

Alberto Ferrari · Marco Russo

Datenanalyse mit Microsoft Power BI und Power Pivot für Excel





Zu diesem Buch – sowie zu vielen weiteren dpunkt.büchern – können Sie auch das entsprechende E-Book im PDF-Format herunterladen. Werden Sie dazu einfach Mitglied bei dpunkt.plus⁺:

www.dpunkt.plus

Datenanalyse mit Microsoft Power BI und Power Pivot für Excel

Alberto Ferrari
Marco Russo

Alberto Ferrari
Marco Russo

Lektorat: Sandra Bollenbacher

Übersetzung & Satz: G&U Language & Publishing Services GmbH, www.gundu.com

Copy-Editing: Petra Heubach-Erdmann, Düsseldorf

Herstellung: Susanne Bröckelmann

Umschlaggestaltung: Helmut Kraus, www.exclam.de

Druck und Bindung: M.P. Media-Print Informationstechnologie GmbH, 33100 Paderborn

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

ISBN:

Print 978-3-86490-510-0

PDF 978-3-96088-250-3

ePub 978-3-96088-251-0

mobi 978-3-96088-252-7

Translation Copyright für die deutschsprachige Ausgabe © 2018 dpunkt.verlag GmbH

Wieblinger Weg 17

69123 Heidelberg

Authorized translation from the English language edition, entitled ANALYZING DATA WITH POWER BI AND POWER PIVOT FOR EXCEL, 1st Edition by ALBERTO FERRARI; MARCO RUSSO, published by Pearson Education, Inc, publishing as Microsoft Press, Copyright © 2017 by Alberto Ferrari and Marco Russo.

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.

German language edition published by dpunkt.verlag GmbH, Copyright © 2018

ISBN of the English language edition: 978-1-5093-0276-5

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 können jedoch für Schäden haftbar gemacht werden, die in Zusammenhang mit der Verwendung dieses Buchs stehen.

5 4 3 2 1 0



Inhaltsverzeichnis

Einleitung	xi
Zielgruppe	xii
Voraussetzungen	xii
Aufbau dieses Buches	xiii
Begleitende Inhalte	xiv
Schreibweisen und Hinweiskästen	xiv
Danksagungen	xv
Errata und Support	xv
 Kapitel 1	
<hr/>	
Einführung in die Datenmodellierung	1
Arbeiten mit einer einzelnen Tabelle	2
Datenmodelle	8
Sternschemata	15
Die Wichtigkeit von Namen	21
Zusammenfassung	23
 Kapitel 2	
<hr/>	
Header/Detail-Tabellen	25
Einführung	25
Werte aus dem Header aggregieren	26
Header/Detail-Tabellen reduzieren	33
Zusammenfassung	35
 Kapitel 3	
<hr/>	
Mehrere Faktentabellen	37
Denormalisierte Faktentabellen	37
Dimensionsübergreifende Filterung	43
Mehrdeutigkeit von Modellen	45

Bestellungen und Rechnungen	48
Gesamtrechnungsbetrag für einen Kunden	53
Gesamtbetrag der Rechnungen für die ausgewählte Bestellung bzw. den ausgewählten Kunden	53
Betrag der in Rechnung gestellten Bestellungen	54
Zusammenfassung	57

Kapitel 4

Datum und Uhrzeit	59
Eine Datumsdimension erstellen	59
Automatische Zeitdimensionen	62
Automatische Zeitgruppierung in Excel	63
Automatische Zeitgruppierung in Power BI Desktop	64
Mehrere Datumsdimensionen	65
Umgang mit Datum und Uhrzeit	71
Zeitinformationsberechnungen	73
Geschäftskalender	75
Berechnungen mit Arbeitstagen	77
Arbeitstage in einer einzigen Region	77
Arbeitstage in mehreren Regionen	80
Besondere Zeiträume im Jahr	84
Sich nicht überlappende Zeiträume	85
Zeiträume relativ zu heute	86
Sich überlappende Zeiträume	89
Wochenkalender	90
Zusammenfassung	96

Kapitel 5

Historische Attribute	99
Einführung in langsam veränderliche Dimensionen	99
Langsam veränderliche Dimensionen verwenden	104
Langsam veränderliche Dimensionen laden	107
Die Granularität der Dimension korrigieren	110
Die Granularität der Faktentabelle korrigieren	113

Schnell veränderliche Dimensionen	114
Die richtige Modellierungstechnik wählen	117
Zusammenfassung	119

Kapitel 6

Snapshots	121
Einführung	121
Snapshots aggregieren	122
Abgeleitete Snapshots	128
Übergangsmatrizen	130
Zusammenfassung	136

Kapitel 7

Datums- und Zeitintervalle	139
Einführung in Zeitdaten	139
Aggregationen mit einfachen Intervallen	141
Datumsübergreifende Intervalle	143
Schichten und Zeitversatz modellieren	148
Laufende Ereignisse analysieren	149
Unterschiedliche Dauern vermischen	159
Zusammenfassung	164

Kapitel 8

m:n-Beziehungen	167
Einführung	167
Das bidirektionale Muster	168
Nicht additive Berechnungen	171
Kaskadierende m:n-Beziehungen	172
Zeitliche m:n-Beziehungen	175
Prozentuale Zuordnungsfaktoren	178
m:n-Beziehungen materialisieren	181
Die Faktentabelle als Brücke verwenden	182
Überlegungen zur Leistung	183
Zusammenfassung	185

Kapitel 9

Unterschiedliche Granularitäten 187

Einführung in Granularität	187
Beziehungen zwischen Tabellen unterschiedlicher Granularität	189
Vorhersagedaten analysieren	189
DAX-Code zum Verschieben von Filtern	192
Filterung über Beziehungen	194
Werte mit falscher Granularität ausblenden	196
Werte einer feineren Granularität mithilfe von Zuordnungsfaktoren berechnen	200
Zusammenfassung	201

Kapitel 10

Segmentierungsmodelle 203

Mehrspaltige Beziehungen	203
Statische Segmentierung	206
Dynamische Segmentierung	208
ABC-Analyse	210
Zusammenfassung	215

Kapitel 11

Währungsumrechnung 217

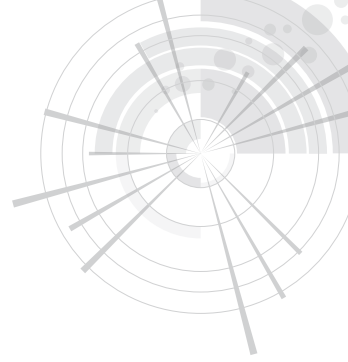
Die möglichen Situationen	217
Mehrere Quellwährungen, eine Berichtswährung	218
Eine Quellwährung, mehrere Berichtswährungen	222
Mehrere Quellwährungen, mehrere Berichtswährungen	227
Zusammenfassung	229

Anhang

Grundlagen der Datenmodellierung 231

Tabellen	231
Datentypen	233
Beziehungen	233

Filter und Kreuzfilter	234
Unterschiedliche Arten von Modellen	237
Sternschema	238
Schneeflockenschema	238
Modelle mit Brückentabellen	240
Additivität berechneter Felder	241
Additive berechnete Felder	241
Nicht additive berechnete Felder	241
Halbadditive berechnete Felder	242
 Index	 243



Einleitung

Excel-Benutzer lieben Zahlen. Vielleicht ist es auch eher so, dass Zahlenliebhaber Excel lieben. Wie dem auch immer sein mag – sobald Sie daran interessiert sind, Erkenntnisse aus irgendeiner Art von Datenmenge zu gewinnen, ist es sehr wahrscheinlich, dass Sie schon viel Zeit mit Excel, Pivottabellen und Formeln verbracht haben.

