

3 Importieren und Modellieren der Daten

Im vorigen Kapitel habe ich mich ausführlich mit den Benutzeroberflächenfunktionen der Ansichten *Bericht* und *Daten* befasst. Dieses Kapitel erörtert die Benutzeroberfläche von Power Query und der Ansicht *Modell*. Am Ende dieses Kapitels werde ich die Daten des Kurses an der (fiktiven) Cool School University in eine Power BI Desktop-Datei importiert haben und dann diese Daten verwenden, um einige der grundlegenden Funktionen von Power Query zur Datenmanipulation und -gestaltung zu demonstrieren.

Bei der Vorstellung der Ansicht *Modell* werde ich zeigen, was Beziehungen in einem Datenmodell sind und was sie bewirken. Zum Schluss gehe ich die Daten durch und probiere aktiv etwas damit aus. Wenn Sie praktisch veranlagt sind, empfehle ich Ihnen, die Beispieldaten zu nehmen und mir zu folgen, um die Schritte und die Ergebnisse nachzuvollziehen.

3.1 Die Daten abrufen

Für diese Übung werde ich mit drei Microsoft Excel-Dateien arbeiten, die Sie von der Website zum Buch herunterladen können (siehe den Abschnitt »Über die Begleitdateien« in der Einleitung dieses Buches). Wir versetzen uns in die Lage des Dozenten eines Kurses zur Datenvisualisierung an der Cool School University. Wir haben Dateien mit den Informationen, die uns die Uni zur Verfügung gestellt hat, Dateien mit Informationen, die wir selbst von unseren Studierenden erhalten haben, und Dateien mit den Noten, die die Studierenden für ihre Arbeiten und Aufgaben erhalten haben.

Ich beginne auf der Registerkarte *Start* im Menüband der Ansicht *Bericht* und verwende die Schaltfläche *Excel-Arbeitsmappe*, um die erste Datei mit dem Namen `CoolSchoolUniversity-DatenVonUniStart.xlsx` zu öffnen. Wenn Sie die Schaltfläche *Excel-Arbeitsmappe* anklicken, zeigt Power BI das Dialogfenster *Öffnen* an, in dem Sie zu dieser Excel-Datei navigieren. Wählen Sie die Datei aus und klicken Sie auf *Öffnen*.

Es öffnet sich das Fenster *Navigator*, das in der Regel so aussieht wie das in Abbildung 3-1. Ich schreibe »in der Regel«, denn wenn Sie eine Datenbank als Datenquelle verwenden, sehen Sie wahrscheinlich eine größere Anzahl von auswählbaren Datenelementen, abhängig von Ihrer Zugriffsebene auf die Datenbank.

Wenn Sie eine Verbindung zu einer Excel-Arbeitsmappe herstellen, sollten Sie alle Arbeitsblätter der Arbeitsmappe sehen.

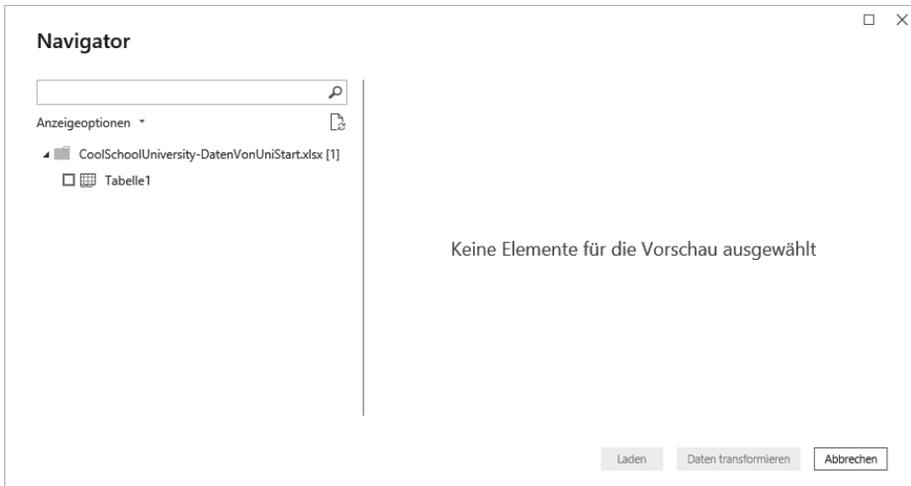


Abb. 3-1 Das Navigator-Fenster sollte sich öffnen, sobald Sie die Datei mit den von der Uni bereitgestellten Daten geöffnet haben.

Von hier aus möchte ich Ihre Aufmerksamkeit auf die untere rechte Ecke lenken, wo Sie Wenn Sie im Menü auf der linken Seite ein Datenelement auswählen, in unserem Fall *Tabelle1*, wird eine Vorschau der Daten angezeigt, wie in Abbildung 3-2 zu sehen. Beachten Sie, dass ich auf das Kästchen neben *Tabelle1* geklickt und es damit für den Import ausgewählt habe. Nur damit Sie es wissen: Sie müssen das Kontrollkästchen nicht aktivieren, um die Vorschau zu sehen.

sehen, dass die Schaltflächen *Laden* und *Daten transformieren* nicht mehr ausgegraut sind. Dies geschieht, nachdem ein Datenelement durch Anklicken des Kästchens ausgewählt wurde. Diese Datei hat zwar nur ein Arbeitsblatt, aber wenn die Arbeitsmappe mehrere Arbeitsblätter hätte, könnten Sie gleichzeitig mehrere Arbeitsblätter zum Importieren auswählen. Wie ich jedoch einmal von zwei weisen glatzköpfigen Männern gelernt habe, die praktischerweise in Würfeln leben, sollten Sie immer *Daten transformieren* wählen. Wenn Sie auf *Laden* klicken, werden die Daten genau so übernommen, wie sie in der Vorschau angezeigt werden, und in das Datenmodell geladen. Nichts hindert Sie daran, später zu Power Query zurückzukehren und die Tabelle zu bearbeiten, aber sobald Sie *Laden* wählen, werden alle anderen Prozesse angehalten, damit Power BI die Daten in den VertiPaq-Datenspeicher bringen kann.

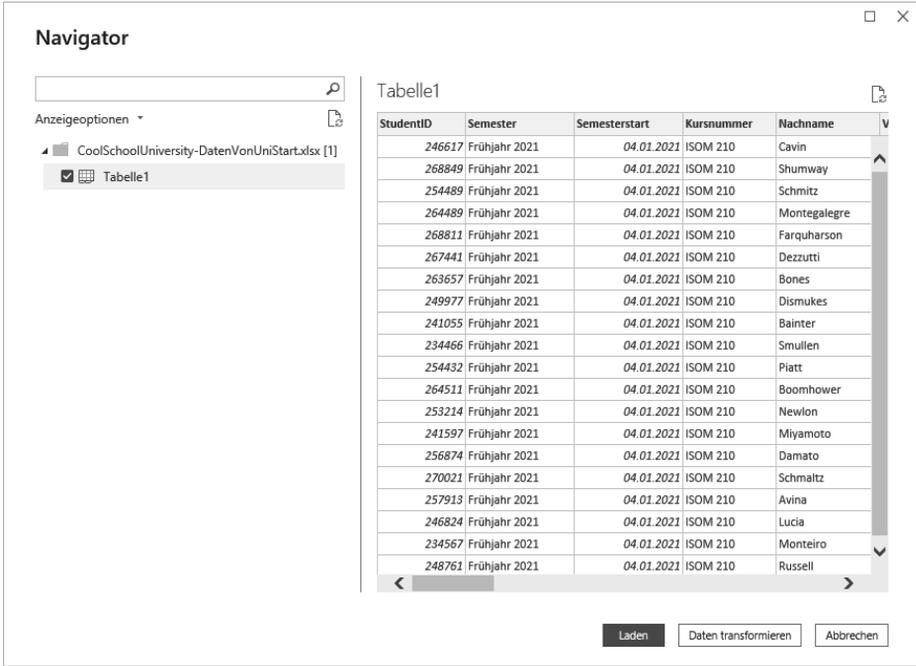


Abb. 3-2 Sobald Sie im Menü auf der linken Seite ein Datenelement auswählen, sehen Sie auf der rechten Seite eine Vorschau der Daten.

Meine Faustregel lautet, dass es nie schadet, sich eine zusätzliche Sekunde Zeit zu nehmen, auf *Daten transformieren* zu klicken und in Power Query noch einmal zu überprüfen, ob die Daten so geformt und eingegeben sind, wie Sie es erwarten. Schauen Sie sich an, was passiert, wenn Sie auf *Daten transformieren* klicken (siehe Abbildung 3-3).

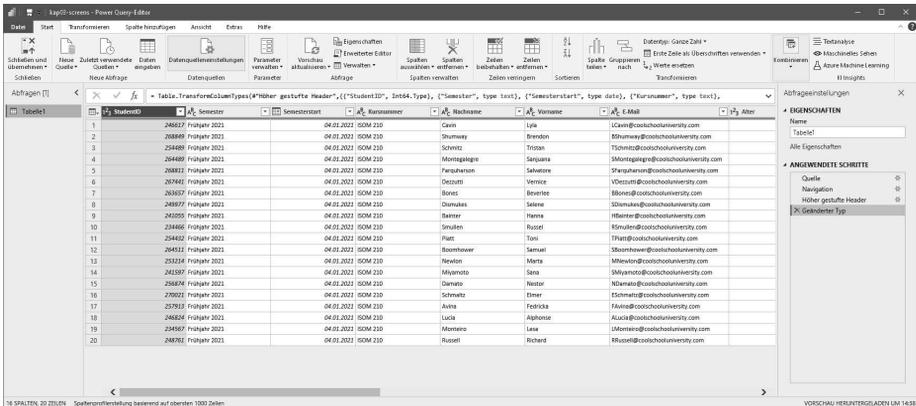


Abb. 3-3 In Power Query werde ich die Daten an unsere Bedürfnisse anpassen.

Ein neues Fenster erscheint. Es zeigt eine Menge Dinge an: unsere zukünftige Tabelle (derzeit *Tabelle1*), eine Vorschau der Daten, eine ganze Reihe von Datenum-

wandlungsoptionen und die derzeit angewandten Umwandlungsschritte für das, was Power BI jetzt eine Abfrage nennt. Eine *Abfrage* ist ein Datenelement, das in unser Datenmodell importiert werden soll.

