CITY OF BOSTON: TRANSIT AND PERFORMANCE - TEAM C
DELIVERABLE 1

Analysis of MBTA Bus Performance and Demographics in Boston

Introduction

Public transportation plays a vital role in cities like Boston by providing mobility and access to jobs, services, healthcare, education, and amenities. However, historically not all communities have been served equally by public transit, often due to inequitable urban planning and infrastructure investments. This report aims to analyze the performance of MBTA bus routes across Boston in relation to the socioeconomic demographics of the communities they serve. The goal is to identify any disparities in service levels or quality and assess whether they may disproportionately impact marginalized groups.

Specifically, this analysis will focus on examining MBTA bus operational data from January 2022, alongside Census demographic data aggregated at the neighborhood level in Boston. Key questions to be explored include: What are the end-to-end travel times for different bus routes across the city? Are there noticeable disparities in on-time performance and reliability between routes? What are the racial, economic, and age demographics of the communities primarily served by each bus route? If clear service disparities emerge from the data, do they appear to adversely and disproportionately impact disadvantaged communities?

To answer these questions, the report will utilize data science techniques including data collection APIs, cleaning and joining of datasets, exploratory data analysis, statistical summaries, data visualization, and geographic mapping. Findings will provide critical insights into potential service inequities along demographic lines. Recommendations will also be made on how to focus future transit investments to improve equity in service quality and accessibility for marginalized groups.

Public transit equity has broad implications for social justice, economic mobility, and environmental sustainability in diverse cities like Boston. This data-driven analysis aims to provide an objective, comprehensive understanding of where service gaps exist across bus routes, who they impact, and how they can be addressed. The goal is to promote discussion and progress towards an affordable, reliable transportation network that connects and serves all Boston communities.

Data Collection and Cleaning

The first step in the data cleaning process was to join and filter the MBTA data to focus only on January 2022. This involved accessing the MBTA API to get real-time location, route, and schedule data for all bus trips during that month. The raw data was then filtered and aggregated to create a consolidated dataset for analysis.

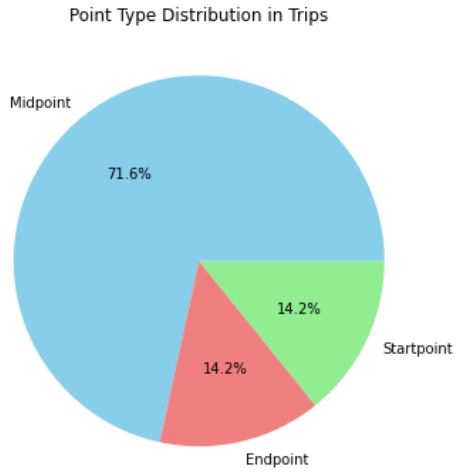
Next, key metrics like route travel times and on-time percentages were calculated. The scheduled and actual departure/arrival times in the MBTA data were used to compute total travel times for each route and trip. These travel times were averaged to get the mean time for each route. The scheduled versus actual times were also compared to calculate on-time performance.

When answering future questions, we will use spatial joins to associate the bus routes and stops with demographic data from the 2020 Census. We will first identify the census tracts that each stop is located in, and then we will link the tract-level population statistics to the stops and routes. This will allow us to append useful demographic data like income levels and race/ethnicity percentages to the MBTA data.

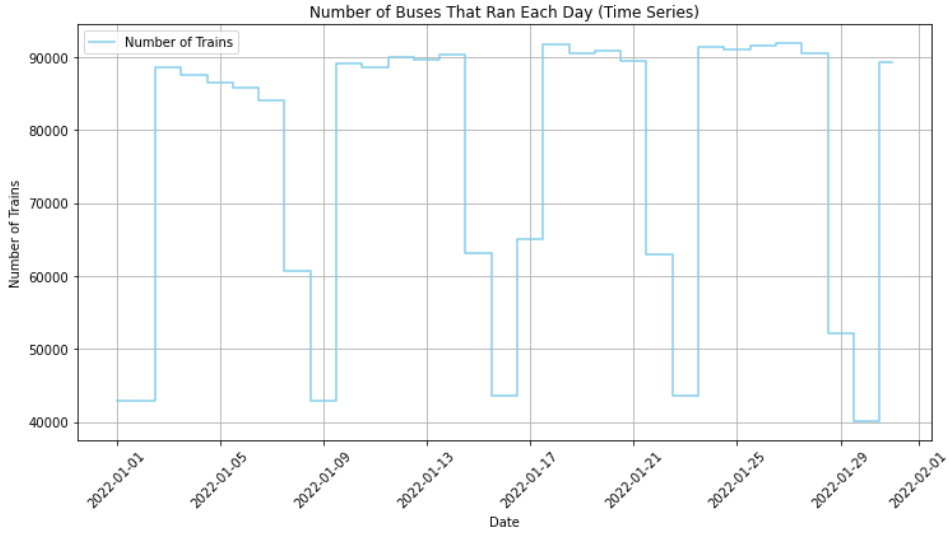
Our MBTA dataset was checked for quality and missing values were handled prior to analysis.

