

**Máster universitario de Ciencia de Datos**

**Práctica 3**

**Diseño y uso de bases de datos analíticas – Explotación de datos.**

Autor:

Mario Ubierna San Mamés

|  |
| --- |
| Índice de Contenido |

[Índice de Contenido 3](#_Toc73218588)

[Índice de tablas 5](#_Toc73218589)

[Índice de ilustraciones 6](#_Toc73218590)

[1. Introducción 10](#_Toc73218591)

[1.1. Presentación 10](#_Toc73218592)

[1.2. Descripción 10](#_Toc73218593)

[2. Creación del modelo OLAP 12](#_Toc73218594)

[2.1. Creación de la estructura física del modelo 12](#_Toc73218595)

[2.2. Creación del proyecto 12](#_Toc73218596)

[2.3. Vistas del origen de datos 18](#_Toc73218597)

[2.3.1. Vista Llamadas112 19](#_Toc73218598)

[2.3.2. Vista Mediciones 21](#_Toc73218599)

[2.4. Creación de cubos 24](#_Toc73218600)

[2.4.1. Cubo Llamadas112 25](#_Toc73218601)

[2.4.2. Cubo Mediciones 28](#_Toc73218602)

[2.5. Jerarquías y dimensiones 35](#_Toc73218603)

[2.5.1. DIM\_Ambito\_Geografico 35](#_Toc73218604)

[2.5.2. DIM\_Fecha 39](#_Toc73218605)

[2.5.3. DIM\_Grupo\_Edad 42](#_Toc73218606)

[2.5.4. DIM\_Medicion 42](#_Toc73218607)

[2.5.5. DIM\_Tipologia 43](#_Toc73218608)

[3. Implementación de la solución 44](#_Toc73218609)

[4. Explotación de la solución 51](#_Toc73218610)

[4.1. Análisis de las provincias con mayor porcentaje de movilidad según datos móviles 51](#_Toc73218611)

[4.2. Análisis del porcentaje de la población que evitaba aglomeraciones según la comunidad autónoma 56](#_Toc73218612)

[4.3. Análisis del promedio de sanciones por habitantes 59](#_Toc73218613)

[4.4. Evolución de las llamadas de urgencia al 112 en Cataluña por tipología de llamada 59](#_Toc73218614)

[4.5. Análisis de las llamadas de urgencia frente al porcentaje de la población española que evitaba las aglomeraciones entre los meses comprendidos entre marzo y junio de 2020 en Cataluña, desglosado por provincia 63](#_Toc73218615)

[4.6. Determinación del día de la semana con menor número de denuncias 65](#_Toc73218616)

[4.7. Análisis de las diez fechas (*top ten*) con mayor número de llamadas de urgencia al 112 con tipología de tránsito registrada, tanto en época de *COVID* como antes 66](#_Toc73218617)

|  |
| --- |
| Índice de tablas |

|  |
| --- |
| Índice de ilustraciones |

[Ilustración 1 - Base de datos DB\_mariousm. 12](#_Toc73218501)

[Ilustración 2 - Creación del proyecto multidimensional. 13](#_Toc73218502)

[Ilustración 3 - Configuración origen de los datos. 14](#_Toc73218503)

[Ilustración 4 - Propiedades del proyecto Visual Studio. 15](#_Toc73218504)

[Ilustración 5 - Configuración destino de datos en Visual Studio. 16](#_Toc73218505)

[Ilustración 6 - Ventana origen de datos. 17](#_Toc73218506)

[Ilustración 7 - Modificación conexión de datos. 18](#_Toc73218507)

[Ilustración 8 - Creación de nueva vista del origen de datos. 19](#_Toc73218508)

[Ilustración 9 - Selección origen de datos vista. 19](#_Toc73218509)

[Ilustración 10 - Selección tablas para la vista Llamadas112. 20](#_Toc73218510)

[Ilustración 11 - Diseño vista vLlamadas112. 21](#_Toc73218511)

[Ilustración 12 - Selección tablas para la vista Mediciones. 22](#_Toc73218512)

[Ilustración 13 - Nombre de la vista vMediciones. 23](#_Toc73218513)

[Ilustración 14 - Diseño vista vMediciones. 24](#_Toc73218514)

[Ilustración 15 - Creación de un cubo. 25](#_Toc73218515)

[Ilustración 16 - Uso de tablas existentes en el cubo. 25](#_Toc73218516)

[Ilustración 17 - Selección vista del origen de datos cLlamadas112. 26](#_Toc73218517)

[Ilustración 18 - Selección métricas de cLlamadas112. 26](#_Toc73218518)

[Ilustración 19 - Selección dimensiones de cLlamadas112. 27](#_Toc73218519)

[Ilustración 20 - Resumen de cLlamadas112. 27](#_Toc73218520)

[Ilustración 21 - Ventana diseño cLlamadas112. 28](#_Toc73218521)

[Ilustración 22 - Selección vista del origen de datos cMediciones. 29](#_Toc73218522)

[Ilustración 23 - Selección métricas de cMediciones. 29](#_Toc73218523)

[Ilustración 24 - Selección dimensiones de cMediciones. 30](#_Toc73218524)

[Ilustración 25 - Resumen de cMediciones. 30](#_Toc73218525)

[Ilustración 26 - Ventana diseño cMediciones. 31](#_Toc73218526)

[Ilustración 27 - Eliminación de DIM\_Ambito\_Geografico\_1 31](#_Toc73218527)

[Ilustración 28 - Eliminación de DIM\_Fecha\_1. 32](#_Toc73218528)

[Ilustración 29 - Agregar dimensiones a cMediciones. 32](#_Toc73218529)

[Ilustración 30 - Selección de dimensiones de cMediciones. 33](#_Toc73218530)

[Ilustración 31 - Ventana diseño cMediciones. 33](#_Toc73218531)

[Ilustración 32 - Relación de DIM\_Ambito\_Geografico. 34](#_Toc73218532)

[Ilustración 33 - Relación de DIM\_Fecha. 34](#_Toc73218533)

[Ilustración 34 - Relaciones dimensiones cMediciones. 35](#_Toc73218534)

[Ilustración 35 - DIM\_Ambito\_Geografico. 35](#_Toc73218535)

[Ilustración 36 - Jerarquía DIM\_Ambito\_Geografico. 36](#_Toc73218536)

[Ilustración 37 - KeyColumns provincia\_nombre. 36](#_Toc73218537)

[Ilustración 38 NameColumn provincia\_nombre. 37](#_Toc73218538)

[Ilustración 39 - KeyColumns comarca. 37](#_Toc73218539)

[Ilustración 40 - NameColumn comarca. 38](#_Toc73218540)

[Ilustración 41 - KeyColumns municipio. 38](#_Toc73218541)

[Ilustración 42 - NameColumn municipio. 39](#_Toc73218542)

[Ilustración 43 - DIM\_Fecha. 39](#_Toc73218543)

[Ilustración 44 - Jerarquía DIM\_Fecha. 40](#_Toc73218544)

[Ilustración 45 - KeyColumns mes. 40](#_Toc73218545)

[Ilustración 46 - NameColumn mes. 41](#_Toc73218546)

[Ilustración 47 - KeyColumns dia. 41](#_Toc73218547)

[Ilustración 48 - NameColumn dia. 42](#_Toc73218548)

[Ilustración 49 - DIM\_Grupo\_Edad. 42](#_Toc73218549)

[Ilustración 50 - DIM\_Medicion. 43](#_Toc73218550)

[Ilustración 51 - DIM\_Tipologia. 43](#_Toc73218551)

[Ilustración 52 - Implementar la solución. 44](#_Toc73218552)

[Ilustración 53 - Implementación de la solución. 45](#_Toc73218553)

[Ilustración 54 - Creación de la dimensión temporal. 46](#_Toc73218554)

[Ilustración 55 - Asignación del tipo para el día. 47](#_Toc73218555)

[Ilustración 56 - Asignación del tipo para el mes. 47](#_Toc73218556)

[Ilustración 57 - Asignación del tipo para el año. 48](#_Toc73218557)

[Ilustración 58 - Asignación del tipo para la fecha. 48](#_Toc73218558)

[Ilustración 59 - Relación de atributos DIM Ambito Geografico. 49](#_Toc73218559)

[Ilustración 60 - Relación de atributos DIM Fecha. 49](#_Toc73218560)

[Ilustración 61 - Establcer la relación de flexible a rígida. 49](#_Toc73218561)

[Ilustración 62 - Relaciones rígidas en DIM Ambito Geografico. 50](#_Toc73218562)

[Ilustración 63 - Relaciones rígidas en DIM Fecha. 50](#_Toc73218563)

[Ilustración 64 - Análisis de las provincias con mayor porcentaje de movilidad. 51](#_Toc73218564)

[Ilustración 65 - Creación de la media. 52](#_Toc73218565)

[Ilustración 66 - Creación de la media. 52](#_Toc73218566)

[Ilustración 67 - Análisis de las provincias con mayor porcentaje de movilidad. 53](#_Toc73218567)

[Ilustración 68 - Análisis de las provincias con mayor porcentaje de movilidad. 53](#_Toc73218568)

[Ilustración 69 - Consulta MDX para el análisis de las provincias con mayor porcentaje de movilidad. 54](#_Toc73218569)

[Ilustración 70 - Análisis de las provincias con mayor porcentaje de movilidad. 54](#_Toc73218570)

[Ilustración 71 - Análisis de las provincias con mayor porcentaje de movilidad. 55](#_Toc73218571)

[Ilustración 72 - Análisis del porcentaje de población que evitaba aglomeraciones según la CCAA. 56](#_Toc73218572)

[Ilustración 73 - Creación número de filas. 56](#_Toc73218573)

[Ilustración 74 - Creación número de filas. 57](#_Toc73218574)

[Ilustración 75- Análisis del porcentaje de población que evitaba aglomeraciones según la CCAA. 57](#_Toc73218575)

[Ilustración 76 - Creación media población aglomeración. 58](#_Toc73218576)

[Ilustración 77 - Análisis del porcentaje de población que evitaba aglomeraciones según la CCAA. 59](#_Toc73218577)

[Ilustración 78 - Evolución de las llamadas según la tipología. 60](#_Toc73218578)

[Ilustración 79 - Evolución de las llamadas según la tipología. 61](#_Toc73218579)

[Ilustración 80 - Evolución de las llamadas según la tipología. 62](#_Toc73218580)

[Ilustración 81 - Evolución de las llamadas según la tipología. 63](#_Toc73218581)

[Ilustración 82 – Análisis de las llamadas en Cataluña en época de COVID. 64](#_Toc73218582)

[Ilustración 83 - Análisis del porcentaje de población que evitaba aglomeraciones en época de COVID. 65](#_Toc73218583)

[Ilustración 84 - Consulta número de llamadas de tipo tránsito en época COVID. 66](#_Toc73218584)

[Ilustración 85 - Análisis número de llamadas de tipo tránsito en época de COVID. 66](#_Toc73218585)

[Ilustración 86 - Consulta número de llamadas de tipo tránsito no época de COVID. 67](#_Toc73218586)

[Ilustración 87 - Análisis número de llamadas de tipo tránsito no época de COVID. 67](#_Toc73218587)

|  |
| --- |
| Introducción |

## Presentación

A partir de la solución oficial de la segunda práctica (PRA2) el estudiante debe implementar los cubos multidimensionales necesarios para la explotación de la información y el posterior análisis de datos. De este modo se facilitará la toma de decisiones de los usuarios potenciales.

Así pues, esta actividad tiene el objetivo de implementar un modelo multidimensional online analytical processing (OLAP) para el análisis multidimensional de la información con el fin de responder a las preguntas definidas en el análisis de requerimientos.

Adicionalmente, se facilitará junto a este enunciado el fichero «*export\_DW\_COVID.sql*», que contiene los scripts de generación y carga de todas las tablas planteadas en la solución, para que el estudiantado pueda partir de la misma base.

## Descripción

Más concretamente esta tercera parte del caso práctico consiste en diseñar un modelo OLAP para el análisis multidimensional de la información disponible en el almacén de datos que permita dar respuesta a las siguientes cuestiones:

* Análisis de las provincias con mayor porcentaje de movilidad según datos móviles.
* Análisis del porcentaje de la población que evitaba las aglomeraciones según la comunidad autónoma.
* Análisis del promedio de sanciones por habitante.
* Evolución de las llamadas de urgencia al 112 en Cataluña por tipología de llamada.
* Análisis de las llamadas de urgencia frente al porcentaje de la población que evitaba las aglomeraciones entre los meses comprendidos entre marzo y junio de 2020 en Cataluña, desglosado por provincia.
* Determinación del día de la semana con menor número de denuncias.
* Análisis de las diez fechas (*top ten*) con mayor número de llamadas de urgencia al 112 con tipología de tránsito registrada, tanto en época de *COVID* como antes.

|  |
| --- |
| Creación del modelo OLAP |

En este apartado vamos a crear tanto el proyecto en *Visual Studio*, como observar el origen de datos, la creación de los cubos y la decisión sobre jerarquías/dimensiones/atributos relacionados.

## Creación de la estructura física del modelo

Lo primero que debemos de hacer es ejecutar el *script* que se nos ha proporcionado “*export\_DW\_COVID.sql*”, ya que haciendo esto obtenemos el diseño tanto de los hechos como de las dimensiones junto a sus datos.

Cabe destacar que para ejecutar el *script* lo hemos tenido que modificar añadiendo las dos primeras líneas siguientes, para que así haga uso de la base de datos correcta:



Ilustración - Base de datos DB\_mariousm.

## Creación del proyecto

Una vez que tenemos todos los datos que vamos a hacer uso de ellos para explotarlos, podemos crear el proyecto en *Visual Studio*.

Para ello abrimos el programa y creamos un nuevo proyecto de tipo “*Analysis Services Multidimensional*”:

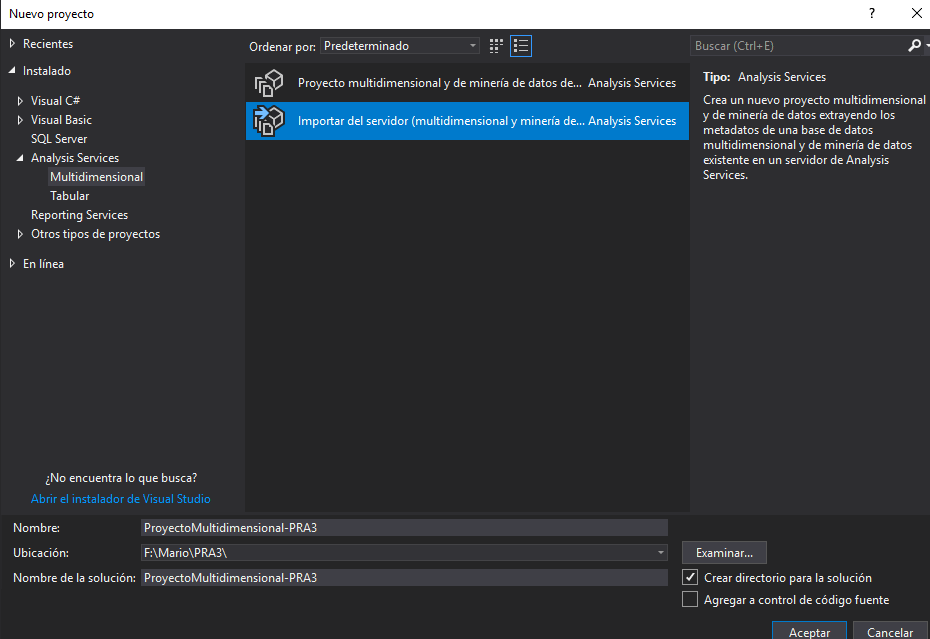


Ilustración - Creación del proyecto multidimensional.

