**Глава 20. Power Query в помощь**

Это продолжение перевода книги Роб Колли. Формулы DAX для Power Pivot. Главы не являются независимыми, поэтому рекомендую начать сначала.

[Предыдущая глава](http://baguzin.ru/wp/?p=20328) [Содержание](http://baguzin.ru/wp/?p=19780#content) Следующая глава

Базы данных являются превосходным источником данных. Люди вкладывают усилия в создание баз данных, чтобы данные были чистыми, полными и оптимизированными до того, как они будут использоваться. Если у вас есть DBA (администратор базы данных), любите его, обнимайте, благодарите. А затем попросите очистить и переформатировать данные для ваших целей))

Но за пределами базы данных, реальные данные... грязные. К сожалению, б***о***льшая часть мира не получает свои данные из баз данных. Б***о***льшая часть мира по уши в текстовых файлах и файлах Excel. В конце концов, кнопка *Экспорт в Excel* является 3-й наиболее распространенной кнопкой во всех приложениях BI (после Ok и Cancel, конечно), и эти кнопки производят такие файлы. Реальные данные часто разделяются на множество файлов, когда вы хотели бы анализировать их все вместе. Реальные данные помещаются в столбцы, когда они нужны в строках. Реальные данные хранятся в двумерной таблице, когда вам нужны отдельные таблицы данных и поиска. Эта глава не для тех, кто живет на суше. Если вы живете на земле баз данных, и у вас есть дружественный DBA, перейдите к следующей главе!



Рис. 20.1. Большинство людей, к сожалению, живут за пределами сказочной страны. Они живут в грязном мире реальных данных

### Power Query: очистка грязных данных

Power Query позволяет очищать, форматировать и преобразовывать данные, сохраняя при этом возможность обновить их в один клик:

* Это идеальное дополнение к Power Pivot. Последний имеет очень ограниченные возможности для очистки, форматирования и преобразования данных.
* Очень легко сделать первые шаги, при этом инструмент обладает невероятной глубиной. Интерфейс интуитивно понятен и управляется обычной лентой. Но вы не ограничены кнопками на ленте – под поверхностью лежит мощный язык программирования.

Power Query может так много, что ему посвящена целая книга [M is for (Data) Monkey](http://ppvt.pro/DataMonkey) Кена Пульса и Мигеля Эскобара (я уже приобрел ее, и планирую перевести. – *прим. Багузина*).

Power Query (в Excel и Power BI Desktop) постоянно обновляется, так что ваш интерфейс и функционал может отличаться от того, что описано здесь. В этой главе авторы использовали скриншоты Excel 2013 (а я – Excel 2019). Примеры файлов доступны в формате Excel 2013 и PBIX (Power BI Desktop).

### Создание одной таблицы в PowerPivot из нескольких файлов из Интернета

Вы хотите создать таблицу поиска продуктов. У вас три завода, и каждый производит свой набор продуктов. У вас три файла, и вы хотели бы объединить их в одну таблицу перед загрузкой в Power Pivot. Кроме того, вы хотели бы, чтобы эта процедура была автоматизирована, и в следующий раз, когда вы получите новые файлы, вам не пришлось бы повторять всё с начала. И чтобы еще немного усложнить, скажем, что исходные файлы публикуются на веб-сайте (рис. 20.2; файл Excel для этого примера будет сформирован в процессе выполнения шагов; его финальный вид представлен в PowerQuery\_01\_Append.xlsx).

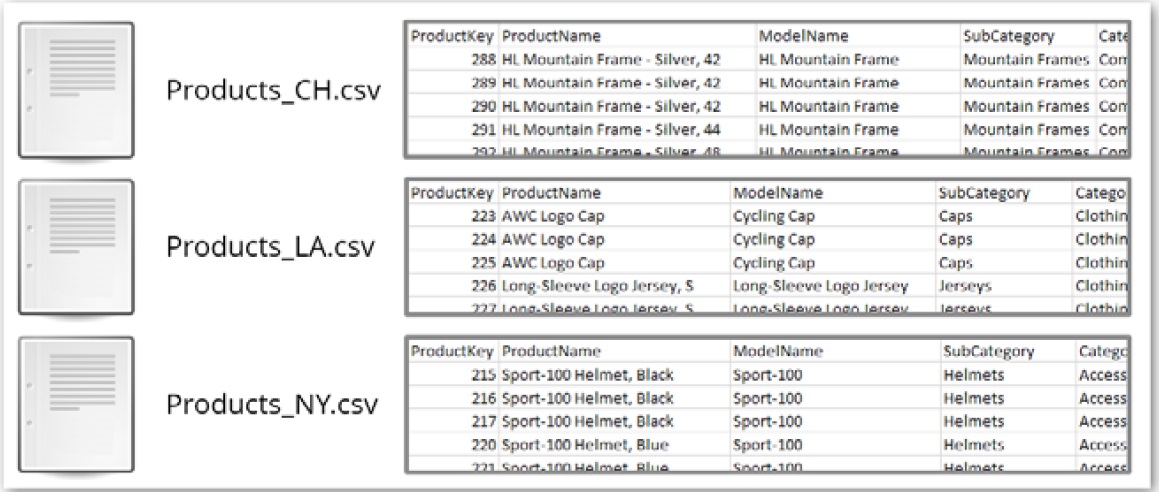


Рис. 20.2. Нужно объединить три CSV-файла одинаковой структуры, расположенные на сайте

Вам не нужно загружать файлы CSV на свой компьютер. Вы можете оставить файлы там, где они есть, и построить полностью автоматическую систему, которая не требует ручного вмешательства для обновления.

После загрузки и установки Power Query на ленте Excel 2013 появится вкладка Power Query (рис. 20.3). В версиях Excel 2016 и 2019 функции Power Query доступны на вкладке *Данные* в области *Получить и преобразовать данные* (рис. 20.4).

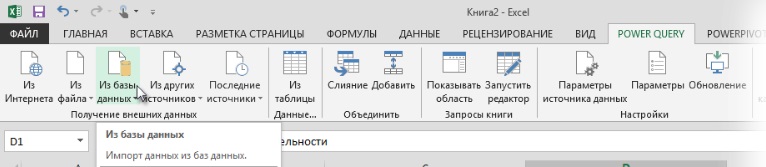


Рис. 20.3. Лента Power Query в Excel 2013

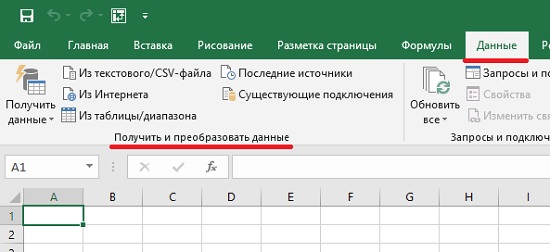


Рис. 20.4. Функции Power Query доступны в Excel 2016 или 2019 на вкладке *Данные*

Кликните кнопку *Из Интернета*, и в открывшемся окне наберите адрес файла: [www.powerpivotpro.com/wp-content/uploads/2015/10/Products\_CH.csv](http://www.powerpivotpro.com/wp-content/uploads/2015/10/Products_CH.csv) (рис. 20.5).

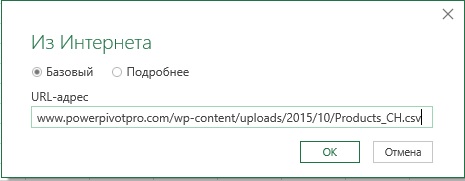


Рис. 20.5. Импорт из Интернета

Откроется окно загрузки Power Query (рис. 20.6). Кликните *Загрузить*.

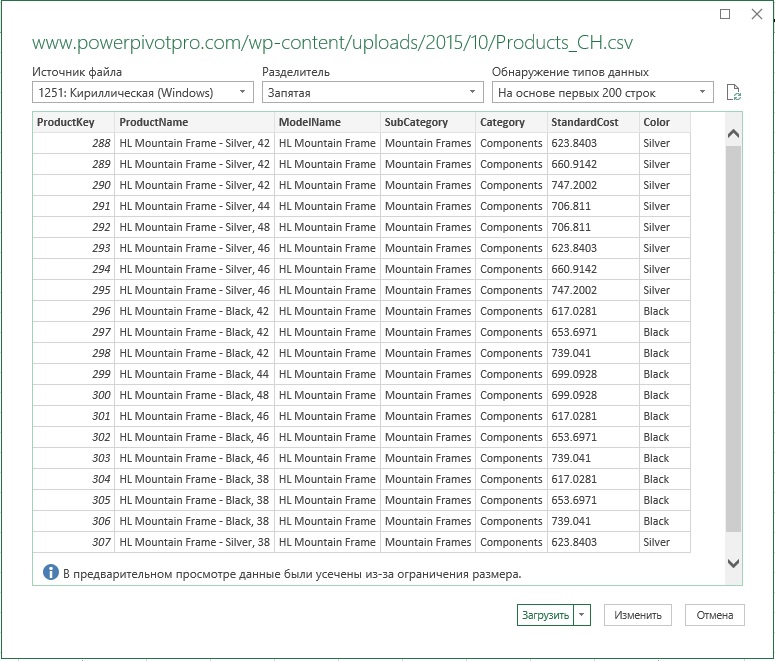


Рис. 20.6. Окно загрузки *Power Query*

Данные загрузятся на лист Excel в Таблицу. На ленте появятся две вкладки: *Работа с таблицами* и *Работа с запросами* (рис. 20.7).

