Analyzing the Effect of Open Street Location Program on the Noise Complaints

Final Project Report

Due 12/13/2023

CIS 4400 Group 9 Section CTRA

Edward Shin | Jerome Galam | Laney Zou | Tsering Lama | Denis Dudkin
edward.shin2@baruchmail.cuny.edu | jerome.galam@baruchmail.cuny.edu
laney.zou@baruchmail.cuny.edu | tsering.lama2@baruchmail.cuny.edu
denis.dudkin@baruchmail.cuny.edu

Introduction

Problem

The objective of this project is to determine if there is a correlation between noise complaints and the NYC Open Street Locations program. During the formulation stage of our project, we hypothesized that there would be a positive correlation between the volume of noise complaints associated and the length/amount of times some specific area has received a noise complaint. Our belief was that when some area is reserved for festivities or public events it is likely that there will be a lot of noise and commotion resulting in some sort of formal complaint and that less complaints or no complaints would be reserved for areas that do not have much reservations. However, it is important to mention that we did expect some regions could have an unusually high volume of noise complaints without having scheduled events and these findings would be overlooked in our test results. Both data sources are from NYC Open Data - and the 311 Complaint/Service Request are from the 311 and DoITT and the Open Street Location dataset is from the Department of Transportation (DOT). All of the data was accessed from the beginning of our project on October 10th, 2023 until our final report which is due December 13, 2023.

NYC Open Data Link Address:
https://opendata.cityofnewyork.us/

What do we aim to do with this information?

Assuming that our potential hypothesis of there being a positive correlation between the amount of noise complaints and open street locations we would hope to amend that issue moving forward. Ultimately, if there are particular locations that have a very strongly positive correlation between complaints and that area being an open street being meaningful we believe that the opportunity to reserve those locations should be terminated. However, if there is no correlation between those two factors it would be good to know that open street locations don't necessarily have a strong impact on the noise complaints of the area so it would be fair to overlook this as an issue in the future if someone thought an open street could cause this type of problem. Finally, the data collected from this information would further bolster the authenticity and severity of noise complaints in concentrated areas as this would imply the chaos and noise that is occurring at some specific location is greater than what a potential open street could cause. This would mean that the city should look into those problems and address them as they have some validity in severity.

Description

For the purposes of this project, we will be utilizing two datasets; the 311 Service Request/Complaint Dataset and the Open Street Locations Dataset will be analyzed across different granular levels to produce findings and draw our conclusions. In consideration of that, we have included a short description of both datasets below:

1. 311 Service Request/Complaint
   311 Service Requests from 2010 to Present | NYC Open Data

In this dataset New York City residents call 311 for information about city services or report issues that they experience for the city to provide resources or resolve. This dataset contains information on reports made to 311 and includes attributes such as location, agency the report pertains to, time/date, type of report, street, and more. For The purposes of our project we will be querying this data to utilize the information regarding Residential Noise Complaints that will be compared to our next dataset.

2. Open Street Locations
   Open Streets Locations

This dataset gives information about the New York City Open Street Program. This program limits traffic in designated areas (closes the street) for some designated and requested period of time. The purpose for these "open streets" is to close street(s) to provide a safer space for city programs, community events, and much more. The attributes that this dataset provides would be the organization that is requesting the open street, the purpose of the open street, the date or time that set of streets are to be deemed "open", and the start to end street of that "open street". We hope to utilize these attributes to draw some conclusions after merging both this database and the relevant information from the 311 Complaint/Service Request dataset.

311 Service Request from 2010 to Present | NYC Open Data Link Address:
https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9

Open Streets Locations Link Address:
https://data.cityofnewyork.us/Health/Open-Streets-Locations/uiay-nctu

