PITTSBURGH DATAWORKS

April 2023 // Issue 9   **www.pghdataworks.org**

The DataJam Download

Official Newsletter of the Pittsburgh DataWorks

DataJam 2023 Events

Final Posters Due – March 31st, 2023 (*ask for time extension if needed)

• See the DataJam Guidebook's instructions for preparing your posters (can be downloaded on the DataJam page of pghdataworks.org website)

• Email your poster to us at datajam@pghdataworks.org as a pdf file

Schedule a Final Presentation – April 21, 24, 25, 26

• Email datajam@pghdataworks.org to let us know when you would like to schedule your team's final presentation to judges

• Times available 4/21/23, 4/24/23, 4/25/23 and 4/26/23 from 10 AM – 2 PM, EST (provide us with your first, second and third choices)

DataJam Finale – Thursday April 27th (5:30-7:00 PM EST)

• All team projects will be presented

• A number of awards will be given to DataJam Teams

New Resources for DataJam Teams

New DataSet Guides

The feedback we have received from many DataJam teams the last few years is that one of the most difficult things is finding the right dataset to analyze to answer research questions that DataJam teams pose. We have been listening and a year ago we started to tackle this problem by starting to develop DataSet Guides. Each guide is designed to help DataJam teams tackle answering a research question about a particular topic. This year we had guides available for several biomedical topic areas including "Asking Questions about Heart Disease (heart attacks, coronary artery disease, high blood pressure)", "Asking Questions about Lung Diseases (Covid-19, asthma, COPD)", "Asking Questions about Brain Diseases (strokes, Alzheimer's Disease, mental health)" and "Asking Questions about Nutritional Issues (obesity, diabetes)". We also had guides available about several economic issues including "Asking Questions about Inflation", "Asking Questions about Supply and Demand" and "Asking Questions about Recessions". For each guide some background information is provided so that you understand the topic area. Example research questions that DataJam teams might ask are provided. For example, for heart attacks you could ask "What time of year do heart attacks occur the most?" or "What cities, counties, parts of the country have the most heart attacks?" or "Does the density of fast-food restaurants or food prices affect heart attack rates?" Accompanying the presentations are links to a variety of datasets that can be analyzed to answer these questions.

This year we have worked with several teams of scientists at the University of Pittsburgh to develop guides on new topic areas. The first of these new DataSet Guides is being released this month! This guide focuses on "Asking Questions about Male Reproduction and Infertility". The example research questions that DataJam teams might ask include: "Does the infertility rate vary between countries of different socioeconomic status?", "Has fertility in the USA been on the decline in recent years?", and "Do environmental contaminants impact human fertility

Asking Questions about the Male Reproductive System and Infertility

rates?" The scientist who worked with us on developing this DataSet Guide is Dr. Georgia Atkins. Learn more about Dr. Atkins in "Meet the Data Science Professional" section of this month's newsletter (below)!
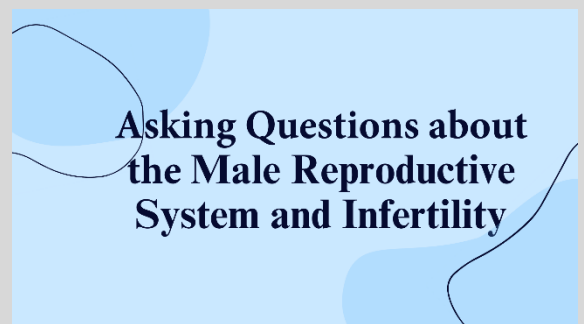
A New "DATASETS" Page on the Pittsburgh DataWorks Website

To make it easier to find our DataSet Guides and other resources we
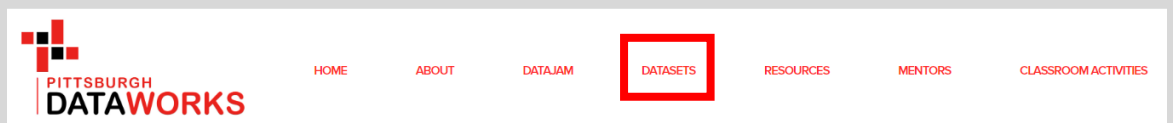
have for helping teams find datasets for analysis, we have revised our website and we now have a new page called "DATASETS"! Read our new introduction on how to find datasets to answer your DataJam question at the top of the page. Then scroll through the pictures of DataSet Guides that are currently available. Click on a guide that looks interesting to you and a brief summary of the guide comes up. If you are interested in it, click on the picture and the whole guide will appear and you can download a copy if you want. Below the DataSet Guides there are several links to national databases that teams will find useful for answering a wide variety of research questions. Finally, at the bottom of our new page are regional databases. Currently there are many DataJam teams in three hub areas: Pennsylvania, New Jersey and California so we are working hard to add regional databases from these areas. However, DataJam is constantly expanding to new parts of the country and with this expansion we will expand the availability of regional databases.

