



Universidad Abierta Interamericana.

INSTRUCTIVO PARA LA APLICACIÓN DE LA TÉCNICA DE MINERÍA DE DATOS.

“CLUSTERING”.

Profesor: Horacio Caballero.

Integrantes:

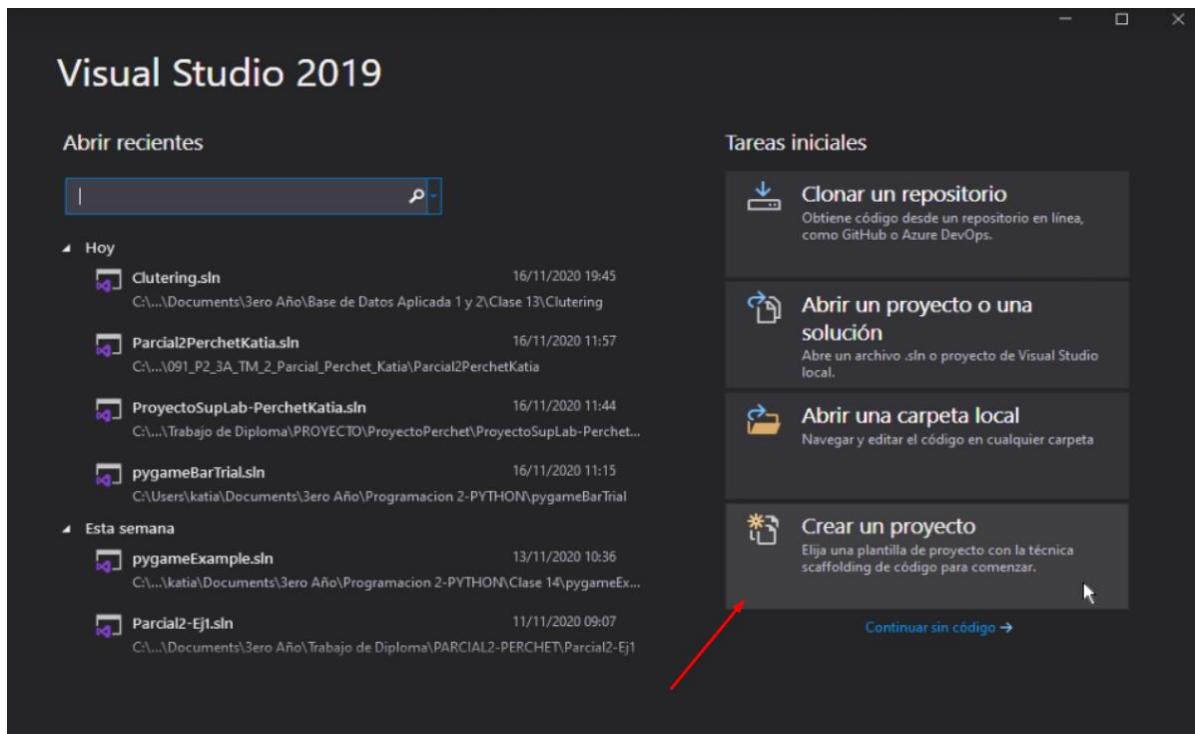
- Chmelik, Lucas.
- Peralta, Flavio.
- Perchet, Katia.
- Sasso, Patricio.

Contenidos del Instructivo.

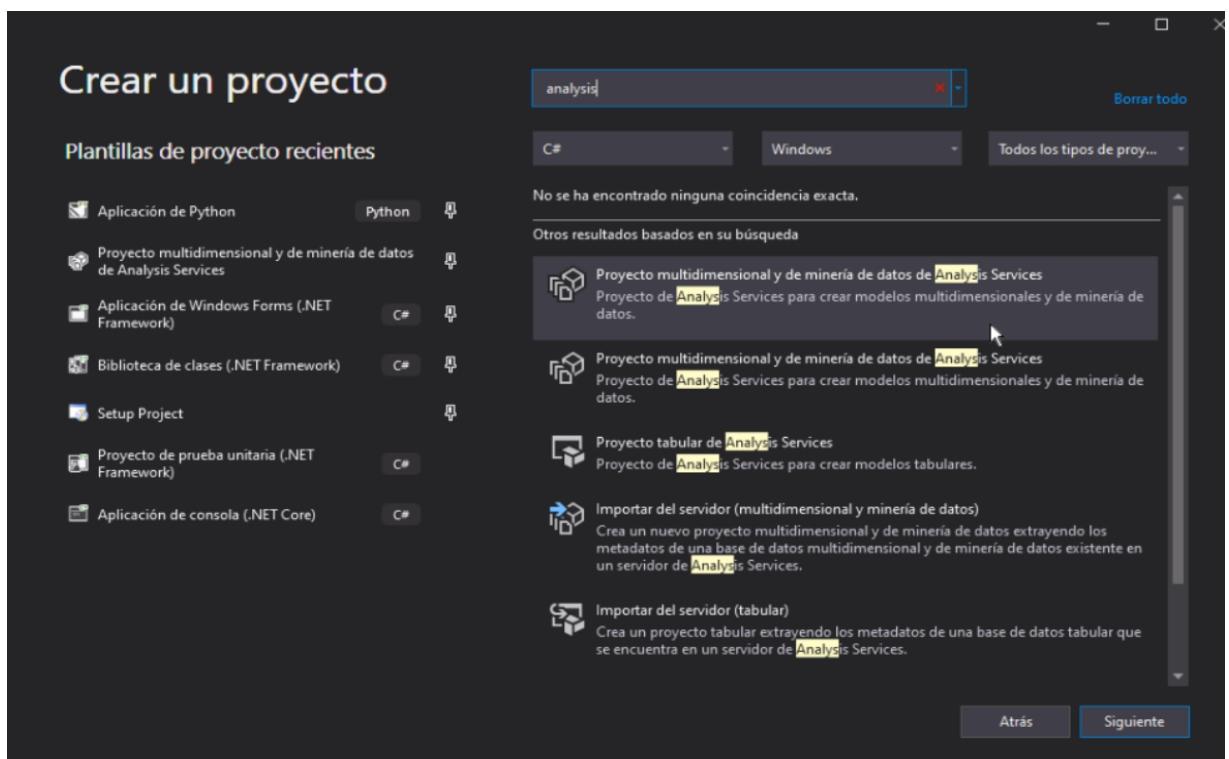
Creación del Proyecto	3
Creación del Origen de Datos	4
Creación de Vista del Origen de Datos.....	8
Creación del Cubo.....	11
Conclusión de la creación de cubos	14
Modificación de Dimensiones y establecimiento de Jerarquías	14
Implementación de la técnica “Clustering”	16
Conclusión de aplicar técnica de clustering para minería de datos.....	26
Errores posibles	27
No usar instancia predeterminar de SQL server	27
Usuario sin permisos necesarios	31

Creación del Proyecto.

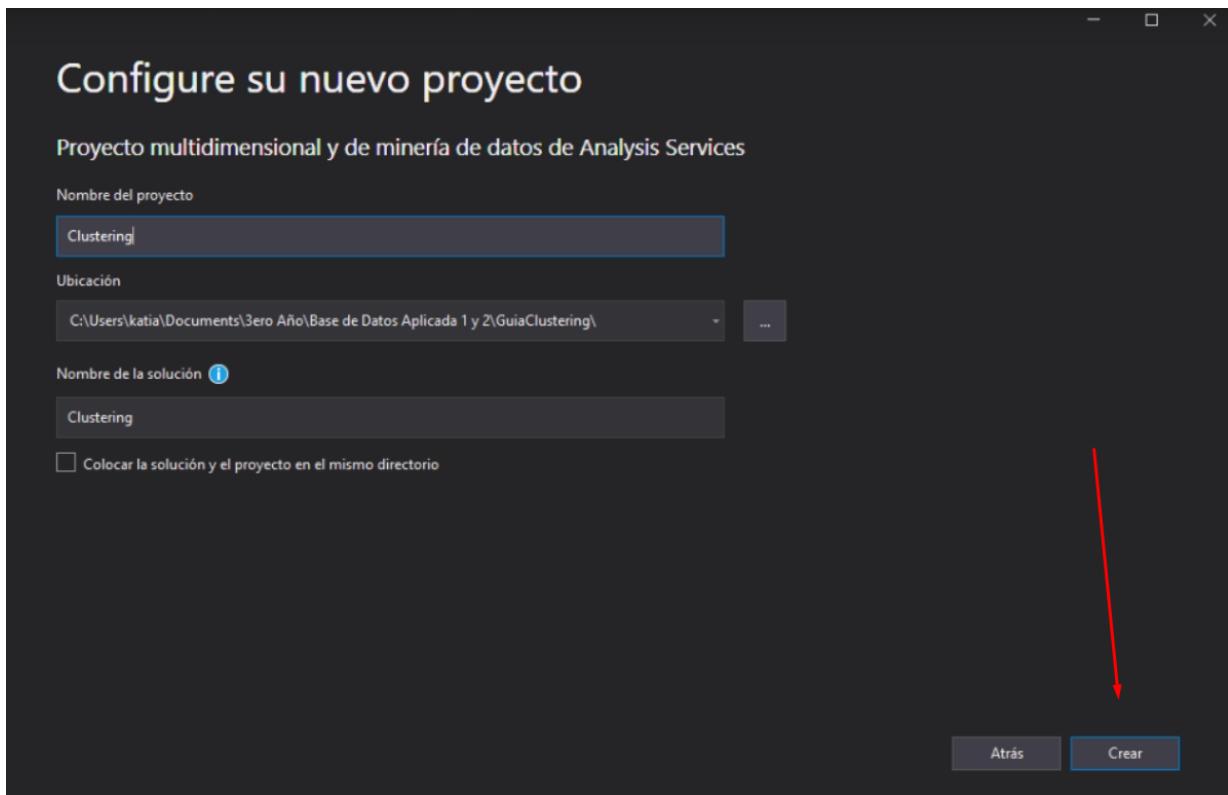
- Iniciar un nuevo proyecto en Visual Studio o en SQL Server Data Tools, nosotros utilizaremos Visual Studio 2019.



- Seleccionar el tipo de proyecto “Proyecto multidimensional y de minería de datos de Analysis Services”.

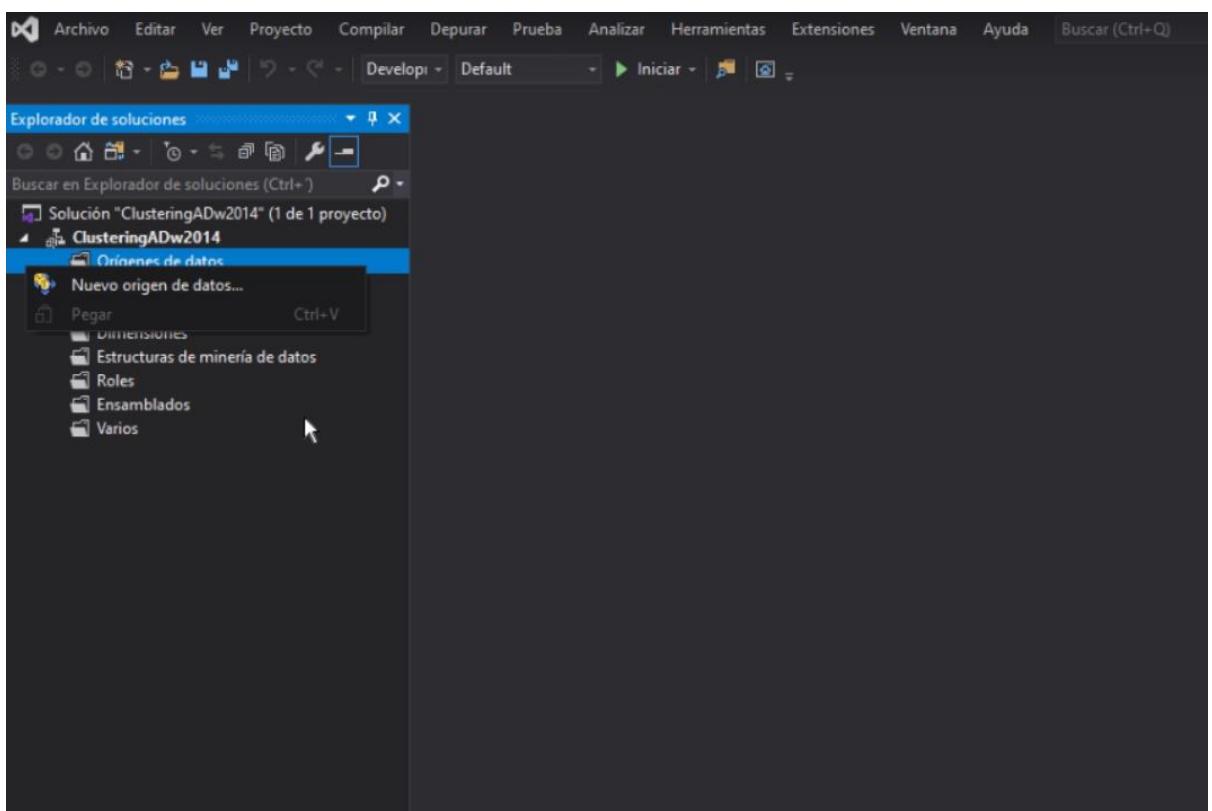


- Ingresar nombre deseado y ubicación, luego presione “Crear”.

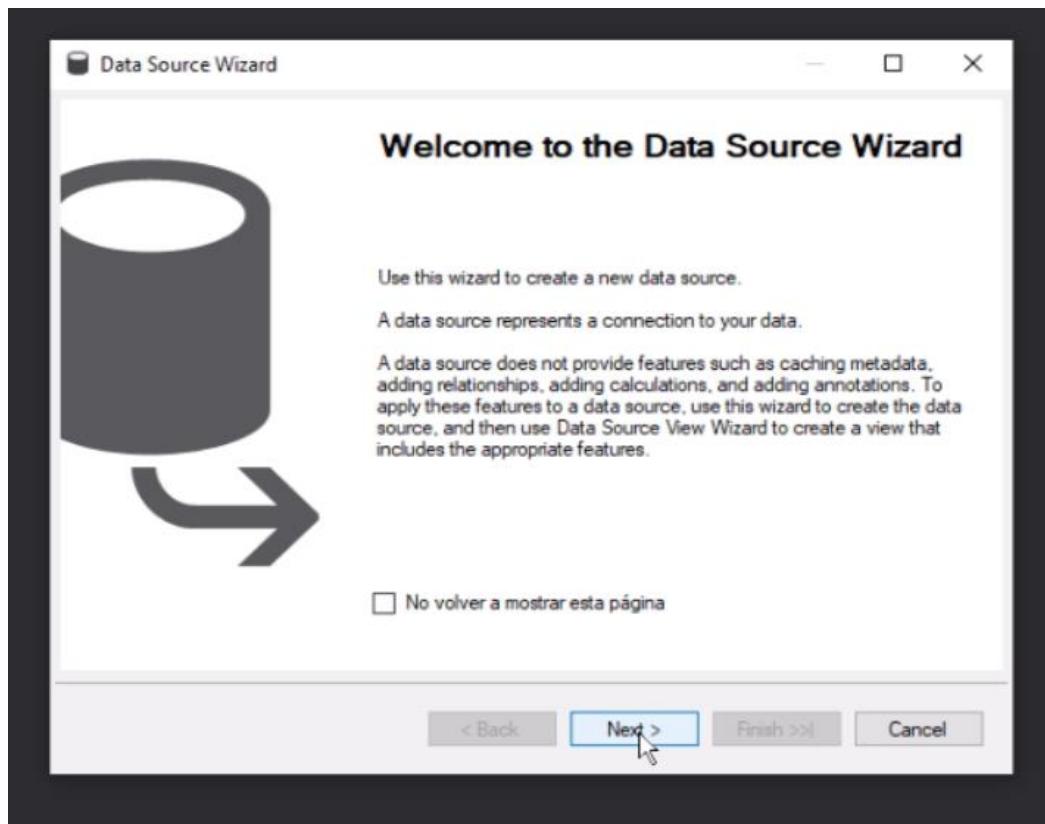


Creación del Origen de Datos.

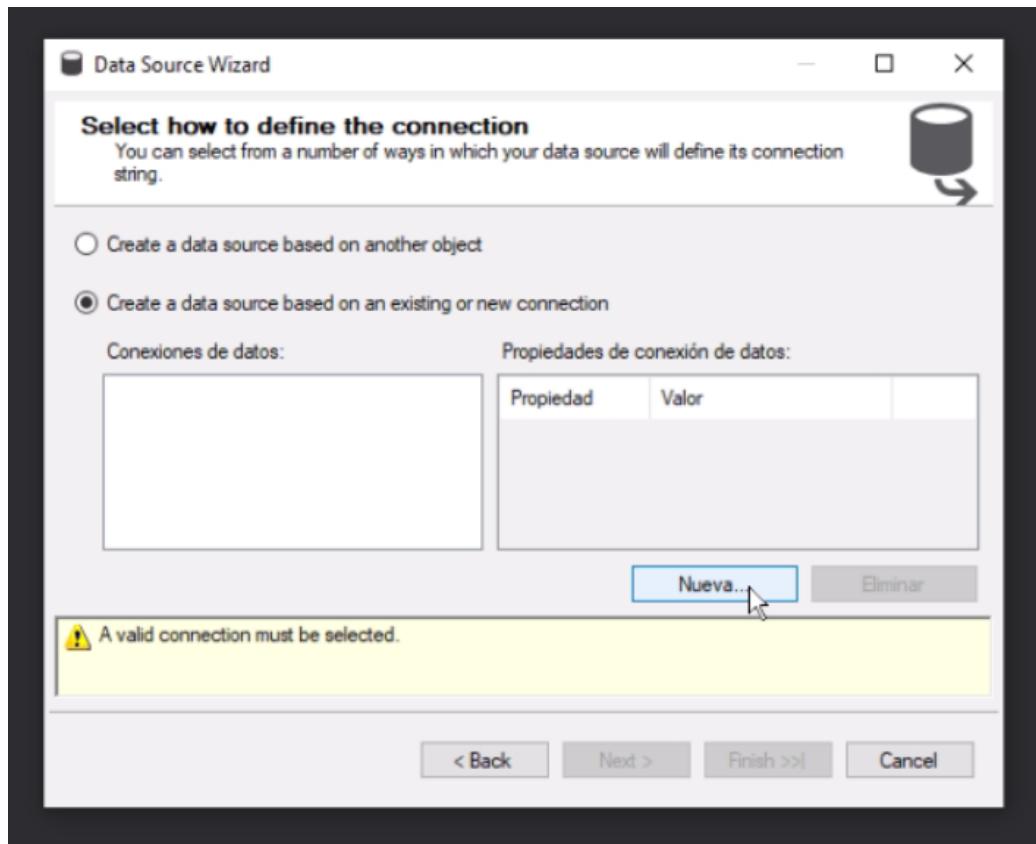
- Una vez creado, presione “Nuevo origen de datos”.



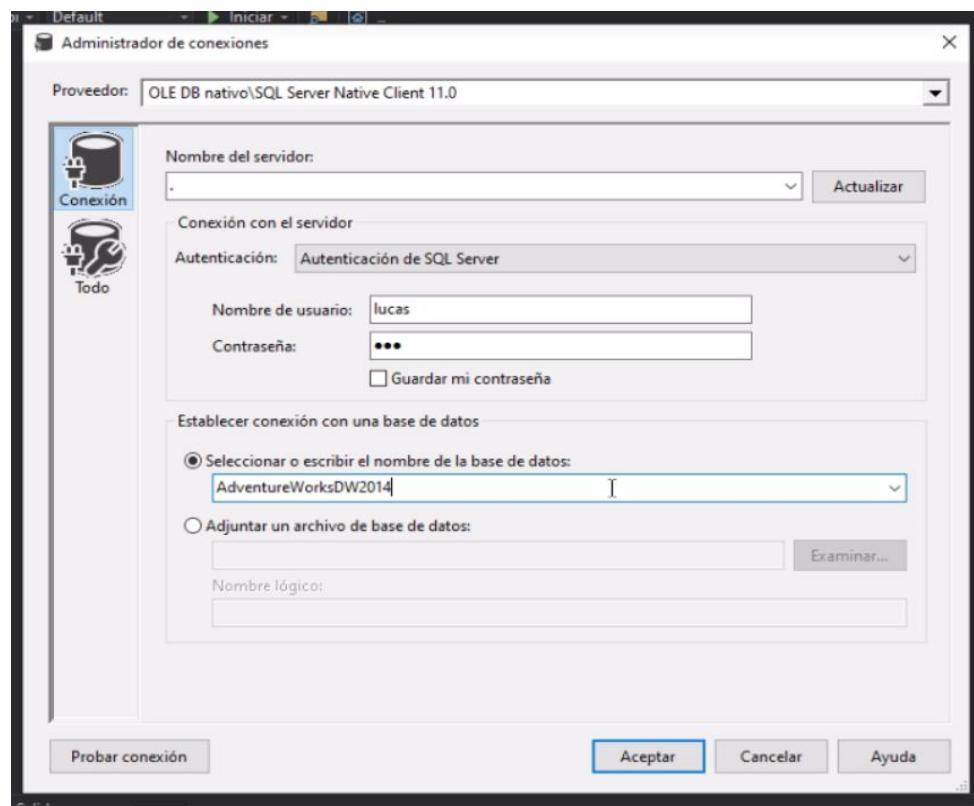
- Presionar “Siguiente” para comenzar con el asistente a crear el origen de datos.



