**Using Data to Make City Transportation Better: A Smarter Way to Solve Traffic and Parking Problems**

Shreejana Ghalan Lama

CUNY School of Professional Studies

Data Acquisition and Management IS 362

Professor Larry Cohen

May 18, 2025

Table of Contents

[1. Introduction 3](#_Toc193559472)

[2. Problem Description 3](#_Toc193559473)

[3. Solution Discussion 4](#_Toc193559474)

[4. Analysis 5](#_Toc193559475)

[4.1 Transit Recovery Trends 5](#_Toc193559476)

[4.2 Weekday vs. Weekend Patterns 5](#_Toc193559477)

[4.3 Year-over-Year Recovery 6](#_Toc193559478)

[5. Conclusion 7](#_Toc193559479)

[6. Recommendations 8](#_Toc193559480)

[References 9](#_Toc193559481)

# **Introduction**

Urban transportation systems are the lifeblood of modern cities, essential for their economic vitality and social connectivity, yet they often grapple with persistent inefficiencies such as congestion, inconsistent service, and significant planning gaps. New York City, home to one of the largest and most complex transit systems globally, exemplifies these challenges, particularly in the wake of unprecedented disruptions like the COVID-19 pandemic. How can the wealth of publicly available data be harnessed to navigate these complexities and restore efficiency? This project explores this question by leveraging public open data through Python-based data analytics to analyze and improve urban transportation efficiency, using MTA daily ridership data from 2020 to 2025 as a compelling case study.

The central goal is to uncover critical trends in post-pandemic transit recovery, highlight evolving behavioral patterns such as weekday versus weekend ridership, and identify clear opportunities for data-informed decision-making. This work draws inspiration from Ben Wellington’s (2016) TED Talk, which demonstrates the power of using urban data to solve practical city problems. Furthermore, it builds upon the conceptual frameworks proposed by Kitchin (2014) and Batty (2013), who discuss the transformative potential of the data revolution and smart city paradigms in reshaping urban planning and infrastructure management through data-driven insights.

# **Problem Description**

Public transit systems are intricate ecosystems, their performance intricately linked to a multitude of variables including rider behavior, service frequency, external events like public health crises, and the existing state of infrastructure. While these systems generate vast amounts of open data, this resource is often characterized by its "messy" nature—unstructured, incomplete, or requiring significant processing—and consequently remains underutilized. NYC’s Metropolitan Transportation Authority (MTA) releases regular updates on ridership, but this raw data demands substantial cleaning and transformation before it can yield actionable insights for strategic planning.

The onset of the COVID-19 pandemic profoundly disrupted traditional commuting patterns and overall transit usage. As remote work became widespread, school schedules shifted, public health guidelines evolved, and social behaviors adapted, city agencies often lacked clear, real-time insights into these rapidly changing transit needs. This information vacuum can lead to services being either under-supplied, causing frustration and hindering mobility, or over-supplied, resulting in inefficient resource allocation. This project directly addresses the challenge of transforming publicly available, yet often unwieldy, data into clear, actionable intelligence that can support smarter, more adaptable, and resilient transportation planning.

# **Solution Discussion**

To explore post-pandemic transit trends in NYC, this project utilized the "MTA Daily Ridership Data (2020–2025)" dataset sourced from the NYC Open Data portal (NYC Open Data, 2025). The entire process of data acquisition, cleaning, transformation, analysis, and visualization was conducted within the Python programming environment, primarily leveraging the Pandas library for data manipulation, and Matplotlib and Seaborn for generating visualizations.

The analysis targeted five key transit modes: Subways, Buses, Long Island Rail Road (LIRR), Metro-North Railroad, and the Staten Island Railway. The dataset included variables such as total estimated ridership and the percentage of comparable pre-pandemic ridership for each mode. Data preparation involved several critical steps:

* Parsing and standardizing date fields into datetime objects using pd.to\_datetime for time-series analysis.
* Creating derived columns, such as the day of the week name, month, and year, from the date objects to facilitate trend identification and grouped analysis.
* **Addressing Data Integrity and Type Conversion Challenges:** A significant challenge encountered was the format of the ridership data columns. These columns, essential for numerical analysis, contained non-numeric entries or were not consistently recognized as numeric types by Pandas upon initial loading. Attempting direct numerical operations would lead to errors. To overcome this, each ridership column (e.g., 'Subways: Total Estimated Ridership', 'Buses: Total Estimated Ridership', etc.) was explicitly converted to a numeric type using pd.to\_numeric(errors='coerce'). The errors='coerce' argument was crucial as it automatically transformed any problematic, non-convertible entries into NaN (Not a Number) values instead of halting the process. Subsequently, these NaN values, along with any pre-existing missing data, were handled by imputing them with 0 using fillna(0). This decision was made based on the assumption that a missing or unreadable value likely indicated no or negligible ridership for that specific entry, ensuring data integrity for subsequent calculations and preventing errors in aggregation and visualization.
* Employing 30-day rolling averages using rolling(window=30).mean() to smooth out daily fluctuations and reveal clearer long-term trends in ridership recovery.
* **Utilizing Advanced Visualization with Seaborn:** For some visualizations, the Seaborn library was employed. Seaborn, built on top of Matplotlib, provides a high-level interface for drawing attractive and informative statistical graphics, which was particularly beneficial for comparing distributions across weekdays and weekends, a feature extending beyond basic Matplotlib functionalities typically covered.

