

## Group 1: Project Proposal

# “311 Illegal Parking Complaint Type and Parking Violations Issued For Fiscal Years 2017-2021”

**Baruch College (CUNY)**  
**CIS 4400 CMWA: Data Warehousing for Analytics**  
**Professor Richard Holowczak**

### **Member's Email Address:**

Danny Huang ([danny.huang1@baruchmail.cuny.edu](mailto:danny.huang1@baruchmail.cuny.edu))

Ayra Akhter ([ayra.akhter@baruchmail.cuny.edu](mailto:ayra.akhter@baruchmail.cuny.edu))

Eric Myagkov ([eric.myagkov@baruchmail.cuny.edu](mailto:eric.myagkov@baruchmail.cuny.edu))

Mariya Tabachnikova ([mariya.tabachnikova@baruchmail.cuny.edu](mailto:mariya.tabachnikova@baruchmail.cuny.edu))

Siming Deng ([siming.deng@baruchmail.cuny.edu](mailto:siming.deng@baruchmail.cuny.edu))

## Type of 311 Complaint: Illegal Parking

### Business Problem:

- Can violation tickets get more expensive after a number of violations?
- Do areas with a higher amount of 311 complaints also exhibit greater amounts of parking and camera violations?
- How many illegal parking complaints are related to “Double Parking”? Is there a larger, city-wide issue with double parking?
- Does a specific zip code have the defective infrastructure (roads, intersections, etc) that could cause drivers to be more prone to the aforementioned violations?
- Could an area with a higher number of reported violations be in closer proximity to a police precinct? (More cops => more tickets)

### Synopsis:

Group 1 is observing historical and current data on parking/motor-vehicle violations in NYC to provide accessible information to drivers and local government officials. Group 1 will use the data to better understand the source of these violations, and if there are external factors that skew this data. Faulty infrastructure and elevated police presence are some examples of factors that can inflate the number of violations in a certain zip code. The integration of the 311 data and parking & camera violations can highlight certain issues, which would then be solved by the city.

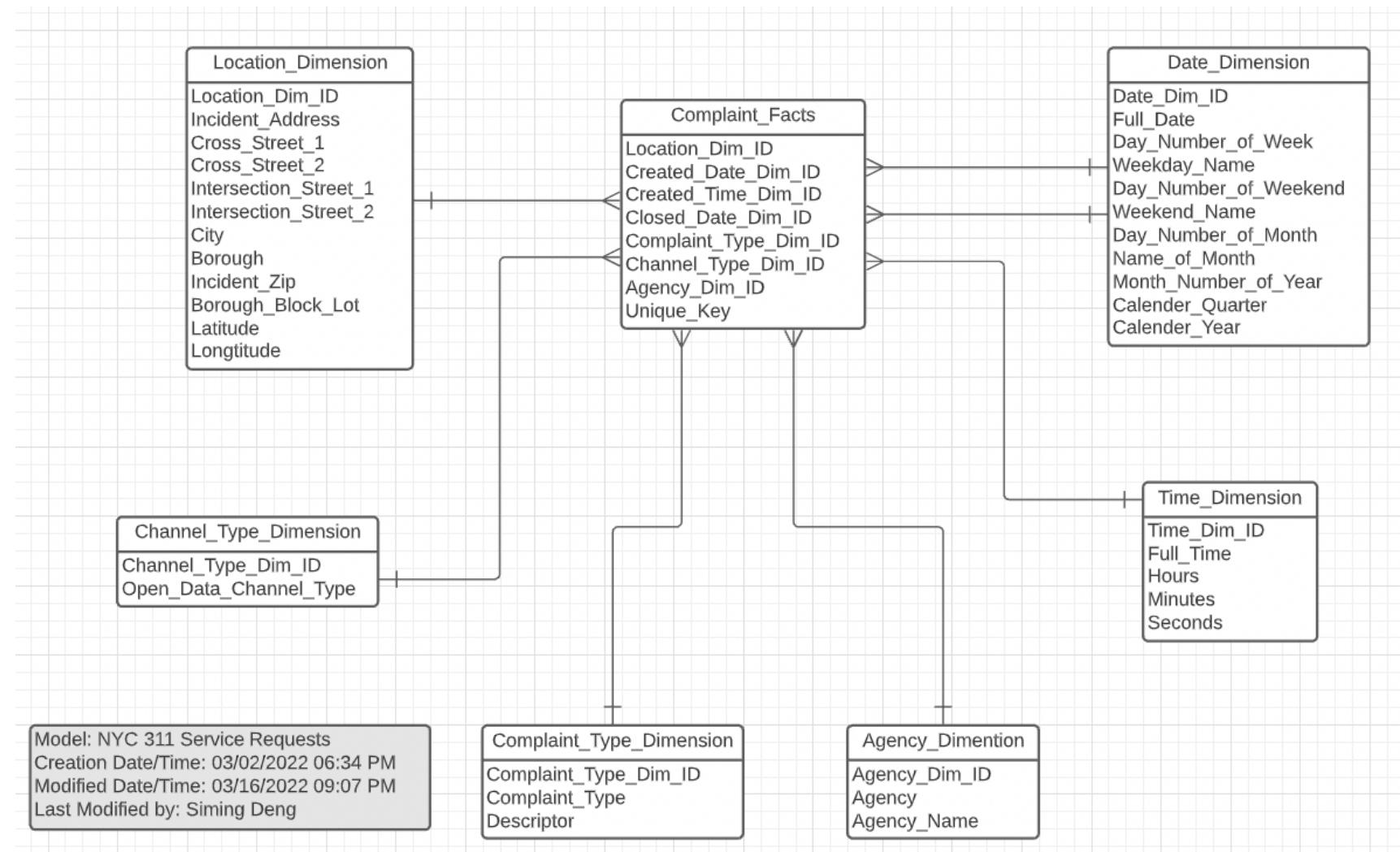
The first data set Group 1 will use comes from 311 complaints. The data includes the department, complaint type, incident zip code, and status. The second data set will be from Open Data, regarding parking and camera violations. This data set includes the violation type, fine in dollars, vehicle plate number, and issuing agency.

To find a solution, Group 1 will observe tables to track records of violations within specific zip codes and try to deduce if external factors relating to infrastructure skew the reported data.

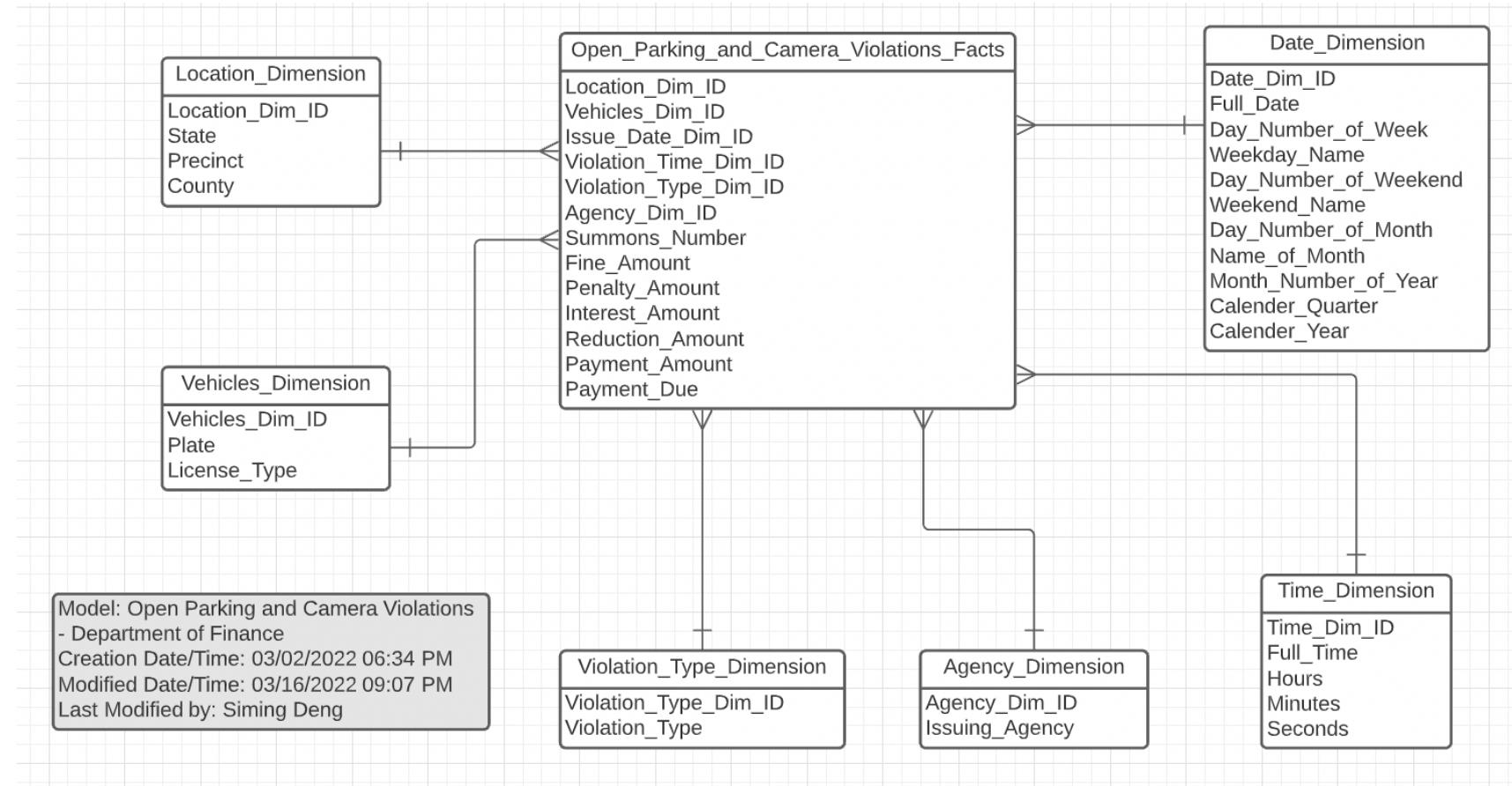
### **List of KPIs (including descriptive details)**

- Percentage increase in the number of issued parking violations in NY
  - Total number of illegal parking complaints each year since 2017 in NY
  - Number of illegal parking complaints per incident zip code in NY
  - Number of illegal parking complaints per Channel Type each year in NY
  - The total fine amount because of parking violation each year since 2017
  - Total penalty amount because of parking violation each year since 2017
  - Total interest amount because of parking violation each year since 2017
  - Total reduction amount because of parking violation each year since 2017
  - Total payment amount because of parking violation each year since 2017
  - Number of violation tickets issued per violation time each day and month
- 
1. Which agency gets the most complaints?
  2. Which month has the highest fine and penalty amount in NY?
  3. Which date is the busiest for the New York Police Department?

