



Data Analysis and Data Operation with Python (Second Edition)

Python数据分析 与数据化运营

(第2版)

宋天龙◎著

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，360° 结合数据场景落地

畅销书大幅升级，资深大数据专家撰写
国内一线数据分析专家和数据分析组织联袂推荐



Data Analysis and Data Operation with Python (Second Edition)

Python数据分析 与数据化运营

(第2版)

宋天龙◎著



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

Python 数据分析与数据化运营 / 宋天龙著. —2 版. —北京: 机械工业出版社, 2019.5
(数据分析与决策技术丛书)

ISBN 978-7-111-62776-0

I. P… II. 宋… III. ①软件工具 - 程序设计 ②企业管理 - 数据管理 IV. ① TP311.561
② F272.7

中国版本图书馆 CIP 数据核字 (2019) 第 096862 号

Python 数据分析与数据化运营 第 2 版

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 李 艺

责任校对: 殷 虹

印 刷: 中国电影出版社印刷厂

版 次: 2019 年 6 月第 2 版第 1 次印刷

开 本: 186mm×240mm 1/16

印 张: 35.5

书 号: ISBN 978-7-111-62776-0

定 价: 129.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88379426 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294

读者信箱: hzit@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

HZBOOKS | 华章IT | Information Technology



为什么要写这本书

随着商业竞争形势的日益严峻，企业需要不断寻找提高利润率、降低成本、提高产出价值的有效方法，而数据化运营恰好是满足企业这一需求的关键武器。数据化运营包含了运营和数据两种要素，前者需要较多的业务经验，而后者对数据分析提出了更高的要求。只有把二者结合起来，在技能、经验和技术的帮助下，数据化运营才能在企业内部真正落地、生根、发芽。

对数据化运营而言，各企业普遍关注的结构化数据分析、挖掘的场景非常丰富，例如销售预测、会员生命周期维护、商品结构分析等，这些普遍的共同认知为本书提供了接地气的基礎。但除了这些“传统内容”外，还有很多非结构化的数据主题，它们在数据化运营过程中的作用越来越重要，例如主题挖掘、图片分析、文本挖掘、图像识别、语音识别等，这些内容拓展了数据化运营发挥价值的场景基础。

Python 作为数据工作领域的关键武器之一，具有开源、多场景应用、快速上手、完善的生态和服务体系等特征，这些特征使其在数据分析与数据化运营的任何场景中都能游刃有余；即使在其为数不多的短板功能上，Python 仍然可以基于“胶水”的特征，引入对应的第三方工具、库、程序等来实现全场景、全应用的覆盖。在海量数据背景下，Python 对超大数据规模的支持性能、数据分析处理能力、建模的专业程度及开发便捷性的综合能力方面要远远高于其他工具。因此，Python 几乎是数据化运营工作的不二之选。

纵观整个国内技术书籍市场，有关 Python 的书籍不少，但普遍都是基于工具层面的介绍，且侧重于工具本身的方法、参数、调用、实例，与真正实践结合的较少；而有关数据化运营的书籍，目前市场上还为数不多，现有的数据化运营方面的书籍大多是基于 Excel 等工具的入门级别的分析类书籍。因此，本书结合 Python 和数据化运营两个方面，在数据分析工作流程和数据化运营主题结合的基础上，通过指标、模型、方法、案例配合工具的形式，详细介绍了如何使用 Python 来支持数据化运营，尤其是传统工具无法满足的应用场景。

