## 

## **Analyzing the Effect of Open Street Location Program on Noise Complaints from 2020 to 2023.**

## 

## Final Report

12/13/2023

CIS 4400 Group 9

Tsering Lama | Edward Shin | Jerome Galam | Laney Zou | 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](mailto:denis.dudkin@baruchmail.cuny.edu)

# **Outline**

Page number

1. Introduction …………………………………………………………………………………………… 3 - 4

* Objective……………………………………………………………………………………. 3
* Data Sources and their Description ……………………………………….……. 4

1. Dimensional Model Diagrams …………………………………………………………………. 5 - 6
2. ETL Process …………………………………………………………………………………………… 7 - 21

* Extraction …………………………………………………………………………………. 7
* Loading …………………………………………………………………………………….. 8
* Transformation with DBT Models ………………………………………….…… 9 - 21

1. Dimensional Schema and Business Analytics (tools used) ………………….….….. 22
2. KPI Visualizations ……………………………………………………………………………..….. 23 – 24
3. Conclusion ……………………………………………………………………………………….….… 25
4. Lessons Learned ……………………………………………………………………………….…… 26
5. References ……………………………………………………………………………………….….... 27

# **Introduction**

## **Objective**

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 /number of times some specific area has received a noise complaint. We believed that when some area is reserved for festivities or public events, there will likely be a lot of noise and commotion resulting in some formal complaints, and that fewer complaints or no complaints would be reserved for areas with few 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 -](https://opendata.cityofnewyork.us/) and the 311 Complaint/Service Request is from the 311 and DOT 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 our potential hypothesis of there being a positive correlation between the number of noise complaints and open street locations we would hope to amend that issue moving forward. Ultimately, if particular locations have a very strong 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 investigate those problems and address them as they have some validity in severity.

## **Description**

For 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](https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9)

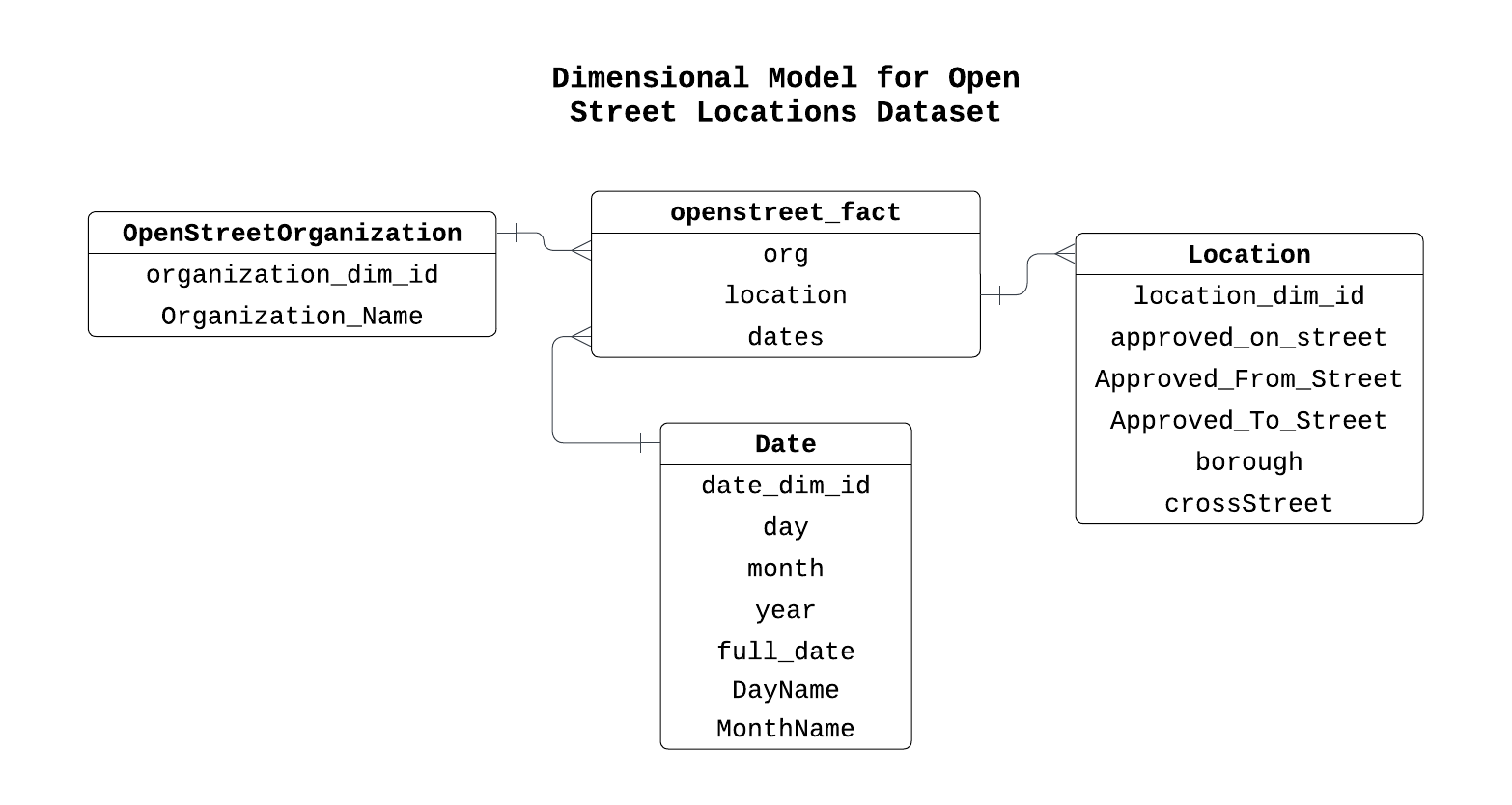
In this dataset, New York City residents call 311 for information about city services or report issues that they experience for the city to provide resources or resolve. This dataset contains information on reports made to 311 and includes attributes such as location, agency the report pertains to, time/date, type of report, street, and more. For 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](https://data.cityofnewyork.us/Health/Open-Streets-Locations/uiay-nctu)

