
Business intelligence y business analytics

PID_00264726

Xavi Font

Tiempo mínimo de dedicación recomendado: 3 horas



Xavi Font

El encargo y la creación de este recurso de aprendizaje UOC han sido coordinados por el profesor: Jordi Ayza Graells (2019)

Primera edición: marzo 2019

© Xavi Font

Todos los derechos reservados

© de esta edición, FUOC, 2019

Av. Tibidabo, 39-43, 08035 Barcelona

Diseño: Manel Andreu

Realización editorial: Oberta UOC Publishing, SL

Ninguna parte de esta publicación, incluido el diseño general y la cubierta, puede ser copiada, reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea éste eléctrico, químico, mecánico, óptico, grabación, fotocopia, o cualquier otro, sin la previa autorización escrita de los titulares del copyright.

Índice

Introducción	5
Objetivos	6
1. Introducción al <i>business intelligence</i>	7
1.1. La importancia de la información.....	9
1.2. Arquitectura BI.....	11
1.3. Modelización del negocio.....	13
2. Introducción al <i>business analytics</i>	14
2.1. ¿Qué ofrece el <i>business analytics</i> ?	14
2.2. Métodos de <i>business analytics</i>	15
2.2.1. Contraste de hipótesis	16
2.2.2. Regresión lineal	16
2.2.3. Regresión logística.....	16
2.2.4. Redes neuronales y métodos de <i>deep learning</i>	17
2.2.5. SVM	18
2.2.6. Series temporales	18
2.2.7. Sistemas de recomendación	19
2.2.8. <i>Topic modelling</i> (LDA)	19
2.2.9. Árboles de decisión	19
2.2.10. Métodos de ensamblaje.....	20
2.2.11. <i>Clustering</i>	21
2.2.12. Reducción de la dimensionalidad	22
2.3. Arquitectura BA	22
3. Soluciones en el entorno industrial	24
3.1. BI en el entorno industrial	25
3.2. <i>Business analytics</i> en el entorno industrial.....	25
3.3. Proveedores de soluciones BA	26
4. Toma de decisiones basada en datos	28
4.1. Evolución de las organizaciones respecto a los datos	28
4.1.1. Organización resistente a los datos	28
4.1.2. Organización consciente de los datos.....	28
4.1.3. Organización guiada por los datos.....	29
4.1.4. Organización conocedora de los datos.....	29
4.1.5. Organización dirigida por los datos.....	30
4.1.6. Caso Challenger	30
Resumen	33

Introducción

En los últimos años, la mayoría de las empresas y organizaciones han hecho uso de un conjunto de herramientas y soluciones de software con el propósito de gestionar, mantener y controlar los diferentes procesos de negocio. El objetivo principal era garantizar la rentabilidad y asegurar un cierto grado de competitividad.

Aunque son posibles diferentes opciones, las dos más utilizadas son: *business intelligence* (BI) y *business analytics* (BA). ¿Cuáles son las diferencias básicas entre ambas soluciones?

Las soluciones BI se centran principalmente en analizar los datos actuales y que permiten visualizar la operativa del negocio. En cambio, las soluciones BA van un paso por delante en cuanto a complejidad y capacidad para generar nuevo conocimiento y valor para la empresa. Utilizan datos históricos para predecir tendencias o para identificar patrones de compra. Esta es una primera aproximación a los dos conceptos introducidos de forma simplista. Muchas de las empresas que ofrecen soluciones tecnológicas (como IBM) proponen un conjunto de herramientas dentro del marco de estas propuestas descritas.

Probablemente los conceptos que en este módulo vamos a tratar ya han sido vistos y escuchados repetidamente en los medios de comunicación. Del conjunto de las noticias que cubren (y han cubierto extensivamente) vamos a centrarnos en los sistemas de *business intelligence* (BI) o también llamados de inteligencia de negocio y los sistemas de *business analytics* (BA) o analítica de negocio. Con un denominador común asociado a garantizar la competitividad de un negocio por la mejora en la toma de decisiones. Esta idea simple está dando lugar a este nuevo concepto de *data driven company*, que es una evolución coherente con los avances en el entorno tecnológico y de la industria 4.0. Los negocios actuales deben afrontar cambios globales en los ámbitos tecnológicos y de comportamiento de los clientes y han de ayudarse de herramientas que permitan digitalizarse para dar respuestas rápidas a este entorno cambiante de una manera eficiente.

Objetivos

Los objetivos que pretende conseguir este módulo son los siguientes:

- 1.** Entender los entornos/sistemas de BI
 - a)** Definición y contexto
 - b)** El impulso de las áreas estratégicas del negocio gracias a los sistemas BI
 - c)** Elementos tecnológicos (*data warehouse*) y soluciones
- 2.** Comprender la importancia de los sistemas BA y por qué se ofrecen en la actualidad como alternativa a los entornos BI tradicionales.
- 3.** Diferenciar entre la gestión y el mantenimiento de la información con las decisiones que se puedan tomar a partir de los datos analizados.
- 4.** Aplicaciones y soluciones en el ámbito industrial y la toma de decisiones basada en datos.

1. Introducción al *business intelligence*

El *business intelligence* es un término que lleva ya más de cinco décadas en el vocabulario de los gestores de las empresas. Lo podríamos traducir como *inteligencia de negocio*, si bien vamos a mantener la terminología anglosajona.

La definición más simple y comprensible es la que define el *business intelligence* (BI) como un sistema software que ayuda a la toma de decisiones a partir de la información disponible en el sistema.

Podemos entender el BI como un proceso interactivo que permite la exploración y el análisis de los datos de nuestro sistema de información. Puede que el sistema resida en un gestor tradicional RDBMS o bien, para facilitar la interacción, en un sistema *data warehouse*. Entre otras, podemos tener diferentes visiones relativas a áreas de negocio:

- Clientes
- Ventas
- Proveedores
- Marketing

Un aspecto importante es la exploración de los datos (EDA, *exploratory data analysis*). Esta exploración preliminar permite interpretar mejor la realidad del problema que estemos analizando o facilita la comprensión del proceso de negocio bajo estudio.

Los sistemas de BI dan a la vertiente del análisis un aspecto clave, dado que inducen a la posible extracción de conocimiento relevante para el negocio. Por ejemplo, relación entre diferentes variables, identificación de tendencias o patrones en los procesos de compra de los clientes.

La mayor parte de los sistemas BI operan en un entorno donde los responsables o mandos se enfrentan a plazos muy ajustados, recursos limitados y con el propósito de responder en tiempo real a cualquier problema o demanda que aparezca (siendo estas evidentemente cambiantes). Estos datos residen en un sistema RDBMS (modelo relacional) o en sistemas *data warehouse* (modelo estrella o copo de nieve), que derivan normalmente de uno RDBMS. Obviamente, todo proyecto dentro de la organización que se preste a utilizar un sistema de BI va a definir unos objetivos de interés para la organización:

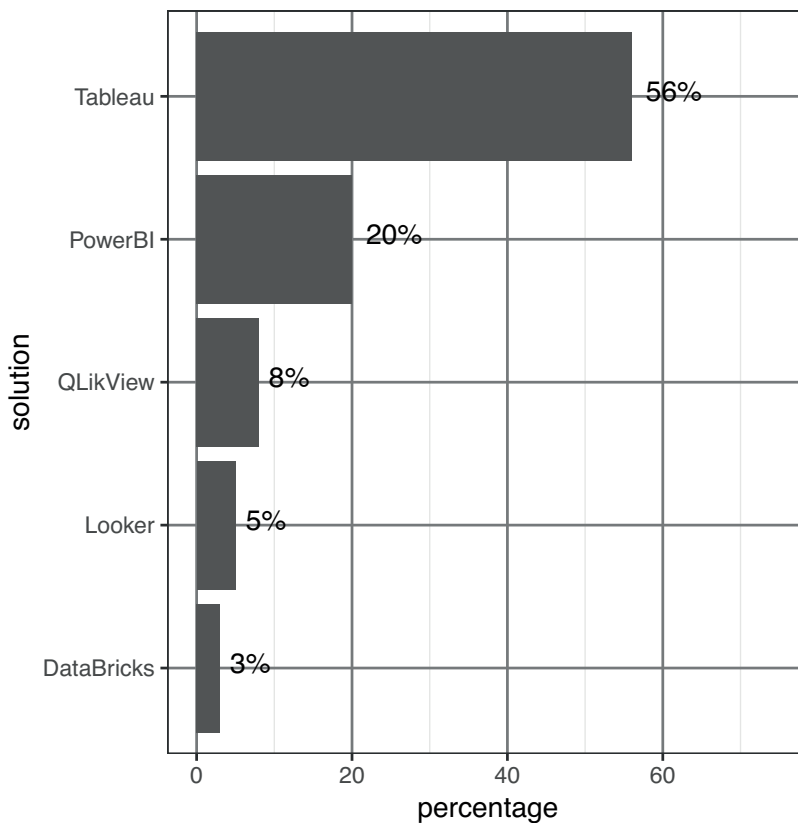
- Análisis de los clientes más rentables.
- Análisis de los clientes menos rentables.
- Productos más vendidos / dónde / margen de beneficio.
- Efecto de una campaña de marketing sobre la ventas.
- Unidades de negocio más eficientes.
- Evolución de los *stocks* y compra a proveedores.

