

USANDO ML PARA RESOLVER PROBLEMAS DE NEGOCIO



FAVIO
VÁZQUEZ
DATA SCIENTIST



@FavioVaz



@faviovaz

H₂O.ai

¿Quién soy?

- Venezolano
- Licenciado en Física e Ingeniero en Computación
- Maestro en Física por PCF-UNAM (Cosmología)
- Científico de datos en **H2O.ai**
- Miembro de la facultad e instructor en **Emeritus**
- Profesor en: UNAM, TEC de Monterrey, Dartmouth, Columbia y Berkeley
-

Favio Vázquez Following ▾

Data scientist, physicist and computer engineer. Love sharing ideas, thoughts and contributing to Open Source in Machine Learning and Deep Learning ;).



Medium member since November 2018 · Editor of Ciencia y Datos · Top writer in Artificial Intelligence

2 Following 8K Followers ·

- Muy activo en LinkedIn ;)
- Director ejecutivo y científico de datos jefe de **Closter**
- Creador de **Ciencia y Datos**
- Escritor en Towards Data Science, Heartbeat, Becoming Human y Planeta Chatbot
- Top escritor de Medium en negocios, inteligencia artificial y tecnología

H2O.ai Democratize AI

22

Kaggle Grandmasters

#1, #2, #5 y #9 del mundo

200K

H2O.ai
Miembros de Meetup

20K

Compañías Open Source



H2O.ai

Fundada en Silicon Valley, 2012

Fondos \$152M | Series D

Inversores Goldman Sachs, Ping An, Wells Fargo, NVIDIA, Capital One, Nexus Ventures

Plataformas de IA de clase mundial

H2O-3 Open Source Machine Learning at Scale

H2O Driverless AI Automatic Machine Learning

H2O Wave Open Source Low-code AI Framework

H2O AI Hybrid Cloud and AI-first App Store

Obsesionados con nuestros clientes

Aetna, CVS, Allergan, AT&T, Capital One, HCA, Citi, Coca Cola, Nationwide, Franklin Templeton, Kaiser Permanente, State Farm, Pepsi, Pacific Life, Roche, Unilever, Walgreens, Bank of America, UPS, PayPal, MasterCard, Progressive, Tokio Marine, Chipotle

Visionario Reconocido

H2O.ai

Homogeneidad de la
plataforma

App Store

Líder de pensamiento de
AutoML

Gartner Magic Quadrant para Cloud AI Developer Services (2021)

Figure 1: Magic Quadrant for Cloud AI Developer Services



Gartner Magic Quadrant para Data Science y Plataformas ML (2021)

Figure 1: Magic Quadrant for Data Science and Machine Learning Platforms



Creación de valor

Ampliación ampliada

Inteligencia artificial
rica y explicable

H2O AI HYBRID CLOUD

Una plataforma. Miles de casos de uso.



MODELO

AutoML
extensible y
abierto



EXPLICAR

Interpretabilidad
automatizada y
AutoDoc



OPERAR

Implementar en
cualquier lugar,
gobernar y
mejorar



DESARROLLAR

Marco de
desarrollo ligero
de Python



CONSUMIR

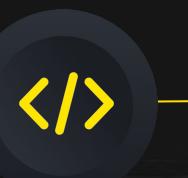
Comunidad
APIs
AppStore

Científicos de
Datos

Comentarios iterativos y desarrollo
rápido a lo largo de todo el ciclo
de vida de ML.



Operaciones



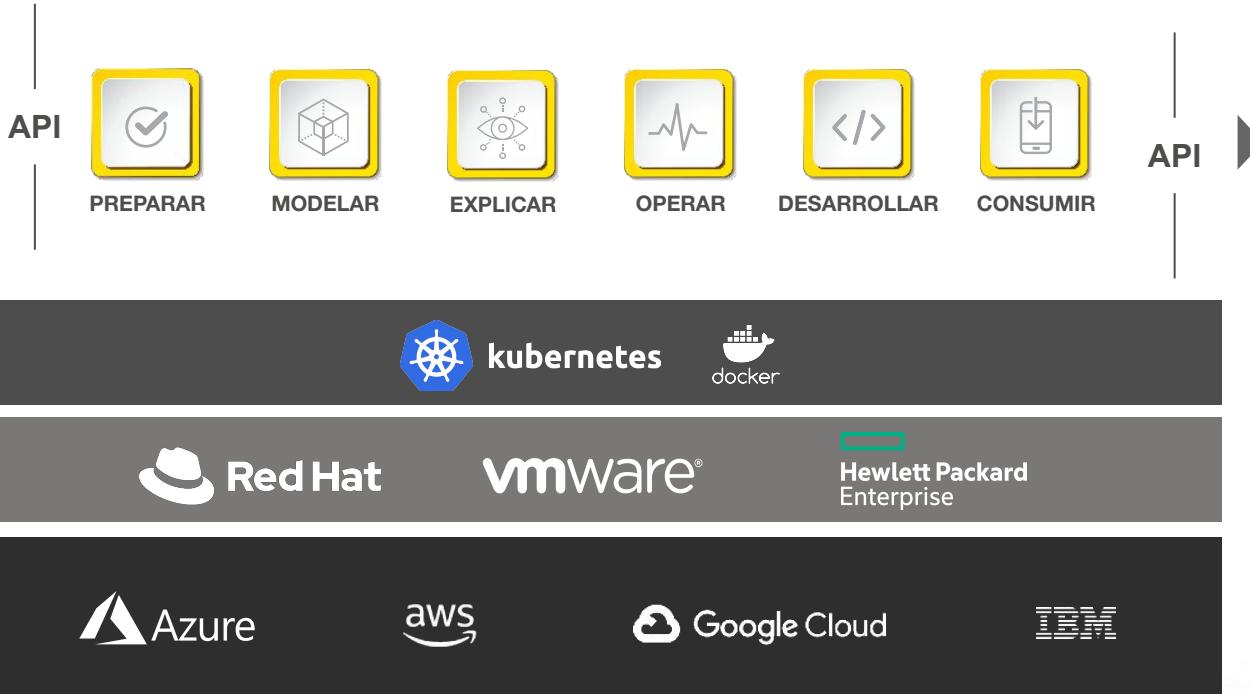
Desarrollador

Usuario
comercial

H2O AI Hybrid Cloud Architecture

snowflake
databricks
CLOUDERA
teradata.

