

Analítica Empresarial para la Toma de Decisiones Efectivas. ¿Cómo usar Big Data?

Rolando de la Cruz

Director Académico Magíster en Data Science

rolando.delacruz@uai.cl

Rolando de la Cruz Mesía

- ▶ Director Académico del Magíster en Data Science de la Facultad de Ingeniería y Ciencias de la Universidad Adolfo Ibáñez e Investigador del UAI Smart Center. Previamente fue Profesor Adjunto, Jefe de Investigación y Jefe de Extensión del Instituto de Estadística de la Pontificia Universidad Católica de Valparaíso, Chile (2015-2018), Profesor Asistente (2007-2010) y Profesor Asociado (2010-2015) de los Departamentos de Salud Pública y Estadística de la Pontificia Universidad Católica de Chile. El prof. De la Cruz obtuvo su Doctorado en Estadística (con máxima distinción) en la Pontificia Universidad Católica de Chile (2005). Consultor, investigador y docente con más de 15 años de experiencia en data science, big data, machine learning con aplicaciones en diversas áreas tales como: logística portuaria, banca, retail, retail financiero, biomedicina, farmacología, marketing, gestión, educación, seguridad, transporte.



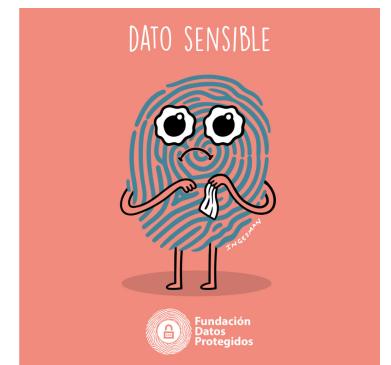
Agenda

- ✓ Introducción
- ✓ Industria 4.0
- ✓ ¿Qué es Data Science? ¿Qué es Big Data?
- ✓ Casos de Uso
- ✓ Oportunidades y Desafíos
- ✓ Actividad

✓ Introducción

¿Qué es un dato?

Hecho individual acerca de algo de interés para alguien.



¿Qué es información?

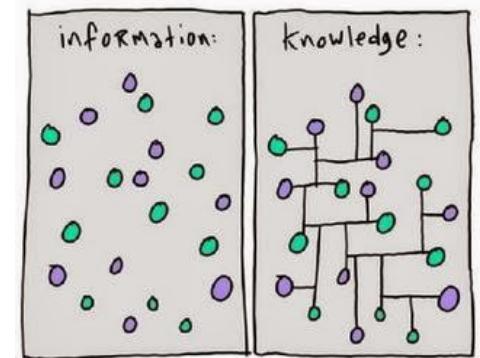
Datos relacionados.

Información es un conjunto organizado de datos procesados.



¿Qué es conocimiento?

Información co-relacionada.
Patrones!



Tipos de Conocimiento

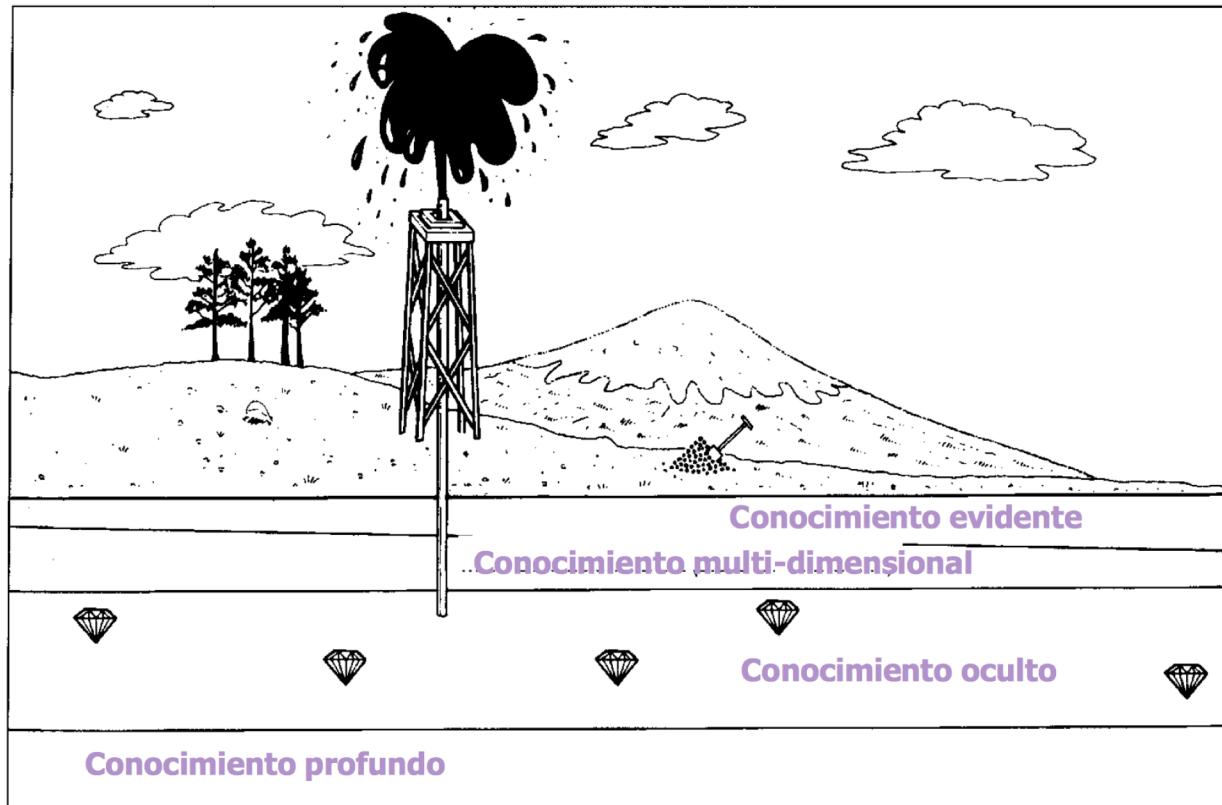


Imagen original: P. Adriaans, D. Zantinge. Addison-Wesley: "Data Mining", 1996.

Generación de Datos

Comercial

- Web
- Supermercados
- Bancos
- Aseguradoras

Científico

- Satélites
- Telescopios
- Microarrays
- Simulación
- Publicaciones

Datos

Almacenados

- o Web
- o Archivos
- o Base de datos

Regulating the internet giants

The world's most valuable resource is no longer oil, but data

The data economy demands a new approach to antitrust rules

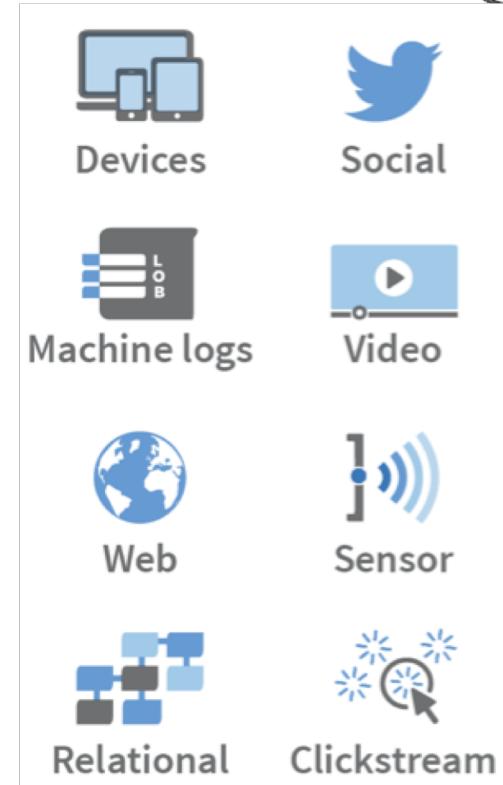


Información

Algo peor que no tener información disponible es tener mucha información.

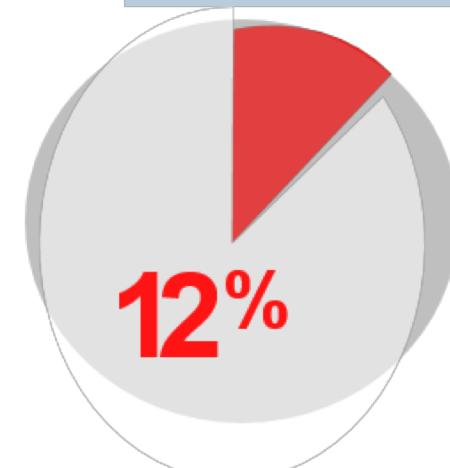
Y no saber qué hacer con ella.

Datos, datos y más datos....

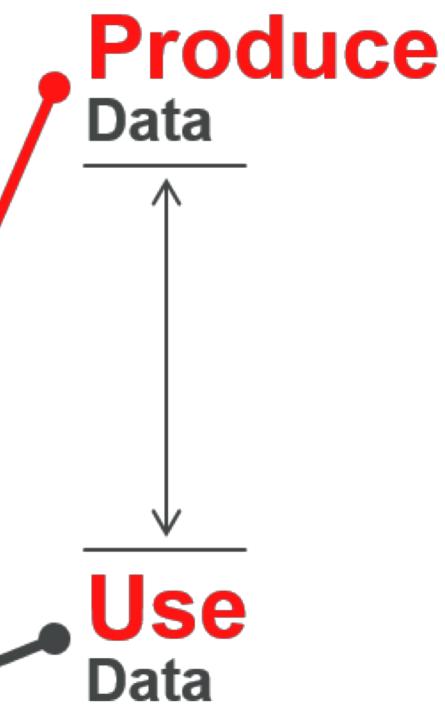


¿Qué hacemos? ¿Seguimos nuestra intuición?

The Useful Data GAP



Executives who feel they understand the impact data will have on their organizations



Fuente: Delphi Analytics

En el día a día estamos enfrentados a:

- . Tomar decisiones.
- . Solucionar problemas.

Ayuda: Tomar decisiones basadas en datos.



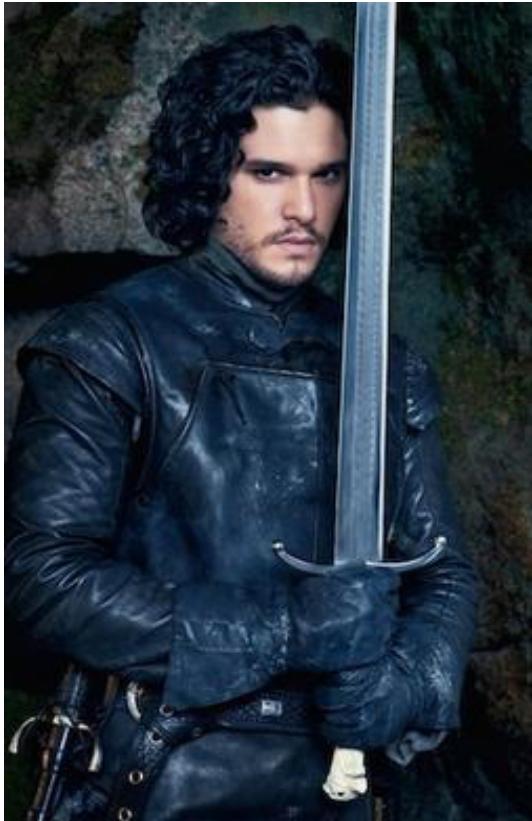
Epidemia de cólera: inicios del big data

- Londres 1854. Barrio Soho.
- Ataque epidemia de cólera.
Aprox. 700 muertes.



