**MTA Daily Ridership Data Project**

**Presented by:**

1. Hany Moussa
2. Beshoy Gergis
3. Amir Ibrahim
4. Ahmed Wahid

Under the supervision of **Eng. Mohamed Attia**

**Table Of Contents**

[1. Overview 3](#_Toc196407192)

[2. Objectives 3](#_Toc196407193)

[3. Data Source 3](#_Toc196407194)

[4. Data Analysis Questions 3](#_Toc196407195)

[5. Data Cleaning 4](#_Toc196407196)

[6. Calculated Measures (DAX)](#_Toc196407196) 8

[7. Data Modelling 8](#_Toc196407197)

[8. Data Visualization (Report Creation) 10](#_Toc196407198)

[9. Conclusion 17](#_Toc196407199)

[10. References 17](#_Toc196407199)

[11. Acknowledgment 20](#_Toc196407200)

# 1. Overview

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.

Our dataset contains daily ridership data for New York’s Metropolitan Transportation Authority (MTA) from March 2020 to October 2024. It includes total estimated ridership counts and the percentage compared to pre-pandemic levels for different transportation modes. With over 1,700 entries, it offers valuable insights into how COVID-19 affected transit usage and how ridership patterns have changed over time across the MTA network.

# 2. Objectives

The dataset aims to:

* Facilitate research and analysis of transportation trends before and after the COVID-19 pandemic.
* Enable exploration of changes in passenger numbers, transportation modes, and usage patterns.
* The intended audience for this Dashboard are:
  + MTA planners and city officials: to support data-driven decision-making.
  + Policy makers: to understand shifts in public behavior.
  + Researchers: investigating urban recovery.

# 3. Data Source

The dataset originates from the [**MTA Daily Ridership Dataset**](https://github.com/HaniMoussa/MTA-Project-DEPI--G2-6-.git), providing comprehensive data on daily public transportation usage. It encompasses information on ridership figures from **March 1, 2020, to October 31, 2024**.

|  |  |
| --- | --- |
| Represents the specific day for which the ridership data is recorded. | Date |
| Note: The following fields represent seven distinct modes of transportation. | |
| The daily total estimated number of this transportation ridership in New York City (NYC) is associated with the ridership data. | Total Estimated Ridership |
| Daily ridership estimates as a percentage of ridership on an equivalent day prior to the COVID-19 pandemic | Percentage of Comparable Pre-Pandemic Day |

The data encompasses various aspects, including

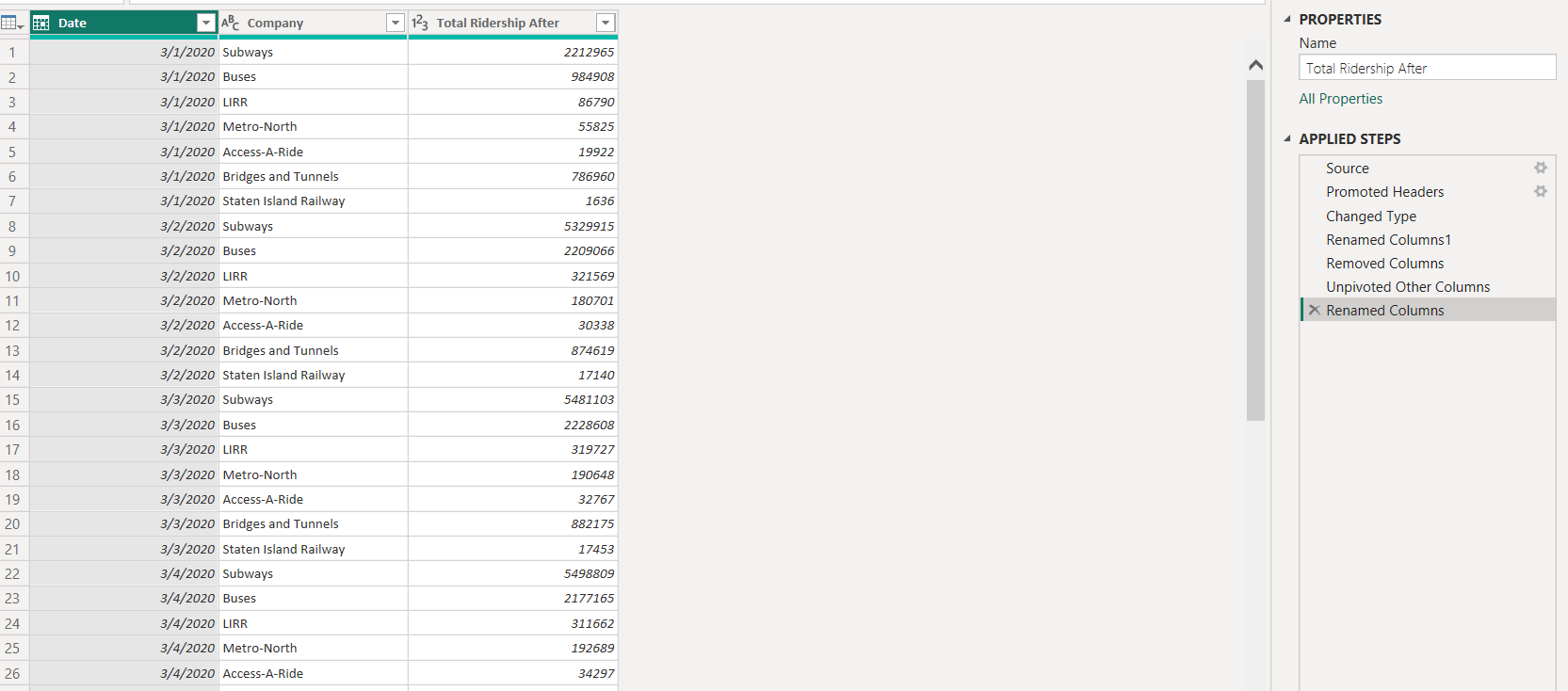
# 4. Data Analysis Questions:

