



Separando las aguas

Business Intelligence, Business Analytics y Big Data

Todo tiene que ver con todo, pero no del todo¹

Por Mauricio Giacometti – Instituto CPE

Business Intelligence, Business Analytics y Big Data son tres conceptos que están relacionados con una forma de gestión basada en evidencia, una manera de tomar decisiones de manera informada. No se trata de eliminar otros inputs (como la experiencia y la intuición) para la toma de decisiones, sino de complementarlas con el análisis de los datos que tenemos disponibles, para poder tomar decisiones de mayor calidad, y aumentar la probabilidad de arribar a soluciones exitosas y rentables.

Los 3 conceptos tienen en común que intentan generar valor a partir de los datos: de obtenerlos, de manipularlos, de limpiarlos, de organizarlos, de ordenarlos y vincularlos. Además, veremos cómo en la utilización de dos de estos 3 conceptos mencionados, podremos detectar patrones que permitan tomar mejores decisiones.

Business Intelligence, Business Analytics y Big Data son un conjunto de técnicas, habilidades y tecnología que permiten la generación de valor dentro de una organización. Dicho valor se genera mediante la creación de productos o la mejora de procesos internos que conducen a un progreso considerable en la eficiencia total de la empresa, que podremos verla reflejada en los procesos en áreas operativas, hasta en el modo de ver la empresa como generadora de valor para la sociedad.

Esto se hace evidente ya que las 5 empresas de mayor valor (medida como capitalización bursátil) son “Data driven companies”. Esto quiere decir que su núcleo es la generación y el uso eficiente de los datos como forma de diferenciarse en el mercado y crear valor. A su vez, el 2017 fue el 2º año consecutivo que “Data Scientist” es elegido como el mejor trabajo del año en Estados Unidos², sin mencionar que varias consultoras internacionales han proyectado que para 2020 habrá una escasez considerable de este tipo de profesionales².

Ahora bien, vemos que estos conceptos generalmente están estrechamente vinculados, al punto de que no logramos diferenciarlos e incluso en algunos casos vemos que se tratan como sinónimos.

Si bien es cierto que están vinculados y que incluso hay áreas grises en donde es difícil distinguir a cuál concepto pertenece cierta técnica o a cuál área pertenece un software específico, debemos hacer un esfuerzo por delimitar estas áreas del conocimiento de manera de poder comenzar a entenderlas, y en un futuro incorporarlas en nuestras organizaciones.

¹ M. Yela

² <https://www.forbes.com/sites/louiscolombus/2017/05/13/ibm-predicts-demand-for-data-scientists-will-soar-28-by-2020/#64eb92a97e3b>

Un orden cronológico situaría en primer lugar al concepto de Business Intelligence, el cual responde a la necesidad de consolidar, dar coherencia a los datos de diferentes fuentes (sistemas de gestión, CRM, email marketing, página web, telefonía, entre otros) y sumarizar (contar, sumar, promediar, etc) dichos datos para dar solución a la función de reporting corporativo, principalmente. Este abordaje implica lidiar con la unificación de diferentes tipos de fuentes (bases de datos relacionales, archivos planos, planillas electrónicas, entre otros) y trae aparejados temas técnicos y de negocio, dado que un mismo dato (ejemplo: fecha de nacimiento) puede estar presente en distintas fuentes con valores diferentes. Otro ejemplo sería que se deben recodificar ciertas variables categóricas (como pasar de departamentos a regiones, asociadas a ingresos promedio de sus habitantes). La decisión de qué y cómo medir es responsabilidad del área negocios y no del área IT.

Esta tecnología es altamente rígida ya que la información se actualiza a una frecuencia predefinida (típicamente, una vez por mes). Además, el perfil del profesional asociado con este tipo de tecnología es típicamente el de un desarrollador especializado en herramientas de BI y los cambios requeridos por el área negocios (calcular una nueva métrica) pueden llegar a tardar semanas en ser implementados, ni que hablar de los errores ocasionados por problemas de comunicación. Otro punto para considerar es que este tipo de proyectos tiene en promedio, una tasa de éxito de alrededor del 20 - 30%³.

