

SMD. Práctica 2.1. Herramientas de usuario final
Power Query en *Excel*

José Samos Jiménez

Departamento de Lenguajes y Sistemas Informáticos
Universidad de Granada

2020 jsamos (LSI-UGR)

Curso 2019-2020

Índice

1. Introducción	3
2. Obtención de los datos	3
3. Transformar los datos en formato de tabla simple	6
3.1. Inicio de <i>Power Query</i>	6
3.2. Importación de datos	6
3.3. Transformar en tabla simple	8
3.4. Codificación del número de habitantes	11
3.5. Salvar los datos	11
4. Obtener las tablas de las dimensiones	12
4.1. Transformaciones generales para cualquier dimensión	12
4.2. Duplicar y adaptar la definición de dimensiones	13
5. Obtener la tabla de hechos	14
6. Enriquecer las dimensiones y los hechos	17
6.1. Enriquecer la dimensión <i>Cuándo</i>	17
6.2. Enriquecer la dimensión <i>Dónde</i>	18
6.3. Enriquecer y simplificar los hechos	21
6.4. Salvar los datos	21

Los objetivos de esta actividad son:

- Aprender a importar a *Power Query* datos desde otras fuentes de datos.
- Entender el funcionamiento de *Power Query*.
- Saber cómo transformar los datos.
- Saber cómo generar nuevas tablas.
- Aprender a obtener un diseño en estrella.

Esta actividad se realizará sobre *Power Query*, herramienta integrada en *Excel*. En principio, se puede realizar sobre cualquiera de sus versiones, siempre que sea igual o superior a 2010. Las capturas de pantalla que se presentan se han realizado sobre la versión 2010.

A continuación, después de una introducción sobre la herramienta, se irán presentando algunas de sus funcionalidades mediante pasos que deberás ir realizando.

1. Introducción

Power Query es una de las herramientas de *Microsoft* orientadas al autoservicio de Sistemas OLAP. Surgió en 2011 como parte de *Microsoft SQL Azure Labs*, con el nombre de *Data Explorer*; con ese mismo nombre se incorporó en 2013 como un plug-in de *Excel 2010* y *2013* y, después, en ese mismo año, se le dio el nombre de *Power Query*. En 2015 se incluyó en *Excel 2016* dentro del apartado «Obtener y transformar».

2. Obtención de los datos

El guión de esta práctica está desarrollado utilizando los datos de la provincia de Granada. Debes realizar los pasos numerados que se indican sobre tus propios datos: **sobre los datos de la provincia que te corresponde**¹.

Los datos de las distintas provincias se puede obtener del INE (Instituto Nacional de Estadística), a partir de <https://www.ine.es/dynt3/inebase/index.htm?padre=525>, cuya página se muestra en la figura 1.

Pulsando sobre el icono «Descarga de ficheros» (figura 1) situado a la izquierda del nombre de la provincia que tenemos asignada, se abre la ventana de la figura 2, que nos permite seleccionar el formato de descarga de los datos. Podríamos seleccionar cualquiera de ellos pero en este guión se ha considerado el formato «Excel: extensión XLSx».

Vamos a trabajar con datos del Padrón municipal para los municipios de la provincia asignada. Para cada municipio, obtenemos los datos del número de mujeres y hombres contabilizados cada 1 de Enero, desde el año 1996 hasta el último año publicado.

1. Descarga de <https://www.ine.es/dynt3/inebase/index.htm?padre=525> el archivo de la provincia que tienes asignada² en formato *Excel* y sitúalo en una carpeta de trabajo.

Si abrimos el archivo que hemos descargado (figura 3, el nombre depende de la provincia), podemos ver que tenemos una tabla dinámica con los municipios de la provincia correspondiente en la primera columna (una combinación del código y el nombre del municipio) y, en el resto de columnas, tenemos la combinación de los valores del sexo y el año. Adicionalmente, la tabla tiene unas líneas de cabecera y, si accedemos a la última línea de la tabla, podemos ver que, a continuación, también hay varias líneas al pie de la misma (figura 4).

Adicionalmente, vamos a completar esos datos con otros comunes que he preparado a partir de datos obtenidos del INE y del CNIG³ (Centro Nacional de Información Geográfica). Están disponibles en el enunciado de esta actividad.

¹<https://pradogrado1920.ugr.es/mod/assign/view.php?id=166481>

²<https://pradogrado1920.ugr.es/mod/assign/view.php?id=166481>

³<http://centrodedescargas.cnig.es/CentroDescargas/catalogo.do?Serie=CAANE>

The screenshot shows the INEbase website interface. At the top, there is a navigation bar with the INE logo and the text 'Instituto Nacional de Estadística'. Below this, there are several colored boxes representing different sections: '01 EI INE', '03 Métodos y proyectos', '05 Prensa', '02 INEbase', '04 Formación y empleo', '06 Productos y servicios', 'Censo electoral', and 'Sede electrónica'. A search bar with the text 'Buscar' is located on the right. Below the navigation bar, there is a banner with the text 'INEbase / Demogr... / Padrón. ... / Cifras ofi... / Cifras oficiales de población resultantes de la revisión del Padrón municipal a 1 de enero'. Below the banner, there is a section titled 'Detalle municipal' with a link to 'Descargar archivo comprimido con los ficheros excel municipales de cada año a nivel nacional'. This section contains a list of 25 provinces, each with a download icon and a link to 'Población por municipios y sexo'. The provinces listed are: 1 Albacete, 2 Alicante/Alacant, 3 Almería, 4 Araba/Álava, 5 Asturias, 6 Ávila, 7 Badajoz, 8 Balears, Illes, 9 Barcelona, 10 Bizkaia, 11 Burgos, 12 Cáceres, 13 Cádiz, 14 Cantabria, 15 Castellón/Castelló, 16 Ciudad Real, 17 Córdoba, 18 Coruña, A, 19 Cuenca, 20 Gipuzkoa, 21 Girona, 22 Granada, 23 Idiazabal, 24 Huelva, and 25 Huesca. A 'Descarga ficheros' button is located at the bottom left of the list.

Detalle municipal (Descargar archivo comprimido con los ficheros excel municipales de cada año a nivel nacional)

- 1 Albacete: Población por municipios y sexo.
- 2 Alicante/Alacant: Población por municipios y sexo.
- 3 Almería: Población por municipios y sexo.
- 4 Araba/Álava: Población por municipios y sexo.
- 5 Asturias: Población por municipios y sexo.
- 6 Ávila: Población por municipios y sexo.
- 7 Badajoz: Población por municipios y sexo.
- 8 Balears, Illes: Población por municipios y sexo.
- 9 Barcelona: Población por municipios y sexo.
- 10 Bizkaia: Población por municipios y sexo.
- 11 Burgos: Población por municipios y sexo.
- 12 Cáceres: Población por municipios y sexo.
- 13 Cádiz: Población por municipios y sexo.
- 14 Cantabria: Población por municipios y sexo.
- 15 Castellón/Castelló: Población por municipios y sexo.
- 16 Ciudad Real: Población por municipios y sexo.
- 17 Córdoba: Población por municipios y sexo.
- 18 Coruña, A: Población por municipios y sexo.
- 19 Cuenca: Población por municipios y sexo.
- 20 Gipuzkoa: Población por municipios y sexo.
- 21 Girona: Población por municipios y sexo.
- 22 Granada: Población por municipios y sexo.
- 23 Idiazabal: Población por municipios y sexo.
- 24 Huelva: Población por municipios y sexo.
- 25 Huesca: Población por municipios y sexo.

Figura 1: Lista de datos por provincia.

The screenshot shows a dialog box titled 'Descarga ficheros'. It contains a section 'Descargar la tabla en formato:' with the following options: 'Pc-Axis', 'Excel: extensión XLS', 'Excel: extensión XLSx', 'CSV separado por tabuladores', 'CSV separado por ,', 'CSV separado por ;', and 'Json'.

Descarga ficheros

Descargar la tabla en formato:

- Pc-Axis
- Excel: extensión XLS
- Excel: extensión XLSx
- CSV separado por tabuladores
- CSV separado por ,
- CSV separado por ;
- Json

Figura 2: Formatos de descarga disponibles.

