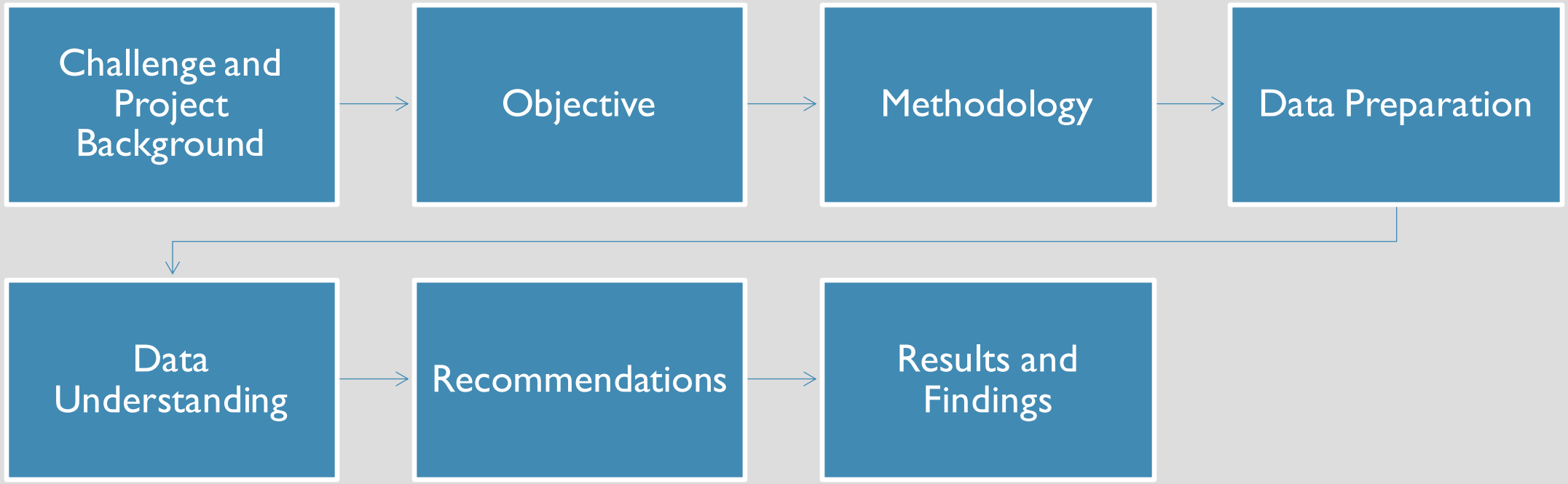
BAIS CAPSTONE PROJECT

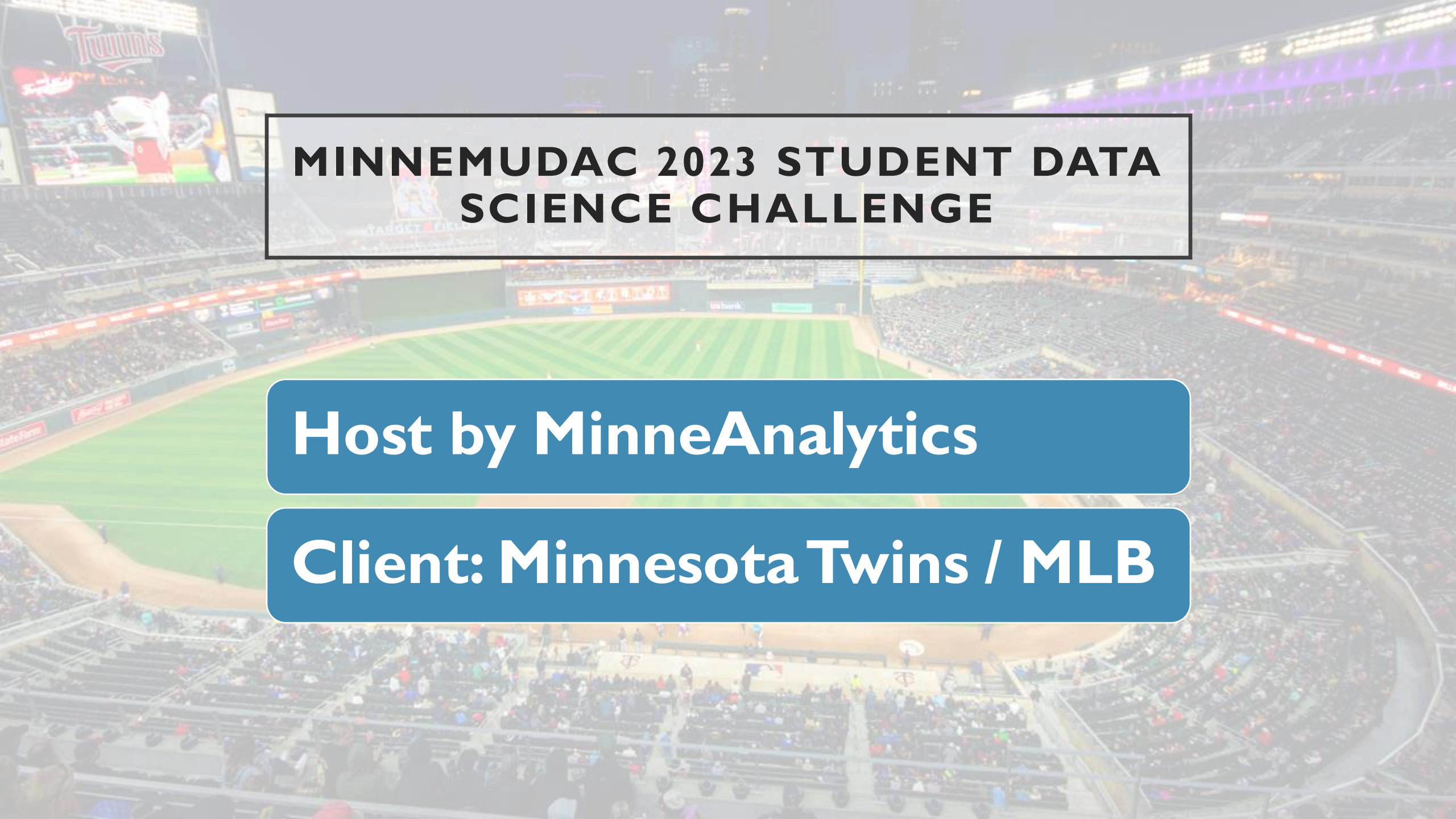
Home Attendance for the Minnesota Twins

MLB Group

Beth Tiggelaar, Katie Zawoyski, Ryan Yuson, Sarah Phillips

AGENDA


Challenge and Project Background
Objective
Methodology
Data Preparation
Data Understanding
Recommendations
Results and Findings

