

Gestión de datos

Motivación

Lorena Etcheverry (lorenae@fing.edu.uy)
Instituto de Computación, FING, UdelAR

¿quién soy yo?

Lorena Etcheverry Venturini

Dra. en Informática (PEDECIBA Informática - UdelaR), Ingeniera en Computación (FING- UdelaR)

Trabajo en el Instituto de Computación de la FING dentro del grupo GEMA (Gestión, Modelado y Análisis de Datos).

Mi área de interés es la gestión de datos, en particular las bases de datos de grafos, la calidad de datos, y los temas vinculados a la ética en ciencia de datos como la privacidad, anonimización y la detección de sesgos.

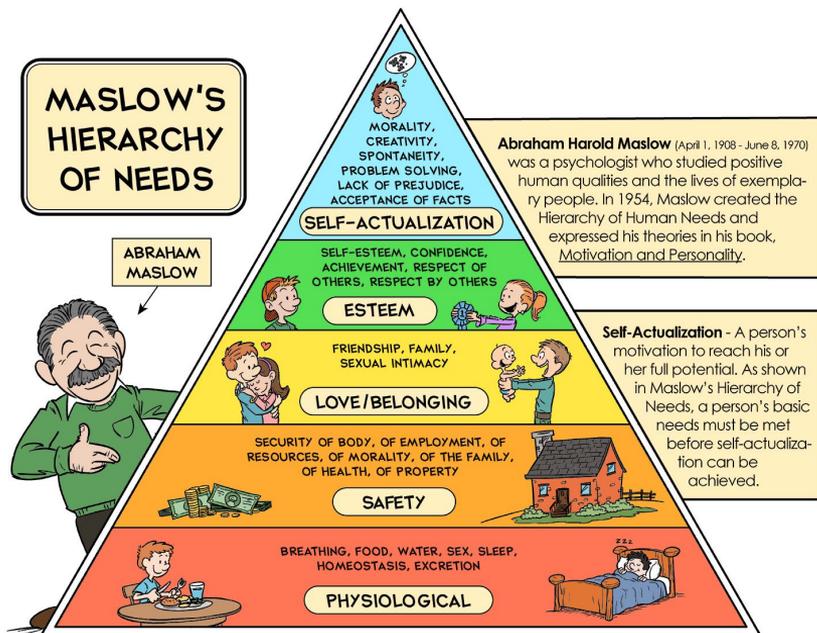
Me gustaría que cada uno de uds se presente

Respondiendo al hilo “Presentación personal” en el foro

Yo soy científic@ de datos

¿por qué debería preocuparme esto de la gestión de datos?

Es imposible hacer ciencia de datos sin datos 😊



www.timvandevall.com | Copyright © 2013 Dutch Renaissance Press LLC.

THE DATA SCIENCE HIERARCHY OF NEEDS

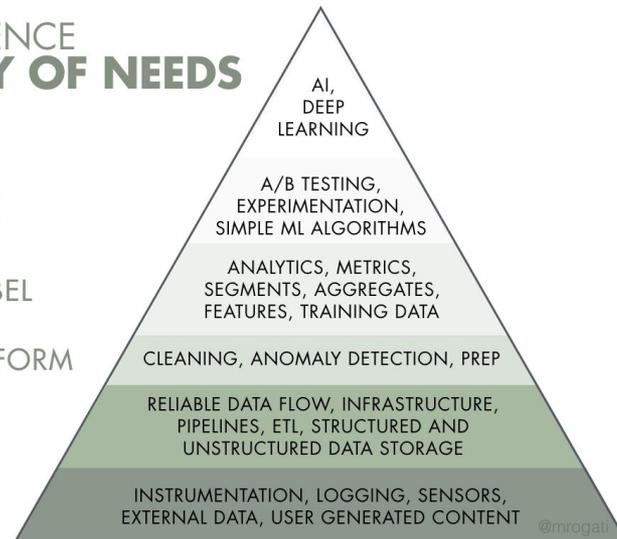
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