Exploratory Data Analysis

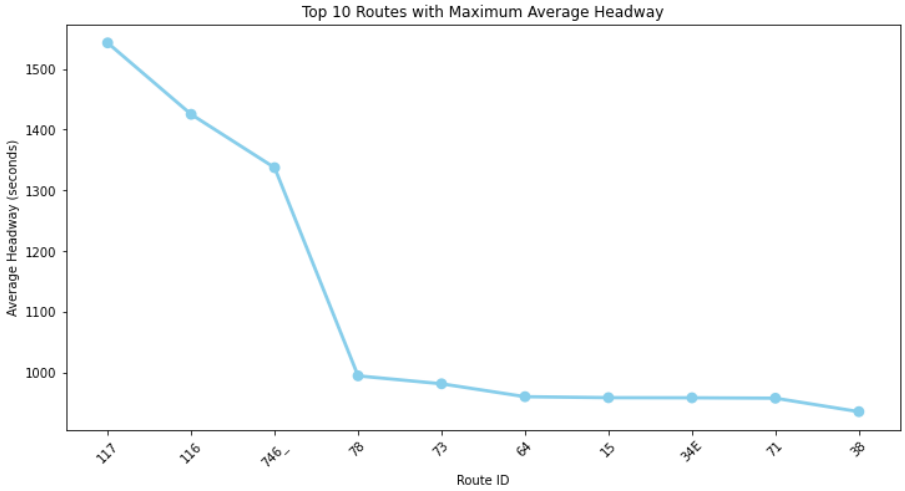
During our EDA, we examined several aspects of our cleaned MBTA dataset, including various distributions of point types in routes, the bus volume throughout the year, the headway of different bus routes, and frequencies of different bus lines.



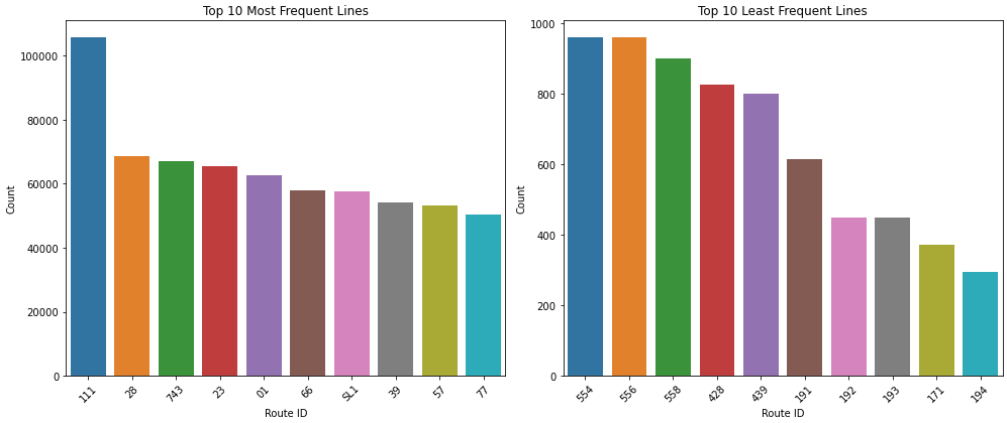
The graph above charts the proportions of measurements that are classified as either starting point, end points, or neither (midpoints). From this, we can see that over 28% of our collected data comes from either a starting or ending location. This possibly varies depending on the route (i.e., shorter routes may have less endpoints); however, it is possible that we have the same number of midpoints (5) for each route, with 1 start and 1 end point.



