

# Ejemplo 3 - Uniendo datos



Lectura CSV Ventas



Lectura CSV Productos

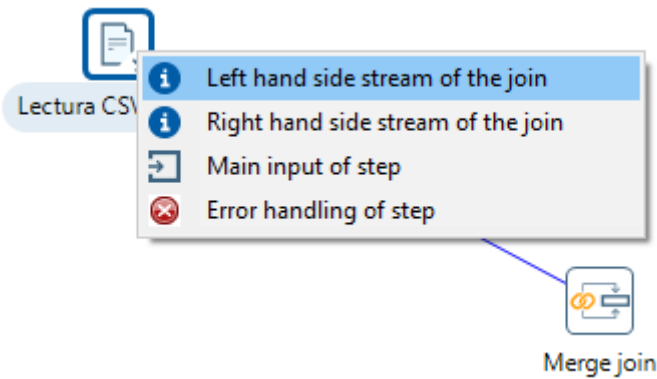
## *Caso de Uso 2 - Lecturas CSV*

En este caso, vamos a leer datos de ventas y de productos (adjuntos), y vamos a unirlos de forma similar a un *join*, para posteriormente, tras agrupar los datos, crear un informe.

### *Merge join*

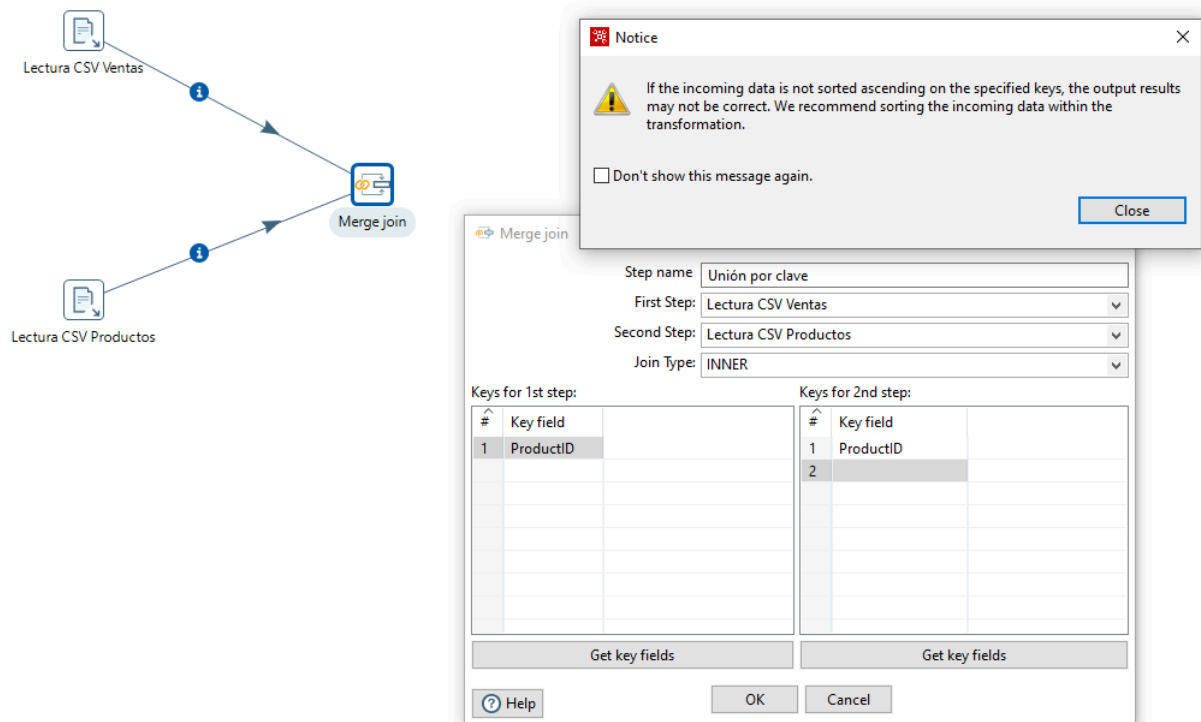
Para ello, primero vamos a leer de forma separada cada archivo mediante un paso de tipo *CSV file input*, utilizando el separador necesarios en cada caso (; o ,). Así pues, tendremos dos pasos, tal como se puede observar en la imagen de la derecha. En el caso del CSV de *ventas*, al tener registros de diferentes países, deberemos cambiar el tipo del ZIP (el código postal) a *String*.

