

A day in the life of a Data Scientist

Gabriela de Queiroz

Sr. Machine Learning Manager, IBM

 [@gdequeiroz](https://twitter.com/gdequeiroz) |  linktr.ee/gdq



Gabriela de Queiroz

Sr. Machine Learning Manager, IBM

- Founder of **R-Ladies** (rladies.org)
- Founder of **AI Inclusive** (ai-inclusive.org)
- Member of **R Foundation** (r-project.org)

- B.S. in Statistics
- MSc. in Epidemiology
- MSc. in Statistics



**Data Scientist + Developer Advocate + Open Source Developer + Manager +
Statistician + Epidemiologist + Community Builder + Mentor + Speaker + Educator**



R-Ladies

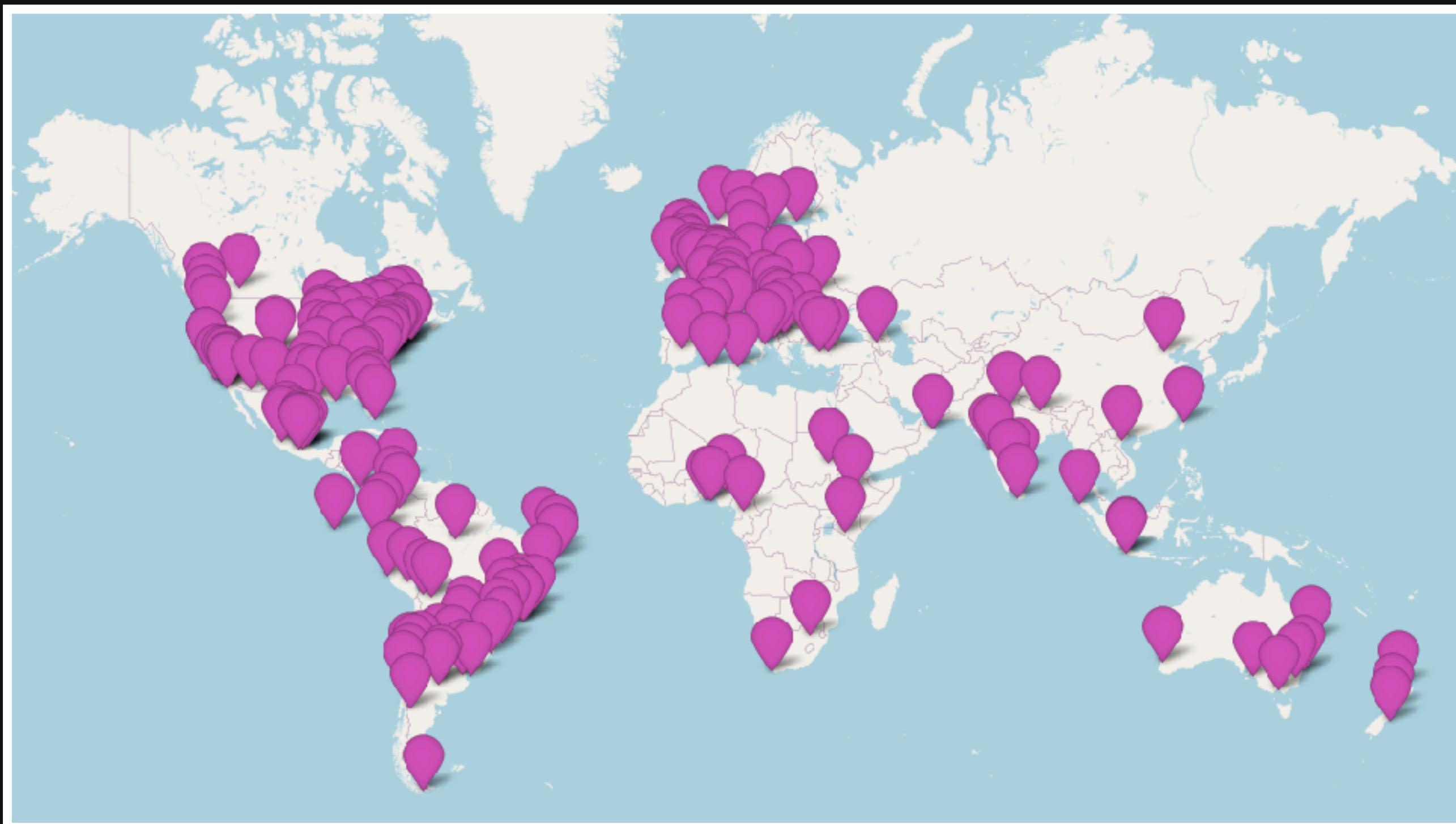
rladies.org

Worldwide organization that promotes

diversity in the R community via meetups

and mentorship in a friendly and safe

environment





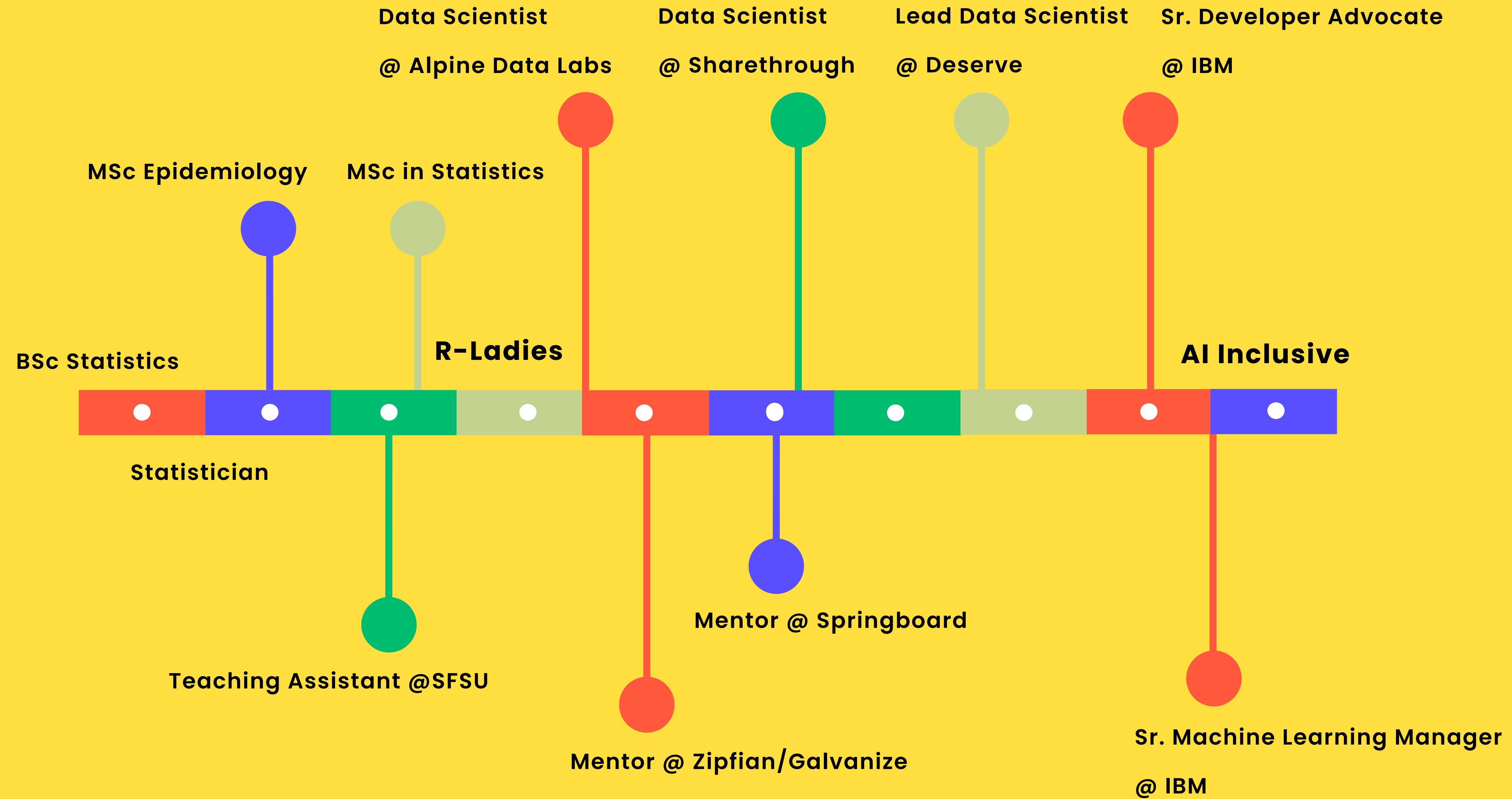
AI Inclusive

Mission: Increase the representation and participation of minority groups in Artificial Intelligence

Together, we are building a community to make **AI** more **inclusive** to everyone.

- Website: ai-inclusive.org
- Twitter: bit.ly/ai-inclusive-twitter
- Facebook: bit.ly/ai-inclusive-facebook
- Instagram: bit.ly/ai-inclusive-instagram
- Youtube: bit.ly/ai-inclusive-youtube





