

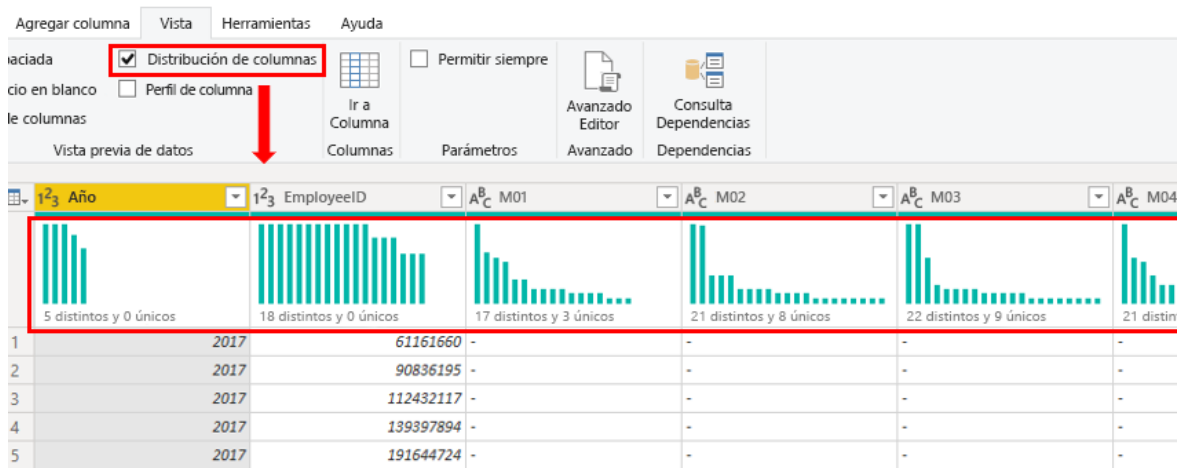
Reducción de la cardinalidad

La cardinalidad es un término que se usa para describir la unicidad de los valores de una columna. La cardinalidad también se usa en el contexto de las relaciones entre dos tablas, donde describe la dirección de la relación.

Identificación de niveles de cardinalidad en columnas

Anteriormente, cuando ha usado el Editor de Power Query para analizar los metadatos, en la opción **Distribución de columnas** de la pestaña **Vista** se mostraban estadísticas de la cantidad de elementos diferentes y únicos de cada columna de los datos.

- **Recuento de valores distintos:** el número total de valores distintos en una columna concreta.
- **Recuento de valores únicos:** el número total de valores que solo aparecen una vez en una columna concreta.



Una columna con muchos valores repetidos en su rango (el recuento de elementos distintos es alto) tendrá un nivel bajo de cardinalidad. Por el contrario, una columna con muchos valores únicos en su rango (el recuento de elementos únicos es alto) tendrá un nivel de cardinalidad alto.

La cardinalidad baja genera un rendimiento más optimizado, por lo que es posible que tenga que reducir el número de columnas de cardinalidad alta del conjunto de datos.

Reducción de la cardinalidad de las relaciones

Al importar varias tablas, es posible que tenga que realizar un análisis con los datos de todas esas tablas. Las relaciones entre esas tablas son necesarias para calcular los resultados de forma precisa y mostrar la información correcta en los informes. Power BI Desktop ayuda a simplificar la creación de esas relaciones. De hecho, en la mayoría de los casos no tendrá que hacer nada; la característica de detección automática se encarga de ello. Pero es posible que en ocasiones tenga que crear relaciones o modificar una relación. En cualquier caso, es importante entender las relaciones en Power BI Desktop y cómo crearlas y editarlas.

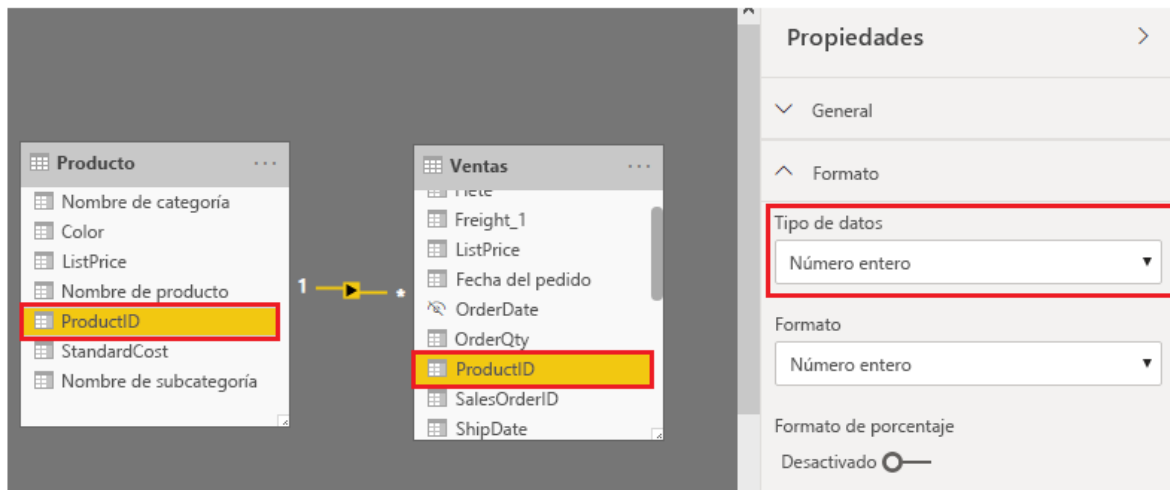
Al crear o editar una relación, puede configurar opciones adicionales. De forma predeterminada, Power BI Desktop configura automáticamente opciones adicionales en función de su mejor aproximación, que puede ser diferente para cada relación según los datos de las columnas.

