A Cost-efficient Approach to Building an NBA Team

Professor Akram M. Almohalwas[1], Euntak Jeong[1], Alex Huo[1], Tim Tianyang Yang[1], Minjie Xia[1], and Ignat Kulinka[1]

[1] *Department of Statistics, University of California Los Angeles,*

June 25, 2017

Abstract

While all players in the NBA are extremely talented athletes, many hone their skills to the point at which their expertise cannot be denied and simply demand massive salaries. This paper presents a possible solution to the old of problem of creating the best possible performing sports team while at the same time minimizing the cost of the team. Through a selection-ranking algorithm our team created based on real game data, we can pinpoint player whose talents are impressive yet salaries are low in comparison. For each position on the court, the algorithm identifies cost efficient and prominent players and ranks them compared to each other. As a proof of concept, we create an optimal theoretical team as well as a backup for each position. In addition, to demonstrate the abilities of the picked performance to cost optimized team we use spatial analytics and graphics to create shot charts to compare it to top teams in the NBA such as Golden State Warriors and Cleveland Cavaliers. While the paper deals specifically with basketball, given the right data, the methodology presented can be easily adapted for any team based sport.

1 Introduction

The best combination of five players in the square field is a key to success in a basketball game. The simplest way to create the best team is to recruit the most well-known, high ranking players in the NBA. However, salaries that each team can offer are limited, and often combinations of less known players can be extremely competitive on the field. Thus, each NBA scouter's mission is to find the most affordable players with highest performance possible.

We used a NBA play-by-play dataset of Season 2015-2016 from Big Data Ball website [1]. We determined the key statistics for each position that would separate top performers. Then, we created a ranking system for each position based on selected statistics. Our system allowed us to identify the players with the best performances-to-salaries ratio. With those players, we created our "Fantasy Team". Finally, we created a collection of visual maps of the field for "sweet" and "unsweet" spots for the our team. These plots demonstrate comparative performance of our dream team with other top teams such as Golden State Warriors and Cleveland Cavaliers.

The long-term goals of our research are 1) to establish a proper ranking system for each position according to their roles, 2) to create a low cost and high performance team based on our ranking system, 3) to search for underestimated players who can perform better than players who are in the same salary brackets. In this paper, our specific aims are to 1) introduce the best players in each position with their salaries and 2) demonstrate visual analytics and spatial analysis that can expose differences and similarities between our theoretical team and other top NBA teams. We present a league-wide case study that attempts to answer one simple yet complex question: who will be a part of the highest performing yet budget efficient team in the NBA.

2 Methodology

2.1 Data Cleaning

The first step of the research is data cleaning and feature engineering. For the 2015-2016 season dataset, we exclude the players with Field Goal Attempts (FGA) lower than 300 first so that the statistics of the remaining players are a good indicator of their performance on the field. Then, we select the key features that are needed for the ranking system: player, Field Goal Percentage (FGP), Defense Rebound (DREB), Offense Rebound (OREB), Assist (AST), Turn Over (TOV), Block (BLK), and Field Goal Made (FGM), and assign the variables to a new empty data frame called features. We add the columns DRED and OREB together to get a new features called Rebound (REB).

Since each observation of features records the statistics for one player in one game, we group the players by names and add the statistics together. The resulted data frame "features" then has variables: player names and their sums of key features for season 15 to 16. Then, we engineer new features by dividing each player's sums of key features by the number of games they played. The new features represent the averages of key features for each player. In order to better evaluate the performance of players, we rank the averages of key features respectively. In this way, each player now has seven rankings for each key features. Then, we scrape the salary and position for each player online (reference), and add these two variables to the data frame "features".

Lastly, we divide the data frame "features" into five sections based on the positions of the players. So now we have five data frames: each one is for one position. And each data frame has variables: player names, the position, the sums of key features (DREB, OREB, AST, TOV and BLK), the averages of key features, the rankings, and the salaries.

2.2 Procedure

In order to find our best team, we devise an algorithm to optimize the players' performance constrained on their salary in the 2015-2016 season. The selection algorithm first separates the top 75% of the players in the NBA based on the FGA. The motivation behind this is that we are only interested in the most active players as they would be the best prospects for our fantasy team. Next the players are separated by the positions they played in the 2015-2016 season. Using the statistics given in our data, we pick out statistics that are both accessible and are the most relevant for each of the positions on the team. The statistics used to rank players in each position are presented in Table 1 below.

