
Deliverable 2

Transit Performance - Team C

Chandahas, Eason, Manushi, Munir, Patrick

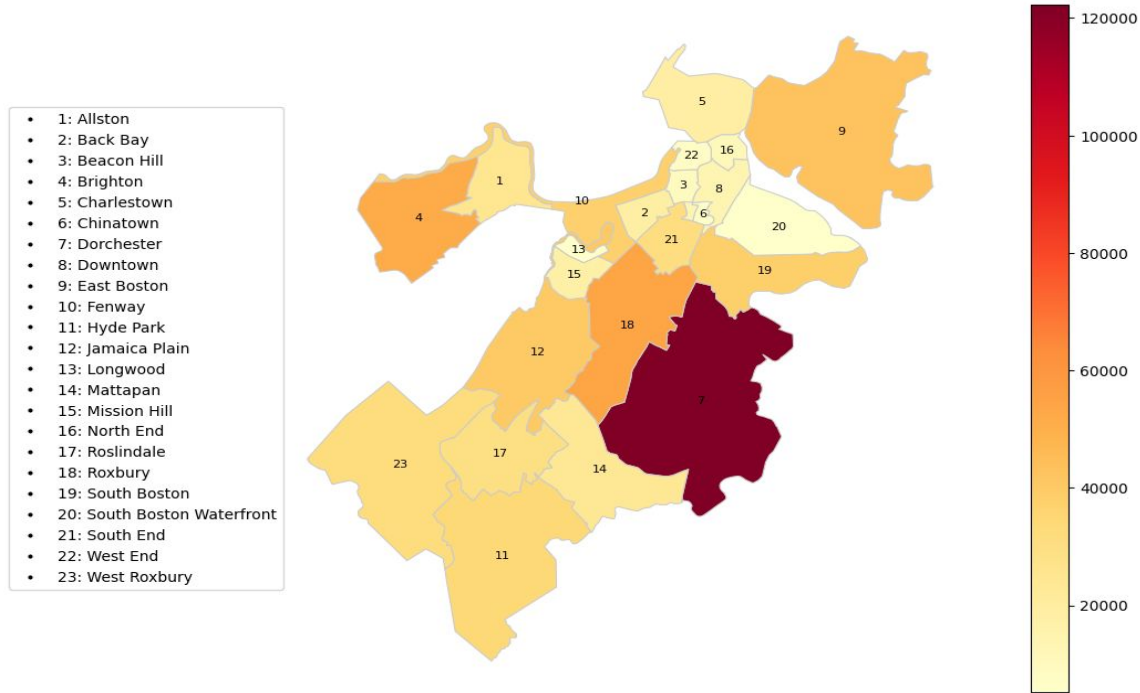
Transit Performance

- Public transit provides vital access to jobs, services, etc in cities like Boston
- Historically, some communities are underserved by public transit due to inequitable planning
- This analysis examines MBTA bus performance and census demographics in Boston
- Goals are to identify service disparities and see if they disproportionately affect marginalized groups
- Focus on MBTA data and neighborhood-level census demographics in Boston
- Key questions:
 - What are the end-to-end travel times for different bus routes?
 - Are there noticeable disparities in on-time performance between routes?
 - What are the demographics of communities served by each route?
 - Do any service disparities appear to adversely impact disadvantaged groups?
- We use basic data science techniques such as APIs, data cleaning, and mapping to gather results
- Aims to provide insights into service inequities and guide investments to improve equity
- Public transit equity has implications for social justice, mobility, sustainability in diverse cities

