

DATA SCIENCE



Federico Baiocco
baioccofed@gmail.com
3512075440



Clase 1 - Agenda

¡ BIENVENIDOS AL CURSO DE DATA SCIENCE !

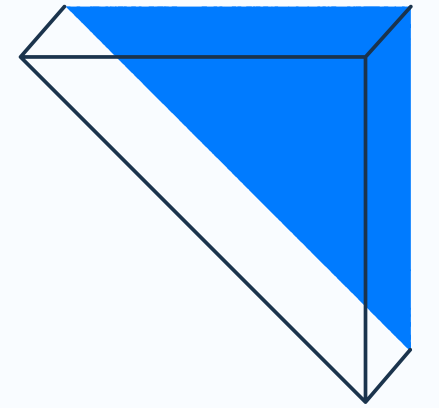
- Nos presentamos
- Herramientas que vamos a utilizar
- Ciencia de datos
- Workflow data science
- El curso - Sprints
- Jupyter notebooks

Nos empezamos a conocer

ICARO 2021



Herramientas



01

Zoom



02

Anaconda



03

Python



04

Slack



05

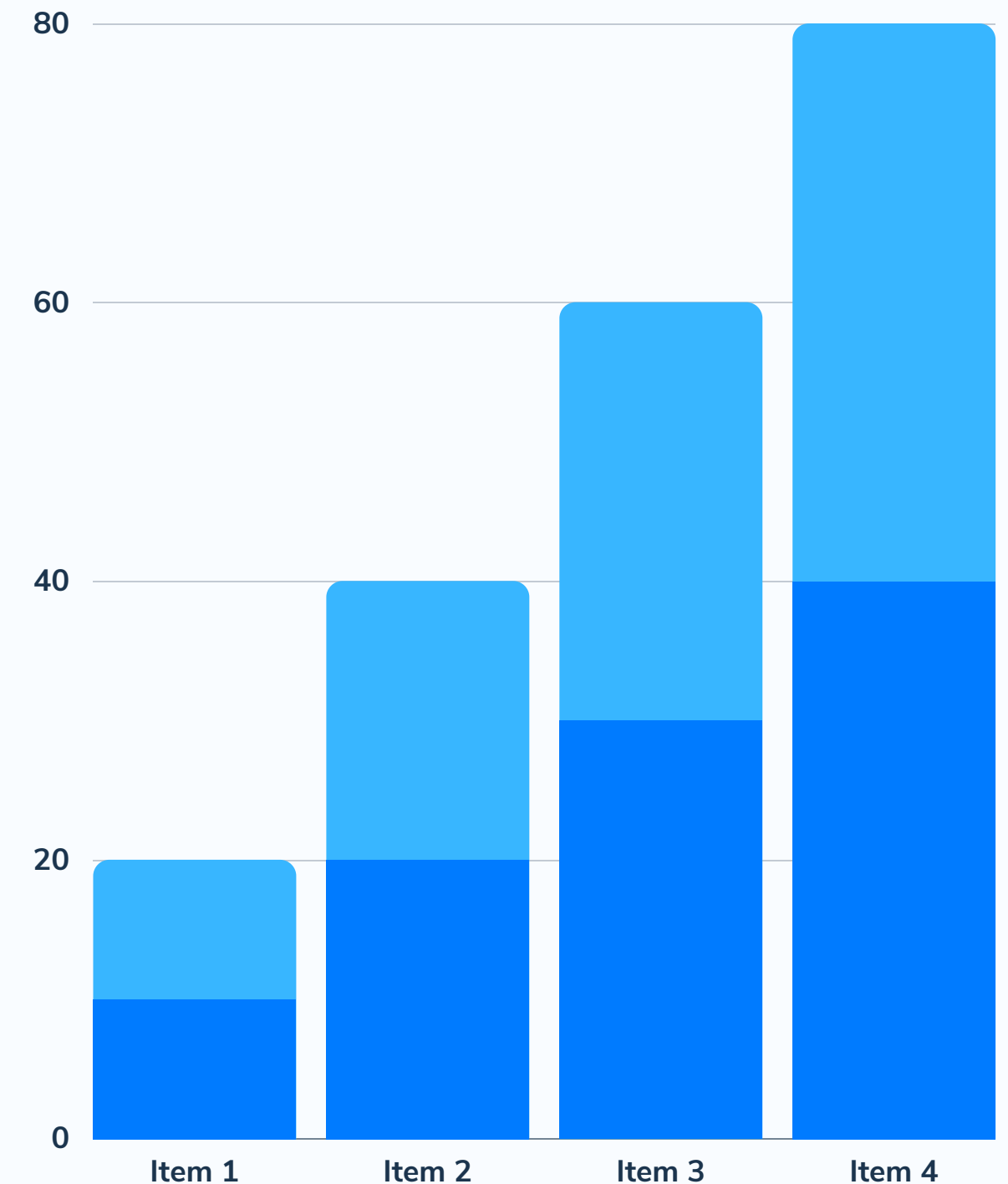
GOOGLE



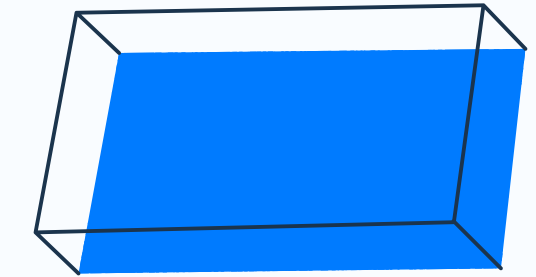
La ciencia de datos

"DATA IS THE OIL OF THE 21ST CENTURY"

- La ciencia de datos combina múltiples campos que incluyen estadísticas, métodos científicos y análisis de datos para extraer el valor de los datos.
- Algunas disciplinas relacionadas:
 - Data engineer
 - Data analyst