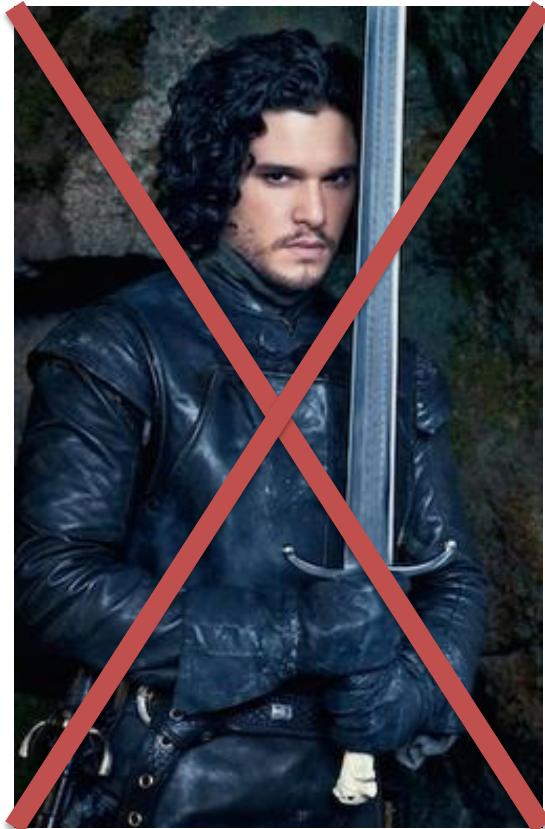
A COURT FOR KING CHOLERA.

Jon Snow



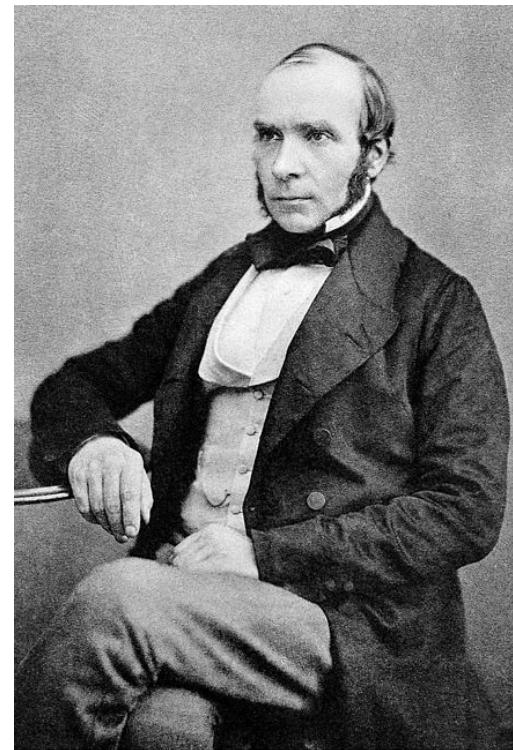
Fuente: Wikipedia

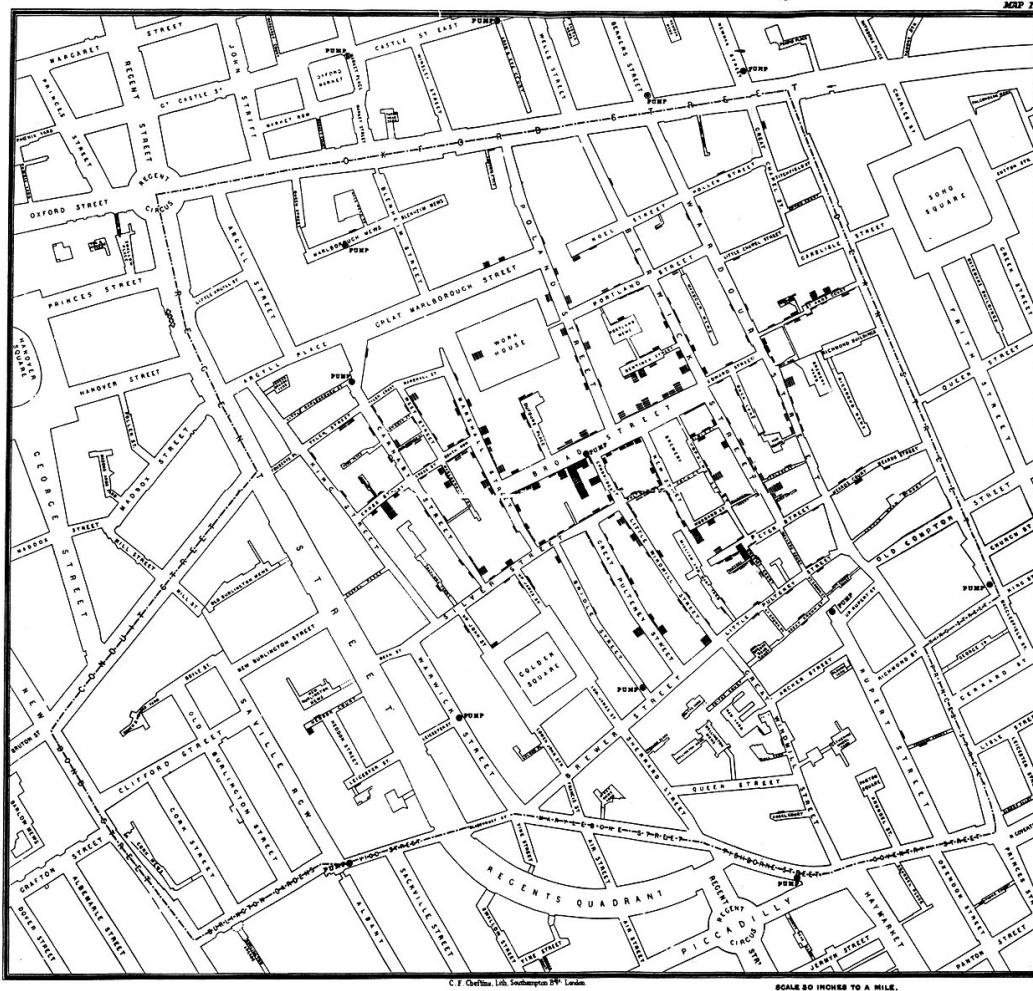
Jon Snow



Fuente: Wikipedia

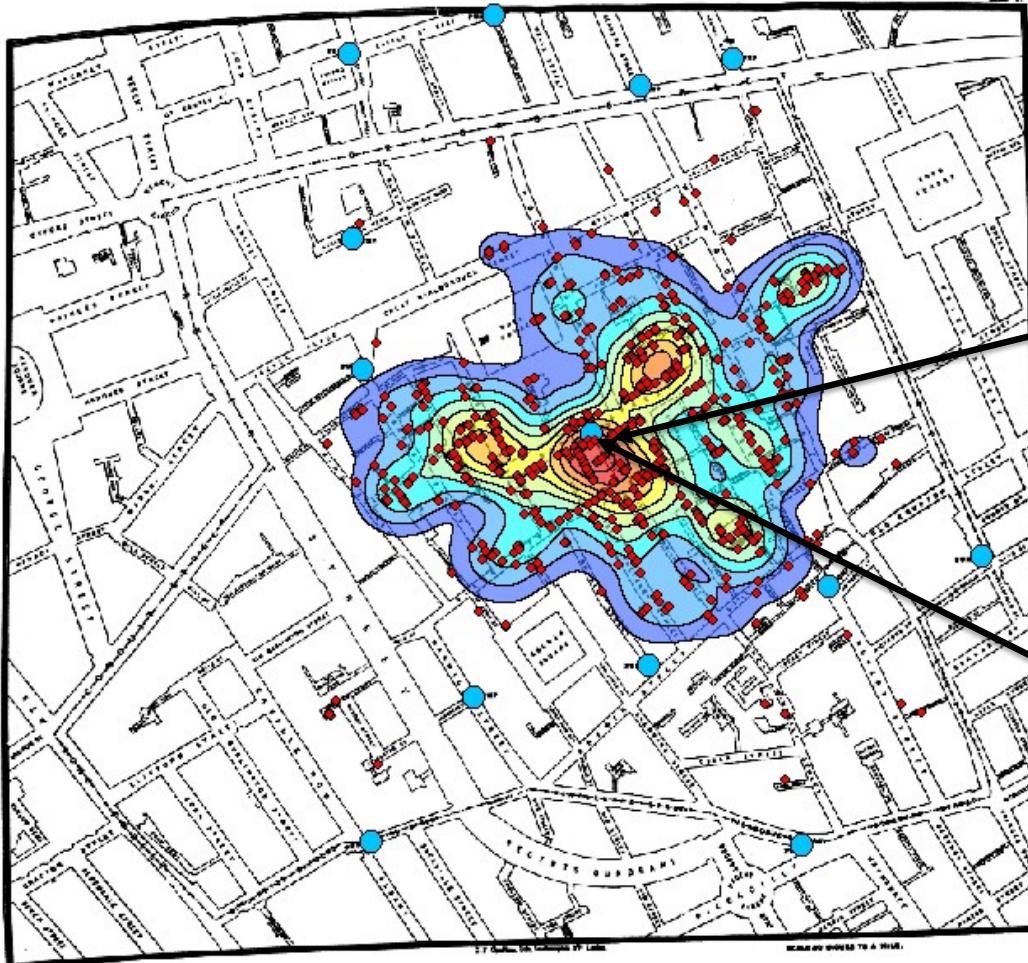
Jhon Snow (1813 – 1858)





Distribución de las primeras 83 muertes.





Más del 70% de las muertes por cólera durante el brote de 1854 se produjeron dentro de esta área de influencia.



