be solved through versioning the data.

Managers would receive reports dealing mainly with their department, and other internal information. The dashboard would further be developed to include department-specific tables that include more information on the closing of the call or ticket. For example, the New York Police Department can use the database to see incidents and replan their beat routes.

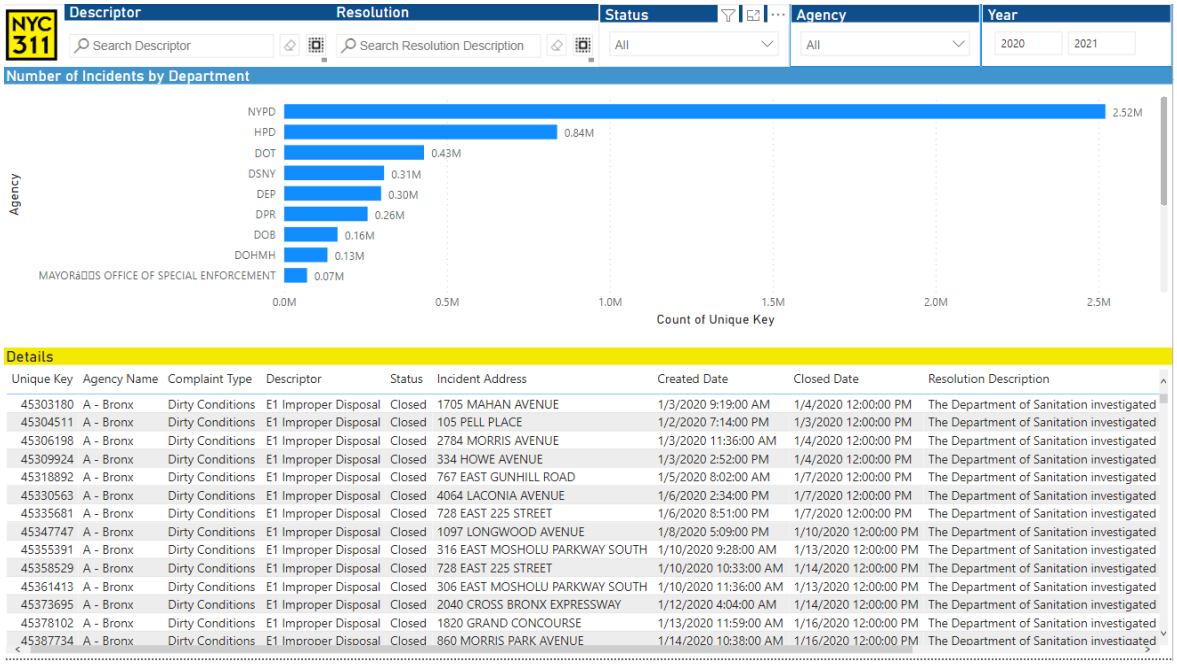


Figure 1:

A simplified dashboard where the end user can browse through incident reports without needing to know SQL or any other programming language.

