

УДК 004.424
ББК 32.372
П88

Кен Пульс, Мигель Эскобар

П88 Приручи данные с помощью Power Query в Excel и Power BI / пер. с англ.
А. Ю. Гинько. – М.: ДМК Пресс, 2022. – 572 с.: ил.

ISBN 978-5-93700-105-4

Иногда нас называют мартышками, работающими с данными, но на самом деле мы чаще походим на волшебников. Наши данные редко появляются на свет в готовом к работе виде, и у нас могут уходить долгие часы на их очистку, фильтрацию и преобразование. Power Query помогает сократить этот процесс при первичной обработке данных, а все последующие обновления позволяет свести к простому нажатию на кнопку. Когда дело касается импорта, очистки и преобразования исходных данных для дальнейшего анализа, освоить Power Query бывает гораздо легче, чем выучить формулы Excel или язык программирования VBA.

Нет сомнений, что Power Query навсегда изменит подход специалистов Excel к работе с данными. Если у вас есть Excel, значит, у вас уже есть Power Query – этот инструмент встроен в Excel 2016 и выше. Эта книга поможет вам извлечь максимум пользы из Power Query.

УДК 004.424
ББК 32.372

Master Your Data with Power Query in Excel and Power BI, published by Holy Macro! Books.
Copyright © 2021, Russian-language edition copyright.

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

ISBN 978-1-61547-058-7 (англ.)
ISBN 978-5-93700-105-4 (рус.)

© Tickling Keys, Inc., 2021
© Перевод, оформление, издание,
ДМК Пресс, 2022

Оглавление

Предисловие от издательства	15
Предисловие	17
Как Power Query изменил НАШИ жизни	17
История Кена: «Кофе и Power Query»	17
История Мигеля: новый старт	18
Благодарности от авторов	19
Благодарности от Кена	20
Благодарности от Мигеля	21
Наши преданные читатели	22
И наконец	22
Глава 0. Революция данных	23
Общий сценарий для аналитиков данных	23
Преимущества и опасности черной магии	24
Будущее изменилось	26
Почему Power Query – это магия?	28
Извлечение	29
Преобразование	29
Загрузка	30
Возможности Power Query и интеграция с другими продуктами	31
Компоненты Power Query	32
Цикл обновлений Power Query	34
Power Query Online	34
Microsoft 365	35
Excel 2016/2019/2021	35
Excel 2010 & 2013	35
Power BI Desktop	35
Как использовать эту книгу	36
Где найти Power Query?	36
Excel 365	36
Power BI Desktop	37
Предыдущие версии Excel	37
Подключение к данным	37
Особые пометки	38
Сопроводительные файлы	38
Глава 1. Основы Power Query	39
Перед началом	39
Изменение настроек Power Query по умолчанию в Excel	40
Изменение настроек Power Query по умолчанию в Power BI	40
Извлечение	41

Настройки подключения (выбор данных).....	41
Аутентификация.....	42
Предварительный просмотр	42
Выбор назначения запроса.....	43
Преобразование.....	44
Редактор Power Query	45
Преобразования по умолчанию.....	46
Источник (Source).....	46
Повышенные заголовки (Promoted Headers).....	47
Измененный тип (Changed Type).....	47
Создание и изменение преобразований	48
Загрузка.....	50
Установка типов данных	50
Переименование запроса	52
Загрузка запроса в Excel.....	52
Загрузка запроса в Power BI.....	53
Обновление запросов	54
Редактирование запросов	55
Запуск редактора Power Query в Power BI.....	56
Запуск редактора Power Query в Excel.....	56
Просмотр шагов.....	57
Настройка шагов	57
Влияние Power Query	60

Глава 2. Управление запросами61

Использование архитектуры со множеством запросов.....	61
Разделение запросов на E, T и L.....	61
Преимущества совмещения запросов	62
Преимущества разделения запросов	63
Влияние разделения запросов на производительность.....	63
Ссылки на запросы.....	65
Создание базового запроса.....	65
Ссылочные запросы.....	66
Визуализация дерева зависимостей запросов	69
Просмотр зависимостей при помощи Monkey Tools	70
Выбор места загрузки запроса	71
Выбор места загрузки запроса в Power BI	72
Выбор места загрузки запроса в Excel	72
Изменение места назначения.....	76
Организация запросов.....	78
Создание папок в Power Query.....	78
Перенос запросов в группы	79
Изменение порядка следования запросов и групп.....	80
Создание подпапок запросов	81
Разделение существующих запросов	81
Заключительные мысли об архитектуре запросов.....	82

