MTA DAILY RIDERSHIP

ANALYSIS & FORECASTING

1- DATA DESCRIPTION

MTA Daily Ridership Data:

Our dataset tracks daily ridership and traffic volumes across multiple NYC transit services starting from March 1, 2020, capturing the impact and recovery from the COVID-19 pandemic.

Included Transit Modes:

- Subways
- Buses
- Long Island Rail Road (LIRR)
- Metro-North Railroad
- Staten Island Railway
- Access-A-Ride (Paratransit)
- Bridges & Tunnels (Vehicle traffic)

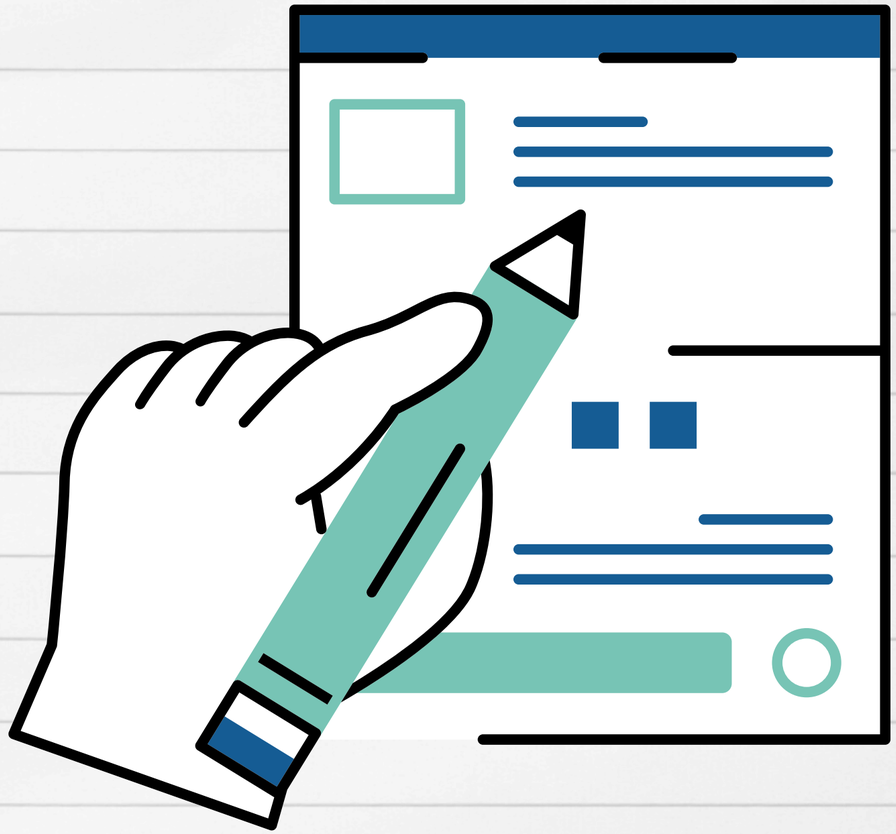
Key Metrics (for each mode):

- Total Estimated Ridership/Traffic
- % of Comparable Pre-Pandemic Day
- (shows how current ridership compares to typical pre-COVID levels)

Timeframe:

- Covers 1,706 daily records with no missing data.

2- EXCEL



Data Cleaning in Excel

- **Import CSV File**

via File > Open or Data > Get External Data

- **Convert Date Column to Date Format**

Format Cells > Date

- **Check for Missing Values (non found)**

Use Filters or Conditional Formatting > Blanks

- **Format Numbers**

Comma style for totals

Percentage style for % columns

- **Renam Columns for Clarity**

Data Visualization in Excel

- **Line Chart for Time Trends**

Plot Date vs. Ridership for each mode

- **Pivot Tables for Summaries**

Aggregate by month, mode, or % recovery

- **Add Slicers or Filters**

Enable interactive exploration

- **Bar/Column Charts**

Compare recovery across transit types

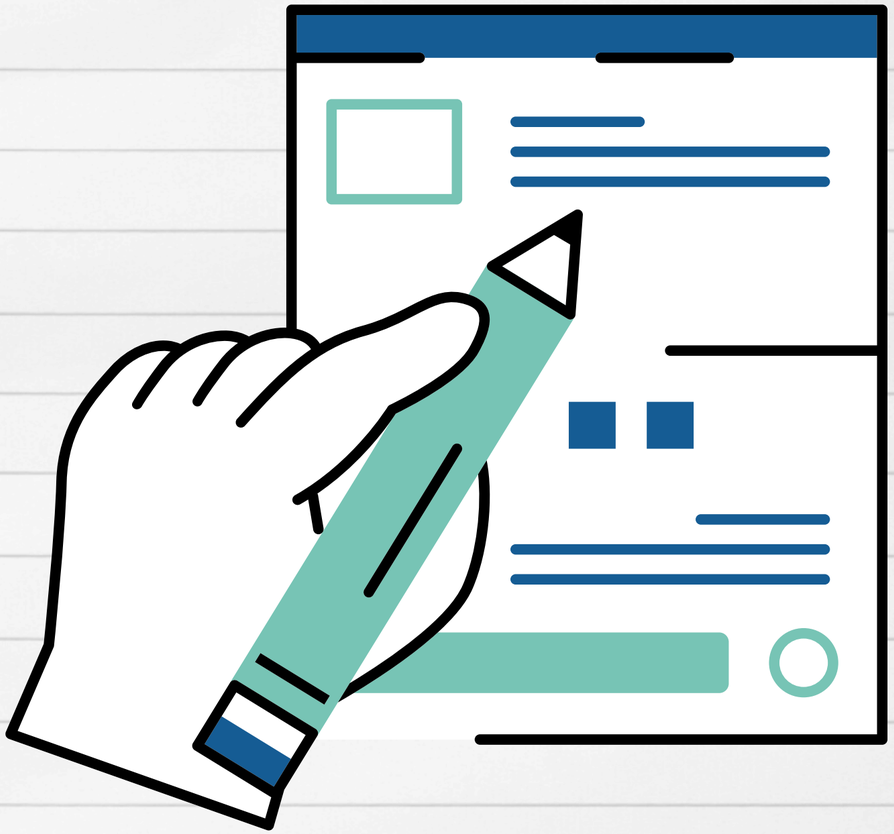
3-SQL

SQL steps

- **Uploaded the csv to** sqliteonline.com
- **ensured that the first row is used as headers having the column names :** founded that it didn't
- **used:** DROP TABLE IF EXISTS your_table_name;
- **Reuploaded the data and inserted first row as header**
- **Viewed the data:** SQL CopyEdit SELECT * FROM your_table_name;
- **Exported the Database:** Go to File → Export DB → Download DB

```
1    SELECT * FROM MTA_Daily_Ridership;
2    DROP TABLE IF EXISTS MTA_Daily_Ridership;
3
```

