Fan Impact on Home Court Advantage
Collin Travasos

Across all sports and competition, location plays a factor. Having home field or in this case home court advantage is an undeniable part of sports. Organization's are constantly looking for ways to improve their team's success, with the primary focus being on the players and coaches. The product that is out on the floor is going to be the greatest factor of the team's success. This report aims to look at the small percent that the front office can change to impact the game outside of the team's personnel – the fans.

From a business perspective, fans are typically thought of as a stream of revenue, not a factor in the game that can be tweaked to improve the chances of victory. Most think that fan attendance is a result of team success, and while that is likely true, team success can also be impacted by fan attendance. The total number of fans is not the only factor either, it's what those fans are doing and how loud they are cheering that can influence the game. The "quality of fan" is just as important as how many fans are in attendance in regards to the impact that they can have on the outcome.

This project dives into the relationship that the fans can have on their team's success. More importantly, it seeks to determine if and what the relationship between the total attendance and who is in attendance. The hypothesis is that fans who pay more for tickets do not cheer as loudly – and do not impact the game as much – as the fans in the cheaper seats. There have been studies conducted on how many fans are in the arena and how that can change the game, but none on how ticket costs can have an impact. This project aims to find, in short, does the cost of attendance affect the outcome of a game.

The inspiration for this project stemmed in part from the Golden State Warriors relocating from Oracle Arena across the bay to the brand new Chase Center. Oracle, located in Oakland, was renowned for its passionate fans. The arena was easily accessible by public transportation and by car to all members of the greater Bay Area. It was right next to the Colosseum, home of the Raiders and A's, who have also since left the city of Oakland. It was by no means a nice arena, nor in a nice part of town, but tickets were cheap and the fans were loud. Prior to the rise of the splash brothers it was coined "Roaracle" during the We Believe days of the early 2000s. Since moving to downtown San Francisco, ticket prices have skyrocketed and the makeup of the fans has changed. It is now a hub for venture capitalists and tech startups rather than families and die-hards. The fans have gone from wearing t-shirts and jerseys to dress shirts and patagonia vests. It is noticeably quieter. Does this affect the game? One would think.

Using this thought process one could hypothesize that over the years small market teams would have a stronger home court advantage than big market teams. If this is true, the Clippers should have a better home court advantage than the Lakers. They share an arena but Lakers tickets are notably more expensive simply because they are the Lakers. Portland, Indiana, and Utah should theoretically have a stronger home court advantage than Miami, New York, and Chicago to name a few. These large market teams do have a higher demand for tickets and as such prices will be higher. But the question remains, are the most passionate fans unable to attend because of price gouging.

The data used for this project was a teams yearly gate receipts (total ticket sales), total season attendance, and season results (ie total, road, and home wins and losses, etc). The gate receipts were obtained from statista and the season attendance and results were scraped from espn. This data was collected from 2011-2024, with the covid season thrown out. These three data frames were manipulated, adjusted, and combined within r studio in a single data set. After wrangling the needed data, it ended up being 355 observations of 28 variables.

Once collected, a handful of models were run. Linear regressions were run looking at the relationship between home win percentage, attendance, and ticket sales and the various combinations of those variables. Those are useful because they are simple and easily interpretable for looking at the relationship between variables. They are however limited to linear relationships and struggle with complex relationships. Next, logistic regression looked at ticket sales and attendance on the chance a team had at winning an individual game as opposed to their record. This is also simple and easily understandable, especially for those with limited technical backgrounds, however they can underperform somewhat if patterns are not linear. Additionally, a random forest model was used to predict single game results and an XG boost model was used to predict home game win percentage for the season, both based on attendance and gate receipts. These are better for non linear relationships. They are both good at handling overfitting and are helpful with large, complex data sets but struggle if parameters are not properly set. Furthermore they can be difficult to interpret and put into practice. Graphs for the models run are located at the end of the report.

