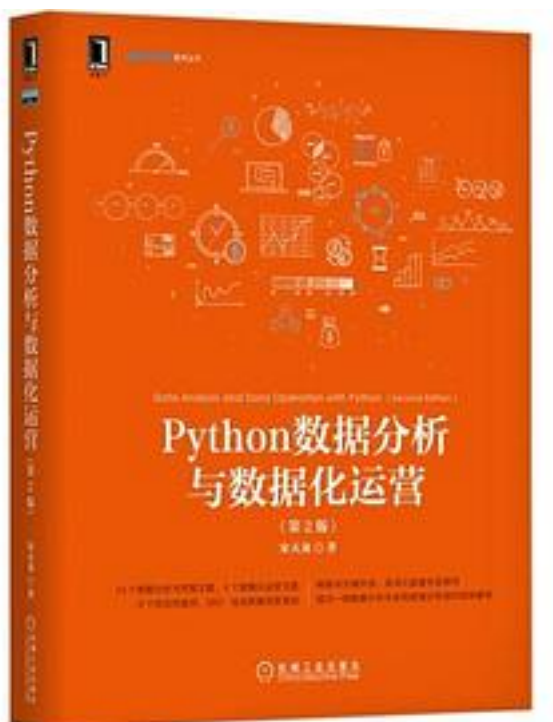


Python数据分析与数据化运营（第2版）



[Python数据分析与数据化运营（第2版）_下载链接1](#)

著者:宋天龙

出版者:机械工业出版社

出版时间:2019-6-1

装帧:平装

isbn:9787111627760

这是一本将数据分析技术与数据使用场景深度结合的著作，从实战角度讲解了如何利用Python进行数据分析和数据化运营。

畅销书全新、大幅升级，第1版近乎100%的好评，第2版不仅将Python升级到了新的版本，而且对具体内容进行了大幅度的补充和优化。作者是有10余年数据分析与数据化运营的资深大数据专家，书中对50余个数据工作流知识点、14个数据分析与挖掘主题、4个数据化运营主题、8个综合性案例进行了全面的讲解，能让数据化运营结合数据使用场景360°落地。

全书一共9章，分为两个部分：

第一部分（第1-4章） Python数据分析与挖掘

首先介绍了Python和数据化运营的基本知识，然后详细讲解了Python数据获取（结构化和非结构化）、预处理、分析和挖掘的关键技术和经验，包含10大类预处理经验、14个数据分析与挖掘主题，50余个知识点。

第二部分（第5~9章） Python数据化运营

这是本书的核心，详细讲解了会员运营、商品运营、流量运营和内容运营4大主题，以及提升数据化运营价值的方法。每个运营主题中都包含了基本知识、评估指标、应用场景、数据分析模型、数据分析小技巧、数据分析大实话以及2个综合性的应用案例。

作者介绍:

宋天龙(TonySong)

大数据技术专家，触脉咨询合伙人兼副总裁，前Webtrekk中国区技术和咨询负责人（Webtrekk，德国的在线数据分析服务提供商）。

擅长数据挖掘、建模、分析与运营，精通端到端数据价值场景设计、业务需求转换、数据结构梳理、数据建模与学习以及数据工程交付。在电子商务、零售、银行、保险等多个行业拥有丰富的数据项目工作经验，参与过集团和企业级数据体系规划、DMP与数据仓库建设、大数据产品开发、网站流量系统建设、个性化智能推荐与精准营销、企业大数据智能等。参与实施客户案例包括联合利华、Webpower、德国OTTO集团电子商务（中国）、Esprit中国、猪八戒网、顺丰优选、乐视商城、泰康人寿、酒仙网、国美在线、迪信通等。

