

Gil Raviv

Power Query

In Excel und Power BI
Daten sammeln, kombinieren
und transformieren





Gil Raviv, Microsoft Data Platform MVP und Power BI-Blogger unter <https://DataChant.com>. Als Senior Program Manager für Microsoft Excel leitete er das Design und die Integration von Power Query als Datenabruf-Data-Wrangling-Technologie der nächsten Generation von Excel 2016 und wurde zum ausgewiesenen Experten von M, der Formelsprache von Power Query. Raviv verfügt über 20 Jahre Erfahrung im Bereich Programmierung und ist Inhaber von vier US-Patenten in den Bereichen soziale Netzwerke, Cybersicherheit und Datenanalyse. Er hat Datenanalysen und Big Data genutzt, um Sicherheitsprodukte, angefangen bei fortgeschrittener Bedrohungserkennung bis hin zum Kinderschutz auf Facebook, bereitzustellen.

In seinem Blog auf *DataChant.com* befasst sich Gil Raviv mit Power BI und Power Query, seit er Anfang 2016 in sein neues Zuhause nach Chicago gezogen ist. Als Group Manager Analytics Practice bei Avade begleitet er Fortune-500-Kunden bei der Entwicklung moderner Self-Service-Analytics-Lösungen auf Basis von Power BI und Azure.

Sie erreichen Gil Raviv über gilra@datachant.com.

Power Query

In Excel und Power BI Daten sammeln, kombinieren und transformieren

Gil Raviv

Gil Raviv

Übersetzung: Rainer G. Haselier

Lektorat: Sandra Bollenbacher

Copy-Editing: Petra Heubach-Erdmann, Düsseldorf

Satz: Gerhard Alfes, mediaService, Siegen, www.mediaservice.tv

Herstellung: Stefanie Weidner

Umschlaggestaltung: Helmut Kraus, www.exclam.de

Druck und Bindung: mediaprint solutions GmbH, 33100 Paderborn

ISBN:

Print 978-3-86490-727-2

PDF 978-3-96088-926-7

ePub 978-3-96088-927-4

mobi 978-3-96088-928-1

1. Auflage 2020

Translation Copyright für die deutschsprachige Ausgabe © 2020 dpunkt.verlag GmbH

Wieblinger Weg 17

69123 Heidelberg

Authorized translation from the English language edition, entitled COLLECT, COMBINE, AND TRANSFORM DATA USING POWER QUERY IN EXCEL AND POWER BI, 1st Edition by GIL RAVIV, ISBN 978-1509307951, published by Pearson Education, Inc, publishing as Microsoft Press, Copyright © 2019 by Gil Raviv

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

ISBN of the English language edition: 9781509307951

German language edition published by DPUNKT.VERLAG GMBH, Copyright © 2020

Hinweis:

Dieses Buch wurde auf PEFC-zertifiziertem Papier aus nachhaltiger Waldwirtschaft gedruckt. Der Umwelt zuliebe verzichten wir zusätzlich auf die Einschweißfolie.



Schreiben Sie uns:

Falls Sie Anregungen, Wünsche und Kommentare haben, lassen Sie es uns wissen: hallo@dpunkt.de

Die vorliegende Publikation ist urheberrechtlich geschützt. Alle Rechte vorbehalten. Die Verwendung der Texte und Abbildungen, auch auszugsweise, ist ohne die schriftliche Zustimmung des Verlags urheberrechtswidrig und daher strafbar. Dies gilt insbesondere für die Vervielfältigung, Übersetzung oder die Verwendung in elektronischen Systemen.

Es wird darauf hingewiesen, dass die im Buch verwendeten Soft- und Hardware-Bezeichnungen sowie Markennamen und Produktbezeichnungen der jeweiligen Firmen im Allgemeinen warenzeichen-, marken- oder patentrechtlichem Schutz unterliegen.

Alle Angaben und Programme in diesem Buch wurden mit größter Sorgfalt kontrolliert. Weder Autor noch Verlag noch Übersetzer können jedoch für Schäden haftbar gemacht werden, die in Zusammenhang mit der Verwendung dieses Buches stehen.

5 4 3 2 1 0



Inhaltsverzeichnis

Vorwort	xiii
Einleitung	xv
Für wen ist dieses Buch?	xvi
Annahmen über Sie, den Leser	xvi
Wie ist dieses Buch aufgebaut?	xvii
Über die Begleitdateien	xix
Systemvoraussetzungen	xxiii
Hinweiskästen	xxiii
Support und Feedback	xxiv
Errata, Updates und Support	xxiv
Wir wollen von Ihnen hören	xxiv
Bleiben Sie am Ball	xxiv

Kapitel 1

Einführung in Power Query	1
Was ist Power Query?	2
Ein kurzer geschichtlicher Überblick zu Power Query	3
Wo finde ich Power Query?	6
Hauptkomponenten von Power Query	8
Daten abrufen und Konnektoren	8
Die Hauptbereiche des Power Query-Editors	10
Übung 1-1: Ein erster Blick auf Power Query	16
Zusammenfassung	22

Kapitel 2

Grundlegende Herausforderungen bei der Datenvorbereitung	23
Die Bedeutung aus codierten Spalten extrahieren	24
Die AdventureWorks-Herausforderung	24
Übung 2-1: Der alte Weg: Excel-Formeln verwenden	26
Übung 2-2, Teil 1: Der neue Weg	27

Übung 2-2, Teil 2: Nachschlagetabellen zusammenführen	31
Übung 2-2, Teil 3: Fakten- und Nachschlagetabellen	35
Das Feature <i>Spalte aus Beispielen</i> verwenden	38
Übung 2-3, Teil 1: Einführung in <i>Spalte aus Beispielen</i>	38
Praktische Verwendung von <i>Spalte aus Beispielen</i>	40
Übung 2-3, Teil 2: Größe in Buckets/Klassen konvertieren	41
Informationen aus Textspalten extrahieren	44
Übung 2-4: Hyperlinks aus Nachrichten extrahieren	44
Datumswerte bearbeiten	53
Übung 2-5: Verarbeitung unterschiedlicher Datumsformate	53
Übung 2-6: Datumswerte aus verschiedenen Gebietsschemata	56
Datum- und Uhrzeitelemente extrahieren	59
Das Modell vorbereiten	60
Übung 2-7: Aufteilen von Daten in Nachschlagetabellen und Faktentabellen	61
Wenn eine Beziehung fehlschlägt	63
Übung 2-8: Durch Trennzeichen getrennte Werte in Zeilen aufteilen	64
Zusammenfassung	66

Kapitel 3

Daten aus mehreren Quellen kombinieren	69
Ein paar Tabellen anfügen	69
Zwei Tabellen anfügen	70
Übung 3-1: Beispiel Fahrräder und Zubehör	70
Übung 3-2, Teil 1: Abfragen als neu anfügen verwenden	72
Übung 3-2, Teil 2: Abfrageabhängigkeiten und Verweise	74
Drei oder mehr Tabellen anfügen	76
Übung 3-2, Teil 3: Fahrräder + Zubehör + Komponenten	76
Übung 3-2, Teil 4: Fahrräder + Zubehör + Komponenten + Kleidung	78
Tabellen in größerem Maßstab anhängen	79
Tabellen aus einem Ordner anfügen	80
Übung 3-3: AdventureWorks-Produkte aus einem Ordner anfügen	80
Ein paar Gedanken zum Importieren aus Ordnern	83
Tabellenblätter aus einer Arbeitsmappe anfügen	84
Übung 3-4: Tabellenblätter anfügen – Die Lösung	85
Ein robuster Ansatz zum Kombinieren mehrerer Tabellenblätter	89
Zusammenfassung	91

Kapitel 4

Kombinieren von Tabellen mit unterschiedlichen Datenstrukturen	93
Das Problem von Tabellen mit unterschiedlichen Datenstrukturen	94

Was sind nicht übereinstimmende Tabellen?	94
Symptome und Risiken von nicht übereinstimmenden Tabellen	95
Übung 4-1: Nicht übereinstimmende Spaltennamen auflösen – Der reaktive Ansatz	96
Nicht übereinstimmende Tabellen aus einem Ordner kombinieren	97
Übung 4-2, Teil 1: Das Symptom der fehlenden Werte demonstrieren ...	98
Übung 4-2, Teil 2: Die Annahme der gleichen Reihenfolge und die Vereinheitlichung der Überschriften als Lösung	100
Übung 4-3: Einfache Normalisierung mit <i>Table.TransformColumnNames</i>	102
Die Konvertierungstabelle	105
Übung 4-4: Transponieren und die Verwendung einer Konvertierungstabelle	107
Übung 4-5: Entpivotieren, Zusammenführen und erneut Pivotieren	112
Übung 4-6: Nur die Spaltennamen transponieren	114
Übung 4-7: M zum Normalisieren von Spaltennamen verwenden	121
Zusammenfassung	124

