

Limpiar datos con PBI

Objetivo	1
Instrucciones	1
Importar los archivos CSV	1
Limpieza de clientes_dirty.csv	2
Problemas detectados	2
Pasos de limpieza	2
Limpieza de pedidos_dirty.csv	2
Problemas detectados	2
Pasos de limpieza	2
Limpieza de productos_dirty.csv y detalle_pedidos_dirty.csv	3
Resumen de herramientas de Power Query usadas	3
Entregables	4

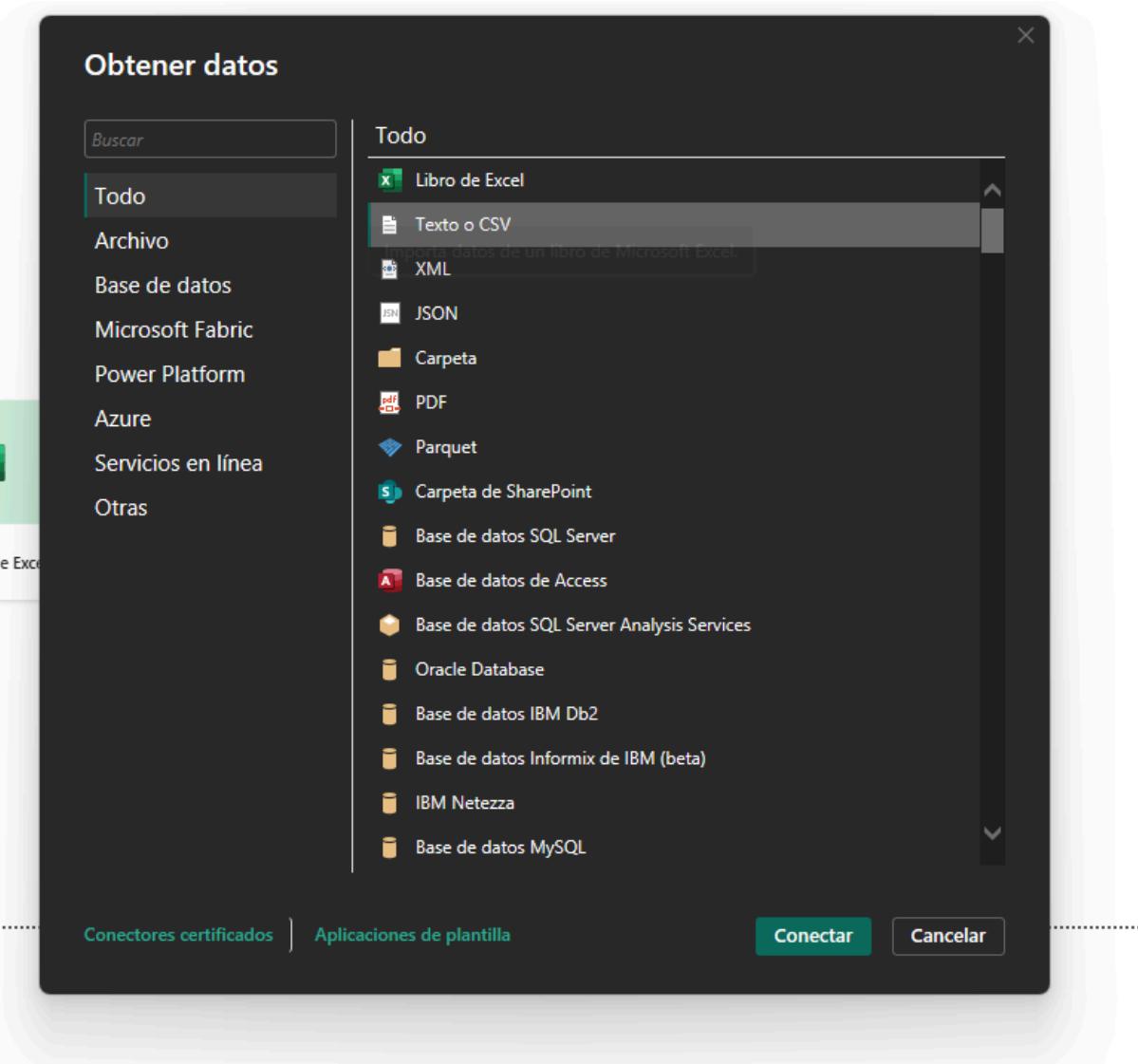
Objetivo

Aprender a **limpiar y preparar datos sucios** con Power BI Desktop.

