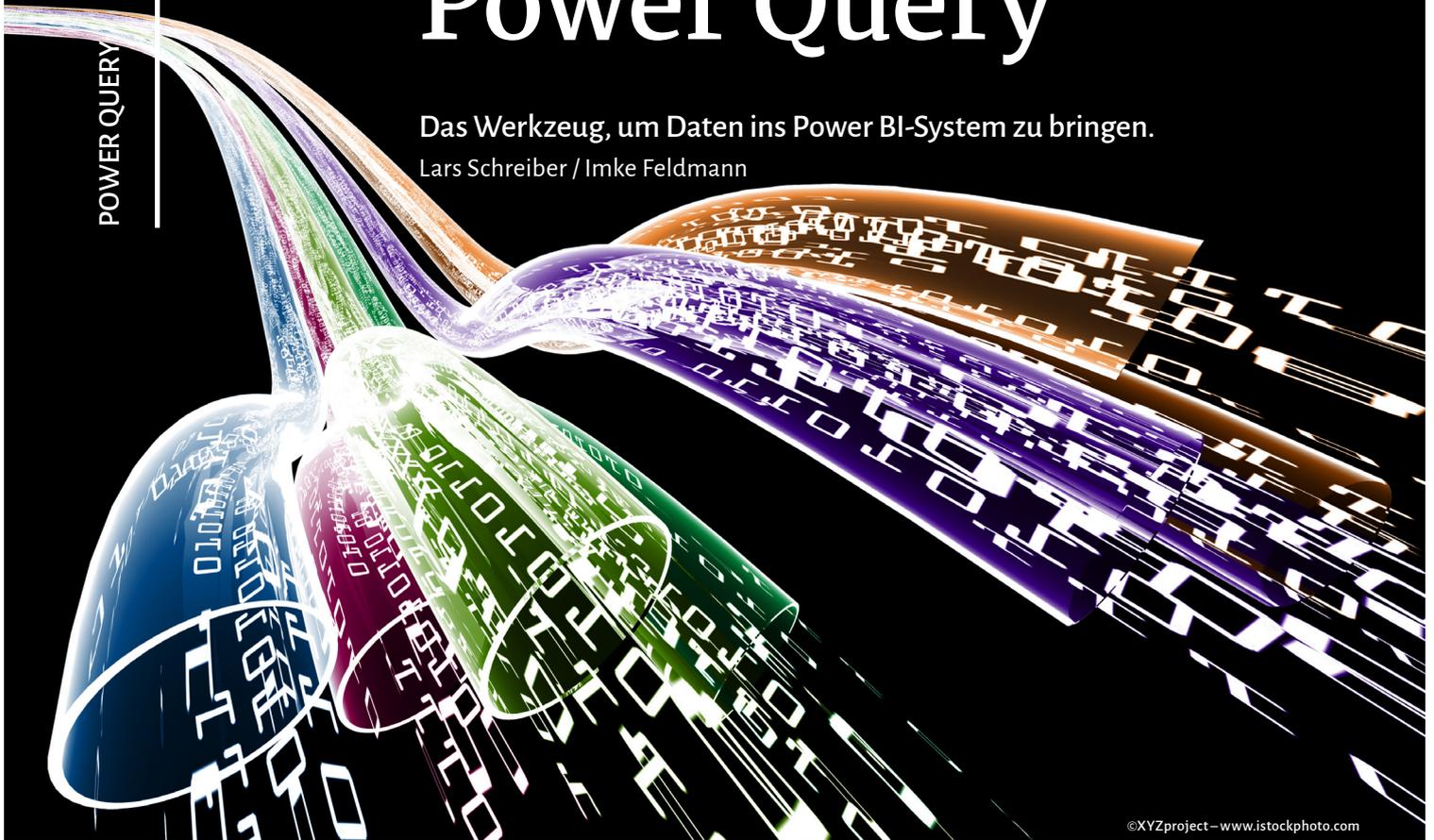


Power Query

Das Werkzeug, um Daten ins Power BI-System zu bringen.

Lars Schreiber / Imke Feldmann

POWER QUERY



©XVZproject – www.istockphoto.com

In der letzten Ausgabe haben wir einen Überblick zu Microsofts moderne Datenanalyse-Plattform Power BI gegeben. In diesem und nachfolgenden Artikeln werden wir Ihnen deren einzelne Bestandteile detaillierter näherbringen. Der aktuelle Artikel beschreibt Zweck und Funktionsweise der Komponente Power Query.

