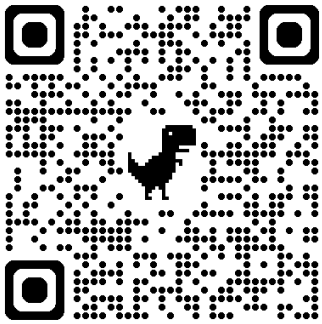
Zimam 2025 Healthcare AI Datathon

QR code for this document



QR code for the Datathon only <u>temporary</u> access

SCAN ME

If you want to continue having access after the datathon, please signal the event organizers.

Table of contents

Dates and Venue:

- November 6, 2025 | Jumeirah Emirates Towers
- November 7-8, 2025 | Center for Innovation & Technology, Dubai Health

Welcome to the Healthcare Datathon!

The Datathon is an intensive, collaborative event that will run over two and a half days. Multidisciplinary teams of healthcare professionals, data scientists, AI researchers, and experts will work together to answer clinical questions utilizing big data, as well as develop innovative AI-driven solutions for real-world healthcare challenges.

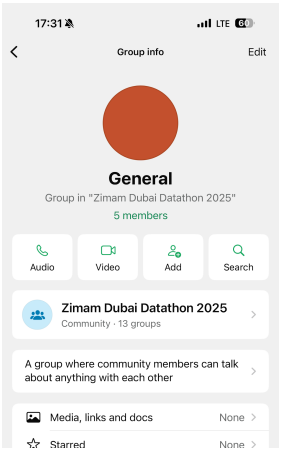
During the Datathon, participants will analyze real-world datasets and apply analytical techniques to answer important clinical questions and develop predictive models that address critical aspects of patient care such as disease prediction, diagnosis, and treatment. We will also focus on important challenges surrounding the use of AI in healthcare, which include data and algorithmic bias, ethical considerations, and the need for equitable and transparent AI solutions that serve diverse patient populations.

In addition to hands-on problem-solving, the datathon will include mentorship from experienced clinicians, AI experts, and data scientists, providing valuable insights into both the technical and clinical aspects of AI in healthcare. The event will conclude with a final presentation, where teams will showcase their findings and proposed solutions to a panel of judges.

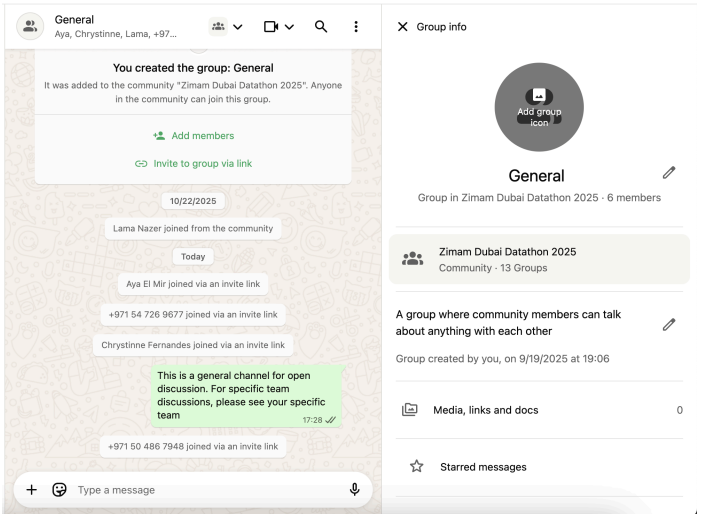
Files

- Dubai Zimam Datathon 2025
  - Welcome Package and FAQs ZIMAM Datathon 2025.docx
  - Schedule ZIMAM Datathon 2025.docx
  - reading material
  - **Find your TEAMS:** ZIMAM Datathon Team Distribution 2025.docx
- welcome to 2025 Zimam datathon.ipynb
- welcome-participants.Zimam Datathon 2025.ipynb
- GitHub: welcome-participants
- GitHub: Zimam Datathon 2025 repos

