**Minnesota Twins Stadium Attendance**

Beth Tiggelaar, Katie Zawoyski, Ryan Yuson, Sarah Phillips

**Executive Summary**

The objective of this project was to build a predictive model for MLB baseball attendance based on twenty years of historical game log data, as well as provide actionable insight into underlying factors that influence home game attendance. Overarchingly, we aim to provide recommendations on operational, revenue, and financial strategy impacts for the Minnesota Twins.

**Project Insights and Recommendations**

Firstly, we have developed and compared the performance of several models following a variety of linear and non-linear approaches. Our final predictive model has shown a relatively high degree of accuracy in forecasting attendance for MLB games. Data exploration contributed to identifying factors such as seasonality, day of week, time of day, opponent, and previous game attendance to all have a significant impact on attendance levels.

Additionally, our analysis has provided insight into several areas of business operation that currently affect and can have a positive impact on stadium attendance and revenue. Moreover, we have collected external sources of stadium-specific data to include factors such as seating capacity, city population, club value, and operating income.

We have begun mapping a set of actionable recommendations to help the Minnesota Twins optimize their operational and financial strategy, with the goal of maximizing attendance and revenue generated by both ticket, concession, and merchandise sales. These recommendations include utilizing dynamic ticket pricing, improving facility/labor processes and amenities, and leveraging targeted marketing through themed games, exclusive ticket bundles, and promotional item giveaways.

There have been several technical risks throughout the project, such as feature engineering and the proper handling of existing and created variables, as well as the risk of an overfit model. Mitigation of these risks has included close project management and regular communication within our group, as well as robust testing and validation of our predictive model.

**Next Steps and Future Scope**

We are currently focused on refining our recommendations and findings and reassessing our project in the scope of our sponsor’s expectations to develop an implementation plan for our final insights and strategies. We are also conducting additional analysis to identify potential areas of improvement for our predictive model and specifically account for progressive team performance. We will continue exploring other MLB organizations and gain understanding of league-wide attendance, strategy, and financial success. We believe that these next steps will help to ensure a successful delivery of the project and provide the Twins with the insight and recommendations necessary to optimize their operational and financial strategy moving into the 2023 season.

**Project Overview & Objectives**

This project was provided to us by MinneAnalytics with a client of the Minnesota Twins. They presented the following case, prompting our team to discover variable trends and factors that influence stadium attendance for the Minnesota Twins and other MLB teams. Ultimately, the goal was to use those factors to predict stadium attendance for the 2023 season. Looking at the significant factors that affect Twins attendance can benefit them in many ways. Having better understanding of factors that affect attendance can help highlight ways of improvement. Comparing these factors to other MLB teams is also beneficial. Every team has slight variance in their attendance and promotional strategies so it can be beneficial to compare strategies to high performing attendance teams. Using these insights, one of our goals was to provide recommendations to the Twins to increase overall attendance.

The purpose of this project is to provide a better understanding of stadium attendance to increase the business to provide value in operations, revenue, and strategic segments. The operational benefits include better predicting staffing levels and increasing supply chain efficiency. If the Twins can gauge what each game's attendance levels are, they can better estimate how much staff to schedule as well as how much food or other supply to have in stock. These metrics will also save the Twins money, helping increase revenue. Increasing overall attendance will also help profit in terms of generating more ticket revenue. Having a better understanding of attendance levels will also assist in dynamic pricing and better promotional strategies. Along with this, the more fans in attendance, the more they will spend on concessions and merchandise. All increasing Twins revenue. Looking at the strategic side, these metrics can help with revenue forecasting, informed budgeting, and improved marketing development.

**Methodology**

The approach we took to this project was mainly focused on how to prepare and execute our regression correctly. To start the project, we ran two t-tests. One t-test was ran to determine what variables had the most significant positive impact on attendance. From here, a second t-test was ran to determine what teams had the largest positive impact on attendance. The purpose of this t-test was more so to aid us later while conducting further research and creating recommendations for the Minnesota Twins. After variable importance was conducted, we ran an initial regression model. Before running the regression model, we determined that random forest with all predictors without cross-validation was the most accurate choice. Our initial regression featured data that was provided to us from the competition that we cleaned and added dummy variables to. From this model, we concluded that external data was needed to make a more accurate model. The data we collected was pertaining to capacity of the stadium and the city’s population. After integrating the external data, we ran three separate random forest regressions. The first regression was just with the external data being integrated from there. The second regression was a subset of the last five-years from the dataset including the external data. The last regression, which we found to be most accurate, was a dynamic regression model that included previous attendance.

**Data Preparation**

Data preparation for this project was split into six steps: cleaning the original data, creating dummy variables, repositioning columns, removing invalid attendance, gathering external data, and creating subsets of the overall dataset.

The first task was cleaning the dataset provided by the MinneMUDAC Competition. The main issues dealt with were the date’s formatting, varied team abbreviations, and games not played in home stadiums. The dates of games were listed as a large integer instead of separate dates. For example, instead of 2000-08-08, the date would be 20000808. There was nothing wrong with the team abbreviations. However, many teams had multiple team abbreviation. One example would be the Washington Nationals had WAS as their main abbreviation. We discovered though that the Washington Nationals were originally the Montreal Expos, but they moved to Washington D.C. in 2005. The data provided started in 2000, so the Nationals were also referred to as MON. Another example would be the Miami Marlins. They used to be called the Florida Marlins, so they had both FLO and MIA as their assigned abbreviations. The key to this was not so much cleaning the data but learning and being more thorough with the data. The last part of data cleaning was fixing the games that were not played in team’s home stadiums. Some games were played in other places outside of a team’s home stadium, like there were games played in Tokyo. For this reason, we had to go back through and match the correct stadium capacity to each game.

The next step in data preparation was creating dummy variable. The dummy variables were created to replace categoric variables. Some examples of categoric variables were used is the day of the week the game was played, the month the game was played, and the game being played at nighttime or daytime.

Following creating dummy variables, we needed to adjust columns positions. With moving variables around, columns and their variables became out of order and inconsistent. For the regression to work, we needed to make sure that column positioning was uniform across each of our datasets.

After repositioning columns within our dataset for consistency, we focus on correcting the given attendance. We first eliminated all instances of the 2020 season from our dataset because the attendance for each game was zero and we did not want outliers or anything to skew our final results. After eliminating 2020, we found more instances of attendance equaling zero. The attendance being zero was mainly due to games being postponed from bad weather, so we took out these instances as well. Lastly, we found a few cases where attendance was listed as a negative number. These were just a few small errors in the dataset, so we eliminated those few games with negative numbers.

The next part of our data preparation for the final regression model was collecting and integrating external dataset. This was a huge factor in the success of our final random forest regression with all predictors and without cross validation regression model. The approach we took for this step was splitting our external data into two categories. One category was meant for putting the external data into the actual regression model. The other category of external data was meant for exploration and to be used in our later recommendations.

For modeling external data, we focused on adding a few key variables. Stadium capacity was one of the main areas because it creates a stopping point for predictions. The next variables for modeling are city population and the number of professional teams within a city. Cities with a higher population and more professional sports teams are more likely to have a higher attendance. The reason being these cities not only have the people to fill up stadiums, but they have people who are passionate about sports and willing to go to games. Take Pittsburgh for example; not only does this city have the Pirates, but also the Steelers and the Original Six team the Penguins. The other two variables that were includedino the model were previous attendance and season. Season refers to whether a game is in the spring, summer, or fall.

For exploration of external data, there were four main areas we looked further into. The first area is league affixation. There are two leagues within the MLB: the American League and the National League. We wanted to see the difference between games where teams were from different leagues and games where teams were from the same league. Another reason was just comparing the two overall leagues in general. The second area of exploration was the specific city population of where the stadium was located. Many times, stadiums are not located in the place where the teams are listed to play. Famously, the New York Jets play in New Jersey and the Dallas Cowboys play in Arlington. It was important to look at the surrounding areas to where a stadium is actually located to determine possible attendance or possible issues relating to attendance.

Cost of attending was another factor we looked further into, as the cost to attend may be a deciding factor on why a person may or may not attend a game. The last areas of focus for external data used for exploration was team promotions. Offering promotions can help increase attendance for a game. We mainly focused on what the Los Angeles Dodgers, the New York Yankees, and the St. Louis Cardinals did differently from the Minnesota Twins when it came to promotions. The reason for these three teams is because of t-test we ran determine those three teams had the largest positive impact on attendance.

The last aspect of data preparation was creating subsets for our final dataset, GameLogs. We first created a subset of the 2022 season attendance, so we could use that to create our test data. For the dynamic model, we had to make a subset of just the Minnesota Twins attendance due to it having previous attendance included. The last subset we created included the past five years of attendance.

**Exploratory Analysis**

Exploratory analysis was an important part to our results and recommendations. The main exploration we did were surrounding topics about Minnesota’s Twin attendance, time comparisons, league comparisons, variable importance, and location.