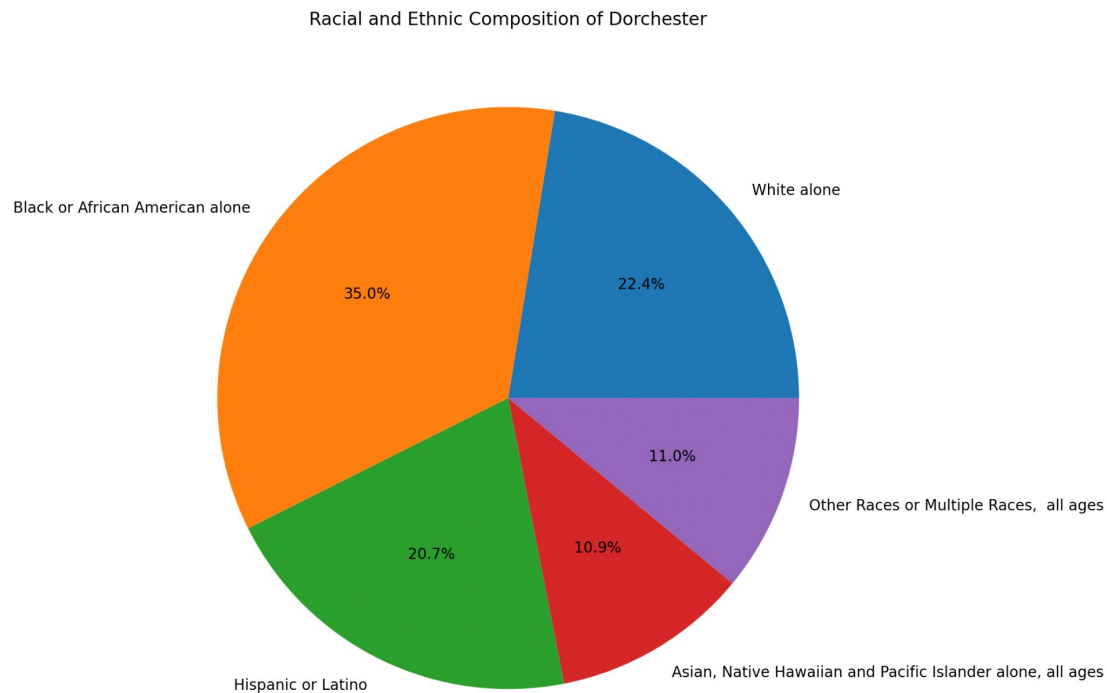
Progress So Far

- What we've done so far:
 - Exploratory Data Analysis (EDA) of bus data - analyzed distributions of points in routes, bus volumes, headway, and frequencies of bus lines.
 - Demographics - analyzed the demographics of each neighborhood based on census data
 - Addressed base questions 1 and 2 - evaluated lateness and punctuality of bus routes to assess disparities in service levels.
 - Addressed base question 3 - analyzed demographic of neighborhoods served by different routes.
- Division of work so far
 - Chandrahas and Manushi - data cleaning, preprocessing, EDA of bus data, addressing base questions 3,4,5.
 - Eason and Munir - addressing base questions 1, 2, 3, and EDA of census data.
 - Patrick - addressing base question 1, extension project EDA and questions, and presentation of results (report and PowerPoint)
- Data-related steps so far
 - Data Collection - accessed MBTA API for real-time bus data; spatial joins with 2020 Census data for demographic insights.
 - Data Cleaning and Preprocessing - focused on January 2022 data; calculated key metrics like route travel times and on-time percentages.