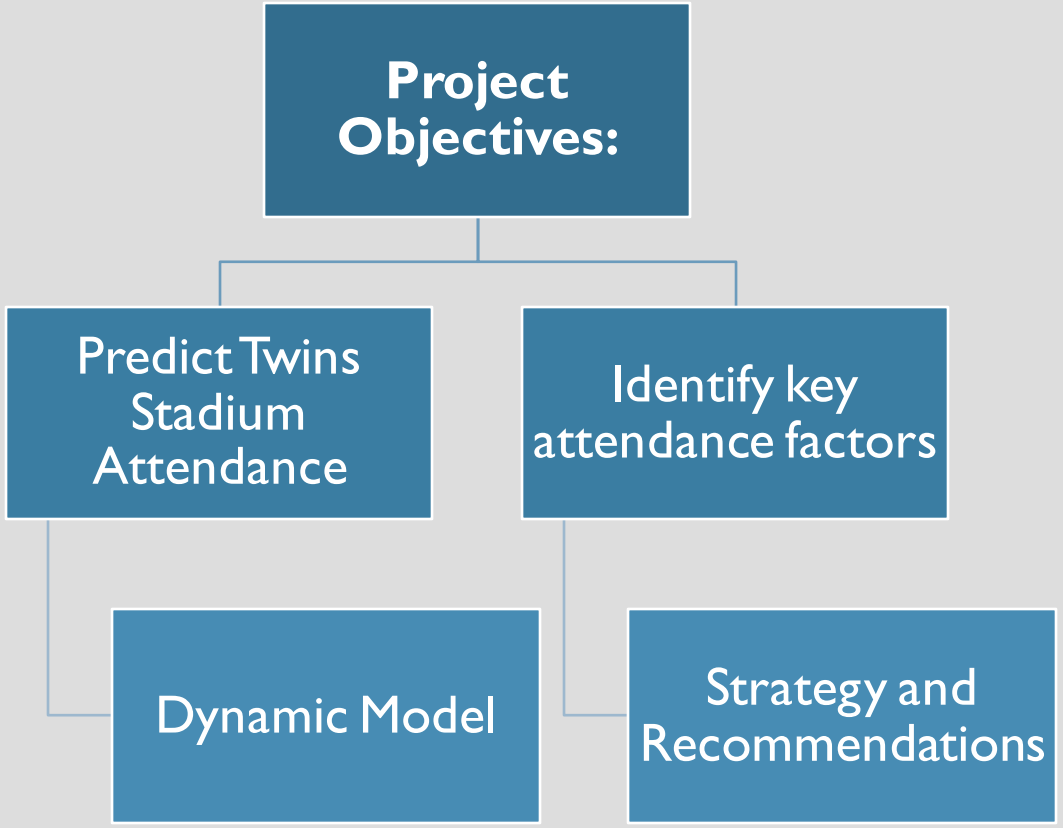


MINNEMUDAC 2023 STUDENT DATA SCIENCE CHALLENGE

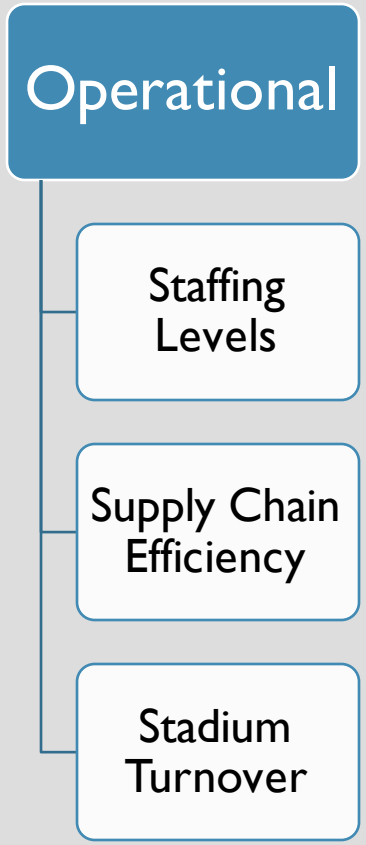
Host by MinneAnalytics

Client: Minnesota Twins / MLB

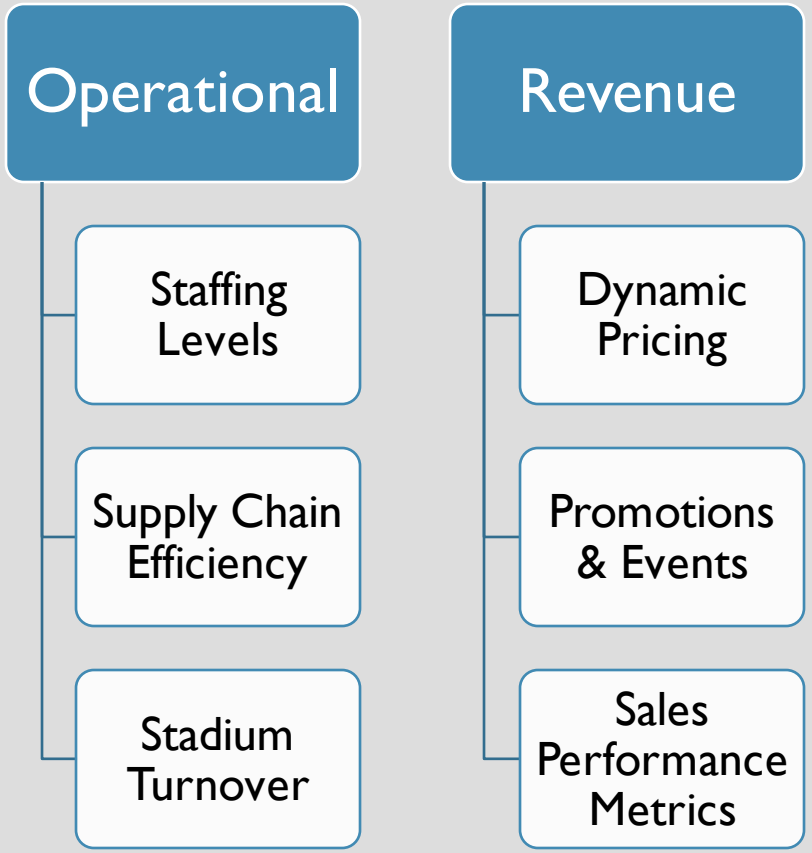
UNDERSTAND MLB HOME GAME FACTORS TO INFORM ATTENDANCE GROWTH



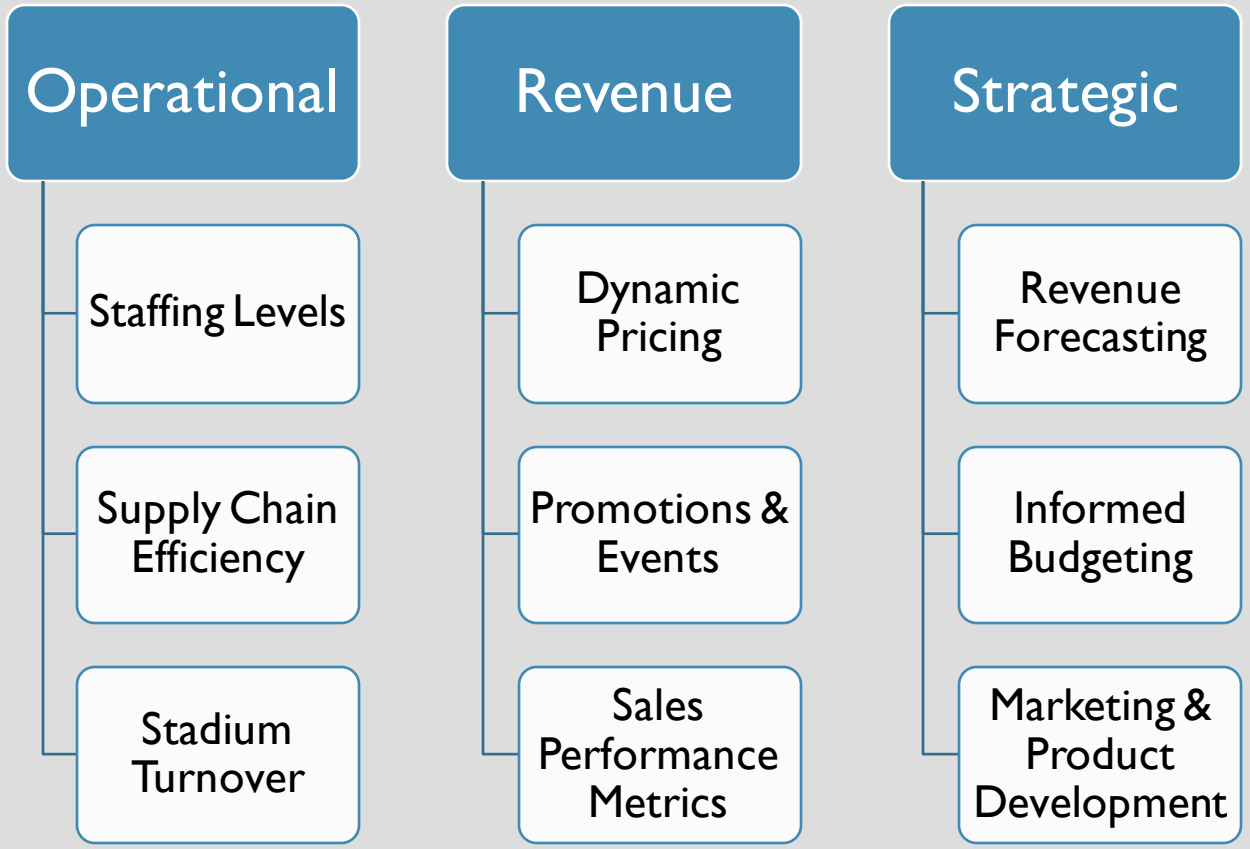
PREDICTING ATTENDANCE FOR BUSINESS PROCESS OPTIMIZATION

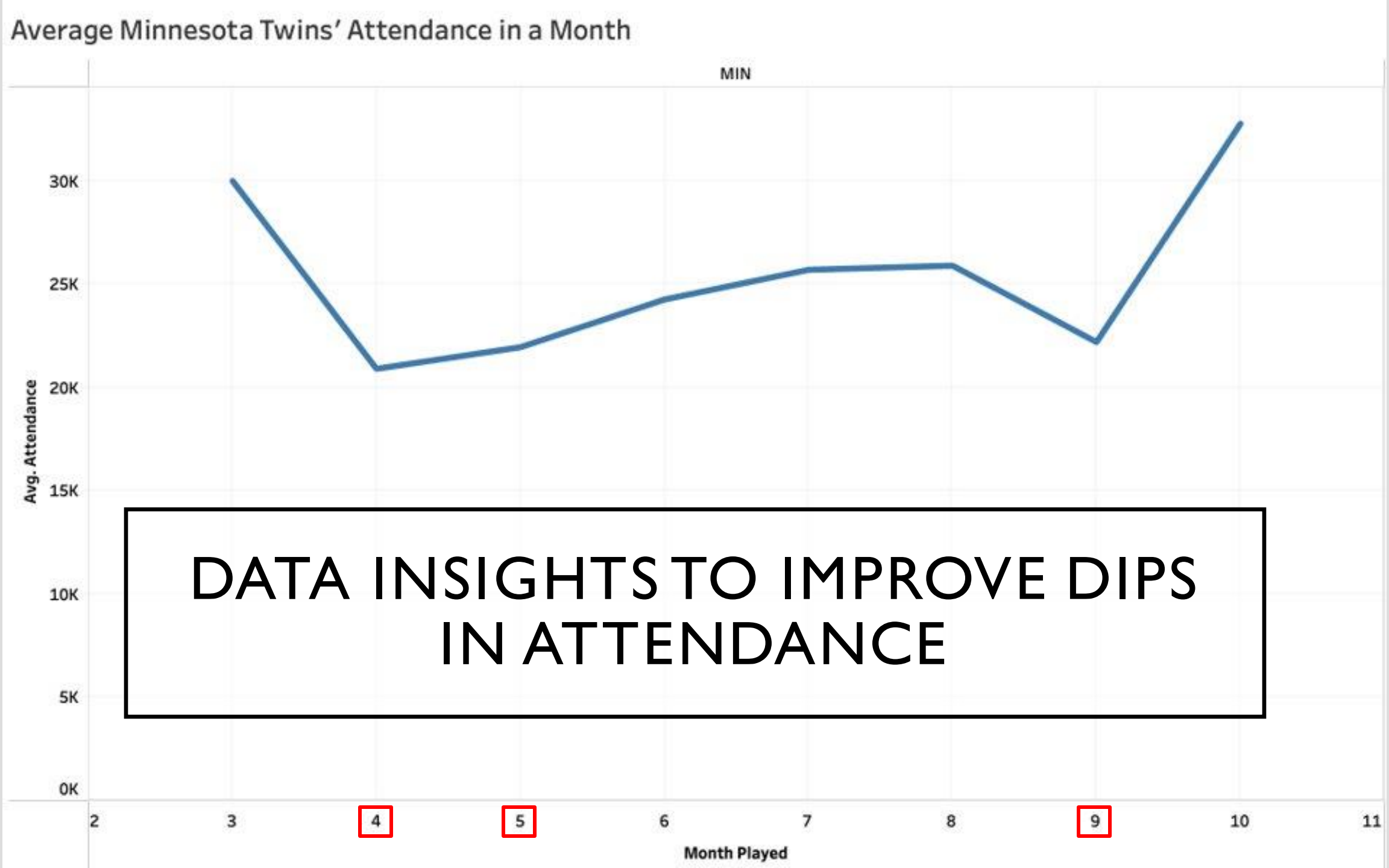


PREDICTING ATTENDANCE FOR BUSINESS PROCESS OPTIMIZATION



PREDICTING ATTENDANCE FOR BUSINESS PROCESS OPTIMIZATION





DATA INSIGHTS TO IMPROVE DIPS IN ATTENDANCE

METHODOLOGY

T-test and Variable Importance

METHODOLOGY

T-test and Variable Importance

Initial regression model

METHODOLOGY

