

ACÀMICA

TEMA DEL DÍA

# Análisis Exploratorio de Datos

¡Cerramos el primer gran tema del programa! Ya conoces todos los ingredientes por separado, es hora de darles una mirada integradora.



# Agenda

---

Daily

Explicación: Paradoja de Simpson

Break

Trello

Hands-on training

Consultas Entrega 01

Cierre



# Daily



Daily



## Sincronizando...

### Toolbox



¿Cómo te ha ido?  
¿Obstáculos?  
¿Cómo seguimos?

### Challenge



¿Cómo te ha ido?  
¿Obstáculos?  
¿Cómo seguimos?

# Análisis Exploratorio de Datos



Un buen Análisis Exploratorio  
de Datos no pierde de vista las  
particularidades del problema  
en el que estamos  
trabajando.<sup>1</sup>

<sup>1</sup> Es decir, depende del problema.

# ABC del Análisis Exploratorio de Datos

1. Fijarse tipos de datos y qué valores toman.
2. Métricas básicas del dataset
3. Graficar la distribución de los datos (histograma de cada columna, etc.).
4. Explorar valores faltantes y outliers<sup>2</sup>
5. Graficar features vs. features. Explorar correlaciones.

<sup>2</sup> Aún no lo vimos, pero pronto.



## Repaso: ¡Un poco de todo!

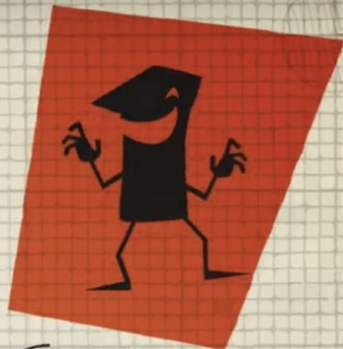


Extra

# ¿Qué es la Paradoja de Simpson?



SIMPSON'S PARADOX



# Paradoja de Simpson

<b>Hospital A</b>	Número de pacientes	Número de pacientes que salvan	Supervivencia
Pacientes No Graves			
Pacientes Graves			
<b>Total</b>	<b>1000</b>	<b>900</b>	<b>90%</b>

# Paradoja de Simpson

<b>Hospital A</b>	Número de pacientes	Número de pacientes que salvan	Supervivencia
Pacientes No Graves			
Pacientes Graves			
<b>Total</b>	<b>1000</b>	<b>900</b>	<b>90%</b>

<b>Hospital B</b>	Número de pacientes	Número de pacientes que salvan	Supervivencia
Pacientes No Graves			
Pacientes Graves			
<b>Total</b>	<b>1000</b>	<b>800</b>	<b>80%</b>

¿Conviene ir al Hospital A  
porque la probabilidad de  
supervivencia es mayor?



¡Necesitamos más  
datos para decidir!

# Paradoja de Simpson

<b>Hospital A</b>	Número de pacientes	Número de pacientes que salvan	Supervivencia
Pacientes No Graves	900	870	96.6%
Pacientes Graves	100	30	30%
<b>Total</b>	<b>1000</b>	<b>900</b>	<b>90%</b>

<b>Hospital B</b>	Número de pacientes	Número de pacientes que salvan	Supervivencia
Pacientes No Graves	600	590	98.3%
Pacientes Graves	400	210	52.5%
<b>Total</b>	<b>1000</b>	<b>800</b>	<b>80%</b>



## Paradoja de Simpson: ¡CUIDADO CON ESTOS VALORES!

<b>Hospital A</b>	Número de pacientes	Número de pacientes que salvan	Supervivencia
Pacientes No Graves	900	870	96.6%
Pacientes Graves	100	30	30%
<b>Total</b>	<b>1000</b>	<b>900</b>	<b>90%</b> 

<b>Hospital B</b>	Número de pacientes	Número de pacientes que salvan	Supervivencia
Pacientes No Graves	600	590	98.3%
Pacientes Graves	400	210	52.5%
<b>Total</b>	<b>1000</b>	<b>800</b>	<b>80%</b> 