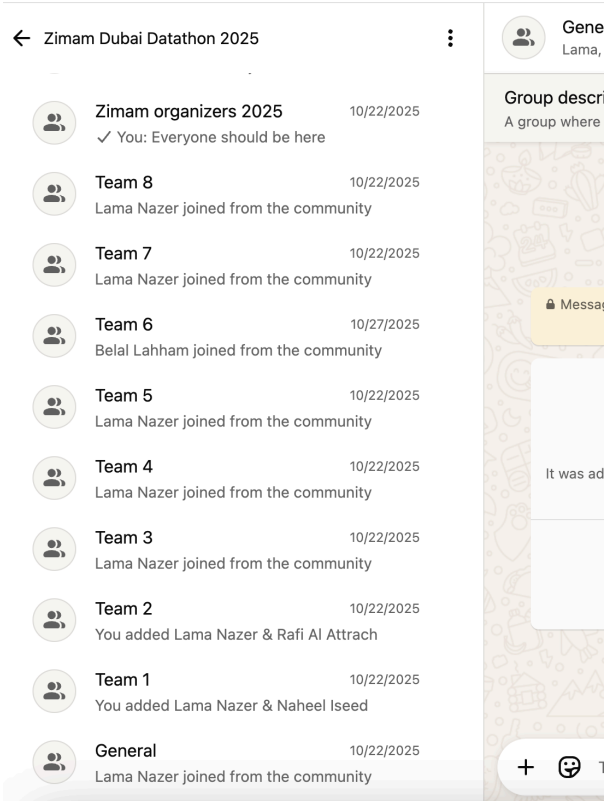
Communication

- **WhatsApp community:** https://chat.whatsapp.com/DcgxTg2lNX34snUc0NP9wN
  - First, join the “General” channel.
  - Second, click on the **“Zimam Dubai Datathon 2025”** channel
    - Mobile: 

- Desktop:



- Then, join your specific team



Datathon Program

ZIMAM Healthcare Datathon/ Dubai, UAE

November 6-8, 2025

Day 1- Thursday, November 6, 2025

Time	Activity
3:00 pm	Welcome & Opening Remarks Overview of Datathon (Structure and Expectations) ***Lama Nazer/ King Hussein Cancer Center/ Jordan***
3:15 pm	Overview of the Datasets and Tools ***Ian Wong/ Duke University/ USA***
3:30 pm	Bias in Artificial Intelligence and Strategies to Mitigate ***Lama Nazer/ King Hussein Cancer Center/ Jordan***
3:50 pm	Healthcare AI and the Agentic Future ***Steve Bethke/ Mayo Clinic Platform/ USA***
4:10 pm	Beyond Accuracy: Building ML Pipelines that See Everyone ***Aya El Mir/ New York University Abu Dhabi***
4:30 pm	Team Distribution and Introductions/ Set-up Access to Datasets ***Chrystinne Fernandes/ Massachusetts Institute of Technology/ USA***
5:00 pm	Closing & End of Day 1

Day 2- Friday, November 7, 2025

9:00 am	Briefing Session
9:15 am	Practical Tips and Tools ***Rajna Fani & Rafi Al Attrach, University of Munich***
9:30 am	Datathon Teamwork
11:00 am	**Checkpoint 1** Team consults with guest instructors and mentors. Be ready to discuss the refined team question with rational and clinical implications.
11:30 am	Datathon Teamwork
12:30 pm	Lunch & Prayer Time
1:30 pm	Datathon Teamwork
2:30 pm	**Checkpoint 2** Team consults with guest instructors and mentors. Be ready to discuss your analytical plan and variables/ features you plan to evaluate. Address sources of bias and include rationale for your plan.
3:00 pm	Datathon Teamwork
3:30 pm	**Team presentations** Consists of 5-minute presentation of the team research question, rational, clinical implications, and analytical plan/ sources of bias.
4:30 pm	Closing & End of Day 2

