gTech
<professional services/>

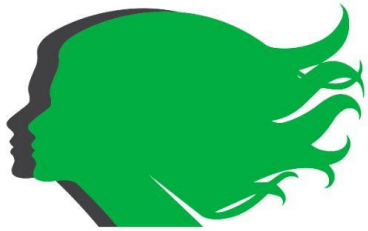
WOMEN IN DATA SCIENCE
@ GOOGLE

kaggle

WiDS@Google Datathon Workshop

WOMEN IN DATA SCIENCE

@ GOOGLE



WOMEN IN **DATA SCIENCE**
@ GOOGLE

Welcome to WiDS @Google

Yuka Abe

Sumedha Menon

WOMEN IN DATA SCIENCE

- A **conference** with 150+ regional events worldwide in more than 60 countries, reaching 100,000 participants annually.
- A **datathon**, encouraging participants to hone their skills using a social impact challenge.
- A **podcast** series, featuring data science leaders from around the world talking about their work, their journeys, and lessons learned along the way.
- An **education outreach** program to encourage secondary school students to consider careers in data science, artificial intelligence (AI), and related fields.
- A **workshop** series to build your data science skills, inspiring women and girls with role model instructors.

WiDS@Google will be spread across 2 days

Today **Technical focus**

- ✓ Tutorials and Hands-on practice
- ✓ Key ML concepts
- ✓ Python code demos
- ✓ Setting you up for the Datathon challenge
- ✓ Mentoring

Feb 23rd **Career focus**

- ✓ Panel discussion with Data Scientists
- ✓ Career Opportunities at Google
- ✓ Networking

Agenda for today

- **9:20 -10:05 EST / 14:20 - 15:05 GMT**

  Introduction to Machine Learning
- **10:15 - 11:05 EST / 15:15 - 16:15 GMT**

  Python code walkthrough for a regression problem
- **11:15 - 11:30 EST / 16:15 - 16:30 GMT**

  Datathon challenge introduction & team formation for hands-on exercise
- **11:30 - 13:30 EST / 16:30 - 18:30 GMT**

  Datathon Challenge hands-on exercise with mentor support

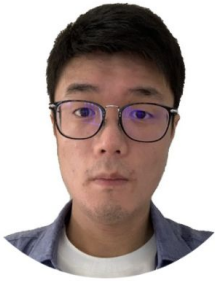
A BIG THANK YOU…

Christiane Ahlheim

Manisha Arora

Alex Chinien

Mike Anderson



Yan Sun

Tom Symonds

Zoe Zhang

A BIG THANK YOU…

Lyla Kuriyan

April Anderson

David Sneddon

Dave Franklin

Semih Oguz

Nikhil Madan

Dirk Nachbar

How much does a polar bear weigh?

Enough to **break the ice!**

Ice Breaker Sessions

- Where are you located?
- What do you do?
- How do you unwind after school/work?
- You favorite data science resources (courses/blogs/books...)
- What do you hope to take away from this event?
- Let’s get connected through email/LinkedIn, etc.🤝

Let’s get started!

Christiane Ahlheim

9:20am - 10:05am EST (2:20pm - 3:05pm GMT)
Introduction to Data Science & Machine Learning

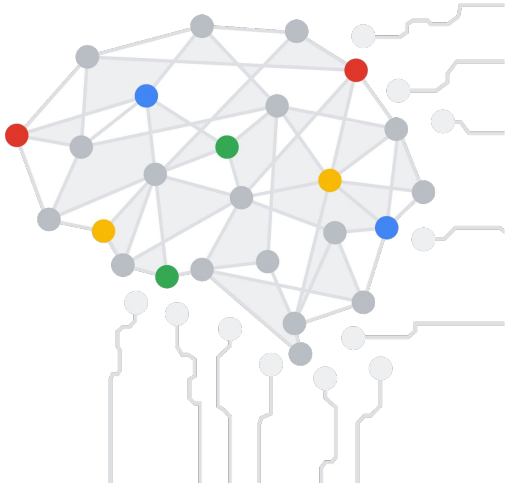
Christiane is a data scientist from gtech, helping Google’s advertisers from Europe to reach peak marketing performance.

One of her focus areas is privacy-first data science.

gTech
<professional services/>

WiDS Datathon

Hands-on exercise



Reminder - Supercharge your career at our follow up session

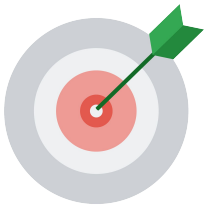
We have a **career focussed event** planned next **Wednesday (Feb 23rd)**

- **Data Scientists panel discussion**
  - DS from Engineering, Cloud, Creative Works, Professional Services
- **Career Opportunities at Google**
  - Hear our recruitment team shed light on career opportunities at Google and how you can apply
- **Networking sessions**

WiDS 2022 Datathon Challenge

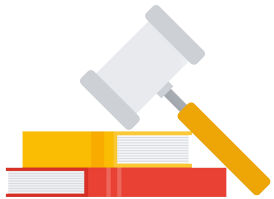
Mitigate the effect of the climate change with a focus on **energy efficiency**

