



KU Leuven
DATATHON 2023
KICK-OFF



Made possible by

RISK
CONCILE



KPMG

pwc

Deloitte.

 **KBC**

WELCOME STUDENTS!

We are thrilled to welcome all of our talented students who have accepted the challenge to torture the data until it confesses!

Without students, no **Datathon**
@KULeuven!

KU LEUVEN





WELCOME SPONSORS!

We are equally excited to welcome our
generous sponsors!

Without sponsors, no datathon @ KULeuven! (or it would be one without food, drinks and prizes...). Take this moment to introduce yourselves.

RISK
CONCILE



KPMG



pwc



Deloitte.



KBC

RiskConcile

Fast-growing FinTech based in the centre of Leuven, looking for talent !

RISK
CONCILE



KPMG in Belgium

Meet our Lighthouse
representatives



**Bart Van
Rompaye**

Senior Manager



**Jorgo
Haezaerts**

Junior Advisor



**Julien
L'hoest**

Junior Advisor

17 years certified Top Employer

Flex

career path

44%
female



1900

employees in Belgium

300

graduates joining us
each year

Flex

compensation

56%
male

8
offices &
hybrid working

Meet us in person

Join us in Ghent on 13 October!



Keep up to date

Join our Talent Community



**Do work that
matters**



**Come as you
are**

**Thrive with
us**



**Learn for a
lifetime**



**Make your
mark**

PwC presentation

What comes to mind when you think of PwC?
Auditing? Tax? Accounting? Consulting? We're
all of those things - and so much more!

An opportunity to **learn with purpose, lead with heart, and make a positive impact** on the world.

PwC is a global **community of solvers!**



Interested in joining
PwC?
www.careers.pwc.be



Wouter Travers



Laurens Vercauterden

Deloitte.

Deloitte Belgium offers value added services in audit, risk advisory, accounting, tax & legal, consulting & financial advisory services.

Some numbers for 2022:

Headcount: 6000+

New Hires: 1642

Revenue: 706 Mio Euro

Net Revenue Growth: +11,1%

Choose your impact!

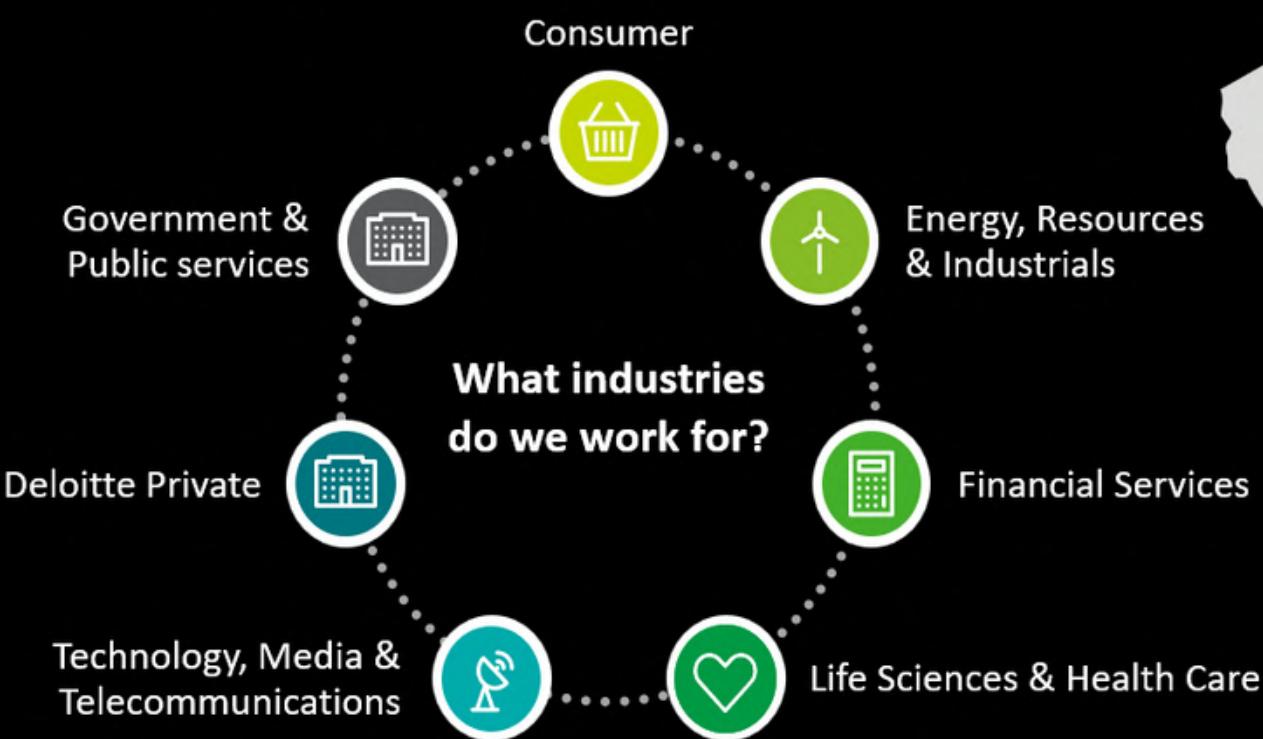
deloitte.com/be/mycareer





Let's connect!

You can apply on our career website





Meet our people



Anthony Coppens

Manager in the Data Science team
of AI & Data



Abbas Nassereddine

Working on Data Science Projects with
an insurance market leader and a
European institution



Dekeyser Gilles

Senior Consultant in the Data
Science Tribe migrating to Data
Architecture

Differently THE NEXT LEVEL

KBC @ KUL Datathon & Jobfair



Moving forward
together.

KBC - 'the reference'

Formed in 1998 after the merger of two Belgian banks (Kredietbank and CERA Bank) and a Belgian insurance company (ABB Insurance).

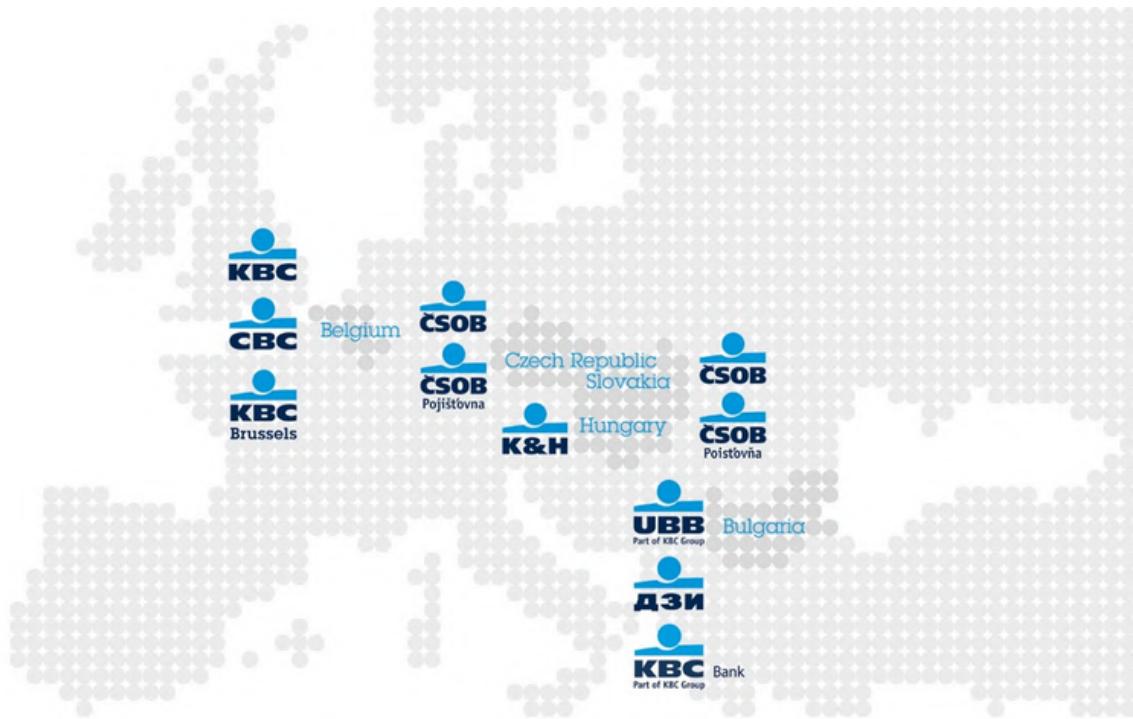
Principal activity: integrated bank-insurance.

Clients: 13 million.

Core markets: Belgium, the Czech Republic, Slovakia, Hungary, Bulgaria. Also present, to a limited extent, in other countries.

Network: ca. 1 200 bank branches, insurance sales via own agents and other channels, various electronic channels.