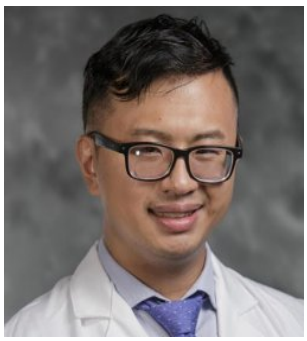
Day 3- Saturday, November 8, 2025

9:00 am	Briefing Session
9:15 am	Datathon Teamwork
11:30 am	**Checkpoint 3** Team consults with guest instructors and mentors. At this point, you should have your dataset cleaned and analysis started. Be ready to discuss your initial findings and analysis.
12:00 am	Datathon Teamwork
12:30 pm	Lunch & Prayer Time
2:00 pm	**Checkpoint 4** Team consults with guest instructors and mentors. Be ready to discuss your presentation.
3:00 pm	**Team Presentations** Consists of 5-minute presentation of the team research question, analytical methods, findings, limitations, clinical implications.
4:00 pm	Judging
4:30 pm	Announcement of Winners Closing of Day 3

Instructors & Moderators

Lama H Nazer, PharmD, BCPS, FCCM

Lama Nazer is the Clinical Affairs Manager and critical care clinical pharmacist at King Hussein Cancer Center. She has extensive contributions in the field of healthcare AI, with emphasis on bias and equity in the development of AI technologies and machine learning algorithms. She serves on the steering committee for the Data Science group of the Society of Critical Care Medicine (SCCM) and was part of the task force that organized the first SCCM Datathon. She has been actively involved in local and international forums, conferences and workshops on digital health and AI, discussing various aspects of AI in healthcare and advocating for its ethical and responsible integration into clinical practice. More recently, she was a member of the working group for the international STANDING Together initiative that developed recommendations for tackling algorithmic bias and promoting transparency in health datasets.



An-Kwok Ian Wong, MD, PhD

Ian Wong is an Assistant Professor of Medicine at Duke University, specializing in Pulmonary, Allergy, and Critical Care Medicine. He also holds a secondary appointment in Biostatistics & Bioinformatics and focuses on using AI and data science to improve clinical decision-making.

His research focuses on developing equitable, interpretable predictive models and AI-driven tools that leverage electronic health records (EHR) to improve patient outcomes. He led efforts to validate and deploy large language models for clinical applications, including EHRmonize, a framework for standardizing medical concept extraction from EHRs. He is also well recognized for his work on racial and ethnic disparities in pulse oximetry which has influenced national policy discussions. Dr Wong developed and shared multiple open datasets and has organized national and regional Datathons, including as chair of the SCCM and Duke Datathons.

Piereandrea Morandini, M.S.

Pierandrea is a Senior Data Scientist with seven years of experience applying machine learning technologies to the healthcare domain. He has a master's degree from Politecnico di Milano in Italy and spent time as a visiting student at the Laboratory of Computational Physiology at MIT, he is currently working at Humanitas Research Hospital in Milan, where he specializes in developing Natural Language Processing tools to extract value from unstructured text data. In addition to his professional work, Pierandrea is part of the MIT Critical Data Alumni network, contributing to projects at the intersection of machine learning and healthcare. He has actively participated in Datathons and global initiatives, leveraging data-driven insights to address real-world challenges in medical research and clinical decision-making.

Adrien Carrel, MSc MEng

Adrien is a researcher specializing in machine learning and statistics, with a particular focus on time series analysis and its applications in healthcare. He holds dual degrees in Mathematics and Computer Science from Imperial College London and CentraleSupélec, where he developed strong expertise in advanced computational methods, statistical modeling, and data-driven problem solving. At the Massachusetts Institute of Technology (MIT), Adrien conducted research at the intersection of machine learning and public health, focusing on statistical methodologies to uncover disparities and improve decision-making in healthcare systems.

Naheel Said, PharmD, BCPS

Naheel is a critical care clinical pharmacist at the King Hussein Cancer Center in Amman, Jordan, where she provides specialized pharmaceutical care to critically ill patients within a multidisciplinary team. She earned her Doctor of Pharmacy degree from the University of Jordan, followed by the completion of a pharmacy residency at KHCC in May 2024. She is a board-certified pharmacotherapy specialist with a strong commitment to advancing evidence-based practice. Actively engaged in critical care at the international level, she contributes as a member of the Society of Critical Care Medicine. Her professional interests

include integration and application of artificial intelligence in healthcare, with ongoing research utilizing large-scale critical care databases such as MIMIC-IV and eICU. Her work focuses on clinical outcome analysis, predictive modeling, and exploring innovative data-driven strategies to improve patient care and healthcare delivery.

