GBH Impact and Equity: Boston Bus Equity

By
Youran Geng (MSDS 2025, yg779xw@bu.edu)
Huihao Xing (MSDS 2025, huxing@bu.edu)
Mi-Ru Youn (MSDS 2025, miruyoun@bu.edu)
Chen Yu Liu (MSDS 2025, erioe@bu.edu)
Namika Takada (MSDS 2025, ntakada@bu.edu)

Analysis Paper
DS701 Fall 2024 at Boston University
BU Spark!

Table of Contents

Introduction

Project Overview:

Boston's public transportation has a rich history spanning nearly 400 years, beginning with ferry services in the 1600s. Today, the MBTA serves over 1 million passengers and contributes an estimated $11.5 billion annually to the Greater Boston economy. Public transit plays a crucial role in Massachusetts and Boston's quality of life, impacting economic development, environmental sustainability, and social equity. While the MBTA is a state agency, not managed by the City of Boston, the city can provide input on community-related decisions such as bus routes.

This project aims to deliver a detailed analysis of the Massachusetts Bay Transportation Authority (MBTA) bus performance between 2019 and 2023. The main focus is to compare the ridership and reliability data for each bus route between the years. This study will identify any areas where bus service could be improved to better meet community needs. By analyzing the performance for each route, this report can inform data-driven decision-making for policymakers and MBTA to enhance overall transit system performance to better serve Boston residents and communities.

Data Collection and Cleaning:

The datasets utilized for this project were sourced from the MBTA Open Data Portal. For our analysis, we focused on bus performance metrics, specifically ridership and bus arrival and departure times, from 2019 and 2022, while utilizing the 2023 System-wide Passenger Survey data and Livable Streets report to gain insights into the demographic characteristics of those most affected by bus service quality.

For ridership, our data came from the MBTA Bus Ridership by Trip, Season, Route/Line, and Stop dataset. The file contains detailed ridership statistics for buses, focusing on average boardings, alightings, and passenger load. The data spans Fall 2016 to Fall 2022 and is organized by various metrics, including trips, seasons, routes, directions, stops, day types, and time periods.

Ridership was calculated by the MBTA through two ways: automatic passenger counts (APC) or automated fare collection (AFC). The APC system uses infrared beams and other technologies to count the number of passengers who enter and leave a vehicle, whereas the AFC counts the number of CharlieCard and ticket taps. In this dataset, ridership is defined as one ride per passenger on each trip. To ensure accuracy, the MBTA counts individual rides, including those involving transfers. This approach provides a more comprehensive picture of system usage. However, it's important to note that the data on boardings and alightings used to calculate ridership may be subject to collection biases, potentially affecting the overall accuracy of the ridership figures.

The data reveals potential inconsistencies in ridership measurement. For instance, some bus routes show higher boarding numbers than alightings, which theoretically should be equal. This discrepancy could be attributed to passengers not tapping their CharlieCards or tickets, leading to inaccuracies in the Automated Fare Collection (AFC) system.

Additionally, it's important to note that this dataset only covers the Fall period. As such, it may not provide a comprehensive representation of ridership patterns throughout the entire year. Seasonal variations in travel behavior, such as changes due to academic schedules or tourism, could significantly impact these figures.

Overall, the ridership dataset was very clean, requiring minimal preprocessing. We split the dataset into multiple subsets, and the only transformation involved was converting average statistics into total counts. To calculate total boardings and alightings, we multiplied the sample size by the average boarding/alighting values and added the results as new columns, 'total_boardings' and 'total_alightings'.

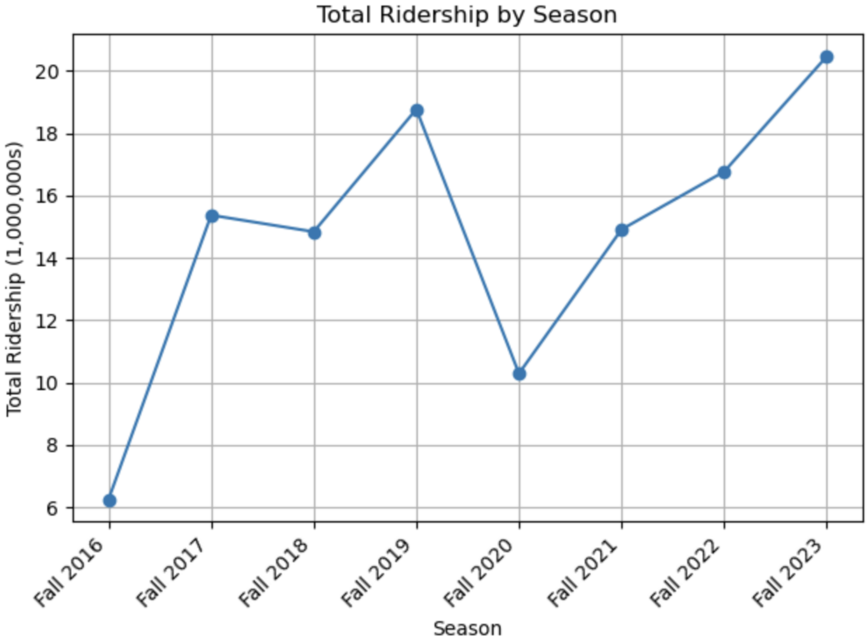
The bus arrival and departure data includes various lines and stations with scheduled and actual departure times, as well as an indicator of whether the trip should be evaluated based on schedule standards or headway standards. Four variables, “half_trip_id”, “time_point_id”, “time_point_order”, and “actual” contain missing values, but the missing portion is less than about 5%. Therefore, these missing values are removed.

We found that the actual departure sequence of buses on a route does not match the scheduled departure sequence. For example, as shown in Appendix B Figure 1, two rows represent the same bus stop on the same route for two consecutive trips. Based on the scheduled departure times, the row 1 bus should have arrived at and departed from the stop before the row 2 bus. However, according to the actual departure times, the row 2 bus arrived and departed first. This can lead to issues when calculating the waiting time at a station or for a route. If we use actual departure time minus scheduled departure time to compute latency, it may not align with passengers' actual waiting times. For example, based on the Figure , the calculated latency for row 1 would be 15 minutes, and for row 2, it would be 6 minutes early. However, for passengers waiting for the row 1 bus, they would actually end up boarding the row 2 bus, resulting in an actual waiting time of 12 minutes. For passengers intending to board the row 2 bus, they might not arrive at the station 3 minutes early as assumed, making it more likely that they would board the next scheduled bus instead.

Therefore we create a new variable called available_bus_depart_time which represents the earliest bus departure time available for passengers waiting at the scheduled departure time (Appendix A Figure 2).

