

Preparación de datos en Power BI Desktop

En este laboratorio, empezará a desarrollar una solución de Power BI Desktop para la empresa Adventure Works. A lo largo del proceso, conectará con datos de origen, obtendrá una vista previa de los datos y usará técnicas de vista previa de datos para comprender las características y la calidad de los datos de origen.

En este laboratorio, aprenderá a:

- Abrir Power BI Desktop
- Establecer las opciones de Power BI Desktop
- Conectar con datos de origen
- Obtener una vista previa de los datos de origen
- Usar técnicas de vista previa de datos para entender mejor los datos

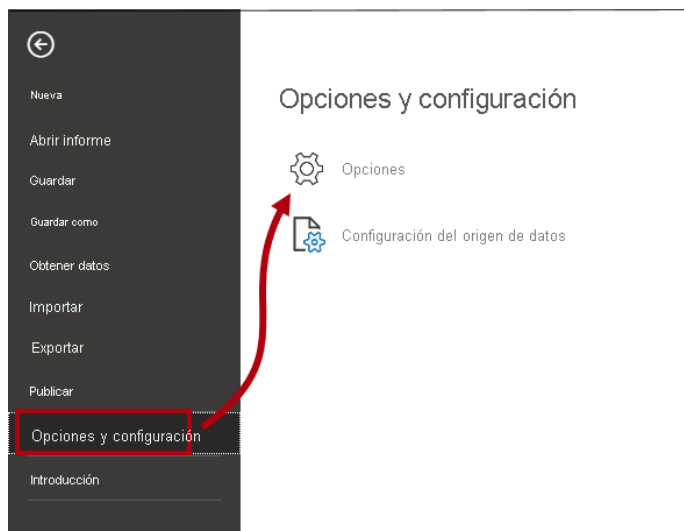
Preparación de los datos

En este ejercicio, creará ocho consultas de Power BI Desktop. Seis de ellas obtendrán datos de SQL Server y dos, de archivos CSV.

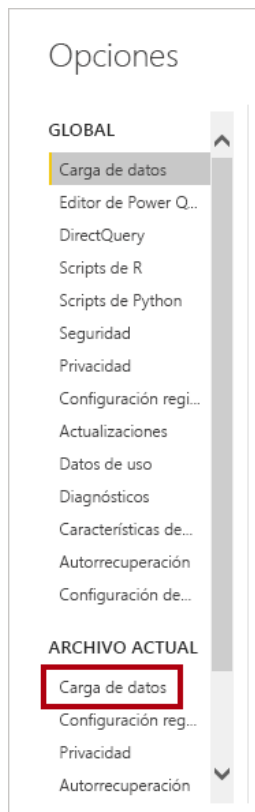
Establecimiento de las opciones de Power BI Desktop

En esta tarea, establecerá las opciones de Power BI Desktop.

1. En Power BI Desktop, haga clic en la ficha de cinta **Archivo** para abrir la vista Backstage.
2. A la izquierda, elija **Opciones y configuración** y seleccione **Opciones**.

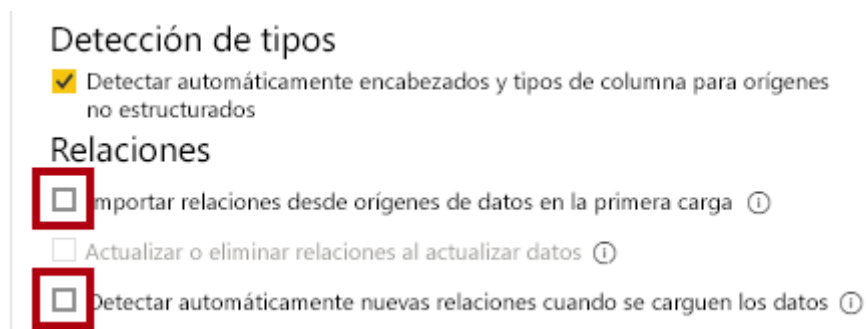


3. En la ventana **Opciones**, a la izquierda, en el grupo **Archivo actual**, seleccione **Carga de datos**.



La configuración de **Carga de datos** para el archivo actual permite una serie de opciones que determinan los comportamientos predeterminados del modelado.

4. En el grupo **Relaciones**, desactive las dos opciones que están activadas.



Aunque tener estas dos opciones habilitadas puede resultar útil a la hora de desarrollar un modelo de datos, las ha deshabilitado anteriormente para admitir la experiencia de laboratorio.

