Inteligência Artificial, Ciência de Dados e Estatística: trajetória, dicas e desafios

---

Gabriela de Queiroz

@gdequeiroz | linktr.ee/gdq

slides: bit.ly/cimatech-2020

Gabriela de Queiroz

Sr. Engineering and Data Science Manager, IBM

- Fundadora do R-Ladies (rladies.org)
- Fundadora do AI Inclusive (ai-inclusive.org)

- Graduação em Estatística (UERJ)
- Mestrado em Epidemiologia (ENSP/Fiocruz)
- Mestrado em Estatística (CSUEB)

Data Scientist + Developer Advocate + Open Source Developer + Manager +

Statistician + Epidemiologist + Community Builder + Mentor + Speaker + Educator

Fundada em outubro de 2012.
A ideia era retribuir à comunidade e criar um local onde as pessoas se sentissem confortáveis, seguras e bem-vindas.
Um lugar onde as pessoas pudessem fazer perguntas, aprender juntas e compartilhar.

31
OCT

Wednesday, October 31, 2012

Introduction to R (beginners and pre-beginners)

Hosted by
Gabriela de Queiroz

Details

Hello R-ladies!

The first meetup will take place on October 31st at the Google office in San Francisco.

For this first meetup, we'll do an introduction to R. We'll go over the following topics:

installing R setting up an R environment (RStudio) basic commands (open files, simple dataset manipulation, simple plots, etc) loading packages the help function and how to read its output

All you need is your laptop and charger.

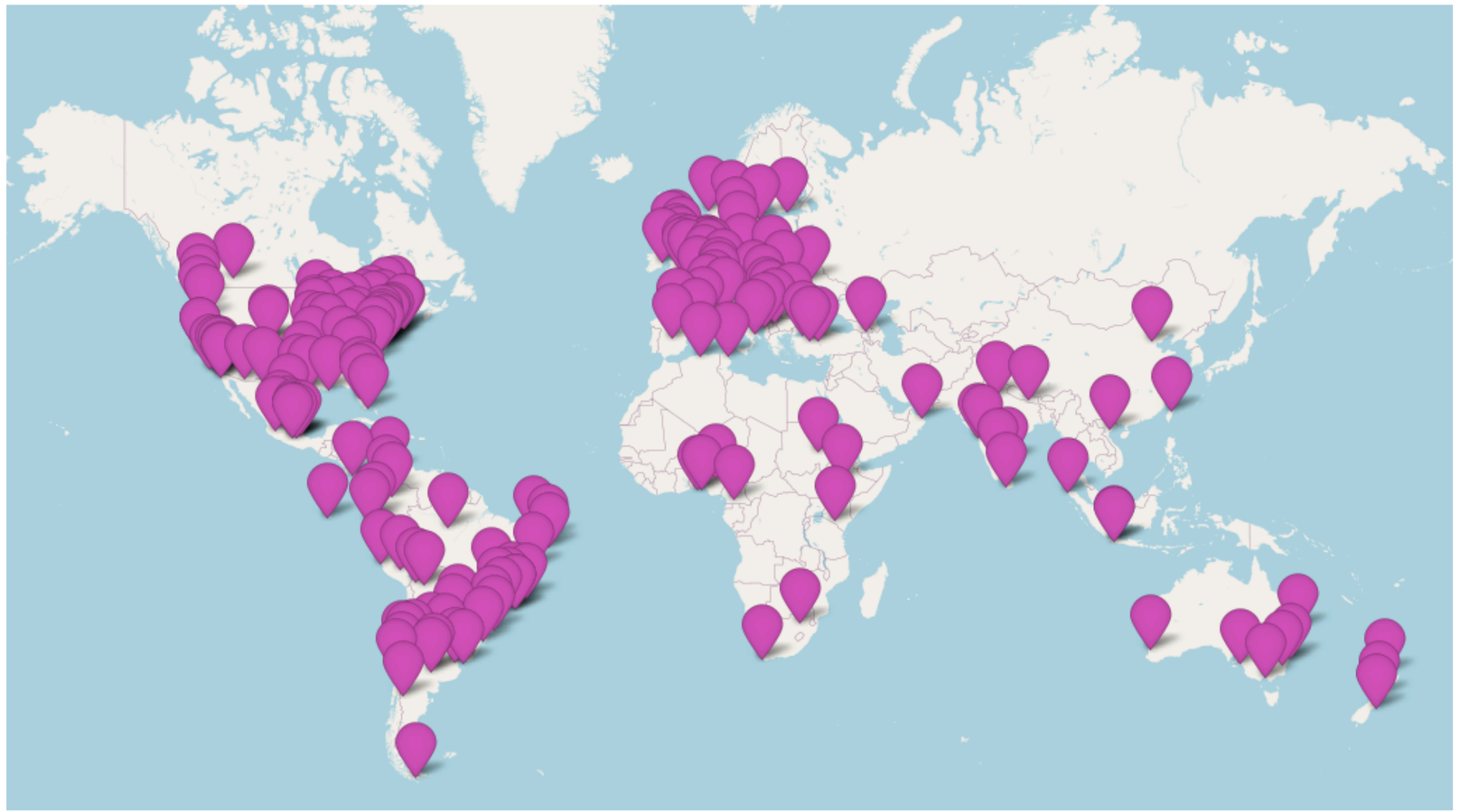
We look forward to seeing you!

R-Ladies

54
R-Ladies Countries

193
R-Ladies Cities

71618
R-Ladies members on meetup.com



	city	state	country	dt_created	members
41	São Paulo		BR	2018-02-10	1004
38	Belo Horizonte		BR	2018-04-20	800
34	Niterói		BR	2018-06-04	580
51	Rio de Janeiro		BR	2017-02-27	579
23	Florianópolis		BR	2019-04-07	441
43	Porto Alegre		BR	2017-10-30	346
7	Natal		BR	2020-06-07	241
32	Salvador		BR	2018-07-23	192
22	Recife		BR	2019 09 02	147
25	Goiania		BR	2019 05 06	146
18	Vitória		BR	2019 09 29	132
6	Fortaleza		BR	2020-06-09	69
26	Lavras		BR	2019-04-16	41
5	Curitiba		BR	2020-06-12	36
14	Ribeirão Preto		BR	2020-03-06	32
21	Manaus		BR	2019-09-11	29

52
R-Ladies groups in Latin America

AI Inclusive

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

Missão: Aumentar a **representatividade** e **participação** de minorias em Inteligência Artificial

- 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

Capítulos no Rio de Janeiro, Salvador e em San Francisco (EUA)

Se tiver interesse em criar um capítulo, é só enviar um email: info@ai-inclusive.org

Siga a gente:

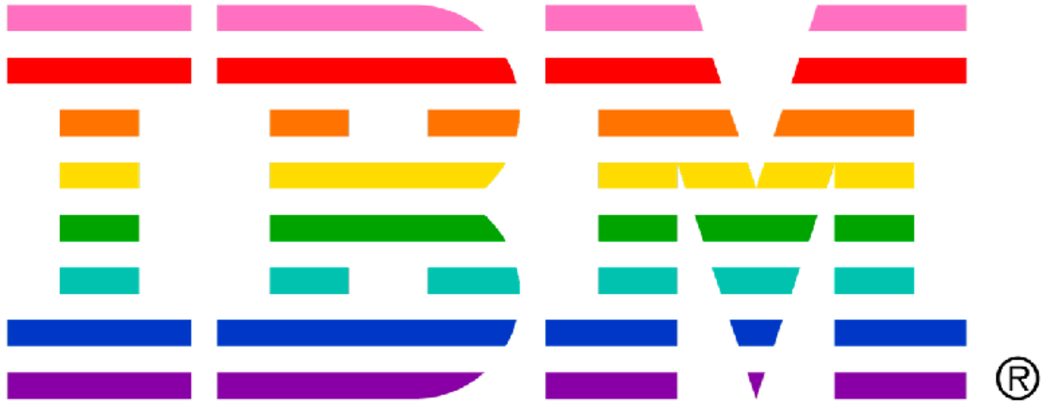
bit.ly/ai-inclusive-instagram

Materiais de Inteligência Artificial, Ciência de Dados, Machine Learning Eventos, Ingressos de graça e muito mais!

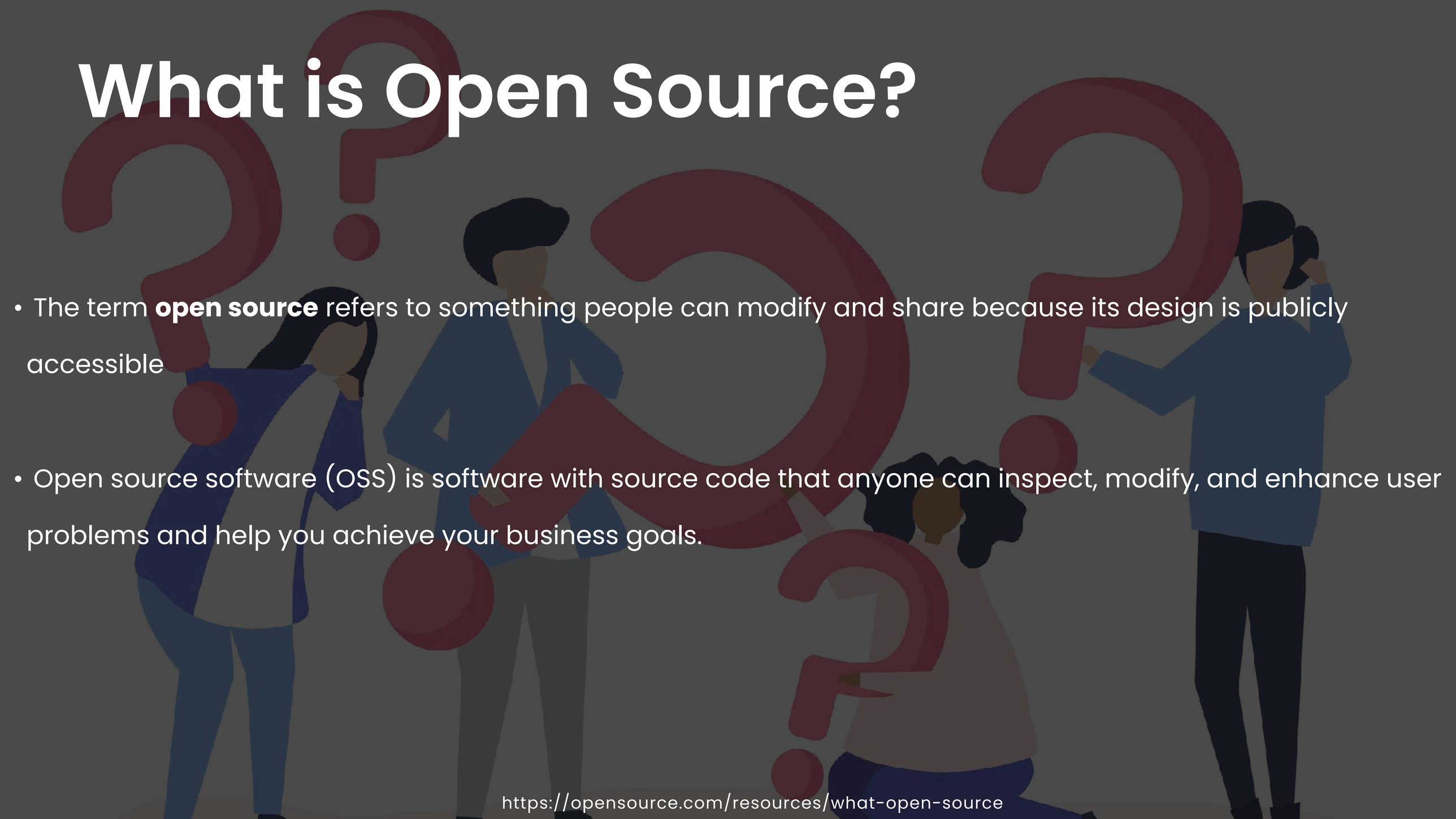
Em breve:

Bolsas de estudo para cursos de R, Python, Ciência de Dados, SQL, e outros.

O que você faz na IBM?



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 user problems and help you achieve your business goals.

https://opensource.com/resources/what-open-source

Open Source @ IBM

A strong history and commitment to open source

Open Source @ IBM

https://www.ibm.com/opensource/story/

IBM open source by the numbers

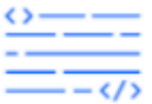
At IBM, we take open source seriously. We train our employees in the best practices for engaging in open source communities and the importance of open governance, and we empower them to create open source projects that solve their business and personal problems.

25+

years in
open source

3,000+

IBMers active
in open source



15,000

commits per
month

1,900

hosted GitHub
repositories

Center for Open Source Data & AI Technologies (CODAIT)

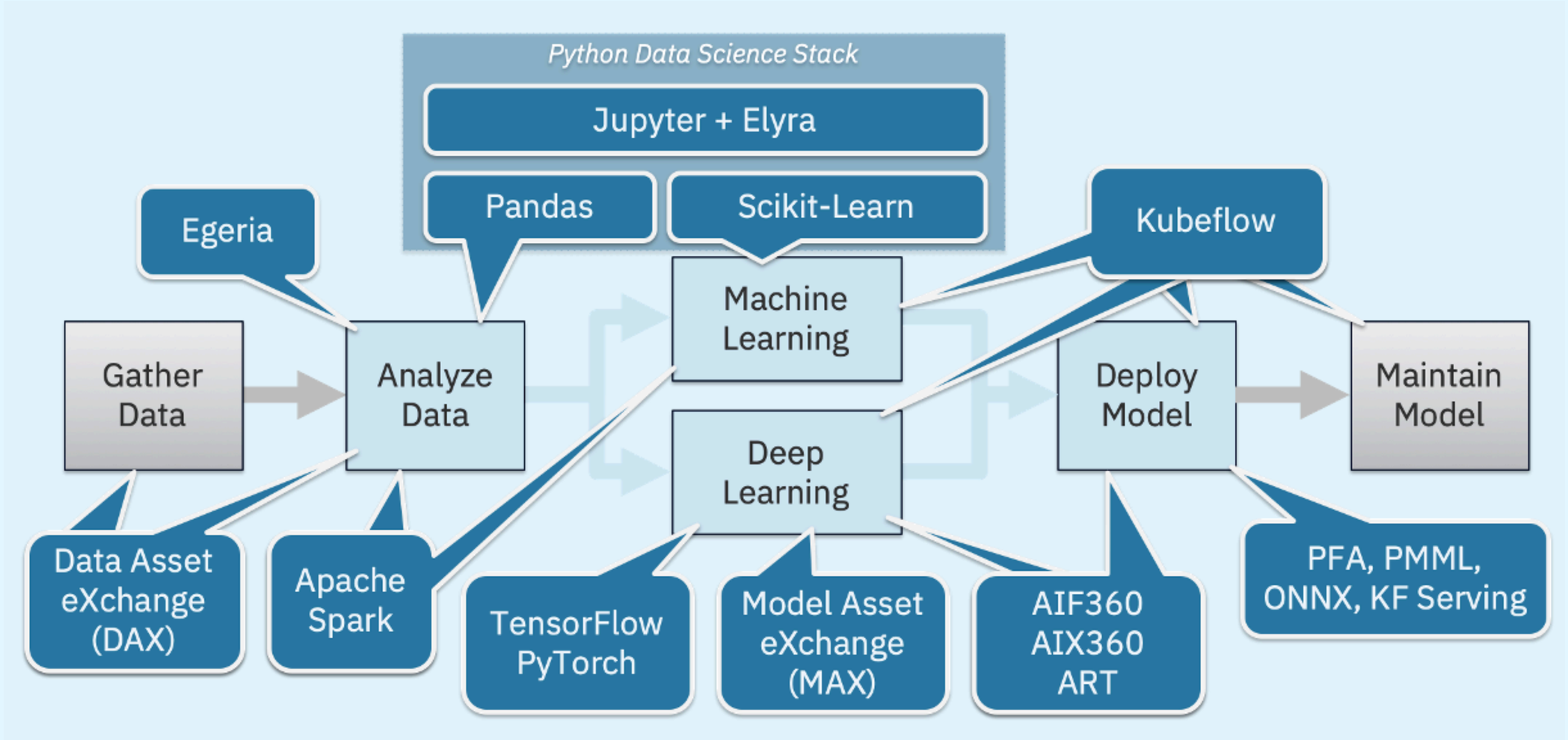
We build tools to make AI accessible and available to everybody

