Transit & Air Quality Team D
Final Report

Table of Contents

Team Members	**2**
Data Sources	**2**
Goal & Motivation	**2**
Challenges Faced	**3**
Overview of Project	**4**
Deliverable 1	4
Deliverable 2	4
Deliverable 3	4
Key Questions Answered	**5**
Base questions	5
Extension questions	9
Conclusion	**17**
Individual Contribution	**18**

Team Members

Mithat Kus (Team Rep), mthtks@bu.edu
Maria Eusse Henao, meusse@bu.edu
Sanath Bhimsen, sanath97@bu.edu
Chengkai Yang, cyang23@bu.edu

Data Sources

AirNow API (https://docs.airnowapi.org/)
AQICN (https://aqicn.org/city/boston/)
Census data (https://data.census.gov/)
CDC Data
(https://data.cdc.gov/500-Cities-Places/PLACES-Local-Data-for-Better-Health-ZCTA-Data-2023/qnzd-25i4/about_data)

Goal & Motivation

Our project's primary goal is to investigate the relationship between public transportation, air quality, demographic factors, and public health in Boston. We are motivated by the pressing need to understand air quality impacts residents of Boston, particularly in the context of certain communities that often bear the results of poor air quality's health consequences. By analyzing data on air quality, transportation infrastructure, and demographic information, we aim to provide actionable insights that can inform policies, foster equity, and ultimately improve the quality of life for all Boston residents.

Our project's extended phase delves into exploratory data analysis, seeking to uncover unexpected connections within this complex dataset. We aspire to unveil hidden patterns and correlations that may offer new perspectives on the intersections between air quality, health outcomes, and transit systems. Our main motivation is to unveil insights that can guide evidence-based decision-making and contribute to the broader discourse on sustainable transportation, public health, and social equity.

Challenges Faced

AirNow API Limitations: The AirNow API posed several challenges due to its rate limit of 500 requests per hour. To collect AQI data for the past five years (2017-2022), the team had to run the data collection script for each year, resulting in approximately 40 hours of continuous execution per year. Overcoming this rate limit required implementing code modifications to save progress and resume data collection from where it left off. Managing the prolonged data collection process and ensuring data continuity was a logistical challenge.

Lack of Data Granularity: The team encountered a significant challenge related to the granularity of the AQI data obtained from the AirNow API. The data from only two out of five sensors were available, all located closely to each other. This resulted in limited variance in AQI values, making it challenging to draw meaningful conclusions and perform in-depth analyses. The low standard deviations and a limited number of unique AQI values for both Ozone and PM2.5 AQI made it difficult to discern trends and patterns across multiple zip codes and years.

Lack of Guidance: The team faced challenges related to the lack of clear guidance from Spark regarding data sources and selection. This lack of guidance made it challenging to make informed decisions about which data to use and where to obtain it, which meant that we had to deal with a lot of uncertainty in the data collection process.

Transition to New Data Source (AQICN): Due to the limitations of the AirNow API data, the team had to transition to a new data source, AQICN, to obtain AQI data for the only available year 2022. Having to use census data from 2021, CDC data from 2023 and air quality data from 2022 raised concerns about the consistency and quality of our dataset

Data Quality and Mean Calculation: The new data source, AQICN, presented additional challenges. Some zip codes had limited data, covering only a few months of the year 2022. Considering that AQI tends to fluctuate between seasons and months, this further complicated the validity of our findings Moreover, AQICN did not provide a daily mean AQI recording, which led us to implement a formula to estimate the mean based on other metrics. However, the accuracy of this estimation was compromised due to non-normal distributions of daily recordings and varying recording patterns which further complicated the analysis.

Data Collection and Processing Efficiency: The sheer volume of data collected and processed, spanning multiple years, zip codes, and health metrics, presented challenges in terms of data management, processing efficiency, and resource allocation. Ensuring that data was appropriately cleaned, merged, and analyzed within the project's timeline required careful planning and execution. Furthermore, given the constraints we faced, changing the scope of the project almost every deliverable was very difficult to deal with and implement.

Addressing Data Discrepancies: The team faced discrepancies between different data sources, including census data, AQI data, and CDC data. Addressing these discrepancies and ensuring data compatibility for meaningful analysis posed a significant challenge.

Overview of Project

Deliverable 1

Before our team got started with any analysis, in order to be able to find patterns of how air quality impacts different areas of Boston, we made the decision to make our project zip code based, meaning we decided that almost all the data collected in our project to be zip-code based. For the first deliverable, we decided to collect air quality data using the AirNow API. This API provides 2 different AQI metrics, PM2.5 AQI and Ozone AQI which have different methods of calculation, meaning that different relationships could be understood using each of them. Upon consulting with our TPM, we decided to fetch data for the past 5 available years (2022, 2021, 2019, 2018, 2017). We wrote a python script that fetches daily Air Quality recordings for each of the zip codes and then calculated the yearly averages for each zip code. The main issue that we faced was that this API provider was collecting data from a very limited number of sensors that are all very close to each other. This caused our data to have very little variation and meant that our data was not granular enough for an analysis that allows for meaningful inferences to be made. Our team also collected zip code specific census data in order to make our analysis more meaningful. To be able to generalize the census data to any year, we specifically downloaded the 2021 5 year estimates. The focus for this deliverable was understanding air quality's relationship with census data.

