# Metropolitan Transportation Authority (MTA) Performance Metrics

By The Integrators (Rachael Gottbetter, William Kinum, & Kyle Tenczar)

## Introduction to the MTA

- **The Metropolitan Transportation Authority** (MTA) is a public-benefit corporation responsible for public transportation in New York State.
- It serves 12 counties in southeastern New York and 2 counties in southwestern Connecticut under a contract with the Connecticut Department of Transportation (CDOT).
- Focuses on **customer experience** over traditional on-time performance metrics.

## Customer-Focused Journey Metrics

- Traditional **terminal on-time performance (OTP)** does not fully reflect rider experience.
- New **journey-based** metrics assess:
  - **Additional Platform Time (APT):** Extra waiting time beyond the schedule.
  - **Additional Train Time (ATT):** Extra onboard time due to delays.
  - **Additional Journey Time (AJT):** Sum of APT and ATT.
  - **Journey Time Performance:** % of trips where AJT < 5 minutes.
- Metrics account for **peak vs. off-peak** service.

## Data Collection & Challenges

- MTA relies on **MetroCard swipes and OMNY taps** for ridership data.
- NYC Transit does not track **exit swipes**, requiring estimation models.
- Data sources include:
  - Subway schedules & planned work adjustments.
  - Real-time train arrival and departure data.
  - Algorithms to estimate missing trip destinations.

## Methodology – Constructing the OD Matrix

- **Origin-Destination (OD) Matrix** estimates rider flows across stations.
- A **representative weekday** (usually a median Tuesday-Thursday) is selected for analysis.
- Steps to estimate missing trip data:
  1. Identify next available tap/swipe within five days.
  2. Discard irrational trips (e.g., same-station entries and exits).
  3. Adjust for **fare evasion rates** (historical rates range from **3.9% (2018) to 13.8% (2024)**).

## Network Model & Trip Assignments

- Trips are converted into **unlinked trips** (each segment taken by a rider).
- A **time-space subway model** is built, incorporating:
  - **Nodes:** Stations, transfer points, train positions.
  - **Arcs:** Train movements, walking, and waiting times.
- Rider behavior is factored in using two categories:
  - **Comfort-First Riders:** Prioritize less crowded trains.
  - **Speed-First Riders:** Take the fastest route possible.

## Path Optimization with Dijkstra’s Algorithm

- **Dijkstra’s Algorithm** is a graph-based method for finding the shortest path between points in a network.
- Applied in two steps for subway trip optimization:
  1. Reverse calculation from **destination nodes** (fewer to process).
  2. Adjustments for **train capacity, crowding penalties**, and transfer difficulties.
- The algorithm assigns **weights** to different paths based on travel time, congestion, and rider behavior.
- Multiple iterations refine estimates, improving accuracy and aligning with real-world conditions.

## Key Insights from the Data

- Customer journey data **highlights bottlenecks** and **delays** more effectively than OTP.
- Example findings:
    - **Peak-hour congestion** significantly increases **APT and ATT**.
    - **Transfer penalties** (e.g., between station complexes) impact total journey time.
    - **Real-world conditions** like service disruptions can be modeled with greater precision.

## Impact on Subway Performance & Planning

- **Benefits of customer-focused metrics:**
  - More accurate representation of delays.
  - Identifies inefficiencies in **transfer points & overcrowded lines**.
  - Supports data-driven **service adjustments & schedule changes**.
- Helps policymakers **optimize infrastructure investments**

# Data Sources

- 5 Primary Data Sources
    - Schedule Data
    - Automatic Train Supervision (ATS)
    - Communications-Based Train Control (CBTC)
    - Programmable Logic Controllers (PLCs)
    - Beacon Train Tracking System

# Data Sources

- Schedule Data
    - Planned train movement for each day
    - Scheduled service changes
- Automatic Train Supervision (ATS)
    - Circuit-based
    - Logs the time a train passes over each track circuit
- Communications-Based Train Control (CBTC)
    - Similar to ATS data
    - Exclusively for the L Train

# Data Sources

- Programmable Logic Controllers (PLCs)
    - Binary data
        - Is train on track switch or not?
    - Relay room computers record circuit activation data
    - Primary data for all of B Division (except L Train)
- Beacon Tracking
    - Bluetooth beacons on the train connected to platform-based sensors
        - Arrival & departure times
    - Actively compared to current schedule information

# Data Errors

- Trains can sometimes be matched to wrong train ID
  - Due to service rerouting & circuit issues
- Bluetooth-based data is not updated if a train is stuck between stations
  - Also not precise enough for accurate arrival and departure times
- Some arrival/departures are input manually
  - Operators can have different reporting styles
  - Needs more checks for errors
- CBTC has small chance for recording two different arrival times for same train
  - Data scanned for possible errors (trains moving above speed limt, or moving backwards

