

# Wann verschmelzen (Blend) und wann verknüpfen (Join)

---

Willkommen bei diesem Video über den Vergleich von Datenverschmelzungen mit datenbankübergreifenden Verknüpfungen.

Sie können die Übungsarbeitsmappe herunterladen und zum Nachverfolgen Ihre eigene Version von Tableau nutzen.

Kurz gesagt, integrieren datenbankübergreifende Verknüpfungen zwei Datenquellen auf Zeilenebene, um eine einzige Datenquelle zu erstellen. Bei der Verschmelzung werden separate Abfrageergebnisse von jeder Datenquelle in der Ansicht zusammengefasst (aggregiert). Weitere Informationen speziell über die datenbankübergreifende Verknüpfung und Datenverschmelzung finden Sie in den jeweiligen Videos zu diesen Themen.

## Einer-zu-viele-Beziehungen

Als Erstes richten wir eine Verschmelzung mit unseren Datenquellen „Produkt 2016“ und „Umsatz 2016“ aus dem Video zu datenbankübergreifenden Verknüpfungen ein. Zunächst einmal wollen wir kurz unseren Umsatz pro Kategorie nach Märkten analysieren.

- Wir rufen die Datenquelle **Produkt** auf und ziehen „Kategorie“ auf den Zeilencontainer.
- Anschließend wechseln wir zur Datenquelle **Umsatz**.
  - Wir klicken auf den Link nach Produktkennung und erstellen die Verschmelzungsbeziehung.
  - Dann ziehen wir „Umsatz“ auf den Spaltencontainer.
- Das sieht prima aus. Wir haben unsere Verschmelzung, und wir können uns einen Eindruck von unserem Gesamtumsatz pro Kategorie verschaffen.
- Aber was geschieht, wenn wir „Markt“ auf „Farbe“ ziehen? Die meisten unserer Balken sind jetzt grau, was ein Sternchen bedeutet. Wie kommt das?

Wir haben die Verschmelzung anhand der Produktkennung ausgeführt. Diese hat eine Einer-zu-viele-Beziehung zu „Markt“. Die blassgrünen Balken zeigen Produktkennungen, die nur auf dem USCA-Markt abgesetzt werden, aber alle grauen Balken beziehen sich auf Produktkennungen, die auf mehreren Märkten abgesetzt werden. Diese werden in der Zusammenfassung durch ein Sternchen dargestellt. Aufgrund der Art und Weise, wie eine Verschmelzung ausgeführt wird, sprich: Zusammenfassung der Ergebnisse aus den einzelnen Datenquellen und zusammengefasste Darstellung in der Ansicht, können wir Daten, für die eine Einer-zu-viele-Beziehung besteht, nicht ohne weiteres visualisieren.

In diesem Fall liefert eine datenbankübergreifende Verknüpfung bessere Ergebnisse als die Verschmelzung. Diese Datenquelle wird auf der Zeilenebene für „Produktkennung“ verknüpft. Wir erstellen die erste Ansicht noch einmal neu.

- „Kategorie“ auf „Zeilen“
- „Umsatz“ auf „Spalten“
- „Markt“ auf „Farbe“

Wir haben jetzt das gewünschte Verhalten: Unser Umsatz pro Kategorie sind nach Märkten aufgeschlüsselt. Als wir die Daten verknüpft haben, haben wir die Umsatzdaten einfach durch die Produktdaten ergänzt. Die ursprüngliche Anzahl der Datensätze blieb dabei unverändert.

## Verknüpfungen, durch die die Anzahl der Zeilen erhöht wird

Ist eine datenbankübergreifende Verknüpfung demnach immer die beste Option? Sehen wir uns ein anderes Datensatzpaar an. Dieses Mal verwenden wir „Office City“ und „Coffee Chain“ aus dem Video zum Thema „Datenverschmelzung“. Es handelt sich um zwei Einzelhandelsketten im Besitz desselben Konzerns. Office City verkauft

in allen 50 US-Bundesstaaten und Coffee Chain verkauft in 20 Bundesstaaten. Nehmen wir einmal an, wir möchten unseren kombinierten Umsatz von beiden Organisationen in allen 50 US-Bundesstaaten ansehen.

Hier haben wir eine verknüpfte Datenquelle. Die Verknüpfung wurde anhand des gemeinsamen Felds „Bundesstaat“ vorgenommen.