5. Haga clic en **Aceptar**.

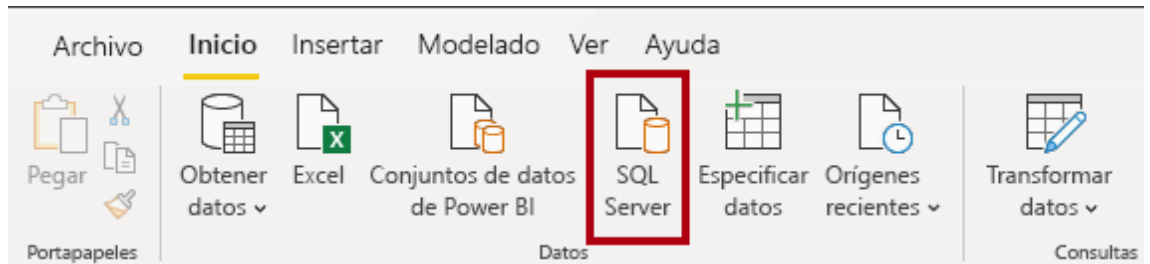


6. Guarde el archivo de Power BI Desktop.

Tarea 3: Obtención de datos de SQL Server

En esta tarea, creará consultas basadas en tablas de SQL Server.

1. En la ficha de cinta **Inicio**, en el grupo **Datos**, haga clic en **SQL Server**.



2. En la ventana **Base de datos de SQL Server**, en el cuadro **Servidor**, escriba **localhost**.



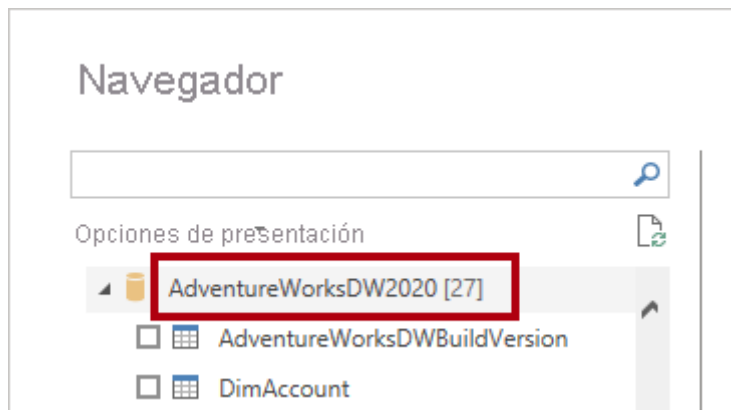
En este laboratorio, se conectará a la base de datos de SQL Server mediante **localhost**. Esto no es una práctica recomendada cuando cree sus propias soluciones. simplemente se debe a que los orígenes de datos de la puerta de enlace no pueden resolver **localhost**.

3. Haga clic en **Aceptar**.

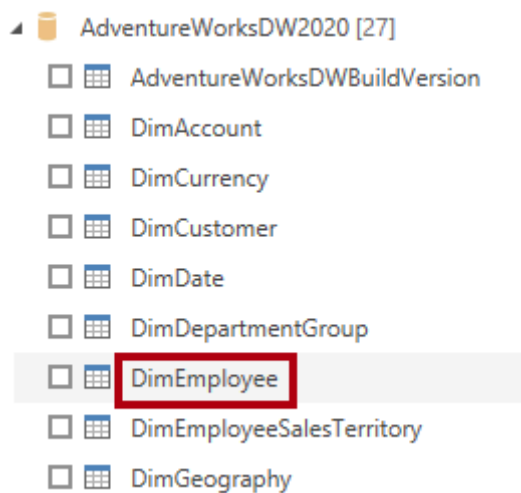


4. En la ventana **Navegador**, a la izquierda, expanda la base de datos **AdventureWorksDW2020**.

La base de datos **AdventureWorksDW2020** está basada en la base de datos de ejemplo **AdventureWorksDW2017**. Se ha modificado para admitir los objetivos de aprendizaje de los laboratorios del curso.



5. Seleccione la tabla **DimEmployee**, pero no active su casilla.



6. En el panel de la derecha, fíjese en que hay una vista previa de los datos de la tabla.

Los datos de la vista previa permiten determinar las columnas y un ejemplo de filas.

