

阿里云总裁**胡晓明**、明道创始人**任向晖**力荐！

帮助我们在日常工作中感知和发掘数据的价值！

Broadview[®]
www.broadview.com.cn

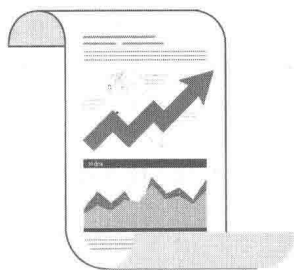


数据化运营 速成手册

胡晨川 著

 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>



数据化运营 速成手册

胡晨川 著

电子工业出版社

Publishing House of Electronics Industry

北京•BEIJING

内 容 简 介

本书用于提升互联网公司员工的数据应用能力,即数据化运营能力。首先,从最常用的数据图表切入,帮助执行层正确地绘图,管理层正确地看图;接着,梳理运营中最基本的数据应用知识,涉及数据获取、数据清洗、数据认知、分析框架、指标体系、运营实验等内容。然后,介绍作者认为必要的统计学知识,包括假设检验、方差分析、回归分析和时间序列分解,并引入了管理科学中的规划求解方法。最后,介绍了数据分析工具的发展趋势,并分享了作者近些年的工作及学习心得。

本书适用于互联网公司的数据分析师、运营人员、产品经理和中层管理人员。

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

数据化运营速成手册 / 胡晨川著. —北京:电子工业出版社, 2017.5

ISBN 978-7-121-31267-0

I. ①数… II. ①胡… III. ①企业管理—数据管理—手册 IV. ①F272.7-62

中国版本图书馆 CIP 数据核字(2017)第 070019 号

策划编辑:郑柳洁

责任编辑:郑柳洁

印 刷:三河市良远印务有限公司

装 订:三河市良远印务有限公司

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:787×980 1/16 印张:16.75 字数:340 千字

版 次:2017 年 5 月第 1 版

印 次:2017 年 5 月第 2 次印刷

印 数:3001~7000 册 定价:55.00 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 zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式:010-51260888-819, faq@phei.com.cn。



推荐序一

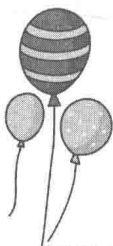
随着信息经济生态的加速形成，数据的力量正在被不断激发。智能设备、物联网、云计算等技术帮助人类构建着一个日益丰富的数据世界，其越来越完整地反映出我们所在的现实世界的面貌，并且通过对数据的进一步分析和归纳，探寻其中蕴含的规律，使我们对世界的认知能够达到过去几千年从未企及的深度。

一直以来，阿里云致力于为客户提供普惠的计算能力和服务能力，我们希望通过阿里云提供的技术和平台，帮助客户把握 DT 时代的入口，让客户的数据得以不断沉淀，能够产生更大的价值。今天，我们很欣喜地看到，越来越多的企业拥有了数据资产，拥有了通往未来无限想象的可能。但是我们也清楚地认识到，只有让数据产生商业价值，数据才能真正为社会、为消费者、为各行各业产生力量。

在浩如繁星的数据之下，一双能够理解和利用数据的深邃明眸显得尤其珍贵，在《数据化运营速成手册》中，我看到了一个热爱数据分析工作的年轻人，如何不断回顾、总结、提炼、打磨自己的工作场景，尽他所能地融合统计科学、商业思维、视觉工具等多领域知识，以最具象的方式与我们分享他的所思所得。他分享的知识也许并非前沿，操作指南也并非殿堂级的严谨，但却很有温度。这些内容从最具体的工作中来，带我们到最普遍也是最实用的应用场景中去，帮助我们在日常工作中感知和发掘数据的价值，相信会吸引并帮助更多的人认识和使用数据。

大数据时代的帷幕刚刚拉开，对于数据资源这座宝贵的矿藏，我们今天的所见也许只是九牛一毛，未来还有无限的可能等待我们去挖掘。但是我相信，与应用相结合，让数据产生商业价值，永远是数据的生命所在。希望有更多像晨川一样对数据抱有热忱的年轻人加入进来，共同探索如何让数据最有效地为商业需求服务，从中激发和传播数据的能量，推进新的变革，期待与你们同行。

阿里巴巴集团资深副总裁、阿里云总裁，胡晓明



推荐序二

和很多互联网公司的 CEO 一样，我也本能地重视数据。在数据化运营的蛮荒时代，所谓的运营工作也许只有一个简单的访问计数。即使这样，无论是个人网站主、网店店主，还是博客作者都会忍不住刷屏，恨不得听到网站运行的每一个脉动。

