



Metropolitan transportation authority "MTA"

DEPI ' round 2 '





Meet Our Team



Christine Ayman



Rawan Rekapy



Mazen Mohamed



Omar Ahmed



Donia Saeed



Afnan Ibrahim

Dr/ Amal Mahmoud



Agenda

1) Introduction

2) Types of transportation

3) objective

4) Methodology

5) insight



Introduction

Before the pandemic, the MTA was the backbone of NYC, moving millions daily. When COVID-19 hit, ridership plummeted as offices closed and travel halted, pushing the system into financial crisis. During lockdowns, empty trains ran for essential workers while the MTA enforced masks and overnight cleanings. Today, ridership is rebounding but remains below pre-pandemic levels, with shifted patterns—weekends busier than weekdays—as hybrid work endures. Challenges like fare evasion and budget gaps persist, yet the MTA continues adapting, proving its irreplaceable role in the city's survival and revival.





Types of transportation:

SUBWAY

The MTA Subway operates 24/7 across New York City with 472 stations and multiple lines.

Long Island rail road

Long Island Rail Road (LIRR): The LIRR connects Manhattan with Long Island, offering commuter rail services.

Staten Island Railroad

Staten Island Railway (SIR): The SIR provides train service exclusively within Staten Island.

Buses

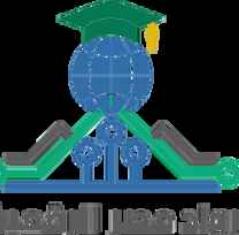
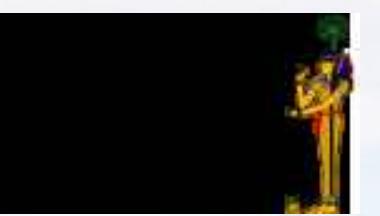
MTA buses run local, express, and Select Bus Service (SBS) routes throughout the five boroughs.

Metro-North Railroad

Metro-North Railroad: Metro-North links New York City with its northern suburbs like Westchester and Connecticut.

Access_A_Ride

Access-A-Ride: This is a paratransit service for individuals with disabilities who cannot use regular buses or subways.



Objective

- Analyze and forecast daily ridership data from the MTA.
- Clean and preprocess raw data to uncover trends and support decision-making.

Answer key business questions (e.g., COVID-19 impact, service-specific changes, peak ridership patterns).
Develop a forecasting model to predict future ridership.
Deliver a Tableau dashboard and comprehensive report.





METHODOLOGY

Dataset:

- About dataset
- Dataset description

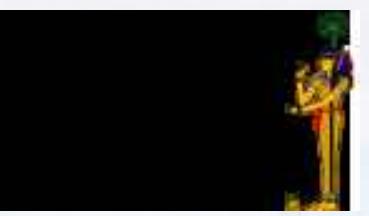
Data preprocessing

- Data Cleaning & Preprocessing
- Data modeling

Data Analysis

- Data visualization



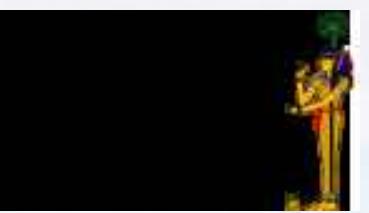


Dataset

ABOUT DATASET:

The dataset is provided by DEPI in one sheet in csv format named by

Name	Size	Packed	Type	Modified	CRC32
..			File folder		
MTA_Daily_Ridership.csv	136,715	58,002	Microsoft Excel Co...	11/4/2024 10:43 AM	047A8932
MTA_DailyRidershipData_Overview.pdf	90,732	85,570	Microsoft Edge PD...	10/30/2024 12:31 PM	B6726C64
MTA_data_dictionary.csv	2,517	516	Microsoft Excel Co...	10/30/2024 12:38 PM	CABC6FE0



Dataset

DATASET DESCRIPTION:

TYPES OF TRANSPORTATION:

Mode_Name	Mode_ID
Subway	1
Bus	2
LIRR	3
Metro-North	4
Access-A-Ride	5
Bridges and Tunnels	6
SIR	7