Una vez que hemos seleccionado la opción “Importar del servidor” y establecemos el nombre del proyecto junto con su ubicación, nos aparecerá un asistente para configurar la base de datos a usar.

Para ello indicamos el servidor y en el menú desplegable de la base de datos, seleccionamos la nuestra, es decir, “DB\_mariousm”:

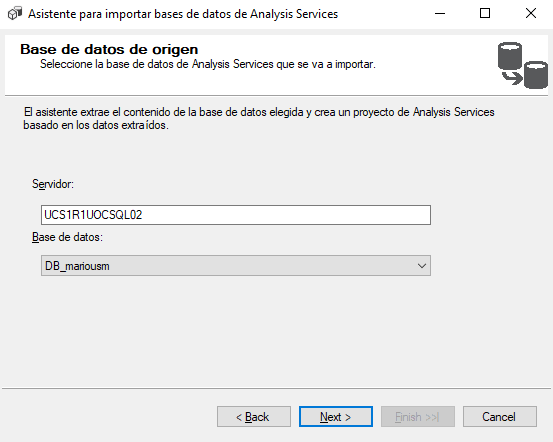


Ilustración - Configuración origen de los datos.

Una vez importada la estructura de la base de datos a nuestro proyecto, hay que configurar el destino de los datos. Para ello, nos dirigimos a las propiedades del proyecto:

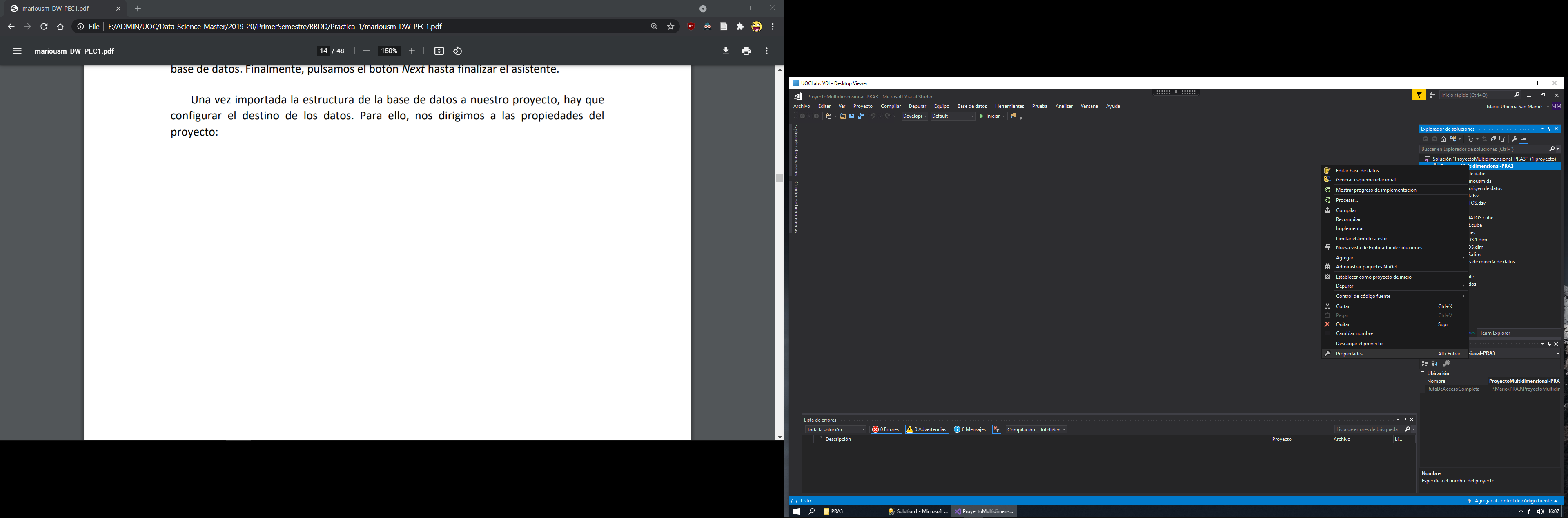


Ilustración - Propiedades del proyecto Visual Studio.

En las propiedades nos vamos al apartado implementación, y en él cambiamos el nombre del servidor tal y como se muestra en la siguiente captura:

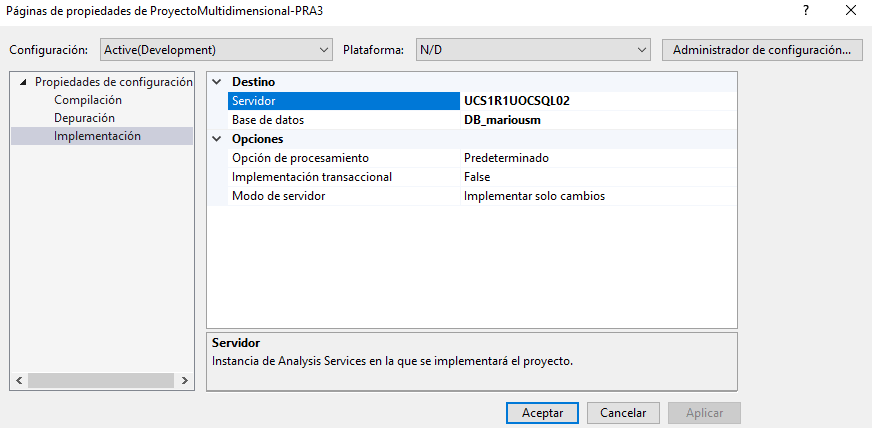


Ilustración - Configuración destino de datos en Visual Studio.

Finalmente hay que configurar el origen de los datos, para ello sobre el nombre de la base de datos en el explorador de soluciones, hacemos doble click y nos mostrará la siguiente ventana:

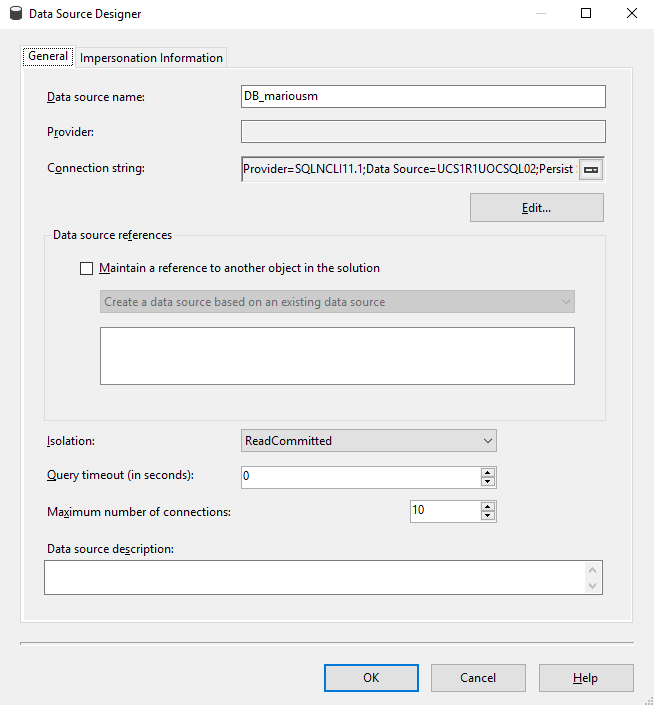


Ilustración - Ventana origen de datos.

Pulsamos sobre el botón “*Edit*”para cambiar la conexión a la base de datos, una vez hemos pulsado veremos la siguiente pantalla:

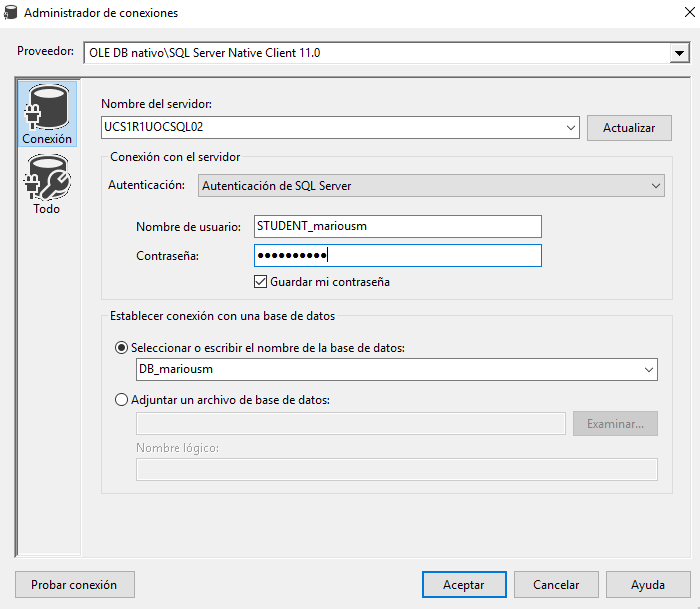


Ilustración - Modificación conexión de datos.

Introducimos el servidor que se nos ha proporcionado, el nombre del usuario y la contraseña que usamos para acceder a *Microsoft SQL Server Management Studio 2017*.

Una vez realizados todos estos pasos podemos dar por concluido este apartado.

## Vistas del origen de datos

Una vez que ya tenemos configurado nuestro proyecto, podemos comenzar a definir las diferentes vistas respecto al origen de los datos.

En nuestro caso vamos a tener dos:

* Vista de Llamadas112: nos va a permitir realizar el acceso a la tabla correspondiente de las llamadas y así hacer un análisis de las mismas.
* Vista de Mediciones: en ella vamos a poder realizar el acceso a la tabla de métricas y hacer un profundo análisis de las mismas.

Para crear las dos vistas debemos acceder al explorador de soluciones del proyecto y hacer click derecho sobre la carpeta “Vistas del origen de datos”, posteriormente pulsaremos sobre la opción “Nueva vista del origen de datos”:

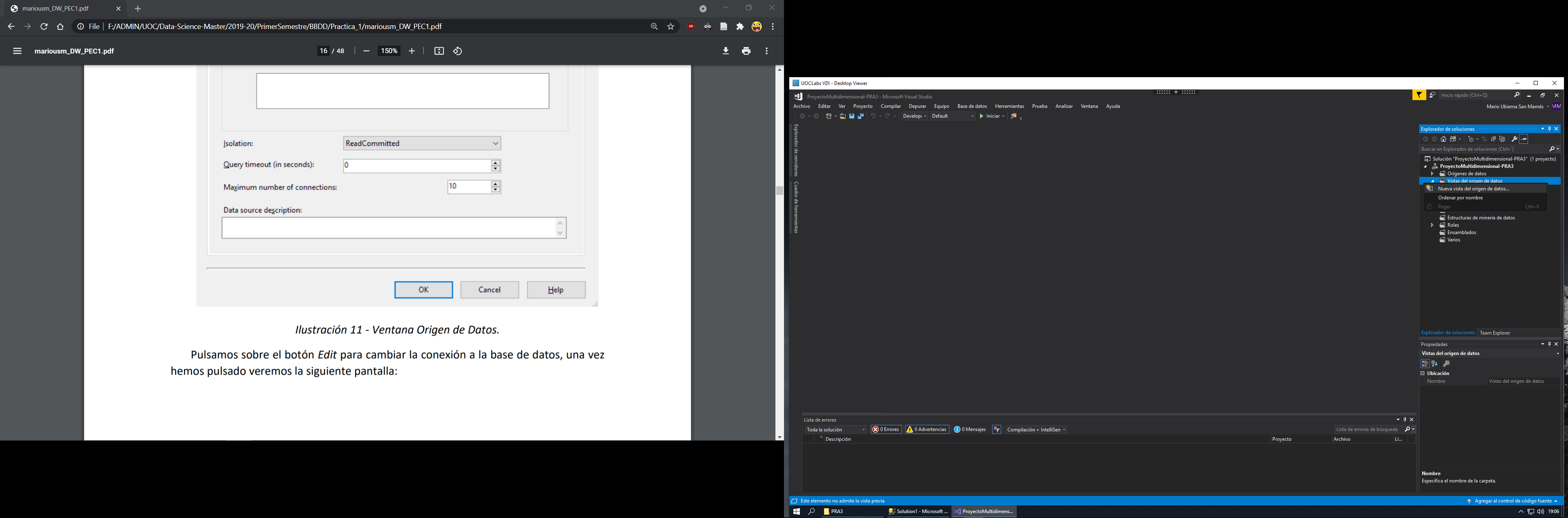


Ilustración - Creación de nueva vista del origen de datos.

Una vez que hemos accedido al asistente para la creación de la vista, nos aparecerá la siguiente ventana en la que tenemos que seleccionar el origen de datos:

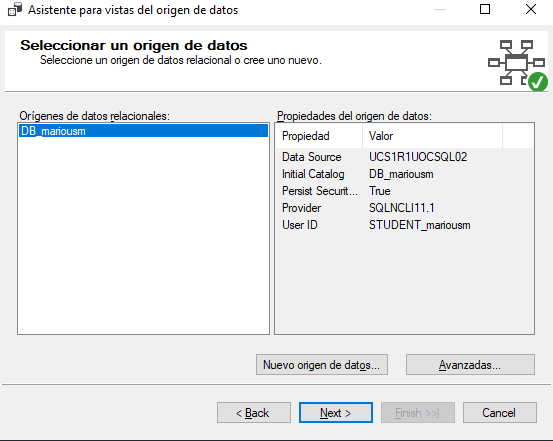


Ilustración - Selección origen de datos vista.

### Vista Llamadas112

Posteriormente seleccionamos tanto la tabla del hecho como de las dimensiones correspondientes, en nuestro caso el hecho es “FACT\_Llamadas112” y las dimensiones: “DIM\_Ambito\_Geografico”, “DIM\_Fecha”, “DIM\_Tipologia”:

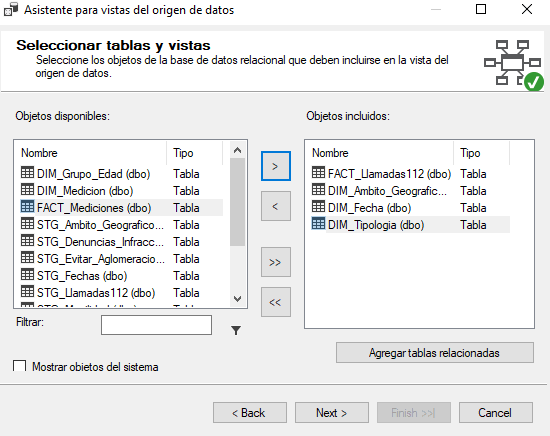


Ilustración - Selección tablas para la vista Llamadas112.

Pulsamos el botón “*Next”*, indicamos el nombre de la vista “vLlamadas112” y finalizamos el asistente.

Al hacer doble click sobre la vista nos muestra el diseño de la misma:

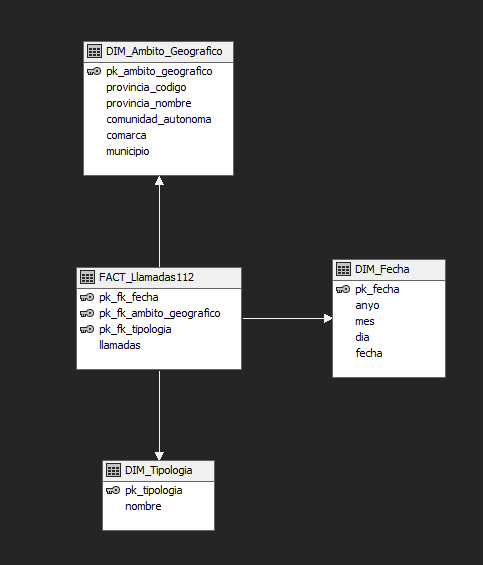


Ilustración - Diseño vista vLlamadas112.

