



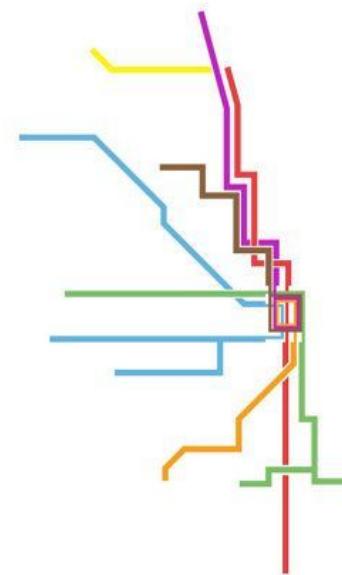
# Impact of Chicago Home Games on CTA 'L' Ridership

Dream Team

Gabriel Bang, Jacquelyn Cleary, Megan Dodson, Konsta Jokipii, Nick Neubauer

T/Th Cohort

February 6, 2021



CHICAGO

# Project Assumption

We hypothesize that Chicago home games will have a causal impact to increases in CTA 'L' ridership.

## Methodology

This project studies the impact of Blackhawks/Bulls, Cubs, Bears, and White Sox home games to subway ridership. To minimize other external factors we consolidate data testing to each team's respective station.

## Project Overview

- Data Sources & Clean Up
- Baseline Data Validation
- Game Day vs. Non-Game Day
- Cubs Ridership
- Same Sport Analysis: Cubs vs. White Sox
- Da Bears
- United Center CTA Analysis: Blackhawks vs. Bulls
- ANOVA Test to Compare Means
- Lessons Learned



# Data Sources & Clean up

## Data Sources

- Pulled CTA 'L' data from Chicago Data Portal
- Each team member pulled their team's home schedules

## Clean up process

- Merged all home schedules into one masterfile
- Converted dates to date.time
- Deleted unwanted rows - years
- Used a merge left to combine both data sources into one file
- Filled NAs with False

```
In [11]: # Convert string date to datetime
teams_raw.rename(columns={"Date":"date"},inplace=True)
teams_raw["date"] = pd.to_datetime(teams_raw["date"],format="%m/%d/%Y")
cta_raw["date"] = pd.to_datetime(cta_raw["date"],format="%m/%d/%Y")
cta_raw.shape
```

```
Out[11]: (1028040, 5)
```

```
In [15]: # delete unwanted rows in each dataset
merge_df = cta_raw.merge(teams_raw, on="date", how="left")
merge_df
```

```
In [23]: merge_df.fillna(value=False, inplace=True)
merge_df = merge_df[(merge_df["date"] > "2016-12-31")]
merge_df.sort_values(["date"])["date"].nunique()
```

```
Out[23]: 1369
```

```
In [24]: merge_df.to_csv("clean_df.csv")
```

# Baseline Data Visualization

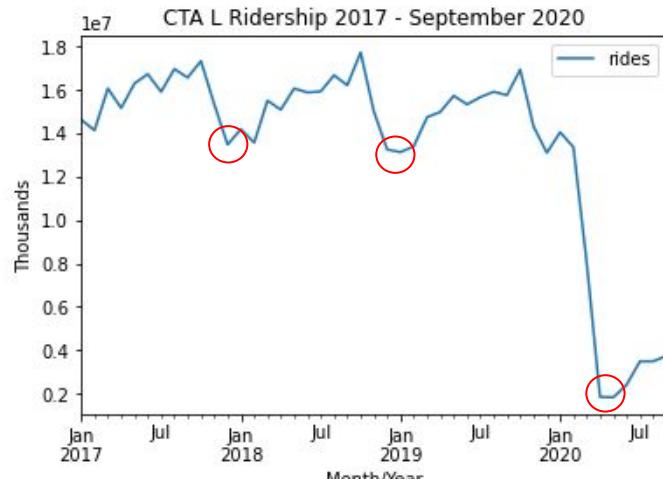
## Visual I:

1. The line graph displays change in Total CTA L Ridership (thousands) between Jan-2017 to Sept-2020.
2. Consistencies:
  - a. Peaks tend to occur in warmer months
  - b. Pits tend to occur in cooler months
3. Inconsistencies:
  - a. In 2020, heading into historically warmer months, ridership experienced an abnormal sharp decrease.  
**Any thoughts on the cause of this occurrence?**

