STATS 100 Project Proposal

What differentiates a "good contract" and a "bad contract" in the NBA?

Motivation

Ever since I watched *The Importance of Undervalued Contracts* by NBA YouTuber MDJ, I became obsessed with determining whether a new free agency signing is an undervalued or overvalued contract. I want to see if I can use data to separate contracts based on value and predict whether a contract would be worthwhile based on previous performance, which would be more valuable in the real NBA setting.

Data Sources

Basketball reference has the salary and team/player season performance data I'll be using. It has the most robust data with very little cleaning necessary, but it does have the downside of not having an API or easy-to-use scraper to pull its data from. I may try to combine data sources like abresler's nbastatR package and this data.world salary data to replicate the Basketball reference data, but it would depend on how much time I would need to do either task.

Project Direction

To differentiate between undervalued and overvalued contracts, I plan to use KNN regression. I can use variables such as years in the league, contribution metrics from previous seasons, lifetime fines to metricize character, etc. For prediction, I can use data from previous seasons into the KNN regression for current contracts or train a new KNN model using previous data.

I've yet to determine which metric to use to determine a player's contribution to a team, but there is a large variety such as win shares, value over replacement player (VORP), or other advanced metrics such as BPM and RAPTOR. I need to factor in the contract value of the player into this metric, which I can do simply by dividing the salary of the player by the metric I choose. I may need to explore the relationship between player salary and contribution to find a better method to incorporate the two. At the minimum, I hope to be able to see clusters of good/mediocre/bad contracts relative to production.

Challenges

There are 2 major challenges in this project.

1. Not all teams can nor should determine a player's value in the same way. For contenders, a player could be valuable if and only if he bolsters the team's playoff success. On the other hand, rebuilding teams might value aging veterans to pass on lessons in leadership to their younger core.
2. A player's value can change drastically over time, especially more so if they're on a longer contract. This can happen due to age (for the better or worse), a team's priorities changing (contending vs. rebuilding), injuries.

The first challenge can be addressed by only including contracts signed by teams in the playoff hunt, but the second challenge imposes a much larger difficulty. The approach I use to address the second challenge will determine my project's complexity, so I may have to find a balance.