Chart, treemap chart

Description automatically generated

Figure 1. Average Attendance at Target Field per Visiting Team

We began our analysis with looking at the Minnesota Twins. So, *Figure 1* is a tree map detailing the average attendance for the Minnesota Twins at Target Field per visiting team. The data used for this tree map is from the provided data after it had been cleaned. The map showed us that the Chicago Cubs had the highest average attended games with an average of 33,577. This makes sense because, even though the Chicago Cubs and the Minnesota Twins are in different leagues, the distance between the two teams is minimal. The Cubs fanbase is also well-established and they are known to be one of America’s most beloved teams. Before baseball was broadcast nationally, the only way to know what was happening live was by listening to the radio and WGN (the Chicago Cubs’ radio channel) was one of the few channels available. This inherently made extremely loyal fans nationwide. The lowest average attended games at Target Field are against the Miami Marlins with an average attendance of only 22,378. The same reason for the Cubs having highly attended games is the same reason for the Marlins having low attendance on average. The Marlins are at a disadvantage with distance. While the Cubs and Twins are both located in the Midwest, the Marlins are all the way in Florida. The Miami Marlins also have not built up a strong fan base. They do not have the advantage of being a nationally beloved team and have consistently played poorly so no bandwagon fans have been created. Like the Cubs, they are in a different league from the Minnesota Twins. So, the only reason anyone would be going to watch their games at Target Field is people just wanting to watch the Twins.

Chart, bar chart

Description automatically generated

Figure 2a. Day vs Night Games Split Up by Days of the Week

Chart, bar chart

Description automatically generated

Figure 2b. Day vs Night Games Split Up by Months

*Figure 2a* and *Figure 2b* present bar charts relating to time comparison. They depict the difference in the MLB’s average attendance between day vs night game. *Figure 2a* presents the average attendance of day and night games by the day of the week. *Figure 2b* presents the average attendance of day and night games by month. Based on the figures, we can conclude day games are attended on average more than night games for both day of week and month the game was played in. The only instance where the average attended games were played at night was on Sundays. We found this interesting since most people have work or school the following day and would normally assume people would not want to attend a game ending late at night. Overall, however, daytime games in MLB have a higher average attendance rate compared to nighttime games.

Chart, bar chart

Description automatically generated

Figure 3a. Teams within the American League’s Attendance

Chart, bar chart

Description automatically generated

Figure 3b. Teams within the National League’s Attendance

When looking a comparing the two leagues within the MLB, we found that the Minnesota Twins are in the lower quartile of average attendance. We concluded this by first looking at the Minnesota Twins’ league’s average attendance. The Minnesota Twins are in the American League, shown in *Figure 3a*. We first noticed that they were in the middle of the American League when it came to average attendance. However, when we decided to look at the National League’s average attendance (shown in *Figure 3b*), we realize they must be in the bottom quartile because National League teams have higher average attended games compared to American League teams.

Chart

Description automatically generated

Figure 4a. T-Test Variable Importance Top and Bottom Three Results

Chart, bar chart

Description automatically generated

Figure 4b. Average Weekday Attendance

We ran two t-test to find variable importance. *Figure 4a* presents the three most impactful and the three least impactful variables on attendance. The three variables with largest positive impact on attendance were games in March, games in July, and games on Saturdays. The reason for March having a positive impact on attendance is it is the beginning of the season. People have not watched the MLB since October of the previous year, so they are excited for the season to start. Games in July have a positive impact due to there being more games played and better weather than say spring and fall games. Saturday games have a large positive impact on attendance because of most people having that day of the week off. The three variables with the most negative impact on attendance are games played on Tuesday, Wednesday, and Thursday. Whether due to work or school, Tuesdays, Wednesdays, and Thursdays are extremely busy for people. People in response to this are less inclined to go to a baseball game. Also, there are less people traveling to attend games. As a result, people living in cities where stadiums are located are the main source of attendance for these games. The bar chart in *Figure 4b* shows the dip in average attendance throughout the MLB in the middle of the week.

Chart, waterfall chart

Description automatically generated

Figure 5. T-Test of Team's Impact on Attendance Results

The second t-test we ran was based on teams’ impact on attendance. In *Figure 5*, the top three teams with largest positive impact on attendance are the Los Angeles Dodgers, the New York Yankees, and the St. Louis Cardinals. The results of this t-test were useful later in the process of our project. We did more research into the Los Angeles Dodgers, the New York Yankees, and the St. Louis Cardinals’ promotional strategies to see what they did differently from the Minnesota Twins.

**External Data**

To build a better model and conduct better analysis, we found it impactful to collect external data. The use of collecting external data around the MLB allowed our team to deepen our analysis by identifying insights that may not have been found with the original data provided.

When collecting data, we split the features collected into two categories: data to improve regression modeling, and data to improve exploration and analysis to provide better insights.

The data collected to improve our regression model included stadium capacity, city population, number of professional teams in the city, the season of the game, and previous games attendance. We were able to use this data to better our model, producing more accurate results. For example, running our original model without external data, there were many predictions that’s values were over the stadium capacity level. Adding the feature of stadium capacity helped lessen the possibility of this error occurring.

Data collected for exploratory analysis included whether the game is a league game or not, stadium location, cost to attend a game. Collecting this data allowed us to deepen our exploratory analysis to provide better insights on variables that play a role in effecting variable attendance.

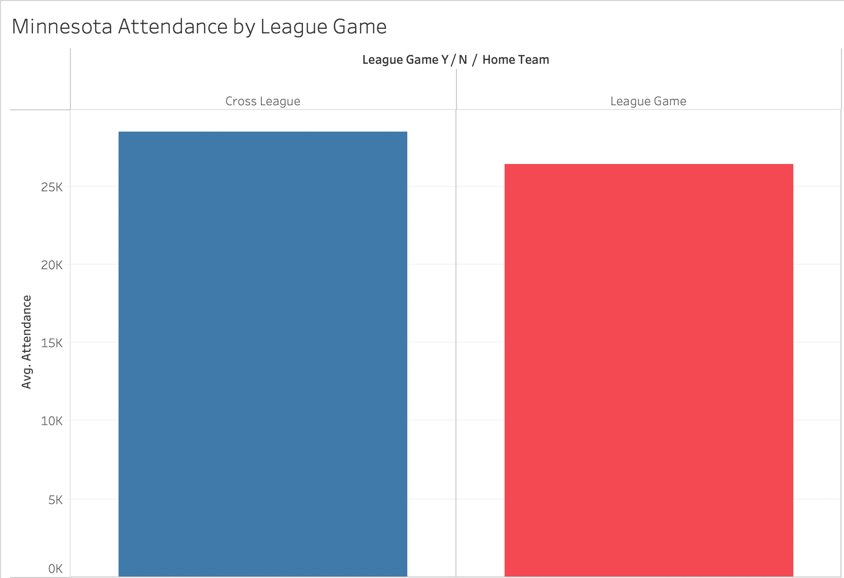


Figure 6. Twins Average Attendance by League Game

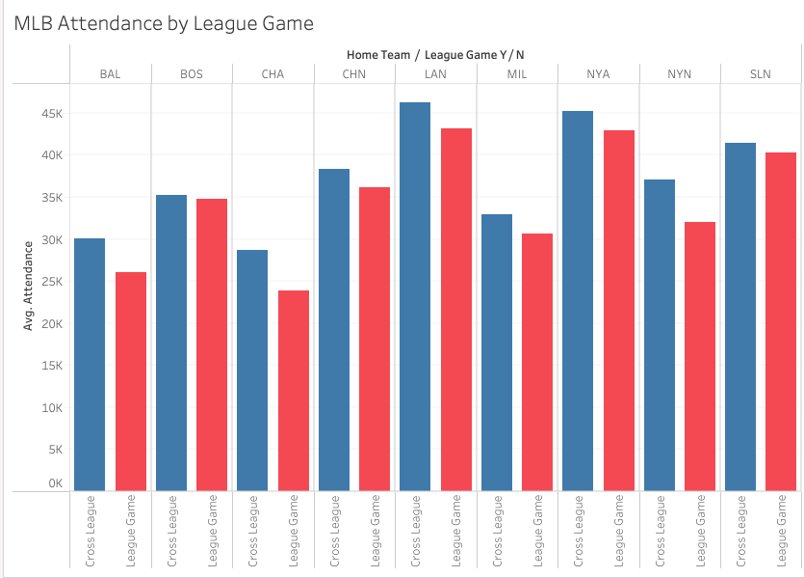


Figure 7. League Game Attendance for Other MLB Teams

The original data set provided displayed variables for each instance of what league the teams playing were in. We performed feature engineering and took those variables to create a new column identifying if the game was a league game or not. After creating that variable, we performed additional analysis on how that variable affected overall game attendance. Above shows two graphs (*Figure 6 and Figure 7*) of the average attendance of teams by whether it was a league game or cross-league. We found that overall, cross-league games averaged slightly higher attendance consistent throughout all of MLB teams. It is also important to note that we see the pattern of cross-league game average attendance being higher than league game attendance no matter if that team has higher or lower attendance levels overall.