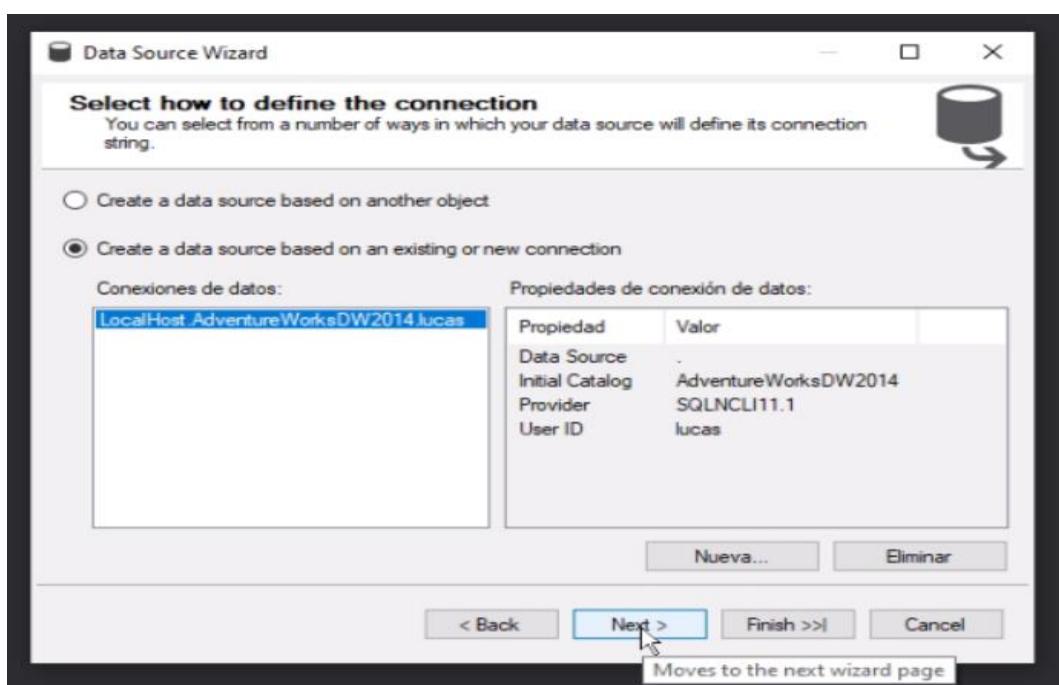
- Presionar “Nueva” para crear una nueva conexión a la base de datos.



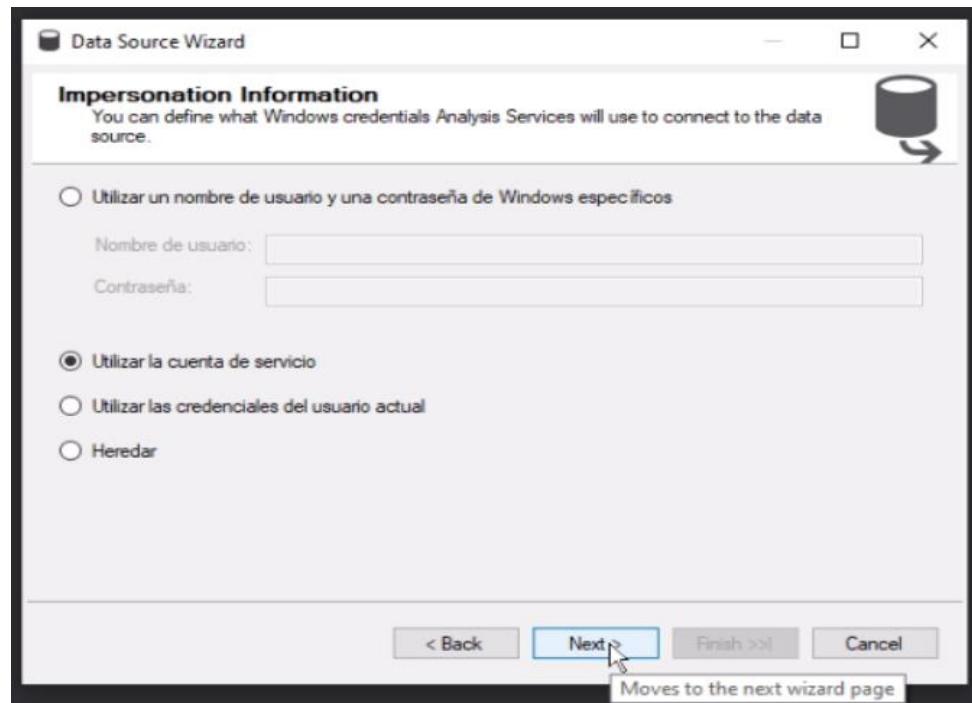
- Colocar “.” en Nombre del servidor para que nos traiga las bases de datos que tenemos en nuestro servidor. Puede ingresar también la palabra “localhost” o ingresar el nombre del servidor.
- Ingresar los datos del usuario de SQL Server para poder utilizar completamente la base de datos. Anteriormente tendría que existir un usuario con todos los permisos para utilizarla.
- Seleccionar la base de datos a utilizar y presionar “Aceptar”.



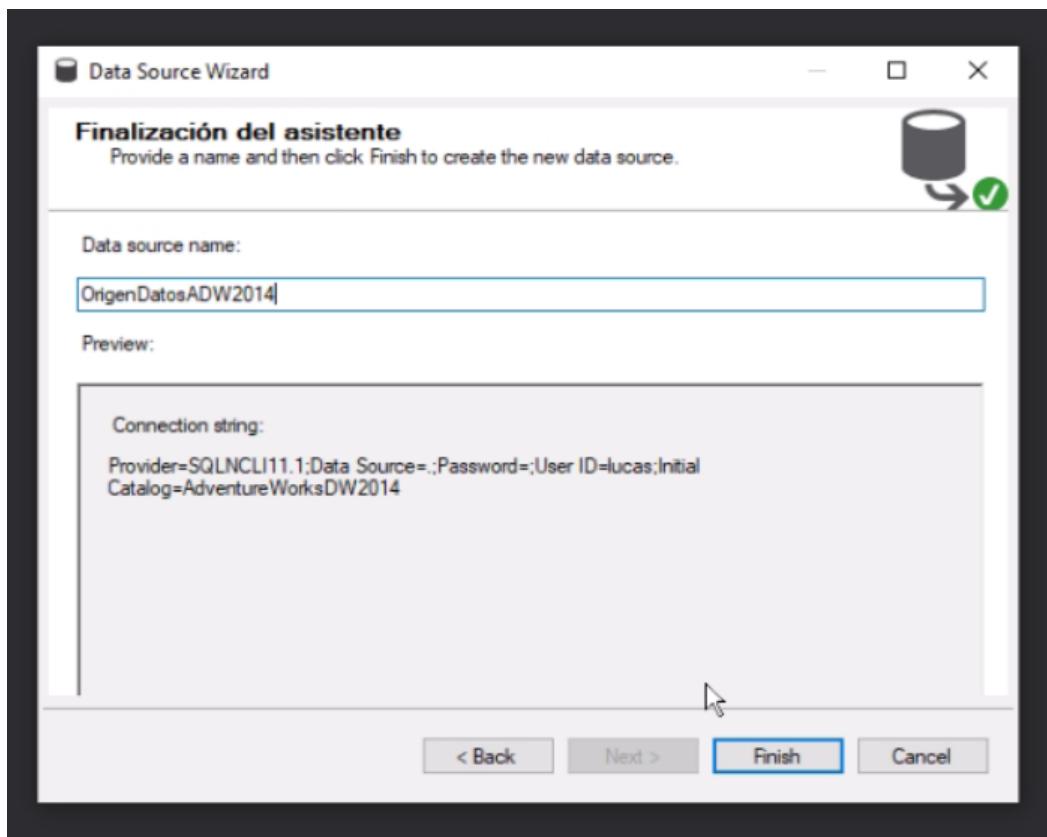
- Una vez creada la conexión con la base de datos, presionar “Next”.



- Seleccionar “Utilizar la cuenta de servicio” para que tome la cuenta de Windows directamente.

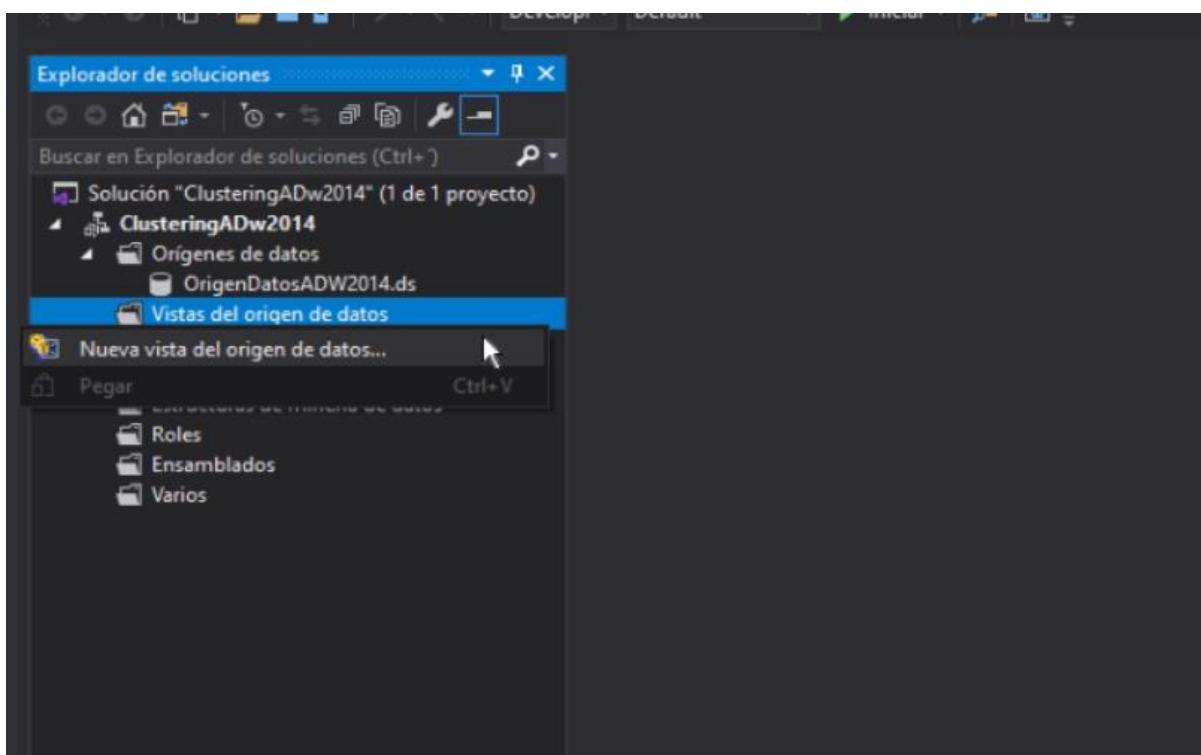


- Ingresar el nombre para el origen de datos y presionar “Finalizar”.

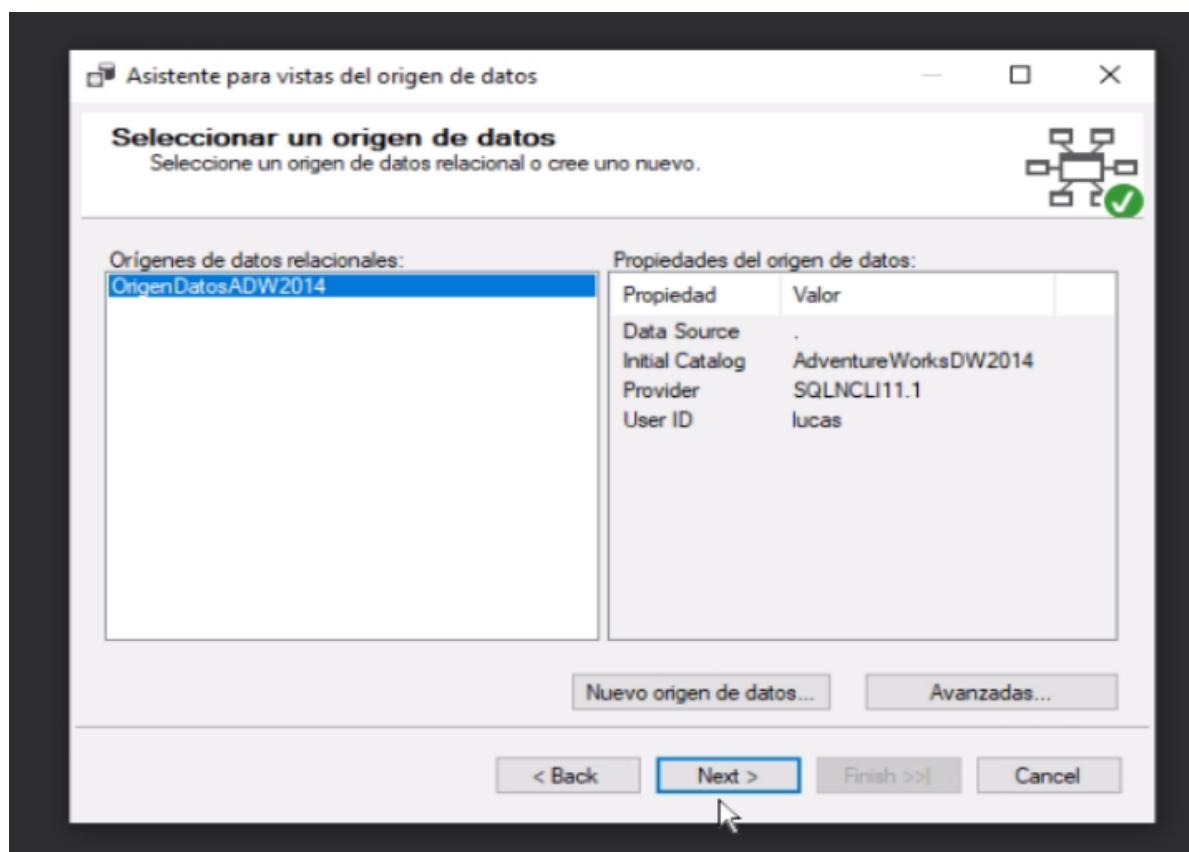


Creación de Vista del Origen de Datos.

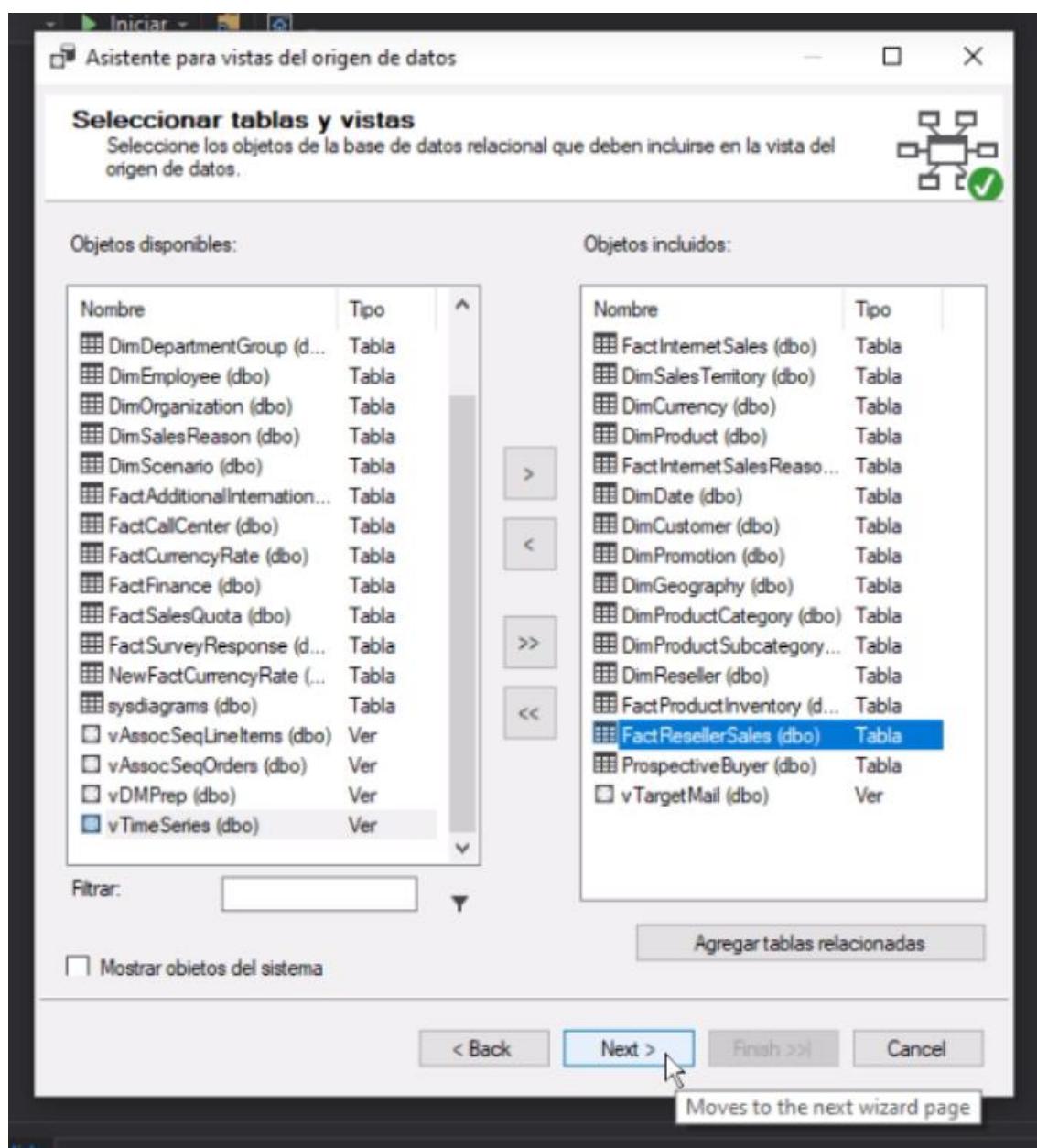
- Luego, deberá crear la vista para poder utilizar el origen de datos, para realizar esto, presionar en “Nueva vista del origen de datos”.



