

Capstone 1 - Understanding the Relationship Between NBA Player Salary and Statistical Performance

•••

Statistical Analysis and Machine Learning

Statement of Purpose

The purpose of this project is to understand the relationship between NBA player performance and salary.

There are two potential groups of clients for this problem:

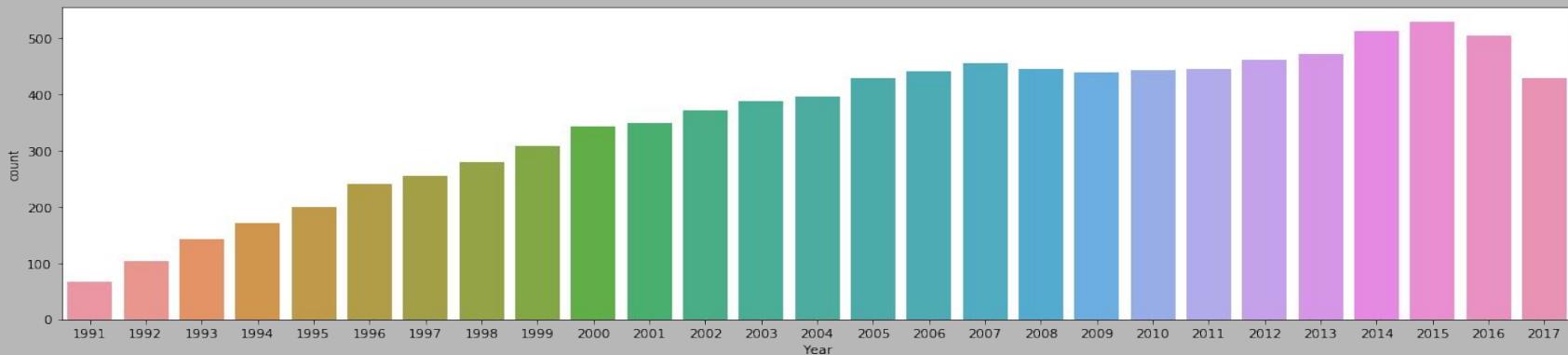
- The first, and primary, clients are NBA players themselves who can use the information to not only determine when their career performances may peak, but also to determine when to pursue maximum contracts during the course of their careers.
- The second group that can find this analysis useful would be NBA owners and teams who may be able to use it to determine when to best invest in a player based on their career performances in particular metrics.

Disclaimer

Before we move on, we need to make a quick note about bias in the dataset that could influence some of our results

Since we were originally interested in analyzing the relationship between statistics and salary over the course of players' entire careers, we didn't include any players whose careers began before 1991.

