12/2022 - DS 4200 - final project group 7: Mass Ave pollution

Data Report: The Effects of Automobiles on Pollution Within the Mass Ave Community

Mia Huebscher, Ashvika Boopathy, Qianyong Hu

Project Summary

Over recent years, researchers have conducted numerous studies on the effects of pollution on human health. These studies have shown that pollution has a significantly negative impact on not just the physical health of human beings, but also their mental health. In May of 2022, a National Geographic article reported that noise pollution can create problems like heart disease, high blood pressure, stress, sleep disturbances, and poor performance in school for children. Another article published by National Geographic

in 2022 reported the extensive effects of air pollution on human health, among which include heart disease, lung cancer, nerve damage, and respiratory disease.

For this reason, our project aims to analyze the levels of air and noise pollution within the Massachusetts Avenue community (colloquially referred to as the Mass Ave community) - a region filled with vibrant and historical neighborhoods whose problems have largely been teased and overlooked by the rest of Boston's districts. More specifically, our project will serve to explain to viewers the effect that automobile use has on air and noise pollution within the Mass Ave community. By illustrating the impact of vehicles on these types of pollution, our project aims to convince members of the Mass Ave community to decrease their automobile consumption to the best of their ability by utilizing alternative modes of transportation.

Data

As our primary data set, our group utilized data offered by the Boston 311 Program, which reports intricate information on every complaint made by citizens around Boston regarding their communities. To create this data set, we utilized the Boston 311 Program's data from 2019, 2020, and 2021. We chose to employ this data in our project to enable us to analyze the extent to which Mass Ave community members feel the effects of pollution by vehicles.

Data subsets

One of the additional data sets we found reports on the daily air quality index values for areas within Suffolk county. These areas include Von Hillern St, Dudley Square Roxbury, and Boston Kenmore Square. This data comes from the United States Environmental Protection Agency. We downloaded three subsets of this data for the years we are focusing on (2019, 2020, and 2021). This data set has information about the pollutants and their pollutant levels for each day. We chose this data set because we believe it will be useful to display air quality issues in Boston to the Mass Ave community, demonstrate to residents that cars are causing air quality issues, and convince people to take other means of transportation to improve not only their well being but also the environment's.

Another supporting data set we found details the number of public facilities around Boston. These public facilities include parking meters, private parking lots, gas stations, and

so on. The number of public facilities in different areas would reflect the different levels of vehicle use around the city. This data set is also helpful because it can provide us insight into popular car routes in Boston, as people likely drive cars through communities that have a lot of facilities to support their driving. Utilizing the parking meters data will help us determine the different densities of parking meters in different areas. Therefore, we can use this data subset to demonstrate the relation between parking meters and air quality.

To analyze the effects of vehicle use on pollution, we collected data on the various types of transportation that Boston citizens use to travel to work. This commute data will enable us to prove a correlation between vehicle use and pollution because we can show how rises and falls in car use create fluctuations in air quality and automotive noise disturbance complaints. Specifically, this commute data reports the number of households in Boston that employ various transportation modes to get to work.

Data Cleaning Process

To clean our primary 311 Program data set, we had to filter the data so that it only included information for neighborhoods within the Mass Ave community. This task served challenging because we had to carefully judge which neighborhoods to include. Another challenge we encountered occurred when we filtered the data based on the type of complaint that was made, as we had to decide if this complaint could be utilized to provide relevant information on pollution. These decisions took some discussion and time, but ultimately our team managed to filter our data in a way we believe will provide helpful insight to our project.

When cleaning the AQI (Air Quality Index) data, one challenge we encountered was getting strange numbers when trying to find the average PM2.5 levels; we realized the issue was that the data types for the columns were all objects. After examining the types for each data, we concluded that we had to change the data types for all the columns; certain columns such as the site needed to be strings rather than objects, PM 2.5 levels had to be integers, and the date column had to be a DateTime data type. After fixing these details, our code functioned properly. Another part of the data cleaning process for this dataset was removing columns we are not going to use. After deciding to focus on one pollutant (PM2.5 levels) we decided to keep the date column, site name, and PM 2.5 level columns.

The parking assets data set contains some redundant information. For example, more than four columns of this data set are used to represent the same things, location. Some columns contain the same value or empty values. For cleaning up, we want to analyze their pay or no pay periods of parking meters, but the columns for this information contain too much information. For example, this dataset includes all information on pay periods such as price, time slot, and day. It is difficult for us to clean this part or extract certain information contained in this column.

