

SESIÓN 2

# FUNDAMENTOS DE CIENCIA DE DATOS

**U** UPAEP

# Agenda

- Temas de interés del área de Ciencia de Datos.
- M1: Introducción a Ciencia de Datos.
  - La gestión de los datos.
  - Actividad 1: Del bit ... al Big Data.
  - El concepto de Big Data (Datos Masivos).
    - ¿Qué es Big Data?
    - Rasgos característicos.
- ClickStream01: Instalación de MySQL y complemento de Excel.

# Temas de interés del área de Ciencia de Datos

***Presentación de noticia sobre Ciencia de Datos  
(Jorge Rubén Vázquez y Ricardo Baez)***

***Foro: Participación 03: El dilema de las redes sociales***

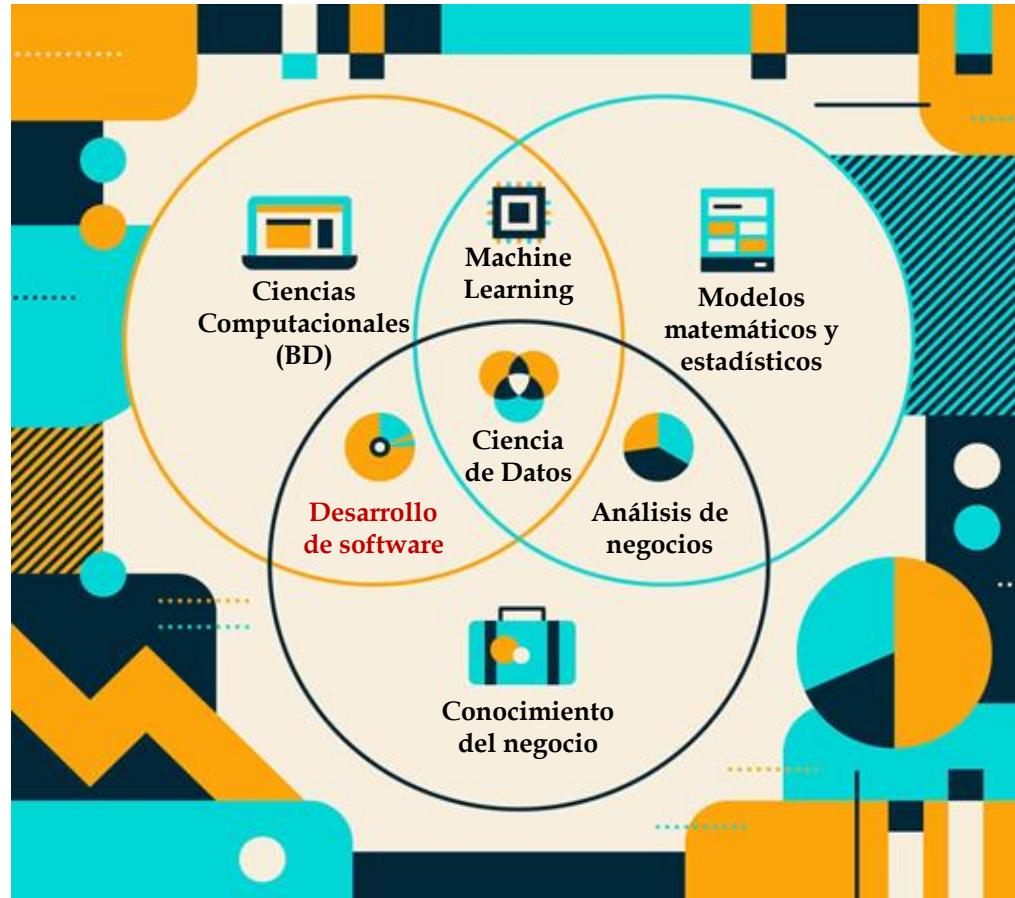
1. Menciona dos ventajas y dos desventajas del uso de las redes sociales.
2. ¿Qué opinas de lo que comparte el creador del botón “Me gusta” de Facebook?
3. Según tu opinión: ¿Cuál es el objetivo principal de las redes sociales?
4. Opcional: ¿Cuál es tu reflexión, con respecto a **tu** uso de redes sociales?



