Gabriel Bernardes-Louis - Final Individual Project

(COMP3125 Individual Project)

Gabriel Bernardes-Louis   
*Wentworth Institute of Technology - Data Science Fundamentals - Final Individual Project (Section 2)*

# **I.Introduction (*Heading 1*)**

The MBTA is well known, at this point, for having very unreliable and slow service. While this has been changing over the past year and things are starting to improve, this project will be looking at where the MBTA is still lacking and how it can improve an aging system with years of deferred maintenance and a stubbornness for the old ways in a city that is always changing. By looking at the demographics and areas served the project will attempt to find the underserved communities and attempt to find solutions to the lack in service. While also discussing what demographics use the MBTA’s current services and try to address the needs of the people who use the MBTA’s services the most. This is a vital and necessary project as the MBTA has been putting off doing a major system route redesign for decades and the city and its surrounding neighborhoods have changed very much since the bus and subway routes were first established. This will allow the MBTA to better serve communities that may have not been an idea in the head of the officials who designed the routes decades ago. The project will mainly look at data published publicly by the MBTA to attempt to solve these problems.

# **II.Datasets**

## ***A.Source of dataset***

I will be using data from the MBTA Website. It was created for public and internal use by the MBTA to help track service patterns, delays, and frequencies. These data sets look at the total number of system trips per month as well as the towns served by the MBTA. These data sets have been updated continuously for at least a year in most cases. They were generated by taking raw data from the MBTA and translating it into visuals for better representation and readability.

## ***B.Character of the datasets***

The format of the dataset is csv files that are over 100 MBs. They include data such as what station the trip originated from as well as what time, day, and what line the station is on. As well as how many trips originated from that station. I cleaned the data to only include the date and number of trips per month instead of per day.

# **III.Methodology**

I used three different methods to analyze this data. These methods were used through python, pandas, and matplotlib. I first had to download the raw csv files from the MBTA resources website and upload the file path to the python file i was writing the code in.

## ***Data Loading***

I first used pandas to load the dataset (MBTA\_Gated\_Station\_Entries.csv) into dataframes. This allows for efficient data manipulation and analysis.

## ***Data Aggregation***

## I next grouped and summed data using groupby in pandas to aggregate the total ridership or entries by specific periods. Then I extracted specific time-based features like month\_year from datetime columns using pandas' datetime capabilities.

## 

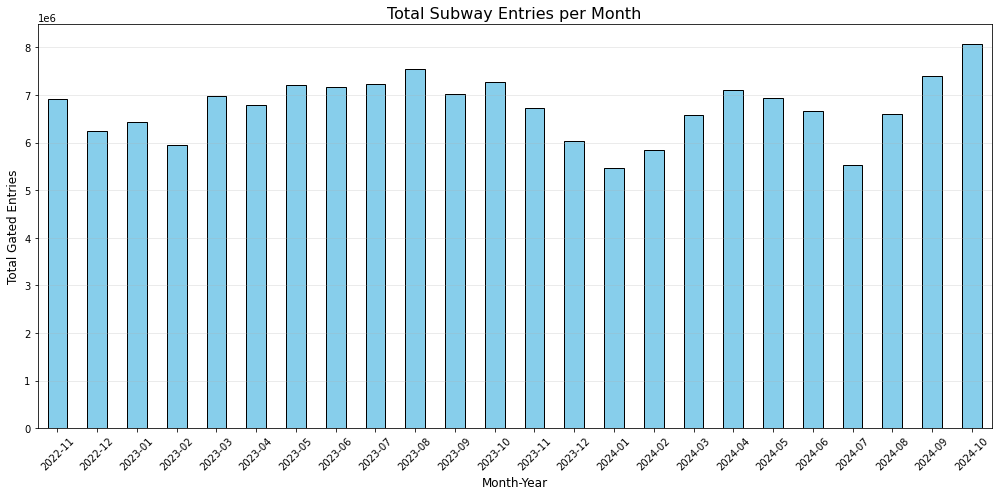
## ***Visualization***

Used matplotlib.pyplot as plt to create bar charts and line plots, showcasing trends and distributions in the data. Enhanced plots with titles, axis labels, and gridlines for better readability and interpretation.

# **IV.Results**

After my analysis my results were as expected. Overall the MBTA has a steady ridership number. And services a large area of greater Boston. When comparing these results to some historic data it is clear that the MBTA is still lacking in service compared to its past and what it requires, but it is slowly improving.