Chrystinne Fernandes, PhD

Chrystinne is a Senior Postdoctoral Researcher at MIT and Instructor at Harvard T.H. Chan School of Public Health, working on the development of the MIMIC database and the PhysioNet data-sharing platform for reproducible research in healthcare. She mentors students in collaborative medical data science and contributes to the organization of international Datathons. She holds a PhD in Computer Science from the Pontifical Catholic University of Rio de Janeiro (2019) and was a visiting PhD student at King's College London (2018). Her background includes over a decade of experience in e-health innovation, with expertise in Machine Learning, Data Science, Multi-agent Systems, and IoT.

Belel Lahham, B.Sc

Belal Lahham is a Data Scientist at the King Hussein Cancer Center. He develops advanced AI-driven solutions to streamline hospital operations, including automated clinical note data

extraction and real-time voice and chat systems powered by agentic AI architectures. One of his flagship projects, KHCC Brain, is an intelligent assistant designed to support medical staff in data analysis and academic writing. It combines technologies like Knowledge Graphs, advanced RAG, NL-to-SQL, deep research tools, chart generation, and medical imaging classification to boost research and streamline data workflows. He is currently building Patient Navigator, a smart system that extracts key information from patient notes and automates appointment synchronization and notifications for staff and patients.

Rajna Fani, B.Sc, M.Sc

Rajna Fani is a machine learning researcher working at the intersection of AI and healthcare. Currently completing her master's in computer science at Technical University of Munich, with thesis research at MIT's Laboratory for Computational Physiology, she develops robust AI models for clinical time-series and NLP applications. Rajna brings experience from roles at SAP, Rohde & Schwarz, and CGI, where she worked on full-stack ML systems, enterprise chatbots, and NLP pipelines. She actively mentors interdisciplinary teams at global healthcare datathons hosted by MIT Critical Data, including events at Duke, SCCM Chicago, and Hospital Clínic Barcelona.

Aya El-Mir, B.Sc

Aya El Mir is a researcher specializing in machine learning, with a strong interest in its applications in healthcare. She graduated from New York University Abu Dhabi (NYUAD) with a BSc in Computer Engineering. During her undergraduate studies, she conducted research across various institutions including the Technical University of Munich Medical School, NYU Abu Dhabi, and NYU New York. Aya applies machine learning across diverse domains, with a primary focus on healthcare. Her bachelor's thesis explored optimizing Multimodal Large Language Models (MLLMs) for use in low-resource healthcare settings.

These experiences led her to discover her passion for the field of ML for healthcare. She participated in the Society of Critical Care Medicine (SCCM) Datathon in NYC and collaborated with various members of the MIT Critical Care Data group to develop a tutorial for the Bumblekite ML Summer School at ETH Zurich. She is also passionate about outreach, having led a seminar in AI, healthcare, and ethics for high school students in Japan through the HLAB program.

Rafi Al Attrach, B.Sc, M.Sc

Rafi is a machine learning researcher working at the intersection of AI and healthcare. He completed his Bachelor's and Master's in Information Systems at TUM with thesis research at MIT's Laboratory for Computational Physiology, where he develops transformer models for clinical data applications and conversational AI systems. Rafi brings experience from roles at Amazon's AGI organization, Capgemini, and various research projects, where he worked on foundation models, enterprise systems, and healthcare AI solutions. He actively mentors interdisciplinary teams at global healthcare datathons hosted by MIT Critical Data, including recent events at Duke, Brown, Hospital Clínic Barcelona, and Zurich.

.