Fundamentos de Ciencias de Datos

## **M1 - Introducción a Ciencia de Datos**





Fuente: Elaboración propia con información de <https://clevertap.com/blog/data-science/>

# M1 – La Gestión de los Datos

## ***Tipos de productos de Ciencia de Datos***

1. Mis datos son valiosos para mi (in  $\rightarrow$  in).

Datos internos útiles para la organización.

Inteligencia empresarial clásica. Existen muchas oportunidades todavía.

2. Esos datos son valiosos para mi (out  $\rightarrow$  in).

Datos externos útiles para la organización.

Medios sociales, Internet, datos abiertos. Crear nuevas oportunidades.

3. Mis datos son valiosos para otros (in  $\rightarrow$  out).

Datos internos útiles para otras organizaciones.

Mis datos tienen utilidad para otros. Crear nuevas oportunidades.

# M1 – La Gestión de los Datos

## ***Tipos de productos de Ciencia de Datos***

4. Esos datos son valiosos para otros (out  $\rightarrow$  out).

Datos externos útiles para otras organizaciones.

Estos datos tienen utilidad para otros. ¡Científico de datos *freelancer*!

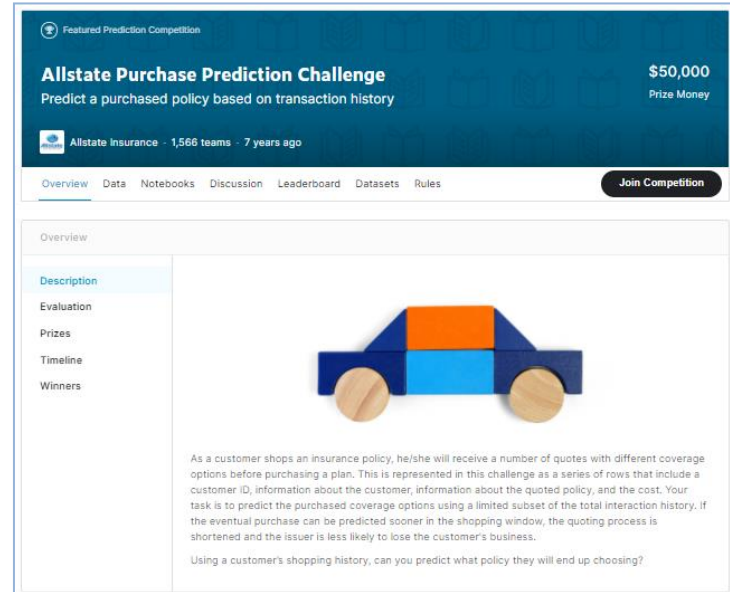
5. Creando datos ( $\emptyset \rightarrow$  out).

Coleccionar datos que pueden tener valor. ¡Emprendedor de datos!

# M1 – La Gestión de los Datos

## ***Ejemplos de productos basados en datos (in → in)***

- Una compañía de seguros de automóviles, Allstate quiere predecir la póliza que se comprará dado el historial de transacciones de sus clientes.



<https://www.kaggle.com/c/allstate-purchase-prediction-challenge>



# M1 – La Gestión de los Datos

## ***Ejemplos de productos basados en datos (in → out )***

- Smart Steps recolecta datos en tiempo real de subsidiarios de Telefónica (Movistar , O<sub>2</sub>, Vivo), con el objetivo de vender a otras compañías los datos, las herramientas y la habilidad de analizarlos y representarlos.