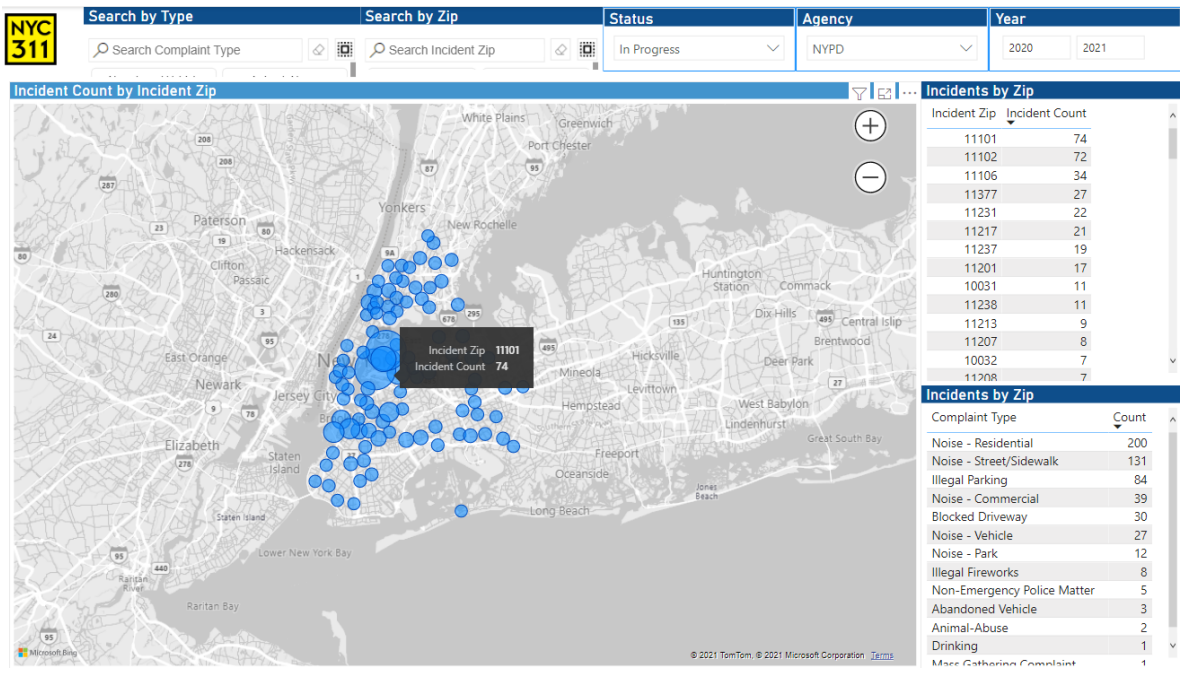


Figure 2:

Used as an example: NYPD managers can quickly browse and see where they have the most open incidents.

The charts are have connected interactions so that you can change the map based on complaint type, and other factors

C-level executives would receive the big picture, and summary level reports. These dashboards would provide more of an overview of their department, and how it plays into the larger call center percentages. To further develop this dashboard, budgeting actuals and projected spending can also be included down the line to coordinate future fiscal year workload predictions.

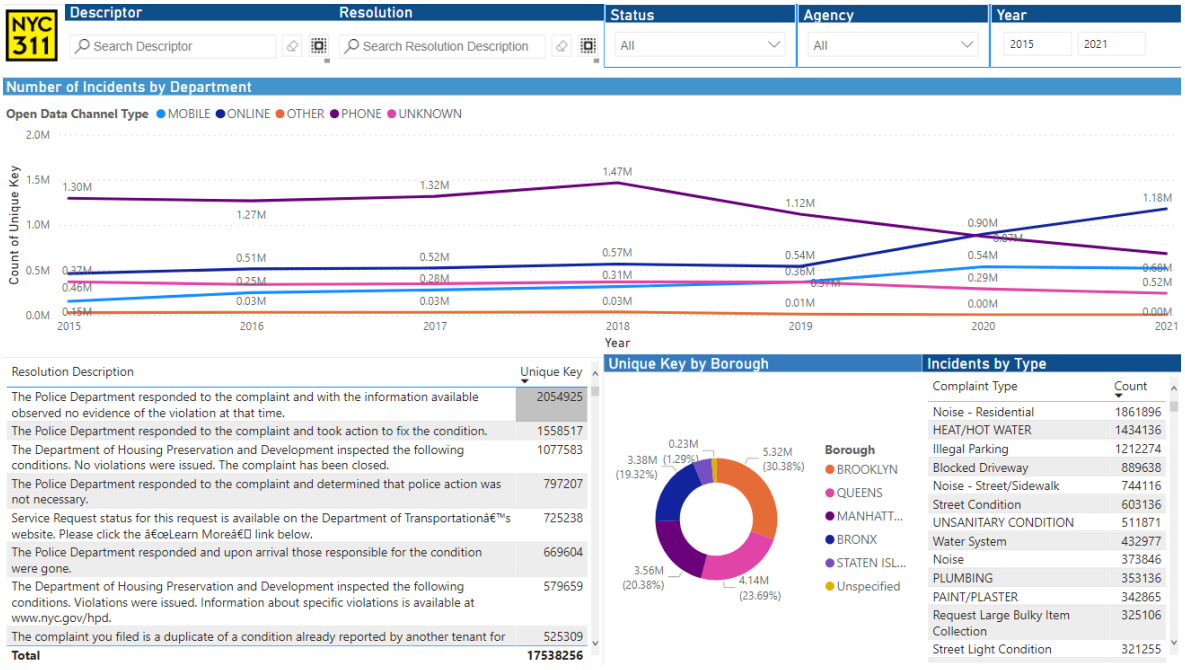


Figure 3:

Used as an example: Just from a quick glance, a C-level executive can see the major drop in call center usage since the start of the pandemic.

The charts are connected interactions so that you can also check by incident type or

resolution.

For redundancy, it would be inevitable that it happens as departments want more control over their data, and hire their own analysts. Another reason for redundancy would be if each department uses different reporting methods. For example, one only purchases licenses for powerBI, but another prefers more open source options such as Metabase.

