

数据分析

思维与可视化

黑马程序员 / 编著



清华大学出版社



数据分析

思维与可视化

黑马程序员 / 编著



清华大学出版社
北京

内 容 简 介

本书以实用为设计理念,并结合数据分析相关理论,系统地介绍了数据分析的相关内容,包括电商数据分析概述、数据分析业务指标、数据准备与处理、数据分析常用方法、常用数据分析工具、数据可视化、数据图表专业化、撰写数据分析报告等,能够帮助读者掌握数据分析的整个流程。

本书附有配套的教学 PPT、题库、教学视频、教学设计等相关资源。同时,为了帮助初学者及时地解决学习过程中遇到的问题,还提供了专业的在线答疑平台。

本书可作为高等院校本、专科相关专业的数据分析课程的教材,也可作为企业数据分析岗位培训教材。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

数据分析思维与可视化/黑马程序员编著. —北京:清华大学出版社,2019

ISBN 978-7-302-53382-5

I. ①数… II. ①黑… III. ①数据处理 IV. ①TP274

中国版本图书馆 CIP 数据核字(2019)第 175954 号

责任编辑:袁勤勇 薛 阳

封面设计:韩 冬

责任校对:白 蕾

责任印制:丛怀宇

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 装 者:三河市吉祥印务有限公司

经 销:全国新华书店

开 本:185mm×260mm

印 张:14.25

字 数:344千字

版 次:2019年9月第1版

印 次:2019年9月第1次印刷

定 价:39.00元

产品编号:084038-01



播妞学姐

播妞学姐——IT技术女神，由传智播客旗下院校服务品牌院校邦推出，专门服务于计算机相关专业的大学生及IT爱好者；添加播妞学姐公众号可获取教材源代码。

播妞学姐的其他资源：

IT技术视频

学习路线图

就业宝典

IT实战项目

软件开发环境

.....



关注播妞学姐公众号：boniu1024

领取教材源代码

教师获取教材配套资源

添加微信/QQ

2011168841

序 言

江苏传智播客教育科技股份有限公司(简称传智播客)是一家致力于培养高素质软件开发人才的科技公司,“黑马程序员”是传智播客旗下高端 IT 教育品牌。

“黑马程序员”的学员多为大学毕业后,想从事 IT 行业,但各方面条件还不成熟的年轻人。“黑马程序员”的学员筛选制度非常严格,包括了严格的技术测试、自学能力测试,还包括性格测试、压力测试、品德测试等。百里挑一的残酷筛选制度确保学员质量,并降低企业的用人风险。

自“黑马程序员”成立以来,教学研发团队一直致力于打造精品课程资源,不断在产、学、研三个层面创新自己的执教理念与教学方针,并集中“黑马程序员”的优势力量,有针对性地出版了计算机系列教材 80 多种,制作教学视频数十套,发表各类技术文章数百篇。

“黑马程序员”不仅斥资研发 IT 系列教材,还为高校师生提供以下配套学习资源与服务。

为大学生提供的配套服务

1. 请登录“高校学习平台”<http://yx.ityxb.com>,免费获取海量学习资源,帮助高校学生解决学习问题。
2. 针对高校学生在学习过程中存在的压力等问题,我们还面向大学生量身打造了“IT 技术女神”——“播妞学姐”,可提供教材配套源码和习题答案,以及更多 IT 学习资源,同学们快来关注“播妞学姐”的微信公众号:boniu1024。



“播妞学姐”微信公众号

为教师提供的配套服务

针对高校教学,“黑马程序员”为 IT 系列教材精心设计了“教案+授课资源+考试系

统+题库+教学辅助案例”的系列教学资源,高校老师请登录“高校教辅平台”<http://yx.ityxb.com> 或关注码大牛老师微信/QQ: 2011168841,获取教材配套资源,也可以扫描下方二维码,加入专为 IT 教师打造的师资服务平台——“教学好助手”,获取最新教师教学辅助资源的相关动态。