Instrucciones

Importar los archivos CSV

1. **Descomprime** el archivo zip adjunto *PowerBI_dirty.zip*.
2. Abre **Power BI Desktop**.
3. Haz clic en **Obtener datos > Texto/CSV**.



4. Selecciona los archivos:

- clientes_dirty.csv

clientes_dirty.csv

Origen de archivo: 65001: Unicode (UTF-8) Delimitador: Coma

IDCliente	Nombre	País
1	Ana	España
2	juan	México
null	Luis	namibia
3	Sara	Chile
4	Luis	Argentina
5	María	españa
3	sara	Argentina
null		

- pedidos_dirty.csv
- productos_dirty.csv
- detalle_pedidos_dirty.csv

5. Haz clic en **Cargar o Transformar datos** para ir al **Editor de Power Query**.

Limpieza de clientes_dirty.csv

Problemas detectados

- Espacios en los nombres.
- Mayúsculas y minúsculas inconsistentes.
- Países con espacios y acentos inconsistentes.

Pasos de limpieza

1. Selecciona la columna *Nombre*.

- Quitamos espacios al principio y al final: **Transformar > Limpiar > Recortar**.

The screenshot shows the Microsoft Power Query Editor interface. A table with columns 'IDCliente' and 'Nombre' is displayed. The 'Nombre' column is selected, indicated by a blue border around its header cell. On the right side of the screen, the 'Transformar' (Transform) ribbon tab is active, and the 'Limpiar' (Clean) option is selected. A context menu is open over the 'Nombre' column, listing various data cleaning and transformation options. The 'Recortar' (Trim) option is highlighted with a dark gray background, indicating it is the current step being applied. The 'PROPIEDADES' (Properties) pane on the right shows the column name as 'Nombre' and the 'Todas las propiedades' (All properties) section expanded. The 'PASOS APPLICADOS' (Applied steps) pane shows the sequence of steps taken so far, with 'Tipo cambiado' (Type changed) at the top and 'Recortar' highlighted in dark gray.

- Quitamos caracteres no imprimibles: **Transformar > Limpiar > Limpiar**.

The screenshot shows the Talend Studio interface with a data grid containing columns 'Nombre' and 'País'. A context menu is open over the 'Nombre' column, with the 'Cleanse' option selected. The 'Cleanse' submenu is expanded, showing options like 'Reemplazar errores...', 'Dividir columna', and 'Cambiar nombre...'. To the right, the 'PROPIEDADES' panel shows 'Nombre' set to 'clientes_dirty' and the 'PASOS APlicados' panel shows 'Tipo cambiado' applied. A dropdown menu on the right lists transformation steps: 'minúsculas', 'MAYÚSCULAS', 'Poner En Mayúsculas Cada Palabra', 'Recortar', 'Limpiar', 'Longitud', 'JSON', and 'XML'. The 'Limpiar' step is highlighted.

- Ponemos en mayúsculas cada palabra: **Transformar > Formato > Poner En Mayúsculas Cada Palabra**.

The screenshot shows the Talend Studio interface with the same data grid and context menu. The 'Format' submenu is now selected, showing options like 'Textos recortados', 'Dividir columna', and 'Cambiar nombre...'. The 'PROPIEDADES' panel shows 'Nombre' set to 'clientes_dirty'. The 'PASOS APlicados' panel shows 'Texto recortado' applied. The same dropdown menu on the right is visible, with 'Poner En Mayúsculas Cada Palabra' highlighted.

2. Selecciona la columna *País*.

- Quitamos espacios al principio y al final: **Transformar > Limpiar > Recortar**.