* How did the actual total ridership compare to the expected total ridership? What were the actual and expected values
* Understand the recovery of total ridership after the pandemic and provide a forecast for upcoming months.
* Which months experienced the highest and lowest impacts on ridership due to the pandemic?
* Have ridership numbers returned to pre-pandemic levels?
* What is the quarter-over-quarter recovery percentage in ridership?
* What is the impact of weekends and holidays on the average ridership per each transportation mode?
* What is the impact of different seasons on the average and total ridership per each transportation mode?
* Are there any remarks or issues that need to be addressed resulting from the pandemic's impact on ridership?

# 5. Data Cleaning

5.1 Total Ridership After COVID (Actual Data)

The dataset was carefully cleaned by going through all the columns and removing any that were not directly useful for the analysis. This step helped make the data easier to understand and work with. After cleaning, pivot tables were created to summarize the data in a clear and organized way. These tables made it possible to group the information by specific categories and calculate totals, helping to highlight patterns and trends in the data.

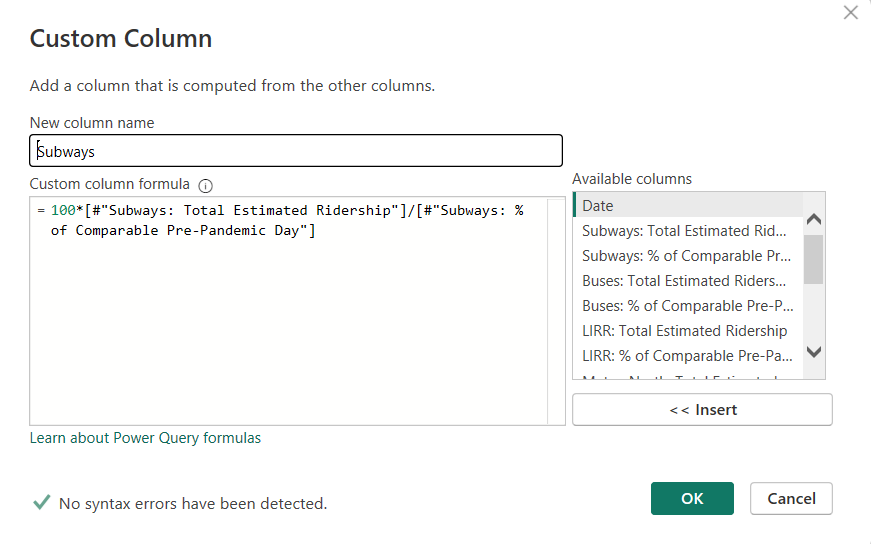
The resulting table columns

|  |  |
| --- | --- |
| Represents the specific date for which the ridership data is recorded. | Date |
| Indicates the transportation company or service provider associated with the ridership data. | Company |
| The daily Total Estimated Ridership After COVID for a Specific Date and company | Total Ridership After COVID |

