

对比Excel，轻松学习

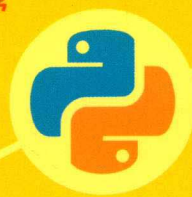
# Python

## 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



电子工业出版社  
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

电子工业出版社

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。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。  
版权所有，侵权必究。

### 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 [zltz@phei.com.cn](mailto:zltz@phei.com.cn)，盗版侵权举报请发邮件至 [dbqq@phei.com.cn](mailto:dbqq@phei.com.cn)。

本书咨询联系方式：010-51260888-819, [faq@phei.com.cn](mailto: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，这些操作在 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，这些操作在 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](http://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

第 1 章 数据分析基础 .....	2
1.1 数据分析是什么 .....	2
1.2 为什么要做数据分析 .....	2
1.2.1 现状分析 .....	3
1.2.2 原因分析 .....	3
1.2.3 预测分析 .....	3
1.3 数据分析究竟在分析什么 .....	4
1.3.1 总体概览指标 .....	4
1.3.2 对比性指标 .....	4
1.3.3 集中趋势指标 .....	4
1.3.4 离散程度指标 .....	5
1.3.5 相关性指标 .....	5
1.3.6 相关关系与因果关系 .....	6
1.4 数据分析的常规流程 .....	6
1.4.1 熟悉工具 .....	6
1.4.2 明确目的 .....	7
1.4.3 获取数据 .....	7
1.4.4 熟悉数据 .....	7
1.4.5 处理数据 .....	7
1.4.6 分析数据 .....	8
1.4.7 得出结论 .....	8
1.4.8 验证结论 .....	8
1.4.9 展示结论 .....	8
1.5 数据分析工具：Excel 与 Python .....	8

## 实践篇

第 2 章 熟悉锅——Python 基础知识 .....	12
2.1 Python 是什么 .....	12

2.2	Python 的下载与安装.....	13
2.2.1	安装教程.....	13
2.2.2	IDE 与 IDLE.....	17
2.3	介绍 Jupyter Notebook .....	17
2.3.1	新建 Jupyter Notebook 文件 .....	17
2.3.2	运行你的第一段代码.....	19
2.3.3	重命名 Jupyter Notebook 文件 .....	19
2.3.4	保存 Jupyter Notebook 文件 .....	19
2.3.5	导入本地 Jupyter Notebook 文件 .....	20
2.3.6	Jupyter Notebook 与 Markdown.....	21
2.3.7	为 Jupyter Notebook 添加目录 .....	21
2.4	基本概念.....	26
2.4.1	数.....	26
2.4.2	变量.....	26
2.4.3	标识符.....	27
2.4.4	数据类型.....	28
2.4.5	输出与输出格式设置.....	28
2.4.6	缩进与注释.....	29
2.5	字符串.....	30
2.5.1	字符串的概念.....	30
2.5.2	字符串的连接.....	30
2.5.3	字符串的复制.....	30
2.5.4	获取字符串的长度.....	30
2.5.5	字符串查找.....	31
2.5.6	字符串索引.....	31
2.5.7	字符串分隔.....	32
2.5.8	移除字符.....	32
2.6	数据结构——列表.....	33
2.6.1	列表的概念.....	33
2.6.2	新建一个列表.....	33
2.6.3	列表的复制.....	34
2.6.4	列表的合并.....	34
2.6.5	向列表中插入新元素.....	34
2.6.6	获取列表中值出现的次数.....	35
2.6.7	获取列表中值出现的位置.....	35
2.6.8	获取列表中指定位置的值.....	36
2.6.9	删除列表中的值.....	36
2.6.10	对列表中的值进行排序.....	37
2.7	数据结构——字典.....	37
2.7.1	字典的概念.....	37
2.7.2	新建一个字典.....	37



