

O'REILLY®



Python для Excel

Современная среда для автоматизации
и анализа данных



Материалы
на www.bhv.ru

Феликс Зумштейн

Феликс Зумштейн

Python для Excel

Современная среда для автоматизации
и анализа данных

Санкт-Петербург
«БХВ-Петербург»

2024

УДК 004.43
ББК 32.973.26-018.1
3-93

Зумштейн Ф.

3-93 Python для Excel: Пер. с англ. — СПб.: БХВ-Петербург, 2024. — 336 с.: ил.
ISBN 978-5-9775-6882-1

Книга посвящена автоматизации Excel с помощью языка программирования Python. Описаны дистрибутив Anaconda Python и современные средства разработки, такие как менеджеры пакетов Conda и pip, блокноты Jupyter и Visual Studio Code. Даны необходимые основы языка Python и введение в анализ данных с помощью библиотеки pandas. Приведены приемы чтения и записи файлов Excel без Excel. Рассмотрено программирование приложений Excel с помощью популярного пакета с открытым исходным кодом xlwings: автоматизация Excel, инструменты на основе технологии Python, трекер пакетов Python, а также функции, определяемые пользователем.

Электронный архив на сайте издательства содержит цветные иллюстрации к книге.

Для опытных пользователей Excel и программистов

УДК 004.43
ББК 32.973.26-018.1

Научный редактор:

Инженер-тестировщик компании КРОК Дмитрий Колфилд

Группа подготовки издания:

Руководитель проекта	<i>Евгений Рыбаков</i>
Зав. редакцией	<i>Людмила Гауль</i>
Редактор	<i>Наталья Смирнова</i>
Компьютерная верстка	<i>Натальи Смирновой</i>
Оформление обложки	<i>Зои Канторович</i>

© 2023 BHV

Authorized Russian translation of the English edition of Python for Excel ISBN 9781492081005

© 2021 Zoomer Analytics LLC.

This translation is published and sold by permission of O'Reilly Media, Inc., which owns or controls all rights to publish and sell the same.

Авторизованный перевод с английского языка на русский издания Python for Excel ISBN 9781492081005

© 2021 Zoomer Analytics LLC.

Перевод опубликован и продается с разрешения компании-правообладателя O'Reilly Media, Inc.

Подписано в печать 03.11.23.

Формат 70×100^{1/8}. Печать офсетная. Усл. печ. л. 27,09.

Доп. тираж 1500 экз. Заказ № 8046.

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

Отпечатано с готового оригинал-макета

ООО "Принт-М", 142300, М.О., г. Чехов, ул. Полиграфистов, д. 1

ISBN 978-1-492-08100-5 (англ.)
ISBN 978-5-9775-6882-1 (рус.)

© Zoomer Analytics LLC, 2021

© Перевод на русский язык, оформление.

ООО "БХВ-Петербург", ООО "БХВ", 2023, 2024

Содержание

Предисловие	11
Почему я написал эту книгу	12
Кому адресована эта книга	12
Структура книги	13
Версии Python и Excel	14
Условные обозначения, используемые в этой книге	15
Примеры использования кода	15
Иллюстрации	16
Онлайн-обучение O'Reilly	17
Как с нами связаться	17
Благодарности	17
Часть I. Введение в Python	19
Глава 1. Зачем нужен Python для Excel?	21
Excel как язык программирования	22
Excel в новостях	23
Передовые методы программирования	24
Современный Excel	29
Python для Excel	31
Читабельность и эксплуатационная пригодность	31
Стандартная библиотека и менеджер пакетов	32
Научные вычисления	34
Особенности современного языка	35
Кросс-платформенная совместимость	36
Заключение	36
Глава 2. Среда разработки	38
Дистрибутив Anaconda Python	39
Установка	39
Anaconda Prompt	40
Python REPL: интерактивная сессия Python	43
Менеджеры пакетов: Conda и pip	44
Среды Conda	46

Jupyter Notebooks.....	47
Запуск блокнотов Jupyter	48
Ячейки блокнота	49
Режим редактирования и командный режим	51
Порядок выполнения имеет значение	52
Завершение работы блокнотов Jupyter	52
Visual Studio Code	53
Установка и настройка	55
Запуск скрипта на Python	57
Заключение	61
Глава 3. Приступая к работе с Python.....	63
Типы данных	63
Объекты	64
Числовые типы	65
Логический тип данных	67
Строки	69
Индексирование и нарезка	70
Индексирование	70
Нарезка (Slicing).....	71
Структуры данных	72
Списки.....	72
Словари	75
Кортежи	76
Множества	77
Управление потоком	78
Блоки кода и оператор <i>pass</i>	78
Оператор <i>if</i> и условные выражения	79
Циклы <i>for</i> и <i>while</i>	80
Анализ списков, словарей и множеств	83
Организация кода.....	84
Функции.....	84
Модули и инструкция по импорту	86
Класс <i>datetime</i>	88
PEP 8: Руководство по стилю для кода Python.....	90
PEP 8 и VS Code.....	92
Подсказки по типам	93
Заключение	94
Часть II. ВВЕДЕНИЕ В PANDAS	95
Глава 4. Основы NumPy	97
Начало работы с NumPy	97
Массив NumPy	97