Dimensional Model Diagrams

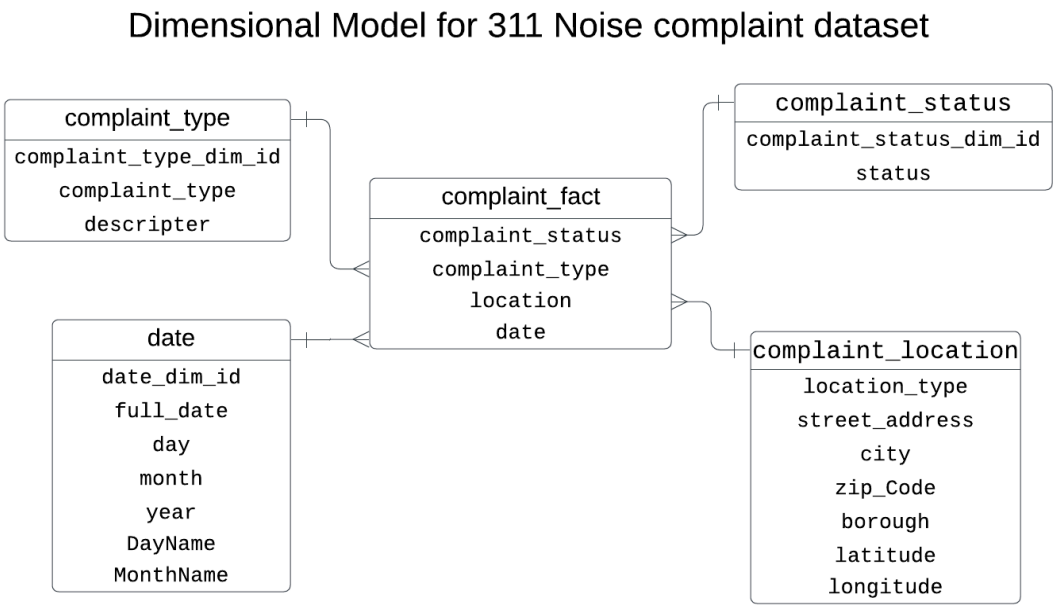


Figure 1 Dimensional Model for 311 Noise Complaint Dataset

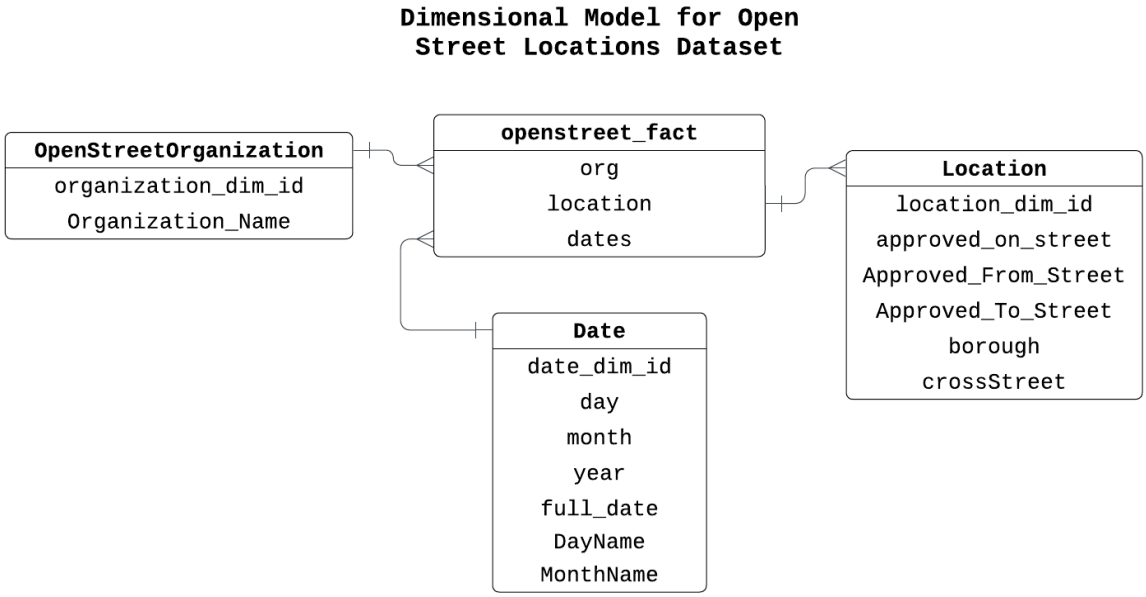


Figure 2 Dimensional Model for Open Street Location Dataset

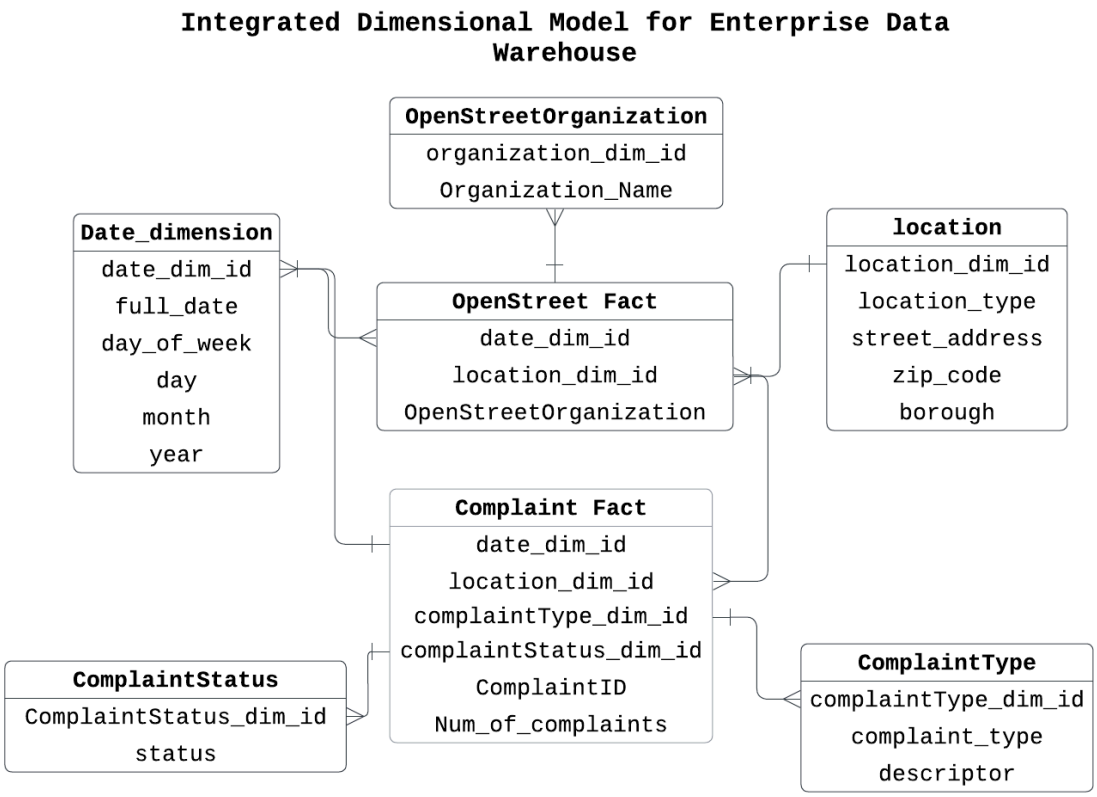


Figure 3 Integrated Dimensional Model

ELT Process

ELT process can be broken down into the 3 parts including extracting the source data using the open socrata API then loading it into our target DBMS, which for this project is Google Big Query and then performing transformation on the data.

Extraction

Before diving into the Extraction, we first performed data profiling to access the quality of both the datasets.

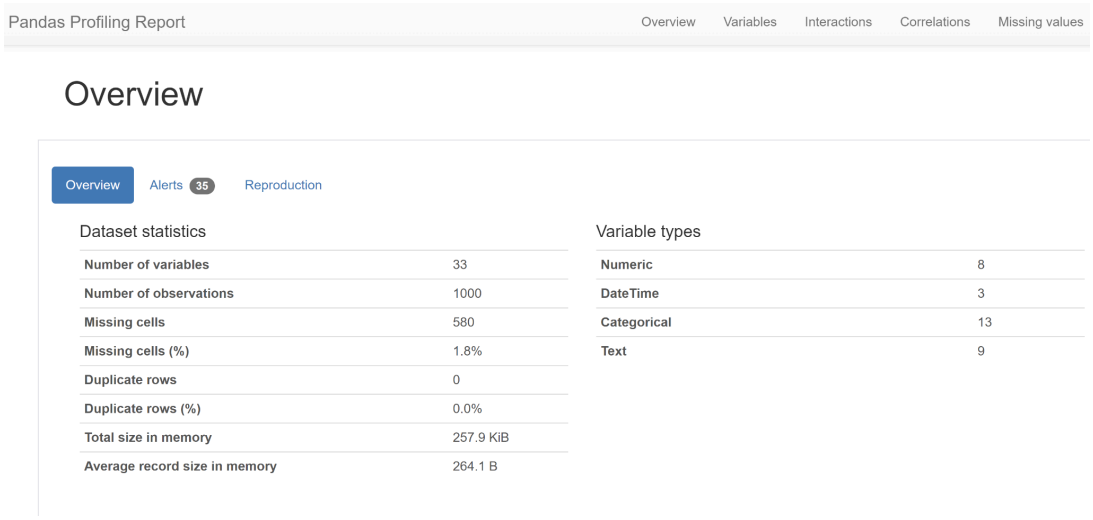


Figure 4 *Data profiling report overview of the 311 Noise Complaint dataset*

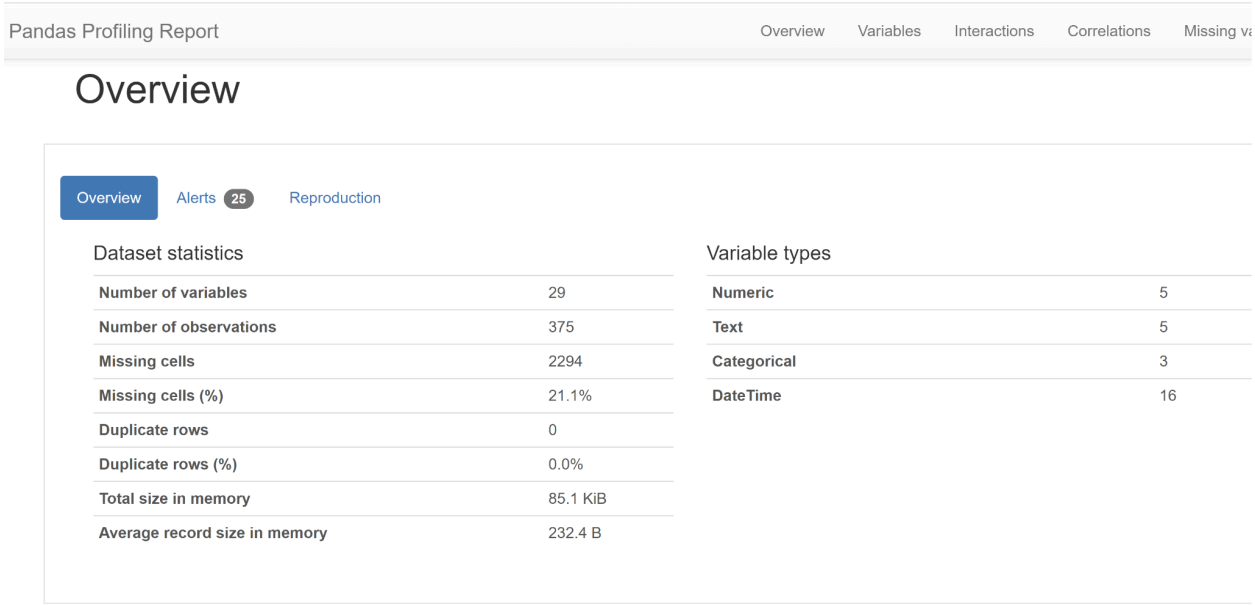


Figure 5 Data profiling report overview of the Open Street Locations dataset

Loading

As part of the loading process, the extracted data was then loaded into Google Big Query, which was connected to DBT for transformation processes.

