**Глава 14. Группировка и суммирование в Power Query**

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

[Предыдущая глава](http://baguzin.ru/wp/?p=21124) [Содержание](http://baguzin.ru/wp/?p=20852#content) 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:

Изображение выглядит как снимок экрана

Автоматически созданное описание

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*.

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 14.2. Заменить или добавить шаг

Импортированная таблица подготовлена для группирования и суммирования данных:

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 14.3. Таблица готова к обработке

### Группировка данных

Выберите столбец *Date* –> Щелкните правой кнопки мыши *Группировать по…*

Изображение выглядит как снимок экрана

Автоматически созданное описание

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*

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 14.5. Добавление уровней группировки и агрегирования столбцов

Кликните Ok. Данные сгруппируются:

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 14.6. Данные сгруппированы по дате и каналу

Обратите внимание, что данные можно агрегировать различными способами:

Изображение выглядит как снимок экрана

Автоматически созданное описание

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* –> *Закрыть 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*.

Изображение выглядит как снимок экрана

Автоматически созданное описание

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.* Откроетсяокно *Группировать по*. Добавьте новый столбец агрегирования – *Details*. Задайте для него операцию *Все строки*. Нажиме Ok.

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 14.9. Добавление столбца агрегирования

В таблице появился новый столбец – *Details* (см. рис. 14.10). Он весьма необычен. Каждая *Table* содержат 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).

Изображение выглядит как снимок экрана

Автоматически созданное описание

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.

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 14.11. Добавление пользовательского столбца

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 14.12. Пользовательский столбце… 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 Ok.

Изображение выглядит как снимок экрана

Автоматически созданное описание

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:

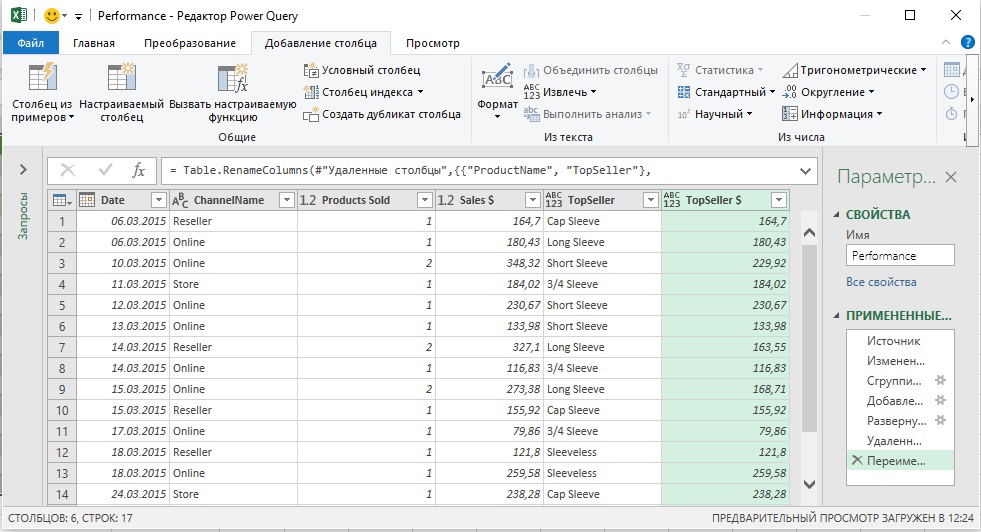


Рис. 14.14. Данные, сгруппированные по дате и каналу с данными о самом продаваемом продукте

Остается определить процентный вклад топового продукта в ежедневные продаж. Перейдите на вкладку *Добавление столбца* –> *Настраиваемый столбец*. Переименуйте его в *TS % of Sales* и введите следующую формулу: =[#"TopSeller $"]/[#"Sales $"].

Изображение выглядит как снимок экрана

Автоматически созданное описание

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:



Рис. 14.16. Итоговый анализ