Оглавление

Рецензия

Вы держите в руках уникальную по нескольким причинам книгу.

Во-первых, это первая книга на русском языке по системе бизнес-аналитики Microsoft Power BI. В течение нескольких последних лет, когда слушатели после тренингов по Excel, Power Pivot и Query спрашивали «что мне почитать про Power BI?», я не знал, что ответить. Англоязычной литературы написано по этой теме уже много, но на русском – полный ноль. Теперь уже нет.

Во-вторых, я очень рад, что в качестве первой ласточки издательство «ДМК Пресс» решило перевести именно эту книгу. Альберто Феррари и Марко Руссо однозначно входят в круг самых достойных авторов в этой области. Они щедро делятся своими знаниями в книгах и статьях, выступают на конференциях и проводят тренинги по Power Pivot, DAX и Power BI ещё с самого начала появления этих технологий и знают о них больше, чем кто бы то ни было. Отдельно, как тренер, хочу отметить их преподавательский талант, стройность и логичность объяснений, красоту примеров – это дорогого стоит.

Бизнес-аналитика (Business Intelligence, BI) давно уже перестала быть уделом гиков-айтишников из миллиардных корпораций. Сегодня она способна принести пользу при принятии управленческих решений в компании любого калибра, помочь визуализировать результаты и непрерывно отслеживать их динамику, собирая данные из разных «вселенных»: бухгалтерских программ, баз данных, файлов, интернета. Сегодня каждый может (и должен!) быть «сам себе аналитик». И эта книга – настоящий клад и огромное подспорье для всех, кто встал на этот путь.

Николай Павлов,
Microsoft Certified Trainer, Microsoft Most Valuable Professional,
автор проекта «Планета Excel», www.planetaexcel.ru

Предисловие от издательства

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв прямо на нашем сайте **www.dmkpress.com**, зайдя на страницу книги, и оставить комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу **dmkpress@gmail.com**, при этом напишите название книги в теме письма.

Если есть тема, в которой вы квалифицированы, и вы заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу **http://dmkpress.com/authors/publish_book/** или напишите в издательство по адресу **dmkpress@gmail.com**.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы удостовериться в качестве наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг – возможно, ошибку в тексте или в коде, – мы будем очень благодарны, если вы сообщите нам о ней. Сделав это, вы избавите других читателей от расстройств и поможете нам улучшить последующие версии этой книги.

Если вы найдете какие-либо ошибки в коде, пожалуйста, сообщите о них главному редактору по адресу **dmkpress@gmail.com**, и мы исправим это в следующих тиражах.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательство «ДМК Пресс» очень серьезно относится к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконно выполненной копией любой нашей книги, пожалуйста, сообщите нам адрес копии или веб-сайта, чтобы мы могли применить санкции.

Пожалуйста, свяжитесь с нами по адресу электронной почты **dmkpress@gmail.com** со ссылкой на подозрительные материалы.

Мы высоко ценим любую помощь по защите наших авторов, помогающую нам предоставлять вам качественные материалы.

Введение

Пользователи Excel любят цифры. А может, те, кто любят цифры, любят Excel. Как бы то ни было, если вам нравится доходить до самой сути при анализе любых наборов данных, скорее всего, вы провели немало времени, работая с Excel, сводными таблицами и формулами.

В 2015 году увидел свет программный продукт Power BI. И сегодня справедливо будет утверждать, что те, кто любят цифры, любят также Power Pivot для Excel и Power BI. Эти средства имеют много общего – в частности, их объединяет движок баз данных VertiPaq, а также язык DAX, унаследованный от SQL Server Analysis Services.

В прежних версиях Excel процесс анализа информации главным образом основывался на загрузке наборов данных, расчете значений в столбцах и написании формул для построения графиков. При этом в своей работе вы сталкивались с серьезными ограничениями – начиная с размера рабочей книги и заканчивая тем, что язык формул Excel не лучшим образом подходит для решения числовых задач большого объема. Новый движок, лежащий в основе Power BI и Power Pivot, стал огромным шагом вперед. С ним в вашем распоряжении оказался полный функционал баз данных, а также потрясающий язык DAX. Но ведь с большой силой приходит и большая ответственность! И если вы хотите воспользоваться всеми преимуществами этих новых средств, вам придется многому научиться. В частности, необходимо будет познакомиться с основами моделирования данных.