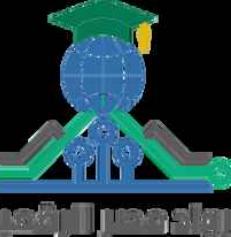
Dataset

Dataset Description:

Recovery Percentage

Fact_ID	Date_ID	Mode_ID	Recovery_Percentage	Riders
1	20200301	1	97.00%	2212965
2	20200301	2	99.00%	984908
3	20200301	3	100.00%	86790
4	20200301	4	59.00%	55825
5	20200301	5	113.00%	19922
6	20200301	6	98.00%	786960
7	20200301	7	52.00%	1636
8	20200302	1	96.00%	5329915
9	20200302	2	99.00%	2209066
10	20200302	3	103.00%	321569
11	20200302	4	66.00%	180701
12	20200302	5	102.00%	30338
13	20200302	6	95.00%	874619
14	20200302	7	107.00%	17140
15	20200303	1	98.00%	5481103
16	20200303	2	99.00%	2228608
17	20200303	3	102.00%	319727
18	20200303	4	69.00%	190648
19	20200303	5	110.00%	32767
20	20200303	6	96.00%	882175
21	20200303	7	109.00%	17453
22	20200304	1	99.00%	5498809
23	20200304	2	97.00%	2177165
24	20200304	3	99.00%	311662
25	20200304	4	70.00%	192689
26	20200304	5	115.00%	34297
27	20200304	6	98.00%	905558

Recovery_Percentage
97.00%
99.00%
100.00%
59.00%
113.00%
98.00%
52.00%
96.00%
99.00%
103.00%
66.00%
102.00%
95.00%
107.00%
98.00%
99.00%
102.00%
69.00%
110.00%
96.00%
109.00%
99.00%
97.00%
99.00%
70.00%
115.00%
98.00%



Data preprocessing

Data cleaning:

Rows: 1706

Column: 15

	Date	Subways: Total Estimated Ridership	Subways: % of Comparable Pre-Pandemic Day	Buses: Total Estimated Ridership	Buses: %
1	2020-03-01	2212965	97	984908	99
2	2020-03-02	5329915	96	2209066	99
3	2020-03-03	5481103	98	2228608	99
4	2020-03-04	5498809	99	2177165	97
5	2020-03-05	5496453	99	2244515	100
6	2020-03-06	5189447	93	2066743	92
7	2020-03-07	2814637	92	1249085	94
8	2020-03-08	2120656	93	957163	96
9	2020-03-09	4973513	89	2124770	95
10	2020-03-10	4857818	87	2111989	94
11	2020-03-11	4697122	84	2112967	94
12	2020-03-12	4149505	75	1938424	86
13	2020-03-13	3484996	63	1715737	77
14	2020-03-14	1670665	54	993287	75
15	2020-03-15	1157711	51	711555	72
16	2020-03-16	2178555	39	1237309	55
17	2020-03-17	1788786	32	1094949	49
18	2020-03-18	1625280	29	1059502	47
19	2020-03-19	1422112	26	933602	42
20	2020-03-20	1309125	24	868602	39
21	2020-03-21	619618	20	411491	31
22	2020-03-22	407732	18	73517	7



