Assessing Audience Impact on NBA Player Performance

Abdo Ahmed - 30053710
Alexander Sewell – 30043188

1. INTRODUCTION

1.1. MOTIVATION

1.1.1. **Context**

Since the start of the COVID-19 pandemic, there has been plenty of uncertainty with restrictions, especially with regards to large events like sports. The NBA was one of the first major sports leagues to shut down in March 2020, and they were one of the first to resume their season in a "bubble" environment with no fans. The bubble took place in Disney World, where all games were played on neutral courts with no fans. The NBA bubble had many crazy moments, with some players having lots of unexpected success. After the bubble season finished, a new NBA season began back in the team's regular arenas. Depending on the jurisdiction, arena capacity ranged anywhere from 0% to 100%. This trend changed with each wave and development of COVID-19. It is well known throughout sports that fans affect the players in different ways, both positively and negatively. We are curious how the attendance levels at arenas affect player performance. More specifically, we want to quantify the impact the audience has on NBA player performance.

1.1.2. **Problem**

The problem we want to explore is the impact of attendance levels on NBA teams and players' performance. We plan on doing this by looking at crowd sizes as well as exploring the impact of 'home-court advantages.' This problem is important to address because the NBA is the premier basketball league in the world, and as such should be focused on providing the best product possible. By researching this topic, the NBA can determine if having a live crowd has negative consequences on the quality of play that they display to the world. Furthermore, this can give general managers and coaches insight into the quality of certain players. If certain players seem to struggle under pressure with lots of fans in attendance but play quite well with limited fans, it might be important to keep in mind for the long-term success of this player. On the other hand, players are constantly graded on their performance in tough environments, specifically when they play on the road against opposing fans. Players who can rise to the occasion in these difficult situations are vital to teams looking for success in the playoffs. Popular media and sports forums tend to make anecdotal statements about players' performance in certain environments, but we intend to quantify this impact more thoroughly.

1.1.3. **Challenges**

There are many confounding variables in this problem. Players are rarely consistent game after game, so determining if a player is struggling or performing well is due to strictly attendance levels will be difficult. For example, some players are known as "streaky", in which they perform well for a period, only to be followed by a period of poor play. Furthermore, attendance levels are heavily related to the team which a player will be playing against. A team situated in an area with more COVID restrictions will have a higher sample size of low-attendance games than teams in less restricted areas. Sometimes players have difficult or favorable matchups depending on the quality of a certain team or the quality of players at certain positions. This will need to be accounted for in determining how impactful attendance levels are to player performance. For example, the Utah Jazz are known for their high-quality defense anchored by center Rudy Gobert. It is likely that players will struggle against this matchup simply due to his presence, regardless of the impact of the crowd. We also need to consider how to best define player performance. This has been attempted to do with different advanced metrics, but we will need to explore which of these metrics best work for our project. Another challenge is this project may be biased towards offensive-minded players. Offensive stats like points, assists, and shooting percentage are very easy to measure and are readily available. While defense has some measurable stats like steals and blocks, the true impact of high-quality defensive play doesn't show up on the box score. Even on nights where a player is struggling to shoot the ball, they can do the "little things" like hustling and boxing out to make a positive impact on the game. Finally, we need to consider that team rosters change from season to season. A player who was in a bad situation one year, may end up on a team the next year where they are utilized more effectively.

1.2. OBJECTIVES

1.2.1. Overview

We intend to examine the impact of an audience on the performance of NBA players. Within this overall idea, we will need to explore multiple different areas of focus, including defining player performance, analyzing matchups, and determining home court advantages.

1.2.2. Goals & Research Questions

Overarching Goal: Assess the audience's impact on NBA player performance.
Overarching Question: How can we use data science techniques to determine the impact that the fans in attendance have on NBA players?
Sub-questions:

1. What are the best metrics we can use to assess player performance?
2. Are certain players "big game performers"? That is, do some players perform better under pressure than others (Regular season vs playoffs games)? Are star players more effective than non-stars?
3. How advantageous is a home crowd? Do players and teams lose the home court advantage with a lack of fans? Are there teams that have a stronger home court advantage?
4. What matchups do certain players struggle with? For example, are there certain teams that most players struggle or do well against? Is a top-level team, regardless of capacity, more difficult for players than a bottom level team?

Some of these questions, while not directly related to the overarching question, are critical in our analysis. For example, if a player consistently struggles against a certain team, it may not be the attendance level affecting their performance. We need to answer these questions to avoid spurious correlations.

1.3. METHODOLOGY

1.3.1. Data

Data Set 1 - Attendance Data: The data providing scheduling and attendance information can be collected using an extract of a csv file provided from basketball-reference.com. The attributes will be Date, Time, Visiting Team, Visitor Points, Home Team, Home Points, and Attendance. Attendance is shown as a number with no reference to the full stadium capacity, so we will also need to collect that data to show attendance as a percentage. We will either scrape that data from espn.com or use the maximum attendance value for each home team as the denominator.

Data Set 2 – Player Game Data: The data providing nightly box score information will be collected via web scraping using a spider we will create in R/python. We need to scrape data as there is no current dataset that compiles this information, so we would have to create one ourselves. Game-by-game data is available on basketball-reference.com so we will collect it from there. The attributes we will scrape include basic box score stats such as Minutes Played, Points, Rebounds, and Assists. We will also scrape advanced box score stats such as True Shooting %, Offensive Rating, and Box Score Plus Minus (BPM). When searching a variety of articles, we determined BPM to be the best measurement for overall player performance. The goal of this stat is to use a variety of data points from the game "to estimate a player's contribution in points above league average per 100 possessions played" [1]. We believe this metric is the best for our analysis because it accounts for many different factors compiled into one final number.

1.3.2. Approach

We want to focus in on the impact of crowds on a player's performance, and we will do this by looking at a few different scenarios which include home game performance vs away game performance, regular season performance vs playoff performance, and non-empty arena performance vs empty arena performance. We plan on running t-tests and proportion tests to analyze the performance in each category. We decided to use statistical testing instead of predictive techniques such as regression because we believed an interactive dashboard was more valuable. Although, if it is deemed to be a valuable addition to the project, we will use predictive tools such as linear regression to look at future performance.

We think this will work well because of the number of different metrics that are being tracked for basketball related data. A player's performance can be quantified by factors such as the number of points they scored, the amount of steals they had, or the number of rebounds they grabbed, among others. This allows us to measure a successful performance in a way that we find to be justifiable. Using statistical tests to analyze these factors, paired with a powerful visualization tool will be allow us to capture these results.

1.3.3. **Workflow**

Workflow:

- Build tools for data collection/preparation (Scraping done in R with the RVEST package)
- Clean and organize scraped data
- Perform Exploratory Data Analysis to gain initial insights
- Conduct statistical experiments and tests
- Create a user-friendly dashboard in R Shiny to present our results

The most challenging portion of the workflow will be conducting statistical experiments, specifically controlling as many factors as possible while trying to isolate the impact of an audience. There are many interactions that we need to consider, and ultimately, we will have to settle on a realistic goal. It will be impossible to control all extenuating factors, but we will have to work to control these factors to the best of our ability. We expect that our statistical analysis will be able to explain only a portion of the variation in player performance as a function of audience impact. If this step does not work out perfectly, we still expect to have insights into how players are affected by different audiences.

1.3.4. **Workload Distribution**

Initially both group members will share the workload evenly, completing the steps together. As our project gets more established, we will take on more defined roles and responsibilities. Projects like these are ever evolving so the more we uncover about our project the more this will change.

1.4. CONTRIBUTIONS

Data analytics has increasingly been used to help gain insights over the past few years. According to Forbes magazine, the field of sports analytics should increase to “almost $4 billion by 2022” [2] This could be attributed to the vast number of useful applications that analytics can present. Analytics can be a tool used to improve the business side of sports by learning what factors affect fan engagement. Furthermore, analytics can have an impact on the actual game itself. Teams can develop strategies that are based on insights created by analytics, which could allow them to attack weaknesses within their system while simultaneously leveraging strong parts of their game to ultimately gain a competitive advantage over their opponents. It will be interesting to see how significant an advantage can be created using sports analytics in the future.