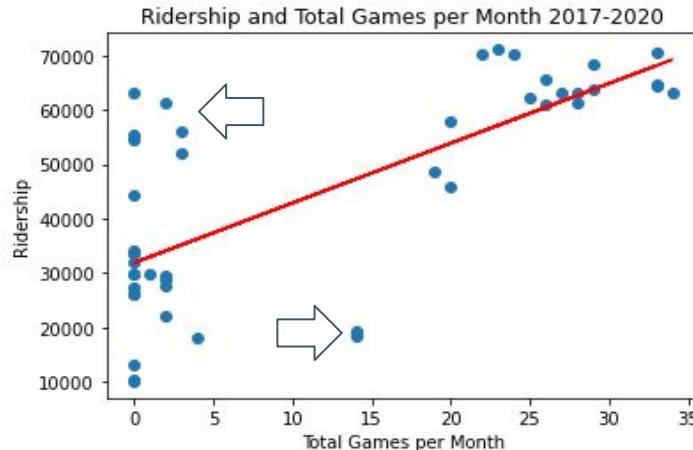
## Visual II:

1. The regression model displays the relationship between Ridership and Total Games per Month Jan-2017 to Sept-2020 ( $PCC = 0.72$ )
2. Consistencies:
  - a. Most of points in the plot show correlation between the two axes.
3. Inconsistencies:
  - a. A number of points fall strikingly above or below the regression line. **Any thoughts on the cause of these occurrences?**

Visual I.



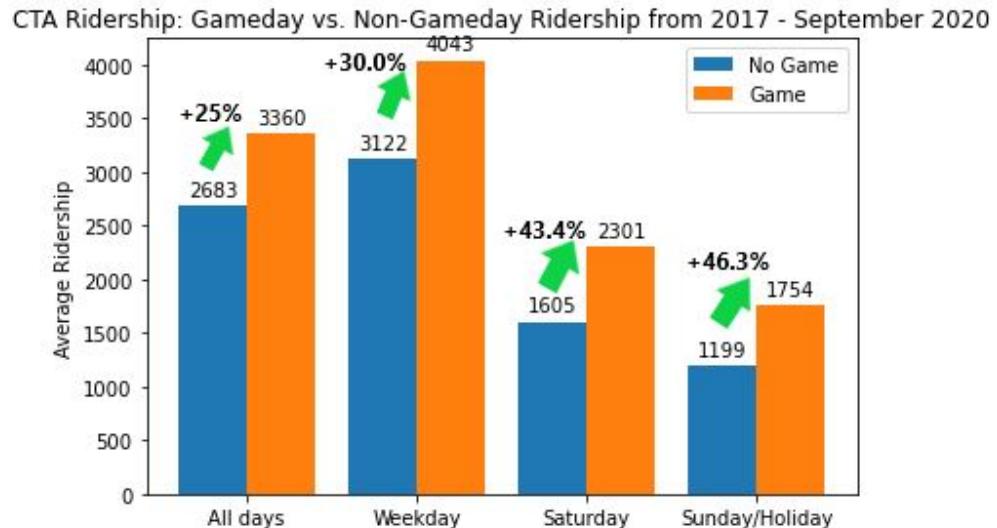
Visual II.



# Game Day vs Non-Game Day

## 2017 - September 2020

- 25% increase in ridership across all teams
  - 30% increase in ridership on weekdays
  - 43.3% increase in ridership on Saturdays
  - 46.3% increase in ridership on Sundays/Holidays
- Biggest increase on Sundays/Holiday
- Smallest increase on Weekdays
- Ridership is positively impacted by gamedays

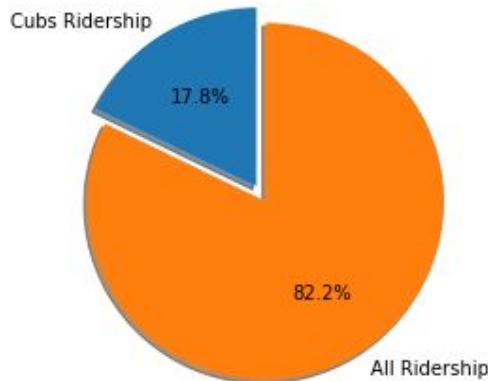




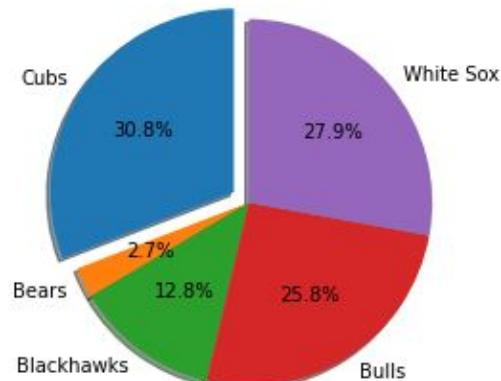
# Cubs Ridership

- Cubs game day ridership (stations near stadium only): 17.8% of total ridership over the three years
- Any game day ridership across all five teams (all stations)
  - Cubs: 30.8% - largest of all 5 teams
  - Largest venue/attendance

