

Pearson

打造大数据团队

从组建到价值创造全流程指导

[美]朱达·菲利普斯 (Judah Phillips) ◎著

王 扬 闫化平◎译

国际数据分析研究所创始人 托马斯·H.达文波特
作序推荐

数据分析师、企业管理者必备，第一本打造大数据分析团队和实现数据分析变现的实用指南

从数据收集、管理、分析到执行，教你打造一流的数据分析团队。

BUILDING A DIGITAL
ANALYTICS ORGANIZATION

CREATE VALUE BY INTEGRATING
ANALYTICAL PROCESSES,
TECHNOLOGY, AND PEOPLE
INTO BUSINESS OPERATIONS

中国人民大学出版社

打造 大数据团队

从组建到价值创造全流程指导

[美]朱达·菲利普斯 (Judah Phillips) 著

王 扬 闫化平 译



**BUILDING A DIGITAL
ANALYTICS ORGANIZATION**

CREATE VALUE BY INTEGRATING ANALYTICAL
PROCESSES, TECHNOLOGY, AND PEOPLE
INTO BUSINESS OPERATIONS

图书在版编目 (CIP) 数据

打造大数据团队：从组建到价值创造全流程指导 / (美) 朱达·菲利普斯 (Judah Phillips) 著；王扬，闫化平译. -- 北京：中国人民大学出版社，2018.7

书名原文：Building a Digital Analytics Organization: Create Value by Integrating Analytical Processes, Technology, and People into Business Operations

ISBN 978-7-300-23734-3

I. ①打… II. ①朱… ②王… ③闫… III. ①数据处理 IV. ① TP274

中国版本图书馆 CIP 数据核字 (2016) 第 289914 号

打造大数据团队：从组建到价值创造全流程指导

[美] 朱达·菲利普斯 著

王 扬 闫化平 译

Dazao Dashuju Tuandui: Cong Zujian Dao Jiazhi Chuangzao Quanliucheng Zhidao

出版发行 中国人民大学出版社

社 址 北京中关村大街 31 号

邮政编码 100080

电 话 010-62511242 (总编室)

010-62511770 (质管部)

010-82501766 (邮购部)

010-62514148 (门市部)

010-62515195 (发行公司)

010-62515275 (盗版举报)

网 址 <http://www.crup.com.cn>

<http://www.ttrnet.com> (人大教研网)

经 销 新华书店

印 刷 天津中印联印务有限公司

规 格 170mm×230mm 16 开本

版 次 2018 年 7 月第 1 版

印 张 17.75 插页 1

印 次 2018 年 7 月第 1 次印刷

字 数 283 000

定 价 69.00 元

版权所有

侵权必究

印装差错

负责调换

推荐序

过去半个多世纪里，与信息及其技术管理相关的学科逐渐发展起来。计算和编程早期主要属于学术活动，之后才逐渐发展为真正的职业。公司首次利用数据处理（data processing）方法处理金融和其他形式的内部信息时，就是在先前无结构化的领域引入正规的程序与结构。计算机中心的操作员通常身穿白色工作服，以展示他们的专业性和科学性。虽然这种高级形式也许与数据处理并不匹配，但是它允许管理结构化的内部信息并最终掌握它，使其发展成专业化的领域。

在过去十年里，信息管理领域进入了全新的时代。信息管理是互联网的产物，即数据来源于网络、邮件、线上内容、移动设备、各种手机应用程序和日渐发展壮大的“物联网”。如同早期的计算机工作一样，初期的数据管理是非正式的，更像某种业余爱好。公司通常都有兼职的“网络人员”去设计、安装和维护公司的网站。数字化业务方面缺少衡量标准，其他方面也一样缺乏管理。一些大型知名企业的网站经常中断，甚至其注册域名有时候也会失效。

然而，这本书将有力地证明当前的数据管理正在不断地走向成熟。分析，即建立标准，进行报告，并预测和优化关键变量，是资源管理的主要功能。关于网络和数据分析的讨论已经持续多年，但直到现在，网络分析对于大多数公司来说也谈不上是主要业务。网络分析主要包括对独特访客和页面浏览量进行统计，这项工作通常由兼职人员来完成。

《打造大数据团队》给出了数据分析所必备的各种专业、严谨的方法。你需要的不仅仅是网络兼职人员，而是需要仔细思考你的标准和关键绩效指标（KPIs）。你需要做的事情也不仅仅是报告、预测、优化和严谨地测试。朱达·菲利普斯一直都是大数据分析相关学科的倡导者，不过如今，全世界都已经做好了接纳这些学科的准备，因此本书的出版适逢其时。