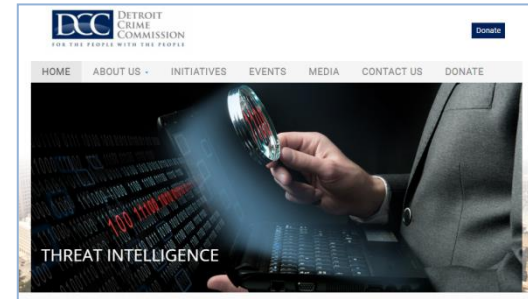


<https://www.business-solutions.telefonica.com/es/products/big-data/business-insights/smart-steps-product/>

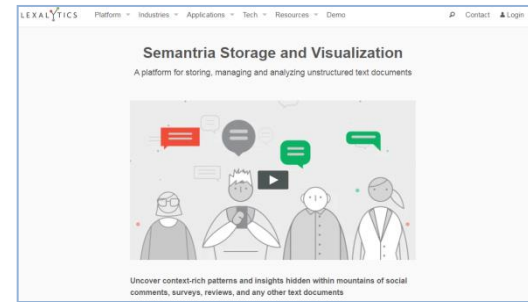
# M1 – La Gestión de los Datos

## ***Ejemplos de productos basados en datos (out → in)***

- La Comisión del Crimen de Detroit (DCC) reconoció que muchos delincuentes publicaban sobre sus crímenes en varias plataformas de redes sociales, anunciando potenciales planes, alardeando de drogas y armas en Facebook, Twitter e Instagram, y organizando su próximo movimiento. DCC decidió aprovechar esta información abierta al asociarse con Semantria para introducir análisis de texto que le permitiría al equipo rastrear criminales, actividades y consecuencias.



<https://detroitcrimecommission.org/>

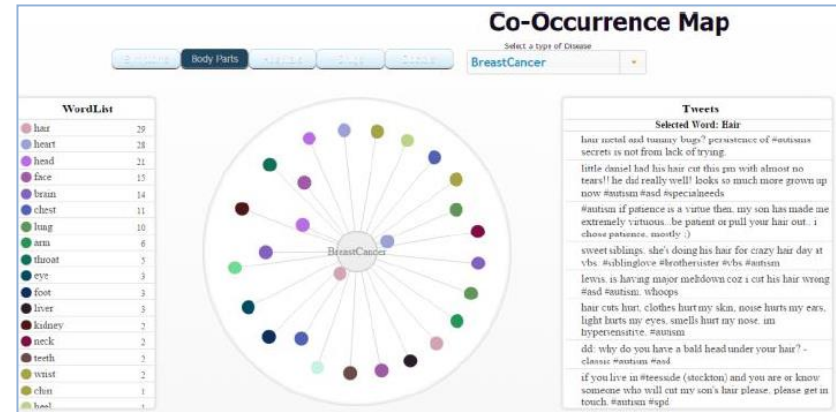


<https://www.lexalytics.com/ssv>

# M1 – La Gestión de los Datos

## ***Ejemplos de productos basados en datos (out → out)***

- Healthcaredataanalysis.org fue un experimento para mostrar que los tweets podrían brindar información valiosa sobre el efecto de las enfermedades y la relación entre los síntomas y las drogas.



# M1 – La Gestión de los Datos

## ***Ejemplos de productos basados en datos ( $\emptyset \rightarrow out$ )***

- Cuando se conduce con la aplicación de Waze abierta, se comparte información en tiempo real que se utiliza para conocer el estado del tráfico y la estructura de las vías. Los usuarios también reportan tráfico, accidentes, controles policiales, vías bloqueadas, condiciones meteorológicas, etc. Waze recoge esta información y la analiza al instante, para proporcionar a otros wazers la mejor ruta para llegar a su destino en cualquier hora del día.



<https://www.waze.com/es/>

# M1 – La Gestión de los Datos

## ***Fundamentos del trabajo con Datos***

¿Qué es un **dato**?

“Antecedente necesario para llegar al conocimiento exacto de algo o para deducir las consecuencias legítimas de un hecho”.

