NBA 2018-2019 Season Data Analysis

Presented by: Anastasia Livaditis, Cooper Hird, Jared Dukes, Anna Pigulko, Vince Hrabe

Link to data: [DM\_Datasheet.xlsx](https://drive.google.com/open?id=1wZGqxIxayLJcDjoOFl7wrPxQCkz0JzjY)

# Introduction:

All figures in the data are from the NBA 2018-2019 season. We are not only predicting salary but introducing other statistics that directly correlate with the amount of salary received. Our goal for this project is to discover which statistics are the best predictors of salary throughout this specific NBA season. Using this, we were able to determine which players are properly valued, and which players are often underrated/overrated. By determining which players provide more value than is captured by their salary, we can generalize by aspects of these players to create a system of locating undervalued players.

# Data Dictionary

* USG% (Usage) - The number of times that a given possession ended with something that the given player did.
* TS%(True Shooting %) - A measure that combines 2P%, 3P%, and FT% in a way that measures how efficient a player is at scoring.
* PER(Player Efficiency Rating) - A catch-all number used to measure how good a player was for a given season. Average is 15, data follows a normal distribution.
* WS - Another catch-all number; this is a raw statistic where the default is 0 and each good thing you do adds to the total
* G (Games) - Games the player played any minutes in
* GS (Games Started) - Games a player was in the starting lineup for (generally means you are one of the best 5 players on your team)
* MP (Minutes Played) - Total minutes played over the season
* 2PA (Two Point Shots Attempted) - Number of two point shots a given player takes in a season
* 3PA (Three Point Shots Attempted) - Number of three point shots a given player takes in a season.
* FTA (Free Throws Attempted) - Number of free throws a given player takes in a season.
* REB (Rebounds) - Total rebounds on the season. A rebound happens when a shot is missed and a player grabs the ball.
* AST (Assists) - Total assists on the season. An assist happens when a pass is made to a player who immediately scores.
* STL (Steals) - Total steals on the season. A steal occurs when a defensive player takes the ball away from an offensive player.
* BLK (Blocks) - Total blocks on the season. A block occurs when an offensive player shoots the ball and the defensive player smacks the ball with his hand to stop the shot.
* TOV (Turnovers) - Total turnovers on the season. A turnover occurs when the offense loses the ball without taking a shot
* PF (Personal Fouls) - Total fouls committed on the season. A foul occurs when a player violates an in-game rule and is thus his team is assessed a penalty.
* PTS (Points) - Total points on the season. A point occurs when an offensive player makes a shot; one point is awarded for a free throw, two for a two point shot, and three for a three point shot.
* FG% (Field Goal %) - Average of 2PT% and 3PT%.
* 2P% (Two Point Shot %) - Percentage of two point shots made.
* 3P% (Three Point Shot %) - Percentage of three point shots made.
* FT% (Free Throw %) - Percentage of free throws made.

## Defining Positions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 |
| Point Guard | Shooting Guard | Small Forward | Power Forward | Center |
| PG | SG | SF | PF | C |

### Goal: To determine what factors most correlate to a player’s upcoming yearly salary.

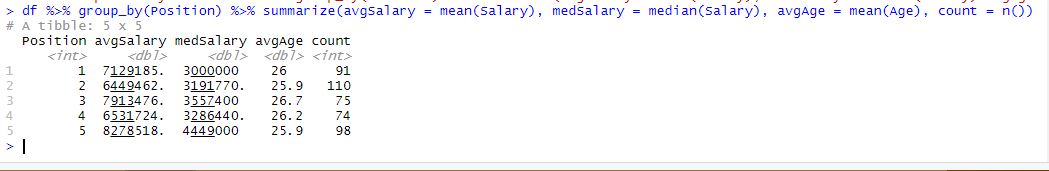
## Notes

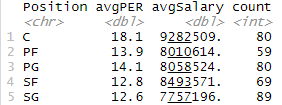
* All figures are from the 2018-19 NBA season (last season.)
* This means we are not predicting salary, rather determining how the other statistics provided correlated to the salary amount the player was currently on. This eliminates the time dissonance from the initial proposal.
* 550 Data Entries in total.
* This data lies outside of the realm of player waivers; if a player agreed to a “buyout” and sacrificed some of his salary in return for his release from the team, it is not recorded in this sheet. Also, if a player was waived with multiple years remaining on his contract, the planned salary from that year will be used as salary.

