

El Dato

Unidad 5

Analítica del dato Data Analytics



Reflexión previa

Todas las personas usamos y analizamos datos para tomar decisiones: ¿qué decisión de compra tomar en función de las características y precios?, ¿qué ropa ponerme en función del tiempo?, ¿qué ruta escoger para ir al trabajo?, ¿debo hacer deporte para mejorar mi salud y calidad de vida?, ¿qué comer?, etc.

Una buena gestión de los datos nos ofrece muchas ventajas. Comprar más barato, ahorrar tiempo y dinero, no pasar frío o no mojarnos, huir de atascos o mejorar nuestra salud. Esto que se ve muy fácil a nivel personal, es trasladable a las organizaciones.

Los datos que las organizaciones pueden tener a su disposición y que deben gestionar no van a parar de crecer. **El volumen, la variabilidad y la velocidad de crecimiento dificultan la gestión de los datos.** Las tecnologías convencionales no permiten atender las necesidades de captura, procesamiento y análisis que la nueva realidad nos exige.

¿Qué necesitamos para poder gestionar tales cantidades de datos heterogéneos?

Del Bit al Big Data

El término Big Data se refiere a un conjunto de datos que es **demasiado grande o demasiado complejo** para que los dispositivos informáticos ordinarios lo procesen.

Si nos fijamos en la historia reciente de los datos, en 1999 teníamos un total de 1,5 exabytes de datos y 1 gigabyte se consideraba big data. Hoy en día, en la era del Zettabyte, estaríamos hablando de Big Data a partir de 1 Terabyte (1000 Gb). Por hacer un símil con la superficie terrestre, podríamos decir que en 1999 la superficie del estado (505.370 km²) era "Big" mientras que hoy en día necesitaríamos la superficie de todo el planeta tierra (510.072.000 km²) para considerar esa superficie como "Big".

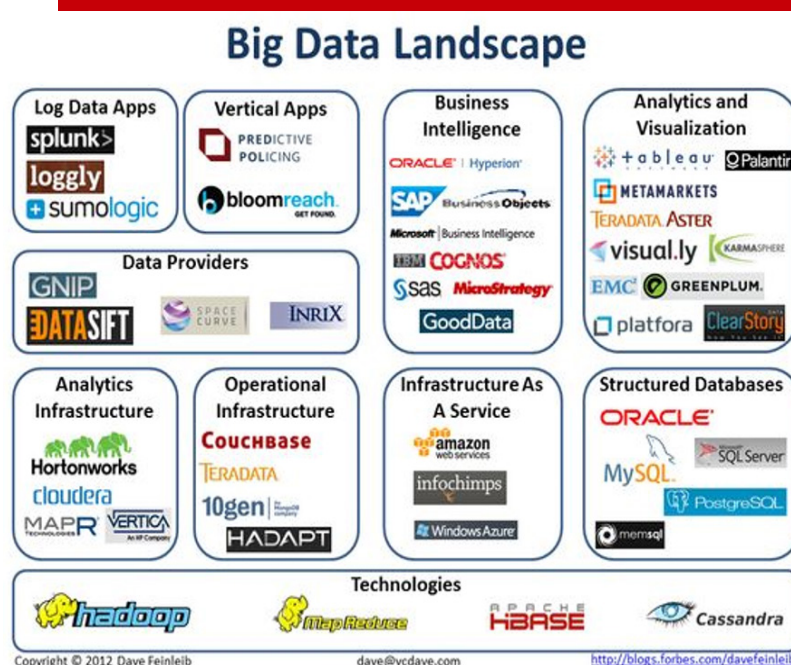
Como vemos, **el volumen de datos que se generan es abismal** y de una **casuística extremadamente compleja** para su análisis.

Además, hay otra característica importante en el Big Data. Las organizaciones cada vez exigen que el **análisis** sea lo más cercano posible al **tiempo real**. Esa inmediatez en el procesamiento de grandes cantidades de datos heterogéneos y en el posterior análisis, hace que el reto sea aún mayor.

Big Data es un término que describe el gran volumen de datos, **tanto estructurados como no estructurados**, que inundan los negocios cada día. Pero no es la cantidad de datos lo que es importante. **Lo que importa con el Big Data es lo que las organizaciones hacen con los datos**. La tecnología existente y la futura permite tratar y analizar esos volúmenes de datos heterogéneos para

poder dar **respuesta a las preguntas y necesidades de las organizaciones y extraer valor**.