2871.xlsx - Microsoft Excel

Cifras oficiales de población resultantes de la revisión del Padrón municipal a 1 de enero

1 Cifras oficiales de población resultantes de la revisión del Padrón municipal a 1 de enero

2 Detalle municipal

3

4 Granada: Población por municipios y sexo.

5 Unidades: Personas

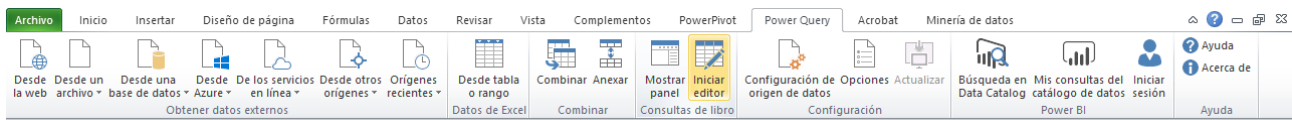
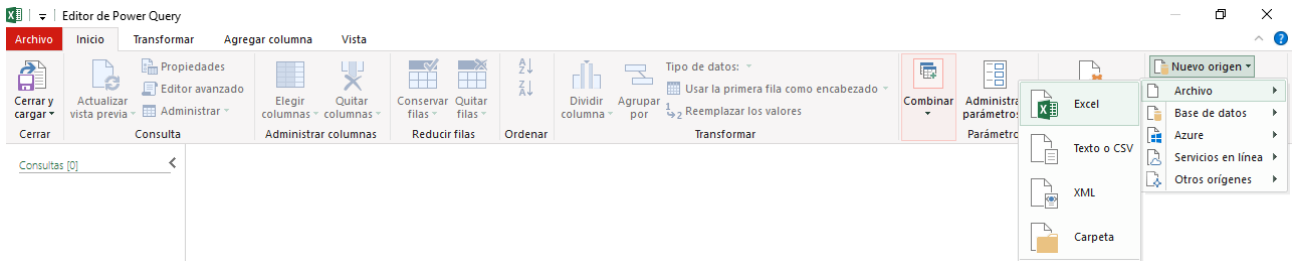
6

	Total	2018	2017	2016	2015	2014	2013	2012	2011
18 Granada		912.075	912.938	915.392	917.297	919.455	919.319	922.928	
18001 Agrón		288	290	297	310	313	324	342	
18002 Alamedilla		599	606	611	637	658	695	673	
18003 Albolote		18.746	18.660	18.497	18.464	18.306	18.082	18.088	
18004 Albondón		729	745	762	782	797	814	889	
18005 Albuñán		415	418	421	420	430	423	430	
18006 Albuñol		7.128	6.998	6.971	6.948	6.853	6.878	6.587	
18007 Albuñuelas		830	836	876	893	889	879	920	
18010 Aldeire		628	635	639	631	644	653	670	
18011 Alfacer		5.378	5.440	5.429	5.445	5.490	5.552	5.513	
18012 Algarinejo		2.591	2.677	2.734	2.813	2.875	2.943	3.029	
18013 Alhama de Granada		5.980	5.981	6.044	6.126	6.191	6.154	6.214	
18014 Alhendín		9.168	8.928	8.740	8.503	8.306	8.020	7.746	
18015 Alicón de Ortega		475	495	511	523	529	506	500	
18016 Almegíjar		346	362	377	376	361	377	413	
18017 Almúñecar		26.377	27.135	27.397	27.391	27.195	25.586	27.703	
18018 Alpuxarra de la Sierra		978	1.021	1.038	1.041	1.065	1.044	1.132	
18018 Alquife		624	665	703	718	701	733	712	
18020 Arenas del Rey		1.178	1.179	1.231	1.865	1.913	1.929	2.038	
18021 Armilla		23.968	24.028	23.901	23.623	23.175	22.904	22.507	
18022 Atarfe		18.554	18.392	18.092	17.792	17.570	17.141	16.843	
18023 Baza		20.519	20.642	20.656	20.668	20.644	21.258	21.407	
18024 Beas de Granada		985	1.000	1.004	1.012	1.005	1.016	1.031	
18025 Beas de Guadix		340	345	354	363	364	366	380	
18027 Benaolúa		3.285	3.286	3.277	3.298	3.298	3.301	3.351	
18028 Benaolúa de las Villas		1.065	1.083	1.107	1.148	1.163	1.252	1.375	
18029 Benamaurel		2.297	2.337	2.359	2.378	2.414	2.445	2.447	
18030 Bérchules		719	715	727	739	762	759	835	
18032 Bubión		302	301	294	309	315	337	335	
18033 Busquistan		279	278	276	289	296	265	292	

Figura 3: Datos de la provincia.

181	18194 Zújar	2.597	2.612	2.678	2.777	2.825	2.930	2.989
182								
183								
184	Notas:							
185	Si en la serie de datos en algún año no existe población es porque a la fecha de referencia no existía el municipio.							
186	No existen cifras a 1 de enero de 1997 porque no se realizó revisión del padrón para ese año							
187	Las cifras de 1996 están referidas a 1 de mayo y las demás a 1 de enero.							
188								
189	Fuente:							
190								

Figura 4: Datos de la provincia (pie).

Figura 5: Inicio de *Power Query*.Figura 6: Importar un archivo *Excel*.

2. Descarga el archivo `datosComunes.zip`, asociado a la actividad, y sitúa los archivos que contiene en la misma carpeta de trabajo.

Estos archivos contienen códigos para la altitud, superficie y tamaño de los municipios (en el archivo `códigos.xlsx`, las hojas `codAltitud`, `codSuperficie`, `codTamMunicipio`), la asociación de cada municipio con la comarca a la que pertenece y la asociación de cada provincia con su comunidad autónoma (en el archivo `orgTerritorio.xlsx`, las hojas `municipioComarca` y `provinciaCA`, respectivamente).

3. Transformar los datos en formato de tabla simple

Los datos del Padrón para cada municipio tienen formato de tabla dinámica, vamos a transformarlos en una tabla simple, con datos sin agregaciones.

3.1. Inicio de *Power Query*

Podríamos trabajar directamente con el archivo original de nuestros datos, pero es recomendable trabajar desde un archivo distinto y no modificar los archivos con los datos originales. Vamos a leer el contenido de ese archivo desde un nuevo archivo de trabajo.

Para iniciar *Power Query* desde un archivo *Excel* tenemos que pulsar sobre [«Power Query», «Consultas de libro», «Iniciar editor»] (figura 5). Se abre en una ventana una nueva aplicación, el *Editor de Power Query*.

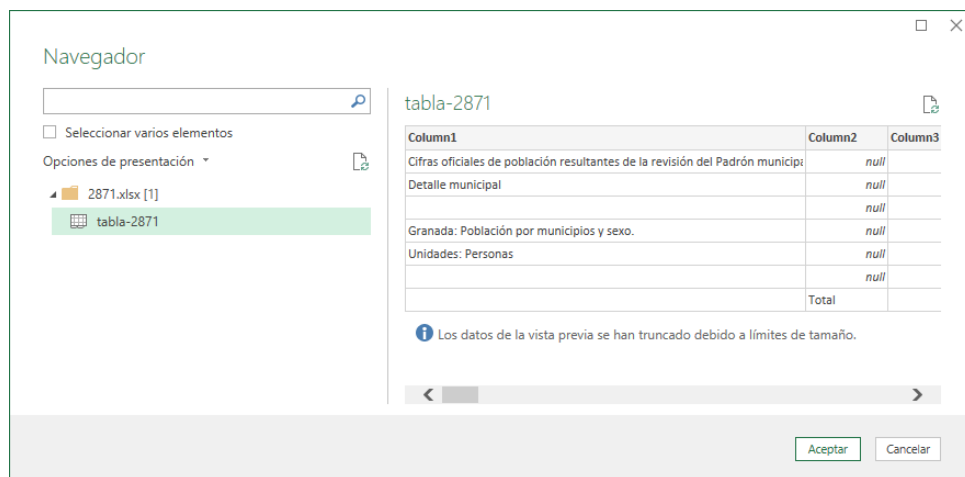
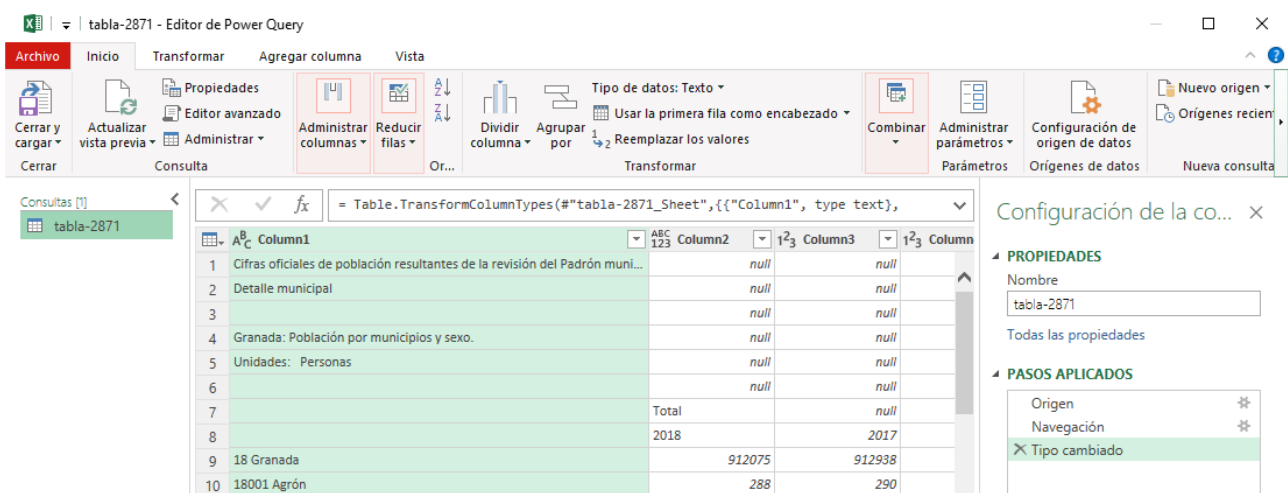
3. En otra carpeta de trabajo, crea un archivo *Excel* cuyo nombre sea el nombre de la provincia que tienes asignada, el literal «-ETL-» y tu nombre de usuario de correo UGR (en mi caso se llamará `granada-ETL-jsamos.xlsx`) e inicia el *Editor de Power Query*.