30+ Open Source Developers

codait.org

We build tools to make AI accessible and available to everybody

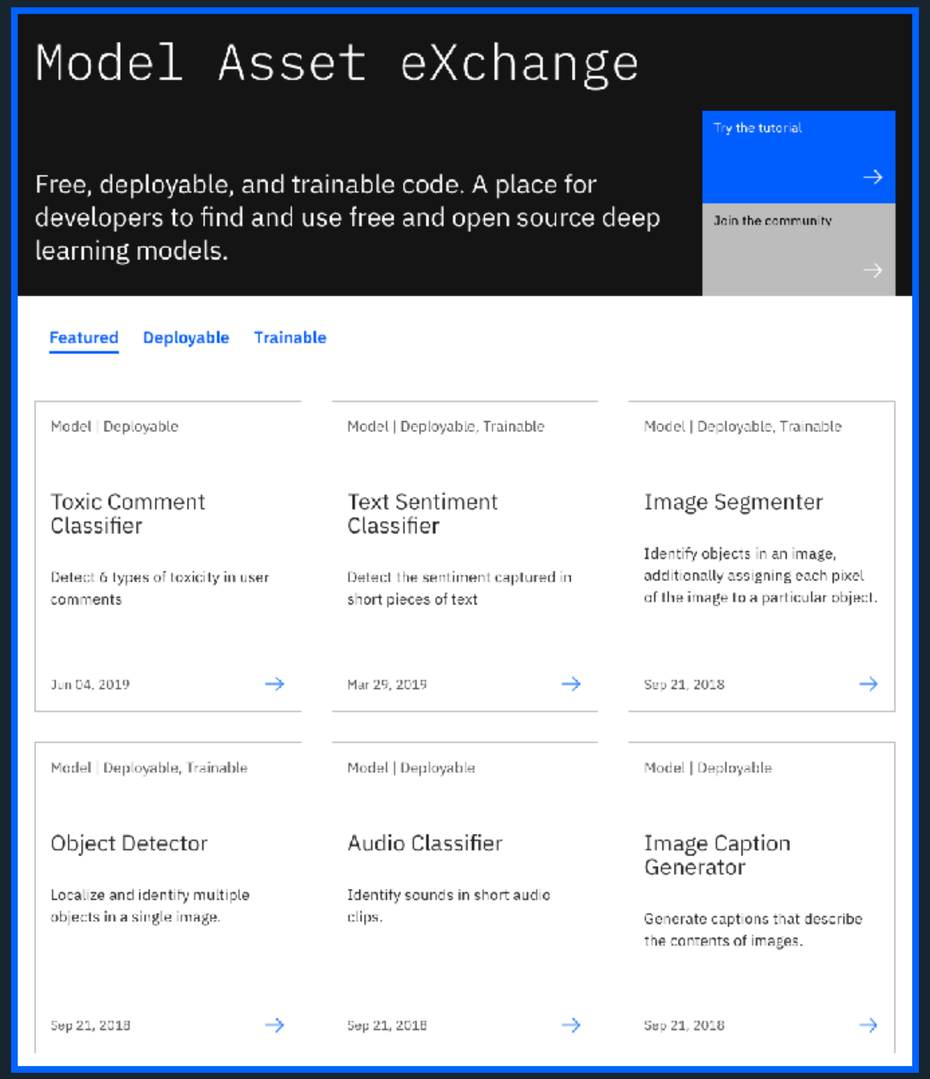
(codait.org)



We contribute to and advocate for the open-source technologies

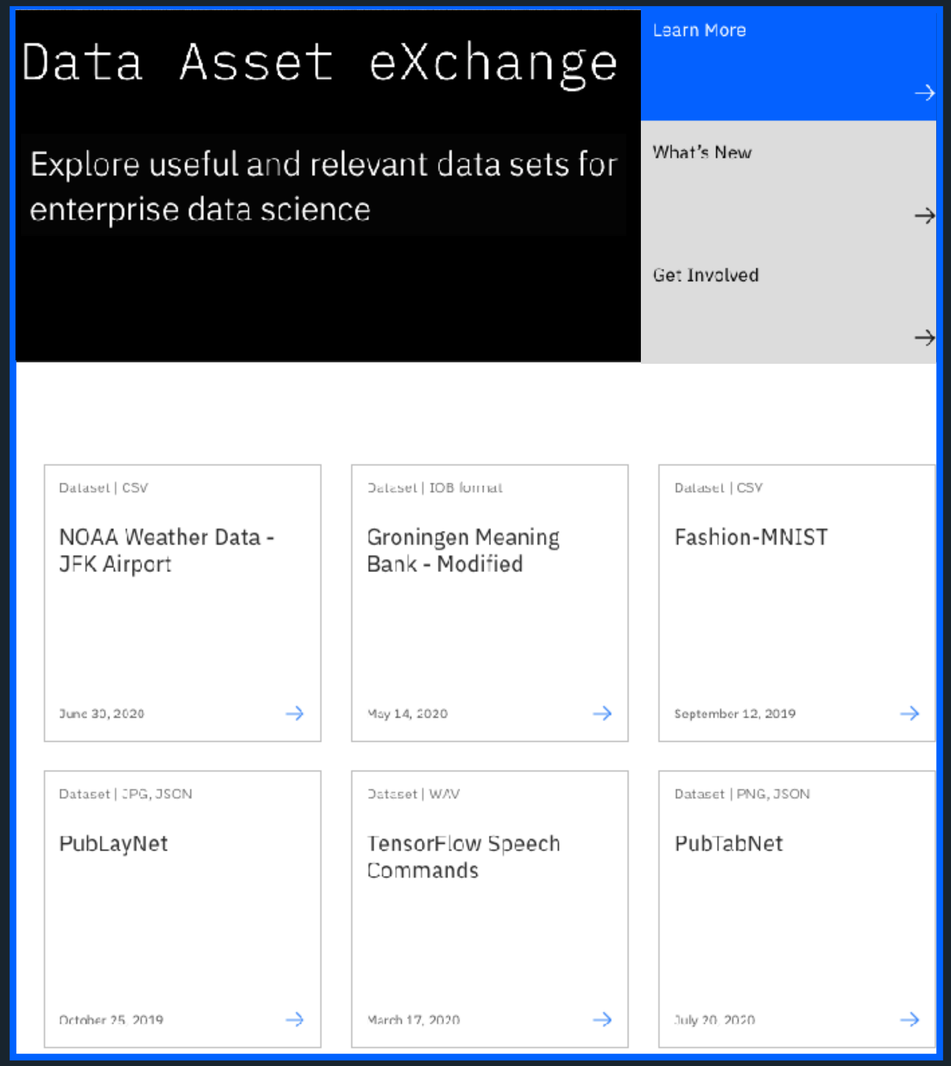
codait.org

Model Asset eXchange



ibm.biz/model-exchange

Data Asset eXchange



ibm.biz/data-exchange

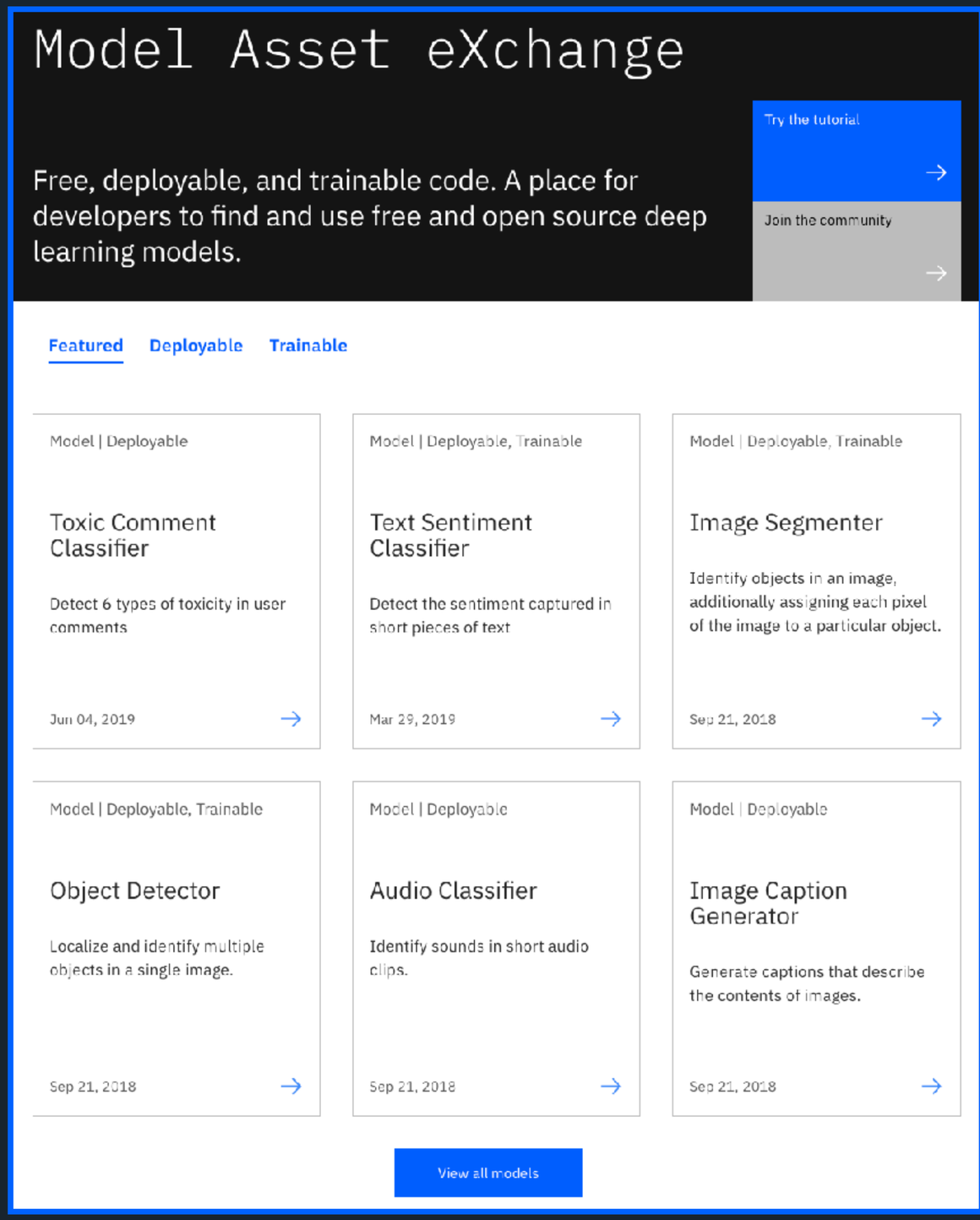
Model Asset eXchange (MAX)

Place for developers/data scientists to find and use

free and **open source** deep learning models

ibm.biz/model-exchange

slides: bit.ly/cimatech-2020



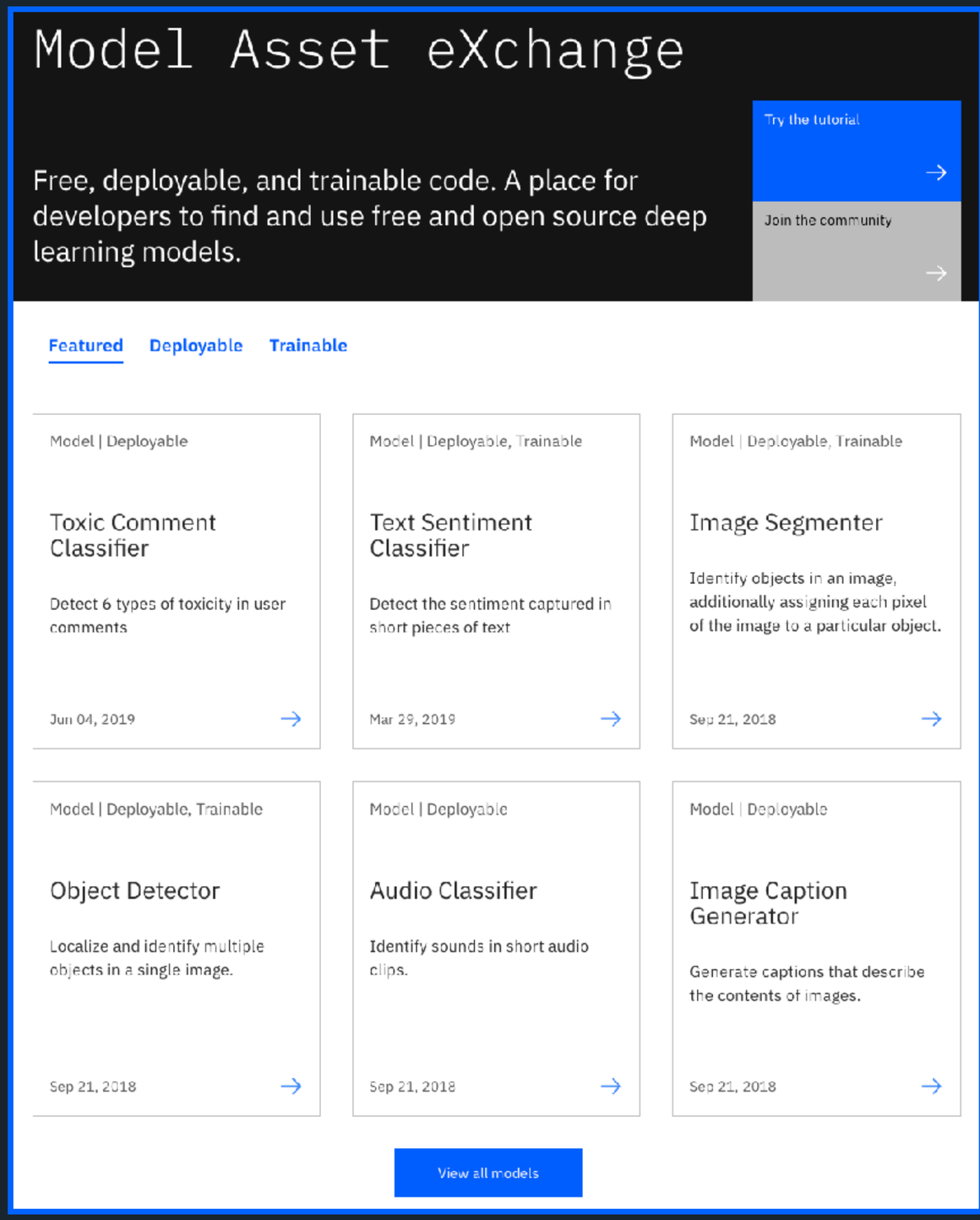
Model Asset eXchange (MAX)