本书第 1 版上市后，得到来自各行各业很多好友和读者的支持和反馈，在此致以感谢！第 2 版在第 1 版基础上做了很多优化甚至重写，还新增了很多内容。

优化以及重写的内容

- ❑ 全部代码基于 Python 3 做优化或重写，书中的 Python 版本是 Python 3.7。
- ❑ 基于 Jupyter 做调试、分析和应用，更适合数据分析师的应用场景，包括探索性分析、数据预处理、结果可视化展示、交互式演示等。
- ❑ 网页数据解析中基于 Class 做功能封装和处理，便于以网页对象为主体的数据工作。
- ❑ PIL/Pillow 的替换和方法应用介绍，多用于图像处理工作。
- ❑ 第 4 章数据挖掘的案例部分，每一个案例都经过重写，并增加了很多知识点，以实际案例为需求，实际应用数据挖掘算法做建模和分析应用。
- ❑ Matplotlib 调用 3D 图形展示多个维度的信息，并可通过拖曳展示不同角度下数据的分布情况。
- ❑ 第 5 章第 1 个案例是 RFM 代码的重构，以及针对不同分组的精细化运营策略的制定。
- ❑ 第 5 章第 2 个案例是基于嵌套 Pipeline 和 FeatureUnion 复合数据工作流的营销响应预测，并基于复合（两层管道）的 Pipeline 做数据 workflow 管理。
- ❑ 第 6 章第 2 个案例是基于集成算法 GBDT 和 RandomForest 的投票组合模型的异常检测，GBDT 和 RandomForest 是两个典型的代表模型“准确度”和“稳定度”的算法，这种兼顾“准”和“稳”的模型搭配更符合实际需求。
- ❑ 第 7 章基于自动节点树的数据异常原因来下探分析树形图的内容和可视化，优化了代码和样式，可视化效果更好，并能获得更多信息，包括维度分解过程、主因子、其他因子和潜在因子等。

新增的内容

- ❑ 基于 Anaconda 的 Python 环境的安装和配置，更方便初学者快速搭建 Python 应用环境。
- ❑ Jupyter 基础工具的用法，包括安装、启动、基础操作、魔术命令、新内核安装和使用、执行 shell 命令、扩展和插件使用、系统基础配置等。
- ❑ 基于 Pandas 的 get_dummies 做标志转换，即 OneHotEncode 转换。
- ❑ 特征选择的降维中新增 feature_selection 配合 SelectPercentile、VarianceThreshold、RFE、SelectFromModel 做特征选择。
- ❑ 特征转换的降维中新增 PCA、LDA、FA、ICA 这几个数据转换和降维的具体方法。
- ❑ 特征组合的降维中新增基于 GBDT、PolynomialFeatures、gplearn 的 genetic 方法做组合特征。
- ❑ 第 4 章的分类算法中新增使用 XGboost 做分类应用，以及配合 graphviz 输出矢量图形。
- ❑ pyecharts 的数据可视化的应用和操作，尤其是关联关系图的应用。
- ❑ Python 通过 rpy2 调用 R 程序，实现关联算法的挖掘，包括直接执行程序文件、代码段、变量使用等。
- ❑ Python 通过 rpy2 调用 auto.arima 实现自动 ARIMA 的应用，降低 Python 在时间序列算法应用时的门槛。
- ❑ 自动化学习：增加了对于自动化数据挖掘与机器学习的理论、流程、知识和应用库介绍，

并基于 TPOT 做自动化回归和分类学习案例演示。

我希望能尽自己的微薄之力，将过往所学、所感、所知提炼出来供更多人了解。如果读者能从本书中感悟一二，我将倍感欣慰；如果读者能将其用于工作实践，这将是本书以及数据工作之福！

读者对象

本书旨在提供有关数据与运营的综合知识的介绍和应用，虽然基础工具是 Python，但本书并没有就 Python 基础规则和语法做详细介绍，因此希望读者具有一定的 Python 基础。相信我，只要你认真看过 Python 教学视频（网络上有很多），只需大概 2 个小时就能具备这种基础。

本书对读者的知识背景没有特定要求，书中的内容都尽量言简意赅、深入浅出。本书适合以下几类读者阅读：

- ❑ 企业运营人员。本书的核心命题就是运营，其中涉及会员运营、商品运营、流量运营和内容运营四大主题，无论运营人员希望获得运营知识，还是希望获得数据分析和挖掘方法，都可以从书中找到。
- ❑ 数据分析师。毫无疑问，数据分析师是本书的核心受众群体之一，书中介绍的数据抽取、预处理和分析挖掘经验一定能为数据分析师带来很多“不一样”的收获，每个运营主题下的小技巧、模型和案例更能激发数据分析师的灵感——原来数据工作还能这样做。
- ❑ Python 工程师。坦白讲，本书不是一本专门介绍 Python 语法、规则的书籍。但 Python 作为一种“万能”工具，在数据分析和挖掘领域具有举足轻重的地位，对于任何一位工作领域中涉及数据（或大数据）的 Python 工程师来说，本书的价值会成倍增长。书中对 Python 数据处理、计算和挖掘库的应用介绍，以及对有关工具库的用法、注意点和小知识介绍一定会使 Python 工程师的工作和能力更上一层楼。
- ❑ 数据挖掘工程师。数据分析与挖掘在实际运营中都是不分家的，本书没有冠以“挖掘”之名，但并不意味着没有挖掘（或机器学习）算法。本书第 4 章基本都是围绕常用算法展开的知识介绍，其中各个算法类的“大坑”都是笔者多年经验的总结；在运营主题的案例中几乎都用到了算法来解决实际问题。除了基本的数据挖掘以外，本书还涉及自然语言处理、日志处理、文本挖掘、图像处理等相关知识。