7. Para crear consultas, active la casilla situada junto a las seis tablas siguientes:

- DimEmployee
- DimEmployeeSalesTerritory
- DimProduct
- DimReseller
- DimSalesTerritory
- FactResellerSales

8. Para aplicar transformaciones a los datos de las tablas seleccionadas, haga clic en **Transformar datos**.

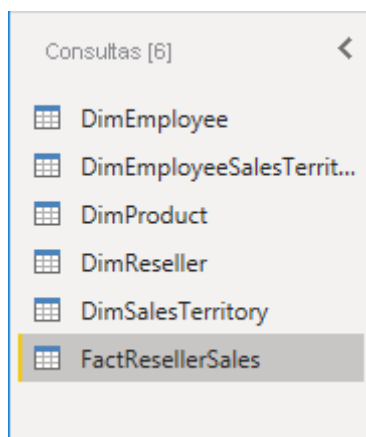
En este laboratorio, no transformará los datos. Los objetivos de este laboratorio se centran en explorar y generar perfiles de los datos en la ventana **Editor de Power Query**.



Tarea 4: Vista previa de las consultas de SQL Server

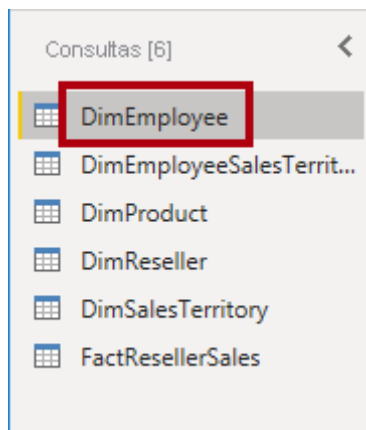
En esta tarea, obtendrá una vista previa de los datos de las consultas de SQL Server. En primer lugar, descubrirá información importante sobre los datos. Además, usará herramientas de calidad, distribución y perfil de las columnas para entender los datos y evaluar su calidad.

1. En la ventana del **Editor de Power Query**, a la izquierda, observe el panel **Consultas**.



El panel **Consultas** contiene una consulta para cada tabla seleccionada.

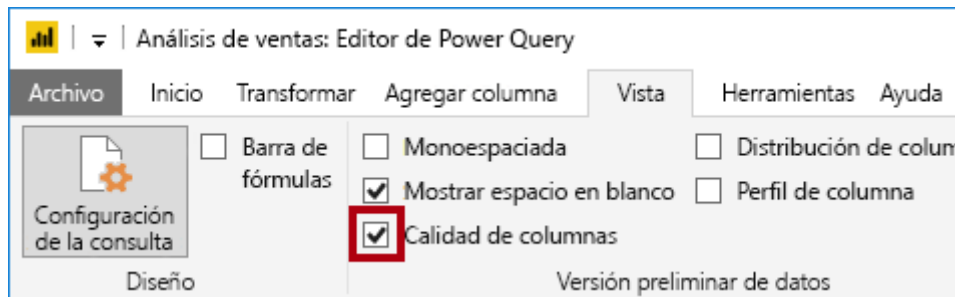
2. Seleccione la primera consulta **DimEmployee**.



La tabla **DimEmployee** de la base de datos de SQL Server almacena una fila por cada empleado. Un subconjunto de las filas de esta tabla representa a los vendedores, que será pertinente para el modelo que va a desarrollar.

3. En la barra de estado de la parte inferior izquierda, observe las estadísticas de la tabla: 33 columnas y 296 filas.

4. En el panel de vista previa de los datos, desplácese horizontalmente para revisar todas las columnas.
5. Observe que las cinco últimas columnas contienen vínculos de tipo **Tabla o Valor**.
6. Para evaluar la calidad de las columnas, en la ficha de cinta **Vista**, en el grupo **Vista previa de los datos**, consulte **Calidad de columnas**.

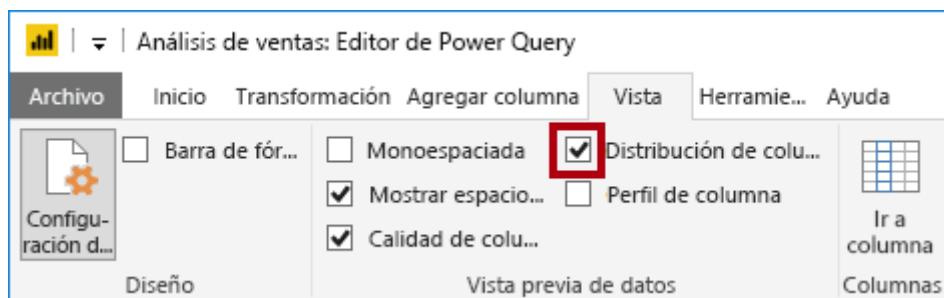


La característica calidad de las columnas permite determinar fácilmente el porcentaje de valores válidos, errores o valores vacíos encontrados en las columnas.