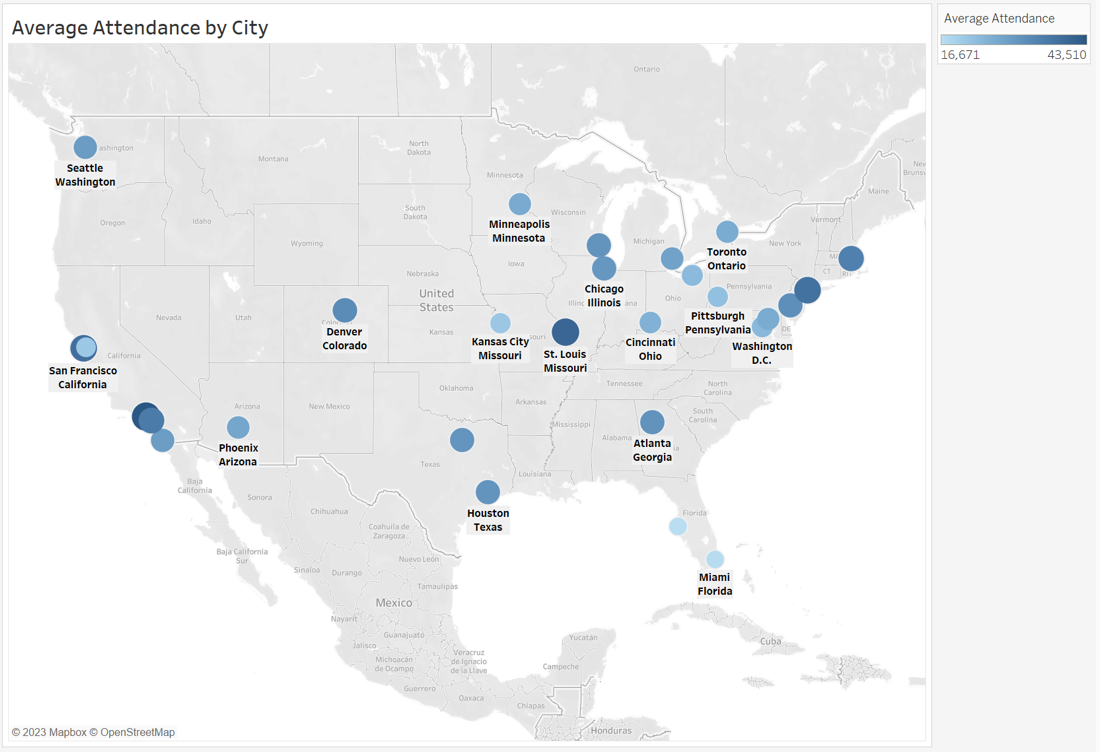


Figure 8. Average Attendance by Stadium City

When looking at each team’s stadium location, we noticed that certain cities have more success at drawing fans to their stadiums than others. This is seen in *Figure 8* above. Most notably, Los Angeles, California and St. Louis, Missouri have high average attendance levels despite being in close proximity to many other MLB teams. On the other hand, St. Petersburg, Florida and Miami, Florida have the lowest averages of stadium attendance in the league, followed by Oakland, California. When comparing these results to the Minnesota Twins, Minneapolis falls in the middle of the average stadium attendance range. There is room for improvement with these results because of the lack of teams and stadiums to the West of Minneapolis. This opens a new market and fanbase that the Twins can cater to in order to draw more people to Target Field. Understanding that the lack of teams near Minneapolis could be a geographical advantage for the Twins is key when making efforts to increase their average attendance levels.

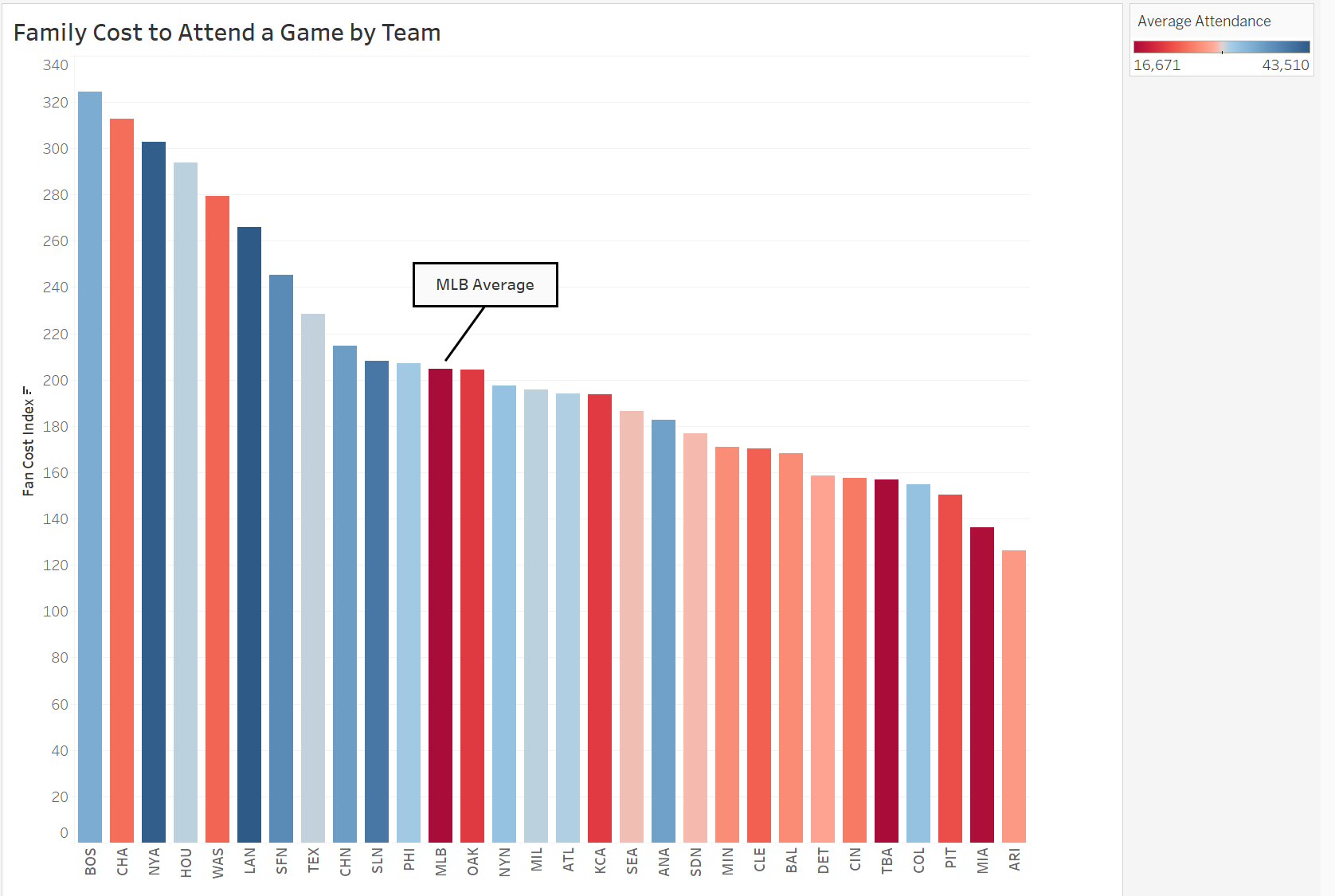


Figure 9. Family Cost to Attend a Game by Team

Another variable that we believed would impact attendance was the cost of attending a game. We found that the cost to attend a game for a family of four varies by team and city. In *Figure 9* the color indicates the team's average attendance, where red is a lower average attendance and blue is a higher average attendance. The Boston Red Sox, New York Yankees, and Los Angeles Dodgers have high family costs to attend their games, yet they still have a large number of fans attending their games on average. Teams like the Chicago White Sox and Washington Nationals have high family costs to attend their games, but they have average attendance numbers that are on the lower end of the range. When looking at the Minnesota Twins, their cost for a family of four to attend their games is $170.89 which is below the MLB average, but their average attendance is also on the lower side. This insight allowed us to understand that we want to make games more attractive for families in order to keep the cost for them to attend low, while improving the average attendance levels for the Twins.

**Promotional Insights**

As mentioned, our group conducted a t-test by each home team in our dataset to identify which organizations were most positively impactful on attendance. Given the wide variety of promotional strategies and approaches across the league, we viewed further research and analysis in this area as an opportunity to explore what strategies highly impactful teams implement and understand what aspects of these strategies have proven to effectively influence home stadium attendance for these organizations. It is important to note that the promotional scheduling data of the following teams only reflects the 2022 season; information was collected from publicly available promotional schedules, as well as announcements from local news articles and other third-party platforms. *Figure 10* presents a snapshot of average attendance by each team’s respective giveaway items.