Kapitel 5

Den Kontext erhalten	127
Den Kontext der Dateinamen und Arbeitsblätter beibehalten	128
Übung 5-1, Teil 1: Die Technik der benutzerdefinierte Spalte	128
Übung 5-1, Teil 2: Umgang mit Kontextinformationen in Dateinamen und Tabellenblattnamen	130
Titel vor dem Anfügen beibehalten	131
Übung 5-2: Die Titel mithilfe von Drilldown erhalten	132
Übung 5-3: Titel beim Import aus einem Ordner erhalten	136
Titel nach dem Anfügen beibehalten	139
Übung 5-4: Aus mehreren Tabellenblättern in derselben Arbeitsmappe die Titel beibehalten	139
Kontexthinweise verwenden	144
Übung 5-5: Verwenden einer Indexspalte als Hinweis	145
Übung 5-6: Identifizieren des Kontexts anhand der Nachbarschaft zu einer Zelle	149
Zusammenfassung	153

Kapitel 6

Tabellen entpivotieren	155
Schlecht gestaltete Tabellen erkennen	156
Entpivotieren – Eine Einführung	159
Übung 6-1: »Spalten entpivotieren« und »Andere Spalten entpivotieren« verwenden	160
Übung 6-2: Nur ausgewählte Spalten entpivotieren	163

Umgang mit Gesamtergebnissen	164
Übung 6-3: Gesamtergebnisse entpivotieren	165
2×2 Hierarchieebenen entpivotieren	168
Übung 6-4: 2×2 Hierarchieebenen mit Datumswerten entpivotieren	168
Übung 6-5: 2×2 Hierarchieebenen entpivotieren	172
Umgang mit Teilergebnissen in entpivotierten Daten	174
Übung 6-6: Umgang mit Teilergebnissen	175
Zusammenfassung	177

Kapitel 7

Fortgeschrittenes Entpivotieren und Pivotieren von Tabellen	179
Tabellen mit mehreren Hierarchieebenen entpivotieren	180
Die virtuelle PivotTable, Zeilenfelder und Spaltenfelder	180
Übung 7-1: In den AdventureWorks-Daten $N \times M$ Hierarchieebenen entpivotieren	181
Die Sequenz für das Entpivotieren verallgemeinern	184
Übung 7-2: Am Ende beginnen	185
Übung 7-3: <i>FnUnpivotSummarizedTable</i> erstellen	187
Die Transformation <i>Spalte pivotieren</i>	199
Übung 7-4: Eine falsch entpivotierte Tabelle wiederherstellen	199
Übung 7-5: Tabellen mit mehrzeiligen Datensätzen pivotieren	201
Zusammenfassung	206

Kapitel 8

Herausforderungen bei der Zusammenarbeit lösen	207
Lokale Dateien, Parameter und Vorlagen	208
Zugriff auf lokale Dateien – auf die verkehrte Weise	208
Übung 8-1: Verwenden eines Parameters für einen Pfadnamen	209
Übung 8-2: Erstellen einer Vorlage in Power BI	211
Übung 8-3: Parameter in Excel verwenden	214
Übung 8-3, Teil 2: Die Datenkombination neu erstellen	218
Mit freigegebenen Dateien und Ordnern arbeiten	222
Daten aus Dateien auf OneDrive for Business oder SharePoint importieren	222
Übung 8-4: Migrieren Ihrer Abfragen, um eine Verbindung mit OneDrive for Business oder SharePoint herzustellen	224
Übung 8-5: Von lokalen Ordnern zu SharePoint-Ordnern migrieren	227
Sicherheitsaspekte	229
Alle Abfragen mithilfe des Dokumentinspektors in Excel entfernen	230
Zusammenfassung	231

Einführung in die Power Query-Formelsprache M	233
M lernen	234
Die Lernstufen	234
Stufe 1: Nur Benutzeroberfläche	235
Stufe 2: Einfache Formeländerungen in der Bearbeitungsleiste	235
Stufe 3: M in benutzerdefinierten Spalten	236
Stufe 4: Benutzerdefinierte Funktionen und der erweiterte Editor	236
Stufe 5: Fortgeschrittene Schleifen (<i>List.Accumulate</i> und <i>List.Generate</i>)	237
Stufe 6: Freie Codeerstellung im Erweiterten Editor	237
Online-Ressourcen	238
Offline-Ressourcen	238
Übung 9-1: Mit <i>#shared</i> die integrierten Funktionen erkunden	239
Die Bausteine von M	240
Übung 9-2: <i>Hello World</i>	240
Der <i>let</i> -Ausdruck	242
Zugriffsbereiche beim Zusammenführen von Ausdrücken aus mehreren Abfragen	245
Typen, Operatoren und integrierte Funktionen in M	248
Grundlegende M-Typen	250
Der Datentyp Zahl (Number)	250
Der Datentyp Uhrzeit (Time)	251
Der Datentyp Datum (Date)	252
Der Datentyp Dauer (Duration)	253
Der Datentyp Text	254
Der Datentyp null	254
Der Datentyp logisch	255
Komplexe M-Typen	256
Der Datentyp Liste (List)	256
Der Typ Datensatz (Record)	260
Der Typ Tabelle (Table)	263
Bedingungen und <i>if</i>-Ausdrücke	265
Benutzerdefinierte Funktionen	268
Der Ausdruck <i>each</i>	271
Fortgeschrittene Themen	272
Fehlerbehandlung	272
Faule und eifrige Auswertung	274
Schleifen	274
Rekursion	275
<i>List.Generate</i>	276
<i>List.Accumulate</i>	277
Zusammenfassung	278

Kapitel 10

Von Stolperfallen zu robusten Abfragen	281
Ursachen und Auswirkungen der Stolperfallen	282
Bewusstsein	284
Bewährte Verfahren	285
Modifikationen der M-Ausdrücke	285
Stolperfälle 1: Die Bearbeitungsleiste ignorieren	286
Übung 10-1: Verwenden der Bearbeitungsleiste, um statische Verweise auf Spaltennamen zu entdecken	287
Stolperfälle 2: Geänderte Typen	289
Stolperfälle 3: Gefährliches Filtern	292
Übung 10-2, Teil 1: Schwarze Produkte filtern	292
Die Logik hinter der Filterbedingung	294
Übung 10-2, Teil 2: Werte im Filterbereich suchen	296
Stolperfälle 4: Spalten neu anordnen	297
Übung 10-3, Teil 1: Eine Teilmenge der Spalten neu anordnen	297
Übung 10-3, Teil 2: Die benutzerdefinierte Funktion <i>FnReorderSubsetOfColumns</i>	300
Stolperfälle 5: Spalten entfernen und auswählen	301
Übung 10-4: Umgang mit den Zufallsspalten in der Wide World Importers-Tabelle	301
Fehlende Spalte ignorieren	302
Spalten basierend auf ihrer Position auswählen oder entfernen	302
Spalten basierend auf ihren Namen auswählen oder entfernen	303
Stolperfälle 6: Spalten umbenennen	303
Übung 10-5: Die Zufallsspalten in der Wide World Importers-Tabelle umbenennen	304
Stolperfälle 7: Eine Spalte in Spalten aufteilen	307
Übung 10-6: Eine falsche Aufteilung durchführen	308
Stolperfälle 8: Spalten zusammenführen	311
Weitere Stolperfallen und Techniken für robuste Abfragen	313
Zusammenfassung	313

Kapitel 11

Einfache Textanalyse	317
In Textspalten nach Schlüsselwörtern suchen	318
Übung 11-1: Einfache Erkennung von Schlüsselwörtern	318
Schlüsselwörter mithilfe eines kartesischen Produkts erkennen	322
Übung 11-2: Ein kartesisches Produkt implementieren	324

Übung 11-3: Erkennen von Schlüsselwörtern mithilfe einer benutzerdefinierten Funktion	332
Welche Methode sollte verwendet werden: statische Suche, kartesisches Produkt oder benutzerdefinierte Funktion?	335
Eine Textspalte in einzelne Wörter aufteilen	335
Übung 11-4: Einfaches Aufteilen in einzelne Wörter	335
Übung 11-5: Stoppwörter herausfiltern	341
Übung 11-6: Mithilfe von »nach Wörtern teilen« nach Schlüsselwörtern suchen	344
Übung 11-7: Wortwolken in Power BI erstellen	353
Zusammenfassung	354

