· 入职数据分析师系列 ·

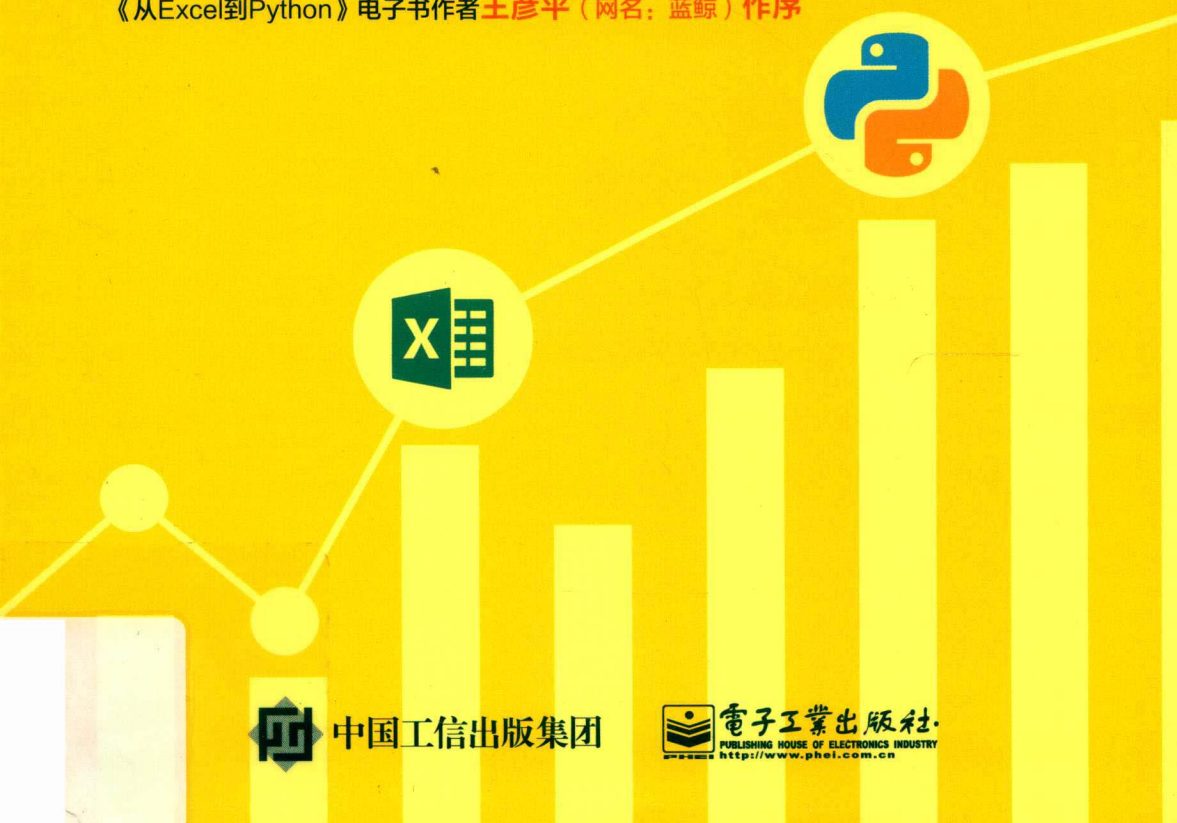
Broadview
www.broadview.com.cn

对比Excel，轻松学习 Python 数据分析

张俊红 著

集Python、Excel、数据分析于一体，是数据从业者案头实操工具书

《从Excel到Python》电子书作者王彦平（网名：蓝鲸）作序



中国工信出版集团

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

入职数据分析师系列 ·

对比Excel，轻松学习 Python 数据分析

张俊红 著

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

内 容 简 介

集Python、Excel、数据分析为一体是本书的一大特色。

本书围绕整个数据分析的常规流程：熟悉工具—明确目的—获取数据—熟悉数据—处理数据—分析数据—得出结论—验证结论—展示结论进行Excel和Python的对比实现，告诉你每一个过程中都会用到什么，过程与过程之间有什么联系。本书既可以作为系统学习数据分析操作流程的说明书，也可以作为一本数据分析师案头必备的实操工具书。

本书通过对比Excel功能操作去学习Python的代码实现，而不是直接学习Python代码，大大降低了学习门槛，消除了读者对代码的恐惧心理。适合刚入行的数据分析师，也适合对Excel比较熟练的数据分析师，以及从事其他岗位想提高工作效率的职场人。

图书在版编目（CIP）数据

对比 Excel，轻松学习 Python 数据分析 / 张俊红著. —北京：电子工业出版社，2019.2
（入职数据分析师系列）
ISBN 978-7-121-35793-0

Ⅰ. ①对… Ⅱ. ①张… Ⅲ. ①软件工具－程序设计 Ⅳ. ①TP311.561

中国版本图书馆 CIP 数据核字(2018)第 279763 号

策划编辑：张慧敏
责任编辑：汪达文
印　　刷：三河市华成印务有限公司
装　　订：三河市华成印务有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编：100036
开　　本：720×1000　1/16　印张：17.75　字数：365 千字　彩插：1
版　　次：2019 年 2 月第 1 版
印　　次：2019 年 2 月第 1 次印刷
印　　数：3000 册　定价：59.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：010-51260888-819，faq@phei.com.cn。