Representación **simbólica** de un atributo o variable **cuantitativa** o **cualitativa**.

# M1 – La Gestión de los Datos

## ***Fundamentos del trabajo con Datos***

### **Clasificación de los datos**

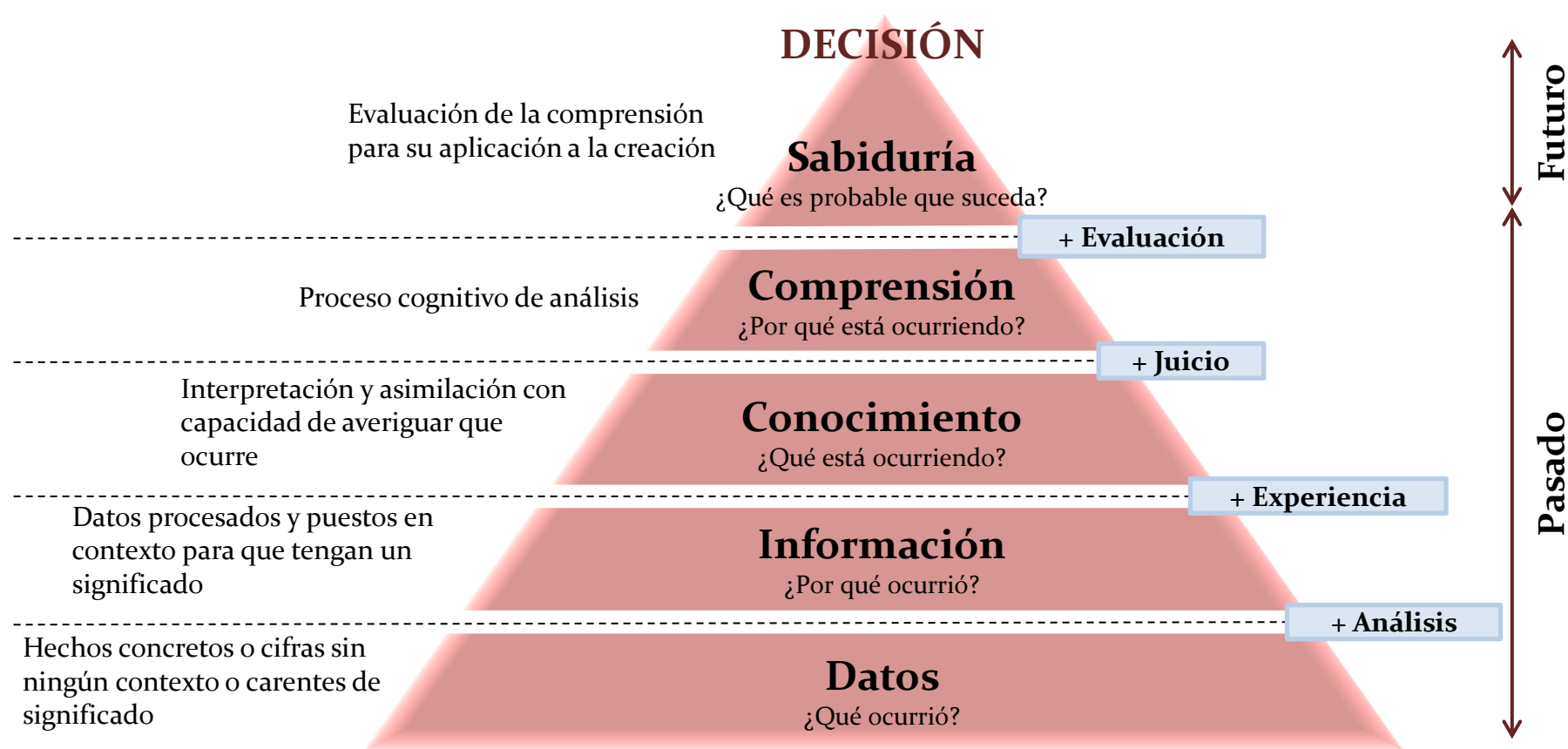
Según el tratamiento al que han sido sometidos:

- Datos en bruto (*raw data*)
- Datos procesados

Según la estructura que presentan:

- Datos estructurados
- Datos no estructurados
- Datos semiestructurados

# Pirámide del conocimiento (DIKW) (Fayyad, Shapiro, 1996)



# M1 – La Gestión de los Datos

## ***Ciclo de Vida de los Datos***

“Lapso de tiempo en el que los datos existen desde el momento de su creación, hasta su transformación: evolución, modificación, reutilización o destrucción”.





***ClickStream01:  
Instalación de MySQL y  
complemento Excel***

# M1 – La Gestión de los Datos

## ***Introducción***

El módulo 1 se centra en el análisis de los tres puntos siguientes:

- La gestión de los datos.
- **El concepto de Big Data.**
- El rol que el Big Data desempeña en las organizaciones.

# M1 – La Gestión de los Datos

## ***Objetivos y resultados de aprendizaje***

- **Objetivo**
  - Introducir el concepto de Big Data.
- **Resultados del aprendizaje**
  - Saber qué es Big Data.
  - Reconocer los retos y oportunidades que plantea Big Data.
  - Identificar cuáles son los componentes típicos de un sistema de Big Data.
    - Conocer la existencia de diferentes arquitecturas.
  - Conocer la interrelación existente entre Big Data y Cloud Computing.

***Actividad 1***  
***Del bit... al Big Data***

# M1 – Introducción a Big Data

## ***Introducción a Big Data***

- La sociedad hiperconectada genera datos en un volumen, velocidad y variedad que impiden su tratamiento mediante las tecnologías tradicionales.
- La necesidad de gestionar mayores datos de estas características, ha dado lugar al desarrollo de nuevas tecnologías, que siguen evolucionando, y que son capaces de hacerlo a unos costes razonables.

# M1 – Introducción a Big Data

## ***Introducción a Big Data***

**Big Data** representa gran volumen, alta velocidad y gran variedad de datos que demandan formas innovadoras de procesamiento para tener un mejor conocimiento, toma de decisiones y automatización de procesos a través de analítica avanzada.



# M1 – Introducción a Big Data

## *¿Qué es Big Data para los gerentes?* [IBM,BBVA]



**18%** Un mayor abanico de información.

**16%** Nuevos tipos de datos y análisis.

**15%** Información en tiempo real.

**13%** Flujos de datos de nuevas tecnologías en medios no tradicionales.

# M1 – Introducción a Big Data

**¿Qué es Big Data?** Hay que identificar Big Data con el tipo de información que pueden aportar los datos, basado en las **cuatro preguntas siguientes**:



- La cuarta es prescriptiva: **¿qué puedo hacer para que eso ocurra?**



# M1 – Introducción a Big Data

## *Introducción a Big Data*

- **Big Data** tiene como objetivo construir **mejores modelos**, que puedan dar recomendaciones o soluciones de **mayor precisión**, que buscan hacer del mundo un lugar mejor.