后来，我们有了 Google Analytics 这样的工具，脚本一载入，就像拥有了上帝视角，把网站的角角落落和用户的起起落落都看得一清二楚。我们曾经认为这已经是传统商业难以企及的运营条件，它能够帮助我们打败对手，建立长久的竞争优势。但互联网商业发展的日新月异，让它的运营水平也跟着一起神速推进。很快事情就不再那么简单，数据运营工作的复杂度和专业度每年都在翻倍，仿佛昨天还在学习一元二次方程，今天就已经是微积分的难度。互联网行业的发展速度在提供创富神话的同时，也给我们的学习能力下了最后通牒。

我们已经无法凭借直觉经验和抓大放小这样的简单原则来应用数据化运营。任何一项网络服务都面临着复杂的数据观测、分析和应用需求，从用户获取、激活、留存、转化和再传播，每一个环节都能够派生出上百个数据点来提供不同维度的指标，我们面临遴选指标、跟踪和汇报指标，以及建立 PDCA 改进行动机制的高难度协作。这既不是一个人能够埋头做完的活儿，也不是人多就力量大的任务。有效的数据化运营工作考验团队的领导力，识别重要任务的能力，沟通和协作的能力；同时也必然需要有团队成员专注在这个环节，完整掌握这一组专业能力。晨川的《数据化运营速成手册》正是从他的专业岗位经验出发，为从业人员提供的一本实务指南。

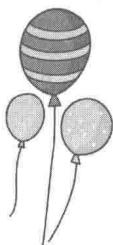
本书的亮点还在于作者对有效学习的理解，以及构建知识框架的基本方式和专业读者的实践需求。从基础图表这个具象的事物开始，再升级到图表的高级特性。然后再介绍数据化运营的基本方法和原则，再抽象到更深入的数据化运营思维模式。这种浅—深—浅—深的节奏也是我特别认可的学习模式。本书也留出了相当的篇幅跳出了基本的数据化运营工作，延



伸到有深度的量化分析和决策科学领域，这对有意愿更深入研究和开发数据化运营工具与模型的专业人员也有启发作用。对于 CEO 来说，我们一直说重视数据，养成基于数据设定目标和决策的习惯，但是这一切需要掌握了基本的数据技能后才能奏效。如果你不够了解数据，自然不会真正信任它。

晨川在明道工作时，用他的专业和热情帮助明道建立了数据化运营的基本框架。如今，他又能为行业留下实务著作，我有必要对他致以双重的感谢，并真诚地推荐读者静下心来，开启这项重要能力的学习。

明道创始人，任向晖



好评袭来

对数据的驾驭能力不足，往往是很多运营人员的能力短板。把抽象的数据封装进生动直观的图表中，调动起更多的感官能力洞察数据，是高明之举。作者有扎实的数据功底和运营实战，从图表切入，向读者展现丰富的数据运营技巧和知识，具有极高的实战价值，力荐！

滴滴企业级事业部战略负责人，郎海舰

认识胡晨川已经有6年了，这些年来，他始终都在做一件事儿——努力成为一流的数据分析专家。大家眼前的这本《数据化运营速成手册》算是他的一个阶段性成果，对于数据这点事儿，他已经形成了自己的一套理论体系，相信本书会对所有希望学习数据分析的朋友有所帮助。

头头是道文化产业基金董事，许维

晨川在互联网数据分析方向具有丰富的实践经验，特别是对数据分析方向的新人可能踩的“坑”，有比较深刻的见解。通过本书，读者可以清晰地了解每种图表展示方式的应用场景及边界，并可以直接应用到自己的数据分析工作中。

神策数据创始人&CEO，桑文锋

每一个优秀的企业都离不开精细的数据化运营，数据化运营离不开正确的数据思维方式和熟练的数据分析技能，而思维方式和技能的培养则需要体系化的“武装”。本书由点及面，从很多数据分析师容易忽略的数据图表开始切入，深入到数据化运营的基础知识和思维方式的讲解，再到实操技能的扩充与真实工作经验的总结，是一本集理论与实践的好书，非常值得推荐！

天善智能创始人，梁勇

前言



尊敬的读者，您能翻开本书，是我的荣幸。虽然书名中含有“速成”两字，但我相信理性的读者都不会期待真的有学习捷径。我发现，数据分析的入门者往往不能快速形成合适的知识框架，导致其走了许多弯路。因此，本书不遵从教科书逻辑，而从实际应用出发，相信能够提高读者的学习效率。总的来说，这是一本关于商业数据应用的书。希望本书的内容能给读者的工作和生活带来实质性的帮助。