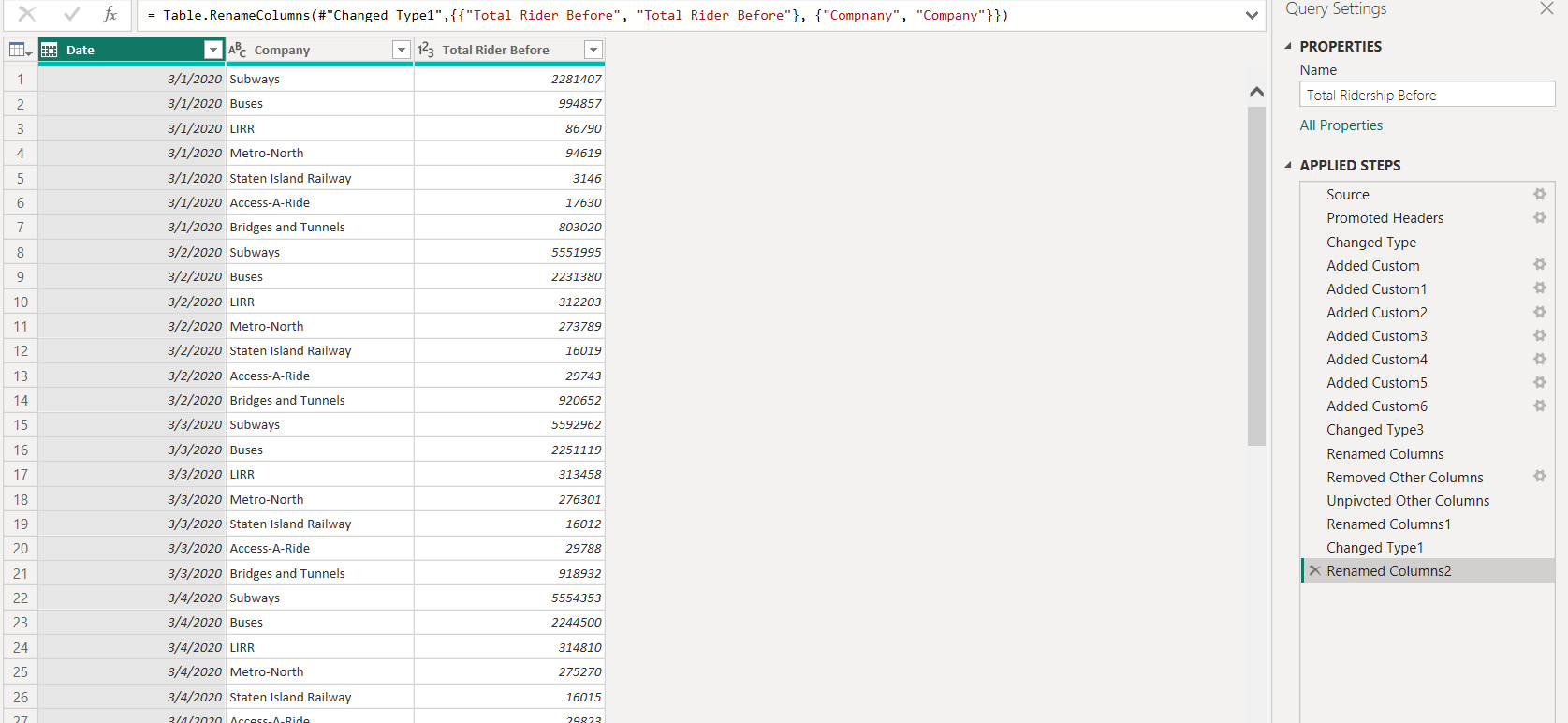
1 Table of Total Ridership After COVID

## 5.2 Expected Total Ridership

We calculated the expected total ridership by dividing the actual total ridership by the expected percentage of a comparable pre-pandemic day. This approach allowed us to estimate how many riders we would normally expect if ridership patterns had fully returned to pre-pandemic levels. In other words, it helped us adjust the current numbers to reflect what they might have been under normal, pre-pandemic conditions. Then we repeat the pivot step with the resulting columns to get the same table layout as the actual ridership table.



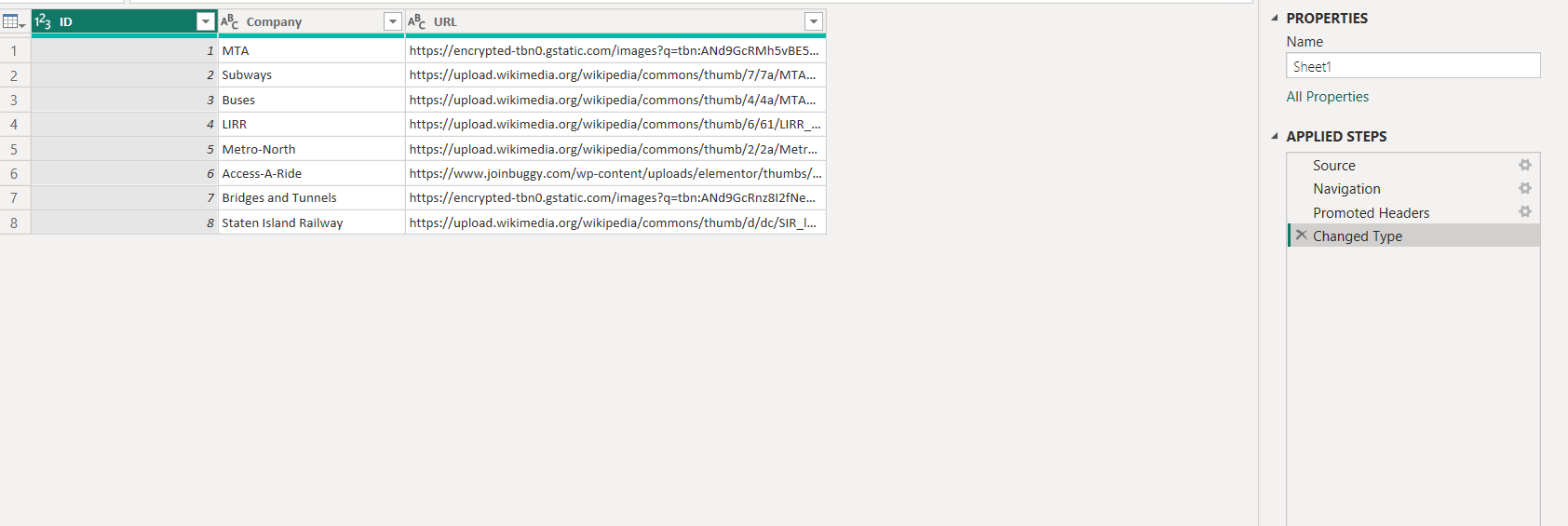
2. Custom Column created For the Subway

The resulting table columns

3. Table of Total Ridership Before COVID

|  |  |
| --- | --- |
| Represents the specific day for which the ridership data is recorded. | Date |
| Indicates the transportation company or service provider associated with the ridership data. | Company |
| The daily Total Estimated Ridership | Total Estimated Ridership |