Ein flexibles Endnutzer-Werkzeug für Datenimport und -transformation

Power Query ist eine Komponente in Power BI, mit der man Daten aus zahlreichen Quellsystemen importieren und bereinigen, filtern, umformen oder anderweitig bearbeiten kann, bevor man sie in das Power BI-Datenmodell lädt. Dabei sind die erstellten Schritte jederzeit auf Basis einer neuen Datenlage wiederholt ausführbar. Manuelles aufwändiges und oft fehleranfälliges „Copy & Paste“ wird so effektiv vermieden. Dies kann gerade bei zeitkritischen Prozessen wie Monats- oder Jahres-

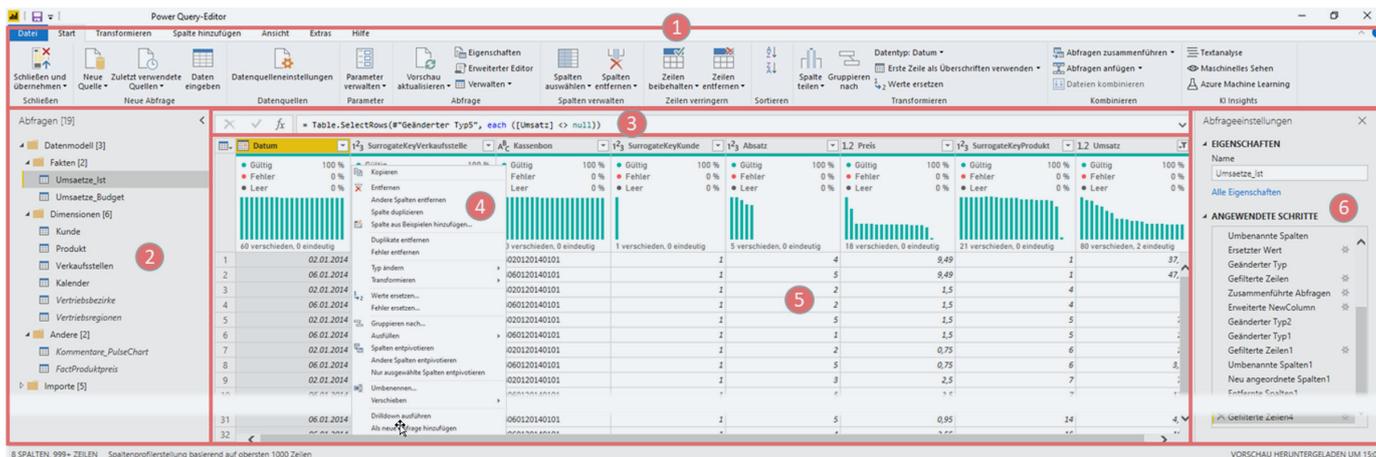
abschlüssen wertvolle Zeitersparnis bringen und gleichzeitig die Qualität der Abschlüsse verbessern.

Die Bedienung ist sehr benutzerfreundlich gestaltet und man kann sehr viele Schritte mit Mausclicks erledigen. Sind jedoch fortgeschrittene Operationen notwendig, für die es keine Schaltflächen auf der Benutzeroberfläche gibt, kann man die bestehende Lösung reibungslos mithilfe von Code anpassen (man spricht hier auch von einer „No code, low code“-Umgebung). Die hier verwendete Sprache nennt sich „Power Query M formula language“, auch kurz „M“ genannt. Es dürfte u. a. der einfachen Benutzbarkeit von Power Query zu verdanken sein, dass dieses Werkzeug mittlerweile in vielen Microsoft-Programmen Verwendung findet. Wir beziehen uns im vorliegenden Artikel auf Power Query in Power BI Desktop.

Power Query ist auch für andere Systeme offen. So kann man in Power BI z.B. etab-

lierte Sprachen aus dem Data-Science-Bereich wie Python oder R mit einbinden. Damit könnten ggf. bereits bestehende Lösungen schnell in Power BI integriert werden oder auf bestehende Module mit z.T. komplexen Berechnungen zurückgegriffen werden, ohne das Rad neu erfinden zu müssen. Außerdem kann man sich aus Power Query heraus mit Webdiensten verbinden, die Teile der Transformationen übernehmen oder anreichern, wie z.B. Übersetzungen oder Anwendungen künstlicher Intelligenz. Für Nutzer des Lizenzmodells Power BI Premium stellt Microsoft einige seiner Cognitive Services sowie Azure Machine Learning bereits voll integriert über die Benutzeroberfläche zur Verfügung.

