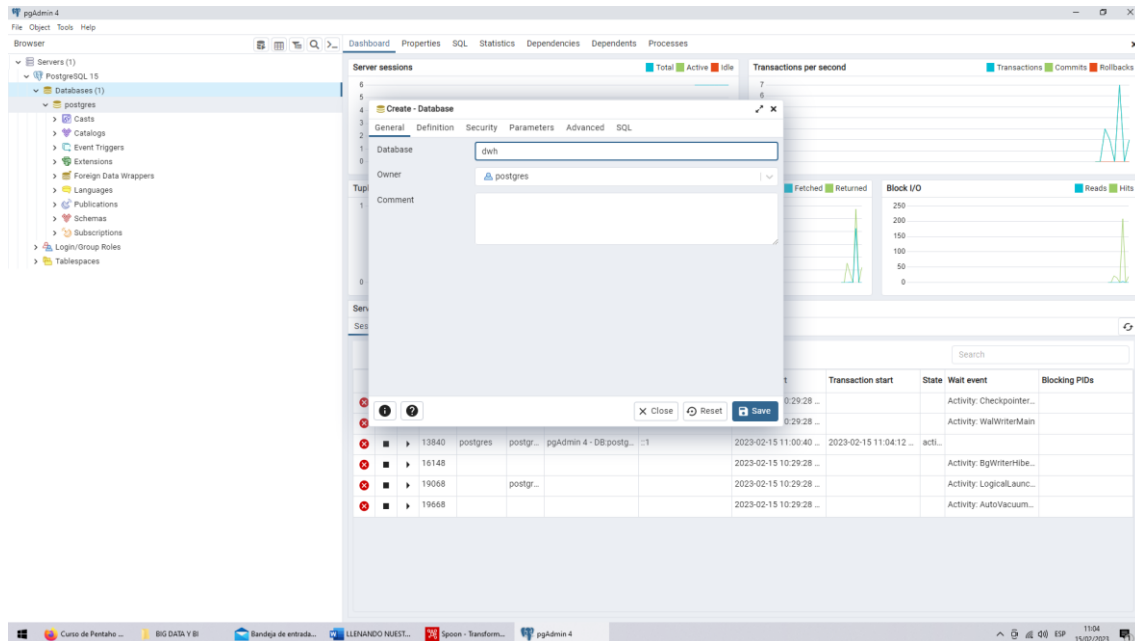


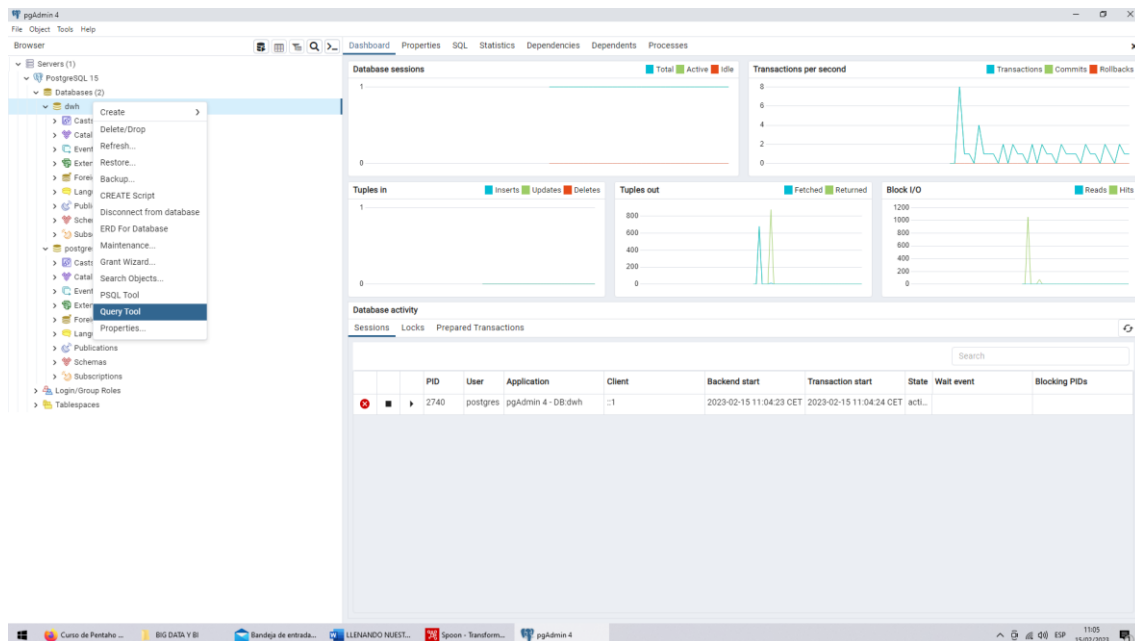
# LLENANDO NUESTRO DWH

Vamos a rellenar nuestro DWH con un proceso ETL leyendo datos desde Excel, CSV y BBDD en SQL. En primer lugar abrimos pgAdmin.

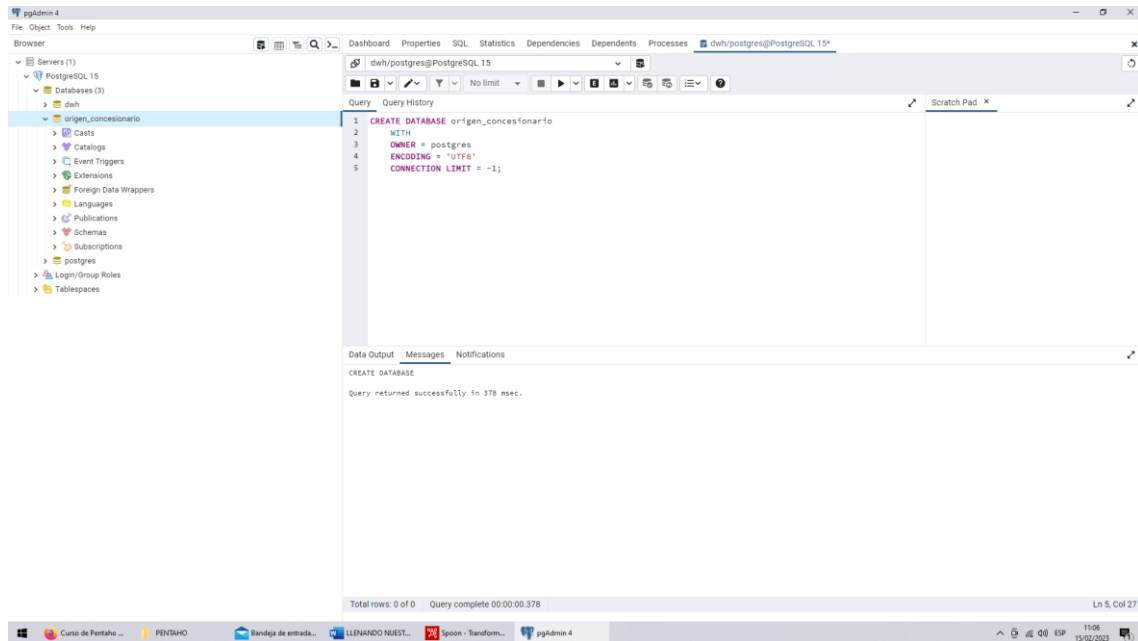
Primero creo la bbdd dwh



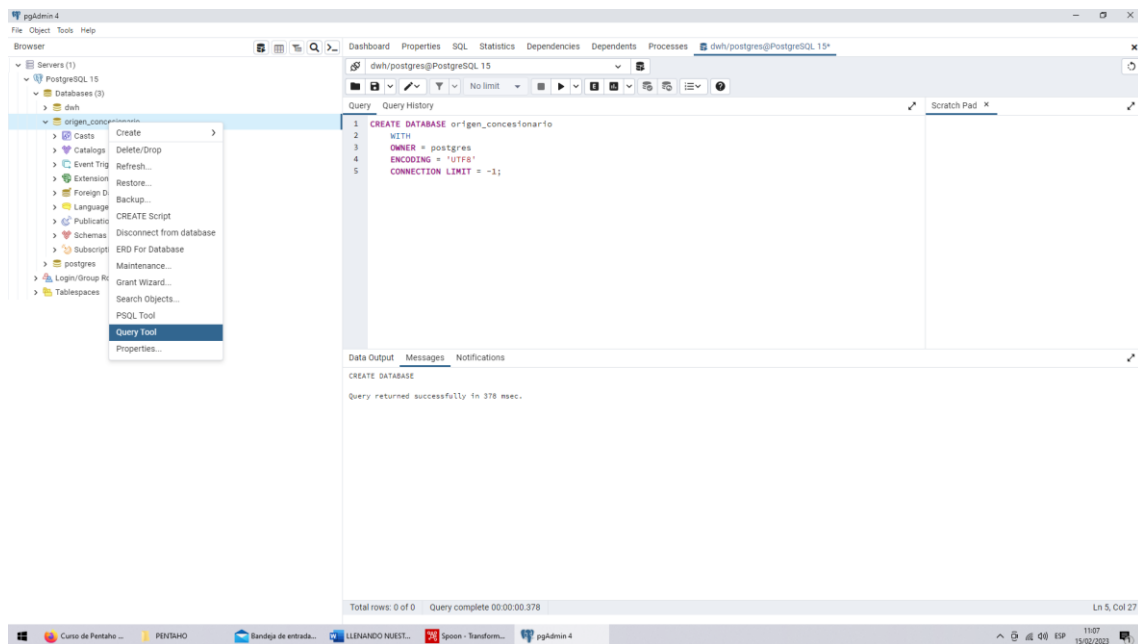
Y ejecutamos el script creacion\_dwh desde el query tool, es donde cargaremos los datos



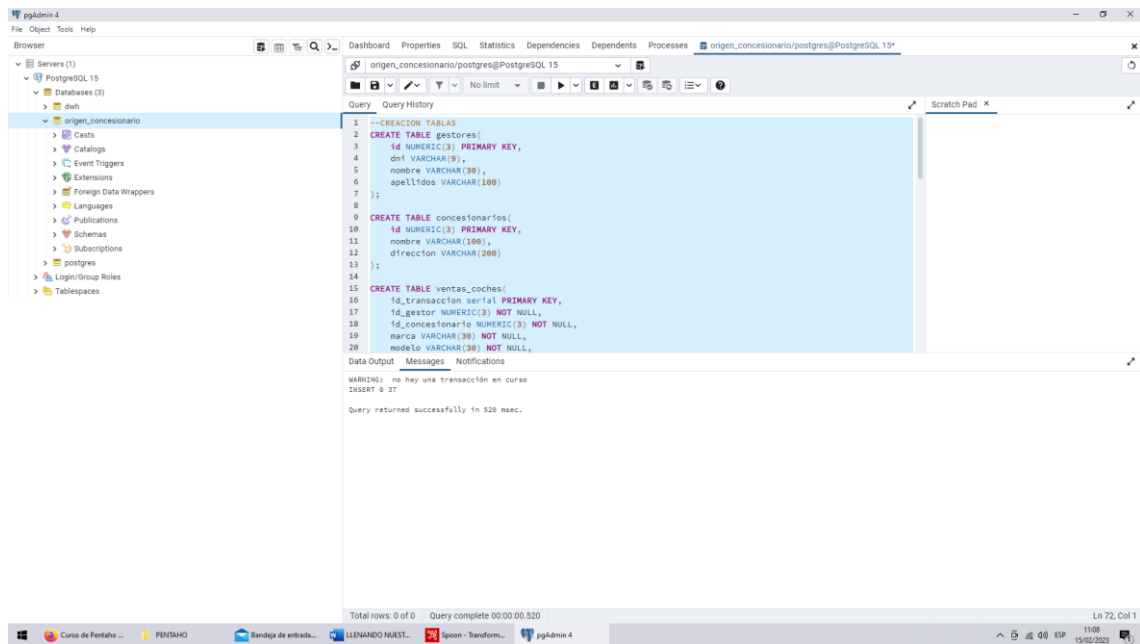
Despues creamos otra bbdd con crear\_origen\_concesionario.sql  
Arrastramos el primer script y ejecutamos en el query tool