尽管有关网络数据分析的书籍不计其数，但我个人认为这本书在很多方面都有其独到之处。这本书涉及的内容比网络数据分析更为广泛，还包括社交媒体、移动终端、行为定位以及其他方面的数字信息。本书建议大多数公司将注意力放在更大规模的数据分析上，而不仅仅局限于关注网站的点击量。

这本书为数据分析领域带来了较为复杂的数据管理和数据分析，这在以往的网络数据分析中并不常见。在管理方面，这本书涉及的话题主要包括：如何提供数据分析功能、如何思考数据管控，以及如何处理公司中数据分析团队与其他类型分析人员之间的关系。数据管控可能吸引不到那些业余爱好者的关注，但它对于一家成熟企业的信息环境来说却是至关重要的。

朱达·菲利普斯在这本书中提出了很多在数据分析方面最为优秀和最具代表性的见解，我非常欣慰。我一直认为统计学家约翰·图基（John Tukey）的探索性数据分析（EDA）是一种非常有效的分析方法，该方法可以让你接近自己所研究的数据，并了解其基本参数，但近期相关的分析书籍中都很少提到这个分析方法了。所以，当我在本书中看到有专门介绍探索性数据分析的内容时，十分兴奋，因为它无疑是探索数字数据的伟大工具。

我十分期待未来我们能够拥有涵盖所有数据类型的分析系统，其中包括这本书中提到的数字类型，以及其他在商业分析功能中通常出现的客户、金融、运营等方面的数据类型。这本书是迈向数据整合的重要一步，与许多其他网络分析书籍不同的是，这本书并没有假定数据分析是唯一的分析类型，而是鼓励在商业分析中采用其他类似的原则和方法。正如我在写作当中会鼓励读者不要再单纯地报告，而是要转向预测分析和测试一样。因此，我非常高兴数据分析领域能有这样的融合趋势。

我推荐这本书是因为作者对自己所表达的内容十分清楚。朱达·菲利普斯曾在一些互联网公司（如 Monster.com 和潮流网店 Karmaloop）和线下公司（如诺基亚和励德·爱思唯尔集团）做过数据分析顾问或主管。如果你将书中所说的方法应用于你的公司，你将会远远超过其他人，这无疑会将你的职业生涯推向一个全新的高度。甚至，某一天你也许会穿上白色工作服，成为数据分析专家！

托马斯·H. 达文波特（Thomas H. Davenport）

哈佛商学院、美国巴布森学院教授

国际数据分析研究所创始人

中文版序

在世界各国的文化中，社区都扮演着重要的角色。在分析学领域，团队是通过组织创建的。最好的工作要通过团队的协作和共同努力，通过从分析和数据科学中创造出积极的业务成果来实现。多年来，我一直在为一些世界上最大的公司打造分析能力和组建分析团队，亲身体会到了这一事实。后来，我有了写一本关于数据分析的书的想法，于是就开始考虑有哪些主题和内容是适合的。当时，有许多数据分析相关的书，但内容都是偏战术性或者技术性的，更适合分析人员和技术人员参考。因此，我决定为商界领袖们写一本与众不同的、独一无二的数据分析书。你现在正在阅读的，正是我的工作成果。

当然，本书也包含了大量足以供分析专家和技术专家们参考的信息和知识，但面向的主要读者还是企业管理者。无论如何，我确实认为每个从事数据分析工作的分析师、数据科学家、技术专家和工程师都应该读一读这本书。因为我认为它解释了如何构建和运作一个专注于解答业务问题，帮助人们做出更好的决策，从而创造商业价值的数据分析团队。这是第一本，几乎也是唯一一本有关这一主题的书。

为什么分析学很有用？目前对数字数据的运用和分析情况如何？本书的讨论就从对这两个问题的研究开始。你知不知道地球上的数据字节比所有的沙粒都多呢？在本书中，我很快就会带你进入对“分析价值链”的回顾，这是我自创的一

个概念，和科特勒的“4P 营销”概念差不多，可以说是“分析学的 N 个 P”。这些概念能帮助读者理解那些位于“做分析工作”水平之上的实操性方法论。通过介绍一种能够成功地开展分析的分阶段方法，并以一些相关的宏观概念对其进行完善，读者可以立即上手并在组织中发挥实际影响。

