CIS 4400 - CMWA

311 Complaint: Illegal Parking

Group 6

Samuel Salzman (samuel.salzman@baruchmail.cuny.edu)
Amy Yeung (amy.yeung@baruchmail.cuny.edu)
Brian Li (brian.li3@baruchmail.cuny.edu)
Eun Joo Choi (eunjoo.choi1@baruchmail.cuny.edu)
Dayeong Min (dayeong.min@baruchmail.cuny.edu)

Narrative Description:

With the increasing amount of problems related to illegal parking throughout the New York City area, we chose illegal parking for our 311 issue complaint. According to the NYC 311 data from 2017, illegal parking is in the top three of the most common complaints, with 146,240 complaints for the year. Illegal parking is a major issue in NYC, which often leads to multiple police-issued tickets, increased traffic, and increased accidents.

In addition, we believe that the weather conditions have a great impact on the amount of complaints issued for illegal parking, and will integrate well with our data on illegal parking. For example, if it is snowing or raining outside, there is a good chance that drivers would park their cars in an illegal spot temporarily so that they can run into the store or building that they need to go into, instead of finding a legal spot a couple of blocks away and then having to walk in the bad weather. It is going to be interesting to work with these datasets and see what the relationship between them is, and how we can build a successful data warehouse.

KPIs:

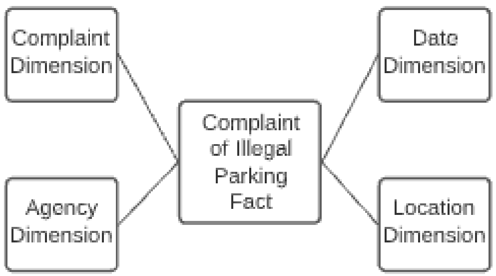
1. Frequency of complaints per day
2. Frequency of complaints per weather type
3. Frequency of complaints per weather condition
4. Number of complaints per city, per neighborhood
5. Number of complaints per type of complaint description (Ex: Blocking Driveway)

Dataset Sources:

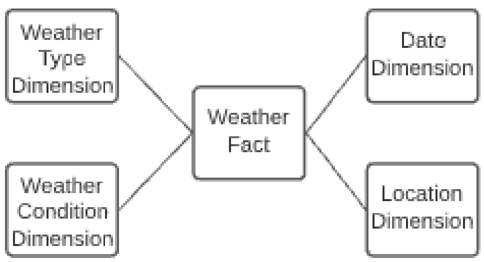
Illegal Parking Dataset

Weather Conditions Dataset

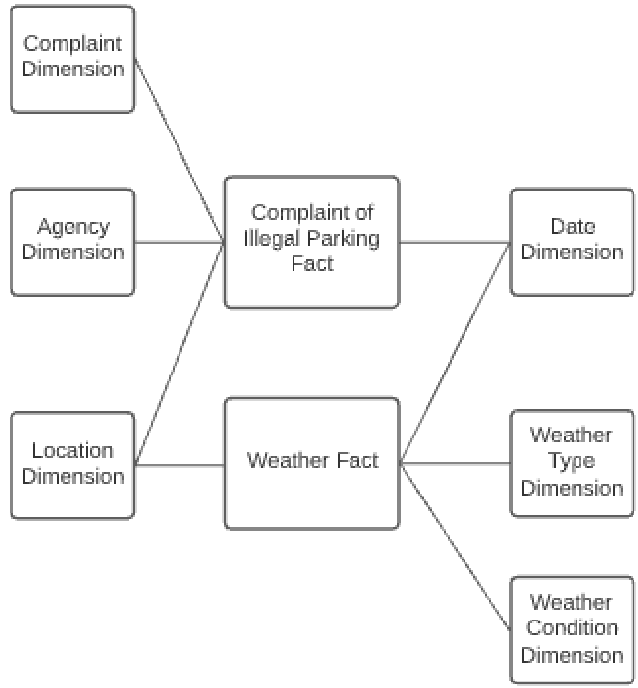
Illegal Parking Data Mart



Weather Data Mart



Enterprise Data Warehouse



Complaint (Complaint_dim_id, Complaint_type, Descriptor)

Date (Date_dim_id, Full_Date, Day_of_Week, Week_day, Day_name, Day_is_weekday, Month, Month_Name, Year, Year_week, Year_day, Fiscal_year, Fiscal_quarter)

Location (Location_dim_id, City, Borough, ZipCode, Latitude, Longitude)

Agency (Agency_dim_id, Agency, Agency_name)

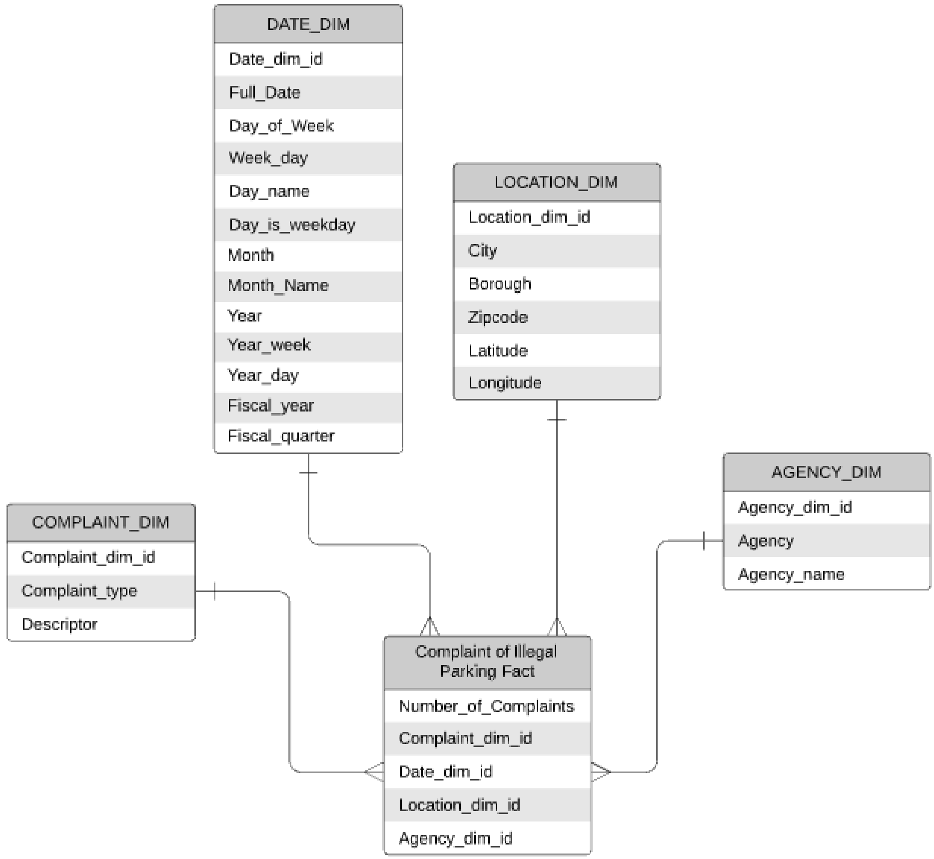
Complaint of Illegal Parking Fact (Number_of_Complaints, Complaint_dim_id, Date_dim_id, Location_dim_id, Agency_dim_id)

