Transit & Air Quality

Introduction

The project aims to analyze the interplay between transportation infrastructure and air quality in Boston, Massachusetts. With a focus on the environmental and public health implications of transportation choices, the study will assess the impact of proximity to public transit and roadways on air quality. Special attention will be given to how these factors affect different demographics, particularly marginalized communities. Utilizing datasets that range from pollution indexes to census demographics, the project seeks to uncover actionable insights to inform sustainable urban transportation policies.

Data Collection

- **Collecting Data**
    - Downloading necessary CSV files from the given sources
    - Using various APIs such as Purple Air and Airnow, implementing code to automate the data collection
- **Creating Metadata for collected data**: brief description of datasets and features
- **Data Filtering**: Selecting key columns related to community type and racial demographics.
- **Data Cleaning**: Removing rows with undefined community types.

Exploratory Data Analysis (EDA)

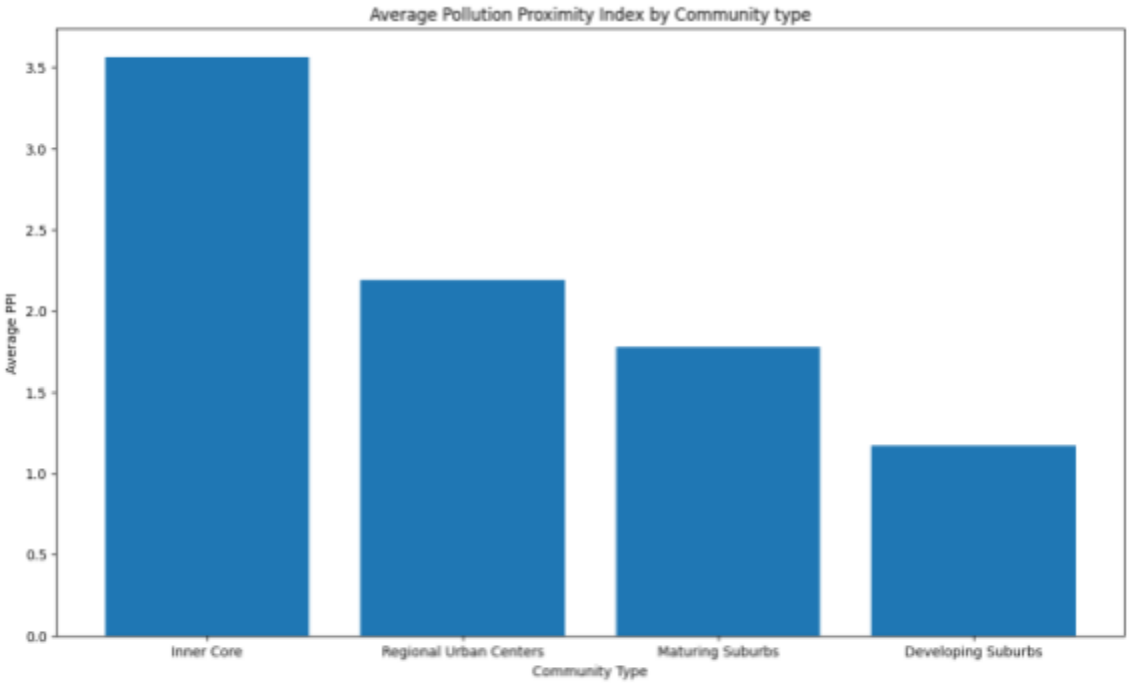

Figure 1. Bar graph showing the mean Pollution Proximity Intensity (PPI) by community type in Boston.

