

Глава 5. Объединение листов Excel в Power Query

Это продолжение перевода книги Кен Пульс и Мигель Эскобар. Язык M для Power Query. Главы не являются независимыми, поэтому рекомендую читать последовательно.

[Предыдущая глава](#) [Содержание](#) [Следующая глава](#)

В предыдущей главе были описаны методы агрегации данных из нескольких .txt или .csv файлов. В настоящей заметке описано, как объединять *Таблицы* или листы Excel из активной или внешней рабочей книги. К сожалению, стандартный пользовательский интерфейс Power Query этого не умеет. Но дополнительные манипуляции не будут слишком сложными. При этом методы работы с данными в активной книге отличаются от методов извлечения данных из внешнего файла.

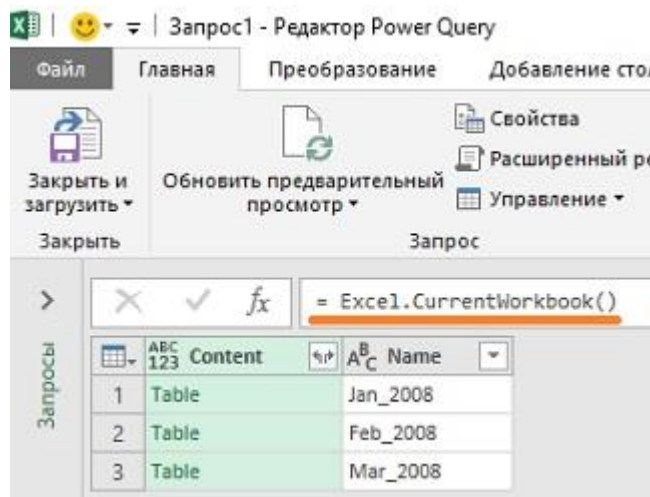


Рис. 5.1. Доступные Таблицы в окне редактора Power Query

Объединение таблиц и диапазонов в текущем файле

Откройте файл с примерами *Consolidate Tables-Start.xlsx*. В файле три листа с информацией о подарочных сертификатах, выданных spa-салонам. Сотрудник, создавший файл, не указывал в Таблицах дату выпуска сертификатов, а размещал данные на отдельных листах по месяцам: Jan 2008, Feb2008, Mar 2008. Для анализа данные желательно объединить, и Power Query справится с этим быстрее, чем ручная обработка.

Создайте пустой запрос: пройдите по меню *Данные* → *Получить данные* → *Из других источников* → *Пустой запрос*. В окне редактора Power Query в строке формул ведите (рис. 5.1):

```
=Excel.CurrentWorkbook()
```

Нажмите Enter. Вы увидите доступные объекты текущей книги: Таблицы, именованные диапазоны, подключения к внешним источникам данных. К сожалению, нельзя получить перечень листов файла.

Как вы узнали из главы 4, можно щелкнуть пробел рядом с зелеными словами в столбце *Content* для предварительного просмотра данных *Table* (рис. 5.2).

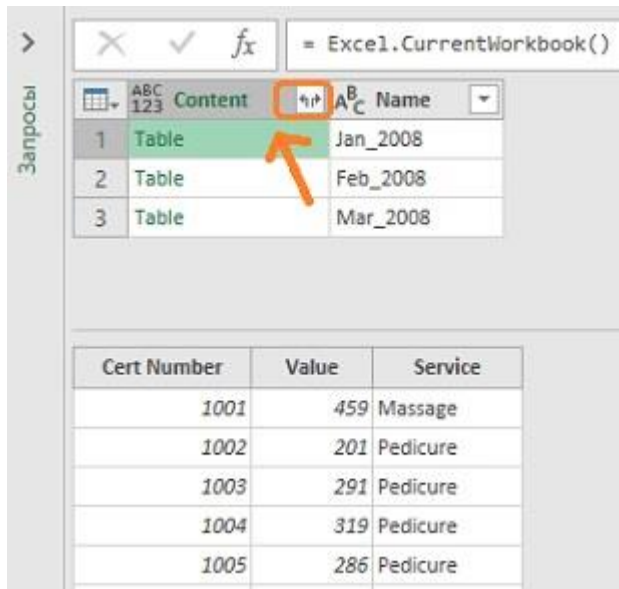


Рис. 5.2. Предварительный просмотр содержимого Table

В главе 4 вы также узнали, что кликнув на двуглавую стрелку в верхней части столбца *Content*, вы развернете содержимое Таблиц, сохраняя сведения из столбца *Name*. Итак, кликните двуглавую стрелку, в открывшемся окне настроек снимите флажок *Использовать исходное имя столбца как префикс*, нажмите *Ок*:

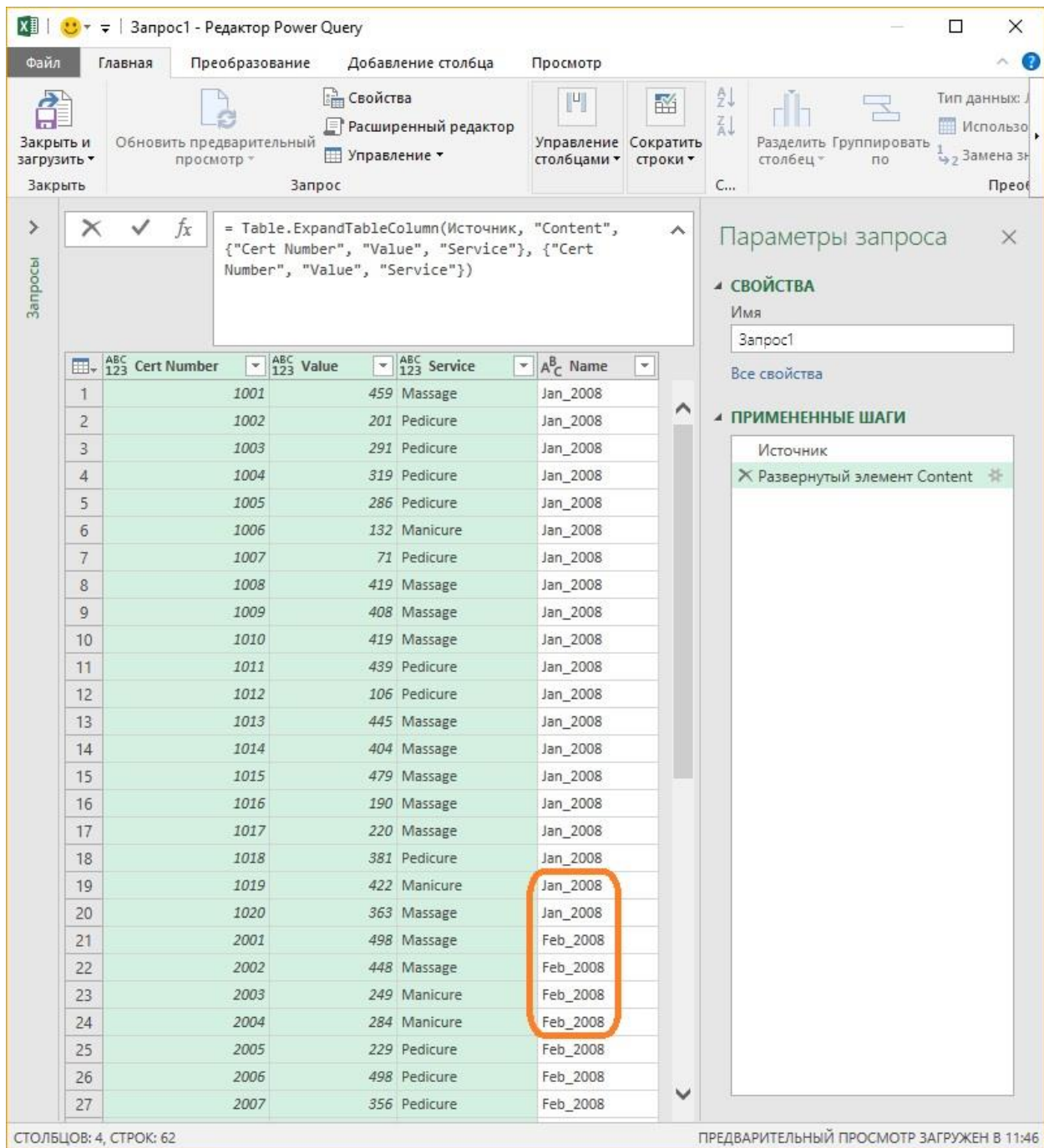


Рис. 5.3. Импорт и объединение данных с добавлением столбца с именем таблицы

Преобразуйте имена таблиц в даты (подробнее см. главу 4):

- Щелкните правой кнопкой мыши столбец *Name* → *Замена значений*
- Заменить символ (подчеркивание) на `_1_` (пробел, единица, запятая, пробел)
- Щелкните правой кнопкой мыши столбец *Name* → *Тип изменения* → *Дата*
- Перейдите на вкладку *Преобразование* → *Дата* → *Месяц* → *Конец месяца*
- Щелкните правой кнопкой мыши столбец *Name* → *Переименовать* → *Конец месяца*
- Измените имя запроса на *Подарочные сертификаты*