El resto de operaciones las realizaremos sobre el *Editor de Power Query* en ese archivo.

3.2. Importación de datos

Power Query permite la importación de datos en muchos formatos. En este caso vamos a importar datos en formato *Excel*. Para ello pulsamos sobre [«Inicio», «Nueva consulta», «Nuevo origen», «Archivo», «Excel»] (figura 6). A continuación, buscamos y seleccionamos en el árbol de carpetas el archivo que queremos importar.

Al seleccionar un archivo, nos muestra los elementos que tiene definidos (figura 7). En el caso de *Excel*, además de las hojas, nos puede mostrar otros elementos adicionales definidos a partir de las hojas como pueden ser tablas. Podemos indicar que queremos importar varias hojas seleccionando

Figura 7: Elementos a importar de un archivo *Excel*.Figura 8: Elementos a importar de un archivo *Excel*.

la opción «Seleccionar varios elementos». Al pulsar sobre cada elemento, se muestra su contenido. Pulsando sobre el botón «Aceptar» se importan los elementos seleccionados.

4. Importa a *Power Query* la hoja del archivo con los datos de los municipios de la provincia que has descargado en el punto 1.

En la figura 8, se puede apreciar que, en la columna a la izquierda, aparece como nombre de una consulta el mismo de la hoja que hemos importado; a la derecha, aparecen los pasos que se han llevado a cabo para acceder y transformar los datos para la consulta; y, en el centro, se pueden ver los datos obtenidos para dicha consulta.

Los pasos **Origen** y **Navegación** determinan el archivo de trabajo y la hoja dentro del mismo, respectivamente; estos pasos son imprescindibles. El paso **Tipo cambiado**, trata de interpretar los tipos de datos de las columnas para asignarlos a los datos importados y transformarlos según el tipo asignado: generalmente es un motivo de problemas; lo añade automáticamente después de varias de las operaciones de transformación que permite; mi recomendación es **eliminarlo siempre que se añade automáticamente**.

Cualquier paso se puede eliminar pulsando sobre la «X» situada a la izquierda de su nombre.

5. Elimina el paso **Tipo cambiado** (elimínalo siempre que se añade automáticamente).

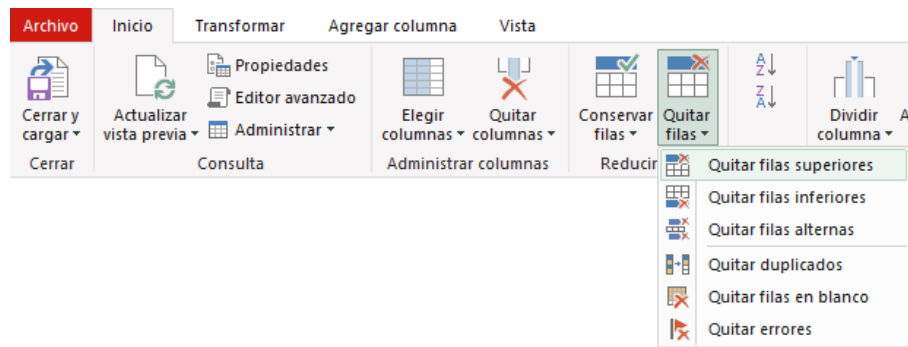


Figura 9: Quitar filas en una consulta.

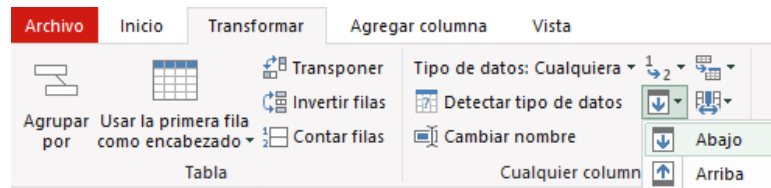


Figura 10: Rellenar hacia abajo.

3.3. Transformar en tabla simple

Los datos disponen de una cabecera (6 líneas) y un pie (9 líneas) situados antes y después de la tabla dinámica, respectivamente (figuras 3 y 4). Podemos eliminar las primeras o últimas filas de una consulta pulsando sobre [«Inicio», «Reducir filas», «Quitar filas», «Quitar filas superiores»] e [«Inicio», «Reducir filas», «Quitar filas», «Quitar filas inferiores»], respectivamente (figura 9).

6. Elimina las filas de cabecera y pie de la tabla dinámica.

Los datos de las dimensiones en la tabla dinámica están en la primera columna y en las dos primeras filas, la combinación de estos últimos, forma la cabecera del resto de columnas. Los datos de la primera columna (datos de municipio) están completos, sin embargo, en los datos de las dos primeras filas se ha adoptado el criterio de no repetir los valores en las combinaciones. Para transformar la tabla dinámica en una tabla simple, deberíamos completar los valores faltantes.

Power Query está orientado principalmente al trabajo con columnas (salvo las operaciones que hemos aplicado de eliminar filas). Si queremos transformar los datos de las filas, tenemos que transformarlas en columnas. La operación que permite transformar las filas en columnas y viceversa es [«Transformar», «Tabla», «Transponer»].

7. Transpón la tabla dinámica.

Una vez transpuesta, podemos completar los valores que faltan en la primera columna. Cada valor aparece una sola vez y el resto de celdas hasta el nuevo valor que aparece deberían estar rellenas con el valor inicial. Esta operación se realiza automáticamente seleccionando las columnas que queremos completar y pulsando sobre [«Transformar», «Cualquier columna», «Rellenar», «Abajo»] (figura 10).

8. Rellena los datos que falta de la primera columna de la tabla.

La primera fila de la tabla contiene las cabeceras de las columnas para la matriz que forma parte de la tabla dinámica. Para transformar la tabla dinámica en una tabla simple, necesitamos que las cabeceras sean los propios nombres de las columnas. Esta operación se puede llevar a cabo pulsando sobre [«Transformar», «Tabla», «Usar la primera fila como encabezado»].

9. Usa la primera fila como encabezado de la tabla (no olvides eliminar el paso **Tipo cambiado** que añade automáticamente).

Consultas [1] < tabla-2871

ABC 123 ABC 123_1 ABC 123 18 Granada ABC 123 18001 Agrón ABC 123 18002 Alamedilla ABC 123 18003 Albolote ABC 123 18004 Albondón

	ABC 123	ABC 123_1	ABC 123 18 Granada	ABC 123 18001 Agrón	ABC 123 18002 Alamedilla	ABC 123 18003 Albolote	ABC 123 18004 Albondón
1	Total	2018		912075	288	599	18746
2	Total	2017		912938	290	606	18660
3	Total	2016		915392	297	611	18497
4	Total	2015		917297	310	637	18464
5	Total	2014		919455	313	658	18306
6	Total	2013		919319	324	695	18082
7	Total	2012		922928	342	673	18088
8	Total	2011		924550	330	704	17892
9	Total	2010		918072	340	699	17637
10	Total	2009		907428	358	721	17089
11	Total	2008		901220	381	757	16879
12	Total	2007		884099	374	777	16450
13	Total	2006		876184	283	785	15978

Configuración de la columna... x

PROPIEDADES

Nombre
tabla-2871

Todas las propiedades

PASOS APLICADOS

- Origen
- Navegación
- Filas superiores quitadas
- Filas inferiores quitadas
- Tabla transpuesta
- Rellenar hacia abajo
- Encabezados promovidos

Figura 11: Resultado parcial de la transformación.