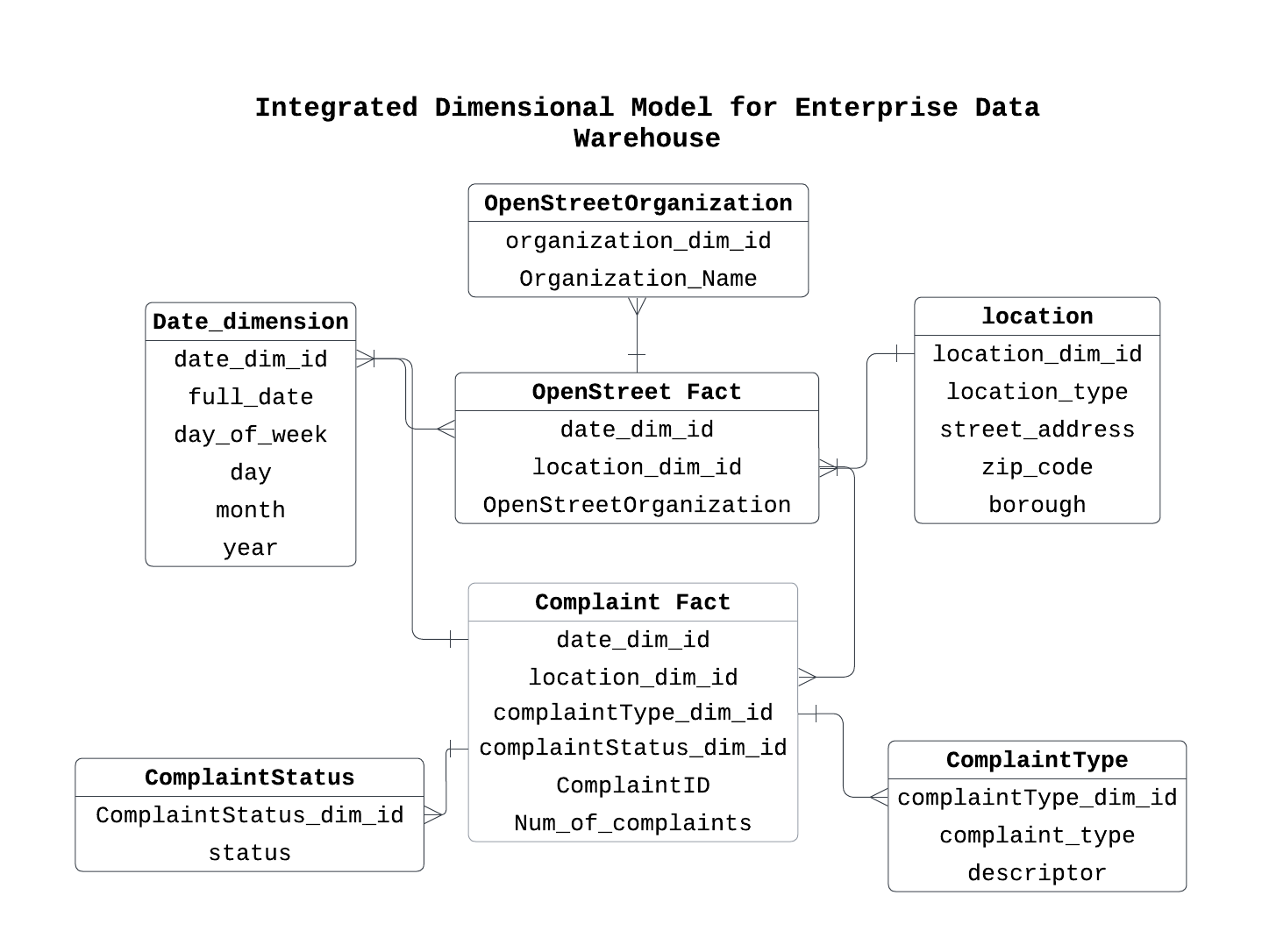
This dataset gives information about the New York City Open Street Program. This program limits traffic in designated areas (closes the street) for some designated and requested time. The purpose of these “open streets” is to close street(s) to provide a safer space for city programs, community events, 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 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**