El data scientist



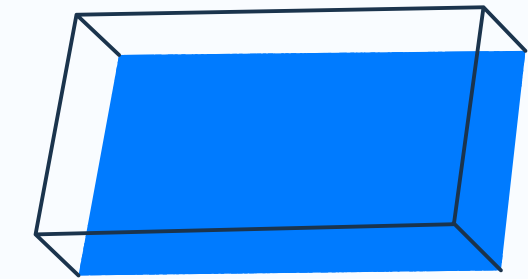
ESTADÍSTICA Y MATEMÁTICAS

PROGRAMACIÓN

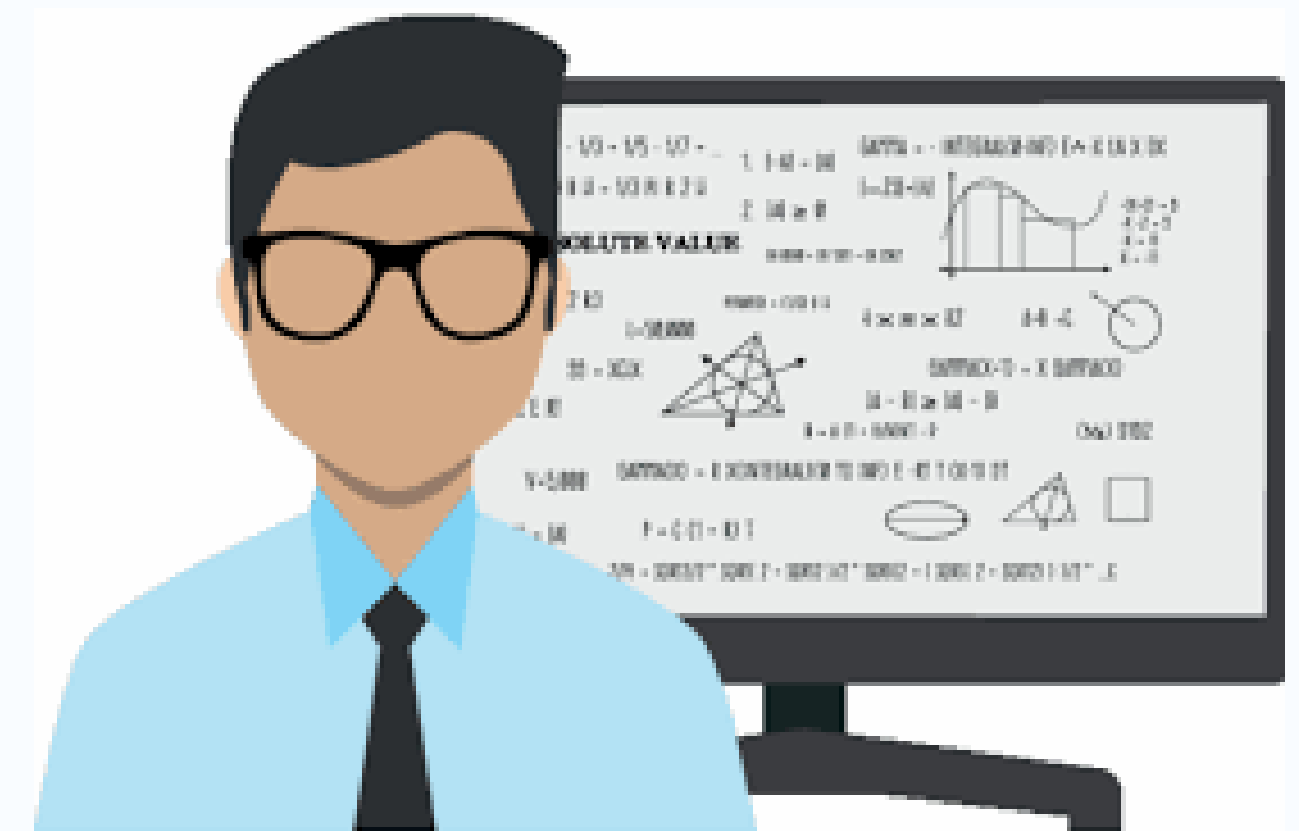
CONOCIMIENTO DEL PROBLEMA

**COMUNICACIÓN DE LOS
RESULTADOS**

El data scientist

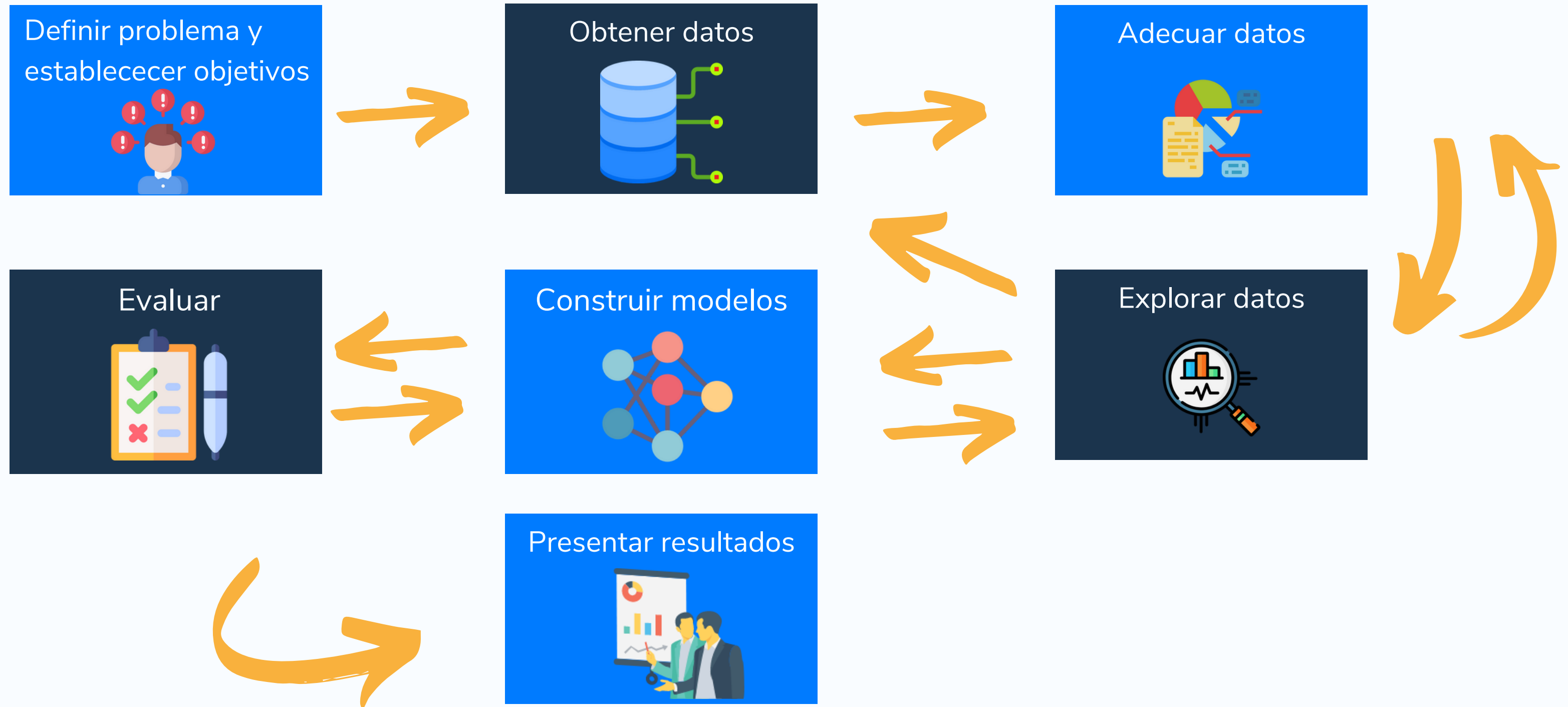
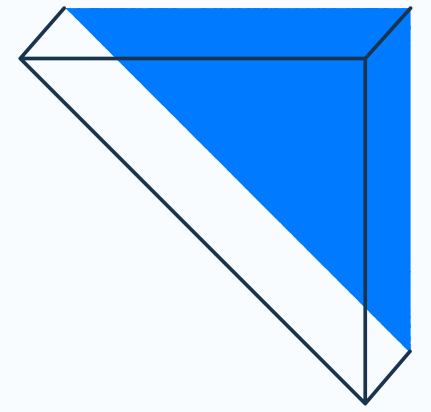


Es una persona que puede agarrar un conjunto de datos, desarrollar un caso de uso para los mismos, crear una hipótesis sobre cómo utilizarlos, desarrollar experimentos utilizándolos, analizar los resultados y llegar a una solución.