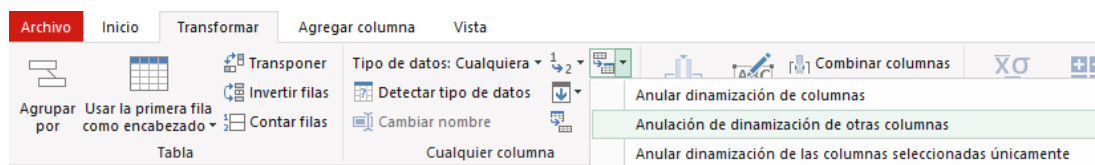


Figura 12: Anular dinamicización de columnas.

El resultado parcial obtenido para los datos de Granada se puede ver en la figura 11. Las dos primeras columnas contienen los datos de los campos **Sexo** y **Año**, respectivamente, que definen el contenido de las filas; la cabecera del resto de columnas contiene los datos de los municipios (el código del municipio y su nombre).

Podemos eliminar la dinamicización de una tabla dinámica seleccionando las columnas que contienen los datos que definen el contenido de las filas (en este caso, seleccionando las dos primeras columnas) y pulsando sobre [«Transformar», «Cualquier columna», «Anular dinamicización de columnas», «Anular dinamicización de otras columnas»] (figura 12) para eliminar la dinamicización de las columnas que contienen las mediciones del Padrón: el número de habitantes para cada municipio, sexo y año.

10. Elimina la dinamicización de las columnas de la tabla dinámica que contienen las mediciones del Padrón.

Ya tenemos una tabla simple como resultado de las transformaciones que hemos aplicado. Sin

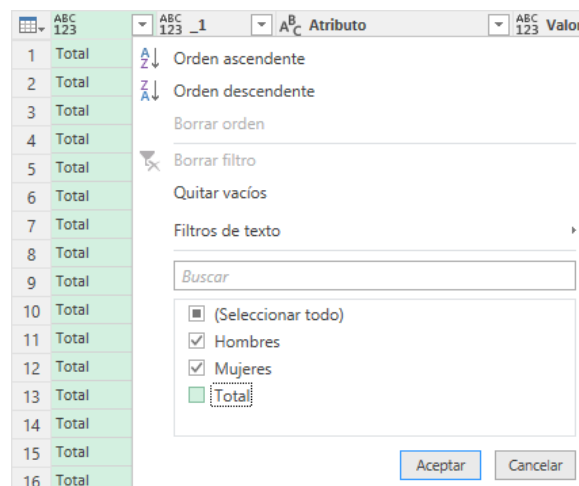


Figura 13: Eliminar registros con determinados valores de un campo.

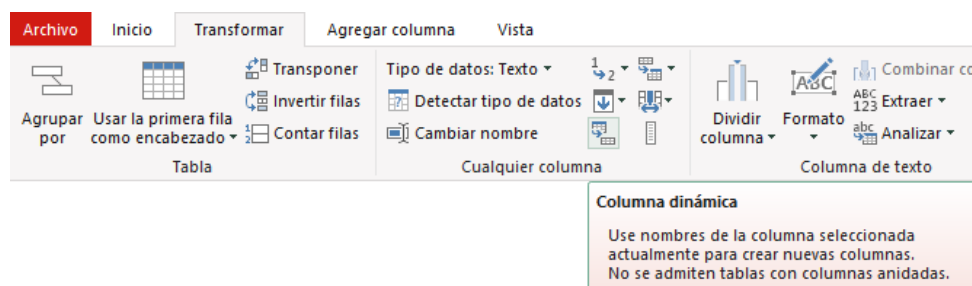


Figura 14: Definir una columna dinámica.

embargo, podemos observar que, además de los datos elementales, tenemos filas con datos agregados, en concreto: el valor **Total** para el sexo y el valor con el código y nombre de la provincia (**18 Granada** en mi caso). Podemos eliminar esos valores deseleccionándolos para cada campo en su cabecera (figura 13).

Por partir de una tabla dinámica, en el caso de las mediciones, podemos tener datos nulos: celdas en las que no hay datos registrados por algún motivo (por ejemplo, se indica en la hoja original que en el año 1997 no se recogieron los datos para ningún municipio). En alguna situación, es posible que nos interese modificar esos datos para que tengan el valor 0. En este caso concreto, si no tenemos datos registrados, nos interesa eliminarlos (es suficiente con deseleccionar el año 1997).

11. Elimina las filas que contengan datos agregados o valores **null** (vacío) en la medición (**se indican en los dos párrafos anteriores**).

El campo que contiene los datos de los municipios, contiene el código y el nombre del municipio. Podemos separar los valores en dos campos seleccionando la columna correspondiente y pulsando sobre [«Transformar», «Columna de texto», «Dividir columna», «Por delimitador»]. En este caso el delimitador es «Espacio» y debemos elegir la opción «Delimitador situado más a la izquierda» para que considere el código por un lado y el resto de valores por otro.

12. Divide en dos campos distintos el código y el nombre del municipio (no olvides eliminar el paso **Tipo cambiado**).

Una vez tenemos todos los campos definidos, podemos asignarles el nombre y tipo adecuado a su contenido. El nombre de una columna se puede modificar pulsando «Doble-clic» sobre la cabecera de la columna e introduciendo el literal correspondiente. El tipo se puede cambiar pulsando sobre el icono situado a la izquierda del nombre de cada columna y seleccionando la opción adecuada: insertará un paso **Tipo cambiado** que ahora no debemos eliminar.

13. Llama a los campos **Sexo**, **Año**, **CodMunicipio**, **Municipio** y **Habitantes**, y asígnale el tipo «Texto» a todos⁴ excepto a **Habitantes** que es de tipo «Número entero».

El campo **Habitantes** puede ser una medición en la que se distingue por sexo. Otra posibilidad es considerar tres mediciones, una por cada valor del campo **Sexo** y el total.

Se puede definir directamente una columna para cada valor del campo **Sexo**, seleccionando la columna y pulsando sobre [«Transformar», «Cualquier columna», «Columna dinámica»] (figura 14). Nos pide que seleccionemos la «Columna de valores», en este caso es el campo **Habitantes**. El resultado es que crea dos campos, **Hombres** y **Mujeres**, podemos definir el campo **Habitantes** como suma de ambos, seleccionándolos y pulsando sobre [«Agregar columna», «De número», «Estándar», «Agregar»].

14. Define una columna **Hombres** y otra **Mujeres** a partir de las columnas **Sexo** y **Habitantes**. A partir de los campos **Hombres** y **Mujeres** define un nuevo campo **Habitantes** como la suma de ambos.

⁴Interesa que solo sean de tipo numérico las mediciones.

Agregar una columna condicional

Agregue una columna condicional que se calcula a partir de las otras columnas o valores.

Nuevo nombre de columna
Personalizado

	Nombre de columna	Operador	Valor	Salida
Si	Habitantes	es menor o igual...	ABC 123 100	Enton... ABC 123 c01
O si	Habitantes	es menor o igual...	ABC 123 200	Enton... ABC 123 c02
O si	Habitantes	es menor o igual...	ABC 123 500	Enton... ABC 123 c03
O si	Habitantes	es menor o igual...	ABC 123 1000	Enton... ABC 123 c04
O si	Habitantes	es menor o igual...	ABC 123 2000	Enton... ABC 123 c05
O si	Habitantes	es menor o igual...	ABC 123 3000	Enton... ABC 123 c06

Agregar regla

De lo contrario
ABC 123 c16

Aceptar Cancelar

Figura 15: Crear un campo condicional.

3.4. Codificación del número de habitantes

Nos interesa estudiar los municipios según su nivel en el número de habitantes de acuerdo con los códigos dados en el archivo `codTamMunicipio.xlsx`. Como este nivel ha podido ir cambiando a lo largo del tiempo, vamos a asociar a cada municipio en cada momento el código que le corresponde.