Deliverable 2

At this point, our team was still trying to understand how problematic our AQI data was. Since there was no available alternative to that data, our team had to continue forward with the same AQI data. We first conducted a thorough analysis on the lack of granularity in our data, and tried to further understand the limitations that our data brings. Since a zip-code based analysis would not have resulted in correct outcomes, we decided to shift the scope of our data, and use Boston-specific air quality data. Furthermore, we analyzed yearly changes in both PM2.5 AQI and Ozone AQI, and tried to find patterns / relationships between air quality and transportation data for Boston in general. For this deliverable, our team looked into finding relationships with more census metrics than the previous deliverable and found important patterns. Lastly, for our extension proposal, acknowledging that the results might not reflect the true picture, our team decided to investigate correlations between zip code based CDC health data and air quality. After a surface level analysis, we found interesting patterns that helped us see that this relationship would be a good area to focus on.

Deliverable 3

For this deliverable, in order to improve the granularity of our data we collected new air quality data. The data is for the year 2022 and includes 8 zip codes where each zip code has its own sensor. With this new data we were able to properly investigate the relationship between air quality and census data as well as the relationship between air quality and health conditions.

Key Questions Answered

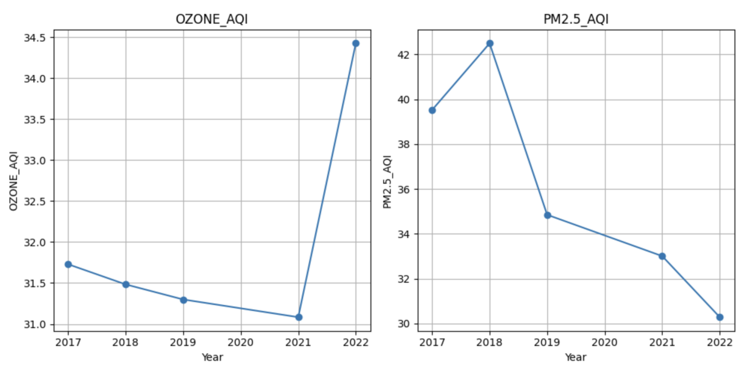
Base questions

1. What are the trends in yearly change in air quality for Boston residents? How could these trends relate to transportation trends?

The upcoming discussion delves into our analysis of the trends in yearly changes in air quality for Boston residents. We have conducted two distinct regression analyses to explore these trends, focusing on two critical air quality metrics: PM2.5 and Ozone levels. By applying linear regression models to these variables, we aim to discern potential trends in air pollution over time. The analysis could shed light on whether air quality is improving, deteriorating, or remaining stable as years progress.

The graphical analysis provided (Figure 1) illustrates the trends in air quality indices for Ozone and PM2.5 in Boston from 2017 through 2022. The Ozone AQI graph indicates a period of improvement followed by a recent upsurge, while the PM2.5 AQI graph shows a steady decline over the same timeframe.

Figure 1

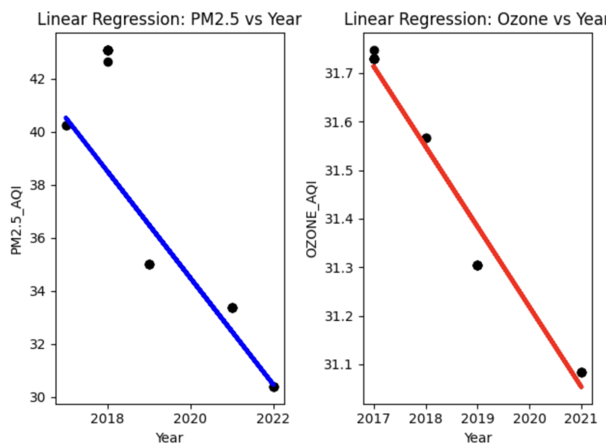


From the Ozone AQI graph, we deduce an initial period of amelioration in air quality, which is subsequently overshadowed by a notable increase. This recent development suggests a potential shift in the factors affecting Ozone levels.

In contrast, the PM2.5 AQI graph conveys a consistent improvement, indicating sustained efforts towards reducing fine particulate pollution. This trend is a positive indicator of air quality management success and possibly reflects broader shifts towards cleaner energy practices and advancements in emissions technology.

According to the U.S. Environmental Protection Agency's (EPA), there has been a significant decrease in PM2.5 levels since 1990. Specifically, the annual concentration of PM2.5 has decreased by 37% from 2000. This can explain the downward PM2.5 AQI trend, yet we will further investigate any additional causes for this. Moreover, EPA's New England regional office confirmed in October 2022 that "Based on preliminary data collected between March and September 2022, there were 24 days when ozone monitors in New England recorded ozone concentrations above levels considered healthy." (EPA). Most of the pollution responsible for ozone formation comes from large combustion sources, vehicles, and various everyday activities such as using household products. In order to investigate if there is any explanation for these AQI changes, we will analyze the relationship between transit and air quality in the next base question.

Figure 2

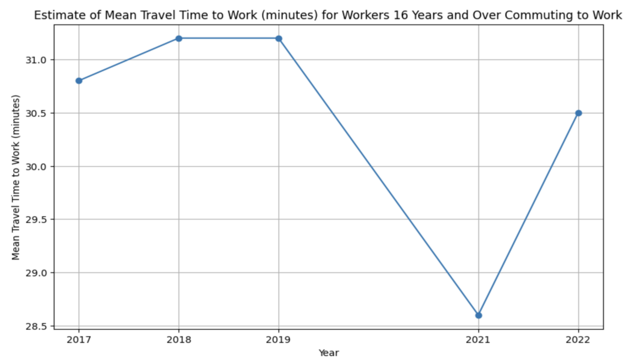


Staying within the context of the current question, in order to understand the larger picture, our team wanted to see the regression line for air quality after the removal of extreme outliers. The revised regression analysis (Figure 2), after the removal of outliers, shows a consistent decrease in both PM2.5 AQI and Ozone AQI over time. This downward trend for both pollutants could suggest that overall air quality in Boston is improving. This improvement could be due to a variety of factors, such as successful environmental policies, a transition to cleaner energy sources, advancements in emissions-reducing technologies, or even changes in public behavior such as reduced usage of personal vehicles or increased use of public transportation. However, it should also be noted that the outliers could also be representing a trend of more unhealthy air quality days and the removal of these data points could be obscuring occasional but significant spikes in pollution levels.

