NBA Championship Teams and their Starters' Valuable Factors.

1. Introduction

Through this data analysis project our group's main goal was to examine the influence certain variables have on the number of games a player started (GS), within the context of championship teams. Specifically, we want to understand two main concepts in the NBA. First, what do coaches value in a starter, free throw percentage (FT%), rebounds (rebounds), or other variables, and how does this translate to a championship level team? Second, did championship level coaches in previous eras of basketball put more emphasis for players in certain categories rather than others, and how has this changed overtime? To carry out these research questions, we examined statistics from the 1985-1986 Boston Celtics, 1996-1997 Chicago Bulls, 2009-2010 Los Angeles Lakers , and the 2016-2017 Golden State Warriors. Moreover, our group decided to observe the statistics for these specific teams because they all won a championship in their respective seasons. Additionally, these teams are from different eras of basketball, and it would provide us with a diverse source of information for our dataset. Furthermore, an analysis of the distinction in time periods that each team played would provide a very unique perspective into the transformation of NBA basketball. Through our analysis we found that the Boston Celtics and Golden State Warriors used turnovers as a predictor variable which determined which players are likely to start. In contrast, the Chicago Bulls starters were best predicted by assists per game, while the Lakers demonstrated steals as their best predictor.

2. Background & Related Work

Since its foundation in 1946, the NBA has not only witnessed large scale growth in regard to both talent and popularity, but the league has taken massive leaps to model their organization like any major business. Such a transformation was caused by the NBA's increasing emphasis on data analytics. Currently, the NBA serves as both an entertainment, but also a business goliath. The most expensive team in the organization is the New York Knicks, which is surprisingly (Press, T. A., 2019) valued at roughly 4.6 billion dollars (Badenhausen et al., 2020).

That being said, the NBA is of clear interest to the data science community. Furthermore after a review of the literature, there have been attempts by many groups to evaluate NBA data in various ways. Specifically, one study examined the effect aging can have on playing performance in the NBA (Vaci et al., 2019). Furthermore there have been other studies where groups observed game performance between all star and non all star players (Sampaio et al., 2015) and European and Non-European players (Paulauskas, 2018). Although there has been extensive research into the field of NBA data analytics, the literature has demonstrated that there are many approaches to understand the future of basketball rather than the past.

3. Data Collection

3.1 Dataset Overview

In this experiment we collected data from the basketball reference website (Sport Reference LLC). On this site there are career and season-by-season summaries of each players' performance. Initially, this raw data was collected by scorekeepers that work for the National Basketball Association, however Sports Reference LLC has an agreement with the NBA association that enables this organization to provide publicly-accessible, archived data from previous seasons. Furthermore, because our focus is specifically on the aforementioned teams, our group loaded each team's season statistics into a csv. The data was then merged from a csv into 4 dataframes.

The rows in this dataframe represent each player and the columns indicate the player's per game statistics. All of these statistics are for a given season. For example in the first row in fig. 1,