Die Verbindung zu Webdiensten öffnet darüber hinaus auch das Tor zur Prozessautomatisierung, denn darüber lassen sich beispielsweise auch Flows in Power Automate starten, wenn manche Daten bestimmten Anforderungen nicht genügen.



- 1) Power Query Editor Menüband
- 2) Navigationsleiste für die Abfragen
- 3) Bearbeitungsleiste (zeigt den M-Code)
- 4) Kontextmenü
- 5) Vorschauenfenster (zeigt Ergebnis für jeden Abfrageschritt)
- 6) Abfrageeinstellungen (listet Namen aller Abfrageschritte)

Abb. 1: Grafische Nutzeroberfläche von Power Query in Power BI Desktop

Import von Daten

Der Import von Daten stellt i.d.R. den ersten Schritt beim Erstellen einer Power BI-Lösung dar. Ein Mausklick auf „Daten abrufen“ eröffnet Zugang zu über 100 verschiedenen Datenkonnektoren wie z.B. Excel, SQL Server, SharePoint, Internetseiten u.v.m.

Tabellenimport

Sobald man sich zur Datenquelle verbunden hat, erhält man eine Auswahl der vorhandenen Objekte. Bei einer Datenbank erscheinen die Tabellen und Sichten und wenn man sich mit Excel verbindet, sieht man die Liste der vorhandenen Arbeitsblätter und Tabellen. Daraus kann man sich eine oder mehrere Tabellen auswählen und erhält als nächstes eine Vorschau der ersten Tabellenzeilen der jeweiligen Objekte. Zusätzlich kann man sich die Profile der einzelnen Spalten grafisch anzeigen lassen und Indikatoren für die Datenqualität erhalten: Dadurch kann man Probleme oder Unregelmäßigkeiten mit den Quelldaten schnell erkennen und erhält eine gute Ausgangsbasis, um notwendige Bereinigungen durchführen zu können.

Sobald die Daten importiert und ggf. transformiert wurden, kann die Abfrage gespeichert und die Daten in das Datenmodell geladen werden. Sollten sich die Quelldaten ändern und man möchte den aktuellen Stand abrufen, lässt sich die Abfrage einfach erneut aufrufen. Dann wird die Datenquelle erneut geladen und die in der Abfrage abge-

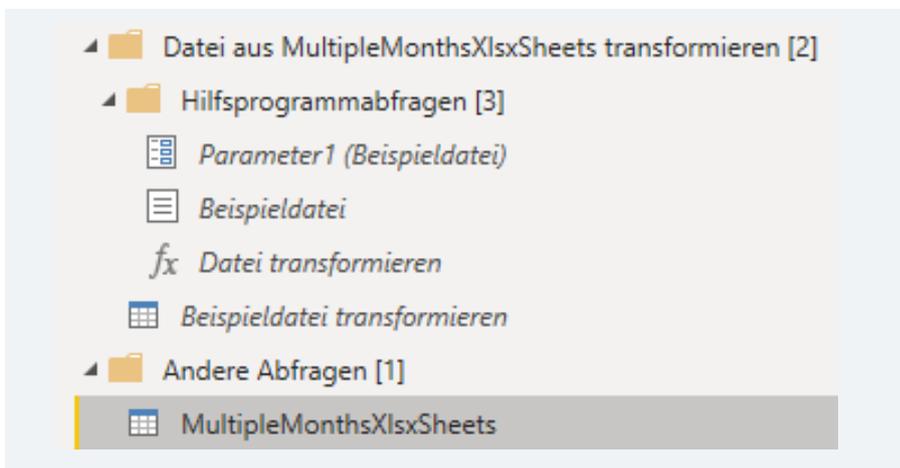


Abb. 2: Mehrere Dateien aus Ordnern importieren und mit einer gemeinsamen Funktion umwandeln