Evidentemente, cada uno de estos objetivos tienen una traducción directa en reducción de costes, incremento de las ventas, conocimiento del comportamiento del cliente y/o del negocio, entre otros.

Un último aspecto no menor al introducirnos en los sistemas BI es la comunicación o herramientas de visualización y/o de generación de informes. Y en este apartado está la explicación del grado de aceptación de estas soluciones. Los usuarios del sistema pueden obtener toda esta información sin intermediarios (vinculados a los departamentos de informática o IT).

En una encuesta reciente realizada por Panoply (Data Warehouse Trends Reports), se establece que las empresas (básicamente pymes) se decantan por utilizar sistemas que destacan por los aspectos más visuales. En la figura 1 se observa el grado de implantación de diferentes soluciones BI.

Figura 1. Uso de herramientas BI por las pymes del mercado americano



Los avances que ofrece la digitalización y que denominamos industria 4.0 tienen un actor de interés que desde el primer instante los consideró estratégicos. El caso de la automoción es uno de estos ejemplos, e incluso podríamos afirmar que ha abanderado la adopción de soluciones de inteligencia empresarial mucho antes que otras industrias más aparentemente próximas a la utilización de soluciones tecnológicas. En este ámbito, la información y la inteligencia son críticas para poder fabricar automóviles, camiones o maquinaria de transporte. En estas situaciones tener una visión clara, precisa y en tiempo real de la cadena de suministro y de los canales de distribución es esencial. De hecho, puede marcar la diferencia entre el éxito y el fracaso. Por estos motivos se puede concluir que una solución de BI no es interpretada como un lujo en esta industria, sino como una necesidad.

No es difícil entender la importancia de la gestión de la cadena de suministro, ya que es absolutamente crítica porque puede contener más de veinte mil piezas necesarias para producir cada automóvil. Estas piezas, además, son producidas a partir de una cadena de suministro que comprende de tres a cinco niveles de casi mil proveedores.

Incluso después de que el vehículo salga de fábrica los sistemas BI pueden recopilar información de venta, tiempos y/o de otros fabricantes para mejorar la calidad de la información que el sistema es capaz de tratar.

1.1. La importancia de la información

Los datos por sí solos no son absolutamente nada. Para que puedan tener algún elemento relevante necesitan un contexto con el fin de que estos se puedan entender y/o interpretar. Un dato con este contexto es lo que llamamos *información*. Imaginad la venta de un vehículo en un concesionario. ¿Cuál es la información más relevante?

- Fecha de compra:
- Fecha de entrega:
- Vendedor:
- Nombre del cliente:
- CP:
- Edad:
- Marca:
- Modelo:
- Motor:
- Color:
- Opciones: lista de las opciones elegidas por el cliente*
- Promoción:
- Precio total
- IVA
- Forma de pago

Enlaces de interés

Recomendamos observar las diferentes soluciones en las siguientes webs:

<https://www.tableau.com/>

<https://bit.ly/1Jupp9D>

<https://databricks.com/>

<https://bit.ly/2RO2ctq>

* Por ejemplo, llantas de aleación, sistema de navegación, telefonía preinstalada, techo panorámico, tapicería de piel...

Observad que hablamos de información porque ya damos por hecho el contexto de los datos. Por ejemplo, fecha de compra representa una fecha en la que formalmente (legalmente) se realiza la compra. Esta fecha está compuesta por tres valores (día, mes y año).

¿Qué información puede extraer un sistema BI de esta colección de atributos (factura de compra)? Es evidente que mucha. Por ejemplo:

- Día de la semana en el que se producen más ventas.
- Día del mes con más ventas.
- Qué comercial realiza más ventas.
- Qué tipo de modelo es el más vendido.
- Qué opciones prefieren los clientes.
- Qué forma de pago utilizan los clientes.
- De dónde son los clientes de un concesionario dado.

Toda esta información que puede tener su base operacional tiene incidencia en otros departamentos. Los departamentos de ventas/compras, financiero y marketing, entre otros, obtendrán información valiosa del conjunto de esta información.

Como regla general, los sistemas de BI permitirán una serie de soluciones que van en la línea de mejorar los ingresos, como por ejemplo:

- Incremento del número de clientes / cuota de mercado por unidad de tiempo o zona geográfica.
- Evitar pérdida de clientes (incluso en el caso de automoción es importante retener clientes).
- Incrementar ventas (ofrecer mejores opciones con precios más optimizados).
- Incrementar la rotación de vehículos (que estén el menor tiempo posible en el concesionario).
- Analizar el efecto de las campañas publicitarias o descuentos.
- Simular nuevas estrategias de venta/descuentos/promociones.

Otra línea de actuación de los sistemas BI pasaría por la otra cara de los negocios, que es la reducción y/o eliminación de costes:

- Mejora en los RR. HH. (buenos vendedores/trabajadores).
- Identificar reclamaciones y seguimiento de su solución.
- Reducir el tiempo de espera.

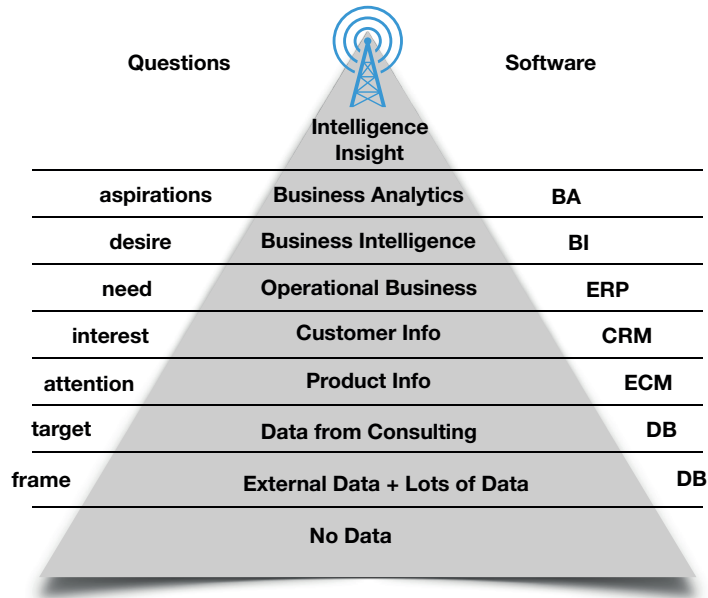
Hay una serie de ventajas que podríamos identificar como intangibles:

- Conocimiento de los hábitos de nuestros clientes / mejora del sistema post-venta.

- Control de la información facilitada y su uso.
- Análisis de la estrategia de precios.

Los niveles de conocimiento asociados a la información pueden visualizarse en la siguiente pirámide del conocimiento (ver figura 2).

Figura 2. Pirámide de los niveles de conocimiento



1.2. Arquitectura BI

La arquitectura más implementada para dar soporte a los procesos de *business intelligence* se basan en el concepto de *data warehouse*, que fue popularizado por Bill Inmon (pero que inicialmente fue desarrollado en IBM por Barry Devlin y Paul Murphy), quien indica que:

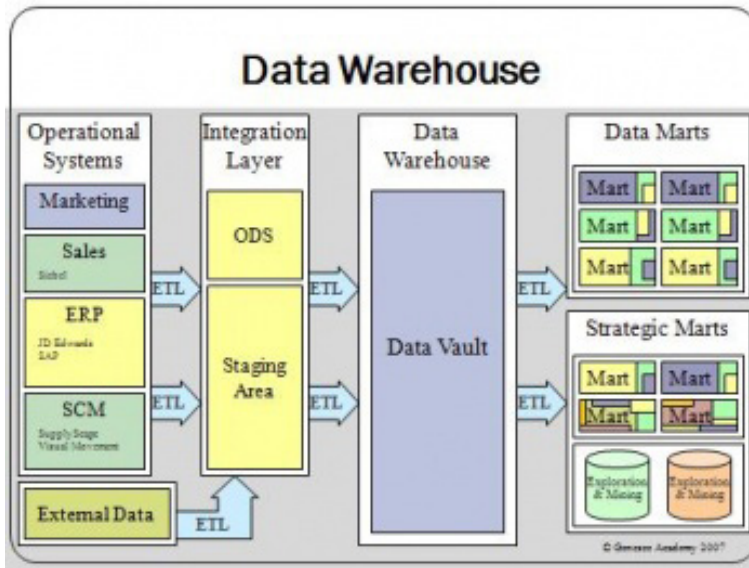
«proporciona una visión corporativa, común e integrada de los datos de la organización, variable en el tiempo, no volátil y que da soporte a la toma de decisiones.»

