Boston Bus Performance

Team F

Jialu Li, Junyi Li, Laksanawist Mutiraj, Qinfeng Li, Yifei Zhou

Project Background and Goal

- ***Background :***

  –Historical Roots: Public transport in Boston dates back nearly 400 years, with ferry services originating in the 1600s.
  –MBTA's Impact: Serving over 1 million people daily, the MBTA contributes an estimated $11.5 billion annually to the greater Boston area's economy.

- ***Goal:***

  - Examining MBTA bus service performance trends by geography with important insights into potential disparities by neighborhood and other demographic factors like race and income.

  - Explore the availability of alternative mobility service in the areas where MBTA services are the least reliable and among the demographic group most impacted by service level disparities.

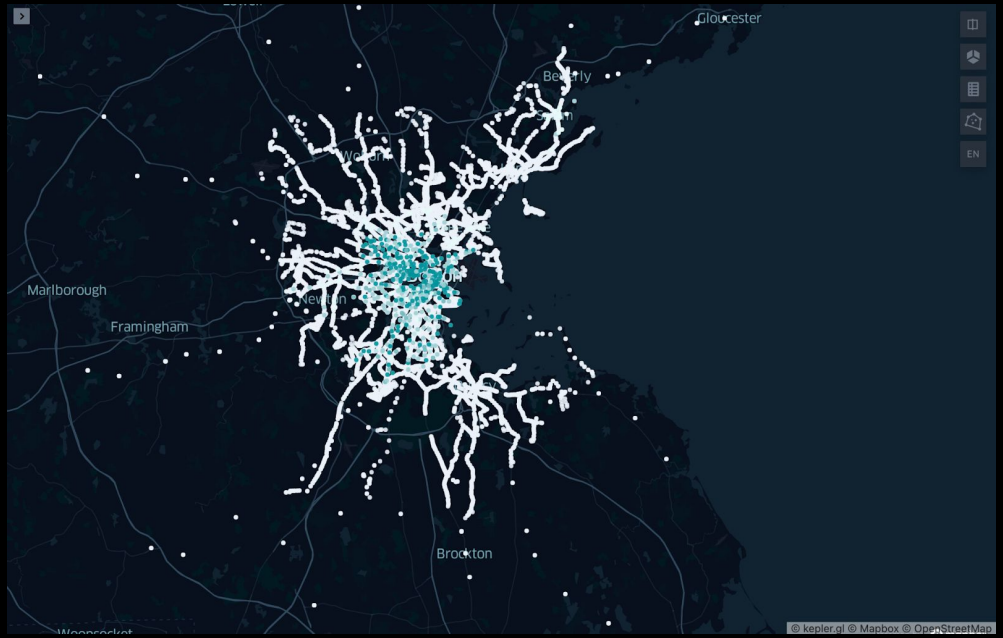
Motivation for Extension Project

Our project aims to enhance the understanding of Boston's transit system by incorporating factors like service disruptions and accessibility, especially for disabled individuals, thereby promoting a more equitable urban mobility landscape.

By analyzing the interaction between various transit modes, such as Blue Bikes and buses, and developing a delay prediction system based on route and temperature, we strive to directly benefit passengers and improve their daily commuting experience.

