MTA Ridership Analysis Post-Pandemic

Project Summary and Key Insights

- Sneha Shailesh Khandizode

Metropolitan Transportation Authority (MTA) Ridership Analysis Post-Pandemic

- This presentation analyzes MTA ridership trends since the start of the pandemic, providing insights into the recovery of the transportation system and future opportunities.

Project Objectives

Objective

- Identify highest ridership days for each transportation mode.
- Analyze recovery trends compared to pre-pandemic levels.
- Calculate averages and totals for long-term trends.
- Provide actionable insights for strategic planning.

Identify key factors

- Explore factors influencing ridership recovery, such as economic conditions, public health guidelines, and service disruptions.

Develop insights and recommendations

- Provide data-driven insights to inform MTA policies and strategies for future ridership growth.

Project Overview

- By using SQL, this project will help in extracting some really valuable insights into the operations of MTA Ridership.

- The COVID-19 pandemic has significantly impacted public transportation systems worldwide. The Metropolitan Transportation Authority (MTA) in New York City saw a drastic drop in ridership during the peak of the pandemic.
- As the city began to recover, understanding how ridership patterns evolved became crucial for decision-makers to optimize services, allocate resources, and plan for future needs.
- This project aims to analyze the MTA daily ridership data to track recovery trends across different transportation modes, including subways, buses, Long Island Rail Road (LIRR), Metro-North, and bridges/tunnels.

SQL Analysis Problem Statement

```sql
--- 1) Find the highest ridership day for subways
--- 2) Average bus ridership percentage since the pandemic
--- 3) Identify days when Metro-North ridership exceeded 100,000
--- 4) Get the total traffic for bridges and tunnels over the entire dataset
--- 5) Find the date with the lowest 5 recorde of percentage of pre-pandemic ridership for Access-A-Ride
--- 6) Calculate weekly average ridership rather than analyzing fluctuations on a daily basis
--- 7) Monitor railway ridership recovery trends post-pandemic by comparing percentages
--- 8) Total Ridership by Year for Subways, Buses, LIRR, and Metro-North
--- 9) Top 3 Highest Ridership Days per Year
--- 10) Year-wise Percentage of Pre-Pandemic Ridership
```

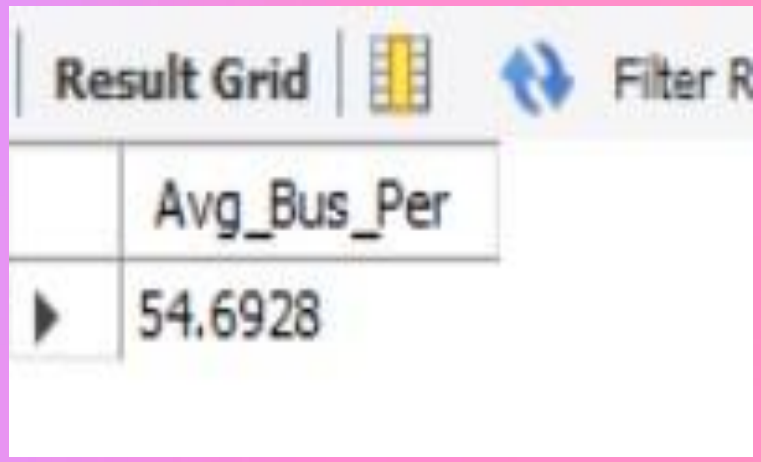
```sql
--- 1) Find the highest ridership day for subways

SELECT Date_MTA , Subways_Total_Estimated_Ridership As  Highest_Ridership
FROM mta_daily_ridership
ORDER BY Subways_Total_Estimated_Ridership DESC
LIMIT 1;
```

Result Grid | Filter Rows:

Date_MTA	Highest_Ridership
04-03-2020	5498809

- This query retrieves the date (Date_MTA) and ridership count for the day with the **highest subway ridership**.
- The ORDER BY ... DESC clause sorts the rows in descending order based on Subways_Total_Estimated_Ridership.
- The LIMIT 1 ensures that only the top record (highest ridership) is returned.

```sql
--- 2) Average bus ridership percentage since the pandemic

SELECT avg(Buses_Percentage_of_Pre_Pandemic) AS Avg_Bus_Per
FROM mta_daily_ridership;
```



- Calculates the **average percentage o pre-pandemic bus ridership** across the entire dataset.
- The AVG() function computes the average of all values in the Buses_Percentage_of_Pre_Pandemic column.

