

EDUARDO
V. ABAD

Data
Analytics



ANÁLISIS DE DATOS: INTEGRIDAD DE DATOS Y CUMPLIMIENTO NORMATIVO



Eduardo V. Abad



[@eduardovabad](https://www.instagram.com/eduardovabad)



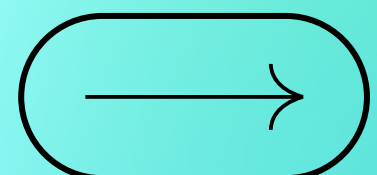
INTEGRIDAD DE DATOS Y CUMPLIMIENTO NORMATIVO

Este documento refleja la importancia de la integridad de datos y brinda un ejemplo con datos de una empresa global.

Escenario: Fechas del calendario de una empresa global

Las fechas del calendario se presentan en una amplia gama de formas abreviadas. Según el lugar donde vives, se puede usar una forma abreviada diferente.

- En algunos países, 12/10/20 (DD/MM/AA) significa 12 de octubre de 2020.

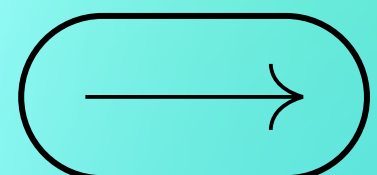


- En otros países, el estándar oficial es AAAA-MM-DD, por lo tanto el 12 de octubre de 2020 pasa a ser 2020-10-12.
- En los Estados Unidos, la forma aceptada es mes, día, año (MM/DD/AA), por lo tanto el 12 de octubre de 2020 será 10/12/20.

Ahora, piensa lo que podría suceder si estás trabajando como analista de datos para una empresa internacional y no tienen en cuenta los diferentes formatos de las fechas.

La integridad de datos resultante puede ser cuestionable. El análisis de los datos sería **inexacto**.

¿Qué pasaría si pidieras stock de más para diciembre cuando en realidad lo necesitaban para octubre?

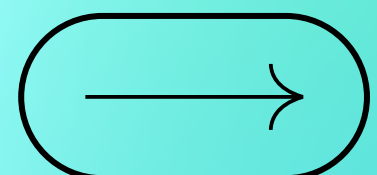


Un buen análisis depende de la integridad de los datos, y la integridad de los datos generalmente depende del uso de un formato único.

Por lo tanto, es importante verificar varias veces el formato de las fechas para asegurarse de que lo que creías que era el 10 de diciembre de 2020 en realidad era el 12 de octubre de 2020 y viceversa.

A continuación, se incluyen algunos temas para tener en cuenta:

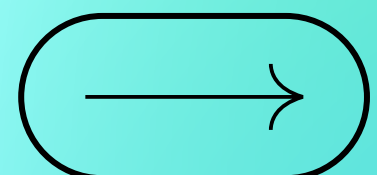
La replicación de datos puede comprometer la integridad de los datos: Siguiendo con el ejemplo anterior, imagina que pides a tus colegas internacionales que verifiquen las fechas y se ajusten a un formato. Entonces, un analista copia un gran conjunto de datos para controlar las fechas.



Sin embargo, por temas relacionados con la memoria, se copia solamente parte del conjunto de datos. El analista podría así verificar y estandarizar datos incompletos. En consecuencia, una parte del conjunto de datos se certificaría como conforme, pero la totalidad del conjunto de datos aún tendría fechas sin verificar. Dos versiones de un mismo conjunto de datos pueden provocar resultados inconsistentes.

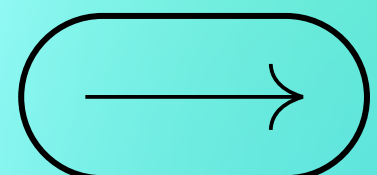
Una auditoría final de los resultados podría ser **necesaria** para descubrir qué fue lo que sucedió y corregir todas las fechas.

La transferencia de datos puede comprometer la integridad de los datos: Otro analista controla las fechas en una hoja de cálculo y elije importar los datos validados y estandarizados de nuevo a la base de datos.



Pero supongamos que el campo de datos de la hoja de cálculo se clasificó erróneamente como campo de texto durante el proceso de importación (transferencia) de datos. Ahora, algunas de las fechas en la base de datos quedaron registradas como cadenas de texto. En este punto, se deben limpiar los datos para poder restaurar su integridad.

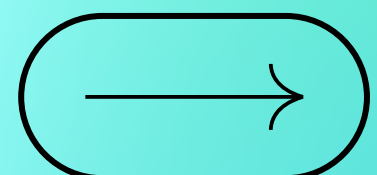
La manipulación de datos puede comprometer la integridad de los datos: Al controlar las fechas, otro analista se da cuenta de lo que parece ser un registro duplicado en la base de datos y lo borra. Pero, en realidad, el analista borró un registro único de una filial de la empresa y no un registro duplicado de la empresa. A tu conjunto de datos ahora le faltan datos y los datos deben ser restaurados para que sean exhaustivos.



CONCLUSIONES

Afortunadamente, la integridad de los datos se puede mantener con un formato de fecha estándar y el cumplimiento normativo por parte de todas las personas y de los sistemas que trabajan con los datos.

Pero más allá de dónde provienen tus datos, siempre **debes asegurarte de controlar si los datos son válidos, están completos y limpios** antes de comenzar cualquier tipo de análisis.





**SÍGUEME
PARA OBTENER
MÁS INFORMACIÓN
Y TIPS.**



Eduardo V. Abad



[@eduardovabad](https://www.instagram.com/eduardovabad)