

DETAILED METHODOLOGY—BLUEBIKES AND MBTA CONNECTIONS

This document provides more detail regarding the network connectivity and Bluebikes trips analyses in the *Bluebikes and MBTA Connections* study, which was supported through the Boston Region Metropolitan Planning Organization's Federal Fiscal Year 2025 Unified Planning Work Program.

Network Connectivity

The Boston Region MPO staff used publicly available Bluebikes station data published in June 2024 and linked it to trips data.¹ Stations that did not serve at least 100 Bluebikes trips between January 2024 and December 2024 were removed. Additional data cleaning consolidated some duplicate stations in the data. The final cleaned version of the dataset had 485 Bluebikes stations considered for analysis.

MBTA station entry points, stops, and the route network were derived from a canonical 2023 General Transit Feed Specification (GTFS) file.² For stations where entry points were not available in the dataset, the representative station point was used. Stops that served only ferry or shuttle routes were excluded from analysis. Additionally, staff only used entry points and stops within the Boston Region MPO area (containing the entire Bluebikes service area) for analysis and reporting. The MPO walk network was retrieved from OpenStreetMap using the OSMnx Python library.³

A Bluebikes station and an MBTA station or stop were considered reachable if they could be accessed within a maximum walking distance of approximately 250 meters. We calculated MBTA reachability for a majority of the Bluebikes stations using the Madina library in Python.⁴ We chose this method to reflect walk network characteristics that are overlooked in analyses that use simple radial buffers around stations.

However, some additional computation was necessary to account for errors and edge cases. For instances where Madina threw an error, we used R's spNetwork package to generate the network reachable within a maximum of 250 meters around the Bluebikes station.⁵ We then created an approximate walkshed by generating a buffered (20 meters) and smoothed concave hull of the network output using the concaveman and smoothr R packages.⁶ An MBTA station or stop was deemed reachable if an access

¹ Bluebikes station data were retrieved from <https://bluebikes.com/system-data>.

² MBTA GTFS data were retrieved from https://cdn.mbta.com/archive/archived_feeds.txt.

³ OSMnx Python library: <https://osmnx.readthedocs.io/en/stable/>.

⁴ Maina Python library: <https://madinadocs.readthedocs.io/en/latest/>. During analysis, staff discovered that the Madina library did not route along single segments that were looped (i.e., its start and end nodes were the same). Upon investigation, staff concluded that this did not contribute to a significant level of error.

⁵ spNetwork R package: <https://cran.r-project.org/web/packages/spNetwork/index.html>.

⁶ concaveman R package: <https://cran.r-project.org/web/packages/concaveman/index.html>.
smoothr R package: <https://cran.r-project.org/web/packages/smoothr/index.html>.

point fell within the walkshed of the Bluebikes station. A final round of manual cleaning was done to further consider other pairs that were not previously captured but were considerably close to a network distance of 250 meters.

The resulting dataset contained 1,554 unique pairs of Bluebikes stations and MBTA stops or stations within a reachable distance of each other.⁷

Bluebikes Trips near Transit

Cleaning Bluebikes Trip Data

Staff retrieved publicly available 2024 trip data from the Bluebikes website. The analysis focused on trips that occurred between January 2024 and December 2024. Trip data were filtered to only consider trips that met the following criteria:

- Sufficient information was available about the trip (e.g., no missing station data).
- The trip started and ended at a station that served at least 100 trips from January 2024 to December 2024.
- The trip had a duration of at least two minutes and no more than two hours.
- The trip did not start and end at the same station.
- The trip started and/or ended near transit (i.e., at least one of the Bluebikes stations was considered reachable).
- The trip occurred during MBTA service hours (simplified as 5:00 AM to 1:00 AM).⁸

Determining Trip Types

We looked at nearly four million Bluebikes trips from 2024 and classified them by their proximity to MBTA bus and rail service according to one of the four trip types:

- **Potential first-mile:** The trip started at a Bluebikes station not near transit, ended at a reachable Bluebikes station near transit, and had a recorded duration between 5 to 20 minutes.
- **Potential last-mile:** The trip started at a Bluebikes station near transit, ended at a Bluebikes station not near transit, and had a recorded duration between 5 to 20 minutes.
- **Supplemental, efficient:** The trip started and ended at a Bluebikes station near transit and had a simulated travel time ratio of at least 1.25 (biking was significantly faster than taking a transit trip).