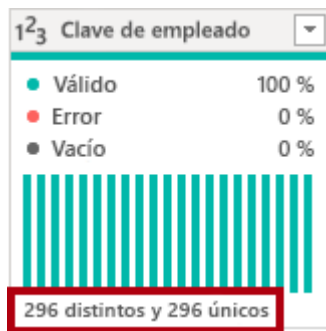
7. En la columna **Position** (sexta columna desde el final), observe que el 94 % de las filas están vacías (NULL).

A ^B C Posición	
● Válido	6 %
● Error	0 %
● Vacío	94 %

8. Para evaluar la distribución de las columnas, en la ficha de cinta **Vista**, en el grupo **Vista previa de los datos**, consulte **Distribución de columnas**.

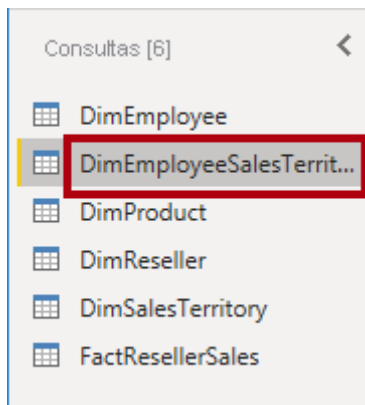


9. Revise de nuevo la columna **Position** y observe que hay cuatro valores distintos y uno único.
10. Revise la distribución de la columna **EmployeeKey** (la primera): 296 valores distintos y 296 únicos.



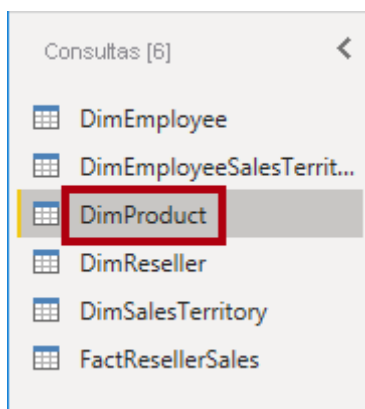
Cuando los recuentos distintos y únicos son iguales, significa que la columna contiene valores únicos. Al realizar el modelado, es importante que algunas tablas del modelo tengan columnas únicas

11. En el panel **Consultas**, seleccione la consulta **DimEmployeeSalesTerritory**.



La tabla **DimEmployeeSalesTerritory** almacena una fila para cada uno de los empleados y las regiones del territorio de ventas que administran. La tabla permite relacionar varias regiones con un solo empleado. Algunos empleados administran una, dos o posiblemente más regiones.

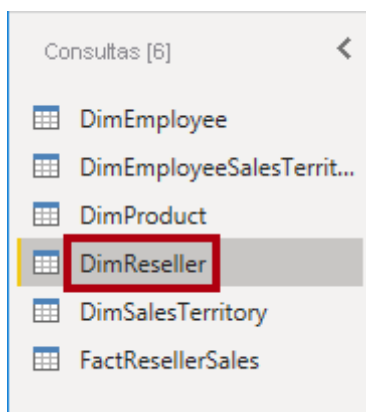
12. En el panel **Consultas**, seleccione la consulta **DimProduct**.



La tabla **DimProduct** contiene una fila por producto vendido por la empresa.

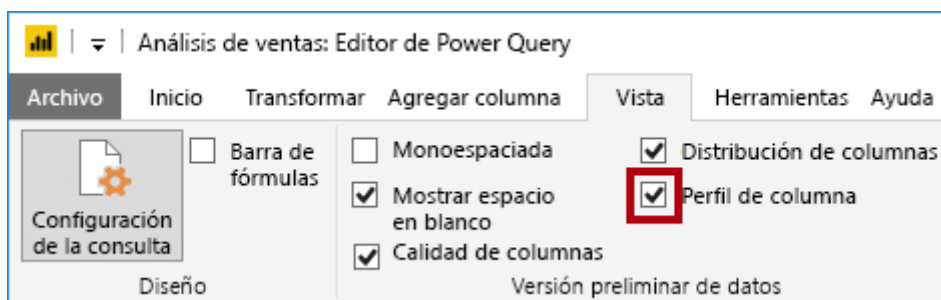
13. Desplácese horizontalmente para mostrar las últimas columnas.