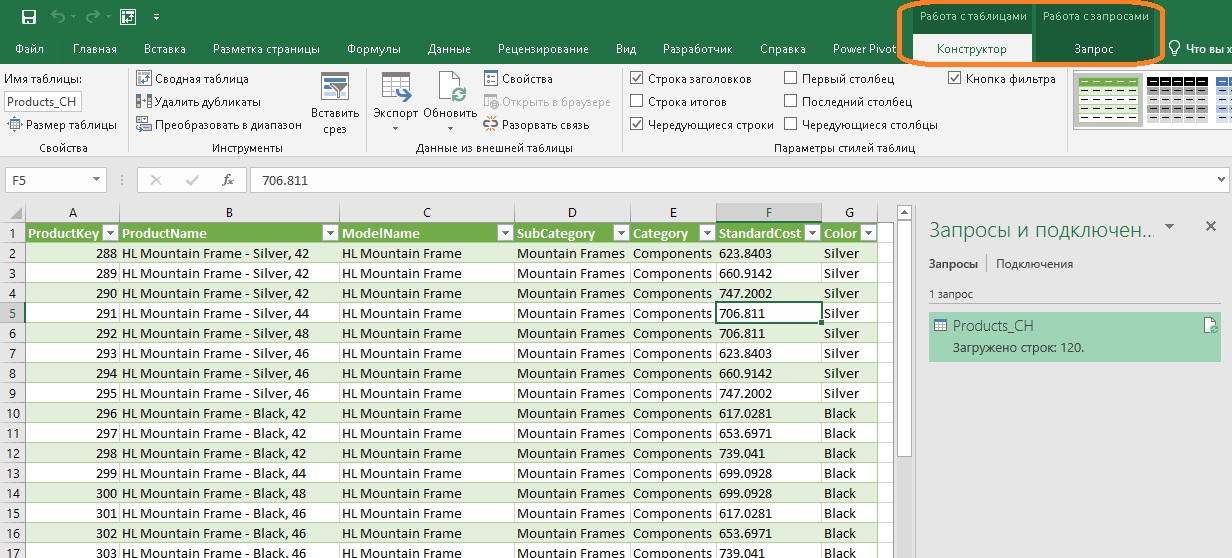


Рис. 20.7. Данные из Интернета загрузились на лист Excel в Таблицу

Поскольку название завода присутствует только в имени файла, добавим столбец-метку принадлежности продукции к Чикагскому заводу. Для этого пройдите по меню *Запрос* –> *Изменить* или дважды кликните на таблице Products\_CH, расположенной в области *Запросы и подключения* (см. правую часть рис. 20.7), или наведите курсор на таблице Products\_CH, и в всплывающей подсказке, нажмите *изменить* (рис. 20.8).

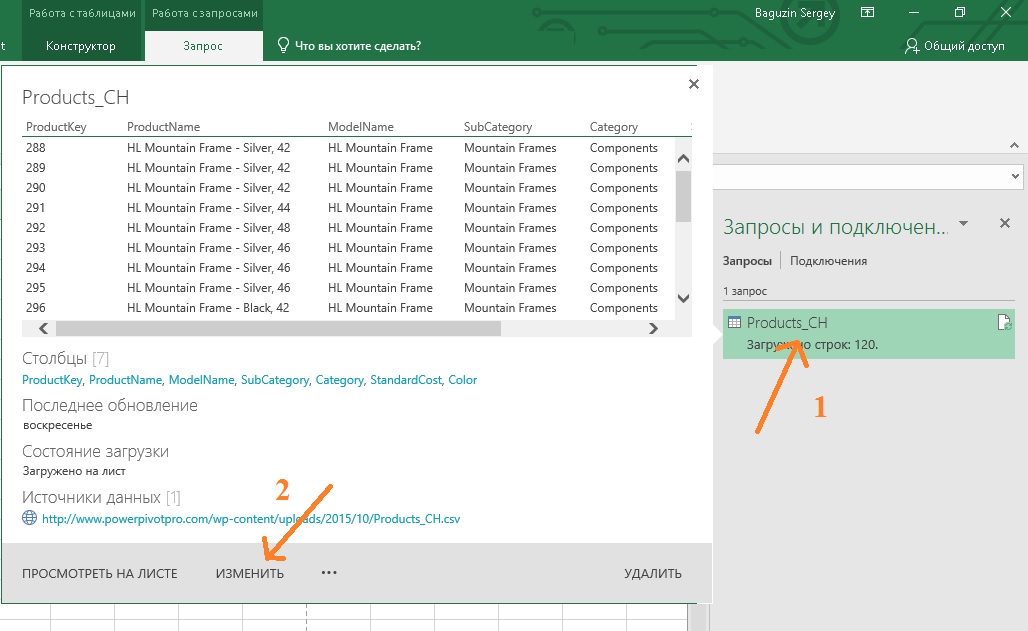


Рис. 20.8. Всплывающая подсказка при наведении курсора на таблицу Products\_CH

Откроется окно *Редактор* *Power Query.* Перейдите на вкладку *Добавление столбца* и нажмите *Настраиваемый столбец.* Укажите имя столбца и введите статическую формулу (="CH"):

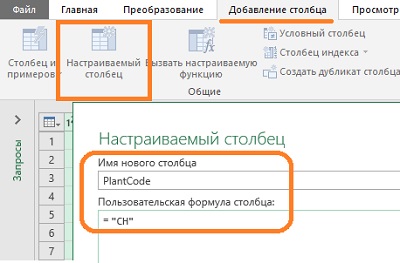


Рис. 20.9. Добавление пользовательского столбца

Обратите внимание (рис. 20.10):

1. Добавлен пользовательский столбец.
2. В области *Примененные шаги* отражается новый шаг – *Добавлен пользовательский объект*. Заметьте, что в этой области отражаются все шаги Power Query, выполненные ранее. Power Query записывает и воспроизводит все шаги преобразования данных. В будущем все шаги будут повторены при нажатии кнопки Обновить.
3. В строке формул показана формула на языке M для этого шага (если у вас не отражается строка формул, перейдите на вкладку *Просмотр*, и поставьте соответствующую галочку. Строка формул дает вам представление о мире M. Однако вам не обязательно изучать и понимать M, чтобы использовать Power Query. Как правило, достаточно кнопок на ленте.
4. Переименуйте запрос и на вкладке *Главная* нажмите кнопку *Закрыть и загрузить*.

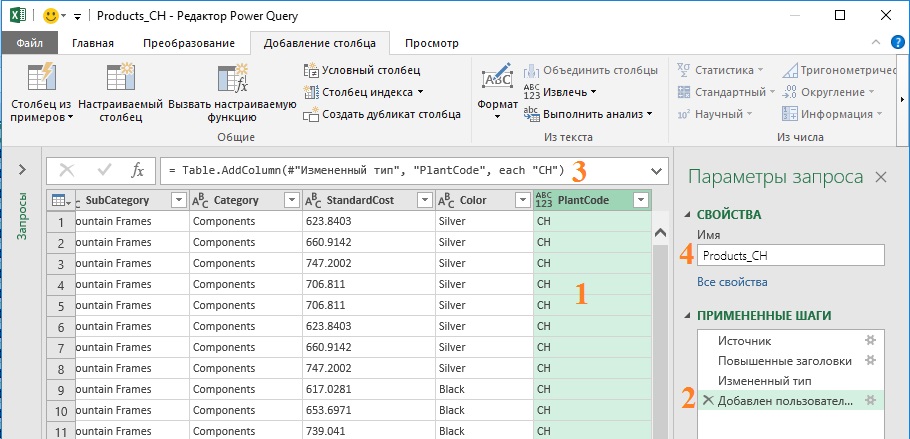


Рис. 20.10. Новый столбец добавлен в таблицу

#### Загрузка данных в Power Pivot

Итак, у вас есть таблица в Excel, но это – не совсем то, что вам нужно. Вы хотели поместить данные в Power Pivot. Кликните правой кнопкой мыши на запрос, и выберите *Загрузить в…* (рис. 20.11)

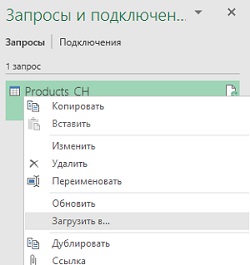


Рис. 20.11. Загрузка запроса в Power Pivot

В открывшемся окне *Импорт данных* выберите *Только создать подключение* и *Добавить эти данные в модель данных* (рис. 20.12).

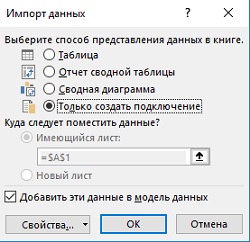


Рис. 20.12. Импорт данных в Power Pivot

Вышеприведенная комбинация загружает данные в Power Pivot, но не в Excel. После того, как вы нажмете Ok, появится предупреждение об этом. Нажмите *Да*, данные загрузятся в Power Pivot и удаляться с листа Excel. Чтобы увидеть таблицу, перейдите в Excel на вкладку *Power Pivot*, и кликните на кнопке *Управление*. Откроется окно Power Pivot с нашими данными.

#### Подключение ко второму и третьему CSV-файлам

В Excel перейдите на вкладку *Данные* в области *Получить и преобразовать данные*. Кликните кнопку *Из Интернета*, и в открывшемся окне наберите адрес второго файла: [www.powerpivotpro.com/wp-content/uploads/2015/10/Products\_LA.csv](http://www.powerpivotpro.com/wp-content/uploads/2015/10/Products_LA.csv). Повторите процедуры, описанные на рис. 20.4–20.9. Только теперь вставьте пользовательский столбец с формулой ="LA", и назовите запрос Products\_LA.

После этого загрузите третий файл (завод NY). Покажу, как выполнить подключение файла к модели данных Power Pivot (без промежуточной загрузки в Excel).