Employees: 42 000

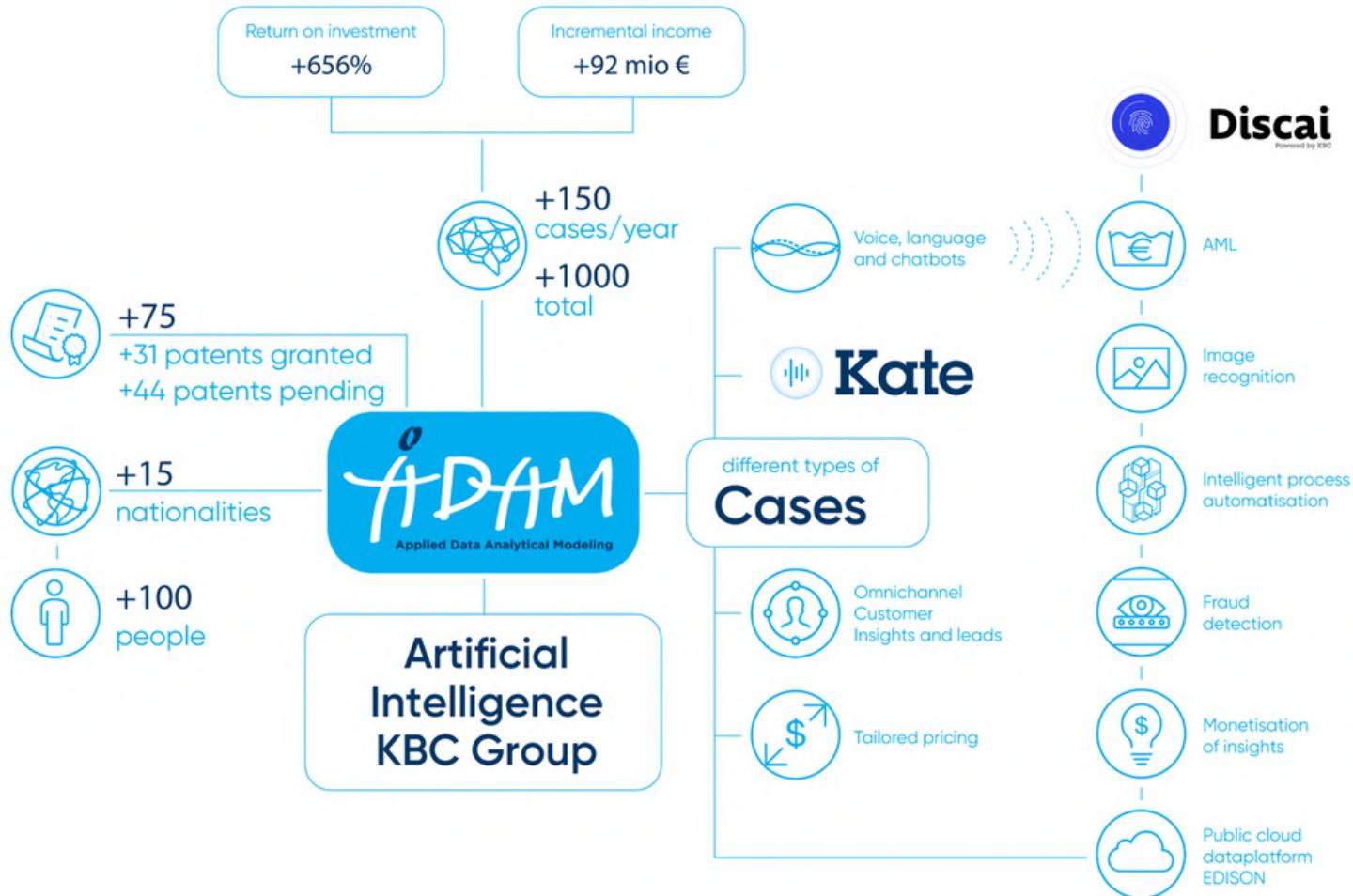




Driven by data
Powered by *ADAM*

Internal

Differently THE NEXT LEVEL



Internal

Differently THE NEXT LEVEL

Multidisciplinary and multicultural team



Internal

Differently THE NEXT LEVEL

Meet my colleagues on the 23rd of February



Karsten Gielis



Kamran Mammadzada



Tamás Baráth

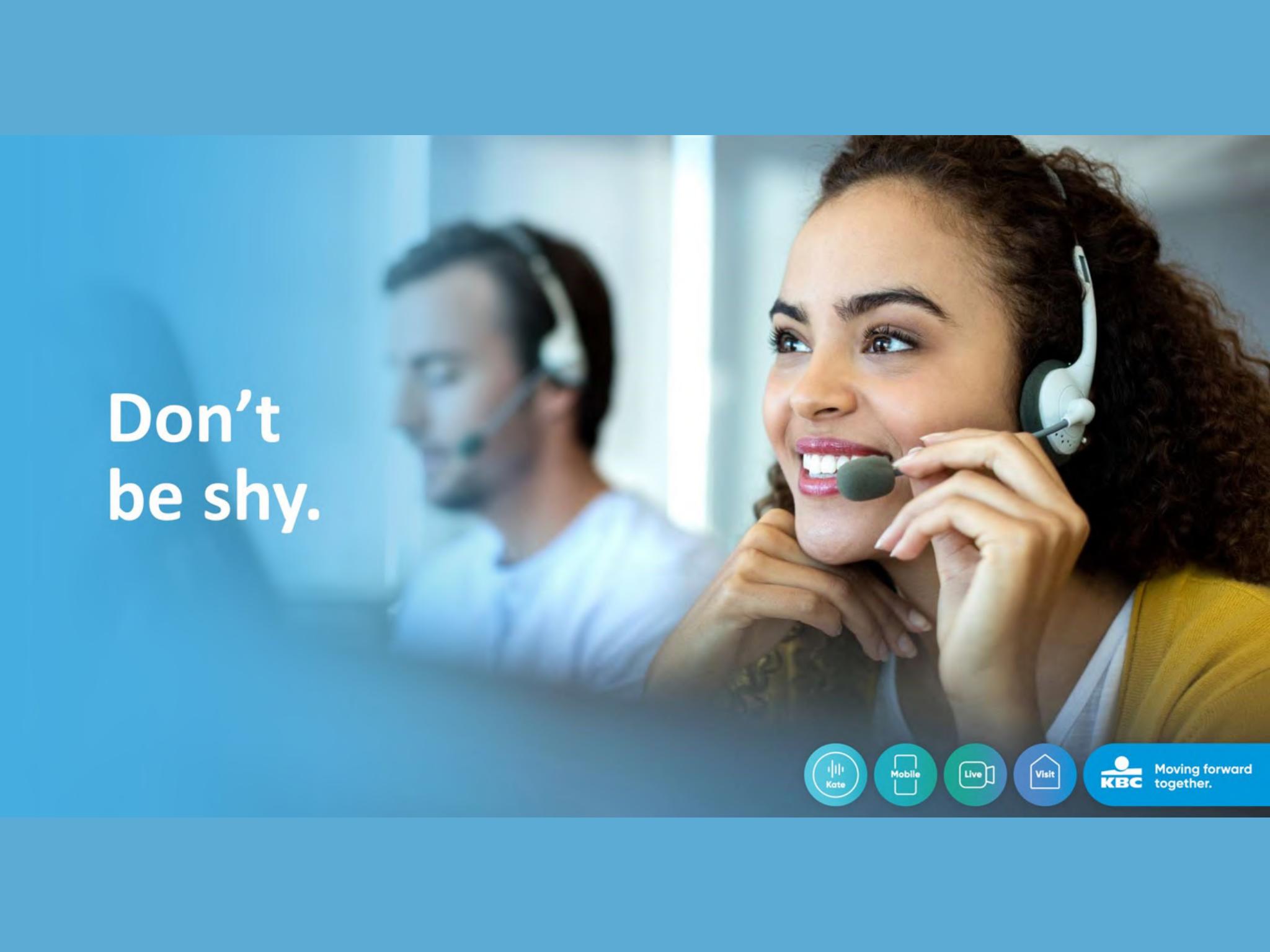
Internal

Differently THE NEXT LEVEL

We are hiring!

Data Scientist





Don't
be shy.



Kate



Mobile



Live



Visit



KBC

Moving forward
together.



LET'S GET PRACTICAL

Timing is key to a successful datathon

MARK YOUR CALENDARS

February 23

14h00 - DEADLINE

- Who? Team captain
- Where? Google Drive (Shared in due time)
- What? Pitch slides + code (or link to GitHub/App), no data

The earlier, the better!!

16h00 - Voucher Pick-Up

16h15 - Start Pitching

19h00 - Award Ceremony

21h00 - End meet-the-sponsors

21h00 - End jobfair



MARK YOUR CALENDARS

February 23

14h00 - DEADLINE

16h00 - Voucher Pick-Up

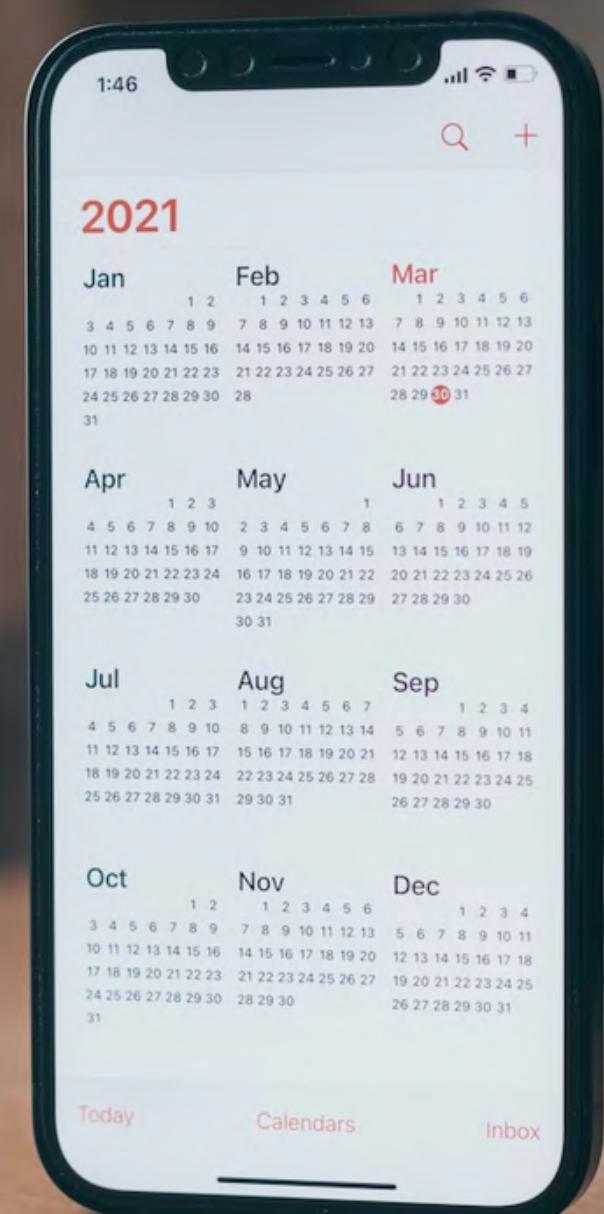
- Team Captain picks up voucher for catering
- Drinks and food are served 16h15-19h30

16h15 - Start Pitching

19h00 - Award Ceremony

21h00 - End meet-the-sponsors

21h00 - End jobfair



MARK YOUR CALENDARS

February 23

14h00 - DEADLINE

16h00 - Voucher Pick-Up

16h15 - Start Pitching

- Teams **pitching**
- Team captain will be informed beforehand where and when.

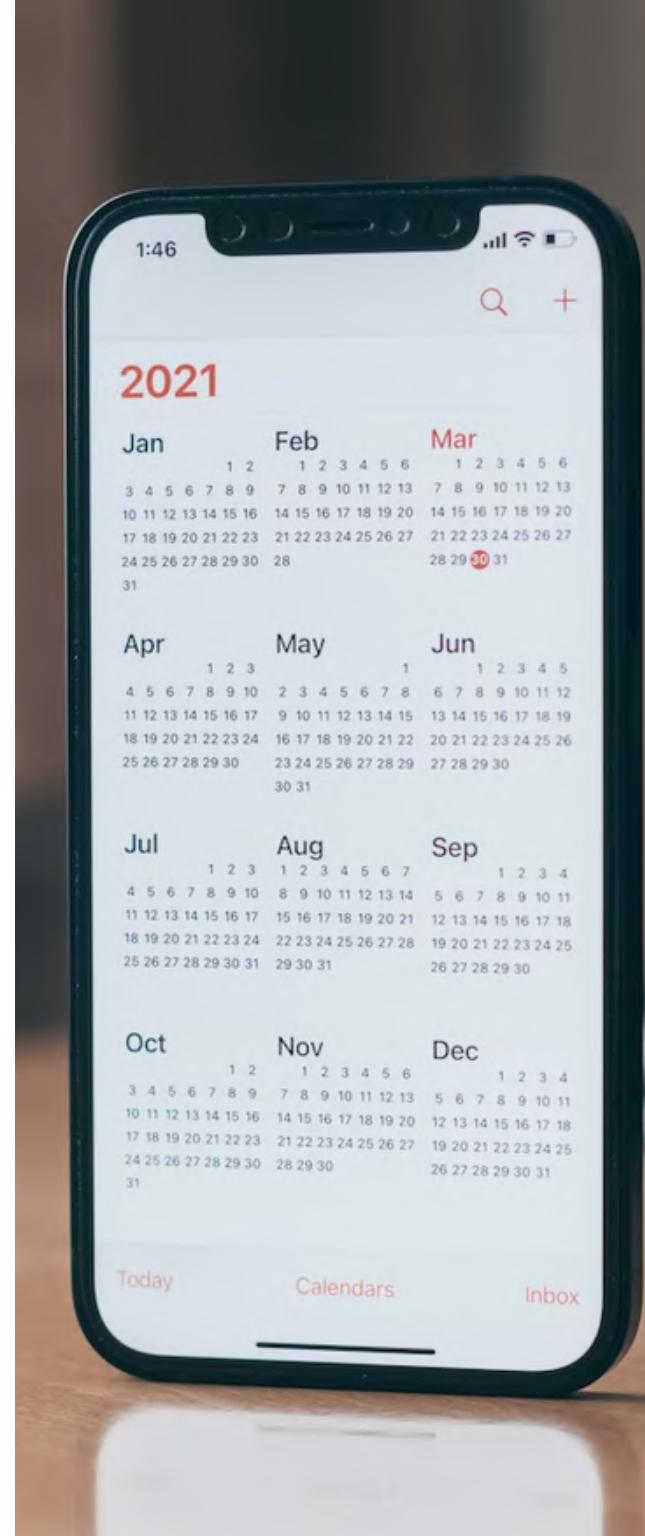
Ground rules:

1. Use your **own laptop**.
2. There are **HDMI** cables (but no HDMI Thunderbolt adaptors)
3. Be at least **5 minutes earlier**
4. Pitching **behind closed doors**
5. Time limited to **5 minutes**
6. Jury Q&A of 5 minutes after pitch
7. Teams are graded based on modeling, pitching, and visualization.

19h00 - Award Ceremony

21h00 - End meet-the-sponsors

21h00 - End jobfair



MARK YOUR CALENDARS

February 23

14h00 - DEADLINE

16h00 - Voucher Pick-Up

16h15 - Start Pitching

19h00 - Award Ceremony

- Each member of winning teams receive a €200 CoolBlue voucher
- Price categories:

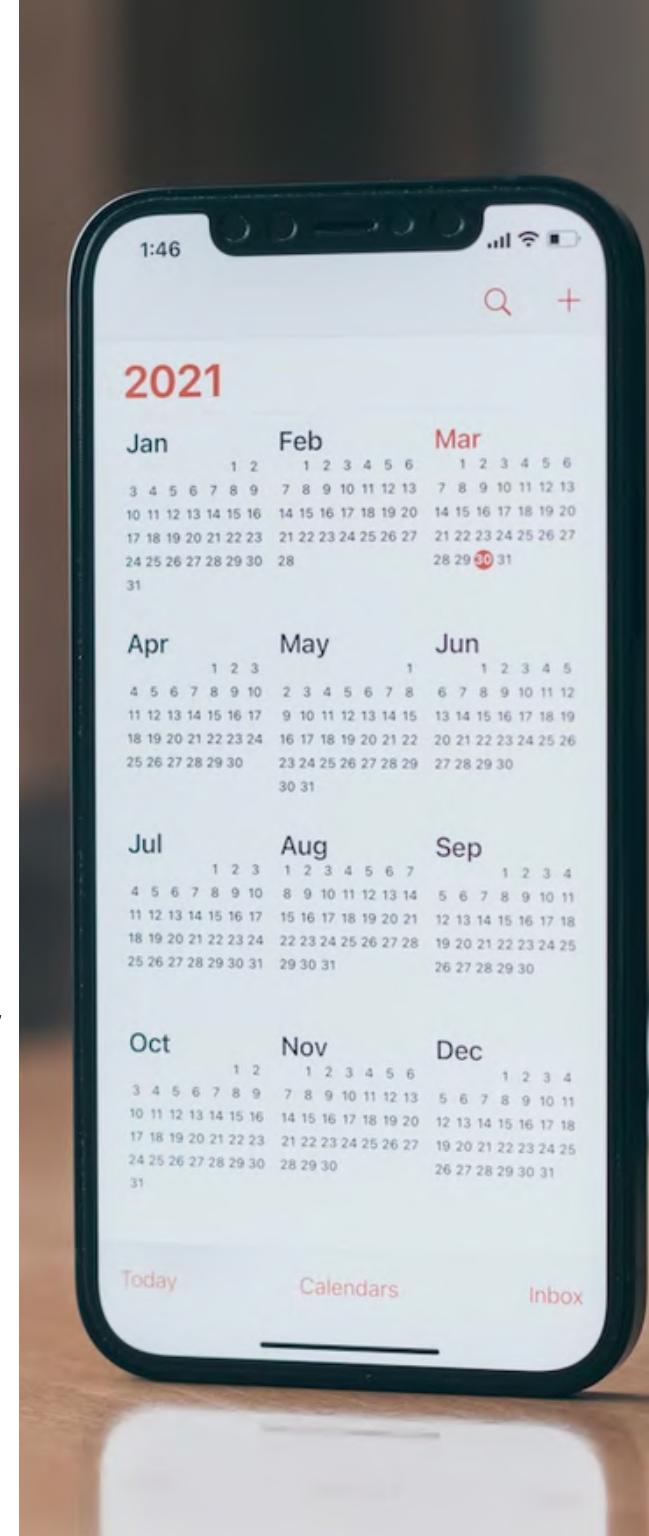


Best Pitch Best Visuals Best Model Best Bachelor Best Master

- Pitch in front of audience that votes for '**Winner of the Datathon 2023'**

21h00 - End meet-the-sponsors

21h00 - End jobfair



MARK YOUR CALENDARS

February 23

14h00 - DEADLINE

16h00 - Voucher Pick-Up

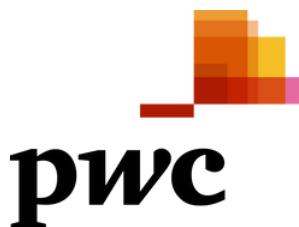
16h15 - Start Pitching

19h00 - Award Ceremony

19h30 - Meet-the-sponsors reception

- Meet our sponsors from 19h30 - 21h00

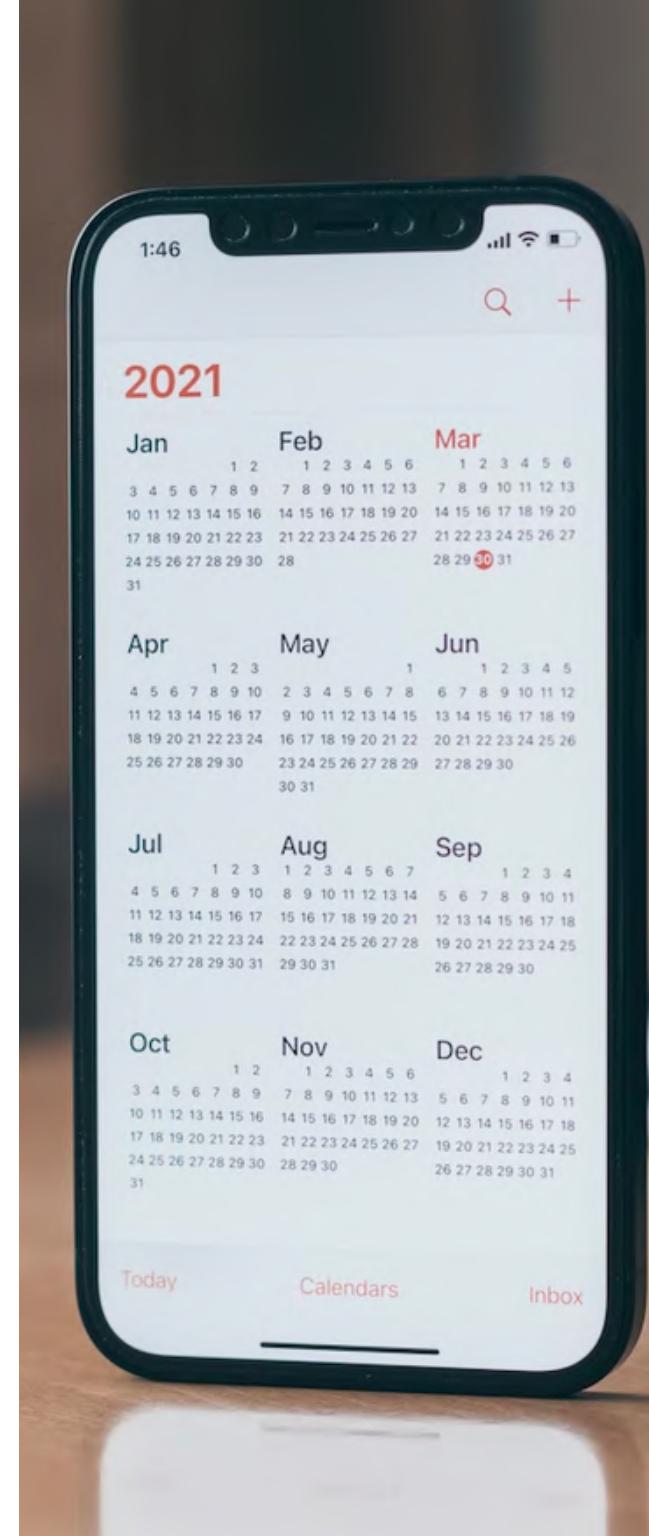
RISK
CONCILE



Deloitte.



21h00 - End jobfair



MARK YOUR CALENDARS

February 23

14h00 - DEADLINE

16h00 - Voucher Pick-Up

16h15 - Start Pitching

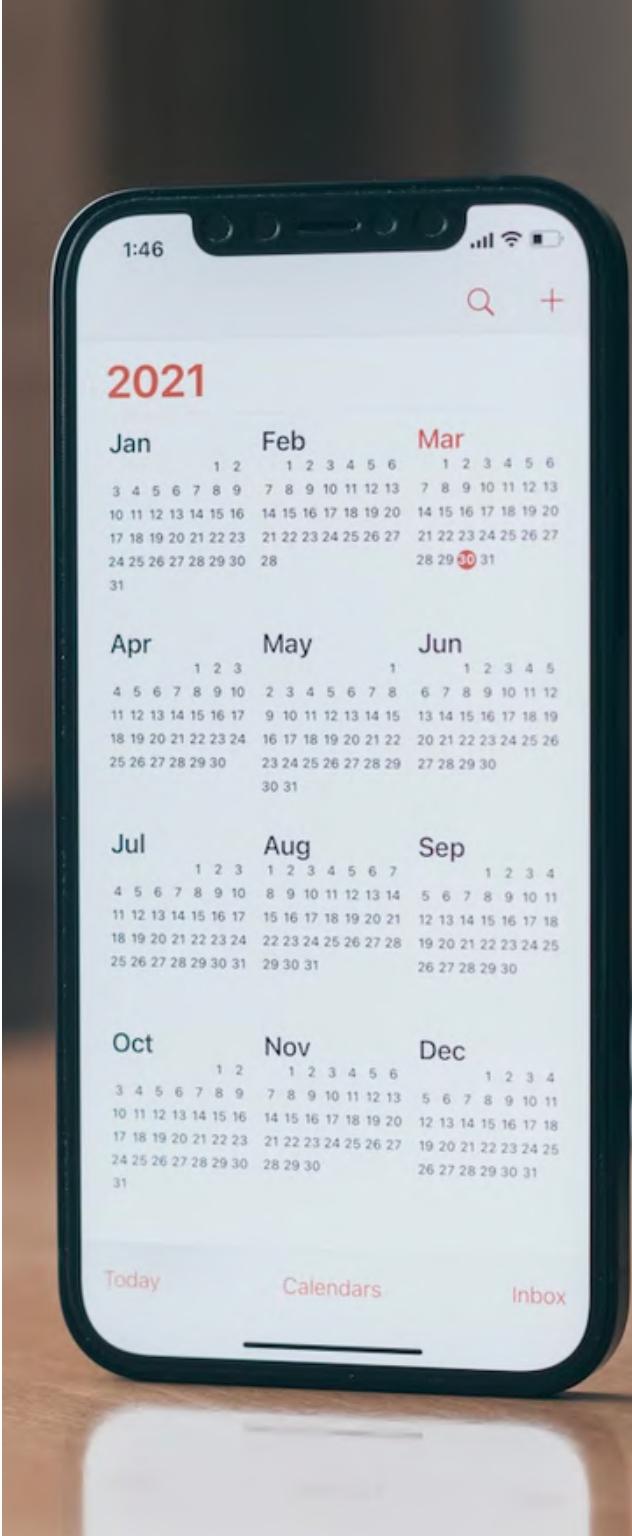
19h00 - Award Ceremony

21h00 - End meet-the-sponsors

21h00 - End jobfair

There is also a jobfair in the Quadrivium 01.100.

You are more than welcome to register.



CHILL(Y) AND COOL

In order to save energy, KU Leuven has decided to shut down the heating at 17h00.

Be sure to wear your favourite sweater!

KU LEUVEN





...AAAND ACTION!

A film crew & photographer will be present at the Datathon, so be sure to strike a pose!

The pictures and after-movie will be shared with you on our website! That way, you can share your experiences with friends and family

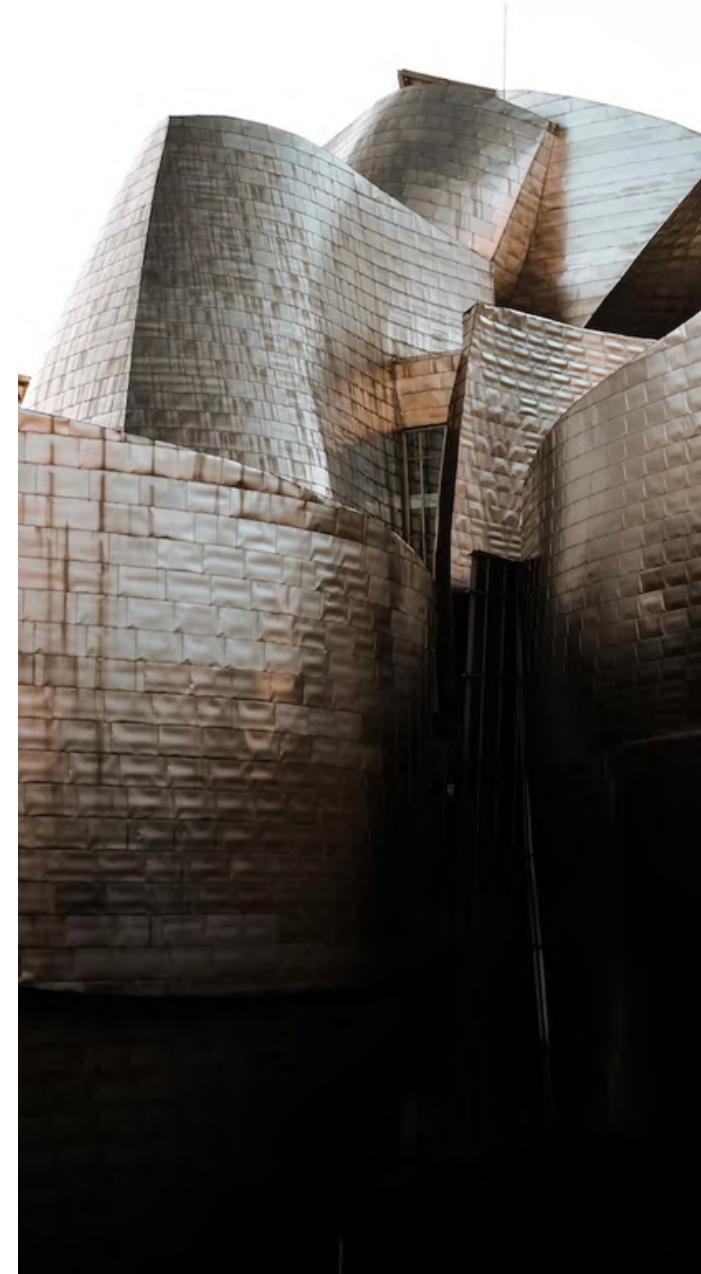
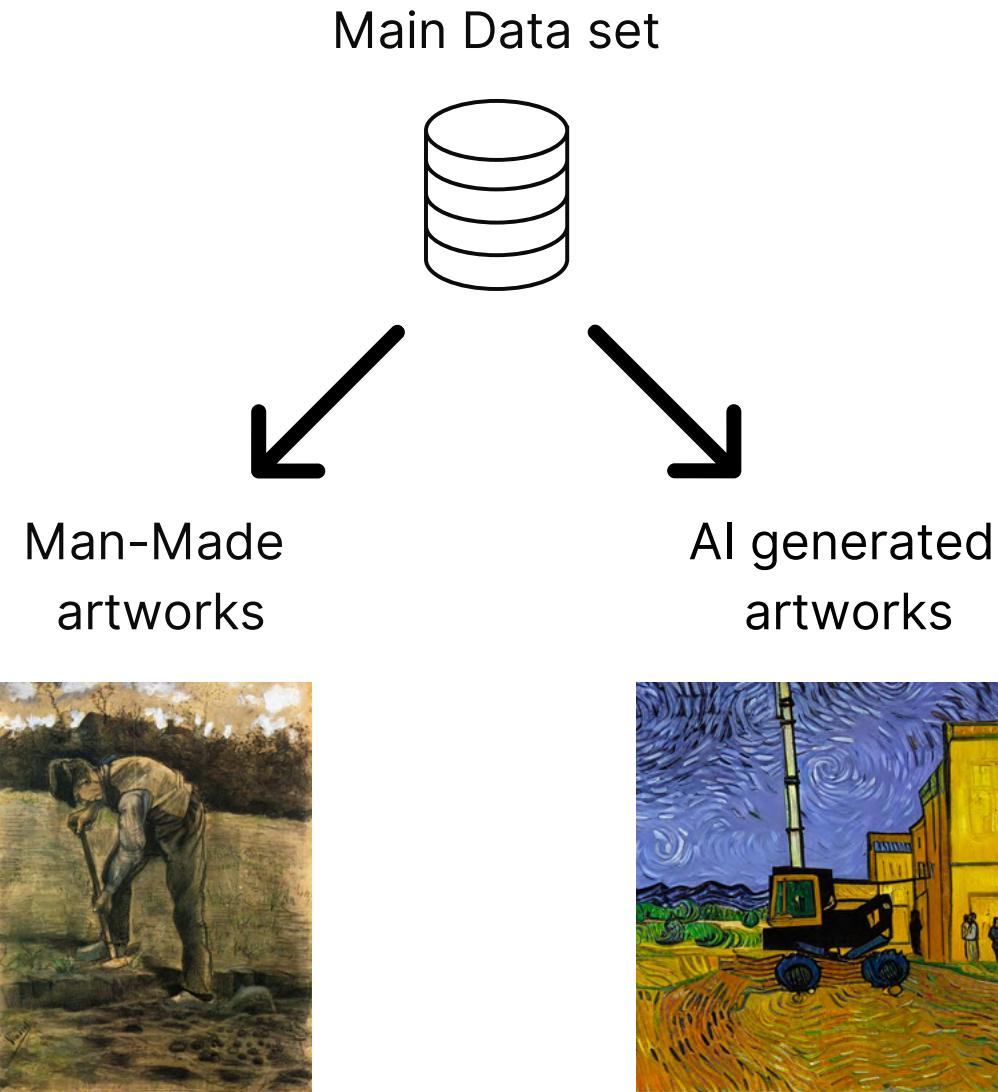
Note: If you wish not to be photographed or have your image recorded, please make your wishes known to the photographer, film crew and/or the event organizers

The background of the entire image is a vibrant, abstract painting in the style of Jackson Pollock. It features a dense, chaotic arrangement of thick, expressive brushstrokes in various colors, primarily red, blue, yellow, and white, creating a dynamic and energetic composition.

