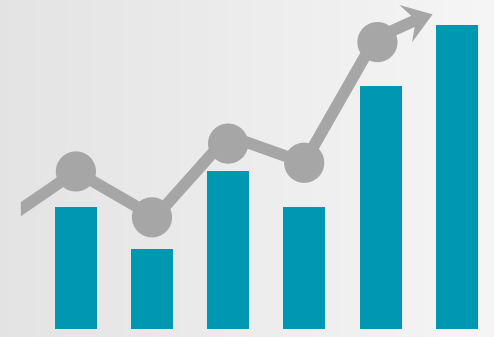
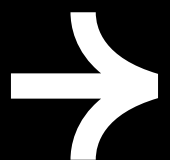


DATA ANALYTICS



LIMPIEZA DE DATOS: ERRORES COMUNES QUE SE DEBEN EVITAR

BY EDUARDO V. ABAD



01

NO HACER UNA COPIA DE SEGURIDAD ANTES DE LIMPIAR LOS DATOS

Siempre es bueno ser precavido y crear una copia de seguridad de tus datos antes de empezar a limpiarlos. Si tu programa colapsa, o si tus cambios provocan un problema en tu conjunto de datos, siempre puedes recurrir a la versión guardada y recuperarla.

El simple procedimiento de hacer una copia de seguridad de tus datos puede ahorrarte horas de trabajo y, fundamentalmente, un dolor de cabeza.



02

NO VERIFICAR LOS ERRORES DE ORTOGRAFÍA

Los errores de ortografía pueden ser errores tan simples como los errores de escritura o de entrada de datos. La mayoría de las veces se pueden detectar las palabras mal escritas o los errores gramaticales comunes, pero se dificulta con datos como nombres o direcciones.

Por ejemplo, si estás trabajando en la tabla de una hoja de cálculo con datos de clientes, puedes encontrarte con un cliente llamado "John" cuyo nombre fue ingresado incorrectamente como "Jon" en algunos lugares. Es muy probable que la revisión ortográfica de la hoja de cálculo no lo marque, por lo tanto, si no verificas los errores de ortografía y lo captas, tu análisis contendrá errores.



03

OLVIDARSE DE DOCUMENTAR ERRORES

Documentar tus errores puede ahorrarte mucho tiempo, ya que te ayuda a evitar esos errores en el futuro mostrándote cómo resolverlos.

Por ejemplo, podrías encontrar un error en una fórmula en tu hoja de cálculo. Descubres que algunas de las fechas en tus columnas no tienen el formato correcto. Si tomas nota de esta corrección, puedes consultarla la próxima vez que tu fórmula se altere y anticiparte a la resolución de problemas.

Documentar tus errores también te ayuda a hacer un seguimiento de los cambios que se producen en tu trabajo, de manera que puedas retroceder si algo no funcionó.



04

NO VERIFICAR LOS VALORES DE CAMPOS ERRÓNEOS

Un valor de campo erróneo ocurre cuando los valores se ingresan en el campo equivocado. Incluso estos valores podrían tener un formato correcto, lo cual dificulta aún más su detección si no tienes cuidado. Por ejemplo, podrías tener un conjunto de datos con columnas para ciudades y países. Son el mismo tipo de datos, por eso, es muy fácil mezclarlos. Pero si intentaras encontrar todas las instancias de España en la columna de país, y España se hubiera ingresado incorrectamente en la columna de ciudad, perderías puntos de datos clave.

Asegurarte de que tus datos se hayan ingresado correctamente es esencial para lograr análisis precisos y completos.



05

PASAR POR ALTO VALORES FALTANTES

Los valores faltantes en tu conjunto de datos pueden crear errores y aportar conclusiones inexactas.

Por ejemplo, si intentas obtener el número total de ventas de los últimos tres meses, pero se omitió una semana de transacciones, tus cálculos serán inexactos.

Como práctica recomendada, intenta mantener tus datos lo más limpios posible conservando su exhaustividad y coherencia.



06

MIRAR SOLO UN SUBCONJUNTO DE DATOS

Es importante pensar en todos los datos relevantes durante la limpieza de datos. Esto te ayuda a asegurarte de comprender el panorama general de los datos, y a prestar atención a todos los posibles errores.

Por ejemplo, si estás trabajando con datos sobre patrones de migración de las aves de diferentes fuentes, pero solo limpias una fuente, podrías no darte cuenta de que algunos datos están repetidos. Esto causará problemas en tu análisis posterior.

Si quieres evitar errores comunes como las duplicaciones, cada uno de los campos de tus datos requiere igual atención.



07

PERDER DE VISTA LOS OBJETIVOS EMPRESARIALES

Cuando limpias datos, podrías hacer descubrimientos nuevos e interesantes sobre tu conjunto de datos, pero no querrás que esos descubrimientos te distraigan de la tarea en cuestión.

Por ejemplo, si estás trabajando con datos meteorológicos para hallar el número promedio de días de lluvia en tu ciudad, también podrías detectar algunos patrones interesantes sobre las nevadas.

Eso es realmente fascinante, pero no se relaciona con la pregunta que estás tratando de responder justo en este momento. ¡Ser curioso es genial! Pero no dejes que eso te distraiga de la tarea que tienes entre manos.



08

NO ANALIZAR EL SISTEMA ANTES DE LIMPIAR LOS DATOS

·Si queremos limpiar nuestros datos y evitar futuros errores, necesitamos comprender la causa raíz de los datos sucios. Imagina que eres un mecánico de autos.

Buscarías la causa del problema antes de empezar a reparar el auto, ¿verdad? Lo mismo ocurre con los datos. Primero, averiguarás de dónde provienen los errores. Quizás provengan de un error en el ingreso de datos, falta de configuración del corrector ortográfico, o duplicaciones.

Luego, una vez que sabes de dónde provienen los datos incorrectos, puedes controlar la situación y mantener tus datos limpios.





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

BY EDUARDO V. ABAD