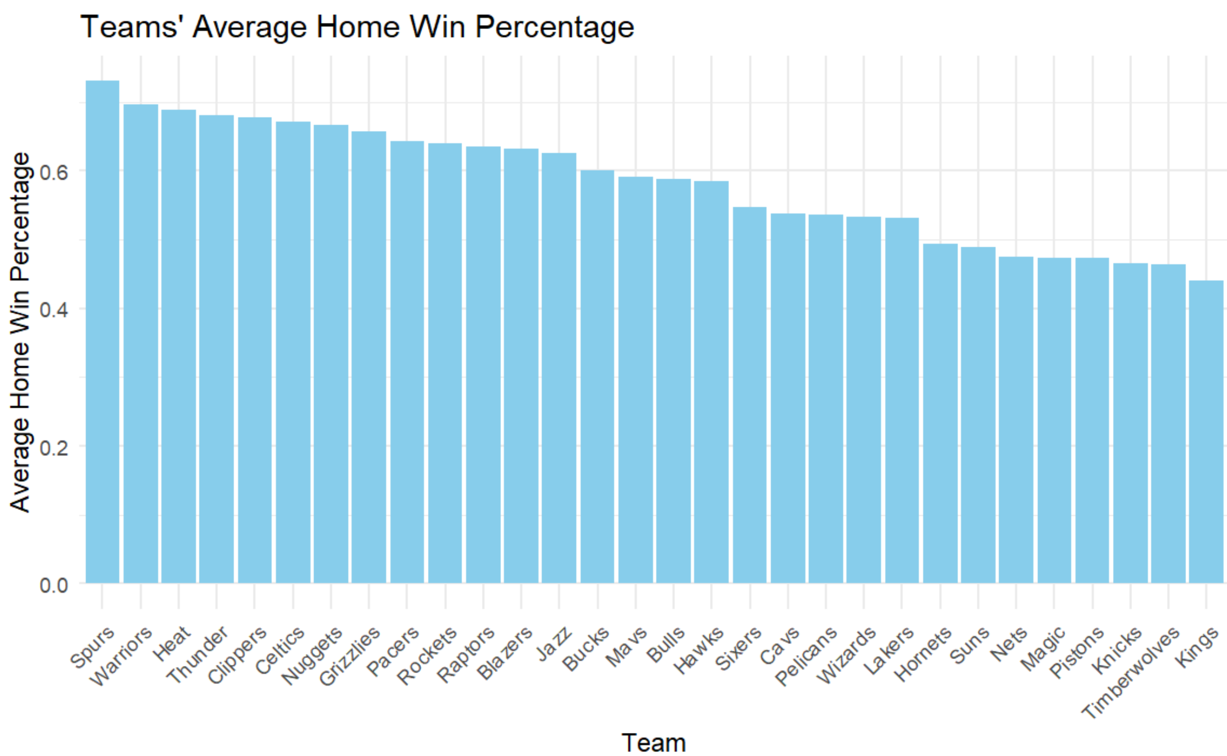
The regressions were moderately successful. Both the linear and logistic models gave statistically significant estimates to the fan attendance on win percentage and likelihood. This confirms that attendance does have a linear relationship with winning. For every person that comes into the game, these models estimated a small but significant increase in the team's performance. The gate receipts were also significant in these models. For every dollar brought in there was a bump in wins. However when both ticket sales and attendance were included, attendance was significant and ticket sales were not. This is likely due to the strong correlation between tickets sold and ticket sales.