Bevor ich mich mit dem ganzen Schnickschnack von Power Query beschäftige, möchte ich die beiden anderen Excel-Dateien in Power Query einbinden und ihnen Namen geben, die sinnvoller sind als *Tabelle1* und *Tabelle2*. Ohne das Power Query-Fenster zu verlassen, wählen Sie *Neue Quelle | Excel-Arbeitsmappe* und fügen Sie *CoolSchoolUniversity-VonDozentErfassteInfosStudenten.xlsx* hinzu. Sind Sie fertig? Wiederholen Sie dies nun mit der Arbeitsmappe *CoolSchoolUniversity-Noten.xlsx*, um so die Noten für das Semester zu erhalten.

Es ist erwähnenswert, dass beim Hinzufügen von Datenquellen aus Power Query heraus im Fenster *Navigator* keine Optionen zum *Laden* und *Transformieren von Daten* angezeigt werden. Es gibt nur eine einzige OK-Schaltfläche und die Schaltfläche *Abbrechen*. Aber keine Sorge: Wenn Sie im *Navigator*-Fenster von Power Query aus auf OK klicken, verhält es sich so, als hätten Sie *Daten transformieren* gewählt.

An dieser Stelle haben Sie wahrscheinlich bemerkt, dass die Datei mit den Noten anders ist als die ersten beiden! Tut mir leid, ich musste hier ein wenig die Kurve kriegen. Diese Datei hat in der Tat zwei Arbeitsblätter. Sie wissen von früher, dass Sie beide Arbeitsblätter einbinden können, indem Sie beide auswählen und auf OK klicken.

Schauen Sie sich jedoch die *NotenPunkte*-Datenvorschau an, und Sie werden feststellen, dass sie zusätzliche Spalten enthält, die nicht benannt und null sind. Ich habe absichtlich ein wenig an dieser Datei herumgepfuscht, um Ihnen das Feature *Vorgeschlagene Tabellen* zu zeigen. Im *Navigator*-Fenster sehen Sie ein Dropdownmenü mit *Vorgeschlagenen Tabellen*. Wenn Sie sie auswählen und die Daten in der Vorschau anzeigen, sehen Sie, dass sie ohne die zusätzlichen Spalten, die Null-Werte enthalten, korrekt aussehen. Wählen Sie nun die *AufgabeDetails* von oben und *Tabelle1* von unten aus und fügen Sie beide ein. Wenn Sie herausfinden, was ich getan habe, um das Excel-Blatt zu verunstalten, gebe ich Ihnen 100 Punkte.

Bevor dieser Abschnitt endet, sollten Sie die Abfragen umbenennen, damit sie einen Sinn ergeben. Die Abfrage *AufgabeDetails* hat ihren Arbeitsblattnamen behalten, was hilfreich ist, aber die anderen sollten Sie noch umbenennen. Denken Sie daran, dass die Umfragedaten aus der Datei *CoolSchoolUniversity-VonDozentErfassteInfosStudenten.xlsx* stammen und die von der Uni bereitgestellten demografischen Daten aus der Datei *CoolSchoolUniversity-DatenVonUniStart.xlsx*. Benennen Sie diese um, indem Sie entweder mit der rechten Maustaste auf die Tabelle in der Abfrageansicht klicken und *Umbenennen* wählen oder den Namen im Eigenschaftsfenster auf der rechten Seite von Power Query anpassen, während die Abfrage ausgewählt ist. Benennen Sie sie in **Demografische Daten**, **Umfragedaten** und **Noten** um.

Wenn Sie alles richtig gemacht haben, sollte Power Query so aussehen, wie es Abbildung 3–4 zeigt. Wenn es bei Ihnen anders aussieht, schlage ich vor, mit der

rechten Maustaste auf jedes Objekt zu klicken, *Löschen* zu wählen und von vorne zu beginnen, um herauszufinden, wo Sie einen Fehler gemacht haben. Wenn Sie alles zusammenhaben, klicken Sie auf die Schaltfläche *Schließen & übernehmen*, um die Daten zu laden. Um von der Ansicht *Bericht* in Power BI Desktop zum Power Query-Fenster zurückzukehren, klicken Sie auf der Registerkarte *Start* auf die Schaltfläche *Daten transformieren*. Sobald Sie auf *Schließen & übernehmen* klicken, werden die Daten geladen und dann in der Speicher-Engine komprimiert.

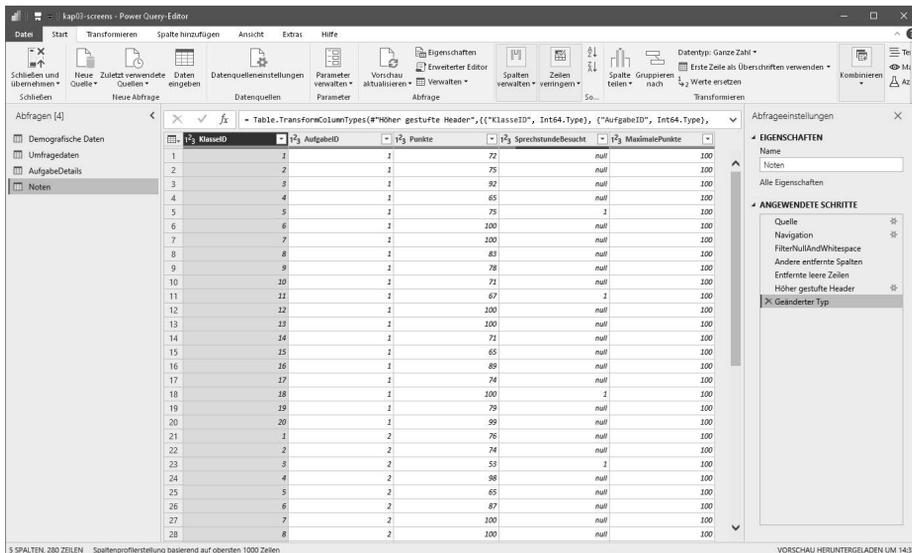


Abb. 3-4 Wenn Sie die einzelnen Schritte nachverfolgen, dann befinden Sie sich an dieser Stelle.

3.2 Das Menüband von Power Query-Editor

Jetzt haben Sie die Daten in Power Query importiert. Wenn Sie sich ein wenig im Menüband umsehen, werden Sie feststellen, dass es viele Möglichkeiten gibt, Ihre Daten zu transformieren. In dieser Einführung konzentriere ich mich im Menüband von Power Query auf die Registerkarten *Start*, *Transformieren* und *Spalte hinzufügen*. Wenn Sie eine Datentransformation durchführen möchten, diese aber nicht im Menüband finden, können Sie sie möglicherweise durch manuelle Codierung anstoßen. Das würde den Rahmen dieser Übung sprengen, aber wenn Sie mehr über M, die Power Query-Formelsprache, erfahren möchten, finden Sie die Dokumentation von Microsoft zu dieser Sprache unter <https://learn.microsoft.com/de-de/powerquery-m>.

Bevor ich fortfahre, hier noch kurz dies: Im folgenden Abschnitt wird häufig auf ausgewählte Spalten verwiesen. Sie können eine Spalte auswählen, indem Sie auf den Spaltentitel klicken. Wie Sie in Abbildung 3-4 sehen, habe ich die Spalte *KlasseID* ausgewählt. Wenn Sie mehrere Spalten auswählen möchten, können Sie eine Spalte auswählen, die -Taste gedrückt halten und auf eine andere Spalte

klicken. Dadurch werden alle Spalten zwischen diesen beiden Spalten ausgewählt. Um mehrere, nicht nebeneinander stehende Spalten auszuwählen, drücken Sie **[Strg]** und klicken Sie, um mehrere einzelne Spalten auszuwählen.

3.2.1 Die Registerkarte »Start«

Im Großen und Ganzen ist die Registerkarte *Start* in Power Query sehr übersichtlich. In Abbildung 3–5 sehen Sie alle Schaltflächen in ihren Gruppen; ich werde sie durchgehen und Ihnen einige Beispiele für ihre Verwendung geben.

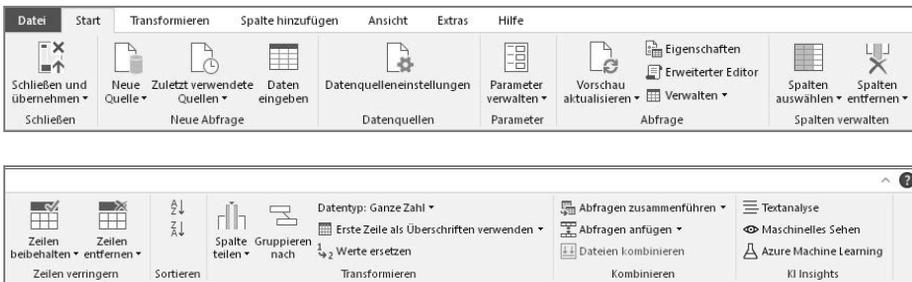


Abb. 3–5 Die Registerkarte *Start* ist der Ort, an dem sich das Herz befindet ... oder das Gehirn?

Die Schaltfläche *Schließen & übernehmen* übernimmt alle Änderungen, die Sie bis zu diesem Zeitpunkt vorgenommen haben, und wendet alle an, wobei alle Datenelemente, die davon betroffen sind, aktualisiert werden. Wenn Sie auf den Pfeil der Schaltfläche klicken, wird ein Dropdownmenü angezeigt, das Ihnen drei Optionen bietet:

- *Schließen & übernehmen*;
- *Übernehmen*, wodurch, wie zu erwarten, die Änderungen übernommen, Power Query aber nicht geschlossen wird; und
- *Schließen*, wodurch keine Änderungen übernommen werden. Wenn Sie sich für *Schließen* entscheiden, werden Sie von Power BI Desktop gewarnt, dass Sie noch nicht übernommene Änderungen haben.

In der Gruppe *Neue Abfrage* können Sie mit der Schaltfläche *Neue Quelle* neue Daten hinzufügen. Wenn Sie auf das Symbol klicken, wird das gesamte Menü zur Auswahl der Datenquelle angezeigt, während Sie durch Klicken auf den Pfeil ein Dropdownmenü mit den am häufigsten verwendeten Daten erhalten.