为什么写本书

长久以来，我都有成为数据科学家的梦想。虽然朝着这个目标奋斗了几年，但是若问我数据科学家到底是什么，我依然没有明确的答案。两件事儿，是数据科学家必须做到的：一是运用合理的数据分析方法，从数据中形成商业决策；二是能让越来越多的人正确地理解并应用数据分析方法。这两件事儿，也就成为我的奋斗目标。所以，当电子工业出版社的编辑联系我写作本书时，我欣然应允了。

谁适合阅读本书

“让读者正确地理解并应用数据分析”是本书的主要目的。广义地说，本书的内容浅显，适合所有对数据分析感兴趣的人。但依我的初衷，本书是写给互联网公司的运营、产品和管理者的。毕竟，我的工作经验均来自互联网背景，因此认知会有一定的局限性，故而这些内容不一定适用于所有行业。

如何阅读本书

首先，建议读者在阅读本书的同时上手操练。本书内容所涉及的实践操作，全部基于Excel 2016。Excel上手简单、功能全面且运行稳定，它的实用性无须多言。

本书的内容结构比较“怪异”。先通过前三章解决“如何正确地作图”这个问题。为什么要将作图放在最前面呢？因为，运营和产品人员要在工作中制作大量图表，管理者要阅读大量图表；若图表做得不正确，难免出现信息扭曲，导致管理者决策失误。所以，作图者应该正确地作图，管理者应该正确地看图。因此，作者从作图切入，进而展开本书的内容，比较符合实际。

接着，在第4章介绍了运营中最基本的数据应用知识。本章涵盖的内容较多。作者的写作逻辑是“缺什么补什么”，即以作者的认知为基础，填补作者认为本书目标读者所欠缺的数据知识（涉及数据库基本结构、数据获取、数据清洗、数据认知、建立分析框架、构建指标体系、运营活动量化等）。

第5章介绍一些简单实用且必不可少的统计学知识。通过这些知识，读者能够培养自己的量化分析思维，并能体会到统计学的强大。第6章篇幅不长，但希望能为读者打开“管理科学”这扇大门，即应用运筹学中的规划求解方法。让读者体会到：做决策绝不仅是定性分析，它可以一步步量化获得最优解。

第7章简单介绍了数据分析的工具和运营工具。里面的部分内容由 GrowingIO 创始人张溪梦先生和神策数据创始人桑文锋先生执笔。第8章写了一些作者近年的心得体会，希望这些经验能帮助读者少走弯路。

作者建议初次阅读本书的读者先按章节顺序快速通读。在通读之后，若在工作中遇到了某些难处，可以有针对性地精读某一部分。另外，本书中所提及的参考书目，希望读者重视，这些书都是很好的学习资源。

建议读者关注作者的微信公众号“川术”。作者会不定期地分享自己的感悟。本书中的数据都是“模拟的”，建议读者将本书所提的方法应用在自己的数据中。若有强烈的需要，读者可以在作者的微信公众号留言或者发邮件至 huchenchuan@hotmail.com。

感谢

首先，感谢出版方电子工业出版社，给予我机会将自己的经验分享给大家。接着，感谢本书的编辑郑柳洁及其他工作人员，在本书的撰写和出版过程中给予的极大帮助。然后，要感谢我的老师、朋友和同事们，没有他们，就没有我经验的积累和认知的提高。还要感谢本书的推荐人和其他专家，是他们的指导和关怀，让我能高效地前进。最后，感谢我的家人，尤其是我有孕在身的妻子，在每一个伏案写作的夜晚，给予我莫大的理解和无微不至的照顾。

作者



目 录

第 1 章 全面认识数据图表	1
1.1 详解数据图表的基本构成	1
1.2 控制数据图表中的信息量	9
1.3 真的需要作图吗	10
第 2 章 建立数据图表的认知	14
2.1 利用散点图探究数据间的关系	14
2.1.1 最基本的散点图样例	15
2.1.2 散点图的制作	16
2.1.3 散点图的变种 1: 添加平滑线	19
2.1.4 散点图的变种 2: 利用气泡图观察更多指标间的关系	20
2.1.5 散点图的变种 3: 用分类矩阵形成决策	22
2.1.6 散点图的局限性	23
2.2 利用柱形图将“对比”做到极致	24
2.2.1 利用累加柱形图对比数据结构的变化	25
2.2.2 多指标组合对比	27
2.2.3 用平均值优化单指标的对比	29
2.2.4 用瀑布图观察总量分解后的对比	32
2.2.5 如何正确对比数值指标与比率指标	33
2.3 用折线图观察时间序列数据	35