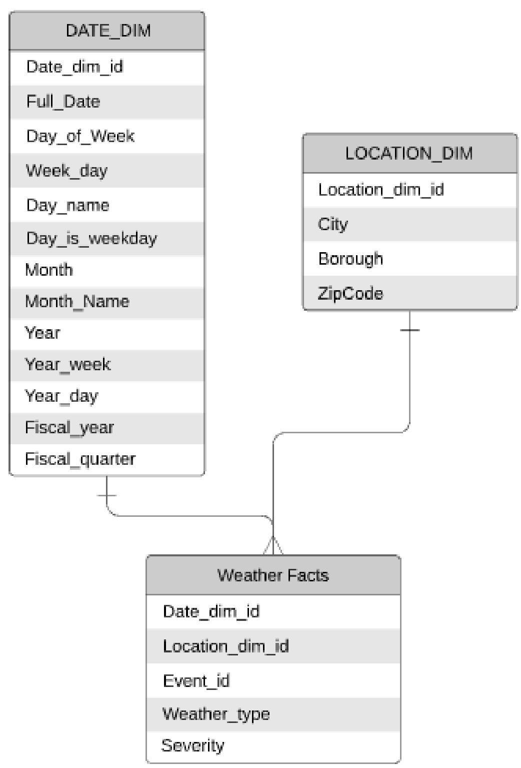
Date (Date_dim_id, Full_Date, Day_of_Week, Week_day, Day_name, Day_is_weekday, Month, Month_Name, Year, Year_week, Year_day, Fiscal_year, Fiscal_quarter)

Location (Location_dim_id, City, Borough, ZipCode)

Weather Fact (Date_dim_id, Location_dim_id, Event_id, Weather_type, Severity)

Dimensional Model:


DATE_DIM
Date_dim_id
Full_Date
Day_of_Week
Week_day
Day_name
Day_is_weekday
Month
Month_Name
Year
Year_week
Year_day
Fiscal_year
Fiscal_quarter
LOCATION_DIM
Location_dim_id
City
Borough
Zipcode
Latitude
Longitude
AGENCY_DIM
Agency_dim_id
Agency
Agency_name
COMPLAINT_DIM
Complaint_dim_id
Complaint_type
Descriptor
Complaint of Illegal Parking Fact
Number_of_Complaints
Complaint_dim_id
Date_dim_id
Location_dim_id
Agency_dim_id


DATE_DIM
Date_dim_id
Full_Date
Day_of_Week
Week_day
Day_name
Day_is_weekday
Month
Month_Name
Year
Year_week
Year_day
Fiscal_year
Fiscal_quarter
LOCATION_DIM
Location_dim_id
City
Borough
ZipCode
Weather Facts
Date_dim_id
Location_dim_id
Event_id
Weather_type
Severity

Agency Dimension Table

```
SELECT ROW_NUMBER() OVER() as agency_dim_id, agency, agency_name
FROM

(SELECT DISTINCT agency, agency_name

    from `focused-history-305115.CIS4400_Project.Illegal_parking_2018-2020`
)
```

agency_dim_id	agency	agency_name
1	NYPD	New York City Police Department
2	NYPD	Traffic Management Center

This is the agency dimension that is a part of our Complaint of Illegal Parking fact table. In this set of SQL code, we are building our model in DBT Cloud, our ETL software, in order to select the two distinct agencies, agency names, and give each output a surrogate key. We selected our data from the data warehouse in Google BigQuery.

Complaint Dimension Table

```
SELECT ROW_NUMBER() OVER() as complaint_dim_id, complaint_type, descriptor
FROM

(SELECT DISTINCT complaint_type, descriptor

    from `focused-history-305115.CIS4400_Project.Illegal_parking_2018-2020`
)
```

complaint_dim_id	complaint_type	descriptor
1	Illegal Parking	Blocked Hydrant
2	Illegal Parking	Blocked Sidewalk
3	Illegal Parking	Detached Trailer
4	Illegal Parking	Blocked Bike Lane
5	Illegal Parking	Unauthorized Bus Layover
6	Illegal Parking	Parking Permit Improper Use
7	Illegal Parking	Commercial Overnight Parking
8	Illegal Parking	Overnight Commercial Storage
9	Illegal Parking	Posted Parking Sign Violation
10	Illegal Parking	Double Parked Blocking Traffic
11	Illegal Parking	Double Parked Blocking Vehicle

This is the complaint dimension that is a part of our Complaint of Illegal Parking fact table. In this set of SQL code, we are building our model in DBT Cloud, our ETL software, in order to select the eleven distinct complaint types, descriptors, and give each output a surrogate key. We selected our data from the data warehouse in Google BigQuery.