Antes de la aparición del Big Data las organizaciones extraían los datos de sus aplicaciones, ERPs, CRMs para procesarlos con herramientas de analítica. Pero hoy en día es posible cargar información adicional que NO se encuentra alojada en los sistemas de información de la organización: datos de redes sociales, datos estadísticos de terceros, datos de operadoras de telecomunicaciones, servicios de datos on-line, o incluso todos los datos que es posible obtener a partir de la sensorica. Para poder gestionar y analizar toda esa cantidad de información es necesario ir un paso más allá con la tecnología.



Dada esta rápida expansión, es necesario hacer frente a nuevas necesidades analíticas derivadas de la cantidad tan ingente y heterogénea de datos que tenemos a nuestro alrededor. Y es que el por qué del Big Data y la analítica de datos lo explica muy bien Mario Benedetti:

"Cuando creíamos que teníamos todas las respuestas, de pronto, nos cambiaron las preguntas".

Mario Benedetti

Test

Para adentrarnos en el mundo de la analítica, hagamos antes un test rápido.

1. 5 robots fabrican 5 tostadoras en 5 minutos, ¿en cuánto tiempo fabricarán 100 robots 100 tostadoras?

- a) en 100 minutos
- b) en 50 minutos
- c) en 20 minutos
- d) en 5 minutos

Si has contestado 100 minutos, no te sientas mal. No es la respuesta correcta, pero si es la más frecuente.

Esta pregunta forma parte de un test de inteligencia realizado a un grupo de 3.000 personas de universidades americanas, incluida Harvard, del que solo el 17% acertó la respuesta correcta.

Según el profesor de la Universidad de Yale, Shane Frederick, **solemos buscar inicialmente la respuesta más rápida**. Sin embargo, responderán acertadamente quienes adopten una actitud más reflexiva y metódica, venciendo la “barrera” de la primera reacción e intuición.

Dicho de otra manera, nuestro cerebro trabaja en dos niveles de profundidad: el pensamiento superficial es el responsable de acciones cotidianas y de las primeras reacciones. **El pensamiento analítico se encarga de procesar y analizar la información** más importante para dar una respuesta.

Si a 5 robots les cuesta 5 minutos hacer 5 tostadoras, entonces a 1 robot le cuesta 5 minutos hacer 1 artículo. Si tenemos 100 robots, cada uno de ellos necesitará 5 minutos para elaborar un artículo. Si los 100 trabajan a la vez, necesitaremos 5 minutos para hacer 100 tostadora. Por lo tanto, la respuesta son 5 minutos.

¿Qué es la analítica del dato?



La analítica es el **análisis computacional sistemático de los datos o las estadísticas** y se utiliza para descubrir, interpretar y comunicar patrones significativos en los datos. También implica la aplicación de patrones de datos para una toma de decisiones eficaz.

¿Recordáis cuando decíamos al principio del curso que la mayoría de las personas comienzan a buscar datos cuando tienen una pregunta en mente?

Bien, pues lo mismo sucede en las organizaciones. **Los datos son las respuestas a las preguntas que surgen en el día a día de una organización. Y la analítica de datos es la herramienta que nos ayuda a encontrar las respuestas a esas preguntas.**

- ¿Dónde está el fraude fiscal?
- ¿Qué actividades debemos subvencionar?
- ¿Qué colectivos precisan asistencia social?
- ¿Qué medidas debemos impulsar para mejorar la movilidad del territorio?
- ¿Qué hacemos para lograr una Bizkaia más sostenible?

Para dar respuesta a todas estas preguntas necesitamos analizar datos. Volúmenes enormes de datos que deben ser analizados sistemáticamente mediante plataformas y aplicaciones informáticas potentes que sean capaces de cruzar datos de diferentes orígenes y fuentes de información y presentar los resultados de manera ágil.

Tipos de analítica

Veamos de manera rápida los tipos de analítica existentes y para qué pueden ser utilizados.

¿Qué ocurrió?

Descriptiva: Ayuda a responder las preguntas sobre algo que ya haya sucedido. Es un tipo de analítica utilizada frecuentemente como medio para explicar algo a las personas interesadas.

Un ejemplo de analítica descriptiva son las memorias anuales de las empresas. Las grandes organizaciones o las propias administraciones, publican cada año un informe descriptivo de las principales métricas del último ejercicio.