1. Данные –> Из Интернета
2. Адрес файла [www.powerpivotpro.com/wp-content/uploads/2015/10/Products\_NY.csv](http://www.powerpivotpro.com/wp-content/uploads/2015/10/Products_NY.csv)
3. Ok
4. В окне предварительного просмотра Power Query сразу жмем *Изменить*
5. В окне редактора Power Query перейти на вкладку *Добавление столбца*
6. В окне *Настраиваемый столбце* указать название PlantCode и формулу ="NY", нажать Ok
7. В окне редактора Power Query перейти на вкладку *Главная*, нажать кнопку *Закрыть и загрузить*
8. Выбрать из двух опций вторую – *Закрыть и загрузить в…*
9. В окне *Импорт данных* выбрать *Только создать подключение* и *Добавить эти данные в модель данных* (см. рис. 20.12).
10. Нажмите Ok.

Теперь у вас в книге Excel определены три запроса, а в Power Pivot есть три листа с данными.

#### Объединение таблиц

В Excel в области *Запросы и подключения* кликните правой кнопкой мыши на любом запросе, и выберите *Добавить*:

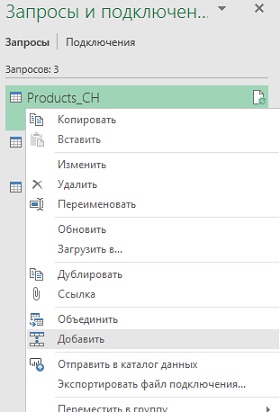


Рис. 20.13. Команда *Добавить*

В открывшемся окне *Добавление* переключитесь на опцию *Три таблицы или больше*, и последовательно добавьте таблицы в правое окно (рис. 20.14). Нажмите Ok.

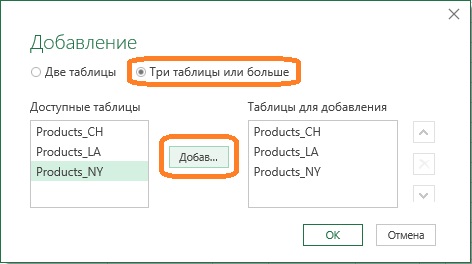


Рис. 20.14. Окно *Добавление*

Редактор Power Query запросит сведения о конфиденциальности данных (рис. 20.15). Нажмите *Продолжить*.

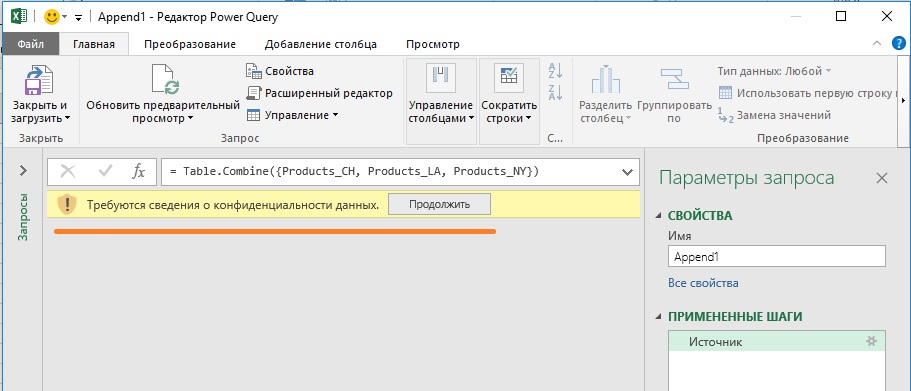


Рис. 20.15. Запрос сведений о конфиденциальности данных

Поскольку эти CSV-файлы находятся на общедоступном сайте, вы можете выбрать *Общий* (рис. 20.16). Нажмите *Сохранить*.

В настройках параметров Power Query можно игнорировать уровни конфиденциальности, подавляя тем самым появление этого предупреждения. Однако мы не рекомендуем делать это. Подробнее см. официальный комментарий Microsoft – [Privacy levels Power Query](https://support.office.com/en-us/article/Privacy-levels-Power-Query-cc3ede4d-359e-4b28-bc72-9bee7900b540)).

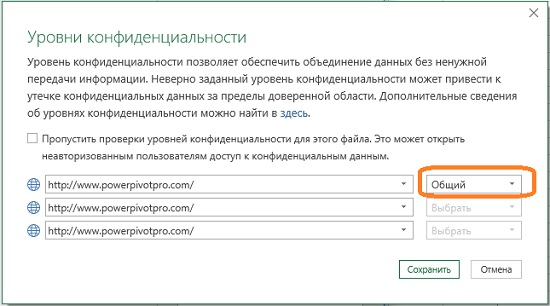


Рис. 20.16. Выбор уровня конфиденциальности

Power Query дал имя новому (объединенному) запросу по умолчанию. Дайте запросу говорящее имя Products. Обратите внимание на формулу в строке формул (рис. 20.17).

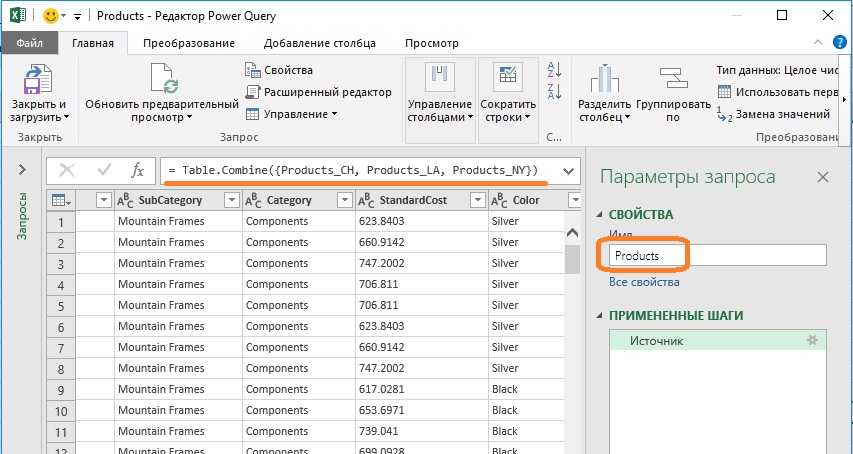


Рис. 20.17. Объединенный запрос

В окне редактора Power Query перейдите на вкладку *Главная*, нажмите кнопку *Закрыть и загрузить.* Выберите опцию *Закрыть и загрузить в…* В окне *Импорт данных* выберите *Только создать подключение* и *Добавить эти данные в модель данных* (см. рис. 20.12). Нажмите Ok. Теперь у вас в Power Pivot есть четыре листа (таблицы): три исходных CSV-запроса плюс результат объединения (рис. 20.18).

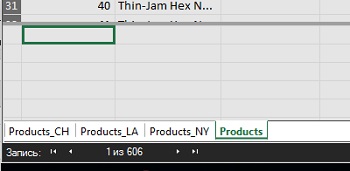


Рис. 20.18. Листы Power Pivot

На самом деле, вам нужна только последняя объединенная таблица. Вернитесь к трем исходным запросам, последовательно кликните на каждом из них правой кнопкой мыши, выберите *Загрузить в…* и снимите флажок *Добавить эти данные в модель данных* (рис. 20.19). Нажмите Ok. Появится предупреждение о потере данных. Не пугайтесь. Подтвердите.

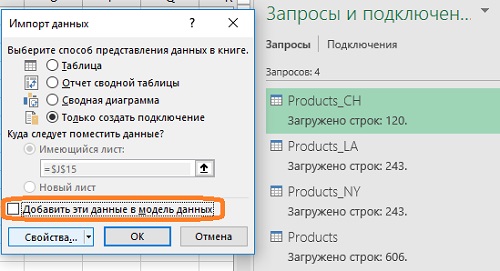


Рис. 20.19. Удаление исходных (ставших ненужными) таблиц из Power Pivot

Обратите внимание: теперь в Power Pivot есть только одна вкладка. Она содержит:

* 120 + 243 + 243 = 606 строк из всех трех файлов, и
* столбец PlantCode (которого нет ни в одном из CSV-файлов) присутствует в результирующей таблице и принимает одно из трех значений: CH, LA, NY.

Диалоговое окно *Импорт данных* позволяет сразу выбрать опцию *Только создать подключение* и не устанавливать флажок *Добавить эти данные в модель данных*. Данные никуда не будут загружены, а запрос будет помечен, как *Только подключение*. На первый взгляд это может показаться странным, но, как вы только что увидели, это очень полезно если исходные запросы будут далее объединены. Также обратите внимание, что параметры Power Query позволяют изменить загрузку по умолчанию и включить опцию *Быстрая загрузка данных*. Чтобы открыть окно Параметры запроса, пройдите в Excel по меню Данные –> Получить данные –> Параметры запроса (рис. 20.20).

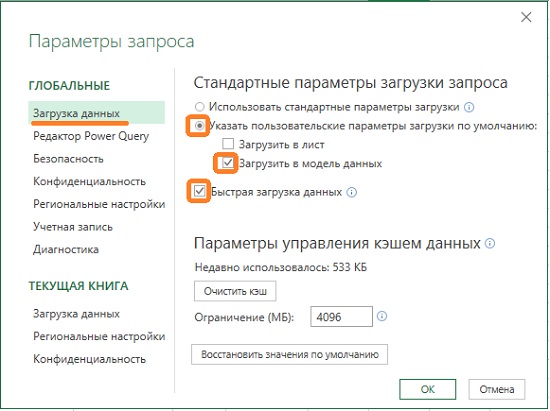


Рис. 20.20. Параметры запроса

Стандартное поведение приводит к загрузке запроса на новый лист при загрузке одного запроса и загрузке нескольких запросов в модель данных в случае загрузки нескольких запросов за один раз. Это поведение можно переопределить для конкретного запроса, используя диалоговое окно *Загрузить в...* Включение быстрой загрузки данных ускорит загрузку запросов, но Excel может не отвечать в течение долгого времени.