Моделирование данных – это отнюдь не ядерная физика, а лишь набор базовых знаний, которым должен овладеть всякий, кто заинтересован в анализе данных. К тому же если вы любите цифры, то вам непременно придется по душе моделирование данных. Освоить эту науку будет несложно, а вместе с тем вы получите массу удовольствия.

В этой книге вы познакомитесь с базовыми концепциями моделирования данных на практических примерах, с которыми наверняка не раз встречались в жизни. В наши планы не входило написание запутанной книги с подробным описанием комплексных решений, необходимых для реализации сложных систем. Вместо этого мы сосредоточились на реальных ситуациях, с которыми ежедневно сталкиваемся в работе в качестве консультантов. Когда к нам обращались за помощью, а мы видели, что имеем дело с типичной задачей, то отправляли ее прямиком в архив. Позже, открыв заветный ящик, мы получили ценные примеры для книги и расположили их в порядке, пригодном для обучения моделированию данных.

Прочитав эту книгу, вы вряд ли станете гуру в области создания моделей данных, но знаний по этой теме у вас существенно прибавится. И если впоследствии в поиске решения очередной задачи на вычисление нужного вам значения вы допустите мысль об изменении модели данных, значит, мы поработали не зря. Кроме того, вы уверенно вступите на путь становления успешного специалиста в области моделирования данных. Но заключительный шаг к вершине вы сможете сделать, только набравшись практического опыта и набив немало шишек. К сожалению, опыт нельзя приобрести, читая книги.

Для кого предназначена эта книга?

Целевая аудитория книги довольно разнообразна. В нее входят и пользователи Excel, применяющие в своей практике Power Pivot, и специалисты по анализу данных в Power BI, и даже новички в области бизнес-аналитики, желающие познакомиться с основами моделирования данных. Все они потенциальные читатели данной книги.

Заметьте, что мы не включили в этот список тех, кто целенаправленно хочет почитать о создании моделей данных. Изначально мы предполагали, что наш читатель может даже не знать, что ему нужно какое-то моделирование каких-то данных. Наша цель – дать вам понять, что проектирование моделей данных – это как раз то, что вам нужно, и познакомить с базовыми принципами этой прекрасной науки. В общем, если вам интересно, что такое моделирование данных и чем оно так полезно, эта книга для вас.

Как мы представляем себе нашего читателя?

Мы предполагаем, что наш читатель обладает базовыми знаниями в области сводных таблиц Excel и/или имеет опыт использования Power BI в качестве средства отчетности и моделирования. Наличие аналитических навыков также приветствуется. В своей книге мы не затрагиваем вопросы интерфейса Excel или Power BI. Вместо этого мы фокусируем свое внимание исключительно на моделях данных – как проектировать и модифицировать их так, чтобы значительно упростить запросы. Так что наша задача – рассказать вам, что делать, а как это делать, вы уж решите сами. Мы не планировали создавать пошаговое руководство, а хотели максимально простым языком объяснить достаточно сложную тему.

Также мы намеренно обошли вниманием описание языка DAX. Было бы невозможно уместить в одной книге и теорию моделирования данных, и DAX. Если вы уже знакомы с этим языком, вам будет проще разобраться с многочисленными примерами кода на DAX, представленными в данной книге. В противном случае советуем вам прочитать книгу «Подробное руководство по DAX» (The Definitive Guide to DAX), являющуюся полноценным

учебником по этому языку и хорошо сочетающуюся с приведенными в нашей книге примерами.

Структура книги

Книга начинается с пары легких вводных глав, за которыми следуют главы, каждая из которых посвящена отдельному виду модели данных. Предлагаем вам краткое описание:

- глава 1 «Введение в моделирование данных». Является вводной частью в базовые принципы моделирования данных. В ней мы расскажем, что из себя представляет модель данных, начнем говорить о понятии гранулярности, определим понятия основных моделей хранилища данных – «звезда» и «снежинка», – а также поговорим о нормализации и денормализации;
- глава 2 «Использование главной/подчиненной таблицы». Описывает наиболее распространенный сценарий с наличием главной и подчиненной таблиц. В этой главе мы обсудим пример с заказами и строками заказов, размещенными в двух отдельных таблицах фактов;
- глава 3 «Использование множественных таблиц фактов». Описывает сценарии, в которых у вас есть множество таблиц фактов, на основании которых необходимо построить единый отчет. В этой главе мы подчеркнем важность создания корректной многомерной модели для облегчения работы с информацией;
- глава 4 «Работа с датой и временем». Это одна из самых длинных глав книги. В ней затронуты вопросы логики расчетов на основании временных периодов. Мы расскажем, как правильно создать таблицу-календарь и работать с функциями времени (YTD, QTA, PARALLELPERIOD и др.). После этого приведем несколько примеров расчетов на основании рабочих дней, поработаем с особыми периодами года и поясним в целом, как правильно работать с датами;
- глава 5 «Отслеживание исторических атрибутов». В этой главе описываются особенности использования в модели данных медленно меняющихся измерений. Также представлено детальное описание трансформаций, которые необходимо выполнить для отслеживания исторических атрибутов, и даны инструкции по написанию корректного кода на DAX, учитывающего медленно меняющиеся измерения;
- глава 6 «Использование снимков». Описывает любопытные аспекты использования снимков (snapshot). В этой главе вы узнаете, что такое снимки, когда и для чего их необходимо использовать, а также как рассчитывать значения при применении снимков. Кроме того, мы посмотрим, как можно использовать мощную модель с применением матрицы переходов;

- глава 7 «Анализ интервалов даты и времени». В этой главе мы пойдем еще на шаг дальше, чем в главе 5. Мы продолжим заниматься временными вычислениями, но на этот раз обратимся к модели данных, в которой события, хранящиеся в таблице фактов, обладают определенной длительностью, а значит, требуют особого подхода для получения корректных результатов;
- глава 8 «Связи многие ко многим». Описывает характерные особенности использования связей «многие ко многим». Такой тип связи играет важную роль в любой модели данных. Мы рассмотрим обычные связи «многие ко многим», связи с каскадными действиями и их использование с учетом факторов перераспределения и фильтров. Также обсудим вопросы производительности таких связей и способы ее улучшения;
- глава 9 «Работа с разными гранулярностями». В этой главе мы углубимся в работу с таблицами фактов с разными уровнями гранулярности. Мы рассмотрим примеры из области бюджетирования, в которых таблицы фактов будут хранить информацию с разной степенью детализации, и предложим несколько альтернативных способов для решения этих ситуаций как при помощи языка DAX, так и непосредственно в модели данных;
- глава 10 «Сегментация данных в модели». В этой главе мы рассмотрим несколько моделей с применением техники сегментации. Начнем с простой сегментации по цене, после чего перейдем к анализу динамической сегментации с использованием виртуальных связей. В конце главы проведем ABC-анализ средствами DAX;
- глава 11 «Работа с несколькими валютами». В этой главе мы рассмотрим особенности работы с несколькими валютами. Взаимодействуя с курсами валют, важно понимать их специфику и в соответствии с ней строить модель данных. Мы проанализируем несколько сценариев с разными требованиями и для каждого из них выработаем оптимальное решение;
- приложение А «Моделирование данных 101». Это приложение можно рассматривать как справочное руководство. Здесь мы кратко опишем на примерах все базовые концепции, использованные в этой книге. При возникновении вопросов вы всегда можете обратиться к приложению, освежить в памяти соответствующую тему и вернуться к чтению.

Сложность моделей и решений будет возрастать на протяжении всей книги, так что мы советуем читать ее последовательно, а не прыгать от главы к главе. Так вы сможете постепенно идти от простого к сложному и осваивать по одной теме за раз. После прочтения книга может стать для вас справочным руководством, и когда вам потребуется построить ту или иную модель данных, вы можете смело открыть нужную главу и воспользоваться предложенным решением.

Условные обозначения

В этой книге приняты следующие условные обозначения:

- **жирным** помечен текст, который вводите вы;
- *курсив* используется для обозначения новых терминов;
- программный код обозначен в книге `моноширинным шрифтом`;
- первые буквы в названиях диалоговых окон, их элементов, а также команд – прописные. Например, в диалоговом окне **Save As...** (Сохранить как...);
- комбинации нажимаемых клавиш на клавиатуре обозначаются знаком плюс (+) между названиями клавиш. Например, **Ctrl**+**Alt**+**Delete** означает, что вы должны одновременно нажать клавиши **Ctrl**, **Alt** и **Delete**.