## Paradoja de Simpson: **MÁS INFORMACIÓN AQUÍ**

<b>Hospital A</b>	Número de pacientes	Número de pacientes que salvan	Supervivencia
Pacientes No Graves	900	870	<b>96.6%</b>
Pacientes Graves	100	30	<b>30%</b>
<b>Total</b>	<b>1000</b>	<b>900</b>	<b>90%</b>

<b>Hospital B</b>	Número de pacientes	Número de pacientes que salvan	Supervivencia
Pacientes No Graves	600	590	<b>98.3%</b>
Pacientes Graves	400	210	<b>52.5%</b>
<b>Total</b>	<b>1000</b>	<b>800</b>	<b>80%</b>

A close-up photograph of a white ceramic cup filled with a latte. The surface of the milk is decorated with intricate latte art, featuring a central heart shape surrounded by concentric, wavy lines. The cup is placed on a matching white saucer. In the background, a white napkin and a silver spoon are visible, though they are out of focus. The overall lighting is soft and even.

**¡BREAK!**

---

HERRAMIENTA DE GESTIÓN DIGITAL

# Trello



## PENDIENTES

- ☐ PENDIENTES BANCOLOMBIA
- ☐ SIGUIENTES PASOS EVENTO EMPLEAB
- ☐ PENDIENTES ~~TRUCES~~ TUCRES, SOCIALIZAR
- ☐ ARAZCLAR KEYDOCS, DOCS, PPT
- ☐ REPORTE B2B, darle la vuelta al requerimiento
- ☒ OBJETIVOS CITY COORDINATOR
- ☒ ONBOARDING CITY COORDINATOR
- ☒ LEER PEDAGOGIA (JULIAN)
- ☒ LEER COU NO. FORMAL (BSU)
- ☐ BSU ORGANIZAR CONTRATOS
- ☐ COORDINAR VENIDA BRIAN
- ☐ BUDGET 2020
- ☐ EJECUTAR FLYERS
- ☒ EJECUTAR P&P
- ☐ CONFIRMAR ESPACIOS 1Q - 2Q
- ☐ INV. + Grupos inicio
- ☐ EDTECH, siguientes pasos
- ☐ BIT + ACÁMICA, next steps → AGENDA BORRADOR

## EXPE

EXPER

Cana

PROGRA

- P

- A

- N

TRABAJAR

EL 3er B

INTERSHIP

MIENTRAS

LA CLASAS

NO FRENA

~~sig~~ sig

↳

↳

↳

↳

COMO

INFORMA

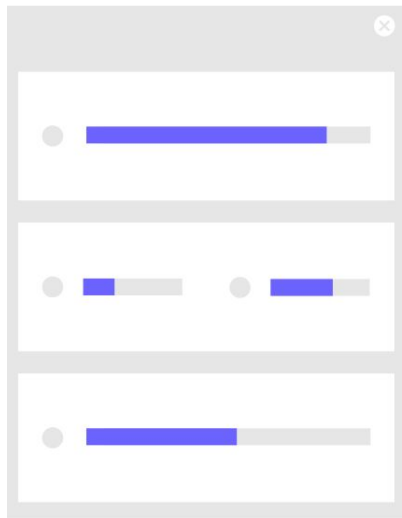
¿Les parece familiar?



Una herramienta de gestión digital que nos ayuda a **ordenar, segmentar y visualizar** nuestros flujos de trabajo de una manera muy práctica y fácil.