Bomba ubicada en Broad Street

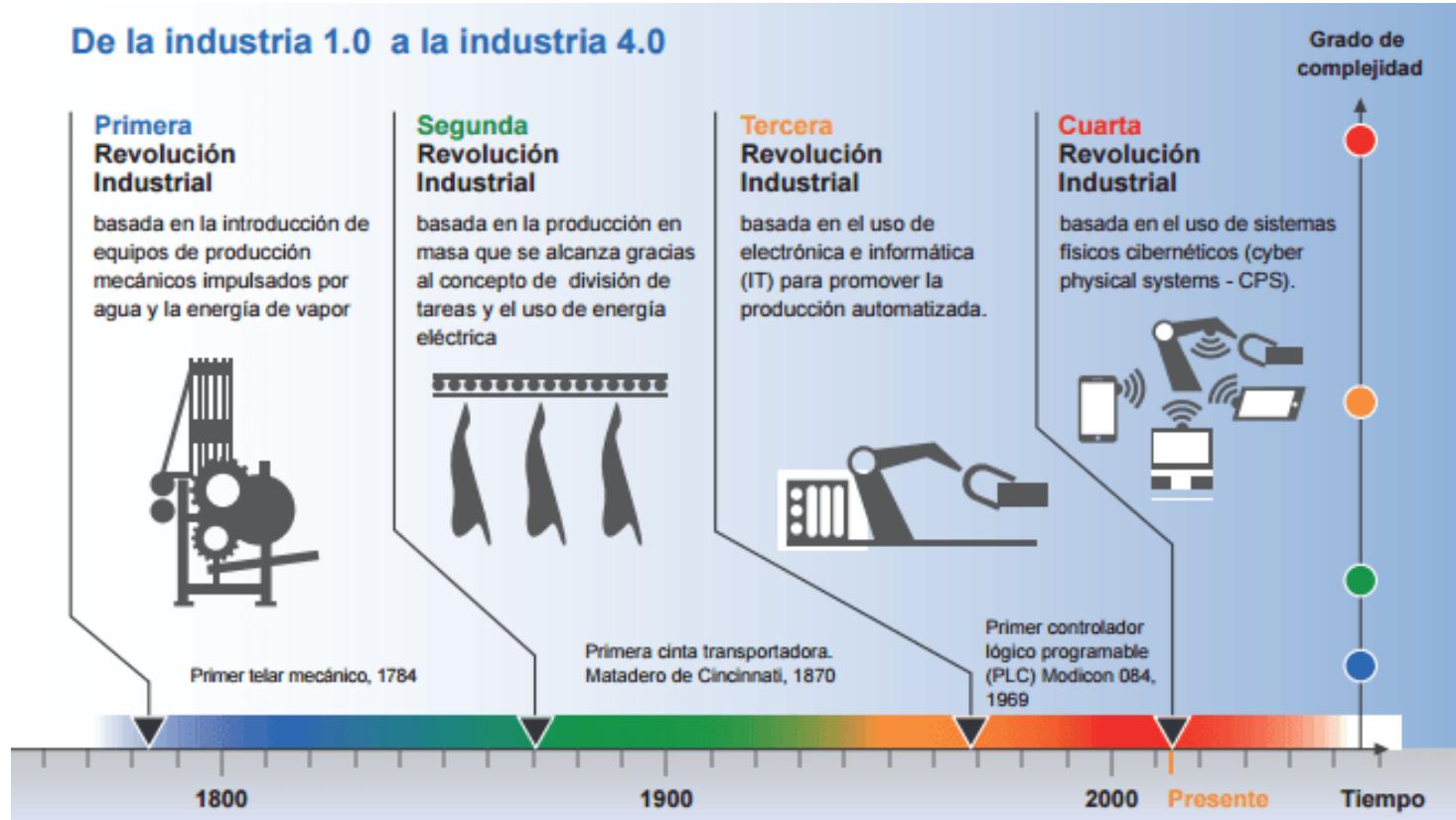
Fuente: <https://medium.com/public-health/john-snow-early-big-data-science-d62b4dacd71b>

A partir de los datos:

- ✓ Demostró que el cólera era causado por el consumo de aguas contaminadas con materias fecales, al comprobar que los casos de esta enfermedad se agrupaban en las zonas donde el agua consumida estaba contaminada con heces.
- ✓ Aportes: Mejoras en la salud pública, entre otros.

✓ La Industria 4.0

De la industria 1.0 a la industria 4.0



<https://papelesdeinteligencia.com/que-es-industria-4-0/>

Industria 4.0

Plantea una revolución entre la automatización de los procesos de producción y la explosión de la era digital que se extiende a todas las industrias.



Fuente: AMETIC

Industria 4.0

- ▶ El reto no estará en conseguir integrar todo esto y que los sistemas o máquinas hablen entre sí para que todo funcione como un reloj suizo.

- ▶ El verdadero reto estará una vez más en las personas, en como liderar el proceso de transformación digital dentro de la organización y en el cambio que supondrá adaptarse y trabajar en los nuevos entornos conectados de la industria 4.0.

Industria 4.0: Organizaciones basadas en datos

- ▶ Tener el dato como epicentro de las acciones.

Las compañías que utilizan los datos como elemento clave en la toma de decisiones son más eficientes y productivas.

Tres pilares de una organización basada en datos:



Industria 4.0: Organizaciones basadas en datos

- ▶ **Tecnología:** abierta, flexible, híbrida, escalable y orientada a la movilidad.
- ▶ **Gobierno de datos:** reglas, normas, políticas que aseguren la integridad de los datos.
- ▶ **Organización:** decisiones basadas en los datos en la que organización participa activamente y es conscientemente de la importancia de los mismos.

Todo lo anterior lleva a un cambio cultural en la organización.

Industria 4.0: Organizaciones basadas en datos

- ▶ Científico de Datos.
- ▶ Ingeniero de Datos.
- ▶ Arquitecto de Datos.
- ▶ Responsable de Protección de Datos.
- ▶ Responsable de Visualización de Datos, etc.

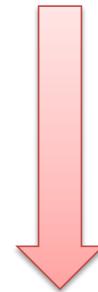
✓ ¿Qué es Data Science? ¿Qué es Big Data?

Minería de Datos

- ▶ Grandes bases de datos contienen información no plenamente explotada.
 - Información oculta
 - Información valiosa: “Conocimiento”
- ▶ Esta información puede ser encontrada entre los datos haciendo uso de **minería en los datos**
 - ✓ Comercial: competencia, servicios con el cliente.
 - ✓ Científico: clasificar y segmentar los datos.

KDD: Descubrimiento de Conocimiento en Bases de Datos

Los datos son la materia prima bruta

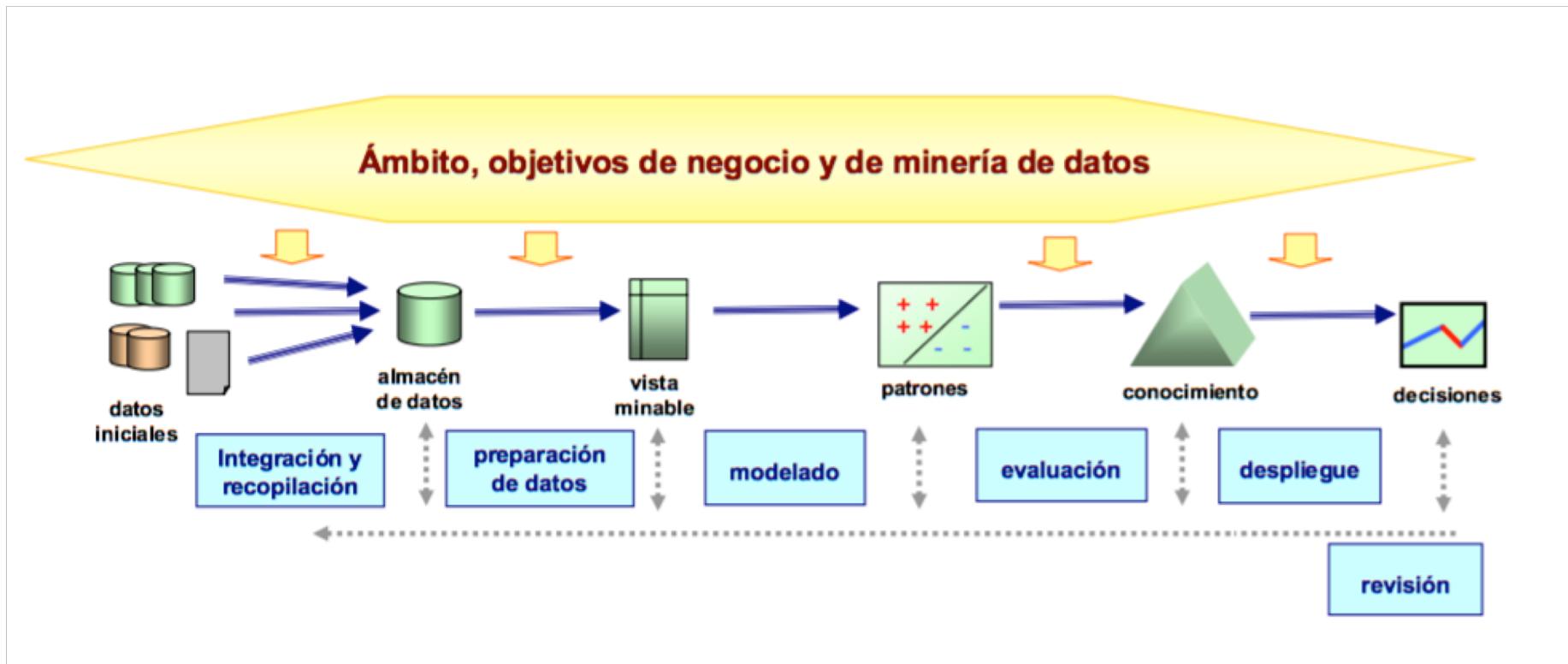


INFORMACIÓN

KDD: Descubrimiento de Conocimiento en Bases de Datos

El KDD es el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos, teniendo como objetivo encontrar conocimiento útil relevante y nuevo sobre un fenómeno o actividad, presentando los resultados de manera visual.

Procesos KDD



¿Qué es Data Science?

- Es un campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, ya sea estructurados o no estructurados.

A. Liu. Data Science and Data Scientist

- Es la ciencia que estudia la extracción de conocimiento a partir de los datos.



SPOTLIGHT ON BIG DATA

Spotlight

ARTWORK Tamar Cohen, *Andrew J Buboltz*
2011, silk screen on a page from a high school
yearbook, 8.5" x 12"

Data Scientist: *The Sexiest Job of the 21st Century*

**Meet the people who
can coax treasure out of
messy, unstructured data.**
by Thomas H. Davenport
and D.J. Patil

70 Harvard Business Review October 2012



Data Scientists: The New Rock Stars of the Tech World

By Troy Sadkowsky, July 2, 2014

Peter Naur (1974). *Concise Survey of Computer Methods.*
Petrocelli Books.