Im Jahr 2015 wurde Power BI veröffentlicht. Heutzutage kann man mit Fug und Recht sagen, dass Zahlenliebhaber sowohl Power Pivot für Excel als auch Power BI lieben. Beide Tools haben viele Merkmale gemeinsam, nämlich die Datenbank-Engine VertiPaq und die von SQL Server Analysis Services geerbte Sprache DAX.

Um in früheren Excel-Versionen Erkenntnisse aus Zahlen zu gewinnen, musste man zunächst einige Datasets laden und dann Spalten berechnen und Formeln schreiben, um Diagramme zu entwerfen. Dabei gab es einige Einschränkungen: Die Größe der Arbeitsmappe war von Bedeutung, und die Formelsprache von Excel war für große Zahlen nicht gerade ideal. Die neue Engine von Power BI und Power Pivot stellt einen Riesenschritt nach vorn dar. Jetzt können Sie sämtliche Möglichkeiten einer Datenbank und eine hervorragende Sprache (nämlich DAX) nutzen. Doch diese neuen Möglichkeiten bringen auch mehr Verantwortung mit sich. Wenn Sie dieses neue Instrument wirklich nutzen wollen, müssen Sie mehr lernen, nämlich die Grundlagen der Datenmodellierung.

Datenmodellierung ist keine höhere Mathematik, sondern eine Grundfertigkeit, die jeder beherrschen sollte, der sich dafür interessiert, Erkenntnisse aus Daten zu gewinnen. Wenn Sie Zahlen mögen, dann werden Sie auch Datenmodellierung lieben. Es ist also nicht nur eine leicht zu erwerbende Fertigkeit, sondern macht auch unglaublich viel Spaß.

In diesem Buch lernen Sie die grundlegenden Konzepte der Datenmodellierung anhand von praktischen Beispielen kennen, denen Sie sehr wahrscheinlich in Ihrem täglichen Arbeitsleben begegnen. Wir wollten kein kompliziertes Buch über Datenmodellierung schreiben, in dem wir die vielen komplexen Entscheidungen, die Sie treffen müssen, um eine komplexe Lösung zu erstellen, im Detail erläutern. Stattdessen konzentrierten wir uns auf Beispiele aus unserer täglichen Beratertätigkeit. Wann immer ein Kunde uns bat, bei der Lösung eines Problems zu helfen, und wir der Meinung waren, dass dieses Problem häufig auftritt, dann hoben wir es auf. Anschließend haben wir dieses Problemarchiv geöffnet und für jedes dieser Beispiele eine Lösung bereitgestellt, die wir anschließend so geordnet haben, dass sie auch als Schulung zur Datenmodellierung dienen können.

Nachdem Sie die Lektüre dieses Buches beendet haben, werden Sie noch kein Experte für Datenmodellierung sein, haben aber bereits eine größere Sensibilität für das Thema erworben. Wenn Sie sich anschließend wieder Ihrer eigenen Datenbank zuwenden, um zu überlegen, wie Sie einen erforderlichen Wert berechnen können, und Sie dabei auf die Idee kommen, dass eine Änderung des Modells hilfreich sein könnte, dann haben wir mit diesem Buch unser Ziel erreicht. Vor allem sind Sie dann auf dem Weg, ein erfolgreicher Datenmodellierer zu werden. Diesen letzten Schritt – also ein großartiger Datenmodellierer zu werden –, können Sie nur mit viel Erfahrung und nach vielen Fehlern gehen. Leider ist Erfahrung nichts, was man in einem Buch lernen kann.

Zielgruppe

Dieses Buch hat eine sehr breite Zielgruppe. Sind Sie vielleicht ein Excel-Benutzer, der Power Pivot für Excel einsetzt, oder ein Datenforscher, der Power BI nutzt? Oder stehen Sie noch am Anfang Ihrer Karriere als Business-Intelligence-Profi und möchten eine Einführung in die Themen der Datenmodellierung lesen? In all diesen Fällen ist dies das richtige Buch für Sie.

Beachten Sie, dass wir Personen, die ein Buch über Datenmodellierung lesen möchten, nicht in diese Liste aufgenommen haben. Tatsächlich sind wir beim Schreiben dieses Buches davon ausgegangen, dass unsere Leser gar nicht wissen, dass sie Datenmodellierung brauchen. Unser Ziel ist es, Ihnen deutlich zu machen, dass Sie Datenmodellierung erlernen müssen, und Ihnen dann einige Einblicke in die Grundlagen dieser wunderbaren Wissenschaft zu geben. Kurz gesagt, wenn Sie wissen möchten, was Datenmodellierung ist und warum sie eine nützliche Fähigkeit darstellt, dann ist dies das richtige Buch für Sie.

Voraussetzungen

Wir gehen davon aus, dass unsere Leser über Grundkenntnisse in Excel-Pivottabellen verfügen oder Power BI als Berichts- und Modellierungswerkzeug einsetzen. Auch einige Erfahrungen mit der Analyse von Zahlen sind sehr hilfreich. Die Benutzeroberfläche von Excel und Power BI behandeln wir in diesem Buch nicht. Stattdessen konzentrieren wir uns auf Datenmodelle und darauf, wie man sie erstellt und abwandelt, damit sich der Code einfacher schreiben lässt. Das heißt, wir behandeln das »Was« und überlassen Ihnen das »Wie«. Wir wollten keine schrittweise Anleitung schreiben, sondern ein Buch, das anspruchsvolle Themen auf einfache Weise erklärt.

Ein Thema, das wir bewusst nicht in dem Buch behandeln, ist die Sprache DAX. Es wäre unmöglich gewesen, Datenmodellierung und DAX zusammen im selben Buch zu erklären. Wenn Sie bereits mit der Sprache vertraut sind, dann können Sie viel Nutzen aus den vielen DAX-Codebeispielen in diesem Buch ziehen. Sollten Sie DAX hingegen noch lernen müssen, dann lesen Sie *The Definitive Guide to DAX*, den umfassendsten Leitfaden zur dieser Sprache, der sich auch gut an die Themen dieses Buches anschließt.

Aufbau dieses Buches

Das Buch beginnt mit ein einigen einfachen, einleitenden Kapiteln, gefolgt von einer Reihe monographischer Kapitel, die jeweils ein bestimmtes Datenmodell behandeln. Die folgende Aufstellung gibt eine kurze Beschreibung der Kapitel:

- **Kapitel 1**, »Einführung in die Datenmodellierung«, gibt eine kurze Einführung in die Grundprinzipien der Datenmodellierung. Hier zeigen wir, was Datenmodellierung ist, erklären Granularität und definieren die grundlegenden Modelle für Data Warehouses, nämlich Sternschema, Schneeflockenschema, Normalisierung und Denormalisierung.
- **Kapitel 2**, »Header/Detail-Tabellen«, behandelt ein sehr weit verbreitetes Modell, nämlich Header/Detail-Tabellen. Hier finden Sie Erklärungen und Lösungen für Situationen, in denen Sie z. B. Bestellungen und deren einzelne Zeilen in zwei getrennten Faktentabellen unterbringen.
- **Kapitel 3**, »Mehrere Faktentabellen«, beschreibt Situationen, in denen Sie mehrere Faktentabellen haben und einen Bericht erstellen müssen, der sie kombiniert. Hier betonen wir, wie wichtig es ist, das korrekte dimensionale Modell aufzustellen, um die Daten richtig durchsuchen zu können.
- **Kapitel 4**, »Datum und Uhrzeit«, ist eines der längsten des Buches. Es behandelt Zeitinformationsberechnungen. Wir erklären, wie Sie eine korrekte Datumstabelle erstellen und grundlegende Zeitinformationsberechnungen durchführen (YTD, QTA, PARALLELPERIOD usw.). Außerdem zeigen wir einige Beispiele für die Berechnung von Arbeitstagen, die Handhabung von besonderen Zeiträumen im Jahr und den richtigen Umgang mit Kalenderdaten im Allgemeinen.
- **Kapitel 5**, »Historische Attribute«, beschreibt die Verwendung von langsam veränderlichen Dimensionen. In diesem Kapitel werden die Transformationsschritte genauer erläutert, die Sie in Ihrem Modell durchführen müssen, wenn Sie veränderliche Attribute verfolgen müssen. Außerdem erfahren Sie, wie Sie Ihren korrekten DAX-Code für langsam veränderliche Dimensionen schreiben.
- **Kapitel 6**, »Snapshots«, behandelt die faszinierenden Aspekte von Snapshots. Wir erklären, was Snapshots sind, warum und wann sie verwendet werden sollten und wie Sie aus einem Snapshot Werte berechnen. Außerdem stellen wir das äußerst leistungsfähige Modell der Übergangsmatrix vor.
- **Kapitel 7**, »Datums- und Zeitintervalle«, geht noch mehrere Schritte über den Inhalt von Kapitel 5 hinaus. Wiederum geht es um Zeitberechnungen, aber diesmal analysieren wir Modelle, bei denen die in den Faktentabellen gespeicherten Ereignisse eine bestimmte Dauer haben und daher eine besondere Behandlung benötigen, um korrekte Ergebnisse zu liefern.
- **Kapitel 8**, »m:n-Beziehungen«, erklärt, wie Sie m:n-Beziehungen einsetzen können. Solche Beziehungen spielen in jedem Datenmodell eine sehr wichtige Rolle. Wir behandeln reguläre m:n-Beziehungen, kaskadierende Beziehungen und die Verwendung von Zuordnungsfaktoren und Filtern, besprechen ihre Leistung und erklären, wie Sie sie verbessern können.

- **Kapitel 9**, »Unterschiedliche Granularitäten«, geht eingehender auf die Arbeit mit Faktentabellen mit unterschiedlicher Granularität ein. Wir zeigen ein Beispiel mit einer Datenvorhersage, bei dem die Faktentabellen unterschiedliche Granularität aufweisen, und stellen verschiedene Lösungsmöglichkeiten mit DAX-Code und durch Anpassung des Datenmodells vor.
- **Kapitel 10**, »Segmentierungsmodelle«, erläutert verschiedene Segmentierungsmodelle. Wir beginnen mit einer einfachen Segmentierung nach Preis, gehen dann zur dynamischen Segmentierung mithilfe virtueller Beziehungen über und erläutern schließlich die ABC-Analyse mit DAX.
- **Kapitel 11**, »Währungsumrechnung«, behandelt den Umgang mit unterschiedlichen Währungen. Bei der Arbeit mit Wechselkursen ist es wichtig, die Anforderungen zu verstehen und das richtige Modell zu erstellen. Wir untersuchen mehrere Fälle mit unterschiedlichen Anforderungen und stellen für jeden davon die jeweils beste Lösung vor.
- **Anhang A**, »Grundlagen der Datenmodellierung«, ist als Nachschlagewerk gedacht. Die in diesem Buch verwendeten Grundbegriffe werden anhand von Beispielen kurz beschrieben. Wann immer Sie sich unsicher sind, können Sie dorthin vorblättern, um sich zu informieren, und dann zur eigentlichen Lektüre zurückkehren.

Die Komplexität der Modelle und der Lösungen nimmt Kapitel für Kapitel zu, sodass es besser ist, dieses Buch von Anfang an zu lesen, anstatt von Kapitel zu Kapitel zu springen. So können Sie dem natürlichen Fluss der Komplexität folgen und ein Thema nach dem anderen erlernen. Nach Beendigung der Lektüre können Sie das Buch jedoch auch als Nachschlagewerk nutzen. Wenn Sie also eine Lösung für ein bestimmtes Modell benötigen, können Sie direkt zu dem Kapitel springen, in dem es behandelt wird, und dort die Einzelheiten der Lösung einsehen.

Begleitende Inhalte

Um die Arbeit mit diesem Buch zu vereinfachen, haben wir begleitende Inhalte zusammengestellt, die Sie von der folgenden Seite herunterladen können:

https://www.dpunkt.de/datenanalyse_excel

Zu diesen Inhalten gehören die Excel- und Power BI Desktop-Dateien für alle Beispiele in diesem Buch. Für jede Abbildung im Buch gibt es eine eigene Datei, sodass Sie die einzelnen Schritte genau nachvollziehen und immer auf dem gleichen Stand ausgehen können, um die Beispiele selbst auszuprobieren. Bei den meisten dieser Beispiele handelt es sich um Power BI Desktop-Dateien. Wenn Sie sie auf Ihrem eigenen PC nachvollziehen wollen, sollten Sie sich die neueste Version dieses Programms von der Power BI-Website herunterladen.

Schreibweisen und Hinweiskästen

In diesem Buch werden folgende Schreibweisen und Hinweiskästen verwendet:

- **Fettschrift** wird für Text verwendet, den Sie eingeben.
- *Kursivschrift* wird für neue Begriffe, für Datenbankelemente wie Tabellen, Spalten usw., für Elemente der Benutzeroberfläche sowie für URLs verwendet.

- Code erscheint in nicht proportionaler Schrift.
- Bei Tastaturkürzeln sind die einzelnen zu drückenden Tasten durch ein Pluszeichen verbunden. Beispielsweise bedeutet `Strg` + `Alt` + `Entf`, dass Sie gleichzeitig die Tasten `Strg`, `Alt` und `Entf` drücken müssen.



Hinweis



Tipp



Zusätzliche Informationen



Achtung

Danksagungen

Bevor wir zum Abschluss dieser Einleitung kommen, möchten wir unserer Lektorin Kate Shoup danken, die uns bei der ganzen redaktionellen Arbeit begleitet hat, sowie unserem Fachgutachter Ed Price. Ohne ihre akribische Arbeit wäre dieses Buch viel schwerer zu lesen! Dass dieses Buch weniger Fehler enthält als unser Originalmanuskript, ist diesen beiden Personen zu verdanken. Wenn es immer noch Fehler enthält, so liegt das natürlich ganz an uns.

Errata und Support

Wir haben alle Anstrengungen unternommen, um die Richtigkeit dieses Buches und der begleitenden Inhalte sicherzustellen. Alle Fehler des englischsprachigen Originals, die seit der Veröffentlichung gemeldet wurden, sind auf der Website von Microsoft Press unter der folgenden Adresse aufgeführt:

<https://aka.ms/AnalyzeData/errata>

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

hallo@dpunkt.de

Bitte beachten Sie, dass über unsere E-Mailadresse kein Software-Support angeboten wird. Für Supportinformationen bezüglich der hier verwendeten Microsoft-Produkte besuchen Sie die Microsoft-Website:

<http://support.microsoft.com>

Einführung in die Datenmodellierung

In diesem Buch geht es um Datenmodellierung. Als Erstes stellt sich die Frage, warum Sie sich überhaupt damit beschäftigen sollten. Schließlich können Sie auch einfach Erkenntnisse aus Ihren Daten gewinnen, indem Sie in Excel eine Abfrage laden und eine Pivottabelle daraus erstellen. Wozu brauchen Sie da Datenmodellierung?

Als Berater werden wir täglich von Einzelpersonen oder Unternehmen beauftragt, die Schwierigkeiten damit haben, die erforderlichen Zahlen zu berechnen. Sie haben das Gefühl, dass die Zahl, nach der sie suchen, existiert und berechnet werden kann, aber entweder die Formeln zu kompliziert sind oder die Zahlen nicht stimmen. In 99 % der Fälle liegt das an einem Fehler im Datenmodell. Wenn Sie das Modell korrigieren, lässt sich die Formel leicht aufstellen und verstehen. Wenn Sie Ihre Analysemöglichkeiten verbessern und sich lieber auf die Entscheidungsfindung konzentrieren möchten anstatt darauf, eine komplizierte DAX-Formel auszutüfteln, müssen Sie daher Datenmodellierung lernen.