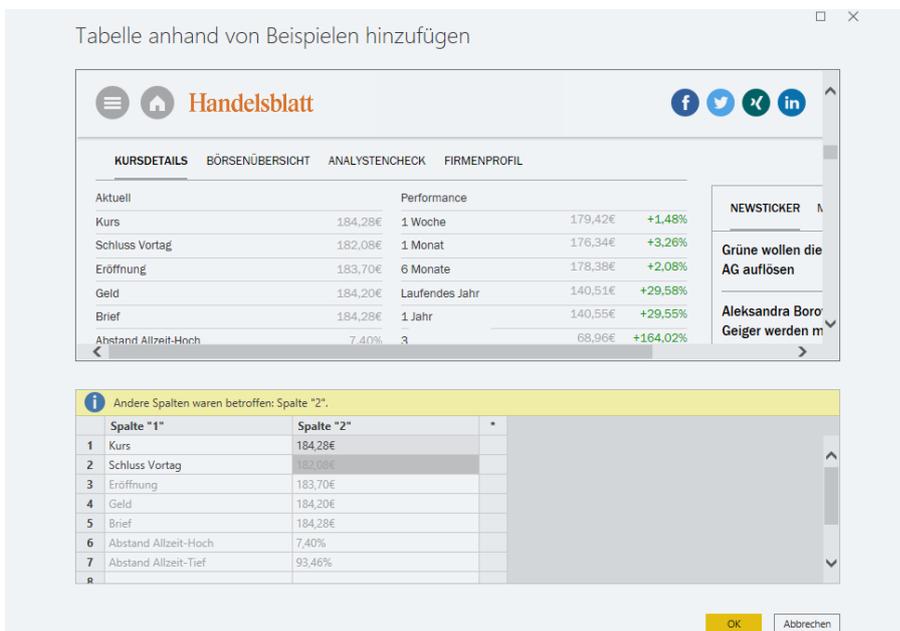


Abb. 3: Import von Daten aus dem Internet durch Assistenten unterstützt

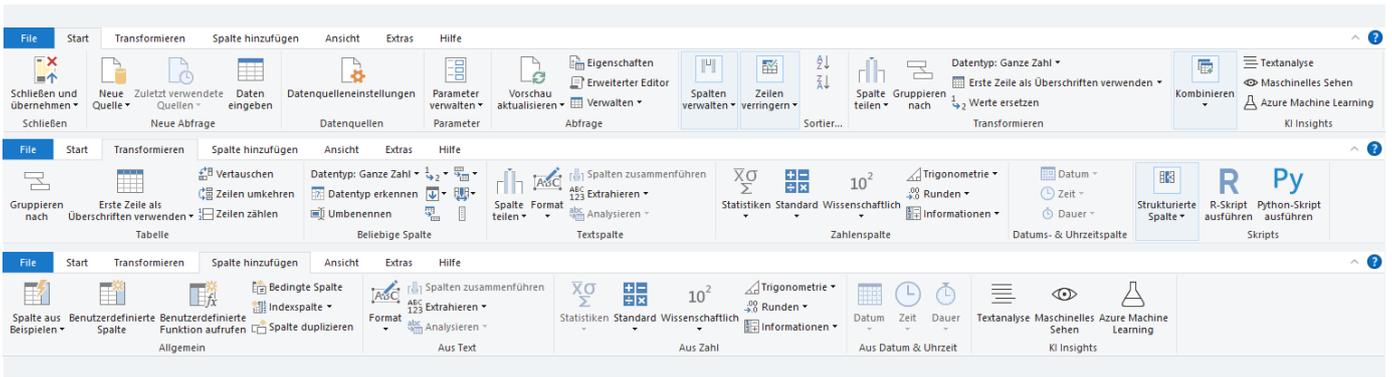


Abb. 4: Menüleisten ähneln dem gewohnten Office-Format

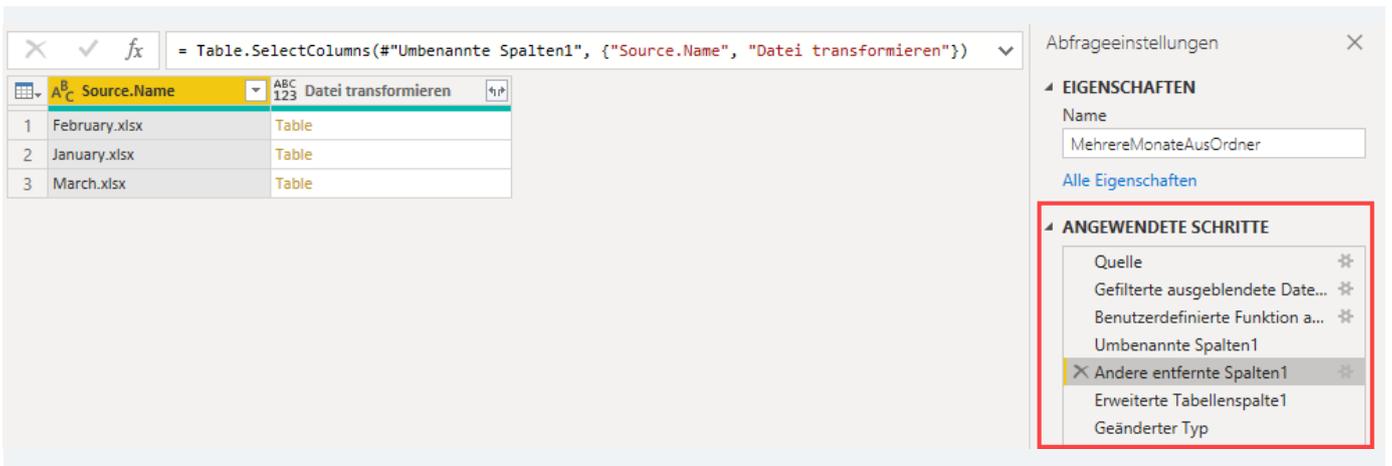


Abb. 5: Jeder Schritt wird mit M-Code aufgezeichnet

Kreuztabellenform

Gute Lesbarkeit, aber schlecht weiterzuverarbeiten

Produkt	Produkt	Kostenart	GJ 2016 gesamt Plan * 1.000 E	GJ 2016 kumuliert Plan * 1.000 E	GJ 2016 kumuliert Ist * 1.000 E
P01001	Produkt 01001	Kosten	-177,00	-157,00	-333,00
P01001	Produkt 01001	bewertete Arbeitsstage	-13,00	-5,00	-85,00
P01001	Produkt 01001	externe Dienstleister	-99,00	-60,00	-83,00
P01001	Produkt 01001	Instandhaltungskosten	-56,00	-62,00	-70,00

Datensatzform

Schlechte Lesbarkeit, aber gut weiterzuverarbeiten

	A	B	C	D	E
1	Produktnummer	Produktbezeichnung	Kostenart	Szenario	Wert
2	P01001	Produkt 01001	Kosten	GJ 2016 gesamt Plan	-177
3	P01001	Produkt 01001	Kosten	GJ 2016 kumuliert Plan	-157
4	P01001	Produkt 01001	Kosten	GJ 2016 kumuliert Ist	-333
5	P01001	Produkt 01001	Kosten	Monat MAI Plan	-209
6	P01001	Produkt 01001	Kosten	Monat MAI Ist	-220
7	P01001	Produkt 01001	bewertete Arbeitsstage	GJ 2016 gesamt Plan	-13