Place for developers/data scientists to find and use

free and **open source** deep learning models

ibm.biz/model-exchange

slides: bit.ly/cimatech-2020



MAX Object Detector

Upload an image

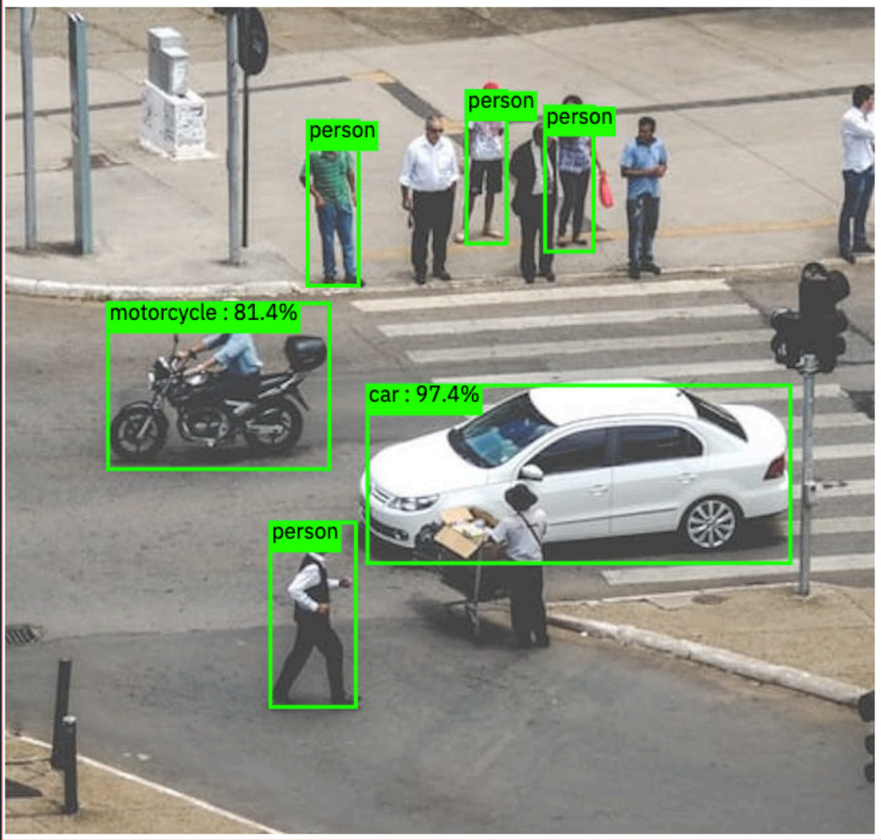
Choose File No file chosen

Submit Use your webcam

Filter detected objects ⓘ

Probability Threshold: 50%

Labels Found ⓘ



MAX Image Caption Generator

Upload A New Image ⓘ

Choose Files No file chosen

Submit

Deselect All Select All Delete Uploaded Images


a man holding a snowboard on top of a snow covered slope .


a person riding a snowboard down a snow covered slope .


a man riding skis down a snow covered slope .


a person riding a motorcycle on a race track


a tenn s racket w th a tennis ball in it

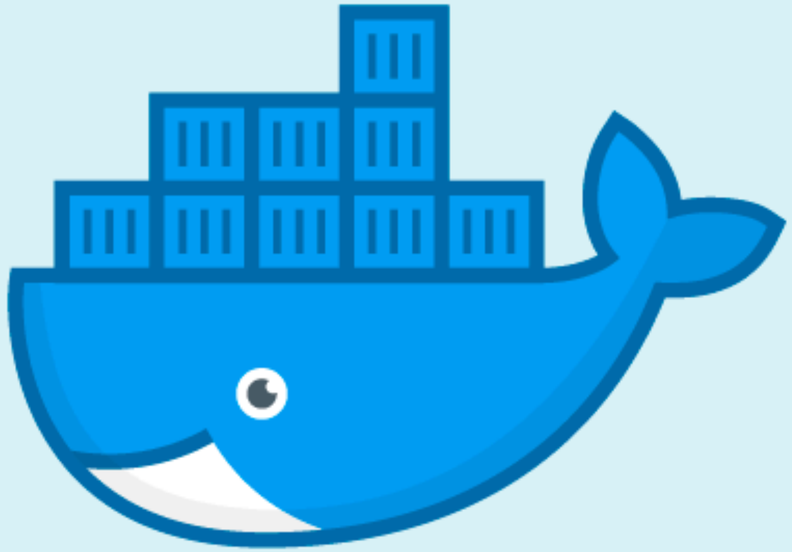

a group of baseball players stand ng on top of a field .


a crowd of people watching a tennis match .


a group of young men playing a game of basketball .

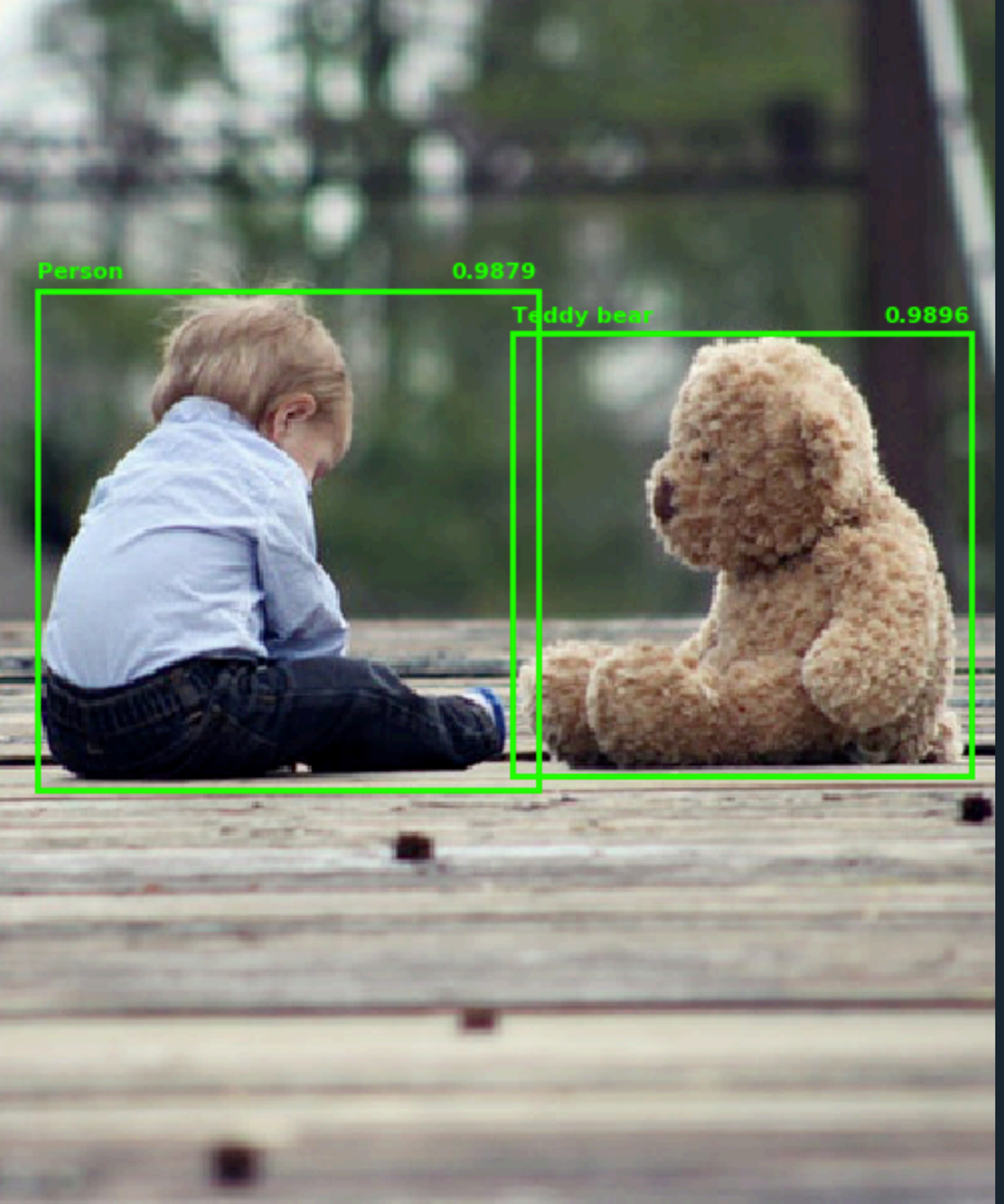

a female tenn s player in action on the court .

What do I need to get started?