Guest Faculty

Ahmed AbuSalah, PhD

Dr. Ahemd AbuSalah is a Professor of Artificial Intelligence and Digital Health, and the Director of the Digital Innovation Hub at King Faisal Specialist Hospital & Research Centre (KFSHRC). He also serves as an adjunct professor of Informatics at both the University of Minnesota and Alfaisal University. Previously, he was a core faculty member in Biomedical Informatics and the Founding Director of the Clinical Informatics Core at the University of Minnesota. He also served as the Chief Informaticist at the U.S. National Center for Interprofessional Practice and Education (NCIPE). Prof. AbuSalah brings extensive experience across hospitals, universities, academic medical centers, and Fortune 5 HealthTech companies in both the U.S. and Saudi Arabia. His work focuses on advancing care excellence, driving operational sustainability, and fostering organizational thought leadership. He holds an Executive Certificate in Corporate Innovation Strategies from MIT, a Ph.D. in Biomedical Informatics from the University of Minnesota, dual master's degrees in Computer Science from Purdue University and the University of Illinois, and dual bachelor's degrees in Engineering and Information Systems from Jordan University of Science & Technology and Philadelphia University.

Siddiq Anwar, MBBS

Dr. Siddiq Anwar is a Consultant Nephrologist and Clinical Professor of Medicine at Sheikh Shakhbout Medical City in Abu Dhabi, recognized as one of the region's leading nephrologists in the region. He holds a Visiting Fellowship at the Abu Dhabi Investment Authority Lab, focusing on applying AI to improve patient outcomes. Dr. Anwar has spearheaded high-impact initiatives including the digital transformation of renal care services, establishment of the UAE's Paired Kidney Donation program, and the region's first global kidney exchanges. He co-founded RenAIssance, an award-winning AI research lab with Khalifa University and MBZUAI, while supporting capacity-building programs for renal care access in India and Africa. Beyond healthcare, he founded Raza Farm, a circular integrated farming initiative in Karnataka, India.

Mohammad Yaqub

Dr Yaqub is an associate professor at Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), Abu Dhabi, UAE. Dr. Yaqub leads the BioMedIA group at MBZUAI. He completed his PhD in Biomedical Engineering from the University of Oxford. His domain knowledge is Artificial Intelligence (AI), machine learning, and biomedical health & image analysis. He has published over 70 peer-reviewed articles, has 6 patents, co-edited 2 books, and secured many national and international grant funding. He is also an honorary research fellow at the Nuffield

Department of Clinical Neurosciences (NDCN) and the Acute Vascular Imaging Centre, University of Oxford. Dr. Yaqub was also a VP of Engineering at a startup and led a team to develop FDA-approved medical device products, which are currently used on millions of patients worldwide. Dr. Yaqub founded Labib AI, MBZUAI's first start-up based in Abu Dhabi.

Steve Bethke

Steve Bethke is the Vice President of the Solution Developer Market for Mayo Clinic Platform. He is responsible for developing and executing strategic plans and overall growth of the solution developer market including top line revenue, go-to-market, packaging and pricing for Mayo Clinic Platform's health tech accelerator program and solution developer program. Bethke manages a cross-functional team across product and solutions and works to ensure all efforts align with Mayo Clinic Platform's business objectives and vision.

Bethke brings deep experience in the healthcare technology business with a cross-disciplinary background in engineering, product development, and product lifecycle management. Before joining Mayo Clinic Platform, he held key executive roles at GE Healthcare and Optum as well as an executive managing VC and PE backed growth stage companies though successful growth and exits.

He earned a bachelor's degree in engineering from Boston University and an MBA from Pepperdine University.

What to Bring?

- Your Laptop will be necessary.
- Data will be made available in the cloud. We'll be using Colab Notebooks, and everything can be done in the cloud. If you prefer, you can also use your local machine.

What are the team distributions and expectations?

