CHICAGO CTA Ridership Analysis

Programming for Analytics
Final Project

Annabella Liang, Minh Vo

01

BUSINESS CASE

02

METHODOLOGY

03

VISUALIZATION

04

CONCLUSIONS

01

BUSINESS CASE

EXECUTIVE SUMMARY

- Chicago Transportation Authority (CTA) would like to evaluate how different weather conditions influence different bus routes and L-lines so that CTA can provide better service and cost-efficiency
- Optimizing frequencies, precise prediction of arrival time, preparing for emergencies, allocating staff, and arranging connecting transportation around stations based on the weather forecast.
- This project will focus on analyzing the CTA ridership between 2018-2022.

BUSINESS CASE

Business objectives

- Enable better and more stable services
- Improve cost efficiency
- Gain long-term development insights

Research focuses

- How weather conditions impact CTA ridership, and at which specific bus routes or L-lines.
- How pandemic shifted ridership, on volume and location

02

METHODOLOGY

Datasets

L-Line data

station_id	stationname	date	daytype	rides
40020	Harlem-Lake	6/30/2022	W	1808
40030	Pulaski-Lake	6/30/2022	W	690
40040	Quincy/Wells	6/30/2022	W	3914
40050	Davis	6/30/2022	W	1506
40060	Belmont-O'Hare	6/30/2022	W	2793
40020	Harlem-Lake	6/29/2022	W	1616
40030	Pulaski-Lake	6/29/2022	W	622
40040	Quincy/Wells	6/29/2022	W	3978
40050	Davis	6/29/2022	W	1540
40060	Belmont-O'Hare	6/29/2022	W	2711

Bus data

route	date	daytype	rides
10	7/31/2022	U	409
1001	7/31/2022	U	93
103	7/31/2022	U	408
106	7/31/2022	U	255
11	7/31/2022	U	417
111	7/31/2022	U	656
111A	7/31/2022	U	93
112	7/31/2022	U	366
115	7/31/2022	U	820
119	7/31/2022	U	1185

Daytypes are as follows:

- W=Weekday
- A=Saturday
- U=Sunday/Holiday

Sources:
https://data.cityofchicago.org/Transportation/CTA-Ridership-Bus-Routes-Daily-Totals-by-Route/jyb9-n7fm
https://data.cityofchicago.org/Transportation/CTA-Ridership-L-Station-Entries-Daily-Totals/5neh-572f

Datasets

Weather data

datetime	tempmax	tempmin	temp	humidity	precip	precipprob	precipcover	preciptype	snow	snowdepth	windgust	windspeed	winddir	sealevelpressure	cloudcover	visibility
7/31/2018	79	68.3	73.5	70.4	0.008	100	4.17	rain	0	0		11.6	38.5	1015.4	80.8	9.9
8/1/2018	83.8	63.6	74.2	66.6	0.029	100	8.33	rain	0	0	29.9	15.9	234.9	1013.8	57.3	9
8/2/2018	85	70.3	77.6	63.6	0.009	100	8.33	rain	0	0	21.9	13.5	222.9	1013.5	70.1	9.9
8/3/2018	86.7	67	76.8	63.5	0	0	0		0	0	18.3	12.1	71.8	1016.1	44.7	9.2
8/4/2018	96.3	72.4	84.4	55.9	0	0	0		0	0	25.4	18.3	206.5	1016.1	54.6	9.9
8/5/2018	94.4	77.5	84.5	56.4	0.047	100	12.5	rain	0	0	30.5	20.3	217.3	1016.7	75.2	9.9
8/6/2018	82	72.6	76.9	77.3	0.153	100	25	rain	0	0	34.4	11.4	53.8	1015.6	91.6	9.7
8/7/2018	84	71.2	75	81.4	1.02	100	25	rain	0	0		9.5	358.2	1013.6	82.5	8.8
8/8/2018	81.7	69.6	75.3	74.3	0.012	100	8.33	rain	0	0		9.6	39.5	1012.6	67.8	9.2
8/9/2018	90.7	68.3	78.1	69.5	0.196	100	12.5	rain	0	0	31.1	15.5	259.8	1012.6	51.7	9.9
8/10/2018	83.2	71.8	77.1	75.3	0	0	0		0	0	21.2	16.1	30.9	1013.1	51.6	9.5

- Temperature (temp): F
- Precipitation (precip): mm
- Snow Depth (snowdepth): mm
- Wind Speed (windspeed): mph
- Visibility: miles