Los cuatro elementos básicos de la definición de Inmon son los siguientes:

- **Orientado a temas:** Aquellos que los gestores del negocio tratan (producto, venta, región/área).
- **Integrado:** Contiene todos los datos de todos los sistemas operacionales de forma consistente.
- **Variante en el tiempo:** Indica que debemos almacenar la dimensión del tiempo (tener históricos).
- **No volátil:** Los datos están pensados para ser leídos (no escritos o actualizados).

Podemos visualizar las propuestas de algunos fabricantes y observar los elementos comunes.

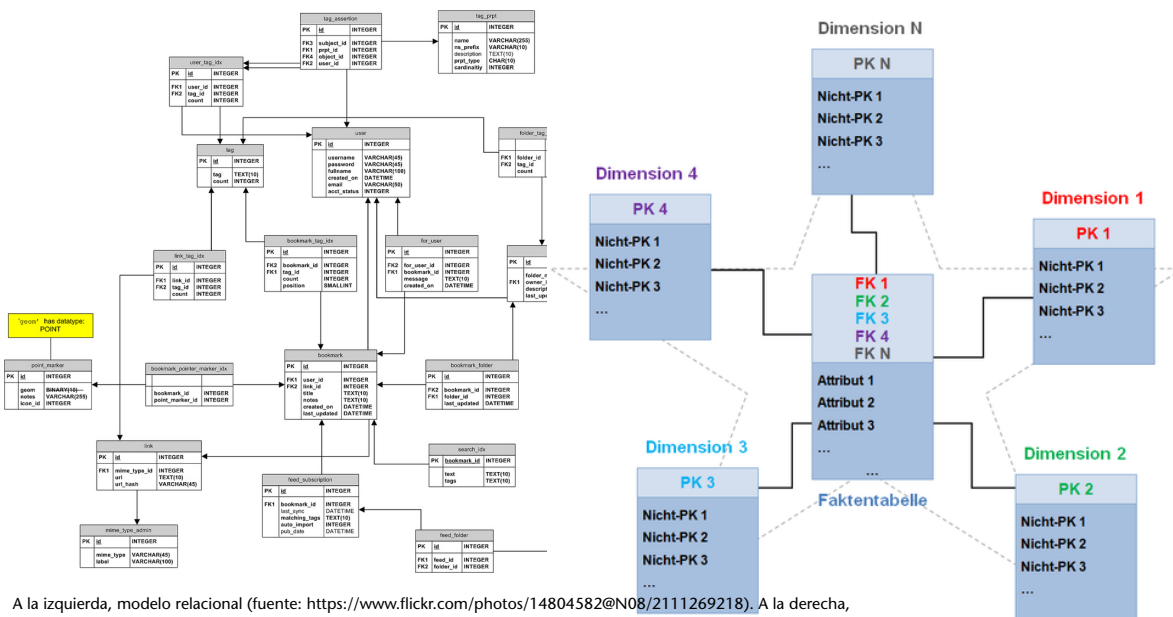
Figura 3. Propuesta de arquitectura *data warehouse*



Fuente: http://www.orafaq.com/wiki/Data_warehouse

Se puede apreciar en la figura 3 que en la parte izquierda del diagrama existe la parte operacional, soportada por un modelo RDBMS. La parte central y la de la derecha son ya elementos que configuran soluciones dirigidas al *business intelligence*. La flechas con las siglas ETL (*extraction, transformation and load*) significan los procesos de extracción, transformación y carga.

Figura 4. El proceso de ETL pretende pasar de un modelo poco comprensible por los responsables del negocio a un modelo donde las tablas tienen una traducción directa en los conceptos que los gestores entienden y gestionan.



A la izquierda, modelo relacional (fuente: <https://www.flickr.com/photos/14804582@N08/2111269218>). A la derecha, modelo estrella (Fuente: <https://de.wikipedia.org/wiki/Sternschema>)

1.3. Modelización del negocio

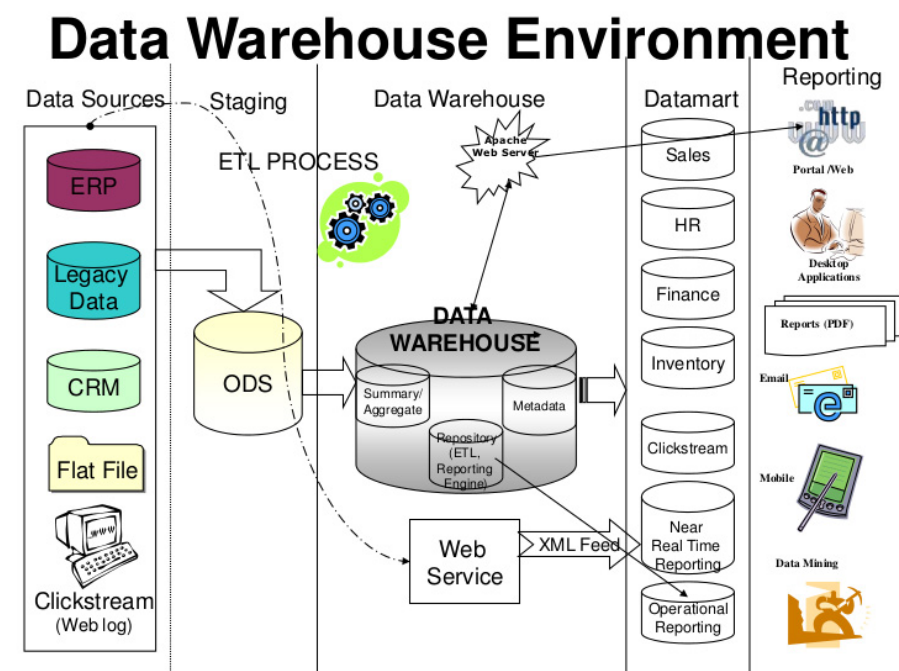
Los sistemas de información tratan de mantener todos los datos operacionales y transaccionales relativos a la operativa del negocio. Tienen, no obstante, un pequeño problema derivado de la complejidad del modelo relacional. El conjunto de tablas que modelan la realidad de la operativa del negocio puede contener del orden de más de doscientas tablas. ¿Qué responsable de negocio puede entender las relaciones entre las tablas, las claves y la integridad referencial que los sistemas RDBMS tienen?

Una solución es modelar los conceptos que los gestores responsables del negocio utilizan. Esto da pie a un nuevo enfoque si somos capaces de entender las preguntas que nos formulamos respecto a nuestro negocio.

Quizá nuestro interés esté en el departamento de ventas y por consiguiente podemos definir una serie de hechos, como la venta y definir medidas KPI (indicadores clave del negocio).

La mayoría de los suministradores de soluciones BI tienen opciones para adaptar el modelo RDBMS a la propia de un DW. En la solución mostrada en la figura 5 se puede observar de nuevo la arquitectura asociada a este tipo de soluciones. En la parte izquierda los datos que nutren al sistema, a continuación las partes relativas a ETL y *data warehouse* con una especialización de la solución a cada unidad de negocio (RRHH, marketing, gestión de *stocks*) y una parte de generación de informes en la parte derecha (*reporting* y visualización).

Figura. 5 Datawarehouse



2. Introducción al *business analytics*

2.1. ¿Qué ofrece el *business analytics*?

Las empresas que ofrecen este tipo de solución utilizan definiciones como la siguiente:

«IBM Analytics ofrece un porfolio completo de soluciones de *big data* y analítica para impulsar su negocio cognitivo. Nuestras soluciones permiten a su organización trabajar con datos para responder a las preguntas de negocio más complicadas, descubrir patrones y poner en práctica ideas innovadoras.»

Este tipo de ofertas es el que vamos a encontrar en la mayoría de los vendedores de software.

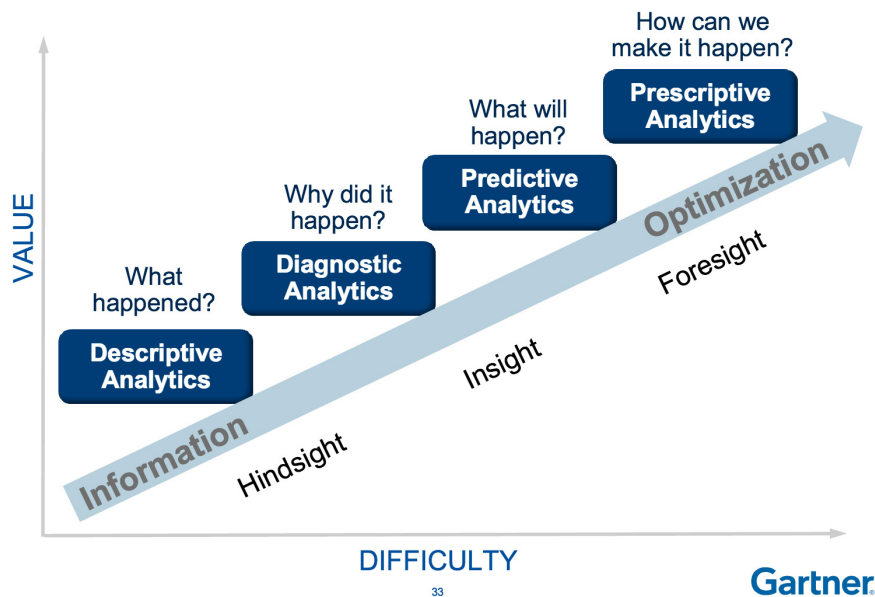
Las herramientas de *business analytics* se caracterizan por:

- una interfaz intuitiva y fácil de usar que se adapte a las necesidades de los usuarios,
- aproximaciones automatizadas al análisis de datos y al uso de modelos complejos,
- algoritmos avanzados de análisis de datos con cierto grado de flexibilidad y autonomía para detectar patrones de interés para el negocio o con capacidad para recomendar visualizaciones,
- facilidad para tratar diferentes fuentes de datos complejos, poderlos integrar o combinar de la manera deseada y sin requerir un esfuerzo o conocimiento sofisticado,
- capacidad para verificar modelos estadísticos sofisticados y complejos de forma iterativa usando algoritmos de *machine learning* y
- facilidad para la interacción y la visualización de datos. Sencillez en la forma de interactuar.

