NYC OpenData

NYC 311 Reporting

New York 311 Service Requests Database System

Group 7: Nuo Chen, Yuelin Guo, Jiaxin Shen,
Xuefan Han, Xinying Li, Xinxin Zhang

Client Scenario

Background

- Majority of 311 calls are handled by customer service representatives who have been trained to send requests to the appropriate departments
- Peak volume -received more than 180k calls per day
- Expensive manpower with limited resources

Goal & Plan

- Automated system to extract relevant information
- Construct a relational database
  - Easier to sort and find information
- Improved efficiency of retrieval and automate certain inquiries

Client

-New York City Government -NYC311

Original Data

```
data.head()
```

	Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...
0	53821441	04/04/2022 12:00:03 AM	04/04/2022 12:10:07 AM	NYPD	New York City Police Department	Noise – Residential	Loud Music/Party	Residential Building/House	10466.0	655 EAST 230 STREET	...
1	53818157	04/04/2022 12:00:07 AM	04/04/2022 01:23:48 AM	NYPD	New York City Police Department	Blocked Driveway	Partial Access	Street/Sidewalk	10006.0	8 ALBANY STREET	...
2	53825706	04/04/2022 12:00:37 AM	04/06/2022 03:02:03 PM	HPD	Department of Housing Preservation and Develop...	WATER LEAK	HEAVY FLOW	RESIDENTIAL BUILDING	10031.0	385 CONVENT AVENUE	...
3	53823230	04/04/2022 12:00:48 AM	04/06/2022 04:17:01 PM	HPD	Department of Housing Preservation and Develop...	HEAT/HOT WATER	ENTIRE BUILDING	RESIDENTIAL BUILDING	11226.0	175 LOTT STREET	...
4	53818619	04/04/2022 12:00:57 AM	04/04/2022 02:45:47 AM	NYPD	New York City Police Department	Illegal Parking	Blocked Hydrant	Street/Sidewalk	11385.0	1711 HARMAN STREET	...

5 rows × 41 columns

28.1 million Rows

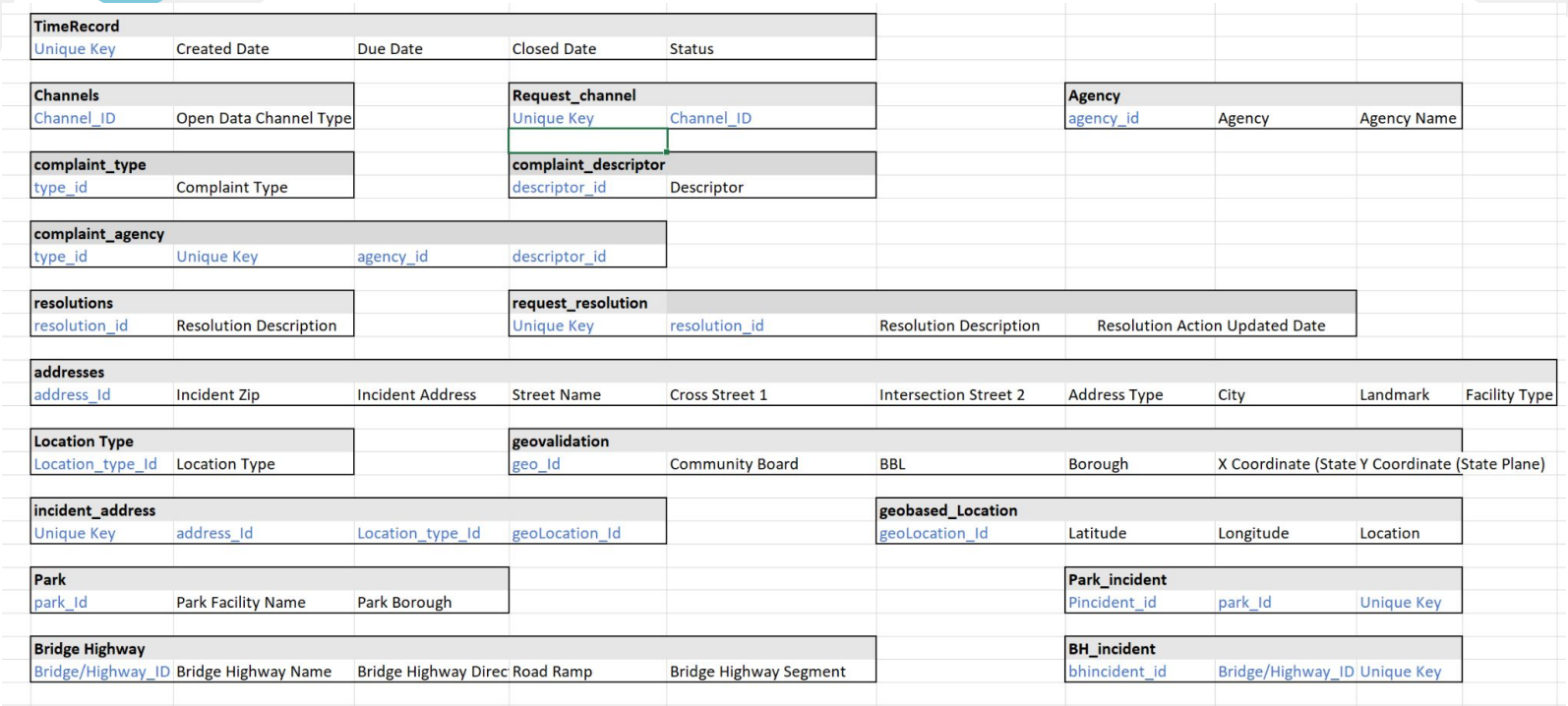
Includes 311 service requests from 2010 to present

