



WiDS Datathon 2024  
#WiDSDatathon

Mervat Abuelkheir  
Data Engineering Course



## WiDS Long-term Vision

We envision a future in which women are fully integrated and represented in all areas of Data Science, and share equally in Decision Making, Economic Prosperity, and Opportunities.



The background of the slide features a large yellow circle on the right side. On the left, there is a collage of images: a woman with long dark hair in a purple top, a woman with long brown hair smiling, a close-up of a blue eye, and a woman with short blonde hair and glasses. There are also solid green and blue rectangular blocks interspersed among the photos.

## WiDS Mission

To change the field of data science across the globe by **elevating and empowering women** and helping the community reach 30% representation of women in education, innovation, and leadership by 2030.

# 30x30: Inspire, Enable, Collaborate and Innovate

## 30% by 2030

### Education

Undergraduate and graduate students in data science and related areas

### Innovation

Influential innovators in industry, academia, NGOs and government

### Leadership

Data science leaders, managers and decision-makers in industry, academia, NGOs and government

Today

Education



30% in '30

Innovation



30% in '30

Leadership



30% in '30





Predicting Deforestation by Identifying Oil Palm Plantation



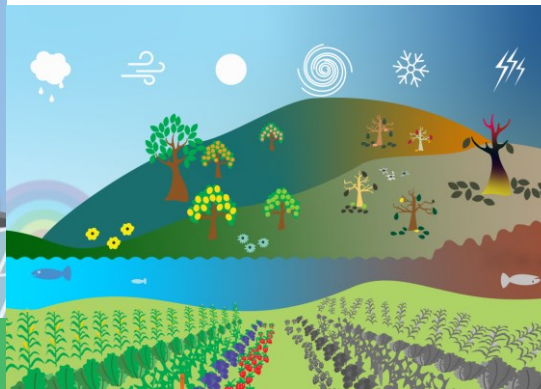
Predicting Patient Survival in Intensive Care Units (ICU)'s



Identifying Diabetes Condition to Provide Better Care for ICU Patients



Using Data Science to Mitigate Climate Change



Adapting to Climate Change by Improving Extreme Weather Forecasts

## WiDS Datathon Goals:

**Lower barriers** to entry into data science

**Foster learning and enhanced knowledge** of data science skills, applications, and career growth for students

Build **connections and community**

*Always focused on a social*

*challenge!*

# WiDS Datathon Impact



Women in  
Data Science  
Worldwide

**78%** **Female**  


**51%** **Beginners**  


**62%** **No Prior  
Experience**  


**73%** **New Skills**  


Join us, and participate in the annual WiDS Datathon  
[widsworldwide.org/learn/datathon](https://widsworldwide.org/learn/datathon)



## 2024 Challenge

### Topic:

Equity in Healthcare  
sponsored by Gilead  
Sciences.

Participants will  
analyze a unique  
oncology dataset to  
identify possible  
inequities in  
patient care.

## The WiDS Datathon – What is it?

**What:** Global predictive analytics competition for applying data science skill to solve a critical social impact challenge

**How:** WiDS provides well-curated, real-world datasets and crafts a data science challenge that is hosted on the Kaggle platform.

**Who:** Targeting participants at Beginners and Advanced levels

**Why:** Designed to reduce barriers to learning data science, in a supportive/ collaborative environment



### RESOURCES PROVIDED:

Webinars and Office Hours



Community (Slack)



Global Workshops



Guides and Tutorials



# 2024 Challenge Topic: Equity in Healthcare

**Gilead Sciences** is the sponsor of this year's datathon.

They provided a rich, real-world oncology dataset with information about demographics, diagnosis and treatment options, and insurance provided about women patients who were diagnosed with breast cancer.

Students will analyze the dataset to identify possible inequities in patient care, such as treatment duration.

Women in Data Science Worldwide





# 2024 Challenges:

## WiDS Datathon++

September 2023 - June 2024

*(advanced)*

## WiDS Datathon #1

January - March 2024

*(beginner)*

## WiDS Datathon #2

April - June 2024

*(beginner)*

Women in Data Science Worldwide





# WiDS Datathon++ Challenge Task

Predict time to treatment of women patients who are diagnosed with metastatic triple negative breast cancers (metastatic TNBC).