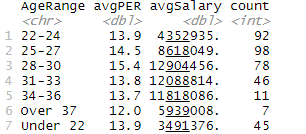
# Initial Predictions:

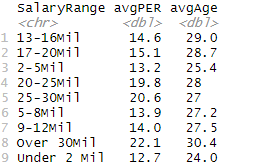
* Heavily Correlated Stats:
  + TS%
  + WS
  + GS
  + MP
  + 3PA
  + FTA
  + G
* Lightly Correlated Stats:
  + USG%
  + PTS
  + PER
  + 3P%
  + 2PA
  + FGA
* Not Really Correlated Stats:
  + 2P%
  + FG%
  + AST
  + REB
* Inversely Correlated Stats:
  + TOV
  + Age

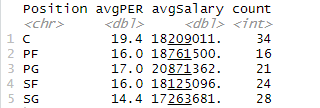
R Insights & Script

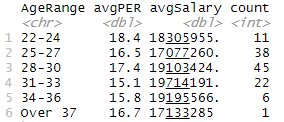


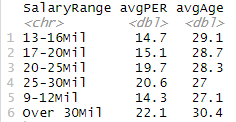
- Average PER and salary by position for entire dataset

- Average PER and salary per age range

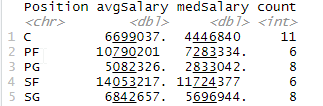
-Average PER and age per salary range

- Average PER and salary by position for salary>$10m

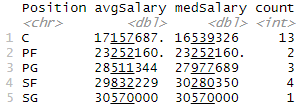
- Average PER and salary by age range

- Average PER and age per salary range over $10m

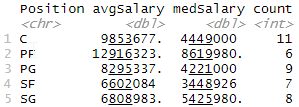
It is worth noting that while both the “20-25Mil” and “Over 30Mil” seem to provide value greater than the ranges below it, there is very little difference in the value provided by players in the “9-12Mil” range and the “13-16Mil” and “17-20Mil” range. It is also interesting to note that the average age of players in the “9-12Mil” range is virtually equivalent to that of the “20-25Mil” range, while players in the “13-16Mil”, “17-20Mil”, and ranges are around two years older on average, and players in the “Over 30Mil” range are on avearge *three* years older on average. It is awfully odd to see what amounts to a 3.4 year difference in average age between players making “25-30Mil” and “Over 30Mil” since those are adjacent ranges.

- Stats for average PER (14-16)

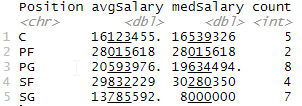
Here we note that across the board, the average salary is less than the median salary, quite considerably so in fact. In fact, at the PG position, we see a 44.3% reduction in when going from median to average salary. This implies a significant skew in this data, with many more players making paltry salaries and only a few making exorbitant salaries.

 - Stats for elite PER (>22)

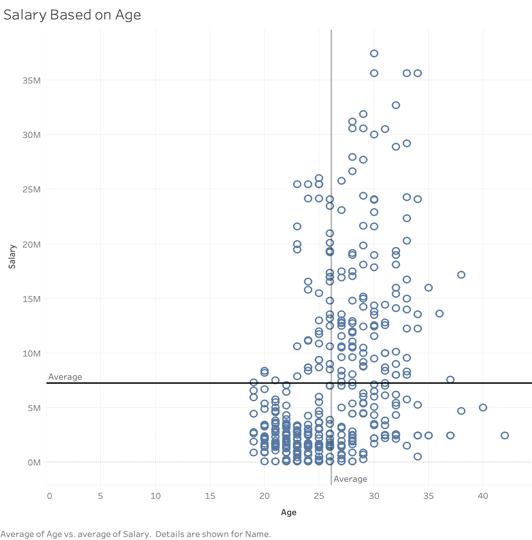
Unlike in the previous category, players with elite PER have a much more normally distributed set of salaries. In fact, we see that the average salary of SF’s is *less* than the median salary - this speaks to the fact that almost all of the elite SF’s are being paid the absolute maximum amount allowed,

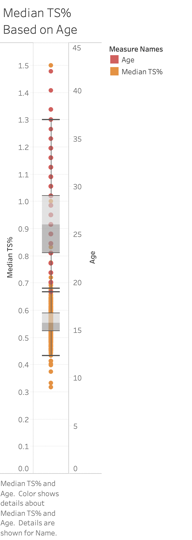
 - Stats for average USG (19-21)

As with average PER, we notice a similar trend of average salary being significantly greater than median salary for average players. Players with average usage at the PF position are paid significantly more than any other position, and C is a clear second; this is likely due to the nature of those positions, which are traditionally less focused on offense and more focused on defense.

 - Stats for elite USG (>28)

The trend of average versus median salary continues for this dataset, outside of the Shooting Guard position - this is likely due to Luka Doncic, who was already performing at an elite USG level in his rookie year, meaning he was on the first year of his rookie contract, making the 168th most money in the league ($6.57 million) despite being a top-30 player in the NBA.