Una vez descritas las principales características de Business Intelligence, pasaremos a comentar las particularidades de Business Analytics: en primer lugar, es de orden mencionar las 2 grandes áreas de Analytics: analítica avanzada y analítica básica. La analítica avanzada refiere a métodos que permiten encontrar regularidades o patrones en los datos de manera automática o semi automática y la analítica básica está muy relacionada con Business Intelligence ya que tiene, en principio, los mismos objetivos que ésta. Lo que diferencia a estas 2 áreas del conocimiento, es la base tecnológica que las soporta y que permite que el uso de las mismas sea radicalmente diferente. En Business Intelligence, los datos están sumariados a priori y guardados en disco. Estas tareas son realizadas por profesionales de IT y la edición de los criterios de cómo se obtiene esta información son difíciles de cambiar y pueden ejecutarse en un tiempo considerable (generalmente semanas o meses, dependiendo de la profundidad del cambio). Cuando hablamos de Business Analytics, nos referimos a una tecnología más ágil, ya que los cálculos se ejecutan en memoria RAM, diferenciándose con la tecnología usada por Business Intelligence. En primer lugar, la creación y los cambios de métricas y dimensiones (regiones, rangos de edades, unidades de negocio y en general cualquier variable categórica) pueden definirse y ejecutarse en el momento, lo que es conocido en el mercado como “responder las preguntas en el momento en que se realizan”. Sumado a esto, estas métricas y dimensiones pueden definirse a nivel de usuario final y ejecutarse con pocos conocimientos de informática (esto se conoce como “self service analytics”). En segundo lugar, podemos trabajar con la totalidad de los datos y no con un subconjunto agregado de los mismos (en BI se tiene la facturación por Año-Mes-Unidad de negocio-Sucursal por ejemplo), lo que permite llegar a ver los datos con el mayor nivel de detalle posible (el mismo que vería un analista realizando consultas a la base de datos relacional del software de gestión). Esto último permite evaluar aspectos que con BI serían directamente imposibles ya que esta tecnología solamente permite ver valores agregados. Por ejemplo, si una factura fue ingresada con un monto 10 veces superior por error en el ingreso de

³ <https://www.cio.com/article/3221430/business-intelligence/4-reasons-most-companies-fail-at-business-intelligence.html>

datos (se ingresó con un cero de más), con BI veríamos que en determinado mes, la facturación es relativamente alta y eso no condice con nuestra percepción. A partir de esto, comenzaremos a investigar y veremos que para una unidad de negocio, la facturación es excepcionalmente alta (la que pertenece el producto de la factura mal ingresada), pero no sabremos cuál es la factura que fue ingresada con errores. Además de la capacidad de reporting, esta funcionalidad permite la de auditoría y la de evaluación de calidad de datos (en su dimensión precisión) que nos puede ahorrar una cuantía considerable en términos de prevención de fraudes internos, entre otros fines.

Para terminar la comparativa entre estos 2 conceptos, me gustaría subrayar que una vez que Analytics aparece en escena, BI se transforma mayoritariamente en una tecnología con objetivos de reporting que es de naturaleza estática, mientras que Analytics toma el papel de inquisidor y generador de valor para la organización.

Reporting está vinculado con la medición y disponibilización de métricas definidas a nivel de la industria (no contemplando las características específicas de la organización que la implementa). Éstas no suponen gran complejidad y son seguidas a largo plazo sin grandes cambios ni actualización frecuente de las definiciones de las métricas o dimensiones utilizadas. Sus usos más comunes son el seguimiento de objetivos (tanto operativos como estratégicos), el monitoreo de indicadores de forma tal que no se alejen de valores de referencia (benchmarking) generalmente con valores comparables del pasado ya que es poco difundida la comparativa entre empresas salvo notables excepciones.

Analytics supone la constante búsqueda de situaciones reveladoras que aporten valor a la organización (actionable insights). En este sentido, el conocimiento de las particularidades de la industria y de la organización para la cual se está trabajando, es fundamental para el planteo de las hipótesis que guiarán la estrategia de análisis (qué fuentes de datos utilizar, qué variables considerar, cómo agrupar, qué medir, en qué tipo de áreas buscar, entre otros). Un ejemplo de ello es detectar que la rotación de personal ha incrementado de manera importante en los últimos meses, y saber que es un motivo de preocupación por parte de la gerencia. Uno de los objetivos del analista de datos es ver cómo y de qué manera puede aportar a mejorar o revertir esa situación desde su puesto. Esta búsqueda puede encaminarse desde la óptica salarial y puede plantearse la siguiente hipótesis: el salario de la empresa está por debajo de lo que ofrece el mercado. No obstante, si entiende que la empresa es totalmente reacia a incrementar los salarios, sabe que, aunque logre demostrar que éstos son mucho más bajos que en el mercado, no va a lograr que se revierta esta situación y no aportará una perspectiva útil para la solución del problema. En cambio, sí podrá analizar los perfiles entre los empleados que deciden continuar trabajando en la empresa y los que no, para que en el proceso de selección se contemplen y se haga hincapié en los perfiles similares a los empleados con más propensión de permanecer más tiempo en la organización.