Due to the covid pandemic, the NBA has had to make the decision to play games in empty arenas with no crowd present. This unprecedented circumstance allows us to gain a more complete insight into how a crowd affects the quality of basketball being played. This will benefit NBA executive leadership because they will have a better understanding of what factors influence their players, thus allowing them to make better decisions that will influence the quality of the product they present. This will also benefit the players/teams that are playing the game, as this information can allow them to prepare for games given the environment they will be playing in.

The insights we generate from this project could also be useful in the sports betting environment. Discovering trends about player performance in different situations can help gamblers gain foresight about a game they decide to bet on. If a certain player seems to consistently struggle in a certain environment, gamblers could find opportunities to bet against that player. Ultimately, since determining direct causality will be difficult to do in this project, predictive power for future games will likely be limited. There may be, however, a potential edge to gain for sports betting uses.

Overall, our R Shiny web app will be of benefit to NBA fans. They will be able to interact with the data to gain insights on their favorite teams and players. Fans are always interested in learning more about how players and teams perform, so giving them an interactive dashboard will be useful for them.

2. RELATED WORK

1.1. TECHNOLOGY SCAN

The question on the impact of fans is something many people have been asking about since the new world of empty arenas became a necessity. In our research, we found many news articles exploring the subject. The Los Angeles Times wrote “sports psychologists wonder if the lack of crowd response might affect focus, effort, and even strategy in the games” [3]. They compared it to a stand-up comic performing to a camera instead of an audience. Players feed off a response from fans, so they didn’t know what to expect with the new bubble environment. Lebron James even said, “play games without the fans? Nah, it’s impossible” [3]. These players are so used to entertaining, so playing without fans was a whole new experience. Something to keep in mind, as pointed out by The Ringer, is that “a lot of the arenas pipe in crowd noise” [4]. The same article investigated some of the statistics of certain teams’ home and away records depending on their crowd, but it wasn’t an in-depth analysis, and we expect our report to be more rigorous. An article from 2015 looked at the effects of an empty stadium in baseball, and it suggested “less experienced players might find it easier to concentrate in an empty stadium” [5]. We will keep this in mind throughout our analysis. Moving away from attendance levels, Yard Barker wrote an article in 2019 ranking the NBA home crowds [6]. Our analysis will include this information and try to follow their methodology to see how the home court advantage has changed in the last 3 years.

Overall, there has been lots of speculation and anecdotal reviews of the impact an audience or lack thereof has on the NBA. We were not able to find any major statistical breakdowns of this problem, so we believe we will be able to factualize the impact, and potentially back up or refute the points provided in the outlined articles. These articles will be good guiding points for us as we make our way through the project.

1.2. ACADEMIC PAPERS

1.2.1. No Fans, No Problem: An Investigation of Audience Effects on Shooting Performance in Professional Basketball [7]

This research paper was conducted by the Department of Kinesiology at the University of Tennessee on December 30, 2021. Like our project, it examined the impact of the lack of audience during the NBA Bubble in 2020. They noted an observational increase in free throw shooting percentage during the spectator-less bubble and wanted to back that observation up with statistics. Their results found that there was a significant increase in free throw shooting percentage during the bubble. As kinesiologists, they noted some potential reasons for this increase. Firstly, they discussed open vs closed skills in sports. An open skill is one where the environment is unpredictable and constantly changing, and the movement required to perform the skill is influenced by that environment. In basketball, this would be jump shots near a defender, or a layup in traffic. A closed skill is the opposite, where the skill is performed in a predictable environment. Free throw shooting would be considered a closed skill, since you always shoot from the same place, on a hoop the same height, with no defenders impacting your shot. Sports research has found that closed skills are performed better when they are done in similar environments to an athlete’s practice. The logical conclusion is that free throw shooting practice for NBA athletes takes place in a relatively empty gym, with no audience noise, so a game environment with no audience noise (such as the Bubble) will be much closer to the practice environment than games with thousands of spectators in attendance. The other consideration is the “choking” phenomena, which means negative performance in high pressure situations. The study noted that there are mixed results with other studies on this subject, where some suggest free throw shooting to be unaffected, while others suggest a link between spectators and choking. A robust study done in the NBA from 2009 to 2016 found that higher levels of spectators led to a worsened free throw percentage in high pressure scenarios. This example of choking was found to be significant regardless of whether these spectators were supporting the shooting team or against the shooting team. Based on this research, the study concluded that the NBA Bubble environment would likely be less conducive to choking due to the absence of spectators. This study only examined these free throw shooting differences in the Bubble. We can build on this analysis and try to understand the impacts post-Bubble, where there were still many games played in empty arenas for most of the next season.

1.2.2. Spectator Effect on Team Performance in College Basketball [8]

The next study was conducted in June 1993, looking at a measles outbreak that happened during the 1988-1989 college basketball season, resulting in empty arenas for a short time. This paper used three dependent variables: Total points scored, field goal percentage, and free throw percentage. There were two teams used in their analysis, as they were the only ones with multiple games played in both a normal environment and a spectator-less environment. The data showed a large observational difference, with improvement in all categories for both teams in the no spectator environment. They conducted t-tests to back that up with statistics, and they found no significant differences. This was mainly due to the low sample size of 11 games. The researchers found that the results would have to be replicated over 20 games for the results to return as significant. They concluded the paper, coincidentally, by saying the next analysis "will have to wait until the next misfortunes", which is exactly the case for our project.

1.2.3. Choking under Pressure: Evidence of the Causal Effect of Audience Size on Performance [9]
This discussion paper, published by the Institute of Labor Economics, analyzed free throw shooting performance under pressure to estimate a causal effect between audience size and free throw success. This study used NBA data from the 2007/08 season to the 2015/16 season, looking at all regular season games. They noted a strong negative effect on free throw shooting caused by an increase in spectators. They begin by explaining the phenomenon known as "choking under pressure", which means that higher pressure situations lead to a decrease in performance. Factors influencing pressure include the importance of certain situations, increased monetary incentives, and the presence of a supportive or hostile audience. While it might be natural to think that playing in a supportive atmosphere would reduce pressure, studies have shown biological stress indicators to increase even when playing in front of a supportive audience. This study goes deeper into quantifying pressure than our project. For example, they used play-by-play data to understand the score difference and the time of the game that the free throw was performed. Our data is only looked at as the total game performance, not considering these intricacies. This is important because pressure to perform is fluid and changes throughout the game. Audience size is one variable of many that can affect a player's free throw shooting success. Interestingly, this study found a negative causal effect between audience size and free throw success for the home team. This effect was mostly driven by shots taken in the first half of the game. Away teams are not significantly affected in a similar way. Contrary to the commonly known home court advantage, this was an interesting case of a home disadvantage. Overall, this study provided us with useful insights to keep in mind when continuing our analysis, particularly the influences on pressure that our data doesn't consider such as timing of the game and importance of the free throw.

1.2.4. Estimating the Effect of Home Court Advantage on Wins in the NBA [10]
The final academic paper we read was written for the Economics Department at Illinois Wesleyan University. The goal was to quantify the effect of the home court advantage in the NBA. The study used a logit regression to study how fan attendance, field goal and free throw percentages, fouls called by the referee, and days of rest effect a team's success in a home court environment. They found that an increase on one standard deviation relates to an increase of the home team's chances of winning the game by 2.7%, suggesting that the supporting fans have a positive impact on a team's success. They also noted a referee bias, speculating that referees tend to favor the home team to avoid getting booed by the home crowd. Overall, the home court advantage was backed up with statistics in this analysis. We believe our project can further quantify the home court advantage, considering the unique environments of empty arenas and the NBA Bubble.