2. How do yearly changes in transit times in Boston relate to air quality, specifically concerning Ozone AQI levels?

As mentioned in our previous answer, most of the pollution responsible for ozone formation comes from large combustion sources such as vehicles. In order to understand the relationship between transit and air quality, we looked deeper into a certain relationship. The only transit related census data that is available for the entirety of Boston is the mean travel time to work. Variations in this commute time are indicative of traffic flow changes, potentially influencing air quality. For example, extended commutes often coincide with increased traffic congestion, augmenting vehicular emissions, and consequently affecting Ozone AQI .

Figure 3



Looking at Figure 3, it can be seen that from 2017 to 2018, there was a slight increase in mean travel time, which may suggest a modest growth in traffic congestion. This period doesn't show a corresponding increase in the Ozone AQI, which could indicate that the increase in travel time was not substantial enough to impact air quality significantly or that it was offset by improvements in vehicle emissions standards and pollution control measures.

In 2019, the mean travel time to work decreases slightly. This could be due to improved traffic management, increased adoption of alternative transportation modes such as cycling or public transport, or enhancements in road infrastructure. The decline in travel time here could contribute to the continued improvement in PM2.5 AQI, as reduced congestion generally means fewer emissions.

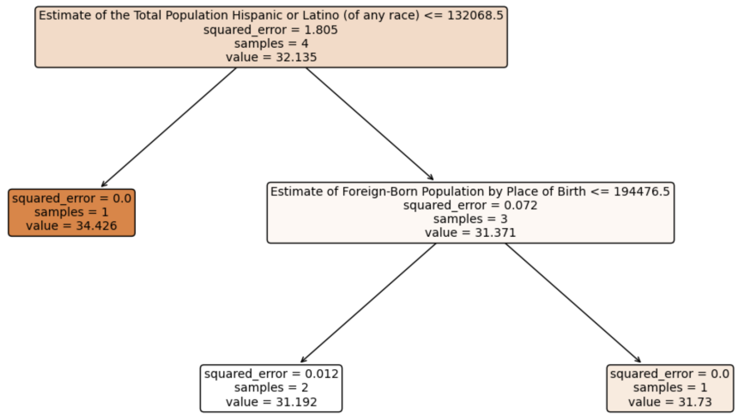
The year 2021 shows a marked decrease in mean travel time which is likely a direct consequence of the COVID-19 pandemic, where lockdowns, restrictions, and a surge in remote work led to a significant reduction in traffic. This reduction in on-road vehicles would be expected to have a positive effect on air quality. Indeed, such a scenario might contribute to the lower Ozone AQI observed in the same year, assuming that reduced traffic directly translates to decreased emissions of ozone.

The travel time graph points to an escalation in travel time in 2022, which is likely indicative of a rebound in traffic congestion as society shifts back to pre-pandemic norms. This resurgence in vehicular activity could be contributing to the increase in Ozone levels, highlighting a direct connection between traffic volume and air pollution.

To summarize, this pattern shows the direct connection between transit times and air quality, affirming that transit time is a valuable proxy for understanding and predicting changes in air pollution levels, particularly Ozone, within an urban environment.

3. How do areas with poor air quality compare to areas with better air quality based on different demographic characteristics?

Figure 4



The decision tree analysis in figure 4 was conducted on our merged zip code based AQI and census data. The decision tree analysis drawn from a broad set of 42 other census metrics, has honed in on the size of the Hispanic or Latino population and the foreign-born population as significant predictors of AQI values. The selection of these specific metrics from a larger pool underscores their potential importance in relation to air quality, as the decision tree algorithm seeks to minimize prediction error and isolate the most impactful variables.

The precision with which the model predicts AQI values in regions with a Hispanic or Latino population of less than or equal to 132,068.5, as well as the low squared error in areas with a foreign-born population by place of birth of less than or equal to 194,476.5, implies a notable relationship between these demographic factors and AQI. This is particularly striking given that these two metrics were distinguished above others, suggesting a more consistent and stronger association with AQI than the remaining variables considered.

The decision tree's identification of the Hispanic or Latino population size as a significant predictor of AQI values may suggest that these communities are situated in areas where air quality is generally poorer. Given the complex factors that contribute to air pollution, such as industrial emissions, traffic congestion, and the presence of fewer green spaces, it is possible that these demographic metrics are highlighting underlying environmental justice issues.

This very likely indicates that these populations are more likely to live in areas with higher pollution where air quality is poorer. This could be due to a variety of socio-economic factors, including housing affordability, residential segregation, and proximity to industrial or high-traffic areas.

Moreover, our findings in our extension analysis support this relationship of poor air quality impacting the Hispanic / Latino community more heavily than other communities in Boston, which suggests that these findings should be taken very seriously by policymakers and urban developers.

