

Ejercicio 4 - Cuestionarios Airbnb

Para el siguiente caso de uso, vamos a utilizar datos de los cuestionarios de Airbnb que se pueden descargar desde

<https://insideairbnb.com/get-the-data/>

En concreto, nos vamos a centrar en los datos de Málaga (**listings.csv**).

Uniendo datos

Ya hemos visto la opción para leer archivos CSV. Cuando queremos cargar varios ficheros en el mismo paso, podemos utilizar, dentro de *Input*, la opción de *Text File Input/Entrada Fichero de Texto*.

En nuestro caso, solo tenemos los datos de Málaga, pero podemos descargar los archivos del resto de ciudades de Andalucía (por ejemplo) para cargarlos juntos en este paso. Esto es opcional, pero aunque carguemos solo un fichero, utilizaremos *Text File Input/Entrada Fichero de Texto*:

[illegible]

Caso de Uso 3 - Filtrado compuesto

Recordad que antes, en la pestaña *Content/Contenido* elegiremos la , como separador de campos y en *Fields/Campos*, tenemos que obtener los campos a leer:

Entrada Fichero de Texto

Nombre paso: **Entrada Fichero de Texto**

Fichero | Contenido | Manejo de Errores | Filtros | Campos | Additional output fields

#	Nombre	Tipo	Formato	Posición	Longitud	Precisión	Moneda	Decimal	Grupo	Nulo si	Por defecto
1	id	Integer	#		15	0	EUR	.	,	-	
2	name	String			50		EUR	.	,	-	
3	host_id	Integer	#		15	0	EUR	.	,	-	
4	host_name	String			22		EUR	.	,	-	
5	neighbourhood_group	String			0		EUR	.	,	-	
6	neighbourhood	String			20		EUR	.	,	-	
7	latitude	BigNumber	#,#####		17	14	EUR	.	,	-	
8	longitude	BigNumber	#,#####		19	16	EUR	.	,	-	
9	room_type	String			15		EUR	.	,	-	
10	price	Integer	#		15	0	EUR	.	,	-	
11	minimum_nights	Integer	#		15	0	EUR	.	,	-	
12	number_of_reviews	Integer	#		15	0	EUR	.	,	-	
13	last_review	Date	yyyy-MM-dd				EUR	.	,	-	
14	reviews_per_month	BigNumber	#,##		4	2	EUR	.	,	-	
15	calculated_host_listings_count	Integer	#		15	0	EUR	.	,	-	
16	availability_365	Integer	#		15	0	EUR	.	,	-	
17	number_of_reviews_ltm	Integer	#		15	0	EUR	.	,	-	
18	license	String			57		EUR	.	,	-	

Traer Campos Minimal width

Caso de Uso 3 - Campos

Nos vamos a quedar con un subconjunto de las columnas y las vamos a renombrar. Para ello, dentro de la categoría *Transform* elegimos el paso *Select values/Selecciona/Renombrar valores* y elegimos y renombramos los siguientes campos:

- **id**, **room_type**, **neighborhood** y **price** pasarán a ser **habitacion_id**, **habitacion_tipo**, **barrio** y **precio**.

Selecciona/Renombrar valores

Nombre paso: **Selecciona/Renombrar valores**

Selecciona & Modifica | Eliminar | Meta-información

Campos:

#	Nombre campo	Renombrar a	Longitud	Precisión
1	id	habitacion_id		
2	room_type	habitacion_tipo		
3	neighbourhood	barrio		
4	price	precio		