3. WORKFLOW

3.1. Step 1: Build Tools for Data Collection/Preparation

- **STRATEGIES:** The possible implementation strategies we had for the attendance dataset were to manually gather the data or use a web crawler to scrape the data. The possible strategies for the box score data were to manually gather the data, use a web crawler to scrape the data, or to use a dataset provided by Sthathead, a subscription service from basketball-reference. Each dataset would be retrieved from basketball-reference.com, one of the leading sources of NBA data. The advantage of pulling data manually is that we can have some autonomy when choosing how our dataset looks like, but a disadvantage is it is very time consuming. An advantage of web scraping is it is a skill that can help improve our overall knowledge of HTML/CSS. A disadvantage would be it is harder to correct errors in the scraped data without restarting the process. Lastly, an advantage of Stathead is that it has most of the data available to us in a clean format, but a disadvantage is that it is a paid subscription to use.
- **IMPLEMENTATIONS:** We compiled the attendance dataset and box score dataset by using a web crawler to scrape the data. We did this because the attendance data was spread across over 30 different links and the box score information was spread across over 3000 links. Using a web crawler to crawl the links for us automatically and retrieve table information was a much more viable option. The web crawler was created in R, made with the capabilities provided from the RVEST package. We added to our data set by scraping additional box scores for the current 2021-22 season.
- **LIMITATIONS:** For the box score dataset, there were a few options for the type of data we could retrieve. The

webpages typically provided us with eight tables to choose from for each game, with each game offering us granular information such as player performance in each quarter, half, and the entire game. We are only interested in the full game results for our analysis. The other options are too granular and do not fit in within the scope of our project.

3.2. Step 2: Clean and Organize Scraped Data

- **STRATEGIES:** Once we have gathered the two datasets, we are able to merge them into one dataset. This would be possible by assigning a unique identifier to both datasets so that we can connect the attendance of a game to the player and team data for that game. We also need to identify any inconsistencies that exist within the dataset, such as the use of null/NA values vs the use of '0' values in the same column. This is important to clean up because quantitative analysis that we conduct usually will not work properly when there are null values present. The cleaning could be done multiple ways, potentially through coding with R or Python, or using software such as Microsoft Excel's Power Query functionality. The advantage of using R/Python is that it is better at handling large amounts of data, while an advantage of using Power Query is that exporting the data is far more convenient due to Power Query's integration to Microsoft Excel.
- **IMPLEMENTATIONS:** We will connect the two datasets by using for-loops to look through each dataset and find a match with the unique identifier. Once a match is found, we can add a new column for attendance to add on to the box score dataset. To identify inconsistencies with null values, we will use the r function is.na() to search quantitative columns and identify any null values. If we do detect any, we will investigate to see if it is appropriate to change the values or leave them the way they are.
- **LIMITATIONS:** No major limitations encountered in this step. Basketball-reference provides clean, high-quality data so we were able to finish the cleaning and organizing process smoothly.

3.3. Step 3: Perform EDA to Gain Initial Insights

- **STRATEGIES:** The purpose of EDA is to further understand our dataset and the meanings behind the values. This helps us determine the context behind the questions we have. Some examples of the type of EDA we can do include creating histograms to understand the distribution of attendance levels for our dataset, as well as boxplots showing different player stats in empty vs non-empty arenas. We can also perform initial statistical tests such as t-tests and proportions tests on players/teams to see if there are statistically significant differences in the dataset. Another portion of this phase is to research different basketball stats and identify certain key metrics we can use to evaluate performance levels. Exploratory visuals can be created in Power BI, or in R using the GGPLOT package. Initial statistical tests can be done in R. Finding the best metrics will be done through online research as well as exploring our dataset.
- **IMPLEMENTATIONS:** Using R, we created both histograms and boxplots to understand our data better. Histograms were useful in looking at attendance levels, as each season had a very different distribution of these numbers (see appendix). We learned that we may need to incorporate data from this current season to show the impact of spectators returning to the audience, as last year's season had more empty arena games than we expected. Boxplots were used to look generally at different metrics and how they differed in empty/non-empty scenarios. Based on this, we ran some basic t-tests, looking at average points scored per game for the Miami Heat in the 2019-2020 season. We wanted to initially explore the 2019-2020 season as our histogram showed this season to have the most balance of empty (bubble) and non-empty (pre-bubble) games. The Miami Heat made it to the NBA Finals, so we have a good sample size of games in each category. Initial tests showed a significant increase in free throw %, while field goal % and points per game were not significantly different, and observationally these stats decreased in the bubble. These results are very preliminary and will be built on further as we move along with the project.
- **LIMITATIONS:** Due to the size of the dataset, we haven't moved as far along this step as we planned to at this point in the project. We are updating the timeline to complete this phase by March 4th, 2022. We currently have only explored a few performance features as they relate to an empty vs non-empty arena. We plan to look at other factors, such as home game vs away game performance, and how audience affects the potential home court advantage.

3.4. Step 4: Conduct Statistical Experiments and Tests

- **STRATEGIES**: The purpose of this step is to conduct the analysis of our data to answer the research questions that we have asked. We will do this on the dataset that we have scraped together, while also adding variables to our dataset based on the dataset. We will then have numerical variables from the results of the game, as well as categorical variables that help us describe the environments that the game was played in. These variables will allow us to conduct statistical tests to examine player performance in different conditions to help see if there are differences in player performance. The possible implementations include a variety of statistical tests such as t-tests and proportion tests.

- **IMPLEMENTATIONS**: Currently we have implemented t-tests and proportion tests to understand the effects of empty arenas. We also may explore creating a logistic regression model to predict whether a game will result in a win based on a variety of inputs. This would allow us to understand the effects of multiple variables at once on a team's chance at winning. We will continue to update this step as we progress with the experiments.
- **LIMITATIONS**: We are slightly limited by our data set. The academic paper, "Choking Under Pressure", described many features that can influence a player's performance that our data set doesn't capture. For example, when measuring free throw shooting performance, the time remaining in the game, the difference in the score, and the importance of the shot can all impact a player's ability to perform well. Our data set doesn't account for these externalities that can affect the pressure of the situation. This isn't a major issue as we have a robust data set to achieve our main goal, and these extra features are outside the scope of our project.

3.5. Step 5: Visualize and Present Results

- **STRATEGIES**: During this step, we will need to communicate the findings of our research. A strategy to accomplish this is building a dashboard in Power BI or Tableau. This would allow users to look at the findings of our research in a digestible and interactive way. The advantage of using Power BI or Tableau is the opportunity to connect to multiple data sources at once, so if we wanted to expand on our project in the future, we would be able to do so while making connections directly in the application. Alternatively, we can build a web app using R Shiny. We have less experience using this package, but it would be a great opportunity to expand our skillsets. An advantage of using R Shiny is that we believe this would give the user extra analytical flexibility without the requirement of knowing how to program with R, while enabling the same insights as an analyst fluent in R.
- **IMPLEMENTATIONS**: We created an R Shiny web app with two tabs: Home Court Advantage and Player Impact Analysis. The Home Court Advantage tab lets users choose the team and the environments they wish to analyze, and data displaying winning rates at home vs away are shown with a proportion test p-value to indicate whether there is a difference between home and away courts. The Player Impact Analysis tab lets users choose a player, statistic, and which seasons they want to analyze. The output is a player summary, showing general performance information. Below that is analysis on the statistic of choice that was selected, including visualizing the distribution and trend of the statistic, and a t-test or proportion test of that statistic between empty and non-empty environments.
- **LIMITATIONS**: Power BI and Tableau are limited in flexibility, particularly when it comes to showing statistical test results. Using R Shiny would result in less data processing to get the results formatted well in Power BI or Tableau. In terms of R Shiny, we are limited in our knowledge of the package because we have not had hands on experience building a functional web app. We believe we can overcome this limitation and learn a new skill in the process. Now that the dashboard is completed, some limitations we found were difficulties organizing the visualizations. While Power BI and Tableau allow you to drag and drop everything, R Shiny visualizations are created by code, which took a while to get used to.