Datenmodellierung gilt gewöhnlich als schwer zu erlernen. Wir werden Ihnen nicht einreden, dass das nicht so wäre. Datenmodellierung ist ein vielschichtiges Thema. Es ist anspruchsvoll und es erfordert einige Anstrengung, um es zu lernen und um Ihr Gehirn darauf zu trainieren, bei der Betrachtung eines Szenarios das Modell im Geiste vor sich zu sehen. Es stimmt, Datenmodellierung ist kompliziert, anspruchsvoll und erweitert den Geist. Mit anderen Worten, es macht viel Spaß!

In diesem Kapitel finden Sie einige einfache Beispiele von Berichten, bei denen das richtige Datenmodell zu einfacheren Formeln führt. Da es sich um Beispiele handelt, lassen sie sich natürlich nicht vollständig auf Ihr Geschäft übertragen. Dennoch hoffen wir, dass sie Ihnen eine gute Vorstellung davon geben, warum Datenmodellierung eine so wichtige Fähigkeit ist. Ein guter Datenmodellierer zu sein, bedeutet im Grunde genommen, Ihr spezifisches Modell einem der vielen verschiedenen Muster zuzuordnen, die bereits von anderen untersucht und eingerichtet worden sind. Ihr Modell unterscheidet sich gar nicht so stark von anderen. Es hat sicherlich einige Eigenheiten, aber es ist sehr wahrscheinlich, dass Ihr Problem bereits von jemand anderem gelöst worden ist. Zu lernen, wie Sie Ähnlichkeiten zwischen Ihren Datenmodellen und denen in den Beispielen finden, ist nicht einfach, aber sehr befriedigend. Die Lösung erscheint dann vor Ihren Augen, und die meisten Probleme mit Ihren Berechnungen verschwinden schlagartig.

Für die meisten unserer Beispiele verwenden wir die Datenbank von Contoso. Dabei handelt es sich um ein fiktives Unternehmen, das über verschiedene Vertriebskanäle elektronische Geräte in aller Welt verkauft. Ihr Geschäft unterscheidet sich sehr wahrscheinlich davon, wes-

halb Sie die Contoso-Berichte und die daraus gewonnenen Erkenntnisse auf Ihren Fall übertragen müssen.

Da dies das erste Kapitel ist, beschäftigen wir uns zunächst mit der Terminologie und den Grundprinzipien. Wir erklären, was ein Datenmodell ist und warum Beziehungen so wichtige Bestandteile davon sind. Außerdem führen wir die Begriffe Normalisierung, Denormalisierung und Sternschema ein. Die Vorgehensweise, Prinzipien anhand von Beispielen vorzuführen, halten wir im ganzen Buch ein, aber hier, in den ersten Schritten, ist sie viel offensichtlicher.

Schnallen Sie sich an und tauchen Sie ein in die Geheimnisse der Datenmodellierung!

Arbeiten mit einer einzelnen Tabelle

Wenn Sie Excel und Pivottabellen verwenden, um Erkenntnisse aus Ihren Daten zu ziehen, laden Sie diese Daten wahrscheinlich mithilfe einer Abfrage aus einer Quelle, gewöhnlich einer Datenbank. Anschließend erstellen Sie eine Pivottable aus diesem Dataset (Datenmenge) und beginnen mit Ihren Nachforschungen. Dabei unterliegen Sie natürlich den üblichen Einschränkungen von Excel, wobei die wichtigste lautet, dass das Dataset nicht mehr als 1.000.000 Zeilen umfassen darf, da es sonst nicht in ein Arbeitsblatt passt. Ehrlich gestanden, als wir zum ersten Mal von dieser Einschränkung hörten, hielten wir sie nicht einmal für eine Einschränkung. Warum um alles in der Welt sollte jemand 1.000.000 Zeilen in Excel laden wollen, anstatt eine Datenbank zu verwenden? Man könnte meinen, der Grund für ein solches Vorgehen liege daran, dass Excel im Gegensatz zu Datenbanken keine Kenntnisse über Datenmodellierung erfordert.

Wenn Sie tatsächlich Excel verwenden wollen, kann dies jedoch eine wirklich schwere Einschränkung darstellen. In der Contoso-Datenbank, die wir für unsere Beispiele verwenden, umfasst die Verkaufstabelle 12.000.000 Zeilen. Damit ist es nicht möglich, sie komplett in Excel zu laden, um mit der Analyse zu beginnen. Für dieses Problem gibt es jedoch eine einfache Lösung: Anstatt alle Zeilen abzurufen, verringern Sie die Anzahl, indem Sie eine Gruppierung durchführen. Wenn Sie beispielsweise an einer Analyse der Verkäufe nach Kategorie und Unterkategorie interessiert sind, laden Sie nicht die Verkaufszahlen für jedes Produkt, sondern gruppieren die Daten nach Kategorie und Unterkategorie, was die Anzahl der Zeilen erheblich reduziert.

Wenn Sie die Verkaufstabelle mit ihren 12.000.000 Zeilen nach Hersteller, Marke, Kategorie und Unterkategorie gruppieren und die Angabe der Verkäufe pro Tag beibehalten, erhalten Sie 63.984 Zeilen, was sich in einer Excel-Arbeitsmappe gut handhaben lässt. Die richtige Abfrage zu schreiben, um eine solche Gruppierung durchzuführen, ist gewöhnlich eine Aufgabe für die IT-Abteilung (sofern Sie nicht selbst SQL gelernt haben). Wenn Sie den Code für die Abfrage haben, können Sie mit der Analyse der Zahlen beginnen. In Abbildung 1–1 sehen Sie die ersten Zeilen der Tabelle nach dem Import in Excel.

Wenn die Tabelle in Excel geladen ist, können Sie sich endlich zu Hause fühlen und eine Pivottable zur Analyse der Daten erstellen. In Abbildung 1–2 sehen Sie als Beispiel die Verkaufszahlen pro Hersteller für eine gegebene Kategorie. Dabei wurden eine gewöhnliche Pivottable und ein Datenschnitt verwendet.

FullDateLabel	Manufacturer	BrandName	ProductSubcategoryName	ProductCategoryName	SalesQuantity	SalesAmount	TotalCost
2007-03-31	Adventure Works	Adventure Works	Coffee Machines	Home Appliances	55	14332.268	7651.84
2008-10-22	Contoso, Ltd	Contoso	Cell phones Accessories	Cell phones	2040	23504.88	12648.94
2009-01-31	Adventure Works	Adventure Works	Televisions	TV and Video	194	51593.106	28146.4
2009-01-21	Fabrikam, Inc.	Fabrikam	Camcorders	Cameras and camcorders	282	163007.2	76709.45
2007-12-31	Adventure Works	Adventure Works	Laptops	Computers	29	14008.43	7944.32
2007-06-22	Contoso, Ltd	Contoso	Cell phones Accessories	Cell phones	680	6107.24	3420.44
2007-06-22	Proseware, Inc.	Proseware	Projectors & Screens	Computers	86	71417.6	30786.94
2007-08-23	Adventure Works	Adventure Works	Laptops	Computers	43	22672.2	9954.6
2009-03-30	The Phone Company	The Phone Company	Touch Screen Phones	Cell phones	198	48500.37	24164.56
2008-03-24	Contoso, Ltd	Contoso	Home & Office Phones	Cell phones	306	7353.594	3914.64
2007-09-30	Fabrikam, Inc.	Fabrikam	Microwaves	Home Appliances	44	4805.604	2824.24
2007-11-13	Adventure Works	Adventure Works	Desktops	Computers	153	47357.97	28256.02
2008-12-06	Contoso, Ltd	Contoso	Projectors & Screens	Computers	32	10790.4	6477.2
2007-11-14	Contoso, Ltd	Contoso	Digital SLR Cameras	Cameras and camcorders	146	55397.5	25876
2009-12-30	Adventure Works	Adventure Works	Desktops	Computers	32	15107.75	7952.97
2009-03-13	Wide World Importers	Wide World Importers	Recording Pen	Audio	42	7990.92	3607.26
2009-08-11	Wide World Importers	Wide World Importers	Recording Pen	Audio	9	1466.1	749.16
2009-09-28	Contoso, Ltd	Contoso	Microwaves	Home Appliances	78	9955.268	5189.27
2008-02-18	A. Datum Corporation	A. Datum	Digital Cameras	Cameras and camcorders	345	70989.93	32872.58
2007-08-15	Litware, Inc.	Litware	Washers & Dryers	Home Appliances	69	112603.8	56472.35

Abbildung 1-1 Durch die Gruppierung von Verkaufsdaten entsteht eine kleine und leicht zu analysierende Tabelle.