- Population Density
    - As shown above in Figure 1, communities that are closer to the inner core of the city have much higher PPIs compared to those farther from the core.
        - PPI from highest to lowest by community type
            - Inner Core, Regional Urban Centers, Maturing Suburbs, Developing Suburbs
        - As we would have guessed, more urbanized areas have higher pollution proximity indices
            - Thus, we can conclude that population density is directly related to the PPI

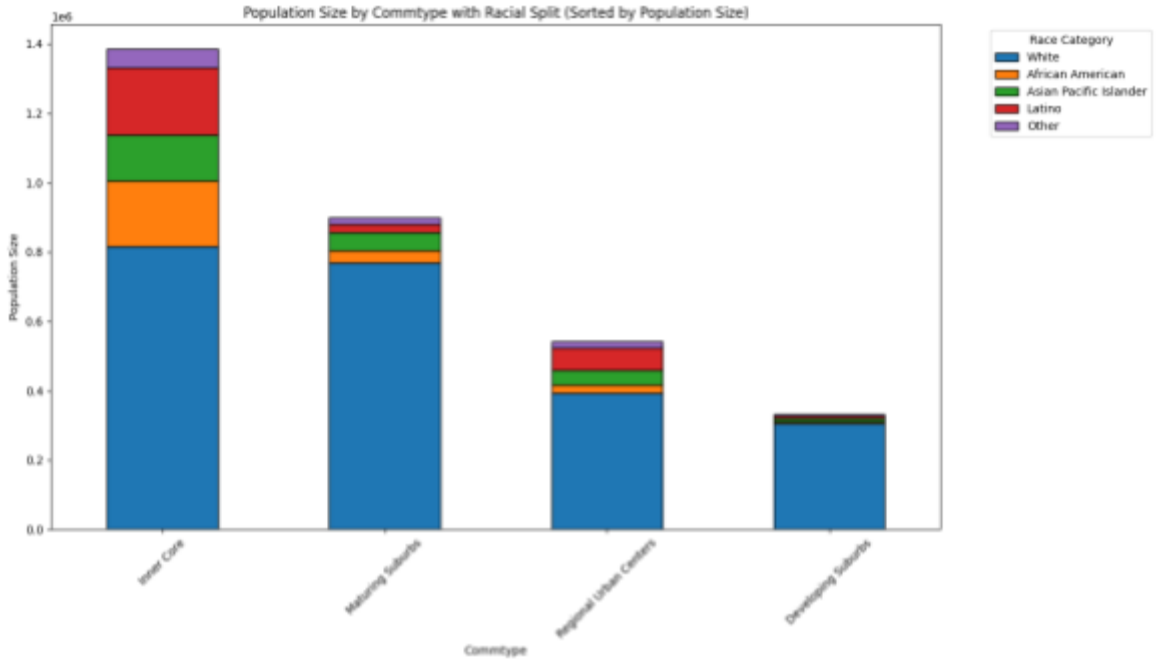

Figure 2. Stacked bar chart showing population distribution by race within each community type, sorted by total population size.

- Race/ethnicity
    - After having looked at Fig. 1, we assumed there would be a strong correlation between population size and the PPI for each community type
        - However, we found a different order for population size from highest to lowest
            - Inner Core, Regional Urban Centers, Maturing Suburbs, Developing Suburbs.
    - Furthermore, the community types most affected by high average Pollution Proximity Intensity have a much higher proportion of minorities compared to those with lower PPI.
        - As seen in Fig. 2, Inner Core and Regional Urban Centers have the highest percentage of minorities. However, even though Maturing Suburbs has a higher population compared to Regional Urban

Centers, Inner Core and Regional Urban Centers have a higher average PPI compared to Maturing Suburbs and Developing Suburbs

- Thus, we can conclude that poor air quality affects racial minorities more than the white population

