

Power Query в Excel и Power BI сбор, объединение и преобразование данных

Гил Равив



Гил Равив

Power Query в Excel и Power BI: сбор, объединение и преобразование данных

Санкт-Петербург

«БХВ-Петербург»

2021

УДК 004.65
ББК 32.973.26-018.2
P13

Равив Г.

P13 Power Query в Excel и Power BI: сбор, объединение и преобразование данных: Пер. с англ. — СПб.: БХВ-Петербург, 2021. — 480 с.: ил.

ISBN 978-5-9775-6725-1

В книге описан Power Query, инструмент для продвинутого бизнес-анализа, встроенный в Excel, Power BI и другие продукты Microsoft и предназначенный для подключения к источникам данных и их преобразования. Рассказано, как обрабатывать таблицы Excel без использования сложных формул и макросов, выбирать данные из нескольких источников, объединять несовпадающие таблицы, сохранять контекст при формировании запросов. Особое внимание уделено базовым и продвинутым методам отмены свертывания столбцов (Unpivot) и совместной работе над отчетами. Дано введение в язык формул M и показано, как правильно формировать запросы, избегая сбоев и «ловушек». Рассмотрены вопросы текстовой аналитики и извлечения знаний из текстов, а также аналитики социальных сетей. В конце книги приведен заключительный проект, объединяющий и закрепляющий полученные навыки.

Для аналитиков данных и пользователей Excel

УДК 004.65
ББК 32.973.26-018.2

Группа подготовки издания:

Руководитель проекта	<i>Евгений Рыбаков</i>
Зав. редакцией	<i>Екатерина Сависте</i>
Перевод с английского	<i>Александра Сергеева</i>
Компьютерная верстка	<i>Ольги Сергиенко</i>
Оформление обложки	<i>Карины Соловьевой</i>

© 2021 BHV

Authorized translation from the English language edition, entitled COLLECT, COMBINE, AND TRANSFORM DATA USING POWER QUERY IN EXCEL AND POWER BI, 1st Edition; ISBN 978-1-5093-0795-1; by GIL RAVIV; published by Pearson Education, Inc, publishing as Microsoft Press, Copyright © 2019 by Gil Raviv.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc. RUSSIAN language edition published by BHV, Copyright © 2021.

© 2021 BHV

Авторизованный перевод англоязычного издания под названием COLLECT, COMBINE, and TRANSFORM DATA USING POWER QUERY IN EXCEL AND POWER BI, 1st Edition; ISBN 978-1-5093-0795-1; GIL RAVIV, опубликованного Pearson Education, Inc и подготовленного к изданию Microsoft Press, Copyright © Gil Raviv, 2019. Все права защищены. Никакая часть этой книги не может быть воспроизведена или передана в любой форме или любыми средствами, будь то электронными или механическими, включая фотокопирование, запись или любую информационно-поисковую систему, без разрешения Pearson Education, Inc. Русский перевод опубликован ООО "БХВ".

Подписано в печать 02.02.21.

Формат 70×100^{1/16}. Печать офсетная. Усл. печ. л. 38,7.

Тираж 1500 экз. Заказ № 7434.

"БХВ-Петербург", 191036, Санкт-Петербург, Гончарная ул., 20.

Отпечатано в ПАО «Можайский полиграфический комбинат».

143200, Россия, г. Можайск, ул. Мира, 93.

www.oaomprk.ru, тел.: (495) 745-84-28, (49638) 20-685



ISBN 978-1-5093-0795-1 (англ.)
ISBN 978-5-9775-6725-1 (рус.)