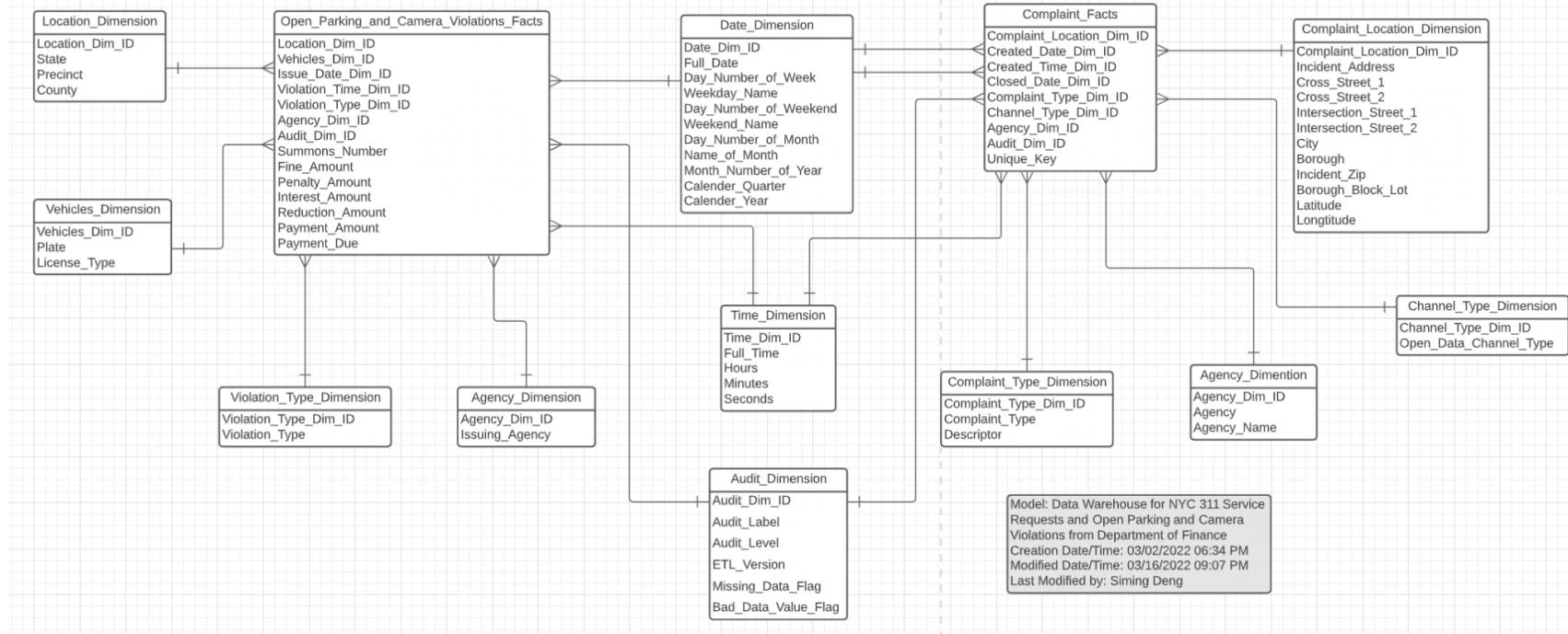
## 311 Service Dimensional Model



## Open Parking and Camera Violations Dimensional Model



## Data Warehouse Dimensional Model



## **Data Profiling**

### **311 Complaint Dataset:**

- We have a total of 967634 rows and 19 columns
- We have missing values for the following columns: intersection\_street\_1 (35.1%), intersection\_street\_2 (35.1%), bbl (16.4%), cross\_street\_2 (5.9%), cross\_stree\_1 (5.8%), incident\_address (5.3%), city (5.1%), latitude (1%), longitude (1%), and closed\_date (0.1%)
- We have high positive correlations between incident\_zip vs bbl, incident\_zip vs longitude, and longitude vs latitude
- We have high negative correlations between incident\_zip vs latitude, and bbl vs latitude

### **Target DBMS:** Google Big Query

We decided to use Google BigQuery for our target DBMS for multiple reasons. For one, we are all familiar with BigQuery because of the assignment we had. This means that we know that we are all able to contribute. Additionally, BigQuery can handle large datasets. For the 311 complaints dataset, we decided to use complaints starting from 2017 to present, which is over four years worth of data. BigQuery is able to handle such large datasets. It is also able to separate data based on the parameters specified. We want to look at certain complaints and zip codes, and with BigQuery we can compare whatever variables we want.

### **ETL Tool:** dbt

For our ETL tool, we chose dbt. This is because it seems similar to previous programming that was done, since it involves SQL. It will also make it easy to collaborate between group members because dbt can be connected to Git. This way everyone can access the most recent update.

## ETL Programming

311\_service\_requests.sql

```
SELECT unique_key,
       created_date,
       closed_date,
       agency,
       agency_name,
       complaint_type,
       descriptor,
       incident_address,
       cross_street_1,
       cross_street_2,
       intersection_street_1,
       intersection_street_2,
       city,
       borough,
       incident_zip,
       bbl,
       latitude,
       longitude,
       open_data_channel_type,
FROM `bigquery-public-data.new_york_311.311_service_requests`
WHERE complaint_type = 'Illegal Parking' AND FORMAT_DATE("%Y", created_date) IN ('2017', '2018', '2019', '2020', '2021')
```

Project.yml

```
Version: 2
Models:
  311_service_requests:
    +materialized = table
```

Tests.yml

```
version: 2

models:
  - name: 311_service_requests
    columns:
      - name: unique_key
        tests:
          - unique
          - not_null
      - name: status
        tests:
          - name: status
            tests:
              - accepted_values:
                  values:
                    - completed
```

```
dim_complaint_type.sql
SELECT
    row_number() OVER () AS Complaint_Type_Dim_ID,
    complaint_type, descriptor
FROM
    (SELECT DISTINCT complaint_type, descriptor
     FROM `fluted-quasar-341922.311_complaint_requests.311_illegal_parking_2017_to_2021`  
)
```

```
dim_agency_311.sql
SELECT
    row_number() OVER () AS Agency_311_Dim_ID,
    agency, agency_name
FROM
    (SELECT DISTINCT agency, agency_name
     FROM `fluted-quasar-341922.311_complaint_requests.311_illegal_parking_2017_to_2021`  
)
```

```
dim_channel_type.sql
SELECT
    row_number() OVER () AS Channel_Type_Dim_ID,
    open_data_channel_type
FROM
    (SELECT DISTINCT open_data_channel_type
     FROM `fluted-quasar-341922.311_complaint_requests.311_illegal_parking_2017_to_2021`  
)
```

```
dim_complaint_location.sql
SELECT
    row_number() OVER () AS Compaint_Location_Dim_ID,
    incident_address,
    city,
    borough,
    incident_zip,
    borough_block_lot,
    latitude,
    longitude
FROM
    (SELECT DISTINCT incident_address,
        city,
        borough,
        incident_zip,
        bbl AS borough_block_lot,
        latitude,
        longitude
    FROM `fluted-quasar-341922.311_complaint_requests.311_illegal_parking_2017_to_2021` )
)
```

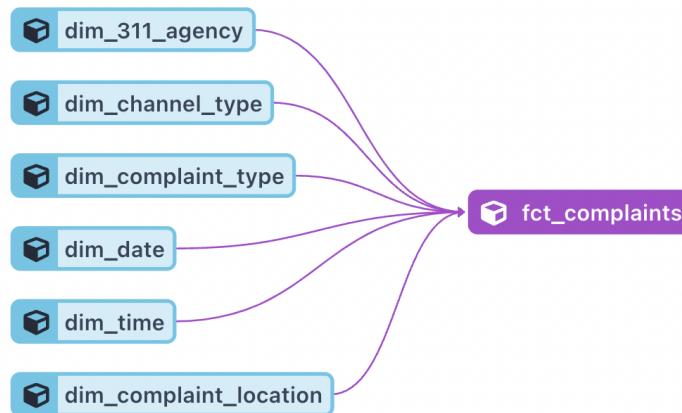
```
fct_complaints.sql
WITH dim_agency AS
(
    SELECT * FROM {{ ref('dim_311_agency') }}
),
dim_channel_type AS
(
    SELECT * FROM {{ ref('dim_channel_type') }}
),
dim_complaint_type AS
(
    SELECT * FROM {{ ref('dim_complaint_type') }}
),
dim_date AS
(
    SELECT * FROM {{ ref('dim_date') }}
),
dim_time AS
(
    SELECT * FROM {{ ref('dim_time') }}
),
dim_complaint_location AS
(
    SELECT * FROM {{ ref('dim_complaint_location') }}
)

SELECT unique_key,
    agency_311_dim_id,
```

```

complaint_type_dim_id,
Channel_Type_Dim_ID,
Complaint_Location_Dim_ID,
date_dim_id AS Created_Date_Dim_ID,
Time_Dim_ID AS Created_Time_Dim_ID
FROM `fluted-quasar-341922.311_complaint_requests.311_illegal_parking` AS sr
    INNER JOIN dim_agency ON dim_agency.agency_name = sr.agency_name
    INNER JOIN dim_complaint_type ON dim_complaint_type.descriptor = sr.descriptor
    INNER JOIN dim_date ON dim_date.full_date = EXTRACT(DATE FROM sr.created_date)
    INNER JOIN dim_time ON dim_time.full_time = EXTRACT(TIME FROM sr.created_date)
    INNER JOIN dim_complaint_location ON dim_complaint_location.borough = sr.borough
        AND dim_complaint_location.incident_zip=sr.incident_zip
        AND dim_complaint_location.latitude = sr.latitude
        AND dim_complaint_location.longitude = sr.longitude
    INNER JOIN dim_channel_type ON dim_channel_type.open_data_channel_type = sr.open_data_channel_type

```



fct\_complaints
 QUERY
 SHARE
 COPY
 SNAPSHOT
 DELETE
 EXPORT

---

|     | SCHEMA     | DETAILS           | PREVIEW               |                     |                           |                     |         |
|-----|------------|-------------------|-----------------------|---------------------|---------------------------|---------------------|---------|
| Row | unique_key | agency_311_dim_id | complaint_type_dim_id | Channel_Type_Dim_ID | Complaint_Location_Dim_ID | Created_Date_Dim_ID | Created |
| 1   | 37174916   | 1                 | 1                     | 2                   | 115341                    | 256                 | 37858   |
| 2   | 37177430   | 1                 | 1                     | 3                   | 259122                    | 256                 | 32405   |
| 3   | 37174122   | 1                 | 1                     | 3                   | 199602                    | 256                 | 39899   |
| 4   | 37172619   | 1                 | 1                     | 1                   | 246257                    | 256                 | 22286   |
| 5   | 37175062   | 1                 | 1                     | 3                   | 242570                    | 256                 | 9242    |
| 6   | 37177900   | 1                 | 1                     | 1                   | 157645                    | 256                 | 15926   |
| 7   | 37172648   | 1                 | 1                     | 1                   | 192658                    | 256                 | 34901   |
| 8   | 37174919   | 1                 | 1                     | 1                   | 175312                    | 256                 | 26486   |
| 9   | 37175049   | 1                 | 1                     | 3                   | 189232                    | 256                 | 25357   |
| 10  | 37174211   | 1                 | 1                     | 3                   | 293402                    | 256                 | 36277   |
| 11  | 37175641   | 1                 | 1                     | 3                   | 2524                      | 256                 | 29826   |
| 12  | 37174909   | 1                 | 1                     | 1                   | 25087                     | 256                 | 41495   |
| 13  | 37177409   | 1                 | 1                     | 2                   | 100961                    | 256                 | 24684   |
| 14  | 37173500   | 1                 | 1                     | 2                   | 259415                    | 256                 | 29875   |
| 15  | 37167571   | 1                 | 1                     | 2                   | 271094                    | 256                 | 2641    |
| 16  | 37175804   | 1                 | 1                     | 3                   | 196406                    | 256                 | 31278   |
| 17  | 37177901   | 1                 | 1                     | 2                   | 287597                    | 256                 | 10839   |