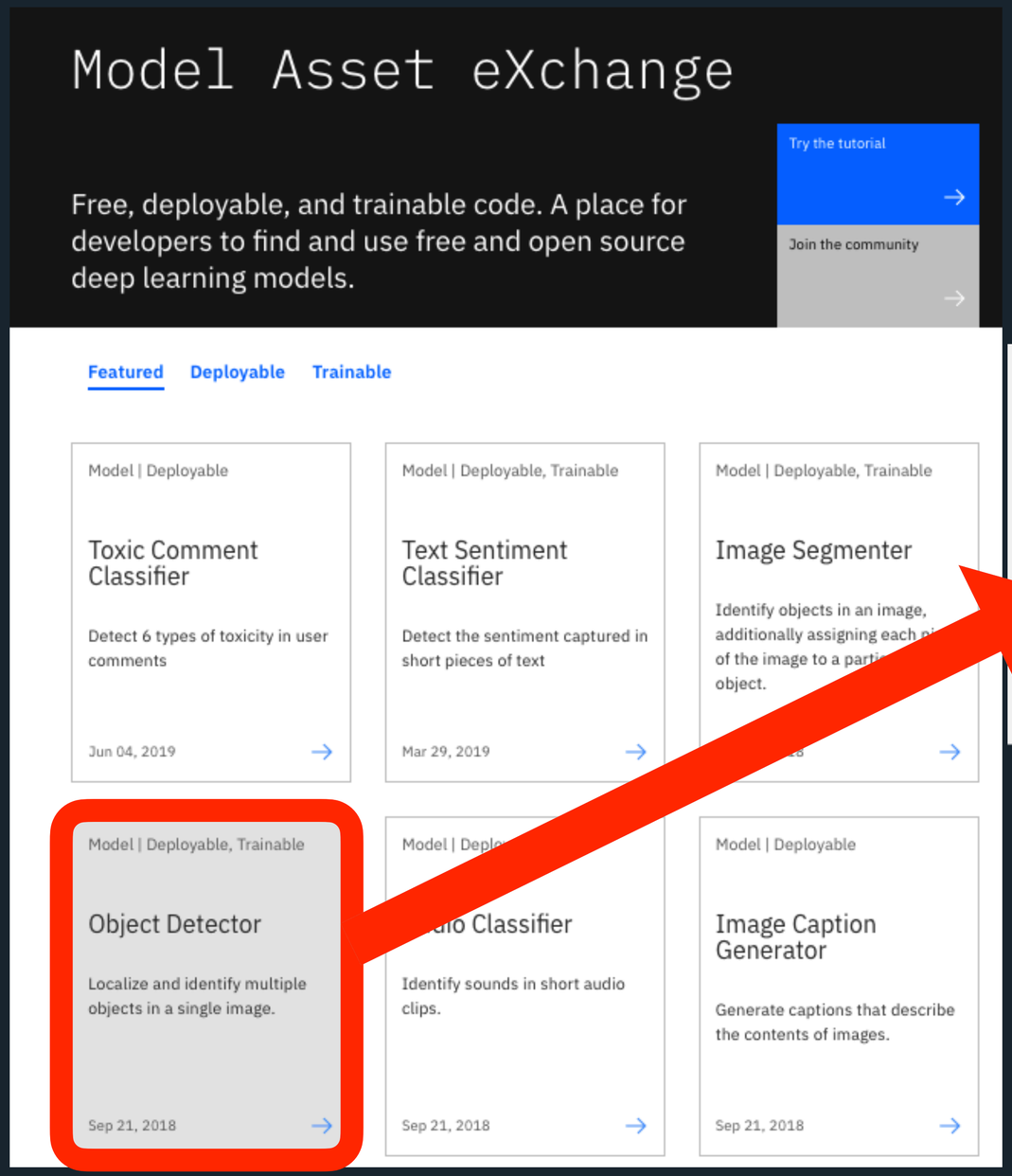
docker

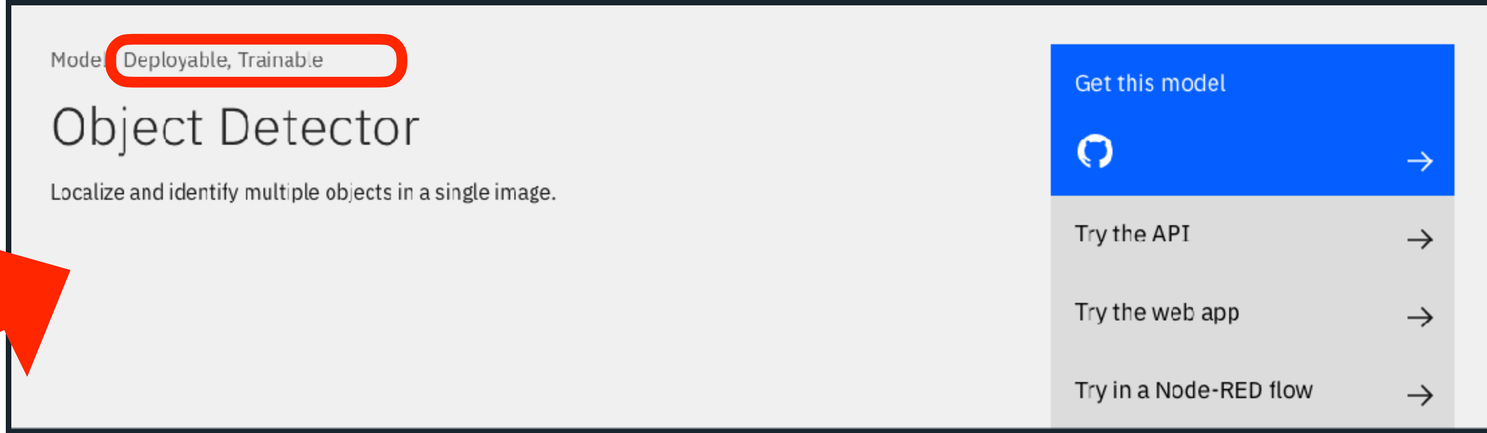
https://www.docker.com



OBJECT DETECTOR

Localize and **identify** multiple objects in a single image

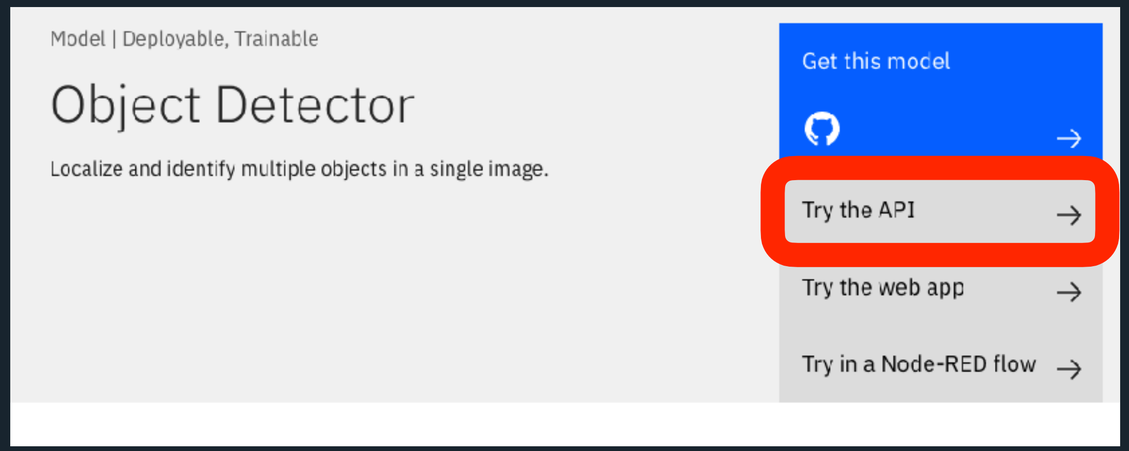




ibm.biz/model-exchange

slides: bit.ly/cimatech-2020

Access the API via Python

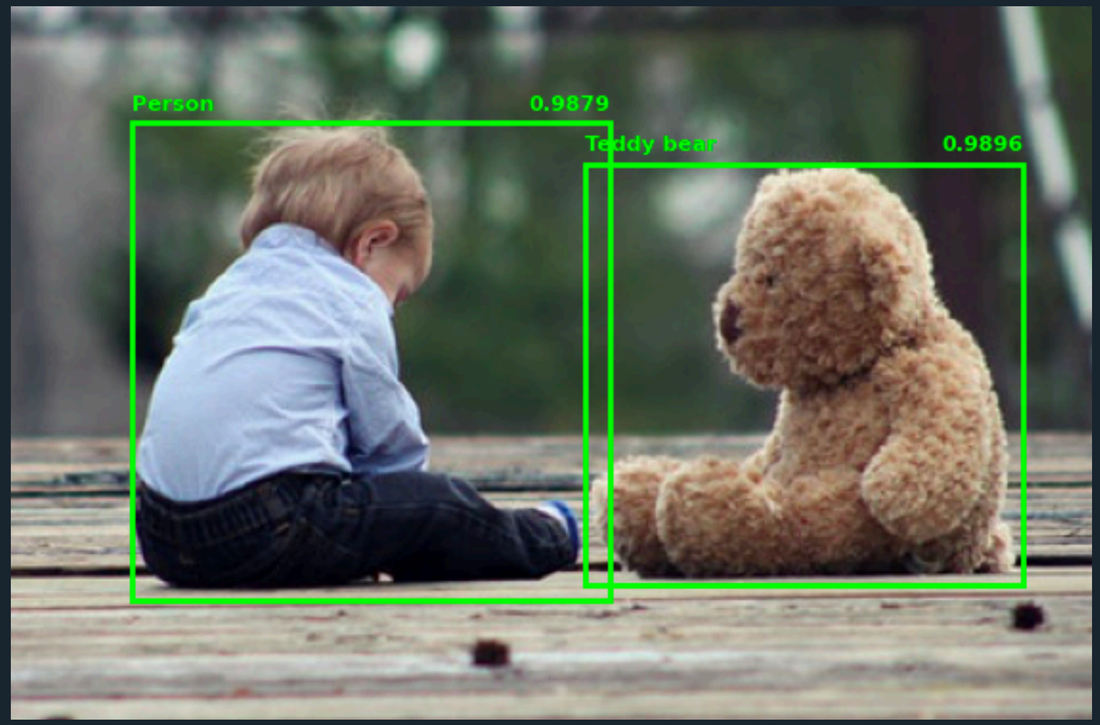



```
# Model
url = 'http://max-object-detector.codait-prod-41208c73af8fca213512856c7a09
db52-0000.us-east.containers.appdomain.cloud/'
model_endpoint =  'model/predict'
complete_url = url + model_endpoint

# 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)
```



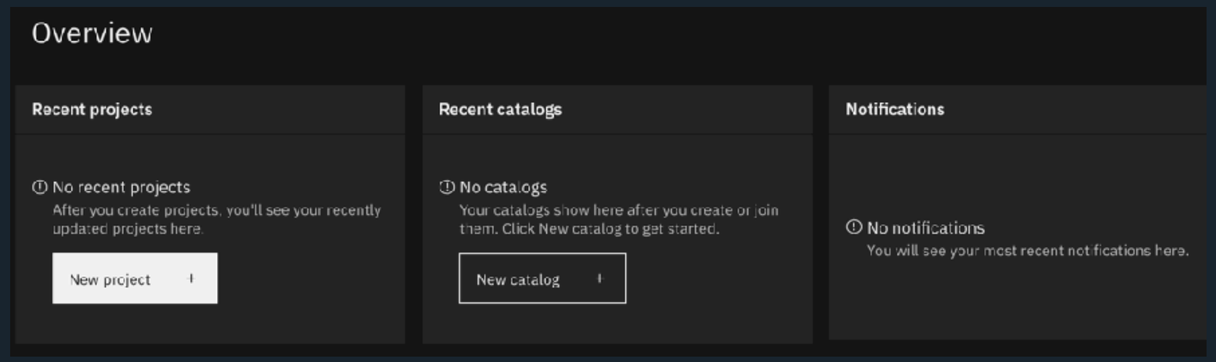
Try yourself here:

ibm.biz/max-notebook

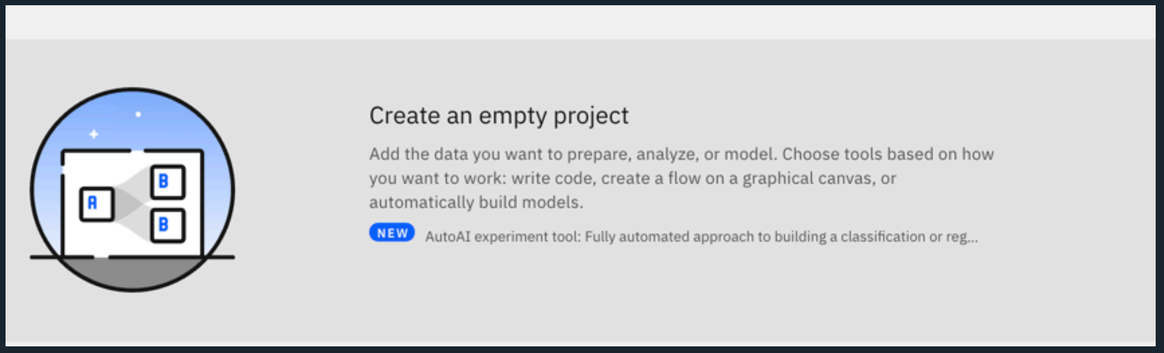
Sua vez ...

1) Clique aqui ibm.biz/Bdq5CG para criar uma conta

2) Login na sua conta e crie um projeto

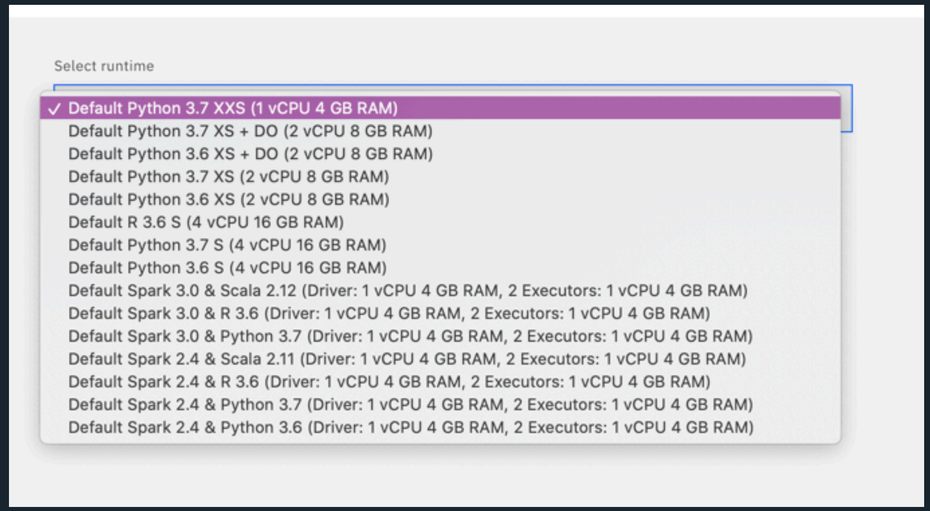


3) Crie um *empty project*



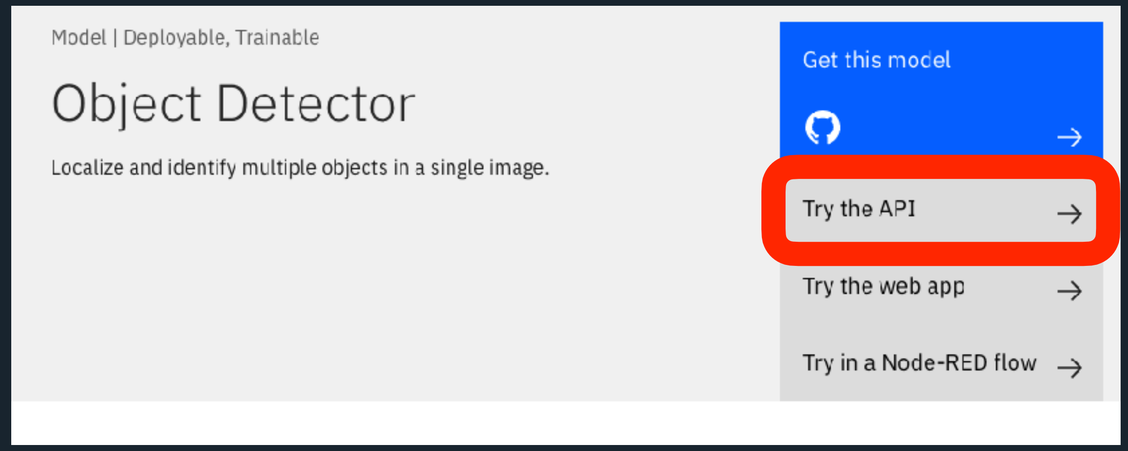
4) Vá no link ibm.biz/max-notebook
e copie o projeto

5) Selecione um *runtime*



Agora você pode rodar o notebook!

Access the API via R




```r
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)
```

Access the API via Swagger

MAX Object Detector 1.4.0

[Base URL: /]
http://max-object-detector.codait-prod-41208c73af8fca213512856c7a09db52-0000.us-east.containers.appdomain.cloud/swagger.json

Localize and identify multiple objects in a single image.

model Model information and inference operations

- GET `/model/labels` Return the list of labels that can be predicted by the model
- GET `/model/metadata` Return the metadata associated with the model
- POST `/model/predict` Make a prediction given input data

POST `/model/predict` Make a prediction given input data

Parameters — Cancel

Name	Description
image * required file *(formData)*	An image file (encoded as PNG or JPG/JPEG) Choose File traffic.jpeg
threshold number *(query)*	Probability threshold for including a detected object in the response in the range [0, 1] (default: 0.7). Lowering the threshold includes objects the model is less certain about. 0.7

Execute | Clear

Responses — Response content type: application/json

Curl

```
curl -X POST "http://max-object-detector.codait-prod-41208c73af8fca213512856c7a09db52-0000.us-east.containers.appdomain.cloud/model/predict?threshold=0.7" -H "accept: application/json" -H "Content-Type: multipart/form-data" -F "image=@traffic.jpeg;type=image/jpeg"
```

Request URL

```
http://max-object-detector.codait-prod-41208c73af8fca213512856c7a09db52-0000.us-east.containers.appdomain.cloud/model/predict?threshold=0.7
```

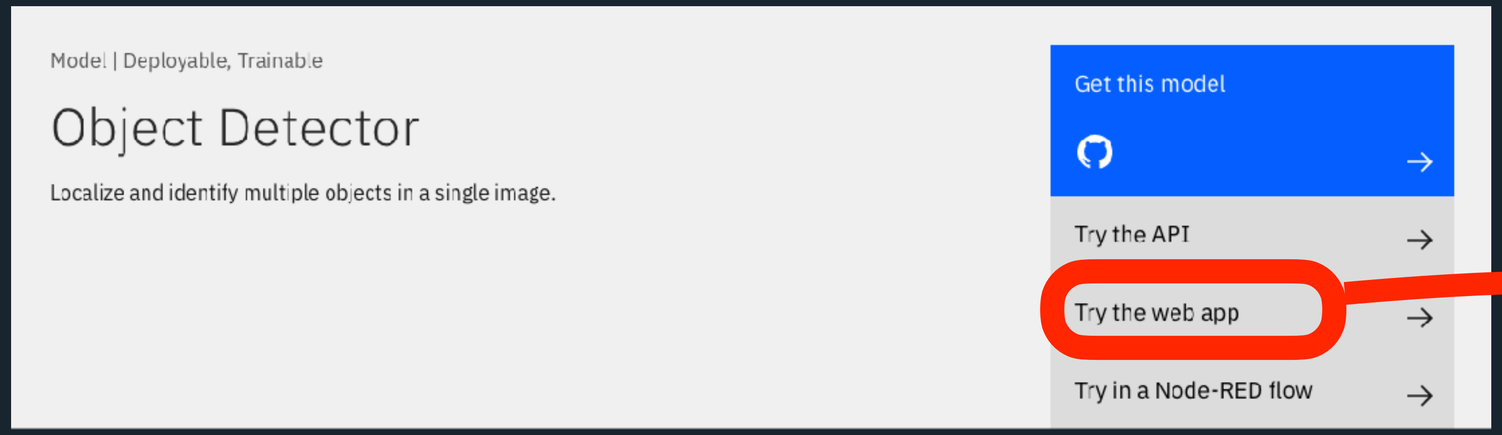
Server response

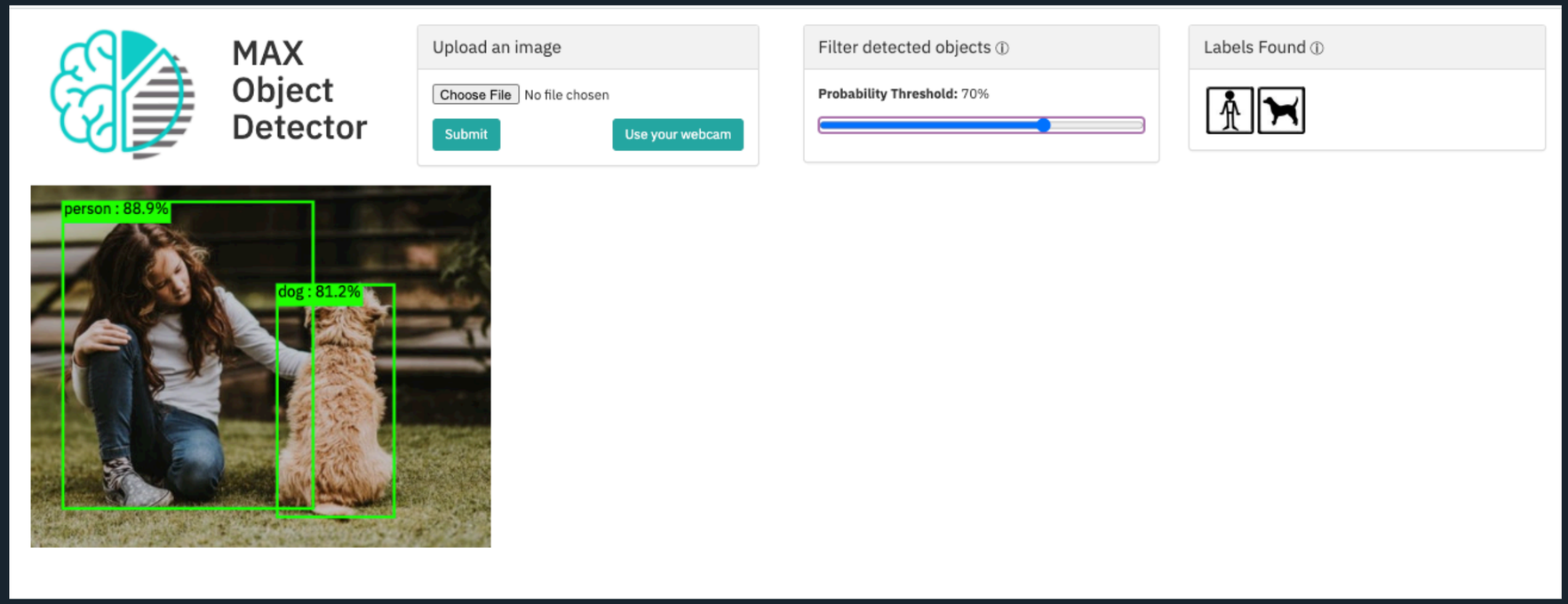
Code: 200 — Details

Response body


```
{
  "status": "ok",
  "predictions": [
    {
      "label_id": "3",
      "label": "car",
      "probability": 0.9741619627685547,
      "detection_box": [
        0.45756405559196472,
        0.41685003042222107,
        0.6725007891654968,
        0.9029390215873718
      ]
    },
    {
      "label_id": "1",
      "label": "person",
      "probability": 0.8824045658111572,
      "detection_box": [
        0.6231690645217896,
        0.30522748827934265,
        0.8462619781494141,
        0.4034259617328644
      ]
    },
    {
      "label_id": "4",
      "label": "motorcycle",
      "probability": 0.8161902685165405,
      "detection_box": [
        0.3583410978317261,
        0.11843161284923553,
        0.5584568117141724,
        0.37282349338531494
      ]
    },
```