著有多部畅销书：

《Python数据分析与数据化运营》

《网站数据挖掘与分析：系统方法与商业实践》

《企业大数据系统构建实战：技术、架构、实施与应用》

目录: 前言

第1章 Python和数据化运营1

1.1 用Python做数据化运营1

1.1.1 Python是什么1

1.1.2 数据化运营是什么2

1.1.3 Python用于数据化运营5

1.2 数据化运营所需的Python相关工具和组件5

1.2.1 Python程序6

1.2.2 Python交互环境Jupyter7

1.2.3 Python第三方库23

1.2.4 数据库和客户端29

1.2.5 SSH远程客户端30

1.3 内容延伸：Python的OCR和tensorflow31

1.3.1 OCR工具：Tesseract-ocr31

1.3.2 机器学习框架：TensorFlow31

1.4 第1个用Python实现的数据化运营分析实例、销售预测	32
1.4.1 案例概述	32
1.4.2 案例过程	32
1.4.3 案例小结	36
1.5 本章小结	37
第2章 数据化运营的数据来源	40
2.1 数据化运营的数据来源类型	40
2.1.1 数据文件	40
2.1.2 数据库	41
2.1.3 API	42
2.1.4 流式数据	43
2.1.5 外部公开数据	43
2.1.6 其他来源	44
2.2 使用Python获取运营数据	44
2.2.1 从文本文件读取运营数据	44
2.2.2 从Excel获取运营数据	55
2.2.3 从关系型数据库MySQL读取运营数据	57
2.2.4 从非关系型数据库MongoDB读取运营数据	64
2.2.5 从API获取运营数据	68
2.3 内容延伸：读取非结构化网页、文本、图像、视频、语音	72
2.3.1 从网页中获取运营数据	72
2.3.2 读取非结构化文本数据	73
2.3.3 读取图像数据	74
2.3.4 读取视频数据	78
2.3.5 读取语音数据	81
2.4 本章小结	85
第3章 10条数据化运营不得不知道的数据预处理经验	87
3.1 数据清洗：缺失值、异常值和重复值的处理	87
3.1.1 数据列缺失的4种处理方法	87
3.1.2 不要轻易抛弃异常数据	89
3.1.3 数据重复就需要去重吗	90
3.1.4 代码实操：Python数据清洗	92
3.2 将分类数据和顺序数据转换为标志变量	100
3.2.1 分类数据和顺序数据是什么	100
3.2.2 运用标志方法处理分类和顺序变量	101
3.2.3 代码实操：Python标志转换	101
3.3 大数据时代的数据降维	104
3.3.1 需要数据降维的情况	104
3.3.2 基于特征选择的降维	105
3.3.3 基于特征转换的降维	106
3.3.4 基于特征组合的降维	112
3.3.5 代码实操：Python数据降维	114
3.4 解决样本类别分布不均衡的问题	123
3.4.1 哪些运营场景中容易出现样本不均衡	124
3.4.2 通过过抽样和欠抽样解决样本不均衡	124
3.4.3 通过正负样本的惩罚权重解决样本不均衡	124
3.4.4 通过组合/集成方法解决样本不均衡	125
3.4.5 通过特征选择解决样本不均衡	125
3.4.6 代码实操：Python处理样本不均衡	125
3.5 数据化运营要抽样还是全量数据	128
3.5.1 什么时候需要抽样	128
3.5.2 如何进行抽样	129
3.5.3 抽样需要注意的几个问题	130
3.5.4 代码实操：Python数据抽样	131