Results per page: 50 ▾ 1 – 50 of 518238 |◀|◀|▶|▶|▶|

## **Open Parking and Camera Violation**

```
CREATE TABLE `rising-matrix-350000.Open_Parking_and_Camera_Violations_2017_to_2021.2017` AS
SELECT Plate,
       State,
       License_Type,
       Summons_Number,
       Violation_time,
       Violation,
       Issue_date,
       Fine_Amount,
       Penalty_Amount,
       Interest_Amount,
       Reduction_Amount,
       Payment_Amount,
       Amount_Due,
       Precinct,
       County,
       Issuing_Agency
FROM `handy-bonbon-142723.nyc_open_parking_and_camera_violations.Open_Parking_and_Camera_Violations_2017`

--Similar Queries are created to create tables to store the data from 2017 to 2021
```

```
CREATE TABLE `rising-matrix-350000.Open_Parking_and_Camera_Violations_2017_to_2021.ALL` AS
SELECT *
FROM `rising-matrix-350000.Open_Parking_and_Camera_Violations_2017_to_2021.2017`
UNION ALL
SELECT *
FROM `rising-matrix-350000.Open_Parking_and_Camera_Violations_2017_to_2021.2018`
UNION ALL
SELECT *
FROM `rising-matrix-350000.Open_Parking_and_Camera_Violations_2017_to_2021.2019`
UNION ALL
SELECT *
FROM `rising-matrix-350000.Open_Parking_and_Camera_Violations_2017_to_2021.2020`
UNION ALL
SELECT *
FROM `rising-matrix-350000.Open_Parking_and_Camera_Violations_2017_to_2021.2021`

--Combined all five tables into one table
```

```
-- Filtering data to be able to convert string to time
CREATE TABLE `rising-matrix-350000.Open_Parking_and_Camera_Violations_2017_to_2021.All_Filtered` AS
SELECT Plate,
       State,
       License_Type,
       Summons_Number,
       PARSE_TIME('%H:%M%p', Violation_time||'M') AS Violation_time,
       Violation AS Violation_Type,
       Issue_date,
       Fine_Amount,
       Penalty_Amount,
       Interest_Amount,
       Reduction_Amount,
       Payment_Amount,
       Amount_Due,
       Precinct,
       County,
       Issuing_Agency
  from `rising-matrix-350000.Open_Parking_and_Camera_Violations_2017_to_2021.ALL`
 WHERE  (violation_time LIKE '%A' OR violation_time LIKE '%P' )
    AND SUBSTR(violation_time,0,2) NOT IN('26', '28', '37', '38', '50', '51', '52', '55', '56', '68', '48', '49', '70')
    AND LENGTH(violation_time) = 6
    AND SUBSTR(violation_time,0,1) NOT IN ('.', ':')
    AND NOT violation_time LIKE '%+%'
    AND NOT violation_time LIKE '%.%'
    AND NOT violation_time LIKE '%-%'
    AND NOT violation_time LIKE '%/%'
    AND NOT violation_time LIKE '% %'
```

```
AND NOT violation_time LIKE '%`%'
AND CAST(SUBSTR(violation_time,0,2) AS INT64) < 24
```

```
dbt_project.yml
name: 'team_01_cis4400_group_project'
version: '1.0.0'
config-version: 2
profile: 'default'
model-paths: ["models"]
analysis-paths: ["analyses"]
test-paths: ["tests"]
seed-paths: ["seeds"]
macro-paths: ["macros"]
snapshot-paths: ["snapshots"]
target-path: "target"
clean-targets:
  - "target"
  - "dbt_packages"
models:
  team_01_cis4400_group_project:
    marts:
      +materialized: table
```

```
packages.yml
packages:
  - package: dbt-labs/dbt_utils
    version: 0.8.4
```

```
dim_location.sql
SELECT
    row_number() OVER () AS Location_Dim_ID,
    State, Precinct, County
FROM
    (SELECT DISTINCT State, Precinct, County
     FROM `rising-matrix-350000.Open_Parking_and_Camera_Violations_2017_to_2021.ALL`  

)
```

```
dim_vehicles.sql
SELECT
    row_number() OVER () AS Vehicles_Dim_ID,
    Plate, License_Type
FROM
    (SELECT DISTINCT Plate, License_Type
     FROM `rising-matrix-350000.Open_Parking_and_Camera_Violations_2017_to_2021.ALL`  

)
```

```
dimViolation_type.sql
SELECT
    row_number() OVER () AS Violation_Type_Dim_ID,
    Violation AS Violation_Type
FROM
    (SELECT DISTINCT Violation
     FROM `rising-matrix-350000.Open_Parking_and_Camera_Violations_2017_to_2021.ALL`  

)
```

```
dim_agency.sql
SELECT
    row_number() OVER () AS Agency_Dim_ID,
    Issuing_Agency
FROM
    (SELECT DISTINCT Issuing_Agency
     FROM `rising-matrix-350000.Open_Parking_and_Camera_Violations_2017_to_2021.ALL`  
)
```

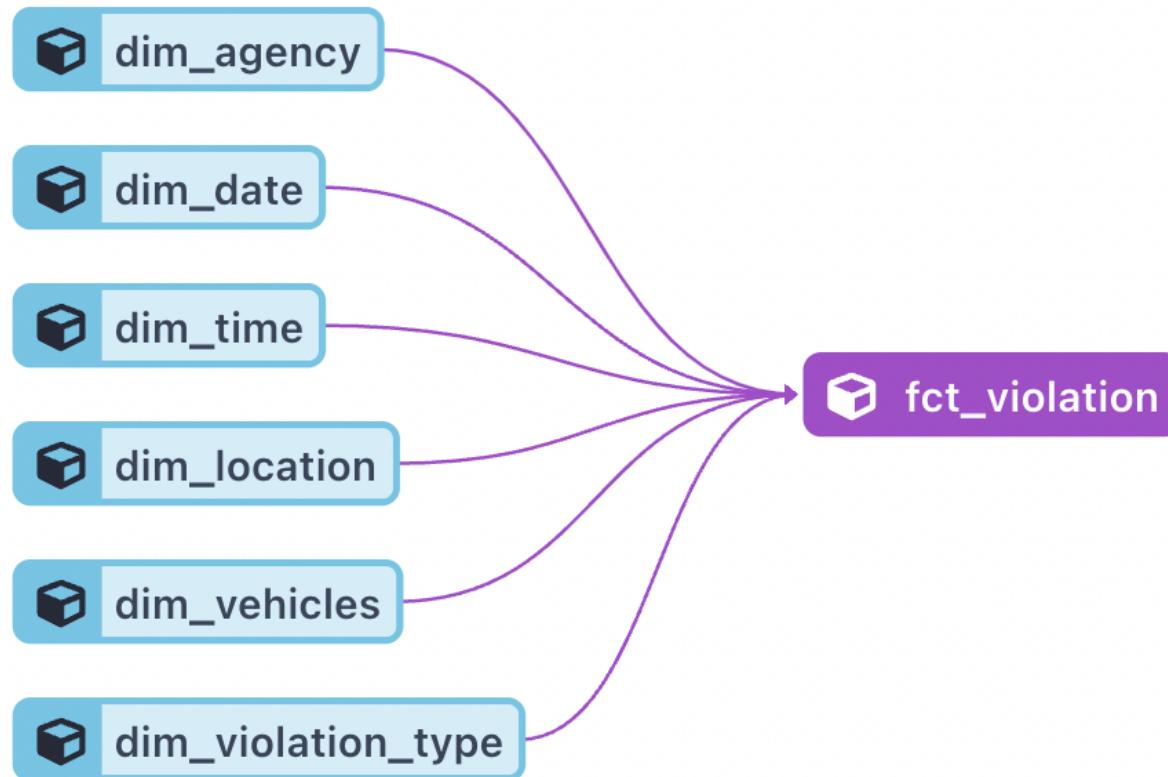
```
dim_date.sql
SELECT
    ROW_NUMBER() OVER() AS Date_Dim_ID,
    FORMAT_DATE("%Y%m%d",d) AS Date_Integer,
    d AS Full_Date,
    EXTRACT(YEAR FROM d) AS Year,
    EXTRACT(WEEK FROM d) AS Week_Number_of_Year,
    EXTRACT(DAY FROM d) AS Day_Number_of_Month,
    FORMAT_DATE('%Q', d) AS Quarter_Number_of_Year,
    EXTRACT(MONTH FROM d) AS Month,
    FORMAT_DATE('%B', d) AS Month_Name,
    FORMAT_DATE('%w', d) AS Day_Number_of_Week,
    FORMAT_DATE('%A', d) AS Day_Name,
    (CASE WHEN FORMAT_DATE('%A', d) IN ('Sunday', 'Saturday')
          THEN 0 ELSE 1 END) AS Day_is_Weekday
FROM(
    SELECT *
    FROM UNNEST(GENERATE_DATE_ARRAY('2017-01-01', '2021-12-31', INTERVAL 1 DAY)) AS d)
```

```
dim_time.sql
SELECT
    ROW_NUMBER() OVER() AS Time_Dim_ID,
    CAST(T as time) AS Full_Time,
    EXTRACT(HOUR FROM T) AS Hour,
    EXTRACT(MINUTE FROM T) AS Minute,
    EXTRACT(SECOND FROM T) AS Second
FROM (SELECT *
      FROM UNNEST(GENERATE_TIMESTAMP_ARRAY('2017-01-01 00:00:00', '2017-01-01 11:59:59', INTERVAL 1 SECOND)) AS T)

fct_violations.sql
WITH dim_agency AS
(
    SELECT * FROM {{ ref('dim_agency') }}
),
dim_date AS
(
    SELECT * FROM {{ ref('dim_date') }}
),
dim_time AS
(
    SELECT * FROM {{ ref('dim_time') }}
),
dim_location AS
(
    SELECT * FROM {{ ref('dim_location') }}
),
dim_vehicles AS
```

```
(  
    SELECT * FROM {{ ref('dim_vehicles') }}  
) ,  
dimViolationType AS  
(  
    SELECT * FROM {{ ref('dim_violation_type') }}  
)  
  
SELECT  
    Summons_Number,  
    Fine_Amount,  
    Penalty_Amount,  
    Interest_Amount,  
    Reduction_Amount,  
    Payment_Amount,  
    Amount_Due,  
    Agency_Dim_ID,  
    Date_Dim_ID AS Issue_Date_Dim_ID,  
    Time_Dim_ID AS Violation_Time_Dim_ID,  
    Location_Dim_ID,  
    Vehicles_Dim_ID,  
    Violation_Type_Dim_ID  
FROM `rising-matrix-350000.Open_Parking_and_Camera_Violations_2017_to_2021.All_Filtered` AS opcv  
    INNER JOIN dim_agency ON dim_agency.issuing_agency = opcv.issuing_agency  
    INNER JOIN dim_date ON dim_date.full_date = opcv.Issue_date  
    INNER JOIN dim_time ON dim_time.Full_Time = opcv.Violation_Time  
    INNER JOIN dim_location ON dim_location.state = opcv.state  
        AND dim_location.precinct = opcv.precinct
```

```
AND dim_location.county = opcv.county  
INNER JOIN dim_vehicles ON dim_vehicles.plate = opcv.plate  
AND dim_vehicles.license_type = opcv.license_type  
INNER JOIN dimViolation_type ON dimViolation_type.Violation_Type = opcv.Violation_Type
```



