**Caso práctico:** Almacén de datos para el análisis de indicadores de crisis en la eurozona

## Solución PRA3- Explotación de datos

**Nombre estudiante: Lukaz Martin Doehne**

**Aula 3**

**Índice de contenidos**

[**1. Diseño del modelo OLAP**](#_heading=h.xsqkpkz4gv5a) **3**

[1.1. Creación del proyecto y revisión del origen de datos](#_heading=h.q1dgzf6a5pcm) 3

[1.2. Creación de la vista del origen de datos](#_heading=h.oyg0frqwd0md) 3

[1.3. Creación e implementación de los cubos](#_heading=h.h05ksqtlmm3a) 3

[1.4. Configuración de jerarquías, dimensiones y atributos](#_heading=h.adhaqdnqp7ii) 3

[1.5. Procesado y resolución de errores](#_heading=h.yory08eqfg51) 3

[**2. Explotación del modelo OLAP**](#_heading=h.nzj5luwgw4q3) **3**

# Diseño del modelo OLAP

Esta actividad consiste en realizar la implementación de cubos multidimensionales para la explotación de la información, que va a servir para el análisis de datos y la posterior toma de decisiones.

En esta PRA3 se plantea cómo realizar el diseño de un modelo multidimensional online analytical processing (OLAP) para el análisis multidimensional de la información disponible en el almacén de datos. Con este fin se va a llevar a cabo lo siguiente:

**1)** Relación de precios de consumo en alimentación por país en marzo del año 2022, ordenados por HICP de mayor a menor.

**2)** Obtener la media del HICP para España por año. Mostrar los resultados redondeados a dos decimales.

**3)** Visualizar la evolución del HICP por año para España, Francia, Italia y Alemania. ¿En qué país ha habido un incremento más alto? Se recomienda utilizar Excel para analizar esta evolución.

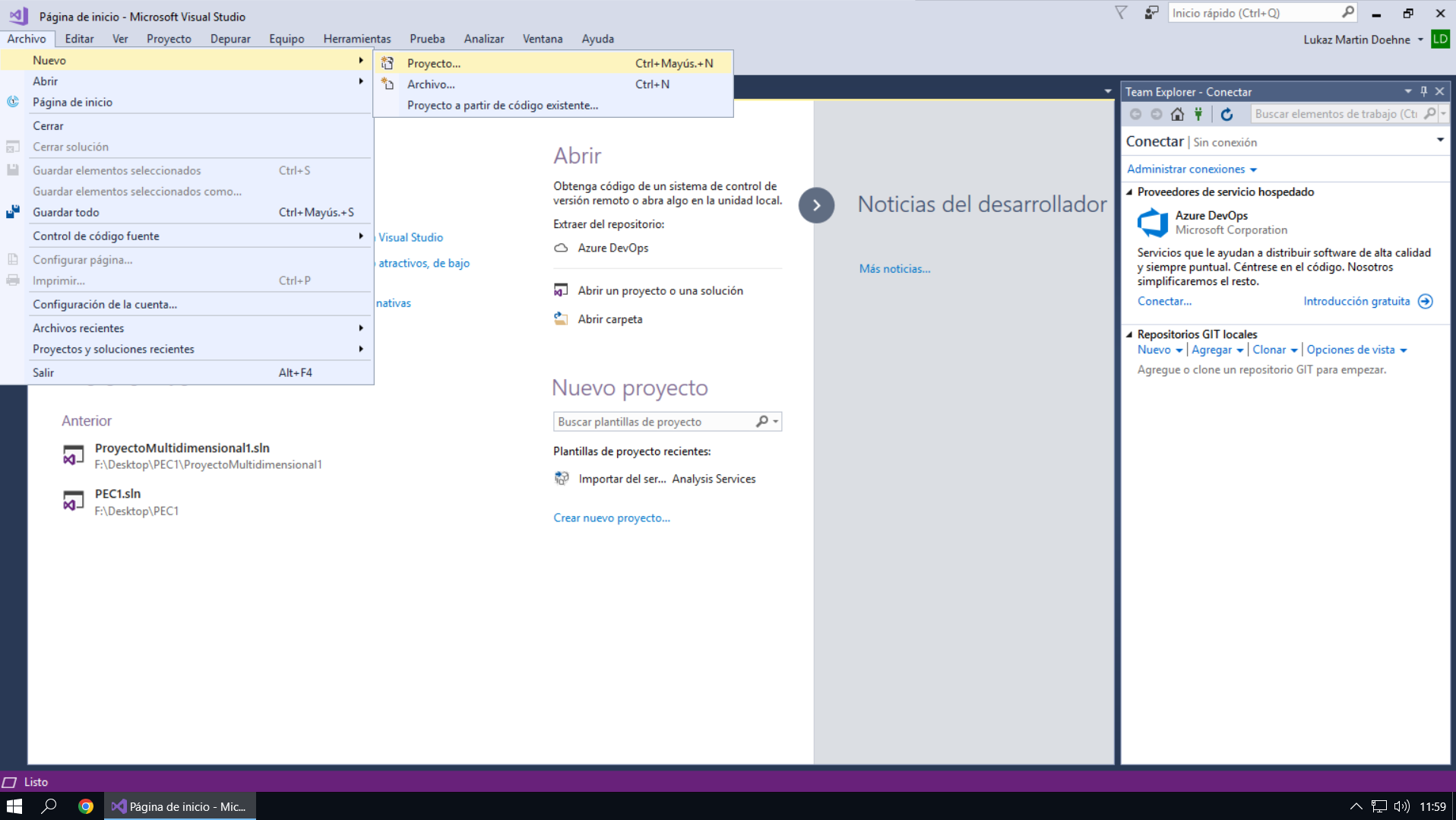
**4)** Obtener el top cinco de fines de consumo (COICOP) durante 2022, según la media de su HICP, ordenados por este valor de mayor a menor.

**5)** Obtener el precio medio del consumo de gas y electricidad por banda de consumo en España durante el año 2022. ¿Varía el precio por banda de consumo? Si es así, ¿qué interpretación se le puede dar a esta variación?

**6)** Visualizar la evolución por año de precios de electricidad para España y Austria. ¿En qué país ha habido un incremento más alto? Se recomienda utilizar Excel para analizar esta evolución.

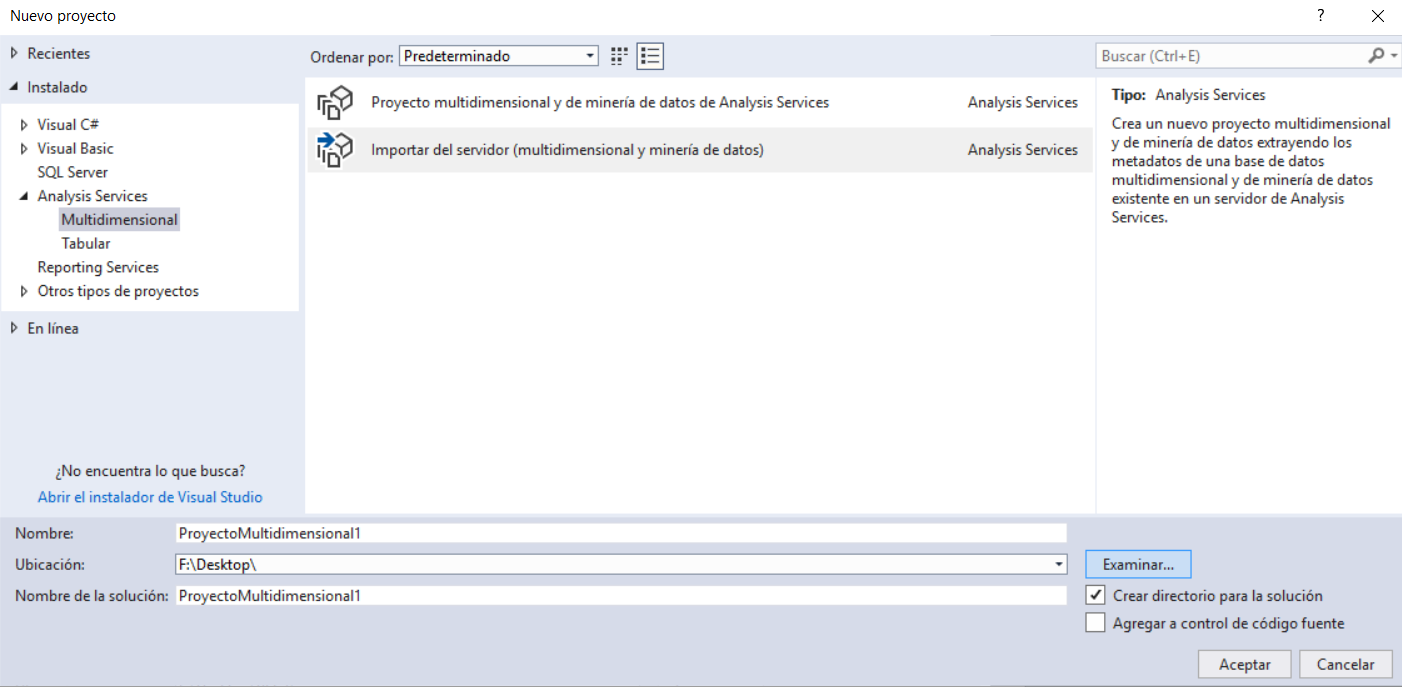
## Creación del proyecto y revisión del origen de datos

Iniciamos Visual Studio 2017 y seleccionamos *Nuevo proyecto*.

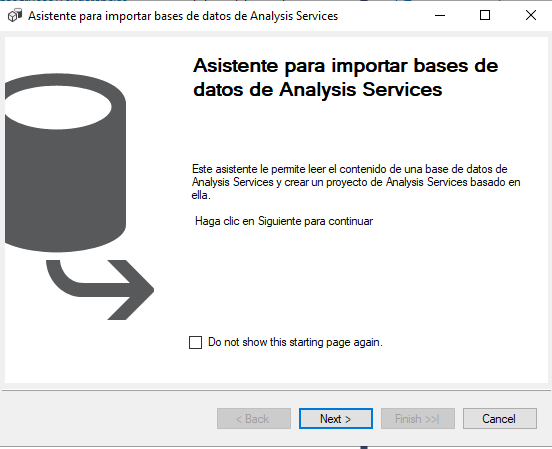


El tipo de proyecto que nos interesa es *Proyecto multidimensional y de minería de datos*, que permite crear cubos analíticos y realizar proyectos de minería de datos.

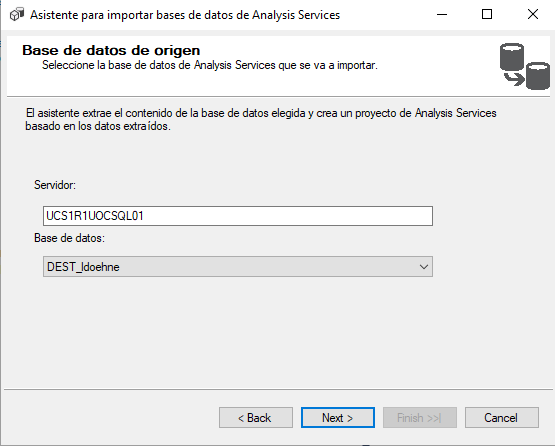
Después hemos de seleccionar la opción *Importar del servidor (multidimensional y minería de datos)*.



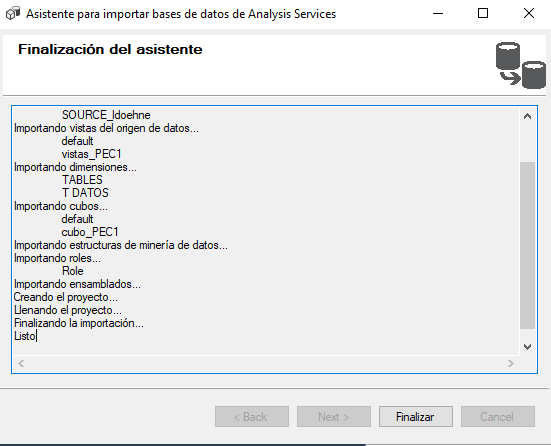
Le ponemos el nombre DBA\_PRA3 y lo ubicamos dentro de la unidad de equipo **F:\** para que persistan los cambios que realizamos. Seleccionamos *Aceptar* para iniciar el asistente e importar la base de datos de Analysis Services.