Por muy útil que sea la analítica descriptiva es recomendable combinarla con otro método, como el diagnóstico, para profundizar en el porqué de algo. El análisis descriptivo señalará lo que ha sucedido, pero todavía tendrá que explorar el razonamiento que hay detrás del evento.

¿Por qué pasaron las cosas?

Diagnóstico: Al igual que un diagnóstico, los análisis de diagnóstico proporcionan una visión de por qué ha sucedido algo. Funcionan junto con los análisis descriptivos para explicar mejor los resultados críticos. Si echas un vistazo a los indicadores clave de rendimiento (KPI) y quieres entender por qué algo ha mejorado o empeorado, los análisis de diagnóstico ayudan a:

- Identificar anomalías en los datos
- Recoger los datos que ayudan a entender los cambios
- Utilizar técnicas estadísticas para ayudar a explicar dichas anomalías
- Como ejemplo concreto, la analítica de diagnóstico puede ayudar a responder por qué tu equipo puede haber incumplido sus objetivos en un periodo determinado.

¿Qué ocurrirá?

Predictivo: Como su nombre indica, la analítica predictiva trabaja para responder a preguntas sobre lo que podría ocurrir en el futuro. Este método analítico aprovecha los datos del pasado para evaluar las tendencias y estimar la probabilidad de que algo se repita. Para que la analítica predictiva funcione se utiliza el análisis estadístico, la regresión y el aprendizaje automático.

¿Qué haremos si ocurre?

Prescriptiva: Si te encuentras en una posición crítica para tomar una decisión sobre el futuro pero sientes inseguridad sobre qué opción tomar, la analítica prescriptiva puede ser un salvavidas. La analítica prescriptiva funciona encontrando patrones a partir de grandes conjuntos de datos y luego estima la probabilidad de diferentes resultados.

Evolución de la analítica de datos



Hasta hace relativamente poco tiempo, la analítica realizada era puramente descriptiva.

Se analizaba de manera objetiva la base de datos históricos de la empresa, y se intentaban extraer conclusiones estadísticas de acciones pasadas que pudieran defender una decisión estratégica determinada con un grado de confianza alto.

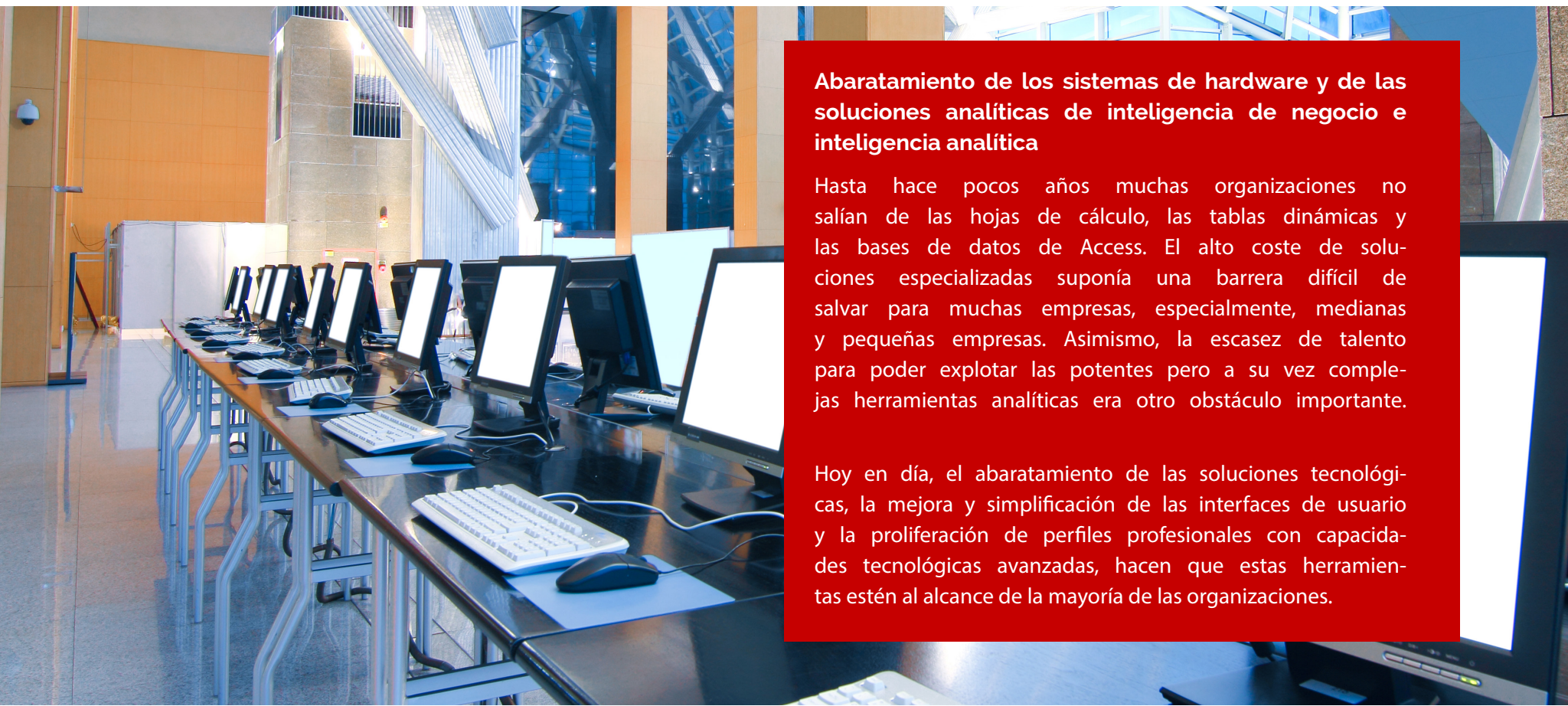
Sin embargo, en los últimos años, gracias a la labor de personal técnico, estadístico y matemático, se han conseguido desarrollar diferentes sistemas de analítica predictiva y prescriptiva, siempre basados en **algoritmos estadísticos y reglas matemáticas**, que no solo analizan los históricos de la empresa, sino