Extension questions

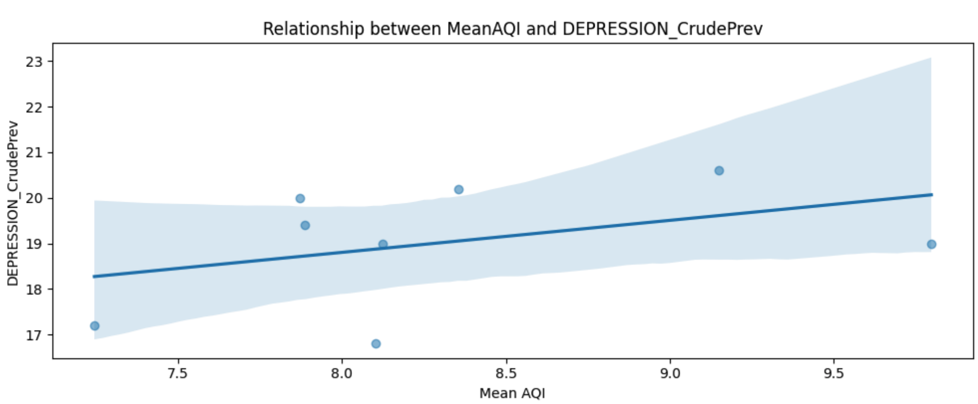
4. What is the relationship between air quality and health conditions for residents in Boston?

In order to analyze this relationship, our team created a zip code based dataset with AQI and CDC data. To give more information about the data, our AQI data is from the year 2022 and our health data is obtained from CDC's Local Data for Better Health 2023 release. Below are the health conditions that turned out to be the most correlated with air quality.

DEPRESSION_CrudePrev	0.409928
CSMOKING_CrudePrev	0.391466
COPD_CrudePrev	0.347405
PHLTH_CrudePrev	0.325904
CHD_CrudePrev	0.313530
GHLTH_CrudePrev	0.305696
ARTHRITIS_CrudePrev	0.261498
HIGHCHOL_CrudePrev	0.258266
BINGE_CrudePrev	0.256889
MHLTH_CrudePrev	0.251783
KIDNEY_CrudePrev	0.249956
STROKE_CrudePrev	0.237206
LPA_CrudePrev	0.235133
DIABETES_CrudePrev	0.222201
TEETHLOST_CrudePrev	0.218312

The analysis reveals a notable association between poor air quality and a range of health issues. The health conditions with the strongest correlations include depression, smoking, COPD (Chronic Obstructive Pulmonary Disease), high blood pressure, coronary heart disease, general health status, arthritis, high cholesterol, binge drinking, mental health, kidney disease, stroke, physical inactivity, diabetes, and tooth loss. These correlations suggest that areas with lower air quality may see higher prevalence rates of these conditions. In order to better understand and visualize these relationships, we created regression lines for these variables, and we will investigate reasonings for the variables that are known to be highly associated with poor air quality.

Figure 5

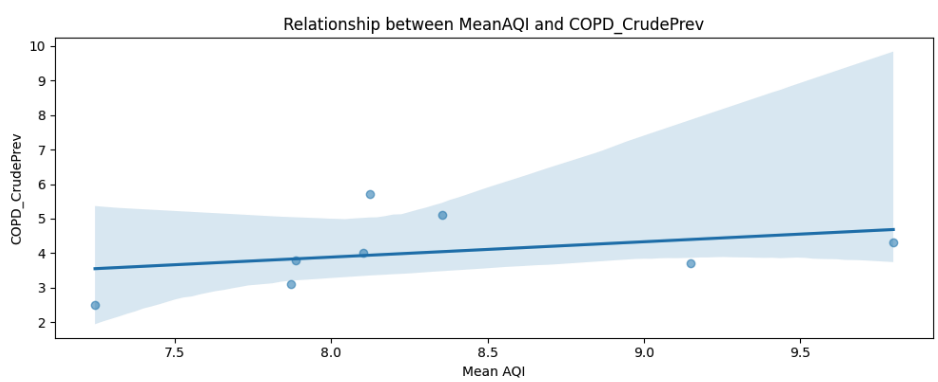


The health metric that has the highest correlation with air quality is depression prevalence. The graph (Figure 5) demonstrates a discernible linear relationship between the prevalence of depression and mean AQI, with depression rates tending to increase as air quality worsens. This trend is depicted through a regression line that steadily ascends with higher AQI values, which suggests that as air quality declines, the burden of depression in the population correspondingly rises.

The analysis of this relationship points to several underlying mechanisms that could explain this association. Poor air quality is known to trigger inflammatory responses in the body, which have been linked to the pathophysiology of depression. Furthermore, exposure to pollutants can disrupt the balance of neurotransmitters, critical in mood regulation, which could lead to depressive symptoms. Stress responses also play a crucial role, as individuals living in areas with poor air quality may experience chronic stress due to concerns about health, environmental conditions, and the economic implications of pollution. This chronic stress is a known risk factor for the development of depression.

Additionally, polluted environments often discourage outdoor physical activities and social interactions due to the discomfort and health risks associated with exposure to pollutants. Physical activity has been widely recognized for its role in preventing and alleviating symptoms of depression, while social interaction can provide emotional support and buffer against mental health issues. The decrease in these protective behaviors due to poor air quality could contribute to the higher incidence of depression.

