Technical Report

Objective: This data analysis project focuses on **MTA daily ridership** trends, comparing pre-pandemic and current usage across all transportation modes. By leveraging tools such as **Power Query, SQL, Tableau, and Power BI,** we aim to identify the most utilized and efficient transit options, providing insights for service optimization and strategic decision-making.

1) Data Cleaning:

We have reshaped the dataset by Power Query through unpivoting for the original attributes to be **(Date - Year - Month - Weekday - Transportation Mode - Type(Ridership, Traffic, Trip) - Values - Percentage of pre-pandemic days)**

| Date	Subways: To	Subways:	Buses: Tot	Buses:	LIRR: T	LIRR: % of	Metro-North:	Metro-North: %	Access-A-Ride: Tota	Access-A-Ride: %	Bridges and Tunnel	Bridges and Tunnels: %	Staten Isl
9/15/2024	2161737	85	705080	64	128174	129	118303	113	24154	141	946517	107	
9/16/2024	3821428	66	1473843	63	254618	78	215601	75	35319	119	934600	98	
9/17/2024	4189311	73	1536675	66	268678	82	232809	81	38667	130	945571	99	
9/18/2024	4232888	73	1539412	66	281624	86	229697	80	39300	132	963160	101	
9/19/2024	4255184	74	1547727	67	272725	83	226944	79	38515	130	997635	105	
9/20/2024	3933048	68	1471585	63	270065	82	215673	75	36687	124	1019966	107	
9/21/2024	2744632	85	950423	68	160150	135	133469	87	23611	138	980137	103	

Original Dataset

Date	Year	Month	Weekday	Transportation Mode	Type	Value_of_Type	Percentage_Pre_Pandemic_
1/1/2021	2021	January	Friday	Subways	Ridership	613692	29
1/2/2021	2021	January	Saturday	Subways	Ridership	988418	37
1/3/2021	2021	January	Sunday	Subways	Ridership	653187	31
1/4/2021	2021	January	Monday	Subways	Ridership	1557977	30
1/5/2021	2021	January	Tuesday	Subways	Ridership	1597518	31
1/6/2021	2021	January	Wednesday	Subways	Ridership	1613629	31
1/7/2021	2021	January	Thursday	Subways	Ridership	1611099	31

Reshaped Dataset

That will aid in analysis and allow us to gain insights more easily, with greater consistency and comparability across all transportation modes. Additionally, we extract the Year, Month, and Weekday to analyze patterns and drill down into the data.

Also, we get additional data for Weather, Gasoline prices, and New York's Events that happened in this period, and what events will occur in the future that may affect ridership
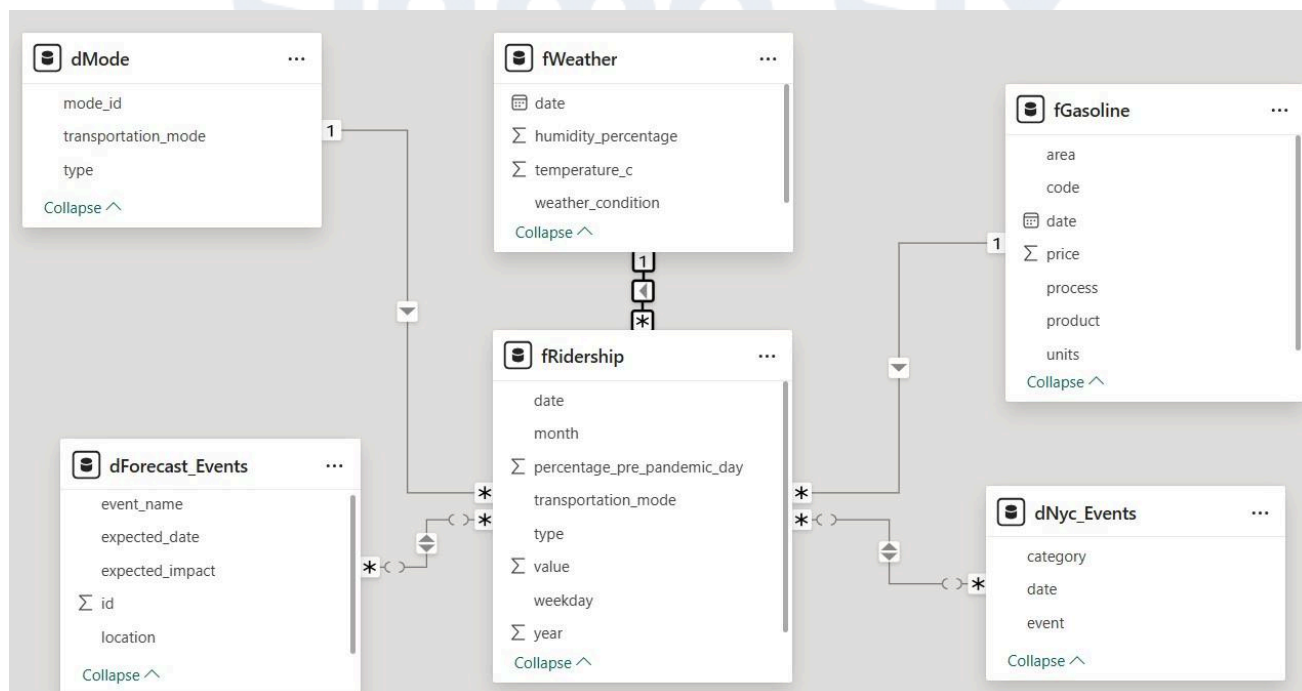
2) The Overall View:

A) *The Most Usage mode* with their Total Ridership at 2024,2020 and overall

```sql
with mta_2024 as (
select transportation_mode as most_usage_2024 , sum(value) as total_ridership24
from mta
where year = 2024 and type != 'Traffic'
Group by 1
order by 2 desc
limit 1 ),
mta_2020 as (
select transportation_mode as most_usage_2020 , sum(value) as total_ridership20
from mta
where year = 2020 and type != 'Traffic'
Group by 1
order by 2 desc
limit 1 ),
mta_all as (
select transportation_mode as most_usage_overall , sum(value) as total_ridership24
from mta
where type != 'Traffic'
Group by 1
order by 2 desc
limit 1 )
select*
from mta_all,
    mta_2020,
    mta_2024
```

Most mode |Overall	Total Ridership	Most mode |2020	Total Ridership	Most mode |2024	Total Ridership
Subways	4,280M	Subways	370M	Subways	987M

That indicates that *Subways* have been the most utilized mode of transportation since the pandemic.

B) _The Lowest Usage mode_ with their Total Ridership at 2024,2020 and overall

```sql
with mta_2024 as (
select transportation_mode as lowest_usage_2024 , sum(value) as total_ridership24
from mta
where year = 2024 and type != 'Traffic'
Group by 1
order by 2 asc
limit 1 ),
mta_2020 as (
select transportation_mode as lowest_usage_2020 , sum(value) as total_ridership20
from mta
where year = 2020 and type != 'Traffic'
Group by 1
order by 2 asc
limit 1 ),
mta_all as (
select transportation_mode as lowest_usage_overall , sum(value) as total_ridership_all
from mta
where type != 'Traffic'
Group by 1
order by 2 asc
limit 1 )
select*
from mta_all,
    mta_2020,
    mta_2024
```

Lowest mode |Overall	Total Ridership	Lowest mode |2020	Total Ridership	Lowest mode |2024	Total Ridership
Staten Island Railway	7.55M	Staten Island Railway	0.72M	Staten Island Railway	1.74M

This shows that _Staten Island Railway_ had the lowest usage mode since the pandemic because not everyone has returned to work, with many still working remotely. This transportation is used for students going to Manhattan for education and tourism.

3) Trend Analysis:

A) Daily ridership changed over time for different transportation modes, as shown in the next Chart:

Trend Usage For Transportation
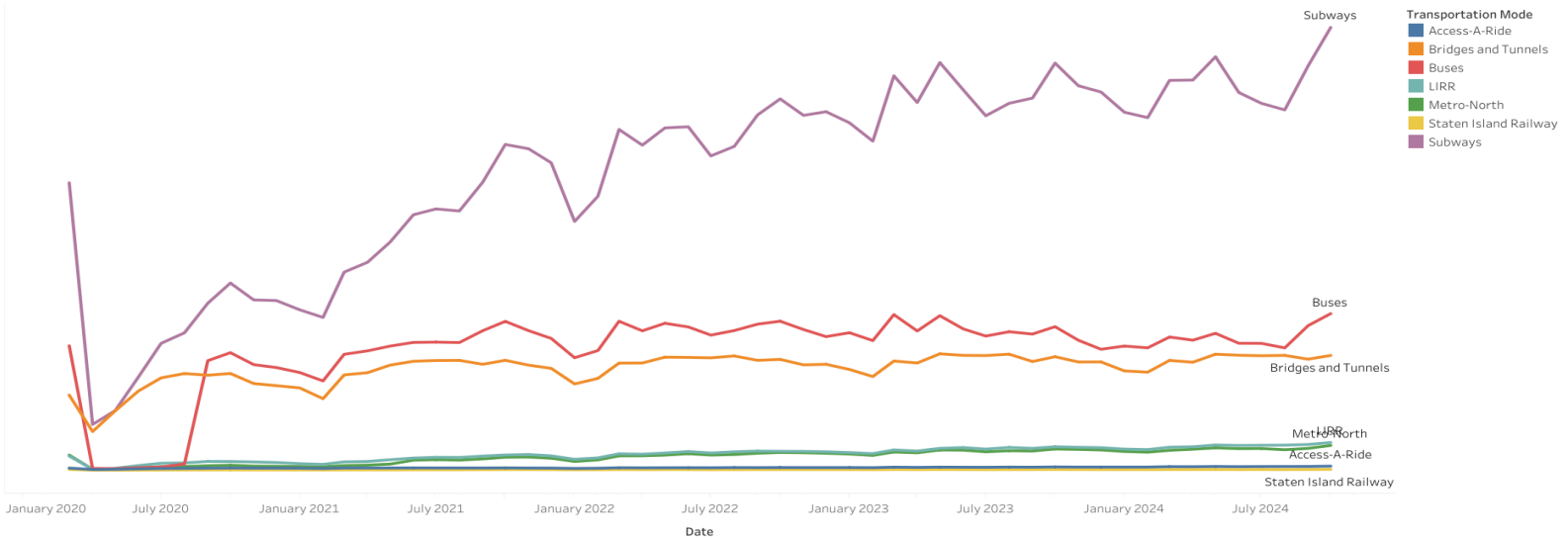
This line chart shows the trend for every transportation mode, which shows the drop in *March 2020* because of the pandemic and how the usage increased after *July 2020* when natural life came back gradually, especially for *subways and bridges*.

