



SMT DATA CHALLENGE 2023

In baseball, defensive analytics present challenges different from pitching and hitting. Not only is spatial information essential, but one fielder's location, movement, and decision-making can materially impact those of other players.

For this year's Data Challenge, we want you to **consider a game situation involving the interaction of two or more defensive teammates**. Moreover, since action is central to all sports, we want you to **frame your analysis as a story**. Show us how your work informs and contributes to others' knowledge of and appreciation for the game.

We encourage you to consider any situation that interests you, but meanwhile, here are a few sample questions that you may use or draw inspiration from:

- When making an outfield assist, does hitting the cutoff man really make a difference?
- Compare double plays using different combos of Players A, B, C, and D. Which are most/least effective? Why?
- How does the range of Infielder A affect surrounding players? Infielder B?

| ELIGIBILITY

- The Data Challenge is open to STUDENTS ONLY. You must be enrolled as a high school, undergraduate, or graduate student for the Fall 2023 semester. Participants are expected to register using a .edu email address or similar. If you have questions, please contact smt.datachallenge@smt.com.
- Participants must be at least 18 years of age.

| REGISTRATION (Deadline: May 31)

Teams of up to four will compete in High School/Undergraduate or Graduate Divisions. You can register on the [CMSAC 2023 website](#) or by going directly to the [2023 SMT Data Challenge Registration Form](#). Data will be made available on or after June 1.

| SUBMISSION (Deadline: August 31)

- A short, narrative paper on your study in PDF format (max: 2000 words)
- A GitHub repo link containing code files and .csv files with results

A panel comprised of people from academia, industry, journalism, and sports (including team executives, players, umpires, scouts etc.) will judge your submissions based on the following criteria:

- How rich and textured (not necessarily big or small) is the question being asked?
- How applicable is the analysis?
- How appropriate were the methods used?
- How specific and narrative was the topic of investigation?
- How well did you communicate your findings? This includes both written text and visualizations. How did the use of facts, data-supported narratives, anecdotes, visual aids, etc. buttress storytelling?

| FINALISTS (Announced: October 4) AND PRESENTATIONS (November 11)

Three finalists from each division will present their work at CMSAC 2023 on November 11. Presentations will undergo a second round of judging, and division winners will be announced at the end of the conference.

| WINNERS AND PRIZES

- Division winners: \$1000
- Runners-up: \$250

| QUESTIONS?

If you have questions, please contact smt.datachallenge@smt.com.

| IMPORTANT DATES

MAY 31: REGISTRATION DEADLINE

DATA AVAILABLE ON OR AFTER JUNE 1.

AUG 31: SUBMISSION DEADLINE

OCT 4: FINALISTS ANNOUNCED

NOV 11: FINALIST CMSAC 2023 PRESENTATIONS

OPTIONAL ANALYSIS DEMO/Q&A SESSIONS:

- JUNE 14
- JULY 5
- JULY 26
- AUGUST 16