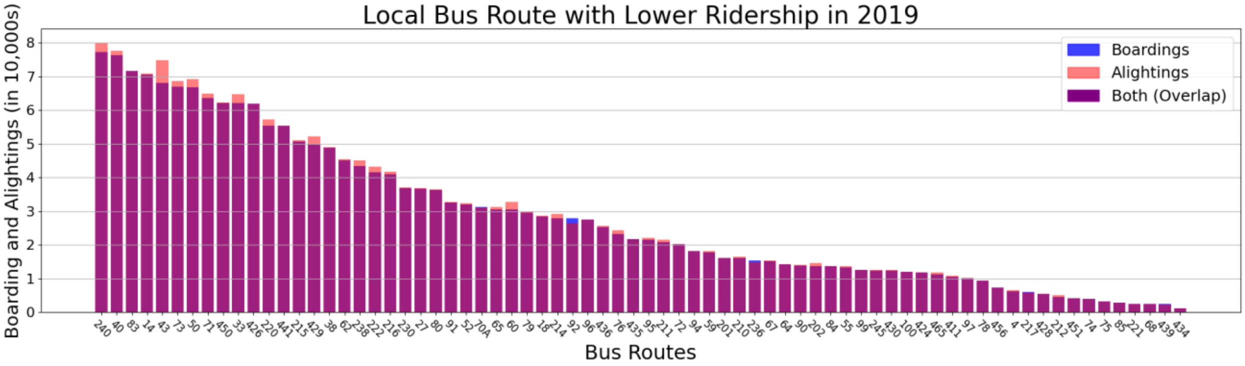
Data Analysis

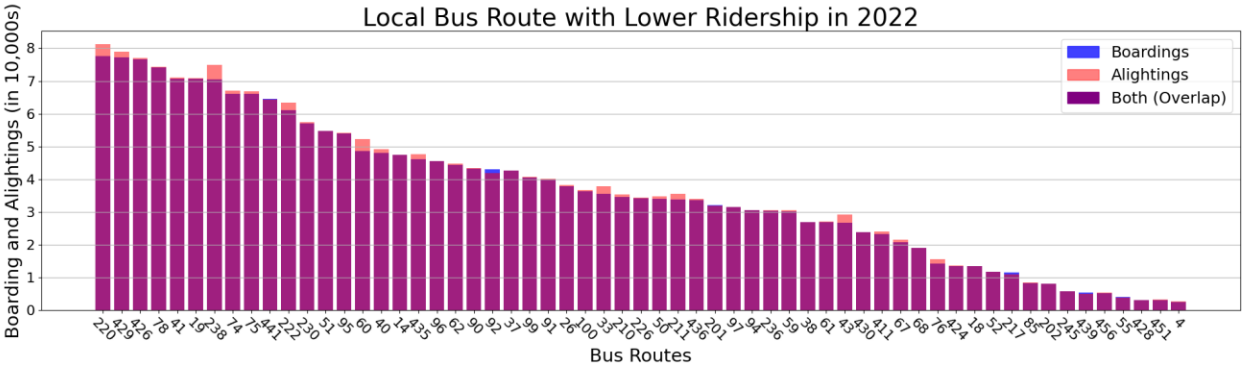
What is the ridership per bus route?

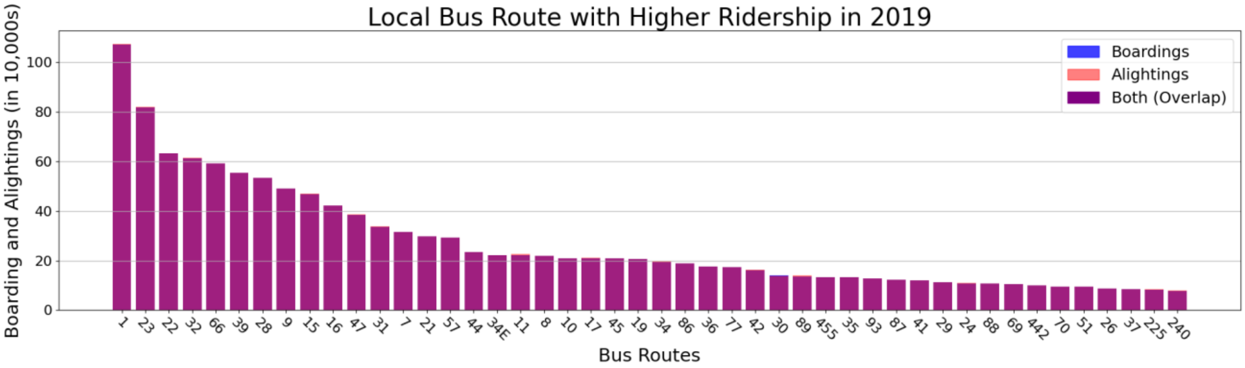


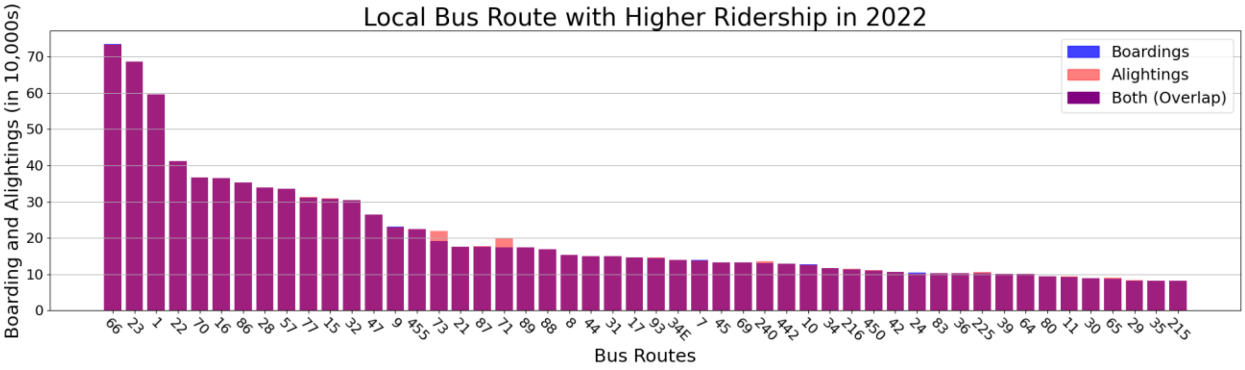
In our initial analysis, we examined the total ridership count for each subset and compared all the available years. The graph above illustrates that ridership steadily increased, first peaking in Fall 2019 at over 18 million boardings. However, a sharp decline occurred in Fall 2020, with ridership dropping to just over 10 million during the height of COVID-19. In the years following, ridership began to recover and eventually surpassed pre-pandemic levels, exceeding 20 million boardings. Because COVID-19 was a significant disruption, we decided to focus on the change between Fall 2019 and Fall 2023. These two years provide a more representative comparison of ridership trends as if the nation had not faced a pandemic crisis.

To further explore ridership patterns, we created bar charts to compare ridership counts per bus route between 2019 and 2022. To reduce clutter, we split local bus ridership into two groups based on a threshold of 80,000 total boardings. Bus routes under the 500 series, with some exceptions for the 300 series, were classified as local bus routes. As shown in the first four figures below, there are clear differences in ridership trends between 2019 and 2022. In 2019, ridership was significantly higher across most routes, with one route exceeding 100,000 boardings and alightings. However, in 2022, while ridership showed signs of recovery, the totals for higher-performing routes were still lower compared to 2019.

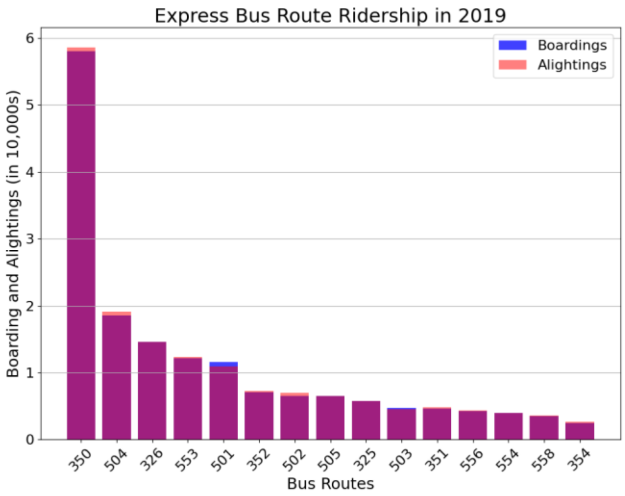

Local Bus Route with Lower Ridership in 2019
Boarding and Alightings (in 10,000s)
0
1
2
3
4
5
6
7
8
Boardings
Alightings
Both (Overlap)
240 40 83 14 43 73 50 71 450 33 426 220 441 215 429 38 62 238 222 216 230 27 80 91 52 70A 65 60 79 18 214 92 96 436 76 435 95 211 72 94 59 201 210 236 67 64 90 202 84 55 99 245 430 100 424 465 411 97 78 456 4 217 428 212 451 74 75 85 221 68 439 434
Bus Routes

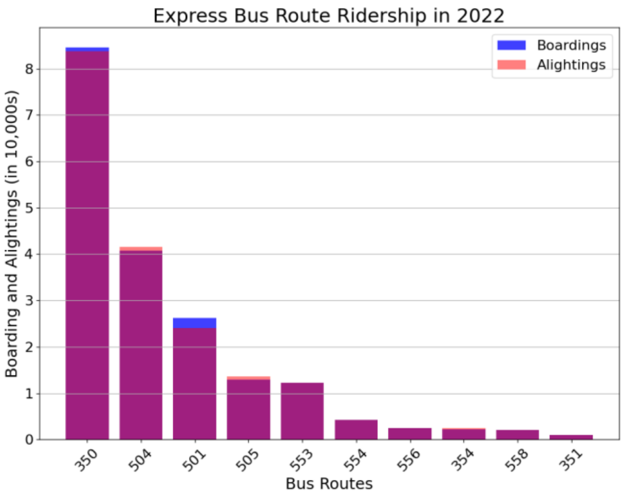

Local Bus Route with Lower Ridership in 2022
Boarding and Alightings (in 10,000s)
0
1
2
3
4
5
6
7
8
Boardings
Alightings
Both (Overlap)
220 429 426 78 41 19 238 74 75 441 222 230 51 95 60 40 14 435 96 62 90 92 37 99 91 26 100 33 210 226 50 211 436 201 97 94 236 59 38 61 43 430 411 67 68 76 424 18 52 217 85 202 245 439 456 55 428 451 4
Bus Routes