+200 fuentes
de datos



Aplicaciones
de negocios



Onda H2O

H2O AI Hybrid Cloud Architecture

H2O.ai

Capa de aplicaciones de IA



Capa de servicios de IA

Machine Learning



Capa de contenedor



Kubernetes



docker

Capa de infraestructura



Contenido

- Motivación
- Conceptos importantes
- Timeline de la ciencia de datos
- Crear un producto de datos (teoría)
- **Problema y solución de ciencia de datos (taller)**

Motivación: Los productos de la Ciencia de Datos son complejos

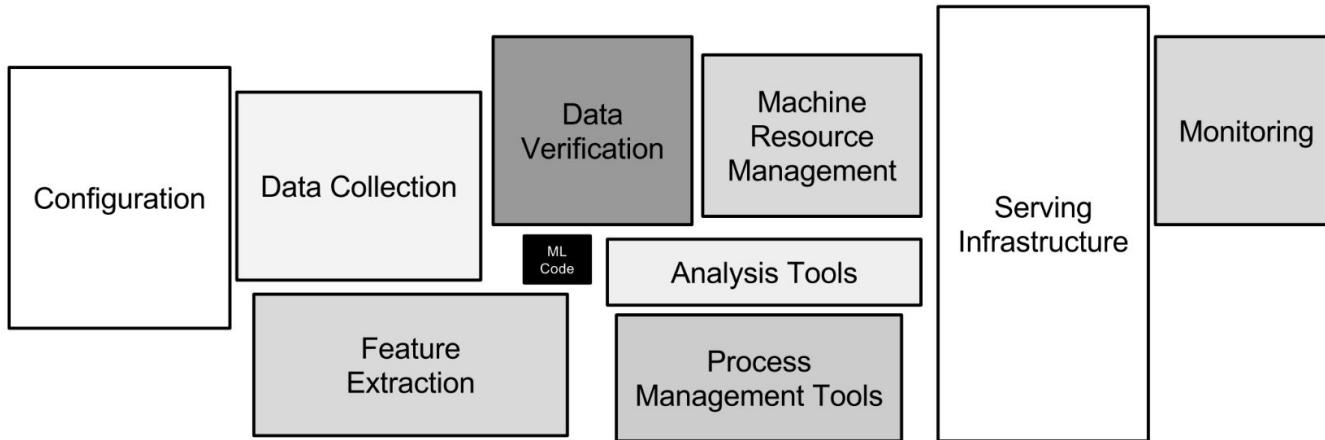
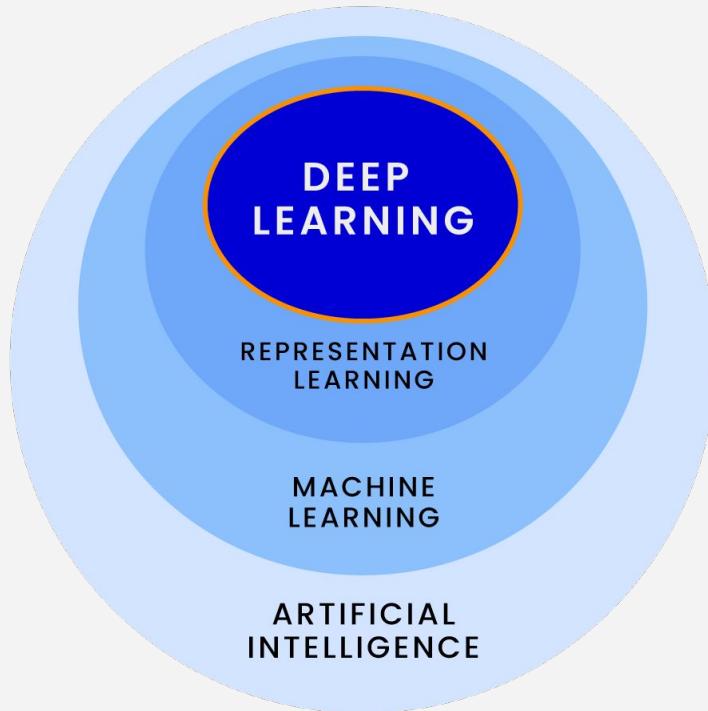


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

Conceptos importantes

Panorama de IA



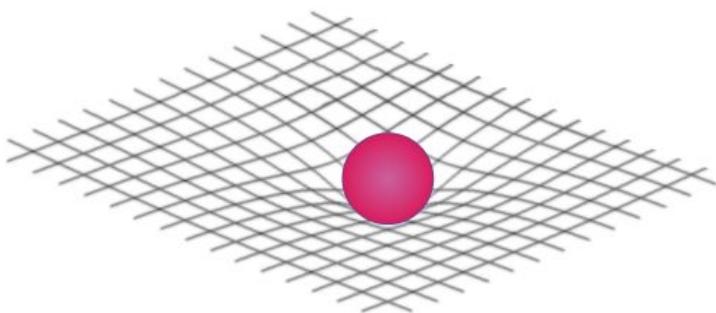
Inteligencia Artificial (IA)



Boston Dynamics

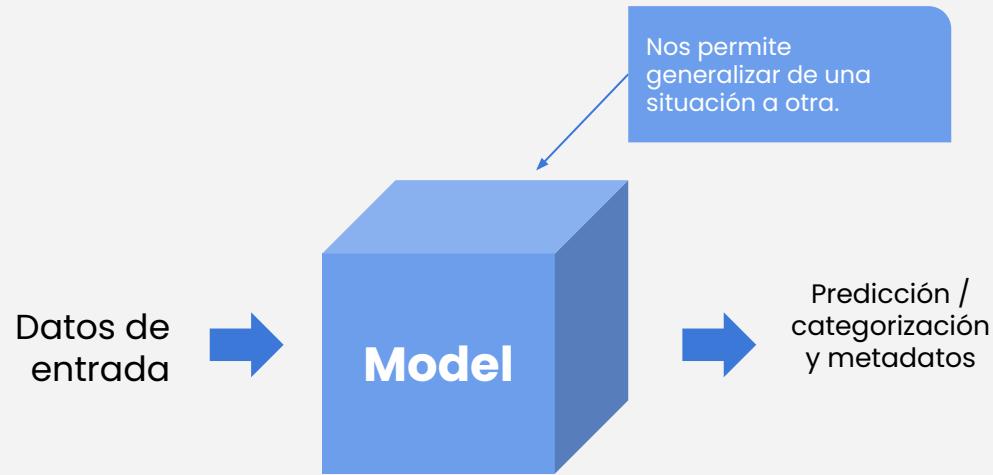
Procesos automatizados para emular las habilidades y capacidades humanas con la computación.