COMMITTED TO
IMPROVING THE STATE
OF THE WORLD

The Jobs Landscape in 2022

emerging
roles,
global
change
by 2022

133
Million

declining
roles,
global
change
by 2022

75
Million

Top 10 Emerging

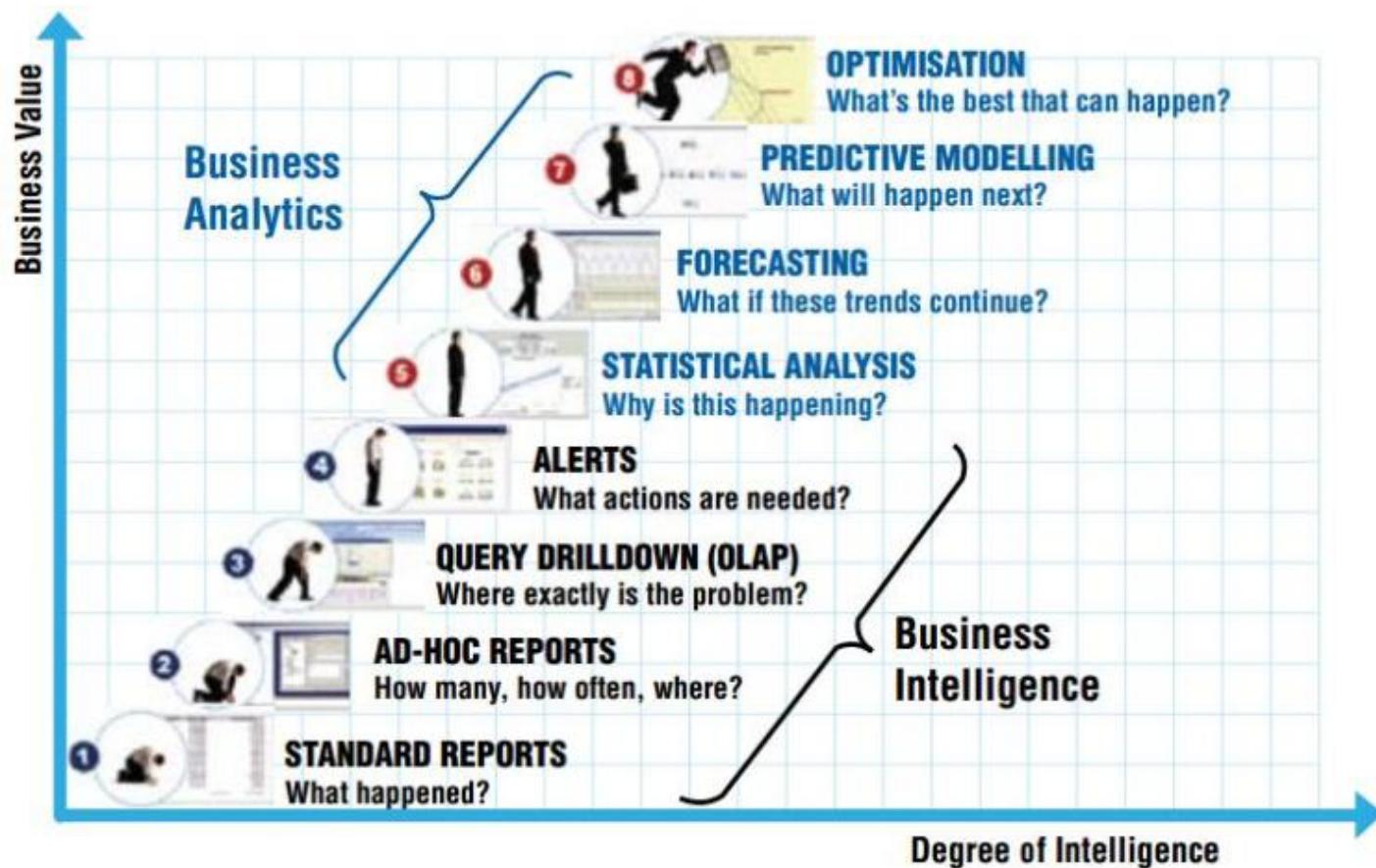
1. Data Analysts and Scientists
2. AI and Machine Learning Specialists
3. General and Operations Managers
4. Software and Applications Developers and Analysts
5. Sales and Marketing Professionals
6. Big Data Specialists
7. Digital Transformation Specialists
8. New Technology Specialists
9. Organisational Development Specialists
10. Information Technology Services

Top 10 Declining

1. Data Entry Clerks
2. Accounting, Bookkeeping and Payroll Clerks
3. Administrative and Executive Secretaries
4. Assembly and Factory Workers
5. Client Information and Customer Service Workers
6. Business Services and Administration Managers
7. Accountants and Auditors
8. Material-Recording and Stock-Keeping Clerks
9. General and Operations Managers
10. Postal Service Clerks

Source: Future of Jobs Report 2018, World Economic Forum

INGENIERÍA

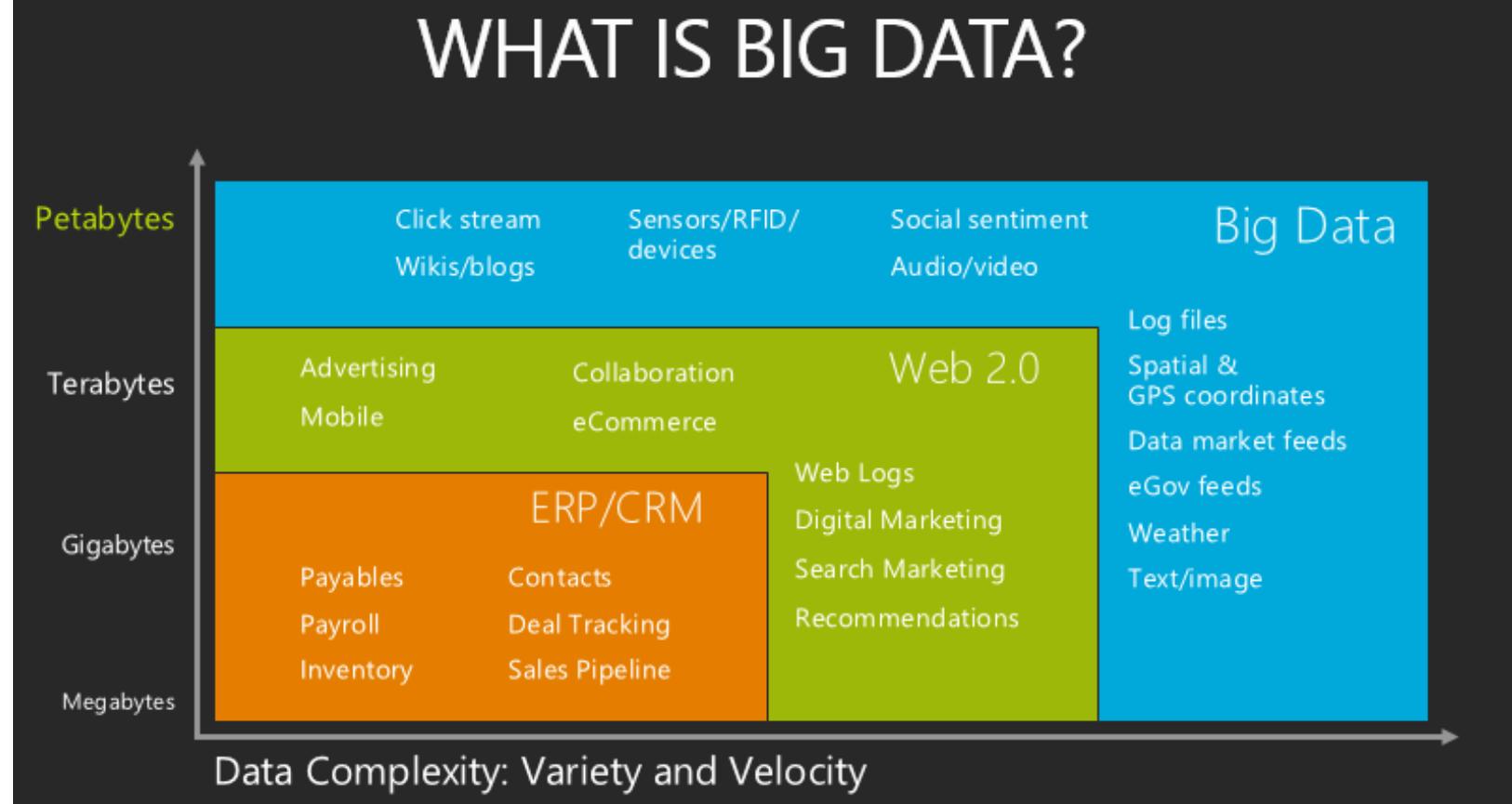


Analítica

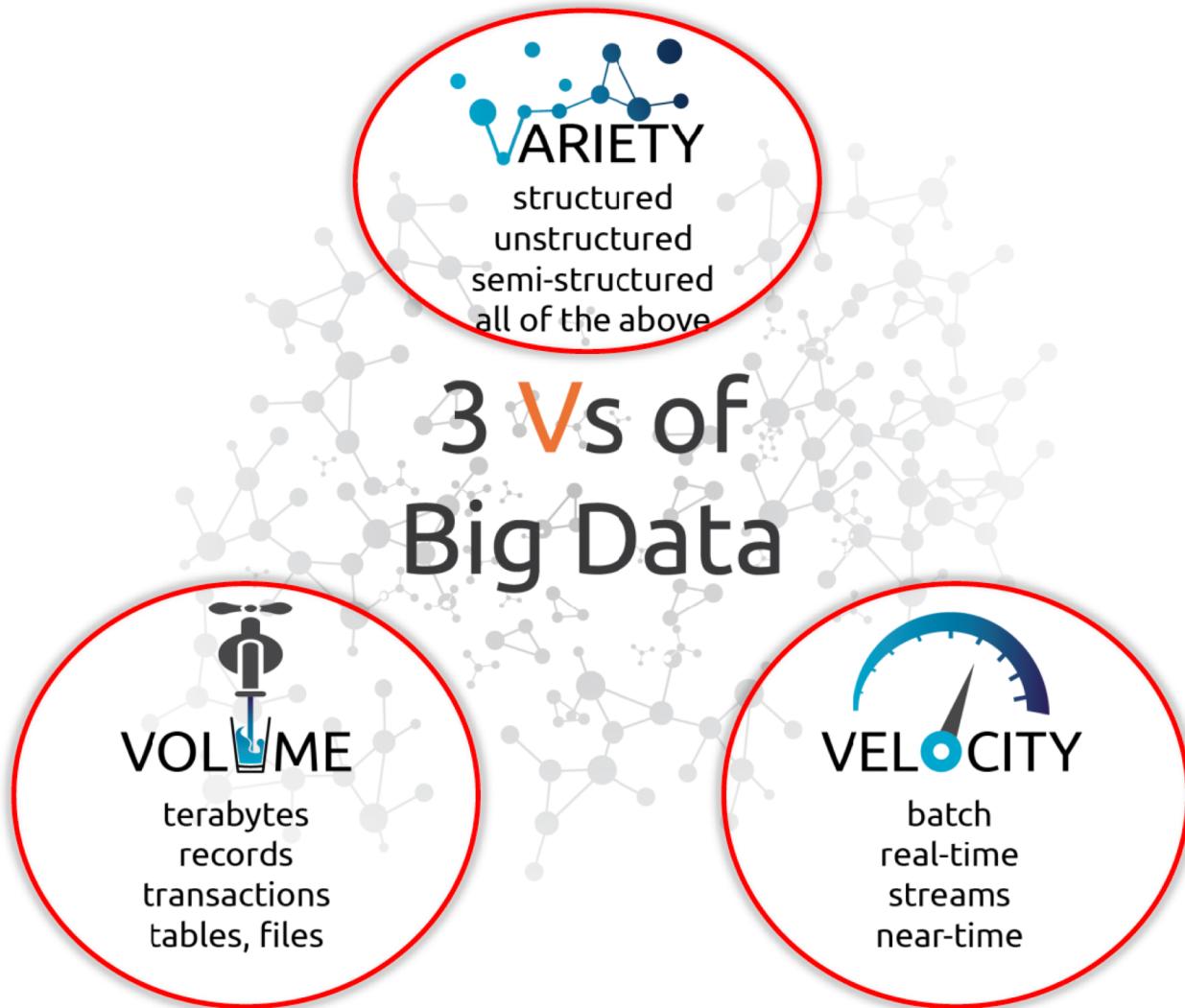
- ▶ La **analítica descriptiva** proporciona información sobre las relaciones entre los datos y sus características. Genera información del tipo:
 - El tabaco y el alcohol son los factores más importantes en la enfermedad X
 - Los clientes sin televisión y con bicicleta tienen características muy diferentes del resto.
- ▶ La **analítica predictiva** responde preguntas sobre datos futuros.
 - ¿Cuáles serán las ventas el año próximo?
 - ¿Es esta transacción fraudulenta?
 - ¿Qué tipo de seguro es más probable que contrate el cliente X?

Analítica

- ▶ La **analítica prescriptiva** tiene en cuenta no solo los datos del negocio, sino cómo impactan las decisiones en las cuentas de costes y beneficios, y qué restricciones y consideraciones deben contemplarse en las acciones que se van a realizar. Esto permite generar automáticamente políticas de acción realistas que tienen un impacto directo en el ROI del negocio.