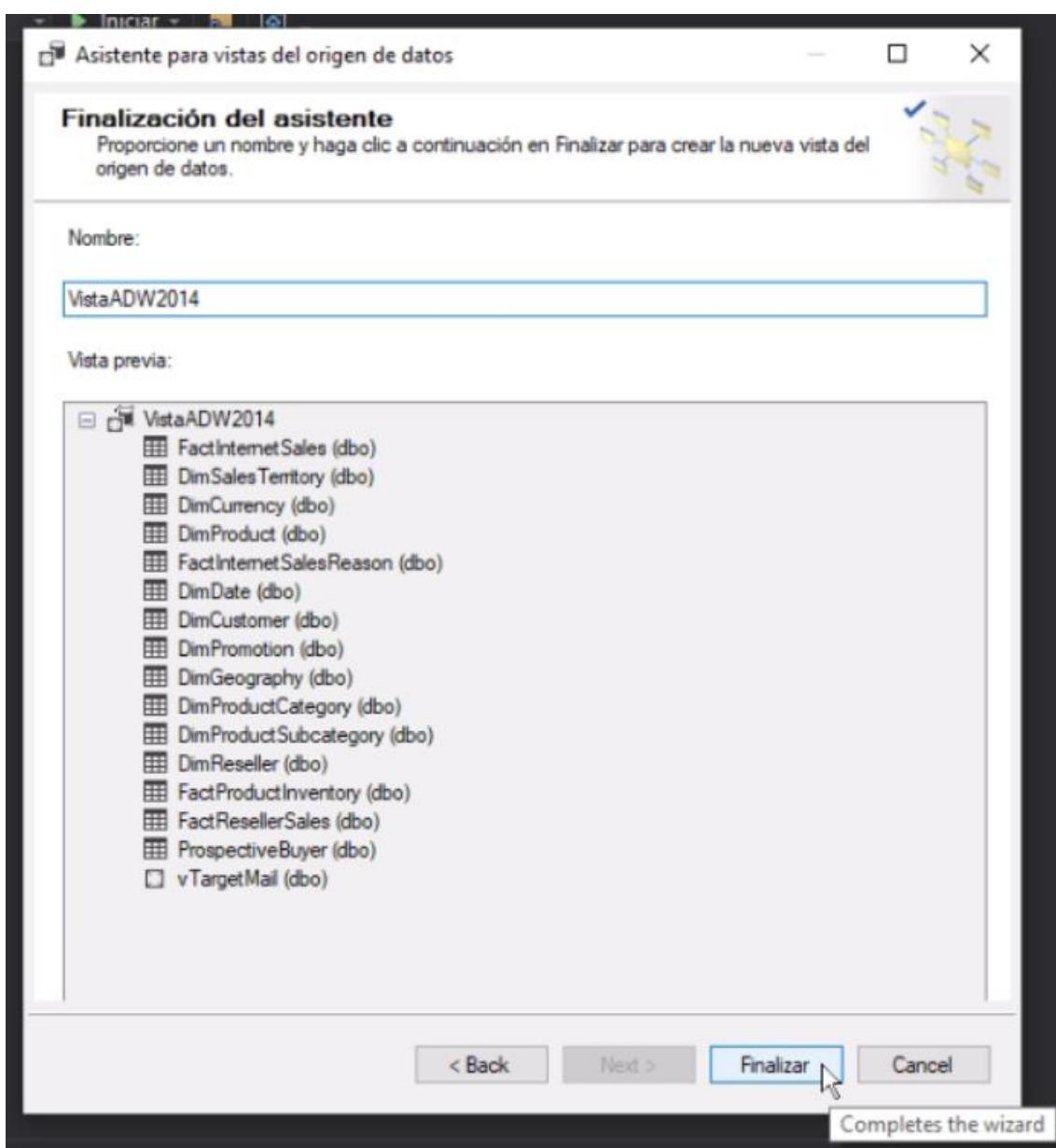
- Seleccionar el origen de datos al que le queremos crear una vista.



- Se debe incluir las tablas que se utilizaran, para este instructivo se agregan las que se visualizan en la imagen y presionar “Siguiente”.

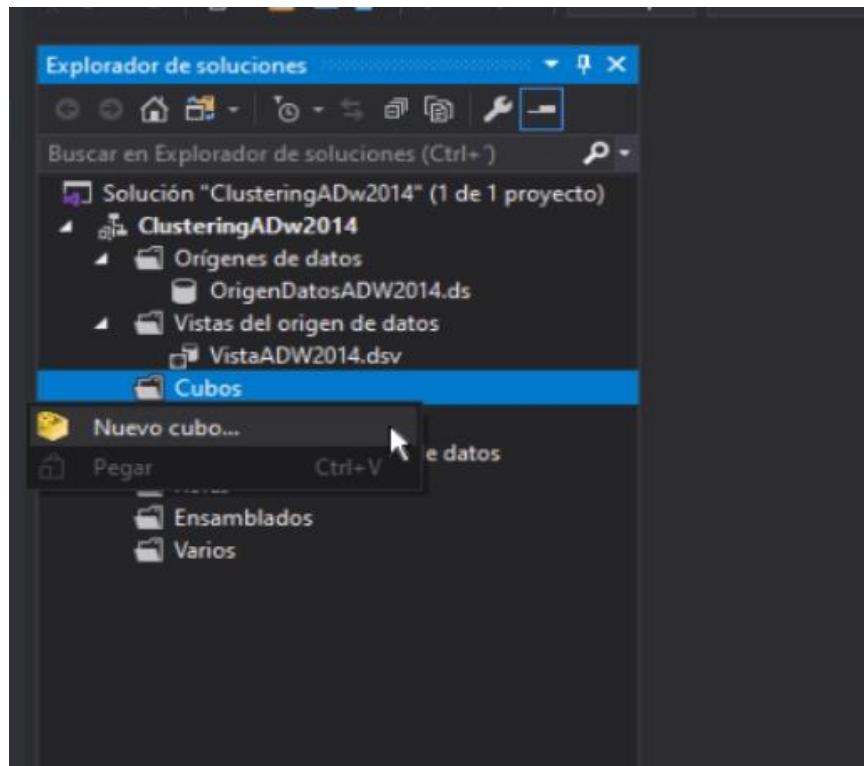


- Ingresar el nombre de la vista del origen de datos y presionar “Finalizar”.

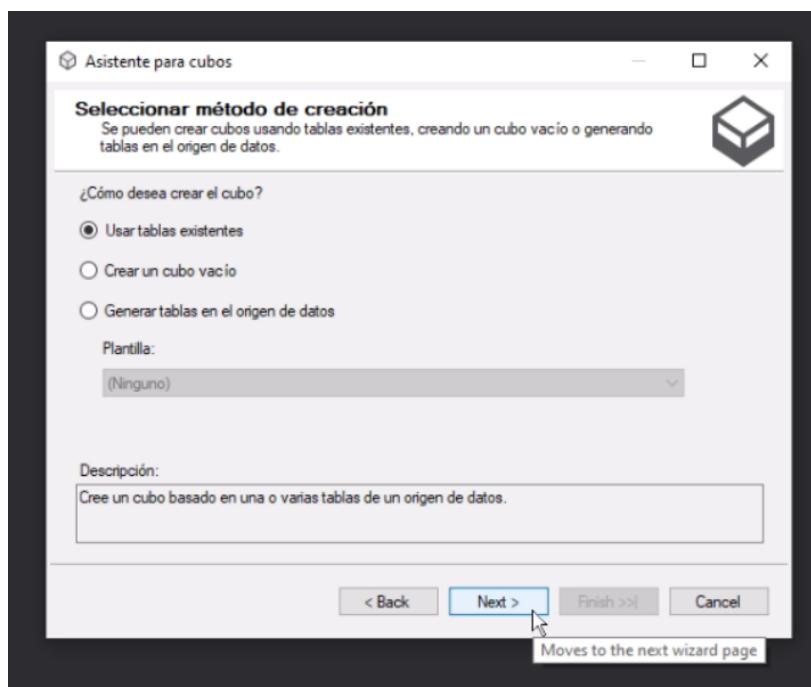


Creación del Cubo.

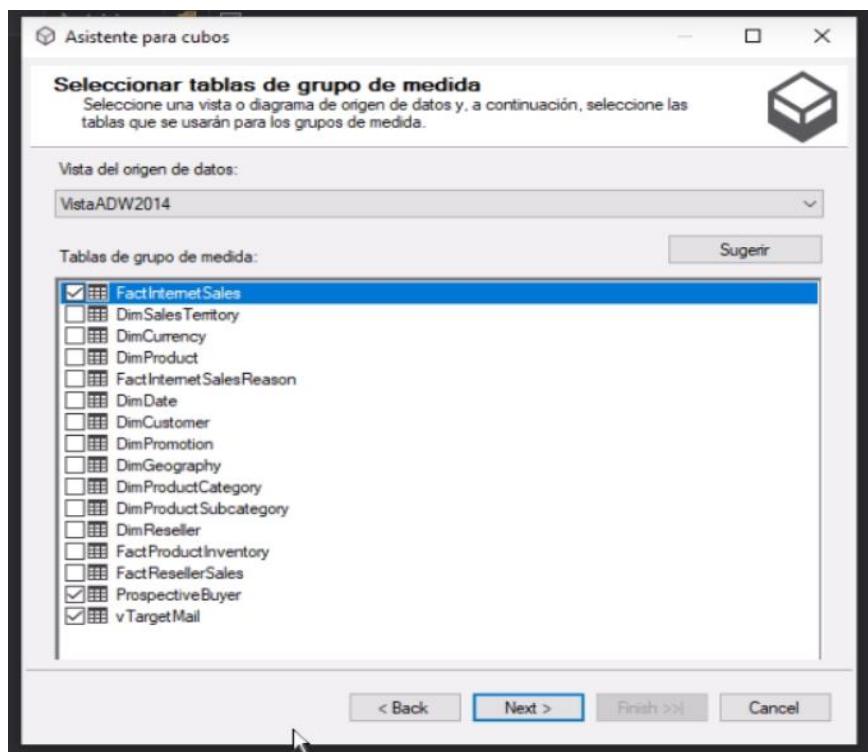
- Presionar en “Nuevo cubo”.



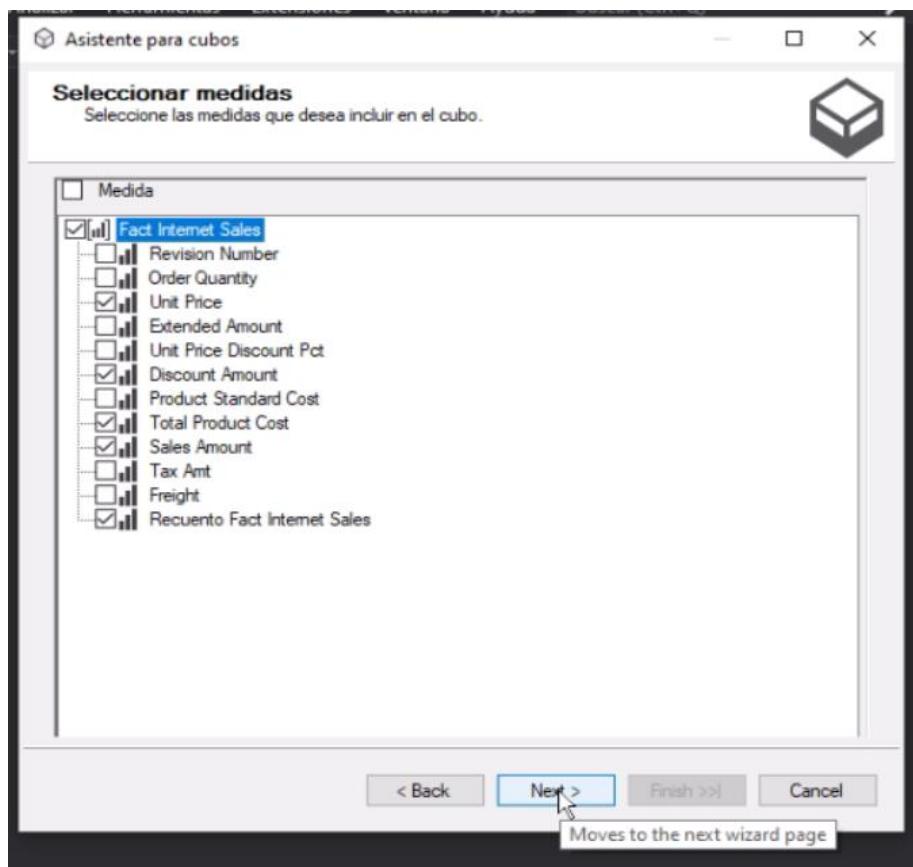
- Seleccionar “Usar tablas existentes” y presionar “Siguiente”.



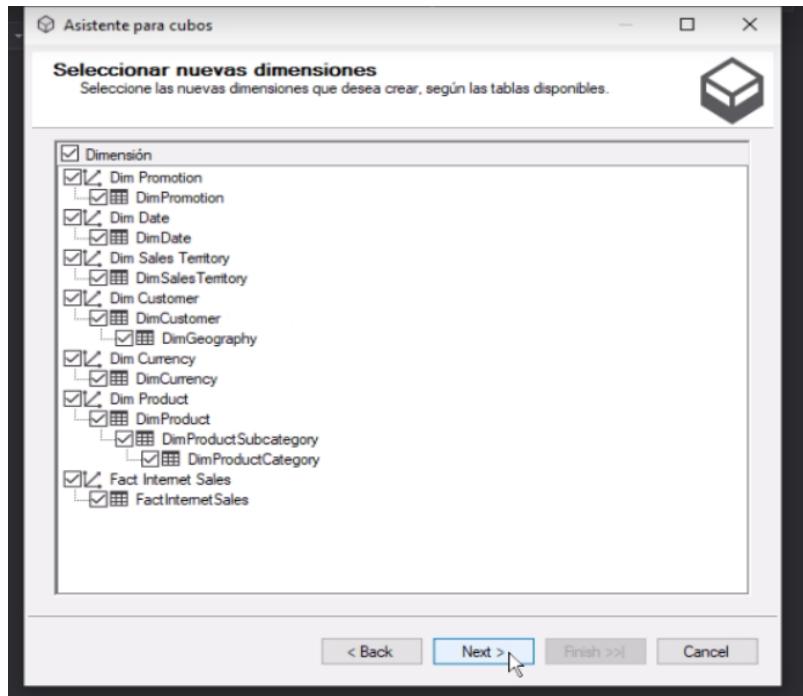
- Para crear el cubo, tildar las tablas que se visualizan en la imagen.



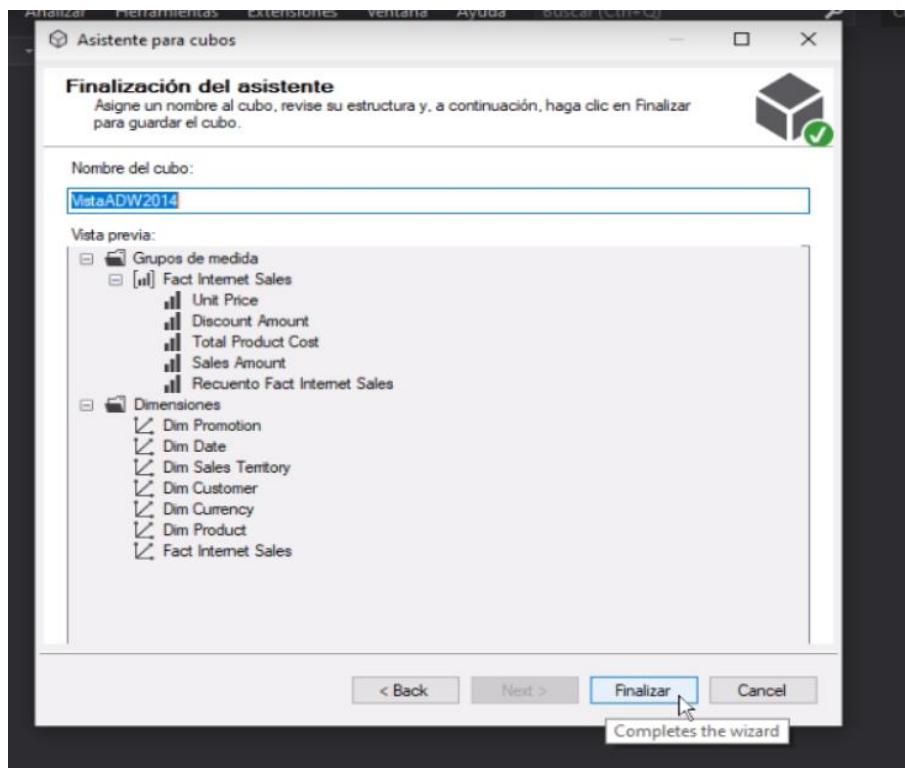
- Luego, tildar las medidas de la imagen o las que sean necesarias para el cubo y presionar "Siguiente".



- Seleccionar las dimensiones deseadas para el cubo y presionar “Next”.



- Ingresar el nombre del cubo, verificamos que las medidas y las dimensiones son las seleccionadas y presionar “Finalizar”.

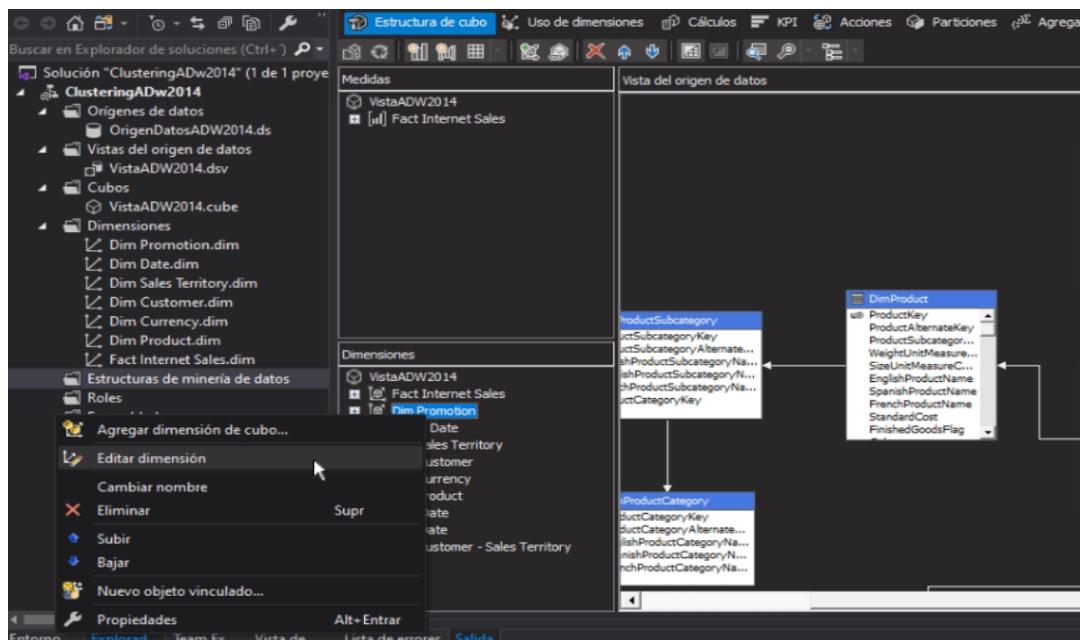


Conclusión de la creación de cubos.

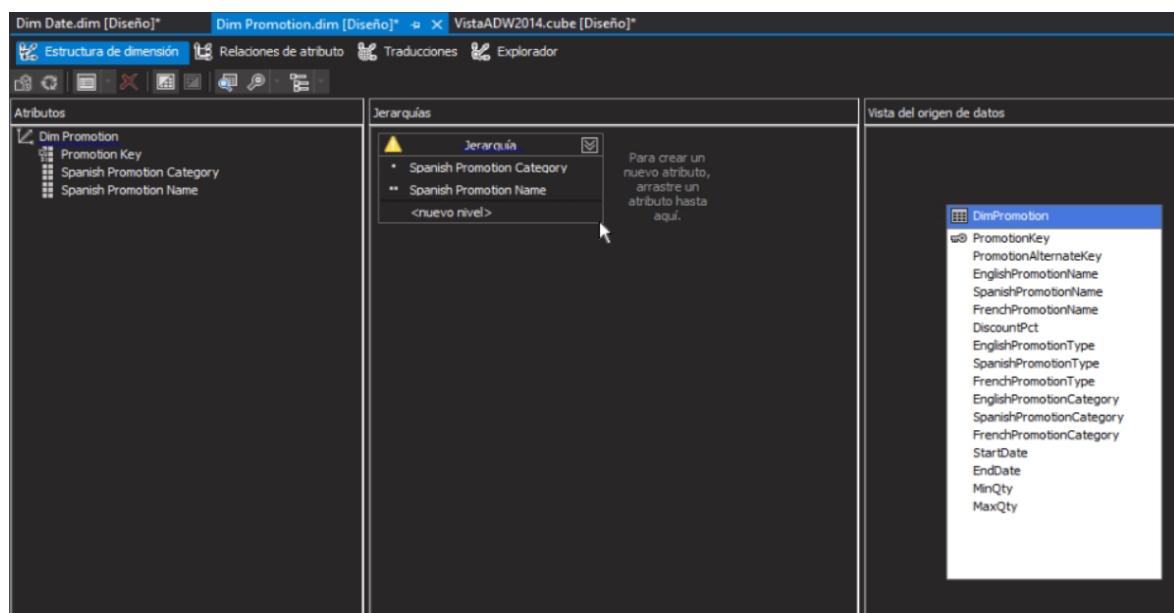
Podemos decir que un Cubo almacena datos de negocio complejos en una estructura multidimensional por la que se puede navegar más fácilmente. Es una herramienta muy útil para analizar el negocio desde un punto de vista gerencial por su fácil utilización. Además, mejora la rapidez en el almacenamiento de datos, optimiza notablemente el rendimiento de la memoria caché y la extracción de datos de este.