Глава 3. Типы данных и ошибки.....	85
Типы и форматы данных.....	85
Форматы	85
Типы данных.....	86
Как устанавливать формат данных в Power Query?	90
Порядок шагов имеет значение	90
Важность определения типов данных	92
Распространенные ошибки в Power Query	93
Ошибки на уровне шага	94
Ошибки источников данных	95
Ошибки вида «столбец X не найден».....	97
Ошибки значений.....	99
Обнаружение ошибок.....	99
Ошибки из-за неправильного приведения типов	101
Ошибки по причине несовместимости типов данных.....	103
Проверка запросов на ошибки	105
Обнаружение источника ошибок	105
Исправление исходного запроса	106
Удаление запроса с ошибками	108
Заключительные мысли о типах данных и ошибках	108
 Глава 4. Перенос запросов между Excel и Power BI.....	 109
Перенос запросов между решениями	109
Перенос запросов Excel в новую рабочую книгу	110
Перенос запросов из Excel в Power BI.....	113
Перенос запросов из Power BI в Excel.....	114
Перенос запросов из Power BI в новый проект Power BI	115
Импорт запросов из Excel в Power BI.....	115
Только внешние источники данных.....	116
Импорт модели данных Excel в Power BI	118
Импорт данных на основе таблиц Excel – копирование	119
Таблицы Excel – сохранение подключения	126
Заключительные мысли о переносе запросов между решениями	128
 Глава 5. Импортирование из плоских файлов.....	 131
Понимание процесса импорта данных.....	131
Определение системных настроек	132
Как программа интерпретирует плоские данные.....	133
Импортирование файлов с разделителями.....	135
Источник данных	136
Извлечение данных	136
Задача.....	137
Использование локали для установки корректных типов данных	138
Импортирование файлов без разделителей	141
Подключение к файлу.....	142
Очистка файлов без разделителей	143

Разделение столбцов по позиции.....	145
Прелесть ошибок в Power Query.....	146
Удаление лишних столбцов	148
Объединение столбцов	149
Разделение столбцов по разделителю	150
Исключение дублирующихся пробелов	151
Минута славы Power Query	152

Глава 6. Импортрование из файлов Excel..... 155

Данные в активной рабочей книге	155
Подключение к таблицам Excel.....	156
Подключение к табличным диапазонам	158
Подключение к именованным диапазонам.....	161
Динамические именованные диапазоны.....	163
Подключение к рабочим листам Excel из той же книги	165
Данные из других рабочих книг.....	165
Подключение к файлу Excel.....	166
Подключение к таблицам	168
Подключение к именованным диапазонам.....	169
Подключение к рабочим листам.....	170
Заключительные мысли о подключении к данным Excel	174

Глава 7. Простые техники преобразования данных..... 177

Снимаем проклятие сводных данных.....	177
Подготовка данных	178
Отмена свертывания других столбцов	179
Повторное сведение данных при помощи сводной таблицы	181
Есть ли жизнь после обновления данных?	182
Разница между различными типами отмены свертывания	183
Сведение столбца	184
Разделение столбцов.....	186
Разделение столбца на несколько столбцов.....	187
Разделение столбца на строки	188
Разделение на столбцы с отменой свертывания против разделения на строки	190
Фильтрация и сортировка	191
Фильтрация значений	192
Применение контекстных фильтров.....	195
Сортировка данных.....	197
Группирование данных	198

Глава 8. Добавление данных..... 203

Базовые операции по добавлению данных.....	203
Добавление двух таблиц.....	205
Добавление дополнительных таблиц.....	208
Объединение запросов с разными заголовками	211

Добавление таблиц и диапазонов в текущем файле	213
Консолидация таблиц	214
Консолидация диапазонов и рабочих листов	218
Используйте =Excel.CurrentWorkbook() с осторожностью	221
Заключительные мысли о добавлении запросов.....	221
Глава 9. Объединение файлов	223
Практический пример	223
Описание процесса	225
Методология объединения файлов	225
Архитектура запросов при объединении файлов	225
Шаг 0: подключение к папке.....	227
Подключение к локальной/сетевой папке	228
Подключение к папке SharePoint.....	229
Подключение к OneDrive для бизнеса.....	231
Подключение к другим файловым системам	231
Шаг 1: фильтрация и страховка на будущее	232
Методология шага 1	232
Применение шага 1 к нашему примеру.....	233
Шаг 2: объединение файлов.....	235
Методология шага 2	236
Применение шага 2 к нашему примеру.....	236
Шаг 3: преобразование данных в запросе примера.....	239
Почему нужно использовать запрос «Преобразовать пример файла»?	239
Использование запроса «Преобразовать пример файла».....	240
Шаг 4: преобразование данных в мастер-запросе	243
Исправление ошибки на уровне шага в мастер-запросе	243
Сохранение свойств файлов	244
Добавление дополнительных шагов	246
Обновление	248
Использование данных	248
Добавление новых файлов	249
Повышение эффективности с помощью сохранения верхних строк.....	250
Глава 10. Объединение данных	253
Основы объединения данных.....	253
Создание подготовительных запросов	254
Выполнение объединения запросов	254
Типы соединений	257
Внешнее соединение слева.....	260
Внешнее соединение справа	262
Полное внешнее соединение.....	264
Внутреннее соединение	265
Антисоединение слева	265
Антисоединение справа.....	266