Mit der Schaltfläche *Zuletzt verwendete Quellen* können Sie schnell eine Verbindung zu allen Datenquellen herstellen, mit denen Sie in letzter Zeit gearbeitet haben, nicht nur zu denen, die sich in Ihrer aktuellen Power BI Desktop-Datei befinden. Dies kann nützlich sein, wenn Sie Tests durchführen und eine zweite Datei generieren und schnell Verbindungen herstellen möchten oder wenn Sie an einem neuen Modell arbeiten und es um etwas ergänzen möchten, mit dem Sie zuvor gearbeitet haben.

Mit *Daten eingeben* wird eine einfache Tabellenstruktur angezeigt, in der Sie Daten manuell einfügen und Spaltennamen erstellen. Ich empfehle Ihnen, *Daten eingeben* nicht sehr ausgiebig zu verwenden, aber es kann nützlich sein, um eine Mapping-Tabelle anzulegen, wenn Sie unterschiedliche Datenquellen haben, oder um eine schnelle zusätzliche Dimension zu erstellen, die möglicherweise an anderer Stelle in Ihren Daten nicht vorhanden ist. Die Oberfläche von *Daten eingeben* sieht zwar ein wenig so aus wie ein Arbeitsblatt in Excel, bietet aber keine Formelunterstützung und keine Regeln zur Datenüberprüfung. Es handelt sich lediglich um Zeilen und Spalten, und je mehr Daten Sie eingeben, desto mehr müssen Sie manuell verwalten. Deshalb warne ich generell davor, *Daten eingeben* zu verwenden, wenn es möglich ist, aber es ist da, wenn Sie es brauchen.

Über die Schaltfläche *Datenquelleneinstellungen* wird ein Fenster wie das in Abbildung 3–6 angezeigt. Von hier aus können Sie die Quellen ändern, eine PBIDS-Datei exportieren (was im Grunde eine Power BI Desktop-Verknüpfung für diese Datenquelle ist, die Sie in einer anderen Power BI-Datei öffnen können), Berechtigungen bearbeiten und Berechtigungen löschen.

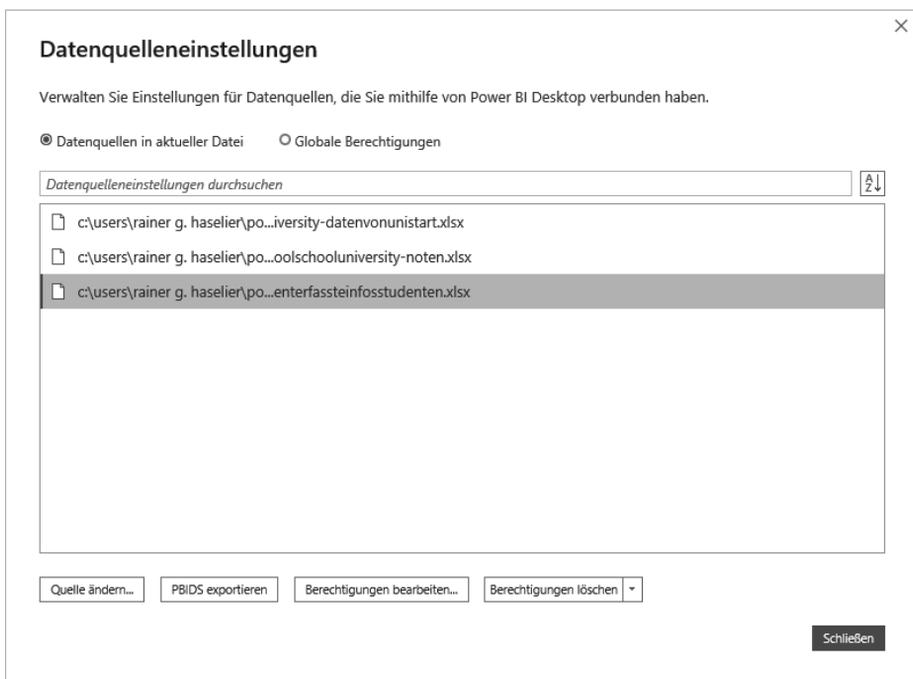


Abb. 3–6 Datenquelleneinstellungen – Dieser Befehl scheint nicht so wichtig zu sein, bis Sie ihn doch mal brauchen.

Mit *Parameter verwalten* können Sie neue Parameter erstellen sowie alle vorhandenen Parameter anzeigen und bearbeiten. In Power Query sind *Parameter* Werte, die in bestimmten Kontexten aufgerufen werden können, um die Art und Weise zu ändern, wie mit bestimmten Daten interagiert wird. Parameter haben eine Vielzahl von Einsatzmöglichkeiten, die ich später in Verbindung mit Was-wäre-

wenn-Analysen, Parameterregeln in Bereitstellungspipelines und mehr erörtern werde. Für den Augenblick reicht es, wenn Sie wissen, dass in Parametern viel Macht steckt und dass ich später darauf zurückkommen werde.

In der Gruppe *Abfrage* der Registerkarte *Start* können Sie die Vorschau aktualisieren, die Eigenschaften anzeigen, den erweiterten Editor öffnen oder eine Abfrage verwalten, wobei nicht ganz offensichtlich ist, was sich dahinter verbirgt.

Mit der Schaltfläche *Vorschau aktualisieren* können Sie die Abfrage, die Sie gerade markiert haben, in einer aktualisierten Vorschau der Daten anzeigen lassen. Dies kann hilfreich bei sich häufig ändernden Datasets sein und Sie sich verwisseln möchten, dass Sie immer noch das gewünschte Transformationsverhalten erhalten, obwohl sich die Daten ändern. Sie können auch auf den Dropdownpfeil klicken, um die Vorschau aller aktuellen Abfragen in Ihrem Modell zu aktualisieren.

Die Schaltfläche *Eigenschaften* öffnet ein Dialogfeld, in dem Sie den Namen der Abfrage, die Beschreibung der Abfrage und die beiden Kontrollkästchen *Laden in Bericht aktivieren* und *In Berichtsaktualisierung einbeziehen* einschalten können; standardmäßig sind diese Optionen immer aktiviert. Im Feld *Beschreibung* ist nichts zu sehen, bis eine Beschreibung hinzugefügt wurde; es kann hilfreich sein, eine Abfrage zu beschreiben.

Wenn *Laden in Bericht aktivieren* ausgeschaltet ist, wird die gesamte Tabelle so behandelt, als ob sie nicht existiert, und alle Visualisierungen, die sich auf diese Tabelle stützen, gehen verloren. Betrachten Sie dies als eine Art sanftes Löschen einer Tabelle aus Ihrem Datenmodell. Die Tabelle ist in Power Query weiterhin vorhanden, wenn Sie sie wieder aktivieren möchten, aber ansonsten behandelt Power BI sie so, als würde sie nicht existieren.

Die Aktualisierungsfunktion *In Berichtsaktualisierung einbeziehen* ist nicht ganz so intuitiv, wie sie sein sollte. Wenn *Laden aktivieren* deaktiviert ist, können Sie diese Tabelle nicht in die Aktualisierung des Berichts einbeziehen. Wenn diese Funktion deaktiviert ist, werden die Daten so belassen, wie sie derzeit im Modell vorhanden sind, aber die Abfrage wird nicht aktualisiert, wenn die Daten neu aktualisiert werden. Ich verwende dies im Allgemeinen, wenn eine Tabelle sehr groß ist und ich einige andere Änderungen bei einer Datenaktualisierung testen möchte, ohne die größte Tabelle hierfür neu zu laden. In anderen Anwendungsfällen werden bestimmte Daten zu einem bestimmten Zeitpunkt für Prüfungszwecke usw. aufbewahrt.

Im *Erweiterten Editor* sehen Sie den tatsächlichen M-Code, der bei jeder von Ihnen vorgenommenen Transformation erzeugt wird! Ich persönlich finde es toll, dass es eine Möglichkeit gibt, den Code zu sehen, der erstellt wird, wenn man eine Änderung vornimmt, vor allem, wenn man daran interessiert ist, die Sprache selbst zu lernen. M als Sprache ist meiner Meinung nach leicht zu erlernen, aber schwer zu beherrschen. Wenn Sie versuchen wollen, M direkt vom Editor aus zu ändern, sollten Sie eine Kopie Ihres aktuellen Codes in einer Textdatei oder einem anderen Backup anlegen, bevor Sie Änderungen vornehmen. Der Grund dafür ist, dass die Änderungen, die Sie im Editor vornehmen, wenn Sie auf *Fertig* klicken,

das ergeben, was Sie erhalten, egal ob Ihr Code funktioniert oder nicht. Eine gute Faustregel ist wie immer bei der Programmierung, dass man ein Backup haben sollte!

Die Schaltfläche *Verwalten* schließlich hat keinen sehr guten und aussagekräftigen Namen. Wenn Sie auf diese Schaltfläche klicken, sehen Sie drei Optionen: *Löschen*, *Duplizieren* und *Verweis*.

Beim *Duplizieren* einer Abfrage wird die Abfrage in ihrer aktuellen Form dupliziert. Das ist einfach und bedeutet tatsächlich, dass Sie die Daten zweimal speichern.

Verweis macht etwas ganz anderes. Dieser Befehl erstellt eine neue Abfrage, die auf den ersten Blick wie ein Duplikat aussieht, es aber nicht ist. Sie übernimmt die Abfrage, auf die sie verweist, in ihrer endgültigen Form und ermöglicht es Ihnen, von dort aus weiterzuarbeiten. Wenn Sie Änderungen an der übergeordneten Abfrage vornehmen, wird sich auch die referenzierte Abfrage ändern, was zu Fehlern in dieser Abfrage führen kann; seien Sie also vorsichtig.

Die Gruppe *Spalten verwalten* hat zwei Schaltflächen: *Spalten auswählen* und *Spalten entfernen*. Beide bringen Sie zum gleichen Ziel, aber auf unterschiedliche Weise.

