High School DataJam Guidebook 2027

DataJam Guidebook

Welcome to The DataJam! We hope this will be a rewarding and enriching experience for you. This guidebook will help guide your journey through the process. The guidebook is laid out as follows:

	Page
Overview of The DataJam	3
DataJam Step by Step Timeline	4-5
Proposal Guidelines	6
Proposal Template	7
Permission Slip	8-9
Poster and Presentation Guidelines	10
Judging Rubric	11
Scoring Guidance	12-13

Overview of The DataJam

What is The DataJam?

The DataJam is a team-based data science learning experience and competition designed to guide high school students through a complete data science project of their choosing. The DataJam provides ample resources to facilitate this journey including access to trained DataJam mentors, a collection of data sources, and a library of guides for each step of The DataJam competition. Detailed information about The DataJam and all of The DataJam Resources are available on the website, thedatajam.org.

What do I need to start a team?

Teams are formed by a minimum of **two participants** (high school students) and an **advisor** (teacher, club director or parent). A good way to get started is to go to our website and on the home page click on either the "Students" or "Teachers" pathway. Mid-way down the page there is a "Forms" section and you can click on "High School Forms". The first form you will see is the "**High School Team Sign-Up Form**".
From there, all you need to do is fill out the form and submit it. Once your sign-up form is received by The DataJam, you will receive an email confirming that we received your sign-up form. You will then want to follow the 2027 DataJam Progression to learn about each step your team should take for a DataJam project – see pages 4 & 5). Teams that request mentoring services will use **The DataJam Mentoring Platform**, a web-based platform hosted on Google Classroom, to communicate with mentors via discussion board/Google Meet videoconferencing and receive DataJam news. To ensure that teams are able to interact with mentors on the platform, **students are required to fill out a Student Permission Slip** (pages 8 & 9). Teams then work through three main steps, (Step 1) putting together a proposal (page 6 & 7), (Step 2) performing data analytics, and (Step 3) preparing a poster and presentation for judging (page 10).

Who are the mentors? And how does a team get connected to a mentor?

Mentors are university students who volunteer to help teams participating in The DataJam. DataJam Mentors are required to attend a training class or take an online training course that prepares them to be a DataJam Mentor.

When a team wants to work with a DataJam mentor, the team should indicate that they would like to receive help from our mentoring services when signing up for the competition (mentors can help with every step of a DataJam project from developing a research question to ask, developing a testable hypothesis, finding a publicly available dataset to analyze to answer their research question, undertaking data analyses, making graphs, preparing and practicing posters & presentations). Teams that request access to **The DataJam Mentoring Platform** will receive an email invite from Joey Stuter, the DataJam Project Manager, **only after the team has turned in a signed permission slip for each member**. On the Mentoring Platform, trained DataJam mentors will be on stand-by from **8:00 AM to 8:00 PM (ET) daily** to assist teams with all aspects of their projects.

2027 DataJam Step by Step Timeline

Step 1: Sign-up

Late August 2026 – January 2027

Each team needs to fill out the online High School Sign-up Form on The DataJam website (thedatajam.org) and student permission slips on page 8 & 9 to be sent to datajam@thedatajam.org.
The sign-up form will ask (1) advisor name and email, (2) team name, (3) student names and emails, and (4) whether the team would like to receive mentoring this year. If you are signing up more than one team with the same advisor, please fill out one form for each team. DO NOT fill out all teams on one form. ***Note that it is best to provide a non-school email address on the permission slip as many schools block outside emails such as those from The DataJam, and school emails will not work with The DataJam Mentoring Platform.**
Upon requesting mentoring, a channel for your team will be created on The DataJam Mentoring Platform and each team member will receive an email inviting them to join the workspace.

Step 2: Submit your proposal

December 4, 2026 – January 29, 2027

There is a rolling deadline for teams to turn in DataJam proposals. Use the proposal template on page 7 of this guidebook and submit it to the High School online proposal form on The DataJam's website. Your team will receive feedback from The DataJam and further guidance on your proposed project within a week after your proposal is turned in.