Conclusion:

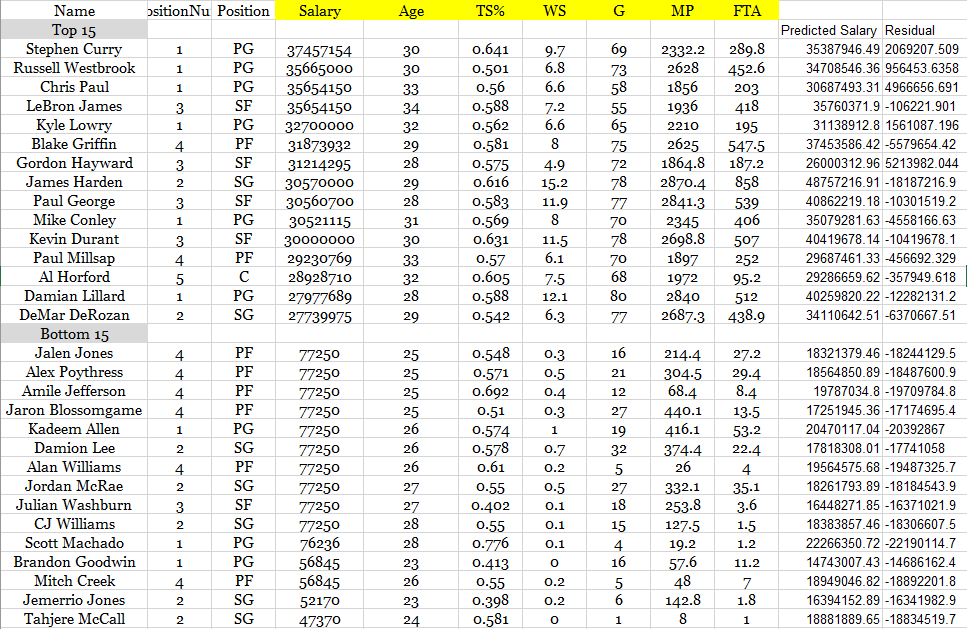
* We were surprised to see center as the highest-paid position on average AND median. Center is often considered the least valuable position in today’s game, but the salaries don’t bear that out.
* The median salary of PG’s is last out of the five positions, but a clear and distinct 3rd in average salary. It appears that there are more outlier salaries in the PG position, which makes sense - there are several notable PG’s who have made.
* We expected SF to be the least common position but was surprised to see that PF and SF are in fact neck and neck. This may mean some traditional PF’s are being labeled as C’s, a clear trend in the modern era. This is even more perplexing given the research I did into 5 star high school recruits (in short, the elite of elite high school athletes) and PF was the most common position for basketball recruits. It would be interesting to study further as to whether it is, as we suspect, that many of these players labeled as PF’s earlier in their career are eventually shoehorned into the C position due to the shifting of the modern game, or whether there is simply a huge proportion of elite PF prospects who do not stick in the NBA.
* 67% of players making over $10 mil next year are between the ages of 25-30. (83 out of 123)
* Proportion of players at each position making over $10 mil
  + PG: 27.1%
  + SG: 31.4%
  + SF: 26.2%
  + PF: 27.1%
  + C: 42.5%
* Proportion of players making over $10 mil at each age range
  + 22-24: 11.9%
  + 25-27: 38.8%
  + 28-30: 57.7%
  + 31-33: 47.85%
  + 34-36: 54.5%
  + Over 37: 14.3%

The two most heavily correlated variables in our dataset appear to be Age and TS%. Age is not necessarily surprising; many factors within the NBA CBA create an environment in which players are more often underpaid on their first contract, and overpaid on later ones. In a way, it is a measure to prevent hasty impatient GM’s from spending huge amounts of money on players that have never even played in the NBA, or have only played a very limited sample. This was a huge deal up until about a decade ago in the NFL, which allowed rookies to negotiate a contract upon being drafted to the team. This culminated in #1 overall pick in the 2010 NFL Draft, Sam Bradford, negotiating a whopping 6 year, $78 million contract, with an astounding $50 million guaranteed - no small feat for a league where even superstars weren’t getting that much guaranteed on a single contract. The NBA, in contrast, has had pre-negotiated rookie scale contracts for some time now, allowing the league to tamp down on contracts in the same vein of the Sam Bradford deal. For reference, Luka Doncic was drafted #3 overall in the 2018 NBA Draft, and now, only a couple of months into his second year in the NBA, is considered one of the faces of the league, one of the best in the game right now. Where does he rank on our list of salaries? A whopping 168th. That’s a few spots down from Lance Thomas of the New York Knicks, who played 33.78% of the minutes Luka did last year, and is not even employed in the NBA this year after being waived by the Knicks and the Nets.