CTA Ridership vs. Cubs Ridership: 2017 - September 2020

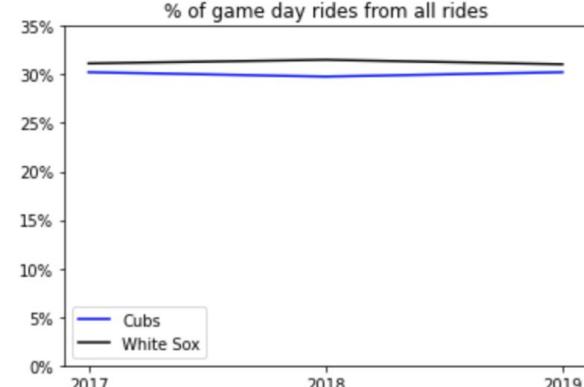
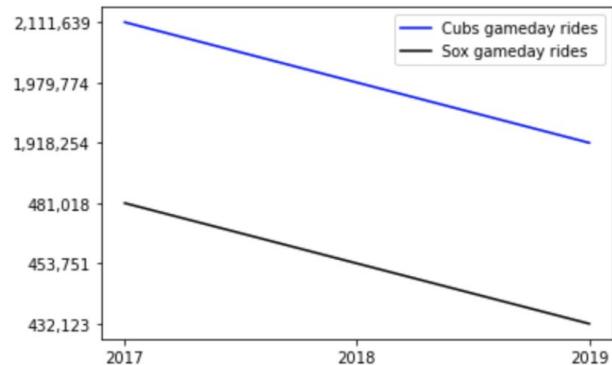
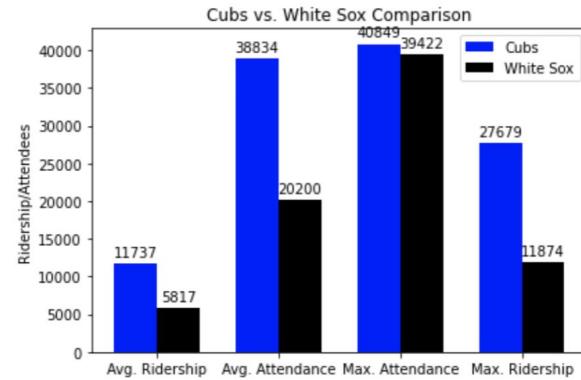


CTA Ridership by Team on Gamedays from 2017 - September 2020



# Same Sport Analysis: Cubs vs. White Sox

- White Sox vs. Cubs games draw the highest attendance AND ridership
- While attendance for these games is similar, ridership at the Cubs station is 133% higher
- When adjusting avg. ridership to avg. attendance, both team's fans 'L' train usage is 28.8%
- Cubs attendance dropped -3.3% while game day ridership contracted -9.2%
- White Sox attendance didn't drop however their ridership also contracted -10.2%
- Increase in popularity of alternate modes of transportation could be affecting game day ridership
- Game days contribute almost 1/3rd of all traffic at Cubs and White Sox stations
- Game days make up approx. 22% of all days in a year so the effect is significant