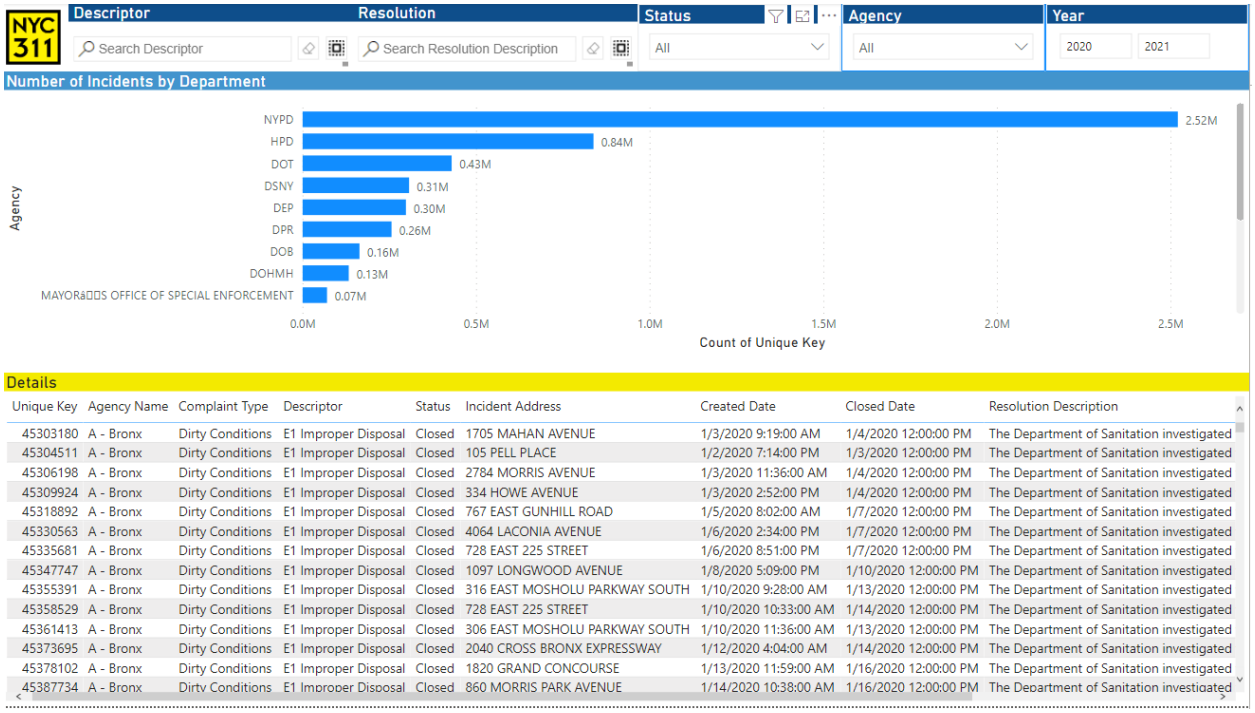
The performance will lag as more data is included in but that can be addressed further with data lakes. Through that, we can have multiple queries, big data analytics, full-text search, and real-time analytics.

Conclusion

Our goals of helping managers, and executives make informed decisions by understanding their data more effectively were met through the dashboards. These decisions included understanding their incidents better, their volume, and other analytical procedures. The RDMS implementation aided the client in not being overwhelmed by the amount of data that comes from years of incident transactions. By breaking the incidents into 4 major categories, the data is easier to comprehend at first glance.