B) Seasonal trends in ridership for different transportation modes, as shown in the next Chart:

Seasonal Usage For Transportation
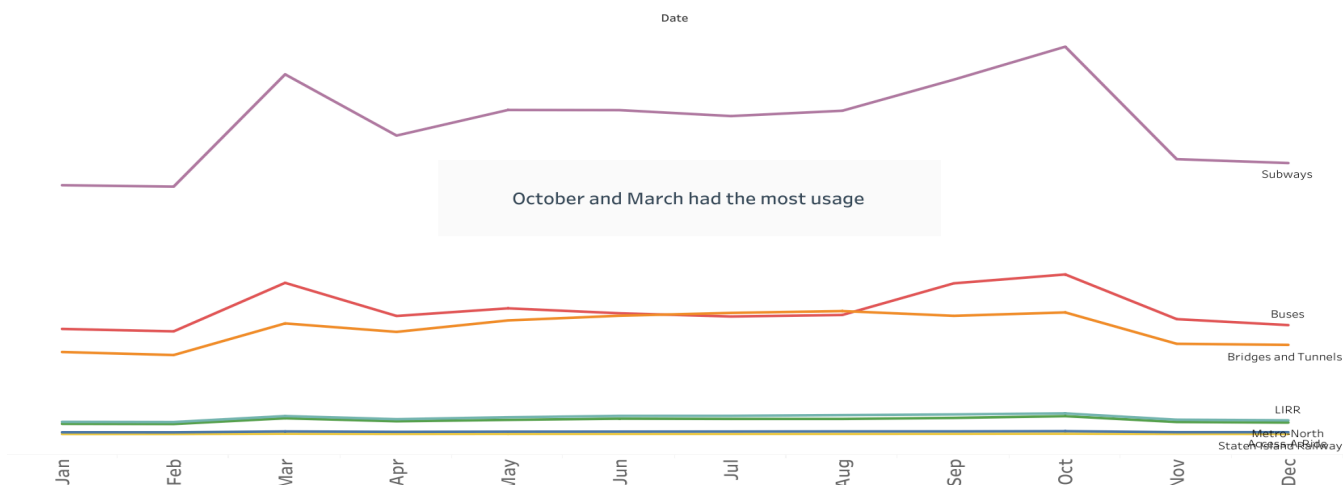
This chart highlights the seasonal ridership peaks in March and October, driven by several key factors:

- _March_ marks the transition from winter to spring, leading to increased movement as the weather improves. Additionally, schools and universities resume full activity, contributing to higher ridership.
- _In March_, companies, shops, and offices return to full operational capacity after the winter slowdown.
- _October_ is one of the most active months for education. Ideal weather conditions encourage more travel across the city.
- Preparations for the NYC Marathon begin in _October_, attracting participants and spectators.
- **Major concerts and many tourists** also contribute to the increased ridership during _October_.

4) Pandemic Impact:

This is the Chart for the average percentage of comparable pre-pandemic days and shows how ridership recovered post-pandemic :