The Data Science Career



Find jobs Company reviews Find salaries

Upload your resume Sign in

What statistician

Where United States

Find jobs

Advanced Job Search

statistician jobs in United States

Sort by: relevance - date

Page 1 of 1,254 jobs



Find jobs Company reviews Find salaries

Upload yo

What

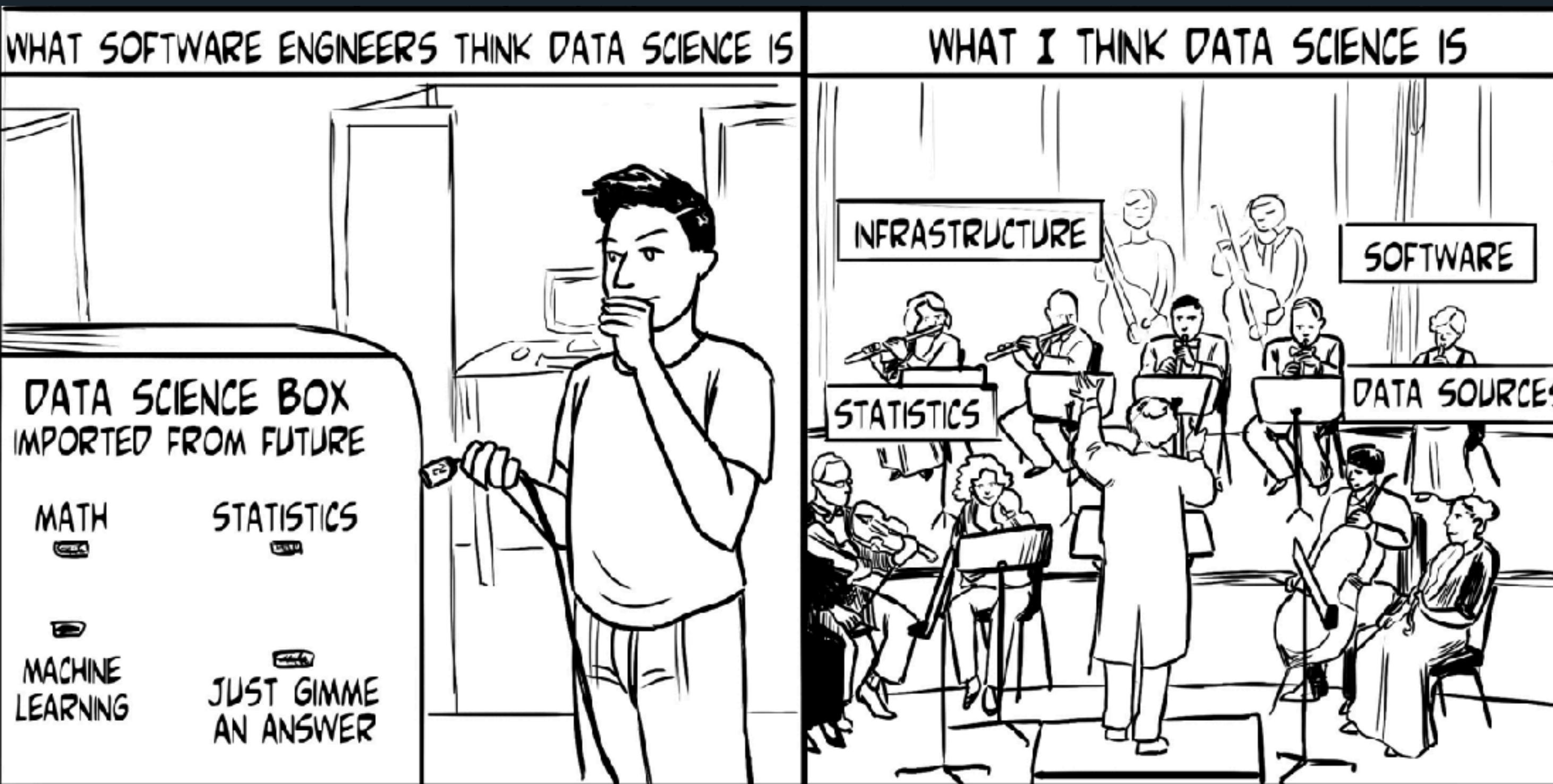
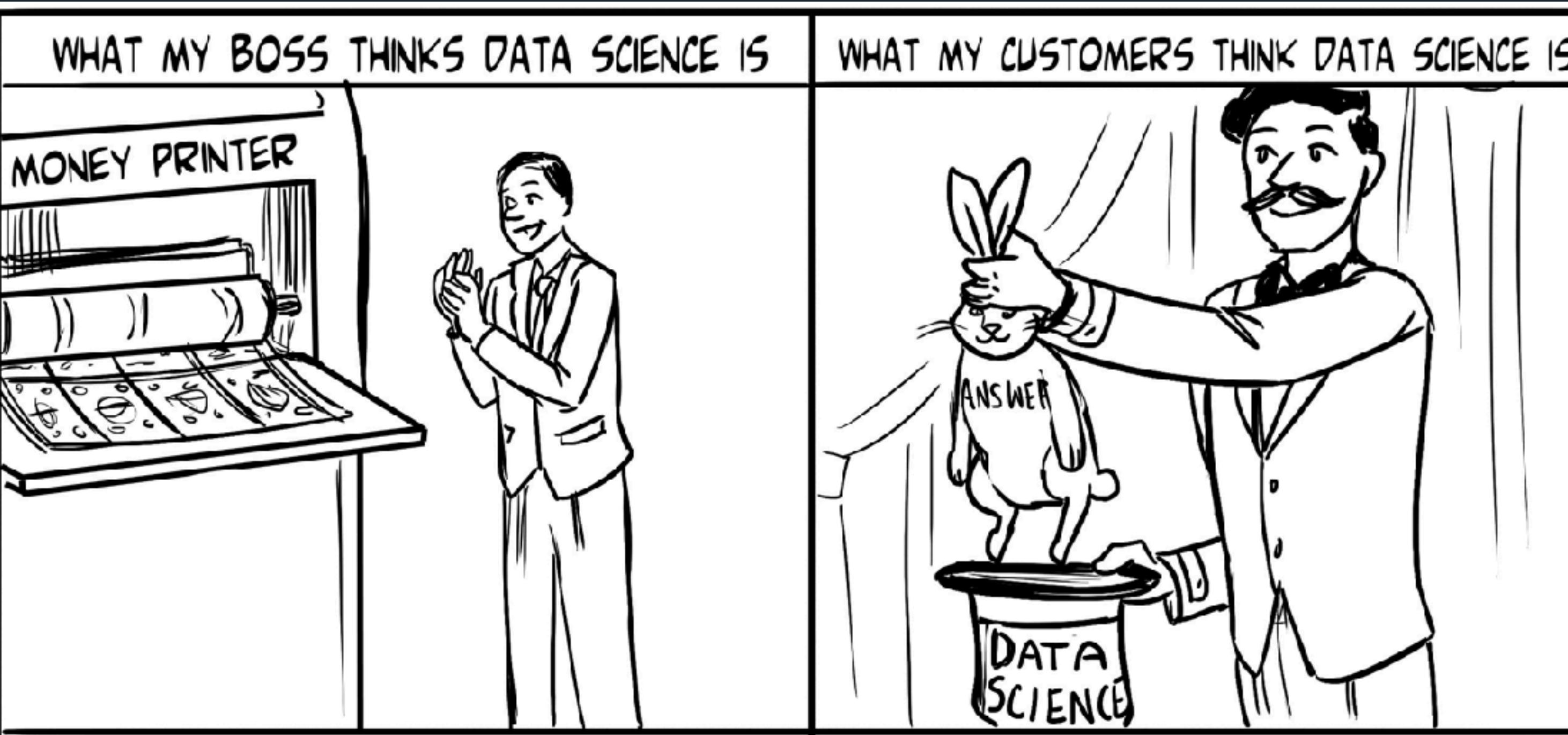
Where United States

Find jobs

data scientist jobs in United States

Sort by: relevance - date

Page 1 of 7,483 jobs



The background of the image features a repeating pattern of cowboy hats arranged in a grid on a light-colored, possibly cream or beige, wall. The hats are a light tan or beige color with dark brown bands. They are positioned in a staggered grid, creating a sense of depth.

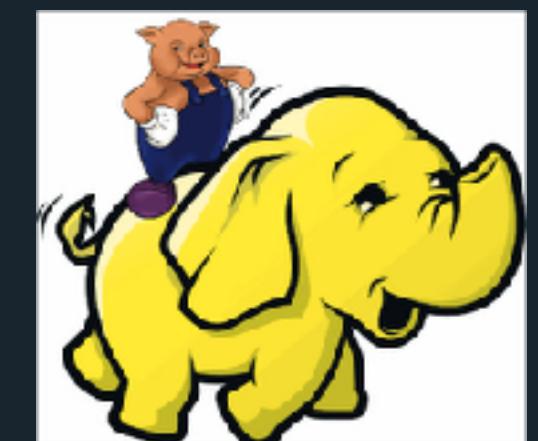
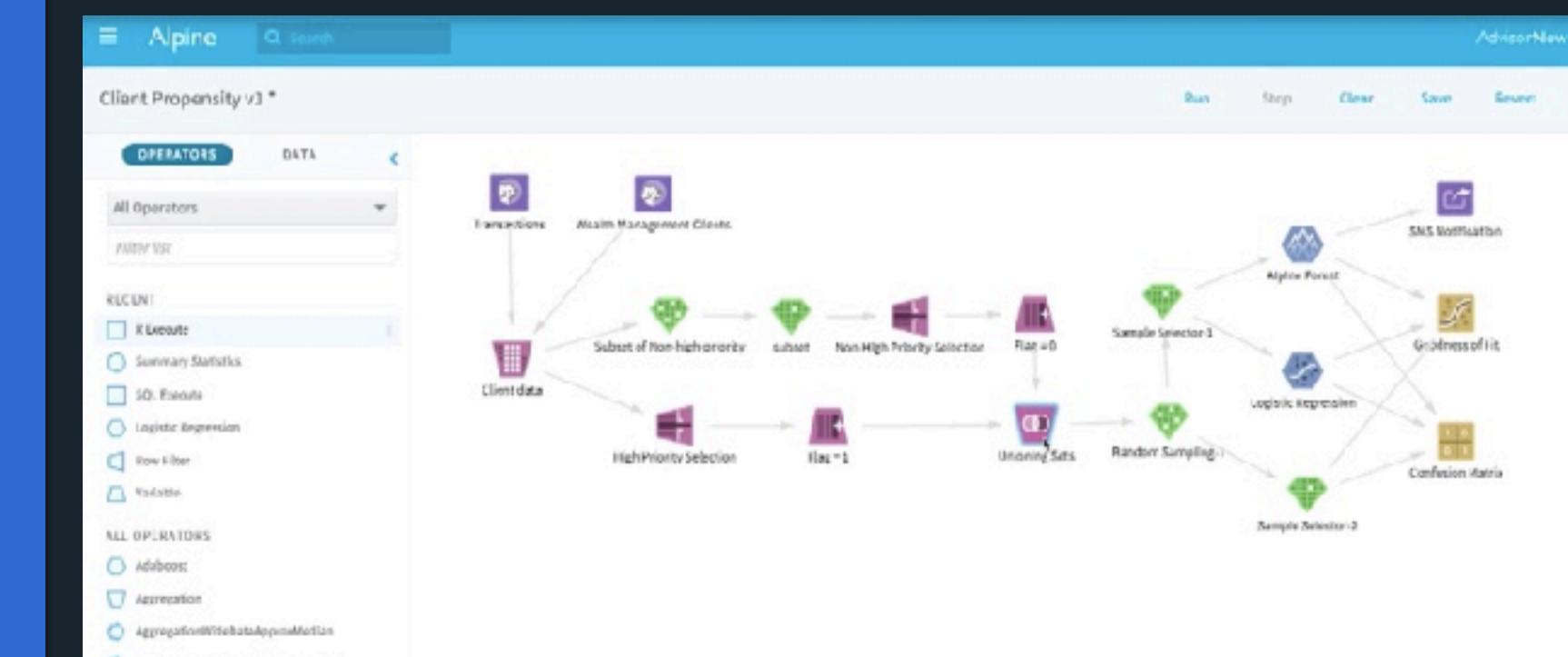
Data Scientists are adaptable
professionals

Different Companies, Different Roles

- ## Key Responsibilities:
- 1) Consultant
 - 2) Train Customers
 - 3) Prioritization
 - 4) Documentation



40 Employees
10 Engineers
5 Data Scientists



Different Companies, Different Roles

sharethrough

Key Responsibilities:

- 1) Infrastructure
- 2) Models/Optimization
- 3) Experimental Design
- 4) Teach



150 Employees

20 Engineers

1 Data Scientist

Spark SQL

Spark Scala

pyspark

MySQL

python

databricks

R

R Studio

jupyter

Different Companies, Different Roles

deserve

Key Responsibilities:

- 1) Infrastructure
- 2) Dashboard
- 3) Marketing
- 4) ML Models

25 Employees

5 Engineers

2 Data Scientists

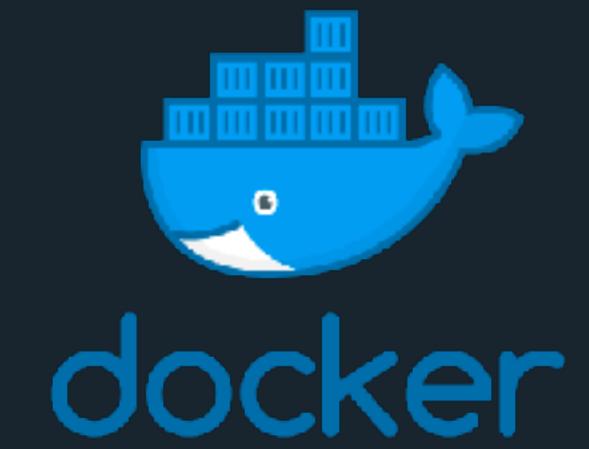


Different Companies, Different Roles



Key Responsibilities:
Several!

