

FACULTAD  
DE CIENCIAS  
ECONÓMICAS



Universidad  
Nacional  
de Córdoba

# Introducción a la Ciencia de Datos con Python

EDICIÓN 2025

UNIDAD 1 – CIENCIA DE DATOS

## PRESENTACIÓN – Edición 2025

### Docentes clases Prácticas/Taller:

Lic. María Inés Ahumada - [ines.ahumada@unc.edu.ar](mailto:ines.ahumada@unc.edu.ar)

Lic. Pablo Romero - [pablo.romero@mi.unc.edu.ar](mailto:pablo.romero@mi.unc.edu.ar)

### Comunicación:

 SLACK - Int. Ciencia de Datos 2025 

 Slack será nuestro espacio pa

Podrás usar Slack en múltiples dispositivos:

### Materiales y actividades:



## Modalidad de cursado

- Clases presenciales teórico-prácticas (taller)
  - Aviso por Slack unos días antes de cada clase para avisar temas
  - No habrá clases grabadas. Es importante asistir a las clases
  - Todos los materiales (por ejemplo, este) y notebooks de Python disponibles en el aula virtual
- 

## Unidad 1: Contenidos

- **Ciencia de datos:** definición y su importancia en la actualidad. Aplicaciones prácticas en el campo de las ciencias económicas.
- **Introducción a la analítica de datos** (Data Analytics): terminología y conceptos fundamentales, categorías de Data Analytics: descriptivo, predictivo, prescriptivo.
- **El proceso de Ciencia de Datos:** roles comunes, datos, obtención de datos, curación e ingeniería de características (feature engineering).
- **Big data:** introducción y nociones básicas. Identificación de grandes volúmenes de datos
- Arquitectura de datos. Business Intelligence.
- Disponibilidad de datos
- Gobernanza de los datos
- Ética en la ciencia de datos.
- Software de programación en ciencia de datos

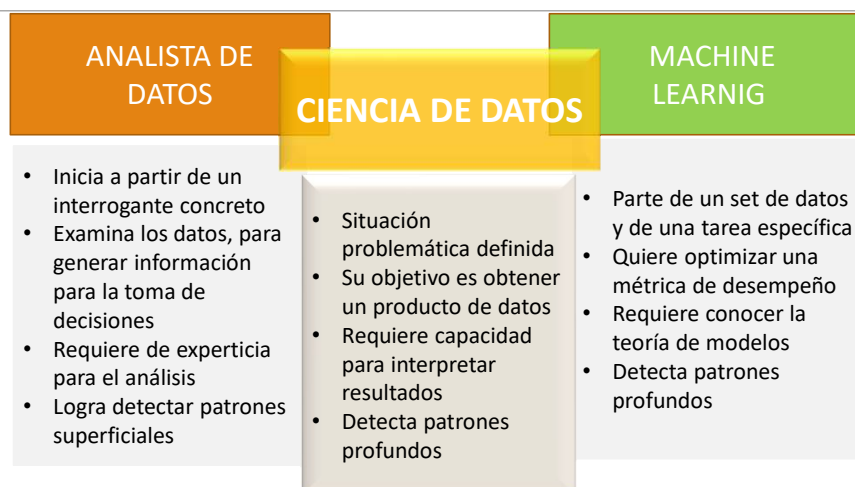
# CONCEPTOS

"Lo que hacen los científicos de datos es hacer descubrimientos mientras nadan en los datos... [su] rasgo dominante es una intensa curiosidad: un deseo de ir más allá de la superficie de un problema, encontrar las preguntas en su centro y destilarlas en un conjunto muy claro de hipótesis, que se puede probar. Esto a menudo implica el pensamiento asociativo que caracteriza a los científicos más creativos en cualquier campo... tal vez se está volviendo claro que la palabra 'científico' se ajusta a este rol emergente... su mayor oportunidad para agregar valor no es crear informes o presentaciones para altos ejecutivos, sino innovar con productos y procesos orientados al cliente".

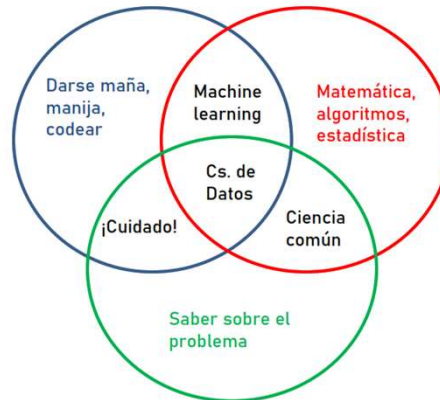
**Ciencia de datos:** Predecir un comportamiento basado en métodos estadísticos, encontrando patrones.

