

目录

第 1 章 数据分析概述	1
1.1 数据分析的背景.....	1
1.2 什么是数据分析.....	2
1.3 数据分析的应用场景.....	3
1.4 数据分析的流程.....	4
1.5 为什么选择 Python 做数据分析	5
1.6 创建新的 Python 环境—Anaconda	6
1.6.1 Anaconda 发行版本概述.....	7
1.6.2 在 Windows 系统中安装 Anaconda	7
1.6.3 通过 Anaconda 管理 Python 包	9
1.7 启用 Jupyter Notebook.....	12
1.7.1 启动 Anaconda 自带的 Jupyter Notebook.....	12
1.7.2 Jupyter Notebook 界面详解.....	13
1.7.3 Jupyter Notebook 的基本使用	16
1.8 常见的数据分析工具.....	20
1.9 本章小结.....	21
1.10 本章习题.....	22
第 2 章 科学计算库 NumPy.....	23
2.1 认识 NumPy 数组对象.....	23
2.2 创建 NumPy 数组.....	25
2.3 ndarray 对象的数据类型	27
2.3.1 查看数据类型.....	27
2.3.2 转换数据类型.....	29
2.4 数组运算	30
2.4.1 矢量化运算	30
2.4.2 数组广播	32
2.4.3 数组与标量间的运算.....	33
2.5 ndarray 的索引和切片	34
2.5.1 整数索引和切片的基本使用.....	34
2.5.2 花式（数组）索引的基本使用.....	37
2.5.3 布尔型索引的基本使用.....	39
2.6 数组的转置和轴对称.....	40
2.7 NumPy 通用函数.....	43
2.8 利用 NumPy 数组进行数据处理.....	47
2.8.1 将条件逻辑转为数组运算.....	47
2.8.2 数组统计运算.....	48
2.8.3 数组排序	49
2.8.4 检索数组元素.....	50
2.8.5 唯一化及其他集合逻辑.....	50
2.9 线性代数模块	51
2.10 随机数模块	53

2.11 案例—酒鬼漫步.....	54
2.12 本章小结	56
2.13 本章习题	56
第 3 章 数据分析工具 Pandas.....	58
3.1 Pandas 的数据结构分析	59
3.1.1 Series	59
3.1.2 DataFrame	62
3.2 Pandas 索引操作及高级索引	66
3.2.1 索引对象.....	66
3.2.2 重置索引.....	67
3.2.3 索引操作.....	70
3.3 算术运算与数据对齐.....	76
3.4 数据排序.....	77
3.4.1 按索引排序.....	77
3.4.2 按值排序.....	80
3.5 统计计算与描述.....	81
3.5.1 常用的统计计算.....	81
3.5.2 统计描述.....	83
3.6 层次化索引.....	85
3.6.1 认识层次化索引.....	86
3.6.2 层次化索引的操作.....	92
3.7 读写数据操作	97
3.7.1 读写文本文件.....	98
3.7.2 读写 Excel 文件	101
3.7.3 读取 HTML 表格数据	103
3.7.4 读写数据库	105
3.8 案例—北京高考分数线统计分析.....	110
2.8.1 案例需求	111
2.8.2 数据准备	111
2.8.3 功能实现	112
3.9 本章小结	116
3.10 本章习题	117
第 4 章 数据预处理.....	119
4.1 数据清洗	120
4.1.1 空值和缺失值的处理.....	120
4.1.2 重复值的处理.....	128
4.1.3 异常值的处理.....	131
4.1.4 更改数据类型.....	137
4.2 数据合并	140
4.2.1 轴向堆叠数据.....	141
4.2.2 主键合并数据.....	144
4.2.3 根据行索引合并数据.....	151
4.2.4 合并重叠数据.....	153
4.3 数据重塑	155