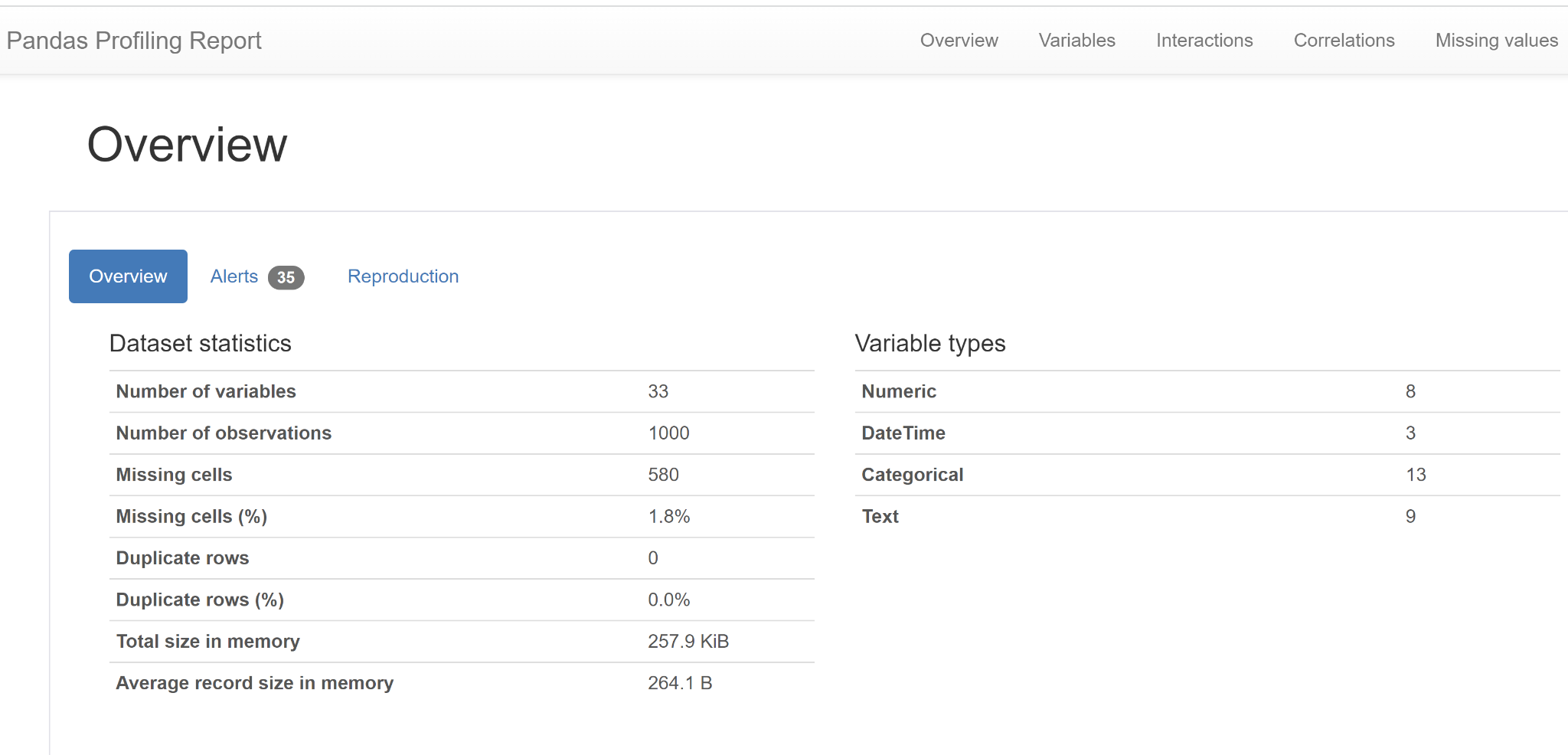
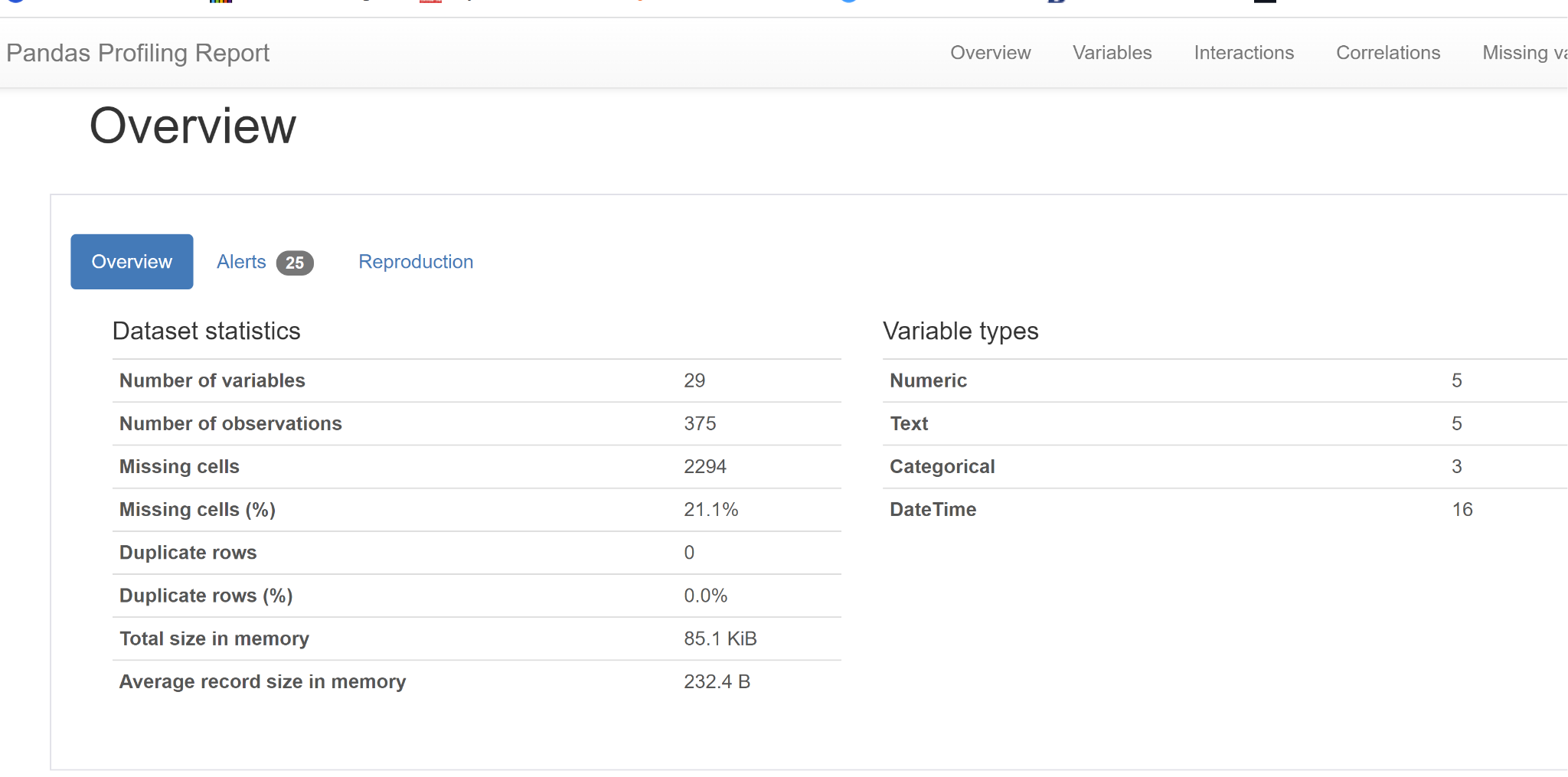
Before diving into the Extraction, we first performed data profiling to assess the quality of both datasets.