The cleaning of the commute data did not prove too difficult, as we were only required to combine the data from 2019, 2020, and 2021. However, we did struggle to find this data, as we conducted many searches to find data that could effectively represent vehicle use. Ultimately, we believe that this commute data will serve as the most efficient data set to aid in our analysis.

Data Types within Our Datasets

Within our primary 311 Program data set, the categorical data resides in the columns that document the speed in which each complaint was resolved, whether the subject of each complaint has been fixed, the methods used to fix each complaint subject, the subject of each complaint, the reason for each complaint, the type of each complaint, the department responsible for resolving each complaint, photos to visualize the reason behind each complaint, and the method used to report each complaint. The columns of the data that report the public works district, city council district, neighborhood, neighborhood services district, street, and zip code in which each complaint was made also contain categorical data. The quantitative data can be found in the columns that describe the year of each complaint and the latitude-longitude coordinates of each complaint. Lastly, the ordinal data resides in the column that reports the exact date and time each complaint was issued.

For the AQI dataset, the first column (displaying the dates) contains ordinal data, the site names are categorical nominal data, and the PM 2.5 level column is quantitative data.

Our parking facilities data set contains two kinds of data types - quantitative data and categorical data. The quantitative data is the location of each Parking meter and the categorical data represents pay rules and status.

Our commute data set includes four columns, the name of the commute type and location for each data point (categorical data), as well as the year associated with each data point and the number of Boston households using each commute type (quantitative data).

Links to Our Data Sources

311 Program Data Set

https://data.boston.gov/dataset/311-service-requests

Air Quality Index Daily Values Report - Ashvika Boopathy

https://www.epa.gov/outdoor-air-quality-data/air-quality-index-daily-values-report

Parking Meter Data Set - Qianyong Hu

https://data.boston.gov/dataset/parking-meters

Commute Type Data Set - Mia Huebscher

https://data.census.gov/cedsci/table?q=transportation%20in%20Boston%20city,%20Massachusetts&tid=ACSDT1Y2021.B08141

Links to Background Information

PM 2.5

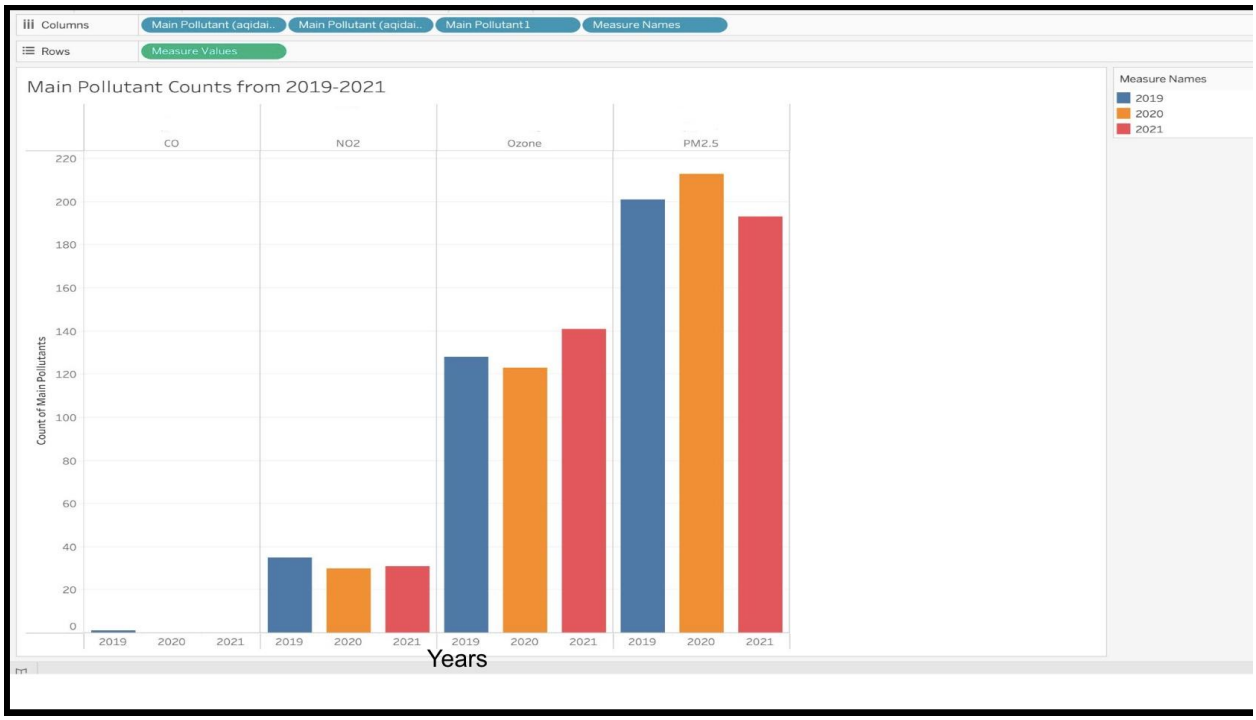
https://www.health.ny.gov/environmental/indoors/air/pmq_a.htm#:~:text=Fine%20particulate%20matter%20

