



Módulo 10

Verificar e informar los resultados de tu limpieza

AGENDA DE LA CLASE

- ✓ Verificar e informar los resultados
- ✓ Herramientas para la verificación y limpieza final
- ✓ Capturar los cambios de la limpieza



01 - Verificar e informar los resultados

La **verificación** es un **proceso** para confirmar que un esfuerzo de **limpieza de datos** se ejecutó correctamente y que los datos **resultantes** son **precisos** y **confiables**.



El **primer paso** en el proceso de verificación es **regresar a tu conjunto de datos original** sin depurar y **compararlo** con lo que tienes ahora.

Revisar si los espacios vacíos, duplicados, errores ortográficos, etc. que identificaste en tu conjunto sucio ya no están en tu conjunto limpio.

Utilizar filtros, formatos condicional, CTRL + B para buscar palabras.

01 - Verificar e informar los resultados



Otro **paso clave** de la verificación incluye tener una **visión general de tu proyecto**.

Es el momento para confirmar que te estás **enfocando** en el **problema del negocio** que necesitas resolver y para asegurarte de que los **datos** son realmente **capaces de resolver ese problema**.



- 1 Confirmar el problema del negocio
- 2 Confirmar el objetivo del proyecto
- 3 Verificar que los datos pueden resolver el problema y que están alineados con el objetivo

01 - Verificar e informar los resultados

Asegúrate de haber identificado y corregido los problemas más comunes:



Fuentes de errores



Errores de coincidencia de tipo de datos



Datos nulos



Cadenas desordenadas (incoherentes)



Palabras mal escritas



Formatos de fecha desordenados (incoherentes)



Números mal escritos



Etiquetas engañosas de variables (columnas)



Espacios y caracteres extra



Datos truncados



Duplicados



Lógica de negocios

02 - Herramientas para la verificación y limpieza final

Algunas herramientas en Excel muy buenas para corregir errores comunes de manera automática que podemos usar son:



Función =ESPACIOS(): quita los espacios, al principio, al final o los repetidos en los datos



Quitar duplicados: busca automáticamente y elimina entradas duplicadas de una hoja de cálculo



Tablas dinámicas: resumen de datos y nos permite encontrar errores que aparecen repetidamente



Buscar y reemplazar: para encontrar un error que se repite varias veces, como por ejemplo una palabra mal y reemplazarla por el valor correcto

03 - Capturar los cambios de la limpieza

La **documentación** es el **proceso** de hacer un **seguimiento de los cambios**, las adiciones, las eliminaciones y los errores involucrados en tu esfuerzo de limpieza de datos

Tener este **registro** permite **tres cosas muy importantes**:



Recuperar errores de la limpieza de datos



Informar a otros usuarios sobre los cambios que has hecho



Determinar la **calidad de los datos** que se van a usar en el análisis

Un **registro de cambios** es un **archivo** que contiene una lista ordenada **cronológicamente** de las **modificaciones** que se hicieron a un proyecto.



Historial de versiones de la hoja



Historial del proyecto

¿PREGUNTAS?





¡Muchas gracias!