水木书荟

# Python 数据分析与实战

吕云翔　李伊琳　王肇一　张雅素　编著

清华大学出版社
北 京

## 内 容 简 介

使用 Python 进行数据分析是十分便利且高效的，因此它被认为是最优秀的数据分析工具之一。本书从理论和实战两个角度对 Python 数据分析工具进行了介绍，并采用理论分析和 Python 实践相结合的形式，按照数据分析的基本步骤对数据分析的理论知识以及相应的 Python 库进行了详细的介绍，让读者在了解数据分析的基本理论知识的同时能够快速上手实现数据分析程序。

本书适用于对数据分析有浓厚兴趣但不知从何下手的初学者，在阅读数据分析的基础理论知识的同时可以通过 Python 实现简单的数据分析程序，从而快速对数据分析的理论和实现两个层次形成一定的认知。



**责任编辑：**魏江江　王冰飞
**封面设计：**
**责任校对：**徐俊伟
**责任印制：**

**出版发行：**清华大学出版社
**网　　址：**http://www.tup.com.cn，http://www.wqbook.com
**地　　址：**北京清华大学学研大厦 A 座　　**邮　　编：**100084
**社 总 机：**010-62770175　　**邮　　购：**010-62786544
**投稿与读者服务：**010-62776969，c-service@tup.tsinghua.edu.cn
**质量反馈：**010-62772015，zhiliang@tup.tsinghua.edu.cn
**课件下载：**http://www.tup.com.cn,010-62795954
**印 刷 者：**
**装 订 者：**
**经　　销：**全国新华书店
**开　　本：**185mm×260mm　　**印　张：**12.25　　**字　　数：**211 千字
**版　　次：**2019 年 1 月第 1 版　　**印　　次：**2019 年 1 月第1次印刷
**定　　价：**　.00 元

产品编号：077077-01

# 前 言

本书是面向初学者的数据分析入门指南。按照数据分析的数据预处理、分析与知识发现和可视化3个主要步骤，本书逐步对数据分析涉及的理论进行讲解，并对实现这些步骤所用到的Python库进行详细介绍。通过理论与实践穿插的讲解方式，本书使读者能够在了解数据分析基础知识的同时快速上手实现一些简单的分析。

全书分为10章，第1、3、6章介绍数据分析理论，按照数据分析的基本流程介绍了理论知识和一些常用方法，穿插在理论章节之间的Python实战章节可以让读者在了解理论之后用相应的Python库来进行实战操作。通过阅读第1～8章的内容，读者已经对数据分析的各主要流程形成了一定的认识，但这些知识可能还未形成一个完整的体系，因此本书在第9和第10章引入了两个完整的数据分析实例，帮助读者建立知识点之间的联系，形成对数据分析整个知识面的清晰认知。建议读者在阅读实战章节时跟随介绍自己动手尝试一下，这样一定会发现数据的魅力所在。

作为一本数据分析入门书籍，本书着重介绍基础知识，对前沿的内容涉及较少，这些内容留待读者在更进一步的学习中深入探索。对于Python语言的知识，本书仅对与数据分析相关的库进行了介绍，如果读者对Python语言本身有兴趣，可以参考Python语言工具书及官方文档等详细了解Python的语法和底层原理等。另外，本书所有数据分析程序的实现均在单机情况下进行，并没有对如何使用Python进行分布式数据分析作介绍，有兴趣的读者可以了解一下Python分布式数据分析的相关库，如pyspark等。

本书主要由吕云翔、李伊琳、王肇一、张雅素编写，曾洪立、吕彼佳、姜彦华也参与了部分内容的编写并进行了素材整理及配套资源制作等。

由于作者的水平和能力有限，本书难免有疏漏之处，恳请各位同仁和广大读者给予批评指正，也希望各位能将实践过程中的经验和心得与我们交流（yunxianglu@hotmail.com）。

作　者

2018年7月

# 目 录