

ASSIGNMENT #2: SIGNAL V. NOISE IN PITCHER OUTCOMES

Your GM has asked you to prepare a report on the top pitchers available in free agency this off-season, predicting their RE24 allowed for next season and making a recommendation on how to prioritize them.

WHY ARE YOU BEING ASKED TO DO THIS?

Evaluating players in free agency is a fundamental task for baseball front offices. Usually, when creating projections, you have multiple years of data, and age is an important consideration. We are simplifying the task to practice the fundamental skill of extracting signal from noise in a single season of data.

WHAT (EXACTLY) ARE YOU BEING ASKED TO DO?

Your task is to put together everything we've learned so far about separating signal from noise to predict next year's RE24 against four of the top free agent pitchers available: **Corbin Burnes**, **Jack Flaherty**, **Max Fried** and **Blake Snell**. For the sake of these predictions, you are to ignore the players' ages and use only data provided in event.csv from the current year. In your report, describe your prediction methodology in sufficient detail that another student in the class could reproduce your work.

After describing your methodology and presenting your results, make a recommendation by ranking the four pitchers in the order that you would prioritize signing them. In other words, if all pitchers were available for the same price, what would be your preference order for signing them? Note any large gaps between consecutive pitchers in this ranking. **This recommendation is a subjective exercise, and you are allowed to consider outside information when making this recommendation.**

The expectation is that you will use some of the tools we have covered in the class (e.g. regression to the mean) in the context of batter evaluation. We have learned how to separate signal from noise for batters, and now you must think critically about how to apply that framework to pitchers. Although you are expected to use some of the tools we have learned in class, there is no right answer. You are also encouraged to think on your own about how to predict pitcher performance. Importantly, you are expected to do some sanity-checking of your results. Just as you would in a real job, make sure your results make sense before you submit them!

Note that you are discouraged from incorporating outside data in your predictions. For example, pitch trajectory information is very useful for pitcher projection, but we will cover that on the next assignment. For now, focus on extracting the signal from the event data.

SUBMISSION REQUIREMENTS

- A PDF report (max 4 pages) summarizing your findings, including at minimum the following:
 - A description of (and reasoning for) your methodology for predicting pitcher performance
 - At least one data visualization that tells a story about your results
- A R script (.R file extension) with all of the code you used to generate your report and the CSV

REMINDERS

- Prepare your report as if your audience is a baseball executive who has not seen the assignment prompt. Write clearly and concisely, and format your report in a way that makes it easy to read.
- In this class we value **critical thinking**! Don't just parrot what you've been taught—bring your own experiences to bear on the assignment. If you disagree with something we've covered in class, that's strongly encouraged! Be sure to explain your reasoning.
- Please **anonymize** your submission by removing any personally identifiable information (including file paths in your R script that contain things like a username!).

HOW WILL YOUR GRADE BE DETERMINED?

You will get feedback on your work product based on several criteria. Within each of those criteria, the feedback will be: Missing (0%), Needs Improvement (70%), Good (85%) or Exceeds Expectations (100%). Your grade on the assignment will be the average of the grades across criteria. The criteria are:

1. **Description of prediction methodology.** Did you explain your prediction methodology with sufficient detail that another baseball data scientist could replicate it?
2. **Implementation of prediction methodology.** Did you correctly implement the prediction methodology (or did you make a mistake)?
3. **Data visualization.** Did you include a data visualization that tells a compelling story?
4. **Critical thinking.** Does your analysis exhibit a depth of thinking about the problem, or are you just applying the methods we've covered in class?
5. **Written communication.** Did you write clearly and concisely? Did you organize your key ideas with the evidence supporting them? Did you format your report in a way that makes it easy to read?