Figure 6

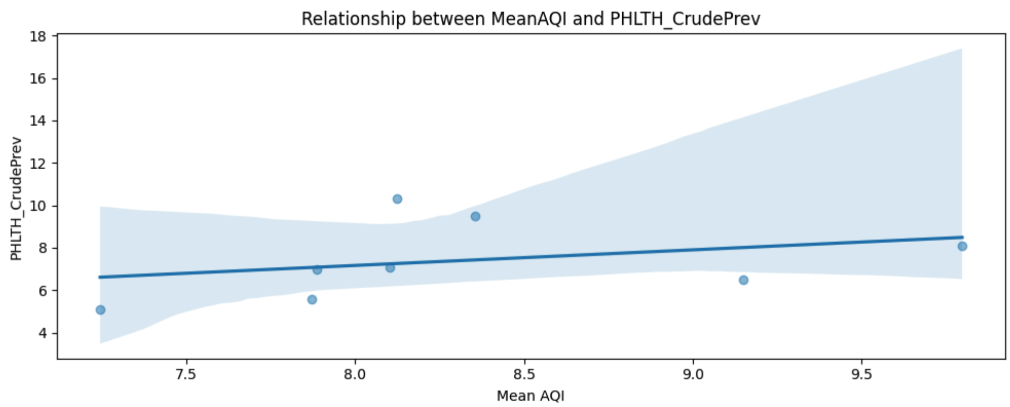


Next we look into the relationship between air quality and Chronic Obstructive Pulmonary Disease (COPD). Similar to our previous graph, this visualization (Figure 6) underscores a critical public health concern: as air quality deteriorates, the incidence of COPD increases.

Poor air quality is a well-documented risk factor for the development and exacerbation of respiratory diseases. Pollutants in the air, such as particulate matter, ozone, nitrogen dioxide, and sulfur dioxide, can cause direct harm to the respiratory tract, leading to decreased lung function and chronic inflammation, both of which are hallmark features of COPD. Furthermore, these pollutants can exacerbate respiratory infections, which can lead to more frequent and severe COPD exacerbations, a decline in lung function, and an increased risk of hospitalization.

This relationship also indicates potential socio-economic disparities, as areas with poor air quality are often economically disadvantaged, which may limit residents' access to preventive measures and healthcare services. These disparities can lead to a higher burden of COPD in communities with lower air quality, emphasizing the need for targeted interventions to improve air quality and respiratory health in these areas.

Figure 7

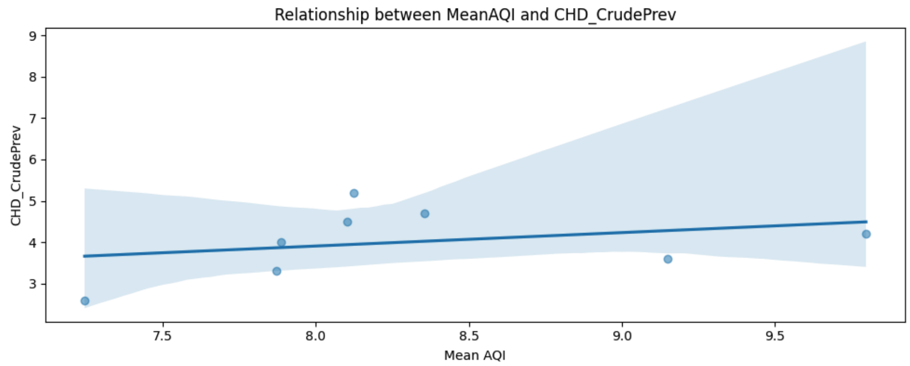


Another health metric that had one of the highest correlations with air quality was the prevalence of adults reporting poor physical health for 14 or more days within a given month. The graph illustrates a statistical correlation between mean AQI and poor physical health. This health metric, indicative of chronic or severe physical ailments, appears to worsen as air quality degrades.

Poor air quality is known to have a long list of adverse effects on physical health. Pollutants can cause acute respiratory symptoms and exacerbate chronic conditions like asthma, bronchitis, and other lung diseases, notably COPD as mentioned previously. These side effects could directly increase the number of days individuals experience poor physical health. Moreover, air pollution has been linked to cardiovascular problems, including heart attacks and strokes, due to the systemic inflammation and oxidative stress it can cause.

Another plausible explanation for the connection between poor air quality and increased reports of bad physical health is the role of air pollution in hindering outdoor activities. When air quality is low, individuals may choose to avoid outdoor exercise, leading to a decrease in physical fitness which is a predictor of poor health.

Figure 8



One last health metric that our team analyzed with a high correlation with air quality is Coronary Heart Disease (CHD) prevalence. The graph (Figure 8) demonstrates a relationship between mean AQI and the prevalence of coronary heart disease across different zip codes, with CHD rates seemingly increasing in areas with poorer air quality. This suggests a link between exposure to pollutants and cardiovascular health.

Poor air quality can lead to CHD through several mechanisms. First, certain air pollutants, such as fine particulate matter (PM2.5) and nitrogen oxides, have been shown to penetrate deep into the lungs and enter the bloodstream. Once in the bloodstream, these pollutants can cause systemic inflammation and oxidative stress, which are known to contribute to the development and progression of atherosclerosis, a primary underlying cause of CHD.

Additionally, these pollutants can have immediate effects on the cardiovascular system, including causing changes in heart rate variability, increasing blood pressure, and impairing endothelial function, which is critical for blood vessel health. Over time, these physiological changes can lead to the stiffening of arteries and contribute to the formation of plaques that narrow and harden coronary arteries, leading to CHD.

Overall, the comprehensive analysis of our Boston zipcode based dataset, integrating AQI data from 2022 with CDC's health data from the 2023 release, has shown significant correlations between air quality and various health conditions. Our investigation has highlighted that residents in areas with lower air quality are more likely to suffer from a range of health issues, including depression, COPD, CHD, and general poor physical health. Our data highlights the critical role of clean air as a determinant of public health and the importance of efforts to mitigate air pollution's impact on the Boston community.

5. What is the connection between demographic statistics and the prevalence of health issues? In what ways does air quality impact various population groups in Boston, and what role might commuting patterns and vehicular emissions play in shaping the city's overall health profile?