### Vista Mediciones

A continuación, vamos a realizar el mismo proceso que el detallado en la creación de la vista anterior, pero en este caso vamos a crear una vista con el fin de explotar la información sobre las denuncias, el porcentaje de población que evitaba aglomeraciones…

Esta vista tiene en cuenta el hecho “FACT\_Mediciones” y las dimensiones: “DIM\_Ambito\_Geografico”, “DIM\_Fecha”, “DIM\_Grupo\_Edad” y “DIM\_Medicion”:

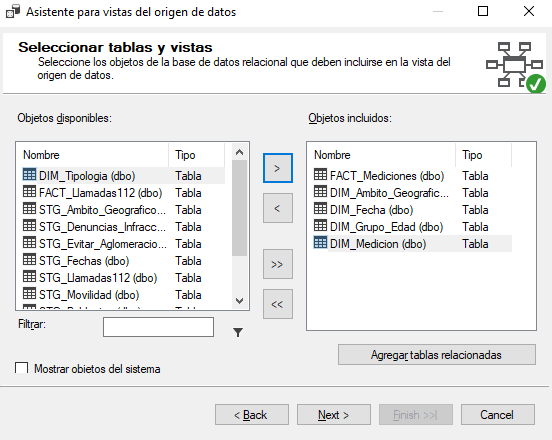


Ilustración - Selección tablas para la vista Mediciones.

Pulsamos el botón “*Next*”, indicamos el nombre de la vista “vMediciones” y finalizamos el asistente:

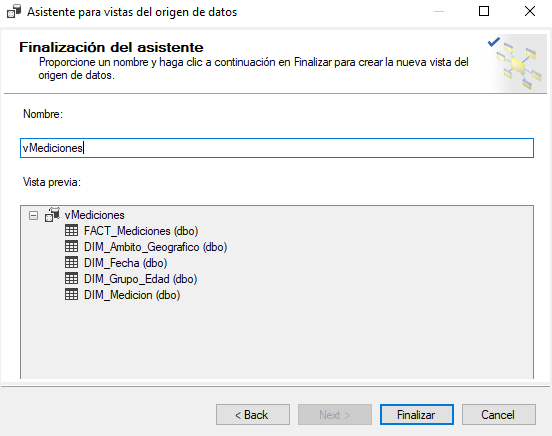


Ilustración - Nombre de la vista vMediciones.

Al hacer doble click sobre la vista nos muestra el diseño de la misma, comprobamos que todo está de forma correcta:

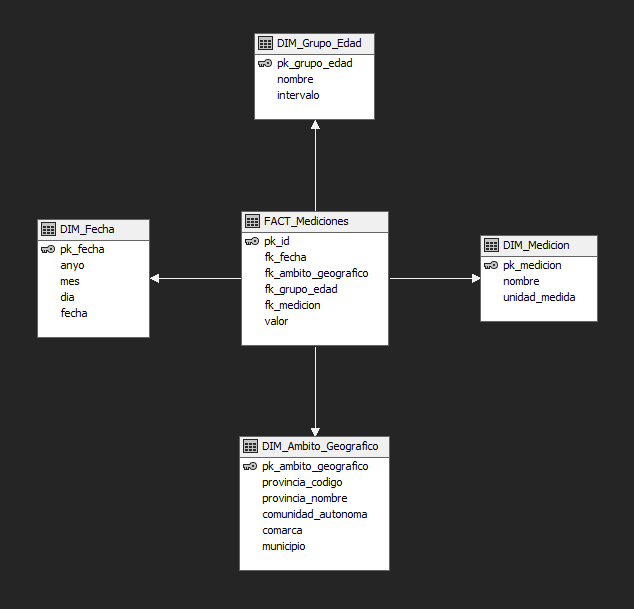


Ilustración - Diseño vista vMediciones.

## Creación de cubos

Una vez creadas las diferentes vistas para así poder realizar el análisis de los datos, tenemos que crear cada uno de los cubos correspondientes.

Para crear un cubo nos debemos dirigir al explorador de soluciones, y sobre la carpeta “Cubos” pulsamos botón derecho, posteriormente seleccionamos la opción “Nuevo cubo”:

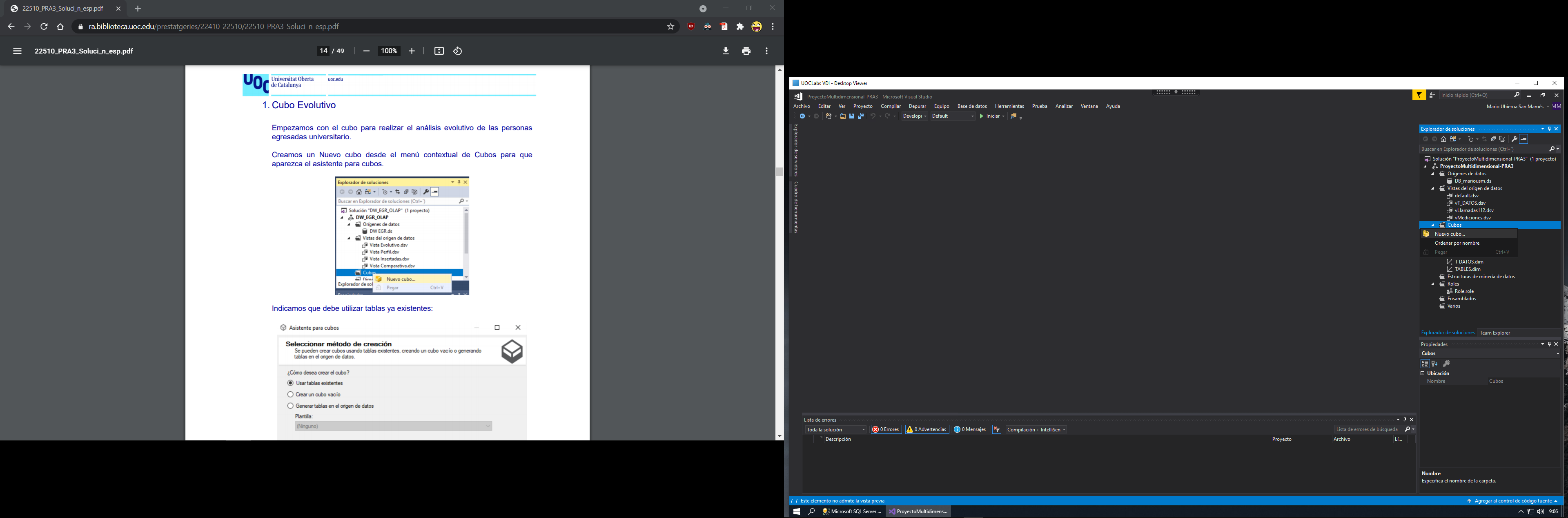


Ilustración - Creación de un cubo.

Una vez seleccionada la opción “Nuevo cubo” nos aparecerá el siguiente asistente, donde debemos indicar que vamos a hacer uso de tablas existentes:

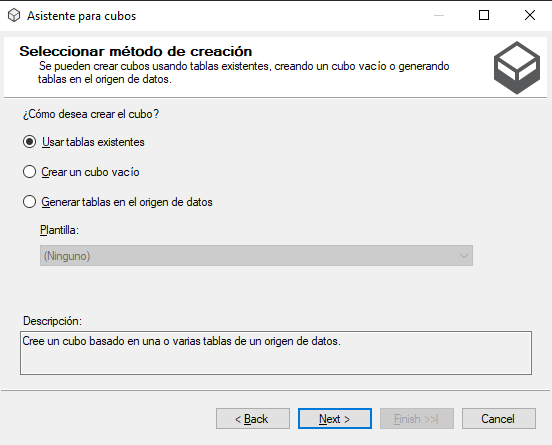


Ilustración - Uso de tablas existentes en el cubo.

### Cubo Llamadas112

Una vez que hemos realizado los pasos anteriores, indicamos que queremos la vista correspondiente a las llamadas al 112, es decir, “vLlamadas112”:

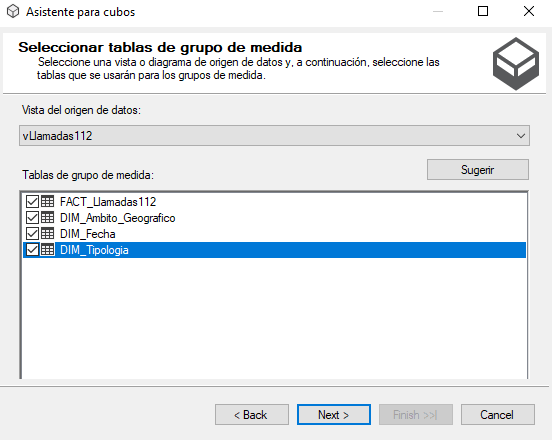


Ilustración - Selección vista del origen de datos cLlamadas112.

Posteriormente, seleccionamos las métricas:

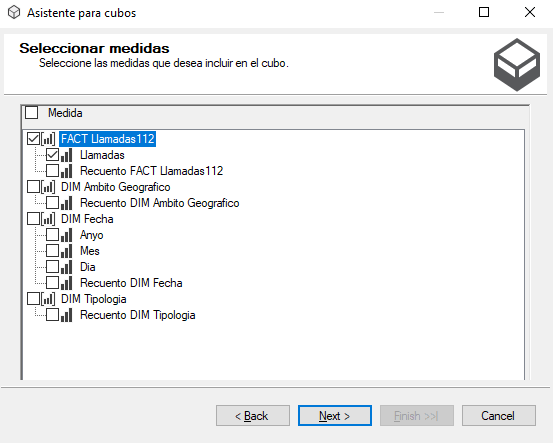


Ilustración - Selección métricas de cLlamadas112.

Por último indicamos las dimensiones del cubo:

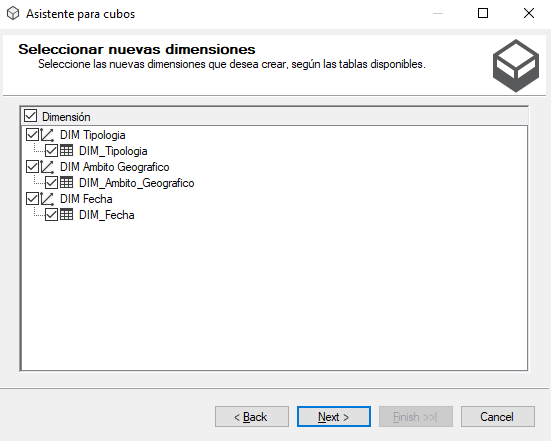


Ilustración - Selección dimensiones de cLlamadas112.

Al terminar el asistente obtenemos el siguiente resumen, también indicamos el nombre del cubo:

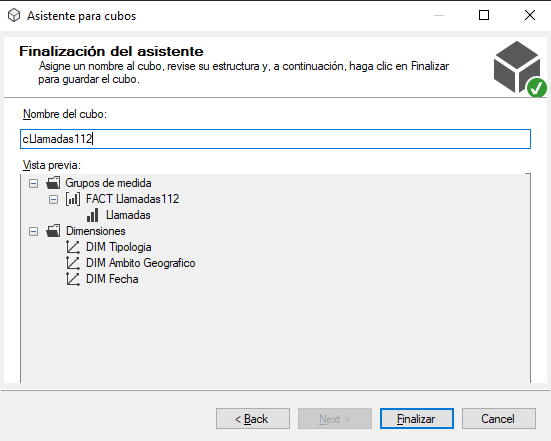


Ilustración - Resumen de cLlamadas112.

Finalmente, cuando hemos creado el cubo tenemos la siguiente ventana de diseño de “cLlamadas112”, a la izquierda observamos las métricas y dimensiones, y a la derecha la vista del origen de datos:

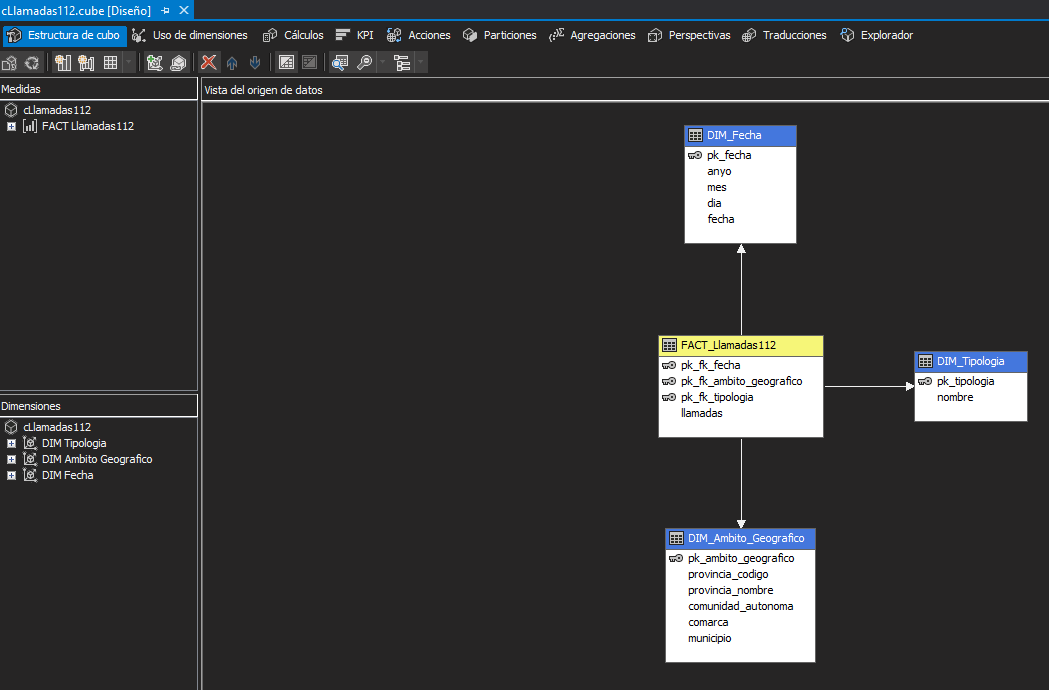


Ilustración - Ventana diseño cLlamadas112.

### Cubo Mediciones

Para crear el cubo de mediciones seguimos los mimos pasos que en el cubo anterior. Una vez que llegamos al asistente para crear el cubo, indicamos que la vista que queremos es la de “vMediciones”:

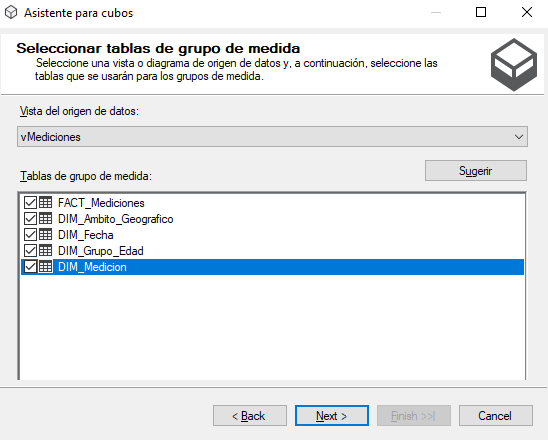


Ilustración - Selección vista del origen de datos cMediciones.

Posteriormente, seleccionamos las métricas:



Ilustración - Selección métricas de cMediciones.

Por último, indicamos las dimensiones del cubo:



Ilustración - Selección dimensiones de cMediciones.

