

商业数据分析

——原理、方法与应用

**Business Analytics Principles, Concepts,
and Applications: What, Why, and How**

马克·J. 施尼德詹斯(Marc J. Schniederjans)

[美] 达拉·G. 施尼德詹斯(Dara G. Schniederjans) 著

克里斯多夫·M. 斯塔基(Christopher M. Starkey)

王忠玉 王天元 王 伟 译



机械工业出版社

针对商业数据分析的教学需求，本书提供了综合而完整的知识体系。本书介绍了最新的商业数据分析理念，阐述了什么是商业数据分析、为什么它如此有价值、如何应用它，并介绍了现在常用的重要统计技术、方法和软件工具。本书阐述了商业分析中的三种数据分析——描述性数据分析、预测性数据分析、规范性数据分析，并指导用户认识理解自己的数据分析项目。本书将基本概念与工具、技术及实际应用结合起来，通过实例，循序渐进地厘清那些企业或其他组织中可以用商业数据分析来解决的共性问题。事实已经证明，商业数据分析的原则、概念和应用提供了行之有效的理论和方法，成功地将它整合到组织当中，能帮助组织有效提升智能决策的竞争优势。

本书适合高校相关专业学生作为教材使用，也可供想了解大数据、数据分析如何帮助商业运营的管理者参考。

试读版本仅供防疫期间线上教学使用

推 荐 序

我们正处于一个变革的时代，面对大数据、人工智能（AI）的崛起所带来的巨大机遇与挑战。一方面，我们需要深刻洞察技术变革的趋势，树立大数据时代的商业思维，特别是看清对商业运营管理和商业模式所带来的冲击，并及时做出应对策略，在管理形式和方法上做出响应和调整；另一方面，在商业、经济及其他领域中，管理决策行为将日益基于数据和数据分析做出，而并非基于经验和直觉，这些正是管理学院或商学院在人才培养中面临的新挑战。

最近，国外一些知名大学的管理学院或商学院纷纷创建了商业数据分析（Business Analytics，又称为商业分析）硕士项目或专业。商业数据分析是融合了商业经济管理、统计学、计算机科学的新兴交叉学科。商业数据分析既注重经济、金融、管理知识的掌握和运用，又关注大数据分析方法及人工智能技术的学习和训练，目的是促使数据分析工作更具有预测潜力，从而提高公司业绩和战略管理能力，最终提高运营效率。简言之，商业数据分析搭建了数据分析与商业管理之间的桥梁，是利用各种分析工具来研究数据（包含大数据），寻找有用信息，进而更好地进行决策的一门新兴交叉学科。

商业数据分析着重培养学生综合运用专业知识解决理论与实际问题的能力。这些能力的获得，通常离不开管理学院或商学院规划的知识体系和实践能力培养，当然也离不开这一领域的优秀教材。

这本书由马克·J. 施耐德詹斯、达拉·G. 施耐德詹斯及克里斯多夫·M. 斯塔基合著，是商业数据分析领域中一本非常优秀的教材。全书以商业数据分析为主线，详细地讲述了商业数据分析三个基本部分——描述性数据分析、预测性数据分析、规范性数据分析的内容。全书结构清晰，观点明确，阐述详尽，语言简练。

正如作者在书中前言所说，“本书为商业数据分析的基础教学而编写，既可服务于大学教育，也可服务于实践者。除了为商业数据分析提供最新的文献和研究外，本书还利用简单的术语及在实践中十分有用的辅助方法，来解释商业数据分析流程。”

这本书具有以下三个特点：①紧跟当代大数据、人工智能变革的发展

前沿，密切追踪业界发展动态；②理论与应用实践更加紧密，在给出有关理论的阐述讲解之后，辅以案例应用，体现学以致用；③知识体系完整，在讲述商业数据分析内容时，突出逻辑主线，深入浅出，图文并茂，并将有关辅助工具的内容放在附录之中。