Modificación de Dimensiones y establecimiento de Jerarquías.

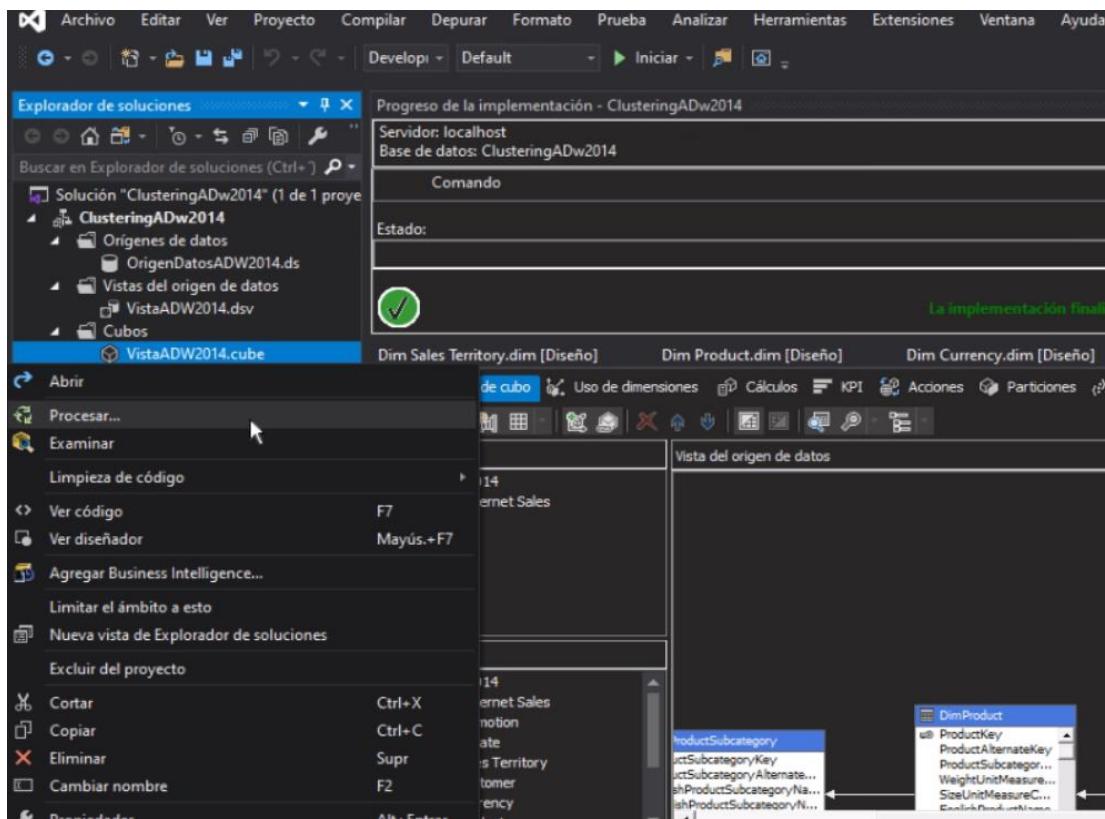
- Una vez generado el cubo, seleccionar la dimensión a modificar, en este caso es “Dim Promotion”.



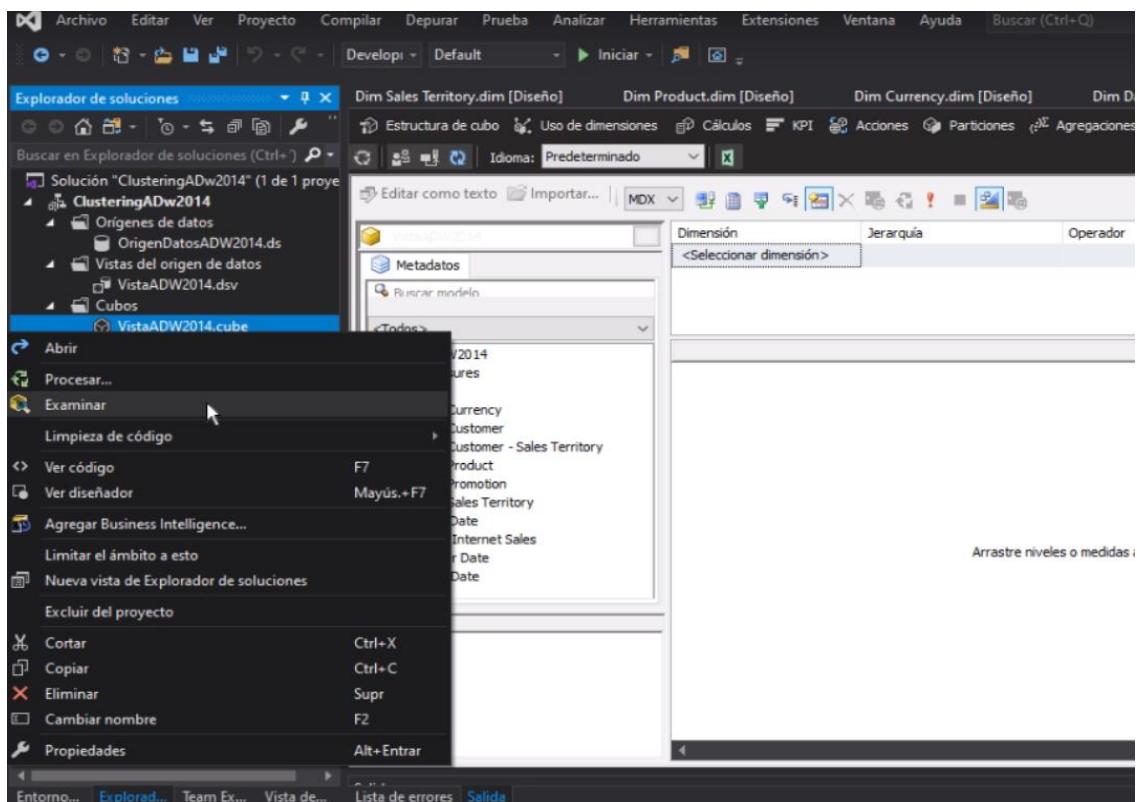
- Para realizar las jerarquías en cómo se mostrarán los datos, arrastrar las columnas de la tabla hacia la pestaña “Atributos” y luego a la pestaña de “Jerarquía” en el orden deseado, ordenando del más genérico a el más específico.



- Posteriormente, debemos procesar el cubo para que se efectúen los cambios.



- Cuando finalice el procesamiento del cubo, presionar en “Examinar” para poder realizar las consultas correspondientes utilizando las jerarquías.



- Para utilizar la jerarquía, arrastrar la jerarquía generada y el Unit Price o la medida a consultar.
- En caso de que no aparezca la grilla, presionar el botón que se encuentra destacado.

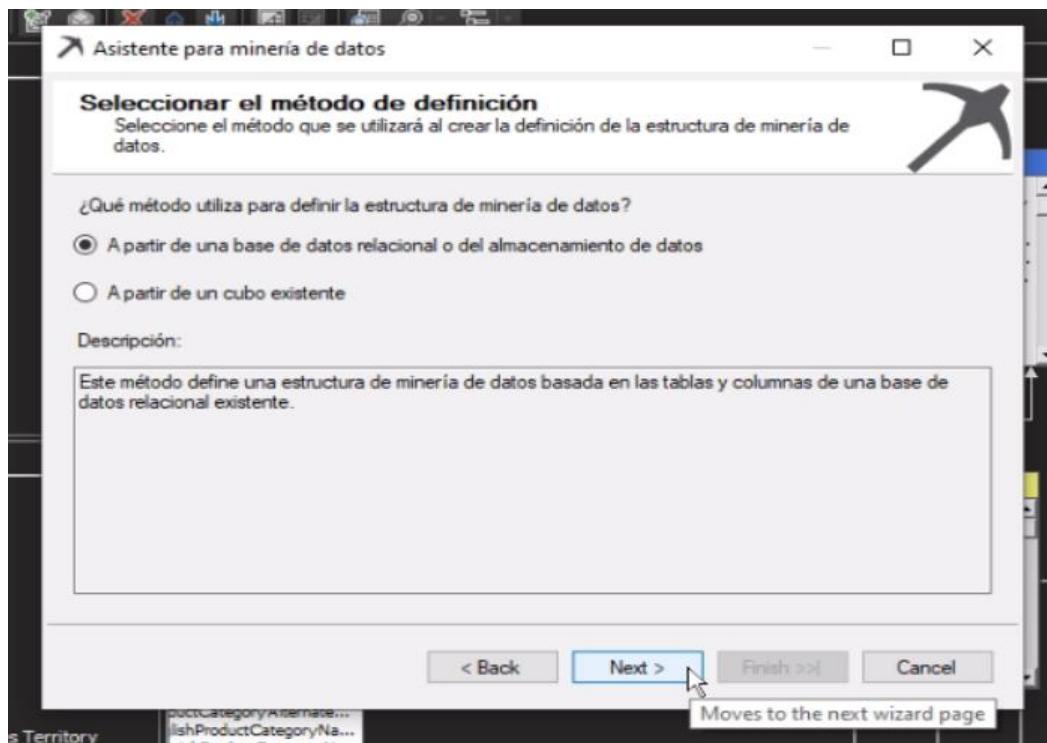
The screenshot shows the Analysis Services Designer application. The top menu bar includes tabs for 'Dim Sales Territory.dim [Diseño]', 'Dim Product.dim [Diseño]', 'Dim Currency.dim [Diseño]', 'Dim Date.dim [Diseño]', and 'VistaADW2014.cube [Diseño]'. The toolbar has various icons, with the second one from the left highlighted in red. The main area contains a table with four columns: 'Dimensión', 'Jerarquía', 'Operador', and 'Expresión de filtro'. The table shows data from the 'Spanish Promotion Category' dimension, including rows for 'Distribuidor' and 'Sin descuento' with their corresponding promotion names and unit prices. On the left, there's a sidebar with 'Metadatos' and a tree view of the cube structure, including dimensions like 'Dim Sales Territory' and 'Dim Product'.

Implementación de la técnica “Clustering”

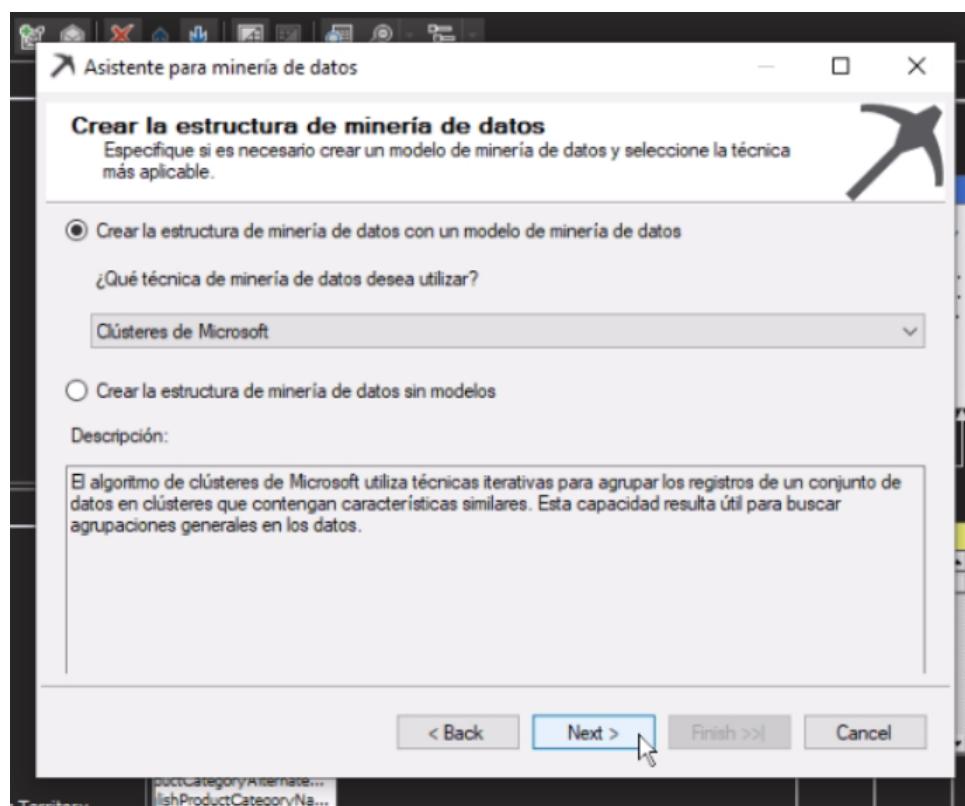
- Nuestro objetivo es agrupar por la cantidad de autos que poseen los clientes. Para implementar la técnica, debemos crear una “Nueva estructura de minería de datos”.

The screenshot shows the Data Mining Designer application. The top menu bar includes 'Archivo', 'Editar', 'Ver', 'Proyecto', 'Compilar', 'Depurar', 'Prueba', 'Analizar', 'Herramientas', 'Extensiones', 'Ventana', 'Ayuda', and 'Buscar (Ctrl+Q)'. The toolbar has various icons. The 'Explorador de soluciones' (Solution Explorer) on the left shows a project named 'Solución "ClusteringADw2014"' containing a 'ClusteringADw2014' folder with 'Orígenes de datos', 'Vistas del origen de datos', 'Cubos', 'Dimensiones', and 'Medidas'. The 'Fact Internet Sales.dim [Diseño]' tab is selected in the center. The right pane shows the 'Vista del origen de datos' (Data Source View) with a diagram of data flows between 'Fact Internet Sales' and dimensions like 'DimProduct', 'DimSubcategory', and 'DimCategory'.

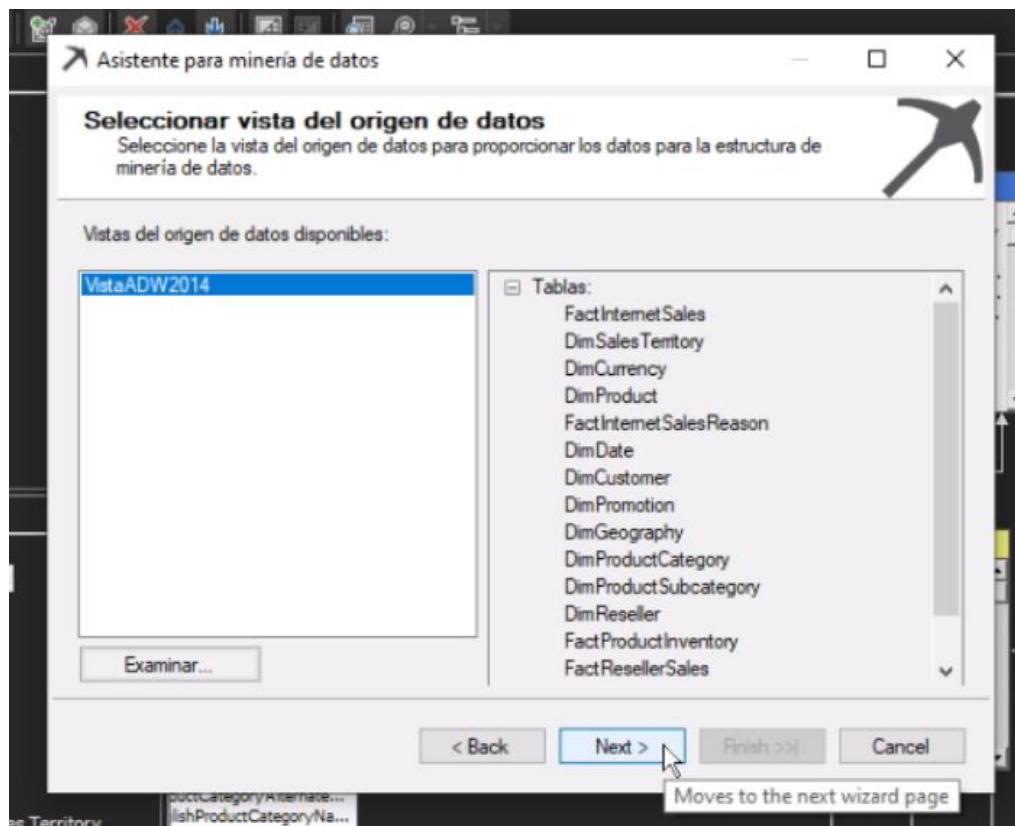
- Seleccionar “A partir de una base de datos relacional o del almacenamiento de datos” en caso de que la tabla a utilizar no se encuentre dentro del cubo o que lo que se desea utilizar sea una vista del origen de datos.



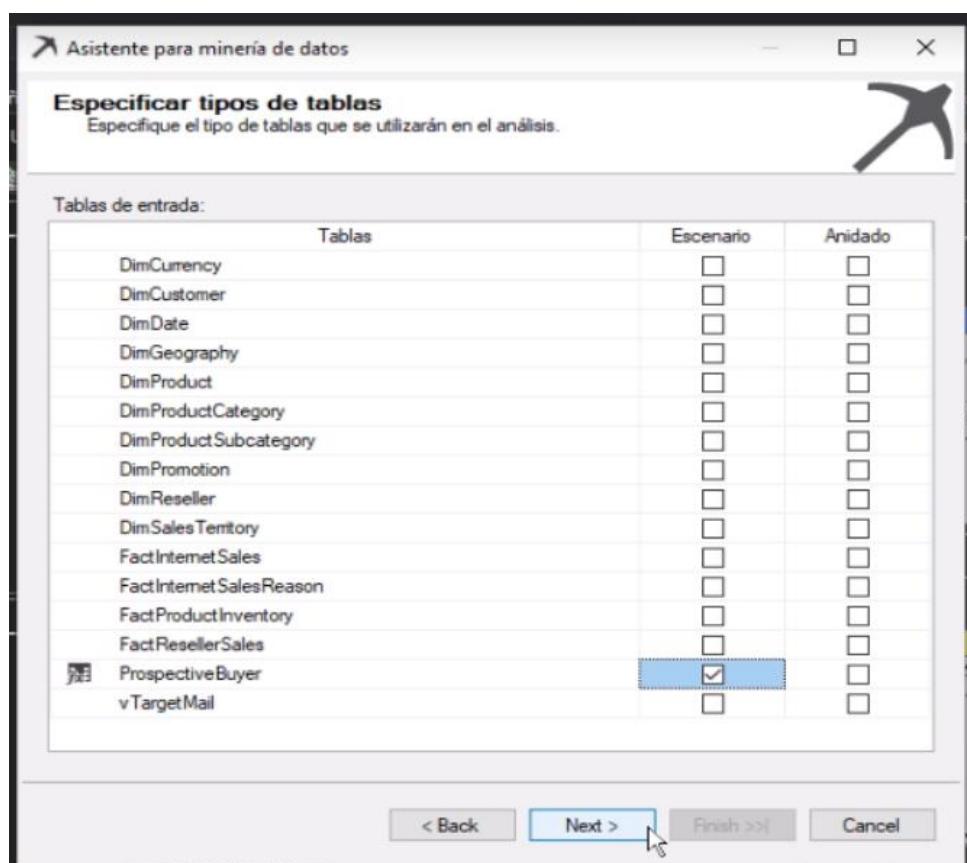
- Seleccionar la técnica de minería de datos a realizar, en este caso, “Clústeres de Microsoft” y continuar.



