Women in Data Science Worldwide

WiDS Datathon 2025
#WiDSDatathon

Kylie Cancilla
Caterina Ponti
Liana Mendoza

Workshop Goals

- Challenge Task Overview
- Timeline and general outline
- Kaggle page
- Categorical variable description
- Functional Connectome Matrices

Women in Data Science Worldwide

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.

**Women in
Data Science
Worldwide**

WiDS Mission

To change the field of data science across the globe by **elevating**, **educating**, and **empowering** women to achieve 30% representation of women in data science by 2030.

Datathon 2025 Goals

- Lowering barriers to entry into data science

- Encourage students to hone their data science skills, in subjects such as: data wrangling, processing and preparing for machine learning along with fitting and testing a machine learning model

- We hope by the end of this challenge you feel more confident to enter the field of data science

- Provide women a sense of belonging

WiDS Datathon Impact

Women in Data Science Worldwide

78% **Female**

versus a typical Kaggle competition with less than 20% female

51% **Beginners**

characterized themselves as beginners, 11% were advanced /experts in the field

62% **No Prior Experience**

have not previously participated in a machine learning or predictive analytics competition

73% **New Skills**

respondents said they gained new skills in data science

Join us, and participate in the annual WiDS Datathon
widsworldwide.org/learn/datathon

WiDS Datathon Challenge Overview

The WiDS Datathon was developed with Ann S. Bowers Women's Brain Health Initiative in collaboration with Cornell University and UC Santa Barbara. Datasets and support are provided by the Healthy Brain Network (HBN), the signature scientific initiative of the Child Mind Institute, and the Reproducible Brain Charts project (RBC).

Neuropsychiatric disorders that occur in development, like anxiety, depression, autism, and attention deficit hyperactivity disorder, or ADHD, often differ in how and to what extent they affect males and females. ADHD occurs in about 11% of adolescents, with around 14% of boys and 8% of girls having a diagnosis. **There is some evidence that girls with ADHD can often go undiagnosed**, as they tend to have more inattentive symptoms which are harder to detect. Girls with ADHD who are undiagnosed will continue suffering from symptoms that burden their mental health and capacity to function.

WiDS Datathon Challenge Task

This datathon challenge aims to answer this question:

What brain activity patterns are associated with ADHD; are they different between males and females, and, if so, how?

To work towards the answer to this question, participants will be tasked with building a **multi outcome model to predict both an individual's sex and their ADHD diagnosis** using functional brain imaging data of adolescents and their socio-demographic, emotions, and parenting information.

A multi-outcome model is designed to predict multiple target variables simultaneously using a single machine learning model.

(female:1, male:0, ADHD: 1 ; no ADHD 0)

Submit Prediction ···

WiDS Datathon 2025

Unraveling the Mysteries of the Female Brain: Sex Patterns in ADHD

Overview   Data   Code   Models   Discussion   Leaderboard   Rules   Team   Submissions   Settings

Overview

Join the WiDS Datathon to hone your data science skills while solving an interesting and critical social impact challenge! Using fMRI data, you will be tasked with building a model to predict both an individual's sex and their ADHD diagnosis.

Competition Host
widsdatathon

Prizes & Awards
Kudos Edit
Does not award Points or Medals

Start
an hour ago

Description

Background / Context:

The WiDS Datathon Global Challenge was developed with Ann S. Bowers Women's Brain Health Initiative (WB... collaboration with Cornell University and UC Santa Barbara. Datasets and support are provided by the Healthy... Network (HBN), the signature scientific initiative of the Child Mind Institute, and the Reproducible Brain Charts... (RBC).

Neuropsychiatric disorders that occur in development, like anxiety, depression, autism, and attention deficit h... disorder, or ADHD, often differ in how and to what extent they affect males and females. ADHD occurs in abou... adolescents, with around 14% of boys and 8% of girls having a diagnosis. There is some evidence that girls w... can often go undiagnosed, as they tend to have more inattentive symptoms which are harder to detect. Girls with ADHD who are undiagnosed will continue suffering with symptoms that burden their mental health and capacity to function.

