Aaron Drexler

10/10/21

DSC 550

Milestone 1

If there is one thing that I have great interest in (other than data science) it is baseball, both as a player and as a fan. I consider myself a baseball nerd, one who follows baseball to an extreme level. One thing did strike me as interesting as further intriguing is how to value each player. This often comes up when signing free-agent contracts or looking back on these contracts reflecting on time. I wanted to look back at recent year then historically, to see what the level of value is, dollars to data, of each player. While my data raw does not have any advanced sabermetric data (new age baseball is heavy into advanced statistics), I will calculator those advanced statistics are compare them to each player’s salary. I plan on using several different tools, such as data visualizations, correlations, and several other different tools that will allow me to determine the average baseline of dollars to performance. For reference, the expected rate is an 8-9 million dollars per win above replacement (one of those advanced statistics) I will test this hypothesis to see if that is indeed the average player’s win against replacement matches the hypothesis. This is an imperfect study, as baseball has a unique and uneven monetary system that leaves young superstars paid way less that most of their veteran teammates despite their better performance. There are a number of different CSV files within this one source, and I will be using more than one as my file sources.