8	P01001	Produkt 01001	bewertete Arbeitsstage	GJ 2016 kumuliert Plan	-5
9	P01001	Produkt 01001	bewertete Arbeitsstage	GJ 2016 kumuliert Ist	-85
10	P01001	Produkt 01001	bewertete Arbeitsstage	Monat MAI Plan	-91

Abb. 6 Kreuztabellenform für die Analyse vs. Datensatzform für das Datenmodell

speicherten Schritte erneut ausgeführt, diesmal eben auf Basis der aktuellen Daten. Das bedeutet, dass man keinerlei zusätzliche Schritte unternehmen muss, um regelmäßig aktuelle Daten im System zu haben, außer die Datenaktualisierung anzustoßen. Zusätzliche, geänderte oder gelöschte Daten bzw. Tabellenzeilen werden automatisch auf den neuesten Stand gebracht.

Import aus Ordern

Häufig ist es jedoch so, dass neue Daten in der Datenquelle nicht als zusätzliche Zeilen in einer vorhandenen Tabelle geführt werden, sondern in neuen Dateien abgespeichert werden. Hat man beispielsweise für jeden Monat eine neue Datei und haben diese Dateien einen jeweils identischen Aufbau (Spaltenlayout), dann kann man die Funktionalität „Import aus Ordner“ verwenden. Dieser Modus sorgt dafür, dass alle ausgewählten Dateien eines Ordners (falls gewünscht inkl. Unterordner) importiert werden. Dann erhält man die Möglichkeit, sämtliche Transformationen, die man durchführen möchte, in einer Abfrage zu erstellen.