Access the API via Web App





Try yourself here:

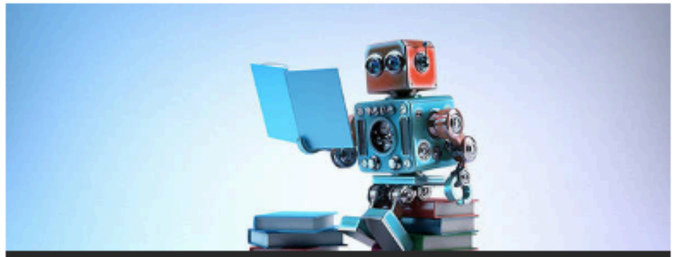
ibm.biz/object-detector-webapp

How do I get started?

ibm.biz/max-tutorial

Examples on how to easily consume MAX models

ibm.biz/max-code-patterns



Code Pattern

Create a machine learning powered web app to answer questions

Nov 05, 2019

Code Pattern

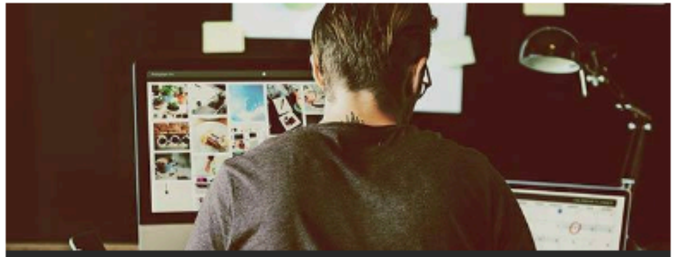
Build a web app that recognizes yoga poses using a model from the Model Asset Exchange

Oct 03, 2019

Code Pattern

Use your arms to make music

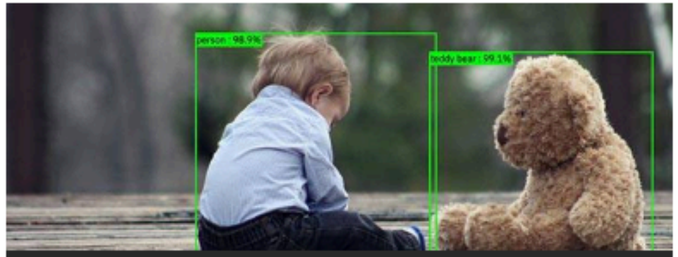
Apr 22, 2019



Code Pattern

Create a web app to interact with machine learning generated image captions

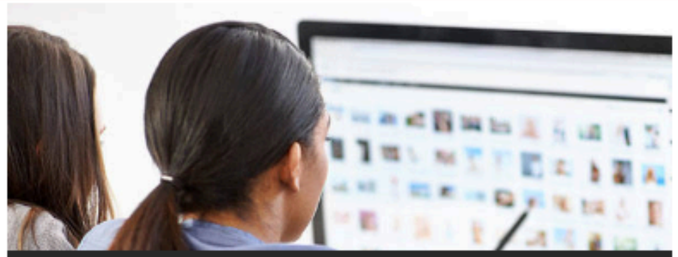
Mar 28, 2019



Code Pattern

Create a web app to visually interact with objects detected using machine learning

Mar 28, 2019



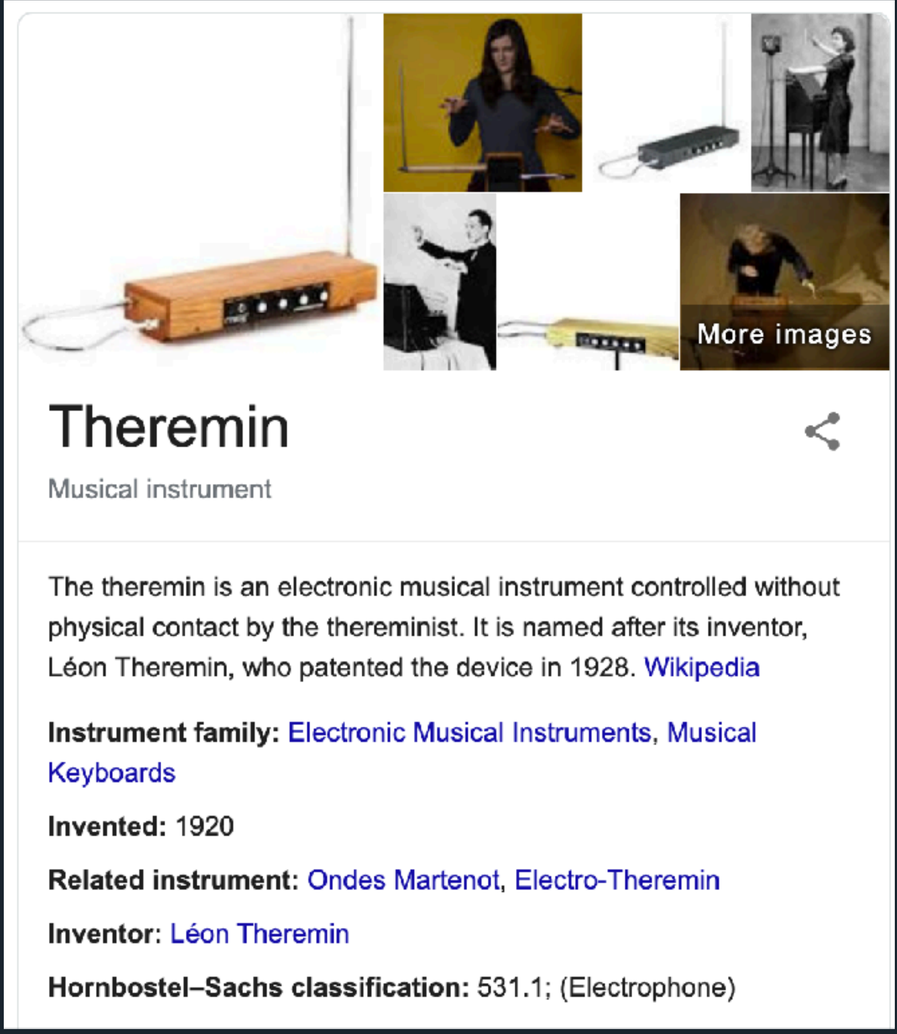
Code Pattern

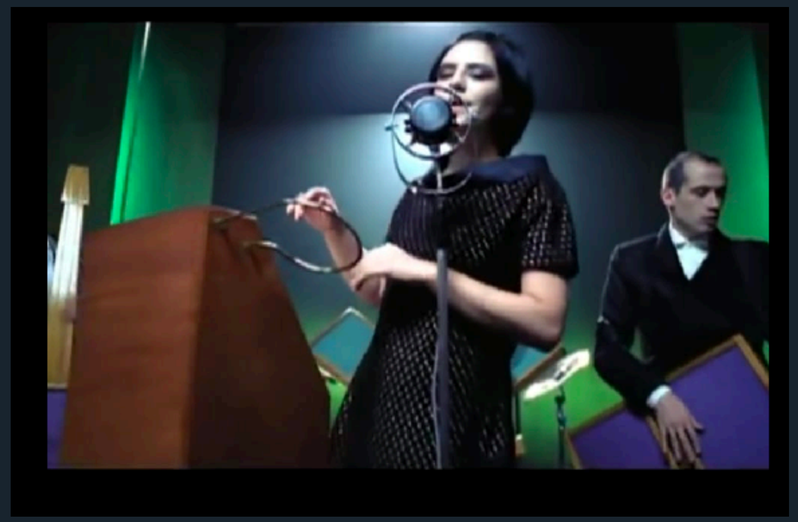
Deploy a deep learning-powered 'Magic cropping tool'

Mar 28, 2019

Use your arms to make music

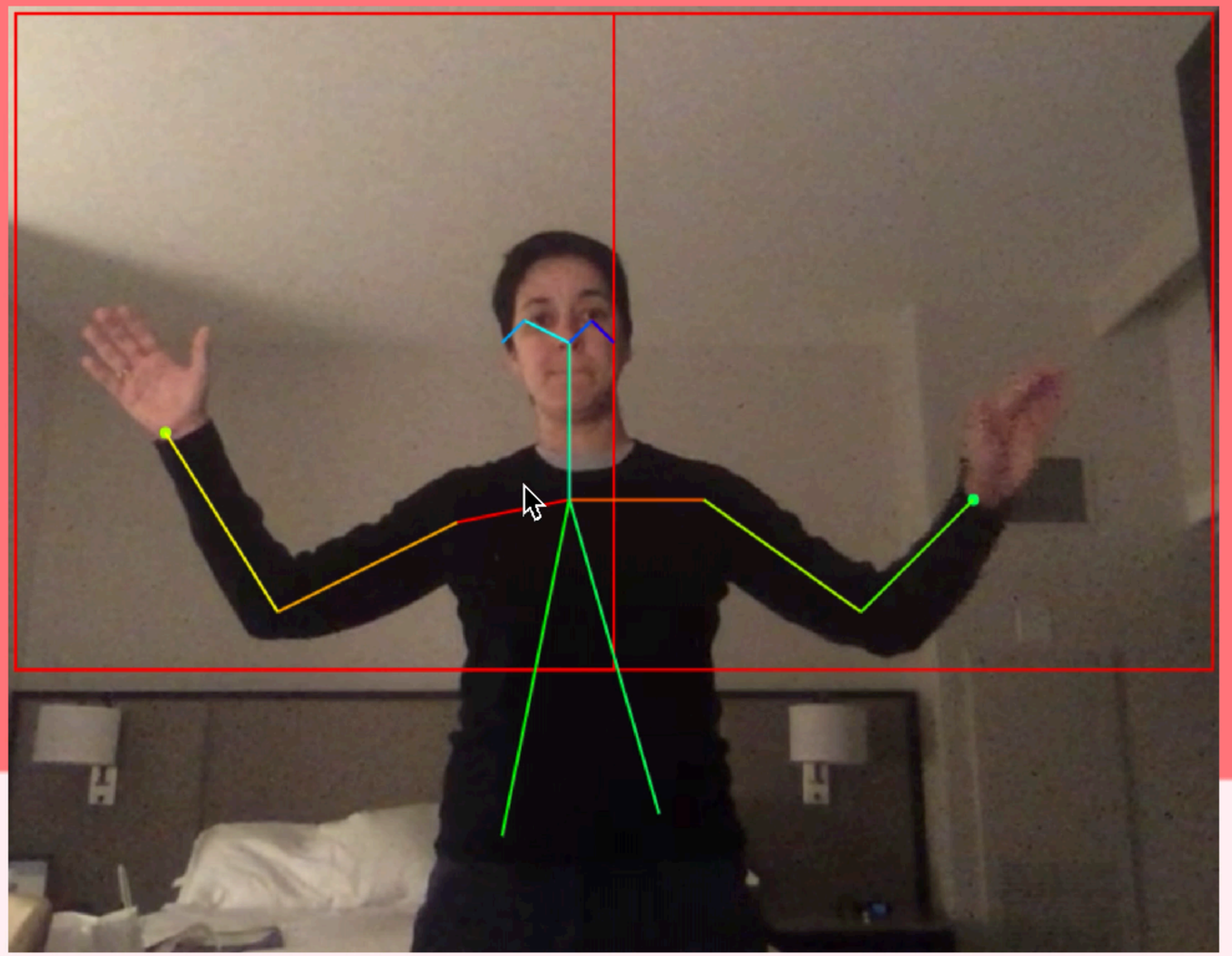
Create music with your arms using the Model Asset eXchange (MAX) **human pose estimator** model and **TensorFlow**





veremax

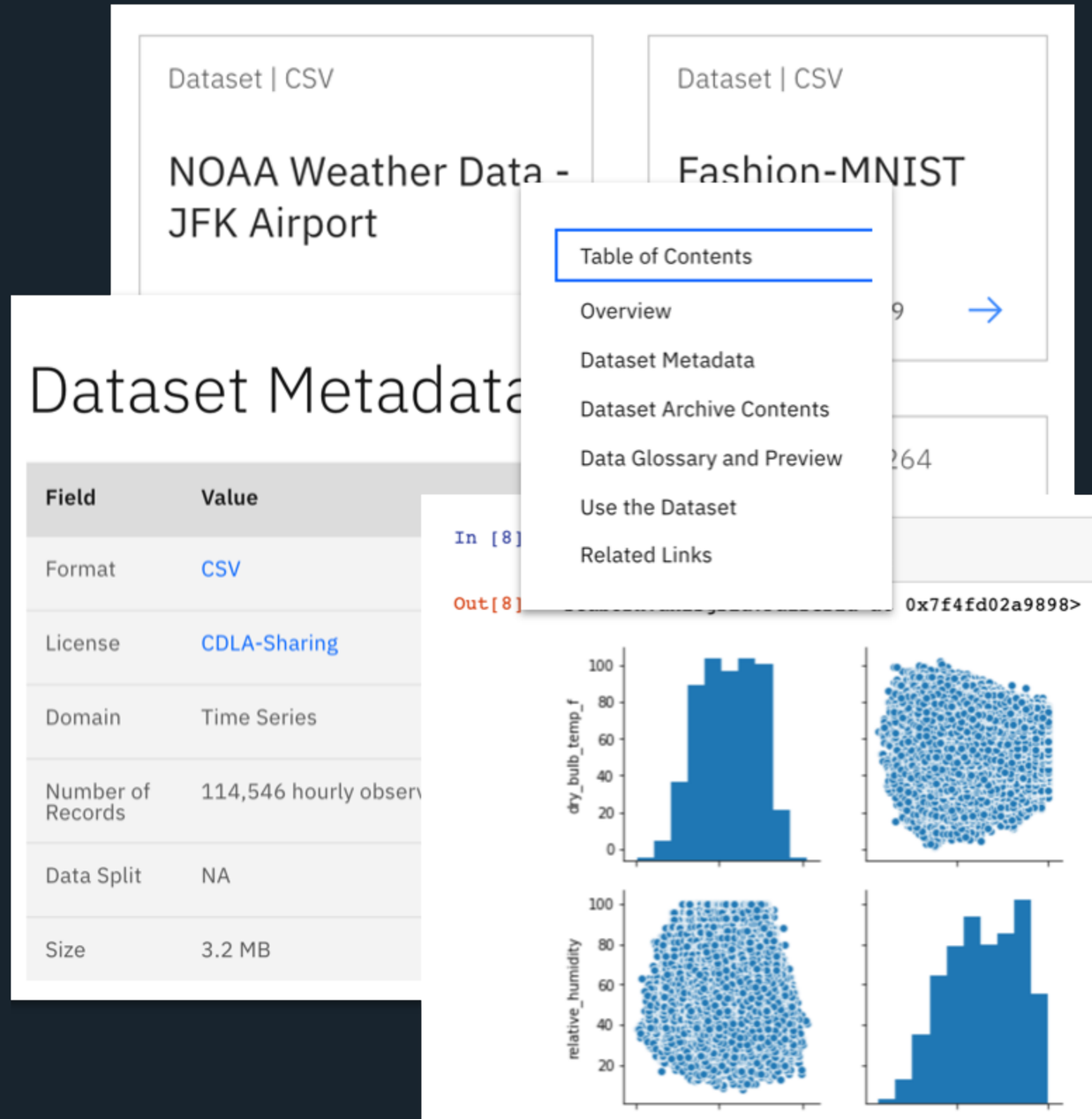
a video theremin using OpenPose



powered by the MAX Human Pose Estimator | made possible with TensorFlow.js | learn more about the code

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

CDLA-Sharing | CSV

NOAA Weather Data – JFK Airport

Local climatological data originally collected at JFK airport.