Las relaciones pueden tener una cardinalidad diferente. La cardinalidad es la dirección de la relación y cada relación del modelo se debe definir con un tipo de cardinalidad. Las opciones de cardinalidad en Power BI son las siguientes:

- **Varios a uno (*:1)**: esta relación es el tipo predeterminado más común. Significa que la columna de una tabla puede tener más de una instancia de un valor y que la otra tabla relacionada, a menudo conocida como tabla de búsqueda, solo tiene una instancia de un valor.
- **Uno a uno (1:1)**: en este tipo de relación, la columna de una tabla solo tiene una instancia de un valor determinado y la otra tabla relacionada solo tiene una instancia de un valor determinado.
- **Uno a varios (1:*)**: en este tipo de relación, la columna de una tabla solo tiene una instancia de un valor determinado y la otra tabla relacionada puede tener más de una instancia de un valor.
- **Varios a varios (:)**: en los modelos compuestos, puede establecer una relación de varios a varios entre las tablas, lo que elimina los requisitos de valores únicos en las tablas. También quita las soluciones alternativas anteriores, como la incorporación de tablas nuevas solo para establecer relaciones.

Durante el desarrollo, creará y editará las relaciones en el modelo, de modo que, cuando cree relaciones en él, independientemente de la cardinalidad que haya elegido, debe asegurarse siempre de que las dos columnas que usa para participar en una relación compartan el mismo tipo de datos. El modelo no funcionará si intenta crear una relación entre dos columnas, donde una tiene un tipo de datos de texto y la otra un tipo de datos entero.

En el ejemplo siguiente, el campo **ProductID** tiene el tipo de datos **Número entero** en las tablas Product y Sales. Las columnas con el tipo de datos **Entero** funcionan mejor que las que tienen el tipo de datos **Texto**.



Mejora del rendimiento mediante la reducción de los niveles de cardinalidad

Power BI Desktop ofrece distintas técnicas que se pueden usar para reducir los datos que se cargan en los modelos de datos, como la de resumen. La reducción de los datos que se cargan en el modelo mejorará la cardinalidad de las relaciones del informe. Por esta razón, es importante que se esfuere por minimizar los datos que se cargarán en los modelos. Esto es especialmente cierto para los modelos grandes o los modelos que prevé que crecerán de tamaño con el tiempo.

Posiblemente la técnica más eficaz para reducir el tamaño de un modelo sea usar una tabla de resumen del origen de datos. Mientras que una tabla de detalles puede contener todas las transacciones, una tabla de resumen contendrá un registro por día, semana o mes. Podría ser un promedio de todas las transacciones diarias, por ejemplo.

Ahí va un ejemplo: una tabla de hechos de ventas de origen almacena una fila por cada línea de pedido. Se puede lograr una reducción significativa de los datos mediante el resumen de todas las métricas de ventas si se agrupa por fecha, cliente y producto, y no se necesitan los detalles de cada transacción individual.

Por tanto, se puede lograr una reducción de datos incluso más significativa mediante la agrupación por fecha a nivel de mes. Podría lograr una reducción posible del 99 % del tamaño del modelo; pero ya no se podrían crear informes en el nivel de día o de pedido individual. La decisión de resumir los datos de tipo de hechos siempre implicará una compensación con los detalles de los datos. Una desventaja es que podría perder la capacidad de profundizar en

los datos porque el detalle ya no existe. Esta contrapartida se podría mitigar mediante el uso de un diseño de modelo mixto.

En Power BI Desktop, un diseño de modo mixto genera un modelo compuesto. Básicamente, permite determinar el modo de almacenamiento de cada tabla. Por tanto, se puede establecer la propiedad **Modo de almacenamiento** de cada tabla en **Importar** o **DirectQuery**.

Una técnica eficaz para reducir el tamaño del modelo consiste en establecer la propiedad **Modo de almacenamiento** para tablas de tipos de hechos mayores en **DirectQuery**. Este enfoque de diseño puede funcionar bien junto con las técnicas que se usan para resumir los datos. Por ejemplo, los datos de ventas resumidos se podrían usar para lograr informes de "resumen" de alto rendimiento. Se podría crear una página de obtención de detalles para mostrar las ventas pormenorizadas de un contexto de filtro específico (y reducido) y mostrar todos los pedidos de ventas en contexto. La página de obtención de detalles incluiría objetos visuales basados en una tabla de DirectQuery para recuperar los datos del pedido de ventas (los detalles).