Existe una evolución entre las soluciones de *business intelligence* y las de *business analytics*. La consultora Gartner lo sintetiza en un gráfico (ver figura 6) donde se visualiza toda la vertiente analítica del análisis y su incidencia en el valor que aporta a la organización.

Figura 6. Modelo analítico de la consultora Gartner (<https://www.gartner.com>)

Gartner Analytic Ascendancy Model



Fuente: <https://www.zdnet.com/article/data-to-analytics-to-ai-from-descriptive-to-predictive-analytics/>

La lista de métodos que podemos encontrar dentro de estas soluciones es completa y se ajusta a los últimos avances que se dan en los ámbitos del *machine learning* o del *deep learning*.

¿Qué tipos de métodos ofrecen las soluciones de *business analytics*? Entre los que podemos encontrar se encuentran los siguientes:

- Contraste de hipótesis
- Regresión lineal
- Regresión logística
- Redes neuronales y métodos de *deep learning*
- SVM
- Series temporales
- Sistemas de recomendación
- *Topic modelling* (LDA)
- Árboles de decisión
- Métodos de ensamblaje
- *Clustering*
- Reducción de la dimensionalidad

2.2. Métodos de *business analytics*

La mayoría de los métodos que se ofrecen en las diferentes plataformas exceden por mucho el tiempo y espacio que tenemos destinado a presentar este tipo de soluciones. En cualquier caso, es conveniente entender la motivación y la utilización de este tipo de técnicas.

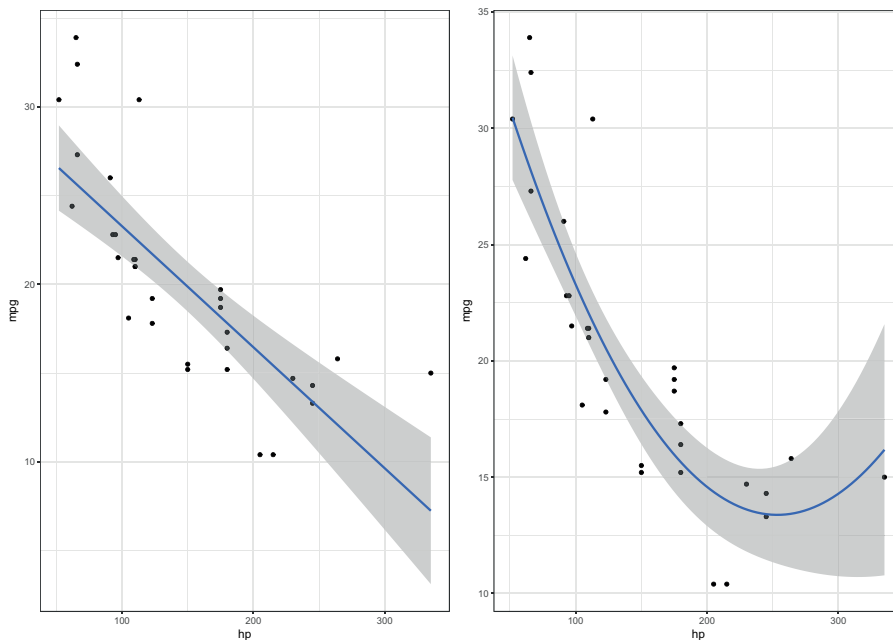
2.2.1. Contraste de hipótesis

La idea es contrastar de manera científica una hipótesis que el responsable del negocio se cuestiona o debe solucionar. Se basa en un procedimiento simple que deriva de la estadística básica.

2.2.2. Regresión lineal

El objetivo de este tipo de técnicas es el de predecir una variable objetivo en función de otra que es explicativa. La relación existente entre ambas puede ser una recta u otra función, como una función cuadrática.

Figura 7. Dos ejemplos de regresión



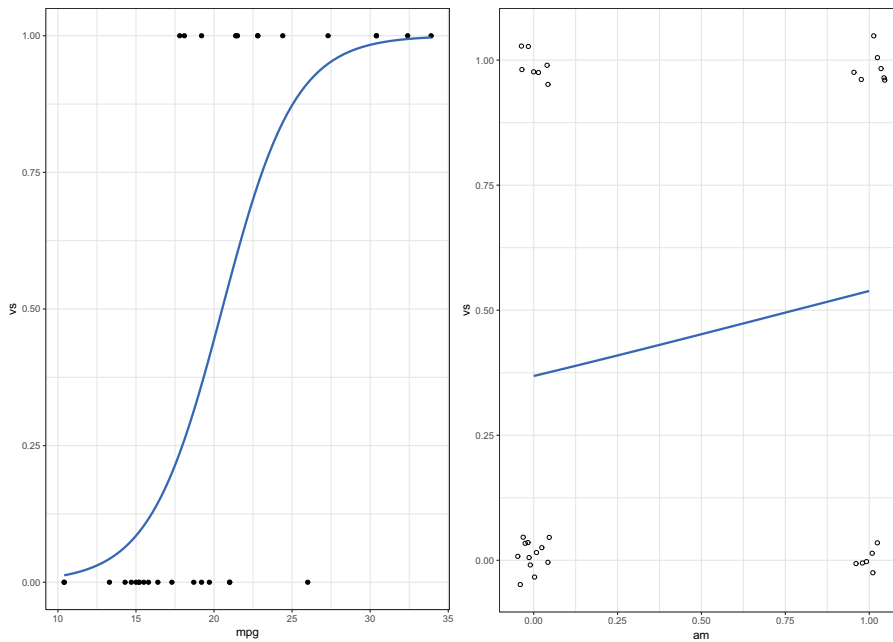
Aclaración

Todas las imágenes mostradas desde la figura 7 hasta la figura 15 son generadas con el software R bajo el IDE de RStudio.

2.2.3. Regresión logística

En este caso el objetivo también es predecir una variable, pero que se interpreta como una probabilidad. En la práctica, entendemos la regresión logística como un método de clasificación. Es decir, nos va a permitir decidir si se da una de las dos clases bajo estudio (sí/no, enfermo/sano, error/correcto).

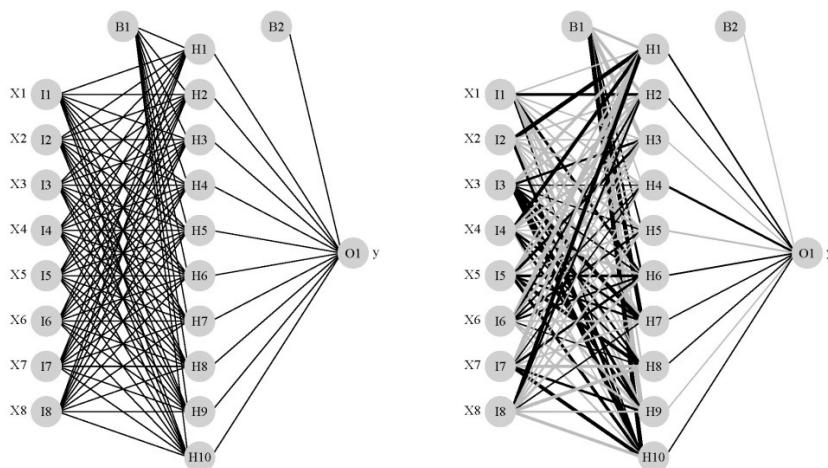
Figura 8. Dos ejemplos de regresión logística, uno con un resultado coherente y otro sin mucho sentido



2.2.4. Redes neuronales y métodos de *deep learning*

Estas son dos propuestas que van en la misma dirección en el sentido de que solucionan problemas de aprendizaje supervisado. Pueden resolver problemas de clasificación. Es decir, asignar una etiqueta (correspondiente a la variable categórica) a partir de una entrada de variables que actúan de variables explicativas o bien predecir un valor numérico (problema de regresión).

