

УДК 004.424
ББК 32.372
М314

М314 Даниил Маслюк

Анализируем данные в Microsoft Power BI. Подготовка к экзамену DA-100/
пер. с англ. А. Ю. Гинько – М.: ДМК Пресс, 2022. – 332 с.: ил.

ISBN 978-5-97060-849-4

В книге рассматриваются все основные темы для подготовки к сертификационному экзамену Microsoft DA-100, посвященному анализу данных при помощи Microsoft Power BI. Материал удобно структурирован: каждая глава охватывает одну тему, в каждом разделе рассматривается конкретный навык. Освещаются вопросы получения данных, их подготовки, преобразования, моделирования и визуализации средствами Power BI Desktop, создания отчетов и дашбордов, углубленного анализа данных и др.

Издание пригодится профессионалам в области бизнес-аналитики, аналитикам данных и специалистам по проектированию и созданию отчетов, желающим проверить или подтвердить свои навыки.

УДК 004.424
ББК 32.372

Original English language edition published with the authorization of Microsoft Corporation by: Pearson Education, Inc. Copyright © 2021 by Pearson Education, Inc. All rights reserved. Russian-language edition copyright © 2021 by DMK Press. All rights reserved.

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.

Материал, изложенный в данной книге, многократно проверен. Но, поскольку вероятность технических ошибок все равно существует, издательство не может гарантировать абсолютную точность и правильность приводимых сведений. В связи с этим издательство не несет ответственности за возможные ошибки, связанные с использованием книги.

ISBN 978-0-13-681968-4 (англ.)
ISBN 978-5-97060-849-4 (рус.)

© Pearson Education, Inc., 2021
© Оформление, перевод на русский язык,
издание, ДМК Пресс, 2022

Оглавление

Предисловие от издательства	13
Благодарности	14
Об авторе	15
Введение	16
Структура книги	16
Подготовка к экзамену	17
Сертификация Microsoft.....	17
Сопроводительные файлы.....	17
Список ссылок на источники в сети	18
Ошибки, обновления и поддержка.....	18
Оставайтесь на связи.....	18
Глава 1. Подготовка данных.....	19
Навык 1.1. Получение данных из различных источников.....	19
Распознавание источника данных и подключение к нему	20
Редактор Power Query	23
Зависимости запросов.....	24
Изменение настроек источника данных.....	25
Выбор между использованием общего набора данных и созданием локального	26
Выбор режима хранения	28
Режим импорта данных.....	28
Режим DirectQuery.....	29
Режим Live Connection.....	30
Составные модели	30
Выбор подходящего типа запроса.....	32
Последствия использования режима DirectQuery.....	32
Поиск проблем с производительностью запросов	35
Просмотр машинных запросов	35
Диагностика запросов.....	37
Дополнительное (инкрементное) обновление	39
Использование облачного хранилища Microsoft Dataverse	40
Применение параметров.....	40
Создание параметров	41
Использование параметров	42
Редактирование параметров	44
Создание функций.....	45
Использование и создание файла PBIDS	46

Использование и создание потока данных	47
Подключение к набору данных с использованием конечных точек XMLA.....	48
Навык 1.2. Профилирование данных.....	50
Распознавание аномалий в данных.....	50
Оценка структуры данных и извлечение свойств столбцов	52
Получение статистических сведений о данных	53
Навык 1.3. Очистка, преобразование и загрузка данных.....	55
Поиск проблем с качеством информации и противоречивостью данных и их устранение путем замены значений.....	56
Замена значений.....	56
Удаление строк.....	58
Поиск первопричины ошибки	58
Определение и преобразование типов данных столбцов.....	59
Использование локали	62
Определение и создание подходящих ключей для объединения	63
Ключи для объединения таблиц в Power Query	63
Ключи для создания связей.....	64
Применение преобразований к структуре таблиц	65
Работа с шагами запроса	67
Уменьшение количества строк и столбцов.....	68
Сведение, отмена свертывания и транспонирование	71
Добавление столбцов.....	73
Комбинирование запросов.....	78
Добавление.....	78
Объединение.....	80
Соглашение об именовании столбцов и запросов	83
Использование расширенного редактора для написания кода на языке M в Power Query	84
Строка формул.....	86
Настройка загрузки данных	88
Решение проблем с импортом данных.....	89
Заключение	91
Мысленный эксперимент.....	93
Ответы.....	95
Глава 2. Моделирование данных	96
Навык 2.1. Проектирование модели данных.....	96
Определение таблиц.....	97
Плоская схема	97
Схема «звезда»	98
Схема «снежинка»	99
Настройка свойств таблиц и столбцов.....	101
Свойства таблиц.....	101
Свойства столбцов.....	102

