Does Team Success Mean Higher Attendance

Nobody who likes sports likes to lose. Nobody who watches sports likes to watch their favorite team lose. But would people actually still pay money to come watch their hometown teams even if they do not win very often? This project will investigate the relationship between a team’s wins and their average home attendance. By viewing the relationship between wins and home average attendance, we can see if people will come watch their hometown teams live regardless of their status within the league. The better teams in the NBA would be expected to bring in larger crowds than the lesser teams in the NBA.

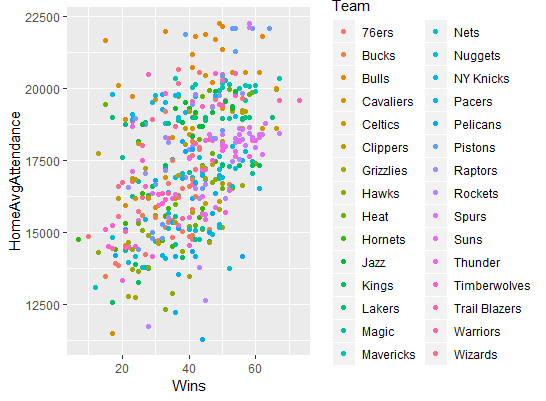
This project is not only interesting to the casual NBA fan, but also could be vital information for NBA team owners and general managers. The goal of NBA team owners is to generate profit while the goal of NBA team executives is to build and form the best teams they can. If winning more games is shown to draw large crowds, team owners and executives would be inclined to take a look. NBA team owners want their teams to bring in as much revenue as possible and if data shows that putting together better teams correlate to bigger crowds, then team owners would be motivated to focus on forming and building good teams. This information could be also be used to determine ticket prices for teams and attempt to maximize revenue from attendance. Team executives could also use this data to their advantage by securing their job by putting forth teams that win more games. It would be in their best interest to satisfy the team owners with their ability to build and create teams that win.

The dataset used in this project is a ranking of the teams that brought in the highest average home attendance throughout a fifteen year period starting from the 2000-2001 season to the 2015-2016 season. The teams are ranked from 1-30 in order of their average home attendance. Other columns included the team’s total home attendance for the season and also away game attendance as well. A separate dataset with each team’s wins during the respective season was created and added to the original dataset in order to compare the relationship between wins and average home attendance. The dataset consisted of 476 rows with the first row being the team that brought in the highest average home attendance during the 2015-2016 season and ending with the team that brought in the lowest average home attendance during the 2000-2001 season. Next to each team was a column indicating their win total for the season. Both datasets were rather clean and did not require much wrangling and cleaning. Although there were some missing values in some columns, there were no missing values in the columns that would be used in this project so they were left alone. The only modifications made to the datasets was merging the two datasets and changing the order of the columns for easier viewing.

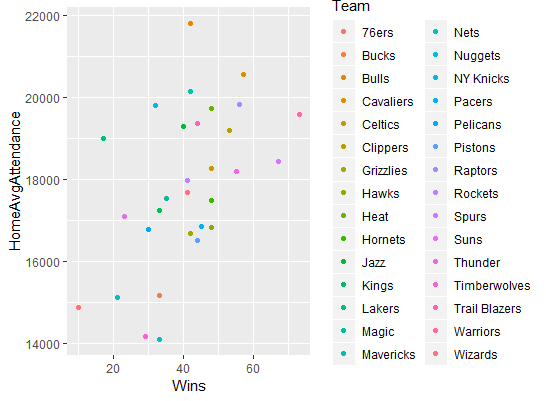
One limitation of this dataset is that it only showed the average home attendance and total home attendance but one thing that this dataset does not account for is the arena capacity of each team. Every team plays in a different arena with a varying amount of seats so it would be interesting to see the data on how many teams actually fill up their stadiums or if some teams fail to do so on a consistent basis. By adding that information to this dataset, we could compare the percentage in which teams fill up their stadium. By looking at a team’s percentage of filling the arena, it could put every team on a level playing field. For example, a team with a bigger arena would be able to bring in bigger crowds than a team with a smaller arena but if the team with a small arena consistently sells every seat while the team with the bigger arena fails to sell out every seat, then the case could be made that the team playing in the smaller arena is actually performing better than the team playing in the bigger arena. This could potentially be a further investigation for this project and it would be something to look into.

By just quickly browsing through the data, it would seem that wins do not affect the average home attendance. My initial hypothesis was that a team’s wins would not greatly affect the average home attendance. At first glance, some teams that did not win many games were amongst the leaders for average home attendance throughout the league. I would have expected to see the top teams in the league based on their wins to be attracting the biggest home audiences on average but that did not seem to be the case at first glance. The 2015-2016 Lakers won only 17 games during the season, one of the worst records in the league, and they still managed to rank in the 11th slot for highest average home attendance. The 17-win Lakers beat out the 67-win Spurs, the 48-win Celtics, and the 55-win Thunder. Alternatively, there are also cases of good teams who seem to struggle to attract people to games. During the 2015-2016 season, the Hawks won a very respectable 48 games during the season but only ranked in the 22nd in the league in terms of average home attendance, behind the 23-win Suns and the 33-win Kings.

However, plotting some of the data tells a different story. By producing scatterplots of the data in an attempt to represent the relationship between wins and average home attendance, a relationship is present. By plotting all the wins over the 15-year period as the independent variable against the average home attendance over the 15-year period as the dependent variable, a positive relationship is seen on the graph. The scatterplot shows the relationship between wins and average home attendance. As seen below, besides a few anamolies, as wins increase it appears average home attendance increases as well.



The positive relationship between wins and average home attendance is also apparent if we take a look at specific years and some years show more of a positive relationship than others. The graph below shows the data of the 30 teams during the 2015-2016 season.



It was very interesting that my initial views seemed to be swayed by what seems like outliers that do not speak for the whole data. After plotting the data and seeing that there does seem to be a relationship between wins and average home attendance,more specific research can also be done. Some questions to further investigate through this dataset would be to see if team location, east or west, plays a role in attendance. We can also look into the relationship between wins and average home attendance for specific teams and see if every team seems to follow the same trend or if they differ by city. Utilizing different kinds of plots could also prove to be beneficial in representing other trends or relationships that could be hiding within the data. Some other correlations may be found by using different plots that scatterplots fail to visualize. After these initial findings, my approach has changed because I know that there is at least some relationship to explore as opposed to my original hypothesis. Originally, I thought that there would not be any relationship between wins and average home attendance but after viewing the preliminary plots, there seems to be relationships and trends that can be further explored.