رواد مصر الرقمية

وزارة الاتصالات وتكنولوجيا المعلومات

MTA Daily Ridership Analysis – Final Report

Submitted by: Data Pioneers

Team Leader: Mohamed Alaa Mahmoud

Team Members:

Mohamed Alaa Mahmoud Mohamed
Haneen Elsayed Mosbah Abdel Aziz
Ahmed Emadeldin Abdelmonim Elboghdady
Ganna Saad Khaled El-Tabakh
Karim Mahmoud Mohamed Ibrahim
Hady Khaled Mohamed Ahmed

Project Title :MTA Daily Ridership Analysis

Project Description

This project presents a comprehensive analysis of the MTA (Metropolitan Transportation Authority) daily ridership data.
Our primary goal is to understand how public transportation usage has changed over time, particularly comparing pre- and post-COVID-19 periods.
Through this analysis, we aim to extract behavioral patterns, identify inefficiencies, and generate actionable insights that can enhance transit operations and support strategic planning.

Primary Objectives:

- Assess how the pandemic affected rider volumes across different stations and time periods.
- Identify behavioral changes in commuter habits during and after COVID-19.
- Evaluate the operational performance of the MTA during crisis and recovery periods.

Key Questions:

- To what extent did ridership decrease during the pandemic?
- Which stations or lines were most and least affected?
- What time frames saw the most significant change in usage?
- Were there areas where service demand remained relatively stable?

Project Significance:

- Data-driven Decision-Making: The findings will support MTA planners and city officials in making informed decisions regarding service adjustments and resource allocation.
- Operational Optimization: By identifying demand shifts, the project helps improve the efficiency of scheduling and operations.
- Rider Behavior Insights: Understanding commuter preferences and habits will guide future improvements in transit accessibility and service quality.

Tools & Technologies Used in the Project

- The successful implementation of the **MTA Daily Ridership Analysis** project relies on a range of integrated tools and technologies specialized in data preparation, analysis, and visualization. Each tool played a key role in streamlining the workflow and producing accurate, meaningful results.

Power BI

- The main platform is used for **visualizing and analyzing data** through dynamic, interactive dashboards. Power BI allowed us to explore usage trends and identify patterns across different time periods using clear and accessible visuals.

Power Query

- Used for **cleaning and transforming the dataset** before analysis. Tasks such as removing duplicates, filtering dates, and reshaping data were efficiently handled to prepare a structured and analysis-ready dataset.

Microsoft Excel

- Served as the **primary data source**, containing daily ridership figures. The data was reviewed and organized in Excel before being imported into Power BI for further transformation and analysis.

DAX (Data Analysis Expressions)

- DAX was used inside Power BI for **custom calculations and advanced analytics**, such as calculating percentage changes in ridership and creating cumulative totals.
Additionally, DAX was used to **generate seasonal classifications (Spring, Summer, Fall, Winter)** based on the ride date, which helped in analyzing how ridership varied across different seasons.

Power BI Service

- A cloud-based platform used to **publish and share reports with team members**, enabling collaboration and scheduled data refreshes to ensure that reports always reflect the most current data.

Objectives

The project's objectives aim to provide a comprehensive analysis of ridership trends, focusing on the changes before and after the **COVID-19** pandemic:

Understand Ridership Trends Over Time

The goal is to identify how public transportation usage has changed across different time periods. Through data analysis, the project will provide a clear picture of the varying trends in ridership.

Analyze Passenger Behavior and Usage Patterns

Examine factors that affect passenger behavior, such as weekday vs. weekend usage and seasonal trends. This analysis will help understand when and where public transportation is used most frequently.

Evaluate Post-Pandemic Recovery Trends

Assess whether ridership has returned to pre-pandemic levels and identify challenges in restoring normal usage levels of public transportation.

Optimize Planning and Decision-Making Based on Data Insights

Utilize the insights derived from the data to enhance service efficiency and improve public transportation strategies. This objective supports decision-making aimed at improving operations and resource allocation.

Planning & Data Understanding

Understanding the Dataset:

The dataset contains several key columns that provide details about ridership across different modes of transportation. Here are the main columns included in the dataset:

Key Columns in the Dataset:

- **Subways**: Ridership data for the subway system.
- **Buses**: Ridership data for buses.
- **LIRR (Long Island Rail Road)**: Ridership data for the Long Island Rail Road trains.
- **Metro-North**: Ridership data for Metro-North trains.
- **Access-A-Ride**: Ridership data for the Access-A-Ride service for people with disabilities.
- **Bridges and Tunnels**: Data for traffic across bridges and tunnels.
- **Staten Island Railway**: Ridership data for the Staten Island Railway.

Total Estimated Ridership

- Represents the absolute number of riders for each mode of transportation.