## fctViolation

[QUERY](#)[SHARE](#)[COPY](#)[SNAPSHOT](#)[DELETE](#)[EXPORT](#)

SCHEMA

DETAILS

PREVIEW

| Row | Summons_Number | Fine_Amount | Penalty_Amount | Interest_Amount | Reduction_Amount | Payment_Amount | Amount_Due | Agency_Din |
|-----|----------------|-------------|----------------|-----------------|------------------|----------------|------------|------------|
| 1   | 8533992725     | 60.0        | 10.0           | 0.0             | 0.0              | 70.0           | 0.0        | 2          |
| 2   | 8497302345     | 65.0        | 0.0            | 0.0             | 13.0             | 52.0           | 0.0        | 2          |
| 3   | 8793215848     | 65.0        | 0.0            | 0.0             | 35.0             | 30.0           | 0.0        | 2          |
| 4   | 8558669351     | 115.0       | 60.0           | 12.32           | 0.22             | 187.1          | 0.0        | 2          |
| 5   | 8658924622     | 65.0        | 60.0           | 0.43            | 0.0              | 125.43         | 0.0        | 2          |
| 6   | 8559016867     | 45.0        | 0.0            | 0.0             | 9.0              | 36.0           | 0.0        | 2          |
| 7   | 1483476212     | 180.0       | 0.0            | 0.0             | 0.0              | 180.0          | 0.0        | 5          |
| 8   | 1447620288     | 60.0        | 10.0           | 0.0             | 0.0              | 70.0           | 0.0        | 4          |
| 9   | 8379024485     | 60.0        | 10.0           | 0.0             | 0.0              | 70.0           | 0.0        | 2          |
| 10  | 8605444629     | 65.0        | 60.0           | 12.84           | 0.41             | 137.43         | 0.0        | 2          |
| 11  | 8481346159     | 65.0        | 60.0           | 1.06            | 0.0              | 126.06         | 0.0        | 2          |
| 12  | 8657658793     | 60.0        | 60.0           | 1.52            | 0.21             | 121.31         | 0.0        | 2          |
| 13  | 8484793424     | 65.0        | 0.0            | 0.0             | 7.0              | 58.0           | 0.0        | 2          |
| 14  | 8673152859     | 165.0       | 30.0           | 0.0             | 0.0              | 195.0          | 0.0        | 2          |
| 15  | 8532718814     | 60.0        | 60.0           | 0.15            | 0.0              | 120.15         | 0.0        | 2          |
| 16  | 8621624102     | 65.0        | 0.0            | 0.0             | 13.0             | 52.0           | 0.0        | 2          |
| 17  | 8606611797     | 60.0        | 10.0           | 0.0             | 0.0              | 70.0           | 0.0        | 2          |

Results per page:

50 ▼

1 – 50 of 55816737

&lt; &lt; &gt; &gt;&gt;

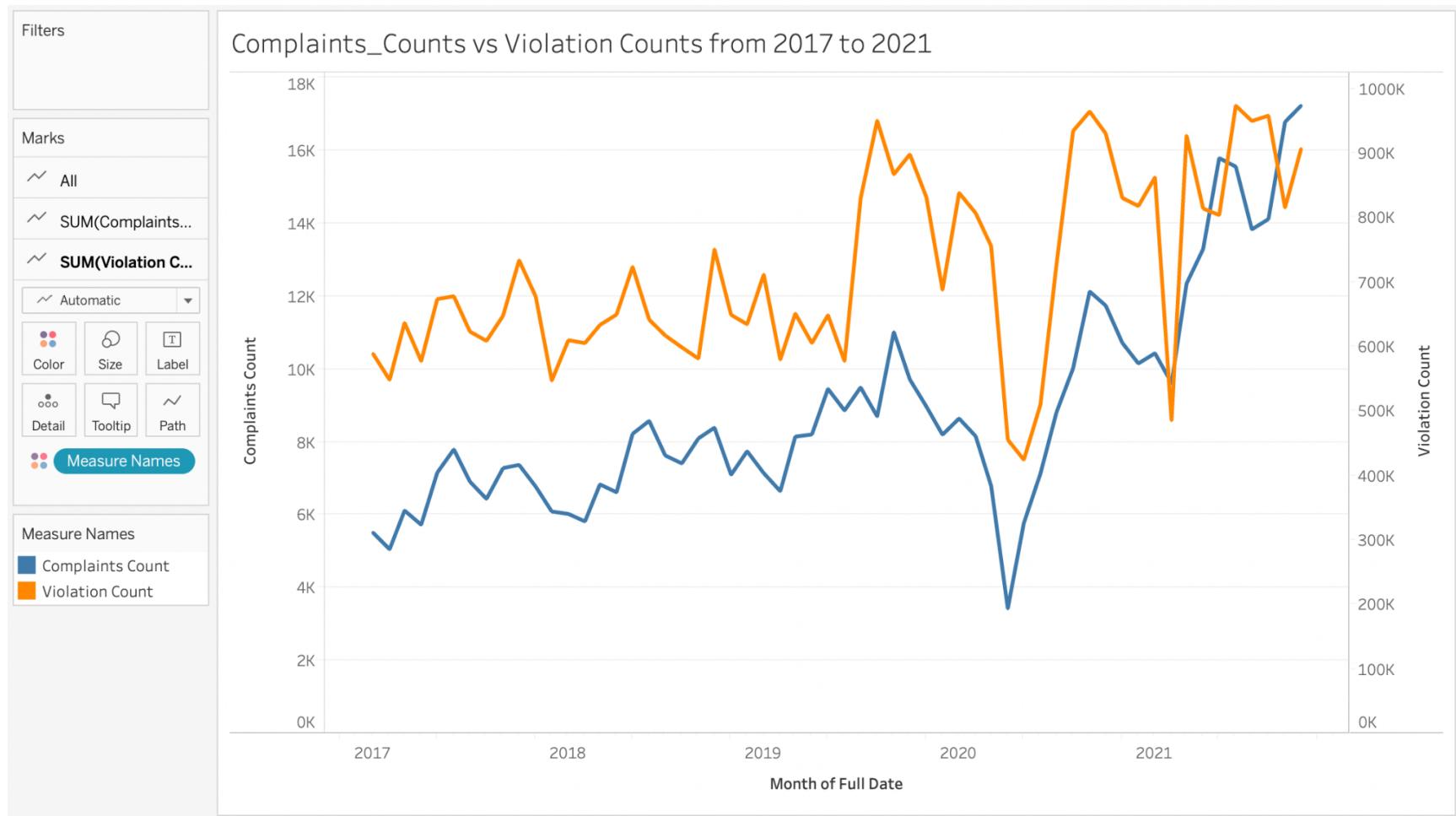
## Tableau Visualizations

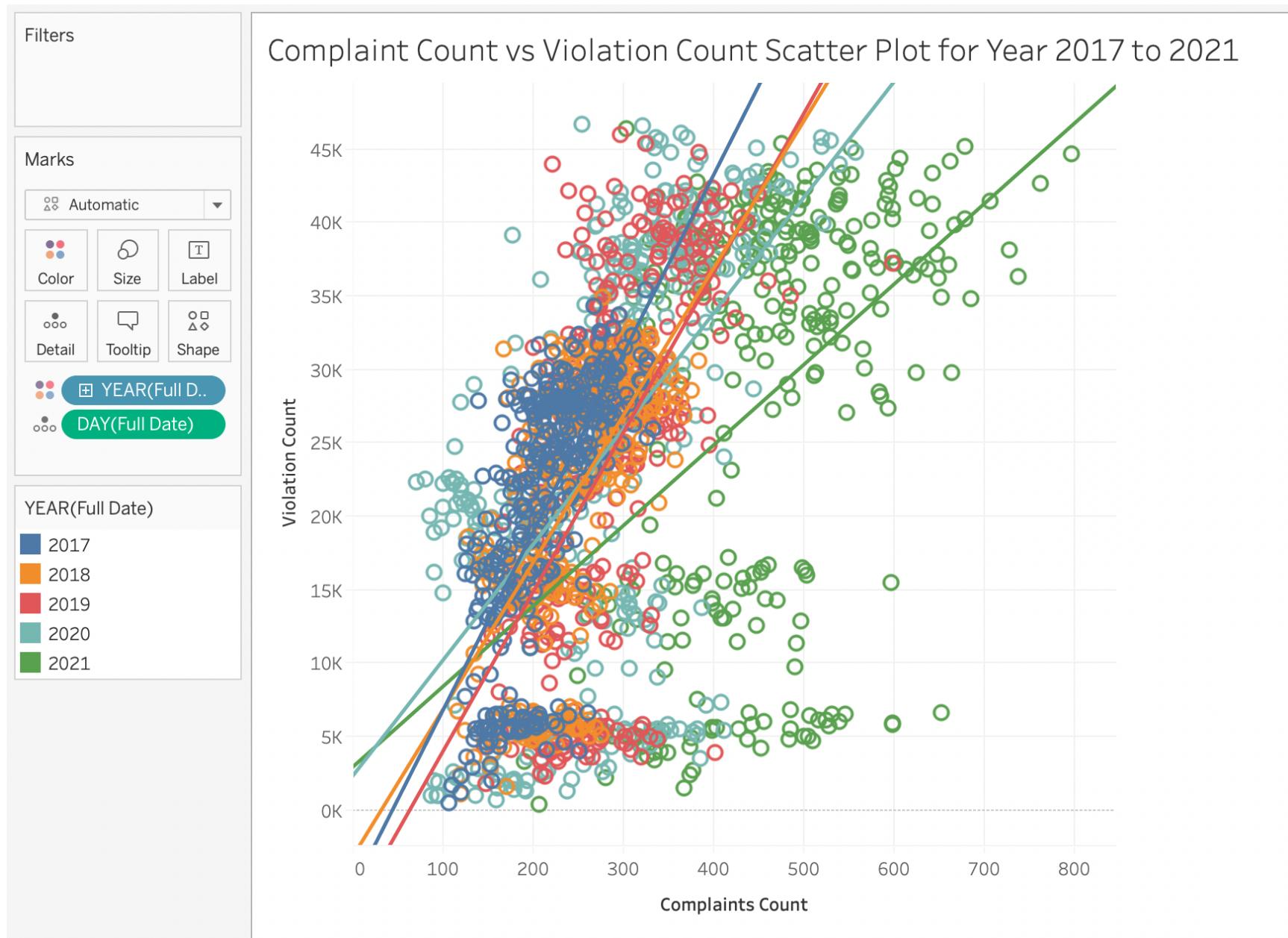
```
WITH complaints AS
(select created_date_dim_id,
 dim_date.Full_Date,
 count(unique_key) as complaints_count
from `rising-matrix-350000.dbt_sdeng.fct_complaints` as fct_complaints
INNER JOIN `rising-matrix-350000.dbt_sdeng.dim_date` as dim_date ON dim_date.Date_Dim_ID =
fct_complaints.created_date_dim_id
WHERE dim_date.Full_Date BETWEEN '2017-01-01' AND '2021-10-30'
group by dim_date.Full_Date, created_date_dim_id
order by dim_date.Full_Date),


violations AS
(
select Issue_Date_Dim_ID,
 dim_date.Full_Date,
 dim_location.State as location,
 count(Summons_Number) as violation_count
from `rising-matrix-350000.dbt_sdeng.fctViolation` as fct_violation
INNER JOIN `rising-matrix-350000.dbt_sdeng.dim_date` as dim_date ON dim_date.Date_Dim_ID =
fct_violation.Issue_Date_Dim_ID
INNER JOIN `rising-matrix-350000.dbt_sdeng.dim_location` as dim_location ON dim_location.Location_Dim_ID =
fct_violation.Location_Dim_ID
INNER JOIN `rising-matrix-350000.dbt_sdeng.dim_agency` as dim_agency ON dim_agency.Agency_Dim_ID =
fct_violation.Agency_Dim_ID
WHERE dim_date.Full_Date BETWEEN '2017-01-01' AND '2021-10-30'
AND dim_location.State = 'NY'
```