Recovery Percentage Trend Line for Transportation Modes
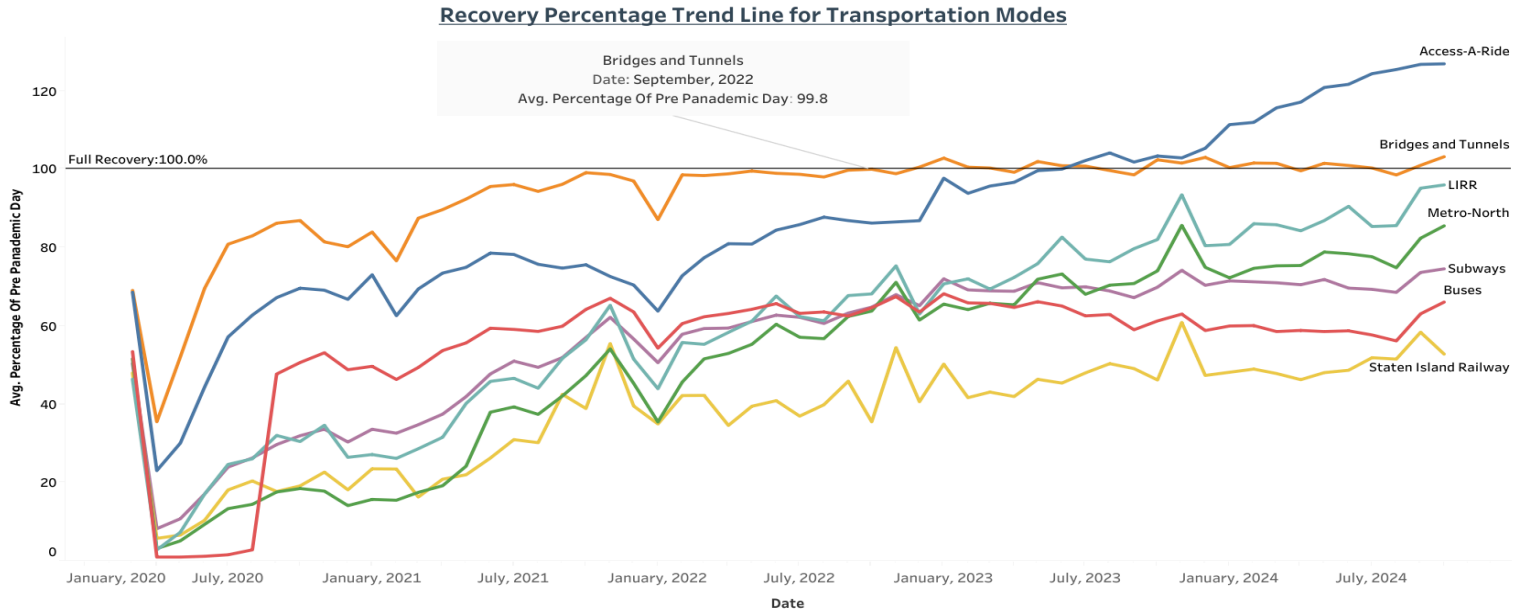
The chart shows that **bridges and tunnels** were **the fastest-recovering** transportation mode by **September 2022**, followed by **Access-a-road**. This is largely due to the increased reliance on **private and personal vehicles for transportation** as many people continue working remotely. Meanwhile, although **Subways** remained the most used mode, they had yet to recover fully.

- ● The overall health of every transportation system Per year:

```
select transportation_mode , year, round(avg(percentage_pre_pandemic_day),2) as overall_health
from ridership
group by 1,2
order by 1
```

The Overall Health of Every transportation mode per year

transportation_mode	2020	2021	2022	2023	2024
Access-A-Ride	55.93	73.35	81.73	100.31	120.23
Bridges and Tunnels	72.53	92.33	98.07	100.97	100.83
Buses	26.14	57.27	62.93	63.59	59.76
LIRR	24.83	43.01	61.66	77.66	87.62
Metro-North	16.57	33.05	56.19	70.84	77.53
Staten Island Railway	18.76	30.84	40.59	47.58	50.25
Subways	26.29	46.43	61.26	70.02	71.17

This table shows that **Bridges and Tunnels** have the highest overall health among transportation modes until **2024. Access-A-Ride** sees a **120%** increase compared to pandemic levels as many people shift away from crowded transportation.

- [The overall health of every Weekday for transportation:](#)

```
select transportation_mode, weekday , round(avg(percentage_pre_pandemic_day),2) as overall_health
from ridership
group by 1,2
order by 1 asc
```

The Overall Health of Every transportation mode per year

Transportation Mode	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Access-A-Ride	82.02	80.27	87.08	90.87	88.64	85.25	89.02
Bridges and Tunnels	93.62	90.39	91.25	93.45	95.66	97.72	91.55
Buses	55.43	54.43	54.77	55.66	54.82	53.7	54.02
LIRR	75.44	54.75	52.37	52.33	52.79	53.53	72.71
Metro-North	61.26	48.18	49.95	49.38	48.73	48.29	51.78
Staten Island Railway	40.99	38.52	37.59	37.79	37.72	35.25	36.81
Subways	61.71	51.58	53.23	54.35	54.24	52.53	60.59

This Table shows Friday is the closest weekday to recovery from the pandemic and return to normal performance by overall Health 97.72%

- **The Overall health per year for Events Day:**

```
select distinct n.event, r.year, round(avg(r.percentage_pre_pandemic_day),2) as OverallHealth
from nyc_events as n
left join ridership as r
on n.date = r.date
group by 1,2
```

Overall Health for Events by Year

Event	2020	2021	2022	2023	2024
2020 Presidential Election	39.14				
2022 Midterm Elections			66.14		
Big Apple Comic Con			66.29	85.07	94.86
Christmas Spectacular	41.73	61.73	70.56	78.61	74.57
Congestion Pricing Announcement				71.71	
COVID-19 Full Lockdown	15.06				
Five Boro Bike Tour	12.71	31.29	74.43	88.57	88.14
Gradual Reopening	37.89	44.25			
Halloween Celebrations	42.71	70.14	63.29	73.14	79.57
Head In The Clouds Festival					91.86
Heat Wave					79.93
Macy's Thanksgiving Day Parade	47.00	79.57	79.57	91.00	
Major Restrictions Lifted		57.83			
New York City Half Marathon				80.57	92.71
New York City Marathon	38.43	77.29	87.00	96.14	
New York Comic Con			75.86	78.68	92.11
New York Fashion Week	43.35	51.44	66.48	74.98	79.75
New York Knicks vs. Brooklyn Nets		48.86			
New York Mets Opening Day		43.00			
New York Rangers vs. New York Islanders		43.43			
New York Yankees Opening Day		43.00			
NYC Broadway Week (Summer)		62.28	70.50	76.11	86.66
NYC Broadway Week (Winter)				73.21	78.04
NYC Culture Week					86.98
NYC Must-See Week				73.21	78.04
NYC Restaurant Week (Summer)		55.90	66.25	76.87	80.67
NYC Restaurant Week (Winter)			56.76	73.21	78.04
Return to Office Policies		48.00			
September 2023 Floods				57.86	
Taylor Swift Dance Party					74.43
The New Yorker Festival					94.10
Tomatito Concert					76.00
Tropical Storm Isaias	23.43				
US Open Tennis Championships	42.22	57.64	69.23	76.69	82.77