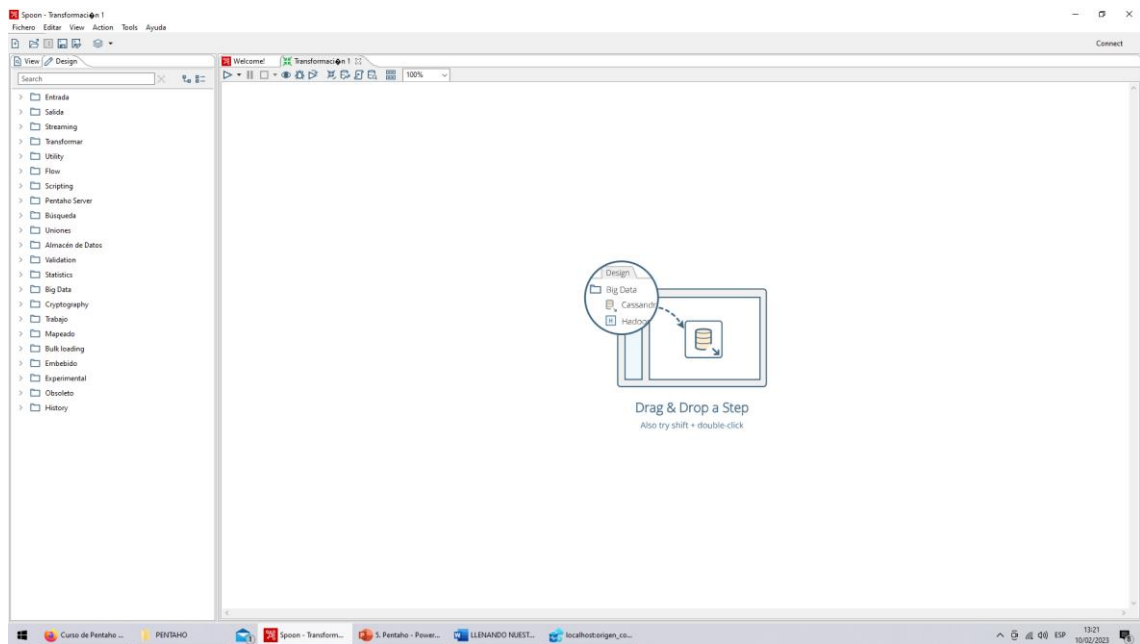
Abrimos el query tool de origen concesionario



Lanzamos el segundo script

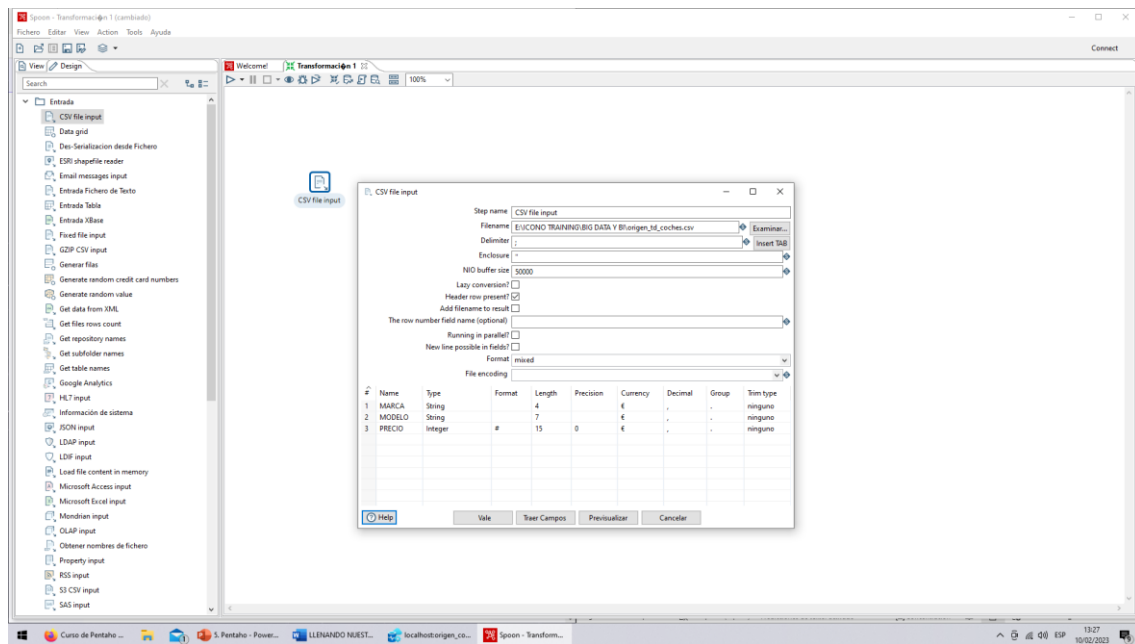


Abrimos PDI ejecutando el Spoon.bat

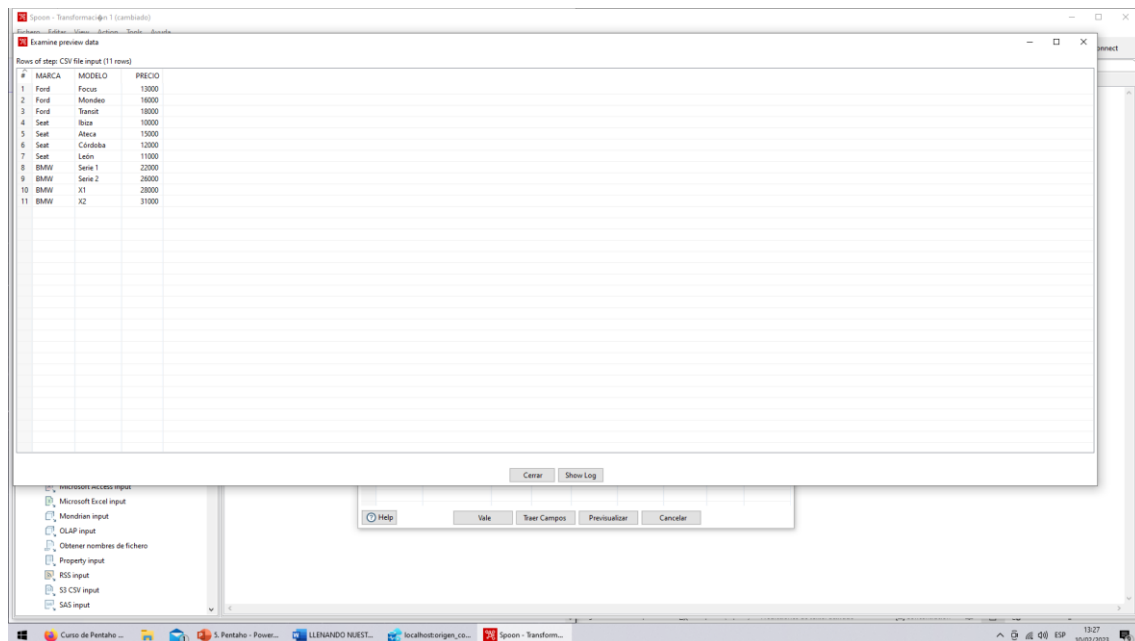


Creamos una primera transformación encargada de cargar las dimensiones. El primer origen de datos es un CSV.

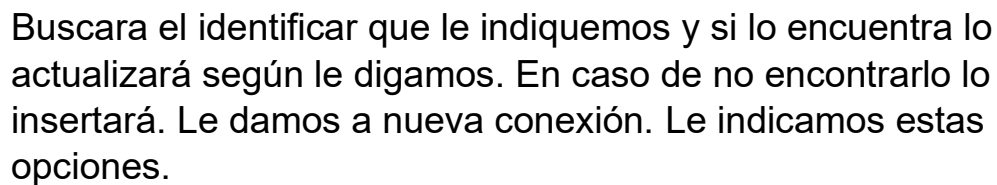
Agregamos el input de CSV y elegimos el archivo origen\_td\_coches.csv con delimitador por ;. Podemos darle a traer campos y ver si lo trae correctamente.