Local Bus Route with Higher Ridership in 2019
Boarding and Alightings (in 10,000s)
0
20
40
60
80
100
Boardings
Alightings
Both (Overlap)
1 23 22 32 66 39 28 9 15 16 47 31 7 21 57 44 34E 11 8 10 17 45 19 34 86 36 77 42 30 89 455 35 93 87 41 29 24 88 69 442 70 51 26 37 225 240
Bus Routes

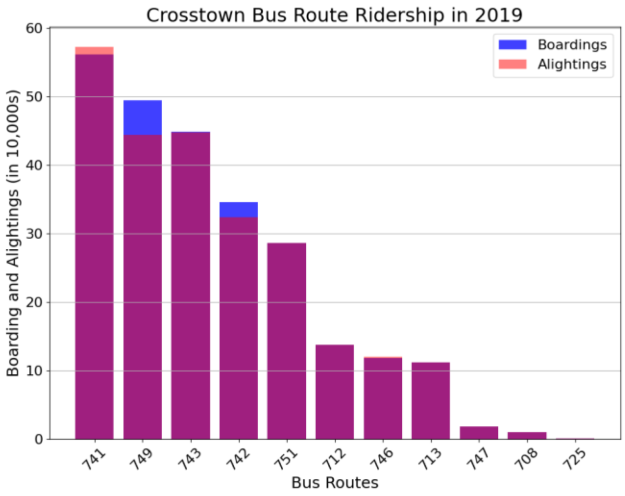

Local Bus Route with Higher Ridership in 2022
Boarding and Alightings (in 10,000s)
0
10
20
30
40
50
60
70
Boardings
Alightings
Both (Overlap)
66 23 1 22 70 16 86 28 57 77 15 32 47 9 455 73 21 87 71 89 88 8 44 31 17 93 34E 7 45 69 240 442 10 34 216 450 42 24 83 36 225 39 64 80 11 30 65 29 35 215
Bus Routes

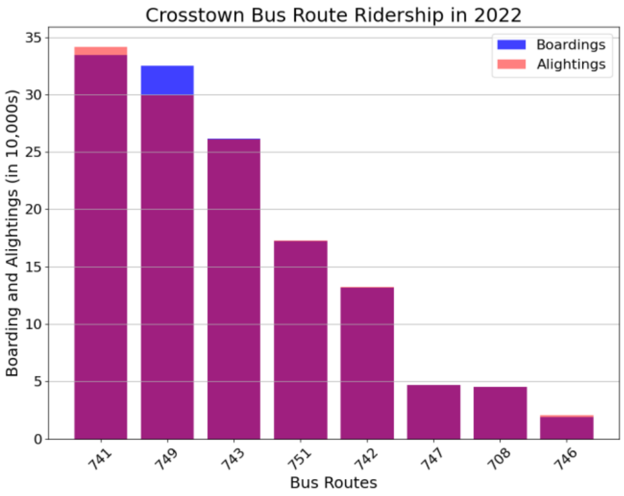
For express bus routes, as shown below, there were fewer routes documented, but overall, they experienced a notable increase in ridership compared to 2019 levels. Route 350 remained the top performer, while routes 504, 501, and 505 also showed some increase in ridership. However, the remaining bus routes either saw no significant change or experienced a decrease in ridership.



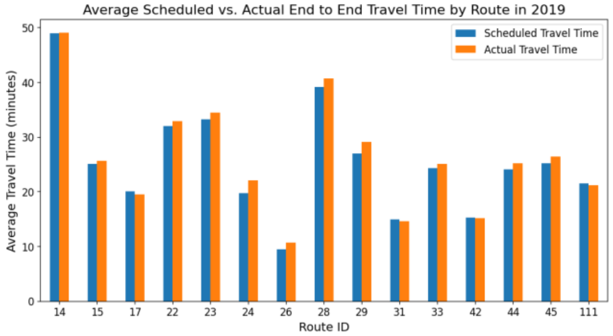


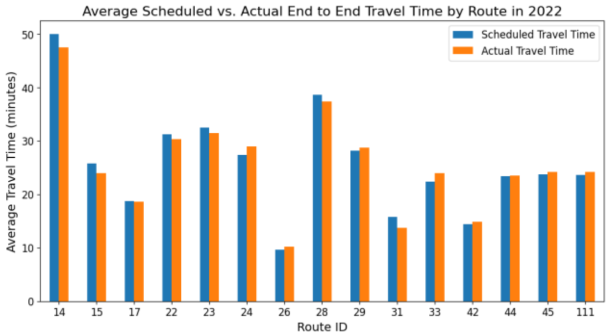
Crosstown bus routes, as displayed below, show an overall decrease in ridership in 2022 compared to 2019. Key routes, including 741, 749, 743, and 751, recorded lower total boardings and alightings. Additionally, the introduction of new routes may have contributed to this decline, as some older routes could have been replaced or made redundant by these newer services.





What are the end-to-end travel times for each bus route in the city?





Due to limited data on bus departure times, our analysis was limited to the bus routes shown in the charts above. End-to-end travel time was calculated by grouping each bus route by its service date and specific trip ID and finding the difference between the start and end times. Our analysis of average year-round scheduled and actual travel times indicates that most bus routes have experienced improved travel times. Bus Routes 23 and 28, specifically, each showed a decrease of approximately 3 minutes. However, there were two exceptions. Bus Route 111 experienced a 3-minute increase, while Bus Route 24 showed a more significant increase of 7 minutes in average travel time.

On average, how long does an individual have to wait for a bus (on time vs. delayed)?

The dataset of Arrival Departure contains the bus's scheduled departure time and actual departure time for individual stops. The datasets we are interested in are those from 2019 and 2022, as one is pre-pandemic and the other is post-pandemic. The insight of how pandemic has influenced the on time and delayed time for this specific problem plays an important role in our analysis.

Since there is no official definition for on time or delayed, we discussed with the client to consider buses with waiting times greater than positive or negative 60 seconds delayed. This buffer time also plays an important role in the following question, as some of the concerns are being raised. For example, the buses leaving before the scheduled time are also within the boundary of buffer time in reverse measure. However, assuming a passenger arrives at the bus stop on time, the passenger will not be able to catch the bus that is leaving earlier than the scheduled time, which will cause a prolonged delay for this passenger. Therefore, we have excluded the buffer time that is negative and adopted a new variable in cleaning. The waiting time for on-time buses and delayed buses are calculated by subtracting the time of the next available bus from the scheduled time from the cleaned dataset. This measure correctly calculates the waiting time for passengers who arrive on time.

Then, we are able to summarize the 2019 and 2022 datasets with the measure above. The average waiting time for on-time buses is 29 seconds(Appendix C, Figure A). This number is not surprising because we have manually set the buffer time as 60 seconds. The Waiting time is 437 seconds on average for delayed buses(Appendix C, Figure B). Also, the average waiting time for delayed buses in 2019 is lower for the first three months, about the same for May, and greater than 2022 for the rest of the month.

What is the average delay time of all routes across the entire city?

Due to the limits of computing resources, we discussed this with the client and decided to move to more specific routes for our analysis. The next question will cover more analysis of target routes listed in the requirement documentation.

What is the average delay time of the target bus routes (22, 29, 15, 45, 28, 44, 42, 17, 23, 31, 26, 111, 24, 33, 14 - from Livable Streets report)?

Following the measure we used for waiting time, we adopted the same calculation measure to get delayed time. The delayed time is calculated by subtracting the time of the next available bus from the scheduled time from the cleaned dataset. The only difference is that buffer time is no longer valuable in this question as we are focusing on delayed time.

The delayed time is aggregated by year and bus routes and sorted by 2019 bus count(Appendix C, Figure C). The left-skewed data distribution is not surprising since we sorted the data in descending order. However, a special route with a significantly high delay time is observed. The summary shows that Route 14 has an average delayed time of over 1,000 seconds for both years.

