## NBA Player Value: Win Shares and Salary

#### Where should an organization put its money to generate the best chance at success? A question asked by businesses and professional sports teams alike; with the most accurate answer leading to favorable outcomes. I took a deeper dive into the NBA and analyzed players based on their productivity and their salary. Can we determine a player's true value to an organization in regards to specifically their on court production? What expectations do various salaries bring? Based upon production of a player, can we predict their value? This analysis aims to answer those questions and more to provide insights for organizations to increase their rate of success.

## Sources of Data and Background Information:

#### 

#### It's common for NBA fans to research stats and information regarding teams and players within the organization. The best part and possibly the least appreciated is the fact that there is SO MUCH data publicly available. I (and many others) find themselves on [Basketball-Reference.com](https://www.basketball-reference.com/) more frequent than others as it is a trusted and reliable source of statistical information for players and teams. For this analysis, I utilized NBA Advanced Player Statistics as well as their contract information. This data is available and kept up to date on a consistent basis.

#### One caveat to my dataset is that is is only up to date through 4/6/2019. This means that it did not include all statistics through the entire regular season.

#### 

#### This analysis is largely relying on one particular data point for NBA Players, Win Shares (WS). This number is something that is calculated by Basketball-Reference and you can find more information regarding that by following [this link](https://www.basketball-reference.com/about/ws.html). In short, it is this: Win Shares is a player statistic which attempts to divvy up credit for team success to the individuals on the team.

#### Utilizing player statistics and contract information, we can determine a lot of information regarding a players value to an organization, in theory.

The Process:

Initially, I began with two datasets; one for the statistics for the active players in the 2018-19 NBA season and one for their contracts and salary data. Cleaning the data mostly included getting the columns in their respective data type for analysis purposes. An additional step included removing a header row that was in the tables from Basketball Reference. To identify and remove those rows that were headers, I found a unique identifier to the header row and called upon that to remove those. The next step was to combine the datasets. This was completed by using the NBA Players’ names as a reference point to combine the statistics with contract data.

One of the interesting things to make note of here is that there are times where players are traded during the regular season and this resulted in the same player having statistics for that particular team. I treated these instances as they were, additional data points. Sometimes a player may contribute and play differently on a new team, and so there may be more than one data point for a player if he were traded midseason.

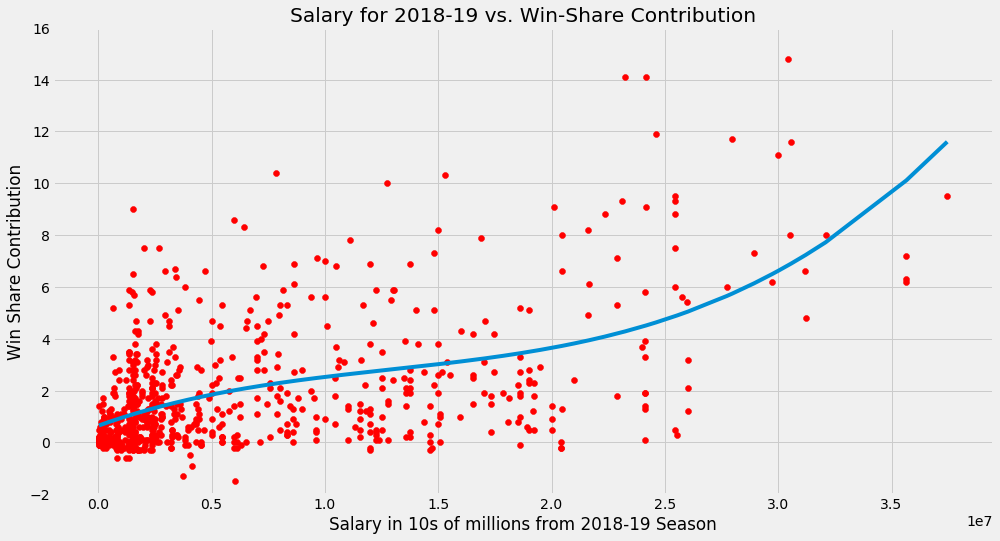
Another notable piece of information regarding the data here is that there were some null values initially after the merging of the two datasets. The number one reason here is that there were players who had short term, 10 day contracts, and may have never played a game. This would result in “contract” data but no statistical information. For these situations, and to make the results as accurate as possible, I removed the players who didn’t have an actual contract (more than 10 days) AND didn’t play any games to accumulate any stats.