Wenn Sie auf die Schaltfläche *Spalten auswählen* klicken, wird ein Dialogfeld angezeigt, in dem Sie auswählen können, welche Spalten Sie in der Abfrage behalten möchten; die anderen werden verworfen. Wenn Sie auf den Pfeil der Schaltfläche *Spalten auswählen* klicken, können Sie das eben erwähnte Dialogfeld *Spalten auswählen* aufrufen oder zu einer bestimmten Spalte in der Abfrage wechseln. Diese zweite Funktion kann nützlich sein, wenn Sie viele Spalten in einer Abfrage haben und sich nicht sicher sind, wo sie sich in der Spaltenreihenfolge befinden.

Wenn Sie auf die Schaltfläche *Spalten entfernen* klicken, wird die Spalte einfach entfernt. Sie erhalten kein Dialogfeld. Es werden alle ausgewählten Spalten in den Abgrund geworfen. Wenn Sie auf den Pfeil der Schaltfläche *Spalten entfernen* klicken, öffnet sich ein Menü, das eine zweite Option anbietet: *Andere Spalten entfernen*. Dabei werden alle nicht markierten Spalten entfernt.

Die Gruppe *Zeilen verringern* enthält zwei Schaltflächen: *Zeilen beibehalten* und *Zeilen entfernen*. Für jede dieser Schaltflächen gibt es eigene Anwendungsfälle für entweder die oberste Anzahl X Zeilen oder für einen Bereich von Zeilen ab Zeile X bis zur Zeile Y. Ich möchte Sie in dieser Gruppe vor der Möglichkeit warnen, mit dem entsprechenden Befehl der Schaltfläche *Zeilen entfernen* Duplikate zu entfernen. Es ist wichtig zu wissen, dass Duplikate auf der Grundlage der von Ihnen ausgewählten Spalten entfernt werden. Wenn Sie also lediglich eine Spalte ausgewählt haben und Duplikate entfernen, könnten Sie Daten verlieren, die Sie ursprünglich behalten wollten. In der Gruppe *Sortieren* können Sie eine Spalte markieren und dann in aufsteigender oder absteigender Reihenfolge sortieren. Dies funktioniert nicht, wenn mehrere Spalten ausgewählt sind.

Als Nächstes komme ich zur Befehlsgruppe *Transformieren*, die Ihnen einige grundlegende Transformationsoptionen anbietet. Mit *Spalten aufteilen* teilen Sie eine ausgewählte Spalte anhand eines von Ihnen gewählten Trennzeichens in

mehrere Spalten auf. Ein klassischer Anwendungsfall ist eine Spalte, die Vor- und Nachnamen enthält, die aber getrennt werden müssen.

Mit *Gruppieren nach* erstellen Sie eine Art Aggregation über Ihre aktuelle Abfrage. Dies kann nützlich sein, wenn Sie z.B. eine große Menge an Daten haben und eine Art Voraggregation vornehmen möchten, um die Anzahl der Zeilen zu reduzieren. In Abbildung 3–7 sehen Sie eine Beispielgruppierung, die unsere Notendaten verwendet, um den Mittelwert pro Aufgabe zu ermitteln.

Abb. 3–7 Gruppieren nach ermöglicht es, Daten mit einer Vielzahl von Operationen zu aggregieren.

Über die Schaltfläche *Datentyp* können Sie den Datentyp einer ausgewählten Spalte ändern. Es gibt viele verschiedene Datentypen und Power BI versucht, diese automatisch zu erkennen, wenn die Daten zum ersten Mal in Power Query eingegeben werden. Power BI erkennt die Datentypen sehr gut, aber manchmal ist es nicht perfekt. Wenn Sie also eine manuelle Anpassung vornehmen müssen, können Sie das hier tun. Den Datentyp einer Spalte erkennen Sie übrigens an dem Symbol in der Spaltenüberschrift sowie durch eine schnelle Überprüfung.

Erste Zeile als Überschriften verwenden ist ziemlich selbsterklärend, und Power BI wird je nach Datenquelle versuchen zu erkennen, ob die erste Zeile Spaltenüberschriften enthält. Wenn Sie auf den Pfeil neben der Schaltfläche *Erste Zeile als Überschriften verwenden* klicken, wird auch eine Option angezeigt, mit der Sie die Überschriften zur ersten Zeile machen, also den umgekehrten Vorgang auslösen.

Mit der Schaltfläche *Werte ersetzen* können Sie Werte in einer bestimmten Spalte suchen und ersetzen. Ein üblicher Anwendungsfall wäre hier, je nach Ihrer Analyse entweder die Zahl Null als Null-Wert oder Null-Werte wie die Zahl Null zu behandeln. Es ist wichtig zu beachten, dass Sie keine einzelnen Zellen bearbeiten können, weil Zellen ein Konzept sind, das Power Query nicht kennt; Power Query versteht nur Spalten.

In der Gruppe *Kombinieren* können Sie *Abfragen zusammenführen*, *Abfragen anfügen* und *Dateien kombinieren*. *Abfragen zusammenführen* funktioniert wie eine SQL JOIN-Anweisung. Angenommen, Sie haben Spalte A in Tabelle A

und Spalte A in Tabelle B und Sie möchten Spalte N aus Tabelle B zu Tabelle A hinzufügen. Eine JOIN-Anweisung besagt, dass, wenn der Wert in Spalte A von Tabelle A und Spalte A von Tabelle B gleich ist, der Wert von Spalte N ausgegeben und dieser hinzugefügt werden soll. Ich mache hier das Gleiche.

Der Verwendungszweck von *Abfragen anfügen* besteht darin, mehrere Abfragen, bei denen die Daten die gleiche Form und das gleiche Format haben, zu einer längeren Abfrage zusammenzufassen, ähnlich wie bei einer UNION-Anweisung in SQL.

Mit dem Befehl *Dateien kombinieren* können Sie dem Problem der Anfrageabfragen zuvorkommen, indem Sie mehrere Dateien in einer einzigen Ansicht vorkombinieren. Wenn Sie Spalten haben, die für jede Tabelle eindeutig sind, wird dieser Befehl alles zusammenführen, um ein einziges Ergebnis zu erhalten, das alle Spalten und Werte zwischen den beiden Tabellen umfasst. Achten Sie also beim Anhängen von Tabellen darauf, dass Sie das gewünschte Ergebnis erhalten.

Schließlich enthält die Gruppe *KI Insights* Elemente, die Ihr Unternehmen verwenden kann, um eine Verbindung zu Azure KI-Diensten herzustellen. Beachten Sie, dass für diese Funktionen Power BI Premium verwendet werden muss, damit sie aktualisiert werden können. Es ist sehr wahrscheinlich, dass Ihr Unternehmen nicht über diese Ressourcen verfügt. Daher werde ich keine Zeit darauf verwenden, außer zu erwähnen, dass es sie gibt. Fragen Sie Ihren Azure-Administrator, ob Ihr Unternehmen Azure Machine Learning Studio oder Cognitive Services verwendet, um mehr zu erfahren.

3.2.2 Die Registerkarte »Transformieren«

Die Registerkarte *Transformieren* enthält einige zusätzliche Optionen für die Transformation Ihrer Daten, die über das hinausgehen, was Sie in der Befehlsgruppe *Transformieren* auf der Registerkarte *Start* gesehen haben. Wie Sie im vorigen Kapitel gesehen haben, hat Microsoft jedoch auch hier einige Redundanzen in das Menüband eingebaut, auf die ich im Folgenden näher eingehen werde. Doch zunächst werfen Sie einen Blick auf Abbildung 3–8, die die gesamte Registerkarte *Transformieren* zeigt.

Zunächst sehen Sie in der Gruppe *Tabelle* die Befehle *Gruppieren nach* und *Erste Zeile als Überschriften verwenden*. Diese sind selbsterklärend.

Vertauschen nimmt Ihre Spalten und macht sie zu Zeilen (ein Vorgang, der auch Transponieren genannt wird). *Zeilen umkehren* nimmt die Reihenfolge der Daten nach Zeilen und dreht sie um, sodass der letzte Datensatz zuerst erscheint und der erste Datensatz der letzte wird. *Zeilen zählen* zeigt Ihnen, wie viele Zeilen Sie in dieser Abfrage haben.

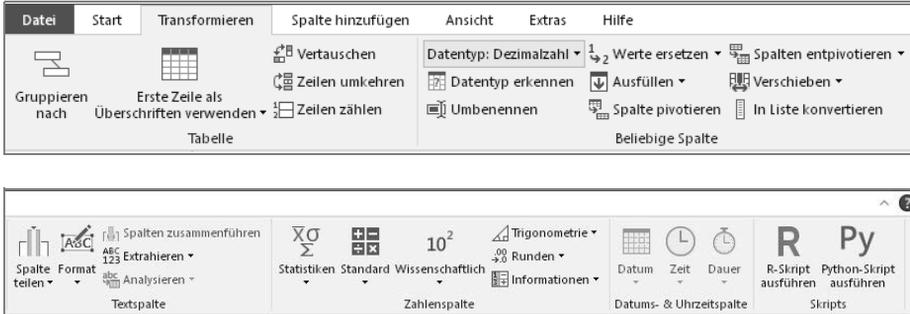


Abb. 3-8 Transformieren Sie Ihre Daten.

Die Gruppe *Beliebige Spalte* bietet die Möglichkeit, den Datentyp einer Spalte zu bearbeiten, wie Sie bereits gesehen haben. Sie können auch Werte ersetzen, wie zuvor beschrieben.

Mit der Funktion *Spalten entpivotieren* werden ausgewählte Datenspalten in eine Attributspalte und eine Wertspalte umgeformt. Manchmal möchten Sie Ihre Daten auf diese Weise umgestalten, wenn Sie beispielsweise viele Spalten in eine kleinere Anzahl von Spalten mit einer größeren Anzahl von Datensätzen konsolidieren möchten.

Die Schaltfläche *Datentyp erkennen* geht jede Spalte durch und versucht, ihr auf der Grundlage der Vorschau der Daten einen geeigneten Datentyp zuzuweisen. Power BI tut dies oft automatisch, wenn Sie die Daten das erste Mal in Power Query hineinholen; aber Sie können dies erneut ausführen lassen, wenn Sie eine Reihe von Transformationen vornehmen, die die Datentypen mehrerer Spalten ändern könnten.

Mit der Schaltfläche *Ausfüllen* können Sie fehlende Werte entweder von oben oder von unten mit einem Wert auf der Grundlage anderer Werte in der Spalte ausfüllen. Ich will ganz ehrlich sein, ich habe diese Funktion noch nie benutzt.