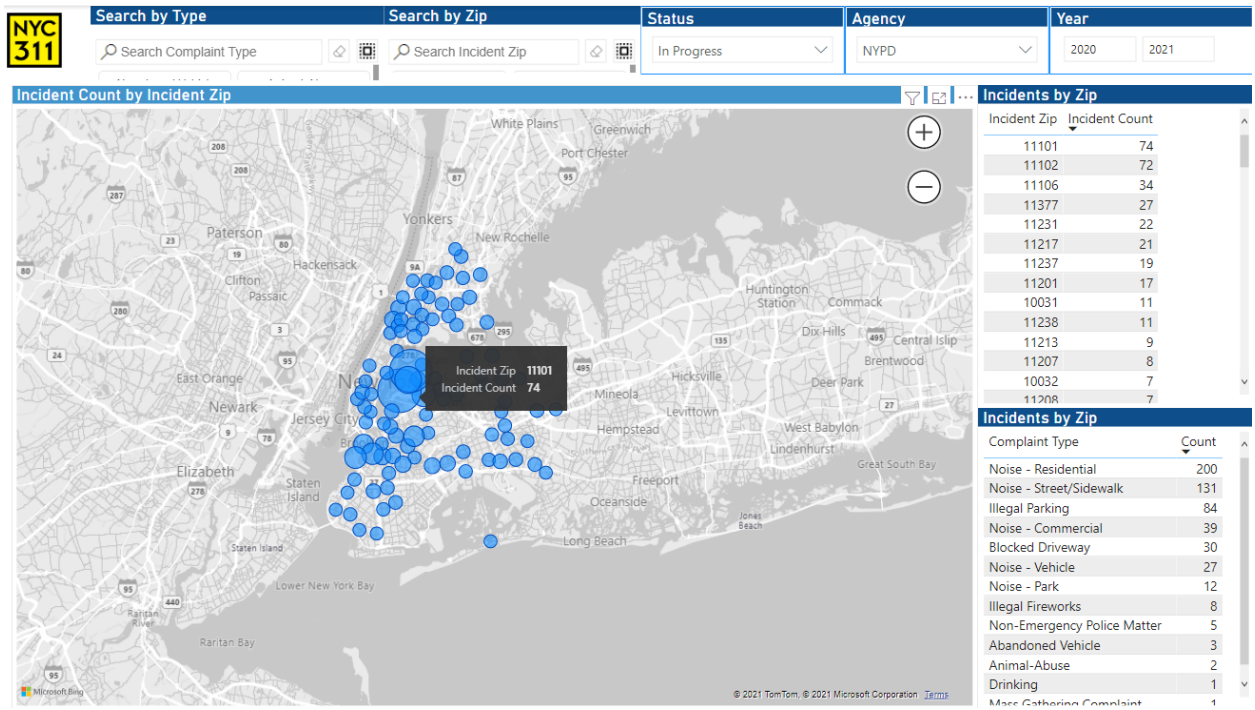
Appendix

Figure 1:



Unique Key	Agency Name	Complaint Type	Descriptor	Status	Incident Address	Created Date	Closed Date	Resolution Description
45303180	A - Bronx	Dirty Conditions	E1 Improper Disposal	Closed	1705 MAHAN AVENUE	1/3/2020 9:19:00 AM	1/4/2020 12:00:00 PM	The Department of Sanitation investigated
45304511	A - Bronx	Dirty Conditions	E1 Improper Disposal	Closed	105 PELL PLACE	1/2/2020 7:14:00 PM	1/3/2020 12:00:00 PM	The Department of Sanitation investigated
45306198	A - Bronx	Dirty Conditions	E1 Improper Disposal	Closed	2784 MORRIS AVENUE	1/3/2020 11:36:00 AM	1/4/2020 12:00:00 PM	The Department of Sanitation investigated
45309924	A - Bronx	Dirty Conditions	E1 Improper Disposal	Closed	334 HOWE AVENUE	1/3/2020 2:52:00 PM	1/4/2020 12:00:00 PM	The Department of Sanitation investigated
45318892	A - Bronx	Dirty Conditions	E1 Improper Disposal	Closed	767 EAST GUNHILL ROAD	1/5/2020 8:02:00 AM	1/7/2020 12:00:00 PM	The Department of Sanitation investigated
45330563	A - Bronx	Dirty Conditions	E1 Improper Disposal	Closed	4064 LACONIA AVENUE	1/6/2020 2:34:00 PM	1/7/2020 12:00:00 PM	The Department of Sanitation investigated
45335681	A - Bronx	Dirty Conditions	E1 Improper Disposal	Closed	728 EAST 225 STREET	1/6/2020 8:51:00 PM	1/7/2020 12:00:00 PM	The Department of Sanitation investigated
45347747	A - Bronx	Dirty Conditions	E1 Improper Disposal	Closed	1097 LONGWOOD AVENUE	1/8/2020 5:09:00 PM	1/10/2020 12:00:00 PM	The Department of Sanitation investigated
45355391	A - Bronx	Dirty Conditions	E1 Improper Disposal	Closed	316 EAST MOSHOLU PARKWAY SOUTH	1/10/2020 9:28:00 AM	1/13/2020 12:00:00 PM	The Department of Sanitation investigated
45358529	A - Bronx	Dirty Conditions	E1 Improper Disposal	Closed	728 EAST 225 STREET	1/10/2020 10:33:00 AM	1/14/2020 12:00:00 PM	The Department of Sanitation investigated
45361413	A - Bronx	Dirty Conditions	E1 Improper Disposal	Closed	306 EAST MOSHOLU PARKWAY SOUTH	1/10/2020 11:36:00 AM	1/13/2020 12:00:00 PM	The Department of Sanitation investigated
45373695	A - Bronx	Dirty Conditions	E1 Improper Disposal	Closed	2040 CROSS BRONX EXPRESSWAY	1/12/2020 4:04:00 AM	1/14/2020 12:00:00 PM	The Department of Sanitation investigated
45378102	A - Bronx	Dirty Conditions	E1 Improper Disposal	Closed	1820 GRAND CONCOURSE	1/13/2020 11:59:00 AM	1/16/2020 12:00:00 PM	The Department of Sanitation investigated
45387734	A - Bronx	Dirty Conditions	E1 Improper Disposal	Closed	860 MORRIS PARK AVENUE	1/14/2020 10:38:00 AM	1/16/2020 12:00:00 PM	The Department of Sanitation investigated