From this graph, we can see the total number of bus routes that were run each day on the MBTA for the month of January. We can see a clear pattern where more bus routes (~90,000) are run during the week than on the weekends, and that Saturday (~60,000) has more bus routes than Sunday (~40,000). We can also see that there is a holiday schedule, as Martin Luther King, Jr. Day had roughly the bus volume of an average Saturday. We can also see some variation between days; however, there is no obvious explanation for these discrepancies yet.



In this plot, we see the "maximum average headway" for the top 10 bus routes. Headway is defined as the "distance or duration between vehicles in a transit system measured in space or time." Thus, we see which routes have the most time between buses when run in this graph. It appears that three routes (the 117, 116, and 746) have far more time between them than all the other routes plotted.

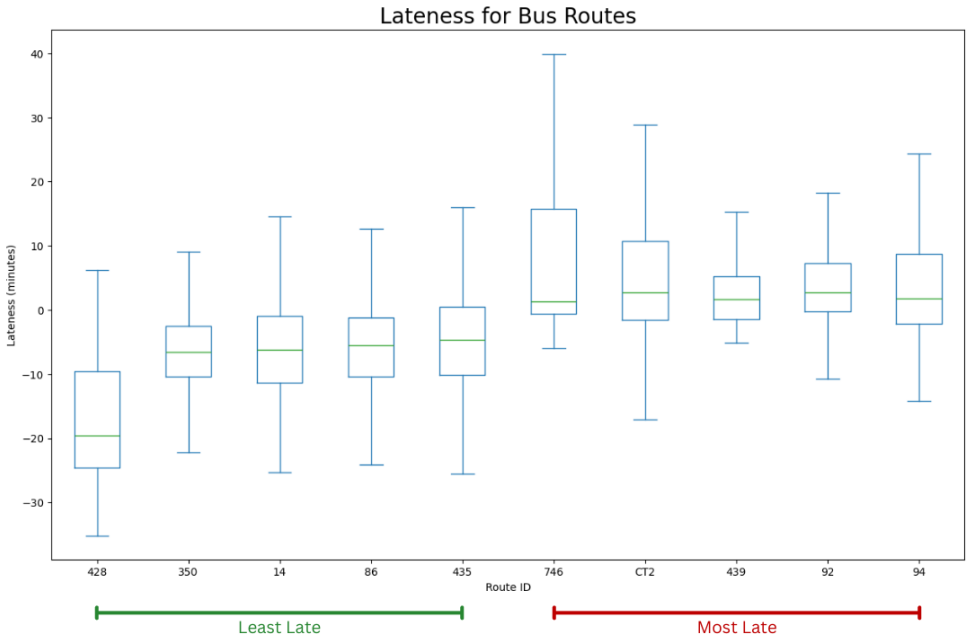


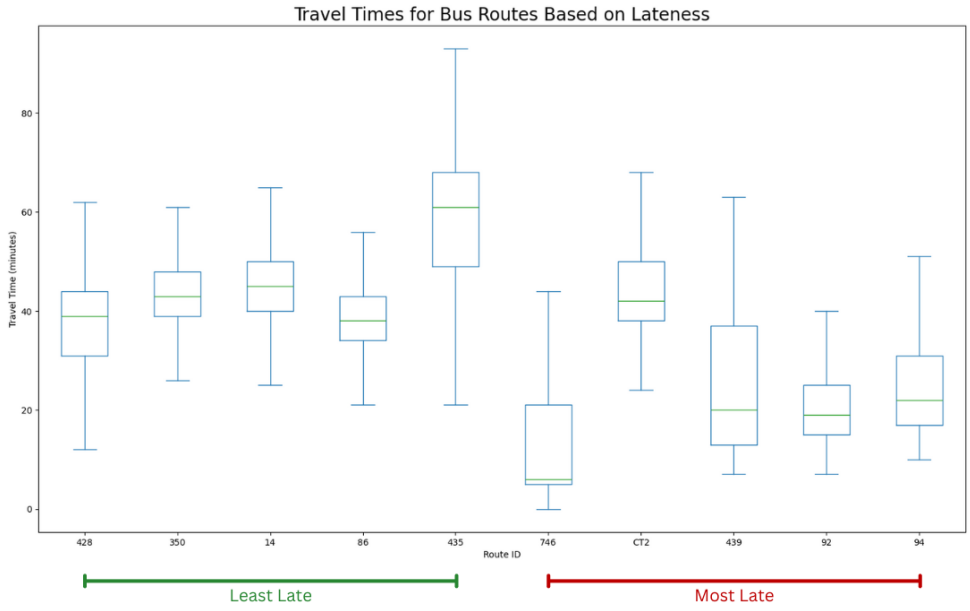
From these frequency plots, we can see the top 10 most frequent and least frequent lines by total count of routes run. We see that the "111" bus route is by far the most frequent, with a roughly 66% increase in total number of routes compared to the "28" route (the second-most frequent). This makes sense; upon looking at the MBTA website, we see that this route has scheduled departures every 5 minutes. This may be because there are less buses running in this area.

