Aaron Drexler

Final

DSC 530

* + **Outcome of your EDA**
  + Overall, I was looking at the comparisons between size, such as height and weight in relation to hitting triples. As an assistant baseball coach, I wondered if the old adage, short small hitters are more likely to hit triples. While I could not tell the player type by the stats, I focused on the correlations between height and weight with number of triples and triple percentage. The following is Pearson correlations that were found. For the full statistics, please check the power point:
    - Weight To Triples: -0.1982152
    - Height To Triples: -0.2512949
    - Weight To Triples/Hits: -0.1175858
    - Height To Triples/Hits: -0.1201294
  + I expected there to be a mid-level negative correlated. While I was correct about the negative correlation, it was relatively to very week. The negative correlation does indicate that people who are light, but especially short are more likely to triple. However, with the weak correlation, I believe that this is not a causation, but rather a shared attribute for those who are quick on the base paths. I also noticed that the triple percentage went down in correlation strength, indicating that those who hit triples are more likely to hit a high rate of hits as well.

**What do you feel was missed during the analysis?/ Were there any variables you felt could have helped in the analysis?**

* + In theory, I should have included a speed consideration in my triples analysis. I could have used steals, as that is the most speed related statistics, but average steals fluctuates so much through the years, as does triples. If I had more time, I would have created a yearly mean, and used the residuals from the mean as a speed measurement to determine if this was the cause for higher triples rate. By comparing and find the correlations between those, we can see if there is a relationship between quick players and triples.
  + **Were there any assumptions made you felt were incorrect?**
  + Coming into this activity, I was under the assumption that being small caused players to be more likely to triple. However, post project, I found that the correlation I though was high actually had a relatively weak correlation (even if I was right about the orientation of the correlation.
  + **What challenges did you face, what did you not fully understand?**
  + One of the biggest challenges that I faced was my attempt to compare two different data banks. This was especially tricky as the people data set had all players, both batters and pitchers, while Batting had just batters in its databank. I finally found a way to combine the two data banks using playerID as an index for combination. .