Modelo

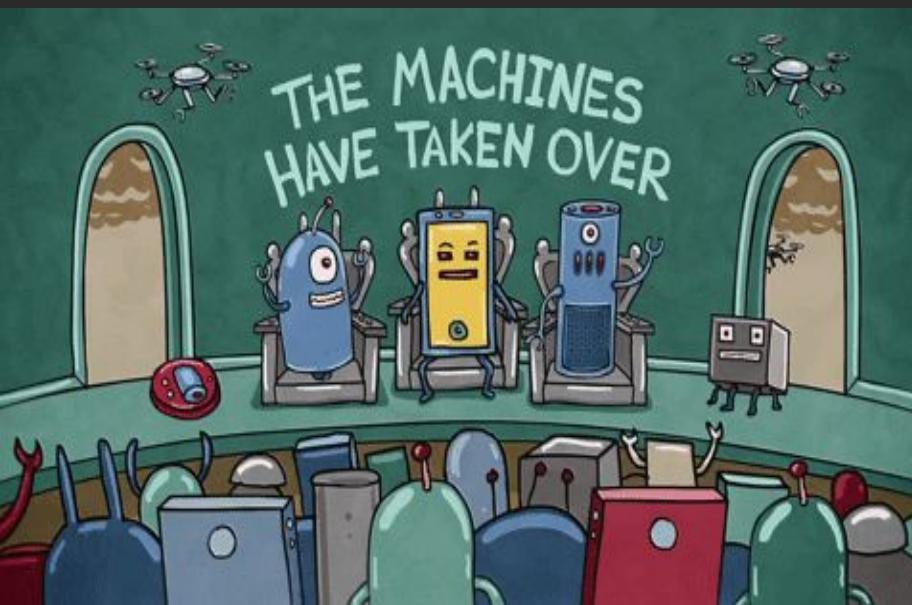


$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$$

Abstracción de la realidad para comprender un proceso o fenómeno utilizando herramientas matemáticas.



Machine Learning

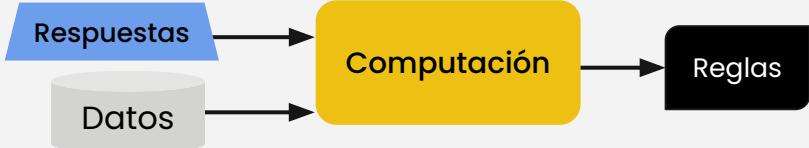


Descubrimiento de patrones a partir de datos para predecir el funcionamiento de un proceso o fenómeno utilizando algoritmos que no están programados explícitamente.

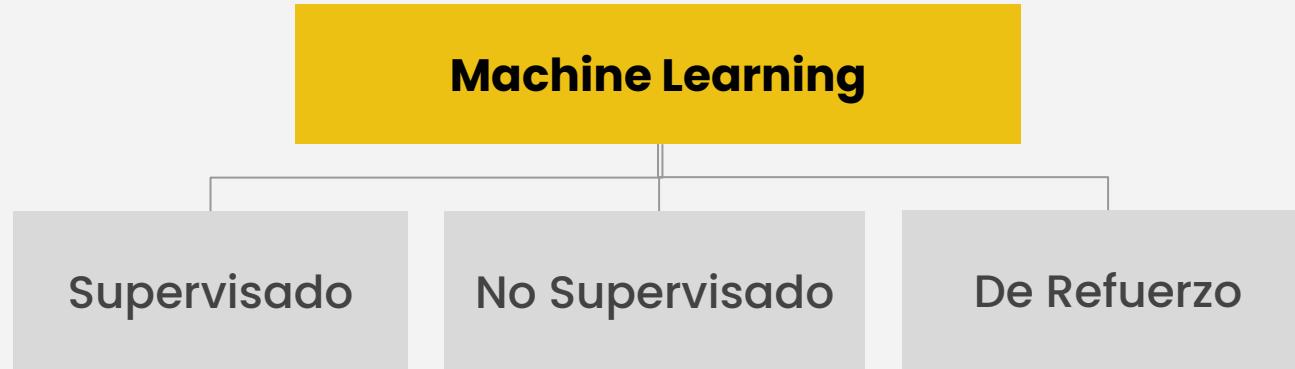
Programación Tradicional



Machine Learning



Tipos de Machine Learning



impulsado por tareas
(predecir el siguiente valor)



impulsado por datos
(identificar clústeres)