MEET THE DATA

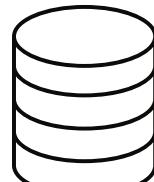
Introducing the main features of the data set

STRUCTURE OF THE DATA



STRUCTURE OF THE DATA

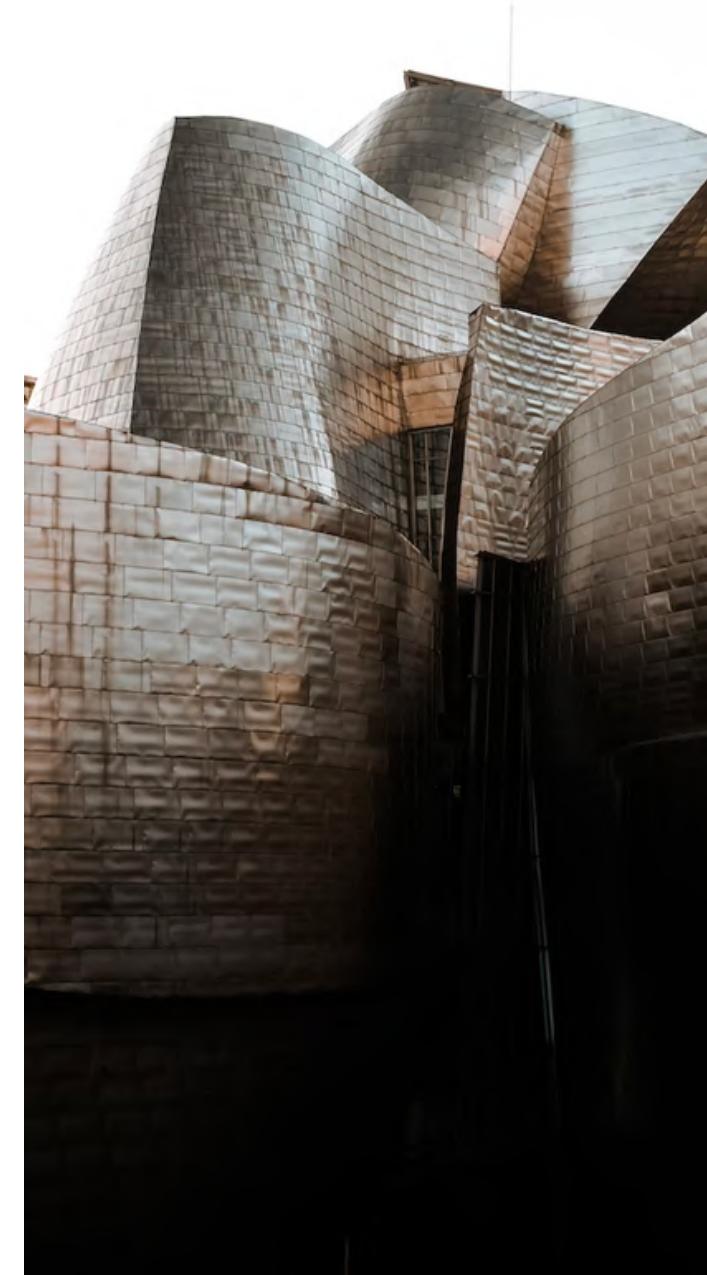
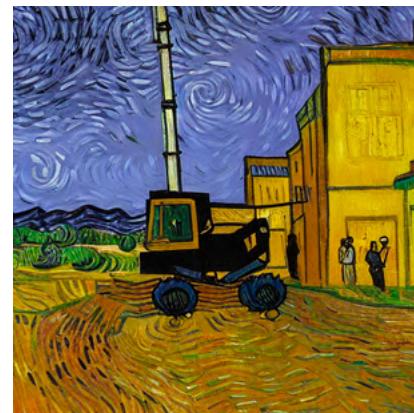
Main Data set



Man-Made
artworks



AI generated
artworks

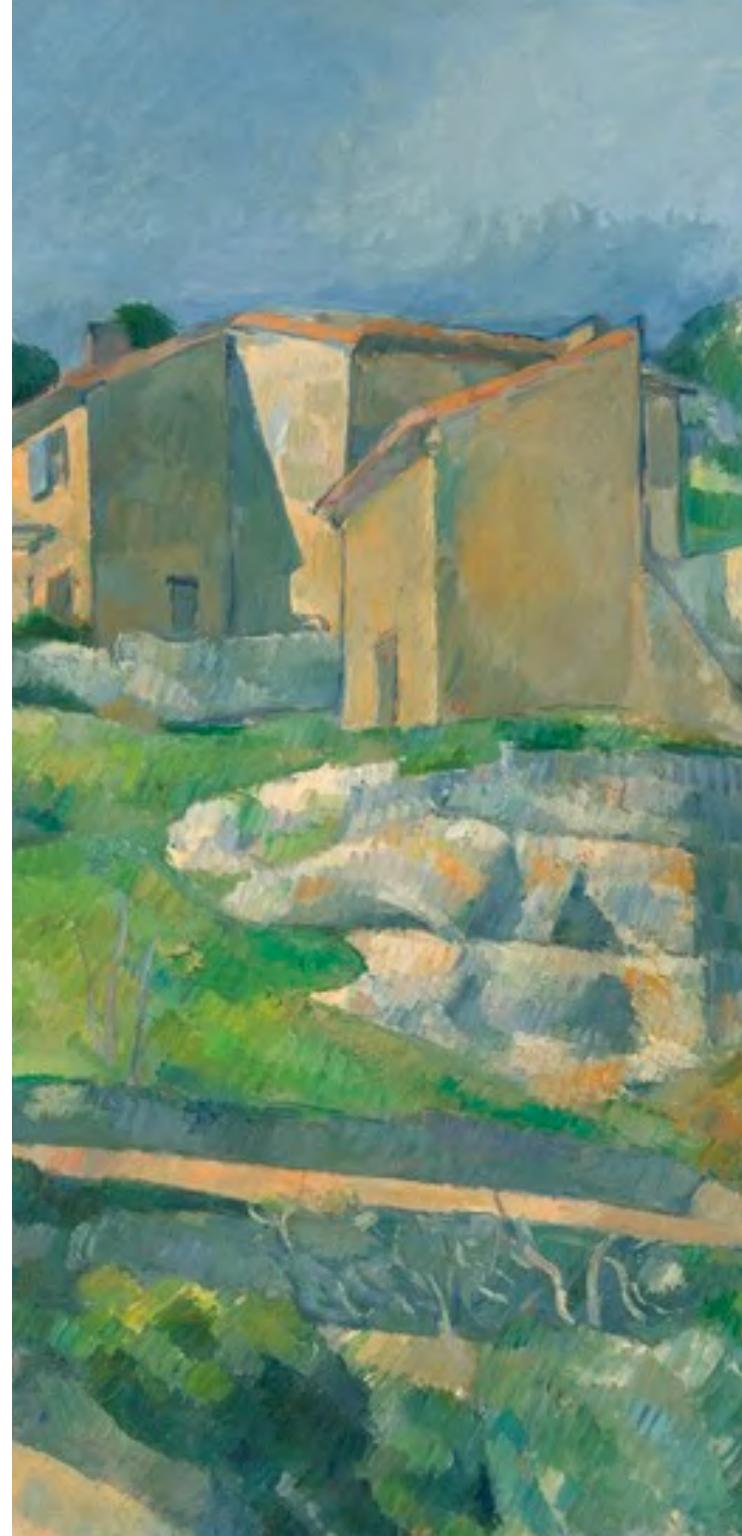


DATA SETS

Main data

The topic for this datathon is art, both real and AI generated. To each artwork, there is either an artist or (in the generated case) an inspiration. The data is not cleaned! Be prepared to preprocess the inputs.

Remember:
Dirty Data = Dirty Results



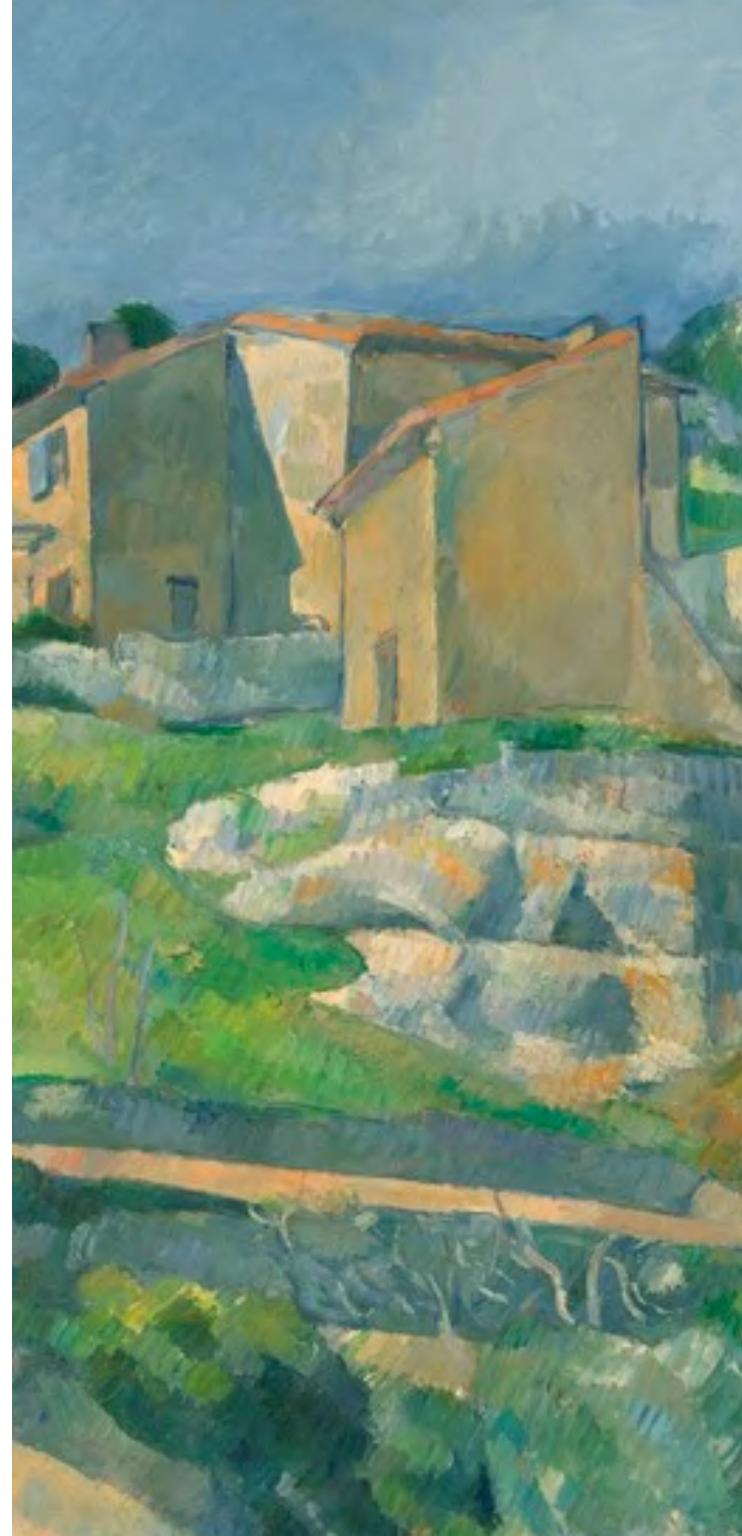
DATA SETS

Main data

Remember:
Dirty Data = Dirty Results

The data set can be approached in many different ways: it can be seen as a tabular data set, a network graph and as an image data set! An interesting answer can only come from an interesting question!

Be creative!