This table indicates that The New York City Marathon played a key role in boosting ridership and restoring it to normal levels. The event day recorded the highest average Comparable Pandemic Percentage, at 96.14% in 2023, compared with other events.

- ## The Impact of Events on Ridership on pre-pandemic days:

```sql
WITH events_ridership AS (
    SELECT SUM(r.value) AS events_ridership
    FROM nyc_events AS n
    JOIN ridership AS r
    ON n.date = r.date
),
ridership AS (
    SELECT SUM(value) AS total_ridership
    FROM ridership
)
SELECT ROUND((e.events_ridership::NUMERIC / r.total_ridership::NUMERIC) * 100, 2) AS
events_percentage
FROM events_ridership AS e, ridership AS r;
```

The total ridership on event days accounts for **74.01%** of the pre-pandemic ridership, underlined
the significant impact of events on MTA transit usage.

```sql
WITH events_ridership AS (
    SELECT transportation_mode, SUM(r.value) AS events_ridership
    FROM nyc_events AS n
    JOIN ridership AS r
    ON n.date = r.date
    GROUP BY transportation_mode),
total_events_ridership AS (
    SELECT SUM(events_ridership) AS total_events
    FROM events_ridership)
SELECT e.transportation_mode,
       ROUND((e.events_ridership::NUMERIC / t.total_events) * 100, 2) AS mode_percentage
FROM events_ridership AS e
JOIN total_events_ridership AS t ON 1=1
ORDER BY mode_percentage DESC;
```

The contribution of each transportation mode to event ridership

Transportation Mode	Percentage of Total ridership
Subways	52.56%
Buses	22.19%
Bridges and Tunnels	19.66%
LIRR	2.77%
Metro-North	2.25%
Access-A-Ride	0.48%
Staten Island Railway	0.09%

Based on the table that shows the percentage of each mode to total ridership of events days,
Subways are the most affected mode from events by 52.56%, followed by buses by 22.19%

5) Correlation & Dependency

- **The Correlation between subway and bus ridership**

```sql
SELECT CORR(subways.value, buses.value) AS correlation
FROM ridership subways
JOIN ridership buses
ON subways.date = buses.date
WHERE subways.transportation_mode = 'Subways'
AND buses.transportation_mode = 'Buses';
```

The result indicates that there is no significant correlation between Subways and buses ridership, as the correlation coefficient is 0.06

- **The Correlation between bridge/tunnel traffic levels impact subway ridership:**

```sql
SELECT CORR(s.value, b.value) AS correlation
FROM ridership AS s
JOIN ridership AS b
ON s.date = b.date
WHERE s.transportation_mode = 'Subways'
AND b.transportation_mode = 'Bridges and Tunnels';
```

The result indicates a strong positive relationship between Subway, Bridge, and Tunnel ridership, as the correlation coefficient is 0.74.

- **The Correlation between bridge/tunnel traffic levels impact Buses ridership:**

```sql
SELECT CORR(t.value, b.value) AS correlation
FROM ridership AS t
JOIN ridership AS b
ON t.date = b.date
WHERE t.transportation_mode = 'Buses'
AND b.transportation_mode = 'Bridges and Tunnels';
```

The result indicates a strong positive relationship between Bridges, tunnels, and Bus ridership, as the correlation coefficient is 0.69.

- **The Correlation between weather temperature and its impact on total ridership:**

```sql
with total_ridership as (
        select date, sum(value) as total_ridership
        from ridership
        group by 1)
SELECT CORR(r.total_ridership, w.temperature_c) AS correlation
FROM total_ridership AS r
left JOIN weather AS w
ON r.date = w.date
```

The results indicate a very weak correlation between temperature and total ridership, with a correlation coefficient of 0.05, suggesting that temperature has a minimal impact on transit usage.

6) Performance Comparison

- The Transportation mode has retained the highest % of pre-pandemic ridership:

```
select transportation_mode as highest_mode, Avg(percentage_pre_pandemic_day) as Retained_Percentage
from ridership
group by 1
Order by 2 desc
limit 1
```

The query results indicate that **Bridges and Tunnels** retained the highest percentage of pre-pandemic ridership at **93.38%**. This is likely due to a shift in transportation preferences, as many people prefer private cars and Uber after the pandemic to avoid crowded public transit.