Position	Statistics
Point Guard	FGP, AST, DREB, TOV
Shooting Guard	FGP, AST, DREB
Power Forward	FGP, BLK, REB
Small Forward	FGP, DREB
Center	FGP, BLK, REB

Table 1: Statistics chosen to rank players in each position

Since we are interested in the typical performance of a player regardless of the time he spends on the court, we rank each player based on the averages of each statistic for the position. This step helps us to make sure that players who play less games are not unfairly penalized. Next, a weighted average of the statistics is calculated in order to guarantee that we are rewarding well balanced players who perform the tasks of their positions well. The weighted average is then recorded for each of the players and serves as their ranking score. The weight of each rank is defined as the statistic divided by the sum of all statistics for that position. Lastly, for each of the positions we create a plot with standardized salary on the vertical axis and the standardized ranking score on the horizontal axis. On this plot, the ideal player is at the origin because it represents a player that is free and has extremely low weighted average ranking score. Thus, to optimize the average performance of a player constrained on their salary we used Euclidean distance to find the player that is the closest to the ideal player at the origin.

3 Results

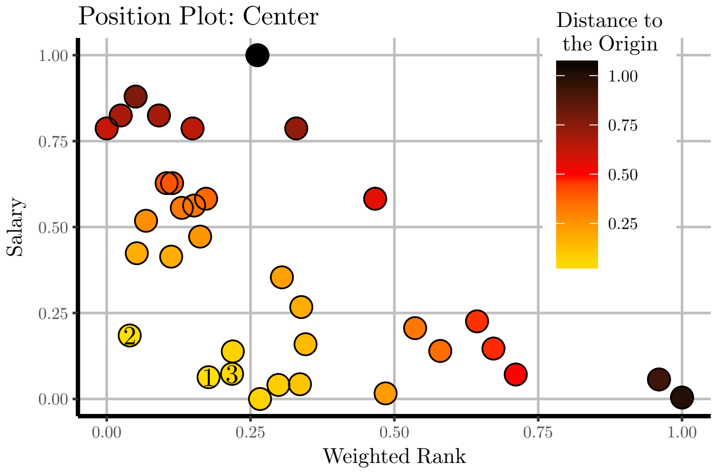
3.1 Selection Plots

The plots below are the results of applying the selection algorithm defined above on the NBA data from the 2015 - 2016 season. The first and second picks for each of the five positions are included in the table below.

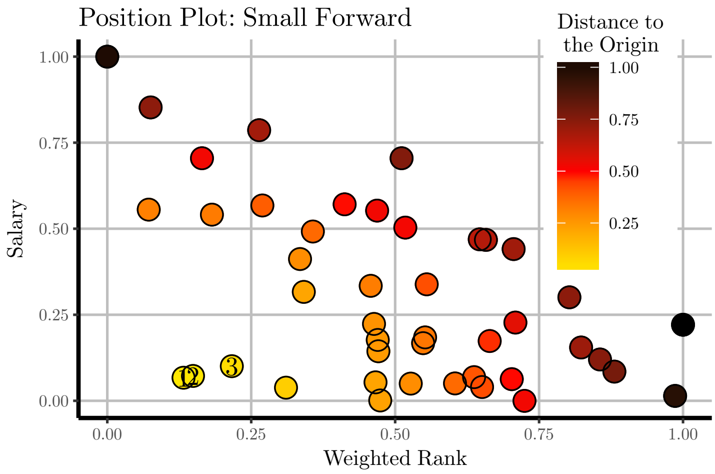
	Center	Small Forward	Power Forward
First pick	Zaza Pachulia	Giannis Antetokounmpo	Nikola Jokic
Second pick	Karl-Anthony Towns	Omri Casspi	Gorgui Dieng

	Point Guard	Shooting Guard
First pick	Darren Collison	CJ McCollum
Second pick	Michael Carter-Williams	Lance Stephenson