Additionally, we see a frequency plot of the least frequent bus routes. The 192, 193, and 194 routes are listed as least frequent, only being run 350-400 times total. Upon searching the MBTA website, these routes are not listed, indicating that they may have been removed for their lack of frequency (and passengers). The "171" route is a special low-service route to Logan Airport.

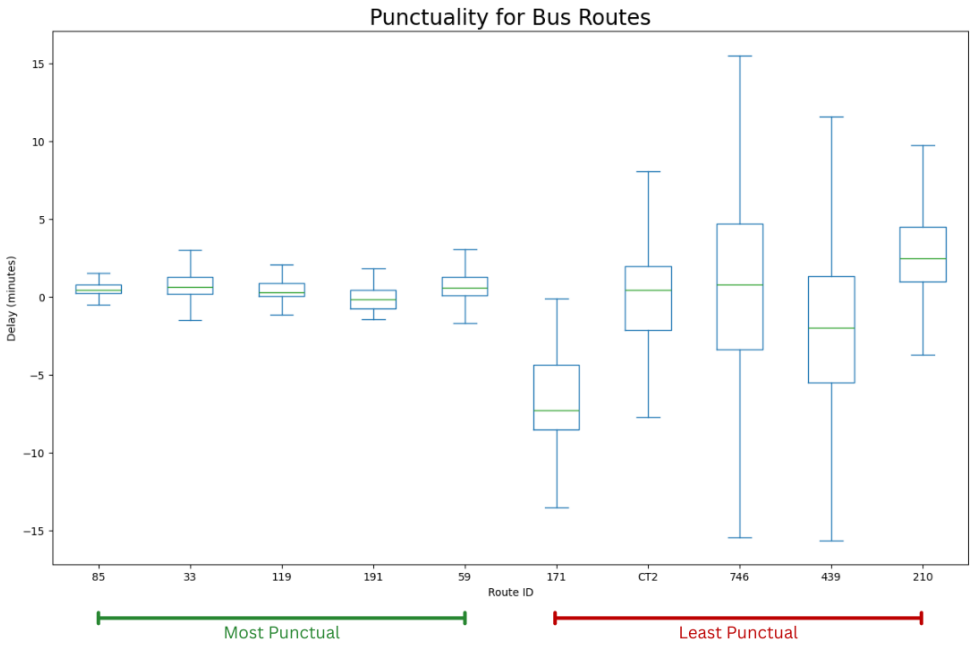
Preliminary Findings

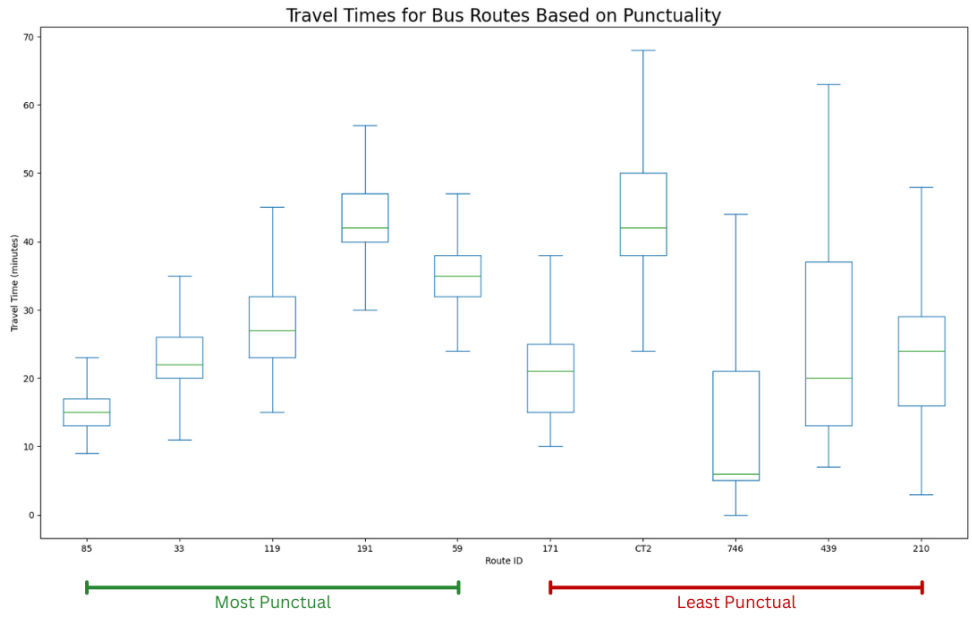
Here, we can start to view images of our preliminary findings. Graphs of lateness measure the difference between scheduled end-to-end travel time and actual end-to-end travel time. Negative values are good as they represent actual travel times that were less than the scheduled ones.





Graphs of punctuality measure the difference between the scheduled departure from the start point and the actual departure from the start point. Negative values and positive values are bad as they represent buses that departed ahead of or behind schedule. Values close to 0 are good as they represent buses that departed as per the schedule.





Base Questions Answered

1. What are the end-to-end travel times for different bus routes?
2. Are there disparities in the service levels of different routes? (which lines are late more often than others?)

Next Steps

For further analysis, we may look at evaluating service quality by time of day and day of week. The preliminary analysis looked at aggregate on-time performance, but breaking this down by peak versus off-peak times or weekends could reveal variations. The MBTA data includes timestamps that could be used to segment performance. Comparing ridership volumes during different time periods would also show if service quality aligns with demand.

Additionally, linking the MBTA data to census demographics could provide insights on the populations impacted by service disparities. The census data for Boston neighborhoods' income levels, age distribution, and other attributes can be connected to the bus routes serving each area. This will help answer questions about whether lower-income or minority areas receive lower bus service quality. Further analysis could focus on areas not served by buses or trains.

