

Глава 14. Группировка и суммирование 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](#)

Иногда при работе с большими наборами данных перед анализом в сводных таблицах необходимо сгруппировать записи в источнике. Power Query позволяет группировать и суммировать данные. Чтобы продемонстрировать, как группировка работает в Power Query, рассмотрим продажи производителя футболок:

	A	B	C	D
1	Date	ProductName	ChannelName	Amount
2	06.03.2015	Cap Sleeve	Reseller	164,7
3	06.03.2015	Long Sleeve	Online	180,43
4	10.03.2015	Short Sleeve	Online	229,92
5	10.03.2015	Sleeveless	Online	118,4
6	11.03.2015	3/4 Sleeve	Store	184,02
7	12.03.2015	Short Sleeve	Online	230,67
8	13.03.2015	Short Sleeve	Online	133,98
9	14.03.2015	Sleeveless	Reseller	163,55
10	14.03.2015	3/4 Sleeve	Online	116,83
11	14.03.2015	Long Sleeve	Reseller	163,55

Рис. 14.1. Продажи футболок

Вы хотите вывести в таблице:

- Объем продаж в день по каждому каналу
- Продукт-лидер ежедневных продаж, в процентах от объема продаж

Подключение к данным

Откройте Grouping.xlsx. Кликните любую ячейку в *Таблице* на листе Sales → *Данные* → *Из таблицы/диапазона*. В редакторе Power Query щелкните правой кнопкой мыши столбец *Date* → *Тип изменения* → *Дата*. Поскольку при импорте таблицы Power Query автоматически присвоил данным в столбце *Date* тип Дата/Время, появится окно с предложением заменить текущее преобразование типов данных или добавить новое. Поскольку автозамена Power Query вам не нужна, кликните *Заменить текущее*.

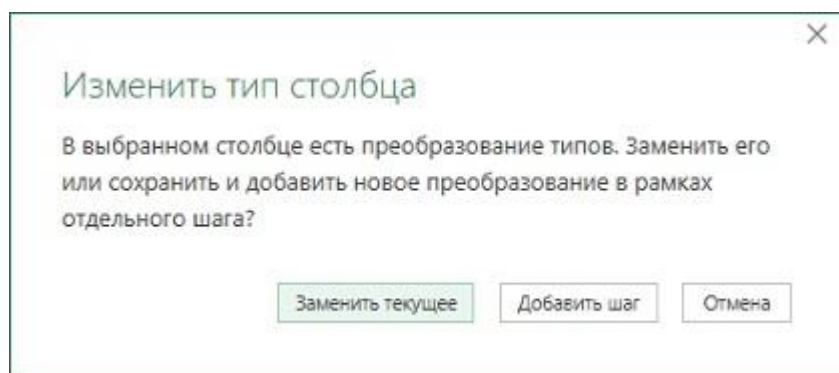


Рис. 14.2. Заменить или добавить шаг

Импортированная таблица подготовлена для группирования и суммирования данных:

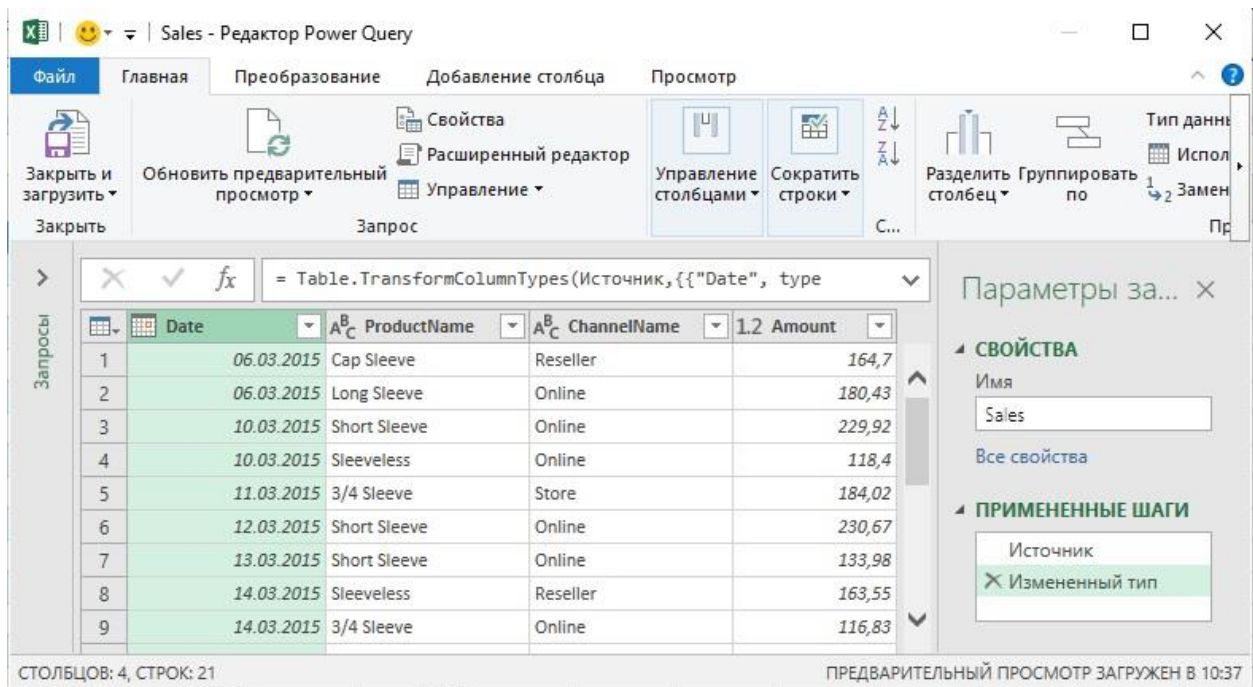


Рис. 14.3. Таблица готова к обработке

Группировка данных

Выберите столбец *Date* → Щелкните правой кнопки мыши *Группировать по...*

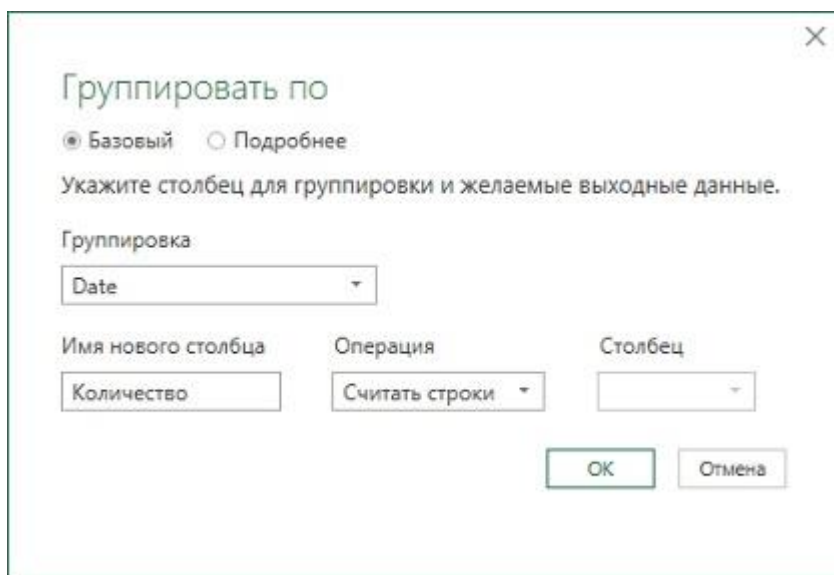


Рис. 14.4. Исходный вид окна *Группировать по*

В этом окне можно определить элементы, которые вы хотите сгруппировать, а также то, как вы хотите их сгруппировать. Power Query по умолчанию предлагает для группировки только столбец *Date*. Если вы оставите эту настройку в таком виде, у вас будут данные, сгруппированные только по дате, но не по каналу, поэтому вам нужно внести изменения. Кликните переключатель *Подробнее*, добавьте группировку по столбцу *ChannelName* (рис. 14.5). Поскольку вы хотите подсчитать общее количество каналов, а также общее количество продуктов по каналам, вам необходимо выполнить следующие шаги:

- Измените имя нового столбца на *Products Sold*
- Кликните кнопку *Добавить агрегирование*
- Дайте новому столбцу имя *Sales \$* и выберите операцию суммирования по столбцу *Amount*