Introduciremos los datos del servidor de Analysis Services (UCS1R1UOCSQL01) y seleccionamos nuestra base de datos **DEST\_ldoehne**. Pulsamos en *Next* y se importará la base de datos.

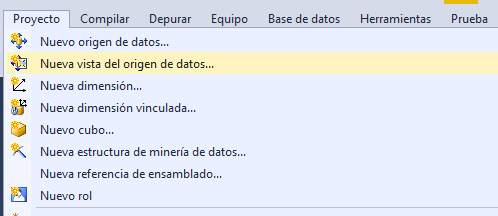


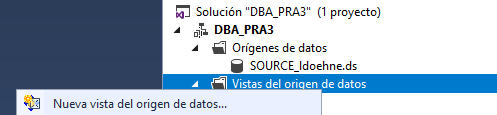
Seleccionamos *Next* y *Finalizar*. Comprobamos que se ha creado un origen de datos **SOURCE\_ldoehne** que se corresponde con vuestra base de datos en el servidor. Además, se crea una vista de datos por defecto, con un cubo default, una dimensión tables y un rol. Estos objetos no deben ser borrados ni utilizados en la realización de la práctica.



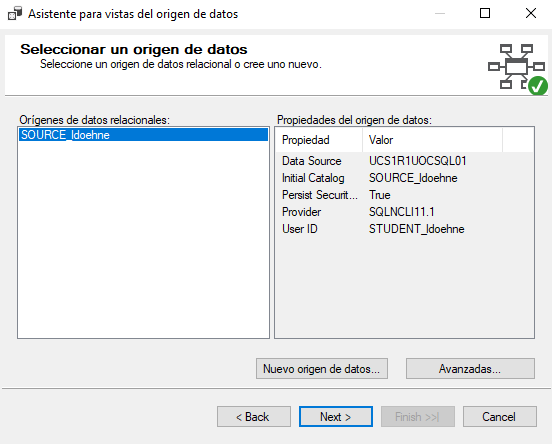
## Creación de la vista del origen de datos

A continuación, se procede a crear las vistas del origen de datos. Para crear cada una de las vistas, se elige *Nueva vista del origen de datos* del menú contextual *Vistas del origen de datos* o a en la opción con el mismo nombre del menú *Proyecto*.





En el asistente, se selecciona *Next*.



Se crea una vista para cada uno de los cubos. Así definimos las vistas:

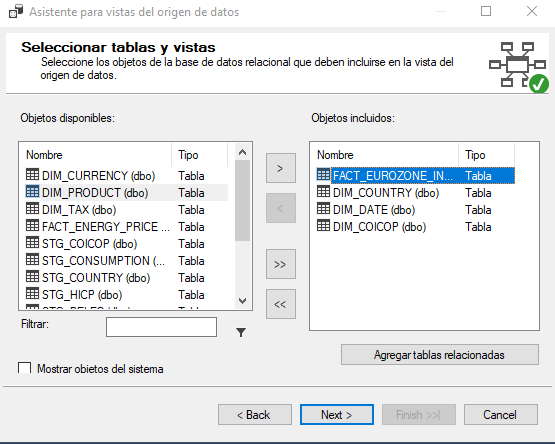
● **Vista\_Indicadores\_Eurozona**: dará acceso a las tablas para diseñar el cubo de análisis de los indicadores de la eurozona (HICP).

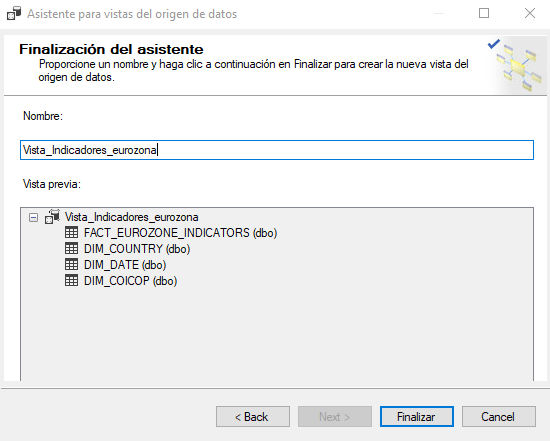
● **Vista\_Precio\_Energia**: dará acceso a las tablas para diseñar el cubo de análisis de los precios de la energía (gas y electricidad).

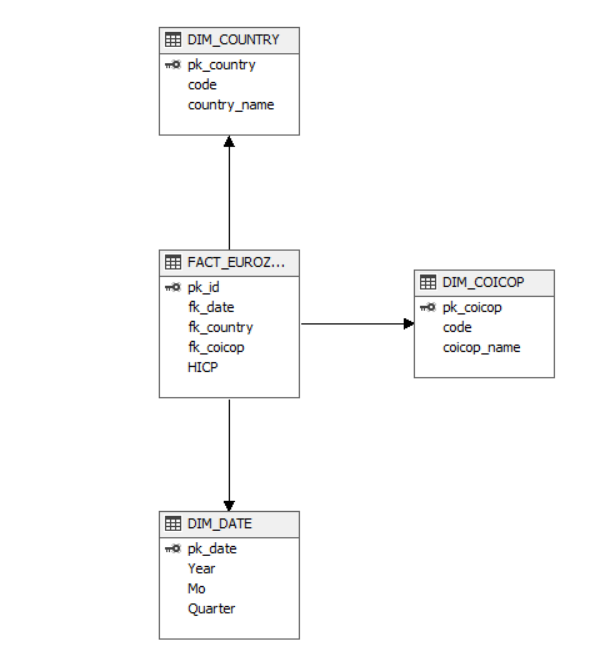
Se seleccionan las tablas necesarias para definir cada una de las vistas de orígenes de datos. Una forma rápida de hacerlo es seleccionar una tabla de hechos y después elegir la opción «*add related tables*»; esta opción utilizará las claves foráneas para traer las tablas relacionadas.

Si se vuelve a seleccionar esta opción, se obtendrán las relacionadas a las ya añadidas. En todo caso, siempre es posible seleccionar las tablas de manera manual.

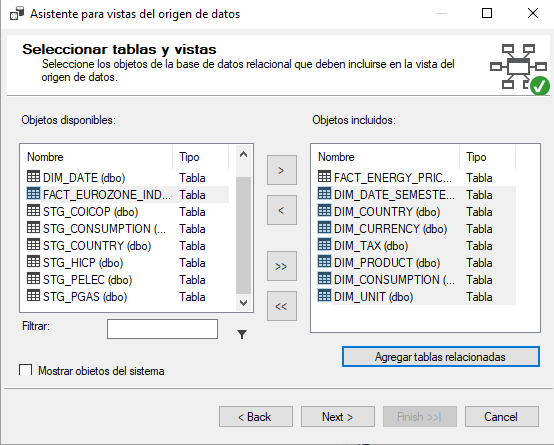
**Vista\_Indicadores\_Eurozona**

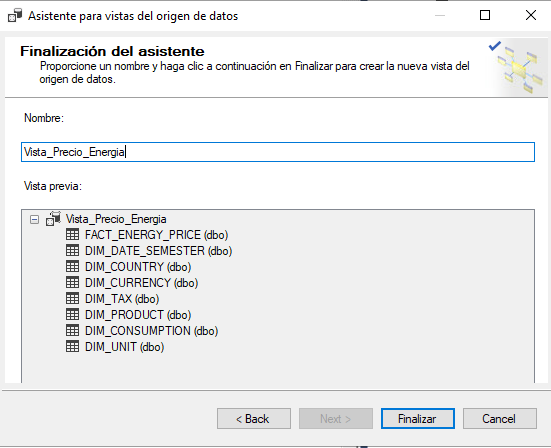


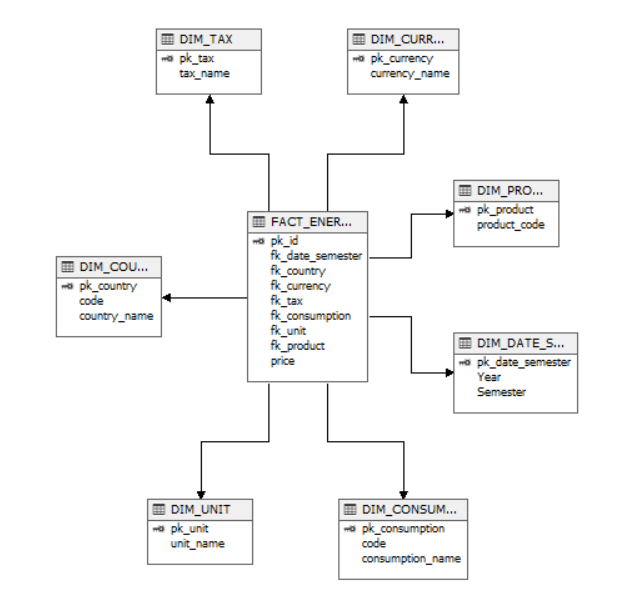




**Vista\_Precio\_Energia**

****

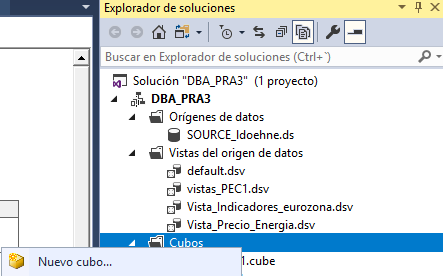
****

****

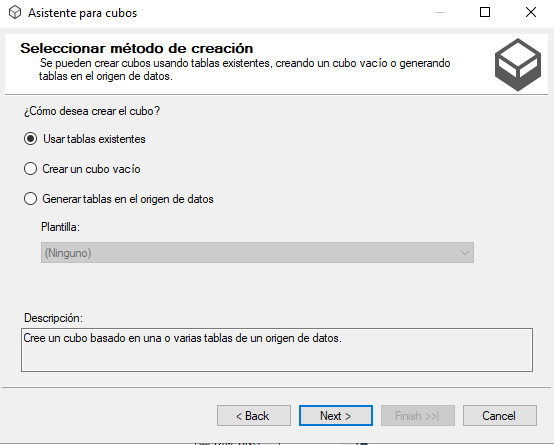
## Creación e implementación de los cubos

Después del paso anterior de creación de las vistas para los orígenes de datos, el siguiente es la creación de los cubos. Este paso nos permitirá realizar los análisis OLAP para la explotación de los datos del almacén de datos.

Crearemos cada cubo desde el menú contextual de cubos, seleccionando la opción «Nuevo cubo».

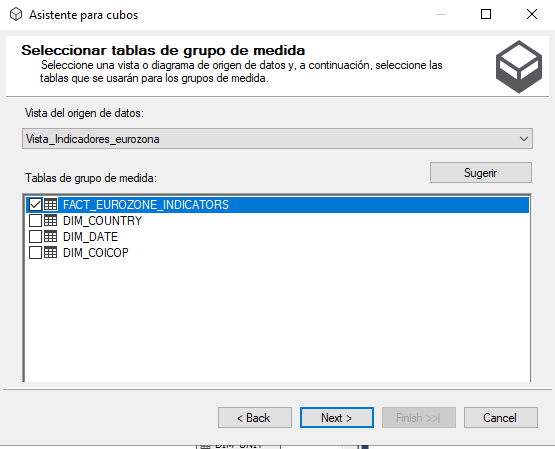


En el asistente, indicamos que debe utilizar tablas ya existentes y escogemos cada una de las vistas creadas en el paso anterior para crear el cubo correspondiente.