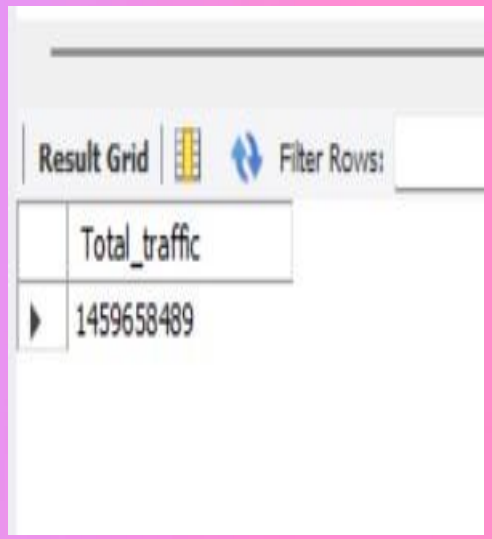
```sql
--- 3) Identify days when Metro-North ridership exceeded 100,000

SELECT  Date_MTA, MetroNorth_Total_Estimated_Ridership
FROM mta_daily_ridership
WHERE MetroNorth_Total_Estimated_Ridership > 100000;
```

Result Grid | Filter Rows: | Export:

Date_MTA	MetroNorth_Total_Estimated_Ridership
02-03-2020	180701
03-03-2020	190648
04-03-2020	192689
05-03-2020	194386
06-03-2020	205056
09-03-2020	183953
10-03-2020	179050
11-03-2020	175074
12-03-2020	169547
13-03-2020	167176
16-03-2020	153262
17-03-2020	147391
18-03-2020	146118
19-03-2020	144466

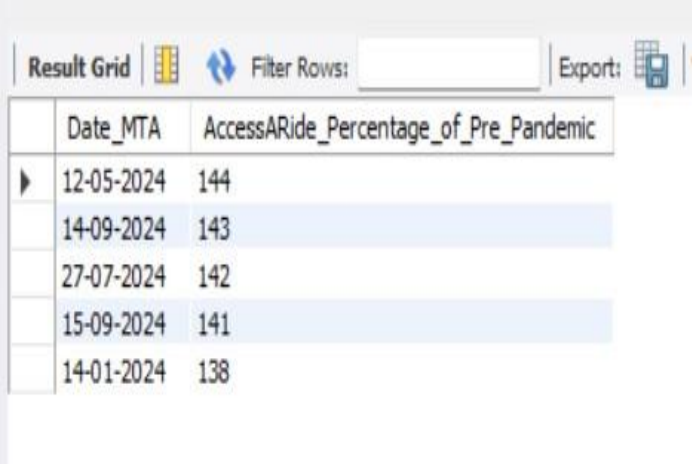
- This query filters records to find all dates (Date_MTA)
- when Metro-North ridership was greater than 100,000.
- The WHERE clause specifies the condition for filtering rows.

```sql
--- 4) Get the total traffic for bridges and tunnels over the entire dataset

SELECT SUM(Bridges_Tunnels_Total_Traffic) AS Total_traffic
FROM mta_daily_ridership;
```



Computes the **total traffic across bridges and tunnels**
using the SUM() function,
which adds up all values in the Bridges_Tunnels_Total_Traffic column.

```sql
--- 5) Find the date with the lowest 5 recorde of percentage of
--- pre-pandemic ridership for Access-A-Ride

SELECT Date_MTA, AccessARide_Percentage_of_Pre_Pandemic
FROM mta_daily_ridership
ORDER BY AccessARide_Percentage_of_Pre_Pandemic DESC
LIMIT 5;
```