Rows: 11913

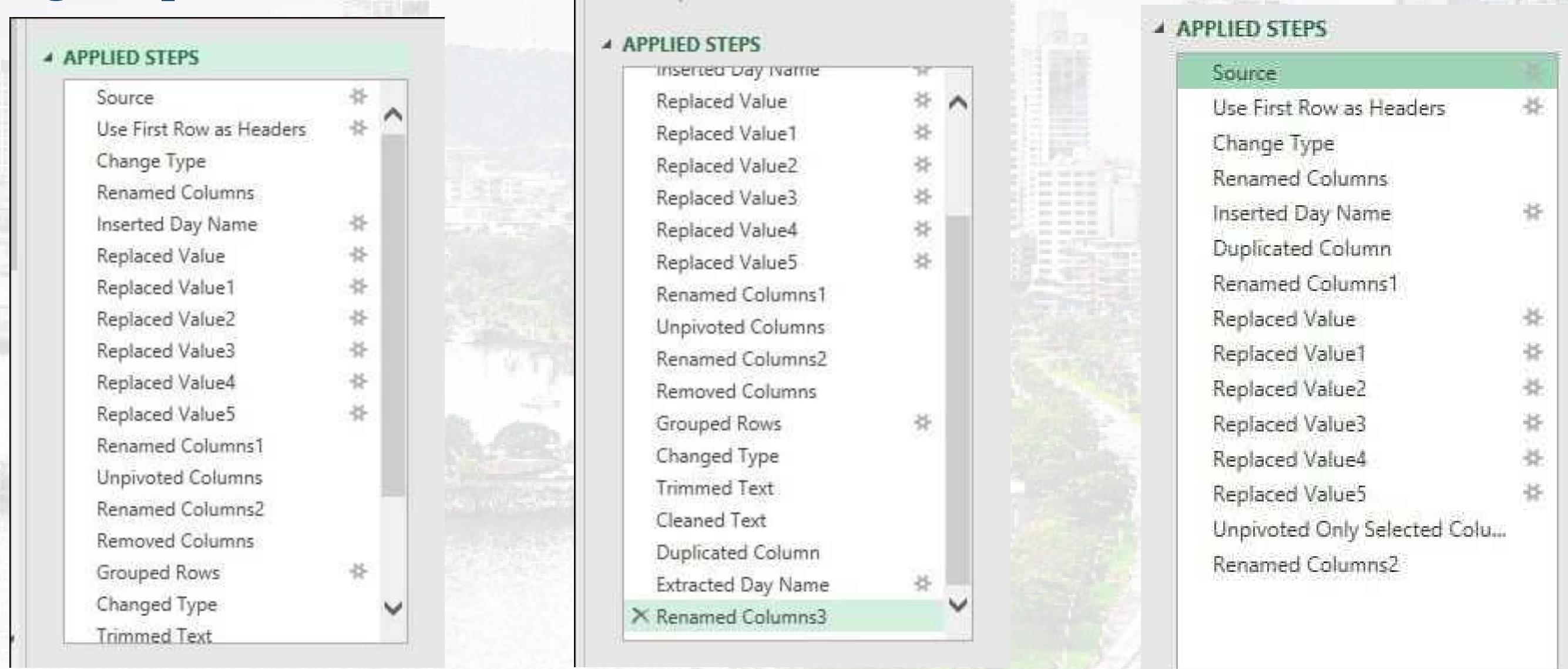
Column: 9

A	B	C	D	E	F	G	H	I	
1	Date	Type of day	Transport mode	Recovery_Percentage	Day	Riders	Year	Month	Quarter
2	3/1/2020	Weekend	Subway	0.97	Sunday	2212965	2020	March	1
3	3/1/2020	Weekend	Bus	0.99	Sunday	984908	2020	March	1
4	3/1/2020	Weekend	LIRR	1	Sunday	86790	2020	March	1
5	3/1/2020	Weekend	Metro-North	0.59	Sunday	55825	2020	March	1
6	3/1/2020	Weekend	Access-A-Ride	1.13	Sunday	19922	2020	March	1
7	3/1/2020	Weekend	Bridges and Tunnels	0.98	Sunday	786960	2020	March	1
8	3/1/2020	Weekend	SIR	0.52	Sunday	1636	2020	March	1
9	3/2/2020	Weekday	Subway	0.96	Monday	5329915	2020	March	1
10	3/2/2020	Weekday	Bus	0.99	Monday	2209066	2020	March	1
11	3/2/2020	Weekday	LIRR	1.03	Monday	321569	2020	March	1
12	3/2/2020	Weekday	Metro-North	0.66	Monday	180701	2020	March	1
13	3/2/2020	Weekday	Access-A-Ride	1.02	Monday	30338	2020	March	1
14	3/2/2020	Weekday	Bridges and Tunnels	0.95	Monday	874619	2020	March	1
15	3/2/2020	Weekday	SIR	1.07	Monday	17140	2020	March	1
16	3/3/2020	Weekday	Subway	0.98	Tuesday	5481103	2020	March	1
17	3/3/2020	Weekday	Bus	0.99	Tuesday	2228608	2020	March	1
18	3/3/2020	Weekday	LIRR	1.02	Tuesday	319727	2020	March	1
19	3/3/2020	Weekday	Metro-North	0.69	Tuesday	190648	2020	March	1
20	3/3/2020	Weekday	Access-A-Ride	1.1	Tuesday	32767	2020	March	1
21	3/3/2020	Weekday	Bridges and Tunnels	0.96	Tuesday	882175	2020	March	1