4- PYTHON



DATA PREPROCESSING

Goal: Clean and structure the MTA Daily Ridership dataset for analysis.

Steps:

1. Loaded dataset using pandas.
2. Converted date strings to datetime objects.
3. Created new features:
   - day_of_week, month, year
   - is_weekend flag
4. Checked for and handled missing values.
5. Standardized column names.

DATA PREPROCESSING

```python
import pandas as pd

# Load and clean dataset
df = pd.read_csv("MTA_Daily_Ridership.csv")
df['date'] = pd.to_datetime(df['date'])
df['day_of_week'] = df['date'].dt.day_name()
df['month'] = df['date'].dt.month_name()
df['year'] = df['date'].dt.year
df['is_weekend'] = df['day_of_week'].isin(['Saturday', 'Sunday'])
```

EXPLORATORY DATA ANALYSIS

Goal: Derive insights by answering key business questions.

INSIGHTS GAINED:

1. **Recovery Post-COVID:** Bridges and Tunnels rebounded fastest.
2. Weekday vs. Weekend: Weekends showed lower but more stable ridership.
3. Least Drop: Access-A-Ride maintained more consistent usage.

INSIGHTS GAINED:

4. Correlations: Access-A-Ride and Subway showed weak correlation.
5. Seasonality: Ridership dips in winter, peaks in fall/summer.
6. Peak Days/Months: Tuesdays and October had high ridership.

EXPLORATORY DATA ANALYSIS

```
import seaborn as sns
import matplotlib.pyplot as plt

# Example: Correlation heatmap
modes = ['subways', 'buses', 'lirr', 'metro_north', 'access_a_ride', 'bridges_
plt.figure(figsize=(10, 6))
sns.heatmap(df[modes].corr(), annot=True, cmap="coolwarm")
plt.title("Correlation Between Modes of Transportation")
plt.show()
```

FORECASTING

Goal: Predict ridership for the upcoming month.
Model Used: Linear Regression
Steps:

1. Created total_ridership metric
2. Trained a model on monthly totals
3. Forecasted next month's ridership

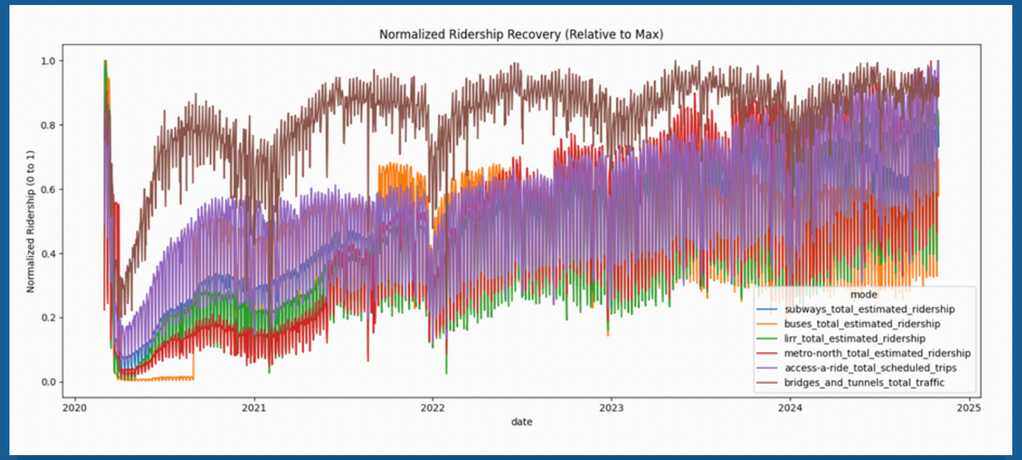
Forecast Insight: Estimated steady increase in subway and bus ridership post-pandemic.

FORECASTING

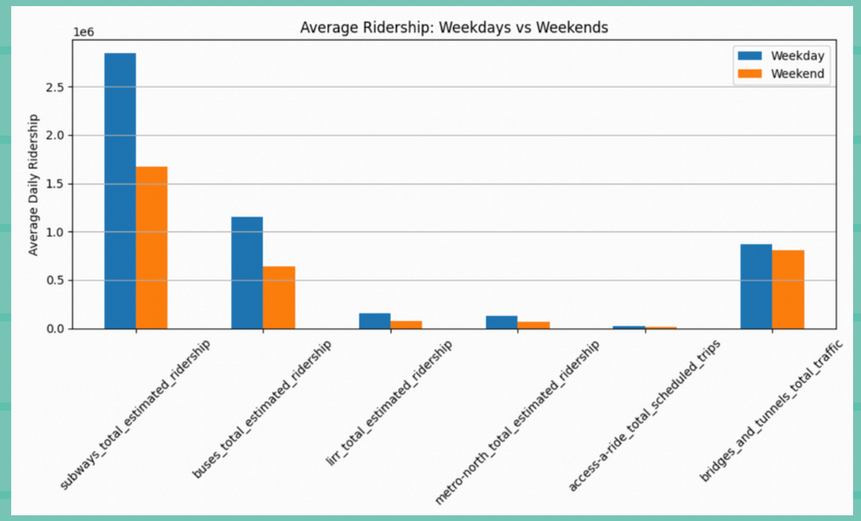
```
from sklearn.linear_model import LinearRegression

monthly_data = df.groupby(['year', 'month']).agg({'subways': 'sum'}).reset_ind
monthly_data['month_num'] = pd.to_datetime(monthly_data['month'], format='%B')
monthly_data['time'] = range(len(monthly_data))

model = LinearRegression()
model.fit(monthly_data[['time']], monthly_data['subways'])
next_month = model.predict([[len(monthly_data)]])
```

