

Разделение данных на таблицы поиска и фактов 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](#)

«Правильная» подготовка данных служит ключом к успеху их анализа. Если данные представлены в единой таблице, лучше разделить их на таблицу транзакций и одну или несколько справочных таблиц. Подробнее о пользе нескольких таблиц в модели данных см. [Альберто Феррари, Марко Руссо. Анализ данных при помощи Microsoft Power BI и Power Pivot для Excel.](#)

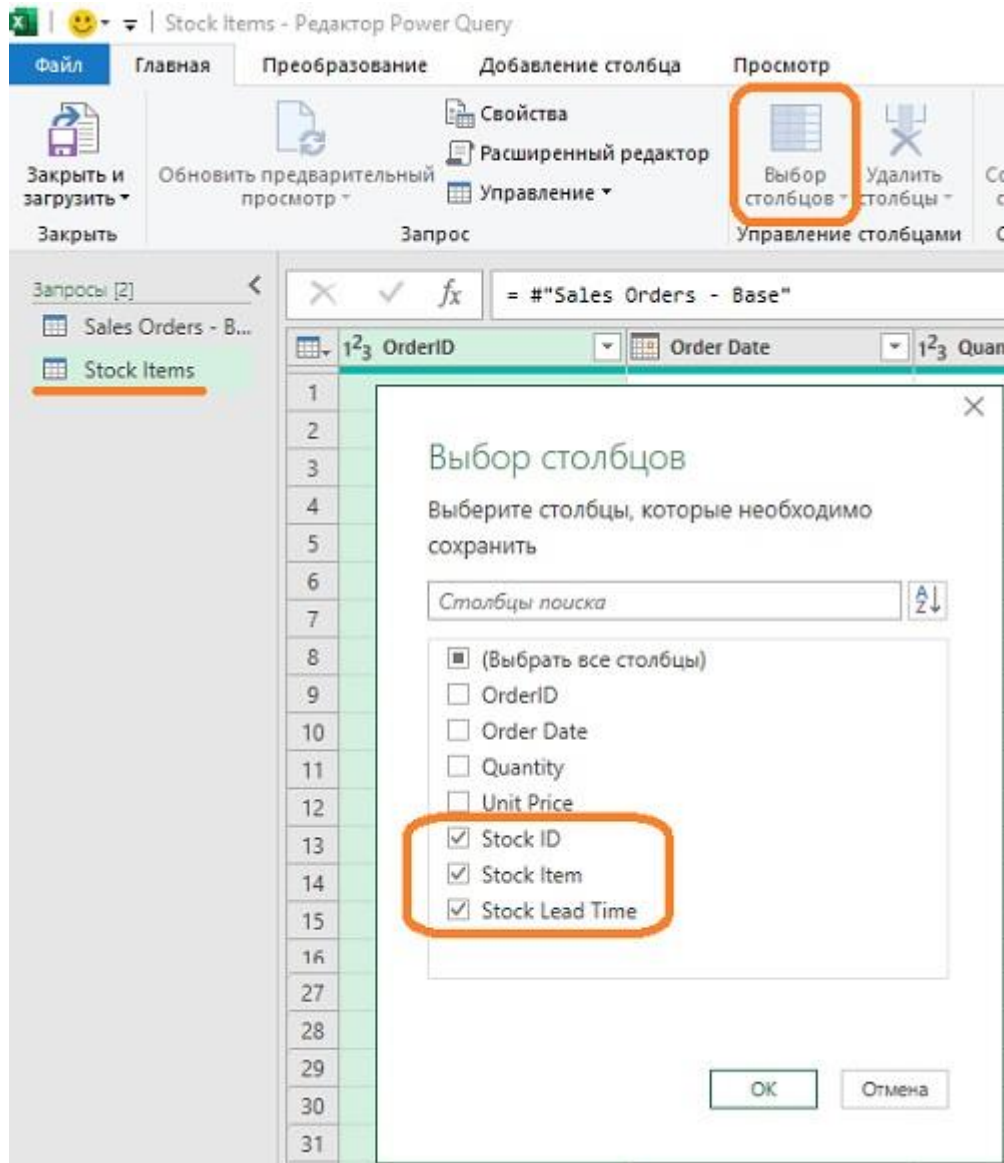


Рис. 1. Сохранить в таблице несколько столбцов

Мы используем Power Query для разделения исходной таблицы на таблицу фактов и таблицу поиска. Загрузите файл C02E07.xlsx с [сайта](#) или из приложенного к заметке архива и сохраните его на диске. Создайте новую книгу Excel. Пройдите по меню *Данные* → *Получить данные* → *Из файла* → *Из книги*. Выберите C02E07.xlsx. Нажмите *Импорт*. В окне *Навигатора* выберите *Sales_Order*, нажмите *Преобразовать данные*. Откроется окно редактора Power Query. Переименуйте запрос *Sales_Order* в *Sales Order - Base*. В левой части окна на панели *Запросы* щелкните правой кнопкой мыши на запросе *Sales Order - Base* и в контекстном меню выберите команду *Ссылка*. Переименуйте новый запрос на *Stock Items*.