Data preprocessing

Data cleaning:

- handling missing data
- detecting duplicated values



The figure consists of three side-by-side screenshots of a data preprocessing interface, likely from a software like Microsoft Power BI or similar. Each screenshot shows a list of 'APPLIED STEPS' on the left and a preview of the data on the right.

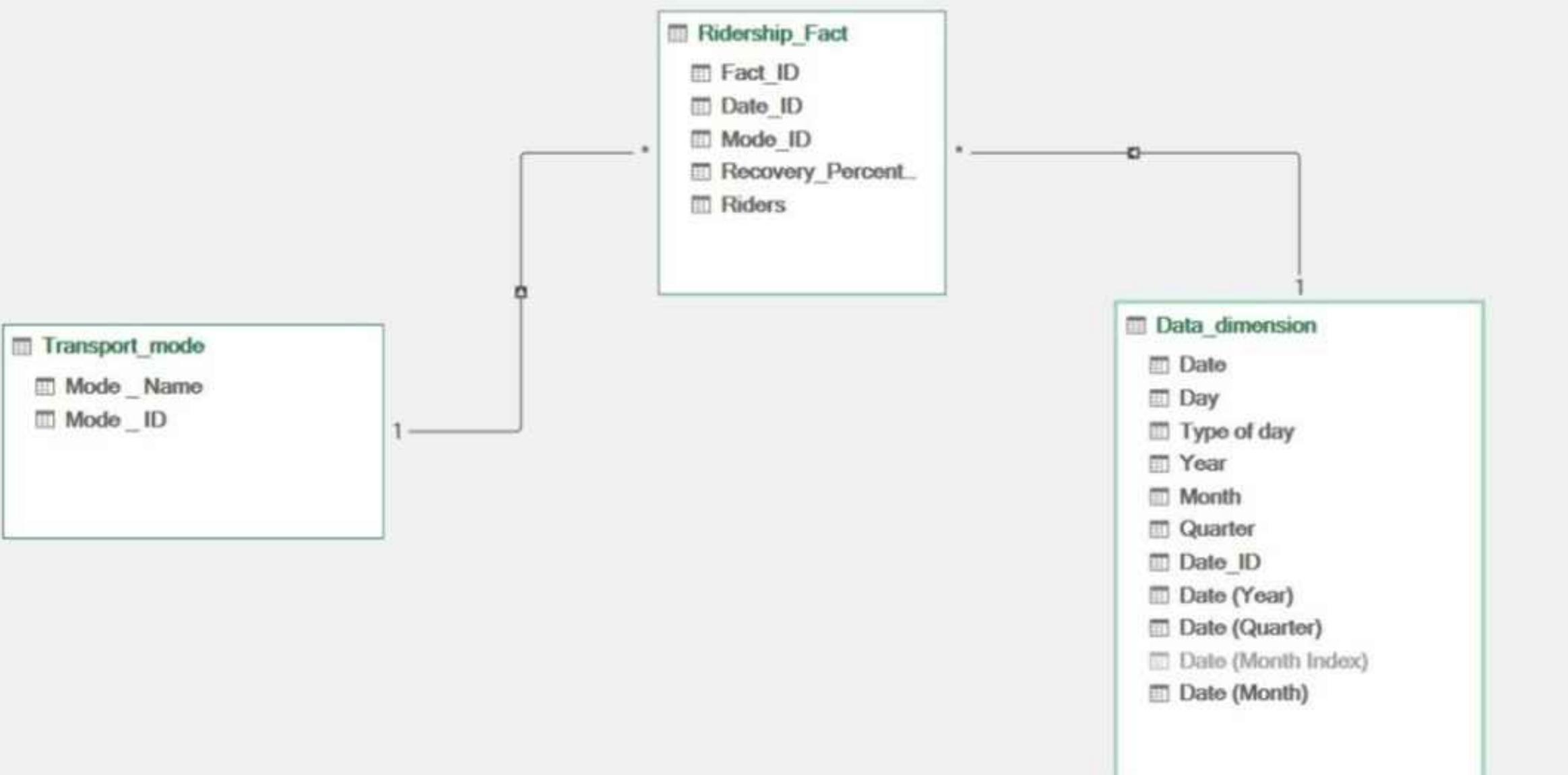
- Left Screenshot (Initial State):** Shows a list of steps including 'Source', 'Use First Row as Headers', 'Change Type', 'Renamed Columns', 'Inserted Day Name', 'Replaced Value', 'Replaced Value1', 'Replaced Value2', 'Replaced Value3', 'Replaced Value4', 'Replaced Value5', 'Renamed Columns1', 'Unpivoted Columns', 'Renamed Columns2', 'Removed Columns', 'Grouped Rows', 'Changed Type', 'Trimmed Text', 'Duplicated Column', 'Extracted Day Name', and 'Trimmed Text'. The last step, 'Trimmed Text', has a red asterisk indicating an error.
- Middle Screenshot (Intermediate State):** Shows the same list of steps, but 'Trimmed Text' is removed. A new step, 'Renamed Columns3', is added at the bottom.
- Right Screenshot (Final State):** Shows the final list of steps: 'Source', 'Use First Row as Headers', 'Change Type', 'Renamed Columns', 'Inserted Day Name', 'Duplicated Column', 'Renamed Columns1', 'Replaced Value', 'Replaced Value1', 'Replaced Value2', 'Replaced Value3', 'Replaced Value4', 'Replaced Value5', 'Renamed Columns2', 'Unpivoted Only Selected Colu...', and 'Renamed Columns3'. The 'Duplicated Column' step has a red asterisk.

