COVID-19 Pandemic and the NBA Isolation Zone

Ethan Sheldon
Georgia Institute of Technology
esheldon6@gatech.edu

Brooks Slocum
Georgia Institute of Technology
bslocum3@gatech.edu

Justin Nielson
Georgia Institute of Technology
jnielson3@gatech.edu

Konrad Richter
Georgia Institute of Technology
krichter6@gatech.edu

1 INTRODUCTION

Statistical models to predict player and team outcomes is of great interest to sports teams, TV networks, advertisers, the sports betting industry, reporters, researchers, fans etc. Therefore, it is a much studied area of data analysis [15]. Especially in popular sports like basketball, large amounts of data are collected and analyzed for team executives and coaches to make decisions on player contracts, trades, style of play, and in-game decisions [3].

In order to keep the game running during a global pandemic, the National Basketball Association (NBA) has implemented a special security concept, the "Bubble". The NBA, in essence, is quarantining itself from the rest of society. It was one of the first professional sports league to discontinue play in mid-March after a player tested positive for COVID-19 and then create a plan to return to play.

The Bubble plan instituted by the NBA was to have 22 teams out of the 30 teams that had a statistical chance to make the playoffs play games beginning in August. All teams would play at the ESPN Wide World of Sports Complex and stay at three different Disney hotels, located at Walt Disney World Resort in Florida. The 22 teams that were invited into the bubble would all play eight "seeding games" that would determine the 16 teams that would go to the playoff rounds. Players, coaches, reporters, and support staff are not allowed out of the bubble for a period of up to three months and family were not allowed in the bubble unless they quarantined beforehand.

These special circumstances changed the traditional everyday life of the players extremely on and off the court. In our project, we want to analyze how the factors that determine player and game outcome in the Bubble have changed without travel [12] [4] or fans at games [1] compared to prior seasons and playoffs using established quantitative methods and metrics. These methods and metrics include but are not limited to basketball possessions, offensive ratings, defensive ratings, plays, per-minute statistics, pace adjustments, true shooting percentage, effective field goal percentage, rebound rates, Four Factors, plus/minus statistics, counterpart statistics, linear weights metrics, individual possession usage and individual efficiency, Pythagorean method and Bell Curve method [5] [6] [9].

A normally very influential factor for the outcome of the game is home advantage. Obviously, this factor does not play a big role in the bubble, since all games take place at the same location and do not require travel. We are interested to see if this is also the case for other factors that are less obvious. Maybe even factors that did not play a role at all in the older seasons are now decisive for the game.

2 PROBLEM DEFINITION

Our approach was to identify and research specific metrics which contributed to each specific's game outcome (win/loss) [11]. Once the metrics were identified we compared them to the previous five seasons (2015-2020). The end goal being how has the global pandemic and an isolation season without travel or fans affected the NBA.

3 INITIAL FINDINGS

Using box scores statistics, basic trends are evident within the Bubble. Offensively, scoring has gone up, shooters have done better than previous playoffs, and free throws are up by 28% using a difference-in-difference calculation to compare the past 5 post-seasons to regular seasons. Using the same difference-in-difference measure implied points per game due to 3 point shooting are up 128%, due to a record high in attempts and near record shooting percentage. In a lot of ways, the

Bubble has exacerbated already existing trends in how offenses run. Specifically, teams are playing in a more analytically optimal way shooting more three-pointers and free throws. Defensively, there is scoring evidence to support that defenses are worse, which may be in part or entirely due to the rise in fouls, free throws, or points attributable to 3 point shots. Teams that are younger also performed better than expected in the Bubble. Overall, the trends that have already existed in terms of how teams are playing and composing their rosters were simply taken further within the context of the Bubble.

4 PROPOSED METHOD

To perform our analysis we have identified and researched five distinct parts of the game: scoring and scoring distributions, hustle and defense, team cohesion, team make-up, and home court advantage(and travel). For each of the five factors identified we will perform a two-step process outlined below as well as analyze the factors themselves using the most appropriate analytical framework.