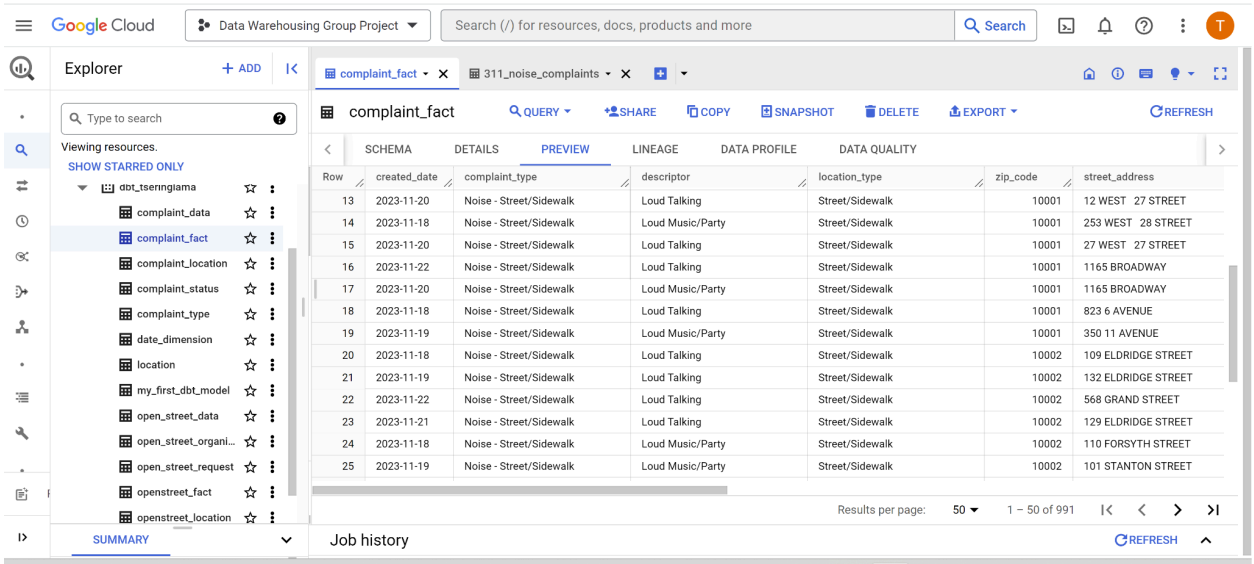


Figure 6 Google Big Query Extracted Data

Transformation

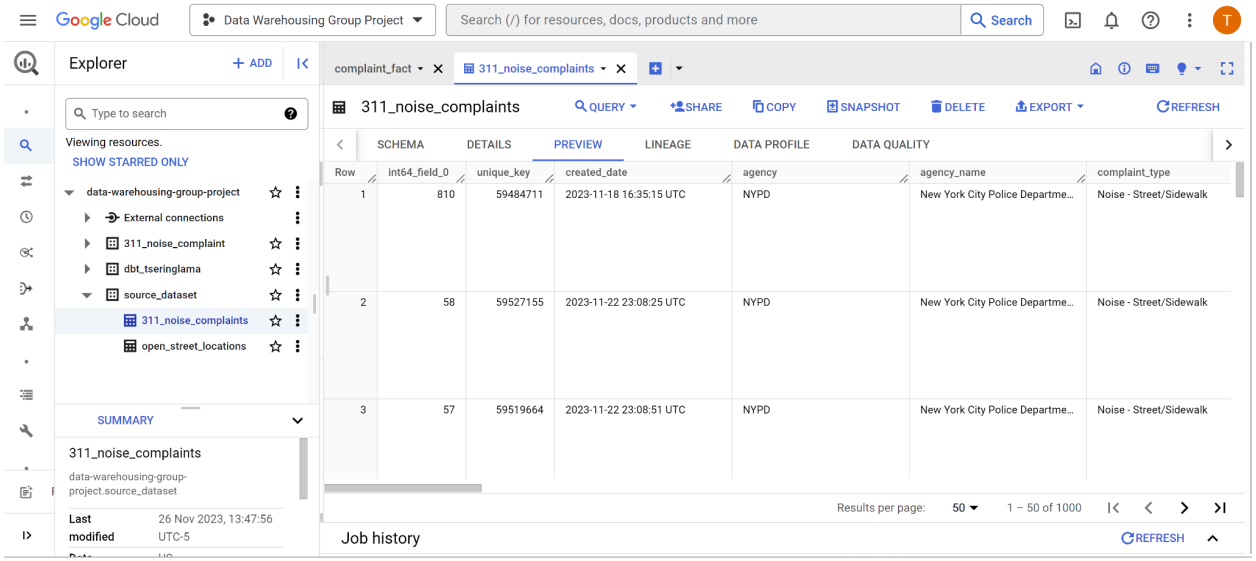


Figure 7 311 Noise-Complaint-Street/Sidewalk (2020 - Present)

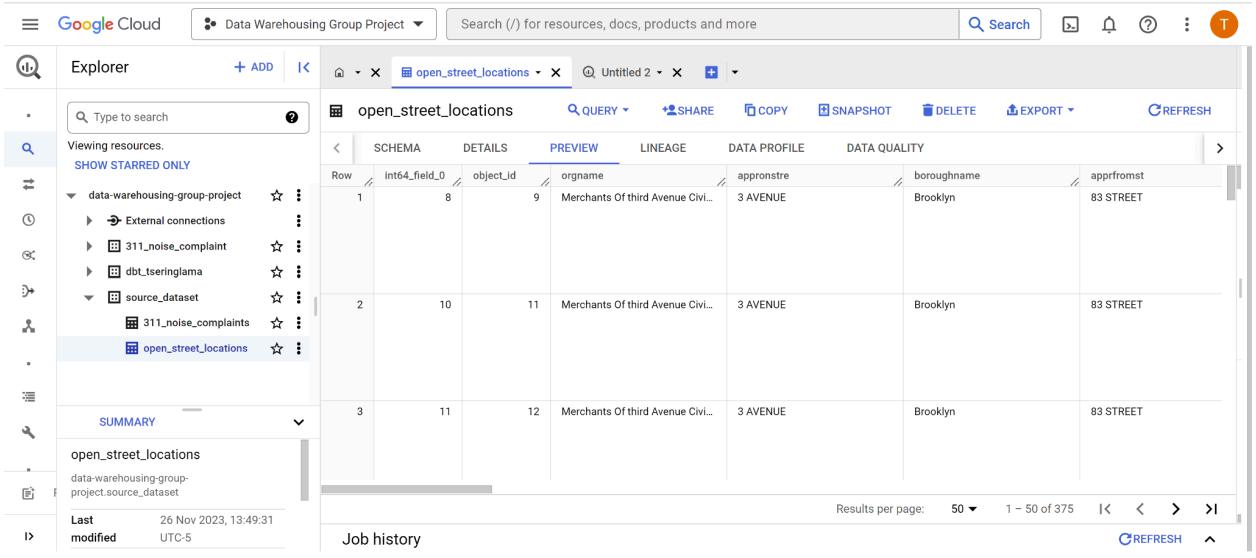


Figure 8 Open-Street locations (2020 - Present)