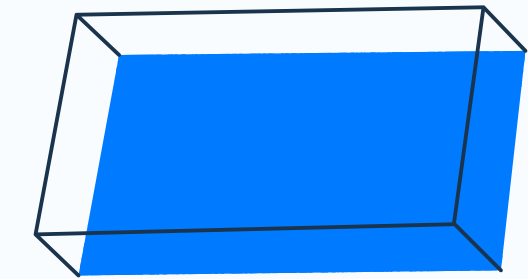


Data science workflow

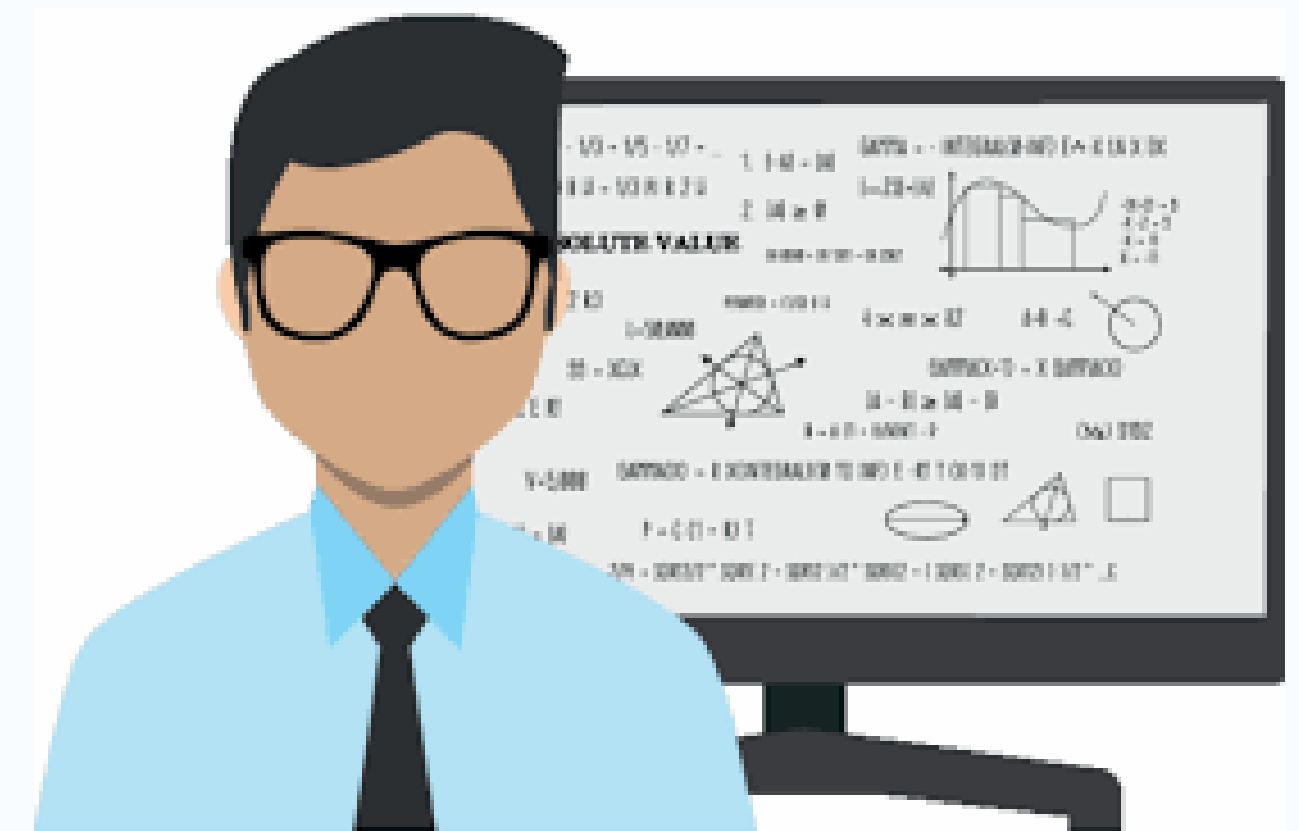
UN PROCESO ITERATIVO



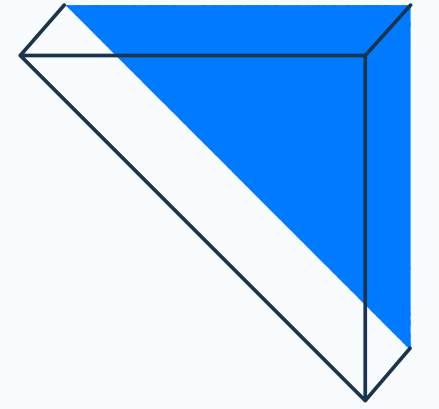
El data scientist



- Las computadoras son las mejores herramientas que tenemos para coleccionar y procesar datos
- Una de las etapas que más tiempo consume es la de obtener y limpiar los datos

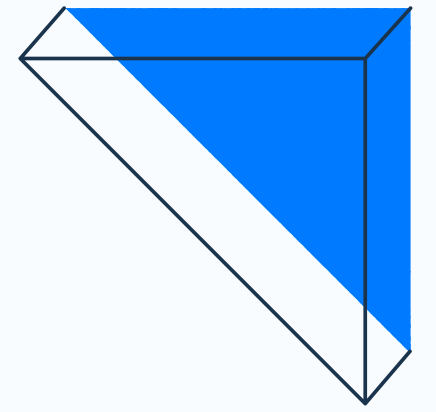


¿ Dónde se aplica ?