#### Тестирование обновления

Перейдите в окно Power Pivot на закладку Products, и нажимаете *Обновить* (даже не *Обновить все*, рис. 20.21). Все три «дочерних» запроса, извлекут три свежих CSV-файла из Интернета, и обновят данные, содержащиеся в них (рис. 20.22).

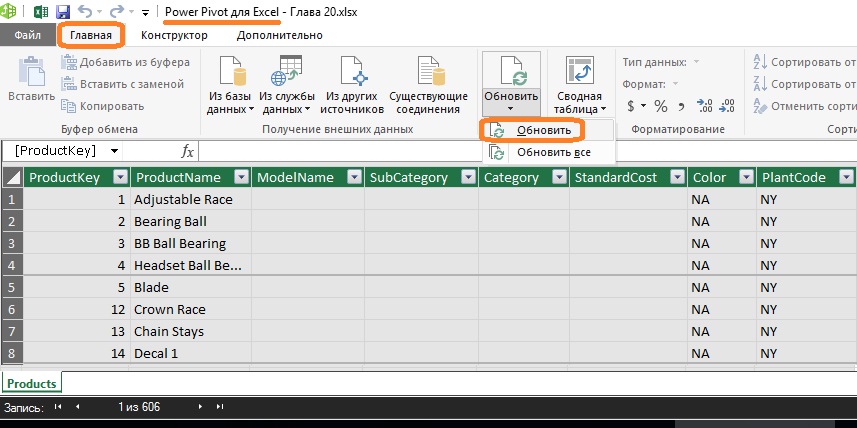


Рис. 20.21. Команда *Обновить*

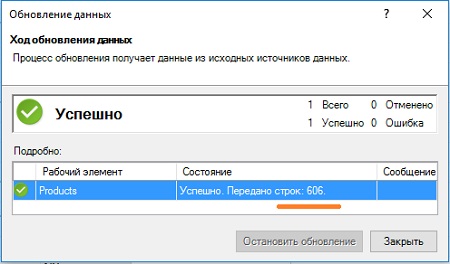


Рис. 20.22. Обновление прошло успешно; все 606 строк из трех файлов обновлены

### Объединение нескольких файлов из папки в одну таблицу

Создайте новую книгу Excel. Пройдите по меню *Данные* –> *Получить данные* –> *Из файла* –> *Из папки* (рис. 20.23). Затем выберите путь к папке, в которой находятся файлы, и нажмите Оk (рис. 20.24). И здесь файл Excel будет сформирован в процессе выполнения шагов; его финальный вид представлен в PowerQuery\_02\_FromFolder.xlsx.

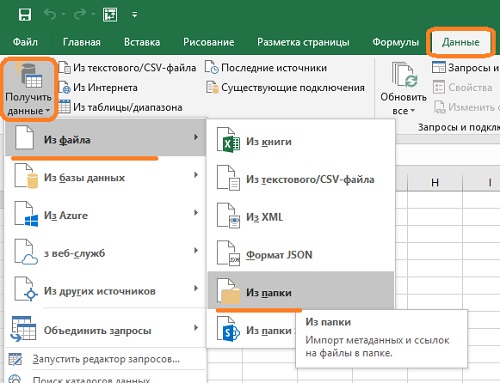


Рис. 20.23. Импорт всех файлов из папки

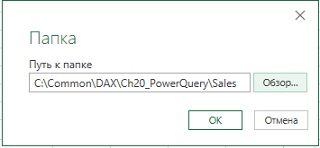


Рис. 20.24. Путь к папке

Power Query отобразит список файлов в папке (рис. 20.25).

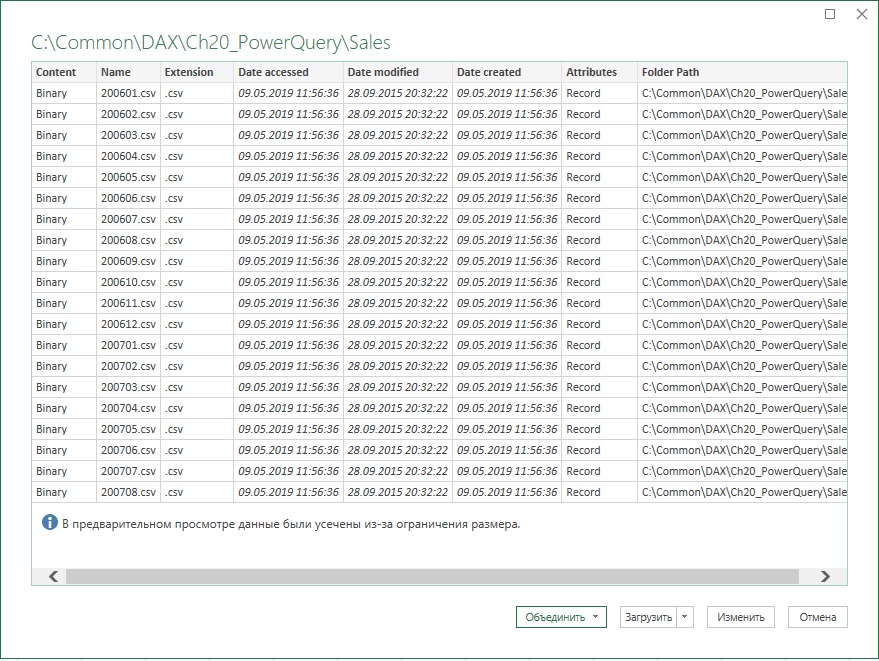


Рис. 20.25. Список файлов в выбранной папке

Если вы уверены в своих данных и полагаетесь на стандартные настройки Power Query (которые, надо отметить, работают довольно хорошо, и распознают типы данных), то можете кликнуть на правый треугольник кнопки *Объединить* и выбрать *Объединить и загрузить в…* После обработки запроса откроется окно *Объединить файлы*. Кликните Ok. Появится окно импорта (как на рис. 20.12). Выберите *Только создать подключение* и *Добавить эти данные в модель данных.* В Power Pivot на лист *Sales* загрузится 58 189 строк из всех файлов.

Однако, Power Query предоставляет также отличные возможности вручную отредактировать запрос. Для этого находясь в окне просмотра файлов (рис. 20.25) кликните *Изменить*. Откроется окно редактора Power Query (рис. 20.26). Загружено 36 файлов (1). Кликните на кнопку действия в заголовке столбца *Content* (2)*.* Это запустит выполнение нескольких шагов запроса. После обработки откроется окно *Объединить файлы*. Кликните Ok.

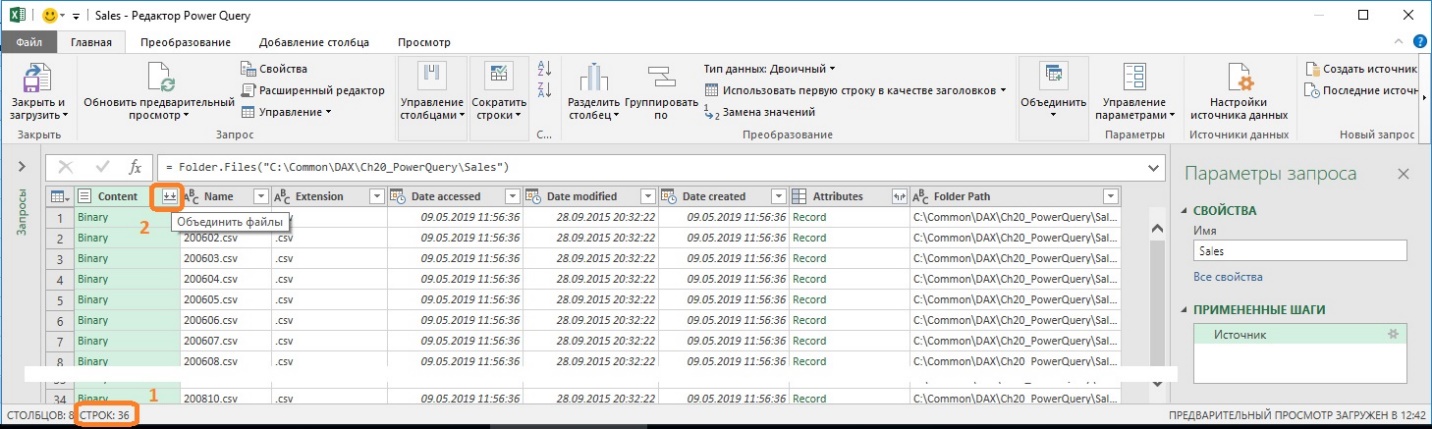


Рис. 20.26. Окно редактора Power Query

Теперь окно редактора выглядит иначе (рис. 20.27). В нем уже не перечень файлов, а их объединенное и преобразованное содержимое. Загружено более 999 строк данных. В области *Параметры запроса* отражаются выполненные шаги. В строке формул указана таблица с названиями столбцом и типами данных (текстовые, дата, числа, …)