- There will be 5 teams of approximately 8-10 participants, which will consist of healthcare professionals and data scientists.
- The datasets that the teams will use are the e-ICU and MIMIC.
- At least one of the data science participants at each table will have had experience at a prior Datathon and substantial experience with the PhysioNet datasets.
- Everyone attending is interested in AI and data science. This is a unique situation where everyone has a pre-selected interest. All the clinical participants have an interest in data science and all the data science participants have an interest in healthcare!
- The clinical participants are expected to shape the meaning of the question (especially making sure that the problem is defined in a clinically relevant fashion), lend context to the clinical data, and shape the results.
- The data science participants should be ready to learn about the clinical question and the datasets, and to understand how to effectively extract the relevant data, wrangle data, and analyze the results.
- During the datathon, all team members are expected to actively contribute and learn from the experiences of others.
- Neither side is more important; the interplay of both is critical for effective analysis. We want participants to learn that this is team science and we really need to engage people with different perspectives and from different areas of expertise to tackle such complex questions.

Frequently Asked Questions (FAQ)

Link to videos

Please review the following videos PRIOR to the Datathon and complete all the outlined requirements.

- How to set up Physionet (MIMIC, eICU) for BigQuery
- How to Register for GitHub (and get a GitHub ID)
- How to analyze PhysioNet EHR data with Jupyter/BigQuery

Access to datasets

The current datasets can be accessed through the links below; please request access PRIOR to the Datathon.

- MIMIC-III
- MIMIC-IV
- eICU-CRD

Affiliated data sets are acceptable, although there are no increased compute resources (e.g., for chest x-ray or waveform data) guaranteed:

- MIMIC-CXR
- MIMIC-IV-ED
- MIMIC-IV Waveform

Derivative datasets are acceptable, e.g.,

- BOLD, a blood-gas and oximetry linked dataset
- Hourly MIMIC-IV data
- Curated Data for Describing Blood Glucose Management in the Intensive Care Unit

All teams must have access to at least the following datasets: MIMIC-III, MIMIC-IV, and eICU-CRD.

How do I sign up for MIMIC-III, MIMIC-IV, eICU?

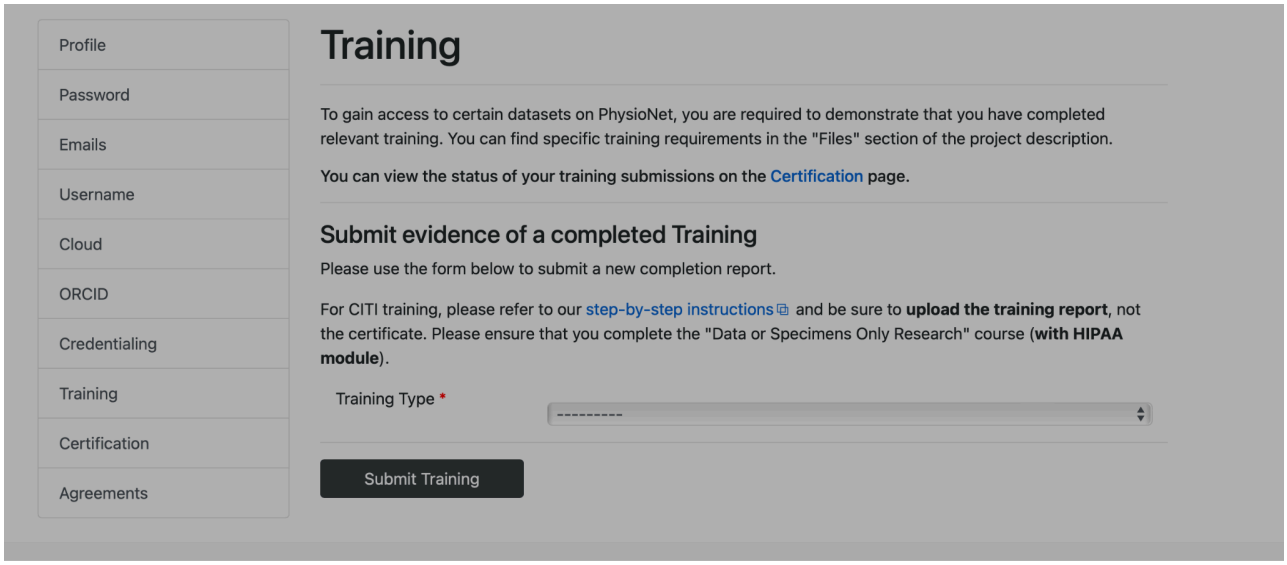
Great question!
We've created a video for you. How to set up Physionet (MIMIC, eICU) for BigQuery

I can't figure out how to sign up for a Code of Conduct on PhysioNet.