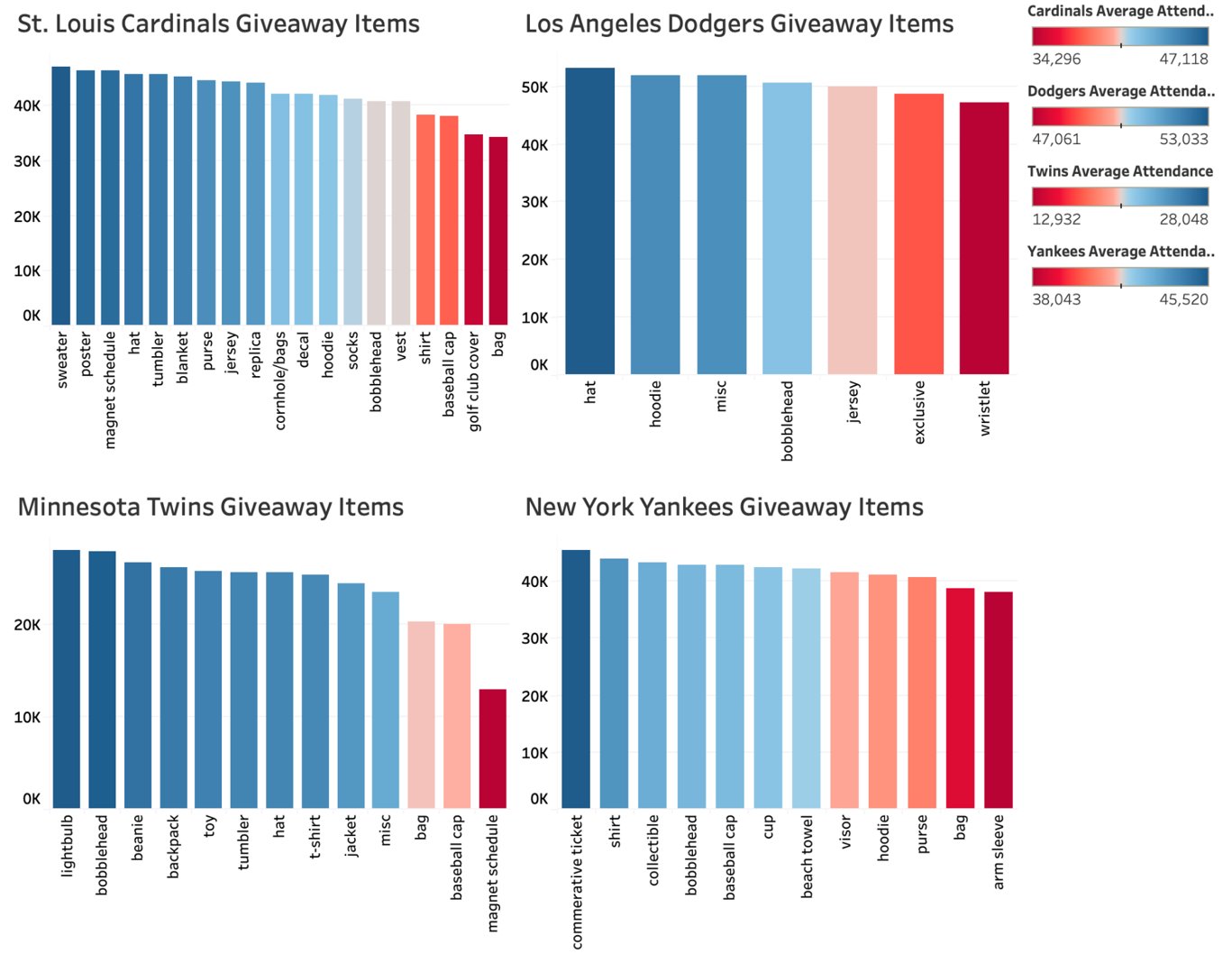


Figure 10. Average Attendance per Giveaway by Team

In 2022, the St. Louis Cardinals employed the most aggressive approach to promotion placement with the most giveaways across 40 total games, and 30 events such as pre/postgame programs, performances, tributes, ceremonies, and partnerships. There were 26 games in which they utilized multiple-item giveaways to incentivize sales of exclusive ticket bundles. Bobblehead giveaways were surprisingly underwhelming, as average attendance was higher at games without bobbleheads. The frequency and diversity of items reveals the effectiveness of several unique personal and clothing items instead, as shown in *Figure 10*. Moreover, child-specific giveaways display lower avg. attendance than 16+/adult targeted giveaways *(Figure 11).* It is also important to note that all giveaways for children are limited to 10K as opposed to 25K for 16+ young adults. From a larger perspective, *Figure 12* indicates a positive correlation between attendance and item quantity limits.

