

Объединение большого числа таблиц в Power Query

Это фрагмент книги [Гил Равив. Power Query в Excel и Power BI: сбор, объединение и преобразование данных](#).

[Предыдущий раздел](#)

[К содержанию](#)

[Следующий раздел](#)

Добавление таблиц из папки

В компании AdventureWorks ежегодно по мере выпуска новых товаров создается новый файл. Сейчас имеется три файла: C03E03 - 2015.xlsx, C03E03 - 2016.xlsx, C03E03 - 2017.xlsx за три года. Но в будущем их может стать больше. Отчет должен включать все таблицы из папки. Скачайте приложенные файлы Excel и поместите их в выделенную папку. Откройте новую книгу в Excel. Пройдите *Данные* → *Получить данные* → *Из файла* → *Из папки*. В окне *Обзор* кликните *Открыть*, хотя никакие файлы не выбраны:

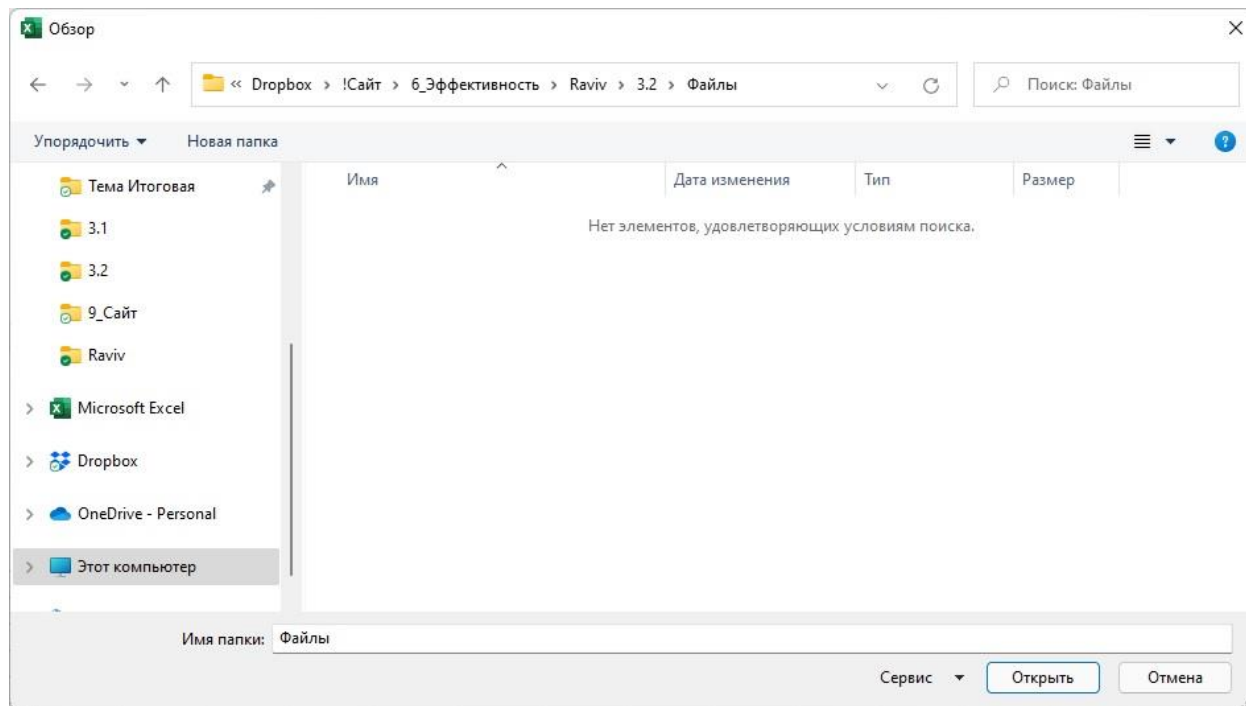


Рис. 1. Выбор папки для импорта файлов

В окне предварительного просмотра RQ вы увидите файлы, находящиеся в паке. Выберите *Объединить* → *Объединить и преобразовать данные*.