⁷ For analysis reporting, entry points from the same station were consolidated to create unique pairs of Bluebikes station to MBTA stop/station reachability.

⁸ For trips that only ended near transit, only trips that ended within these service hours were included. For trips that only started near transit, only trips that started within these service hours were included. For trips that started and ended near transit, only trips that started and ended within these service hours were included.

- **Supplemental, less efficient:** The trip started and ended at a Bluebikes station near transit and had a simulated travel time ratio less than 1.25 (biking was slower or similarly as fast as taking a transit trip).

More information on determining travel time ratios can be found in the following section.

Determining Supplemental Travel Time Efficiency

To determine if a supplemental trip that started and ended near transit was efficient or less efficient in travel time, staff compared the estimated transit and biking travel times for all observed origin-destination pairs. The travel-time ratio was calculated by dividing the transit travel time by the biking time. A ratio of less than 1.25 suggested that biking was slower or similarly as fast as taking transit, while a ratio of at least 1.25 suggested that biking was significantly faster than taking transit.

Estimated travel times were calculated using the r5r R package.⁹ Staff analyzed a walking speed of 4.8 kilometers per hour (3 miles per hour) and a biking speed of 14.5 kilometers per hour (9 miles per hour) based on previous MPO analyses. Additionally, staff used a land elevation dataset to slow walking and biking speeds near hills with a Minetti cost function.¹⁰ Biking was only permitted on network segments with a maximum level of traffic stress of 3 as determined by the r5r algorithm.¹¹

To calculate the travel times by transit, staff analyzed walk-to-transit trips with no more than two transfers. Staff grouped trips by a representative time, day, and season to associate a typical transit schedule with the observed Bluebikes trips. Specifically, staff used GTFS files to represent the seasons and chose days from those files to represent a typical weekday, Saturday, and Sunday. These days were chosen to reflect optimal service with limited transit disruptions. Trip times were grouped to a 5-minute interval to account for computational memory constraints and to limit computation times. The days modeled for trips during 2024 are shown in Table 1.

⁹ r5r R package: <https://cran.r-project.org/web/packages/r5r/index.html>.

¹⁰ Elevation data was sourced from the elevatr R package: <https://cran.r-project.org/web/packages/elevatr/index.html>. This data comes from United States 3DEP (formerly NED) and global GMTED2010 and SRTM terrain data, courtesy of the U.S. Geological Survey: <https://github.com/tilezen/joerd/blob/master/docs/data-sources.md>.

¹¹ The level of traffic stress is a metric that defines biking suitability of a roadway segment on a scale of 1 (most suitable) to 4 (least suitable). The r5r R package calculates the level of traffic stress based on available data from the OpenStreetMap network. Future analysis could improve the level of traffic stress accuracy by incorporating other datasets.

Table 1
Dates for Simulating Transit Trips

Season	Season Dates	GTFS File(s) for Routing	Typical Weekday	Typical Saturday	Typical Sunday
Winter 2023/2024	January 1 – April 6	https://cdn.mbtace.com/archive/20240111.zip	January 30	January 13	January 14
Spring 2024	April 7 – June 16	https://cdn.mbtace.com/archive/20240416.zip (weekday); https://cdn.mbtace.com/archive/20240412.zip (weekend)	April 16	April 13	April 14
Summer 2024	June 16 – August 24	https://cdn.mbtace.com/archive/20240709.zip (weekday); https://cdn.mbtace.com/archive/20240709.zip (weekend)	July 9	July 6	July 7
Fall 2024	August 25 – December 14	https://cdn.mbtace.com/archive/20240828.zip (weekday), https://cdn.mbtace.com/archive/20241101.zip (weekend)	September 3	November 2	November 3
Winter 2024/2025	December 15 – December 31	https://cdn.mbtace.com/archive/20241227.zip	December 30	December 28	December 29

It is important to note that the results derived from this methodology, particularly as they relate to MBTA service, are estimations. Bluebikes trip data and MBTA service data alone do not detail exactly how Bluebikes users may be connecting to or substituting a bicycle ride for transit. Future data collection, integration of more data sources, and enhanced methodologies may provide further insight into how MBTA service characteristics relate to Bluebikes trip patterns and rider behavior.