The screenshot shows the Power BI Data Editor interface. A context menu is open over the 'País' column header. The menu includes options like Copiar, Quitar, Quitar otras columnas, Duplicar columna, Agregar columna a partir de los ejemplos..., Quitar duplicados, Quitar errores, Cambiar tipo, Transformar, Reemplazar los valores..., Reemplazar errores..., Dividir columna, Agrupar por..., Rellenar, Anulación de dinamización de columnas, Anulación de dinamización de otras columnas, Anular dinamización de las columnas seleccionadas únicamente, Cambiar nombre..., Mover, Rastrear desagrupando datos, and Agregar como consulta nueva. On the right, the ribbon shows the 'Transformar' tab is selected. The 'PROPIEDADES' pane shows the column name as 'clientes_dirty'. The 'PASOS APlicados' pane lists several steps: Origen, Encabezados promovidos, Tipo cambiado, Texto recordado, minúsculas, MAYÚSCULAS, Poner En Mayúsculas Cada Palabra, Recortar (which is highlighted), Limpiear, Longitud, JSON, and XML.

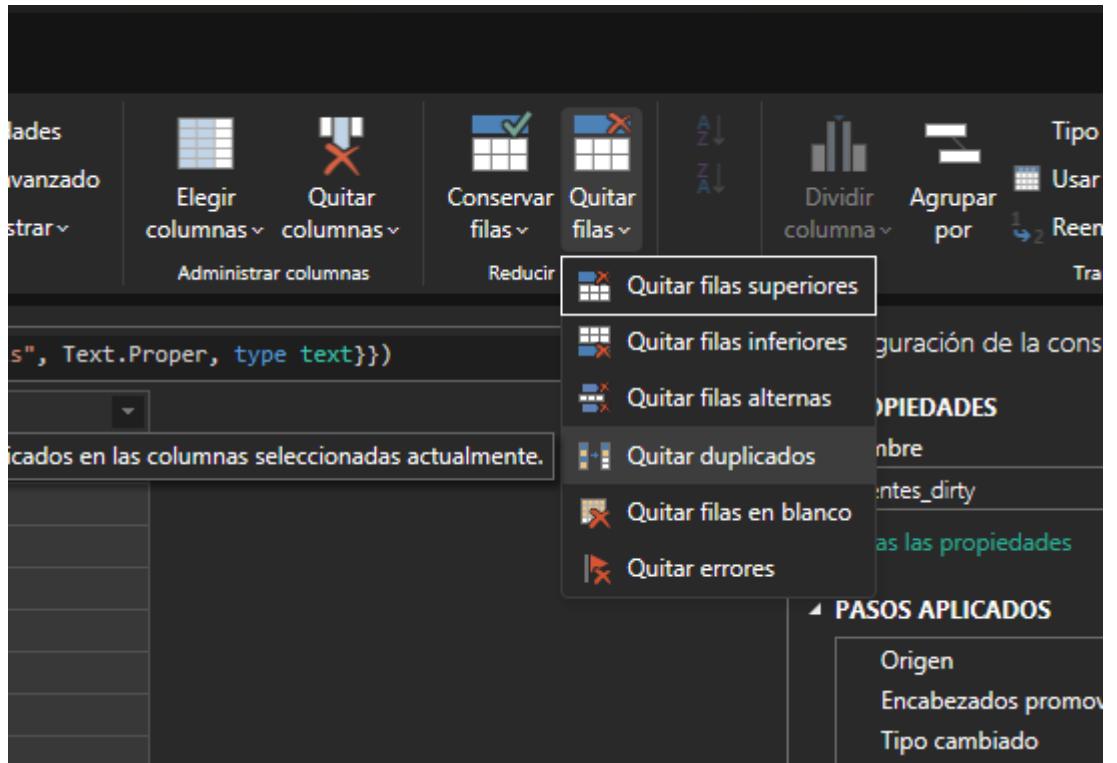
- Quitamos caracteres no imprimibles: **Transformar > Limpiar > Limpiar**.

The screenshot shows the Talend Data Integration environment. A context menu is open over a column named 'País' in a table. The menu includes options for copying, deleting, and transforming data, as well as specific steps for handling duplicates and errors. To the right, the 'PASOS APlicados' (Applied Steps) panel lists several steps: 'minúsculas', 'MAYÚSCULAS', 'Poner En Mayúsculas Cada Palabra', 'Recortar', 'Limpiar', 'Longitud', 'JSON', and 'XML'. The 'PROPIEDADES' (Properties) panel shows the step is named 'clientes_dirty'. The main workspace displays a table with data rows like 'España', 'México', 'namibia', 'Chile', 'Argentina', 'españña', and 'Argentina'.