<https://bitsnbytesofbigdata.wordpress.com/2015/05/21/big-data-whats-the-big-deal/>



Y siguen sumando V's:

- Veracidad
- Valor



Cada 60 segundos

<http://onlinemarketingservicesllc.com/wp-content/uploads/2013/07/20130725-11414769-Qmee-Online-In-60-Seconds2.png>

La inundación DE DATOS en los vehículos autónomos

Radar
~10-100 kB
Por segundo

SonAr
~10-100 kB
Por segundo

GPS
~50kB
Por segundo

Cámaras
~20-40 mb
Por segundo

Vehículos autónomos
4000 GB
Por día... cada día

Lidar
~10-70 mB
Por segundo



Fuentes de Datos

Visión general de las fuentes de datos



Datos no estructurados o semi-estructurados



Instagram

Redes sociales



Sensorización



Información médica

Análisis en el mundo Big Data

- ▶ Lo relevante no es tener la capacidad para recolectar y almacenar una gran cantidad de datos. El síndrome de Diógenes de los datos!!!
- ▶ Con la acumulación de los datos no se alcanzan ventajas competitivas: es necesario saber organizarlos, refinarlos, y convertirlos en información relevante que permita ganar posiciones en el mercado.
- ▶ Los datos tienen sólo valor potencial, es su análisis y sistematización el que permite incrementar la capacidad de innovar y obtener ventajas en las organizaciones.

Tecnología en el mundo Big Data

- ▶ Sea escalable de forma masiva a petabytesde datos.
- ▶ Soporte y acceso a datos de baja latencia y toma de decisiones.
- ▶ Tenga análisis integrado para acelerar el modelado de análisis avanzado y de los procesos.

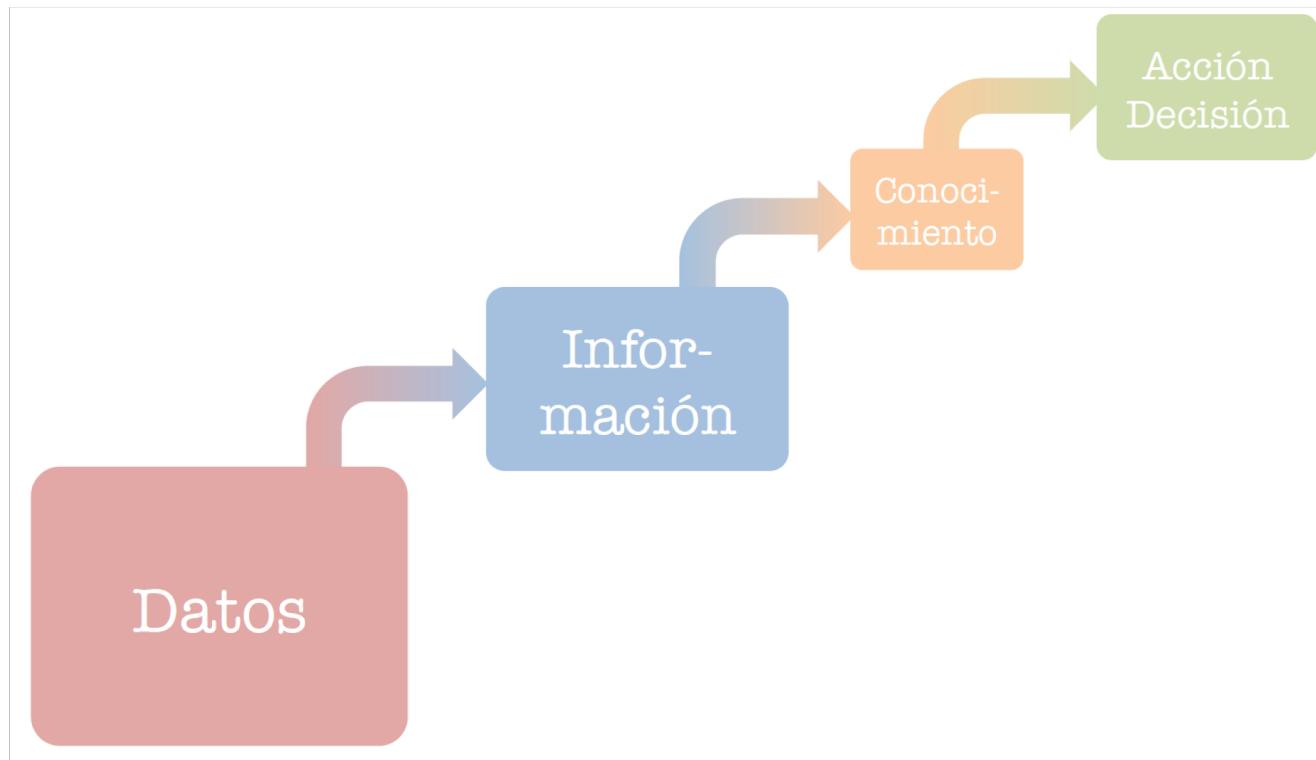
Valor

- Creando transparencia.
- Reducir ineficiencias.
- Permitiendo la experimentación para descubrir necesidades, exponer variabilidad y mejorar el rendimiento de procesos.
- Reemplazar o ayudar decisiones humanas por algoritmos automatizados.
- Crear nuevos modelos de negocio, productos o servicios.

McKinsey Global Institute Big Data Report 2011

Valor

Datos ≠ Información ≠ Valor

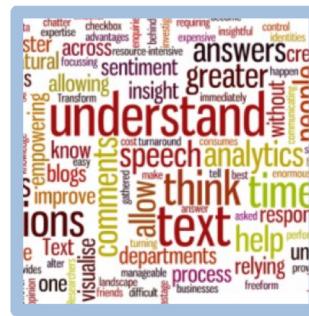


Valor

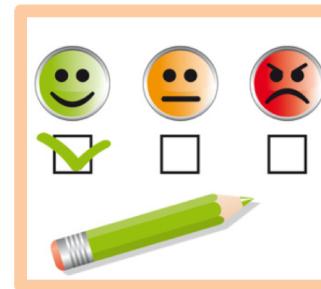
- Datos ≠ Información ≠ Valor



Tweets sobre marca/evento/tema



Análisis lingüístico del contenido



Clasificación del contenido. Generación de alertas



Interacción con usuarios



✓ Usos



<https://www.iebschool.com/blog/big-data-salud-big-data/>



Framingham localidad ubicada cerca de Boston, USA.

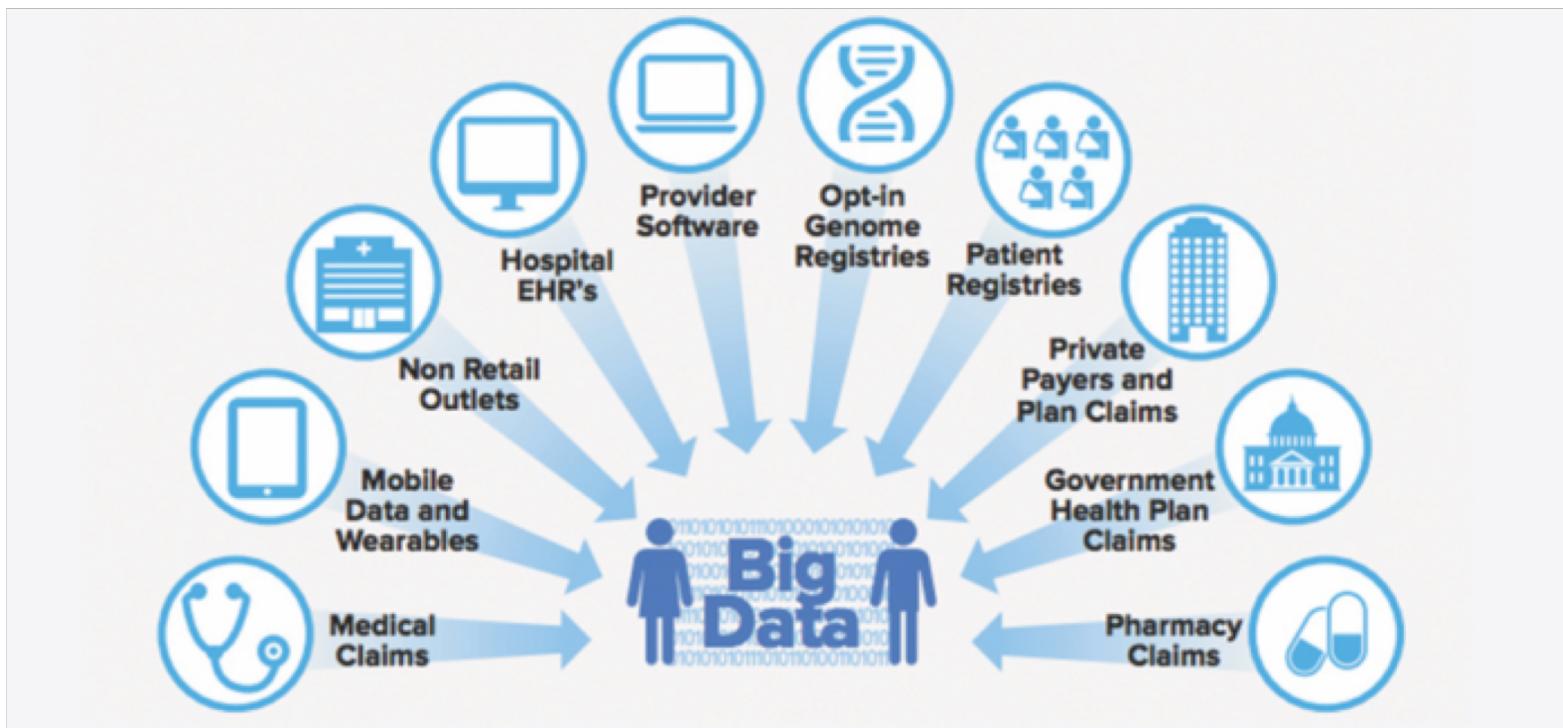
Cohorte Framingham

Las causas de las enfermedades cardiacas y accidentes cerebrovasculares no eran conocidas

Factores de riesgo:

Presión arterial alta, niveles altos de colesterol, fumar, obesidad, diabetes, inactividad física...

Big Data en Salud



Fuente: <https://www.medgadget.com/2018/04/big-data-in-healthcare-market-value-share-of-20-69-with-cerner-co-cognizant-dell-philips-siemens-and-business-forecast-to-2022.html>

Big Data en Salud

Medicina de las 4P:

- Predictiva
- Preventiva
- Personalizada
- Participativa

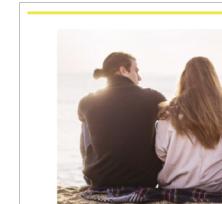
google.org ?