% of Comparable Pre-Pandemic Day

- This percentage compares current ridership to pre-pandemic levels, helping assess the impact of the pandemic on public transportation usage.

Data Structure:

The dataset consists of:

- **Measures (14 columns)**: Includes total ridership and percentage recovery for each transit mode.
- **Dimensions (2 columns)**: Includes date and transit mode

Data Dictionary

Below is a description of the key fields used in the dataset and the additional fields generated during the data preparation process:

Model Measures

Name	Expression	IsHidden	Table	State
Post Pandemic	SUM(Fact_MTADailyRidership[Total Ridership])	False	_Measures	Valid
Pre Pandemic	DIVIDE(SUM('Fact_MTADailyRidership'[Total Ridership]), AVERAGE('Fact_MTADailyRidership'[%Pre-Pandemic Ridership]) / 100, 0)	False	_Measures	Valid
Recovery Rate	DIVIDE([Post Pandemic], [Pre Pandemic], 0)	False	_Measures	Valid
Ridership Difference	[Weekday Ridership] - [Weekend Ridership]	False	_Measures	Valid
Weekday Ridership	CALCULATE([Post Pandemic], FILTER('Dim_Calender', WEEKDAY('Dim_Calender'[Date], 2) IN { 1, 2, 3, 4,	False	_Measures	Valid

Model Measures

Name	Expression	IsHidden	Table	State

	'Dim_Calender',
	WEEKDAY(
	'Dim_Calender'[Date],
	2
	) IN {
	1,
	2,
	3,
	4,
	5
	}
	)
	)			

Weekend Ridership	CALCULATE([Post Pandemic], FILTER(Dim_Calender, Dim_Calender[WeekEnd?] = "Yes"))	False	_Measures	Valid
Weekend Ridership Rate	DIVIDE([Weekend Ridership], [Weekday Ridership], 0) * 100	False	_Measures	Valid

Model Columns

DataCategory	DataType	FormatString	IsHidden	IsAvailableInMDX
DayOfMonth	Integer		True	True
MonthOfYear	Integer		True	True
Months	Text		True	True
PaddedDateTableDates	Date		True	True
QuarterOfYear	Integer		True	True
Quarters	Text		True	True
Regular	Date	mm/dd/yyyy	False	True
Regular	Integer	0	False	True
Regular	Integer	0	True	True
Regular	Text		False	True
Regular	True/False	"TRUE";"TRUE";"FALSE"	False	True
RowNumber	Integer		True	True
Years	Integer		True	True

Model Table

Name	IsHidden	DataCategory	Expression
_Measures	False	Regular	Row("Column", BLANK())
DateTableTemplate_00a4c597-a513-4dfd-b117-75281b9aa2dc	True	Regular	Calendar(Date(2015,1,1), Date(2015,1,1))
Dim_Calender	False	Regular	CALENDARAUTO()
Dim_TransportMode	False	Regular	
Fact_MTADailyRidership	False	Regular	
LocalDateTable_ea03c049-59dc-4b74-8b56-9e9c411ce6ce	True	Regular	Calendar(Date(Year(MIN('Dim_Calender'[Date])), 1, 1), Date(Year(MAX('Dim_Calender'[Date])), 12, 31))
Model_COLUMNS	False	Regular	INFO.VIEW.COLUMNS()
model_measures	False	Regular	INFO.VIEW.MEASURES()
Model_Relationship	False	Regular	INFO.VIEW.RELATIONSHIPS()
Model_Table	False	Regular	INFO.VIEW.TABLES()

Model Relationship

Relationship	State	SecurityFilteringBehavior	ToColumn	IsActive	ToCardinality
'Dim_Calender'[Date] *[<-]1 'LocalDateTable_ea03c049-59dc-4b74-8b56-9e9c411ce6ce'[Date]	Ready	Single	Date	True	One
'Fact_MTADailyRidership'[Date] *[<-]1 'Dim_Calender'[Date]	Ready	Single	Date	True	One
'Fact_MTADailyRidership'[Transport Mode] *[<-]1 'Dim_TransportMode'[Transport Mode]	Ready	Single	Transport Mode	True	One

Business Questions & Analysis

1. Ridership Trends Over Time

- **How has total ridership in subways and buses changed before and after the pandemic?**

Analyzing the change in ridership across public transit before and after the impact of the pandemic.

- **Are there seasonal patterns in public transit usage (e.g., differences between summer and winter)?**

Identifying whether there are variations in ridership based on seasons.

- **What is the growth or decline rate of ridership over time?**

Analyzing the yearly or monthly trends in ridership growth or decline.

- **What are the busiest months for ridership?**

Identifying the months with the highest ridership.

2. Rider Behavior & Usage Patterns

- **What are the top 5 busiest and least busy days in the subway?**

Analyzing the daily patterns to identify the busiest and least busy days.

- **Do public holidays affect ridership numbers?**

Analyzing the relationship between public holidays and ridership numbers.