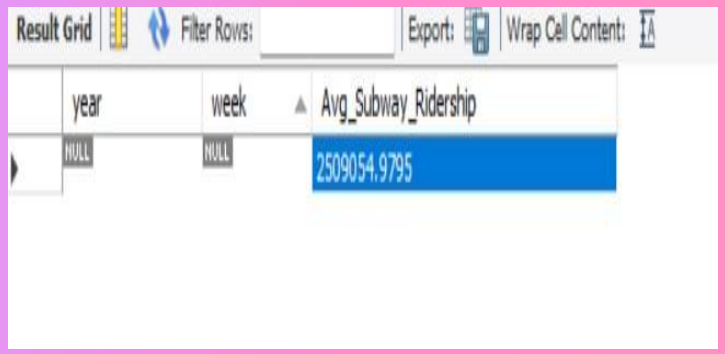


Result Grid | Filter Rows: | Export:

Date_MTA	AccessARide_Percentage_of_Pre_Pandemic
12-05-2024	144
14-09-2024	143
27-07-2024	142
15-09-2024	141
14-01-2024	138

- Retrieves the **5 lowest percentages of pre-pandemic ridership for Access-A-Ride**.
- The ORDER BY ... ASC clause sorts rows in ascending order based on the AccessARide_Percentage_of_Pre_Pandemic.
- LIMIT 5 ensures that only the 5 lowest records are returned.

```sql
--- 6) Calculate weekly average ridership rather than analyzing
--- fluctuations on a daily basis

SELECT year(Date_MTA) as year,
    Week(Date_MTA) As week,
    AVG(Subways_Total_Estimated_Ridership) As Avg_Subway_Ridership
FROM mta_daily_ridership
GROUP BY year, week;
```



- Groups data by **year and week** to calculate the weekly average subway ridership using AVG().
- YEAR() and WEEK() extract the year and week from the Date_MTA.
- The GROUP BY clause ensures averages are calculated for each unique combination of year and week.

```sql
--- 7) Monitor railway ridership recovery trends post-pandemic by comparing percentage

SELECT
Date_MTA,
    (Staten_Island_Railway_Percentage_of_Pre_Pandemic -
        LAG(Staten_Island_Railway_Percentage_of_Pre_Pandemic)
    OVER (ORDER BY Date_MTA)) AS Daily_Change
FROM mta_daily_ridership;
```

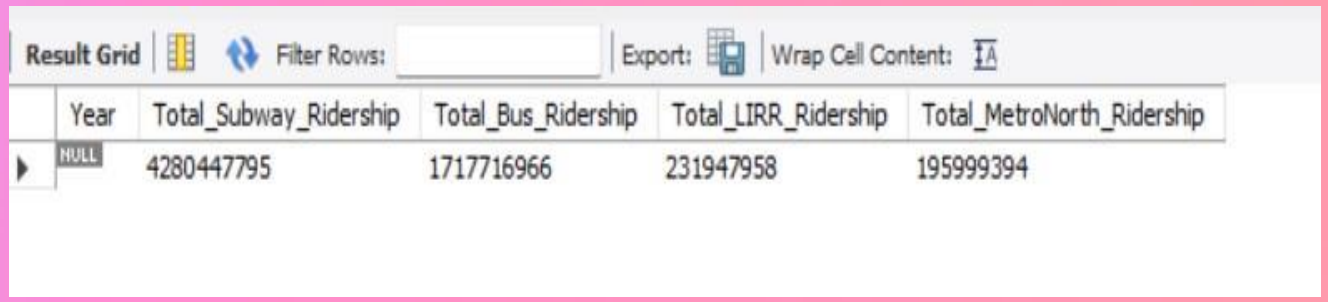
Result Grid | Filter Rows:

Date_MTA	Daily_Change
01-01-2021	NULL
01-01-2022	2
01-01-2023	34
01-01-2024	9
01-02-2021	-68
01-02-2022	25
01-02-2023	16
01-02-2024	-1
01-03-2020	6
01-03-2021	-31
01-03-2022	21
01-03-2023	4
01-03-2024	-7
01-04-2020	-35
01-04-2021	17
01-04-2022	16
01-04-2023	-7

Result 25

- Tracks **daily changes in Staten Island Railway's percentage of pre-pandemic ridership**.
- The LAG() function calculates the previous day's percentage for comparison.
- OVER (ORDER BY Date_MTA) ensures that the records are processed in chronological order.