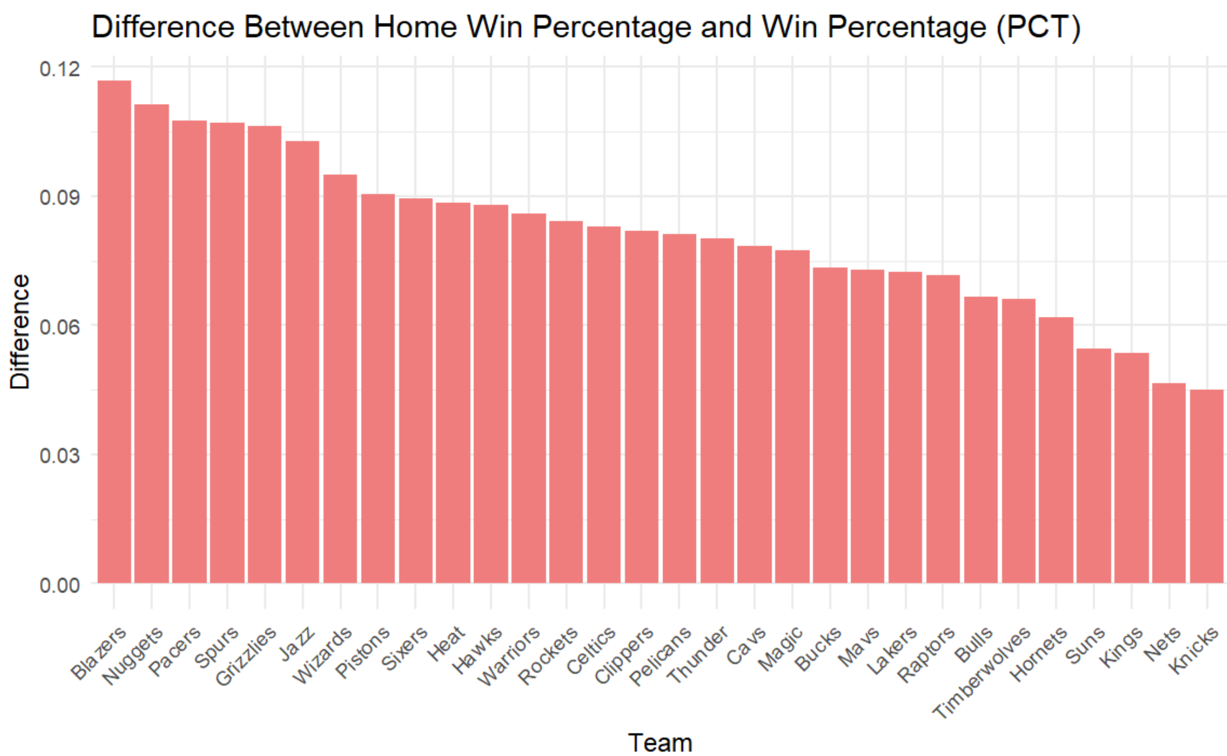
The random forest technique was used a handful of times with moderately intriguing results. First, models were ran with attendance, gate receipts, and attendance plus gate receipts. These gave an OOB estimate of error rates of 38%, 34%, and 31% respectively which is higher than one would like. The dataset was then split into training and testing sets, with 80% included in the train set and the remainder in the test set. Using the most successful of the prior three random forest models, attendance and gate receipts were used to predict the chance of victory. This also did not yield fantastic results with an accuracy of .67 which is less than the no information rate of .70. The p value of .67 indicates low significance and the kappa of .18 is also low. The model struggled with predicted losses, and predicted far more wins than the test set had.

The XG boost did not prove much better when looking at the binary win-loss outcome. The accuracy and no information rate were both .70, the p value was extremely low, and the kappa value was 0. In fact, the model did not predict any losses, which is not very helpful. However, using an XG boost regression model gave a promising RMSE of .16. These are far from what one would hope for and could benefit from more adjustment, but still hold value within the scope of this project.

With the results of these models, one can confidently conclude that teams that win more have more fans and bring in more revenue. People want to see good teams and are willing to spend more to do so. This makes sense and is not revolutionary by any means. Unfortunately using these models it is not possible to accurately predict how a team will do in an individual game or over the course of the season. It is also not possible to determine if teams have extra success with additional fans and revenue or if it is the other way around. There is undoubtedly correlation between them, but there is no clear causation one way or the other.

Even though the predictive models were not particularly successful, the original hypothesis was corroborated by the data. The first of the following graphs show teams ranked by home win percentage and the second shows the difference between home and total win percentage. In the first graph, home court success is displayed, but it is heavily impacted by how good the teams are. The Spurs, Warriors, and Heat have all been great franchises, and the graph is nearly identical to that of total win percentage. The second visualization reveals the teams that have a true boost at home. The teams towards the top are not the most successful teams, rather they are the teams who have the biggest difference at home compared to on the road, or in other words, home court advantage. These teams are, as hypothesized, small market organizations with traditionally cheaper ticket prices. Conversely, the teams at the bottom are big market teams with higher ticket costs.