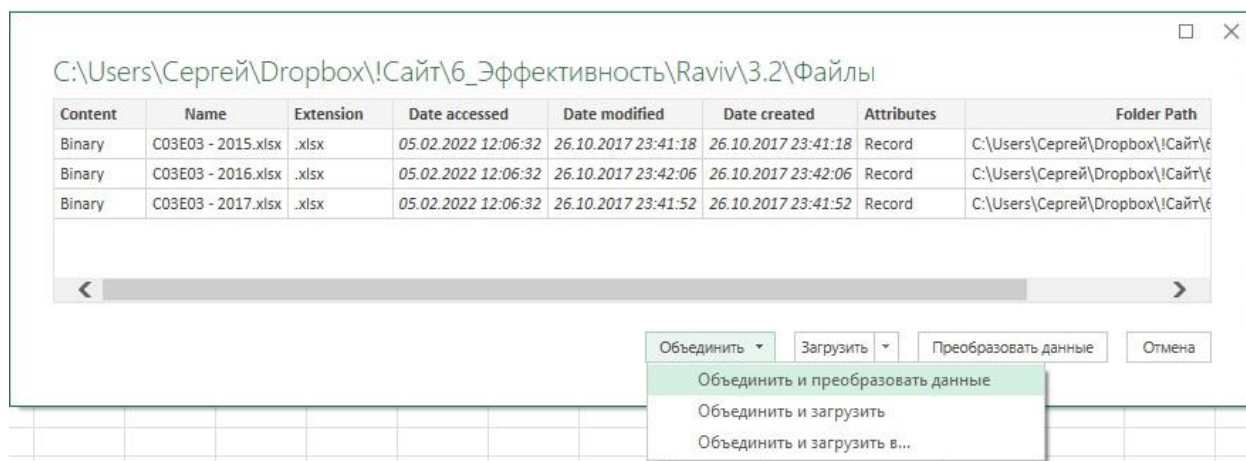


Рис. 2. Объединить и преобразовать данные

Если у вас нет уверенности, что все файлы из папки следует объединять, лучше добавить еще один шаг в сценарий. Вместо *Объединить* → *Объединить и преобразовать данные* кликните *Преобразовать данные*. После этого в редакторе Power Query откроется папка загрузки. Здесь можно применять фильтры. Например, если в папке имеются другие типы документов, можно

отфильтровать файлы по расширению .xlsx. Затем можно объединить файлы, кликнув кнопку *Объединить файлы*, расположенную в заголовке столбца *Content*.