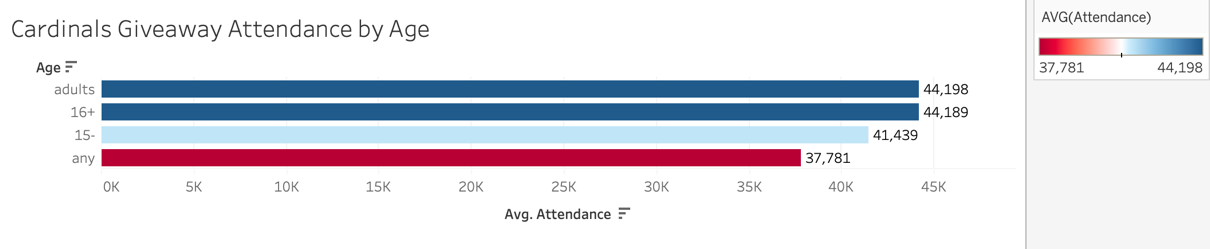


Figure 11. Average Attendance by Age - Cardinals

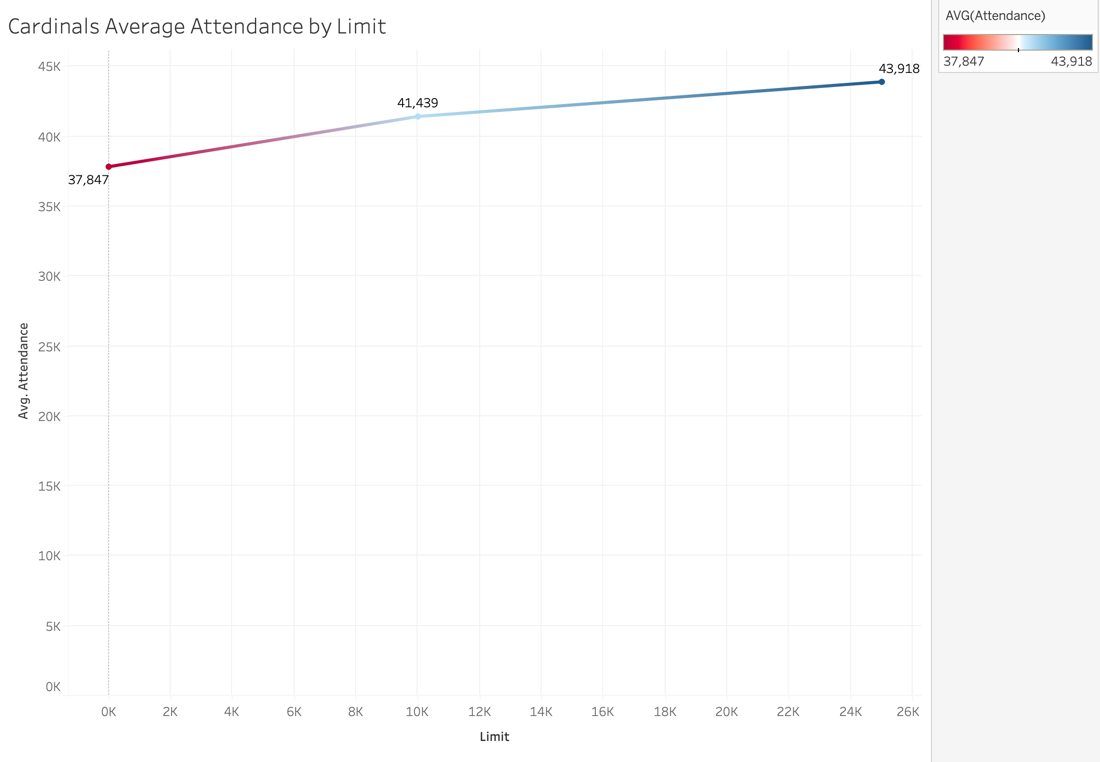


Figure 12. Average Attendance by Quantity Limit – Cardinals

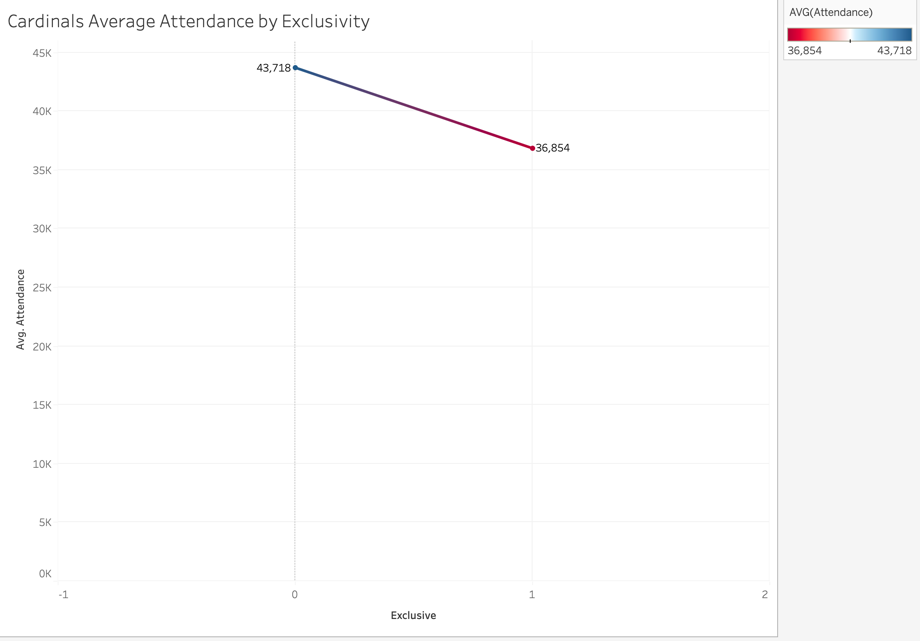


Figure 13. Average Attendance by Giveaway Type – Cardinals

On the contrary, the Los Angeles Dodgers do not follow the use of age restrictions as part of their giveaway criteria, as each giveaway opportunity was available to any guest in attendance. Moreover, the organization uniquely embraces loose giveaway limits. Specifically, giveaway quantities are either limited to the first 40,000 guests or not limited at all; surprisingly, average attendance is marginally higher for giveaway with no-limit giveaways (40,688) than the use of a large limit in this fashion (38,324). Throughout the duration of the season, twenty-six total giveaways were provided, including thirteen exclusive giveaways that all require a special ticket bundle. We see lower attendance averages with these exclusive giveaways compared to standard giveaways, as shown in *Figure 13;* most notably, games in which there is no giveaway opportunity whatsoever still slightly outperform exclusive giveaways.

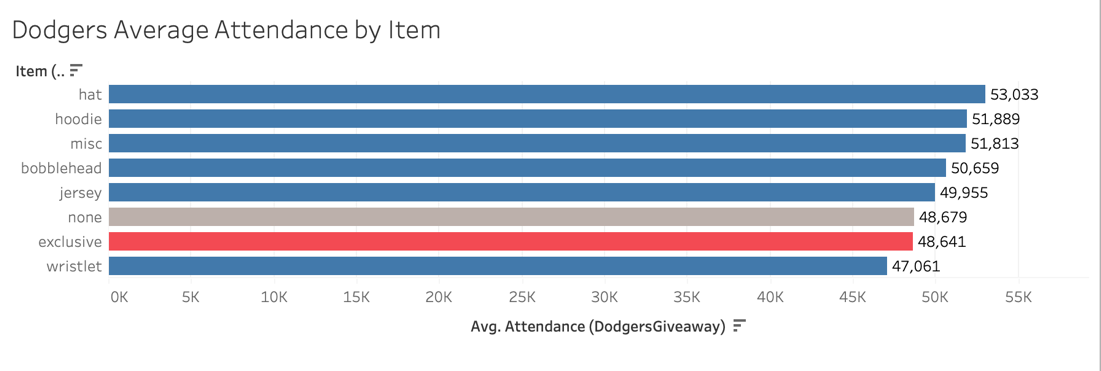
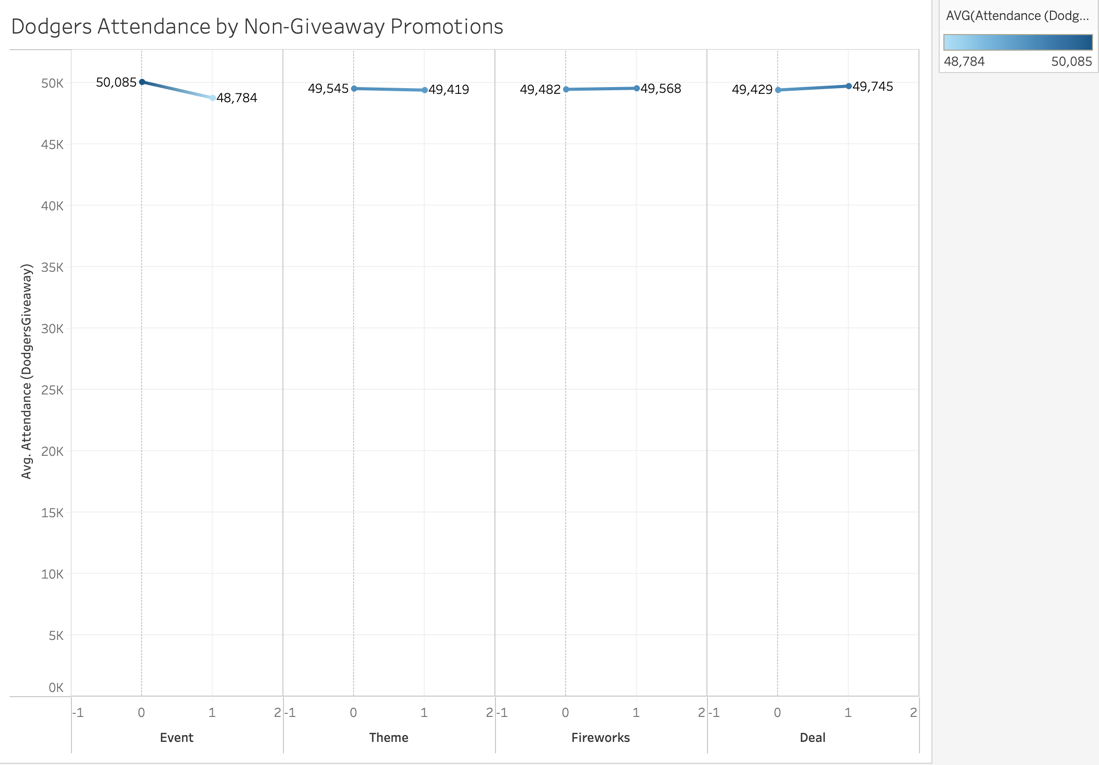
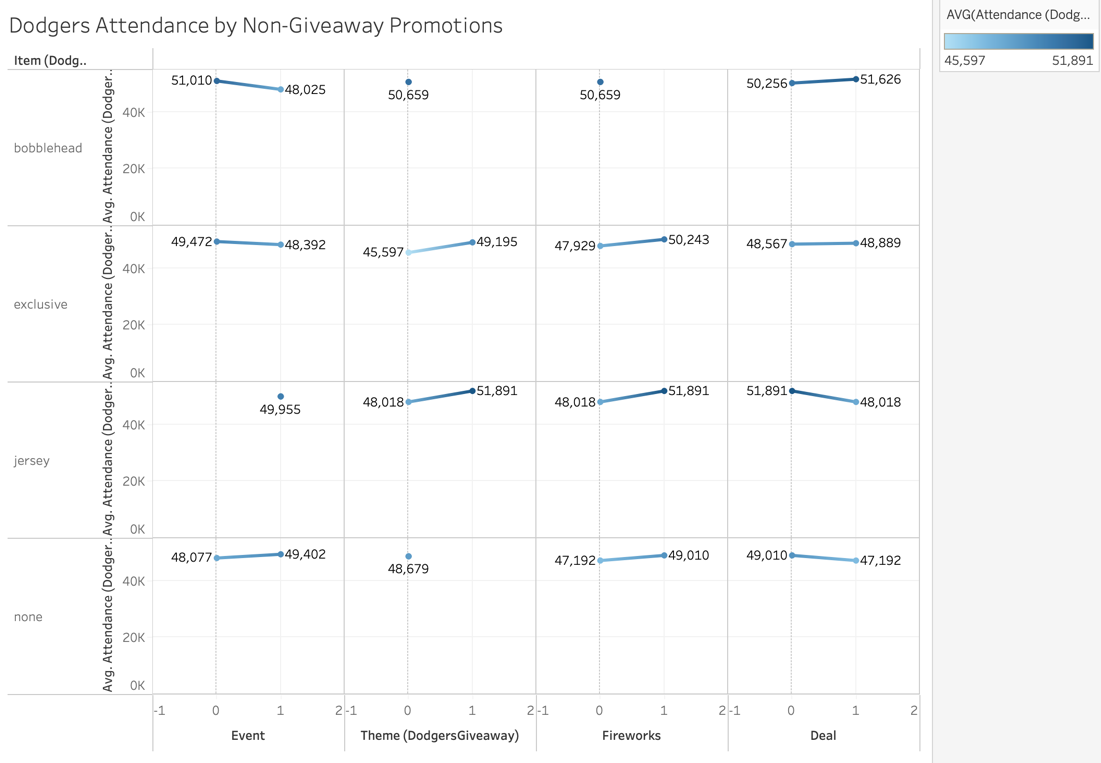


Figure 14. Average Attendance by Giveaway – Dodgers

What stands out in the Dodgers 2022 promotional schedule, however, is that it is far more entertainment-oriented than other teams we assessed in this context, with 13 Taco Tuesdays, 13 Friday Night Fireworks, and 6 Viva Los Dodgers nights; the Dodgers additionally provided a plethora of other promotions, such as heritage nights and other special celebrations, that go beyond a simple game theme. *Figure 14* displays average attendance for Events, Game Themes, Fireworks, and Deals across the promotional season. Observably, the placement of game themes, fireworks, and food deals consistently have a positive impact on attendance, while event placement surprisingly does not. *Figure 15* provides a more detailed picture of average attendance by item when placed alongside other promotional activities. It is worth noting that some giveaways (hat, hoodie, wristlet, miscellaneous) were excluded from this figure, as these instances were placed individually and were not placed alongside/impacted by other events.