# Data Calibration

- Goal: Estimate “true” arrival & departure times
  - “Wheel stop time” vs “wheel start time”
- ○ For circuit-level data, difference between true and recorded times is a result of station and train length
  - Train could start moving and stop while still in station, circuity still triggered
- Circuit-level data calibrated manually by watching trains and comparing to automatic data
- Some missing data filled in with average dwell times for each stop, line, and track

# Methodology - Daily Assignment Model

Assigns each simulated passenger trip to the actual train they would have taken, adjusting for real-world delays, reroutes, and congestion while ensuring the simulation reflects realistic travel choices.

# Methodology - Daily Assignment Model

The model assigns each unlinked trip (individual subway ride) to both the scheduled train a passenger is expected to take and the actual train they likely boarded based on real-world conditions.

The model assumes riders understand the train schedule, adjusts for delays, overcrowding of stations, and reroutes.

# Methodology - Daily Assignment Model (Further notes)

- **Alternative routes** (transfers, backtracking) are only assigned if they save at least **10 minutes**.
- Riders stay within the **same corridor** (e.g., 1/2/3 trains instead of switching to A/C/E), cutting **computational time from 2 days to 45 minutes**.
- **Crowding is tracked**—if a train is full, passengers wait for the next one.

The goal is the ensure a **realistic simulation** of passenger decisions under changing subway conditions.

# Methodology - Generate Metrics

Measures passenger experience by comparing **actual** vs. **scheduled** travel times, ensuring data accuracy before reporting

**Key Metrics Calculated:**

- **APT (Additional Platform Time):** Extra waiting time.
- **ATT (Additional Train Time):** Extra onboard time.
- **AJT (Additional Journey Time):** Total extra travel time (APT + ATT).
- **CJTP (Customer Journey Time Performance):** % of trips completed **within 5 minutes** of the schedule.

# Methodology - Generate Metrics

Metrics are **aggregated by line, time period, and day type** using weighted averages.

**Late-night & weekend data excluded** due to inconsistent patterns.

**Validation checks** ensure accuracy—**analysts investigated anomalies** (e.g., extreme delays, missing data) and returned models if needed.

Provides a **data-driven view** of subway reliability
from the passenger perspective

# Hypothesized Predictive Power

**Passenger Wait Times (APT)**:

- Predicts the likelihood of delays based on historical service patterns.
- APT data can help forecast areas of congestion or high wait times.

**Train Assignment (ATT)**:

- Predicts which trains are likely to be assigned to passengers under specific service conditions.
- Helps optimize scheduling and reduce passenger wait times.

**Journey Times (AJT)**:

- Predicts how variations in service disruptions or schedule changes affect overall travel time.

**CJTP (Customer Journey Time Percentage)**:

- Predicts the proportion of riders likely to experience minimal delays (AJT < 5 minutes).
- Useful for assessing overall system reliability.

# Hypothesized Predictive Power Cont.

**Addressing Service Disruptions**

**Predictive Modeling of Disruptions**:

- The dataset can forecast disruptions' effects on passenger journey times, helping to mitigate the impact of delays.

**Dynamic Rerouting**:

- Predicts the most efficient reroute options for passengers in real-time, reducing inconvenience.

**Transfer Optimization**:

- Data helps predict which transfer points will be most effective in reducing travel time during service disruptions.

# Why the Dataset is Important

**Operational Efficiency**:

- The dataset helps subway systems optimize service delivery, reducing delays, improving passenger satisfaction, and maximizing capacity.

**Real-time Decision Making**:

- Provides predictive insights that can inform decisions during service disruptions or heavy passenger load times.

**Long-term Infrastructure Planning**:

- Data-driven analysis allows transit authorities to plan for future expansions or changes based on trends in passenger behavior.

# Potential Applications of the Dataset

**Improved Customer Experience**:

- Anticipates delays and reroutes to minimize disruption, leading to a smoother travel experience.

**Enhanced Scheduling**:

- Enables more accurate prediction of peak vs. off-peak load, improving train frequency and reducing overcrowding.

**Resource Allocation**:

- Predicts times of high demand or delay, optimizing staffing and train availability for smoother operations.

## Summary

- **Dataset Overview**: Tracks subway journey metrics (wait times, travel times, delays) across multiple lines.
- **Key Assumptions**: Passengers rely on schedules, but delays and rerouting impact journey times.
- **Predictive Power**: The dataset predicts wait times, travel times, and service disruptions, aiding in scheduling and re-routing decisions.
- **Importance**: Improves operational efficiency, enhances passenger experience, and informs long-term infrastructure planning.