National Geographic Articles on the Effects of Pollution

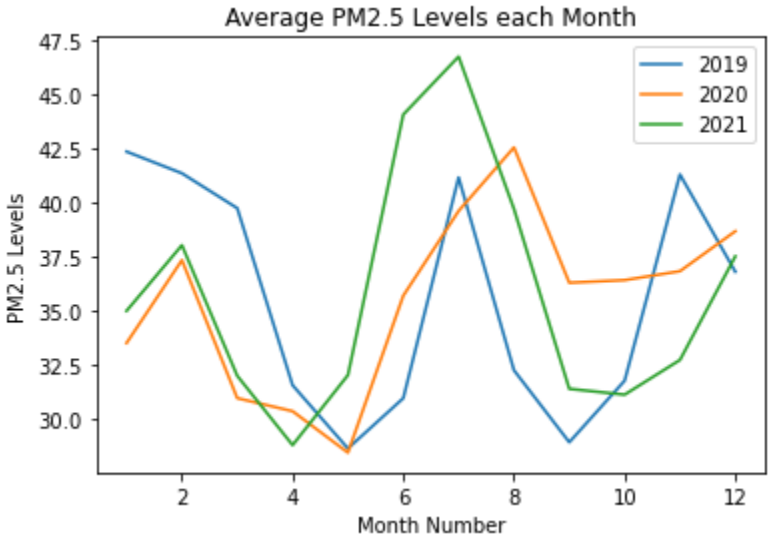
https://education.nationalgeographic.org/resource/noise-pollution