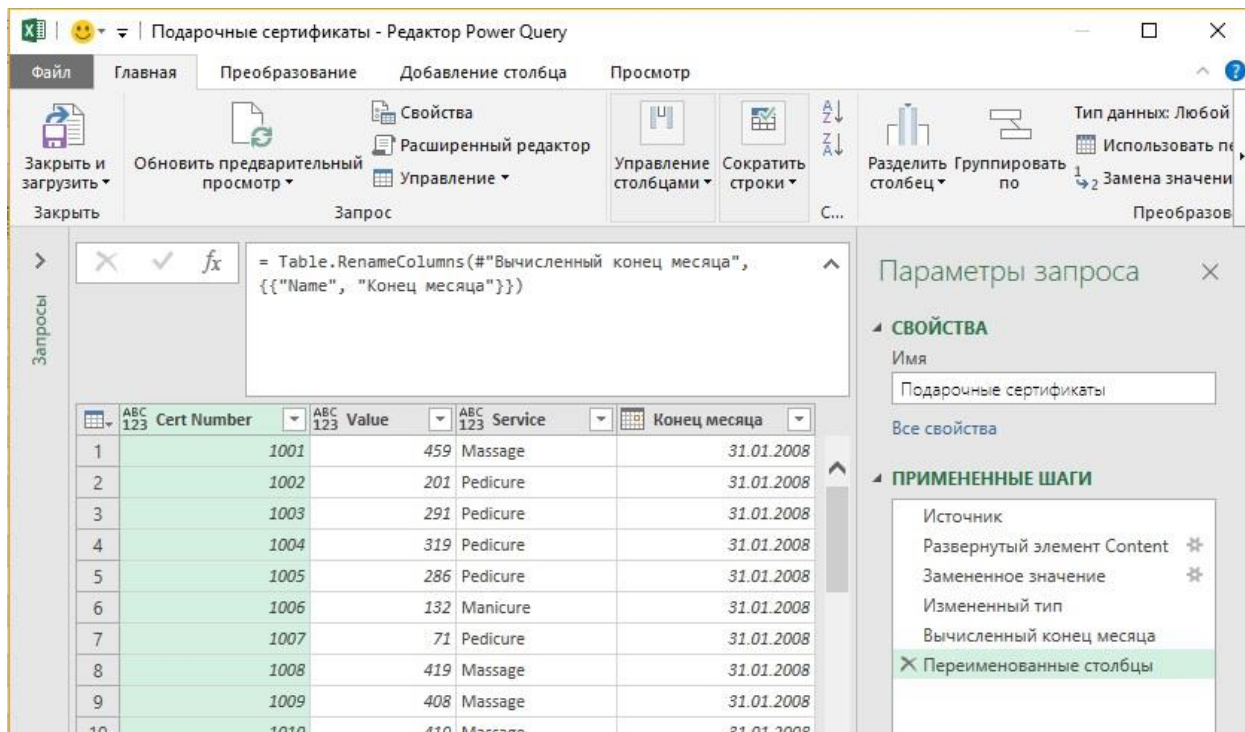


Рис. 5.4. Финальный вид запроса

Перейдите на вкладку *Главная* и кликните *Закреть и загрузить*. Объединенные данные будут размещены на новом листе Excel. К сожалению, запрос содержит кучу ошибок:

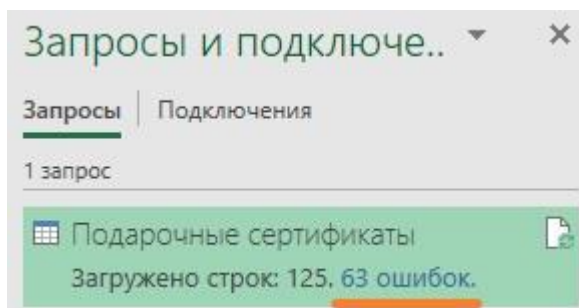


Рис. 5.5. Данные объединены на листе Excel; правда, вкрались ошибки

Не страшно. Наведите курсор мыши на запрос *Подарочные сертификаты*, и во всплывающем окне кликните *Изменить*. Снова откроется редактор Power Query. В области ПРИМЕНЕННЫЕ ШАГИ выберите первый шаг – *Источник* (цифра 1 на рис. 5.6). Вы заметите, что теперь у вас четыре таблицы в списке. Добавилась таблица *Подарочные сертификаты*, которая была создана в результате запроса. Чтобы избавиться от дублирования, необходимо добавить фильтрацию таблиц, участвующих в запросе. Кликните стрелочку возле названия столбца *Name* (2), выберите опцию *Текстовые фильтры* → *Не содержит*. Подтвердите, что вы хотите вставить шаг. В окне *Фильтрация строк* выберите *Не содержит* – *Подарочные* (3):