DATA SETS

Main data

Remember:

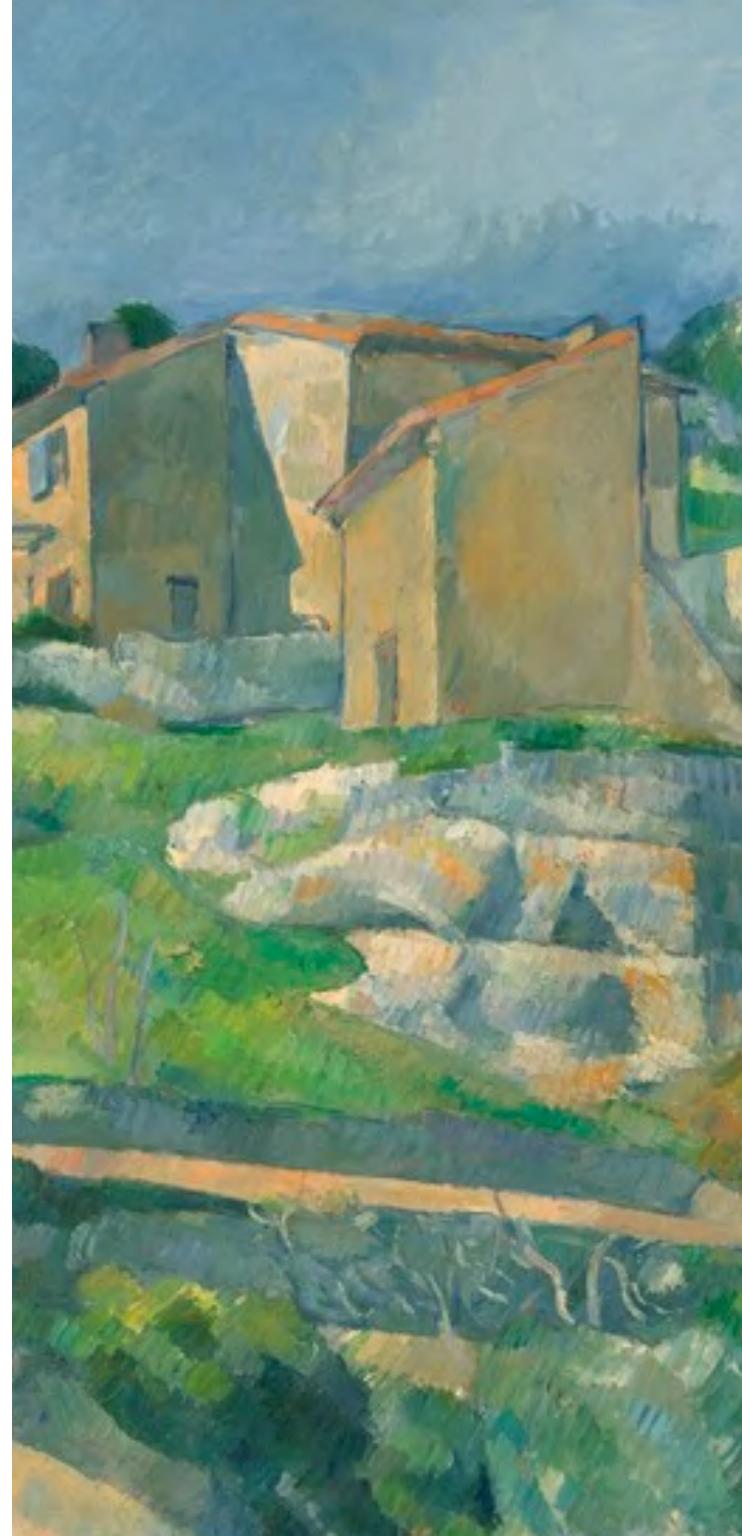
Dirty Data = Dirty Results

Be creative!

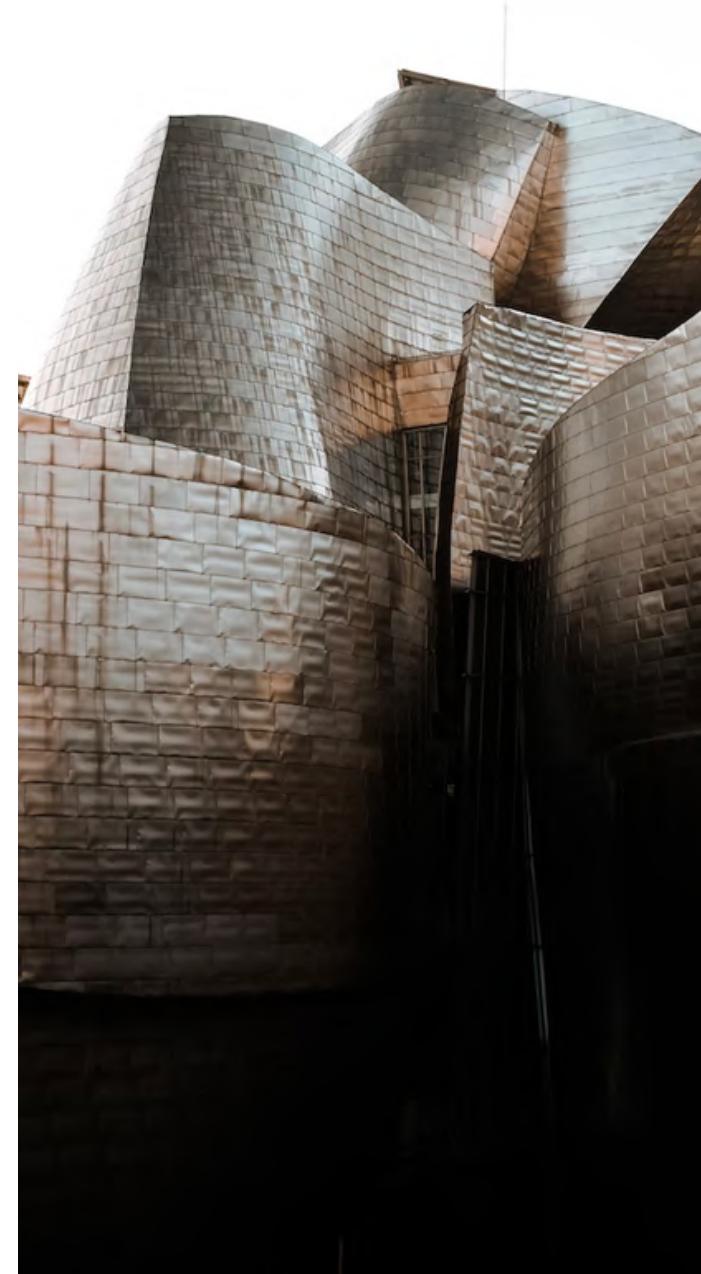
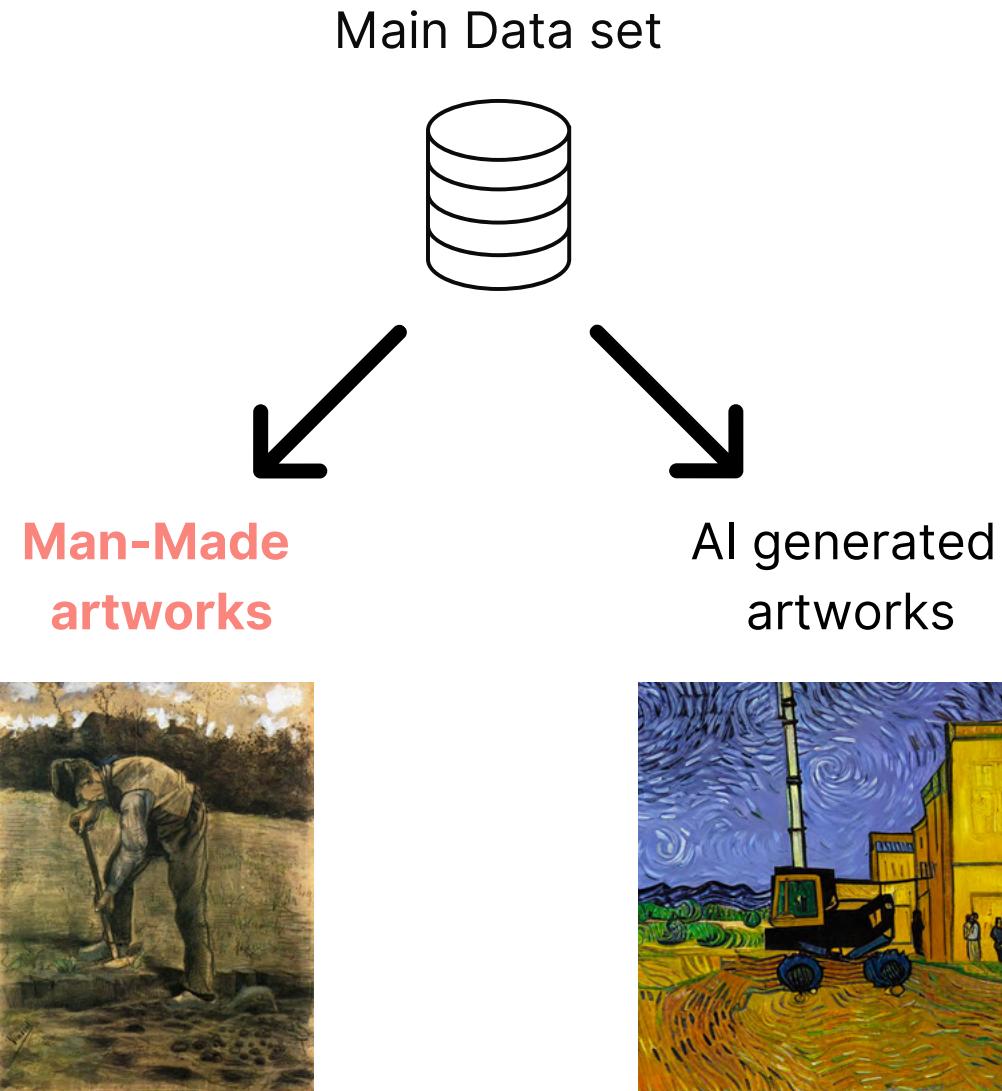
Any good data analysis task takes a lot of man-hours, but you're not alone!

Making an ambitious, but realistic plan of attack is the best place to start.

Work as a team!



STRUCTURE OF THE DATA



DATA SETS

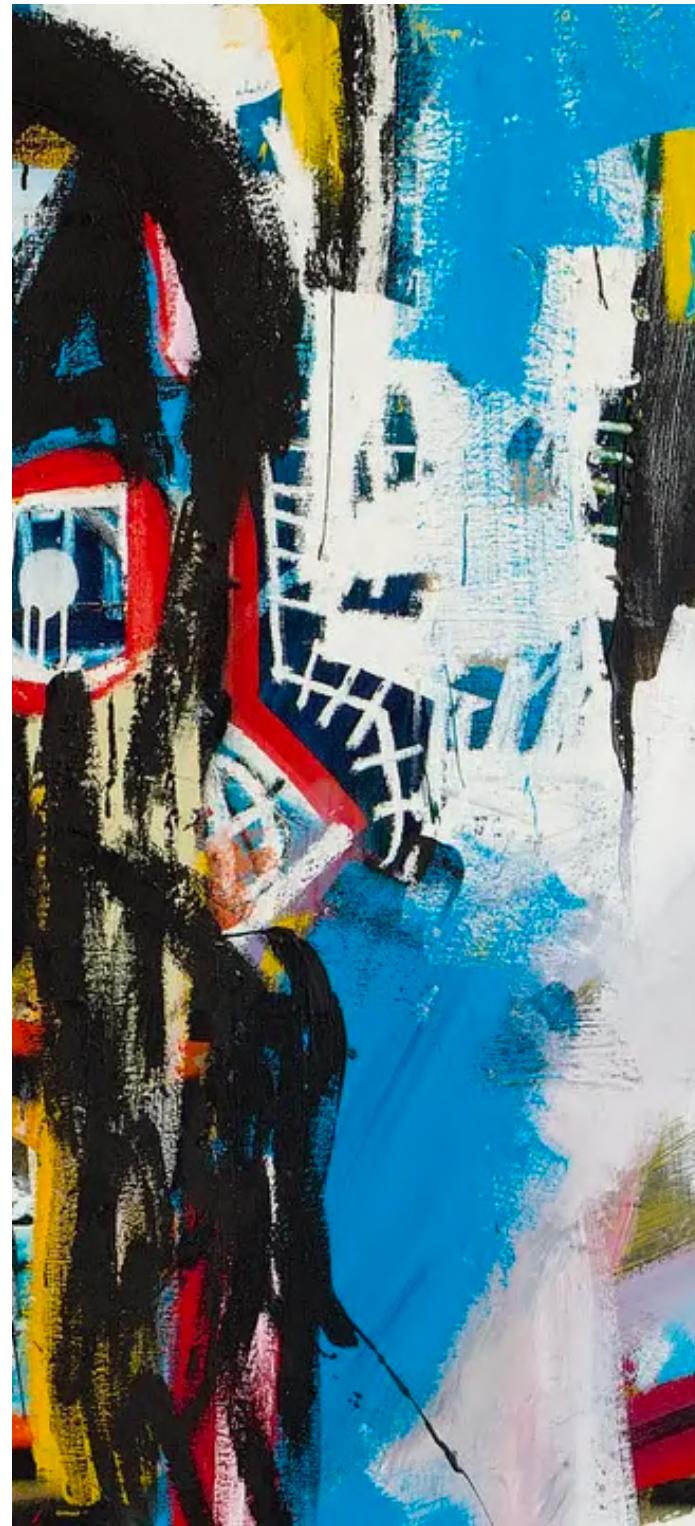
Man-Made Artworks

Multifaceted data set full of potential questions waiting to be analyzed:

- Multiple data sets that can be approached as tabular data. Examples of tables: artworks, artists, (artistic) movements, ...
- Relationships that give the data a graph structure. Relationships include apprenticeships, recommendations, ...
- Filled with image data as well