41 Columns

Consists 32 text columns, 4 number columns, 4 date column, 1 location column

The original data was retrieved from NYC Opendata.com using the link: https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9/data

Normalization Plan



TimeRecord				
Unique Key	Created Date	Due Date	Closed Date	Status

Channels	
Channel_ID	Open Data Channel Type

Request_channel	
Unique Key	Channel_ID

Agency		
agency_id	Agency	Agency Name

complaint_type	
type_id	Complaint Type

complaint_descriptor	
descriptor_id	Descriptor

complaint_agency			
type_id	Unique Key	agency_id	descriptor_id

resolutions	
resolution_id	Resolution Description

request_resolution			
Unique Key	resolution_id	Resolution Description	Resolution Action Updated Date

addresses									
address_Id	Incident Zip	Incident Address	Street Name	Cross Street 1	Intersection Street 2	Address Type	City	Landmark	Facility Type

Location Type	
Location_type_Id	Location Type

geovalidation						
geo_Id	Community Board	BBL	Borough	X Coordinate (State	Y Coordinate (State Plane)	

incident_address			
Unique Key	address_Id	Location_type_Id	geoLocation_Id

geobased_Location			
geoLocation_Id	Latitude	Longitude	Location

Park		
park_Id	Park Facility Name	Park Borough

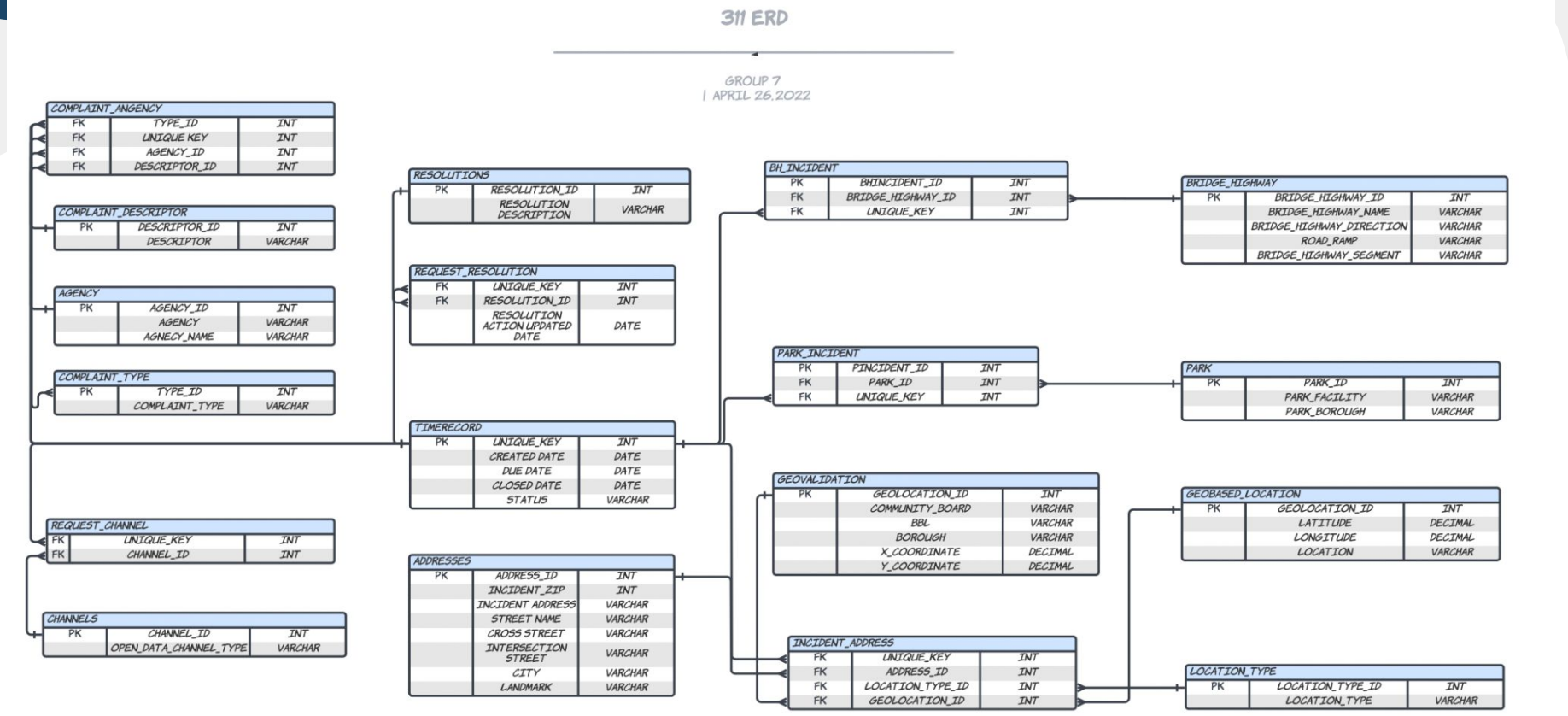
Park_incident		
Pincident_id	park_Id	Unique Key

