

DATA SCIENCE

Analítica Predictiva y Evolución de BI



PI DATA STRATEGY & CONSULTING

PATRICIO DEL BOCA



PI Co-founder

@pdelboca

pdelboca@piconsulting.com.ar

RODRIGO BOBADILLA



Data Scientist en **PI**

@rodrib_08

rbobadilla@piconsulting.com.ar

AGENDA

Objetivo

Introducción

¿Qué es Data Science?

Soluciones y usos más frecuentes.

Ciclo de Vida de un Proyecto

Demo

Thumb Rules

Big Data

Conclusiones

Preguntas

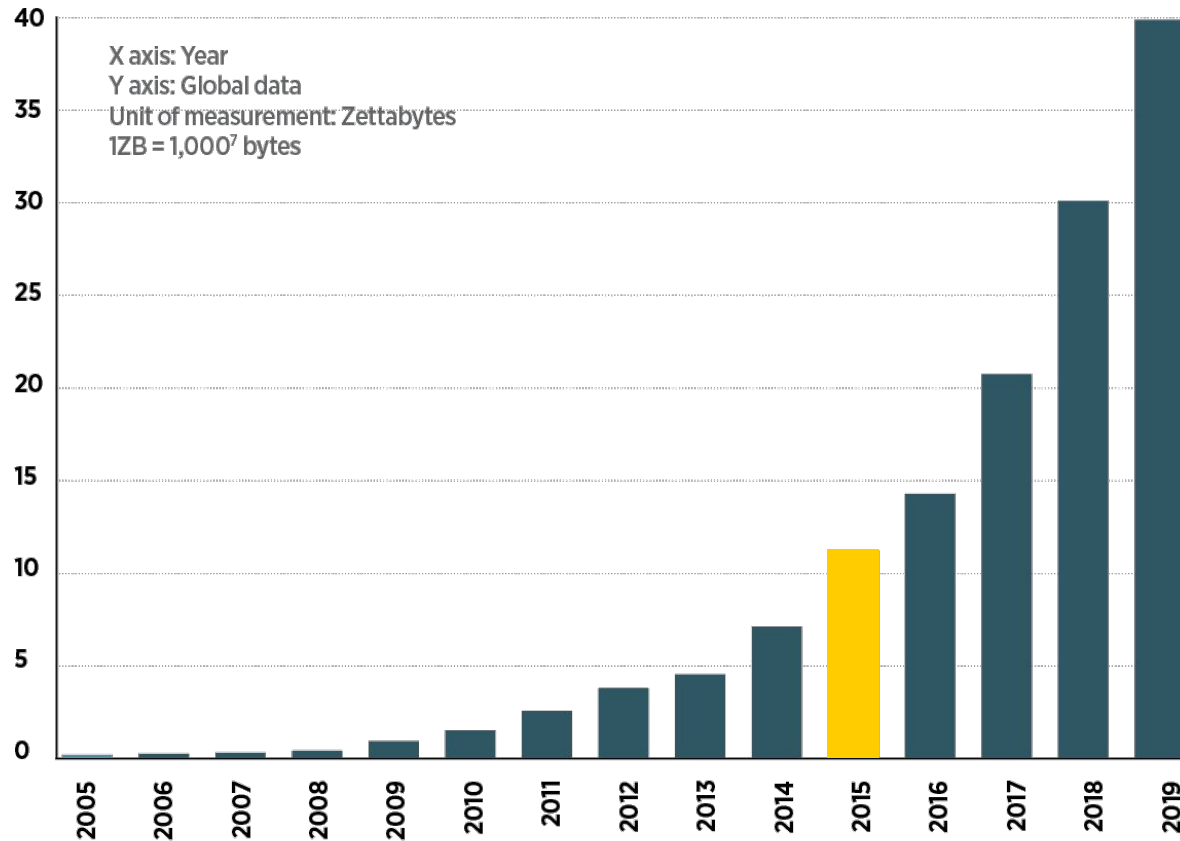
OBJETIVO

Dar una introducción a Data Science para conocer los conceptos y características fundamentales de esta disciplina.

Data Science?