*Figure 15. Average Attendance by Promotional Activity – Dodgers*



*Figure 16. Average Attendance by Promotional Activity by Item– Dodgers*

On the other hand, the Yankees promotional approach provides a unique perspective on how to better market promotions. Specifically, they strategically market "collectible" items to strengthen the placement of dedicated events, such as commemorative opening weekend tickets and replica statues/plaques for special ceremonies (e.g., jersey retirements and player celebrations). Additionally, they modify standard items by integrating game themes so to offer limited edition/exclusive giveaways, such as Baby Yoda and Scooby-Doo bobbleheads, at no extra cost to attendees. Regarding age criteria, average attendance is similar between adult-specific and general giveaways, with child-targeted items exhibiting the lowest average attendance. Uniquely, quantity limits reveal attendance fluctuations between each limit level (0, 5, 10, 18, 25, 40), though they still hold a relatively positive correlation with one another, as shown in *Figure 18*. Overall, the Yankees frequently strive to provide promotional opportunities that are more reflective of the game experience, rather than providing promotions that simply supplement a game.

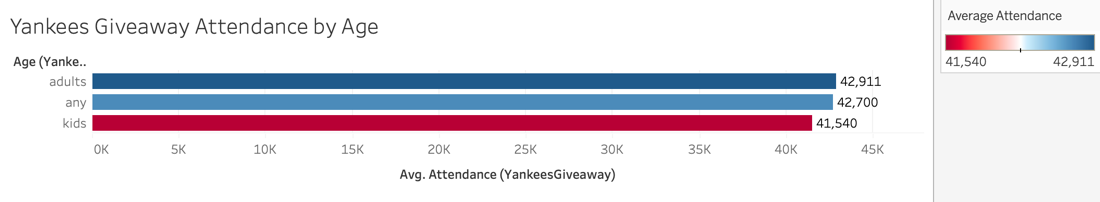


Figure 17. Average Attendance by Age – Yankees

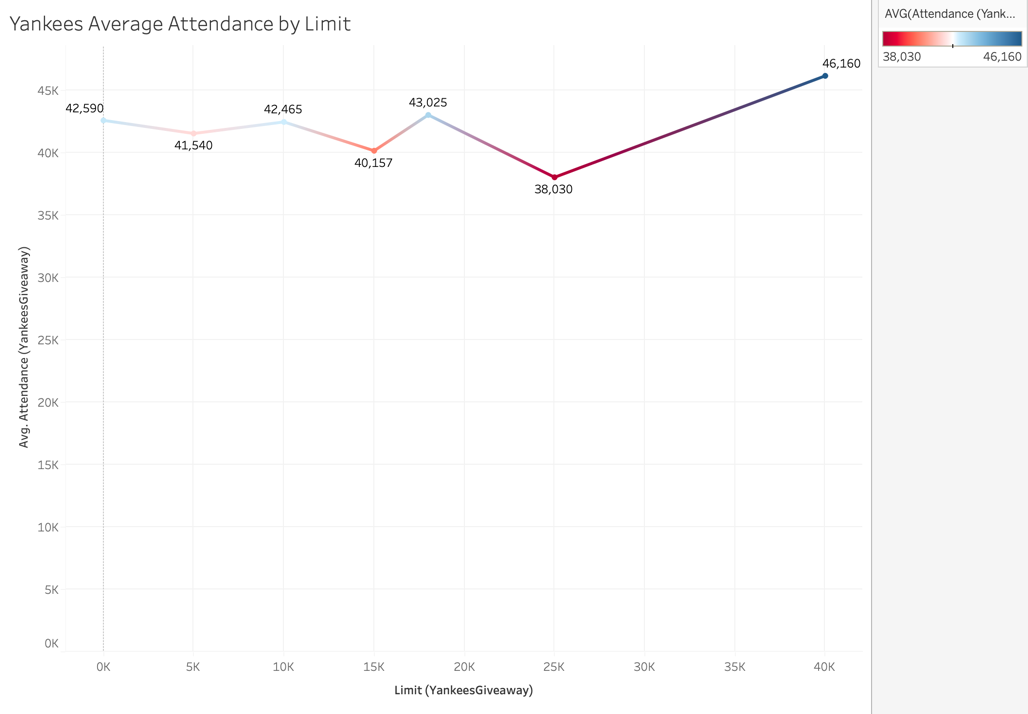


Figure 18. Average Attendance by Quantity Limit– Yankees

Despite not being ‘highly impactful’ on attendance, there are many observable insights and trends that can be derived from promotional strategies pursued by the Minnesota Twins. *Figure 19* provides a snapshot of the organization’s observed giveaway opportunities through the 2022 season. Note that some items are not reflected in this dataset, as games in which these giveaways were scheduled were cancelled/postponed and/or attendance for these games was unavailable. One characteristic of Twins promotional scheduling is that games with child-specific giveaway items observe higher attendance than traditional any-age giveaways *(Figure 20)* which is not observable among any of the other teams we explored. Additionally, attendance exhibits a positively correlated relationship with quantity limits, though giveaways limited to the first 10,000 fans observe considerably lower attendance than at any other level (Figure 21).