Step 3: Projects Underway

January – Late March 2027

Teams work on DataJam projects. DataJam Mentors are able to provide guidance to teams using The DataJam Mentoring Platform. Trained DataJam mentors will be on stand-by from 8:00 AM to 8:00 PM (ET) daily to assist teams with all aspects of their projects using text-based communication via discussion board and videoconferencing via Google Meet.

Step 4: Submit your Poster

Friday, March 26, 2027

DataJam posters are due. Follow the instructions on page 10 of this guidebook. Posters should be 24" x 36" in size and saved as a pdf file. They should be submitted to the High School Poster Submission Form on The DataJam website.

Step 5: Submit your Presentations and Attend a Judging Session

April 4 - April 20, 2027

Final slide presentations will be scheduled for each team at a time that works for both the team and the 3-judge panel, who will be listening to each team. The DataJam will email each team's advisor to find out potential days/times the team will be available for their judging session. The final presentations will be held on Zoom. Teams are to submit their presentation slides to the Presentation Submission Form on The DataJam's website at least 24 hours before their presentation is scheduled.

Step 6: Finale

Thursday, April 22, 2027

High School DataJam 2027 Finale! The finale will be held on Thursday, April 22, 2027 from 5:30-7:00 PM EST (3:30-5:00 PST), on Zoom link: https://pitt.zoom.us/j/94389494426 . DataJam teams, parents, teachers, mentors, and all who are interested are invited to attend. DataJam team projects will be presented, and a variety of awards will be given!

Proposal Guidelines

Create a one-page proposal describing the question your team is asking or the problem your team is addressing. The proposal should follow the proposal template on page 5 of this guide and should describe:

- A **problem or question** your team's project will address,
- **A statement of why the problem or question is interesting**,
- The team's **hypothesis**,
- What dataset(s) you will use to address your question (**provide links to datasets**),
- How you **plan to analyze the data** to address the problem. What types of **data visualizations** you expect to use.

Data: It is the responsibility of each team to find and download the dataset(s) needed for their project. The goal is not to have a really big dataset to work with (terabytes of data are not needed), but rather to have enough data to gain insights into the problem. Finding datasets can be difficult; DataJam mentors are available to help. Lots of help can be found on the ***Resource page*** of The DataJam website, thedatajam.org.

Visualization Tools: Simple visualizations can be made with programs such as Google Sheets and Excel. More complex visualizations can be made with Tableau Public, which is a freely available program, and in R and python. See videos for how to make visualizations on our ***Resources page*** of The DataJam website, thedatajam.org.

Statistical Analysis: Each team will decide the appropriate statistical method to apply to their problem. We have lots of resources available to help. See the ***many videos and tip sheets on the Resources page*** of The DataJam website, thedatajam.org**.** DataJam mentors are available to help think about what analyses to use and to teach teams how do various analyses.

Judging Criteria for the Final Project: Participants are encouraged, but not required, to consider tackling a problem in their community. The following criteria will be used by judges in evaluating the final poster and presentation that each team prepares:

- Form a research question based upon the relevant data
- Uncover and use appropriate data sources
- Analyze data/information
- Data Visualization
- Draw logical and defensible conclusions
- Communicate clearly and effectively findings and conclusions
- Answering Questions
- Recognize Limitations

See the judging criteria on page 11, with guidance for judging each criteria on pages 12 and 13.

DataJam Project Proposal Template

School Name (& Team Number if there is more than one team/school):

Project Title:

Team Member Names & Email Addresses:

Problem:
IN THE FORM OF A QUESTION STATE WHAT PROBLEM YOUR DATAJAM PROJECT WILL ADDRESS?

Why is it Important?:
IN A PARAGRAPH EXPLAIN WHY THIS IS AN IMPORTANT QUESTION AND WHO WOULD BE INTERESTED IN THE ANSWER.

Hypothesis:
A STATEMENT SHOULD BE MADE ABOUT WHAT THE TEAM THINKS THE ANSWER TO THEIR QUESTION WILL BE.

Data Required:
PROVIDE A LIST OF DATASETS YOU HAVE FOUND THAT YOU WILL ANALYZE TO ANSWER YOUR QUESTION. PLEASE INCLUDE THE **FULL LINKS** SO WE CAN LOOK AT THE DATASETS WHILE WE ARE LOOKING AT YOUR PROPOSAL.