## 5.3 Table of Logos

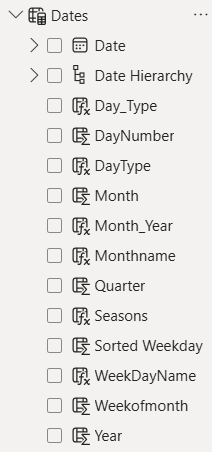
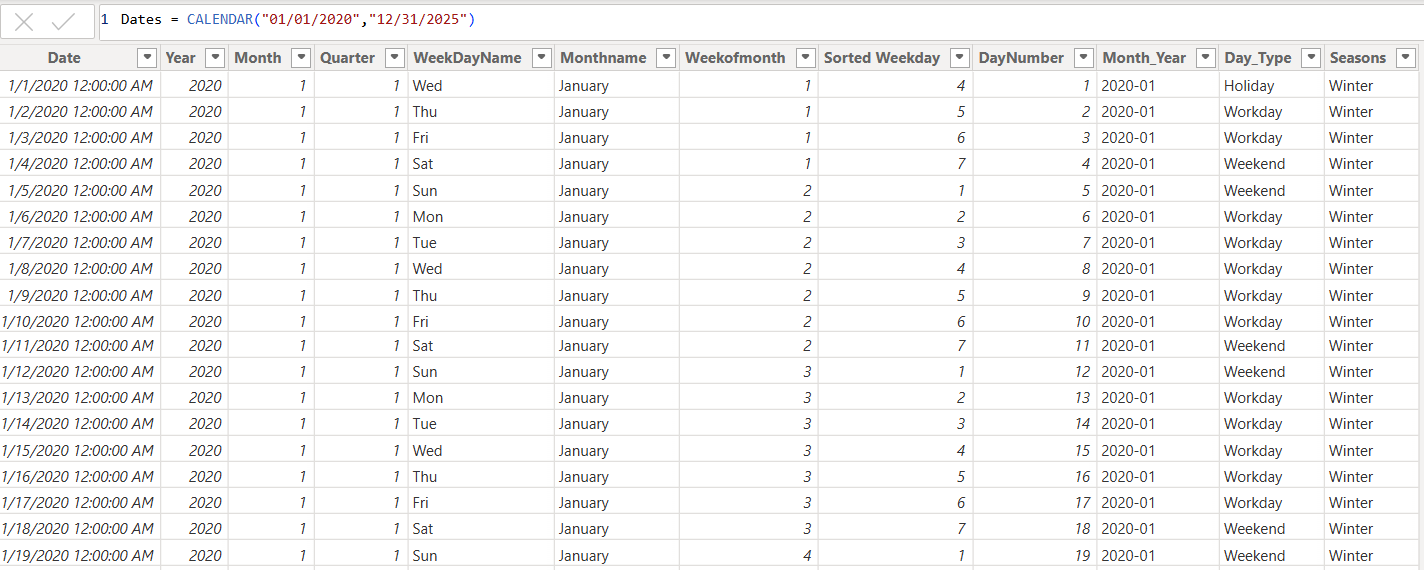
A new table with the URLs of company logos was added to the dataset. This made it possible to include each company’s logo in the visuals, which improved the overall presentation and made the data easier to understand.

4. Logos Table

## 5.4 Dates Table

## Dates Calculated by DAX:

This table was created using DAX functions to enhance data analysis. It features a structured date hierarchy, including year, quarter, month, and day. Additionally, it provides detailed attributes like sorted weekdays and week numbers, allowing for advanced analytical scenarios and time-based comparisons.



5. Dates calculated by DAX

|  |  |
| --- | --- |
| Columns in Dates table calculated by DAX | |
| **Year** | The four-digit year (e.g., 2024). |
| **Quarter** | Indicates the quarter of the year (Q1 to Q4), based on the month. |
| **Month** | The numerical value of the month (e.g., 1 for January, 12 for December). |
| **Monthname** | Full or abbreviated name of the month (e.g., January, Feb). |
| **WeekDayName** | Name of the day in the week (e.g., Monday, Saturday). |
| **Weekofmonth** | Indicates which week of the month a specific date falls in (e.g., 1st week, 2nd week). |
| **Sorted Weekday** | days sorting logically (e.g., "Sunday 1", "Monday 2")—helps in chronological ordering. |
| **DayNumber** | The numerical value of the day within the month (e.g., 1 to 31). |
| **Month\_Year** | A combined field showing both month and year (e.g., "Mar 2024"), useful for sorting and filtering. |
| **Seasons** | Categorizes months into seasons (e.g., Winter, Spring, Summer, Autumn). |
| **Day\_Type** | Categorizes days based on type (e.g., Weekend, Weekday, Holiday). |

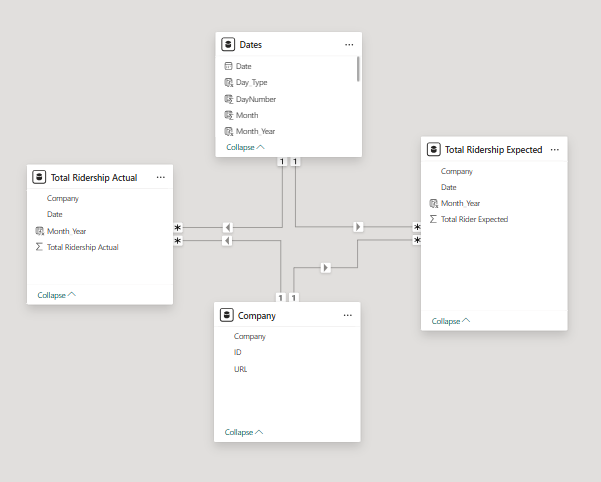
## 6. Calculated Measures (USING DAX):

|  |  |
| --- | --- |
| **Measure** | **Function** |
| Actual Total Ridership Qrt Over Qrt% | Calculates the **percentage change in ridership between consecutive quarters**. It helps analyze how ridership levels are improving or declining from one quarter to the next, |
| Total Ridership Actual average per DayNumber | Calculates the **average number of rides for each day of the month**. It is useful for identifying patterns or variations that occur on specific days, such as weekends or holidays. |
| Total Ridership | Calculates the **total number of rides** across the entire dataset. It gives an overall view of ridership volume, which is crucial for assessing the impact of changes or disruptions in public transport usage. |
| Max\_Ridership\_By\_Month | Identifies the **maximum ridership recorded within a single month**. It helps highlight peak usage periods, which can be essential for capacity planning and identifying high-demand times. |
| Min\_Ridership\_By\_Month | Finds the **minimum ridership recorded in a given month**. It can indicate the lowest point of ridership. |
| Line Chart Address | Creates a **dynamic title for a line chart** that automatically displays the date range of the data. It takes the **earliest date (MIN)** and the **latest date (MAX),** formats them as a long date. The result is a title that looks like this:  **Total Ridership From [Earliest Date] To [Latest Date]** |
| Avg\_Ridership\_By\_Month | Calculates the **average number of rides within each month**, allowing for a comparison between months to identify consistent patterns or anomalies. |
| Max\_Ridership\_Month & Min\_Ridership\_Month | These measures return the month with the highest and Lowest total ridership based on aggregated monthly data. |