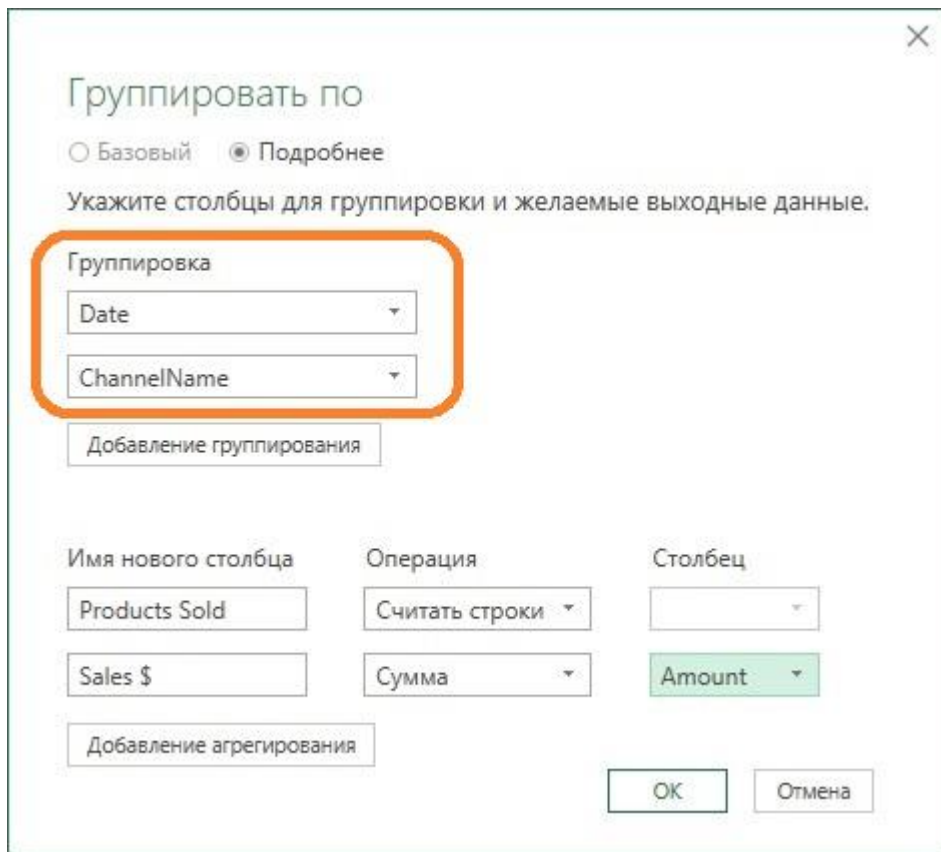


Рис. 14.5. Добавление уровней группировки и агрегирования столбцов
Кликните Ok. Данные сгруппируются:

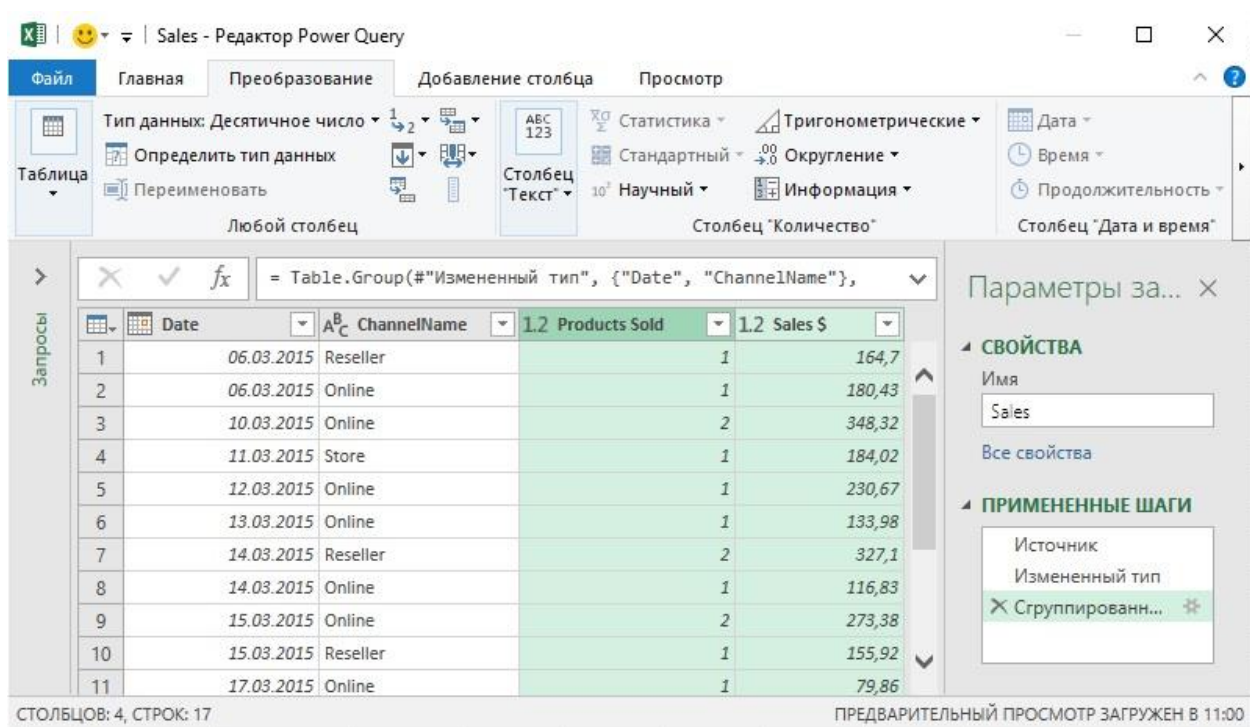


Рис. 14.6. Данные сгруппированы по дате и каналу

Обратите внимание, что данные можно агрегировать различными способами:

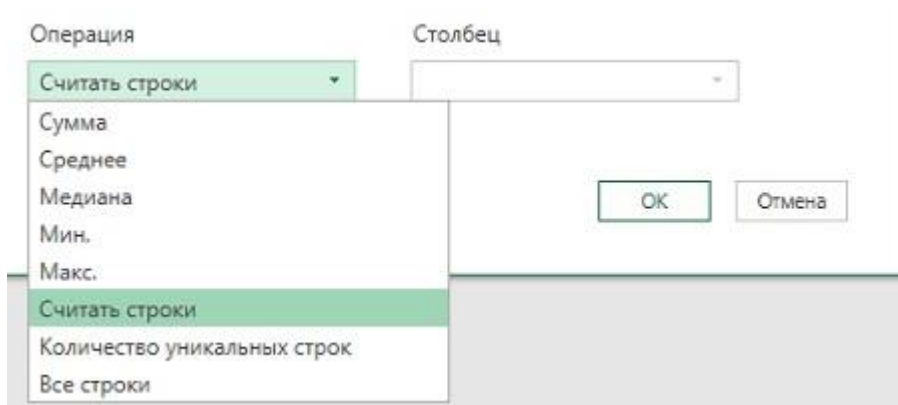


Рис. 14.7. Доступные функции агрегирования данных

В качестве дополнительного бонуса функция *Группировать по* удаляет из таблицы все столбцы, которые не были указаны в области *Группировка* (выделено на рис. 14.5). Теперь вы можете загрузить эти данные в *Таблицу* Excel. Измените имя запроса на *Grouped*. *Главная* → *Заккрыть и загрузить*.

Создание сводной статистики

Теперь вам нужно найти самый продаваемый продукт в каждом сегменте и какую долю в процентах от общего объема продаж он составляет. Чтобы подстраховаться на случай ошибок, создайте копию запроса. В книге Excel на панели *Запросы и подключения* щелкните правой кнопкой мыши запрос *Grouped* → *Дублировать*. Запрос откроется в редакторе Power Query. Переименуйте его *Performance*.

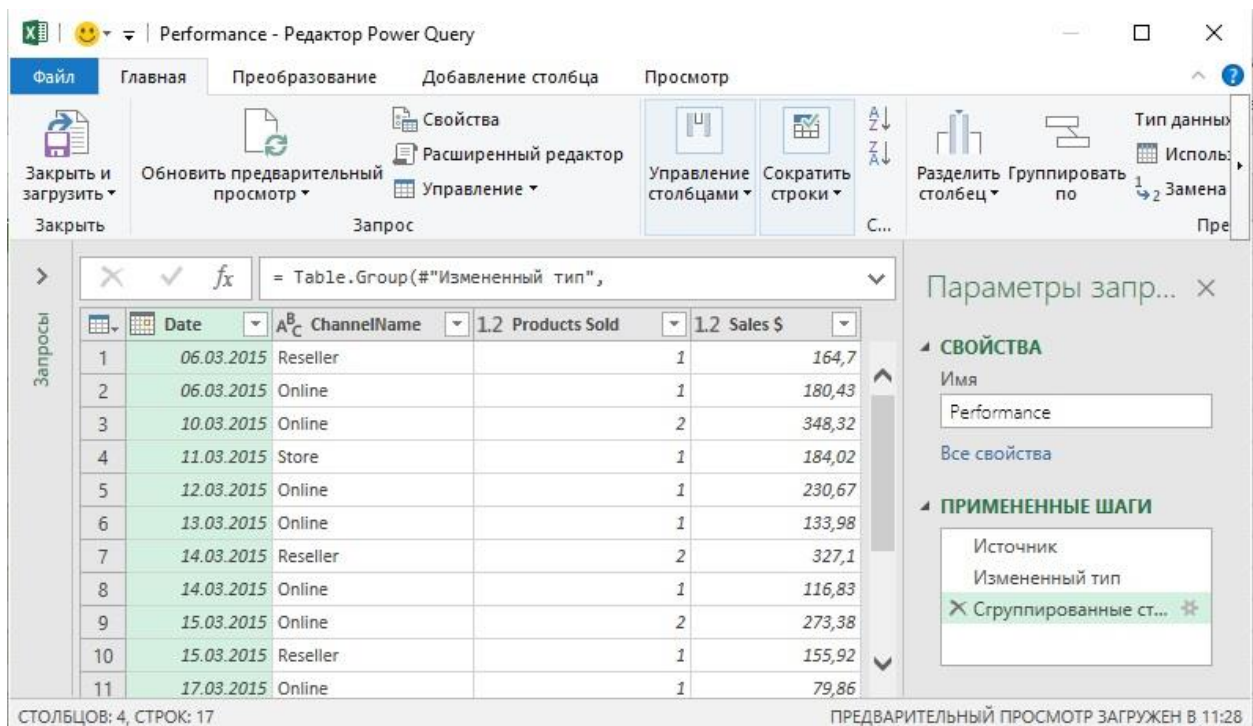


Рис. 14.8. Точная копия запроса *Grouped*

Определение топового продукта

Для этого нужно изменить шаг *Сгруппированные строки* и добавить еще один шаг. Нажмите на шестеренку рядом с шагом *Сгруппированные строки*. Откроется окно *Группировать по*. Добавьте новый столбец агрегирования – *Details*. Задайте для него операцию *Все строки*. Нажиме Ок.