Analysis Plan:
IN ONE OR TWO PARAGRAPHS DESCRIBE HOW YOU PLAN TO ANALYZE YOUR DATA, WHAT ANALYTICAL STRATEGIES YOU WILL USE AND WHAT DATA VISUALIZATIONS YOU THINK YOU WILL USE? (No work needs to have been done at the time the proposal is submitted)

DataJam 2027

Parent Permission Slip, Waiver of Liability, Assumption of Risks, Indemity Agreement

The DataJam (thedatajam.org) is the educational nonprofit that runs the High School DataJam each year to introduce high school students to learning about data science, how critical it is for understanding everything in our current world, and how to work with big data sets from a data visualization and statistics perspective.

We encourage all high schools to put together a DataJam team (2 or more students constitute a team; generally teams have 2-8 students). Each high school team is asked to have a teacher at their high school to serve as a DataJam Advisor. Teams are mentored by university students who are trained as DataJam Mentors in how to teach data science, best practices in mentoring, working with diverse communities, and how to give scientific presentations. Mentors are advised to use The DataJam Mentoring Platform on Google Classroom for mentoring and NOT share personal contact information, "friend" youth on social media, or have direct contact with youth.

The DataJam will make available The DataJam Mentoring Platform, hosted on Google Classroom, where all teams can easily get assistance from DataJam mentors for helping them think about the research question they want to study, how to find data sets, how to analyze data, how to make graphs and how to put together a presentation. They will also be able to ask to have a Zoom meeting with a mentor so they can get assistance at any time that works for them. All interactions between mentors and students will be online and by videoconference. Mentors will be advised to NOT hold Zoom conferences with single students.

The timeline for this year is (1) Fall 2026 through January 2027 – teams sign up to participate in High School DataJam 2027, (2) December 4, 2026 – January 29, 2027 – team proposals should be submitted, (3) Jan-Mar 2027 – teams work on their projects with help of mentors, (4) March 26, 2027 - posters due, (5) April 22, 2027 – High School DataJam Finale on Zoom – teams give presentations about their projects and many awards are given.

Waiver: In turn for being permitted to participate in DataJam 2027 I do hereby release, waive, discharge, and promise not to sue The DataJam, or any of the colleges or universities training DataJam mentors, from any and all claims, including negligence and sexual misconduct in connection with my child's participation in High School DataJam 2027.

Assumption of Risks: Participation in DataJam 2027 carries with it certain inherent risks that can not be eliminated regardless of the care taken to avoid risks. Risks vary from hurt feelings to exposure to inappropriate language or behaviors.

Indemnification and Hold Harmless: I also agree to indemnify and hold The DataJam and all universities training DataJam mentors harmless from any and all claims, actions, suits, procedures, costs, expenses, damages and liabilities, including attorney's fees, arising out of my child's involvement in DataJam 2027.

Severability: I further agree that this Waiver of Liability, Assumption of Risk, and Indemnity Agreement is intended to be as broad and inclusive as permitted by law, and that if any portion is held invalid the remaining portions will continue to have full legal force and effect.

I am the parent of _____, who is a current high school student at _____High School. I give my permission for my child to participate in online videoconferences and use The DataJam Mentoring Platform on Google Classroom to participate in the High School DataJam 2027. I also give permission for The DataJam to post pictures from videoconferences of my child participating in High School DataJam 2027 activities.

Acknowledgment of Understanding: I have read this Waiver of Liability, Assumption of Risk, and Indemnity Agreement, fully understand its terms, and **understand that I am giving up substantial rights, including my right to sue.** I confirm that I am signing the agreement freely and voluntarily, and **intend my signature to be a complete and unconditional release of all liability** to the greatest extent allowed by law.

_____ _____

Parent Signature Date

_____ _____

Printed Parent Name Student contact email address (a publicly available email address works best as many high schools block emails from outside organizations such as The DataJam).

Scan and EMAIL TO datajam@thedatajam.org

Poster and Final Presentation Guidelines

Data analytics consists of two broad tasks: (a) extracting meaningful information from raw data and (b) explaining the impact and value of that information. Your poster and final presentation for The DataJam should show how your DataJam project accomplished these two tasks.

