MTA Subway Trains Delayed
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 reflects the number of subway trains delayed each month, broken down by subway line, division of the line (A Division or B Division), day type (weekday or weekend) based on scheduled service, broken down into delay reporting category.

A train is defined as delayed if it arrives at its destination terminal more than five minutes late, if it did not make any scheduled station stops, or if it was scheduled to run but did not operate.

Holidays are assigned to the Weekday/Weekend day type based on what schedule ran.

The categories of delay available in the dataset, and a definition of what they mean, are as follows:

- Crew Availability: Delays caused by lack of crew (train operator, conductor, and/or tower operator).
- Planned ROW Work: Delays due to planned right-of-way work for maintenance and capital projects.
- Infrastructure & Equipment: Delays due to infrastructure and equipment failures on the system. These failures may include issues with signaling, track, car equipment, stations, and more— provided that it is found that a group within the Department of Subways was responsible for the problem. This category also includes any delays that resulted from the activities of internal personnel.
- Police & Medical: Delays resulting from actions taken by the public. This category encompasses delays that relate to customer misconduct, the police response to incidents, sick and injured passengers, unauthorized persons on the roadbed, or persons come in contact with trains.

- Operating Conditions: Delays due to congestion, crowding, and from trains skipping stops to manage other delays.

- External Factors: Delays caused by factors outside the agency's control, such as power outages, fires outside MTA property, bridge openings, external debris that fall on our tracks, and inclement weather.

This dataset was published through [MTA's Open Data Program](), which is committed to increasing transparency through high-quality open data and accompanying resources. 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
Overview

Delayed trains are defined by their association with an incident. An incident is a singular event or operating condition that took place in the subway system. Incidents can exist without associated delays or train intervals. Each incident is described by an incident letter, which describes the event and the responses to it.

Data on train arrivals at terminals, train abandonments, train cancellations, and trains skipping regular stops are automatically computed from the Automatic Train Supervision system for the 123456 lines and the 42nd Street Shuttle (S 42nd). For all other lines, these data are manually entered into the Integrated Train Register Activity Console (I-TRAC) digital dispatching tool by train dispatchers.

Manual process

Dispatchers create incident letters for each incident. Each incident is assigned a primary trouble code (a set of codes used to describe an incident), up to four secondary trouble codes, and a Charge Department, which identifies the primary party that was responsible for the incident. Incidents can be charged to divisions in the Department of Subways, other units of the MTA, such as Construction & Development, to external parties like Police, or for conditions, like Weather.

Following the creation of an incident letter, dispatchers assign delayed trips to the incident letter of the most significant event that affected the trip. The assignment of delays to incidents, and incidents to

trouble codes, depends upon the subjective judgement of the dispatcher, and is managed through comments in I-TRAC, and the dispatcher's recollection of events. Comments are entered in I-TRAC for stations where a train dwells for over a minute, a previously on-time train arrives over five minutes late, and when a late train lost additional time since its last comment. Delayed trains are only tagged to one incident letter.

Dispatchers are allowed to review incident letters up to 45 days after an incident occurs. While the number of delays typically does not change after the month is over, and while this dataset is reported with a one-month lag, data are subject to change with a two-month lag. An automatic process later maps incidents to the reporting category based on the trouble code and/or the charge department.

Statistical and Analytic Issues

Limitations

There are some key limitations to the usage of delay and incident data given how they are collected. When analyzing this dataset, it is critically important that analysts keep these limitations in mind:

- This dataset describes general categories of events that cause delays.
- A variety of different types of events are captured under each of the delay categories. For example, car equipment delays include delays due to trains with mechanical issues, and trip cancellations due to maintenance-related car shortages.
- A train can be delayed by multiple events at different locations but will only be assigned to the incident letter of the most impactful event. As a result, other events that may have delayed a train are not captured in the data.
    - For example, in many cases, underlying operating conditions (e.g. dispatching practices and schedule design) will contribute to delays during a service disruption. However, our delay data do not provide any mechanism to capture that compounding effect.
    - More broadly, operational conditions (e.g. dispatching practices and schedule design) are often not explicitly captured in delay data since they do not manifest as events.
- Changes in how delays are attributed to delay categories can evolve over time, meaning that the increase or decrease in any one delay category might not fully explain changes in performance.
    - For example, a NYCT internal initiative in mid-2023 more rigorously linked Operating Conditions delays (delays that cannot be tied to a specific event) to their root-cause

incidents. That successful effort reduced the number of Operating Conditions delays but made it more likely that groups of delays would all be tied to a single incident.

There are a few other minor data limitations that address less-frequently occurring issues.

- Train arrival times are recorded manually on the B Division, which means that they are subject to human recording errors. Department of Subways reviews all delay data to check for errors (e.g. very early trains or backwards movements) and corrects these data points to the best of its ability. However, some infrequent and minor irregularities may still occur.
- The attribution of trains to incident letters is a manual process that is typically accurate but is subject to dispatcher judgement in the assignment of delayed trains to incidents.

Data completeness

The Essential Service Plan implemented at the outbreak of the COVID-19 pandemic meant that the electronic systems loaded with the schedules did not properly reflect the service or schedules that were operated after March 22. Therefore, data for March 2020 is based on actual delays for March 1 to March 22, and data for April 2020 is based on actual delays from April 14 to April 30. Data in April was provided after the Essential Service schedules were phased in, from April 14.

To maintain comparisons with historical data, for March 2020, the totals for the 15 weekdays and 7 weekend days with data were factored up using a daily average to the projected total for the 22 weekdays and 9 weekend days in March. For April 2020, the totals for the 13 weekdays and 4 weekend days with data were factored up using a daily average to the projected totals for the 22 weekdays and 8 weekend days in April. The detail of delays by type and major incidents categories was not provided for March or April 2020. Additionally, there was no projection for weekend incidents for March 2020, though they remained at zero through March 22.

The surge in COVID cases due to the Omicron variant required service and schedule changes on multiple lines between December 29, 2021 and January 24, 2022, and performance was measured against the schedule operated each day.

Data categorization

Data for the J and Z lines are combined and show up under JZ, and data for the N and W lines are combined and show up under N.

Limitations of Data Use

There are no limitations on the data at this time.

Release Notes
Version 1.0.0 release notes instituted with new pipeline and systemwide rows removed (11/13/2024)

Version 1.0.1 documentation edit for consistency across all years (11/14/2024)

Version 2.0.0 temporary removal of the subcategory field, documentation edits to note methodological

issues and clarify combined lines (09/29/2025)