Tenemos que crear un nuevo campo que, en función del valor del campo **Habitantes**, tome como valor el código que le corresponda de entre los definidos. Esta operación la podemos llevar a cabo creando un campo condicional, pulsando sobre [«Agregar columna», «General», «Columna condicional»] y añadiendo las condiciones necesarias sobre el campo **Habitantes** (figura 15). Si queremos asignar códigos que empiecen por 0, es recomendable añadir algún separador delante que podamos después eliminar fácilmente (en la figura 15 se ha añadido la letra c). Si no se hace así, añade un campo numérico y elimina los ceros al inicio.

Para obtener un campo que contenga solo el código numérico de dos dígitos, una posibilidad es, teniendo seleccionada la nueva columna añadida, pulsar sobre [«Agregar columna», «De texto», «Extraer», «Últimos caracteres»] e indicar los dos últimos caracteres. El resultado es que añade una nueva columna con los datos deseados.

Podemos eliminar la columna inicialmente añadida (**Personalizado**) pulsando sobre ella y seleccionando la opción «Quitar» en el menú de contexto. La columna con el código la podemos renombrar como **CodTamMunicipio** y, «Pulsando-arrastrando-soltando», la podemos situar a continuación de la columna **Municipio**.

15. Añade a continuación del campo **Municipio** un nuevo campo de tipo texto llamado **CodTamMunicipio** que contenga un código de dos dígitos definido a partir del campo **Habitantes**, de acuerdo a los códigos del archivo `codTamMunicipio.xlsx`.

3.5. Salvar los datos

Previamente vamos a cambiar el nombre a la consulta. Se le puede cambiar pulsando «Doble-clic» sobre el nombre de la consulta (en la columna de la izquierda) o bien directamente en el campo «Nombre» del apartado «PROPIEDADES» (en la columna de la derecha).

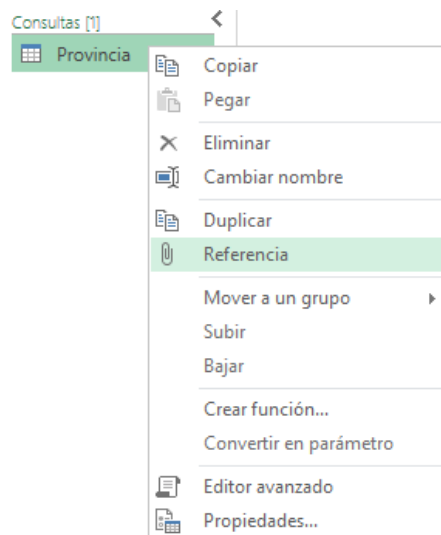


Figura 16: Crear una consulta a partir de otra mediante referencia.

Pulsando sobre [«Inicio», «Cerrar y cargar»], podemos ver el resultado de la transformación en una hoja de *Excel*.

16. Llama a la consulta **Provincia** y pasa los datos a *Excel*, asigna a la hoja el mismo nombre que a la consulta y guarda los cambios en el archivo *Excel*.

En *Excel*, pulsando «Doble-clic» sobre el nombre de la consulta en la columna de la derecha, accedemos directamente a la ventana de definición de consultas de *Power Query*.

4. Obtener las tablas de las dimensiones

Una vez tenemos los datos en formato de tabla simple, vamos a estructurarlos en forma de hechos y dimensiones. Para realizar el diseño, podemos considerar los datos agrupados de la forma siguiente para las dimensiones:

- *Cuándo*: Año.
- *Dónde*: CodMunicipio, Municipio, CodTamMunicipio.

La dimensión *Cuándo* es muy simple ya que tiene un solo campo; la dimensión *Dónde* tiene tres campos: en este caso no tenemos versiones en los nombres de los municipios (por la estructura inicial de los datos) pero sí tenemos posible evolución en el tamaño del municipio relativo al número de habitantes. En cada caso, deberemos obtener las combinaciones de valores únicos para los campos de cada dimensión.

4.1. Transformaciones generales para cualquier dimensión

Partimos del resultado obtenido en la consulta **Provincia**. Podemos crear una nueva consulta que utilice directamente el resultado final de otra consulta creando una referencia a partir de ella. Cualquier modificación en la consulta original se transmitirá automáticamente a las consultas definidas mediante referencia a partir de ella.

Para crear una referencia a partir de la consulta **Provincia**, en su menú contextual, elegimos la opción «Referencia» (figura 16).

Renombramos la consulta con el nombre de la dimensión (*Cuándo* o *Dónde*). A continuación, debemos elegir los campos que componen la dimensión con la que estamos trabajando; esta operación se puede realizar pulsando sobre [«Inicio», «Administrar columnas», «Elegir columnas»] (figura 17). En

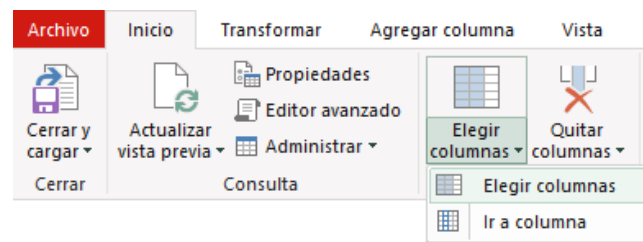


Figura 17: Elegir columnas en una consulta.

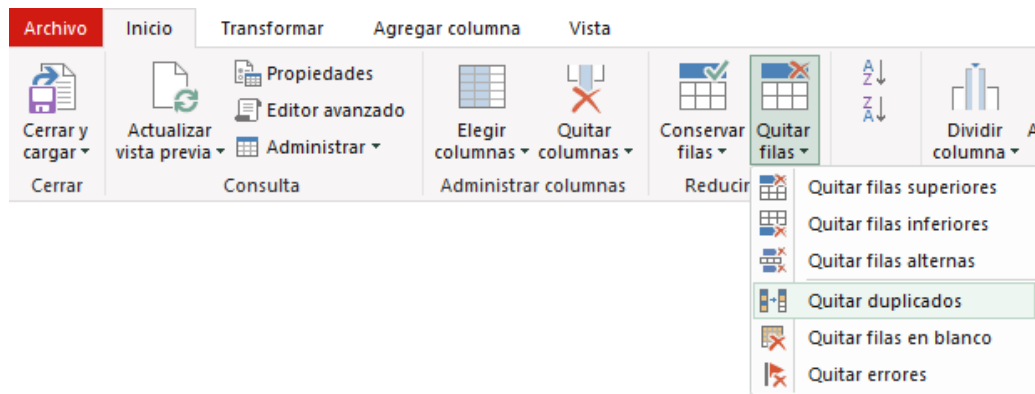


Figura 18: Quitar duplicados en las filas.

la ventana que se abre, seleccionamos las columnas que contienen datos de la dimensión que estamos definiendo.

Obtenemos como resultado una tabla que contiene solo los campos de la dimensión pero con valores duplicados en las filas. Para eliminar los duplicados, pulsamos sobre [«Inicio», «Reducir filas», «Quitar filas», «Quitar duplicados»] (figura 18).

Una vez tenemos los valores únicos, añadimos una llave primaria generada pulsando sobre [«Agregar columna», «General», «Columna de índice», «Desde 1»] (figura 19). La columna añadida es de tipo «Número decimal», se puede ver en el icono a la izquierda del nombre **Índice** que asigna automáticamente a la columna. Pulsando sobre ese icono, podemos cambiar el tipo de la columna, en este caso debería ser de tipo «Número entero». Con esto tenemos los elementos de cualquier dimensión definidos en una consulta.

4.2. Duplicar y adaptar la definición de dimensiones

Los pasos descritos hasta ahora son genéricos para cualquier dimensión. Podemos definir una nueva consulta para obtener otra dimensión seleccionando la opción «Duplicar» del menú contextual de la consulta que acabamos de definir (figura 20).

Lo único que tenemos que modificar son los campos seleccionados en cada caso. Esta operación se puede hacer directamente pulsando sobre la rueda situada a la derecha del paso **Otras columnas quitadas** (figura 21) y seleccionando los campos adecuados para cada dimensión.

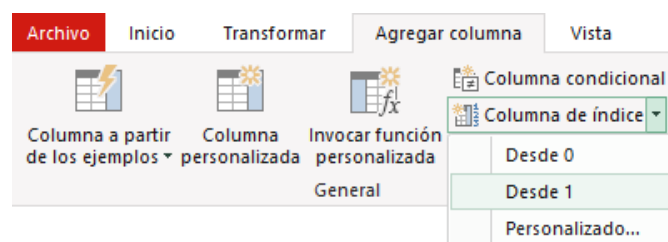


Figura 19: Añadir columna de índice.