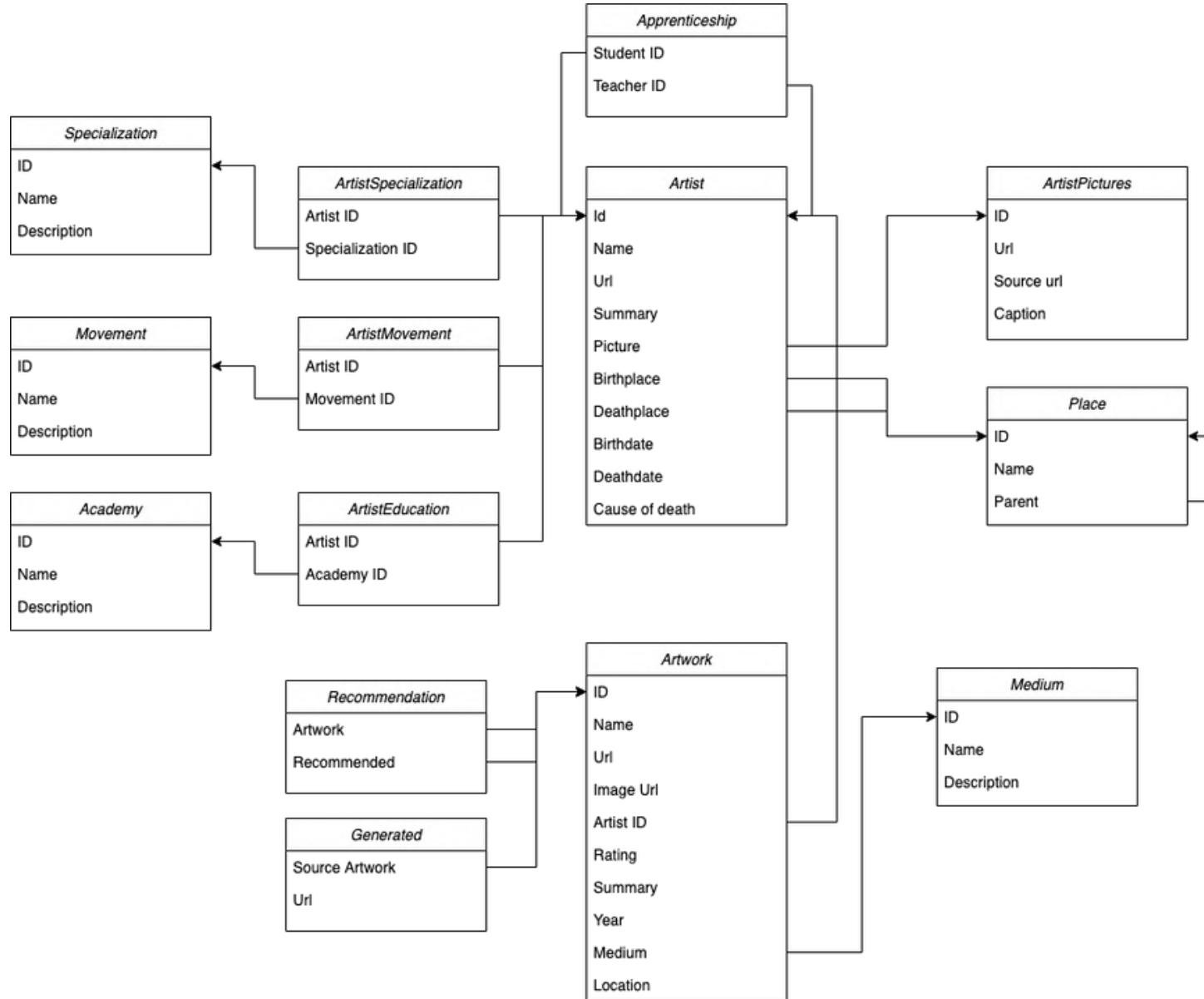
Over 10.000 artworks are included in the data set

Data obtained from wikigallery.com and Wikipedia



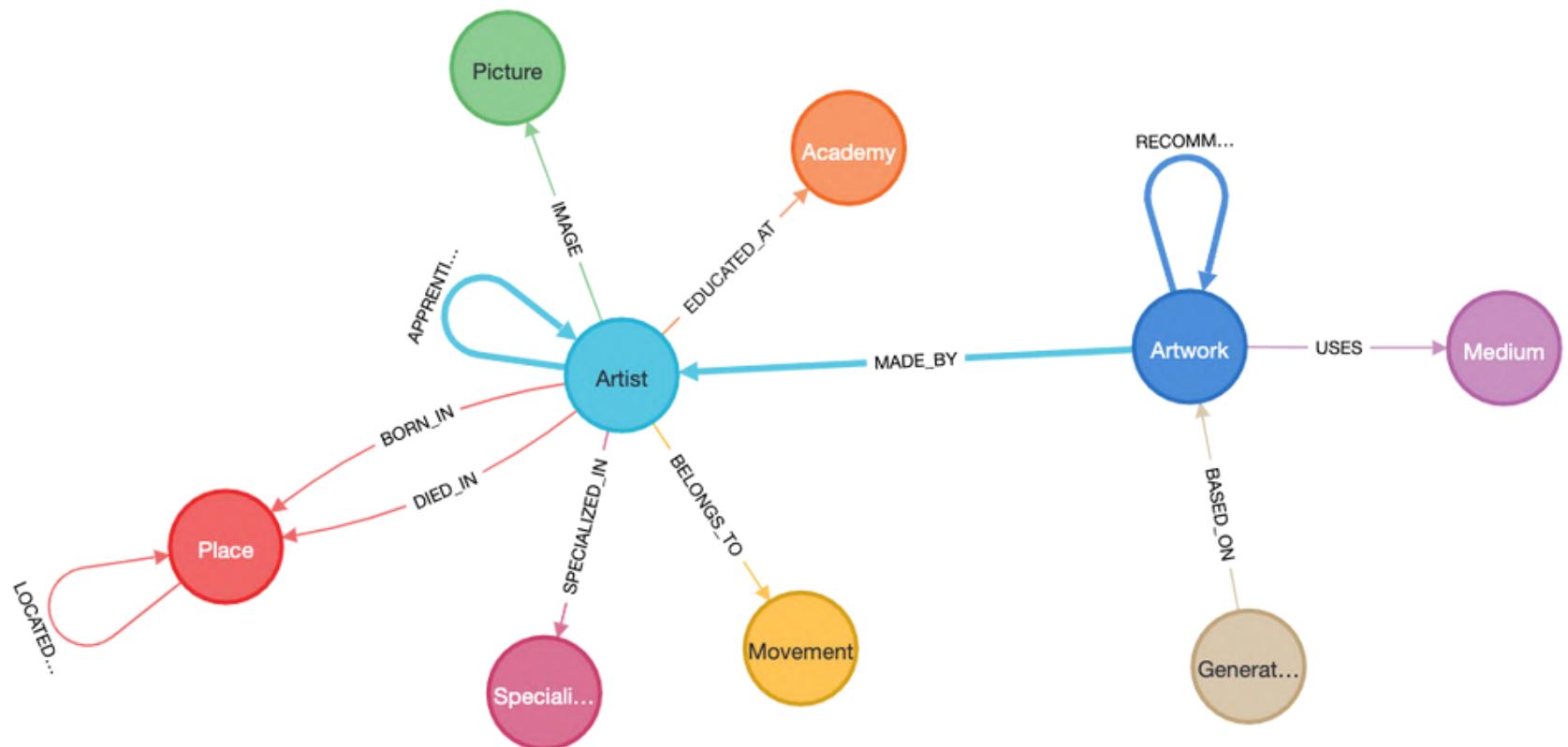
DATA SETS

Man-Made Artworks - ER Schema



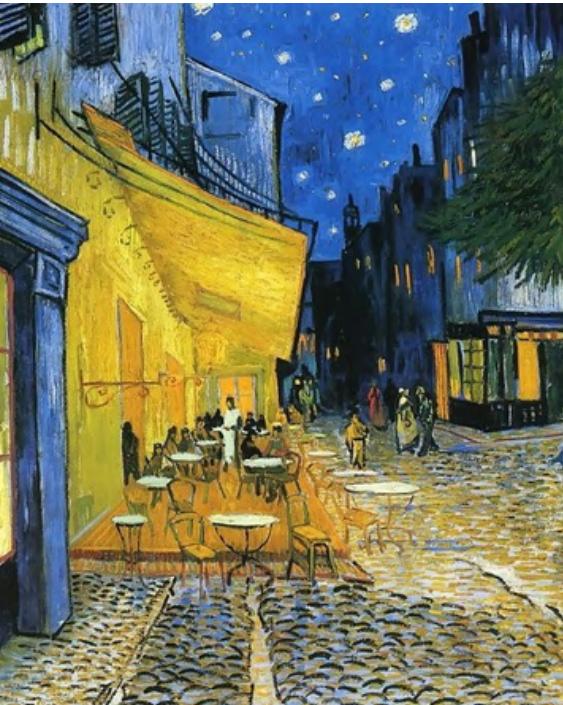
DATA SETS

Man-Made Artworks - Graph Schema

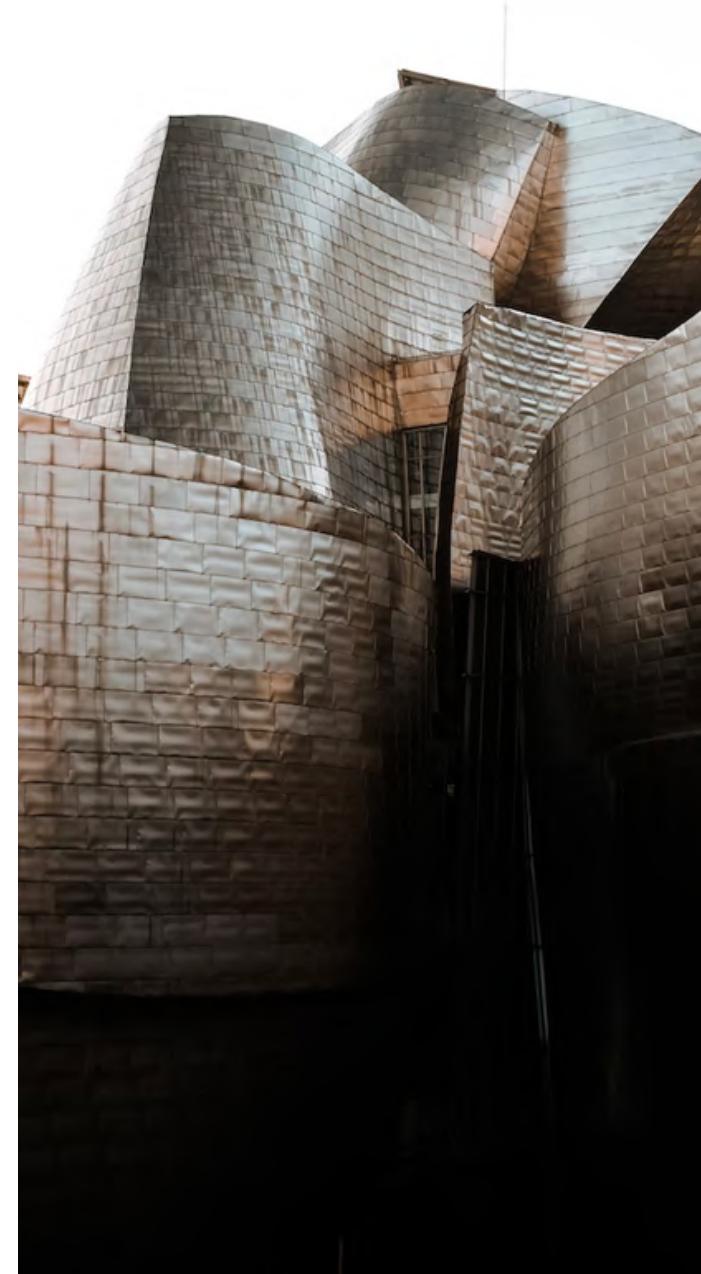
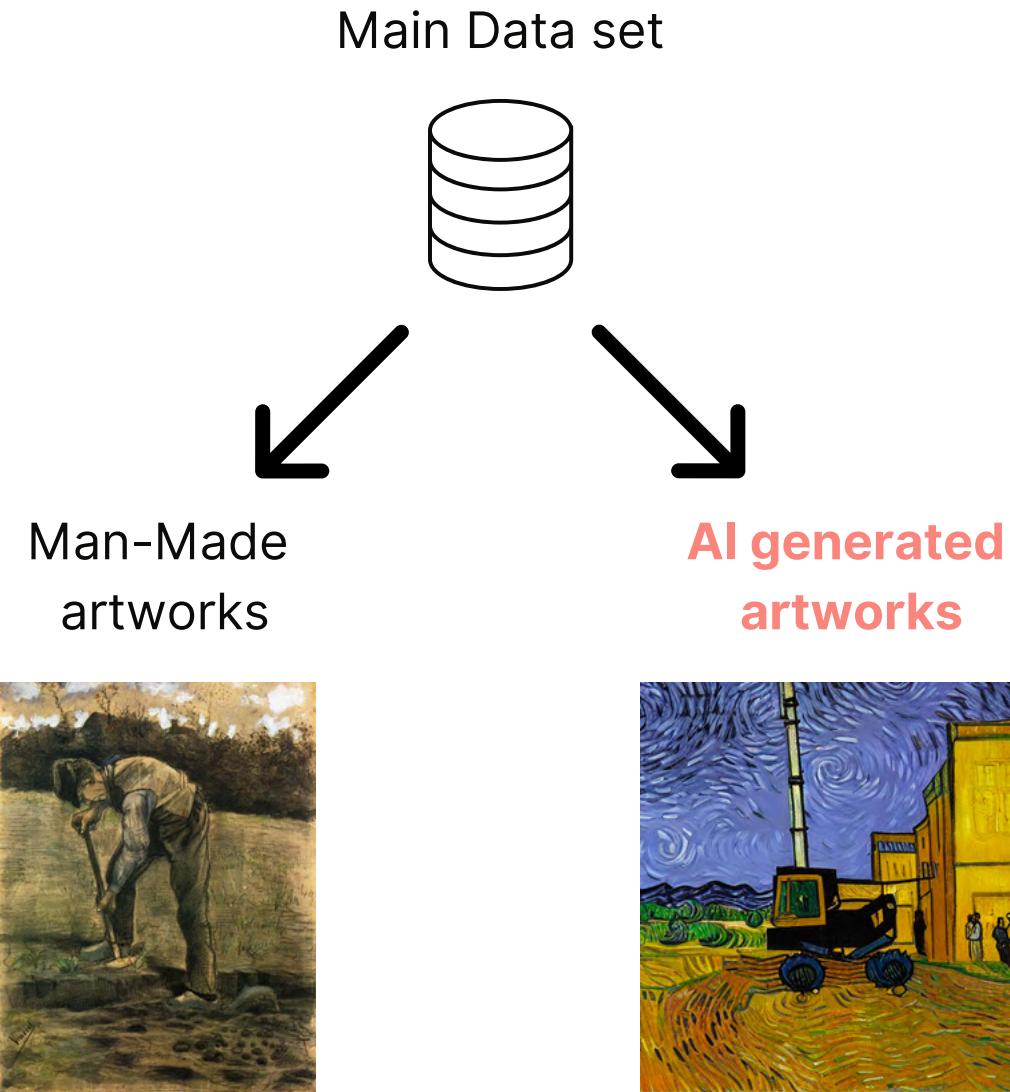