Population Distribution For The City of Boston

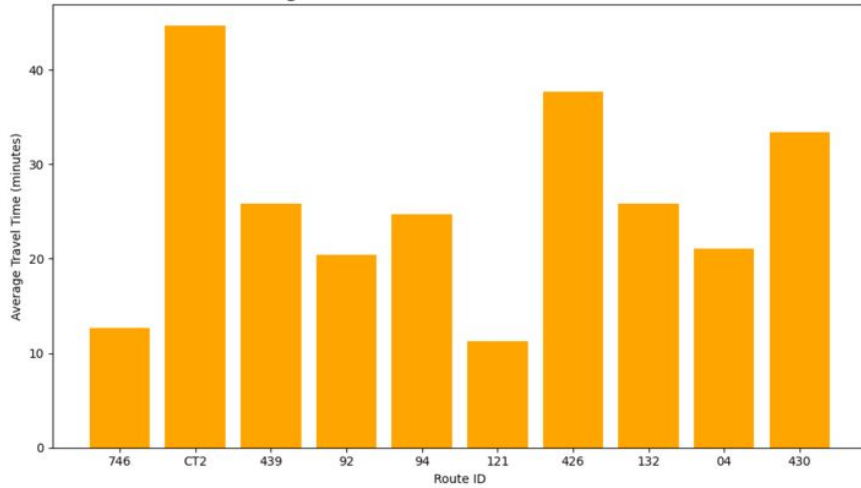


Ethnic Distribution of the Most Populated Neighborhood

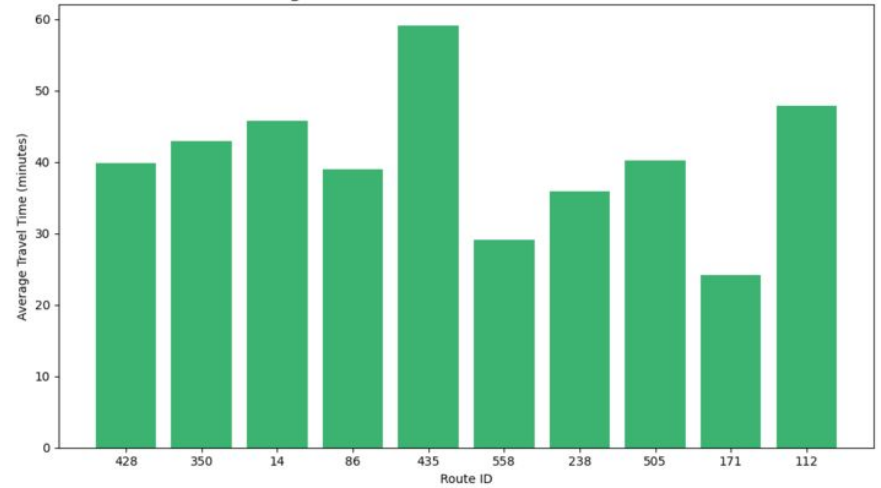


Average Travel Times for Bus Routes

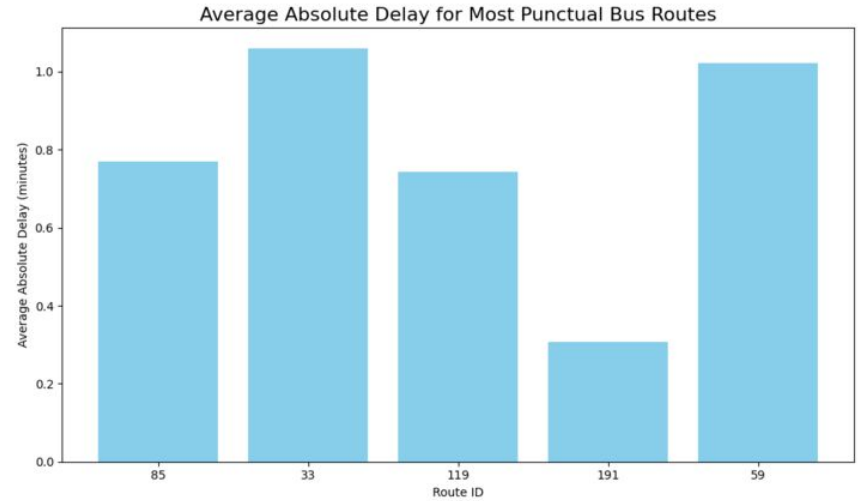
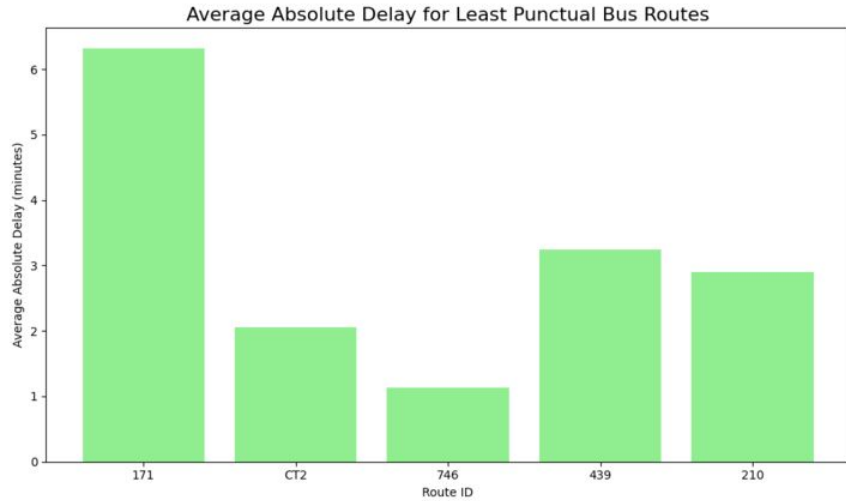
Average Travel Times for Most Late Bus Routes



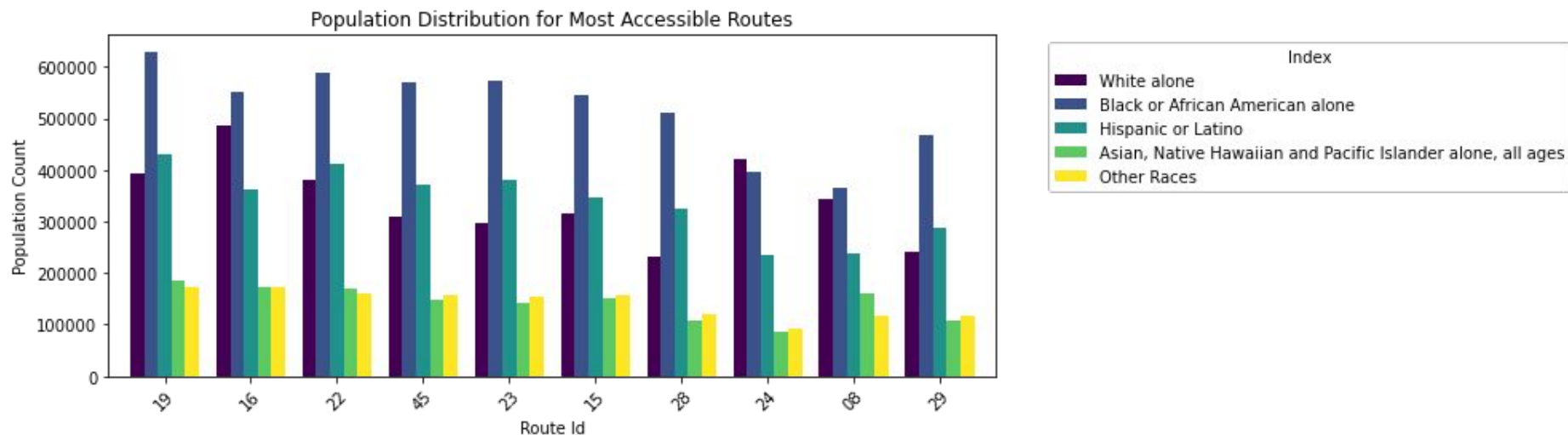
Average Travel Times for Least Late Bus Routes