These preparatory steps were crucial for transforming the raw dataset into a structured format suitable for robust analysis and the generation of meaningful insights. The subsequent visualizations, such as line charts for recovery trends and bar charts for comparative usage, were designed to clearly communicate the findings.

**A graph of a number of people

AI-generated content may be incorrect.**

**Line chart of 30-day rolling averages for transit recovery (2020–2025)**  
Smoothed Recovery Trends in NYC Transit Ridership by Mode

# **Analysis**

## **4.1 Transit Recovery Trends**

The processed data, visualized in Figure 1, reveals a dramatic and immediate drop in transit usage across all modes in early 2020, coinciding with the pandemic's onset. Recovery patterns, however, varied significantly by mode. Buses demonstrated a relatively quicker initial rebound throughout 2020 and 2021. In contrast, the commuter rails, LIRR and Metro-North, experienced a slower start to their recovery, with ridership gradually but steadily increasing through 2023 and 2024. By late 2024, LIRR impressively reached over 90% of its pre-pandemic ridership levels, closely followed by Metro-North. Subways maintained a more consistent, albeit slower, recovery trajectory throughout the period. The Staten Island Railway, however, consistently lagged behind all other modes in its recovery.

These divergences likely reflect shifting work patterns and lifestyle changes. Commuter railroads, heavily reliant on traditional 9-to-5 suburban commuters, were more significantly impacted by the sustained adoption of remote and hybrid work models. Subways and buses, catering to a broader range of travel needs including essential workers and diverse intra-city movements, showed more resilience, as noted by Batty (2013) in discussions on urban mobility patterns.

## **4.2 Weekday vs. Weekend Patterns**

To delve deeper into behavioral shifts, ridership data was aggregated and analyzed by day of the week. This analysis, illustrated in Figure 2, uncovered distinct patterns. Subways and buses exhibited relatively consistent usage throughout the weekdays, underscoring their continued role as essential modes for daily commuting, errands, and general intra-city travel.

Conversely, LIRR and Metro-North displayed noticeable spikes in ridership during weekends, particularly on Saturdays. This trend suggests a significant rise in recreational, leisure, or non-work-related travel on these commuter lines. This shift aligns with evolving commuter habits in a hybrid work environment, where weekends might be increasingly utilized for activities previously constrained by traditional workweeks. Concurrently, the discernible dip in weekday usage for LIRR and Metro-North, compared to pre-pandemic norms (though not directly shown, implied by recovery rates), likely indicates the persistence of remote or flexible work arrangements among their primary suburban rider base.

**A graph of different colored bars

AI-generated content may be incorrect.**

**Bar chart:** *Transit Mode Usage Across Weekdays and Weekends*

## **4.3 Year-over-Year Recovery**

An analysis of yearly average ridership percentages, relative to pre-pandemic levels (visualized in Figure 3), revealed distinct recovery phases and trajectories across the transit modes from 2020 to 2024 (with early 2025 data reflecting continuation of these trends). In 2020, all systems operated at dramatically reduced capacity, with average ridership hovering between 20%–30% of pre-pandemic figures.

Buses showed the strongest initial recovery in 2021, reaching nearly 60% of pre-pandemic levels, possibly due to their perceived safety (open-air environment, less crowding compared to subways initially) and continued service for essential workers. However, as hybrid work policies solidified and public confidence gradually returned, Subways, LIRR, and Metro-North began to gain significant momentum.

By 2023 and 2024, both LIRR and Metro-North had surpassed 85% recovery, with LIRR nearly achieving full pre-pandemic ridership levels. Subways demonstrated consistent year-over-year growth, steadily rebuilding their passenger base. Buses, after their initial strong rebound, saw their recovery plateau slightly compared to the rapid gains of commuter rails. The Staten Island Railway, however, consistently remained the outlier, struggling to surpass 60% recovery throughout the analyzed period. This persistent lag could be attributed to a combination of factors, such as its geographical isolation, reliance on ferry connections, limited direct routes to major employment hubs outside Staten Island, or perhaps service frequencies not aligning effectively with new demand patterns. These findings strongly suggest that differentiated, mode-specific strategies are necessary, reflecting their unique recovery timelines, evolving user bases, and operational contexts.