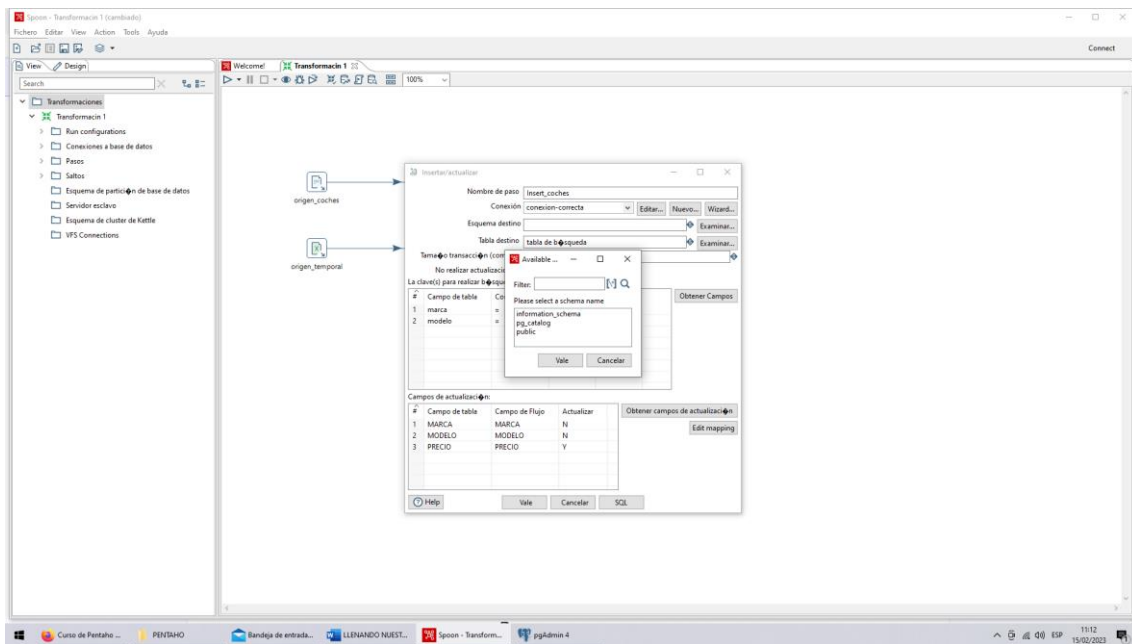


Y con previsualizar podemos ver los registros

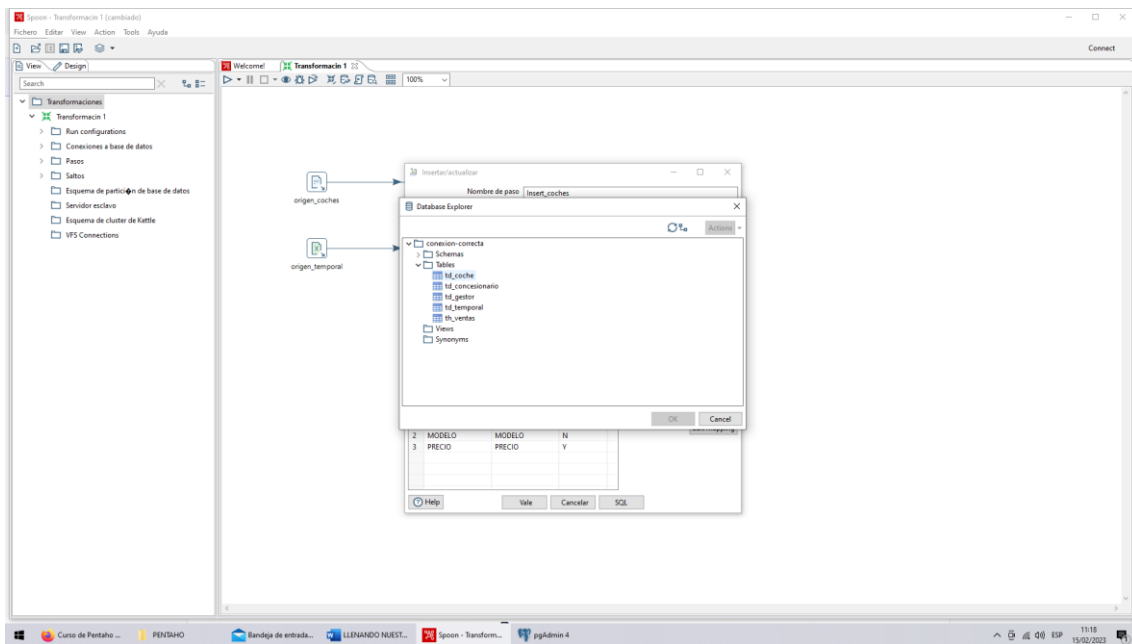


Insertamos un nuevo paso de Insertar/Actualizar

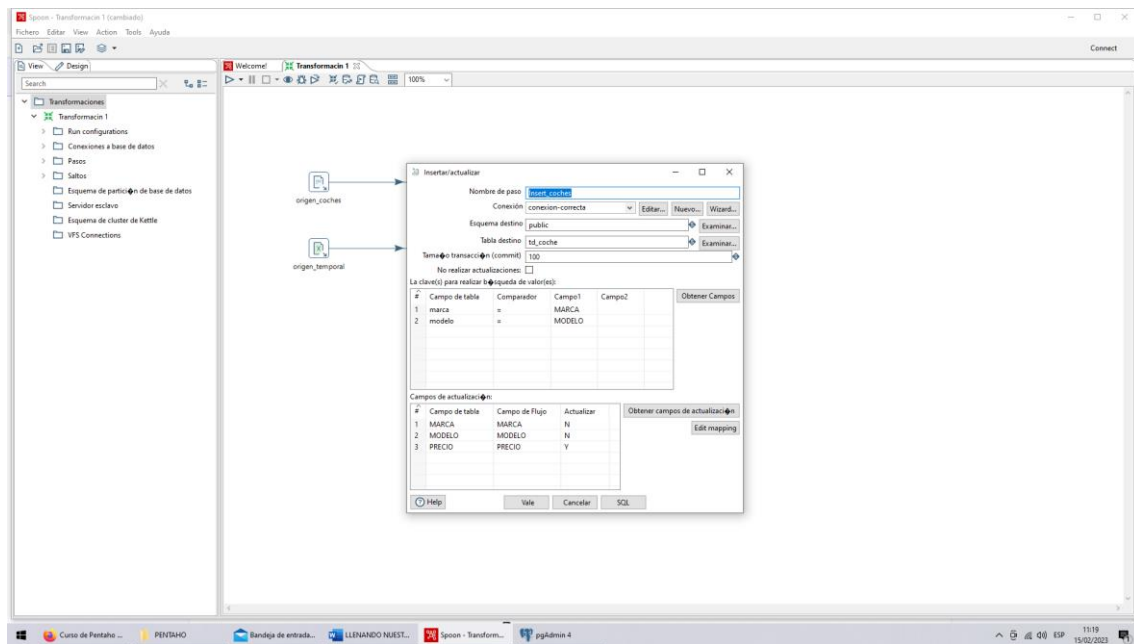




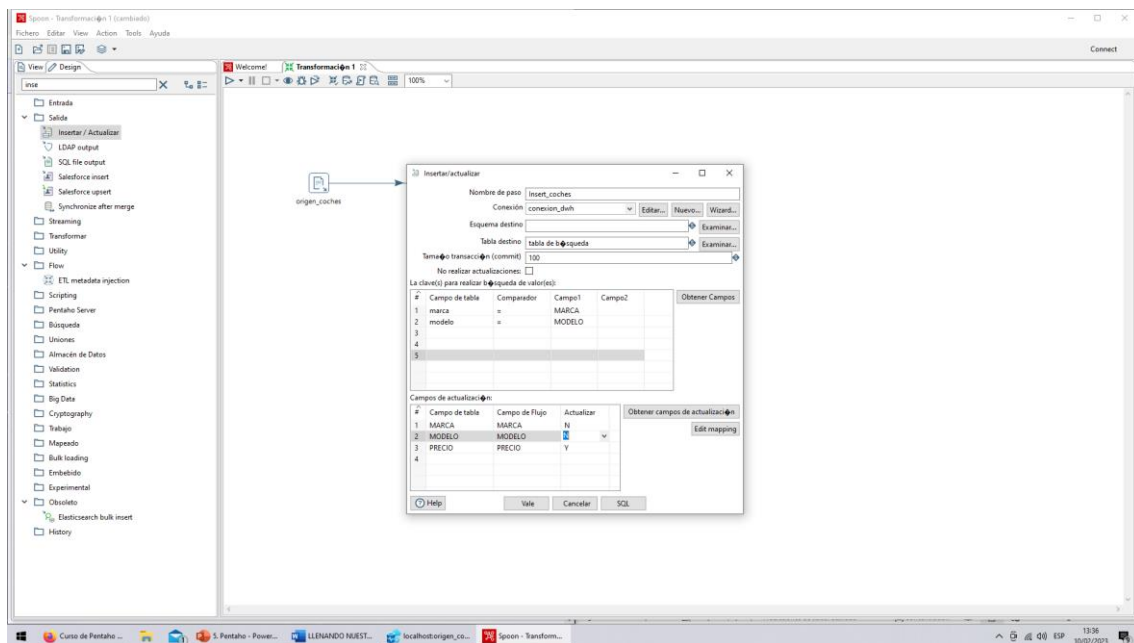
Le damos a la tabla td\_coche



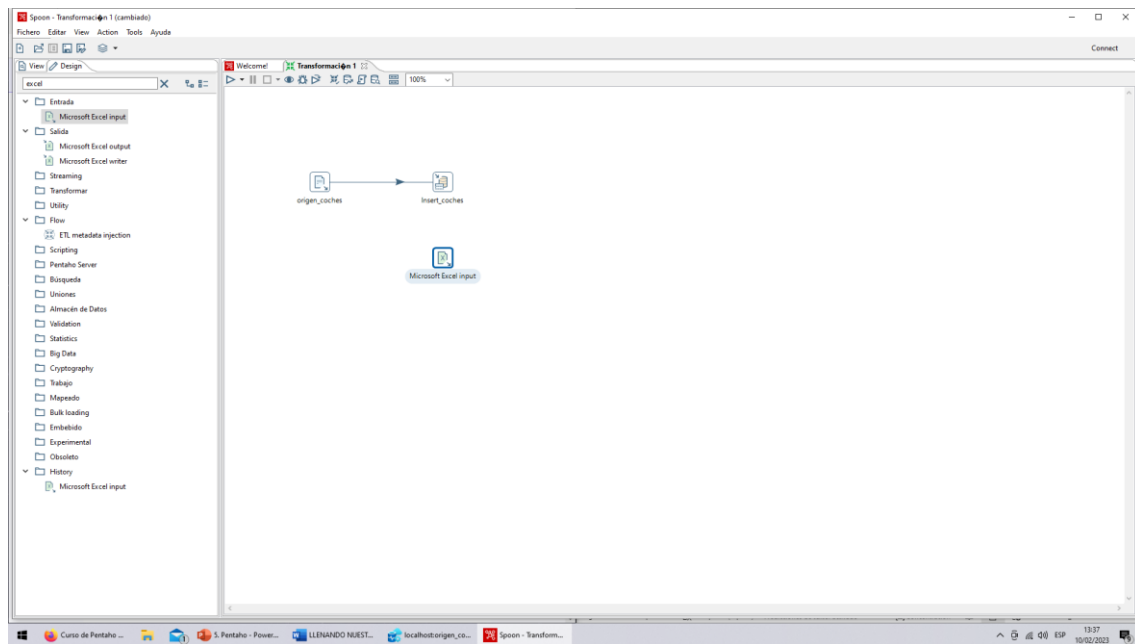
Despues le damos a obtener campos y dejamos la siguiente modificacion



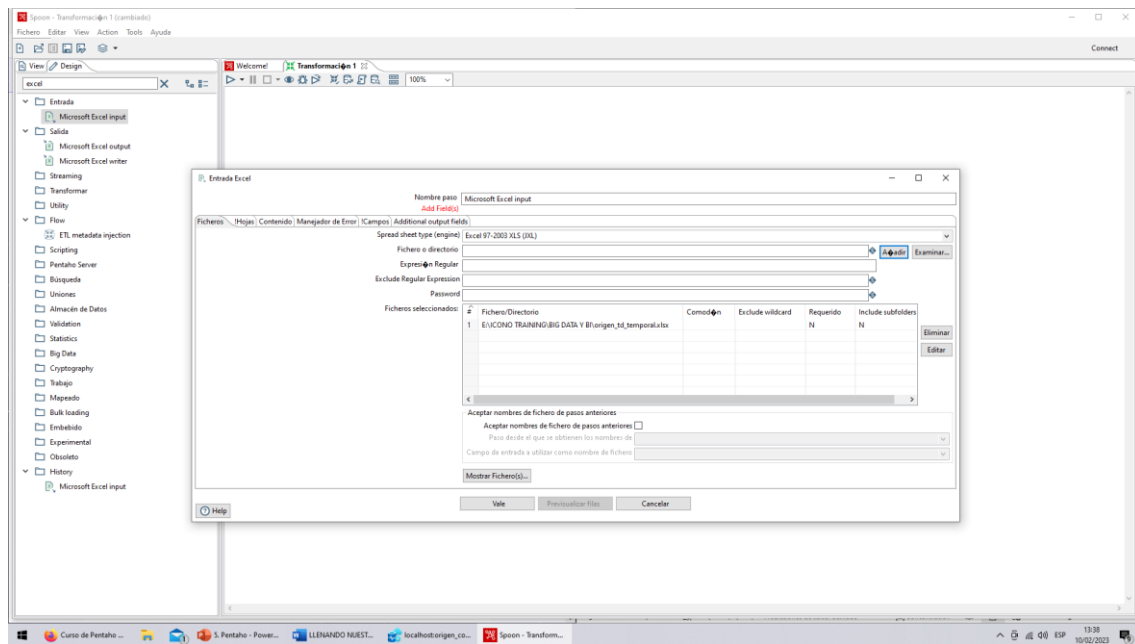
Y le damos a obtener campos actualizados donde tendremos que cambiar dos Y por dos N y dejar solo en Y el precio



Agregamos ahora el siguiente origen que es un excel que incluye la dimensión temporal