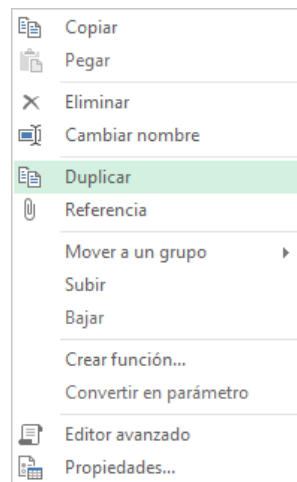


Figura 20: Crear una consulta a partir de otra mediante referencia.

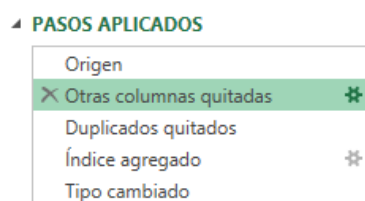


Figura 21: Modificar los pasos aplicados.

Una vez tenemos definidas de esta forma todas las dimensiones, podemos personalizar los nombres de los campos índice y situarlos al principio de la tabla. Para renombrar un campo, basta con pulsar «Doble-clic» sobre su nombre e introducir el nuevo nombre. En este caso, para las dimensiones **Cuándo** y **Dónde** renombramos el campo **Índice** como **IdCuándo** e **IdDónde**, respectivamente. Para mover un campo, «Pulsamos-arrastramos-soltamos» el campo en la posición que deseemos; en el caso de mover al principio de la tabla, desde el menú contextual del campo, podemos elegir la opción [«Mover», «Al principio»].

17. Define las consultas **Cuándo** y **Dónde** asociadas a las dimensiones correspondientes (muestra el resultado para cada consulta).
18. Para la consulta **Dónde** comprueba si hay municipios cuyo nombre se repite y, en caso de darse esta situación, explica las razones que la provocan. Explica qué alternativa tendríamos si no quisiéramos que se repitieran los nombres de municipio, sin perder ninguna información.

5. Obtener la tabla de hechos

Una vez hemos definido las dimensiones, podemos sustituir sus campos en la tabla original por la llave generada de cada dimensión.

Como hemos definido referencias a partir de la tabla **Provincia**, no debemos modificarla para que las referencias definidas a partir de ella no se vean afectadas. Definiremos una nueva consulta también como referencia a partir la tabla **Provincia**, la consulta **Padrón**; esta será la consulta donde llevaremos a cabo las sustituciones.

Para sustituir los valores de los campos de cada dimensión por su llave generada, tenemos que unir las tablas correspondientes. Para ello, seleccionamos la consulta **Padrón** y pulsamos sobre [«Inicio», «Combinar», «Combinar consultas», «Combinar consultas»] (figura 22).

En la ventana que se abre (figura 23), seleccionamos la consulta con la que queremos hacer la operación *Join* (la de una de las dimensiones, en la figura 23, se ha seleccionado la consulta **Dónde**) y, para cada consulta de la operación se seleccionan los campos que intervienen y el orden de estos.

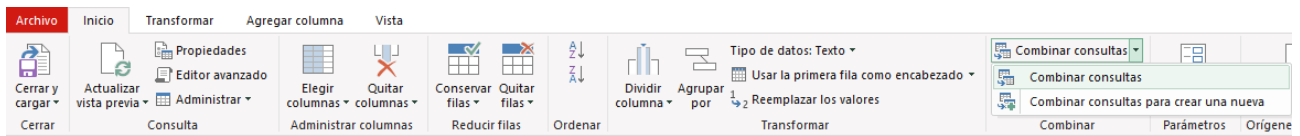


Figura 22: Combinar consultas.

Combinar

Seleccione una tabla y las columnas coincidentes para crear una tabla combinada.

Padrón

Año	CodMunicipio 1	Municipio	CodTamMunicipio 2	Hombres	Mujeres	Habitantes
1996	18001	Agrón	03	228	212	440
1996	18002	Alamedilla	05	508	499	1007
1996	18003	Albolote	09	6259	6196	12455
1996	18004	Albondón	05	588	604	1192
1996	18005	Albuñán	04	256	253	509

Dónde

IdDónde	CodMunicipio 1	Municipio	CodTamMunicipio 2
1	18001	Agrón	02
2	18002	Alamedilla	03
3	18003	Albolote	08
4	18004	Albondón	03
5	18005	Albuñán	03

Tipo de combinación

Externa izquierda (todas de la primera, coincidencias...)

Aceptar **Cancelar**

Figura 23: Ventana de definición de la combinación de consultas.

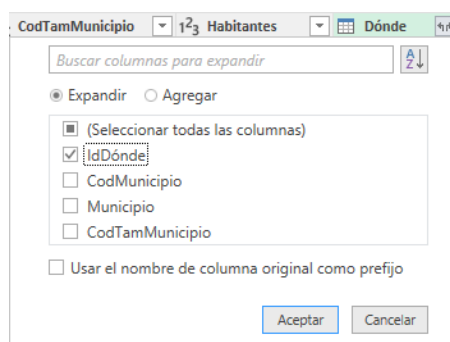


Figura 24: Selección de campos de una consulta.

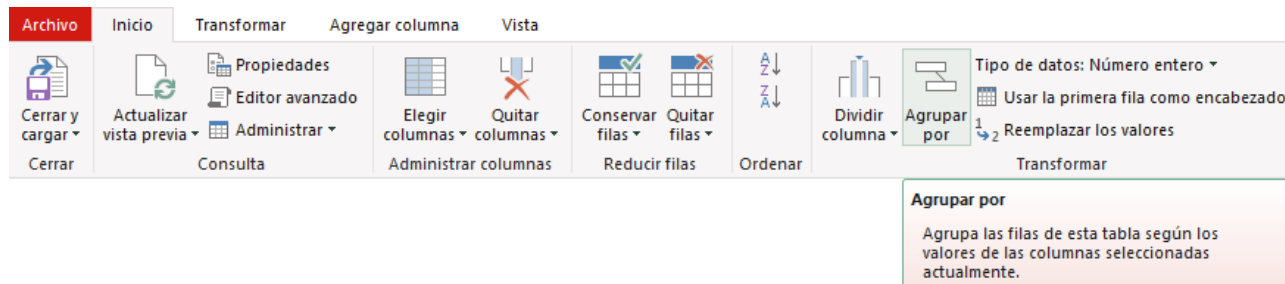


Figura 25: Agrupar los registros de una tabla.

Cada campo se selecciona pulsando sobre él. En el caso de utilizar más de un campo de cada consulta, se seleccionan manteniendo pulsada la tecla «Ctrl» al pulsar sobre el campo: en la parte derecha de la cabecera, se indica mediante un número el orden en el que se consideran los campos en cada consulta. En el campo «Tipo de combinación» se puede definir el criterio de unión entre las tablas, la opción por defecto es adecuada para la operación que pretendemos hacer.

NOTA IMPORTANTE: los campos de las dimensiones que intervengan en la unión para definir la tabla de hechos deberán mantenerse siempre en las dimensiones, en caso contrario, al reevaluar la consulta fallará la definición de los hechos (por no encontrar los campos utilizados en su definición).

El resultado de esta operación es que a la tabla original (Padrón) se le añade una columna con el nombre de la otra consulta donde podemos seleccionar los campos que queremos añadir. Pulsando sobre el icono con dos flechas a la derecha del nombre de la nueva columna (figura 24) podemos seleccionar los campos a incluir en la tabla: en este caso nos interesa incluir solo la llave generada sin necesidad de que le añada ningún prefijo.

Una vez añadida la nueva llave externa a la tabla Padrón, podemos eliminar los campos incluidos en la dimensión asociada, representados por dicha llave externa. Podemos seleccionar los campos de la consulta pulsando sobre [«Inicio», «Administrar columnas», «Elegir columnas»] (figura 17).

Una vez sustituidos los campos de todas las dimensiones, podemos reordenar los campos restantes (llaves externas y mediciones). El orden no es relevante aunque se suelen poner al principio de la tabla los campos que la identifican, en este caso las llaves externas.

La combinación de las llaves externas debería identificar a los hechos. Para asegurarnos de que efectivamente es así, es recomendable agrupar los registros de la tabla utilizando dichas llaves. Podemos realizar esta operación pulsando sobre [«Inicio», «Transformar», «Agrupar por»] (figura 25). En la ventana que se abre, definimos los campos por los que se realiza la agrupación (las llaves externas) y, para cada medición, definimos la función de agregación que le corresponde (figura 26). En este caso esta operación no sería necesaria, pero es bueno conocerla.

19. Define la consulta Padrón y sustituye en ella los valores de las columnas de las dimensiones Cuándo y Dónde por las llaves generadas de cada una.