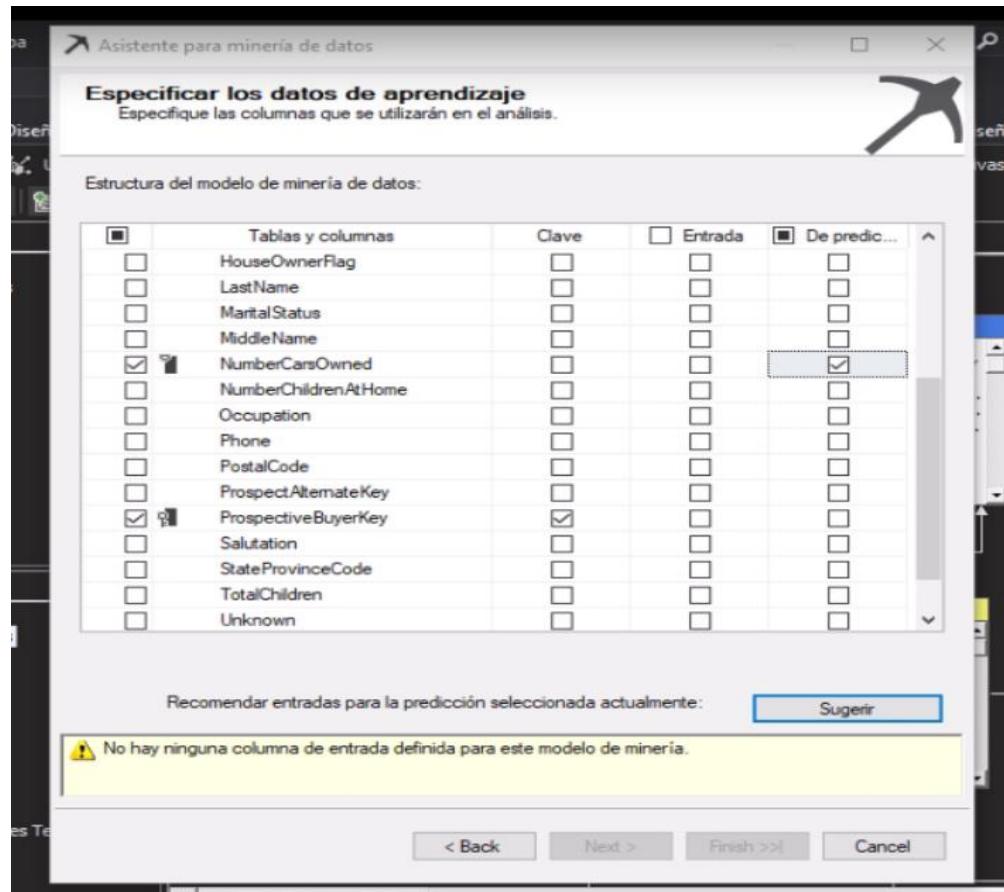
- Seleccionar la vista deseada o el cubo, sus tablas y el criterio en caso de que sea necesario.



- Tildar el escenario o la tabla que será nuestro punto a enfocar.



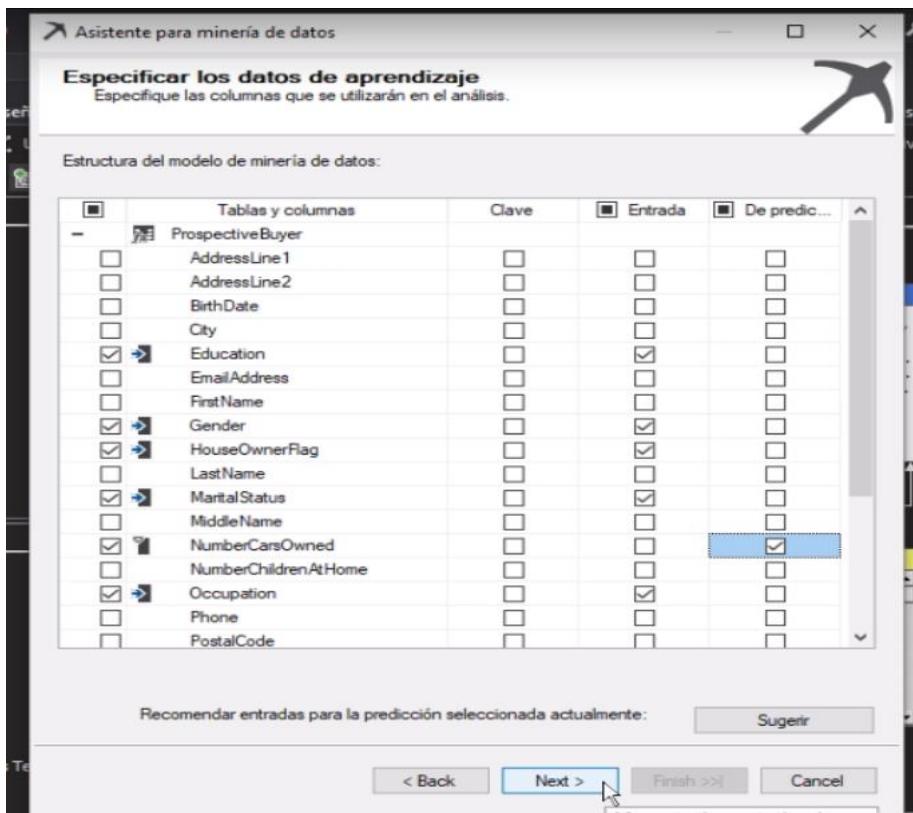
- Seleccionar el campo de predicción que queremos estimar y luego presionamos el botón “Sugerir” para que nos traiga los campos que ayudarían a facilitar la predicción. En este ejemplo, la predicción será la cantidad de autos que posee la media de la población.



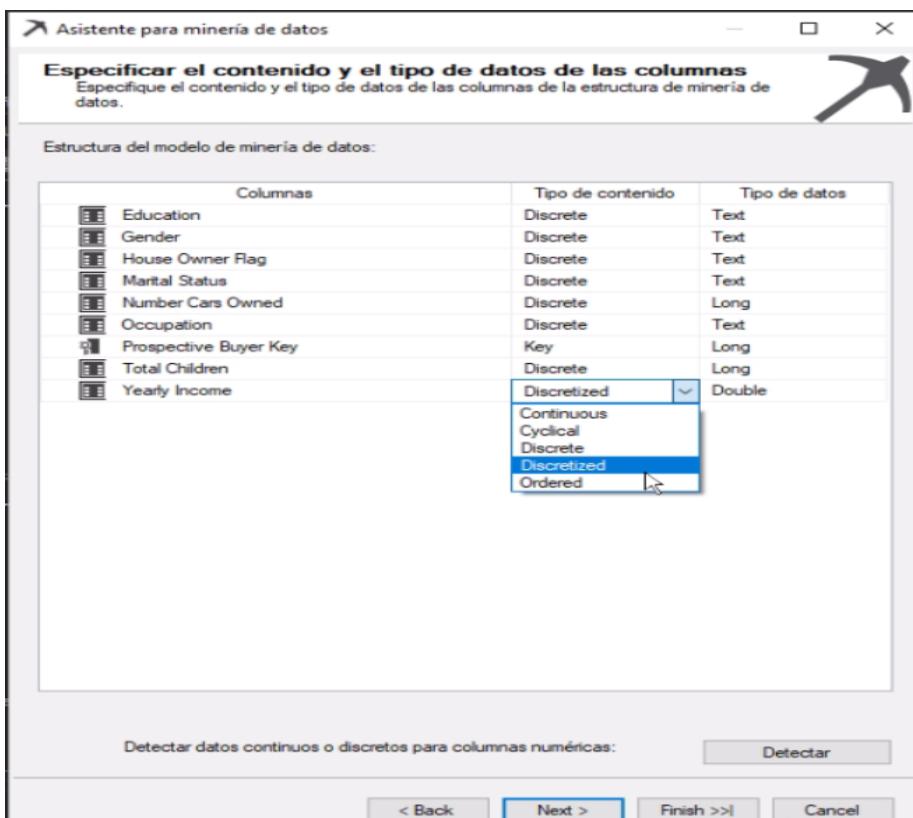
- Marcamos los datos de entrada que consideremos óptimos para la predicción.

Nombre de columna	Puntuación	Entrada
YearlyIncome	0,144	x
Education	0,120	x
TotalChildren	0,115	x
NumberChildrenAtHome	0,103	x
Occupation	0,088	x
StateProvinceCode	0,053	
AddressLine2	0,026	
MiddleName	0,012	
Unknown	0,005	
MaritalStatus	0,001	x
HouseOwnerFlag	0,001	x
Salutation	0,001	
Gender	0,001	x
Phone		
PostalCode		
City		
AddressLine1		
EmailAddress		

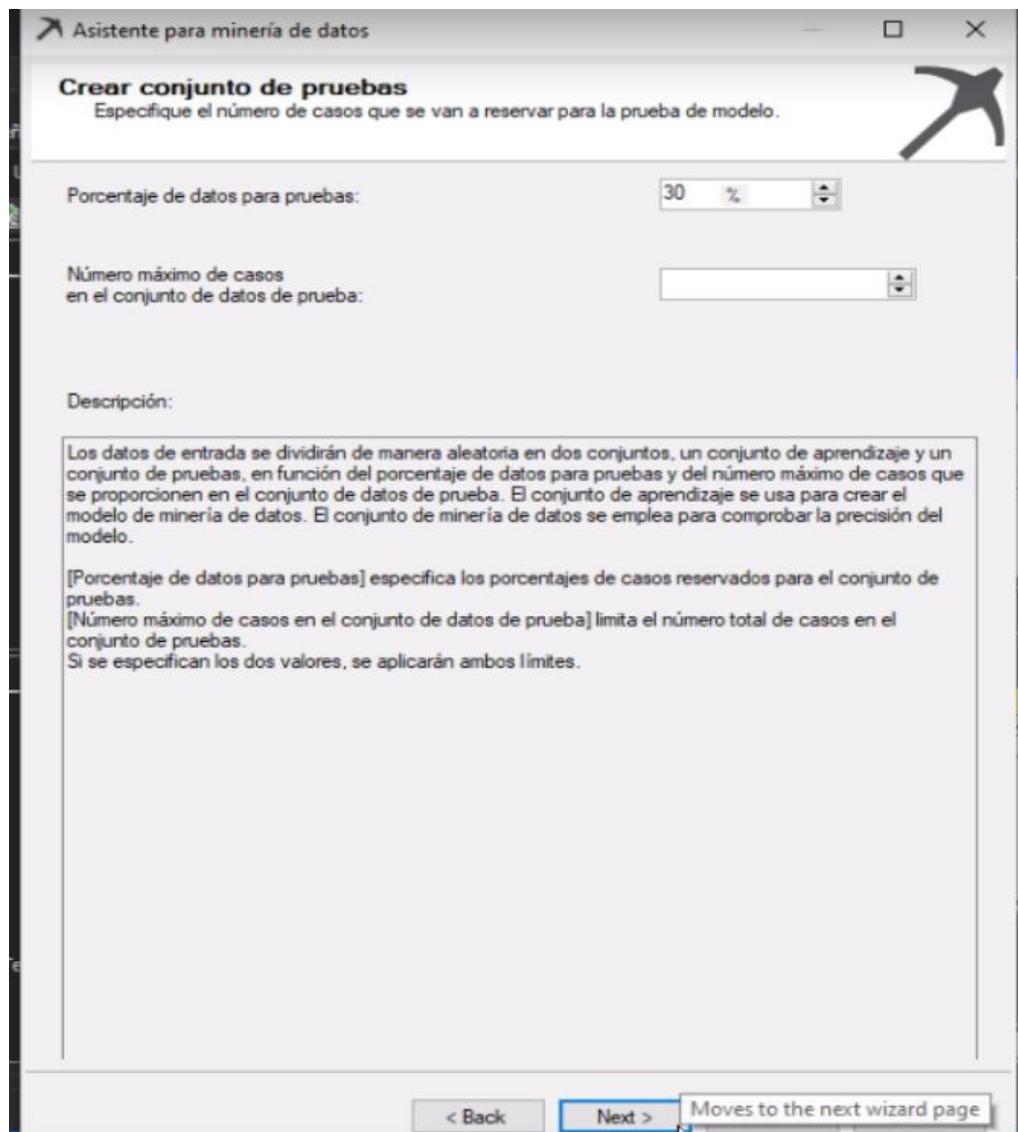
- Visualizar los datos seleccionados como entrada de la sugerencia y presionar “Next”.



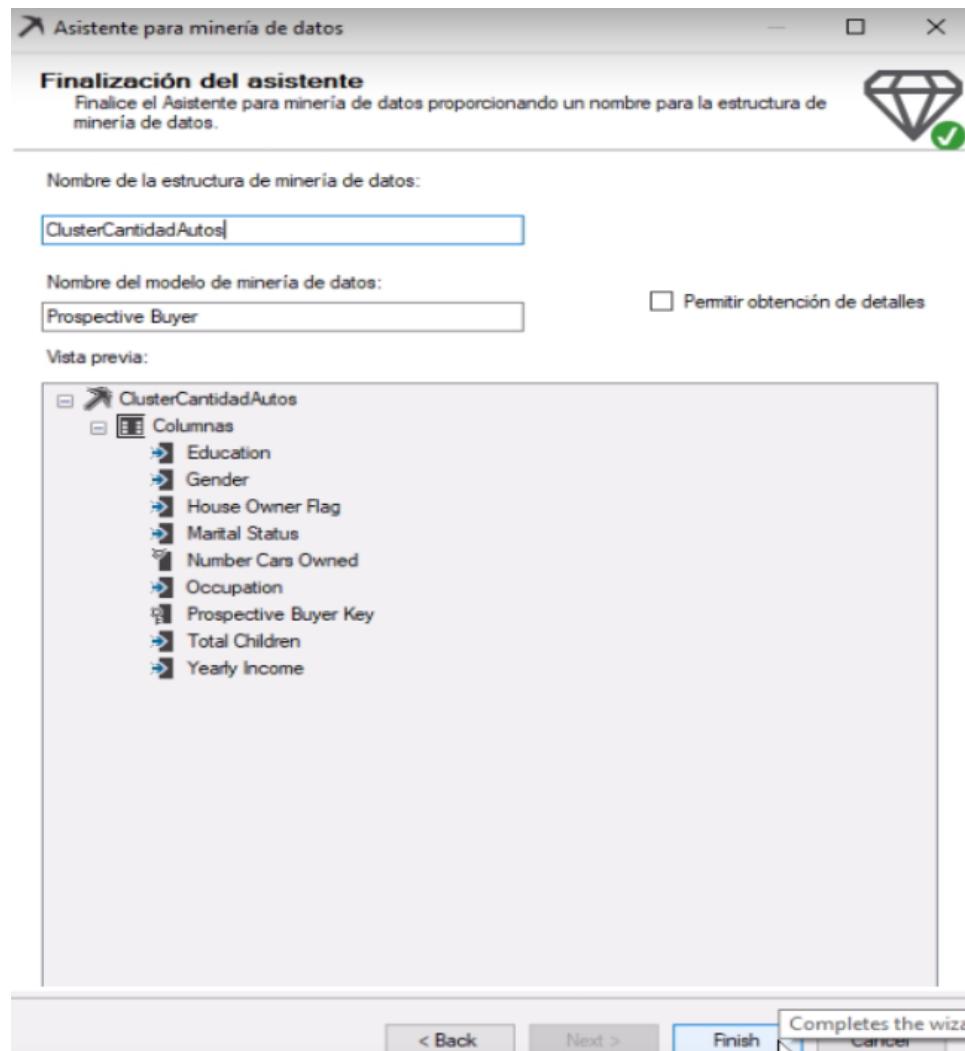
- Pueden aparecer variables continuas, a estas debemos cambiarle al tipo Discrete a los valores que sean enteros y Discretized a los que pueden ser decimales, para que la predicción sea lo más exacta posible.



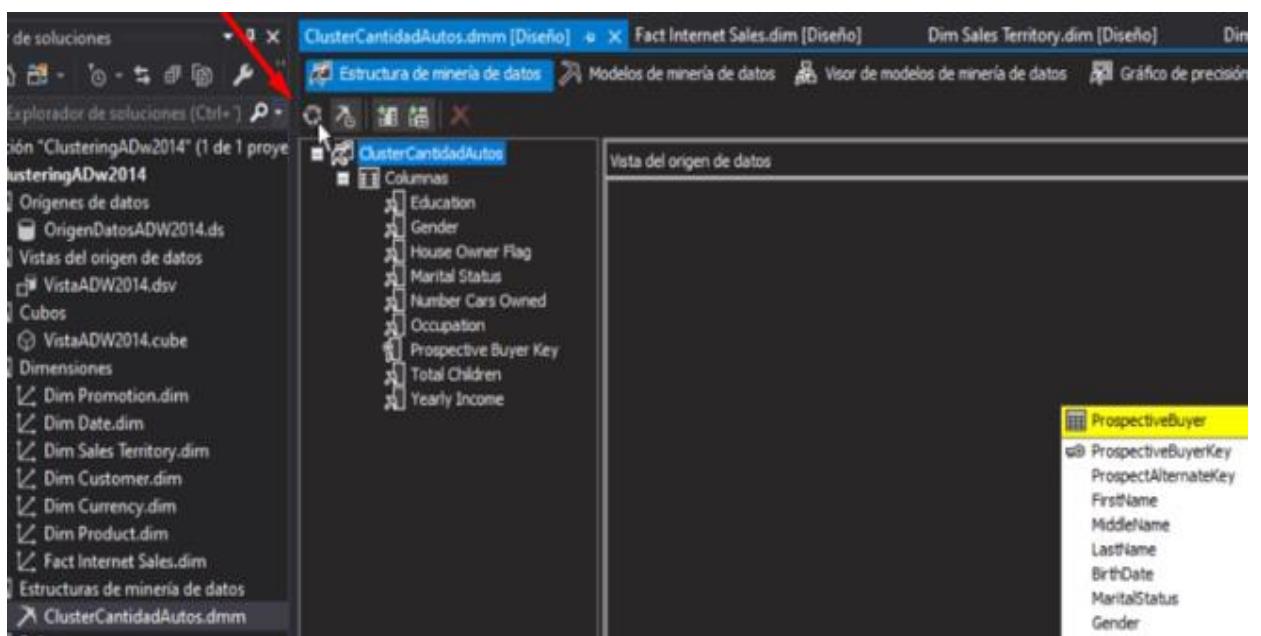
- Dejar el porcentaje de datos por defecto, este porcentaje muestra lo que representa la muestra de la cantidad de datos totales. Puede ser un porcentaje o una cantidad.



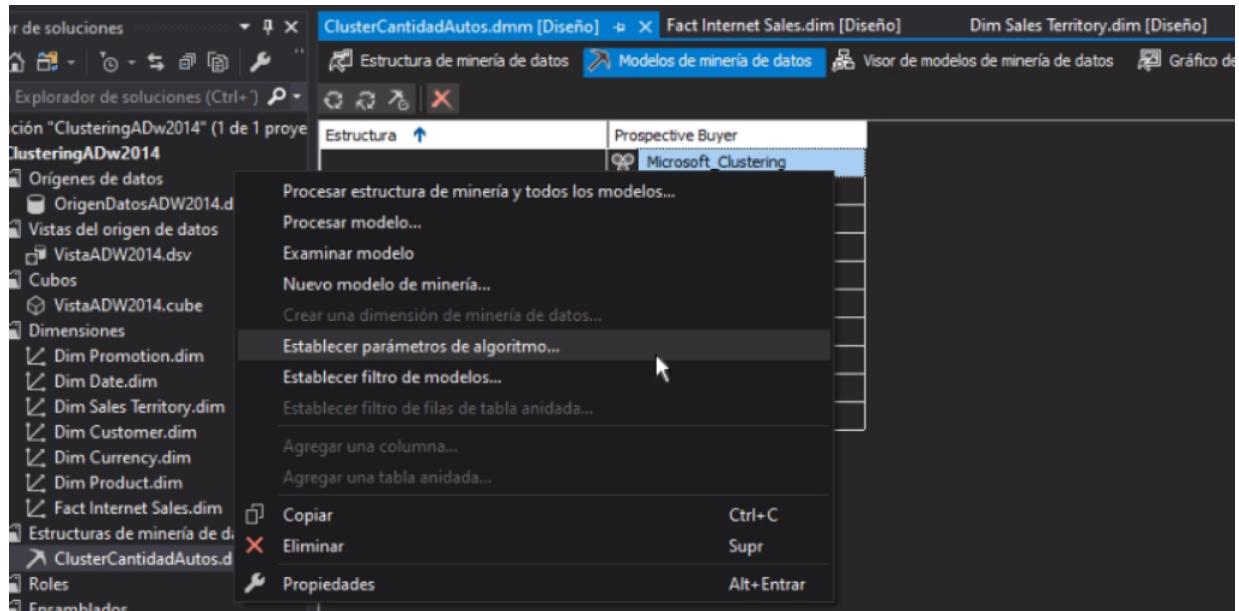
- Ingrese el nombre deseado y finalizar.