ProductCategoryName	Sum of SalesAmount
Audio	141,178,573.89
Cameras and camcorders	85,468,758.14
Cell phones	44,940,846.17
Computers	173,760,754.90
Games and Toys	16,092,228.97
Home Appliances	140,433,368.67
Music, Movies and Audio Bo...	
TV and Video	
Grand Total	601,874,530.73

Abbildung 1-2 Aus einer Excel-Tabelle lässt sich leicht eine Pivottable erstellen.

Ob Sie es glauben oder nicht – damit haben Sie schon ein Datenmodell erstellt! Auch wenn es nur eine einzige Tabelle umfasst, ist es doch ein Datenmodell. Sie können nun seine analytischen Möglichkeiten erkunden und möglicherweise verbessern. Das Datenmodell in diesem Beispiel ist stark eingeschränkt, da es weniger Zeilen aufweist als die Quelltabelle.

Als Anfänger sind Sie vielleicht der Meinung, dass der Grenzwert von 1.000.000 Zeilen in einer Excel-Tabelle nur die Anzahl der Zeilen betrifft, die Sie zur Analyse abrufen können. Das stimmt zwar, aber diese Größeneinschränkung führt auch zu einer Einschränkung des Datenmodells und damit der analytischen Möglichkeiten Ihrer Berichte. Um die Anzahl der Zeilen zu verringern, mussten Sie die Daten schon in der Quelle gruppieren, sodass Sie nur die nach Spalten geordneten Verkäufe abrufen konnten, in diesem Beispiel nach Kategorie, Unterkategorie und einigen anderen Spalten.

Dadurch beschränken Sie implizit Ihre Analysemöglichkeiten. Wenn Sie beispielsweise einen Datenschnitt nach Farbe durchführen wollen, ist die Tabelle schon nicht mehr als Quelle geeignet, da sie keine Spalte für die Produktfarbe enthält. Eine Spalte zu der Abfrage hinzuzufügen ist kein großes Problem; das wirkliche Problem besteht darin, dass die Tabelle mit jeder Spalte größer wird, und zwar nicht nur, was die Breite (also die Anzahl der Spalten) angeht, sondern auch die Länge (die Anzahl der Zeilen). Tatsächlich wird aus einer einzigen Zeile mit den Verkäufen für eine gegebene Kategorie – z. B. Audio – ein Satz mehrerer Zeilen, jede davon mit der Kategorie Audio, aber mit Werten für die verschiedenen Farben.

Wenn Sie im Extremfall nicht im Voraus entscheiden wollen, welche Spalten Sie für den Datenschnitt verwenden möchten, müssen Sie alle 12.000.000 Zeilen laden – und damit können Sie keine Excel-Tabelle mehr verwenden. Das ist es, was wir meinten, als wir schrieben, dass die Modellierungsmöglichkeiten von Excel eingeschränkt sind. Nicht in der Lage zu sein, viele Zeilen zu laden, bedeutet implizit, nicht in der Lage zu sein, eine fortgeschrittene Analyse an umfangreichen Datenvolumen vorzunehmen.

Hier kommt Power Pivot ins Spiel. Damit haben Sie nicht mehr mit der Beschränkung auf 1.000.000 Zeilen zu leben. Es gibt praktisch keinen Grenzwert für die Anzahl der Zeilen, die Sie in eine Power Pivot-Tabelle laden können. Mit Power Pivot können Sie die komplette Verkaufstabelle in das Modell laden und eine tiefeschürfende Analyse der Daten durchführen.



Power Pivot ist seit Excel 2010 als externes Add-In in Excel verfügbar und seit Excel 2013 Teil des Produkts. Seit Excel 2016 verwendet Microsoft die neue Bezeichnung Excel-Datenmodell für ein Power Pivot-Modell, allerdings wird der Begriff Power Pivot ebenfalls noch verwendet.

Da Ihnen jetzt alle Verkaufsinformationen in einer einzigen Tabelle vorliegen, können Sie eine ausführlichere Analyse an den Daten ausführen. In Abbildung 1–3 sehen Sie beispielsweise eine Pivottable aus dem Datenmodell (d. h. aus Power Pivot), in der alle Spalten geladen sind. Jetzt können Sie Datenschnitte nach Kategorie, Farbe und Jahr durchführen, da all diese Informationen vorhanden sind. Mehr verfügbare Spalten in der Tabelle bedeuten mehr Analysemöglichkeiten.

ProductCategoryName	Sum of SalesAmount			
	Column Labels			
Row Labels	2007	2008	2009	Grand Total
Audio				
Cameras and camcorders				
Cell phones				
Computers				
Games and Toys				
Home Appliances				
Music, Movies and Audio Bo...				
TV and Video				
Black	\$59,783,936.63	\$63,073,257.64	\$70,489,997.35	\$193,347,191.62
Blue	\$5,128,405.12	\$6,516,691.17	\$7,715,161.42	\$19,360,257.71
Brown	\$6,301,722.60	\$7,183,128.75	\$4,904,738.33	\$18,389,589.68
Gold		\$39,185.40	\$122,755.63	\$161,941.03
Green	\$1,747,237.19	\$1,566,822.01	\$2,159,190.14	\$5,473,249.35
Grey	\$6,318,419.68	\$5,924,762.67	\$6,410,704.05	\$18,653,886.41
Orange		\$12,800.63	\$84,766.80	\$97,567.43
Pink	\$20,078.44	\$43,870.83	\$228,526.86	\$292,476.13
Red	\$6,926,045.96	\$7,872,382.02	\$11,495,613.50	\$26,294,041.47
Silver	\$43,375,399.72	\$39,082,679.67	\$36,471,502.90	\$118,929,582.29
White	\$64,963,894.39	\$64,142,854.42	\$70,639,615.97	\$199,746,364.78
Yellow	\$260,512.33	\$330,165.82	\$537,704.67	\$1,128,382.82
Grand Total	\$194,825,652.07	\$195,788,601.04	\$211,260,277.62	\$601,874,530.73

Abbildung 1–3 Wenn sämtliche Spalten verfügbar sind, können Sie aus Ihren Daten interessantere Pivottabellen erstellen.