- How does weekday vs. weekend ridership differ

```
SELECT
    CASE
        WHEN weekday IN ('Saturday', 'Sunday') THEN 'Weekend'
        ELSE 'Workday'
    END AS day_type,
    year,
    ROUND(AVG(percentage_pre_pandemic_day), 2) AS OverallHealth,
    SUM(value) AS TotalRiderShip
FROM Ridership
GROUP BY 1,2;
```

Overall Health of WorkDay and Weekend					
Day Category	2020	2021	2022	2023	2024
Workday	34.85	52.85	65.07	73.56	77.46
Weekend	33.39	56.04	68.52	81.52	90.22

Usage of WorkDay and Weekend					
Day Category	2020	2021	2022	2023	2024
Workday	606 M	1204 M	1496 M	1631 M	1390 M
Weekend	152 M	316 M	376 M	415 M	346 M

These tables indicate that **weekend** days recover faster than **weekdays**, even though weekdays have the highest ridership due to work and study commitments.

- The percentage contribution of each mode to total ridership:

```
select
    transportation_mode,
    sum(value) AS total_ridership,
    round((sum(value) * 100 / (select sum(value) from ridership)),2) AS
percentage_contribution
from ridership
group by 1
order by 3 DESC;
```

Transportation mode with their Contribution of Total Ridership		
Transportation Mode	**Total Ridership**	**% of Mode Contribution**
Subways	4280 M	**53.97%**
Buses	1718 M	**21.66%**
Bridges and Tunnels	1460 M	**18.41%**
LIRR	232 M	**02.92%**
Metro-North	196 M	**02.47%**
Access-A-Ride	37 M	**00.47%**
Staten Island Railway	8 M	**00.10%**

This table shows that Subways account for 53.97% of total ridership in New York. They are followed by buses, which contribute 21.66%.

7) Forecasting
- We will calculate the function of line regression between every Mode (Ridership - %Pandemic) and Date:

    a) Ridership Forecasting:

Function of Line Regression for each mode:

Modes	Ridership Line Regression Function
All Modes	Total Ridership = 2262.82*Day of Date + -9.65979e+07
Access - A- Ride	Total Ridership = 10.9779*Day of Date + -469247
Bridges and Tunnels	Total Ridership = 173.654*Day of Date + -6.91427e+06
Buses	Total Ridership = 390.55*Day of Date + -1.64677e+07
LIRR	Total Ridership = 107.654*Day of Date + -4.68085e+06
Metro-North	Total Ridership = 105.594*Day of Date + -4.60977e+06
Staten-Railway	Total Ridership = 2.43258*Day of Date + -104413
Subways	Total Ridership = 1471.96*Day of Date + -6.33516e+07

We will use this function to forecast the Total ridership for every mode on upcoming events:

After running the forecasting model in Tableau, the prediction estimates that total ridership will reach 10.26 million in June 2025. We recommend considering the likelihood of an increase due to the World Cup, which is expected to drive higher ridership.

Strategic Analysis

a) Service Optimization & Efficiency
  - Based On Ridership Patterns, MTA could optimize transportation as follows:
    -In Level Weekday :

```sql
select
     weekday,
     sum(value) as total_ridership
from ridership
where weekday not in ('Sunday','Saturday')
group by 1
order by 2 desc
```

Usage in Workdays	
Workday	Total Ridership
Wednesday	1317 M
Thursday	1299 M
Tuesday	1292 M
Friday	1245 M
Monday	1174 M

The table indicates that 'Wednesday' experiences the highest ridership on weekdays, while 'Monday' has the lowest. (rather than Weekends) Based on this:

- MTA could schedule maintenance on 'Monday' to minimize disruptions.
- Reducing work hours on 'Monday' and reallocating them to other days could help save power and resources.
- Enhancing transit services during peak 'Wednesday' hours, avoiding maintenance, and strengthening emergency control measures can help prevent breakdowns.

-In Level Months :

```sql
WITH highest_month AS (
    SELECT
        month AS month_name,
        SUM(value) AS total_ridership,
        'Highest' AS category
    FROM ridership
    GROUP BY month
    ORDER BY total_ridership DESC
    LIMIT 3
),
lowest_month AS (
    SELECT
        month AS month_name,
        SUM(value) AS total_ridership,
        'Lowest' AS category
    FROM ridership
    GROUP BY month
    ORDER BY total_ridership ASC
    LIMIT 3
)
SELECT * FROM highest_month
UNION ALL
SELECT * FROM lowest_month;
```

Months with Highest Usage		
Month	**Total Ridership**	**Category**
October	817 M	Highest
September	762 M	Highest
March	757 M	Highest
Months with Lowest Usage		
Month	**Total Ridership**	**Category**
February	522 M	Lowest
January	530 M	Lowest
December	572 M	Lowest

The table shows the three months (October - September - March) with the highest usage due to Education and the normal weather as many events in New York are based on this table :

- MTA should increase transits during these months and avoid any maintenance.
- Enhanced ventilation systems to improve air quality.
- Introduce dynamic scheduling to handle peak and off-peak hours efficiently.