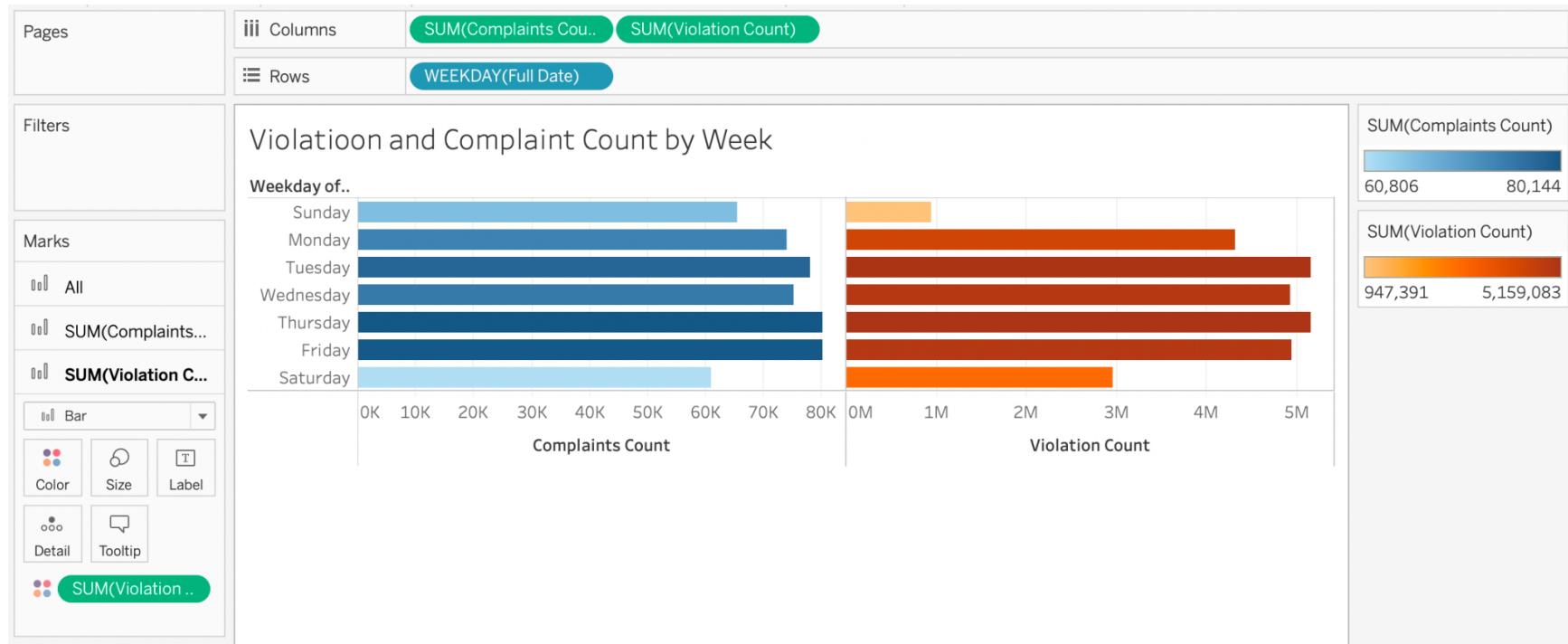
```
    AND dim_agency.Issuing_Agency IN ('PARKING CONTROL UNIT', 'TRAFFIC', 'POLICE DEPARTMENT')
group by dim_date.Full_Date, Issue_Date_Dim_ID, dim_location.State
order by dim_date.Full_Date
)

select
Full_Date,
location,
complaints_count,
violation_count
from complaints inner join violations using(Full_Date)
order by Full_Date
```



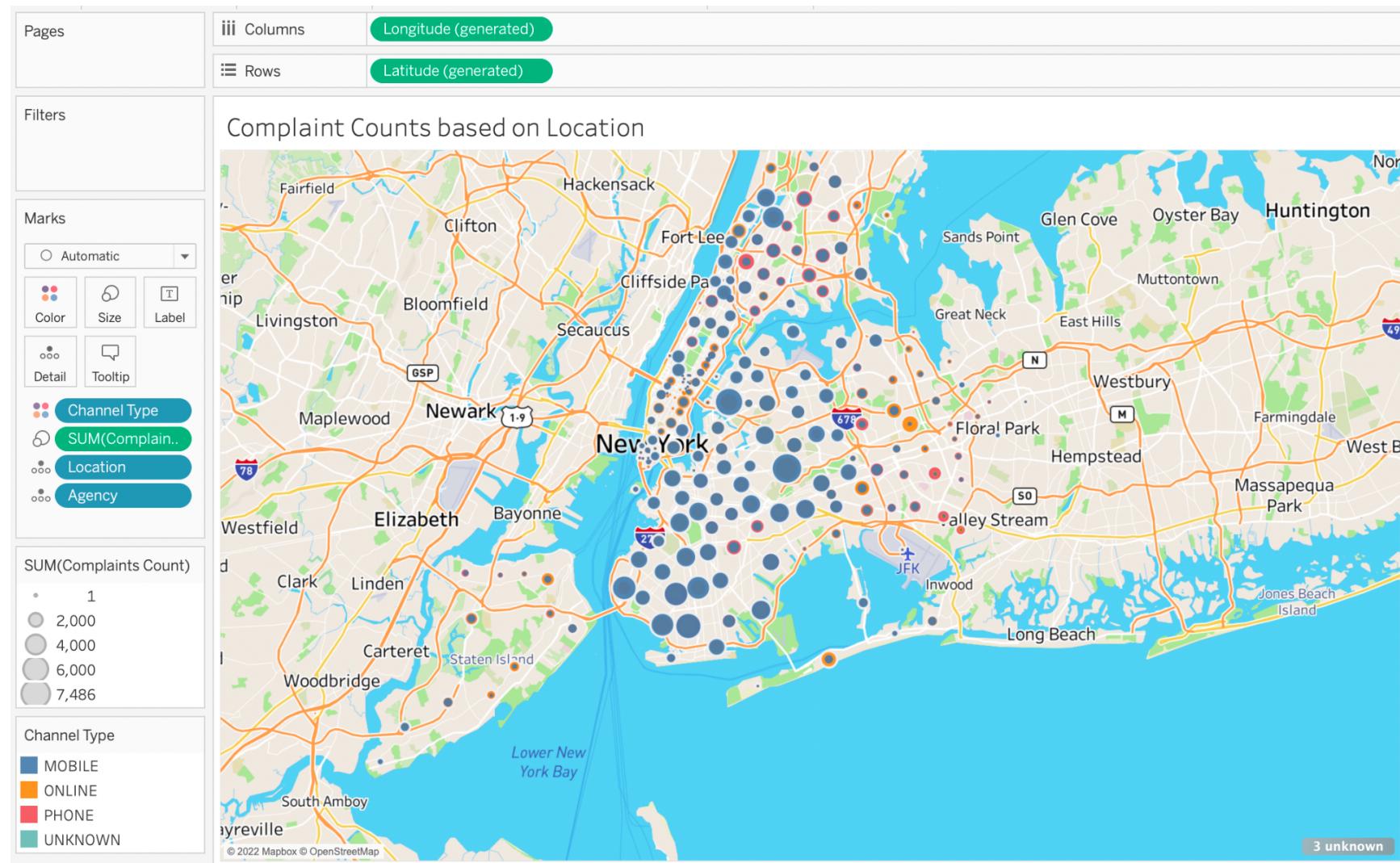


## Which day of the week is the busiest for the New York Police Department?



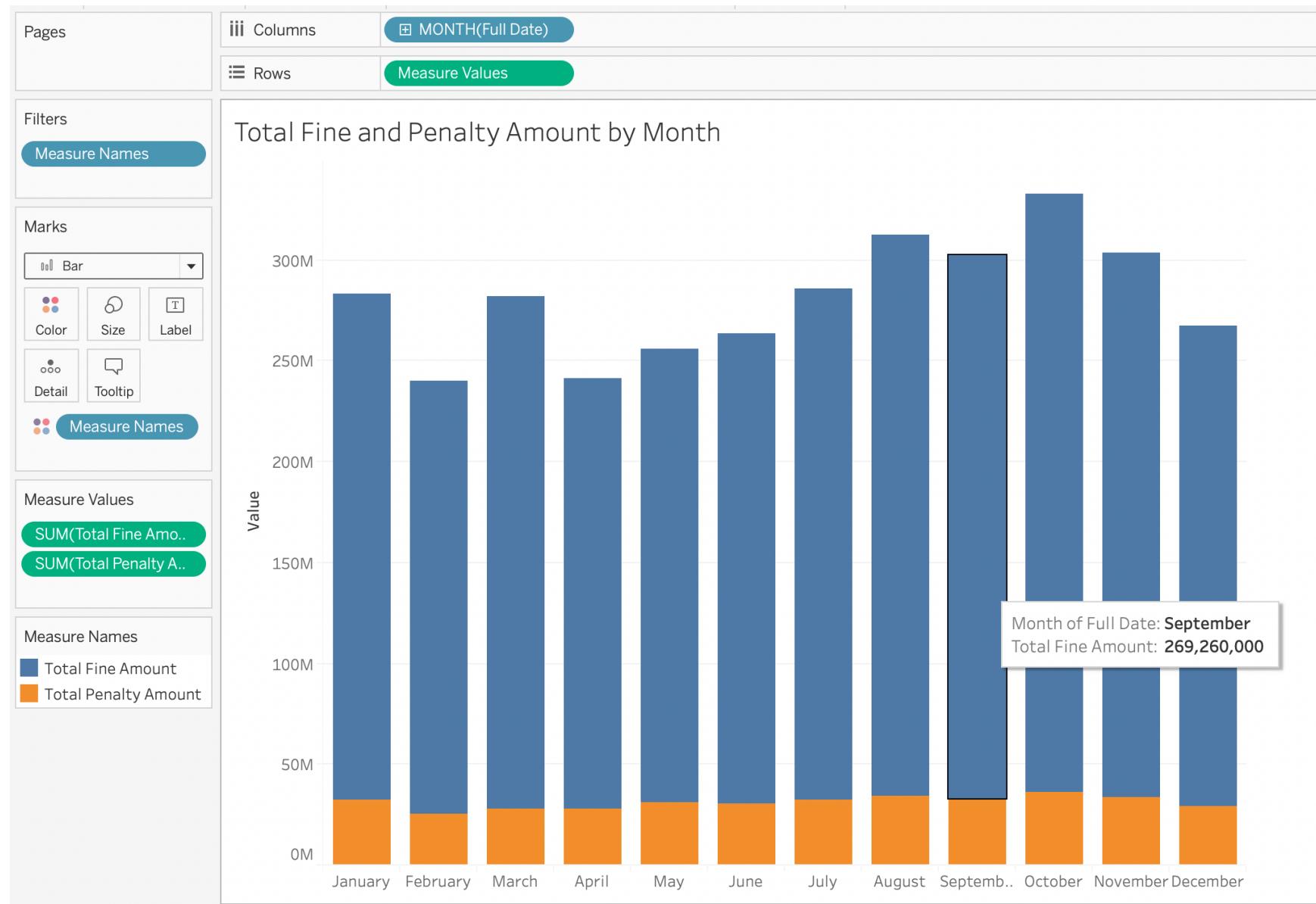
```
WITH complaints AS
(select
    dim_complaint_location.incident_zip as location,
    dim_channel_type.open_data_channel_type as channel_type,
    dim_311_agency.agency_name as agency,
    count(unique_key) as complaints_count
from `rising-matrix-350000.dbt_sdeng.fct_complaints` as fct_complaints
INNER JOIN `rising-matrix-350000.dbt_sdeng.dim_complaint_location` as dim_complaint_location ON
dim_complaint_location.Complaint_Location_Dim_ID = fct_complaints.Complaint_Location_Dim_ID
INNER JOIN `rising-matrix-350000.dbt_sdeng.dim_channel_type` as dim_channel_type ON
dim_channel_type.Channel_Type_Dim_ID = fct_complaints.Channel_Type_Dim_ID
INNER JOIN `rising-matrix-350000.dbt_sdeng.dim_311_agency` as dim_311_agency ON dim_311_agency.Agency_311_Dim_ID =
fct_complaints.agency_311_dim_id
group by location, channel_type, agency)

select
    complaints_count,
    location,
    channel_type,
    agency
from complaints
```