For more domain context see the following videos: The Role of AI in Advancing Women's Brain Health Research and Sex-specific differences in the healthy and disordered brain

Tutorials & Resources

Team Formation

Awards & Recognition

Women in Data Science Worldwide

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

# Challenge Timeline

Women in Data Science Worldwide
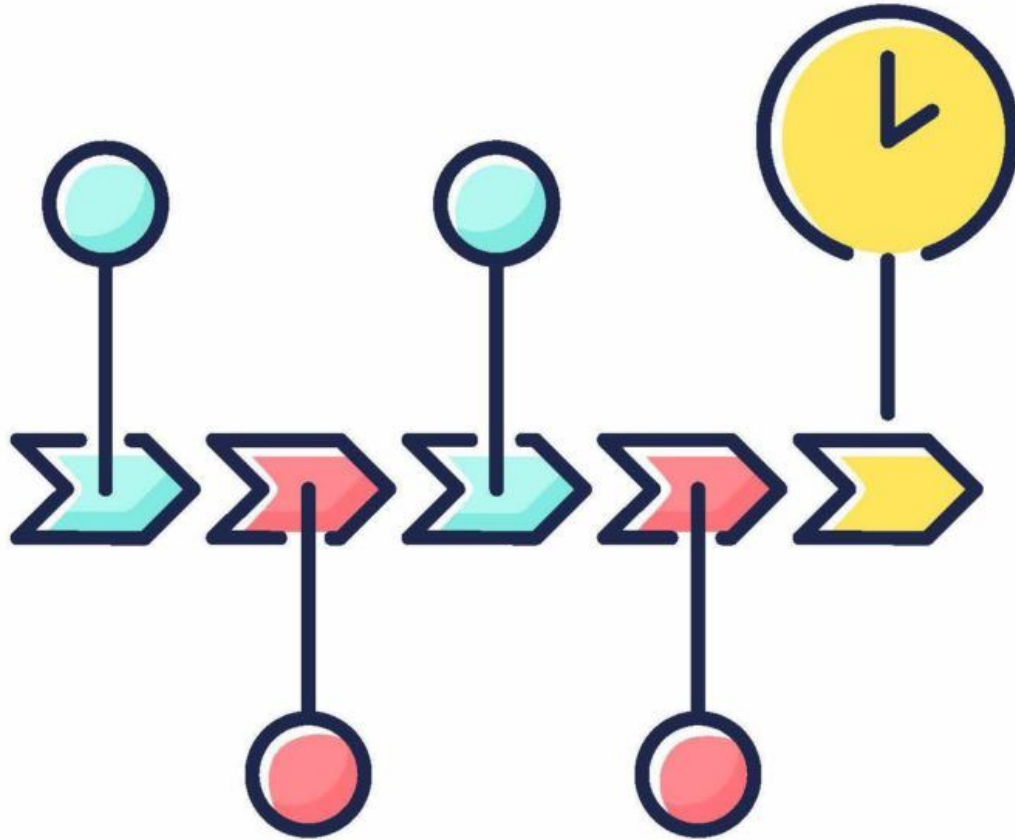
- **Challenge launches January 7th!**
- **You have until April 27th to submit your teams**
- **You have until April 30th to submit your results on kaggle**

The Dataset

- **Training folder** (1,200+ subjects) :
  - **Targets:**
    - ADHD Diagnosis
    - Sex
  - **Functional MRI Connectome Matrices**
  - **Socio-demographic Information:**
    - Quantitative Metadata
    - Categorical Metadata