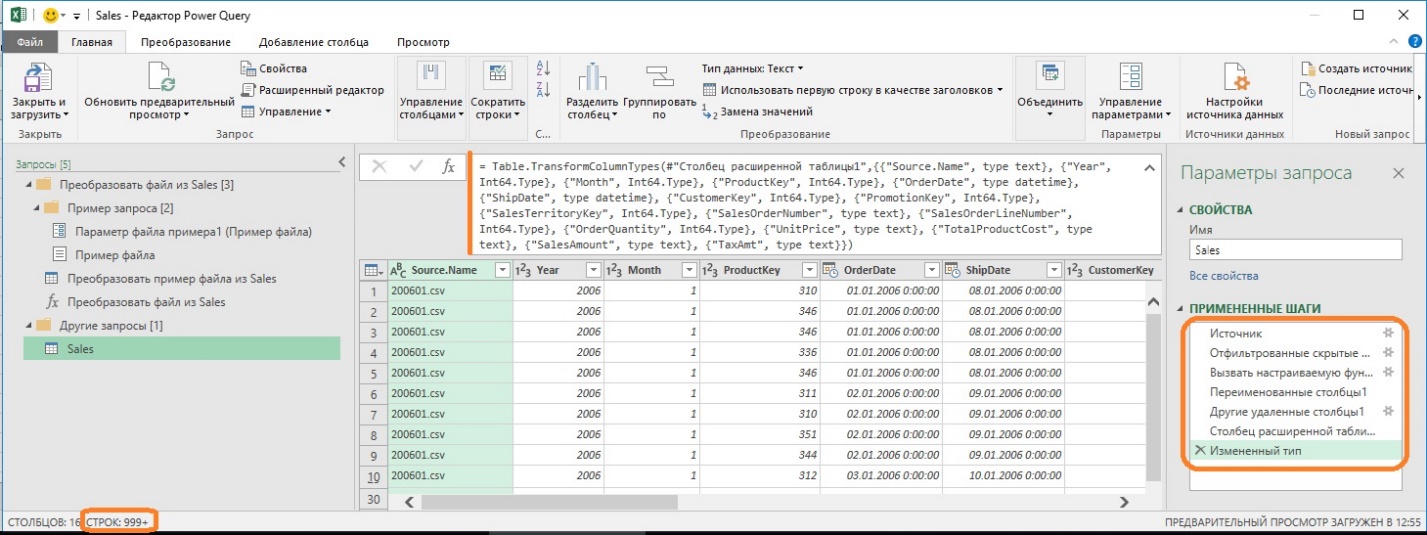


Рис. 20.27. Объединенные и преобразованные данные в окне редактора Power Query

Некоторые данные импортировались некорректно. Например, все числовые столбцы не были распознаны, и импортировались как текст (рис. 20.28). Это связано с тем, что региональные настройки на моем ПК – русские, а автор подготовил файлы в американской традиции с разделителем десятичных знаков точкой.

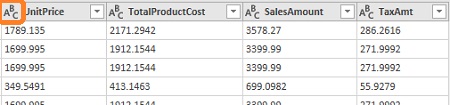


Рис. 20.28. Десятичные числа импортировались некорректно

Чтобы исправить ситуацию выберите столбец UnitPrice, кликните правой кнопкой мыши, и пройдите по меню *Тип изменения* –> *Используя локаль* (рис. 20.29).

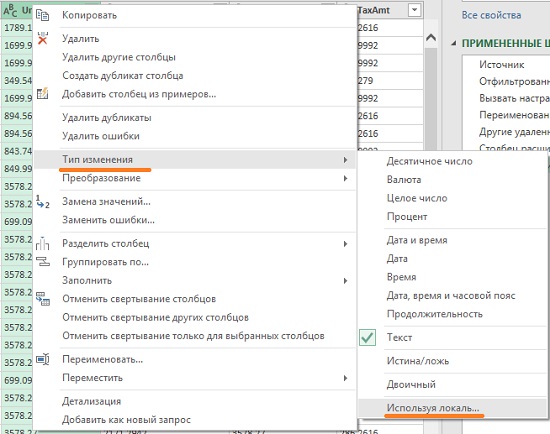


Рис. 20.29. Преобразование десятичного разделителя

В открывшемся окне *Изменение типа по локали* введите тип данных для выбранного столбца в исходном файле – *Десятичное число* (рис. 20.30). А также языковый стандарт исходного файла – *Английский (США)*.

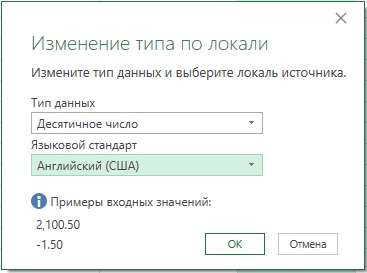


Рис. 20.30. Изменение типа данных в запросе на основании региональных настроек исходного файла (см. также справку [Microsoft](https://support.office.com/ru-ru/article/%D0%98%D0%BD%D1%82%D0%B5%D1%80%D0%BD%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F-power-query-d42b9390-1fff-413f-8120-d7df0ced20b9?ui=ru-RU&rs=ru-RU&ad=RU))

Нажмите Ok. Также суммы принято показывать с точностью до двух знаков после запятой. Выберите столбец UnitPrice, кликните правой кнопкой мыши, и пройдите по меню *Преобразование* –> *Округление* –> *Округление*. В открывшемся окне укажите 2. Нажмите Ok. Повторите операцию для еще трех числовых столбцов.

Наконец, два столбца содержат избыточное указание времени 0:00:00 (рис. 20.31). Выделите столбец, кликните правой кнопкой мыши, и пройдите по меню *Преобразование* –> *Только дата.*

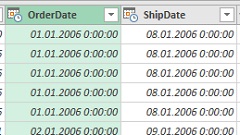


Рис. 20.31. Дата и время

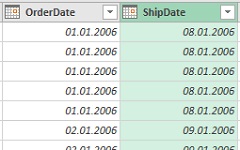


Рис. 20.32. Только дата

После окончания всех преобразование не забудьте поместить данные в Power Pivot. Для этого, как обычно, в окне редактора Power Query пройдите по меню *Главная* –> *Закрыть и загрузить* –> *Загрузить в...* Выберите *Только создать подключение* и *Добавить эти данные в модель данных.*

Если в папке Sales появятся новые файлы, достаточно в модели данных Power Pivot пройти по меню Главная –> Обновить, и в модель данных добавятся новые строки, пройдя всю предварительную обработку.

### Добавление пользовательских столбцов в таблицы поиска

В главе о [производительности](http://baguzin.ru/wp/?p=20351) вы прочитали совет, что импортированные столбцы обычно лучше, чем вычисляемые, и вы хотите добавить один или несколько пользовательских столбцов в таблицу поиска. В нашем примере, это столбец PriceBucket для сегментации продуктов по уровню цены: $ – массовые, $$ – средней категории, $$$ – хай энд, $$$$ – премиум.

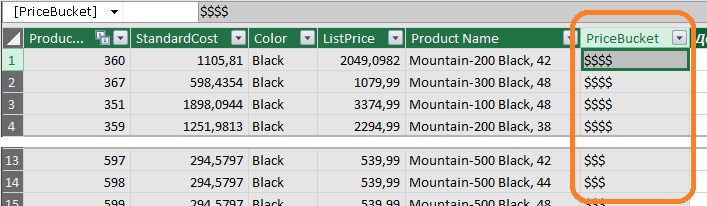


Рис. 20.33. Добавление пользовательского столбца с ценовой сегментацией

Откройте файл PowerQuery\_03\_CustomColumn.xlsx (убедитесь, что исходная база данных BookData.accdb находится в той же папке, иначе связь будет утеряна, что не позволит выполнить шаги примера). В модель данных загружена таблиц Product (см. рис. 20.33, но столбец PriceBucket пока отсутствует).

В редакторе Power Query на вкладке *Добавить столбец* имеется довольно много вариантов добавления пользовательских столбцов с использованием кнопок ленты (рис. 20.34). Конкретный набор возможностей будет зависеть от типа данных в выбранном столбце. Иногда вам потребуется изменить тип данных, чтобы получить доступ к нужным опциям.

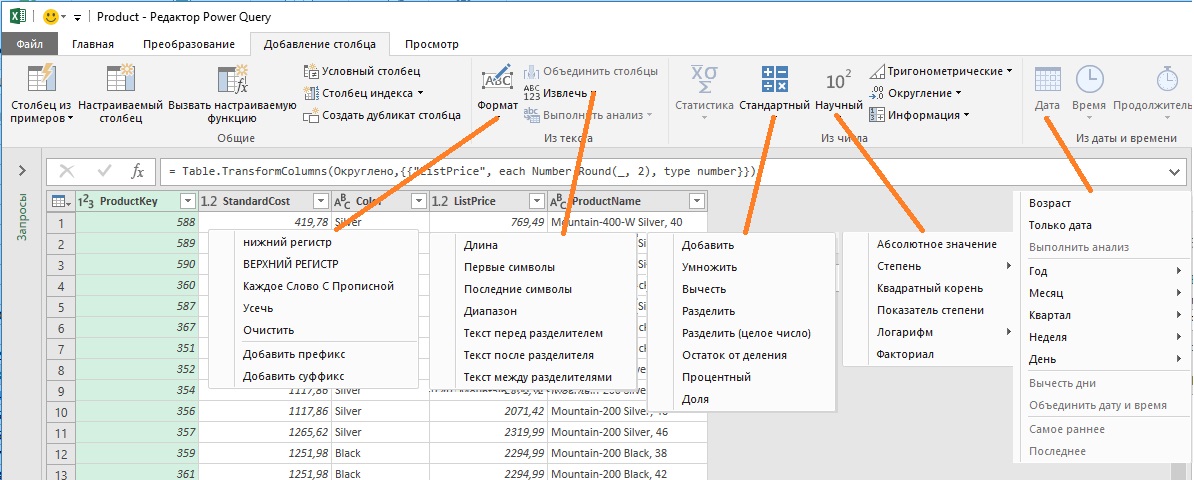


Рис. 20.34. Возможности для пользовательского столбца в Power Query

Для нашей задачи ни один из стандартных вариантов не подходит, поэтому начнем с нажатия кнопки *Настраиваемый столбец*. Введите формулу:

if [ListPrice] < 100 then "$"

else if [ListPrice] < 500 then "$$"

else if [ListPrice] < 1000 then "$$$"

else "$$$$"

