



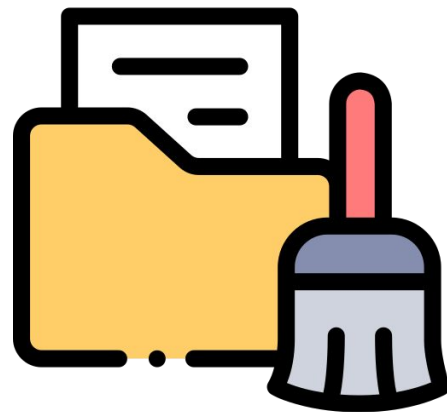
# Módulo 7

## **La importancia de la Integridad. Datos impecables.**



## AGENDA DE LA CLASE

- ✓ Integridad de los Datos
- ✓ Datos Insuficientes
- ✓ Limpieza de datos
- ✓ Herramientas para la limpieza de datos en hojas de cálculos



# 01 - Integridad de los Datos

La **integridad de los datos** consiste en la:

**exactitud,**

**integridad,**

**consistencia y**

**confiabilidad**

de los datos a lo largo de su **ciclo de vida**.



Si la integridad de los datos está comprometida, un análisis entero puede estar mal hecho porque los datos con los que trabajabas eran erróneos desde el principio.

# 01 - Integridad de los Datos

Formas de comprometer la integridad:

1

**Al replicar los datos:** Al archivar en distintas locaciones pueden perder **sincronización**.



2

**Al transferir los datos:** Si el proceso se **interrumpe**, los datos quedan incompletos.



3

**Al manipular los datos:** En el proceso de hacer los datos más eficientes, **un error puede comprometer la integridad**.



4

**Otros:** Errores humanos, Virus, Malware, Hackeos.



## 02 - Datos insuficientes

Datos **limitados geográficamente**

**Solución:** buscar set de datos globales

Datos **se actualizan constantemente.**

**Solución:**

- Esperar por más datos.
- Realizar estudios preliminares para luego ser ampliados.



Ejemplos de datos  
insuficientes

Datos de **un solo origen o fuente.**

**Solución:**

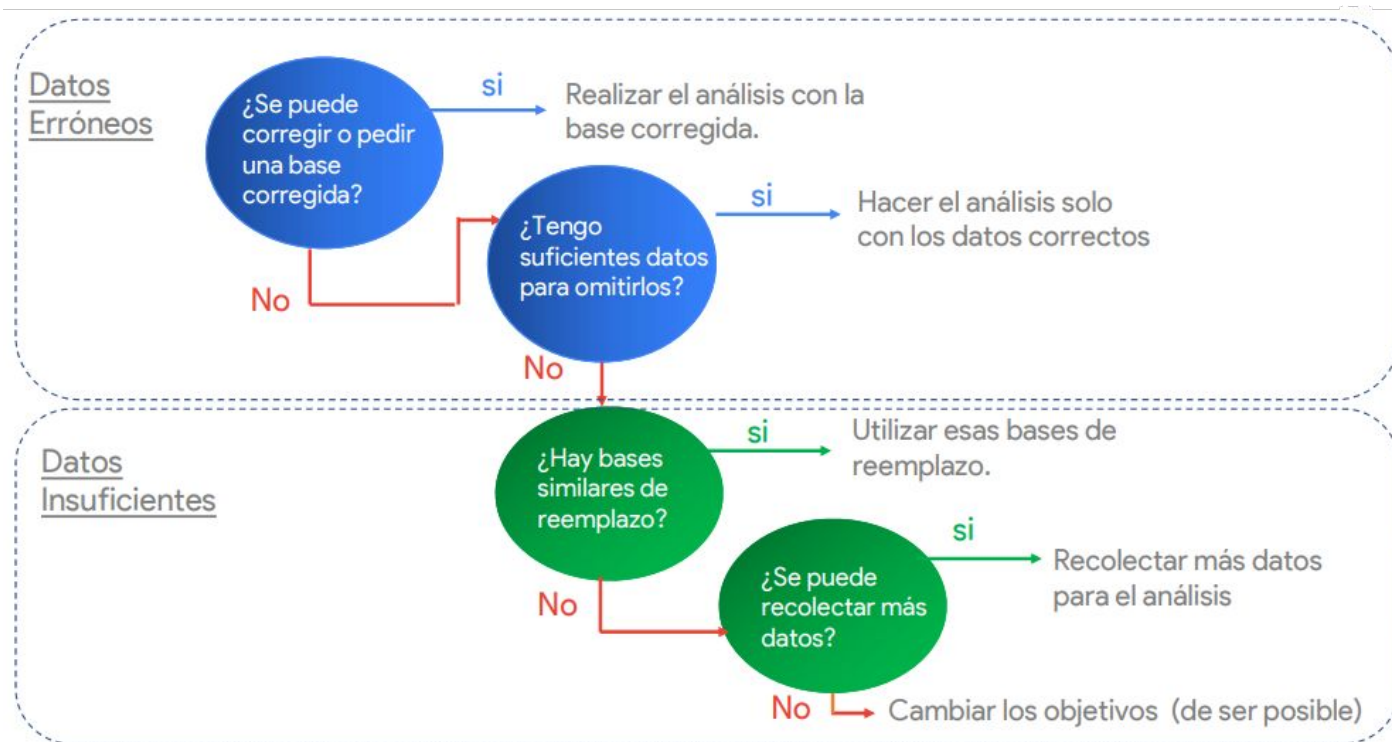
- Buscar set de datos externos para complementar.
- Identificar tendencias generales con lo disponible.
- Buscar datos de reemplazo