In order to find patterns between different demographics and health issues, we merged our CDC data with our census data from Deliverable 2, and looked at certain relationships. First, we analyzed how areas with high percentages of each racial group are affected by certain health conditions. Below are graphs of certain metrics that we believe are very striking

Figure 9

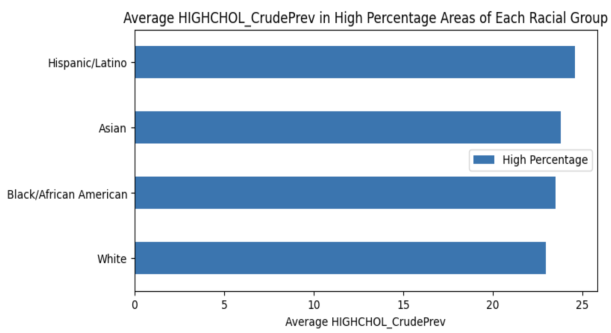


Figure 10

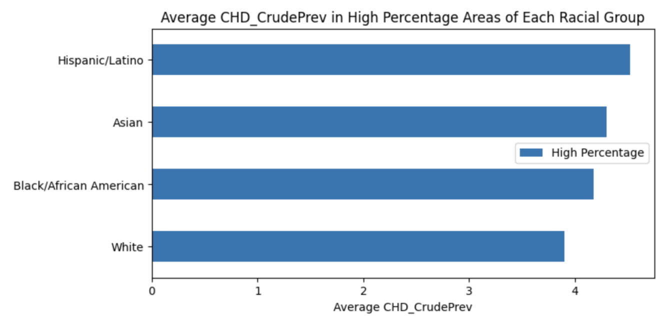


Figure 11

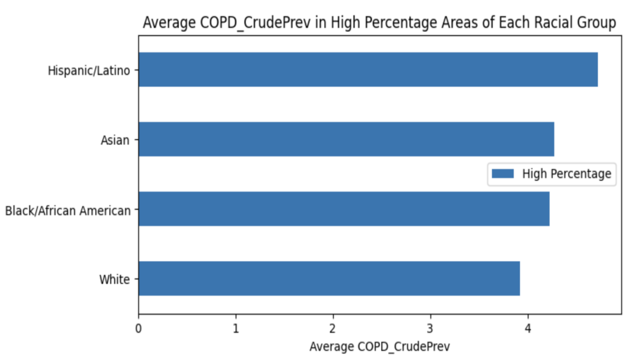
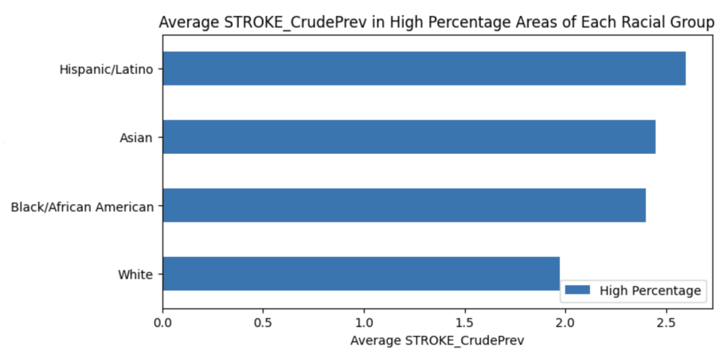


Figure 12



The series of graphs provided (Figure 9-12) each depict the average prevalence of specific health conditions—COPD (Chronic Obstructive Pulmonary Disease), high cholesterol (HIGHCHOL), stroke, and CHD (Coronary Heart Disease)—across different racial groups in areas with high percentages of each group. The graphs offer a stark visual representation of health disparities among racial groups and suggest a potential relationship with air quality.

COPD prevalence is notably higher in areas with a high percentage of Hispanic/Latino populations, which may reflect air pollution exposure. High cholesterol prevalence is similarly elevated in these communities. Stroke prevalence is higher among Black/African American populations, and CHD prevalence is substantially higher in areas with a higher percentage of White populations. These conditions are critical markers of chronic health issues that have been linked to environmental factors, including air quality.

Air pollution is known to exacerbate respiratory diseases like COPD by causing airway inflammation and decreasing lung function. For cardiovascular conditions like high cholesterol, stroke, and CHD, air pollutants can accelerate atherosclerosis—the buildup of fats, cholesterol, and other substances in and on the artery walls—which can lead to heart attacks and strokes. Particulate matter and other pollutants can also cause systemic inflammation and oxidative stress, which are known to contribute to cardiovascular diseases.

Moreover, the disparity in health conditions among different racial groups could be reflective of underlying socio-economic and environmental injustices. These include differential access to healthcare, varying levels of exposure to pollutants due to residential zoning, proximity to industrial areas, and disparities in occupational hazards. Areas with high percentages of certain racial groups may be more likely to face these challenges, which can be exacerbated by poor air quality leading to higher rates of the health conditions depicted in the graphs.

In order to better understand the bigger picture of how certain communities are more heavily affected by poor air quality, it's also important to revisit the decision tree graph discussed earlier. This graph highlighted a statistical relationship between the size of the Hispanic or Latino population and the foreign-born population within certain areas and the AQI levels.

Specifically, for the Hispanic/Latino community and foreign-born populations, the decision tree indicates that as their numbers increase in a given area, the AQI levels tend to be higher. This could imply that these communities are more likely to reside in areas with poorer air quality. When we connect this finding with the observed health disparities, particularly the higher prevalence of COPD and other chronic conditions in areas with high percentages of Hispanic/Latino and foreign-born populations, a concerning pattern emerges. It suggests that not only are these communities potentially experiencing worse air quality, but they are also exhibiting higher rates of certain health conditions that can be exacerbated by pollution.

