

Cuándo hacer combinaciones y cuándo hacer uniones

Bienvenido a este video en el que se comparan las combinaciones de datos con las uniones entre bases de datos.

Puede descargar el libro de trabajo de ejercicios para seguir la presentación en su propia copia de Tableau.

Las uniones entre bases de datos integran dos fuentes de datos en el nivel de fila para crear una única fuente.

La combinación agrega en la vista resultados de consultas independientes a cada fuente de datos. Para obtener información específica sobre la unión entre bases de datos y la combinación de datos, vea los videos individuales correspondientes.

Relaciones de uno a varios

Para comenzar, haremos una combinación con las fuentes de datos Products 2016 y Sales 2016 del video sobre unión entre bases de datos. Hagamos un análisis rápido de las ventas de cada categoría por mercado.

- Desde la fuente de datos de **productos**, arrastramos Categoría hasta Filas.
- Después, vamos a la fuente de datos de **ventas**.
 - Hacemos clic en el vínculo junto a ID de producto para establecer la relación de combinación.
 - Y arrastramos Ventas hasta Columnas.
- Todo se ve bien. Tenemos la combinación y podemos observar las ventas generales por categoría.
- Pero ¿qué sucede si arrastramos Mercado hasta Color? Ahora, vemos el color gris en la mayor parte de las barras. Eso indica que hay asteriscos. ¿Por qué?

Hicimos la combinación por ID de producto, que tiene una relación de uno a varios con Mercado. La parte azulada clara de las barras corresponde a ID de productos que solo se venden en el mercado USCA. Lo que se ve en gris se vende en varios mercados y está representado con un asterisco en la agregación. En la combinación, los resultados de cada fuente de datos se agregan por separado y se muestran juntos en la vista. Por lo tanto, no se pueden visualizar los datos fácilmente cuando hay una relación de uno a varios.

En este caso, la unión entre bases de datos es más apropiada que la combinación. Esta fuente de datos está unida en el nivel de fila por ID de producto. Recreemos esa primera vista.

- Categoría a Filas
- Ventas a Columnas.
- Mercado a Color.

Ahora obtenemos el comportamiento deseado. Las ventas de cada categoría se dividen por mercado. Cuando unimos los datos, complementamos la información de ventas con la información de productos, sin modificar el número original de registros de Ventas.

Uniones que inflan el número de filas

Entonces ¿la unión entre bases de datos es siempre la mejor opción? Usemos otro par de conjuntos de datos: Office City y Coffee Chain del video sobre combinación de datos. Estas son dos cadenas minoristas de la misma empresa matriz. Office City vende sus productos en los 50 estados de Estados Unidos, y Coffee Chain, en 20 estados. Supongamos que queremos ver las ventas combinadas de ambas organizaciones en los 50 estados.

Aquí tenemos una fuente de datos unida por el campo común Estado.

- Usaremos Estado de la conexión Office City de esta fuente de datos, ya que contiene los 50 estados.

- Si hacemos clic con el botón secundario en el cálculo “Ventas agregadas” y lo editamos, vemos que hemos agregado las ventas de Office City y Coffee Chain.
 - Hicimos esa agregación con la función Zero Null (cero-nulo). Por lo tanto, en los estados en que Coffee Chain no ha realizado ventas, el valor nulo se trata como cero.
- Arrastremos Ventas agregadas hasta Columnas.
- Oh. Estas empresas son realmente prósperas si en algunos estados las ventas alcanzan los cientos de millones de dólares. Estoy segura de que existe algún error. Hagamos un análisis más profundo.

Este dashboard muestra las ventas de Office City en 4 estados y las ventas de Coffee Chain en los mismos 4 estados. Las ventas combinadas deberían ser las que vemos en la ecuación del centro, que simplemente agrega los valores.

- Las ventas combinadas (amarillo) proporcionan la respuesta correcta. La fuente de datos unida (púrpura) contiene valores muy exagerados.

Si observamos el número de registros de Office City, Coffee Chain y la fuente de datos unida, vemos que ese valor está claramente inflado.

Dado que hicimos la unión por Estado, Tableau creó una fila para cada combinación de Estado con información única: Producto y Fecha de Coffee Chain e ID de fila de Office City. Con esta multiplicación, contamos los valores de las ventas varias veces.

Por el contrario, la combinación ofrece los valores correctos. Esta proporciona los valores de ventas de una fuente de datos para Estado y los valores de ventas de la otra fuente de datos para esos estados. Nosotros los agregamos. Ese es el comportamiento deseado para este caso.

Como puede ver, el campo elegido para hacer la unión o la combinación tienen un ENORME impacto sobre el análisis. Debemos elegir el campo correcto. Además, debemos comprender cómo se crean las combinaciones y las uniones para saber cuál elegir en cada situación.

Beneficios de cada una

Cuando se realiza una unión entre bases de datos para crear una fuente de datos integrada, se obtienen muchos beneficios.

- Dado que la unión se hace en el nivel de fila, no aparecen las limitaciones que afectan a la combinación de datos. Es decir, no se generan asteriscos.
- A diferencia de los datos combinados, la fuente de datos única puede extraerse, guardarse, publicarse y compartirse fácilmente. Se usa como cualquier otra fuente de datos.

Sin embargo, si la granularidad del campo de unión no es la correcta, la unión puede inflar el tamaño del conjunto de datos y proporcionar información engañosa.

La combinación de datos envía consultas independientes a cada fuente de datos y, después, hace la agregación con el nivel de granularidad deseado para la vista. Eso puede ser exactamente lo que su análisis necesita.

Conclusión

Gracias por ver este video sobre la combinación y la unión de bases de datos. Lo invitamos a continuar viendo los videos gratuitos de capacitación y obtener más información acerca de Tableau.