A continuación, para unir los datos, dentro de la categoría *Join/Unión* utilizamos el paso *Merge join/Unión por clave*. En este caso tras arrastrar el paso al área de trabajo, vamos a enlazar desde el merge hacia los dos orígenes, en un caso como *left hand side stream of the join* y el otro como *right hand side stream of the join*:



### Caso de Uso 2 - Entradas de Merge

A continuación, editamos el *Merge* y le decimos que el campo de unión es *ProductID*. Cuando le damos a aceptar recibimos un mensaje de advertencia avisándonos que si los datos no están ordenados podemos obtener resultados incorrectos.



### Caso de Uso 2 - Aviso Merge

Así pues, vamos a añadir previo al merge un paso de ordenación a cada entrada, ordenando los datos por *ProductID* de manera ascendente. Para comprobar su funcionamiento, podemos hacer un *preview* del *merge*.

## Agrupando datos

Como el resultado final es un informe de ventas con el total de unidades vendidas y la cantidad recaudada agrupada por país y categoría del producto, necesitamos agrupar los datos.

Para ello, dentro de la categoría *Statistics*, utilizaremos el paso *Group by/Agrupar por*. En nuestro caso queremos agrupar por país (*Country*) y categoría (*Category*), y los datos que vamos a agregar son la suma de unidades (*Units*) y la suma recaudada (*Revenue*). Así pues, nuestra agrupación quedaría así:

Group by

Step name:

Include all rows? ☐

Temporary files directory:

TMP-file prefix:

Add line number, restart in each group ☐

Line number field name:

Always give back a result row ☐

The fields that make up the group:

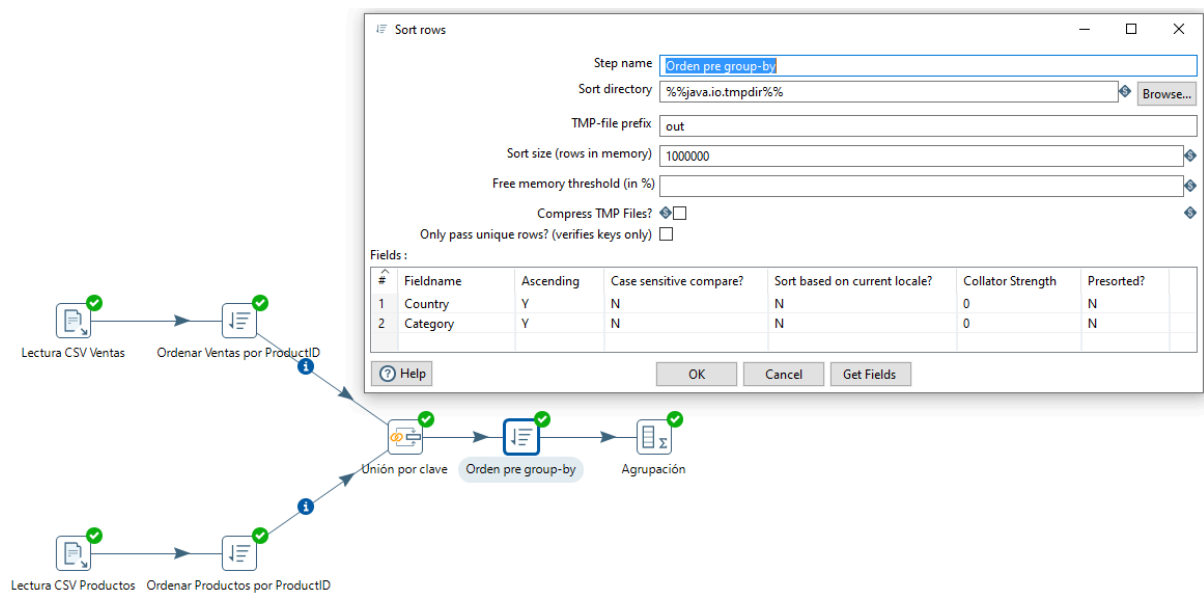
#	Group field
1	Country
2	Category

Aggregates:

#	Name	Subject	Type
1	TotalUnidades	Units	Sum
2	Recaudacion	Revenue	Sum

### Caso de Uso 2 - Agrupamos por país y categoría

En este caso, sucede lo mismo que antes, que este paso necesita los datos ordenados. Así pues, vamos a ordenar las salidas del merge por país y categoría.



### Caso de Uso 2 - Ordenamos antes de agrupar

El último paso que nos queda es exportar los datos a un fichero de texto mediante el paso *Text file output*.