que buscan **predecir escenarios hipotéticos futuros**, y tener la capacidad de, en dichos escenarios, modificar diferentes indicadores para analizar su incidencia directa en las decisiones principales y **poder tomar la decisión más óptima**.

Esto ha comenzado a suceder en muchas organizaciones debido a, entre otros, los siguientes **factores**:

Abaratamiento de los costes de almacenamiento y computación de la información

El almacenamiento en la nube (cloud) ha supuesto otra pequeña revolución en el almacenamiento de la información. Las organizaciones, acostumbradas hasta hace bien poco a destinar partidas económicas importantes y espacios físicos grandes para la ubicación de sus servidores informáticos físicos, han encontrado una oportunidad en la nube. Los datos están alojados en espacios de almacenamiento virtualizados, por lo general aportados por terceros. Las compañías de alojamiento operan enormes centros de procesamiento de datos. Las personas y entidades usuarias que requieren estos servicios compran, alquilan o contratan la capacidad de almacenamiento necesaria. Las operadoras de los centros de procesamiento de datos, a nivel servicio, virtualizan los recursos según los requerimientos de las y los clientes. Algunos ejemplos son: Dropbox, iCloud de Apple, Google Drive o OneDrive de Microsoft



Abaratamiento de los sistemas de hardware y de las soluciones analíticas de inteligencia de negocio e inteligencia analítica

Hasta hace pocos años muchas organizaciones no salían de las hojas de cálculo, las tablas dinámicas y las bases de datos de Access. El alto coste de soluciones especializadas suponía una barrera difícil de salvar para muchas empresas, especialmente, medianas y pequeñas empresas. Asimismo, la escasez de talento para poder explotar las potentes pero a su vez complejas herramientas analíticas era otro obstáculo importante.

Hoy en día, el abaratamiento de las soluciones tecnológicas, la mejora y simplificación de las interfaces de usuario y la proliferación de perfiles profesionales con capacidades tecnológicas avanzadas, hacen que estas herramientas estén al alcance de la mayoría de las organizaciones.

Transformación digital de las organizaciones

La digitalización ofrece demasiadas oportunidades de crecimiento e innovación en todos los ámbitos, sectores, regiones y cualquier tipo de empresa como para desaprovecharla. No obstante, la transformación digital es imposible sin una adecuada gestión de los datos, es decir, de la materia prima principal para poder llevar a cabo dicha transformación.

Sistemas integrados de información que facilitan el acceso y la explotación de la información

Explotar y analizar datos de diferentes sistemas de información no integrados es una tarea que requiere un esfuerzo enorme debido a la necesidad de depurar, estandarizar, transformar, ordenar y estructurar los datos antes de poder empezar a analizarlos. Los sistemas integrados de información impulsados por la mayoría de las organizaciones parten de un punto de partida en el que esas labores son menos exigentes en costo y tiempo.