Figura 9. Arquitectura de una red neuronal



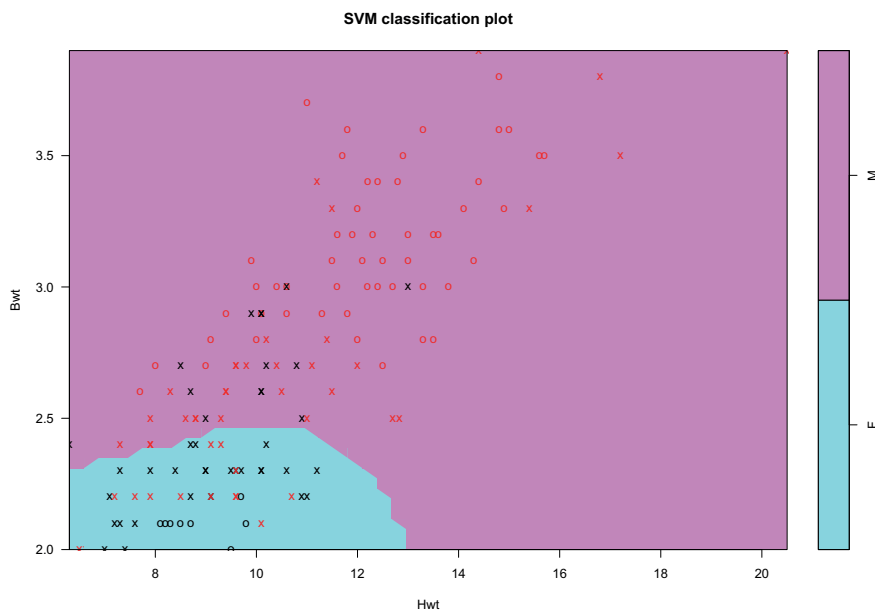
Los modelos de *deep learning* tienen una complejidad superior por el hecho de utilizar en sus arquitecturas un número muy superior de capas intermedias (las que encontramos entre la entrada y la salida y las que definen el grado de complejidad de la arquitectura de la red). Este tipo de métodos complejos son

muy utilizados gracias al uso de librerías como la de TensorFlow de Google y la API de keras.

2.2.5. SVM

Este conjunto de métodos denominados *support vector machines* (máquinas de vectores de soporte) se basan en un proceso de optimización que intenta maximizar el margen entre la línea (plano o hiperplano) de división de dos grupos de observaciones. Este tipo de aproximaciones ofrece una gran flexibilidad gracias a la parametrización del método. El uso de diferentes funciones *kernel* ayuda a este objetivo.

Figura 10. Ejemplo de clasificación utilizando SVM



2.2.6. Series temporales

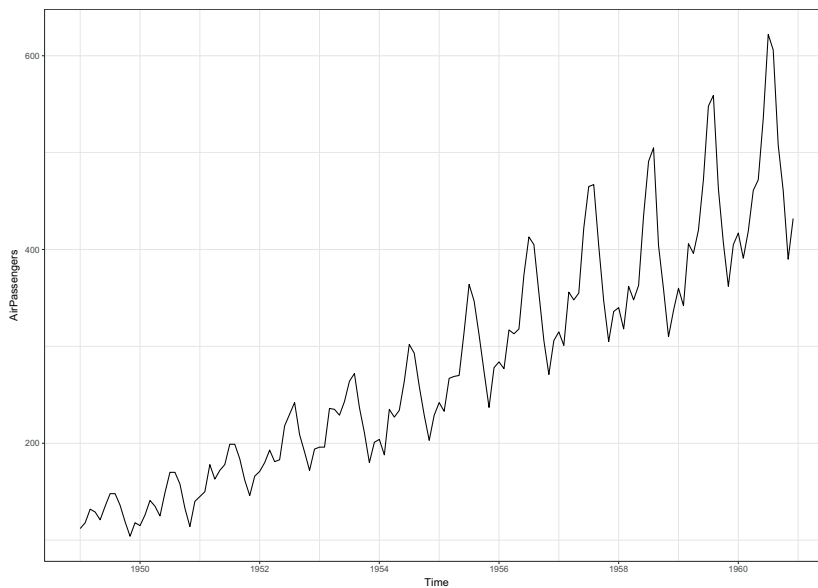
Este es un conjunto de métodos sofisticados que son importantes en los modelos predictivos donde tengamos como elemento clave la unidad de tiempo.

Existe una gran variedad de técnicas para tratar series temporales, unas muy descriptivas y otras más sofisticadas y basadas en modelos de procesos estocásticos. Entre estos modelos son utilizados los siguientes:

- AR
- MA
- ARMA
- ARIMA
- SARIMA

y también modelos de volatilidad (ARCH, GARCH).

Figura 11. Ejemplo de serie temporal que persigue predecir el número de pasajeros para el próximo año



2.2.7. Sistemas de recomendación

Tienen una gran importancia por su aplicación en muchos campos. Es un elemento estratégico para muchas empresas actuales, como Netflix, Amazon y Facebook, entre otras.

Matemáticamente, los problemas de recomendación se pueden asimilar a problemas de tratamiento de matrices esparsas (con muchos ceros) y el objetivo es predecir estos valores en función de unos criterios (estableciendo diferentes métodos).

Este tipo de soluciones es muy habitual en lo que se denomina inteligencia colectiva, que está muy ligada al auge de las redes sociales.

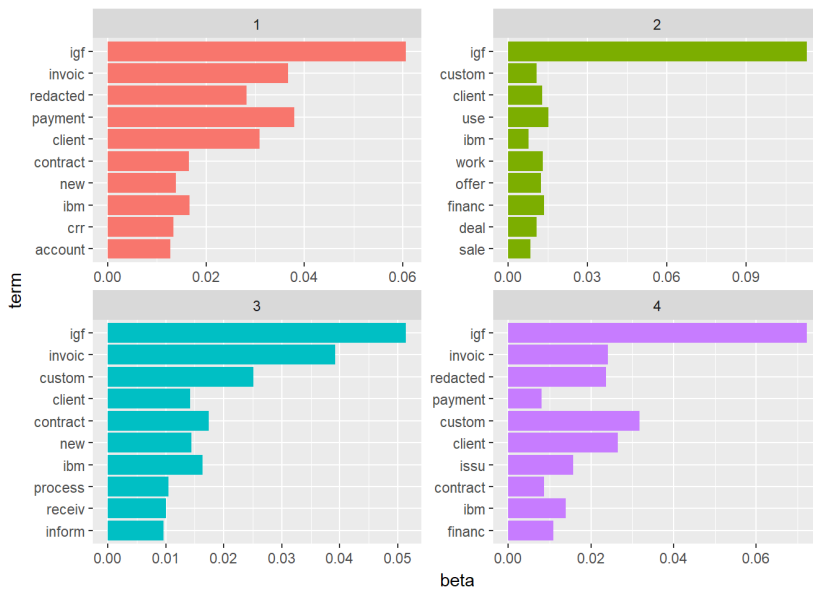
2.2.8. Topic modelling (LDA)

Este es otro tema importante que en los últimos años ha adquirido relevancia por verse impulsado por los avances en el procesamiento de lenguaje natural (NLP). El análisis de los textos permite identificar una serie de características en principio no aparentes y que esta aproximación pone al descubierto.

2.2.9. Árboles de decisión

Este tipo de modelo se puede adaptar a problemas de regresión (dan como resultado un valor numérico) o a un problema de clasificación (da como resultado una etiqueta). Una de las características apreciadas de este tipo de técnicas

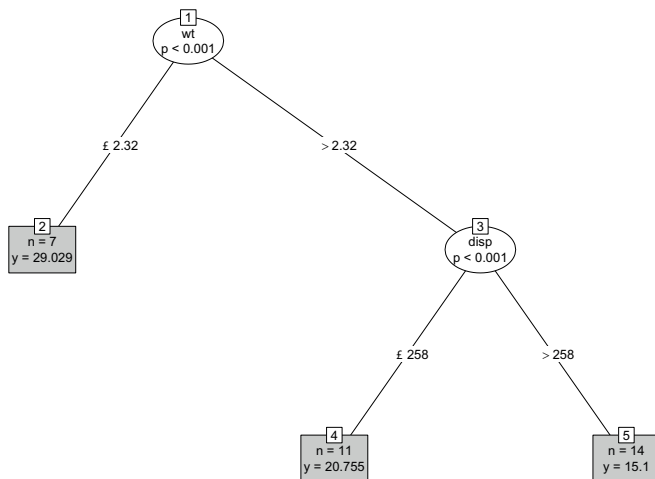
Figura 12. Ejemplo de modelización LDA con cuatro términos



(muy clásicas) es su facilidad para ser interpretadas por cualquier usuario. La visualización resultante del modelo ayuda a hacer uso de este tipo de modelos.

No obstante, debemos recalcar que este tipo de técnica no es muy recomendable en problemas con gran volumen de datos.