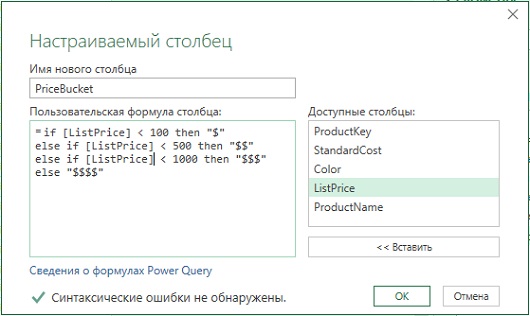


Рис. 20.35. Формула для пользовательского столбца

После загрузки таблицы в PowerPivot новый пользовательский столбец можно использовать для создания мер, или просто вывести в сводную таблицу:

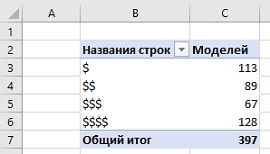


Рис. 20.36. Распределение моделей по сегментам

### Использование опции Unpivot

Даже если вы не знакомы с термином Unpivot, вы узнаете проблему, когда увидите ее.[[1]](#footnote-1) Это одна из самых распространенных и самых неприятных задач формирования данных. Предположим, у вас есть бюджетная таблица:

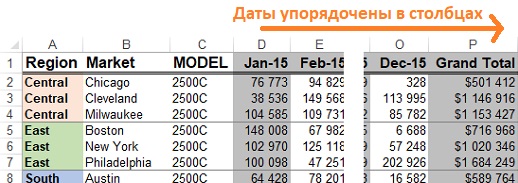


Рис. 20.37. Данные расположены в столбцах, а не строках

К сожалению, значительная часть экспорта выполняется из отчетов, которые вроде бы должны были представить финальные данные, но не справились с поставленной задачей. А поскольку наши отчеты разработаны, чтобы быть дружелюбными человеку, мы часто выводим периоды (даты) в столбцы. Но Power Pivot – не человек. Это машина для обработки данных, и она предпочитает найти даты в строках, а не в нескольких столбцах:

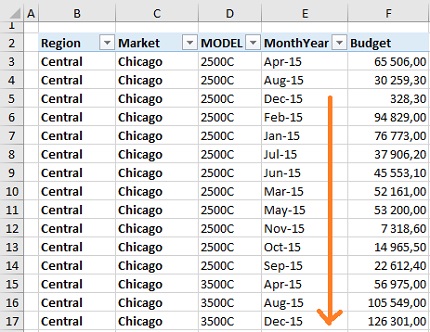


Рис. 20.38. Именно в таком виде желательно иметь данные в Power Pivot – даты упорядочены в строках

Когда в исходных данных у вас есть несколько столбцов с датами вы должны преобразовать данные, чтобы столбец с датами был один. Т.е., превратить широкую и короткую таблицу в узкую и высокую.

На этот раз исходные данные живут в Excel. Создайте новую книгу в Excel, пройдите по меню *Данные* –> *Получить данные* –> *Из файла* –> *Из книги Excel*. Выберите файл Budget.xlsx, нажмите *Импорт*. В окне *Навигатор* будет представлен список всех таблиц, доступных в этой книге (рис. 20.39). Выберите таблицу Budget, нажмите *Изменить*.

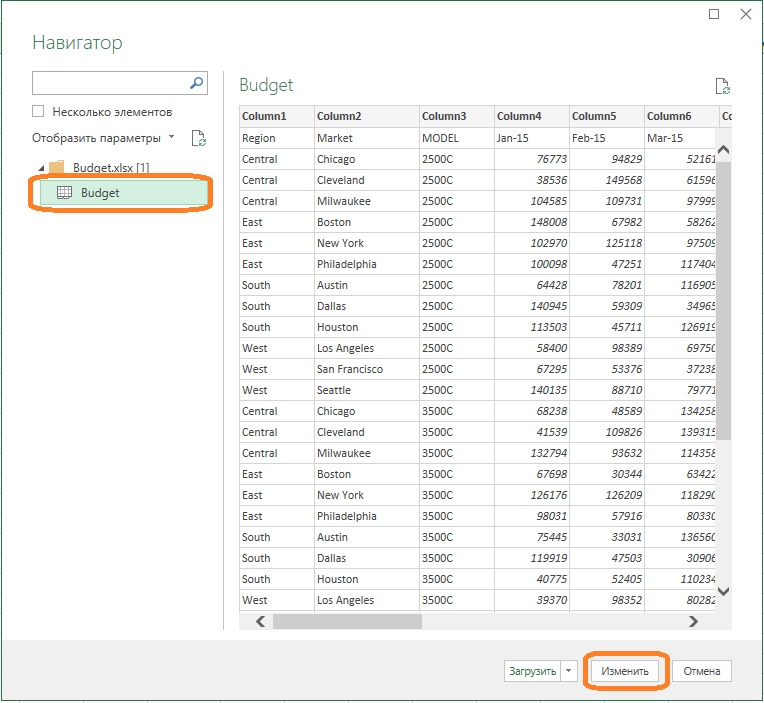


Рис. 20.39. Подключение Power Query к файлу Excel

Откроется окно редактора Power Query. Первоначально Power Query обрабатывает первую строку таблицы как данные, а не как строку заголовка. На закладке *Главная* нажмите кнопку *Использовать первую строку в качестве заголовков*. Столбец Grand Total (общий итог), последний в таблице, является избыточным, поскольку у нас уже есть данные за каждый месяц. Выберите столбец и пройдите по меню *Управление столбцами* –> *Удалить столбцы*.

Операция Unpivot (*Отменить свертывание столбцов*) находится на вкладке *Преобразование*. Вы можете выбрать все столбцы месяцев и пройти по меню *Отменить свертывание столбцов* –> *Отменить свертывание столбцов*. Но можем выбрать три столбца слева от Столбцов месяца (регион, рынок, модель), а затем пройти по меню *Отменить свертывание столбцов* –> *Отменить свертывание других столбцов* (рис. 20.40). Чтобы выбрать несколько столбцов, удерживайте клавишу Ctrl и щелкните заголовки столбцов, или выберите первый столбец, и, удерживая клавишу Shift, щелкните по последнему столбцу диапазона выделения.

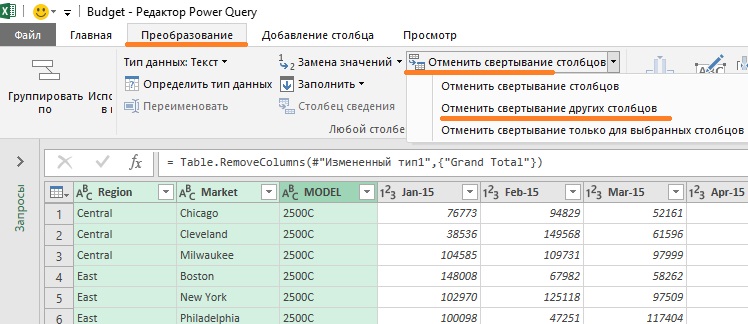


Рис. 20.40. Команда Unpivot – Отменить свертывание других столбцов

Свертывание столбцов выполнено:

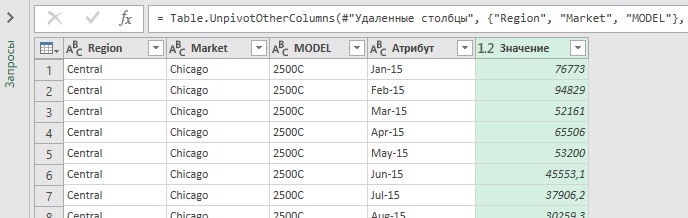


Рис. 20.41. Итог работы команды Unpivot

Преимущество второго подхода – *Отменить свертывание других столбцов* (над просто *Отменить свертывание столбцов*)в том, что в будущем могут добавиться столбцы с новыми датами. При обновлении данных второй подход отработает корректно, а вот первый – нет.

Команда Unpivot создала два новых столбца: Атрибут и Значение. Переименуйте их в MonthYear и Value. Округлите значения в столбце Value до 2 знаков после запятой. Загрузите запрос в модель данных Power Pivot, кликнув в окне редактора Power Query на вкладке *Главная* кнопку *Закрыть и загрузить* и далее *Закрыть и загрузить в…*

В Power Pivot мы имеем «узкую» таблицу с 5 столбцами и 576 строками (наш исходный «широкий» набор данных имел только 48 строк и 15 столбцов). B вспомните, что одна из тем, рассмотренных в главе о [производительности](http://baguzin.ru/wp/?p=20351), – это операция Unpivot.

### Использование Power Query для создания таблицы поиска из таблицы данных

Необходимо создать отдельные таблицы данных и поиска в модели данных Power Pivot в соответствии с [рекомендациями](http://baguzin.ru/wp/?p=20351). Однако в качестве источника данных часто вы получаете единую таблицу (рис. 20.42).