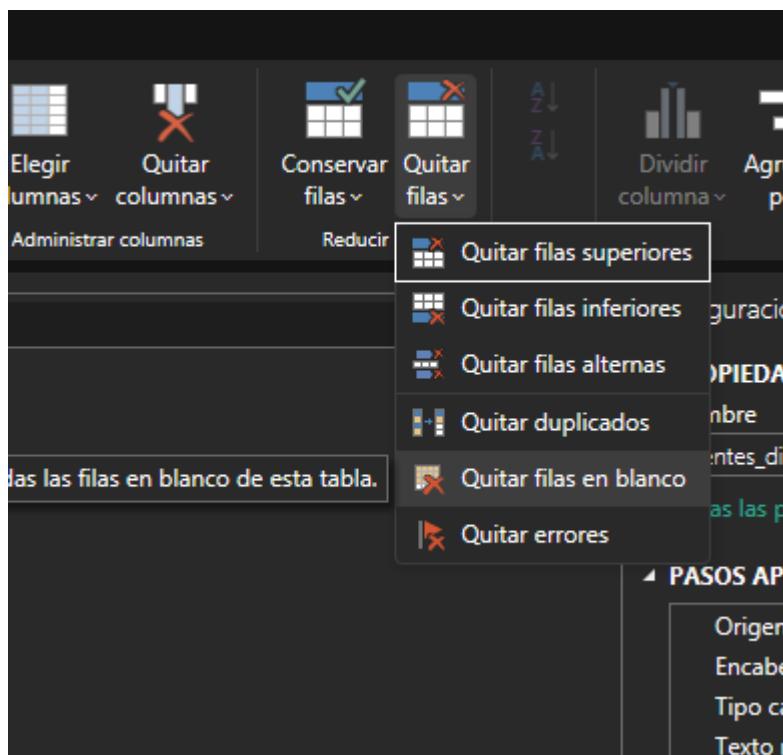
- Ponemos en mayúsculas cada palabra: **Transformar > Formato > Poner En Mayúsculas Cada Palabra**.

3. Selecciona todas las columnas.

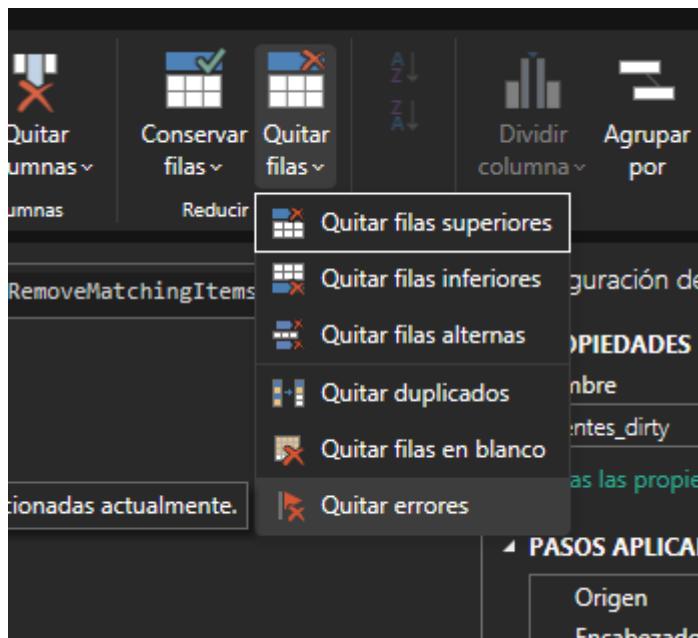
- Eliminamos las filas duplicadas: Inicio > Quitar filas > Quitar duplicados.



- Eliminamos las filas vacías: Inicio > Quitar filas > Quitar filas en blanco.



- Eliminamos las filas con errores: **Inicio > Quitar filas > Quitar errores**.



4. Selecciona la columnas *IDCliente*.

- Selecciona el filtro: **Quitar vacíos**.