For the classification problem many different algorithms can be tried out. From simple procedures, like a logistic regression or a decision tree, to complex procedures like XGBoost, random forest or even neural networks. It should be noted that although the prediction accuracy may be better with complex models, the interpretation of the influencing factors is more difficult than with simple algorithms. In a logistic regression, the influence of the parameters can be assessed by the coefficients and the p-values. In a random forest model it is possible to evaluate the variable importance, but it only reflects the absolute influence of the variable and not the direction in which it acts. For the more complex procedures it might be helpful to include so-called SHAP values in the analysis. They are a helpful tool to understand the reasons for a certain model prediction. For tasks related to data visualization we rely on Tableau on the one hand and on various Python packages on the other. Especially Seaborn and the Pandas plotting functions have proven to be very helpful for creating fast and easy visualizations. For more complex and interactive visualizations we use bokeh and Plotly. The separate visualization software Tableau allows us to create interactive dashboards based on an exported data set and bring the data to life.

5 DESIGN OF EXPERIMENTS

Using each of the five factors there are really two problems to solve. The first is a classification problem inside the bubble- simply put what measurables contributed to determining the outcome (win or loss) of games? The second problem is, are the measurables identified in part one the same things that have contributed in previous iterations of the NBA playoffs. Or, if they are different, why so and can that be directly attributed to the environment of the bubble itself. Within that two-step framework we will analyze the five factors in different statistical ways.

For data, we used the NBA_API and basketball-reference.com. There are approximately 133 different endpoints and each endpoint varies from 12-60+ data-points with 10 thousand to 100 thousand plus records. Altogether, 80 teams have participated in the playoffs in the past 5 years and played 412 games total.

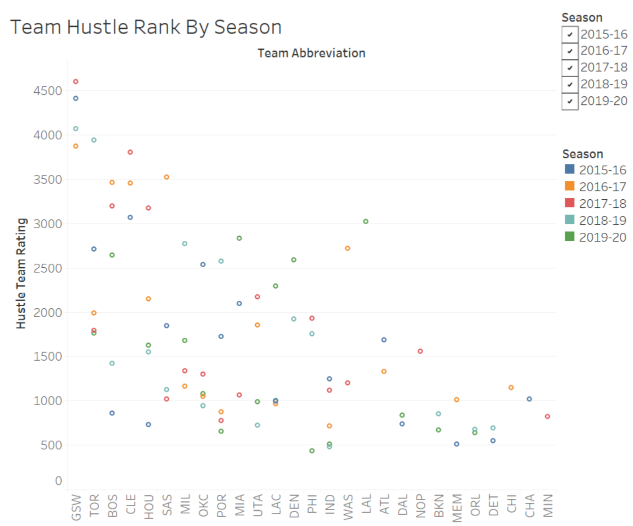

Figure 1: team hustle ranks

As you can see in figure 1, there is a good correlation of hustle stats ranking and the eventual NBA champion each season. Hustle team rating[2] is a weighted calculation of box score hustle and defense stats into a single combined data point: 2-Point Field Goal Contested: A contest decreases the probability of a 2-Point Field Goal going in by .05 x 2.0 Point Value of a 2-Point Field Goal = .1 Hustle Points. 3-Point Field Goal Contested: A contest decreases the probability of a 3-Point Field Goal going in by .05 x 3.0 Point Value of a 3-Point

Field Goal = .15 Hustle Points. Deflection: We estimate that the team whose player deflected the ball has a .25 chance of recovering the ball x 1.0 Expected Point Value for gaining possession = .25 Hustle Points. Loose-Ball Recovered: Recovering a loose ball increases the probability of possession by .5 x 1.0 Expected Point Value for gaining possession = .5 Hustle Points. Charge: 1.5 Expected Point Value for creating a turnover when the opponent otherwise would have likely taken a very good shot (as most charges take place near the basket) + .25 Expected Point Value by drawing a foul on the opposing player (gets team closer to the bonus and limits opposing player's ability to play) = 1.75 Hustle Points.