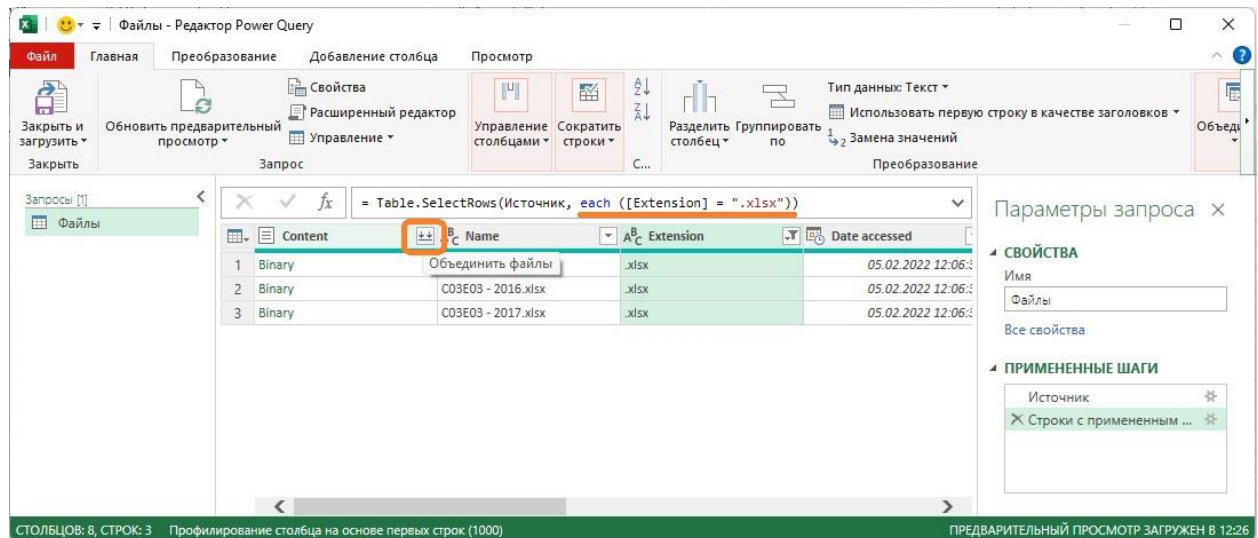


Рис. 3. Управление файлами при объединении

Не важно, каким путем вы пошли: кликнув *Объединить* → *Объединить и преобразовать данные* или через фильтрацию и кнопку *Объединить файлы*, у вас откроется окно *Объединить файлы*. Выберите лист *Sheet1*, кликните *Ok*.

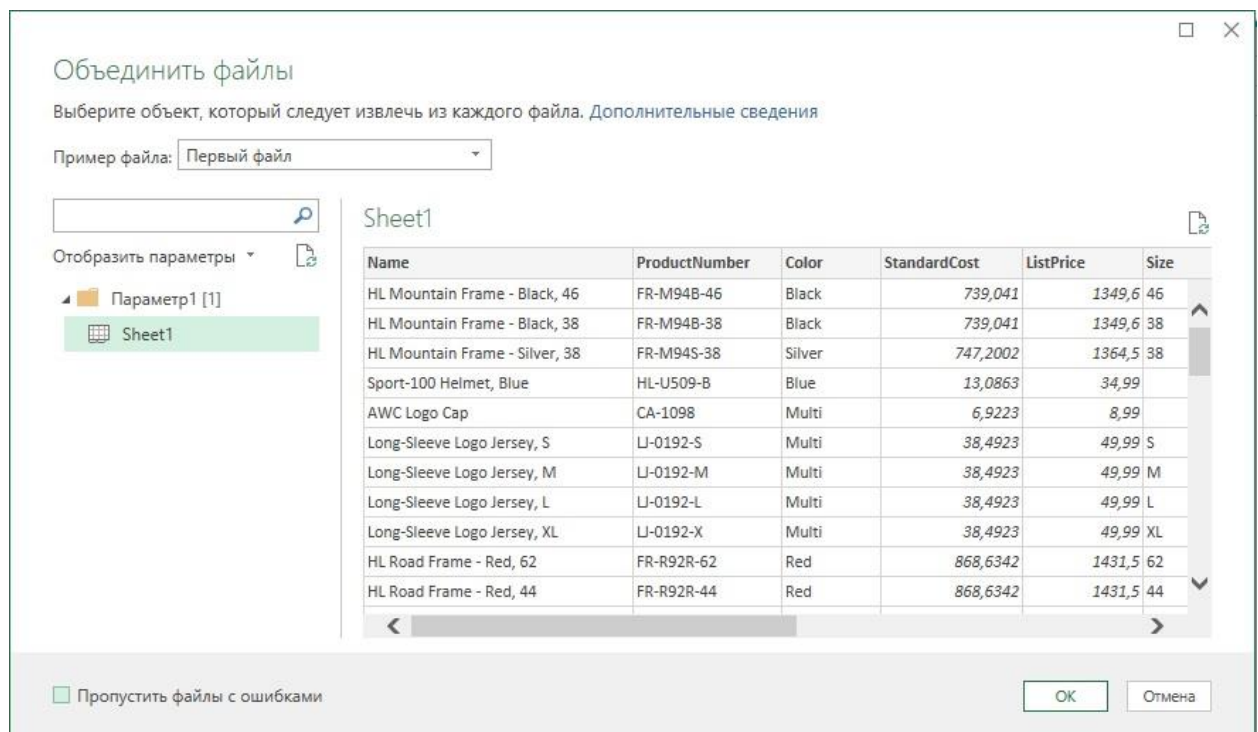


Рис. 4. Объединить файлы

Окно *Объединить файлы* является эквивалентом окна *Навигатор*, которое открывается при загрузке одной книги Excel. В окне *Навигатор* можно выбрать рабочий лист или таблицу для изменения или для загрузки в отчет, в окне *Объединить файлы* можно выбрать рабочий лист или таблицу. Выбор, сделанный в отношении одной рабочей книги, применится ко всем рабочим книгам в папке.