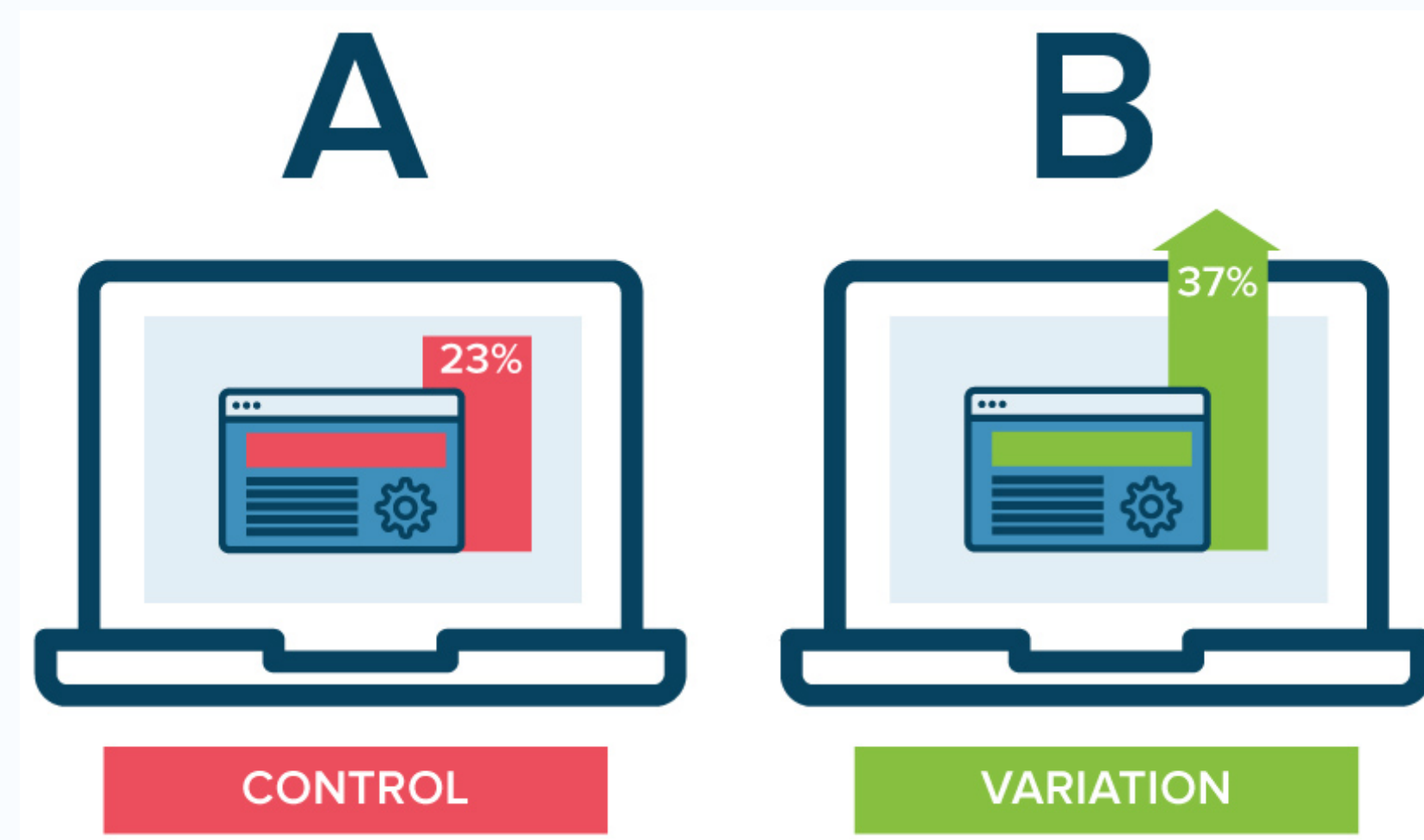


- Sistemas de recomendación
- Detección de fraude
- Computer vision
- Segmentación de clientes
- Customer churn prevention
- Procesamiento del lenguaje natural