Die Schritte, die in dieser Abfrage definiert wurden, werden dann automatisch auf alle anderen importierten Dateien angewendet. So wird eine einheitliche Transformation für sämtliche Daten gewährleistet. Danach kann man die Daten mit einem Mausklick in einer großen Gesamttabelle zusammenfassen, die dann über das Monatsfeld unterscheidbar wird. Und auch hier handelt es sich um eine nachhaltige Lösung: Kommen neue Dateien in diesem Ordner dazu, werden diese automatisch bei der nächsten Datenaktualisierung in Power BI berücksichtigt.

Importe nicht nur auf Tabellen beschränkt

Auch für Daten, die nicht in Tabellenform vorliegen, gibt es Standardkonnektoren. Textdateien können eingelesen werden und sofern sie einer festgelegten Struktur folgen, kann man auch hier Extrakte in Tabellenform überführen. Für Importe aus dem Internet gibt es sogar einen Assistenten, der die Auswahl der auszulesenden Daten grafisch unterstützt: Man klickt die Elemente an, die man haben möchte und gibt ihnen einen Feldnamen. Diese Zuordnung kann man danach auf eine Liste von Webseiten anwenden, die gleichartig gestaltet sind und man erhält eine strukturierte Zusammenfassung der Informationen von mehreren Webseiten, die sich regelmäßig aktualisieren lässt.

Sollte sich die benötigte Datenquelle nicht unter den vorhandenen Datenkonnektoren befinden, kann man sich einen eigenen Konnektor schreiben (lassen). Viele solcher sog. "custom connectors" kann man sich aber auch bereits aus dem Internet herunterladen, wie z.B. für die Microsoft Graph API, mit der man an zahlreiche Office 365-Objekte herankommt.

Häufig ist es jedoch nicht damit getan, sich nur einfach mit den Daten des Quellsystems zu verbinden, denn die Daten sind mitunter in ihrer ursprünglichen Form nicht für die Weiterverarbeitung in Power BI geeignet. So kann es erforderlich sein, dass die Datenstruktur angepasst werden muss oder inhaltliche Änderungen vorgenommen werden müssen. Power Query bietet dafür umfangreiche Transformationsmöglichkeiten.

Transformationen

Nachdem man die Daten in Power Query importiert hat und die Vorschau sieht, kann man Transformationen durchführen. Dies kann für viele Zwecke über die grafische Oberfläche mit Mausklicks durchgeführt werden. Die Menüs sind an die Office-Programme angelehnt:

Jeder Mausklick wird dabei aufgezeichnet (ähnlich wie beim VBA-Makro-Rekorder) und ist als einzelner Schritt unter „Angewendete Schritte“ sichtbar:

So kann man jederzeit zu einem früheren Schritt zurückgehen und ihn ggf. modifizieren oder auch lö-

sen. Ebenso lassen sich Schritte zwischen bestehenden Schritten einfügen. Sehr viele Operationen, die man aus Excel kennt, kann man auch in Power Query vornehmen. Mathematische Operationen, Formatkonvertierung, Suchen, Filtern, Verkettungen u.v.m. Darüber hinaus bietet Power Query jedoch auch nützliche Funktionalitäten, die man aus Excel so nicht kennt und für die man in Excel aufwändig VBA programmieren müsste:

Entpivotieren

Während beim Pivotieren Feldattribute in Spaltenüberschriften verwandelt werden, geschieht beim Entpivotieren genau das Gegenteil: Spaltenüberschriften werden zu Werten in einzelnen Zeilen. So ein Schritt ist häufig erforderlich, um die Daten in das für das Power BI-Datenmodell erforderliche Format zu konvertieren. Hat man z.B. eine Tabelle, in der für jeden Monatswert eine neue Spalte erzeugt wurde, kann man durch Entpivotieren eine 2-spaltige Tabelle erzeugen: Die erste Spalte enthält die Monatsnamen und die zweite die entsprechenden Werte.

