Lai Tsun Hin

CSCI 4831

Sabermetrics

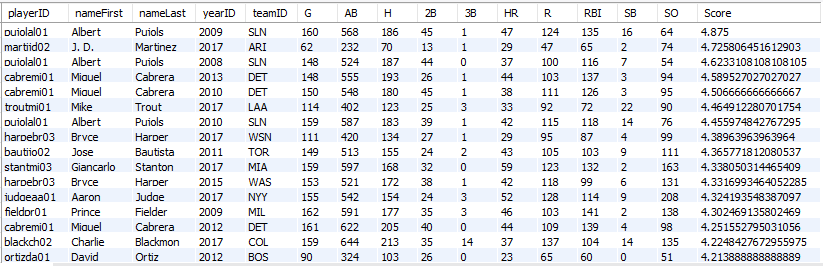
Final Project

The statistic I chose to evaluate deals with how efficient a player’s recovery from injury is. This was extremely hard, because it was hard to create just a category based on injured players. There wasn’t a category in which the INJ players were shown in a csv. To look at the injured players, and come up with a recovery percent, I had to look at a player’s career statistics versus the post injury statistics to come up with the amount the player had come back from injury.

I initially planned to use the Lahman database to get my data since it provided all the players information and the statistics from 1991 – 2017, but to look at when a player was injured, it didn’t provide a way of categorizing players regarding injury, then I tried looking at injury datasets online, but there weren’t any data sets that could be provided.

I then looked into specific player and position injuries, and decided to narrow the topic down to the recovery percentage of injured batters/ hittes. To compare how each player was doing pre injury and their recovery process post injury. I looked at how a player was performing before his injury, and how the player performed in the next season, to see how they are performing compared to their previous season.

To look at how a player was performing. I came up with an equation to evaluate the player’s performance. I took into account of the player’s Hits, Runs, 1B, 2B, 3B, HR as well as the RBI, SB and SO. With that in mind, I took all players from 2007 to 2017, to provide a 10 year range of baseball play to compare the performances.

Similar to how a fantasy baseball team can be drafted, I took all the above metrics into account to find a top scorer in the past 10 years. I also set the restriction of AB, being over 200, to take away minimal play of players. I then got a sense of the top player in the hitting category, with a score of 4.875. 

With that in mind, I set this highest score as the maximum in the statistical category. Next I divided all the score by that max value and times it by a 100, to get a sense of how a player performed regularly out of a 100.

After coming up with the scoring system, I realized that since players were coming back from injuries, a lot of them would be playing half or even a quarter of a season, so I divided the previous equation by the amount of games played, to get a per game average of the player. Giving it a better sense of the recovery process in a per game aspect.

When looking through different baseball statistics, I was never able to find one that identified the players as injured, so to examine injured players, we have to input the player’s name and injury year, then I would filter out the past two seasons’ score and compare it to their score after coming back from the injury. I think this would be a good statistic because it shows how the player is progressing on in the next year and can give others a sense of recovery and also a value to the team.

I go in a lot more detail in my demonstration you will be able to see the value of each player and also his performance against teammates. So, you will be able to see if players are recovering well and worth their value.

Most of my analysis was done on the juypter notebook and the values can be changed around and tailored to fit whatever player that is injured. Lastly I enjoyed this project a lot and I want to thank you for such a wonderful year.