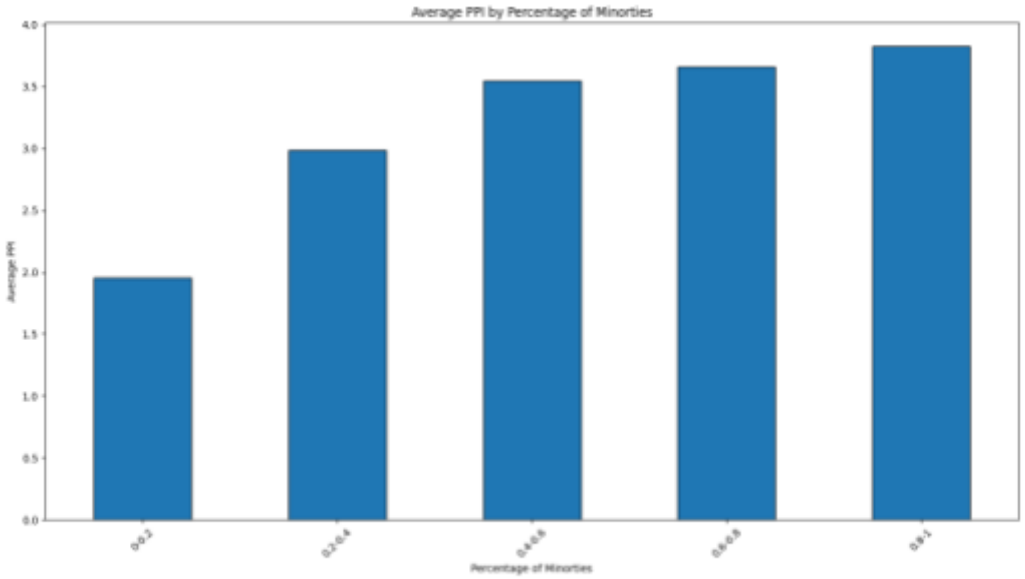

Figure 3. Bar graph depicting the average Pollution Proximity Index (PPI) across varying percentages of minority populations.

- Racial Disparity in Air Quality
  - Fig. 3 affirms the correlation we found between the percentage of minorities and the quality of air
    - Higher percentage of minorities = Higher Average PPI
  - Relation
    - Average PPI increases logarithmically as the percentage of minorities increases

Figure 4. Bar chart ranking Boston neighborhoods by housing density, highlighting variations in residential unit concentration across the city.

Figure 5. Stacked bar chart displaying the composition of combined PM2.5 and PM10 values in Boston neighborhoods, ordered by total particulate matter concentration.

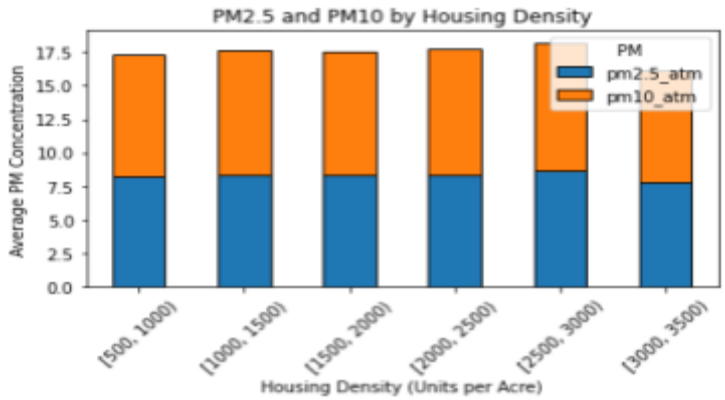

Figure 6. Stacked bar chart comparing average concentrations of PM2.5 and PM10 across bins of housing density, highlighting the relationship between residential crowding and particulate matter levels.

- Housing Density
    - After having compared Fig. 4 and 5, we saw little to no correlation between the combined PM values of a neighborhood and its housing density
    - Fig. 6 confirmed our hypothesis that housing density of a neighborhood doesn't have a significant effect on the particle pollution concentration on that given neighborhood