https://education.nationalgeographic.org/resource/air-pollution

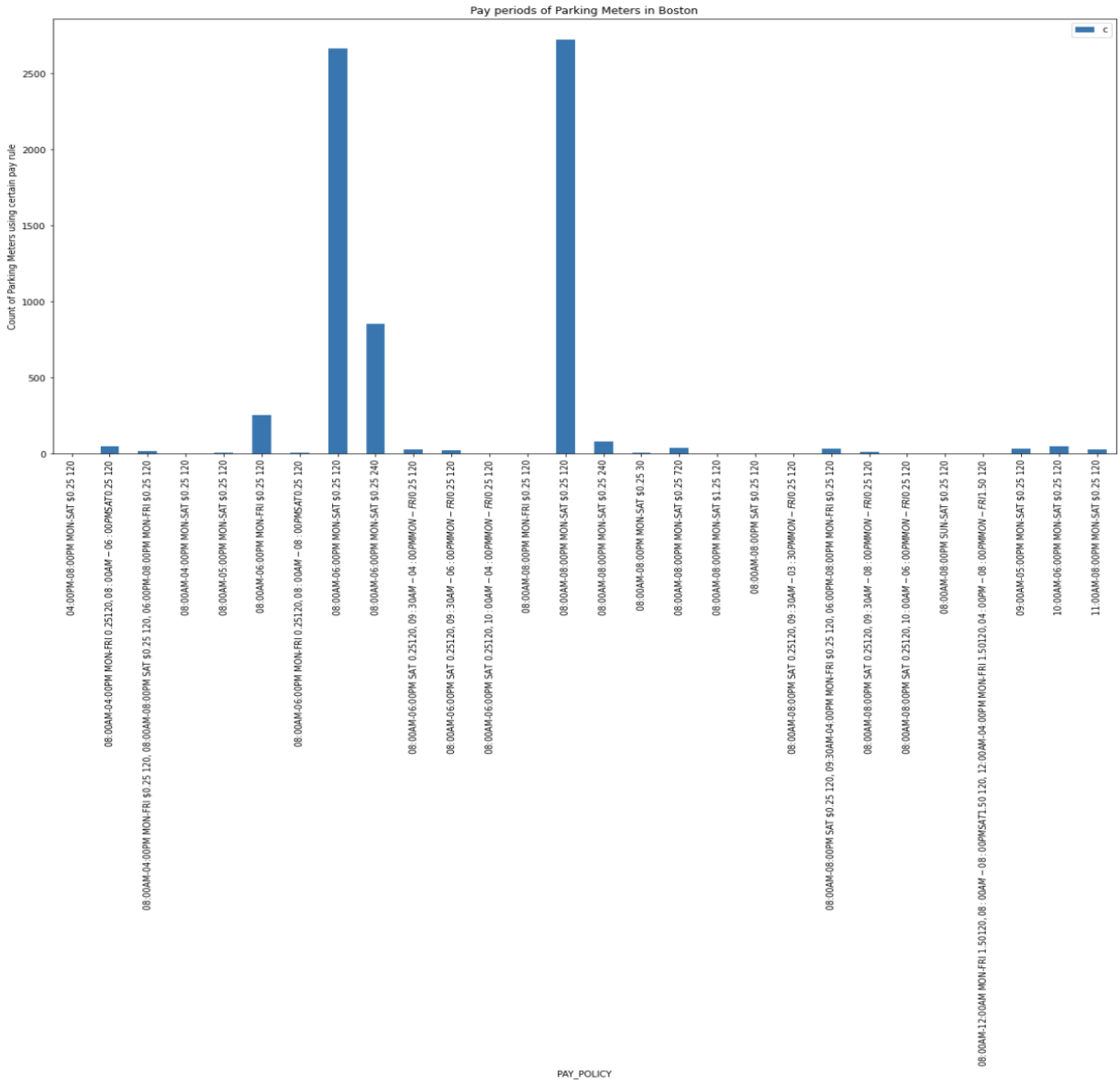
Observations & Insights



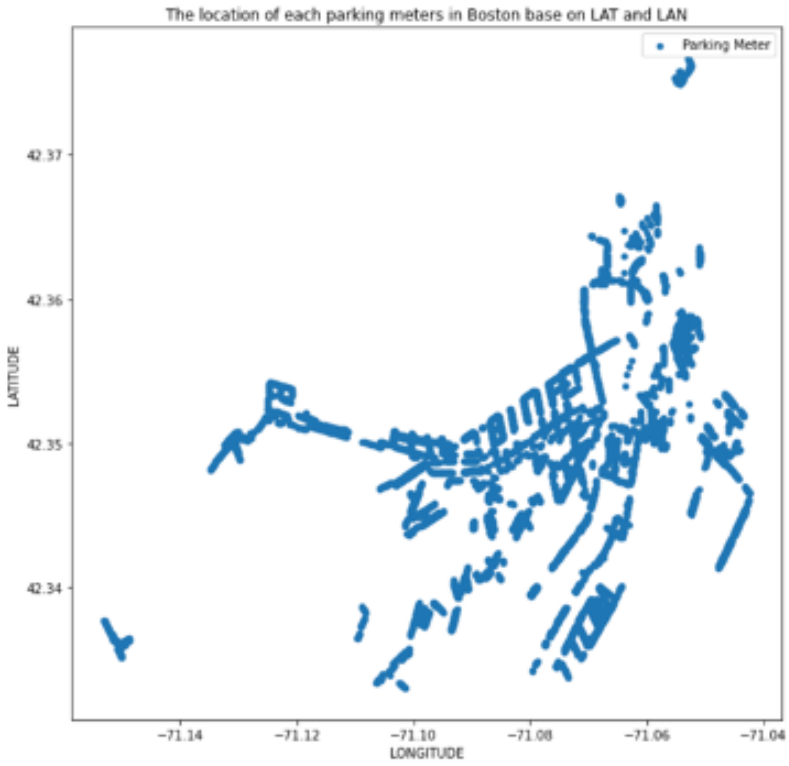
When exploring the AQI dataset, we were interested in learning about the pollutants' impact over the last three years. This bar graph displays the number of occurrences that each pollutant was labeled as the main pollutant for that year in Suffolk county, from 2019-2021. The pollutants include CO, NO2, Ozone, and PM 2.5. Based off of this bar graph we can conclude that PM 2.5 is the most common main pollutant each year. PM 2.5 can negatively impact our health if found in very high levels; a contributor to this pollutant is vehicles. Increased PM 2.5 leads to poor air quality, which in turn harms our health.