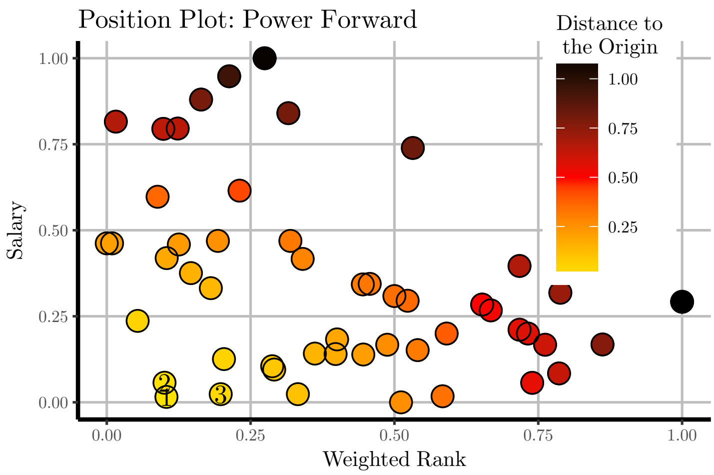
Table 2: First and second picks for the team based on the selection algorithm



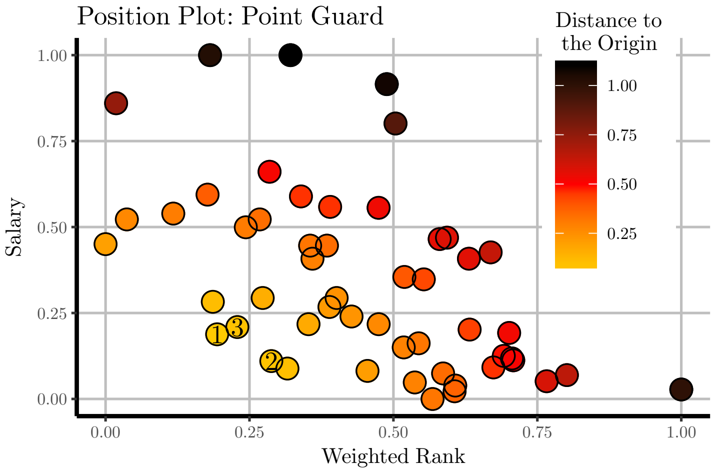
Rank	Player
1	Zaza Pachulia
2	Karl-Anthony Towns
3	Steven Adams



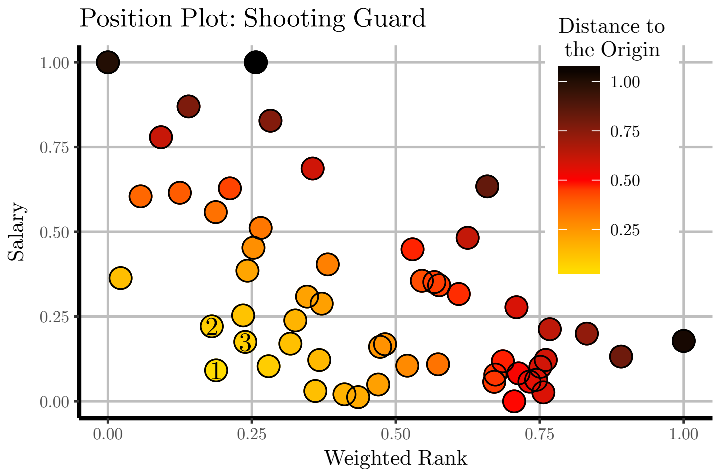
Rank	Player
1	Giannis Antetokounmpo
2	Omri Casspi
3	Thabo Sefolosha