SOME OF OUR VISULIZATIONS

01

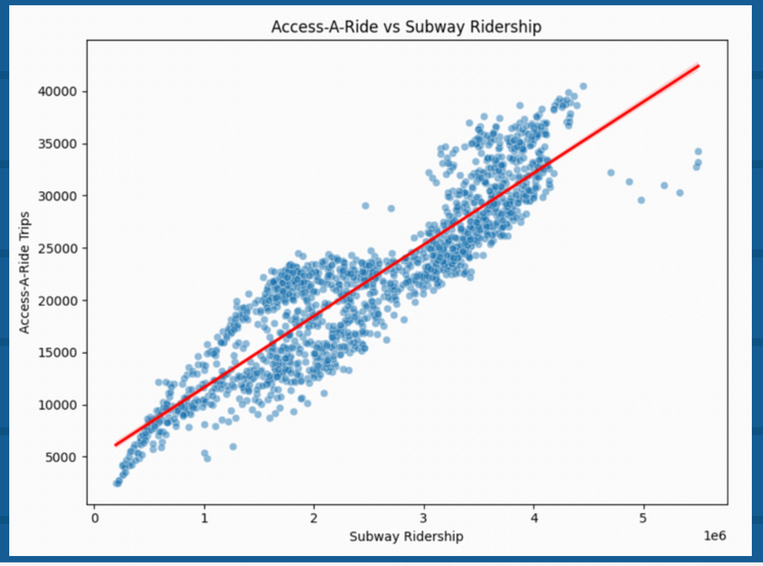


02

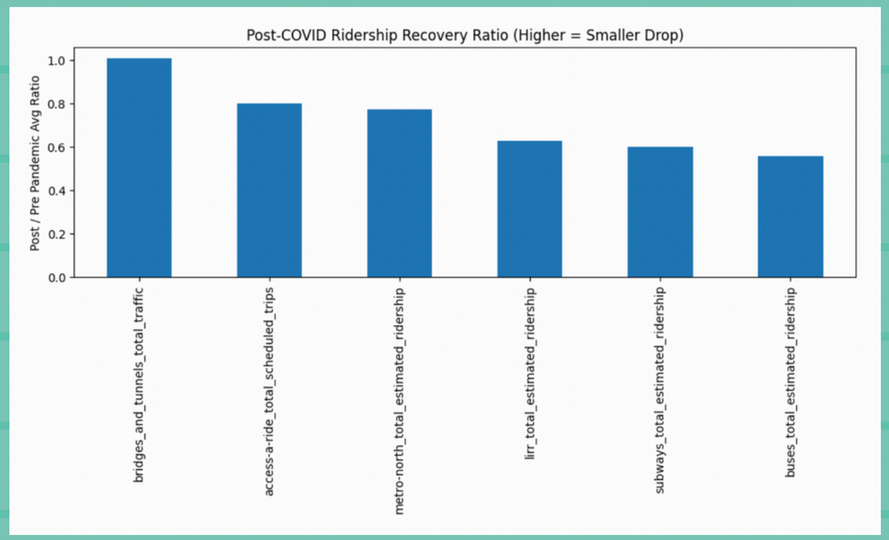


SOME OF OUR VISULIZATIONS

03

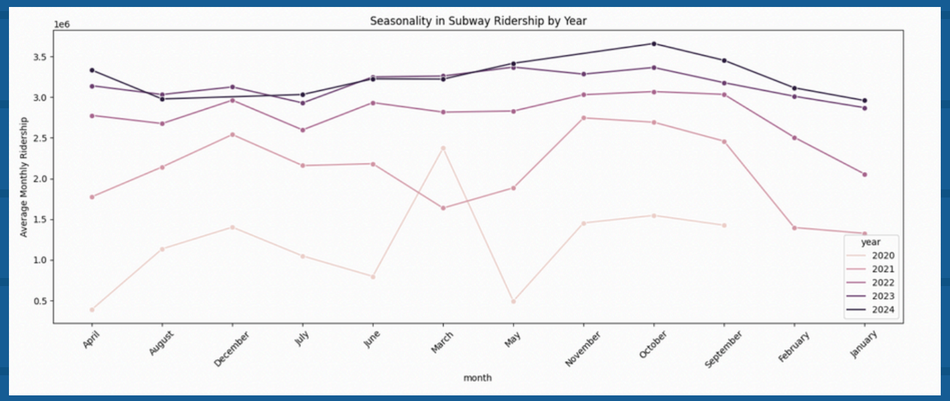


04

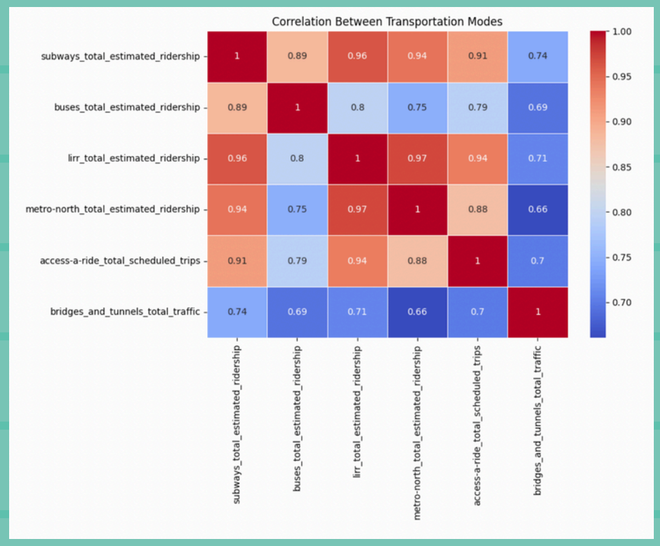


SOME OF OUR VISULIZATIONS

05



06



SUBWAY RIDERSHIP PREDICTION USING GRU MODEL

FORECASTING DAILY MTA PASSENGER COUNTS

DATASET DESCRIPTION

Data Overveiw:

- The dataset contains daily subway rideship counts in New York City
- It covers the period from 2020 to 2023

Data Preprocessing:

- Handled missing values (df.dropna())
- Converted time columns into usable features(pd.to_datetime())
- Applied MinMaxScaler to normalize the data (Used: MinMaxScaler() from sklearn.preprocessing)

Data Splitting: the data was split into: (Used: train_test_split())

- Training set
- Validation set
- Test set

PROBLEM STATEMENT

- Predicting the number of subway riders each day is a surprisingly smart challenge. In this project, we aim to harness the power of deep learning to forecast daily subway ridership using historical data. Accurate predictions can help the city optimize subway services, ease overcrowding, and create a smoother, more efficient commute for millions of New Yorkers. It’s not just about numbers — it’s about making everyday travel smarter and more reliable.

MODEL SELECTION

Gated Recurrent Unit (GRU)
we selected GRU as deep learning model for time series forecasting.It’s type of RNN(recurrent neural network).

EFFICIENCY

Has only two gates (update and reset).This design results in fewer trainable parameters, which speeds up the training process and reduces memory.

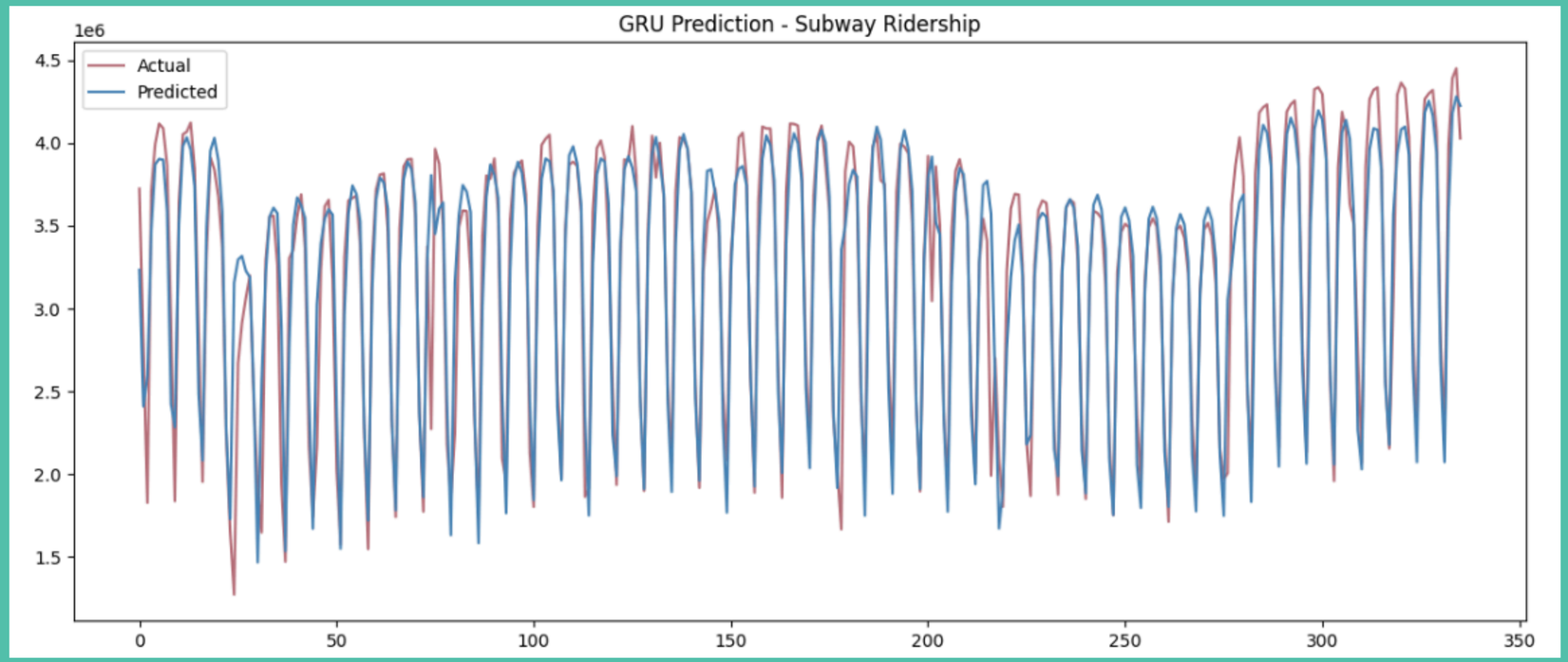
PERFORMANCE

GRU delivers strong performance in time series forecasting tasks.

MODEL ARCHITECTURE
ARCHITECTURE OF GRU

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import GRU, Dense, Dropout, BatchNormalization
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import EarlyStopping, ReduceLROnPlateau

model = Sequential()


model.add(GRU(128, return_sequences=True, input_shape=(seq_length, 1)))
model.add(Dropout(0.2))
model.add(BatchNormalization())

model.add(GRU(64, return_sequences=False))
model.add(Dropout(0.2))
model.add(BatchNormalization())


model.add(Dense(1))