Source: https://www.visualcrossing.com/weather/weather-data-services

METHODOLOGY

Preparation

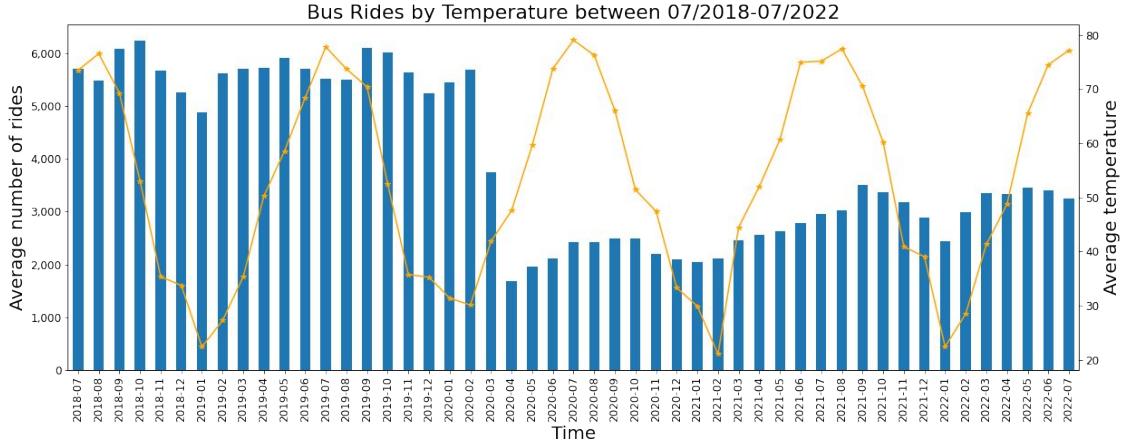
- Make assumptions on metrics like weather, events, pandemic, etc., that would impact CTA ridership.
- Clean up datasets of CTA ridership, bus stops and railway stations, and weather based on assumptions.

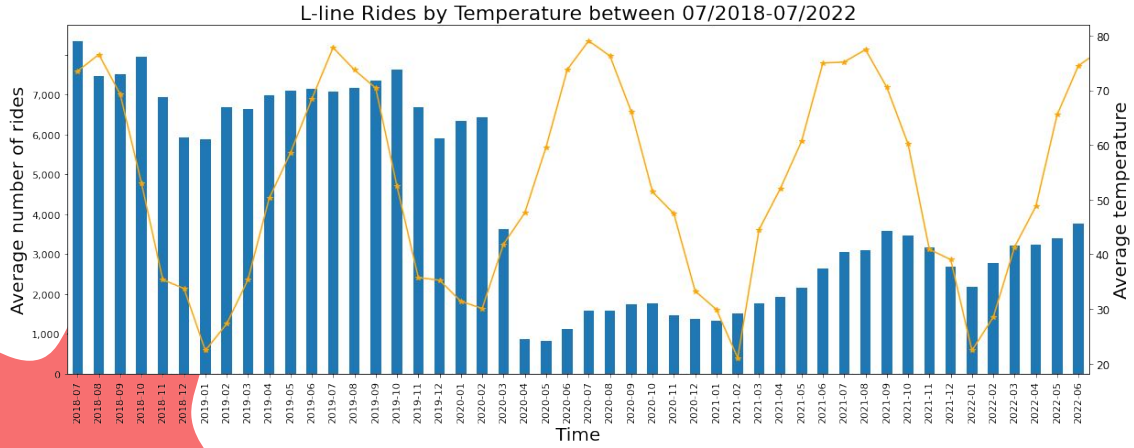
Solutions

- Identify trends of weather conditions and ridership over time by visualizing both datasets on graphs.
- Categorize ridership data into pre-, in-, and post-pandemic (within the past 5 years).
- Summarize traffic distribution along lines on graphs
- Merge datasets to identify correlations between conditions and ridership

03

DATA ANALYSIS & VISUALIZATION





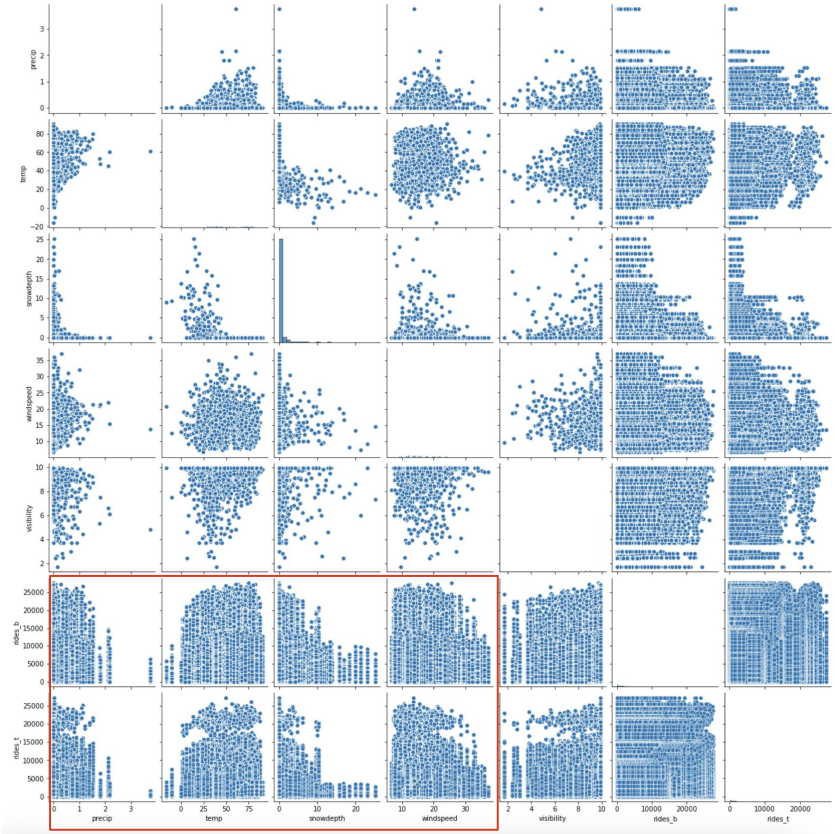
Ridership over time by temperature

- Temperature directly affects ridership: number of riders decreases in January yearly as the temperature drops.
- Both average bus rides and L-line rides started to decrease significantly since 04-2020 due to pandemic.
- Ridership started to increase slowly since 05-2021