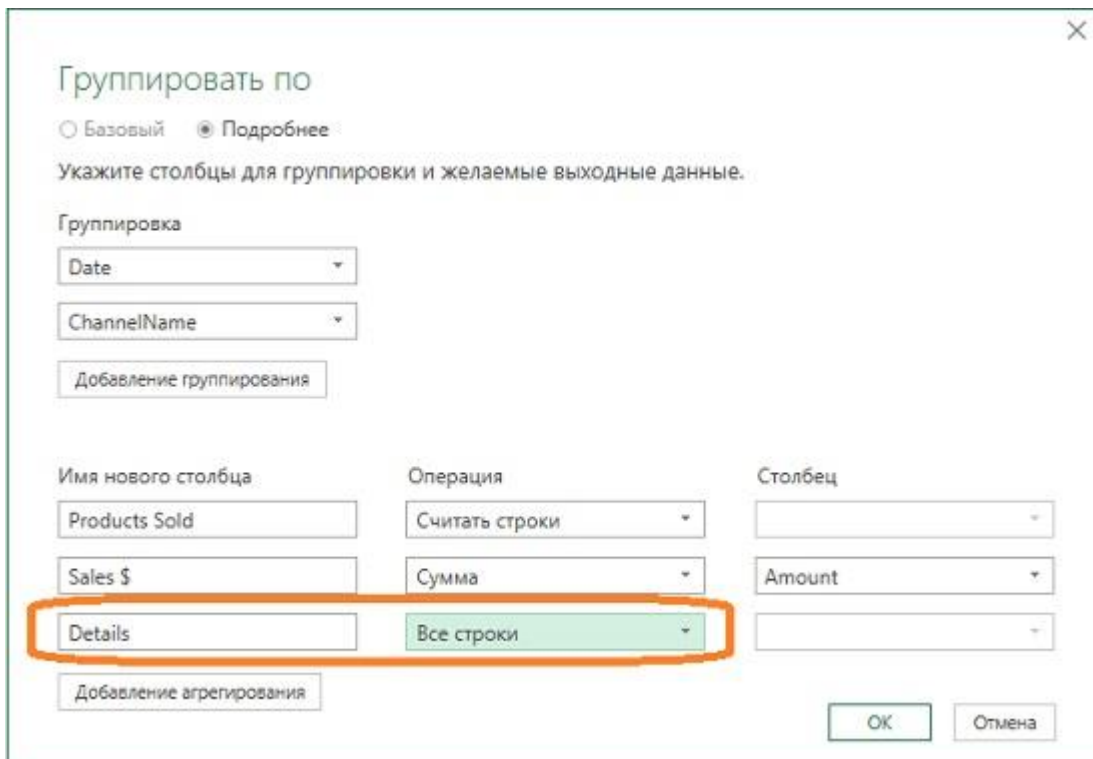


Рис. 14.9. Добавление столбца агрегирования

В таблице появился новый столбец – *Details* (см. рис. 14.10). Он весьма необычен. Каждая *Table* содержит сведения о том, какие строки из предыдущего шага были суммированы. Чтобы увидеть это, кликните пробел справа от слова *Table* (1). Внизу появится поле с новой таблицей (2).