300.000 Employees
? Engineers
? Data Scientists



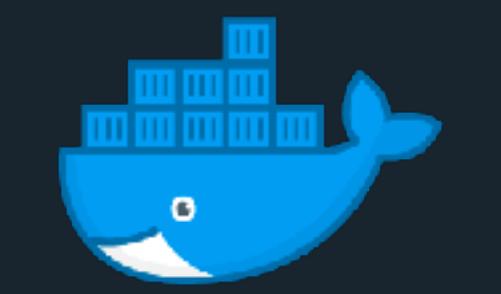
IBM Senior Open Source Developer - DL/ML/AI Developer and Advocate

Key Responsibilities:

- 1) Write great code in key open source communities
- 2) Democratize AI by building tools, launching new open source projects, and improving existing core open source frameworks
- 3) Improve the productivity of developers and domain experts by expanding usage of AI in industries
- 4) Gain eminence in the community by socializing your work, and speaking at open source developer and IBM events.
- 5) Work with offering managers and product teams on applications
- 6) Guide and mentor clients to become self-sufficient open source developers
- 7) Be authentic; mentor others, and be open to mentoring by others (lifelong learning)
- 8) Read and comment on more code than you write; fix bugs, test cases, and documentation

CODAIT

~ 40 open source
developers



docker



Sr. Machine Learning Manager

Responsibilities:
(MANY!)

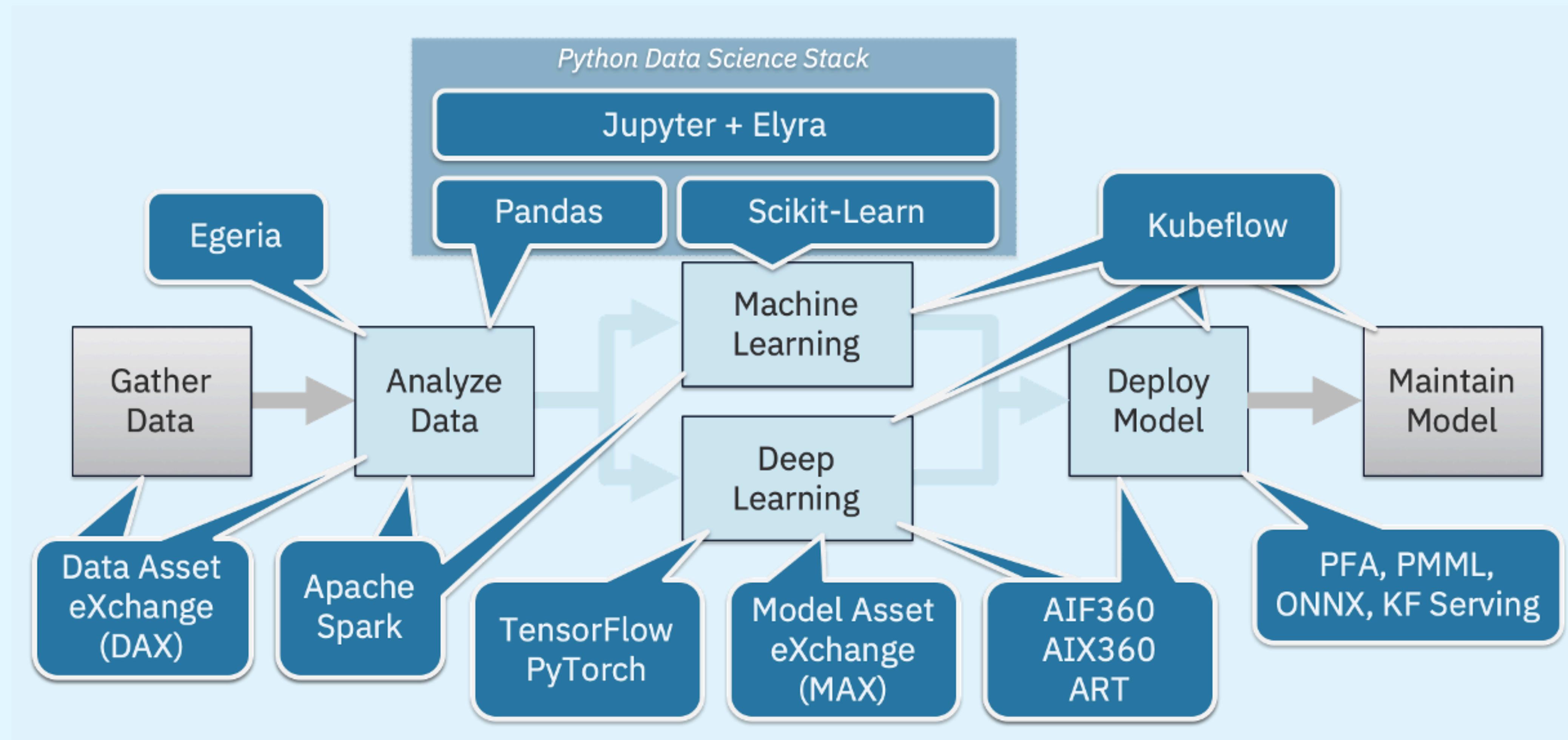
Mon 3/23	Tue 3/24	Wed 3/25	Thu 3/26	Fri 3/27
9 – 10 Meeting with X Team	9 – 10 Meeting with X Team	9 – 10 Meeting with Z Team	9 – 10 Meeting with X Team	9 – 10 Sprint Retrospective
10 – 11 Project Planning Session	10 – 11 Press Interview	10 – 11 Project Planning Session	10 – 11 Corporate Event update	10 – 11 Project Planning session
11 – 12p Call with East Coast office	11 – 12:30p Presentation to executive team	11 – 12p Coaching session	11 – 12p Presentation to Mobile team	11 – 12:30p Training session
12p – 1p Management Luncheon	12:30p – 1:30p Call with Tony	12p – 1p Lunch with CEO	12p – 1p Team Lunch	12:30p – 1:30p Conference Call
1p – 2p 1 on 1 with Tony	1:30p – 3p White Board session with Antoine	1p – 2p Project Z Meeting	1p – 2p 1 on 1 with Chris	1:30p – 2:30p Project Z Meeting
2p – 3p Review session	2p – 3p 1 on 1 with Amanda	2p – 3p 1 on 1 with Tara	2:30p – 3:30p Check in with Rich	2:30p – 3:30p Senior leadership checkin
3p – 5p Leadership team planning session	3p – 4p Meeting with Design team	3p – 5p Company All Hands	3p – 4p Meeting with engineering	3:30p – 5p New project Kickoff
	4p – 5p HR Mandated training			

ML TEAM

10 open source developers

We build tools to make AI accessible and available to everybody

(codait.org)



What is open source???



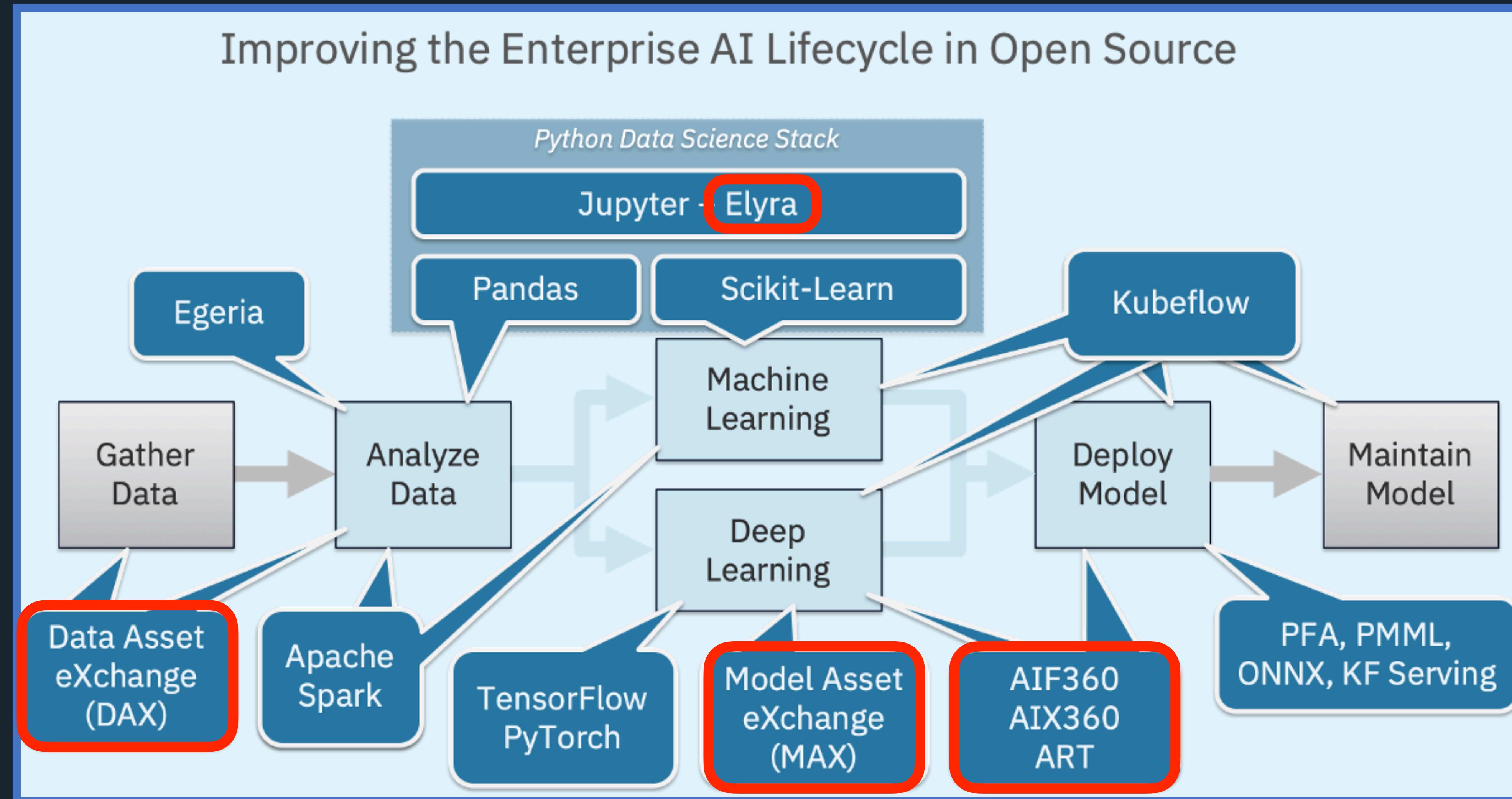
What is open source?

