

УДК 004.4
ББК 32.372
Г49

Г49 Гинько А. Ю.

Анализ и визуализация данных в Yandex DataLens. Подробное руководство: от новичка до эксперта. – М.: ДМК Пресс, 2023. – 356 с.: ил.

ISBN 978-5-93700-171-9

Современный мир невозможно представить без данных, и от их визуального представления во многом зависит эффективность работы с ними. Средств анализа и визуализации данных великое множество, и в этой книге мы подробно поговорим об одном из них – Yandex DataLens.

Читая эту книгу, вы совершите увлекательный путь от самых азов до вершин мастерства в обращении с этим мощнейшим облачным инструментом. Мы вместе построим несколько красочных дашбордов, а помимо этого вы узнаете все о подключениях, датасетах, чартах, функциях, включая оконные, и LOD-выражениях.

Издание будет полезно тем, кто хочет освоить искусство визуализации данных с нуля, а также тем, кто уже имеет опыт работы с Yandex DataLens и желает расширить свои знания и навыки.

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 978-5-93700-171-9

© Гинько А. Ю., 2022

© Оформление, издание, ДМК Пресс, 2023

Оглавление

Предисловие от издательства	8
Об авторе.....	9
Предисловие	10
Введение	11
Цель написания книги	11
Для кого предназначена эта книга	11
Структура книги	12
Сопроводительные материалы и исходные файлы.....	13
Глава 1. Основы Yandex DataLens.....	14
История DataLens: от корпоративной системы к массовому селф-сервису ...	14
Причины перехода на Yandex DataLens	16
Облачное хранение данных	17
Как устроено облако.....	19
Модели предоставления облачных услуг.....	20
С места в карьер: строим первый дашборд	21
Загрузка компонентов из Marketplace	21
Интерфейс и терминология Yandex DataLens	23
Мой первый дашборд.....	25
Заключение	55
Глава 2. Подключение к данным	56
Организация хранения данных.....	56
Собственные сервера	56
Облако	56
Управляемые БД против виртуальной машины	57
Создание кластера базы данных MySQL.....	59
Подключение к базам данных	63
Подключение к базе данных на примере MySQL.....	66
Особенности подключения к другим БД.....	68
Подключение к CSV-файлам, Google Sheets и Excel	71
Подключение к CSV-файлам	71
Подключение к Google Sheets.....	72
Загрузка данных из Excel	73
Подключение к Яндекс Метрика и AppMetrica.....	73
Яндекс Метрика.....	74
AppMetrica	74
Права доступа к подключениям	75
Заклучение	77
Глава 3. Подготовка данных: датасеты	78
Введение в датасеты.....	78
Создание датасета	78

Объединение данных из нескольких таблиц.....	80
Поля данных.....	82
Создание поля данных.....	82
Удаление поля данных.....	87
Изменение поля данных.....	87
Дублирование поля данных.....	88
Обновление полей в датасете.....	88
Запись формул в вычисляемых полях.....	89
Типы данных.....	90
Типы данных в Yandex DataLens	90
Таблица соответствий типов данных.....	91
Типы агрегации данных.....	93
Параметры уровня датасета	94
Фильтрация датасета.....	100
Создание фильтра	100
Изменение и удаление фильтра.....	102
Управление доступом к датасету и строкам.....	102
Права доступа ко всему датасету	102
Права доступа к данным на уровне строк.....	104
Датасеты на основе SQL-запросов	106
Заключение	108
Глава 4. Элементы визуализации: чарты.....	109
Чарты на основе датасета	109
Создание чарта на основе датасета	110
Разновидности чартов	112
Линейная диаграмма	112
Линейная диаграмма с двумя осями	124
Накопительная диаграмма с областями.....	125
Нормированная диаграмма с областями	126
Столбчатая диаграмма.....	127
Нормированная столбчатая диаграмма	129
Линейчатая диаграмма.....	130
Нормированная линейчатая диаграмма	132
Точечная диаграмма	134
Круговая диаграмма.....	136
Кольцевая диаграмма	137
Индикатор	139
Условный индикатор.....	140
Древовидная диаграмма.....	140
Таблица	141
Сводная таблица.....	144
Карта	145
QL-чарты	148
Создание QL-чарта.....	149
Добавление параметров в QL-чарт	150

Настройки чартов	151
Общие настройки	152
Настройки измерений и показателей.....	153
Настройки секций	155
Настройки секций на карте	157
Параметры уровня чарта	159
Иерархии в чартах	161
Диагностика чартов.....	163
Публикация чартов.....	165
Встраивание чартов	166
Управление доступом к чарту.....	166
Заключение	167
Глава 5. Панели визуализации: дашборды	168
Составляющие дашборда	168
Наш первый осмысленный дашборд	169
Создание дашборда.....	169
Добавление чартов на дашборд	171
Создание вкладок.....	178
Добавление селекторов на дашборд	180
Связи и алиасы	189
Добавление чартов из других датасетов.....	195
Текстовые виджеты и язык разметки Markdown	201
Добавление виджета с заголовком.....	206
Настройки дашборда	206
Публикация дашборда.....	208
Управление доступом к дашборду	209
Примеры дашбордов	210
Заключение	211
Глава 6. Стандартные функции и примеры их использования.....	212
Агрегатные функции.....	213
Прямая агрегация.....	214
Условная агрегация	216
Статистическая агрегация	220
Приблизительная агрегация.....	225
Строковая агрегация	225
Логические функции	226
CASE.....	226
IF.....	227
IFNULL.....	228
ISNULL.....	229
ZN.....	229
Операторы.....	229
Математические операторы.....	229
Логические операторы.....	231

Математические функции	233
Функции округления.....	234
Функции сравнения	234
Арифметические функции	235
Тригонометрические функции	236
Строковые функции	236
Функции поиска в строке	236
Функции строковых преобразований.....	239
Функции для работы с регулярными выражениями	243
Функции преобразования типов	246
Функции даты и времени.....	249
Функции работы с датами	249
Функции возвращения части даты	251
Функции для работы с временными рядами	251
Функции для работы с массивами.....	254
Функции разметки	255
Географические функции.....	256
Заключение	256
Глава 7. Оконные функции и LOD-выражения.....	257
Оконные функции	257
Агрегатные оконные функции	259
Оконные функции смещения.....	282
Ранжирующие оконные функции	287
Вложенные оконные функции	295
BEFORE FILTER BY	297
Собираем дашборд-шпаргалку по оконным функциям.....	299
LOD-выражения	300
Фиксированный уровень детализации.....	302
Добавление уровней детализации.....	304
Исключение уровней детализации	310
Собираем дашборд-шпаргалку по LOD-выражениям	311
Заключение	311
Глава 8. Инструменты ETL и обработка данных с помощью Python и машинного обучения	312
Инструменты и технологии ETL	312
Пример 1. Из базы в базу с очисткой и агрегацией	314
Пример 2. Поиск ближайших магазинов на карте Москвы.....	321
Обработка данных с применением машинного обучения	331
Пример 1. Прогнозирование на основе линейной регрессии.....	331
Пример 2. Кластеризация магазинов методом k-средних.....	338
Заключение	351
Предметный указатель	352