Cincinnati Reds Hackathon 2024

MLB's Freaky Friday: Pitcher Role Reversal

Background:

In 2023, the average starting pitcher threw 5.1 innings; in 2013 that number was 5.9 innings. As a result, the average Major League Baseball team had 526 relief appearances in 2023, compared to 478 in 2013. Appearances alone do not tell the story of the additional innings relievers had to cover — in 2013, across the league there were 3,048 multi-inning relief appearances (~102 per team); in 2023 this was up to 3,941 multi-inning appearances (~131 per team).

Given these changes, MLB pitching staffs are becoming increasingly amorphous in nature, and the types of roles available to pitchers are more fluid than ever before. Particularly, pitchers are now utilized across specific (and often multiple) innings, leverage situations, and game circumstances that can enable teams to get the most out of them, even though that utilization may not conform to tradition. Despite these recent developments, there are sure to still be pitchers currently serving in a role that is suboptimal for maximizing their value or performance.

Question:

With the provided data, identify characteristics of pitchers that lead to their success in specific roles, such as a traditional "starting pitcher", a traditional "relief pitcher", or any other pitching roles that you establish within your own analysis. Then, using your findings, identify some MLB pitchers who could be better off in a different role – perhaps as a starter, as a reliever, as a swingman, or as something else entirely! Finally, provide an in-depth analysis of 2 or 3 of these players. Why might these role switches be better for them?

How you answer this question is entirely up to you, and we encourage you to be creative with your data-driven approaches, player selection, and role proposals within your submission. But at a minimum, our team is expecting to see a(n):

- Clear, defined analytical process;
- Detailed explanation of your model(s) and modeling decisions;
- Explanation of your findings about characteristics of successful starters, relievers, and/or other pitching roles; and
- Detailed description of 2-3 MLB pitchers who are candidates to switch roles.

Submission:

Your submission should consist of **a PDF write-up** of **no more than 1,500 words and 8 figures or tables**, as well as **a separate documented notebook or script** containing the complete code that can be used to create and reproduce your project. Write-up submissions will not be penalized for any number of words or figures below these restrictions. Additionally, please try to limit any code in the write-up that does not directly enhance your analysis (e.g. leave relevant statistical formulas in the write-up but restrict any visualization code to the notebook).

Submissions will be evaluated based on three equally weighted criteria:

- **Data Science**: Is the approach fundamentally sound? Are the selected statistical models appropriate given the data and the problem at hand? Do the analytical applications serve to answer the questions in a technical and thoughtful manner?
- **Baseball**: Are the conclusions and proposed role-changes derived from the data? Does the analysis demonstrate knowledge of baseball-specific data and nuances? Are the suggestions applicable for an MLB team to use and implement?
- **Communication**: Is the write-up easy to follow? Do the figures enhance the report and analysis? Are you able to clearly and concisely communicate your process, proposals, and justifications?

Final submissions are due at **11:59pm EST on Monday, February 5th**. Please submit your PDF write-up and code notebook to the Reds Hackathon Coderbyte link (https://coderbyte.com/sl-candidate?promo=cincinnatireds-sqyjd:technical-assessment-tgkkq8xar7&invb=userweij5h2w). To submit your assignment, **please collect your files in a shareable directory on a platform like Google Drive, Dropbox, or GitHub**. There will be a submission box where you can paste in the URL of the shared directory. We will download and review the files you include in your submission.

Data:

Our team has provided you with two datasets, along with a file containing data documentation and explanations of the included columns, described below.

- savant_pitch_level.csv – This Baseball Savant dataset includes every MLB pitch thrown since 2021. It includes information about the situation, the pitch, and the batted ball.
- fangraphs_season_level.csv – This FanGraphs dataset includes season-level data for pitchers since 2021. It includes information about player performance – including traditional metrics, advanced statistics, and pitch-type specific measurements – split out by pitcher role (i.e. as a "starting pitcher" and as a "relief pitcher").
- data_dictionary.xlsx – Contains documentation about the columns found in the two datasets.

You are free to use other publicly available datasets, but we feel you do not need to go beyond the data provided. If you do so, please provide all relevant information about your sources and collection process.

Please reach out to hackathon@reds.com with any additional questions. Good luck!