# 7. Data Modelling

**Data Modeling Steps:**

1. **Identify Key Entities:** Recognize the primary subjects or concepts represented in the data. Based on the diagram, the key entities appear to be:
   * Dates (containing date-related attributes)
   * Total Ridership Actual (likely representing total ridership after pandemic)
   * Total Ridership Expected (likely representing total ridership % of Comparable Pre-Pandemic date).
   * Company (containing transportation companies’ Logos URLs)
2. **Establish Relationships Between Entities:** Determine how the different entities are connected and define their relationships (one-to-one, one-to-many).
   * The Dates table has a one-to-many relationship (1 : \*) with the Total Ridership Actual table based on the Date column. Multiple daily ridership records belong to a single day.
   * The Dates table has a one-to-many relationship (1 : \*) with the Total Ridership Expected table based on the Date column. Multiple daily ridership records belong to a single day.
   * Total Ridership Actual has a many-to-one relationship (\* : 1) with the Company table based on the Company column. Multiple total ridership records belong to a single company.
   * Total Ridership Expected has a many-to-one relationship (\* : 1) with the Company table based on the Company column. Multiple total ridership records belong to a single company.
3. **Identify Primary and Foreign Keys:** Specify the attributes that uniquely identify records within each entity (primary keys) and the attributes used to establish relationships between entities (foreign keys).
   * **Dates:** Date column is likely the primary key, Date Column acts as a foreign key in other tables.
   * **Company:** Company column is likely the primary key. Company Column acts as a foreign key in other tables.
   * **Total Ridership Actual:** Date and Company act as foreign keys linking to the Dates and Company tables, respectively.
   * **Total Ridership Expected:** Date and Company act as foreign keys linking to the Dates and Company tables, respectively.
4. **Normalize the Data Model (Implicit):** While not explicitly shown, ensure the data model is structured to minimize redundancy and improve data integrity. The separation of dates and company information into their own tables suggests a level of normalization.



6. Data Modelling

# 8. Data Visualization (Report Creation)

## 8.1 Home Page

This dashboard analyzes MTA ridership during and after the COVID-19 lockdown. It shows how ridership trends compare to expectations over time.

The dashboard also analyzes performance and traffic for vehicles and bridges versus expected levels.

It explores individual company performance, including seasonal, weekday, weekend, and holiday effects.

It identifies the main factors affecting total ridership. It displays ridership distribution by dates and Day-Type.

The MTA logo is included, reinforcing the data's source and context

A screenshot of a web page

AI-generated content may be incorrect.

7 Home Page

## 8.2 Answers of Pre-asked Data Analysis Questions:

## Q1: How did the actual total ridership compare to the expected total ridership? What were the actual and expected values?

**A graph of a person

AI-generated content may be incorrect.**Represented by a column chart to visually compare the actual total ridership against the expected total ridership, its purpose is to quickly show the variance between these two values.

One column represents the actual ridership, and the other the expected ridership, allowing for an immediate understanding of how the actual figures compare to the anticipated ones.

This straightforward comparison aids in quickly assessing the overall ridership performance and identifying deviations from projections

8 Actual VS Expected

## Q2: Understand the recovery of total ridership after the pandemic and provide a forecast for upcoming months.

This line chart illustrates the trend of total ridership following a lockdown period and provides a forecast for the upcoming year. The blue line represents the actual ridership over time, starting with a significant drop after the lockdown, followed by a fluctuating but generally upward trend.

Key data points are labeled along the line, showing ridership values at different times, a dashed red line indicates a general upward trend over the observed period.

The right side of the chart extends into a "Forecast for coming Year" represented by a black wavy line within a shaded grey area, suggesting a predicted range for ridership in the near future.

A graph showing the growth of the lockdown

AI-generated content may be incorrect.The goal of this visual is to depict the recovery of total ridership after a significant event like a pandemic-induced lockdown and to offer an outlook on potential ridership levels in the coming months, highlighting both the historical trend and future expectations.

8 Forecasting

A screenshot of a phone

AI-generated content may be incorrect.Q3: Which months experienced the highest and lowest impacts on ridership due to the pandemic?

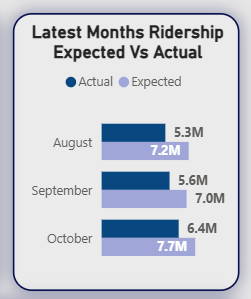
These Cards present the months with the highest and lowest ridership figures.

This visual clearly highlights the extreme points in ridership data.

These cards are designed to automatically update whenever the company or timeframe of analysis changes.

9 Highest and Lowest Month

Q4: Have ridership numbers returned to pre-pandemic levels?

The bar chart compares the actual ridership in the most recent months against the expected ridership for those months.