**Científico de datos:** "los científicos de la información y la computación, las bases de datos y ingenieros y programadores de software, expertos disciplinarios, curadores y anotadores expertos, bibliotecarios, archivistas y otros, que son cruciales para la gestión exitosa de una colección de datos digitales" (Data Collections: Enabling Research and Education in the 21st Century- 2005) Computadoras emulando comportamientos humanos específicos.

## La ciencia de datos



# La ciencia de datos, desde la práctica



# Todo deja huellas... datos y más datos

LN **Página12**

**ClarínX**



Google



edenor



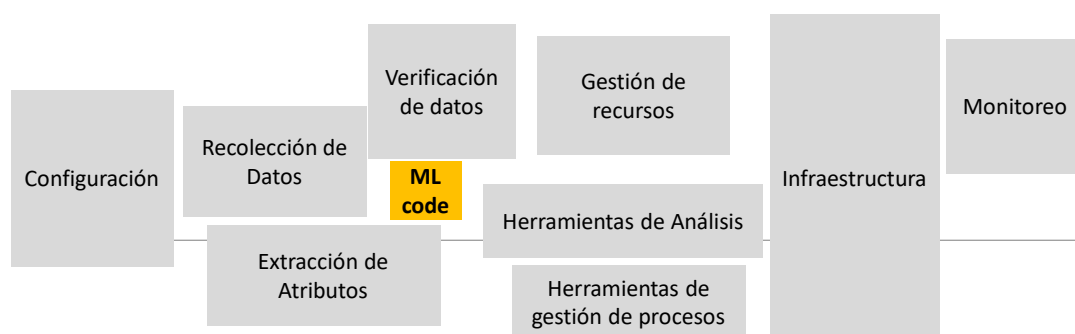
# CONCEPTOS

**Inteligencia Artificial:** capacidad de las máquinas de mostrar habilidades y comportamientos “inteligentes”. Computadoras emulando comportamientos humanos específicos.

**Machine Learning:** (=Aprendizaje automático) se **considera un subconjunto** de IA. Se trata de una de las maneras de las que disponemos para lograr que las máquinas aprendan y "piensen" como los humanos. Las técnicas de ML se utilizan cuando queremos que las máquinas aprendan de la información que les proporcionamos. Predecir un comportamiento basado en métodos estadísticos, encontrando patrones, etc.

**Deep Learning:** Subconjunto de ML, sin necesidad de definir características pero requiere de mucho mayor volumen de datos.