2015

DATA GROWTH

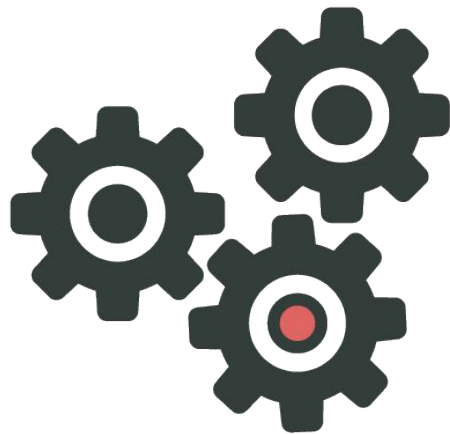


Note: Post-2013 figures are predicted. Source: UNECE

Bajos costos de procesamiento,



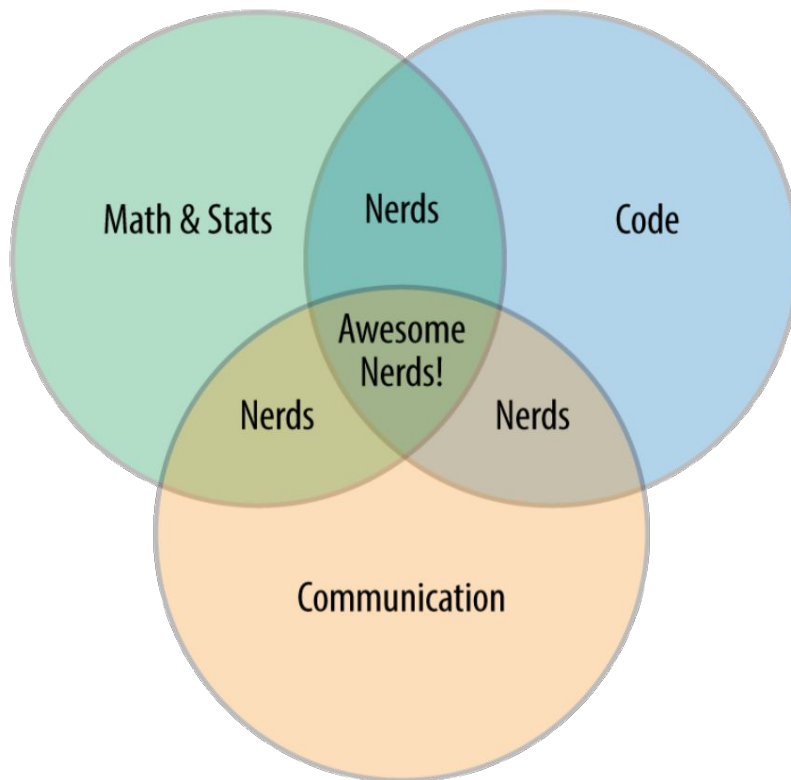
algoritmos,



y muchos datos.



Nuevo rol, Data Scientist.



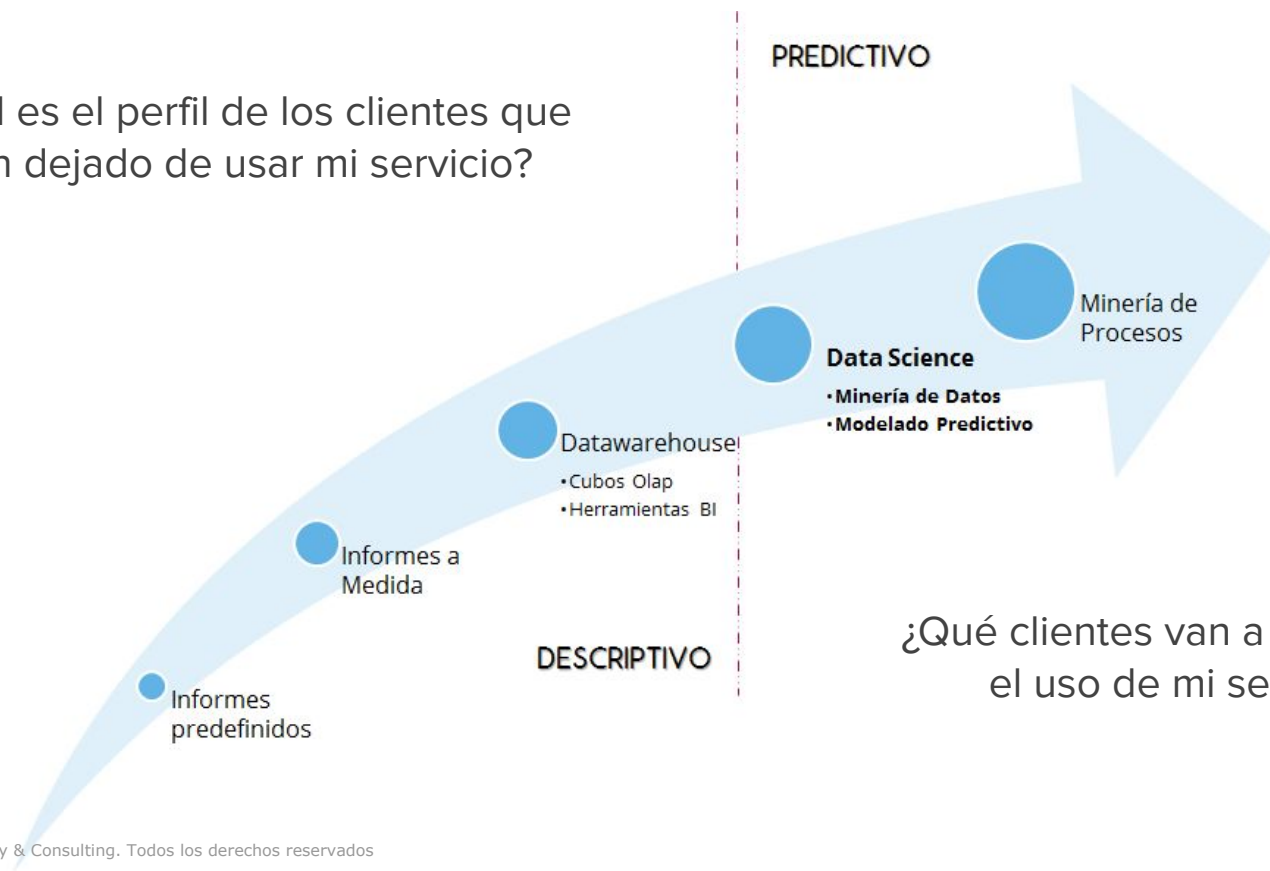


We define Data Science as managing the **process** that can transform hypothesis and data into **actionable** predictions.