In summary, improving services is essential during the busiest months, while maintaining reduced hours in the least busy months helps power saving. Additionally, it's important to consider the workday patterns previously discussed.

- *The strategies can be used to increase public transit adoption post-pandemic:*

To encourage more people to use public transit after the pandemic, transit agencies can implement the following strategies:

1. Improve Safety & Hygiene Measures
   - Regular deep cleaning of buses, trains, and stations.
   - Installing hand sanitizers and touchless payment systems.
   - Enhanced ventilation systems to improve air quality.

2. Optimize Schedules & Reduce Wait Times
   - Adjust the frequency based on demand using ridership data analysis.
   - Provide real-time tracking through mobile apps.
   - Introduce dynamic scheduling to handle peak and off-peak hours efficiently.

3. Offer Incentives & Affordable Pricing
   - Implement discounts for students, seniors, and frequent riders.
   - Introduce loyalty programs or free-ride days.
   - Offer flexible subscription plans (e.g., pay-per-use instead of monthly passes).

4. Enhance Connectivity & Accessibility
   - Improve integration with micro-mobility options (e.g., bikes, scooters).
   - Ensure seamless transfers between different transit modes.
   - Expand transit routes to underserved areas.

5. Leverage Data & AI for Smarter Transit Decisions
   - Use predictive analytics to anticipate demand shifts.
   - Optimize fleet distribution and service planning.
   - Implement AI chatbots for customer support and trip planning.

6. Promote Sustainable & Green Initiatives
   - Transition to electric or hybrid buses.
   - Offer incentives for eco-friendly commuting.
   - Launch public awareness campaigns on the environmental benefits of transit.

7. Improve Passenger Experience & Marketing
   - Enhance station aesthetics and seating areas.
   - Provide free Wi-Fi and charging ports onboard.
   - Run social media campaigns to rebuild trust in public transit.

b) Financial and Customer Experience:

- *The congestion pricing play a role in shifting car users to public transportation:*

```sql
SELECT
    r.date,
    g.price,
    sum(r.value) as total_ridership
FROM gasoline as g
JOIN ridership as r ON g.date = r.date
group by r.date,g.price
ORDER BY r.date;
```

Impact Gasoline with Ridership

Ridership increases when fuel prices rise, indicating a shift to public transit. However, the 2023 decline suggests other factors, like remote work or economic conditions. The 2020 drop aligns with COVID-19 impacts.

