Win Games To Fill The Stands?

Capstone Report

Alex Yom

Table of Contents

Introduction

In sports, having the home-field advantage can be the difference between a win or a loss. But at the same time, having a successful winning team may be a prerequisite to have a cheering, large home crowd. There is a reason why teams try their hardest throughout the season to attain home-field advantage. Players thrive off the energy from a supportive crowd. Team owners would also love to sell out every home game to maximize revenue and boost team morale. Of course every team executive would like to see their team amongst the most popular team in terms of highest average home attendance but they know that their team may play in one of the smaller arenas in the league, making it near impossible to compete with the teams with larger arenas. For that reason, team executives may want to look at the percent of the maximum capacity they fill on average throughout the season. That way, every team would be based on the same point. Some teams who experience huge capacities may only fill eighty-five percent of their stadium while another team that draws an average capacity each night might be selling out every single seat throughout the season.

What factors separate the most popular teams from the lower echelon teams? Are people more inclined to fill the seats of the stadium if their home team wins more often? If it can be proved that wins on the court can lead to fans in the seats, it would be in the best interest of team owners and basketball executives to assemble the best teams they can.

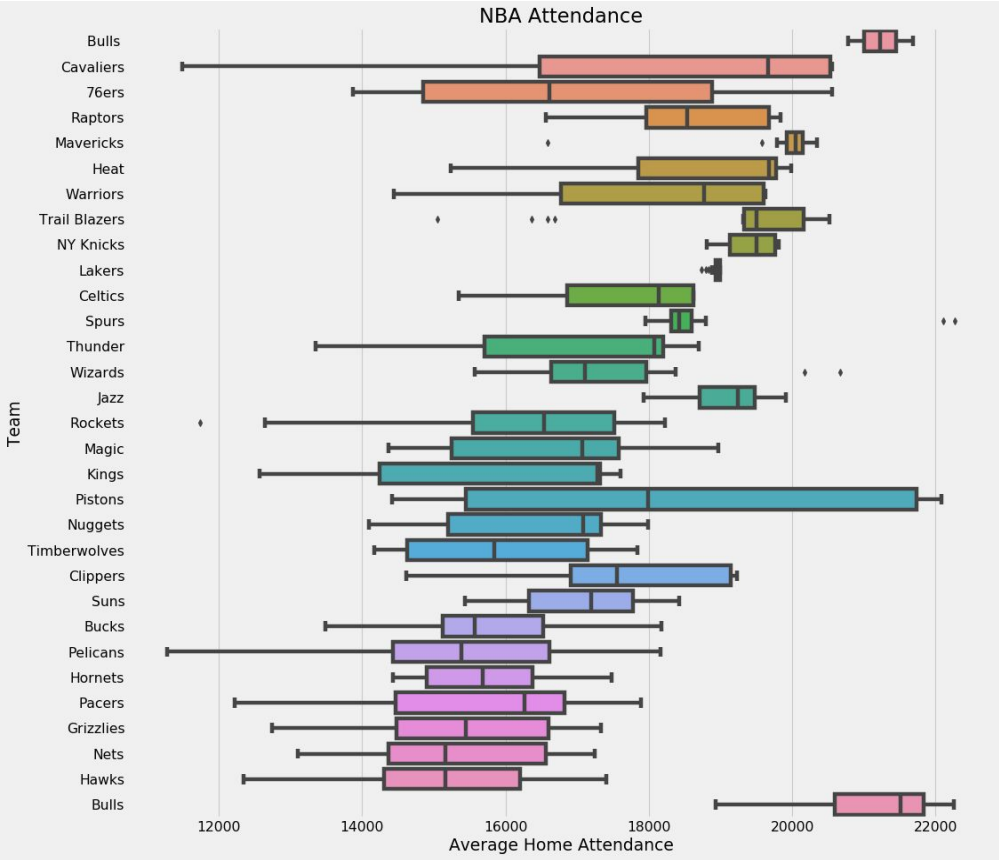
This project explored a dataset that chronicled the ranking of every NBA team based on average home attendance starting from the 2000-2001 season to the 2015-2016 season. The teams were ranked from highest average home attendance to the lowest. The teams were ranked from 1-30 with 1 being the team with the highest average attendance. The dataset also includes the number of wins that each team accumulated throughout the season to see if winning had a correlation to higher attendance. For a more in-depth analysis, I added more features to this dataset by combining it with another dataset that was assembled. Instead of just comparing the raw average numbers of attendance, I compared the percentage of the arena that each team filled per year and compare the percentages. Since each team plays in a different home arena which have varying capacities, it would be more uniform to compare the percentages. By computing the percentage of arena filled and inputting those values to a new column, we were able to compare the teams with a more unbiased perspective. Teams that