**Which month has the highest fine and penalty amount in NY over the past 5 years?**

```
SELECT
    SUM(Fine_Amount) AS Total_Fine_Amount,
    SUM(Penalty_Amount) AS Total_Penalty_Amount,
    dim_date.Full_Date,
    dim_location.State
FROM `rising-matrix-350000.dbt_sdeng.fctViolation` AS fctViolation
INNER JOIN `rising-matrix-350000.dbt_sdeng.dim_date` AS dim_date
ON fctViolation.Issue_Date_Dim_ID = dim_date.Date_Dim_ID
INNER JOIN `rising-matrix-350000.dbt_sdeng.dim_location` AS dim_location
ON dim_location.Location_Dim_ID = fctViolation.Location_Dim_ID
INNER JOIN `rising-matrix-350000.dbt_sdeng.dim_agency` AS dim_agency
ON dim_agency.Agency_Dim_ID = fctViolation.Agency_Dim_ID
WHERE dim_date.Full_Date BETWEEN '2017-01-01' AND '2021-10-30'
    AND dim_location.State = 'NY'
    AND dim_agency.Issuing_Agency IN ('PARKING CONTROL UNIT', 'TRAFFIC', 'POLICE DEPARTMENT')
GROUP BY dim_date.Full_Date, dim_location.State
```



## Reference

1. (DOF), Department of Finance. “Open Parking and Camera Violations: NYC Open Data.” *Open Parking and Camera Violations | NYC Open Data*, 12 Feb. 2022,  
<https://data.cityofnewyork.us/City-Government/Open-Parking-and-Camera-Violations/nc67-uf89>.
2. 311, DoITT. “311 Service Requests from 2010 to Present: NYC Open Data.” *311 Service Requests from 2010 to Present | NYC Open Data*, 16 Feb. 2022,  
<https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9>.

## Appendix

### **311 Service Request**

```
SELECT unique_key,  
  
       created_date,  
       closed_date,  
       agency,  
       agency_name,  
       complaint_type,  
       descriptor,  
       incident_address,  
       cross_street_1,  
       cross_street_2,  
       intersection_street_1,  
       intersection_street_2,  
       city,  
       borough,
```

```

incident_zip,
bbl,
latitude,
longitude,
open_data_channel_type,
FROM `bigquery-public-data.new_york_311.311_service_requests`
WHERE complaint_type = 'Illegal Parking' AND FORMAT_DATE("%Y", created_date) IN
('2017', '2018', '2019', '2020', '2021')

```

Query results

SAVE RESULTS ▾
 EXPLORE DATA ▾

| JOB INFORMATION |            |                         | RESULTS | JSON                    | EXECUTION DETAILS |                                 |                 |                 |          |
|-----------------|------------|-------------------------|---------|-------------------------|-------------------|---------------------------------|-----------------|-----------------|----------|
| Row             | unique_key | created_date            |         | closed_date             | agency            | agency_name                     | complaint_type  | descriptor      | incident |
| 1               | 50790209   | 2021-06-06 18:31:48 UTC |         | 2021-06-06 21:16:42 UTC | NYPD              | New York City Police Department | Illegal Parking | Blocked Hydrant | 15-26    |
| 2               | 50798809   | 2021-06-06 17:41:32 UTC |         | 2021-06-06 18:37:19 UTC | NYPD              | New York City Police Department | Illegal Parking | Blocked Hydrant | 87-44    |
| 3               | 50790852   | 2021-06-06 12:26:27 UTC |         | 2021-06-06 17:08:49 UTC | NYPD              | New York City Police Department | Illegal Parking | Blocked Hydrant | 149-30   |
| 4               | 50790861   | 2021-06-06 19:28:01 UTC |         | 2021-06-06 20:13:35 UTC | NYPD              | New York City Police Department | Illegal Parking | Blocked Hydrant | 104-30   |
| 5               | 50791546   | 2021-06-06 20:16:32 UTC |         | 2021-06-06 20:36:08 UTC | NYPD              | New York City Police Department | Illegal Parking | Blocked Hydrant | 89-40    |

Results per page: 50 ▾ 1 – 50 of 967634 |< < > >|

```
In [1]: pip install pandas-profiling

Requirement already satisfied: pandas-profiling in /opt/anaconda3/lib/python3.8/site-packages (3.1.0)
Requirement already satisfied: jinja2>=2.11.1 in /opt/anaconda3/lib/python3.8/site-packages (from pandas-profiling) (2.11.2)
Requirement already satisfied: visions[type_image_path]==0.7.4 in /opt/anaconda3/lib/python3.8/site-packages (from pandas-profiling) (0.7.4)
Requirement already satisfied: tqdm>=4.48.2 in /opt/anaconda3/lib/python3.8/site-packages (from pandas-profiling) (4.50.2)
Requirement already satisfied: missingno>=0.4.2 in /opt/anaconda3/lib/python3.8/site-packages (from pandas-profiling) (0.5.1)
Requirement already satisfied: joblib==1.0.1 in /opt/anaconda3/lib/python3.8/site-packages (from pandas-profiling) (1.0.1)
Requirement already satisfied: markupsafe==2.0.1 in /opt/anaconda3/lib/python3.8/site-packages (from pandas-profiling) (2.0.1)
Requirement already satisfied: scipy>=1.4.1 in /opt/anaconda3/lib/python3.8/site-packages (from pandas-profiling) (1.5.2)
Requirement already satisfied: multimethod>=1.4 in /opt/anaconda3/lib/python3.8/site-packages (from pandas-profiling) (1.8)
Requirement already satisfied: tangled-up-in-unicode==0.1.0 in /opt/anaconda3/lib/python3.8/site-packages (from pandas-profiling) (0.1.0)
```

```
In [2]: import pandas_profiling
import pandas as pd
```

```
In [3]: df = pd.read_csv('/Users/simengdeng/Desktop/CIS 4400/311_illegal_parking_2017_to_2021_without_table_name.csv', low_memo
```

```
In [4]: data_report = pandas_profiling.ProfileReport(df)
```

```
In [5]: data_report.to_file('311_illegal_parking_2017_to_2021_data_report.html')
```

Summarize dataset: 100%  59/59 [01:17<00:00, 1.31s/it, Completed]

Generate report structure: 100%  1/1 [00:02<00:00, 2.94s/it]

Render HTML: 100%  1/1 [00:00<00:00, 1.03it/s]

Export report to file: 100%  1/1 [00:00<00:00, 46.63it/s]

# Overview

| Overview                      | Alerts <span>42</span> | Reproduction |
|-------------------------------|------------------------|--------------|
| <b>Dataset statistics</b>     |                        |              |
| Number of variables           |                        | 19           |
| Number of observations        |                        | 967634       |
| Missing cells                 |                        | 1073105      |
| Missing cells (%)             |                        | 5.8%         |
| Duplicate rows                |                        | 0            |
| Duplicate rows (%)            |                        | 0.0%         |
| Total size in memory          |                        | 140.3 MiB    |
| Average record size in memory |                        | 152.0 B      |
| <b>Variable types</b>         |                        |              |
| Numeric                       |                        | 5            |
| Categorical                   |                        | 14           |

# Variables

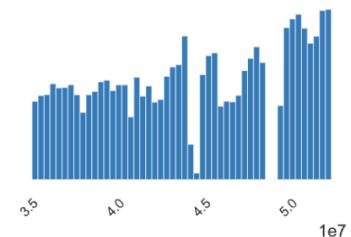
## unique\_key

Real number ( $\mathbb{R}_{\geq 0}$ )

UNIQUE

|              |             |
|--------------|-------------|
| Distinct     | 967634      |
| Distinct (%) | 100.0%      |
| Missing      | 0           |
| Missing (%)  | 0.0%        |
| Infinite     | 0           |
| Infinite (%) | 0.0%        |
| Mean         | 44511226.44 |

|              |          |
|--------------|----------|
| Minimum      | 35137568 |
| Maximum      | 52454392 |
| Zeros        | 0        |
| Zeros (%)    | 0.0%     |
| Negative     | 0        |
| Negative (%) | 0.0%     |
| Memory size  | 7.4 MiB  |



[Toggle details](#)

## created\_date

Categorical

HIGH CARDINALITY

UNIFORM

|              |         |
|--------------|---------|
| Distinct     | 963455  |
| Distinct (%) | 99.6%   |
| Missing      | 0       |
| Missing (%)  | 0.0%    |
| Memory size  | 7.4 MiB |

2021-05-04 22:27:59 UTC 4  
2021-04-15 21:59:15 UTC 3  
2021-07-27 12:30:38 UTC 3  
2018-06-24 23:28:02 UTC 3  
2020-09-27 07:46:07 UTC 3  
Other values (963450) 967618

[Toggle details](#)

**closed\_date**

Categorical

**HIGH CARDINALITY**  
**UNIFORM**

|                     |         |
|---------------------|---------|
| <b>Distinct</b>     | 903138  |
| <b>Distinct (%)</b> | 93.4%   |
| <b>Missing</b>      | 486     |
| <b>Missing (%)</b>  | 0.1%    |
| <b>Memory size</b>  | 7.4 MiB |

2019-07-12 15:08:07 UTC 23

2019-07-12 12:08:24 UTC 17

2019-07-12 11:25:46 UTC 16

2019-07-12 15:08:02 UTC 14

2019-07-12 11:26:21 UTC 14

Other values (903133) 967064

[Toggle details](#)**agency**

Categorical

**CONSTANT**  
**HIGH CORRELATION**  
**REJECTED**

|                     |         |
|---------------------|---------|
| <b>Distinct</b>     | 1       |
| <b>Distinct (%)</b> | < 0.1%  |
| <b>Missing</b>      | 0       |
| <b>Missing (%)</b>  | 0.0%    |
| <b>Memory size</b>  | 7.4 MiB |

NYPD

967634

[Toggle details](#)