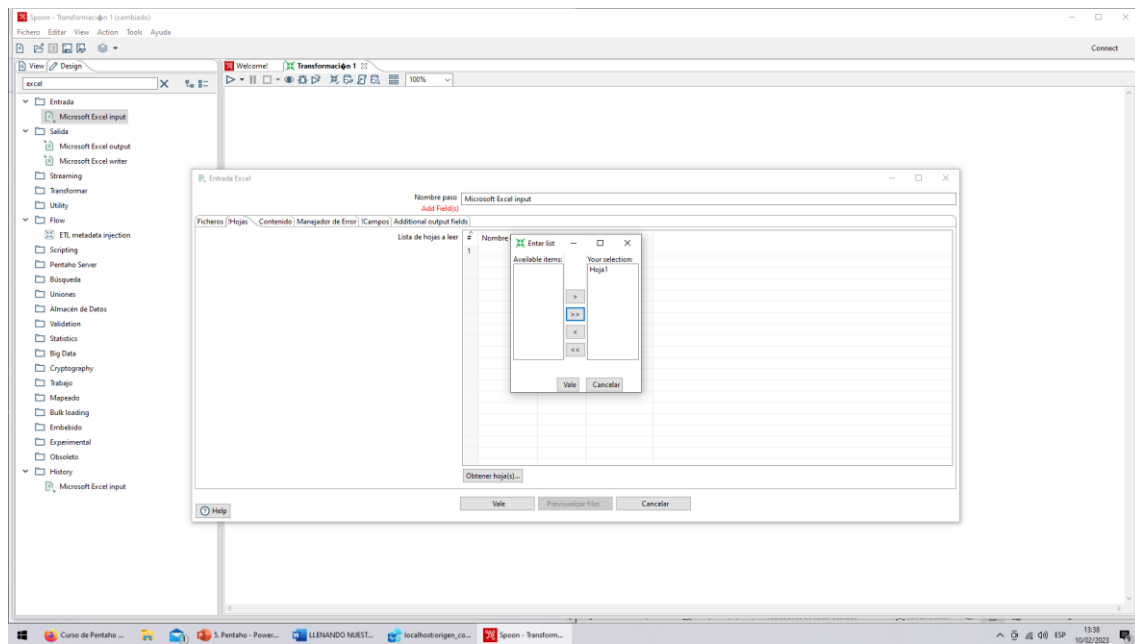


Añadimos importante poner que es de 2007

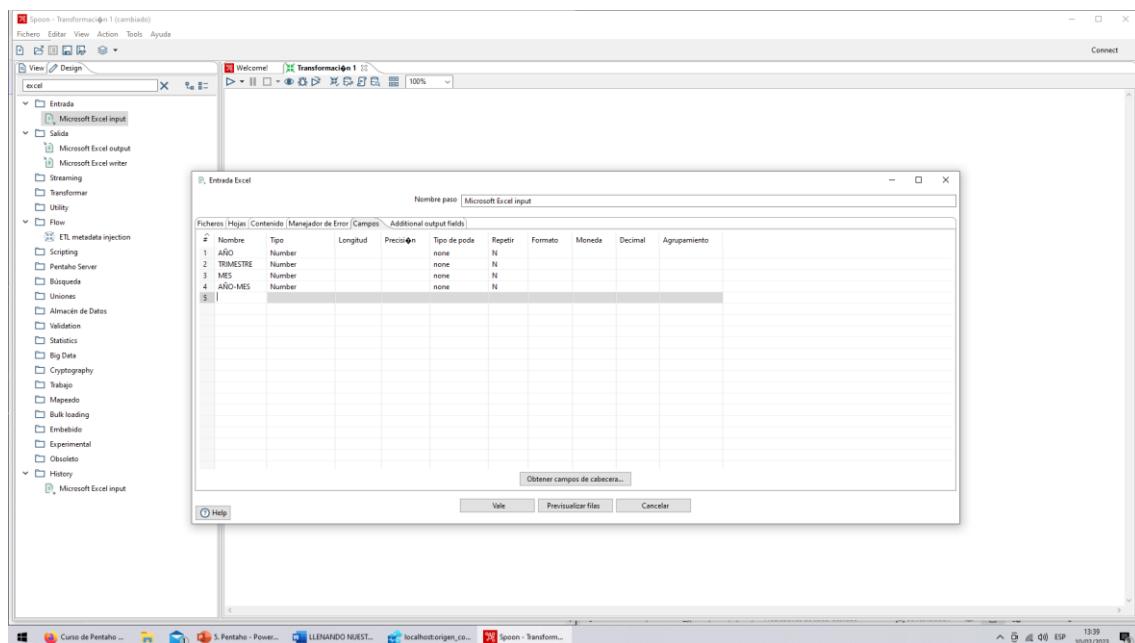


Obtenemos hojas

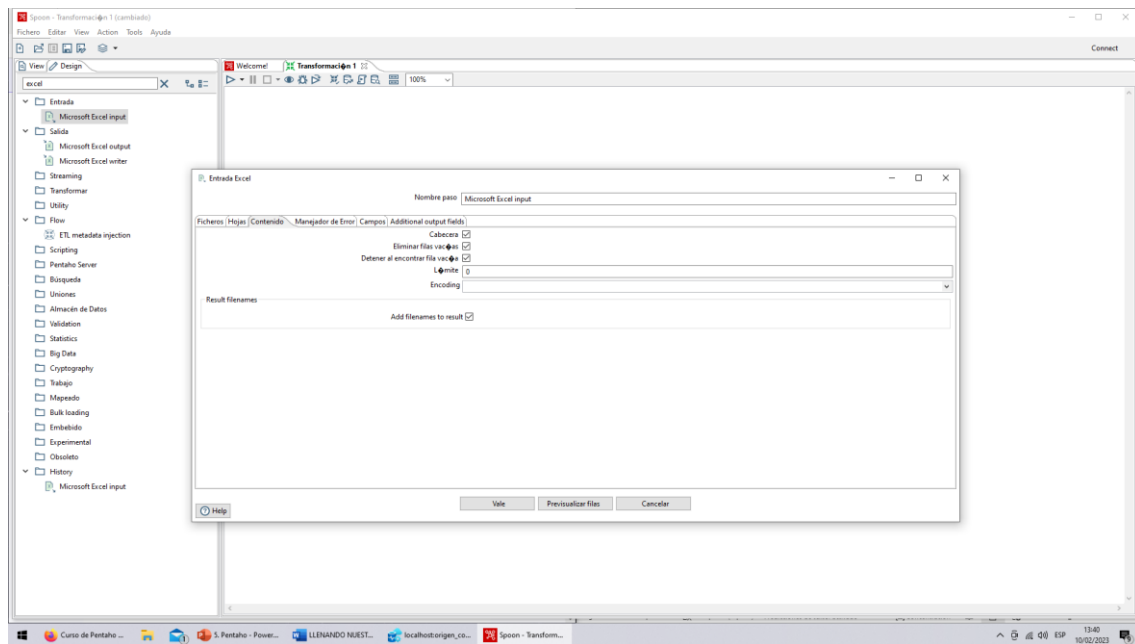




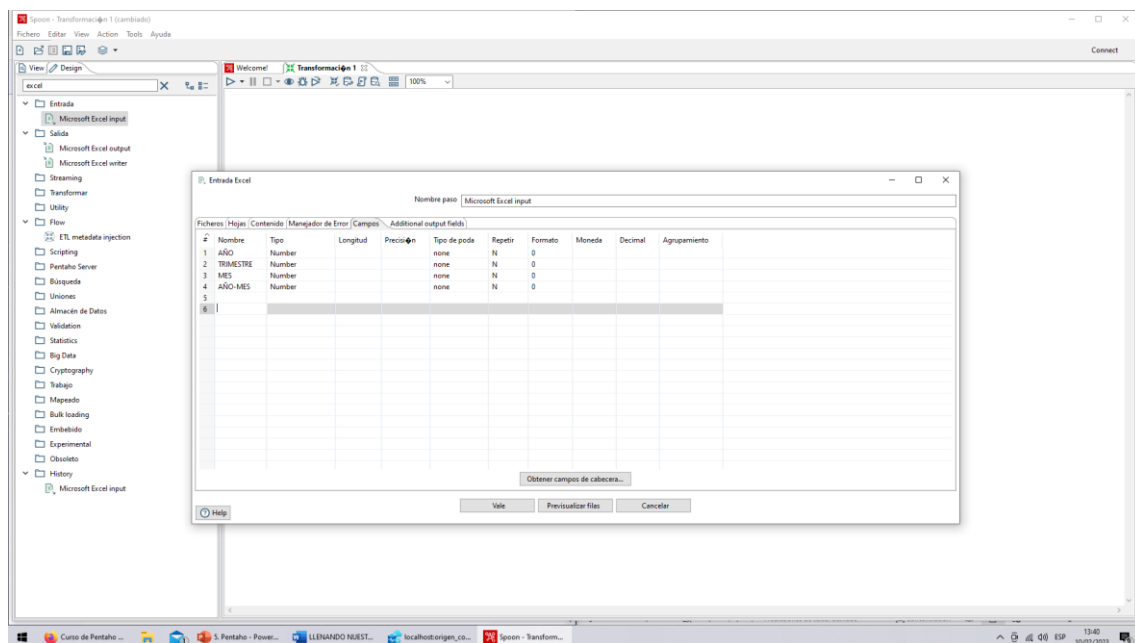
Obtenemos los campos



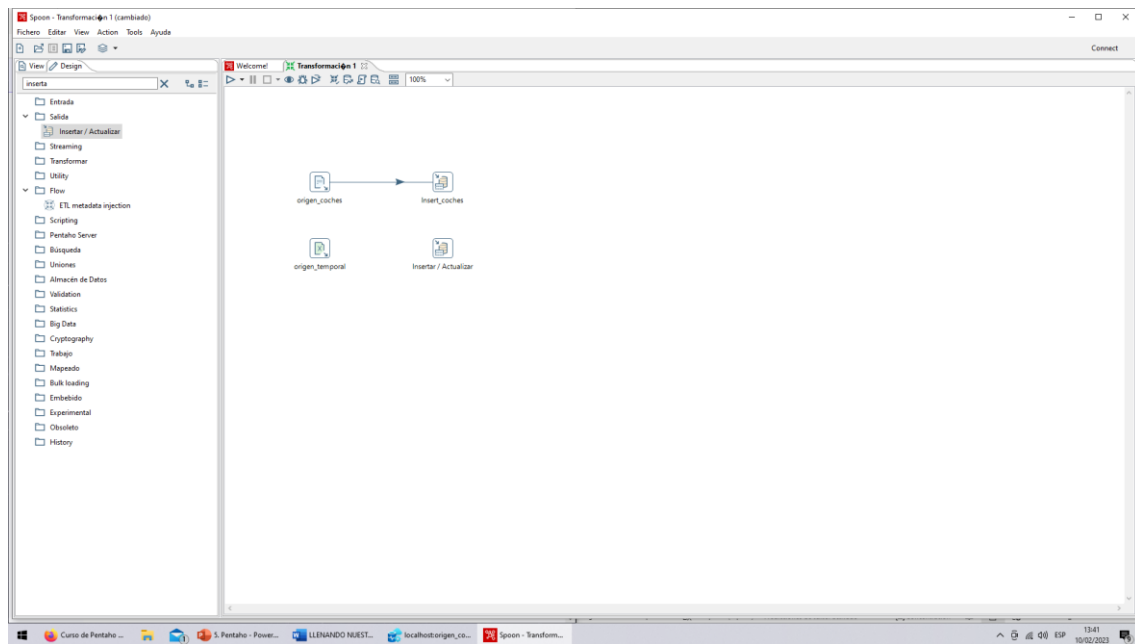
Y dejamos las siguientes opciones



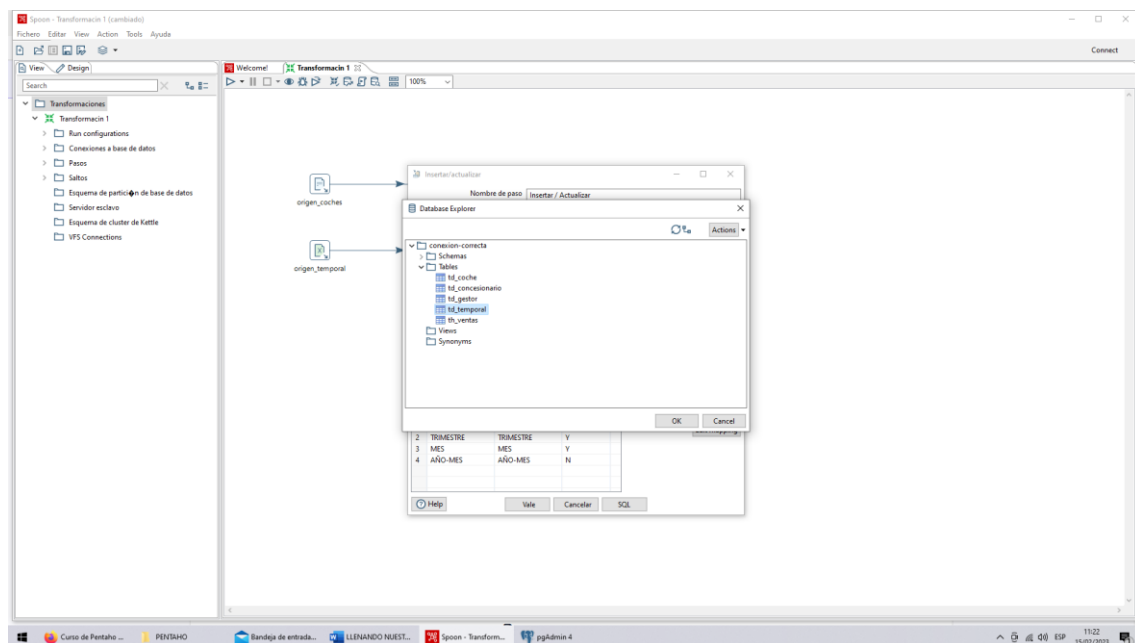
Le ponemos un 0 al formato para quitar el decimal