Mit der Schaltfläche *Verschieben* verschieben Sie die Position einer Spalte in einer Tabelle. Sie können sie nach links, rechts, an den Anfang oder an das Ende verschieben. Sie können dies auch mit mehreren ausgewählten Spalten tun.

Umbenennen ermöglicht Ihnen, eine ausgewählte Spalte umzubenennen.

Mit *Spalte pivotieren* können Sie die Werte einer Spalte nehmen, diese Werte in Spalten umwandeln und die Werte über die neue Kombination neu berechnen. Dies kann nützlich sein, wenn Sie Attribute in einer Spalte und Werte in einer anderen Spalte haben und Sie die Werte mit den Attributen als Spalten sehen möchten.

In Liste konvertieren ist eine einzigartige Funktion. Sie nimmt eine Spalte und wandelt sie in einen speziellen Tabellentyp in Power Query um, der Liste genannt wird. Listen können sozusagen als Parameter an spezielle benutzerdefinierte Funktionen übergeben werden. In diesem Buch werde ich mir nicht allzu viele Gedanken über Listen machen.

Die Gruppe *Textspalte* hat einen etwas seltsamen Namen, weil Sie diese Funktionen nicht nur für Textspalten verwenden können. Zum Glück sind sie größtenteils recht einfach.

Mit *Spalte teilen* können Sie eine Spalte auf der Grundlage eines ausgewählten Trennzeichens in zwei Spalten aufteilen.

Mit *Format* können Sie den Datentyp nicht bearbeiten. Sie können jedoch beispielsweise alle Daten in Klein- oder Großbuchstaben umwandeln oder Präfixe oder Suffixe hinzufügen. Mit dieser Funktion wird diese Spalte in eine Textspalte umgewandelt.

Spalten zusammenführen fügt zwei oder mehr Spalten zusammen und ermöglicht es Ihnen, ein Trennzeichen zu wählen.

Mit *Extrahieren* können Sie eine oder mehrere Spalten ändern, damit in der Spalte nur noch bestimmte Zeichen übrig bleiben. Ich verwende diese Funktion nicht sehr oft, aber wenn, dann meist, um Nicht-Unicode-Zeichen zu entfernen.

Analysieren macht etwas anderes als alle anderen Befehle in diese Gruppe. *Analysieren* nimmt eine JSON- oder XML-Datei, also semistrukturierte Daten, und bringt sie für die Analyse in ein klassisch strukturiertes Format. Ich verwende in den Beispielen keine JSON- oder XML-Dateien, aber wenn Sie mit NoSQL-Datenbanken arbeiten oder Ergebnisse von API-Aufrufen erhalten, müssen Sie diese Daten möglicherweise parsen, bevor Sie sie analysieren, und der Befehl *Analysieren* übernimmt diese Aufgabe.

Mit den Funktionen für *Zahlenspalten* können Sie Statistiken über eine Spalte als separaten Wert abrufen, Standardberechnungen mit einer Spalte durchführen, um deren Werte zu bearbeiten, wissenschaftliche Berechnungen mit einer Spalte durchführen, trigonometrische Funktionen mit einer Spalte durchführen, eine Spalte runden oder feststellen, ob ein Wert gerade oder ungerade ist, oder das Vorzeichen der Zahl ermitteln. Mit Ausnahme der Befehle, die sich hinter der Schaltfläche *Statistiken* verbergen, empfehle ich, eine Spalte zu duplizieren, bevor Sie Änderungen an den Daten vornehmen; dann steht Ihnen ab da sowohl der unveränderte Originalwert als auch der geänderte Wert zur Verfügung.

Die Gruppe *Datums- und Uhrzeitspalte* funktioniert auf die gleiche Weise wie die Gruppe *Zahlenspalte*, stellt jedoch zahlreiche Datums- und Zeittransformationen zur Verfügung.

Die letzte Gruppe auf der Registerkarte *Transformieren* ist die Gruppe *Skripts*, die die Möglichkeit bietet, sowohl R- als auch Python-Skripte auf Ihre Daten anzuwenden. Um diese Funktionen zu nutzen, müssen Sie R oder Python installiert haben und Power BI muss wissen, wo sich Ihre Installation dieser Sprachbibliotheken befindet. Wenn Sie mit Data Frames in R oder Python besser zurechtkommen und Transformationen lieber dort durchführen und diese in einem einzigen Schritt verwenden möchten, können Sie das tun. Sie können die Arbeit auch in Power Query vorbereiten, eine Skriptaufgabe anwenden und dann weitere Änderungen in Power Query vornehmen. Sie können diese Funktionen miteinander kombinieren, und wenn Sie beide verwenden möchten, geht das auch. Das Erlernen von R und Python würde den Rahmen dieses Buches sprengen, aber wenn Sie eine Karriere im Bereich Data Analytics anstreben, könnten sich diese beiden Sprachen als sehr nützlich erweisen.

3.2.3 Die Registerkarte »Spalte hinzufügen«

Das Erste, was auf der Registerkarte *Spalte hinzufügen* zu beachten ist, ist, dass Sie bis auf die Gruppe *Allgemein* alle anderen Funktionen bereits an anderen Stellen des Menübands gesehen haben. Die Gruppe *Allgemein* enthält jedoch so viel Features, dass sie einen eigenen Abschnitt im Buch verdient. Ein Blick auf Abbildung 3–9 macht dies ein wenig deutlicher.

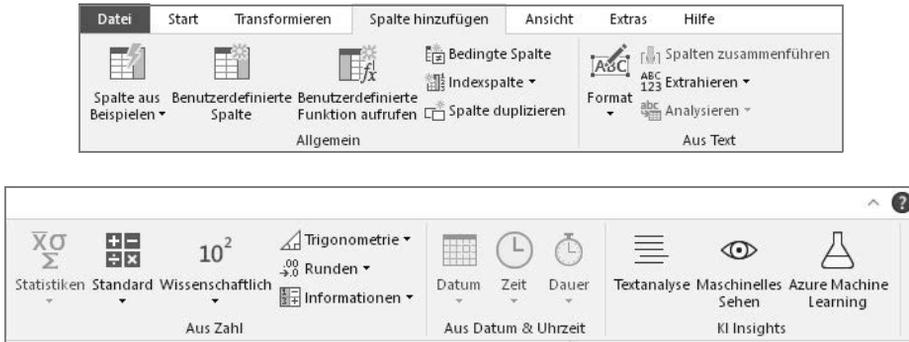


Abb. 3–9 Das Power Query-Menüband hat oft mehrere Stellen, an denen bestimmte Funktionen zu finden sind, aber zumindest sind sie immer nach Themen sortiert.

Spalte aus Beispielen ist tatsächliche, faktische, Schwarze Magie. Sie können eine oder mehrere Spalten auswählen, aus denen Sie Daten abrufen möchten. Dann müssen Sie einige Beispiele für die Ausgabe angeben, und Power BI versucht, die Spalte automatisch auszufüllen, während Sie Beispiele zur Verfügung stellen, mit denen Sie arbeiten. In unsere Abfrage *Noten* füge ich eine Spalte mit *Spalte aus Beispielen* hinzu, die, wenn die Spalte *SprechstundeBesucht* null ist, *nein* enthält und anderenfalls *ja*. Sie sehen dies in Abbildung 3–10. Als weiteren Bonus sehen Sie die Logik in M, die sich hinter dem Befehl verbirgt. Dies ist eine weitere Möglichkeit, zu sehen, wie die Formelsprache M die verschiedenen Dinge behandelt.

Benutzerdefinierte Spalte öffnet ein Dialogfeld, in dem Sie eine benutzerdefinierte Spalte mit M erstellen. Dies kann nützlich sein, wenn Sie Transformationen durchführen möchten, die vielleicht nicht von der Benutzeroberfläche unterstützt werden oder die Sie effizienter mittels Code durchführen könnten. Das Dialogfeld zeigt Ihnen zumindest an, ob ein Syntaxfehler im Code vorliegt, aber wenn Sie M gerade erst lernen, sind die Fehler nicht unbedingt sehr hilfreich.

Benutzerdefinierte Funktion aufrufen ist ein unglaublich leistungsfähiges Werkzeug, mit dem Sie eine Funktion nehmen und darüber für eine bestimmte Spalte iterieren, um einen bestimmten Wert zu finden und dann eine Art von Logik anzuwenden. Ein einfaches Beispiel ist die Erstellung einer Funktion, die eine neue Spalte erstellt und dann den Wert einer Spalte mit einem anderen Wert multipliziert, um einen dritten Wert zu erhalten, der in diese neue Spalte eingefügt wird. Haben Sie diese Logik verstanden? Dies ist ein sehr leistungsfähiges Werkzeug, aber es kann auch schwierig sein, die Funktionen zu verstehen, wenn Sie zum ersten Mal damit arbeiten.

Spalte aus Beispielen hinzufügen

Geben Sie die Beispielswerte zum Erstellen einer neuen Spalte ein (STRG+EINGABE zum Übernehmen).
 Transformieren: `if [SprechstundeBesucht] = null then "nein" else if [SprechstundeBesucht] = 1 then "ja" else null`

OK Abbrechen

	123	AufgabeID	123	Punkte	123	SprechstundeBesucht	123	MaximaleP	Benutzerdefiniert
1	1	1	1	72		null		nein	
2	2	1	1	75		null		nein	
3	3	1	1	92		null		nein	
4	4	1	1	65		null		nein	
5	5	1	1	75	1	1	1	ja	
6	6	1	1	100		null		nein	
7	7	1	1	100		null		nein	
8	8	1	1	83		null		nein	
9	9	1	1	78		null		nein	
10	10	1	1	71		null		nein	
11	11	1	1	67	1	1	1	ja	
12	12	1	1	100		null		nein	
13	13	1	1	100		null		nein	
14	14	1	1	71		null		nein	
15	15	1	1	65		null		nein	

Abb. 3-10 Ein paar Beispiele für das, was ich will, ein bisschen KI, und voilà, eine neue Spalte!