The screenshot shows the 'Filtros' (Filters) dialog box for the 'IDCliente' column. The left pane lists filter options: 'Orden ascendente', 'Orden descendente', 'Borrar orden', 'Borrar filtro', 'Quitar vacíos', and 'Filtros de número'. The 'Quitar vacíos' option is currently selected. The right pane shows a table with two columns: 'IDCliente' and 'Nombre'. The 'Nombre' column contains values: Ana, Juan, Luis, Sara, Luis, María, and Sara. The 'IDCliente' column has a dropdown menu with the same seven names. Below the table, there is a 'Buscar' (Search) input field and a list of numerical filters from 1 to 5, all of which are checked. At the bottom are 'Aceptar' (Accept) and 'Cancelar' (Cancel) buttons.

The screenshot shows the Power Query Editor interface. A context menu is open over the first column of a table. The table has three columns: 'IDCliente' (containing values 1, 2, 3, 4, 5), 'Nombre' (containing values 'A', 'B', 'C', 'D', 'E'), and 'País' (containing values 'España', 'Francia', 'Alemania', 'Reino Unido', 'Italia'). The menu options include:

- Copiar (Copy)
- Quitar (Delete)
- Quitar otras columnas (Delete other columns)
- Duplicar columna (Duplicate column)
- Agregar columna a partir de los ejemplos... (Add column from examples...)
- Quitar duplicados (Remove duplicates)
- Quitar errores (Remove errors)
- Cambiar tipo (Change type)
- Transformar (Transform)
- Reemplazar los valores... (Replace values...)
- Reemplazar errores... (Replace errors...)
- Agrupar por... (Group by...)
- Rellenar (Fill)
- Anulación de dinamización de columnas (Cancel dynamic column mapping)
- Anulación de dinamización de otras columnas (Cancel dynamic mapping of other columns)
- Anular dinamización de las columnas seleccionadas únicamente (Cancel dynamic mapping of selected columns only)
- Cambiar nombre... (Change name...)
- Mover (Move)
- Rastrear desagrupando datos (Trace ungrouped data)
- Agregar como consulta nueva (Add as new query)

Limpieza de pedidos_dirty.csv

Problemas detectados

- Fechas en formatos diferentes (10/01/2023, 12-03-2023, etc.).
- Importe con algunos decimales y otros enteros.

Pasos de limpieza

Selecciona la columna *Fecha*.

- Haz clic en **Transformar > Tipo de datos > Fecha**.
- Power Query convertirá automáticamente distintos formatos de fecha.

Selecciona la columna *Importe*.

- Haz clic en **Transformar > Tipo de datos > Número decimal**.

Verifica que *IDCliente* tenga el mismo tipo de dato que en *clientes_dirty.csv* (Número entero).

- Elimina filas duplicadas, vacías y con errores por si en el futuro las hubiera.

Limpieza de productos_dirty.csv y detalle_pedidos_dirty.csv

productos_dirty.csv está casi limpio:

- Asegúrate de que IDProducto sea Número entero.
- Precio como Número decimal.
- Elimina filas duplicadas, vacías y con errores por si en el futuro las hubiera.

detalle_pedidos_dirty.csv:

- IDPedido y IDProducto como Número entero.
- Cantidad como Número entero.
- Elimina filas duplicadas, vacías y con errores por si en el futuro las hubiera.

Resumen de herramientas de Power Query usadas

Acción	Ubicación
Quitar espacios	Transformar > Limpiar > Recortar
Quitar caracteres no imprimibles	Transformar > Limpiar > Limpiar
Capitalizar	Transformar > Formato > Poner En Mayúsculas Cada Palabra
Cambiar tipo de dato	Transformar > Tipo de datos
Quitar filas duplicados	Inicio > Quitar filas > Quitar duplicados
Quitar filas vacías	Inicio > Quitar filas > Quitar duplicados
Quitar filas con errores	Inicio > Quitar filas > Quitar errores
Quitar columnas vacías o con errores	Filtrar columnas y quitar vacíos
Separar columnas	Transformar > Columna > Dividir columna (si aplica)

Entregables

- Añadir como comentarios la tarea un enlace a Github, de acceso público, a un repositorio que contenga el archivo **Power BI**: *nombreyapellidos-A3-2-Limpiar datos con PBI.pbix*.
- Archivo **PDF**: *nombreyapellidos--A3-2-Limpiar datos con PBI.pdf*. Hazlo a partir de este archivo añadiendo, en el apartado correspondiente:
 - Capturas de pantalla de **autoría inequívoca**.
 - Comentarios y observaciones.