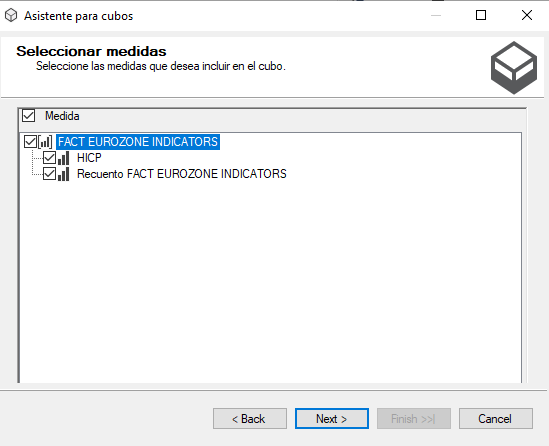


**Cubo\_Indicadores\_Eurozona**

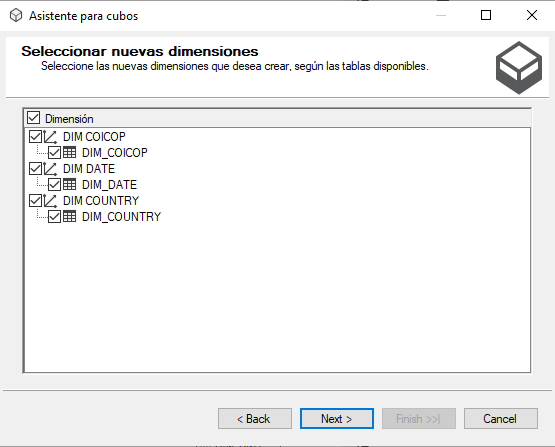
Seleccionamos la vista «Vista\_Indicadores\_Eurozona» y escogemos la tabla de hechos.

****

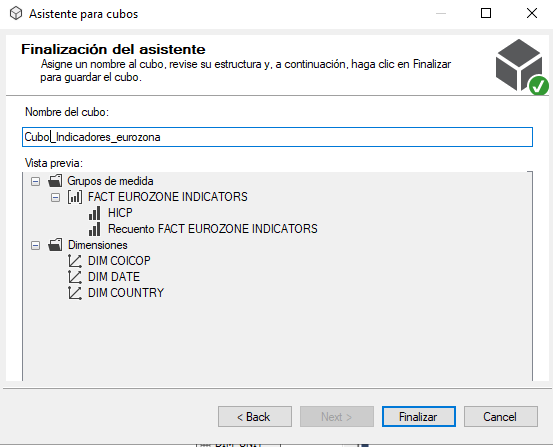
A continuación, seleccionaremos las métricas. Además de las definidas, se añade el recuento para utilizarlo en los análisis y conocer, por ejemplo, el número de métricas realizadas al día.



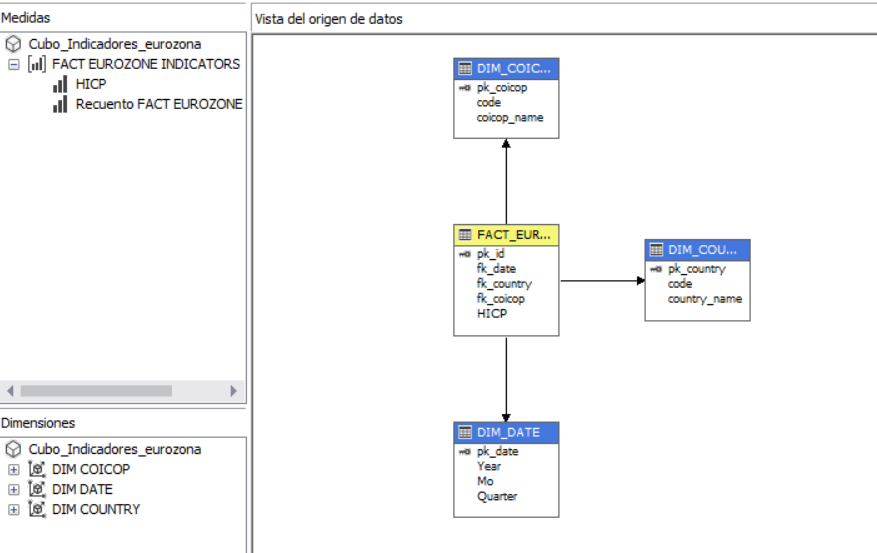
A continuación, seleccionaremos las dimensiones.



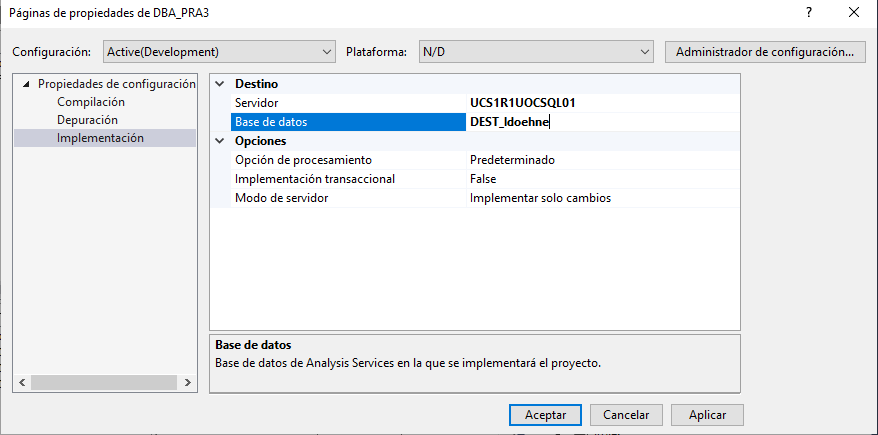
Al finalizar el asistente, visualizamos la pantalla con el resumen del cubo que queréis crear y le asignamos un nombre.



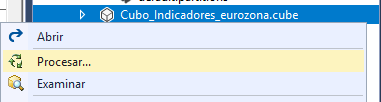
Seleccionamos «Finalizar» y en la pantalla se mostrará el modelo multidimensional del cubo que se acaba de crear. En la parte izquierda de la pantalla podemos visualizar en forma de árbol las medidas y dimensiones que componen el modelo OLAP del cubo.



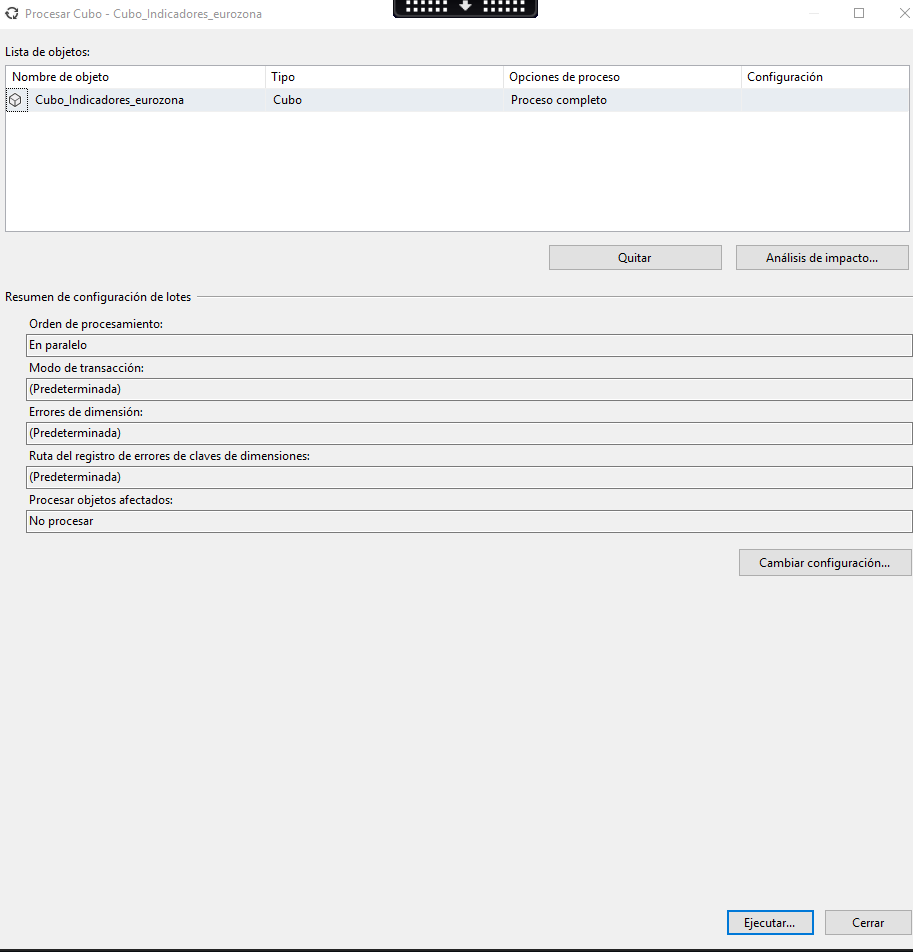
Antes de procesar el cubo vamos a especificar correctamente el servidor al que apuntamos que por defecto está *localhost*. Para ello vamos a las propiedades del proyecto, Implementación.



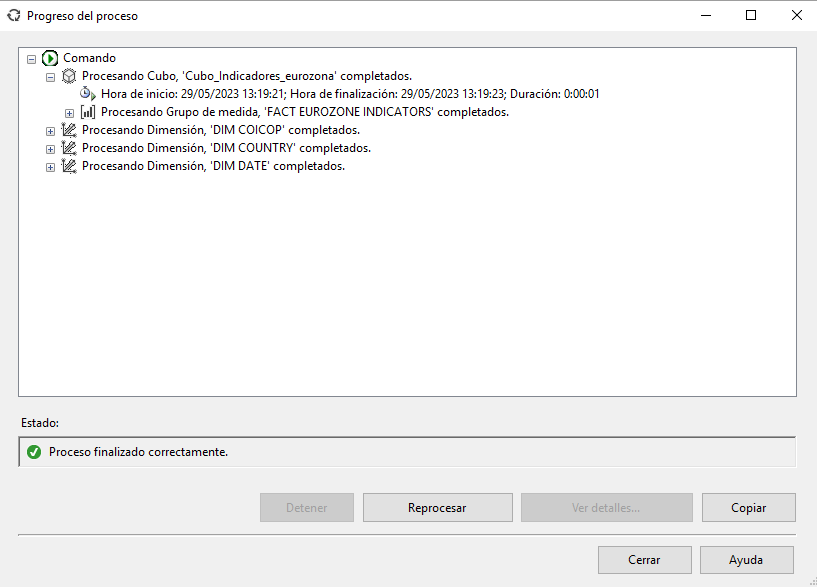
Para poder visualizar los datos del cubo deberéis implementarlo en el servidor de SSAS. Para ello, seleccionaremos la opción «Procesar» en el menú contextual.