- **What is the difference in ridership between weekdays and weekends?**

Studying passenger behavior on weekdays vs. weekends.

- **Does an increase in car traffic reduce public transit ridership?**

Analyzing the relationship between car traffic and public transit usage.

3. Post-Pandemic Recovery

- **Has public transit usage recovered to pre-pandemic levels?**

Analyzing the recovery of public transit usage after the COVID-19 pandemic.

- **How did the COVID-19 pandemic impact ridership?**

Understanding the direct impacts of the COVID-19 pandemic on ridership patterns.

4. Planning & Decision-Making Improvement

- **What is the average daily ridership in the subway?**

Calculating the average daily ridership in the subway.

- **What is the total number of passengers across all transportation modes for each day?**

Calculating the total number of passengers across all modes of public transport on a given day.

- **Which mode of transportation has the highest ridership?**

Identifying the most used transportation mode among all public transit options.

Data Cleaning & Preparation

Tools Used:

- Power BI
- Power Query
- Microsoft Excel

Cleaning Steps:

1. Unpivoting Data

- The data was restructured for better organization, making it easier to filter and analyze. The **Unpivoting** technique was used to transform columns with repeated data into rows, enabling more efficient handling of the data.

2. Removing Duplicates

- Ensured that only unique values remained in the dataset by removing duplicate records, ensuring data accuracy and integrity.

3. Creating Fact & Dimension Tables

- **Fact Tables** were created to store numeric values like ridership counts, while **Dimension Tables** were used to store descriptive information such as date and transit mode. This structure enables more flexible analysis and improves performance when querying the data.

KPIs (Key Performance Indicators)

Public Transit Usage Metrics

- **Total Ridership**

The total number of passengers over time. This indicator provides an understanding of the overall demand for public transportation.

- **Daily/Weekly/Monthly Ridership**

Usage trends over different time periods such as days, weeks, and months, helping to identify patterns and changes in ridership.

- **Average Ridership per Station**

Identifying the busiest and least-used stations, which is important for efficiency analysis and resource allocation.

COVID-19 Impact & Recovery Trends

- **Pre vs. Post-Pandemic Ridership**

Comparing ridership before and after the pandemic to determine whether public transit usage has recovered to pre-pandemic levels.

- **Monthly Ridership Growth (%)**

Tracking monthly ridership trends and improvements, assessing the recovery or decline in usage.

Seasonality & Passenger Behavior

- **Weekday vs. Weekend Ridership**

Analyzing the differences between public transit usage on weekdays versus weekends.

- **Seasonal Impact on Ridership**

Studying how different seasons impact ridership patterns.

Service Optimization & Decision-Making

- **Top Busiest Stations**

Recognizing high-demand locations for service improvements and better resource planning.

Data Model & Architecture

Star Schema Model

The **Star Schema** model was used for structured data representation. This model helps in providing a data structure that allows for accurate and fast analysis by creating relationships between different tables.

Tables Included:

- **Dim_Calendar**

Contains time-based details such as day, month, year, and seasons. This table provides flexibility in analyzing the data based on different time periods.

- **Dim_TransportMode**

Contains information about different modes of transportation (e.g., subway, buses, trains, etc.). It is used to analyze data by transit mode.

- **Fact_MTADailyRidership**

Holds daily ridership data for each transport mode. This is the main table where the data is stored to determine ridership counts over time.

Structured Relationships

- The relationships between the tables are properly structured, enabling accurate and fast analysis. This ensures smooth access to the data through complex queries.

Overview

This report presents an in-depth analysis of MTA daily ridership data using Power BI, focusing on:

Ridership trends before and after COVID-19.

Public transportation usage patterns.

Impact of seasonal variations and weekdays/weekends.

Key insights to support decision-making and optimize transit operations.

Key Performance Indicators (KPIs)

1.Total Ridership After the Pandemic:

- 7.93 billion riders (compared to 25 billion pre-pandemic).

2.Recovery Rate:

- 31.31% (current ridership compared to pre-pandemic levels).

3.Busiest Day:

- Wednesday.

4.Busiest Season:

- Autumn.

5.Most Used Transport Mode:

- Subways.

Transport Mode Comparison

1.Rapid Transit:

- Holds the highest post-pandemic ridership share at 59.46%.

2.Public Transit:

- Public transit usage increased from 18.73% to 22.13% after the pandemic.

3.Infrastructure Usage:

- Showed a slight increase.

Ridership Trends by Days

1.Wednesday:

- Wednesday remains consistently the busiest day, followed by Tuesday and Thursday.

2.Weekend Ridership:

- Weekend ridership remains significantly lower than weekdays.

3. Trends:

- Wednesday had the highest usage rate at 25.95%.

Seasonal Impact on Ridership

- **Autumn:** Recorded the highest ridership post-pandemic at 27.29%.