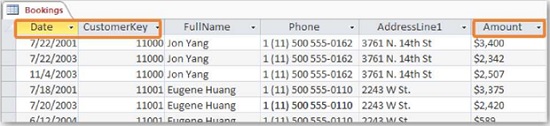


Рис. 20.42. Исходная таблица данных

Каждая строка в таблице содержит три основных элемента информации: когда, кто и сколько (выделено на рис. 20.42), но также содержит дополнительную информацию о каждом клиенте – полное имя, телефон и адрес. Это перебор, нет нужды хранить в одной таблице всю информацию о бронировании и о клиентах. Лучше выделить информацию о клиентах в отдельную таблицу поиска. Вот желаемый результат:



Рис. 20.43. Бронирование и клиенты в виде отдельных таблиц данных и поиска

Наша таблица бронирования живет в Access, поэтому создайте новый файл Excel, и пройдите по меню *Данные* –> *Получить данные* –> *Из базы данных* –> *Из базы данных Microsoft Access*. В открывшемся окне *Импорт данных* выберите файл DataAndLookupTogether.accdb и нажмите *Импорт*. Затем выберите таблицу Booking и нажмите *Изменить* для запуска редактора Power Query. Выделите и удалите столбцы, которые не нужны вам в таблице клиентов (дата и сумма). Отсортируйте таблицу по столбцу CustomerKey, выбрав этот столбец и нажав кнопку A-Z. Этот шаг предназначен исключительно для демонстрации, чтобы лучше было видно, как работает команда *Удалить дубликаты*. Для построения таблиц поиска сортировать не обязательно.

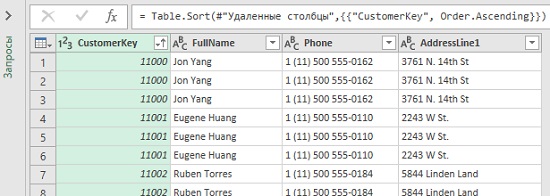


Рис. 20.44. Таблица после удаления двух столбцов и сортировки

Выделите столбец CustomerKey и нажмите кнопку *Удалить строки* –> *Удалить дубликаты*:

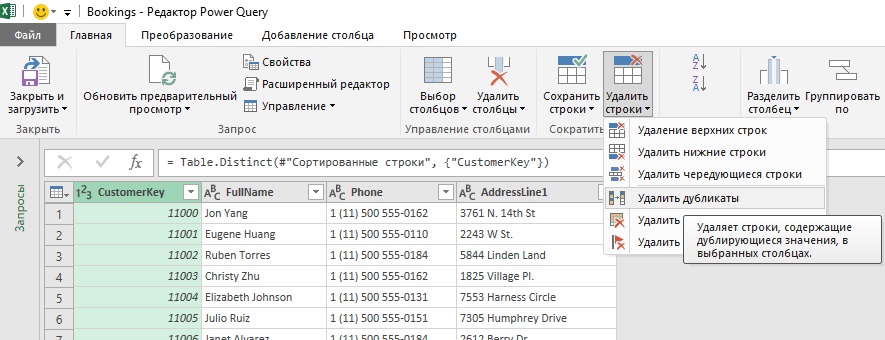


Рис. 20.45. Таблица после команды *Удалить дубликаты*

Переименуйте запрос в Customers, и загрузите его в модель данных PowerPivot. В этой таблице ключ клиента уникален.

Повторите импорт таблицы Booking из базы данных DataAndLookupTogether.accdb. На этот раз удалите столбцы свойств клиента (FullName, Phone и Address). Загрузите таблицу в PowerPivot. Свяжите две таблице в модели данных с помощью столбца CustomerKey. Получится результат, к которому мы стремились (как на рис. 20.43; см. также PowerQuery\_05\_Unflatten.xlsx).

### Создание таблицы календаря

Создание таблицы календаря связано с теми же проблемами, что и создание таблицы поиска; в конце концов, таблица календаря – это просто особый вид таблицы поиска. Если вы вручную создаете таблицу календаря, вам нужно часто обновлять ее, чтобы включать последние даты (а иногда, исключать старые). Или вы можете создать таблицу календаря, которая распространяется в будущее, но это может выводить в ваши срезы и сводные таблицы ненужные значения дат и сделать некоторые из расчетов «странными» в отношении неполных месяцев или недель.

Наилучший вариант – иметь таблицу календаря, которая «идет в ногу» с вашей таблицей данных – такой, которая начинается с первой даты, для которой у вас есть данные, и заканчивается на последнюю дату, для которой у вас есть данные. К сожалению, в редакторе Power Query до сих пор отсутствует функциональность «создать календарь» (в Excel 2019 я ее тоже не нашел). Хорошая новость – в Power Query есть мощный язык M. Плохая новость – M совершенно не похож на Excel и DAX. Хорошая новость – если вы выполняли приведенные выше примеры, вы уже писали на языке M.

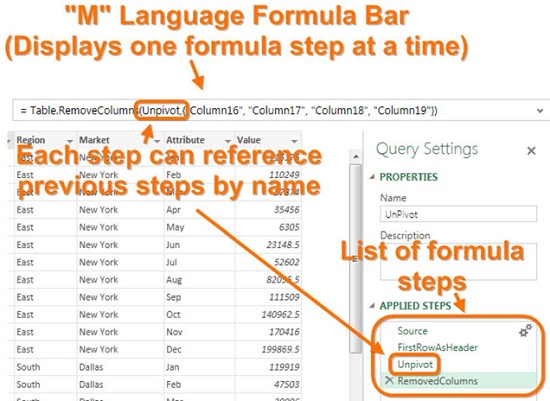


Рис. 20.46. Запись, которую вы видите в строке формул, написана на языке M, и создана кнопками ленты. Запрос может содержать несколько шагов формулы, и каждый шаг будет ссылаться на предыдущие шаги по имени

Кнопки на ленте Power Query похожи на Macro Recorder для макросов VBA – они помогают вам писать M, не зная языка, но затем вы можете редактировать M вручную, если хотите.

Создайте новую книгу Excel. Пройдите по меню *Данные* –> *Получить данные* –> *Из других источников* –> *Пустой запрос*:

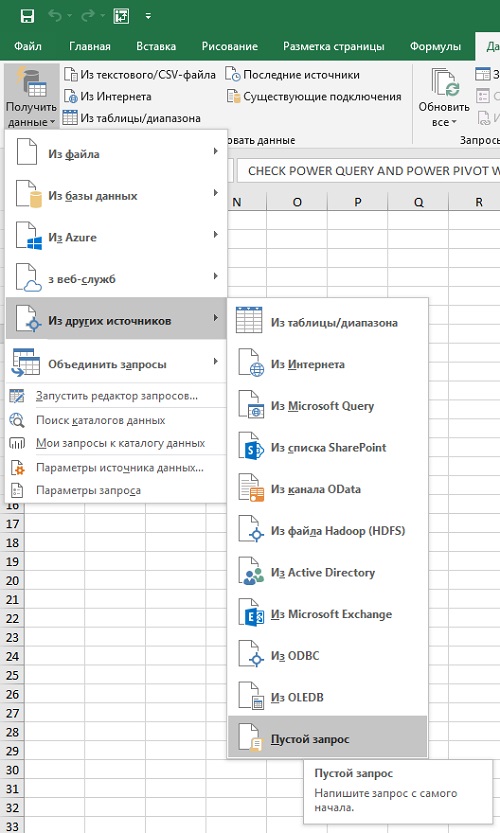


Рис. 20.47. Создание запроса с помощью языка М

В открывшемся окне *Редактор Power Query* кликните на кнопке *Расширенный редактор*:

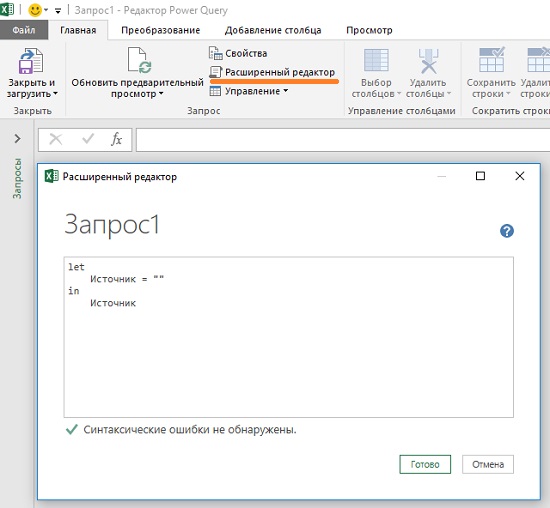


Рис. 20.48. Окно расширенного редактора

Удалите шаблон-заготовку, хранящуюся в расширенном редакторе и вставьте следующий текст:

//Calendar M Query

let

ChangedType = Table.TransformColumnTypes(Bookings,{{"Date", type date}}), //Rename Bookings, Date as needed for your dataset

MaxDate = Record.Field(Table.Max(ChangedType, "Date"),"Date"), //Rename Date as needed for your dataset

MinDate = Record.Field(Table.Min(ChangedType, "Date"),"Date"), //Rename Date as needed for your dataset

DaysElapsed = Number.From(MaxDate-MinDate),

DatesList = List.Dates(MinDate, DaysElapsed+1,Duration.From(1)),

RawDatesTable = Table.FromList(DatesList, Splitter.SplitByNothing(), {"Date"}, null, ExtraValues.Error),

ChangedType1 = Table.TransformColumnTypes(RawDatesTable,{{"Date", type date}}),

InsertedDay = Table.AddColumn(ChangedType1, "Day", each Date.Day([Date]), type number),

InsertedMonth = Table.AddColumn(InsertedDay, "Month", each Date.Month([Date]), type number),

InsertedYear = Table.AddColumn(InsertedMonth, "Year", each Date.Year([Date]), type number),

#"Inserted Year" = Table.AddColumn(InsertedYear, "Year.1", each Date.Year([Date]), type number),

#"Removed Columns" = Table.RemoveColumns(#"Inserted Year",{"Year.1"})

in

#"Removed Columns"

Переименуйте запрос: Запрос1 –> Calendar. Код можно найти в файле PowerQuery\_06\_Calendar.xlsx (убедитесь, что файл DataAndLookupTogether.accdb находится в той же папке; в противном случае связь с исходной базой данных будет потеряна; это не помешает увидеть код, но не позволит его обновить). Для этого откройте файл в Excel. Пройдите по меню *Данные* –> *Запросы и подключения*. Справа кликните правой кнопкой мыши на запросе Calendar, и выберите *Изменить*. В открывшемся окне *Редактор Power Query* на вкладке *Главная* кликните *Расширенный редактор*.