Kapitel 12

Fortgeschrittene Textanalyse: Bedeutung extrahieren	357
Microsoft Azure Cognitive Services	358
API-Schlüssel und Ressourcenbereitstellung in Azure	360
Vor- und Nachteile der Verwendung der Cognitive Services in Power Query	363
Textübersetzung	366
Referenz der Textübersetzungs-API	367
Übung 12-1: Einfache Übersetzung	367
Übung 12-2: Mehrere Nachrichten übersetzen	373
Stimmungsanalyse	378
Wie lautet der API-Aufruf für die Stimmungsanalyse?	379
Übung 12-3: Implementieren von <i>FnGetSentiment</i> , einer benutzerdefinierten Funktion für die Stimmungsanalyse	381
Übung 12-4: Stimmungsanalyse für große Datasets ausführen	393
Schlüsselwörter extrahieren	395
Übung 12-5: Die Logik der Stimmungsanalyse zum Extrahieren von Schlüsselwörtern verwenden	396
Unterstützung für mehrere Sprachen	398
Übung 12-6: Die Logik der Stimmungsanalyse zur Erkennung der Sprache verwenden	400
Zusammenfassung	401

Kapitel 13

Abschlussprojekt: Alles miteinander kombinieren	403
Übung 13-1: Retten Sie Wide World Importers	403
Tipps	405
Teil 1: Start der Lösung	405
Teil 2: Die Funktion für das Entpivotieren aufrufen	408

Teil 3: Die Pivotieren-Sequenz für die Umsätze 2018	409
Teil 4: Die Umsätze 2018 und 2015 bis 2017 kombinieren	410
Übung 13-2: Tabellen vergleichen und den Hacker aufspüren	411
Tipps	411
Übung 13-2: Die Lösung	412
Die verborgene Nachricht des Hackers in der kompromittierten Tabelle entdecken	414
Zusammenfassung	415
Abbildungsverzeichnis	416
Index	417

Vorwort



Als wir uns daran machten, das ursprüngliche Power Query-Add-In für Excel zu erstellen, hatten wir eine einfache, aber ehrgeizige Mission: Verbindungen zu den Daten der Welt herzustellen und sie zu transformieren. Fünf Jahre später haben wir das anfängliche Excel-Add-In hinter uns gelassen und PowerQuery ist direkt in Excel, Power BI, PowerApps und eine wachsende Reihe von Produkten, die Daten extrahieren und transformieren müssen, integriert. Jedoch bleibt unsere ursprüngliche Mission weitestgehend unverändert. Mit der ständig zunehmenden Heterogenität der Daten scheint unsere Mission in vielerlei Hinsicht noch ehrgeiziger und herausfordernder zu sein als jemals zuvor. Ein Großteil der heutigen Computerlandschaft ist auf Daten ausgerichtet, aber die Daten sind nicht immer da, wo oder wie Sie sie benötigen – wir werden Power Query mit dem Ziel weiterentwickeln, diese Lücke zwischen den rohen und den gewünschten Datenzuständen zu überbrücken.

Während der gesamten Entwicklung und Weiterentwicklung von Power Query hat die Anwender-Community durch Vorschläge und Feedback eine entscheidende Rolle bei der Gestaltung des Produkts gespielt. Die Community hat auch eine zentrale Rolle bei der Entwicklung wertvoller Lerninhalte gespielt. Als einer der treibenden Kräfte der nativen Integration von Power Query in Excel 2016 ist Gil gut aufgestellt, um wertvolle Einblicke und Tipps für eine Vielzahl von Szenarien zu geben. Auch nach seiner Zeit bei Microsoft ist Gil ein aktives und einflussreiches Mitglied der Power Query-Community geblieben. Viel Spaß beim Erstellen von Abfragen!

– Sid Jayadevan, Engineering Manager für Power Query, Microsoft Corporation

Für Leser, die mit Power Query nicht vertraut sind, ist es eine unglaublich leistungsstarke und erweiterbare Engine, die das Herzstück der Microsoft BI-Tools darstellt. Es verbessert Self-Service Business Intelligence (BI) mit einer intuitiven und konsistenten Benutzeroberfläche zum Erforschen, Kombinieren und Verfeinern von Daten über eine Vielzahl von Quellen hinweg. Da die Datenvorbereitung in der Regel ungefähr 80% jeder BI-Lösung ausmacht, ist ein gutes Verständnis von Power Query der erste Schritt bei jeder Art von Berichtserstellung und Datenanalyse. Neben den zentralen Power Query-Features behandelt Gil auch fortgeschrittene Themen, wie zum Beispiel die Verwendung von Power Query zur Automatisierung der Datenvorbereitung und -bereinigung, die Verbindung zu sozialen Netzwerken, um zu erfassen, was Ihre Kunden über Ihr Unternehmen sagen, und wie Sie Power Query für Dienste aus dem Bereich maschinelles Lernen nutzen, um Sentiment-Analysen durchzuführen, und wie man die Sprache M verwendet, um praktisch jede Art von Rohdaten zu einer Quelle von Erkenntnissen zu machen, aus denen Sie Nutzen ziehen können. Dieses Buch zeichnet sich dadurch aus, dass es zusätzliche Begleitinhalt mit fertigen Beispielen, Datenquellen und Schritt-für-Schritt-Anleitungen bietet.

Gil ist ein ehemaliges Mitglied des Excel-Teams und des Microsoft Data-Teams. Er hat einen direkten Beitrag zu den Features und dem Design von Power Query geliefert, verfügt über eine erstaunliche Fülle an Wissen zu Power Query und zeigt, wie mit Power Query schwierige Probleme bei der Datenintegration vereinfacht werden können. Trotz des inhärent erweiterbaren und benutzerfreundlichen Designs von Power Query kann es dennoch schwierig sein, es für Unternehmensszenarien einzusetzen. Glücklicherweise ist Gil Raviv als engagiertes Community-Mitglied, Verfasser von Forenbeiträgen, Konferenzmoderator, Peer-Mentor und Power BI MVP ein Meister darin, komplexe Konzepte in sehr einfache Schritte zu zerlegen, um so den Leser erfolgreich werden zu lassen. Daher ist dieses Buch ein Muss für jeden BI-Spezialisten, Datenanalytiker oder Geschäftsmann, der aus den Daten, mit denen er zu tun hat, Mehrwert erzielen möchte.

– Charles Sterling, Senior Program Manager, Microsoft Corporation



Einleitung

Wussten Sie, dass in Microsoft Excel, Power BI und anderen Produkten eine Datentransformationstechnologie integriert ist, mit der Sie an Ihren Daten wahre Wunder vollbringen, sich wiederholende manuelle Arbeit vermeiden und bis zu 80% Ihrer Zeit einsparen können?

- Jedes Mal, wenn Sie ähnliche Daten in Ihre Arbeitsmappe kopieren/einfügen und manuell bereinigen, verschwenden Sie wertvolle Zeit, möglicherweise ohne sich der Alternativen bewusst zu sein, mit denen Sie dies besser und schneller erledigen können.
- Jedes Mal, wenn Sie sich darauf verlassen, dass andere Ihre Daten in die richtige Form und den richtigen Zustand bringen, sollten Sie wissen, dass es eine einfachere Möglichkeit gibt: Gestalten Sie Ihre Daten einmal um und freuen Sie sich darüber, dass das in Zukunft automatisiert funktioniert.
- Jedes Mal, wenn Sie schnelle fundierte Entscheidungen treffen müssen, Sie sich aber massiven Herausforderungen bei der Datenbereinigung stellen müssen, wissen Sie, dass Sie diese Herausforderungen jetzt leicht bewältigen können und ein nie da gewesenes Potenzial erhalten, um die für das Erhalten von Einblicken benötigte Zeit zu verkürzen.

Sind Sie bereit für Veränderung? Sie sind dabei, die wahnsinnige Frustration der sich wiederholenden manuellen Datenbereinigung durch schiere Aufregung und Spaß zu ersetzen, und durch diesen Prozess können Sie sogar Ihre Datenqualität verbessern und neue Erkenntnisse gewinnen.

Excel, Power BI, Analysis Services und PowerApps nutzen eine bahnbrechende Technologie zum Abrufen und Transformieren von Daten, Power Query, die es jeder Person mit grundlegenden Excel-Kenntnissen ermöglicht, den Datenimport, die Umgestaltung von Daten und die Datenbereinigung zu automatisieren. Mit ein paar einfachen Klicks in der Benutzeroberfläche und einer einheitlichen Benutzererfahrung für eine Vielzahl von Datenquellen und Formaten können Sie jede Herausforderung bei der Datenvorbereitung lösen und zum Meister im Data-Wrangling werden.

In diesem Buch werden Sie echte Datenherausforderungen angehen und lernen, wie Sie diese mit Power Query lösen können. Im Rahmen von mehr als 70 Herausforderungen und 200 Übungsdateien im Begleitmaterial zu diesem Buch importieren Sie chaotische und unzusammenhängende Tabellen und arbeiten sich durch die Erstellung automatisierter und gut strukturierter Datensätze, die für die Analyse bereitstehen. Die meisten Techniken sind einfach zu befolgen und können leicht in Ihrem eigenen Arbeitsalltag wiederverwendet werden.

Für wen ist dieses Buch?