Сопутствующий контент

Для подкрепления ваших навыков на практике мы снабдили книгу сопутствующим контентом, который можно скачать по ссылке: https://aka.ms/AnalyzeData/downloads.

Представленный архив содержит файлы в форматах Excel и/или Power BI Desktop для всех примеров из этой книги. Каждому рисунку соответствует отдельный файл, чтобы вы имели возможность анализировать разные шаги и присоединиться к выполнению примера на любой стадии. Для большинства примеров представлены файлы в формате Power BI Desktop, так что мы настоятельно рекомендуем вам установить этот программный пакет с сайта Power BI.

Благодарности

В конце вводной главы мы бы хотели выразить благодарность нашему редактору Кейт Шуп (Kate Shoup), которая помогала нам на протяжении всей книги, и техническому редактору Эду Прайсу (Ed Price). Если бы не их дотошность, читать эту книгу было бы гораздо труднее. Если книга содержит меньше ошибок, чем наша первоначальная рукопись, это только их заслуга. А во всех оставшихся неточностях виноваты лишь мы.

Список опечаток и поддержка

Мы сделали все возможное, чтобы текст и сопутствующий контент к этой книге не содержали ошибок. Все неточности, которые были обнаружены после публикации издания, перечислены на сайте Microsoft Press по адресу: https://aka.ms/AnalyzeData/errata.

Если вы нашли опечатку, которая не указана в перечне, вы можете оповестить нас на той же странице.

Если вам требуется дополнительная помощь, направьте письмо в Microsoft Press Book Support по адресу: mspinput@microsoft.com.

Отметим, что услуги по поддержке программного обеспечения Microsoft по этому адресу не оказываются.

Обратная связь

Ваше удовлетворение от книги – главный приоритет для Microsoft Press, а ваша обратная связь – наш самый ценный актив. Пожалуйста, выскажите свое мнение об этой книге по адресу: https://aka.ms/tellpress.

Пройдите небольшой опрос, и мы прислушаемся ко всем вашим идеям и пожеланиям. Заранее благодарим за ваши отзывы!

Оставайтесь с нами

Давайте продолжим общение! Заходите на наш Twitter: @MicrosoftPress.

Глава 1

Введение в моделирование данных

Книга, которую вы держите в руках, посвящена *моделированию данных* (data modeling). Но перед тем как приступать к чтению, неплохо бы понять, зачем вам вообще нужно изучать моделирование данных. В конце концов, вы можете просто загрузить нужные данные в Excel и построить на их основе сводную таблицу. Так зачем вам еще что-то знать о моделировании данных?

К нам как к консультантам в этой области часто обращаются частные лица и компании, которые не могут рассчитать какие-то нужные им показатели. При этом они понимают, что все исходные данные для расчета у них есть, но либо формула получается чересчур сложной и запутанной, либо цифры не сходятся. В 99 % случаев причиной является неправильно спроектированная *модель данных* (data model). Если ее поправить, формула станет простой и понятной. Так что вам просто необходимо научиться моделировать данные, если вы хотите улучшить свои аналитические навыки и предпочитаете концентрироваться на принятии правильных решений, а не на поиске замысловатой формулы в справочнике по DAX.

Обычно считается, что моделирование данных – непростая тема для изучения. И мы не станем этого отрицать. Это действительно сложная область. Она потребует от вас серьезных усилий, к тому же вам нужно будет постараться перестроить сознание так, чтобы сразу мыслить категориями модели данных, рассуждая о возможных сценариях. Так что да, моделирование данных – тема непростая, ресурсоемкая и требующая немалых усилий в освоении. Иными словами, сплошное удовольствие!

В этой главе мы покажем вам несколько примеров того, как правильно спроектированная модель данных помогает облегчить написание итоговых формул. Конечно, это всего лишь примеры, и они могут не относиться напрямую к стоящим перед вами задачам. Но мы надеемся, что их будет достаточно для понимания того, почему стоит изучать моделирование данных. Быть хорошим специалистом по моделированию данных – значит уметь подгонять актуальную модель под шаблоны, изученные и решенные

другими. Ваша модель данных ничем не отличается от других. Да, в ней есть свои особенности, но высока вероятность, что до вас с подобными задачами уже кто-то сталкивался. Научиться выявлять сходства между вашим примером и моделями, описанными в книге, не так просто, но в то же время очень приятно. Когда вы достигнете успеха в этом, решения задач начнут появляться перед вами сами, а большинство проблем с расчетом нужных вам показателей просто исчезнут.

