

Capstone 1 In-Depth Analysis

For the in-depth analysis for my first capstone project I didn't too much and really tried to keep things simple. The following steps were taken in order to get the material that I really needed:

1. First I took the information I had previously gathered in order to determine what I wanted to do my regressions on.
 - a. I found after looking further into the scatterplots I had previously created that not only did the wholistic statistics effect win rate more reliably as far a trends go, but the other individualistic statistics couldn't seem to fit any sort of regression reliably.
 - b. The conclusion I came to was that I needed for now to only create regression models for win shares as the y value vs. value over replacement player, box plus/minus score, and player efficiency ratings as the x values.
2. The next step was to reshape the data in order to be able to set up my regressions. The data was in a 1d array so this was a necessary step for being to use regression models from scikit-learn.
3. The fourth step I took was preparing the different regression models for the three statistics I was looking at more in depth. I set up a linear model for the three major statistics as they relate to win shares.
 - a. My findings were interesting for the regression models I created. I found the following:

- i. The first thing I found was that while all three of the statistics had a positive impact on win shares, player efficiency rating had by far the least significant impact of the three.
- ii. Next I realized that the trend lines compared between box plus/minus score and value over replacement player while being similar were not nearly as similar as I thought they'd be. Those two statistics are similar in that value over replacement player is directly related to box plus/minus score as far as the formulas go. Value over replacement player was the statistic, however, that was clearly the most significant.

After completing the analysis and looking more into the data I was working with I further developed my hypothesis. Originally I had thought that teams were undervaluing individualistic statistics based on the correlation between those statistics and salary and that teams should look more into the needs of the team individually rather than just getting the best overall players possible. After doing more research I found that individualistic statistics had basically no effect on how much a player was contributing to a team's wins. I realized that wholistic statistics were much more important. After doing more research though I found that two of the wholistic statistics I was looking at in box plus/minus score and value over replacement player were based mostly on the individualistic statistics I looked at. This leads me to believe that not only are the wholistic statistics more important, but in addition how well rounded you are as a player

also will have a relatively significant impact on how much you contribute to your team's win rate.