T-test and variable importance

Initial regression model

External data

METHODOLOGY

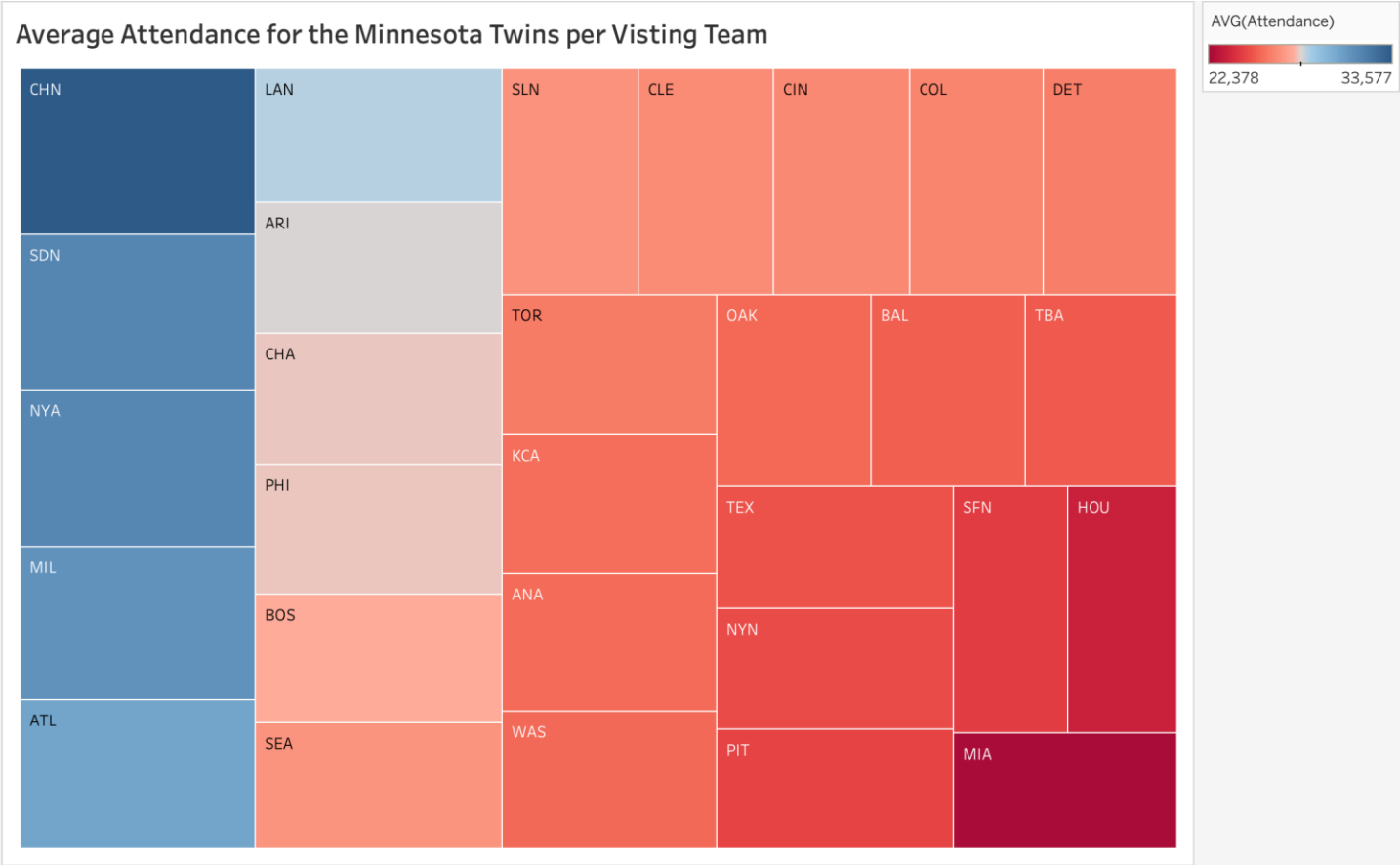
- T-test and Variable Importance
- Initial Regression Model
- Collect External Data
- Integrate into Final Regression

DATA PREPARATION

1. Properly format data types
2. Create dummy variables from categoric features
3. Handle column position for consistency
4. Remove instances with invalid attendance
5. Gather and integrate external data
6. Create subsets from GameLogs

AREAS OF DATA EXPLORATION

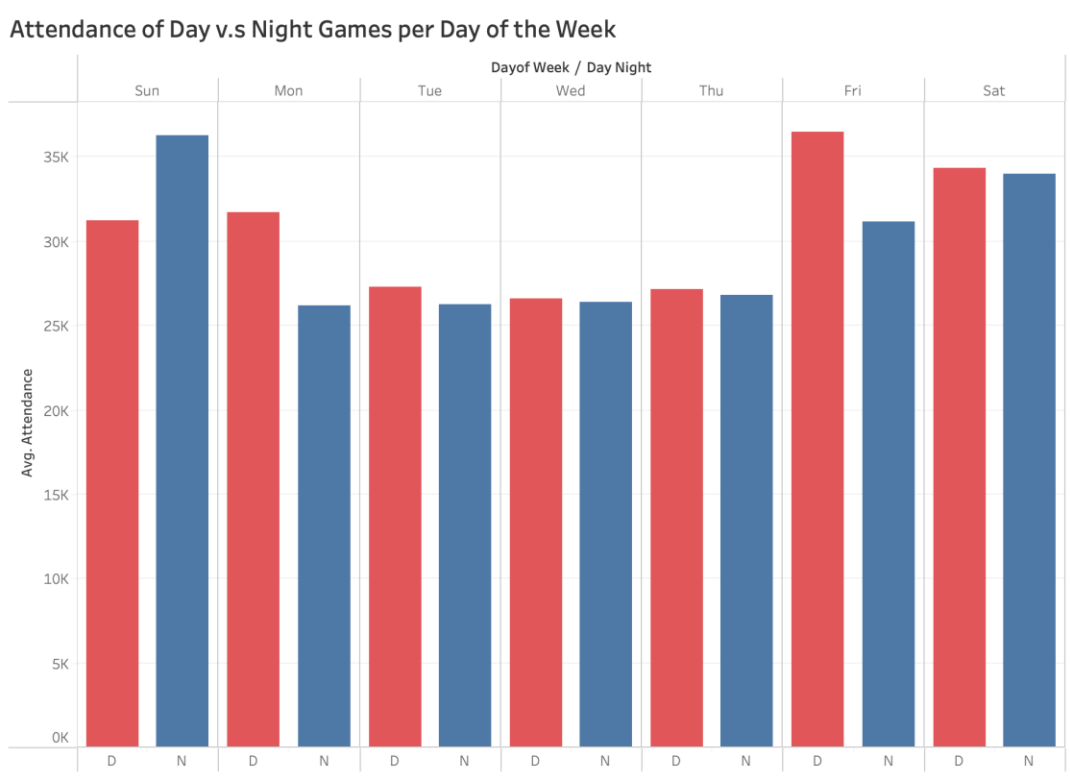
- Minnesota Twin's Average Attendance
- MLB Attendance Time Comparisons
- League Attendance Comparisons
- Variable Importance
- Attendance by Location