В окне редактора Power Query столбец *Source.Name* содержит имена файлов из папки. Из этих имен можно извлечь важный контекст – год. Например, так: замените "C03E03 - " пустой строкой, затем замените ".xlsx" пустой строкой. Чтобы выполнить замену выделите столбец *Source.Name* и пройдите *Преобразование* → *Замена значений* → *Замена значений*. Или разделите столбец

Source.Name разделителем "-", а затем "." Удалите лишние столбцы. Переименуйте столбец "Year". Загрузите три объединенных файла на лист Excel.

Если теперь вы добавите файл C03E03 - 2018.xlsx в папку, то вам будет достаточно в книге Excel с запросом, использующим импорт из папки, кликнуть *Обновить*, и данные из нового файла будут добавлены. В какой-то момент, выполняя запрос выше в редакторе PQ был создан ряд вспомогательных запросов. Мы изучим их смысл позже.

Добавление листов из книги

Иногда сходные таблицы сохраняются на разных листах в одной книге Excel. Сведем все рабочие листы в одну таблицу, сохраняя контекст года выпуска для каждого продукта. Загрузите файл C03E04. Откройте новую книгу Excel. Пройдите *Данные* → *Получить данные* → *Из файла* → *Из книги*. Выберите файл C03E04.xlsx, кликните *Импорт*.

В окне *Навигатор* выберите строку, где находится значок папки (в нашем примере, C03E04.xlsx). Не выбирайте определенные листы в окне *Навигатор*. Если это сделать, то появление новых рабочих листов в рабочей книге потребует изменить запрос. Вместо этого выбирается вся книга. Кликните *Преобразовать данные*. Откроется редактор Power Query