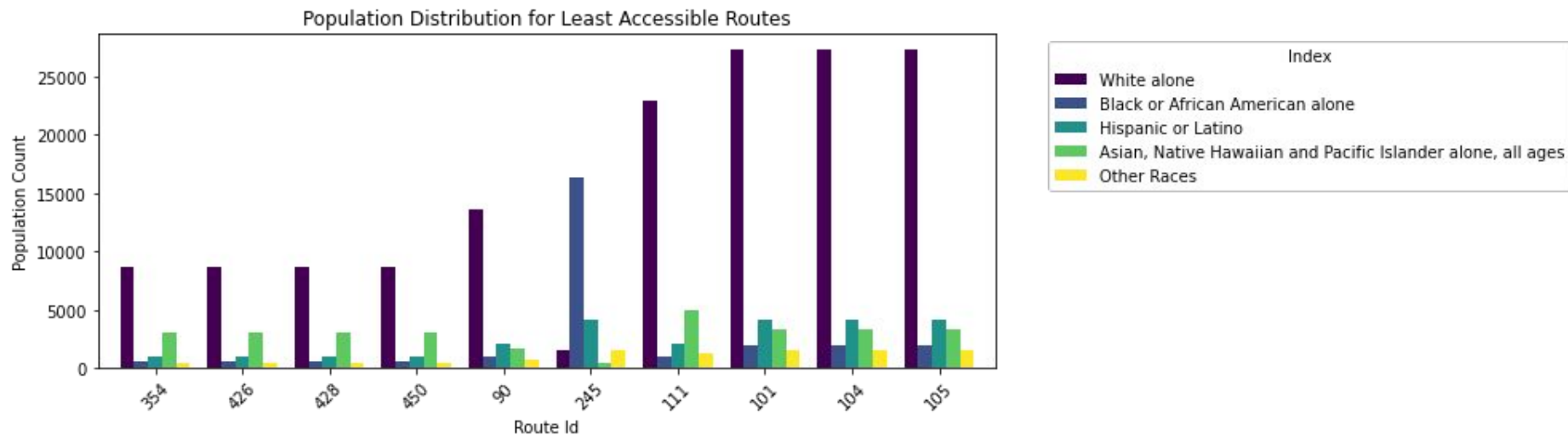
Service Level Disparities for Bus Routes



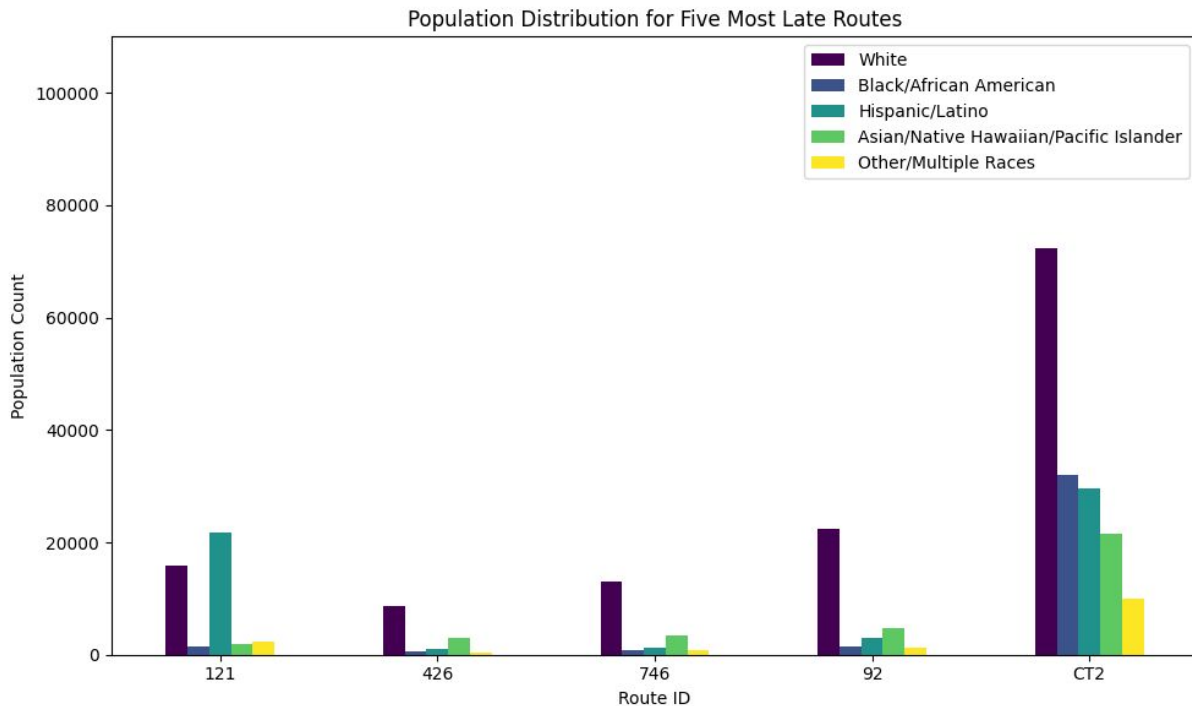
Population Distribution for Most Accessible Routes