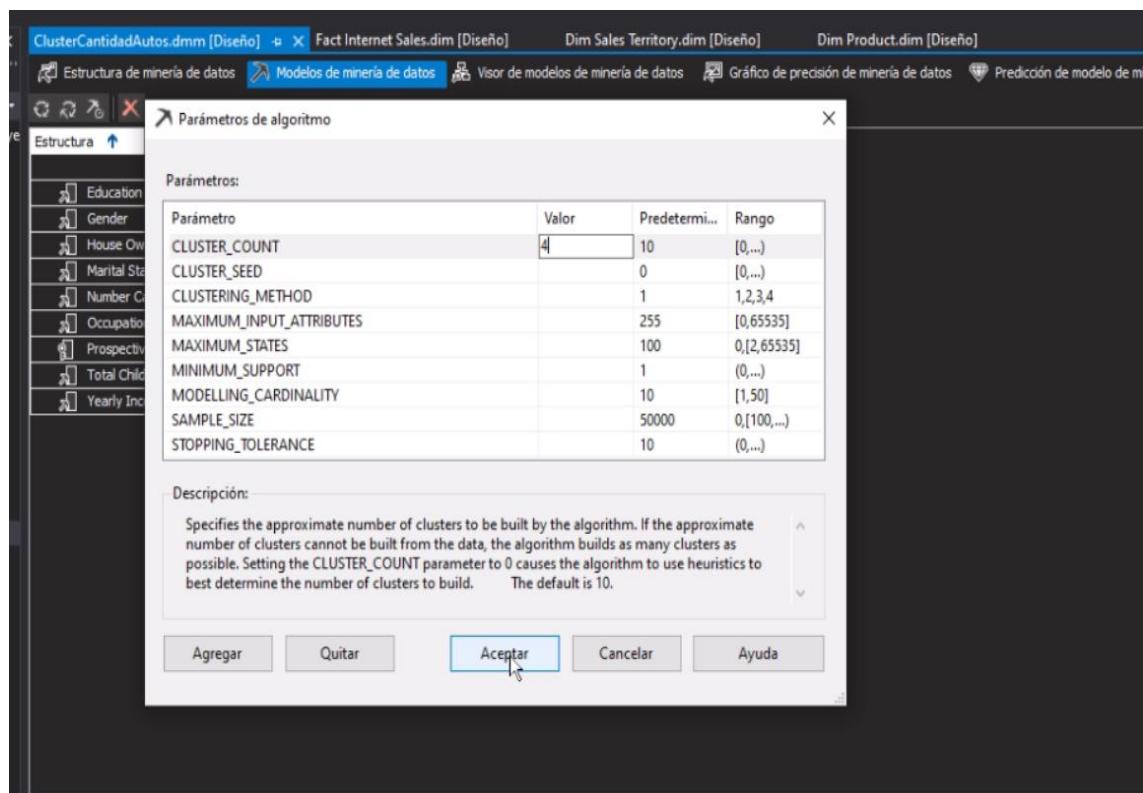
- Para poder utilizar la minería de datos creada, deberán procesarla y para eso deberán presionar el botón indicado en la imagen.



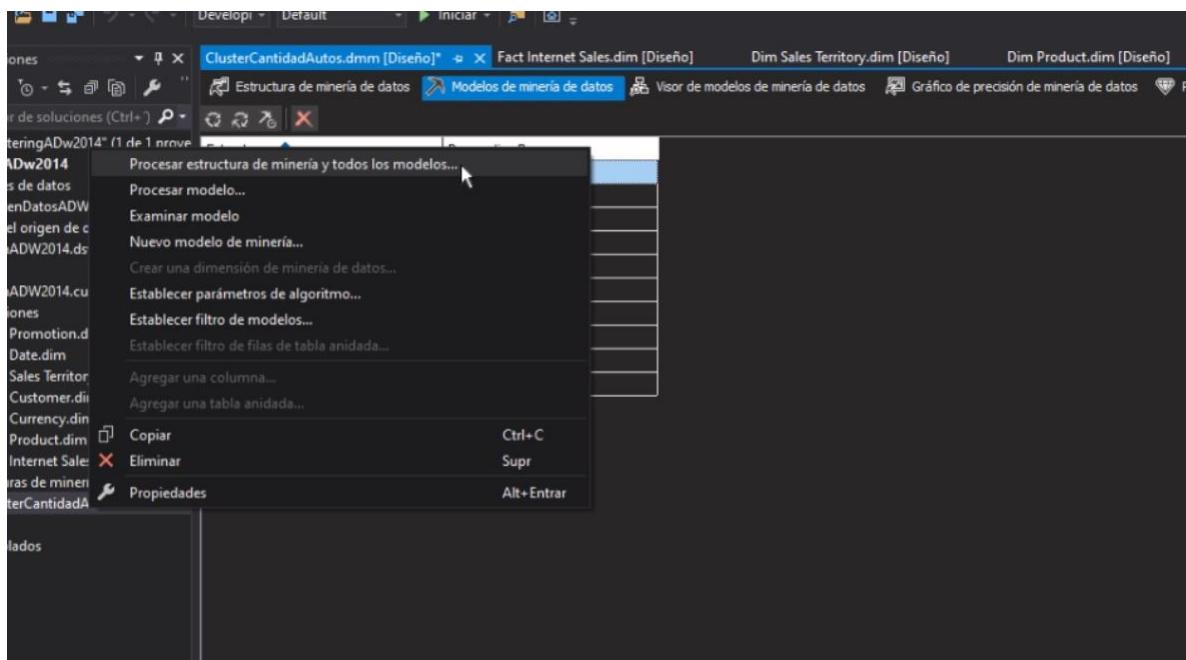
- En la pestaña de Modelos de minería de datos seleccionar con el click derecho Microsoft_Clustering y luego en Establecer parámetros de algoritmo para definir la cantidad de grupos que queremos obtener de la población.



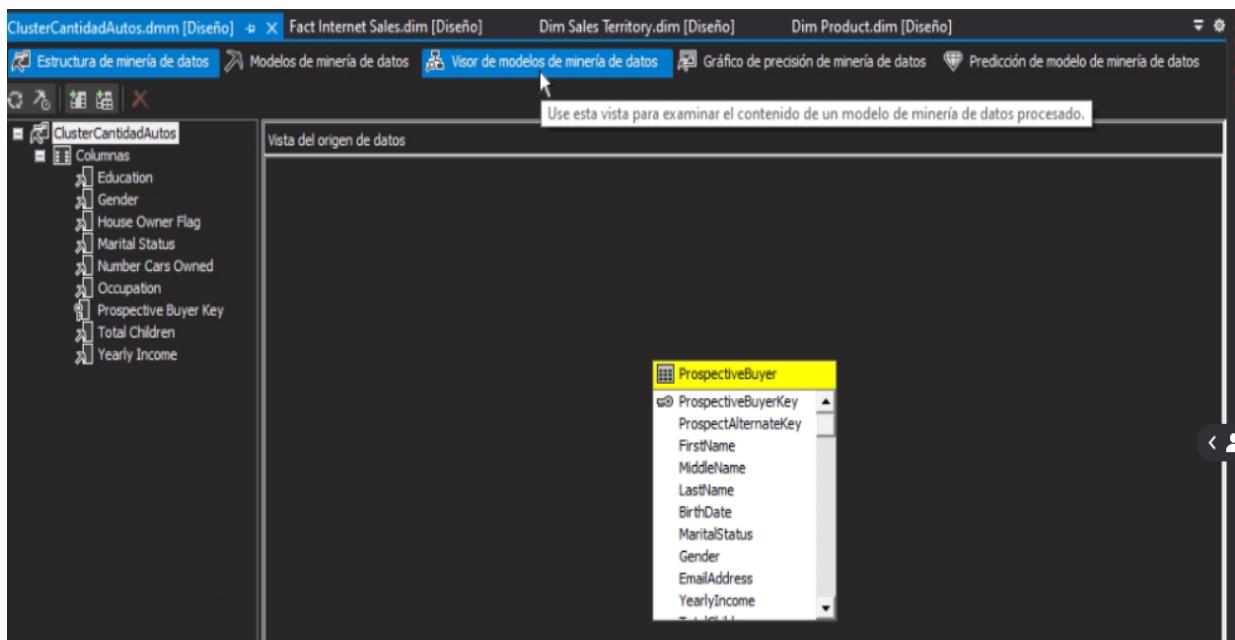
- A continuación, colocar la cantidad de clústeres que queremos visualizar.



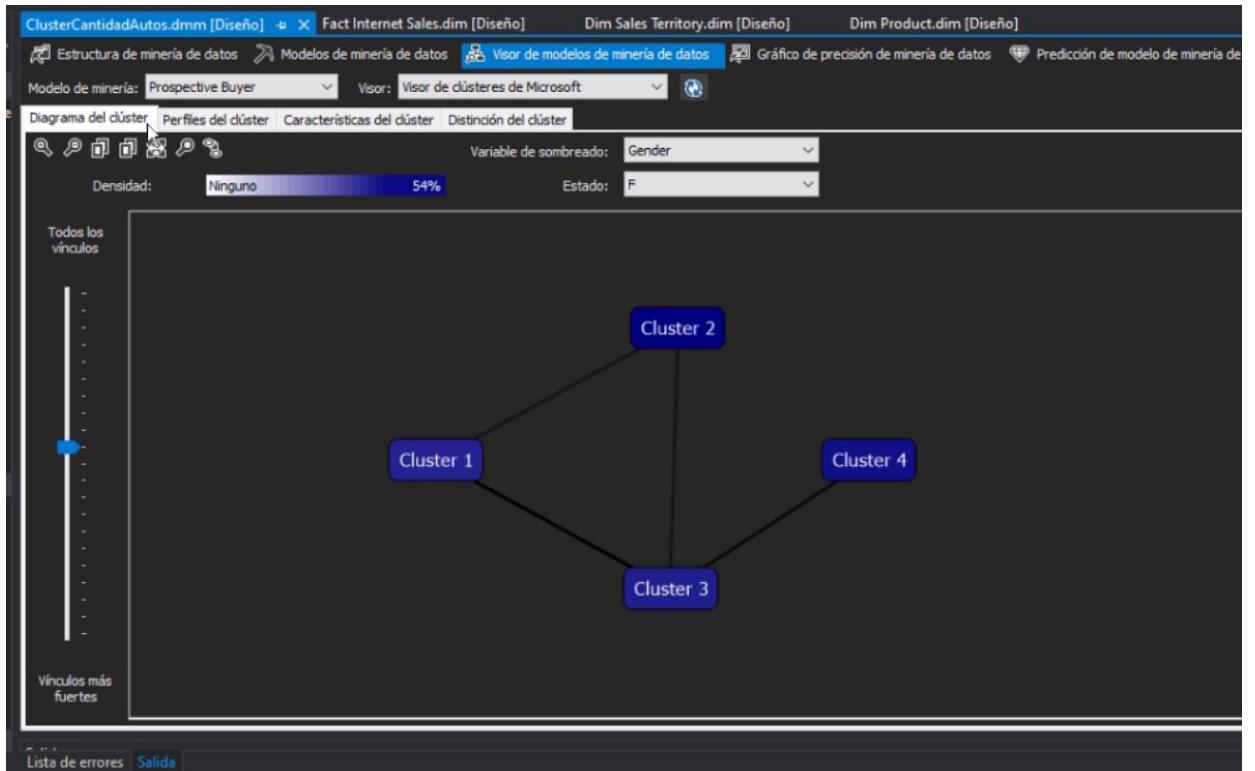
- Posteriormente, deberán procesar la estructura de minería y todos los modelos para efectuar los cambios.



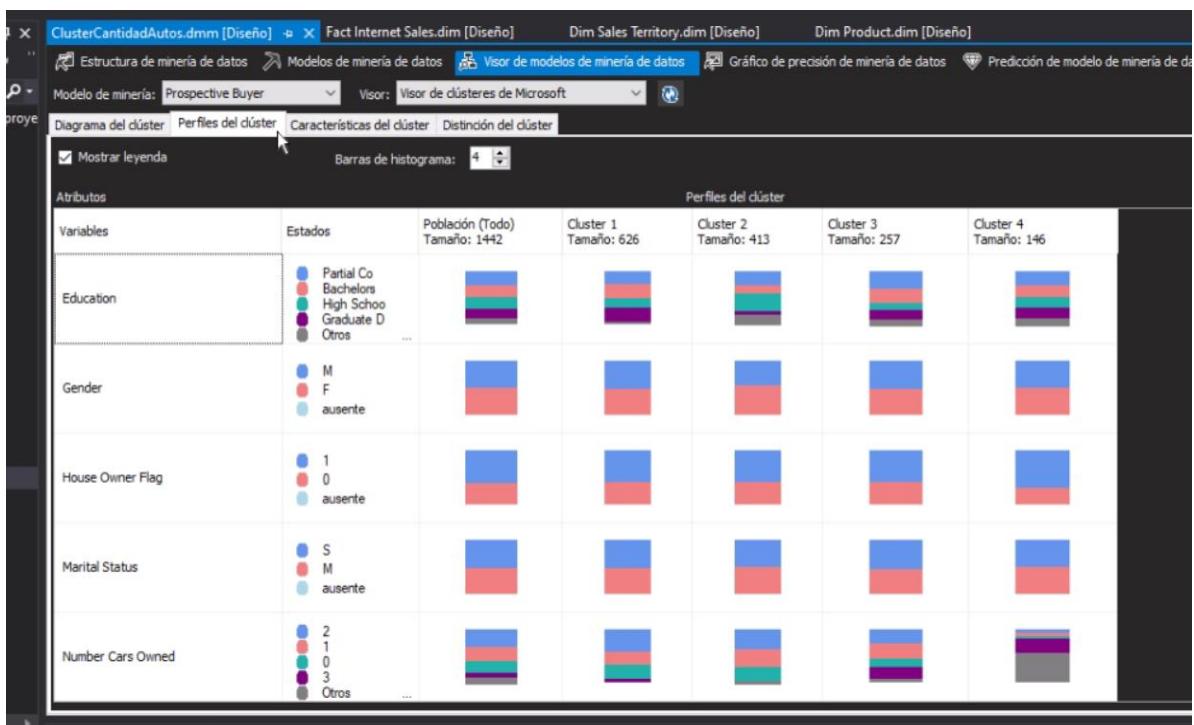
- Seleccionar la pestaña “Visor de modelos de minería de datos” para ver los clústeres representados en forma de diagrama.



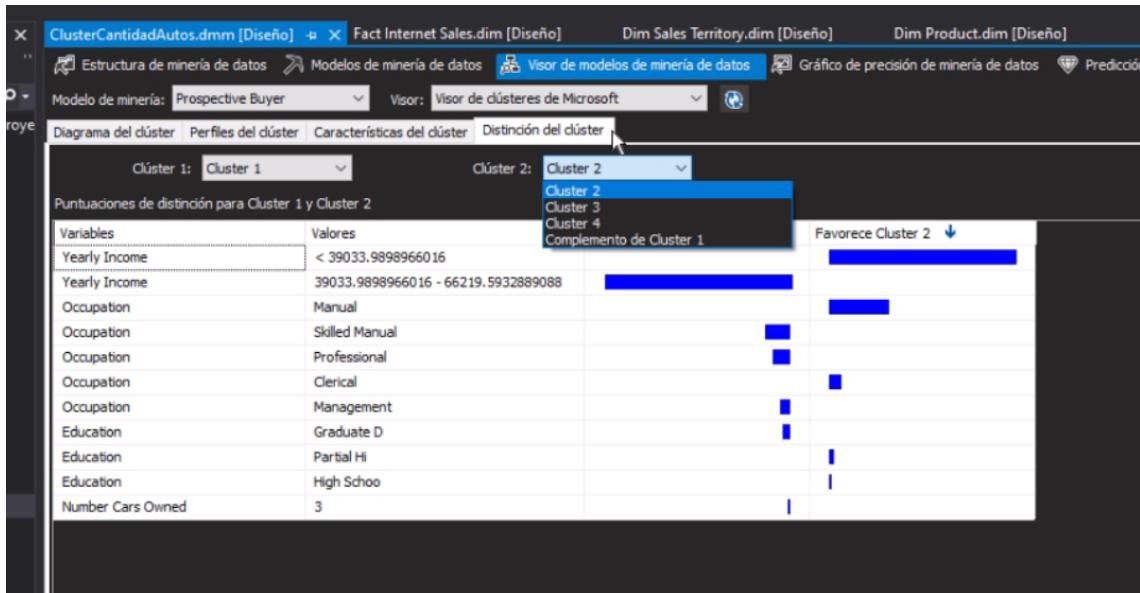
- Una vez procesado, aparecerá un diagrama en el cual podrán visualizar la relación/vínculos entre los clústeres.
- Cuanto más definida sea la unión, más similares serán las poblaciones.



- Si seleccionan la pestaña “Perfiles del clúster” visualizarán diferentes diagramas de barras según los valores de entrada que seleccionaron previamente.



- En la pestaña “Distinción del clúster” podrán comparar entre las distintas variables de los clústeres.



Conclusión de aplicar técnica de clustering para minería de datos.

Esta técnica nos permite identificar de manera automática una cantidad de grupos provenientes de un conjunto de datos de entrada, basándonos en cierta medida de similitud. Buscamos patrones o una estructura de los datos de interés con el objetivo de predecir un resultado.

Los grupos de datos contiene miembros que son muy similares entre sí, pero estos miembros deben ser diferentes a los objetos que pertenecen a otro grupo.

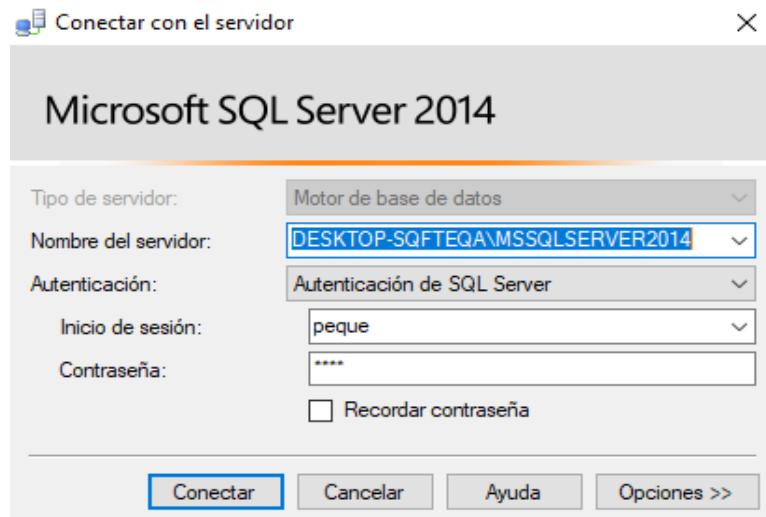
Las técnicas de Clustering son apropiadas para los casos de aprendizaje no supervisado en los que se quiere agrupar y tener conocimiento a un alto nivel de cómo se han generado los datos y como están organizados, sin tener conocimiento a priori de estos.

Errores posibles

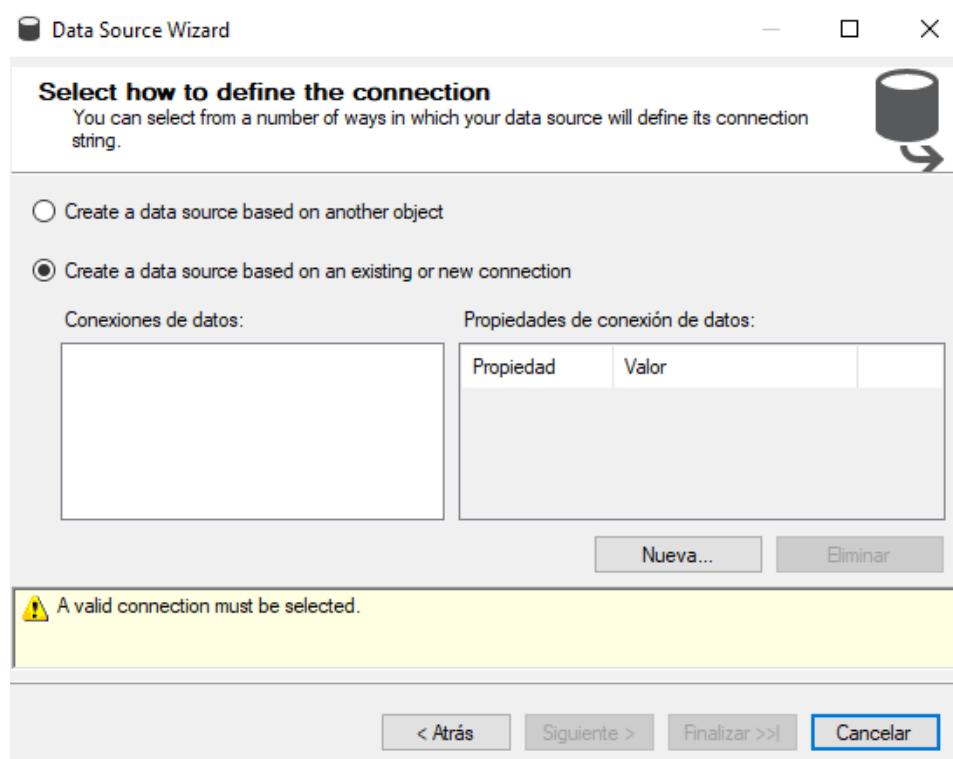
No usar instancia predeterminar de SQL server.