DBT Models

1. complaint _type:

```
SELECT row_number() OVER () AS complaint_type_dim_id, *
FROM
   (SELECT DISTINCT complaint_type, descriptor
   FROM {{ref('complaint_data')}}
   )
```

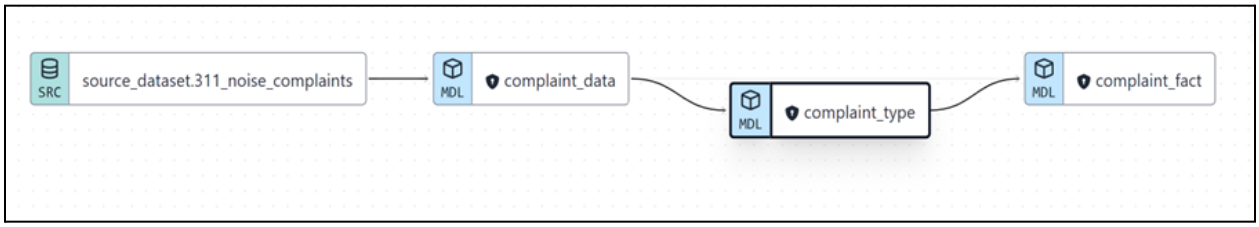


Figure 9 Complaint Type

2. complaint_status:

```
SELECT row_number() OVER () AS complaint_status_dim_id, *
FROM
   (SELECT status
   FROM {{ref('complaint_data')}}
   )
```

Figure 10 Complaint Status

3. open_street_location:

```
with openstreet_location as (
   SELECT
   DISTINCT
   approved_on_street,
   Approved_From_Street,
   Approved_To_Street,
   borough
   FROM {{ref('open_street_data')}}
)
select row_number() OVER () AS openstreet_location_dim_id, * from
openstreet_location
```

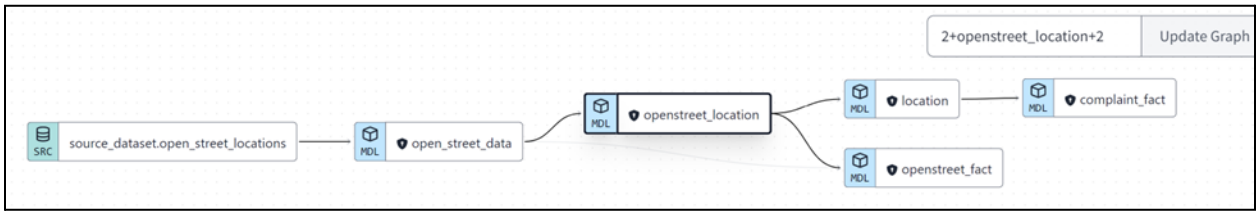


Figure 11 Open Street Location

4. complaint_location:

```
with complaint_location as (
   SELECT DISTINCT location_type, street_address,
   city, zip_code, borough, cast(latitude as STRING) as latitude,
longitude
   FROM {{ref('complaint_data')}}
)
select * from complaint_location
```

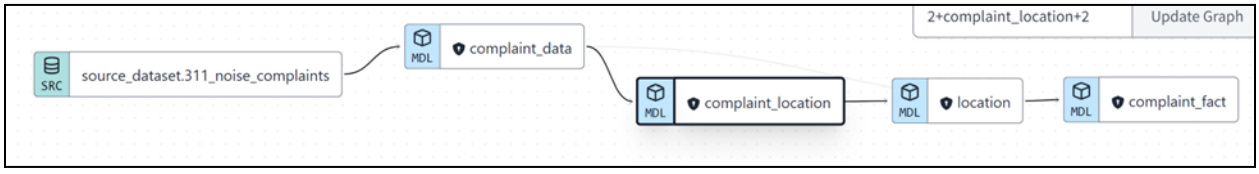


Figure 12 Complaint Location

5. location:

```
with complaint_loc as(
   select * from {{ref('complaint_location')}}
),
openstreet_loc as (
   select * from {{ref('openstreet_location')}}
),
location as (
   select
   location_type,
   street_address,
   zip_code,
from complaint_loc
full join openstreet_loc
on complaint_loc.borough = openstreet_loc.borough
where location_type IS NOT NULL AND street_address IS NOT NULL
)
select row_number() OVER () AS location_dim_id, *
from location
```

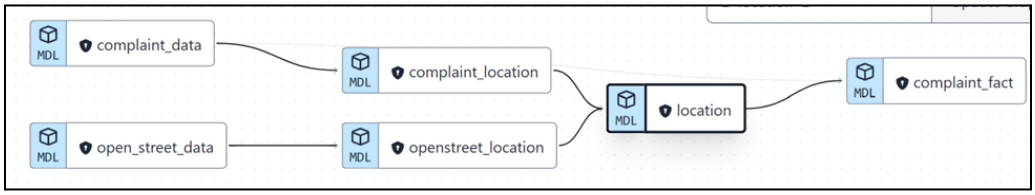


Figure 13 Location

6. OpenStreetOrganization:

```
SELECT
row_number() OVER () AS openstreet_organization_dim_id,
Organization_Name
FROM
(SELECT DISTINCT Organization_Name
  FROM {{ref('open_street_data')}}
)
```

Figure 14 Open Street Organization

7. Date_dimension:

```
/* generating dates using the macro from the dbt-utils package */
with dates_dim as (
{{ dbt_utils.date_spine(
   datepart="day",
   start_date="cast('2020-01-01' as date)",
   end_date="cast(date_add(current_date(), interval 1 day) as date)")
}})

/* extracting some date information*/
select row_number() OVER () AS date_dim_id,
   date_day as full_date,
   extract(dayofweek from date_day) as day_of_week,
   extract(day from date_day) as day,
   extract(month from date_day) as month,
   extract(year from date_day) as year
from dates_dim
```

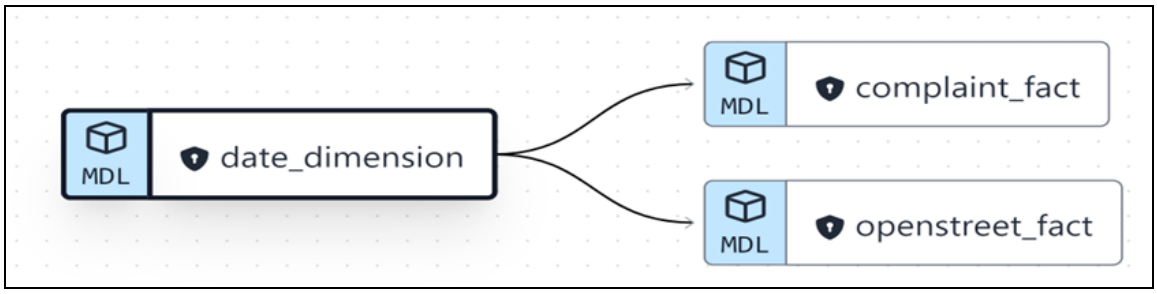


Figure 15 Date Dimension

8. vw_openstreet_startdate Dimension:

```
with dates as (
   select * from {{ref("date_dimension")}}
)
select date_dim_id as openstreet_startdate_dim_id,
full_date as start_fullDate,
day_of_week as start_dayOfWeek,
day as start_day,
month as start_month,
year as start_year
from dates
WHERE full_date IS NOT NULL
```

Figure 16 Open Street Start Date

9. vw_openstreet_enddate Dimension:

```
with dates as (
   select * from {{ ref("date_dimension")}}
)
select date_dim_id as openstreet_enddate_dim_id,
full_date as end_fullDate,
day_of_week as end_dayOfWeek,
day as end_day,
month as end_month,
Coalesce(year, 0000) as end_year
from dates
WHERE full_date IS NOT NULL
```

