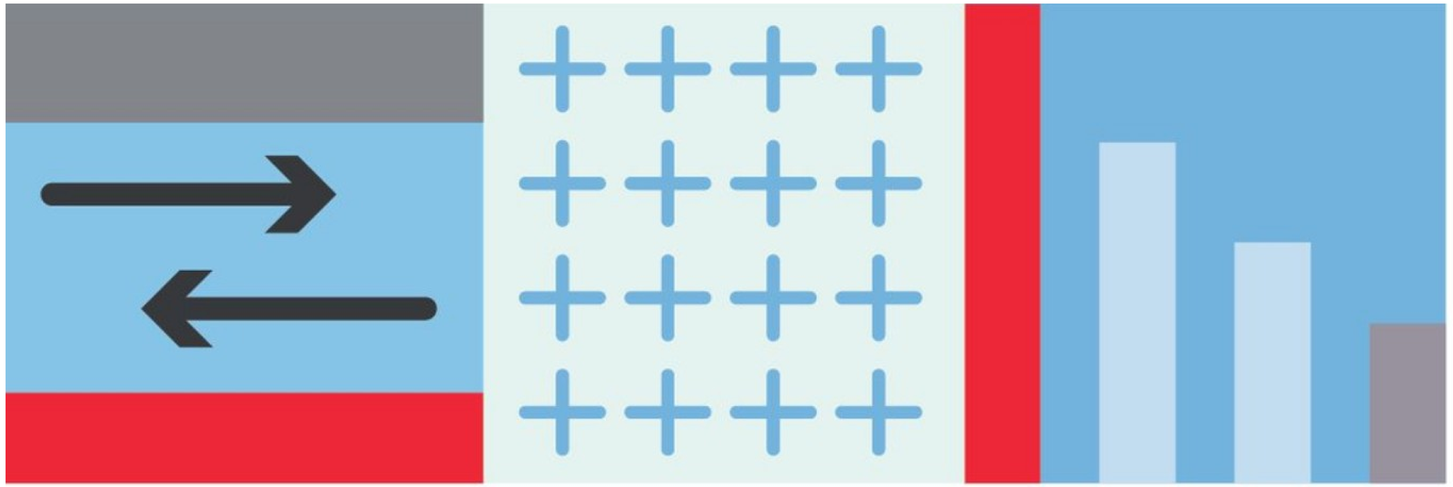
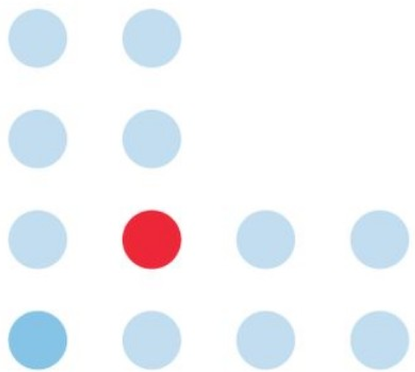