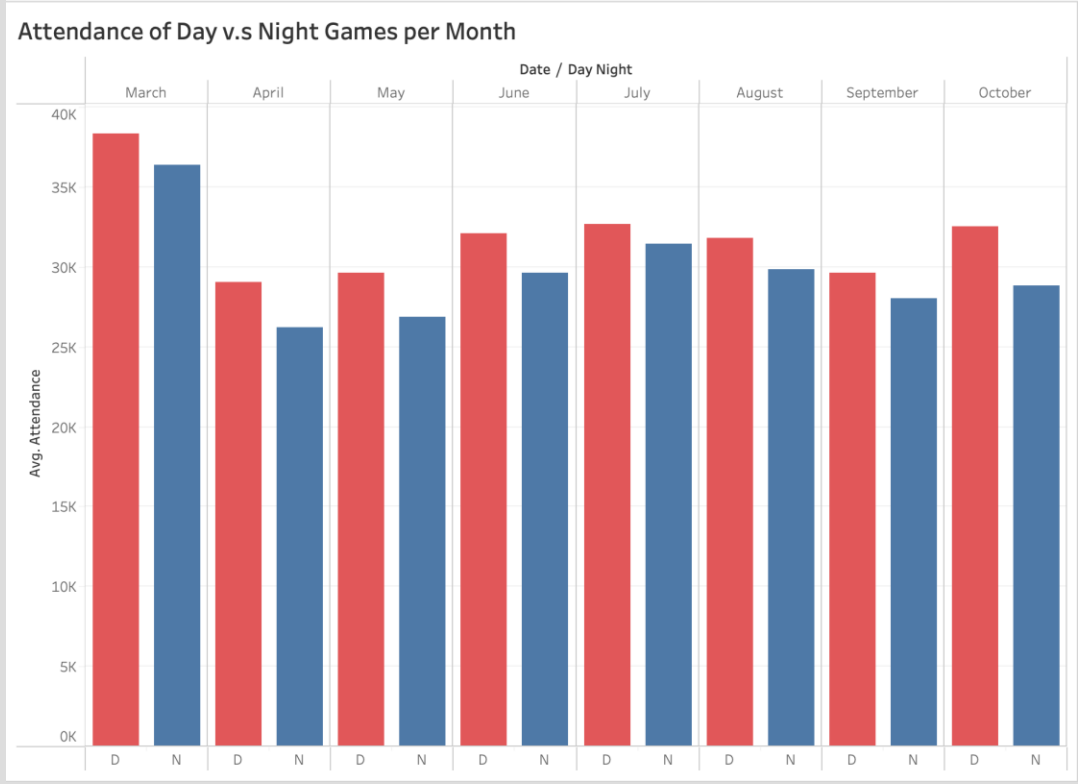


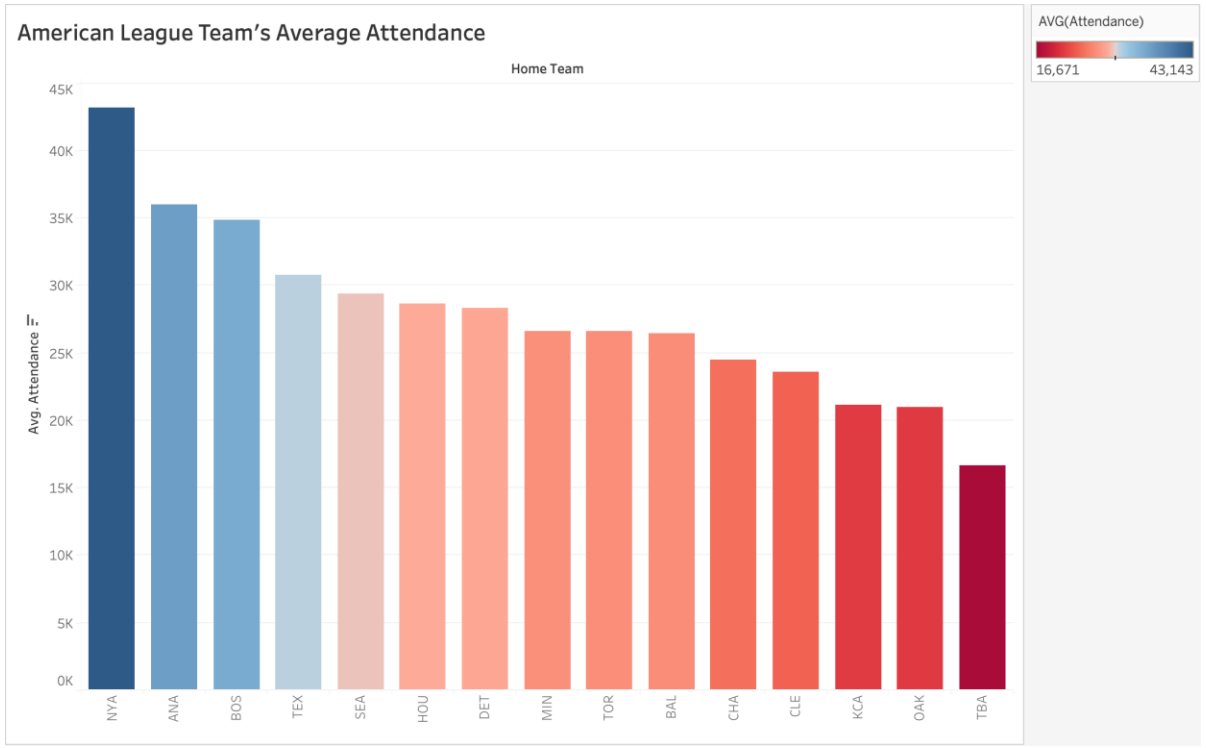
Chicago Cubs have the **highest** attended games on average
Miami Marlins have the **lowest** attended games on average