# ¿Para qué me sirve?

seguimiento  
personal del  
proyecto



respaldar  
nuestro  
trabajo

mostrar status  
a colegas

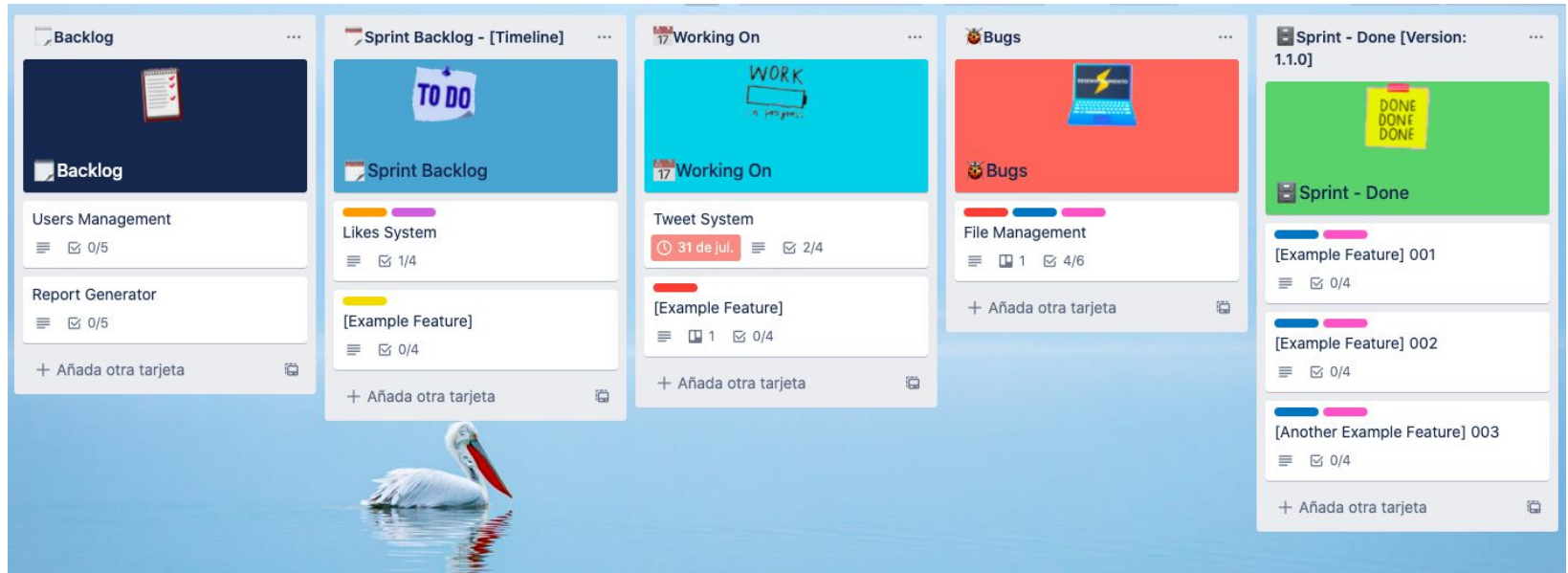


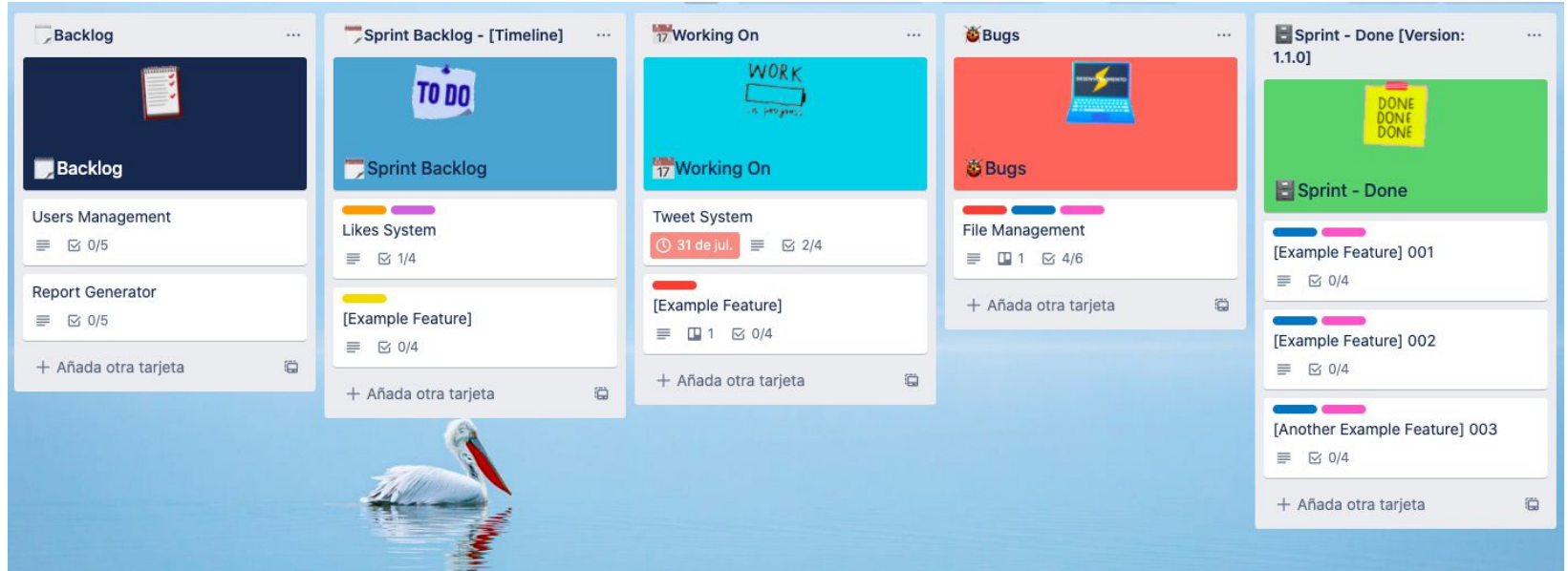
# ¿Cómo funciona?