- **Winter:** Had the lowest ridership at 20.48%.
- **Spring and Summer:** Were nearly equal at around 26%.

The chart highlights ridership fluctuations by season, helping optimize transit services for peak periods.

Weekday vs. Weekend Ridership

- **Total Weekday Ridership:** 6.33 billion riders.
- **Total Weekend Ridership:** 1.61 billion riders.
- **Difference:** 4.72 billion riders (indicating significantly lower weekend usage).
- **Staten Island Railway:** Had the highest weekend ridership rate at 207.52%.
- **Bridges and Tunnels:** Had the lowest weekend ridership rate at 74.75%.
- **Ridership Trends Over the Years**

Gradual recovery in ridership from 2020 to 2024.

Clear evidence of post-pandemic transit usage recovery, but still below pre-pandemic levels.

Key Insights & Recommendations

1. Significant Drop in Ridership Post-Pandemic:

- Post-pandemic ridership is only **8 billion**, compared to **25 billion** pre-pandemic — that's a recovery rate of just **31.31%!**
- This indicates a slow return to public transport usage.

2. Subways Remain the Most Used Mode:

- Subways had the highest ridership both pre- and post-pandemic.
- However, the recovery rate is around **40%,** while modes like Staten Island Railway have less than **25%** recovery.

3.Wednesday Has the Highest Ridership:

- Likely due to being the mid-point of the work week, offering a balance between productivity and routine.

4.Autumn Leads in Seasonal Ridership:

- The most ridership occurs in Autumn, possibly due to pleasant weather and seasonal events.

5.Massive gap between weekday and weekend ridership

- Weekday Ridership: **6.33 billion**
- Weekend Ridership: **1.61 billion**
- Difference: **4.72 billion**.

6.Bridges and Tunnels dominate weekend usage

- Weekend ridership rate: **146.31%** compared to weekdays!
  This may indicate more leisure or intercity travel during weekends using cars via bridges/tunnels.

7.Island Railway has the lowest weekend ridership rate

- Only **42.59%** of its weekday ridership. Suggests low demand or connectivity issues.

8. impact still present on weekend travel

- Based on the chart, around **80%** of people did **not** return to their pre-pandemic weekend travel behavior.

9.(especially Monday, Wednesday, and Friday) have higher total ridership

- The daily analysis chart confirms stronger usage midweek, aligning with standard work patterns.

Recommendations

1. Rebuild Trust in Public Transport Post-COVID:

- Launch awareness campaigns on hygiene, safety, and the benefits of using public transport again.
- Target groups that heavily relied on transit before the pandemic.

2. on Underperforming Transport Modes:

- Improve services and promote awareness for LIRR and Staten Island Railway to boost their usage and recovery.

3. Leverage High Ridership Days and Seasons:

- Offer mid-week promotions or themed travel events on Wednesdays and during Autumn to increase engagement.

4. Public Transit through Infrastructure Development:

- While Rapid Transit dominates, Public Transit has room to grow with strategic expansion and improved user experience.

5. weekday service levels

- As ridership is significantly higher, optimizing bus/rail frequencies and reliability is key.

6. weekend public transport usage

- Launch marketing campaigns, discount fares, or partner events to increase ridership during weekends.

7. Enhance Staten Island Railway services

- Investigate low weekend usage and invest in improving schedules, comfort, or last-mile connectivity.

8. Bridges and Tunnels operations on weekends

- With high weekend demand, consider smoother toll systems or congestion management strategies.

9.. Leverage post-pandemic insights for better planning

- Understand the behavior shift and develop strategies to regain public trust and attract users back, especially on weekends.

Conclusion

In this project, we successfully analyzed **MTA daily ridership data** to uncover key trends, patterns, and insights. Using **Power BI, DAX, and Power Query**, we transformed raw data into meaningful visualizations that provide a clear understanding of public transit usage.

Key Findings

Ridership Trends Over Time:

We observed fluctuations in subway and bus usage, with noticeable dips during weekends and holidays.

Post-Pandemic Recovery: While ridership is gradually recovering, it has yet to reach pre-pandemic levels.

Seasonal & Behavioral Insights: Winter months show lower ridership, while summer and weekdays tend to have higher passenger numbers.

Impact on Decision-Making: The analysis highlights the need for **data-driven strategies** to optimize public transit services, ensuring efficiency and better resource allocation.

Final Thoughts

This project demonstrates the power of **data analytics in public transportation**. By leveraging **Power BI and DAX**, we provided actionable insights that can support better urban mobility planning. As data pioneers, we believe that **data-driven decision-making is the future of smarter cities!**

Next Steps? Future research could **explore real-time ridership predictions, impact of new policies, and AI-driven optimizations** to further enhance public transit efficiency.

" Data speaks louder than words – let's keep exploring! "