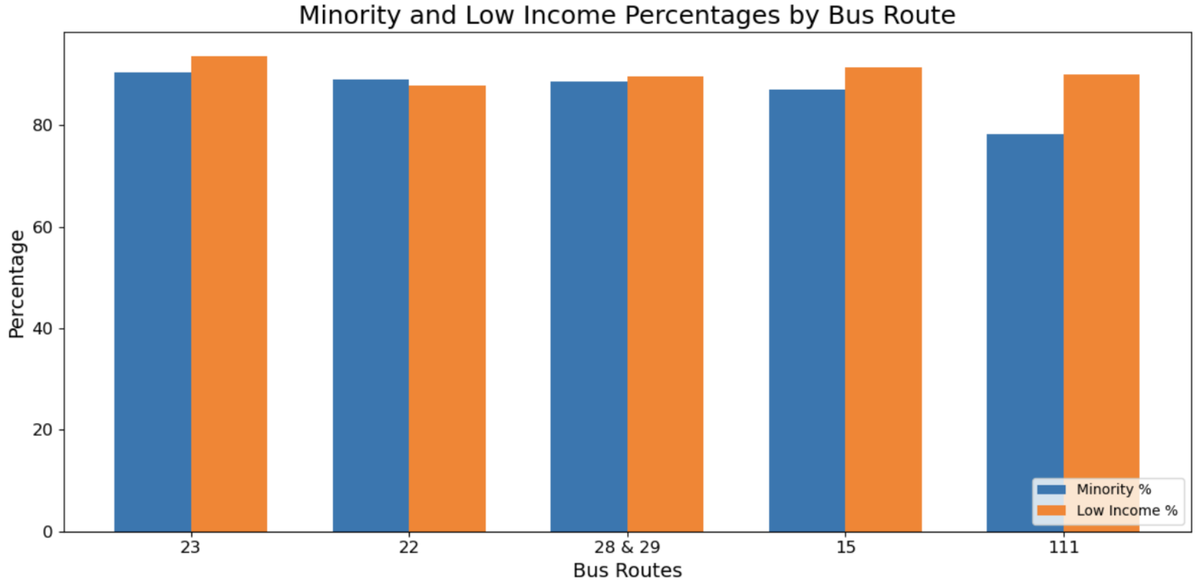
To understand how this occurs, we further examined the dataset and counted the number of bus routes to check if any outliers are influencing the analysis. After counting, route 14 appears to be a route with few bus schedules(Appendix C, Figure D). This increases the suspicion that the summary is influenced by outliers. Upon closer inspection of the outliers, some routes appear to have an especially huge gap between the scheduled time and the next available bus departure time(Appendix C, Figure E). Subsequent analysis of the original dataset revealed that Route 14, with its limited daily trips, contributes to extended wait times. For instance, missing a bus at 13:31 necessitates waiting until the next departure at 15:21(Appendix C, Figure F).

Are there disparities in the service levels of different routes (which lines are late more often than others)?

The percentage of delayed buses across different routes is aggregated to form a chart (Appendix C, Figure G). The differences between routes are not significant, and the level of service has not changed uniformly over the years. No clear evidence supports the claim that disparities exist in the service levels across different routes.

If there are service level disparities, are there differences in the characteristics of the people most impacted (e.g. race, ethnicity, age, income, etc.)?

Based on the data from the previous analysis, we observed that the characteristics of the most impacted bus routes were quite similar. However, a key limitation arose from the fact that many of the bus routes most affected by delays were not included in the 2023 System-Wide Passenger Survey, which provides demographic insights into MBTA ridership. As a result, we were only able to calculate the percentages of minority and low-income riders for the following bus routes: 23, 22, 28 & 29, 15, and 111.



Upon examining the data, it is evident that the bus routes most impacted by delays tend to serve higher percentages of minority and low-income populations. A closer look at the specific routes and the neighborhoods they traverse reveals that these delayed routes predominantly operate in high-demand transit areas and Environmental Justice Communities, which are often the most reliant on public transportation.

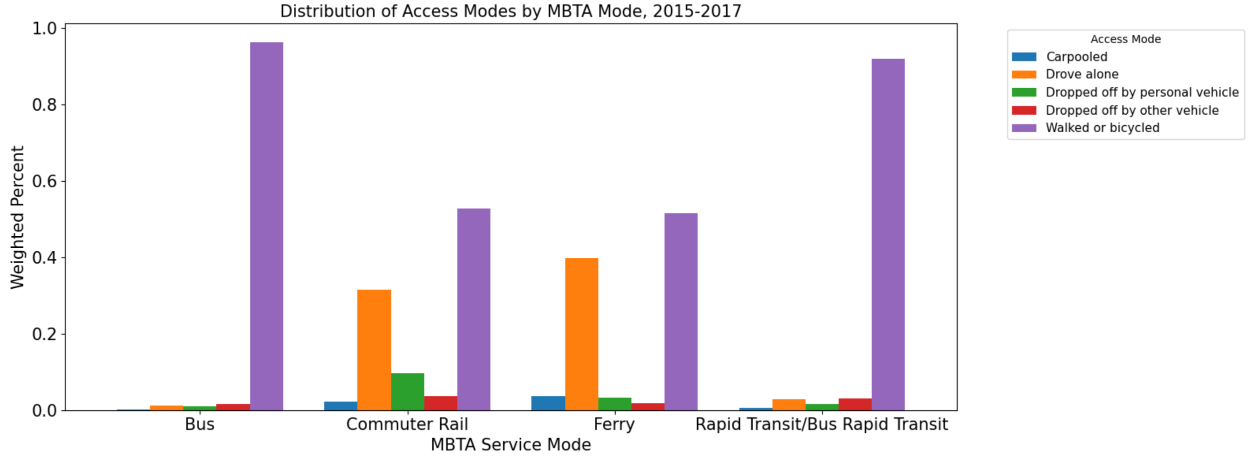
According to Mass.gov, a neighborhood is defined as an Environmental Justice population if one or more of the following are true: the annual median household income is not more than 65 percent of the statewide annual median household income; minorities comprise 40 percent or more of the population; 25 percent or more of the household identify as speaking English less than "very well"; minorities make up 25 percent or more of the population and the annual median household income of the neighborhood does not exceed 150 percent of the statewide annual median household income.

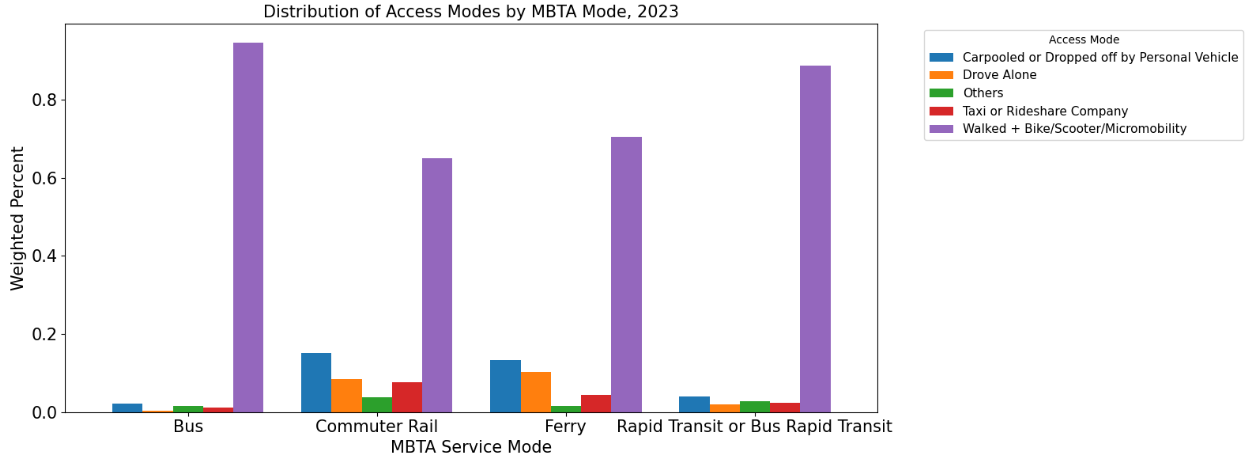
When examining this analysis in the context of Massachusetts' definition of Environmental Justice (EJ), it becomes clear that areas with the highest concentrations of EJ residents—such as Roxbury, Dorchester, Chinatown, and Lynn—are likely experiencing significant latent demand for transit services. (Appendix B)

Delays across the bus network, both citywide and on specific routes, have shown clear disparities. Routes serving Environmental Justice communities, which often have higher concentrations of low-income and minority populations, appear to be most impacted by service delays. This inequity is concerning, as it suggests that the residents who rely most heavily on public transportation for access to jobs, education, and healthcare are also the most underserved in terms of timely service.