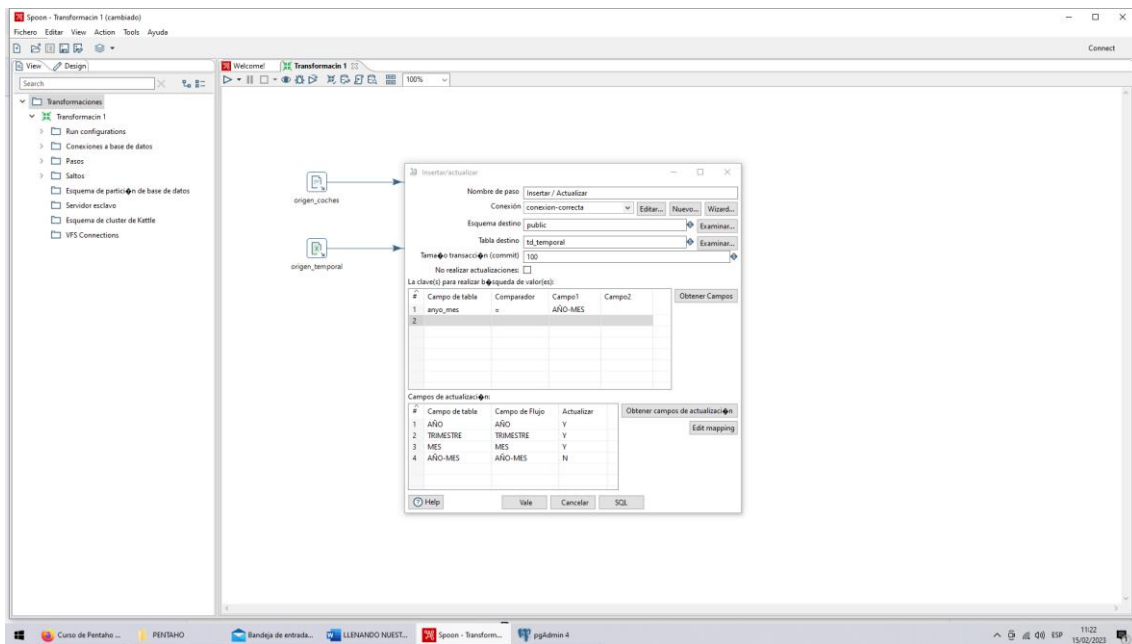


Agregamos otro insertar/actualizar

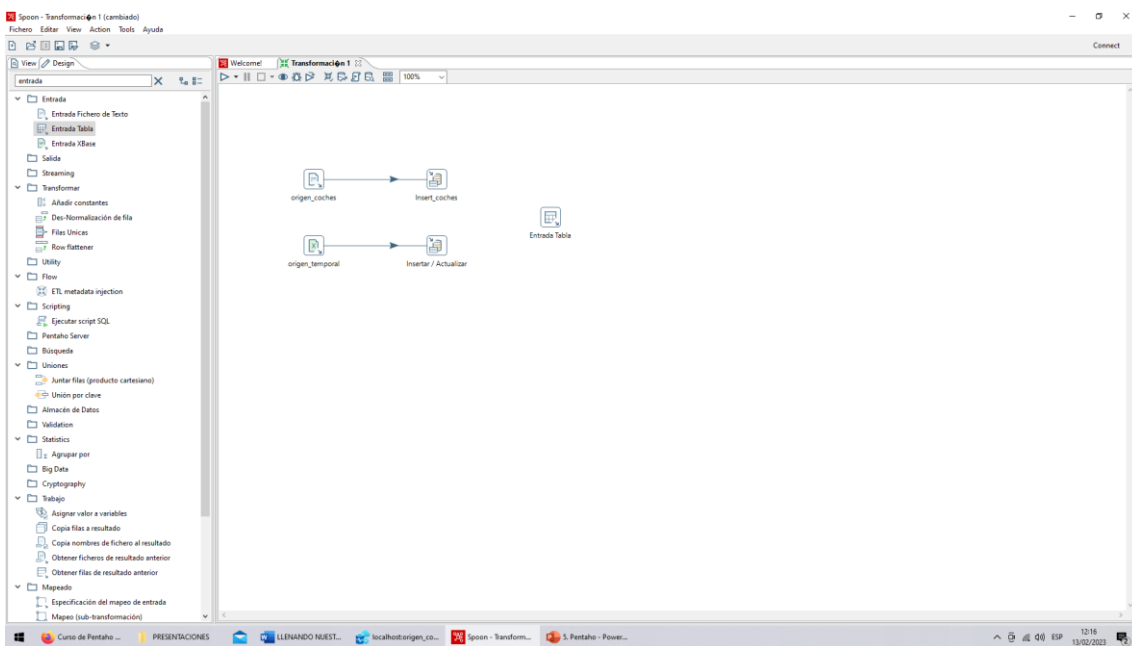


Con las siguientes condiciones

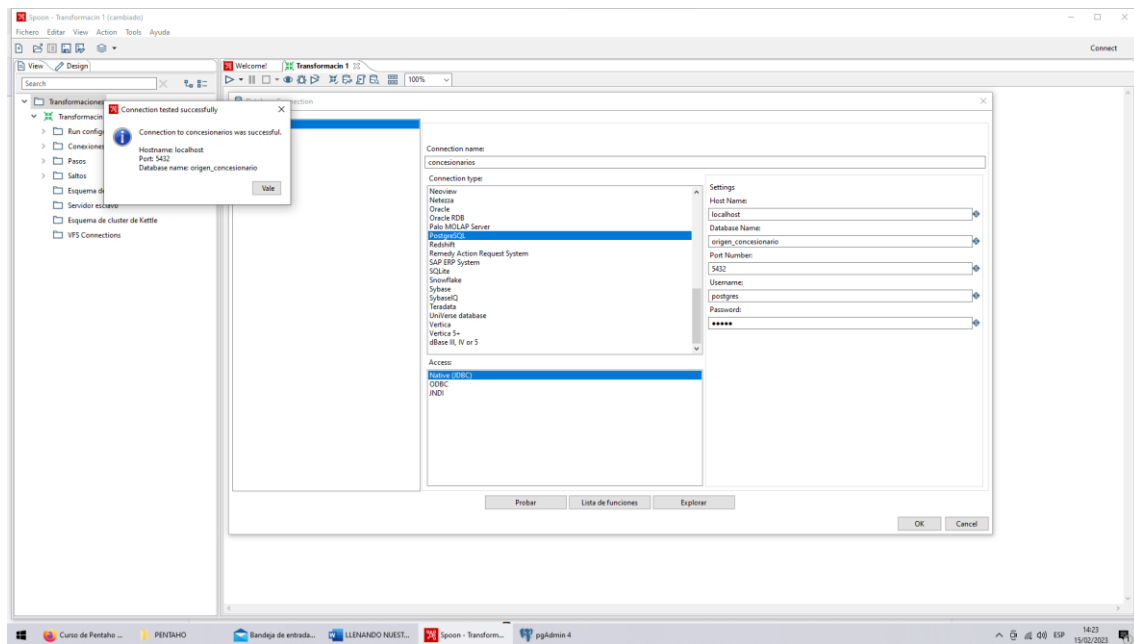




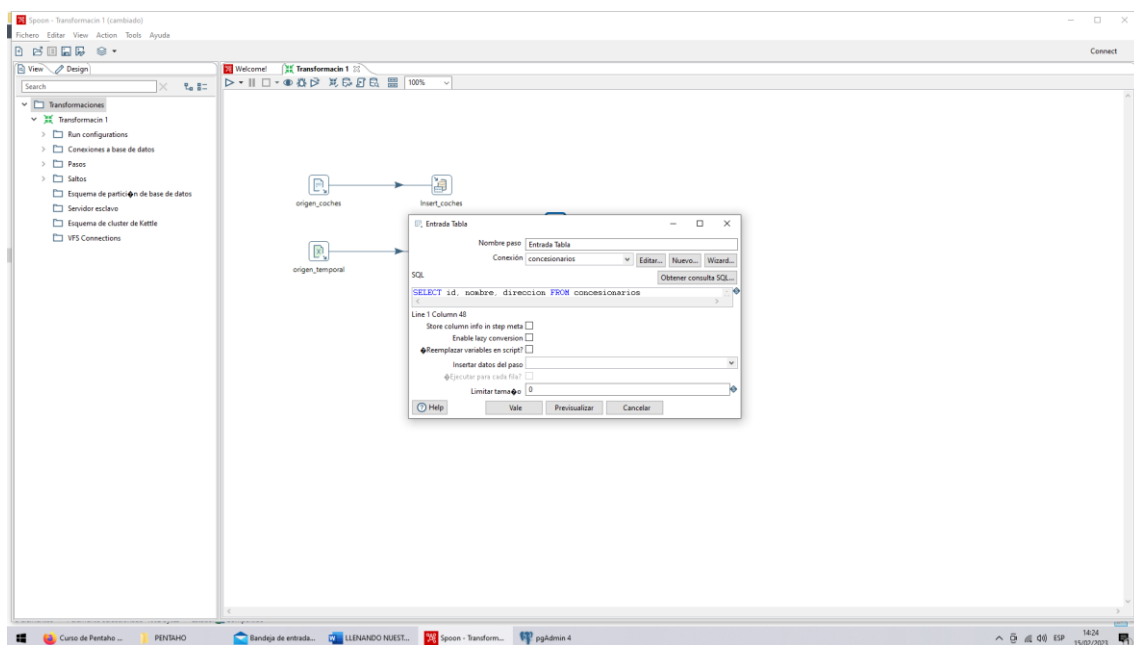
Buscamos el paso de entrada de tabla



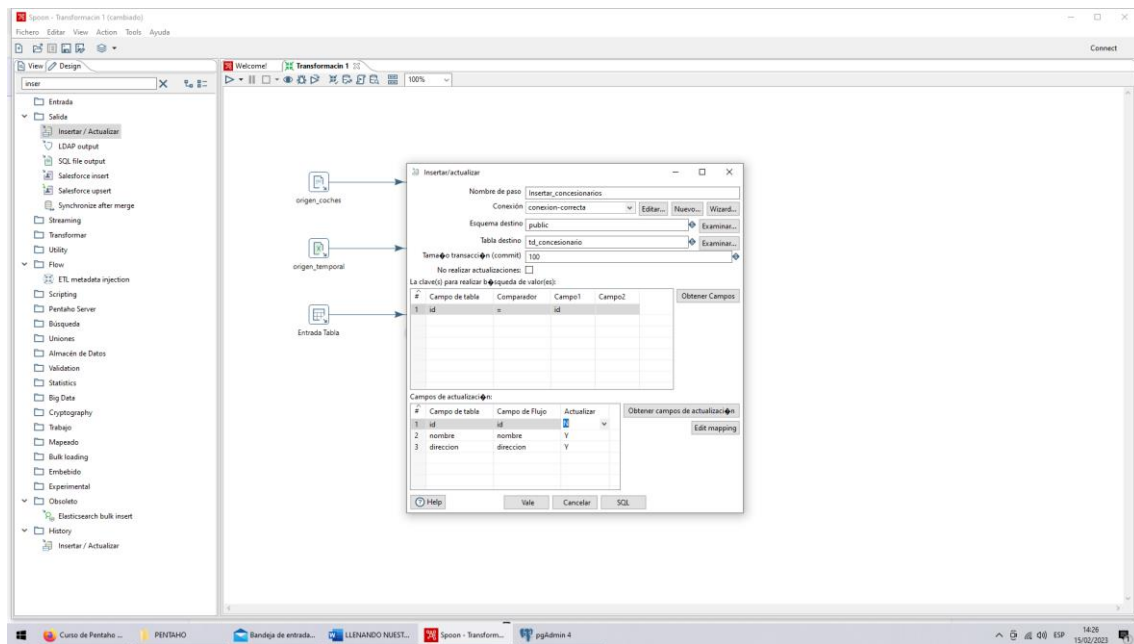
Creamos en el una nueva conexión que apunte a la tabla de concesionarios