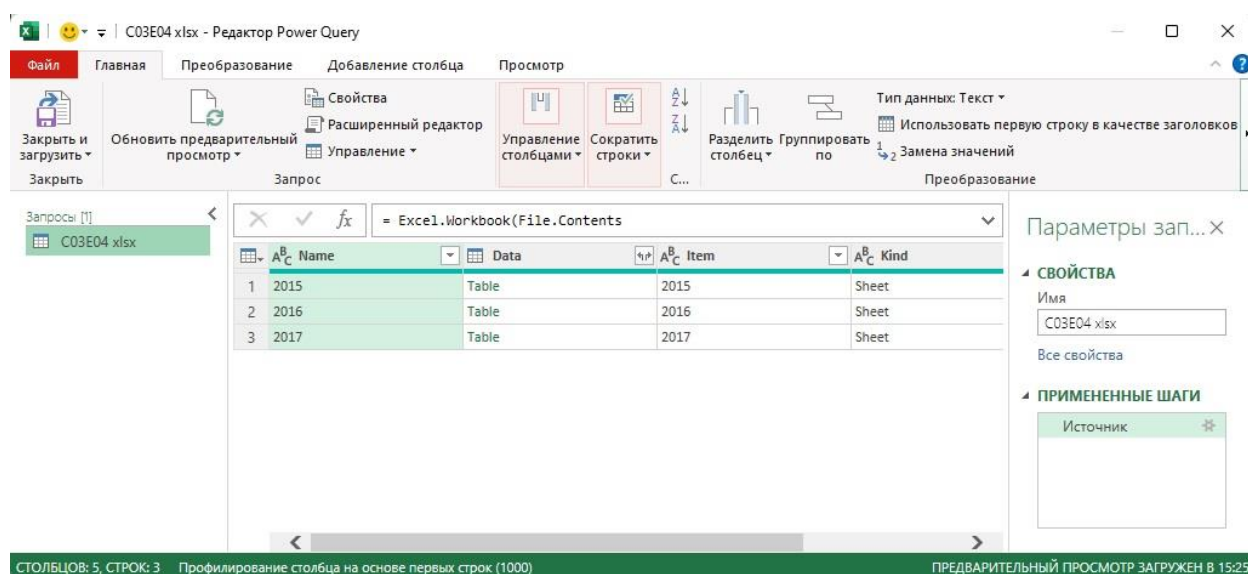


Рис. 5. Три листа в одном запросе

Переименуйте запрос в Products. Вы видите таблицу с рабочими листами в отдельных строках. Фактическое содержание каждого листа инкапсулировано в столбце *Data*. Перед тем как комбинировать рабочие листы, оставьте только столбцы *Name* и *Date*.

Если в исходной книге имеются скрытые рабочие листы или некоторые рабочие листы с несвязанными данными, то можно отфильтровать их до момента удаления ненужных столбцов. Например, можно применить фильтр к столбцу *Hidden* для того, чтобы оставить только строки, содержащие значение FALSE, что позволит исключить скрытые рабочие листы.

В заголовке столбца *Data* щелкните мышью на кнопке *Развернуть*. В окне снимите галочку *Использовать исходное имя столбца как префикс*. Нажмите *Ок*. Обратите внимание, что объединенная таблица содержит заголовки всех трех таблиц.

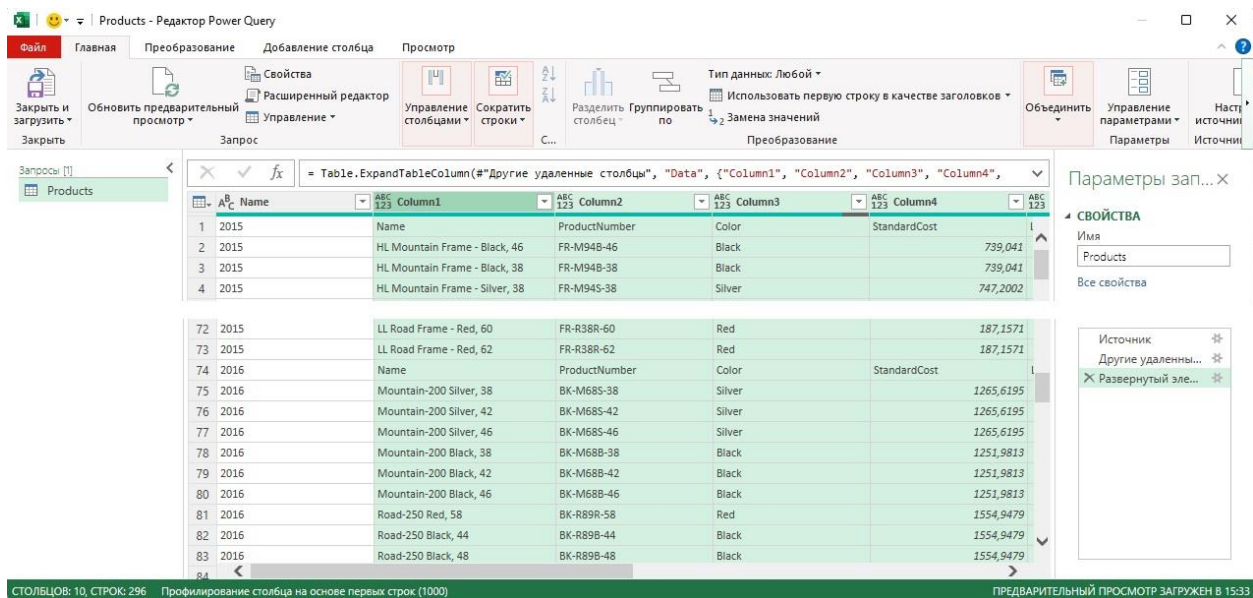


Рис. 6. Объединенная таблица

Пройдите *Преобразование* → *Использовать первую строку в качестве заголовков*. Кликните фильтр в заголовке столбца *Name*. Снимите галочку с *Name*. Проверьте что в районе строки 73 исчезла строка заголовка второго листа. Измените заголовок столбца *2015* на *Year*. Загрузите объединенную таблицу в книгу Excel.