¿Qué valor nos aporta la analítica de datos?

La gran variedad de datos que generan las empresas contiene información valiosa, y la analítica de datos es la forma de descubrirla. **La analítica de datos puede ayudar a una organización en todo**, desde la personalización de una propuesta de marketing para cliente individual hasta la identificación y mitigación de riesgos para su negocio. Echemos un vistazo a algunos de los beneficios del uso de la analítica de datos.



Personalizar la experiencia de cliente

Las organizaciones recopilan datos de clientes a través de muchos canales diferentes, como el canal presencial, el canal electrónico o las redes sociales. Al utilizar la analítica de datos para crear **perfiles completos de clientes** a partir de estos datos, las organizaciones pueden obtener información sobre su comportamiento para ofrecer una experiencia o un servicio más personalizado.

Esto es algo que las empresas privadas llevan tiempo realizando, especialmente las grandes empresas que atienden a cliente final. Pero, ¿y por qué las administraciones públicas no han comenzado a explotar esta **vía de innovación**? ¿Acaso ofrecer servicios personalizados a la ciudadanía no se podría ver como una **obligación de la administración** para poder dar respuesta a gran parte de los principios que deben ser respetados por el sector público en sus actuaciones y relaciones con la ciudadanía y resto de administraciones?

Facilitar la toma de decisiones

Las organizaciones pueden utilizar el análisis de datos para orientar las decisiones empresariales y minimizar las pérdidas financieras o el deterioro de la calidad de atención a sus clientes. La analítica predictiva puede sugerir lo que podría ocurrir en respuesta a los cambios en el negocio previstos, y la analítica prescriptiva puede indicar **cómo debería reaccionar el negocio** ante estos cambios.

Racionalizar las operaciones

Las organizaciones pueden **mejorar la eficiencia operativa** mediante el análisis de datos. La recopilación y el análisis de datos sobre la cadena de suministro pueden mostrar dónde se originan los retrasos o cuellos de botella en la producción y ayudar a predecir dónde pueden surgir los problemas en el futuro. Si una previsión de la demanda muestra que un proveedor específico no podrá manejar el volumen necesario para la temporada de vacaciones, una empresa podría complementar o sustituir a este proveedor para evitar retrasos en la producción.

¿Y si hiciéramos un ejercicio análogo con el flujo de tramitación de un determinado procedimiento administrativo? Los datos podrían ayudar a las administraciones a mostrar dónde se originan los cuellos de botella y ayudar a prescribir acciones para mitigarlos o incluso eliminarlos. E incluso detectar probables atascos administrativos e incluso incumplimiento de plazos ante la entrada de un montón de solicitudes en periodo de vacaciones o con personal de baja.



Mitigar el riesgo y manejar los contratiempos

Los riesgos están en todas las áreas de todas las organizaciones. En el sector privado podríamos citar la pérdida de talento o de clientes, los impagos, la seguridad informática o incluso física del personal y/o las instalaciones, la responsabilidad legal...

La analítica de datos nos puede **ayudar a conocer y entender los riesgos**, a predecir situaciones o eventos futuros y a prescribir acciones concretas para mitigar los riesgos y/o prepararse para manejar los contratiempos.

Las caídas de ventas o de ingresos en un determinado sector pudiera ser un indicador de una futura caída en la recaudación fiscal, un aumento de los aplazamientos y embargos, pérdida de empleo, etc. Este tipo de indicadores podrían ser utilizados para establecer medidas preventivas que pudieran minimizar los efectos negativos a medio y largo plazo.

Mejorar la seguridad informática

Todas las empresas se enfrentan a amenazas a la seguridad de los datos. Las organizaciones pueden utilizar la analítica de datos para diagnosticar las causas de anteriores violaciones de datos mediante el procesamiento y la visualización de los datos pertinentes. Por ejemplo, el departamento de tecnologías de la información (TI) puede utilizar aplicaciones de análisis de datos para analizar, procesar y visualizar sus registros de auditoría para determinar el curso y los orígenes de un ataque. Esta información puede ayudar al departamento de TI a localizar las vulnerabilidades y aplicarles parches.