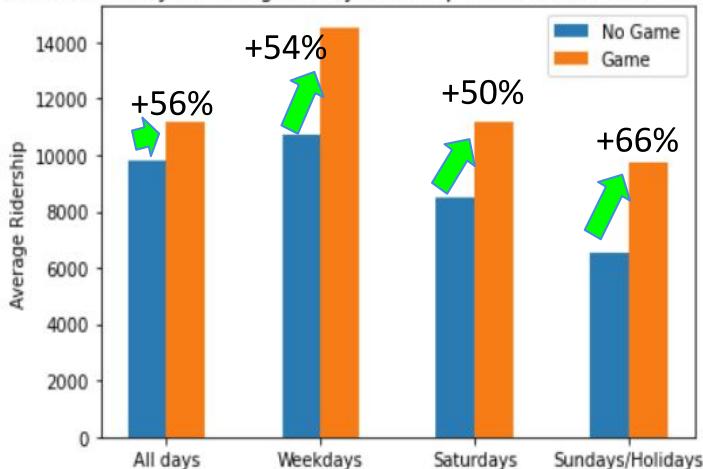




## DA BEARS

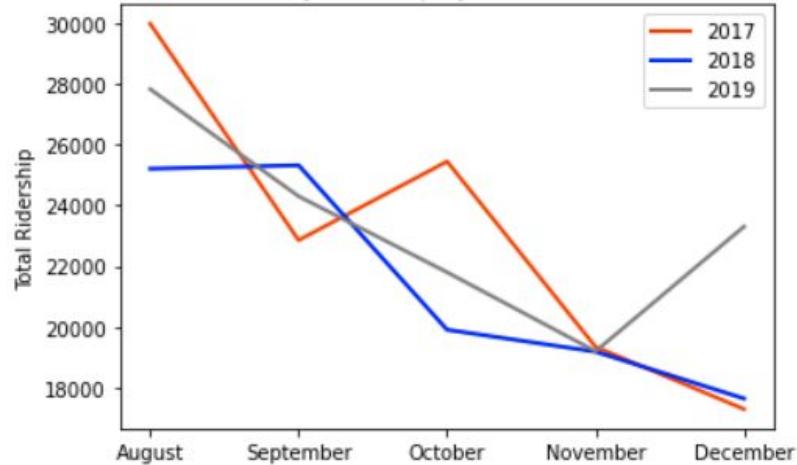
- Bears home games have a positive impact on ridership at the Roosevelt station
- On Bears home games, there was at least a 50% increase on average ridership - no matter the day of the week

Bears Gameday vs. Non-gameday Ridership at Roosevelt Station 2017-2019



- Ridership decreases as the season progresses
- Weekday games (Monday and Thursday) represent spikes in later months and account for the difference in ridership season to season

Bears Gameday Ridership by Month Over the Seasons



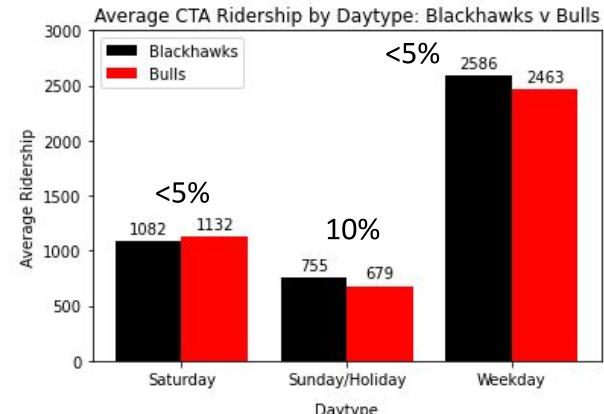
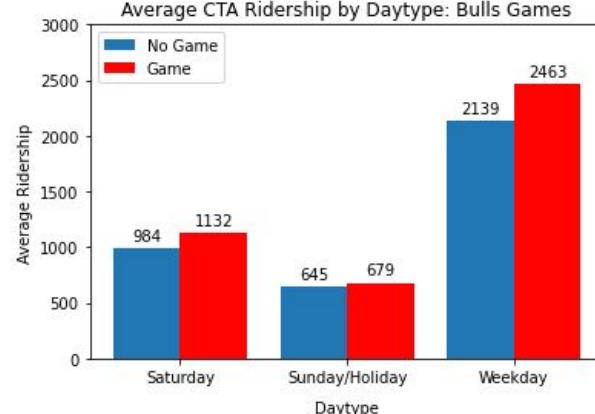
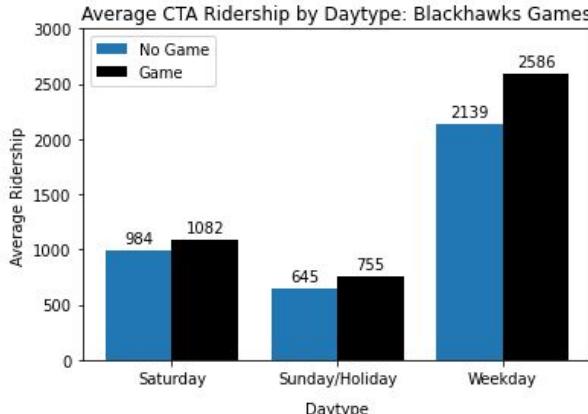
- Bears fans and their loyalty continue to be difficult to predict