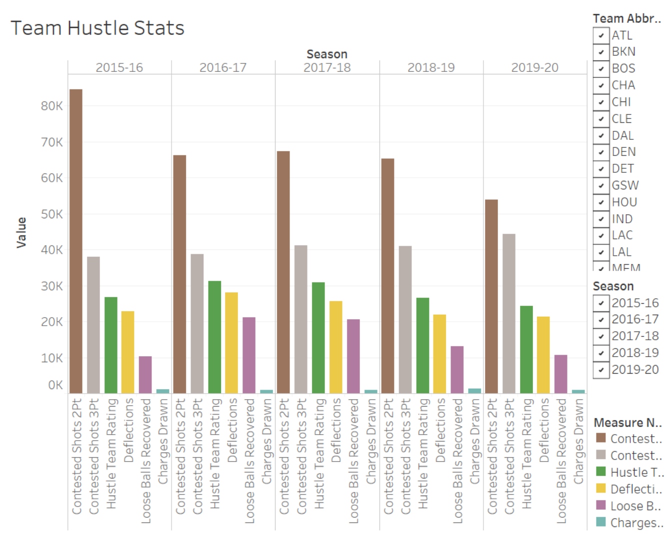

Figure 2: team hustle stats

Total Points and Hustle Rating By Season

	Season				
	2015-16	2016-17	2017-18	2018-19	2019-20
PTS	189,716	184,560	181,086	190,064	186,155
Hustle Team Rating	26,733	31,275	30,857	26,488	24,258
Hustle Team Rating /PTS	0.14	0.17	0.17	0.14	0.13

PTS, Hustle Team Rating and Hustle Team Rating /PTS broken down by Season. The data is filtered on variance of Plus Minus and Team Name. The variance of Plus Minus filter includes everything. The Team Name filter has multiple members selected.

Figure 3: total points and hustle stats

However, when looking at hustle stats in figures 2 and 3, over the past five seasons it saw a sharp decline in the bubble playoffs of 2019-20 and prior 2018-19 playoffs as a percent of total points scored falling from 0.17 to 0.14 to 0.13. One explanation is the rise in 3-pt shot attempts and style of play that has moved to more quick pace that is offense focused over defense. The counter argument is championship teams in the last two playoffs the Toronto Raptors and the Los Angeles Lakers both ranked high in hustle stats. The NBA adage "defense wins championships" still holds true.

6 EVALUATION

The classification models can be evaluated using common metrics such as accuracy or F-score on a test data set. In order to assess the significance of the different predictors, the corresponding standard errors must be examined.

Additionally, the uncertainty of COVID-19 pandemic, and the potential for other emergent diseases, means that there is a high potential for further basketball "Bubbles" at the professional and amateur level [8]. If the evaluation proves to be significant, knowing how game dynamics change in a closed environment could be an advantage for an organization [10].

With regard to the game rotation data, the following two metrics were examined at team and player level:

- the average playing time per player
- the average continuous playing time per player

Furthermore, both metrics were examined for the entire team and exclusively for the Starting 5. It turns out that the continuous playing time per player of the Starting 5 increased for most teams (69%) in the bubble. The average continuous playing time increased by an average of 30 seconds for these players. Related to the players of the whole team, the increase amounts to 24 seconds. The total playing time per player has also increased very slightly but insignificantly. If you look at the data at player level, there are some exceptions to this trend. Probably the most prominent exception is the player LeBron James. During the Bubble, he played significantly more rotations, so his average continuous playing time was almost halved from 16 to 9 minutes, see figure 4. Unfortunately, none of the analyzed variables before or during the bubble showed a significant predictive power for the outcome of the game. Logistic regression models were used to model the game outcome, but the p-values of all predictors were very high.