如何阅读本书

本书内容从逻辑上共分为两大部分，第一部分是有关数据分析类的主题，第二部分是有关数据化运营的主题。

第一部分包括第 1~4 章和附录，主要介绍了 Python 和数据化运营的基本知识、数据来源获取、数据预处理，以及数据分析和挖掘的关键经验。其中：

- ❑ 第 1 章是 Python 和数据化运营的基本知识介绍。
- ❑ 第 2 章对传统的结构化和非结构化数据来源及获取、读取和解析进行了介绍，包括数据

文件、数据库、API、流式数据、外部公开数据等，也提到了如何读取网页、文本、图片、视频、语音等类型的数据。

- ❑ 第3章总结了10个有关数据化运营过程的数据预处理经验，涵盖了常见的数据清洗、标志转换、数据降维、样本不均衡、数据源冲突、抽样、共线性、相关性分析、数据标准化、数据离散化等内容，并在扩展内容中简单介绍了有关网页、日志、图像、自然语言的文本预处理工作，作为拓展知识阅读。
- ❑ 第4章总结了数据分析、挖掘和网站分析方法的8个主题类，内容涵盖聚类、回归、分类、关联、异常检测、时间序列、路径分析、漏斗分析、归因分析、热力图分析及其他统计分析相关话题，各个类别中都以关键经验为基础，围绕应用误区、注意事项和应用方法展开详细介绍。

第二部分包括第5~9章，分别介绍了会员运营、商品运营、流量运营和内容运营四大主题，以及提升数据化运营价值的方法。在每个数据化运营主题中都包含了基本知识、关键指标、应用场景、数据分析模型、数据分析小技巧、数据分析“大实话”及两个应用案例。

- ❑ 基本知识：有关运营主题的基本内涵、价值、用途等方面的介绍。
- ❑ 关键指标：运营主题的主要评估和矩阵指标，按类别拆分和归纳。
- ❑ 应用场景：总结数据对于运营的价值落地在哪些场景中。
- ❑ 数据分析模型：“大型”的数据分析方法，包括统计分析、数据挖掘、网站分析、数学模型。
- ❑ 数据分析小技巧：“小型”的数据分析方法，看起来相对简单但非常有效。
- ❑ 数据分析“大实话”：有关运营或数据分析的潜在规律的解释及介绍。
- ❑ 应用案例：每个运营主题都包含两个应用案例，基本上每个案例的应用算法和技巧都不相同，目的是呈现不同算法在不同场景下的差异化应用。

除了以上内容外，以下信息是对本书中涉及特定内容的解释和说明。

渐进式的内容：本书的Python代码和实现部分，在不同章节可能会具有不同的代码风格，包括定义规则、注释、功能实现等，这是因为笔者试图遵循循序渐进的原则，先介绍功能实现，然后再介绍其他的备选方案，以及规范、原则等来辅助Python的实现。这样做一方面是希望尽量多地展示解决同一类问题的不同方法，让读者能根据自身实际情况选择最“合适”的用法示例；另一方面是因为，考虑到可能有很多读者不具备较强的Python基础知识，笔者不希望一上来就让这些读者感觉到使用Python会受到各种“条条框框”的限制，从而打击他们使用Python的信心，毕竟，能实现功能需求是第一位的。

内容延伸：本书第1~4章都有内容延伸章节，其内容是有关非结构化主题的读取、分析、处理。由于每个主题展开来都能写成一本书，因此仅在内容延伸中抛砖引玉，有兴趣的读者可以了解和学习。

相关知识点：本书很多章节中都有“相关知识点”部分，其内容是关于特定工具、知识、算法、库等方面的较为详细的介绍，充当了本书的知识堡垒。

本章小结：每章的结尾都有“本章小结”，在小结中包含 4 部分内容。

- **内容小结：**有关本章内容的总结。
- **重点知识：**本章需要读者重点掌握的知识和内容。
- **外部参考：**本章提到但是无法详细介绍的内容，有兴趣的读者可以基于外部参考构建自己的知识图谱。
- **应用实践：**基于本章内容给出在实践中落地的建议。