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**

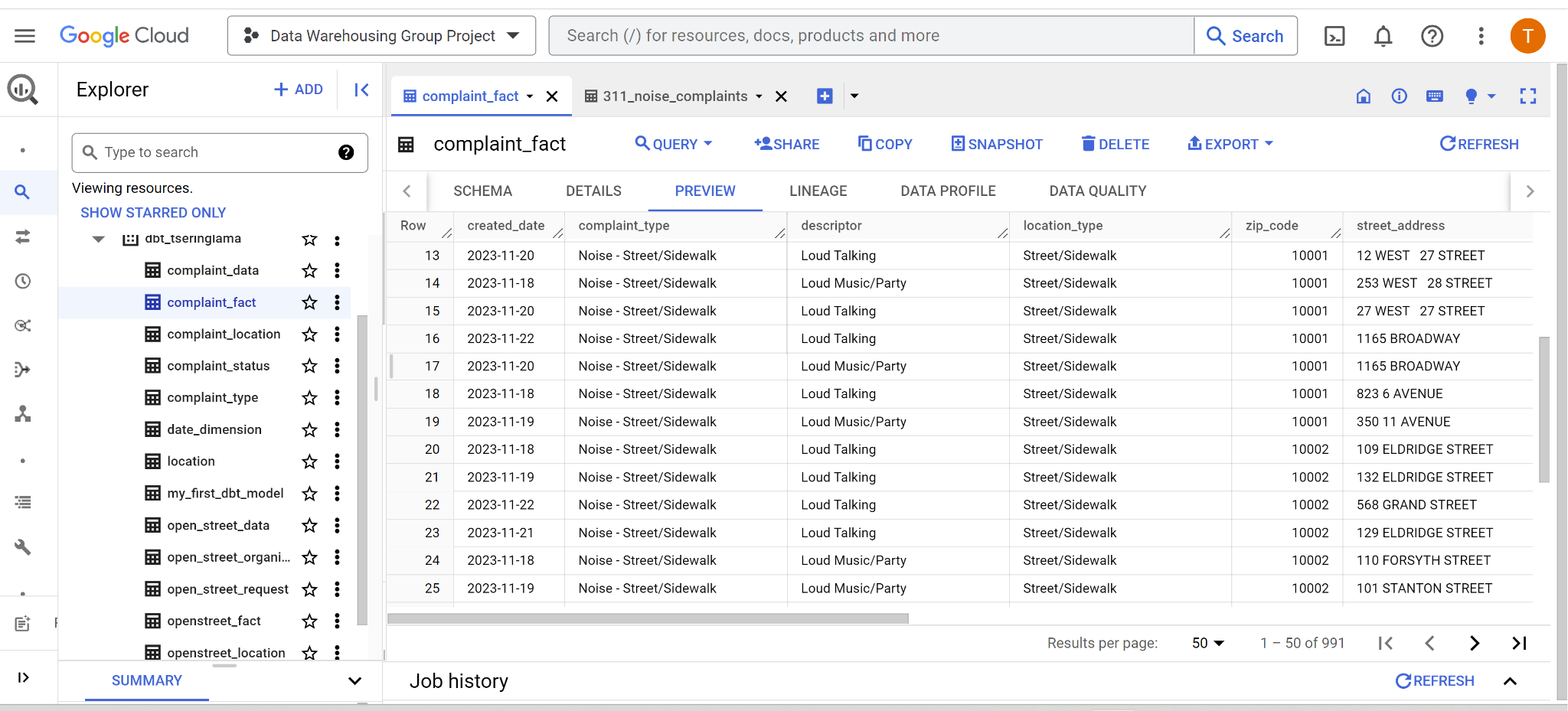
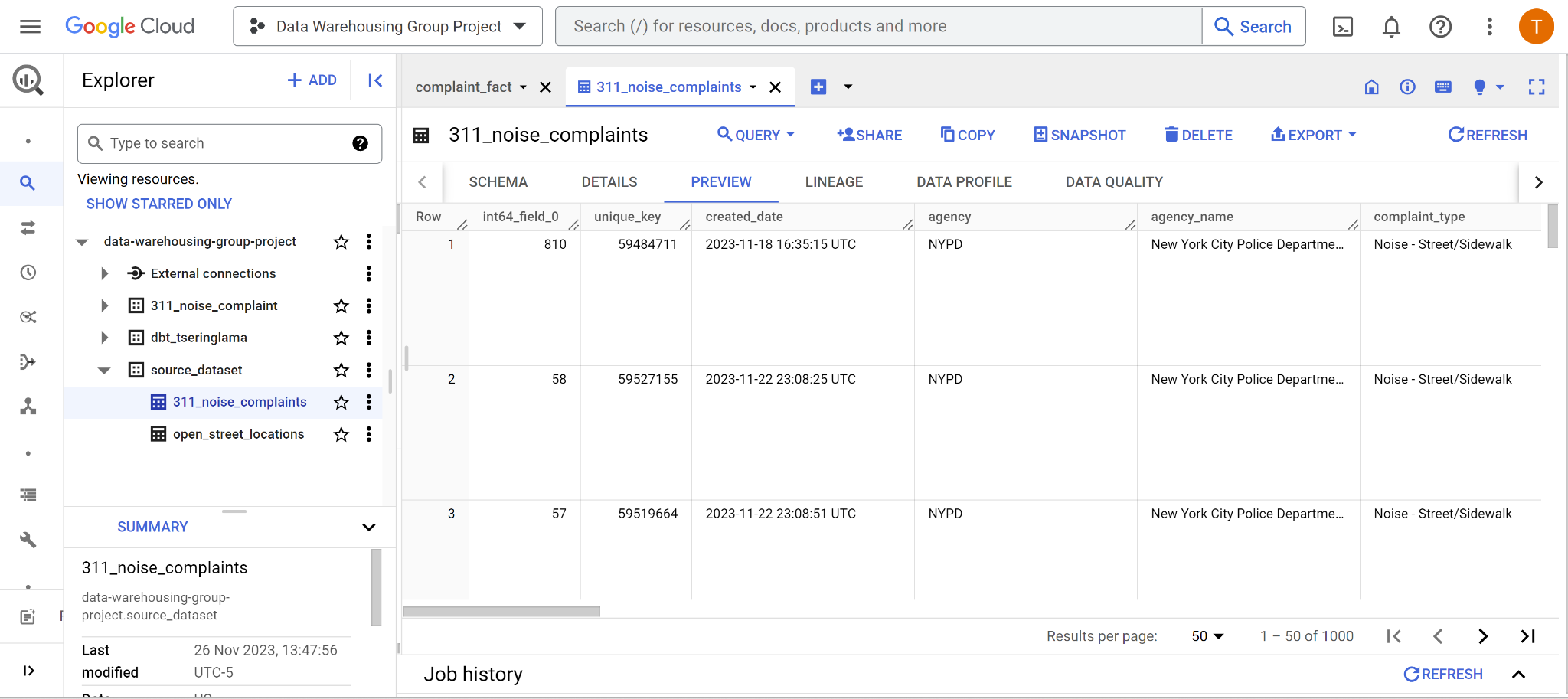
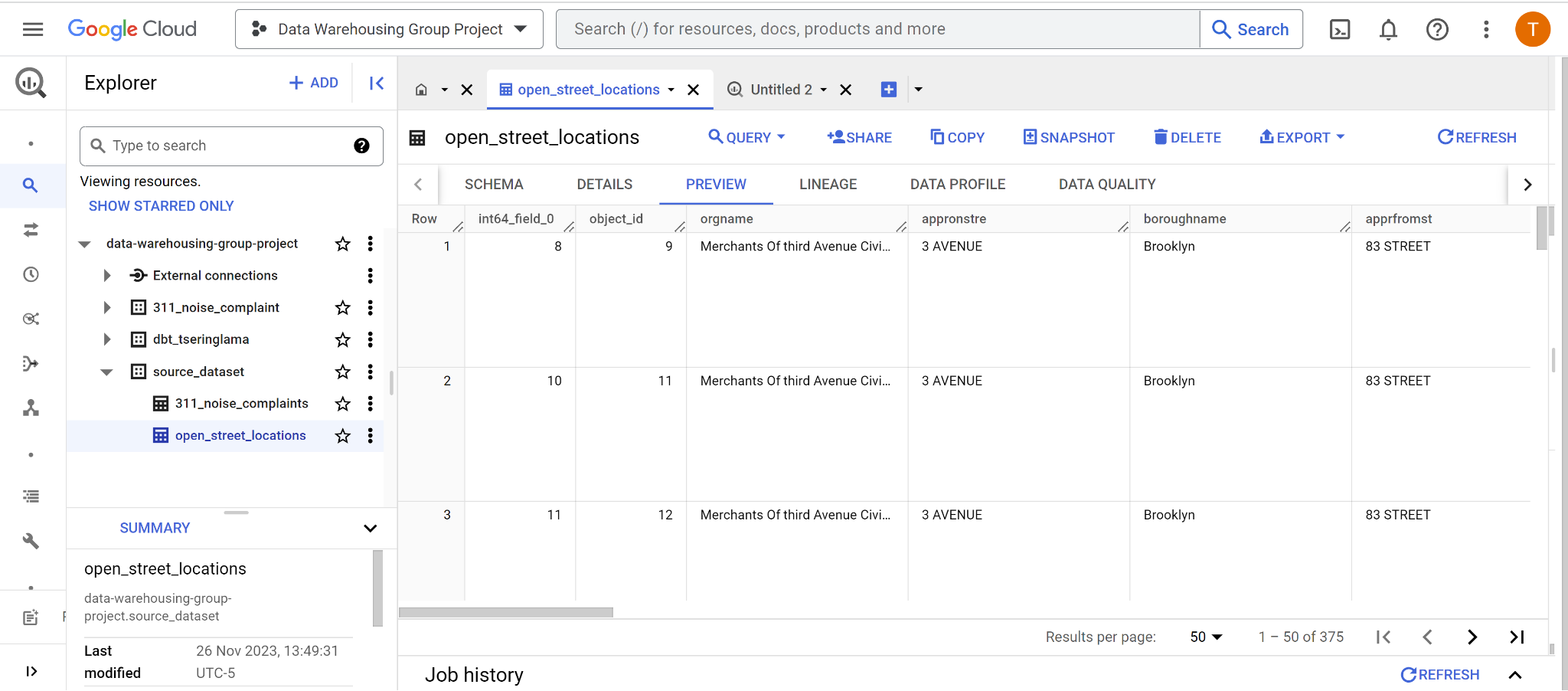
As part of the loading process, the extracted data was then loaded into Google Big Query, which was connected to DBT for transformation processes.

