Kevin Nguyen

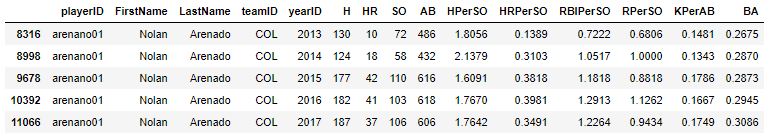
CSCI 4831

Final Project

The statistic I chose to evaluate players with mainly deals with how efficient the player is with regard to how many time that player strikeouts. I was really interested in this statistic because when I watch guys like Aaron Judge or Giancarlo Stanton for example I either see a massive home run or they get a strikeout. However, I will not only be looking at home runs, but also other statistics such as hits, runs, and runners batted in. This is so I can see all the other ways a player can contribute compared to the number of times they strikeout. I will be using the Lahman database to get my data because it has all the players information and the statistics that I will need to use. I will only be using the data from 1991 – 2017 just because my crashed when I ran the sql query for more than those amounts of years.

To compare how players are doing I will look at the league average for that year and see who is above and below the average to see how efficient they have been compared to the average.

This statistic is comparable to batting average except instead of dividing by the number of at bats I am dividing by the number of strikeouts. I wanted to see the ratio a player has for different offensive statistics to number of strikeouts. I will be using a correlation test between Hits Per Strikeout and Batting Average to evaluate my Strikeout Efficiency statistic.

When looking through different baseball statistics I was never able to find one that uses strikeouts as the main focus, so I wanted to see what I could do with it. I think this would be a good statistic because it shows how efficient the player is based on the number of strikeouts they received in the year. Also with my demonstration you will be able to see if players have improved or declined throughout the years. So, you will be able to see if players are becoming more efficient per strikeout, or getting worse. An example of Nolan Arenado is shown below.

As you can see he is been improving his HPerSO each year, and also his HRPerSO, but in 2017 he had an off year and his numbers started to decline. However, even though he was in a decline he was still one of the most efficient players for the year as he was above average for each statistics and in the top 25 for HRPerSO.

In the Jupyter Notebook that I created you will be able to filter the data I have gotten based on playerID, First and Last Name, Team, Year, the minimum At Bats, and the Statistic of your choosing, I hope you enjoy looking at my statistic. Sorry if it is not too in depth, I do not know much about baseball and was having a hard time coming up with a statistic.