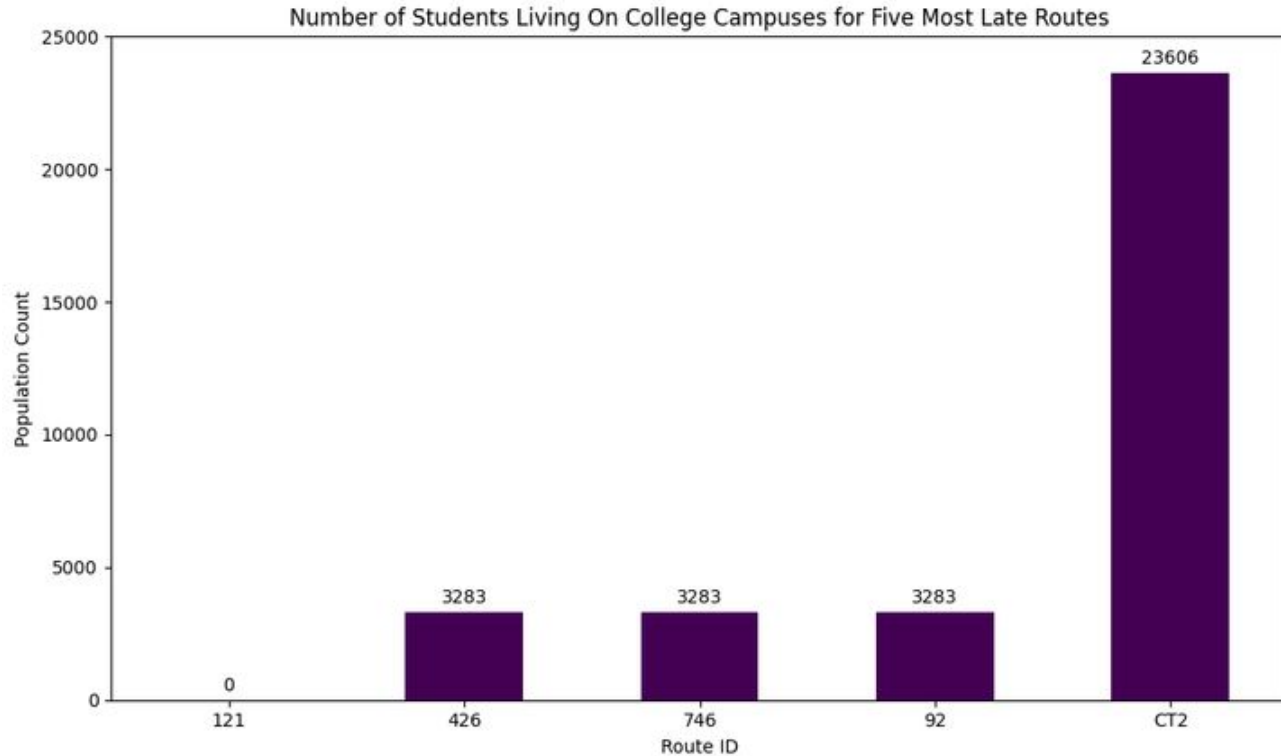
Population Distribution for Least Accessible Routes



Ethnic Distribution of Routes with Maximum Lateness



Contributing Variables affecting Bus Routes



Challenges & Limitations

List of challenges & limitations

- Mapping bus routes to neighborhoods: bus routes and their stops are given by municipality, not by neighborhood. We had to work with the geographic data provided for each stop to assign them by neighborhoods
- Lack of localization: the census data provided was not granular enough to give context on each neighborhood. We had to make assumptions about uniform population and ethnic density when drawing conclusions based on our dataset
- The MBTA data was for the entire state of Massachusetts but the census data was limited to the city of Boston so that created a challenge in merging the two datasets.