Metastatic TNBC is considered the most aggressive TNBC, and women who are diagnosed with it are among those who are in the most urgent need for timely treatment.

Students will develop a model to predict how many days it takes for a patient to receive the first treatment for their cancer diagnosis based on patients' characteristics.

**Data now available on Kaggle for download:**

<https://www.kaggle.com/t/09a124705ae44da2ab0912e858639645>

# Steps to begin the challenges



1. Sign up

2. Create Your  
Kaggle Account

3. Form a Team

4. Start the  
Challenge!



## WiDS Datathon 2024 Registration

Sign up to participate in this year's WiDS Datathon 2024! In this year's datathon challenge participants will analyze a unique oncology dataset to identify possible inequities in patient care, such as treatment duration and time to adoption of treatment.

Completing this form confirms your registration with the WiDS Worldwide team for the WiDS Datathon. You only need to complete registration once, despite which datathons you choose to compete in. All participants must register to participate in the challenge.

After submitting this form you'll receive a confirmation email with next steps including skill-building resources, and ways that you can connect with global participants.

More information: [widsworldwide.org/learn/datathon](https://widsworldwide.org/learn/datathon)  
Questions? [datathon@widsworldwide.org](mailto:datathon@widsworldwide.org)

First Name \*

Last Name \*

Email \*

By providing your email address, you are consenting to receive communications about the WiDS Datathon, including datathon updates.

## Sign up here to begin:

[https://bit.ly/datathon2024\\_registration](https://bit.ly/datathon2024_registration)

n

Registration is required for all WiDS Datathon participants. You only need to complete registration once, despite which datathons you choose to compete in.

Follow these steps for your participation to be counted towards the **WiDS Datathon University Program**, and to be eligible for University recognition:

What kind of organization do you work for? \*  
Academia / Student

Select Academia/Student

What is your role in Academia? \*  
Undergraduate Student

Select your most appropriate role from the list

WiDS Datathon University Program

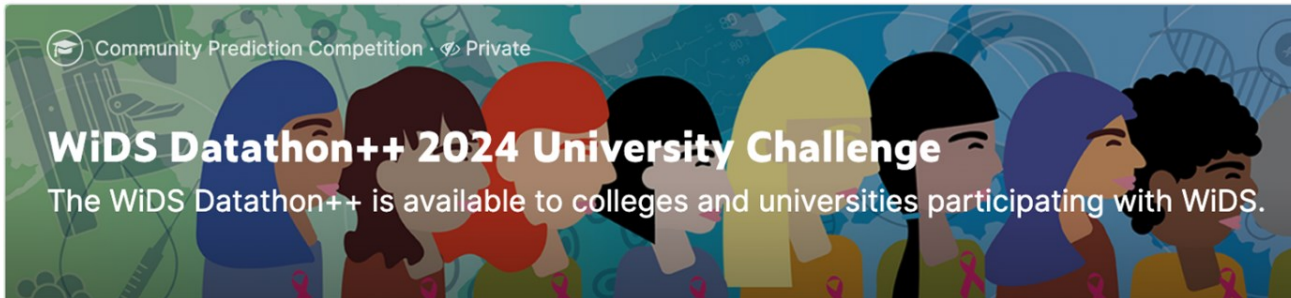
Are you participating in the datathon directly with your college or university via the WiDS Datathon University Program? If so please check the box.

☒ <— Be sure to check this box

Please select your college or university. \*  
+ Add

Select your university from drop-down list





Host [Overview](#) Data Code Discussion Leaderboard Rules Team

## Overview

Join the WiDS Datathon++ University Challenge to hone your data science skills while solving an interesting and critical social impact challenge! Utilizing a unique oncology dataset, you will develop a model to predict time to treatment of women patients who are diagnosed with metastatic triple negative breast cancer.

### Start

Set start date via the [launch checklist](#).