Mit der Funktion *Bedingte Spalte* können Sie eine Spalte nach dem Muster einer Wenn-dann-sonst-Anweisung erstellen. Wenn Spalte X kleiner als Spalte Y ist, dann Ja, sonst Nein. Sie können auch mehrere Klauseln für kompliziertere oder iterative Wenn-Funktionen haben. Falls Sie mit SQL vertraut sind, können Sie sich dies wie eine CASE WHEN-Anweisung vorstellen. Abbildung 3-11 zeigt das Dialogfeld zum Hinzufügen einer bedingten Spalte.

Bedingte Spalte hinzufügen

Fügt eine bedingte Spalte hinzu, die aus den anderen Spalten oder Werten berechnet wird.

Neuer Spaltenname
Benutzerdefiniert.1

Spaltenname	Operator	Wert	Ausgabe
Wenn		ABC 123	Dann ABC 123

Klausel hinzufügen

Sonst
ABC 123

OK Abbrechen

Abb. 3-11 Manchmal brauchen wir nur eine IF-Anweisung.

Mit *Indexspalte* wird eine Spalte erstellt, die bei 0 oder 1 beginnt und die eine eindeutige Liste von Werten anlegt, die in der Regel der Zeilennummer entspricht. Das Wichtigste an einer Indexspalte ist, dass sie völlig eindeutig ist und bei Bedarf als improvisierter Schlüsselwert verwendet werden kann. Sie können diese Spalte möglicherweise mit anderen zusammenführen, um den Index- oder Schlüsselwert an andere Tabellen weiterzugeben, um so Beziehungen zu erstellen. Im Idealfall haben Ihre Daten bereits Schlüssel, und ein Index ist nicht erforderlich. Wenn Sie jedoch eine Indexspalte erstellen müssen, ist das sehr einfach.

Spalte duplizieren nimmt die aktuelle Spalte und kopiert sie. Das war's schon. Dazu gibt es keinen witzigen Kommentar. Es handelt sich hierbei einfach um Kopieren und Einfügen.

Zum Schluss möchte ich noch die Registerkarte *Ansicht* erwähnen. Die Elemente auf dieser Registerkarte ermöglichen es Ihnen, einige kleine Anpassungen an der Art und Weise vorzunehmen, wie Sie mit Power Query-Editor interagieren. In der Gruppe *Datenvorschau* möchte ich die Kontrollkästchen *Spaltenqualität*, *Spaltenverteilung* und *Spaltenprofil* hervorheben, da sie es Ihnen ermöglichen, mehr Details über Ihre Daten auf Spaltenebene zu sehen. Dazu gehören eine prozentuale Aufschlüsselung der gültigen, fehlerhaften und leeren Werte in einer Spalte, die numerische Verteilung der Werte in einer Spalte und eine Verteilung der Daten, einschließlich der Anzahl der eindeutigen Werte in einer Spalte.

Wenn Sie alle Änderungen vorgenommen haben und bereit sind, Ihre Daten mit den Änderungen zu laden, klicken Sie auf die Schaltfläche *Schließen und übernehmen*. Dadurch wird Ihr Power Query-Fenster geschlossen und Sie kehren zur Power BI Desktop-Arbeitsfläche zurück.

3.3 Die Ansicht »Modell« in Power BI Desktop

Kehren wir zu Power BI zurück und werfen Sie einen Blick auf die dritte Ansicht in Power BI Desktop, die Ansicht *Modell*. Warum sind Beziehungen wichtig? Was bewirken sie eigentlich? Kann ich Beziehungen einfach ignorieren und eine große Tabelle erstellen, die den Welthunger löst? In diesem Abschnitt werde ich mich mit diesen Fragen beschäftigen.

Werfen Sie einen kurzen Blick auf die Ansicht *Modell* der aktuellen Datei in Abbildung 3–12.

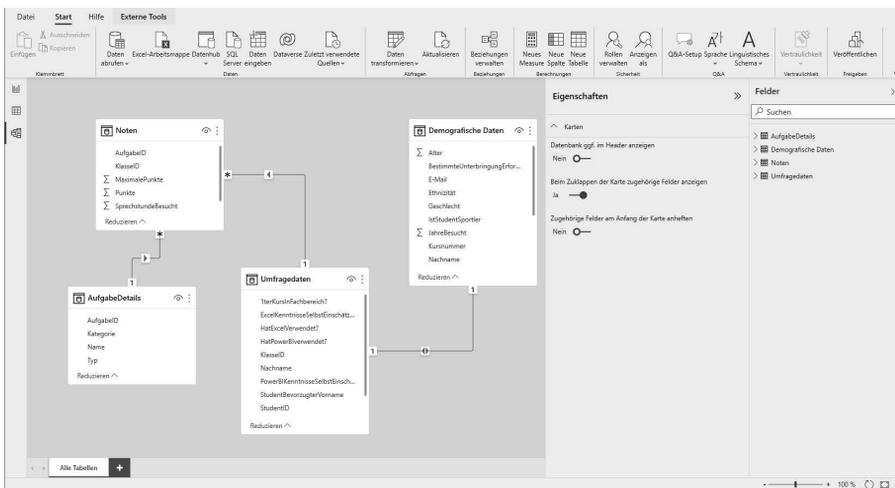


Abb. 3–12 Das ist es, unser Datenmodell.

Standardmäßig versucht Power BI, Beziehungen zu erkennen, wenn Daten importiert werden und die Daten so aufgebaut sind, dass es die Beziehungen richtig erkennt. Leider gelingt dies nicht immer gut. Daher sollten Sie die Beziehungen nach dem Datenimport immer noch einmal überprüfen, ob Ihr Modell wirklich so aufgebaut ist, wie Sie es sich vorgestellt haben. Ihr Modell sieht nach dem ersten Importvorgang vielleicht nicht so aus wie in Abbildung 3–12, wenn Power BI zum ersten Mal versucht, Beziehungen zu erkennen. Wie Sie hier und in Kapitel 6 sehen werden, kommt es wirklich darauf an, welche Form die Beziehungen annehmen!

3.3.1 Was ist eine Beziehung?

Eine Beziehung ist, einfach ausgedrückt, eine Möglichkeit, Power BI mitzuteilen, welche Tabellen miteinander verbunden sind und wie. Eine Beziehung sollte so aufgebaut sein, dass die Werte einer Spalte in einer Tabelle mit den Werten einer anderen Tabelle übereinstimmen, sodass Sie bei einer Abfrage sinnvolle Ergebnisse erhalten. Eine Beziehung ist eine Kombination aus mehreren Faktoren – Spalten, Richtung und Kardinalität. Wenn Sie den Mauszeiger über die Linien in unserem Datenmodell bewegen, werden sie leuchtend gelb, und wenn Sie darauf doppelklicken, sehen Sie das Dialogfeld *Beziehung bearbeiten* wie in Abbildung 3–13. In diesem Beispiel zeige ich die Beziehung zwischen *AufgabeDetails* und *Noten*.

Lassen Sie uns dieses Dialogfeld genauer ansehen und erörtern, was es dort gibt. Zunächst sehen Sie, dass zwei Tabellen ausgewählt sind. Da ich in diesem Fall die Beziehung in der Ansicht *Modell* ausgewählt habe, wurden die betreffenden Tabellen bereits ausgefüllt. Sie sehen, dass in jeder Tabelle eine Spalte hervorgehoben dargestellt wird. Dies ist die Spalte, die der Beziehung zugrunde liegt.

Hier teilen Sie Power BI Folgendes mit: »Wenn du eine Abfrage ausführst und Daten sowohl aus *Noten* als auch aus *AufgabeDetails* abrufen musst, kannst du die Tabellen über diese Spalte verknüpfen.« Dies ist in jedem Datenbanksystem wichtig, und bei Power BI ist es nicht anders. Was allerdings anders ist, ist, dass Sie nicht mehrere Spalten auswählen können, um eine Beziehung herzustellen.

»Aber Jeremy!« Ich kann schon hören, wie einige von Ihnen beim Lesen laut rufen: »Ich kann eine SQL-Anweisung schreiben, die mehrere Verknüpfungsbedingungen zwischen zwei Tabellen enthält!« Ja, das können Sie. Power BI führt jedoch keine SQL-Anweisungen aus, wenn es seine interne Datenbank abfragt: Es führt DAX aus. Wenn Sie eine Beziehung zwischen zwei Tabellen mit mehreren Spalten anlegen müssen, schlage ich vor, eine Spalte mit einem zusammengesetzten Wert der Spalten zu erstellen, die Sie in beiden Tabellen verbinden möchten, und dann Ihre einzelne Beziehung mit Ihrer neuen zusammengesetzten Spalte zu erstellen.

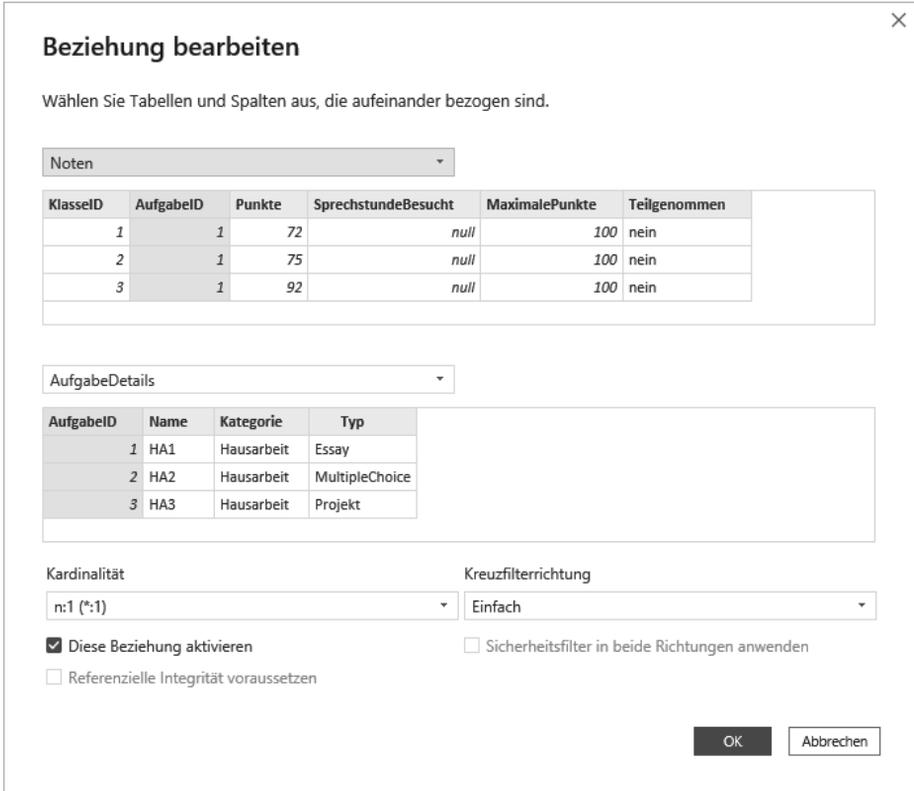


Abb. 3-13 Das ist kein Zufall: Beziehungen sind grundlegend.