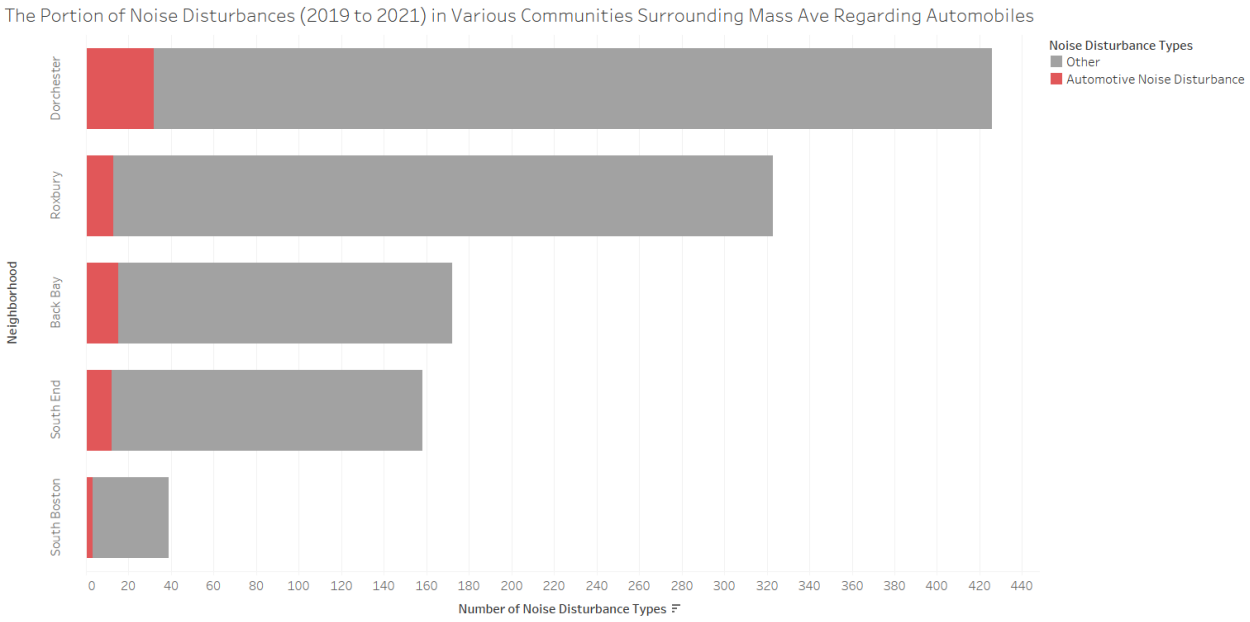
As previously stated, one source of PM 2.5 is vehicles such as cars. In this line plot we wanted to see changes over time in average PM 2.5 levels for each month across each year. The x-axis represents the month number; for example 2 is February. We specifically wanted to display this graph because we wanted to find out if this pollutant level decreased in 2020 due to Covid restrictions; since many people stayed at home and many people were most likely not traveling to Boston for work, we believed that line for this pollutant would be less steep than the blue and green lines representing 2019 and 2021. There are some parts in the graph where 2020 has lower PM 2.5 I'm levels compared to 2019 and 2021; for instance month number 5, or May, is the lowest point for 2020. During May MA had strict restrictions and people were only going out for the essentials. However, towards August some of the restrictions started relaxing, meaning more people were traveling, which contributed to poor air quality from vehicles.