## Description

### Background: Equity in Healthcare

Women in Data Science (WiDS) Worldwide is on a mission to increase participation of women in data science to benefit societies worldwide. We need to support the advancement of women's healthcare w

Addressing healthcare inequity is critical and has an extensive positive impact on health, social, and economic economy. Women often encounter distinct health challenges and stand to gain significantly from efforts to ensure equity in access to healthcare. Achieving health equity involves addressing disparate treatments based on demographic and societal factors so that in the future women can achieve their highest level of health and well-being.



# Hosted on Kaggle

Kaggle is one of the most popular machine learning platforms in the world. Users can share code, data, and ideas, and collaborate on projects.

Check it out: [kaggle.com](https://www.kaggle.com)

# Resources and Tutorials



Kaggle Tutorials:

- [Introduction to Kaggle Competitions](#)
- [Video: Getting Started with Kaggle](#)

Skill-Building Tutorials (challenge preparation materials):

- [Intro to Jupyter and Data Structures](#)
- [Beginners Data Science Modules](#)

**\*NEW!\*** WiDS Datathon Starter Code:

- [WiDS Datathon++ challenge](#)

# Team Guidelines



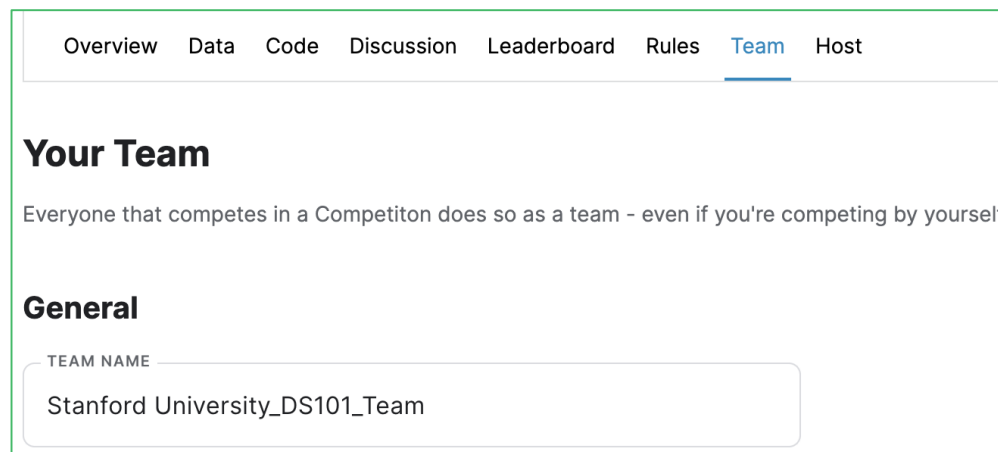
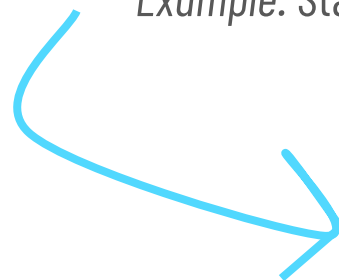
*The WiDS Datathon++ is open to individuals or teams of up to 4; at least half of each team must be individuals who identify as women.*

## Team Formation naming convention on Kaggle:

Please name your Kaggle team with the following formula to easily track submissions and results:

*Kaggle team name:* University Name\_Course Name\_Team Name

*Example:* Stanford University\_DS101\_Team WiDS'ers



**#university\_students** Slack  
channel available for  
collaboration!





**Discussion  
Questions for  
WiDS  
Datathon++**

# WiDS Datathon++ Discussion Question #1



What are the social, economic, and geographic factors that show the strongest correlation with time to receive treatments of patients?

(e.g., whether it is race/ethnicity, or locations, or incomes or something else)

# WiDS Datathon++

## Discussion Question #2



Can you develop some graphs/dashboards that show the analytics of those factors? What stories can you tell with these data visualizations?



# WiDS Datathon++ Discussion Question #3



Are there interactions between the variables and/or interesting clusters? What might be the implications of these patterns on access to healthcare?

# WiDS Datathon++

## Discussion Question #4



What are different approaches to address NAs in the data? What are the pros and cons of these approaches?

# WiDS Datathon++

## Discussion Question #4



What additional information can you gain by performing NLP on different columns of the data? What can you observe about the right or left or quadrants of the breast?