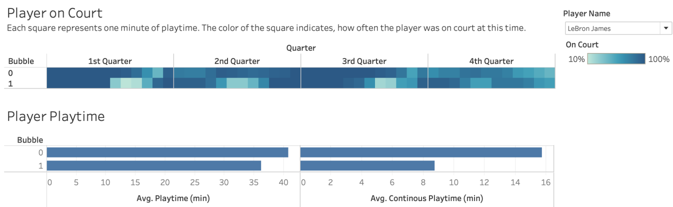

Figure 4: LeBron James on court statistics

To analyze the impact of age, linear regression and difference in difference calculations were used to assess the difference in how aged impacted the regular season vs. post-season results of each team [14]. Across the past 5 years age has been a strong predictor of regular and post-season performance. Older teams are better than younger teams with each year of average roster age contributing over 1.3 wins per year. This year was no different and the bubble was no different with a coefficient of 1.2 and the Los Angeles Lakers (29.1) winning the championship. Of note, however, was the performance of two younger teams the Denver Nuggets (25.9) and Boston Celtics (25.1) who reached the Conference Finals in each division. As well as the shortcomings of veteran favorites Los Angeles Clippers (28.1), Houston Rockets (29.9) and Milwaukee Bucks (29.5). On the individual level, the disparity between the ages of the best players has never been greater. Compare Luka Doncic and Lebron James, respectively among the youngest and oldest major contributors in the bubble and by every measure among the best, with Player Efficiency Ratings of 30.33 and 28.3 respectively, or third and fifth overall. In conclusion, analysis of player age revealed that the player and team age projection curves are flattening potentially because individual players are reaching their peaks earlier and can maintain them for longer. Using the same statistical methodology teams seem to be adding more wins to their rosters through free-agency and trades than ever before. The Miami Heat, for example, went from 99% roster continuity between 2018 and 2019 to just 54% between 2019 and 2020. Overall, the Bubble showed that there are two ways to compose a roster. A younger team via the draft or an older team via free agency and trades. Older teams are better but not all franchises can pursue this strategy due to the allure of each franchise and city.

The shot selection analysis indicated shot distance increased by 6% to 14.61 feet. As indicated in figure 5, the shot location also switched back to the left side of the court opposed to the right side in the past 3 seasons. The Classification model KNN for the Bubble produced a 2% decrease in accuracy compared the the previous 4 seasons being exactly the same. The shot analysis suggests the move to farther shots decreases the shot accuracy. Lastly, the move to the left hand side (defensive perspective) indicates the majority of offensive players (right handed) either chose or defensive strategies were focused on shifting their opponent to the left-hand side of the court.

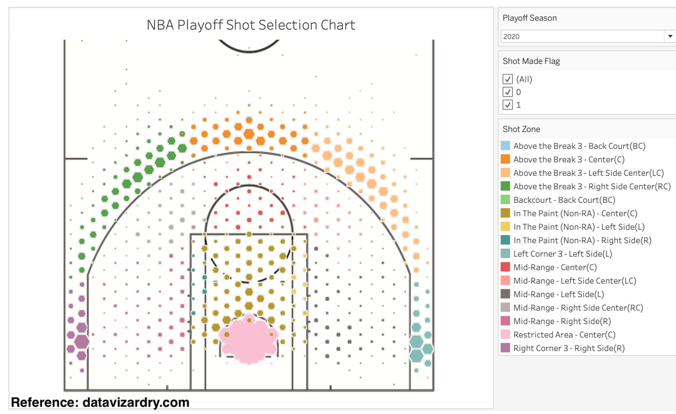

Figure 5: shot selection chart

7 INNOVATIONS

One innovation is using hustle stats as a gauge of player value to the team in winning the NBA championship [2].