La aplicación de técnicas de análisis de datos que se diferencian de Analytics en el carácter automatizado y en la continuidad en la obtención de los indicadores, merece una mención especial: programas informáticos como R, SPSS, Stata, Matlab, E-views, entre otros.

Asimismo, en un artículo que se trabaja esta temática, no se puede dejar de mencionar Excel, que es conocido como el software más utilizado de análisis de datos, pero no por ello el más adecuado para la gran mayoría de los objetivos perseguidos por esta disciplina.

Un punto que me gustaría dejar en claro es la diferencia entre analytics y data analysis ya que son muy similares y son continuamente confundidas. Generalmente las técnicas de análisis de

datos (data analysis) están vinculadas con el sector académico, no con el sector corporativo dado el carácter transitorio de los fenómenos que estudia (“movimientos migratorios entre Uruguay y España entre 1945 y 1955” por citar un ejemplo). Sin embargo, es bastante común encontrar empresas que realizan análisis de datos y no analytics, lamentablemente.

Por último, debemos contemplar el fenómeno de Big Data que es un concepto que está altamente relacionado con los anteriores, pero que no es comparable por algunas razones que mencionaré a continuación.

En primer lugar, Big Data refiere a un tipo de datos y no a técnicas que permitan obtener información a partir de datos. Un conjunto de datos debe reunir varios atributos para ser considerado como Big Data e inclusive si los reuniera, podría no ser tratado como Big Data.

Comencemos por definir qué es Big Data y lo haremos a través de los atributos que el conjunto de datos debe poseer para ser considerado como tal. En este apartado es que mencionaremos las ya “populares” 3 V’s (aunque hay autores que mencionan 4 o incluso 5 V’s).

Las mismas son:

- Volumen
- Velocidad
- Variedad

Desarrollaremos el atributo *volumen* en términos relativos ya que debemos tomar un punto de comparación para determinar que la cantidad de datos es “grande” como para ser considerada Big Data. La respuesta es más esquivada de lo que en principio se presenta ya que muchas empresas recogen, administran y analizan una gran cantidad de datos, pero no los tratan como Big Data. Es necesariamente arbitrario definir una cantidad de bytes para presentarla como el punto de referencia que determine que una cantidad mayor es Big data y que menos que esa cantidad de bytes no lo es. Algunos consideran que una cantidad por encima de un par de decenas de terabytes ya es suficiente para considerar ese conjunto de datos como Big data, mientras otros autores mencionan que una cantidad menor a un petabyte (1.024 terabytes) no debe considerarse como Big data.

El segundo atributo es la *velocidad* y esto implica que las fuentes de datos Big data, lejos de ser una colección de hechos del pasado, que no son pasibles de actualización o adición de datos, son, de hecho, un conjunto de datos “vivos”, es decir, de datos en continuo cambio y en aumento. El advenimiento de las redes sociales otorga a diferentes organizaciones una gran potencialidad de detectar y apropiarse de valor. Las redes sociales están “alimentadas” por más de mil millones de usuarios a nivel mundial que solamente en Facebook generan 4 millones de likes cada minuto, o que 35 millones de personas actualizan su perfil por día⁴. Si bien las redes sociales son un excelente ejemplo de velocidad en cuanto al crecimiento de ese conjunto de datos denominado Big data, hay otros ejemplos que valen la pena mencionar. Cabe destacar el caso de otro fenómeno como *internet of things* o simplemente IOT, que refiere a que las cosas (zapatos, autos, relojes, heladeras, etc) tengan sensores que midan sus diferentes atributos y las comuniquen a una gran base de datos para su posterior análisis. El incremento de la popularidad de IOT, ha generado bases de datos que tienen la característica de velocidad propia de un set de datos que perfectamente puede describirse como Big data.

⁴ <https://www.socialpilot.co/blog/social-media-statistics>

De esta manera, llegamos a la última característica de Big data: la *variedad*. Personalmente creo que este atributo ha sido el más importante para que Big data no haya sido en el pasado un problema para los profesionales informáticos. Una vez que esta característica se hizo presente de manera más frecuente en los sets de datos utilizados por las empresas, se empezó a generar la necesidad de crear un conjunto de software y hardware específico para su tratamiento. Para entender esta característica debemos primero comprender que la tecnología estándar para la gestión de datos en las organizaciones es la de bases de datos relacionales (que operan en su mayoría mediante el lenguaje SQL). En estas bases de datos, cada elemento de información tiene su lugar propio y debe seguir ciertas reglas, ya que, de lo contrario, no podrá formar parte de esa base de datos (tablas, campos, tipos de datos). Si, por ejemplo, tengo en la tabla clientes, un campo llamado: fecha de nacimiento y tiene como tipo de datos admitidos, una fecha, esa tabla nunca va a poder contener el registro de un cliente con fecha de nacimiento “ab27/13/1&90” ya que por diferentes motivos no es reconocida como una fecha. Digamos que la base de datos “espera”⁵ un conjunto de datos que sigan ciertas reglas y que, si el dato no las cumple, no puede formar parte de ella. Big data es exactamente lo opuesto a esta lógica ya que los datos pueden encontrarse en formatos no estructurados (texto, audio, video) o pueden provenir de fuentes con estructuras cambiantes (por ejemplo, los datos que provienen de scraping⁶ de páginas web externas a la organización y que cambian sin previo aviso).