Aprender de los errores

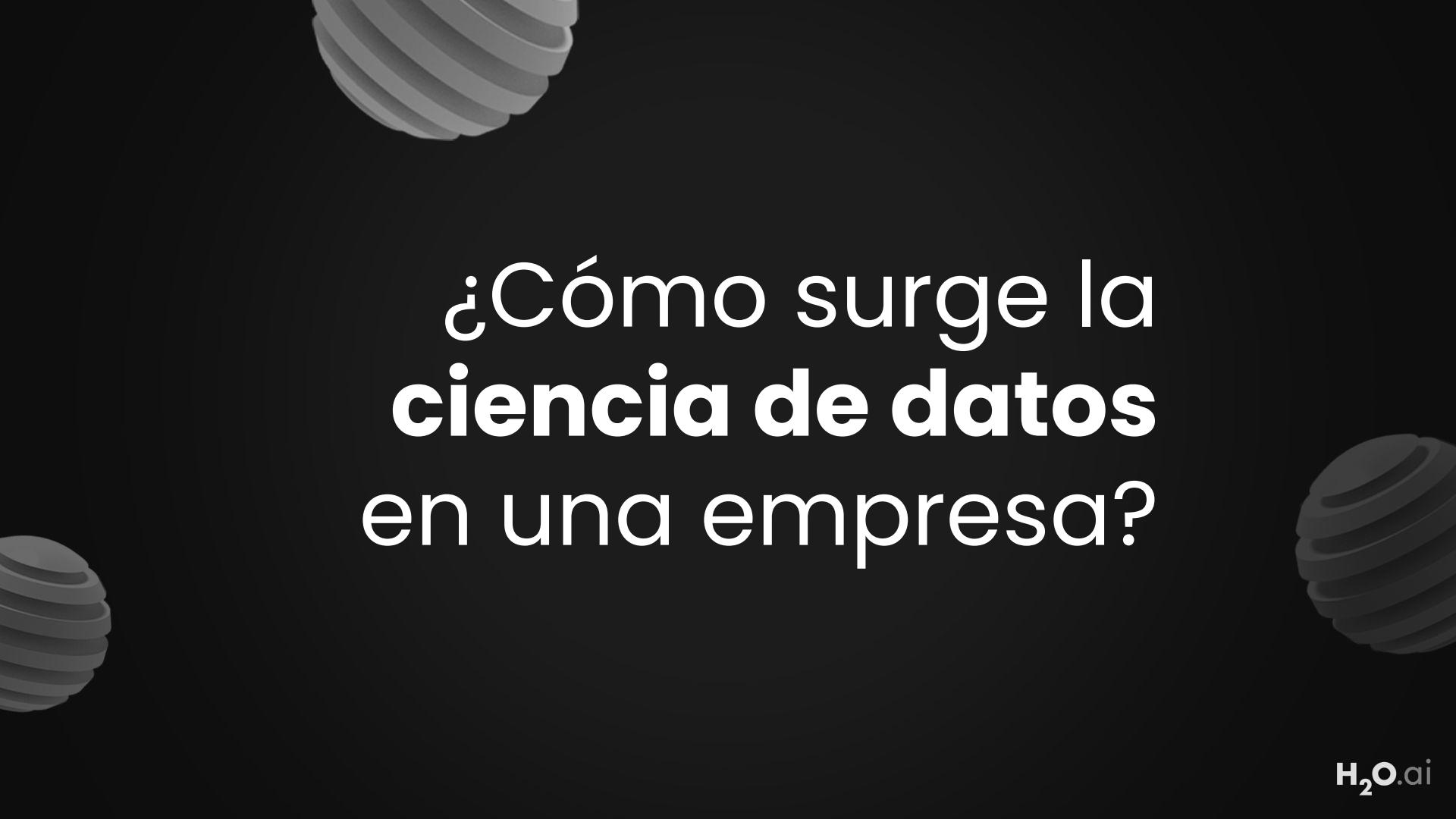
methodology
sciences development effort areas
evaluation **support project**
software center engineering campus
collaboration space **director**
impact **working** career
students reproducibility
studio **science** groups
budget evaluation
institute faculty work tools statistics
university environment **scientists**
activities new domain
education **research**
graduate program

EL PAPEL DE LA CIENCIA DE DATOS

La ciencia de datos es el mediador en el camino de usar la IA para impactar en negocios.



Tal vez el nombre "ciencia de datos" no sea muy cómodo para todos, pero intentaré demostrar que podemos sacarle provecho por ahora.



¿Cómo surge la **ciencia de datos** en una empresa?

NATURALEZA



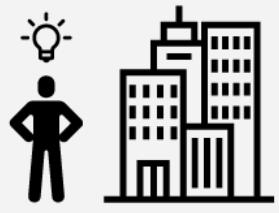
Comunicación



10100
00101
10100



GENTE DE TI



JEFE DE JEFES

ILUSTRATIVO

Tenemos
muchos datos,
tenemos que
hacer algo al
respecto.

ILUSTRATIVO

NATURALEZA



...



10100
00101
10100



GENTE DE TI

CREACIÓN DE NUEVAS ÁREAS



JEFE DE JEFES



TRANSFORMACIÓN DIGITAL



GERENTE DE CIENCIA DE DATOS



CDO



LABORATORIOS DE BIG DATA



JEFE DEL CeO DE CIENCIA DE DATOS

ILUSTRATIVO

NATURALEZA



...