As you can see in figure 6, the top hustle players each playoffs have generally been members of the winning NBA championship team. Having multiple players on the hustle stats ranking is a good indicator of teams winning playoff games and the championship over the past five seasons.

For example, Draymond Green, Klay Thompson and Stephen Curry, who all play for the Golden State Warriors, were the top hustle players in the 2017-18 playoffs and ended up with them winning the championship. However, in the 2018-19 playoffs, while the same three GSW players ranked high in hustle stats two players Kyle Lowry and Kwahi Leonard from the eventual championship team the Toronto Raptors where also near the top of the hustle rankings. Injuries in the NBA Finals

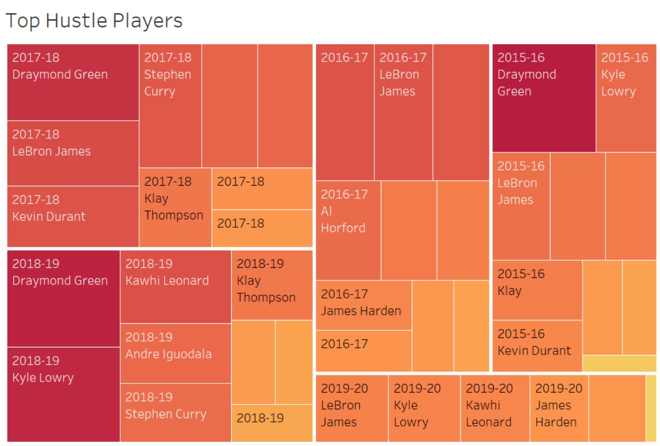

Figure 6: player hustle ranks

to Stephen Curry and Klay Thompson played a critical role in the Warriors not repeating as champion.

More recently, the Los Angeles Lakers won the championship in the 2019-20 bubble playoffs behind perennial All-Star and four-time league MVP Lebron James. In past seasons, Lebron James ranked high in hustle stats with his prior team the Cleveland Cavaliers.

Based on analysis of hustle stats including contested shots, deflections, screen assists, box outs, charges drawn and loose balls recovered were generally more predictive of outcomes in bubble vs. prior seasons using classification algorithms including logistic regression, SVC, KNN, random forest. The average cross validation f1-scores of the classification models ranged from a low of 0.49 to a high of 0.64 with logistic regression, SVC and random forest models performing the best and the KNN model performing the worst in both bubble and prior seasons.

In addition, the top hustle players were highly correlated to wins in the bubble and in prior seasons with championship teams having one or multiple players represented in ranking.

Another innovation is the analysis of home court advantage, no travel, and how we can gauge the overall impact it has on a team winning the NBA championship [7] [13]. For example, the Los Angeles Lakers were the 1 seed in the Western Conference, had home court advantage throughout the entire "bubble" playoffs, and ended up winning the championship in the 2019-20 bubble playoffs.

A third innovation is the development of the player age contribution metric used in conjunction with scoring analysis, hustle analysis, and home court analysis this should give help to determine the role player and team age had on outcomes in the bubble and outside it.

8 CONCLUSION

The NBA Bubble was a unique experience for fans and players. The scoring percentages and scoring increased slightly. Compared to prior seasons where the 2pt shot was the most prevalent, the 3pt shot is now the most common shot. The younger teams were allowed more training during the break and were able to outperform expectations. The old adage of 'defense and hustle wins championships' was demonstrated and is well alive. Knowing how game dynamics change in a closed Bubble environment are a significant advantage for an organization. [10] Overall, the Bubble exhibited and exacerbated many of the same trends in prior seasons.

Distribution of team member effort - all team members have contributed similar amount of effort.