4. RESULTS

Question 1: What are the best metrics we can use to assess player performance?

Although not a quantitative testing question, this question is important to answer because it is going to be how we measure performance for our statistical tests going forward. In terms of player performance, the three metrics we decided to use were BPM, points scored, and free throw percentage.

- BPM: Box Plus Minus (BPM) is a metric that is calculated by basketball-reference.com. It uses the entire box score to determine a player's rate of production per 100 possessions played. It is advantageous to use this metric because it considers the players position, since different positions are going to be valued differently in terms of box score metrics. (i.e. Point guards will be weighted more heavily on points scored and assists passed, while centers will be more heavily weighted on rebounds scored and shots blocked.) This way, we can use a metric that standardizes performance across the board. The scale of BPM is scaled so that a score of 10 is an excellent performance, while a score of -2 or below is consistent with performance of players who come off the bench. (**See Appendix**)

- PTS scored: We decided on using points scored as a metric since it is the most influential stat in determining a player's performance, and the one that most people seem to be focused on. We want to see if the average points scored for players is different between the different environments.
- Free throw percentage: We decided on using free throw percentage since it is the only shot in the game of basketball where the player has no interference from other players. We hypothesize that because of this, free throws are one of the best ways to determine crowd impact since the crowd is the only factor that could distract a player that is shooting free throws.

EDA: Exploring the relationship between performance and attendance levels with the Miami Heat

4.1. EXPERIMENTS: Three statistical tests (one t-test, two proportions tests) with data for the Miami Heat during the 2019-2020 season. We tested points per game, free throw %, and field goal % in games with fans in attendance vs games with no fans in attendance (NBA Bubble games). Boxplots of these numbers are shown in the appendix.

4.2. QUANTITATIVE:

4.2.1. T-test for points:

H0: The mean points scored with and without fans are the same
Ha: The mean points scored with and without fans are different
Alpha: 0.1

- Mean points with no fans – 110.31
- Mean points with fans – 112.15
- p-value: 0.4135

4.2.2. Prop-test for free throw % (FT%)

H0: The proportion of made free throws with and without fans are the same
Ha: The proportion of made free throws with and without fans are different
Alpha: 0.1

- FT% with no fans – 82.41%
- FT% with fans – 77.77%
- p-value: 0.0106

4.2.3. Prop-test for field goal % (FG%)

H0: The proportion of made field goals with and without fans are the same
Ha: The proportion of made field goals with and without fans are different
Alpha: 0.1

- FG% with no fans – 45.73%
- FG% with fans – 46.96%
- p-value: 0.325

4.3. INSIGHTS:

Overall, this initial testing shows a significant improvement in free throw shooting percentage for the Miami Heat. Points per game and field goal shooting percentage was not significantly different between the Bubble environment and the regular environment. This seems to align with the assumptions of the No Fans, No Problem study, where closed skills are likely to benefit in environments similar to practice while open skills may not. As these are just initial tests, we will not draw strong conclusions either way. To improve this example, we would need to look at the differences in the teams they played against. Since the Bubble consisted mostly of playoff games, the Miami Heat would have played nearly all their games against strong competition, whereas the regular season would have many games with weaker competition. Normalizing for the defensive quality of the other team will be important in the next steps of our project.

Question 2: Are certain players "big game performers"? That is, do some players perform better under pressure than others (Regular season vs playoffs games)? Are star players more effective than non-stars?

4.1. EXPERIMENTS:

We used t-tests and proportion tests to compare the means and proportions respectively of players in playoff games and non-playoff games. To fulfill the condition of normality, we only tested players who had more than 25 playoff and non-playoff games, because of this only 56 players were eligible for testing. The three metrics we tested for were Points, BPM and Free Throw Percentage.

4.2. QUANTITATIVE:

4.2.1. Test 1: T-Test for Performance by BPM per Player for Playoff and Non-Playoff Games

H0: The mean BPM in playoff and non-playoff games are the same. ($\mu_{playoffBPM} - \mu_{non\text{-}playoffBPM} = 0$)
Ha: The mean BPM in playoff and non-playoff games are different. ($\mu_{playoffBPM} - \mu_{non\text{-}playoffBPM} \neq 0$)
Alpha: 0.1

4.2.1.1. Results: Number of players in each category

- Nearly equal BPM: 3 Players
- Non-significantly decreased BPM (p-value between 0.1 and 0.9, decrease in empty arenas): 24 players
- Non-significantly increased BPM (p-value between 0.1 and 0.9, increase in empty arenas): 24 players
- Significantly decreased BPM (p-value <= 0.1, decrease in empty arenas): 3 players
- Significantly increased BPM (p-value <= 0.1, increase in empty arenas): 2 players

4.2.2. Test 2: T-Test for Performance by Points per Player for Playoff and Non-Playoff Games

H0: The mean BPM in playoff and non-playoff games are the same. ($\mu_{playoffPTS} - \mu_{non\text{-}playoffPTS} = 0$)
Ha: The mean BPM in playoff and non-playoff games are different. ($\mu_{playoffPTS} - \mu_{non\text{-}playoffPTS} \neq 0$)
Alpha: 0.1

4.2.2.1. Results: Number of players in each category

- Nearly equal PTS (p-value >=0.9): 4 players
- Non-significantly decreased PTS (p-value between 0.1 and 0.9, decrease in empty arenas): 18 players
- Non-significantly increased PTS (p-value between 0.1 and 0.9, increase in empty arenas): 10 players
- Significantly decreased PTS (p-value <= 0.1, decrease in empty arenas): 19 players
- Significantly increased PTS (p-value <= 0.1, increase in empty arenas): 5 players

4.2.3. Test 3: Proportion Test for Performance by Free Throw Percentage per Player for Playoff and Non-Playoff Games

H0: The mean BPM in playoff and non-playoff games are the same. ($p_{playoffFT\%} - p_{non\text{-}playoffFT\%} = 0$)
Ha: The mean BPM in playoff and non-playoff games are different. ($p_{playoffFT\%} - p_{non\text{-}playoffFT\%} \neq 0$)
Alpha: 0.1

4.2.3.1. Results: Number of players in each category

- Nearly equal FT% (p-value >=0.9): 14 players
- Non-significantly decreased FT% (p-value between 0.1 and 0.9, decrease in empty arenas): 21 players
- Non-significantly increased FT% (p-value between 0.1 and 0.9, increase in empty arenas): 17 players
- Significantly decreased FT% (p-value <= 0.1, decrease in empty arenas): 3 players
- Significantly increased FT% (p-value <= 0.1, increase in empty arenas): 1 player

4.3. INSIGHTS:

In terms of BPM, the majority of players tested did not have a significant difference between performance in playoff games and non-playoff games. However, player performance was not completely equivalent between the two groups either as the vast majority of players had some sort of insignificant increase or decrease. Furthermore, when looking at player performance in terms of points, we see that around 42% of players tested saw a significant difference in points scored, most of them being decreases in points scored. To contrast this, when looking at free throw percentage, we see that only 4 out of the 56 players tested had a significant difference. 14 players even saw no change at all in their free throw percentage, indicating that players can make free throws consistently no matter the stakes of the game being played. Overall, the only performance metric that had a notable difference in performance were the points scored. BPM was found to be an overall non-significant factor, meaning that if a player were playing like a superstar caliber player in the regular season, they would also play like a superstar caliber player in the playoff. Similarly, free throw percentage production was consistent in both environments, meaning that in general, the sample of players tested were not bothered by the playoff crowd's intensity.

4.4. OUTCOMES:
Success would be measured if we were able to successfully determine whether playoff games impacted performance or not. This would be done if we were able to successfully conduct experiments (i.e., have enough data to work with for the conditions of the tests), as well as have results that we can derive insights from that make sense.

