

MTA Subway Origin-Destination Ridership Estimate: 2023

Overview

General Description

The Metropolitan Transportation Authority (MTA) is a public-benefit corporation responsible for public transportation in the state of New York serving 12 counties in southeastern New York, along with two counties in southwestern Connecticut under contract to the Connecticut Department of Transportation (CDOT). The MTA is the largest transportation network in North America.

Subway service within New York City is operated by MTA New York City Transit (NYCT).

This dataset provides an estimate of subway travel patterns based on scaled-up OMNY and MetroCard return tap and swipe data for 2023. It will provide estimated passenger volumes for all populated origin-destination (OD) pairs aggregated by month, day of the week, and hour of day. It also provides the name, ID, and approximate latitude and longitude of the origin and destination subway complexes.

This dataset only looks at subway taps and swipes, but there may be some trips with a large number of bus trips in between subway taps or swipes. As such, the data cannot provide any information on whether the destination subway stop was the final destination of the journey or whether there was a transfer to another mode of transportation. It also does not provide any information on what route a rider took to get from a given origin to a given destination.

This dataset was published during the first phase of the MTA's commitment to increasing transparency. This is one of a [series of annual datasets](#) with origin-destination ridership estimates. We continually examine all our published and publishable data with a view to both providing datasets that can be effectively utilized by our customers and the public at large, and to providing regular, automated updates to these datasets efficiently and sustainably. Consequently, this dataset may be restructured and/or combined with other similar datasets in the future.

Data Collection Methodology

The process for pairing taps and swipes and for removing flagged trips is the same as the one used in the calculation of [subway customer journey-focused metrics](#).

Loading data

First, all subway trips in a time period are loaded. The process then removes trip records with null values for either origins or destinations, which represent a very small number of taps and swipes, and are the result of taps or swipes at irregular fare control areas (FCAs), such as the MetroCard bus, MetroCard van, the customer service center at 2 Broadway, and the MetroCard window at the Westchester County Center.

Pairing taps and swipes

Since New York City Transit's Automated Fare Collection (AFC) system does not require riders to tap or swipe to exit, an algorithm is used to connect trips under the assumption that a passenger's next or previous tap or swipe location generally predicts their destination.

The algorithm looks first for the next subway tap or swipe by the same card on the same day. If there is such a tap or swipe, it is assumed that it took place at the destination of the current trip. If no next tap or swipe is found on the same day, the first tap or swipe by the same card the next day is looked for, and if there is such a tap or swipe, it is assumed it took place at the destination of the current trip.

The algorithm continues this process of looking for the next tap or swipe on the same card up to a period of five days. If no next tap or swipe is found within five days, it is assumed this is the last trip of the period, and the rider was returning home, so it is assumed that the destination is the first tap or swipe of the period. Taps and swipes are linked together if the following one occurs within five days instead of just for 24 hours since it was found to provide better matches.

Build OD table

After taps and swipes are paired, an OD table is built. Irrational trips—trips for which the tap or swipe pairing appears to have incorrectly guessed the destination of the rider—are flagged. Irrational trips can occur when a rider takes a one-way trip on the subway, returns home by another mode, like a taxi, and does not travel by subway in the following five days. Since no next tap or swipe would be found for this rider, the first tap or swipe of the day would be used as the destination, leading the only tap or swipe of the day to appear as both the origin and destination of the trip.

Other examples of irrational trips include trips missing origins or destinations, trips with more than five days between the origin tap or swipe and the inferred destination tap or swipe, trips with identical origin and destination fare control areas, and trips with origin and destination FCA combinations in an excluded trip table. The excluded trip table includes stations that are considered too near to each other for a logical subway trip—for which the walking time is less than six minutes and is faster than the

expected travel time using the subway-and pairs of FCA pairs that are within the same station or station complex.

These flagged trips are screened out, with the volume of remaining trips at a given FCA scaled up using a simple scaling factor so the total volume of subway trips in that FCA is still correct. The scaling factor is the total number of taps and swipes divided by the number of rational trips in the time period for a given origin FCA, assuming that irrational trips, which are trips with un-inferable destinations, likely followed the same distribution as the rational trips from that FCA. Finally, trips are aggregated to the hour for each OD pair at the station complex level and averaged across the month by day of the week.

Statistical and Analytic Issues

Data users should be mindful that travel patterns can vary significantly if there are holidays, and that ridership is different for different holidays (since some employers allow their workers to take different days off). These days include MLK Day, Presidents Day, Columbus Day, Veterans Day, Thanksgiving, and Jewish holidays like Yom Kippur. The typically lower ridership on holidays can pull down the monthly ridership average for the particular day of the week with the holiday.

Data users should also keep in mind that these ridership numbers represent our best estimate for each origin-destination pair. While the overall volume for each origin station complex will tie to that station complex's true ridership due to the scaling factor outlined above (within a margin of error due to rounding), the volume for each OD pair should not be treated as an exact value. The scaling process, along with the monthly aggregation, will often result in fractional ridership estimates.

In addition, the matching process is not perfect. For example, someone heading to work in Lower Manhattan from Kew Gardens–Union Turnpike might get off in the morning at one station (Bowling Green) but get on a train at another station (Whitehall Street) to head home. The matching process would record the OD pair for the first of these trips as Kew Gardens–Union Turnpike to Whitehall Street.

Because the matching process does not include bus trips, this data set will also contain examples like the following: a rider takes the 7 from Junction Boulevard to Flushing–Main Street and then transfers to a bus to continue their journey to a final destination in Flushing. The return journey is the reverse: starting with a bus trip to Flushing–Main Street and continuing by subway to Junction Boulevard. The matching process would record the OD pair for the first of these trips as Junction Boulevard to Flushing–Main Street without providing any indication of the second half of the journey by bus.

Limitations of Data Use

There are no limitations on the data at this time.

Release Notes

Version 1.0.0 initial release (07/23/2024)