Can we chart changes over TIME?

We analyzed the passenger census data on both the 2015-17 survey and the 2023 survey and they revealed several interesting changes. The most significant change in ridership patterns is the distribution of access modes. In particular, we observed a significant drop in the ratio of people driving to commuter rail stations and ferry stations.

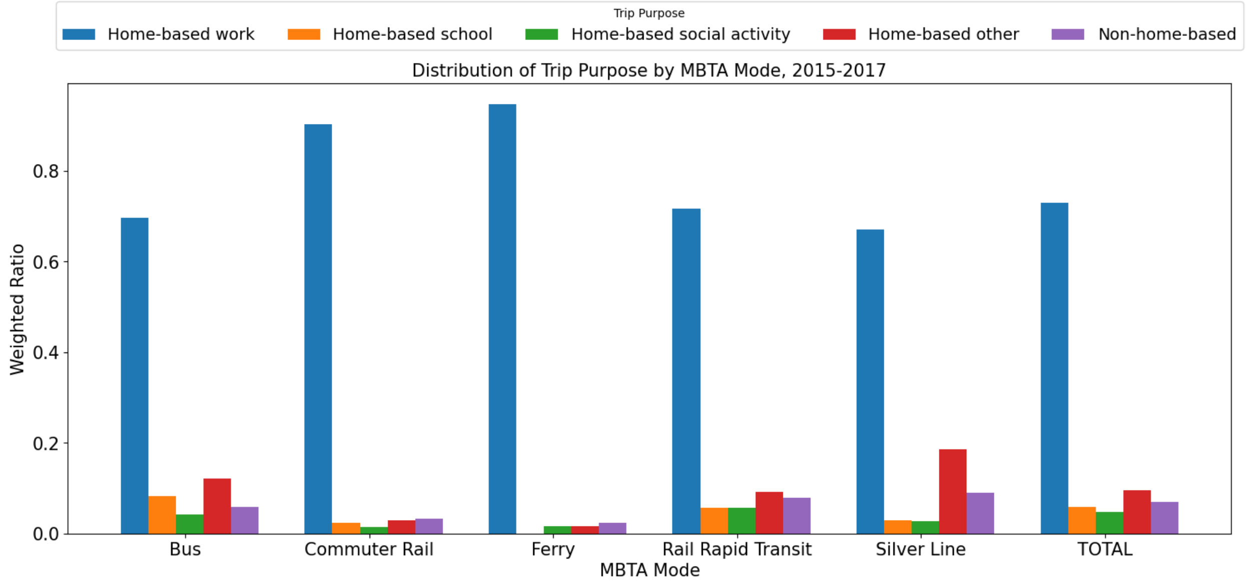


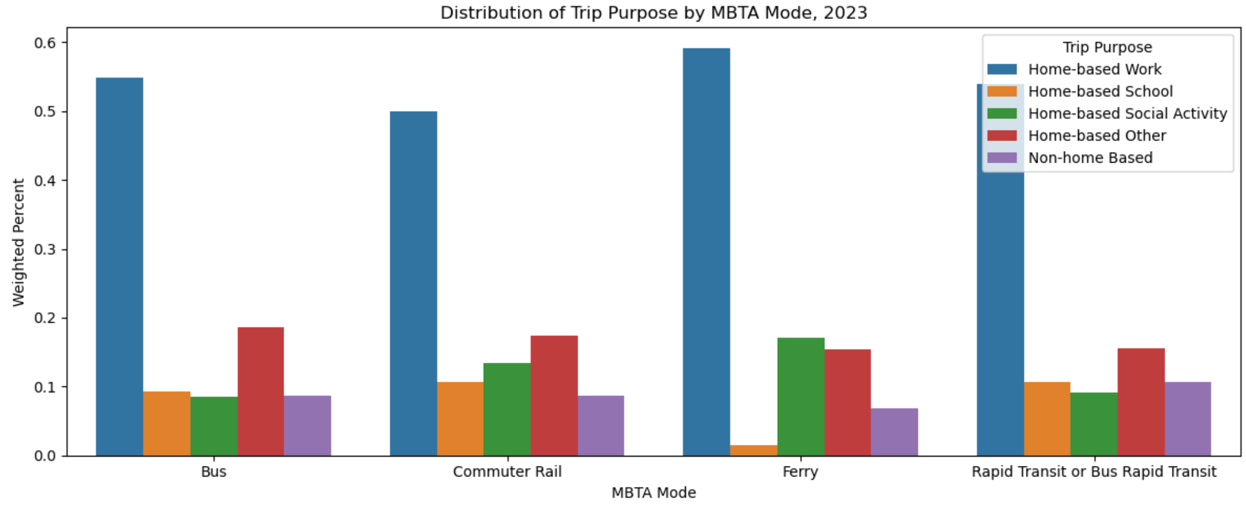


The census data reveals that over 70% of the MBTA trip purposes are work-related (more details will be mentioned later). Therefore, such a significant drop can be attributed to pandemic-driven transformations in the workforce and mobility behavior.

First, the pandemic has significantly accelerated the adoption of remote work policies. Many jobs have shifted to fully or partially remote setups, including those previously requiring access to downtown offices, undoubtedly. Thus, people are willing to switch to remote work to avoid inconvenient and time-consuming daily commutes, resulting in the drop of people driving to these stations. Second, the economic impact of the pandemic may have caused job losses or compelled workers to relocate closer to home. For those who previously relied on commuter rails or ferries, these circumstances eliminate the necessity of traveling to distant workplaces. Commuter rails are critical for workers residing far from the central Boston area, while ferries are indispensable for individuals needing to cross water bodies for work. Although much less people rely on them in the post-pandemic era, demands for them do not disappear. Therefore, such a change in distribution underscores the need to improve the MBTA network to efficiently adapt people's travel needs.

The following shift of trip purposes between pre-pandemic and post-pandemic era further confirms the shift in work patterns and transit dependency, as mentioned above. We can see that during the years 2015-2017, over 70% of people take the MBTA services for work, with over 80% for commuter rails and ferries; but in 2023, this ratio drops to below 60% for all MBTA modes. These changes may indicate a potential long-term shift in how people utilize public transportation, thereby highlighting the need for a more adaptive and flexible MBTA network.





Conclusion

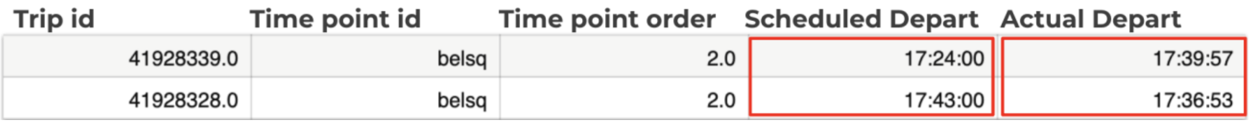
This project has provided a comprehensive analysis of the Massachusetts Bay Transportation Authority bus system performance between 2019 and 2022, focusing on ridership, travel times, delays, and service disparities across various bus routes in Boston. Through this analysis, we have identified key trends and patterns in bus performance and service equity which we can then inform future improvements to the MBTA system.

To address these disparities, it is crucial for policymakers and the MBTA to prioritize improvements in the routes serving high-demand areas and Environmental Justice communities. Investments in operational improvements, such as better scheduling, more frequent service, and targeted resources for delayed routes, can help reduce service gaps and ensure that all Boston residents have access to reliable public transportation.

Overall, this analysis underscores the importance of data-driven decision-making in transportation planning. By continuing to monitor bus performance and making adjustments based on community needs and equity considerations, the MBTA can enhance its service offerings, reduce disparities, and contribute to a more equitable and sustainable public transit system in the Greater Boston area.