Figura 13. Ejemplo de árbol de decisión para el caso de predecir consumo (en miles per gallon)



2.2.10. Métodos de ensamblaje

Permiten ensamblar diferentes modelos para mejorar la capacidad predictiva. Las estrategias más utilizadas dentro de este grupo de técnicas son las siguientes:

- **Bagging**: Construye múltiples modelos (normalmente del mismo tipo) de diferentes muestras del conjunto de datos de entrenamiento.
- **Boosting**: Construye múltiples modelos (normalmente del mismo tipo), cada uno de los cuales aprende a corregir los errores de predicción del modelo anterior del conjunto de modelos.
- **Stacking**: Construye múltiples modelos (normalmente de diferente tipo) y un modelo supervisor que aprende cómo es mejor combinar los modelos restantes.

2.2.11. Clustering

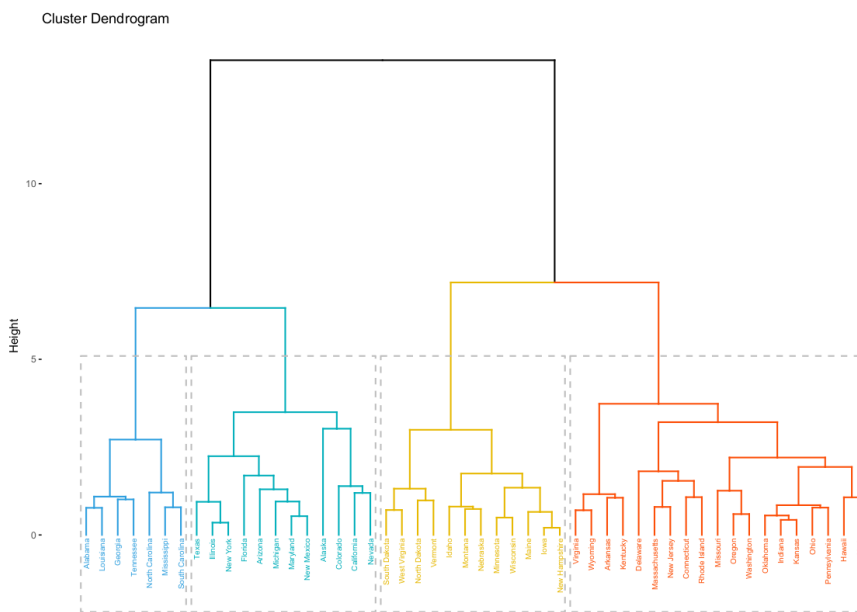
Los métodos de *clustering* son el principal ejemplo de métodos de aprendizaje no supervisado. En general, no hay ningún atributo en la base de datos que actué como *target* y, por tanto, estos métodos tienen como objetivo agrupar. Estas agrupaciones persiguen disminuir la varianza dentro de los grupos y maximizar la varianza fuera de ellos.

Figura 14. Ejemplo de agrupación con técnicas de *clustering*



Existe una gran variedad de aproximaciones a los métodos de *clustering*. Entre estos, podemos citar:

- *Partitioning methods*
- *Hierarchical clustering*
- *Fuzzy clustering*
- *Density-based clustering*
- *Model-based clustering*

Figura 15. Ejemplo de agrupación con técnicas de *clustering*

2.2.12. Reducción de la dimensionalidad

Los métodos de reducción de la dimensionalidad de los datos intentan conservar su naturaleza intrínseca pero reduciendo el número de variables presentes. Existen diferentes estrategias para reducir el número de dimensiones de nuestro conjunto de datos.

De entre las diferentes propuestas existentes, a continuación detallamos algunas de las más utilizadas:

- PCA (*principal component analysis*)
- ICA (*independant component analysis*)
- T-SNE (*t-distributed stochastic neighbor embedding*)
- Autoencoder (utilizado principalmente en modelos de *deep learning*)
- DCT (*discrete cosine transform*)

2.3. Arquitectura BA

La descripción del modelo propuesto por la compañía IBM recoge los elementos descritos anteriormente. Si bien la propuesta es una evolución de los entornos BI, esta propuesta se caracteriza por:

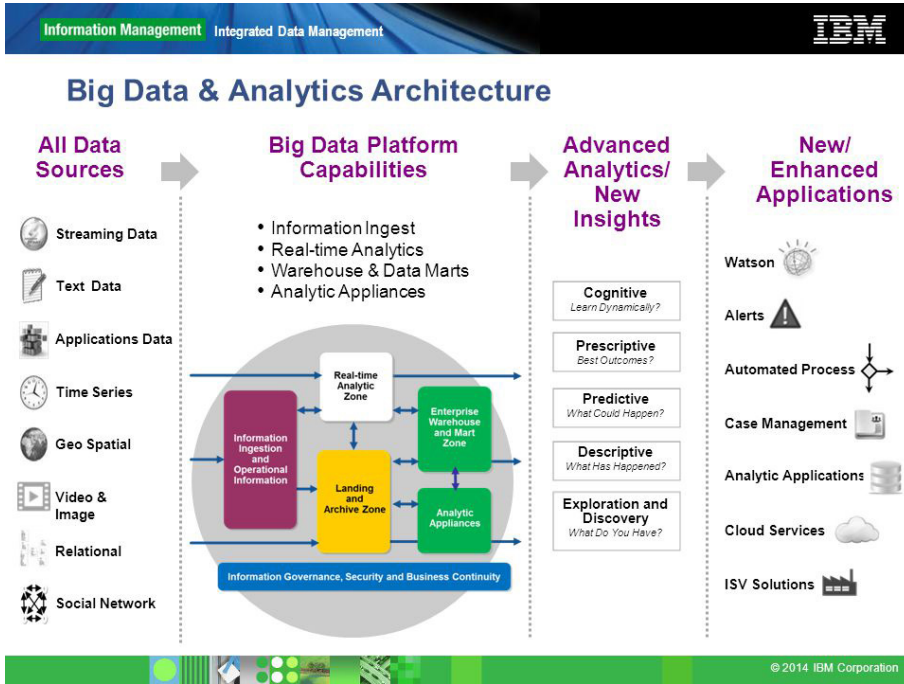
- Permitir el análisis de cualquier tipo de dato. Es decir, podemos analizar y tratar datos en *streaming*, texto, de aplicaciones, de series temporales, información geoespacial, vídeo e imágenes y redes sociales, entre otros.

Ved también

Las arquitecturas que dan soporte a las funciones propias del *business analytics* son las que se describen en el módulo "Introducción al *big data*".

- Disponer de plataforma de soporte, que ingiere toda la información y permite que se pueda procesar y utilizar convenientemente por herramientas analíticas avanzadas y de descubrimiento de nuevo conocimiento.
- Dar soporte a un conjunto de aplicaciones.

Figura 16. Propuesta de arquitectura *business analytics* de IBM Fuente: www.ibm.com



Ved también

Los nuevos entornos de *big data* se trata en el módulo "Introducción al *big data*" de esta asignatura.

Enlace de interés

Para complementar la información que hemos tratado aquí, podéis consultar el siguiente enlace:
<https://ibm.co/2WTxGSz>

La realidad de las últimas propuestas de arquitecturas BA se centra en los nuevos entornos de *big data*.

3. Soluciones en el entorno industrial

En el entorno industrial existe un interés evidente por integrar y fomentar todas estas tecnologías para tener una ventaja competitiva en el mercado. Podemos pensar en una empresa como GE (General Electric), fundada en 1892 y por tanto con unas inercias que podríamos catalogar de contrarias o con resistencia a cambios radicales. De hecho, GE ha impulsado una sensorización de sus motores que, junto con modelos físicos, algoritmos de inteligencia artificial y analítica de datos, construye uno o diversos modelos que permiten:

- ahorrar costes,
- mejorar la seguridad de los clientes,
- evitar roturas innecesarias y
- detectar y corregir funcionamientos defectuosos o peligrosos.

La capacidad que este tipo de solución ofrece a GE es incalculable. General Electric puede averiguar el desgaste de sus motores ya sea en aeronaves, en locomotoras, en turbinas de gas o en turbinas eólicas utilizando datos de sensores en lugar de suposiciones o estimaciones. Esta aproximación permite predecir cuándo necesitarán mantenimiento o si hay peligro inminente de rotura.

La solución va un paso más allá de la previsión de la esperanza de vida de la máquina, ya que puede optimizar su funcionamiento y mejorar el rendimiento.

Según datos de GE, estas soluciones permiten:

- Un incremento del 20 % de la producción eléctrica en parques eólicos.
- Una reducción anual del consumo de gasolina y de las emisiones de carbono en una de sus locomotoras en más de 120.000 litros y 170.000 toneladas al año respectivamente.

Esta tecnología depende de algoritmos de inteligencia artificial que continuamente se están actualizando. Si los datos obtenidos se identifican como erróneos, corruptos y que faltan, entonces una batería de test de *machine learning* resuelven esta disyuntiva de manera automática y sin la intervención humana.