Dieses einfache Beispiel vermittelt Ihnen schon einen ersten wichtigen Grundsatz der Datenmodellierung: *Größe ist wichtig, da sie mit der Granularität zusammenhängt*. Aber was ist *Granularität*? Da es sich um einen der wichtigsten Begriffe handelt, den Sie in diesem Buch kennenlernen werden, führen wir ihn hier möglichst früh ein. Im weiteren Verlauf werden wir die Erklärung noch vertiefen, aber zur Einführung wollen wir zunächst eine einfache Erklärung geben. In dem ersten Dataset haben Sie die Informationen nach Kategorie und Unterkategorie gruppiert und dabei auf einige Einzelheiten verzichtet, um die Größe zu verringern. Technisch ausgedrückt haben Sie eine Granularität auf der Ebene von Kategorien und Unterkategorien

gewählt. Sie können sich die Granularität als die Detailliertheit oder Feinheit Ihrer Tabellen vorstellen: je größer die Granularität, umso detaillierter die Informationen. Mit mehr Einzelheiten können Sie auch eine detailliertere Analyse durchführen. Beim letzten Dataset – also demjenigen, das wir in Power Pivot geladen haben – befindet sich die Granularität nicht mehr auf Kategorie- und Unterkategorie-, sondern auf Produktebene. (Tatsächlich ist sie noch feiner, nämlich auf der Ebene der einzelnen Verkäufe eines Produkts.) Ihre Möglichkeiten für Datenschnitte und zum Drehen (Slice and Dice) hängen von der Anzahl der Spalten in der Tabelle, also von deren Granularität ab. Wie Sie bereits wissen, haben Sie mit einer erhöhten Anzahl von Spalten auch mehr Zeilen.

Die richtige Granularität zu wählen, ist immer eine schwierige Aufgabe. Wenn Ihre Daten die falsche Granularität aufweisen, wird es fast unmöglich, Formeln zu schreiben, da die Informationen entweder verloren sind (wie in dem vorherigen Beispiel, wo es keine Farbinformationen mehr gab) oder über die Tabelle verstreut und falsch gegliedert. Es ist daher auch nicht richtig zu sagen, dass eine höhere Granularität immer gut ist. Die Daten müssen die *richtige* Granularität aufweisen, also diejenige, die sich am besten für den vorliegenden Zweck eignet.

Ein Beispiel für verlorene Informationen haben Sie bereits gesehen. Aber was bedeutet »verstreute« Informationen? Das lässt sich nicht ganz so einfach erkennen. Nehmen wir an, Sie wollen das durchschnittliche jährliche Einkommen der Kunden berechnen, die eine bestimmte Auswahl Ihrer Produkte kaufen. Diese Information ist vorhanden, denn in der Verkaufstabelle sind alle Informationen über die Kunden verfügbar. Das können Sie in Abbildung 1–4 erkennen, die einige Spalten der Tabelle zeigt, mit der wir arbeiten. (Um den Inhalt der Tabelle sehen zu können, müssen Sie das Power Pivot-Fenster öffnen.)

ProductCategoryName	ProductSubcategoryName	ProductName	SalesAmount	FirstName	LastName	YearlyIncome
Cameras and camcorders	Digital SLR Cameras	A. Datum SLR Camera X137 Grey	\$627.00	Katrina	Xie	€ 20,000.00
Cameras and camcorders	Digital SLR Cameras	A. Datum SLR Camera X137 Grey	\$627.00	Seth	Rodriguez	€ 80,000.00
Cameras and camcorders	Digital SLR Cameras	A. Datum SLR Camera X137 Grey	\$627.00	Evelyn	Arun	€ 10,000.00
Cameras and camcorders	Digital SLR Cameras	A. Datum SLR Camera X137 Grey	\$627.00	Christy	Beck	€ 40,000.00
Cameras and camcorders	Digital SLR Cameras	A. Datum SLR Camera X137 Grey	\$627.00	Alejandro	Nara	€ 40,000.00
Cameras and camcorders	Digital SLR Cameras	A. Datum SLR Camera X137 Grey	\$627.00	Leah	Lu	€ 30,000.00
Cameras and camcorders	Digital SLR Cameras	A. Datum SLR Camera X137 Grey	\$627.00	Robyn	Torres	€ 20,000.00
Cameras and camcorders	Digital SLR Cameras	A. Datum SLR Camera X137 Grey	\$627.00	Jimmy	Moreno	€ 30,000.00
Cameras and camcorders	Digital SLR Cameras	A. Datum SLR Camera X137 Grey	\$627.00	Rafael	Cai	€ 20,000.00
Cameras and camcorders	Digital SLR Cameras	A. Datum SLR Camera X137 Grey	\$627.00	Jenny	Ferrier	€ 110,000.00
Cameras and camcorders	Digital SLR Cameras	A. Datum SLR Camera X137 Grey	\$627.00	Levi	Arun	€ 70,000.00
Cameras and camcorders	Digital SLR Cameras	A. Datum SLR Camera X137 Grey	\$627.00	Randall	Torres	€ 40,000.00

Abbildung 1–4 Die Produkt- und die Kundeninformationen stehen in derselben Tabelle.

In jeder Zeile der Tabelle *Sales* befindet sich auch eine Spalte mit dem Jahreseinkommen des Kunden, der das betreffende Produkt gekauft hat. Für einen einfachen Versuch, das durchschnittliche Jahreseinkommen der Kunden zu berechnen, können wir wie folgt ein berechnetes Feld (*Measure*) aus DAX verwenden:

```
AverageYearlyIncome := AVERAGE ( Sales[YearlyIncome] )
```

Das berechnete Feld funktioniert sehr gut. Sie können es in einer Pivottabelle wie in Abbildung 1–5 einsetzen, die das durchschnittliche Jahreseinkommen der Kunden zeigt, die Haushaltsgeräte verschiedener Marken kaufen.

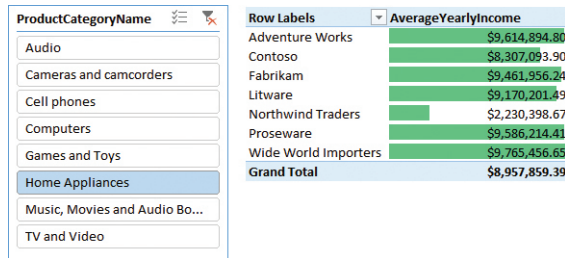


Abbildung 1–5 Analyse des durchschnittlichen Jahreseinkommens von Kunden, die Haushaltsgeräte kaufen.

Der Bericht sieht gut aus, aber leider ist die berechnete Zahl falsch, nämlich viel zu hoch. Was Sie hier berechnen, ist der Durchschnitt über die Verkaufstabelle, die eine Granularität auf der Ebene einzelner Verkäufe aufweist. Das heißt, die Tabelle enthält eine Zeile für jeden Verkauf, also möglicherweise mehrere Zeilen für denselben Kunden. Kauft ein Kunde also beispielsweise drei Produkte an drei verschiedenen Tagen, wird er bei der Bildung des Durchschnitts dreimal gezählt, was zu einem falschen Ergebnis führt.

Man könnte meinen, dadurch würde ein gewichteter Durchschnitt berechnet, aber auch das ist nicht ganz korrekt. Um einen gewichteten Durchschnitt zu berechnen, müssen Sie eine Gewichtung definieren, und dazu ziehen Sie nicht einfach die Anzahl der Kaufvorgänge heran, sondern die Anzahl der Produkte, den Gesamtbetrag oder irgendeinen anderen bedeutungsvollen Wert. Außerdem hatten wir in diesem Beispiel ja ohnehin vorgehabt, nur einen einfachen Durchschnitt zu berechnen, und das erledigt das berechnete Feld nun einmal nicht korrekt.

Es ist nicht so leicht zu erkennen, aber auch hier haben wir mit dem Problem einer falschen Granularität zu kämpfen. Die Informationen sind zwar verfügbar, aber nicht mit einem einzelnen Kunden verknüpft, sondern über die ganze Verkaufstabelle verstreut, weshalb es schwer ist, die Berechnungsformel zu schreiben. Um den richtigen Durchschnitt zu erhalten, müssen Sie für Granularität auf Kundenebene sorgen, indem Sie entweder die Tabelle neu laden oder eine kompliziertere DAX-Formel einsetzen.