	2017	2018	2019
Record	5 - 11	12 - 4	8 - 8
Total home game ridership	114,892	107,232	116,395
# of home games	10	10	11

# United Center CTA Analysis: Blackhawks vs Bulls

Comparison of average ridership on game days vs non-game days from 2017-2019 seasons

- CTA "L" Stations included in analysis @ Ashland-Lake & Medical Center stations by the United Center
  - *Blackhawks Game Days (% difference from baseline)*
    - Weekday = 20.9%
    - Sunday/Holidays = 17.1%
    - Saturday = 10.0%
  - *Bulls Game Days (% difference from baseline)*
    - Weekday = 15.2%
    - Saturday increase = 15.0%
    - Sunday/Holiday = 5.3%



# ANOVA Test to Compare Means

Team: Blackhawks



Saturday:

F\_onewayResult(pvalue=0.6701349562162419)

Sunday:

F\_onewayResult(pvalue=0.03965867712688939)

Weekday:

F\_onewayResult(pvalue=6.549434308072871e-09)

Team: Bulls



Saturday:

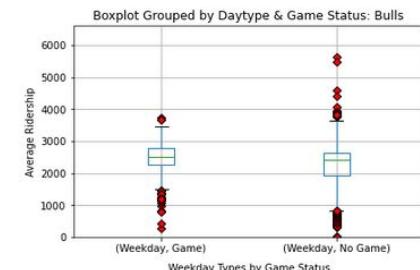
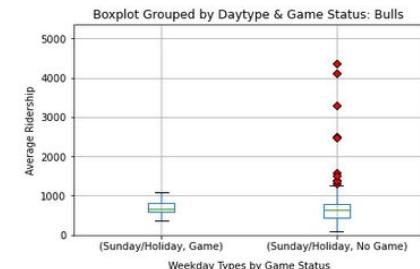
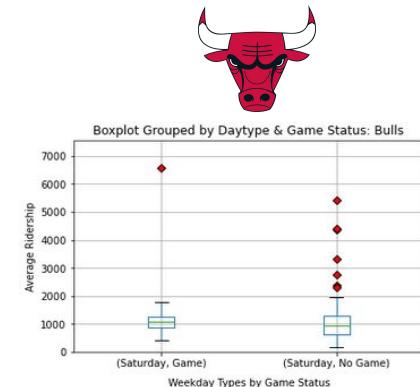
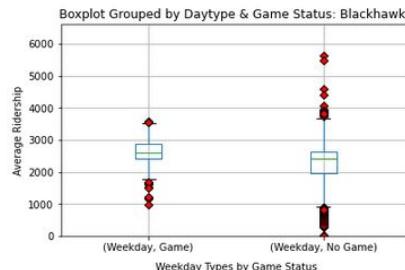
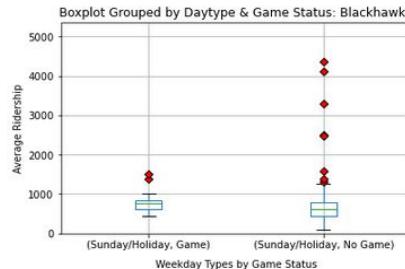
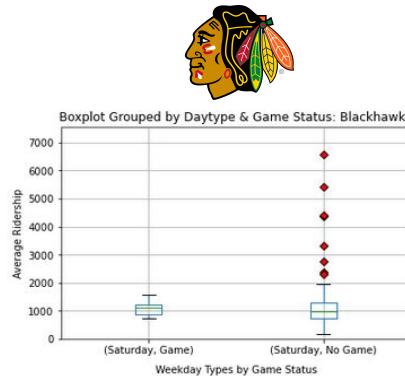
F\_onewayResult(pvalue=0.049312564058412535)

Sunday:

F\_onewayResult(pvalue=0.7759471601159668)

Weekday:

F\_onewayResult(pvalue=2.5937901182414625e-09)





# Data Assumptions and Missing Information

## Unfair Assumptions:

- Uptick in station ridership is largely due to game day traffic
- Not accounting for other events in the areas
- Assuming all gameday traffic only goes through the closest 'L' stop to stadium
- Attendance and ridership are directly correlated

## Additional data sources to tell a better story

- Gameday attendance
- Fan demographics
- Weather data
- Rideshare data: Uber/Lyft
- CTA 'L' destination and time of ride
- CTA Bus/Divvy/Scooter