Figura 26: Ventana de definición de la agrupación.

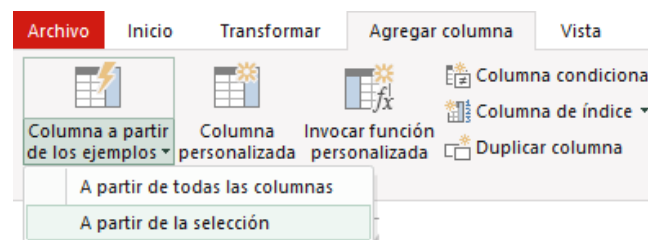


Figura 27: Definir una nueva columna mediante ejemplos.

6. Enriquecer las dimensiones y los hechos

Actualmente tenemos en las dimensiones exclusivamente los datos incluidos en la tabla original. A partir de estos datos básicos podemos derivar o asociar otros datos que pueden resultar interesantes para analizar los hechos.

Adicionalmente, podemos considerar mediciones adicionales que se pueden usar en la definición de mediciones calculadas.

6.1. Enriquecer la dimensión *Cuándo*

En la consulta *Cuándo* tenemos los datos del año al que se refieren las mediciones. Para realizar consultas, además del año, nos puede interesar disponer de la década correspondiente. Para definir la década a partir del año, podemos usar la funcionalidad de definición de una columna mediante ejemplos.

Para ello, seleccionamos las columnas en las que se basan los ejemplos (en este caso solo necesitamos solo la columna *Año*) y pulsamos sobre [«Agregar columna», «Columna a partir de los ejemplos», «A partir de la selección»] (figura 27).

Considerando los ejemplos que facilitamos, la herramienta deduce la función necesaria para obtener el resultado a partir de los datos de entrada. Por ejemplo, si introducimos la década para los años 2018, 2009 y 1999, deduce correctamente la función que la obtiene a partir del año (figura 28).

20. Añade el campo *Década* a la consulta *Cuándo*.

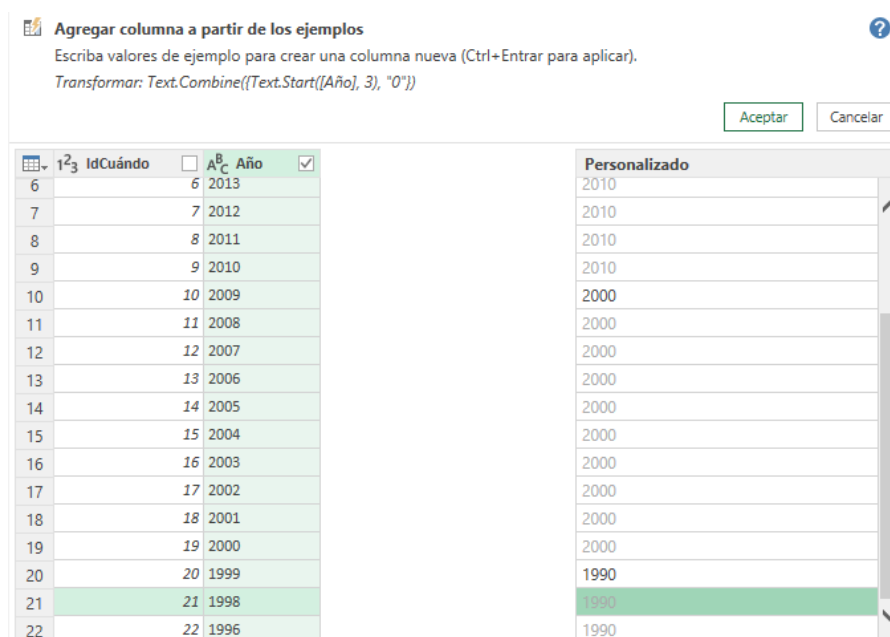


Figura 28: Definición de una nueva columna mediante ejemplos.

6.2. Enriquecer la dimensión *Dónde*

Para la dimensión *Dónde*, a partir del municipio también podemos obtener datos asociados relevantes para el proceso de toma de decisiones. En este caso no se obtienen aplicando una fórmula, como hemos hecho para obtener la década a partir del año (en el apartado 6.1), sino que necesitamos alguna fuente de datos donde se asocie el municipio con el resto de datos. Para simplificar el desarrollo, se facilitan en los archivos `orgTerritorio.xlsx` y `códigos.xlsx` algunos datos que pueden ser relevantes.

21. Importa a *Power Query* las hojas `municipioComarca` y `provinciaCA` del archivo `orgTerritorio.xlsx`, y las hojas `codAltitud`, `codSuperficie` y `codTamMunicipio` del archivo `códigos.xlsx`. No olvides eliminar el paso *Tipo cambiado* en todas las consultas generadas (una por cada hoja importada).

En la consulta `municipioComarca` los municipios se identifican mediante el campo `CODIGOINE` cuyos dos primeros dígitos hacen referencia a la provincia y los tres siguientes identifican al municipio dentro de la provincia, es decir, las cinco primeras posiciones se corresponden con el campo `CodMunicipio` de la consulta *Dónde*. Adicionalmente, tenemos el código de la provincia del municipio, `COD_PROV`; el nombre del municipio, `NOMBRE_ACTUAL`; su superficie en Ha, `SUPERFICIEHa`; y su capital, `CAPITAL`⁵; para la capital, tenemos su altitud, `ALTITUD`; también tenemos el código y el nombre de la comarca a la que pertenece el municipio, `CO_COMARCA` y `DS_COMARCA`.

Nos interesa disponer del campo `CodMunicipio` en la consulta `municipioComarca` para poder unirla con la consulta *Dónde*. Lo podemos definir a partir de las cinco primeras posiciones del campo `CODIGOINE` mediante ejemplos o bien pulsando sobre [«Agregar columna», «De texto», «Extraer», «Primeros caracteres»].

22. En la consulta `municipioComarca`, define una nueva columna llamada `CodMunicipio` con el contenido de las cinco primeras posiciones del campo `CODIGOINE`.

En la consulta `provinciaCA`, para cada provincia, `COD_PROV`, `PROVINCIA` y `CAPITAL`, tenemos los datos de la comunidad autónoma en la que se encuentra: `COD_CA` y `COMUNIDAD_AUTONOMA`.

⁵En la mayoría de los casos la capital coincide con el nombre del municipio, pero si el municipio está compuesto por varios núcleos no necesariamente es así, por ejemplo, en Granada, tenemos los municipios de *Nevada*, *Alpujarra de la Sierra* o *Lecrín*, entre otros, cuyas capitales son distintas del nombre del municipio.

Podemos completar los datos de la consulta `municipioComarca` con los de la consulta `provinciaCA` realizando la unión por los campos de código de provincia correspondientes: `COD_PROV`.

23. Añade a la consulta `municipioComarca` los campos `PROVINCIA`, `COD_CA`, `COMUNIDAD_AUTONOMA` y `CAPITAL` de la consulta `provinciaCA`, y selecciona exclusivamente los datos de la provincia que tienes asignada (en mi caso *Granada*).

Podemos usar la columna `CodMunicipio` para unir las consultas `Dónde` y `municipioComarca` y, de esta forma, enriquecer la dimensión `Dónde` con el resto de datos asociados al municipio.

24. Añade a la consulta `Dónde` los campos `COD_PROV`, `SUPERFICIEHa`, `CAPITAL`, `ALTITUD`, `CO_COMARCA`, `DS_COMARCA`, `PROVINCIA`, `COD_CA`, `COMUNIDAD_AUTONOMA` Y `CAPITAL.1` de la consulta `municipioComarca`.

Asociados a los campos relativos a la superficie y altitud, tenemos las consultas `codSuperficie` y `codAltitud` donde se asocian códigos a intervalos de valores. La superficie en la consulta `codSuperficie` viene dada en Km^2 mientras que en la consulta `Dónde` está en Ha .

25. En la consulta `Dónde`, sustituye la columna `SUPERFICIEHa` por una nueva columna `Superficie` en la que exprese la superficie en Km^2 .
26. Asocia a los valores de las columnas `Superficie` y `ALTITUD` de la consulta `Dónde` los códigos de las consultas `codSuperficie` y `codAltitud`, según el tramo en el que se encuentren (en la sección 3.4 se ha hecho para el número de habitantes⁶). Llama a los nuevos campos definidos `CodSuperficie` y `CodAltitud`, respectivamente.
27. Incluye en la consulta `Dónde` las descripciones asociadas a los códigos `CodTamMunicipio`, `CodSuperficie` y `CodAltitud` que se pueden encontrar en las consultas `codTamMunicipio`, `codSuperficie` y `codAltitud`, respectivamente.