Certificación
Universitaria

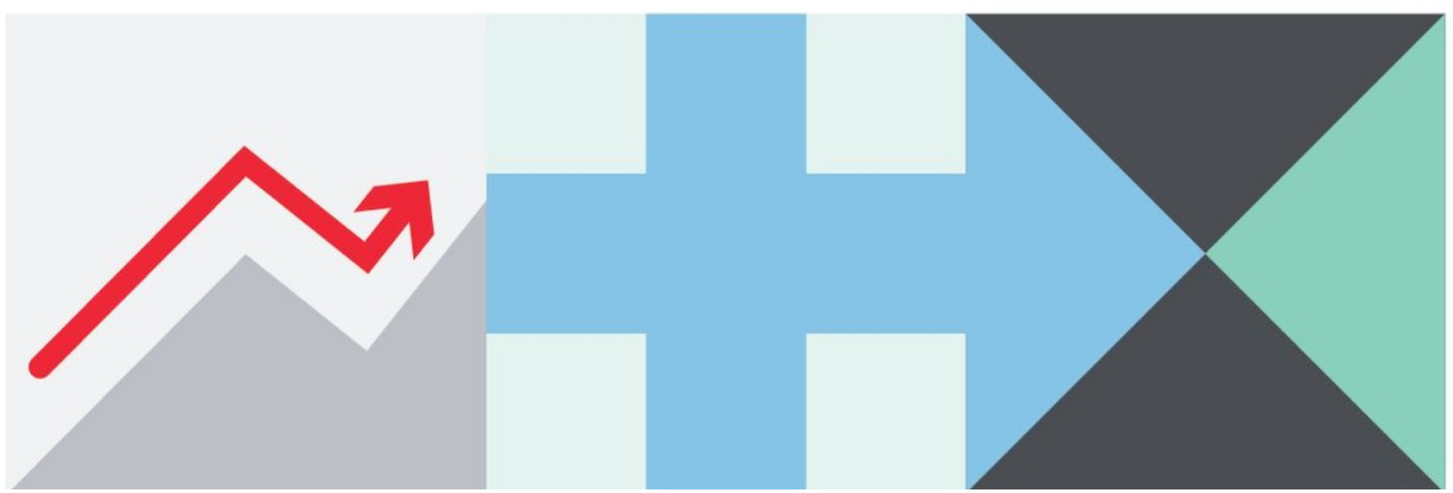
DATA SCIENCE



E

Introducción a la Ciencia de Datos

Sol Represa



CC

UNC FCEFyN

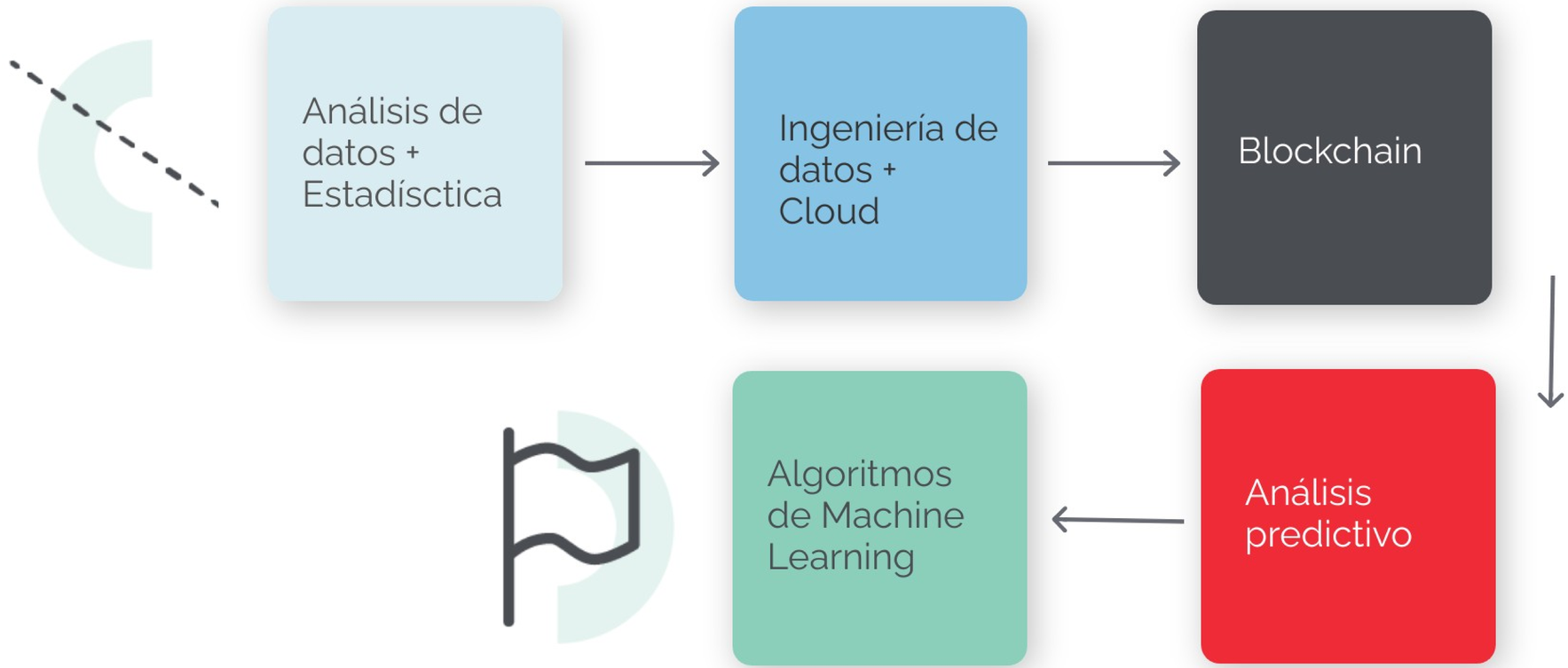
mundosE
PEOPLE & BUSINESS SCHOOL

PROFESORA
Sol Represa

Actualmente

- Líder en Ciencias de Datos - NTT Data
- Docente universitaria UNLP
- Consultora en Ciencia de Datos

Nuestro recorrido



¿Qué es un hackatón?

- Los hackatones son **eventos lúdicos**
- Nos juntamos para **trabajar de forma colaborativa** en desafíos.

Son **maratones** con base digital para **desarrollar soluciones**, rápidas y efectivas, y sirven como **disparadores de ideas** colectivas y creativas.

Hackatón

- En equipos de **máximo 4 personas** se trabajará sobre **problemáticas reales**.
- El objetivo es que cada estudiante pueda utilizar su proyecto final como parte de su **portfolio personal**.
- La presentación será mediante un **pitch de 4 minutos**.
- Debe ser completamente **documentado y de libre acceso** en un repositorio de Github.

