Final Write-up

Yash Manne

My project explored NBA player salaries as compared to their overall ratings. Specifically, I wanted to find out which teams are generally better at evaluating talent and leveraging monetary spending effectively. Originally, I also wanted to investigate contracts and what year would be the best time to sign someone to an extension but I wasn't able to find historical contract histories and only could find data for salaries by season. Gathering salary data was also quite challenging in that no site had the full data and I had to scrape it from individual player pages from basketball-reference.com. This gave me an opportunity to learn basic web-scraping. I did run into a roadblock when the tables on the pages didn't contain values but were actually views of a backend database. Thankfully, basketball-reference provided a link to a widget that shows the actual values and the links were mutable enough so that I could plug in player names and gather output.

Well over 50% of my project was simply gathering data from various sources and joining them together in a usable format. I severely underestimated the amount of data munging I would have to do. Thankfully, I had enough data where I could filter out enough to get to a manageable size without any significant loss of information. Once I had the data, I had to transform certain values so that they would be more comparable. The raw salaries had to be standardized by the league salary cap for that season. These scaled salaries were lognormally distributed and had to be transformed to be normally distributed so that I could adequately compare with player rating, which was also normally distributed.

I think I made sufficient plots to highlight that the San Antonio Spurs organization is remarkably good at evaluating talent and other organizations like the Sacramento Kings are generally pretty poor. Other organizations like the Los Angeles Clippers, Memphis Grizzlies, and Toronto Raptors have had a steady upward trend. Still, my analysis was limited. Currently, my method of "Value" drastically favors average players who get paid very poorly. While "value" might be high, having a team of average players will not win championships. As such, I'll have to incorporate player accolades and postseason team success into the model. Additionally, this model had to filter out players based on minutes played and thus doesn't account for players like John Wall who have massive contracts but barely played any games due to injuries. These cases are

really where the magic happens and I'll need far more data and experience to fully answer my question.

Through the course of this project, I've learned that it's far more important what data is plotted and how the data was created rather than the plots themselves. I've made relatively easy data plots that effectively communicate my data analysis but the hardest part was figuring out how to parse through the data in order to analyze my question.