**ML** Las dependencias de datos cuestan más que las dependencias de código



## ¿cuáles son sus herramientas?



### Análisis / estadística



### Procesamiento, análisis RT



### Almacenamiento / procesamiento

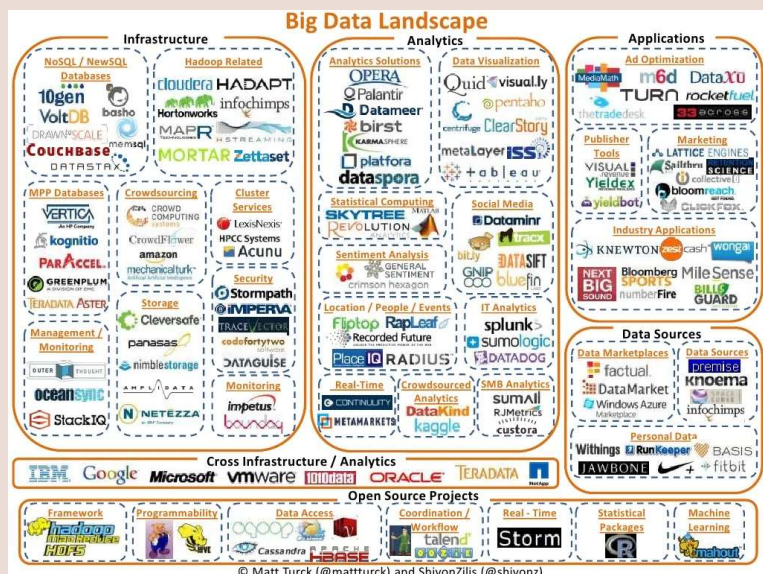


### Visualización



## Un gráfico del ecosistema de big data

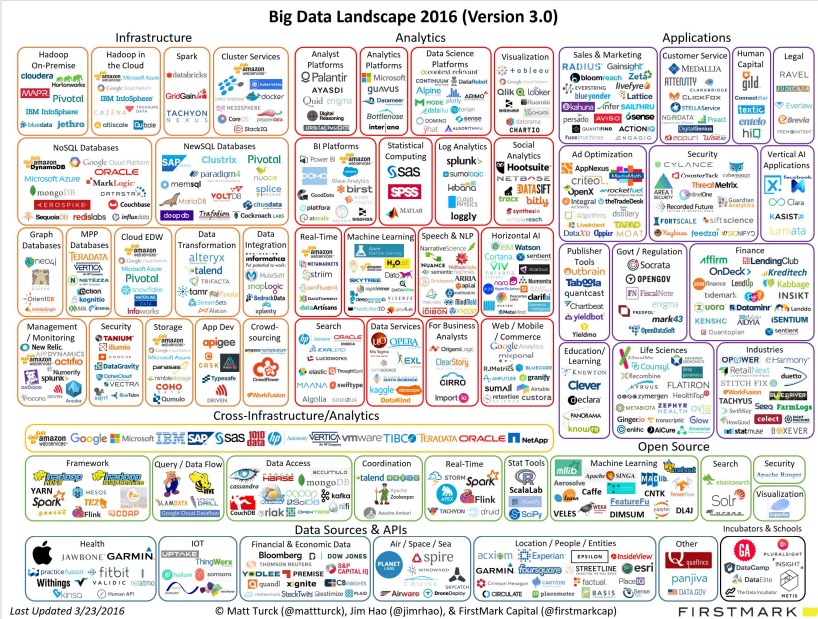
Año 2012





# Un gráfico del ecosistema de big data

Año 2016

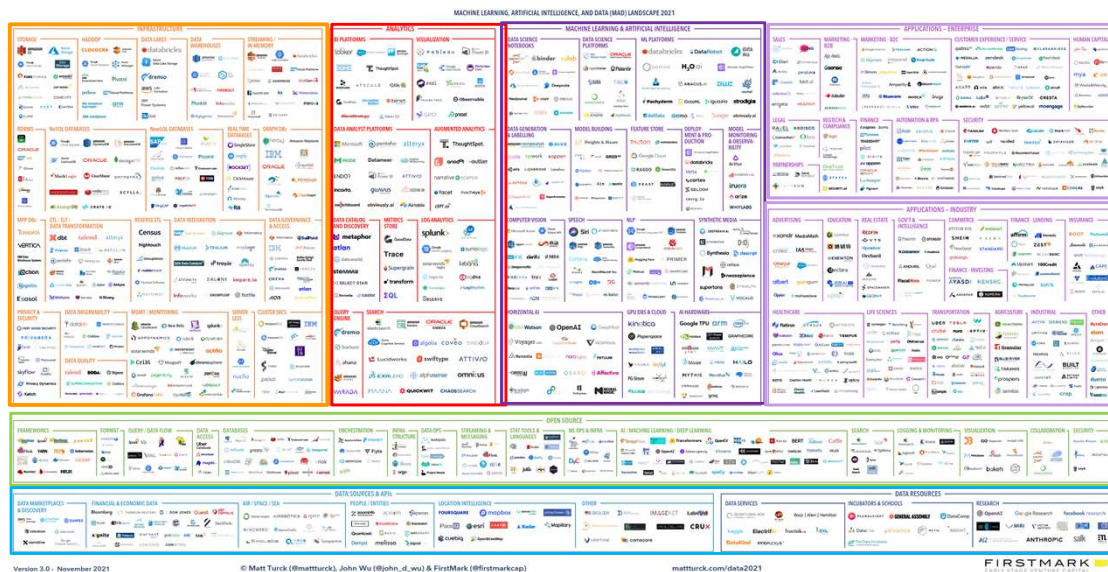


17/3/2025

LIC. MAARÍA INÉS AHUMADA FCE - UNC

13

## Machine learning, Artificial intelligence and Data (MAD) 2021

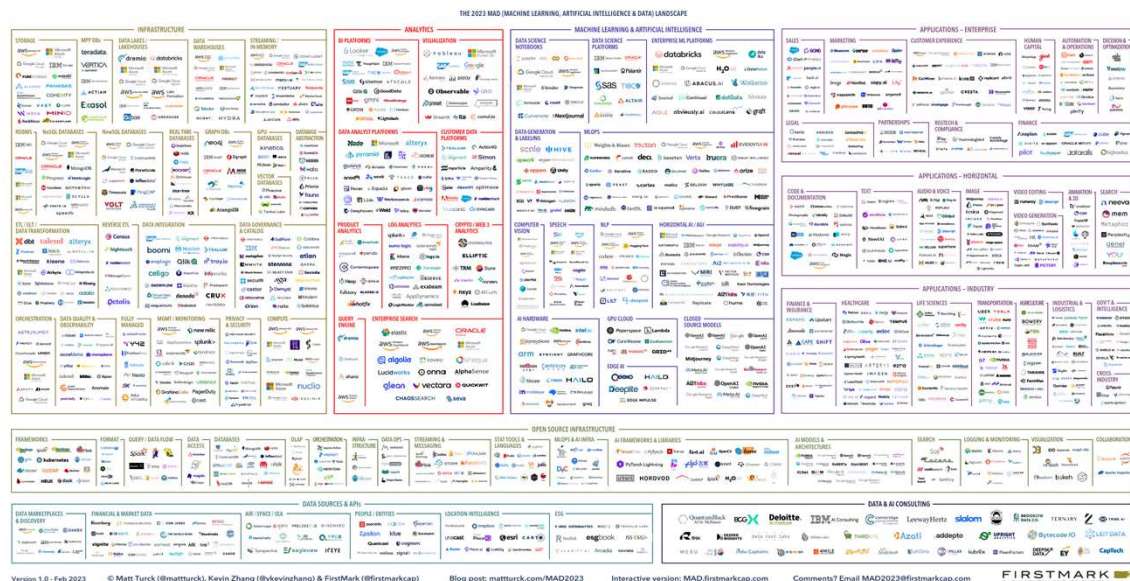


17/3/2025

LIC. MAARÍA INÉS AHUMADA FCE - UNC

14

# Machine learning, Artificial intelligence and Data (MAD) 2023



17/3/2025

LIC. MAARÍA INÉS AHUMADA - FCE - UNC

15

## La visión macro: dar sentido a la complejidad del ecosistema

Históricamente, y aún hoy en muchas organizaciones, los datos han significado **datos transaccionales** almacenados en **bases de datos relacionales**, y quizás algunos tableros para el análisis básico de lo que sucedió con el negocio en los últimos meses.