Question 3: How advantageous is a home crowd? Do players and teams lose the home court advantage with a lack of fans? Are there teams that have a stronger home court advantage?

4.1. EXPERIMENTS:
Proportion tests for each individual team and the overall league on winning percentage in six different environments. Significance level of 5% was used. The environments we looked at include:

4.1.1. All Games: Analysis using each game in our dataset
4.1.2. Empty Arena Games: Only games with no fans in attendance
4.1.3. Non-Empty Arena Games: Only games with fans in attendance
4.1.4. NBA Bubble Games: All games played during the NBA Bubble. This condition has no fans and no travel, playing in a completely neutral arena every game.
4.1.5. Non-Bubble Playoff Games: The Bubble was mostly playoff games (suggesting higher quality of competition). Comparisons to playoffs outside the Bubble is needed, since the home court advantage is expected to shrink when playing against stronger competition
4.1.6. Non-Bubble Empty Arena Games: Unlike the Bubble, these games involved away team travel to opposing arenas. This condition includes more factors likely to increase the home court advantage than the NBA Bubble environment.

4.2. QUANTITATIVE:
Each category follows the same hypothesis:

H0: The proportion of games won at home and away are the same ($p_{homewin\%} - p_{awaywin\%} = 0$)
Ha: The proportion of games won at home and away are different ($p_{homewin\%} - p_{awaywin\%} \neq 0$)
Alpha: 0.1

4.2.1. All Games

4.2.1.1. Individual Team Tests
- Significant home court advantages – 30 teams
- Non-significant home court advantages – 0 teams
- Significant away court advantages – 0 teams
- Non-significant away court advantages – 0 teams

4.2.1.2. Overall League Tests
- Overall Home Team Winning Percentage – 55.9%
- Overall Away Team Winning Percentage – 44.1%
- p-value: < 2.2e-16

4.2.2. Empty Arena Games

4.2.2.1. Individual Team Tests
- Significant home court advantages – 1 team
- Non-significant home court advantages – 18 teams
- No specific advantage (p-value of 1) – 7 teams
- Significant away court advantages – 0 teams
- Non-significant away court advantages – 4 teams

4.2.2.2. Overall League Tests
- Overall Home Team Winning Percentage – 54.1%
- Overall Away Team Winning Percentage – 45.9%
- p-value: 0.002

4.2.3. Non-Empty Arena Games

4.2.3.1. Individual Team Tests

- Significant home court advantages – 30 teams
- Non-significant home court advantages – 0 teams
- Significant away court advantages – 0 teams
- Non-significant away court advantages – 0 teams

4.2.3.2. Overall League Tests

- Overall Home Team Winning Percentage – 56.3%
- Overall Away Team Winning Percentage – 43.7%
- p-value: < 2.2e-16

4.2.4. NBA Bubble Games

4.2.4.1. Individual Team Tests

- Significant home court advantages – 1 team
- Non-significant home court advantages – 6 teams
- No specific advantage (p-value of 1) – 12 teams
- Significant away court advantages – 1 team
- Non-significant away court advantages – 2 teams

4.2.4.2. Overall League Tests

- Overall Home Team Winning Percentage – 52.3%
- Overall Away Team Winning Percentage – 47.7%
- p-value: 0.45

4.2.5. Non-Bubble Playoff Games

4.2.5.1. Individual Team Tests

- Didn't test individual teams since playoff teams change every year

4.2.5.2. Overall League Tests

- Overall Home Team Winning Percentage – 54.8%
- Overall Away Team Winning Percentage – 45.2%
- p-value: 0.01

4.2.6. Non-Bubble Empty Arena Games

4.2.6.1. Individual Team Tests

- Significant home court advantages – 1 team
- Non-significant home court advantages – 16 teams
- No specific advantage (p-value of 1) – 8 teams
- Significant away court advantages – 0 teams
- Non-significant away court advantages – 5 teams

4.2.6.2. Overall League Tests

- Overall Home Team Winning Percentage – 54.5%
- Overall Away Team Winning Percentage – 45.5%
- p-value: 0.002

4.3. INSIGHTS:

After performing this series of testing, we can confidently say that there is a home court advantage in the NBA, and that advantage is affected by audience levels. Compared to all non-empty games, there are 29 less teams with a significant home court advantage in empty games played outside of the NBA Bubble. While there is still an overall home court advantage at the league level, the lack of fans in attendance clearly improves the away team's chances of winning on the road. While empty arenas impact the home court advantage, it seems that other factors such as travel and environmental familiarity. This testing answers our second data science question, explaining the home court advantage and the impacts of no fans in attendance. This question can also be answered for specific teams by using the R Shiny dashboard.

4.4. OUTCOMES:

Success would be measured if we were able to successfully determine whether home court advantages impacted the win percentage of teams. This would be done if we were able to successfully conduct experiments (i.e., have enough data to work with for the conditions of the tests), as well as have results that we can derive insights from that make sense.

Overarching Question: How can we use data science techniques to determine the impact that the fans in attendance have on NBA players?

4.1. EXPERIMENTS:

We answered this question by running t-tests and proportion tests on each player in our dataset with 10 minutes played in 25 or more games in both empty and non-empty arenas. We also only used data from the second and third seasons in our dataset in attempts to equalize the sample size in each group. Outliers were removed by using a boxplot, the normality assumption was relaxed due to sample sizes, and we used a Welch's t-test so the assumption of equal variance could be relaxed. The statistics we tested were BPM, Points, and Free Throw %. We also tested overall league point averages and free throw percentages using the same methods.

4.2. QUANTITATIVE:

4.2.1. Test 1: t-test of Player BPM in non-empty vs empty arenas

H0: The mean BPM in each group is equal ($\mu_{emptyBPM} - \mu_{non\text{-}emptyBPM} = 0$)
Ha: The mean BPM in each group is not equal ($\mu_{emptyBPM} - \mu_{non\text{-}emptyBPM} \neq 0$)
Alpha: 0.1

4.2.1.1. Results: Number of players in each category

- Nearly equal BPM (p-value >=0.9): 24 players
- Non-significantly decreased BPM (p-value between 0.1 and 0.9, decrease in empty arenas): 91 players
- Non-significantly increased BPM (p-value between 0.1 and 0.9, increase in empty arenas): 89 players
- Significantly decreased BPM (p-value <= 0.1, decrease in empty arenas): 13 players
- Significantly increased BPM (p-value <= 0.1, increase in empty arenas): 22 players

4.2.2. Test 2: t-test of Player PTS in non-empty vs empty arenas

H0: The mean PTS in each group is equal ($\mu_{emptyPTS} - \mu_{non\text{-}emptyPTS} = 0$)
Ha: The mean PTS in each group is not equal ($\mu_{emptyPTS} - \mu_{non\text{-}emptyPTS} \neq 0$)
Alpha: 0.1

4.2.2.1. Results: Number of players in each category

- Nearly equal PTS (p-value >=0.9): 14 players
- Non-significantly decreased PTS (p-value between 0.1 and 0.9, decrease in empty arenas): 63 players
- Non-significantly increased PTS (p-value between 0.1 and 0.9, increase in empty arenas): 78 players
- Significantly decreased PTS (p-value <= 0.1, decrease in empty arenas): 44 players
- Significantly increased PTS (p-value <= 0.1, increase in empty arenas): 40 players

4.2.3. Test 3: Proportion test of Player FT% in non-empty vs empty arenas

H0: The proportion of made free throws in each group is equal ($p_{emptyFT\%} - p_{non\text{-}emptyFT\%} = 0$)
Ha: The proportion of made free throws in each group is not equal ($p_{emptyFT\%} - p_{non\text{-}emptyFT\%} \neq 0$)
Alpha: 0.1