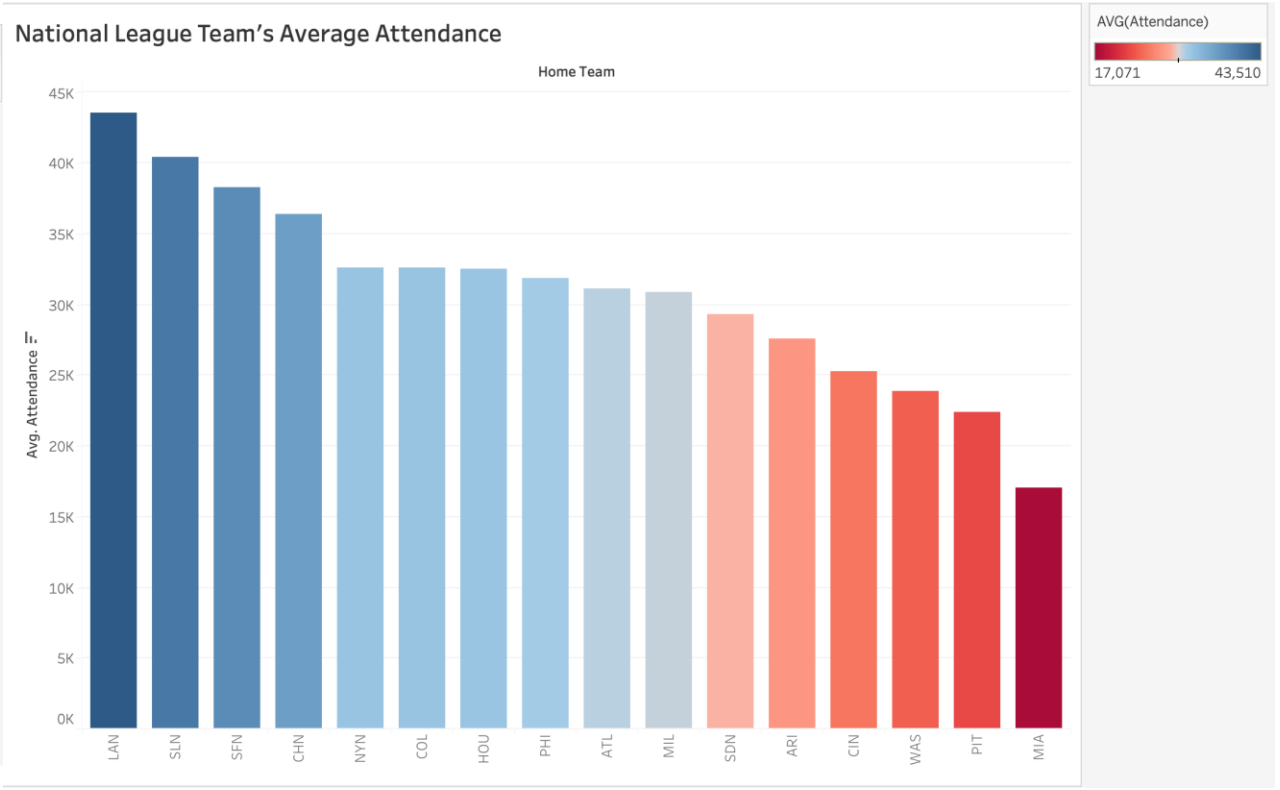
MINNESOTA TWINS ATTENDANCE COMPARED TO VISITING TEAM ATTENDANCE

DAY GAMES AVERAGE **HIGHER** ATTENDANCE THROUGHOUT MLB



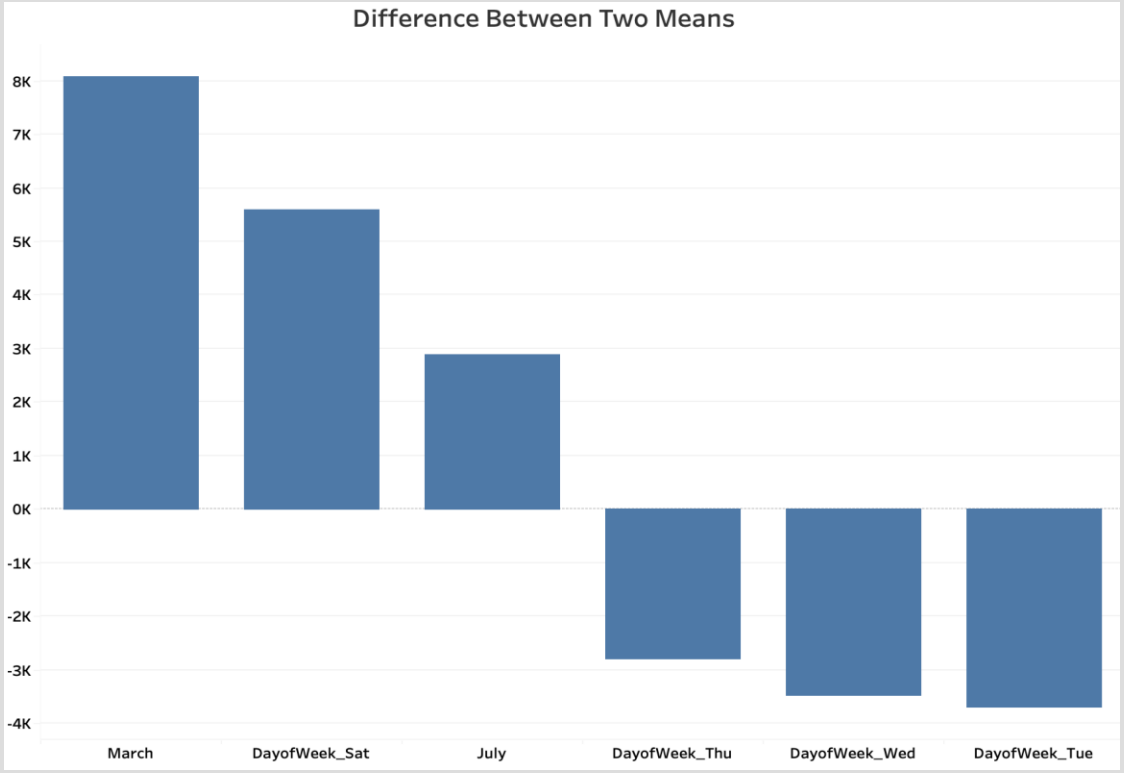




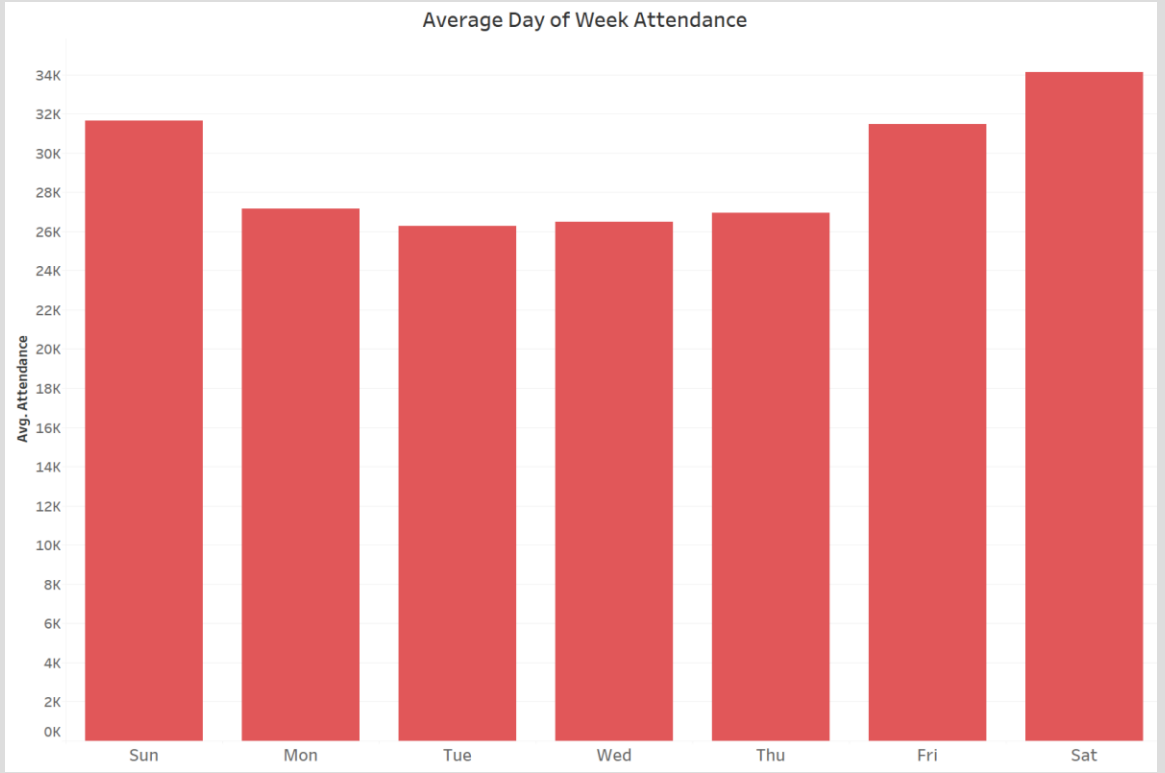


ATTENDANCE VARIANCE BY LEAGUE

VARIABLE IMPORTANCE PROVIDES DATA INSIGHTS



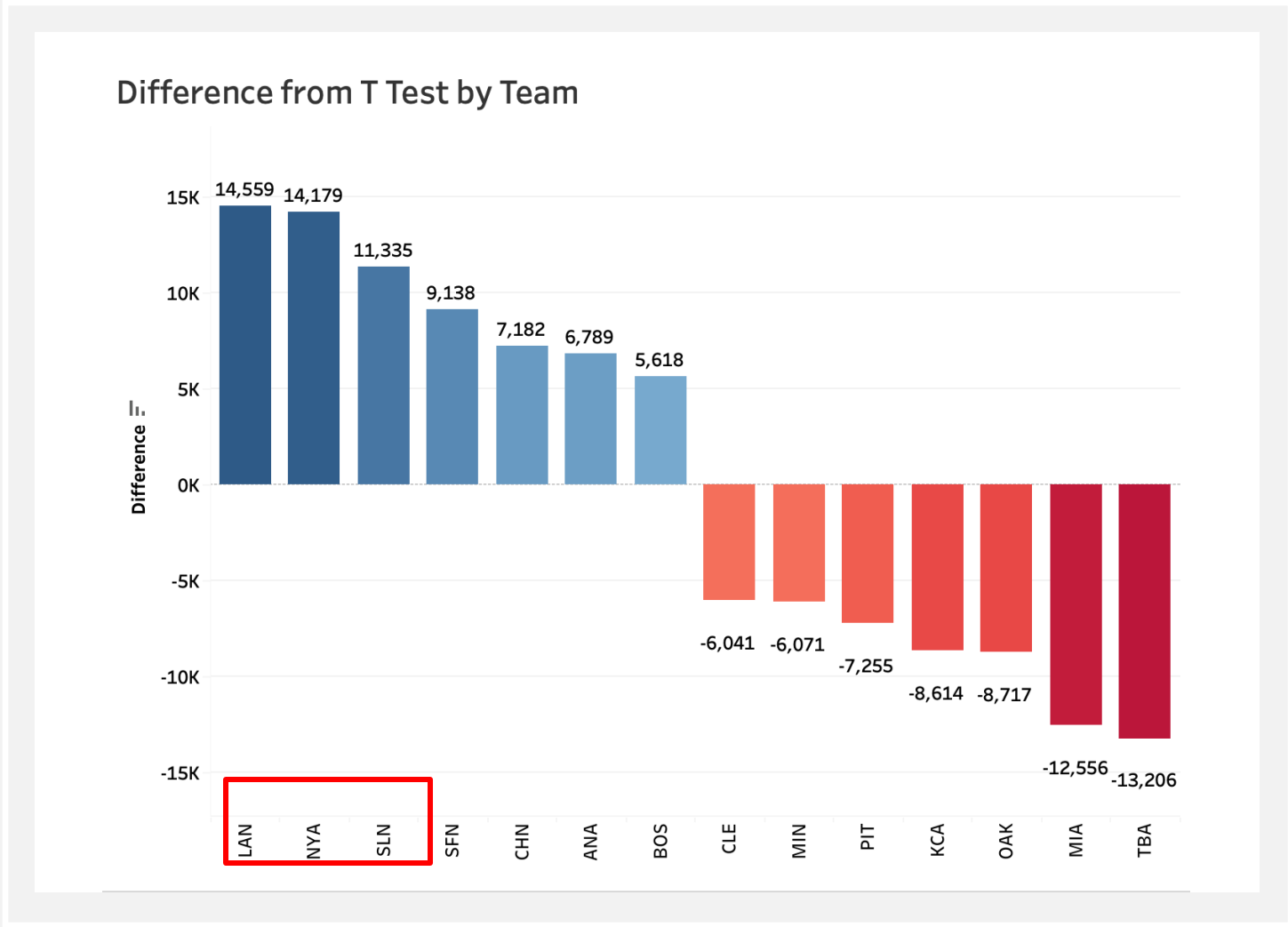
March, July, and **Saturday** games have a **positive impact** on attendance



Weekdays average the **lowest** Attendance

HIGHLY IMPACTFUL TEAMS HELP INFORM BUSINESS STRATEGY

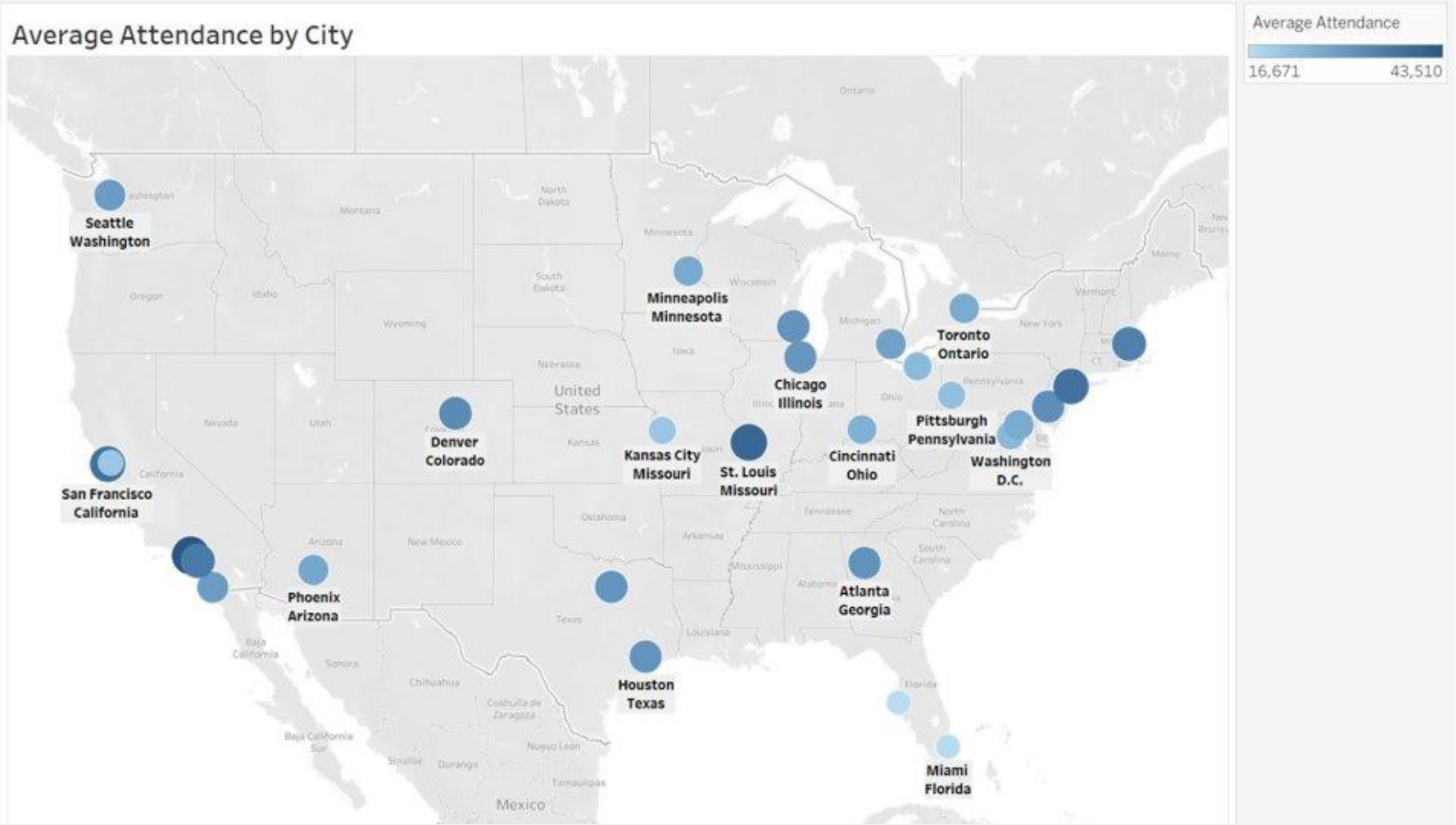
- **Los Angeles Dodgers, New York Yankees,** and **St. Louis Cardinals** have largest positive impact on attendance