- **Test folder** (300+ subjects):
  - **Functional MRI Connectome Matrices**
  - **Socio-Demographic Information**
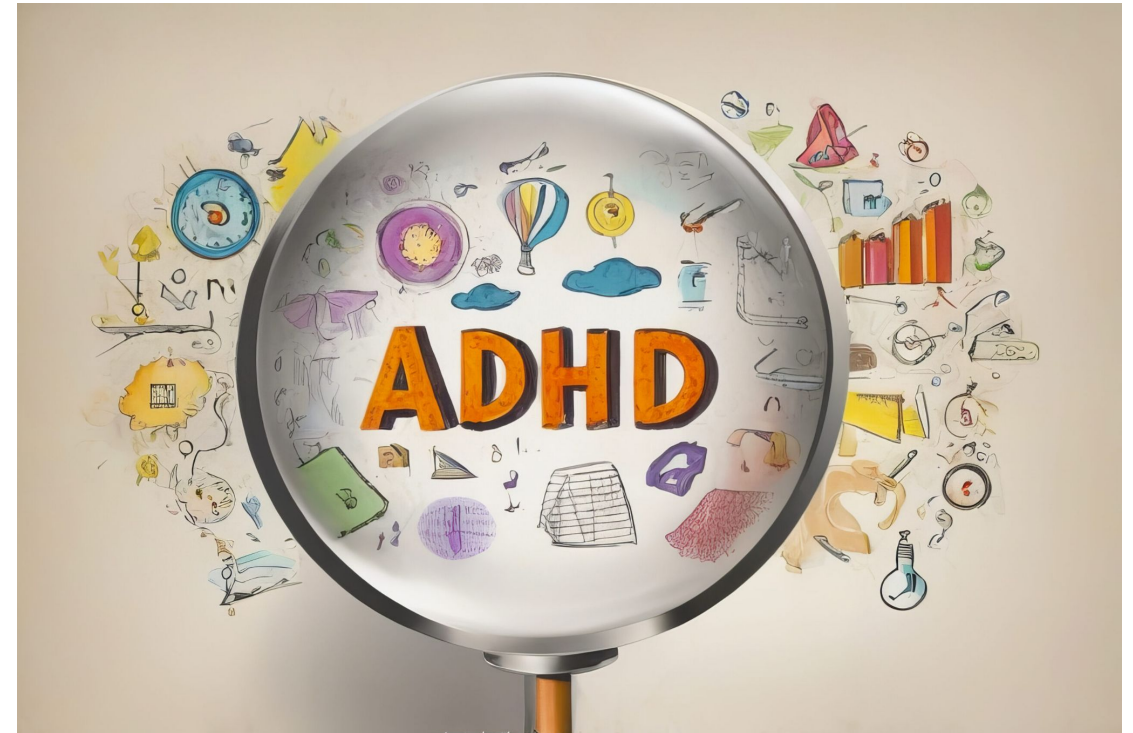
Functional MRI Connectome Matrices

What is fMRI?

- Resting-state Functional Magnetic Resonance Imaging (fMRI)
- Measures brain activity using **Blood Oxygen Level-Dependent** (BOLD) contrast.
  - Active brain regions consume more oxygen → blood flow increases to those regions → fMRI detects changes in blood oxygen levels to infer neural activity.

What is Functional Connectivity?

- Describes interactions between different brain regions.
- Originally represented as a **Matrix**, where rows and columns represent brain regions and the cells the **correlation** between activity in paired regions.
- The matrix was **transformed into a dataset**, with each column representing the interaction between two brain regions (e.g. Region 0 ↔ Region 1)

What do the interactions mean?