Extension Proposal	
Extension Pitch	We propose to extend our current geospatial analysis of Boston's transit systems by integrating additional datasets that consider factors like service disruptions and accessibility features. This effort aims to provide a more inclusive story of the city's transit dynamics, particularly focusing on the distribution and development of resources for disable accessibility around bus stops. We also propose to delve deeper into the intersection of public transit data, specifically focusing on Blue Bike and bus data in Boston. The goal is to uncover insights into how different modes of transit interplay and affect urban mobility. In addition to analysis, we decide to move further by doing a practical move, which is a delay prediction system based on route id and temperature inputs. With the use of this system, the passengers could gain real benefits.
Rationale	Our project aims to extend the geospatial analysis of Boston's transit systems, integrating data on service disruptions and accessibility, especially focusing on disabled access at bus stops. This approach seeks to provide a comprehensive view of the city's transit dynamics and improve urban mobility by analyzing how different modes, like Blue Bikes and buses, interact. Additionally, we plan to develop a practical delay prediction system based on route ID and temperature, offering direct benefits to passengers by enhancing their commuting experience.
Questions for Analysis	What are the accessibility gaps in the current transit network, and how might they affect riders with disabilities? What area does bus stops and blue bike stations covered, and how could that make blue bike a possible alternate for bus? How does the proximity of Blue Bike stations to bus stops affect ridership and transit efficiency? For two basic inputs, bus route id and temperature, how would a predicted delay minutes be?
Data Sets & Sources	Blue_Bike_Stations.geojson Community_Centers.csv PATI_Bus_Stops.csv Hospitals.csv Bus arrival and departure time 2022 Boston temperature data from 2013-2023
Additional Information	Our preliminary analysis has shown interesting trends in transit usage across different Boston neighborhoods. This extended analysis will provide a more comprehensive view, potentially guiding city planners and policymakers in enhancing urban transit systems.

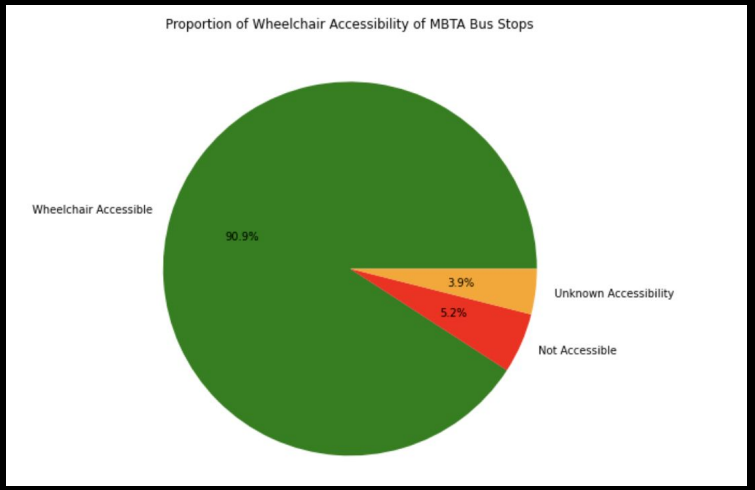
Blue Bike Stations & Bus Stops Dense map