- The term **open source** refers to something people can modify and share because its design is publicly accessible
- Open source software (OSS) is software with source code that anyone can inspect, modify, and enhance

Some Projects



Center for Open Source Data and AI Technologies (CODAIT)



Machine Learning Team

Model Asset eXchange

Model Asset eXchange

Free, deployable, and trainable code. A place for developers to find and use free and open source deep learning models.

Try the tutorial →
Join the community →

Featured Deployable Trainable

Model Deployable Toxic Comment Classifier Detect 6 types of toxicity in user comments Jun 04, 2019 →	Model Deployable, Trainable Text Sentiment Classifier Detect the sentiment captured in short pieces of text Mar 29, 2019 →	Model Deployable, Trainable Image Segmenter Identify objects in an image, additionally assigning each pixel of the image to a particular object. Sep 21, 2018 →
Model Deployable, Trainable Object Detector Localize and identify multiple objects in a single image. Sep 21, 2018 →	Model Deployable Audio Classifier Identify sounds in short audio clips. Sep 21, 2018 →	Model Deployable Image Caption Generator Generate captions that describe the contents of images. Sep 21, 2018 →

Data Asset eXchange

Learn More →

Data Asset eXchange

Explore useful and relevant data sets for enterprise data science

What's New →
Get Involved →

Dataset CSV NOAA Weather Data - JFK Airport June 30, 2020 →	Dataset IOB format Groningen Meaning Bank - Modified May 14, 2020 →	Dataset CSV Fashion-MNIST September 12, 2015 →
Dataset JPG, JSON PubLayNet October 25, 2019 →	Dataset WAV TensorFlow Speech Commands March 17, 2020 →	Dataset PNG, JSON PubTabNet July 20, 2020 →

ibm.biz/model-exchange

ibm.biz/data-exchange

Model Asset eXchange

Place for developers/data scientists to find and use
free and **open source** deep learning models

ibm.biz/model-exchange

Model Asset eXchange

[Try the tutorial](#)



[Join the community](#)



Free, deployable, and trainable code. A place for developers to find and use free and open source deep learning models.

[Featured](#) [Deployable](#) [Trainable](#)

Model | Deployable

Toxic Comment Classifier

Detect 6 types of toxicity in user comments

Jun 04, 2019

Model | Deployable, Trainable

Text Sentiment Classifier

Detect the sentiment captured in short pieces of text

Mar 29, 2019

Model | Deployable, Trainable

Image SegmenTer

Identify objects in an image, additionally assigning each pixel of the image to a particular object.

Sep 21, 2018

Model | Deployable, Trainable

Object Detector

Localize and identify multiple objects in a single image.

Sep 21, 2018

Model | Deployable

Audio Classifier

Identify sounds in short audio clips.

Sep 21, 2018

Model | Deployable

Image Caption Generator

Generate captions that describe the contents of images.

Sep 21, 2018

[View all models](#)

Model Asset eXchange

30+ ready to use deep learning models

- Wide variety of domains (text, audio, image, etc)
- Multiple deep learning frameworks (TensorFlow, PyTorch, Keras)
- Trainable and Deployable versions

ibm.biz/model-exchange

Model Asset eXchange

Free, deployable, and trainable code. A place for developers to find and use free and open source deep learning models.

Try the tutorial →

Join the community →

[Featured](#) [Deployable](#) [Trainable](#)

Model | Deployable

Toxic Comment Classifier

Detect 6 types of toxicity in user comments

Jun 04, 2019



Model | Deployable, Trainable

Text Sentiment Classifier

Detect the sentiment captured in short pieces of text

Mar 29, 2019



Model | Deployable, Trainable

Image SegmenTer

Identify objects in an image, additionally assigning each pixel of the image to a particular object.

Sep 21, 2018



Model | Deployable, Trainable

Object Detector

Localize and identify multiple objects in a single image.

Sep 21, 2018



Model | Deployable

Audio Classifier

Identify sounds in short audio clips.

Sep 21, 2018



Model | Deployable

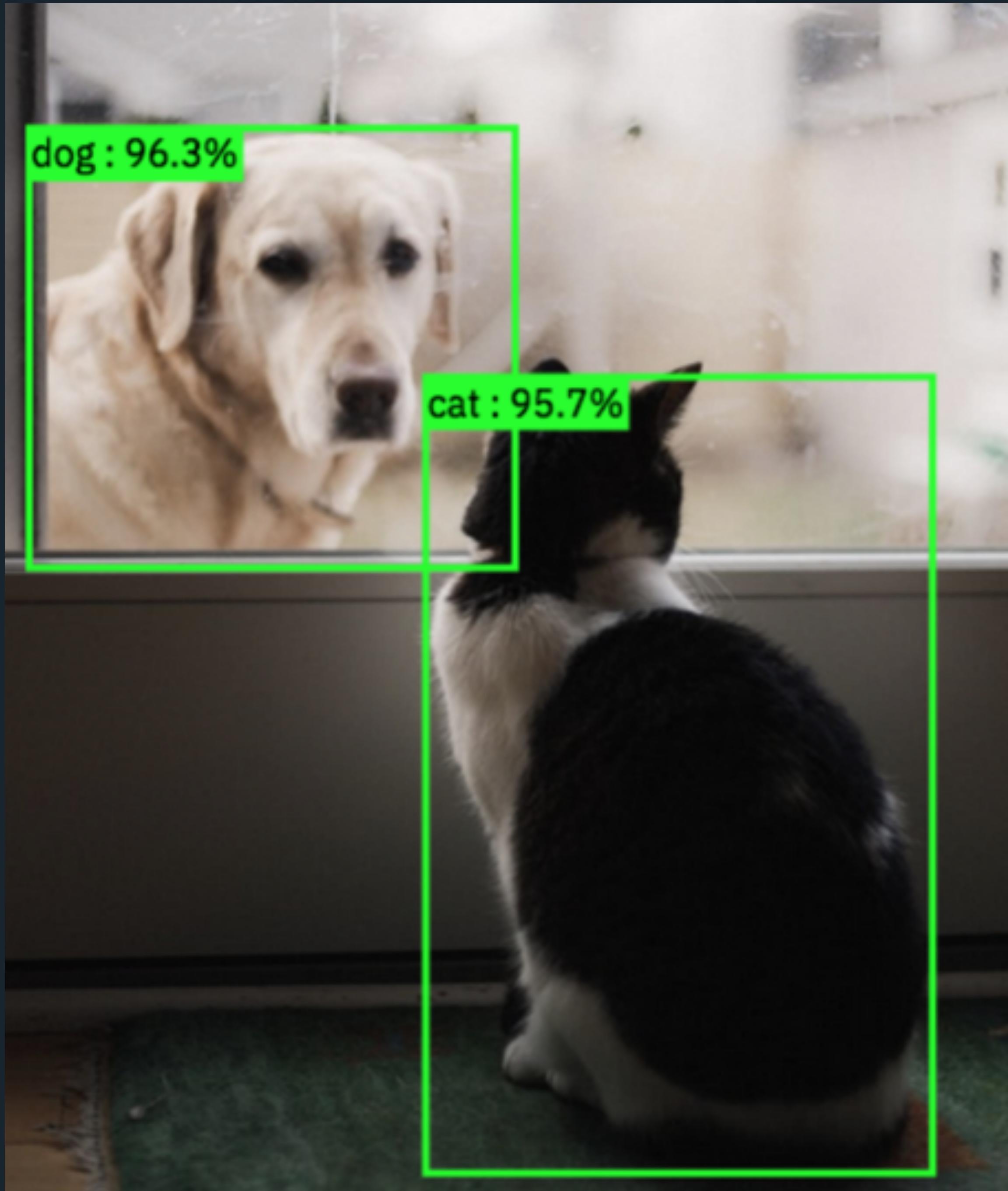
Image Caption Generator

Generate captions that describe the contents of images.

Sep 21, 2018



[View all models](#)



OBJECT DETECTOR

Localize and identify multiple objects in a single image



Signup for IBM Cloud: ibm.biz/gguniversity

Model Asset eXchange

Free, deployable, and trainable code. A place for developers to find and use free and open source deep learning models.

Try the tutorial →
Join the community →

Featured Deployable Trainable

Model | Deployable
Toxic Comment Classifier
Detect 6 types of toxicity in user comments
Jun 04, 2019 →

Model | Deployable, Trainable
Text Sentiment Classifier
Detect the sentiment captured in short pieces of text
Mar 29, 2019 →

Model | Deployable, Trainable
Image Segmente
Identify objects in an image, additionally assigning each pixel of the image to a particular object.
Mar 29, 2019 →

Model | Deployable, Trainable
Object Detector
Localize and identify multiple objects in a single image.
Sep 21, 2018 →

Model | Deployable
Audio Classifier
Identify sounds in short audio clips.
Sep 21, 2018 →

Model | Deployable
Image Caption Generator
Generate captions that describe the contents of images.
Sep 21, 2018 →

ibm.biz/model-exchange

Model Deployable, Trainable

Object Detector

Localize and identify multiple objects in a single image.