Bridge Highway				
Bridge/Highway_ID	Bridge Highway Name	Bridge Highway Direc	Road Ramp	Bridge Highway Segment

BH_incident		
bhincident_id	Bridge/Highway_ID	Unique Key

- 1NF: the original data is already in 1NF. All table attributes are atomic and without repeating attributes
- 2NF: divided the original dataset into 12 separate table to make sure every non-key attribute are fully dependent on the key
- 3NF: divided 2NF tables into 18 separate table to insure all attributes are ONLY depend on the key

ERD



ERD link:
https://lucid.app/lucidchart/58b520fd-55df-468f-9c7a-a8b137ebee48/edit?invitationId=inv_7f9b4cba-4cdd-451f-a34e-33ee0ca3106f

ETL Process

01 Summary

A flow chart of our ETL process

02 Extract

- Create connection to PostgreSQL

- Extract data from original data source

Create Connection

```
[ ]  # Define the connection URL:
     conn_url = 'postgresql://postgres:123@localhost:5432/5310_test3'

     # Create an engine that connects to PostgreSQL:
     engine = create_engine(conn_url)

     # Establish a connection:
     connection = engine.connect()
```

Insert Table

```
[ ]  data = pd.read_csv('311_Service_Requests_from_2010_to_Present.csv',encoding='latin-1')
```

ETL Process

03 Transform

- Rename the original columns
- Remove duplicates
- Add incrementing integers to the subset data frame

9.Geobased_location table

```
col_names2 = ['Latitude', 'Longitude', 'Location']
data_geobased_location = pd.concat([data_geovalidation['geolocation_id'],
                                    data[col_names2].reset_index(drop=True)],axis = 1)
data_geobased_location = data_geovalidation.rename(columns = {'Latitude':'latitude','Longitude':'longitude','Location':'location'})
data_geobased_location.drop_duplicates()
```

	geolocation_id	community_board	bbl	borough	x_coordinate	y_coordinate
0	1	12 BRONX	2.048330e+09	BRONX	1022911.0	264242.0
1	2	01 MANHATTAN	1.000538e+09	MANHATTAN	980494.0	197708.0
2	3	09 MANHATTAN	1.020610e+09	MANHATTAN	999318.0	240050.0
3	4	17 BROOKLYN	3.051360e+09	BROOKLYN	997004.0	174498.0
4	5	05 QUEENS	4.034320e+09	QUEENS	1008154.0	196140.0

04 Load

- Load the transformed data into SQL database

```
data_geobased_location.to_sql(name='geobased_Location', con=engine, if_exists='append', index=False)
```

Analytical Procedures

Needs

- Monitor overall incident cases
- Make accurate record and follow-up
- Detect changes and take timely actions
- Use most updated information to make data-driven decisions

Sample Procedures

- Number of complaints and trend
- Agencies and related cases
- Complaint types in terms of region and time
- Resolution

Number of complaints
1. What are the top 3 areas that have the most complaints in New York?
2. Which bridge has the most complaints?
3. Which boroughs have the most call incidents?