This correlation has profound implications. It suggests that Hispanic/Latino and foreign-born populations may be disproportionately affected by the health impacts of poor air quality, possibly due to a variety of factors including socio-economic status, residential zoning, and access to healthcare. This disparity may reflect broader systemic issues such as environmental injustice, where these populations are situated in neighborhoods that are more exposed to pollutants.

Furthermore, in order to investigate the relationship between commuting patterns, vehicular emissions and health conditions, our team merged transit related variables from the zip code based census data with our new granular air quality data. The four features selected from our census data for this analysis are percentage of population working from home, percentage of population commuting to work by car, average travel time to work and percentage commuting to work. Some of the most notable relationships are graphed below.

Figure 13

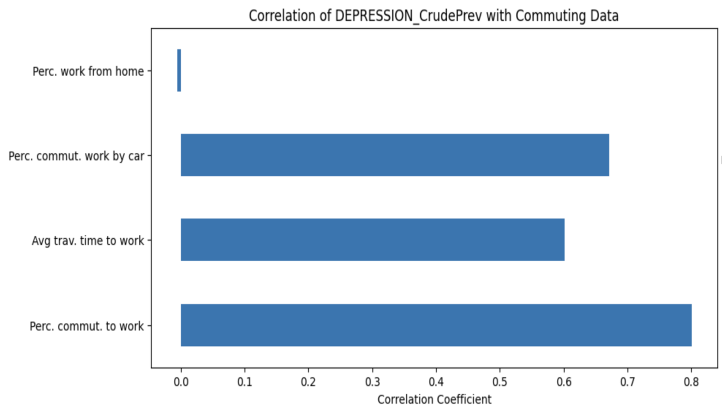


Figure 14

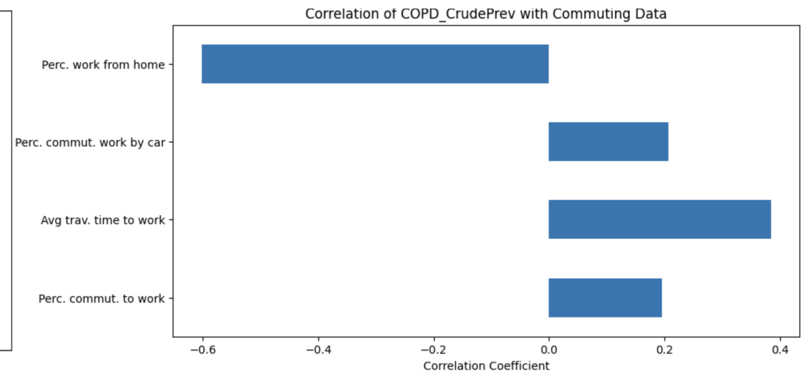


Figure 15

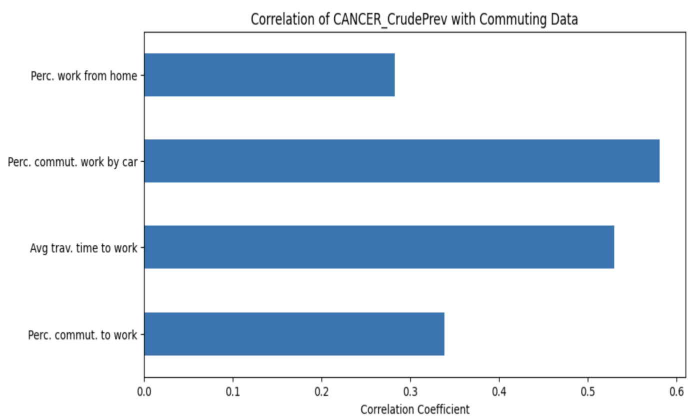
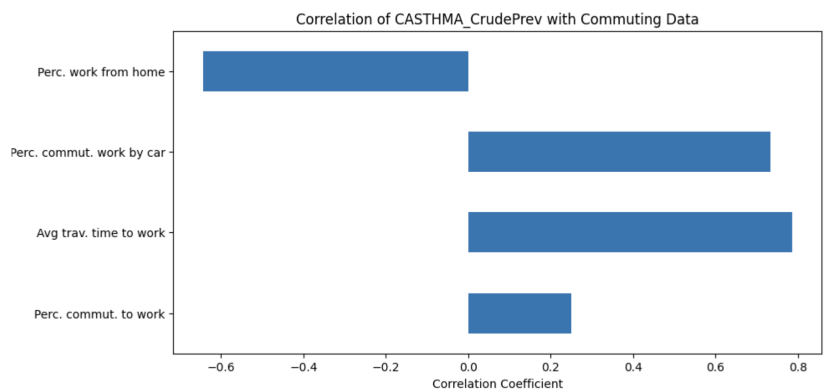


Figure 16



The series of graphs (Figure 13 - 16) display the complex relationship between commuting habits, vehicular emissions, and the prevalence of certain health conditions within Boston. This relationship helped our team understand the broader implications of air quality on public health.

For depression, the data indicates a considerable positive correlation with the percentage of individuals commuting by car and the average travel time to work. This relationship may reflect the psychological impact and physiological stress associated with prolonged exposure to harmful pollutants during extended commutes in dense traffic. As mentioned previously, such conditions are known to contribute to the development and exacerbation of depressive symptoms, potentially due to the direct effect of air pollutants on brain health and the body's stress response.

COPD, as mentioned previously, is a respiratory condition highly sensitive to air quality and here it shows a strong positive correlation with car commuting. This aligns with the understanding that prolonged exposure to vehicle emissions can exacerbate respiratory ailments, as cars emit a variety of pollutants known to impair lung function and increase chronic respiratory inflammation.

When examining cancer prevalence, the positive correlation with average travel time to work might indicate a linkage between longer exposure to vehicle exhaust and increased risk of certain types of cancer. Exhaust contains carcinogens like benzene and formaldehyde, which have the potential to initiate or promote cancer development over time.