4.2.3.1. Results: Number of players in each category

- Nearly equal FT% (p-value >=0.9): 55 players
- Non-significantly decreased FT% (p-value between 0.1 and 0.9, decrease in empty arenas): 67 players
- Non-significantly increased FT% (p-value between 0.1 and 0.9, increase in empty arenas): 91 players
- Significantly decreased FT% (p-value <= 0.1, decrease in empty arenas): 10 players
- Significantly increased FT% (p-value <= 0.1, increase in empty arenas): 16 players

4.2.4. Test 4: t-test of overall league point averages (average points per player per game, aggregated for total league) in non-empty vs empty arenas

H0: The mean PTS in each group is equal ($\mu_{overallemptyPTS} - \mu_{overallnon\text{-}emptyPTS} = 0$)
Ha: The mean PTS in each group is not equal ($\mu_{overallremptyPTS} - \mu_{overallnon\text{-}emptyPTS} \neq 0$)

Alpha: 0.1

4.2.4.1. Results:

Mean in Non-Empty: 10.49 PTS
Mean in Empty: 10.58 PTS
p-value = 0.32,
Therefore, we FTR the null hypothesis and assume empty arenas do not impact the average points scored by players overall

4.2.5. Test 5: t-test of overall league point averages (average points per team per game, aggregated for total league) in non-empty vs empty arenas

H0: The mean PTS in each group is equal ($\mu_{overallemptyPTS} - \mu_{overallnon\text{-}emptyPTS} = 0$)
Ha: The mean PTS in each group is not equal ($\mu_{overallremptyPTS} - \mu_{overallnon\text{-}emptyPTS} \neq 0$)
Alpha: 0.1

4.2.5.1. Results:

Mean in Non-Empty: 111.52 PTS
Mean in Empty: 112.41 PTS
p-value = 0.02,
Therefore, we reject the null hypothesis and assume empty arenas positively impact the average points scored by teams overall. While statistically significant, a difference of less than 1 point per game is not significant in practice.

4.2.6. Test 6: Proportion test of overall league free throw percentages in non-empty vs empty arenas

H0: The proportion of made free throws in each group is equal ($\mu_{overallemptyFT\%} - \mu_{overallnon\text{-}emptyFT\%} = 0$)
Ha: The proportion of made free throws in each group is not equal ($\mu_{overallremptyFT\%} - \mu_{overallnon\text{-}emptyFT\%} \neq 0$)
Alpha: 0.1

4.2.6.1. Results:

Proportion in Non-Empty: 0.773 PTS
Proportion in Empty: 0.782 PTS
p-value = 0.00,
Therefore, we reject the null hypothesis and assume empty arenas impact the proportion of free throws made in the league overall.

4.3. INSIGHTS:

Tests 1-3 looked at each individual player and their performance in non-empty vs empty arenas. The results in each test were varied. There wasn't a particular trend in any of the categories, and the players with significant differences weren't constant. Only one player had a significant decrease in each category, and no players had a significant increase in each category. This seems to show that empty arenas affect each player differently, and these differences could be for a variety of reasons. To understand each player and their differences, users are best served using the R Shiny app. Tests 4-6 looked at aggregated league performance. There was no statistical difference in average player points per game, however there were significant increases in empty arenas for team points per game and free throw shooting percentages. These significant increases were overall quite small. This suggests that, while team points per game and free throw shooting percentages are consistently shown the increase in empty arenas, these increases are relatively small and do not affect the NBA's product significantly.

4.4. OUTCOMES:

Success would be measured if we were able to successfully use data science techniques to answer our overarching question. Since this is almost like a summary of our other questions, this would be done if we were able to successfully conduct experiments (i.e., have enough data to work with for the conditions of the tests), as well as have results that we can derive insights from that make sense.

5. DISCUSSION

5.1 APPROACH. Overall, we are very satisfied with our approach and progress so far. We have greatly improved our skills in collecting, organizing, cleaning, and analyzing data. We were able to get hands on experience with the rvest package when we collected the data from basketball-reference.com. This step required lots of planning to ensure we collected and organized the data in an accessible format. The final data set came from multiple parts of the website, so proper organization was critical. Cleaning the data wasn't too difficult since basketball-reference.com is a polished source, but we were able to experiment with creating our own variables which ended being useful in the analysis phase. When it came to analyzing the data, we made sure to start with a single player or team so that we understood what we needed to look for before repeating the process across the entire data set. This agile approach was effective because we could iterate our experiments in small chunks, saving time and improving our analysis. While learning R Shiny was a valuable experience for us, an alternative approach could have been creating our dashboard in Power BI. We are more comfortable with Power BI, and we may have been able to explore predictive techniques with our data given the saved time. Predictive analysis, such as using logistic regression to classify empty and non-empty games, could have given us interesting insights on multiple variables.

5.2 FUTURE WORK. For the remained of this term, we intend on finishing our statistical testing and building a functional R Shiny app. Completing these steps is very achievable, and we will be satisfied with the progress we made this semester. We believe our analysis can be a starting point for future insights on the topic. Data sets with play-by-play data would bring in external features that we could not account for. These externalities would be useful in further isolating the impact that audience has on NBA player performance.

6. TIMELINE

Task	**Due Date**
Submit Proposal	February 7, 2022
Present Proposal	February 11, 2022
Complete Data Collection and Cleaning/Organizing	February 18, 2022
Complete Initial Exploratory Data Analysis	March 4, 2022
Submit Progress Report 1	February 28, 2022
Finalize Required Statistical Tests	March 4, 2022
Submit Progress Report 2	March 28, 2022
Complete Statistical Testing	April 1, 2022
Create R Shiny App to Showcase Results	Beta Version: April 8, 2022 Final Version: April 25, 2022
Final Presentation	April 8, 2022
Submit Final Report	April 25, 2022

The listed dates are estimates and may be subject to change. The task list above is also non-exhaustive, so if there was a task that we deem necessary for this research project, it will be added to the timeline and all proceeding dates will be adjusted accordingly. The definitions of tasks (blue) are as follows:

- **Complete Data Collection and Cleaning/Organizing:** This task will be considered complete once we have scraped the data needed to perform our analysis. Accurate, clean, organized data is vital for every part of the project we work on after this. We are currently nearing completion of this portion of the project.
- **Completed Initial Exploratory Data Analysis:** This task will be completed when we have conducted analysis on the data to better understand the information we are working with. This task could potentially allow us to develop follow-up questions for further research. This step is necessary to finalize the statistical testing requirements.
- **Finalize Required Statistical Tests:** This task will help us finalize the statistical testing we want to run on the

data. This step is necessary for us because we might gain some new insights from the Exploratory Data Analysis stage, so this allows us to make a final decision on the types of tests we want to run.

- **Complete Statistical Testing:** We will complete our initial statistical testing by the end of this task. The statistical testing will allow us to determine how empty arenas affected a variety of performance metrics
- **Create R Shiny App to Showcase Results:** Finally, in this task we will showcase our insights in an easy-to-use format. An R Shiny app to summarize the data will be critical in preparing our presentation and finalizing our project.

7. CONCLUSIONS

Through the duration of this project, we expect to learn how to identify the right data source that will most accurately allow us to conduct our research. Additionally, we will also learn how to gather data from multiple sources and compile them into a central dataset using web scraping. Improving our exploratory data analysis skills will be crucial for this project. We will be analyzing a lot of data and it is critical the right EDA steps are taken to ensure success. Moreover, we expect to learn how to select the right statistical tests to answer the research questions that we are asking. While we have some experience in advanced analytics, we are excited to get more hands-on with data that interests us as basketball fans. Exploring different techniques to best present our data will be a fun challenge that we look forward to.