Al terminar el asistente obtenemos el siguiente resumen, también indicamos el nombre del cubo:

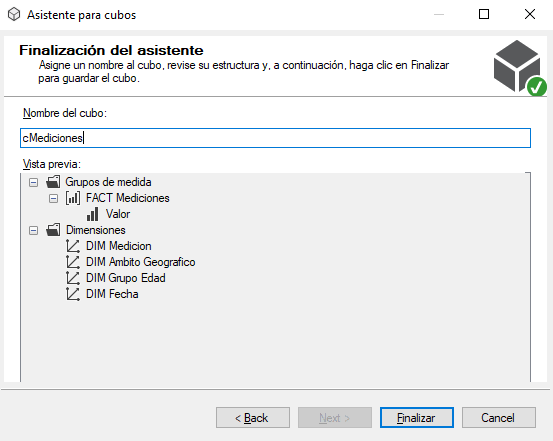


Ilustración - Resumen de cMediciones.

Finalmente, cuando creamos el cubo nos salta la ventana de diseño del mismo, a la izquierda observamos las métricas y las dimensiones, y a la derecha la vista del origen de datos:

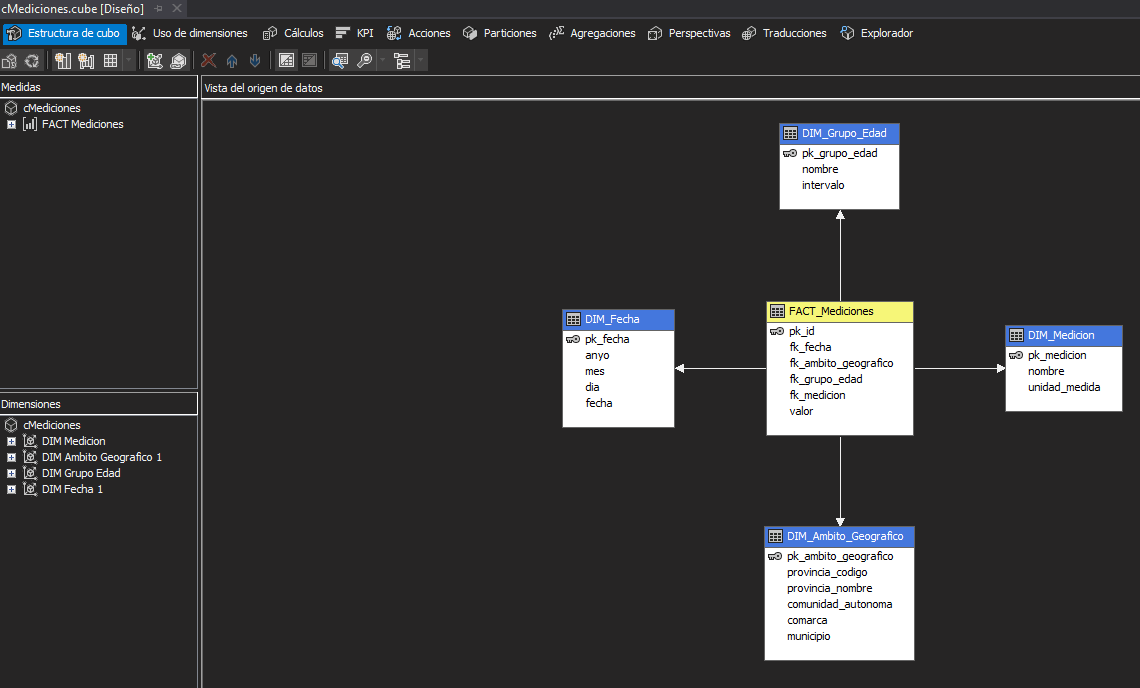


Ilustración - Ventana diseño cMediciones.

Tal y como podemos observar en la anterior ilustración, para las dimensiones comunes “DIM\_Ambito\_Geografico” y “DIM\_Fecha”, nos ha vuelto a crear la misma dimensión, es decir, tenemos duplicadas ambas dimensiones. Esto no es correcto, ya que tenemos que usar las mismas dimensiones en todos los cubos.

Por lo tanto, eliminamos dichas dimensiones que ha creado nuevas, para ello en el explorador de soluciones nos dirigimos a la carpeta “Dimensiones” y en ella eliminamos tanto “DIM\_Ambito\_Geografico\_1” como “DIM\_Fecha\_1”:

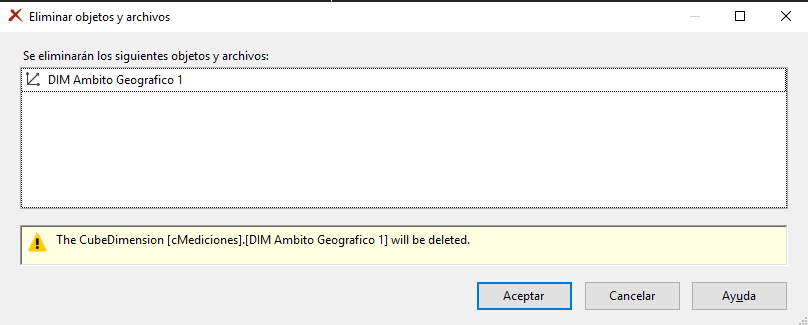


Ilustración - Eliminación de DIM\_Ambito\_Geografico\_1

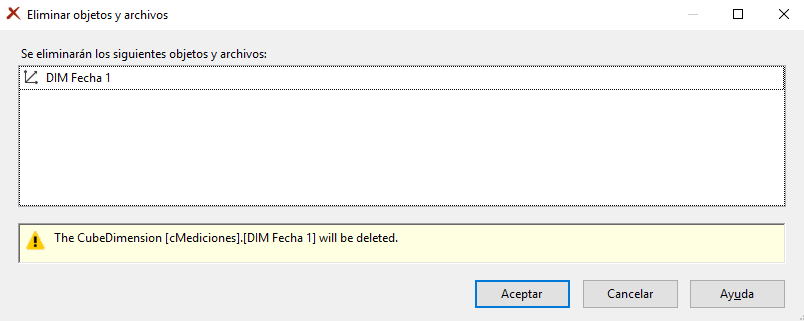


Ilustración - Eliminación de DIM\_Fecha\_1.

Al eliminar dichas dimensiones se eliminan también del cubo, por lo que tenemos que añadir de nuevo dichas dimensiones, pero las que ya habíamos usado en la creación del cubo anterior. Para ello, en el diseño del cubo seleccionamos “Agregar dimensión al cubo”:

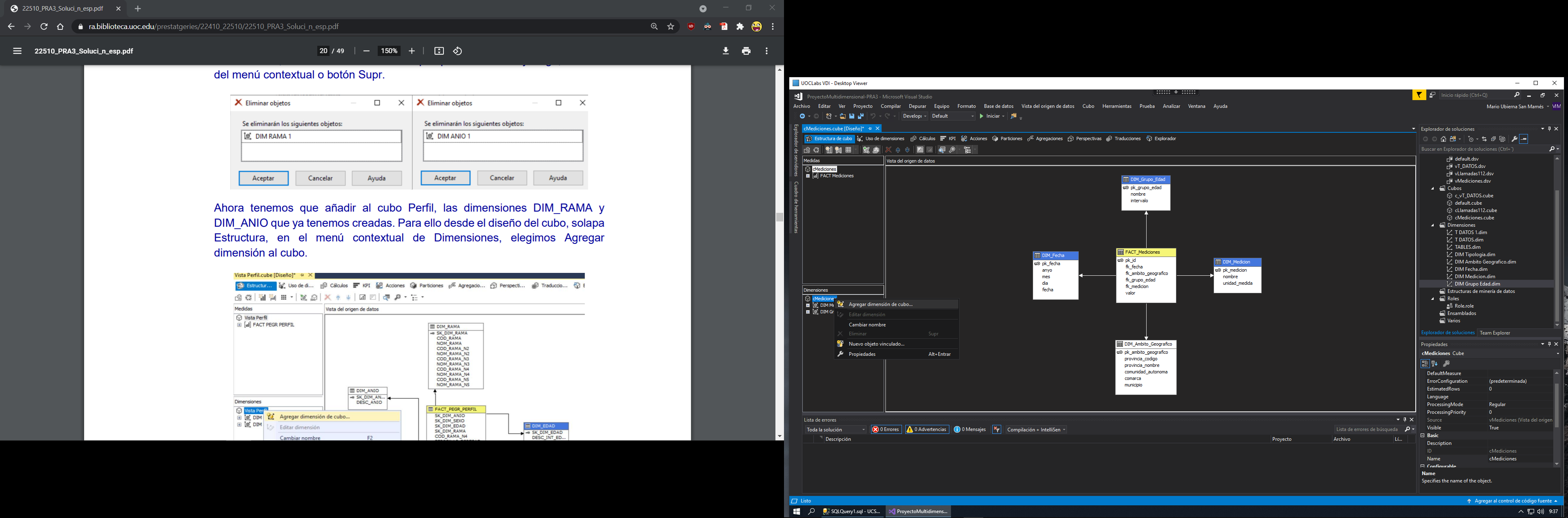


Ilustración - Agregar dimensiones a cMediciones.

Seleccionamos la dimensión “DIM\_Ambito\_Geografico” y “DIM\_Fecha”:

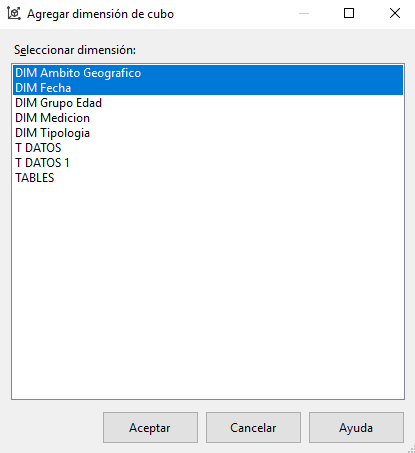


Ilustración - Selección de dimensiones de cMediciones.

Finalmente, el cubo “cMediciones” nos queda de la siguiente forma:

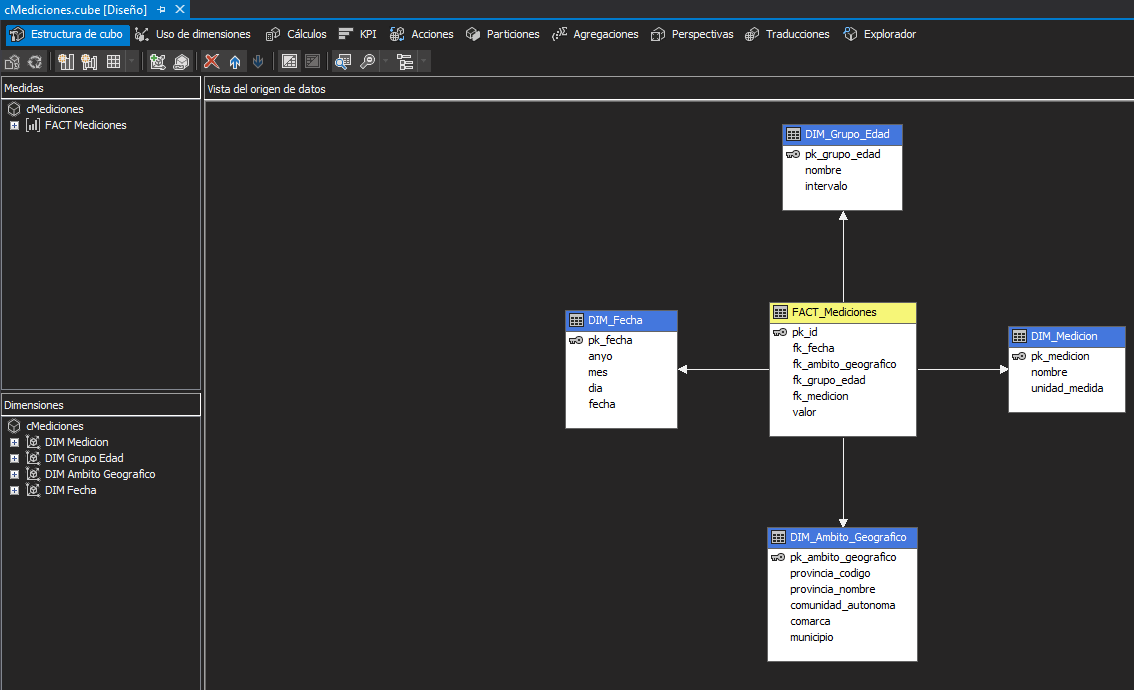


Ilustración - Ventana diseño cMediciones.

Al borrar las dimensiones y añadirlas de nuevo hemos perdido las relaciones de ambas dimensiones, por lo que nos dirigimos a “Uso de dimensiones” y establecemos la relación de ambas:

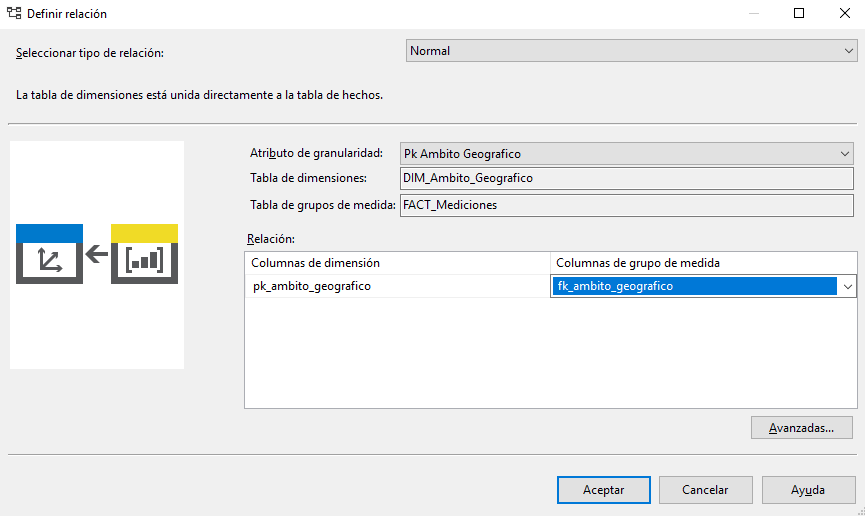


Ilustración - Relación de DIM\_Ambito\_Geografico.

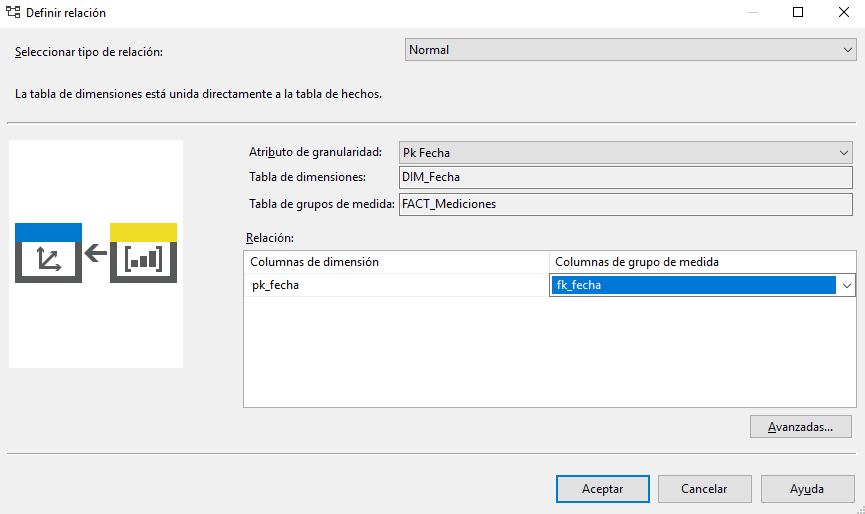


Ilustración - Relación de DIM\_Fecha.