Dieses Buch wurde geschrieben, um geschäftliche Anwender und Berichtsautoren in Microsoft Excel und Power BI zu befähigen, das Potenzial von Power Query optimal auszuschöpfen. Das Buch ist auch für SQL Server- oder Azure Analysis Services-Entwickler relevant, die ihre ETL-Entwicklung (ETL = Extract, Transform, Load) beschleunigen möchten. Benutzer, die Apps mit Microsoft PowerApps erstellen, können ebenfalls dieses Buch nutzen, um komplexe Datasets in ihre Geschäftslogik zu integrieren.

Ganz gleich, ob Sie für sich wiederholende Datenvorbereitungsaufgaben in Excel zuständig sind oder Power BI-Berichte für Ihr Unternehmen entwickeln, dieses Buch ist genau für Sie. Analysten, Business Intelligence-Spezialisten und ETL-Entwickler können ihre Produktivität durch die Techniken in diesem Buch steigern. Da die Power Query-Technologie zum primären Datenstapel in Excel geworden ist und der Einsatz von Power BI enorm zugenommen hat, wird dieses Buch Ihnen helfen, den Weg in Ihrem Unternehmen zu ebnen und eine größere Wirkung zu erzielen.

Das Buch wurde geschrieben, damit alle Power Query-Benutzer neue Fähigkeiten erlernen können. Gleichgültig, ob Sie ein neuer, ein fortgeschrittener oder ein Benutzer mit moderaten Kenntnissen sind, Sie werden nützliche Techniken finden, die Ihnen helfen, die nächste Stufe zu erklimmen.

Annahmen über Sie, den Leser

Grundkenntnisse in Excel oder Power BI werden vorausgesetzt. Während jeder Excel-Benutzer von diesem Buch profitieren kann, würden Sie viel mehr davon profitieren, wenn Sie eines der folgenden Kriterien erfüllen. (Beachten Sie, dass die Erfüllung eines einzelnen Kriteriums ausreichend ist.)

- Sie kopieren und fügen regelmäßig Daten aus den gleichen Quellen in Excel ein und müssen diese Daten häufig bereinigen.
- Sie erstellen Berichte in Excel oder Power BI, die mit externen Datenquellen verbunden sind, und möchten diese verbessern.
- Sie sind mit PivotTables in Excel vertraut.
- Sie sind mit Power Pivot in Excel vertraut und möchten Ihre Datenmodelle vereinfachen.
- Sie sind mit Power Query vertraut und möchten auf die nächste Stufe wechseln.
- Sie entwickeln Geschäftsanwendungen mit PowerApps und müssen sich mit Datenquellen verbinden, deren Datensätzen wenig aufgeräumt sind.
- Sie sind Entwickler von Analysis Services und möchten Ihre ETL-Entwicklung beschleunigen.

Wie ist dieses Buch aufgebaut?

Das Buch ist in 13 Kapitel unterteilt, die bei allgemeinen und einfacheren Datenherausforderungen beginnen und zu fortgeschrittenen und spezifischen Szenarien übergehen, die es zu meistern gilt. Es ist vollgepackt mit praktischen Übungen und Schritt-für-Schritt-Lösungen, die die notwendigen Techniken für die Bewältigung realer Herausforderungen bei der Datenvorbereitung bereitstellen und als langfristige Lernressource dienen, unabhängig davon, wie viele neue Funktionen in Zukunft in Power Query veröffentlicht werden.

In Kapitel 1, »Einführung in Power Query«, wird Ihnen Power Query vorgestellt und Sie erhalten die grundlegenden Kenntnisse, um die nachfolgenden Übungen durchführen zu können.

In Kapitel 2, »Grundlegende Herausforderungen bei der Datenvorbereitung«, erfahren Sie, wie Sie relativ einfache typische Datenherausforderungen bewältigen. Wenn Sie im Rahmen Ihres Arbeitsalltags häufig Datenbereinigungsaufgaben durchführen, werden Sie dieses Kapitel sehr hilfreich finden. Sie lernen die einfachsten Techniken zur Automatisierung Ihrer Datenreinigungsaufgaben kennen, die Sie mit wenigen Mausklicks und ohne Programmierkenntnisse durchführen können. Wenn Power Query noch neu für Sie ist, sparen Sie bereits Zeit, wenn Sie die Techniken in diesem Kapitel befolgen.

In Kapitel 3, »Daten aus mehreren Quellen kombinieren«, erfahren Sie, wie Sie getrennte Datasets kombinieren und mehrere Tabellen im Power Query-Editor anfügen. Sie erfahren, wie Sie mehrere Arbeitsmappen aus einem Ordner anfügen und mehrere Arbeitsblätter auf robuste Weise kombinieren: Wenn später neue Arbeitsblätter hinzugefügt werden, genügt eine einzelne Aktualisierung des Berichts, um die neuen Daten an Ihren Bericht anzuhängen.

In Kapitel 4, »Kombinieren von nicht übereinstimmenden Tabellen«, erklimmen Sie die nächste Ebene und lernen, wie Sie nicht übereinstimmende Tabellen, also Tabellen, die zwar die gleichen Daten enthalten, jedoch unterschiedlich aufgebaut sind, kombinieren. In realen Szenarien sind Ihre Daten segmentiert und isoliert und das Format und die Struktur der Daten sind oft nicht konsistent. Wenn Sie lernen, wie Sie nicht übereinstimmende Tabellen normalisieren, erhalten Sie neue Einblicke in strategische Geschäftsszenarien.

In Kapitel 5, »Den Kontext erhalten«, erfahren Sie, wie Sie externen Kontext in Ihren Tabellen extrahieren und beibehalten und Titel sowie andere Metainformationen, wie Dateinamen und Arbeitsblattnamen, kombinieren, um die angehängten Tabellen mit diesen Informationen zu bereichern.

In Kapitel 6, »Tabellen entpivotieren«, erfahren Sie, wie Sie die Tabellenstruktur verbessern können, um eine bessere Darstellung der Entitäten zu erhalten, die die Daten darstellen. Sie erfahren, dass die Transformation Entpivotieren ein Eckpfeiler für die Bereinigung der Adressierung schlecht gestalteter Tabellen ist, und nutzen die Leistungsfähigkeit von Entpivotieren, um Ihre Tabellen für eine bessere Analyse neu zu strukturieren. Außerdem erfahren Sie, wie Sie mit verschachtelten Tabellen umgehen und warum und wie Sie Summen und Zwischensummen in Ihren Quelldaten ignorieren.

In Kapitel 7, »Fortgeschrittenes Entpivotieren und Pivotieren von Tabellen«, setzen Sie Ihre Reise zur Transformation Entpivotieren fort und verallgemeinern eine Lösung, die Ihnen hilft, beliebige zusammengefasste Tabelle zu entpivotieren, unabhängig davon, wie viele Hierarchieebenen auf Zeilen- und Spaltenebene vorliegen. Anschließend erfahren Sie, wie Sie Pivo-

tieren anwenden, um mehrzeilige Datensätze zu verarbeiten. Die in diesem Kapitel gezeigten Techniken ermöglichen es Ihnen, eine Vielzahl von Transformationen durchzuführen und überstrukturierte Datasets in eine leistungsstarke und agile Analyseplattform umzuformen.

Als Berichtsauteur geben Sie Ihre Berichte häufig an andere Autoren in Ihrem Team oder Unternehmen weiter. In Kapitel 8, »Herausforderungen bei der Zusammenarbeit lösen«, erfahren Sie mehr über grundlegende Herausforderungen bei der Zusammenarbeit und wie Sie diese mithilfe von Parametern und Vorlagen lösen können.

In Kapitel 9, »Einführung in die Power Query-Formelsprache M«, werden Sie tiefer in die Abfragesprache M eintauchen, die verwendet werden kann, um Ihre Abfragen anzupassen, um mehr zu erreichen und Ihre Transformation in einem größeren Umfang von Herausforderungen wiederzuverwenden. In diesem Kapitel lernen Sie die wichtigsten Bausteine von M kennen – Syntax, Operatoren, Typen und eine Vielzahl integrierter Funktionen. Wenn Sie kein fortgeschrittener Benutzer sind, können Sie dieses Kapitel überspringen und es zu einem späteren Zeitpunkt lesen. M vollständig zu meistern ist keine Voraussetzung, um ein Experte für das Data-Wrangling zu werden, jedoch kann die Fähigkeit, die M-Formeln bei Bedarf zu ändern, Ihre Möglichkeiten erheblich steigern.

Die Benutzererfahrung mit dem Power Query-Editor in Excel und Power BI ist sehr erträglich, da sie Ihre alltäglichen, aber entscheidenden Datenvorbereitungsaufgaben in einen automatisierten Aktualisierungsfluss verwandeln kann. Leider treten auf dem Weg zum Beherrschen des Data-Wrangling typische Fehler auf, die Sie möglicherweise im Power Query-Editor machen, was zur Erstellung anfälliger Abfragen führt, die nicht aktualisiert werden können oder zu falschen Ergebnissen führen, wenn sich die Daten ändern. In Kapitel 10, »Von Stolperfallen zu robusten Abfragen«, lernen Sie die häufigsten Fehler oder Fallstricke kennen und wie Sie diese durch robuste Abfragen vermeiden können, deren Aktualisierung nicht fehlschlägt und die nicht zu falschen Ergebnissen führen.