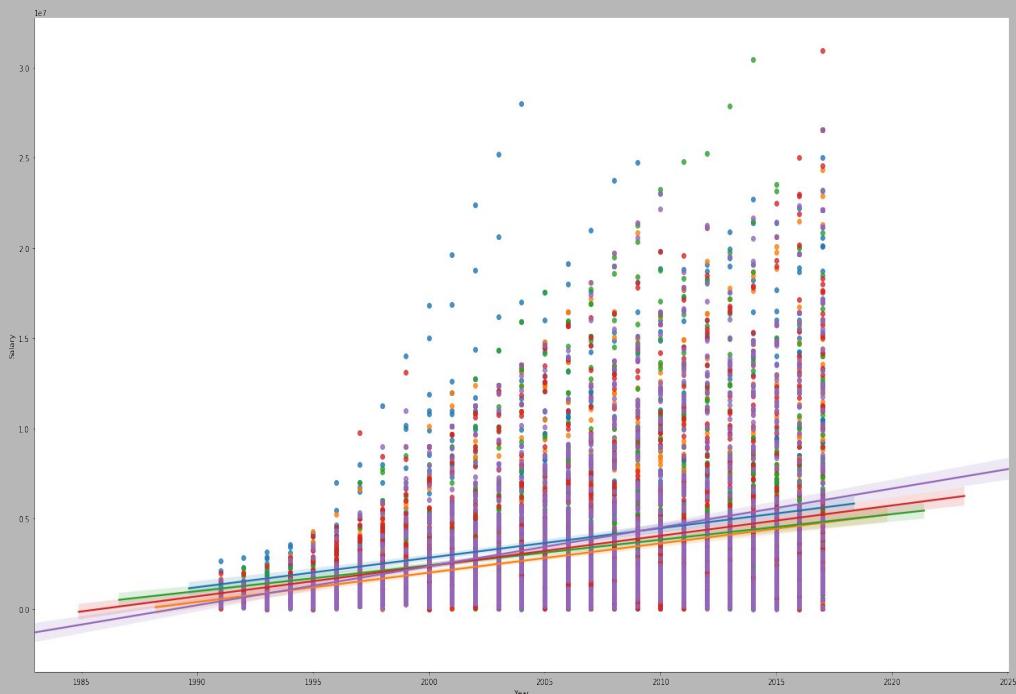
As a result, when we count the number of players in any given year, the stats in the 1990's contain fewer players than subsequent years. The population stabilizes around 2005. Consequently, there will be some bias, especially when we track relationships against years.



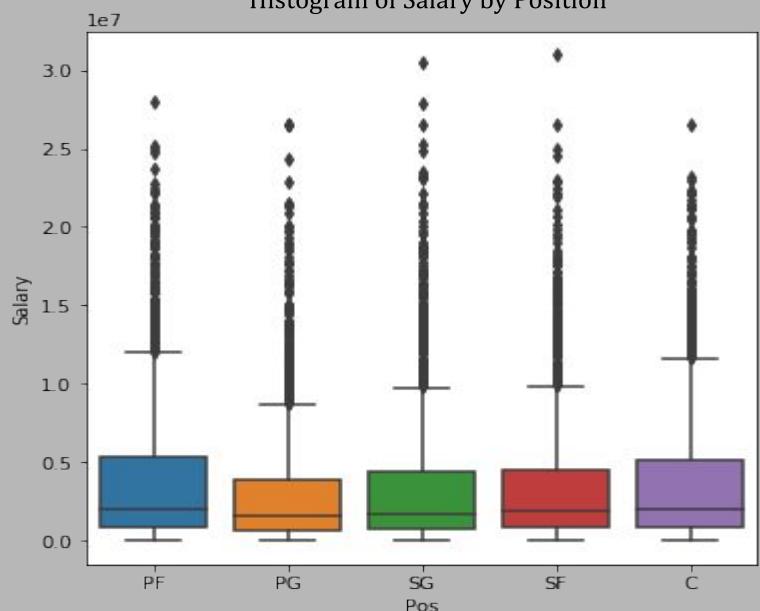
Initial Findings

Salary by Position

Salary Over Time (color coded by position)



Histogram of Salary by Position



Initial Findings cont.

Salary by Position cont.

From the previous graphs, we can see not only that salaries for all players (and all positions) have increased steadily since 1991, but that point guards are historically paid less than other positions, particularly centers and power forwards.

A statistical analysis finds that the difference in salary by position is statistically significant.

It's not immediately clear as to why point guards are paid so much less, and our dataset doesn't have the necessary data for us to develop a testable hypothesis.