14. Observe la columna **DimProductSubcategory**.
15. En el panel **Consultas**, seleccione la consulta **DimReseller**.

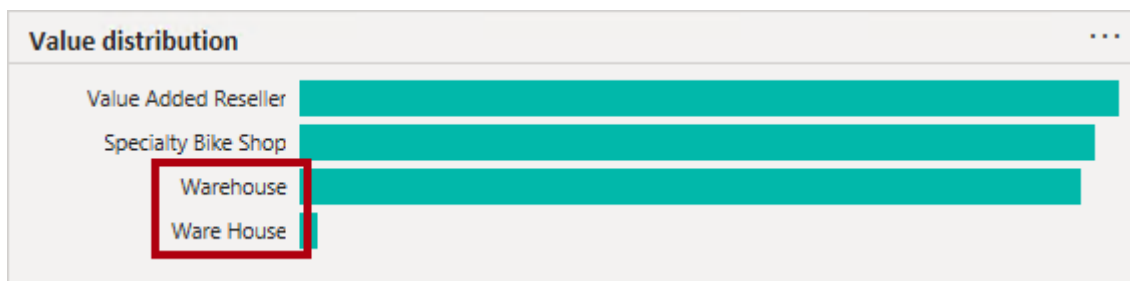


La tabla **DimReseller** contiene una fila por distribuidor. Los revendedores venden, distribuyen o agregan valor a los productos de Adventure Works.

16. Para ver los valores de las columnas, en la ficha de cinta **Vista**, en el grupo **Vista previa de los datos**, consulte **Perfil de columna**.

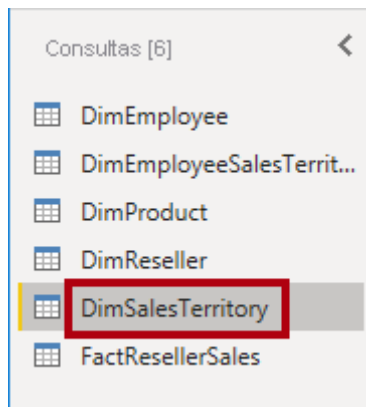


17. Seleccione el encabezado de columna **BusinessType**.
18. Observe el nuevo panel que hay debajo del panel de vista previa de los datos.
19. Revise las estadísticas de columna y la distribución de valores en el panel de vista previa de los datos.
20. Observe el problema en la calidad de los datos: hay dos etiquetas para el almacén (**Warehouse** y **Ware House**, este último mal escrito).



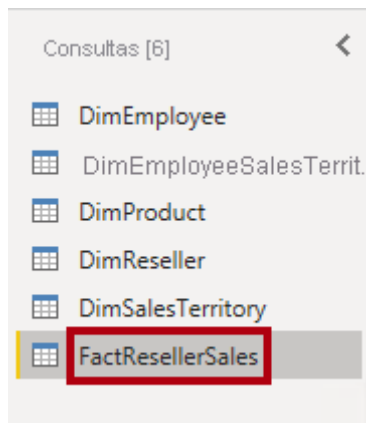
21. Si mantiene el puntero sobre la barra de **Ware House**, verá que hay cinco filas con este valor.

22. En el panel **Consultas**, seleccione la consulta **DimSalesTerritory**.



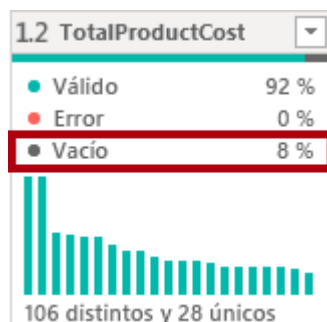
La tabla **DimSalesTerritory** contiene una fila por región de ventas, que incluye **Corporate HQ** (Sede corporativa). Las regiones se asignan a un país y los países se asignan a los grupos.

En el panel **Consultas**, seleccione la consulta **FactResellerSales**.



La tabla **FactResellerSales** contiene una fila por línea de pedido de venta. Un pedido de venta contiene uno o varios elementos de línea.

