



# TESTING PARA CALIDAD DE DATOS

Jornadas de Calidad de Software y Agilidad 2025  
FCEQyN - UNaM | FaCENA - UNNE | FRRe - UTN

# .self()

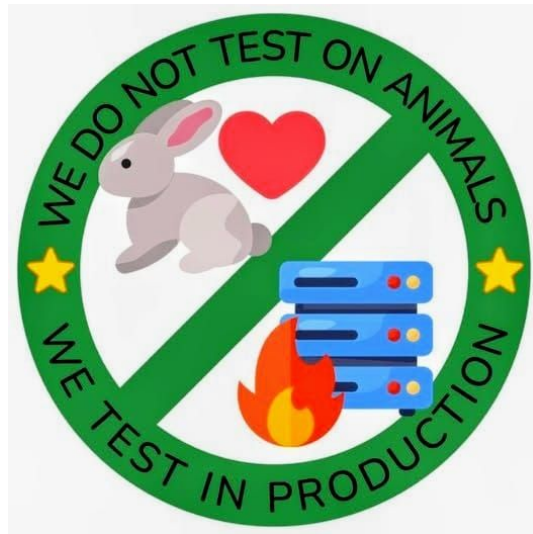
## Martín Rey

Graduado FCEQyN - UNaM x2.

Docente en FCEQyN - UNaM hace 15+ años en las áreas de Ingeniería de Software | Auditoría Informática | Programación (SL1 y SL2) | Datos + Agilidad (*líneas de investigación*)

Nodocente en FCE - UNaM en la Dirección de Tecnologías para la Gestión trabajando con Software + Datos + *otras cosas*





# AGENDA

01

Intro

02

Propuestas

03

Demo

04

Cierre (?)



01

# INTRODUCCIÓN



## Solo software?

En general, cuando se habla de testing se suele asociar a probar el comportamiento de un producto software.

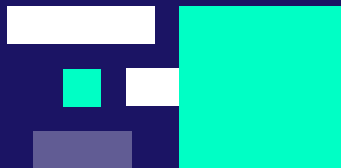
Ahora bien, los modelos de calidad dicen que el **software es más que solo su código.**

# ¿Por qué hacer testing de los datos?

Los datos van a ser **usados en el software** para brindar reportes que se van a utilizar para la toma de decisiones.

E incluso, podrían usarse en soluciones más elaboradas que involucren **BI, ML o IA**.

# Calidad de los datos



## Exactitud

Semántica (valores) |  
Sintáctica (formato)

## Consistencia

Si  $X = 1$ , entonces  $Z = 0$

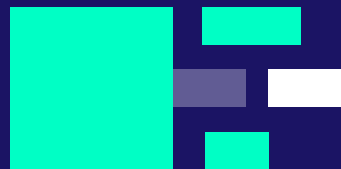


## Completitud

Existencia (o no) de valores  
nulos en los atributos

## +Otros

Ver **ISO25012**







02

ALGUNAS PROPUESTAS

# ¿Qué herramientas podemos usar?

Si bien hay mucho que se puede hacer en los sistemas base (*validaciones y verificaciones antes de guardar los datos*). Existen casos donde se tiene que **hacer un análisis externo** a esos sistemas.

# ¿Qué herramientas podemos usar?

Para esos casos se puede recurrir al mismo (*o casi*) conjunto de **herramientas que usamos para hacer testing** del software.