A modo de conclusión, me parece destacable el hecho de que estos conceptos estén presentes en las conversaciones de los profesionales de hoy. Esto es consecuencia de que no hay sectores dentro de las organizaciones que no estén siendo profundamente transformados por esta área del conocimiento. Estas transformaciones son documentadas cada vez más en publicaciones de diversa índole: desde artículos académicos y de revistas especializadas en una industria, hasta los diarios nacionales e internacionales. En segundo lugar, atribuyo la falta de claridad y la confusión entre éstos y otros conceptos, a la academia, que es quien tiene el rol social de definir y delimitar cada nueva área del conocimiento. La falta de flexibilidad de este tipo de instituciones se ve reflejada en que recién en los últimos 2-3 años, está creando a nivel internacional las primeras carreras y maestrías en Analytics, Data science, Machine learning y similares.

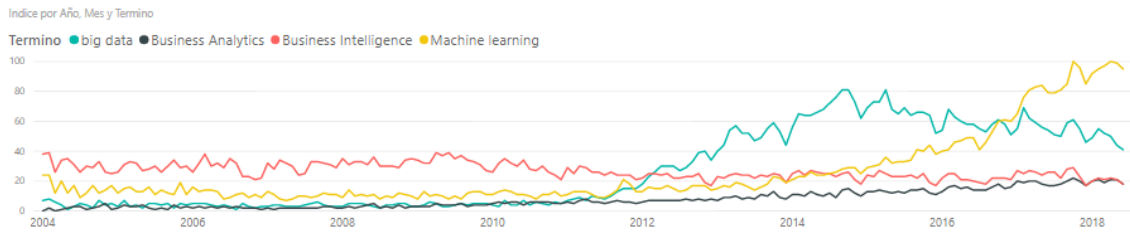
Por otro lado, veo que los 2 primeros conceptos tienen por objetivo tomar los datos, convertirlos en información y distribuirla a los encargados de tomar decisiones dentro de las organizaciones, mientras que Big data es una tipología de conjuntos de datos y que por lo tanto son pasibles de ser transformados en información. Son conceptos muy vinculados, pero claramente no son sinónimos.

Finalmente, me gustaría destacar cómo evolucionó en Estados Unidos, la popularidad⁷ de 4 términos que se encuentran fuertemente asociados a datos, análisis y valor. Estos 4 términos de búsqueda son: Big Data, Business Intelligence, Business Analytics y Machine Learning (traducido comúnmente como aprendizaje automático y vinculable a una versión avanzada de business Analytics), tres de los cuales ya he mencionado, y el último lo dejo como una excusa para volver a comentar sobre este tipo de fenómenos que está revolucionando todos los aspectos de nuestra vida.

⁵ La base de datos se nutre de algún sistema de generación y/o extracción de datos. Un ejemplo típico sería un sistema de gestión empresarial

⁶ El scraping web puede definirse como la técnica de extraer datos de páginas web

⁷ Medida como el índice de búsquedas en Google



Como se observa en la presente gráfica, Business Intelligence fue el término de búsqueda más popular hasta 2012 y luego lo supera ampliamente y de manera más intensa el término Big data. Sin embargo, a partir de fines de 2016, el término Machine Learning es el que lleva la delantera y es, además, el que desde 2004, logra el índice de popularidad más alto de los 4.

Espero haber esclarecido estos conceptos y haber transmitido la importancia que tienen en el mundo de los negocios en tiempos de la 4ª revolución industrial. A su vez, espero haber creado un poco de incertidumbre sobre el concepto y el alcance de otro de los términos comúnmente asociados a estos 3 como ser el fenómeno de Machine Learning.

Enlaces vinculados:

<https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>

<https://www.logianalytics.com/bi-trends/business-intelligence-vs-analytics-whats-the-difference/>

<https://rapidminer.com/blog/summarizing-differences-business-intelligence-advanced-analytics/>