

The DataJam Download

Official Newsletter of the Pittsburgh DataWorks



***DataJam 2023
Finale Winner
Presentation***

***Middle School
Data Science Day***

New DataSet Guide

***Meet the
Data Science
Professional***

***DataJam Mentor
Re-Connects with
DataJam Team***

***DataJam Mentor
Wins National
Award***

***National Workshop
on Data Science
Education***

***DataJam 2023 Finale Winning Team Presents at Google in Pittsburgh
See Video of the Team Presentation on pghdataworks.org Homepage***



North Allegheny High School – Team #1

Middle School Data Science Day

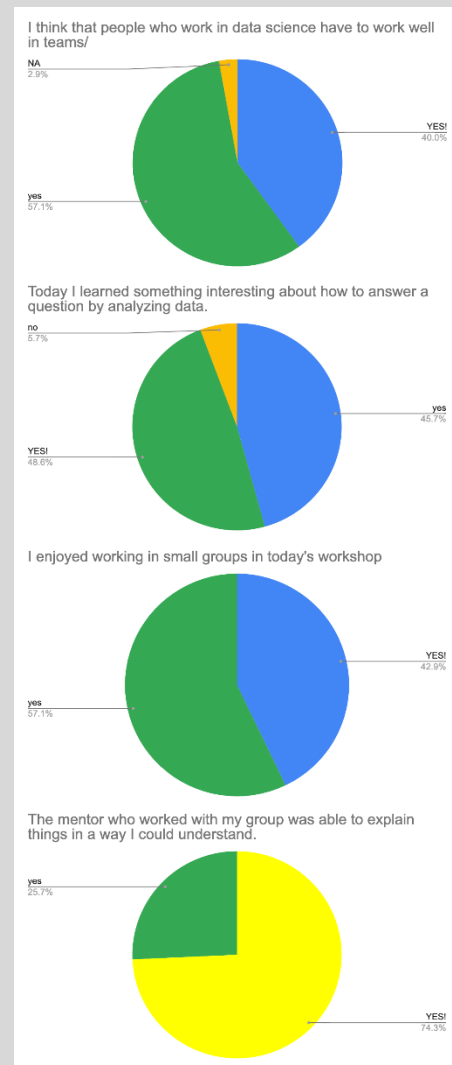
By Jackson Filosa, DataJam Mentor

Since 2014, the DataJam has been held as a competition for high school students interested in learning about data science and leveraging its power for community good. Through the efforts of Pittsburgh DataWorks, community sponsors, partnering schools, DataJam mentors, teachers, and most importantly the students themselves, the DataJam has continued to expand, impact, and lead the way as a data science educator in an increasingly data-driven world.

This year, the DataJam took another exciting step in its efforts to increase accessibility and awareness of data science. It all started at the Science Olympiad held at the University of Pittsburgh in January 2023, when Zach Lindquist and I had the opportunity to give the DataJam Overview presentation for Science Olympiad teams who were attending the annual competition that was held on Pitt's campus. In doing so we sought to connect with a new audience, middle school students, and gauge whether they would be interested in learning more about data science. The answer from surveys collected from more than 100 attendees was a resounding YES! As a result, plans were developed to conduct a Data Science Day for local middle school students as a full-fledged introduction to data science and the DataJam. Through the tireless efforts of local middle school teachers led by Dr. Melissa Survinski of Hampton Middle & High School and DataJam Mentors, a Data Science Day was hosted on March 31st at the North Allegheny High School by Pittsburgh DataWorks. The event included an introductory data science presentation, followed by small group activities planned by DataJam mentors, and ended with a brief final presentation about the DataJam.

DataJam Mentors attending the event were myself, Anant Bajwa, Erin O'Neill, Tony Robol, Jay Singh and Yunge Xiao. Mentors did an incredible job planning fun introductory data science projects for their small groups. Topics chosen were specifically tailored to a middle school audience in terms of subject and difficulty, covering topics such as changes in artificial intelligence over time, DC vs. Marvel movie comparisons, popular dog names in Allegheny County, sports team wins vs. a team's payroll, air quality measured against human activity, and UFO sightings around the United States. Using custom-picked datasets, groups were then led through hands-on activities that included data exploration, analysis, and visualization, finishing with a five-minute presentation to the other groups.

I am happy to report that this day was an incredible success. From the perspective of a DataJam mentor, it was incredibly exciting to be working with middle school students and seeing their enthusiasm for data science. Teachers remarked that the students were very engaged and seemed to benefit a lot from the workshop. We were all blown away by the final presentations and what the groups were able to accomplish in just one morning. Of course, as DataJam mentors, we know that the success of an event should be measured in data! And so, we collected survey responses from the thirty-six middle school students attending the event as seen to the right. We are happy to report that 91.4% of students expressed continued interest in learning about data science, with 34.3% of these students choosing the highest degree of enthusiasm listed on the survey. We are inspired to offer more events for middle school students in the upcoming year!



New DataSet Guide

We have another new DataSet Guide to guide DataJam teams as they choose research questions to study next year and search for datasets. These are now available on the DATASETS page of pghdataworks.org!

The DataSet Guide on ***Asking Questions about Sexually Transmitted Diseases (STDs)*** provides an overview about sexually transmitted diseases, what the symptoms of each are, and how safe sex practices decrease the rate of STD transmission. Potential research questions DataJam teams could ask are outlined and there are links to datasets that can be used to answer questions.