Complaint Date Dimension Table

```
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 year_week,
EXTRACT(DAY FROM d) AS year_day,
EXTRACT(YEAR FROM d) AS fiscal_year,
FORMAT_DATE('%Q', d) as fiscal_qtr,
EXTRACT(MONTH FROM d) AS month,
FORMAT_DATE('%B', d) as month_name,
FORMAT_DATE('%w', d) AS week_day,
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('2015-01-01', '2023-01-01', INTERVAL 1 DAY) ) AS d )
```

date_dim_id	date_integer	full_date	year	year_week	year_day	fiscal_year	fiscal_qtr	month	month_name	week_day	day_name	day_is_weekday
1	20150101	2015-01-01	2015	0	1	2015	1	1	January	4	Thursday	1
2	20150102	2015-01-02	2015	0	2	2015	1	1	January	5	Friday	1
3	20150103	2015-01-03	2015	0	3	2015	1	1	January	6	Saturday	0
4	20150104	2015-01-04	2015	1	4	2015	1	1	January	0	Sunday	0
5	20150105	2015-01-05	2015	1	5	2015	1	1	January	1	Monday	1
6	20150106	2015-01-06	2015	1	6	2015	1	1	January	2	Tuesday	1
7	20150107	2015-01-07	2015	1	7	2015	1	1	January	3	Wednesday	1
8	20150108	2015-01-08	2015	1	8	2015	1	1	January	4	Thursday	1
9	20150109	2015-01-09	2015	1	9	2015	1	1	January	5	Friday	1
10	20150110	2015-01-10	2015	1	10	2015	1	1	January	6	Saturday	0

This is the date dimension that is a part of our Complaint of Illegal Parking fact table. In this set of SQL code, we are building our model in DBT Cloud, our ETL software, in order to select various outputs for looking at the dates of all of the complaints, and give each output a surrogate key. We selected our data from the data warehouse in Google BigQuery.

Complaint Location Dimension Table

```
SELECT ROW_NUMBER() OVER() as location_dim_id, city, borough, incident_zip,
latitude, longitude
FROM

(SELECT DISTINCT city, borough, incident_zip, latitude, longitude

    from `focused-history-305115.CIS4400_Project.Illegal_parking_2018-2020`
)
```

location_dim_id	city	borough	incident_zip	latitude	longitude
1	NULL	Unspecified	NULL	NULL	NULL
2	NEW YORK	MANHATTAN	10001	40.75284936	-74.00287301
3	NEW YORK	MANHATTAN	10001	40.75540475	-74.00111171
4	NEW YORK	MANHATTAN	10001	40.75304146	-74.00397026
5	NEW YORK	MANHATTAN	10001	40.75544591	-74.00221981
6	NEW YORK	MANHATTAN	10001	40.75294542	-74.0030968
7	NULL	MANHATTAN	10001	40.74970635	-73.99156182
8	NEW YORK	MANHATTAN	10001	40.74865481	-73.98809723

This is the location dimension that is a part of our Complaint of Illegal Parking fact table. In this set of SQL code, we are building our model in DBT Cloud, our ETL software, in order to select the information about each complaint's location, including city, borough, zip code, latitude and longitude, and give each output a surrogate key. These results were essential in forming a lot of our KPIs. We selected our data from the data warehouse in Google BigQuery.

Weather Date Dimension Table

```
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 year_week,
EXTRACT(DAY FROM d) AS year_day,
EXTRACT(YEAR FROM d) AS fiscal_year,
FORMAT_DATE('%Q', d) as fiscal_qtr,
EXTRACT(MONTH FROM d) AS month,
FORMAT_DATE('%B', d) as month_name,
FORMAT_DATE('%w', d) AS week_day,
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('2015-01-01', '2023-01-01', INTERVAL 1 DAY) ) AS d )
```

date_dim_id	date_integer	full_date	year	year_week	year_day	fiscal_year	fiscal_qtr	month	month_name	week_day	day_name	day_is_weekday
1	20150101	2015-01-01	2015	0	1	2015	1	1	January	4	Thursday	1
2	20150102	2015-01-02	2015	0	2	2015	1	1	January	5	Friday	1
3	20150103	2015-01-03	2015	0	3	2015	1	1	January	6	Saturday	0
4	20150104	2015-01-04	2015	1	4	2015	1	1	January	0	Sunday	0
5	20150105	2015-01-05	2015	1	5	2015	1	1	January	1	Monday	1
6	20150106	2015-01-06	2015	1	6	2015	1	1	January	2	Tuesday	1
7	20150107	2015-01-07	2015	1	7	2015	1	1	January	3	Wednesday	1
8	20150108	2015-01-08	2015	1	8	2015	1	1	January	4	Thursday	1
9	20150109	2015-01-09	2015	1	9	2015	1	1	January	5	Friday	1
10	20150110	2015-01-10	2015	1	10	2015	1	1	January	6	Saturday	0

This is the date dimension that is a part of our Weather fact table. In this set of SQL code, we are building our model in DBT Cloud, our ETL software, in order to select various outputs for looking at the dates of all of the complaints, and give each output a surrogate key. We selected our data from the data warehouse in Google BigQuery.

Weather Location Dimension Table

```
SELECT ROW_NUMBER() OVER() as location_dim_id, city, borough, zipcode
FROM

(SELECT DISTINCT city, borough, zipcode

    from `focused-history-305115.CIS4400_Project.Weather_2018-2020`
)
```

location_dim_id	city	borough	zipcode
1	New York	New York	10024
2	New York	Brooklyn	11204
3	New York	Queens	11377
4	New York	Brooklyn	11231
5	New York	Queens	11433
6	New York	Bronx	10457
7	New York	Queens	11379
8	New York	Brooklyn	11214

This is the location dimension that is a part of our Weather fact table. In this set of SQL code, we are building our model in DBT Cloud, our ETL software, in order to select the information about each complaint's location, including city, borough, zip code, latitude and longitude, and give each output a surrogate key. These results were essential in forming a lot of our KPIs. We selected our data from the data warehouse in Google BigQuery.

Complaint Fact Table

```sql
SELECT unique_key, agency_dim_id, complaint_dim_id, location_dim_id, date_dim_id

FROM `focused-history-305115.CIS4400_Project.Illegal_parking_2018-2020` AS sr

           INNER JOIN {{ref ('stg_agency')}} AS  agency_dim ON agency_dim.agency
= sr.agency AND agency_dim.agency_name = sr.agency_name
           INNER JOIN {{ref ('stg_complaint')}} AS complaint_dim ON
complaint_dim.complaint_type = sr.complaint_type AND complaint_dim.descriptor =
sr.descriptor
           INNER JOIN {{ref ('stg_complaint_date')}} AS date_dimension  ON
date_dimension.full_date = sr.created_date
           INNER JOIN {{ref ('stg_complaint_location')}}  AS location_dim  ON
location_dim.city = sr.city AND location_dim.borough = sr.borough AND
location_dim.incident_zip=sr.incident_zip AND location_dim.latitude = sr.latitude
AND location_dim.longitude = sr.longitude
```

unique_key	agency_dim_id	complaint_dim_id	location_dim_id	date_dim_id
42971073	1	1	2	1626
43283353	1	1	3	1656
43512291	1	1	4	1683
43553326	1	1	5	1688
43061238	1	1	6	1635
43281602	1	1	3	1656
43848685	1	1	6	1714
43907999	1	1	8	1731
43992604	1	1	6	1739

This is the Complaint Fact Table. In this model, we join all of the surrogate keys from each of the complaint related models (complaint, date, location, and agency dimensions) in order to make our fact table. We had to make sure to join each of them on their respective columns. After successfully running this code, we were able to produce all of the complaints.