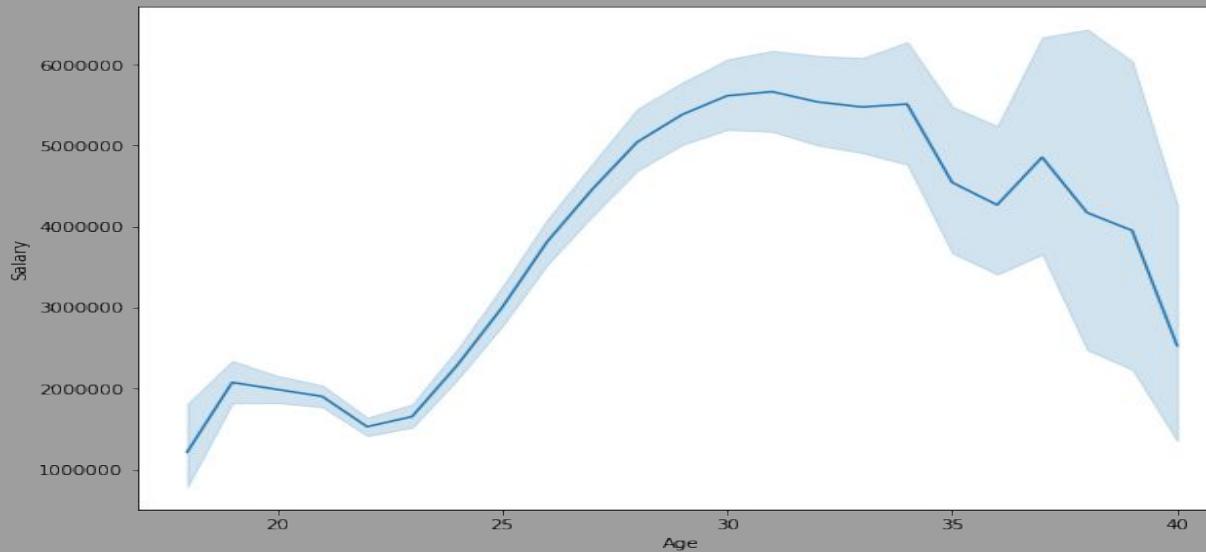
The main takeaway for players is that being a center or power forward is, on average, more lucrative than being a point guard and that one's position does influence one's salary.

Initial Findings cont.

Salary, Age, and Performance

We know that salaries generally increase with age, though there is a drop as players get closer to the age of 40. Salaries peak in the late 20's to early 30's before dropping. Yet, the salaries of veterans are still far greater than those of very young players.