Y ponemos la siguiente select

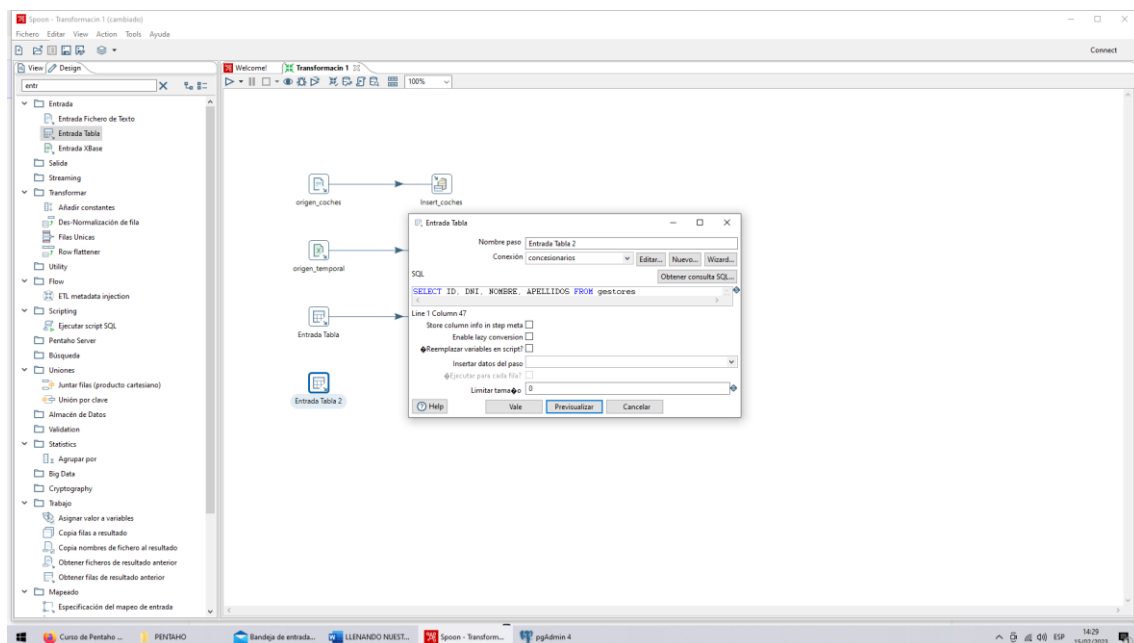


Agregamos otro paso de insertar/actualizar



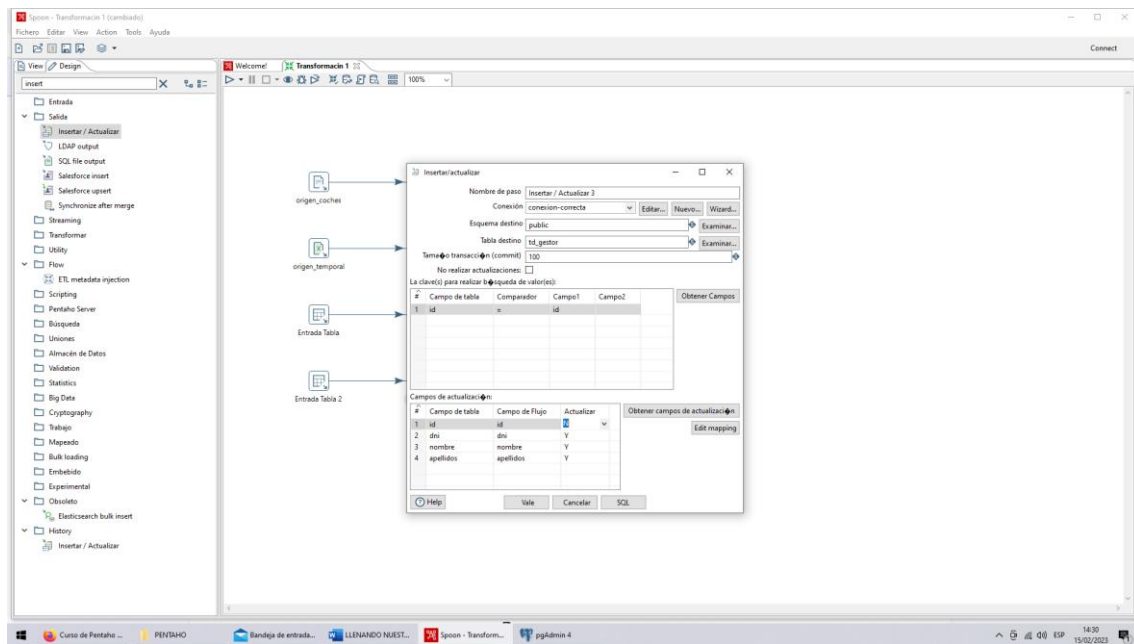
Con conexión al dwh esquema public, la tabla concesionario, solo el id, e importante el N!

Repetimos lo mismo para la ultima dimension de gestores

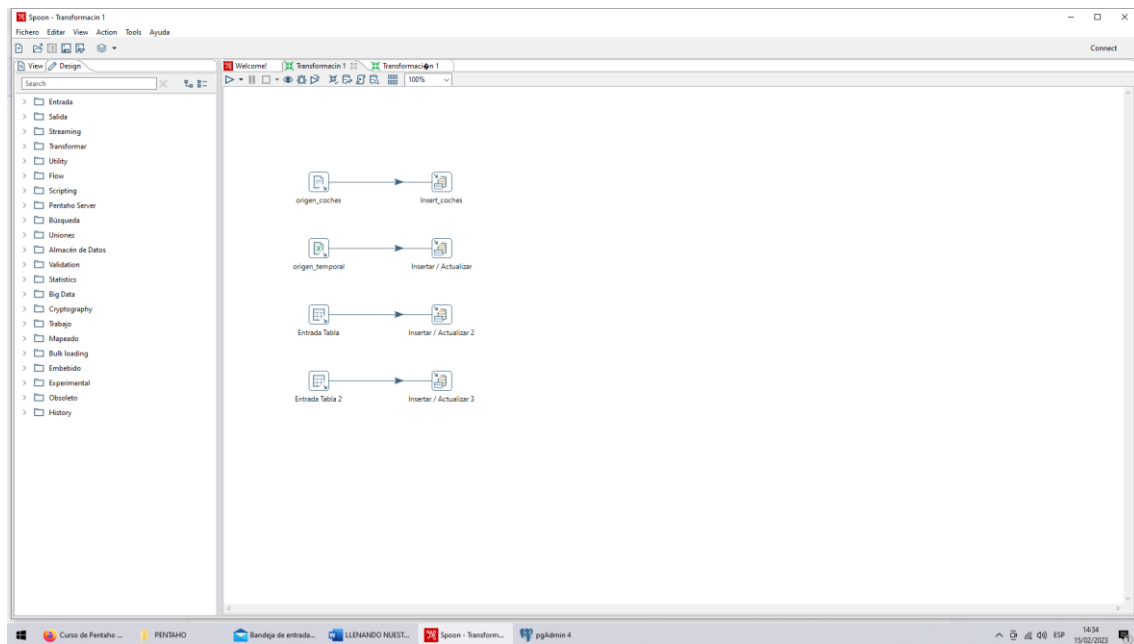


Con la conexión del paso anterior y esta SELECT.

De nuevo insertar actualizar



Con esto ya tendríamos cargadas todas nuestras dimensiones  
Creamos una nueva transformación



Si revisamos nuestra base de datos tenemos una tabla ventas  
coches con información sobre las ventas que se realizan

Query

```
1 SELECT * FROM public.ventas_coches
2 ORDER BY 1d.transaccion ASC
```

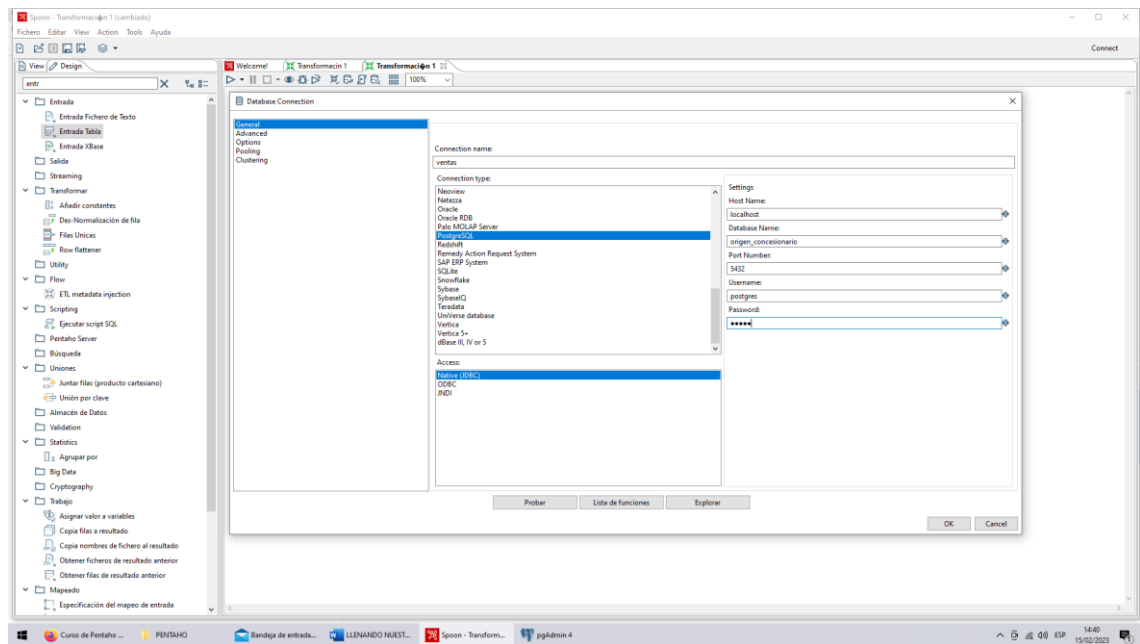
Data Output Messages Notifications

id_transaccion [PK] integer	id_gestor numeric (3)	id_concesionario numeric (3)	marca character varying (30)	modelo character varying (30)	porcentaje_financiado numeric (3)	fecha_venta date
1	1	12	Ford	Focus	56	2019-01-05
2	2	12	Seat	Ibiza	67	2019-01-12
3	3	12	BMW	X1	12	2019-03-03
4	4	12	Seat	Altea	0	2019-08-23
5	5	12	Ford	Mondeo	0	2019-06-30
6	6	12	Seat	León	78	2019-07-01
7	7	12	Seat	Córdoba	67	2019-07-22
8	8	23	Seat	Altea	23	2018-08-08
9	9	23	Ford	Focus	12	2018-03-11
10	10	23	Seat	Córdoba	0	2019-01-14
11	11	23	BMW	Serie 2	43	2019-01-18
12	12	23	BMW	Serie 1	0	2018-08-30
13	13	23	Seat	Ibiza	78	2019-07-23