Theory: Joseph W. Harder, in his article *Play for Pay: Effects of Inequity in a Pay-for-Performance Context,* posits that highly paid players will tend to have higher stats in terms of non-scoring performance (more team-based) and underpaid players will tend to have more shots, but score fewer points (more selfish play). Our data backs up this theory as we can see in the TS% stat being highly correlated to salary. The higher the salary, the higher the TS%. This means that higher paid players are more efficient with their shots.

Some key takeaways from this project:

* Focus on youth: As we mentioned earlier, rookie-scale contracts and other CBA rules artificially deflate the value of young players on their first contract; the best deals in the NBA are almost always young stars on rookie contracts.
* Avoid $13-20 contracts: As we noted earlier, there is virtually no discrepancy between performance of guys in the $10-12mil range and both $13-16mil and $17-20mil.
* Spend more on starters: As we noted in the Games and Games Started vs Salary graphs, there is a much stronger correlation between starting games than simply playing. This means finding a guy who starts 30 games is FAR more valuable than a player who simply plays 30 games.
* Avoid giving out big contracts to SG’s: We noted that amongst players paid over $10mil that SG was, on average, providing by far the least amount of value.
* Elite C’s: Given the avg/median is under $20mil, teams should be looking to pay elite centers more like $15-18 million, as opposed to paying a max contract (anywhere from $27-$35 mil depending on age). There are simply more elite centers on the market (getting paid less money, too) than at any other position.



This table shows a snapshot of our regression results. It displays predicted salary of the top 15 paid players along with the bottom 15 paid players. To arrive at this conclusion, we did a few regression analyses until we got down to just statistically significant predictors of salary (age, TS%, WS, G, MP, and FTA). It is clear that within the top 15 paid players, some are overvalued and some are still undervalued. Within the bottom 15 players, they are all shown as vastly undervalued. However, our data does not account for things like name recognition and previous year's performance. There are other factors that cannot be quantified that lead to salary decisions. Though if we go based purely on statistics for specifically the 2018-2019 season, 439 players were undervalued and only 9 players were overvalued. Average amount undervalued was $15,155,033.7, average amount overvalued was $2,138,237.88. In conclusion, this could be viewed as a predictor for future salary changes. For example, if one of the bottom 15 paid players continues to perform consistently to the 2018-19 season stats, they may fulfill their predicted salary. This is especially true if you look at the age difference between these two groups. The bottom 15 generally have a lot of time left in their careers.

Sources

Basketballinsiders.com. (2019). NBA Team Salaries - By Team | Basketball Insiders | NBA

Rumors And Basketball News. [online] Available at:

http://www.basketballinsiders.com/nba-team-salaries-at-a-glance/ [Accessed 30 Sept. 2019].

Basketball Reference. (2019). 2018-19 NBA Player Stats: Per Game |

Basketball-Reference.com. [online] Available at:

https://www.basketball-reference.com/leagues/NBA\_2019\_per\_game.html [Accessed 30

Sept. 2019].

Coon, L. (2019). NBA Salary Cap FAQ. [online] Cbafaq.com. Available at:

http://www.cbafaq.com/salarycap.htm [Accessed 18 Oct. 2019].

Cosmic-s3.imgix.net. (2017). The Collective Bargaining Agreement NBA. [online] Available at:

https://cosmic-s3.imgix.net/3c7a0a50-8e11-11e9-875d-3d44e94ae33f-2017-NBA-NBPA-Collective-Bargaining-Agreement.pdf [Accessed 18 Oct. 2019].

ESPN. (2019). 2019-20 NBA Stat Leaders | ESPN. [online] Available at:

https://www.espn.com/nba/stats [Accessed 30 Sept. 2019].

Gaines, C. (2018). The 24 highest-paid players in the NBA for the 2018-19 season. [online]

Business Insider. Available at:

https://www.businessinsider.com/nba-highest-paid-players-2018-10 [Accessed 28 Oct.

2019].

Harder, J. (1992). *Administrative Science Quarterly*, [online] 37(2), pp.321-335. Available at:

https://www.jstor.org/stable/2393227?seq=1#metadata\_info\_tab\_contents [Accessed 18

Nov. 2019].

HoopsHype. (2019). These are the 2018/19 salaries of all NBA teams. [online] Available at:

https://hoopshype.com/salaries/2018-2019/ [Accessed 24 Sept. 2019].

Nbastuffer.com. (2019). 2018-2019 NBA Player Stats | NBAstuffer. [online] Available at:

https://www.nbastuffer.com/2018-2019-nba-player-stats/ [Accessed 28 Oct. 2019].

Sky Sports. (2019). NBA News, Highlights and Videos | Sky Sports. [online] Available at:

https://www.nba.com/ [Accessed 30 Sept. 2019].

Spotrac.com. (2019). NBA Rankings. [online] Available at:

https://www.spotrac.com/nba/rankings/ [Accessed 30 Sep. 2019].