In Kapitel 11, »Grundlegende Textanalyse«, nutzen Sie Power Query, um grundlegende Einblicke in Textfeeds zu gewinnen. Viele Tabellen in Ihren Berichten enthalten möglicherweise bereits reichlich Textspalten, die bei der Analyse häufig ignoriert werden. Sie lernen, wie Sie allgemeine Transformationen anwenden, um Bedeutung aus Wörtern zu extrahieren, wie Sie Schlüsselwörter erkennen und häufig auftretende Wörter, die in der Regel keine Relevanz für den Dokumentinhalt besitzen, sogenannte Stoppwörter, ignorieren. Außerdem werden Sie lernen, wie Sie das kartesische Produkt für komplexe Textsuchen anwenden.

In Kapitel 12, »Fortgeschrittene Textanalyse: Bedeutung extrahieren«, werden Sie von grundlegenden zu erweiterten Textanalysen übergehen und lernen, wie Sie Sprachübersetzungen, Stimmungsanalysen und Schlüsselworterkennung mithilfe von Microsoft Cognitive Services anwenden. Mit dem Power Query-Webconnector und einigen grundlegenden M-Funktionen können Sie ohne die Hilfe von Datenwissenschaftlern oder Softwareentwicklern die wirkliche Bedeutung aus Text extrahieren und die Leistungsfähigkeit künstlicher Intelligenz nutzen.

Schließlich werden Sie in Kapitel 13, »Das finale Projekt: Alles miteinander kombinieren«, vor die letzte Herausforderung des Buches treten und Ihr ganzes Wissen auf die Probe stellen, indem Sie Ihre neuen erworbenen Data-Wrangling-Kenntnisse auf eine große Herausforderung anwenden. Wenden Sie die Techniken aus diesem Buch an, um Dutzende von Arbeitsblättern aus mehreren Arbeitsmappen zu kombinieren, die Daten zu entpivotieren und zu pivotieren und um Wide World Importers vor einem groß angelegten Cyberangriff zu retten!

Über die Begleitdateien

Die Begleitdateien zum Buch können Sie unter <https://www.dpunkt.de/powerquery.html> kostenlos herunterladen.

In den Begleitdateien finden Sie:

- Excel-Arbeitsmappen und CSV-Dateien, die als chaotische und schlecht formatierte Datenquellen für alle Übungen im Buch verwendet werden. Es ist daher nicht nötig, eine externe Datenbank zu installieren, um die Übungen durcharbeiten zu können.
- Excel-Arbeitsmappen und Power BI-Berichte mit den Lösungen, die die erforderlichen Abfragen enthalten, um die einzelnen Datenherausforderungen zu lösen.

In der folgenden Tabelle sind die Übungsdateien aufgeführt, die zum Ausführen der Übungen in diesem Buch erforderlich sind.

Kapitel	Dateien
Kapitel 1: Einführung in Power Query	K01U01.xlsx K01U01 - Lösung.xlsx K01U01 - Lösung.pbix
Kapitel 2: Grundlegende Herausforderungen bei der Datenvorbereitung	K02U01.xlsx K02U01 - Lösung.xlsx K02U02.xlsx K02U02 - Lösung - Teil 1.xlsx K02U02 - Lösung - Teil 2.xlsx K02U02 - Lösung - Teil 3.xlsx K02U02 - Lösung - Teil 1.pbix K02U02 - Lösung - Teil 2.pbix K02U02 - Lösung - Teil 3.pbix K02U03 - Lösung.xlsx K02U03 - Lösung - Teil 2.xlsx K02U03 - Lösung.pbix K02U03 - Lösung - Teil 2.pbix K02U04.xlsx K02U04 - Lösung.xlsx K02U04 - Lösung.pbix K02U05.xlsx K02U05 - Lösung.xlsx K02U05 - Lösung.pbix K02U06.xlsx K02U06 - Lösung.xlsx K02U06 - Lösung.pbix K02U07.xlsx K02U07 - Lösung.xlsx K02U07 - Lösung.pbix K02U08.xlsx K02U08 - Lösung.xlsx K02U08 - Lösung.pbix
Kapitel 3: Daten aus mehreren Quellen kombinieren	K03U01 - Fahrräder.xlsx K03U01 - Kleidung.xlsx K03U01 - Komponenten.xlsx K03U01 - Zubehör.xlsx K03U03 - Produkte.zip

Kapitel	Dateien
	K03U03 - Lösung.xlsx K03U03 - Lösung.pbix K03U04 - Jahr pro Arbeitsblatt.xlsx K03U04 - Lösung 01.xlsx K03U04 - Lösung 02.xlsx K03U04 - Lösung 01.pbix K03U04 - Lösung 02.pbix
Kapitel 4: Kombinieren von nicht übereinstimmenden Tabellen	K04U01 - Fahrräder.xlsx K04U01 - Zubehör.xlsx K04U02 - Produkte.zip K04U02 - Lösung.xlsx K04U02 - Lösung.pbix K04U03 - Produkte.zip K04U03 - Lösung.xlsx K04U03 - Lösung.pbix K04U04 - Produkte.zip K04U04 - Konvertierungstabelle.xlsx K04U04 - Lösung - Vertauschen.xlsx K04U04 - Lösung - Vertauschen.pbix K04U05 - Lösung - Entpivotieren.xlsx K04U05 - Lösung - Entpivotieren.pbix K04U06 - Lösung - Überschriften transponieren.xlsx K04U06 - Lösung - Überschriften transponieren.pbix K04U07 - Lösung - M.xlsx K04U07 - Lösung - M.pbix
Kapitel 5: Den Kontext erhalten	K05U01 - Fahrräder.xlsx K05U01 - Fahrräder & Zubehör.xlsx K05U01 - Zubehör.xlsx K05U01 - Lösung.xlsx K05U01 - Lösung 2.xlsx K05U01 - Lösung.pbix K05U01 - Lösung 2.pbix K05U02 - Fahrräder.xlsx K05U02 - Lösung.xlsx K05U02 - Lösung.pbix K05U03 - Produkte.zip K05U03 - Lösung.xlsx K05U03 - Lösung.pbix K05U04 - Produkte.xlsx K05U04 - Lösung.xlsx K05U04 - Lösung.pbix K05U05 - Produkte.xlsx K05U05 - Lösung.xlsx K05U05 - Lösung.pbix K05U06 - Produkte.xlsx K05U06 - Schnellstart.xlsx K05U06 - Schnellstart.pbix K05U06 - Lösung.xlsx K05U06 - Lösung.pbix

Kapitel	Dateien
Kapitel 6: Tabellen entpivotieren	K06U01.xlsx K06U02.xlsx K06U03.xlsx K06U03 - Falsche Lösung.pbix K06U03 - Lösung.xlsx K06U03 - Lösung.pbix K06U04.xlsx K06U04 - Lösung.xlsx K06U04 - Lösung.pbix K06U05.xlsx K06U05 - Lösung.xlsx K06U05 - Lösung.pbix K06U06.xlsx K06U06 - Lösung.xlsx K06U06 - Lösung.pbix
Kapitel 7: Fortgeschrittenes Entpivotieren und Pivotieren von Tabellen	K07U01.xlsx K07U01 - Lösung.xlsx K07U01 - Lösung.pbix K07U02.xlsx K07U02.pbix K07U03 - Lösung.xlsx K07U03 - Lösung.pbix K07U04.xlsx K07U04 - Lösung.xlsx K07U04 - Lösung.pbix K07U05 - Lösung.xlsx K07U05 - Lösung.pbix
Kapitel 8: Herausforderungen bei der Zusammenarbeit lösen	K08U01.xlsx K08U01 - Alice.xlsx K08U01 - Alice.pbix K08U01 - Lösung.xlsx K08U01 - Lösung.pbix K08U02 - Lösung.pbit K08U02 - Lösung 2.pbit K08U03 - Lösung.xlsx K08U03 - Lösung 2.xlsx K08U04 - Lösung.xlsx K08U04 - Lösung.pbix K08U05.xlsx K08U05.pbix K08U05 - Ordner.zip K08U05 - Lösung.xlsx K08U05 - Lösung.pbix
Kapitel 9: Einführung in die Power Query-Formelsprache M	K09U01 – Lösung.xlsx K09U01 – Lösung.pbix