Rank	Player
1	Nikola Jokic
2	Gorgui Dieng
3	David Lee



Rank	Player
1	Darren Collison
2	Michael Carter-Williams
3	Shaun Livingston



Rank	Player
1	CJ McCollum
2	Victor Oladipo
3	Lance Stephenson

3.2 Comparisons to Top NBA Teams

As a means to visualize the favorable and not so favorable shooting areas (so called "sweet" and "unsweet" spots) for each player and for the whole "fantasy team" chosen by our ranking algorithm, we plot the boundaries of an NBA 50 by 47 feet half court based on the official dimensions published by NBA. First, we convert the XY coordinates data into our own coordinates, where we make the center of the whole court the origin and rotate the points in the upper half court around the origin to the lower half court. Thus, we are making the assumption that each of the players will shoot approximately the same on either side of the court. Then, we cut the half court into 50×47 squares (1×1 square foot each), and calculate the density of made or missed shots in each square. Finally, we visualize the sweet and unsweet spots for each player chosen by the ranking algorithm. In order to compare the "spots" with other top tier teams, we plot the graphs containing favorable and unfavorable shooting areas for Golden State Warriors and Cleveland Cavaliers top 2 players in for each position based on FGP. For Golden State Warriors this includes Draymond Green, James Michael McAdoo, Andrew Bogut, Marreese Speights, Brandon Rush, Klay Thompson, Shaun Livingston, Stephen Curry, Harrison Barnes, Andre Iguodala. And for Cleveland Cavaliers: Channing Frye, Kevin Love, Timofey Mozgov, Tristan Thompson, Iman Shumpert, J.R. Smith, Kyrie Irving, Matthew Dellavedova, LeBron James, Richard Jefferson. These plots are included below in the Additional Charts and Tables section. In addition, we included tables that describe the individual, average and total costs of all of the players in the Fantasy Team as well their counterparts, in the Golden State Warriors and Cleveland Cavaliers. As we can see from the shooting areas charts, the three teams are comparable while the cost tables show that the Fantasy Team players are on average twice as cheap as players on the other two teams.

Furthermore, we introduce four features as interpretations for the three plots above. They are "Number of Sweet Spots", "Number of Unsweet Spots", "Number of Total Spots", and "Sweet Spots Percentage". The feature "Number of Total Spots" measures the spread, which is related to the spatial diversity of shooting spots of teams. The higher number total spots means the better the spread is for a team. While the rest three features measure the accuracy. The higher percentage of sweet spots a team has, the more accurate that team is comprehensively.

Team	# Sweet Spots	# Unsweet Spots	# Total Spots	Sweet Spots (%)
Fantasy Team	864	960	1824	47.4
Golden State Warriors	894	945	1839	48.6
Cleveland Cavaliers	821	950	1771	46.4

Table 3: Comparison on the spread and sweet spot percentage

From these data, we can observe that our Fantasy Team has more sweet spots and total spots, and a higher sweet spot percentage than Cleveland Cavaliers. Although our Fantasy Team has less favorable data compared to Golden State Warriors, these numbers of our Fantasy Team are still highly competitive.

3.3 Discussion

As with most statistical analyses, the main limitation is the data used. While implementing our algorithm we worked under the assumption that the data is complete and is a good reflection of players' strength and weaknesses. The limitations of the data is one of the biggest drawbacks for this research. We are constrained by analyzing only features or qualities of players that can be quantified and collected, such as the number of blocks or free throws made. It would be beneficial for a more indepth analyses to be able to find out other indicators of performance such as number of successful screens or team chemistry. Both of which are without a doubt as important to a team's success as the number of three point shots or slam-dunks scored by individual players. Our team identified the lack of reliable data on the performance of players on defence as one of the biggest weaknesses of our methods and most prospective and interesting direction for new research. The ability to gauge the performance of players on both sides of the court gives future researchers and coaches a more complete picture of both the strengths and weaknesses of teams and players alike.

4 Conclusions

4.1 Applications

Basketball is a spatial sport. The team that contains all well known players that cover every spot in the field with high field goal percentage may become the strongest. However, every team financial manager must take budgets into consideration. In this paper, we evaluate all players in season 2015-2016 by different positions and then rank them by our own ranking algorithm that can be a reference for each team manager that he/she may use it to find players with the best performance-salary ratio. We also define the "sweet" and "unsweet" spots, and present the graphs for better visualization. From the graphs, we can easily identify players' shooting performance in 1 by 1 foot square. With our method of visualization "sweet" spots and "unsweet" spot, the team coach may use this approach to design both offensive strategies and defensive ones. For the offensive strategies, the coach could use the visualization to better understand the strength and weakness of the team and train the players accordingly. While for the defensive strategies, the coach can conversely learn the strength of the opponent teams and tell the players to actively defend on the spots. For example, to defend a player, the defensive team can push him to this area and let him shoot, and he will most likely miss the shot.