3.6 解决运营数据的共线性问题	135
3.6.1 如何检验共线性	135
3.6.2 解决共线性的5种常用方法	136
3.6.3 代码实操：Python处理共线性问题	137
3.7 有关相关性分析的混沌	139
3.7.1 相关和因果是一回事吗	139
3.7.2 相关系数低就是不相关吗	139
3.7.3 代码实操：Python相关性分析	140
3.8 标准化，让运营数据落入相同的范围	141
3.8.1 实现中心化和正态分布的Z-Score	141
3.8.2 实现归一化的Max-Min	142
3.8.3 用于稀疏数据的MaxAbs	142
3.8.4 针对离群点的RobustScaler	142
3.8.5 代码实操：Python数据标准化处理	142
3.9 离散化，对运营数据做逻辑分层	145
3.9.1 针对时间数据的离散化	145
3.9.2 针对多值离散数据的离散化	146
3.9.3 针对连续数据的离散化	146
3.9.4 针对连续数据的二值化	147
3.9.5 代码实操：Python数据离散化处理	147
3.10 内容延伸：非结构化数据的预处理	151
3.10.1 网页数据解析	151
3.10.2 网络用户日志解析	159
3.10.3 图像的基本预处理	164
3.10.4 自然语言文本预处理	169
3.11 本章小结	172
第4章 跳过运营数据分析和挖掘的“大坑”	174
4.1 聚类分析	174
4.1.1 当心数据异常对聚类结果的影响	175
4.1.2 超大数据量时应该放弃K均值算法	175
4.1.3 聚类不仅是建模的终点，更是重要的中间预处理过程	177
4.1.4 高维数据上无法应用聚类吗	178
4.1.5 如何选择聚类分析算法	179
4.1.6 案例：客户特征的聚类与探索性分析	179
4.2 回归分析	196
4.2.1 注意回归自变量之间的共线性问题	197
4.2.2 相关系数、判定系数和回归系数之间是什么关系	197
4.2.3 判定系数是否意味着相应的因果联系	197
4.2.4 注意应用回归模型时研究自变量是否产生变化	198
4.2.5 如何选择回归分析算法	198
4.2.6 案例：大型促销活动前的销售预测	199
4.3 分类分析	206
4.3.1 防止分类模型的过拟合问题	207
4.3.2 使用关联算法做分类分析	207
4.3.3 用分类分析来提炼规则、提取变量、处理缺失值	208
4.3.4 类别划分：分类算法和聚类算法都是好手	209
4.3.5 如何选择分类分析算法	210
4.3.6 案例：用户流失预测分析与应用	210
4.4 关联分析	221
4.4.1 频繁规则不一定是有效规则	221
4.4.2 不要被啤酒尿布的故事禁锢你的思维	222
4.4.3 被忽略的“负相关”模式真的毫无用武之地吗	223
4.4.4 频繁规则只能打包组合应用吗	224
4.4.5 关联规则的序列模式	224

4.4.6 使用Python和R分别实现商品交叉销售分析	225
4.5 异常检测分析	240
4.5.1 异常检测中的“新奇检测”模式	241
4.5.2 将数据异常与业务异常相分离	241
4.5.3 面临维度灾难时，异常检测可能会失效	241
4.5.4 异常检测的结果能说明异常吗	241
4.5.5 案例：网站广告流量的异常识别与检测	242
4.6 时间序列分析	253
4.6.1 如果有自变量，为什么还要用时间序列	253
4.6.2 时间序列不适合商业环境复杂的企业	254
4.6.3 时间序列预测的整合、横向和纵向模式	255
4.6.4 案例：未来10天的整体销售预测	255
4.7 路径、漏斗、归因和热力图分析	267
4.7.1 不要轻易相信用户的页面访问路径	269
4.7.2 如何将路径应用于更多用户行为模式的挖掘	270
4.7.3 为什么很多数据都显示了多渠道路径的价值很小	271
4.7.4 点击热力图真的反映了用户的点击喜好吗	271
4.7.5 为什么归因分析主要存在于线上的转化行为	272
4.7.6 漏斗分析和路径分析有什么区别	273
4.8 其他数据分析和挖掘的忠告	274
4.8.1 不要忘记数据质量的验证	274
4.8.2 不要忽视数据的落地性	274
4.8.3 不要把数据陈列当作数据结论	275
4.8.4 数据结论不要产生于单一指标	276
4.8.5 数据分析不要预设价值立场	276
4.8.6 不要忽视数据与业务的需求冲突问题	277
4.9 内容延伸1：非结构化数据的分析与挖掘	278
4.9.1 词频统计	278
4.9.2 词性标注	281
4.9.3 关键字提取	284
4.9.4 文本聚类	286
4.10 内容延伸2：AutoML—自动化数据挖掘与机器学习	289
4.10.1 自动化数据挖掘与机器学习概述	289
4.10.2 TPOT自动化学习简介	290
4.10.3 使用TPOT做自动化学习	292
4.11 本章小结	298
第5章 会员数据化运营	300
5.1 会员数据化运营概述	300
5.2 会员数据化运营关键指标	301
5.2.1 会员整体指标	301
5.2.2 会员营销指标	301
5.2.3 会员活跃度指标	303
5.2.4 会员价值度指标	305
5.2.5 会员终生价值指标	305
5.2.6 会员异动指标	306
5.3 会员数据化运营应用场景	307
5.3.1 会员营销	307
5.3.2 会员关怀	307
5.4 会员数据化运营分析模型	307
5.4.1 会员细分模型	307
5.4.2 会员活跃度模型	309
5.4.3 会员价值度模型	309
5.4.4 会员流失预测模型	310
5.4.5 会员特征分析模型	311