SQL-Joins und Union

Häufig besteht die Notwendigkeit, existierende Tabellen um Daten aus anderen Tabellen anzureichern. Excel verfügt über die Funktionen SVERWEIS oder INDEX/VERGLEICH, um Werte aus verschiedenen Tabellen miteinander in Beziehung zu bringen. Power Query verfügt über ähnliche Optionen, deren Funktionalität jedoch sehr nah an SQL JOINS angelehnt ist. So wird beispielsweise bei einem LEFT OUTER JOIN, der in der rechten Tabelle mehrere Entsprechungen findet, ungleich zum SVERWEIS nicht nur der erste gefundene Wert zurückgegeben, sondern alle Werte. Sehr nützlich sind auch INNER JOINS, mit denen man effizient Schnittmengen mehrerer Tabellen bilden kann und LEFT ANTI JOINS z.B., der nur solche Zeilen enthält, die es in der anderen Tabelle nicht gibt. Dies erleichtert z.B. Kontenabstimmungen enorm.

Unterstützung durch Künstliche Intelligenz

Manchmal hat man es mit Daten zu tun, bei denen man gar nicht so richtig weiß, welche Schritte man anwenden muss, um an die gewünschten Ergebnisse zu kommen. Hier helfen dann Funktionalitäten, die Algorithmen aus der künstlichen Intelligenz anwenden. „Spalte aus Beispielen“ z.B. ermöglicht die Extraktion von Werten aus einer Spalte anhand von einigen richtigen Trainingswerten. Werden die Muster daraus richtig erkannt, können diese automatisch auf die restlichen Werte angewendet werden. Typische Anwendungsfälle hierfür sind Aufspaltung von Adressen, die in einem einzigen Feld erfasst wurden. Ein großer Vorteil hierbei ist, dass die Logik der künstlichen Intelligenz nicht im Verborgenen bleibt, sondern durch M-Code dokumentiert wird. So kann man den erzeugten Code überprüfen und falls nötig, manuell anpassen.



Dipl.-Kfm. (FH) Lars Schreiber

ist freiberuflicher Berater, Entwickler und Trainer für Microsoft Power BI, Microsoft MVP, Hamburg. Lars@ssbi-blog.de



Dipl.-Kffr. Imke Feldmann

ist freiberufliche Beraterin, Entwicklerin und Trainerin für Microsoft Power BI und Microsoft MVP. info@thebiccountant.com

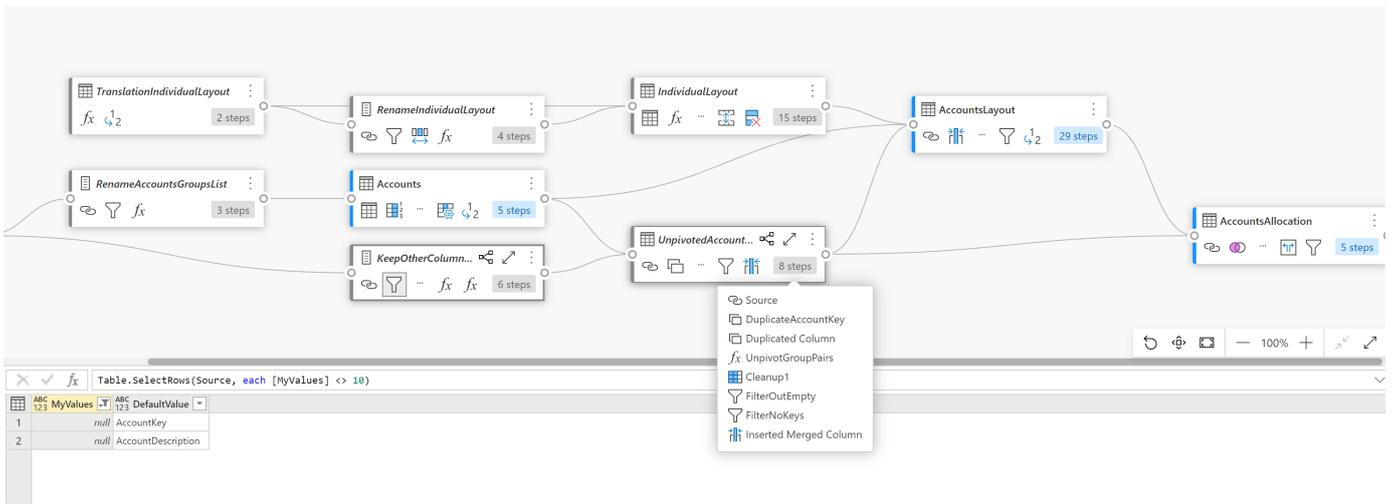


Abb. 7: Neue Diagrammsicht stellt Transformationen komfortabel und anschaulich dar