Flu Trends



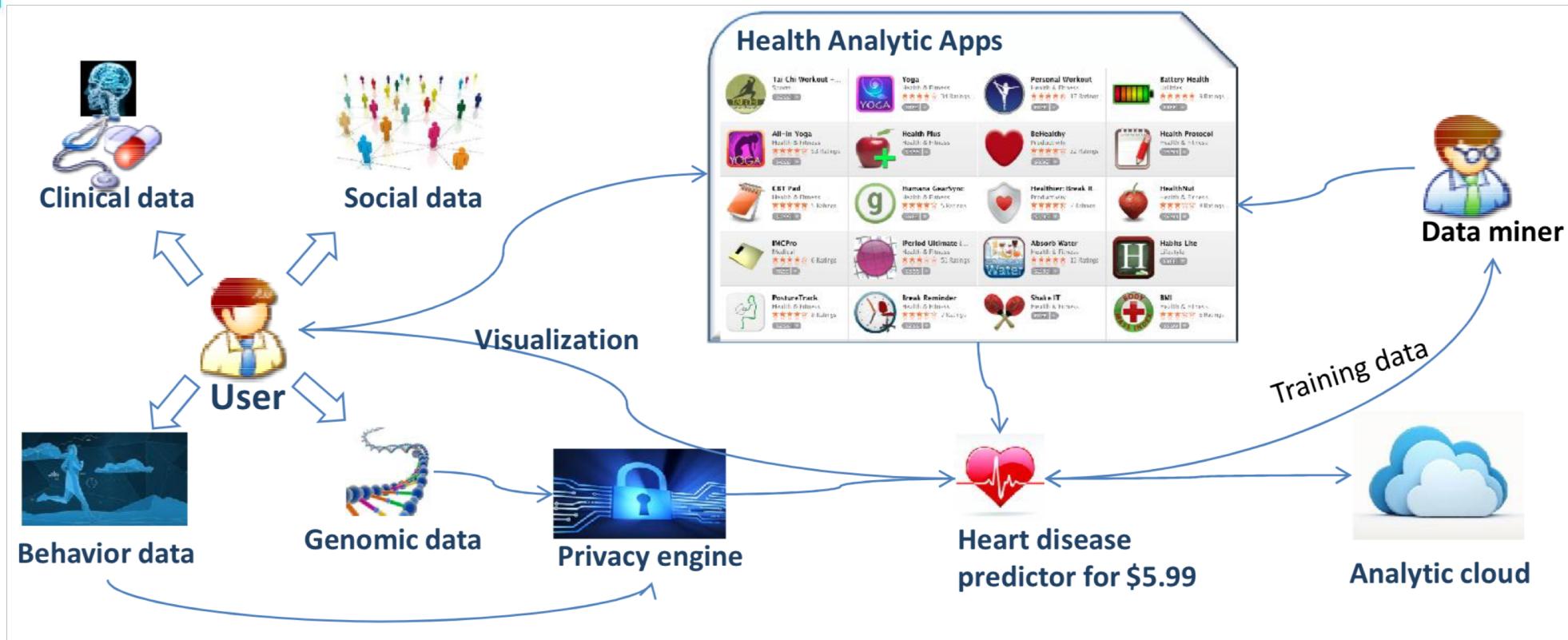
IBM Watson Health

- tecnología de computación cognitiva



MEDICINA PARTICIPATIVA
Llegan los e-pacientes. ¿Eres uno de ellos?

El futuro...



Fuente: J. Sun & C. Reddy. Big data analytics for healthcare.