Векторизация и транслирование	99
Универсальные функции (ufunc)	101
Создание и манипулирование массивами	102
Получение и установка элементов массива	102
Полезные конструкторы массива	103
Представления и копирование	103
Заключение	104
Глава 5. Анализ данных с помощью pandas	105
DataFrame и Series	105
Индекс	108
Столбцы	110
Манипулирование данными	111
Выбор данных	111
Изменение данных	117
Отсутствующие данные	120
Дубликаты данных	122
Арифметические операции	123
Работа с текстовой колонкой	125
Использование функции	126
Просмотр и копирование	127
Объединение DataFrames	127
Объединение	128
Объединение и слияние	129
Описательная статистика и агрегация данных	132
Описательная статистика	132
Группировка	133
Pivoting и Melting	134
Построение графиков	135
Matplotlib	135
Plotly	137
Импорт и экспорт DataFrames	140
Экспорт CSV файлов	141
Импорт CSV-файлов	142
Заключение	144
Глава 6. Анализ временных рядов с помощью pandas	145
DatetimeIndex	146
Создание DatetimeIndex	146
Фильтрация DatetimeIndex	148
Работа с часовыми поясами	150
Общие манипуляции с временными рядами	151
Смещение и процентные изменения	151
Пересчет и корреляция	153

Повторная выборка	156
Скользящее окно	157
Ограничения при работе с pandas	158
Заключение	159

ЧАСТЬ III. ЧТЕНИЕ И ЗАПИСЬ ФАЙЛОВ EXCEL БЕЗ EXCEL 161

Глава 7. Манипулирование файлами Excel с помощью pandas..... 163

Тематическое исследование: отчетность в Excel	163
Чтение и запись файлов Excel с помощью pandas	167
Функция read_excel и класс ExcelFile	167
Метод to_excel и класс ExcelWriter	173
Ограничения при работе pandas с файлами Excel.....	174
Заключение	175

Глава 8. Манипулирование файлами Excel с помощью пакетов

reader и writer 176

Пакеты reader и writer	176
В каких случаях какой пакет используется	177
Модуль excel.py.....	178
OpenPyXL	180
XlsxWriter	184
pyxlsb.....	186
xlrd, xlwt, and xlutils	187
Работа с xlwt.....	189
Расширенный круг задач для reader и writer.....	190
Работа с большими файлами Excel.....	191
Форматирование данных в Excel.....	195
Тематическое исследование (повторное): отчетность в Excel.....	200
Заключение	201

ЧАСТЬ IV. ПРОГРАММИРОВАНИЕ ПРИЛОЖЕНИЯ EXCEL

С ПОМОЩЬЮ XLWINGS..... 205

Глава 9. Автоматизация Excel 205

Начало работы с xlwings.....	206
Использование Excel в качестве средства просмотра данных.....	206
Объектная модель Excel	208
Запуск кода VBA.....	215
Конвертеры, опции и коллекции	216
Работа с DataFrames	216
Конвертеры и опции	217

Диаграммы, рисунки и определенные имена	220
Случай из практики (повторный анализ): отчетность в Excel	223
Расширенные темы xlwings	225
Основы xlwings	225
Улучшение производительности	227
Как действовать при отсутствии недостающих функций	229
Заключение	230
Глава 10. Инструменты Excel на основе технологии Python	231
Использование Excel в качестве интерфейса xlwings	231
Настройка Excel	232
Команда Quickstart	233
Run Main	234
Функция RunPython	235
Развертывание	240
Зависимости Python	240
Автономные рабочие книги: избавление от настройки xlwings	241
Иерархия конфигурации	242
Настройки	243
Заключение	244
Глава 11. Трекер пакетов Python	245
Что мы будем создавать	245
Основной функционал	247
Web APIs	248
Базы данных	251
Исключения	260
Структура приложения	263
Внешний интерфейс	264
Внутренний интерфейс	268
Отладка	271
Заключение	273
Глава 12. Функции, определяемые пользователем (UDFs)	274
Начало работы с UDF	274
UDF Quickstart	275
Тематическое исследование: Google Trends	280
Введение в Google Trends	280
Работа с DataFrames и динамическими массивами	282
Получение данных из Google Trends	287
Построение графиков с помощью UDF	291
Отладка UDFs	293
Дополнительные вопросы по UDF	294
Базовая оптимизация производительности	295

Кэширование	297
Декоратор Sub	299
Заключение	301
Часть V. Приложения.....	303
Приложение А. Среда Conda.....	305
Создание новой среды Conda.....	305
Отключение автоматической активации.....	307
Приложение В. Расширенные функциональные возможности VS Code.....	308
Отладчик	308
Блокноты Jupyter в VS Code.....	310
Запуск блокнотов Jupyter	310
Сценарии Python с ячейками кода.....	311
Приложение С. Дополнительные концепции Python	313
Классы и объекты.....	313
Работа с объектами datetime с учетом временной зоны	315
Изменяемые и неизменяемые объекты Python	316
Вызов функций с изменяемыми объектами в качестве аргументов	317
Функции с изменяемыми объектами в качестве аргументов по умолчанию.....	319
Об авторе	321
Обложка	323
Предметный указатель	325