Obtener campos a seleccionar Edit Mapping

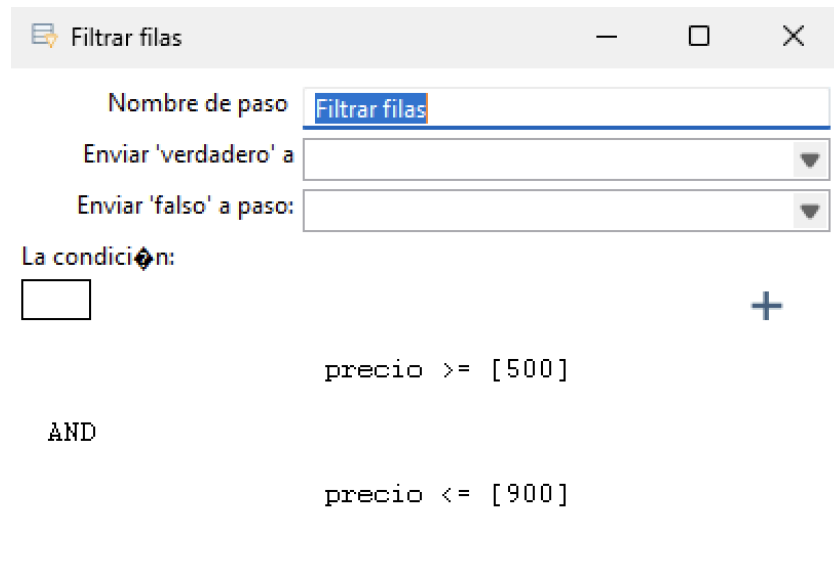
Include unspecified fields, ordered by name ☐

Help Vale Cancelar

Caso de Uso 3 - Selección y nombrado de campos

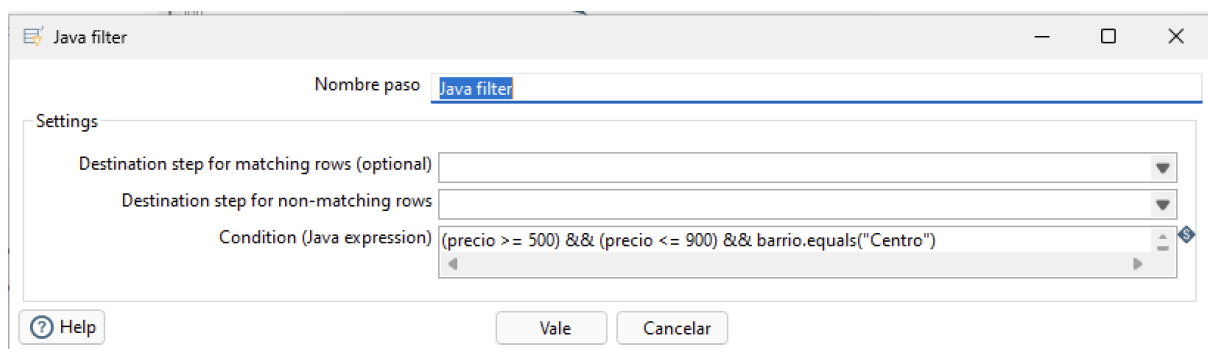
Filtrado compuesto

El siguiente paso que vamos a hacer es quedarnos con aquellos registros cuyo precio esté entre 500 y 900. Así pues, con el paso *Filter Rows/Filtrar filas* realizaremos:



Caso de Uso 3 - Filtrado compuesto

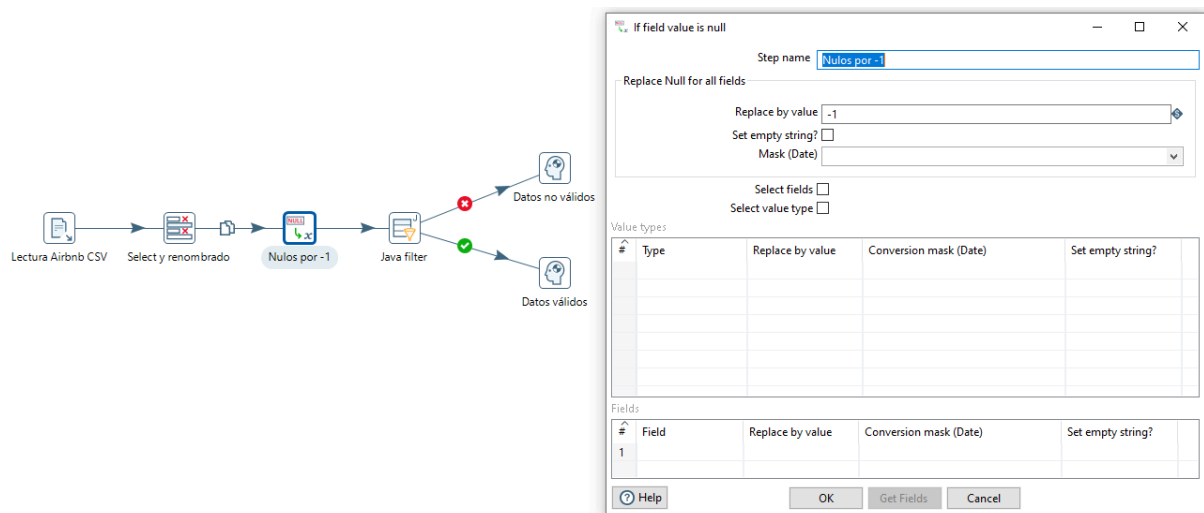
Si el filtrado fuese con condiciones más complejas, en ocasiones es más sencillo utilizar el paso *Java filter* (de la categoría *Flow*), el cual utilizando la notación de Java, podemos indicar la condición a cumplir. Por ejemplo, vamos a filtrar los registros cuyo precio esté entre 500 y 900, y que su barrio sea "Centro" → `(precio >= 500) && (precio <= 900) && barrio.equals("Centro")`:



Caso de Uso 3 - Filtrado Java

Para comprobar su funcionamiento, vamos a añadir un par de pasos *dummy/transformación simulada* (no realizan nada, pero sirven para finalizar tareas). Al ejecutarlo, veremos que nos da un error. Si algún dato es nulo, el filtrado Java provocará un error de transformación.

Una posibilidad es que introduzcamos un paso de la categoría *Utility* denominado *If value is null*. Con este paso, podemos indicar el valor a tomar a todos los campos o hacerlo de forma concreta en los campos que queramos. En nuestro caso, vamos a indicar que cambie todos los nulos por **-1**.



Caso de Uso 3 - Cambiando nulos por -1

Debemos tener en cuenta que como ahora podemos tener precios con **-1**, para evitar recogerlos en el filtrado Java, deberíamos indicarlo en las condiciones. En nuestro ejemplo no tiene sentido, ya que los valores van a estar comprendidos entre 500 y 900, pero en otros filtros (por ejemplo, si solo buscamos los barrios del centro), habría que añadir:

```
(precio>=0) && barrio.equals("Centro")
```

Generando JSON

JSON output

Step name: fichero JSON

General Fields

Operation: Write to file

Settings

Json bloc name: datos

Nr rows in a bloc: 0

Output Value: outputValue

Compatibility mode: ☐

Output File

Filename: \${Internal.Entry.Current.Directory}/a Browse...

Append: ☐

Create Parent folder: ☐

Do not open create at start: ☐

Extension: json

Encoding: UTF-8

Pass output to servlet: ☐

Include date in filename?: ☒

Include time in filename?: ☐

Show filename(s)...