Weather Fact Table

```
SELECT date_dim_id, location_dim_id, eventid, type, severity

FROM `focused-history-305115.CIS4400_Project.Weather_2018-2020` AS wr

          INNER JOIN {{ref ('stg_weather_complaint_date')}} AS date_dimension  ON
date_dimension.full_date = wr.StartTime_UTC_
          INNER JOIN {{ref ('stg_weather_location')}}  AS location_dim  ON
location_dim.city = wr.city AND location_dim.borough = wr.borough AND
location_dim.zipcode=wr.zipcode
```

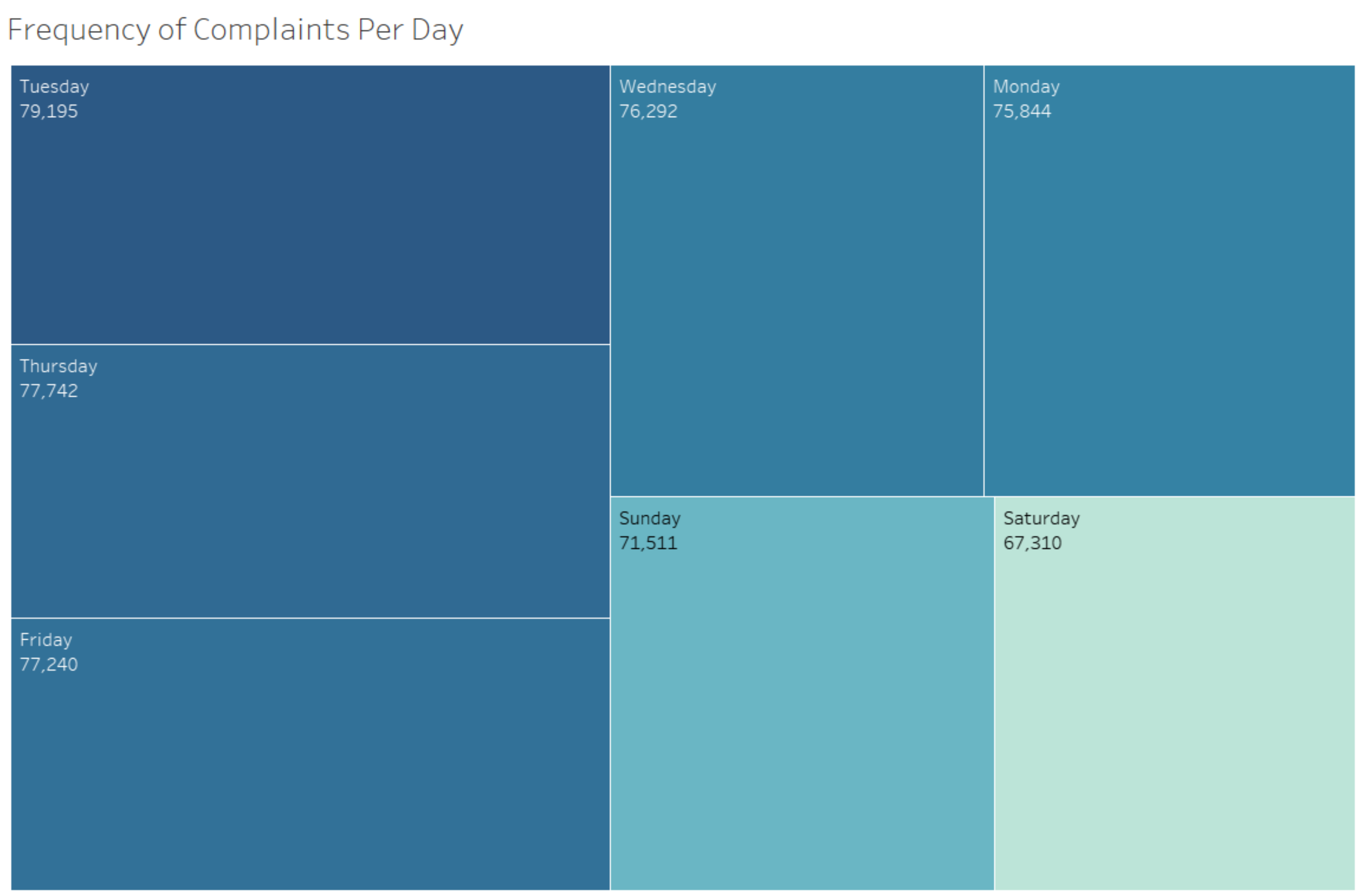
date_dim_id	location_dim_id	eventid	type	severity
1229	1	W-715101	Rain	Light
1229	1	W-715102	Rain	Light
1229	2	W-717612	Rain	Light
1229	2	W-717613	Rain	Light
1229	3	W-720123	Rain	Light
1229	3	W-720124	Rain	Light
1229	4	W-722634	Rain	Light
1229	4	W-722635	Rain	Light
1229	5	W-725145	Rain	Light

This is the Weather Fact Table. In this model, we join all of the surrogate keys from each of the weather related models (date and location dimensions) in order to make our fact table. We had to make sure to join each of them on their respective columns.

