

データブレンディング

データブレンディングのトレーニングへようこそ。付属の練習ワークブックをダウンロードして、実際に Tableau を操作してみてください。

データブレンディングとは

まず、Office City と Coffee Chain という 2 つの小売チェーンを所有しているとしましょう。Office City と Coffee Chain のデータは、それぞれ異なるシステムに収集されています。そのような複数のデータソースのデータを 1 つのビューにまとめる一手段が、データブレンディングです。データブレンディングは、クロスデータベース結合のように行レベルでデータを結合するのではなく、別々のクエリをそれぞれのデータソースに送り、その結果を Tableau で共通したレベルに集計します。

ブレンドリレーションシップ

ここでは、[データ] ペインに 2 つのデータソースがあります。[Office City] と [Coffee Chain] です。この 2 つは簡単に切り替えられるので見てみると、ディメンションやメジャーに違いがあることがわかります。データブレンディングには、2 つのデータソースに共通したフィールドが 1 つ以上必要です。この場合は [州] が共通しています。フィールド名が異なる場合でも、共通したメンバーがあればリレーションシップを手動で定義することができます。

この例では、[Office City] の [地域] と [Coffee Chain] の [地区] は、どちらも [中部]、[東部]、[南部]、[西部] という値を持っているので、この 2 つのフィールドが同等だと定義することにしましょう。[データ] メニューで [リレーションシップの編集] を選択し、[Office City] をプライマリにして [カスタム] を選択し、リレーションシップを追加します。指定するのは [地域] と [地区] です。

すると、Tableau はこの 2 つのフィールドの間にリレーションシップを設定するだけでなく、[州] にも自動的にリレーションシップを設定しました。これ以外にも、一方のデータソースで、もう一方と一致するようにフィールド名を変更する方法があります。[Coffee Chain] の [地区] をゆっくりダブルクリックして、名前を「地域」に変更しましょう。フィールド名を同じにすると、Tableau は 2 つのデータソースの間に自動的にリレーションシップを設定します。

これで共通のフィールドの定義が終わったので、データブレンディングを始めましょう。

ブレンディングの方法

まず、データソースの [Office City] を選択します。[Office City] の [売上] を [列] シェルフに、[州] を [行] シェルフにドラッグします。すると [データ] ペインで、[Office City] の横に青いチェックマークが付きました。Tableau で複数のデータソースに接続した場合は、ビューにドラッグした最初のデータソースが必ずプライマリになり、このように青いチェックマークが付けられます。では、2 番目のデータソースに切り替えましょう。すると、[州] の横にはオレンジ色のリンクアイコンが表示されています。このオレンジ色のリンクは、すでに [州] がビューに追加されたため、Tableau は [州] で自動的にブレンドすることを示します。

また、[地域] フィールドの横には灰色のリンク切れアイコンが表示されています。[地域] は両方のデータソースに共通のフィールドなので、リンクフィールドになる可能性があります。現在のビューでは使用されていないという意味です。[地域] フィールドでもブレンドする場合は、この灰色のリンクをクリックすればオレンジ色になり、リレーションシップが作成されます。では、[Coffee Chain] の [売上] を [列] シェルフにドラッグして、データブレンディングを完成させましょう。すると、オレンジ色のチェックマークが、[Coffee Chain] データソースとビューのフィールドの横に表示されます。これは、セカンダリデータソースであることを示します。

このビューはどのように作成されているのでしょうか。Tableau は、州別の売上の合計をクエリとして各データソースで実行し、その値をまとめてビューに表示しています。[Office City] のすべての州で売上の情報が表示されているのは、セカンダリデータソースにデータがあるかどうかにかかわらず、Tableau はプライマリデータソースの全フィールドメンバー、つまりすべての州の情報を返すからです。

プライマリとセカンダリのデータソースの切り替え

ここで重要なのは、プライマリとセカンダリのソースはワークシートごとに指定するものであり、ワークブック全体で設定されるのではないという点です。ですから新しいシートを表示すると、[データ] ペインのデータソースに、プライマリとセカンダリを区別するオレンジ色や青いチェックマークは付いていません。前のワークシートで設定したリレーションシップは引き継がれていないからです。

では、データソースを入れ替えて [Coffee Chain] をプライマリソースに、[Office City] をセカンダリにするとどうなるのでしょうか。まず、[Coffee Chain] の [売上] をビューにドラッグした後、[州] もドラッグします。先ほどの例と比べると、このビューには半分ほどの州しか表示されていません。これは、[Coffee Chain] のデータセットでは、州の数が [Office City] より少ないためです。Tableau は [Coffee Chain] のすべての州を表示した後、[Office City] で対応する州の情報だけを表示します。ですから、[Office City] にあっても [Coffee Chain] にない州は、このビューに表示されません。

最後に

データブレンディングのトレーニングビデオをご視聴いただき、ありがとうございます。「その他のデータブレンディングに関するトピック」のビデオなど、Tableau の使用方法について引き続き無料のトレーニングビデオをご覧ください。