提示：对于知识点的重要提示和应用技巧，相对“相关知识点”而言，每条提示信息内容量较少，一般都是经验类的总结。

注意：特定知识需要引起注意的点，这些注意点是应用过程中需要避免的“大坑”。

特定名词的混用：本书提到了库和包、模型和算法等词，虽然有含义差异，但书中并没有划清它们的界限，因此在很多时候它们都是等价的。

关于附录的使用方法：除了第 9 章外（该章节没有代码），本书的每一章都有对应源数据和完整代码，该内容可在本书附件中找到，附件可以在华章网站（<http://www.hzbook.com>）或者笔者网站——数据常青藤（http://www.dataivy.cn/book/python_book_v2.zip）下载。需要注意的是，为了让读者更好地了解每行代码的含义，笔者在注释信息中都使用了中文标注，每个程序文件的编码格式都是 UTF-8。

勘误和支持

由于笔者水平有限，加之撰写时间也有限，书中难免会出现一些错误或者不准确的地方，恳请读者批评指正。读者可通过以下途径联系并反馈建议或意见：

- **微信沟通：**本书已建立讨论群，读者可先添加笔者个人微信（TonySong2013），后续笔者会将读者加入本书的讨论群中。
- **网站讨论区：**在笔者网站——数据常青藤的书籍讨论区（<http://www.dataivy.cn/python-data-analysis-and-data-operations/>）留言。
- **电子邮箱：**发送 E-mail 到 517699029@qq.com。

致谢

在本书撰写过程中，得到了来自多方的指导、帮助和支持。

首先，感谢彭亮先生和史研先生。彭亮先生使我感受到什么是高度和专注，并促使我的数据工作从真正意义上步入正途。史研先生对于大数据的广博认知和敏锐洞察力，让我有更多机会深入到不同的数据分支，去探索未曾了解的领域。

其次，感谢机械工业出版社华章分社第二事业部的杨福川老师，杨老师在我出版了两本书之后鼓励我继续撰写本书，并为本书的撰写提供了方向和思路指导。另外，感谢全程参与审核、校验等工作的李艺老师及其他背后默默支持的出版工作者，他们的辛勤付出保证了本书的顺利面世。

再次，感谢在各个数据项目 and 工作中提供支持的领导、朋友、伙伴，尤其是田学锋，他是我的良师益友，他有着非比寻常的视野、胸怀和独到的见解，在我的人生道路上给予了我非常多的指导和启迪。还要感谢跟我一起工作的小伙伴（排名不分先后）：庞程程、徐子东、赵光娟、王成、吕兆星、郑传峰、杨晓鹏、陈骏、江涛、曹佳佳、麻建昕、史晓春、杨勇等。

最后，感谢我的父母、家人和朋友，尤其是我的夫人姜丽女士，是她在我写书的这段期间把家里的一切料理得井井有条，使得我有精力完成本书的全部撰写工作。

谨以此书献给热爱数据工作并为之奋斗的朋友们，愿大家身体健康、生活美满、事业有成！

宋天龙 (Tony Song)

Contents 目 录

前言	1.4.3 案例小结	36
	1.5 本章小结	37
第 1 章 Python 和数据化运营	第 2 章 数据化运营的数据来源	40
1.1 用 Python 做数据化运营	2.1 数据化运营的数据来源类型	40
1.1.1 Python 是什么	2.1.1 数据文件	40
1.1.2 数据化运营是什么	2.1.2 数据库	41
1.1.3 Python 用于数据化运营	2.1.3 API	42
1.2 数据化运营所需的 Python	2.1.4 流式数据	43
相关工具和组件	2.1.5 外部公开数据	43
1.2.1 Python 程序	2.1.6 其他来源	44
1.2.2 Python 交互环境 Jupyter	2.2 使用 Python 获取运营数据	44
1.2.3 Python 第三方库	2.2.1 从文本文件读取运营数据	44
1.2.4 数据库和客户端	2.2.2 从 Excel 获取运营数据	55
1.2.5 SSH 远程客户端	2.2.3 从关系型数据库 MySQL	
1.3 内容延伸: Python 的 OCR	读取运营数据	57
和 tensorflow	2.2.4 从非关系型数据库 MongoDB	
1.3.1 OCR 工具: Tesseract-ocr	读取运营数据	64
1.3.2 机器学习框架: TensorFlow	2.2.5 从 API 获取运营数据	68
1.4 第 1 个用 Python 实现的数据化运营	2.3 内容延伸: 读取非结构化网页、	
分析实例、销售预测	文本、图像、视频、语音	72
1.4.1 案例概述	2.3.1 从网页中获取运营数据	72
1.4.2 案例过程		