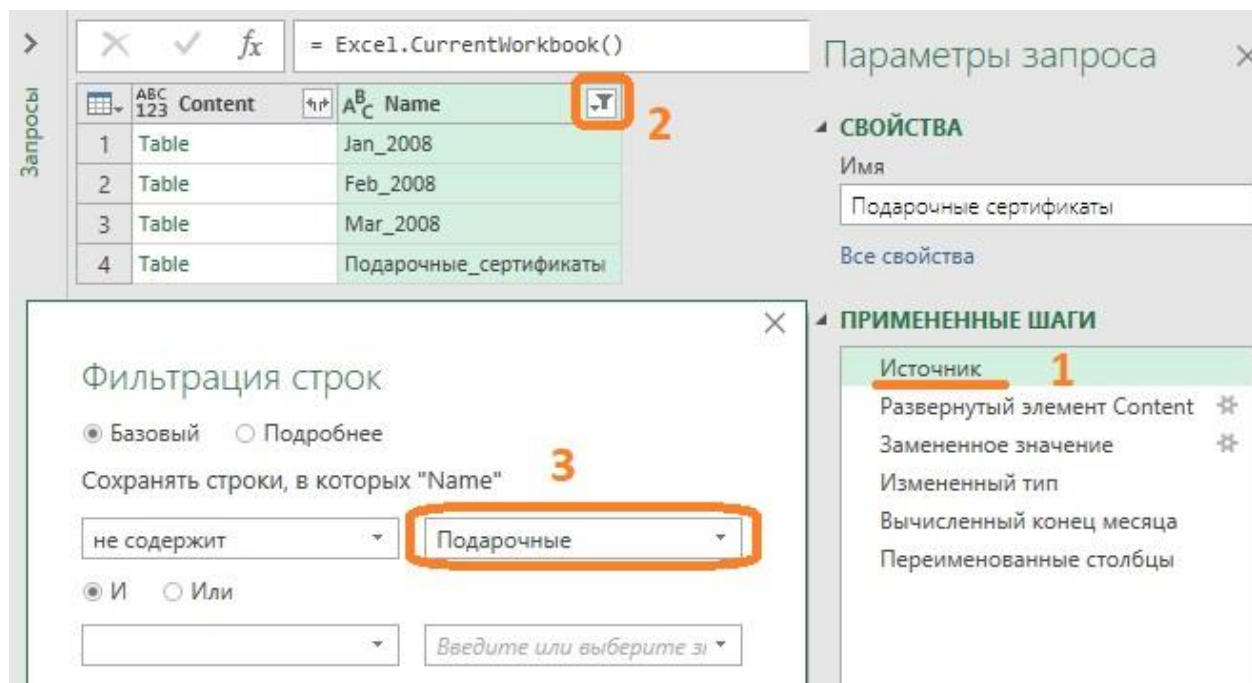


Рис. 5.6. Фильтрация таблиц по имени

Нажмите **Ок**. В редакторе Power Query перейдите на вкладку *Главная*. Кликните кнопку *Заккрыть и загрузить*. Теперь запрос содержит 62 строки; ошибок нет.

Существует и вторая возможность избавиться от ошибок – убрать дубли. Откройте редактор Power Query. Перейдите на шаг *Измененный тип*. Выберите столбец *Name* и на вкладке *Главная* кликните *Удалить строки* → *Удалить ошибки*. Подтвердите, что вы хотите вставить новый шаг в середину запроса. На вкладке *Главная* кликните *Заккрыть и загрузить*.

Объединение диапазонов и листов

Данные на листах Excel могут располагаться не в *Таблицах*. Напомню, что Power Query «не видит» листы Excel. Поэтому исходные данные можно организовать в именованные диапазоны. Это можно сделать, например, с помощью определения области печати. Трюк работает потому, что имя области печати является именем динамического диапазона.

Перейдите на лист Jan 2008. На вкладке *Разметка страницы* кликните *Печатать заголовки*. На закладке *Лист* введите A:D в поле *Выводить на печать диапазон*, кликните **Ок**.