En caso de no usar la instancia predeterminada de SQL Server, se pondría encontrar ciertos errores, para evitarlos debemos realizar los siguientes pasos.

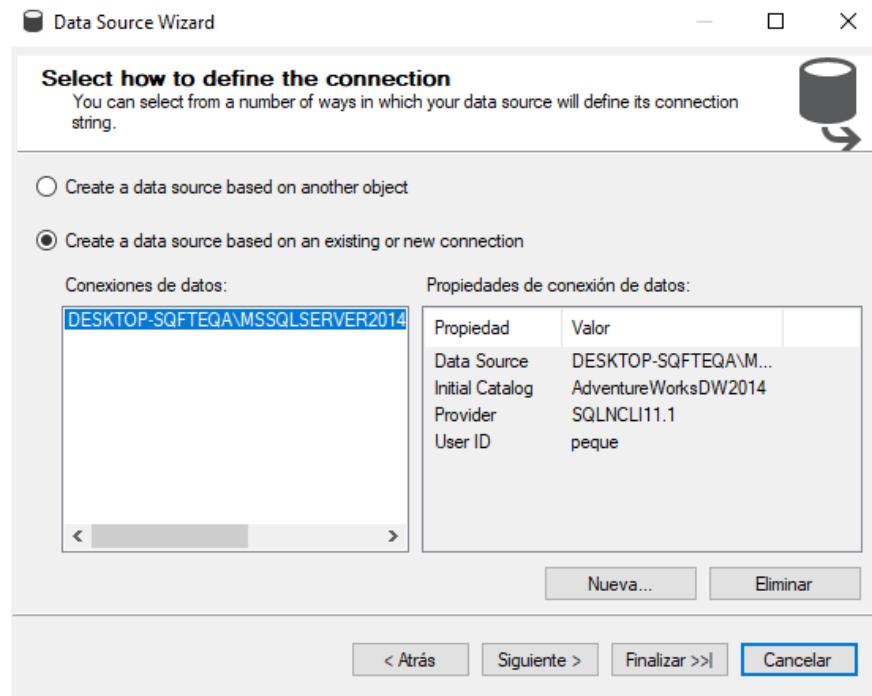
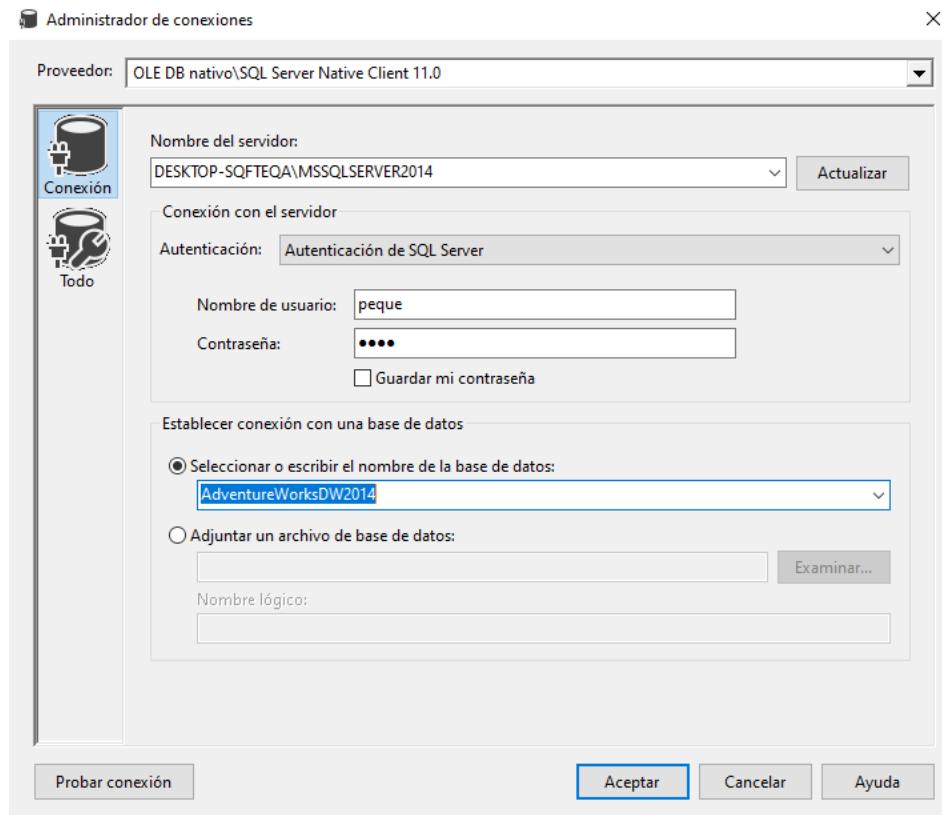
- Copiar el nombre del servidor deseado.



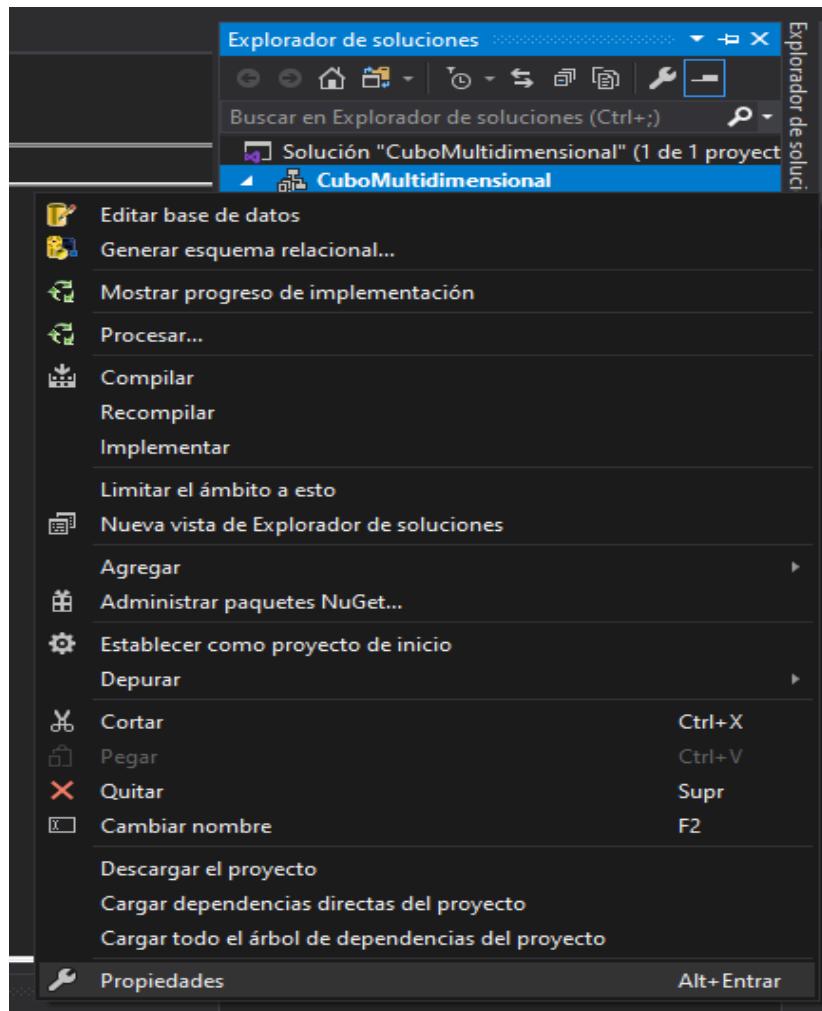
- Crear un nuevo origen de datos.



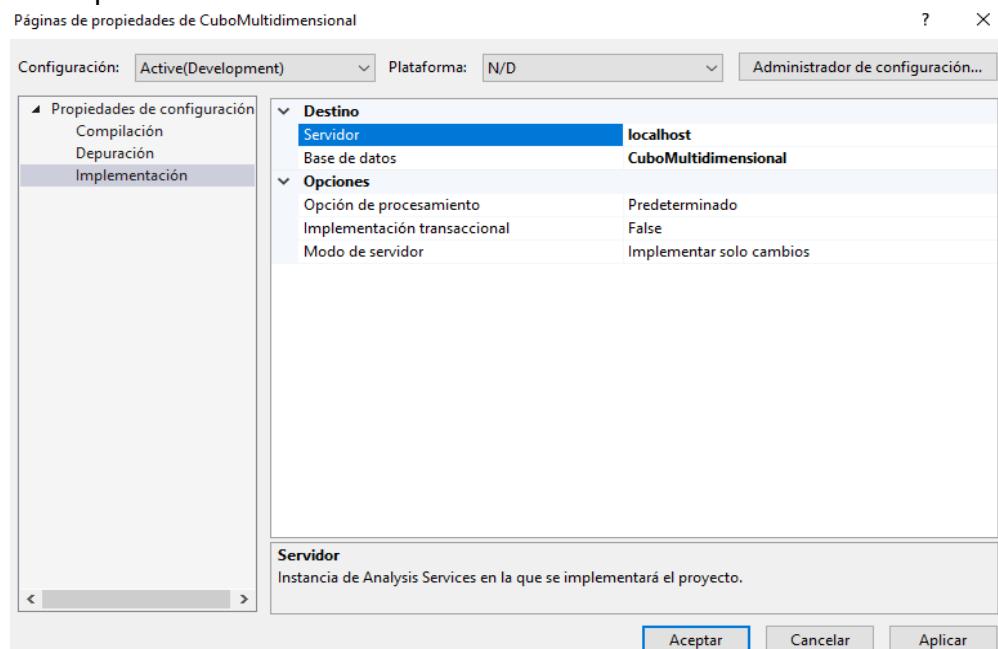
- Colocar el nombre del servidor deseado y hacer click en “Aceptar”. Para verificar que esté correcta la conexión, puede presionar en “Probar Conexión”.



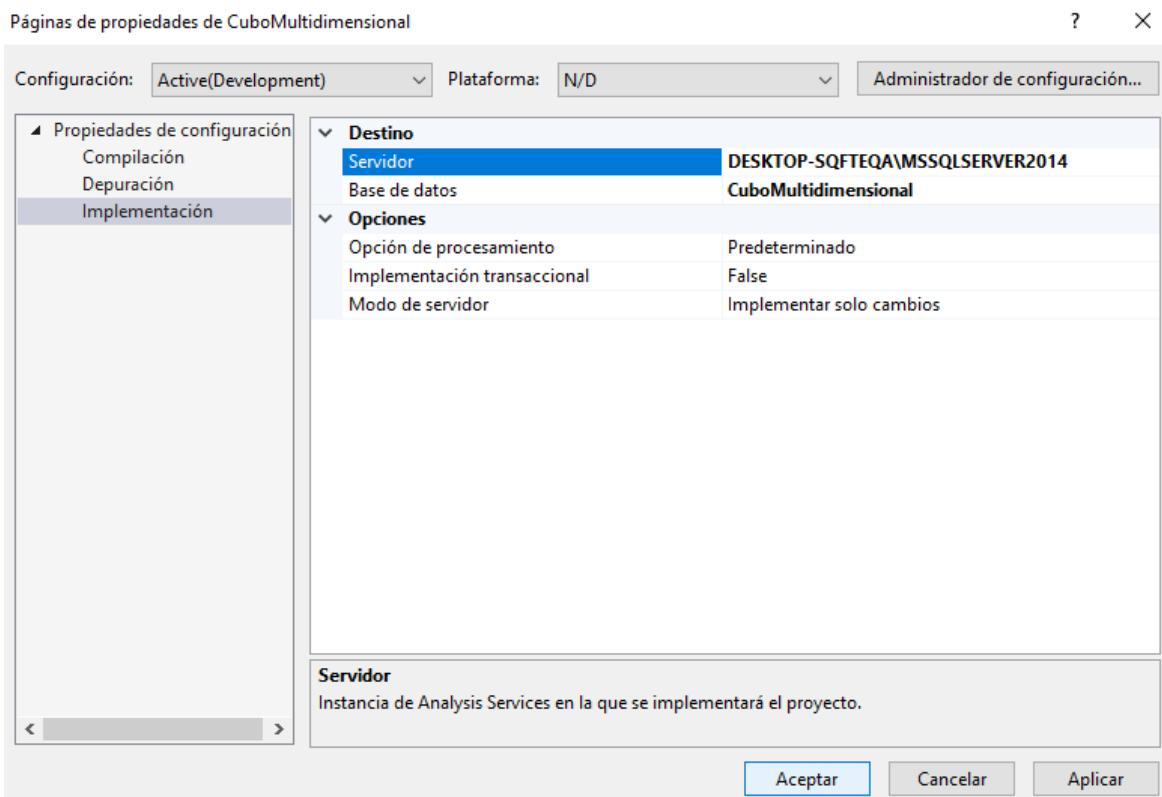
- Luego se deberá asignar el servidor en la ventana de propiedades del proyecto. Para eso es necesario hacer clic derecho en el proyecto y seleccionar “Propiedades”.



- A continuación, se muestra en pantalla la ventana de propiedades y elegir la opción de “Implementación”

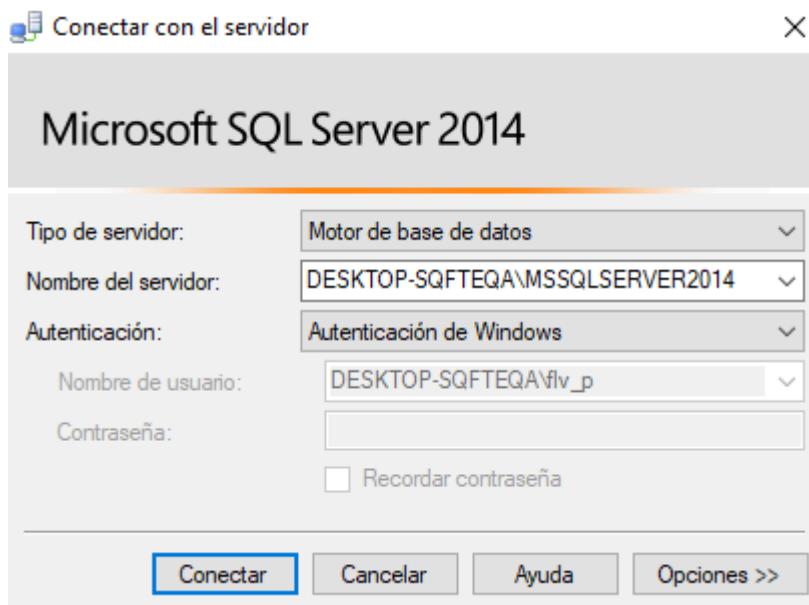


- Escribir el nombre del mismo servidor del origen de datos y dar clic en “Aceptar”.

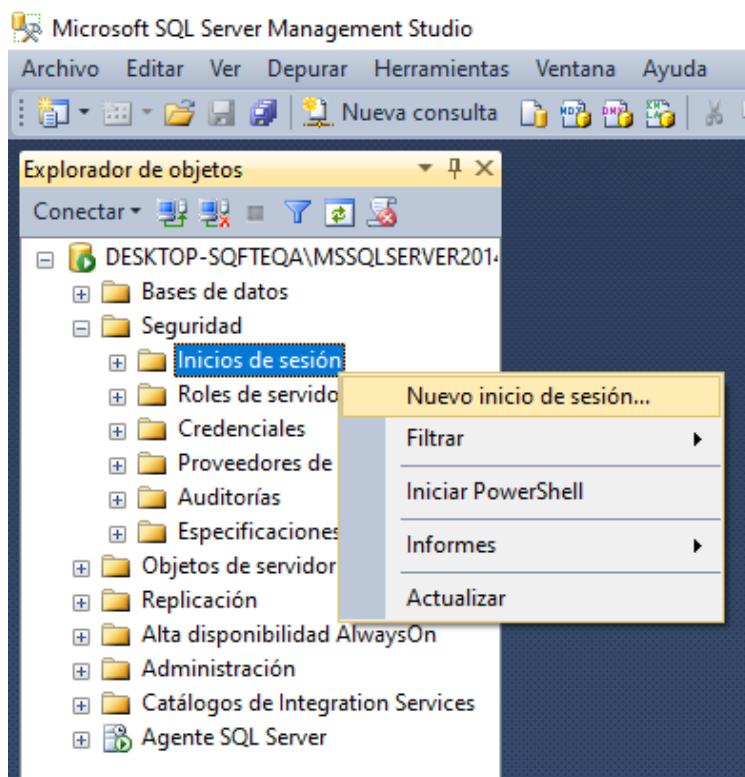


Usuario sin permisos necesarios

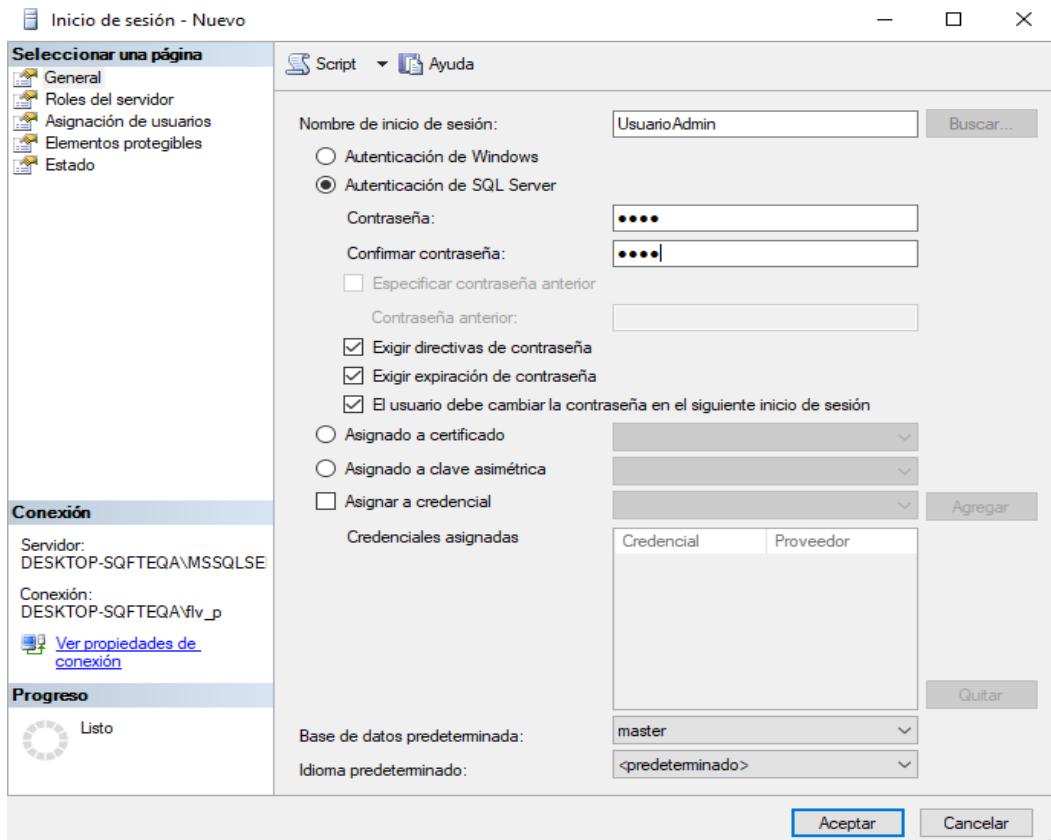
- Nos conectamos a la instancia de SQL que utilizamos para el proyecto



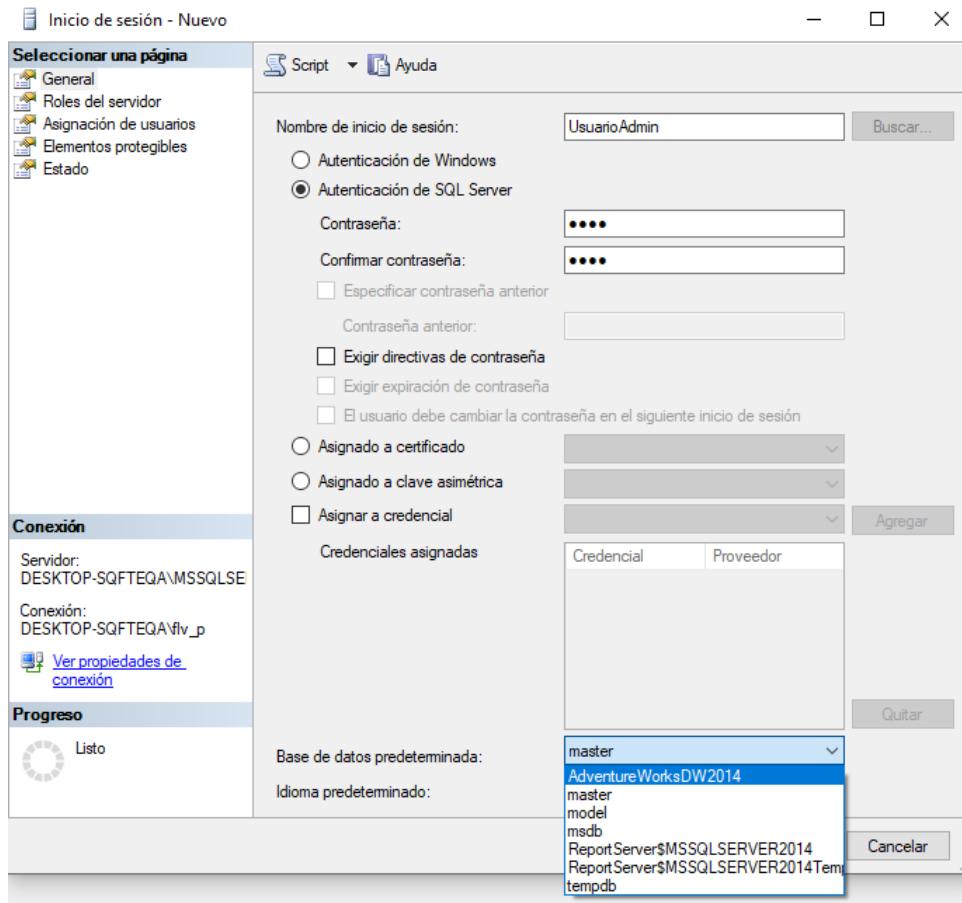
- Presionar click derecho en Inicios de Sesión y seleccionar “Nuevo inicio de sesión”, esto se utiliza para crear un nuevo usuario.