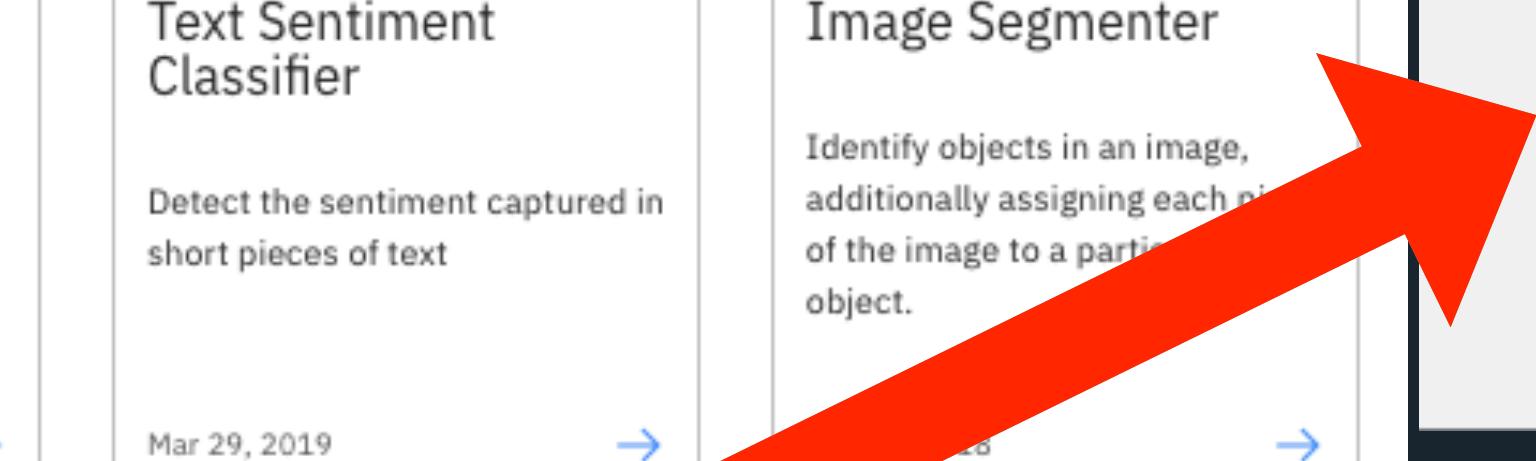
Get this model



Try the API →

Try the web app →

Try in a Node-RED flow →

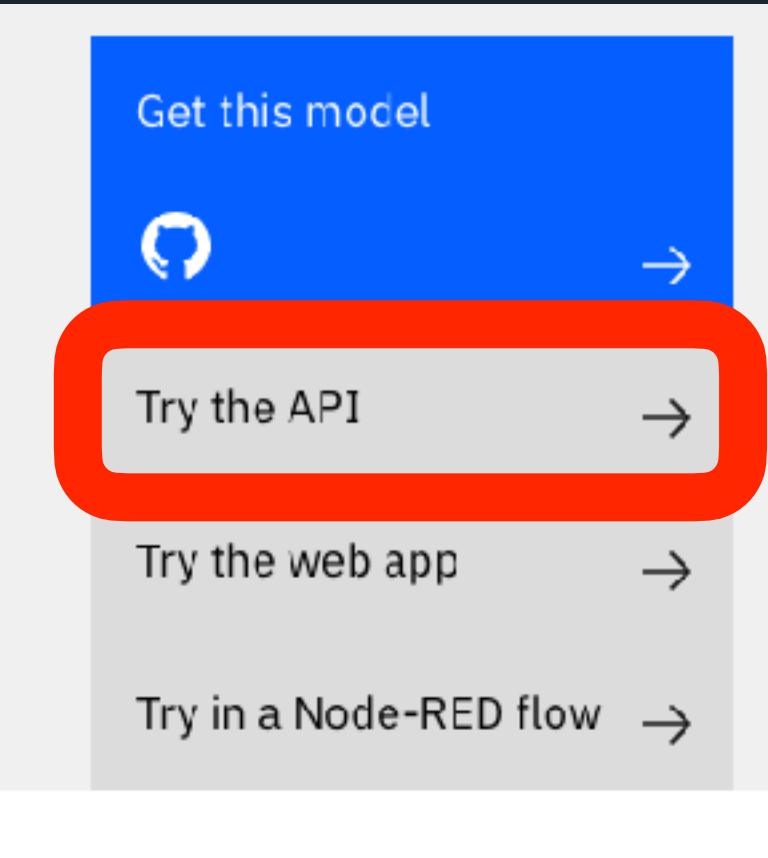


Access the API via Python

Model | Deployable, Trainable

Object Detector

Localize and identify multiple objects in a single image.



response

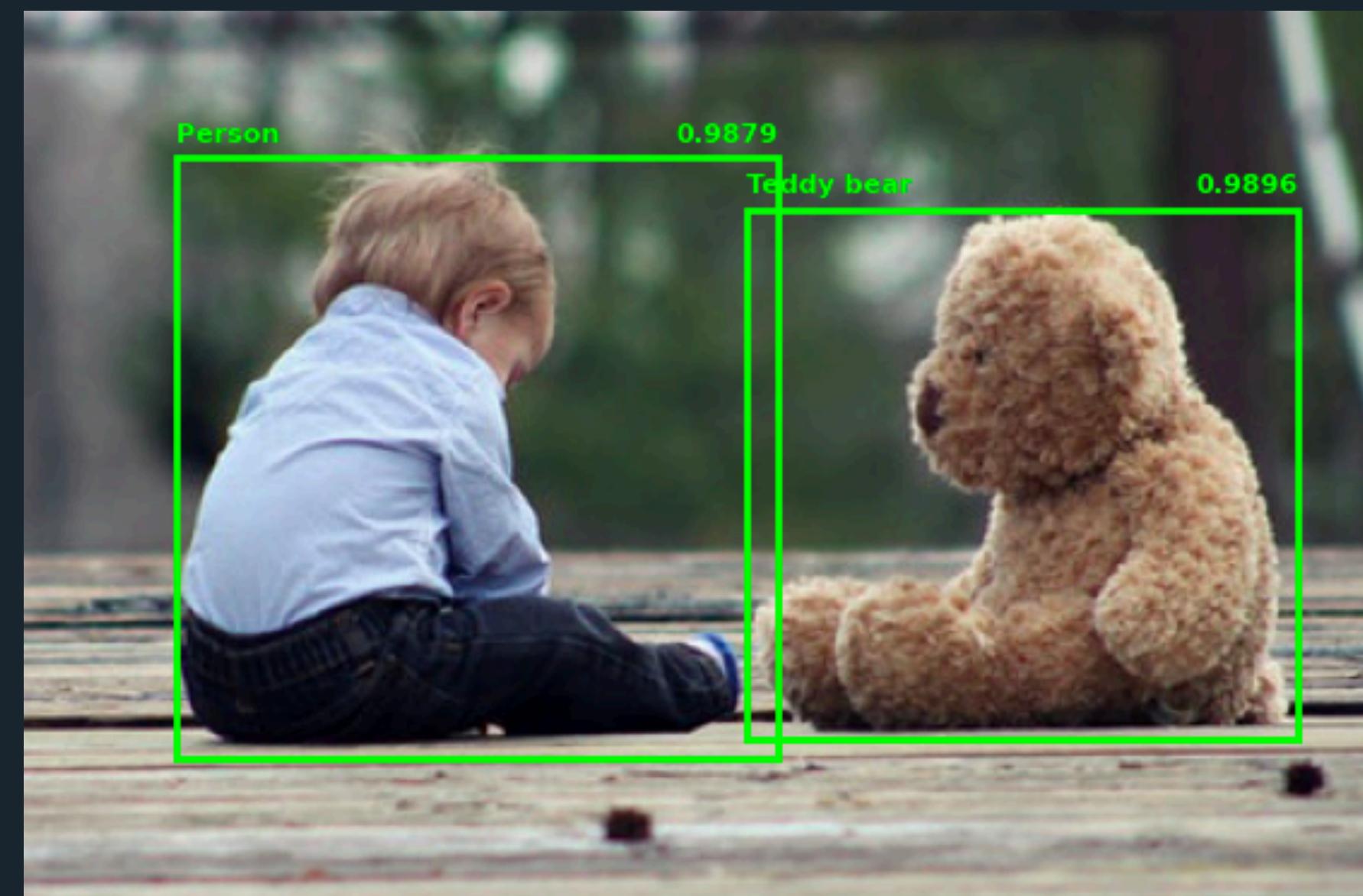
```
{'status': 'ok',
'predictions': [ {'label_id': '88',
'label': 'teddy bear',
'probability': 0.9896332025527954,
'detection_box': [0.27832502126693726,
0.5611844062805176,
0.643224835395813,
0.8432191610336304]}, {
'label_id': '1',
'label': 'person',
'probability': 0.9879012107849121,
'detection_box': [0.24251867830753326,
0.26926860213279724,
0.655893087387085,
0.5768759250640869]}]}
```

```
# Model
url = 'http://max-object-detector.codait-prod-41208c73af8fca213512856c7a09
db52-0000.us-east.containers.appdomain.cloud/'
model_endpoint = 'model/predict'
complete_url = url + model_endpoint

# Upload an image to the MAX model's rest API
path_to_input_image = 'baby-bear.jpg'

with open(path_to_input_image, 'rb') as file:
    file_form = {'image': (path_to_input_image, file, 'image/jpeg')}
    # Post the image to the rest API using the requests library
    r = requests.post(url=complete_url, files=file_form)
    # Return the JSON
    response = r.json()

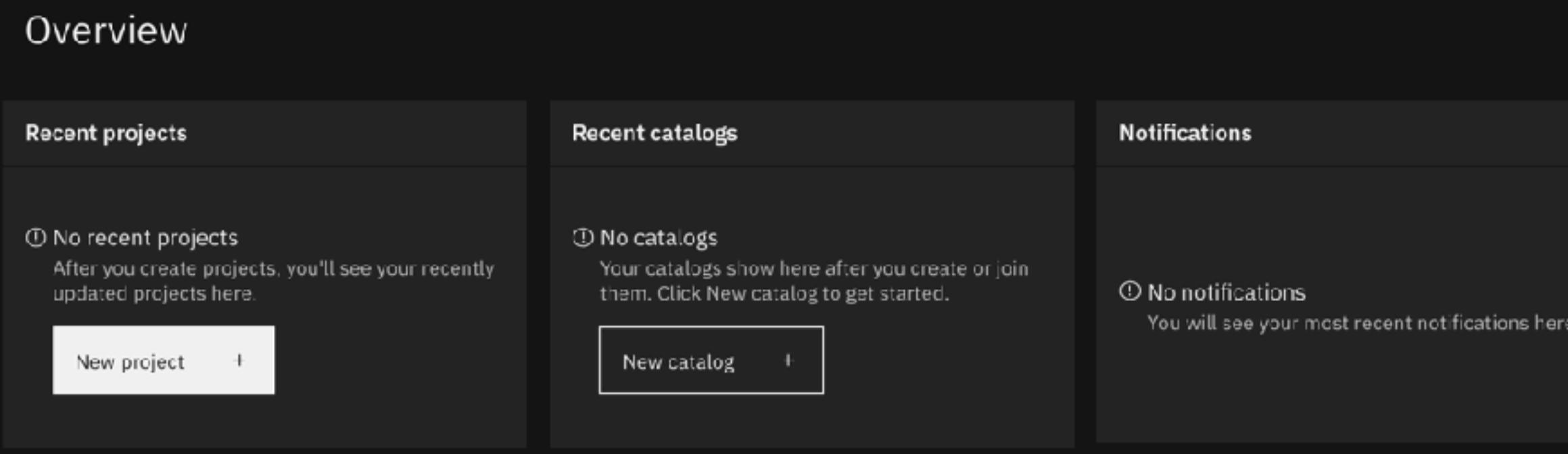
IPython.display.Image(path_to_input_image, width = 450)
```