Poster Guidelines:

You should follow the guidelines below for the development of your poster. However, we encourage you to exercise your creativity in designing your poster and how you would like to communicate the details of your project.

These three guidelines are essential:

- Poster size: 24" x 36". The poster orientation can be vertical or horizontal.
- Your poster should clearly display: (1) Title of your project, (2) The full names of the all team members, (3) the name of your school.
- Submit your poster as a .pdf file to the **poster form** on The DataJam's website by March 26, 2027.

In your poster you will want to have information addressing the questions below. Remember, pictures, graphs, tables and diagrams are very helpful ways of presenting your results.

- What is your research question?
- What was your hypothesis when you started the project?
- What dataset(s) did you use?
- Briefly describe the analyses you did to answer your question. Show graphs and diagrams if relevant.
- What challenges did you face in finding or using the data?
- Summarize your results.
- Can you make a recommendation based on your results?

Except for the essential guidelines, which each team must follow, you can tailor your poster according to your project. Remember, judges will view your poster without you being there to explain it.

Presentation Guidelines:

From April 4 – April 20, 2027 your team will be scheduled to give a 10-minute oral presentation, using slides to a panel of three judges. This presentation will be scheduled at a time that is mutually possible for both your team and the judges. The presentation will be conducted on Zoom.

The oral presentation should address the same information listed above for the poster. Remember to not have too many slides for a 10-minute talk; 10-13 slides is appropriate.

It is best if each member of your team can have some speaking role in the presentation and question answering. A 10-minute period will be allotted for questions.

DATAJAM JUDGING RUBRIC - 2026-27	Poster Weighting	Presentation Weighting
Uncover and use appropriate data sources.	**15.0%**	**10.00%**
Variables in the dataset are directly relevant to answering the question		
The sample size is large enough to answer the question, dataset is from a reliable source		
Form a research question based upon the relevant data.	**20.0%**	**10.00%**
Question is specific and can be answered with data analysis, Question is meaningful to stakeholders		
Analyze data/information.	**25.0%**	**30.00%**
Sophistication of the analyses,		
Was the appropriate methodology used?		
Data Visualization	**15.0%**	**10.00%**
Does the visualization help explain the answer to the question? Is the visualization easy to understand and interpret quickly? Is the visualization well-explained?		
Draw logical and defensible conclusions.	**13.0%**	**15.00%**
Conclusions answer the question and are supported by the results		
Multiple iterations of questioning based on initial results		
Communicate clearly and effectively findings and conclusions.	**12.0%**	**10.00%**
All categories required on poster are present		
Sections of poster are easy to find and understand		
Analyses done are well explained		
Overall quality of the poster/slides		
Answering Questions		**10%**
All team members participate		
Answers are well thought out (not guessing)		
Limitations		**5%**
Limitations about relevance of the findings to the specific data used are recognized		
Total	**100%**	**100%**