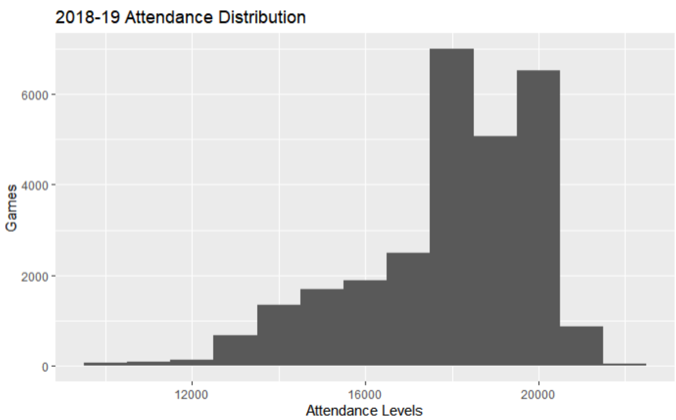
Professionally, we are expecting to improve our presentation skills, and work on presenting data science conclusions with minimal jargon. This skill is critical for workplace success since data scientists tend to work cross-functionally. Being able to explain our technical solution in layman's terms is something we want to excel at.

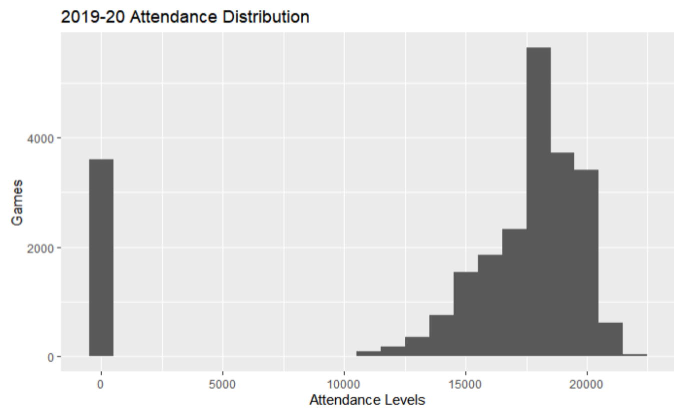
8. REFERENCES

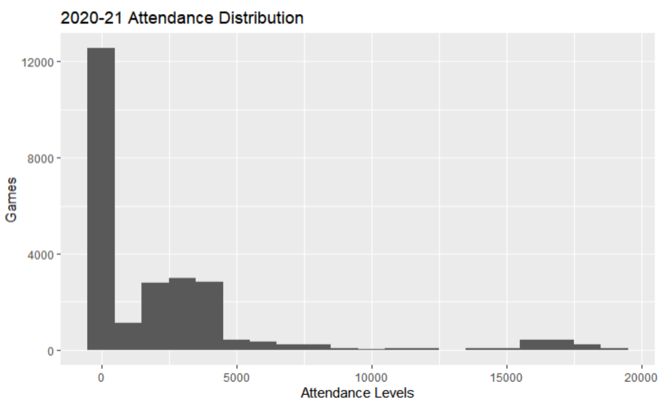
- [1] Daniel Myers, About Box Plus/Minus (BPM), February 2020, Basketball Reference
- [2] Abhas Ricky, How Data Analysis In Sports Is Changing The Game,Jan 2019, Forbes
- [3] David Wharton, No adrenaline buzz? No more choking? How no fans in the stands affects athletes, July 2020, Los Angeles Times
- [4] Paolo Uggetti, How Much Are Fans – or a Lack Thereof – Affecting NBA Games This Season?, Feb 2021, The Ringer
- [5] Joe DeLessio, How Playing in an Empty Stadium Affects Athletes, Apr 2015, The Cut
- [6] Pat Heery, Ranking the NBA Home Crowds, March 2019, Yard Barker
- [7] Logan Markwell, Andrew Strick, and Jared M. Porter, No Fans, No Problem: An Investigation of Audience Effects on Shooting Performance in Professional Basketball, Dec 2021, Journal of Motor Learning and Development
- [8] James Moore and Jody Brylinsky, Spectator effect on team performance in college basketball, June 1993, Journal of Sport Behaviour
- [9] René Böheim, Dominik Grübl, and Mario Lackner, Choking Under Pressure: Evidence of the Causal Effect of Audience Size on Performance , August 2018, Institute of Labor Economics
- [10] Jason Kotecki, Estimating the Effect of Home Court Advantage on Wins in the NBA, 2014, Illinois Wesleyan University
- [11] Bryan Kalbrosky, What is the best advanced statistic for basketball? NBA executives weigh in, Sep 2021, Hoops Hype

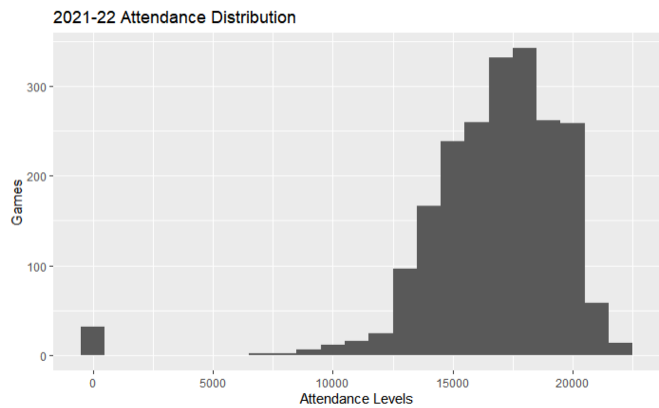
APPENDIX

Histograms of attendance distributions by season

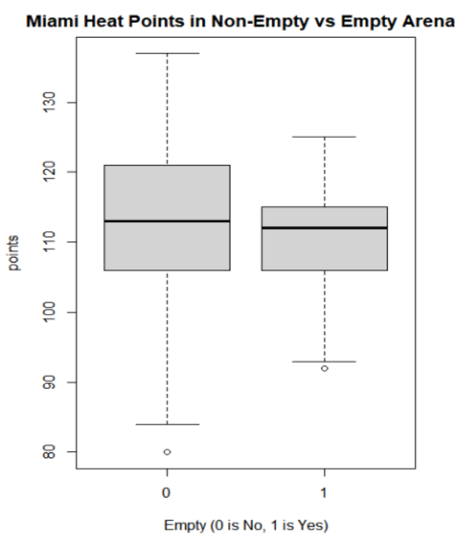


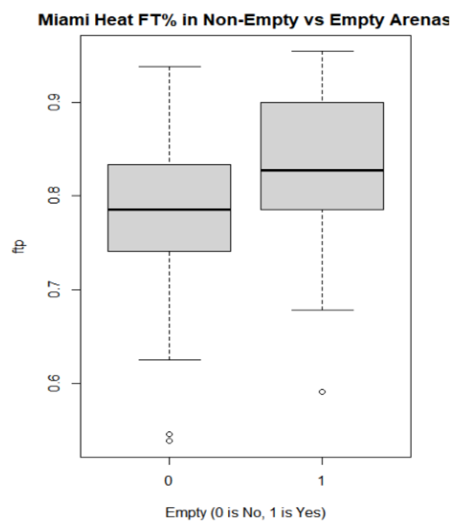


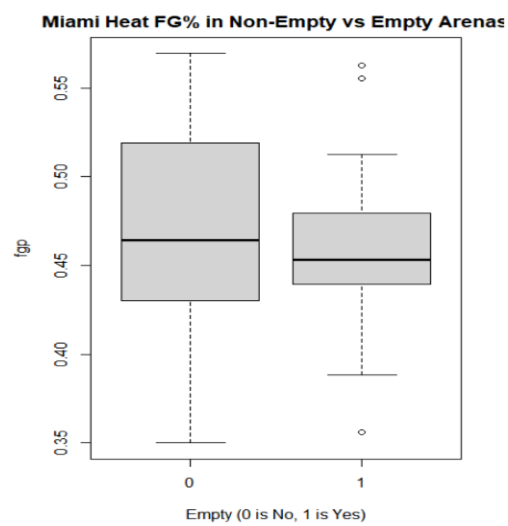




Boxplots of the Miami Heat Data







Link to R Shiny Web App: [R Shiny Web App]