En las herramientas de consulta es frecuente que los literales se presenten ordenados alfabéticamente. En algunos casos esto es adecuado (por ejemplo, los nombres de municipios) pero en otros, cuando representan tramos ordenados o cantidades, sería más adecuado ordenarlos mediante la cantidad que representan que generalmente no coincide con el orden alfabético. El orden correcto suele venir dado por el código asociado, siempre que el código quede ordenado alfabéticamente (por ejemplo, si tiene dos posiciones y utilizamos *01* en lugar de *1*). Por este motivo, en estos casos es adecuado fusionar el código y el literal que lo describe. Se da esta situación para los descriptores del tamaño del municipio, de la superficie y de la altitud.

Podemos fusionar dos campos en una consulta seleccionándolos y pulsando sobre [«Transformar», «Columna de texto», «Combinar columnas»], en la ventana que se abre podemos indicar el separador a utilizar y el nombre del nuevo campo. Como resultado, los campos seleccionados para combinar desaparecen de la consulta y son sustituidos por el nuevo campo generado. No hay ningún problema con que desaparezcan los campos `CodSuperficie` y `CodAltitud`, sin embargo, el campo `CodTamMunicipio` debemos mantenerlo porque ha intervenido en la definición de los hechos (mira la *nota importante* en la sección 5). Por este motivo, previamente tenemos que generar una copia del campo `CodTamMunicipio` seleccionándolo y pulsando sobre [«Agregar columna», «General», «Duplicar columna»]; este nuevo campo podemos fusionarlo con la descripción asociada.

28. En la consulta `Dónde`, define los campos `Tamaño_habitantes`, `Superficie_Km2` y `Altitud_m` respectivamente a partir de la fusión de los campos `CodTamMunicipio`, `CodSuperficie` y `CodAltitud`, y la descripción asociada a cada uno de los códigos.

En los nombres de los nuevos campos se ha incluido la unidad en la que están expresados los intervalos, de esta forma podemos eliminar las unidades de cada uno de los valores que aparecen en

⁶En este caso es más sencillo porque el código solo tiene un dígito.

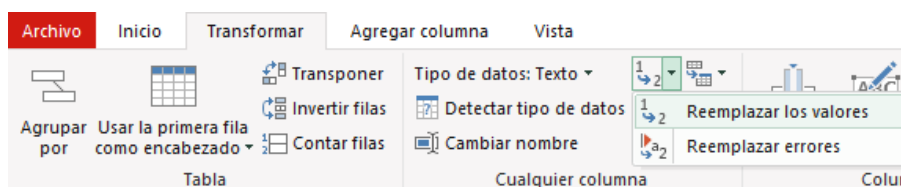


Figura 29: Reemplazar los valores.

las filas. Por ejemplo, en lugar de tener el valor de 101 a 200 habitantes podemos tener de 101 a 200, de esta manera se pueden presentar de una forma más compacta en los informes. En cada caso, podemos sustituir el literal que indica la unidad de medida por el literal vacío (« habitantes», « Km2» y « m.») seleccionando la columna correspondiente y pulsando sobre [«Transformar», «Cualquier columna», «Reemplazar los valores»] (figura 29).

Si miramos los valores de los distintos campos, resalta el campo DS_COMARCA cuyos valores están en mayúsculas. Podemos transformar los valores de manera que solo la primera letra de cada palabra sea mayúscula seleccionando la columna del campo y pulsando sobre [«Transformar», «Columna de texto», «Formato», «Poner En Mayúsculas Cada Palabra»]. Adicionalmente, si se desea, se pueden sustituir los artículos o preposiciones entre palabras por sus equivalentes sin que empiecen por mayúscula.

29. En la consulta Dónde, para los campos Tamaño_habitantes, Superficie_Km2 y Altitud_m, reemplaza los nombres de las unidades respectivas en los valores por el literal vacío. Transforma los valores del campo DS_COMARCA de manera que se escriban en minúscula y comenzando cada palabra en mayúscula.

Para acabar con la definición de la consulta de la dimensión, vamos a renombrar los campos de manera que tengan nombres uniformes.

30. Renombra los campos siguientes de la consulta Dónde:

- COD_PROV → CodProvincia.
- CAPITAL → CapitalMunicipio.
- ALTITUD → Altitud.
- CO_COMARCA → CodComarca.
- DS_COMARCA → Comarca.
- PROVINCIA → Provincia.
- COD_CA → CodCA.
- COMUNIDAD_AUTONOMA → CA.
- CAPITAL.1 → CapitalProvincia.

Reordena los campos para que aparezcan en el orden siguiente:

- IdDónde
- CodMunicipio
- Municipio
- CapitalMunicipio
- CodComarca
- Comarca
- CodProvincia
- Provincia
- CapitalProvincia

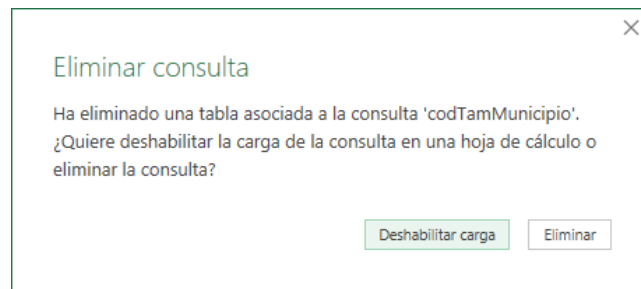


Figura 30: Deshabilitar la carga de una consulta.

- CodCA
- CA
- CodTamMunicipio
- Tamaño_habitantes
- Altitud
- Altitud_m
- Superficie
- Superficie_Km2

6.3. Enriquecer y simplificar los hechos

Nuestro foco de atención son los datos de padrón, la medición que nos interesa principalmente es el número de habitantes. Asociada al número de habitantes podría interesarnos también la densidad de población (número de habitantes por Km^2).

Disponemos de los datos de la superficie de los municipios en la consulta **Dónde**. A diferencia de la medición **Habitantes**, la medición **Superficie** no depende de la dimensión **Cuándo**, por eso es correcto que esté en la dimensión **Dónde**.

Hay herramientas que permiten definir mediciones derivadas utilizando datos ubicados en las dimensiones, otras necesitan que todos los datos estén en los hechos. En cualquier caso, si todos los datos están en los hechos, la definición es más sencilla. Por este motivo, vamos a enriquecer los hechos añadiendo la medición **Superficie** que actualmente se encuentra en la consulta **Dónde**.

Por otro lado, podemos obtener la medición **Habitantes** a partir de las mediciones **Mujeres** y **Hombres**.

31. Añade a la consulta **Padrón** el campo **Superficie** de la consulta **Dónde**. Por ser derivado de los campos **Mujeres** y **Hombres**, elimina el campo **Habitantes**.

6.4. Salvar los datos

Pulsando sobre [«Inicio», «Cerrar y cargar»], podemos ver el resultado de las transformaciones definidas en hojas de *Excel*.

En *Excel*, en la parte derecha, aparece una ventana titulada «Consultas de libro». Pulsando sobre el nombre de cada consulta, se activa la hoja *Excel* correspondiente con los resultados obtenidos. Las hojas conservan sus nombres genéricos (**Hoja1**, **Hoja2**, ...), para trabajar posteriormente con ellas es más adecuado asignar a las hojas el nombre de las consultas. Esto hay que hacerlo manualmente.

Los resultados de algunas consultas no los vamos a necesitar después, podemos eliminar las hojas correspondientes. Al borrar una hoja, nos pide confirmación indicando que se perderán los datos de la hoja; al aceptar que se borre la hoja, nos pregunta si también queremos eliminar la consulta o bien solo deshabilitar su carga (figura 30). En este caso, solo queremos **deshabilitar la carga**, si la borramos perderemos parte de nuestro trabajo. Si posteriormente queremos habilitar la carga de una consulta deshabilitada, podemos hacer esta operación desde el menú contextual de la consulta en la ventana «Consultas de libro» de *Excel*.

32. Para las consultas **Provincia**, **Cuándo**, **Dónde** y **Padrón**, asigna a cada hoja de *Excel* el nombre de la consulta que la genera, elimina el resto de hojas y desactiva las consultas asociadas a ellas, y guarda los cambios en el archivo *Excel*.

Los datos obtenidos como resultado de esta actividad se necesitan para las prácticas siguientes (sin ellos no se pueden realizar).

Una posibilidad para no olvidarlos, es subirlos a PRADO junto al resultado de esta actividad.