Christian Geer

Professor Bonham-Carter

CS 390: Data Analytics

October 29, 2017

The data set that I have selected for my final project is the Lahman Baseball Database. This database contains baseball statistics, player information, and award history for the major leagues from 1871 to 2016. The questions that I have come up with all revolve around the central theme of a player’s value, and how that value is evaluated. My first question aims to find which stats most directly correlate with a whether or not a player is elected to the Hall of Fame. Finding a concrete answer to this question will show if the Hall of Fame voters tend to value certain statistical categories more than others. The second question is similar to the first: what statistical categories correlate with All-Star selections. While you might expect that the answer to this question would mirror that of the first, it could be the case that some stats ‘age’ better than others. It is also important to keep in mind that from 1947-1957 and 1970-present the fans have voted in at least the starters for the game. This could change things as fans could place more or less emphasis on certain stats compared to the BBWAA. The third question turns its focus towards salary. This question is who are/were the most under and over paid players. This question will combine the statistical and salary numbers available to determine the best and worst value players. This question also ties into the following one: what was the best team for its salary? The process for this question will be similar to that last one, but with the added element of combining all of the players of a team together, and will, in a sense, show which front office was able to put together the best team with their available resources. Finally, I will research how much individual awards correlate with a player being named to the Hall of Fame. Along with answer this question, this will also show how much the BBWAA weights stats compared to awards, as a stronger correlation for one would imply it being more important.