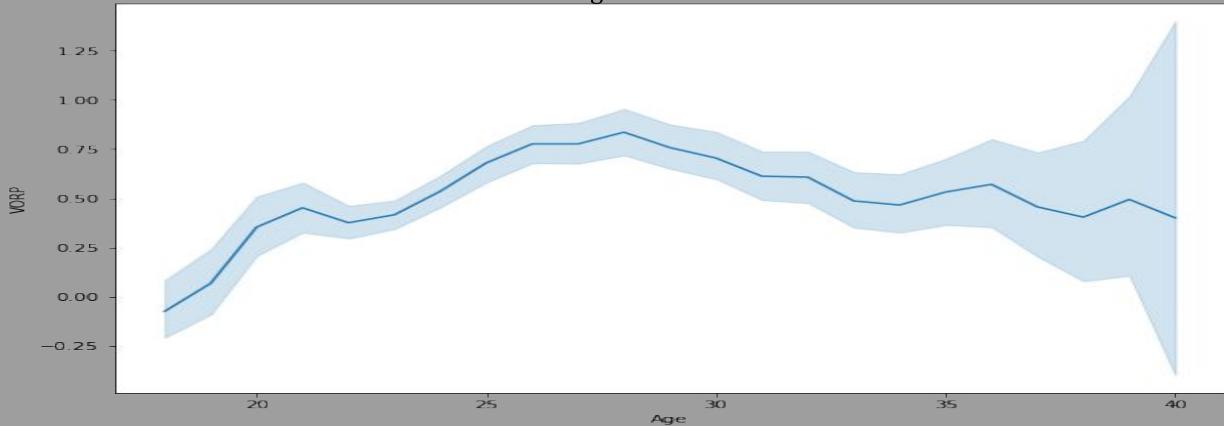
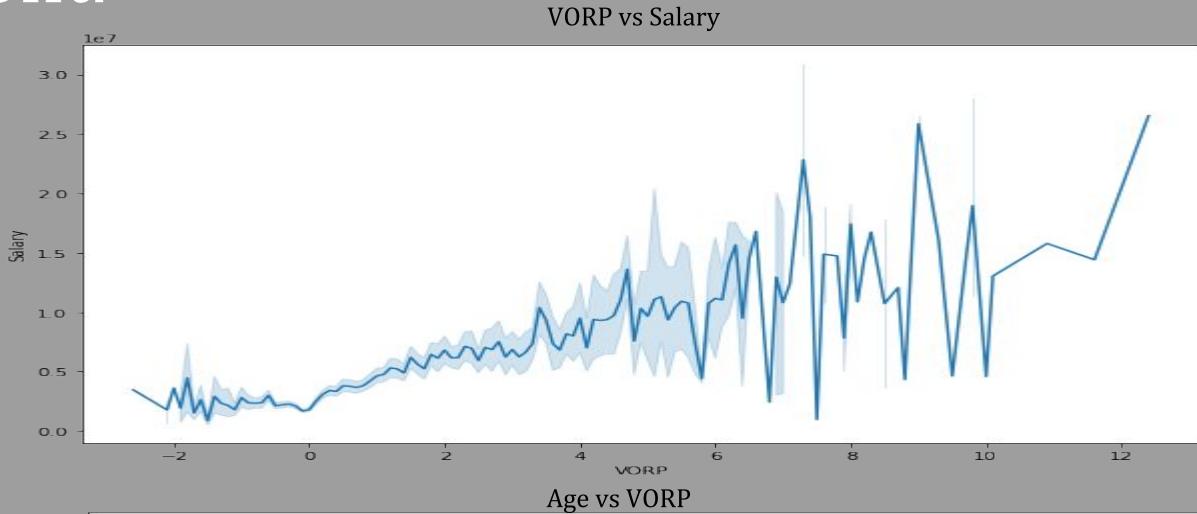
This may partly be due to veterans having demonstrated that they are worthy of high salaries. However, it may also be due to veterans being guaranteed greater pay due to rules negotiated with the NBA by their players' union.



Initial Findings cont.

In terms of salary, VORP correlates fairly well with salary. With regard to age, VORP tends to peak in the mid to late 20's.