play in smaller arenas will not be put in a disadvantage and teams with larger arenas will not have an advantage.

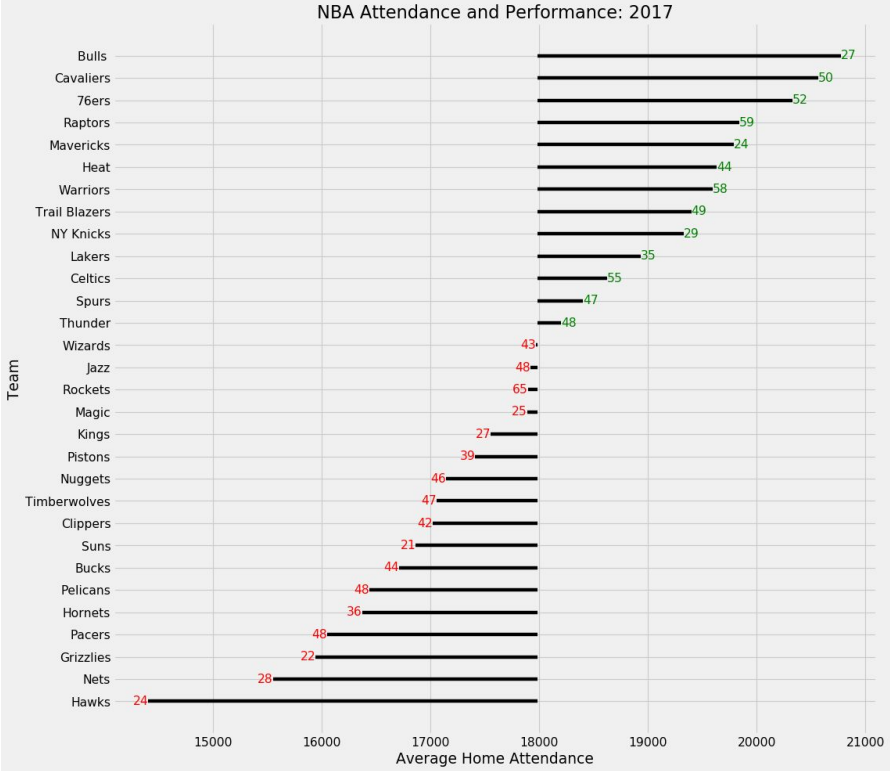
The data wrangling and cleaning process consisted of many steps due to the fact that three different datasets were being concatenated and merged together. The biggest data consisting of the data from years 2000-2015 served as the foundation because it was the largest dataset and had already been cleaned. The new dataset had the information from the years 2016-2017 and had to be merged with the foundational dataset. The values of the dataset were converted from integers to floats in the transition from excel to the notebook so the values had to be converted back. To ensure that these two datasets could be merged without complications, some columns were renamed and other columns were dropped. Lastly, the dataset containing the maximum capacity of every individual team's arena had to be merged with the dataset. After merging all three datasets, it was time to inspect the the dataset and make sure all the values were present and all values were the correct data type. During the merge, some values did not concatenate correctly and resulted in NaN values so those values had to be located and restored to their correct values. The final step in making sure our dataset was ready to be explored was the addition of a brand new column. The 'HomePercentCapacity' column would be the new feature added to this dataset. This column would take the percentage of the 'HomeAvgAttendance' of each team and divide that value over the 'HomeCapacity' value of each team. The new column would give us the percent that each team filled up their home arenas for that season.