Ahora, las empresas marchan hacia un mundo en el que los datos y la inteligencia artificial se integran en innumerables procesos internos y aplicaciones externas, tanto con fines analíticos como operativos.

**Este es el comienzo de la era de la empresa inteligente y automatizada** : donde las métricas de la empresa están disponibles en tiempo real, las solicitudes de hipotecas se procesan automáticamente, los chatbots de IA brindan atención al cliente las 24 horas del día, los 7 días de la semana, se predice la rotación, las amenazas cibernéticas se detectan en tiempo real, las cadenas de suministro se ajustan automáticamente a las fluctuaciones de la demanda, etc.



Matt Turk  
@matturk

2013: Big Data

2017: ML/AI

2020: Automation

Not overhyped fads. Different stages of the same, massive secular trend.

5:58 PM · 6/14/20 · Twitter for iPhone

||| View Tweet activity

39 Retweets 6 Quote Tweets 234 Likes

17/3/2025

LIC. MAARÍA INÉS AHUMADA - FCE - UNC

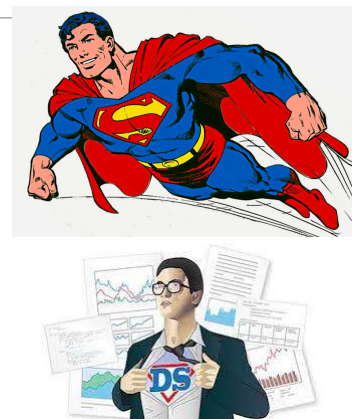
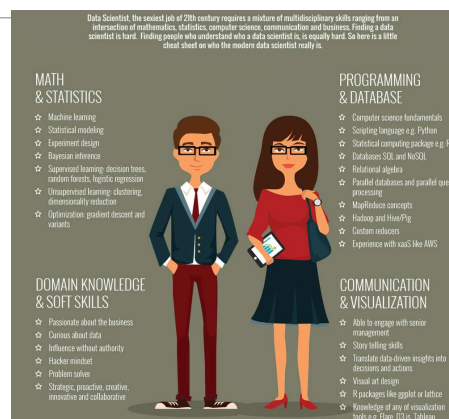
16