Lassen Sie mich nun über Kardinalität und Richtung schreiben. Diese beiden Dinge gehen Hand in Hand. Sie können in Power BI drei Arten von Beziehungen erstellen: Eins-zu-eins-Beziehungen, Eins-zu-viele-Beziehungen und Viele-zu-viele-Beziehungen.

Ich mag keine Viele-zu-viele-Beziehungen. Bei der Erstellung eines Datenmodells werden Viele-zu-viele-Beziehungen nach Möglichkeit vermieden, da sie nicht leistungsfähig sind und manchmal zu merkwürdigen Ergebnissen im Vergleich zu dem, was Sie erwarten, führen. Viele-zu-viele-Beziehungen sind eher mit einem FULL OUTER JOIN vergleichbar, wie Sie es von SQL kennen. Sie erkennen schnell, dass es sich um eine Viele-zu-viele-Beziehung handelt, wenn Sie sich die Linie ansehen, die die beiden Tabellen verbindet, und prüfen, ob sich an beiden Linienenden Sternchen befinden.

Bei einer Eins-zu-eins-Beziehung sind alle Werte in Spalte A von Tabelle A eindeutig, und außerdem sind alle Werte in Spalte A von Tabelle B eindeutig. Sie erkennen eine Eins-zu-eins-Beziehung schnell, wenn Sie an den Enden der Linien, die die beiden Tabellen verbindet, eine 1 sehen.

Bei einer Eins-zu-viele-Beziehung sind alle Werte in Spalte A von Tabelle A eindeutig, jedoch sind die Werte in Spalte A von Tabelle B nicht eindeutig. Sie erkennen diese Art von Beziehung in Ihrem Modell schnell daran, dass in der An-

sicht *Modell* das eine Ende der Linie, die die Tabellen verbindet, mit einer 1 und das andere Ende mit einem Sternchen versehen ist. Der Vorteil dieser beiden Arten von Beziehungen ist, dass ich nach einem Wert suche und diese Werte in der anderen Spalte vergleiche, anstatt nach jeder Kopie dieses Wertes zu suchen und sie mit jeder anderen Kopie dieses Wertes zu vergleichen.

Wenn es um die Richtung einer Beziehung geht, geht es darum, welche Tabelle die andere filtern kann. Können sie in beide Richtungen filtern oder nur in eine? Eine Eins-zu-eins-Beziehung hat standardmäßig die Filterrichtung *Beide*. Dies wird in der Mitte der Verbindungslinie zwischen zwei Tabellen durch Pfeile dargestellt, die in beide Richtungen zeigen. Eine Viele-zu-viele-Beziehung kann nur die Filterrichtung *Beide* verwenden. Eine Eins-zu-viele-Beziehung hat standardmäßig die Filterrichtung *Einfach*, die von der Tabelle mit dem eindeutigen Wert zu der Tabelle mit den nicht eindeutigen Werten führt. Eine einfach gerichtete Beziehung ist daran zu erkennen, dass ein Pfeil auf die Tabelle zeigt, die gefiltert wird.

Lassen Sie mich anhand einiger Beispiele zeigen, was passiert. Zunächst werde ich eine Tabelle anzeigen, die Werte aus diesen beiden Tabellen enthält, ohne dass eine Beziehung besteht, und dann mit der bestehenden Beziehung, um zu zeigen, wie die Richtung funktioniert. Abbildung 3–14 zeigt, was passiert, wenn keine Beziehung definiert ist.

AufgabeID	Durchschnitt von Punkte
1	77,13
2	77,13
3	77,13
4	77,13
5	77,13
6	77,13
7	77,13
8	77,13
9	77,13
10	77,13
11	77,13
12	77,13
13	77,13
14	77,13
Gesamt	77,13

Abb. 3–14 Sie sind noch kein Experte, aber Sie wissen, dass hier etwas nicht stimmt, und die Beziehung ist der Grund dafür.

Hier, in Abbildung 3–14, habe ich die *AufgabeID* der Aufgabe aus der Tabelle *AufgabeDetails* genommen und den Mittelwert der Punktzahl für diese Aufgaben berechnet. Der Wert 77,13 ist der Mittelwert der Punktzahl für alle, aber nicht

die durchschnittliche Punktzahl pro Aufgabe. Ohne eine Beziehung weiß Power BI nicht, wie es eine Abfrage erstellen soll, die das Gesuchte liefert.

Jetzt füge ich die Beziehung wieder ein und erhalte die Ausgabe, die Abbildung 3–15 zeigt.

AufgabeID	Durchschnitt von Punkte
1	82,75
2	79,60
3	82,30
4	79,60
5	78,20
6	80,50
7	78,25
8	81,50
9	79,45
10	79,60
11	77,90
12	80,60
13	78,55
14	41,00
Gesamt	77,13

Abb. 3–15 Das ergibt mehr Sinn, sehr viel mehr Sinn, denn die Beziehung macht die internen Abfragen funktional.

Sie werden feststellen, dass der Gesamtmittelwert immer noch korrekt ist; dieser Wert erfordert keine Interaktion mit *AufgabeID*. Für jede einzelne *AufgabeID* kann ich jedoch eine Abfrage erstellen, die im Grunde besagt: »WHERE AufgabeID = 1 ist, ignoriere alle anderen Werte in der Tabelle Noten, bei denen die AufgabeID nicht 1 ist, und berechne dann die Durchschnittspunkte für genau diese AufgabeID und führe das dann für jede AufgabeID erneut durch.« Beziehungen machen dies möglich.

Lassen Sie mich nun zu den Filterrichtungen kommen. In Abbildung 3–16 habe ich einen Datenschnitt (eine Art von Visualisierung, die andere Visuals auf der Berichtseite filtert, indem sie einen oder mehrere Werte auf die anderen Visuals anwendet, ähnlich wie eine WHERE-Klausel in SQL) auf Punkte, und ich möchte sehen, welche *AufgabeID* diese Punkte enthalten.

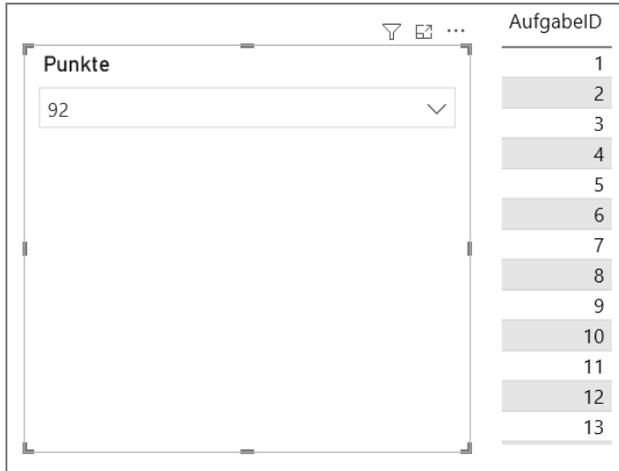


Abb. 3–16 Ein nutzloser Filter ist nutzlos. Kardinalität und Beziehungsrichtung sind wichtig!

Denken Sie daran, dass unser Datenmodell eine Eins-zu-viele-Beziehung zwischen *AufgabeDetails* und *Noten* enthält und nur in eine Richtung filtert, nämlich von der einen in *AufgabeDetails* zu den vielen in der Tabelle *Noten*. Die Punktezahl aus *Noten* kann also nicht nach *AufgabeDetails* filtern.

Drehen Sie das Ganze nun um und setzen Sie die *AufgabeID* auf den Slicer und die Liste der *Punkte* in die Tabelle. Dann sehen Sie, was Sie in Abbildung 3–17 erhalten. Anschließend sehen Sie in Abbildung 3–18, was passiert, wenn die Kreuzfilterrichtung auf beide eingestellt ist.

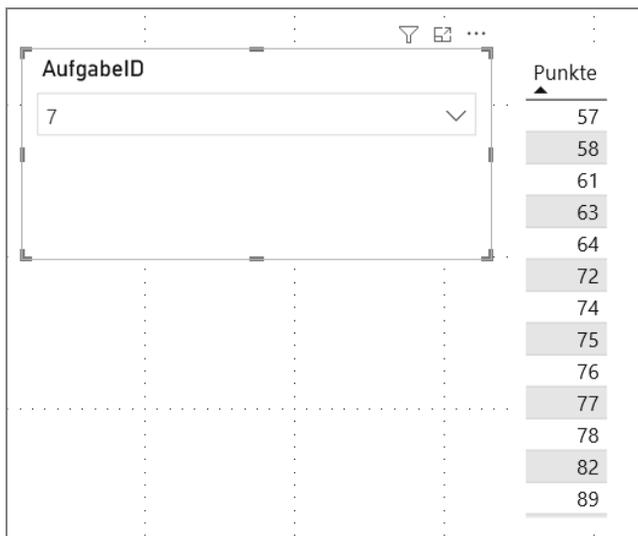


Abb. 3–17 Ein nützlicher Filter ist nützlich!