# Ejemplos

## Python

No requiere que el software de base lo use

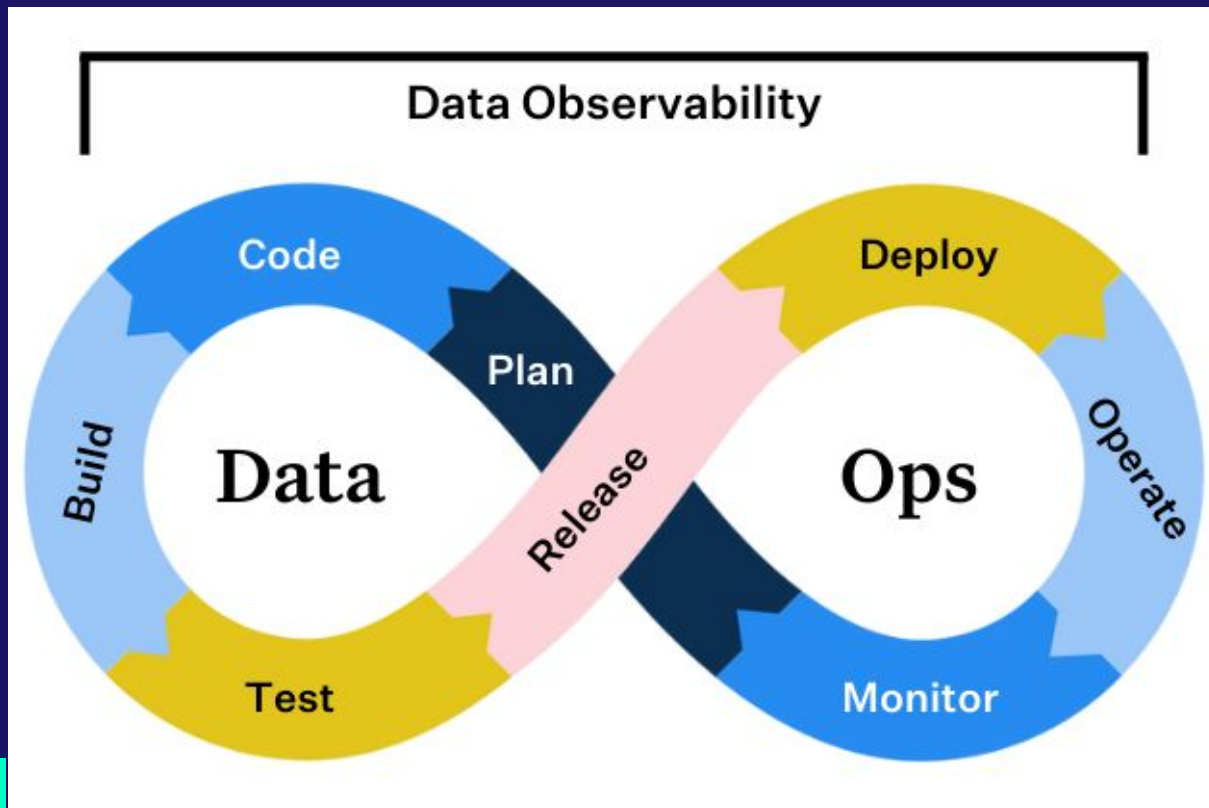
- **PyTest (clásico)**
- **Pandera (librería vinculada a Pandas)**
- **Great Expectations (~similar a BDD)**

# ¿No es algo de ciencia de datos?

**Sí, también.**

En este tipo de proyectos, las prácticas más modernas hablan de verificar la calidad de los datos con los que se entrenan los modelos constantemente.

# DataOps





03

DEMOS

## ¿Qué ejemplos?

Para cada herramienta un caso  
para probar

## Repo

Espacio con todo el contenido  
y ejemplos para ejecutarlos







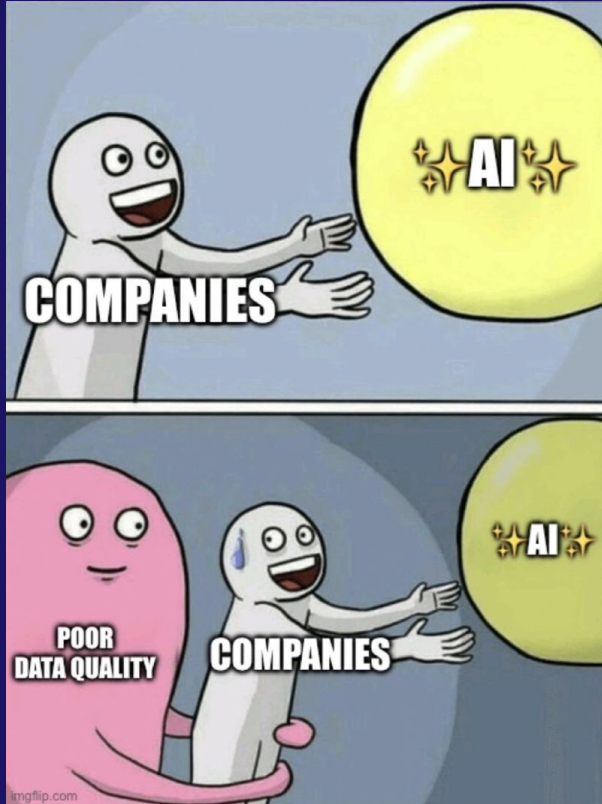
Empecemos



04

CIERRE

# Cierre



Datos de calidad llevan a decisiones basadas en información de calidad.

Además, en el contexto actual de **ponerle IA a todo** (?) los datos con los que esa solución de IA sea entrenada cobran mayor relevancia y la automatización de estos controles nos puede ahorrar tiempo y problemas.



Preguntas?



# Muchas gracias!

CREDITS: This presentation template was created by  
Slidesgo, including icons by Flaticon, and  
infographics & images by Freepik.

**Contacto: [martinrey@fceqyn.unam.edu.ar](mailto:martinrey@fceqyn.unam.edu.ar)**