## A todo vapor: el panorama MAD (aprendizaje automático, inteligencia artificial y datos) de 2024

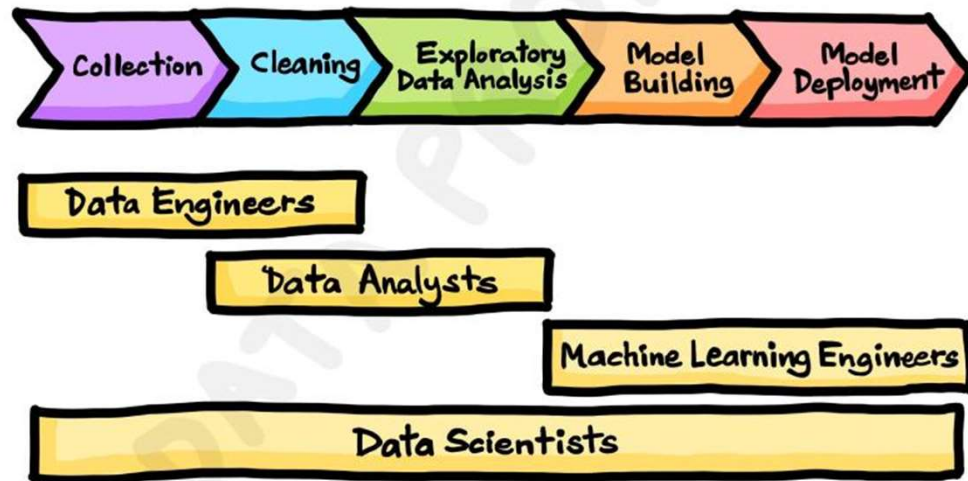
El ecosistema MAD (ML, AI & Data) ha pasado de ser un nicho técnico a convertirse en algo convencional . El cambio de paradigma parece estar acelerándose con implicaciones que van mucho más allá de las cuestiones técnicas o incluso comerciales, y afectan a la sociedad, la geopolítica y quizás la condición humana.

## El científico de datos



<https://visual.ly/community/Infographics/business/what-makes-good-data-scientist>

## Roles en ciencia de datos



17/3/2025

LIC. MAARÍA INÉS AHUMADA FCE - UNC

<https://www.youtube.com/dataprofessor>

## El Proceso de Ciencia de Datos

Mercado. "Datos son negocio"

Desafío. Diseñar estrategias *Big Data* a la medida de los objetivos de la organización para impulsar la innovación y el crecimiento de tu empresa.

Aportar un alto valor en los proyectos de Análisis y Explotación de datos en todos los ámbitos

**Descriptivo:** Data Warehouse, Data Marts, Cuadros de Mando

**Predictivo:** Modelos de Machine Learning / Deep Learning

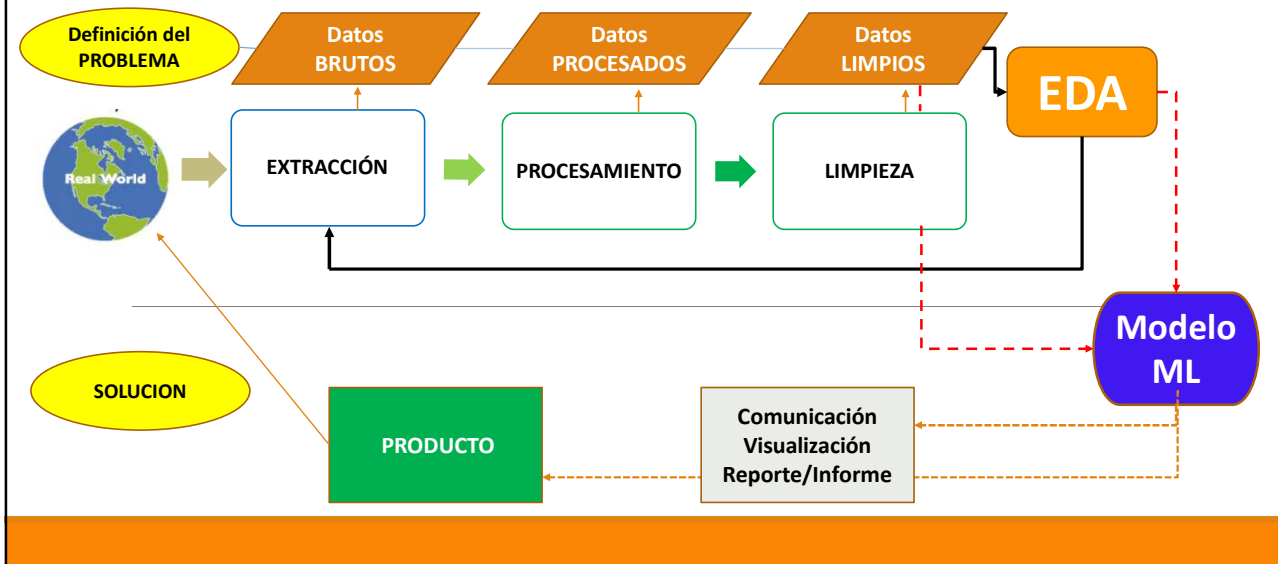
**Prescriptivo:** Recomendaciones sobre estrategias y acciones recomendadas en base a los resultados de los análisis descriptivo y predictivo