5.4.6 营销响应预测模型	312
5.5 会员数据化运营分析小技巧	313
5.5.1 使用留存分析做新用户质量分析	313
5.5.2 使用AARRR做App用户生命周期分析	314
5.5.3 借助动态数据流关注会员状态的轮转	315
5.5.4 使用协同过滤算法为新会员分析推送个性化信息	316
5.6 会员数据化运营分析的“大实话”	318
5.6.1 企业“不差钱”，还有必要做会员精准营销吗	318
5.6.2 用户满意度取决于期望和给予的匹配程度	319
5.6.3 用户不购买就是流失了吗	319
5.6.4 来自调研问卷的用户信息可信吗	320
5.6.5 不要盲目相信二八法则	321
5.7 案例：基于RFM的精细化用户管理	322
5.7.1 案例背景	322
5.7.2 案例主要应用技术	322
5.7.3 案例数据	323
5.7.4 案例过程	323
5.7.5 案例数据结论	337
5.7.6 案例应用和部署	341
5.7.7 案例注意点	341
5.7.8 案例引申思考	342
5.8 案例：基于嵌套Pipeline和FeatureUnion复合数据工作流的营销响应预测	342
5.8.1 案例背景	342
5.8.2 案例主要应用技术	343
5.8.3 案例数据	343
5.8.4 案例过程	344
5.8.5 案例数据结论	353
5.8.6 案例应用和部署	353
5.8.7 案例注意点	354
5.8.8 案例引申思考	355
5.9 本章小结	355
第6章 商品数据化运营	357
6.1 商品数据化运营概述	357
6.2 商品数据化运营关键指标	357
6.2.1 销售指标	357
6.2.2 促销活动指标	360
6.2.3 供应链指标	361
6.3 商品数据化运营应用场景	363
6.3.1 销售预测	363
6.3.2 库存分析	363
6.3.3 市场分析	363
6.3.4 促销分析	364
6.4 商品数据化运营分析模型	364
6.4.1 商品价格敏感度模型	364
6.4.2 新产品市场定位模型	365
6.4.3 销售预测模型	366
6.4.4 商品关联销售模型	367
6.4.5 异常订单检测	368
6.4.6 商品规划的最优组合	368
6.5 商品数据化运营分析小技巧	370
6.5.1 使用层次分析法将定量与定性分析结合	370
6.5.2 通过假设检验做促销拉动分析	373
6.5.3 使用BCG矩阵做商品结构分析	374
6.5.4 巧用4P分析建立完善的商品运营分析结构	376