2.7.3	字典的 keys()、values()和 items()方法	37
2.8	数据结构——元组	38
2.8.1	元组的概念	38
2.8.2	新建一个元组	38
2.8.3	获取元组的长度	38
2.8.4	获取元组内的元素	39
2.8.5	元组与列表相互转换	39
2.8.6	zip()函数	39
2.9	运算符	40
2.9.1	算术运算符	40
2.9.2	比较运算符	40
2.9.3	逻辑运算符	41
2.10	循环语句	41
2.10.1	for 循环	41
2.10.2	while 循环	42
2.11	条件语句	43
2.11.1	if 语句	43
2.11.2	else 语句	44
2.11.3	elif 语句	45
2.12	函数	46
2.12.1	普通函数	47
2.12.2	匿名函数	48
2.13	高级特性	49
2.13.1	列表生成式	49
2.13.2	map 函数	50
2.14	模块	50
<b>第 3 章 Pandas 数据结构</b>		<b>51</b>
3.1	Series 数据结构	51
3.1.1	Series 是什么	51
3.1.2	创建一个 Series	52
3.1.3	利用 index 方法获取 Series 的索引	53
3.1.4	利用 values 方法获取 Series 的值	53
3.2	DataFrame 表格型数据结构	53
3.2.1	DataFrame 是什么	53
3.2.2	创建一个 DataFrame	54
3.2.3	获取 DataFrame 的行、列索引	56
3.2.4	获取 DataFrame 的值	56

第4章 准备食材——获取数据源.....	57
4.1 导入外部数据.....	57
4.1.1 导入.xlsx文件.....	57
4.1.2 导入.csv文件.....	60
4.1.3 导入.txt文件.....	63
4.1.4 导入.sql文件.....	65
4.2 新建数据.....	67
4.3 熟悉数据:.....	67
4.3.1 利用 head 预览前几行.....	67
4.3.2 利用 shape 获取数据表的大小.....	68
4.3.3 利用 info 获取数据类型.....	69
4.3.4 利用 describe 获取数值分布情况.....	71
第5章 淘米洗菜——数据预处理.....	73
5.1 缺失值处理.....	73
5.1.1 缺失值查看.....	73
5.1.2 缺失值删除.....	75
5.1.3 缺失值填充.....	77
5.2 重复值处理.....	78
5.3 异常值的检测与处理.....	81
5.3.1 异常值检测.....	81
5.3.2 异常值处理.....	82
5.4 数据类型转换.....	83
5.4.1 数据类型.....	83
5.4.2 类型转换.....	84
5.5 索引设置.....	86
5.5.1 为无索引表添加索引.....	86
5.5.2 重新设置索引.....	87
5.5.3 重命名索引.....	88
5.5.4 重置索引.....	89
第6章 菜品挑选——数据选择.....	91
6.1 列选择.....	91
6.1.1 选择某一列/某几列.....	91
6.1.2 选择连续的某几列.....	92
6.2 行选择.....	93
6.2.1 选择某一行/某几行.....	93
6.2.2 选择连续的某几行.....	94
6.2.3 选择满足条件的行.....	95

6.3 行列同时选择.....	96
6.3.1 普通索引+普通索引选择指定的行和列.....	97
6.3.2 位置索引+位置索引选择指定的行和列.....	97
6.3.3 布尔索引+普通索引选择指定的行和列.....	98
6.3.4 切片索引+切片索引选择指定的行和列.....	98
6.3.5 切片索引+普通索引选择指定的行和列.....	99
<b>第7章 切配菜品——数值操作</b> .....	<b>100</b>
7.1 数值替换.....	100
7.1.1 一对一替换.....	100
7.1.2 多对一替换.....	102
7.1.3 多对多替换.....	103
7.2 数值排序.....	104
7.2.1 按照一列数值进行排序.....	104
7.2.2 按照有缺失值的列进行排序.....	106
7.2.3 按照多列数值进行排序.....	106
7.3 数值排名.....	108
7.4 数值删除.....	110
7.4.1 删除列.....	110
7.4.2 删除行.....	111
7.4.3 删除特定行.....	112
7.5 数值计数.....	113
7.6 唯一值获取.....	114
7.7 数值查找.....	115
7.8 区间切分.....	116
7.9 插入新的行或列.....	119
7.10 行列互换.....	120
7.11 索引重塑.....	121
7.12 长宽表转换.....	122
7.12.1 宽表转换为长表.....	123
7.12.2 长表转换为宽表.....	125
7.13 apply()与 applymap()函数.....	126
<b>第8章 开始烹调——数据运算</b> .....	<b>127</b>
8.1 算术运算.....	127
8.2 比较运算.....	128
8.3 汇总运算.....	129
8.3.1 count 非空值计数.....	129