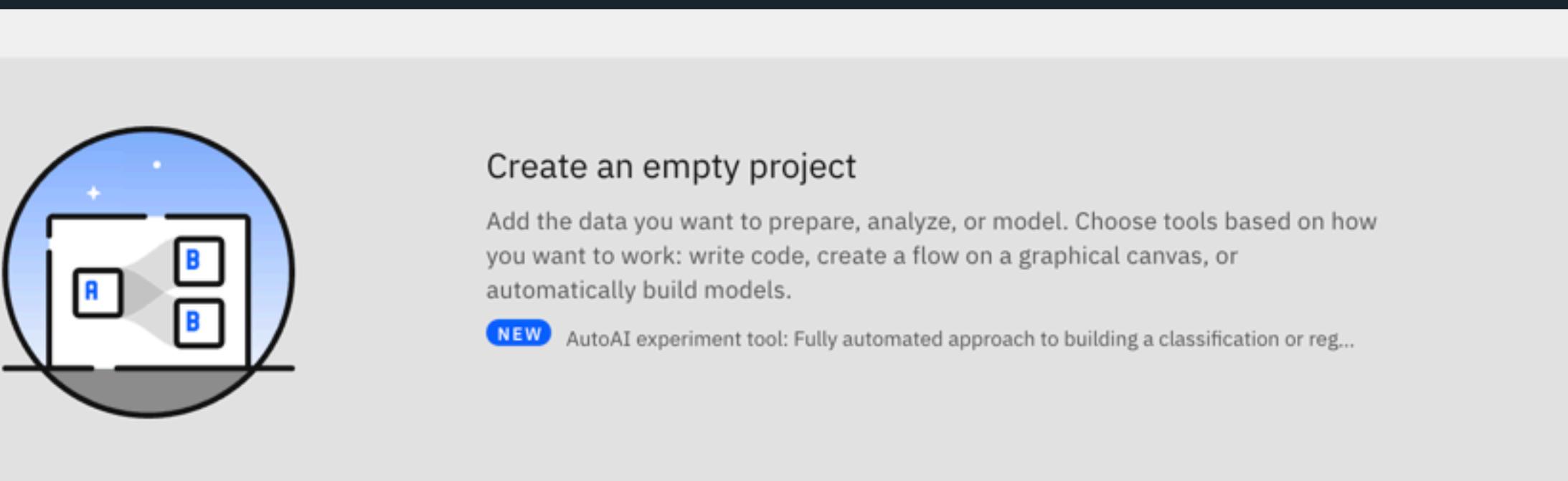
Let's try together ...

1) Go to ibm.biz/gguniversity and create an account

2) Login to your account and create a project



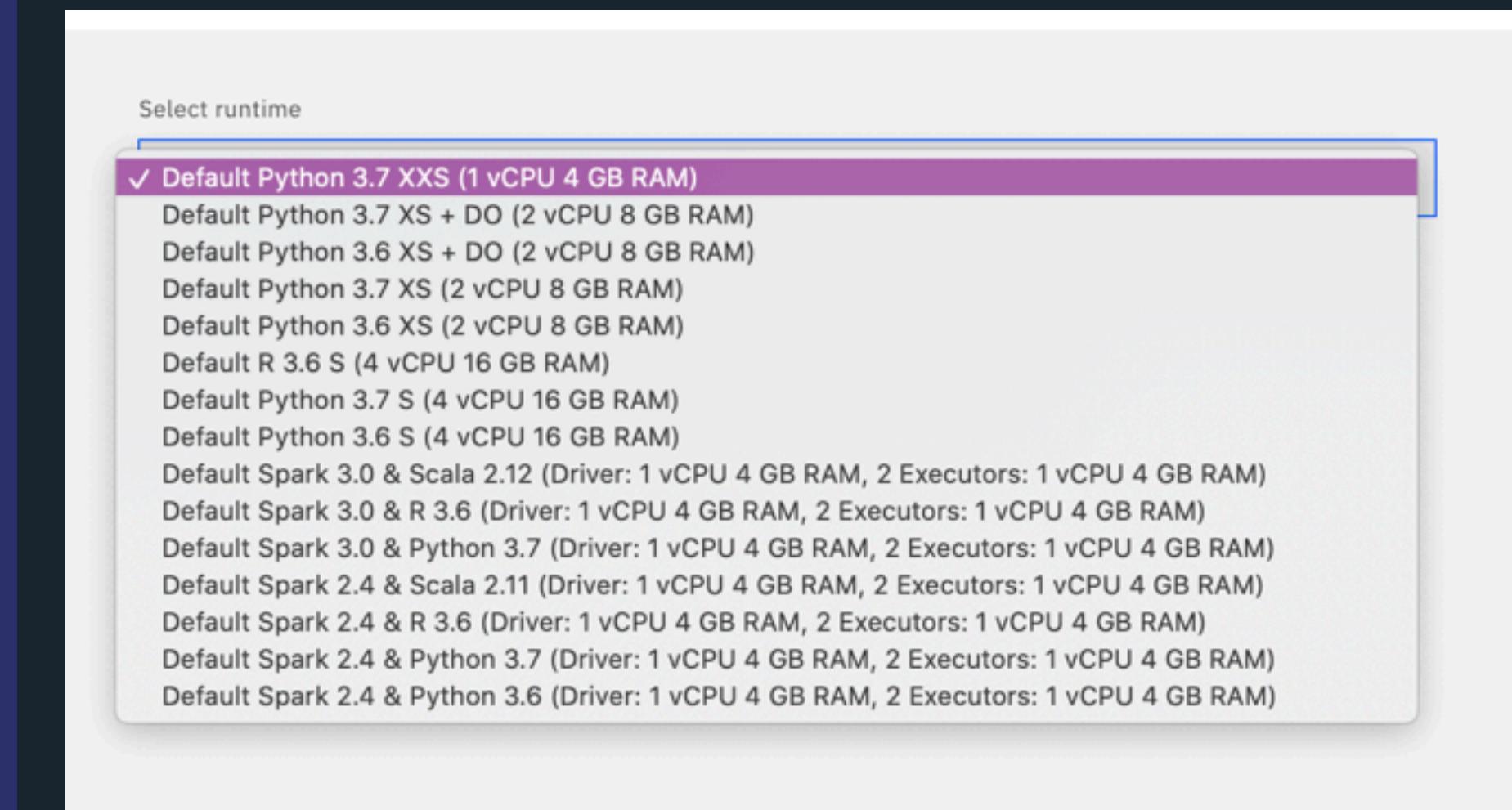
3) Create an *empty project*

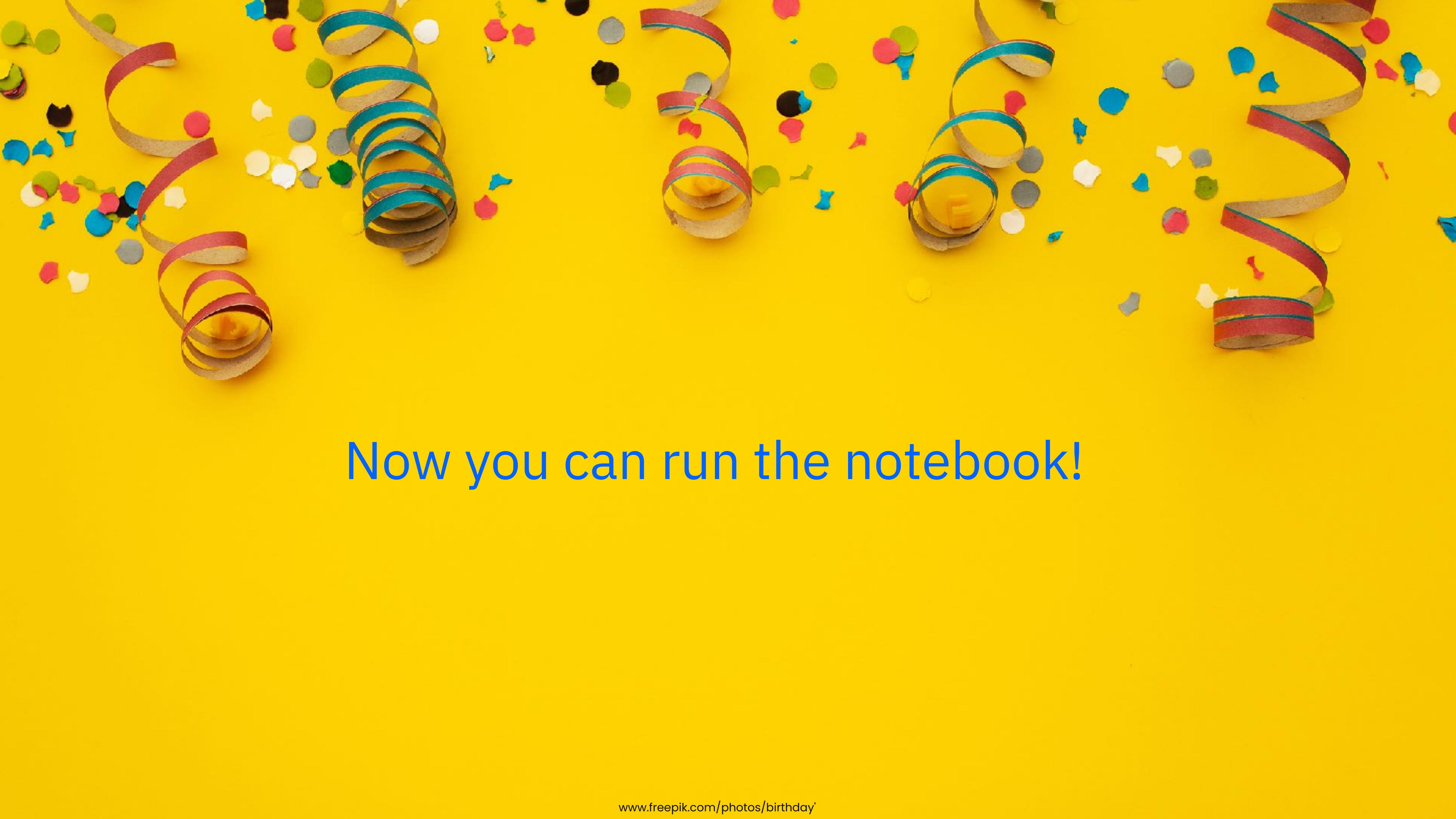


4) Go to the link ibm.biz/max-notebook and copy the project



5) Change the select runtime





Now you can run the notebook!