Datos **desactualizados**

**Solución:** buscar set de datos más actualizados

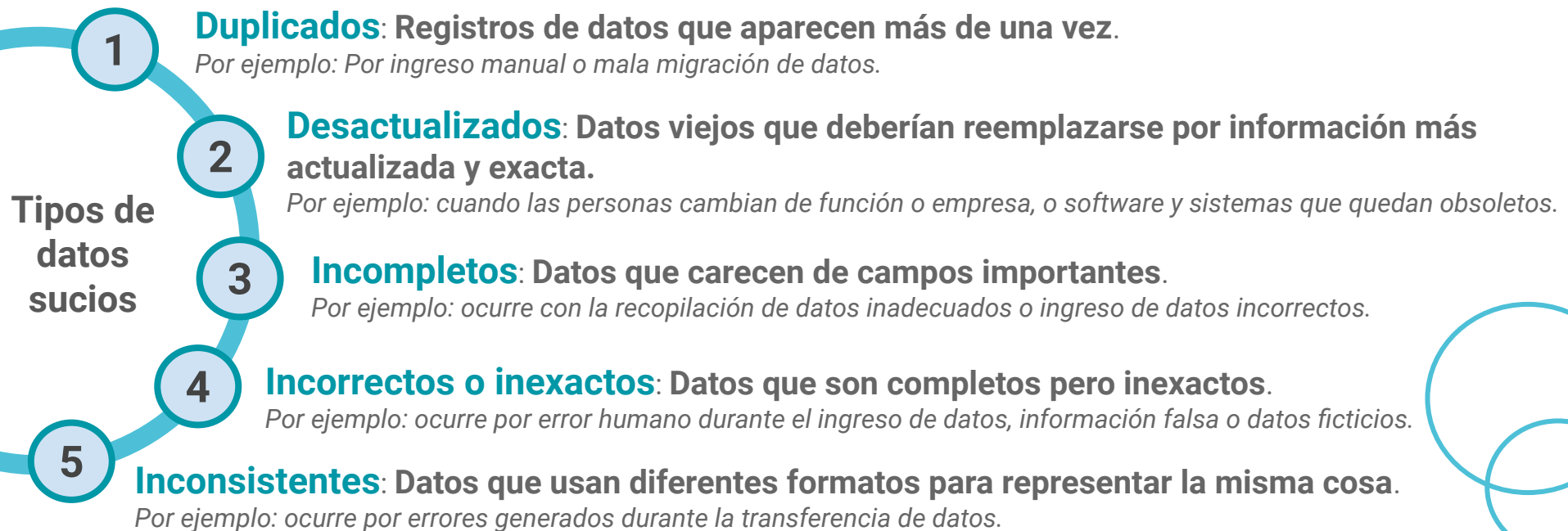
## 02 - Datos insuficientes

¿Qué hacemos cuando los datos son erróneos o insuficientes?



## 03 - Limpieza de datos

Los **datos sucios** son datos **incompletos**, **incorrectos** o **irrelevantes** respecto del problema que intentas resolver.



## 03 - Limpieza de datos

Pautas a tener en cuenta:





## 03 - Limpieza de datos

### Errores comunes de la limpieza de datos:



No chequear errores de tipeo



Perder el seguimiento de los objetivos de negocio



No documentar errores



No tener en cuenta los tiempos de limpieza en el cronograma de trabajo



Mirar solo un conjunto de datos y no la película completa



No solucionar la fuente del error



No hacer backup previo antes de la limpieza



Pasar por alto valores faltantes o nulos



## 04 - Herramientas para la limpieza de datos en hojas de cálculos

### FORMATO CONDICIONAL

Se trata de una herramienta que cambia la forma en que se ven las celdas cuando los valores cumplen ciertas condiciones.

### QUITAR DUPLICADOS

Se trata de una herramienta que permite quitar duplicados de una tabla o conjunto de celdas.

### FORMATO FECHA

Hay distintos formatos de fechas para seleccionar.

### MANIPULEO DE TEXTO

Una **cadena de texto** es un grupo de caracteres dentro de una celda.

Una característica importante de una cadena de texto es su **longitud**, que es el número de caracteres en ella.

Una **subcadena** es un subconjunto más pequeño de una cadena de texto.



¿PREGUNTAS?





**¡Muchas gracias!**