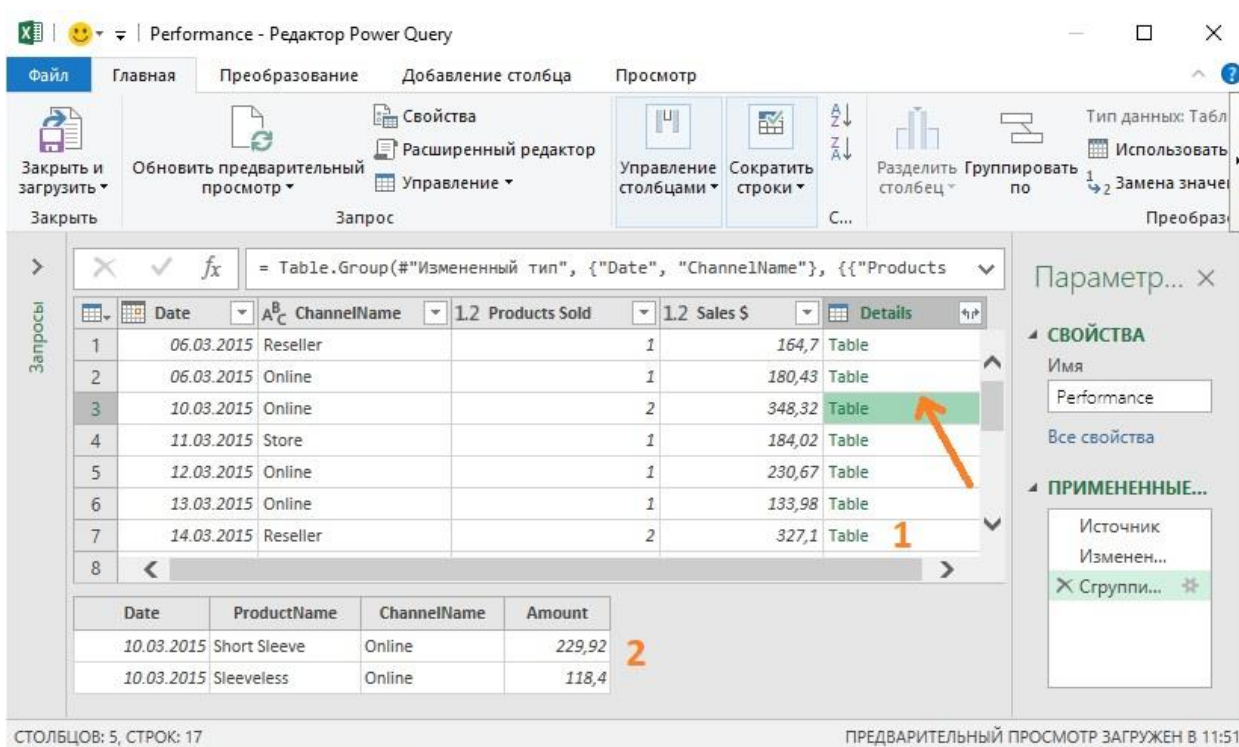


Рис. 14.10. Таблица внизу окна показывает, какие строки использовались для создания сгруппированных значений

Создайте пользовательский столбец с формулой. В редакторе Power Query перейдите на вкладку *Добавление столбца* → *Настраиваемый столбец*. Назовите столбец *MaxRecord*, используйте для него следующую формулу `=Table.Max([Details], "Amount")`. Нажимать *Ok*.