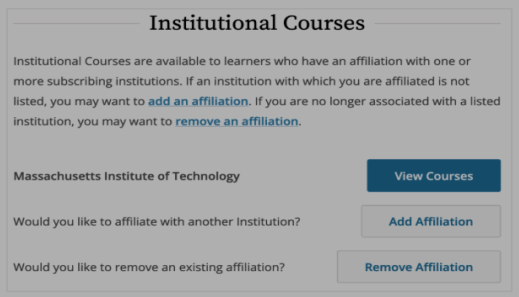
You don't need it. Here is a screenshot of Ian's PhysioNet DUAs and Code of Conduct page as an example. Lots of DUAs. No code of conduct form seems to be needed.

I don't have an MIT email address, can I still do the CITI under MIT affiliates?

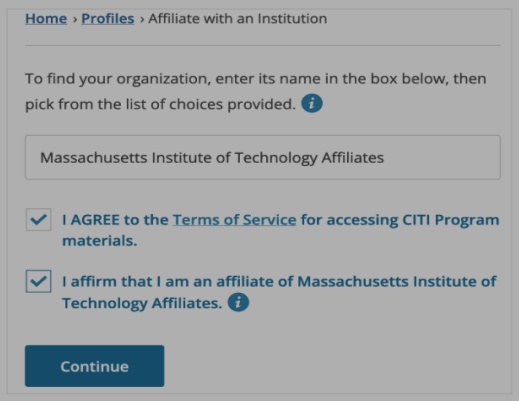
Yes! It asks for a preferred *institutional* account (e.g., Duke.EDU, unc.EDU, etc.), but your personal account could work if you have no other options.



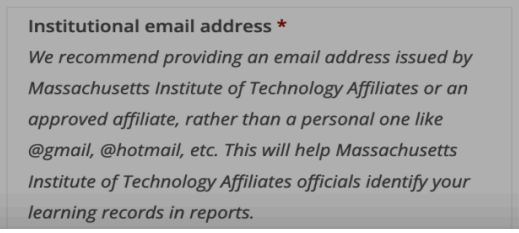
1. Once you have created an account, go to 'My Courses' on the top of the page. Click "Add affiliation"



2. Search for "Massachusetts Institute of Technology Affiliates". Click the checkmarks to agree. You are affiliating with MIT for the purpose of accessing data hosted on MIT servers.



3. Fill out the form. Please use your institutional address, not a personal address, where possible (note, you do not require an MIT address).



Can I bring my own dataset?

- During the Datathon, we will only be working on the datasets stated above. This is to ensure that all team members can have access to all data for the event.

Access to GitHub

I don't have a GitHub ID

Please register for one at GitHub.com. We've created a video for you if you're facing any difficulties: How to Register for GitHub (and get a GitHub ID)

Datathon operations

Can my team change what topic we research?

Yes! The initial topic provided is a *suggestion*. The topic is not guaranteed to be 100% novel; it may have been published before in some fashion. The team's feelings may change based upon the data accessed - perhaps the data doesn't support that analysis. That's all okay. Refining the question and/or pivoting is all part of the process.

I'm clinical, and have never done this before. Can I join?

Yes! Your interest, but not experience, is required. What you *are* required to do, however, is register for all the datasets and run the tutorial videos. You're expected to get elbow-deep in all the data and code as best you can with everyone else on the team. While you're not going to be an expert-level coder or data scientist overnight, this datathon is intended to reduce the activation energy - show you that EHR data science isn't so scary after all.

Everybody codes. Nobody quits.

Can I pick what team I'm in?

No, teams will be assigned. The intent is to build *new* connections and ensure a diversity of experience within a team, so you are likely to be split into a team from different divisions/institutions/companies.