Total rows: 37 of 37 Query complete 00:00:01.196 Ln 1, Col 1

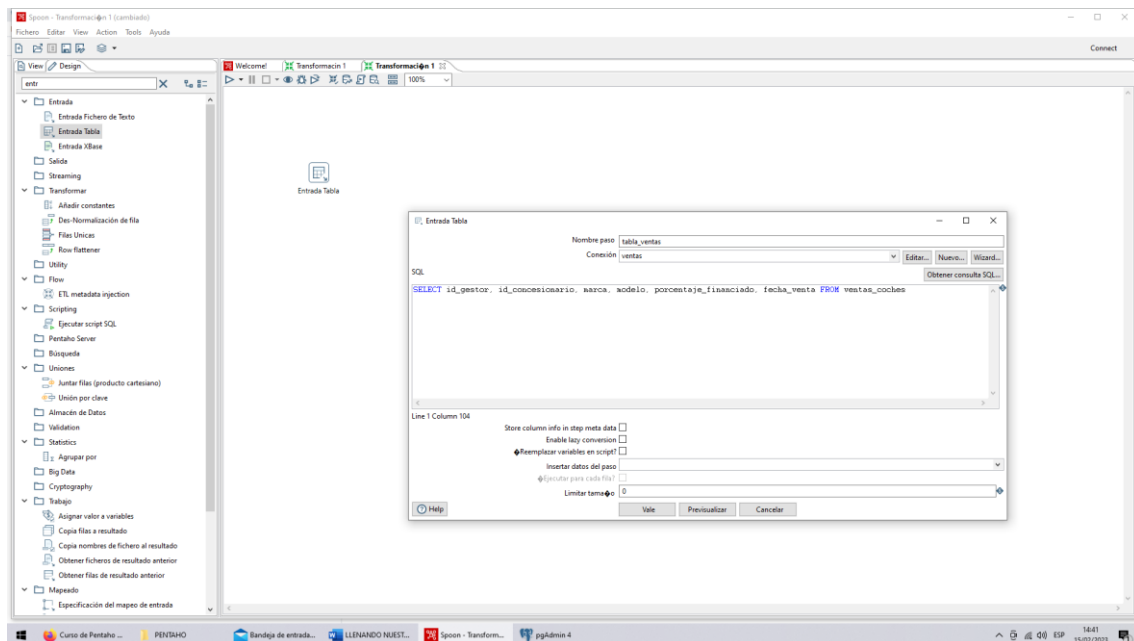
Vamos a cargarla como primera entrada

Primero tendremos que crear una nueva conexión



Y luego seleccionamos los datos con esta select

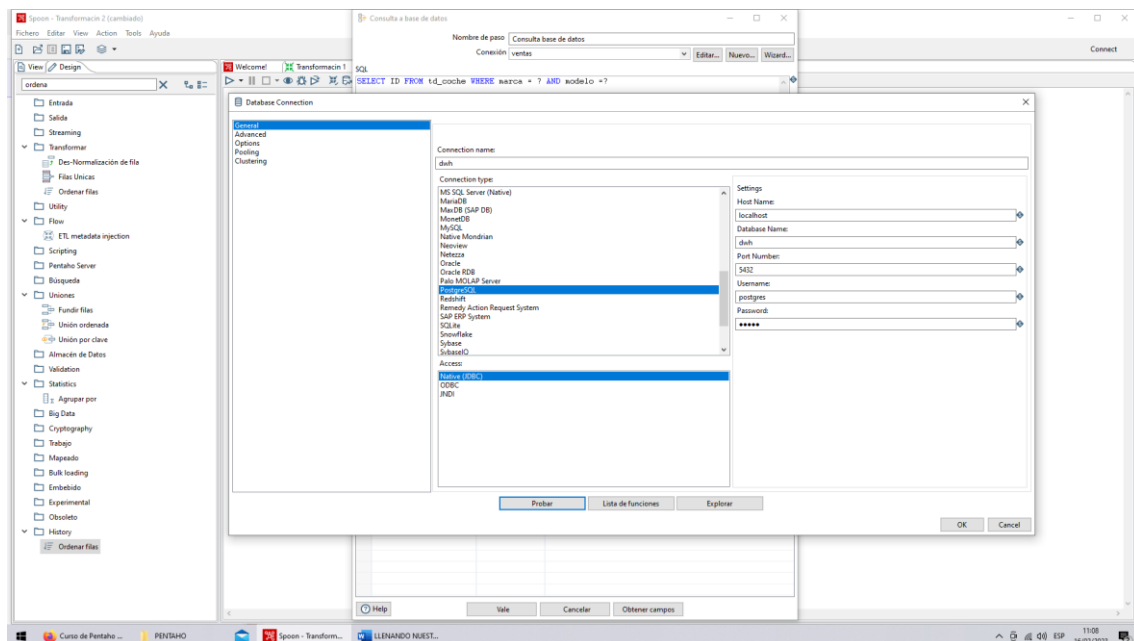




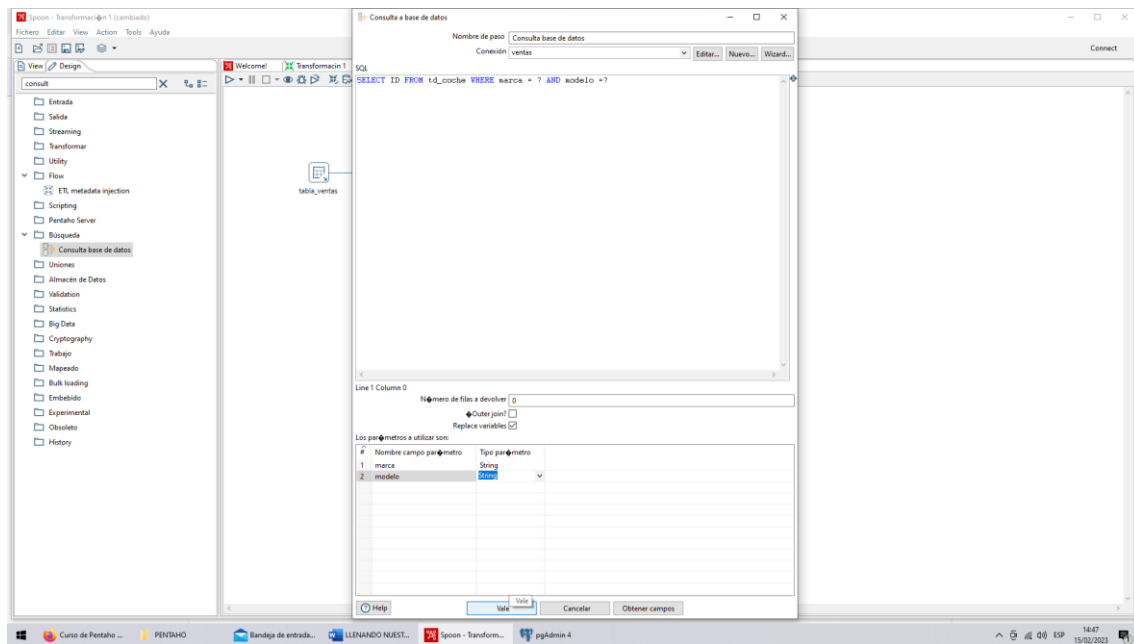
Ahora vamos a hacer una consulta a base de datos para extraer el id de cada dimension para formar nuestra tabla de hechos.

Por ejemplo el id de coche no existe solo tenemos marca y modelo, por lo que hay que consultarlo a la tabla dimensional

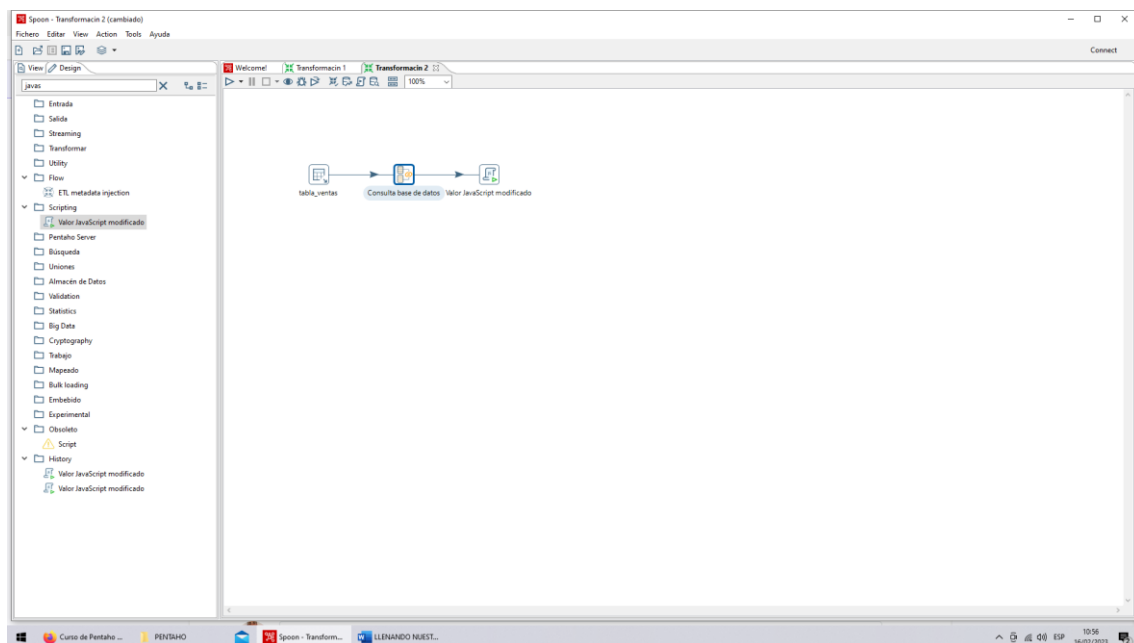
Necesitaremos una nueva conexión



Ponemos lo siguiente y le indicamos abajo los parametros por los que tiene que remplazar y marcando el tick!!

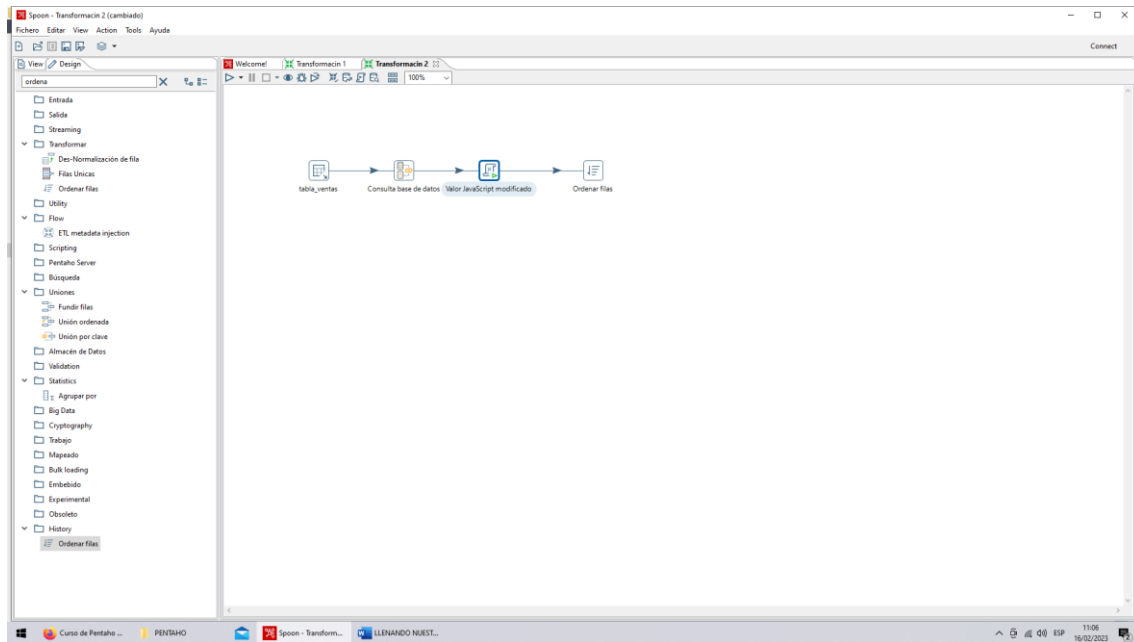


La fecha no esta en el formato que nos interesa, vamos a agregar un paso para modificarlo. Agregamos el paso valor de Javascript modificado.