Kapitel	Dateien
Kapitel 10: Von Stolperfallen zu robusten Abfragen	K10U01.xlsx K10U01 - Lösung.xlsx K10U02 - Lösung.xlsx K10U02 - Lösung.pbix K10U03 - Lösung.xlsx K10U03 - Lösung.pbix K10U04 - Lösung.xlsx K10U04 - Lösung.pbix K10U05.xlsx K10U05 - Lösung.xlsx K10U05 - Lösung.pbix K10U06.xlsx K10U06 - Lösung.xlsx K10U06 - Lösung.pbix K10U06-v2.xlsx
Kapitel 11: Grundlegende Textanalyse	Schlüsselwörter.txt Stoppwörter.txt K11U01.xlsx K11U01 - Lösung.xlsx K11U01 - Lösung.pbix K11U02 - Lösung.xlsx K11U02 - Aktualisierung Vergleich.xlsx K11U02 - Lösung.pbix K11U03 - Lösung.xlsx K11U04 - Lösung.xlsx K11U04 - Lösung.pbix K11U05 - Lösung.xlsx K11U05 - Lösung.pbix K11U06 - Lösung.xlsx K11U06 - Lösung.pbix K11U07 - Lösung.pbix
Kapitel 12: Fortgeschrittene Textanalyse: Bedeutung extrahieren	K12U01 - Lösung.xlsx K12U01 - Lösung.pbix K12U02.xlsx K12U02 - Lösung.xlsx K12U02 - Lösung.pbix K12U02 - Lösung.pbit K12U03 - Lösung.xlsx K12U03 - Lösung.pbix K12U04.xlsx K12U04.pbix K12U04 - Lösung.xlsx K12U04 - Lösung.pbix K12U05 - Lösung.pbix K12U06 - Lösung.xlsx K12U06 - Lösung.pbix

Kapitel	Dateien
Kapitel 13: Das finale Projekt: Alles miteinander kombinieren	K13U01 - Ziel.xlsx K13U01.zip K13U01 - Lösung.xlsx K13U01 - Lösung.pbix K13U02 - Kompromittiert.xlsx K13U02 - Lösung.xlsx K13U02 - Lösung.pbix

Systemvoraussetzungen

Sie benötigen die folgende Software, um die Codebeispiele für dieses Buch zu erstellen und auszuführen:

- Betriebssystem: Windows 10, Windows 8, Windows 7, Windows Server 2008 R2 oder Windows Server 2012
- Software: Office 365, Excel 2016 oder eine neue Version von Excel, Power BI Desktop, Excel 2013 mit Power Query Add-In oder Excel 2010 mit Power Query Add-In

Hinweiskästen



Hinweis



Tip



Zusätzliche Informationen

Support und Feedback

Die folgenden Abschnitte enthalten Informationen zu Errata, Buchsupport, Feedback und Kontaktinformationen.

Errata, Updates und Support

Wir haben uns sehr um die Richtigkeit der in diesem Buch enthaltenen Informationen bemüht. Fehler, die seit der Veröffentlichung der englischen Originalausgabe des Buches bekannt geworden sind, werden auf der Microsoft Press-Website aufgelistet:

<https://www.microsoftpressstore.com/powerquery>

Sollten Sie einen Fehler finden, der noch nicht aufgeführt ist, würden wir uns freuen, wenn Sie uns auf dieser Seite darüber informieren (in englischer Sprache).

Falls Sie zusätzlichen Support benötigen, können Sie sich an den englischsprachigen Buchsupport von Microsoft Press wenden. Sie erreichen ihn unter dieser E-Mail-Adresse:

mspinput@microsoft.com

Mit Anmerkungen, Fragen oder Verbesserungsvorschlägen zu diesem Buch können Sie sich aber auch an den dpunkt.verlag wenden:

hallo@dpunkt.de

Bitte beachten Sie, dass über diese E-Mail-Adressen kein Software- oder Hardware-Support angeboten wird. Für Supportinformationen bezüglich der Software- und Hardwareprodukte von Microsoft besuchen Sie bitte die Microsoft-Website:

<http://support.microsoft.com>

Wir wollen von Ihnen hören

Bei Microsoft Press steht Ihre Zufriedenheit an oberster Stelle. Daher ist Ihr Feedback für uns sehr wichtig, Lassen Sie uns auf dieser englischsprachigen Website wissen, wie Sie dieses Buch finden:

<https://aka.ms/tellpress>

Wir wissen, dass Sie viel zu tun haben. Darum finden Sie auf der Webseite nur wenige Fragen. Ihre Antworten gehen direkt an das Team von Microsoft Press. (Es werden keine persönlichen Informationen abgefragt.) Im Voraus vielen Dank für Ihre Unterstützung.

Über Ihr Feedback per E-Mail freut sich außerdem der dpunkt.verlag über *hallo@dpunkt.de*.

Bleiben Sie am Ball

Falls Sie News, Updates und so weiter zu Microsoft Press-Büchern erhalten möchten, wir sind auf Twitter:

https://twitter.com/dpunkt_verlag

Einführung in Power Query

Wir leben in einem Zeitalter mit permanenten Updates und ständig verfügbaren Technologien. Daher hört es sich vielleicht etwas merkwürdig an, auf Aktualisieren klicken zu müssen. Und dennoch, wenn es richtig betrieben wird und sich Menschen und Kulturen neu erschaffen und beleben, kann das der Auftakt einer Renaissance darstellen.

– Satya Nadella

In diesem Kapitel

- Erhalten Sie eine Einführung in Power Query und erfahren Sie, wie alles begann
- Lernen Sie die Hauptkomponenten von Power Query und dem Power Query-Editor kennen
- Erkunden Sie das Tool und bereiten Sie Beispieldaten für die Analyse vor

In diesem Buch lernen Sie, wie Sie die Features von Power Query nutzen, um Ihre Herausforderungen bei der Datenanalyse zu bewältigen, und wie Sie dabei bis zu 80% der für die Datenvorbereitung benötigten Zeit einsparen. Dieses Kapitel beginnt mit einer offiziellen Einführung. Power Query verdient es. Wissen Sie, während Sie diese Zeilen lesen, gibt es wahrscheinlich eine halbe Million Anwender, die gerade jetzt, genau in diesem Moment, die Zähne zusammenbeißen, während sie in Excel sich wiederholende, aber wichtige Datenvorbereitungsaufgaben durchführen. Sie tun es jeden Tag, jede Woche oder jeden Monat.

Wenn Sie mit dem Lesen dieses Buches fertig sind, werden etwa 50 Millionen Menschen manuell Datenvorbereitungsaufgaben durchgeführt haben, ohne zu wissen, dass ein Tool, das sich in Excel versteckt, nur darauf wartet, ihnen dabei zu helfen, das Leben einfacher zu machen. Einige von ihnen haben bereits gelernt, wie man erweiterte Tools wie Python und R verwendet, um ihre Daten zu säubern; andere verlassen sich auf ihre IT-Abteilungen und warten monatelang darauf, dass ihre Anfragen bearbeitet werden; die meisten von ihnen wollen nur die Arbeit erledigen und haben sich damit abgefunden, Hunderte oder sogar Tausende von Stunden damit zu verbringen, ihre Daten für die Analyse vorzubereiten. Wenn Sie oder Ihre Freunde unter diesen 50 Millionen sind, ist es Zeit, mehr über Power Query und darüber zu erfahren, wie Sie damit Ihre Datenanalysen für immer verändern.

Gleichgültig, ob Power Query neu für Sie ist oder ob Sie ein versierter Praktiker sind, dieses Kapitel hilft Ihnen bei der Vorbereitung auf die vor Ihnen liegende Reise. Im Laufe der Reise werden Sie zu einem Meister beim Data-Wrangling, bei der Datenaufbereitung und in der Lage sein, selbstständig neue Erkenntnisse aus Ihren Daten zu gewinnen.

Was ist Power Query?

Power Query ist eine bahnbrechende Technologie für das Verbinden zu unterschiedlichsten Datenquellen und für die Transformation der Daten, die in Microsoft Excel, Power BI und anderen Microsoft-Produkten zur Verfügung steht. Es ermöglicht jeder Person eine Verbindung zu einem umfangreichen Satz externer Datenquellen und sogar lokaler Daten in einer Kalkulationstabelle herzustellen und die Daten mithilfe einer einfachen Benutzeroberfläche zu sammeln, zu kombinieren und zu transformieren. Sobald die Daten gut vorbereitet sind, können sie in Excel und Power BI in einen Bericht geladen oder als Tabelle in anderen Produkten gespeichert werden. Wenn die Daten aktualisiert werden, können Benutzer ihre Berichte aktualisieren und sich an der automatisierten Transformation ihrer Daten erfreuen.