Certain weather conditions have impacts on ridership

Use pair plot to visualize possible correlation between varied weather conditions and riderships

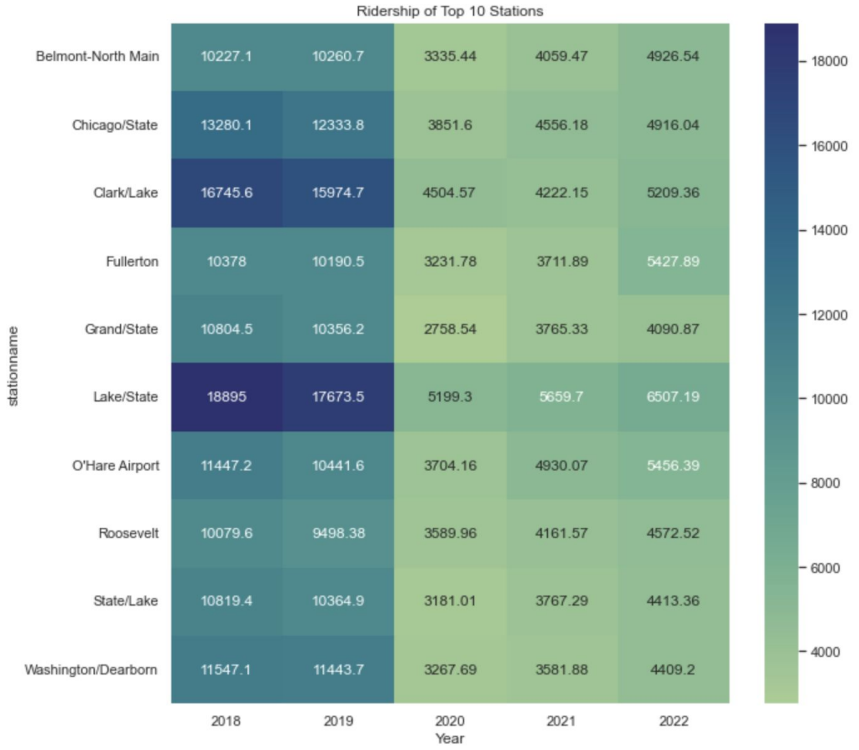
- When precip > 2, riderships of both buses and trains significantly decreased
- When avg. temp dropped below 0, riderships of both buses and trains significantly decreased
- When snow depth > 10, riderships of both buses and trains significantly decreased, trains' riderships were impacted more
- When wind speed > 30, riderships of both buses and trains dropped



Riderships dropped drastically during COVID and haven't risen back to pre-COVID level

Select the top 10 stations with the highest riderships and visualize them on a heatmap to analyze how COVID impacts major public transportation in Chicago

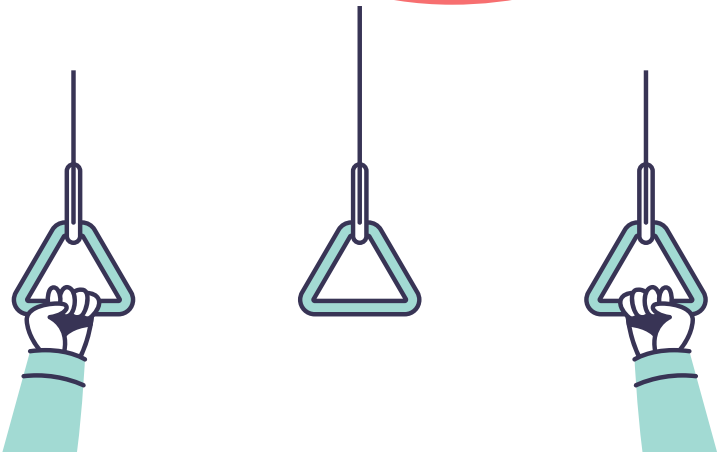
- Riderships dropped by around ⅔ compared to pre-COVID during pandemic
- Riderships slightly picked up in 2022 when all COVID restrictions were lifted but far from pre-COVID level
- COVID has significantly changed how people commute in Chicago city and the impact could be permanent



CONCLUSION

- Certain weather conditions impact riderships of both bus and train: temperature, precipitation, snow depth and wind speed. Severe weather conditions discourage riders to travel and use public transportation. CTA can reduce frequency based on weather forecast to cut down operation cost

- COVID has a longer impact on how people in Chicago use CTA service. Ridership dropped 60-70% across all stations and failed to rise back to pre-COVID level even when all pandemic restrictions were lifted



THANKS!