Por lo tanto, las relaciones de todas las dimensiones nos quedarían de la siguiente forma:

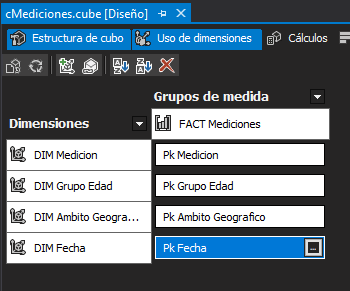


Ilustración - Relaciones dimensiones cMediciones.

## Jerarquías y dimensiones

En este apartado vamos a definir tanto las jerarquías y dimensiones que tenemos en nuestro modelo de datos.

### DIM\_Ambito\_Geografico

Para definir cada una de las dimensiones nos vamos a la carpeta “Dimensiones” dentro del explorador de soluciones, y hacemos doble click sobre la dimensión que vamos a definir, en nuestro caso “DIM\_Ambito\_Geografico”.

Para editar los atributos de la dimensión solamente tenemos que arrastrar desde la vista del origen de datos hacia la dimensión en sí, tal y como se observa en la siguiente ilustración:

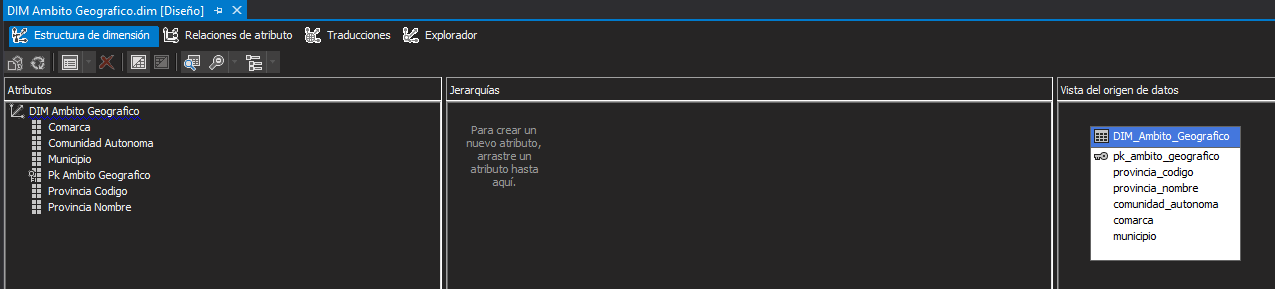


Ilustración - DIM\_Ambito\_Geografico.

Por otro lado, definimos una jerarquía para que nos resulte luego más fácil realizar las consultas respecto al ámbito geográfico. La jerarquía va a ser “Comunidad autónoma” > “Provincia nombre” > “Comarca” > “Municipio”:

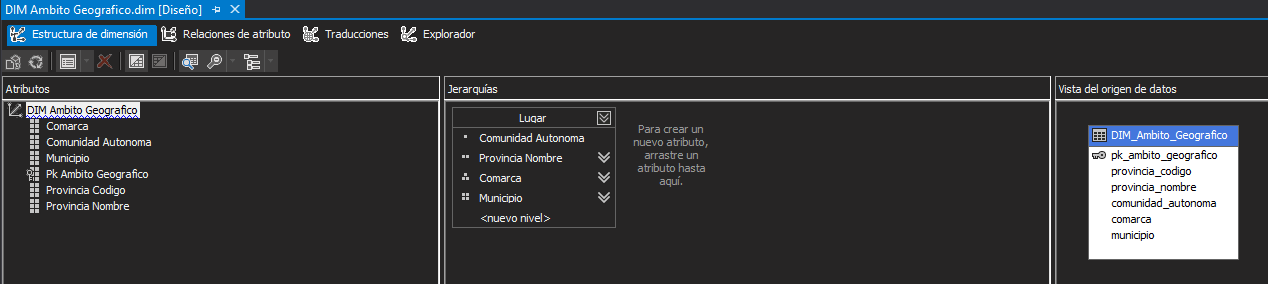


Ilustración - Jerarquía DIM\_Ambito\_Geografico.

Al definir la jerarquía tenemos que cambiar tanto el atributo “*KeyColumns”* como el atributo “*NameColumn*” de los campos pertenecientes a la jerarquía con la excepción de la jerarquía más alta, en nuestro caso el atributo “Comunidad Autonoma”.

Por lo tanto, vamos a definir dichas propiedades para los atributos “Provincia Nombre”, “Comarca” y “Municipio” que se encuentran dentro de la dimensión:



Ilustración - KeyColumns provincia\_nombre.

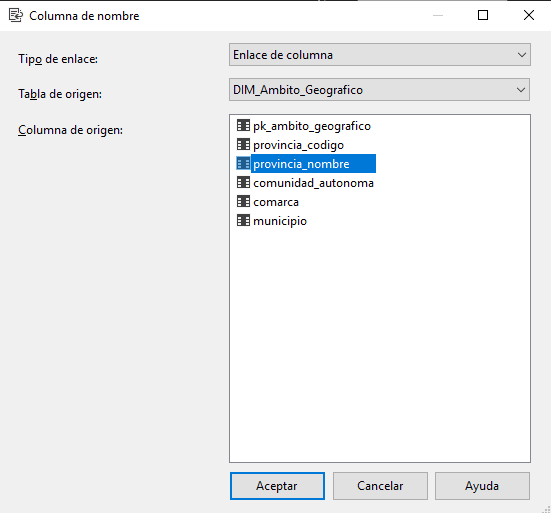


Ilustración NameColumn provincia\_nombre.

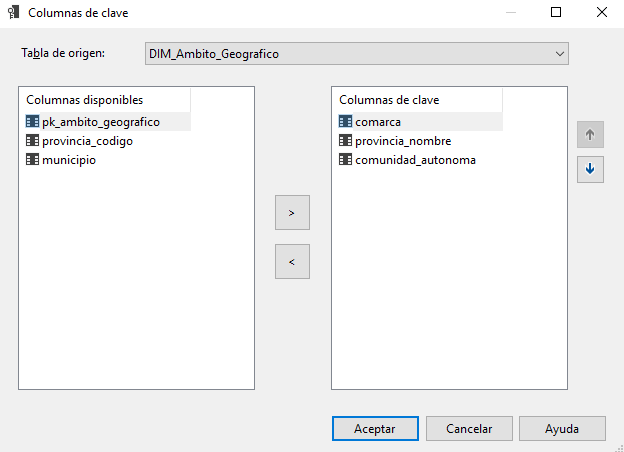


Ilustración - KeyColumns comarca.

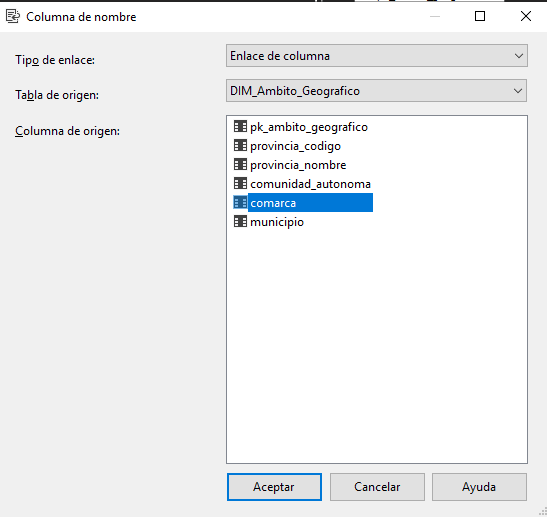


Ilustración - NameColumn comarca.

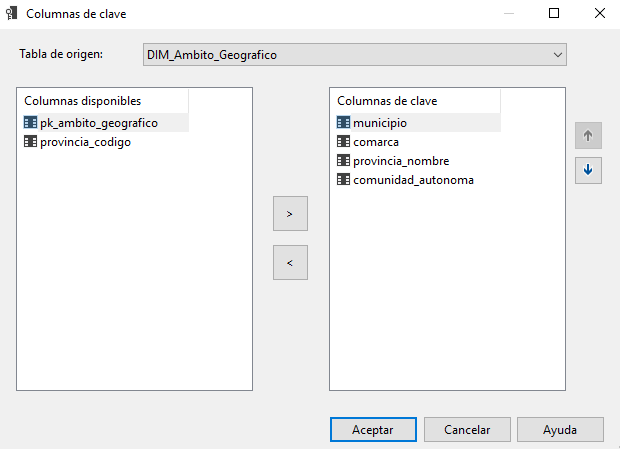


Ilustración - KeyColumns municipio.

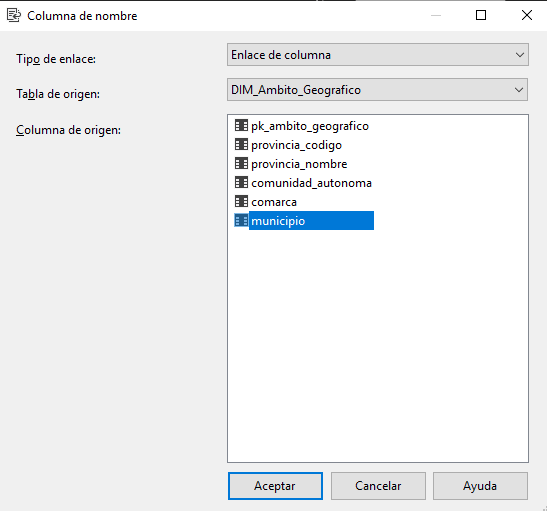


Ilustración - NameColumn municipio.

### DIM\_Fecha

Realizamos el mismo proceso para la dimensión “DIM\_Fecha”, arrastramos los atributos desde la vista del origen de datos a los atributos de la dimensión:

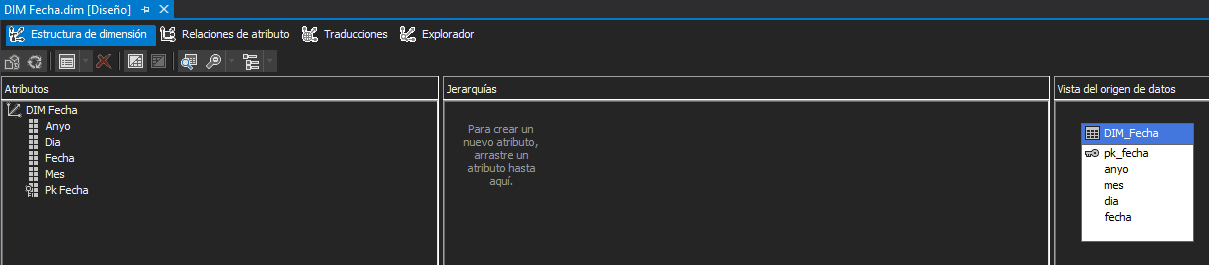


Ilustración - DIM\_Fecha.

Al igual que sucedía con la dimensión anterior, definimos una jerarquía para facilitar la construcción de las consultas después. La jerarquía va a ser “Anyo” > “Mes” > “Dia”:

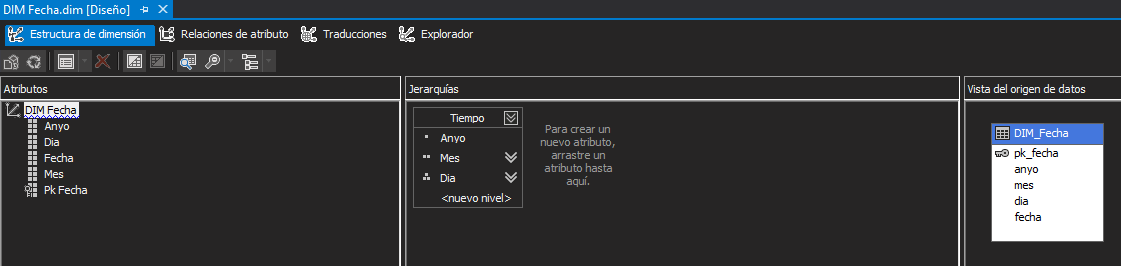


Ilustración - Jerarquía DIM\_Fecha.

Cuando definimos la jerarquía tenemos que cambiar las propiedades “*KeyColumns*” y “NameColumn” al igual que sucedía con la anterior dimensión.

Por lo tanto, vamos a definir dichas propiedades para los atributos “Mes” y “Dia” que se encuentran dentro de la dimensión:

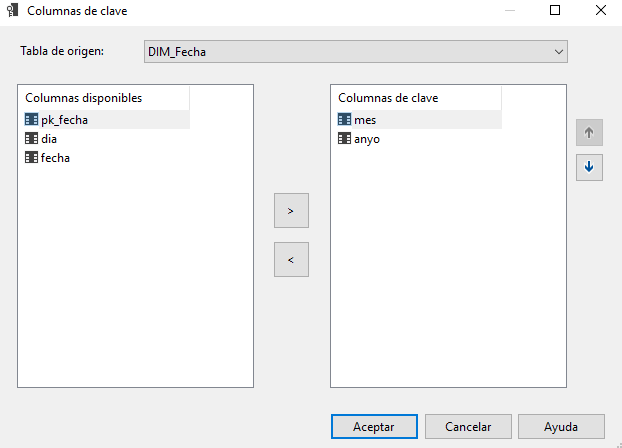


Ilustración - KeyColumns mes.

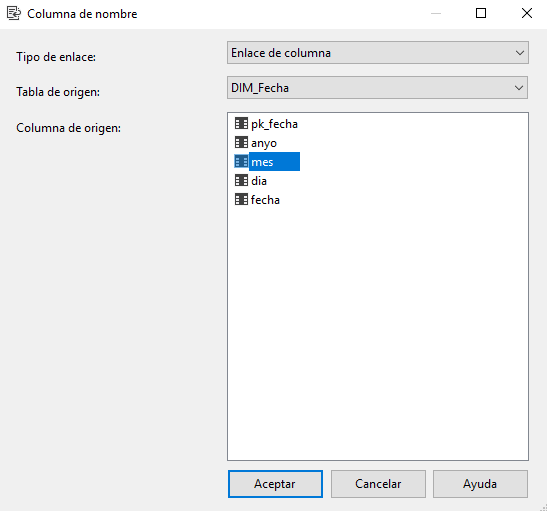


Ilustración - NameColumn mes.



Ilustración - KeyColumns dia.

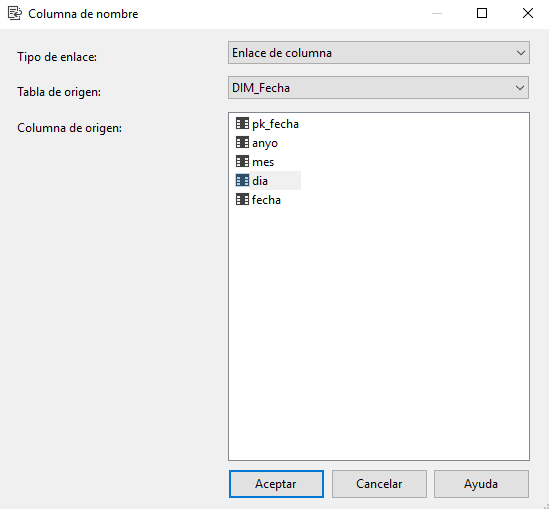


Ilustración - NameColumn dia.