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**

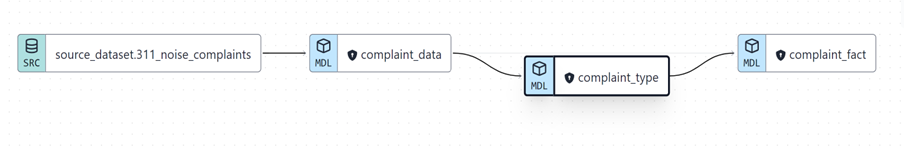
1. complaint \_type:

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,

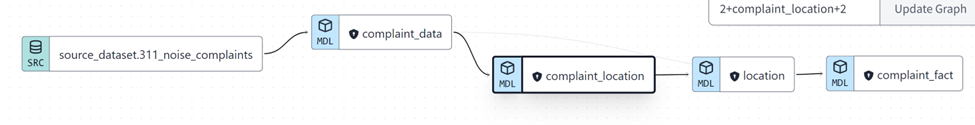
city, zip\_code, borough, cast(latitude as STRING) as latitude,

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

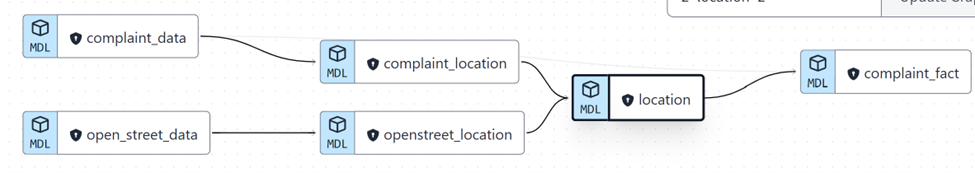
full join openstreet\_loc

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,

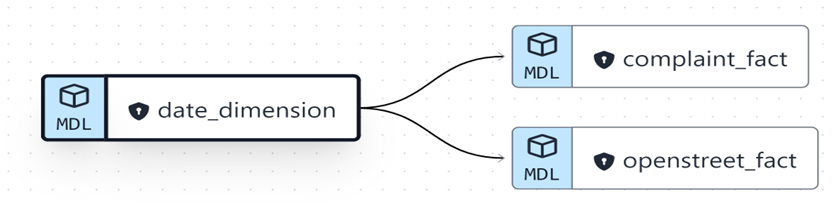
date\_day as full\_date,

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,

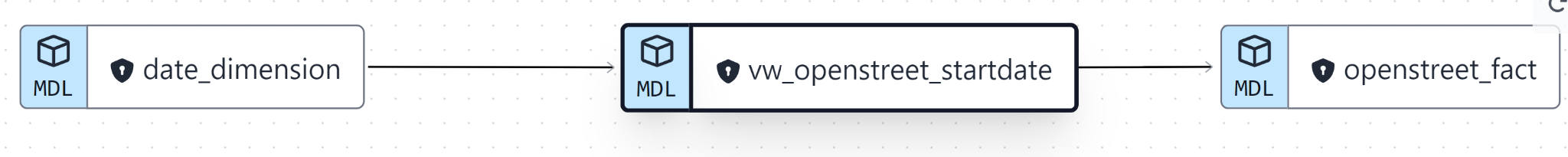
day as start\_day,

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)

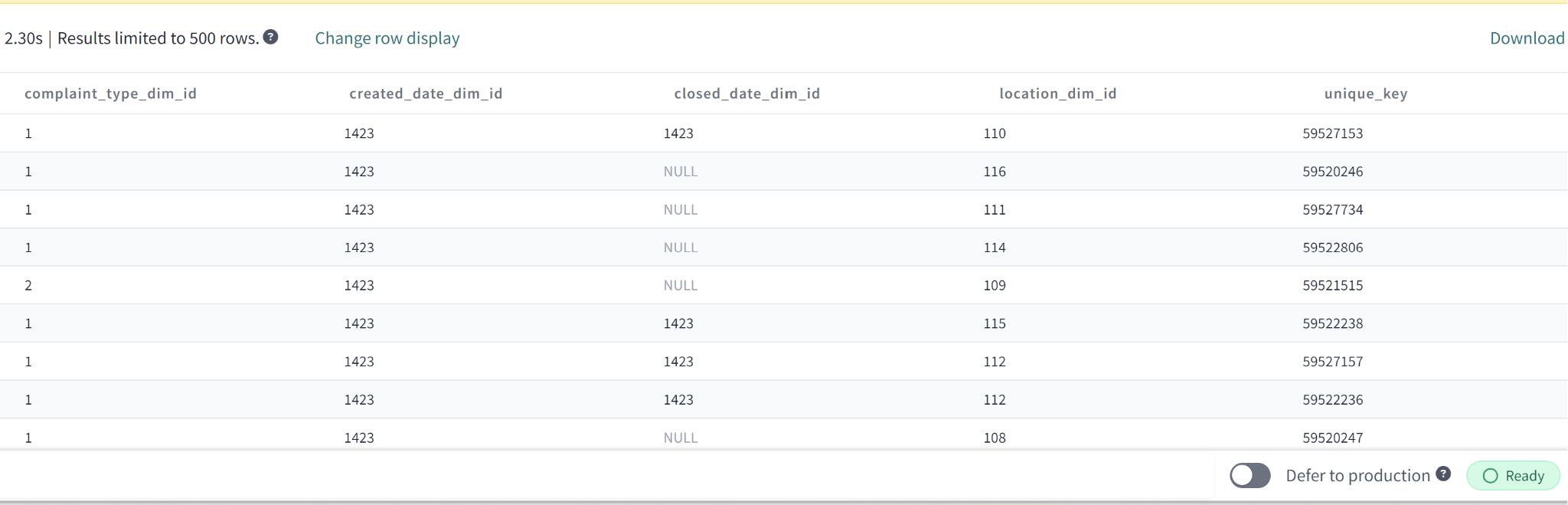
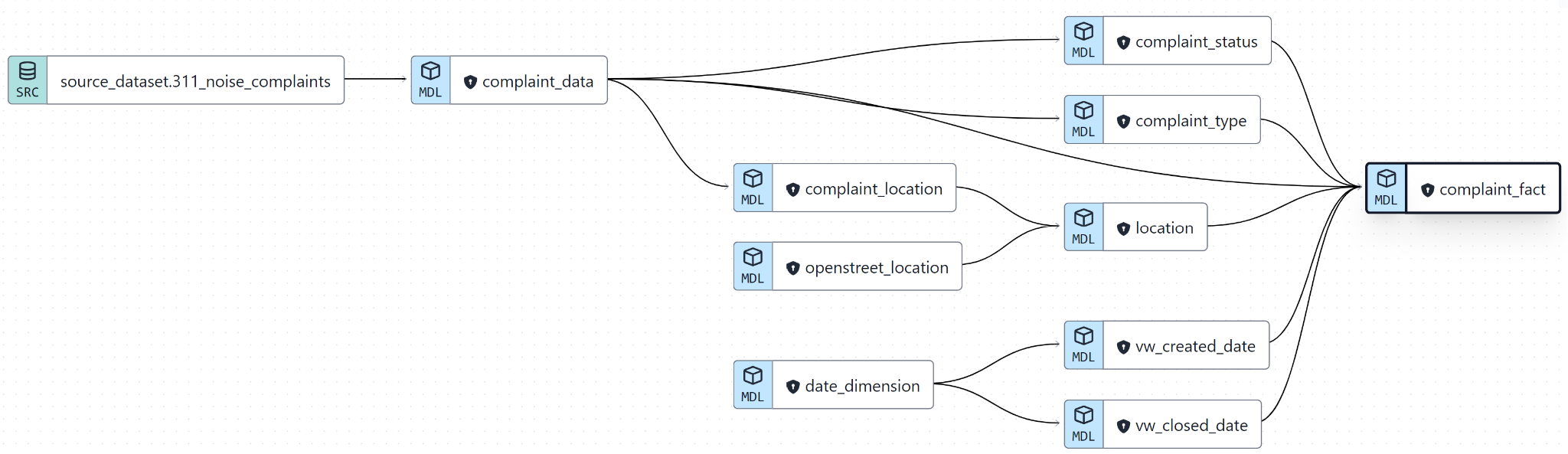
left join location on (all\_complaints.latitude=location.latitude AND all\_complaints.longitude=location.longitude)