Finally, evaluating fare revenue and ridership by route would give a sense of the financial impacts of the service variations. Routes with more riders and revenue may warrant greater priority for improvement. Linking ridership to on-time performance could also show if delays discourage use of certain routes. Overall, these next steps will provide a more thorough understanding of who depends on MBTA buses and how service quality affects different communities.

Conclusion

This preliminary analysis indicates potential disparities in MBTA bus service that adversely affect marginalized communities. Additional data collection, modeling, and investigation of root causes is needed. Improving service equity will require considering demographic factors in transit planning and investments. Ongoing analysis can guide data-driven decisions to ensure affordable, reliable public transportation for all Boston residents.

Public transportation is vital for cities, but historically underserved communities exist. This analysis of MBTA bus performance will aim to identify service disparities and demographic

impacts in Boston. Using January 2022 MBTA data alongside census demographics, we will explore key questions as our project progresses.

So far, we have calculated route travel times and on-time percentages to assess performance variations. Preliminary findings have included evaluating the lateness and punctuality of different bus routes to answer our first two base questions. Further analysis is required to confirm findings and pinpoint causes. Next steps will segment data by time-of-day, connect census factors, and evaluate ridership and revenue differences. The goal is a comprehensive understanding of service gaps and community impacts. Recommendations can then guide investments to improve affordability, accessibility and reliability for all Boston residents.

Transportation equity has broad social justice and mobility implications. While more work is needed, this project exemplifies fact-based, data-driven solutions to serve diverse communities equitably. We hope it spurs discussion and progress towards connecting people across Boston.

Team Member Contributions

Chandrahas and Manushi : Performed data cleaning, pre-processing and EDA. Worked on generating graphs to better understand the trends and patterns in the dataset.

Eason, Patrick and Munir : Worked on addressing the first two base questions, get visualizations to analyze results and formulate the report.