Перейдите на вкладку *Главная*, кликните *Выбор столбцов*. В окне *Выбор столбцов* оставьте активными столбцы *Stock ID*, *Stock Item* и *Stock Lead Time* (см. рис. 1). Кликните *Ok*. Выберите

столбец *Stock ID*, и пройдите по меню *Главная* → *Удалить строки* → *Удалить дубликаты*. Вы сформировали таблицу поиска для товаров на складе. На самом деле, некоторые *Stock ID* всё еще могут быть не уникальными. Об этом ниже.

Теперь создадим таблицу фактов. Снова щелкните правой кнопкой мыши на запросе *Sales Order - Base* и выберите *Ссылка*. Переименуйте новый запрос в *Sales Orders*. Повторите операцию выбора столбцов, оставив все, кроме *Stock Item* и *Stock Lead Time*.

На вкладке *Главная* выберите команду *Заккрыть и загрузить в...* Выберите две опции: *Только создать подключение* и *Добавить эти данные в модель данных*:

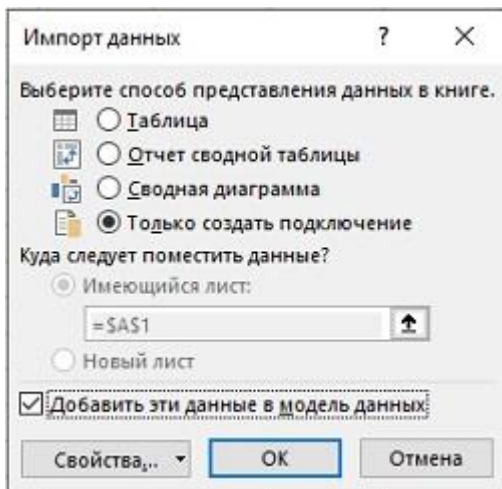


Рис. 2. Импорт в модель данных

Поскольку запрос *Sales Order - Base* является промежуточным, исключим его из модели данных. В Excel, в области *Запросы и подключения* кликните на запрос *Sales Order - Base* правой кнопкой мыши, и выберите *Загрузить в...* В окне *Импорт данных* снимите галочку *Добавить эти данные в модель данных*.

В Excel пройдите по меню *Power Pivot* → *Управление*. На вкладке *Главная* кликните *Представление диаграммы*. Протяните связь по полю *Stock ID* от таблицы *Sales Orders* к *Stock Items*.