EXTERNAL DATA PROVIDES A HOLISTIC APPROACH TO STADIUM INSIGHTS

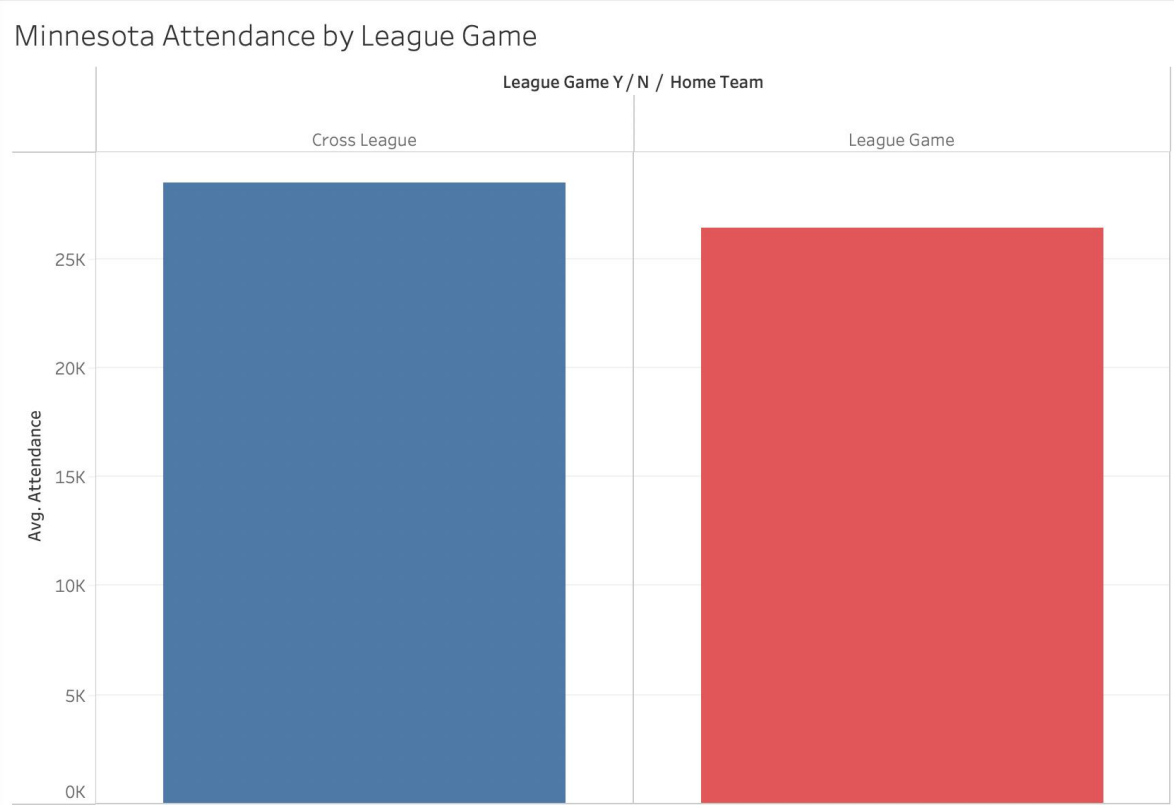
Modeling

STADIUM LOCATION AFFECTS AVERAGE ATTENDANCE

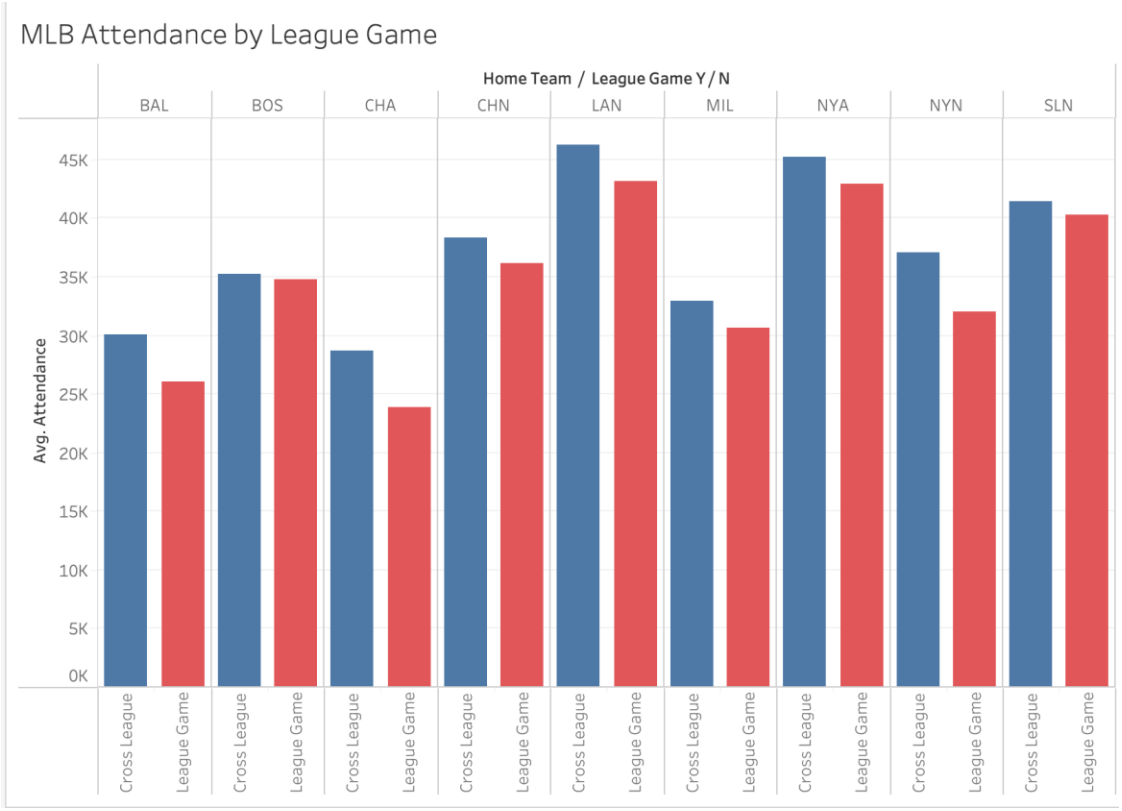


- **Los Angeles, CA** and **St. Louis, MO** have the highest average attendance
- **St. Petersburg, FL** and **Miami, FL** have the lowest average attendance followed by **Oakland, CA**