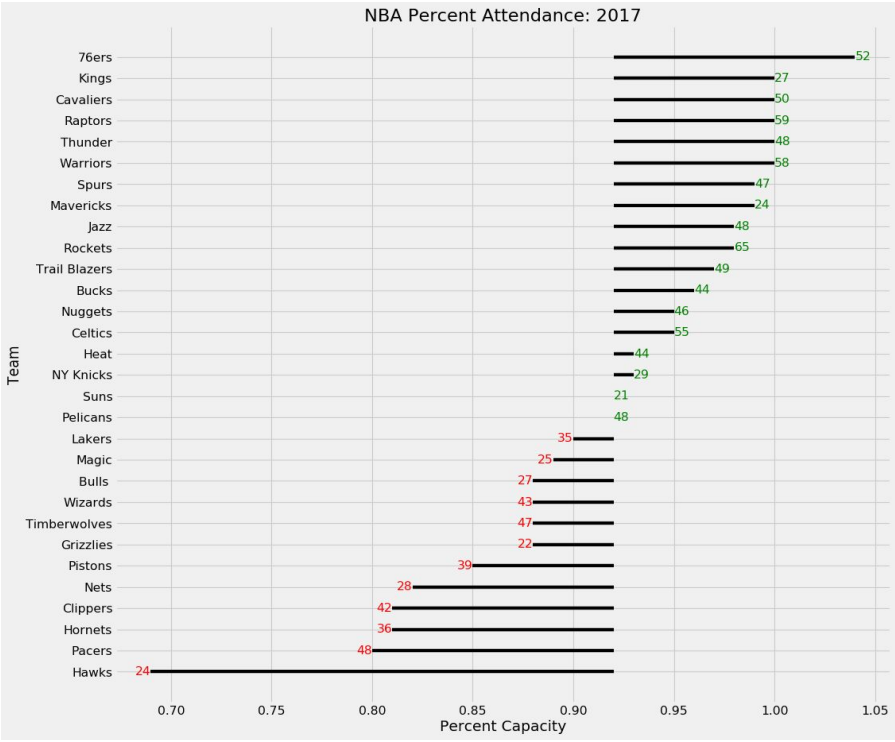
Analysis

Some preliminary data exploration revealed interesting information.