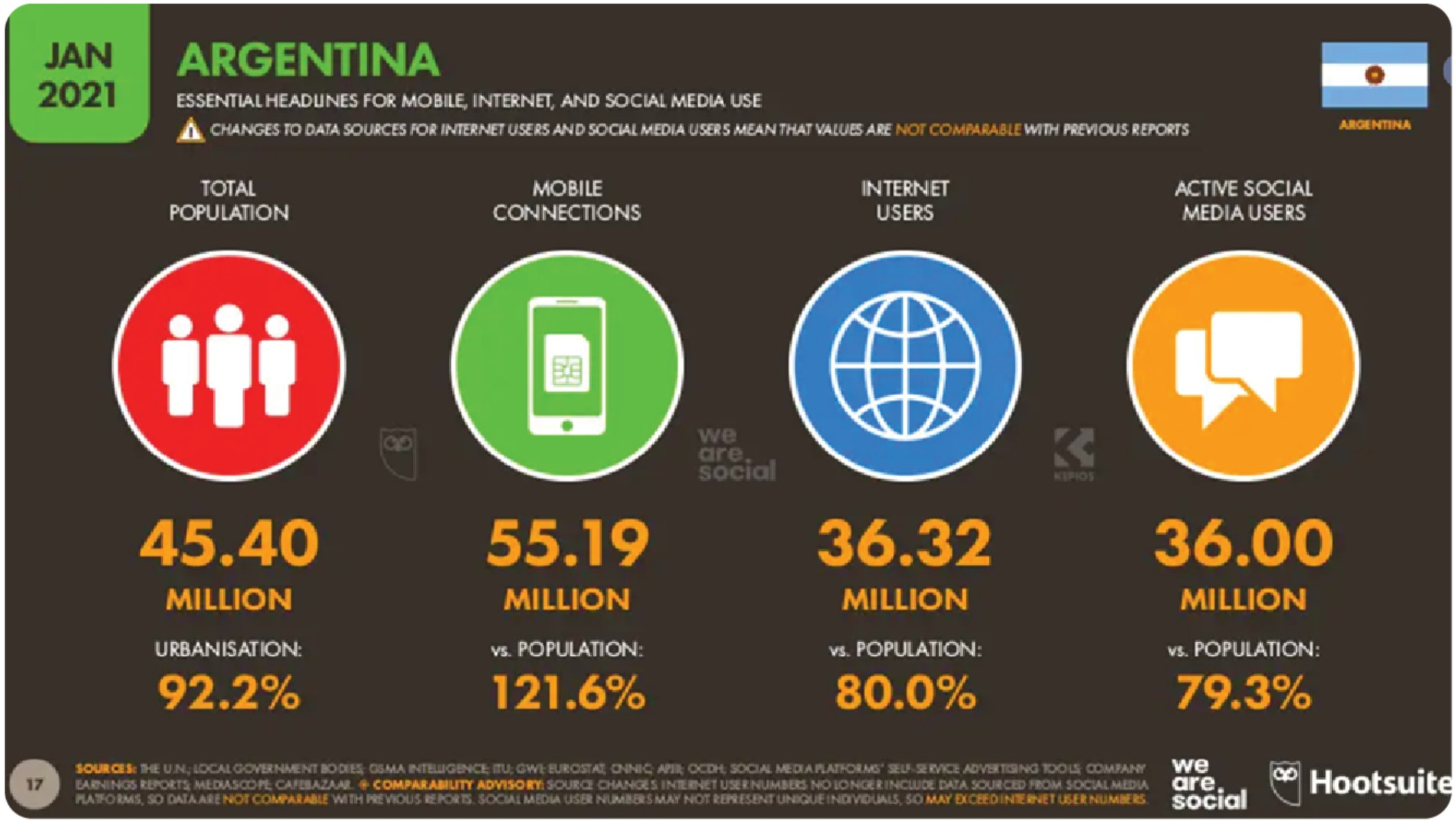
Introducción al Data Science

mE

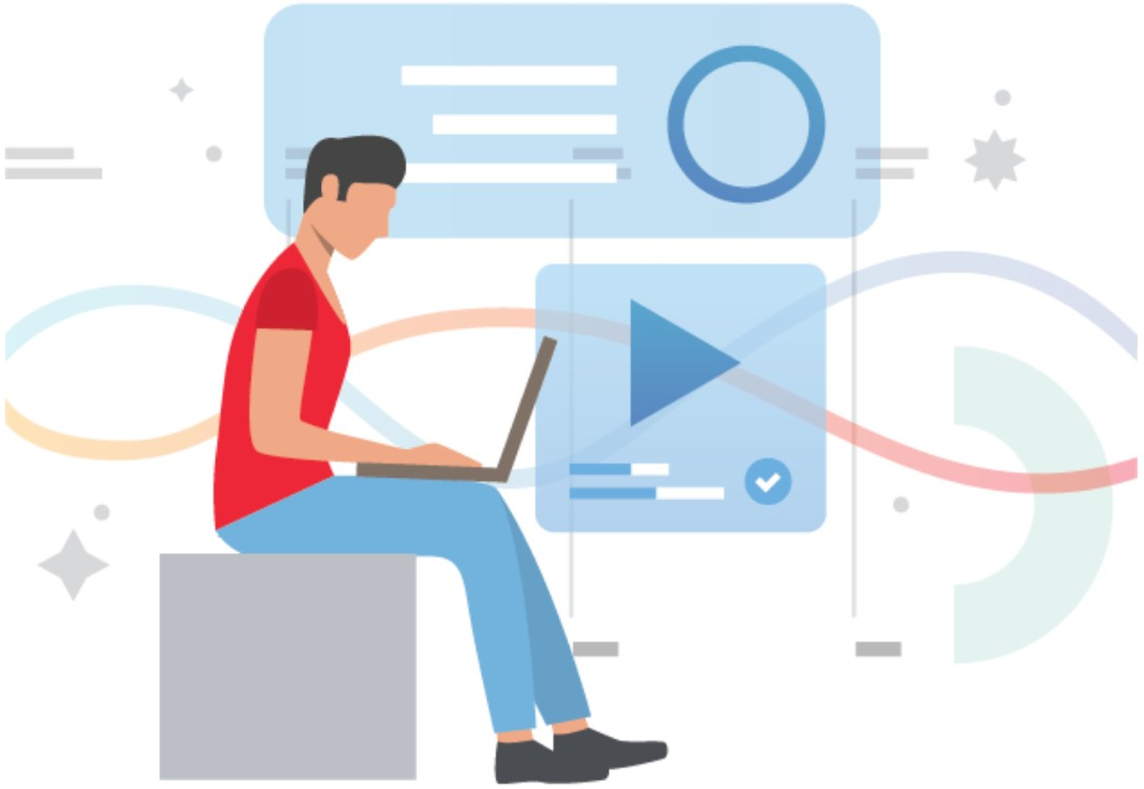
Temas a desarrollar

La sociedad de la información y la 4ta Revolución industrial → Gobernanza sobre los datos. Ciclo de vida de los datos. → Diseño de un proyecto de Data Science. Roles.

¿Cuál es el % de argentinxs que usan redes sociales?


JAN 2021
ARGENTINA
ESSENTIAL HEADLINES FOR MOBILE, INTERNET, AND SOCIAL MEDIA USE
CHANGES TO DATA SOURCES FOR INTERNET USERS AND SOCIAL MEDIA USERS MEAN THAT VALUES ARE NOT COMPARABLE WITH PREVIOUS REPORTS
ARGENTINA
TOTAL POPULATION
45.40 MILLION
URBANISATION: 92.2%
MOBILE CONNECTIONS
55.19 MILLION
vs. POPULATION: 121.6%
INTERNET USERS
36.32 MILLION
vs. POPULATION: 80.0%
ACTIVE SOCIAL MEDIA USERS
36.00 MILLION
vs. POPULATION: 79.3%
17
SOURCES: THE U.N.; LOCAL GOVERNMENT BODIES; GSMA INTELLIGENCE; ITU; GWI; EUROSTAT; CNNIC; APJII; OCDH; SOCIAL MEDIA PLATFORMS' SELF-SERVICE ADVERTISING TOOLS; COMPANY EARNINGS REPORTS; MEDIASCOPE; CAFEBAZAAR. COMPARABILITY ADVISORY: SOURCE CHANGES. INTERNET USER NUMBERS NO LONGER INCLUDE DATA SOURCED FROM SOCIAL MEDIA PLATFORMS, SO DATA ARE NOT COMPARABLE WITH PREVIOUS REPORTS. SOCIAL MEDIA USER NUMBERS MAY NOT REPRESENT UNIQUE INDIVIDUALS, SO MAY EXCEED INTERNET USER NUMBERS.
we are social
Hootsuite

¿Cuántas horas al día usamos internet?