- The goal of the challenge is to create models to predict energy building consumption
- Each row in the data corresponds to a single building observed in a given year
- **Features**:
  - Building characteristics e.g. floor area, facility type etc
  - Local climate and weather variables e.g. annual average temperature, annual total precipitation **etc**
- **Target** : Site Energy Usage Intensity (Site EUI)
- **Evaluation Criteria** : RMSE



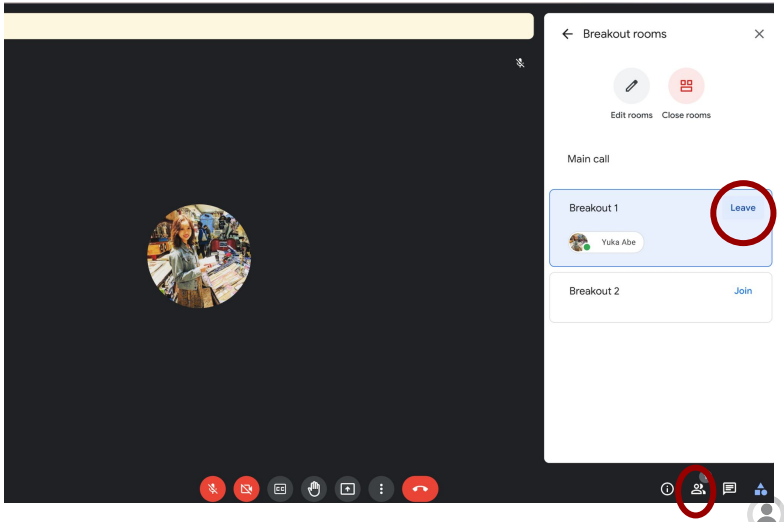
Important Datathon Rules

- The datathon is open to all individuals or teams of up to 4
  - At least half of each team must be women (individuals identifying as female participants)
- **Submission Limits**
  - You may submit a maximum of 15 entries per day.
  - You may select up to 2 final submissions for judging.
- **Competition Timeline**
  - Entry and Team Merger Deadline: February 23, 2022 11:59 PM UTC
  - End Date: February 26, 2022 11:59 PM UTC



Group Assignment & Hands-on exercise

- We will use breakout rooms to divide you into groups
- A dedicated mentor will come to your rooms to help you with team formation.



Hands-on exercise session is a great opportunity to connect with our mentors!

Bhaktipriya R.
Software Engineer Lead
Responsible ML Infra

Dirk Nachbar
Head of Applied Data Science
gTech

Elizabeth Lapo
Software Engineer
Youtube Knowledge Graph

JL Maréchaux
Marketing Science Lead
Google Marketing Platform

Kun Chang
Data Scientist
gTech

Lekshmi Santhosh
Product Analyst
ML Fleet Metrics

Megan Godfrey
Marketing Science Lead
Google Marketing Platform

gTech
<professional services/>

Hands-on exercise session is a great opportunity to connect with our mentors!

Mike Anderson
Data Scientist
gTech

Qin Cao
Machine Learning Engineer
Google AI

Sabina Przioda
Data Scientist
Data Science Trust & Safety

Sara Mourad
Software Engineer
Corp Eng Dev Infrastructure & ML

Simon Holgate
Technical Program Manager
Geo Data Operations MI

Zoe Zhang
Data Scientist
gTech

gTech
<professional services/>

Group Assignment & Hands-on exercise

Hands-on exercises resources:

- **2 python colab notebooks:**
  - 1 Exercise notebook: template_datathon_no_solution.ipynb
- **1 Exercise notebook:**
  - 1 solution notebook: template_datathon_with_solution.ipynb
- **Instruction document** covers how to use Google colab notebook, how to access all the slides and notebooks and how to attend Kaggle competition.

Top tips from 2 Kaggle Wizards in our team

Dirk Nachbar
Head of Applied Data Science
gTech

Alessandro Mariani
Applied Data Scientist
gTech

How to do well in competitive ML?

1. Invest time.
2. Learn from others (improve on other people's solutions).
3. Be explorative, spend good time on engineering features. Do some reading on the domain.
4. Try many diverse approaches, later you can average/ensemble them.
5. Start simple (linear) and then add complexity incrementally.
6. Partition your data to mirror the competition setup. Use the loss/objective function to select models.
7. If the data is big and slow, play with samples.

WOMEN IN DATA SCIENCE
@ GOOGLE

Tips to succeed on Kaggle (advice from Ale & Dirk*)

How to do well in competitive ML?

1. Invest time.
2. Learn from others (improve on other people's solutions).
3. Be explorative, spend good time on engineering features. Do some reading on the domain.
4. Try many diverse approaches, later you can average/ensemble them.
5. Start simple (linear) and then add complexity incrementally.
6. Partition your data to mirror the competition setup. Use the loss/objective function to select models.
7. If the data is big and slow, play with samples.

Others

1. Perseverance (invest time) - I used to dedicated ~3 hours a day when competing
2. Understand what you're doing - forget the competition and focus on learning how trees, neural networks and linear models works and how you need to prepare data differently. You can come back to competition later.
3. You need to have a solid cross-validation setup to understand if your experiments works (don't rely on public leaderboard feedback, this is how I won my first competition!)
4. Team up - but don't team up before you ran out of ideas! Teaming up is great way to share what each other learnt
5. Read the forum - especially past competitions! Threads are full of knowledge

*Both top 15 before Kaggle got really big :)