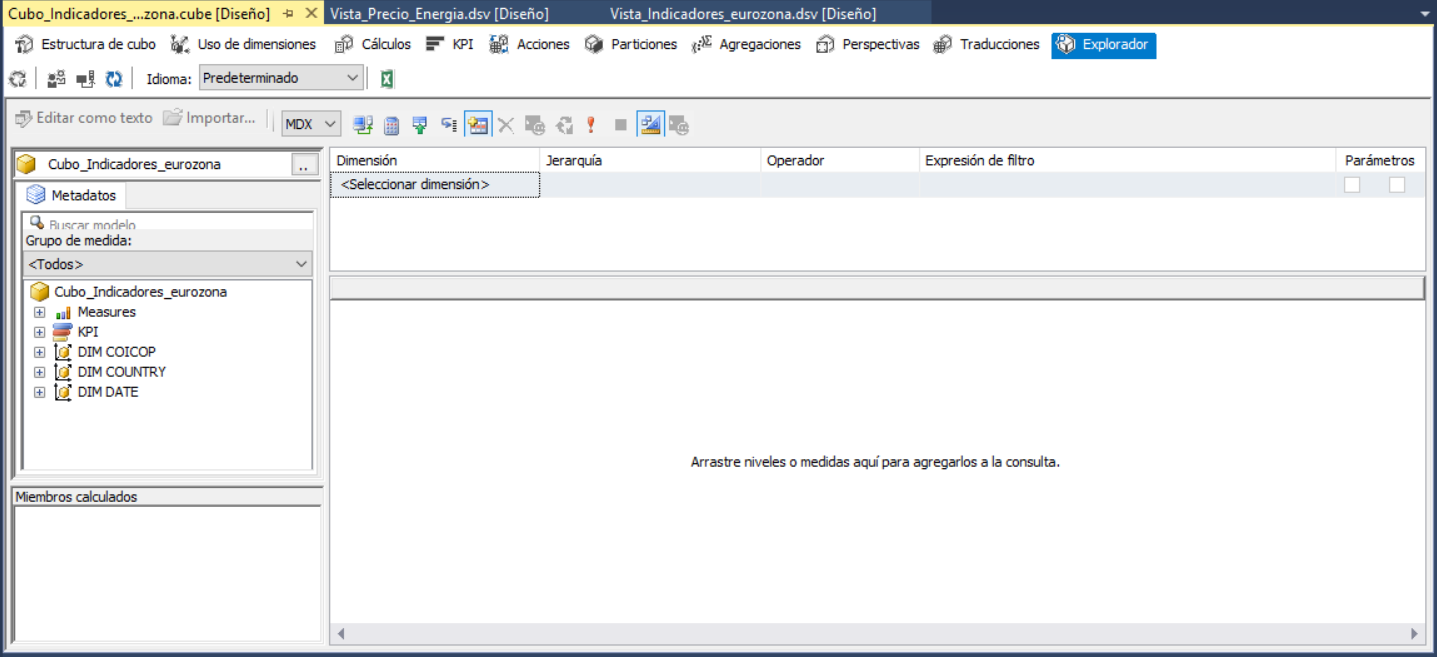
Transcurrido el proceso de implementación, aparecerá la ventana de ejecución del cubo. Seleccionamos «Ejecutar».



Se mostrará la pantalla confirmando que el cubo se ha procesado correctamente.



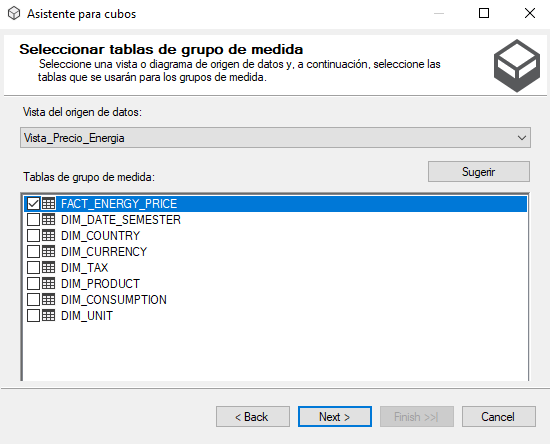
Una vez hecho esto, se puede navegar por el cubo a través de la pestaña del «Explorador».



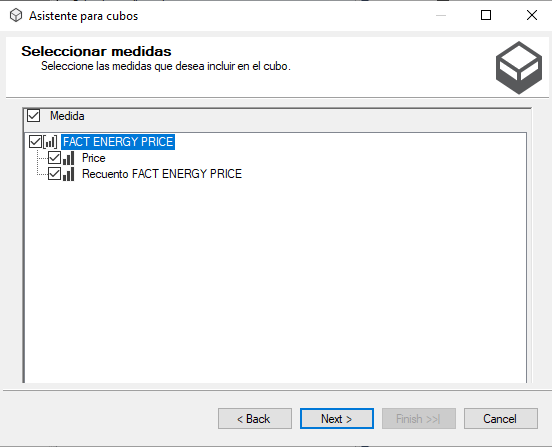
**Cubo\_Precio\_Energia**

Realizamos el mismo proceso para el cubo con la vista de precio energético (electricidad y gas).

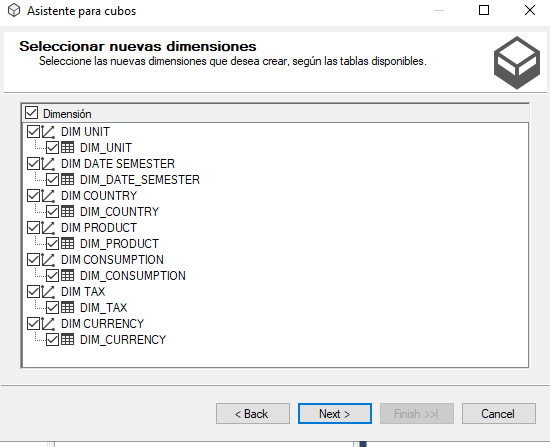
Seleccionamos la vista «Vista\_Precio\_Energia» y pinchamos en la tabla de hechos.



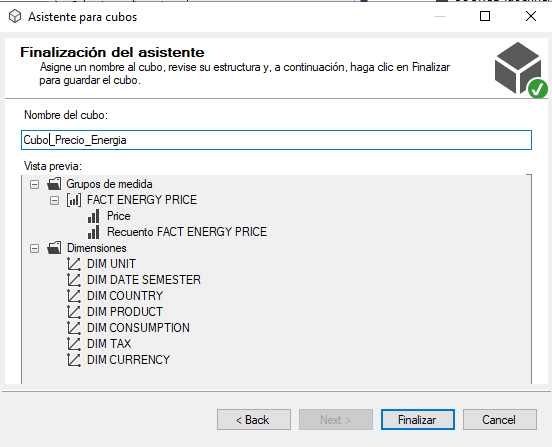
Seleccionamos las medidas que se desean tener implementadas en el cubo.

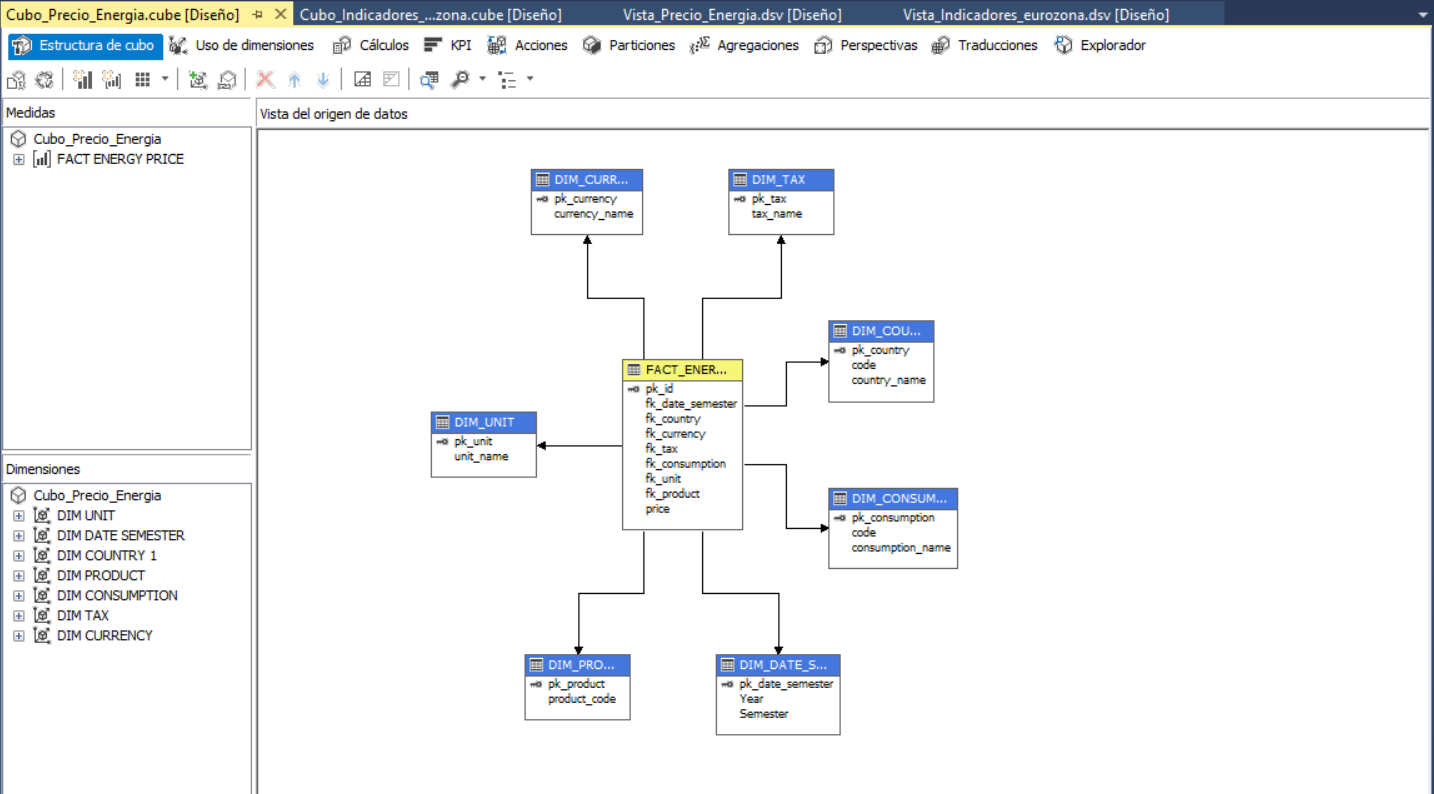


Después escogemos las dimensiones.

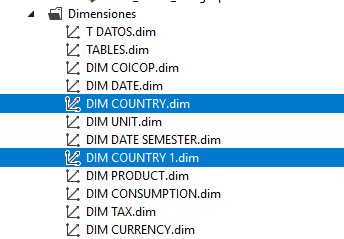


Y, finalmente, le asignamos un nombre adecuado al cubo y finalizamos el asistente.

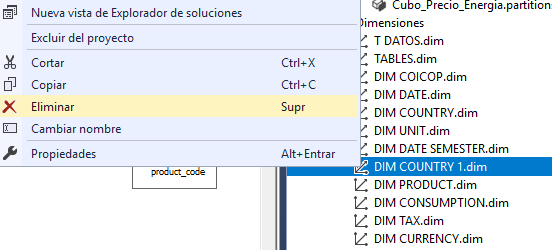




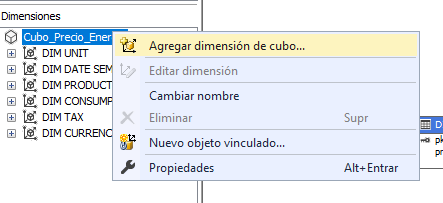
Tras realizar este segundo cubo podréis apreciar cómo algunas dimensiones se han duplicado. Para configurar correctamente las dimensiones conformadas será necesario resolver esta duplicidad.



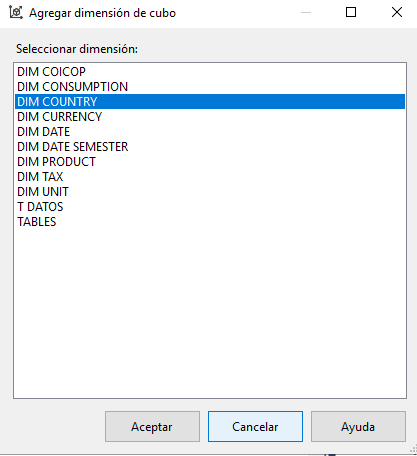
Para ello primero se procede a borrar las dimensiones duplicadas desde el menú contextual.



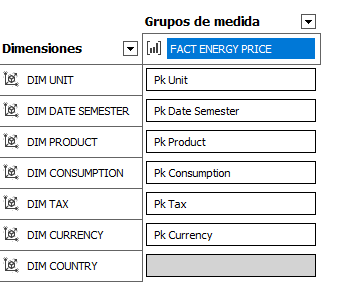
A continuación, deberéis añadir al «Cubo\_Precio\_Energia» la dimensión «DIM COUNTRY» que ya tenía creada. Para ello, desde el diseño del cubo, en la pestaña «Estructura de cubo», se escoge la opción «Agregar dimensión al cubo» desde el menú contextual de las dimensiones.