Can I change teams at the datathon?

No, teams are fixed.

Judging criteria

How will we be judged?

Judging has two parts:

- Peer judging: an Eurovision-style judging, where each team judges every other team.
- Formal judges: will judge your eligibility for categories

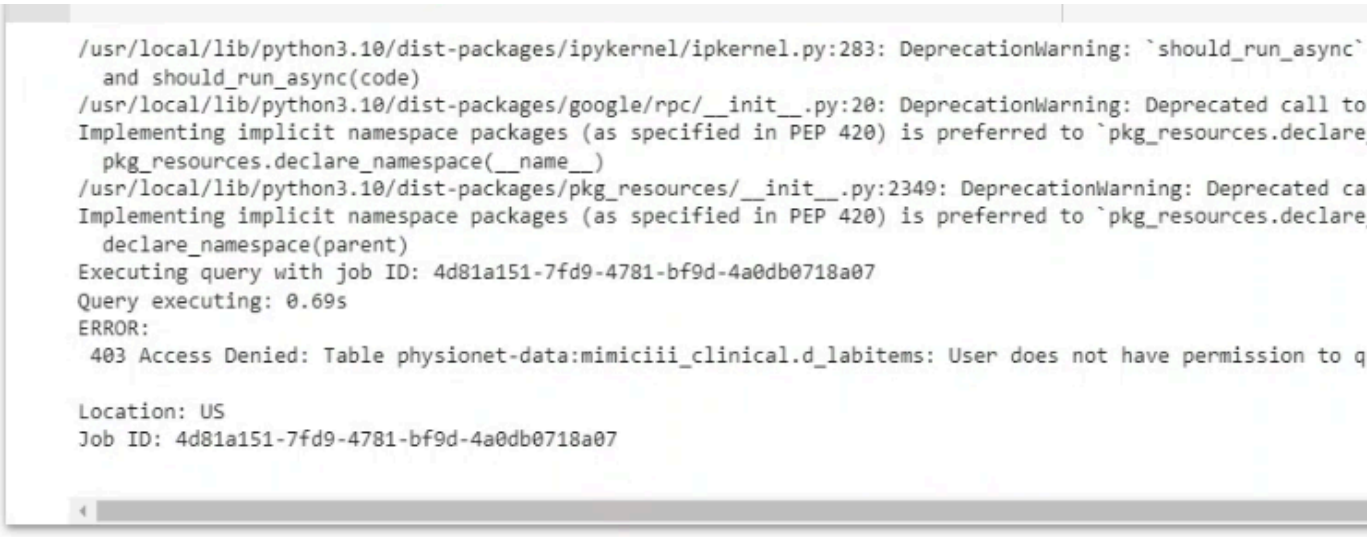
Running the example Colab notebook

I can't run the cells

You probably received an error because you'll need to run the 'defines' cell first.
Consider following this instructional video. How to analyze PhysioNet EHR data with Jupyter/BigQuery

The Google Colab doesn't have access to MIMIC / eICU

You probably received an error like this.



Be sure you:

- Logged into the right Gmail account (it's easy to have a different one selected)
- Followed the How to set up Physionet (MIMIC, eICU) for BigQuery

What is the presentation format?

- Each team will get up to 5 minutes to present their project.
- The presentation should address the following:
    - The clinical question (<45s)
    - Team name / team members
    - Clinical motivation/background/ why is this question important to answer
    - Methods / approach which should address the datasets used, how the data was handled and what analytical analysis was used.
    - Table 1 Patient characteristics/ Flow diagram
    - Results/ Findings
    - Biases / Data problems
    - GitHub code link

The judging will take into account the following major points:

- Originality and innovation: how unique and creative is the team's approach to the problem?
- Relevance to clinical practice and impact: how can the team's solution influence or improve
- Appropriate method selection: did the team choose the most suitable method for their project, regardless of complexity?
- Team's plan for moving forward with the project: how well did the team articulate their future plans for the project beyond the Datathon?
- The contribution to knowledge about Health Equity?