

Quando combinar e quando unir

Bem-vindo a este vídeo que compara combinações de dados e uniões entre bancos de dados.

Você pode baixar a pasta de trabalho de exercícios para acompanhar os procedimentos usando sua própria cópia do Tableau.

Resumidamente, uma união entre bancos de dados consiste em unir as linhas de duas fontes de dados para criar uma única fonte de dados. Já a combinação agrega na exibição diversos resultados das consultas feitas em cada fonte de dados. Para obter mais informações sobre uniões entre bancos de dados e combinação de dados, assista aos seus respectivos vídeos.

Relações um para muitos

Para começar, farei uma combinação com as fontes de dados Produtos 2016 e Vendas 2016 que usamos no vídeo sobre uniões entre bancos de dados. Vamos analisar rapidamente as vendas das categorias por mercado.

- Selecione a fonte de dados **Produtos** e arraste a dimensão Categoria para Linhas.
- Agora, selecione a fonte de dados **Vendas**.
 - Clique no símbolo de vínculo ao lado de ID do produto para criar a relação da combinação
 - e arraste Vendas para Colunas.
- Isso parece ótimo! Temos nossa combinação e podemos ver o total das vendas por categoria.
- E se arrastássemos Mercado para Cor? A maioria das barras agora está em cinza, o que indica um asterisco. Por quê?

Fizemos a combinação com a dimensão ID do produto, que tem uma relação de um para muitos com Mercado. A parte das barras em azul claro mostra os IDs dos produtos vendidos apenas no mercado USCA, e a parte em cinza se refere àqueles que foram vendidos em vários mercados, representados em uma agregação com um asterisco. Devido à forma como uma combinação é feita, agregando os resultados de cada fonte de dados separadamente e mostrando-os juntos na exibição, não conseguimos visualizar facilmente os dados quando há uma relação de um para muitos.

Nesse caso, a união entre bancos de dados é uma opção melhor do que a combinação. A união desta fonte de dados foi feita no nível da linha por ID do produto. Vamos recriar essa primeira exibição.

- Categoria para Linhas.
- Vendas para Colunas.
- Mercado para Cor.

Agora temos o resultado que queríamos; as categorias estão detalhadas por mercado. Quando unimos os dados, simplesmente complementamos as informações dos produtos com as informações das vendas, mantendo o número de registros de vendas inalterado.

Uniões aumentam o número de linhas

A união entre bancos de dados é sempre a melhor opção? Vamos usar para outro par de conjuntos de dados. Desta vez, as fontes de dados Office City e Coffee Chain que usamos no vídeo sobre combinação dos dados. Essas fontes de dados contêm informações de duas redes de varejo de uma mesma empresa.

Office City efetua vendas em todos os 50 estados dos EUA, e Coffee Chain em 20 estados. Vamos supor que queremos ver as vendas combinadas dessas duas organizações nos 50 estados.

Esta fonte de dados é uma união das fontes de dados Office City e Coffee Chain, criada a partir do campo Estado presente nas duas.

- Arrastaremos o campo Estado de Office City dessa fonte de dados, porque ele contém todos os 50 estados.
- E se clicarmos com o botão direito do mouse no cálculo “Vendas adicionadas” e selecionarmos Editar, podemos ver que adicionamos as vendas de Office City e Coffee Chain
 - com a função Zero Null. Assim, nos estados em que nenhuma venda de Coffee Chain foi efetuada, o valor nulo será substituído por zero.
- Agora, arrastaremos Vendas adicionadas para Colunas.
- Uau. O desempenho dessas empresas é excelente! Podemos ver centenas de milhões em vendas em alguns estados. O problema é que sei que essas informações não estão corretas. Vamos analisar isso mais a fundo.

Este painel mostra as vendas de Office City em quatro estados e as vendas de Coffee Chain nesses mesmos estados. As vendas combinadas são o que vemos aqui no meio: a soma dos valores das vendas.

- As vendas combinadas (em amarelo) fornecem informações precisas, mas a fonte de dados criada com a união (em roxo) mostra valores muito mais altos.

Se analisarmos o número de registros de Office City, de Coffee Chain e da fonte de dados que é a união dessas duas fontes, vemos um aumento significativo.

Como fizemos a união por Estado, o Tableau criou uma linha para cada combinação de Estado com uma informação única: as dimensões Produto e Data de Coffee Chain e ID da linha de Office City. A repetição desses produtos significa que estamos contabilizando os valores das vendas diversas vezes.

Já a combinação forneceu os valores corretos, porque retornou os valores das vendas de uma fonte de dados para os estados, o valor das vendas da outra fonte de dados para esses mesmos estados e apenas somamos esses valores. Esse é o resultado que desejamos neste caso.

Como podemos ver, o campo que usamos para fazer uma união ou uma combinação pode afetar BASTANTE a análise. Escolher o campo certo e entender como a união ou a combinação será realizada é muito importante para saber quando usar cada um desses recursos.

Benefícios da união e da combinação

Ao realizar uma união entre bancos de dados para criar uma fonte de dados integrada, você obtém diversos benefícios.

- Como as uniões são feitas no nível da linha, elas não são limitadas pela relação um para muitos presente na combinação de dados, ou seja, nenhum asterisco é gerado.
- A fonte de dados resultante da união pode ser extraída, salva, publicada e facilmente compartilhada com outras pessoas, diferentemente da combinação de dados. É como se essa fonte de dados fosse original.

No entanto, se o nível de detalhamento do campo a partir do qual a união é feita não estiver correto, criar uma união pode aumentar artificialmente o tamanho do conjunto de dados e resultar em informações incorretas.

A combinação de dados envia as consultas individuais para cada fonte de dados e as agrega, com o nível de detalhamento desejado, na exibição. Isso pode ser exatamente o que a análise precisa.

Conclusão

Obrigado por assistir a este vídeo sobre como criar fontes de dados com combinações ou uniões entre bancos de dados. Sugerimos que você continue assistindo aos vídeos de treinamento gratuitos para aprender mais sobre como usar o Tableau Public.