- *How can MTA improve accessibility for Access-A-Ride users:*

```sql
SELECT
transportation_mode,year , round(avg(percentage_pre_pandemic_day),2) as Overall_health
from ridership
where transportation_mode = 'Access-A-Ride'
group by 1,2
```
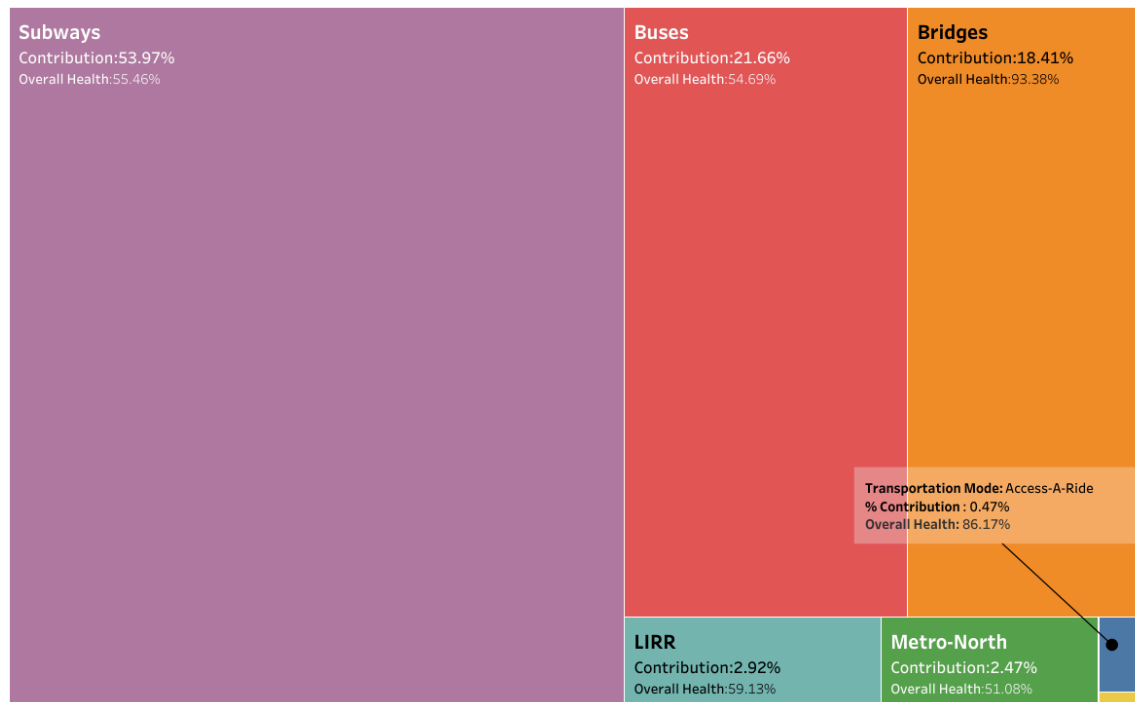
Flow Recovery of Access-A-Ride Mode
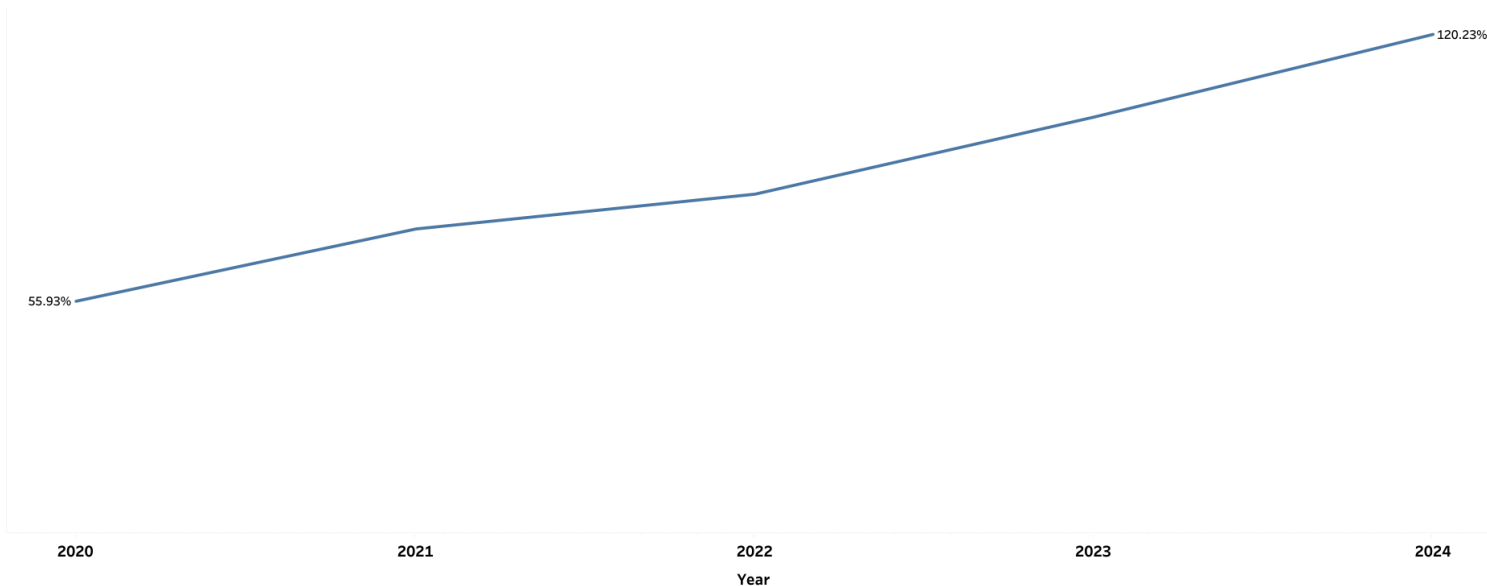
Access-A-Ride experienced a significant recovery growth, surpassing normal levels with **an overall increase of 120.23%**. This suggests a rising shift towards its usage in the past year. Therefore, MTA should implement improvement measures, such as:

- Enhance booking efficiency with a user-friendly system
- Optimize routes for faster and more reliable service
- Upgrade accessibility features for better user experience.
- Integrate with public transit for seamless travel.

- *The Measures that MTA Must Take to Improve Passenger Comfort and Satisfaction:*

Contribution Of Transportation in Riderships

Subways
Contribution:53.97%
Overall Health:55.46%

Buses
Contribution:21.66%
Overall Health:54.69%

Bridges
Contribution:18.41%
Overall Health:93.38%

Transportation Mode: Access-A-Ride
% Contribution : 0.47%
Overall Health: 86.17%

LIRR
Contribution:2.92%
Overall Health:59.13%

Metro-North
Contribution:2.47%
Overall Health:51.08%

Based on this chart, which illustrates the contribution of each transportation mode along with their overall health, MTA should implement the following measures:

- Enhance Subway Conditions: Since subways account for 53.97% of ridership, improving cleanliness and seating can greatly impact passenger satisfaction.

- Optimize Bus Services: With a 21.66% contribution, ensuring more frequent, reliable, and less crowded buses can improve the overall experience.

- Maintain Bridges and Roads: Bridges **contribute 18.41% and have 93.38% overall health**, indicating they are well-maintained. Keeping this standard high is essential.

- Improve Accessibility for Access-A-Ride Users: Although it has **a low contribution of 0.47%**, its overall health of 86.17% indicates increased reliance. MTA should enhance **scheduling, coverage, and vehicle comfort to better serve these users.**

- Enhance LIRR and Metro-North Services: **These modes contribute 2.92% and 2.47%, respectively. MTA may have to postpone any enhancements and save money until the other modes recover.**

- Improve Customer Communication: Providing real-time updates, better signage, and transparent service changes can increase passenger trust and satisfaction.