

Глава 3. Простое объединение таблиц в Power Query

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

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

Перед профессионалами Excel часто встают задачи объединения данных из нескольких однотипных таблиц.¹ Power Query может делать это автоматически.

В папке примеров есть три CSV-файла: Jan 2008.csv, Feb 2008.csv и Mar 2008.csv. Начните с импорта первого файла:

- Создайте новую книгу Excel
- Создайте запрос *Данные* → *Из текстового/CSV-файла*
- Выберите файл Jan 2008.csv

	Date	Account	Dept	Amount
2027	31.01.2008	86330	150	100,02
2028	31.01.2008	86350	150	20,5
2029	31.01.2008	86360	150	282,82
2030	31.01.2008	86400	170	0,03
2031	31.01.2008	86450	170	-105,16
2032				89790,94
2033				

Рис. 3.1. Импортированный CSV-файл Jan 2008.csv содержит одну ошибку

Power Query импортирует файл и автоматически выполнит следующие действия:

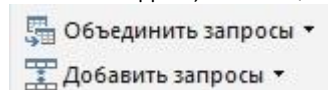
- Продвинет первую строку в заголовки.
- Задаст типы данных.

Поскольку исходный файл Jan 2008.csv содержит данные в стандарте США, следует удалить шаг *Измененный тип*, и повторно назначить типы данных, используя для столбцов *TranDate* и *Sum of Amount* локальные установки США (а не РФ, действующие по умолчанию; подробнее см. [предыдущую главу](#)). Переименуйте столбцы *TranDate* → *Date* и *Sum of Amount* → *Amount*. Нажмите кнопку *Заккрыть и загрузить*. Данные будут импортированы на лист Excel в виде, как на рис. 1.

Данные всё еще содержат одну ошибку – общие итоги. Вернитесь в редактор Power Query. Выделите столбец *Date*, кликните *Удалить строки* → *Удалить ошибки*. Нажмите *Заккрыть и загрузить*. Строка с итогами будет удалена.

Повторите операции для импорта Feb 2008.csv и Mar 2008.csv. Когда вы закончите, у вас будет три таблицы в книге Excel, каждая на своем листе. Чтобы объединить таблицы создайте новый запрос. Пройдите по меню *Получить данные* → *Объединить запросы* → *Добавить*:

¹ На самом деле, Power Query поддерживает два типа объединений:



В английском варианте, это Merge Queries и Append Queries. Первая опция позволяет объединять таблицы, исключая строки-дубли и проводя иные интеллектуальные операции с данными. Вторая опция просто добавляет каждый последующий набор в конец существующего. Пиктограммы довольно неплохо иллюстрируют это. Настоящая заметка посвящена второй опции.