DATA SETS

Man-Made Artworks - Examples



STRUCTURE OF THE DATA



DATA SETS

AI-generated artworks

- Some of the original artworks have an AI-generated copy
- For ethical reasons, the model has an NSFW filter. Generated artworks containing nudity have been replaced with a black square
- No guarantee that the generated artwork resembles the original, or even has the same style

Content warning

Images contained in this data set might be disturbing, offensive, and might misrepresent the original artworks



DATA SETS

AI-generated artworks

All copies have been generated using Stable Diffusion v1-4, a Text-to-Image model:

- Each copy is generated using the prompt:
- A painting of <Title artwork> in the style of <Name artist>.
- All images have been generated using Google Colab
- You can generate your own artworks using this [link](#)
- You are more than welcome to use this code to generate your own images, or use another model!
- Useful resource: <https://huggingface.co/models>

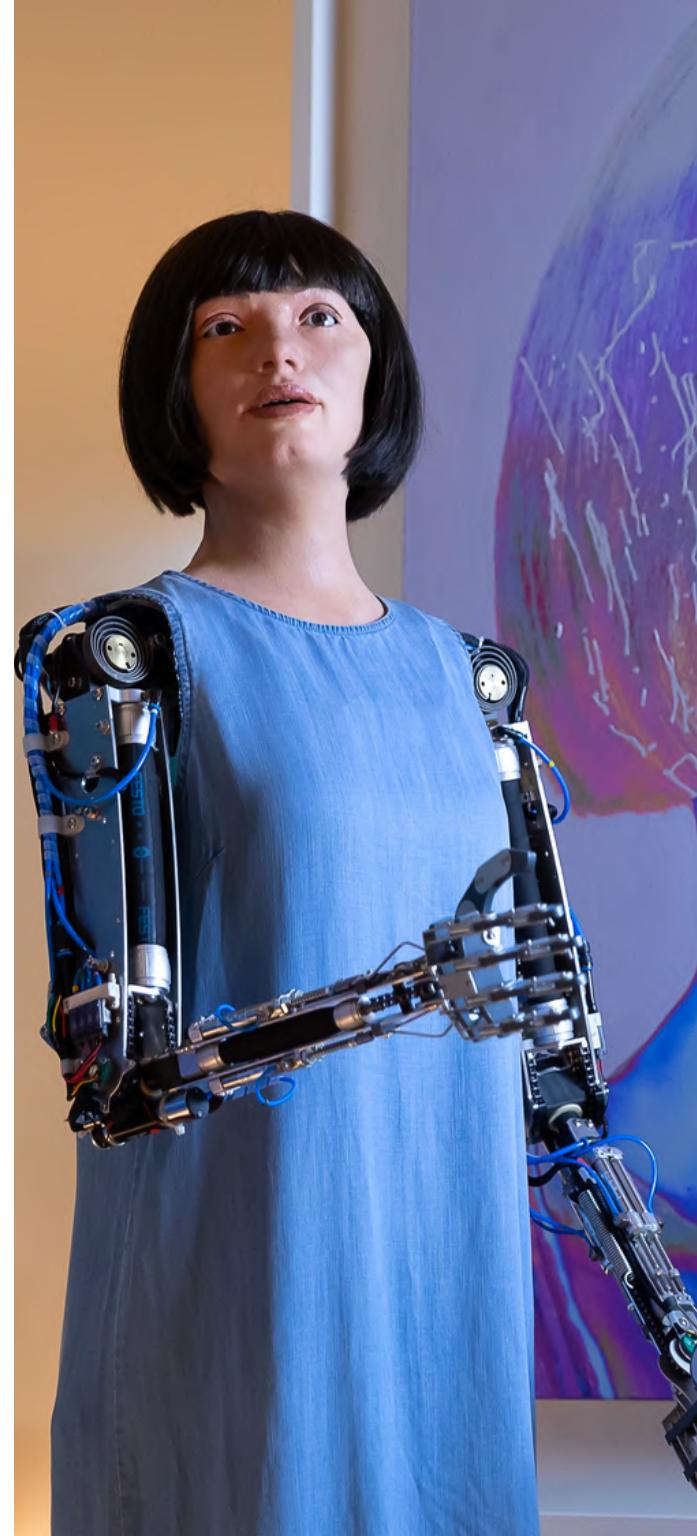
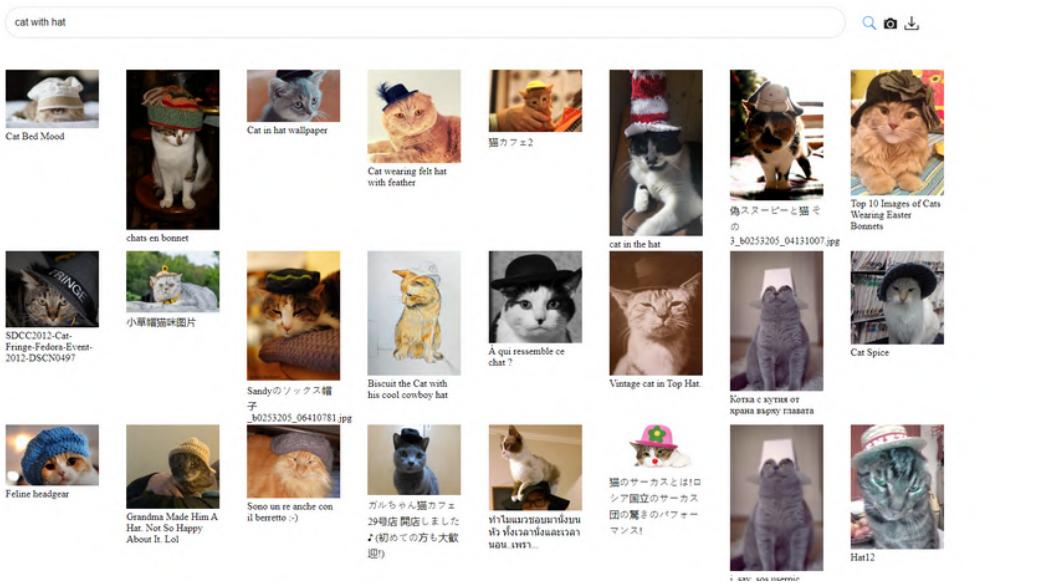


DATA SETS

AI-generated artworks

Stable Diffusion v1-4 has been trained on LAION-5B

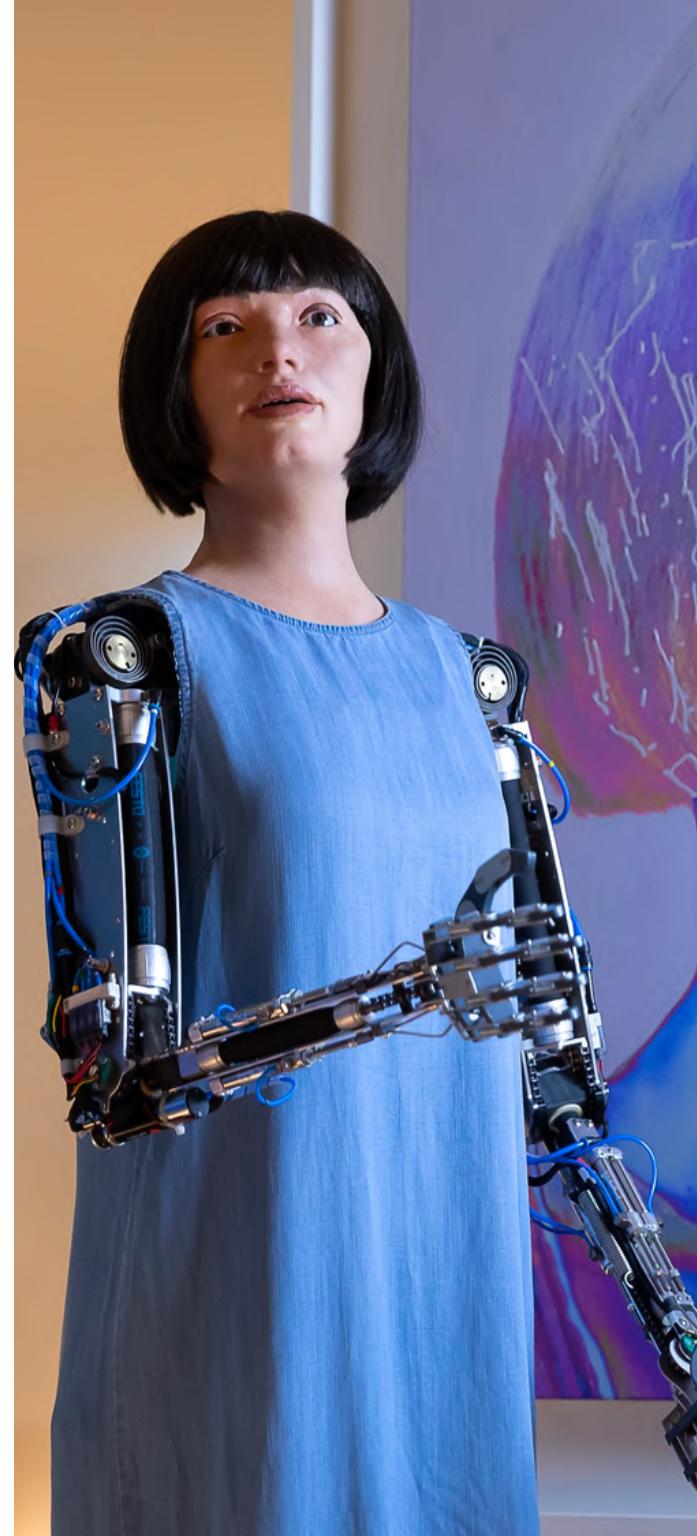
- Contains over 5.85 billion image-text pairs
- Filtered out pictures that do not match the text (CLIP)
- Filtered out pictures containing NSFW content or watermarks
- Contains mostly English data (2.3 billion)



DATA SETS

AI-generated artworks - Model drawbacks

- The model struggles to generate images with readable text in it
- The model sometimes fails to generate anatomically correct bodies, especially hands
- The model has difficulties with understanding spatial relations, such as 'A cat in a box'
- Important for us is that very famous images might be in the training data so frequently that the model starts to remember the artworks, see for example starry night
- Biased towards western culture due to the nature of the data set
- Struggles with non-English prompts
- Fixed size!



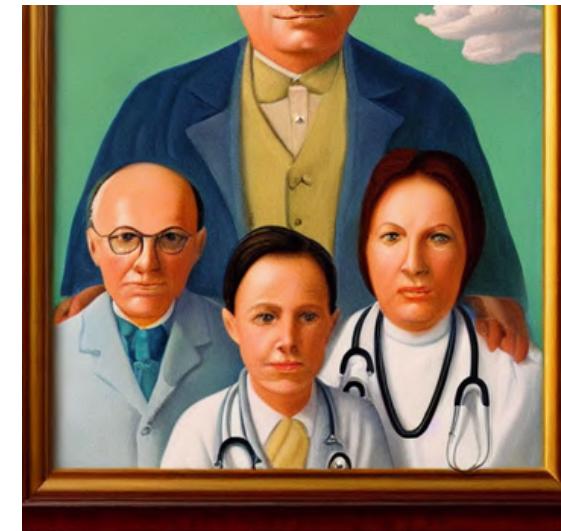
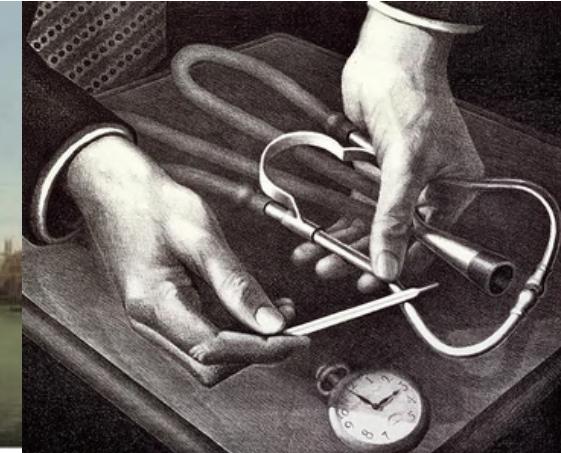
DATA SETS