model.compile(optimizer='adam', loss='mse')
```

RESULTS & EVALUATION



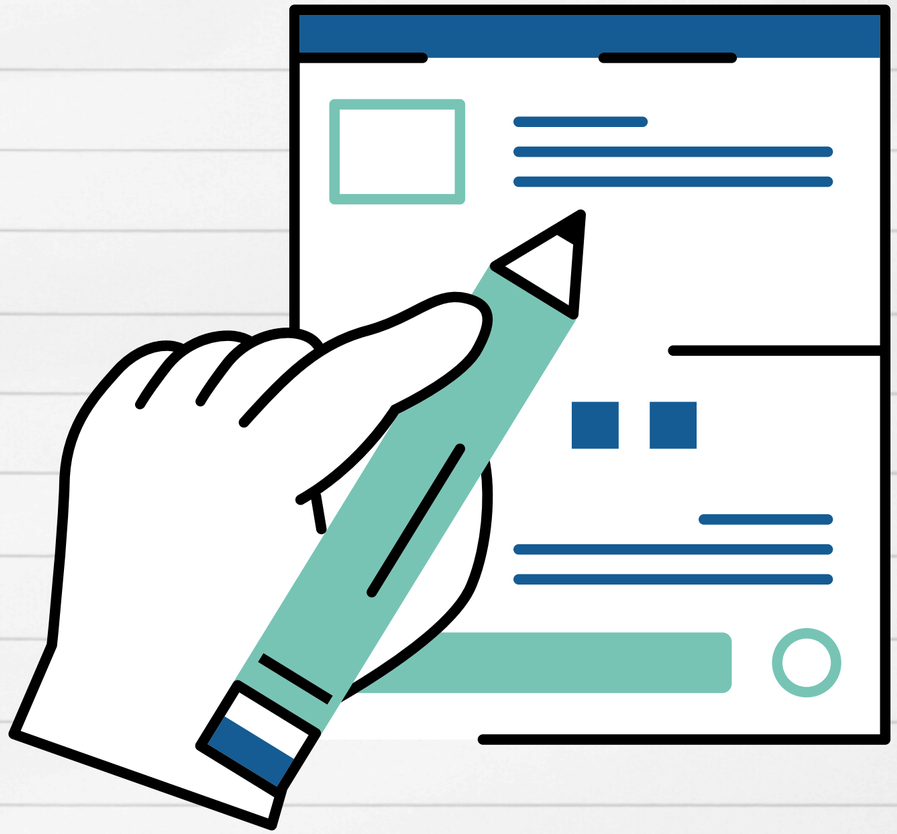
Acutual

PINK LINE

Predicted

BLUE LINE

5-TABLEAU



PROJECT SUMMARY (IN 5 STEPS):

Project Summary (in 5 steps):

1. Imported Data: NYC daily ridership CSV file (Subways, Buses, etc.).

2. Cleaned Data: Selected key columns like date, ridership, and % of pre-pandemic use.

3. Trend Analysis: Created line charts to show how ridership changed over time.

4. Impact Comparison: Used bar charts to compare how each transport mode was affected by COVID.

5. Built Dashboard: Combined everything in an interactive Tableau dashboard with filters.

EXPLORATORY DATA ANALYSIS

Goal: Derive insights by answering key business questions.

INSIGHTS GAINED:

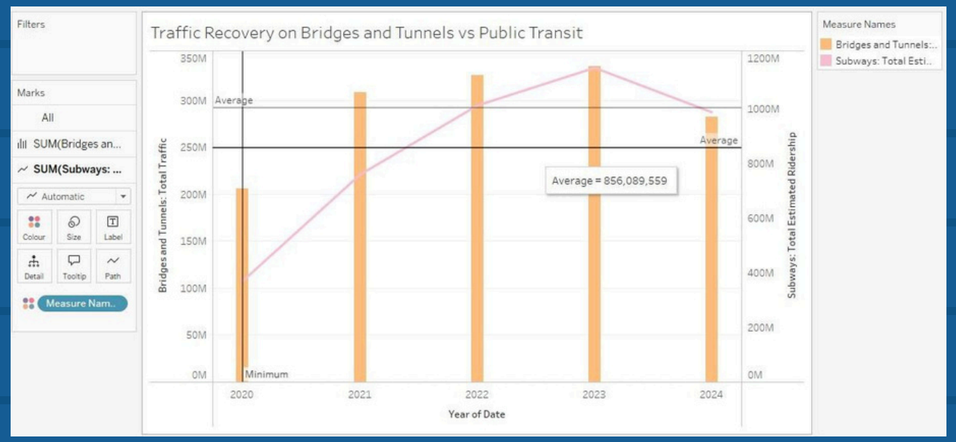
1. Subways were the most affected
Ridership dropped significantly and took longer to recover compared to other modes.
2. Buses recovered faster
Bus usage returned to a higher % of pre-pandemic levels quicker than subways.

INSIGHTS GAINED:

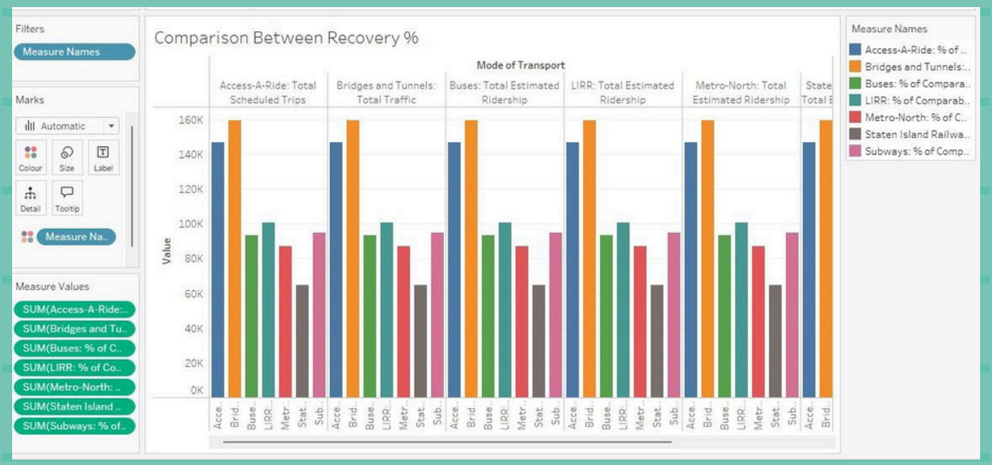
3. LIRR and Metro-North stayed low
These commuter rail systems showed consistently low recovery, indicating continued remote work trends.
4. Pandemic impact peaked around early 2020
Across all modes, the steepest drop in ridership was between March–April 2020.