Conclusions so far

Racial Diversity: The most populated regions (Dorchester and Roxbury) have a higher racial and ethnic diversity with more balanced proportions among different groups, whereas the least populated regions (South Boston Waterfront and Longwood) are predominantly white with fewer individuals from other racial and ethnic backgrounds. These patterns can reflect various factors, including historical settlement patterns, economic opportunities, and social dynamics within these regions.

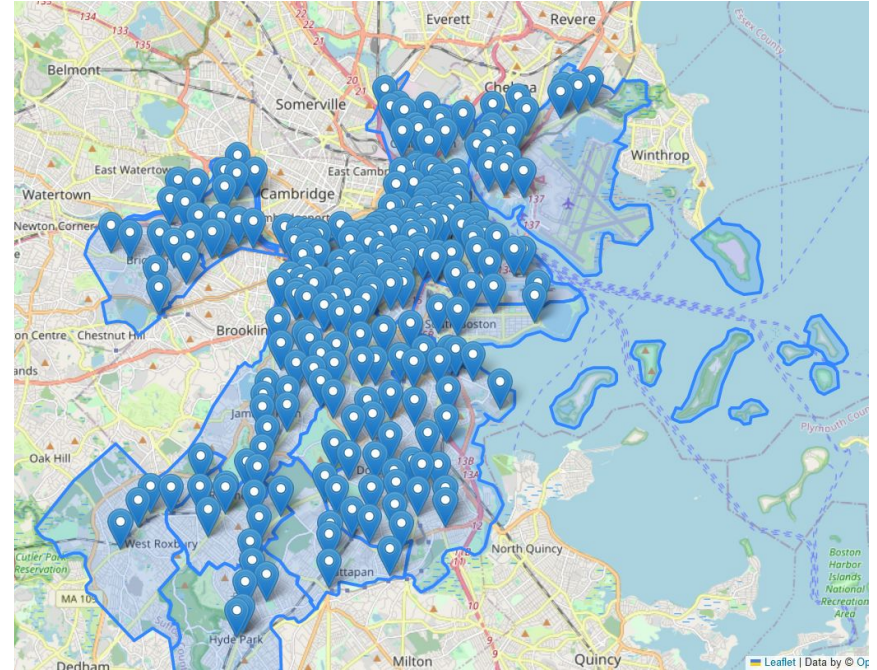
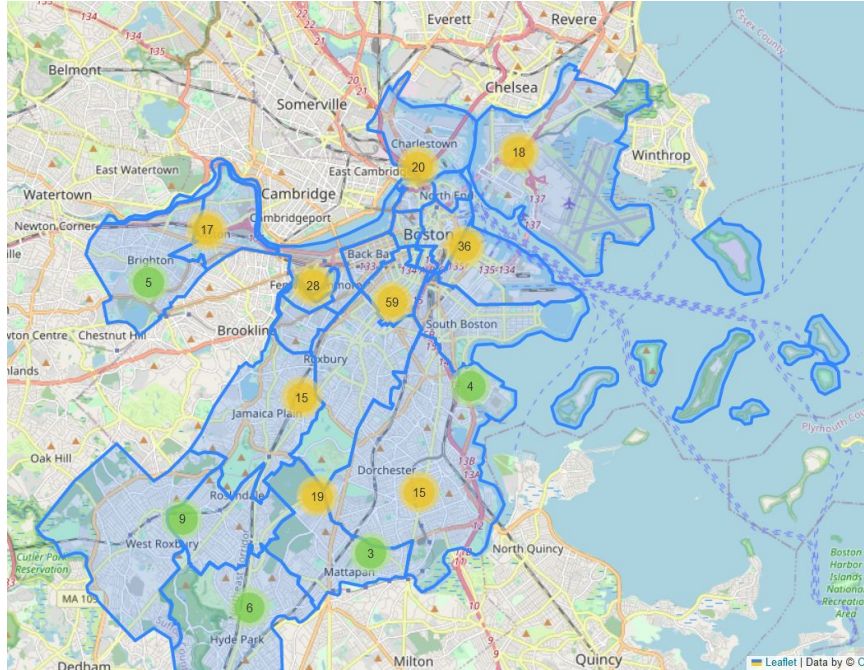
Age Diversity: The most populated regions have a higher percentage of children below the age of 18 (~20%), whereas the least populated regions have a lower percentage of children below the age of 18 (<6%)

Racial Disparities: The largest racial disparities occur on the bus routes that serve the least number of people. When considering how to address racial disparities, the MBTA should look at potentially replacing bus routes that serve the least number of people with routes that serve more people in more diverse parts of the city, improving accessibility in areas that need it.