A/B tests

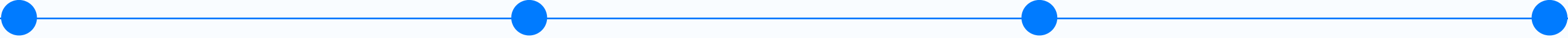


- Consiste en comparar 2 (o más) versiones de una web o aplicación para medir cuál es la más eficiente
- Se muestra de forma aleatoria distintas versiones a distintos usuarios
- Una vez que se recolectaron datos, se miden resultados y se elige un ganador



Planificación

4 sprints



Machine Learning desde cero: mi primer modelo

Dado una problemática vas a tener que armar el caso de uso, definir el/los objetivos e identificar si cuentas con los datos requeridos para esbozar la solución, también vas a tener realizar un análisis exploratorio del set de datos y utilizar un modelo simple de Machine Learning como solución propuesta.

Pasando a un modelo más complejo

Para este sprint vamos a profundizar sobre la estadística y matemática que está detrás, dando un salto hacia modelos más complejos de Machine Learning. El objetivo es poder hacer foco en la interpretación de resultados y la correcta evaluación y comparación de los mismos.

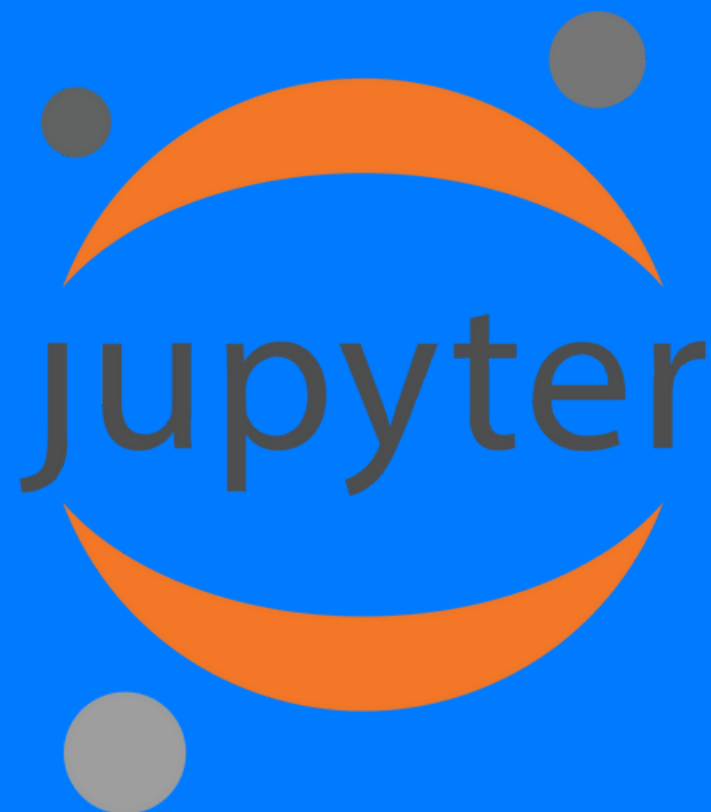
Advanced Machine Learning: potenciando mi solución

El desafío consiste en aplicar las técnicas de ingeniería de features enriqueciendo la información disponible, así como utilizar y desarrollar algoritmos avanzados de ML y tener la capacidad de interpretar los resultados. ¿cuáles serían tus conclusiones, se puede mejorar aún más los resultados?

Desarrollando aplicaciones modernas

En este proyecto se propone que seas capaz de adentrarte y poder resolver problemas muy actuales y en donde el machine learning ha marcado un antes y un después con sus novedosas soluciones.

Empecemos a conocer algunas herramientas...



colab



Preguntas?

Para la proxima clase:

- OPCIONAL: Instalar Anaconda

<https://docs.anaconda.com/anaconda/install/>