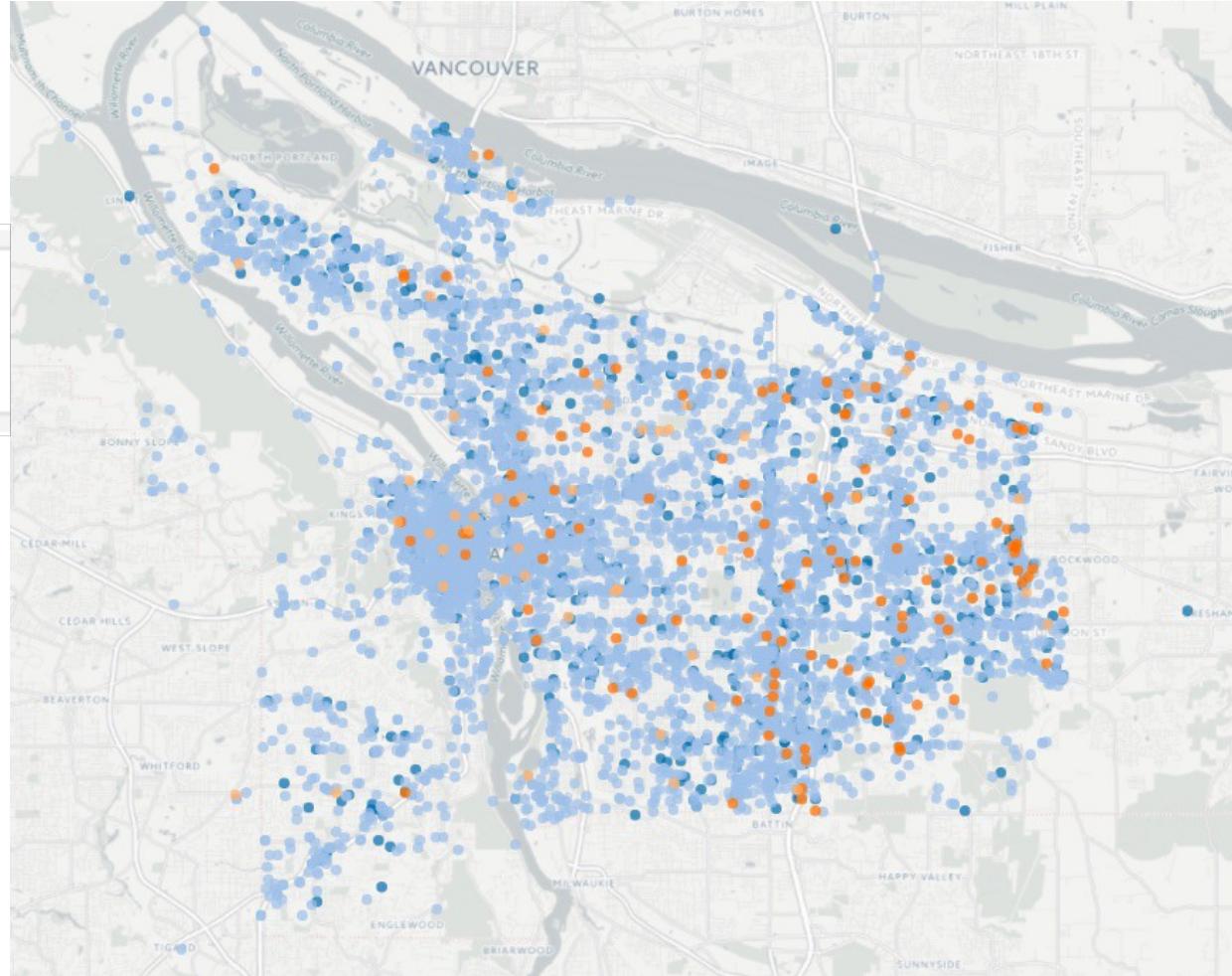


BIG DATA

& Seguridad Ciudadana

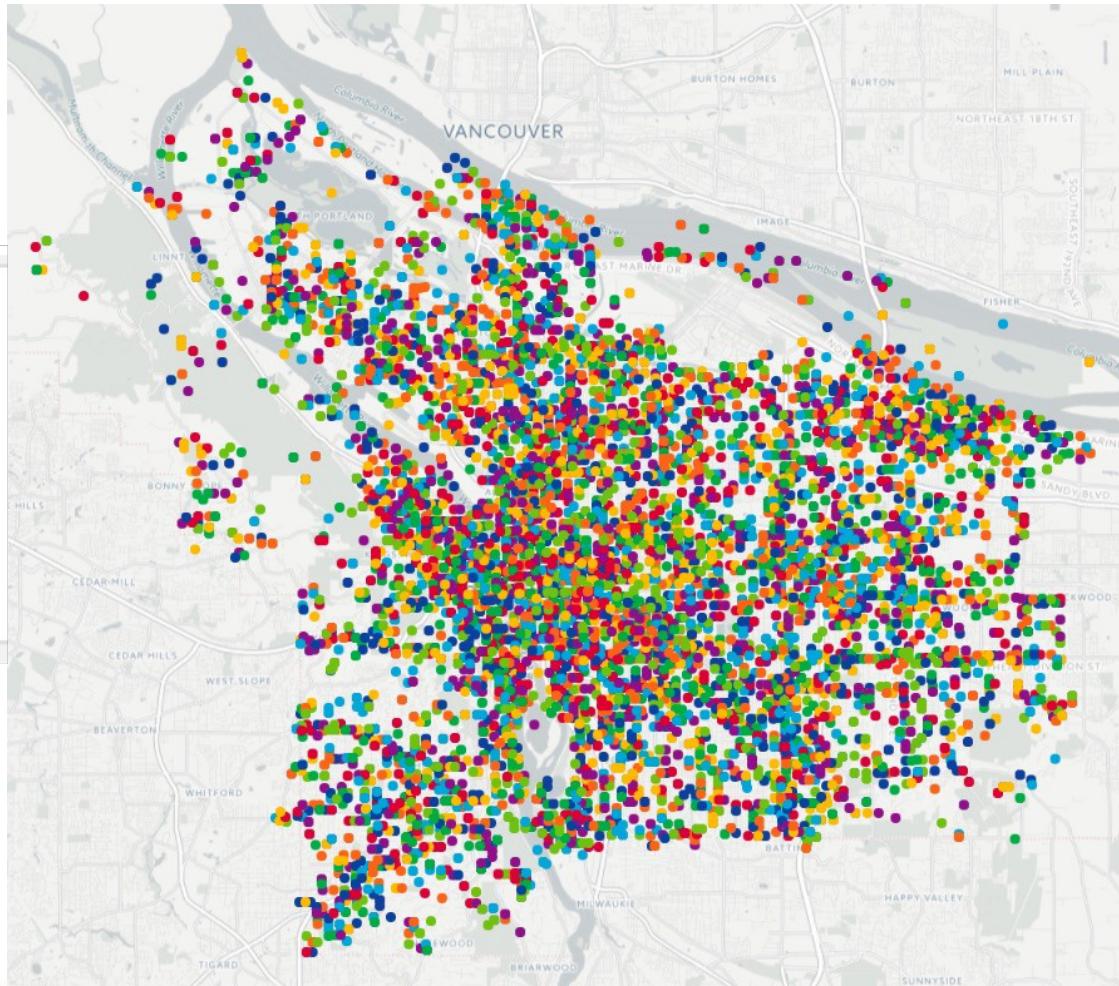
Delitos

- STREET CRIMES
- OTHER
- MOTOR VEHICLE THEFT
- BURGLARY



Delitos

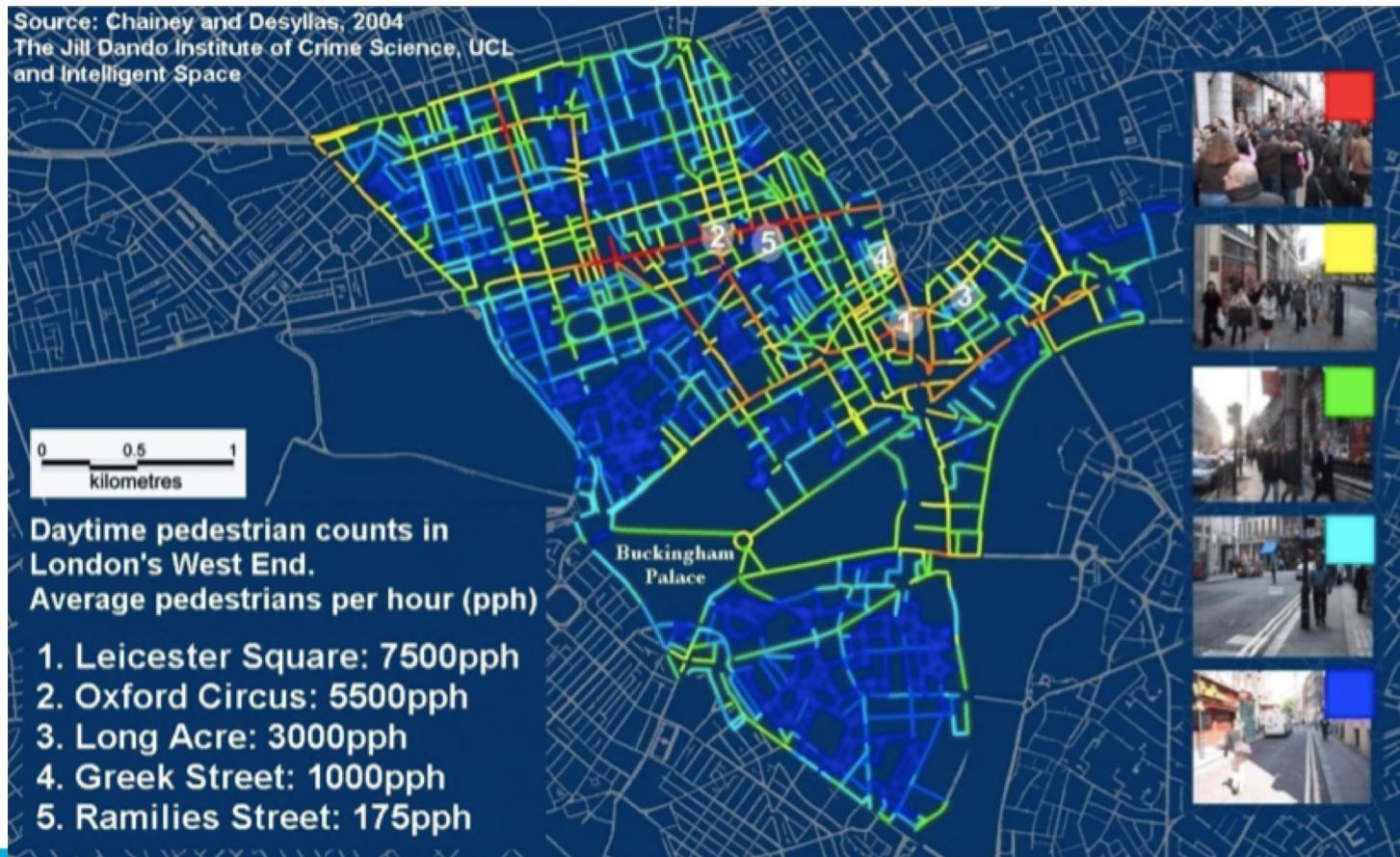
- Event
- NightlifeSpot
- CollegeUniversity
- ShopService
- OutdoorsRecreation
- ArtsEntertainment
- Residence
- food_foursquare



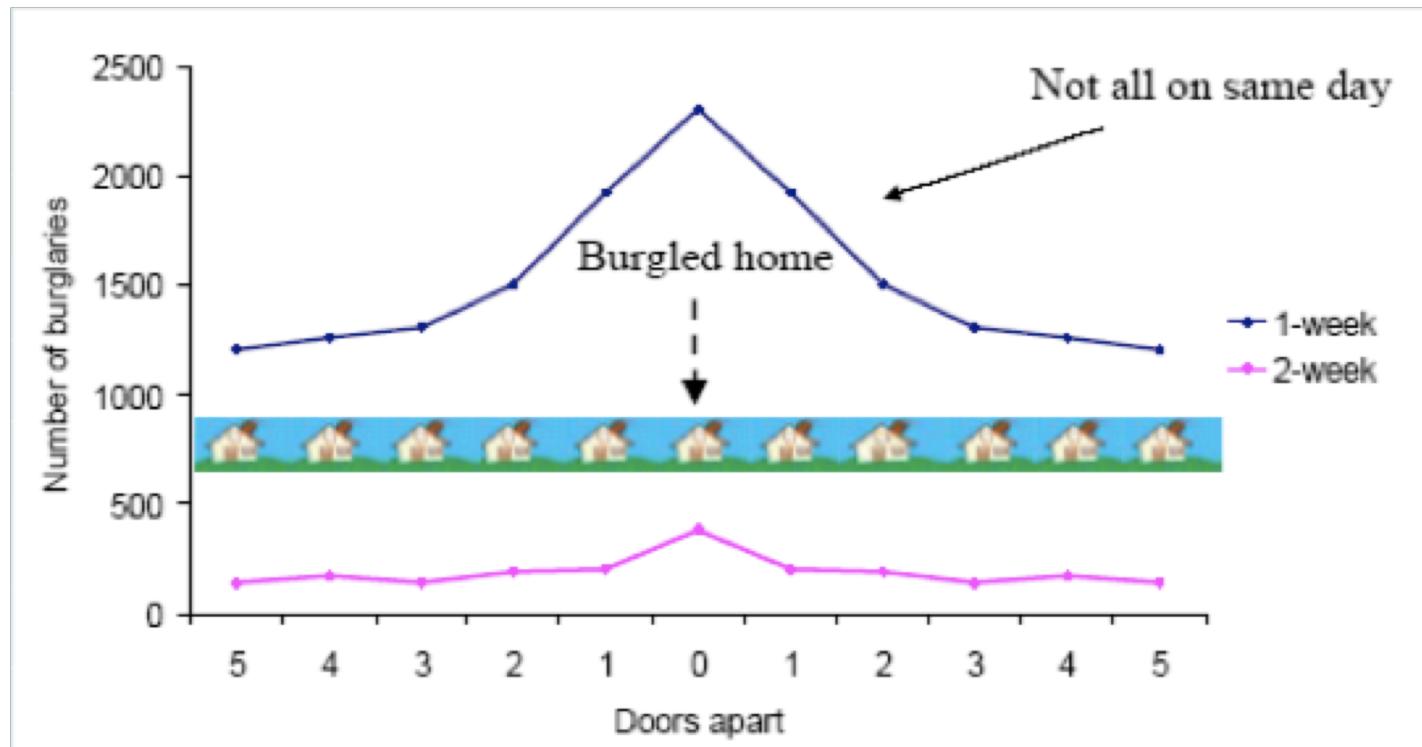
Seguridad Ciudadana

- ▶ Predicción de delitos por zona geográfica
 - ▶ Victimización cercana
 - ▶ Victimización repetida
- ▶ Zonas donde prevenir, cómo y cuándo

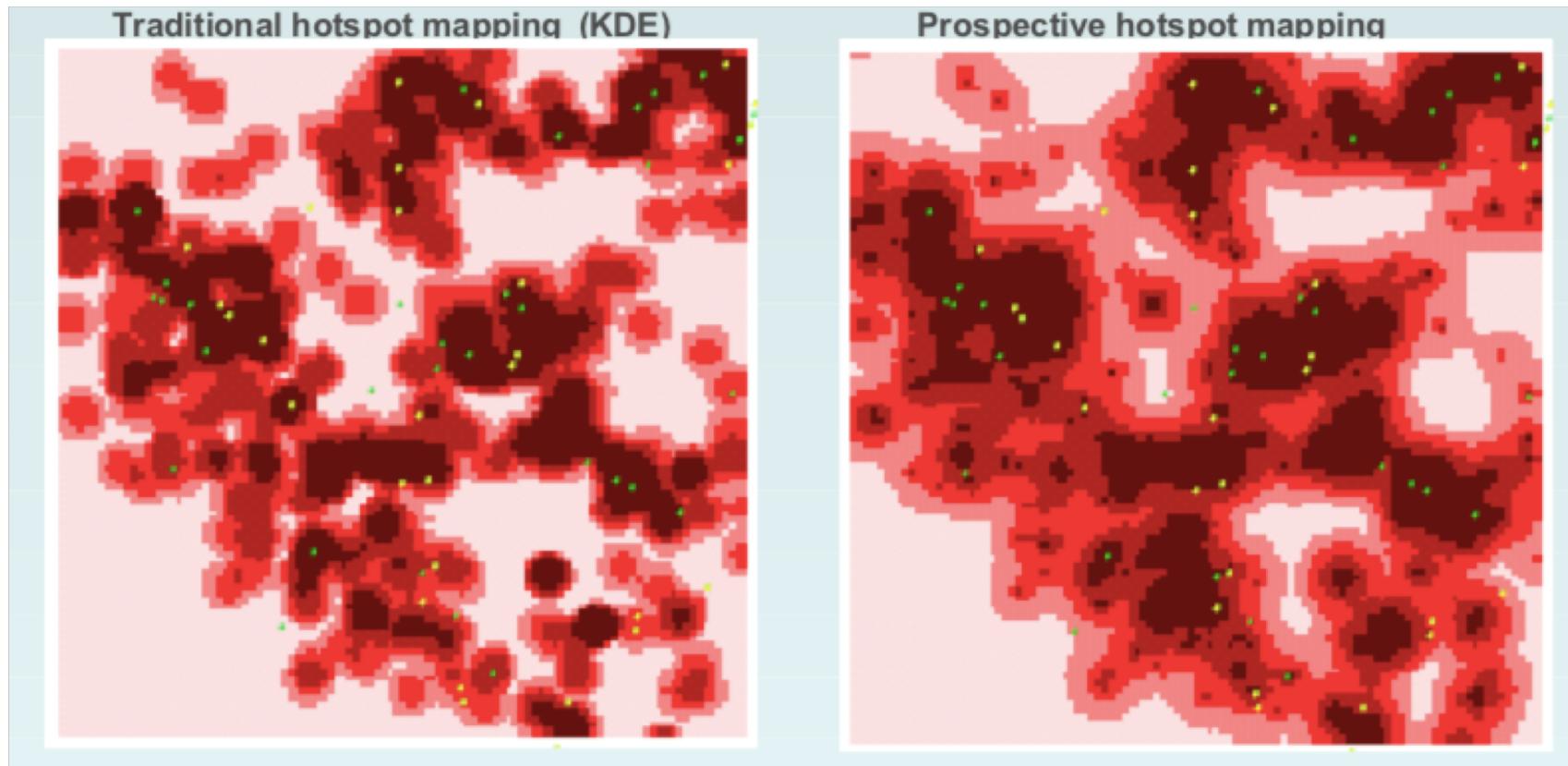
Generadores, Atractores, Habilitadores del Crimen