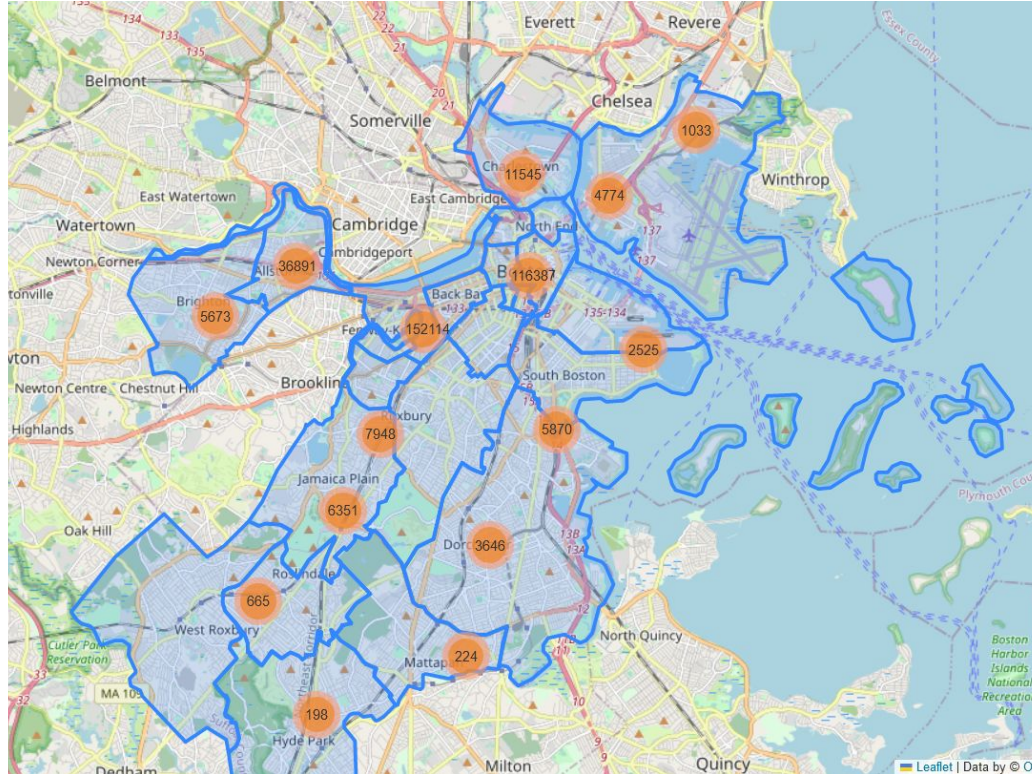
Extension Proposal

Extension Pitch	<i>With our original topic, we analyzed MBTA trip data and Boston census data to draw conclusions about service disparities in different neighborhoods. For our extension, we will analyze potential disparities in Bluebike availability by neighborhood, and determine if there are enough stations to meet potential demand from different communities.</i>
Rationale	<i>After analyzing availability of MBTA bus data, it is important to consider alternative transportation options. Bluebikes are a public option that people can rent at an affordable price.</i>
Questions for Analysis	<i>We are curious about the number and locations of different stations, whether they are dispersed evenly throughout the city and/or concentrated in areas of high-population density, or if there are some areas with a lack of stations. H</i>
Datasets & Sources	<u><i>Bluebikes Stations</i></u> - https://s3.amazonaws.com/hubway-data/current_bluebikes_stations.csv <u><i>Blue Bikes Comprehensive Trip Histories</i></u> - https://s3.amazonaws.com/hubway-data/202310-bluebikes-tripdata.zip
Data Visualizations	<i>Proposed graphs include:</i> <ul style="list-style-type: none">• <i>Map of locations of each station within Boston area</i>• <i>Bar graphs of number of stations per neighborhood</i>• <i>Clustering of stations and/or trip data</i>
Additional Information	<i>Information on Bluebikes data - https://bluebikes.com/system-data</i> <i>Example references:</i> <ul style="list-style-type: none">• https://cambridge-intelligence.com/geospatial-data-visualization-regraph-redwoodjs/• https://www.chaossearch.io/blog/blue-bikes-data-dive-part-1