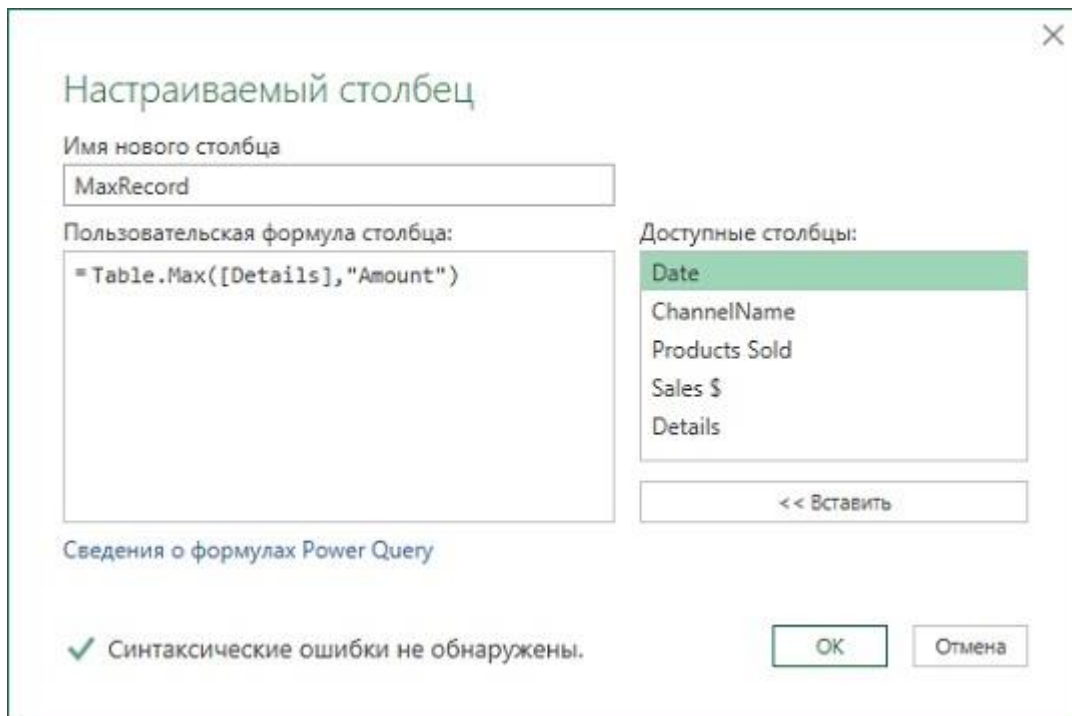


Рис. 14.11. Добавление пользовательского столбца

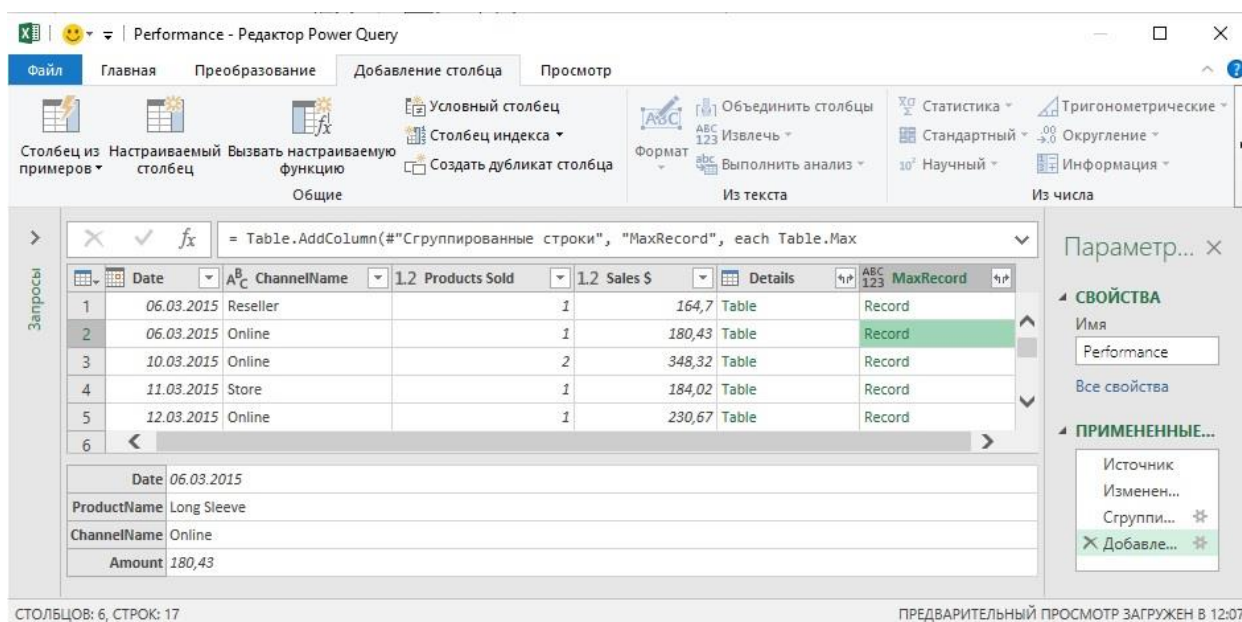


Рис. 14.12. Пользовательский столбец... Но что он означает?

Источник

Давайте вернемся на мгновение назад. Используемая формула (рис. 14.11) проверяет таблицу и извлекает максимальное значение. Поиск ведется в столбце *Details*, поскольку он содержит таблицы, которые необходимо проверить. Максимум берется по столбцу *Amount* таблицы. Имя *Amount* взято в кавычки по правилам синтаксиса функции. Однако, функция *Table.Max* возвращает не значение, а запись. Эта запись не просто содержит максимальное значение, она содержит все детали, соответствующие максимальному значению (вы узнаете больше о записях в главе 19).

Столбец *MaxRecord* можно развернуть с помощью кнопки с двуглавой стрелкой, чтобы поместить отдельные компоненты в столбцы таблицы запроса. Щелкните стрелку *Развернуть* в столбце *MaxRecord*. Выберите *ProductName* и *Amount*, отключите *Использовать исходное имя столбца как префикс*. Нажмите *Ок*.

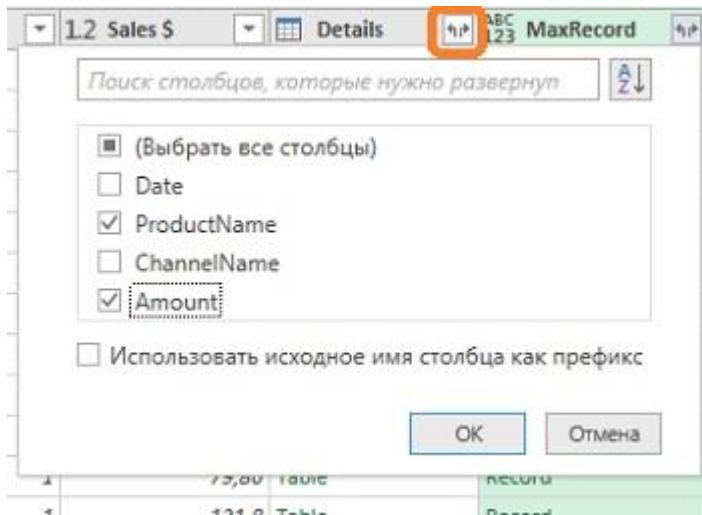


Рис. 14.13. Разворачивание столбца *MaxRecord* в столбцы *ProductName* и *Amount*

Щелкните правой кнопкой мыши столбец *Details* → *Удалить*. Щелкните правой кнопкой мыши столбец *ProductName* → *Переименовать* → *TopSeller*. Щелкните правой кнопкой мыши столбец *Amount* → *Переименовать* → *TopSeller \$*. В результате – у вас таблица, которая суммирует продажи по дате и каналу и показывает самый продаваемый товар, а также сумму его продаж:

	Date	ChannelName	1.2 Products Sold	1.2 Sales \$	TopSeller	TopSeller \$
1	06.03.2015	Reseller	1	164,7	Cap Sleeve	164,7
2	06.03.2015	Online	1	180,43	Long Sleeve	180,43
3	10.03.2015	Online	2	348,32	Short Sleeve	229,92
4	11.03.2015	Store	1	184,02	3/4 Sleeve	184,02
5	12.03.2015	Online	1	230,67	Short Sleeve	230,67
6	13.03.2015	Online	1	133,98	Short Sleeve	133,98
7	14.03.2015	Reseller	2	327,1	Long Sleeve	163,55
8	14.03.2015	Online	1	116,83	3/4 Sleeve	116,83
9	15.03.2015	Online	2	273,38	Long Sleeve	168,71
10	15.03.2015	Reseller	1	155,92	Cap Sleeve	155,92
11	17.03.2015	Online	1	79,86	3/4 Sleeve	79,86
12	18.03.2015	Reseller	1	121,8	Sleeveless	121,8
13	18.03.2015	Online	1	259,58	Sleeveless	259,58
14	24.03.2015	Store	1	238,28	Cap Sleeve	238,28