Letzteres können Sie wie folgt tun, allerdings ist diese Formel nicht ganz einfach zu verstehen:

```
CorrectAverage :=
AVERAGEX (
    SUMMARIZE (
        Sales,
        Sales[CustomerKey],
        Sales[YearlyIncome]
    ),
    Sales[YearlyIncome]
)
```

Als Erstes müssen Sie die Verkäufe auf Kundenebene aggregieren (um eine Granularität auf Kundenebene zu bekommen). Erst dann können Sie eine AVERAGE-Operation auf der resultierenden Tabelle durchführen, in der jeder Kunde nur einmal vorkommt. In diesem Beispiel ver-

wenden wir SUMMARIZE, um die Vorabaggregation auf Kundenebene in einer temporären Tabelle durchzuführen, und ermitteln dann den Durchschnitt von *YearlyIncome* in dieser temporären Tabelle. Wie Sie in Abbildung 1–6 sehen, unterscheidet sich die korrekte Zahl erheblich von dem zuvor berechneten falschen Wert.

ProductCategoryName	AverageYearlyIncome	CorrectAverage
Adventure Works	\$9,614,894.80	\$535,593.62
Contoso	\$8,307,093.90	\$262,307.94
Fabrikam	\$9,461,956.20	\$361,924.73
Litware	\$9,170,201.49	\$265,677.30
Northwind Traders	\$2,230,398.67	\$151,583.50
Proseware	\$9,586,214.41	\$491,908.56
Wide World Importers	\$9,765,456.65	\$1,035,131.95
Grand Total	\$8,957,859.39	\$260,183.91

Abbildung 1–6 Die Gegenüberstellung der Daten für den richtigen und den falschen Durchschnitt zeigt, wie weit wir danebengelegt haben.

Es lohnt sich, etwas mehr Zeit zu investieren, um sich voll und ganz mit der folgenden einfachen Wahrheit vertraut zu machen: Das Jahreseinkommen ist eine Information, die auf der Ebene des einzelnen Kunden von Bedeutung ist. Auf der Ebene der einzelnen Verkäufe dagegen ist diese Zahl fehl am Platze. Anderes ausgedrückt, Sie können einen Wert, der eine Bedeutung auf Kundenebene hat, nicht mit derselben Bedeutung auf der Ebene der einzelnen Verkäufe verwenden. Um das richtige Ergebnis zu erhalten, mussten wir hier die Granularität reduzieren, wenn auch nur in einer temporären Tabelle.

Aus diesem Beispiel können Sie zwei wichtige Dinge lernen:

- Die richtige Formel ist weit komplizierter als ein einfaches AVERAGE. Sie müssen Werte vorübergehend aggregieren, um die korrekte Granularität der Tabelle zu erhalten, da die Daten in der Tabelle verstreut und nicht auf geeignete Weise geordnet sind.
- Es ist sehr wahrscheinlich, dass Sie einen solchen Fehler übersehen, wenn Sie mit den Daten nicht vertraut sind. Ein Blick auf den Bericht in Abbildung 1–6 lässt schon ahnen, dass das Jahreseinkommen viel zu hoch ist, um wahr zu sein – als ob keiner der Kunden weniger als 2.000.000 Dollar pro Jahr verdienen würde! Bei komplizierteren Berechnungen jedoch kann es weit schwieriger sein, den Fehler zu erkennen, und das kann zu Berichten mit falschen Zahlen führen.

Sie müssen die Granularität erhöhen, um Berichte mit dem gewünschten Detaillierungsgrad hervorzurufen, aber wenn Sie sie zu weit erhöhen, wird die Berechnung mancher Zahlen schwieriger. Wie wählen Sie daher die korrekte Granularität? Das ist eine knifflige Frage, deren Beantwortung wir uns für später aufheben. Wir hoffen, Ihnen die erforderlichen Kenntnisse vermitteln zu können, um die richtige Granularität der Daten in Ihren Modellen erkennen zu können. Allerdings ist die Auswahl der richtigen Granularität eine Fähigkeit, die sich selbst erfahrene Datenmodellierer nicht so leicht aneignen können. Vorläufig jedoch begnügen wir uns mit der Erkenntnis, was Granularität ist und wie wichtig es ist, die richtige Granularität für jede Tabelle in Ihrem Modell zu wählen.

Das Modell, an dem wir bis jetzt gearbeitet haben, leidet in Wirklichkeit unter einem viel größeren Problem, das jedoch auch in gewissem Sinne mit der Granularität zu tun hat. Das

größte Problem dieses Modells besteht darin, dass es nur über eine einzige Tabelle verfügt, die sämtliche Informationen enthält. In einem solchen Modell müssen Sie die Granularität der Tabelle wählen und dabei alle möglichen berechneten Felder und Analysen berücksichtigen, die Sie benutzen bzw. durchführen wollen. Wie sehr Sie sich auch anstrengen, die gewählte Granularität wird niemals ideal für alle berechneten Felder sein. In den nächsten Abschnitten führen wir die Verwendung mehrerer Tabellen ein, bei denen Sie unterschiedliche Granularitäten auswählen können.

Datenmodelle

Im vorherigen Abschnitt haben Sie gelernt, dass ein Modell mit einer einzigen Tabelle Probleme bei der Festlegung der richtigen Granularität hervorruft. Excel-Benutzer verwenden oft Einzeltabellenmodelle, da dies vor der Version 2013 die einzige Möglichkeit zum Erstellen von Pivottabellen war. In Excel 2013 hat Microsoft das Excel-Datenmodell eingeführt, mit dem Sie viele Tabellen laden und über Beziehungen verbinden können. Dadurch können Sie viel leistungsfähigere Datenmodelle erstellen.

Was aber ist ein *Datenmodell*? Dabei handelt es sich einfach um einen Satz von Tabellen, die durch Beziehungen verknüpft sind. Ein Einzeltabellenmodell ist bereits ein Datenmodell, aber kein sehr interessantes. Wenn Sie mehrere Tabellen haben, machen die Beziehungen dazwischen das Modell viel leistungsfähiger und interessanter zu analysieren.

Ein Datenmodell zu erstellen, ist ein ganz natürlicher Vorgang, sobald Sie mehr als eine Tabelle laden. Überdies laden Sie gewöhnlich Daten aus Datenbanken, die von Profis gepflegt werden und bereits über ein Datenmodell verfügen. Das bedeutet, dass Ihr Datenmodell sehr wahrscheinlich das in der Quelldatenbank bereits vorhandene Modell nachstellen wird. In gewissem Sinne vereinfacht das Ihre Arbeit.

Doch wie Sie in diesem Buch noch sehen werden, ist es leider sehr unwahrscheinlich, dass das Quelldatenmodell die ideale Struktur für die Art von Analyse aufweist, die Sie durchführen möchten. Wir werden Ihnen anhand von Beispielen mit zunehmender Komplexität zeigen, wie Sie von einer beliebigen Datenquelle ausgehend Ihr eigenes Modell aufbauen können. Um Ihnen das Lernen zu vereinfachen, beschreiben wir diese Techniken nach und nach im weiteren Verlauf dieses Buches. Zunächst einmal beginnen wir mit den Grundlagen.

Um das Prinzip eines Datenmodells kennenzulernen, laden Sie die Tabellen *Product* und *Sales* der Contoso-Datenbank in das Excel-Datenmodell. Anschließend sehen Sie die Diagrammansicht aus Abbildung 1–7, die die beiden Tabellen mit ihren Spalten zeigt.



Das Beziehungsdiagramm steht in Power Pivot zur Verfügung. Um darauf zuzugreifen, klicken Sie im Excel-Menüband auf die Registerkarte *Power Pivot* und dann auf *Verwalten*. Klicken Sie anschließend auf der Registerkarte *Start* des Power Pivot-Fensters in der Gruppe *Ansicht auf Diagrammansicht*.