10100
00101
10100



GENTE DE TI



Jefe de Jefes

PROBLEMAS DE NEGOCIO



PRIORIZACIÓN

TRANSFORMACIÓN DIGITAL



GERENTE DE CIENCIA DE DATOS



LABORATORIOS DE BIG DATA



GENTE DE TI

Jefe del CeO de CIENCIA DE DATOS

CDO

ILUSTRATIVO

NATURALEZA



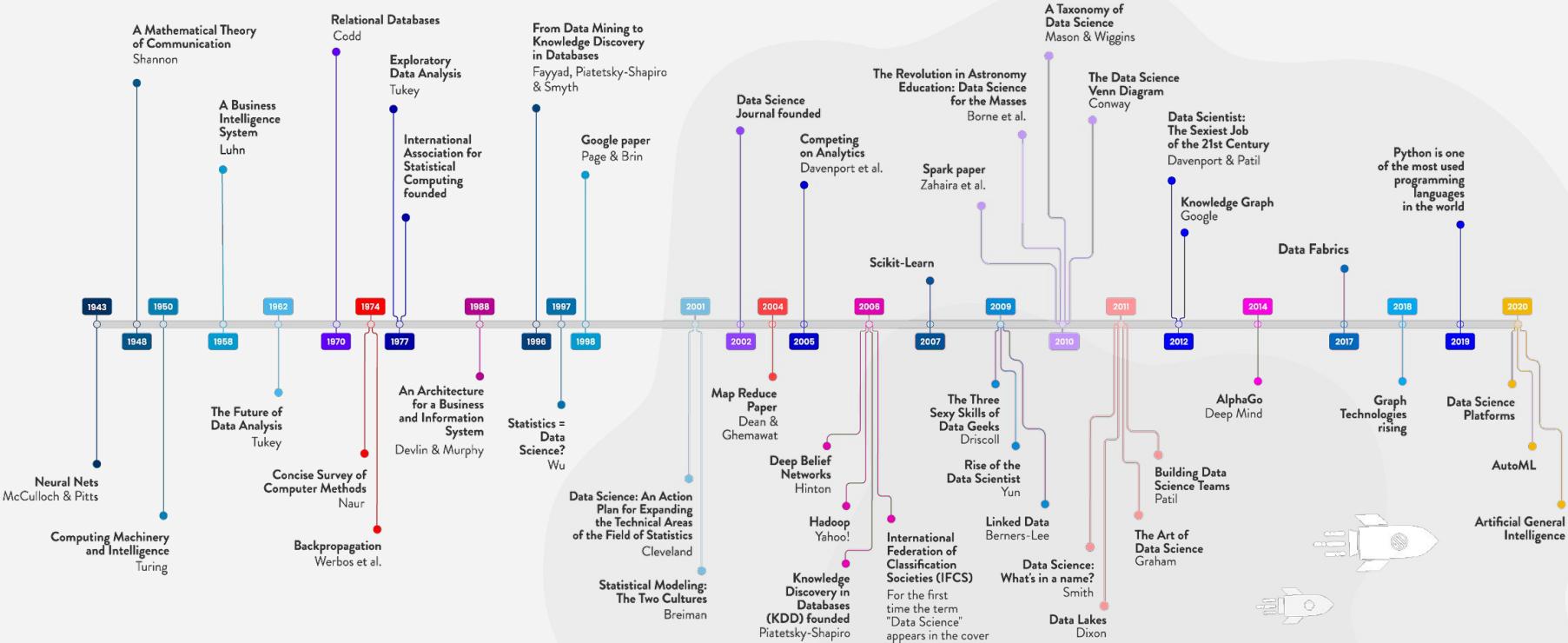
10100
00101
10100

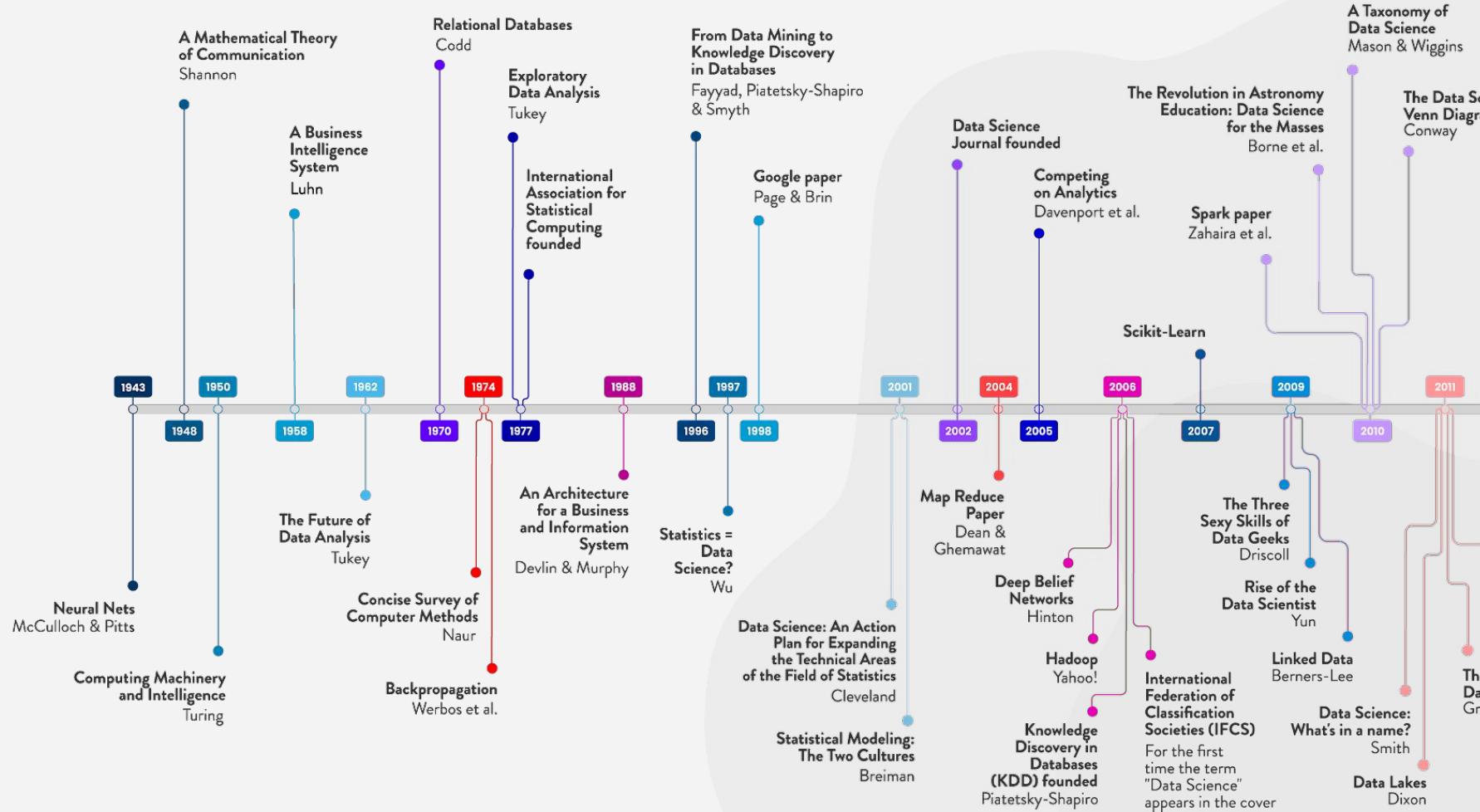


GENTE DE TI



DATA SCIENCE TIMELINE





From Data Mining to
Knowledge Discovery
in Databases

Fayyad, Piatetsky-Shapiro
& Smyth

Google paper
Page & Brin

Data Science
Journal founded

Competing
on Analytics
Davenport et al.

The Revolution in Astronomy
Education: Data Science
for the Masses
Borne et al.

Spark paper
Zaharia et al.

A Taxonomy of
Data Science
Mason & Wiggins

The Data Science
Venn Diagram
Conway

Data Scientist:
The Sexiest Job
of the 21st Century
Davenport & Patil

Python is one
of the most used
programming
languages
in the world

1996
1998

Statistics =
Data
Science?
Wu

Data Science: An Action
Plan for Expanding
the Technical Areas
of the Field of Statistics
Cleveland

Statistical Modeling:
The Two Cultures
Breiman

Map Reduce
Paper
Dean &
Ghemawat

Deep Belief
Networks
Hinton

Hadoop
Yahoo!