Meet the Mentor

Hello everyone! My name is Kevin Xu, and I'm a senior stats major here at the University of Pittsburgh. I remember the first time I took a statistics course was during my junior year of high school. Statistics was one of my favorite AP courses, but as I reflect upon it now, I had been doing "stats" since way before then. I remember when I was walking to elementary school, I would time the duration of the different paths I could take to ascertain the most efficient route. I remember meticulously cataloging my iTunes library in middle school, wondering which albums had the longest run time, which tracks I would listen to the most, and other simple analytics that would provide insight into my listening habits. When I finally took my AP statistics class, my interest in this field was cemented. Statistics has always been a tool that I've used to represent and visualize the world around me, and finally receiving formal education on the topic opened my eyes to a potential career path I had not considered before.

As I entered my first few years of undergrad at the University of Pittsburgh, I was still a bit overwhelmed with the number of majors I could choose from. I have always had a wide variety of interests and this fact made it difficult to commit to a single field. Psychology, pharmacy, and public health were all areas that I considered focusing on, but the realization that the research-based nature of these topics allowed for all of them to have statistical applications helped me finally decide on becoming a stats major.

Additionally, as I considered the current climate of our digital world, I realized how fundamental big data is to the world today. The internet has completely changed the way the world works by creating a new, unique set of issues for younger generations to tackle. Just as data helped me gain insight into the minutia of my everyday life, the goal of my statistics education is to fully harness the power of stats and big data, to utilize this tool in a way that would help inform decisions for positive change. Since there is already a strong culture of medicine and public health at Pitt, I wish to use the statistical analysis skills I have not only aid the development of research in these fields, but also to help those who make decisions at the government level to properly allocate resources to areas that need them the most.

Meet the Editor

Greetings everyone, my name is Lucas Troy, and I have the honor of serving as the Website Designer and Newsletter Editor at Pittsburgh DataWorks. I am currently pursuing a major in Computer Engineering at the University of Pittsburgh, and when I'm not engaged in my studies, I devote my time to personal coding projects, data analysis, weightlifting, and spending time with my close friends.

During my freshman year, I had the privilege of serving as a DataJam mentor, an experience that truly instilled in me the value of imparting knowledge to young individuals. Data literacy has become a crucial aspect of our society, and it's fantastic to see Pittsburgh DataWorks paving the way for young students to learn how to use it responsibly. It was this experience that led Dr. Cameron to ask me to take over the role of managing Pittsburgh DataWorks' website, which I have been doing since the summer of 2021. As the Website Designer and Newsletter Editor, my primary responsibilities include designing educational resources, organizing content, and promoting the DataJam. I also create monthly analytics reports that allow us to track the organization's progress and share it with the board. In just a year, we've made incredible strides, and it has been an honor to be a part of this growth.

Together with Dr. Cameron, we have developed a highly informative newsletter that has reached a vast audience. It serves as a platform to share resources developed by students in Dr. Cameron's class, important dates for the DataJam timeline, awards won by the organization, and more. We have received positive feedback from our readers, and it has been gratifying to see our efforts make an impact.

Thank you for taking the time to learn more about me and my work at Pittsburgh DataWorks. I look forward to continuing to serve the organization and contributing to its continued success.