- **Positive values**: Brain regions are <u>more</u> active together
- **Negative values**: Brain regions are <u>less</u> active together
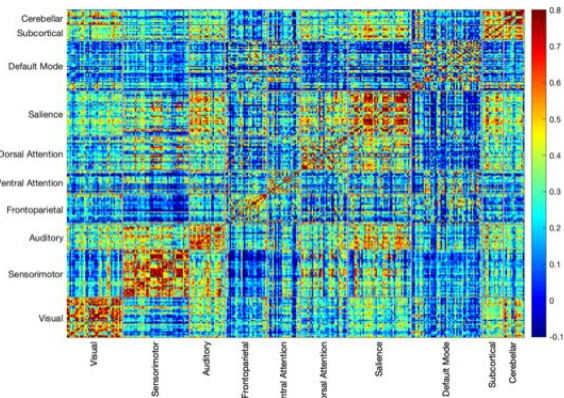
participant_id	0throw_1thcolumn	0throw_2thcolumn	0throw_3thcolumn	0throw_4thcolumn
70z8Q2xdTXM3	0.0934730255256336	0.1469023438487970	0.0678925633843087	0.0151411151994212
WHWymJu6zNZi	0.029579748150484	0.1793233783007680	0.1129332102999940	0.0382909950599437
4PAQp1M6EyAo	-0.0515796811041808	0.1397339475203460	0.0682950218190208	0.046990703137188
obEacy4Of68l	0.0162734004730087	0.2047015410840510	0.1159800052901390	0.043102511558928
s7WzzDcmDOhF	0.0657705599446428	0.0987136527971691	0.0976039908636495	0.1129884882306060

Metadata and Descriptions of Variables

Healthy Brain Network (HBN) is the community-based initiative of the Child Mind Institute. HBN openly shares phenotypic and neuroimaging data to study mental health and learning disorders in the developing brain.

- There are qualitative (categorical) and quantitative data types
- Use data as predictor of ADHD diagnosis and sex

Demographics

- Participants include both females and males, aged 5-22 years
- Race, ethnicity, handedness, year of participation, color vision score

MRI information

- Location and age of participant during the MRI scan process

Social Status

- **Barratt Simplified Measure of Social Status**
- Parents' levels of education and occupation

Metadata and Descriptions of Variables

Parenting

- **Alabama Parenting Questionnaire**
- Domains of parenting relevant to etiology and treatment of child externalizing problems
- Positive involvement, supervision/monitoring, use of positive discipline, consistency in the use of such discipline, use of corporal punishment

Emotions

- **Strengths and Difficulties Questionnaire**
- Measures emotional and behavioral well-being of participants
- Asks the caregiver if the participant has experienced a problem
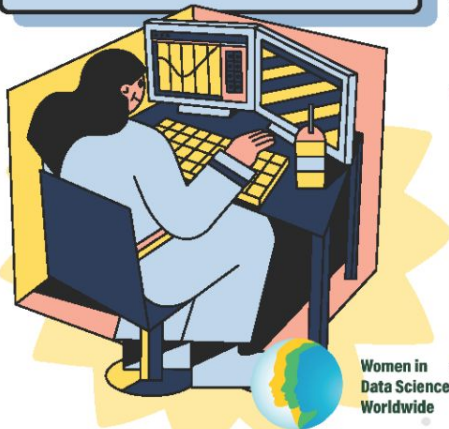- Problem's degree of chronicity, distress, social impairment, and burden to others

HBN DATA
WiDS Datathon Global Challenge 2025

The **Healthy Brain Network (HBN)** is the **community-based research initiative** of the **Child Mind Institute.** HBN openly shares phenotypic and neuroimaging data to study **mental health and learning disorders** in the developing brain.

Child Mind Institute
Healthy Brain Network

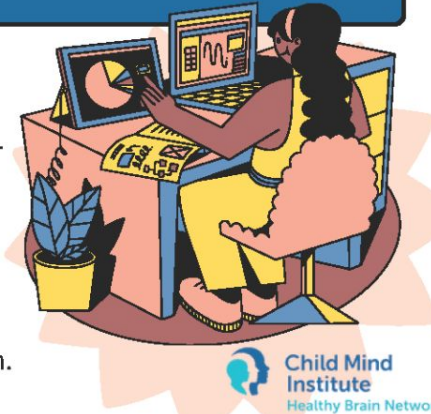**Women in Data Science Worldwide**

MRI
HBN collects **Magnetic Resonance Imaging (MRI)** data, including structural MRI, functional MRI (fMRI), and diffusion tensor MRI (DTI).