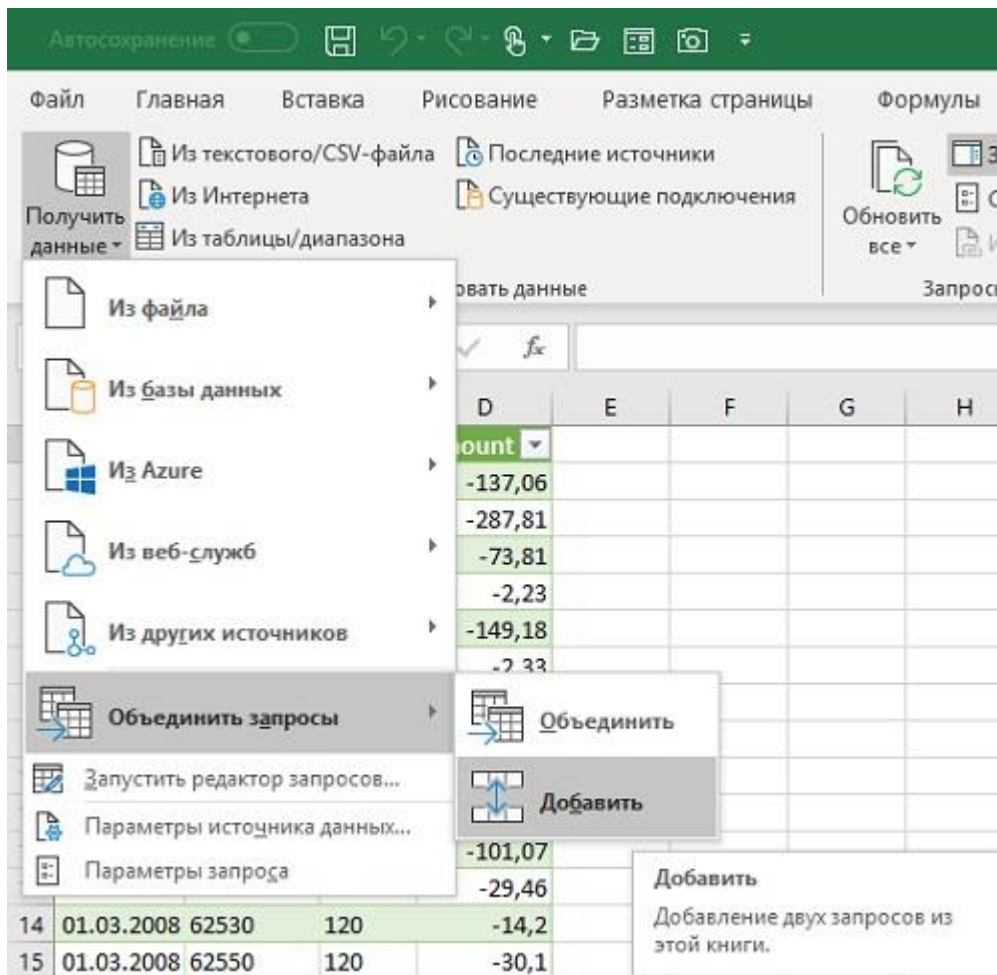


Рис. 3.2. Меню объединения запросов

Откроется диалоговое окно *Добавление* (рис. 3.3). Доступ к окну *Добавление* можно получить и из редактора Power Query. Для этого в редакторе перейдите на вкладку *Главная* и пройдите по меню *Добавить в запросы* → *Добавить запросы в новый*. (рис. 3.4).