4.3.1 重塑层次化索引.....	155
4.3.2 轴向旋转	160
4.4 数据转换	161
4.4.1 重命名轴索引.....	162
4.4.2 离散化连续数据.....	164
4.4.3 哑变量处理类别型数据.....	166
4.5 案例—预处理部分地区信息.....	168
4.5.1 案例需求.....	168
4.5.2 数据准备.....	169
4.5.3 功能实现.....	170
4.6 本章小结	184
4.7 本章习题	184
第 5 章 数据聚合与分组运算.....	188
5.1 分组与聚合的原理.....	188
5.2 通过 <code>groupby()</code> 方法将数据拆分成组.....	189
5.3 数据聚合	198
5.3.1 使用内置统计方法聚合数据.....	198
5.3.2 面向列的聚合方法.....	199
5.4 分组级运算	204
5.4.1 数据转换	204
5.4.2 数据应用	208
5.5 案例—运动员信息的分组与聚合.....	211
5.5.1 案例需求	212
5.5.2 数据准备	212
5.5.3 功能实现	213
5.6 本章小结	223
5.7 本章习题	224
第 6 章 数据可视化.....	226
6.1 数据可视化概述.....	227
6.1.1 什么是数据可视化.....	227
6.1.2 常见的图表类型.....	228
6.1.3 数据可视化的工具.....	233
6.2 Matplotlib—绘制图表.....	234
6.2.1 通过 <code>figure()</code> 函数创建画布.....	234
6.2.2 通过 <code>subplot()</code> 函数创建单个子图	237
6.2.3 通过 <code>subplots()</code> 函数创建多个子图	240
6.2.4 通过 <code>add_subplot()</code> 方法添加和选中子图	241
6.2.5 添加各类标签.....	243
6.2.6 绘制常见图表.....	245
6.2.7 本地保存图形.....	252
6.3 Seaborn—绘制统计图形	253
6.3.1 可视化数据的分布.....	254
6.3.2 用分类数据绘图.....	261
6.4 Bokeh—交互式可视化库.....	268

6.4.1 认识 Bokeh 库	268
6.4.2 通过 Plotting 绘制图形	269
6.5 案例—画图分析某年旅游景点数据	271
6.5.1 案例需求	271
6.5.2 数据准备	272
6.5.3 功能实现	272
6.6 本章小结	278
6.7 本章习题	278
第 7 章 时间序列数据分析	280
7.1 时间序列的基本操作	281
7.1.1 创建时间序列	281
7.1.2 通过时间戳索引选取子集	283
7.2 固定频率的时间序列	286
7.2.1 创建固定频率的时间序列	286
7.2.2 时间序列的频率、偏移量	290
7.2.3 时间序列的移动	293
7.3 时间周期及计算	295
7.3.1 创建时期对象	295
7.3.2 时期的频率转换	297
7.4 重采样	298
7.4.1 重采样方法（resample）	298
7.4.2 降采样	301
7.4.3 升采样	302
7.5 数据统计—滑动窗口	304
7.6 时序模型—ARIMA	308
7.7 案例—股票收盘价分析	310
7.7.1 案例需求	310
7.7.2 数据准备	310
7.7.3 功能实现	311
7.8 本章小结	322
7.9 本章习题	322
第 8 章 文本数据分析	325
8.1 文本数据分析工具	326
8.1.1 NLTK 与 jieba 概述	326
8.1.2 安装 NLTK 和下载语料库	328
8.1.3 jieba 库的安装	330
8.2 文本预处理	331
8.2.1 预处理的流程	331
8.2.2 分词	333
8.2.3 词性标注	335
8.2.4 词形归一化	337
8.2.5 删除停用词	340
8.3 文本情感分析	342
8.4 文本相似度	345

8.5 文本分类	350
8.6 案例—商品评价分析.....	354
8.6.1 案例需求	354
8.6.2 数据准备	355
8.6.3 功能实现	356
8.7 本章小结	361
8.8 本章习题	361
第 9 章 实战—北京租房数据统计分析	363
9.1 数据来源	364
9.2 数据读取	365
9.3 数据预处理	366
9.3.1 重复值和空值处理.....	366
9.3.2 数据转换类型.....	369
9.4 图表分析	371
9.4.1 房源数量、位置分布分析.....	372
9.4.2 户型数量分析.....	384
9.4.3 平均租金分析.....	388
9.4.4 面积区间分析.....	392
9.5 本章小结	395