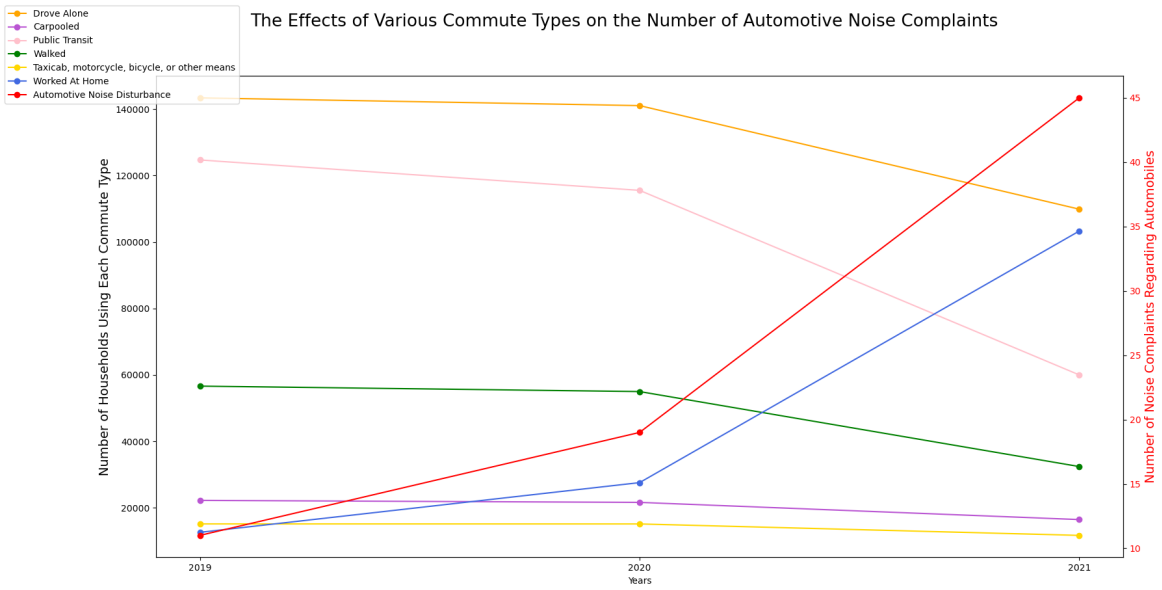
This visualization shows the number of Parking Meters that are followed by the same pay rule. From this diagram, we can distinguish each Parking Meter in Boston based on their pay rules. This could reflect whether the pay period would influence whether people are willing to use some certain parking meters, since more people would drive through some communities to the certain parking meters that have the longest period free period. Then, we can relate this to the air quality PM2.5 to see whether the pay rules are the factor that influences the air quality by using the different colors to identify these meters based on their pay rules in our future map diagram.



This visualization generates the locations of all Parking Meters based on their LON and LAT. This visualization could also be used to generate a map visualization and combine with the locations of the PM2.5 testing sites and mark these sites with different colors based on the average PM2.5 value of a year. Therefore, we can observe whether there is a relation between PM2.5 and the density of parking meters.



This visualization enables viewers to see which neighborhoods along Mass Ave have most experienced the effects of noise pollution, as well as the portion of these noise pollution complaints that were a result of automobiles. From this horizontal stacked bar chart, users can see that in the past three full years (2019, 2020, and 2021), Dorchester residents have experienced the most problems with noise pollution from vehicles and noise pollution in general. This plot also has a functionality that enables users to hover over each stack in the bars and see the percent of total noise disturbances that it takes up in the neighborhood.



This visualization displays a multi-series line chart that can be used to analyze the relationship between various commute types and the number of automotive noise disturbance complaints. From this visualization, users can see that the years 2020 and 2021 saw a significant decrease in commute transportation and an increase in the number of households working from home (likely a result of the pandemic). However, an unusual trend apparent in this graph can be seen through the gradual rise in the number of automotive noise complaints. One possible cause for this trend was discovered after we analyzed our 311 Program data further. In our 311 data, if a case is closed, a closure reason is given which can tell us whether the case was closed because it was a duplicate/invalid. However, we found that most of the complaints made in 2020 and 2021 were never closed, meaning that the large number of complaints in these years could actually be a result of invalid/duplicate reports. This visualization informs our group that we may need to find better data and/or incorporate data for more years.

Task Analysis

The visualizations in our project will primarily be developed for discovery. In our interview with Carol Blair, she discussed a lot about how she has witnessed the overconsumption of cars having an increased impact on pollution in the Mass Ave community. Therefore, the visualizations in our project will aim to quantify the effects of cars on pollution and discover

the accuracy of Carol's hypothesis. The primary consumer of our visualizations is the general public, however, as we hope that our visualizations will prove a strong enough correlation between cars and pollution so that they may convince community members to incorporate greener transportation modes into their everyday lives.

Task Table

Rank	"Domain" Task	Analytic Task	Search Task	Analyze Task
1	Examine the effect that fluctuations in the popularity of various types of commutes have on air quality in Boston	Identify	Browse	Discover
2	Examine the portion of noise pollution created by vehicles in various areas around Mass Ave	Compare	Browse	Discover
3	Quantify the impact of vehicles on the pollution of streetscapes within Mass Ave communities	Identify	Browse	Discover

Conclusion

Based on our initial analysis and observations, we can see that air and noise pollution are very much prevalent in Boston due to automobiles. Our next step is to create a final visualization design that combines our individual analyses to provide comprehensive information on the relationship between pollution and vehicle use in Boston. This visualization will be designed using Altair and Tableau and will fulfill each of our tasks listed in the table above in a thorough and organized manner.