<https://larazon.co/apuntes/futurecraft-4d-los-primeros-zapatos-3d-adidas/>

# Evolución de las librerías





## Evolución de las librerías



# Evolución de las librerías







# Evolución de las librerías





## Evolución de las librerías

# M1 – Introducción a Big Data

<p><b>NETFLIX</b></p> <p><b>NO MATÓ A BLOCKBUSTER</b></p> <p>Los ridículos cargos por pagos atrasados lo hicieron</p>	<p>U B E R</p> <p><b>NO MATÓ A LOS TAXIS</b></p> <p>El acceso limitado a taxis, el mal servicio y el control de tarifas lo hicieron</p>
<p>Apple</p> <p><b>NO MATÓ A LA INDUSTRIA DE LA MÚSICA</b></p> <p>El obligar a comprar álbumes completos lo hizo</p>	<p>amazon</p> <p><b>NO MATÓ A OTROS MINORISTAS</b></p> <p>El mal servicio y experiencia del cliente lo hicieron</p>
<p>airbnb</p> <p><b>NO ESTÁ MATANDO A LA INDUSTRIA HOTELERA</b></p> <p>La disponibilidad limitada y las opciones de precios lo hicieron</p>	<p>La tecnología por sí misma no es el verdadero disruptor</p> <p><b>NO CENTRARSE EN EL CLIENTE ES LA MAYOR AMENAZA PARA CUALQUIER NEGOCIO</b></p> <p><i>#Viral</i></p>

# M1 – Introducción a Big Data

## *¿Qué hace valioso a Big Data?*

- Mercadotecnia personalizada
- Motores de recomendaciones
- Análisis de sentimientos
- Publicidad móvil
- Aplicaciones biomédicas
- Personalizar tratamientos médicos (cáncer)
- Smart Cities (Medellín, Colombia y Curitiba, Brasil)



"Your recent Amazon purchases, Tweet score and location history makes you 23.5% welcome here."



# M1 – Introducción a Big Data

***Industrias que usan Big Data para transformar su modelo de negocio:***

<b>Retail</b>	<b>Fabricación</b>
CRM (Customer Relationship Management) Diseño y ubicación de tiendas Detección y prevención de fraude Optimización de la cadena de suministro	Investigación de producto Analítica de ingeniería Mantenimiento predictivo Análisis y calidad de procesos Optimización de la distribución
<b>Servicios Financieros</b>	<b>Telecomunicaciones y Medios</b>
Algoritmia para comercio Análisis de riesgos Detección de fraude Análisis de cartera	Optimización de redes Scoring de clientes Prevención de rotación de clientes Prevención de fraude
<b>Publicidad</b>	<b>Energía</b>
Publicidad dirigida Señalización de la demanda Análisis de sentimientos Adquisición de usuarios	Redes inteligentes Procesos de exploración Modelización operativa Sensores
<b>Administraciones</b>	<b>Salud y ciencias de la vida</b>
Gestión de mercados Sistemas armamentistas y antiterrorismo Econometría Informática para la salud	Farmacogenómica Bioinformática Investigación farmacéutica Investigación de resultados clínicos

# M1 – Bibliografía

- ✓ Schroeck, M., Analytics: el Uso de Big Data en el Mundo Real. Consultado el 8 de diciembre de 2015. IBM Institute for Business Value, 2012.
- ✓ BBVA Innovation Edge, Big Data: Es Hora de Generar Valor de Negocio con los Datos. Consultado el 27 de junio de 2013.
- ✓ McGahan, A., Unlocking the Big Promise of Big Data, En: Rotman Magazine, 2013.
- ✓ Dasu T., Johnson T., Exploratory Data Mining and Data Cleaning, Wiley, 2003.
- ✓ Fayyad, U., Shapiro, G., From Data Mining to Knowledge Discovery in Databases, En: AI Magazine 17.3 (1996), págs. 37-54.

Imágenes, todas las páginas consultadas en línea el 20 de enero de 2021.

- ✓ <http://www.computing.es/mercado-ti/informes/1109855046401/gasto-global-crecera-32-ciento-2019.1.html>
- ✓ <http://terabitio.com/internet/no-chrome-firefox-5-navegadores-alternos/>
- ✓ <https://www.ebay.com/>
- ✓ [https://www.freepik.es/vector-gratis/coleccion-logotipo-circular-redes-sociales-populares\\_1843265.htm](https://www.freepik.es/vector-gratis/coleccion-logotipo-circular-redes-sociales-populares_1843265.htm)
- ✓ <https://prometeusgs.com/preparado-internet-las-cosas-impacto/>