Подразумевается, что исходная таблица называется *Bookings*, и она содержит столбец с именем *Date*. Места, где эти имена отображаются в формулах выше, выделены цветом. Вы можете изменить их, чтобы соответствовать вашему собственному набору данных.

Нажмите *Готово* в окне *Расширенный редактор* (убедитесь, что синтаксические ошибки не обнаружены):

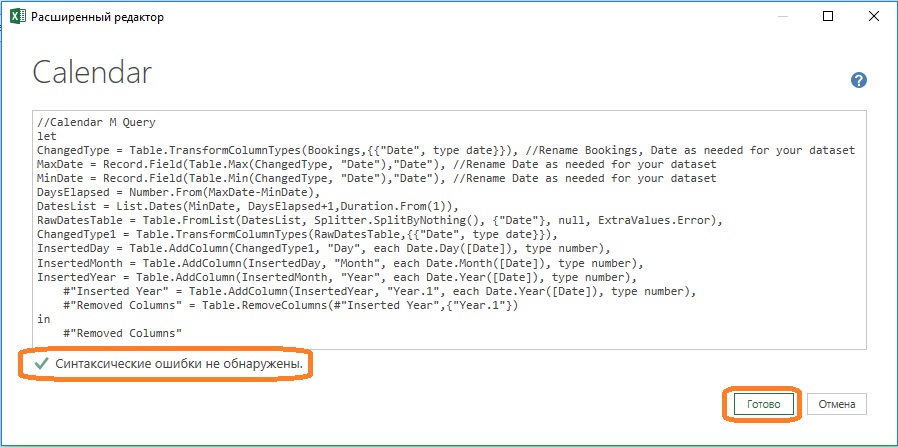


Рис. 20.49. Код на языке М в окне Расширенный редактор

Если загрузить этот запрос в модель данных Power Pivot, получим таблицу Calendar:

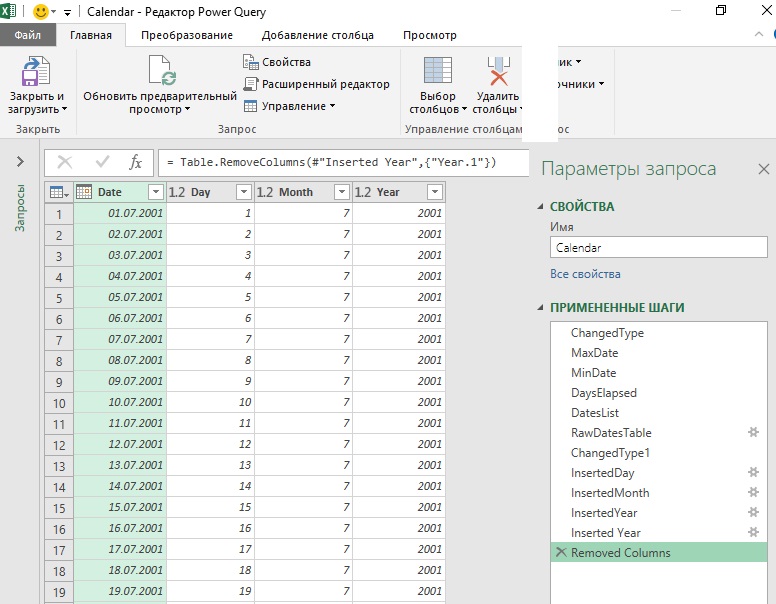


Рис. 20.50. Таблица Calendar в окне Редактор Power Query

Мы получили автоматически созданный (из таблицы данных) календарь, который начинается с нашей самой старой даты из таблицы данных и продолжается до (и включая) самую последнюю дату в нашей таблице данных. Каждая дата отображается только один раз, и никакие даты не пропущены. Все это важные качества календарной таблицы. И, конечно же, все это обновляется автоматически каждый раз, когда мы нажимаем кнопку *Обновить*.

Завершите обработку, пройдя в Редакторе Power Query по меню *Главная* –> *Закрыть и загрузить* –> *Закрыть и загрузить в…* В открывшемся окне *Импорт данных* выберите *Только создать подключение* и *Добавить эти данные в модель данных*. В Power Pivot свяжите таблицы Bookings и Calendar по полю Date.

Поскольку Calendar это таблица календаря, довольно просто создавать дополнительные столбцы по мере необходимости. В Excel в области *Запросы и подключения* наведите мышь на запрос Calendar. Во всплывающем окне кликните *Изменить*. Выделите столбец Date и на вкладке ленты *Добавление столбца* нажмите кнопку *Дата*, затем *День* и *День недели* (рис. 20.51).

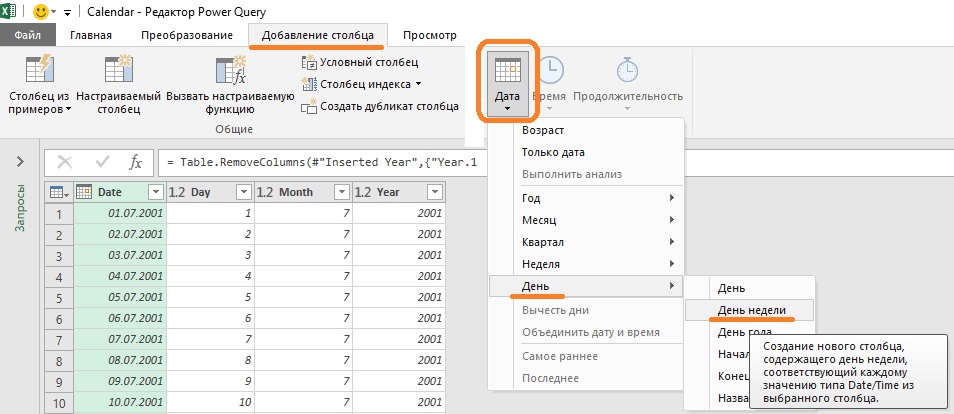


Рис. 20.51. Создание в запросе нового столбца *День недели*

### Как НЕ использовать Power Query

**Не используйте Power Query без Power Pivot.** Вы можете столкнуться с несколькими разовыми задачами, где вы используете только Power Query, но когда дело доходит до создания любой масштабируемой, долговечной, гибкой системы отчетности, Power Pivot является обязательным. Power Query – отличный инструмент, но он не заменяет Power Pivot. Хорошо построенная модель данных Power Pivot имеет возможность отвечать не только на вопросы, которые вы задаете прямо сейчас, но и на вопросы, которые вы можете задать в будущем. У вас есть меры (определите один раз, используйте везде), которые вы можете комбинировать любым способом, и у вас есть связи таблиц. Вы не получите это только с Power Query.

Power Query и Power Pivot дополняют друг друга. Используйте Power Query для очистки, формирования и преобразования данных до их импорта в модель данных. Продолжайте использовать Power Pivot для построения моделей данных, определения связей и написания мер.

**Не используйте вычисления Power Query в качестве замены мер DAX.** Это, по сути, продолжение предыдущего совета. Как только вы обучитесь языку M, вычислительная машина Power Query начинает искушать вас. Использование M и Power Query для добавления вычисляемых столбцов весьма эффективно, и мы настоятельно рекомендуем пользоваться этим. Но использование M в качестве замены мер DAX является ошибкой. Меры более динамичны, они автоматически корректируются и пересчитываются в ответ на срезы и фильтры сводной таблицы. Язык M только повторно вычисляет столбцы, когда вы обновляете данные. И для любых агрегаций, таких как SUM и т.п. язык DAX гораздо более эффективный и быстрый.

**Не используйте Power Query для объединения таблицы данных и таблицы поиска.** Из предыдущих глав вы уже знаете, что плоские таблицы, которые помещают данные и поиск в одну таблицу, не годятся для Power Pivot. Да, с Power Query просто объединить данные в одну таблицу. Но это лишнее. Вы получите оптимальную модель с помощью отдельных таблиц данных и поиска.

**Не используйте Power Query для объединения двух таблиц данных.** Каждая таблица данных обычно собирает информацию из отдельных бизнес-процессов. Например, процесс разработки бюджета сильно отличается от сбора трансакций продаж. Из того, что вам нужно проанализировать их вместе, не следует, что вы помещаете их в одну таблицу. Опять же, Power Query позволяет легко создавать такие объединенные таблицы. Но это не значит, что так следует делать. Оставьте таблицы данных отдельными, свяжите их вместе с помощью общих таблиц поиска, а затем напишите несколько мощных мер. Это даст гораздо более гибкую и мощную модель, чем любые плоские таблицы, которые вы можете создать в Power Query.

1. Если pivot – сводная таблица, то unpivot – что-то, типа разложения сводной таблицы. [↑](#footnote-ref-1)