5.Ridership trends reflect public confidence
Recovery patterns align with reopening phases and public health developments.

SOME OF OUR VISULIZATIONS

01

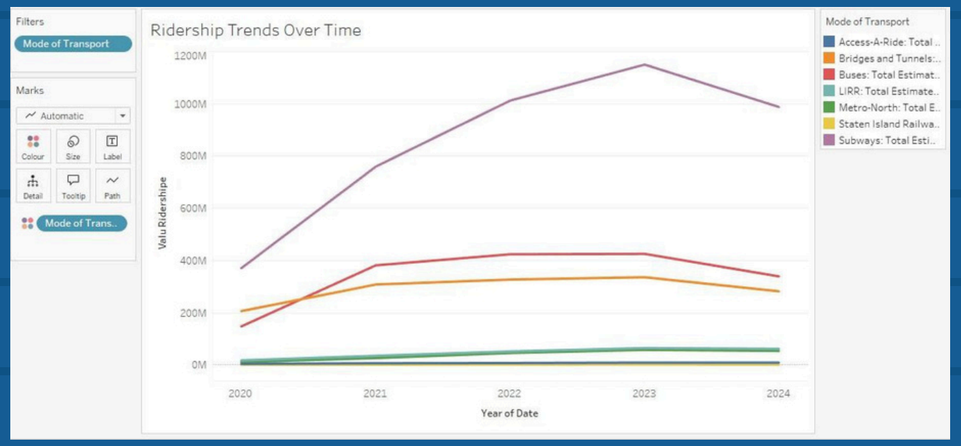


02

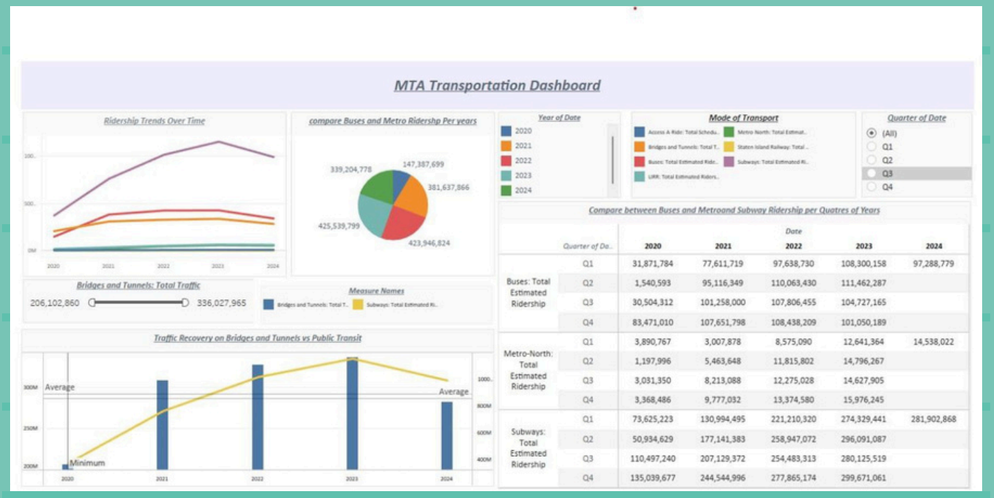


SOME OF OUR VISULIZATIONS

03



04



Quarter of Da..		2020	2021	2022	2023	2024
Buses: Total Estimated Ridership	Q1	31,871,784	77,611,719	97,638,730	108,300,158	97,288,779
	Q2	1,540,593	95,116,349	110,063,430	111,462,287	
	Q3	30,504,312	101,258,000	107,806,455	104,727,165	
	Q4	83,471,010	107,651,798	108,438,209	101,050,189	
Metro-North: Total Estimated Ridership	Q1	3,890,767	3,007,878	8,575,090	12,641,364	14,538,022
	Q2	1,197,996	5,463,648	11,815,802	14,796,267	
	Q3	3,031,350	8,213,088	12,275,028	14,627,905	
	Q4	3,368,486	9,777,032	13,374,580	15,976,245	
Subways: Total Estimated Ridership	Q1	73,625,223	130,994,495	221,210,320	274,329,441	281,902,868
	Q2	50,934,629	177,141,383	258,947,072	296,091,087	
	Q3	110,497,240	207,129,372	254,483,313	280,125,519	
	Q4	135,039,677	244,544,996	277,865,174	299,671,061	

FINAL VISULIZATIONS

MTA Transportation Dashboard

% of Comparable Pre-Pandemic Day

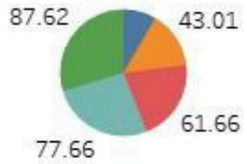
Quarter of Date

(All)

Measure Names

- Subways: % of Compara..
- Access-A-Ride: % of Com..
- Buses: % of Comparable ..
- Bridges and Tunnels: % o..