Agency
4. What is the daily average number of 311 calls received by each agency?
5. Which agency has the most incidents?

Complaint types
6. Which complaint type has the most calls in each region?
7. What are the most complaints reported in Manhattan?
8. What is the accumulated sum for each type of complaint by time series?

Resolutions
9. What's the average resolution speed for each complaint type?
10. Which agency solves the problem fastest?

Version 4 pgAdmin

python™

Analytical Procedures

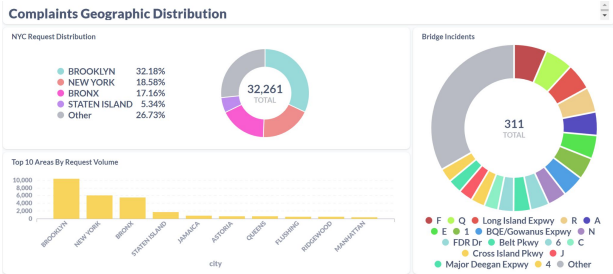
Analyst Level

- Interact directly through PostgreSQL, PgAdmin, and self-serve reporting capabilities on Metabase
- Manipulate information and data according to organizational needs
- Analyze current situations to deliver insights to refine strategies

```
select complaint_type, count(*) as total from timerecord tt
left join incident_address ia
on tt.unique_key = ia.unique_key
left join complaint_agency ca
on tt.unique_key = ca.unique_key
left join complaint_type ct
on ca.type_id = ct.type_id
left join address add
on ia.address_id = add.address_id
where city = 'MANHATTAN'
group by complaint_type
order by total desc;
```

C-Level

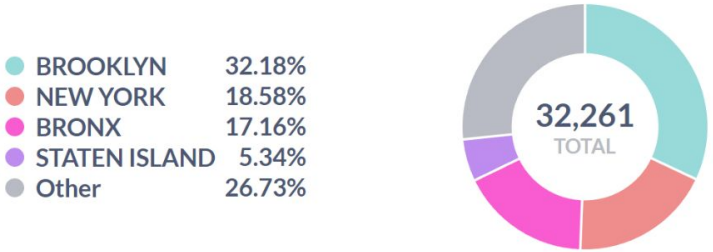
- Monthly, quarterly, annually reports via Metabase
- High level, aggregated metrics in clear presentations
- Resource relocations and staffing strategies