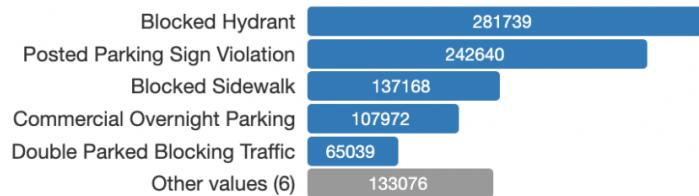
|                         |                     |         |                                 |                                |
|-------------------------|---------------------|---------|---------------------------------|--------------------------------|
| <b>agency_name</b>      | <b>Distinct</b>     | 3       | New York City Police Department | 965906                         |
| Categorical             | <b>Distinct (%)</b> | < 0.1%  | Traffic Management Center       | 1723                           |
| <b>HIGH CORRELATION</b> | <b>Missing</b>      | 0       | NYPD                            | 5                              |
|                         | <b>Missing (%)</b>  | 0.0%    |                                 |                                |
|                         | <b>Memory size</b>  | 7.4 MiB |                                 |                                |
|                         |                     |         |                                 | <a href="#">Toggle details</a> |
| <b>complaint_type</b>   | <b>Distinct</b>     | 1       | Illegal Parking                 | 967634                         |
| Categorical             | <b>Distinct (%)</b> | < 0.1%  |                                 |                                |
| <b>CONSTANT</b>         | <b>Missing</b>      | 0       |                                 |                                |
| <b>HIGH CORRELATION</b> | <b>Missing (%)</b>  | 0.0%    |                                 |                                |
| REJECTED                | <b>Memory size</b>  | 7.4 MiB |                                 |                                |
|                         |                     |         |                                 | <a href="#">Toggle details</a> |

**descriptor**

Categorical

**HIGH CORRELATION**

|                     |         |
|---------------------|---------|
| <b>Distinct</b>     | 11      |
| <b>Distinct (%)</b> | < 0.1%  |
| <b>Missing</b>      | 0       |
| <b>Missing (%)</b>  | 0.0%    |
| <b>Memory size</b>  | 7.4 MiB |

**Toggle details****incident\_address**

Categorical

**HIGH CARDINALITY****MISSING**

|                     |         |
|---------------------|---------|
| <b>Distinct</b>     | 256415  |
| <b>Distinct (%)</b> | 28.0%   |
| <b>Missing</b>      | 51038   |
| <b>Missing (%)</b>  | 5.3%    |
| <b>Memory size</b>  | 7.4 MiB |

**Toggle details**

### [cross\\_street\\_1](#)

Categorical

HIGH CARDINALITY  
MISSING

|                     |         |
|---------------------|---------|
| <b>Distinct</b>     | 9160    |
| <b>Distinct (%)</b> | 1.0%    |
| <b>Missing</b>      | 56469   |
| <b>Missing (%)</b>  | 5.8%    |
| <b>Memory size</b>  | 7.4 MiB |

|          |       |
|----------|-------|
| BEND     | 15418 |
| DEAD END | 10501 |
| 5 AVENUE | 8998  |
| 3 AVENUE | 7927  |
| BROADWAY | 7517  |

Other values (9155) 860804

[Toggle details](#)

### [cross\\_street\\_2](#)

Categorical

HIGH CARDINALITY  
MISSING

|                     |         |
|---------------------|---------|
| <b>Distinct</b>     | 9248    |
| <b>Distinct (%)</b> | 1.0%    |
| <b>Missing</b>      | 56761   |
| <b>Missing (%)</b>  | 5.9%    |
| <b>Memory size</b>  | 7.4 MiB |

|          |       |
|----------|-------|
| BEND     | 14891 |
| DEAD END | 13939 |
| BROADWAY | 6449  |
| 3 AVENUE | 6077  |
| 78 ROAD  | 6030  |

Other values (9243) 863487

[Toggle details](#)

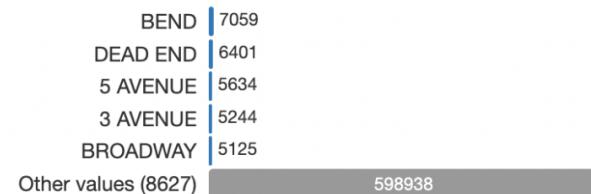
### intersection\_street\_1

Categorical

HIGH CARDINALITY

MISSING

|              |         |
|--------------|---------|
| Distinct     | 8632    |
| Distinct (%) | 1.4%    |
| Missing      | 339233  |
| Missing (%)  | 35.1%   |
| Memory size  | 7.4 MiB |



[Toggle details](#)

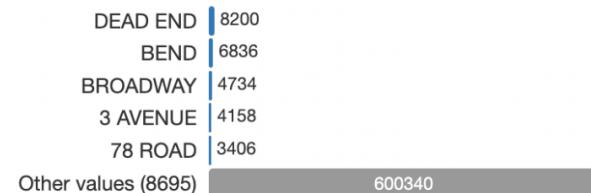
### intersection\_street\_2

Categorical

HIGH CARDINALITY

MISSING

|              |         |
|--------------|---------|
| Distinct     | 8700    |
| Distinct (%) | 1.4%    |
| Missing      | 339960  |
| Missing (%)  | 35.1%   |
| Memory size  | 7.4 MiB |



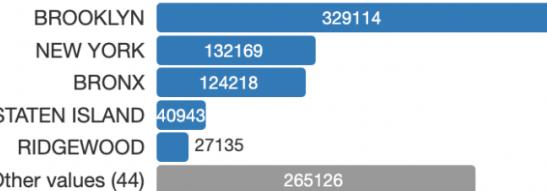
[Toggle details](#)

**city**

Categorical

**HIGH CORRELATION****HIGH CORRELATION****MISSING**

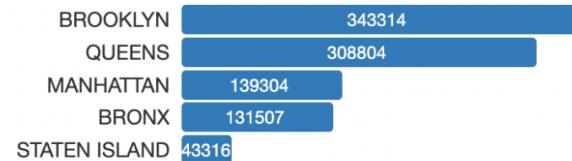
|                     |         |
|---------------------|---------|
| <b>Distinct</b>     | 49      |
| <b>Distinct (%)</b> | < 0.1%  |
| <b>Missing</b>      | 48929   |
| <b>Missing (%)</b>  | 5.1%    |
| <b>Memory size</b>  | 7.4 MiB |

**Toggle details****borough**

Categorical

**HIGH CORRELATION****HIGH CORRELATION**

|                     |         |
|---------------------|---------|
| <b>Distinct</b>     | 6       |
| <b>Distinct (%)</b> | < 0.1%  |
| <b>Missing</b>      | 0       |
| <b>Missing (%)</b>  | 0.0%    |
| <b>Memory size</b>  | 7.4 MiB |

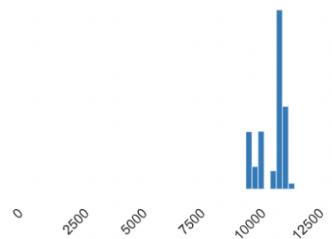
**Toggle details**

**incident\_zip**Real number ( $\mathbb{R}_{\geq 0}$ )

HIGH CORRELATION  
 HIGH CORRELATION  
 HIGH CORRELATION  
 HIGH CORRELATION

|                     |             |
|---------------------|-------------|
| <b>Distinct</b>     | 230         |
| <b>Distinct (%)</b> | < 0.1%      |
| <b>Missing</b>      | 1353        |
| <b>Missing (%)</b>  | 0.1%        |
| <b>Infinite</b>     | 0           |
| <b>Infinite (%)</b> | 0.0%        |
| <b>Mean</b>         | 10949.61138 |

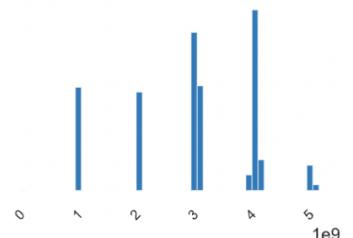
|                     |         |
|---------------------|---------|
| <b>Minimum</b>      | 83      |
| <b>Maximum</b>      | 13207   |
| <b>Zeros</b>        | 0       |
| <b>Zeros (%)</b>    | 0.0%    |
| <b>Negative</b>     | 0       |
| <b>Negative (%)</b> | 0.0%    |
| <b>Memory size</b>  | 7.4 MiB |

[Toggle details](#)**bbl**Real number ( $\mathbb{R}_{\geq 0}$ )

HIGH CORRELATION  
 HIGH CORRELATION  
 HIGH CORRELATION  
 HIGH CORRELATION  
 MISSING

|                     |            |
|---------------------|------------|
| <b>Distinct</b>     | 185871     |
| <b>Distinct (%)</b> | 23.0%      |
| <b>Missing</b>      | 158872     |
| <b>Missing (%)</b>  | 16.4%      |
| <b>Infinite</b>     | 0          |
| <b>Infinite (%)</b> | 0.0%       |
| <b>Mean</b>         | 3018472333 |

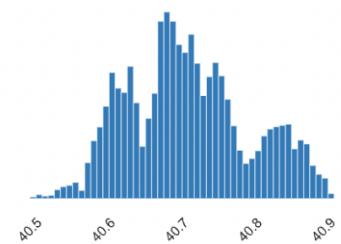
|                     |            |
|---------------------|------------|
| <b>Minimum</b>      | 0          |
| <b>Maximum</b>      | 5270000501 |
| <b>Zeros</b>        | 132        |
| <b>Zeros (%)</b>    | < 0.1%     |
| <b>Negative</b>     | 0          |
| <b>Negative (%)</b> | 0.0%       |
| <b>Memory size</b>  | 7.4 MiB    |

[Toggle details](#)

**latitude**Real number ( $\mathbb{R}_{\geq 0}$ )**HIGH CORRELATION**  
**HIGH CORRELATION**  
MISSING

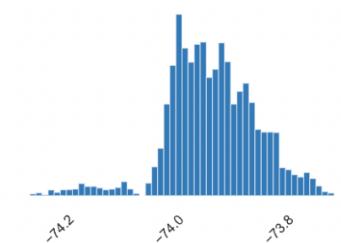
|                     |             |
|---------------------|-------------|
| <b>Distinct</b>     | 282547      |
| <b>Distinct (%)</b> | 29.5%       |
| <b>Missing</b>      | 10002       |
| <b>Missing (%)</b>  | 1.0%        |
| <b>Infinite</b>     | 0           |
| <b>Infinite (%)</b> | 0.0%        |
| <b>Mean</b>         | 40.71507064 |