Определение быстрых мер	103
Выравнивание иерархий типа родитель – потомок	107
Определение ролевых измерений	110
Определение кратности связей и настройка двунаправленной перекрестной фильтрации	114
Кратность связей	115
Направление кросс-фильтрации	116
Оптимизация модели данных	119
Режим хранения данных	119
Организация связей	120
Создание агрегаций	120
Настройка кратности связей	120
Установка связей типа «многие ко многим»	121
Прямые связи типа «многие ко многим»	122
Связи типа «многие ко многим» с использованием таблиц-мостов	123
Создание общей таблицы дат	125
Создание календаря в Power Query	126
Создание календаря при помощи DAX	127
Определение приемлемого уровня гранулярности данных	128
Навык 2.2. Разработка модели данных	131
Установка направления кросс-фильтрации и применение фильтра безопасности	131
Создание вычисляемых таблиц	132
Клонирование таблиц	132
Создание таблицы на основе данных из разных источников	133
Предварительный расчет мер для повышения эффективности отчетов	134
Создание иерархий	134
Создание вычисляемых столбцов	135
Применение ролей безопасности на уровне строк	138
Создание ролей в Power BI Desktop	138
Выбор роли в Power BI Desktop	141
Динамическая безопасность на уровне строк	142
Настройка инструмента Вопросы и ответы (Q&A)	145
Обучение визуального элемента Вопросы и ответы	147
Синонимы	149
Навык 2.3. Создание мер с использованием языка DAX	150
Использование языка DAX для написания сложных мер	151
Использование функции CALCULATE для манипулирования фильтрами	155
Добавление фильтров	155
Удаление фильтров	156
Изменение фильтров	157
Преобразование контекста	158
Логика операций со временем с применением DAX	161
Замена числовых столбцов на меры	164

Использование базовых статистических функций для обогащения данных	165
Создание полуаддитивных мер	166
Навык 2.4. Оптимизация производительности модели данных.....	169
Удаление ненужных строк и столбцов	169
Удаление ненужных строк.....	169
Удаление ненужных столбцов.....	170
Определение мер, визуальных элементов и связей, негативно влияющих на производительность.....	170
Улучшение ситуации с кратностью столбцов путем изменения типов данных.....	172
Улучшение ситуации с кратностью столбцов при помощи суммирования данных	173
Создание и управление агрегациями.....	174
Заключение	176
Мысленный эксперимент.....	178
Ответы.....	181
Глава 3. Визуализация данных	184
Навык 3.1. Создание отчетов.....	184
Добавление элементов визуализации в отчет.....	185
Выбор подходящего типа элемента визуализации	186
Линейчатая диаграмма.....	187
График и диаграмма с областями.....	188
Комбинированная диаграмма	189
Ленточная диаграмма	190
Каскадная диаграмма	191
Воронка	192
Точечная диаграмма	193
Круговая и кольцевая диаграммы	194
Диаграмма дерева.....	195
Карта	196
Датчик	197
Карточка, многострочная карточка и KPI	198
Визуальные элементы с искусственным интеллектом.....	199
Форматирование и настройка визуализаций	200
Импортирование пользовательских визуализаций.....	200
AppSource	201
Из файла	202
Настройка условного форматирования.....	202
Применение срезов и фильтров.....	204
Срезы.....	204
Панель фильтров	206
Добавление визуальных элементов R и Python.....	208
Настройка страницы отчета.....	211
Настройка специальных возможностей в отчетах.....	212
Имена страниц, заголовки и метки	213