Полное антисоединение.....	267
Декартовы произведения.....	268
Методология	268
Пример.....	268
Случайные декартовы произведения	271
Объединения с приблизительными совпадениями.....	272
Методология	273
Пример.....	273
Поиск нечетких соответствий	277
Основы нечеткого поиска	278
Таблицы преобразования	279
Управление порогом подобия.....	281
Стратегии поддержки решений с нечетким поиском	283

Глава 11. Источники данных в интернете 285

Подключение к файлам данных в интернете	285
Подключение к веб-страницам	287
Подключение к данным на веб-странице	287
Естественные и предлагаемые таблицы.....	288
Добавление таблиц с использованием примеров.....	289
Подключение к страницам без таблиц.....	291
Предостережения при работе с данными из интернета	296
Сбор данных.....	296
Целостность данных	296
Надежность решения.....	297

Глава 12. Реляционные источники данных 298

Подключение к базам данных	298
Соединение с БД.....	298
Управление учетными данными	301
Не можете подключиться к нашей базе данных?.....	302
Использование навигатора.....	303
Исследование данных.....	304
Свертывание запросов.....	307
Что такое свертывание запросов?	308
Какие технологии поддерживает механизм свертывания запросов? ..	310
Распространенные мифы относительно свертывания запросов.....	311
Уровни конфиденциальности	313
Объявление уровней конфиденциальности данных	315
Управление уровнями конфиденциальности данных	315
Конфиденциальность и производительность.....	316
Отключение движка конфиденциальности	317
Оптимизация	320

Глава 13. Преобразование табличных данных..... 323

Сложные шаблоны сведения данных	323
---------------------------------------	-----

Сведение сгруппированных данных.....	323
Сведение вертикально сгруппированных данных	330
Сведение горизонтально сгруппированных данных	332
Сложные шаблоны отмены свертывания данных.....	337
Отмена свертывания данных с подкатегориями	337
Эффективная отмена свертывания данных с подкатегориями.....	345
Изменение запроса отмены свертывания данных с подкатегориями	346
Сохранение значений null при отмене свертывания данных.....	349
Продвинутое техники группирования данных.....	352
Процент от целого.....	352
Ранжирование данных	355
Нумерация сгруппированных строк (номера строк по секциям).....	359
Глава 14. Условная логика в Power Query	364
Основа условной логики.....	364
Описание набора данных.....	364
Подключение к данным	365
Создание условной логики при помощи интерфейса пользователя.....	366
Условная проверка в ручном режиме	369
Воспроизведение функции Excel ЕСЛИОШИБКА (IFERROR)	372
Работа с несколькими условиями	376
Сравнение со следующей/предыдущей строкой	379
Столбцы из примеров	383
Глава 15. Значения в Power Query	389
Типы значений в Power Query	389
Таблицы	391
Списки.....	392
Синтаксис	392
Создание списков	392
Преобразование списка в таблицу	394
Создание списка из столбца таблицы	395
Создание списка списков.....	397
Записи	400
Синтаксис	400
Создание записи.....	401
Преобразование записи в таблицу	402
Создание нескольких записей.....	402
Преобразование нескольких записей в таблицу	403
Доступ к записям таблиц по позиции (индексирование строк)	405
Доступ к записям таблиц по критерию.....	407
Создание записей из каждой строки таблицы.....	410
Значения	412
Двоичные данные.....	413
Ошибки	413

Ошибки на уровне строки.....	413
Ошибки на уровне шага	414
Функции.....	415
Ключевые слова в Power Query	418
#binary.....	419
#date, #datetime и #datetimezone	420
#time	421
#duration	422
type	423
#table	426