### DIM\_Grupo\_Edad

Para la dimensión “DIM\_Grupo\_Edad” arrastramos al igual que antes todos los atributos:

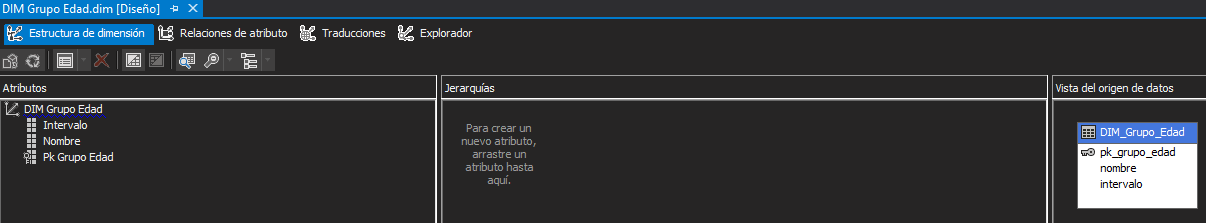


Ilustración - DIM\_Grupo\_Edad.

### DIM\_Medicion

Realizamos el mismo proceso para “DIM\_Medicion”:

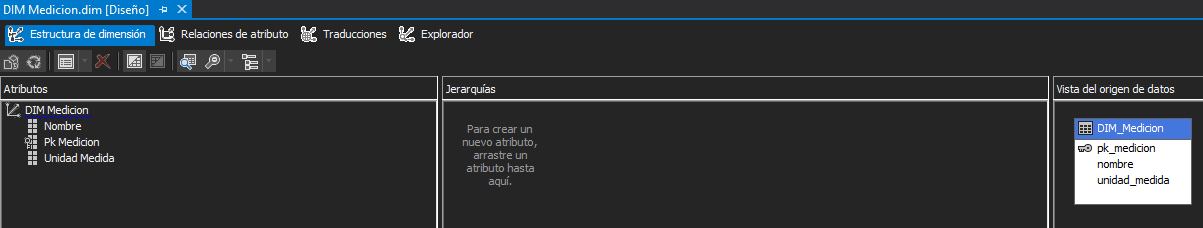


Ilustración - DIM\_Medicion.

### DIM\_Tipologia

Finalmente, definimos la dimensión “DIM\_Tipologia” a partir de los atributos de la vista del origen de datos:

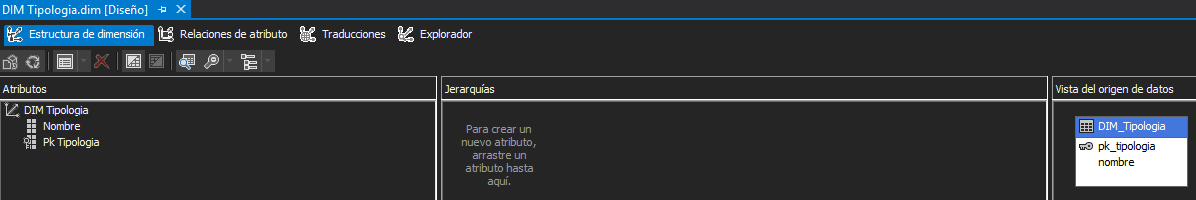


Ilustración - DIM\_Tipologia.

|  |
| --- |
| Implementación de la solución |

Antes de comenzar con la explotación de los datos, tenemos que implementar la solución definida en el punto anterior.

Para poder implementarla, nos dirigimos al menú “Compilar” y seleccionamos la opción “Implementar solución”:

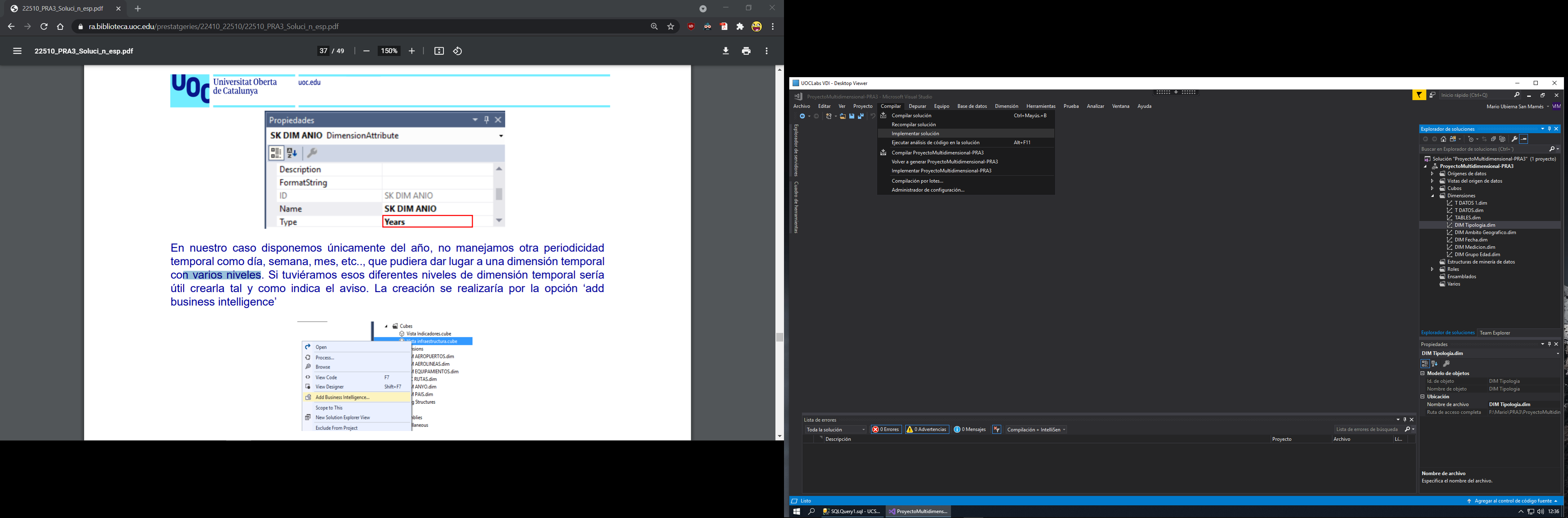


Ilustración - Implementar la solución.

Como resultado de implementar la solución obtenemos la siguiente ventana:

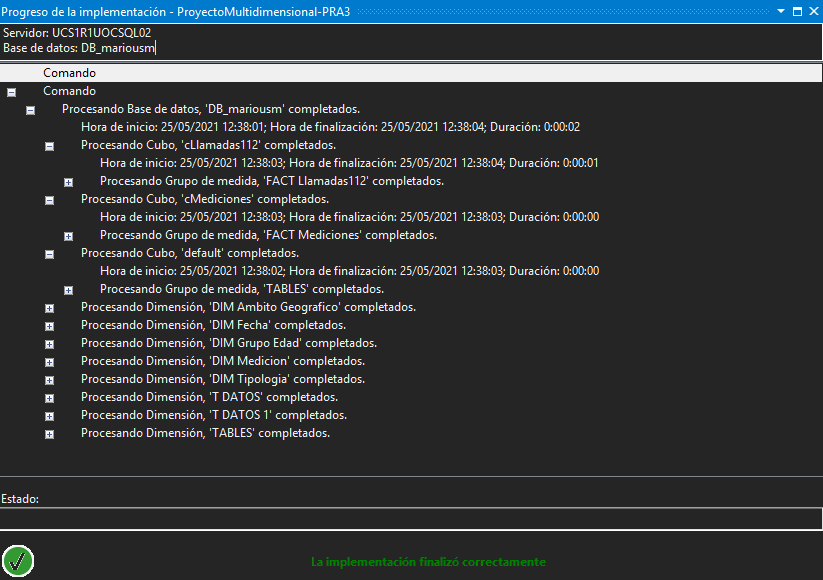


Ilustración - Implementación de la solución.

Cabe destacar que hay un cubo “default” y tres dimensiones: “T DATOS”, “T DATOS 1” y “TABLES” que no son necesarias para la resolución de la práctica. Sin embargo, para no perder permisos hemos decido mantener todos los elementos que teníamos cuando hemos creado el proyecto desde cero.

Finalmente, vemos las recomendaciones que nos sugiere *Visual Studio*:

* Definir una dimensión temporal.
* Uso de relaciones rígidas en lugar de flexibles, tanto en “DIM Fecha” como “DIM Ambito Geografico”.
* Ocultación de atributos si se usan jerarquías, tanto en “DIM Fecha” como “DIM Ambito Geografico”
* No existe una relación entre atributos entre uno o más niveles de jerarquía, tanto en “DIM Fecha” como “DIM Ambito Geografico”.
* Crear jerarquías en dimensiones secundarias no principales, esto sucede para “DIM Grupo Edad”, “DIM Medicion”, “DIM Tipologia”.

Para solventar la dimensión temporal nos dirigimos a la dimensión “DIM Fecha”, la cual se encuentra dentro de la carpeta “Dimensiones” en el explorador de soluciones, e indicamos que dicha dimensión a partir de tu propiedad “*Type*” sea “*Time*” en vez de “Regular”:

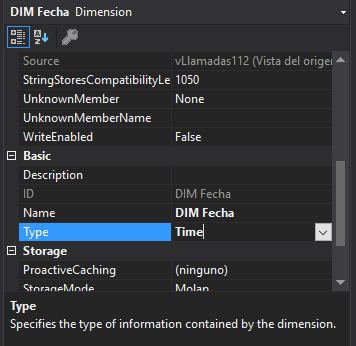


Ilustración - Creación de la dimensión temporal.

Al tener los días, los meses, los años, e incluso la fecha en sí, modificamos los atributos de la dimensión para indiciar que dichos campos son de tipo “*days*”, “*months*”, “*years*” y “*date*” respectivamente:

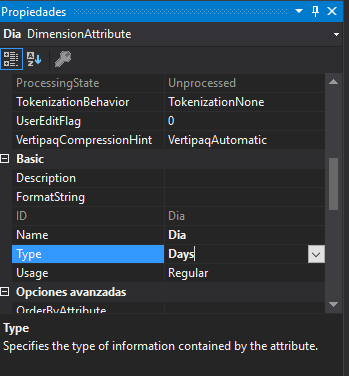


Ilustración - Asignación del tipo para el día.

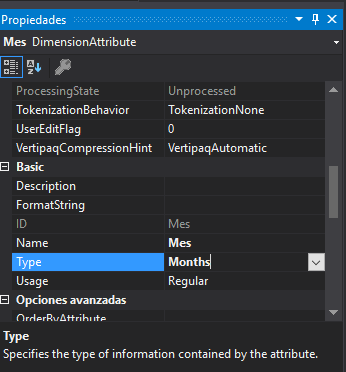


Ilustración - Asignación del tipo para el mes.

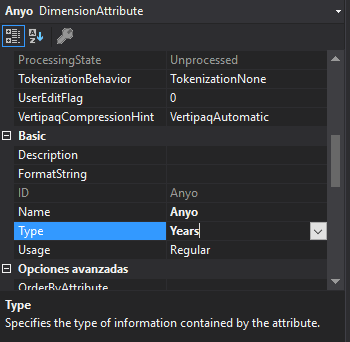


Ilustración - Asignación del tipo para el año.



Ilustración - Asignación del tipo para la fecha.

Ahora tenemos que solventar los *warnings* relativos a las dimensiones “DIM Fecha” y “DIM Ambito Geografico”. Lo primero de todo es analizar la recomendación de ocultar atributos si se usan jerarquías, para ello podemos cambiar la visibilidad de los atributos, pero en nuestro caso no vamos a realizar este cambio ya que así podemos hacer uso de ellos directamente en la explotación de los datos.

La siguiente recomendación es definir una relación entre atributos entre uno o más niveles de jerarquía, para ello definimos dicha relación tanto en la dimensión “DIM Fecha” como “DIM Ambito Geografico”, como aporte cabe destacar que en diseño de las dimensiones se ha explicado la jerarquía y el cómo la hemos definido:



Ilustración - Relación de atributos DIM Ambito Geografico.



Ilustración - Relación de atributos DIM Fecha.

Finalmente, para terminar con las advertencias de estas dimensiones tenemos que definir relaciones rígidas en vez de flexibles entre los atributos de las jerarquías, a modo de ejemplo se muestra el cómo se define para un atributo:

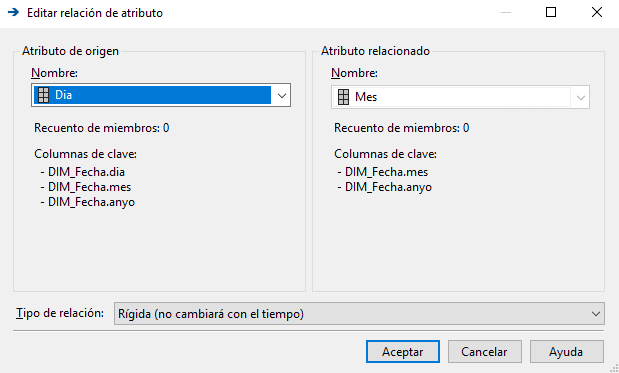


Ilustración - Establcer la relación de flexible a rígida.

Como podemos apreciar, una vez seleccionada la relación modificamos abajo el tipo de relación.

Para terminar con este punto mostramos el cómo quedan finalmente las relaciones rígidas tanto en “DIM Fecha” como “DIM Ambito Geografico”:

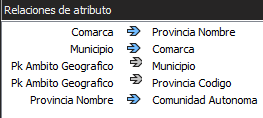


Ilustración - Relaciones rígidas en DIM Ambito Geografico.

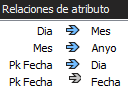


Ilustración - Relaciones rígidas en DIM Fecha.

Cabe destacar, que solo hemos definidos como relaciones rígidas aquellas en los que los atributos pertenecían a la jerarquía, ya que los que no pertenecen no es necesario definir dicha relación como rígida.

Para terminar con una correcta implementación, cabe mencionar que nos salta un *warning* para las dimensiones “DIM Grupo Edad”, “DIM Medicion” y “DIM Tipologia”, esta recomendación es la de definir una jerarquía, pero en nuestro caso no nos hace falta ya que dentro de cada dimensión tenemos relaciones uno a uno y no uno a varios, por lo que todos los valores son únicos y podemos hacer uso de ellos directamente para la explotación de los datos.

|  |
| --- |
| Explotación de la solución |

Una vez que tenemos el diseño de los cubos y los hemos creado, podemos hacer uso de ellos para obtener información sobre los datos.

Para realizar cada una de las consultar que se van a ver a continuación nos tenemos que dirigir a la pestaña “Explorador”, la cual se encuentra dentro de la ventana del diseño del cubo correspondiente.

## Análisis de las provincias con mayor porcentaje de movilidad según datos móviles

Para realizar esta consulta lo primero que debemos de indicar es el filtrado de la dimensión “DIM\_Medidas” con el atributo “Movilidad de la población durante el estado de alarma”, para así definir que lo que vamos a analizar es el porcentaje de movilidad de la población durante el estado de alarma. Además, debemos de filtrar las fechas, es decir, que solo vamos a estudiar la movilidad durante el estado de alarma, para ello seleccionamos el año 2020 y los meses 3, 4, 5 y 6:

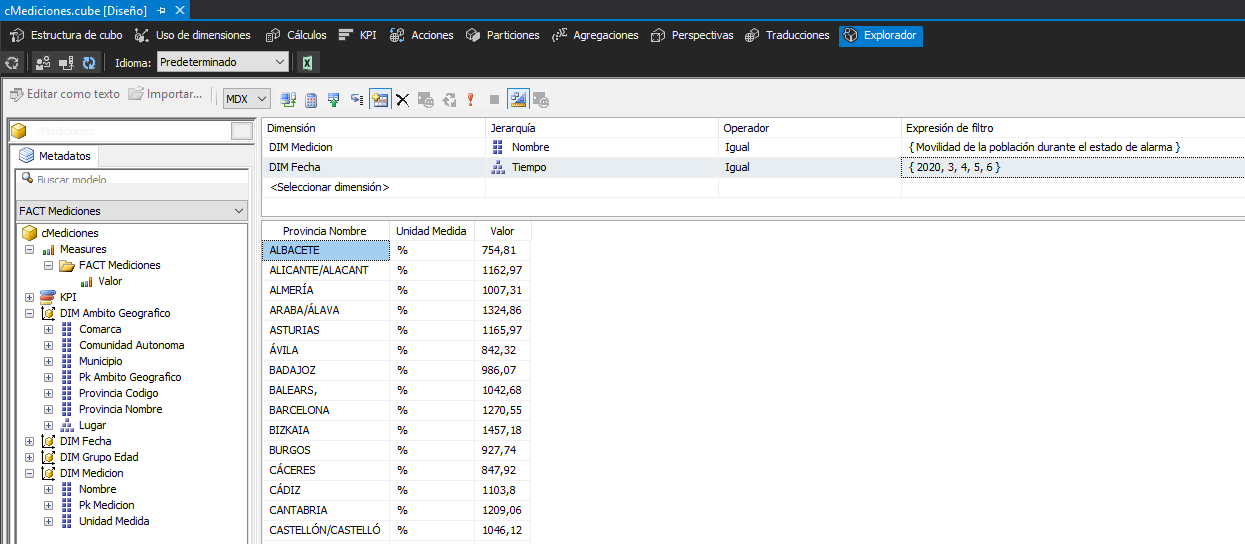


Ilustración - Análisis de las provincias con mayor porcentaje de movilidad.