- LIRR: % of Comparable P..
- Staten Island Railway: % ..
- Metro-North: % of Comp..

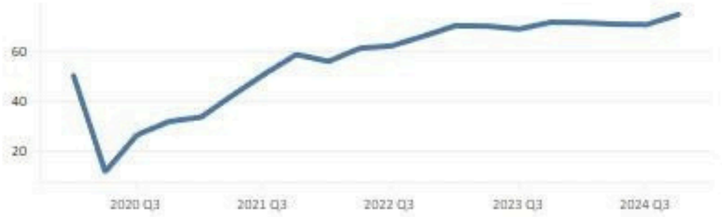
Compare Recovery of Bus vs Subway vs LIRR vs Metro



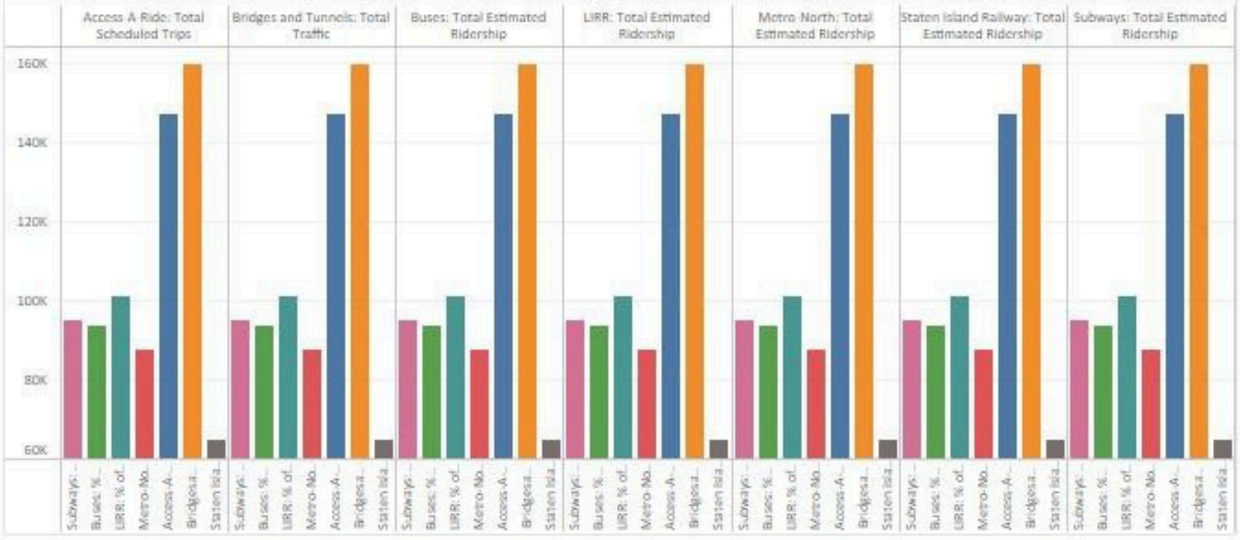
Year of Date

- 2020
- 2021
- 2022
- 2023
- 2024

Subway % Recovery Over Time



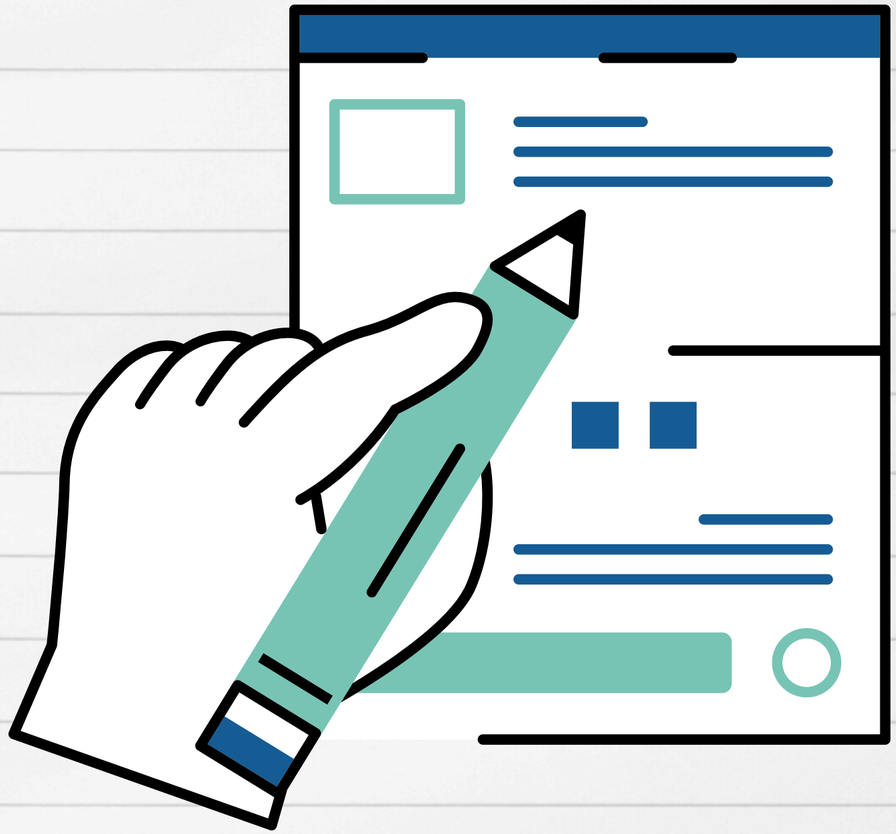
Comparison Between Recovery



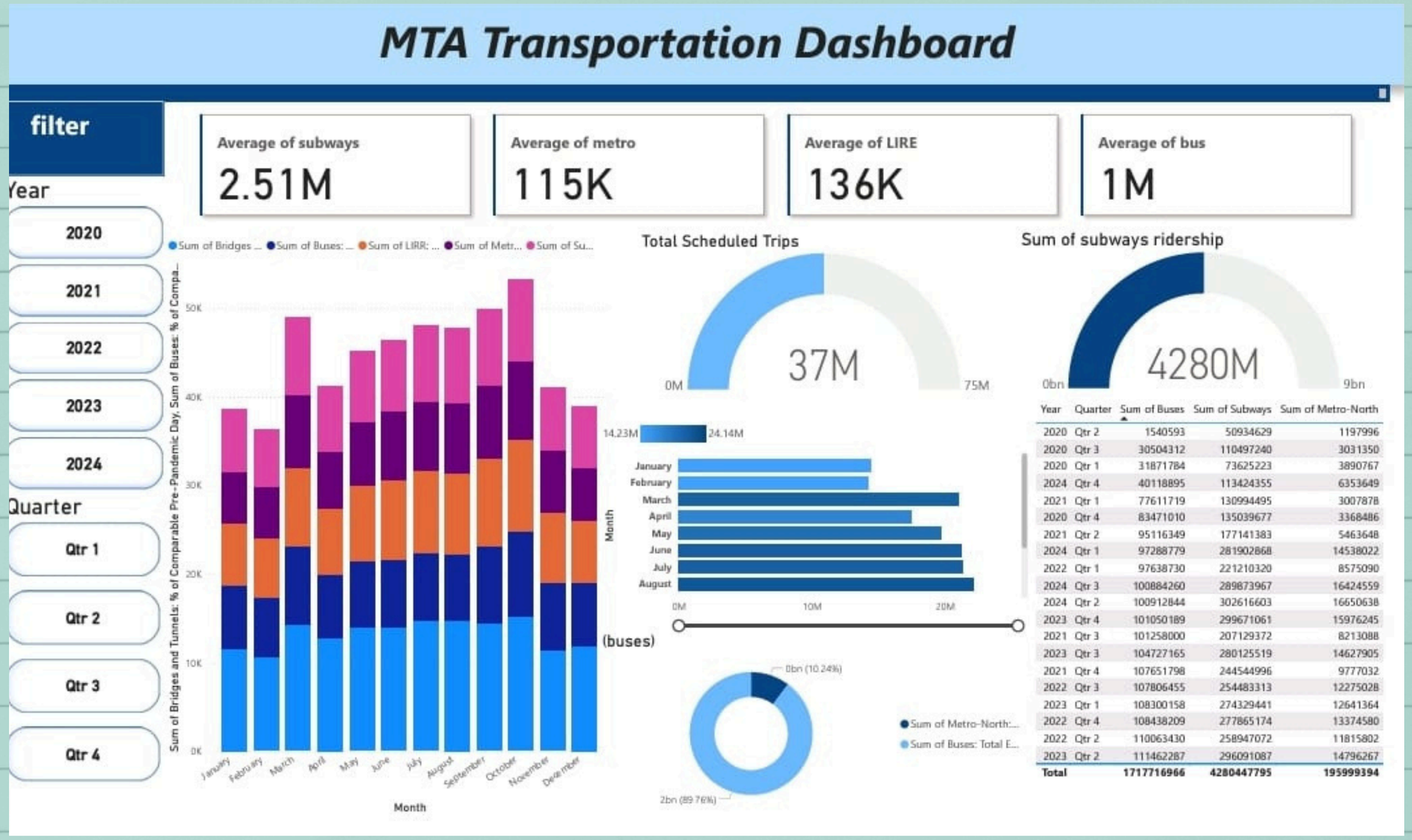
compare recovery per years

	Quarter of Date	Date 2020	2021	2022	2023	2024
Access-A-Ride: % of Comparable Pre-Pandemic Day	Q1	2,125	6,169	6,414	8,621	10,283
	Q2	2,955	6,884	7,469	8,987	10,908
	Q3	5,733	7,014	7,987	9,450	11,545
	Q4	6,301	6,704	7,961	9,555	3,933
Bridges and Tunnels: % of Comparable Pre-...	Q1	2,140	7,459	8,509	9,110	9,203
	Q2	4,767	8,420	9,017	9,162	9,161
	Q3	7,664	8,786	9,089	9,169	9,192
	Q4	7,622	9,037	9,180	9,413	3,198
Buses: % of Comparable Pre-Pandemic Day	Q1	1,656	4,368	5,312	5,996	5,416
	Q2	97	5,118	5,856	5,944	5,342
	Q3	1,568	5,445	5,804	5,659	5,421
	Q4	4,677	5,971	5,996	5,612	2,048
LIRR: % of Comparable Pre-Pandemic Day	Q1	1,438	2,463	4,638	6,359	7,660
	Q2	824	3,568	5,675	7,002	7,936
	Q3	2,533	4,363	5,866	7,146	8,153
	Q4	2,804	5,304	6,326	7,840	2,975
Metro-North: % of Comparable Pre-Pandemic Day	Q1	1,599	1,464	3,977	5,867	6,741
	Q2	534	2,465	5,115	6,387	7,058
	Q3	1,390	3,642	5,401	6,416	7,198