Figure 1

Rk		Age	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	eFG%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS/G
1	Michael Jordan	33	82	82	37.9	11.2	23.1	.486	1.4	3.6	.374	9.9	19.5	.507	.516	5.9	7.0	.833	1.4	4.5	5.9	4.3	1.7	0.5	2.0	1.9	29.6
2	Scottie Pippen	31	82	82	37.7	7.9	16.7	.474	1.9	5.2	.368	6.0	11.5	.522	.531	2.5	3.5	.701	2.0	4.5	6.5	5.7	1.9	0.5	2.6	2.6	20.2
3	Dennis Rodman	35	55	54	35.4	2.3	5.2	.448	0.1	0.3	.263	2.2	4.9	.461	.456	0.9	1.6	.568	5.8	10.2	16.1	3.1	0.6	0.3	2.0	3.1	5.7
4	Toni Kukoč	28	57	15	28.2	5.0	10.6	.471	0.9	2.6	.331	4.1	8.0	.518	.512	2.4	3.1	.770	1.6	2.9	4.6	4.5	1.1	0.5	1.6	1.7	13.2
5	Luc Longley	28	59	59	24.9	3.7	8.2	.456	0.0	0.0	.000	3.7	8.2	.458	.456	1.6	2.0	.792	2.1	3.6	5.6	2.4	0.4	1.1	1.9	3.2	9.1
6	Ron Harper	33	76	74	22.9	2.3	5.3	.436	0.9	2.5	.362	1.4	2.9	.500	.520	0.8	1.1	.707	0.6	1.9	2.5	2.5	1.1	0.5	0.7	1.8	6.3
7	Steve Kerr	31	82	0	22.7	3.0	5.7	.533	1.3	2.9	.464	1.7	2.8	.604	.651	0.7	0.8	.806	0.4	1.2	1.6	2.1	0.8	0.0	0.5	1.2	8.1
8	Jason Caffey	23	75	19	18.7	2.7	5.1	.532	0.0	0.0	.000	2.7	5.1	.534	.532	1.9	2.8	.659	1.8	2.2	4.0	1.2	0.3	0.1	1.3	2.0	7.3
9	Bison Dele	27	9	0	15.3	2.9	7.0	.413	0.0	0.0		2.9	7.0	.413	.413	1.2	1.7	.733	1.6	2.1	3.7	1.3	0.3	0.6	1.2	2.2	7.0
10	Randy Brown	28	72	3	14.7	1.9	4.6	.420	0.1	0.3	.182	1.9	4.3	.437	.426	0.8	1.2	.679	0.5	1.1	1.5	1.8	1.1	0.2	0.8	1.6	4.7
11	Bill Wennington	33	61	19	12.8	1.9	3.9	.498	0.0	0.0	.000	1.9	3.9	.502	.498	0.7	0.9	.830	0.8	1.4	2.1	0.7	0.2	0.2	0.5	2.2	4.6
12	Robert Parish	43	43	3	9.4	1.6	3.3	.490	0.0	0.0		1.6	3.3	.490	.490	0.5	0.7	.677	1.0	1.1	2.1	0.5	0.1	0.4	0.7	0.9	3.7
13	Jud Buechler	28	76	0	9.3	0.8	2.1	.367	0.2	0.7	.333	0.5	1.4	.385	.424	0.1	0.2	.357	0.6	1.1	1.7	0.8	0.3	0.3	0.4	0.7	1.8
14	Dickey Simpkins	24	48	0	8.2	0.6	1.9	.333	0.0	0.1	.250	0.6	1.9	.337	.339	0.6	0.8	.700	0.8	1.2	1.9	0.6	0.1	0.1	0.7	0.9	1.9
15	Matt Steigenga	26	2	0	6.0	0.5	2.0	.250	0.0	1.0	.000	0.5	1.0	.500	.250	0.5	1.0	.500	0.0	1.5	1.5	1.0	0.5	0.5	1.0	0.5	1.5

2P% is Michael Jordan's 2 point field goal percentage per game for the 96'-97' season. For an in depth explanation of all the statistics please reference the appendix.

3.2 Preprocessing

Initially, the csv file did not have a column name for the name of the NBA players. To fix this our group named the column with the names of the players “Player.”

3.3 Dataset exclusions

All actions by players in games are recorded in this dataset. Any gaps in the data were caused by exogenous factors such as physical injuries for specific players, NBA lockouts, or player suspensions. An NBA lockout is a temporary suspension of all NBA games within a season. However, within our datasets there are no examples of such occurrences.

4. Method

In our analysis we normalized our variables and ran multiple linear regression models where we used games started (GS) as the predicted variable and the following averages as the predictor variables for each team:

- Field Goal Percentage (FG%)
- Free Throw Percentage (FT%)
- Total Rebounds (TRB)
- Assists (AST)
- Steals (STL)
- Blocks (BLK)

- Turnovers (TOV)

After identifying the best predictor variable for each team we sorted our original dataframe based on two variables. We sorted it once based on that best predictor variable from most to least. And then we sorted it based on the players' games started (GS) from most to least. After storing those sorted versions into two different data frames, we compared the top 5 players of each dataframe. This further reassures our results of concluding the predictor variable as the best predictor for games started. This is because if the same players with the most games started are also the same players with the highest value for the predictor variable, it indicates that variable as a valid predictor.

5. Results

Through our analysis we found the best predictor variable (4.1) for the predicted variable (GS) for each championship team.