Knowledge
Discovery in
Databases
(KDD) founded
Piatetsky-Shapiro

Scikit-Learn

The Three
Sexy Skills of
Data Geeks
Driscoll

Rise of the
Data Scientist
Yun

Linked Data
Berners-Lee

Data Science:
What's in a name?
Smith

Data Lakes
Dixon

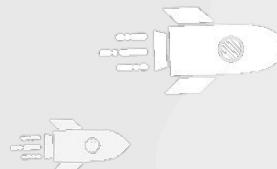
International
Federation of
Classification
Societies (IFCS)
For the first
time the term
"Data Science"
appears in the cover

Data Fabrics

Graph
Technologies
rising

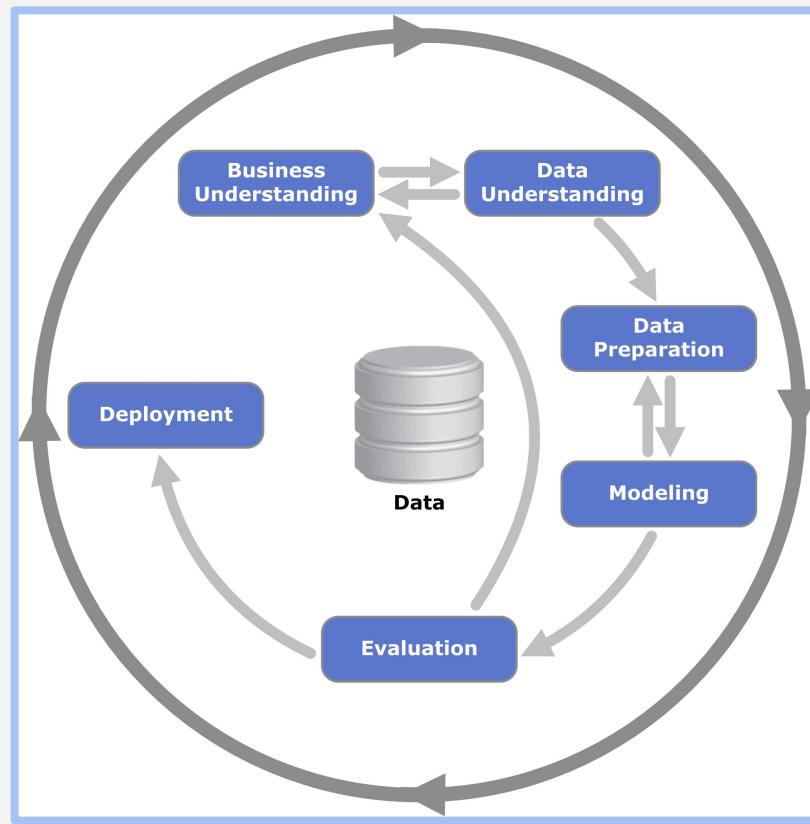
Data Science
Platforms

AutoML
Artificial General
Intelligence

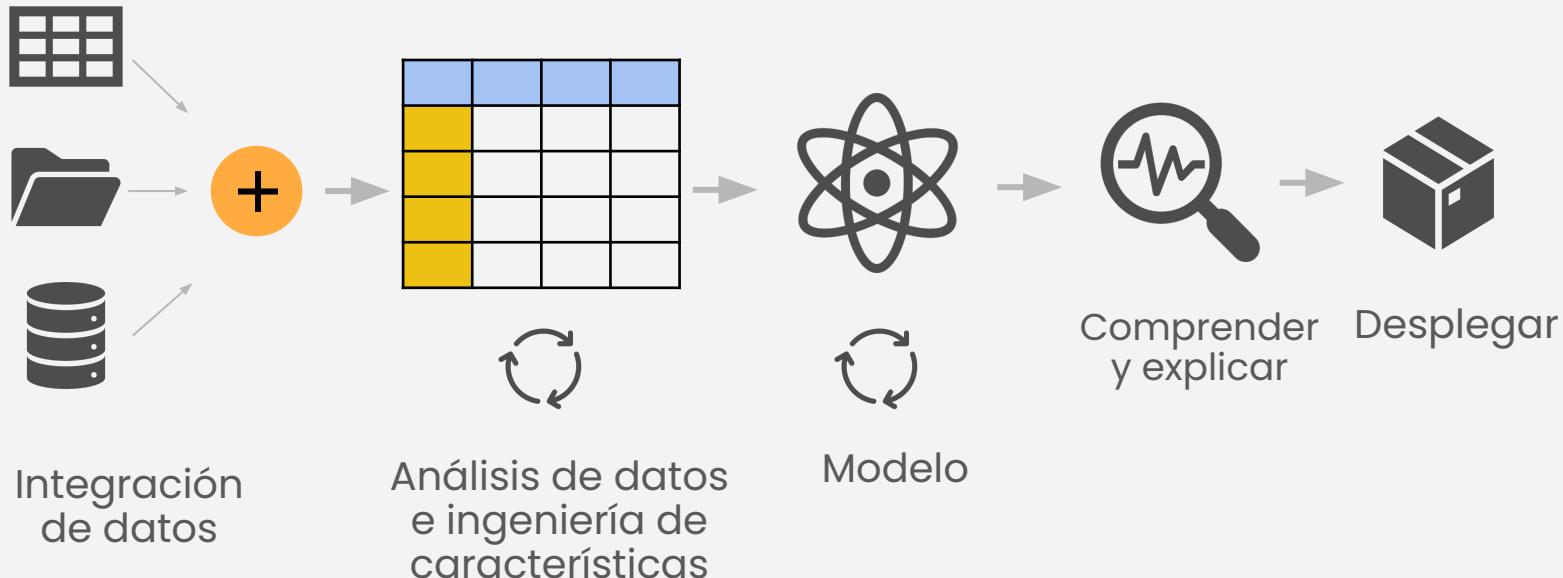


Productos de Ciencia de Datos

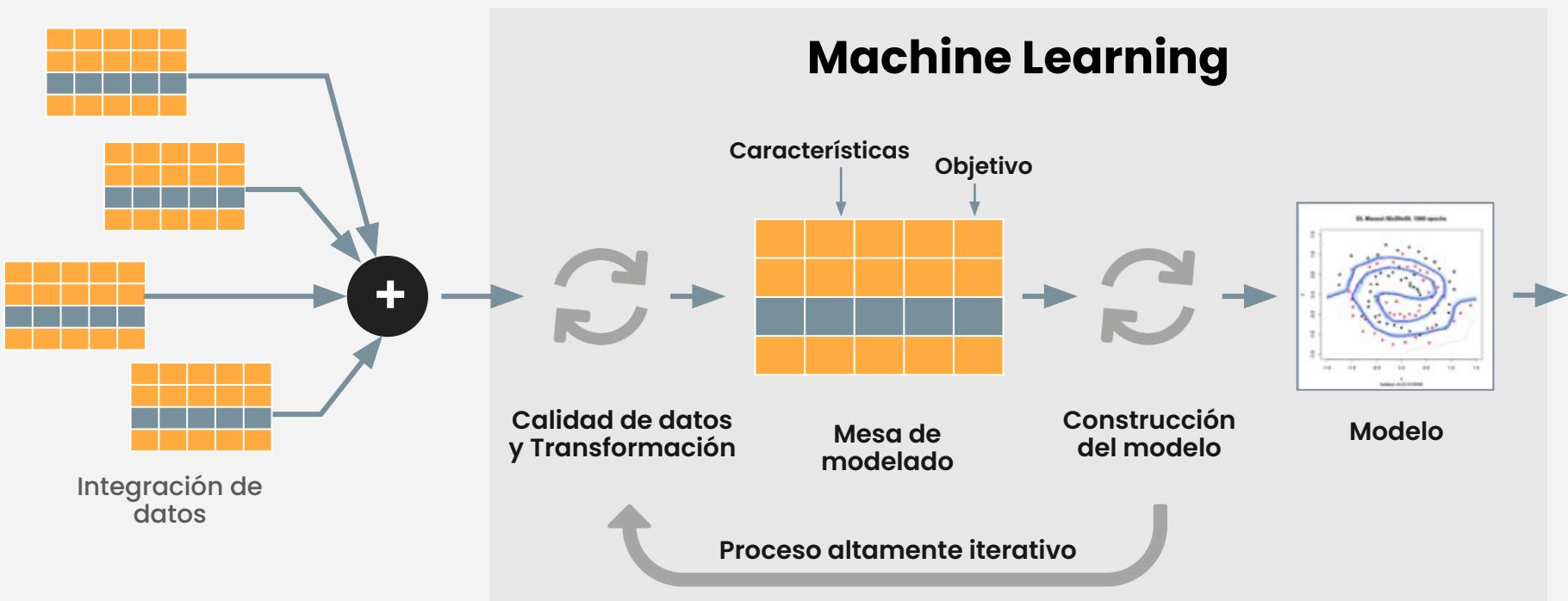
CRISP-DM (Uno de los padres de la CD)