Visualization of KPIs

Frequency of complaints per day (Samuel Salzman)

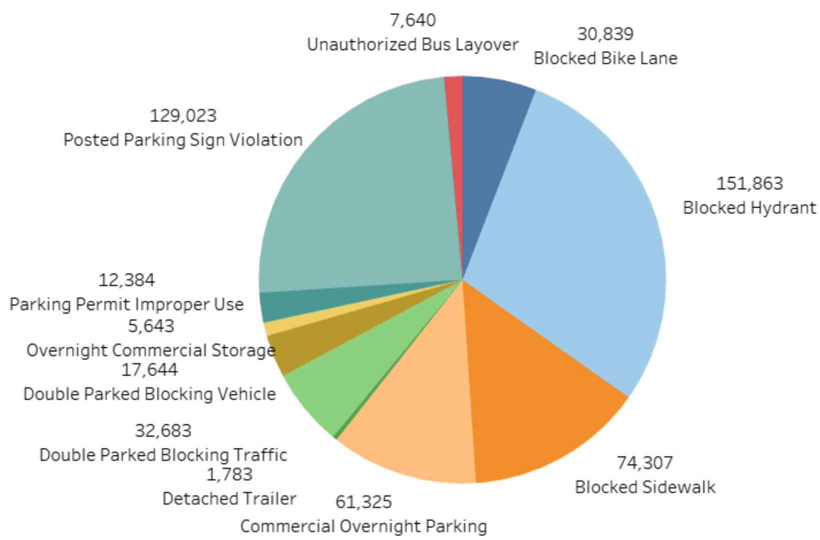
- Here is a KPI visualization showing a breakdown of the frequency of complaints per day. From this visualization, we can see that illegal parking complaints seem to be spread evenly across the 7 days of the week with the most on Tuesday and the least on Saturday with a range of 12,000.



Number of complaints per type of complaint description (Brian Li)

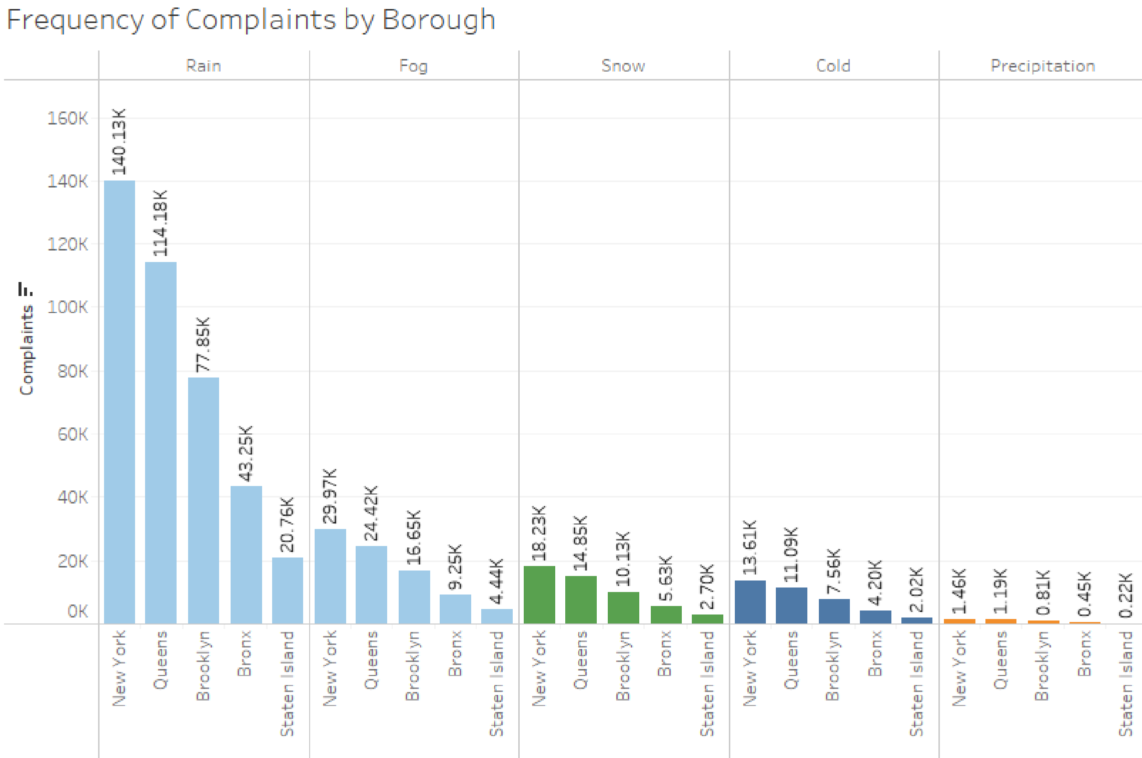
- Here is a KPI visualization showing the distribution of the number of complaints per type of complaint description. There were a total of 11 different complaint descriptions. From the visualization we can see that most of the illegal parking complaints were made for blocked hydrants, posted parking sign violations, followed by blocked sidewalks and commercial overnight parking.

Complaints per Type of Complaint Description



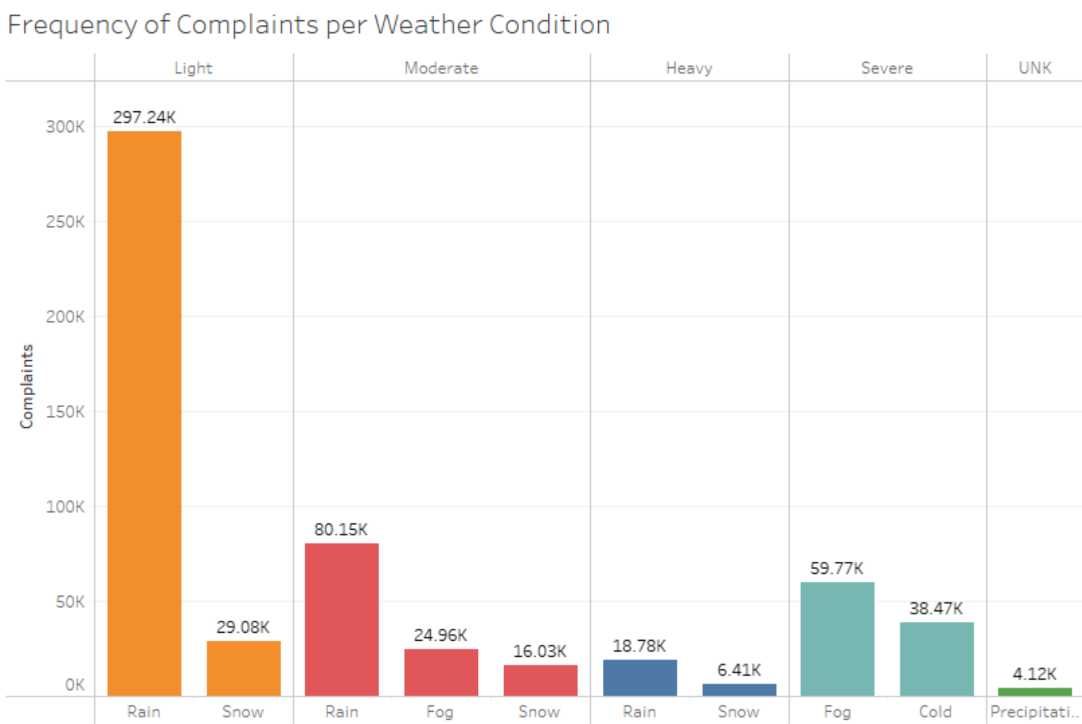
Frequency of complaints by borough per weather type (Amy Yeung)