Save Like

Get this dataset →

Run dataset notebooks →

Preview the data & notebooks →

DAX Dataset Preview | Notebook Preview | Run Notebook in Watson Studio | Dataset Homepage

NOAA Weather Data – JFK Airport

Dataset Metadata | Dataset Preview | Dataset Glossary

Format	CSV
License	CDLA-Sharing
Domain	Time Series
Number of Records	114,546 hourly observations
Data Split	NA
Size	3.2 MB
Data Origin	National Oceanic and Atmospheric Administration (NOAA)
Dataset Version	Version 2 – September 12, 2019 Version 1 – July 16, 2019
Dataset Coverage	Location: New York City Dates: 2010-01-01 through 2018-07-27 Note: To download raw data from NOAA for a different region or date span, follow the steps outlined in the data archive's README.txt.
Business Use Case	*Agriculture* Detect unseasonal temperature change and alert farmers about potential damage to plants. Energy Regulate solar cell charging hours based on weather type condition and temperature. Regulate wind turbine operation based on wind speed and wind direction. Generate energy demand alerts based on temperature. Remotely adjust air conditioning configs to boost energy efficiency based on temperature shifts. *Retail* Estimate outdoor retail foot traffic based on weather condition and temperature predictions.

DAX Dataset Preview | Notebook Preview | Run Notebook in Watson Studio | Dataset Homepage

NOAA Weather Data – JFK Airport

Part 1 - Data Cleaning | Part 2 - Data Analysis | Part 3 - Time Series Forecasting

In [1]:
```
# @hidden_cell
# The project token is an authorization token that is used to access proje
ct resources like data sources, connections, and used by platform APIs.
from project_lib import Project
project = Project(project_id='...', project_access_token='...')
```

Cleaning NOAA Weather Data of JFK Airport (New York)

This notebook relates to the NOAA Weather Dataset - JFK Airport (New York). The dataset contains 114,546 hourly observations of 12 local climatological variables (such as temperature and wind speed) collected at JFK airport. This dataset can be obtained for free from the IBM Developer Data Asset Exchange.

In this notebook, we clean the raw dataset by:

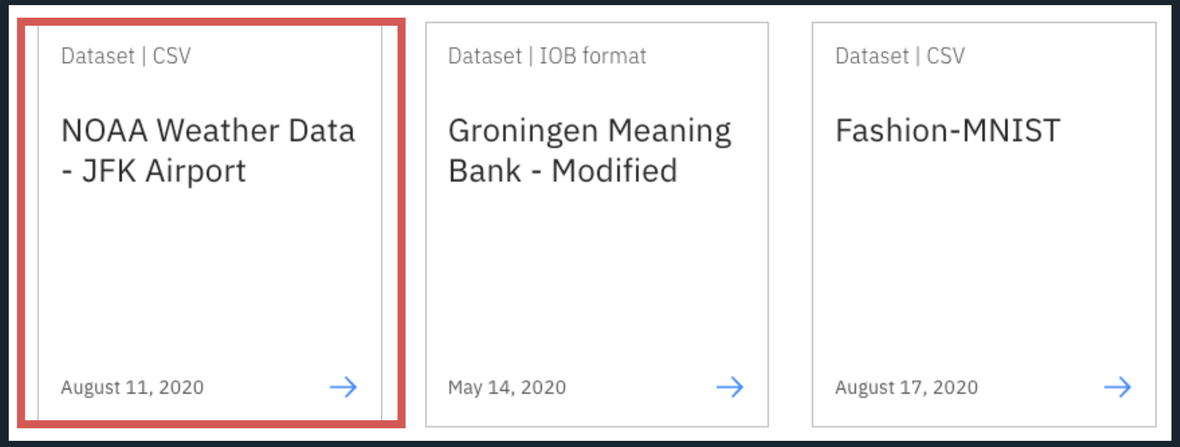
- removing redundant columns and preserving only key numeric columns
- converting and cleaning data where required
- creating a fixed time interval between observations (this aids with later time-series analysis)
- filling missing values
- encoding certain weather features

Table of Contents:

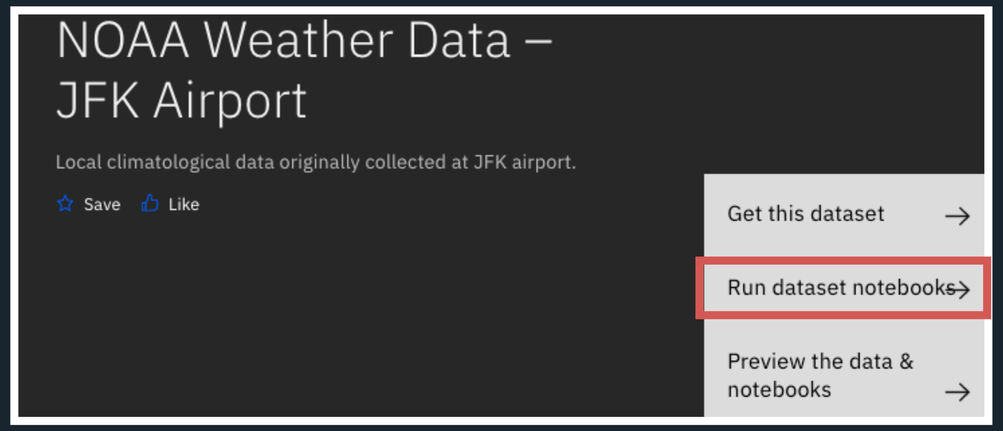
Sua vez ....

1) Clique aqui [ibm.biz/Bdq5CG](ibm.biz/Bdq5CG) para criar uma conta

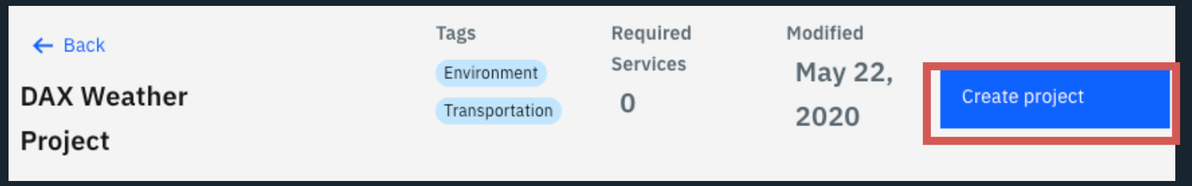
2) Vá até o Data Asset eXchange ibm.biz/data-exchange, escolha um banco de dados



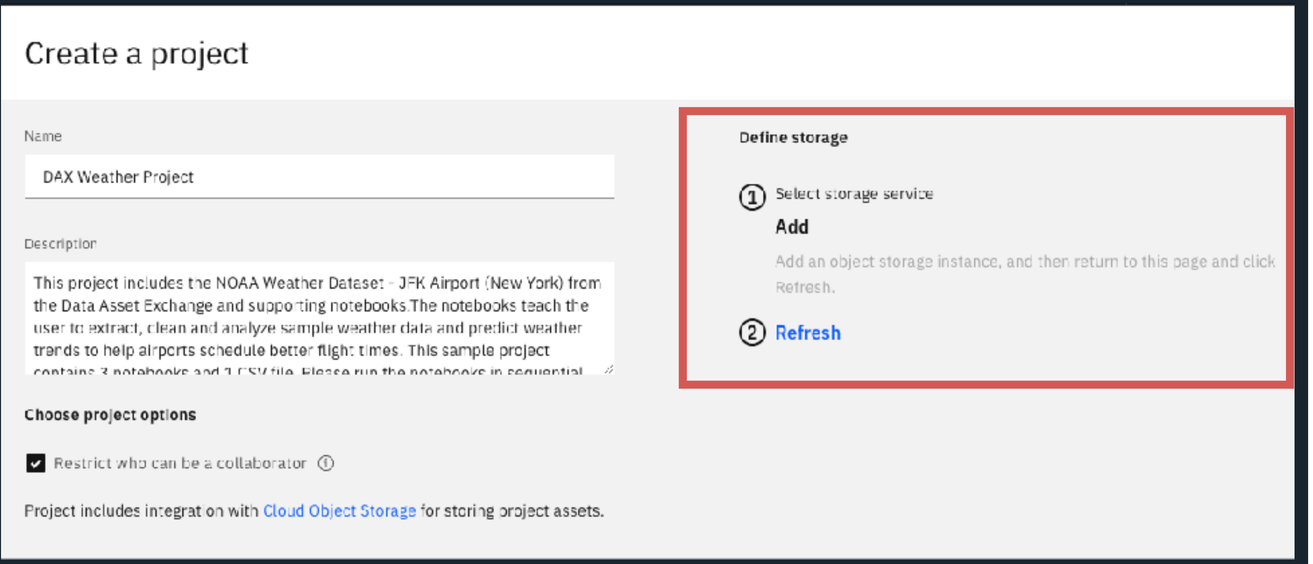
3) Click em “Run dataset notebooks”



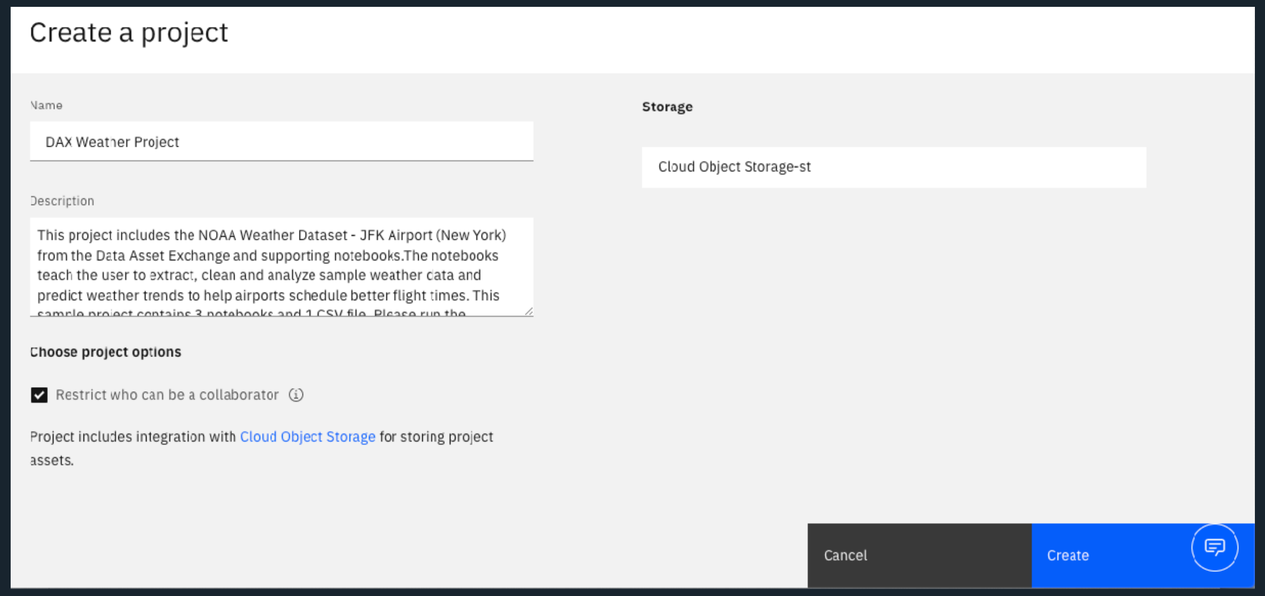
4) Click em “Create a Project”



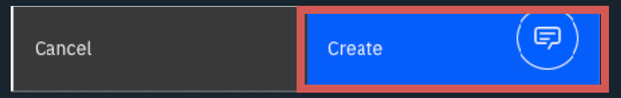
5) Selecione um “Storage Service”

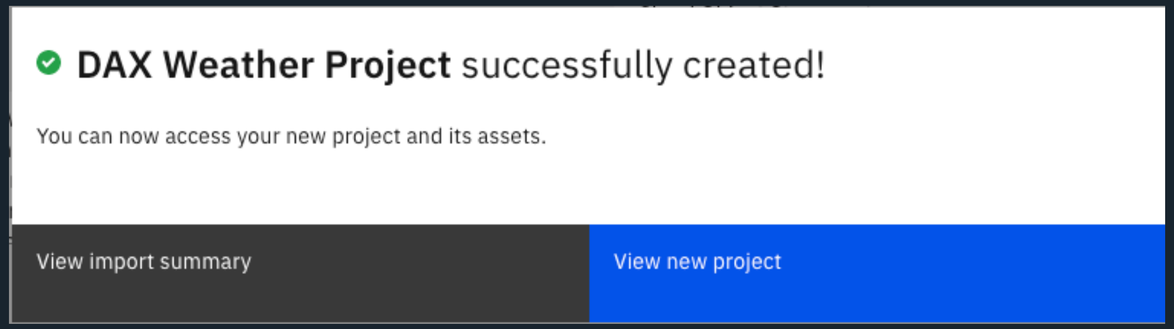


6) Clique em “Refresh”. Você deve ter algo assim:

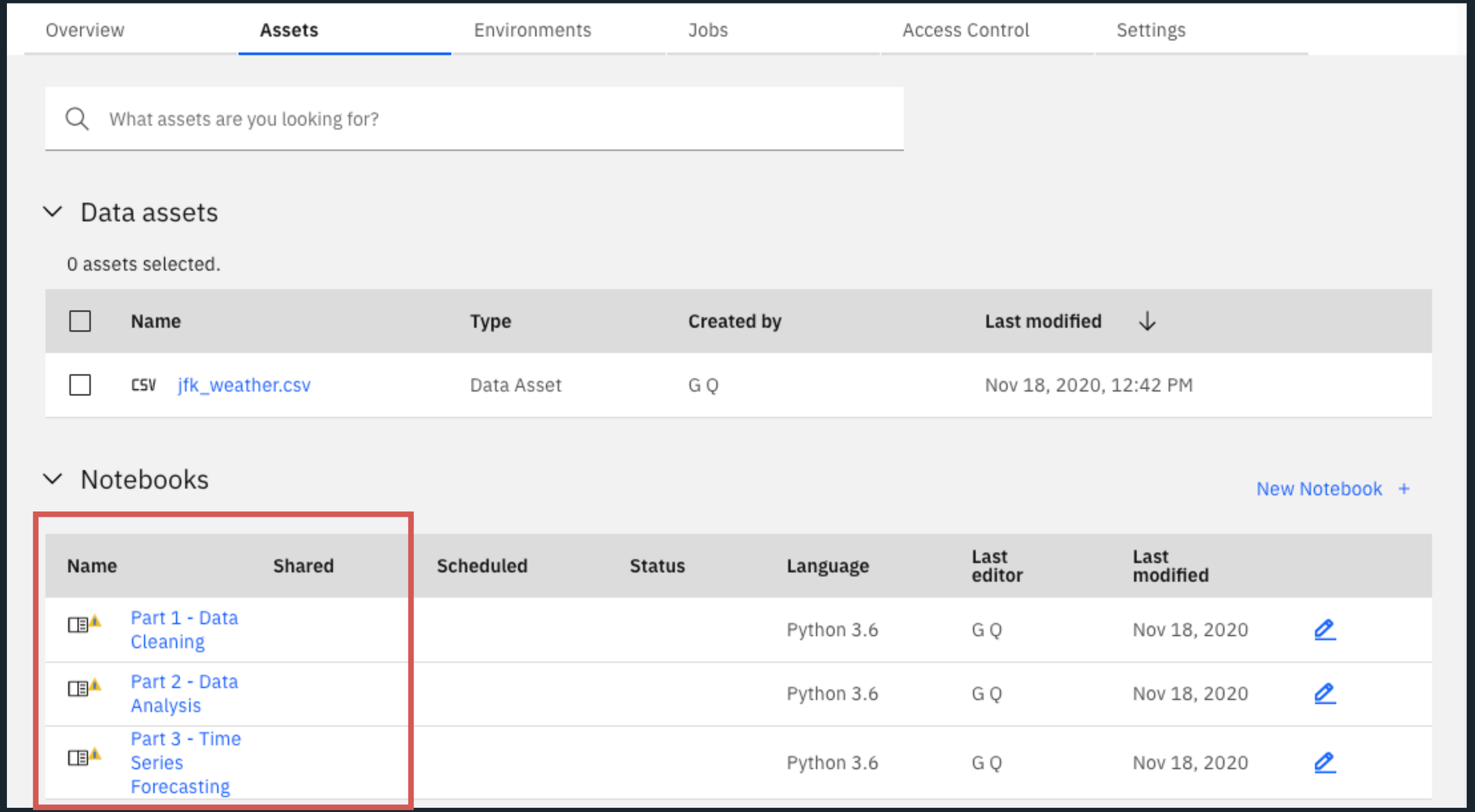


7) Clique em “Create”





Agora você pode rodar os Notebooks!