The visual indicates that in the latest months for which data is available, the actual ridership Subceeded / exceeded the expected ridership. However, these months' performance, even if it surpasses expectations, is not sufficient to definitively conclude whether overall ridership numbers have returned to pre-pandemic levels.

10 Latest Month

Q5: What is the quarter-over-quarter recovery percentage in ridership?

This Column chart displays the percentage change in ridership for each quarter following the lockdown. Each column represents a specific quarter, labeled along the x-axis,The height of each blue column indicates the percentage change in ridership compared to the previous quarter.

For example, the first column shows a significant increase of 136.3% in Quarter 2 of 2020. This is followed by a smaller increase of 60.1% in the subsequent quarter. Some quarters show negative percentages, indicating a decrease in ridership compared to the previous quarter, such as the -66.3% in Quarter 4 of 2024.

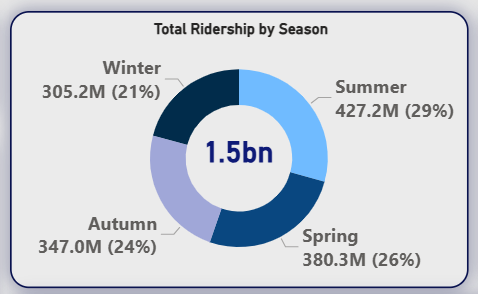
The goal of this visual is to illustrate the rate of recovery (or decline) in ridership on a quarterly basis after the lockdown. It allows for the identification of periods with significant growth and those with setbacks in ridership numbers.

A graph with numbers and a number of blue squares

AI-generated content may be incorrect.

11 QOQ after Lockdown

Q6: What is the impact of different seasons on the total ridership?

The donut chart illustrates the distribution of total ridership across different seasons.

This chart provides the total ridership and the percentage for each season. Summer accounts for the highest ridership.

The goal of this visual is to show the seasonal variation in total ridership. It highlights which times of the year experience the highest and lowest transportation usage, providing insights into seasonal travel patterns.

12 Ridership by Season

Q7: What is the impact of weekdays/weekends/holidays on the average ridership?

A screenshot of a graph

AI-generated content may be incorrect.The multi-gauge chart illustrates the average ridership on different types of days: Workday, Weekend, and Holiday

The goal of this visual is to clearly demonstrate how average ridership varies across different day types. It highlights that weekdays experience the highest average ridership, followed by weekends, with holidays having the lowest average ridership. This information is valuable for understanding daily ridership patterns and can inform operational planning and resource allocation based on the anticipated demand for transportation services on different days of the week and during holidays.

13 Ridership according Day Type

8.3: Supporting Visuals

V1: Heat Calendar

It displays a monthly view, organized by weeks, with the days of the week. The numbers within the cells represent the day of the month

The intensity of the blue shading corresponds to the total ridership on that day. Darker shades indicate higher ridership, while lighter shades indicate lower ridership.

A calendar with numbers and numbers

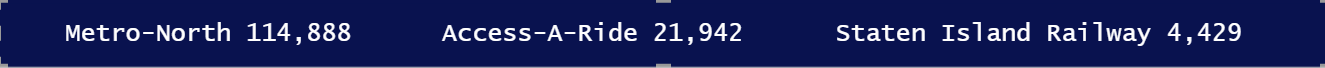
AI-generated content may be incorrect.The goal of this visual is to provide a quick and intuitive overview of daily total ridership. By using color intensity, it allows for the easy identification of days or periods with high or low ridership, potentially highlighting trends related to weekdays, weekends, holidays, or specific events.

14 Heat Calendar

V2: Scroller Bar

The scroller bar displays the average ridership for different transportation modes. Each mode is listed with its corresponding average ridership number. From left to right

The goal of this visual element is to provide a quick overview of the average number of riders for each transportation method.



15 Scroller Bar

V3: Decomposition tree

The Decomposition tree visualizes the breakdown of total ridership based on a hierarchical structure. Starting from the left, the total ridership is initially segmented by different transportation companies or modes.

The visual's goal is to illustrate how total ridership is distributed across these selected categories, providing a detailed drill-down view based on the applied filters. It allows for the examination of ridership volume for a specific company during a particular year, season (, month, and by different day types. The flow lines clearly connect the overall total to its constituent parts under the active filter conditions.

A screenshot of a computer

AI-generated content may be incorrect.

16 Decomposition Tree

A screenshot of a phone

AI-generated content may be incorrect.V4: Slicers and filters

Interactive slicers represent different transportation companies, indicated by a radio button and a label.

The purpose of these slicers is to allow users to filter the data displayed in other associated visuals based on the selected company. By choosing one or multiple options, the user can focus on the ridership data specific to those selections.

These interactive elements enhance data exploration and allow for a more granular analysis of ridership trends for different parts of the transportation system.

17 Company Slicer

Tooltip

The Tooltip displays the "Actual" and "Expected" ridership for various transportation modes across "All/Selected Period.

A screenshot of a graph

AI-generated content may be incorrect.The goal of this visual is to provide insights into the comparison between actual and expected ridership for each company over the entire/selected period. When a user hovers over a specific set of visuals.