© Gil Raviv, 2019
© Перевод на русский язык, оформление.
ООО "БХВ-Петербург", ООО "БХВ", 2021

Оглавление

Вступительное слово	13
Об авторе	15
Благодарности	17
Введение	19
Для кого предназначена эта книга	20
Что нужно знать	20
Структура книги.....	21
О сопровождающем контенте.....	23
Требования к системе.....	28
Как получить поддержку и оставить отзыв	29
Ошибки и поддержка книги	29
Будьте на связи	29
Глава 1. Введение в Power Query	31
Знакомство с Power Query.....	32
Краткий очерк истории становления Power Query	33
Где "прячется" Power Query?.....	37
Основные компоненты Power Query	38
Интерфейс <i>Получить данные</i>	38
Основные панели редактора Power Query	40
Панель просмотра.....	41
Лента.....	42
Панель <i>Запросы</i>	43
Панели <i>Параметры запроса</i> и <i>Примененные шаги</i>	44
Строка формул, расширенный редактор и язык запросов M.....	44
Окна диалога <i>Параметры запроса</i> в Excel и <i>Параметры</i> в Power BI.....	45
УПРАЖНЕНИЕ 1.1. Знакомство с Power Query	46
Резюме	52
Глава 2. Основные проблемы, возникающие при подготовке данных	53
Извлечение значения из кодированных столбцов.....	54
База данных AdventureWorks.....	55
УПРАЖНЕНИЕ 2.1. Старый способ: применение формул Excel.....	56
УПРАЖНЕНИЕ 2.2. Часть 1. Новый способ.....	57
УПРАЖНЕНИЕ 2.2. Часть 2. Объединение таблиц поиска.....	61
УПРАЖНЕНИЕ 2.2. Часть 3. Таблицы фактов и таблицы поиска.....	66

Применение набора функций из раскрывающегося меню <i>Столбец из примеров</i>	68
УПРАЖНЕНИЕ 2.3. Часть 1. Знакомство с функцией <i>Столбец из примеров</i>	69
Практическое применение набора функций <i>Столбец из примеров</i>	71
УПРАЖНЕНИЕ 2.3. Часть 2. Преобразование размера в группы/диапазоны	72
Выборка информации из текстовых столбцов	75
УПРАЖНЕНИЕ 2.4. Извлечение гиперссылок из сообщений	76
Работа с датами	84
УПРАЖНЕНИЕ 2.5. Обработка нескольких форматов для дат	85
УПРАЖНЕНИЕ 2.6. Работа с датами при использовании двух локальных настроек	87
Выборка элементов даты и времени	90
Подготовка модели данных	91
УПРАЖНЕНИЕ 2.7. Разделение данных на таблицы поиска и таблицы фактов	92
Когда связи некорректны	94
УПРАЖНЕНИЕ 2.8. Разбиение на строки значений с разделителями	95
Резюме	98
Глава 3. Объединение данных, полученных из нескольких источников	99
Объединение нескольких таблиц	99
Объединение двух таблиц	100
УПРАЖНЕНИЕ 3.1. Пример с таблицами <i>Vikes</i> и <i>Accessories</i>	100
УПРАЖНЕНИЕ 3.2. Часть 1. Применение команды <i>Объединить запросы в новый</i>	103
УПРАЖНЕНИЕ 3.2. Часть 2. Окна диалога <i>Зависимости запроса</i> и <i>Ссылки</i>	104
Объединение трех и большего числа таблиц	107
УПРАЖНЕНИЕ 3.2. Часть 3. <i>Vikes + Accessories + Components</i>	107
УПРАЖНЕНИЕ 3.2. Часть 4. <i>Vikes + Accessories + Components + Clothing</i>	109
Объединение большого количества таблиц	110
Добавление таблиц из папки	110
УПРАЖНЕНИЕ 3.3. Добавление таблицы <i>AdventureWorks Products</i> из папки	111
Команда <i>Импорт из папки</i>	113
Добавление листов из книги	114
УПРАЖНЕНИЕ 3.4. Добавление рабочих листов — решение	114
Надежный подход к объединению нескольких листов	119
Резюме	120
Глава 4. Объединение несовпадающих таблиц	123
Трудности, вызванные несоответствием таблиц	124
Что такое несоответствующие таблицы?	124
Особенности и риски, связанные с несоответствующими таблицами	124
УПРАЖНЕНИЕ 4.1. Устранение несоответствующих названий столбцов: простейший подход	126
Объединение несоответствующих таблиц из одной папки	127
УПРАЖНЕНИЕ 4.2. Часть 1. Иллюстрация особенности, связанной с пропущенными значениями	127
УПРАЖНЕНИЕ 4.2. Часть 2. Предположение о том же порядке столбцов и обобщение заголовка	130
УПРАЖНЕНИЕ 4.3. Простая нормализация, выполняемая с помощью функции <i>Table.TransformColumnNames</i>	132