Глава 16. Изучаем язык M..... 429

Структура запроса на языке M	429
Структура запроса	430
Область определения запроса и идентификаторы	432
Обобщенные идентификаторы.....	434
Комментарии к коду	435
Собираем все воедино	437
Понимание процесса выполнения запроса.....	438
Что такое ленивое вычисление?	439
План выполнения запроса	440
Итераторы (построчное выполнение).....	443
Ремарка по поводу рекурсивных функций в Power Query	443
Ключевые слова each и _.....	444
Другие техники	449
Получение первого значения из столбца таблицы	449
Замена на null при ошибке навигации	451
Создание динамического списка заголовков типизированных столбцов	452
Создание динамического списка заголовков нетипизированных столбцов	456

Глава 17. Параметры и пользовательские функции 461

Воссоздание метода объединения файлов.....	461
Создание примера файла (Sample File).....	462
Создание параметра Sample File Parameter	463
Создание преобразования файла (Transform Sample)	465
Создание функции Transform Function.....	466
Вызов функции Transform Function.....	467
Обновление функции Transform Function.....	467
Ключевые выводы	468
Создание настраиваемых функций с помощью параметров	469
Создание параметра FilePath	470
Создание запроса Timesheet Transform	471
Создание функции Timesheet Function.....	473
Обновление запроса Timesheet.....	473

Создание настраиваемых функций вручную	477
Построение сценария разового применения	477
Преобразование запроса в функцию	478
Вызов функции	481
Отладка настраиваемых функций	482
Восстановление функциональности	484
Таблицы динамических параметров	485
Проблема с динамическими путями к файлам	485
Реализация таблицы динамических параметров	487
Создание таблицы параметров	487
Реализация функции fnGetParameter	489
Вызов функции	490
Применение таблиц параметров	492
Глава 18. Техники работы с датой и временем	494
Определение границ календаря	494
Динамическое создание границ календаря	495
Корректировка начальной и конечной дат для нестандартных финансовых периодов	497
Корректировка начальной и конечной дат для 364-дневного календаря	499
Календари с последовательными датами	501
Создание календаря	501
Обогащение календаря за счет дополнительных столбцов	503
Столбцы для финансовых периодов в 12-месячном календаре	503
Столбцы-идентификаторы периодов для 364-дневного календаря ...	504
Столбцы финансовых периодов для календарей 4-4-5 (и их разновидностей)	506
Что находится в файле с примерами?	509
Заполнение особых диапазонов даты и времени	510
Заполнение определенного количества дат	510
Заполнение определенного количества часов по каждой дате	512
Заполнение определенного количества дат с заданными интервалами	513
Разнесение данных на основе таблиц с датами	515
Разнесение данных по дням	515
Разнесение данных по целым месяцам	518
Разнесение данных по заданному количеству месяцев от начальной даты	522
Заключительные штрихи к разнесению данных	525
Глава 19. Оптимизация запросов	527
Оптимизация настроек Power Query	527
Глобальные параметры загрузки данных	527
Глобальные параметры редактора Power Query	528
Глобальные параметры безопасности	528

Глобальные параметры конфиденциальности.....	529
Настройки текущей книги (файла) – фоновые данные.....	529
Настройки текущей книги (файла) – другие	531
Использование функций буферизации.....	531
Форсирование вычисления значения	532
Буферизация вычисления значения	534
Снижение временных лагов во время разработки.....	537
Стратегия уменьшения временных лагов	538
Пример борьбы с временными лагами при разработке	539
Адаптация решения для снижения временных лагов.....	541
Изменение данных в предпросмотре	544
Ошибка Formula Firewall	544
Ошибка Formula.Firewall №1: несовместимость уровней конфиденциальности	545
Ошибка Formula.Firewall № 2: доступ к источнику данных	545
Вызов ошибки перестроения сочетания данных	546
Перестроение сочетания данных против создания цепочек запросов.....	548
Перестроение сочетания данных против выравнивания запросов	551
Перестроение сочетания данных при передаче значений в SQL	553
Заключительные мысли об ошибках Formula.Firewall	555

Глава 20. Автоматизация обновлений..... 557

Варианты автоматического обновления в Excel.....	557
Обновления в Excel без VBA.....	557
Фоновое обновление	558
Обновление каждые X минут	558
Обновление при открытии рабочей книги	559
Быстрая загрузка данных.....	559
Автоматизация обновлений запросов в Excel с помощью VBA	560
Обновление одного подключения.....	560
Обновление в определенном порядке	563
Обновление всех запросов.....	565
Проблемы с синхронным обновлением	565
Расписание обновлений в Power BI	566

Предметный указатель..... 568