17/3/2025

LIC. MAARÍA INÉS AHUMADA FCE - UNC

20

## El proceso de Ciencia de Datos



## Identificar el problema

- ¿Cómo resolverlo con datos?
- Hablar con los especialistas y entender su relación con los datos
- Entender qué se sabe y qué no se sabe sobre el problema
- Si el problema todavía no fue abordado, ¿por qué?
- ¿Cuáles son los beneficios de trabajar en este problema y no en otro?
- ¿Existe realmente un problema?

## Recolectar datos

- ¿Qué fuentes de datos conocen los especialistas?
  - ¿Existen otras fuentes de datos que puedan servir para resolver el problema?
  - Evaluar la calidad de los datos
  - ¿Quién es el propietario de los datos?
  - Si los datos no existen, ¿cómo podemos conseguirlos?
  - ¿Cuántos datos necesitamos? ¿Cuáles son los posibles sesgos de las fuentes de datos? Los datos, ¿son independientes entre sí? ¿cuál es el espacio muestral?
- 

## Exploración de datos

Al explorar los datos en su conjunto

### **Considere:**

- Todas las herramientas de análisis que conoce
  - Técnicas más complejas para el análisis multivariado de datos
  - Estudiar la relación entre variables
  - Técnicas de visualización
-

## Curación de datos

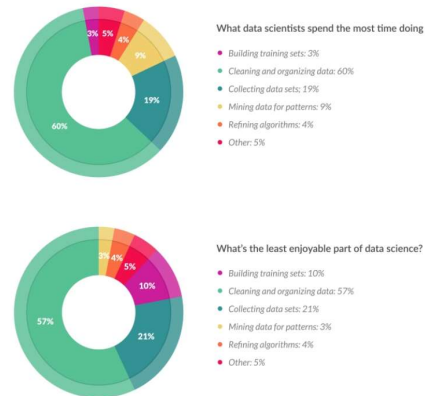
Los científicos de datos dedican el 60% de su tiempo a limpiar y organizar datos. La recopilación de conjuntos de datos ocupa el segundo lugar con el 19 % de su tiempo

*El 76% de los científicos de datos ven la preparación de datos como la parte menos agradable de su trabajo.*

El 57 % de los científicos de datos considera que limpiar y organizar los datos es la parte menos agradable de su trabajo y el 21 % dice esto sobre la recopilación de conjuntos de datos.

### Data Scientists Spend Most of Their Time Cleaning Data

Posted on May 3, 2016



## Curación de datos

Problema	Datos	Decisiones de curación
Predecir el precio de una propiedad en la ciudad de Córdoba	Base de datos gubernamental con registros de transacciones inmobiliarias. Tiene precio, fecha y ubicación.	<ul style="list-style-type: none"> <li>Eliminar día y mes de la transacción.</li> <li><i>Escrapear</i> sitios de compra/venta para extraer información adicional sobre cada propiedad.</li> <li>Imputar los valores faltantes utilizando estimaciones en base a ejemplos similares</li> </ul>



## Desarrollar modelo

¿Tenemos una hipótesis? ¿O estamos explorando a ver qué encontramos?

¿Qué variables importantes podríamos no estar teniendo en cuenta?