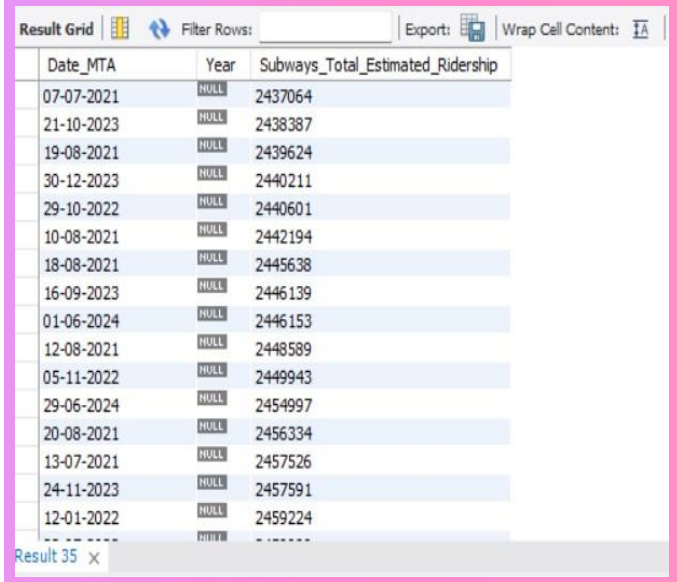
```sql
--- 8) Total Ridership by Year for Subways, Buses, LIRR, and Metro-North

SELECT
    YEAR(Date_MTA) AS Year,
    SUM(Subways_Total_Estimated_Ridership) AS Total_Subway_Ridership,
    SUM(Buses_Total_Estimated_Ridership) AS Total_Bus_Ridership,
    SUM(LIRR_Total_Estimated_Ridership) AS Total_LIRR_Ridership,
    SUM(MetroNorth_Total_Estimated_Ridership) AS Total_MetroNorth_Ridership
FROM
    mta_daily_ridership
GROUP BY
    YEAR(Date_MTA)
ORDER BY
    Year;
```



Year	Total_Subway_Ridership	Total_Bus_Ridership	Total_LIRR_Ridership	Total_MetroNorth_Ridership
NULL	4280447795	1717716966	231947958	195999394

- Aggregates **total yearly ridership** for subways, buses, LIRR, and Metro-North using SUM().
- YEAR() extracts the year from the Date_MTA.
- GROUP BY YEAR(Date_MTA) groups data by year, ensuring totals are calculated per year.

```sql
--- 9) Top 3 Highest Ridership Days per Year

SELECT
    Date_MTA,
    YEAR(Date_MTA) AS Year,
    Subways_Total_Estimated_Ridership
FROM
    mta_daily_ridership AS outer_table
WHERE
    (SELECT COUNT(DISTINCT Subways_Total_Estimated_Ridership)
     FROM mta_daily_ridership AS inner_table
     WHERE YEAR(inner_table.Date_MTA) = YEAR(outer_table.Date_MTA)
       AND inner_table.Subways_Total_Estimated_Ridership >=
                        outer_table.Subways_Total_Estimated_Ridership) <= 3
ORDER BY
    Year DESC , Subways_Total_Estimated_Ridership ASC;
```