La utilización de herramientas sofisticadas de *machine learning*, de *deep learning* y de *reinforcement learning* ha permitido una gestión de los sistemas de energía mucho más eficiente (como el que se aprecia en la fotografía de la figura 17).

Figura 17. Turbina de gas



Fuente: http://www.publicdomainfiles.com/show_file.php?id=14019791416948

3.1. BI en el entorno industrial

En el ámbito industrial hay una gran lista de ejemplos de utilización de soluciones BI. No obstante, en la actualidad la mayoría de las soluciones que se implementan van encaminadas hacia el ámbito analítico. Se da también la situación de que empresas que ofrecen soluciones de BI en realidad están ofreciendo una solución BA.

Comprobad por ejemplo la publicidad de la empresa Sisense de consultoría y soluciones BI: <https://www.sisense.com/solutions/> y también: <https://bit.ly/2tgwUS6>

3.2. Business analytics en el entorno industrial

La industria del *business analytics* (BA) ha crecido de modo exponencial en los últimos años. Los rápidos avances en áreas como el *big data*, en el aprendizaje automático o en inteligencia artificial (IA) están impulsando una nueva era de inteligencia empresarial que incide de modo directo en la industria de manufacturación.

Las empresas de fabricación están entre las industrias más implicadas en la generación intensiva en datos (sensorización) que existen en la actualidad. La naturaleza compleja de estas organizaciones, que incluyen procesos que van desde la contratación, la gestión y mejora de la cadena de suministro hasta la organización de pedidos (de los clientes en sus centros de producción) y la

entrega puntual de los pedidos, hace que la fabricación sea un proceso muy intensivo en datos.

Dada la gran cantidad de datos procedentes de todos los procesos involucrados en la fabricación, existen muchas oportunidades para la mejora en diversas áreas de interés:

- eficiencia en la producción,
- flexibilidad de los procesos,
- visibilidad y monitorización de los procesos,
- evaluación de la monitorización y
- mantenimiento predictivo.

Y ello gracias no tanto a la gran cantidad de datos que se generan, como a lo que hacemos con ellos gracias a las herramientas utilizadas para extraer este conocimiento.

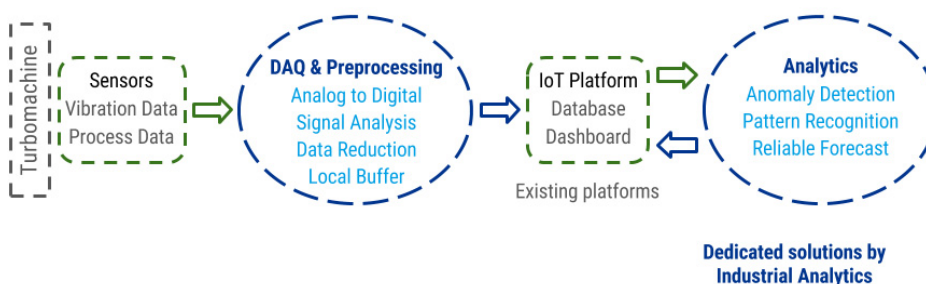
3.3. Proveedores de soluciones BA

Los actores más activos en cuanto a ofrecer soluciones de *business analytics* y los que tienen mayor cuota de mercado (tanto nacional como internacional) son los siguientes:

- IBM
- TABLEAU
- QLIKVIEW
- SAP
- SAS
- MicroSoft
- MicroStrategy
- ORACLE

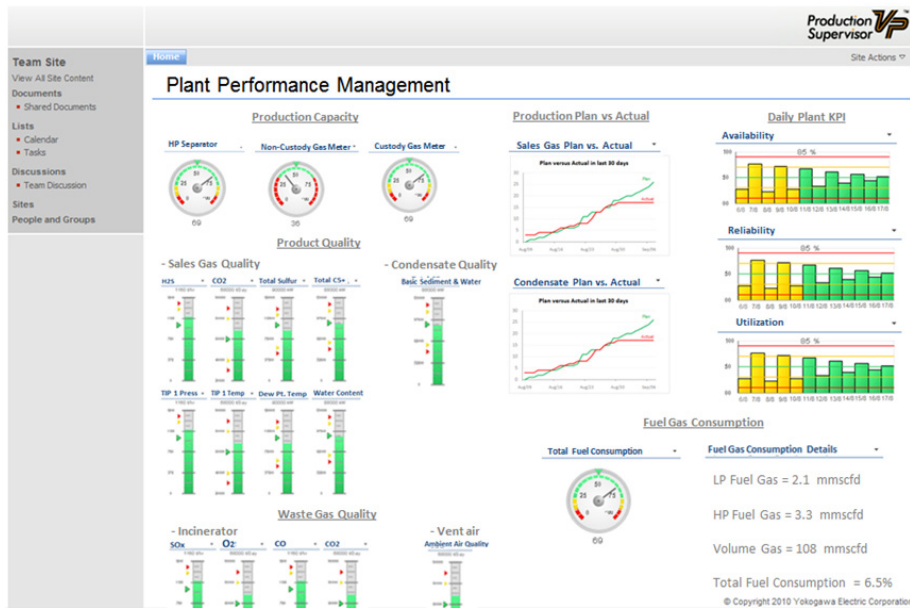
Los fabricantes perciben el análisis de datos como un elemento clave en la implantación exitosa de una estrategia IOT en sus operaciones.

Figura 18. Solución propuesta por una consultora industrial



Las soluciones con una orientación más visual suelen tener mayor aceptación por parte de los clientes. En el siguiente diagrama se puede observar un ejemplo de *dashboard*. Como consecuencia de que los procesos industriales se vuelven más complejos y las demandas de los clientes cambian rápidamente y sin previo aviso, los gerentes no pueden esperar los KPI mensuales y, además, necesitan unos KPI diarios para reaccionar más rápidamente. La solución de gestión de la producción de Yokogawa ofrece este *performance dashboard* que se puede visualizar en la figura 19.

Figura 19. Solución propuesta por una consultora industrial



Fuente: <https://www.yokogawa.com/solutions/solutions/production-management/performance-dashboard/>

4. Toma de decisiones basada en datos

4.1. Evolución de las organizaciones respecto a los datos

Las empresas han tenido una evolución lenta respecto a la rápida implantación de sistemas tecnológicos. Si nos centramos en el concepto de los datos, podemos ver una evolución y diferenciar diferentes estadios:

- Organización resistente a los datos.
- Organización consciente de los datos.
- Organización guiada por los datos.
- Organización conocedora de los datos.
- Organización dirigida por los datos.

4.1.1. Organización resistente a los datos

La tipología de estas organizaciones se resume en una aversión a realizar cualquier cambio en el modo de operar. Respuestas de responsables a consultores/vendedores de soluciones tecnológicas son "siempre lo hemos hecho así", o bien "no hay necesidad de cambiar nada, todo funciona perfectamente". Pero la verdadera razón es que hay miedo. Miedo a que los datos revelen ciertas premisas que han estado permanentemente en la organización. Los datos podrían...

- descubrir problemas de rendimiento no conocidos,
- visualizar contribuciones del personal y proyectar una mala imagen de la empresa,
- socavar el mensaje de la marca y
- mostrar que la estrategia no está alineada.

Es responsabilidad del personal de la empresa cambiar esta actitud frente a los datos. Cuanta más responsabilidad tenga uno en la organización, más fácil será su efecto hacia el cambio.

4.1.2. Organización consciente de los datos

Esta situación describe el caso en el que la organización sabe de la existencia de datos dentro de su marco operativo. También es conocedora de la importancia

de tener los datos y el valor de estos. Este enfoque facilita que se centre en el almacenamiento de datos que vienen suministrados por la operativa del negocio.

En muchas ocasiones la organización no es capaz de ver el potencial de los datos, a no ser que disponga de alguna herramienta suministrada por un proveedor tecnológico. Estas soluciones cubren diferentes áreas funcionales de la organización, y entre ellas podemos encontrar:

- gestión de clientes (CRM),
- gestión empresarial (ERP) y
- contabilidad y planificación financiera.

La transición de una organización resistente a los datos a una consciente viene dada principalmente por la necesidad de sacar provecho a los datos recopilados.

4.1.3. Organización guiada por los datos

El objetivo de este tipo de organización es sacar el máximo provecho de los datos y darles valor. Es obvio que el elemento clave es el análisis, y ello implica que los datos se ponen en producción. Este tipo de aproximación se sustenta con algunas soluciones tecnológicas como las que hemos visto anteriormente:

- *datawarehousing* (ETL) y
- *cloud computing*.

Bajo esta premisa las organizaciones desbloquean el valor táctico de los datos. Los datos entran en producción, empiezan a hablar y a aportar más información.

4.1.4. Organización conocedora de los datos

En este ámbito las organizaciones añaden a los datos un elemento adicional. Los datos son un elemento estratégico y, por consiguiente, son valorados como un capital que va más allá del enfoque táctico. No basta con resolver cuestiones relativas al qué, sino al porqué. Estas preguntas más complejas ofrecen una nueva capa de percepciones antes no disponibles.