Figure 2 Largest Regression coefficients per team

Season	Team	Statistic	Coefficient
1985-1986	Boston Celtics	Turnovers (TOV)	0.8128367423672338
1996-1997	Chicago Bulls	Assists (AST)	0.7281806917936827

2009-2010	Los Angeles Lakers	Steals (STL)	0.7909445512835194
2016-2017	Golden State Warriors	Turnovers (TOV)	0.8366492444954238

Furthermore, after comparing the top 5 players of two sorted versions of each team's data frame (once based on the "GS" and the other on the team's most predictor variable), we found % of starters who also led the team in the predictor variable for the corresponding team, which is shown in figure 3.

Figure 3

Team	Predictor variable	% of starters who led the team in predictor variable
Boston Celtics	TOV	80%
Chicago Bulls	AST	80%
Los Angeles Lakers	STL	60%
Golden State Warriors	TOV	80%

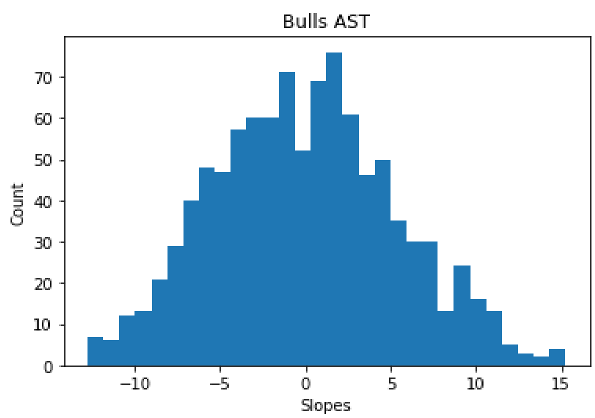
In each team, as our linear regression model predicted, the predictor variable with the largest regression coefficient played the biggest factor in determining a player's number of games started. In addition, the top 5 players with the most games started also held the top 5per game stats for that corresponding predictor variable. For example, taking a look at Figure 3 we can see that 80% (4/5) of the Chicago Bulls' starters (players with the highest amount of game starts on the team within a given season) were also in the top 5 in assists on the team. Essentially this means that assists, with respect to the 96'-97' Bulls, was the greatest indicator for who would start games during the season.

In order to further ensure our conclusions are accurate we performed 1000 random permutations to compare the slope of the relationships to simple randomness. The original slope for the best predictor variable for each team is shown in figure 4. On the following page we have provided a graph of the distributions for the 1000 slopes generated from the linear regressions with "GS" sorted differently 1000 times.

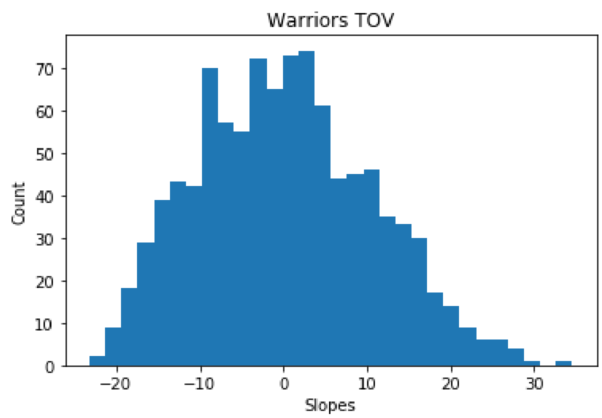
Figure 4. Actual Regression Slopes

Team	Slope	Variable
Boston Celtics	34.594548840358904	TOV
Chicago Bulls	14.912849990728725	AST
Los Angeles Lakers	58.65200225606317	STL
Golden State Warriors	35.09813194608654	TOV

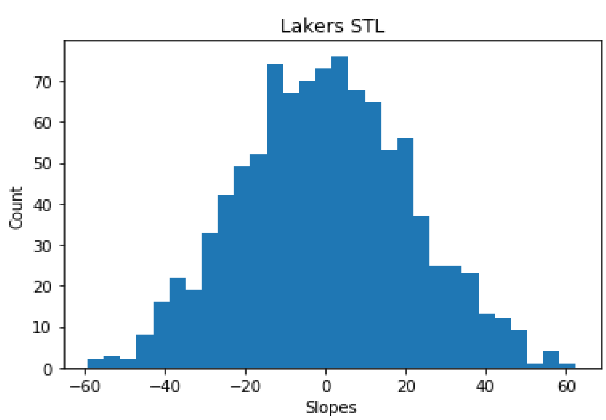
Figure 5. Example Random Regression Slopes