JAN 2021

OVERVIEW OF INTERNET USE

KEY INDICATORS OF INTERNET ADOPTION AND USE

⚠ USER NUMBERS NO LONGER INCLUDE DATA SOURCED FROM SOCIAL MEDIA PLATFORMS, SO VALUES ARE NOT COMPARABLE WITH PREVIOUS REPORTS

ARGENTINA

TOTAL NUMBER OF INTERNET USERS (ANY DEVICE)	INTERNET USERS AS A PERCENTAGE OF TOTAL POPULATION	ANNUAL CHANGE IN THE NUMBER OF INTERNET USERS	AVERAGE DAILY TIME USERS AGED 16 TO 64 SPEND USING THE INTERNET	PERCENTAGE OF USERS ACCESSING THE INTERNET VIA MOBILE DEVICES
36.32 MILLION	80.0%	+3.5% +1.2 MILLION	9H 39M	94.5%

24

SOURCES: KEPIOS (JAN 2021) BASED ON EXTRAPOLATIONS OF DATA PUBLISHED BY: THE ITU; LOCAL GOVERNMENT BODIES; GWI; GSMA INTELLIGENCE; EUROSTAT; APJII; CNNIC; THE UNITED NATIONS. DATA FOR TIME SPENT AND MOBILE INTERNET SHARE FROM GWI (Q3 2020). SEE GLOBALWEBINDEX.COM FOR MORE DETAILS. ✦ COMPARABILITY ADVISORY: SOURCE AND BASE CHANGES. INTERNET USER NUMBERS NO LONGER INCLUDE DATA SOURCED FROM SOCIAL MEDIA PLATFORMS. FIGURES ARE NOT COMPARABLE WITH DATA PUBLISHED IN PREVIOUS REPORTS.

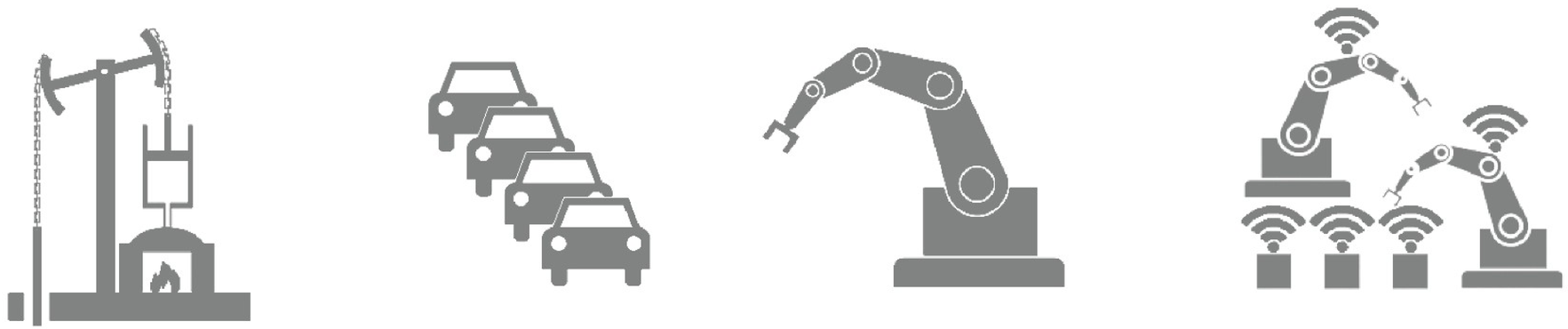
we are social Hootsuite

Sociedad de la información

/ Estamos atravesando una **revolución digital** basada en:

- la **popularización de internet**
- la **masificación del uso de celulares** con alta capacidad de procesamiento
- la aparición de las **redes sociales**