final table



Data preprocessing

Data modeling:



Data analysis

Exploratory data analysis(4 W's):

- **What happened? (Descriptive Analytics)**
- **Why did it happen? (Diagnostic Analytics)**
- **What will happen? (Predictive Analytics)**
- **What should we do? (Prescriptive Analytics)**





Data analysis

Exploratory data analysis

1) what happened?

How did COVID_19 pandemic affect public transportation?

2) why did it happen?

why did public transport usage drop significantly in 2020?

3) what will happen?

what will happen to public transport ridership in the coming period?

4) what should we do?

what should we do to help the public transport sector recover?





Data analysis



Exploratory data analysis

- 1) How the problem happened?**
- 2) How long it may last?**
- 3) How to solve the problem?**





Data analysis

Exploratory data analysis

1) How the problem happened?

How did the ridership decline start and worsen?

2) How to solve the problem?

How long will the effects of COVID_19 on public transport last?

3) How long it may last?

How did COVID_19 pandemic affect public transportation?



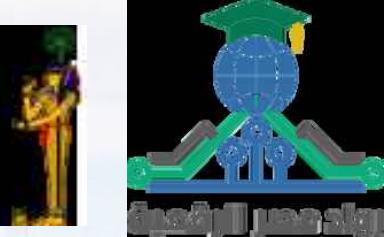
Data analysis



Data visualization:

Excel:

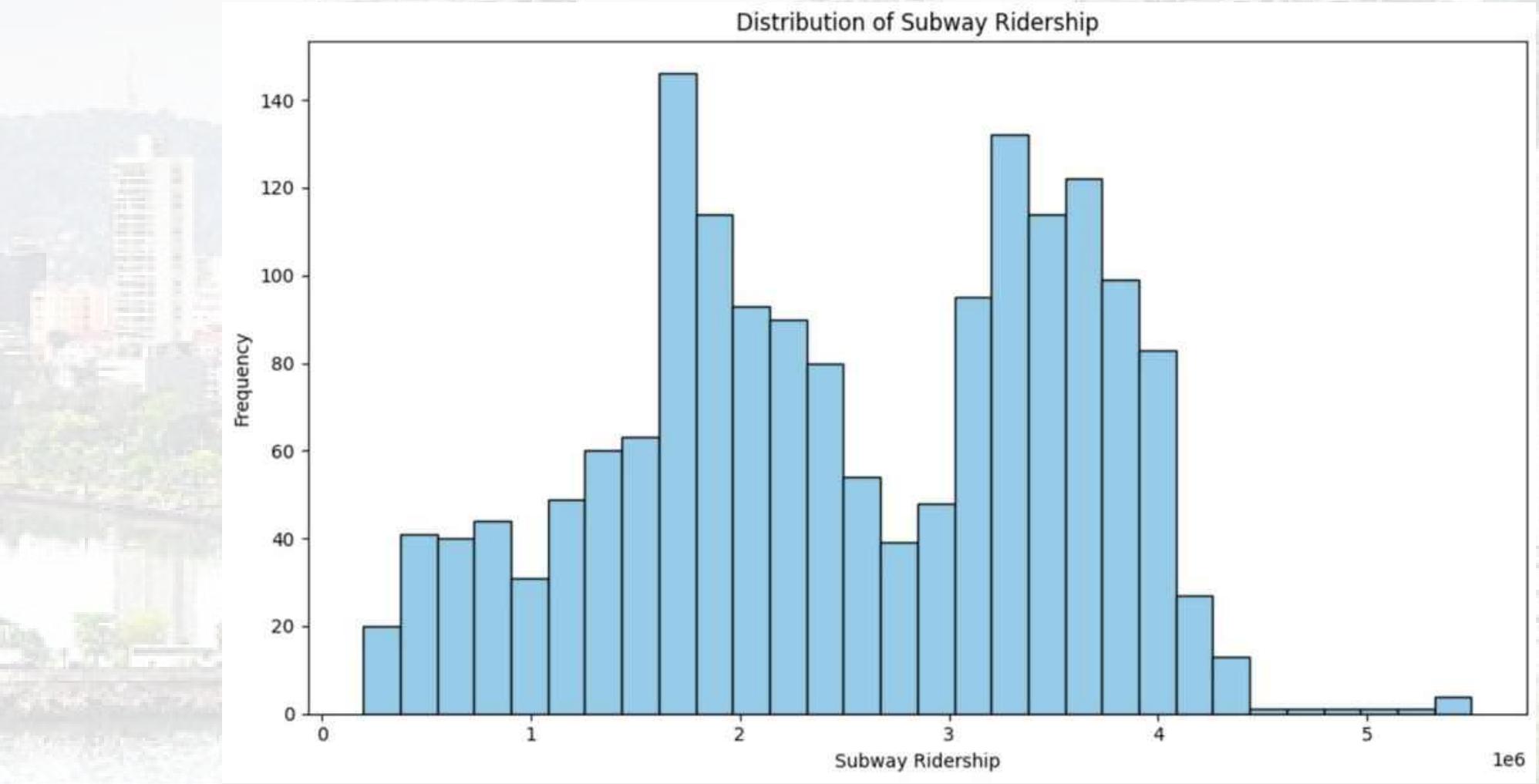
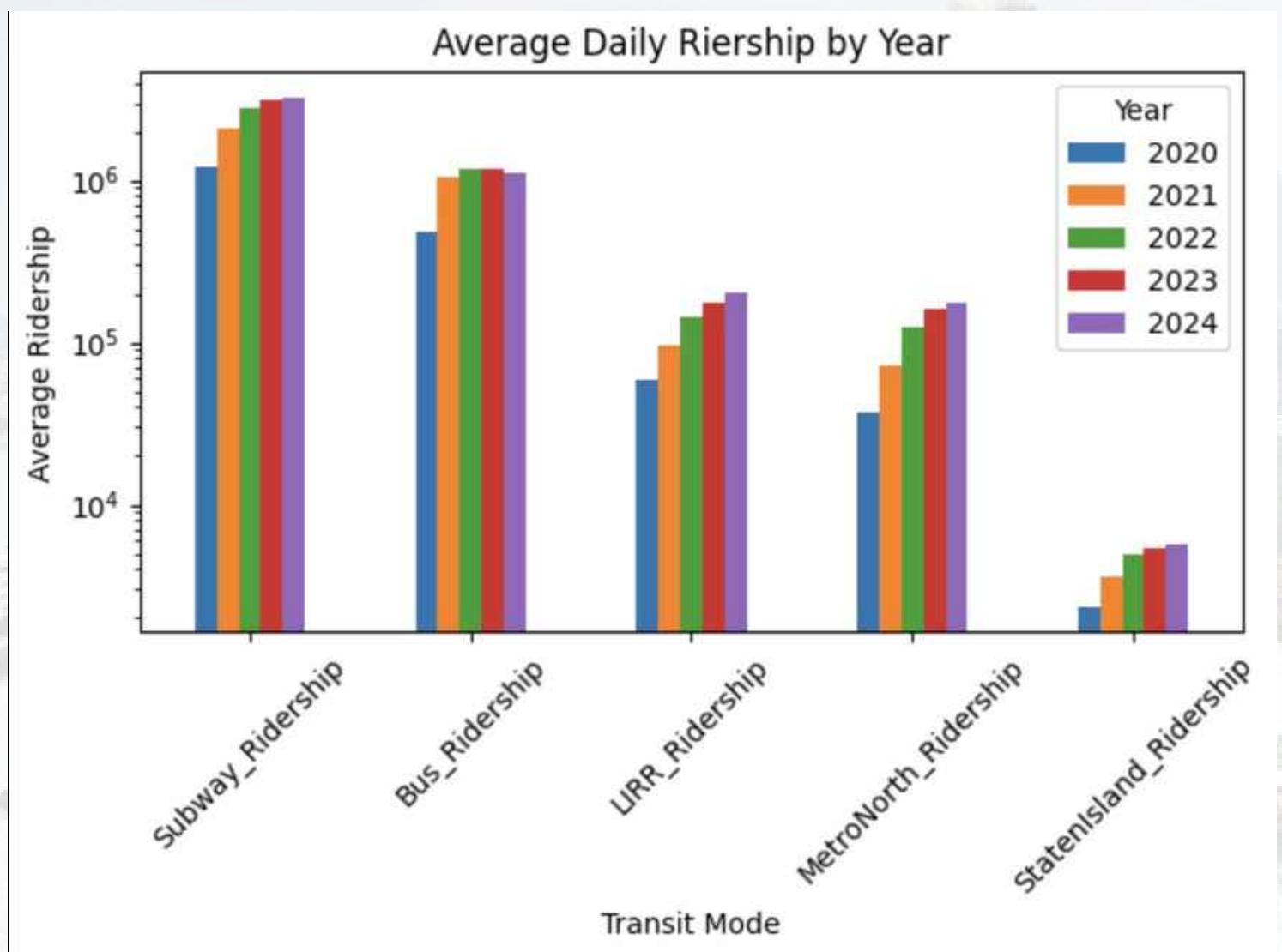




Data analysis

Data visualization:

Python:





Data analysis



Data visualization:

Tableau:





Insights

1. What are the most used transportation modes?

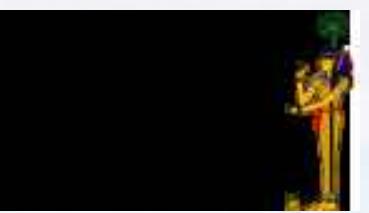
The subway is the most used mode, accounting for 53.97% of total riders, followed by buses at 21.66%.

The least used: SIR (Staten Island Railway) at 0.09%.

2. How did the COVID-19 pandemic affect ridership?

Sharp drop in 2020 (758 million riders) compared to 2021 (1.5 billion), followed by a gradual recovery reaching 2.04 billion in 2023.

In 2024, there is a slight decline (1.73 billion), possibly due to shifts in work patterns (e.g., hybrid work).



Thank You

FOR YOUR ATTENTION

