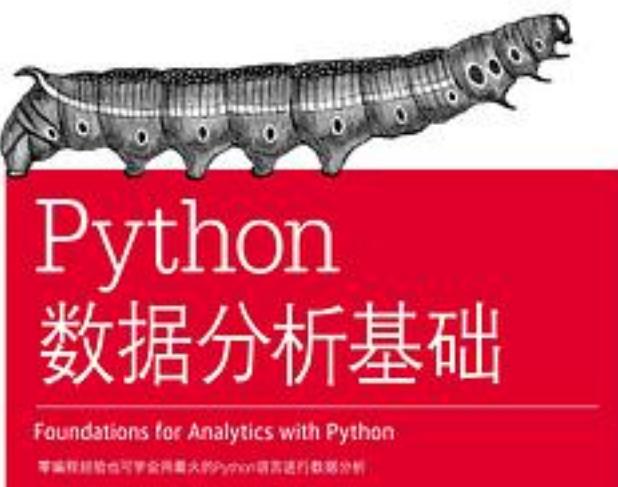


# Python数据分析基础

O'REILLY®

TURING 图灵程序设计丛书



[美] Clinton W. Brownley 著  
陈光欣 译

中国工信出版集团 人民邮电出版社

[Python数据分析基础 下载链接1](#)

著者:[美] Clinton W. Brownley

出版者:人民邮电出版社

出版时间:2017-8

装帧:平装

isbn:9787115463357

想深入应用手中的数据？还是想在上千份文件上重复同样的分析过程？没有编程经验的非程序员们如何能在最短的时间内学会用当今炙手可热的Python语言进行数据分析？

来自Facebook的数据专家Clinton Brownley可以帮您解决上述问题。在他的这本书里，读者将能掌握基本Python编程方法，学会编写出处理电子表格和数据库中的数据的脚本，并了解使用Python模块来解析文件、分组数据和生成统计量的方法。

学习基础语法，创建并运行自己的Python脚本

读取和解析CSV文件

读取多个Excel工作表和工作簿

执行数据库操作

搜索特定记录、分组数据和解析文本文件

建立统计图并绘图

生成描述性统计量并估计回归模型和分类模型

在Windows和Mac环境下按计划自动运行脚本

作者介绍:

Clinton W. Brownley

博士，Facebook数据科学家，负责大数据流水线、统计建模和数据可视化项目，并为大型基础设施建设提供数据驱动的决策建议。

目录: 前言 xi

第1章 Python 基础 1

1.1 创建 Python 脚本 1

1.2 运行 Python 脚本 3

1.3 与命令行进行交互的几项技巧 6

1.4 Python 语言基础要素 10

1.4.1 数值 10

1.4.2 字符串 12

1.4.3 正则表达式与模式匹配 16

1.4.4 日期 19

1.4.5 列表 21

1.4.6 元组 26

1.4.7 字典 27

1.4.8 控制流 30

1.5 读取文本文件 35

1.5.1 创建文本文件 36

1.5.2 脚本和输入文件在同一位置 38

1.5.3 读取文件的新型语法 38

1.6 使用 glob 读取多个文本文件 39

1.7 写入文本文件 42

1.7.1 向 first\_script.py 添加代码 42

1.7.2 写入 CSV 文件 45

1.8 print 语句 46

1.9 本章练习 47

第2章 CSV文件 48

2.1 基础 Python 与 pandas 50

2.1.1 读写 CSV 文件 (第1部分) 50

2.1.2 基本字符串分析是如何失败的 56

2.1.3 读写 CSV 文件 (第2部分) 57

2.2 筛选特定的行 58

2.2.1 行中的值满足某个条件 59

2.2.2 行中的值属于某个集合 60

2.2.3 行中的值匹配于某个模式/ 正则表达式 62

2.3 选取特定的列 64

2.3.1 列索引值 64

2.3.2 列标题 65

2.4 选取连续的行 67

2.5 添加标题行 69

2.6 读取多个 CSV 文件 71

2.7 从多个文件中连接数据 75

2.8 计算每个文件中值的总和与均值 78

2.9 本章练习 81

第3章 Excel 文件 82

3.1 内省 Excel 工作簿 84

3.2 处理单个工作表 88

3.2.1 读写 Excel 文件 88

3.2.2 筛选特定行 92

3.2.3 选取特定列 98

3.3 读取工作簿中的所有工作表 101

3.3.1 在所有工作表中筛选特定行 102

3.3.2 在所有工作表中选取特定列 104

3.4 在 Excel 工作簿中读取一组工作表 106

3.5 处理多个工作簿 108

3.5.1 工作表计数以及每个工作表中的行列计数 110

3.5.2 从多个工作簿中连接数据 111

3.5.3 为每个工作簿和工作表计算总数和均值 113

3.6 本章练习 117

第4章 数据库 118

4.1 Python 内置的 sqlite3 模块 119

4.1.1 向表中插入新记录 124

4.1.2 更新表中记录 128

4.2 MySQL 数据库 131

4.2.1 向表中插入新记录 135

4.2.2 查询一个表并将输出写入 CSV 文件 140

4.2.3 更新表中记录 142

4.3 本章练习 146

第5章 应用程序 147

5.1 在一个大文件集合中查找一组项目 147

5.2 为 CSV 文件中数据的任意数目分类计算统计量 158

5.3 为文本文件中数据的任意数目分类计算统计量 167

5.4 本章练习 174

第6章 图与图表 175

6.1 matplotlib 175

6.1.1 条形图 175

6.1.2 直方图 177

6.1.3 折线图 178

6.1.4 散点图 180

6.1.5 箱线图 181

6.2 pandas	183
6.3 ggplot	184
6.4 seaborn	186
第7章 描述性统计与建模	192
7.1 数据集	192
7.1.1 葡萄酒质量	192
7.1.2 客户流失	193
7.2 葡萄酒质量	194
7.2.1 描述性统计	194
7.2.2 分组、直方图与 t 检验	195
7.2.3 成对变量之间的关系和相关性	196
7.2.4 使用最小二乘估计进行线性回归	198
7.2.5 系数解释	200
7.2.6 自变量标准化	200
7.2.7 预测	202
7.3 客户流失	203
7.3.1 逻辑斯蒂回归	205
7.3.2 系数解释	207
7.3.3 预测	208
第8章 按计划自动运行脚本	209
8.1 任务计划程序 (Windows 系统)	209
8.2 cron 工具 (macOS 系统和 Unix 系统)	215
8.2.1 cron 表文件：一次性设置	216
8.2.2 向 cron 表文件中添加 cron 任务	216
第9章 从这里启航	220
9.1 更多的标准库模块和内置函数	221
9.1.1 Python 标准库 (PSL)：更多的标准模块	221
9.1.2 内置函数	222
9.2 Python 包索引 (PyPI)：更多的扩展模块	222
9.2.1 NumPy	223
9.2.2 SciPy	227
9.2.3 Scikit-Learn	230
9.2.4 更多的扩展包	232
9.3 更多的数据结构	232
9.3.1 栈	233
9.3.2 队列	233
9.3.3 图	233
9.3.4 树	234
9.4 从这里启航	234
附录A 下载指南	236
附录B 练习答案	245
作者介绍	247
封面介绍	247
· · · · · (收起)	

[Python数据分析基础](#) [下载链接1](#)

标签

Python

数据分析

python

编程

计算机

数据科学

工程

入门

评论

数据分析是Python的几大应用场景之一，挺好的入门书。内容涵盖了Python语言基础、CSV/Excel文件的读写和数据处理、SQLite数据库的基本操作、Matplotlib可视化图表简介，以及一些分析实例。其中对CSV/Excel文件操作部分颇为经典，先讲一遍使用基本模块的操作方法，再讲一遍使用Pandas的操作方法，讲解的很细致，对初学者颇为友好。

---

和另外一本python基础的内容格式没有统一，对于初学者会造成一定的困扰。甚至还需要一些linux作为背景。comments全都写在后面，不方便查阅。

---

过了一遍…… 太啰嗦了 原生python有必要讲那么多么？？

---

一开始觉得太简单了，但是硬着头皮照着代码一行行执行，再按书上文字一行行画线读下来，还是有收获的，特别是对CSV、SQLite、Excel、glob的处理套路(几个for循环)；Application应用章节对Map两层嵌套统计、输出的应用，以及对Matplotlib的介绍，很

实用（立马就能照着画图、统计用）。但是机器学习、回归那一章，说的过深，并且讲的库有点过时了。总体还是值得读的，这本虫书，和蜥蜴书我都买了，但似乎得分都没有《利用Python进行数据分析》评分高。具体笔记详见《Python数据分析基础》.mm freemind格式思维导图笔记。

每行的解释很详细。

应该利用电脑点击操作经验，把每一句直接理解成菜单式操作，把编程理解成机械式操作，这样学习事半功倍。

Python数据分析IO基础。

非常一般，不明了。

我，概率统计这四个字都不会写了，第六章看不懂。这本儿，绝大部分内容很基础，很贫嘴，很无聊，很没用。

统计建模值得看看

一般般，里面内容还是挺简单的，在日常工作中，通过实践基本都能掌握。

篇幅合理，难度不大适合入门

第一章说实话对于零基础的人来说容易被吓到，成功的从入门到放弃，可以跳过第一章，书内容不错，但是并没有一种循序渐进的感觉。

最有用处的是第二章和第三章，其他的要么过浅，要么过深都不适合。

非常的基础，推荐一点也不懂 Python 的同学看。

不错，极个别代码有误

最基础的入门python数据分析的书籍，除了python的一个入门，还介绍了常用的数据录入，Pandas，以及一些基本的matplotlib画图。整体浅尝辄止，没有具体案例，适合作为了解python数据分析的书

适合数据分析新手，但需要掌握一定的python语法基础才可以。不然因为作者很粗心，文中有许多小bug。比如，明明说是python3编程，但书中很多代码print没加括号。。。另外，明明是讲解pandas的，但是里面涉及的新函数，语法讲解不清，或者根本就不给介绍。读者要么就得囫囵吞枣，要么就得一遍读，一遍查阅资料才可以。。。不然真会把你搞晕。。唯一的好处就是书中代码都是一遍基础python编，一遍pandas编，让你能明显体会到pandas的好处。这三星是给思路的。但是作者的责任心真是“呵呵哒。”

有些代码过时了，如果版本不一致，需要自己摸索着修改，还是有一部分没能摸索出…书中存在好多印刷的错误，总之还是涨了一些知识。

对很多工具用法进行了逐行解释，比较浅显，适合于初学者

[Python数据分析基础 下载链接1](#)

书评

想深入应用手中的数据？还是想在上千份文件上重复同样的分析过程？没有编程经验的非程序员们如何能在最短的时间内学会用当今炙手可热的Python语言进行数据分析？来自Facebook的数据专家Clinton Brownley可以帮您解决上述问题。在他的这本书里，读者将能掌握基本Python编程方

法...

---

[Python数据分析基础 下载链接1](#)