Teams' Average Home Win Percentage
Average Home Win Percentage
0.0
0.2
0.4
0.6
Spurs
Warriors
Heat
Thunder
Clippers
Celtics
Nuggets
Grizzlies
Pacers
Rockets
Raptors
Blazers
Jazz
Bucks
Mavs
Bulls
Hawks
Sixers
Cavs
Pelicans
Wizards
Lakers
Hornets
Suns
Nets
Magic
Pistons
Knicks
Timberwolves
Kings
Team

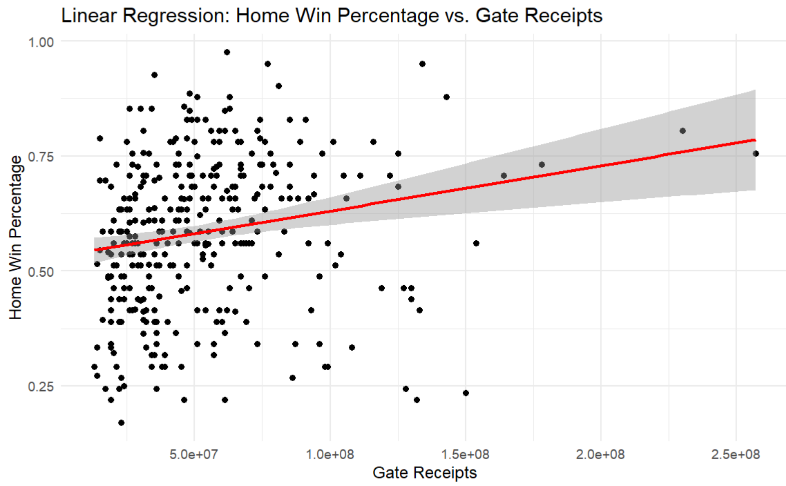

Difference Between Home Win Percentage and Win Percentage (PCT)
Difference
0.00
0.03
0.06
0.09
0.12
Blazers
Nuggets
Pacers
Spurs
Grizzlies
Jazz
Wizards
Pistons
Sixers
Heat
Hawks
Warriors
Rockets
Celtics
Clippers
Pelicans
Thunder
Cavs
Magic
Bucks
Mavs
Lakers
Raptors
Bulls
Timberwolves
Hornets
Suns
Kings
Nets
Knicks
Team

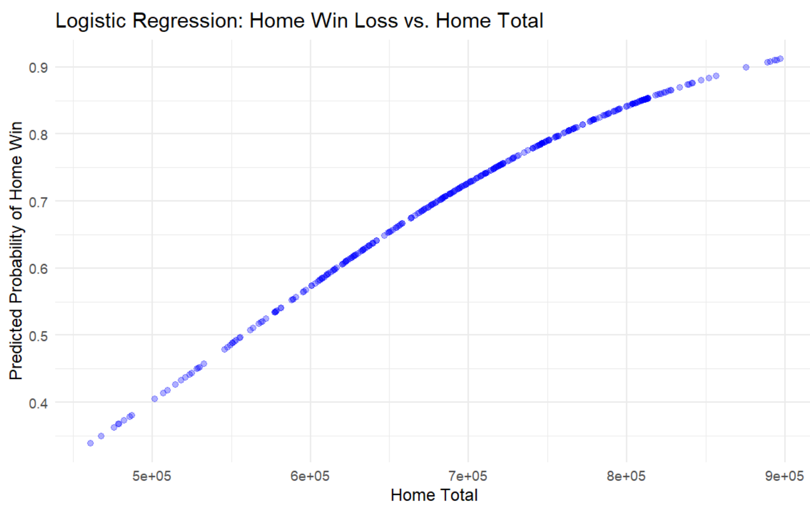
It is important to remember that, regardless of the findings of these models and this project, the primary goal of a team is not necessarily to win games. Ultimately, the team is a business and is looking to make money. You could argue that winning games brings in more money, but the team does not set ticket prices with the sole goal of helping win games, they also do so to bring in the maximum revenue. So even if this project provided indisputable evidence that lower ticket prices combined with higher attendance positively impacted games, teams may still opt to keep or raise their ticket prices to meet financial goals.