Cumulative Visits

— Cumulative Total 21-22   — Cumulative Total 22-23
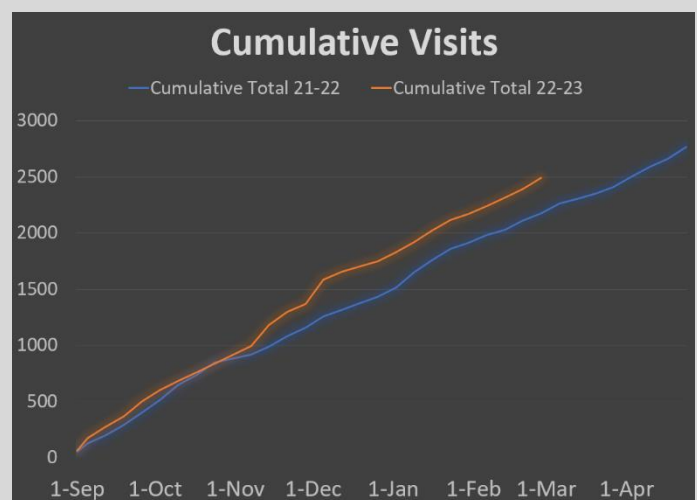
Meet the Data Science Professionals

Georgia Atkins

Hi everyone! I'm Georgia and I'm a second year PhD student with the University of Pittsburgh. I'm originally from Las Vegas, NV and moved to Pittsburgh to go to graduate school. I love going on hikes around the city, grabbing coffee, going to yoga, and trying out new places. This is my first year helping with DataJam but I'm so excited to be here!

I had a non-traditional path to pursuing a career in science. I grew up thinking I wanted to be an elementary teacher until I started college. I had some life changing teachers in my senior year of high school that taught me science and math through the lens of real-world applications, rather than just a textbook page. I was understanding the material on a deeper level, and it became easier to grasp. I went my whole life thinking I was bad at science and math but now I realized I enjoyed solving the puzzle. That made me think "what if I can pursue a career in STEM? What if my struggles in the past were due to poor teaching and not my ability to understand the material?" Throughout college, I continued to seek strong mentors that helped make science challenging but fun.

I decided to get a PhD so that I have the option to teach. I single-handedly have seen that success in STEM is due to supportive mentors who want to help you reach ambitious goals. Throughout my career, I've realized that you don't have to be the smartest person in the world to make a discovery in science, you just need a passion for what you do. Data science helps make science accessible to everyone. There are thousands of datasets collected through public programs that are just waiting to be analyzed. You can make a discovery with these datasets without having to work in a lab and the knowledge learned is just as important.

I encourage you to use DataJam to have fun and get a taste of what a career in STEM could be. Don't be afraid to come up with a hard question or do something really out of the box. Yes, you'll be challenged, and it may be hard at times, but DataJam is full of mentors who are eager to guide you through this process!

Brian Macdonald

As a Master Principle Cloud Architect and Data Scientist with Oracle, I have had the opportunity to work on many innovative projects throughout my career. Although I have spent most of my career at Oracle, I have also worked as a Data Scientist at both Teradata and IQVIA, gaining valuable experience in the field of data science.

I received my bachelor's degree in Information and Decision Systems from Carnegie Mellon University and went on to earn a Master's degree from Northwestern. During my studies, I developed a strong passion for analytics and data science that has only grown stronger over time.

One of my earliest projects involved analyzing my personal record collection using an Apple II+ computer. This experience taught me the power of data analysis and how it can be used to uncover valuable insights. Today, I am still passionate about using data science to solve complex problems, and I have had the opportunity to work with many Fortune 500 companies throughout my career.

One of my favorite areas to apply data science is in sports analytics. I have worked with several MLS teams, helping them leverage data science to improve team performance. I have also worked with the English Premier League, analyzing location and event data to create The Most Powerful Goal and Most Improbable Comeback awards presented by Oracle. An example of the analysis can be seen in the Milestone Awards. I am also proud to work with the Oracle Red Bull Racing eSports team to help them analyze racer performance in the F1 game.

Outside of my professional work, I am passionate about helping others get excited about data science. As a member of the advisory board of the DataJam, I am dedicated to promoting the field of data science and encouraging the next generation of data scientists to pursue their interests.

Throughout my career, I have remained committed to innovation and creativity, using data science to create value for my clients and help them achieve their goals. Whether it's analyzing record collections or improving team performance, data science has the power to unlock valuable insights and drive meaningful change. I am grateful for the opportunities I have had to apply my skills and knowledge in this exciting and ever-evolving field whilst leveraging the newest technologies.

We are looking forward to seeing you at DataJam 2023!
April 27th, 2023