**A graph of different colored bars

AI-generated content may be incorrect.Bar chart – Yearly Average % Recovery per Mode**  
*Transit Mode Recovery Over Time (2020–2025)*

# **5. Conclusion**

This project successfully demonstrates that even with publicly accessible tools like Python and open data portals, meaningful and actionable insights can be extracted to inform and guide public transit decisions in a complex urban environment. The analysis of MTA ridership data from 2020 to 2025 clearly shows that each transit mode has experienced a unique recovery curve and exhibits distinct usage patterns post-pandemic. These findings underscore the necessity for data-informed planning to be mode-specific, highly adaptive, and reflective of actual rider behavior and evolving demand.

Subways and buses, forming the backbone of daily intra-city travel, may benefit most from continued investment in service consistency and frequency. Commuter rails (LIRR and Metro-North) should continue to adapt their services and marketing to capitalize on the observed growth in weekend and off-peak travel, alongside catering to new hybrid work schedules. The persistent underperformance of the Staten Island Railway signals a clear need for targeted investigation, potentially involving service re-evaluation or community-specific studies to understand and address its unique challenges.

Overall, this project reinforces the vision of smart cities, as articulated by Kitchin (2014) and Batty (2013), where real-world data actively drives responsive, equitable, and efficient transportation systems. The insights derived here could empower transit authorities to make more targeted investments and operational adjustments, ultimately fostering a more resilient and user-centric transit network. Furthermore, the methodology employed offers a replicable framework for other cities facing similar urban mobility challenges, or for analyzing other facets of urban life through the lens of open data.

# **6. Recommendations**

Based on the analysis, the following recommendations are proposed to enhance NYC's transportation system:

* **Enhance Weekend Commuter Rail Service:**Given the observed rise in weekend travel, particularly on LIRR and Metro-North, consider increasing service frequency, offering targeted fare promotions, or improving connections to recreational destinations to meet this growing non-work travel demand.
* **Sustain and Optimize Weekday Subway and Bus Frequencies:** Continue to support essential worker and daily commuter traffic on subways and buses by maintaining robust weekday service, while using granular data to optimize routes and schedules based on new commuting hotspots or times.
* **Conduct Targeted Investigation of Staten Island Railway Underperformance:**Initiate a focused study on the Staten Island Railway, incorporating passenger surveys, community outreach, and analysis of alternative transportation options to understand the root causes of its lagging recovery and identify potential interventions (e.g., service adjustments, improved connectivity, infrastructure upgrades).
* **Leverage Open Data Analytics for Dynamic Planning and Operations:** Continue and expand the use of open data analytics to develop and maintain real-time performance dashboards. Such dashboards could display current ridership levels against historical averages, predict demand surges, monitor service disruptions, and track key performance indicators. This would enable swift adjustments to service allocation, proactive communication with riders, and more agile long-term planning.
* **Invest in Data Literacy and Cross-Agency Collaboration:** Foster greater data literacy within transit planning departments and encourage collaboration with other city agencies to enrich transit data with other relevant urban datasets (e.g., employment statistics, housing development, event schedules) for more holistic and predictive modeling.

# References

Batty, M. (2013). Big Data, Smart Cities and City Planning. Dialogues in Human Geography, 3(3), 274-279. [https://journals.sagepub.com/doi/full/10.1177/2043820613513390](https://www.google.com/url?sa=E&q=https%3A%2F%2Fjournals.sagepub.com%2Fdoi%2Ffull%2F10.1177%2F2043820613513390)

Kitchin, R. (2014). The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. SAGE Publications. [https://se.moevm.info/lib/exe/fetch.php/courses:smart\_data:the\_data\_revolution\_big\_data\_open\_data\_data\_infrastructures\_and\_their\_consequences\_by\_rob\_kitchin\_z-lib.org\_.pdf](https://www.google.com/url?sa=E&q=https%3A%2F%2Fse.moevm.info%2Flib%2Fexe%2Ffetch.php%2Fcourses%3Asmart_data%3Athe_data_revolution_big_data_open_data_data_infrastructures_and_their_consequences_by_rob_kitchin_z-lib.org_.pdf)

NYC Open Data. (2025). MTA Daily Ridership Data (2020–2025). Retrieved from [https://data.cityofnewyork.us/](https://www.google.com/url?sa=E&q=https%3A%2F%2Fdata.cityofnewyork.us%2F)

Wellington, B. (2016, April). How We Found the Worst Place to Park in New York City – Using Big Data . TED Conferences. [https://www.ted.com/talks/ben\_wellington\_how\_we\_found\_the\_worst\_place\_to\_park\_in\_new\_york\_city\_using\_big\_data](https://www.google.com/url?sa=E&q=https%3A%2F%2Fwww.ted.com%2Ftalks%2Fben_wellington_how_we_found_the_worst_place_to_park_in_new_york_city_using_big_data)