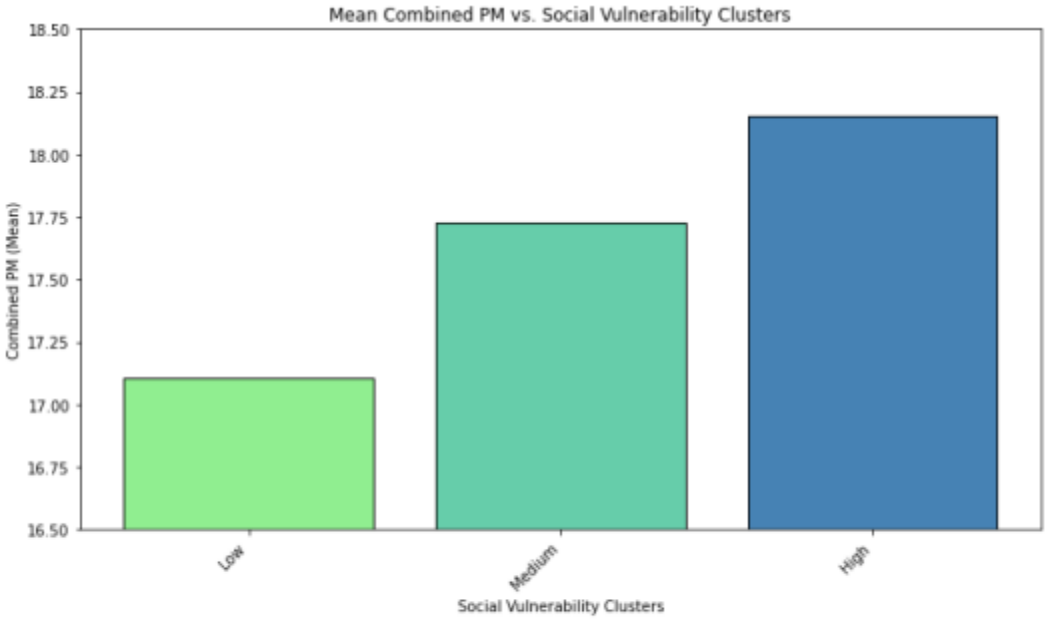

Figure 7. Bar chart comparing mean combined PM levels across three clusters of social vulnerability, revealing potential disparities in air quality related to social factors.

- Social Vulnerability
    - For Fig. 7, we explored different features of social vulnerability such as the senior population, total children, people of color, limited english proficiency, low income, disabled individuals and medical illnesses.
        - After having clumped up Boston communities into low, medium, and high social vulnerability groups, we found that particle pollution affected communities with higher social vulnerability compared to those with lower social vulnerability.

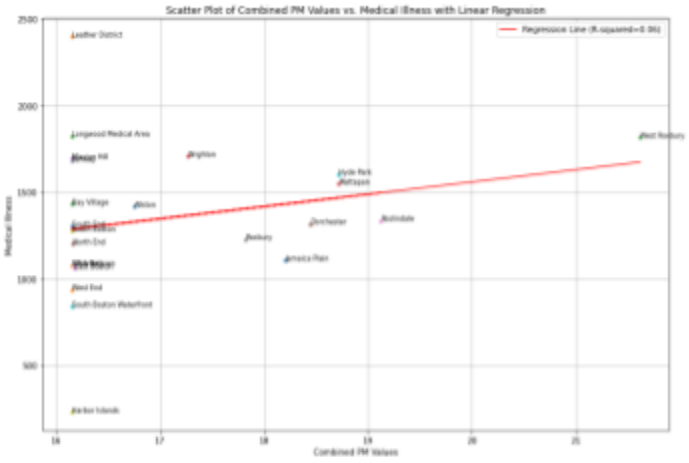

Figure 8. Scatter plot showing the relationship between combined PM levels and medical illness rates in Boston neighborhoods, with a linear regression line indicating the correlation strength.

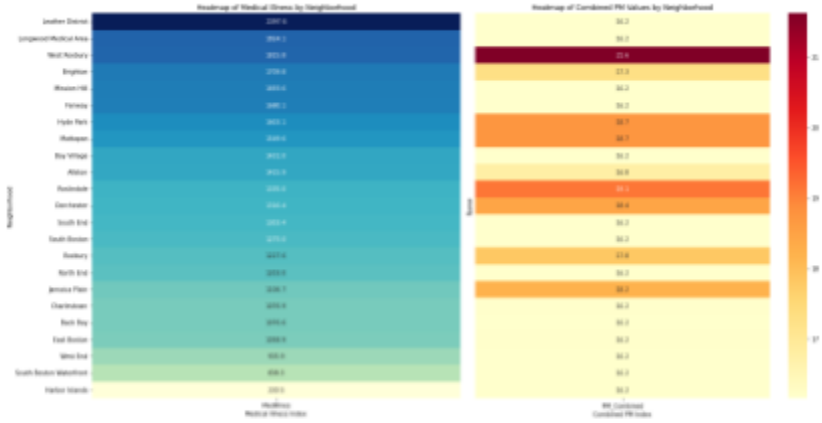

Figure 9. Twin heatmaps presenting average medical illness rates and combined PM values in Boston neighborhoods, sorted by illness rates for direct comparison.