The negative correlation between asthma prevalence and the percentage of people working from home suggests that individuals who are not exposed to outdoor air pollutants, which are common triggers for asthma, may have fewer symptoms or episodes. This could indicate that for many, the indoor environment is less of a factor in asthma exacerbation than outdoor air quality, which highlights the significant impact of urban air pollution on respiratory health. This proves the importance of improving outdoor air quality to reduce asthma prevalence and suggests that efforts to clean the air could have a direct benefit for individuals with asthma.

Overall, the evidence from this analysis calls for a new highly detailed public health approach that addresses both the immediate and underlying causes of these health disparities. Efforts to improve air quality through better urban planning, enhanced public transportation systems, and support for remote work arrangements can be powerful ways to improve public health, particularly in pollution-impacted communities. Addressing these issues holistically could lead to significant improvements in the health and well-being of all Boston's residents, with particular benefits for those currently at a disproportionate risk.

Conclusion

Throughout the course of our project, our team, Transit & Air Quality Team D, has undertaken a comprehensive analysis of the intricate relationship between public transportation, air quality, demographic factors, and public health in the city of Boston. Our efforts have uncovered a multitude of insights, and we draw several key conclusions from our findings.

First and foremost, our investigation has highlighted the critical importance of clean air as a determinant of public health. We have established a clear connection between poor air quality and a range of health conditions, including depression, Chronic Obstructive Pulmonary Disease (COPD), Coronary Heart Disease (CHD), high cholesterol, stroke, and general physical health. The data underscores that residents in areas with lower air quality are more likely to suffer from these health issues, which, in turn, calls for urgent attention to improve air quality to safeguard the well-being of the Boston community.

Furthermore, our analysis has proven that there is a disproportionate impact of poor air quality on specific population groups, particularly the Hispanic/Latino and foreign-born communities. These groups tend to reside in areas with poorer air quality and exhibit higher rates of certain health conditions that can be exacerbated by pollution. This disparity proves to everyone the need for targeted interventions and policies to address environmental justice issues and reduce health inequities.

Additionally, our investigation into commuting patterns and vehicular emissions has revealed important relationships between transportation choices and health outcomes. Prolonged exposure to harmful pollutants during commutes, especially by car, is associated with increased risks of depression, COPD, and cancer. This emphasizes the importance of promoting alternative transportation modes, reducing vehicular emissions, and supporting remote work arrangements to improve public health and reduce the burden of pollution-related health conditions.

In conclusion, our project has not only advanced our understanding of the relationships between public transportation, air quality, demographics, and health in Boston but has also provided valuable insights for evidence-based decision-making and policy development. As we move forward, it is imperative to prioritize initiatives that address air quality improvements, promote environmental justice, and foster healthier commuting options to enhance the overall quality of life for all Boston residents. Our findings serve as a call to action and emphasize the need for collaborative efforts among policymakers, urban developers, and public health officials to create a more sustainable and equitable future for the city of Boston.

Individual Contribution

Mithat Kus: Firstly, Mithat wrote the python scripts that were used in fetching the data through the AirNow API. He took on the responsibility of analyzing CDC (Centers for Disease Control and Prevention) data and its correlation with air quality index (AQI), census data, and other relevant factors. He conducted a detailed analysis to identify and understand the relationships between AQI and various health conditions, including depression, COPD, high blood pressure, coronary heart disease, and more. He also played a critical role in uncovering correlations between AQI and specific health metrics, which shed light on how air quality impacts public health in Boston. His work involved creating visualizations to clearly illustrate the associations between air quality and health outcomes. Mithat's analysis helped establish the importance of clean air for public health and contributed to the project's broader understanding of the connections between air quality, demographics, and health in the Boston area. He was also in charge of communicating the team’s challenges and problems to our professor Lance Galetti, to our TPM and to the Slack channel for our project. Finally, he took on the role of writing the reports and preparing the presentations.

Maria Eusse Henao: Maria was deeply involved in the data collection process, helping retrieve air quality data using the AirNow API. She also conducted extensive exploratory data analysis, focusing on the limitations of the initial AQI data, and was pivotal in the decision to switch to more granular data from AQICN. Maria also found very important correlations that shifted the focus of our project, such as discovering the relationship between poor air quality and Hispanic/Latino populations. She played a significant role in analyzing the relationships between air quality, transportation data, and various census metrics. Her work was instrumental in uncovering patterns and correlations within the data and contributed to the project's overall analysis. Maria also contributed to the preparation of our previous presentations as well as the reports.

Sanath Bhimsen: Sanath contributed to the data collection phase by assisting in fetching air quality data using the AirNow API. He was actively involved in the analysis of air quality trends over the years, focusing on both PM2.5 AQI and Ozone AQI. His work included the identification of trends, outliers, and potential factors influencing air quality. His contributions to this analysis helped the team understand the patterns and fluctuations in air quality in Boston and how they relate to public health.most importantly, his data visualization skills allowed us to understand the problems with the data obtained from the AirNow API data, which was a crucial moment in our project as it led us to start searching for different air quality sources.

Chengkai Yang: Chengkai played a significant role in data analysis, particularly in exploring the correlations between air quality and health conditions in Boston. He made substantial contributions to the development of regression models aimed at understanding the relationships between air quality metrics (PM2.5 AQI and Ozone AQI) and various health conditions. Chengkai was instrumental in visualizing these complex relationships, creating informative charts and graphs that helped the team grasp the impact of air quality on public health. His data visualization skills were very important in conveying the project's findings effectively. He also continuously gave feedback throughout the project which helped the team constantly improve and develop.