5 Additional Charts and Tables

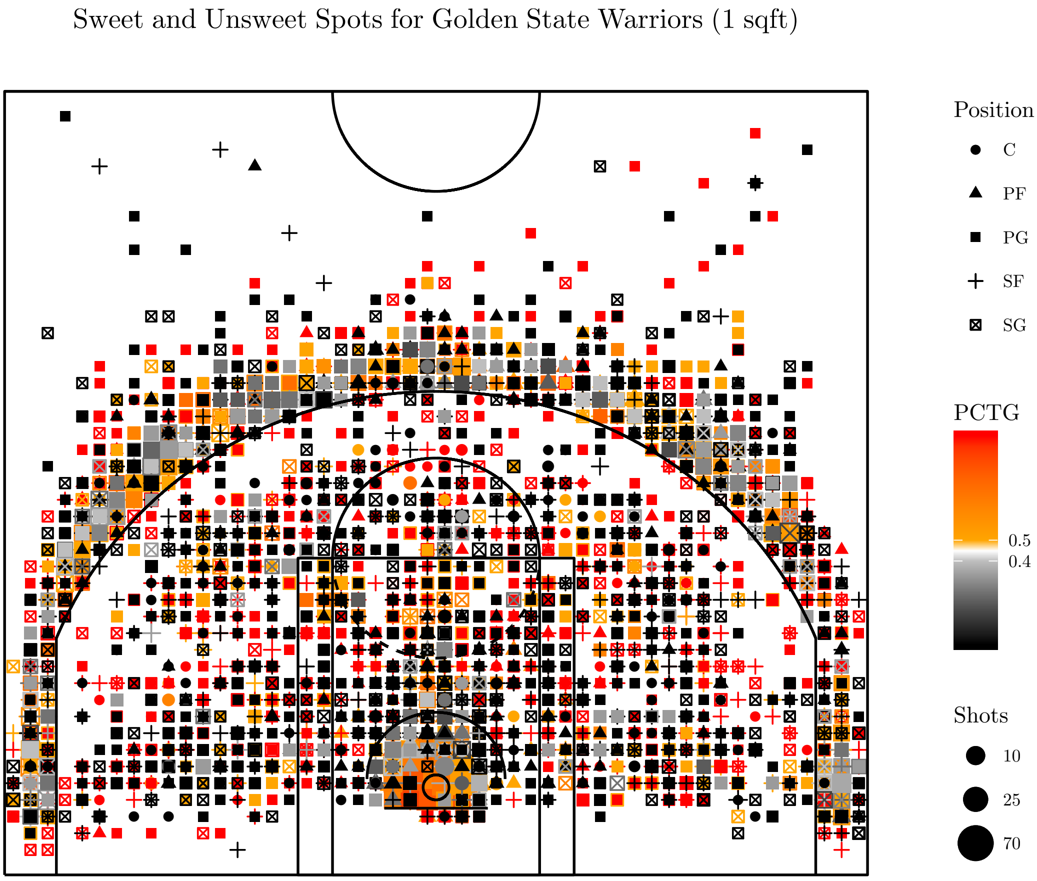


Figure 1: The sweet spots for the Golden State Warriors

Sweet and Unsweet Spots for Cleveland Cavaliers (1 sqft)