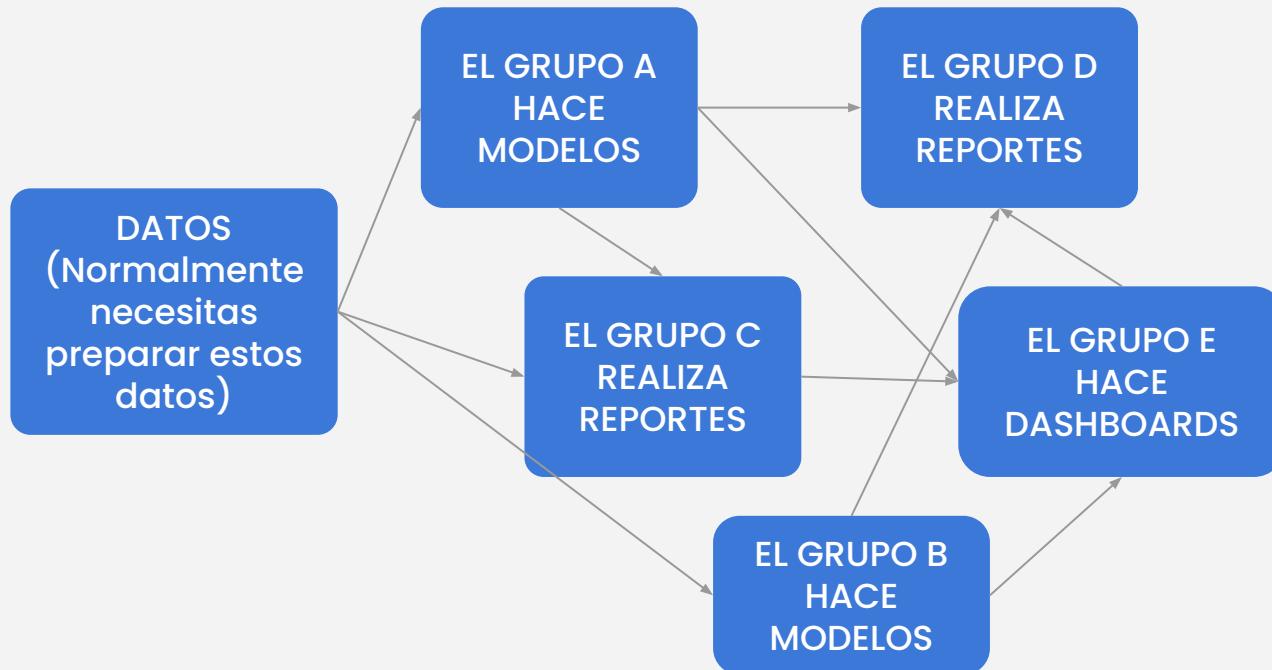
Workflow de la Ciencia de Datos



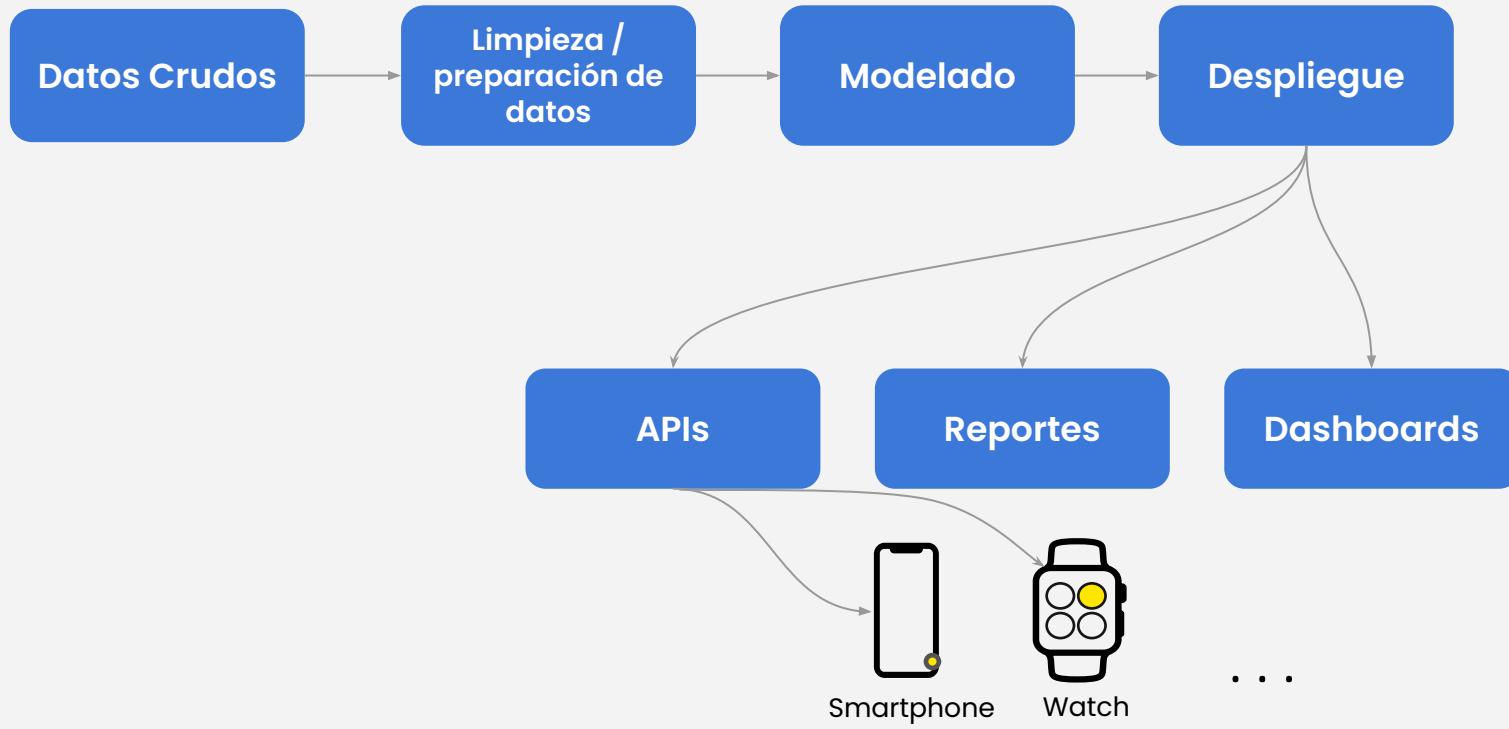
Workflow de la Ciencia de Datos



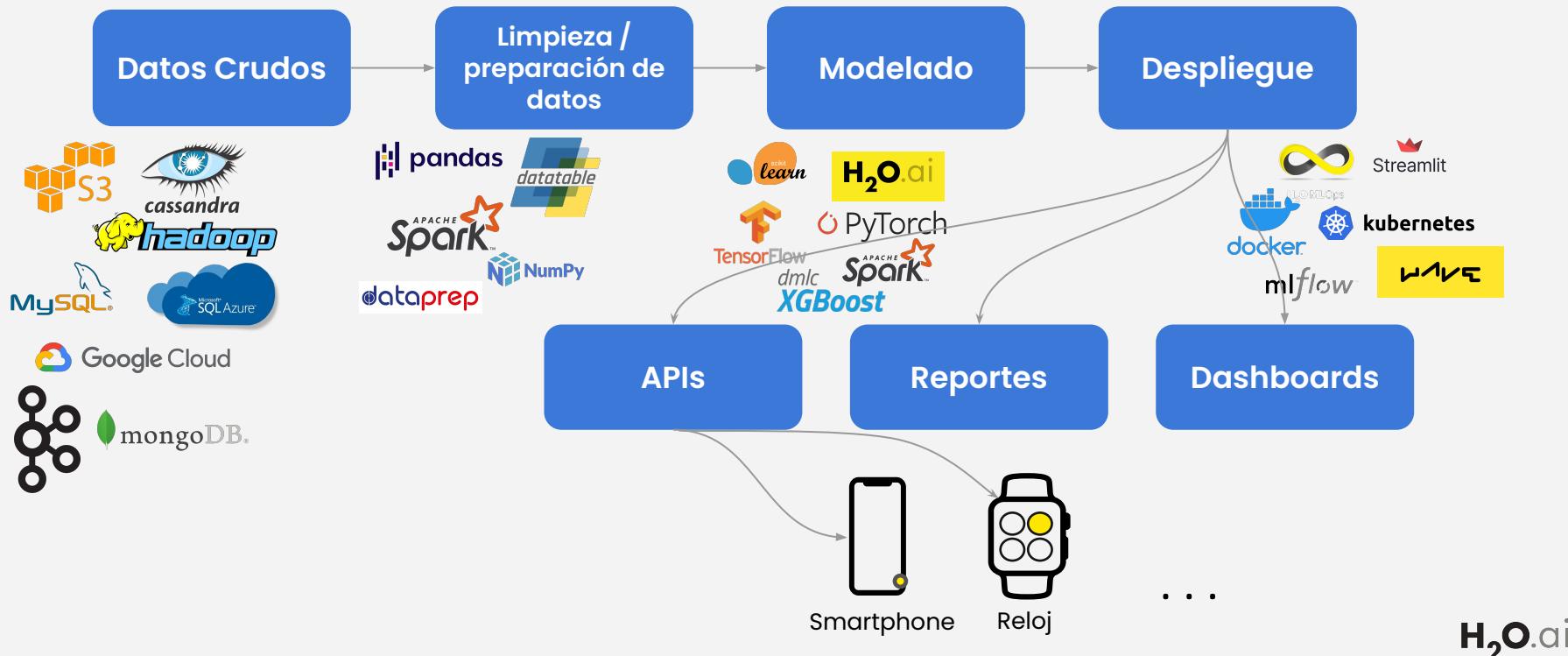
Los productos de la Ciencia de Datos son complejos



Data science products are complex



Data science products are complex



Data science products are complex

Raw Data

Data cleaning
/ preparation

Modeling

Deploy

Además debemos agregar:



**Versionado
Tracking
Métricas
Evaluación**

APIs

Reports

Dashboards



Smartphone

Watch



Demo links

<https://github.com/FavioVazquez/ML-H2O-Webinar-0821>

<https://aquarium.h2o.ai/>