4ta Revolución Industrial



1era

Mecanización, máquinas a vapor

2da

Producción en masa, electricidad

3era

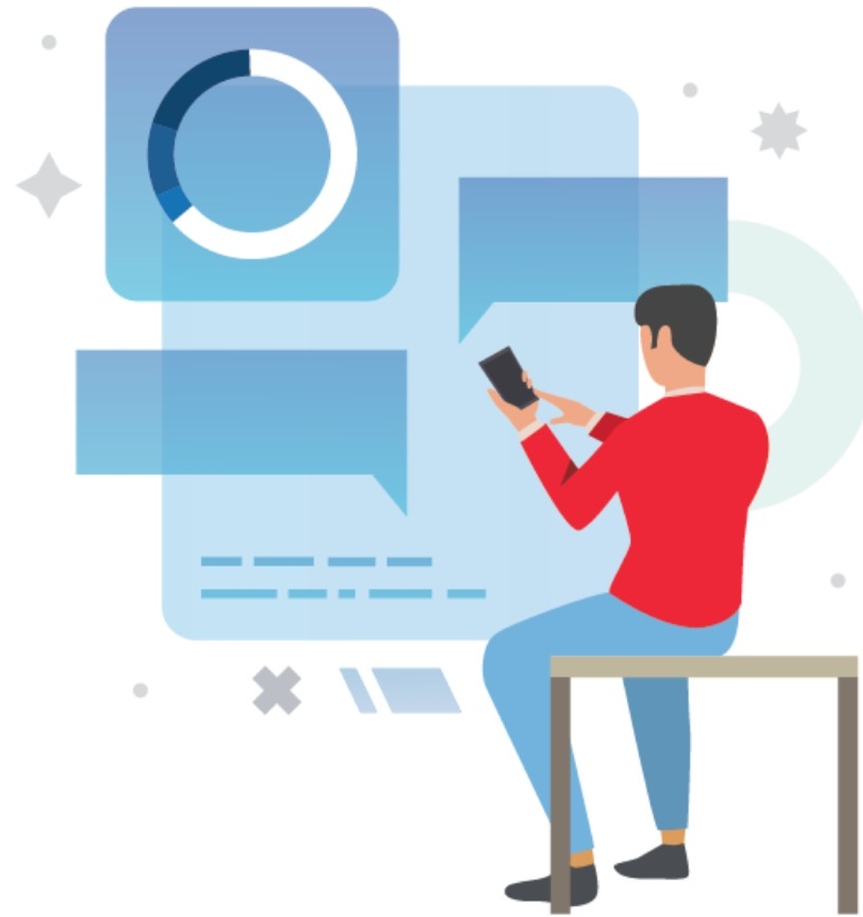
Computación y automatización

4ta

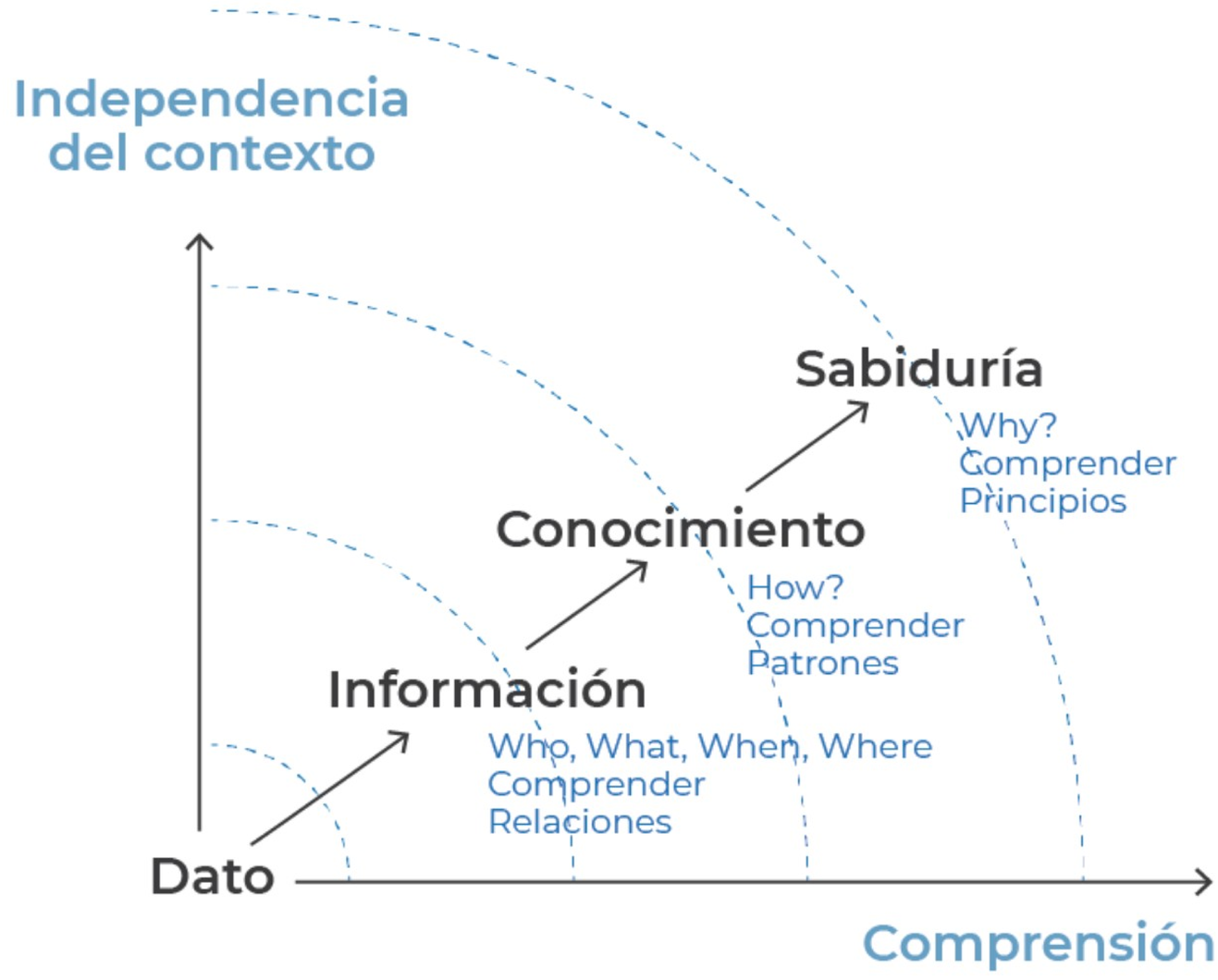
Sistemas ciberfísicos IOT + Big Data

DATA
LOTS AND LOTS OF DATA
makeameme.org

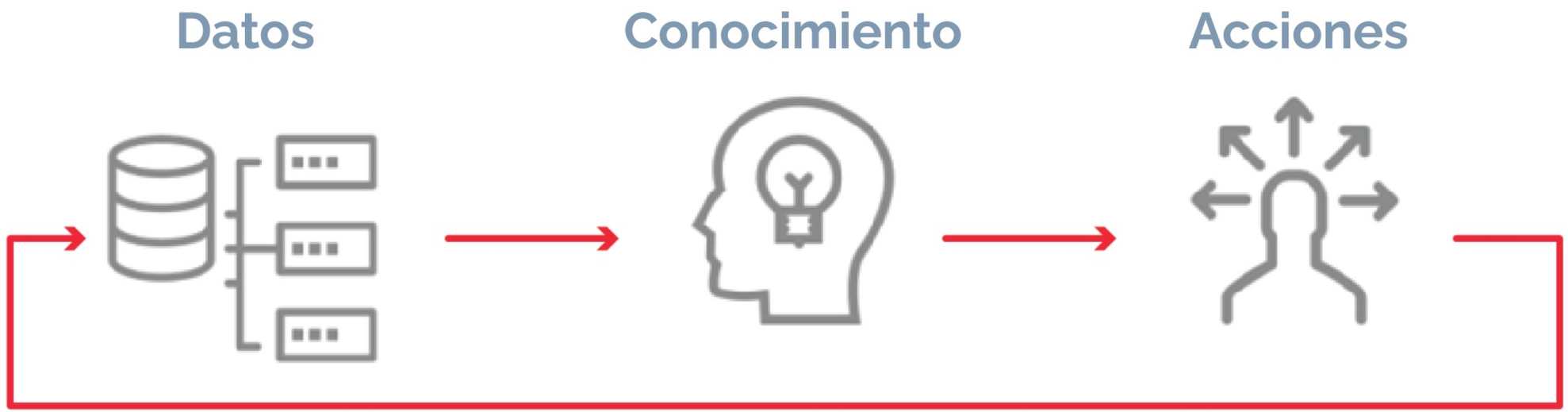
¿Contar con datos es lo mismo que tener información?



