

# Puliendo diamantes

Los datos cada vez cobran mayor relevancia. De hecho, se podría llegar a decir que los datos son el nuevo lenguaje que tienen que hablar los negocios hoy en día. Los datos conducen a la comprensión y gracias a ello las organizaciones disponen de nuevas herramientas para poder tomar decisiones de negocio.

Las personas creamos datos constantemente aunque no nos demos cuenta de ello. Ya no solo a través de las redes sociales, sino que estamos constantemente recibiendo y enviando información. (Incluso cuando nuestros dispositivos están en estado de reposo)

La cantidad de datos que producimos cada día es realmente sorprendente. Cada día generamos aproximadamente 2,5 quintillones de bytes pero con el crecimiento del Internet de las Cosas (IoT) esta cifra pronto quedará obsoleta. Sólo en los dos últimos años hemos generado el 90 por ciento de todos los datos generados a lo largo de la historia. [1]

De acuerdo a un análisis realizado por la consultora Domon durante el año 2018 [2], cada minuto se envían medio millón de tuits a través de la red y se reproducen aproximadamente 750.000 canciones de Spotify. Gestionar este volumen de datos resulta bastante complicado.

La obtención de los datos y su preparación para el análisis es una de las tareas más tediosas a las que se enfrentan las organizaciones en la actualidad. Los analistas dedican mucho tiempo a buscar y reunir los datos adecuados. De hecho, por promedio, los analistas dedican entre el 60 y el 80 por ciento de su tiempo a la preparación de los datos en lugar del análisis.

Además, todos los datos no poseen el mismo valor ni presentan un formato adecuado. Es por ello que a la hora de analizar los

mismos es necesario atender a un aspecto importante: la **calidad**.

### **Pero... ¿qué es la calidad de los datos?**

La calidad de los datos o “Data Quality”, en inglés, hace referencia a una percepción o una evaluación de la idoneidad de los datos para cumplir su propósito en un contexto dado. Esta puede estar determinada por factores como la exactitud, la integridad, la confiabilidad, la relevancia o cómo de actualizados se encuentran. [3]

De hecho, hacer uso de datos de baja calidad puede conllevar a la elaboración de informes erróneos o a estrategias mal planteadas. El daño económico provocado por estos problemas puede ir desde gastos adicionales o hasta multas por la elaboración de informes financieros inadecuados.

Según datos de IBM [4], en el año 2016 el coste anual de los problemas generados por la mala calidad de los datos en Estados Unidos se situó alrededor de los 3,1 trillones de dólares. La razón por la cual los datos con poca calidad cuestan tanto es que estos en muchas ocasiones presentan errores y ante una fecha límite crítica, muchos trabajadores simplemente hacen correcciones ellos mismos para completar la tarea en cuestión.

En este mundo plagado de datos, lo importante no es solo tener una gran cantidad de datos, sino contar con datos de calidad que permitan llevar a cabo análisis exhaustivos con los que obtener los mejores beneficios. En un panorama saturado de datos como el que tenemos actualmente, buscar datos de calidad resulta imprescindible.

Debido a este fenómeno, incluso han comenzado a proliferar start-ups encargadas del procesamiento de los datos. Liberando así a las compañías de realizar estas arduas tareas. De hecho, disponer de datos de calidad resulta esencial a fin de poder realizar análisis correctos.

Puede que al principio los datos no parezcan relevantes pero puliendolos un poco pueden llegar a convertirse en diamantes.

## Referencias

[1] "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read", Forbes, consultado el 15 de enero de 2020, <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/>

[2] "Data Never Sleeps 6.0", Domo, acceso el 15 de enero de 2020, <https://www.domo.com/learn/data-never-sleeps-6>

[3] "¿Qué es Data Quality y por qué es importante?", Elternativa, acceso el 15 de enero de 2020, <https://www.elternativa.com/blog/index.php/2019/04/03/que-es-data-quality-y-por-que-es-importante/>

[4] "Bad data costs the US 3 trillion per year", Harvard Business Review, acceso el 15 de enero de 2020, <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>