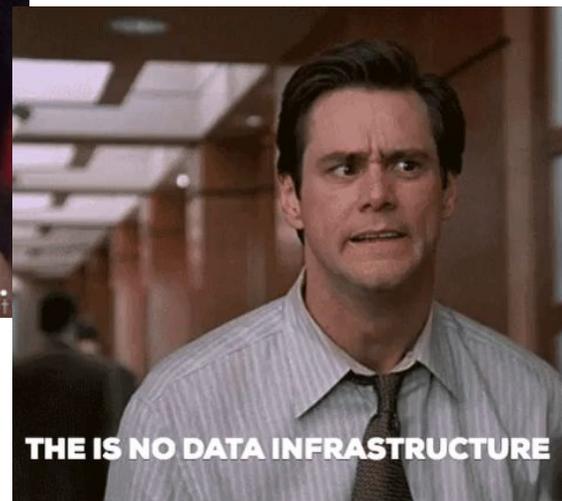


[Mónica Rogatti, The AI hierarchy of needs](#)

LAS EXPECTATIVAS EN CIENCIA DE DATOS



LAS REALIDAD EN CIENCIA DE DATOS



¿ENTONCES?

Muchas veces, las empresas/organizaciones intentan comenzar por la punta de la pirámide debido a diferentes razones

- la presión por obtener resultados ya!!
- la poca madurez y/o la gestión de datos desorganizada (a veces bajo nivel de *[data literacy](#)*)
- la falta de conocimiento

Los científicos de datos quedan en el lugar de 



Do data scientists spend 80% of their time cleaning data? Turns out, no?

 Leigh Dodds  Data  January 31, 2020  4 Minutes

<https://blog.ldodds.com/2020/01/31/do-data-scientists-spend-80-of-their-time-cleaning-data-turns-out-no/>

The Data Scientist Unicorn



Joel Quesada [Follow](#)

Jan 10, 2019 · 3 min read



A Data Scientist Unicorn during his interview

[Joel Quesada, The Data Scientist Unicorn](#)

Conclusión

Como científic@s de datos, su capacidad de extraer valor de los datos está directamente relacionada con la madurez de la plataforma de datos de la organización/empresa.

*Si la organización no tiene una infraestructura de datos madura, van a necesitar **habilidades de gestión de datos** (a.k.a data engineering).*

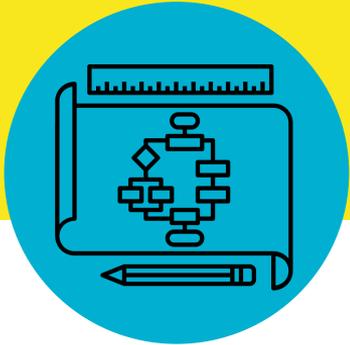


Objetivos de esta parte del curso

Brindarles definiciones y conceptos generales sobre gestión de datos y *data engineering*.

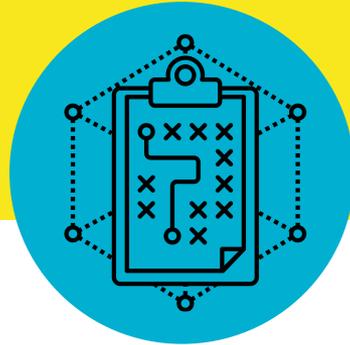
Desarrollar capacidades básicas, y proveer referencias para que uds. puedan profundizar.

Temas de gestión de datos que abordaremos



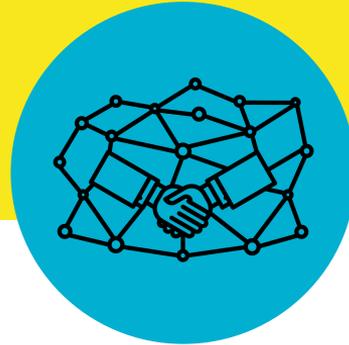
Modelos de datos y modelado

Por que no todo es una planilla de cálculo



Calidad de datos y limpieza

Data profiling
Conceptos de
Calidad de Datos



Integración de datos

¿cómo se integran datos de diferentes fuentes?
Algunas ideas y herramientas



Aspectos éticos, privacidad y sesgo

Privacidad y anonimización
Modelos y sesgo