- Here is a KPI visualization showing the frequency of complaints broken down by each borough, according to each weather type. This was a really interesting KPI to visualize, as it showed a really clear breakdown of the weather's impact on each borough. From here we can see that most of the parking complaints are made in New York (Manhattan), followed by Queens.



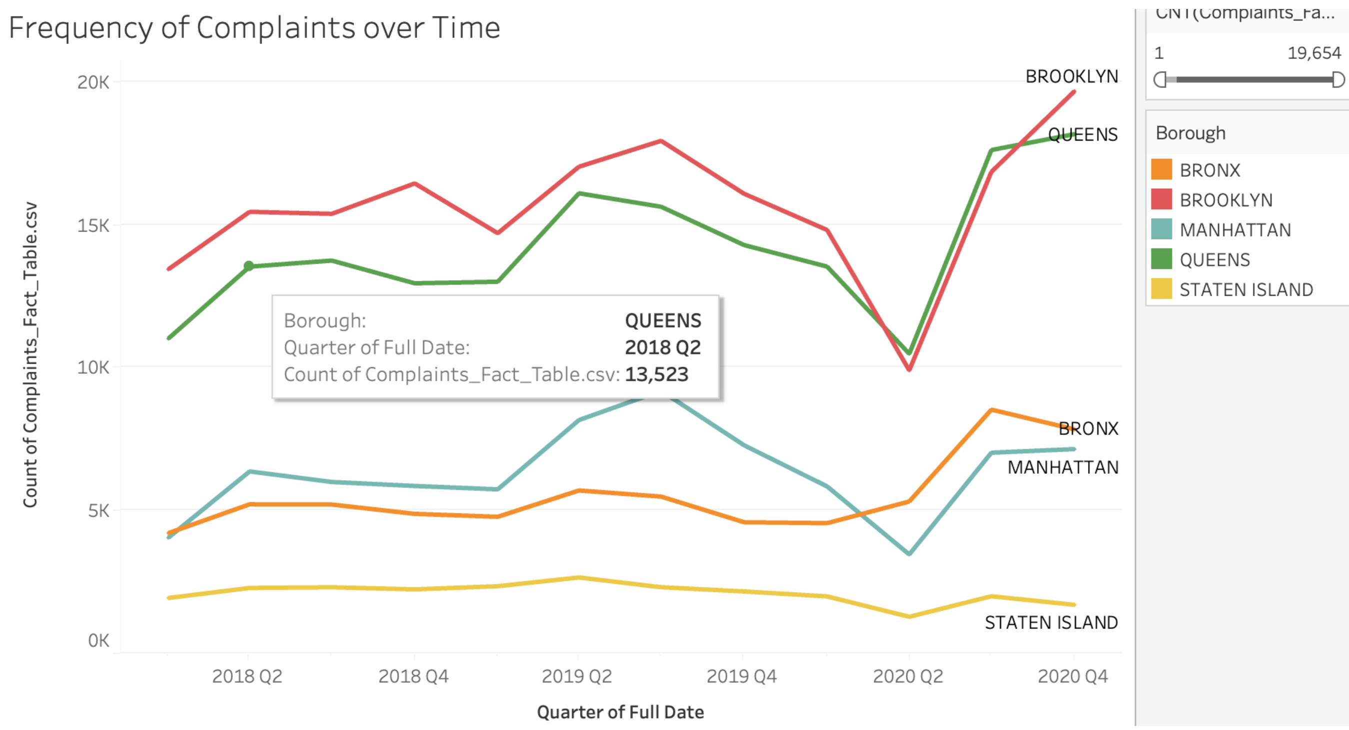
Frequency of complaints per weather condition (Eun Joo Choi)

- Here is a KPI visualization of the frequency of complaints per weather condition. It was interesting to see that the frequency of parking complaints occurs on most light rainy days, followed by moderate rainy days.