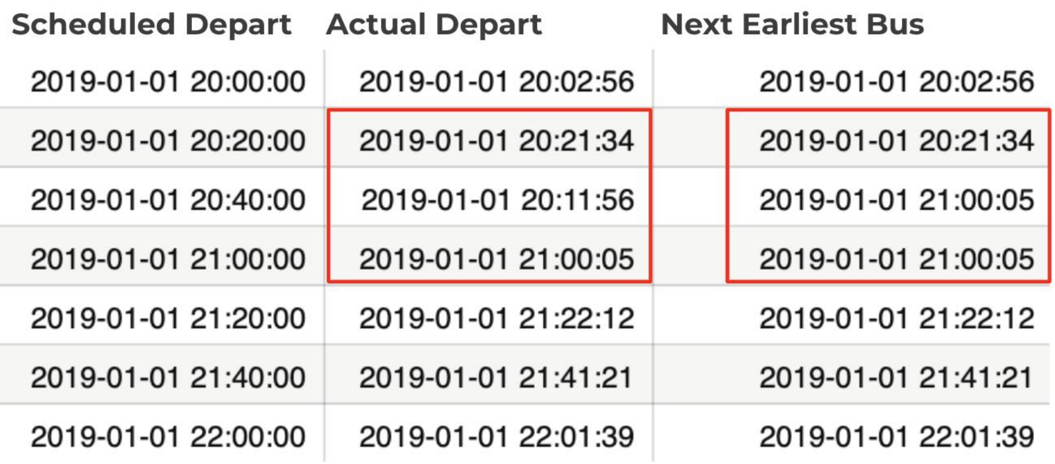
In conclusion, the insights from this report can guide future efforts to improve the MBTA bus network, ensuring that it meets the needs of all residents and helps advance social equity, environmental sustainability, and economic opportunity in Boston and beyond.

Appendix A



Trip id	Time point id	Time point order	Scheduled Depart	Actual Depart
41928339.0	belsq	2.0	17:24:00	17:39:57
41928328.0	belsq	2.0	17:43:00	17:36:53

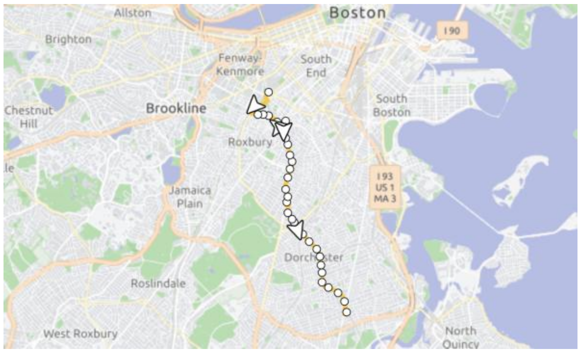
Figure 1



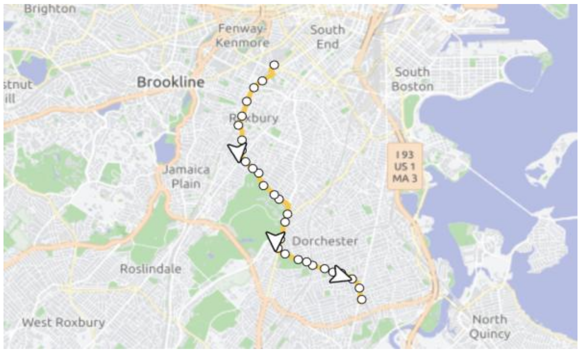
Scheduled Depart	Actual Depart	Next Earliest Bus
2019-01-01 20:00:00	2019-01-01 20:02:56	2019-01-01 20:02:56
2019-01-01 20:20:00	2019-01-01 20:21:34	2019-01-01 20:21:34
2019-01-01 20:40:00	2019-01-01 20:11:56	2019-01-01 21:00:05
2019-01-01 21:00:00	2019-01-01 21:00:05	2019-01-01 21:00:05
2019-01-01 21:20:00	2019-01-01 21:22:12	2019-01-01 21:22:12
2019-01-01 21:40:00	2019-01-01 21:41:21	2019-01-01 21:41:21
2019-01-01 22:00:00	2019-01-01 22:01:39	2019-01-01 22:01:39