序　言

有幸收到张俊红的做序邀请，我非常高兴。

从 PC 时代到移动互联网时代一路走来，每个人都感受到了数据爆炸性的增长，以及其中蕴含的巨大价值。

从 PC 时代开始，我们用键盘、扫描仪等设备使信息数据化。在移动互联网时代，智能手机通过摄像头、GPS、陀螺仪等各种传感器将我们的位置、行动轨迹、行为偏好，甚至情绪等信息数据化。截至 2000 年，全人类存储了大约 12EB 的数据，要知道 1PB=1024TB，而 1EB=1024PB。但是到了 2011 年，一年所产生的数据就高达 1.82ZB（注：1ZB=1024EB），数据已经变成了一种人造的“新能源”。

在商业领域，从信息到商品，从商品到服务，越来越多我们熟悉的事物被标准的数据所度量。无论是在线广告的精准营销，还是电子商务的个性化推荐，又或者是互联网金融的人脸识别，互联网的每一次效率提升都依赖于对传统信息、物品，甚至人的数据化。

在使用数据进行效率变革及商业化的道路上，Excel 和 Python 扮演了关键的角色，它们帮助数据分析师高效地从海量数据中发现问题，验证假设，搭建模型，预测未来。

作为一本数据分析的专业书籍，作者从数据采集、清洗、抽取，以及数据可视化等多个角度介绍了日常工作中数据分析的标准路径。通过对比 Excel 与 Python 在数据处理过程中的操作步骤，详细说明了 Excel 与 Python 间的差异，以及用 Python 进行数据分析的方法。

虽与作者素未谋面，但是对于 Python 在处理海量数据和建模上的高效性与便捷性，以及 Python 在机器学习中的重要性，我们的观点是一致的。同时我们也相信对于数据分析从业者来说，掌握一种用于数据处理的编程语言是非常必要的，而从 Excel 到 Python 的学习方法则是一条学好数据分析的“捷径”。

王彦平

（网名“蓝鲸”，电子书《从 Excel 到 Python——数据分析进阶指南》《从 Excel 到 R——数据分析进阶指南》《从 Excel 到 SQL——数据分析进阶指南》的作者）

2019 年 1 月 8 日

前　言

为什么要写这本书

本书既是一本数据分析的书，也是一本 Excel 数据分析的书，同时还是一本 Python 数据分析的书。在互联网上，无论是搜索数据分析，还是搜索 Excel 数据分析，亦或是搜索 Python 数据分析，我们都可以找到很多相关的图书。既然已经有这么多同类题材的书了，为什么我还要写呢？因为在我准备写这本书时，还没有一本把数据分析、Excel 数据分析、Python 数据分析这三者结合在一起的书。

为什么我要把它们结合在一起写呢？那是因为，我认为这三者是一个数据分析师必备的技能，而且这三者本身也是一个有机统一体。数据分析让你知道怎么分析以及分析什么；Excel 和 Python 是你在分析过程中会用到的两个工具。

为什么要学习 Python

既然 Python 在数据分析领域是一个和 Excel 类似的数据分析工具，二者实现的功能都一样，为什么还要学 Python，把 Excel 学好不就行了吗？我认为学习 Python 的主要原因有以下几点。

1．在处理大量数据时，Python 的效率高于 Excel

当数据量很小的时候，Excel 和 Python 的处理速度基本上差不多，但是当数据量较大或者公式嵌套太多时，Excel 就会变得很慢，这个时候怎么办呢？我们可以使用 Python，Python 对于海量数据的处理效果要明显优于 Excel。用 Vlookup 函数做一个实验，两个大小均为 23MB 的表（6 万行数据），在未作任何处理、没有任何公式嵌套之前，Excel 中直接在一个表中用 Vlookup 函数获取另一个表的数据需要 20 秒（我的计算机性能参数是 I7、8GB 内存、256GB 固态硬盘），配置稍微差点的计算机可能打开这个表都很难。但是用 Python 实现上述过程只需要 580 毫秒，即 0.58 秒，是 Excel 效率的 34 倍。

2．Python 可以轻松实现自动化

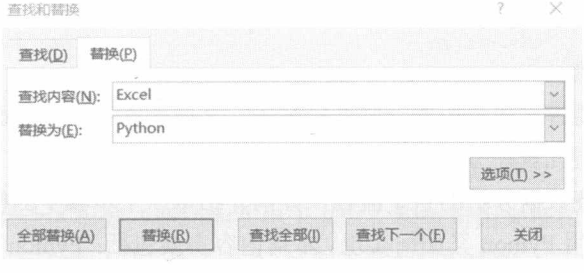
你可能会说 Excel 的 VBA 也可以自动化，但是 VBA 主要还是基于 Excel 内部的自动化，一些其他方面的自动化 VBA 就做不了，比如你要针对本地某一文件夹下面的文件名进行批量修改，VBA 就不能实现，但是 Python 可以。