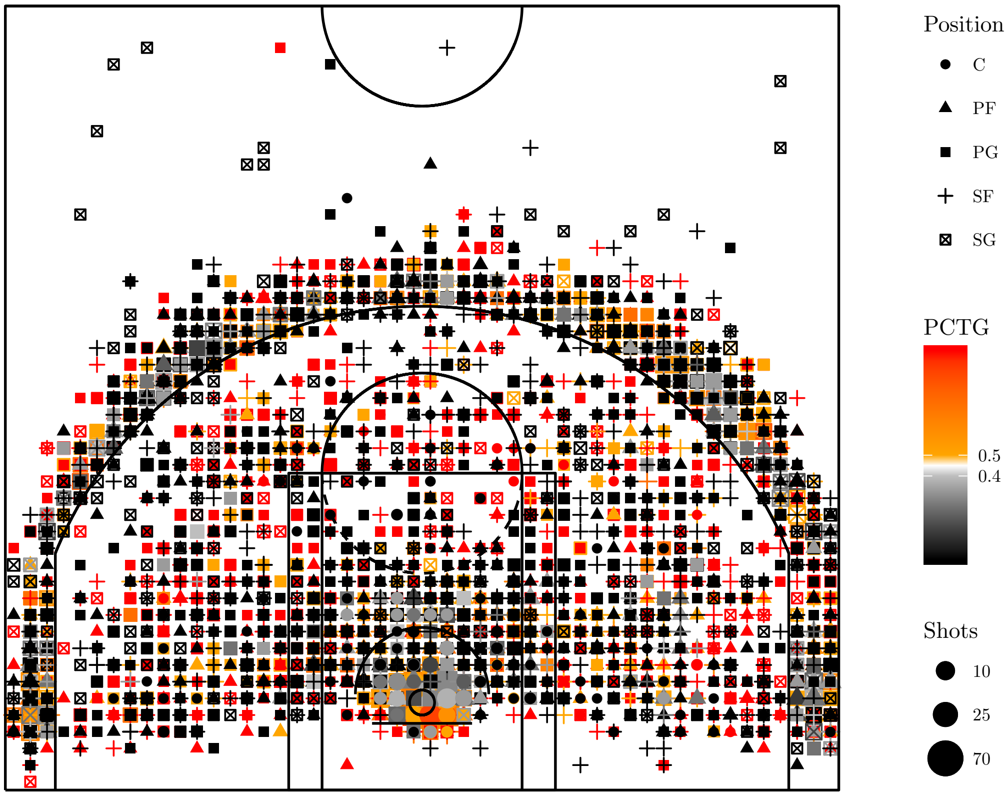


Figure 2: The sweet spots for the Cleveland Cavaliers

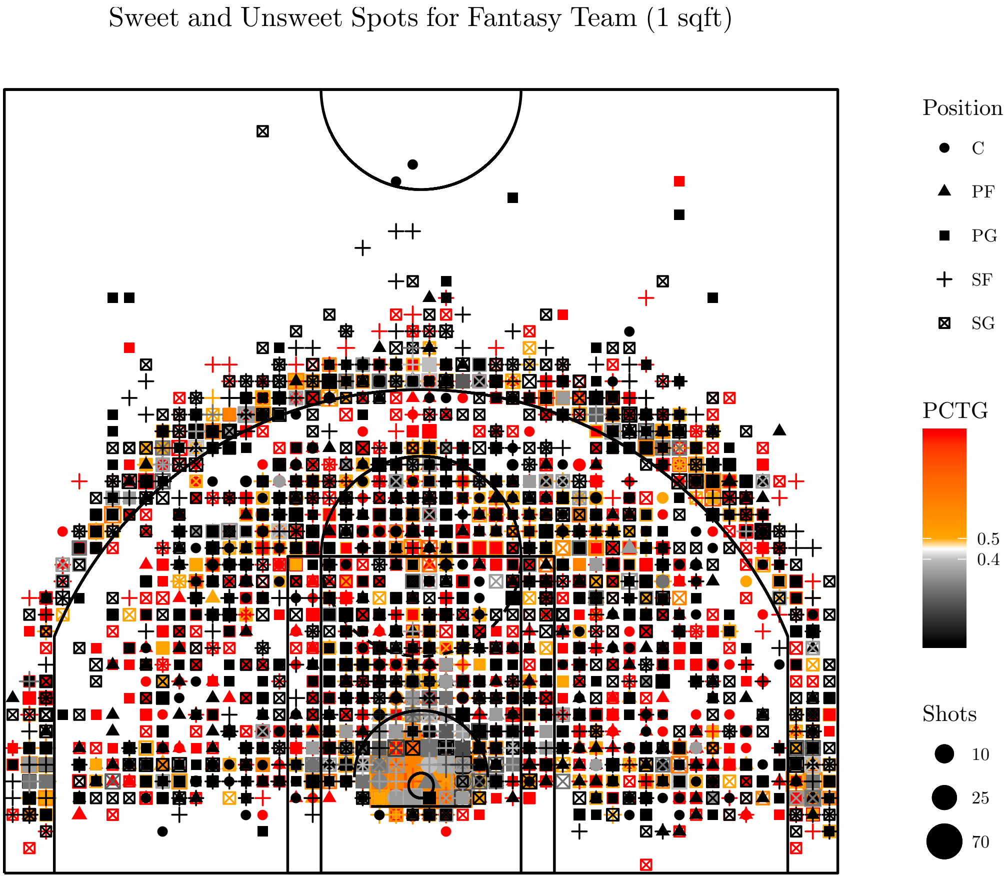


Figure 3: The sweet spots for the Fantasy Team

Player	ESPN Ranking	Salary
Anderson Varejao	179	10,851,659
Andre Iguodala	56	11,131,368
Andrew Bogut	66	11,410,378
Brandon Rush	353	3,500,000
Draymod Green	19	15,330,435
Festus Ezeli	208	7,400,000
Ian Clark	NA	1,015,696
James Michael McAdoo	346	980,431
Harrison Barnes	80	22,116,750
Jason Thompson	207	6,825,000
Kevon Looney	NA	1,142,880
Klay Thompson	16	16,663,575
Leandro Barbosa	248	4,000,000
Marresse Speights	222	1,403,611
Shaun Livingston	117	5,782,450
Stephen Curry	4	2,898,000
Total	2121	122,452,233
Average	151.5	7,653,265

Table 4: Golden State Warriors ESPN Ranking and Salary[2]

Player	ESPN Ranking	Salary
Channing Frye	172	7,806,971
Dahntay Jones	NA	NA
Iman Shumpert	124	9,662,922
J.R. Smith	138	12,800,000
James Jones	362	1,551,659
Jared Cunningham	NA	NA
Joe Harris	340	980,431
Jordan McRae	NA	874,636
Kevin Love	21	21,165,675
Kyrie Irving	18	17,638,063
LeBron James	1	30,963,450
Matthew Dellavedova	198	9,607,500
Mo Williams	213	NA
Richard Jefferson	328	2,500,000
Sasha Kaun	NA	1,333,420
Timofey Mozgov	78	16,000,000
Tristan Thompson	59	15,330,435
Total	2231	148,215,162
Average	159.3571	10,586,797

Table 5: Cleveland Cavaliers ESPN Ranking and Salary[2]

Player	ESPN Ranking	Salary
Zaza Pachulia	209	2,898,000
Karl-Anthony Towns	75	5,960,160
Giannis Antetokounmpo	40	2,995,420
Omri Casspi	242	3,149,524
Nikola Jokic	294	1,358,500
Gorgui Dieng	154	2,348,782
Darren Collison	145	5,229,454