Таблица преобразования	135
УПРАЖНЕНИЕ 4.4. Методика транспонирования, реализуемого с помощью таблицы преобразования	137
Преобразование транспонирования	138
Последовательность объединения	139
Обратное транспонирование	141
УПРАЖНЕНИЕ 4.5. Отмена свертывания, слияние и повторное сведение	142
УПРАЖНЕНИЕ 4.6. Транспонирование лишь имен столбцов	144
УПРАЖНЕНИЕ 4.7. Применение языка M для нормализации имен столбцов	150
Резюме	153
Глава 5. Сохранение контекста	155
Сохранение контекста в именах файлов и листов	156
УПРАЖНЕНИЕ 5.1. Часть 1. Методика, предусматривающая применение настраиваемого столбца	156
УПРАЖНЕНИЕ 5.1. Часть 2. Обработка контекста, состоящего из имен файлов и листов	158
Предварительное добавление для сохранения заголовков	159
УПРАЖНЕНИЕ 5.2. Сохранение заголовков с помощью детализации	160
Применение детализации к ячейке таблицы (объяснение с помощью M-формулы)	161
Объединение в запросе результатов детализации	163
УПРАЖНЕНИЕ 5.3. Сохранение заголовков из папки	165
Сохранение заголовков при постдобавлении контекста	167
УПРАЖНЕНИЕ 5.4. Сохранение заголовков из рабочих листов в одной книге	168
Использование подсказок контекста	173
УПРАЖНЕНИЕ 5.5. Применение столбца индекса в качестве подсказки	174
УПРАЖНЕНИЕ 5.6. Идентификация контекста с учетом близости ячейки	177
Резюме	182
Глава 6. Отмена свертывания столбцов таблицы	185
Идентификация плохо спроектированных таблиц	186
Знакомство с преобразованием отмены свертывания столбцов	189
УПРАЖНЕНИЕ 6.1. Применение преобразований отмены свертывания столбцов и отмены свертывания других столбцов	190
УПРАЖНЕНИЕ 6.2. Использование команды <i>Отменить свертывание только выделенных столбцов</i>	193
Обработка итогов	195
УПРАЖНЕНИЕ 6.3. Применение преобразования отмены свертывания столбцов к общим итогам	195
Удаление общих итогов	198
Применение преобразования отмены свертывания столбцов на уровнях иерархии 2×2	199
УПРАЖНЕНИЕ 6.4. Применение преобразования отмены свертывания таблиц для уровней иерархии 2×2 с датами	200
УПРАЖНЕНИЕ 6.5. Применение преобразования отмены свертывания столбцов к таблицам с уровнями иерархии 2×2	203
Обработка промежуточных итогов при работе с данными, для которых отменено свертывание столбцов	206
УПРАЖНЕНИЕ 6.6. Обработка промежуточных итогов	206
Резюме	209