В основном в своих примерах мы будем использовать базу данных Contoso. Это вымышленная компания, торгующая электроникой по всему миру с использованием различных каналов продаж. Вероятно, вы ведете совершенно иной бизнес – в этом случае вам придется адаптировать отчеты под свои нужды.

Поскольку это первая глава, начнем мы с описания общей терминологии и концепции. Мы расскажем, что такое модель данных и почему в ней так важны связи. Также мы познакомимся с понятиями нормализации/денормализации и схемой «звезда». На протяжении всей книги мы будем описывать новые концепции на примерах, но в первой главе это будет наиболее заметно.

Пристегните ремни! Пришло время узнать все тайны о моделировании данных.

Работа с одной таблицей

Если вы используете Excel и сводные таблицы для анализа данных, велика вероятность, что вы загружаете информацию посредством запроса из какого-то источника – обычно из базы данных. После этого строите сводную таблицу и приступаете к анализу. Разумеется, при этом вы вынуждены мириться с некоторыми ограничениями Excel, главным из которых является лимит на количество строк в таблице, равный одному миллиону. Больше записей просто не поместится на рабочем листе. Честно говоря, в начале своего пути мы не рассматривали эту особенность как серьезный сдерживающий фактор. В самом деле, зачем кому-то может понадобиться загружать в Excel миллион строк, если можно воспользоваться базой данных? Причина может быть в том, что работа с Excel не требует от пользователя знаний в области моделирования данных, а с базой данных – требует.

Так или иначе, эта особенность Excel является существенным ограничением. В базе данных Contoso, которую мы используем в примерах, таблица продаж содержит 12 млн записей. Так что мы не можем просто взять и поместить их все на лист Excel. Но эта проблема легко решается. Вместо того чтобы загружать данные целиком, вы можете сгруппировать их, чтобы сократить количество строк. Если, допустим, вам необходимо проанализировать продажи в разрезе категорий и подкатегорий товаров, вы можете наложить соответствующие группировки, что существенно снизит объем загружаемой информации.

К примеру, разделение исходной таблицы из 12 млн строк на группы по производителю, бренду, категории и подкатегории с сохранением детализации продаж до дня позволило нам сократить количество записей до 63 984, что вполне приемлемо для загрузки на лист Excel. Написание запроса для выполнения подобной группировки – это задача для отдела ИТ или подходящего редактора запросов, если вы, конечно, не знаете язык SQL. Выполнив получившийся запрос, вы можете приступать к анализу. На рис. 1.1 можно видеть первые несколько строк после импорта данных в Excel.

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

Рис. 1.1. Данные о продажах, сгруппированные для облегчения анализа

После загрузки таблицы в Excel вы можете наконец почувствовать себя как дома, создать сводную таблицу и приступать к анализу. На рис. 1.2 мы представили продажи по производителям для выбранной категории посредством обычной сводной таблицы и среза.

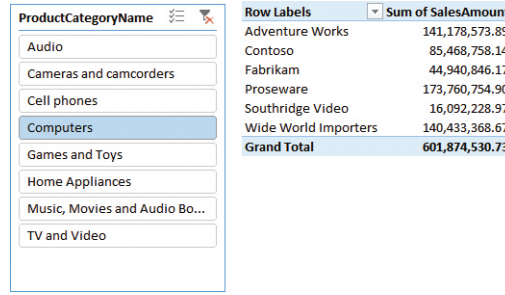


Рис. 1.2. На основании данных в Excel легко можно создать сводную таблицу

Верите вы или нет, но только что вы построили свою первую модель данных. Да, она состоит всего из одной таблицы, но тем не менее это модель данных. А значит, вы можете исследовать ее аналитический потенциал и искать способы для его повышения. У представленной модели есть одно серьезное ограничение – она содержит меньше строк, чем исходная таблица.

Будучи новичком в Excel, вы могли бы подумать, что лимит в миллион строк распространяется только на исходные данные, которые вы загружаете для дальнейшего анализа. И хотя это верно, важно также понимать, что данное ограничение автоматически переносится и на модель данных, что негативно сказывается на аналитическом потенциале отчетов. Фактически, для того чтобы сократить количество строк, вы вынуждены были производить группировку на уровне исходных данных и извлекать продажи, сгруппированные по определенным столбцам.