3．Python 可用来做算法模型

虽然你是做数据分析的，但是一些基础的算法模型还是有必要掌握的，Python 可以让你在懂一些基础的算法原理的情况下就能搭建一些模型，比如你可以使用聚类算法搭建一个模型去对用户进行分类。

为什么要对比 Excel 学习 Python

Python 虽然是一门编程语言，但是在数据分析领域实现的功能和 Excel 的基本功能一样，而 Excel 又是大家比较熟悉、容易上手的软件，所以可以通过 Excel 数据分析去对比学习 Python 数据分析。对于同一个功能，本书告诉你在 Excel 中怎么做，并告诉你对应到 Python 中是什么样的代码。例如数值替换，即把一个值替换成另一个值，对把“Excel”替换成“Python”这一要求，在 Excel 中可以通过鼠标点选实现，如下图所示。



在 Python 中则通过具体的代码实现，如下所示。

```
df.replace("Excel","Python")    # 表示将表 df 中的 Excel 替换成 Python
```

本书将数据分析过程中涉及的每一个操作都按这种方式对照讲解，让你从熟悉的 Excel 操作中去学习对应的 Python 实现，而不是直接学习 Python 代码，大大降低了学习门槛，消除了大家对代码的恐惧心理。这也是本书的一大特色，也是我为什么要写本书的最主要原因，就是希望帮助你不再惧怕代码，让你可以像学 Excel 数据分析一样，轻松学习 Python 数据分析。

本书的学习建议

要想完全掌握一项技能，你必须系统学习它，知道它的前因后果。本书不是孤立地讲 Excel 或者 Python 中的操作，而是围绕整个数据分析的常规流程：熟悉工具—明确目的—获取数据—熟悉数据—处理数据—分析数据—得出结论—验证结论—展示结论，告诉你每一个过程都会用到什么操作，这些操作用 Excel 和 Python 分别怎么实现。这样一本书既是系统学习数据分析流程操作的说明书，也是数据分析师案头必备的实操工具书。

大家在读第一遍的时候不用记住所有函数，你是记不住的，即使你记住了，如果在工作中不用，那么很快就会忘记。正确的学习方式应该是，先弄清楚一名数据分析师在日常工作中对工具都会有什么需求（当然了，本书的顺序是按照数据分析的常规分析流程来写的），希望工具帮助你达到什么样的目的，罗列好需求以后，再去研究工具的使用方法。比如，要删除重复值，就要明确用 Excel 如何实现，用 Python 又该如何实现，两种工具在实现方式上有什么异同，这样对比次数多了以后，在遇到问题时，你自然而然就能用最快的速度选出最适合的工具了。

数据分析一定是先有想法然后考虑如何用工具实现，而不是刚开始就陷入记忆工具的使用方法中。

本书写了什么

本书分为三篇。

入门篇：主要讲数据分析的一些基础知识，介绍数据分析是什么，为什么要做数据分析，数据分析究竟在分析什么，以及数据分析的常规流程。

实践篇：围绕数据分析的整个流程，分别介绍每一个步骤中的操作，这些操作用 Excel 如何实现，用 Python 又如何实现。本篇内容主要包括：Python 环境配置、Python 基础知识、数据源的获取、数据概览、数据预处理、数值操作、数据运算、时间序列、数据分组、数据透视表、结果文件导出、数据可视化等。

进阶篇：介绍几个实战案例，让你体会一下在实际业务中如何使用 Python。具体来说，进阶篇的内容主要包括，利用 Python 实现报表自动化、自动发送电子邮件，以及在不同业务场景中的案例分析。此外，还补充介绍了 NumPy 数组的一些常用方法。

本书适合谁

本书主要适合以下人群。

- Excel 已经用得熟练，想学习 Python 来丰富自己技能的数据分析师。

- 刚入行对 Excel 和 Python 都不精通的数据分析师。
- 其他常用 Excel 却想通过学习 Python 提高工作效率的人。

Python 虽然是一门编程语言，但是它并不难学，不仅不难学，而且很容易上手，这也是 Python 深受广大数据从业者喜爱的原因之一，因此大家在学习 Python 之前首先在心里告诉自己一句话，那就是 Python 并没有那么难。

致谢

感谢我的父母，是他们给了我受教育的机会，才有了今天的我。

感谢我的公众号的读者朋友们，如果不是他们，那么我可能不会坚持撰写技术文章，更不会有这本书。

感谢慧敏让我意识到写书的意义，从而创作本书，感谢电子工业出版社为这本书忙碌的所有人。

感谢我的女朋友，在写书的这段日子里，我几乎把所有的业余时间全用在了写作上，很少陪她，但她还是一直鼓励我，支持我。

读者服务

轻松注册成为博文视点社区用户（www.broadview.com.cn），扫码直达本书页面。

- **提交勘误**：您对书中内容的修改意见可在 提交勘误 处提交，若被采纳，将获赠博文视点社区积分（在您购买电子书时，积分可用来抵扣相应金额）。
- **交流互动**：在页面下方 读者评论 处留下您的疑问或观点，与我们和其他读者一同学习交流。

页面入口：http://www.broadview.com.cn/35793

目　录

入门篇

实践篇

进阶篇