Глава 7. Дополнительные вопросы, связанные с применением преобразования сведения столбцов и отмены свертывания столбцов	211
Применение преобразования отмены свертывания столбцов к таблицам с несколькими уровнями иерархии	212
Виртуальная сводная таблица, определение полей строк и столбцов	212
УПРАЖНЕНИЕ 7.1. Применение преобразования отмены свертывания столбцов к таблице AdventureWorks с уровнями иерархии $N \times M$	214
Обобщение последовательности преобразований отмены свертывания столбцов	217
УПРАЖНЕНИЕ 7.2. Начнем с конца	217
УПРАЖНЕНИЕ 7.3. Создание функции <i>FnUnpivotSummarizedTable</i>	219
Часть 1. Создание отчетов RowFields, ColumnFields и ValueField	220
Часть 2. Удаление шагов <i>Измененный min</i>	220
Часть 3. Функции <i>Table.ColumnNames</i> , <i>List.FirstN</i> и <i>List.Count</i>	222
Часть 4. Переименование столбцов и функция <i>List.Zip</i>	225
Часть 5. Преобразование запроса в функцию	227
Часть 6. Тестирование функции	230
Выводы	230
Преобразование сведения в столбец	231
УПРАЖНЕНИЕ 7.4. Обращение таблицы с некорректной отменой свертывания столбцов	231
УПРАЖНЕНИЕ 7.5. Применение преобразования сведения в столбец к таблицам с многострочными записями	234
Часть 1. Работа с фиксированным числом атрибутов и применение целочисленного деления	235
Часть 2. Обработка нефиксированного числа атрибутов с помощью команд <i>Условный столбец</i> и <i>Заполнение вниз</i>	236
Резюме	239
Глава 8. Совместная работа	241
Локальные файлы, параметры и шаблоны	242
Ошибка при доступе к локальным файлам	242
УПРАЖНЕНИЕ 8.1. Применение параметра для имени пути	243
УПРАЖНЕНИЕ 8.2. Создание шаблона Power BI	245
УПРАЖНЕНИЕ 8.3. Часть 1. Применение параметров в Excel	248
УПРАЖНЕНИЕ 8.3. Часть 2. Изменения в комбинациях данных	252
Работа с общедоступными файлами и папками	256
Импортирование данных из файлов в OneDrive for Business или SharePoint	256
УПРАЖНЕНИЕ 8.4. Миграция запросов для подключения к OneDrive for Business или SharePoint	259
Различия между OneDrive for Business и SharePoint Online	260
УПРАЖНЕНИЕ 8.5. От локальной папки к общедоступной	261
Вопросы безопасности	263
Удаление всех запросов в Excel с помощью инспектора документов	264
Резюме	265
Глава 9. Введение в язык формул M Power Query	267
Изучаем возможности языка M	268
Изучая "продвинутые" возможности языка M	268
Этап 1. Только интерфейс пользователя	268
Этап 2. Простое редактирование в строке формул	269

Этап 3. Формулы M в настраиваемых столбцах	270
Этап 4. Настраиваемые функции в окне расширенного редактора	270
Этап 5. Расширенные вычисления (<i>List.Accumulate</i> , <i>List.Generate</i>)	271
Этап 6. Кодирование в произвольной форме в окне расширенного редактора	271
Ресурсы в Интернете	272
Локальные ресурсы	272
УПРАЖНЕНИЕ 9.1. Применение переменной <i>#shared</i> для изучения встроенных функций	274
Компоненты языка M	275
УПРАЖНЕНИЕ 9.2. Hello World	275
Выражение <i>let</i>	277
Объединение выражений из нескольких запросов и областей доступа	280
Типы, операторы и встроенные функции в языке M	283
Основные типы данных языка M	284
Числовой тип данных	285
Функции времени	286
Тип даты	287
Тип длительности	287
Текстовый тип данных	288
Тип данных <i>Null</i>	289
Логический тип данных	290
Сложные типы данных	291
Тип списка	291
Операторы списка	292
Функции <i>List.X</i>	293
Записи	295
Операторы записи	296
Функции <i>Record.X</i>	298
Табличный тип	298
Условия и выражения <i>if</i>	300
Выражение <i>if-then-else</i>	301
Выражение <i>if</i> внутри выражения <i>let</i>	302
Пользовательские функции	303
Вызов функции	305
Выражение <i>each</i>	306
Дополнительные темы	307
Обработка ошибок	307
Отложенные и немедленные вычисления	308
Циклы	309
Рекурсия	309
Функция <i>List.Generate</i>	310
Функция <i>List.Accumulate</i>	311
Резюме	313
Глава 10. От ловушек к робастным запросам	315
Причины и следствия ловушек	316
Осведомленность	318
Лучшие практические методики	319
Модификации кода M	319