Figure 2

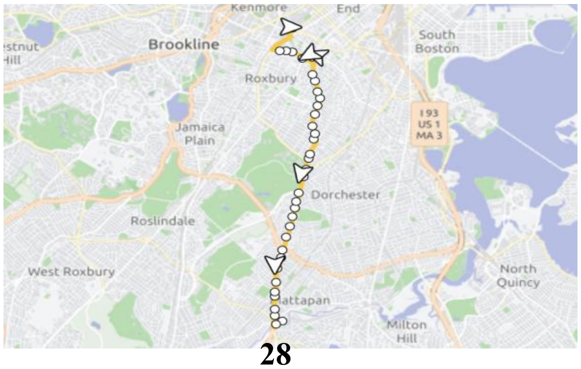
Appendix B



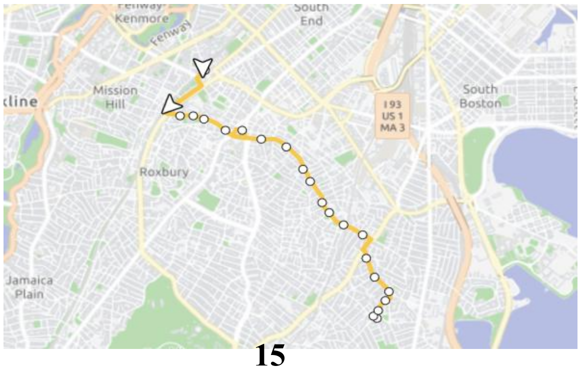
23



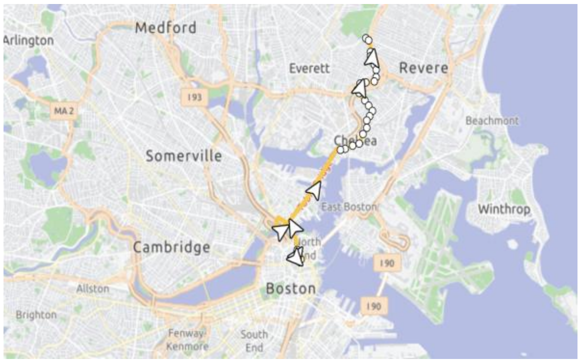
22



28



15



111

Appendix C

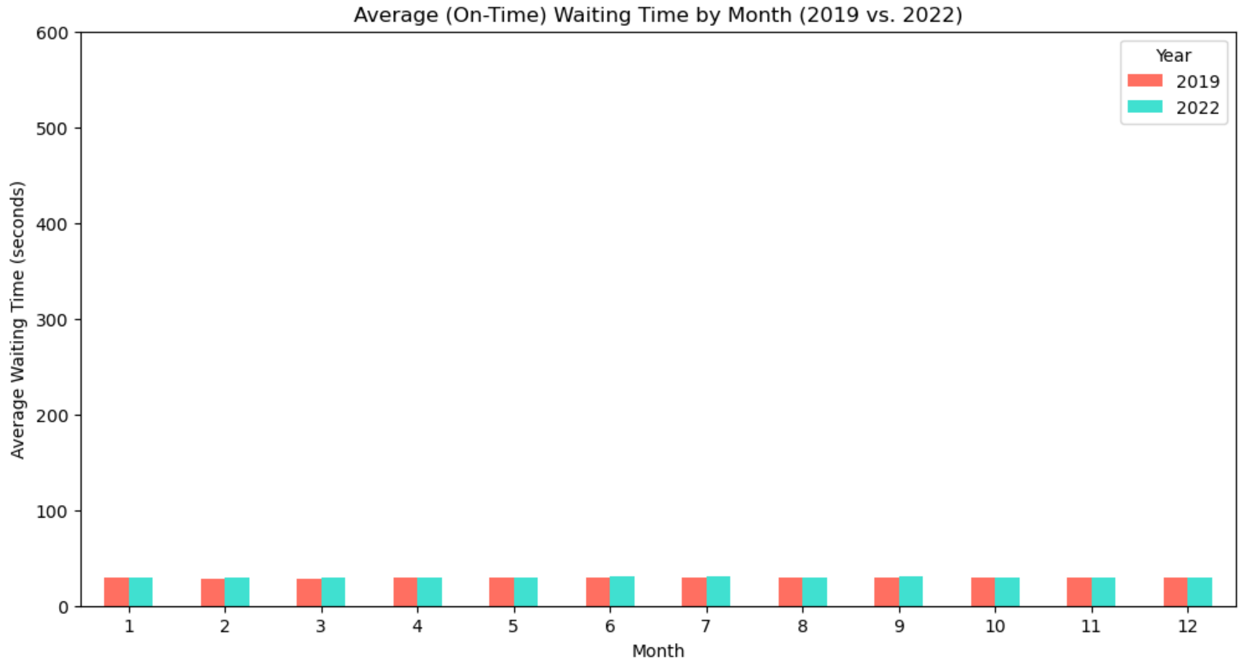


Figure A

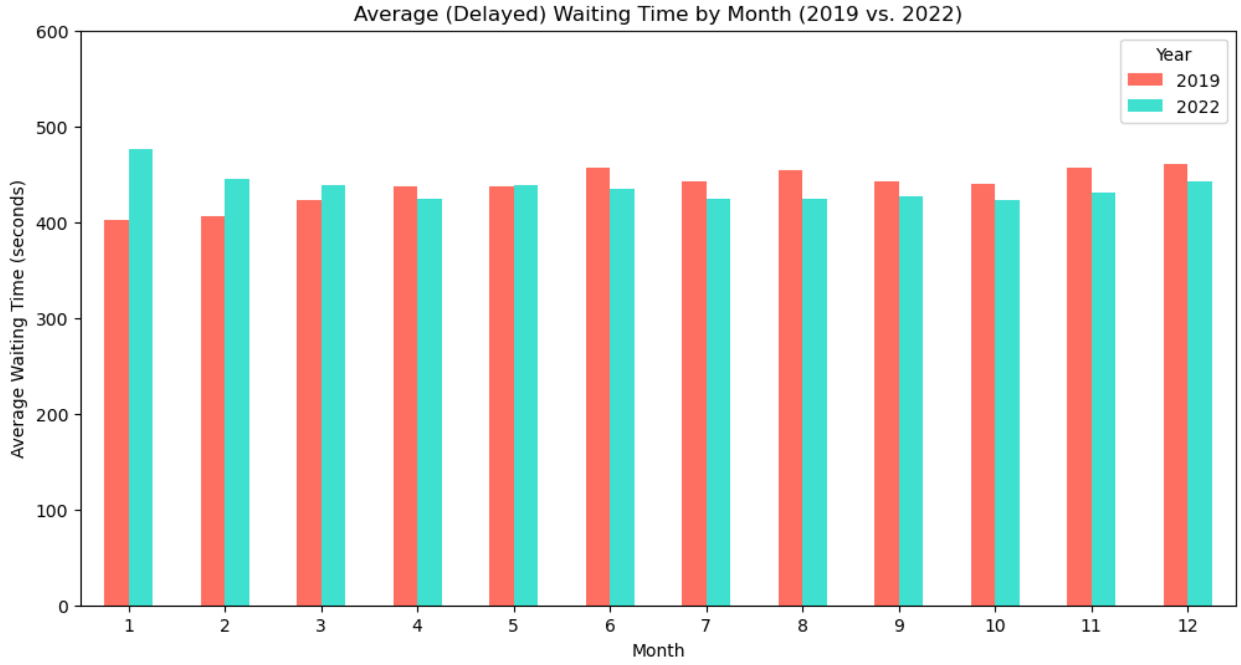


Figure B

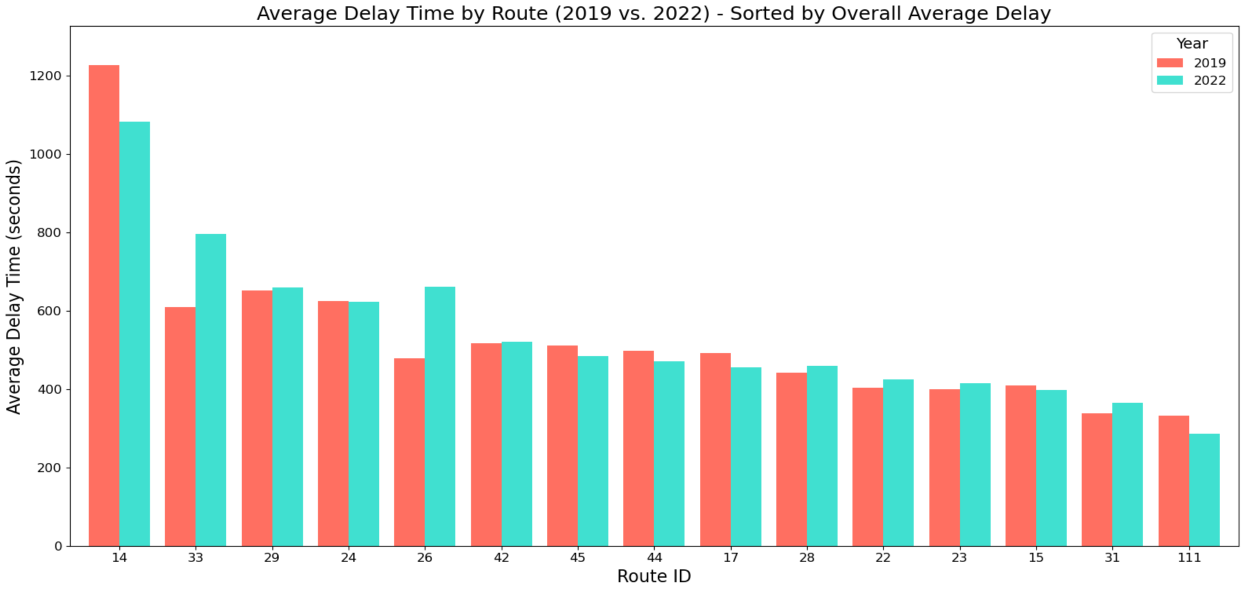


Figure C

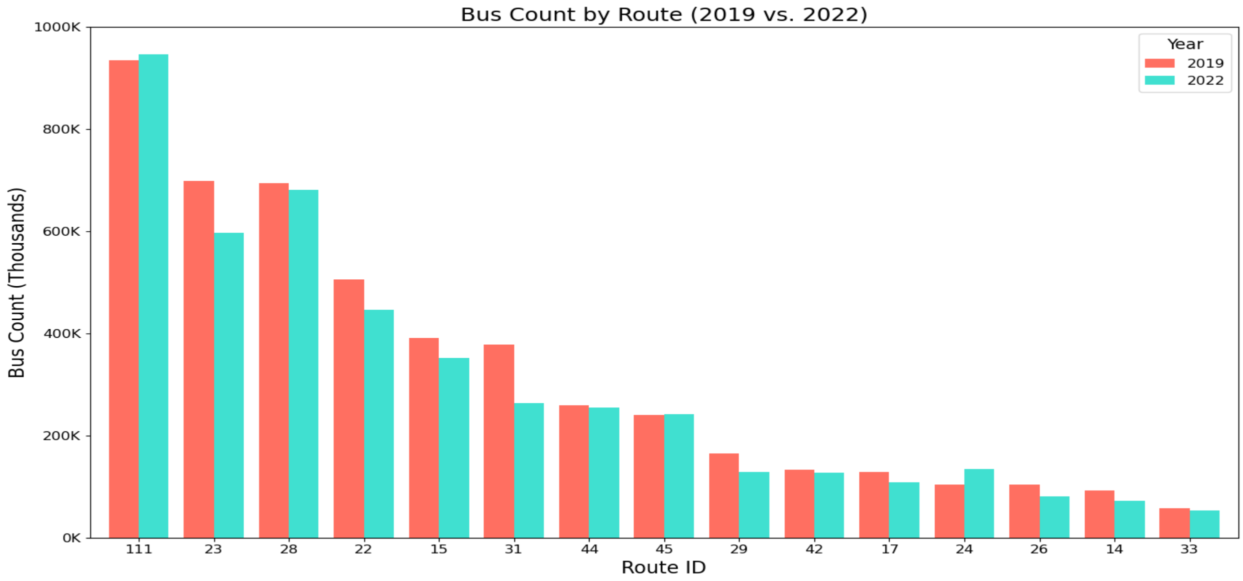
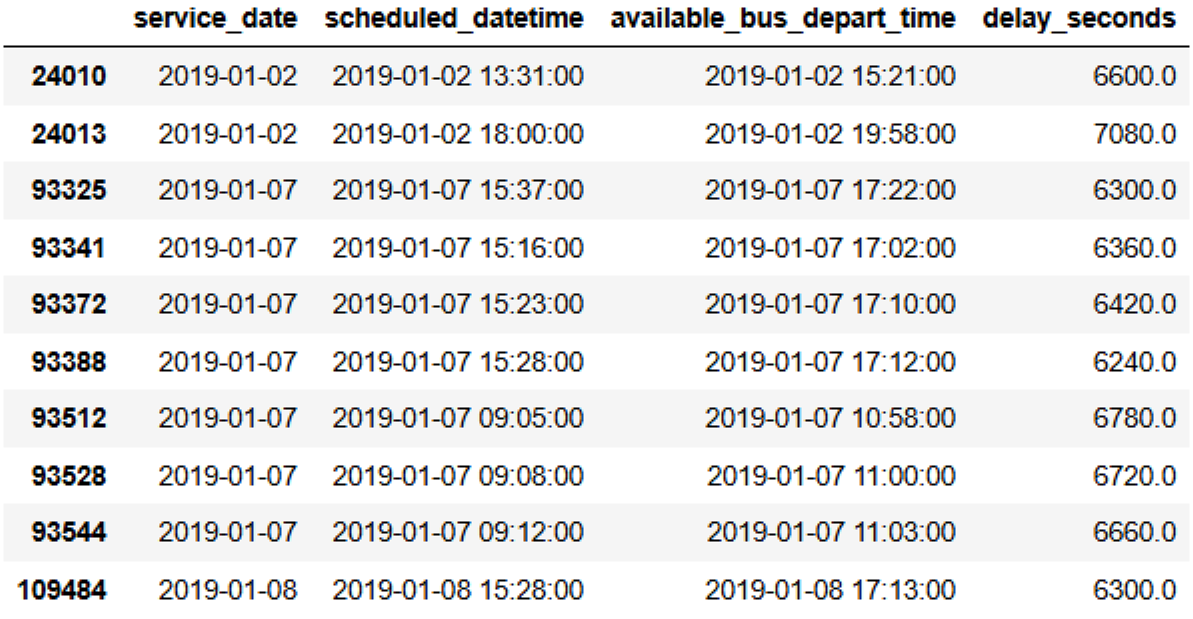
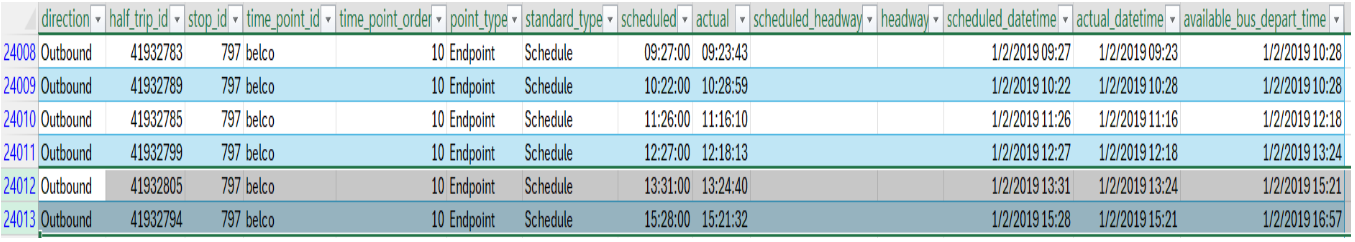


Figure D



	service_date	scheduled_datetime	available_bus_depart_time	delay_seconds
24010	2019-01-02	2019-01-02 13:31:00	2019-01-02 15:21:00	6600.0
24013	2019-01-02	2019-01-02 18:00:00	2019-01-02 19:58:00	7080.0
93325	2019-01-07	2019-01-07 15:37:00	2019-01-07 17:22:00	6300.0
93341	2019-01-07	2019-01-07 15:16:00	2019-01-07 17:02:00	6360.0
93372	2019-01-07	2019-01-07 15:23:00	2019-01-07 17:10:00	6420.0
93388	2019-01-07	2019-01-07 15:28:00	2019-01-07 17:12:00	6240.0
93512	2019-01-07	2019-01-07 09:05:00	2019-01-07 10:58:00	6780.0
93528	2019-01-07	2019-01-07 09:08:00	2019-01-07 11:00:00	6720.0