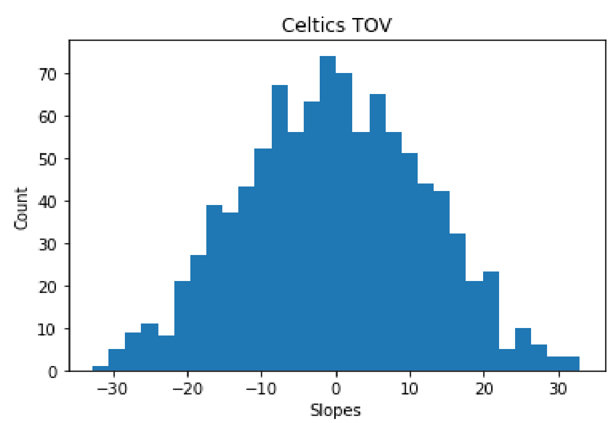
Number of larger slopes: 3



Number of larger slopes: 1



Number of larger slopes: 1



Number of larger slopes: 2

6. Analysis

From the presented data we found two main insights. First, we found that each team has a variable that is a strong predictor for who starts on each respective team. The details of each team will be discussed below. Furthermore, this finding is demonstrated in figure 3. Second, our group found that the number of larger slopes from the random permutations, example results in figure 5, for each team compared to their original linear regression model slopes, figure 4, were all relatively small. From this data we can conclude that our slopes are likely not the result of a random chance and are reasonable.

85'-86' Boston Celtics: The Boston Celtic's greatest predictor for games started was turnovers. This means that, the more turnovers a player had, the more likely they were to start on this respective team, Although it is rather counterintuitive, this result is congruent with KC Jones's coaching style. Jones was known (see Staff, 2017) for being an easy-going coach. Sources indicate he was known to implement a free-reign coaching style, where he let many of the players play as they liked. This coaching style ended up in starters taking more chances and being creative on the court, which in turn resulted in more turnovers. The more player's move the ball around, the more likely they will turn it over. However these turnovers were well rewarded, with the Celtics ranking second in assists in the NBA for that season. This demonstrates that this team liked to move the ball around, and for good reason. One of the team's stars, Larry Bird was 4th in the league in points per game. Clearly though, the Boston Celtics were a dominant team in the league as they won the NBA championship for that season.

96'-97' Bulls: The 96'-97' Bulls' main predictor variable for games started was assists. This can be explained by the Bulls former head coach's coaching style. Phil Jackson heavily emphasized team play during his time at the Chicago Bulls (Williams, 2012). Essentially, Jackson wanted his team to be able to play together, without his help, during games. This coaching style led to cooperative play of the team. In addition, every starter on this Bull's team was a dangerous scorer. Michael Jordan led the league in points per game, Scottie Pippen was a very efficient shooter in the two point range, and Dennis Rodman was a great post player. This can explain why the Bulls were ranked second in points and assists. The talent and cooperativeness of each of these players can provide a good explanation to the corresponding regression model.

09'-10' Los Angeles Lakers: The Los Angeles Lakers' largest predictor for games started was steals. Again, this championship team was coached by Phil Jackson. On this Laker's team was Derek Fisher, Kobe Bryant, and Metta World Peace, who all could perform very well on defense. Overall, the Lakers were second in the league in total rebounds. Though they did not lead other categories, the NBA Lakers were a very good defensive team, which was what led them to the NBA championship.

16'-17' Golden State Warriors: The Golden State Warriors' greatest predictor variables were turnovers. This may be because the team was coached by Steve Kerr, former Chicago Bulls players, who adopted a similar coaching style to KC Jones. Like Jones, Kerr encourages players on his team to self-organize. In other words, he makes his team play together as one unit without having to answer to a chain of command (Wright, 2020). Because of this loose coaching style, the warriors were able to move the ball around more, trying to make more plays. In the process of making plays, the player's are more at risk for losing possession. However, the Warriors were at the top of the league in assists (AST) and field goal percentage (FG%). This means that although the starters turned the ball over, they were very good at moving the ball around and efficient at scoring.

