

## Lección 8: Formateo de Datos – Limpieza y Transformación para Análisis

Bienvenido de nuevo. En esta lección, nos centramos en el formateo de datos: el proceso de limpiar, estructurar y convertir información cruda para que sea usable. Imagina datos "sucios" como un rompecabezas desorganizado – el formateo lo arma. Cubriremos conversión de CSV a tablas, funciones para texto (como SPLIT, JOIN, MATCH), y problemas comunes en transformación (ej: formatos de fecha MM/DD vs DD/MM, decimales con coma vs punto, espacios extras). Usaremos ejemplos del documento adjunto (links a Instagram para visuales) y el texto sobre bucles implícitos, ya que funciones como SPLIT o ARRAYFORMULA ayudan en formateo repetitivo.

Basado en búsquedas comunes, problemas frecuentes incluyen conversiones automáticas (ej: texto a fecha), nulls inconsistentes, y encoding. Soluciones: Fórmulas nativas, Power Query, o scripts. Todo en secciones cortas, con pasos y tablas para claridad.

### Sección 1: Introducción – ¿Qué es Formateo y Por Qué Importa?

Formateo es limpiar datos: eliminar espacios, convertir tipos (texto a número), estandarizar formatos (fechas, monedas), y estructurar (CSV a tabla). Evita errores en análisis (ej: suma falla por texto disfrazado de número).

- **Problemas comunes** (de búsquedas):

- Fechas inconsistentes (MM/DD/YYYY vs DD/MM/YYYY, o auto-conversión a números seriales).
- Decimales (coma en Europa vs punto en USA, causando errores en cálculos).
- Espacios/blanks extras (datos "sucios" de imports).
- Nulls/bad values (ej: "NA" como texto, o celdas vacías que rompen fórmulas).
- Encoding (caracteres especiales corruptos en CSV).
- Archivos grandes: Lentitud o pérdida de formato al guardar.

**Por qué aprender:** Datos limpios ahorran tiempo (evita VLOOKUP fallidos) y mejoran precisión. En equipo, estandariza (ej: fechas uniformes).

### Diferencias clave:

- Sheets/Excel: Fórmulas como TRIM, Power Query para batch.
- DAX: Formatea en medidas (FORMAT, DATEVALUE).
- VBA/AppScript: Loops para custom (ej: Trim en rango).

**Pausa:** Abre Sheets, ingresa datos "sucios" (ej: " 123,45 " en A1). Sigue para limpiarlo.

## Sección 2: De CSV a Tabla – Estructurando Datos Importados

CSV es texto plano; conviértelo a tabla para filtros y fórmulas.

- **Pasos en Excel:** Datos > De Texto/CSV > Selecciona archivo > Delimitador (coma) > Cargar a tabla (Insertar > Tabla).
- **Pasos en Sheets:** Archivo > Importar > Subir CSV > Importar como hoja > Formato > Convertir en tabla.
- **Problemas comunes:** Delimitadores erróneos (punto y coma vs coma), encoding (UTF-8 para acentos).
  - Solución: En Power Query (Excel), ajusta delimitador/encoding; en Sheets, usa =IMPORTDATA(URL) para web CSV.

**Ejemplo visual:** Ver Instagram: to columns para texto a columnas, o otro para splits.

**Diferencias:**

- Sheets: =IMPORTDATA para auto-update.
- Excel: Power Query refresca dinámico.
- DAX: Importa CSV en Power BI > Transformar datos.

**Tabla de Conversión:**

Paso	Sheets	Excel	AppScript/VBA	○
Importar CSV	Archivo > Importar	Power Query > De Texto/CSV	AppScript: Utilities.parseCsv; VBA: QueryTables.Add	
A Tabla	Formato > Tabla	Insertar > Tabla	VBA: ListObjects.Add	

**Pausa:** Importa un CSV y conviértelo. Usa Instagram: PDFExcel si es PDF.

### Sección 3: Funciones para Texto – Manipulación Básica

Usa fórmulas para limpiar y formatear texto. Integra bucles implícitos (del texto adjunto) como ARRAYFORMULA para rangos.

- **TRIM/REMOVE BLANKS:** Elimina espacios extras.

- Ej: =TRIM(A1) para " Hola " → "Hola".

- Problema: Blanks invisibles (CHAR(160)); solución: =SUBSTITUTE(A1, CHAR(160), "") + TRIM.

- **SPLIT/JOIN:** Divide/unir texto.

- SPLIT (del texto): =SPLIT("a,b,c", ",") → Array ["a","b","c"] (implícito loop por delimitador).

- JOIN: =JOIN(", ", {"a","b","c"}) → "a, b, c".

- **MATCH/SEARCH:** Busca posición.