93544	2019-01-07	2019-01-07 09:12:00	2019-01-07 11:03:00	6660.0
109484	2019-01-08	2019-01-08 15:28:00	2019-01-08 17:13:00	6300.0

Figure E



	direction	half_trip_id	stop_id	time_point_id	time_point_order	point_type	standard_type	scheduled	actual	scheduled_headway	headway	scheduled_datetime	actual_datetime	available_bus_depart_time
24008	Outbound	41932783	797	belco	10	Endpoint	Schedule	09:27:00	09:23:43			1/2/2019 09:27	1/2/2019 09:23	1/2/2019 10:28
24009	Outbound	41932789	797	belco	10	Endpoint	Schedule	10:22:00	10:28:59			1/2/2019 10:22	1/2/2019 10:28	1/2/2019 10:28
24010	Outbound	41932785	797	belco	10	Endpoint	Schedule	11:26:00	11:16:10			1/2/2019 11:26	1/2/2019 11:16	1/2/2019 12:18
24011	Outbound	41932799	797	belco	10	Endpoint	Schedule	12:27:00	12:18:13			1/2/2019 12:27	1/2/2019 12:18	1/2/2019 13:24
24012	Outbound	41932805	797	belco	10	Endpoint	Schedule	13:31:00	13:24:40			1/2/2019 13:31	1/2/2019 13:24	1/2/2019 15:21
24013	Outbound	41932794	797	belco	10	Endpoint	Schedule	15:28:00	15:21:32			1/2/2019 15:28	1/2/2019 15:21	1/2/2019 16:57

Figure F

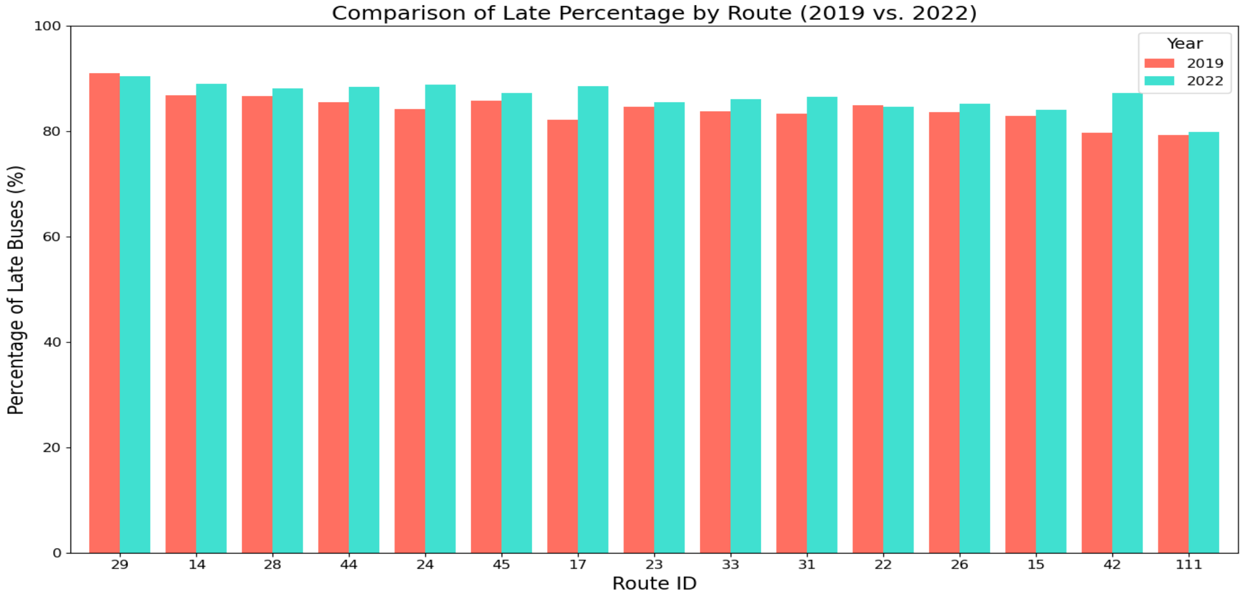


Figure G