What was left was a clean dataset that had null values in the contract section more than anything. There is a simple explanation here: 40% of the NBA Players will be free agents after this season ended. Therefore, there are many players who did not have contract data after this current season.

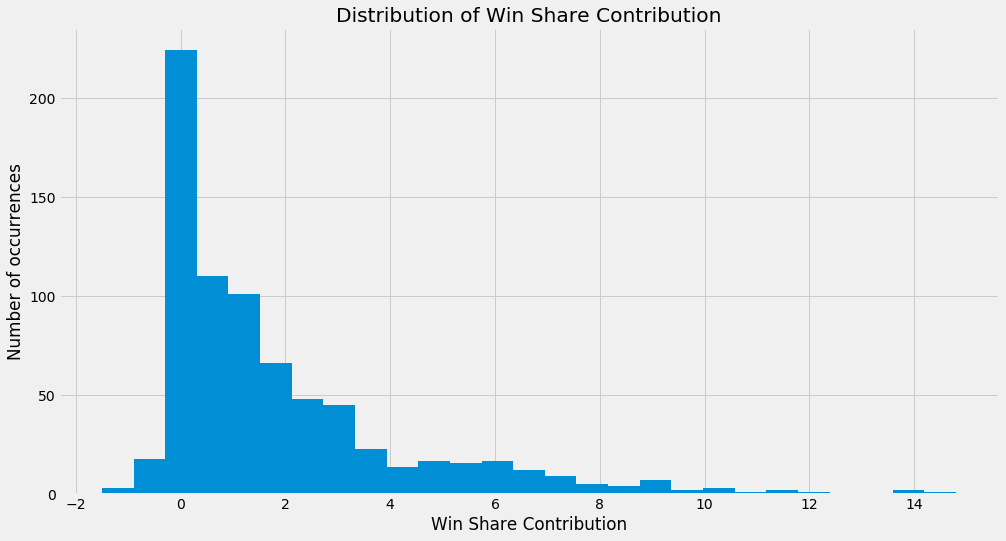
Initial Analysis:

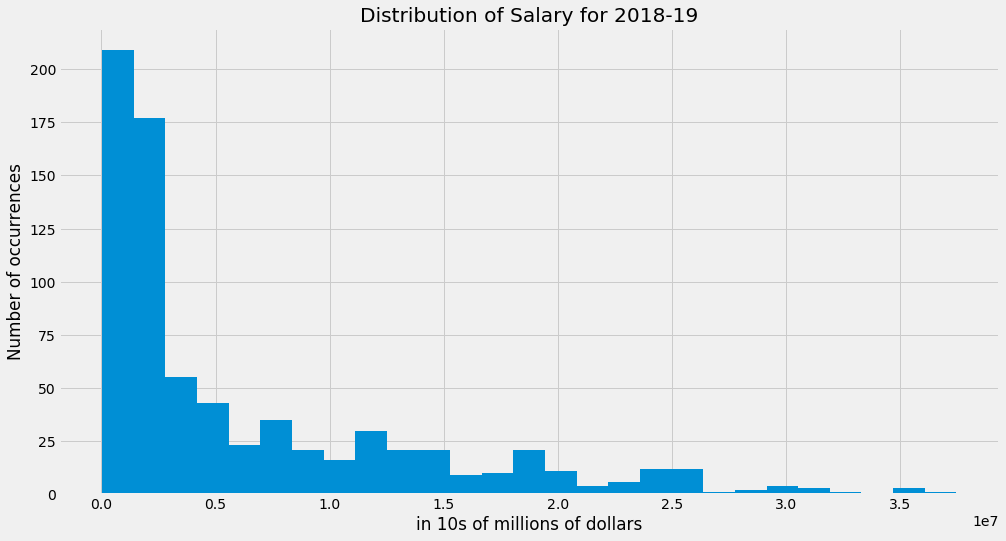
The goal is to utilize the players statistics, primarily Win Share contribution, to predict the salary that these players should have. Utilizing the information in this dataset to train and test a linear regression model to predict the 2019-20 contracts that players will have is going to be able to be tested on a hold out set, the current 2020 salary information.

An initial plot (below) of the Win Share and 2018-19 salary data can show that there is a correlation between the two.



Something to note here is the amount of the data that is in the early portion of the scatter plot above. To further emphasize the skew, the following histogram plots show the data for Win Shares and Salary.





These plots are showing a skew as many of the players are not receiving contracts greater than $5 million/year and have a Win Share contribution of less than 2. A further dive into testing and machine learning will follow.