¿Información es lo mismo que conocimiento?


Independencia del contexto
Sabiduría
Why?
Comprender
Principios
Conocimiento
How?
Comprender
Patrones
Información
Who, What, When, Where
Comprender
Relaciones
Dato
Comprensión

Modelo Data Driven



¿Qué es la **Ciencia de datos**?

¿Qué es la **Ciencia de Datos**?

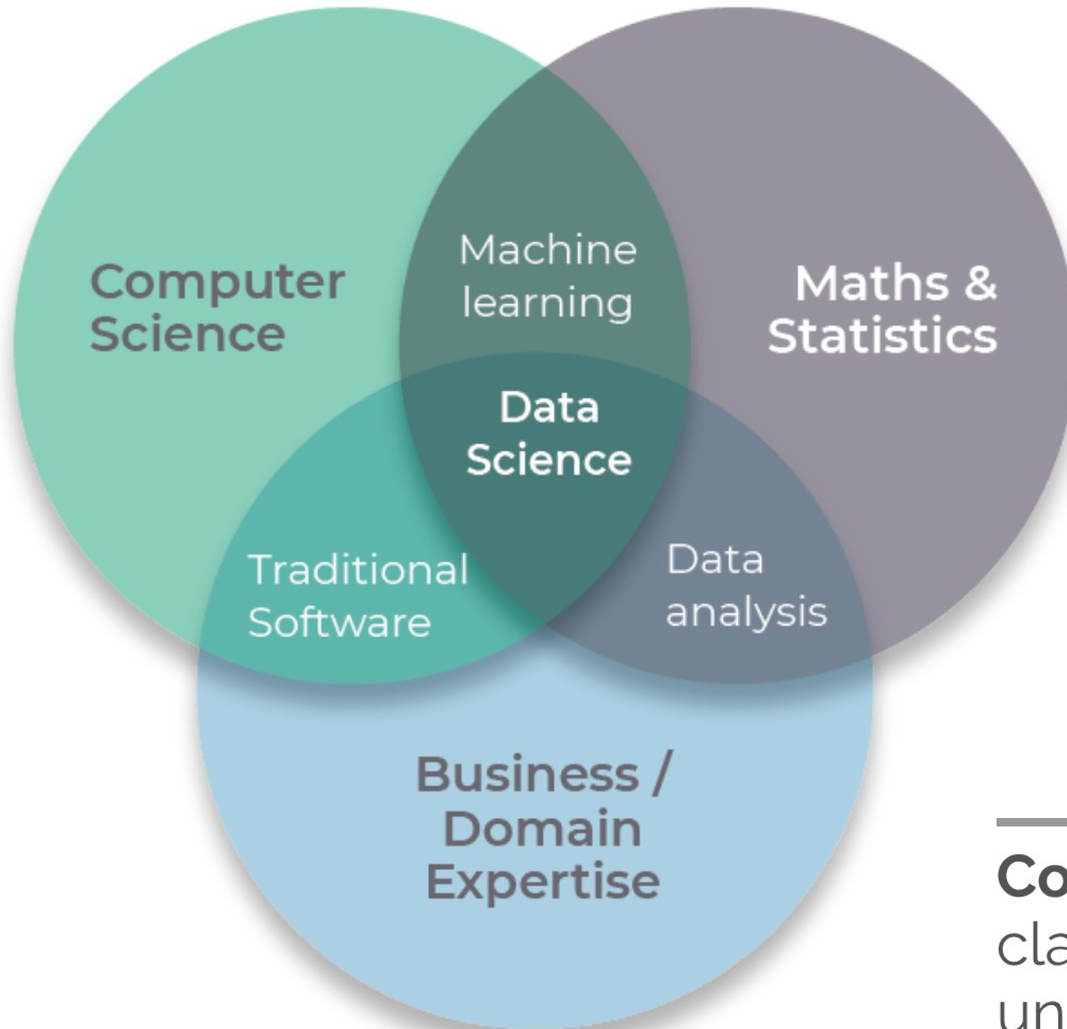
Es un **campo interdisciplinario** que utiliza métodos, procesos, algoritmos y sistemas científicos para **extraer información y generar conocimiento** de datos estructurados y no estructurados.

*¿Qué es la **ciencia de datos**?*

"Es la disciplina que vuelve **útil** los **datos**."

Cassie Kozyrkov

¿Qué es la ***ciencia de datos****?*



Conocer el negocio es clave para ser un **buen Data Scientist**.

Ciclo de vida de los datos

UN GRAN PODER CONLLEVA
UNA GRAN RESPONSABILIDAD

Gobernanza y protección de datos

Conjunto de procesos, funciones, políticas, normas y mediciones que garantizan el uso eficaz y eficiente de la información

- Asegura la **calidad** y **seguridad** de la información
- Permite definir **quién** puede emprender acciones, **sobre qué datos**, **en qué circunstancias** y **mediante qué métodos**.

Reglamento General de Protección de Datos (RGPD)

Marco legal en la **Unión Europea** (2016).

Establece **normas muy estrictas** para las organizaciones que **violen la privacidad de los datos**.

Afecta a **organizaciones de la UE**, o **con empleados en la UE** o **clientes en la UE**.

Datos personales

Podemos determinar **la identidad de la persona de forma directa o indirectamente**.

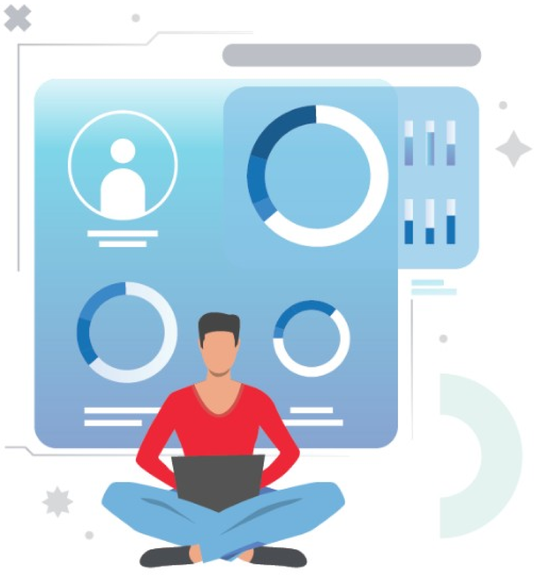
Ej. Nombre completo, DNI, pasaporte.