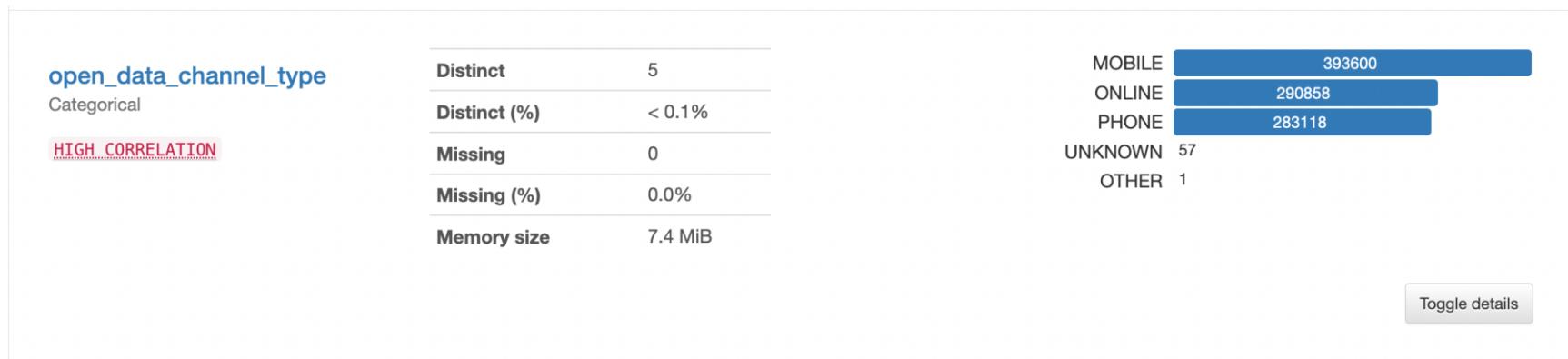
|                     |             |
|---------------------|-------------|
| <b>Minimum</b>      | 40.49912144 |
| <b>Maximum</b>      | 40.91345653 |
| <b>Zeros</b>        | 0           |
| <b>Zeros (%)</b>    | 0.0%        |
| <b>Negative</b>     | 0           |
| <b>Negative (%)</b> | 0.0%        |
| <b>Memory size</b>  | 7.4 MiB     |

[Toggle details](#)**longitude**Real number ( $\mathbb{R}$ )**HIGH CORRELATION**  
**HIGH CORRELATION**  
MISSING

|                     |              |
|---------------------|--------------|
| <b>Distinct</b>     | 282543       |
| <b>Distinct (%)</b> | 29.5%        |
| <b>Missing</b>      | 10002        |
| <b>Missing (%)</b>  | 1.0%         |
| <b>Infinite</b>     | 0            |
| <b>Infinite (%)</b> | 0.0%         |
| <b>Mean</b>         | -73.92353761 |

|                     |              |
|---------------------|--------------|
| <b>Minimum</b>      | -74.25453162 |
| <b>Maximum</b>      | -73.70059685 |
| <b>Zeros</b>        | 0            |
| <b>Zeros (%)</b>    | 0.0%         |
| <b>Negative</b>     | 957632       |
| <b>Negative (%)</b> | 99.0%        |
| <b>Memory size</b>  | 7.4 MiB      |

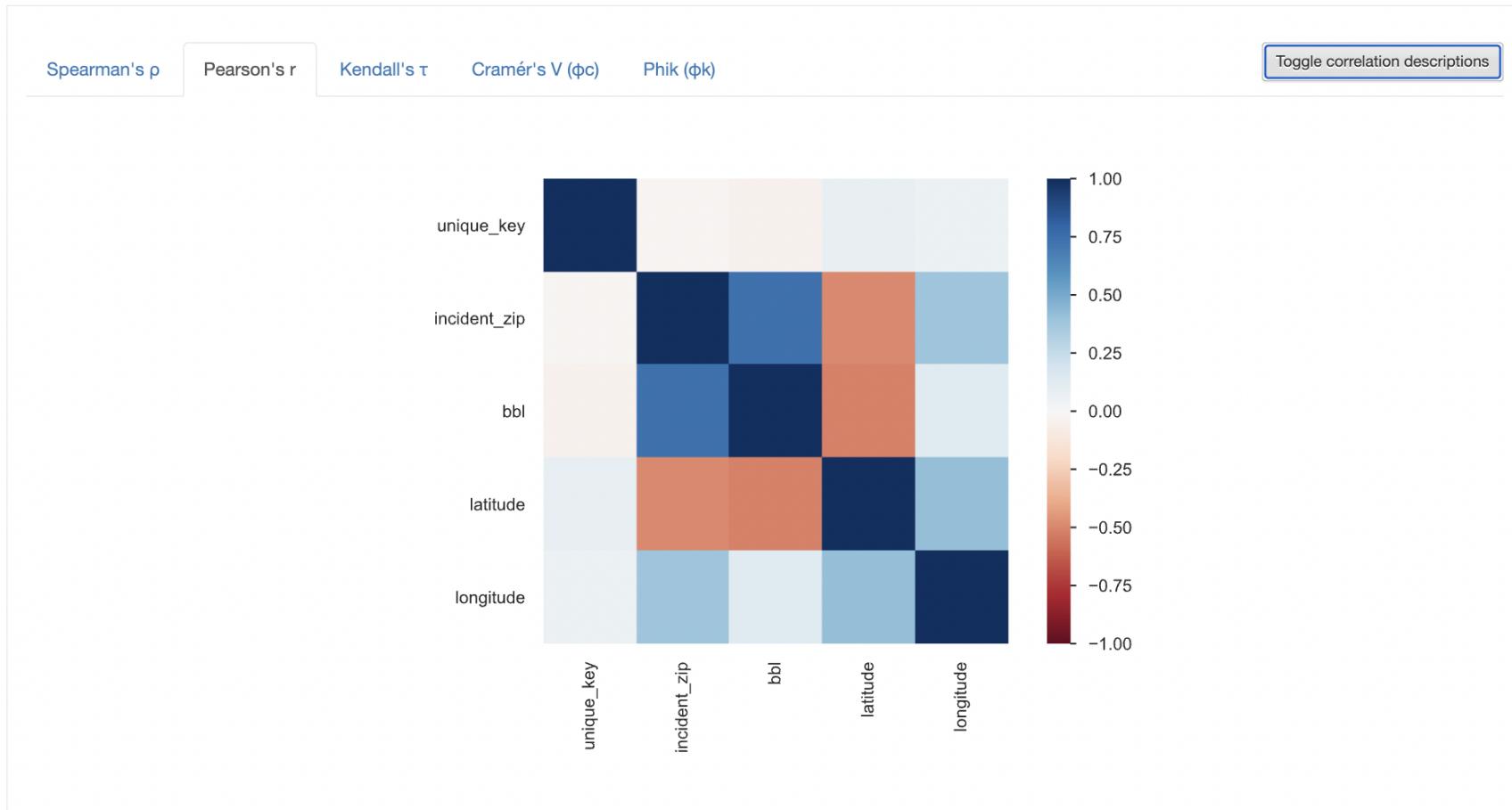
[Toggle details](#)



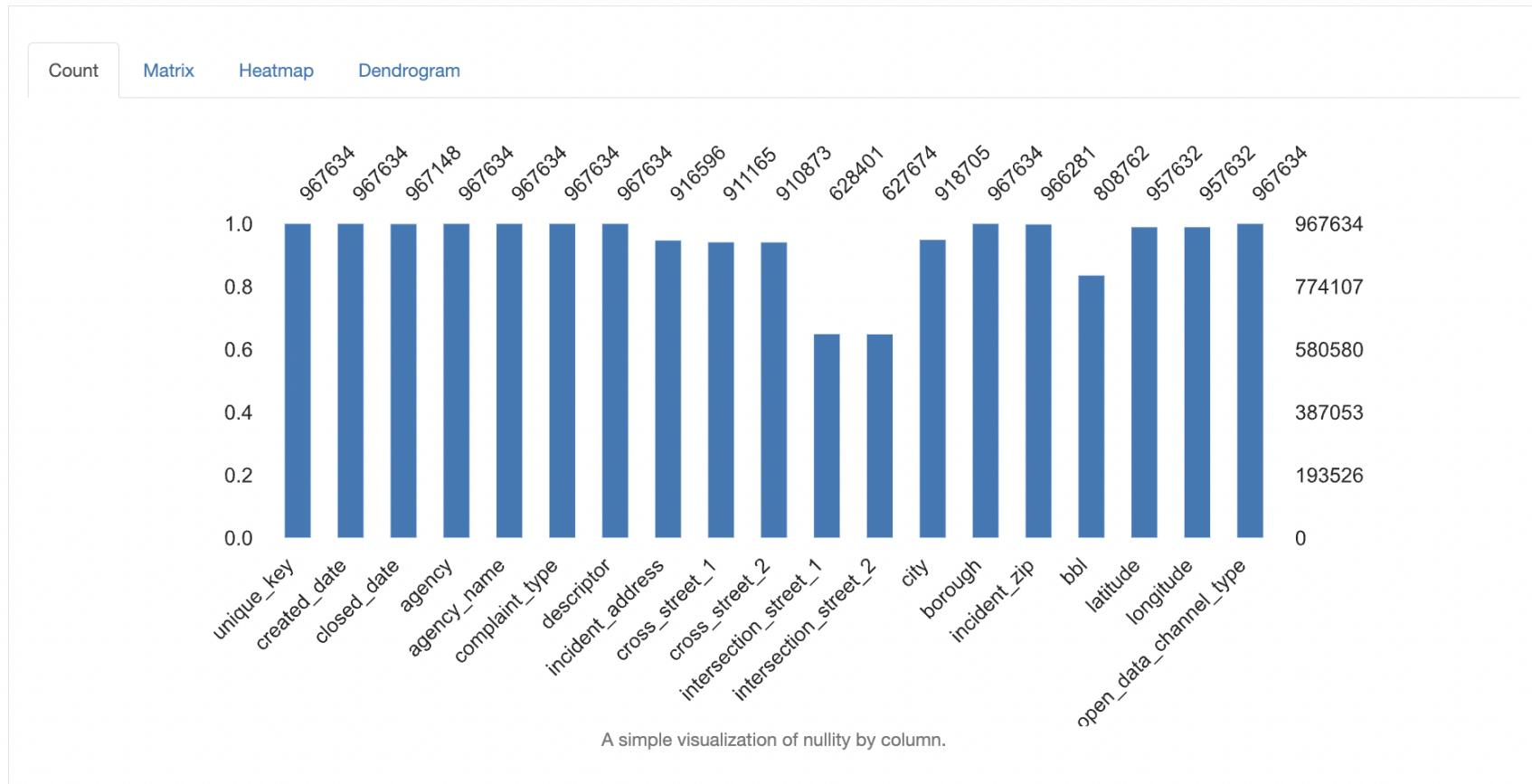
## Interactions



# Correlations



# Missing values



**Narrative conclusion:**

a) the software and database tools the group used to coordinate and manage the project as well as carry out the programming tasks (list of bullet points with software or service and one sentence of what it was used for)

- **Google BigQuery:** We extracted data from datasets and connected to DBT.
- **Jupyter Notebook:** We created a data profile for both datasets.
- **Tableau:** We created a data visualization for both datasets and created a dashboard.
- **DBT Tools:** We created models and fact tables for both datasets to transport to Tableau.

b) the group's experience with the project (which steps were the most difficult? Which were the easiest? what did you learn that you did not imagine you would have? if you had to do it all over again, what would you have done differently?)

- ETL modeling was the most difficult- converting time data from string in the Open Camera and Violations data set
- Visualization was the easiest
- Learned to work with new tools like BigQuery and dbt, learned to work through problems as a group, data profiling to get an idea of what exactly is in the dataset
- Learned the intricacies of ETL pipeline, realized how these pipelines are built and what that would look like in a business setting
- Spent more time on ETL modeling
- We realized that the data in nyc dataset is scarce, and recognized that we should work on snapshot grain instead of transactional grain for time efficiency.

c) if the proposed benefits can be realized by the new system

- By cleaning data, we provide accessible and easy to read information for drivers and local government officials. Through this project, we were able to provide insight about the faulty infrastructure on elevated police presence and the effect it can have on violations.

d) any final comments and conclusions

Overall, we enjoyed and learned a lot through this semester for the group project. Definitely working on the assignments prior to working on project milestones are very helpful in gaining the understanding of the project.