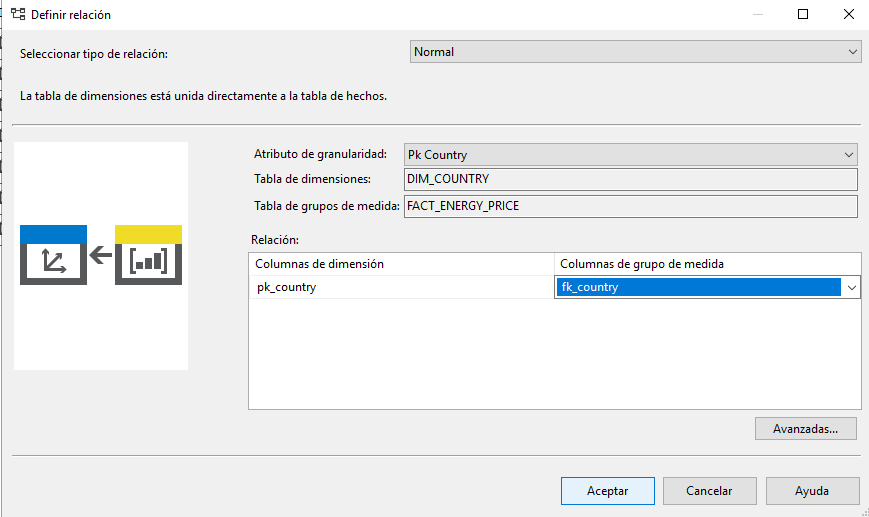
En la ventana que aparece seleccionamos la dimensión que se tiene que agregar al cubo para tener el modelo dimensional completo.



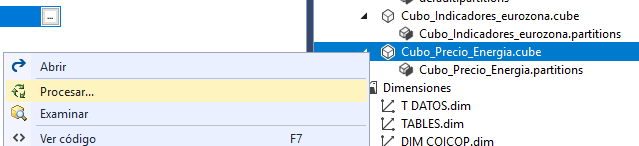
Finalmente, habrá que definir las relaciones entre la tabla de hechos y las nuevas dimensiones. Para ello, se debe ir a la pestaña «Uso de dimensiones» del diseño del cubo.



Se observa que para las nuevas dimensiones no existe relación alguna. Para definirlas, se hace clic en los puntos suspensivos que aparecen al final del cuadro de texto, donde se especifica el campo que se utilizará para relacionar la tabla de hechos con cada dimensión. Como tipo de relación se escoge «Normal» y se seleccionan los atributos y columnas correspondientes.



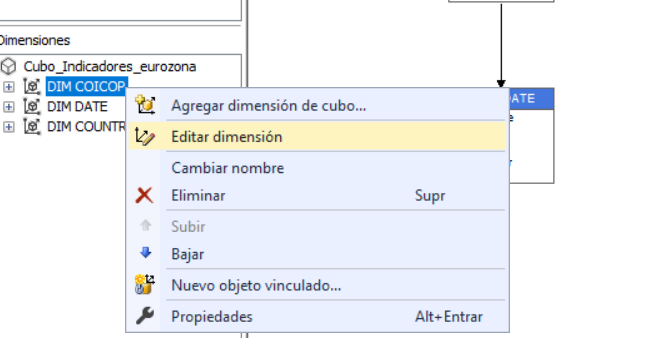
Igual que con el cubo anterior, es necesario procesar el cubo para poder explorarlo.



## Configuración de jerarquías, dimensiones y atributos

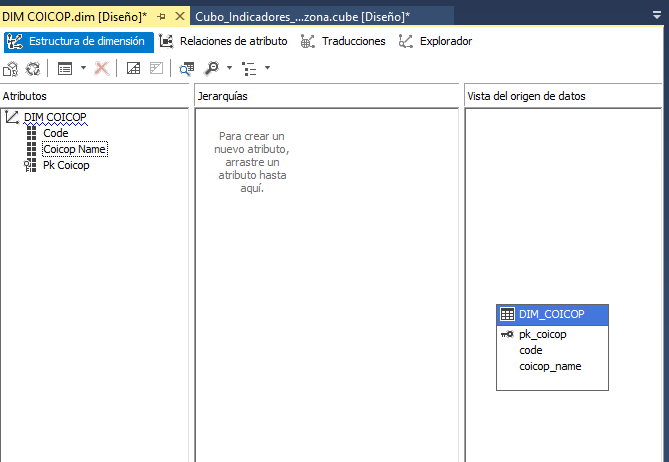
Una vez definidas todas las relaciones del cubo, se trabajará con las dimensiones del proyecto con el objetivo de definir sus atributos y jerarquías. En algunos casos, al tratarse de dimensiones conformadas, es decir, aquellas dimensiones que están compartidas por varias tablas de hechos permitiendo su integración, los cambios realizados tendrán efectos en todos los cubos que las utilicen.

Para definir los atributos y jerarquías de cada dimensión podemos hacer doble clic sobre la dimensión en el apartado «Dimensiones» del «Explorador de soluciones» o elegir la opción «Editar dimensión» desde el menú contextual en la pestaña «Estructura» de cubo.



**DIM COICOP**

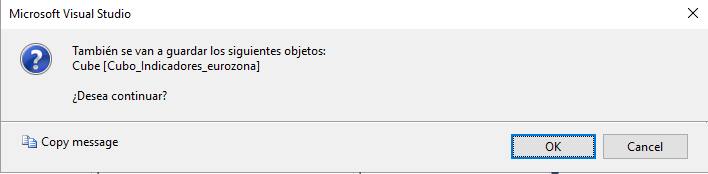
Para editar la dimensión arrastramos los atributos que queremos utilizar desde la zona «Vista del origen de datos» hasta «Atributos».



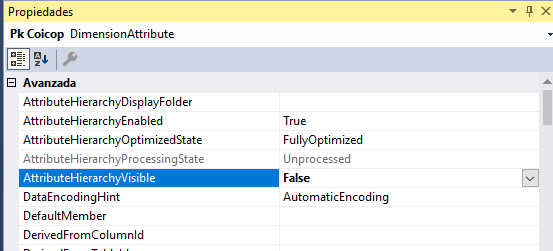
No es necesario seleccionar todos los atributos disponibles, sino sólo aquellos que se utilizarán en el análisis de datos. Opcionalmente, es posible renombrar los atributos de la dimensión.

Tras guardar los cambios y volver a procesar la dimensión, se puede ver una advertencia informando de que debemos crear una jerarquía.

Esto no es necesario para que el cubo funcione, pero, si no se quiere que aparezca la advertencia, se puede seguir la recomendación proporcionada y crear jerarquías sencillas. Sin embargo, en nuestro caso se va a ignorar, siempre teniendo en cuenta que puede haber algún caso en el que puede ser interesante crearla.

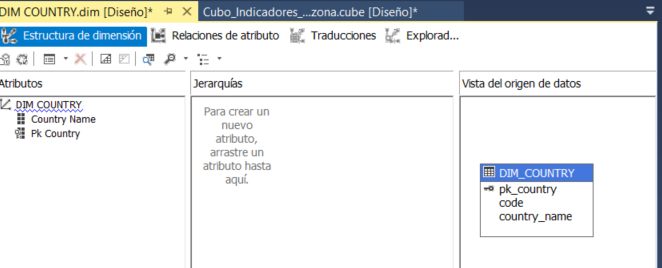


A la hora de analizar los datos los analistas no deberían poder trabajar con valores de tipo clave primaria, dado que estas son de uso interno para la base de datos y, en ocasiones, pueden cambiar entre carga y carga. Además, al tratarse de claves subrogadas, no ofrecen un valor que sea fácil de entender por los analistas. Por todo ello, se oculta seleccionando las propiedades del atributo y cambiando la propiedad a «AttributeHierarchyVisible».

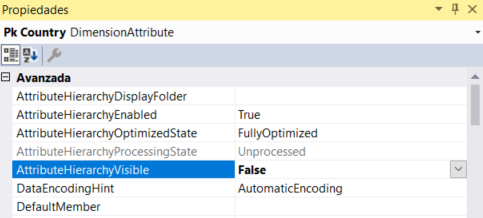


**DIM COUNTRY**

Al igual que en la dimensión anterior, seleccionamos los atributos que queremos y los renombramos si es necesario.

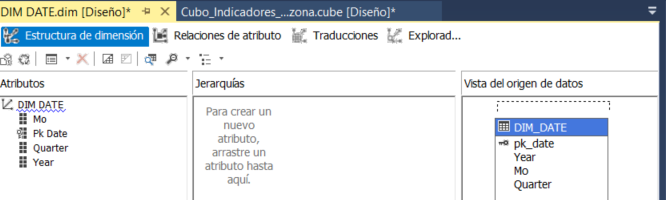


Pk Country se ocultan también. No necesitamos el atributo Code.

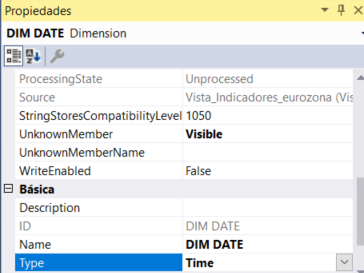


**DIM DATE**

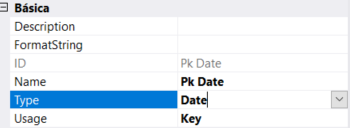
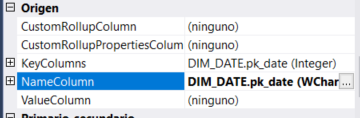
Para la dimensión DATE haremos uso de las variables Mo, Quarter, Year y la clave privada.



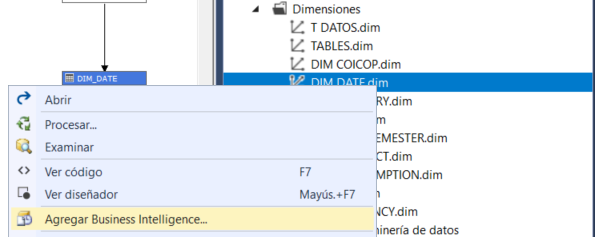
Después, desde la sección «Atributos» de la pestaña de «Estructura de dimensión» visualizamos las propiedades de la dimensión y cambiaremos el tipo a «Time».

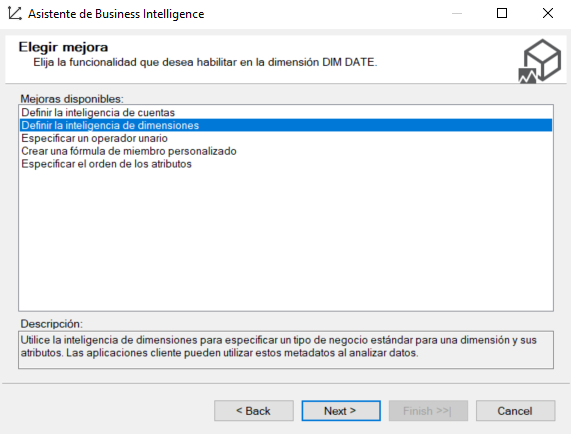


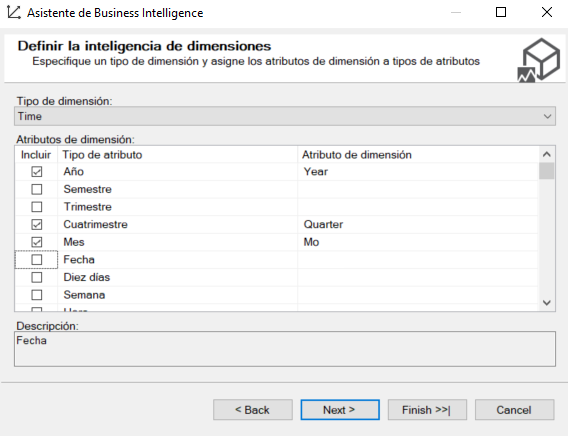
Modificamos las propiedades de «Pk Date» para cambiar el tipo a «date» y «NameColumn» a «pk\_date».