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



*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

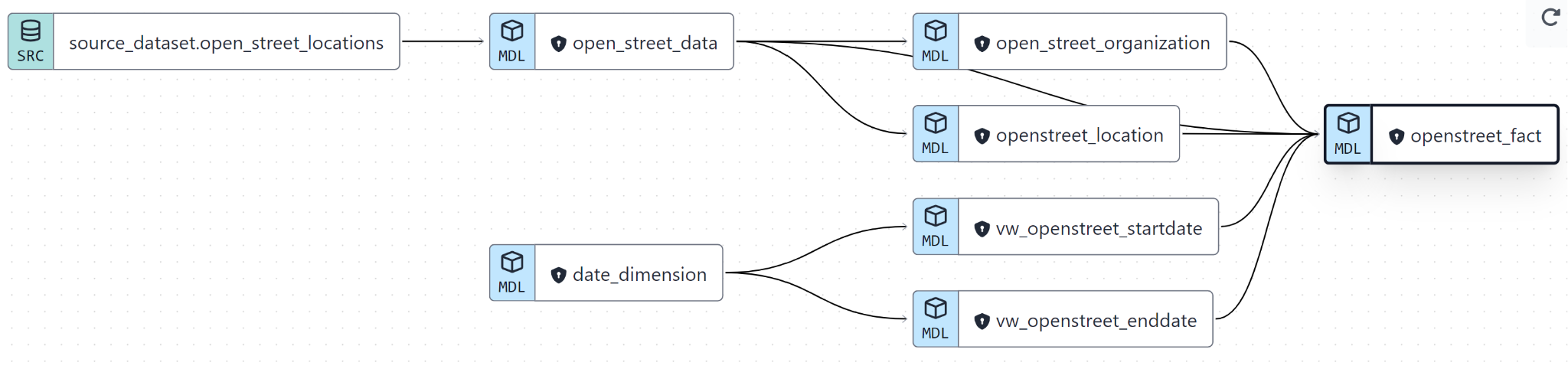
LEFT JOIN openstreet\_location ON all\_openstreet.approved\_on\_street = openstreet\_location.approved\_on\_street

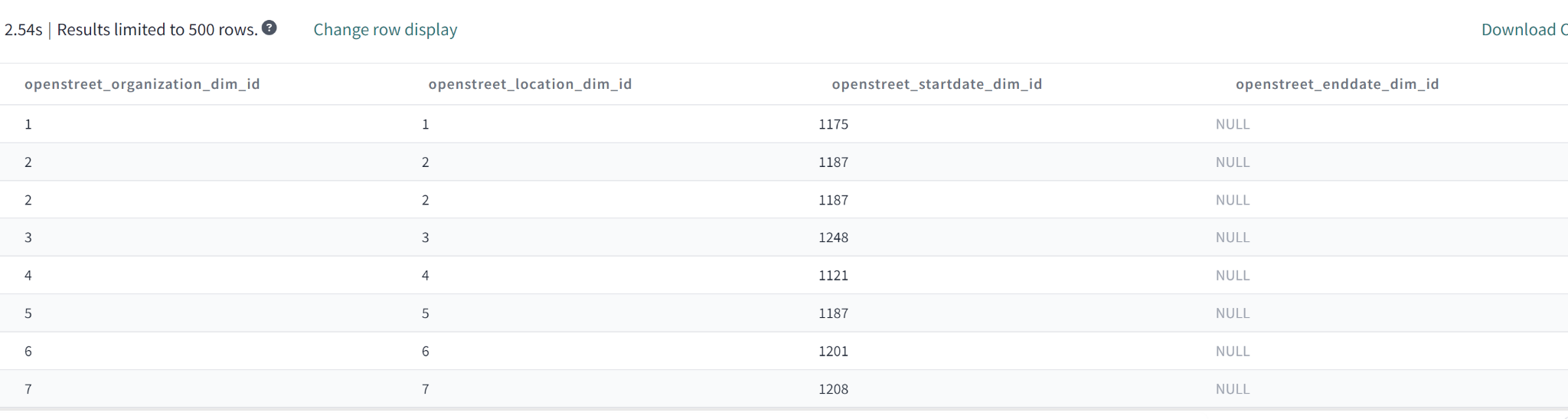
LEFT JOIN start\_date on (all\_openstreet.start\_date= start\_date.start\_fullDate)

LEFT JOIN end\_date on (all\_openstreet.end\_date = cast(end\_date.end\_fullDate as date))

)