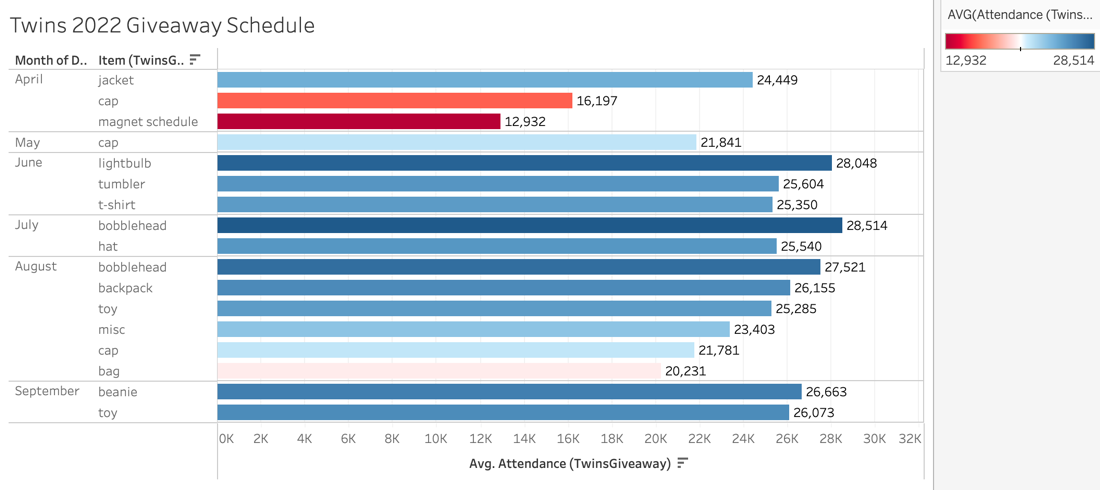


Figure 19. Average Attendance per Item by Month- Twins

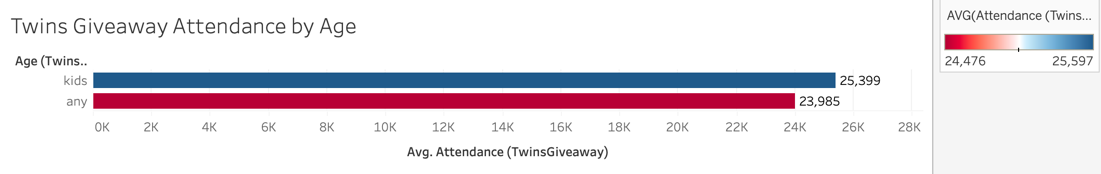


Figure 20. Average Attendance by Age - Twins

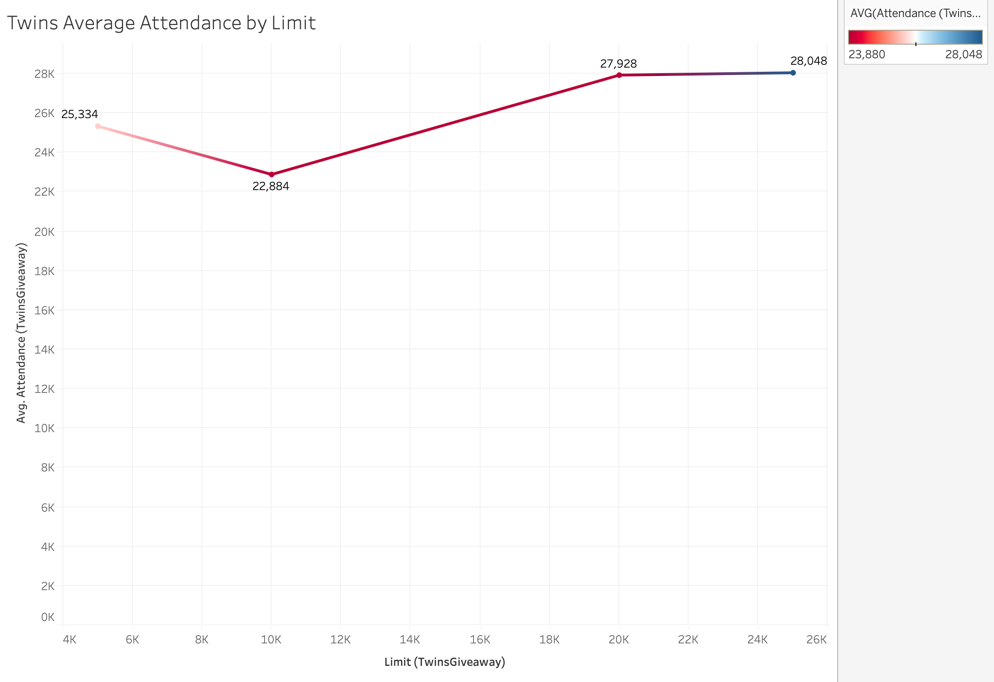


Figure 21. Average Attendance by Giveaway Limit Quantity - Twins

As mentioned, many aspects of promotional success are unique to each of the teams discussed above; item preferences are subject to be influenced by location/weather, giveaway criteria, and the placement of other events. Despite the wide variety in promotional strategies between organizations, there are observable trends in several areas which the Twins can further explore to enhance the strength of promotional scheduling in a way that will positively impact attendance.

Creating a sense of exclusivity using ‘limited edition’ items, either collectible or uniquely modified, may promote higher demand for these items and, subsequently, increase attendance. However, it is essential to pursue this without the use of strict/smaller quantity limits nor truly ‘exclusive’ giveaways, as neither of these strategies are particularly effective in increasing attendance.

Similarly, teams can partner with local businesses and offer co-branded giveaway items, which can promote both the team and the partner. Collaboration with external organizations and businesses can be expanded to deals and events as well. For example, the Cardinals utilize local restaurants to promote free ice-cream and other food deals before games, and the Dodgers and Yankees utilize a wide variety of sponsors to help further promote special occasions.

In the context of promotional marketing, opportunities are frequently placed in a manner that specific fan groups, such as children, women, families, and veterans, are particularly receptive to. Understanding key fan demographics provides a great opportunity to enhance the effectiveness of promotional placement, allowing for a more targeted approach in how promotions (items, themes, appreciation games, and celebrations) are placed so to effectively capitalize on these groups.

**Modeling – Results and Findings**

After gathering insights and making recommendations we wanted to utilize all that information in our regression models to predict the 2023 home game attendance for the Minnesota Twins. Each model we ran was trained on 20 years of data and tested on the 2022 season. We evaluated our models using Mean Absolute Error (MAE), where the smaller the difference between the Test MAE and Train MAE, the more accurate the model. Then we utilized Mean Absolute Percentage Error (MAPE) when comparing model effectiveness, where the smaller the MAPE percentage the more accurate the model predictions. After running a total of eight regression tests, we determined that the Random Forest – all predictors without cross validation model had the strongest results.

We first ran the model with the original data that was provided to us in the challenge. This produced an MAE difference of about 89, but our MAPE was 152.3%. This indicates that the model predictions are not accurate. After examining this further, we noticed that most of the model predictions were too high, with some being over three times the actual attendance recorded for that game. To improve the accuracy of our model we added external data to our input variables. This included stadium capacity, city population, number of professional teams in the city, and season. Doing so raised the MAE difference but decreased the MAPE to only 80.3%. While this number was still not the accuracy level we wanted, it was an improvement on the previous model. Our next model was run with the same external data as before, but we only trained the model on the previous five years of data. This lowered the MAE difference and lowered the MAPE percentage to 62.9%. Overall, the accuracy of our model was improving, but we were striving to reach an MAPE that was less than 25% so that our model produced accurate predictions with low error. This was achieved through our dynamic model results. The dynamic model we ran included all the external data from before, but we also added a column for previous home game attendance. Doing so lowered our MAE difference, but most notably it lowered our MAPE percentage to only 21.2%. These results indicate that our dynamic model produces the most accurate predictions and is the best of the four models we ran.

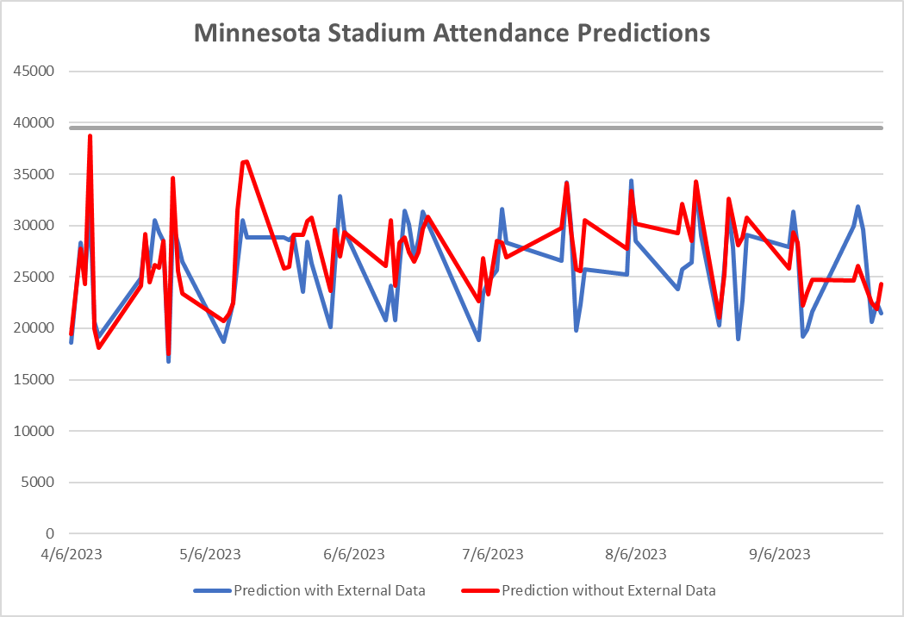


Figure 22. Predictions without Previous Attendance with Line for Capacity of Target Field

Comparing model results is also another way to visually see trends within attendance and can be useful for evaluating model effectiveness. When looking at our model results in *Figure 22*, there is a lot of overlap, but the model without external data produced a larger range of attendance values while the model with the external data had a smaller range of values. The model with the external data also produced lower dips in attendance in the middle of the season when compared to the initial model. We also thought it would be useful to add in a Target Field capacity level line so that the viewer can visually see how full Target Field would be based on the model predictions.

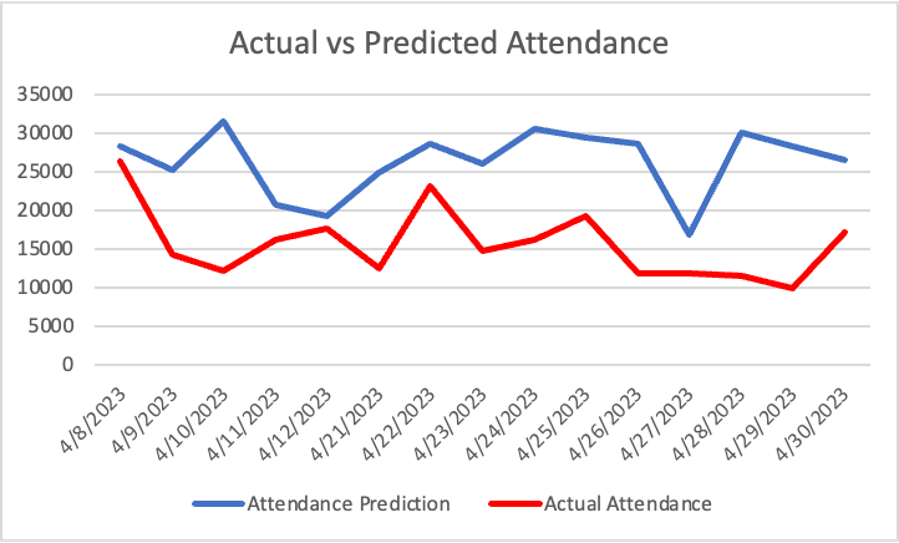


Figure 23. Random Forest Model with External-Integrated Data

Next, we wanted to compare our model results with the actual attendance levels of the games that the Twins have played so far this season. Our model that included the external data seemed to predict attendance numbers that were higher than the actual attendance for each of the games this season. In *Figure 23*, the most notable difference was on April 10th and our closest prediction came on April 12th. Something that our model did not consider was holidays. With April 9th being Easter, we think this could be the explanation for our model spiking on the 10th, while the actual attendance dips.