Zwei unverbundene Tabellen wie in diesem Beispiel sind noch kein echtes Datenmodell, sondern nur zwei Tabellen. Um sie in ein sinnvolles Modell umzuwandeln, müssen Sie Beziehungen zwischen ihnen herstellen. In diesem Beispiel verfügt sowohl die Tabelle *Sales* als auch die

Tabelle *Product* über die Spalte *ProductKey*. In *Product* ist diese Spalte der *Primärschlüssel*. Das bedeutet, dass sie in jeder Zeile einen anderen Wert aufweist und daher verwendet werden kann, um ein Produkt eindeutig zu identifizieren. In der Tabelle *Sales* dagegen dient diese Spalte einem anderen Zweck, nämlich das verkaufte Produkt anzugeben.

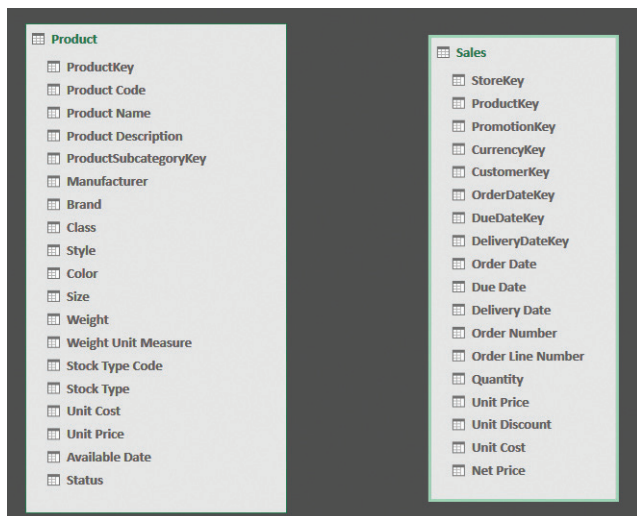


Abbildung 1–7 Mit dem Excel-Datenmodell können Sie mehrere Tabellen laden.

Der **Primärschlüssel** einer Tabelle ist eine Spalte, die in jeder Zeile einen anderen Wert aufweist. Wenn Sie daher einen Wert aus dieser Spalte kennen, können Sie ihn eindeutig einer Zeile zuordnen. Es kann mehrere Spalten mit solchen eindeutigen Werten geben, wobei alle diese Spalten Schlüssel sind. Der Primärschlüssel ist nichts Besonderes. Technisch gesehen ist er lediglich die Spalte, die Sie zur eindeutigen Bezeichnung einer Zeile verwenden. In einer Kundentabelle kann dies beispielsweise die Kundennummer sein, auch wenn die Werte in der Namensspalte möglicherweise alle eindeutig sind.



Wenn Sie in einer Tabelle einen eindeutigen Bezeichner haben und eine Spalte in einer anderen Tabelle darauf verweist, können Sie eine Beziehung zwischen den beiden Tabellen herstellen. Für eine gültige Beziehung müssen beide Bedingungen gelten. Wenn der gewünschte Schlüssel für die Beziehung in keiner der beiden Tabellen Ihres Modells ein eindeutiger Bezeichner ist, müssen Sie das Modell mit einer der Techniken abwandeln, die Sie in diesem Buch kennenlernen werden. Zunächst einmal wollen wir aber anhand unseres Beispiels einige grundlegende Tatsachen über Beziehungen feststellen:

- **Die Tabelle *Sales* ist die *Quelltabelle*** Die Beziehung geht von *Sales* aus, denn um ein Produkt abzurufen, beginnen Sie stets bei *Sales*: Sie ermitteln den Wert der Produktnummer in *Sales* und suchen anschließend in *Product* danach. Danach kennen Sie das Produkt und all seine Attribute.

- **Die Tabelle *Product* ist das Ziel der Beziehung** Das liegt daran, dass Sie bei *Sales* anfangen und dann zu *Product* übergehen. Daher ist *Product* das Ziel Ihrer Suche.
- **Eine Beziehung verläuft von der Quelle zum Ziel** Anderes ausgedrückt: Eine Beziehung hat eine Richtung. Aus diesem Grund werden Beziehungen oft in Form von Pfeilen von der Quelle zum Ziel abgebildet. Allerdings werden in Softwareprodukten unterschiedliche grafische Darstellungen von Beziehungen verwendet.
- **Die Quelltable ist die *n-Seite* der Beziehung** Diese Bezeichnung rührt daher, dass es für ein Produkt sehr wahrscheinlich viele Verkaufsvorgänge gibt, für einen Verkaufsvorgang aber immer nur ein Produkt. Daher haben wir hier eine *n:1-Beziehung*, bei der die Zieltabelle die 1-Seite und die Quelltable die n-Seite ist.
- **Die Spalte *ProductKey* ist sowohl in *Sales* als auch in *Products* vorhanden** In *Product* ist *ProductKey* ein Schlüssel, aber nicht in *Sales*. Daher wird diese Spalte in *Product* als Primärschlüssel bezeichnet, in *Sales* dagegen als *Fremdschlüssel*, also als eine Spalte, die auf einen Primärschlüssel in einer anderen Tabelle verweist.

All dies sind gebräuchliche Begriffe in der Datenmodellierung, weshalb wir sie auch in diesem Buch verwenden. Machen Sie sich aber keine Sorgen, denn wir werden die Definitionen in den ersten Kapiteln noch einige Male wiederholen, damit Sie sich damit vertraut machen können.

Mit Excel und Power BI können Sie eine Beziehung zwischen zwei Tabellen herstellen, indem Sie den Fremdschlüssel (in unseren Beispielen *ProductKey* in *Sales*) auf den Primärschlüssel (*ProductKey* in *Product*) ziehen. Dabei werden Sie feststellen, dass weder Excel noch Power BI Pfeile zur Darstellung von Beziehungen verwenden. Stattdessen wird die Beziehung in der Diagrammansicht durch eine Zahl (1) für die 1-Seite und ein Sternchen für die n-Seite angezeigt, wie Sie in Abbildung 1–8 sehen. In der Mitte steht zwar auch ein Pfeil, aber er gibt nicht die Richtung der Beziehung an, sondern der Filterweiterleitung, was etwas völlig anderes ist. Damit werden wir uns weiter hinten in diesem Buch noch beschäftigen.



Sollte die Power Pivot-Registerkarte verschwinden, liegt das wahrscheinlich daran, dass Excel ein Problem festgestellt und das Add-In ausgeschaltet hat. Um es wieder zu aktivieren, klicken Sie auf die Registerkarte *Datei*, dann im linken Bereich auf *Optionen* und im linken Teil des Optionenfensters auf *Add-Ins*. Öffnen Sie das Listenfeld *Verwalten* am unteren Rand der Seite, wählen Sie *COM-Add-Ins* und klicken Sie auf *Los*. Wählen Sie anschließend im Fenster *COM-Add-Ins* das Add-In *Microsoft Power Pivot for Excel*. Sollte es bereits ausgewählt sein, wählen Sie es ab. Klicken Sie auf *OK*. Haben Sie die Auswahl von Power Pivot aufgehoben, müssen Sie das Fenster *COM-Add-Ins* erneut aufrufen und das Add-In wieder auswählen. Anschließend sollte die Power Pivot-Registerkarte wieder im Menüband erscheinen.

Wenn eine Beziehung vorhanden ist, können Sie die Werte in der Tabelle *Sales* summieren und mit einem Datenschnitt nach Spalten in der Tabelle *Products* filtern. Wie Sie in Abbildung 1–9 sehen, können Sie die Summe der Stückzahlen (Spalte *Quantity* in *Sales*; siehe Abbildung 1–8) nach der Farbe (Spalte *Color* in *Product*) filtern.