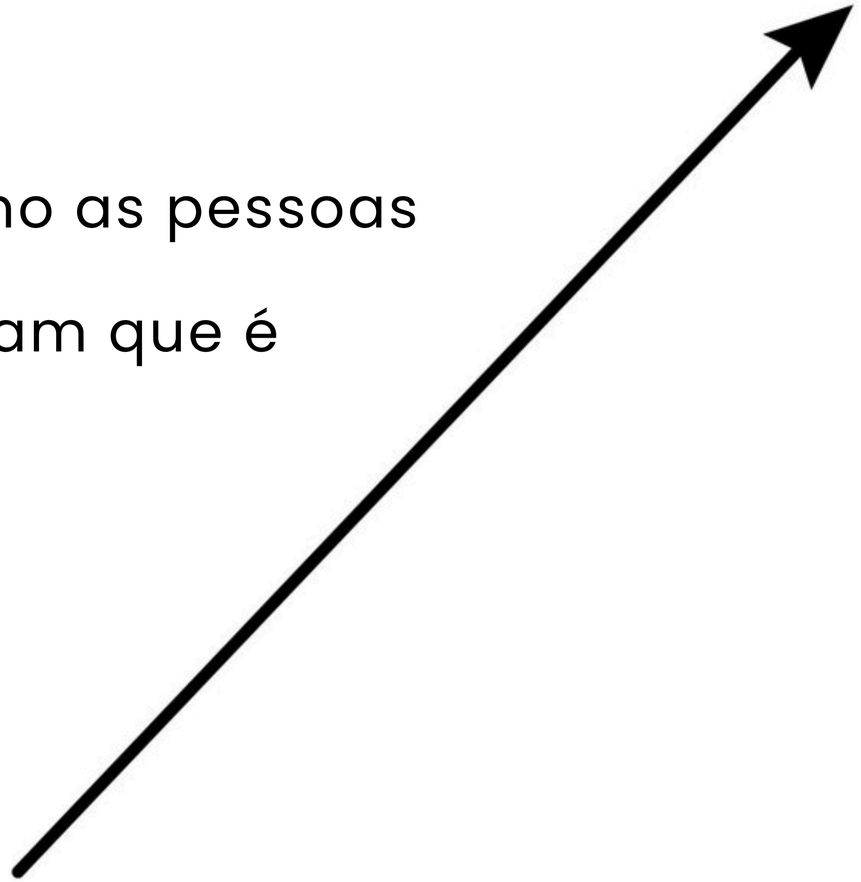
Trajetória

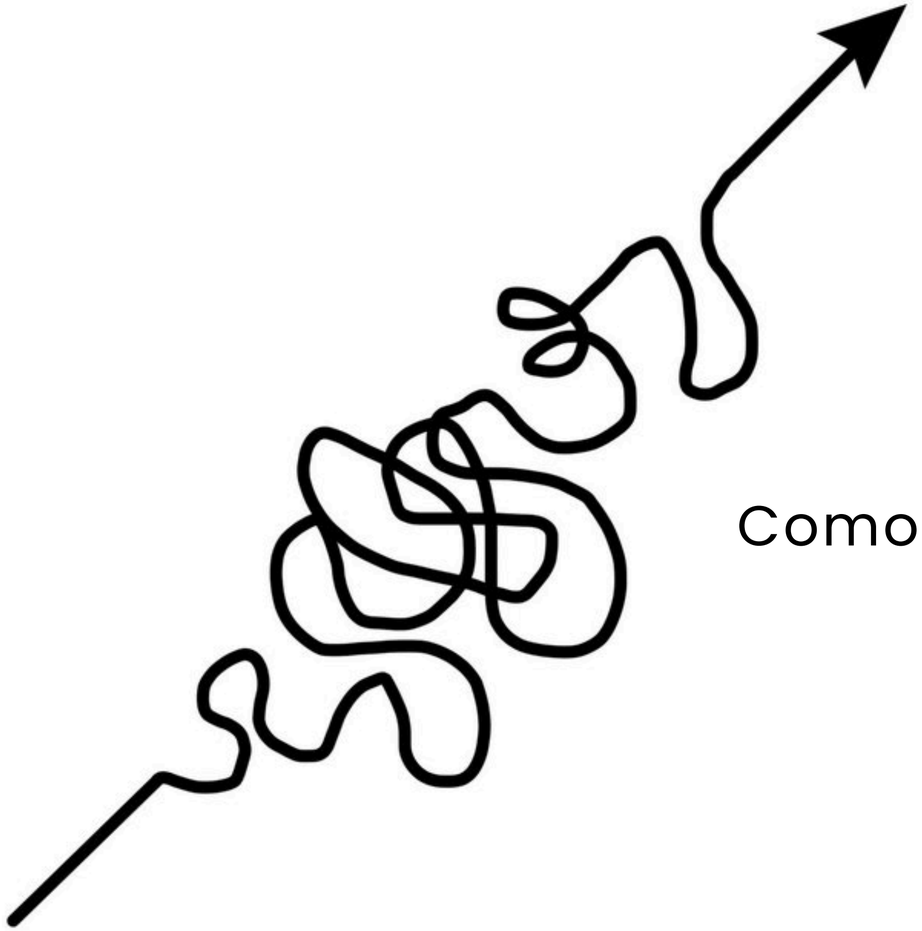
Tudo que você fez ou faz ao longo da sua vida importa.