- Correlation between Medical Illness and Particle Pollution
    - By simply looking at the scatterplot, the graph doesn't necessarily indicate a strong relationship between Medical Illness and the air quality
    - However, by charting a linear regression line we can find that there is a positive correlation between the two variables.
        - As seen in Fig. 8, West Roxbury which has the highest combined PM value also has one of the highest rates in medical illnesses
    - The relation between medical illness and PM is clearly represented in the side by side comparison of heatmaps
        - From Fig. 9, we can see that the neighborhoods with the lowest medical illnesses such as Harbor Islands, South Boston Waterfront, West End and East Boston all have low PM values
        - On the other hand, neighborhoods such as West Roxbury, Hyde Park and Mattapan that are on the higher end in terms of medical illnesses also have high PM values

Results

2. How do areas with poor air quality compare to areas with better air quality based on different demographic characteristics, specifically:

- Race/ethnicity
    - As seen in Fig. 3, areas with a higher percentage of minorities have a higher average PPI.
        - More urbanized communities have higher percentages of minorities, as seen in the segmentation of each community type in figure 2. Those community types have a higher average PPI even when the population size is not larger.
- Housing density
    - No significant correlation between housing density and air quality
- Population density
    - Community types that are more densely populated have a higher average PPI, as shown in Fig. 1.
- Social vulnerability
    - Areas where combined PM 2.5 and PM 10 is higher also have a higher social vulnerability.

3. What is the relationship between health data and What are the trends in yearly change in air quality for Boston residents by neighborhood, zip code.

- We found a positive correlation between health data and air quality as seen in Fig. 8 through the linear regression line
- In addition, neighborhoods with more medical illnesses tend to have higher combined PM values
- However, given more time we hope to explore the relationship between health data and air quality in terms of yearly change instead of just exploring data from 2019.

Individual Contributions

- Individual contributions of each team member. We recommend that each team member writes 3-4 lines about their contributions, which can then be compiled into the report.

- Sungjun Park
    - Metadata for the datasets we used (README.md file under data file)
    - Worked with Dhurv on data visualization for the correlation analysis between Pollution Proximity Intensity and race/ethnicity
    - Exploring the effects of population density on the air quality with Dhruv
    - Finding the relation between different levels of social vulnerability and air quality with Emily
    - Heat Maps to find correlation between medical illnesses and particle pollution

- Dhruv Chandwani
    - Exploring the effects of population density on the air quality with Sungjun
    - Did initial exploring of datasets with Sungjun and Azad
    - Worked with Sungjun on data visualization for the correlation analysis between Pollution Proximity Intensity and race/ethnicity
    - Worked with Azad to extract data from the PurpleAir API: Collected sensor data for Purple API and filtered for geographically relevant outdoor sensors which were public.
    - Worked with Azad to script that computed and added the closest sensor to each row in a Climate Ready Boston Social Vulnerability dataset
    - Worked with Azad to determine how to extract centroids from polygons given by the GeoJSON and CSV

- Emily Sun
    - Analyzed the relationship between PPI and race
    - Found the relation between different levels of social vulnerability and air quality with with Sungjun
    - Data visualization for relationship between housing density and combined PM values
    - Linear regression for correlation between medical illness and combined PM values

- Azad Ellafi
    - Did initial exploring of datasets with Sungjun and Dhruv.
    - Wrote a python script that automated data extraction from the Airnow API.
    - Worked with Dhruv to extract data from the PurpleAir API: Collected sensor data for Purple API and filtered for geographically relevant outdoor sensors which were public.
    - Worked with Dhruv to write a python script that computed and added the closest sensor to each row in the Climate Ready Boston Social Vulnerability dataset.
    - Worked with Dhruv to determine how to extract centroids from polygons given by the Social Vulnerability GeoJSON and CSV.

- Juntao Yan
    - Finding alternative dataset from outside sources
    - Worked on visualization of alternative datasets to determine viability