传智人

2019年6月

责任编辑: 李义忠 封面设计: 李

传智教育 北京传智教育科技股份有限公司

北京传智教育科技股份有限公司 地址: 北京市海淀区中关村大街100号10层1001室 电话: 010-62772017

北京传智教育科技股份有限公司 地址: 北京市海淀区中关村大街100号10层1001室 电话: 010-62772017



北京传智教育科技股份有限公司 地址: 北京市海淀区中关村大街100号10层1001室 电话: 010-62772017

北京传智教育科技股份有限公司 地址: 北京市海淀区中关村大街100号10层1001室 电话: 010-62772017

传智教育 北京传智教育科技股份有限公司

产品编号: 050008-1

北京传智教育科技股份有限公司 地址: 北京市海淀区中关村大街100号10层1001室 电话: 010-62772017

前言

随着数字化技术的广泛应用,我们正在逐步迈入大数据时代。数据已经成为企业的一种新的战略资产,越来越多的企业开始重视数据资产的管理。伴随着数据重要作用的凸显,数据分析也愈发重要,大力拓展数据分析与应用,成为企业制定新的战略目标,突破传统壁垒的重要手段。

为什么要学习这本书

在移动互联网时代,信息的获取成本越来越低,这也导致拿来主义和实用主义盛行。数据分析学科是一个需要紧密联系生活实际的学科,而数据分析也是企业日常工作的必要环节。目前,国内许多高校开设了相关课程,例如一些经济管理类、信息类专业以及电子商务专业。但是,从相关课程内容来看,已经开设的课程内容更偏向理论知识的介绍,对于数据分析综合能力的培养相对薄弱。而通过对本书的学习,可以让读者在实战中掌握数据分析理论和方法,学完之后能够独立完成日常数据分析工作。

如何使用本书

本书从数据分析基础讲起,共8章,详细介绍了数据分析概述、数据准备与处理、数据分析常用方法、常用数据分析工具、数据可视化、数据图表专业化、撰写数据分析报告的相关知识。下面分别介绍各章的主要内容,以帮助读者更好地了解本书的知识架构体系。

第1章:主要介绍电商数据分析概述,包括数据分析基础理论、数据分析流程、数据分析的常见误区、数据分析岗位的职业发展、数据分析常用的指标及术语。

第2章:主要讲解数据分析业务指标,包括网站流量指标、商品数据化运营关键指标、站外营销推广指标、会员数据化运营指标、仓储管理指标、物流配送指标。

第3章:主要讲解数据准备与处理,包括数据准备、数据处理。

第4章:主要讲解数据分析常用方法,包括常用数据分析方法论、数据分析法。

第5章:主要讲解常用数据分析工具,包括常用本地数据分析工具、常用电商数据分析工具、常用网站数据分析工具。

第6章:主要讲解数据可视化的相关知识,包括通过图表展现数据、通过表格展现数据。

第7章:主要讲解数据图表专业化,包括制作严谨的数据图表、图表美化方法、提高图表制作效率。

第8章:主要讲解撰写数据分析报告的相关知识,包括初步认识数据分析报告、数据分析报告的准备与撰写流程、数据分析报告的结构、撰写数据分析报告的注意事项。

第1章和第2章主要介绍数据分析基础知识和分析业务指标;第3章主要介绍如何准备和处理数据;第4章和第5章主要介绍数据分析常用的方法和工具;第6章主要介绍如何通过图表展现数据;第7章主要介绍如何展现数据以及怎样让图表更加专业化;第8章主要介绍如何撰写数据分析报告。

通过本书的系统学习,读者能够掌握基本的数据分析方法和常用工具,可以独立完成日常数据分析工作,成为市场需求的既懂理论、又懂技术的应用型人才。

致谢

本书的编写和整理工作由传智播客教育科技有限公司研究院内容与资源组完成,主要参与人员有吕春林、张鑫、王哲,全体人员在这近一年的编写过程中付出了很多辛勤的汗水,在此表示衷心的感谢。

意见反馈

尽管我们尽了最大的努力,但书中仍难免疏漏和不妥之处,欢迎各界专家和读者朋友来信来函提出宝贵意见。读者在阅读本书时,如发现任何问题或有不认同之处,可以通过电子邮件与我们联系。

请发送电子邮件至: itcast_book@vip.sina.com。

黑马程序员

2019年6月

于北京

图书资源支持

感谢您一直以来对清华版图书的支持和爱护。为了配合本书的使用,本书提供配套的资源,有需求的读者请扫描下方的“书圈”微信公众号二维码,在图书专区下载,也可以拨打电话或发送电子邮件咨询。

如果您在使用本书的过程中遇到了什么问题,或者有相关图书出版计划,也请您发邮件告诉我们,以便我们更好地为您服务。

我们的联系方式:

地 址: 北京市海淀区双清路学研大厦 A 座 701

邮 编: 100084

电 话: 010-62770175-4608

资源下载: <http://www.tup.com.cn>

客服邮箱: tupjsj@vip.163.com

QQ: 2301891038 (请写明您的单位和姓名)

资源下载、样书申请



书圈



扫一扫, 获取最新目录

用微信扫一扫右边的二维码,即可关注清华大学出版社公众号“书圈”。



关注播妞学姐公众号：boniu1024
提供教材源代码

目 录

第 1 章 电商数据分析概述	1
1.1 数据分析基础理论	2
1.1.1 认识数据和数据的价值	2
1.1.2 认识数据分析	4
1.1.3 认识电商数据分析	5
1.2 数据分析的流程	6
1.2.1 明确目的	6
1.2.2 数据收集	7
1.2.3 数据处理	8
1.2.4 数据分析	8
1.2.5 数据展现	9
1.2.6 撰写报告	9
1.3 数据分析的常见误区	10
1.4 数据分析岗位的职业发展	11
1.4.1 数据分析师的发展前景	11
1.4.2 数据分析师的职业要求	11
1.4.3 数据分析师的基本素质	13
1.5 数据分析常用指标及术语	14
1.5.1 平均数	14
1.5.2 绝对数与相对数	14
1.5.3 百分比与百分点	15
1.5.4 比例与比率	15
1.5.5 频数与频率	15
1.5.6 倍数与番数	16
1.5.7 同比和环比	16
小结	16
第 2 章 数据分析业务指标	17
2.1 网站流量指标	18
2.1.1 网站流量数量指标	19

专属于教师及学生的在线教育平台
<http://yx.ityxb.com/>

让 IT 教学更简单

教师获取教材配套资源



添加微信/QQ
2011168841

让 IT 学习更有效

学生获取配套源码

关注微信公众号“播妞学姐”
获取教材配套源码



专属大学生的圈子

2.1.2	网站流量质量指标	21
2.2	商品数据化运营关键指标	27
2.2.1	销售类指标	28
2.2.2	促销活动指标	32
2.3	站外营销推广指标	34
2.4	会员数据化运营指标	39
2.4.1	会员整体指标	39
2.4.2	会员营销指标	40
2.4.3	会员活跃度指标	42
2.4.4	会员价值度指标	43
2.4.5	会员终生价值指标	45
2.4.6	会员异动指标	45
2.5	仓储管理指标	46
2.5.1	库存量	46
2.5.2	库存金额与平均库存金额	47
2.5.3	库存可用天数	47
2.5.4	库存周转率	47
2.5.5	库存周转天数	48
2.5.6	库龄	48
2.5.7	滞销金额	48
2.5.8	残次商品数量、残次商品金额、残次占比	49
2.5.9	缺货率	49
2.6	物流配送指标	49
	小结	50
第3章	数据准备与处理	51
3.1	数据准备	53
3.1.1	认识数据表	53
3.1.2	获取数据	60
3.2	数据处理	66
3.2.1	数据清洗	66
3.2.2	数据加工	76
3.2.3	数据抽样	86
	小结	87
第4章	数据分析常用方法	88
4.1	常用数据分析方法论	89
4.1.1	PEST分析法	89
4.1.2	4P营销理论	92

4.1.3	逻辑树分析法	93
4.1.4	用户行为理论	94
4.1.5	5W2H 分析法	94
4.2	数据分析法	96
4.2.1	对比分析法	96
4.2.2	结构分析法	98
4.2.3	分组分析法	99
4.2.4	平均分析法	100
4.2.5	矩阵关联分析法	100
4.2.6	交叉分析法	103
4.2.7	综合评价分析法	104
4.2.8	漏斗图分析法	106
	小结	111
第 5 章	常用数据分析工具	112
5.1	常用本地数据分析工具	114
5.1.1	SPSS 简介	114
5.1.2	SPSS 的优势	115
5.1.3	SPSS 的窗口介绍	116
5.2	常用电商数据分析工具	118
5.2.1	生意参谋	118
5.2.2	京东商智	122
5.3	常用网站数据分析工具	128
5.3.1	百度统计	128
5.3.2	CNZZ	134
	小结	139
第 6 章	数据可视化	140
6.1	通过图表展现数据	141
6.1.1	认识图表	141
6.1.2	常用的图表类型	143
6.1.3	通过数据关系选择合适的图表	153
6.1.4	统计图制作流程	157
6.2	通过表格展现数据	157
6.2.1	突出显示单元格	157
6.2.2	项目选取	159
6.2.3	添加数据条	160
6.2.4	添加图标集	162
6.2.5	迷你图	164