- MATCH: =MATCH("pera", A1:A10, 0) → Posición (implícito loop en rango).

- **Otras:** CLEAN (elimina no imprimibles), SUBSTITUTE (reemplaza), TEXT (formatea números como texto).

**Ejemplo con bucle implícito:** =ARRAYFORMULA(SPLIT(A1:A10, ",")) – Divide rango entero (como for each en código).

**Problemas comunes:** Texto como números (ej: "001" pierde ceros); solución: =TEXT(A1, "000").

#### Diferencias:

- Sheets: ARRAYFORMULA envuelve para rangos.
- Excel: Spill en 365; Ctrl+Shift+Enter en viejo.
- DAX: SEARCH, REPLACE en medidas.
- VBA/AppScript: Trim(str), split/join nativos.

#### Tabla de Funciones:

Función Sheets/Excel

DAX

AppScript/VBA



TRIM	=TRIM(A1)	TRIM([Col])	str.trim() / Trim(str)
SPLIT	=SPLIT(A1, ",")	No directo; SPLIT en Power Query	str.split / Split(str)
JOIN	=JOIN(", ", A1:A3)	CONCATENATE	str.join / Join(arr)
MATCH	=MATCH(valor, rango, 0)	SEARCH	indexOf / InStr

Pausa: Prueba =TRIM(" Texto con espacios ") en Sheets. Ver Instagram: Trim.

#### Sección 4: Problemas Comunes en Transformación y Soluciones

Basado en búsquedas: Conversiones automáticas, inconsistencias, lentitud en grandes sets.

- **Fechas (MM/DD vs DD/MM):** Problema: Excel asume locale; "01/02/2023" es Jan 2 o Feb 1.
  - Solución: =DATEVALUE(TEXT(A1, "DD/MM/YYYY")) o Power Query > Cambiar Tipo > Usando Locale.
- **Decimales (coma vs punto):** Problema: "1,23" vs "1.23"; errores en suma.
  - Solución: =SUBSTITUTE(A1, ",", ".") + VALUE; o Configuración > Opciones > Avanzado > Separador decimal.
- **Espacios/Blanks:** Problema: Datos invisibles rompen fórmulas.
  - Solución: =TRIM(CLEAN(A1)); ver Instagram: Remove blanks.
- **Encoding/Caracteres Especiales:** Problema: ñ o á corruptos en imports.
  - Solución: Power Query > Transformar > Encoding UTF-8.
- **Números Largos/Científicos:** Problema: 1234567890123456 → 1.23E+15 ( pierde dígitos).
  - Solución: Formato > Texto antes de ingresar.
- **Fechas en Celda:** Ver Instagram: fecha en celda para parseo.

**Usa bucles implícitos:** =ARRAYFORMULA(IF(A1:A10<>"", TRIM(A1:A10), "")) – Limpia rango

entero.

### Diferencias:

- Sheets: Locale auto; usa QUERY para filtros.
- Excel: Power Query resuelve batch.
- DAX: FORMAT para output.

### Tabla de Problemas/Soluciones:

Problema	Solución Sheets/Excel	DAX	AppScript/VBA
Fechas inconsistentes	=TEXT(A1, "DD/MM/YYYY")	FORMAT([Date], "DD/MM/YYYY")	DateSerial / new Date
Decimales	=SUBSTITUTE(A1, ",", ".")	REPLACE	replace(str, ",", ".")
Blanks	=TRIM(A1)	BLANK() check	Trim / trim

**Pausa:** Corrige una fecha errónea con TEXT.

### Sección 5: Formateo Avanzado con Bucles Implícitos

Del texto adjunto: Funciones como MAP, BYROW, REDUCE simulan loops para formateo repetitivo (ej: TRIM en rango).

- **MAP/BYROW:** Aplica función por elemento/fila.
  - Ej: =BYROW(A1:A10, LAMBDA(row, TRIM(row))) – Limpia columna.
- **REDUCE/SCAN:** Acumula (ej: JOIN implícito).
- **ARRAYFORMULA con IF:** Condicional en rango (ej: formateo si no blank).

**Diferencias:** Sheets usa BYROW directo; Excel MAP en 365.

**Pausa:** Prueba =MAP(SEQUENCE(5), LAMBDA(x, TEXT(x, "00"))) en Excel.

### Sección 6: Consejos Finales

- Usa Power Query para batch; fórmulas para quick fixes.

- Verifica locale (Archivo > Opciones > Idioma).
- Evita auto-conversión: Formato > Texto para IDs.

**Conclusión:** Formateo hace datos accionables. Integra con imports previos para flujos limpios.

**Pausa final:** Limpia un CSV. ¿Tema 9? Ej: "Análisis Avanzado". ¡Dime! 😊