Locations of Blue Bike Stations

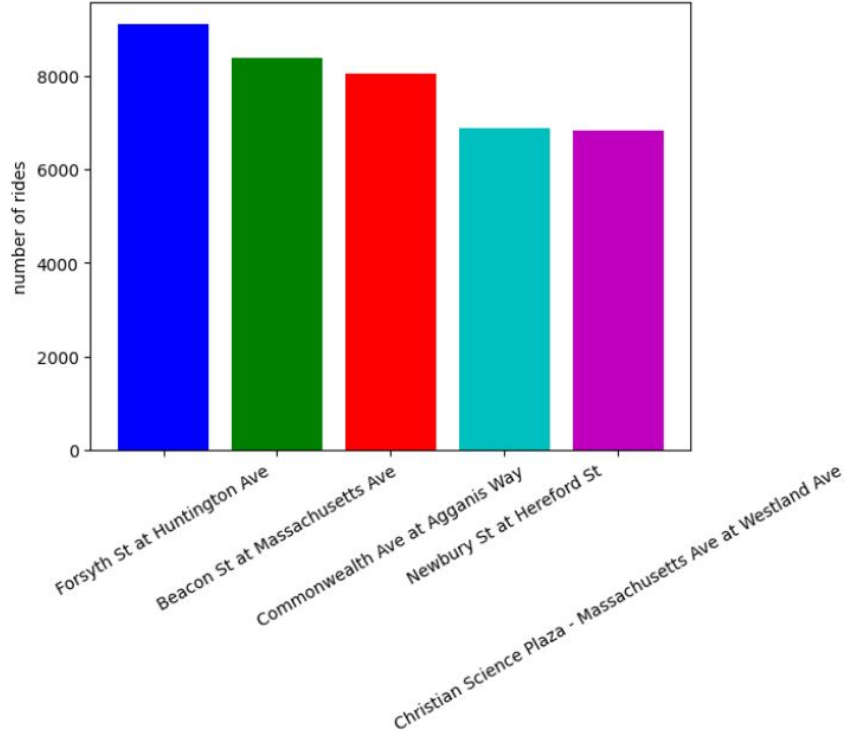


Locations of Blue Bike Rides

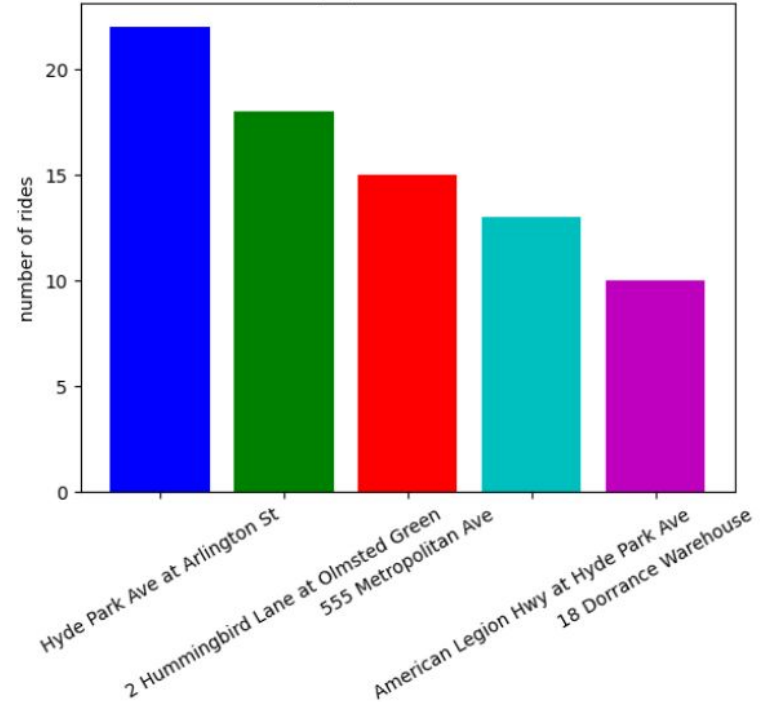


Most and least popular Blue Bike stations

Five most popular Blue Bike stations



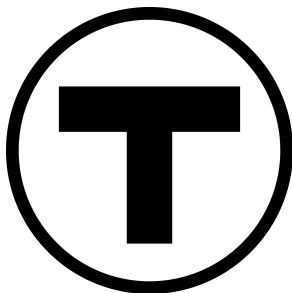
Five least popular Blue Bike stations



Extension Project Questions

Some of the questions we're considering for our extension project include:

- How accessible are Blue Bike stations?
- How many rides are made with Blue Bikes, and where do they start?
- Which neighborhoods have more access to Blue Bikes, and which ones have less?
- What are the demographics of people who have access to Blue Bikes vs. those who do not?
- Are the areas where people have less access to Blue Bikes the same as those where people have less access to MBTA buses?



BLUEbikes SM

Extension Project - Next Steps

- **Geographic information:** We will get the specific neighborhoods for each Blue Bikes station, rather than just the city that they are located in.
- **Demographic data:** We will join our station and ride data with the census data to obtain demographic and diversity statistics for Blue Bikes riders.
- **New data:** Our rides data is currently for the month of September 2022, as this month had the largest number of rides. We are planning to analyze different months outside of September 2022 to see if there are seasonal trends in the data that may affect our conclusions.
- **Rider status:** We will partition rides by subscribers vs. customers and analyze the ride data from that perspective. This will show which neighborhoods have the financial means to subscribe to Blue Bikes and thus take advantage of the program more.