Протестируем решение, добавив в исходный файл CO3E04.xlsx лист с данными за 2018 г. Для этого просто продублируйте лист 2017 и переименуйте его в 2018.

Вы уже заметили, что, пока активирован редактор Power Query, нельзя получить доступ к другим рабочим книгам. Если при работающем редакторе Power Query вам всё же нужно поработать в Excel, запустите новый экземпляр Excel с помощью панели задач или меню Пуск. Новый экземпляр Excel не будет блокироваться текущим окном редактора Power Query.

Обновите запрос *Products* в вашей рабочей книге Excel и удостоверьтесь, что товары, продублированные из рабочего листа 2018, теперь добавляются со значением 2018 в качестве даты выпуска. Это работает!

Усложним задачу. Продублируйте лист 2015, переименуйте его в 2014 и поместите первым в книге. Теперь обновление завершится ошибкой:

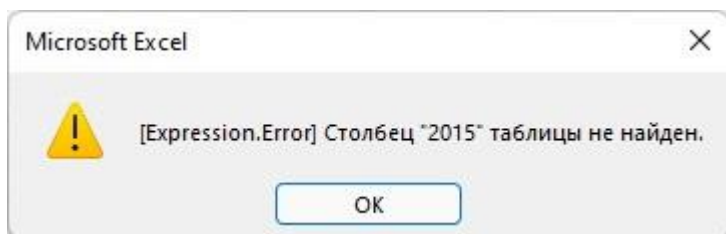


Рис. 7. Ошибка обновления

Чтобы быстро устранить эту ошибку, откройте редактор Power Query, выберите запрос *Products* и выберите в панели *Примененные столбцы* шаг *Изменение типа*. В строке формул отобразится:

```
= Table.TransformColumnTypes("#Повышенные заголовки",{"2015", Int64.Type}, {"Name", type text}, {"ProductNumber", type text}, {"Color", type text}, {"StandardCost", type any}, {"ListPrice", type any}, {"Size", type text}, {"Weight", type any}, {"ParentProductCategoryName", type text}, {"ProductCategoryName", type text})
```

Замените "2015" на "2014".

Выберите шаг *Переименованные столбцы*. В строке формул отобразится:

```
= Table.RenameColumns("#Строки с примененным фильтром",{"2015", "Year"})
```

Замените "2015" на "2014".

Закройте редактор Power Query и обновите отчет. Товары, выпущенные в 2014 году, теперь будут правильно отображаться в отчете.

Как усовершенствовать запрос, чтобы при добавлении нового листа в качестве первого не нужно было вносить исправления?

Продвинутый подход к объединению нескольких листов

Чтобы исправить две ошибки с явным указанием года первого из импортируемых листов, откройте запрос Products в редакторе Power Query и удалите шаг *Измененный тип*. Выберите шаг *Переименованные столбцы*. В строке формул отобразится:

```
= Table.RenameColumns("#Строки с примененным фильтром",{"2014", "Year"})
```

Замените ее на:

```
= Table.RenameColumns("#Строки с примененным фильтром", {Table.ColumnNames("#Строки с примененным фильтром"){0}, "Year"})
```

Здесь вы заменили инструкцию "Переименуй столбец 2015 в Year" на более универсальную: "Переименуй первый столбец таблицы в Year". Ссылка на первый столбец выполнена с помощью функции Table.ColumnNames. Эта функция возвращает список имен столбцов таблицы, которую получает в качестве аргумента. В данном случае таблица называется "Строки с примененным фильтром", она была названа так на предыдущем шаге, когда выполнялось фильтрование таблицы. Для направления Power Query к первому элементу в списке имен столбцов можно задать в фигурных скобках нулевой индекс. Этот метод в языке M позволяет получить доступ к определенным элементам в списке или в определенной строке в таблице. Индекс в языке M начинается с нуля вместо единицы. Таким образом, индекс первого элемента равен нулю, а для n -го элемента будет равен $n - 1$.

Теперь можно закрыть редактор Power Query и обновить отчет. После этого товары будут корректно комбинироваться независимо от того, какой год указан в качестве первого листа.