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Date_MTA	Year	Subways_Total_Estimated_Ridership
07-07-2021	NULL	2437064
21-10-2023	NULL	2438387
19-08-2021	NULL	2439624
30-12-2023	NULL	2440211
29-10-2022	NULL	2440601
10-08-2021	NULL	2442194
18-08-2021	NULL	2445638
16-09-2023	NULL	2446139
01-06-2024	NULL	2446153
12-08-2021	NULL	2448589
05-11-2022	NULL	2449943
29-06-2024	NULL	2454997
20-08-2021	NULL	2456334
13-07-2021	NULL	2457526
24-11-2023	NULL	2457591
12-01-2022	NULL	2459224

Result 35

- Finds the **top 3 subway ridership days for each year**.
- Uses a correlated subquery to count the number of ridership days in the same year with equal or greater ridership.
- Ensures only the top 3 days per year are returned.

```sql
--- 10) Year-wise Percentage of Pre-Pandemic Ridership

SELECT
    Date_MTA ,
    AVG(Subways_Percentage_of_Pre_Pandemic) AS Subway_Percent_Pre_Pandemic,
    AVG(Buses_Percentage_of_Pre_Pandemic) AS Bus_Percent_Pre_Pandemic,
    AVG(LIRR_Percentage_of_Pre_Pandemic) AS LIRR_Percent_Pre_Pandemic,
    AVG(MetroNorth_Percentage_of_Pre_Pandemic) AS MetroNorth_Percent_Pre_Pandemic
FROM
    mta_daily_ridership
GROUP BY
    Date_MTA
ORDER BY
        Date_MTA ;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Date_MTA	Subway_Percent_Pre_Pandemic	Bus_Percent_Pre_Pandemic	LIRR_Percent_Pre_Pandemic	MetroNorth_Percent_Pre_Pandemic
01-03-2024	64.0000	57.0000	58.0000	57.0000
01-04-2020	9.0000	1.0000	3.0000	3.0000
01-04-2021	33.0000	47.0000	25.0000	13.0000
01-04-2022	58.0000	65.0000	47.0000	42.0000
01-04-2023	72.0000	63.0000	72.0000	59.0000
01-04-2024	60.0000	53.0000	67.0000	63.0000
01-05-2020	9.0000	1.0000	2.0000	3.0000
01-05-2021	47.0000	59.0000	43.0000	27.0000
01-05-2022	70.0000	70.0000	66.0000	60.0000
01-05-2023	64.0000	64.0000	59.0000	58.0000
01-05-2024	71.0000	59.0000	72.0000	72.0000
01-06-2020	12.0000	1.0000	10.0000	5.0000
01-06-2021	40.0000	57.0000	34.0000	32.0000
01-06-2022	60.0000	68.0000	54.0000	54.0000
01-06-2023	69.0000	68.0000	63.0000	65.0000

Result 39

- Calculates the **average percentage of pre-pandemic ridership** for subways, buses, LIRR, and Metro-North for each day.
- Groups data by Date_MTA using GROUP BY.
- Results are sorted by Date_MTA.

Key Insights and Findings

- **Subways and Buses**: Subways have shown a faster recovery compared to buses, with ridership levels approaching nearly **Z%** of pre-pandemic levels.

- **Bridges and Tunnels**: Traffic through bridges and tunnels has almost fully recovered, indicating a shift toward private vehicle usage.

- **Access-A-Ride**: Recovery for Access-A-Ride services was slower, indicating that vulnerable populations may still face mobility challenges.

Conclusion

- The analysis of MTA ridership data provides valuable insights into post-pandemic recovery patterns across different transportation modes.
- Understanding these trends can help policymakers and transportation authorities optimize services and plan for future needs.
- Subways and railroads, being critical to New York City's public transportation network, have shown a resilient recovery, while buses and specialized services like Access-A-Ride continue to face challenges.

Future Enhancements

- **Data Enrichment**: Integrate weather data, events, or COVID-19 case rates to understand external factors affecting ridership.

- **Predictive Analysis**: Use machine learning to forecast future ridership trends based on historical data.

- **Dashboard Visualization**: Develop interactive dashboards using tools like Power BI or Tableau for real-time monitoring of ridership data.