18 Tooltip

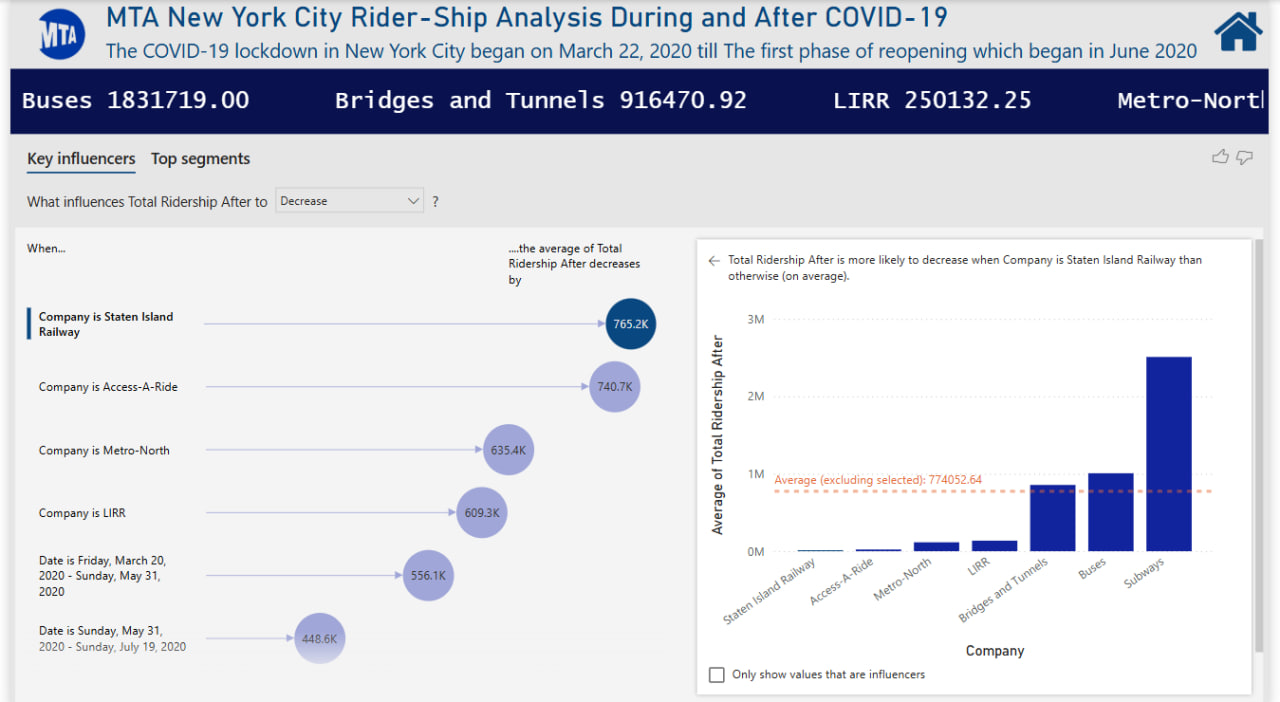
8.4 Key Influencer & Top Segment Page

This page identifies the primary factors that influence total ridership. Key visual elements and information include:

1. Key Influencers Visual “Decrease”:

This visual analyzes "What Influences Actual Total Ridership" It uses a combination of column charts and text descriptions to show the impact of different factors.

* + Company: "Total Ridership is more likely to decrease when Company is Staten Island Railway than otherwise (on average).
  + The visual also indicates the impact of "Access-A-Ride", "Metro-North" , and "LIRR" on decreasing ridership as well.
  + Date: "two main periods “from Friday, March 20, 2020, to Sunday, May 31, 2020", and "from Sunday, May 31, 2020, to Sunday, July 19, 2020" are also shown as factors impacting ridership.



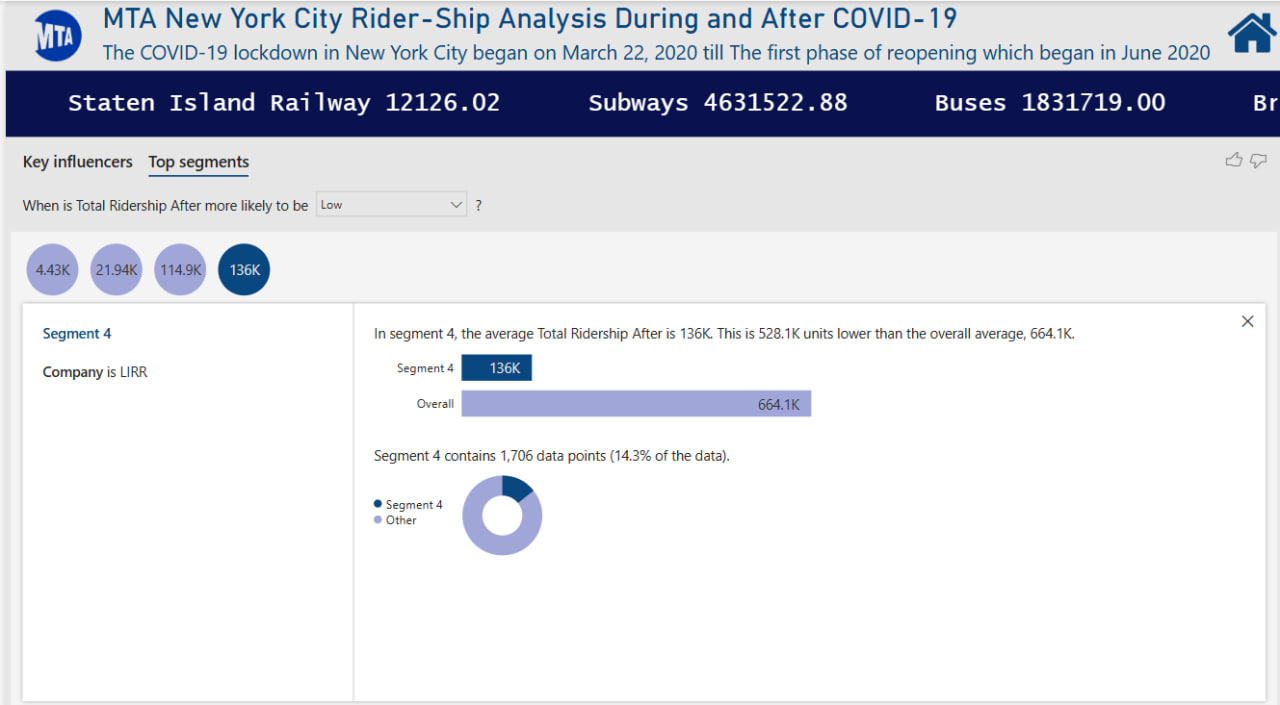
19. Key Influencer

.