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					
3.1 数据清洗：缺失值、异常值和重复值的处理					
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					
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					
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					
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					
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					
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					
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					
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					
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	4.2.4 注意应用回归模型时研究 自变量是否产生变化	198
3.9.4 针对连续数据的二值化	147	4.2.5 如何选择回归分析算法	198
3.9.5 代码实操: Python 数据 离散化处理	147	4.2.6 案例: 大型促销活动前的 销售预测	199
3.10 内容延伸: 非结构化数据 的预处理	151	4.3 分类分析	206
3.10.1 网页数据解析	151	4.3.1 防止分类模型的过拟合问题	207
3.10.2 网络用户日志解析	159	4.3.2 使用关联算法做分类分析	207
3.10.3 图像的基本预处理	164	4.3.3 用分类分析来提炼规则、 提取变量、处理缺失值	208
3.10.4 自然语言文本预处理	169	4.3.4 类别划分: 分类算法和 聚类算法都是好手	209
3.11 本章小结	172	4.3.5 如何选择分类分析算法	210
第 4 章 跳过运营数据分析和 挖掘的“大坑”	174	4.3.6 案例: 用户流失预测分析 与应用	210
4.1 聚类分析	174	4.4 关联分析	221
4.1.1 当心数据异常对聚类 结果的影响	175	4.4.1 频繁规则不一定是有效规则	221
4.1.2 超大数据量时应该放弃 K 均值算法	175	4.4.2 不要被啤酒尿布的故事禁锢 你的思维	222
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	4.4.3 被忽略的“负相关”模式 真的毫无用武之地吗	223
4.1.4 高维数据上无法应用聚类吗	178	4.4.4 频繁规则只能打包组合应用吗	224
4.1.5 如何选择聚类分析算法	179	4.4.5 关联规则的序列模式	224
4.1.6 案例: 客户特征的聚类与 探索性分析	179	4.4.6 使用 Python 和 R 分别实现 商品交叉销售分析	225
4.2 回归分析	196	4.5 异常检测分析	240
4.2.1 注意回归自变量之间的 共线性问题	197	4.5.1 异常检测中的“新奇检测” 模式	241
4.2.2 相关系数、判定系数和回归 系数之间是什么关系	197	4.5.2 将数据异常与业务异常相分离	241
4.2.3 判定系数是否意味着相应的 因果联系	197	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	4.9 内容延伸 1：非结构化数据的分析与挖掘	278
4.6 时间序列分析	253	4.9.1 词频统计	278
4.6.1 如果有自变量，为什么还要用时间序列	253	4.9.2 词性标注	281
4.6.2 时间序列不适合商业环境复杂的企业	254	4.9.3 关键字提取	284
4.6.3 时间序列预测的整合、横向和纵向模式	255	4.9.4 文本聚类	286
4.6.4 案例：未来 10 天的整体销售预测	255	4.10 内容延伸 2：AutoML——自动化数据挖掘与机器学习	289
4.7 路径、漏斗、归因和热力图分析	267	4.10.1 自动化数据挖掘与机器学习概述	289
4.7.1 不要轻易相信用户的页面访问路径	269	4.10.2 TPOT 自动化学习简介	290
4.7.2 如何将路径应用于更多用户行为模式的挖掘	270	4.10.3 使用 TPOT 做自动化学习	292
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	4.11 本章小结	298
4.7.4 点击热力图真的反映了用户的点击喜好吗	271	第 5 章 会员数据化运营	300
4.7.5 为什么归因分析主要存在于线上的转化行为	272	5.1 会员数据化运营概述	300
4.7.6 漏斗分析和路径分析有什么区别	273	5.2 会员数据化运营关键指标	301
4.8 其他数据分析和挖掘的忠告	274	5.2.1 会员整体指标	301
4.8.1 不要忘记数据质量的验证	274	5.2.2 会员营销指标	301
4.8.2 不要忽视数据的落地性	274	5.2.3 会员活跃度指标	303
4.8.3 不要把数据陈列当作数据结论	275	5.2.4 会员价值度指标	305
4.8.4 数据结论不要产生于单一指标	276	5.2.5 会员终生价值指标	305
4.8.5 数据分析不要预设价值立场	276	5.2.6 会员异动指标	306
4.8.6 不要忽视数据与业务的需求冲突问题	277	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	销响应预测	342	
5.5	会员数据化运营分析小技巧	313	5.8.1	案例背景	342
5.5.1	使用留存分析做新用户		5.8.2	案例主要应用技术	343
	质量分析	313	5.8.3	案例数据	343
5.5.2	使用 AARRR 做 App 用户		5.8.4	案例过程	344
	生命周期分析	314	5.8.5	案例数据结论	353
5.5.3	借助动态数据流关注会员		5.8.6	案例应用和部署	353
	状态的轮转	315	5.8.7	案例注意点	354
5.5.4	使用协同过滤算法为新会员		5.8.8	案例引申思考	355
	分析推送个性化信息	316	5.9	本章小结	355
5.6	会员数据化运营分析的		第 6 章	商品数据化运营	357
	“大实话”	318	6.1	商品数据化运营概述	357
5.6.1	企业“不差钱”，还有必要		6.2	商品数据化运营关键指标	357
	做会员精准营销吗	318	6.2.1	销售指标	357
5.6.2	用户满意度取决于期望和		6.2.2	促销活动指标	360
	给予的匹配程度	319	6.2.3	供应链指标	361
5.6.3	用户不购买就是流失了吗	319	6.3	商品数据化运营应用场景	363
5.6.4	来自调研问卷的用户信息		6.3.1	销售预测	363
	可信吗	320	6.3.2	库存分析	363
5.6.5	不要盲目相信二八法则	321	6.3.3	市场分析	363
5.7	案例：基于 RFM 的精细化		6.3.4	促销分析	364
	用户管理	322	6.4	商品数据化运营分析模型	364
5.7.1	案例背景	322	6.4.1	商品价格敏感度模型	364
5.7.2	案例主要应用技术	322	6.4.2	新产品市场定位模型	365
5.7.3	案例数据	323	6.4.3	销售预测模型	366
5.7.4	案例过程	323	6.4.4	商品关联销售模型	367
5.7.5	案例数据结论	337	6.4.5	异常订单检测	368
5.7.6	案例应用和部署	341	6.4.6	商品规划的最优组合	368
5.7.7	案例注意点	341	6.5	商品数据化运营分析小技巧	370
5.7.8	案例引申思考	342	6.5.1	使用层次分析法将定量与	
5.8	案例：基于嵌套 Pipeline 和 FeatureUnion 复合数据工作流的营			定性分析结合	370