Einmal definieren, mehrfach verwenden

Nicht selten bestehen Transformationen aus mehreren Schritten, die man auch gerne für andere Zwecke wiederverwenden möchte. Hat man z.B. eine Lösung gefunden, die einem die Anzahl der Werktage zwischen einem Anfangs- und Enddatum ausrechnet, möchte man diese Logik vielleicht auch in einer anderen Abfrage, Power BI-Datei oder auch in Power Query in Excel verwenden. Hierfür bietet Power Query eine sehr komfortable Möglichkeit, diese Logik in eine eigene Funktion zu überführen. Diese könnte das Anfangs- und Enddatum als Parameter verwenden, die man dann mit individuellen Werten aus der neuen Umgebung füllt.

Darüber hinaus kann man auch einfach komplette Abfragen per Copy & Paste aus einer Power BI-Datei in andere Anwendungen kopieren, die Power Query verwenden. Viele solcher Funktionen kann man sich auch schon im Internet herunterladen, wie z.B. hier: <https://bit.ly/3nNyyVO>.

Einmal lernen, mehrfach anwenden

Power Query wurde auch noch in zahlreiche andere Microsoft-Programme integriert, wie z.B. Excel, Dataverse (ehemals Common Data Service), Azure Data Factory (dort als "Wrangling dataflows" bezeichnet) sowie in den SQL-Konnektor in Power Automate (ehemals „Flow“). Der Funktionsumfang von Power Query ist in diesen anderen Anwendungen aktuell jedoch mehr oder weniger eingeschränkt, verglichen zu den hier beschriebenen Funktionen von Power Query in Power BI Desktop.

zuerst in Power BI Desktop gesehen. Dies hat sich jüngst allerdings etwas geändert und einige neue beeindruckende Features sind zuerst in Dataflows - der Power Query-Online-Version - aufgetaucht: Die Diagrammsicht z.B., in der nicht nur die Beziehungen der einzelnen Abfragen zueinander abgebildet und modifiziert werden, sondern auch einzelne Schritte innerhalb der Abfragen angezeigt werden können:

Fazit

Power Query ermöglicht es Fachanwendern, regelmäßig wiederholbare Prozeduren zu entwickeln, die Daten aus unterschiedlichen Quellsystemen laden, bereinigen und umformen. Dadurch lässt sich ein hoher Automatisierungsgrad für wiederkehrende Tätigkeiten erreichen, der nicht nur viel Zeit sparen sondern auch helfen kann, Fehler zu vermeiden.

Viele Transformationen können über benutzerfreundliche Schaltflächen in einer grafischen Oberfläche durchgeführt werden. Für fortgeschrittene Anforderungen stehen Integrationen mit anderen Lösungen und eine leistungsfähige Programmiersprache voll integriert zur Verfügung. So erhält man eine saubere Datenbasis für das Power BI-Datenmodell, welches wir in der nächsten Ausgabe präsentieren werden. ■



Summary

Power Query ist eine Komponente in Power BI, mit der man Daten aus Quellsystemen importieren, bereinigen, filtern, umformen oder anderweitig bearbeiten kann, bevor man sie in das Power BI-Datenmodell lädt. Das ermöglicht Fachanwendern, regelmäßig wiederholbare Prozeduren zu entwickeln, die Daten aus unterschiedlichen Quellsystemen laden, bereinigen und umformen. Das ermöglicht einen hohen Automatisierungsgrad für wiederkehrende Tätigkeiten, der nicht nur viel Zeit sparen sondern auch helfen kann, Fehler zu vermeiden.

In Power BI gibt es darüber hinaus auch eine Power Query Anwendung in der Cloud, die sich Dataflows nennt. Dort kann man die Ergebnisse der Transformationen in allgemein zugänglichen Zielen wie dem Azure Data Lake Storage oder Dataverse abspeichern. Dadurch stehen diese Daten dann auch anderen Programmen zur Verfügung. Aber auch hier gibt es noch Einschränkungen im Funktionsumfang.

Ausblick

Power Query wird von Microsoft intensiv weiterentwickelt und nicht nur mit zusätzlichen Funktionalitäten ausgestattet, sondern auch in weitere Anwendungen integriert. Bislang haben wir die Neuigkeiten immer