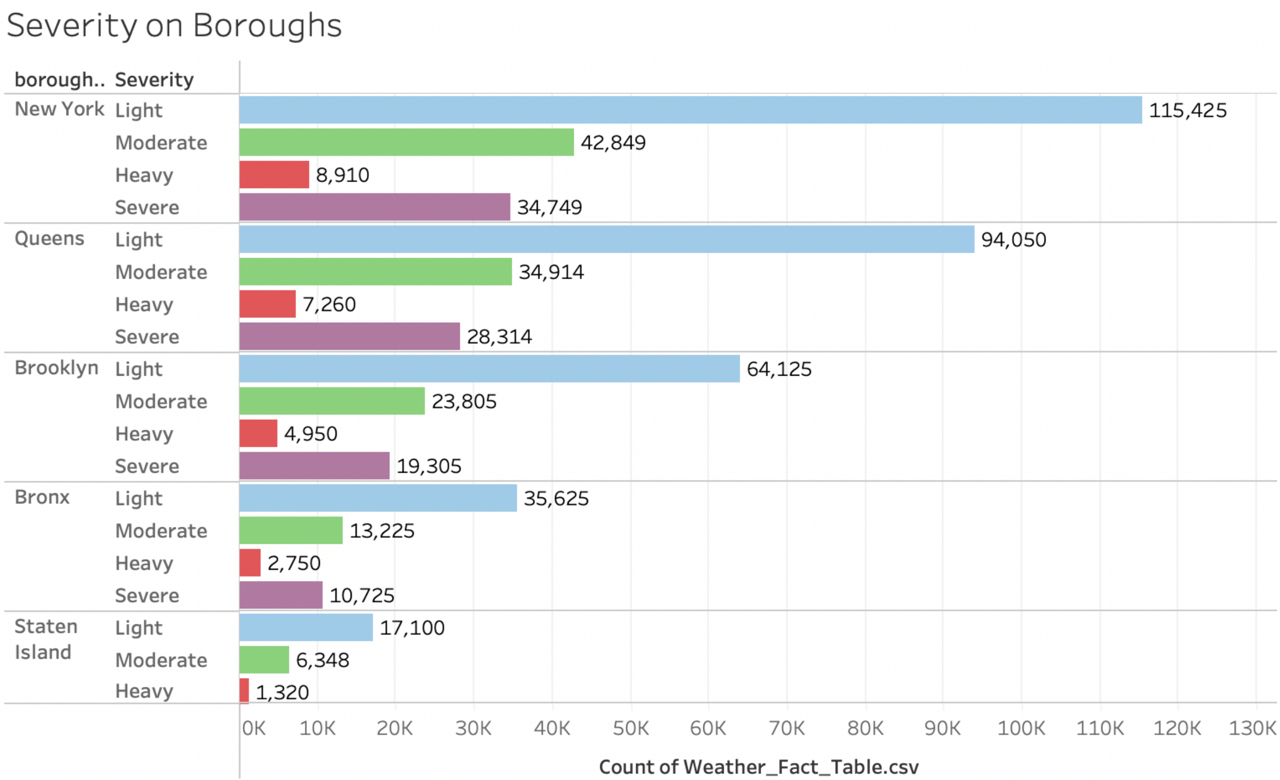
Frequency of Complaints over Time (in Quarters)

- This is a KPI visualization showing the frequency of complaints over time for the three years of our collected data. One interesting point that we witnessed was that the count of complaints was the lowest during the first outbreak of COVID19, and then reached the peak in Quarter 4 of 2020.



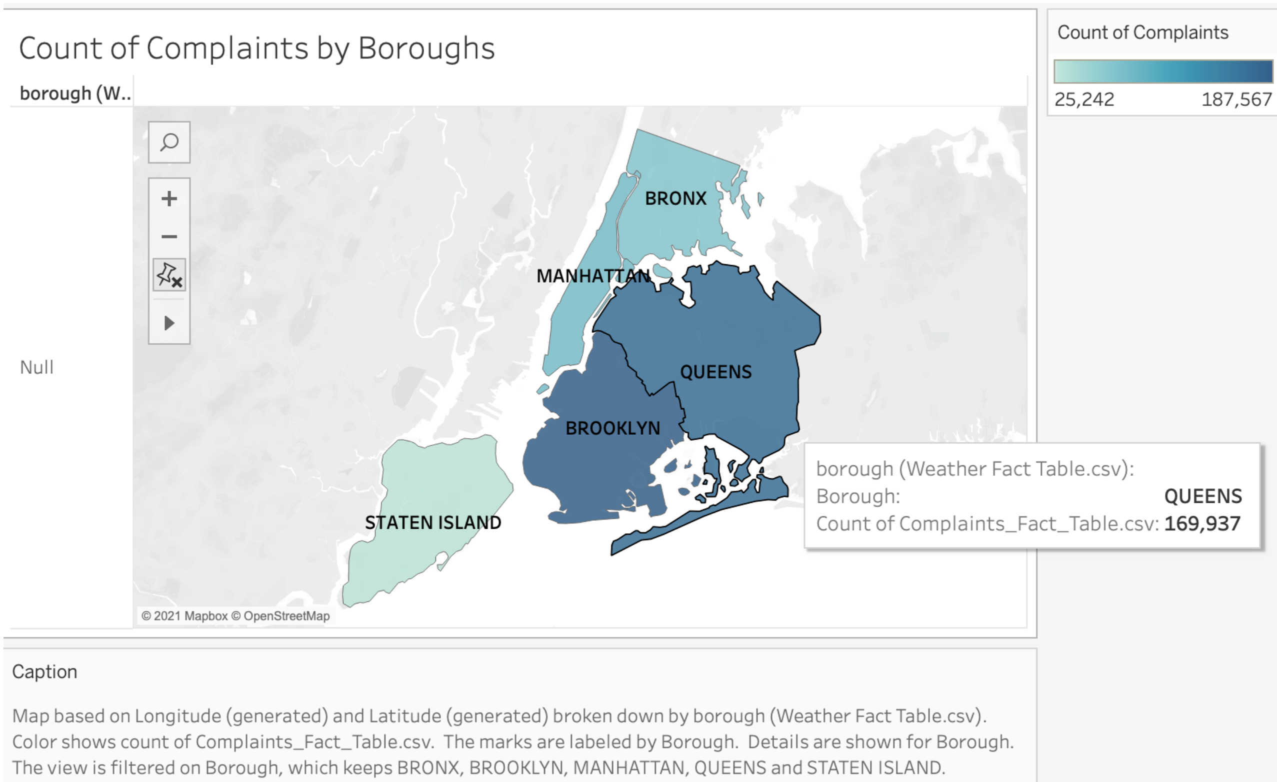
Severity per Boroughs (Dayeong Min)

- Here is a KPI visualization showing the severity of each type of weather condition broken down by borough. The severity level of "Light" is dominant for all boroughs, and frankly, more "Severe" levels are found than "Heavy" levels.