Ловушка 1. Игнорирование строки формул	319
УПРАЖНЕНИЕ 10.1. Применение строки формул для обнаружения статических ссылок на имена столбцов	320
Ловушка 2. Измененные типы	323
Ловушка 3. Небезопасная фильтрация	326
УПРАЖНЕНИЕ 10.2. Часть 1. Фильтрация товаров, окрашенных в черный цвет	327
Логика, определяющая условия фильтрации	329
УПРАЖНЕНИЕ 10.2. Часть 2. Поиск значений на панели <i>Фильтр</i>	331
Ловушка 4. Переупорядочение столбцов	332
УПРАЖНЕНИЕ 10.3. Часть 1. Переупорядочение подмножества столбцов	333
УПРАЖНЕНИЕ 10.3. Часть 2. Пользовательская функция <i>FnReorderSubsetOfColumns</i>	335
Ловушка 5. Удаление и выделение столбцов	336
УПРАЖНЕНИЕ 10.4. Обработка случайных столбцов в таблице <i>Wide World Importers</i>	337
Игнорирование пропущенного столбца	337
Выделение или удаление столбцов на основе их расположения	338
Выделение или удаление столбцов на основе их имен	339
Ловушка 6. Переименование столбцов	339
УПРАЖНЕНИЕ 10.5. Переименование столбцов <i>Random</i> в таблице <i>Wide World Importers</i>	340
Настраиваемая функция <i>FnRenameColumnsByIndices</i>	340
Функция <i>Table.TransformColumnNames</i>	342
Ловушка 7. Разбиение столбца на другие столбцы	343
УПРАЖНЕНИЕ 10.6. Выполнение некорректного разбиения	344
Ловушка 8. Слияние столбцов	347
Дополнительные ловушки и методы для создания надежных запросов	349
Резюме	349
Глава 11. Основы текстовой аналитики	353
Поиск ключевых слов в текстовых столбцах	354
УПРАЖНЕНИЕ 11.1. Основная методика, применяемая для поиска ключевых слов	354
Применение декартова произведения для определения ключевых слов	359
УПРАЖНЕНИЕ 11.2. Реализация декартова произведения	360
Часть 1. Начальная подготовка	361
Часть 2. Декартово произведение	362
Часть 3. Создание соответствий	364
Часть 4. Повышение производительности	366
УПРАЖНЕНИЕ 11.3. Определение ключевых слов с помощью настраиваемой функции	368
Выбор метода: статический поиск, декартово произведение или настраиваемая функция	371
Разбиение слов	372
УПРАЖНЕНИЕ 11.4. Простое разбиение слов	372
Часть 1. Разбиение слов с помощью пробелов	372
Часть 2. Отделение слов от знаков препинания	374
Часть 3. Разделение слов пробелами и удаление пунктуации	376
УПРАЖНЕНИЕ 11.5. Фильтрация по стоп-словам	377
УПРАЖНЕНИЕ 11.6. Поиск по ключевым словам с помощью разделенных слов	380
Часть 1. Использование внутренних запросов на слияние для сопоставления столбцов	380
Часть 2. Обнаружение составных ключевых слов	382