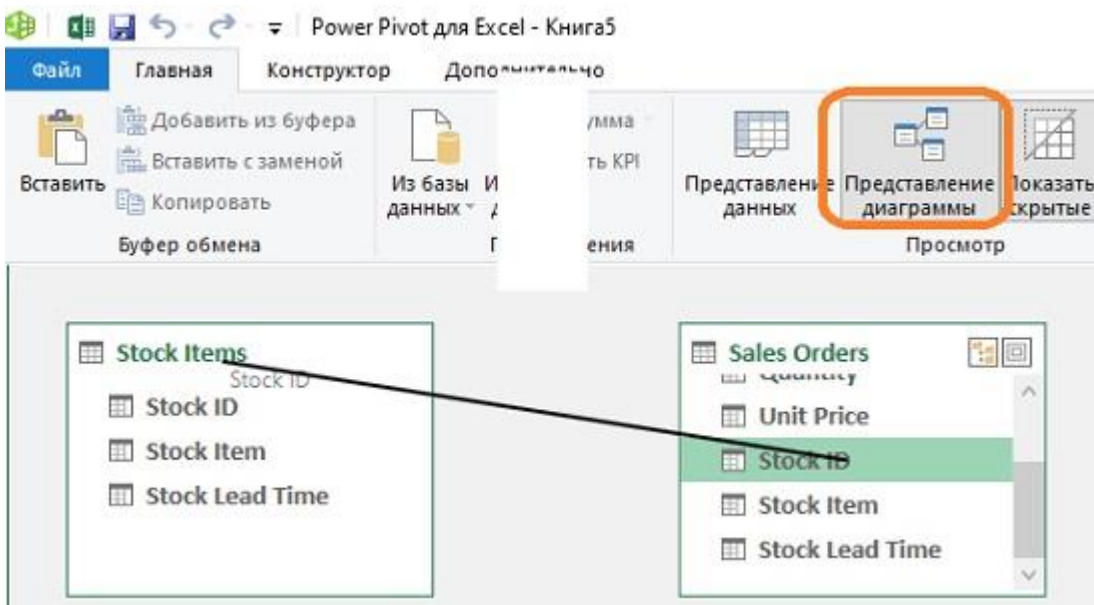


Рис. 3. Связывание таблицы фактов с таблицей поиска

Когда связи некорректны

Если при формировании таблицы поиска в Power Query в качестве ключа для связи используется текстовый столбец (а *Stock ID* в примере выше именно такой) необходимо выполнить два дополнительных шага, прежде чем удалять дубликаты.

Power Query и Power Pivot по-разному чувствительны к регистру. Для PQ записи "Stock_1", "STOCK_1" и "Stock_1 " (с завершающим пробелом) – разные, а для Power Pivot – одинаковые. Т.о., PQ оставит такие записи после команды *Удалить дубликаты*. А PP не поддерживает связь многие к одному, сочтя, что в таблице *Stock Items* записи не уникальны.

Чтобы устранить проблему, вернитесь в окно Excel. В области *Запросы и подключения* кликните на запрос *Sales Order - Base* правой кнопкой мыши и выберите *Изменить*. В окне редактора PQ выделите столбец *Stock ID*, на вкладке *Преобразование* кликните на кнопку *Формат* и далее *нижний регистр*. Можно также выбрать *ВЕРХНИЙ РЕГИСТР* или *Каждое Слово С Прописной*. Любая из этих команд гарантирует, что ключи будут нормализованы.

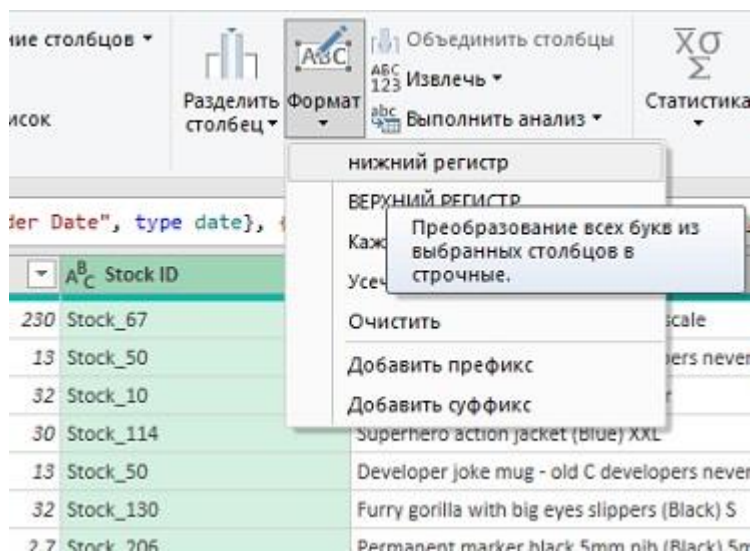


Рис. 4. Преобразование регистра

Выделите столбец *Stock ID* и пройдите по меню *Преобразование* → *Формат* → *Усечь*. Все пробелы в начале и в конце ячеек в столбце *Stock ID* будут удалены. На вкладке *Главная* кликните кнопку *Заккрыть и загрузить*. Вы вернетесь в окно Excel. На вкладке *Данные* кликните кнопку *Обновить все*.