Los departamentos de TI también pueden utilizar modelos estadísticos para **prevenir futuros ataques**. Los ataques suelen implicar un comportamiento de acceso anormal, especialmente en el caso de los ataques basados en la carga, como un ataque de denegación de servicio distribuido (DDoS). Las organizaciones pueden configurar estos modelos para que se ejecuten de forma continua, con sistemas de supervisión y alerta superpuestos para detectar y señalar las anomalías, de modo que los profesionales de la seguridad puedan actuar inmediatamente.



Algunos ejemplos relevantes:

UBER

Una de las primeras ventajas en analítica de datos se observó en los servicios de localización a través de GPS. Los teléfonos inteligentes con posicionamiento satelital han permitido desarrollar mapas avanzados y algoritmos para poder optimizar las rutas de circulación en las ciudades. De igual manera, la exitosa y revolucionaria empresa que proporciona a sus clientes una red de transporte privado, a través de su software de aplicación móvil, utiliza el **análisis de la oferta y la demanda en localización** para cambiar las tarifas según evoluciona la demanda a lo largo del día con lo cual potencian aún más sus ingresos.

Uber

Amazon

En el campo del retail, las empresas analizan de sus transacciones, así como el comportamiento de sus clientes para personalizar las ofertas. El objetivo principal es crear **modelos predictivos**, capaces de detectar las futuras necesidades del cliente o que productos se venderán mejor.

Quienes compran en Amazon, pueden constatar cómo la empresa le sugiere productos que le puedan interesar como resultado de entender las preferencias del consumidor. Las empresas amplían sus datos con la información de las redes sociales, logs de navegación, análisis de textos y datos



de sensores para obtener una imagen de cliente que analizar y computar.

Mediante la aplicación de este tipo de herramientas, hay empresas que han sido capaces de detectar cuándo sus clientes esperaban un bebé.

Google Fit y Apple Watch

Los procesos de Data Analytics están muy ligados con los fenómenos de Big Data e Internet of Things, que permiten y permitirán aún más en el futuro, extraer enormes cantidades de información para analizar y generar conocimiento. Sin embargo, los sistemas de análisis de datos no son solo beneficiosos para las grandes empresas, también las personas que los usan pueden verse beneficiadas con la integración de procesos de análisis de sus datos.

Así, los dispositivos “wearables” como Smart watches o pulseras registran automáticamente datos de consumos de calorías, niveles de actividad y condición física, o patrones de sueño, permitiendo con su uso obtener un **análisis**



detallado de su salud, y, aún más importante un análisis del conjunto de datos de todo el colectivo, que permitirá desvelar insights beneficiosos para el conjunto.

BBVA

Las finanzas y el sector bancario son otro de los grandes beneficiados por la extracción de información útil de la gran cantidad de datos y transacciones que generan.

La entidad BBVA ya ha realizado varios proyectos de extracción y análisis de información financiera de las que ha podido obtener tanto las diferentes actividades comerciales de las grandes urbes (Urban Discovery), el impacto económico de la celebración de grandes eventos o Commerce360 que permite conocer datos que hasta ahora solo eran accesibles a la gran distribución, y que les proporciona una ventaja competitiva a la hora de **diseñar sus estrategias comerciales y de marketing**.



La reelección de Obama

Como último gran ejemplo de las posibilidades de aplicación del Data Analytics, la incorporación del uso del análisis de datos en las campañas políticas, como una variante de las campañas de marketing.

Así, Obama decidió utilizar la analítica de datos para su reelección en 2012, con hasta 20 personas encargadas exclusivamente en la interpretación de los datos recibidos, extrayendo y generando estrategias en función de las preferencias de sus votantes y sus canales de comunicación preferidos.



Potencial de la analítica de datos en la administración pública

Nos dirigimos a un mundo de información y posibilidades casi ilimitados. El caso de Estonia es un claro ejemplo de que lo que puede parecer ficción es posible. Es evidente que no es un camino fácil pero **es un camino que hay que recorrer**.

Debemos analizar cómo estamos utilizando los datos para encontrar caminos de mejora.

Los datos predictivos nos ayudan a saber con anticipación si traer un paraguas al trabajo o coger el autobús.



Los datos del tráfico se utilizan para sincronizar los semáforos, predecir los horarios de llegada de los trenes y autobuses y ayudarnos a encontrar la vía más rápida para llegar al ensayo de nuestro hijo o hija a tiempo.