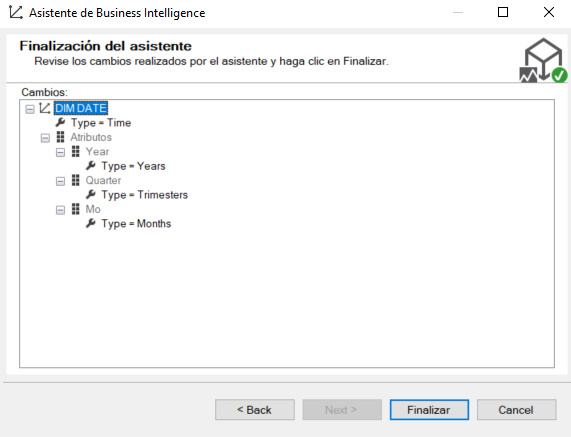


Finalmente, creamos la jerarquía de días, meses y años, con la opción «Agregar Business Intelligence…».

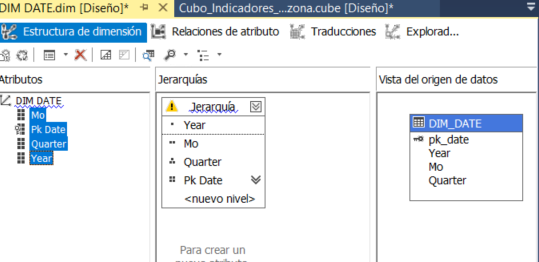




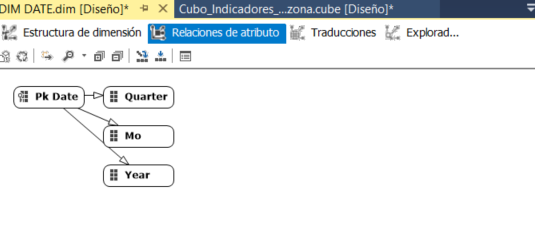


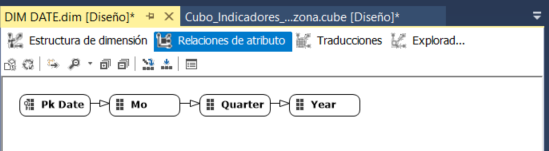


A continuación, se definirá la jerarquía temporal. Usamos «Pk date» para el gránulo de la dimensión.



Sin embargo, aparecerá la advertencia sobre la necesidad de crear relaciones de atributos, como en el caso de la «DIM COUNTRY». Por ello, y de la misma manera que antes, se crearán las relaciones pertinentes.

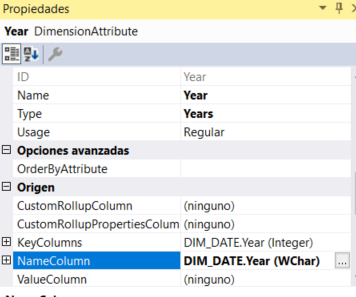
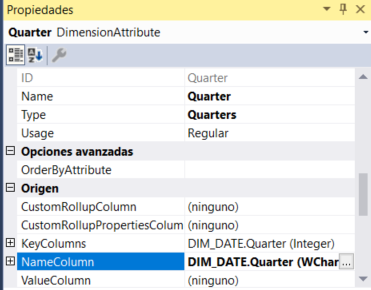
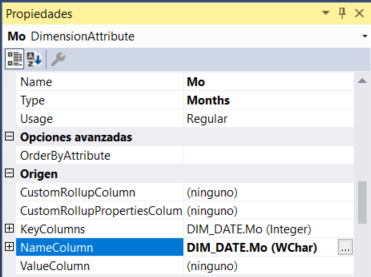




Después se definirán las relaciones como «Rígido».

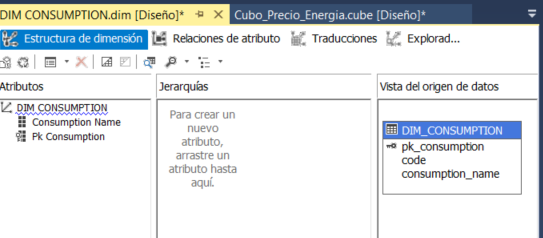


Si vuelven a aparecer errores al implementar, se agregan los atributos necesarios en la propiedad «KeyColumns» y se le asignará el «NameColumn» correspondiente.



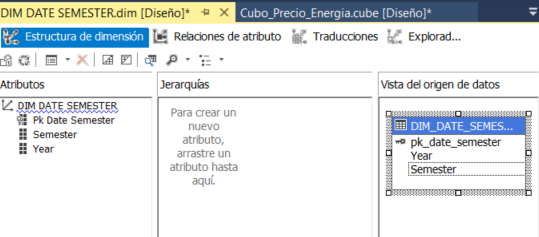
**DIM CONSUMPTION**

Para la dimensión Consumption solo añadiremos el atributo consumption\_name y especificamos PK consumption con «AttributeHierarchyVisible» a FALSE.

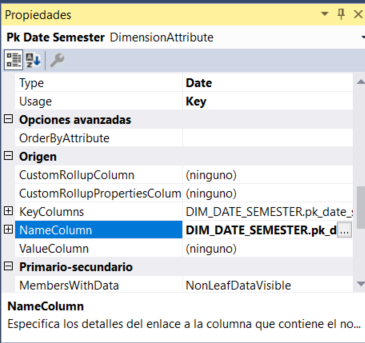


**DIM DATE SEMESTER**

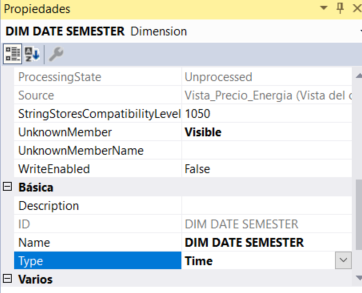
De manera similar importamos todos los atributos a la dimensión.



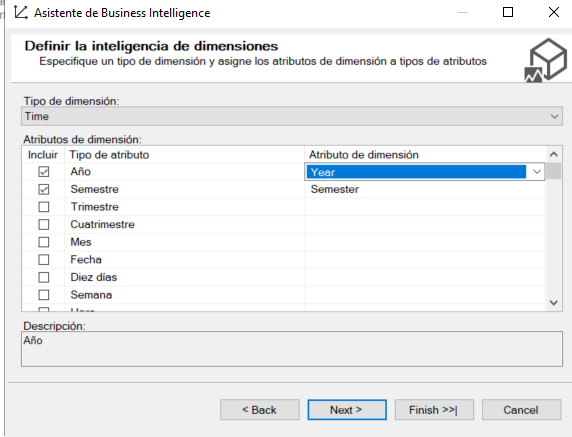
Asignamos «AttributeHierarchyVisible» de la variable PK Date Semester a FALSE, el tipo (Type) como Date y el NameColumn cómo pk date semester.

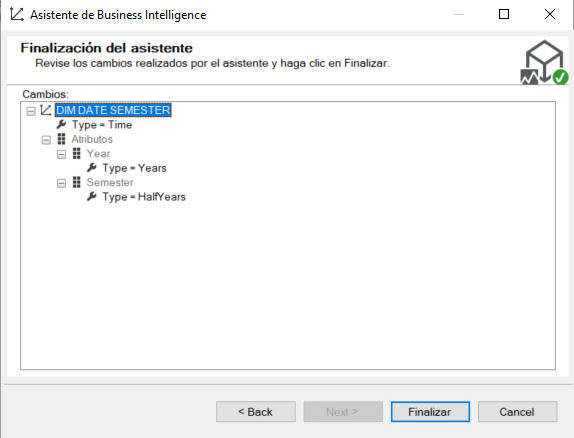


En las propiedades de DIM SEMESTER escogemos como Tipo TIME.

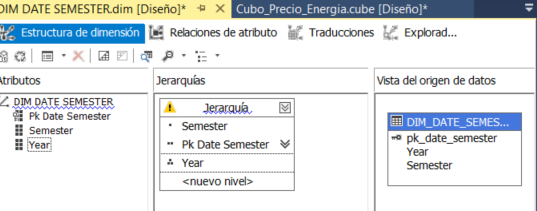


Similar al apartado de DIM DATE seleccionamos el Asistente de Business Intelligence y seleccionamos los atributos Año y Semestre.

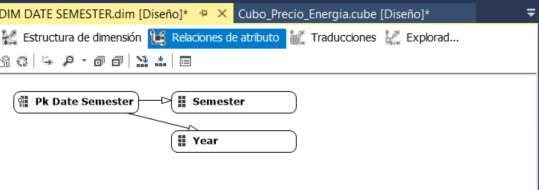




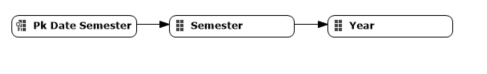
A continuación, se definirá la jerarquía temporal. Usamos «Pk date semester» para el gránulo de la dimensión.



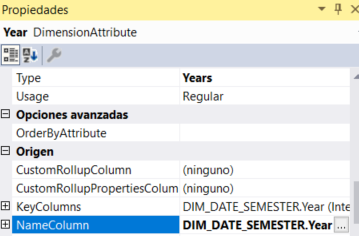
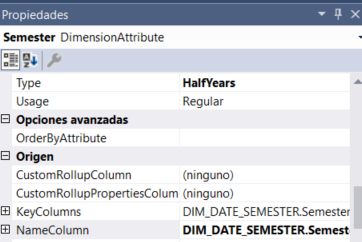
Similar al apartado de DIM DATE, aparecerá la advertencia sobre la necesidad de crear relaciones de atributos. Por ello, y de la misma manera que antes, se crearán las relaciones pertinentes.



Seleccionamos las relaciones como Rígido y nos queda el siguiente diagrama.



Especificamos y comprobamos los valores de Type y NameColumn.



## Procesado y resolución de errores

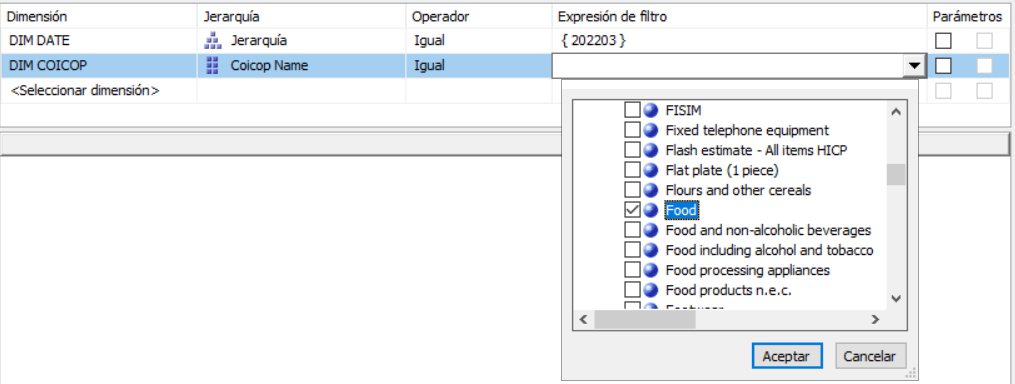
# Explotación del modelo OLAP

**1) Relación de precios de consumo en alimentación por país en marzo del año 2022, ordenados por HICP de mayor a menor.**