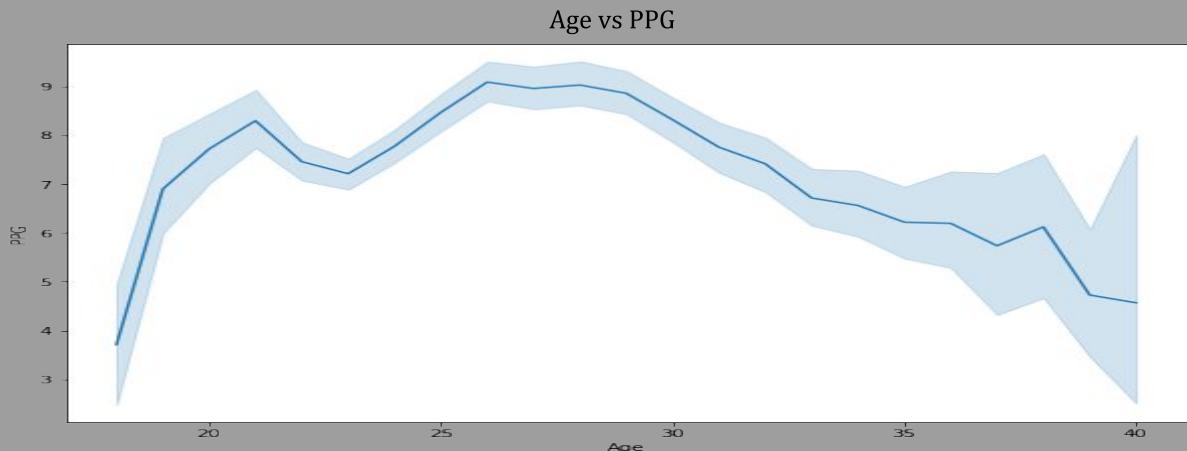
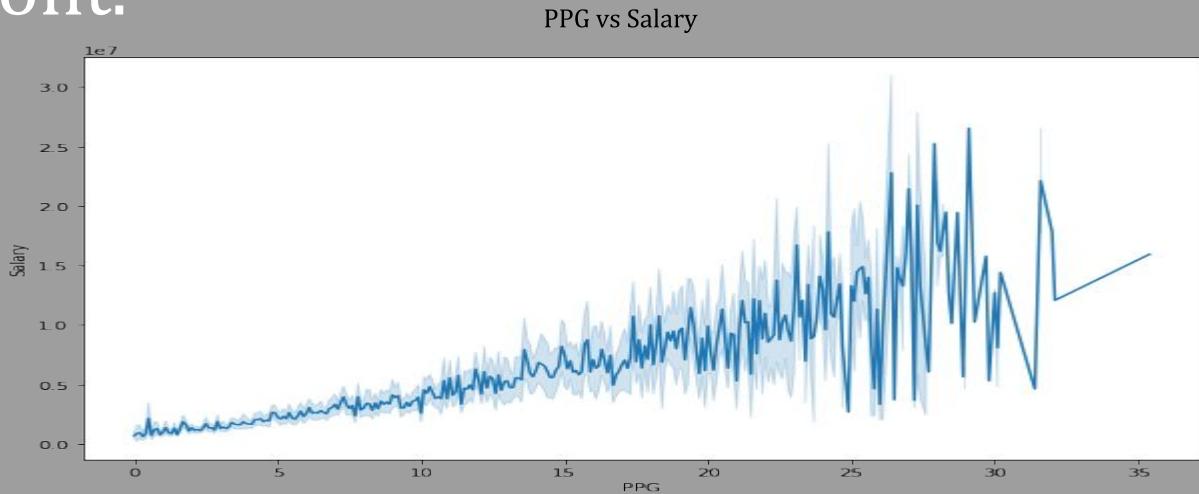
Since VORP is calculated using several offensive and defensive stats, the decrease in VORP values after 30 suggests that players' overall performances may begin to decrease in their 30's.



Initial Findings cont.

PPG correlates rather positively with salary. As with VORP, PPG peaks during players' mid to late 20's. However, there is a much sharper drop after 30.