2.3.1	如何观察趋势	36
2.3.2	探寻趋势变化的原因	38
2.4	利用面积图观察数据结构的变化趋势	44
2.4.1	观察动态的数据结构变化：堆积面积图	44
2.4.2	用于队列分析：堆积面积图	45
2.5	用雷达图进行静态的多维对比	48
2.6	其他类型图表概述	50
2.6.1	使用饼图的 6 个“坑”	50
2.6.2	提升视觉冲击力：树状图	52
2.6.3	量化流程各环节间的转化率：漏斗图	52
第 3 章 数据图表进阶		54
3.1	数据图表到底是什么	55
3.2	如何正确地选择图表	59
3.3	数据图表中的细节	60
3.3.1	图表背景和绘图区背景	60
3.3.2	坐标轴	65
3.3.3	灵活使用辅助线	68
3.3.4	线性趋势线的应用	69
3.3.5	应用移动平均趋势线做时间序列的预测	72
3.3.6	添加信息增强线	74
3.3.7	用标注线指示必要的信息	75
3.4	能让图表升级的高级技巧	76
3.4.1	运用组合图表增加信息承载量	76
3.4.2	运用子母图增加图表中的信息量	79
3.4.3	条件格式中的几项实用功能	80
3.4.4	使用迷你图表压缩空间	85
3.4.5	用不等宽技术优化柱形图和条形图	86
3.4.6	使用 Bullet 图进行绩效评价的可视化	89



第4章 数据化运营的基础知识.....	93
4.1 最基本的数据获取能力	94
4.1.1 认知数据库的一般构造.....	94
4.1.2 能够阅读最基本的取数代码.....	95
4.1.3 用 Excel 获取数据	98
4.2 快速认知数据	105
4.2.1 仔细审核数据源的质量.....	106
4.2.2 提升数据集的质量	107
4.2.3 统一数据类型和单位	111
4.2.4 描述统计分析.....	111
4.2.5 利用相关系数理解数据之间的关系.....	122
4.2.6 通过多维交叉深入认知数据集	125
4.3 几套有用的分析思维框架	132
4.3.1 66 法则与 SQVID 原则	133
4.3.2 麦肯锡的“七步成诗”	135
4.4 创造指标, 应用指标	141
4.4.1 什么是指标.....	142
4.4.2 如何设计高质量的指标.....	143
4.4.3 指标组合: 综合指数	143
4.4.4 需要关注哪些核心指标.....	144
4.5 运营活动的量化	148
4.5.1 优秀的运营活动应具备哪些要素	148
4.5.2 需要哪些过程型和结果型指标	149
4.5.3 如何评价运营活动	149
4.5.4 从纵向与横向两个角度全面对比	151
4.5.5 一种更严谨的测试效果量化方法: DID	154
4.5.6 相似活动间效果的对比.....	155
4.5.7 关于运营活动量化的小结	156
4.6 数据化运营的思维方式	156
4.7 运营数据报告的必备要素	167



4.8 小结	170
第5章 快速提升量化分析能力	171
5.1 用朴素贝叶斯模型进行预测	171
5.1.1 利用全概率公式的一个例子	174
5.1.2 让大数定律给你自信	175
5.1.3 窥一斑而见全豹：中心极限定理	176
5.2 使用假设检验进行理性的推断	177
5.2.1 统计分布是一切推断的基础	179
5.2.2 以正态分布为例，阐述假设检验的过程	181
5.2.3 双侧检验与单侧检验	183
5.2.4 假设检验的细节补充	184
5.3 利用方差分析辨别方案的有效性	185
5.3.1 用户激活措施的有效性判断	186
5.3.2 运用置信区间增强数值估计的可靠性	189
5.3.3 两两比较寻找最精确的结论	190
5.3.4 理解方差分析的思维	191
5.4 浅谈回归技术的应用	191
5.4.1 因变量与自变量的相关关系是回归的基础	191
5.4.2 线性回归建模的详细过程	192
5.4.3 线性回归分析中的注意点	204
5.5 用时间序列分解模型观察波动	204
5.5.1 怎样观察时间序列数据	205
5.5.2 何为时间序列分解	206
5.5.3 时间序列分解的步骤解析	208
5.5.4 时间序列分解方法的应用局限性	212
5.6 如何优化调查问卷	213
5.6.1 态度型问题，增加选项以支撑量化分析	214
5.6.2 问题要有必要且贴合业务需求	215
5.6.3 设置过滤器，识别无效回答	215