DataJam Judging - Scoring Guidance	0	25	50	75	100
Uncover and use appropriate data sources.					
Variables in the dataset are directly relevant to answering the question	No relevant variables; data does not connect to research question.	Few variables loosely related; connection to question unclear.	Some variables relevant but not fully aligned with question.	Most variables are relevant and help partially address the question.	All variables are highly relevant, clearly aligned with and essential to answering the research
Is the sample size large enough to answer the question	Sample size is too small to provide meaningful results.	Sample size is very limited; results are unreliable.	Sample size is somewhat adequate but restricts confidence in results.	Sample size is mostly sufficient for meaningful conclusions with some caveats.	Sample size is clearly sufficient, providing robust and reliable results.
Dataset is from a reliable source	Dataset source is unknown, fabricated, or clearly unreliable.	Dataset comes from questionable or biased sources.	Dataset source is partially reliable but lacks credibility or transparency.	Dataset comes from a generally reliable source with minor concerns.	Dataset is from a highly reliable, authoritative, and verifiable source.
Form a research question based upon the relevant					
Question is specific and can be answered with data analysis	No clear research question stated.	Vague or overly broad question; difficult to analyze.	Question is somewhat specific but still ambiguous in scope.	Question is mostly specific and suitable for analysis.	Question is very clear, specific, and perfectly structured for data
Question is meaningful to community (school, town, state, etc)	No connection to community relevance.	Minimal relevance; weak or superficial connection.	Somewhat relevant; partially meaningful to a community issue.	Clearly relevant to the community with noticeable value.	Highly meaningful, impactful, and directly connected to an important
Analyze data/information.					
Sophistication of the analyses	No analysis performed.	Very basic descriptive statistics only.	Some intermediate analyses; limited depth or creativity.	Advanced analysis attempted with solid execution.	Highly sophisticated and creative analysis applied; industry-grade quality.
Was the appropriate methodology used?	No clear methodology applied.	Inappropriate or incorrect methodology used.	Methodology partially appropriate but not well-justified.	Appropriate methodology used with good reasoning.	Excellent methodology, well-justified and aligned with best practices.
Data Visualization					
Does the visualization help support the answer to the question?	No visualizations provided or unrelated to question.	Visualizations present but do not support the research question.	Some visualizations relevant but confusing or incomplete.	Most visualizations are clear and help explain the question.	Visualizations are highly effective, insightful, and directly support the research question.
Is the visualization easy to understand and interpret quickly?	No attempt at clarity; confusing or unreadable visuals.	Difficult to interpret; poorly labeled or cluttered.	Somewhat understandable with effort; lacks polish.	Mostly clear with minor readability issues.	Extremely clear, intuitive, and professional-level visuals.
Is there a good explanation of the visualization?	No explanations provided.	Minimal explanations with little clarity.	Some explanations present but incomplete or confusing.	Mostly clear explanation with minor readability issues.	Very clear explanation.
Draw logical and defensible conclusions.					
Do the conclusions answer the question?	No conclusions provided or unrelated to question.	Conclusions presented but not connected to the data.	Conclusions somewhat related but lack direct evidence.	Conclusions mostly supported by the data and analysis.	Conclusions directly and fully answer the research question with strong

Conclusions are supported by the results	Conclusions are unsupported or contradicted by the results.	Weak support; results only loosely tied to conclusions.	Partial support; some conclusions fit but others don't.	Strong support with minor inconsistencies.	Conclusions fully supported by results with clear logical alignment.
Multiple iterations of questioning based on initial results	No evidence of iteration or refinement.	Minimal iteration; one superficial follow-up.	Some iteration present but limited depth.	Clear iterative approach with meaningful refinements.	Extensive iteration and refinement leading to deeper insights and stronger conclusions.
Communicate clearly and effectively findings and					
All categories required on poster are present:	Many required sections missing.	Several sections missing or incomplete.	Most sections present but a few weak or	Nearly all required sections complete and	All required sections present, thorough, and
Sections of poster are easy to find and understand	Poster is disorganized; sections are missing or unclear.	Poster organization is confusing; sections hard to find.	Some organization present but difficult to follow.	Mostly well-organized with minor navigation issues.	Highly organized and intuitive; sections are clear and easy to follow.
Analyses done are well explained	No explanations provided.	Minimal explanations with little clarity.	Some explanations present but incomplete or confusing.	Analyses mostly well explained with good clarity.	Analyses fully explained in a clear, professional manner.
Overall quality of the poster/slides	Poster/slides missing or not usable.	Poster/slides very poor in quality and clarity.	Poster/slides of average quality with notable weaknesses.	Poster/slides are good quality with only minor weaknesses.	Poster/slides are excellent, polished, and highly professional.
Answering Questions					
All team members participate	No participation from team members.	One member participates minimally.	Some members participate, others excluded.	Most team members participate actively.	All team members equally engaged and contributing.
Answers are well thought out (not guessing)	Answers show no thought or accuracy; pure guessing.	Answers show minimal thought; mostly guessing.	Some answers reasonable but lacking depth.	Most answers thoughtful with good reasoning.	All answers show deep thought, clear reasoning, and strong evidence.
Limitations					
Limitations about relevance of the findings to the specific data used are recognized	No recognition of limitations.	Minimal recognition of limitations; superficial mention.	Some limitations identified but not deeply explained.	Most relevant limitations identified and explained clearly.	Limitations thoroughly identified, well explained, and thoughtfully integrated into