SELECT \* FROM join\_tbl





*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,

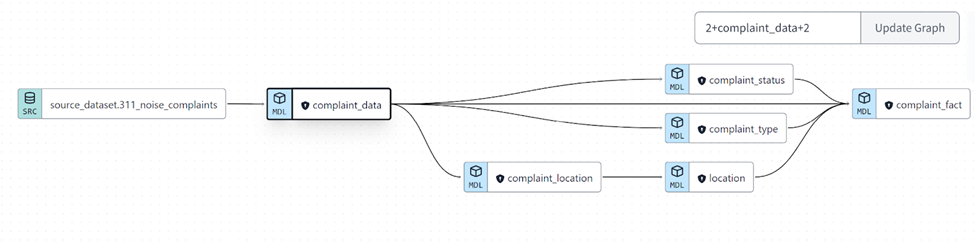
status

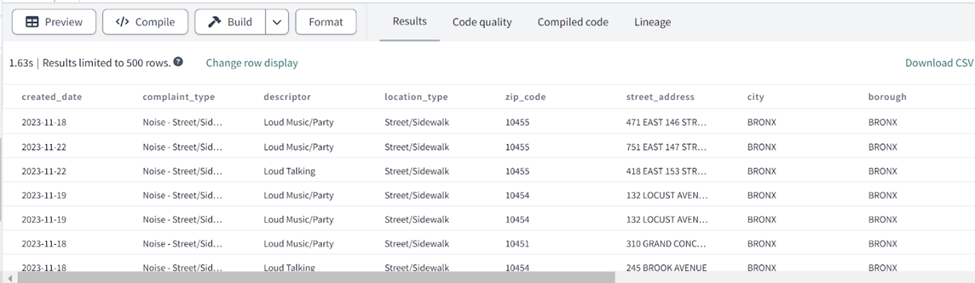
from {{source('source\_dataset','311\_noise\_complaints')}}

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

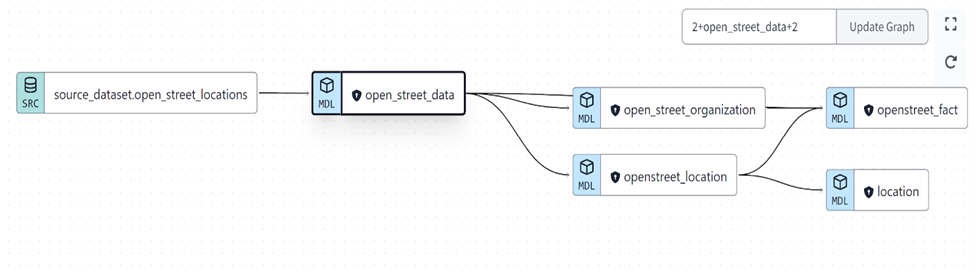
location on all\_openstreet.approved\_on\_street = location.approved\_on\_street

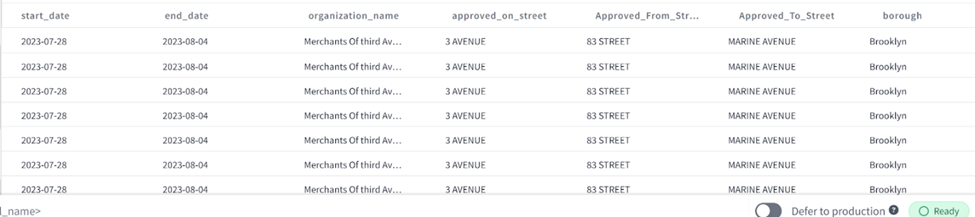
left join dates on all\_openstreet.start\_date = dates.full\_date

)