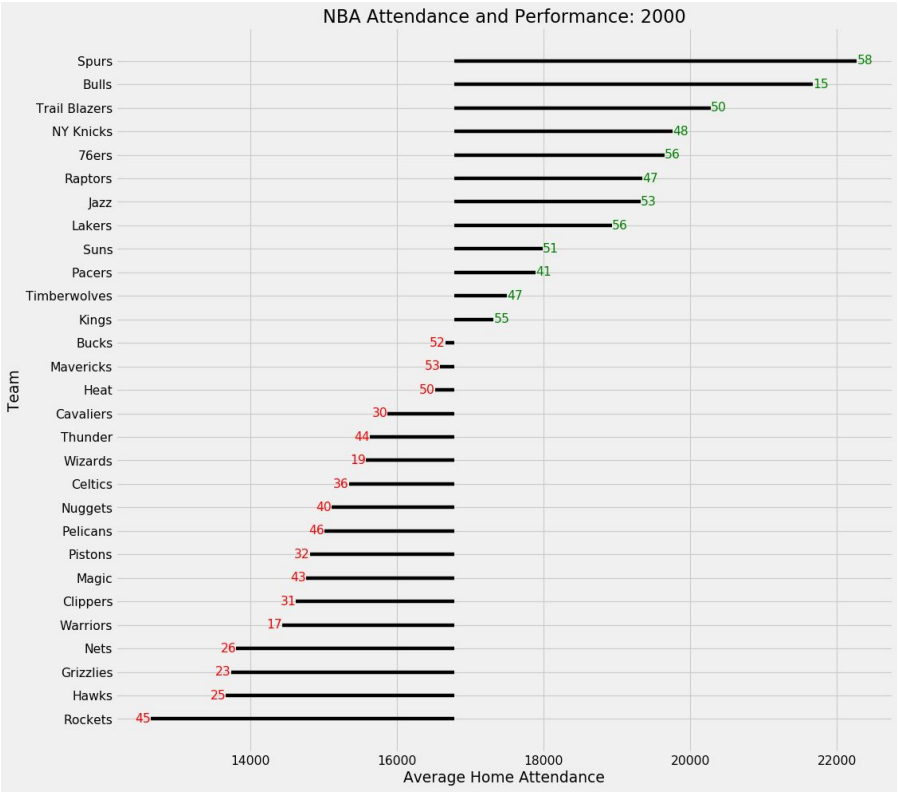


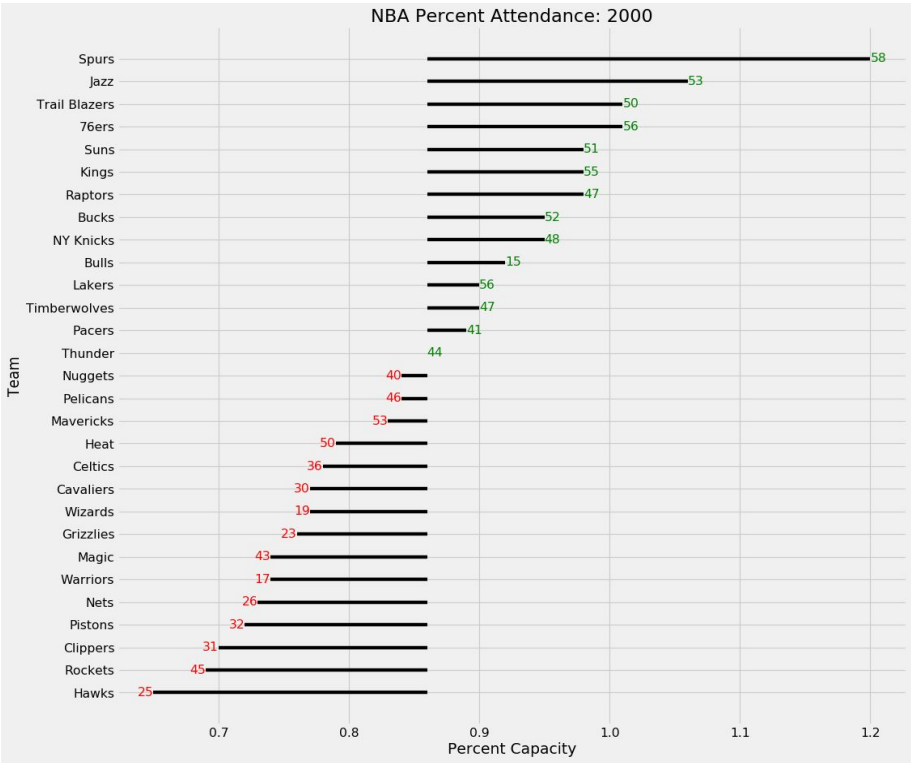
The boxplot above shows the culmination of the average home attendance for each team throughout the 17 seasons. There are several findings that this boxplot can highlight, that would not have been easy to see just by looking at the dataset. Some teams experience a great spectrum in terms of their average home attendance. Teams such as the Cavaliers, Pistons, and the 76ers all exhibit large disparities in their home attendance while other teams such as the Lakers and Spurs are amongst the league's most consistent in terms of average home attendance. There is little change in their attendance. There is a lot of information displayed on this one graph and each team can look at their own box plot and see the variance in their attendance. Some teams experience no variance in their attendance through the years while some teams experience huge waves of variance in their attendance. The one other feature that fluctuates or remains consistent year by year is the amount of wins for each team.


NBA Attendance and Performance: 2017
Team
Bulls 27
Cavaliers 50
76ers 52
Raptors 59
Mavericks 24
Heat 44
Warriors 58
Trail Blazers 49
NY Knicks 29
Lakers 35
Celtics 55
Spurs 47
Thunder 48
Wizards 43
Jazz 48
Rockets 65
Magic 25
Kings 27
Pistons 39
Nuggets 46
Timberwolves 47
Clippers 42
Suns 21
Bucks 44
Pelicans 48
Hornets 36
Pacers 48
Grizzlies 22
Nets 28
Hawks 24
15000
16000
17000
18000
19000
20000
21000
Average Home Attendance


