

ACÀMICA

TEMA DEL DÍA

Análisis Exploratorio de Datos

¡Cerramos el primer tema de nuestra carrera! Ya viste todos los ingredientes por separado, es hora de darles una mirada integradora.



Agenda

Daily

Explicación: Paradoja de Simpson

Break

Hands-on training

Consultas Entrega 01

Cierre



Daily



Daily



Sincronizando...

Bitácora



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

Challenge



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

Repaso de la bitácora



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

¹ 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 outliers²
5. Graficar features vs. features. Explorar correlaciones.

² Aún no lo vimos, pero pronto.

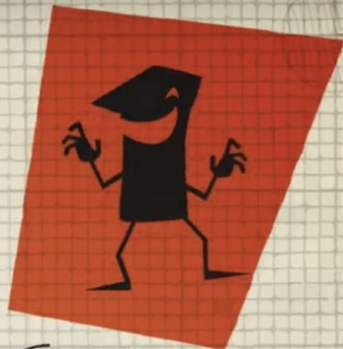
Repaso: ¡Un poco de todo!



Paradoja de Simpson



SIMPSON'S PARADOX



Paradoja de Simpson

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

Paradoja de Simpson

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

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

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



¡Necesitamos más
datos para decidir!

Paradoja de Simpson

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

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

Paradoja de Simpson: ¡CUIDADO CON ESTOS VALORES!

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

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

Paradoja de Simpson: **MÁS INFORMACIÓN AQUÍ**

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

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

Proyecto 1



Encontrá tu próxima
casa en Properati



PROPERATI
EL MEJOR CAMINO A TU PRÓXIMA CASA

Q Buscar

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

Para realizar el proyecto debes usar como referencia el contenido de las bitácoras, de los notebooks trabajados y las presentaciones vistas en clase.

También te será de mucha utilidad consultar la documentación de las librerías y de 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 empezaremos a trabajar
en 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 bitácora 09 y carga las dudas que tengas al Trello
- Resuelve el Challenge.

En el encuentro que viene uno/a de ustedes será seleccionado/a para mostrar cómo resolvió el challenge de la bitácora. De esta manera, ¡aprendemos todos/as de (y con) todas/as, así que vengan preparados/as.

ACÀMICA