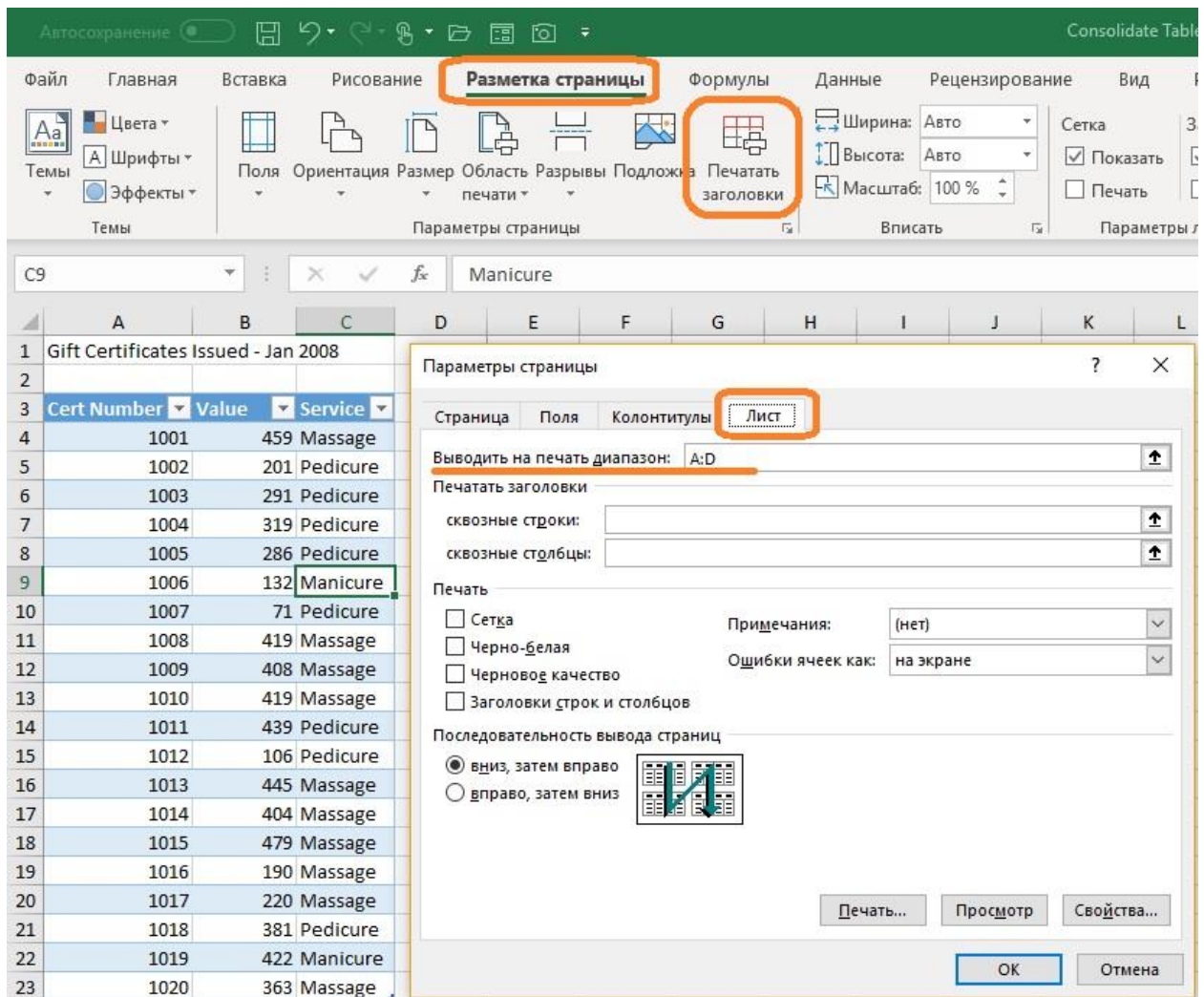


Рис. 5.7. Выбор области печати

Повторите процедуру для листов Feb 2008 и Mar 2008. Создайте пустой запрос, и в строке формул введите: =Excel.CurrentWorkbook(). Нажмите Enter. Вы увидите список трех таблиц и трех именованных диапазонов:



Рис. 5.8. Объекты книги Excel, доступные для импорта в Power Query

Чтобы не дублировать данные, отфильтруйте столбец *Name*: *Заканчивается на – печати*. Нажмите кнопку *Развернуть* в верхней части столбца *Content*. Оставьте выбранными все столбцы, снимите флажок *Использовать исходное имя столбца как префикс*. Обратите внимание на отличия от предыдущего случая. Показаны все строки в выбранных столбцах диапазона печати:

	ABC 123 Column1	ABC 123 Column2	ABC 123 Column3	ABC 123 Column4	A ^B _C Name
1	Gift Certificates Issued - Feb 2008	null	null	null	'Feb 2008'!Область_печати
2	null	null	null	null	'Feb 2008'!Область_печати
3	Cert Number	Value	Service	null	'Feb 2008'!Область_печати
4	2001	498	Massage	null	'Feb 2008'!Область_печати
5	2002	448	Massage	null	'Feb 2008'!Область_печати
6	2003	249	Manicure	null	'Feb 2008'!Область_печати
7	2004	284	Manicure	null	'Feb 2008'!Область_печати
8	2005	229	Pedicure	null	'Feb 2008'!Область_печати
9	2006	498	Pedicure	null	'Feb 2008'!Область_печати
10	2007	356	Pedicure	null	'Feb 2008'!Область_печати
11	2008	133	Manicure	null	'Feb 2008'!Область_печати
12	2009	270	Pedicure	null	'Feb 2008'!Область_печати
13	2010	353	Pedicure	null	'Feb 2008'!Область_печати
14	2011	493	Pedicure	null	'Feb 2008'!Область_печати
15	2012	312	Massage	null	'Feb 2008'!Область_печати
16	2013	197	Massage	null	'Feb 2008'!Область_печати
17	2014	306	Pedicure	null	'Feb 2008'!Область_печати
18	2015	295	Manicure	null	'Feb 2008'!Область_печати
19	2016	363	Massage	null	'Feb 2008'!Область_печати
20	null	null	null	null	'Feb 2008'!Область_печати
21	null	null	null	null	'Feb 2008'!Область_печати
22	null	null	null	null	'Feb 2008'!Область_печати
23	null	null	null	null	'Feb 2008'!Область_печати
24	null	null	null	null	'Feb 2008'!Область_печати