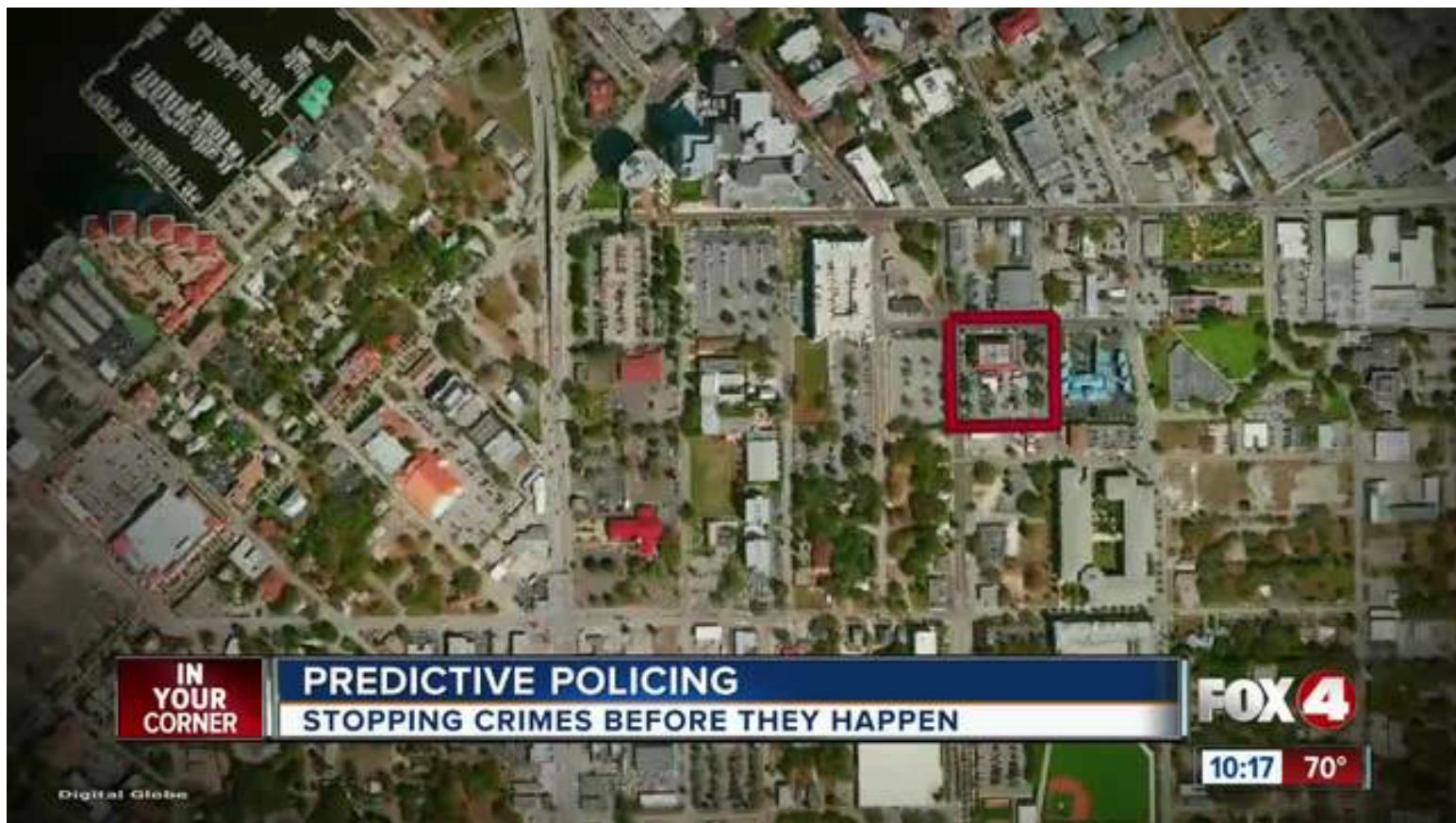
Propiedades dentro de un radio de 400m tienen mayor riesgo de ser asaltadas dentro de 1 mes



Predicción de zonas de delito con mayor precisión



Prevención



Video Analytics

- ▶ **Video Analytics** es una tecnología que se usa para analizar video para datos específicos, comportamiento o actitudes.
- ▶ Video Analytics ofrece la posibilidad de generar alertas preventivas según parámetros predefinidos o incluso realizar comparaciones de rostros con listas previamente definidas a través del uso de firmas biométricas, entregando valiosa información para un accionar eficiente de las instituciones de seguridad

Prevención



¿Con qué fuentes de datos contamos?

Experiencias:

- Proyecto LUCA – NYC: 78 variables; 4,2 TB de información. 72% de precisión.
- **PredPol.** Usa tipo de crimen, lugar y fecha exacta.

¿Con qué fuentes de datos contamos?

- Integrar datos públicos, demográficos, datos móviles, noticias, redes sociales, etc.
 - Determinar factores de por qué hay zonas calientes.
- + información y de calidad ayuda a mejorar la calidad predictiva.



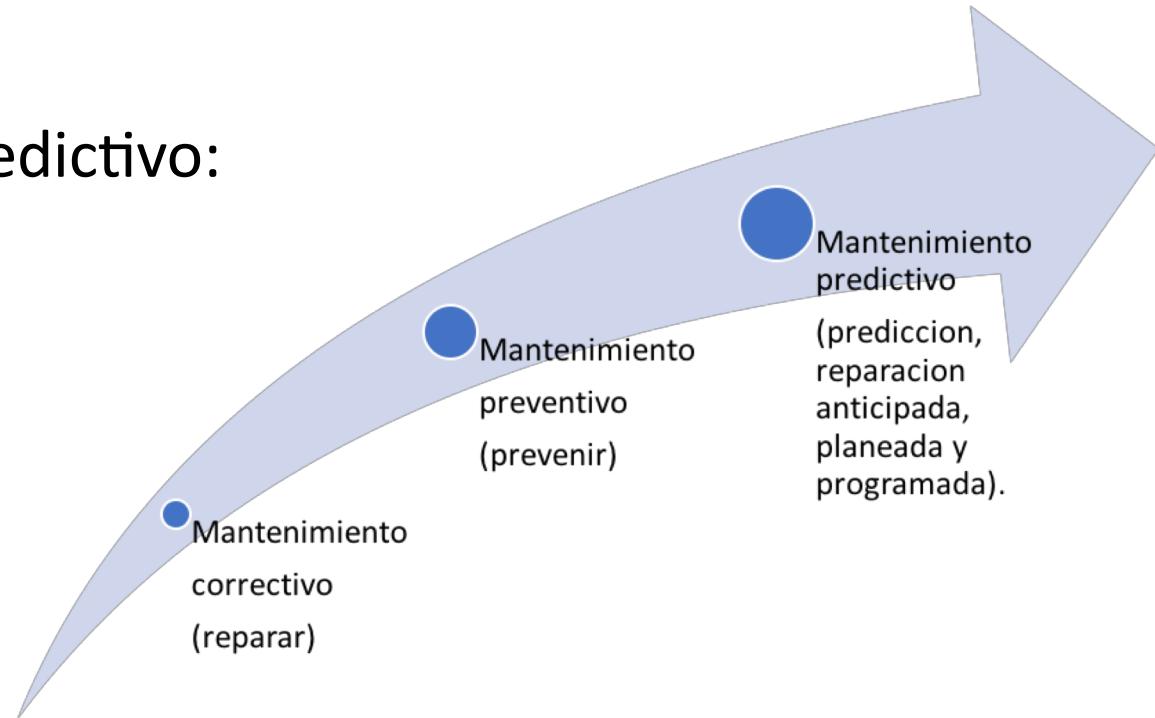
¿Qué ha generado la Industria 4.0?

- ▶ Cambio acelerado de los procesos industriales.
- ▶ Fábricas inteligentes combinan flexibilidad y tecnología.

¿Qué ha generado la Industria 4.0?

► Mantenimiento predictivo:

- Anticipación.
- Eficiencia.
- Productividad.



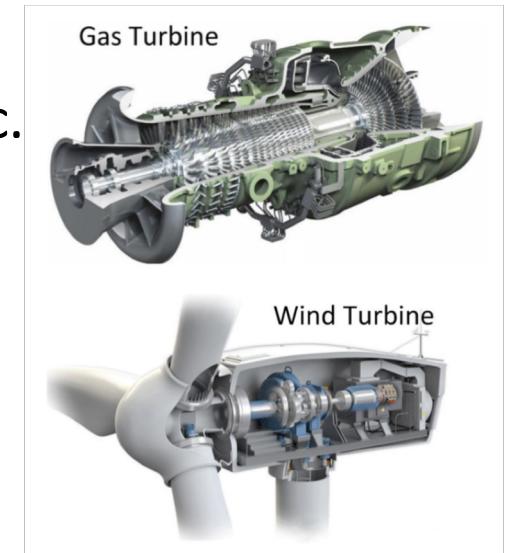
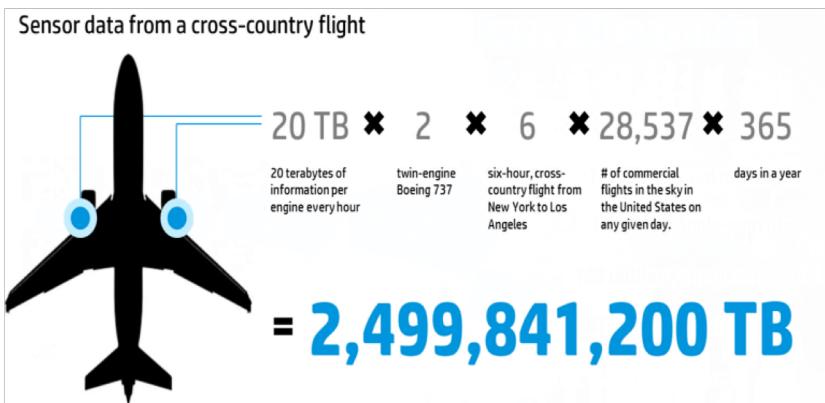
Ventajas del mantenimiento predictivo

- Enfoque proactivo para la gestión de activos.
- Planificación y programación automática.
- Aumento de eficiencia y productividad.
- Aumento de la rentabilidad.
- Mejora de satisfacción al cliente.
- Etc.

Industrial Internet of Things (IIoT)

¿Con qué fuentes de datos contamos?

- Datos de sensores: temperatura, vibración, presión, etc.
- Datos de uso del operador.



✓ Oportunidades y Desafíos

Big Data en Salud

- Potencial de desarrollo
- Hay iniciativas, pero no suficientes



Source: MaRSHealthKick

Oportunidades

Solo el 17 % de los países se sirven de Big Data en sus sistema sanitario, según la Organización Mundial de la Salud (OMS)

Fuente: https://www.balancesociosanitario.com/Big-Data-en-Sanidad-Hacia-la-medicina-personalizada-y-predictiva_a5603.html

- Prevención de errores de medicación
- Identificación de pacientes de alto riesgo
- Reducción de los costos hospitalarios y de los tiempo de lista de espera
 - Predecir la tasa de admisión
- Detectar Fraude
- Detectar riesgos relacionados con el patrón genético, que permitan desarrollar medidas de prevención y tratamiento personalizados

- Medicina personalizada:
 - Farmacogenética
 - Cura de cáncer



Desafíos

- Ficha Clínica Electrónica.
 - descifrar y obtener valor predictivo a partir del texto de los informes clínicos que acompañan a cada paciente.
- Extraer conocimiento de fuentes heterogéneas y complejas.
- Desarrollo de metodologías para datos complejos: imágenes, etc.

✓ Actividad

- ▶ Identifique un caso o problema de negocio.
- ▶ ¿Qué datos usaría? Fuentes de datos.
- ▶ Identifique los tipos de analítica: Descriptiva, Predictiva, Prescriptiva.
- ▶ ¿Cuál es el valor?

Analítica Empresarial para la Toma de Decisiones Efectivas. ¿Cómo usar Big Data?

Rolando de la Cruz

Director Académico Magíster en Data Science

rolando.delacruz@uai.cl