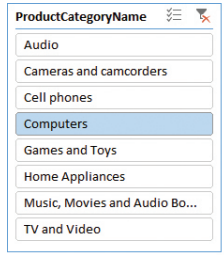
Таким образом, вы косвенно ограничили свои аналитические возможности. К примеру, вы не сможете провести аналитику по цвету товаров на основании полученной таблицы, поскольку информация об этой характеристике просто отсутствует. Добавить столбец к таблице – не проблема. Проблема в том, что при добавлении столбцов будет автоматически увеличиваться размер таблицы как в ширину (в количестве столбцов), так и в длину (в количестве строк). На практике одна строка для отдельной категории – например, аудиотехники (Audio) – превратится в несколько записей, каждая из которых будет содержать свой цвет для этой категории.

А если вы не сможете заранее решить, какие столбцы вам пригодятся для выполнения срезов, то вам придется загружать все 12 млн строк, а с таким объемом Excel не справится. Именно это мы имели в виду, когда говорили, что потенциал Excel в отношении моделирования данных невелик. Ограничение на количество импортируемых строк делает невозможным проведение анализа больших объемов данных.

Здесь вам на помощь приходит Power Pivot. Используя Power Pivot, вы не будете ограничены миллионом строк. Фактически количество записей, загружаемых в таблицу Power Pivot, ничем не ограничено. А значит, вы легко сможете импортировать в свою модель все продажи и проводить на их основании более глубокий анализ.

Примечание. Power Pivot доступен в Excel с версии 2010 в качестве внешней надстройки, а начиная с Excel 2013 включен в основной пакет. В Excel 2016 и следующих версиях Microsoft ввела новый термин для описания моделей Power Pivot: *модель данных Excel* (Excel Data Model). Однако термин Power Pivot по-прежнему широко используется.

Располагая полной информацией о продажах в одной таблице, вы можете проводить более детализированный анализ. К примеру, на рис. 1.3 вы видите сводную таблицу, построенную на основе модели данных Power Pivot со всеми загруженными столбцами. Теперь вы можете осуществлять срезы по категории товара, цвету и году, поскольку вся эта информация находится в модели. Чем больше столбцов, тем выше аналитический потенциал.



Sum of SalesAmount	Column Labels			
Row Labels	2007	2008	2009	Grand Total
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**

Рис. 1.3. Если в модель данных загружены все столбцы, можно строить более интересные сводные таблицы

Этого примера достаточно, чтобы усвоить первый урок, касающийся модели данных: *размер имеет значение, поскольку он напрямую связан с гранулярностью*. Но что такое *гранулярность*? *Гранулярность* (granularity) – одна из важнейших концепций, описываемых в этой книге, и мы постараемся познакомить вас с ней как можно раньше. Далее в книге мы углубимся в изучение этой концепции, а сейчас позвольте дать простое описание термина гранулярность. В первом наборе данных вы сгруппировали информацию по категории и подкатегории, пожертвовав детальными данными ради уменьшения размера таблицы. Говоря техническим языком, вы установили гранулярность таблицы на уровне категории и подкатегории. Можете думать о *гранулярности* как об уровне детализации данных. Чем выше гранулярность, тем более детализированная информация будет доступна для анализа. В последнем рассмотренном наборе данных, загруженном в Power Pivot, гранулярность установлена на уровне товара (на самом деле она даже выше – на уровне каждой отдельной продажи), тогда как в предыдущем примере была на уровне категории и подкатегории. Возможности для детального анализа напрямую связаны с количеством доступных столбцов в таблице, а значит, с ее гранулярностью. Вы уже знаете, что увеличение количества столбцов непременно ведет к увеличению количества строк.

Выбрать правильный уровень гранулярности всегда непросто. При неверном выборе практически невозможно будет извлечь нужную информацию при помощи формул. У вас либо попросту не будет этих данных в таблице (как в примере с отсутствующим цветом товаров), либо эти данные будут разбросаны по всему набору. При этом неправильно будет говорить, что более высокий уровень гранулярности таблицы – это всегда хорошо. Нужно стремиться, чтобы гранулярность была установлена на оптимальном уровне с учетом ваших требований к дальнейшему анализу данных.

