MTA Daily Ridership – Power BI Project Documentation

1. Project Overview:

The Metropolitan Transportation Authority (MTA) is the largest transit network in North America, serving over 12 counties across New York State and parts of Connecticut. This project analyzes the **MTA Daily Ridership Dataset**, which includes subway, bus, LIRR, Metro-North, Access-A-Ride, bridges & tunnels traffic, and Staten Island Railway data from **March 2020 to 2025**.

The dashboard aims to visualize ridership trends, recovery rates after the COVID-19 pandemic, and year-over-year changes for operational and analytical insights.

---

2. Dataset Overview:

A) Data Source:

The dataset comes from the **MTA Daily Ridership (deprecated 2025)** dataset, originally published as part of MTA's transparency initiative.

B) Data Columns:

Each daily record contains the following metrics:

Category	Description
Total Estimated Ridership / Traffic	Daily ridership or traffic volume
% of Comparable Pre-Pandemic Day	How current ridership compares to the equivalent weekday in 2019
Date	Daily granularity

Transit modes included:

- Subways
- Buses
- Long Island Rail Road (LIRR)

- Metro-North Railroad (MNR)

- Access-A-Ride (AAR)

- Bridges & Tunnels traffic

- Staten Island Railway (SIR)

C) Important Notes from Source Documentation:

- Subway and Bus values are derived from OMNY/MetroCard swipes.

- LIRR & Metro-North ridership is estimated from ticket sales.

- Access-A-Ride values represent **scheduled trips**, not completed trips.

- Recovery percentages use the updated February 2023 methodology.

- Pre-pandemic comparison was discontinued starting **January 2025**.

---

3. Data Preparation & Modeling:

A) Data Cleaning

Steps performed in Power Query:

- Converted all date fields to proper Date type

- Removed null or inconsistent values

- Unpivoted ridership and recovery percentage columns into a fact structure

- Standardized rider type names

- Checked data consistency across modes

B) Data Model Design

We implemented a **Star Schema** containing:

-Fact Table

- Fact_Riders

    o Date

    o Rider Type

    o Total Estimated Ridership

    o % of Comparable Pre-Pandemic

    o Pre-pandemic baseline (calculated)

-Dimension Tables

- **Calendar** table (generated using DAX CALENDAR())

- **Rider Type** dimension

C) Relationships

- Fact_Riders[Date] → Calendar[Date] (One-to-Many)

Fact_Riders[Rider Type] → RiderType[Rider Type] (One-to-Many)

---

4. DAX Measures:

1) Estimated Ridership

Estimated Ridership = SUM(Fact_Riders[Total Estimated Ridership])

2) Avg Recovery

avg.recovery = AVERAGE(Fact_Riders[% of Comparable Pre-Pandemic Day])

3) Pre-Pandemic Ridership

Pre-Pandimc Ridership =

DIVIDE(

   SUM('Fact_Riders'[Total Estimated Ridership]),

   AVERAGE('Fact_Riders'[% of Comparable Pre-Pandemic Day]),

   0

)

4) Recovery Rate

Recovery Rate =

DIVIDE(

   [Estimated Ridership],

   [Pre-Pandimc Ridership],

   0

)

5) YoY %

YoY % =

VAR PrevYear =

   CALCULATE ([Estimated Ridership], DATEADD ('calendar'[Date], -1, YEAR))

RETURN

DIVIDE ([Estimated Ridership] - PrevYear, PrevYear)

---

5. Power BI Dashboard Design:

The dashboard contains multiple analytical perspectives:

- KPI Cards:

- **Top Rider Type**

- **Pre-Pandemic Ridership (Total)**

- **Estimated Ridership (Total)**

- **Recovery Rate**

- **YoY % Change**

- Slicers:

- **Year** (2020–2025)

- **Rider Type** (Subways, Buses, LIRR, MNR, AAR, SIR, Bridges & Tunnels)

- Visuals Created:

Line Charts:

1. **Estimated Ridership by Year & Quarter**

2. **Recovery Rate by Year & Month**

3. **Estimated Ridership by Month**

4. **Estimated vs Pre-Pandemic Ridership by Year**

Clustered Column Chart:

- Estimated vs Pre-Pandemic Ridership by Rider Type

Table:

- Formatted Date

- Rider Type

- Estimated Ridership

- Pre-Pandemic Ridership

6. Key Insights:

- Ridership dropped sharply during the early COVID-19 wave (March 2020).

- Recovery is uneven across rider types:

  - **Subways and Buses** show the highest recovery.

  - **Access-A-Ride and SIR** have lower recovery percentages.

- Seasonal patterns reveal ridership peaks between May–October.

- Year-over-year ridership shows consistent improvement through 2023–2024.

Pre-pandemic comparison data stops at **January 2025** per MTA policy.

---

7. Tools & Technologies Used:

- Power BI Desktop

- Power Query (M Language)

- DAX (Data Analysis Expressions)

Excel (Data staging & validation)

---

8. Conclusion:

This Power BI project provides a comprehensive view of MTA ridership trends during and after the COVID-19 pandemic. Through effective data modeling, DAX calculations, and intuitive visualization, the dashboard delivers insightful analysis to understand transit recovery and rider behavior over time.

This documentation summarizes the technical and analytical work performed in the project.