23. Revise la calidad de la columna **TotalProductCost** y observe que el 8 % de las filas están vacías.



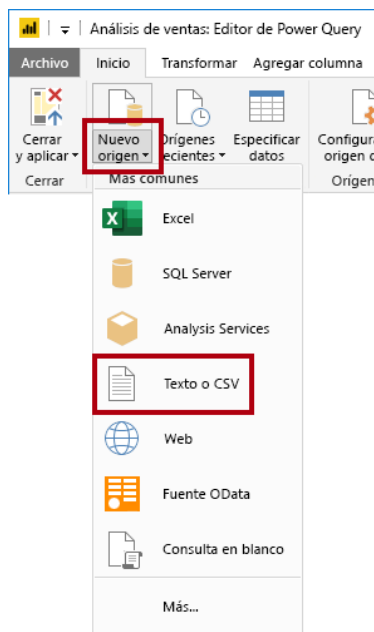
El hecho de que falten valores en la columna **TotalProductCost** supone un problema de calidad de los datos. Para solucionarlo en el laboratorio **Carga de**

datos en Power BI Desktop, aplicará transformaciones a fin de rellenar los valores que faltan con el costo estándar del producto, que se almacena en la tabla relacionada **DimProduct**.

Tarea 5: Obtención de datos de un archivo CSV

En esta tarea, creará una consulta basada en un archivo CSV.

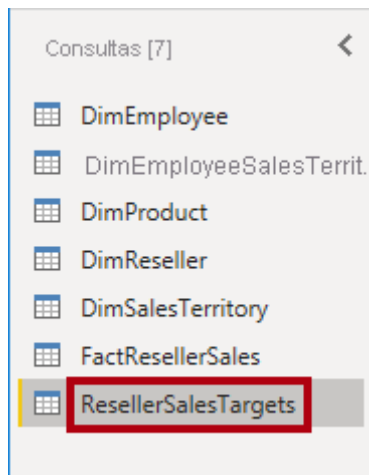
1. Para agregar una nueva consulta, vaya a la ventana **Editor de Power Query**. En la ficha de cinta **Inicio**, en el grupo **Nueva consulta**, haga clic en la flecha desplegable de **Nuevo origen** y luego seleccione **Texto o CSV**.



2. En la ventana **Abrir**, vaya a la carpeta **D:\PL300\Resources** y seleccione el archivo **ResellerSalesTargets.csv**.
3. Haga clic en **Abrir**.
4. En la ventana **ResellerSalesTargets.csv**, revise la vista previa de los datos.
5. Haga clic en **Aceptar**.



6. En el panel **Consultas**, observe la adición de la consulta **ResellerSalesTargets**.

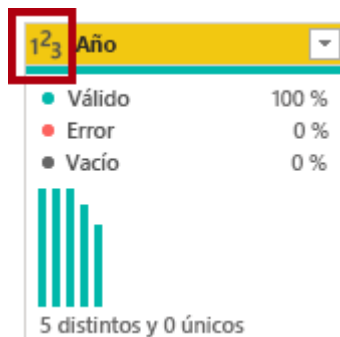


El archivo CSV **ResellerSalesTargets** contiene una fila por vendedor y al año. Cada fila registra 12 objetivos de ventas mensuales (expresados en miles). Tenga en cuenta que el año comercial de la empresa Adventure Works empieza el 1 de julio.

7. Observe que ninguna columna contiene valores vacíos.

Cuando no hay un objetivo de ventas mensual, se almacena un carácter de guion.

8. Revise los iconos de cada encabezado de columna, a la izquierda del nombre de columna.



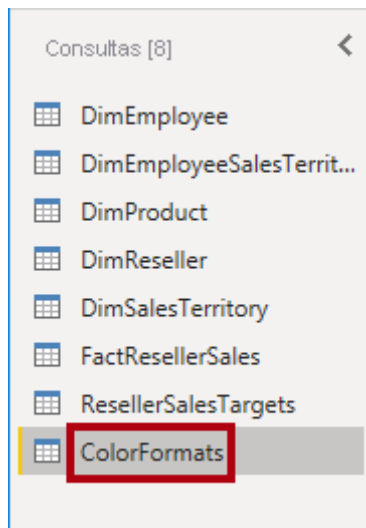
Los iconos representan el tipo de datos de la columna. **123** es un número entero y **ABC** es texto.

Aplicará muchas transformaciones para lograr un resultado con forma diferente que solo conste de tres columnas (**Date**, **EmployeeKey** y **TargetAmount**)

Tarea 6: Obtención de datos adicionales de un archivo CSV

En esta tarea, creará una consulta adicional basada en un archivo CSV diferente.

1. Siga los pasos de la tarea anterior para crear una consulta basada en el archivo **D:\PL300\Resources\ColorFormats.csv**.



El archivo CSV **ColorFormats** contiene una fila por color de producto. Cada fila registra los códigos hexadecimales para dar formato a los colores de fondo y de fuente.