Рис. 5.9. Необработанный рабочий лист

Выполним дополнительную очистку данных:

- Главная → Удалить строки → Удалить верхние строки → 2
- Главная → Использовать первую строку в качестве заголовков
- Столбец Cert Number → Фильтр → снимите флажок с null
- Щелкните правой кнопкой мыши столбец Cert Number → Тип изменения → Целое число
- Выберите столбец Cert Number
- Закладка Главная → Удалить строки → Удалить ошибки
- Выберите столбец Cert Number. Удерживайте нажатой клавишу Shift выберите столбец Service
- Щелкните правой кнопкой мыши один из выбранных заголовков столбцов → Удалить другие столбцы
- Измените имя запроса на Все листы
- Главная → Закрывать и загрузить

При работе с областями печати рекомендуется ограничивать область печати необходимыми строками и столбцами. В примере выше мы выбрали целиком столбцы, что привело к импорту в Power Query около 3 млн. строк с трех листов. Наверное, вы заметили, как медленно выполнялись некоторые команды!

Агрегирование данных из других книг

Вам нужно создать список книг Excel и извлечь их содержимое, аналогично тому, что вы сделали в главе 4, когда вы извлекли содержимое файлов CSV.

Создайте новую книгу Excel. Создать новый запрос: Данные → Получить данные → Из файла → Из папки. Выберите папку Source Files. В списке есть как файлы Excel, так и иные файлы:

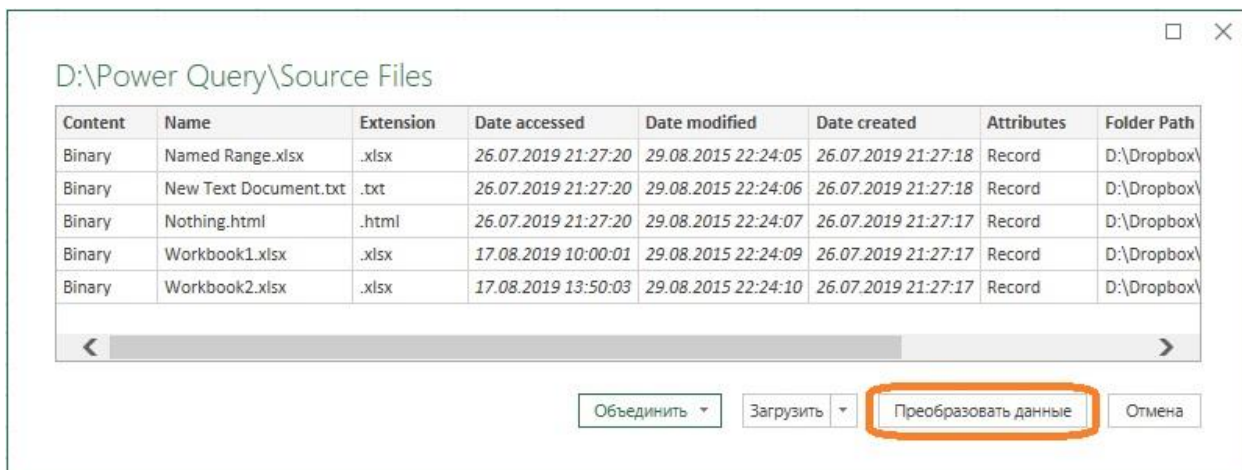


Рис. 5.10. Файлы, доступные в папке *Source Files*

Нажмите *Преобразовать данные*, и отфильтруйте файлы Excel:

- Щелкните правой кнопкой мыши столбец *Extension* → *Преобразование* → *нижний регистр*
- Фильтр столбца *Extension* → *Текстовые фильтры* → *Начинается с...* → *.xlsx*
- Выберите столбцы *Content* и *Name* → щелкните правой кнопкой мыши → *Удалить другие столбцы*

У вас может возникнуть соблазн нажать кнопку *объединить файлы...*

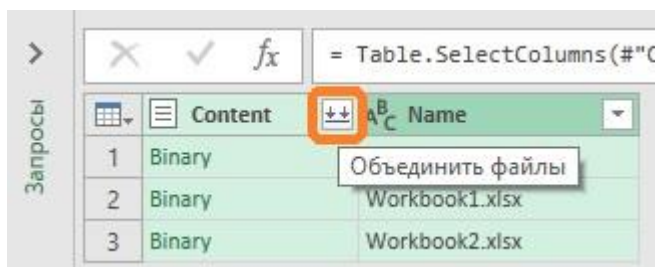


Рис. 5.11. Объединить файл

... и, к сожалению, Power Query позволит вам это сделать. Однако, вы обнаружите, что Power Query делает что-то весьма странное. Чтобы откатить импорт, перейдите в область ПРИМЕНЕННЫЕ ШАГИ и удалите все шаги после шага *Другие удаленные столбы*.