Michael Carter-Williams	103	3,183,526
CJ McCollum	144	3,219,579
Lance Stephenson	149	5,371,672
Total	1555	35,714,617
Average	155.5	3,571,462

Table 6: Fantasy Team ESPN Ranking and Salary[2]

References

[1] "BigDataBall." *Access database friendly and exportable NBA stats, box scores.* N.p., n.d. Web. 04 June 2017. https://www.bigdataball.com/

[2] "HoopsHype" *"These Are the Salaries of All NBA Players This Season."* N.p., n.d. Web. 23 June 2017. http://hoopshype.com/salaries/players/

[3] "The Ultimate Guide to Playing Shooting Guard." *ZUSA Basketball - National Governing Body for Men's and Women's Basketball in the United States.* N.p., 15 June 2011. Web. 04 June 2017.

[4] "The Ultimate Guide to Playing Small Forward." *USA Basketball - National Governing Body for Men's and Women's Basketball in the United States.* N.p., 17 Mar. 2015. Web. 04 June 2017.

[5] "The Ultimate Guide to Playing Point Guard." https://ACTIVE.com N.p., n.d. Web. 04 June 2017.

[6] Gels, Dr. James. "Basketball Fundamentals - Point Guard, Coach's Clipboard Basketball Coaching and Playbook." *Coach's Clipboard Basketball Coaching.* N.p., n.d. Web. 04 June 2017.

[7] "The Ultimate Guide to Playing Power Forward." *USA Basketball - National Governing Body for Men's and Women's Basketball in the United States.* N.p., 24 Mar. 2015. Web. 04 June 2017.

[8] "Ultimate Guide to Playing Center." *USA Basketball - National Governing Body for Men's and Women's Basketball in the United States.* N.p., 25 Jan. 2016. Web. 04 June 2017.

[9] "The Game Court." NBA.com N.p., 17 Sept. 2002. Web. 04 June 2017.