Add File to result filenames: ☐

Help OK Cancel

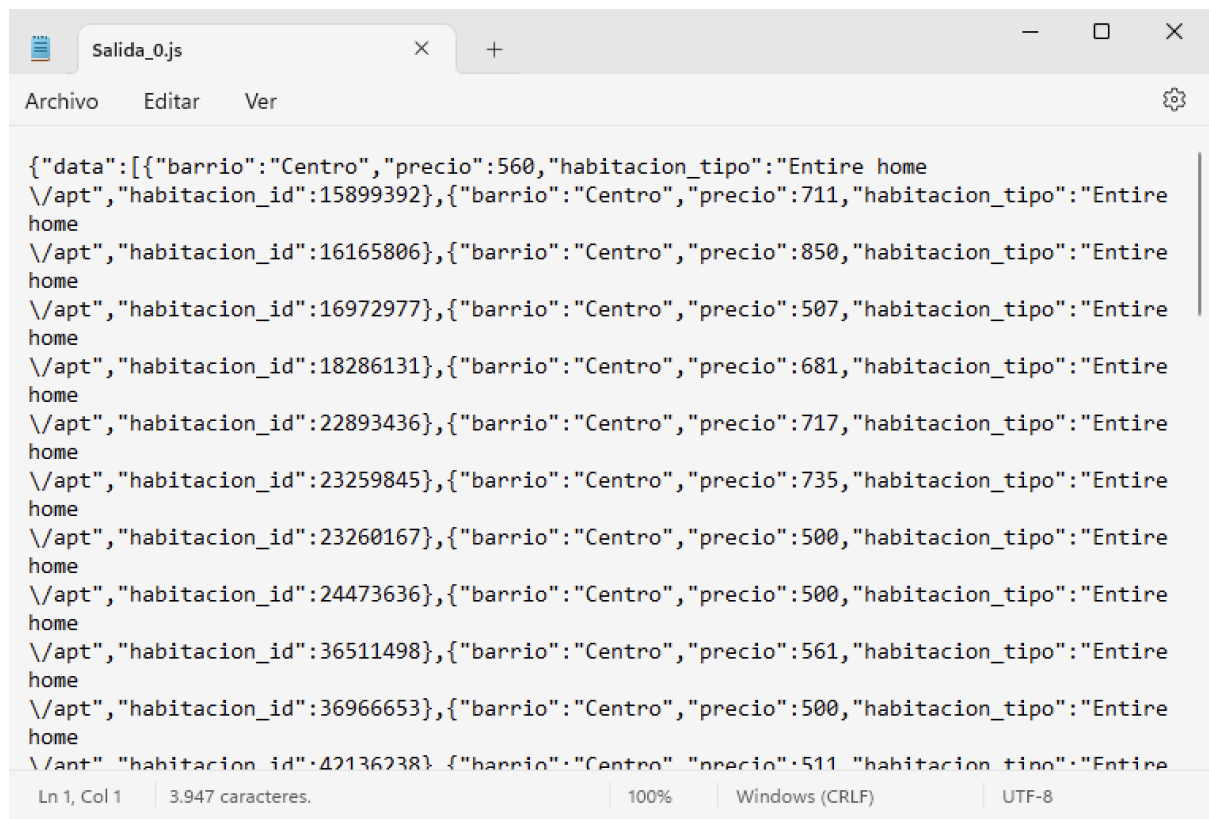
Caso de Uso 3 - Configuración JSON

Finalmente queremos almacenar los datos que cumplen el filtro en un fichero JSON.

Para ello, sustituimos el *dummy* del camino exitoso por un paso *JSON output*. Tras seleccionar los campos que nos interesan, configuraremos:

- *Filename*: La ruta y el nombre del archivo
- *Json bloc name*: nombre de la propiedad que contendrá un objeto o un array de objetos con los datos.
- *Nr rows in a bloc*: Cantidad de datos del archivo. Si ponemos 0, coloca todos los datos en el mismo fichero. Si ponemos 1, generará un fichero por cada registro.

También es importante que, en la pestaña Fields, traigamos los campos. De lo contrario, el archivo JSON estará vacío.