Sample of Attendance Data

Date	Start (ET)	Visitor/Neutral	PTS	Home/Neutral	PTS			Attend.	Notes
Tue, Oct 19, 2021	7:30p	Brooklyn Nets	104	Milwaukee Bucks	127	Box Score		17,341	
Tue, Oct 19, 2021	10:00p	Golden State Warriors	121	Los Angeles Lakers	114	Box Score		18,997	
Wed, Oct 20, 2021	7:00p	Indiana Pacers	122	Charlotte Hornets	123	Box Score		15,521	
Wed, Oct 20, 2021	7:00p	Chicago Bulls	94	Detroit Pistons	88	Box Score		20,088	
Wed, Oct 20, 2021	7:30p	Boston Celtics	134	New York Knicks	138	Box Score	2OT	19,812	
Wed, Oct 20, 2021	7:30p	Washington Wizards	98	Toronto Raptors	83	Box Score		19,800	
Wed, Oct 20, 2021	8:00p	Cleveland Cavaliers	121	Memphis Grizzlies	132	Box Score		15,975	
Wed, Oct 20, 2021	8:00p	Houston Rockets	106	Minnesota Timberwolves	124	Box Score		16,079	
Wed, Oct 20, 2021	8:00p	Philadelphia 76ers	117	New Orleans Pelicans	97	Box Score		12,845	
Wed, Oct 20, 2021	8:30p	Orlando Magic	97	San Antonio Spurs	123	Box Score		16,697	
Wed, Oct 20, 2021	9:00p	Oklahoma City Thunder	86	Utah Jazz	107	Box Score		18,306	
Wed, Oct 20, 2021	10:00p	Sacramento Kings	124	Portland Trail Blazers	121	Box Score		17,467	
Wed, Oct 20, 2021	10:00p	Denver Nuggets	110	Phoenix Suns	98	Box Score		16,074	
Thu, Oct 21, 2021	7:30p	Dallas Mavericks	87	Atlanta Hawks	113	Box Score		17,162	
Thu, Oct 21, 2021	8:00p	Milwaukee Bucks	95	Miami Heat	137	Box Score		19,600	
Thu, Oct 21, 2021	10:00p	Los Angeles Clippers	113	Golden State Warriors	115	Box Score		18,064	
Fri, Oct 22, 2021	7:00p	New York Knicks	121	Orlando Magic	96	Box Score		18,846	
Fri, Oct 22, 2021	7:00p	Indiana Pacers	134	Washington Wizards	135	Box Score	OT	15,407	
Fri, Oct 22, 2021	7:00p	Charlotte Hornets	123	Cleveland Cavaliers	112	Box Score		17,116	
Fri, Oct 22, 2021	7:30p	Toronto Raptors	115	Boston Celtics	83	Box Score		19,156	

Sample of game-by-game player data

Rk	Age	Pos	Date	Tm		Opp		Player	GS	MP	FG	FGA	FG%	2P	2PA	2P%	3P	3PA	3P%	FT	FTA	FT%	TS%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS	GmSc	BPM
1	26-095	C	2019-10-23	OKC	@	UTA	L	Steven Adams	1	28	1	8	.125	1	8	.125	0	0		1	6	.167	.141	3	8	11	2	1	0	2	4	3	-0.9	-11.2
2	26-097	C	2019-10-25	OKC		WAS	L	Steven Adams	1	29	3	9	.333	3	9	.333	0	0		1	2	.500	.354	4	10	14	0	1	1	1	1	7	7.6	-2.1
3	26-099	C	2019-10-27	OKC		GSW	W	Steven Adams	1	25	4	8	.500	4	8	.500	0	0		0	2	.000	.450	5	5	10	3	0	0	1	2	8	8.5	-1.8
4	26-100	C	2019-10-28	OKC	@	HOU	L	Steven Adams	1	27	2	7	.286	2	7	.286	0	0		2	4	.500	.342	2	10	12	1	0	2	1	3	6	5.4	-5.3
5	26-108	C	2019-11-05	OKC		ORL	W	Steven Adams	1	26	5	7	.714	5	7	.714	0	0		1	2	.500	.698	4	7	11	3	0	1	1	1	11	14.0	8.2
6	26-110	C	2019-11-07	OKC	@	SAS	L	Steven Adams	1	27	5	6	.833	5	6	.833	0	0		0	4	.000	.644	2	3	5	3	0	0	3	2	10	6.8	-4.8
7	26-112	C	2019-11-09	OKC		GSW	W	Steven Adams	1	25	6	7	.857	6	7	.857	0	0		1	3	.333	.781	4	4	8	2	1	2	1	0	13	16.5	8.3
8	26-118	C	2019-11-15	OKC		PHI	W	Steven Adams	1	31	3	6	.500	3	6	.500	0	0		1	2	.500	.509	4	3	7	4	1	2	1	2	7	10.7	4.2
9	26-121	C	2019-11-18	OKC	@	LAC	L	Steven Adams	1	27	4	8	.500	4	8	.500	0	0		1	2	.500	.507	0	10	10	6	2	0	2	2	9	11.0	9.9
10	26-122	C	2019-11-19	OKC	@	LAL	L	Steven Adams	1	24	1	3	.333	1	3	.333	0	0		0	0		.333	2	4	6	4	0	0	2	2	2	2.9	-3.0
11	26-125	C	2019-11-22	OKC		LAL	L	Steven Adams	1	27	9	10	.900	9	10	.900	0	0		4	4	1.000	.935	2	4	6	4	0	0	3	4	22	19.4	8.5
12	26-128	C	2019-11-25	OKC	@	GSW	W	Steven Adams	1	28	4	6	.667	4	6	.667	0	0		2	4	.500	.644	4	6	10	5	0	3	2	1	10	14.4	7.5
13	26-130	C	2019-11-27	OKC	@	POR	L	Steven Adams	1	25	4	7	.571	4	7	.571	0	0		0	0		.571	1	5	6	3	0	0	1	1	8	7.6	-1.9
14	26-132	C	2019-11-29	OKC		NOP	W	Steven Adams	1	24	7	11	.636	7	11	.636	0	0		0	0		.636	3	9	12	2	1	2	0	2	14	16.9	9.8
15	26-134	C	2019-12-01	OKC	@	NOP	W	Steven Adams	1	28	6	7	.857	6	7	.857	0	0		5	6	.833	.882	2	8	10	2	1	2	1	3	17	19.5	13.1
16	26-137	C	2019-12-04	OKC		IND	L	Steven Adams	1	29	8	8	1.000	8	8	1.000	0	0		4	4	1.000	1.025	5	4	9	3	0	2	2	2	20	23.0	16.4
17	26-139	C	2019-12-06	OKC		MIN	W	Steven Adams	1	36	9	13	.692	9	12	.750	0	1	.000	4	8	.500	.666	6	5	11	2	2	4	2	4	22	23.2	5.6
18	26-141	C	2019-12-08	OKC	@	POR	W	Steven Adams	1	28	2	6	.333	2	6	.333	0	0		2	2	1.000	.436	4	5	9	4	0	0	1	1	6	8.3	2.9
19	26-142	C	2019-12-09	OKC	@	UTA	W	Steven Adams	1	28	4	4	1.000	4	4	1.000	0	0		3	4	.750	.955	2	11	13	4	0	4	2	1	11	17.3	13.1
20	26-144	C	2019-12-11	OKC	@	SAC	L	Steven Adams	1	28	5	10	.500	5	10	.500	0	0		2	5	.400	.492	5	5	10	4	0	2	1	1	12	13.6	4.8

BPM Scale:

To give a sense of the scale:

- +10.0 is an all-time season (think peak Jordan or LeBron)
- +8.0 is an MVP season (think peak Dirk or peak Shaq)
- +6.0 is an all-NBA season
- +4.0 is in all-star consideration
- +2.0 is a good starter
- +0.0 is a decent starter or solid 6th man
- -2.0 is a bench player (this is also defined as "replacement level")
- Below -2.0 are many end-of-bench players

End of Final Report