select \*

from all\_openstreet



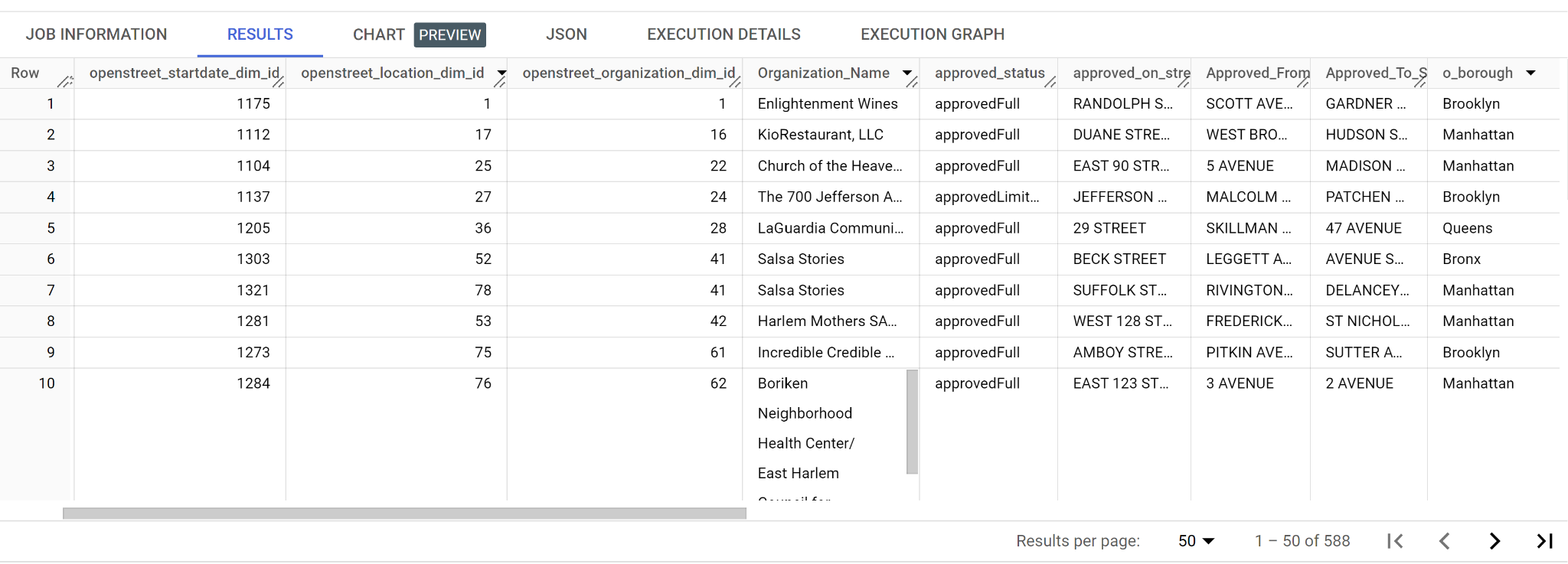


*Figure 22 Open Street SQL*

# **Dimensional Schema and Business Analytics**

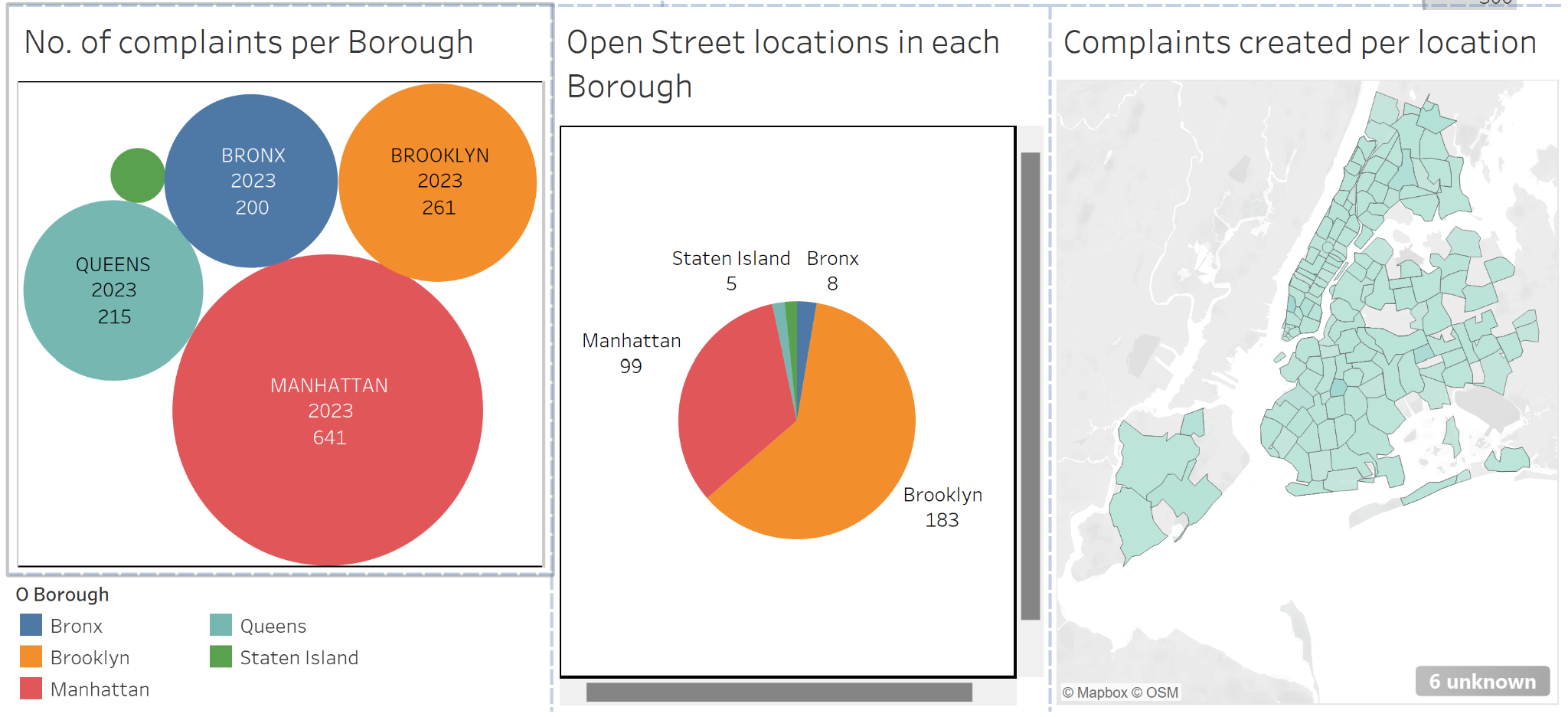
Once the ELT process was complete, we were able to utilize our final dimensional schema to perform our analysis. To access our dimensional scheme, we must query this data after the ELT process and join all of the relevant tables through the conformed dimensions in a data warehouse. In our project, our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



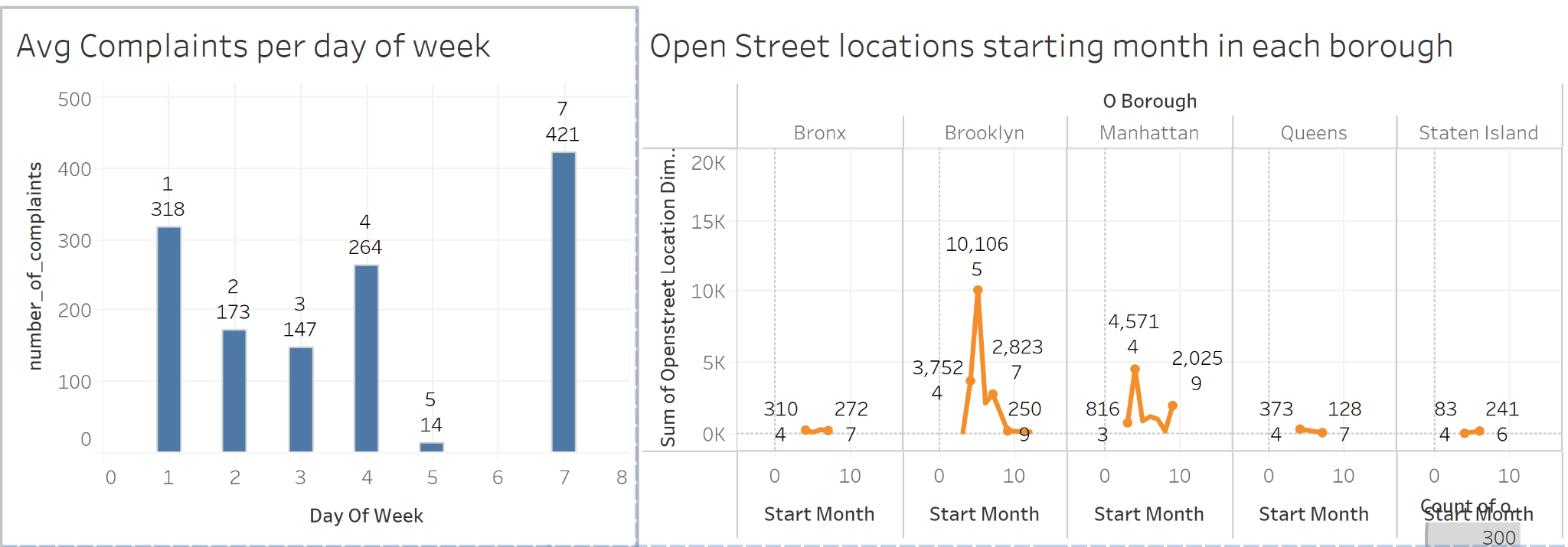


*Figure 23 Final dimensional schemas for 311 noise complaint (top) and open street locations datamart (bottom)*

# **KPI Visualizations**

1. The declared grain of our data warehouse was accumulating. By using our data warehouse, 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 the Bronx has the least of these locations. Complaints could also be traced at a more granular level of zip code as visualized in the map box below:

*Figure 24 Three visualizations of our KPI’s*

1. Another KPI of measure of complaints on the day of the 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. After that, we integrated 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 the information and making our visualizations we have made this conclusion on the proposed benefits.  
  
We can’t confidently confirm that there is a strong correlation between open street locations and noise complaints as 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 number of open street locations requested and the number 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.

**Lessons Learned**

Throughout the project, we had many technical and logistical challenges we had to overcome to synthesize our final report. Concerning 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 implemented our ELT 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 considerations or improvements we could’ve made to our KPIs, 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 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, 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.

Regarding what we all would’ve done differently, they are mostly technically and logistics-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 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