Sucesso

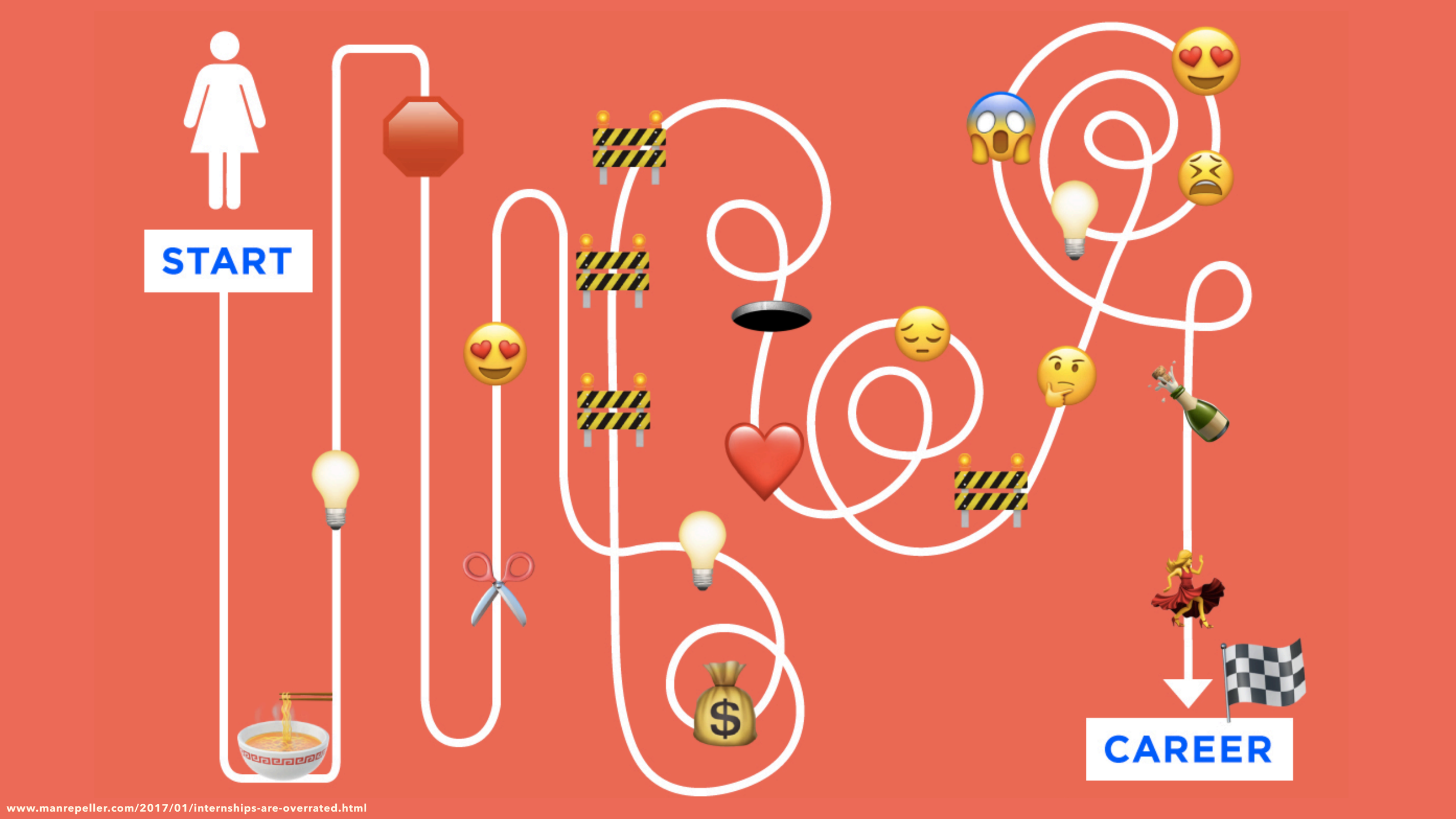
Como as pessoas

acham que é



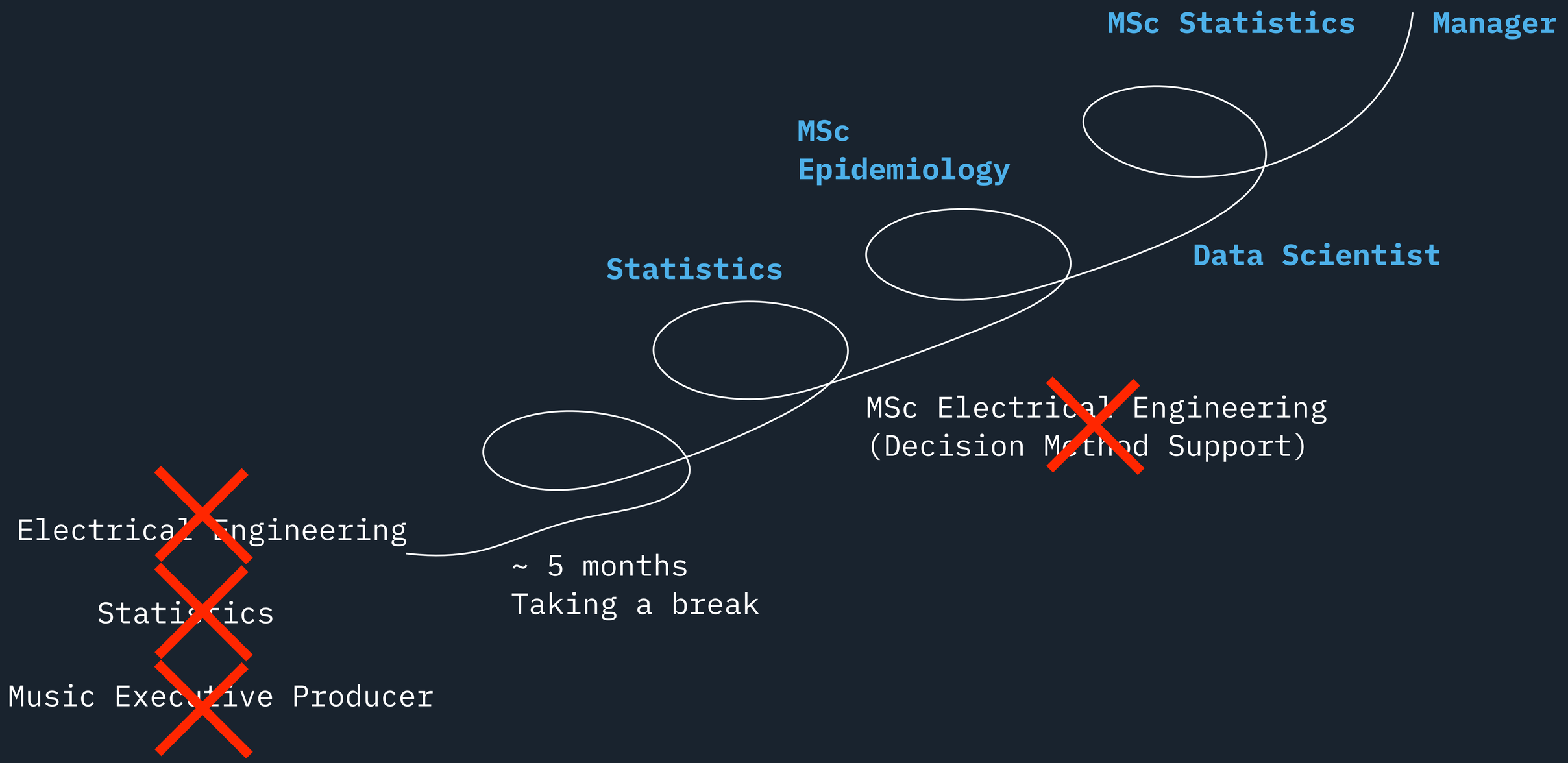


Como é na verdade



www.manrepeller.com/2017/01/internships-are-overrated.html

Trajetória

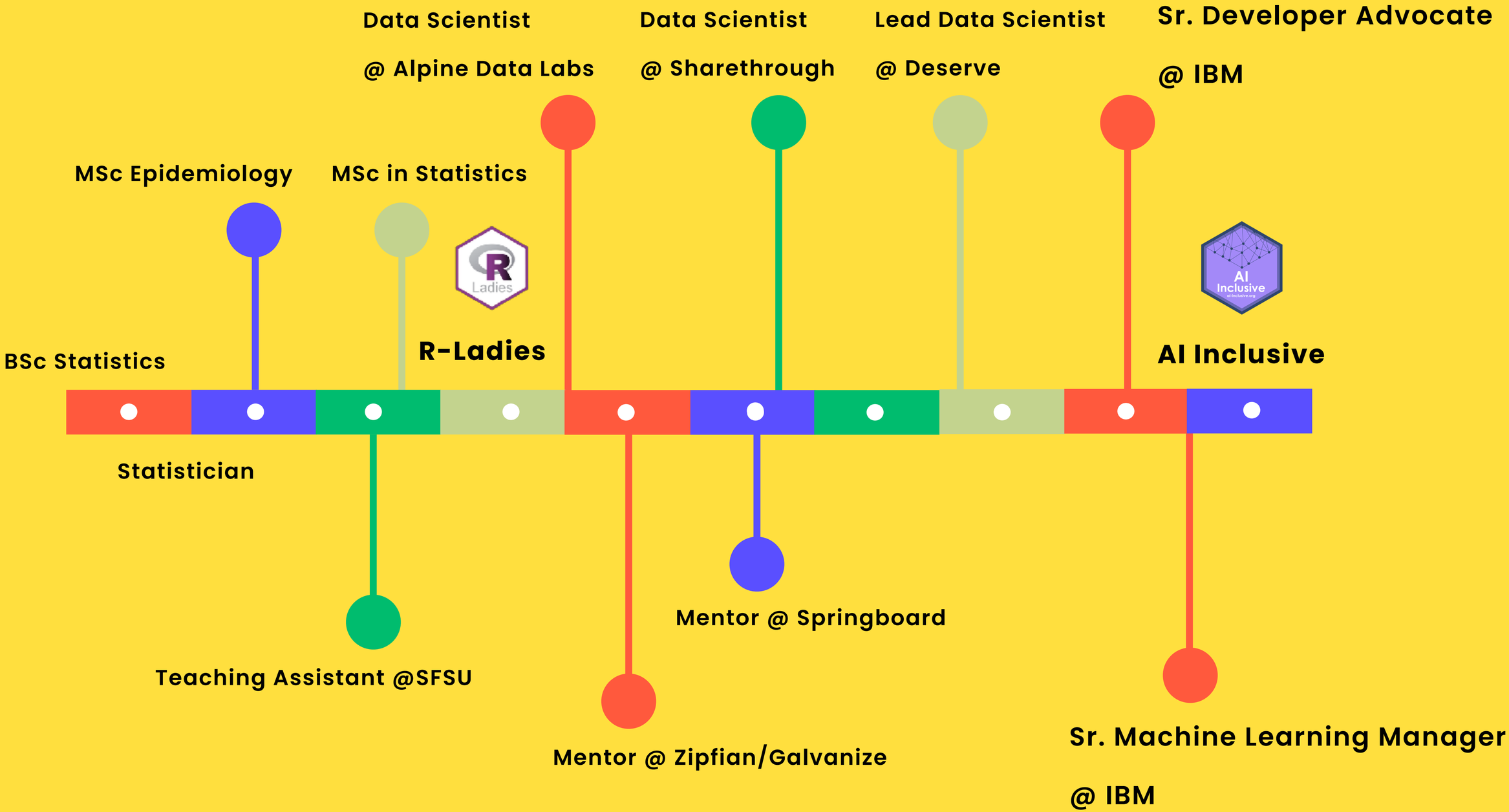

MSc Statistics
Manager
MSc
Epidemiology
Data Scientist
Statistics
MSc Electrical Engineering
(Decision Method Support)
Electrical Engineering
~ 5 months
Taking a break
Statistics
Music Executive Producer

Success is not a straight line, it’s much more of a dance and being open to possibilities.

O sucesso não é uma linha reta, é muito mais uma dança e estar aberto a possibilidades.

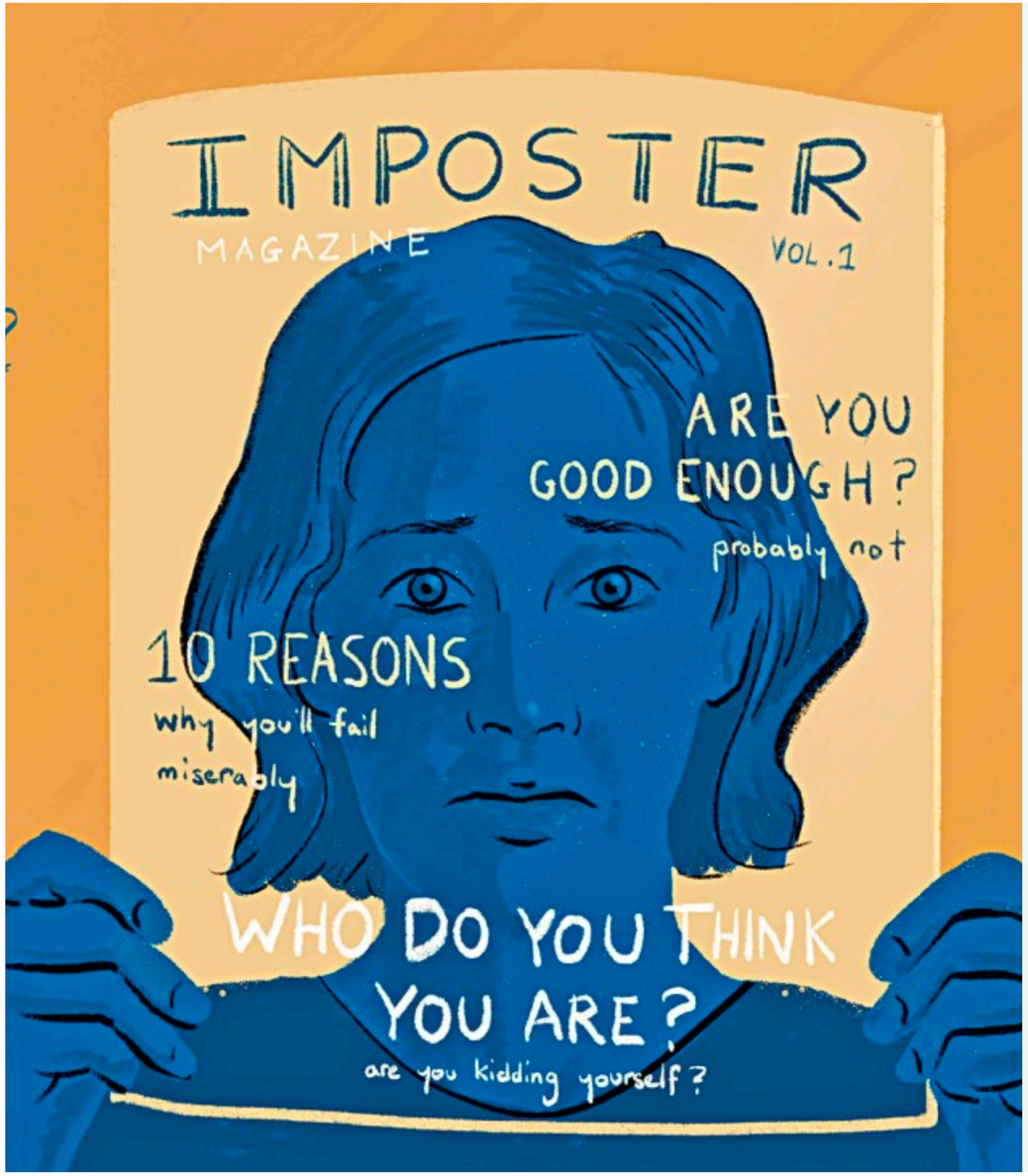
Arianna Huffington

AUTHOR, CO-FOUNDER OF THE HUFFINGTON POST, FOUNDER AND CEO OF THRIVE GLOBAL


Data Scientist
@ Alpine Data Labs
Data Scientist
@ Sharethrough
Lead Data Scientist
@ Deserve
Sr. Developer Advocate
@ IBM
MSc Epidemiology
MSc in Statistics
R Ladies
R-Ladies
AI Inclusive
AI Inclusive
BSc Statistics
Statistician
Teaching Assistant @SFSU
Mentor @ Springboard
Mentor @ Zipfian/Galvanize
Sr. Machine Learning Manager
@ IBM

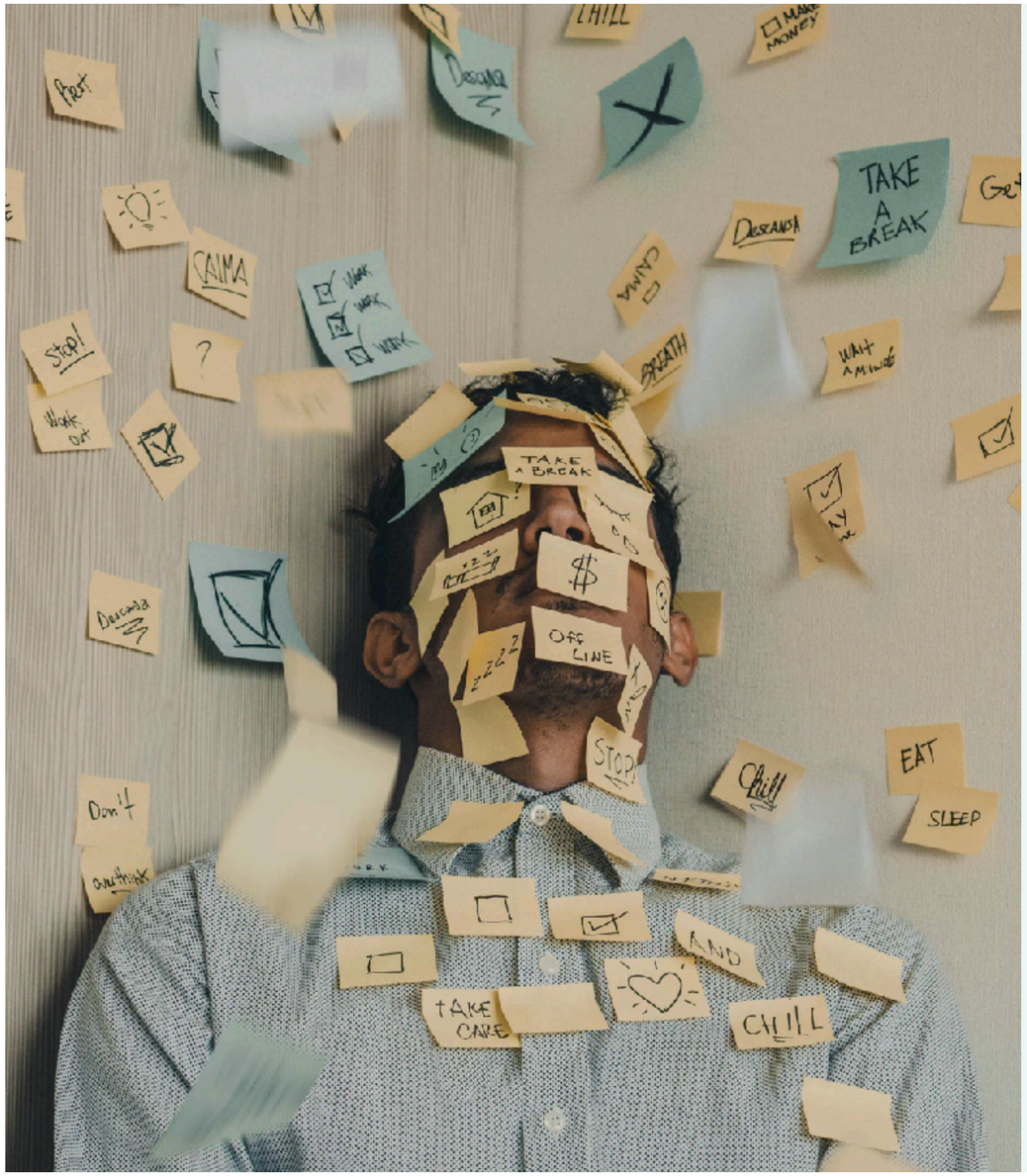
Desafios

There is no failure, only feedback



Síndrome do Impostor

É muito mais comum do que imaginamos.



Muito o que aprender

Como se manter por dentro de tudo que acontece?

THE MORE WE LEARN, THE LESS WE KNOW

Emprego

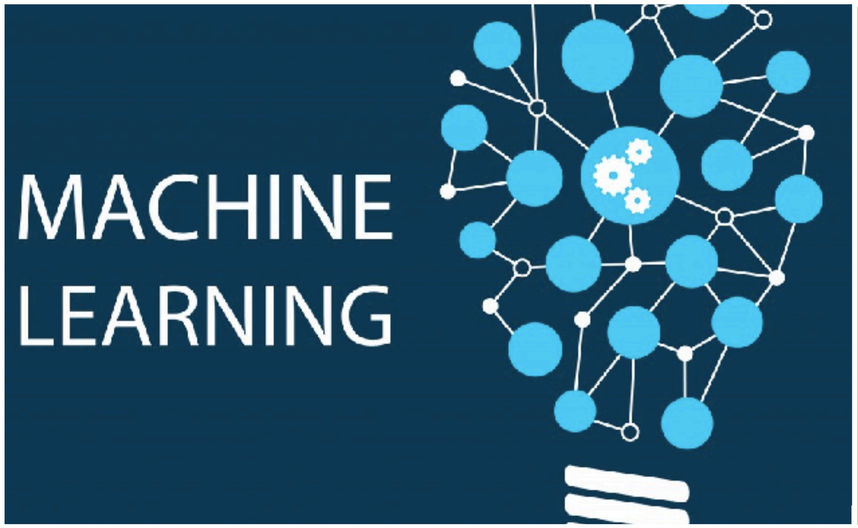
Você vai ser rejeitado, você vai receber muitos nãos, vão te dizer que você não é sabe o suficiente.

Dizer NÃO

Não significa que você vai perder oportunidade ou fechar portas, significa que você vai focar no que é importante para você e naquele momento.

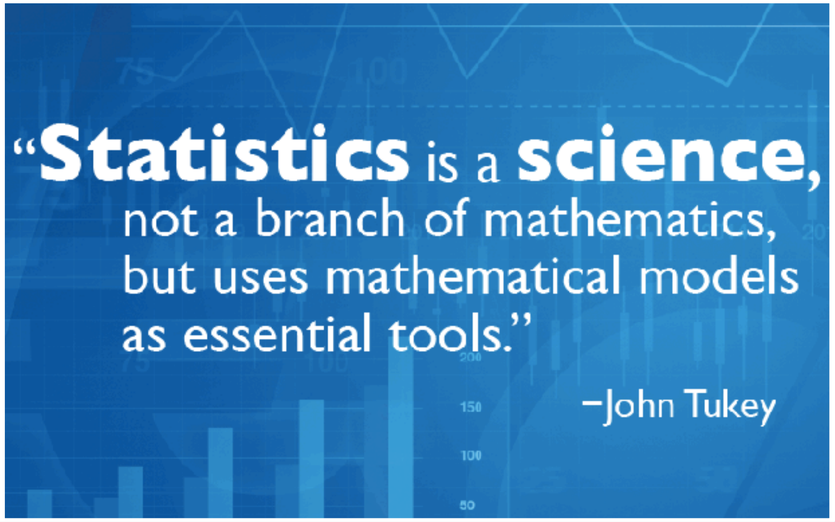
Dicas

No meio de tanta informação, de tanto material, de tanta novidade, por onde seguir? O que fazer?



MACHINE LEARNING

ESTATÍSTICA



PROGRAMAÇÃO

Comunicar

Flexibilidade

Curiosidade

Pensamento Crítico

Faça parte de grupos, comunidades e organizações

Mentores, Role Models, Sponsors, Advocates

Seja quem você é

Siga a gente:

bit.ly/ai-inclusive-instagram

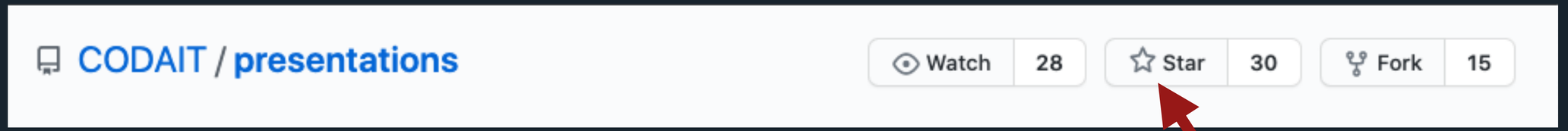
Materiais de Inteligência Artificial, Ciência de Dados, Machine Learning Eventos, Ingressos de graça e muito mais!

Em breve:

Bolsas de estudo para cursos de R, Python, Ciência de Dados, SQL, e outros.

Obrigada!

slides e material: bit.ly/cimatech-2020



Salve o nosso repositório com estrela

ai-inclusive.org

@gdequeiroz | linktr.ee/gdq