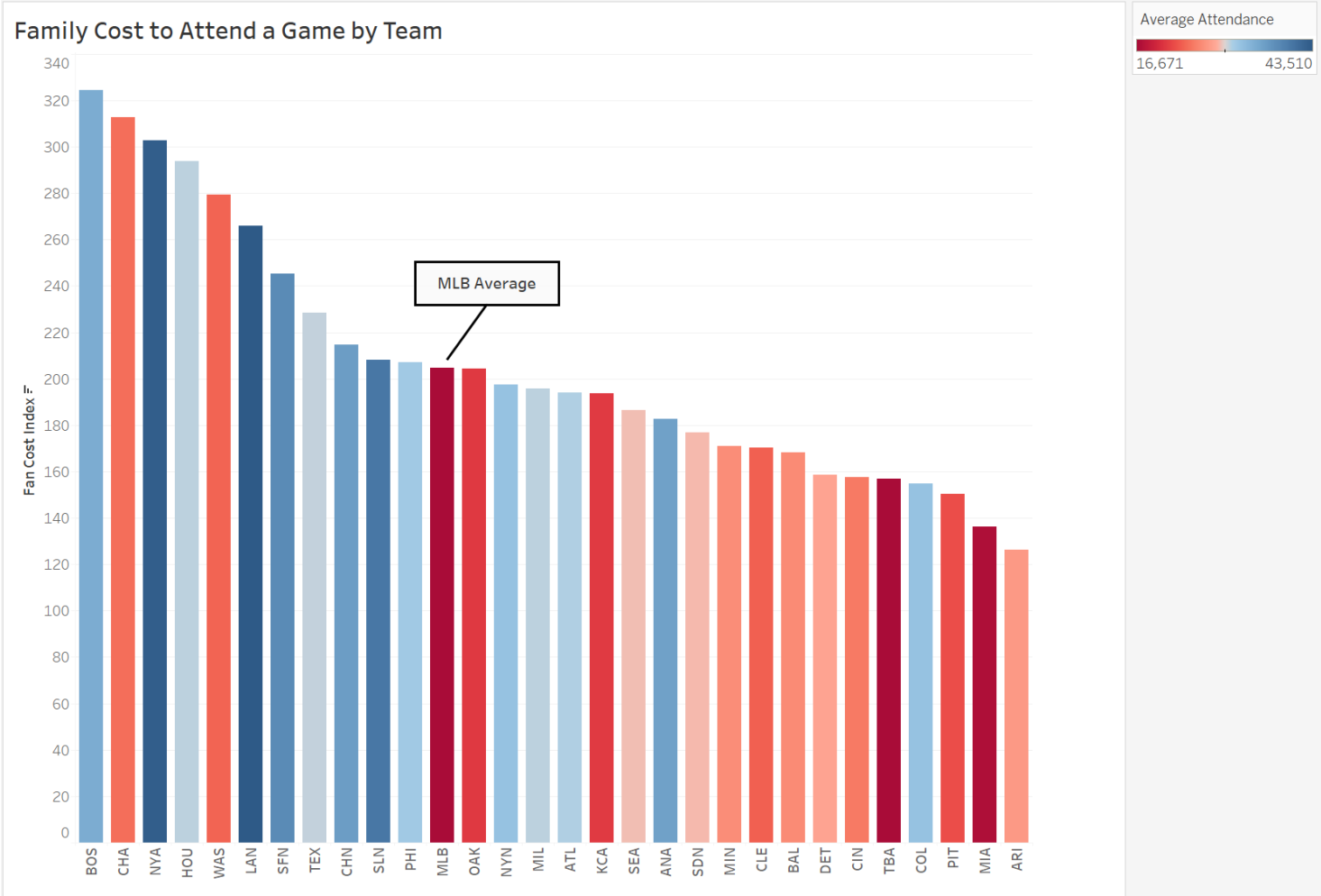
Minnesota Average Attendance by League



Other MLB Average Attendance by League



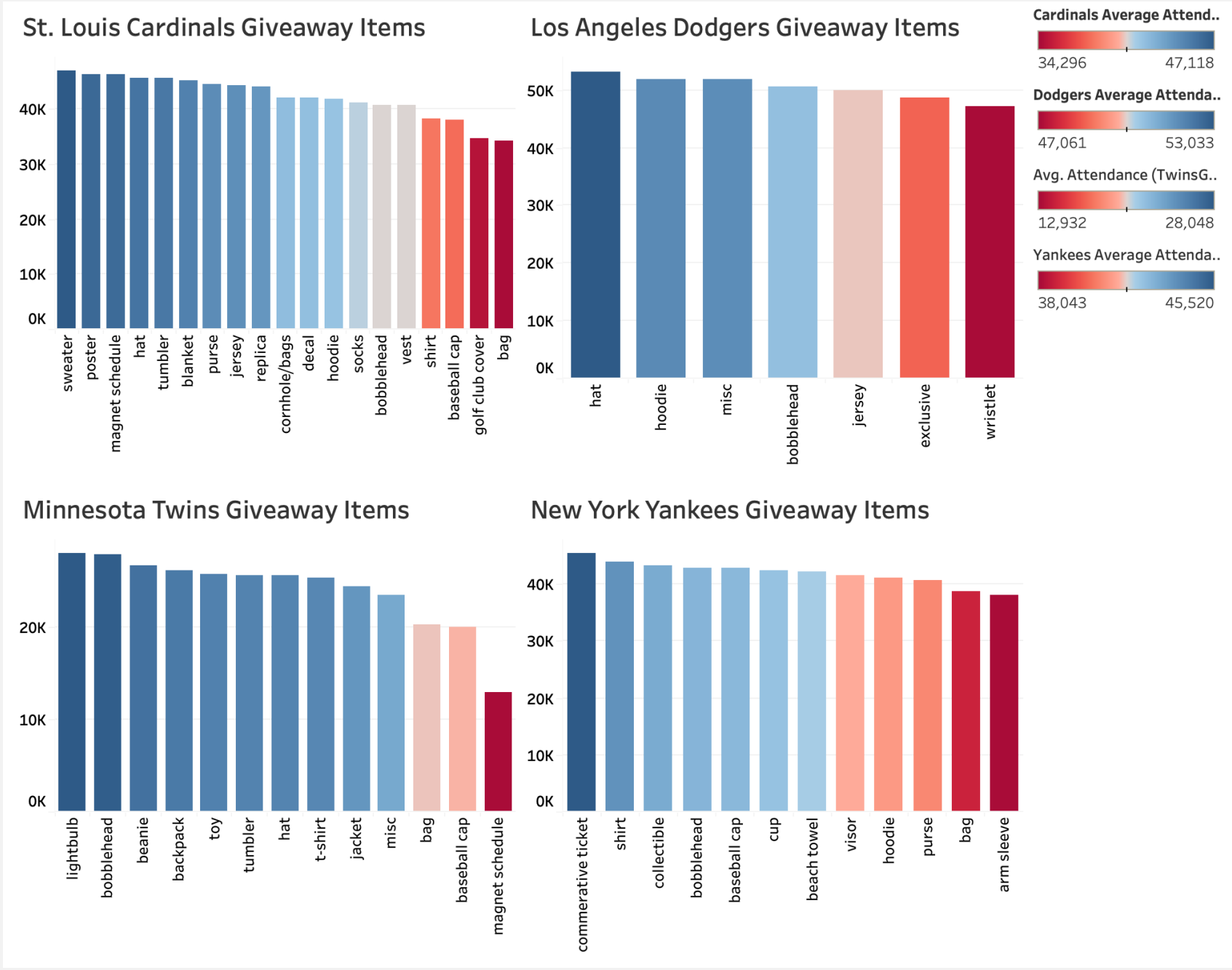
CROSS LEAGUE GAME AVERAGES **HIGHER** ATTENDANCE RATES



COST TO ATTEND A GAME VARIES BY TEAM AND CITY

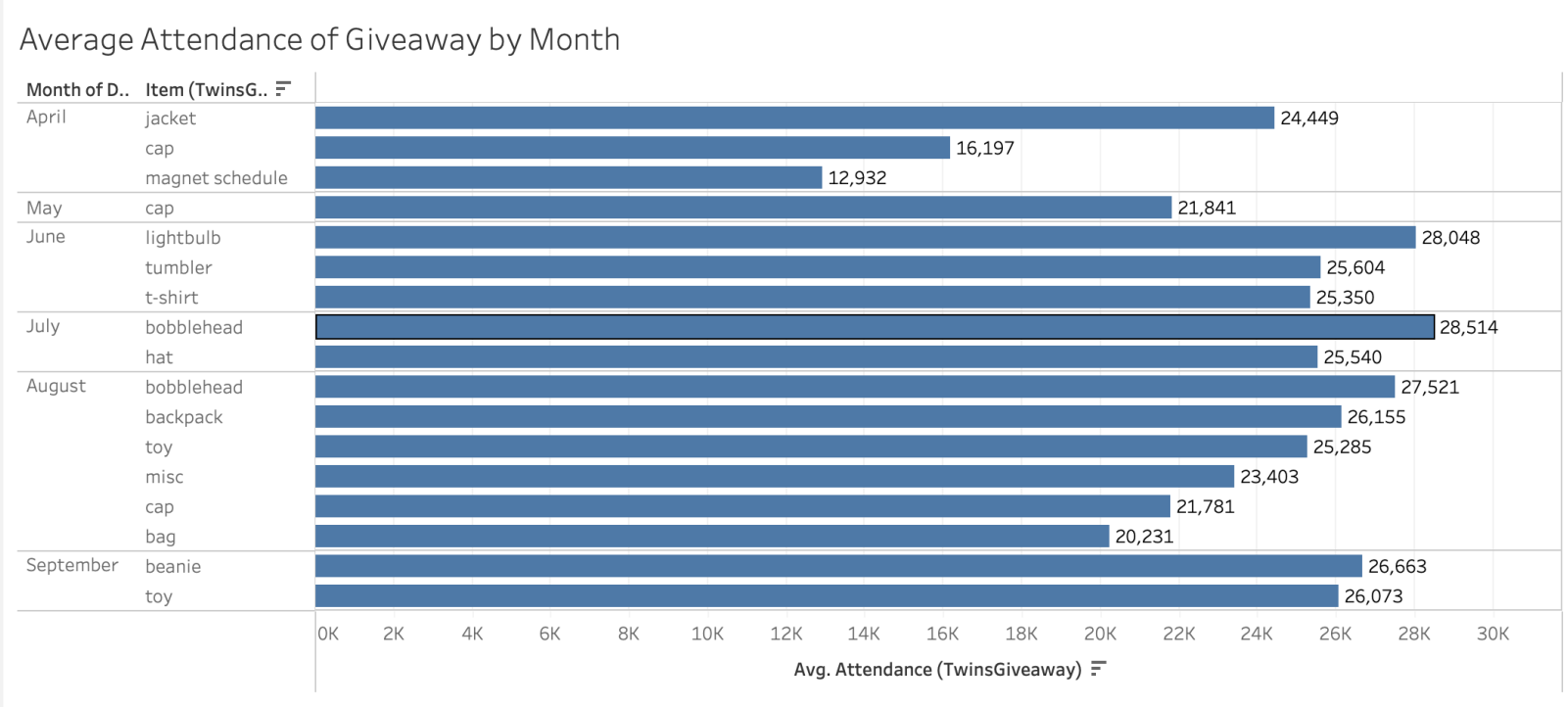
LEAGUE WIDE PROMOTIONAL TRENDS

- **St. Louis Cardinals**
  - Personal clothing items, bundled items, adult 16+ items average higher attendance
- **Los Angeles Dodgers**
  - Bobbleheads highest count/account for most sum of attendance
- **New York Yankees**
  - Commemorative/collectible items draw attendance for special occasions



TWINS PROMOTIONAL INSIGHTS

Giveaways are heavily saturated in summer months	Increase Spring/early Summer giveaway opportunities
Bobbleheads are utilized ineffectively	Increase bobblehead variety and giveaway opportunities
Child-specific giveaways observe higher attendance	Explore and offer more family-oriented promotions