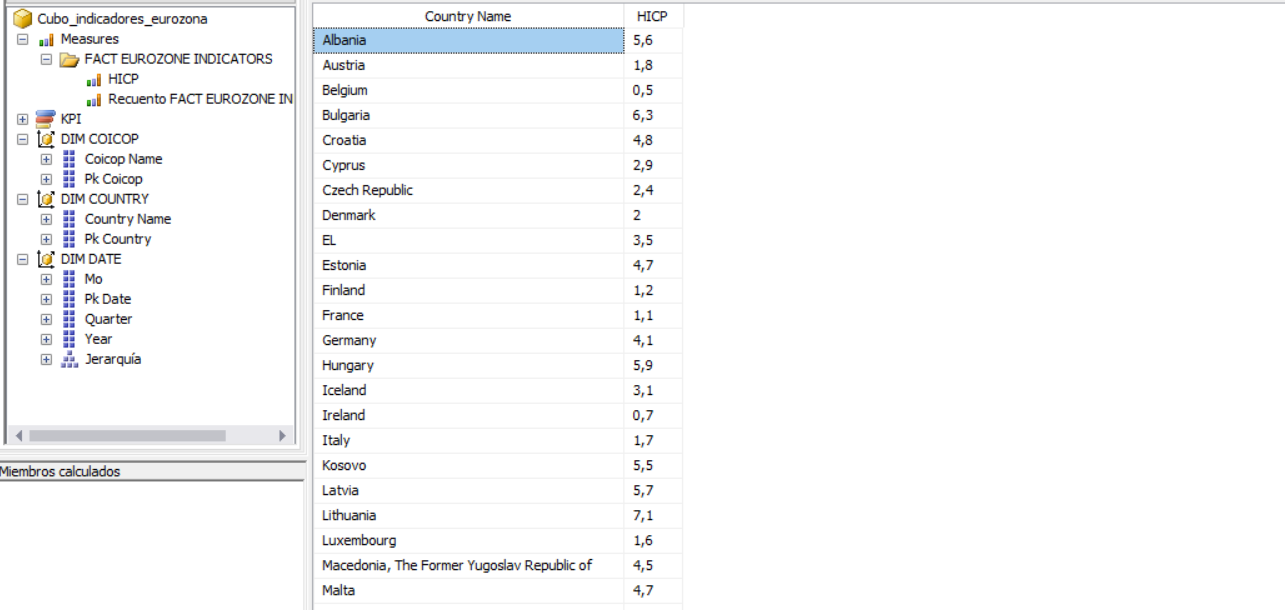
Para realizar este apartado usaremos la pestaña del explorador de cubos de Visual Studio. Desde el Explorador del cubo arrastramos la dimensión DIM DATE y elegimos la expresión de filtro 202203 (marzo del año 2022).



Seguidamente, arrastramos la dimensión COICOP para seleccionar los productos de alimentación. Para ello seleccionamos la expresión de filtro *Food*.



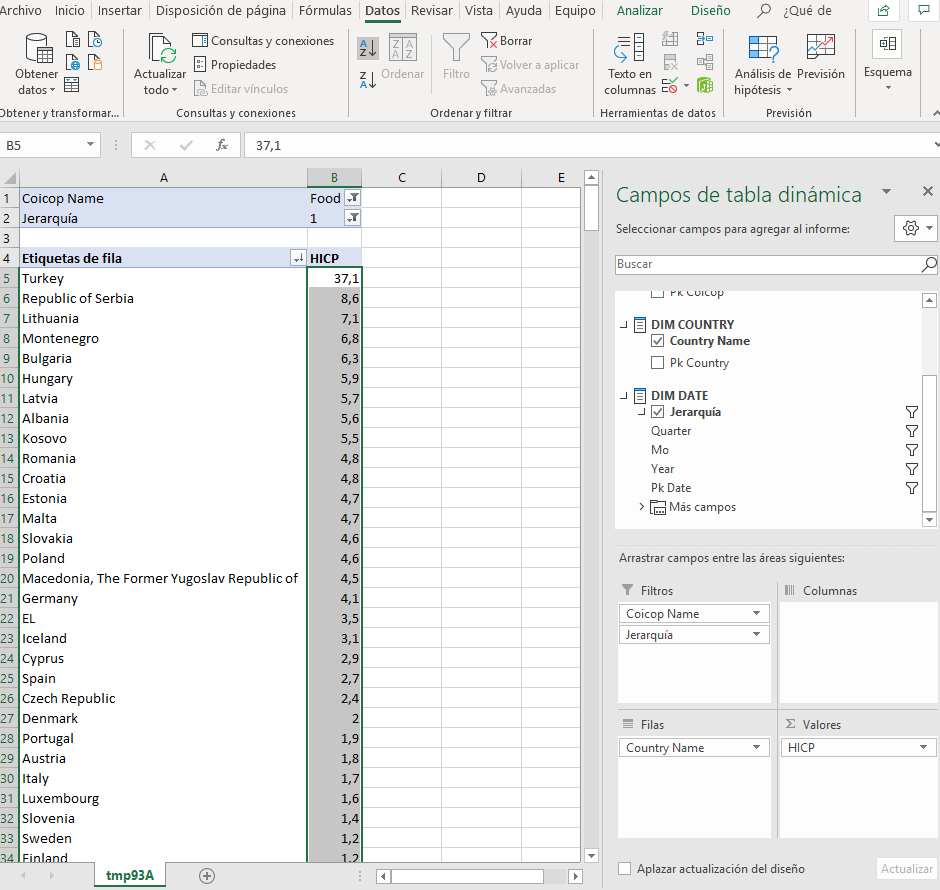
Finalmente, arrastrando el atributo country name de la dimensión COUNTRY al apartado de consultas y añadiendo la métrica HICP, obtenemos el resultado de la primera consulta.



Lo mismo podemos hacer sencillamente desde EXCEL. Para ello clicamos el botón Analizar en Excel.

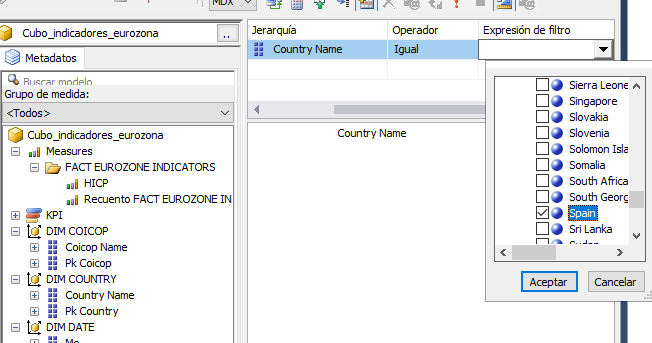


E insertamos los filtros y atributos deseados igual que en MDX. Podemos ordenar fácilmente las columnas con la built-in función de Excel. Esto nos deja a Turquía se sitúa en el nivel más alto de inflación (37.1) y al final de la tabla, se pueden observar países que experimentan deflación (Suiza y Noruega). España obtuvo una inflación de 2.7.

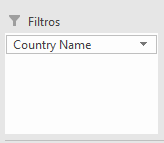
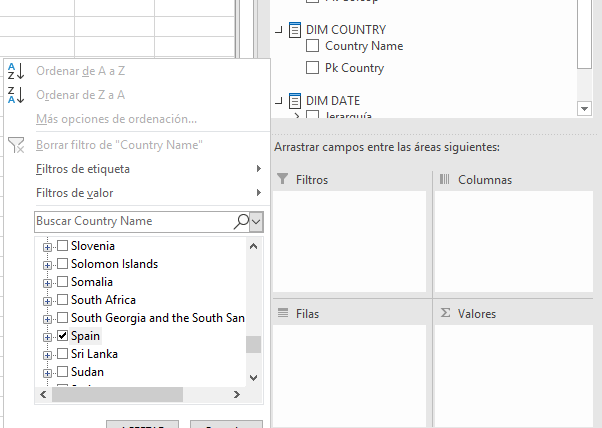


**2) Obtener la media del HICP para España por año. Mostrar los resultados redondeados a dos decimales.**

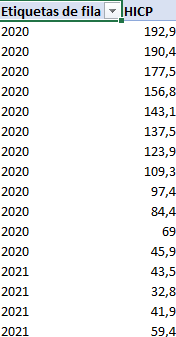
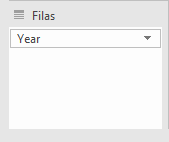
Para esta consulta primero filtramos el país España. Para ello seleccionamos la dimensión COUNTRY y la arrastramos al apartado superior de VS y seleccionamos la expresión de filtro *Spain*.

****

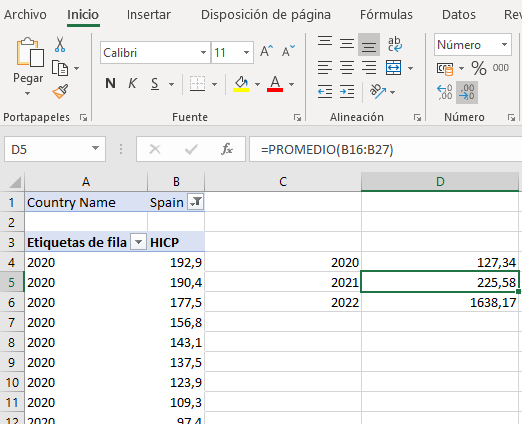
Realizamos el mismo proceso en Excel para utilizar funciones ya incluidas en Excel. Seleccionamos el atributo España para el país y lo arrastramos a los filtros.



Arrastramos el atributo Año para las filas.

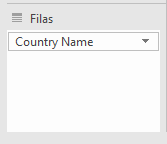
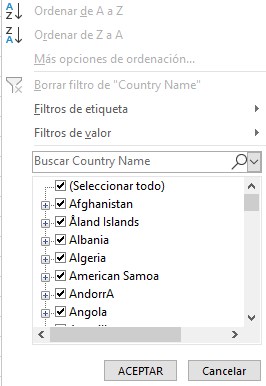


Finalmente, hemos de agrupar por año y utilizamos la función integrada en excel Promedio. Para redondear a 2 decimales podemos utilizar Inicio->Número->Quitar decimales (Marcado en la imagen)

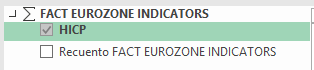


**3) Visualizar la evolución del HICP por año para España, Francia, Italia y Alemania. ¿En qué país ha habido un incremento más alto? Se recomienda utilizar Excel para analizar esta evolución.**

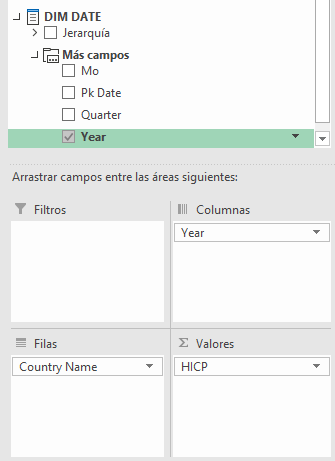
Desde Excel el primer paso será filtrar los países que realmente nos interesan. Para ello seleccionamos España, Francia, Italia y Alemania y arrastramos el campo country name al apartado Filas.

****

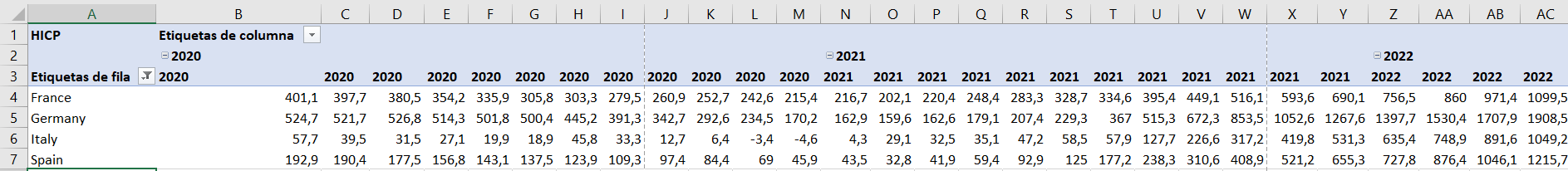
Además, seleccionamos la métrica que nos interesa. En este caso HICP.



Por último para la tabla dinámica seleccionamos el atributo Year como columnas.

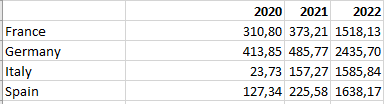


Ahora nos podemos dirigir a la tabla dinámica creada que nos debería de quedar en un formato así.