要建立一个分析团队，你首先需要一家愿意对你和分析职能进行投资的公司。如何才能做到这一点呢？你要学习如何说明和证明对分析团队投资的“要求”是合理的，以及如何与同行和利益相关者一起跟踪验证本团队的投资回报率。当然，一部分投资是用于工具和技术的。你会在本书中发现一些对重要主题的论述，比如你是应该创建还是购买分析技术，以及你如何选择、部署、维护和停用你的工具。

对于那些从未上过统计学课，或者很久以前上过，现在需要“回炉”复习的人来说，本书涵盖了你们应该在数学或者统计学课堂上学习的许多内容，但重点还是放在商业用途上。我们会通过讲故事的方式，来理解分析和数据科学。毕竟，最好的分析学就应该是讲述有关数据的故事，并用这些数据回答业务问题，帮助人们做出决策。要成功地进行分析，你首先需要准确、优质的数据。要获得优质的数据，就必须定义和管理数据。在本书中，你将看到一些数字数据管理方法和管理数据以保持一致性和准确性的实例。

我会涵盖报告、分析和绩效分析——重点强调这些概念之间的差异以及用于分析交付成果的一些好方法。要记住报告而不是分析。分析是在报告数据之前要做的事情。你要将分析和报告放在管理仪表盘上。使用 KPIs 可以帮助一家公司引导数据，因此我提供了对数字分析有用的 KPIs 实例。

长期以来，我一直倡导对数字体验进行实验和测试，所以我介绍了 AB 测试和多变量 / 变量测试方法。2017 年，这些概念经常被应用到“转化率优化”（conversion rate optimization）的科学中，这样你就能了解到在“优化”中应用的基本技术。测试当然需要一个计划和过程，你需要借助技术来完成，书中也对此进行了讨论。

因为数据分析需要定量和定性数据，所以我努力描述和深入挖掘定性数据和客户之声（VOC）数据。和书中的所有主题一样，重点是团队如何成功合作。要做到这一点，了解定性和研究数据会有所帮助。你将读到对市场研究和定性数据

收集技术的综述。当然，这里也会包括竞争性分析，所以你将了解到竞争情报与数字情报之间的不同之处，而且我还会用实际案例来补充说明。本书也会讨论用于整合竞争性数字数据的工具、方法、流程和途径。

2017年，人们对目标定位的概念比5年前有了更好的理解。本书介绍了很多不同类型的目标定位的案例，以及使用数字数据进行定位和重新定位的经验 and 技巧——当然，还介绍了如何将这些数据统一到行动中。2017年的另一个流行语是“全渠道”。然而最初我就这个概念写一些文章时，听说过它的人并不多，尤其是在把它用于了解客户、受众和媒体的效果方面，所知者更是寥寥。如今，那些世界上最成功、最大的公司正大举投资于“全渠道”和目标定位。在本书中，你将了解到如何提高周转率，并提升“全渠道”数据和项目的成功概率。如果你以前从未听说过“全渠道”，那现在就能搞清楚这个术语的意思以及如何实现“全渠道”愿景。

在本书最后，我对分析学的未来做了预测。现在回顾这一章节的内容，我还认为这个做法是正确的。我的许多预期和设想全部或部分实现了。从预测性建议到闭环数据系统，到感知和响应的数据架构，到使用数据与客户和他人进行互动或给予提醒，再到目标定位、服务和交付的自动化，还有数据交互的购物和客户体验，这些目的已全部或部分实现了。最后，我以一个超越国家、文化和地理的概念——隐私和道德来结束这本书。在这个属于大数据、物联网以及实时和流动数据的收集与分析的美丽新世界中，我们必须问的一个问题是，所有这些都是如何影响人们的。我们该如何保护我们的数据以及连接到互联网的所有人的数据？关于数据隐私、安全和公平使用的对话促成了一个全球性社区的形成；分析师、技术专家和管理者也完全拥有发声的权力，能在对话中发表意见，制定规则和指导立法。

我想让大家都知道的是，我写本书的某个灵感也是属于马云的。我还从来没有去过中国——但我希望能早日去访问和观赏这个美丽而历史悠久的国家（手中拿着我的中文版著作）。我读过很多关于中国和中国历史的文章。我认为中文很优美，中国的历史和文化在持续地影响着我们共同的世界和全球经济。我非常荣幸我的第一本书能在中国出版。感谢你能阅读我的书。