NBA Percent Attendance: 2017
Team
76ers 52
Kings 27
Cavaliers 50
Raptors 59
Thunder 48
Warriors 58
Spurs 47
Mavericks 24
Jazz 48
Rockets 65
Trail Blazers 49
Bucks 44
Nuggets 46
Celtics 55
Heat 44
NY Knicks 29
Suns 21
Pelicans 48
Lakers 35
Magic 25
Bulls 27
Wizards 43
Timberwolves 47
Grizzlies 22
Pistons 39
Nets 28
Clippers 42
Hornets 36
Pacers 48
Hawks 24
0.70
0.75
0.80
0.85
0.90
0.95
1.00
1.05
Percent Capacity

The two graphs above, paint an interesting distinction between wins and filling up the seats. The first graph shows each team in terms of average home attendance in relation to the league average attendance. The numbers represent the number of wins that each team had during the season and the red coloring means that the team is below the league average while the green means that the team is above league average. Some interesting points to note are that the Bulls, despite winning a measly 27 games in the 2017-2018 season, led the league in average home attendance. The Mavericks and Knicks also placed in the upper echelon of attendance despite winning less than 30 games. On the other hand, the Rockets actually led the league with 65 wins, were below average in terms of attendance. But, the second graph paints a different picture and puts some perspective on reality. The second graph shows the same data as the first graph except it shows the percentage of the maximum capacity that the teams averaged for the season. When compared side-by-side, some teams are repositioned. The Rockets, led the league with 65 wins but finished below average in terms of average attendance, actually filled 98% of their stadium. Similarly, the Kings who were below average in terms of average attendance strictly by the numbers, actually averaged a sell-out crowd every home game, meaning their stadium was completely full for their home games. The Bulls, who led the league in average attendance, dropped below average in percentages as they came in at filling 88% of their seats on average.

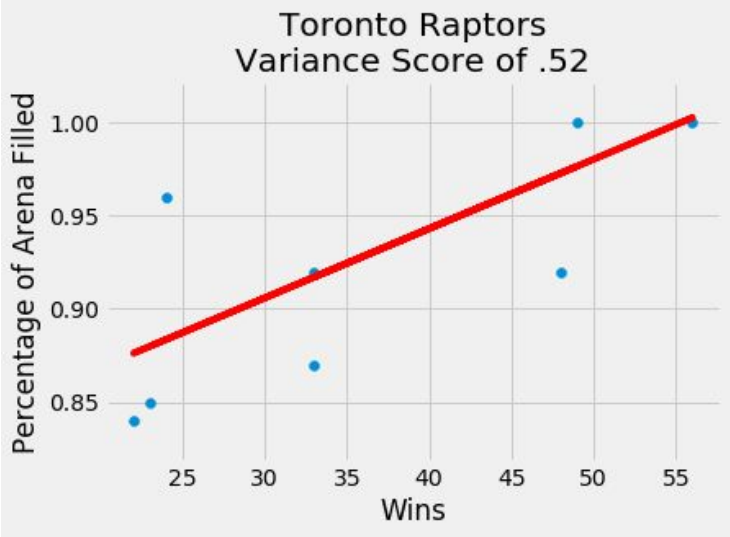


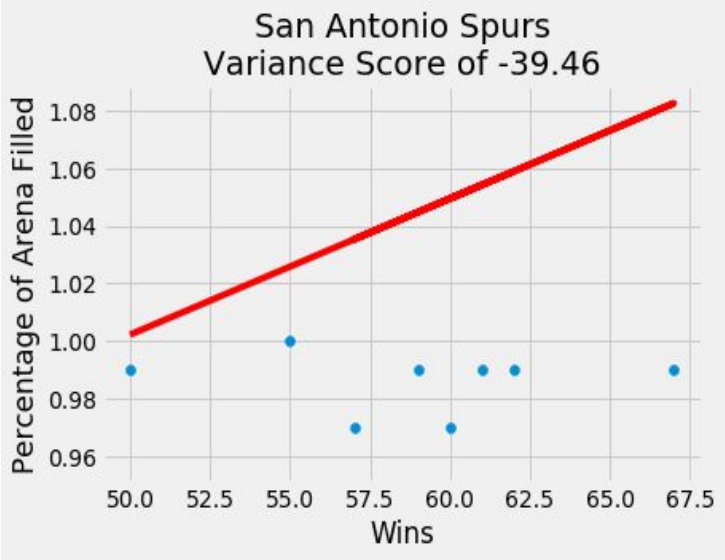


The two graphs above, show the data from the year 2000 and a lot has changed over the past 18 seasons. Throughout the 2000-2001 season, the Spurs enjoyed the largest crowd