8.3.2	sum 求和	130
8.3.3	mean 求均值	130
8.3.4	max 求最大值	131
8.3.5	min 求最小值	132
8.3.6	median 求中位数	132
8.3.7	mode 求众数	133
8.3.8	var 求方差	134
8.3.9	std 求标准差	134
8.3.10	quantile 求分位数	135
8.4	相关性运算	136
<b>第 9 章</b>	<b>炒菜计时器——时间序列</b>	<b>138</b>
9.1	获取当前时刻的时间	138
9.1.1	返回当前时刻的日期和时间	138
9.1.2	分别返回当前时刻的年、月、日	138
9.1.3	返回当前时刻的周数	139
9.2	指定日期和时间的格式	140
9.3	字符串和时间格式相互转换	141
9.3.1	将时间格式转换为字符串格式	141
9.3.2	将字符串格式转换为时间格式	141
9.4	时间索引	142
9.5	时间运算	145
9.5.1	两个时间之差	145
9.5.2	时间偏移	145
<b>第 10 章</b>	<b>菜品分类——数据分组/数据透视表</b>	<b>148</b>
10.1	数据分组	148
10.1.1	分组键是列名	150
10.1.2	分组键是 Series	151
10.1.3	神奇的 aggregate 方法	152
10.1.4	对分组后的结果重置索引	153
10.2	数据透视表	154
<b>第 11 章</b>	<b>水果拼盘——多表拼接</b>	<b>158</b>
11.1	表的横向拼接	158
11.1.1	连接表的类型	158
11.1.2	连接键的类型	160
11.1.3	连接方式	163
11.1.4	重复列名处理	165
11.2	表的纵向拼接	165

11.2.1	普通合并	166
11.2.2	索引设置	167
11.2.3	重叠数据合并	167
<b>第 12 章 盛菜装盘——结果导出</b>		<b>169</b>
12.1	导出为.xlsx 文件	169
12.1.1	设置文件导出路径	170
12.1.2	设置 Sheet 名称	170
12.1.3	设置索引	170
12.1.4	设置要导出的列	171
12.1.5	设置编码格式	171
12.1.6	缺失值处理	172
12.1.7	无穷值处理	172
12.2	导出为.csv 文件	173
12.2.1	设置文件导出路径	173
12.2.2	设置索引	174
12.2.3	设置要导出的列	174
12.2.4	设置分隔符号	174
12.2.5	缺失值处理	174
12.2.6	设置编码格式	175
12.3	将文件导出到多个 Sheet	175
<b>第 13 章 菜品摆放——数据可视化</b>		<b>176</b>
13.1	数据可视化是什么	176
13.2	数据可视化的基本流程	176
13.2.1	整理数据	176
13.2.2	明确目的	177
13.2.3	寻找合适的表现形式	177
13.3	图表的基本组成元素	177
13.4	Excel 与 Python 可视化	179
13.5	建立画布和坐标系	179
13.5.1	建立画布	179
13.5.2	用 add_subplot 函数建立坐标系	180
13.5.3	用 plt.subplot2grid 函数建立坐标系	182
13.5.4	用 plt.subplot 函数建立坐标系	183
13.5.5	用 plt.subplots 函数建立坐标系	184
13.5.6	几种创建坐标系方法的区别	185
13.6	设置坐标轴	185
13.6.1	设置坐标轴的标题	185
13.6.2	设置坐标轴的刻度	187