УПРАЖНЕНИЕ 11.7. Создание "облаков" слов в Power BI	390
Резюме	392
Глава 12. Расширенная текстовая аналитика: извлечение смысла	393
Службы Microsoft Azure Cognitive Services	393
Развертывание ключей и ресурсов API в Azure	396
Развертывание API Translator Text	396
Развертывание API Text Analytics	398
Плюсы и минусы доступа к Cognitive Services, получаемого через Power Query	399
Чрезмерные вызовы API	399
Конфиденциальность и безопасность	401
Неподдерживаемое обновление по расписанию	401
Перевод текста	402
Справочник по Translator Text API	402
УПРАЖНЕНИЕ 12.1. Простой перевод	403
Часть 1. Формирование контента JSON для тела запроса	403
Часть 2. Создание параметра <i>API Key</i>	404
Часть 3. Создание веб-запроса	405
Совместное использование отчета без ключа API	408
УПРАЖНЕНИЕ 12.2. Перевод нескольких сообщений	408
Совместное использование отчета без ключа API: удобный способ	412
Анализ настроений	414
Вызов API Sentiment Analysis	414
Подготовка данных	415
УПРАЖНЕНИЕ 12.3. Реализация настраиваемой функции анализа настроений	
<i>FnGetSentiment</i>	417
Настраиваемая функция <i>FnGetSentiment</i>	417
Часть 1. Загрузка данных	418
Часть 2. Подготовка сообщений	419
Часть 3. Конвертирование таблицы в формат JSON	420
Часть 4. Установка API Key	421
Часть 5. Отправка веб-запроса	421
Часть 6. Обработка ответа	423
Часть 7. Создание настраиваемой функции	423
Часть 8. Вызов функции <i>FnGetSentiment</i>	426
Как избежать ошибок	428
УПРАЖНЕНИЕ 12.4. Выполнение анализа настроений для больших наборов данных	429
Извлечение ключевых фраз	431
УПРАЖНЕНИЕ 12.5. Преобразование логики настроений в ключевые фразы	432
Поддержка нескольких языков	435
Замена языкового кода	435
Динамическое обнаружение языков	435
УПРАЖНЕНИЕ 12.6. Преобразование логики настроений в обнаружение языка	436
Резюме	437
Глава 13. Аналитика социальных сетей	439
Начало работы с соединителем Facebook	440
УПРАЖНЕНИЕ 13.1. Поиск страниц, которые вам понравились	440

Анализ предпочтений ваших друзей.....	446
УПРАЖНЕНИЕ 13.2. В поисках друзей Power BI и их друзей.....	446
УПРАЖНЕНИЕ 13.3. Найдите страницы, которые понравились вашим друзьям.....	449
Анализ страниц Facebook.....	452
УПРАЖНЕНИЕ 13.4. Извлечение сообщений и комментариев из страниц Facebook, основная возможность.....	452
Сокращение обзора: фильтрация результатов по времени.....	456
УПРАЖНЕНИЕ 13.5. Анализ вовлеченности пользователей путем подсчета комментариев и общих ресурсов.....	457
УПРАЖНЕНИЕ 13.6. Сравнение нескольких страниц.....	460
Резюме.....	464
Глава 14. Заключительный проект: объединяем все вместе.....	465
УПРАЖНЕНИЕ 14.1. Спасите данные компании Wide World Importers.....	465
Подсказки.....	467
Часть 1. Приступим к решению.....	467
Часть 2. Вызов функции отмены сведения столбца.....	469
Часть 3. Последовательность свертывания столбцов для доходов 2018 года.....	471
Часть 4. Объединение доходов за 2018 и 2015–2017 годы.....	472
УПРАЖНЕНИЕ 14.2. Сравнение таблиц и отслеживание действий хакеров.....	473
Подсказка.....	473
УПРАЖНЕНИЕ 14.2. Решение.....	474
Обнаружение следов деятельности хакеров во "взломанной" таблице.....	475
Резюме.....	476
Предметный указатель.....	478

* * *

Сведения об авторах рисунков

Перечисленные далее рисунки предоставлены с любезного разрешения компании Microsoft:

- | | | |
|------------|-------------|----------------------------|
| ◆ 1.1–1.9 | ◆ 6.1–6.9 | ◆ 11.1–11.14, 11.16, 11.17 |
| ◆ 2.1–2.16 | ◆ 7.1–7.5 | ◆ 12.3–12.8, 12.10–12.14 |
| ◆ 3.1–3.8 | ◆ 8.1–8.10 | ◆ 13.1–13.11 |
| ◆ 4.1–4.11 | ◆ 9.2–9.10 | ◆ 14.1–14.3 |
| ◆ 5.1–5.11 | ◆ 10.2–10.9 | |