Рис. 14.14. Данные, сгруппированные по дате и каналу с данными о самом продаваемом продукте

Остается определить процентный вклад топового продукта в ежедневные продажи. Перейдите на вкладку *Добавление столбца* → *Настраиваемый столбец*. Переименуйте его в *TS % of Sales* и введите следующую формулу: $=\frac{[#"TopSeller \$"]}{[#"Sales \$"]}$.

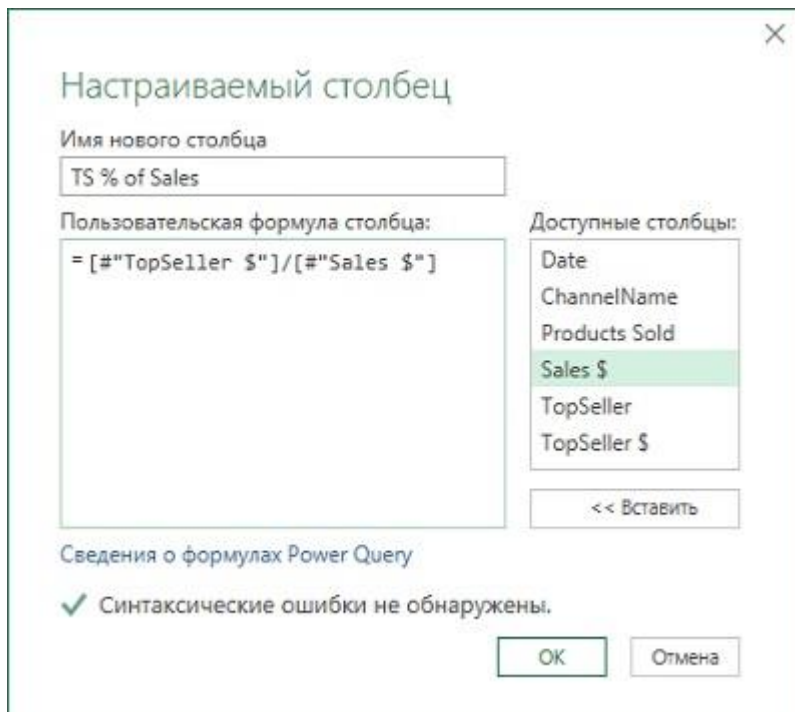


Рис. 14.15. Настраиваемый столбец *TS % of Sales*

Самый простой способ построить формулу – выбрать имя поля справа, и нажать *Вставить* (больше о вводе формул вы узнаете в главе 21). Округлите значения. Кликните правой кнопкой мыши на столбце *TS % of Sales* → *Преобразование* → *Округление* → *Округление* → *Число десятичных знаков* → 2.

Загрузите запрос в таблицу Excel. Назначьте столбцу *TS % of Sales* процентный формат:

	A	B	C	D	E	F	G
1	Date	ChannelName	Products Sold	Sales \$	TopSeller	TopSeller \$	TS % of Sales
2	06.03.2015	Reseller	1	164,7	Cap Sleeve	164,7	100%
3	06.03.2015	Online	1	180,43	Long Sleeve	180,43	100%
4	10.03.2015	Online	2	348,32	Short Sleeve	229,92	66%
5	11.03.2015	Store	1	184,02	3/4 Sleeve	184,02	100%
6	12.03.2015	Online	1	230,67	Short Sleeve	230,67	100%
7	13.03.2015	Online	1	133,98	Short Sleeve	133,98	100%
8	14.03.2015	Reseller	2	327,1	Long Sleeve	163,55	50%
9	14.03.2015	Online	1	116,83	3/4 Sleeve	116,83	100%
10	15.03.2015	Online	2	273,38	Long Sleeve	168,71	62%
11	15.03.2015	Reseller	1	155,92	Cap Sleeve	155,92	100%
12	17.03.2015	Online	1	79,86	3/4 Sleeve	79,86	100%
13	18.03.2015	Reseller	1	121,8	Sleeveless	121,8	100%
14	18.03.2015	Online	1	259,58	Sleeveless	259,58	100%
15	24.03.2015	Store	1	238,28	Cap Sleeve	238,28	100%
16	25.03.2015	Store	1	17,87	3/4 Sleeve	17,87	100%
17	25.03.2015	Online	1	216,69	Short Sleeve	216,69	100%
18	26.03.2015	Reseller	2	35,99	Short Sleeve	29,06	81%

Рис. 14.16. Итоговый анализ