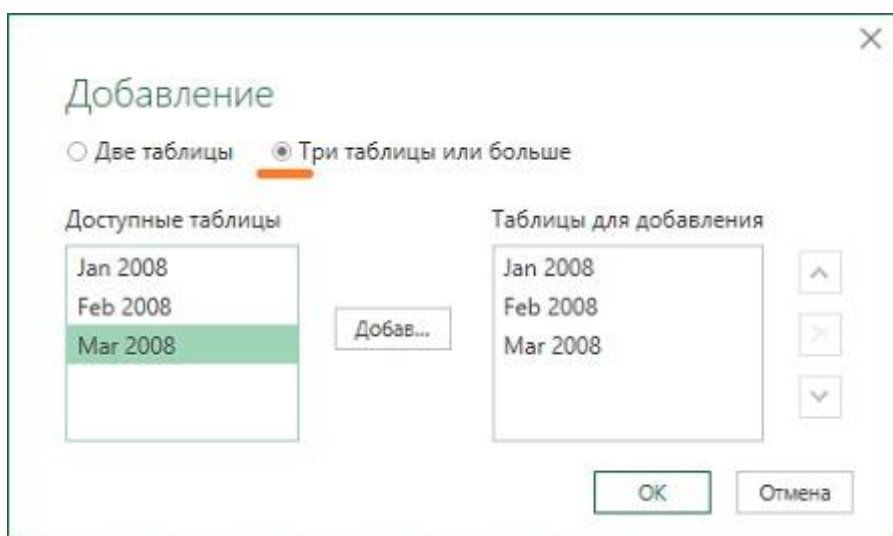


Рис. 3.3. Окно *Добавление*

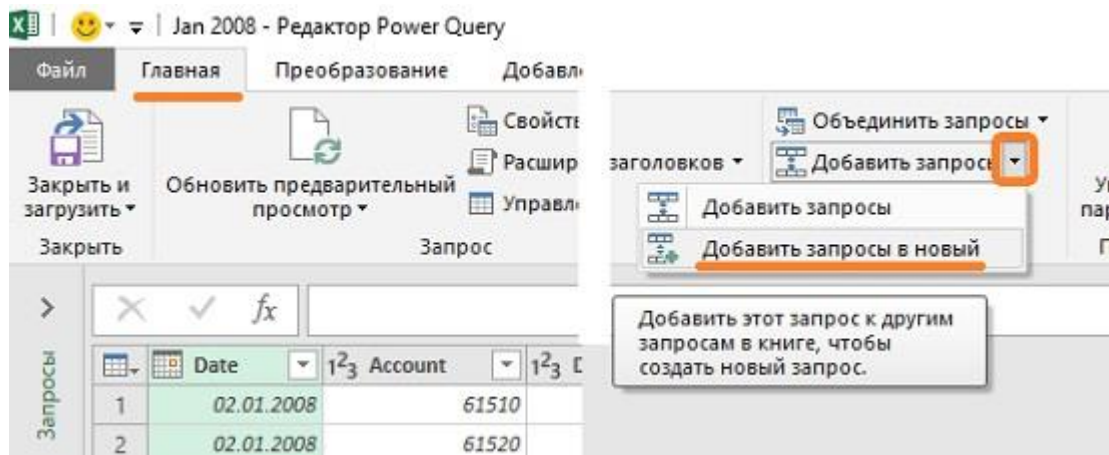


Рис. 3.4. Доступ к окну *Добавление* из редактора Power Query

Диалоговое окно *Добавление* объединяет запросы Power Query, а не таблицы Excel. Упорядочьте запросы в правом окне, чтобы данные располагались последовательно. Нажмите *Ok*. Power Query создаст новый запрос *Append1*, который включает один шаг:

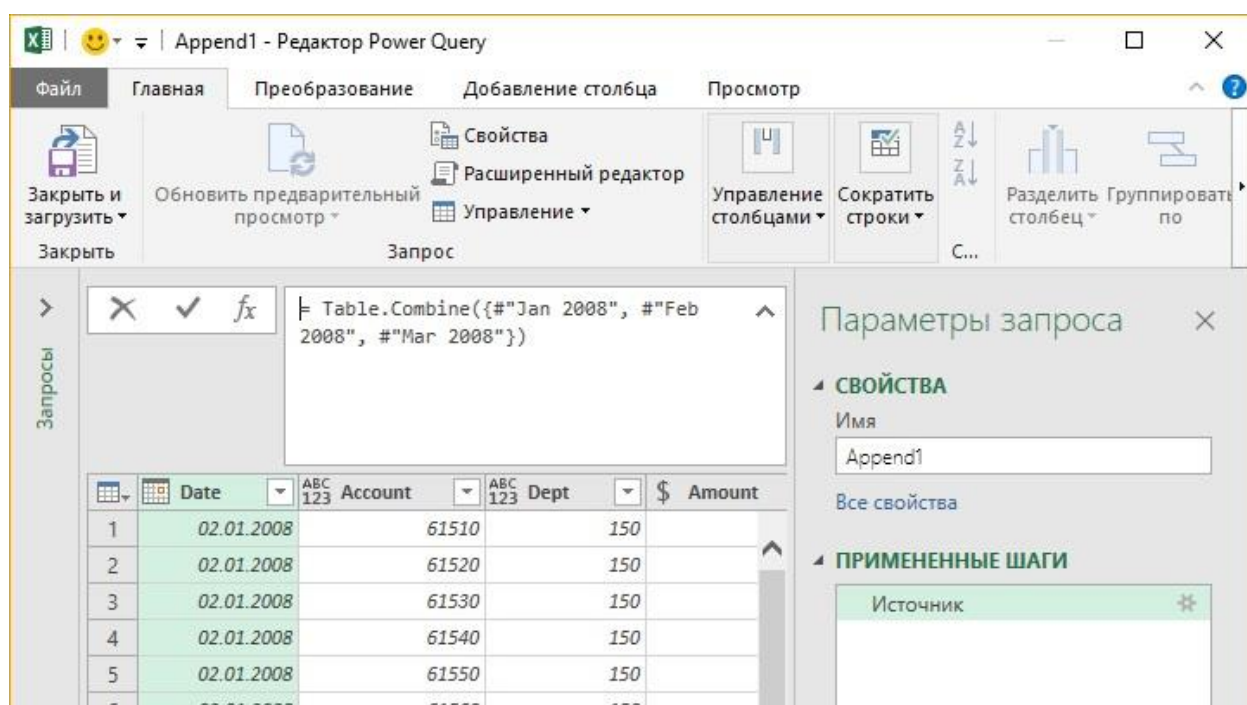


Рис. 3.5. Новый объединенный запрос *Append1*

У вас может возникнуть соблазн прокрутить запрос вниз, чтобы увидеть, все ли ваши записи вошли в него. К сожалению, это займет много времени, так как бегунок работает не так как вы привыкли в Excel. При перемещении вниз новые строки будут подгружаться довольно медленно. Причина в том, что Power Query может использоваться для обработки больших наборов данных. Представьте, что вы подключаетесь к набору данных, из 5 миллионов строк, но хотите вытащить записи только для отдела №150. Power Query осуществляет как бы «предварительный просмотр», который должен дать достаточно информации для определения ключевой структуры данных. Вы выполните преобразования в данных предварительного просмотра и создаёте шаблон. Во время загрузки всех строк Power Query обрабатывает этот шаблон, извлекая только необходимые записи. Это намного эффективнее, чем загрузка всех данных в книгу и последующая обработка каждой строки и столбца.

Но если вы не видите все данные, как вы проверите, что объединенный запрос корректен? Переименуйте запрос *Append1* → *Transactions*. Кликните *Закреть и загрузить*.

	A	B	C	D	
1	Date	Account	Dept	Amount	<p>Запросы и подкл..</p> <p>Запросы Подключения</p> <p>Запросов: 4</p> <p>Jan 2008 Загружено строк: 2 030.</p> <p>Feb 2008 Загружено строк: 1 857.</p> <p>Mar 2008 Загружено строк: 2 197.</p> <p>Transactions Загружено строк: 6 084.</p>
2	02.01.2008	61510	150	-26,03	
3	02.01.2008	61520	150	-55,07	
4	02.01.2008	61530	150	-10,6	
5	02.01.2008	61540	150	-0,29	
6	02.01.2008	61550	150	-48,02	
7	02.01.2008	61560	150	-1,35	
8	02.01.2008	61570	150	-77,04	
9	02.01.2008	62010	150	-305,95	
10	02.01.2008	62020	150	-95,15	
11	02.01.2008	62099	150	8,79	
12	02.01.2008	62510	120	-56,74	
13	02.01.2008	62520	120	-41,22	
14	02.01.2008	62530	120	-3,48	
15	02.01.2008	62550	120	-8,97	
16	02.01.2008	62560	120	-0,46	

Рис. 3.6. Новый запрос суммирует все строки трех запросов

Вы также можете создать сводную таблицу, и убедиться, что Excel корректно объединил запросы:

	A	B	C	D	E	F	G
1	Date	Account	Dept	Amount			
2	02.01.2008	61510	150	-26,03	Месяц	Amount	
3	02.01.2008	61520	150	-55,07	январь	\$89 791	
4	02.01.2008	61530	150	-10,6	февраль	\$56 211	
5	02.01.2008	61540	150	-0,29	март	\$23 820	
6	02.01.2008	61550	150	-48,02	Итого	\$169 822	
7	02.01.2008	61560	150	-1,35			
8	02.01.2008	61570	150	-77,04			

Рис. 3.7. Сводная таблица на основе данных из запроса *Transactions*

Объединение запросов с разными заголовками

Ниже показана ситуация, когда пользователь забыл переименовать столбец *TranDate* в запросе Mar 2008. При объединении запросов Jan 2008 и Mar 2008 получится:

	Date	Account	Dept	Amount	TranDate
2024	31.01.2008	86290	150	37,41	null
2025	31.01.2008	86310	150	44,96	null
2026	31.01.2008	86330	150	100,02	null
2027	31.01.2008	86350	150	20,5	null
2028	31.01.2008	86360	150	282,82	null
2029	31.01.2008	86400	170	0,03	null
2030	31.01.2008	86450	170	-105,16	null
2031	null	61510	150	-137,06	01.03.2008
2032	null	61520	150	-287,81	01.03.2008
2033	null	61530	150	-73,81	01.03.2008
2034	null	61540	150	-2,23	01.03.2008
2035	null	61550	150	-149,18	01.03.2008
2036	null	61560	150	-2,33	01.03.2008
2037	null	61570	150	-191,37	01.03.2008

Рис. 3.8. Столбец *TranDate*, полный нулевых значений в январе, и столбец *Date*, полный нулевых значений в марте

Чтобы исправить это, откройте запрос Mar 2008, переименуйте столбец *TranDate* → *Date*. Сохраните запрос Mar 2008. Откройте запрос *Transactions*. Как только вы откроете запрос, вы увидите, что он уже исправлен – столбец *TranDate* отсутствует. Чтобы поправить таблицу *Transactions* на листе Excel, просто обновите ее.