AI-generated artworks - Model drawbacks



DATA SETS

Original VS AI generated





TIPS & TRICKS

Helpful comments for a succesful datathon

TIPS AND TRICKS

Tooling

Choose the correct tool for the job

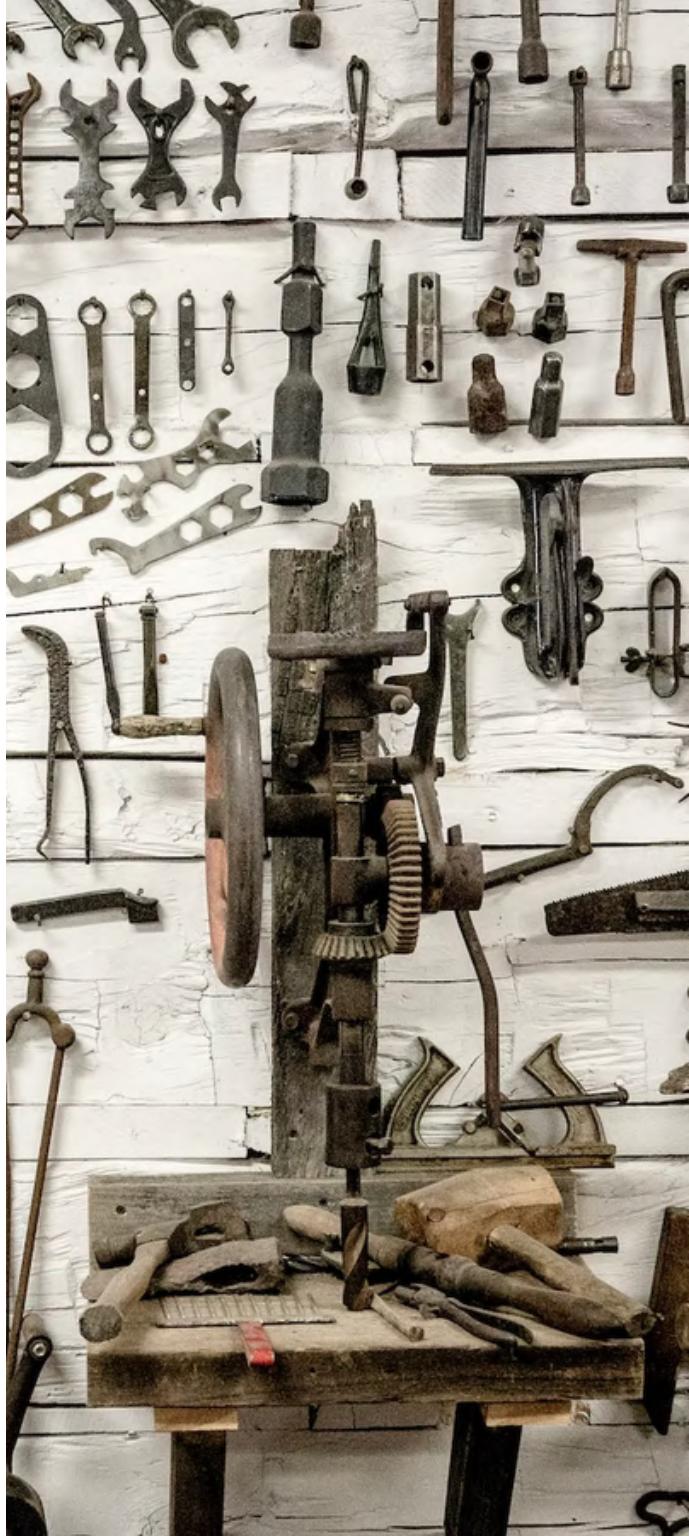
Consider several programming languages and packages before starting your analysis. Using the 'usual' packages is not a bad idea, but might not give you an edge over the competition

(e.g., Gephi for graph visualization, Plotly for interactive plots, ...)

Choose the platform you're going to work on wisely

Apps are appreciated

Have fun!



TIPS AND TRICKS

Tooling

Choose the correct tool for the job

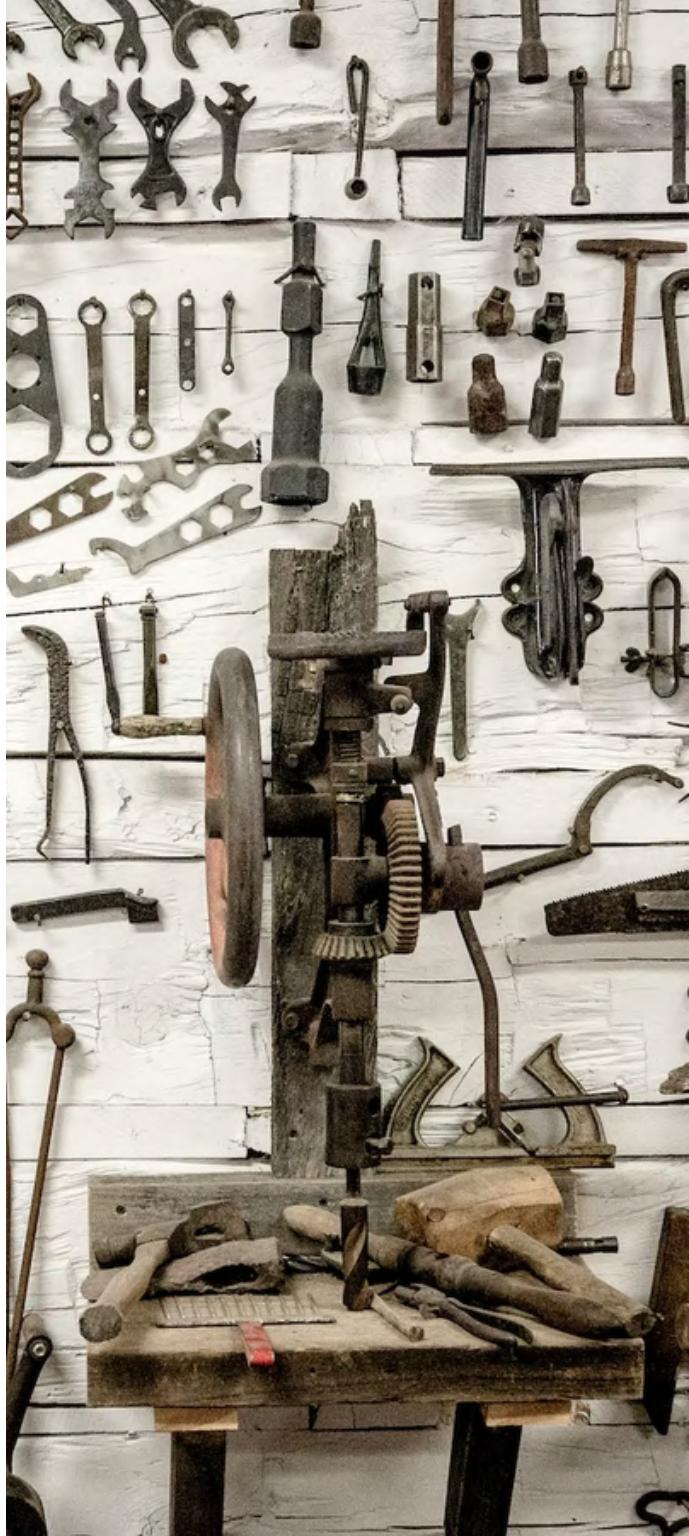
Choose the platform you're going to work on wisely

Working together on the same code is hard, find a solution that fits your whole team!

(e.g., Google colab for Jupyter notebooks. Helpful for their free (but limited) GPU computation services)

Apps are appreciated

Have fun!



TIPS AND TRICKS

Tooling

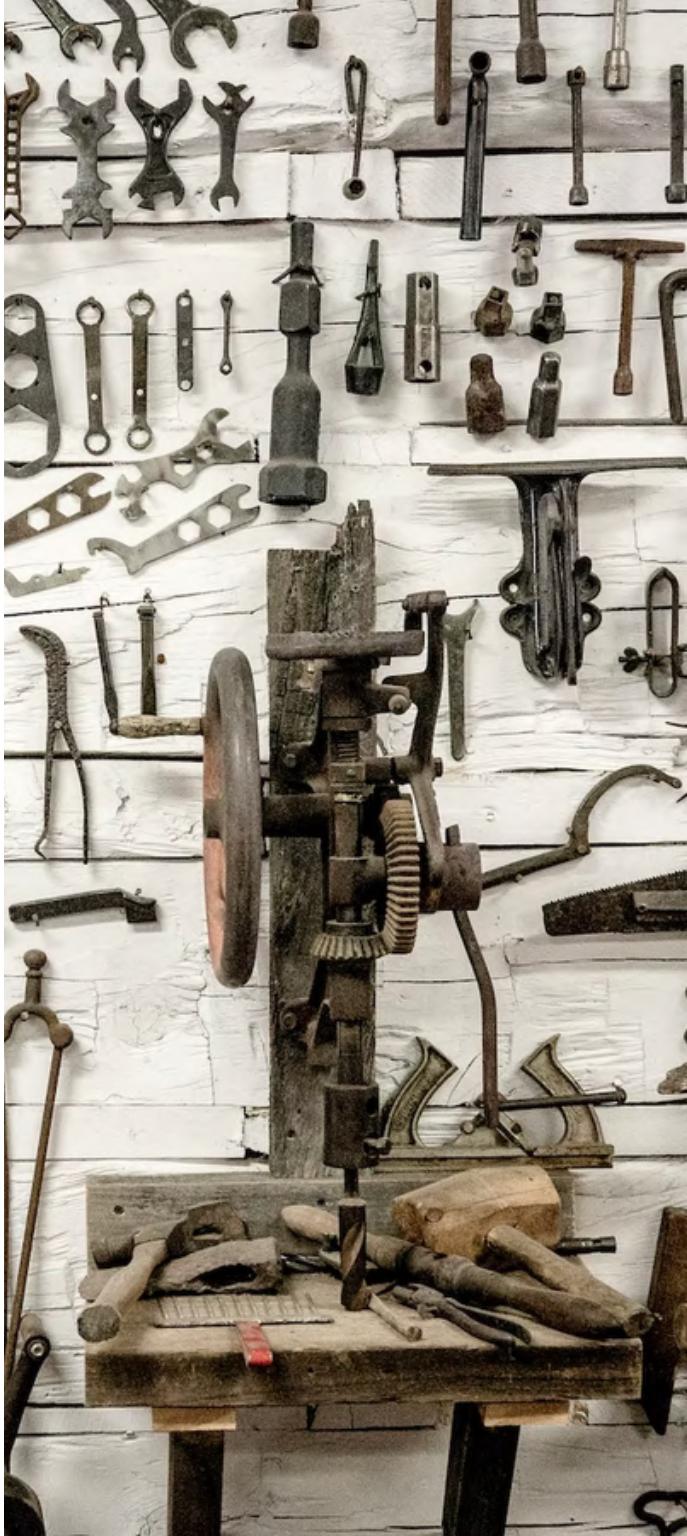
Choose the correct tool for the job

Choose the platform you're going to work on wisely

Apps are appreciated

An app is not required, but appreciated. Make sure the app is clean, intuitive and works on other devices as well!

Have fun!



TIPS AND TRICKS

Tooling

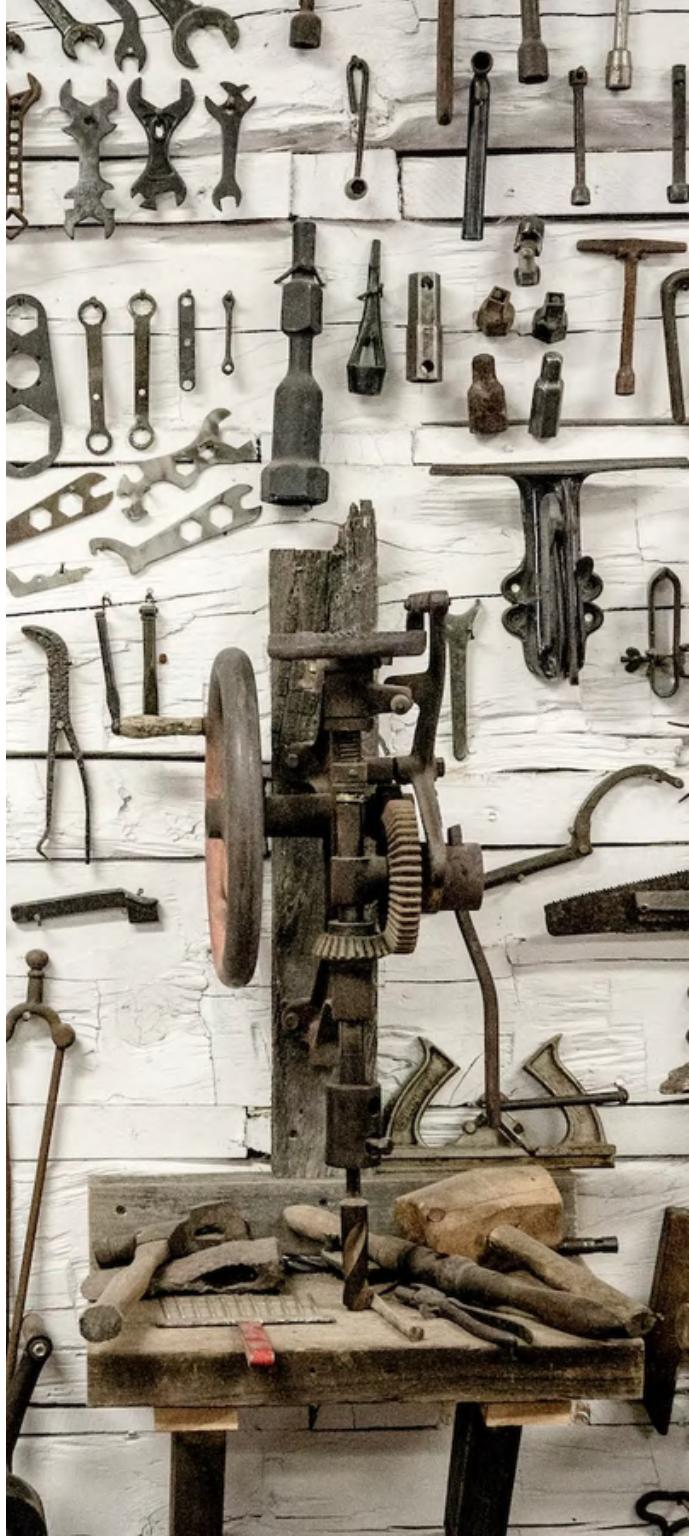
Choose the correct tool for the job

Choose the platform you're going to work on wisely

Apps are appreciated

Have fun!

Remember that this is not graded! Communication is key!
... and most importantly, have fun!



Good Luck!

Made possible by