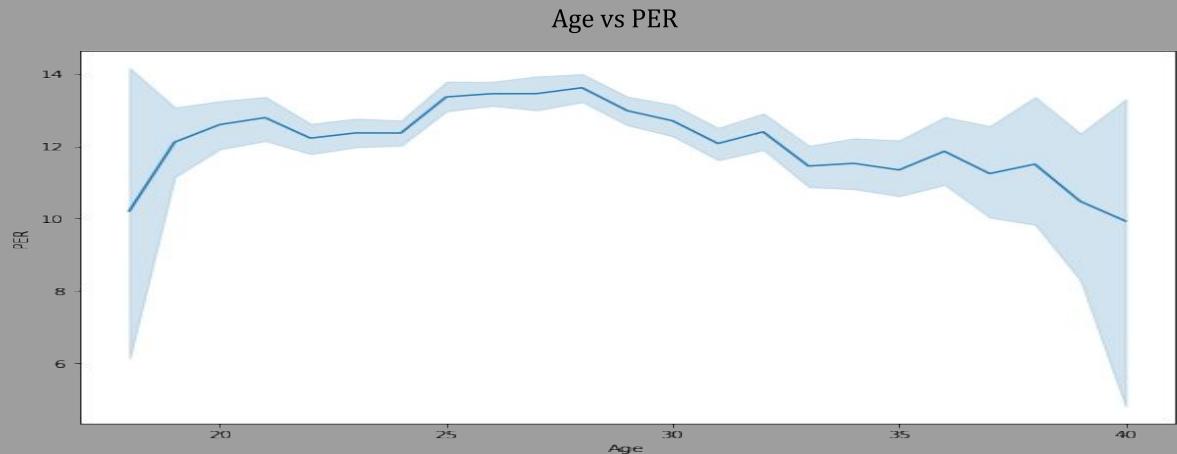
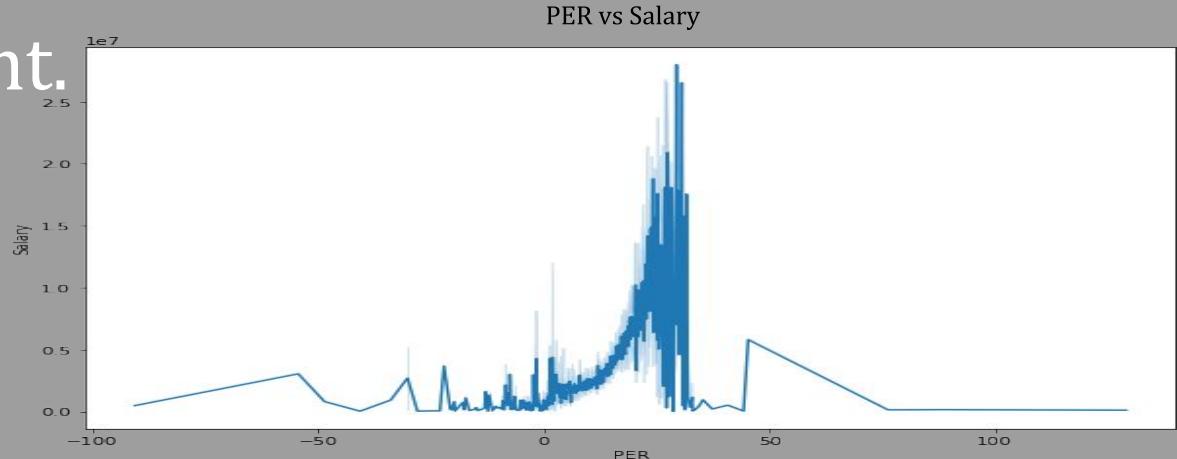
If a player who is primarily known for their PPG prowess, wishes to continue making a high salary into their 30s, it would be wise to diversify their skillset due to how sharply PPG drops for older players.



Initial findings cont.

PER is a rather odd metric, as we'll discuss in more detail shortly. However, just in terms of salary, PER seems to cluster around 15-20 for a broad range of salaries. This particular graph doesn't seem to suggest much of a correlation between PER salary.

With regard to age, as with VORP and PPG, it peaks in the mid to late 20's. Though, it doesn't drop as sharply as PPG after the age of 30.



Initial Findings cont.