	Q4	1,546	4,494	6,016	7,187	2,651
Staten Island Railway: % of Comparable Pre-...	Q1	1,487	1,892	3,576	4,060	4,396
	Q2	689	2,096	3,489	4,058	4,338
	Q3	1,727	3,172	3,758	4,522	4,956
	Q4	1,837	4,096	3,994	4,726	1,637
Subways: % of Comparable Pre-Pandemic Day	Q1	1,562	3,033	5,028	6,306	6,483
	Q2	1,093	3,856	5,559	6,358	6,431
	Q3	2,449	4,670	5,705	6,321	6,483
	Q4	2,941	5,389	6,067	6,572	2,311

6-POWER BI



FINAL VISULIZATIONS



Year	Quarter	Sum of Buses	Sum of Subways	Sum of Metro-North
2020	Qtr 2	1540593	50934629	1197996
2020	Qtr 3	30504312	110497240	3031350
2020	Qtr 1	31871784	73625223	3890767
2024	Qtr 4	40118895	113424355	6353649
2021	Qtr 1	77611719	130994495	3007878
2020	Qtr 4	83471010	135039677	3368486
2021	Qtr 2	95116349	177141383	5463648
2024	Qtr 1	97288779	281902868	14538022
2022	Qtr 1	97638730	221210320	8575090
2024	Qtr 3	100884260	289873967	16424559
2024	Qtr 2	100912844	302616603	16650638
2023	Qtr 4	101050189	299671061	15976245
2021	Qtr 3	101258000	207129372	8213088
2023	Qtr 3	104727165	280125519	14627905
2021	Qtr 4	107651798	244544996	9777032
2022	Qtr 3	107806455	254483313	12275028
2023	Qtr 1	108300158	274329441	12641364
2022	Qtr 4	108438209	277865174	13374580
2022	Qtr 2	110063430	258947072	11815802
2023	Qtr 2	111462287	296091087	14796267
Total		**1717716966**	**4280447795**	**195999394**

THANK YOU VERY MUCH!