- ¿Por qué las ventas bajaron el primer semestre?
- ¿Por qué los clientes no compran el producto principal de nuestro catálogo?
- ¿Por qué los usuarios hicieron un pico de reclamaciones/preguntas el segundo trimestre?

- ¿Por qué la campaña publicitaria funcionó?
- ¿Por qué la zona 1 supera en porcentaje las ventas de la zona 5?
- ¿Por qué las previsiones de venta no se han cumplido?

4.1.5. Organización dirigida por los datos

Este enfoque caracteriza todo el espectro empresarial conocedor de la capacidad transformadora de los datos y el análisis de estos. El objetivo que se persigue es convertir este conjunto de datos y análisis en valor tangible y estratégico para la empresa.

Las empresas son capaces de crear y capturar datos que en el contexto adecuado permiten transformar el negocio para hacerlo más flexible y próximo a los clientes (internos y externos). La representación distintiva de lo que los datos y la analítica son capaces de ofrecer define el panorama de las empresas punteras en la actualidad. El binomio *data* y *analytics* es el denominador común de estas organizaciones.

Obviamente, la tecnología es un elemento clave, ya que nos ha permitido:

- un incremento en la potencia de cálculo,
- la computación distribuida + nube y
- la reducción de costes en todos los niveles (hw + sw).

4.1.6. Caso Challenger

El accidente del transbordador espacial Challenger (misión STS-51-L) se produjo el martes 28 de enero de 1986, justo después de 73 segundos de vuelo (lanzamiento número 25). Inmediatamente el Gobierno creó una comisión para averiguar las causas de este grave incidente.

La desintegración del vehículo entero comenzó después de que una junta tórica (*O-ring*: encargada de sellar y prevenir la liberación de gases durante la combustión) de su cohete acelerador sólido (SRB) derecho fallara durante el despegue. El fallo de la junta tórica causó la apertura de una brecha, que permitió que el gas caliente presurizado del interior del motor del cohete sólido saliera al exterior y contactara con la estructura adyacente, lo que provocó el resultado fatídico.

Cada propulsor consta de tres anillos con un total de seis en el transbordador. Como se disponía de los datos de los anteriores vuelos (23), se analizaron estos datos con la intención de averiguar si existía relación entre la temperatura de despegue y el fallo de los anillos.

Evolución de las organizaciones respecto a los datos

Existen varias aproximaciones y referencias relativas a esta exposición. Recomendamos ver:

<https://ibm.co/2l8CtAg>
<https://mck.co/2JKQKif>
<https://bit.ly/2THDVXQ>
<https://bit.ly/2x9ZHzy>

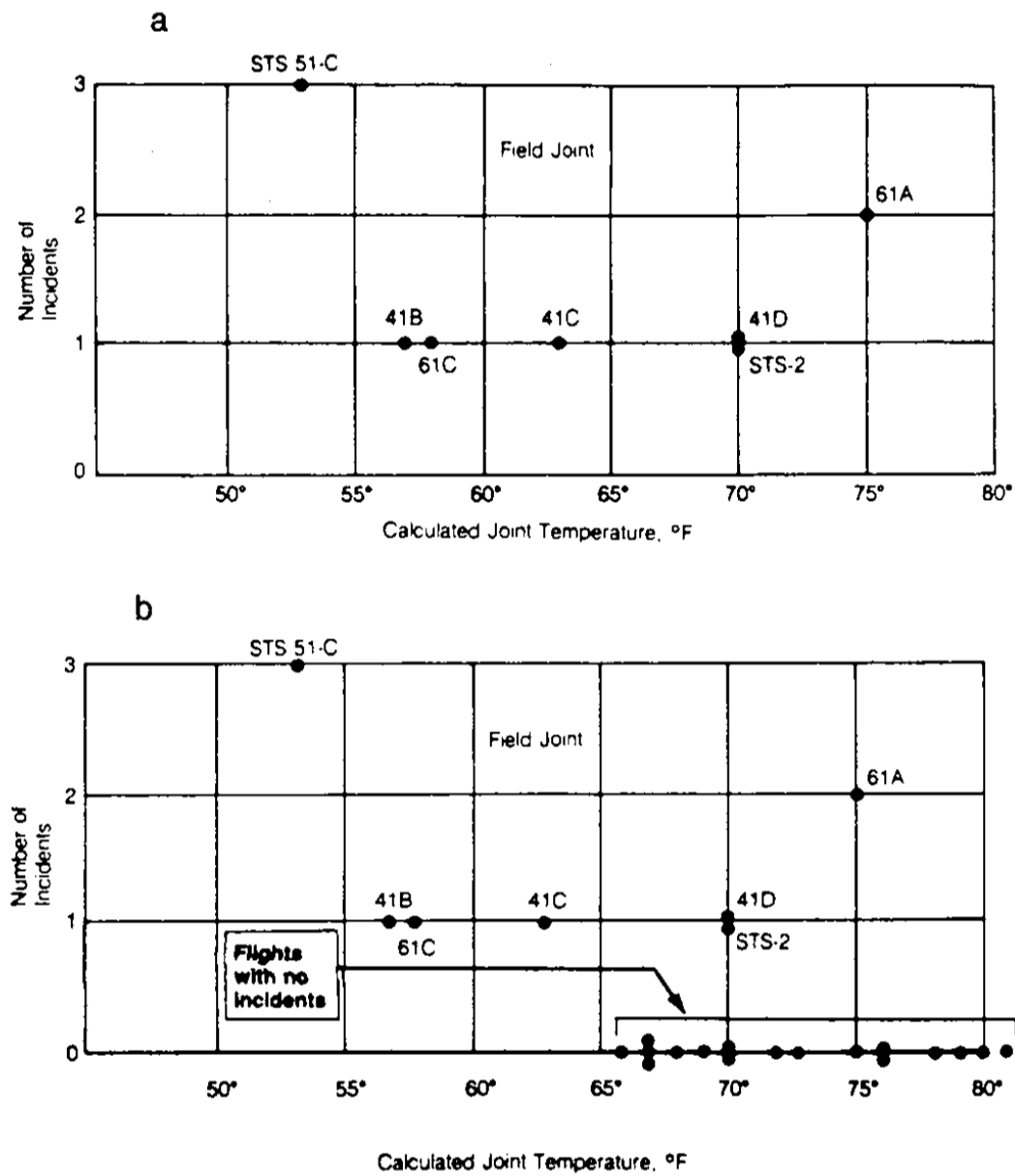


Explosión del transbordador espacial Challenger
 Fuente: <https://bit.ly/2USWKHH>

La realidad es que los ingenieros analizaron los datos donde había problemas con alguno de los anillos. En total, seis vuelos. ¿Qué sucede si analizamos estos datos? Realmente no parece que veamos nada. Es decir, que la temperatura parece no tener incidencia en la presencia de fallo en los anillos. Lo curioso del caso es que los ingenieros de la NASA no tuvieron en cuenta la información precedente de los vuelos donde no hubo ningún problema (como si estos datos no fueran relevantes).

¿Cómo es posible que una organización como la NASA no analizara los datos convenientemente? Lo cierto es que en las comisiones iniciales se criticó que no había ningún experto en estadística. Veamos los gráficos que se manejaron en el caso del Challenger.

Figura 20. Representaciones gráficas de los datos bajo estudio



Fuente: <https://priceconomics.com/the-space-shuttle-challenger-explosion-and-the-o/> que provienen de <https://history.nasa.gov/rogersrep/v1p146.htm>

La conclusión que se debería extraer de esta situación es que en ocasiones no es suficiente con tener los datos, ni tampoco con visualizarlos si esto no se traduce en resaltar el aspecto importante del problema. Los datos correctamente analizados constatan una relación entre la temperatura y la probabilidad de fallo en las estructuras de anillo.

Resumen

Las partes que debemos interiorizar por su importancia son las siguientes:

- 1) Los entornos de BI son entornos que han sido ejemplo de sistemas de ayuda a la toma de decisiones. Estos cubren necesidades de diversas organizaciones.
- 2) Destaca la adaptación de los sistemas RDBMS a sistemas *data warehouse* para alinearse con las necesidades reales de la empresa.
- 3) En la actualidad, las empresas necesitan ser proactivas y no reactivas. Las soluciones que provienen de entornos BA facilitan este enfoque.
- 4) Cualquier solución BI o BA da un peso importante a la visualización. La importancia de que los datos cuenten una historia.
- 5) Los métodos detrás de las soluciones BA tienen una importancia e impacto en la mejora de la extracción de conocimiento de nuestras fuentes de datos.
- 6) No necesitamos ser expertos en BA, sino en cómo utilizarlo para ayudar a ser más competitivos dentro de nuestra empresa.
- 7) Dejar que las decisiones sean dirigidas por razones objetivas, procedentes del correcto análisis de los datos (dejar a un lado las sensaciones).

Este módulo pone también de manifiesto la evolución de las soluciones tecnológicas en un plazo de tiempo breve.