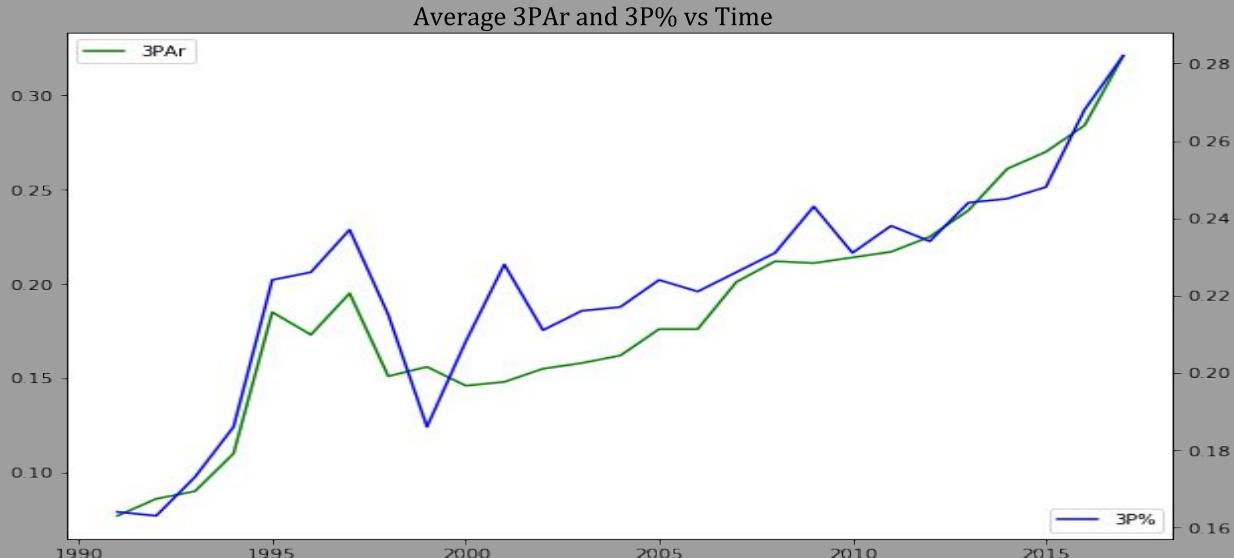
It should be noted that while VORP, PPG, and PER performance tends to peak in the mid to late 20's, salaries don't generally peak until the early 30's. There may be some possible explanations:

- Young players often have to demonstrate to teams that they are worthy of a larger investment. As such, once a player has shown that they can play well for a sustained amount of time, they usually end up signing either with their current teams or with new teams for larger contracts. But in this relationship, salary is almost always playing catch up to statistical performance.
- On the other hand, veterans have "proven their worth" so to speak, with years of evidence to demonstrate that they are worthy of high salaries. In addition, veteran players are guaranteed greater pay due to rules negotiated with the NBA by their players' union.

Initial Findings cont.

3-Point Trends and Salary

While various 3-point metrics didn't correlate highly with salary in our initial dataset, they did show a high correlation when we averaged our data by year. As we see below, both 3PAR and 3P% have increased rather rapidly over the last 10 years. It's indicative of the recent NBA trend in which teams are shifting to schemes that are dedicated to making 3-pointers a greater part of their offense.



Initial Findings cont.

For 3PAR, 3P%, and PER we ran a series of bootstrap analyses to better understand the relationship between them and salary. In a bootstrap analysis we try to determine whether the difference between two groups could be due to chance.

What we find is that there is a statistically significant difference in salary between the players who are above and below the average in each of the above statistical categories.

In other words, it's highly likely that there is a connection between performing well in each of these categories and being paid a higher salary.

Initial Findings cont.

PER is particularly important due to the range of statistics that go into its calculation:

- Players should strive to have high PER values. But given the range of variables that go into such a calculation, a player needs to be fairly well rounded and do well in a number of offensive and defensive categories. I
- That's not to say that players can't be one dimensional and still make high salaries, but being multidimensional can significantly increase one's chances of becoming well compensated financially.
- On average, the PER of younger players is not much lower than that of older players. But is it worth it for a team to invest more money in older players if they could get similar PER ratings from younger players? If PER was the sole factor in determining wins and losses, then perhaps. However, veterans tend to have less statistically quantifiable assets, such as experience, that may be more important for a successful season.

Machine Learning Analysis cont.

We also performed a machine learning analysis in which we attempted to predict salaries just from statistical performance. We were able to achieve 72% accuracy, which is not accurate enough to be able to predict salaries reliably. There may be several reasons why:

- Stats haven't increased at the same rate as salaries. As such, our model would have a difficult time projecting stats to salary values because players are being paid substantially higher wages for only slightly better stats.
- There may also be a variety of factors related to players being underpaid or overpaid. Teams competing with one another over particular players in a given situation drive up the price of a player such that an increase may not be warranted outside of that context (such as playoff contenders needing to pick up last minute help, a team looking to rebuild during the offseason with very few strong free agents to choose from, etc).
- Player popularity may also play a role as teams may be eager to cash in on advertising and ticket sales from a popular player, often resulting in a player getting paid more than his statistical output may warrant. Additionally, as the NBA has become a global brand over the last decade, more money has flowed into the organization, resulting in much of that money being distributed to players.