Meet the Data Science Professional

Michelle Malizio

Hello everyone! I'm Michelle, and I graduated from the University of Pittsburgh this spring with a double major in Biology and Psychology and a minor in Chemistry. I will be starting my master's in Genetic Counseling this fall at the Massachusetts General Hospital Institute of Health Professions, located in Boston. For the past two years, I've worked in Dr. Alexander Yatsenko's lab at Magee-Womens Research Institute, studying the genetic causes of male infertility.

Through my classes and research, I've discovered a love for genetics and have developed a great appreciation for scientific research and data science. Not only have I grown my own knowledge, but I hope to use that knowledge to educate others and make a meaningful impact in my community. As I pursue genetic counseling, data science will provide me with better insight when analyzing genomic data and will help improve patient outcomes.

In my free time, I enjoy traveling, running, cooking, and exploring new places! I also spend time volunteering as a direct service crisis advocate at the Women's Center & Shelter of Greater Pittsburgh, where I advocate for and support survivors of domestic violence. During my time at Pitt, I served as Chair of Pitt Best Buddies' Inclusion Committee, which gave me the opportunity to create life-long friendships with individuals with and without disabilities. I'm passionate about fostering more inclusive communities and increasing access to care for those in need.

DataJam is a wonderful way to explore the importance of data science, while also discovering your own scientific interests. My advice for you is to stay curious, embrace challenges, and follow your passions!



DataJam Mentor Re-Connects with the DataJam Team She Mentored in 2022

Louise Hicks was the DataJam mentor for DataJam team from the Pala Native American Reservation outside of San Diego, CA in the 2022-23 academic year. The Pala team won the "Best New Team" award at DataJam 2022. Louise presented the team's work and her mentoring at the "Practice and Experience in Advanced Research Computing 2022" or PEARC '22 conference in Boston, MA in July 2022 (see September 2022 newsletter). Louise mentored remotely by Zoom, but she recently had the opportunity to visit the team in person in May 2023. We are delighted to have her report of her visit in this month's newsletter below.

“This past month, in early May, I got the pleasure of being able to visit a set of Pala students, whom I mentored for DataJam 2022, in collaboration with advisors from the San Diego Supercomputing Center and Pittsburgh DataWorks. After having spent months during the DataJam season spanning Fall 2021 - Spring 2022 meeting with these students on Zoom, I was quite excited to meet for the first time in person. Plus, I was lucky enough to have been in the San Diego area during a celebration which happens only once a year: Cupa Days. I attended the 48th Annual Cupa Days celebration. This event featured a traditional performance called Birdsinging. Not only were there performances but there was food and booths which had items such as handmade jewelry and textiles. Lastly, the festival featured the reveal of a children’s book being written by members of the Cupa Cultural Center which will aim to educate members of the community about their native language Cupa. As I soaked in the warm Californian sun and all the new things I was learning about the community, I realized that on that day I wasn’t the mentor but instead the student.” – Louise Hicks



DataJam Mentor Wins National Award

Yunge Xiao was the DataJam Mentor for the DataJam team that won the 2023 DataJam – North Allegheny Team 1, with their project “Standardized Testing: A Reflection on Intelligence or Economic Environment?”. She is also the winner of a prestigious national award, “The Newman Civic Fellowship”, that recognizes standout students who are committed to creating positive change in communities locally and around the world, proving them with a year of learning and networking opportunities that emphasize person, professional and civic growth.

Yunge graduated from the University of Pittsburgh this spring (2023) with a Bachelor’s degree in Statistics. She was born in Singapore but primarily grew up in Berwyn, Pennsylvania. She became very interested in using data for community change in her first year at Pitt when she interned for the Pittsburgh Human Rights City Alliance (HRCA) and Pittsburghers for Public Transit. She started serving as a DataJam mentor in the Fall of 2022, when she took the DataJam Mentor training course at Pitt, and she mentored the winning team in DataJam 2023: North Allegheny Team 1.

Yunge will begin a master’s degree in epidemiology at Harvard’s School of Public Health this Fall. The Newman Civic Fellowship will provide her with virtual training and networking opportunities to give her the skills and connections to create large-scale positive change for the world. We are so proud of Yunge, an outstanding DataJam Mentor and an outstanding data scientist --- using her training to improve the lives of those around her!



National Workshop on Data Science Education

Pittsburgh DataWorks has participated in the planning and organization of the 2023 National Workshop on Data Science Education that will be held in Berkeley California (and virtually by Zoom) on June 20-23, 2023. The morning of Thursday, June 22nd will feature a discussion of project-based learning for data science, and the development of a national pipeline for project-based learning. The **DataJam** will be a key exemplar in this discussion. Anyone interested in attending this session June 22nd is welcome to register at this [link](#). Registration is free. Dr. Cameron and Dr. Cramer from the DataWorks advisory committee will be presenting, as well as schools that are collaborating in offering the DataJam and DataJam mentor training.



The screenshot shows the registration page for the 2023 National Workshop on Data Science Education. The page header includes the Berkeley logo and navigation links for About, Academics, Research, Equity & Inclusion, and News & Events. The main content area features a blue background with a cityscape and a hexagonal pattern. The text on the page includes: "2023 National Workshop on Data Science Education", "This workshop is organized by UC Berkeley's Division of Computing, Data Science, and Society with support from Microsoft and the West Big Data Innovation Hub", a "REGISTER HERE" button, "In Person - Hybrid format for 2023", "Location: Berkeley, California", "Dates: Jun 20-23, 2023", and a "Registration" section stating "The workshop is free for attendees. Register for the 2023 National Workshop for Data Science Education here" and "Hybrid / Online will be made available through a Zoom event".

We are looking forward to seeing you at DataJam 2024 next year!

If you would like the DataJam Newsletter sent to your personal email address, just email us at

DataJam@pghdataworks.org.