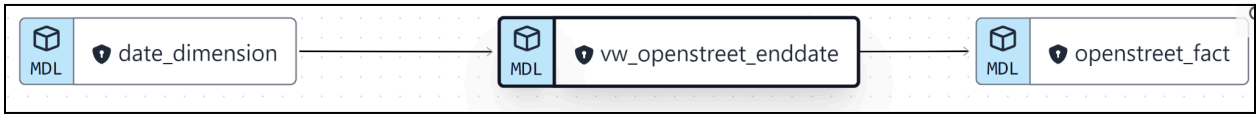


Figure 17 Open Street End Date

10. **vw_created_date Dimension:**

```
with dates as (
  select * from {{ref("date_dimension")}}
)
select date_dim_id as created_date_dim_id, full_date, day_of_week, day,
month, year
from dates
```

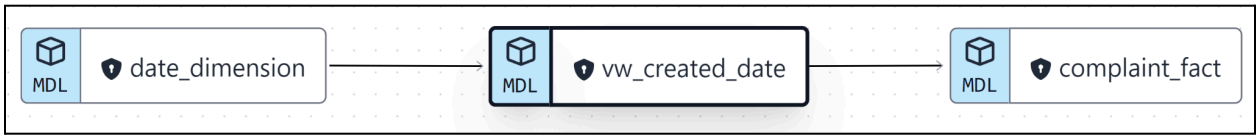


Figure 18 Created Date

11. **vw_closed_date Dimension:**

```
with dates as(
  select * from {{ref("date_dimension")}}
)
select date_dim_id as closed_date_dim_id, full_date, day_of_week, day,
month, year
from dates
```

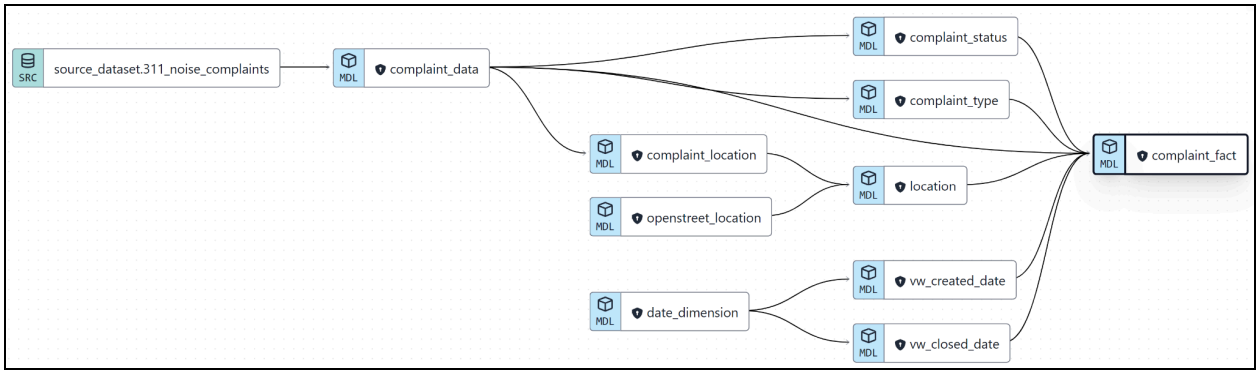
Figure 19 Closed Date

Fact Tables

1. complaint_fact

```
with
   all_complaints as
   (select * from{{ ref("complaint_data") }}),
   complaint_status as (select * from{{ ref("complaint_status") }}),
   complaint_type as (select * from{{ ref("complaint_type") }}),
   location as (select * from {{ ref('location') }}),
   created_date as (select * from{{ ref("vw_created_date")}}),
   closed_date as (select * from{{ ref("vw_closed_date") }}),

join_tbl as (
   select
      complaint_type.complaint_type_dim_id,
      created_date.created_date_dim_id,
      closed_date.closed_date_dim_id,
      location.location_dim_id,
      all_complaints.unique_key
   from all_complaints
   left join complaint_type on (all_complaints.complaint_type=complaint_type.complaint_type
AND all_complaints.descriptor= complaint_type.descriptor)
   left join location on (all_complaints.latitude=location.latitude AND
all_complaints.longitude=location.longitude)
   left join created_date on (all_complaints.created_date= created_date.full_date)
   left join closed_date on (all_complaints.closed_date = cast(closed_date.full_date as date))
)
select * from join_tbl
```



2.30s | Results limited to 500 rows. Change row display Download

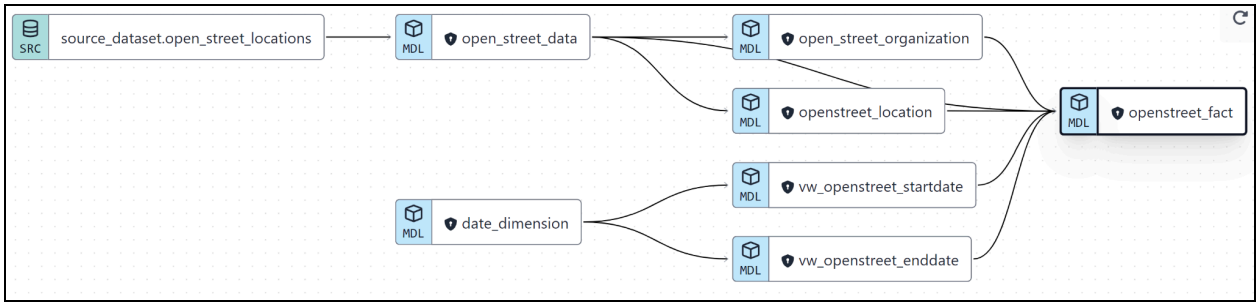
complaint_type_dim_id	created_date_dim_id	closed_date_dim_id	location_dim_id	unique_key
1	1423	1423	110	59527153
1	1423	NULL	116	59520246
1	1423	NULL	111	59527734
1	1423	NULL	114	59522806
2	1423	NULL	109	59521515
1	1423	1423	115	59522238
1	1423	1423	112	59527157
1	1423	1423	112	59522236
1	1423	NULL	108	59520247

Defer to production Ready

Figure 20 Complaint Fact Table

2. openstreet_fact

```
WITH
   all_openstreet AS (SELECT * FROM {{ ref("open_street_data") }}),
   open_street_organization AS (SELECT * FROM {{ ref("open_street_organization") }}),
   openstreet_location AS (SELECT * FROM {{ ref("openstreet_location") }}),
   start_date AS (SELECT * FROM {{ ref("vw_openstreet_startdate") }}),
   end_date AS (SELECT * FROM {{ref("vw_openstreet_enddate")}}),
join_tbl AS (
   SELECT
      open_street_organization.openstreet_organization_dim_id,
      openstreet_location.openstreet_location_dim_id,
      start_date.openstreet_startdate_dim_id,
      end_date.openstreet_enddate_dim_id
   FROM all_openstreet
   LEFT JOIN open_street_organization ON all_openstreet.organization_name =
open_street_organization.organization_name
   LEFT JOIN openstreet_location ON all_openstreet.approved_on_street =
openstreet_location.approved_on_street
   LEFT JOIN start_date on (all_openstreet.start_date= start_date.start_fullDate)
   LEFT JOIN end_date on (all_openstreet.end_date = cast(end_date.end_fullDate as date))
)
SELECT * FROM join_tbl
```



2.54s | Results limited to 500 rows. Change row display Download C

openstreet_organization_dim_id	openstreet_location_dim_id	openstreet_startdate_dim_id	openstreet_enddate_dim_id
1	1	1175	NULL
2	2	1187	NULL
2	2	1187	NULL
3	3	1248	NULL
4	4	1121	NULL
5	5	1187	NULL
6	6	1201	NULL
7	7	1208	NULL