Power Query wird seit seiner Veröffentlichung von Millionen Benutzern verwendet. Aufgrund seiner erheblichen Auswirkungen auf die Möglichkeiten für Information Worker und Datenanalysten hat Microsoft beschlossen, Power Query in weitere Produkte zu integrieren. Hierzu gehören unter anderem die folgenden:

- **Microsoft SQL Server Data Tools (SSDT) für SQL Server 2017 Analysis Services und Azure Analysis Services (siehe <https://docs.microsoft.com/de-de/sql/analysis-services/what-s-new-in-sql-server-analysis-services-2017>)**
- **Common Data Service for Apps, der von Microsoft PowerApps verwendet wird (siehe <https://docs.microsoft.com/en-us/powerapps/maker/common-data-service/data-platform-cds-newentity-pq>)**

Beim Vorbereiten der Daten mithilfe des Power Query-Editors, der zentralen Benutzeroberfläche Power Query, zeichnet das Tool eine Sequenz von Transformationsanweisungen auf, so dass Sie die Datenvorbereitungssequenz in Zukunft wiederholen können. Fortgeschrittene Excel-Benutzer vergleichen dies oft mit der Makroaufzeichnung, aber es ist einfacher als das. Wenn Sie die Transformationsschritte abgeschlossen haben, können Sie die vorbereiteten Daten in einen Bericht laden. Ab diesem Zeitpunkt werden jedes Mal, wenn Sie den Bericht aktualisieren, die gleichen Transformationsschritte ausgeführt, sodass Sie viele Stunden und Tage wiederholter manueller Datenvorbereitungsarbeiten sparen.

Power Query ist wirklich einfach zu bedienen. Es stellt eine einheitliche Benutzeroberfläche zur Verfügung – unabhängig davon, aus welcher Datenquelle Sie die Daten importieren oder über welches Format die Daten verfügen. Power Query ermöglicht es Ihnen, komplexe Datenvorbereitungsszenarien mittels einer Reihe kleiner Schritte zu erstellen, die editierbar und leicht zu verstehen sind. Bei komplexeren Einsatzszenarien können fortgeschrittene Anwender jeden Schritt in der Bearbeitungsleiste oder im erweiterten Editor ändern, um die Transformationsausdrücke anzupassen (mit der M-Abfragesprache, die in Kapitel 9, »Einführung in die Power Query-Formelsprache M«, erläutert wird). Jede Abfolge von Transformationsschritten wird als Abfrage gespeichert, die in einen Bericht geladen oder von anderen Abfragen wiederverwendet werden kann, um so eine Pipeline von Transformationsbausteinen zu erstellen.

Bevor wir die einzelnen Hauptkomponenten von Power Query untersuchen, gehen wir ein paar Jahre zurück und schauen uns an, wie alles begann. Dieser kurze geschichtliche Überblick zu Power Query wird Ihnen helfen, zu verstehen, wie lange es diese Technologie bereits gibt und wie sie sich zum heutigen Status weiterentwickelt hat.

Ein kurzer geschichtlicher Überblick zu Power Query

Power Query wurde ursprünglich 2011 als Teil von Microsoft SQL Azure Labs entwickelt. Es wurde auf dem PASS Summit im Oktober 2011 unter dem Microsoft-Codennamen »Data Explorer« angekündigt. Abbildung 1-1 zeigt die ursprüngliche Benutzeroberfläche.

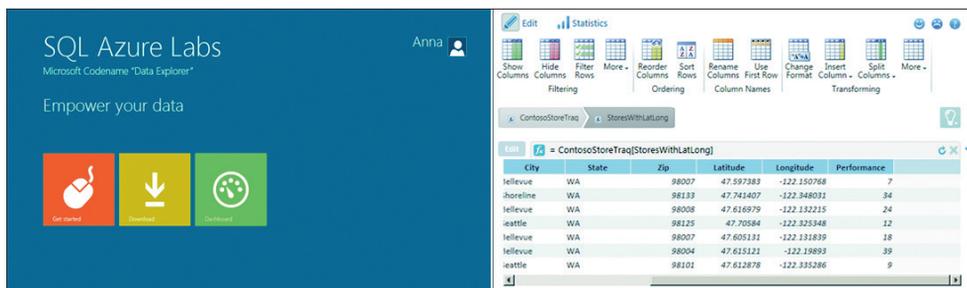


Abbildung 1-1 Hinter dem Produkt mit dem Microsoft-Codennamen *Data Explorer* verbirgt sich eine frühe Version von Power Query.

Im Februar 2013 hat Microsoft das Tool als Excel-Add-In neu gestaltet und es von SQL Azure Labs abgekoppelt. Das Tool, das jetzt den Namen *Data Explorer Preview for Excel* trägt, wurde positioniert, um die Self-Service-BI-Erfahrung in Excel zu verbessern, indem die Untersuchung von Daten und der Zugriff auf eine breite Palette von Datenquellen für umfassendere Erkenntnisse vereinfacht wird.

Bereits von Beginn an stellte Data Explorer als Excel-Add-In eine intuitive und konsistente Benutzeroberfläche für die Erforschung, Kombination und Verfeinerung von Daten zur Verfügung, und zwar für Daten aus einer Vielzahl von Quellen, einschließlich relationaler, strukturierter und halbstrukturierter Datenbanken, OData, Web, Hadoop, Azure Market Place und mehr. Mit Data Explorer war es auch möglich, nach öffentlichen Daten auf Quellen wie Wikipedia zu suchen (eine Funktionalität, die später entfernt wurde).

Nach der Installation der Add-Ins in Excel 2010 oder 2013 war Data Explorer Preview for Excel auf der Registerkarte *Data Explorer* sichtbar. Diese Registerkarte in Excel hatte das gleiche Look & Feel wie das heutige Power Query-Add-In. Der Power Query-Editor wurde zu diesem Zeitpunkt als *Neue Abfrage* bezeichnet, und es gab keine spezifischen Power Query-Registerkarten. Wenn Sie sich die Ankündigung von Data Explorer anschauen und sich über die ursprüngliche Benutzeroberfläche als Excel-Add-In informieren wollen, schauen Sie sich dieses Video an: <https://blogs.msdn.microsoft.com/dataexplorer/2013/02/27/announcing-microsoft-data-explorer-preview-for-excel>.

Abbildung 1-2 zeigt Statistiken über die zunehmende Adaption von Data Explorer und seinen Übergang von SQL Azure Labs zu Excel. Laut MSDN-Profil des Data Explorer Teams bei Microsoft (<https://social.msdn.microsoft.com/Profile/Data%2bExplorer%2bTeam>) startete das

Team seine erste Community-Aktivitäten im Oktober 2011, als Data Explorer zum ersten Mal in SQL Azure Labs veröffentlicht wurde. Seit Februar 2013, als Data Explorer als Excel-Add-In veröffentlicht wurde, hat das Community-Engagement deutlich zugenommen; der Umstieg auf Excel hatte sich eindeutig gelohnt.

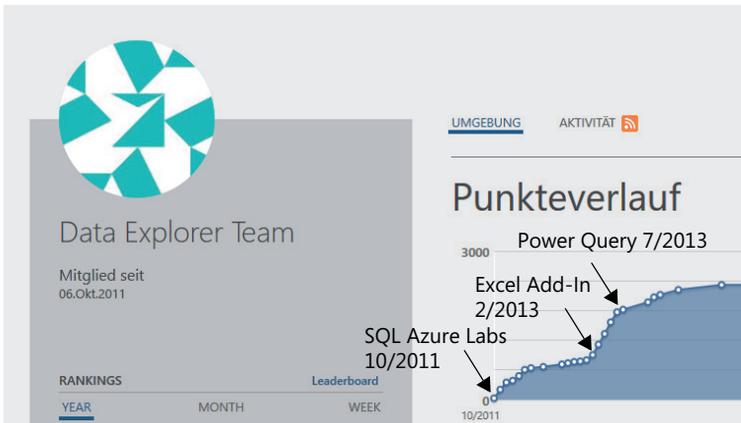


Abbildung 1-2 Der Punkteverlauf des Daten-Explorer-Teams auf MSDN zeigt die zunehmende Adaption von Data Explorer, nachdem das Team von SQL Azure Labs auf Excel umgeschwenkt ist.

Wie Sie an der Trendlinie im Punkteverlauf in Abbildung 1-2 sehen können, begann die Aktivität des Data Explorer Teams im Juli 2013 an Schwung zu verlieren. Es war jedoch kein negativer Moment in der Geschichte des Data Explorers – nur eine Wiedergeburt des Tools unter einem neuen Namen. Im Juli 2013 gab Microsoft die allgemeine Verfügbarkeit des Add-Ins unter dem neuen Namen Power Query-Add-In für Excel bekannt. Zu diesem Zeitpunkt bot das Add-In die gleiche Benutzererfahrung wie die neueste Version von Power Query.

Das Power Query-Team begann, monatlich Updates für das Power Query-Add-In zu veröffentlichen. Diese Entwicklungsgeschwindigkeit führte zu rascher Innovation und ständigem Wachstum der Community. Viele Benutzer und Fans halfen, das Produkt durch direktes Feedback, Foren und Blogs zu gestalten.