6.6 商品数据化运营分析的“大实话”	376
6.6.1 为什么很多企业会以低于进价的价格大量销售商品	376
6.6.2 促销活动真的是在促进商品销售吗	378
6.6.3 用户关注的商品就是要买的商品吗	379
6.6.4 提供的选择过多其实不利于商品销售	380
6.7 案例：基于Gradient Boosting的自动超参数优化的销售预测	381
6.7.1 案例背景	381
6.7.2 案例主要应用技术	381
6.7.3 案例数据	381
6.7.4 案例过程	382
6.7.5 案例数据结论	389
6.7.6 案例应用和部署	389
6.7.7 案例注意点	389
6.7.8 案例引申思考	389
6.8 案例：基于集成算法GBDT和RandomForest的投票组合模型的异常检测	390
6.8.1 案例背景	390
6.8.2 案例主要应用技术	390
6.8.3 案例数据	391
6.8.4 案例过程	391
6.8.5 案例数据结论	397
6.8.6 案例应用和部署	398
6.8.7 案例注意点	398
6.8.8 案例引申思考	399
6.9 本章小结	400
第7章 流量数据化运营	402
7.1 流量数据化运营概述	402
7.2 八大流量分析工具	402
7.2.1 Adobe Analytics	403
7.2.2 Webtrekk Suite	403
7.2.3 Webtrends	403
7.2.4 Google Analytics	404
7.2.5 IBM Coremetrics	404
7.2.6 百度统计	404
7.2.7 Flurry	404
7.2.8 友盟	405
7.2.9 如何选择第三方流量分析工具	405
7.3 流量采集分析系统的工作机制	408
7.3.1 流量数据采集	409
7.3.2 流量数据处理	411
7.3.3 流量数据应用	412
7.4 流量数据与企业数据的整合	413
7.4.1 流量数据整合的意义	413
7.4.2 流量数据整合的范畴	414
7.4.3 流量数据整合的方法	414
7.5 流量数据化运营指标	415
7.5.1 站外营销推广指标	415
7.5.2 网站流量数量指标	417
7.5.3 网站流量质量指标	418
7.6 流量数据化运营应用场景	421
7.6.1 流量采购	421
7.6.2 流量分发	423
7.7 流量数据化运营分析模型	423
7.7.1 流量波动检测	423
7.7.2 渠道特征聚类	425

7.7.3 广告整合传播模型	425
7.7.4 流量预测模型	428
7.8 流量数据化运营分析小技巧	429
7.8.1 给老板提供一页纸的流量dashboard	429
7.8.2 关注趋势、重要事件和潜在因素是日常报告的核心	431
7.8.3 使用从细分到多层下钻数据分析	431
7.8.4 通过跨屏追踪解决用户跨设备和浏览器的访问行为	432
7.8.5 基于时间序列的用户群体过滤	434
7.9 流量数据化运营分析的“大实话”	436
7.9.1 流量数据分析的价值其实没那么大	436
7.9.2 如何将流量的实时分析价值最大化	437
7.9.3 营销流量的质量评估是难点工作	438
7.9.4 个性化的媒体投放仍然面临很多问题	438
7.9.5 传统的网站分析方法到底缺少了什么	439
7.10 案例：基于自动节点树的数据异常原因下探分析	444
7.10.1 案例背景	444
7.10.2 案例主要应用技术	444
7.10.3 案例数据	445
7.10.4 案例过程	446
7.10.5 案例数据结论	459
7.10.6 案例应用和部署	460
7.10.7 案例注意点	460
7.10.8 案例引申思考	461
7.11 案例：基于自动K值的KMeans广告效果聚类分析	461
7.11.1 案例背景	461
7.11.2 案例主要应用技术	462
7.11.3 案例数据	462
7.11.4 案例过程	463
7.11.5 案例数据结论	474
7.11.6 案例应用和部署	475
7.11.7 案例注意点	476
7.11.8 案例引申思考	477
7.12 本章小结	477
第8章 内容数据化运营	480
8.1 内容数据化运营概述	480
8.2 内容数据化运营指标	481
8.2.1 内容质量指标	481
8.2.2 SEO类指标	481
8.2.3 内容流量指标	482
8.2.4 内容互动指标	482
8.2.5 目标转化指标	483
8.3 内容数据化运营应用场景	484
8.3.1 内容采集	484
8.3.2 内容创作	484
8.3.3 内容分发	485
8.3.4 内容管理	485
8.4 内容数据化运营分析模型	485
8.4.1 情感分析模型	485
8.4.2 搜索优化模型	486
8.4.3 文章关键字模型	486
8.4.4 主题模型	487
8.4.5 垃圾信息检测模型	488
8.5 内容数据化运营分析小技巧	488
8.5.1 通过AB测试和多变量测试找到最佳内容版本	488