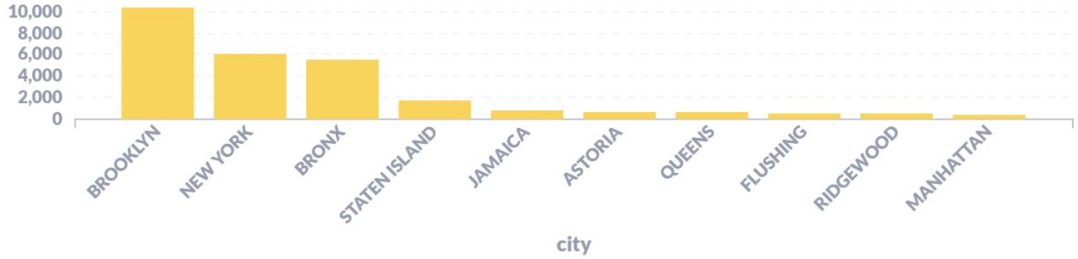
Metabase Demo

Complaints Geographic Distribution

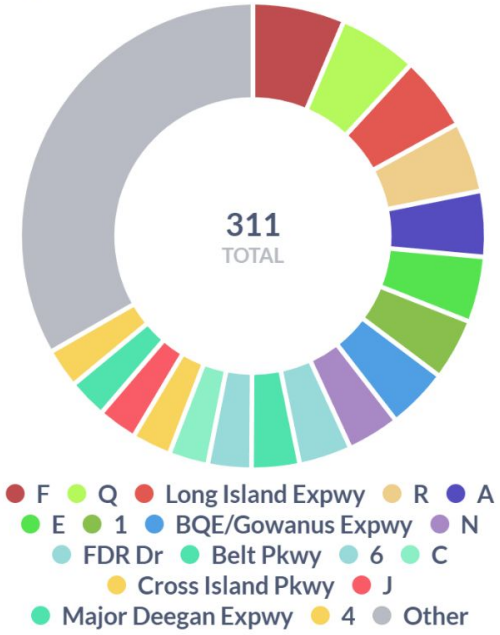
NYC Request Distribution



Top 10 Areas By Request Volume



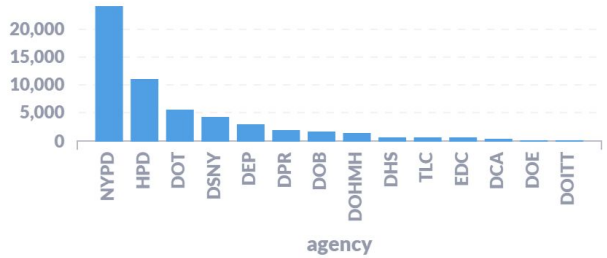
Bridge Incidents



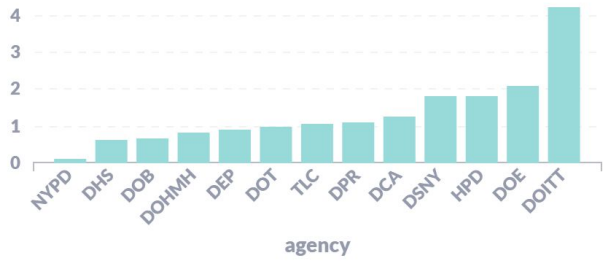
Metabase Demo

Agency Performance

Agency Complaint Count



Agency Resolution Speed



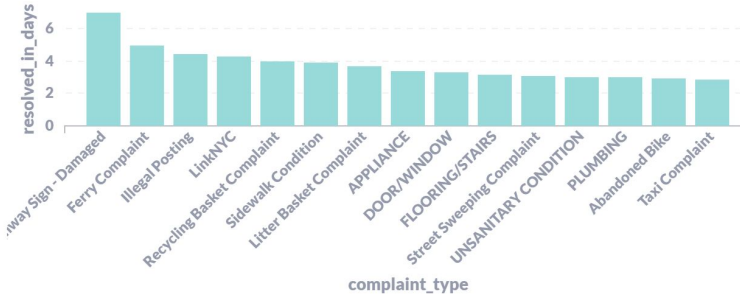
Agency for Each Complaint Type

complaint_type	agency_name
Abandoned Bike	Department of Sanitatio
Abandoned Vehicle	New York City Police De
AHV Inspection Unit	Department of Buildings
Air Quality	Department of Environn
Animal-Abuse	New York City Police De
Animal Facility - No Permit	Department of Health a
Animal in a Park	Department of Parks an
APPLIANCE	Department of Housing
Asbestos	Department of Environn
Asbestos	Department of Health a
Beach/Pool/Sauna Complaint	Department of Health a
BEST/Site Safety	Department of Buildings

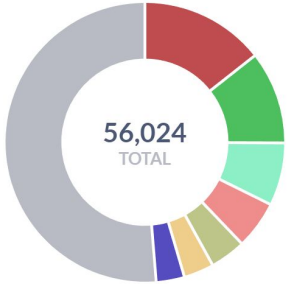
Rows 1-12 of 153

Request Evaluation

Top 15 Time-Consuming Complaints



Most Frequent Complaint Type



Illegal Parking	14.42%
Noise - Residential	10.66%
HEAT/HOT WATER	7.12%
Blocked Driveway	5.38%
Street Condition	3.93%
Noise - Street/Sidewalk	3.28%
UNSANITARY CONDITION	3.05%
Other	52.17%

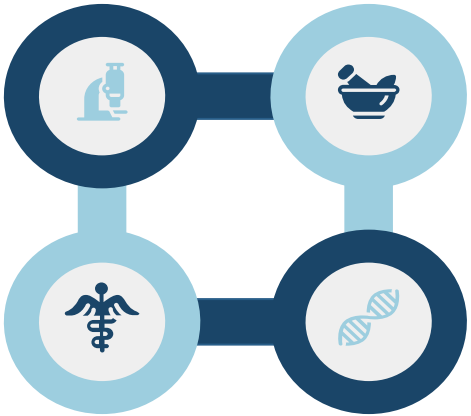
Conclusion

01 Brooklyn Area

-Highest volume of complaints, 32.18%
-Gov needs to focus on the regulation

02 Complaints Type

-Noise complaints
-Illegal parking
-Heat/hot water issue



03 Agency Performance

-Most busy agency: NYPD (20k+)
-Fastest agency resolution speed: DOITT

04 Time-consuming Complaints

-Sign-damaged (up to 6 days)
-Ferry complaint (4-6 days)
-Illegal posting (around 4 days)

Our project gives the NY Gov a clear and vivid information on the complaints that NY citizens having and facing. The result can help government to adjust current policy to benefit the whole state better.

Thank You!
Questions?