¿Cuáles son nuestros propios sesgos a la hora de analizar los datos?

Si el objetivo es elaborar un informe cualitativo o cuantitativo, ¿están las conclusiones justificadas por nuestro análisis? ¿Estamos influenciados por las expectativas de quién va a leer el informe? ¿Ocultamos resultados que contradicen las conclusiones?

Si desarrollamos un modelo predictivo o de machine learning, ¿es el modelo trivial? ¿se basa el modelo únicamente en datos anteriores a lo que queremos predecir?

¿Entendemos cómo funciona el modelo? ¿es ético?

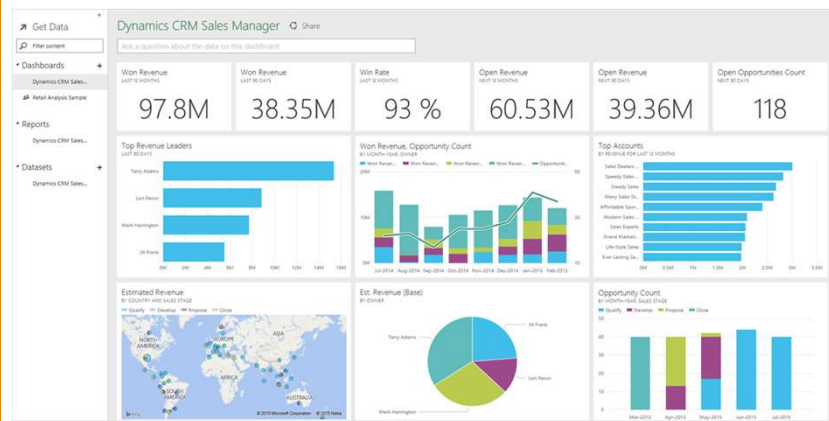
## Solución de datos: Visualizar y comunicar resultados

- Conocer al destinatario ¿Entendemos a quién le estamos comunicando, cuáles son sus conocimientos e intereses y cómo difieren de los nuestros?
  - Considerar aspectos de estética, eficiencia y eficacia al presentar los resultados
  - Énfasis en aquellos que estamos convencidos en comunicar al otro ¿Qué y por qué?
  - ¿Estamos siendo honestos? ¿Estamos diciendo lo que el otro quiere escuchar?
  - ¿Comprendemos el alcance de nuestro reporte?
  - ¿Qué respuesta esperamos de nuestra audiencia?
-

## Visualización de datos

### Dashboard

transformando los datos en información

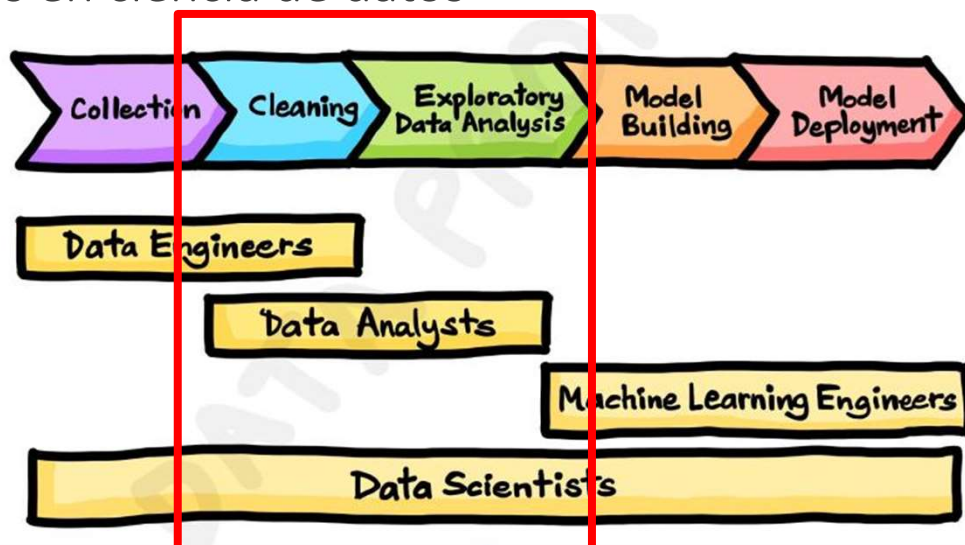


17/3/2025

LIC. MAARÍA INÉS AHUMADA FCE - UNC

29

## Roles en ciencia de datos



17/3/2025

LIC. MAARÍA INÉS AHUMADA FCE - UNC

<https://www.youtube.com/datoscientific>

## CONCEPTO

**Big Data:** “Se refiere a aquellos conjuntos de datos cuyo tamaño excede la capacidad de las herramientas actuales de captura, almacenamiento, carga, y análisis” ([McKinsey Global Institute- Mayo 2011](#)). Disponibilidad de grandes volúmenes de data y la capacidad de procesarlos.