FUNCTIONAL CONNECTIVITY
As part of the RBC initiative, functional connectivity matrices are created from **resting-state fMRI data,** by correlating brain activity across different regions.

PHENOTYPES
HBN collects phenotypic data from >**140 assessments** across several domains. Check out the **comprehensive list of HBN assessments**!

DEMOGRAPHICS
HBN participants include both **females** and **males**, aged **5-22 years.** Participant's **race, ethnicity, handedness, year of participation,** and **color vision score** are provided.

EMOTIONS
The **Strength and Difficulties Questionnaire** measures **emotional and behavioral well-being** of children and adolescents across core areas.

PARENTING
The **Alabama Parenting Questionnaire** measures different **domains of parenting** that are relevant to the etiology and treatment of child externalizing problems.

SOCIAL STATUS
The **Barratt Simplified Measure of Social Status** is a proxy for socio-economic status. It lists parents' levels of **education** and **occupations.**

DIAGNOSES
HBN provides **no-cost diagnostic evaluations** to children and adolescents. Each participant is provided with up to 10 diagnoses. Check out the **HBN diagnostic process**!

ADHD
Attention-deficit/hyperactivity disorder (ADHD) is defined by **inattention, and/or hyperactivity/impulsivity** symptoms that can impact functioning or development.

TYPES
The DSM-5 classifies ADHD into three types based on on the predominant symptoms: **Inattentive Type, Hyperactive/Impulsive Type,** and **Combined Type.**

CERTAINTY
As part of the HBN protocol, clinicians assign a **level of certainty** to diagnoses: **Confirmed, Presumptive, Requires Confirmation, Rule-Out,** and **By-History.**
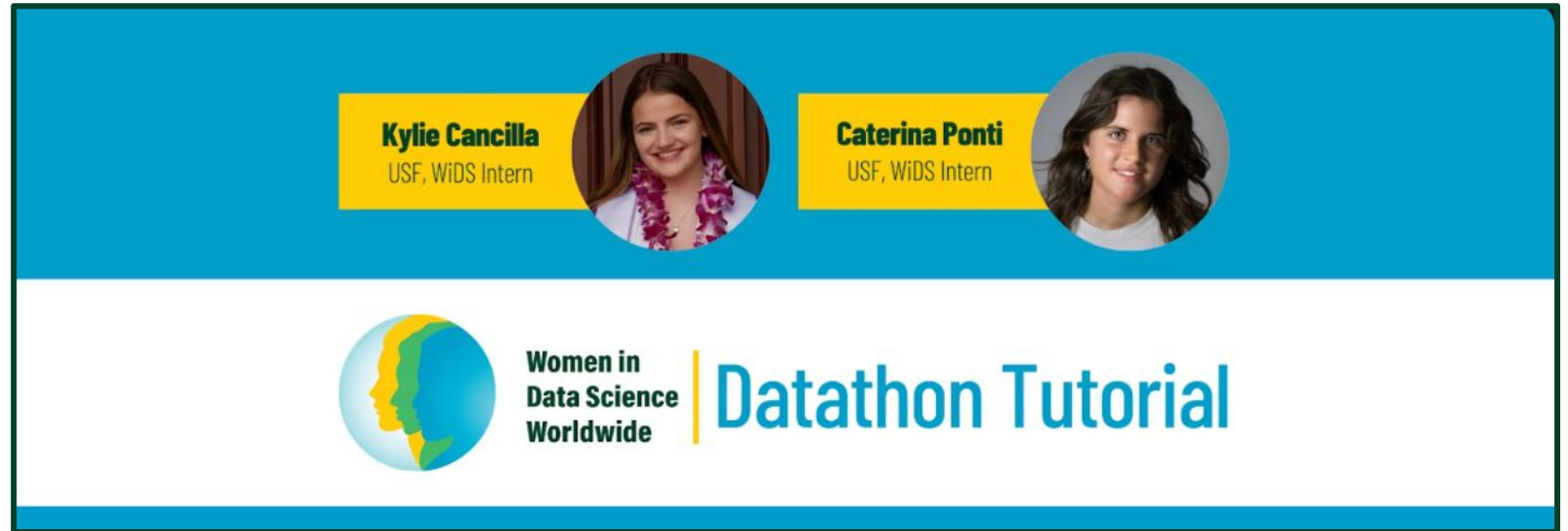
Child Mind Institute 2024