Access the API via R

The R script is available here: bit.ly/ggu-bnda2020

```
library(dplyr)
library(httr)

# Endpoint
endpoint <- 'http://max-object-detector.codait-prod-41208c73af8fca213512856c7a09db52-0000.us-east.containers.appdomain.cloud/'
# endpoint <- 'http://localhost:5000' # if running docker locally or docker hub

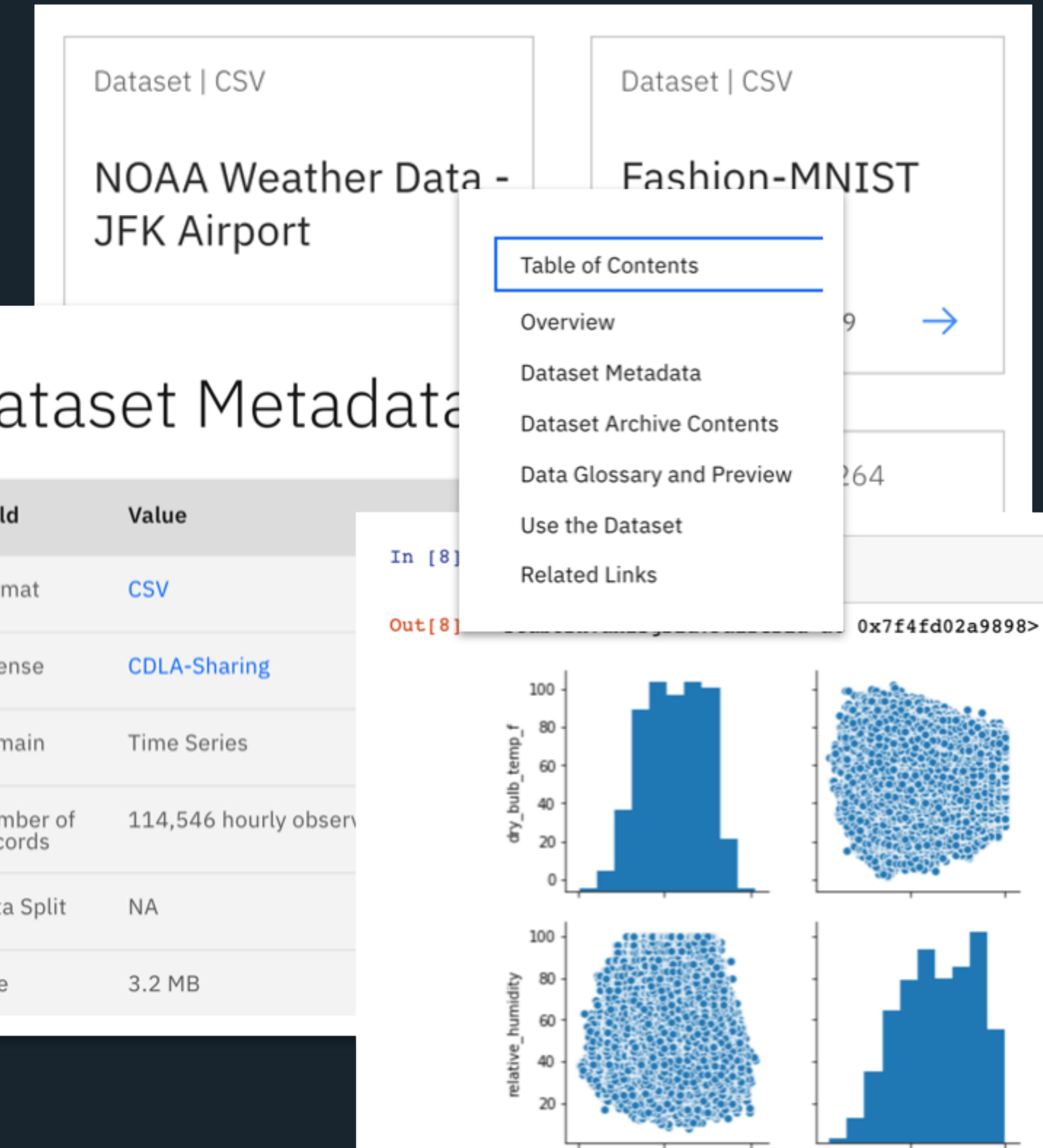
object_detector <- function(path_to_img, endpoint) {
  model_endpoint <- paste0(endpoint, 'model/predict') # Model endpoint
  # POST
  response <- httr::POST(url = model_endpoint,
                          body = list(image = upload_file(path_to_img,
                                                          type = "image/jpeg")),
                          encode = c("multipart"))
  ) %>% content()
  response$predictions
}

# Get the image file from GH
download.file(url = "http://github.com/IBM/MAX-Object-Detector/blob/master/samples/baby-bear.jpg?raw=true",
              'baby-bear.jpg', mode = 'wb')

object_detector("baby-bear.jpg", endpoint)
```

Data Asset eXchange (DAX)

- Curated repository for **open** datasets from IBM Research and third-parties
- Published under data friendly licenses
- Standardized dataset formats and metadata
- Many data sets include starter notebooks (cleansing, data exploration, analysis)



ibm.biz/data-exchange

Data Asset eXchange (DAX)

MIT | CSV

Fashion-MNIST

A dataset of standardized images of fashion items from 10 classes

☆ Save

Like

- Get this dataset →
- Run dataset notebooks →
- Preview the data & notebooks →



By Zalando

Updated August 17, 2020 | Published August 22, 2019

Overview

The Fashion-MNIST dataset contains 60,000 training images (and 10,000 test images) of fashion and clothing items, taken from 10 classes. Each image is a standardized 28×28 size in grayscale (784 total pixels). Fashion-MNIST was created by Zalando as a compatible replacement for the original MNIST dataset of handwritten digits. This version of the dataset has been converted to CSV to enable easier loading in common data science tools (see <https://www.kaggle.com/zalando-research/fashionmnist>).

Legend ⓘ

Categories

- Artificial intelligence
- Classification
- DAX - Data Asset eXchange
- Feature Extraction
- Image

Table of Contents

ibm.biz/data-exchange

Dataset Metadata

BONUS

Overview

Recent projects

① No recent projects

After you create projects, you'll see your recently updated projects here.

New project +

Deployment spaces

① No deployment spaces

After you create spaces, you'll see them here.

New deployment space +

Recent catalogs

① No catalogs

Your catalogs show here after you create or join them. Click New catalog to get started.

New catalog +

Notifications

① No notifications

You will see your most recent notifications here.

New in gallery

NOTEBOOK

Dialog skill analysis for Watson Assistant

AUTHOR
IBM

MODIFIED
Oct 27, 2020

Communications

Explore

Gallery ①

All filters 🔍

Search

Featured

NOTEBOOK

Deploying a Decision Optimization Model with...

AUTHOR
IBM

MODIFIED
Sep 03, 2020

Watson Machine Learning V4

DATA

MNIST handwritten digits - PKL format

AUTHOR
IBM

MODIFIED
Jan 09, 2020

Science & Technology

DATA

Greenhouse Gas Emissions worldwide

AUTHOR
IBM

MODIFIED
May 22, 2016

Environment

All Content

NOTEBOOK

Dialog skill analysis for Watson Assistant

AUTHOR
IBM

MODIFIED
Oct 27, 2020

Communications

NOTEBOOK

Supply Chain Disruption

AUTHOR
IBM

MODIFIED
Oct 20, 2020

Economy & Business

NOTEBOOK

Explore particulate matter data using IBM...

AUTHOR
IBM

MODIFIED
Oct 19, 2020

All filters

Reset filters

Format

Data

Notebook

Sample Project

Topic

Show me more!



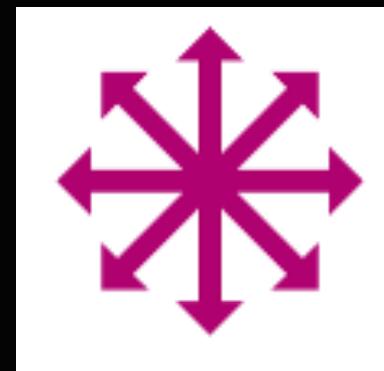
Trusted AI Lifecycle through Open Source

Pillars of trust, woven into the lifecycle of an AI application



IBM and LFAI move forward on
trustworthy and responsible AI
IBM donates Trusted AI toolkits to the Linux Foundation AI

Did anyone tamper
with it?



ROBUSTNESS

Adversarial Robustness 360
↳ (ART)

DEMO: art-demo.mybluemix.net

Is it fair?



FAIRNESS

AI Fairness 360
↳ (AIF360)

DEMO: aif360.mybluemix.net

Is it easy to understand?



EXPLAINABILITY

AI Explainability 360
↳ (AIX360)

DEMO: aix360.mybluemix.net

Is it accountable?



LINEAGE

AI FactSheets 360

DEMO: aifs360.mybluemix.net

Trusted-AI

This GitHub org hosts LF AI Foundation projects in the category of Trusted and Responsible AI.

IBM @LFAI_Foundation info@lfaifoundation.org

Repositories 4 Packages People Projects

Pinned repositories

adversarial-robustness-toolbox
Adversarial Robustness Toolbox (ART) - Python Library for Machine Learning Security - Evasion, Poisoning, Extraction, Inference
Python 1.7k 480

AIF360
A comprehensive set of fairness metrics for datasets and machine learning models, explanations for these metrics, and algorithms to mitigate bias in datasets and models.
Python 1k 340

AIX360
Interpretability and explainability of data and machine learning models
Python 621 136