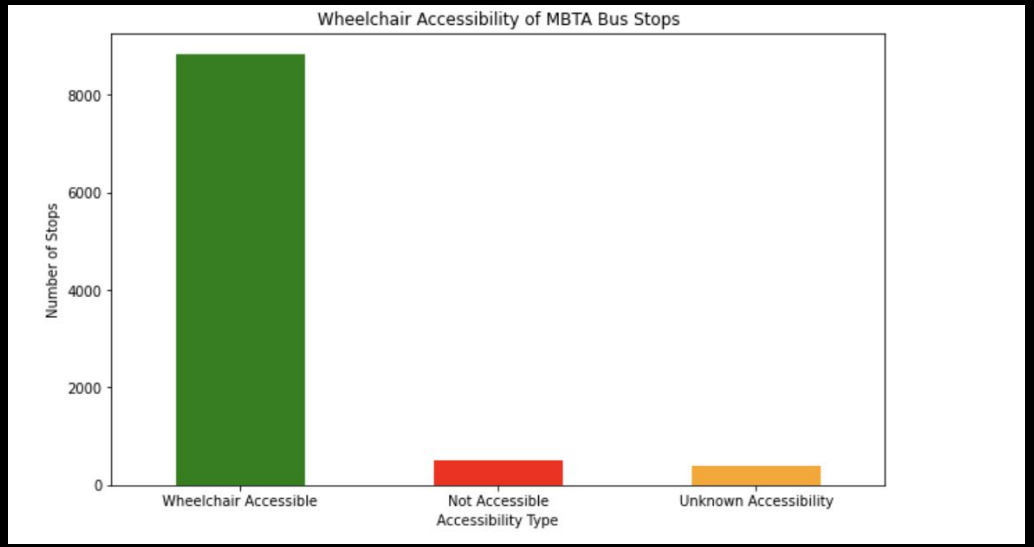


·white dots for bus

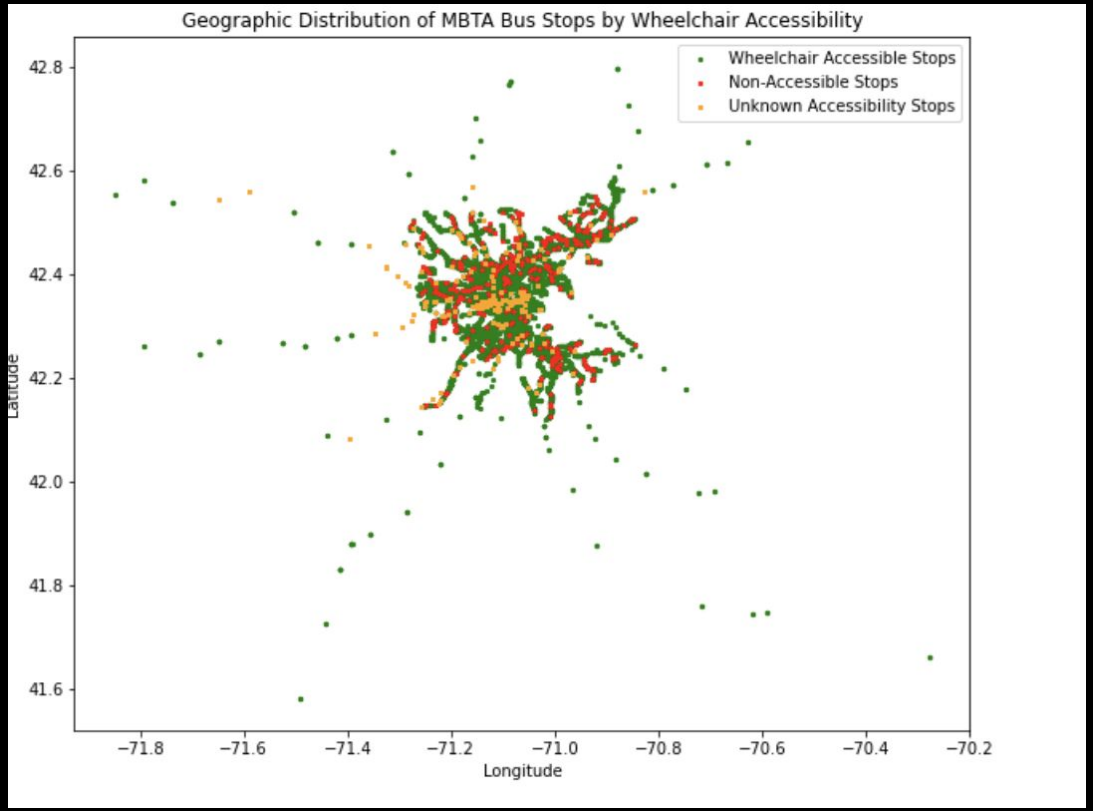
·the other color for blue bike stations

Wheelchair accessibility around bus stops

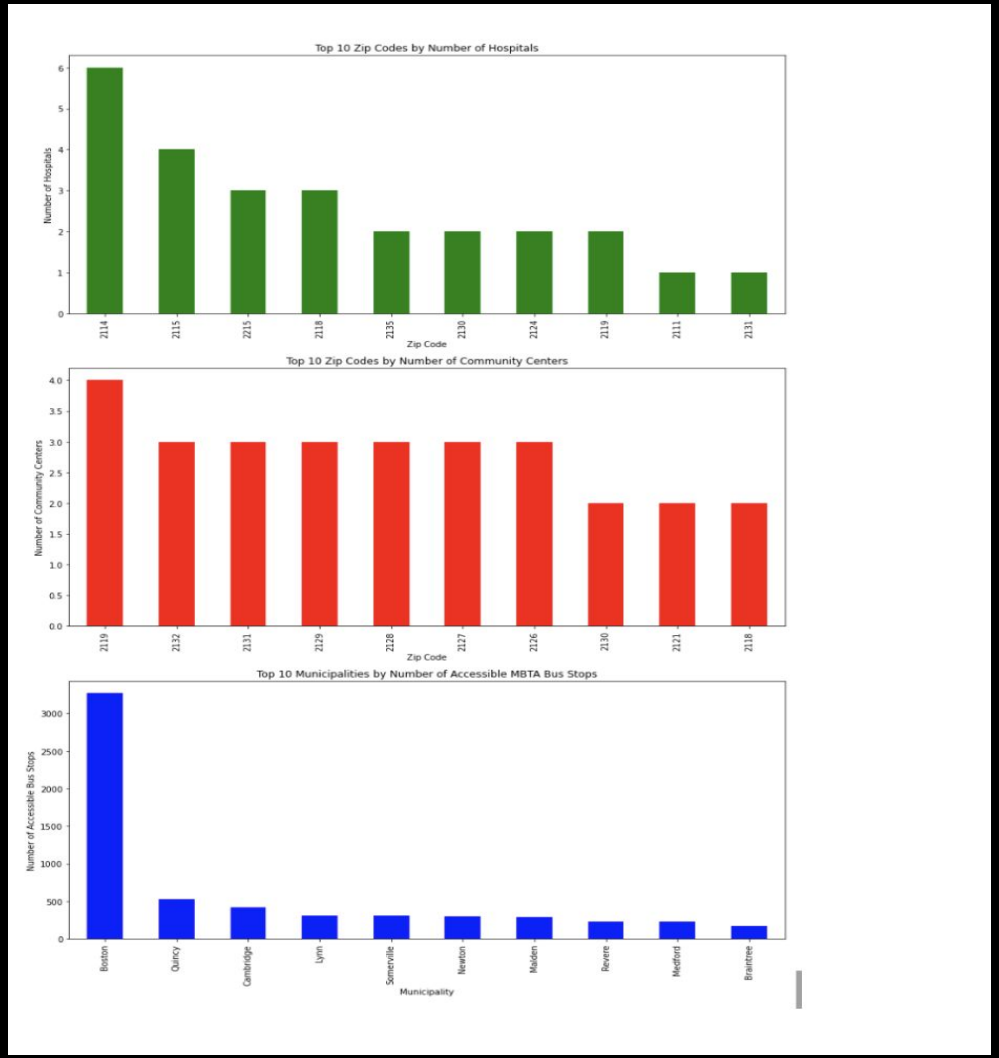




Geographic Distribution of Bus Stops by Wheelchair Accessibility



Analysis of Healthcare, Community, and Transit Accessibility by Location



Delay Prediction System based on temperature and bus route

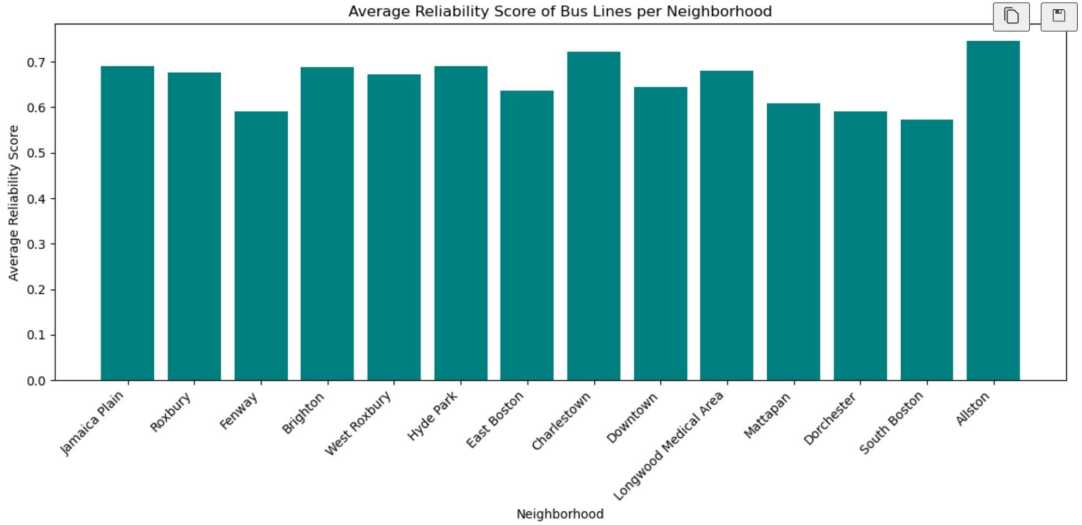
```python
# Load the weather data
weather_data = pd.read_csv('/Users/lijunyi/Downloads/cs506/bus performance/weather/boston_weather_data_2022.csv')

# Define the months to be included in the analysis
months = [1,2,3,4,5,6,7,8,9,10,11,12]

# Initialize an empty DataFrame for selected bus data
selected_bus_data = pd.DataFrame()

# Iterate over the selected months
for month in months:
    month_str = str(month).zfill(2)  # Ensure month is two digits
    file_name = f'MBTA-Bus-Arrival-Departure-Times_2022-{month_str}.csv'
    bus_data_month = pd.read_csv(file_name)

    # Convert scheduled and actual times to datetime
    bus_data_month['scheduled'] = pd.to_datetime(bus_data_month['scheduled'], format='%Y-%m-%d %H:%M:%S.%f')
    bus_data_month['actual'] = pd.to_datetime(bus_data_month['actual'], format='%Y-%m-%d %H:%M:%S.%f')

    # Calculate delay in minutes and handle NaN values in the 'delay' column
    bus_data_month['delay'] = (bus_data_month['actual'] - bus_data_month['scheduled']).dt.total_seconds() / 60

    # Append to the main DataFrame
    selected_bus_data = pd.concat([selected_bus_data, bus_data_month])
```