前 言

多年前，我发现我遇到的大多数人并不清楚我到底从事什么职业，即使我把工作头衔告诉他们，他们也还是不太清楚。了解我到底从事哪种职业的人，都是在互联网行业里工作。我遇到的和我一样在做数据分析工作的人其实还是很少的。我曾一度认为，自己仅仅是一个为某个品牌工作、管理数据分析，并把数据分析同传统分析结合起来的数据分析工作者。当然，无论是品牌公司、代理商还是咨询公司，都会有分析团队，但很少有数据分析人员有机会管理涉及技术、人力、流程和整体分析成果的跨国公司的核心业务数据分析团队。其实仅仅在几年前，我所知道的创建全新数据分析系统或者是传承已有分析系统的人少之又少，这些分析团队通常只专注于理解数据行为，运用数据来驱动战略和战术决策。无论是在私企还是在上市公司，很少有人会进行数据分析，并向公司高管和利益相关者们进行汇报。因为在公司高管看来，数据是不能出现任何“错误”的，否则市场就会有所反应，而股东们对于数据会有很高的期望。这也使我意识到，在这个复杂且高度矩阵的环境里，组建数据分析型组织可能是一条康庄大道，也可能是一条羊肠小径。无论公司是第一次组建数据分析团队，还是从以往的尝试中获得了某些“经验”，都很有可能进入这两条相似却又有细微差别的道路中去。

我开始为品牌组建数据分析团队的时候，很少有人从事过类似工作，也无多少先例可供参考，所以我和同事们必须弄明白如何开展工作才能取得成功。在安

排数据分析活动和组建团队时，我的理念就是：我们的工作要专注于降低成本或增加收益。这样一来，如果数据分析能够影响到企业创收和成本控制，那么这个数据分析团队就能在公司获得稳固的地位。有时候的确如此，但也不排除某些外部因素带来的影响，如大萧条。

在从事数据分析的职业生涯中，我也逐渐对如何在组织中进行数据分析有了自己的见解。多年以来，很多人都问过我准备何时出书，正是因为受到了他们的鼓舞，我才终于下定决心写了这本书。这本书等于是为各行各业的数据分析师、管理者和企业各级主管准备的实用手册，帮助他们学习数据分析的组织因素，去了解和欣赏数据分析的过程、组建数据分析团队的必要性，以及运用严谨的数据分析技术和方法确保数据准确的重要性。另外，读者还可以获得额外的知识，学会欣赏报表、KPIs、数据管控，了解市场调查、定性数据以及其他类型的竞争性数据和商业智能数据，并且明白技术如何提升数据分析及数据分析决策水平。希望读者能够在本书中找到对自己有用的内容，并运用我的观点去调整和制定各自的决策——我曾在自己的职业生涯中运用过他人的观点。

写书绝非易事，它不仅要花费大量的时间和精力，还必须写出独到、客观、正确的内容。对数据分析专业人员来说，其内容又必须是相关且有实用价值的。本书就以上述要求为目的，甚至做得更为全面。这本书能够顺利完成，离不开多年来一起共事、合作以及如今在互联网数据分析领域有所成就的朋友们的帮助，在此一并致谢。

目 录

第 1 章 运用数据分析创造商业价值

大数据和数据科学需要数据分析 • 2

定义数据分析 • 6

第 2 章 分析价值链和数据分析中的 9P 理论

数据分析中 P 的解释 • 10

分析价值链：战略战术成功的流程 • 29

分析数据并对分析进行阐述和社会化 • 35

分析工作请求流程 • 40

第 3 章 创建分析团队

证明对分析团队的投资是合理的 • 52

没有可用投资 • 53

费用过高 • 54

投资分析的商业价值缺乏证据 • 54

克服由组织局限性和处理不当带来的投资阻力 • 57

- 创建分析团队的商业评估和投资建议书 • 58
- 向管理者和其他团队报告分析价值创造 • 59
- 设定分析团队目标 • 63
- 确定分析团队的角色和职责 • 66

第 4 章 何为数据分析工具

- 创建还是购买 • 78
- 平衡分析技术的管理：企业运营还是 IT 部门运营 • 85
- 选择分析工具 • 86
- 成功的工具部署 • 93

第 5 章 数据分析方法和技巧

- 观察数据：数据的形状 • 106
- 运用统计学和机器学习分析数字数据 • 114

第 6 章 数据的定义、规划、收集和管理

- 如何定义数据 • 127
- 什么是数据的商业定义 • 128
- 数据的操作性定义是什么 • 129
- 数据的技术定义是什么 • 129
- 创建和维护数据定义 • 130
- 规划数字数据：你应该做什么 • 132
- 收集你需要了解的数据 • 134
- 数据管理的作用 • 136
- 数据管理团队做什么 • 137
- 跨计划、工程项目和团队的数据管理流程 • 139
- 测试和核实数据的困难之处 • 140