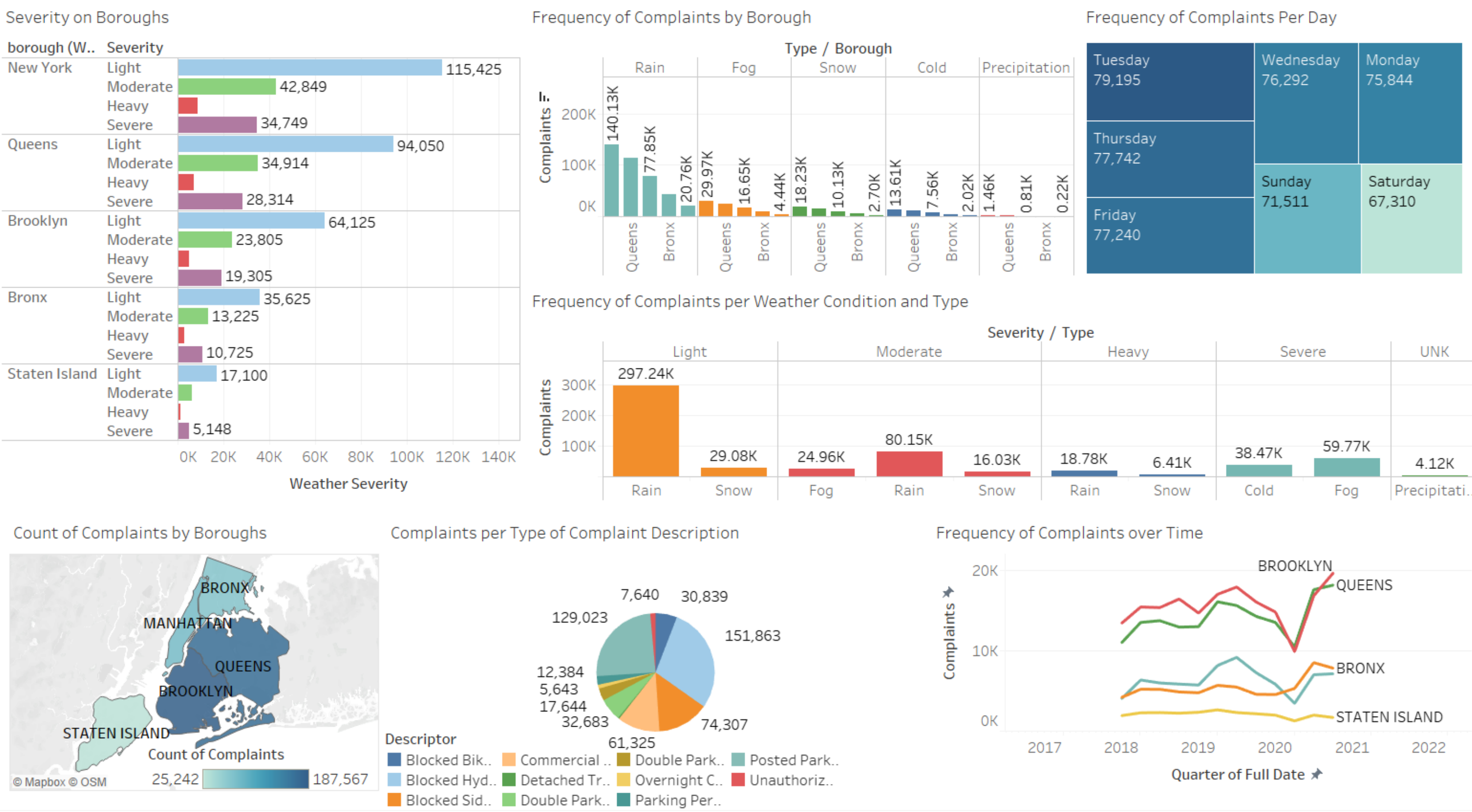
Count of Complaints per Boroughs (Dayeong Min)

- Here is a KPI visualization showing the complaint count broken down by borough. The map is shaded according to the key on the top right. From this visualization, we can see that Brooklyn has the most number of complaints while Staten Island has the least.



Dashboard of KPI Visualizations:

- This is a dashboard that we made through Tableau showing a compilation of all of our KPI visualizations. This is a great way for the business user to see a large amount of different charts that are analyzing the data in their data warehouse.



Conclusion:

Throughout the course of this project, we used three software and database tools to help integrate and conduct our project and programming tasks. First, we used Google BigQuery as our Data Warehouse tool to store our fact tables and dimension models. Next, we used DBT Cloud as our ETL programming tool so that we could extract, transform, and load our data. We did this by creating our dimensional models, and running the dbt code in order to build out our models, which connected with our BigQuery. Finally, we used Tableau as our data visualization tool in order to synthesize our data into visuals to display our KPIs and create our dashboard. We used SQL as our coding language in both BigQuery and DBT Cloud.

While the overall project experience was very rewarding, there were a few challenges that we faced as a group. We had the most difficult time in the ETL process, trying to load data into BigQuery and sharing it with the group. It was a very time consuming process. In addition, while working on the ETL part of the project on DBT Cloud, we had issues with “location not found in the US” as well as other errors when it came to creating the fact tables for both the complaints and the weather. We discussed several times and tested different methods to process the data with the best possible efficiency. Through trial and error, we were able to successfully develop the right code and create output that helped us answer our overall questions, and calculate our KPIs. The easiest part of the project were the KPI visualizations. We used different types of graphs on Tableau to portray our data and explain the KPIs that we calculated.

We mainly communicated through Zoom meetings and an app called “WhatsApp” while working on the project. Our Zoom meetings often involved one person sharing their screen, while the others tried their best to give their input. If we were to do this project over, we would have made changes to the way that we organized the meetings, and distributed the parts accordingly. With everyone on the same page and feeling comfortable with their task, it will make each step of the process easier to complete for everyone in a timely fashion. Over the course of this project, we learned a vast amount about actually creating a data warehouse and the extremely powerful tool that it is. Learning about it in class was very interesting, and putting it to work was simply the cherry on top. We didn’t think that we, as a group, would be able to make such a project, especially due to the fact that this was really new to all of us. It goes to show you what hard work and determination can do. It was also a great experience for us to work as a group, even though there were some challenges with time scheduling and task delegation.

Overall, we believe that we all learned how to take responsibility for our actions, and accept the roles given to us in order to succeed.

While we see a correlation between weather and illegal parking complaints, there is nothing that we can do in order to decrease the amount of complaints. Regardless, we were able to see that there was a correlation between illegal parking complaints and the weather that was present at the time of the complaint. Now that we realized the correlation, the next step would be to create a plan of action to decrease the number of illegal parking complaints. This way, we can use our system to evaluate if there is any significant improvement.

Sources:

Calgary, Open. “311 Service Requests from 2010 to Present.” *NYC Open Data*, data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9/data.

“LSTW: Large-Scale Traffic and Weather Events Dataset.” *Sobhan Moosavi*, 1 Jan. 2021, smoosavi.org/datasets/lstw.