REFERENCES

[1] Alex Bryson, Peter Dolton, J. James Reade, Dominik Schreyer, and Carl Singleton. 2020. *Experimental effects of an absent crowd on performances and refereeing decisions during Covid-19.* Economics Discussion Papers em-dp2020-18. Department of Economics, Reading University. https://ideas.repec.org/p/rdg/emxxdp/em-dp2020-18.html

[2] Jason Concepcion. 2015. How Do You Measure Hustle in the NBA? There's a Stat for That. https://grantland.com/the-triangle/how-do-you-measure-hustle-in-the-nba-theres-a-stat-for-that/

[3] Thomas H Davenport. 2014. Analytics in sports: The new science of winning. *International Institute for Analytics* 2 (2014), 1–28.

[4] Thomas Huyghe, Aaron T. Scanlan, Vincent J. Dalbo, and Julio Calleja-González. 2018. The Negative Influence of Air Travel on Health and Performance in the National Basketball Association: A Narrative Review. *Sports* 6, 3 (2018). https://doi.org/10.3390/sports6030089

[5] Justin Kubatko, Dean Oliver, Kevin Pelton, and Dan Rosenbaum. 2007. A Starting Point for Analyzing Basketball Statistics. *Journal of Quantitative Analysis in Sports* 3 (02 2007), 1–1. https://doi.org/10.2202/1559-0410.1070

[6] Hans Manner. 01 Mar. 2016. Modeling and forecasting the outcomes of NBA basketball games. *Journal of Quantitative Analysis in Sports* 12, 1 (01 Mar. 2016), 31 – 41. https://doi.org/10.1515/jqas-2015-0088

[7] Matt Moore. 2020. Moore: Why are Home Teams Performing So Well in the Bubble? https://www.actionnetwork.com/nba/nba-betting-home-teams-are-outperforming-road-teams-and-lines-in-the-bubble

[8] Allen G.P. Ross, Suzanne M. Crowe, and Mark W. Tyndall. 2015. Planning for the Next Global Pandemic. *International Journal of Infectious Diseases* 38 (2015), 89 – 94. https://doi.org/10.1016/j.ijid.2015.07.016

[9] Jaime Sampaio, Tim McGarry, Julio Calleja Gonzalez, Sergio Saiz, Xavi Alcázar, and Mindaugas Balciunas. 2015. Exploring Game Performance in the National Basketball Association Using Player Tracking Data. *PLoS ONE* 10 (01 2015). https://doi.org/10.1371/journal.pone.0132894

[10] Vangelis Sarlis and Christos Tjortjis. 2020. Sports analytics — Evaluation of basketball players and team performance. *Information Systems* 93 (2020), 101562. https://doi.org/10.1016/j.is.2020.101562

[11] Kai Song, Qingrong Zou, and Jian Shi. 2018. Modelling the scores and performance statistics of NBA basketball games. *Communications in Statistics - Simulation and Computation* 0, 0 (2018), 1–13. https://doi.org/10.1080/03610918.2018.1520878 arXiv:https://doi.org/10.1080/03610918.2018.1520878

[12] Kyle Steenland and James A. Deddens. 1997. Effect of Travel and Rest on Performance of Professional Basketball Players. *Sleep* 20, 5 (05 1997), 366–369. https://doi.org/10.1093/sleep/20.5.366 arXiv:https://academic.oup.com/sleep/article-pdf/20/5/366/8596564/sleep-20-5-366.pdf

[13] Masaru Teramoto and Chad L. Cross. 19 Jul. 2010. Relative Importance of Performance Factors in Winning NBA Games in Regular Season versus Playoffs. *Journal of Quantitative Analysis in Sports* 6, 3 (19 Jul. 2010). https://doi.org/10.2202/1559-0410.1260

[14] Zachary Terner and Alexander Franks. 2020. Modeling Player and Team Performance in Basketball. arXiv:stat.AP/2007.10550 "https://arxiv.org/abs/2007.10550"

[15] Fabian Wunderlich and Daniel Memmert. 2020. Forecasting the outcomes of sports events: A review. *European Journal of Sport Science* 0, 0 (2020), 1–14. https://doi.org/10.1080/17461391.2020.1793002 PMID: 32628066.

List of Figures