Раз вы не можете объединить и импортировать файлы простым методом, пойдём трудным способом:

- В редакторе Power Query перейдите на вкладку *Добавление столбца* → *Настраиваемый столбец*
- Введите формулу `=Excel.Workbook([Content])`
- Нажмите кнопку *Ok*
- Щелкните правой кнопкой мыши столбец *Content* → *Удалить*

Новый пользовательский столбец содержит все объекты, к которым можно подключиться, включая все Таблицы Excel, именованные диапазоны и даже листы:

✕ ✓ *fx* = Table.RemoveColumns("#Добавлен поль

	A ^B C Name	ABC 123 Пользовательская
1	Named Range.xlsx	Table
2	Workbook1.xlsx	Table
3	Workbook2.xlsx	Table

Name	Data	Item	Kind	Hidden
Jan 2008	Table	Jan 2008	Sheet	FALSE
Feb 2008	Table	Feb 2008	Sheet	FALSE
Mar 2008	Table	Mar 2008	Sheet	FALSE
Jan_2008	Table	Jan_2008	Table	FALSE
Feb_2008	Table	Feb_2008	Table	FALSE
Mar_2008	Table	Mar_2008	Table	FALSE

Рис. 5.12. Объекты, доступные для импорта

Столбец *Пользовательская* имеет двуглавую стрелку (значок расширения), поэтому его можно развернуть. Нажав на значок вы получаете список всех объектов трех файлов, доступных для импорта

✕ ✓ *fx* = Table.ExpandTableColumn("#Удаленные столбцы", "Пользовательская", {"Name",

	A ^B C Name	ABC 123 Name.1	ABC 123 Data	ABC 123 Item	ABC 123 Kind	ABC 123 Hidden
1	Named Range.xlsx	Sheet1	Table	Sheet1	Sheet	FALSE
2	Named Range.xlsx	Sheet2	Table	Sheet2	Sheet	FALSE
3	Named Range.xlsx	rngValue	Table	rngValue	DefinedName	FALSE
4	Workbook1.xlsx	Jan 2008	Table	Jan 2008	Sheet	FALSE
5	Workbook1.xlsx	Feb 2008	Table	Feb 2008	Sheet	FALSE
6	Workbook1.xlsx	Mar 2008	Table	Mar 2008	Sheet	FALSE
7	Workbook1.xlsx	Jan_2008	Table	Jan_2008	Table	FALSE
8	Workbook1.xlsx	Feb_2008	Table	Feb_2008	Table	FALSE
9	Workbook1.xlsx	Mar_2008	Table	Mar_2008	Table	FALSE
10	Workbook2.xlsx	Apr 2008	Table	Apr 2008	Sheet	FALSE
11	Workbook2.xlsx	May 2008	Table	May 2008	Sheet	FALSE
12	Workbook2.xlsx	Jun 2008	Table	Jun 2008	Sheet	FALSE
13	Workbook2.xlsx	Apr_2008	Table	Apr_2008	Table	FALSE
14	Workbook2.xlsx	May_2008	Table	May_2008	Table	FALSE
15	Workbook2.xlsx	Jun_2008	Table	Jun_2008	Table	FALSE

Рис. 5.13. Объекты, доступные для импорта

Столбец *Kind* показывает, что у вас есть Листы, определенное имя и Таблицы. Если не отфильтровать этот перечень объектов, у вас будет много дублей:

- Отфильтруйте столбец *Kind*, оставив только *Sheet*
- Отфильтруйте столбец *Name*, удалив файл *Named Range.xlsx*
- Выберите столбцы *Name*, *Name.1* и *Data* → щелкните правой кнопкой мыши на заголовке одного из этих столбцов → *Удалить другие столбцы*
- Кликните кнопку *Развернуть* у заголовка столбца *Data* (снимите настройки префикса)

Запрос теперь выглядит следующим образом:

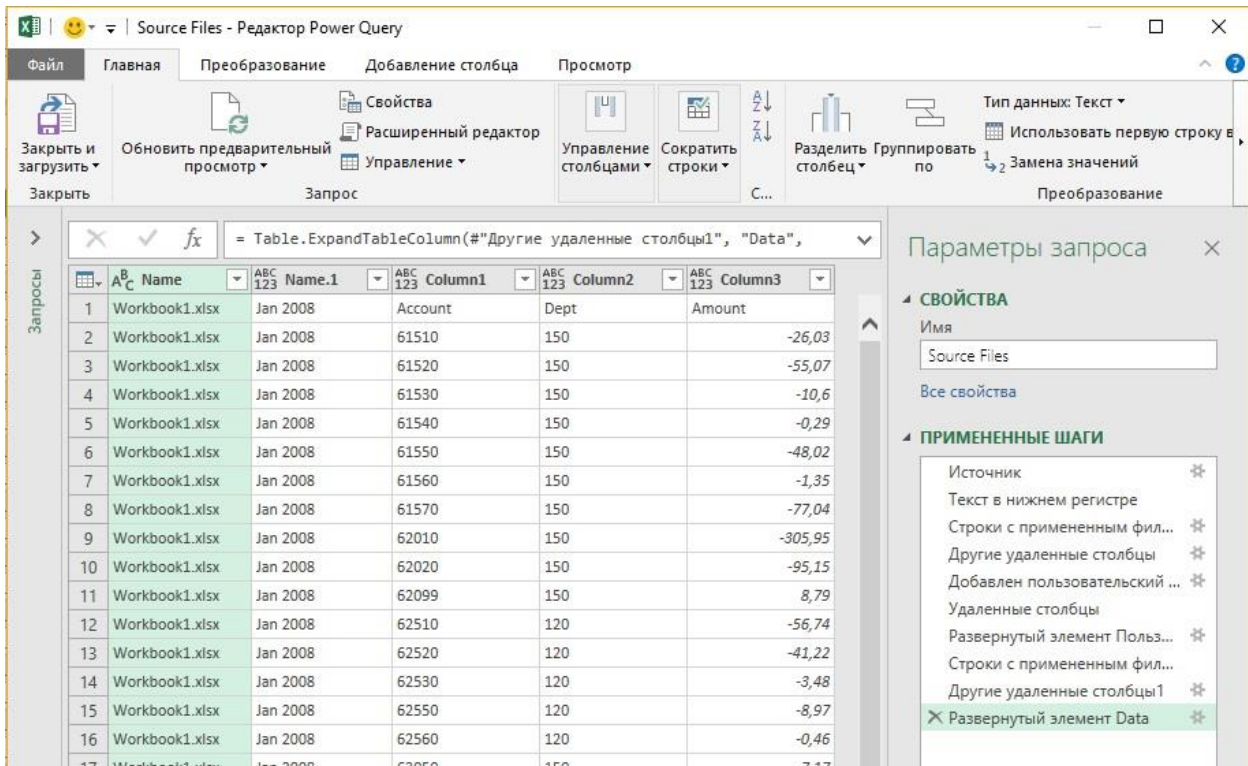


Рис. 5.14. Запрос с шестью импортированными объектами Sheet

Преобразуем данные:

- Главная → *Использовать первую строку в качестве заголовков*
- Щелкните правой кнопкой мыши на заголовке столбца Workbook1.xlsx → *Переименовать* → Source File
- Щелкните правой кнопкой мыши заголовке столбца Jan 2008 column → *Переименовать* → Month
- Выберите столбец Amount → Главная → *Удалить строки* → *Удалить ошибки*
- Измените имя запроса *FromExcelFiles*
- Главная → *Закреть и загрузить*

Данные загружаются в Таблицу на листе Excel. На их основе можно создать сводную таблицу, чтобы увидеть, что вы смогли извлечь из внешних файлов Excel:

Названия строк	Количество по полю Account	Сумма по полю Amount
Workbook1.xlsx	6 084	169 822
+ янв	2 030	89 791
+ фев	1 857	56 211
+ мар	2 197	23 820
Workbook2.xlsx	6 084	169 822
+ апр	2 030	89 791
+ май	1 857	56 211
+ июн	2 197	23 820
Общий итог	12 168	339 645

Рис. 5.15. Сводная позволяет проверить, что же вы импортировали

Видно, что вы успешно извлекли данные из двух Excel-файлов, каждый из которых содержит по три листа. В общей сложности извлекли более 12 000 записей.

Подытожим

Функция Excel.CurrentWorkbook() считывает все объекты текущего файла. Поскольку она является первой в применяемых шагах запроса, вы получаете эффект рекурсии. При обновлении Power Query добавит объекты, созданные в процессе выполнения запроса, к тем, что существовали

первоначально. Стратегии защиты заключается в фильтрации объектов по имени или фильтрации ошибок в ключевых столбцах. Тестируйте метод фильтрации с помощью нескольких команд *Обновить все*.

Функция `Excel.Workbook([Content])` не вызывает проблем с рекурсией, поскольку исходные данные считываются из внешних книг, а результаты запроса сохраняются в текущей книге. Но функция `Excel.Workbook([Content])` создает иную проблему: она извлекает листы в дополнение к диапазонам и таблицам. Это может приводить к дублированию данных. Обратите особое внимание на фильтрацию столбца *Kind* (вид), чтобы избежать этой проблемы. Даже если при построении запроса в книге имеется только один вид данных, полезно применить фильтрацию, чтобы защитить решение от будущих изменений.