Datos sensibles

Son datos personales que revelan:

- el **origen étnico o racial**
- las **opiniones políticas**
- las **convicciones religiosas o filosóficas**
- la **afiliación sindical**
- los **datos genéticos**
- los **datos biométricos** dirigidos a identificar de manera unívoca a una persona física
- los **datos relativos a la salud**
- los **datos relativos a la vida sexual**

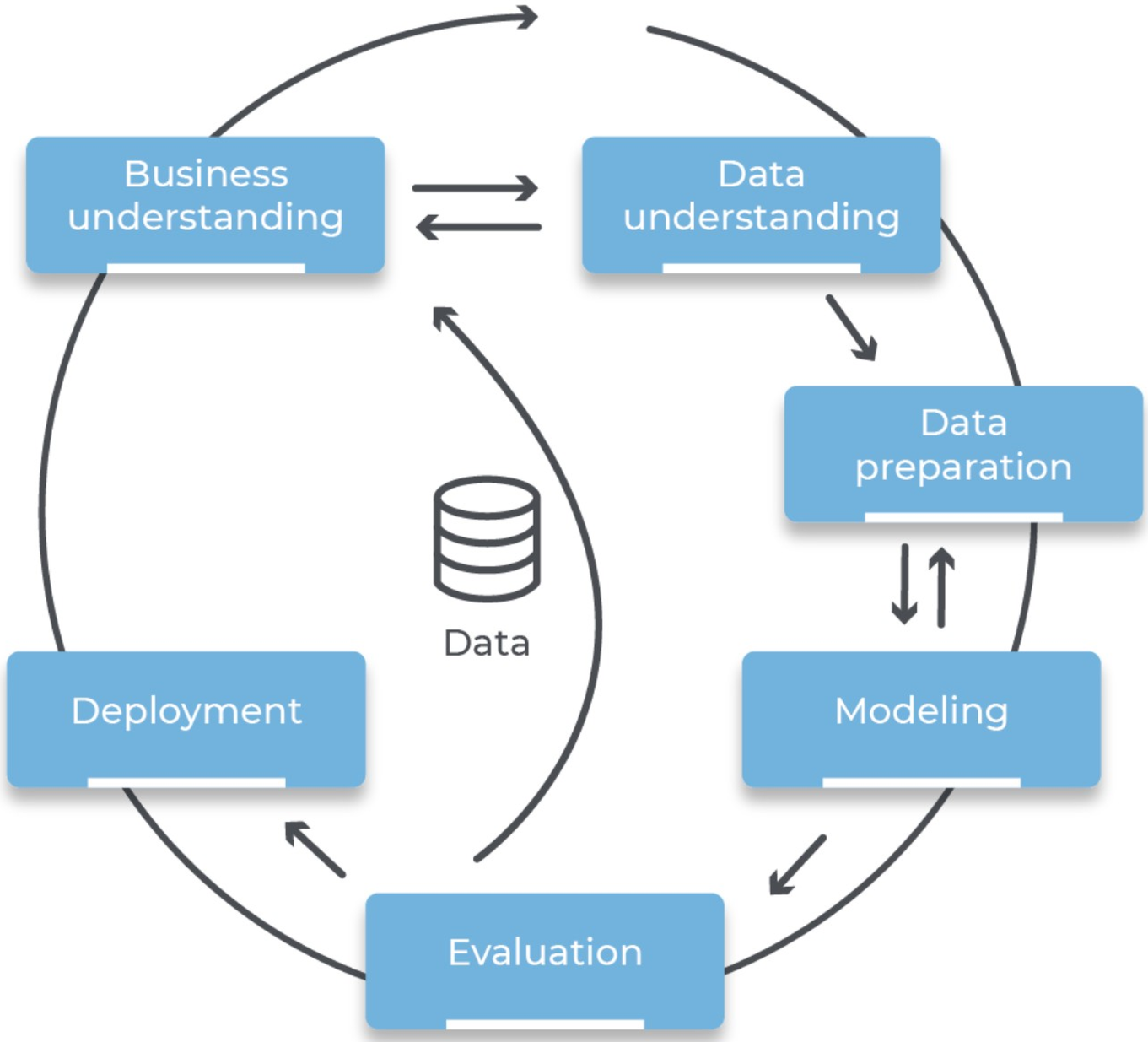
Es decir, datos que puedan ser utilizados como un elemento discriminatorio.