13.6.3	设置坐标轴的范围	190
13.6.4	坐标轴的轴显示设置	191
13.7	其他图表格式的设置	191
13.7.1	网格线设置	191
13.7.2	设置图例	193
13.7.3	图表标题设置	195
13.7.4	设置数据标签	197
13.7.5	图表注释	198
13.7.6	数据表	199
13.8	绘制常用图表	201
13.8.1	绘制折线图	201
13.8.2	绘制柱形图	204
13.8.3	绘制条形图	208
13.8.4	绘制散点图	209
13.8.5	绘制气泡图	211
13.8.6	绘制面积图	212
13.8.7	绘制树地图	213
13.8.8	绘制雷达图	215
13.8.9	绘制箱形图	217
13.8.10	绘制饼图	218
13.8.11	绘制圆环图	220
13.8.12	绘制热力图	221
13.8.13	绘制水平线和垂直线	223
13.9	绘制组合图表	224
13.9.1	折线图+折线图	224
13.9.2	折线图+柱形图	225
13.10	绘制双坐标轴图表	226
13.10.1	绘制双 y 轴图表	227
13.10.2	绘制双 x 轴图表	228
13.11	绘图样式设置	228

## 进阶篇

第 14 章	典型数据分析案例	234
14.1	利用 Python 实现报表自动化	234
14.1.1	为什么要进行报表自动化	234
14.1.2	什么样的报表适合自动化	234
14.1.3	如何实现报表自动化	235
14.2	自动发送电子邮件	239
14.3	假如你是某连锁超市的数据分析师	241



- 14.3.1 哪些类别的商品比较畅销 ..... 242
- 14.3.2 哪些商品比较畅销 ..... 242
- 14.3.3 不同门店的销售额占比 ..... 243
- 14.3.4 哪些时间段是超市的客流高峰期 ..... 244
- 14.4 假如你是某银行的数据分析师 ..... 245
  - 14.4.1 是不是收入越高的人坏账率越低 ..... 246
  - 14.4.2 年龄和坏账率有什么关系 ..... 247
  - 14.4.3 家庭人口数量和坏账率有什么关系 ..... 248
- 第 15 章 NumPy 数组 ..... 250
  - 15.1 NumPy 简介 ..... 250
  - 15.2 NumPy 数组的生成 ..... 250
    - 15.2.1 生成一般数组 ..... 251
    - 15.2.2 生成特殊类型数组 ..... 251
    - 15.2.3 生成随机数组 ..... 253
  - 15.3 NumPy 数组的基本属性 ..... 255
  - 15.4 NumPy 数组的数据选取 ..... 256
    - 15.4.1 一维数据选取 ..... 256
    - 15.4.2 多维数据选取 ..... 257
  - 15.5 NumPy 数组的数据预处理 ..... 259
    - 15.5.1 NumPy 数组的类型转换 ..... 259
    - 15.5.2 NumPy 数组的缺失值处理 ..... 260
    - 15.5.3 NumPy 数组的重复值处理 ..... 260
  - 15.6 NumPy 数组重塑 ..... 261
    - 15.6.1 一维数组重塑 ..... 261
    - 15.6.2 多维数组重塑 ..... 261
    - 15.6.3 数组转置 ..... 262
  - 15.7 NumPy 数组合并 ..... 262
    - 15.7.1 横向合并 ..... 262
    - 15.7.2 纵向合并 ..... 263
  - 15.8 常用数据分析函数 ..... 264
    - 15.8.1 元素级函数 ..... 264
    - 15.8.2 描述统计函数 ..... 264
    - 15.8.3 条件函数 ..... 266
    - 15.8.4 集合关系 ..... 266