AI Fairness 360 (AIF360) R Package
CRAN 0.1.0 CRAN 0.1.0

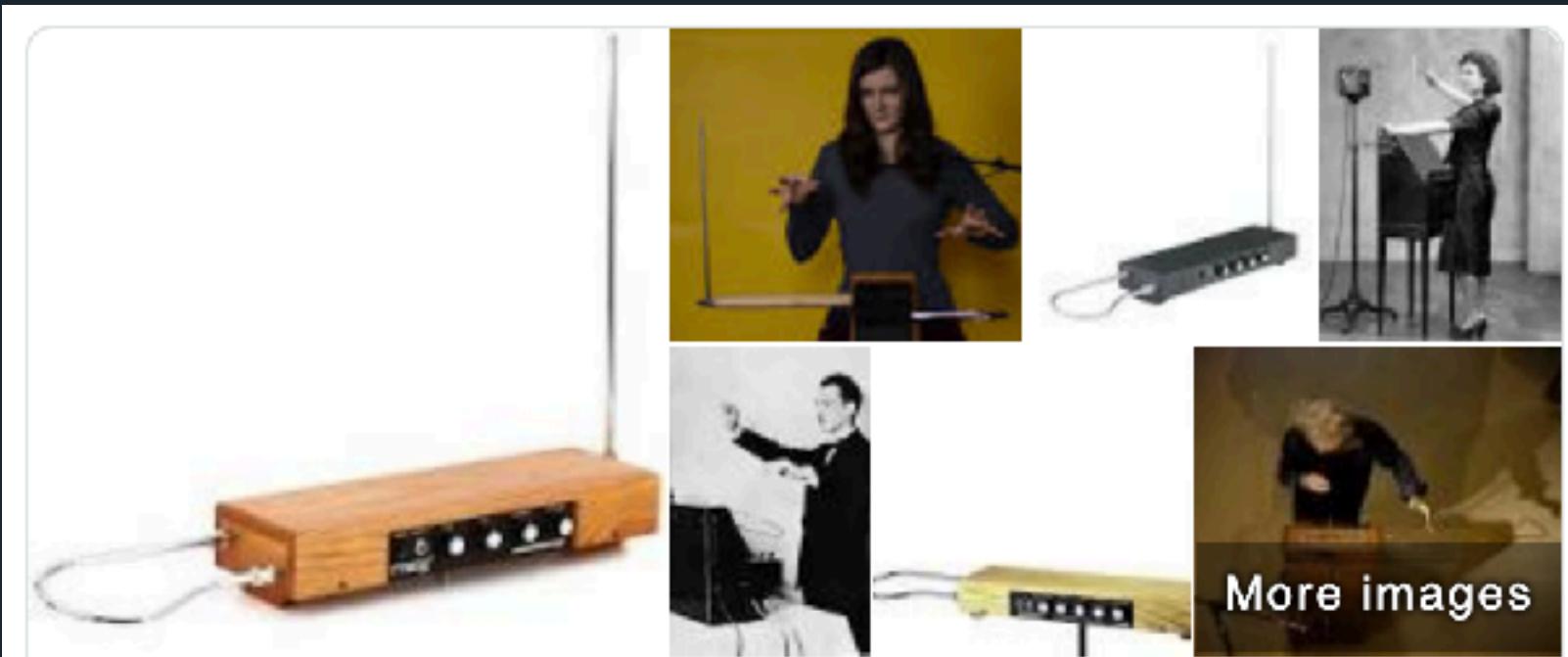
Available in R too!

Any other cool thing?



Use your arms to make music

Create music with your arms using the **Model Asset eXchange (MAX) human pose estimator** model and **TensorFlow**



Theremin

Musical instrument

The theremin is an electronic musical instrument controlled without physical contact by the thereminist. It is named after its inventor, Léon Theremin, who patented the device in 1928. [Wikipedia](#)

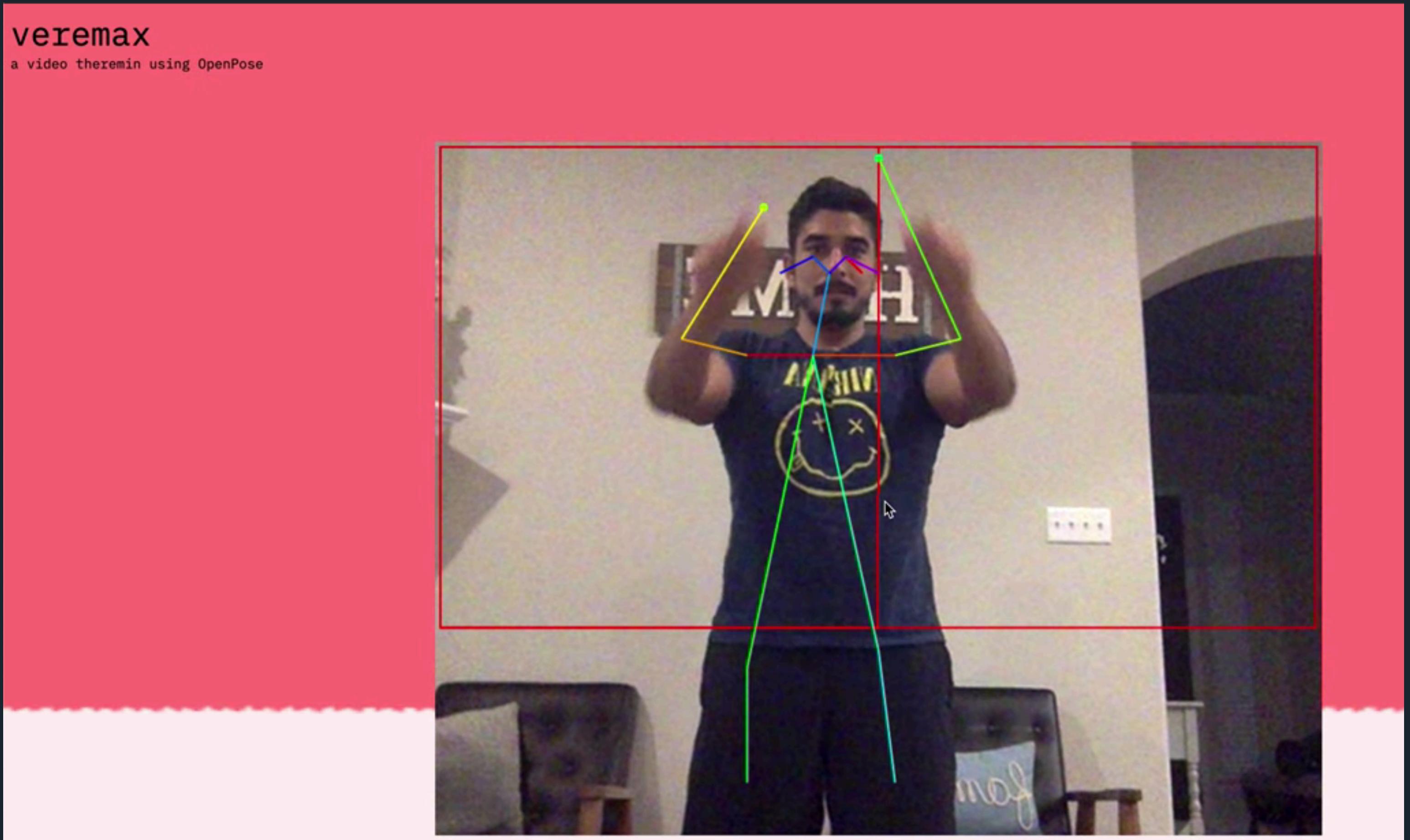
Instrument family: [Electronic Musical Instruments](#), [Musical Keyboards](#)

Invented: 1920

Related instrument: [Ondes Martenot](#), [Electro-Theremin](#)

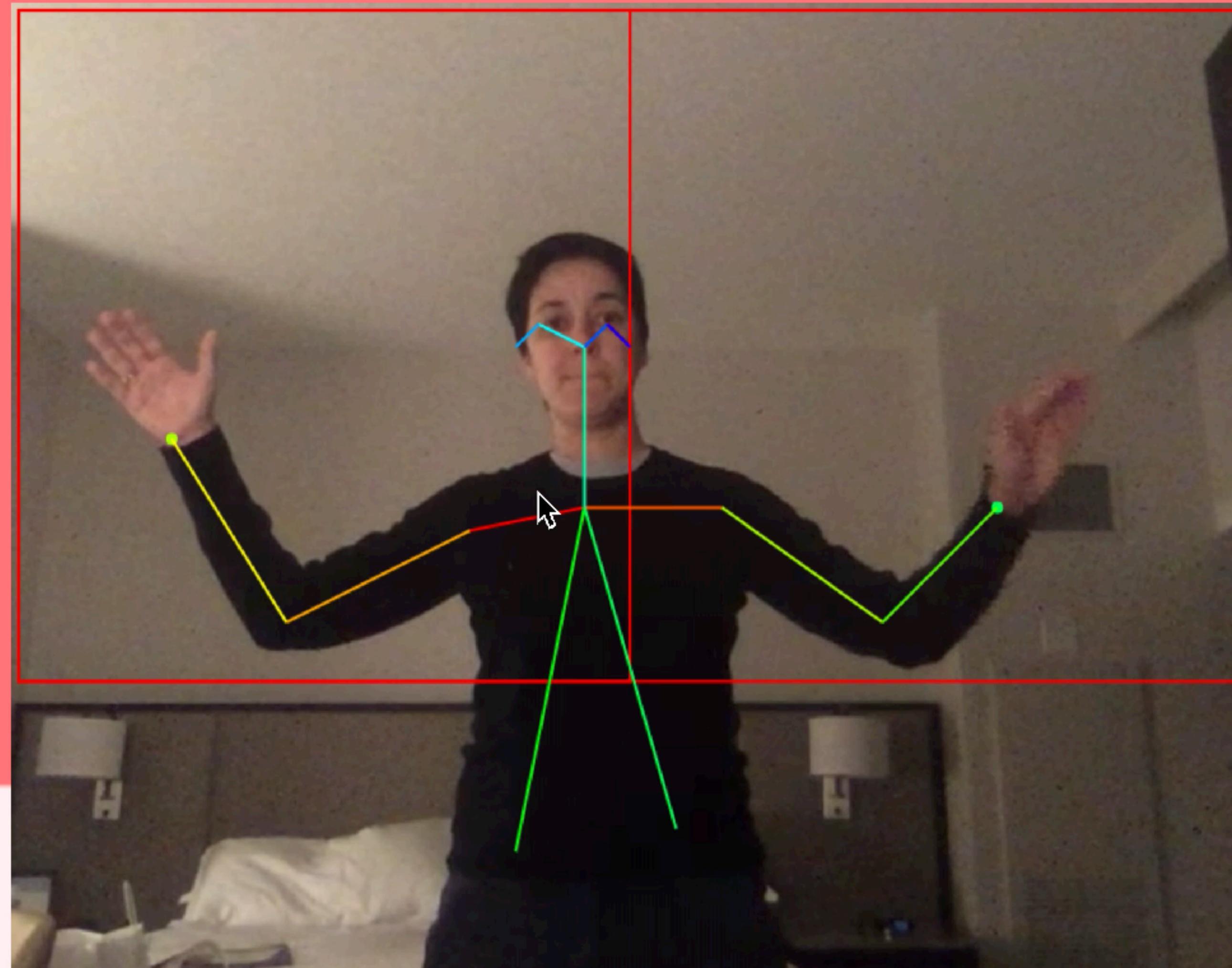
Inventor: [Léon Theremin](#)

Hornbostel-Sachs classification: 531.1; (Electrophone)



veremax

a video theremin using OpenPose



Thank you!



ai-inclusive.org

Follow us:

bit.ly/ai-inclusive-instagram



Resources on AI, DS, ML
Events, Free Tickets and much more

@gdequeiroz | linktr.ee/gdq