Figure 21 Open Street Fact Table

Staging

1. complaint_data.sql

```
with Complaints as
   (select
      cast(created_date as date) as created_date,
      complaint_type,
      descriptor,
      location_type,
      incident_zip as zip_code,
      incident_address as street_address,
      city,
      borough,
      cast(latitude as STRING) as latitude,
      longitude,
      open_data_channel_type,
      unique_key,
      status
      from {{source('source_dataset','311_noise_complaints')}}
      WHERE latitude IS NOT NULL AND LONGITUDE IS NOT NULL
   )
   select* from Complaints
```

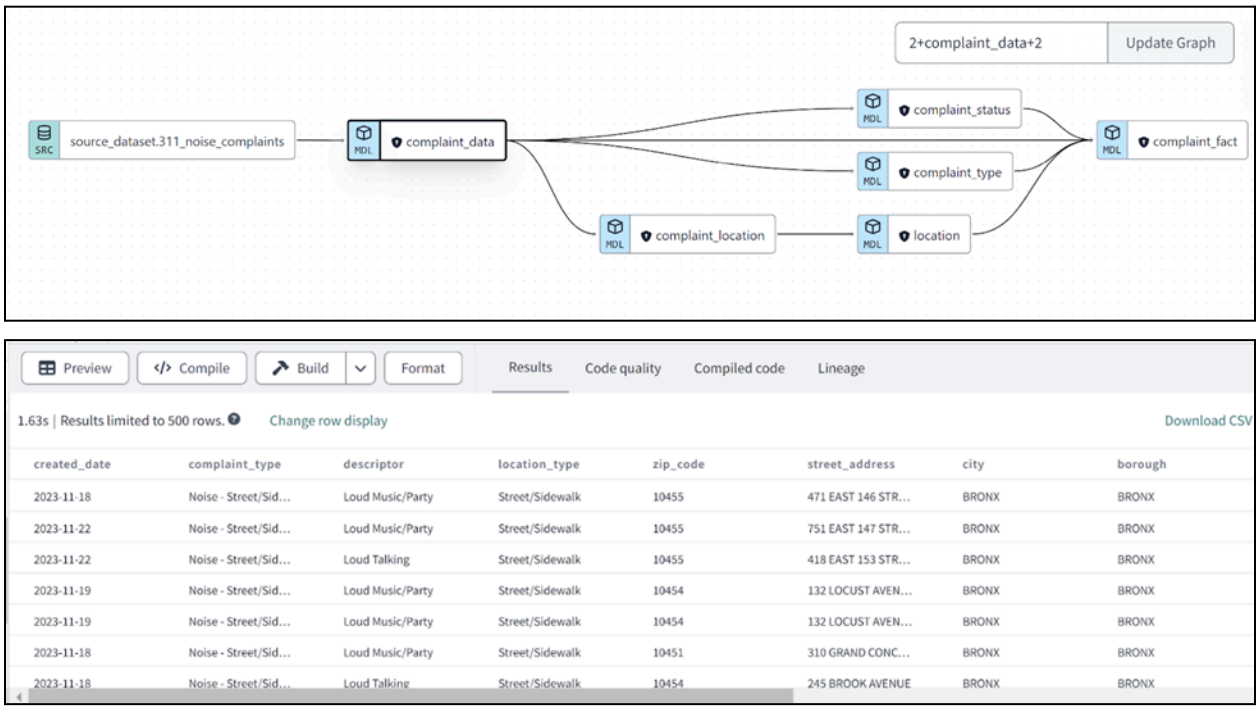
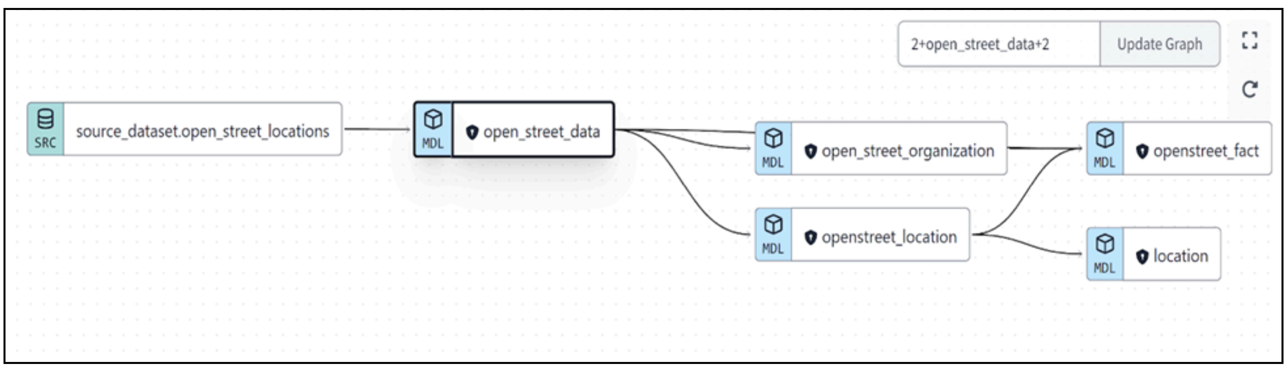


Figure 21 Complaint SQL

2. open_street_data.sql

```
with
   all_openstreet as (select * from {{ ref("open_street_data") }}),
   org as (select * from {{ ref("open_street_organization") }}),
   location as (select * from {{ ref("openstreet_location") }}),
   dates as (select * from {{ ref("date_dimension") }}),
   join_tbl as (
      select
         org.openstreet_organization_dim_id,
         location.openstreet_location_dim_id,
         dates.date_dim_id as date_id
      from all_openstreet
      left join org on all_openstreet.organization_name = org.organization_name
      left join
         location on all_openstreet.approved_on_street = location.approved_on_street
      left join dates on all_openstreet.start_date = dates.full_date
   )
select *
from all_openstreet
```



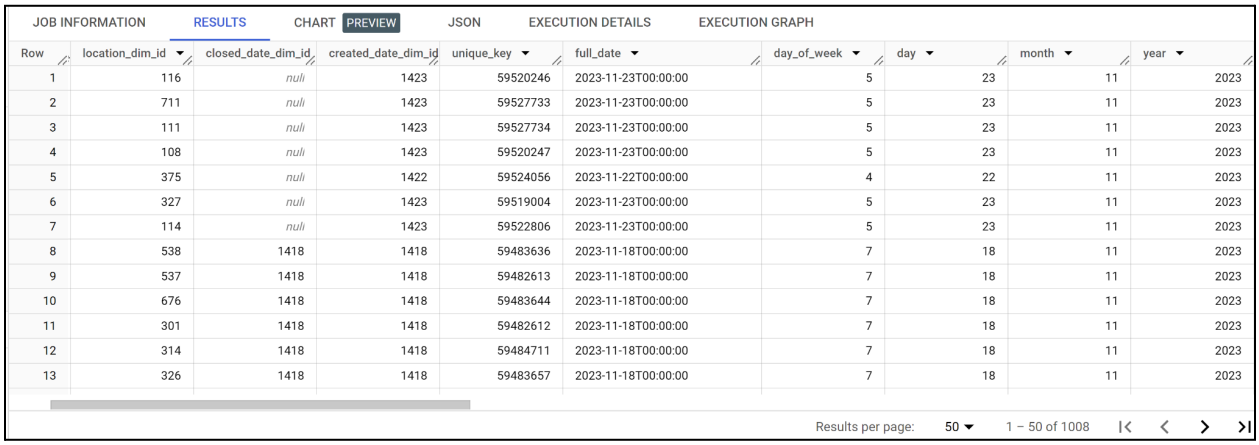
start_date	end_date	organization_name	approved_on_street	Approved_From_Str...	Approved_To_Street	borough
2023-07-28	2023-08-04	Merchants Of third Av...	3 AVENUE	83 STREET	MARINE AVENUE	Brooklyn
2023-07-28	2023-08-04	Merchants Of third Av...	3 AVENUE	83 STREET	MARINE AVENUE	Brooklyn
2023-07-28	2023-08-04	Merchants Of third Av...	3 AVENUE	83 STREET	MARINE AVENUE	Brooklyn
2023-07-28	2023-08-04	Merchants Of third Av...	3 AVENUE	83 STREET	MARINE AVENUE	Brooklyn
2023-07-28	2023-08-04	Merchants Of third Av...	3 AVENUE	83 STREET	MARINE AVENUE	Brooklyn
2023-07-28	2023-08-04	Merchants Of third Av...	3 AVENUE	83 STREET	MARINE AVENUE	Brooklyn
2023-07-28	2023-08-04	Merchants Of third Av...	3 AVENUE	83 STREET	MARINE AVENUE	Brooklyn