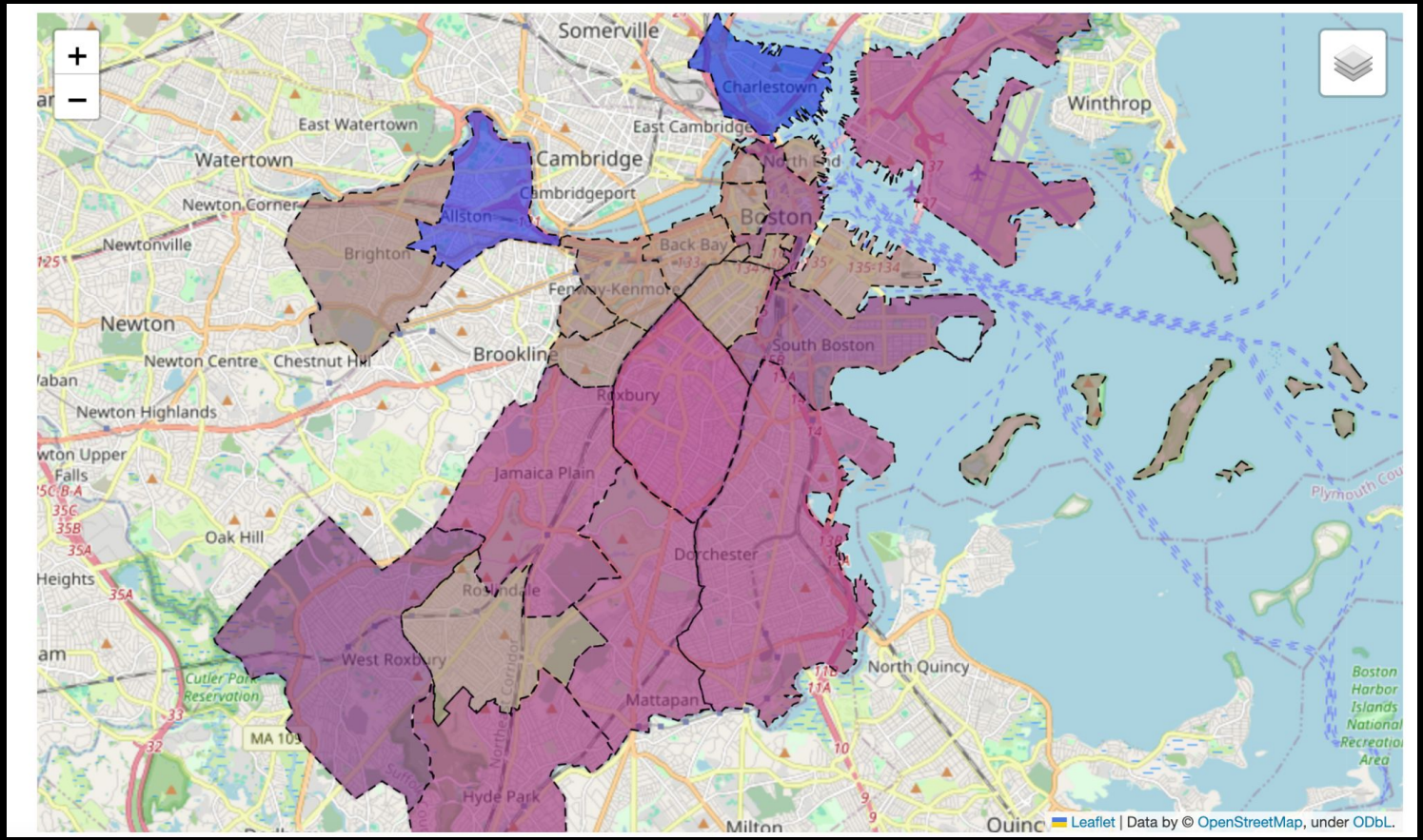
```python
# Example usage:
route_id_input = '57'  # example route_id as string
temperature_input = -10  # example temperature in Celsius

# Predict delay
predicted_delay = predict_delay(route_id_input, temperature_input)
print(f"The predicted delay for route {route_id_input} at {temperature_input}°C is approximately {predicted_d
```

```
The predicted delay for route 57 at -10°C is approximately 5.46 minutes.
```

Reliability vs Income

	Neighborhood	Income	Average Reliability Score
0	Allston	64748.0	0.746070
1	Charlestown	128403.0	0.722110
2	Dorchester	59209.0	0.591057
3	Downtown	119716.0	0.644411
4	East Boston	60579.0	0.636689
5	Hyde Park	69262.0	0.690540
6	Jamaica Plain	66516.5	0.690790
7	Mattapan	55024.0	0.608524
8	Roxbury	32200.0	0.676490
9	South Boston	139868.5	0.572159
10	West Roxbury	103337.0	0.671909




Somerville
Charlestown
East Watertown
East Cambridge
Winthrop
Watertown
Cambridge
Cambridgeport
Newton Corner
Allston
Boston
Newtonville
Brighton
Back Bay
Fenway-Kenmore
Newton
Newton Centre
Chestnut Hill
Brookline
South Boston
Roxbury
Newton Highlands
Jamaica Plain
Oak Hill
Dorchester
Roxbury
West Roxbury
North Quincy
Cutler Park Reservation
Mattapan
MA 103
Hyde Park
Milton
Boston Harbor Islands National Recreation Area
Leaflet | Data by © OpenStreetMap, under ODbL.

Challenge & Limitation

- There only exists detailed and public data for temperatures but not weather. So we can only make the prediction based on temperature.
- For Departure and Arrival time, we cannot use the data from 2023 since data of several months has not been updated.
- The ridership data doesn’t include demographic data such as race of passenger, making it hard to analyze which demographic group is impacted the most.

Individual Contribution

Junyi Li: Analysis of Blue Bikes, Disability accessibility and construct the prediction system

Jialu Li: Analysis on income and service reliability

Yifei Zhou, Qinfeng Li, Laksanawisit Mutiraj: contributed to raise the extension questions that we are investigating now