Los **tableros**, las **listas** y las **tarjetas** te permiten organizar y priorizar tus proyectos.



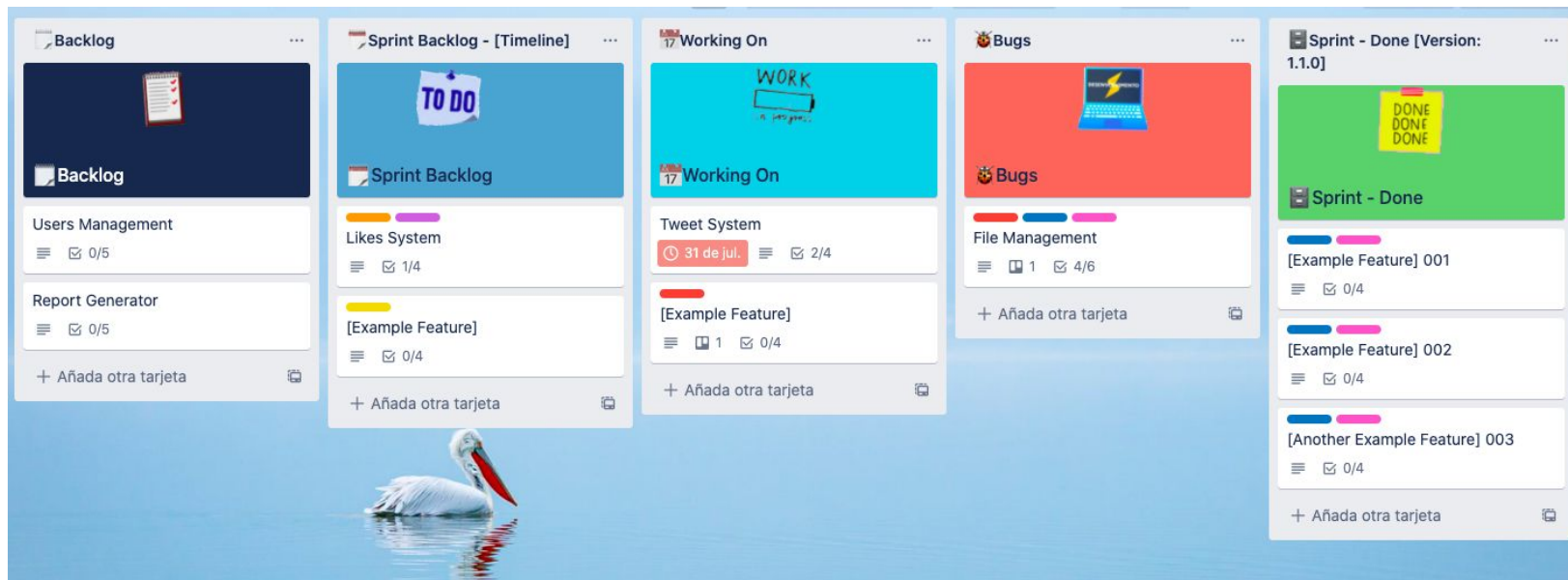






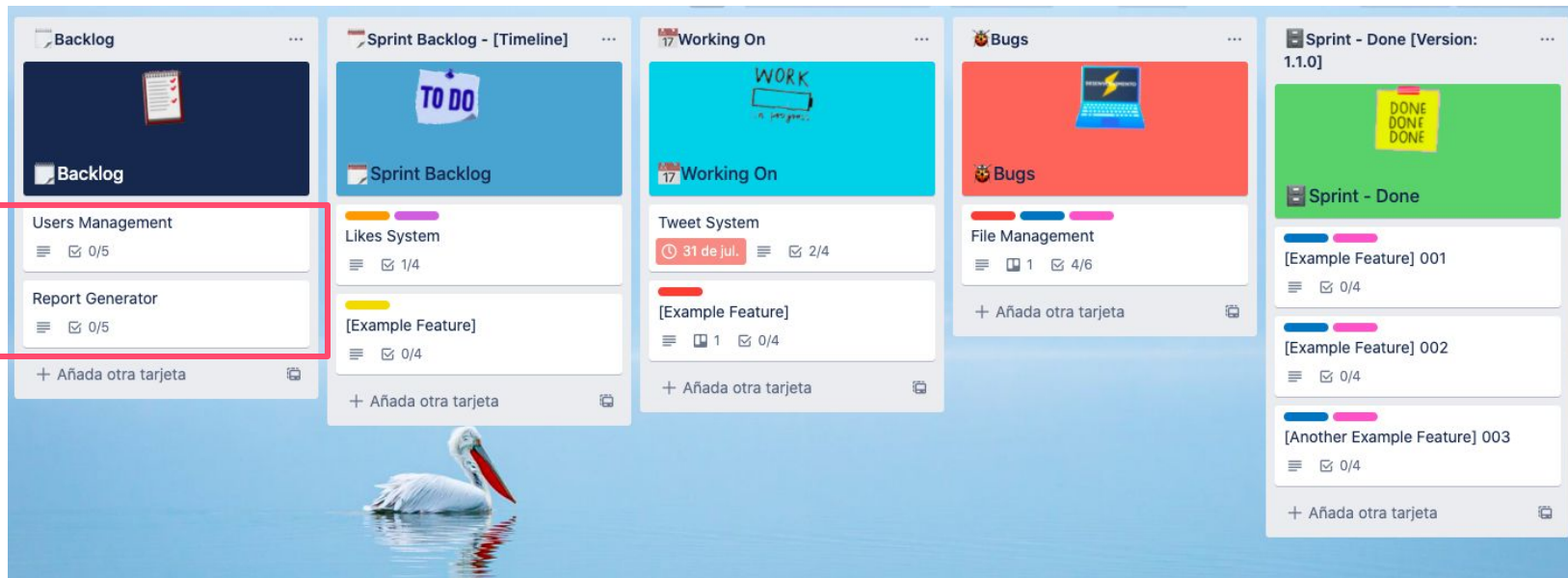
Tablero

Listas



Tablero

Listas

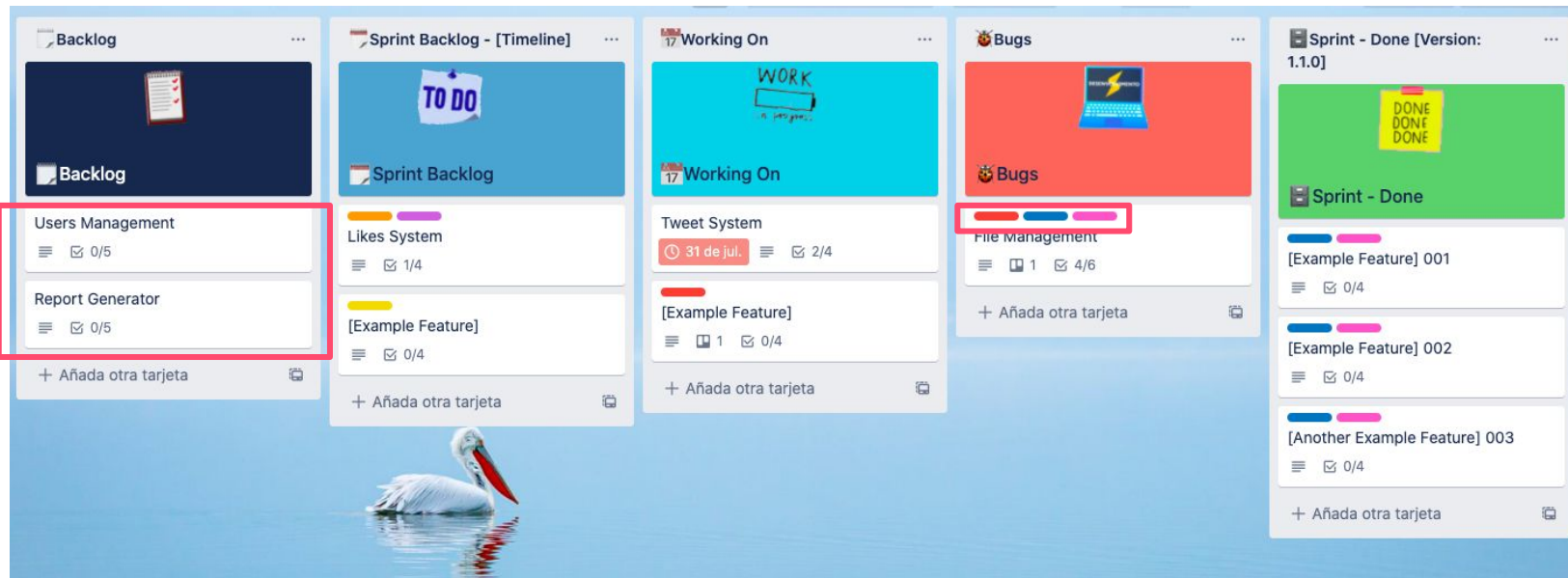


Tarjetas

Tablero

Listas

Etiquetas



Tablero

## ¿Por qué usamos algo así?

1. Tenemos todas las tareas centralizadas
2. Mejora la organización y comunicación a nivel equipo
3. Queda registro de lo que decidimos y de lo que hacemos
4. Tiene muchas sincronizaciones
5. Da visibilidad para las demás áreas
6. Permite pensar a corto, mediano y largo plazo