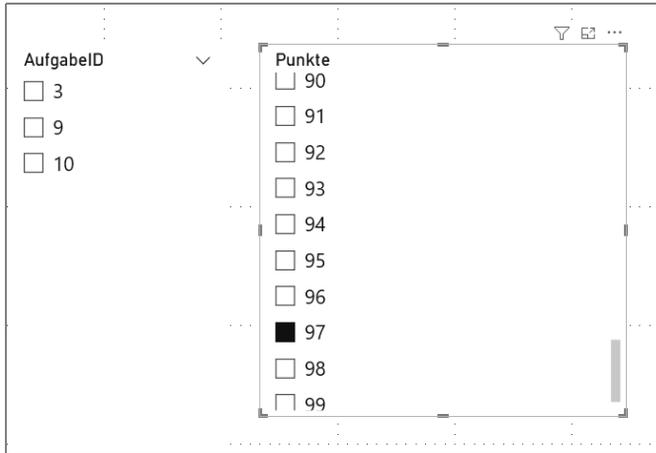


Abb. 3-18 Bidirektionale Filterung

Ich höre schon wieder die Rufe aus dem Äther: »Aber Jeremy, wenn die bidirektionale Filterung das kann, warum machen wir das nicht einfach immer zum Standard?« Die kurze Antwort ist, dass dies nicht so performant ist und dass Sie Beziehungen schaffen können, bei denen *Mehrdeutigkeit* auftreten kann, was die Verwendung anderer Beziehungen verhindert. Im Allgemeinen wäre die beste Datenmodellierungspraxis ein *Sternschema*, bei dem Sie eine Tabelle mit Ihren Fakten haben und dann andere kleinere Tabellen, die sie umgeben wie ein Stern, der diese Tabelle filtert.

Okay, was ist mit der einen Tabelle, die den Welthunger löst? Wäre das nicht auch in Ordnung? Nun, die kurze Antwort lautet nein. Eine große Einzeltabelle nutzt nicht die Vorteile eines dimensionalen Modells, das mehrere Vorteile für die Datenverwaltung und die Benutzerfreundlichkeit bietet. Außerdem kann eine einzelne große Tabelle schnell unhandlich werden, da Sie viele komplexe Bearbeitungen in Power Query durchführen, um die Erstellung dieser einzelnen Tabelle zu erleichtern. Es ist wirklich einfacher, mit einer Tabelle mit 5 Spalten und einer Tabelle mit 10 Spalten als mit einer einzigen Tabelle mit 15 Spalten umzugehen.

Außerdem sind Einzeltabellen anfälliger dafür, bei einzelnen Problemen zu zerbrechen, im Gegensatz zu einem Ansatz mit mehreren einzelnen Tabellen, die ihrerseits mehrere einfachere Elemente enthalten. Wenn Sie nach der Kurzversion eines guten Datenmodells suchen, dann ist das eine Tabelle, die Ihre Fakten enthält und von Tabellen umgeben ist, die diese Fakten mit mehrfachen Eins-zu-viele-Beziehungen filtern können. Wenn Sie das tun, sind Sie auf dem richtigen Dampfer.

3.3.2 Der Arbeitsbereich »Eigenschaften«

Der Arbeitsbereich *Eigenschaften* in der Ansicht *Modell* ist sehr kontextbezogen. Sie finden den Arbeitsbereich *Eigenschaften* immer auf der rechten Seite des Fensters. Je nachdem, welches Datenelement Sie ausgewählt haben, ändert sich der Arbeitsbereich *Eigenschaften* entsprechend dem Typ des ausgewählten Datenobjekts.

Wenn eine Tabelle ausgewählt ist, sehen Sie den Namen, die Beschreibung, die Synonyme, die Zeilenbeschriftung, die Schlüsselspalte, die Flags *Ist verborgen* sowie *Ist eine empfohlene Tabelle*. Im Abschnitt *Erweitert* zeigt dieser Arbeitsbereich den *Speichermodus* an. Der *Speichermodus* kann *Importieren*, *DirectQuery* oder *Dual* sein. Beachten Sie, dass der Modus *Importieren* nicht mehr geändert werden kann, wenn er einmal eingestellt ist. Ein Beispiel für den Bereich *Eigenschaften* in Aktion finden Sie in Abbildung 3–19.

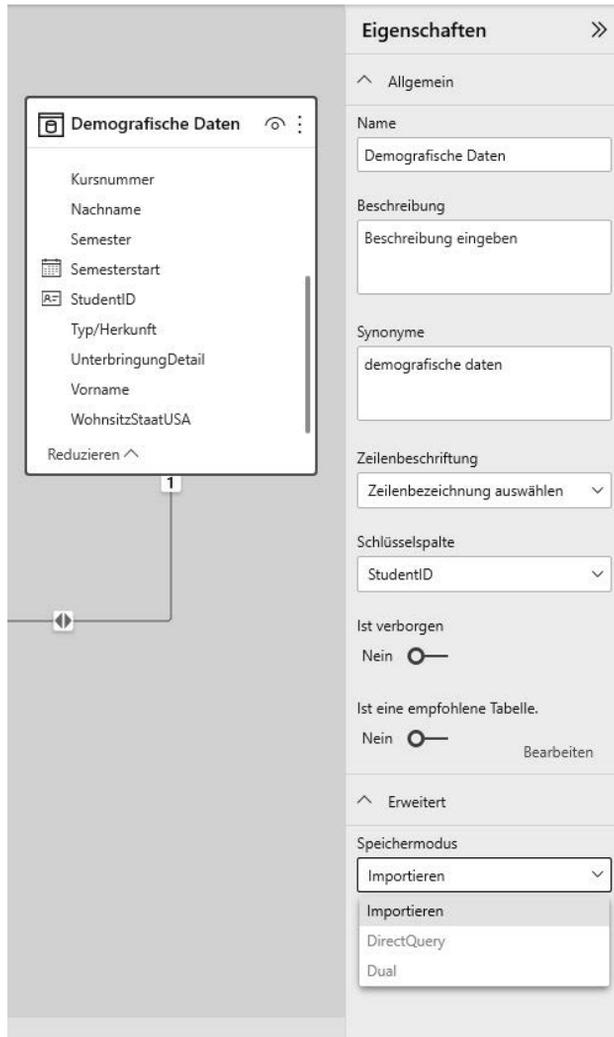


Abb. 3–19 Der Bereich *Eigenschaften* enthält Informationen, die für die Organisation, die Linguistik, die Schlüssel und den Speichermodus einer Tabelle nützlich sind.

Das Feld *Name* ist selbsterklärend. Das Feld *Beschreibung* ist leer und kann mit einer Zusammenfassung des Tabelleninhalts oder seines Zwecks in Ihrem Datenmodell ausgefüllt werden. Die *Synonyme* werden zum Zweck der Frage- und Ant-

wortfunktion aufgelistet, die ich bereits kurz angeschnitten habe, aber später in diesem Buch noch ausführlicher behandeln werde. Die *Zeilenbeschriftung* ermöglicht es Ihnen, eine Spalte zu identifizieren, die als vorgeschlagene Spalte bei der Durchführung von Q&A mit einer bestimmten Tabelle dient.

Die *Schlüsselspalte* ermöglicht es Ihnen, eine Spalte zu identifizieren, die nur eindeutige Werte enthält. Ich hebe gerne die Spalte hervor, die als Grundlage für die Beziehungen dieser Tabelle zu anderen Tabellen dient. Wenn eine Spalte als Schlüsselspalte identifiziert wird, erhält sie außerdem ein spezielles Symbol im Aufgabenbereich *Felder*.

Wenn *Ist verborgen* auf *Ja* eingestellt ist, wird die Tabelle in der Ansicht *Bericht* ausgeblendet, sodass Sie keine Elemente aus dieser Tabelle zu Berichtselementen hinzufügen können. Beachten Sie, dass, wenn Sie bereits Objekte mit dieser Tabelle erstellt haben, diese Objekte sichtbar bleiben, aber Sie können keine Datenelemente aus dieser Tabelle zu neuen Visualisierungen hinzufügen.

Wenn es sich bei einer Tabelle um eine ausgewählte (empfohlene) Tabelle handelt, können Personen in einer Organisation, in der ausgewählte (empfohlene) Tabellen aktiviert sind, über den Power BI-Dienst nach dieser Tabelle suchen.

Wenn eine einzelne Spalte ausgewählt ist, werden die Felder *Name*, *Beschreibung*, *Synonyme*, *Anzeigeordner* und *Ist verborgen* angezeigt. Im Abschnitt *Formatierung* des Aufgabenbereichs sehen Sie den Datentyp der Spalte und das Format der Daten. Im Abschnitt *Erweitert* schließlich können Sie eine andere Spalte auswählen, nach der sortiert werden soll, gegebenenfalls eine Datenkategorie angeben (denken Sie an diese Kennzeichnungen für sehr spezifische Datentypen, die Power BI in sehr spezifischen Situationen verwenden muss) und eine Standardzusammenfassung festlegen. In diesem Abschnitt des Aufgabenbereichs geben Sie auch an, ob eine Spalte NULL-Werte enthalten darf. Es gelten alle Hinweise aus dem vorangegangenen Abschnitt über Tabellen, aber ich möchte ein wenig auf die Anzeigeordner eingehen.

Anzeigeordner sind eine gute Möglichkeit, ein Datenmodell sauber zu halten, indem Datenelemente in einem Unterordner innerhalb einer Tabelle gruppiert werden. Hier werden keine eigentlichen Daten verändert oder manipuliert; dies dient ausschließlich der optischen Aufgeräumtheit. Sie können auch Measures in Anzeigeordnern ablegen, was sich als besonders hilfreich erweist, wenn Sie bestimmte Elemente gruppieren möchten, die in der größeren alphabetischen Sortierung, die Power BI im Fenster *Felder* vornimmt, möglicherweise nicht geeignet sind.

3.4 Zusammenfassung

Zusammenfassend lässt sich sagen, dass Sie in diesem Kapitel eine ganze Menge geschafft haben. Sie haben Ihre ersten Schritte in Power Query unternommen, Daten in ein Modell geladen, Beziehungen zur Unterstützung der Analyse erstellt und nun kann ich mit der Erstellung des Dashboards beginnen, das dabei helfen wird, mehr über die Leistungen der Studierenden im Laufe des Semesters im Kurs zu erfahren.

Um dies zu erleichtern, werde ich im nächsten Kapitel die Visualisierungsoptionen in Power BI näher erläutern und Anwendungsfälle für jede Art von Visualisierung vorstellen. Sie werden auch üben, aus diesen Visualisierungen Rückschlüsse auf die Daten zu ziehen, um ein wenig mehr über die Studierenden zu erfahren. Gut, dass Sie mir bis hierher gefolgt sind. Sie sind definitiv auf dem besten Weg, sich ein Cool-School-University-Sweatshirt zu verdienen.