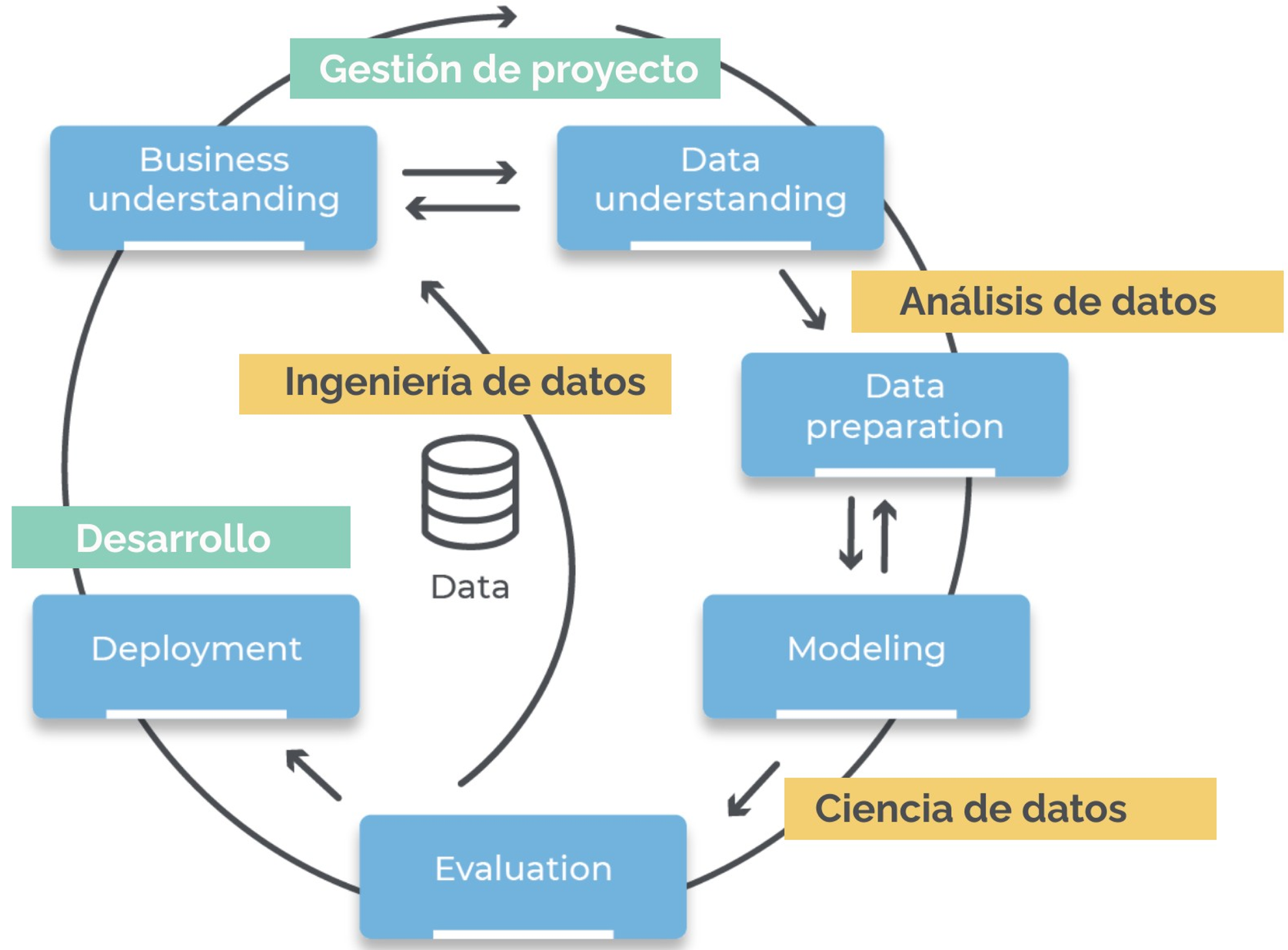
Metodología para realizar **proyectos de ciencia de datos**

mE

Proyecto de ciencia de datos



Proyecto de ciencia de datos

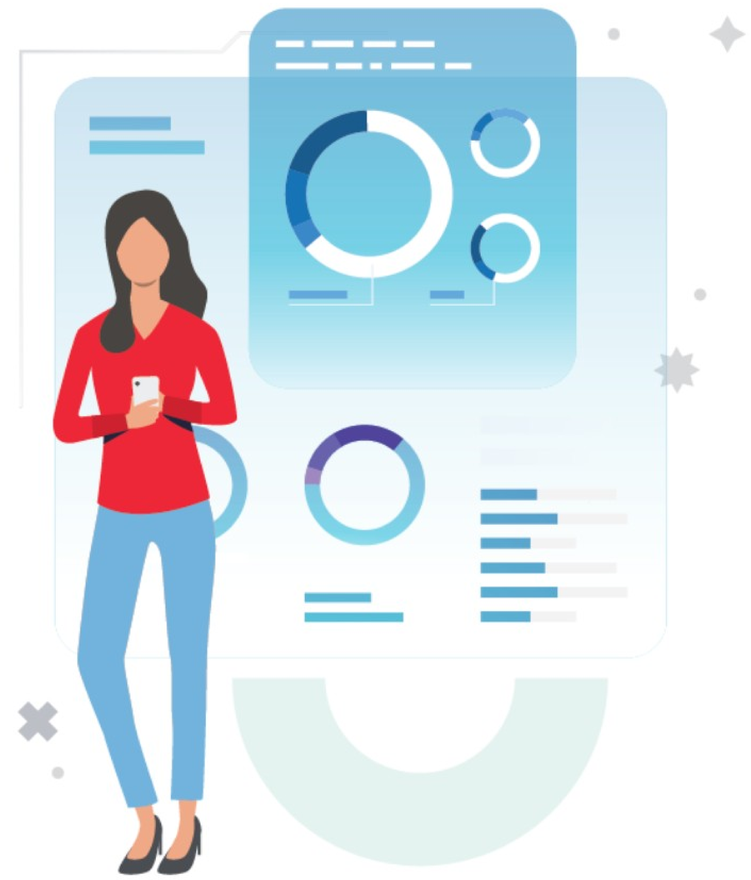


Analista de datos

- **Limpieza de los datos:** Realizan la limpieza de datos.

- **Minería de datos:** Ayudan a las organizaciones a tomar mejores decisiones mediante el análisis descriptivo de los datos.

- **Reporting y desarrollo de dashboards:** Generan informes para los clientes reportando el comportamiento de la organización.



Ingenierx de datos

- **Arquitectura de datos:** Diseña, desarrolla, monitorea y mantiene la arquitectura del almacenamiento de los datos.

- **Proceso ETL:** Colabora en la automatización de los procesos ETL (extracción, transformación y carga de los datos).

Ingenierx de Big Data: hace lo mismo pero con sistemas para Big Data.

Científicx de datos

- **Promueven la generación de conocimiento:** desarrollan las preguntas, diseñan los experimentos, implementan modelos de machine learning e interpretan los resultados para encontrar soluciones enfocadas en el negocio.

*Un buen data scientist puede **asesorar a los otros roles** porque **conoce las necesidades del negocio** y cuenta con un **amplio repertorio de herramientas**.*

Ok guys, it's showtime.

MUCHAS GRACIAS!

UNC

FCEFyN

mundosE
PEOPLE & BUSINESS SCHOOL