PROJECT RECOMMENDATIONS

April, May, September

PROJECT RECOMMENDATIONS

April, May, September

Spring Games

Spring Games

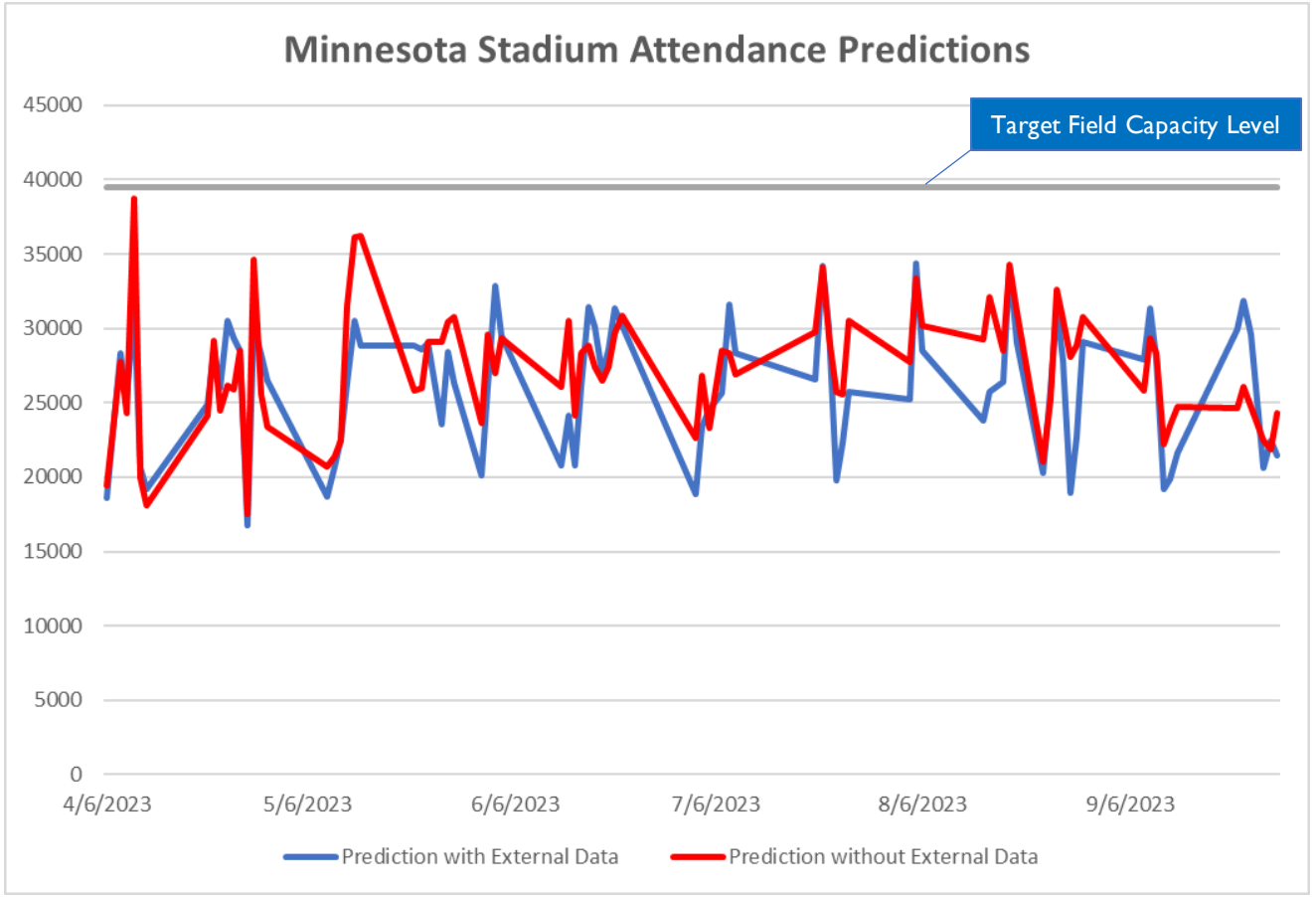
Spring Games

UTILIZING REGRESSION MODELS TO PREDICT 2023 MLB STADIUM ATTENDANCE

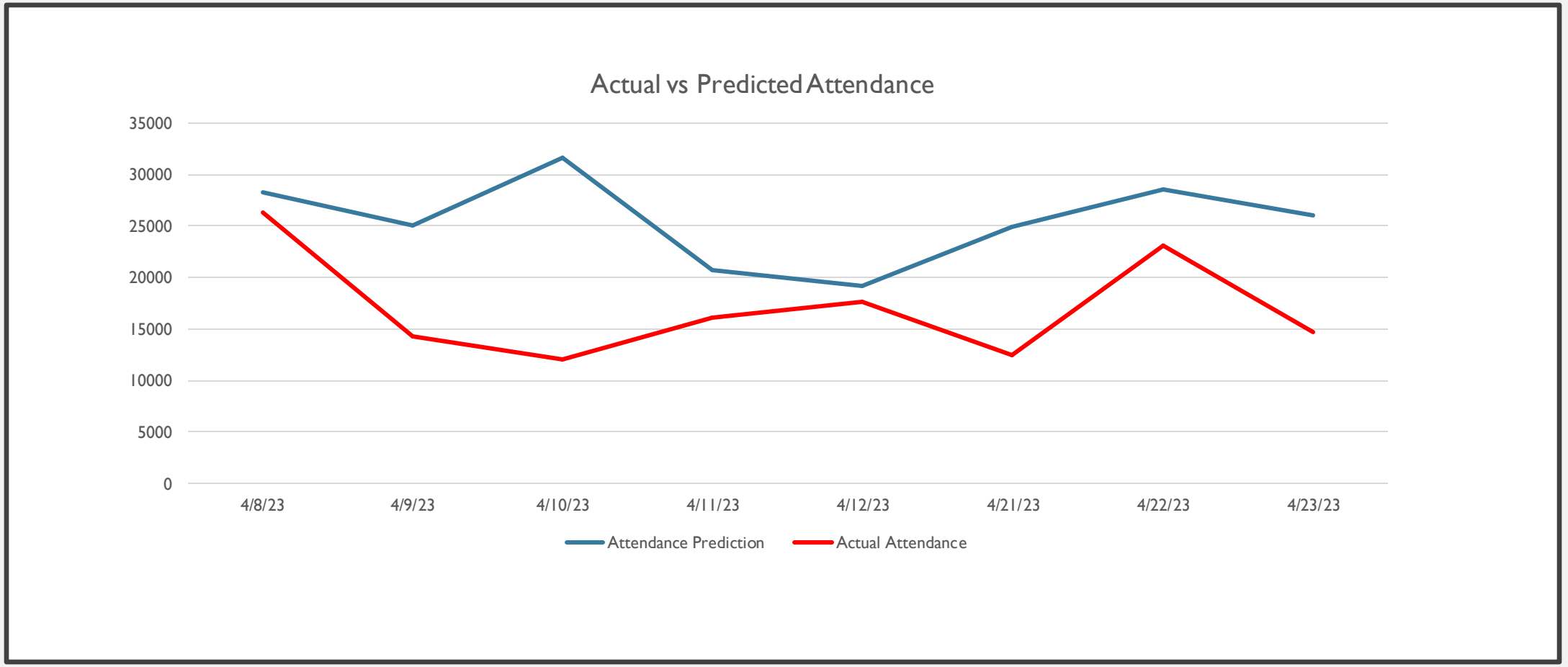
- Eight regression tests ran
- **Trained** on 20 years of data, **tested** on the 2022 season
- Evaluated on Mean Absolute Error
- **Random Forest** – All predictors without Cross Validation had the strongest results

Data Set	Model Approach	Train MAE	Test MAE	Difference
Original Data	RF - All predictors wo CV	1951.817	2041.051	89.234
External Data	RF - All predictors wo CV	1871.633	2239.578	367.945
External Data with Previous Attendance	RF - All predictors wo CV	1579.74	1384.599	-195.141

STADIUM PREDICTION RESULTS

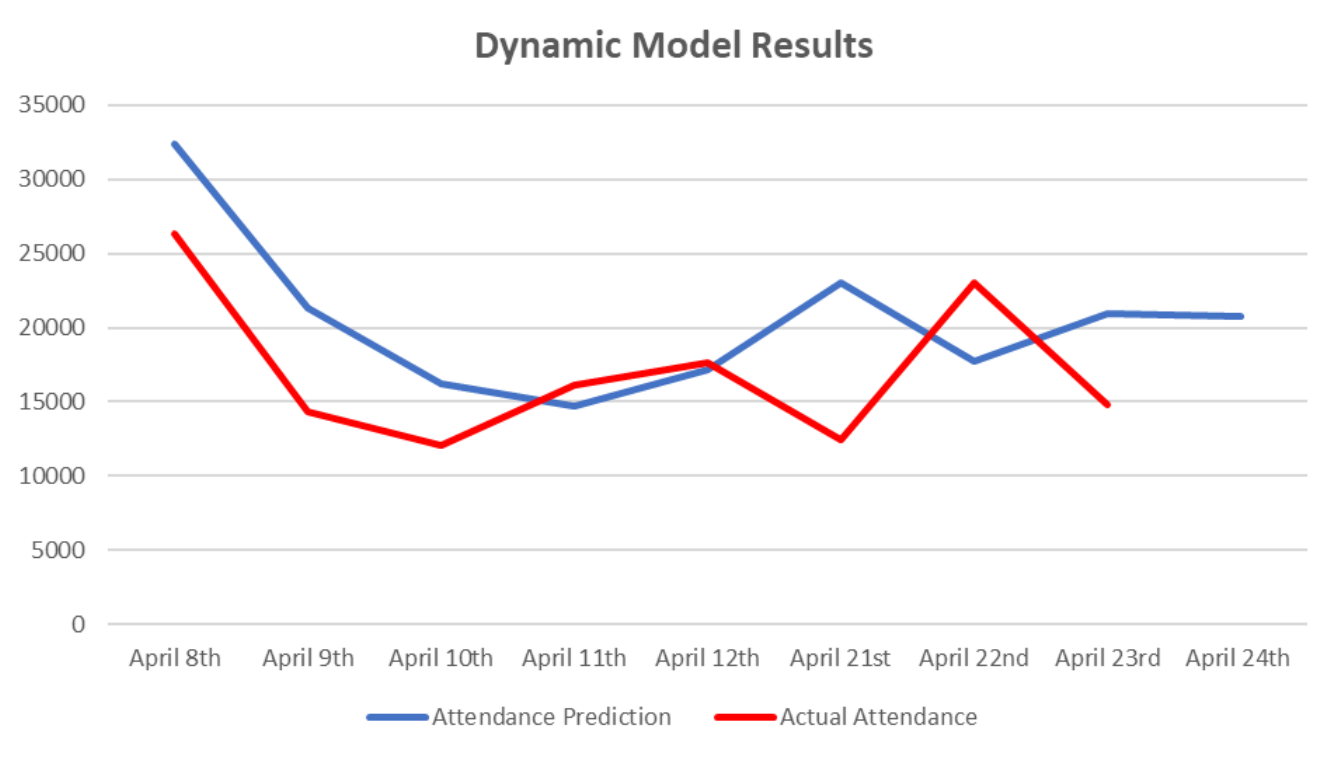


TWINS ATTENDANCE MODEL RESULTS FROM PLAYED GAMES



- *4/9 was Easter – not noted in model

THE **DYNAMIC MODEL** PRODUCES THE MOST ACCURATE RESULTS



OUTCOMES AND STUDENT TAKEAWAYS

Communication is the key factor to team success

OUTCOMES AND STUDENT TAKEAWAYS

Communication is the key factor to team success

Being thorough when going through original data is crucial

OUTCOMES AND STUDENT TAKEAWAYS

Communication is the key factor to team success

Being thorough when going through original data is crucial

Setting a standard meeting time leads to higher quality of work

APPENDIX

- https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.mlb.com%2Fnews%2Ftwins-announce-upcoming-brand-refresh-staff-promotions&psig=AOvVaw1AaDZNB-akK-KI5PSHVpzj&ust=1682004594179000&source=images&cd=vfe&ved=0CBAQjRxqFwoTCMD30MShtv4CFQAAAAAdAAAAABAR
- https://parkmobile.io/blog/target-field-ball-game/