- Wir verwenden „Bundesstaat“ aus der „Office City“-Verbindung in dieser Datenquelle, weil sie alle 50 Bundesstaaten enthält.
- Und wenn wir mit der rechten Maustaste auf „Zusätzlicher Umsatz“ klicken und die Berechnung bearbeiten, sehen wir, dass wir unseren Office City-Umsatz und unseren Coffee Chain-Umsatz addiert haben,
  - und zwar mit einer „keine Werte gleich null“-Funktion. Daher wird das Fehlen von Werten für Bundesstaaten, in denen Coffee Chain keinen Umsatz hat, als 0 behandelt.
- Dann ziehen wir „Zusätzlicher Umsatz“ zu „Spalten“.
- Fantastisch. Dies sind die BOOMENDEN Geschäfte, die Bundesstaaten, in denen unser Umsatz Hunderte Millionen betragen. Zufällig weiß ich aber, dass dies nicht stimmen kann. Gehen wir der Sache also auf den Grund.

Dieses Dashboard zeigt den Umsatz von Office City für vier Bundesstaaten und den Umsatz von Coffee Chain für dieselben vier Bundesstaaten. Unser kombinierter Umsatz sollte den Wert ergeben, den wir in den Gleichungen in der Mitte sehen, sprich: Die Werte sollten einfach addiert werden.

- Die verschmolzenen (gelben) kombinierten Umsätze liefern die richtige Antwort, aber die verknüpfte Datenquelle (lila) zeigt enorm übertriebene Werte.

Wenn wir uns die Anzahl der Datensätze für Office City, Coffee Chain und die verknüpfte Datenquelle ansehen, sehen wir, dass die Anzahl bei der verknüpften Datenquelle deutlich überhöht ist.

Da wir die Verknüpfung anhand von „Bundesstaat“ vorgenommen haben, hat Tableau eine Zeile für jede Kombination von Bundesstaaten erstellt, die eigene Daten enthalten: Produkt und Datum aus Coffee Chain sowie Zeilennummer aus Office City. Dieses Überkreuzprodukt bedeutet, dass wir unsere Umsatzwerte mehrmals gezählt haben.

Die Verschmelzung ergab hingegen die korrekten Werte, weil sie den Umsatzwert für eine Datenquelle für „Bundesstaat“ und den Umsatzwert für die andere Datenquelle für denselben Bundesstaat ergab und wir beide einfach addieren konnten. In diesem Fall ist dies das gewünschte Verhalten.

Wie Sie sehen, hängt das Analyseergebnis ENTSCHEIDEND davon ab, anhand von welchem Feld wir die Verknüpfung oder die Verschmelzung durchführen. Dass wir das richtige Feld auswählen und wissen, wie die betreffende Verschmelzung oder Verknüpfung erstellt wird, ist sehr wichtig für die Entscheidung darüber, wann wir welches Feld nutzen.

## Vorteile je nach Methode

Bei der Ausführung einer datenbankübergreifenden Verknüpfung zum Erstellen einer integrierten Datenquelle gibt es eine ganze Reihe von Vorteilen.

- Die Verknüpfung wird auf der Zeilenebene durchgeführt. Darum ist sie nicht mit all den Beschränkungen verbunden, die für die Datenverschmelzung charakteristisch sind, sprich: keine Sternchen mehr.
- Die einzelne Datenquelle kann im Gegensatz zur Datenverschmelzung extrahiert, gespeichert, veröffentlicht und ohne großen Aufwand mit anderen geteilt werden. Sie verhält sich einfach wie eine weitere Datenquelle.

Bei der Durchführung einer Verknüpfung kann allerdings die Größe Ihres Datensatzes künstlich aufgeblasen werden, wenn die Granularität des Verknüpfungsfelds nicht richtig ist. Dies kann zu irreführenden Informationen führen.

Eine Datenverschmelzung sendet separate Anfragen an jede Datenquelle und aggregiert diese auf der gewünschten Granularitätsebene in der Ansicht. Dies kann den Analyseanforderungen exakt entsprechen.

### **Fazit**

Vielen Dank, dass Sie sich dieses Schulungsvideo zur Kombination von Datenquellen durch Verschmelzungen oder datenbankübergreifende Verknüpfungen angesehen haben. Sehen Sie sich auch unsere anderen kostenfreien Schulungsvideos zur Nutzung von Tableau an.