Мы уже рассматривали пример с потерянными данными. А что значит выражение «данные разбросаны по всему набору»? Проиллюстрировать такое поведение информации несколько сложнее. Представьте, к примеру, что вам необходимо получить средний годовой доход клиентов, покупаю-

щих определенный набор товаров. Такая информация в таблице присутствует – у нас ведь есть все сведения о наших покупателях. На рис. 1.4 показан фрагмент таблицы с нужными нам столбцами (необходимо открыть окно Power Pivot, чтобы увидеть содержимое таблицы).

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

Рис. 1.4. Информация о покупателях и товарах содержится в одной таблице

В каждой строке таблицы продаж в отдельном столбце указывается величина годового дохода клиента, купившего этот товар. В попытке вычислить средний годовой доход покупателя мы можем попробовать создать меру при помощи следующего кода на DAX:

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

Созданная мера отлично работает, и вы можете использовать ее в сводной таблице, как это показано на рис. 1.5. Здесь мы видим средний годовой доход покупателей бытовой техники (Home Appliances) разных брендов.

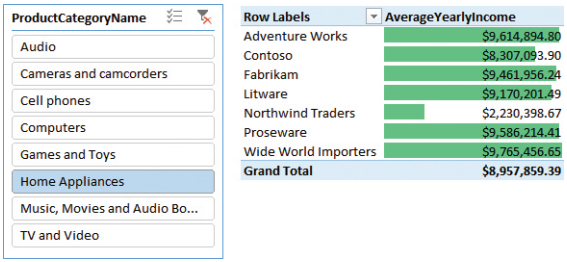


Рис. 1.5. Анализ среднего годового дохода покупателей бытовой техники

Отчет выглядит замечательно, но, к сожалению, цифры в нем не соответствуют действительности – они чересчур завышены. Фактически вы вычисляете среднее значение по таблице продаж с гранулярностью, установленной на уровне каждой продажи. Иными словами, в этой таблице содержатся строки для каждой продажи, а значит, покупатели в ней будут повторяться. Так, если покупатель приобрел три товара в разные дни, при подсчете среднего значения годовой доход для него будет учтен трижды, что приведет к ошибочным результатам.

Вы могли бы сказать, что таким образом получили средневзвешенную величину годового дохода. Но это не совсем так. Для того чтобы рассчитать средневзвешенное, нам необходимо было бы задать вес для каждой составляющей, а брать в качестве веса количество покупок было бы неправильно. Более логично было бы определить как вес количество купленных товаров, сумму покупки или еще какой-то значимый показатель. Кроме того, в данном примере мы планировали вычислять обычное среднее значение годового дохода покупателей, и созданная мера нам в этом ничуть не помогла.

И хотя это не так просто заметить, здесь мы также столкнулись с проблемой некорректно выбранной гранулярности. Получается, что информация, которая нам нужна, доступна, но не привязана к конкретному покупателю, а вместо этого разбросана по таблице продаж, что значительно затрудняет вычисления. Чтобы получить корректный результат, необходимо изменить гранулярность до уровня покупателя – либо путем повторной загрузки таблицы, либо воспользовавшись сложной формулой на языке DAX.

Если вы решите пойти по пути DAX, можно для вычисления среднего годового дохода воспользоваться следующей формулой, довольно сложной для понимания:

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

В этой не самой простой формуле мы сначала агрегируем продажи на уровне (гранулярности) покупателя, после чего применяем к результирующей таблице, в которой каждый покупатель появляется только один раз, функцию *AVERAGEX*. В примере мы применяем функцию *SUMMARIZE* для предварительной агрегации на уровне покупателя во временной таблице, а затем вычисляем среднее значение по `YearlyIncome`. Как видно по рис. 1.6, итоги правильного расчета среднего годового дохода сильно отличаются от наших прежних расчетов.

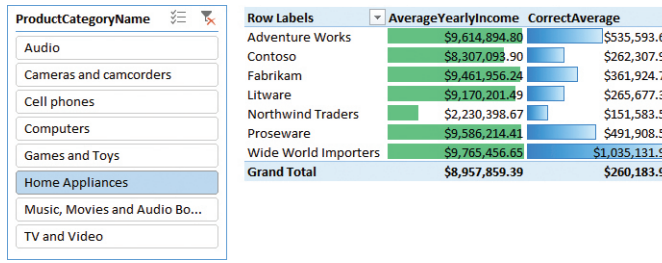


Рис. 1.6. При взгляде на результаты вычислений видно, как далеки мы были от истины