Маркеры.....	214
Альтернативный текст.....	214
Порядок табуляции.....	214
Тема отчета и выбор цвета.....	215
Настройка автоматического обновления страниц.....	216
Фиксированный интервал.....	217
Обнаружение изменений.....	217
Создание отчетов с разбивкой на страницы.....	219
Навык 3.2. Создание дашбордов.....	221
Управление плитками на дашборде.....	221
Настройка мобильного отображения.....	223
Конфигурирование оповещений о данных.....	225
Использование визуального элемента Вопросы и ответы	226
Добавление темы дашборда.....	227
Закрепление динамической страницы отчета на дашборде.....	228
Навык 3.3. Повышение удобства использования отчетов.....	229
Настройка закладок.....	230
Создание пользовательских подсказок.....	233
Настройка и конфигурирование взаимодействий между визуальными элементами.....	236
Настройка навигации по отчету.....	237
Применение сортировки.....	238
Синхронизация срезов.....	240
Использование панели Выбор	242
Применение детализации и кросс-филтрации.....	243
Углубление в данные с использованием интерактивных визуальных элементов.....	245
Экспорт данных отчета.....	247
Разработка отчетов для мобильных устройств.....	247
Заключение.....	249
Мысленный эксперимент.....	252
Ответы.....	253
Глава 4. Анализ данных.....	255
Навык 4.1. Повышение эффективности отчетов за счет выделения важных аналитических выводов.....	255
Применение условного форматирования.....	256
Таблицы и матрицы.....	256
Другие визуальные элементы.....	260
Анализ ведущих N элементов.....	261
Фильтр уровня визуализации.....	261
Вопросы и ответы.....	261
DAX.....	262

Исследование статистической сводки	263
Добавление результатов краткой аналитики (Quick Insights) на дашборд	265
Создание опорных линий с помощью панели Аналитика (Analytics).....	266
Использование оси воспроизведения (Play Axis) и проведение анализа на основе временных рядов	268
Персонализация визуальных элементов.....	270
Навык 4.2. Выполнение углубленного анализа.....	271
Определение выбросов в данных.....	272
Группирование данных и разделение на ячейки	274
Группирование данных.....	274
Разделение на ячейки	276
Использование элемента Ключевые факторы влияния для анализа по измерениям	277
Применение дерева декомпозиции для анализа мер	279
Применение аналитики ИИ	280
Заключение	282
Мысленный эксперимент.....	284
Ответы.....	285
Глава 5. Развертывание и поддержка	287
Навык 5.1. Управление наборами данных.....	287
Настройка запланированного обновления набора данных	288
Доступ к локальным данным.....	288
Настройка запланированного обновления	290
Настройка членства в группах безопасности на уровне строк	291
Присвоение ролей в службе Power BI.....	291
Просмотр в качестве роли в службе Power BI	293
Обеспечение доступа к наборам данных.....	294
Доступ через рабочую область.....	294
Доступ через приложение.....	294
Управлением разрешениями наборов данных	295
Анализ влияния	297
Настройка инкрементного обновления наборов данных.....	297
Создание параметров RangeStart и RangeEnd.....	298
Фильтрация с использованием созданных параметров.....	299
Определение политики инкрементного обновления.....	301
Свертывание запросов	302
Рекомендация и сертификация содержимого в Power BI	303
Настройка формата хранения крупных наборов данных.....	305
Навык 5.2. Создание и управление рабочими областями.....	307
Создание и настройка рабочей области	308
Рекомендованная стратегия жизненного цикла разработки	310
Параметры	310
Конвейеры развертывания	311

12 ❖ Оглавление

Присвоение ролей рабочей области.....	313
Настройка и публикация приложения рабочей области.....	314
Установка.....	315
Переходы.....	316
Разрешения.....	317
Вид приложения	317
Обновление опубликованного приложения.....	318
Отмена публикации приложения	319
Публикация, импорт и обновление элементов рабочей области	319
Применение меток конфиденциальности к содержимому рабочей области.....	320
Настройка подписок	321
Подписка на содержимое.....	321
Управление подписками.....	323
Заключение	323
Мысленный эксперимент.....	325
Ответы.....	327
Предметный указатель	328