_name> Defer to production Ready

Figure 22 Open Street SQL

Dimensional Schema and Business Analytics Tools

Once the ELT process is complete, we were able to utilize our final dimensional schema to perform our analysis. In order to access our dimensional scheme we must query this data after the ELT process and join all of the relevant tables through the conformed dimensions in a data warehouse. In our project our conformed dimensions were the date and location dimensions.



JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	location_dim_id	closed_date_dim_id	created_date_dim_id	unique_key	full_date	day_of_week	day	month	year
1	116	null	1423	59520246	2023-11-23T00:00:00	5	23	11	2023
2	711	null	1423	59527733	2023-11-23T00:00:00	5	23	11	2023
3	111	null	1423	59527734	2023-11-23T00:00:00	5	23	11	2023
4	108	null	1423	59520247	2023-11-23T00:00:00	5	23	11	2023
5	375	null	1422	59524056	2023-11-22T00:00:00	4	22	11	2023
6	327	null	1423	59519004	2023-11-23T00:00:00	5	23	11	2023
7	114	null	1423	59522806	2023-11-23T00:00:00	5	23	11	2023
8	538	1418	1418	59483636	2023-11-18T00:00:00	7	18	11	2023
9	537	1418	1418	59482613	2023-11-18T00:00:00	7	18	11	2023
10	676	1418	1418	59483644	2023-11-18T00:00:00	7	18	11	2023
11	301	1418	1418	59482612	2023-11-18T00:00:00	7	18	11	2023
12	314	1418	1418	59484711	2023-11-18T00:00:00	7	18	11	2023
13	326	1418	1418	59483657	2023-11-18T00:00:00	7	18	11	2023

Results per page: 50 | 1 – 50 of 1008

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	openstreet_startdate_dim_id	openstreet_location_dim_id	openstreet_organization_dim_id	Organization_Name	approved_status	approved_on_stre	Approved_From	Approved_To_S	o_borough
1	1175	1	1	Enlightenment Wines	approvedFull	RANDOLPH S...	SCOTT AVE...	GARDNER ...	Brooklyn
2	1112	17	16	KioRestaurant, LLC	approvedFull	DUANE STRE...	WEST BRO...	HUDSON S...	Manhattan
3	1104	25	22	Church of the Heave...	approvedFull	EAST 90 STR...	5 AVENUE	MADISON ...	Manhattan
4	1137	27	24	The 700 Jefferson A...	approvedLimit...	JEFFERSON ...	MALCOLM ...	PATCHEN ...	Brooklyn
5	1205	36	28	LaGuardia Communi...	approvedFull	29 STREET	SKILLMAN ...	47 AVENUE	Queens
6	1303	52	41	Salsa Stories	approvedFull	BECK STREET	LEGGETT A...	AVENUE S...	Bronx
7	1321	78	41	Salsa Stories	approvedFull	SUFFOLK ST...	RIVINGTON...	DELANCEY...	Manhattan
8	1281	53	42	Harlem Mothers SA...	approvedFull	WEST 128 ST...	FREDERICK...	ST NICHOL...	Manhattan
9	1273	75	61	Incredible Credible ...	approvedFull	AMBOY STRE...	PITKIN AVE...	SUTTER A...	Brooklyn
10	1284	76	62	Boriken Neighborhood Health Center/ East Harlem	approvedFull	EAST 123 ST...	3 AVENUE	2 AVENUE	Manhattan

Results per page: 50 | 1 – 50 of 588

Figure 23 Final dimensional schemas for 311 noise complaint (top) and open street locations datamart (bottom)

KPI Visualizations

1. The declared grain of our data warehouse was accumulating. By using our data warehouse, the business users can easily track each event of noise complaint that was created until its resolution. Furthermore, this user can track the start and opening of open street locations until closure on the end date.
2. As a part of our KPI visualization, we measured the number of noise complaints as well as the open street locations running at the borough level. Manhattan has the most number of complaints reported, whereas Staten Island has the least number of reported noise complaints. The borough with the highest number of open street locations is Brooklyn and Bronx has the least of these locations. Complaints could also be traced at a more granular level of zip code as visualized the map box below:

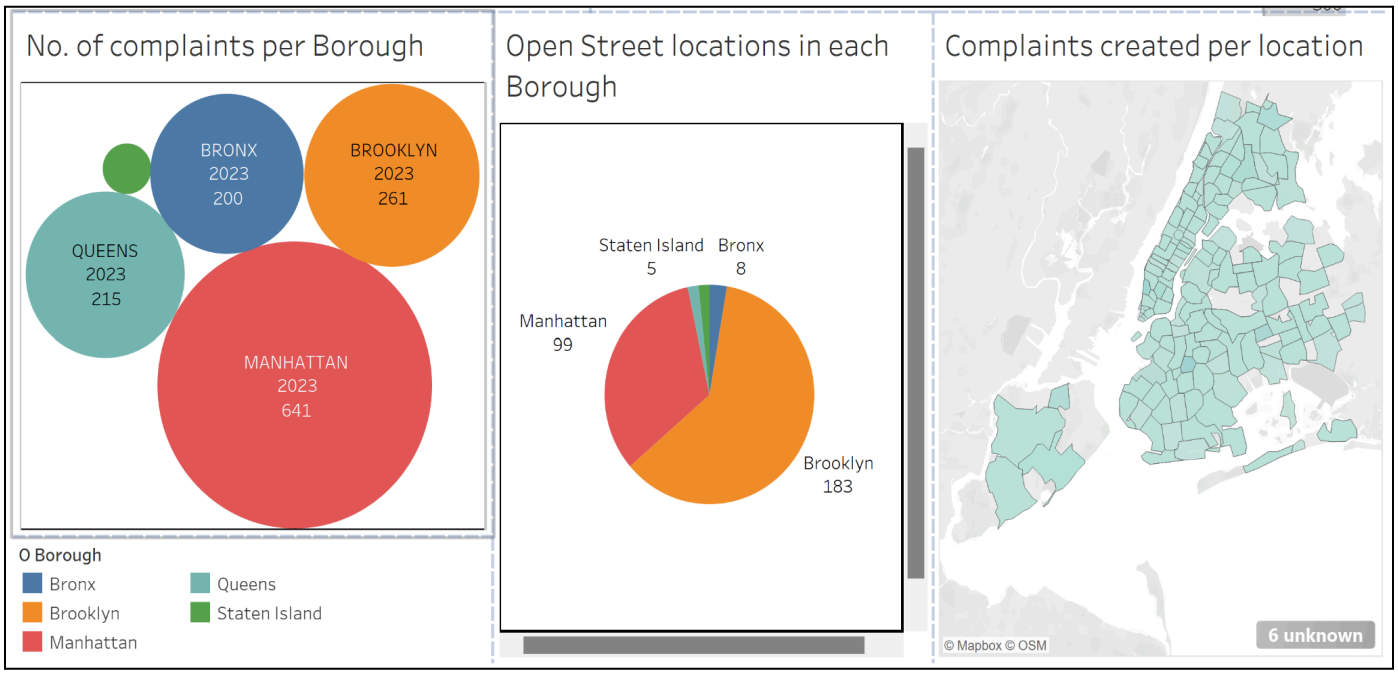


Figure 24 Three visualizations of our KPI's

3. Another KPI of measure of complaints on day of week can also be seen visualized below. Significantly more complaints were reported on Sunday compared to any other day of the week. Similarly, from the last visualization chart of the dashboard, we measured open_street location opening on different months of the year in each borough. Brooklyn has the most number of open street locations and summer months from May to August have the highest number of open street locations.