Das Power Query-Add-In wird immer noch ständig aktualisiert und steht als Add-In für Excel 2010 und Excel 2013 zum Download zur Verfügung. Nach der Installation wird Power Query als neue Registerkarte in Excel angezeigt, und Sie können von dieser Registerkarte aus eine Verbindung zu neuen Datenquellen herstellen.

Im Dezember 2014 veröffentlichte Microsoft eine Vorschau von Power BI Designer (<https://powerbi.microsoft.com/en-us/blog/new-power-bi-features-available-for-preview>). Der Power BI Designer war ein neues Clienttool für die Erstellung von Berichten, das es Business Intelligence-Experten ermöglichte, interaktive Berichte zu erstellen und sie im Power BI-Dienst zu veröffentlichen, der sich ebenfalls noch in der Previewphase befand. Power BI Designer vereinheitlichte drei Excel-Add-Ins (Power Query, Power Pivot und Power View) und war für den Erfolg von Power BI entscheidend. Im Power BI Designer behielt Power Query alle Funktionen des Excel-Add-Ins bei. Während die meisten Benutzererfahrungen identisch waren, wurde der Begriff Power Query in Power BI Designer nicht mehr verwendet. Sieben Monate später, im Juli 2015,

änderte Microsoft den Namen von Power BI Designer in Power BI Desktop und gab dessen allgemeine Verfügbarkeit bekannt (<https://powerbi.microsoft.com/en-us/blog/what-s-new-in-the-power-bi-desktop-ga-update>).

In dieser Phase lieferte das Power Query-Team weiterhin monatliche Updates von Power Query für Excel und Power BI Desktop aus, während es mit dem Excel-Team zusammenarbeitete, um die standardmäßige Benutzererfahrung für das Abrufen von Daten in Excel vollständig zu überarbeiten.

Während das Power Query-Add-In ursprünglich von Excel getrennt war, entschied sich Microsoft, es als systemeigene Komponente zu integrieren und das Power Query-Modul in Excel als primären Data Stack zu verwenden. Im September 2015 veröffentlichte Microsoft Excel 2016 mit der integrierter Power Query-Funktionalität; die Installation eines Add-Ins war nicht mehr erforderlich. Microsoft platzierte die Power Query-Funktionalität zunächst auf die Registerkarte *Daten* in die Befehlsgruppe *Abrufen und transformieren*, die später in *Daten abrufen und transformieren* umbenannt wurde.

Damit war die Power Query-Technologie zum ersten Mal für die massenhafte Verwendung verfügbar und unterstützte native Excel-Funktionen wie Rückgängig und Wiederholen, Kopieren und Einfügen von Tabellen, Makroaufzeichnung und VBA. Weitere Informationen zur Power Query-Integration in Excel 2016 finden Sie unter <https://blogs.office.com/en-us/2015/09/10/integrating-power-query-technology-in-excel-2016>.

Im März 2017 veröffentlichte Microsoft ein Update für Office 365, das weitere Verbesserungen am Data Stack enthielt. Die Power Query-Technologie wurde damit tatsächlich zum primären Data Stack von Excel (<https://support.office.com/en-us/article/unified-get-transform-experience-ad78befd-eb1c-4ea7-a55d-79d1d67cf9b3>). Das Update umfasste eine Vereinheitlichung der Erfahrungen zwischen Abfragen und Arbeitsmappenverbindungen und verbesserte die Unterstützung für ODC-Dateien. Darüber hinaus wurde der Haupteinstiegspunkt zu Power Query, das Dropdownmenü *Daten abrufen*, als erster Befehl auf die Registerkarte *Daten* in den Abschnitt *Daten abrufen und transformieren* verschoben.

Im April 2017 veröffentlichte Microsoft SQL Server Data Tools (SSDT) und kündigte an, dass die moderne Oberfläche für den Abruf von Daten auch für tabellarische Modelle mit dem Kompatibilitätsgrad 1400 der Analysis Services zur Verfügung steht (<https://blogs.msdn.microsoft.com/ssdt/2017/04/19/announcing-the-general-availability-ga-release-of-ssdt-17-0-april-2017>). Mit SSDT 17.0 können Sie Power Query verwenden, um Daten in Ihren Tabellenmodellen in SQL Server 2017 Analysis Services und Azure Analysis Services zu importieren und vorzubereiten. Wenn Sie mit Analysis Services vertraut sind, können Sie hier erfahren, wie Sie Power Query hierfür verwenden:

<https://docs.microsoft.com/en-us/sql/analysis-services/Tutorial-tabular-1400/as-lesson-2-get-data?view=sql-analysis-services-2017>.



Obwohl sich dieses Buch auf Excel und Power BI Desktop konzentriert, sind die meisten Kapitel und Übungen des Buches auch für die Arbeit mit Analysis Services relevant, insbesondere in frühen Phasen Ihrer Projekte, in denen Sie mit chaotischen Datasets umgehen müssen.

Im März 2018 kündigte Microsoft Common Data Service (CDS) for Apps (<https://powerapps.microsoft.com/en-us/blog/cds-for-apps-march>) an und hat Power Query zusammen mit Microsoft Flow als eines der wichtigsten Datenimporttools integriert (siehe Abbildung 1-3). Microsoft hat Power Query über seinen ursprünglichen Zweck hinaus erweitert, um BI-Szenarien zu adressieren, sodass Power Query jetzt als einfaches ETL-Tool (Extract Transform Load) verwendet werden kann, mit dem Business-Anwender Geschäftsanwendungen für Microsoft Office 365 und Dynamics 365 entwickeln können, ohne dass hierfür Programmierkenntnisse erforderlich sind.

Ebenfalls im März 2018 hat Microsoft den Begriff Power Query in Power BI Desktop und Excel wieder eingeführt, indem der Titel des Fensters *Query-Editor* in *Power Query-Editor* geändert wurde. Um es zu starten, können Sie jetzt im Dropdownmenü *Daten abrufen* den Befehl *Power Query-Editor starten* auswählen. Im Juli 2018 gab Microsoft bekannt, dass die Onlineversion von Power Query Teil einer neuen Self-Service-ETL-Lösung, Dataflows, sein wird, mit der Sie Datenvorbereitungen in Power Query problemlos durchführen, die Ergebnisse in Azure speichern und in Power BI oder anderen Anwendungen (<https://www.microsoft.com/en-us/businessapplicationssummit/video/BAS2018-2117>) nutzen können.

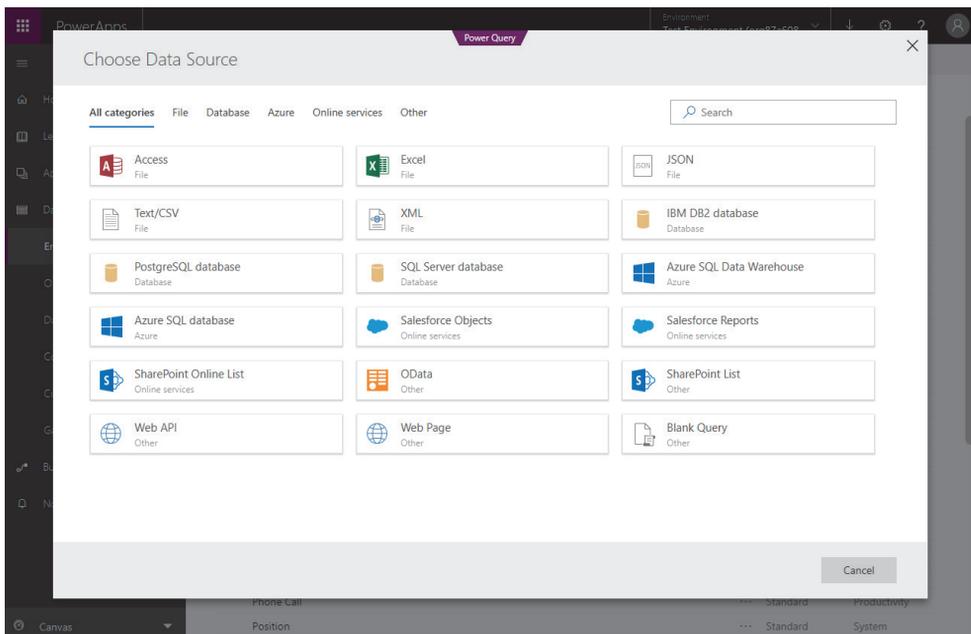


Abbildung 1-3 Power Query in CDS für Apps, die im März 2018 angekündigt wurde

Wo finde ich Power Query?

Power Query in Excel und Power BI Desktop zu finden ist eine Herausforderung, wenn Sie nicht wissen, wonach Sie suchen müssen. Während diese Zeilen geschrieben werden, gibt es keinen zentralen Einstiegspunkt mit dem Namen »Power Query«, um den Power Query-Editor zu starten. Abbildung 1-4 fasst die wichtigsten Einstiegspunkte für Power Query in Excel und Power BI Desktop zusammen.