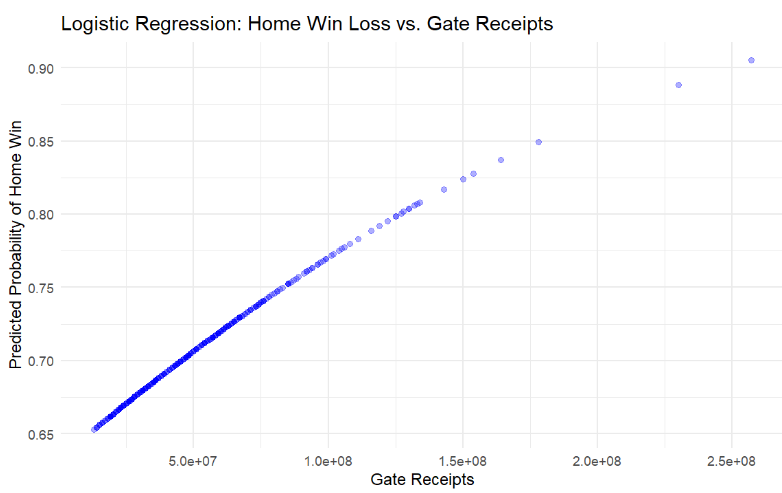
Given more time, and more importantly more information, this project has the potential to yield legitimate actionable results. Ultimately, using season averages was too broad. The scope of the project made it challenging, but these models did not have enough data points to properly predict outcomes. Using unique game attendance and get in price would undoubtedly improve the models. This was the original plan but the data was unobtainable or behind a paywall. Additionally, as one might expect factoring in the opponent would prove useful to the prediction abilities. On the other side of the equation, it would be interesting to look at other statistics as opposed to just wins and losses. Home team stats like points and shooting percentage as well as away team stats like turnovers and free throws would likely prove noteworthy. All that included, it is still possible that results could yield very little. Home court advantage could be primarily impacted by travel distance, or altitude, or even the comfort of players getting to sleep in their own beds. Playing on the road brings more difficulties than just fans. However, with the findings shown in the data, and the teams that appear to have the best true home court advantage, it does not eliminate the possibility that the original hypothesis is accurate and teams with cheaper ticket prices do in fact have a better home court advantage.

Regardless, most can agree that ticket prices to all sporting events are rising and limiting the fans that can come. The goal of this project was to give teams a reason to lower ticket prices. We can all agree that all fans should have the opportunity to watch their favorite teams play. While it is not reasonable to confidently say that lowering ticket prices will help the teams win, teams should still lower prices even if for no reason other than community engagement.

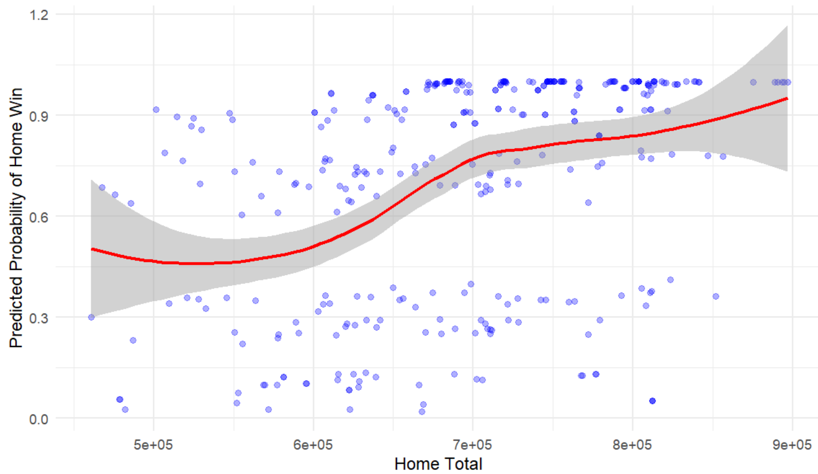
Model Graphs:



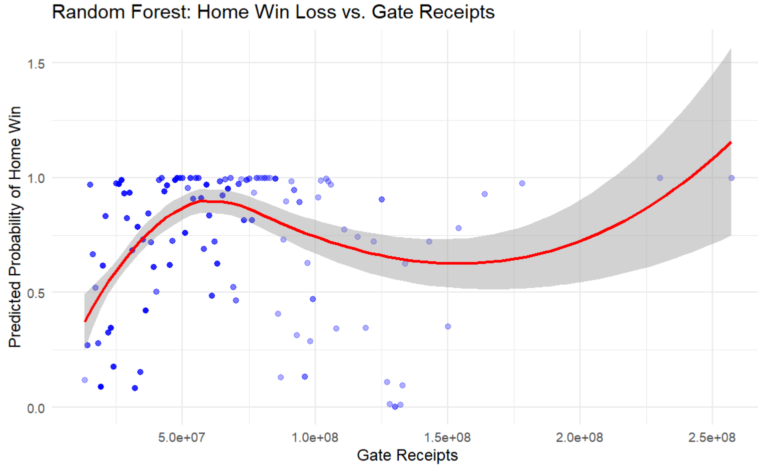




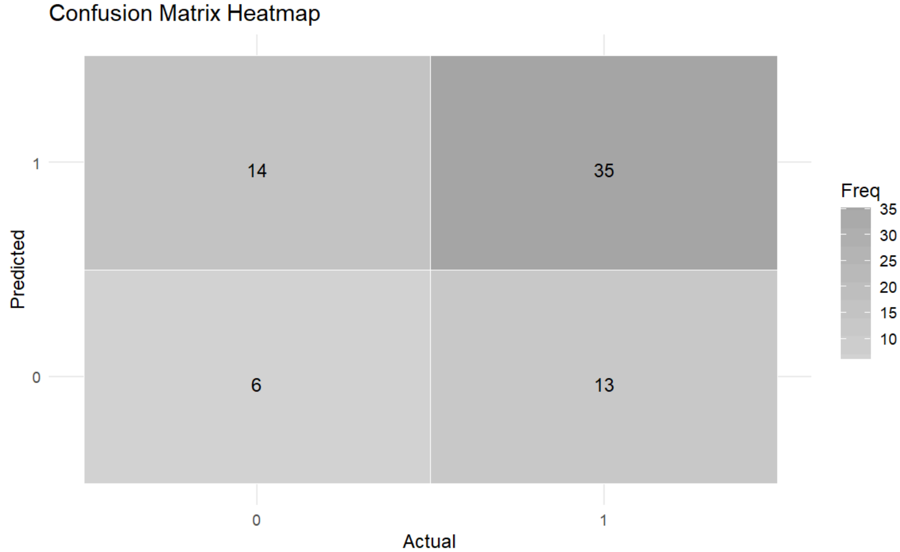
Random Forest: Home Win Loss vs. Home Total

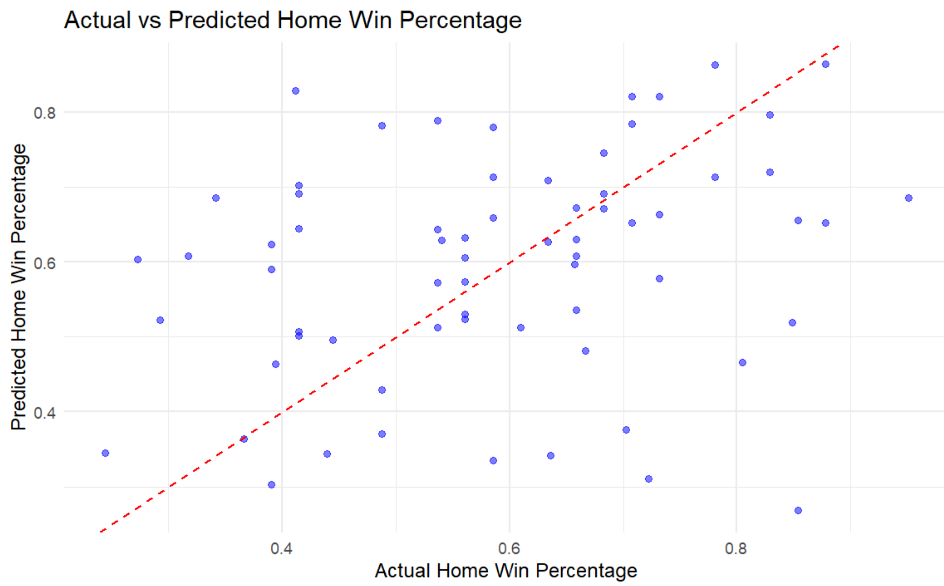


Random Forest: Home Win Loss vs. Gate Receipts



Confusion Matrix Heatmap




Actual vs Predicted Home Win Percentage
Predicted Home Win Percentage
0.8
0.6
0.4
0.4
0.6
0.8
Actual Home Win Percentage

Bibliography

ESPN, ESPN Internet Ventures, www.espn.com/nba/attendance/_/year/2024. Accessed 24 Sept. 2024.

“NBA Standings - 2023-24 Regular Season League Standings.” *ESPN*, ESPN Internet Ventures, www.espn.com/nba/standings/_/season/2024/group/league. Accessed 24 Sept. 2024.

Published by Statista Research Department, and Nov 2. “Golden State Warriors Gate Receipts 2023.” *Statista*, 2 Nov. 2023, www.statista.com/statistics/381541/gate-receipts-of-the-golden-state-warriors-national-basketball-association/.