5.6.4 避免双重问题和一重半问题	216
5.6.5 动态地调查, 设置问题库以保障多次调查的质量	216
第 6 章 科学地决策	217
6.1 从数据中形成决策	217
6.2 线性规划是什么	219
6.3 线性规划建模的操作过程	220
6.3.1 建立逻辑清晰的表格	220
6.3.2 设置输出单元格、目标单元格与可变单元格之前的运算关系	221
6.3.3 设置线性规划建模参数	222
6.4 如何从数据中形成决策	223
6.5 4 类典型且实用的线性规划模型	223
6.5.1 资源分配模型	224
6.5.2 成本收益平衡模型	226
6.5.3 网络配送模型	227
6.5.4 混合模型	229
6.6 线性规划模型小结	231
第 7 章 应用优秀的工具	233
7.1 互联网数据分析工具的演进	234
7.1.1 流量时代	234
7.1.2 用户时代	235
7.1.3 订单时代	235
7.2 行为事件分析工具的简单介绍	236
7.3 数据分析平台简介及趋势	238
7.3.1 数据展现	238
7.3.2 数据处理	238
7.3.3 数据收集	239
7.4 值得推荐的小工具	240
7.4.1 团队协作工具	240
7.4.2 其他小工具	241

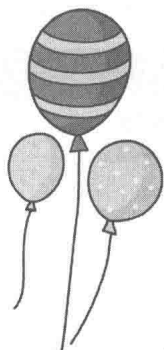
第 8 章 工作经验杂谈	242
8.1 这些年犯过的错误	242
8.1.1 迷信业务模型，浪费公司资源	243
8.1.2 活跃率陷阱	244
8.1.3 不加选择地进行数据追踪	245
8.1.4 为了 KPI 而做数据分析	246
8.1.5 忽略数据质量的保障机制	247
8.1.6 轻视业务执行，重视数据表现	248
8.1.7 不重视数据认知，盲目建模	248
8.1.8 只重视完成任务，忽略了团队成员的个人发展需求	249
8.2 认清数据分析的边界	250
8.2.1 数据库并不能记录一切	250
8.2.2 不可能分离多重因素影响	251
8.2.3 数据不能替代逻辑推理	251
8.2.4 预测的根基未必牢固	251
8.2.5 大多数人会因数据而变懒	252
8.3 我们需要读些什么书	252

轻松注册成为博文视点社区用户（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/31267>





第 1 章

全面认识数据图表

不管你在公司中是何角色，数据图表也许都是你与数据分析的最大交集。翻开前两章，你会觉得这是一本数据图表的应用书；再往下翻，你又会觉得它是一本数据分析的入门书；看完整本书，作者希望你能认为它是一本学习数据化运营的好书。本书从直观的数据图表切入，告诉大家怎么理解数据、应用数据，让数据真正帮助到运营。

1.1 详解数据图表的基本构成

“由浅入深、循序渐进”是学习的固有规律。作者不是一个特立独行的人，因此，本书从最基础的内容开始讲起。别担心，本书不会网罗一批历史故事、名人警句，作者从自己用得最多的一张图开始，解构图表，看看一张统计图表需要具备哪些元素。

若你已经是一名数据分析师，并对用 Excel 作图有相当的了解，那作者建议你跳过本章的内容。若你是一个管理者，只需要解读图表中的信息，不需要亲手制作数据图表，那么作

者也建议你跳过本章。

图 1-1 所示为作者平时用得最多的一种图表形式，有哪些元素构成了这张数据图表呢？我们来一一分解。



图 1-1 制作你的第一个数据图表

1. 定义数据中的维度和指标

你的注意力可能会最先集中在标题上，但它并不是最基础的元素。所有数据图表元素都依托于数据，而数据并不单单是数字，我们所指的数据是量化数据。这个“数据”是统计学范畴的概念，需要区别于计算机范畴所指的“数据”。在计算机范畴中，数据不仅仅是数字和维度，还涵盖了文本、图片、视频等能被存储介质记录的信息。

数据图表的基础数据分为两个部分：指标 (Measure) 和维度 (Dimension)。举个例子，体重、年龄是数字，就是指标；性别、肤色是类别，就是维度。若只告诉你“平均体重 75kg”，你一定会感到困惑，而说“白人男性的平均体重是 75kg”时，你便明白了描述的内容。说明维度和指标相互组合，才会形成有意义的数据。

年龄不仅是数字，有时会变成年龄段（幼年、青年、老年），即变成了类别，也就成了维度。所以，维度和指标是可以相互转化的。换个角度讲，指标是可以加减乘除做四则运算的，而维度则不行，维度只能做排序、包含、与、或、非等集合运算（时间维度是个特例）。

在图 1-1 中你会发现，“segment”、“日期”等筛选项对应着维度，而柱形、折线等图形