Brett Sutow

Final Project

DSC530

**MLB Players Peak Ages**

The overall goal of this research study was to take statistical data from the five best homerun hitters of all-time in Major League Baseball and to analyze it. The question we sought to answer was where a player’s career peak, looking at data spanning over 100 years, with the purpose to dive deeper into understanding. During this journey we observed many unique statistics and correlations. Which allowed us a better understanding overall of the data we were working with.

The first thing that was learned during this analysis was the fact that there is a correlation between homeruns and RBIs with age. This correlation is a negative correlation as a player gets older their homerun and RBI production decreases significantly from earlier in their careers. However, we do not see such a relationship with hits and batting average. Though it is observable that there is some sort of relationship between the two, it cannot be guaranteed.

Now looking at our main question regarding a player’s peak performance and where it comes to into play. We see from our data that a player seems to peak two times during their career. The first time is around 25 years old. Where a player typically hits their peak performance for statistics especially homeruns. They then have another peak, which is smaller a little over 30 years old around 31 or 32. After this age however a player’s career typically goes down in the performance realm. Overall, if you are looking to sign a player to a contract the ideal age to sign them would be from 25 to 33. Which is a key observation in today’s sports world of large contracts.

The final large observation that I took away from my research, was the fact that the player Barry Bonds in himself is a statistical outlier. Bonds peaked during his mid-30s where he famously hit 70hrs, as well as had one of the greatest all-around statistical seasons in history. Barry Bonds was an anomaly in the data and did not properly represent the rest of the players.

What was missed during this analysis were players who were not pure power hitters. Which may have created a different outcome for peak age. As we were seeing statistically power hitters have two peaks, however we can see consistently hitting and batting average stayed pretty stable. If we were to look at more well-rounded contact hitters would they have a different peak area.

Looking at the variables I think I did a pretty good job of handling those that were NA and outliers. My only concern is how to handle the Willie Mays missed years. Because this came during part of his “prime” or “youth” would this have made a difference in his peak age. Which was one of my biggest challenges when working with the data. I believe besides this all my assumptions were good and made sense in the context of the data I was working with.

Overall, this was a great study to fully understand a player’s career, and when they peak. This is study for those sports teams looking to sign or extend the contracts of their players because it allows them to gain a better understanding of when these signings take place.