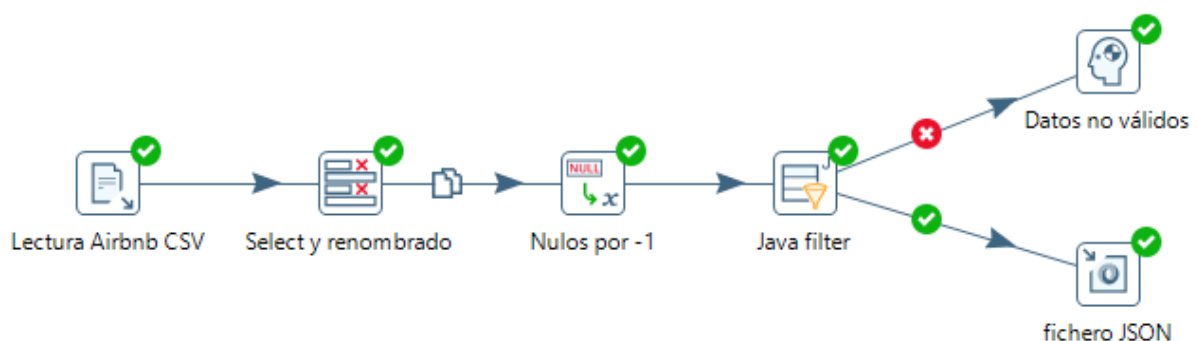


The screenshot shows a code editor window titled 'Salida_0.js'. The editor contains a JSON array of 15 Airbnb listings. Each listing is an object with fields: 'barrio' (all 'Centro'), 'precio' (various values), and 'habitacion_tipo' (all 'Entire home'). The status bar at the bottom indicates 'Ln 1, Col 1', '3.947 caracteres', '100%', 'Windows (CRLF)', and 'UTF-8'.

```
{
  "data": [
    {
      "barrio": "Centro",
      "precio": 560,
      "habitacion_tipo": "Entire home",
      "habitacion_id": 15899392
    },
    {
      "barrio": "Centro",
      "precio": 711,
      "habitacion_tipo": "Entire home",
      "habitacion_id": 16165806
    },
    {
      "barrio": "Centro",
      "precio": 850,
      "habitacion_tipo": "Entire home",
      "habitacion_id": 16972977
    },
    {
      "barrio": "Centro",
      "precio": 507,
      "habitacion_tipo": "Entire home",
      "habitacion_id": 18286131
    },
    {
      "barrio": "Centro",
      "precio": 681,
      "habitacion_tipo": "Entire home",
      "habitacion_id": 22893436
    },
    {
      "barrio": "Centro",
      "precio": 717,
      "habitacion_tipo": "Entire home",
      "habitacion_id": 23259845
    },
    {
      "barrio": "Centro",
      "precio": 735,
      "habitacion_tipo": "Entire home",
      "habitacion_id": 23260167
    },
    {
      "barrio": "Centro",
      "precio": 500,
      "habitacion_tipo": "Entire home",
      "habitacion_id": 24473636
    },
    {
      "barrio": "Centro",
      "precio": 500,
      "habitacion_tipo": "Entire home",
      "habitacion_id": 36511498
    },
    {
      "barrio": "Centro",
      "precio": 561,
      "habitacion_tipo": "Entire home",
      "habitacion_id": 36966653
    },
    {
      "barrio": "Centro",
      "precio": 500,
      "habitacion_tipo": "Entire home",
      "habitacion_id": 42136238
    },
    {
      "barrio": "Centro",
      "precio": 511,
      "habitacion_tipo": "Entire home",
      "habitacion_id": 42136238
    },
    {
      "barrio": "Centro",
      "precio": 511,
      "habitacion_tipo": "Entire home",
      "habitacion_id": 42136238
    },
    {
      "barrio": "Centro",
      "precio": 511,
      "habitacion_tipo": "Entire home",
      "habitacion_id": 42136238
    },
    {
      "barrio": "Centro",
      "precio": 511,
      "habitacion_tipo": "Entire home",
      "habitacion_id": 42136238
    }
  ]
}
```

Resultado final

En la siguiente imagen puedes visualizar la transformación completa:



Caso de Uso 3 - Transformación final