7. Conclusion

What we can glean from this analysis is that each team had a distinct playing style. The starters on the Boston Celtics lost the ball a lot, but they made up for it in efficient shooting by making the team second in the league in FG%. The Chicago Bulls would feed off of each other's talent, which led them to become an all around team. The Bulls were at the top of the league in several categories in the NBA including points and assists. This demonstrates how dominant this team was at the time. In contrast, the Los Angeles Lakers were not a very good offensive team however, they made up for this with their great defense. The team was at the top of the league in rebounds. Lastly, the

Golden State Warriors starters did turn the ball over more than their benchmates. However, they balanced the scale out by making the team best in the league in FG% as well as assists and steals. Specifically, what this data tells us is that, to be an NBA championship team, giving your players freedom to be creative and make mistakes with the ball is important. This will improve their performance in other areas. We know this from the linear regression models for the Celtics and the Warriors. From the Bulls we learned that assists can improve other performances including field goal percentage and points. From the Los Angeles Lakers regression model demonstrated that you don’t have to be an efficient shooting team, as long as you tighten up defensively. To conclude, it is difficult to say what variables can predict the starters on each Championship team for different eras, as more analysis is needed from each decade. However, we found that different coaching styles contribute to winning the championship by taking very distinct approaches to playing basketball.

Acknowledgements

Badenhausen, K. (2020, February 15). NBA Team Values 2020: Lakers And Warriors Join Knicks In Rarefied $4 Billion Club. Retrieved from https://www.forbes.com/sites/kurtbadenhausen/2020/02/11/nba-team-values-2020-lakers-and-warriors-join-knicks-in-rarefied-4-billion-club/#572fadbe2032

Press, T. A. (2019, February 14). Knicks End Franchise-Worst Losing Streak With Win Over Hawks. Retrieved from https://www.nytimes.com/2019/02/14/sports/basketball/knicks-hawks.html

Paulauskas, R., Masiulis, N., Vaquera, A., Figueira, B., & Sampaio, J. (2018). Basketball Game-Related Statistics that Discriminate Between European Players Competing in the NBA and in the Euroleague, *Journal of Human Kinetics*, *65*(1), 225-233. doi: https://doi.org/10.2478/hukin-2018-0030

Sampaio J, McGarry T, Calleja-González J, Jiménez Sáiz S, Schelling I Del Alcázar X, Balciunas M. Exploring Game Performance in the National Basketball Association Using Player Tracking Data. *PLoS One*. 2015;10(7):e0132894. Published 2015 Jul 14. doi:10.1371/journal.pone.0132894

Staff, N. B. A., & NBA.com. (2017, August 24). Legends profile: K.C. Jones. Retrieved from https://www.nba.com/history/legends/profiles/kc-jones

Vaci, N., Cocić, D., Gula, B. *et al.* Large data and Bayesian modeling—aging curves of NBA players. *Behav Res* 51, 1544–1564 (2019). https://doi.org/10.3758/s13428-018-1183-8

Wright, D., Uefa, & Pdp. (2020, March 18). What Can Coaches Learn from Steve Kerr? Retrieved from https://playerdevelopmentproject.com/what-can-coaches-learn-from-steve-kerr/

Williams, S. (2012, November 13). 7 Legendary Leadership Lessons from Phil Jackson's Coaching Career. Retrieved from https://insider.catalystleader.com/read/7-legendary-leadership-lessons-from-phil-jacksons-coaching-career

Zuccolotto, P., Manisera, M., & Sandri, M. (2018). Big Data analytics for modeling scoring probability in Basketball: The effect of shooting under high-pressure conditions. International Journal of Sports Science & Coaching, 13(4), 569-589

Appendix

Number	Title	Link	Description
#1	Glossary	https://drive.google.com/open?id=1LnArlbADRI8U6m9LOsHELlh8jid2voAj8p5qt5jNLWQ	In depth descriptions of all statistics.