小结	168
第7章 数据图表专业化	169
7.1 制作严谨的数据图表	170
7.1.1 专业图表中的元素	170
7.1.2 专业图表制作的注意事项	171
7.2 图表美化方法	181
7.2.1 图表美化的原则	182
7.2.2 图表美化技巧	182
7.2.3 图表的颜色搭配	185
7.3 提高图表制作效率	188
7.3.1 创建图表模板	188
7.3.2 添加标签小工具	188
7.3.3 修剪超大值	192
小结	194
第8章 撰写数据分析报告	195
8.1 初步认识数据分析报告	198
8.1.1 什么是数据分析报告	198
8.1.2 数据分析报告的作用	198
8.1.3 数据分析报告的种类	199
8.1.4 数据分析报告写作原则	200
8.2 数据分析报告的准备与撰写流程	201
8.2.1 数据分析报告的准备	201
8.2.2 数据分析报告撰写流程	202
8.3 数据分析报告的结构	202
8.3.1 标题页	202
8.3.2 目录	204
8.3.3 前言	204
8.3.4 报告正文	206
8.3.5 结论与建议	207
8.3.6 附录	207
8.4 撰写数据分析报告的注意事项	213
小结	214

第 1 章

电商数据分析概述

【学习目标】

知识目标	➢ 了解电商数据分析的基础知识 ➢ 熟悉数据分析常用指标
技能目标	➢ 掌握数据分析的流程

【案例引导】

数据分析帮助公司提高销售额

叮叮网是一家成立时间不长的电商网站,公司主推的一款洁面霜上半年销售情况一直不是很好,已经连续 3 个月没有完成目标销售额,按照这个趋势,很难完成 12 月 1000 万元的销售额。

开会时,老板让负责数据分析工作的小王针对上半年的销售数据做一下分析,找到可以提升下半年销售额的方法,如图 1-1 所示为这款洁面霜上半年的销售额及广告投放情况。

月份	销售额/万元	目标售额/万元	广告投放费用/万元		产品 单价/元
			社交媒体广告	信息流广告	
1	574.57	547.80	120	30	249
2	597.20	597.60	120	30	249
3	615.58	610.05	100	50	249
4	622.08	647.40	100	50	249
5	607.51	669.20	100	30	239
6	628.04	717.00	80	70	239

图 1-1 洁面霜上半年的销售额及广告投放情况

通过对这款洁面霜上半年的销售数据、广告投放费用和营销活动数据的分析,小王发现上半年的销售中存在的问题如下。

(1) 1~6 月公司的销售额整体呈增长趋势,但是增幅很小,而且与目标销售额的差距在变大。

(2) 针对女性群体投放的社交媒体广告投放虽然一直保持在较高的水平,但是女性购买用户数量并没有增加;而针对男性消费者投放的信息流广告虽然不多,但是男性购买用户数量尤其是中老年男性的购买数量却随之增加了,男性用户购买这款洁面霜大多是因为其可以作为须后水使用,缓解剃须后的脸部过敏问题。

(3) 除了降价和广告投放之外,上半年并没有做过其他专门的营销活动以提升销量。

针对分析中发现的问题,小王在第二次会议上建议从以下几个方面对下半年的营销策略进行改进。

(1) 增加针对中老年男性的信息流广告的投放力度,降低社交媒体广告的投放比例。

(2) 在针对中老年男性的信息流广告中,突出洁面霜可以作为须后水使用这一卖点,激发其购买欲望。

(3) 针对这款洁面霜设计专门的优惠活动,例如品牌日半价等,刺激产品销量提升,从而带动销售额上涨。

根据小王的建议,老板让运营和营销部门对下半年的运营及营销策略做了相应调整。最终,这款洁面霜每个月都达到了目标销售额,12月份也成功达成了1000万元销售额的目标,如图1-2所示。

月份	销售额/万元	目标销售额/万元	广告投放费用/万元		产品单价/元
			社交媒体广告	信息流广告	
7	698.62	675.66	80	70	249
8	806.01	783.73	75	75	249
9	838.08	797.90	60	90	249
10	959.67	880.19	60	90	249
11	960.49	898.67	65	85	249
12	1001.85	992.74	60	90	249