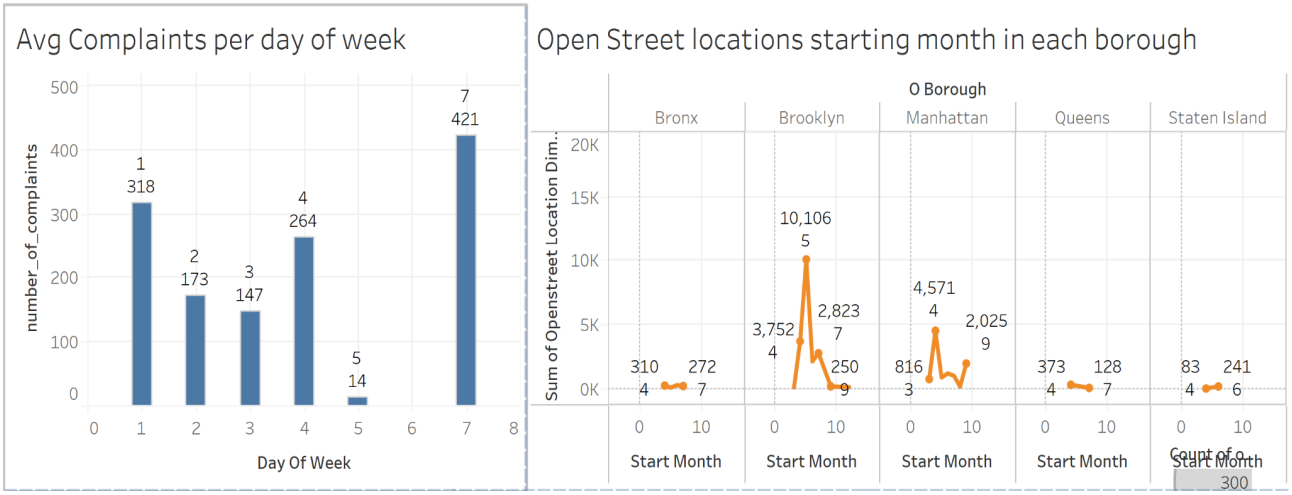


Figure 25 Two additional visualizations

Conclusions

Software and Tools Used

We utilized different technologies to extract the datasets that we used for this project as we encountered different issues or found better ways to resolve problems as we progressed through this project. Initially, we utilized Socrata Open Data API to extract the data from the database and determined that Google BigQuery would be a good tool to host the datasets. Afterwards we interfaced Google BigQuery and DBT which is where we did our transformations. Finally, we utilized Tableau to create all the relevant visualizations in our report.

-Socrata Open Data API
We used the Socrata Open Data api to extract the data from the website.
-Google BigQuery
Utilized to host the database.
-DBT
Utilized to transform the data that was located in our Google BigQuery and build fact tables.
-Tableau
Utilized to visualize the data for our analysis and draw our conclusions.

Proposed Benefits

After having collected all of the information and making our visualizations we have made this final conclusion on the proposed benefits.

We can't confidently confirm that there is a strong correlation between open street locations and noise complaints as the most complaints are from Manhattan and Brooklyn has the most open street reservations. Additionally, Bronx, Brooklyn, and Queens have the same number of complaints but Brooklyn has ten times the sum of Bronx, Brooklyn, and Queens combined.

That being said, the situation in which we mentioned in our hypothesis of there being little to no correlation between the amount of open street locations requested and the amount of noise complaints. Thus, for future reference one should feel confident that they will not risk being the cause of a noise complaint because they wish to host some sort of community event while utilizing an open street. Furthermore, this should demonstrate that there is a potentially large amount of noise issues within Manhattan however the volume of complaints despite low open streets in areas like the Bronx would suggest there are other issues the city should address as well.

Challenges in Execution

Throughout the project we had many technical and logistical challenges we had to overcome in order to synthesize our final report. In regards to logistical challenges we had trouble organizing meetings because all of us had differing schedules and although we were all working on the same project we all had different tasks and responsibilities outside of the scope of this project. Typically, in the other deliverables it wasn't too difficult to circumvent this issue asynchronously but when we did the implementation for our ETL this became a very big challenge. Aside from logistical issues, this portion of the project was very difficult and typically would require all the users working on the same database but given financial and technical constraints we all had to work in parallel on the same "system". Furthermore, the implementation of the ETL brought to light many other inconsiderations or improvements we could've made to our KPI's, data models, prospective fact tables, and more which we had to address over a shorter period of time. This section of the course also brought to light some gaps in knowledge both in a technical and theoretical capacity that we were able to resolve after consulting external resources like class notes, our professor, and other internet resources.

On the other hand, the easiest part was definitely every prerequisite deliverable given that we thought we did everything correctly. For those milestones, it wasn't as difficult because much of the content could be done asynchronously and relied much less on what each of us made as individuals and didn't require as much collaboration other than ideas without a way to check if it "worked" as we weren't implementing it.

Having to utilize DBT, SQL, and Google BigQuery to the level that we ended up utilizing for this project was not something all of us were expecting to have to undertake. Thankfully, each of our different members had a wide variety of skills and the will to learn and collaborate to reach our final solution.

In regards to what we all would've done differently they are mostly technically and logistically related. First, we didn't have a way to unify all of the work that we did as we didn't pay to have multiple users on the same project for certain elements like Google BigQuery and DBT which made it very challenging to have consistent implementation. Secondly, more meetings would definitely be better but as we are all college students with different schedules and responsibilities it would be challenging to match this to a real world implementation of a project like this where every member is working on the same thing and it is their primary responsibility. Finally, as lightly touched on before having the project be our sole and primary responsibility would make execution dramatically simpler both in technical implementation and theoretical implementation as we would be able to focus more.

References

https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9 -
-Requests/Noise Complaint Data Source

https://data.cityofnewyork.us/Health/Open-Streets-Locations/uiay-nctu
-Open Street Location Data Source

https://courses.getdbt.com/courses/fundamentals
-Information on using DBT

Meeting Log for Milestone 6

Date/Time Meeting Started and Ended
Start Time: 12/13/23 at 4:00 PM
End Time: 12/13/23 at 11:00 PM
Work Was Completed/Communicated Asynchronously Via Text/Email

Attendees

Edward Shin, Jerome Galam, Laney Zou, Tsering Lama, Denis Dudkin

Main Topic Discussed
Synthesis and compilation of data for conclusion in final report.

Roles and Responsibilities
It was everybody's responsibility to edit and revise the paper for submission.

Tasks Completed

1) Continued paper synchronously
2) Finished paper
3) Peer edited paper
4) Submit