Takeaways for Players and Teams

- The NBA isn't as position specific as it used to be. Players at all positions have to be more versatile than in the past, and the more multidimensional a player is, the more valuable they become. It's crucial for players to be aware of that for the sake of their career longevity and earnings.
- On the other side of the coin, NBA teams can also benefit from this analysis. By realizing when player performance tends to peak, and by understanding the cost/benefit relationship between salary, performance, and age, teams can make better decisions about how much to pay players.
- This analysis seems a bit dismal for older players - they get paid high salaries, and in many cases, they don't necessarily play much better than their younger, more affordable counterparts. So why shouldn't teams just load up their rosters with younger players who cost less and can potentially offer similar results? Luckily, for older players, this isn't the only factor that teams consider, or should consider, when putting together a roster. Veterans may possess other factors, such as experience, temperament, the ability to perform under pressure, and the ability to get along with teammates, that aren't as easily quantifiable and that may be in greater abundance in older players.

Takeaways for Players and Teams cont.

- Finally, our machine learning analysis also has important insights for players. The fact that we weren't able to predict players' salaries anywhere close to 90% accuracy based on their statistical performance is worth noting. What it suggests is that there's more to how much a player gets paid than just their on-court performance.
- They, for example, can learn to better market themselves. Even if they happen to have mediocre stats, by marketing themselves properly, they can make themselves more financially lucrative to a team. This could be in the form of advertising deals, both for the team and player. The increased popularity can also lead to increased ticket sales and, of course, a greater salary for that player than his on-court performance warrants.

Links to References

Final dataset:

https://github.com/kjd999/Springboard-files/blob/master/Capstone%20Project/nba_final_dataset.csv

Project Proposal:

<https://github.com/kjd999/Springboard-files/blob/master/Capstone%20Project/Documents/Capstone%201%20Project%20Proposal.pdf>

Data Wrangling Report and Notebook:

<https://github.com/kjd999/Springboard-files/blob/master/Capstone%20Project/Documents/Capstone%201%20Data%20Wrangling%20Summary.pdf>

https://github.com/kjd999/Springboard-files/blob/master/Capstone%20Project/nba_data_cleanup%20.ipynb

Data Story Report and Notebook:

<https://github.com/kjd999/Springboard-files/blob/master/Capstone%20Project/Documents/Capstone%201%20Data%20Story.pdf>

https://github.com/kjd999/Springboard-files/blob/master/Capstone%20Project/nba_data_story.ipynb

Exploratory Data Analysis Report and Notebook:

<https://github.com/kjd999/Springboard-files/blob/master/Capstone%20Project/Documents/Capstone%201%20Exploratory%20Data%20Analysis.pdf>

https://github.com/kjd999/Springboard-files/blob/master/Capstone%20Project/exploratory_data_analysis.ipynb

Machine Learning Analysis Report and Notebook:

<https://github.com/kjd999/Springboard-files/blob/master/Capstone%20Project/Documents/Capstone%201%20In-Depth%20Analysis%20.pdf>

<https://github.com/kjd999/Springboard-files/blob/master/Capstone%20Project/nba%20machine%20learning%20analysis.ipynb>

Final Report:

<https://github.com/kjd999/Springboard-files/blob/master/Capstone%20Project/Documents/Capstone%201%20Final%20Report.pdf>

Slide Deck:

<https://github.com/kjd999/Springboard-files/blob/master/Capstone%20Project/Documents/Capstone%201%20Slide%20Deck.pdf>