# El tablero del Proyecto 1





## Vamos a armar salas reducidas



20 min.

En equipos reducidos de 3-4 personas, creen su backlog para el primer Proyecto.

**Tip:** usen el PRD como referencia.

### Recuerden, a medida que actualizan su tablero:

- Marcar el estado de cada tarea.
- Estipular los tiempos de ejecución de cada tarea y marcar las *deadlines* en el tablero.
- Identificar posibles futuros *pain points*.



# Proyecto 1



Encontrá tu próxima  
casa en Properati



**PROPERATI**  
EL MEJOR CAMINO A TU PRÓXIMA CASA

 Buscar

¡Bienvenido/a a tu  
primer flujo de  
trabajo de  
Data Science!

## **Apóyate en:**

- las toolboxes
- los notebooks
- las PPTs de cada meeting.

## **Y además en:**

- la documentación de las librerías
- la comunidad (Stack Overflow y Google).

# Flujo de trabajo

---

PARTE 1 Pensando como un/a Data Scientist.

PARTE 2 Análisis Exploratorio de Datos.

PARTE 3 Primer Modelo de Machine Learning.

¡Hoy empecemos con  
la Parte 1 y 2!

# Recursos



# Análisis Exploratorio de Datos (EDA)

- [Exploratory Data Analysis: Baby Steps](#)
- [Exploratory data analysis in Python](#)

# Para la próxima

---

- Continúa trabajando en el proyecto y/o resolviendo notebooks anteriores.
- Lee la Toolbox 09.
- Resuelve el Challenge.



ACÀMICA