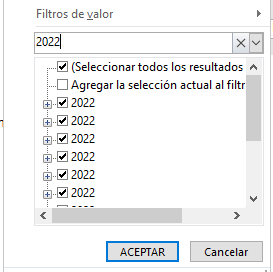
Similar al ejercicio anterior podemos hacer uso de la función PROMEDIO de Excel para escoger el promedio HICP por años y visualizarlos en una nueva celda.

****

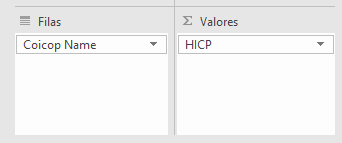


**4) Obtener el top cinco de fines de consumo (COICOP) durante 2022, según la media de su HICP, ordenados por este valor de mayor a menor.**

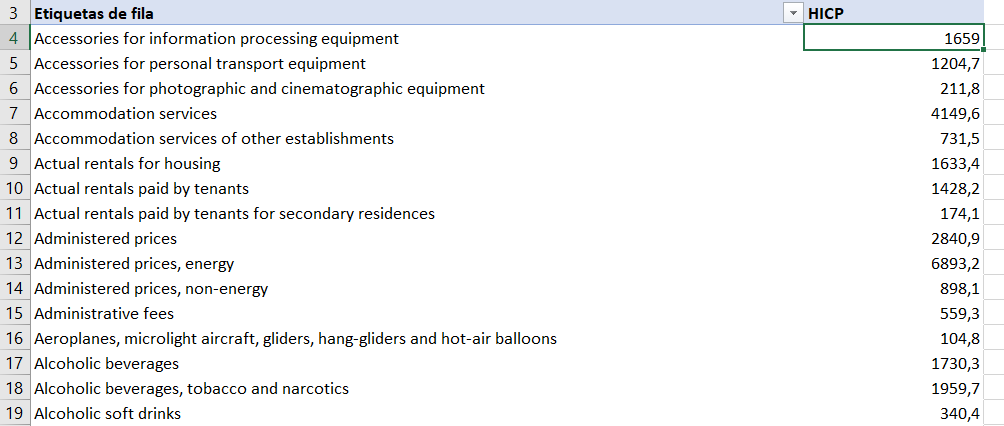
Para comenzar filtraremos por año (2022). Para ellos seleccionamos el atributo year y escribimos 2022. Añadimos el campo al apartado filtros.

****

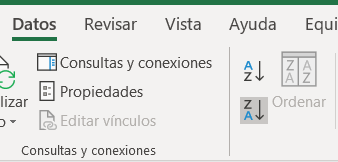
La métrica que usamos es HICP y las filas serán los fines de consumo.



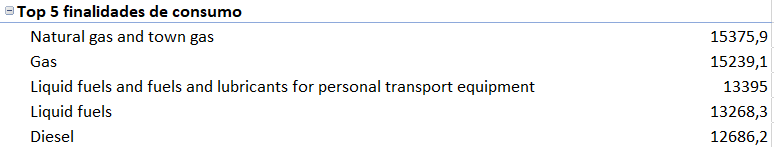
Con ello deberíamos de obtener una tabla tal que así:



En el apartado datos podemos ordenar la tabla dinámica de mayor a menor. Para ello seleccionamos la fila HICP y el botón Z-A que vemos en la imagen.

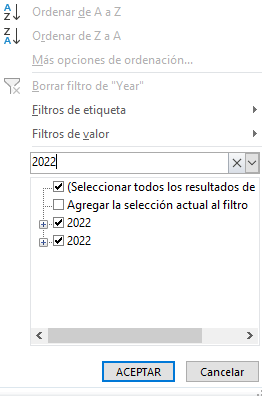
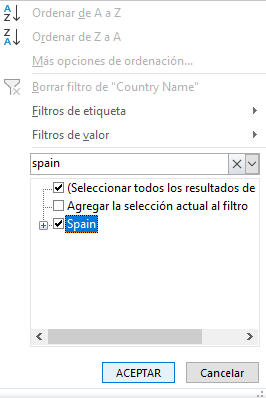


Podemos eliminar las filas posteriores a la 5 o agrupar por selección y añadir un título como se muestra a continuación.

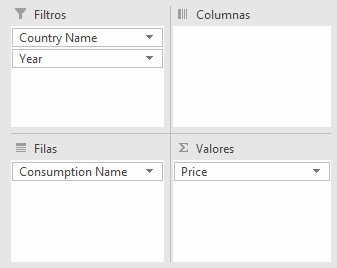


**5) Obtener el precio medio del consumo de gas y electricidad por banda de consumo en España durante el año 2022. ¿Varía el precio por banda de consumo? Si es así, ¿qué interpretación se le puede dar a esta variación?**

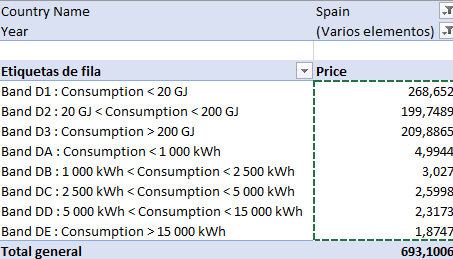
Desde Excel seleccionamos el atributo España del campo country name y lo arrastramos a Filtros. Lo mismo para Year y el año 2022.



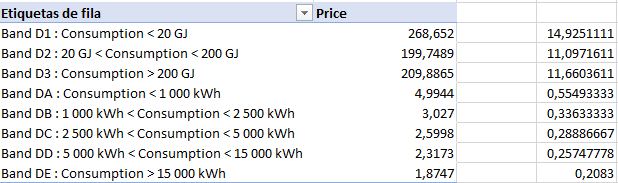
Para las filas seleccionamos Consumption Name y de métrica el valor Price. Obteniendo una tabla dinámica de este formato.



Con ello obtenemos la siguiente tabla dinámica:



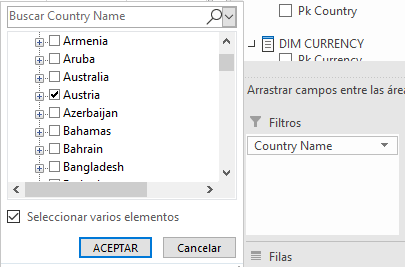
Pero esto no es el precio medio sino la suma de todos. Para conseguir el precio medio dividimos cada valor por la cantidad de elementos. Para valores de energía hay que dividir entre 18 (3 tipos de moneda \* 3 tipos de impuesto \* 2 ocurrencias cada año). Para valores de gas entre 36 (3 tipos de moneda \* 3 tipos de impuesto \* 2 ocurrencias cada año \* 2 unidades para captura de datos).



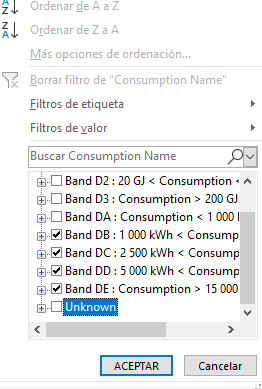
Podemos concluir que sí varía el precio por banda de consumo. Los primeros 3 elementos (D1, D2 y D3) hacen referencia a consumo de gas (se puede observar a partir del .tsv o desde las tablas), además que está tomado en otra medida GJ y la electricidad en kWh.

**6) Visualizar la evolución por año de precios de electricidad para España y Austria. ¿En qué país ha habido un incremento más alto? Se recomienda utilizar Excel para analizar esta evolución.**

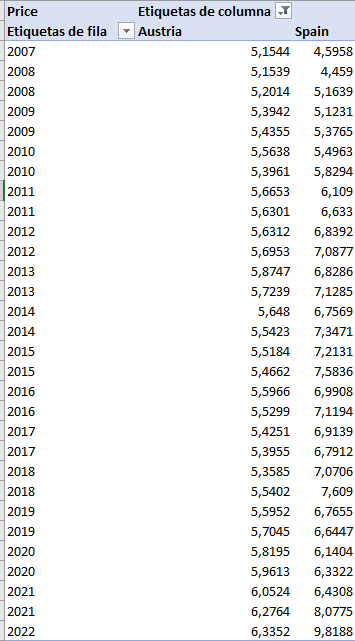
Empezamos seleccionando los países España y Austria y añadiendoles en el apartado filtros.



Para filtrar solo los valores de electricidad (y no obtener los del gas), podemos seleccionar en consumption name los valores de DB, DC, DD y DD (los que tienen kWh como medida), ya que son aquellos utilizados por la electricidad. Aunque en el .tsv de energía de gas también encontramos valores asociados a kWh estos no están asociados con los consumption name escogidos.



Además, añadimos el valor Year y la métrica Price y obtenemos la siguiente tabla dinámica:



Dado que cada valor (año) sale duplicado se podría hacer uso de la función integrada en Excel PROMEDIO, pero en este caso no lo veo necesario ya que podemos observar por año y por semestre.

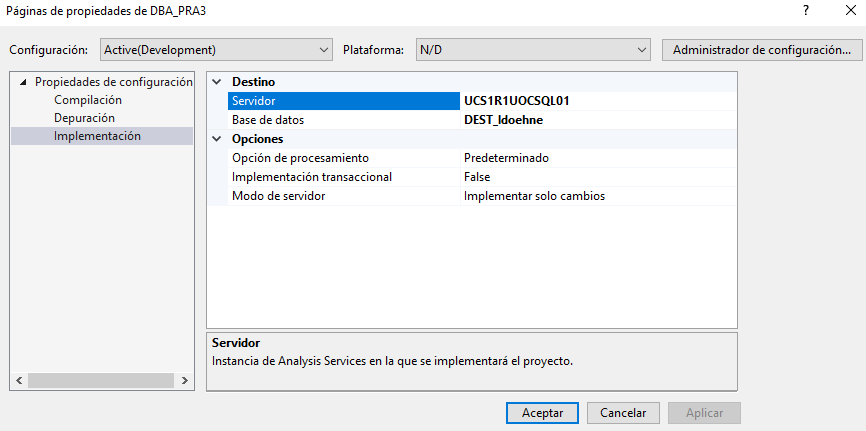
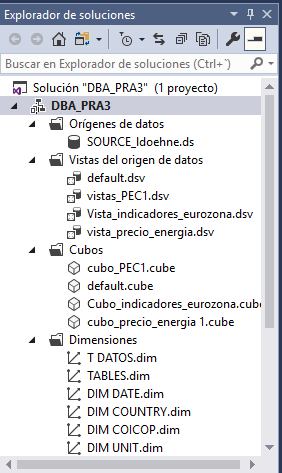
Finalmente, nos queda responder a la pregunta: ¿En qué país ha habido un incremento más alto?

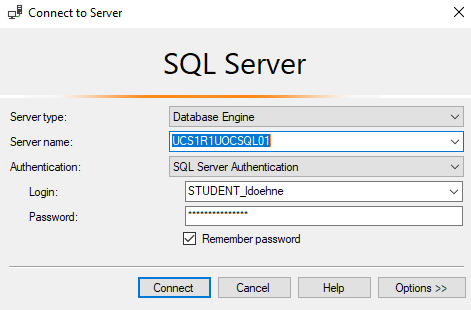
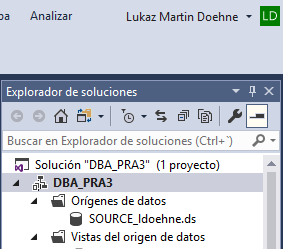
Fácilmente se puede observar que es en España. Que empieza con un valor menor al de Austria y termina superando considerablemente.

—

Además, de manera explícita, se pide incluir en el documento algunas capturas de pantalla (añado las que no tenía incluidas hasta ahora):

● Test de la conexión, donde se vea el servidor, usuario y base de datos.







● Fechas de inicio/fin del proceso.





● En Visual Studio, captura de los objetos creados tras la implementación del proyecto: DataSources, DataSourceViews, Cubes y Dimensions.