Al ejecutar la anterior consulta, vemos que el porcentaje que obtenemos es mayor que el 100%, esto se debe a que está haciendo un sumatorio y no está haciendo la media, por lo que creamos un nuevo campo a partir del campo “Valor”, y éste se encarga de hacer la media:

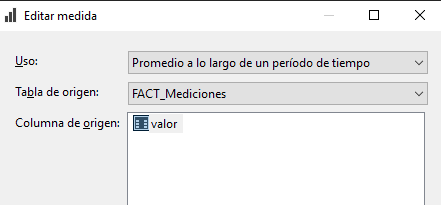


Ilustración - Creación de la media.

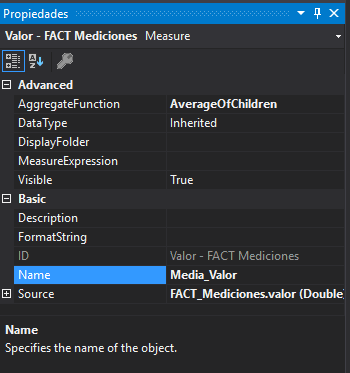


Ilustración - Creación de la media.

Ahora volvemos a crear la consulta con el mismo filtrado que antes tanto para el ámbito geográfico, como para la fecha, con la excepción de que vamos a visualizar el nuevo campo creado “Media\_Valor” en vez de “Valor”:

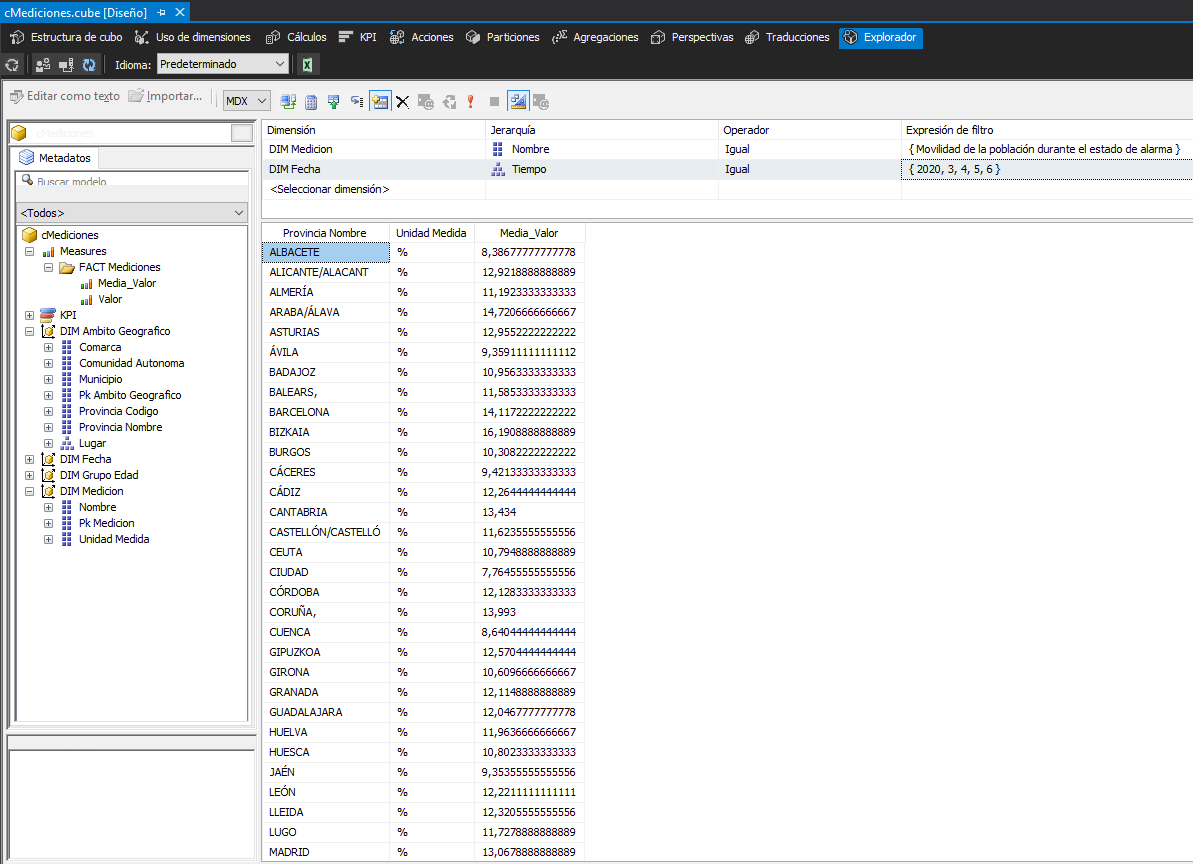


Ilustración - Análisis de las provincias con mayor porcentaje de movilidad.



Ilustración - Análisis de las provincias con mayor porcentaje de movilidad.

De la anterior ejecución vemos que ya tenemos toda la información que queremos analizar, sin embargo, se busca analizar las provincias con mayor porcentaje de movilidad, es por ello que modificamos la consulta manualmente para ordenar el campo “Media\_Valor”:

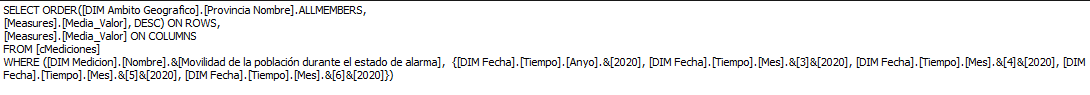


Ilustración - Consulta MDX para el análisis de las provincias con mayor porcentaje de movilidad.

Si ejecutamos dicha consulta obtenemos el siguiente resultado:



Ilustración - Análisis de las provincias con mayor porcentaje de movilidad.

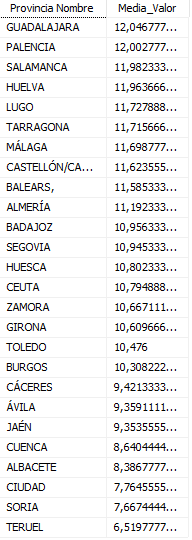


Ilustración - Análisis de las provincias con mayor porcentaje de movilidad.

De la anterior consulta observamos algo curioso, y es que por norma general las provincias que contienen a la capital de la comunidad tienden a tener un mayor porcentaje de movilidad que las demás. Esto puede llegar a ser lógico ya que hay más población en dichas ciudades y, por lo tanto, más servicios básicos se necesitaban durante el estado de alarma.

También es interesante ver que todas las provincias de Galicia se encuentra en los puestos más altos del ranking, es decir, que en todas ellas el porcentaje de movilidad es elevado, no tenemos la información suficiente como para determinar la causa de ésta pero es algo llamativo.

## Análisis del porcentaje de la población que evitaba aglomeraciones según la comunidad autónoma

Lo primero que debemos hacer es filtrar la medición para que solo tenga en cuenta el atributo “Porcentaje de la población que evitaba aglomeraciones”, por otro lado, seleccionamos tanto la comunidad como las métricas para obtener la información:

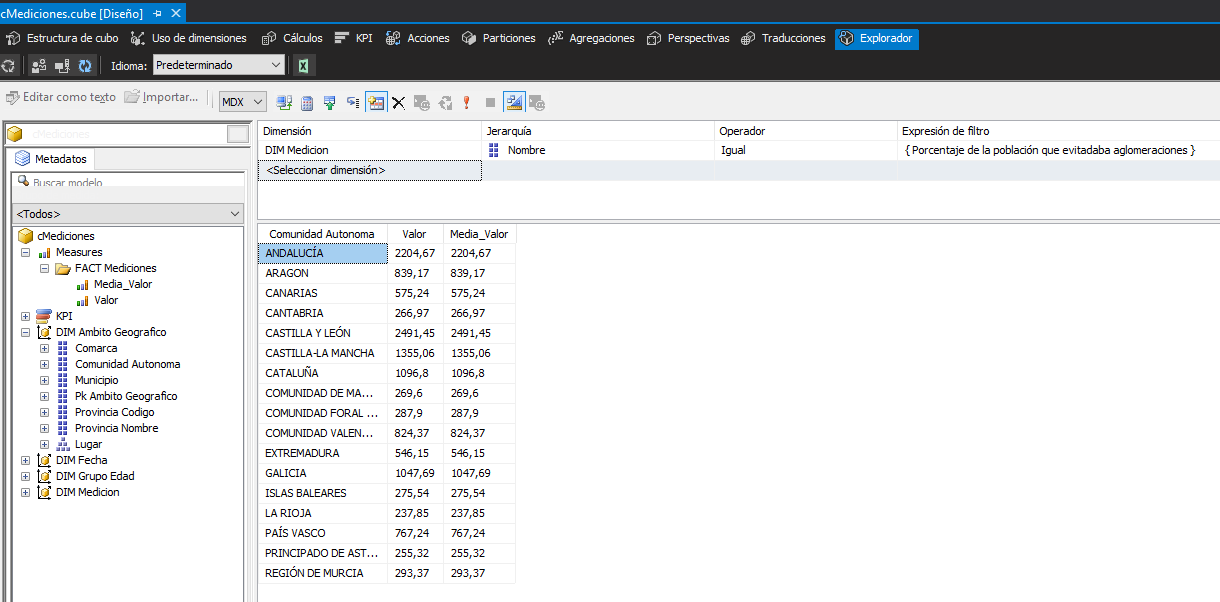


Ilustración - Análisis del porcentaje de población que evitaba aglomeraciones según la CCAA.

Como podemos apreciar, sí que nos devuelve la información pero la media no la está realizando correctamente, ya que dicha métrica realiza el promedio a lo largo de un periodo de tiempo, en nuestro caso, dicho periodo es siempre el mismo 30 de septiembre del 2020.

Por lo tanto, necesitamos hacer el cálculo del valor entre (número de provincias de la CCAA \* 6 (tenemos 6 grupos de edad)), para ello definimos una nueva métrica que va a ser el número de filas, ésta nos dará como resultado el valor de número de provincias de la CCAA \* 6 grupos de edad:

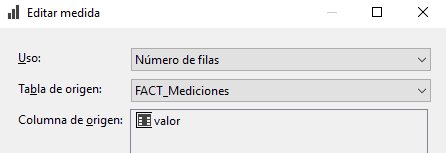


Ilustración - Creación número de filas.

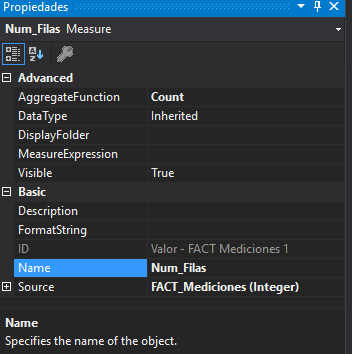


Ilustración - Creación número de filas.

De tal forma que, si volvemos a ejecutar la consulta anterior, obtenemos el siguiente resultado:



Ilustración - Análisis del porcentaje de población que evitaba aglomeraciones según la CCAA.

Ahora lo que nos faltaría es dividir el “Valor” entre “Num\_Filas”, de esa forma tendríamos unos datos coherentes relativos al porcentaje de población que evitaba aglomeraciones por cada comunidad. Para ello, creamos un miembro calculado el cual llamamos “Media\_Pob\_Aglo”:

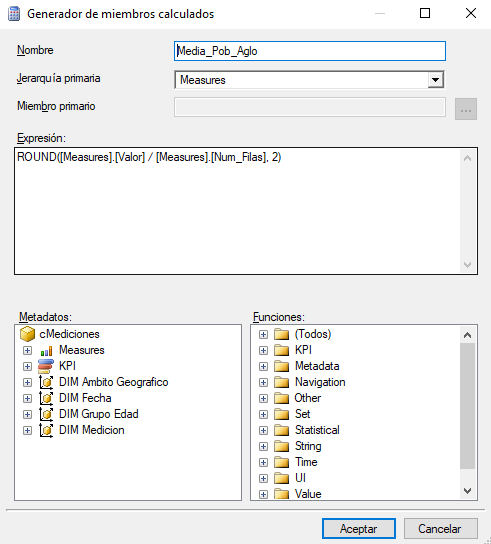


Ilustración - Creación media población aglomeración.

Cuando ya tenemos el nuevo campo creado lo incluimos en el diseño de la consulta, de esta forma nos proporciona el siguiente resultado:

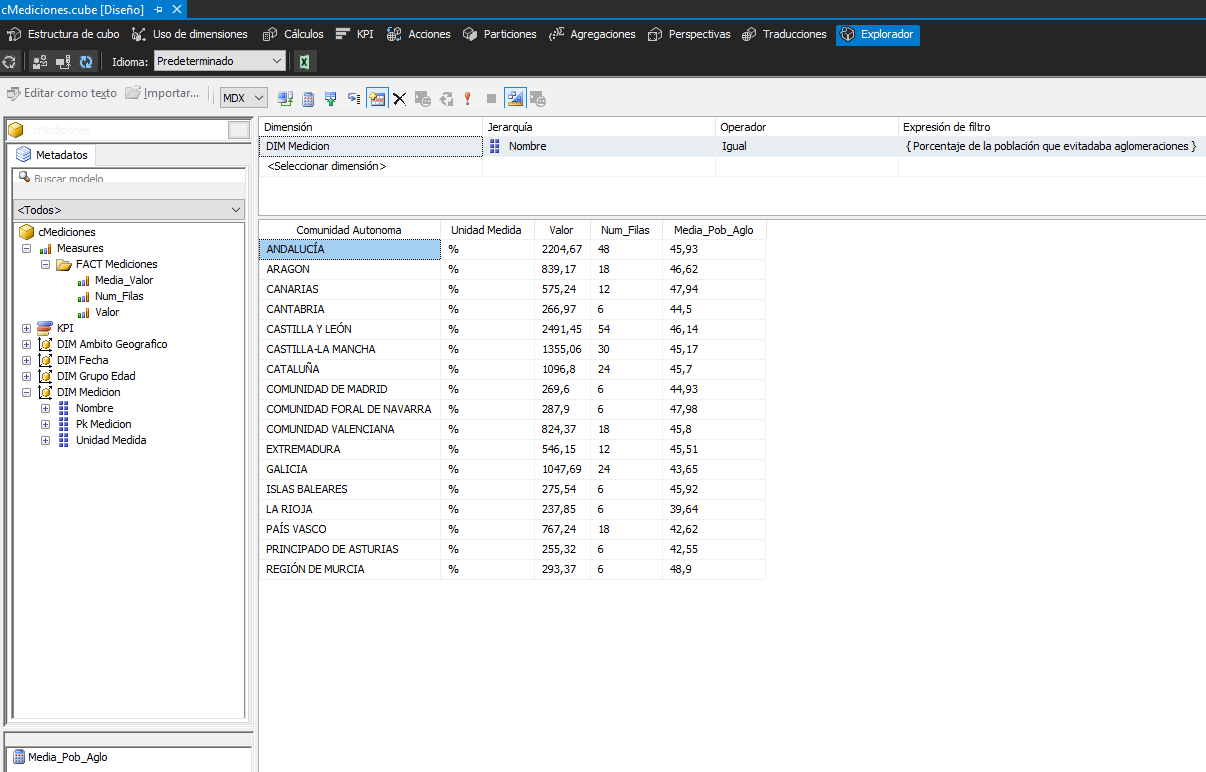


Ilustración - Análisis del porcentaje de población que evitaba aglomeraciones según la CCAA.

