

Datenverschmelzung

Willkommen bei diesem Video zur Datenverschmelzung. Sie können die Übungsarbeitsmappe herunterladen und zum Nachverfolgen Ihre eigene Version von Tableau nutzen.

Was ist eine Datenverschmelzung?

Stellen Sie sich vor, Sie sind Eigentümer von zwei Einzelhandelsketten: „Office City“ und „Coffee Chain“. Sie erfassen Ihre Daten für „Office City“ in einem System und für „Coffee Chain“ in einem anderen. Die Datenverschmelzung bietet nun eine Möglichkeit zur Kombination von Daten aus mehreren Datenquellen in einer einzigen Ansicht. Bei der Datenverschmelzung werden die Daten nicht wie bei einer datenbankübergreifenden Verknüpfung auf Zeilenebene verbunden. Stattdessen werden gesonderte Abfragen an die einzelnen Datenquellen gestellt und die Ergebnisse auf einer allgemeinen Ebene zurück in Tableau aggregiert.

Die Beziehungen der Datenverschmelzung

Im Datenbereich befinden sich zwei Datenquellen: „Office City“ und „Coffee Chain“. Wir können auf einfache Weise zwischen den beiden wechseln und sehen dabei, wie sich die Dimensionen und Kennzahlen ändern. Eine Datenverschmelzung erfordert mindestens ein gemeinsames Feld zwischen beiden Datenquellen, z. B. „Bundesstaat“. Wenn die Feldnamen unterschiedlich sind, die Felder aber gemeinsame Elemente enthalten, können wir diese Beziehung manuell definieren.

Wir wissen beispielsweise, dass „Region“ aus „Office City“ und „Gebiet“ aus „Coffee Chain“ die Werte „Mitte“, „Ost“, „Süd“ und „West“ enthalten. Somit sind beide Felder äquivalent. Wir rufen das Menü „Daten“ auf und wählen „Beziehungen bearbeiten“ aus. Dann ändern wir die primäre Datenquelle in „Office City“, wählen „Benutzerdefiniert“ aus und fügen eine Beziehung hinzu. Für diese wählen wir „Region“ und „Gebiet“ aus.

Beachten Sie, dass Tableau eine Beziehung zwischen den beiden Feldern eingerichtet hat und die automatische Beziehung „Bundesstaat“ anzeigt. Alternativ kann auch das Feld einer Datenquelle in den Namen des Feldes der anderen umbenannt werden. Wir doppelklicken auf „Gebiet“ aus „Coffee Chain“ benennen das Feld in „Region“ um. Wenn die Namen identisch sind, erstellt Tableau automatisch eine Beziehung zwischen den beiden Datenquellen.

Da jetzt die gemeinsamen Felder definiert sind, können wir mit der Verschmelzung beginnen.

So werden Daten verschmolzen

Als Erstes wählen wir die Datenquelle „Office City“ aus. Wir ziehen den Umsatz von „Office City“ in den Spalten-Container und „Bundesstaat“ in den Zeilen-Container. Beachten Sie, dass nun im Datenbereich ein blaues Häkchen neben „Office City“ angezeigt wird. Bei einer Verbindung mit mehreren Datenquellen in Tableau wird immer die erste Datenquelle, die wir in die Ansicht aufnehmen, zur primären. Dies wird durch dieses blaue Häkchen angezeigt. Wechseln wir nun zur zweiten Datenquelle. Beachten Sie das orange Link-Symbol neben „Bundesstaat“. Da wir „Bundesstaat“ bereits in die Ansicht eingefügt haben, verwendet Tableau dieses Feld automatisch als Basis für die Datenverschmelzung. Dies wird die orangefarbene Verknüpfung angezeigt.

Es ist auch eine graue, unterbrochene Verknüpfung neben dem Feld „Region“ vorhanden. Da das Feld „Region“ in beiden Datenquellen enthalten ist, kann es ebenfalls für die Verschmelzung verwendet werden. In der aktuellen Ansicht wird es aber nicht genutzt. Wenn wir auch anhand dieses Feldes Daten verschmelzen wollen, können wir auf die graue Verknüpfung klicken, um sie zu aktivieren und die Beziehung zu erstellen. Wir schließen nun die Verschmelzung ab. Dazu ziehen wir „Umsatz“ von „Coffee Chain“ in den Spalten-Container. Die Ansicht enthält nun ein orangefarbenes Häkchen neben der Datenquelle „Coffee Chain“ sowie im Feld „SUM(Umsatz)“ in der Ansicht. Dieses Häkchen zeigt an, dass das Feld aus der zweiten Datenquelle stammt.

Wie wurde diese Ansicht erstellt? Tableau führt für jede Datenquelle eine Abfrage der Umsatzsumme nach Bundesstaat durch und stellt diese Werte zusammen in der Ansicht dar. Wir sehen nun Umsatzinformationen für jeden Bundesstaat aus „Office City“, da Tableau die Informationen für alle Feldelemente (Bundesstaaten) der primären Datenquelle zurückgibt. Dies ist unabhängig davon, ob die sekundäre Datenquelle entsprechende Daten enthält.

Primäre und sekundäre Datenquellen austauschen

Beachten Sie, dass die primäre und die sekundäre Datenquelle auf Basis der Arbeitsblätter festgelegt werden und nicht global in die Arbeitsmappe übernommen werden. In einem neuen Blatt verfügen die Datenquellen im Datenbereich nicht über orangefarbene und blaue Häkchen, die sie als primäre und sekundäre Datenquelle kennzeichnen. Die eingerichteten Beziehungen des vorherigen Arbeitsblattes werden nicht übernommen.

Was passiert nun, wenn wir unsere Datenquellen tauschen und „Coffee Chain“ zur primären und „Office City“ zur sekundären Datenquelle machen? Wir ziehen zuerst „Coffee Chain“ und dann „Bundesstaat“ in die erste Ansicht. In der Ansicht sind nun nur noch etwa halb so viele Bundesstaaten wie im vorherigen Beispiel enthalten. Der Grund dafür ist, dass der Datensatz von „Coffee Chain“ weniger Bundesstaaten enthält als der Datensatz von „Office City“. Tableau zeigt alle Bundesstaaten aus „Coffee Chain“ an und ruft dann nur die relevanten Informationen aus „Office City“ ab. Das bedeutet, dass Bundesstaaten in „Office City“, die nicht in „Coffee Chain“ vorkommen, nicht in dieser Ansicht angezeigt werden.

Fazit

Vielen Dank, dass Sie sich dieses Schulungsvideo über die Datenverschmelzung angesehen haben. Sehen Sie sich auch unser Video „Zusätzliche Themen zur Datenverschmelzung“ sowie unsere anderen kostenlosen Schulungsvideos zur Nutzung von Tableau an.