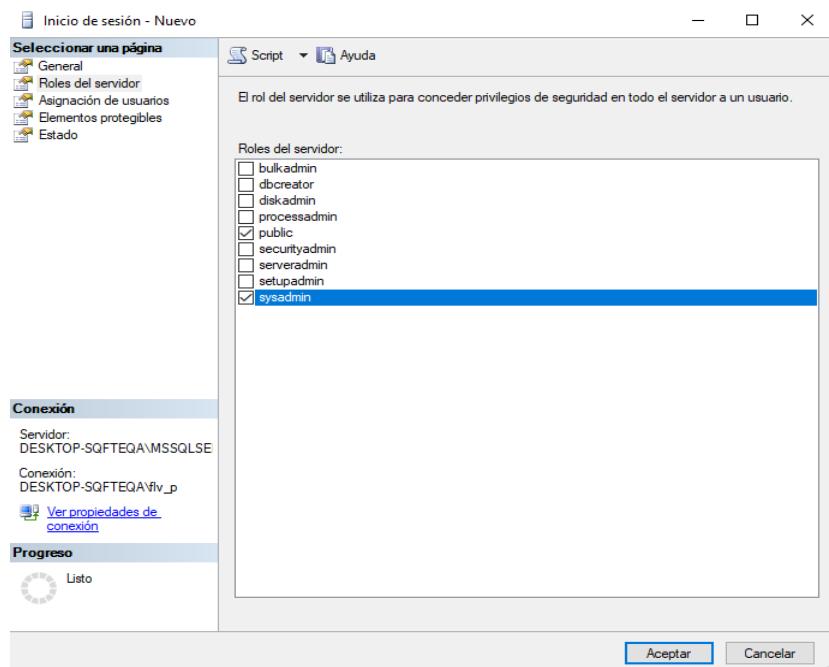
- Definir el nombre del inicio de sesión, seleccionar “Autenticación de SQL Server” e ingresar la contraseña del usuario nuevo.



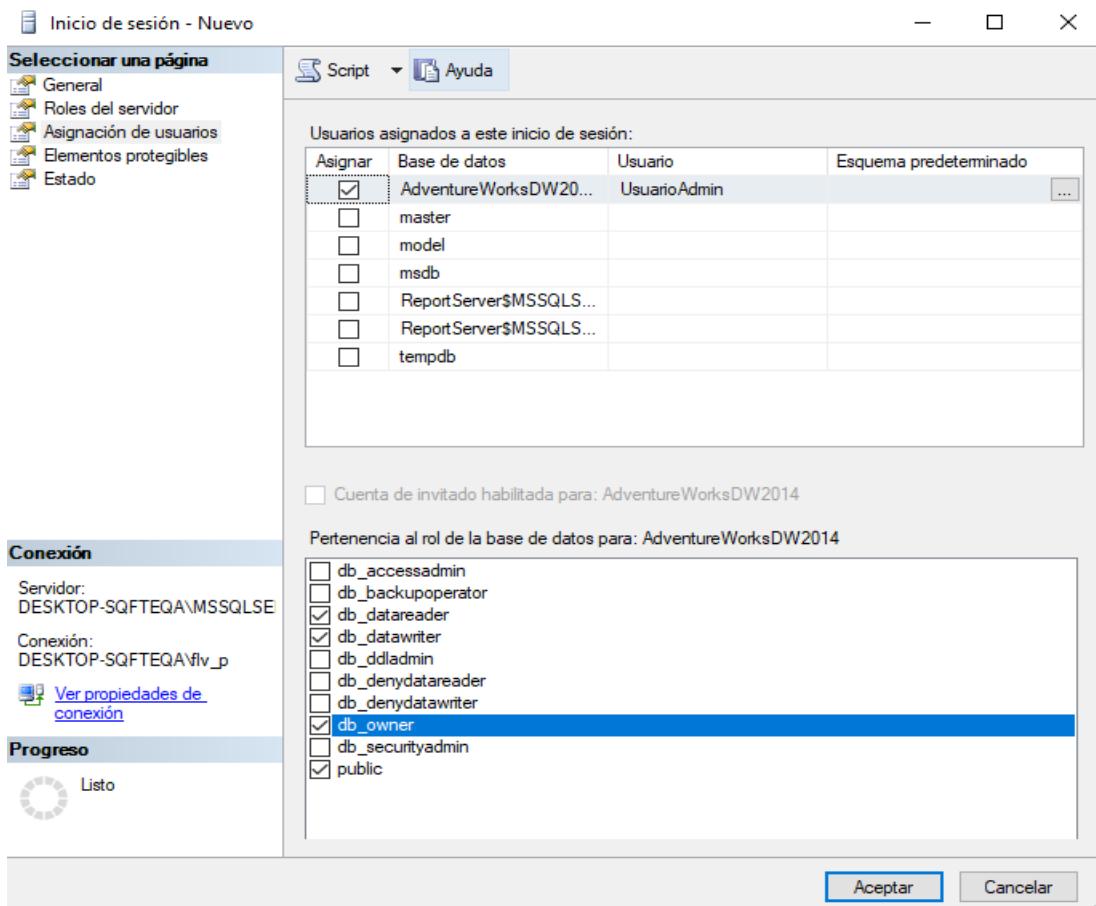
- Desmarcar la opción de “Exigir directiva de contraseña”.
- Seleccionar como base predeterminada AdventureWorksDW2014 que es la utilizada en el proyecto.



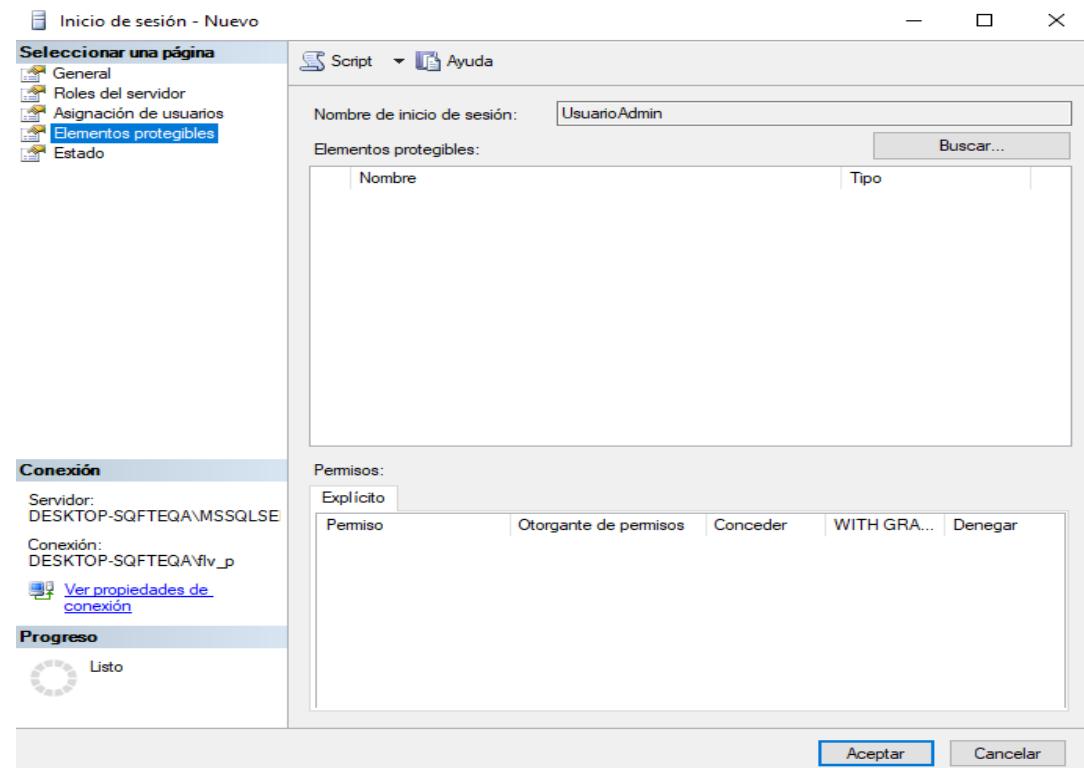
- Avance a la página “Roles del servidor” y tildar “sysadmin”.



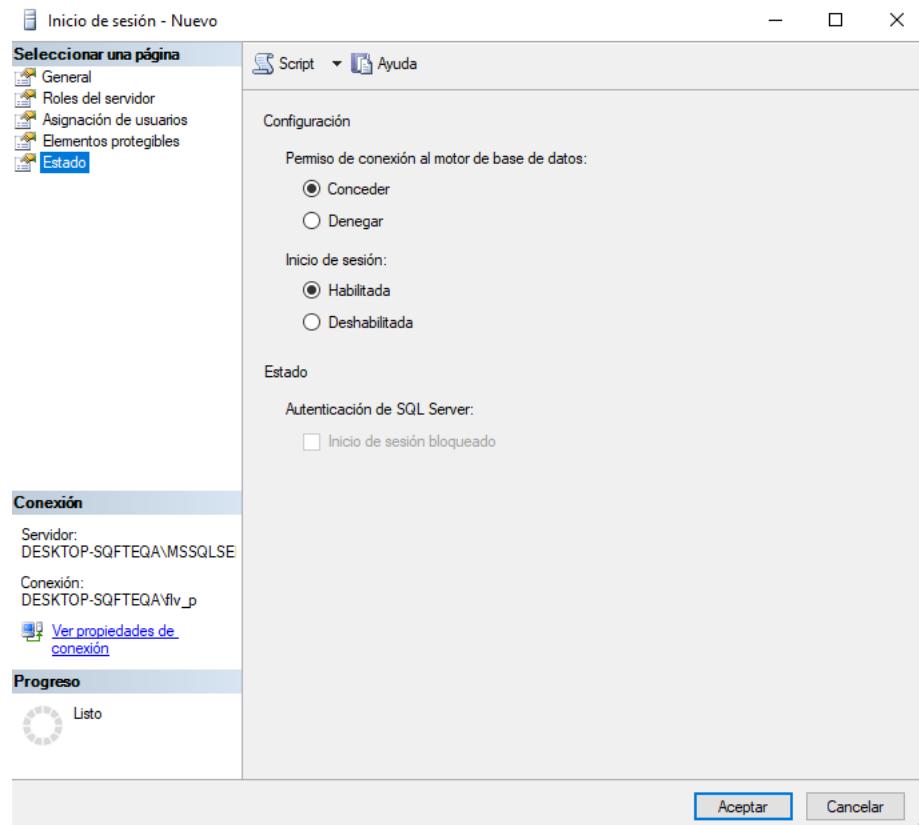
- Ahora avance a la página de “Asignación de usuario”, seleccionar la base de datos y en la grilla de abajo, tildar las opciones que se ven en la imagen.



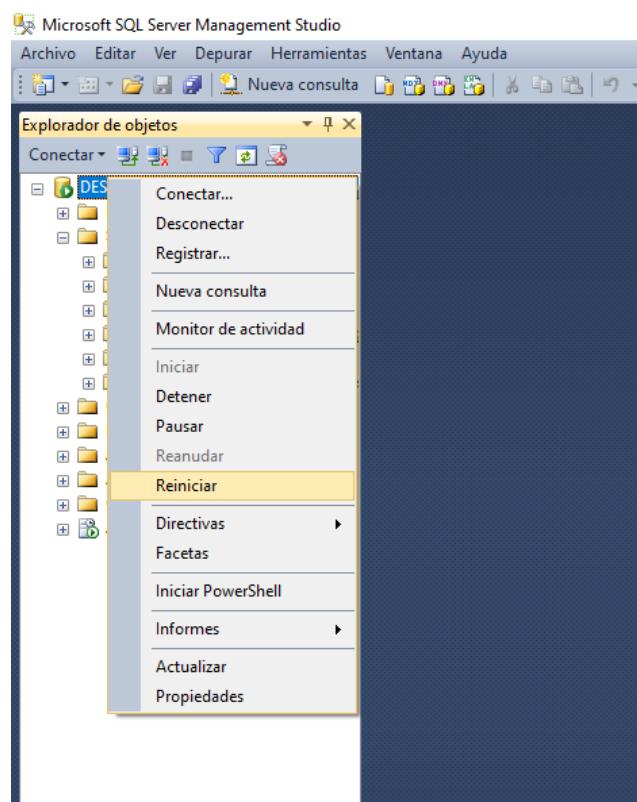
- Verifique que la página de “Elementos protegibles” se vea igual que la imagen.



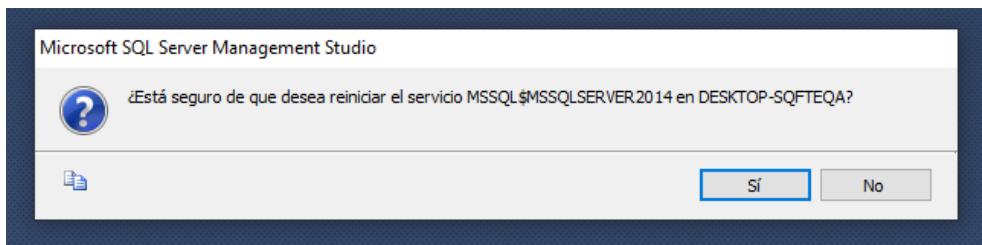
- En la página de “Estado” verificamos que las opciones sean las mismas que la imagen.



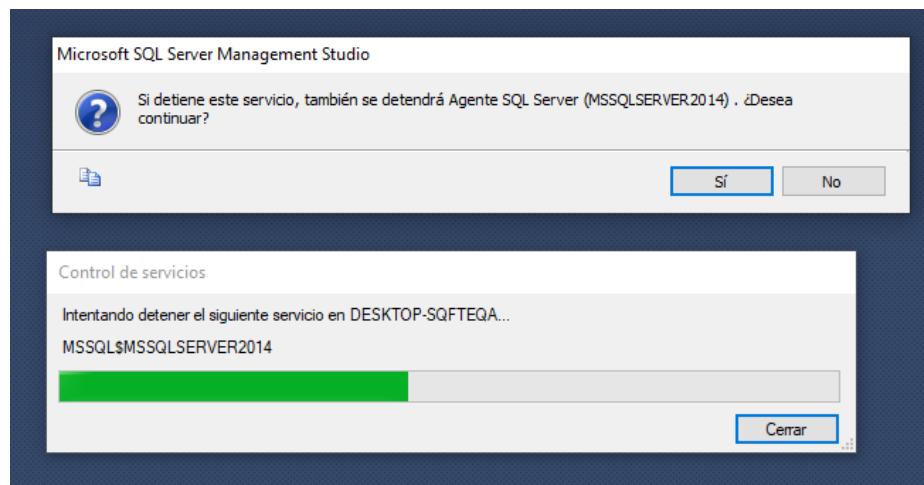
- Reiniciar la conexión a la base de datos para efectuar los cambios.



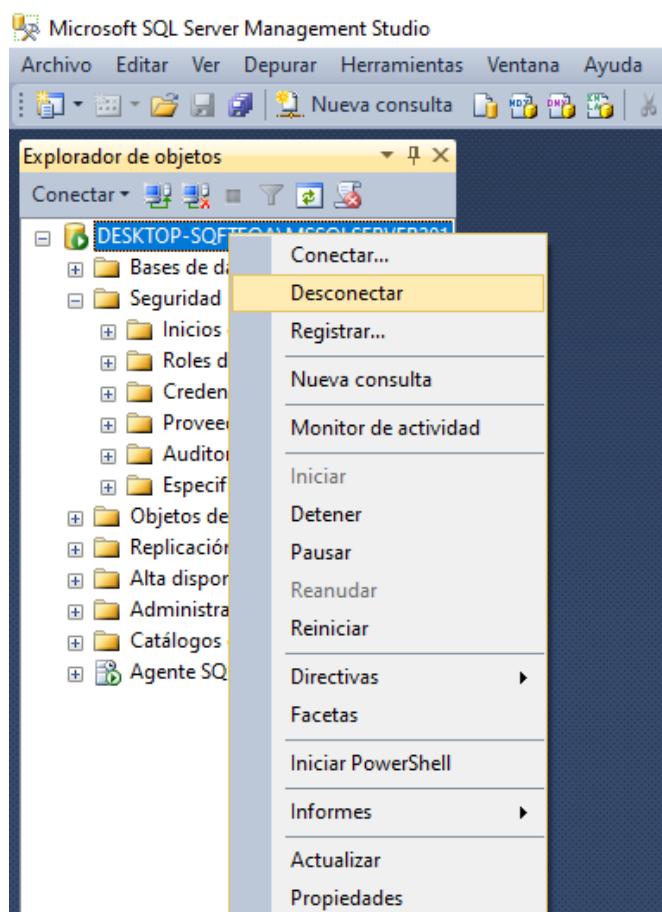
- Seleccionar “si” para reiniciar.



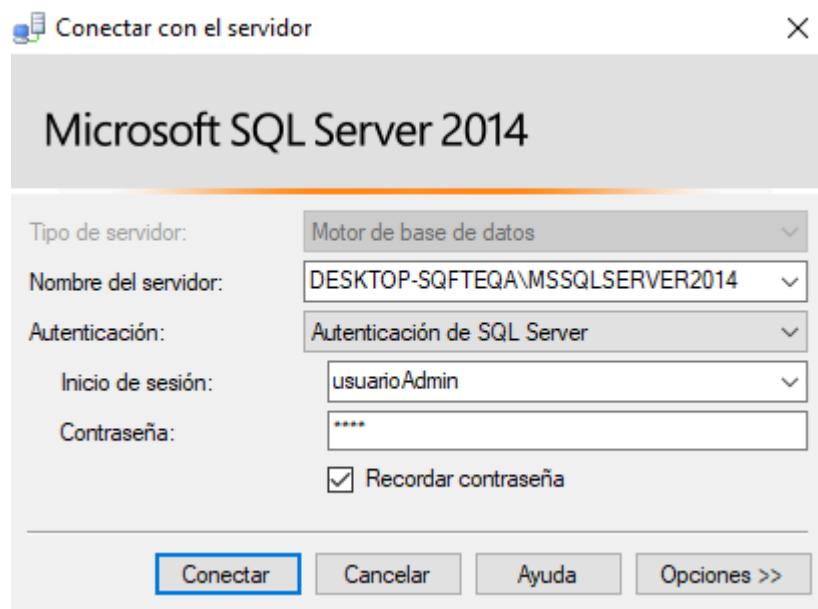
- Seleccionar “si” y esperar a que se vuelva a abrir la conexión.



- Para probar que funcione el nuevo usuario, se deberá desconectar de la instancia.



- Conéctese con el usuario recién creado para verificar que funciona.



- En caso de que el logo de la base de datos no esté verificado con el punto verde se deberá reiniciar el SQL Server Management