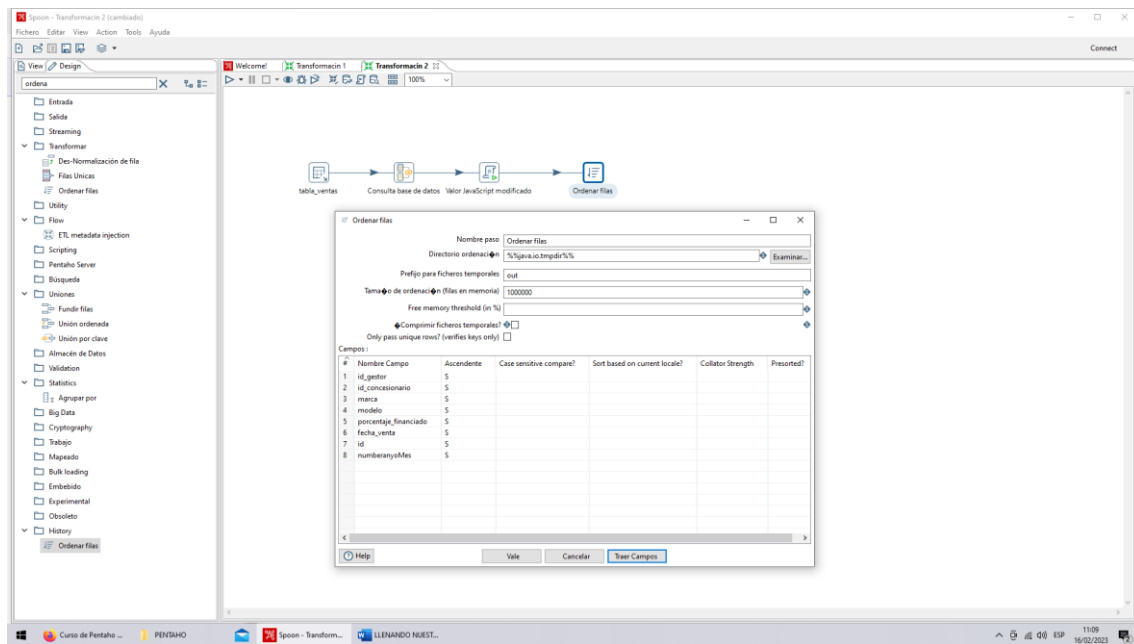


Podremos acceder a las variables del flujo como variables en el script

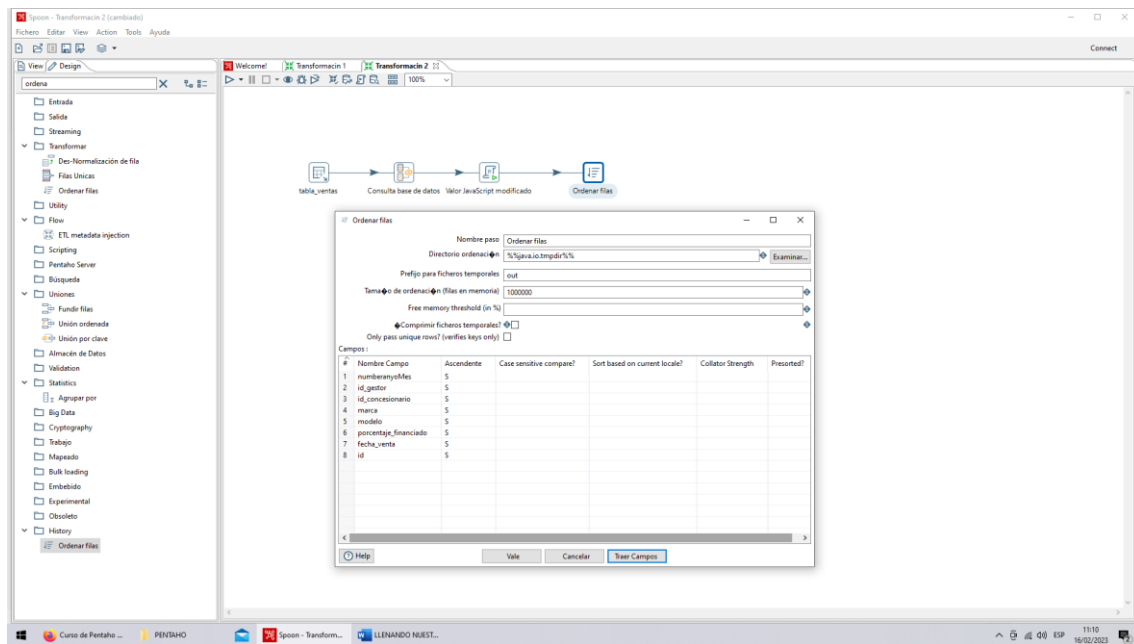




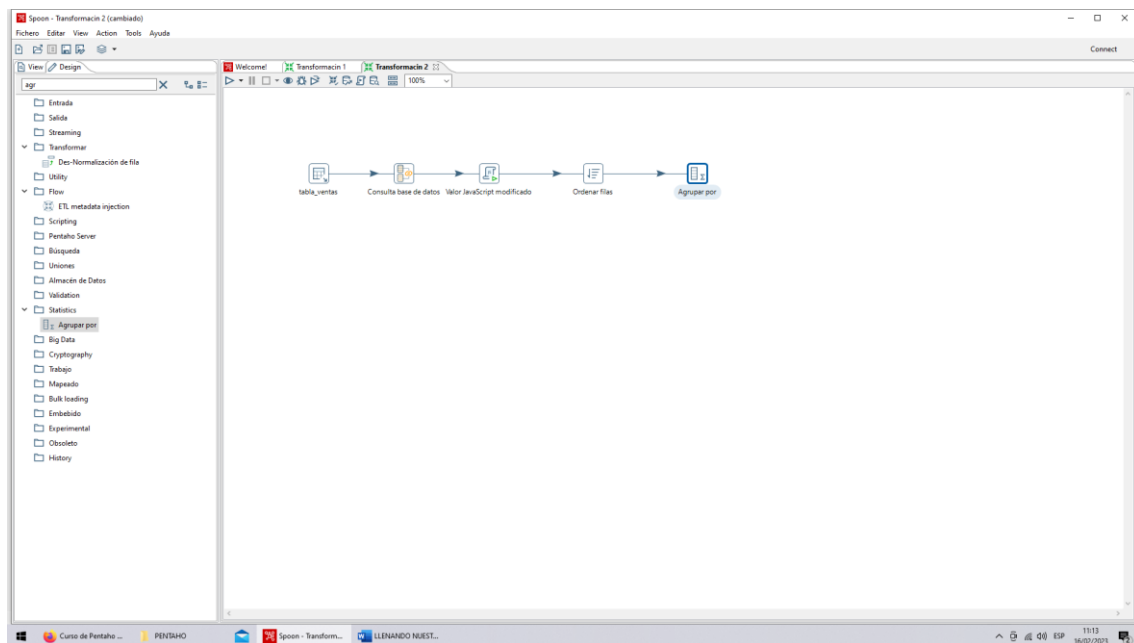
Traemos los campos



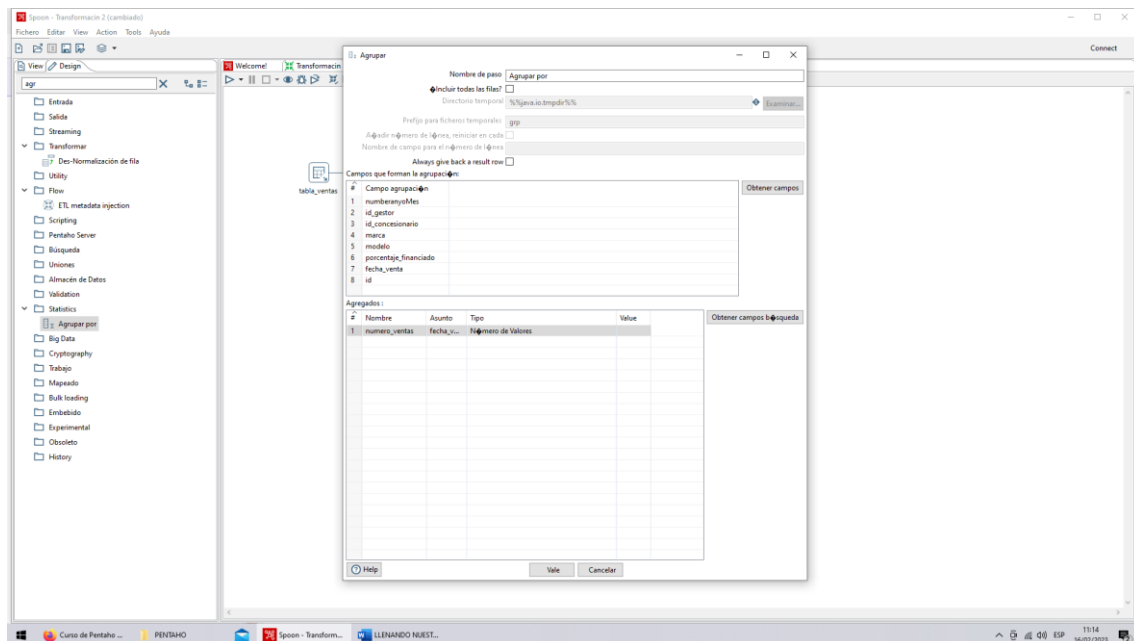
Y le indicamos que ordene por numeroanyomes y luego el resto



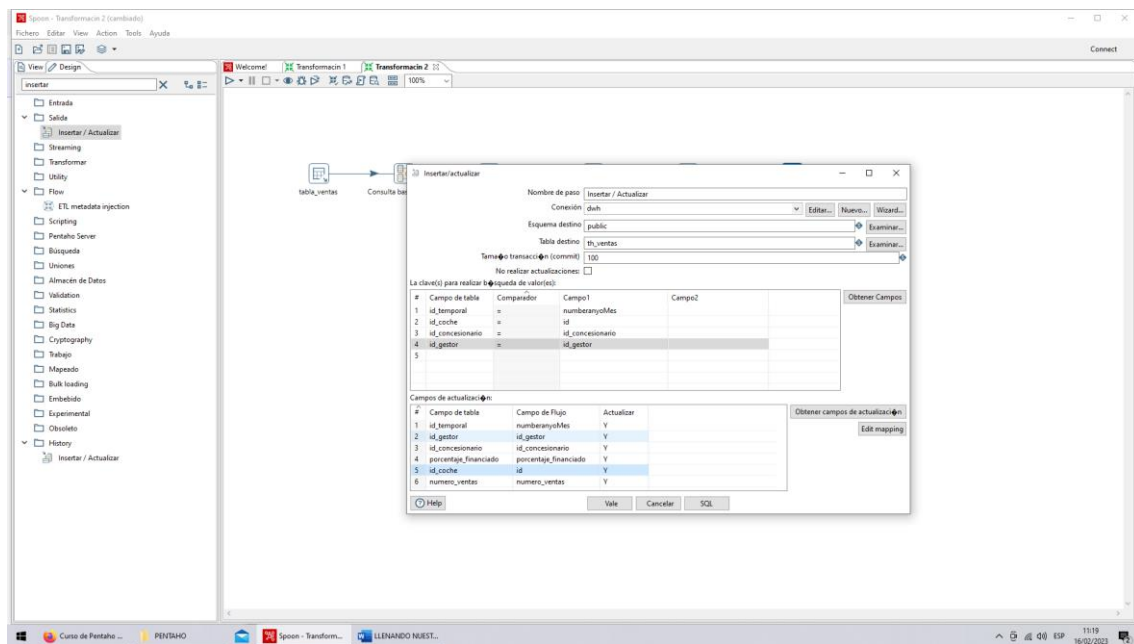
Ahora agregamos un paso de agrupacion para sacar el numero de ventas totales por grupo



Ponemos en los campos que forman la agrupacion todos con el numberanyomes primero y creamos el siguiente valor agregado



El ultimo paso es insertar en el dwh todos estos campos generados en el flujo de datos. Agregamos un insertar actualizar con las siguientes condiciones



Vamos a ejecutar las transformaciones y ver que funcionan correctamente!!