Разбиение на строки значений с разделителями

Рассмотрим специфическое разбиение исходных данных на таблицы фактов и поиска. В таблице *Products* у товара может быть один или несколько цветов. Цвета перечислены через запятые:

	A	B	C	D
1	ProductCodeNew	Average Cost	Average Price	Colors
2	T44	755,1508	1214,85	Blue
3	T79	1481,9379	2384,07	Yellow, Blue
4	T18	461,4448	742,35	Blue, Yellow
5	R79	1082,51	1700,99	Yellow
6	R19	343,6496	539,99	Black
7	M38	419,7784	769,49	Silver
8	M18	301,3988	552,49	Silver, Black
9	M68	1258,8004	2307,49	Silver, Black
10	R89	1541,387338	2443,35	Red, Black
11	R64	713,0798	1120,49	Yellow
12	M47	598,4354	1079,99	Black
13	R93	2171,2942	3578,27	Red
14	R68	884,7083	1457,99	Red
15	M82	1905,1244	3387,49	Silver, Black
16	R50	486,7066	782,99	Red, Black

Рис. 5. Исходная таблица *Products*

Загрузите и сохраните на диске файл *C02E08.xlsx*. Создайте новую книгу Excel. Пройдите по меню *Данные* → *Получить данные* → *Из файла* → *Из книги*. Выберите *C02E08.xlsx*. Нажмите *Импорт*. В окне *Навигатора* выберите *Products*, нажмите *Преобразовать данные*. В левой части окна

редактора Power Query на панели *Запросы* щелкните правой кнопкой мыши на *Products* и выберите *Ссылка*. Переименуйте новый запрос в *Products and Colors*. Перейдите на вкладку *Главная*, кликните *Выбор столбцов*. Оставьте столбцы *ProductCodeNew* и *Colors*. Нажмите *Ok*. Для столбца *Colors* пройдите по меню *Преобразование* → *Разделить столбец* → *По разделителю*:

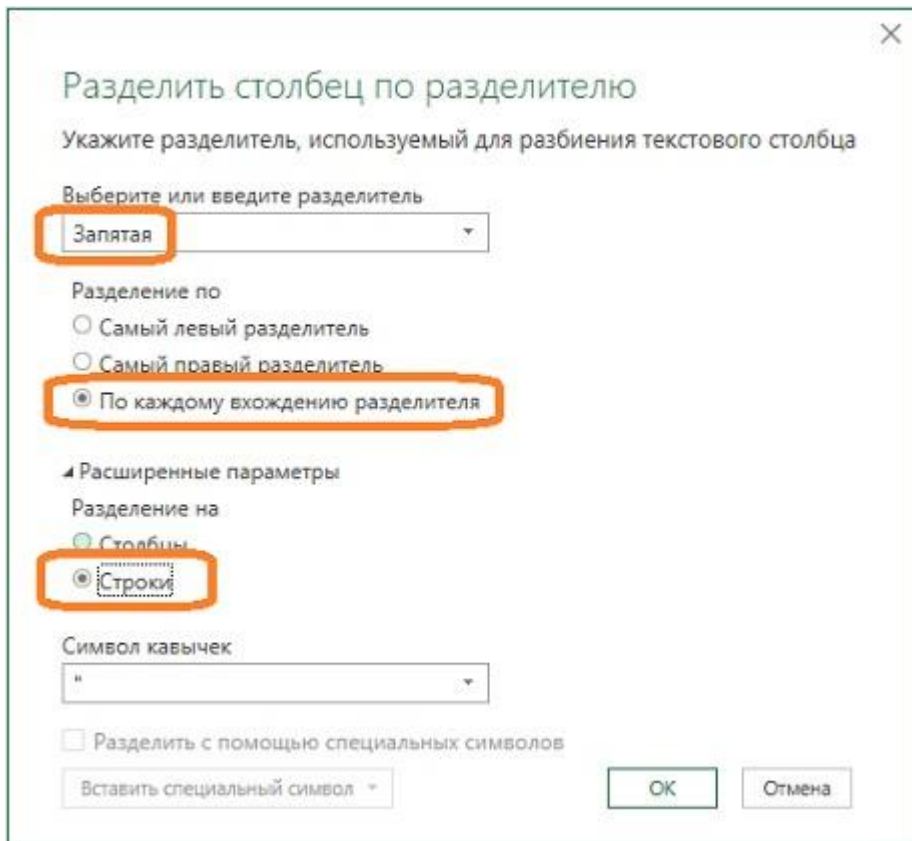


Рис. 6. Настройка *Разделить столбец по разделителю*

Для столбца *Colors* выполните команду *Преобразование* → *Формат* → *Усечь*. Загрузите обе таблицы в модель данных. В Power Pivot свяжите две таблицы:

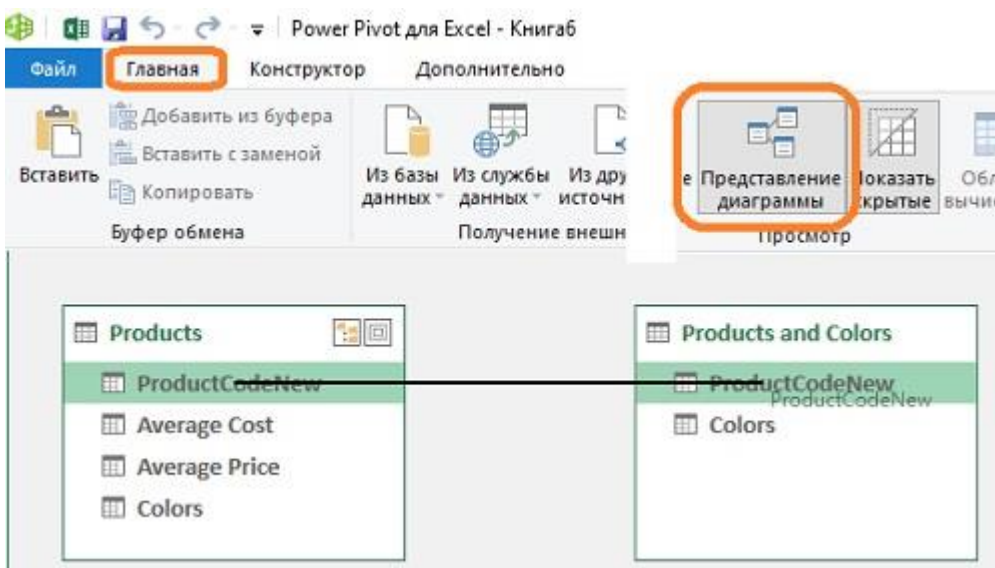


Рис. 7. Настройка связей в Power Pivot

Для уменьшения размеров отчета можно удалить столбец *Colors* из запроса *Products*. Однако это нельзя сделать прямо сейчас, поскольку на запрос *Products* ссылается запрос *Products and Colors*. Это затруднение можно обойти, если создать дубль запроса *Products*, назвав его, например, *Products1*. В этом новом запросе удалите столбец *Colors*, и именно этот запрос загрузите в модель данных, установив связь с таблицей *Products and Colors*. А для запроса *Products* оставьте опцию *Только создать подключение*.

Теперь мы можем визуализировать наши данные, например, так:

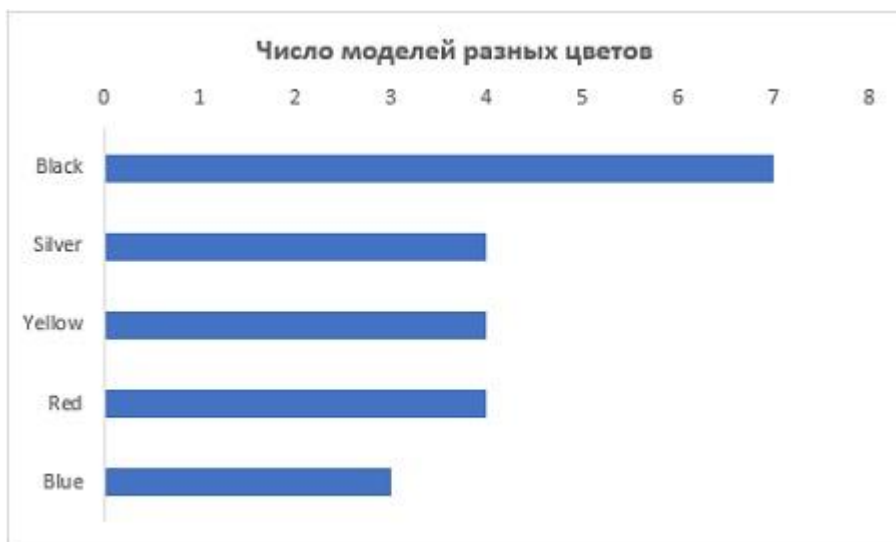


Рис. 8. Количество моделей разных цветов