Save the Date: January 22, 2025
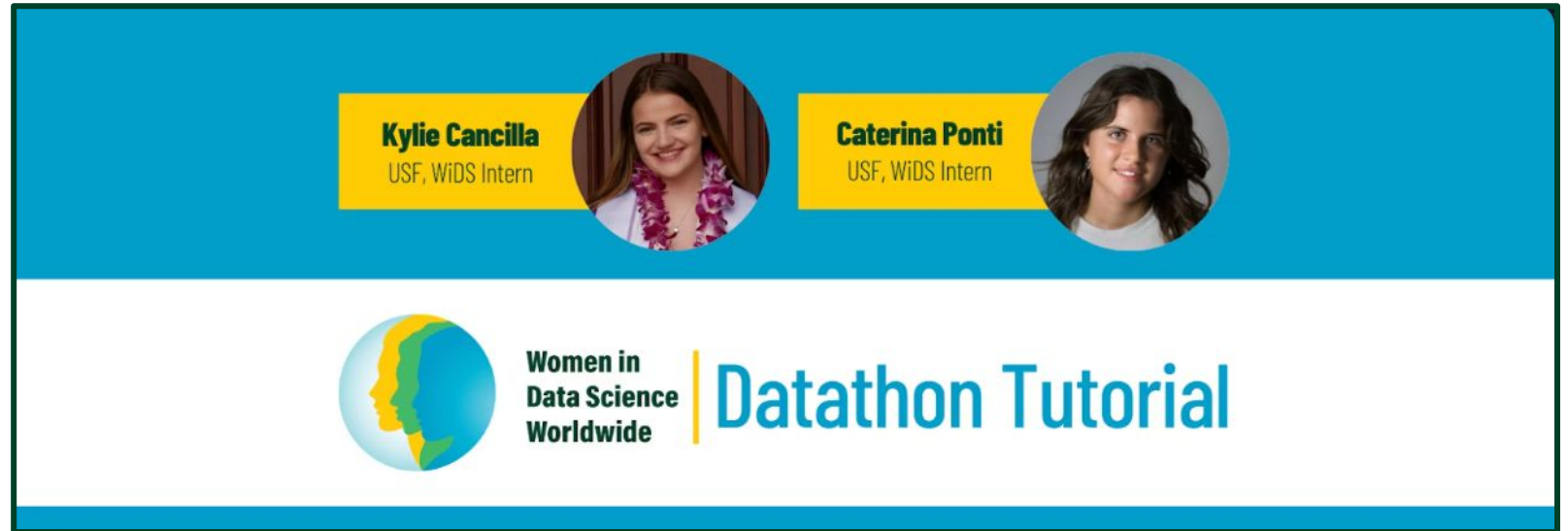
Workshop Topic: Preprocessing and Preparation of the Dataset

A. Encoding categorical variables

B. Merging Metadata and Functional Connectivity Matrix on indices

C. NaNs values

Save the Date: February 5, 2025

Workshop Topic: Building and Evaluating a Machine Learning Model

A. Building the Multi-Output Model

B. Accuracy score

C. Explaining F1-score as the metric for model evaluation

THANK YOU!

Questions?

Women in Data Science Worldwide