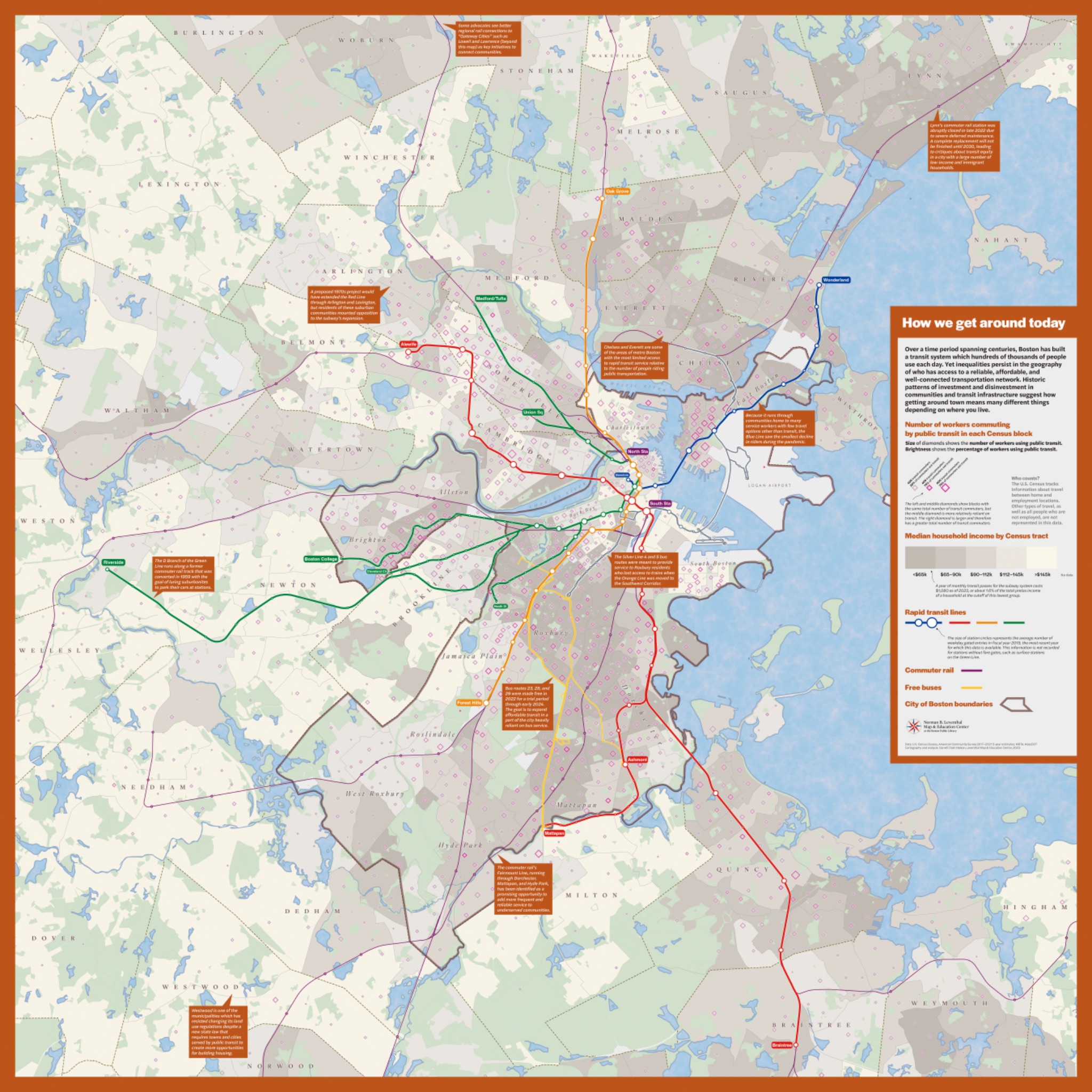
## ***A.Result A: MBTA Has steady ridership***



* The MBTA has kept a steady number of riders when looking at the average ridership per month over the past 2 to 3 years. With the reduced slow zones the MBTA saw its highest monthly trip number in three years this fall. A good improvement
* The trend seems to be that the ridership is going back up slowly, however it is still far from its pre-COVID-19 ridership numbers.

## ***B.Results B: The MBTA Serves a Large Community***

* The MBTA serves a large number of communities around the Greater Boston Area.
* While increased access and frequency are always good to have, the MBTA currently does a good job reaching most towns in a 25 mile radius.



<https://collections.leventhalmap.org/search/commonwealth:0k228v82p>

# **V.Discussion**

While this project was able to find a lot of recent data to analyze, it would have been very nice if I had more time, to be able to go back and look at historical data over the past few decades and look at ridership trends and compare them with events across the MBTA system to see what spiked the most ridership and see how we could recreate that now to help improve the system even more. This would've been even more difficult as the data set may not exist digitally and I may have to go physically to the Massachusetts Archives and look at the data on paper to see this.

# **VI.Conclusion**

In conclusion, my project showed that the MBTA has proven to be able to return to reliable service with great affects on its riderships finally treading towards increased ridership within the next couple years. Also looking at the ridership population and demographics it is clear that the MBTA should continue to improve the frequency of trains and plan more expansions, but they do cover a large area of Eastern Massachusetts already. With more access to the west of the state, increased frequency/relaibalty, and newer/improved technology the system is finally seeming to come into the 21st century full steam ahead.

##### **References**

1. <https://collections.leventhalmap.org/search/commonwealth:0k228v82p>
2. <https://www.mbta.com/performance-metrics/ridership-the-t>
3. <https://www.mbta.com/performance-metrics/speed-restrictions>
4. <https://dashboard.transitmatters.org/>
5. <https://www.mbta.com/projects>
6. <https://www.mass.gov/info-details/mbta-communities-law-qa>