- *Practical Data Science with R* -

EVOLUCION DE BI

¿Cuál es el perfil de los clientes que han dejado de usar mi servicio?



¿Qué clientes van a abandonar el uso de mi servicio?



Detección de Fraude



redefining / insurance

Detección de Patrones y Segmentación de Clientes

NETFLIX

Sistemas de Recomendación



Predicción de Abandono de Marca



Predicción de Demanda



Scoring y Riesgo Crediticio



Mantenimiento Predictivo

ALGORITMOS SUPERVISADOS

Clasificación

Regresión

Motores de Recomendación

Detección de Anomalías

ALGORITMOS NO SUPERVISADOS

Clustering

Clustering Jerárquico

Mapas auto organizados

Descubrimiento de Patrones

AGENDA

Objetivo ✓

Introducción ✓

¿Qué es Data Science? ✓

Soluciones y usos más frecuentes ✓

Ciclo de Vida de un Proyecto

Demo

Thumb Rules

Big Data

Conclusiones

Preguntas

CICLO DE VIDA

- CRISP-DM, el “SCRUM” de Data Science.



1. COMPRENSIÓN DEL NEGOCIO

- ¿Cuál es el valor que aporta al negocio este proyecto?
- ¿Cómo puede saber que sus resultados son precisos o efectivos?

2. COMPRENSIÓN DE LOS DATOS

- Análisis de completitud de los datos
- Búsqueda de errores
- Análisis de valores extremos

3. PREPARACIÓN DE LOS DATOS

- Limpieza de datos
- *Feature engineering*

4. MODELADO

- Generalización que busca explicar y predecir el comportamiento de los datos
- *“Essentially all models are wrong, but some are useful.” George E. P. Box*

5. EVALUACIÓN

- ¿Resuelve el modelo el objetivo que planteamos?
- Análisis de los errores y planificación de próximos pasos
- Si es demasiado bueno para ser verdad, probablemente lo sea! (Overfitting)

6. IMPLEMENTACIÓN

- Presentación de los resultados a todos los stakeholders
- Implementación del modelos en producción
- Tiempo Real vs Procesamiento en Lote
- Cloud vs On-Premise



At the end of the day, it is usually **how we frame the problem**, not the tools and techniques that we use to answer it, that determines how valuable our work is.

AGENDA

Objetivo ✓

Introducción ✓

¿Qué es Data Science? ✓

Soluciones y usos más frecuentes ✓

Ciclo de Vida de un Proyecto ✓

Demo

Thumb Rules

Big Data

Conclusiones

Preguntas

AGENDA

Objetivo ✓

Introducción ✓

¿Qué es Data Science? ✓

Soluciones y usos más frecuentes ✓

Ciclo de Vida de un Proyecto ✓

Demo ✓

Thumb Rules

Big Data

Conclusiones

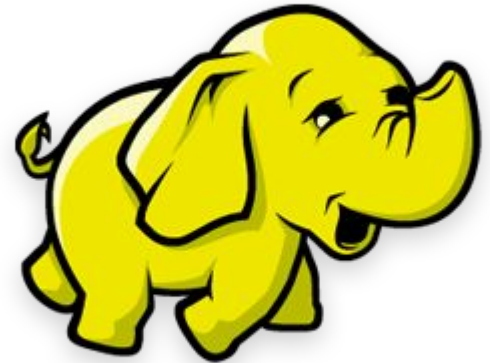
Preguntas

THUMB-RULES

- ¿Hay datos disponibles? ¿Etiquetados? ¿Cuántos?
- ¿Qué tipo de datos?
- Garbage In, Garbage Out.
- Iniciar siempre con una prueba de concepto.
- ¿Está el negocio comprometido con un trabajo de calidad?

BIG DATA

“Big Data es cuando el tamaño comienza a ser un problema y los datos ya no pueden ser procesados en una sola terminal de cómputo.” - *Consenso Popular* -



CONCLUSIONES

- Data Science es un proceso iterativo que requiere de muchas idas y vueltas entre cada etapa
- Lo más importante en un Proyecto es la pregunta que se hace para abordar una problemática concreta, más aún que las herramientas que se usen.

PREGUNTAS?



MUCHAS GRACIAS!



<https://github.com/PiConsulting>



info@piconsulting.com.ar



PI DATA STRATEGY & CONSULTING