1. Analysis of Top Segments Influencing Low Ridership

This page identifies specific segments with low total ridership. Key elements include:

* Example of Top Segments Visual: **Segment 4:** The average Total Ridership After in this segment is 136 K.
* Power BI Context: This page pinpoints a specific condition (LIRR) associated with lower ridership, crucial for identifying areas for improvement.
* Interpretation/Key Takeaways**:** When the company is LIRR, the average ridership is significantly lower than the overall average. This segment's size (14.3% of data) highlights its importance for understanding low ridership trends.

****

20. Top Segment

.

8:5 Are there any remarks or issues that need to be addressed resulting from the pandemic's impact on ridership? (Key Findings)

1. Initial Lockdown Impact (Q2 2020):

The second quarter of 2020, which aligned with strict lockdown measures, brought about a significant downturn in how many people used most forms of transportation. We saw a noticeable drop in ridership for subways, buses, and the various train lines (LIRR, Metro-North, and Staten Island Rail) compared to the first three months of that year. This widespread decrease underscores the immediate and considerable impact of the lockdown, as people generally stayed home and limited their travel to essential trips only.

Interestingly, Access-A-Ride and the use of bridges and tunnels didn't follow this pattern. This could make sense because Access-A-Ride often provides crucial transportation for individuals with specific needs who might not have other ways to get around. Also, bridges and tunnels likely continued to be used by people driving their own cars for necessary journeys, even when public transit was being avoided.

1. Day Type Influence:

The type of day (weekday, weekend, or holiday) demonstrably affects the average ridership across most transportation companies. We observe distinct ridership patterns based on whether it's a typical workday, a less commute-heavy weekend, or a holiday with potentially different travel purposes.

This variability in demand is a critical consideration for the efficient operational planning and resource allocation of services such as subways, buses, LIRR, Metro-North, Staten Island Rail, and Access-A-Ride.

Conversely, the usage of bridges and tunnels appears to be less associated with the typical weekday/weekend/holiday.

1. Seasonal Ridership Similarities with Summer Peak:

While total ridership across all four seasons is relatively close, **Summer consistently exhibits the highest ridership**. suggests that factors such as tourism, school holidays, and warmer weather contribute to increased transportation usage during the summer months compared to Spring, Autumn, and Winter. Although the differences might not be dramatic, the summer peak is a notable trend for capacity planning and service adjustments.

1. Recent Performance Exceeding Expectations (Vehicles & Bridges/Tunnels):

In the most recent full month for which data is available, we've seen that both Access-A-Ride and the use of bridges & tunnels were higher than what we anticipated.

This suggests a positive short-term trend for these specific transportation services. It's an indicator that more people are using them than we had expected.

This recent increase could be happening for a few reasons, such as people possibly depending more on their own vehicles after the pandemic, and restrictions being in place.

**In summary,** while ridership has shown a recovery trend since the initial impact of the lockdown, with some months reaching or slightly exceeding expectations, it's crucial to consider seasonal and daily variations and to compare overall ridership across all modes to pre-pandemic levels for a complete picture of the recovery.

**After all, Subways are still the dominant mode in terms of ridership.**

# 9. Conclusion

* **Significant Initial Impact:** The COVID-19 lockdown, which began in March 2020, had a drastic and immediate effect on ridership companies. The lowest month recorded during the analyzed period was April 2020, with a ridership of only 12.7 million. This starkly contrasts with the average monthly ridership of 114.89 million.
* **Gradual Recovery Post-Lockdown:** Following the initial phase of reopening in June 2020, total ridership began a gradual upward trend. The "Ridership Over Time after Lockdown" chart clearly illustrates this recovery, showing a consistent increase in ridership from the low point in 2020 through 2024.
* **Highest Ridership in Late 2024:** By October 2024, total ridership had reached its highest point in the post-lockdown period, at 167.2 million.
* **Positive Forecast for the Coming Year:** The dashboard includes a forecast indicating continued positive growth in subway ridership for the coming year (2025), projecting a further increase from the 2024 peak.

In summary, the MTA New York City subway system experienced a severe drop in ridership due to the COVID-19 pandemic. However, a consistent recovery has been observed since the initial lockdown, with ridership reaching a post-lockdown peak in late 2024. While forecasts indicate continued growth, the current actual ridership still lags behind pre-pandemic expectations.

# 10. References

**Data Source:** [**MTA Daily Ridership Dataset**](https://github.com/HaniMoussa/MTA-Project-DEPI--G2-6-.git)

**Used Tool:** Power Bi

# 11. Acknowledgement

We extend our heartfelt appreciation to the **Ministry of Communications and Information Technology** (MCIT), the “**Digital Egypt Pioneers**” initiative, and **CLS** Company for their exceptional support and dedication to fostering professional excellence. We are especially grateful to **Eng. Mohamed Attia** for his outstanding mentorship, guidance, and unwavering commitment throughout this initiative and project. His invaluable expertise played a pivotal role in developing our capabilities and ensuring the successful completion of our work.