8.5.2 通过屏幕浏览占比了解用户到底看了页面多少内容	490
8.5.3 通过数据分析系统与CMS打通实现个性化内容运营	491
8.5.4 将个性化推荐从网站应用到App端	494
8.6 内容数据化运营分析的“大实话”	495
8.6.1 个性化内容运营不仅是整合CMS和数据系统	495
8.6.2 用户在着陆页上不只有跳出和继续两种状态	495
8.6.3 “人工组合”的内容运营价值最大化并非不能实现	496
8.6.4 影响内容点击率的因素不仅有位置	497
8.7 案例：基于潜在狄利克雷分配的内容主题挖掘	498
8.7.1 案例背景	498
8.7.2 案例主要应用技术	498
8.7.3 案例数据	498
8.7.4 案例过程	498
8.7.5 案例数据结论	506
8.7.6 案例应用和部署	507
8.7.7 案例注意点	507
8.7.8 案例引申思考	508
8.8 案例：基于多项式贝叶斯的增量学习的文本分类	508
8.8.1 案例背景	508
8.8.2 案例主要应用技术	508
8.8.3 案例数据	509
8.8.4 案例过程	509
8.8.5 案例数据结论	517
8.8.6 案例应用和部署	517
8.8.7 案例注意点	517
8.8.8 案例引申思考	517
8.9 本章小结	518
第9章 数据化运营分析的终极秘籍	520
9.1 撰写出彩的数据分析报告的5个建议	520
9.1.1 完整的报告结构	520
9.1.2 精致的页面版式	521
9.1.3 漂亮的可视化图形	523
9.1.4 突出报告的关键信息	524
9.1.5 用报告对象习惯的方式撰写报告	525
9.2 数据化运营支持的4种扩展方式	526
9.2.1 数据API	526
9.2.2 数据模型	526
9.2.3 数据产品	527
9.2.4 运营产品	527
9.3 提升数据化运营价值度的5种途径	528
9.3.1 数据源：不只有结构化的数据	528
9.3.2 自动化：建立自动工作机制	529
9.3.3 未卜先知：建立智能预警模型	529
9.3.4 智能化：向BI-AI的方向走	530
9.3.5 场景化：将数据嵌入运营环节之中	531
9.4 本章小结	531
附录	533
附录A 公开数据集	533
附录B Python数据工具箱	540
• • • • •	(收起)

标签

数据分析

Python

数据运营

编程

实践者解答

大数据

计算机

当当

评论

挺多实用案例，适合有算法基础和python基础的同学

跨界的书不好写。这本书找到一个还算好的平衡点：在具体运营场景下，怎么使用Python来辅助数据分析，进而解决问题。

之前在图书馆读过这本书的第一版，当时就推荐给周围的朋友们了，这并不是一本纯讲技术的书，书中在讨论技术原理应用的同时加入了很多运营过程中总结的运营经验，当时在做电商的项目，用书中给的方法少走了很多弯路。读和用这本书需要有一定的编程基础，当然，如果是研究算法的大牛，建议去读一些纯技术类的书，这本书更适合入门不久的数据分析师或者其他运营数据分析人员。总的来说，书是很不错的，并且书中的源代码实用性比较强。

[Python数据分析与数据化运营（第2版）_下载链接1](#)

书评

这是作者本人的点评，在正式内容之前的一点感想，供大家参考。

《Python数据分析与数据化运营》是我的第三本书，前两本分别是《网站数据挖掘与分析》、《企业大数据系统构建实战》（合著）。

坦白讲，写这本书蛮有压力，原因有三： 1. 在之前的两本书基础上，很多读者已经...

一开始很期待这种业务经验+代码实操的书。

不过书本这么厚，目前看了前面4章基础内容，200页，感觉干货比较少。

面对对象不明确，书中很多经验说了都是点到即止，真正有点经验的大家都懂。

而如果没有经验的人，这本书又不适合入门。

很多概念性的问题，其他原理书都有讲了，放在...

[Python数据分析与数据化运营（第2版）_下载链接1](#)