图1-2 洁面霜下半年的销售额及广告投放情况

【案例思考】

在本案例中,叮叮网通过对上半年销售数据的分析发现问题,并针对这些问题调整下半年的运营及营销策略,在下半年不仅每个月都完成目标销售额,并且在12月份最终达到1000万元的目标。从该案例可见,数据分析能产生巨大的价值。然而什么是数据分析?它的作用是什么?电商数据分析的流程是怎样的?本章将对电商数据分析的基础知识做详细讲解。

1.1 数据分析基础理论

数据分析是现代社会的不可或缺的一门学科,在各行各业中有着广泛的应用。创业者通过数据分析可以对产品进行优化;运营人员可以根据数据分析结论调整运营策略;产品经理可以使用数据分析洞察用户习惯;金融从业人员可以通过数据分析来规避投资风险;程序员可以通过数据分析进一步挖掘用户的价值。可见,无论从事哪一行业,从业者善于通过数据分析发现行业问题,都会让自身在行业中更具竞争力。本节将对数据分析基础理论进行介绍。

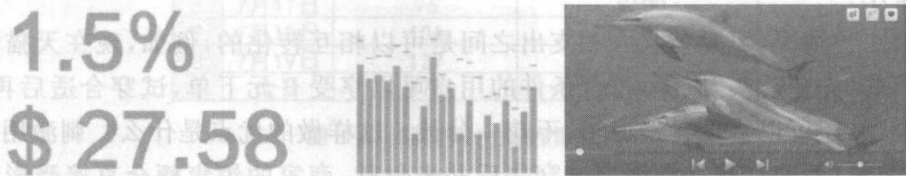
1.1.1 认识数据和数据的价值

现在很多人开口必谈大数据,许多人都说自己在研究大数据,到处宣称数据可以创造价值。那么,到底什么是数据?数据又有哪些价值?

1. 什么是数据

什么是数据?这个问题看似简单其实并不好回答。通常来说,数据可以理解为“观测

值”，是通过观察结果、实验或测量的方式获得的结果，通常是以数量的形式来展现。但是，数据不仅指狭义上的数字，还可以是具有一定意义的文字、字母、数字符号的组合、图形、图像、视频、音频等，也是客观事物的属性、数量、位置及其相互关系的抽象表示。这样一来，数据的范畴就要大得多了，绝不仅局限于数字。如图 1-3 所示即数据的真实体现。



(a) 数字是数据

(b) 声音是数据

(c) 视频是数据

图 1.3 数据的真实体现

2. 数据的价值

增加收入、减少支出、控制风险是关乎企业生存发展的重要问题，而通过对数据的分析，可以对这三方面进行优化改进，使企业良性地发展，这就是数据价值的体现。

1) 增加收入

作为一名专业的数据从业者，你是否想过，你的数据产品能否给公司带来额外的收入呢？下面以 100 套护肤品的价值为例进行说明，如例 1-1 所示。

例 1-1 100 套护肤品的价值

叮叮网是主做护肤品的电商企业，在没使用数据分析之前，每天能卖 100 套。使用了数据分析后，每天能卖多少套呢？如果还是卖 100 套，那么数据分析的价值在哪里呢？如果现在每天能卖 200 套甚至更多，那么数据分析的价值就体现出来了。这个价值的大小就是每天额外多卖的 100 套商品。

2) 减少支出

很多做数据分析的人会觉得自己的工作离市场销售端有点儿远，并不能直接为公司增加收入。但是换个角度思考一下，如果帮助公司节约不必要的支出，也就是成本，这比为公司直接增加收入更有意义。因为公司的收入增长往往会面临很多的不确定性，但是成本却在自己的控制范围内，相对而言可控性更强。下面以叮叮网的客服中心升级改进为例进行说明，如例 1-2 所示。

例 1-2 叮叮网客服中心运营改进

叮叮网现在负责客服工作的有 20 人，但是经过系统升级、数据分析、合理排班，发现只要 10 个客服就可以了。直接节省了 10 个客服人员的人工成本，这是非常确定的事情。所以，如果数据分析可以节省开支会更好，因为更靠谱、可控性更强。

3) 控制风险

有人可能会有这样的疑问，我的数据既不能给公司增加收入，也不能直接节省成本，但是可以控制风险，这样的数据有商业价值吗？当然有。实际上，风险的衡量有两种情况：第一种情况是风险根本无法通过货币衡量，是独立于收入或者支出的另外一个维度；第二种情况是风险可以作为连接收入与支出的一个转化器，通过风险把控，或者可以增加收入，或者