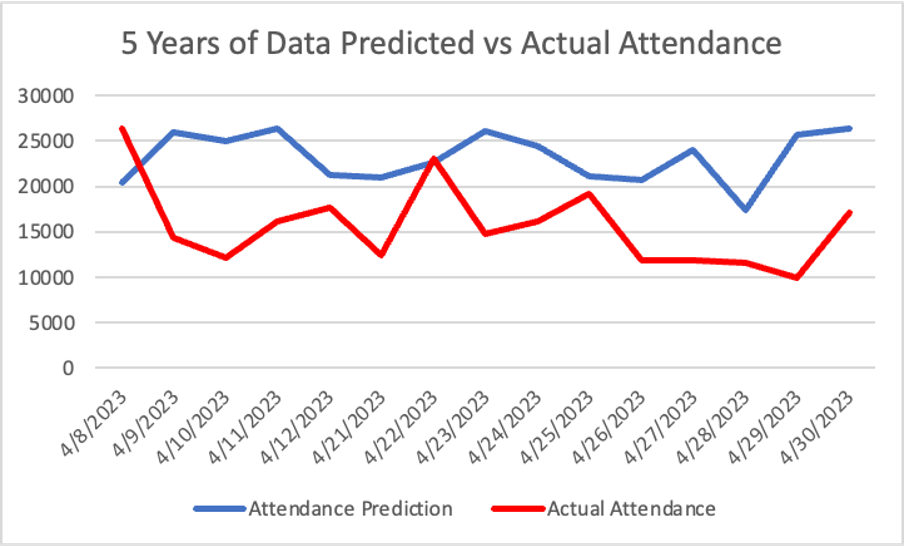


Figure 24. Five-Year Subset of Random Forest Model with External Data

The model that included external data but was only trained on the previous five years of data produced better predictions than the model that included external data and was trained on the previous 20 years. This is seen in *Figure 24* where the values were much closer to the actual attendance and there was some overlap between the predictions and actual values. From this model, our closes prediction came on April 22nd and April 25th.

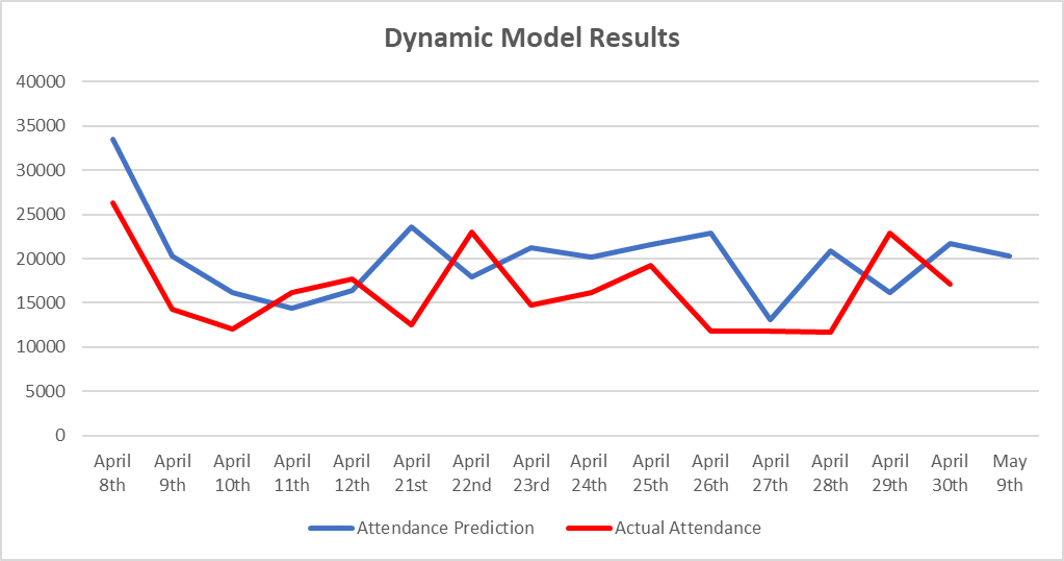


Figure 25. Random Forest Model with Previous Attendance

Lastly, when comparing the dynamic model predictions to the actual attendance we found that the results were the most accurate and aligned with the model’s low MAPE. In *Figure 25,* the lines follow the same trend, especially at the beginning of the season. The model also consistently produced close predictions instead of having only one or two close predictions like the previous models. Our April 12th and April 27th predictions were the most accurate, being only about 1,200 people off each.

Building and running regression models allowed us to examine what inputs were most important for producing accurate attendance predictions. Our findings indicated that the dynamic model produces the most accurate results and would be the most useful model for the Twins going forward. Being able to utilize this model at any point during the season to predict the next home game attendance can help guide when to implement certain promotions and can serve as a guide for many organizational decisions.

**Recommendations**

The overarching goal of this project is to improve the Minnesota Twins stadium attendance. We believe the best way to do so is to identify specific areas of weakness along with variables that negatively impact attendance to provide specific recommendations that target them. Doing this can help flatten out any dips that occur in attendance.

The first area we identified that negatively impacts attendance rates are the months of April, May, and September. As the months of April and May mark the beginning of the season, it is important to build fan excitement to draw more fans to games. To do this, the Twins can host season opener events and advertise these games more. In our promotional insights, we found that the Twins offer most of their promotions and giveaways in the summer months. We recommend revising the promotional schedule, offering more giveaways, and holding more theme games during these beginning months to increase attendance. For the month of September, school is starting up and fall is starting. The Twins can help increase attendance in September by holding back-to-school days or fall-themed events to draw in more fans. An additional way to increase attendance in these months is to offer ticket packages. This would include selling groups of tickets in these months specifically at a slightly lower rate.

After running a T-test looking at all variables, we found that weekdays including Tuesday, Wednesday, and Thursdays negatively impact attendance the most of all variables. There are many ways the Twins can help change this to bring in a larger crowd during these days. To begin, the Twins can expand and place a focus on concession deals for these days. For example, hold happy hour deals for night games. Continuing with food, Twins can partner with local food businesses in the Twin City area to offer their food at games. To increase weekday game attendance, the Twins can host various companies to Target Field and hold more corporate events to draw in a crowd. If people attend a game for work, and have a positive experience, they are more likely to come back on their own. In analysis, we found that kid targeted event performed well overall. During the summer months, the Twins can offer Family summer themed nights to draw in whole families, increasing attendance.

When evaluating the effectiveness of promotional strategies among highly impactful teams, there are several crucial factors to consider for future promotional scheduling. For the Twins, it is imperative to restructure their promotional schedule to ensure an even distribution of giveaways throughout the season. The current schedule for the 2022 season is heavily concentrated in July and August, resulting in lower attendance during the early summer months and a surge in attendance towards the end of the regular season. To address this, the Twins should prioritize the inclusion of bobbleheads as a regular promotional item. Bobbleheads consistently attract higher attendance than other items, and the team should offer a wider variety of both active and retired players, along with an increased number of giveaways throughout the season. In addition to these strategies, the Twins should also consider incorporating more family-oriented promotions. Child-specific giveaways and all-inclusive promotions have proven to be particularly effective in drawing higher attendance. By implementing these approaches, the Twins can enhance the effectiveness of their promotional placement and maximize the potential positive impacts their scheduling can have on attendance.

When it comes to promotional trends across the league, exclusive giveaways that require special ticket packages do not effectively boost attendance. Therefore, if exclusive promotions are to be pursued in the future, it is advisable to make them supplementary to other promotions to avoid the pitfalls of standalone exclusive promotions. Moreover, incorporating elements of game themes and other pre/post-game events into giveaway items is highly encouraged as it adds extra value to the game experience and widens the range of promotional opportunities. Ultimately, the most effective promotions are those that provide value to attendees by complementing the game experience without incurring additional costs. It is crucial for the Twins to ensure that there are low or no barriers for all fans to take advantage of these opportunities, regardless of the type of promotion. By doing so, the Twins can enhance their fan engagement and increase attendance.

In examining the attendance trends across spring, summer, and fall, it's evident that spring games attract the lowest attendance. To combat this issue, the Twins can implement ticket bundle packages or targeted discounts to incentivize fans to attend spring games. Moreover, incorporating more promotional events during early-season games can be an effective strategy to boost attendance during this period.

When comparing attendance trends between league and cross-league games, it is clear that league games draw lower attendance figures. To address this, we recommend adding more events and themed games specifically tailored to league games. Additionally, the Twins can foster rivalries among teams within the same league, which can further enhance the game experience and attract higher attendance figures.

We found that the cost of attending games is a significant barrier for fans who want to attend more games. To alleviate this concern, the Twins can offer concession discounts, including food and beverage tickets and dinner packages. By providing more affordable options, the Twins can encourage more fans to attend their games and increase overall attendance figures.

Implementing these targeted recommendations can effectively increase stadium attendance for the Twins. By addressing the specific weaknesses in their seasonal attendance figures and promotional strategy, the Twins can improve the game experience for fans and strengthen their brand.