Del análisis de la anterior consulta podemos concluir que está bastante nivelado el porcentaje de población que evitaba aglomeraciones según la comunidad autónoma. Cabe destacar que la comunidad autónoma que más porcentaje de población evitaba las aglomeraciones era la Región de Murcia con un 48.9%, sin embargo, La Rioja es la peor comunidad con un 39.64%, estos datos son significativos porque vemos que da igual si la comunidad autónoma es más grande o no, ya que La Rioja y la Región de Murcia son similares en cuanto a extensión y sus porcentajes son muy dispares.

Por otro lado, aunque no hay mucha diferencia entre las comunidades sí que podemos ver una ligera tendencia a que cuanto menor población en la comunidad autónoma menor es también el porcentaje de población que evitaba aglomeraciones.

## Análisis del promedio de sanciones por habitantes

A

## Evolución de las llamadas de urgencia al 112 en Cataluña por tipología de llamada

Para realizar un análisis evolutivo lo que debemos de tener en cuenta es la dimensión del tiempo. Por otro lado, nos pide que se quiere analizar el número de llamadas según la tipología de la misma.

Por lo tanto, a la hora de hacer la consulta en el cubo vamos a añadir la fecha, el nombre de la tipología y la medida para obtener el número de llamadas:



Ilustración - Evolución de las llamadas según la tipología.

Como podemos apreciar en la anterior consulta, al incluir el campo fecha nos muestra la evolución del número de llamadas según su tipología para todos los meses que hay desde el 2014 hasta el 2020. Esto nos da como resultado una consulta un poco grande, y como en el enunciado no se pide si la evolución de las llamadas se tiene que estudiar con granularidad de mes o año, a decisión personal se decide estudiar la evolución con granularidad anual ya que nos da un análisis más general, de tal forma sustituimos el campo “Fecha” por “Anyo”:

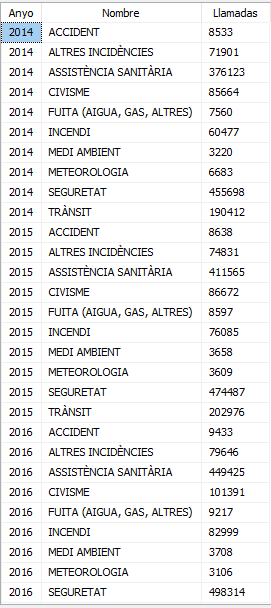


Ilustración - Evolución de las llamadas según la tipología.

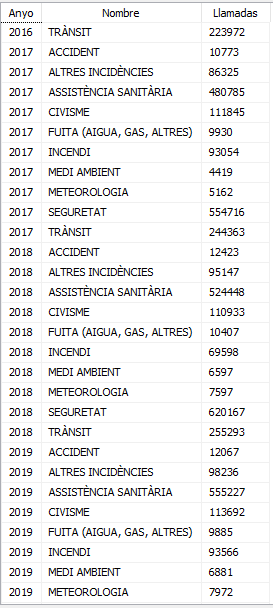


Ilustración - Evolución de las llamadas según la tipología.

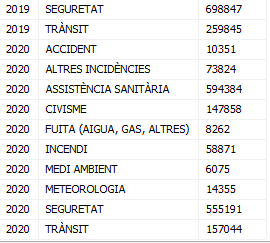


Ilustración - Evolución de las llamadas según la tipología.

Si analizamos los datos de forma general todas las tipologías de llamadas presentan una tendencia al alza, es decir, a medida que va pasando el tiempo se producen un mayor número de llamadas en casi todas las tipologías.

Al estar analizando el impacto del *COVID* en nuestra sociedad, tenemos que fijarnos más detalladamente en las incidencias que se producen en el año 2020, ya que es en éste cuando comienza el problema de dicha enfermedad. Si observamos las incidencias producidas en este año vemos que la gran mayoría van a la baja, es decir, comparando el año 2020 con los demás suele haber un menor número de incidencias. Sin embargo, en este año hay más incidencias respecto a la asistencia sanitaria y de civismo, lo cual tiene todo el sentido del mundo ya que fue un momento difícil a nivel sanitario y también a nivel de comportamiento de la sociedad.

## Análisis de las llamadas de urgencia frente al porcentaje de la población española que evitaba las aglomeraciones entre los meses comprendidos entre marzo y junio de 2020 en Cataluña, desglosado por provincia

Para solventar dicho requisito vamos a realizar primero el análisis de las llamadas de urgencia entre los meses comprendidos entre marzo y junio de 2020 desglosado por provincia.

Por lo tanto, lo primero que debemos hacer es que nos muestre el atributo “Provincia Nombre” y la métrica “Llamadas”, de esta forma conseguimos saber el número de llamadas que se han producido en cada provincia de Cataluña. Finalmente, añadimos un filtro para solo contabilizar las fechas que están comprendidas entre marzo y junio del 2020:

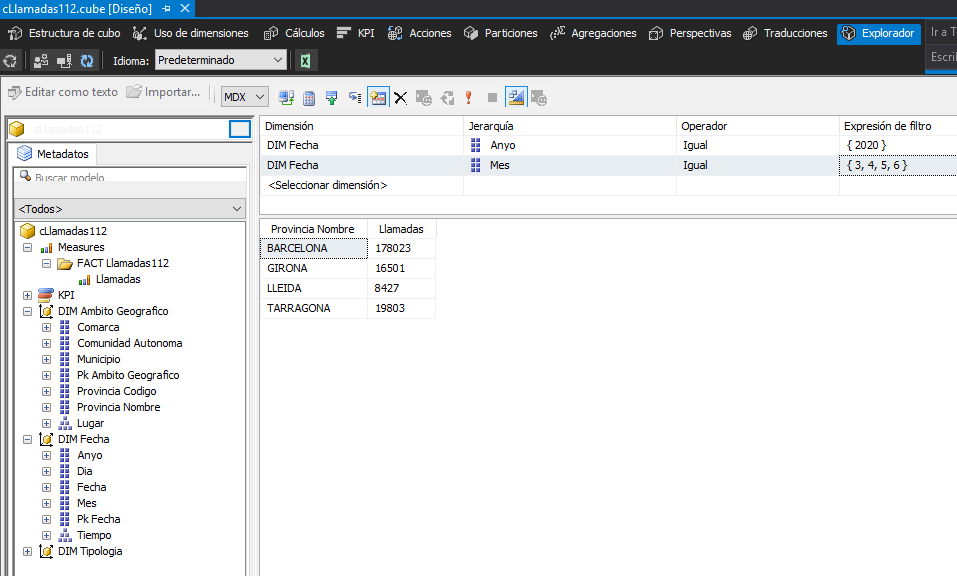


Ilustración – Análisis de las llamadas en Cataluña en época de COVID.

Por otro lado, tenemos que hacer lo mismo para el porcentaje de población que evitaba las aglomeraciones, cabe destacar que los datos relativos a este análisis tienen todos la misma fecha, el 30 de septiembre del 2020, como en la fuente de datos original no indica de qué fecha son, vamos a considerar que se encuentran en el año natural 2020.

Una vez aclarado esto diseñamos la consulta, indicamos que nos desglose la información por el nombre de la provincia. Al igual que sucedía con la consulta del apartado dos, la métrica nos realiza un sumatorio de todos los grupos de porcentajes que tenemos, en nuestro caso hay 6 grupos, es por ello que realizamos la media para cada provincia, para así tener datos más coherentes y poder extraer la información de forma adecuada.

Antes de ejecutar la consulta tenemos que definir los filtros, es decir, indicamos que queremos obtener la métrica correspondiente a “Porcentaje de la población que evitaba aglomeraciones”, indicamos que nos muestre solo las provincias de Cataluña y establecemos que la fecha vaya desde marzo a junio del 2020:

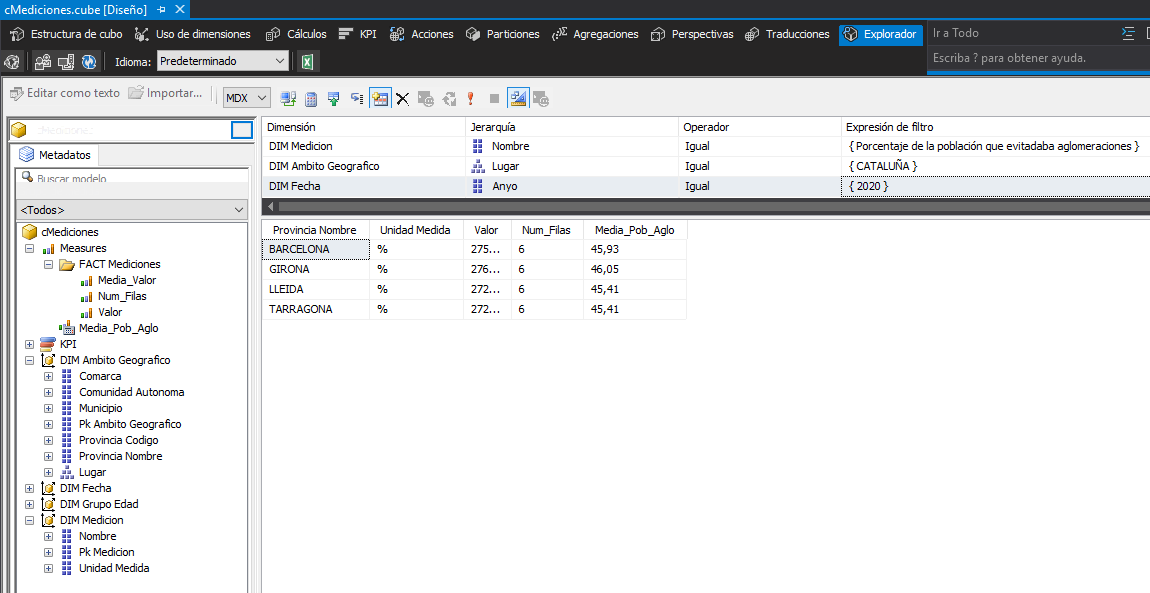


Ilustración - Análisis del porcentaje de población que evitaba aglomeraciones en época de COVID.

Analizando las dos anteriores consultas vemos que Barcelona es la provincia con mayor número de llamadas telefónicas de urgencias (lógico porque es la que tiene mayor población de todas ellas), además el porcentaje de población que evitaba aglomeraciones es mayor que en el caso de Lleida y Tarragona, lo cual es llamativo ver que provincias con menos población y muchas menos llamadas no solían evitar las aglomeraciones, es decir, había un comportamiento menos disciplinado en las provincias pequeñas que en Barcelona. Otra caso a destacar es Girona que tiene un número de llamadas de urgencia bajo y el porcentaje de población que evitaba aglomeraciones es el más alto, esto nos hace indicar que en dicha provincia se tomaron las medidas adecuadas y la gente era más consciente de la crisis sanitaria que se estaba viviendo.

## Determinación del día de la semana con menor número de denuncias

A

## Análisis de las diez fechas (*top ten*) con mayor número de llamadas de urgencia al 112 con tipología de tránsito registrada, tanto en época de *COVID* como antes

Para realizar este análisis vamos a hacer realmente dos consultas, una para las llamadas de urgencia con tipología “*TRÀNSIT*” en época de *COVID*, y la misma consulta pero cuando no era época *COVID*.

Por lo tanto, lo primero de todo es analizar el número de llamadas para la época del *COVID,* como no nos indican cuándo empieza la época de *COVID* hemos considerado que todo el año 2020 es época de *COVID,* es decir, aunque el estado de alarma fue en marzo realmente desde finales del año 2019 / comienzo del 2020 dicha enfermedad ya estaba en circulación por nuestro país. Además, vamos a realizar la consulta a partir de MDX para así obtener un *top ten* ordenado:

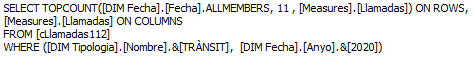


Ilustración - Consulta número de llamadas de tipo tránsito en época COVID.

Al ejecutar la anterior consulta obtenemos el siguiente resultado:

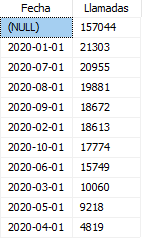


Ilustración - Análisis número de llamadas de tipo tránsito en época de COVID.

Cabe destacar que la fila “*NULL*” se corresponde con todos los registros que hay en el cubo y no son de la tipología “TRÀNSIT”, como éstos son mayoritarios ocupan la primera posición del ranking pero dicha información es innecesaria, es por ello que hacemos un ranking de 11 en vez de 10, para que así nos muestre las 10 fechas reales no 9 fechas más 1 de “*NULL*”.

Analizado este resultado vemos que lógicamente durante el estado de alarma, las llamadas de urgencias al 112 de tipo tránsito son menores que en el resto de meses, esto se debe a que durante los meses de marzo, abril, mayo y junio solo se podían mover los ciudadanos bajo un permiso. Además, una vez finalizado el permiso vemos que los siguientes meses tienen un número elevado de llamadas bajo este motivo.

Ahora realizamos el mismo proceso para todos los años desde el 2014 al 2019 incluidos:

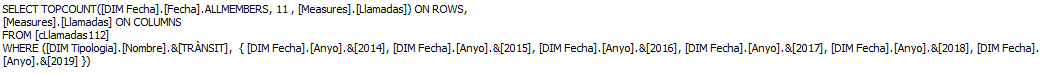


Ilustración - Consulta número de llamadas de tipo tránsito no época de COVID.

Al ejecuta la anterior consulta obtenemos la siguiente información:

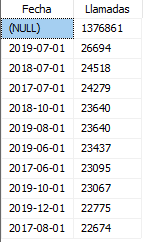


Ilustración - Análisis número de llamadas de tipo tránsito no época de COVID.

Al igual que sucedía antes, la primera fila nos recoge todas aquellas llamadas que no son de tipo tránsito ni en esas fechas, por lo que al ser mayoría ocupa la primera posición, es por ello que para calcular el top lo hacemos con un ranking de 11 al igual que en el caso anterior.

Analizando los datos vemos que tanto si hay *COVID* como si no lo hay, las mayores incidencias de tránsito se producen en los meses de verano y en los primeros de otoño, esto tiene todo el sentido del mundo ya que es en esa época en la que los ciudadanos más se mueven debido a las vacaciones.

También cabe destacar que el top de fechas son del 2017 en adelante, cuando hemos analizado desde el 2014, esto nos refleja que una de dos, o antes se recogían menos datos, o este tipo de incidencia sigue una tendencia positiva, es decir, va aumentado año tras año.