第 7 章 数据报表及 KPIs 的运用

- 什么是报表以及它是如何生成的 • 147
- 一份优秀报表的五个要素: RASTA • 153
- 报表与仪表盘的区别 • 154
- 仪表盘是什么, 它是如何产生的 • 155
- 优秀仪表盘的五个要素: LIVES • 157
- 理解 KPIs • 158
- 报表与仪表盘如何适应分析价值链 • 163
- KPIs 实例 • 164
- 实时数据 vs 及时数据: 从业者的角度 • 170

第 8 章 用数据分析进行优化和测试

- AB 测试: 从此开始 • 176
- 进一步扩展到多变量测试 • 178
- 制订测试优化计划 • 178
- AB 测试和多变量测试的操作过程 • 180
- AB 测试和多变量测试结果的测量、分析与汇报技术和方法 • 182
- 测试支持的优化类型 • 183
- 建立数字优化项目 • 185
- 开发控制实验和数字数据科学 • 186
- 测试和优化数字体验的技巧 • 187

第 9 章 用户反馈数据定性分析

- 倾听顾客声音比以往任何时候都更加重要 • 193
- 交易工具: 市场调研及定性数据收集方法和技术 • 195
- 进行在线、电话或者邮件调查 • 196
- 建立线上和线下社区 • 197

- 面对面采访顾客 • 198
- 建立和利用焦点小组 • 198
- 建立顾客反馈体系 • 199
- 确定顾客的态度、信仰和动机 • 203
- 理解顾客要求和需求 • 203
- 追踪顾客体验 • 204
- 定性数据团队做什么，数据分析是如何起作用的 • 204
- 融合数字行为的数据和定性数据 • 206
- 成功的商业合作 • 208

第 10 章 竞争性情报和数字化分析

- 竞争情报 vs 数字情报 • 214
- 数字竞争情报实例 • 215
- 数字竞争情报的工具和方法 • 217
- 数字竞争情报流程 • 219
- 整合数字行为数据和竞争情报 • 220

第 11 章 定位目标和数据分析自动化

- 定位的类型 • 226
- 定位发生的数字领域 • 228
- 什么是重新定位 • 230
- 重新定位的类型 • 231
- 数据分析团队如何助力定位和重新定位 • 232
- 关于定位和重新定位的建议 • 233

第 12 章 融合全方位渠道，整合数据，以理解客户、用户和媒体

- 全渠道数据类型 • 238

全渠道数据度量	• 240
通过全渠道数据整合定义客户分析	• 241
用客户的数据和你的分析向客户提问	• 242
统一的客户生命周期	• 246
通过全渠道数据整合进行客户分析的工作	• 251
客户分析面临的挑战	• 252
通过全渠道整合进行客户分析时需要什么	• 253

第 13 章 数据分析的未来

预测的个性化	• 257
闭环的行为反馈系统	• 257
多屏统一发送实时可访问的相关内容和广告	• 258
感知和反应	• 259
相互作用和警示	• 259
特定地理位置和目标定位	• 260
自动化服务和产品配送	• 260
数据交互的购物者和顾客体验	• 260
未来的分析业需要注重隐私和道德	• 262

运用数据分析创造商业价值

如今，商业组织必须运用数据分析才能造就新的增量价值。数字体验涵盖了互联网、社交网络及移动应用等领域，是分析所用数据最显著和最重要的来源。因此，在当今商业世界中，只有充分了解了数据收集和数据分析的方式，才能生成全新或增量的价值，抑或降低成本，或者完成其中一项。这在今天的商界是至关重要的。

尽管当前全球市场竞争激烈，任何领域或行业均可使数据分析的利益最大化，然而创造和组建能够充分发挥数据分析功能的团队却并非易事，需要全方位考虑。数据分析人员、流程以及技术均需反思并重组。毕竟，许多公司仍然将数据分析等同于工具和技术（还有数据收集，类似于“监控”）。这种想法其实并不准确。

尽管支持数据分析的技术和工具至关重要且十分必要，但其本身不足以创造商业价值。简单将某个免费的网页分析工具的标准基础 JavaScript（一种程序语言）页面标记增加到数字体验中，进而提供报告所需的信息，这种方式并不能带来任何数字驱动型的决策，或者让我们轻易得出什么深刻的见解。有些公司认为，