Figure 2:



Incident Zip	Incident Count
11101	74
11102	72
11106	34
11377	27
11231	22
11217	21
11237	19
11201	17
10031	11
11238	11
11213	9
11207	8
10032	7
11208	7

Incidents by Zip

Complaint Type	Count
Noise - Residential	200
Noise - Street/Sidewalk	131
Illegal Parking	84
Noise - Commercial	39
Blocked Driveway	30
Noise - Vehicle	27
Noise - Park	12
Illegal Fireworks	8
Non-Emergency Police Matter	5
Abandoned Vehicle	3
Animal-Abuse	2
Drinking	1
Mass Gathering Complaint	1

Figure 3:

NYC 311

Descriptor Search Descriptor

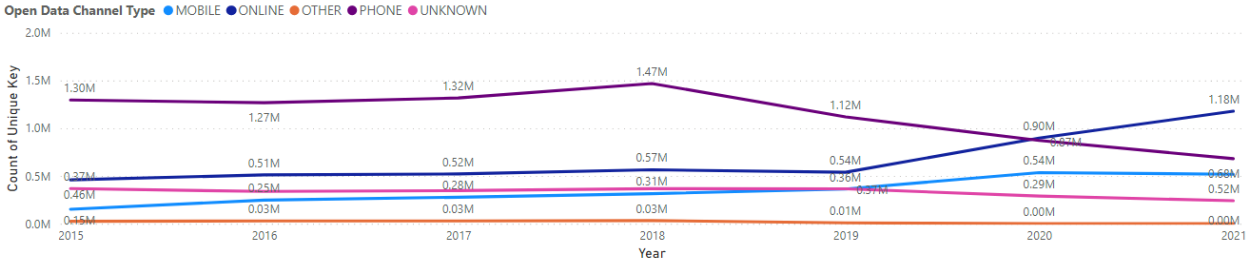
Resolution Search Resolution Description

Status All

Agency All

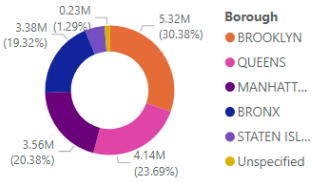
Year 2015 2021

Number of Incidents by Department



Resolution Description	Unique Key
The Police Department responded to the complaint and with the information available observed no evidence of the violation at that time.	2054925
The Police Department responded to the complaint and took action to fix the condition.	1558517
The Department of Housing Preservation and Development inspected the following conditions. No violations were issued. The complaint has been closed.	1077583
The Police Department responded to the complaint and determined that police action was not necessary.	797207
Service Request status for this request is available on the Department of Transportationâ€™s website. Please click the â€œLearn Moreâ€ link below.	725238
The Police Department responded and upon arrival those responsible for the condition were gone.	669604
The Department of Housing Preservation and Development inspected the following conditions. Violations were issued. Information about specific violations is available at www.nyc.gov/hpd.	579659
The complaint you filed is a duplicate of a condition already reported by another tenant for	525309
Total	**17538256**

Unique Key by Borough



Incidents by Type

Complaint Type	Count
Noise - Residential	1861896
HEAT/HOT WATER	1434136
Illegal Parking	1212274
Blocked Driveway	889638
Noise - Street/Sidewalk	744116
Street Condition	603136
UNSANITARY CONDITION	511871
Water System	432977
Noise	373846
PLUMBING	353136
PAINT/PLASTER	342865
Request Large Bulky Item Collection	325106
Street Light Condition	321255