Los dispositivos como las **pulseras de actividad (fit-band)** o los **relojes inteligentes (smartwatch)** nos ayudan a realizar un seguimiento de nuestro estado físico, de modo que podamos elegir opciones para tener vidas más largas y saludables, y el mundo científico está analizando terabytes de información genética para encontrar curas nuevas y desarrollar tratamientos más efectivos y personalizados.

Como podemos ver, los datos que las organizaciones van a atener a su disposición no va a parar de crecer. La tecnología nos permite analizarlos para extraer valor de ellos.



Idea Fuerza

Las tecnologías actuales nos permiten **recopilar, almacenar, tratar, explotar y analizar cantidades enormes de datos que podríamos utilizar para mejorar y crear nuevos servicios como nunca antes habríamos imaginado. ¿A qué estamos esperando?**

Ejemplos reales en el sector público

Veamos ejemplos reales de cómo las organizaciones, incluidas las del sector público, se han beneficiado de la analítica de datos para mejorar sus procesos:



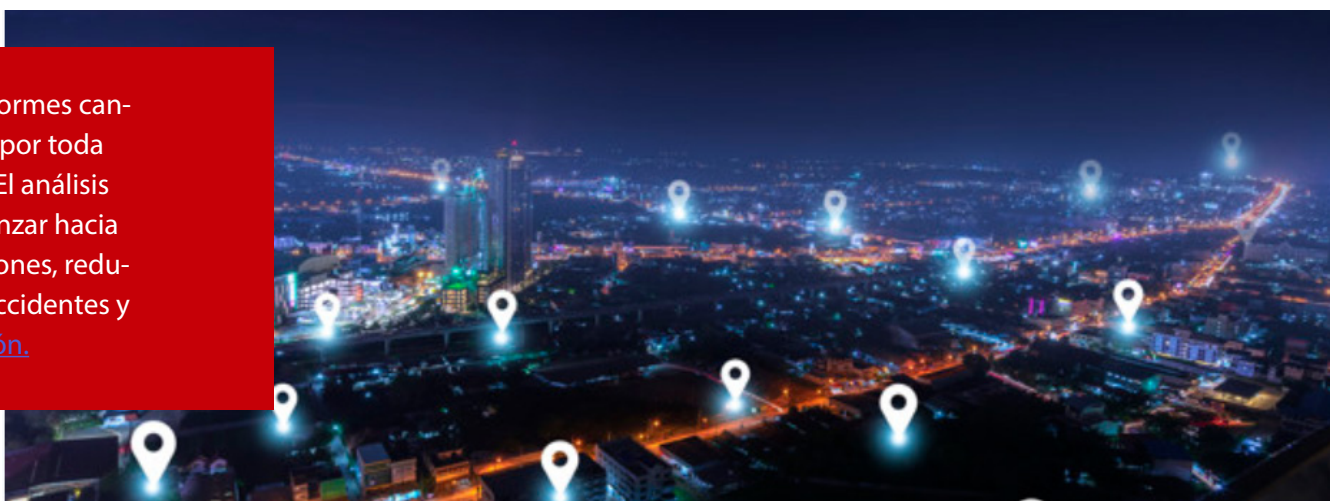
En Nueva Orleans, los datos sobre la naturaleza, la ubicación y el momento de las emergencias pasadas sirvió para predecir dónde y cuándo podrían ocurrir futuras emergencias. Este análisis ayudó a identificar los lugares óptimos para estacionar las ambulancias en espera y para reducir los tiempos de respuesta. [Más información.](#)

El análisis del Ayuntamiento de Newcastle sobre las personas que son "ninis" ayuda a la autoridad local a identificar a los niños con mayor riesgo de no tener empleo, educación o formación. [Más información.](#)



Gracias al análisis de los datos en tiempo real de la movilidad de la flota de vehículos de una empresa de limpieza y recogida de residuos urbanos, es posible adaptarse a contratiempos o situaciones excepcionales como un confinamiento general o una tormenta de nieve, encontrando rutas alternativas que permitan continuar prestando servicio. [Más información.](#)

En el mundo de las Smart Cities, éstas generan enormes cantidades de datos gracias a los sensores repartidos por toda las ciudades, que son procesados en tiempo real. El análisis de esos datos puede ayudar a los gobiernos a avanzar hacia una movilidad más sostenible, disminuir las emisiones, reducir la huella de carbono, disminuir el número de accidentes y un largo etcétera de posibilidades. [Más información.](#)





Bizkaia