6.5.2	通过假设检验做促销拉动 分析	373	6.8.4	案例过程	391
6.5.3	使用 BCG 矩阵做商品结构 分析	374	6.8.5	案例数据结论	397
6.5.4	巧用 4P 分析建立完善的商品 运营分析结构	376	6.8.6	案例应用和部署	398
6.6	商品数据化运营分析的 “大实话”	376	6.8.7	案例注意点	398
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	6.8.8	案例引申思考	399
6.6.2	促销活动真的是在促进商品 销售吗	378	6.9	本章小结	400
6.6.3	用户关注的商品就是要买的 商品吗	379	第 7 章 流量数据化运营		402
6.6.4	提供的选择过多其实不利于 商品销售	380	7.1	流量数据化运营概述	402
6.7	案例：基于 Gradient Boosting 的 自动超参数优化的销售预测	381	7.2	八大流量分析工具	402
6.7.1	案例背景	381	7.2.1	Adobe Analytics	403
6.7.2	案例主要应用技术	381	7.2.2	Webtrekk Suite	403
6.7.3	案例数据	381	7.2.3	Webtrends	403
6.7.4	案例过程	382	7.2.4	Google Analytics	404
6.7.5	案例数据结论	389	7.2.5	IBM Coremetrics	404
6.7.6	案例应用和部署	389	7.2.6	百度统计	404
6.7.7	案例注意点	389	7.2.7	Flurry	404
6.7.8	案例引申思考	389	7.2.8	友盟	405
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	7.2.9	如何选择第三方流量分析 工具	405
6.8.1	案例背景	390	7.3	流量采集分析系统的工作机制	408
6.8.2	案例主要应用技术	390	7.3.1	流量数据采集	409
6.8.3	案例数据	391	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