in the league in terms of both raw numbers and percentages. The bulls who won 15 games, the lowest in the league, finished second in the league in average attendance and were also above league average in percent capacity as well. The 15-win Bulls were able to fill their stadium more than the 56-win Lakers during this season. During the 2000-2001 season, the Rockets won a respectable 45 games but finished second-to-last in percent capacity and last in average home attendance, lower than teams that won half as many as the Rockets.

After looking at the patterns of the league over the past 17 seasons, I planned to look ahead and see if a predictive model could be made to predict the percent capacity a given team could fill depending on their win totals. By using a linear regression model, I attempted to see how each team's patterns fluctuate and see how predictable each team's percent capacity may be.





The Toronto Raptors posted the highest variance value with .52 and the San Antonio Spurs posted the lowest negative value with -39.46. One interesting result was that slightly more than half of the teams posted a negative variance value. The wide range of variance scores leads us to believe that each team experiences a different relationship between their wins and filling their stadiums.

Discussion

The data analysis and visualizations shed more light into the league's attendance and also hints at the notion that wins may not be the only indicator for fillings the seats. The bar graphs showing the raw attendance numbers and the percent of the capacity filled showed an interesting relationship between the two. If a team executive was just looking at their team's rankings based on their average attendance, it would be misleading without also looking at the graph for percent capacity.

The NBA has gone through interesting changes during this time span from the 2000-2001 season to the 2017-2018 season. The average home attendance has increased from 16,783 during the 2000-2001 season to 17,988 during the 2017-2018 season. The average percent capacity increased from 86% during the 2000-2001 season to 92% throughout the 2017-2018 season. The overall growth of the NBA as whole would be an encouraging sign and each team would have to look at their specific data to glean more detailed information. During the 2017-2018 season, the Sacramento Kings, despite winning just 27 games, managed to fill 100% of their stadium throughout the season. This interesting trend seems to start from the 2013-2014 season, where the Kings were able to place within the upper half of the league in terms of percent capacity filled despite not winning many games. This would suggest that the fans support their team regardless of wins. Another interesting team would be the Atlanta Hawks who seem to struggle to fill their stadium. As of recently, they have consistently found themselves towards the very bottom of the league in both attendance and percent capacity filled. Even when they won an astonishing 60 games during the 2014-2015 season, they found themselves in the lower echelon in both attendance and percent capacity, behind the 25-win Orlando Magic and the 38-win Indiana Pacers. Teams can recognize trends in their percent capacity throughout the years as well. The Golden State Warriors found themselves towards the bottom in terms of percent capacity but they have been or at 100% of their stadium capacity since the 2012-2013 season.

The linear regression models with the variance scores illustrate the wide spectrum of trying to predict team's percent capacity. Some teams such as the Cleveland Cavaliers and Atlanta Hawks follow a very linear pattern in terms of their wins and percent capacity. It seems that these teams are able to fill more of their stadiums when they win more games. On the other hand, teams like the Dallas Mavericks and Utah Jazz appear to fill their stadiums in similar rates regardless of how many wins the team wins throughout the season. The different

circumstances scattered throughout the league would suggest that each team's home fans have different personalities with some teams having to produce exciting teams that win a lot of games to fill their stadiums while other teams have the luxury of having fans that come out to support the team regardless of winning.

Because each team experiences differences amongst the home crowds it would be interesting to explore further into each specific team to see other factors that influence attracting fans to the games. There are many different and interesting avenues to continue this research. Maybe teams could poll their home audience and produce a list of the top reasons their fans come to the game. It would obviously vary from team to team. Maybe some stadiums offer more amenities to continue pulling fans to the games. Maybe other teams just find themselves in advantageous locations with loyal fans who support their home team. Another possible avenue to explore would be the potential effect a popular player can have. Some teams are able to develop or attract one of the best players in the league and could certainly boost attractiveness to audiences. This certainly seems plausible and would be interesting to see if data exploration and analysis yield similar results.