El éxito de Big Data no se trata de implementar una pieza de tecnología (como Hadoop o cualquier otra), sino que requiere armar una línea de ensamblaje de tecnologías, personas y procesos.

Se necesita **capturar datos, almacenar datos, limpiar datos, consultar datos, analizar datos, visualizar datos**. Parte de esto lo harán los productos, y parte lo harán los humanos. Todo debe integrarse a la perfección. En última instancia, para que todo esto funcione, toda organización (empres/estado), comenzando por la alta gerencia, debe comprometerse a construir una cultura basada en datos.

## Principales usos del Big Data

- Mejora en el **conocimiento de cliente**: análisis del sentimiento y experiencia del cliente a través de todos los canales, segmentación de clientes para la personalización de las acciones comerciales, cálculo del valor potencial del cliente
- Apoyo en la **toma de decisiones**, al disponer de un mayor detalle de la información disponible, fuentes más veraces y obtención de predicciones mediante algoritmos automatizados.
- **Optimización** de los procesos de negocio, mediante el análisis de la información generada durante el proceso, y ahorro de costes, con el uso de infraestructuras más ágiles y baratas
- **Nuevos modelos de negocio**, innovadores, de productos y servicios, basados en el conocimiento aportado por los datos.

## ¿Qué es distinto en Big Data?

Generadores de Datos



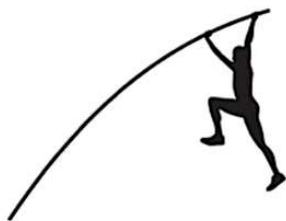
Tipos de formatos



Tipos de análisis



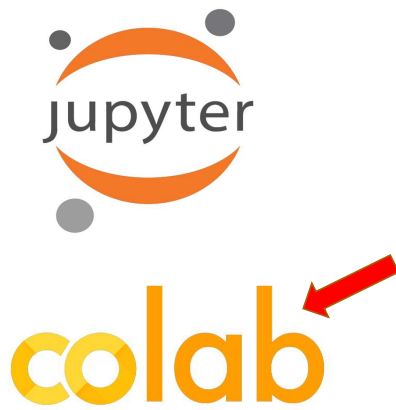
## Qué más es distinto?



**Nuevo paradigma!!!**



## Aplicaciones



**Jupyter notebooks:** Aplicación web combinación de código, texto e imágenes que se pueden correr desde un navegador

[www.jupyter.org](http://www.jupyter.org)

**Notebooks de Google Colab:** es un entorno de desarrollo basado en la nube. Combina código, texto e imágenes que se pueden correr desde un navegador en la nube de Google

<https://colab.research.google.com/>

## Notebooks de Python

```
[ ] x0 = 1 # condicion inicial
    N = 1000
    dt = 2/1000
    x_tiempo = np.zeros(N) # crea un vector de longitud N donde vamos a ir poniendo las soluciones
    x_nuevo = x0
    x_tiempo[0] = x_nuevo
```

Ahora el bloque que se repite desde 1 hasta N lo corremos con una función llamada "for" que itera los números en una lista desde 1 hasta N, que creamos con el comando de numpy `np.arange(1,N)`

```
[ ] for i in np.arange(1,N):
    x_viejo = x_nuevo
    x_nuevo = x_viejo + x_viejo*dt # aca ponemos f(x_viejo)*dt. En este caso, f(x_viejo) = x_viejo
    x_tiempo[i] = x_nuevo
```

Entonces, en `x_tiempo` quedaron las soluciones de  $x(t)$  en 1000 puntos separados por  $dt$  entre  $t=0$  y  $t=2$ . Podemos graficar esto usando el comando `plt.plot`, donde antes tenemos que importar la librería `matplotlib.pyplot`

Cada clase-taller va a estar acompañada por una Notebook

Los ejercicios o trabajos prácticos se entregan como Notebook, con explicaciones (texto) y código, que al ejecutarse cumplen la consigna (más sobre esto hoy)



---

Muchas gracias !!