本书由机械工业出版社正式出版中译本，译者之一王忠玉老师曾翻译过多本经济管理方面的书籍，比如前不久出版的《管理数据分析——原理、方法、工具及实践》中译本（机械工业出版社，2017年出版）也聚焦于商业数据分析领域的，只是侧重点有所不同，将其作为补充阅读更为有益。

李一军

教授、博士生导师

“长江学者”特聘教授

于 哈尔滨工业大学经济与管理学院

译者的话

当今，大数据（Big Data）或数据（Data）正以爆炸方式涌现在社会经济生活的方方面面，各行各业都在产生和利用着数据。比如人们在网络上购物、通信、浏览新闻、收听在线音乐、搜索信息，或在网上表达观点，这些行为都会生成原始数据而被逐一记录。当前，人们不仅拥有大量的数据（包括大数据），而且计算机的运算能力也有了快速的提升，这些为利用数据从事商业研究，特别是从数据挖掘和提取有价值的信息，提供了必要的基本条件和良好的环境。

随着大量丰富多彩的数据源不断地涌现的新时代的来临，数据成为一种新的资源。关于如何从数据中挖掘和提取有价值的信息的例子，这里选取2012年沃尔玛依据飓风预报而安排备货的案例来说明。

在飓风 Frances 来临的一周之前，沃尔玛公司的首席信息官（CIO）Linda M. Dillman 督促她的团队，根据几周之前飓风 Charley 来袭期间沃尔玛的销售数据，对新飓风来袭时的销售进行预测。之前的海量销售数据被保存在数据仓库中，公司利用这些数据预测销售，目的是提前备货，进而提高公司的销售额。公司分析人员对那些销售数据进行挖掘分析，发现了一些不同寻常的需求特征。他们发现，人们确实更多地购买了某些特定的产品，但不是主观猜测的手电筒等。比如，他们发现，飓风到来前，草莓馅饼的销售量出现了增长，是平时销量的7倍左右，而最畅销的产品则是啤酒。据此，他们提前配备货物，极大地提高了公司的销售额。

当今，信息技术，尤其是互联网技术、计算机技术的不断进步和发展降低了人们获取数据、存储数据和传输数据的成本，使得越来越多的企业和政府机构等社会组织有能力从自身的业务系统，或者通过互联网等其他途径，获取规模日益庞大的数据。对于公司来说，数据的价值越来越高，人们更加重视对历史数据的积累，进行数据分析，挖掘和提取有价值的信息，以便高效、正确地认识和掌握事物的发展变动趋势，帮助决策者更好地进行决策。

要想认识和理解商业数据分析的理论及方法，一定离不开当前所处的时代背景，更离不开数据科学的基本理论和技术的支持。下面，首先简略介绍有关数据科学的基础知识，然后给出作为数据科学中一个特定的应用

领域——商业数据分析的有关内容。

1. 数据科学

什么是数据科学 (Data Science)? 依据维基百科的释义, 数据科学是一门利用数据来学习知识的学科, 其目标是从数据中提取出有价值的信息。它结合了诸多领域中的理论和技术, 包括数学、统计学、计算机科学, 特别是计算机科学领域的数据库、大数据、机器学习、数据挖掘、数据可视化等子领域的相关技术和方法。

从数据科学的内涵来看, 它是一个全新的概念, 试图把统计方法和数据分析方法统一起来, 目的是分析和理解客观现象产生的数据。数据科学的核心任务是从数据中抽取信息、发现知识。其研究对象是各种各样的数据及其特性。

为了认识和理解数据科学的内涵, 首先从数据概念开始介绍。如果从数据科学的研究对象来看, 数据是指利用测量工具所获得的未经处理的事实或数字。这些数据的类型多种多样, 不仅包括传统的结构化数据, 也包括网页、文本、图像、视频、语音等非结构化数据。这样定义的范畴既超出数学关于“数”的定义, 又扩展了统计学关于“数据”的定义。实际上, 比较常见的数据包括以下几种类型:

(1) 表格。这是经典的数据类型。在表格数据中, 通常行表示样本, 列表示特征。

(2) 点集 (Point Set)。许多数据被认为是某空间中的点的集合, 还有点云 (Point Cloud) 数据。

(3) 时间序列。文本、通话和 DNA 序列等都可以看成是时间序列。它们也是某种变量 (通常认为是时间) 的函数。

(4) 图像。可以看成是有两个变量的函数。

(5) 视频。可以看成以时间和空间为坐标的函数。

(6) 网页和报纸。虽然网页或报纸上的每篇文章都可以看成是时间序列, 但整个网页或报纸又具有空间结构。

(7) 网络数据。网络本质上是图, 由节点和联系节点的边构成。

(8) 模糊数据。本质上由各种模糊数构成。

除了前述几种基本数据类型之外, 还存在更高层次的数据, 例如图像集、时间序列集、表格序列等。

数据科学的理论与方法通常包括一组概念、原则、过程、技术/方法及工具, 为其核心任务服务。其中, 概念和基本原则给予人们观察问题、解决问题的一套完整的思想框架, 而大量的数据分析技术/方法和工具则帮助人们切

实实现数据科学的目标。一般来说，研究人员在进行数据分析时，要对数据的产生机制做某些基本假设，例如假设数据都是由某类相应的模型产生的。因此，数据分析的基本问题是找出这个模型，如表1所示。由于数据采集收集过程中不可避免会引入噪声，因此这些模型常常都是随机模型。

表1 常用数据类型及相应的数据模型

数据类型	相应的数据模型
点集	概率分布
时间序列	随机过程
图像	随机场
网络数据	图模型、贝叶斯模型
模糊数据	模糊概率分布

美国雪城大学 (Syracuse University) 信息学院的杰弗里·斯坦顿 (Jeffrey Stanton) 在《数据科学导论》(An Introduction to Data Science) 一书中提出：数据科学是一个新兴领域，它包括对海量数据信息的收集、预处理、分析、可视化、管理及保存等工作。

许多企业和机构之所以收集数据，原因是数据带来的价值或者潜在的价值超出了收集数据和管理数据的成本。通常，特定行业或领域的数据中会蕴含着规律，即数据中包含价值。

2. 数据科学的属性和应用过程

从数据科学的交叉特性来看，可将数据科学看成计算机科学、数学和统计学（含有现代数据分析）、特定领域的实务知识交叉融合而形成的新兴学科。相关的理论和技术源于不同学科的研究方向，相互之间存在比较大的差异，比如研究的基本假设等，而数据科学则试图在此基础上构建和谐自洽的理论体系。

德鲁·康韦 (Drew Conway) 利用韦氏图说明研究数据科学需要的技能，如图1所示，进而揭示出数据科学十分重要的属性——交叉性。可以看出，数据科学的特征之一是多学科交叉。

从实践应用来看，数据科学研究并不是某一个领域的专项问题，而是多学科综合性研究问题，应用时要具



图1 数据科学的多学科交叉属性

备三个条件：第一个条件是底层构架开发或使用能力；第二个条件是程序开发能力；第三个条件是数据建模和解决问题能力。

数据科学家需要什么样的能力呢？美国加州大学伯克利分校统计系的郁彬（Bin Yu）教授提出，一个合格的数据科学家应具备的基本素质和技能可概括为SDC³：

- Statistics (S)：统计学。
- Domain (Science) Knowledge (D)：坚实深厚的科学知识。
- Computing (C)：计算技术。
- Collaboration (Team Work) (C)：团队的合作能力。
- Communication (to Outsiders) (C)：与外界的沟通能力。

从现实来看，数据科学家收集数据、清洗数据、创建数据集、分析数据然后提出新观点，尝试用现有的数据预测未来，帮助提高产品、服务的质量和顾客黏性。更好的质量意味着更能取悦顾客、获得收益。因此，数据科学家最应该具备的三个特质是：①知道如何提出好问题；②理解所要处理数据的结构；③很好地解读这些数据。换句话说，数据科学家必须能够完成从数据到信息、知识、洞察力的过程，如图2所示。概括地讲，数据科学家就是提出合适的问题，然后提出有意义的见解来指导决策。

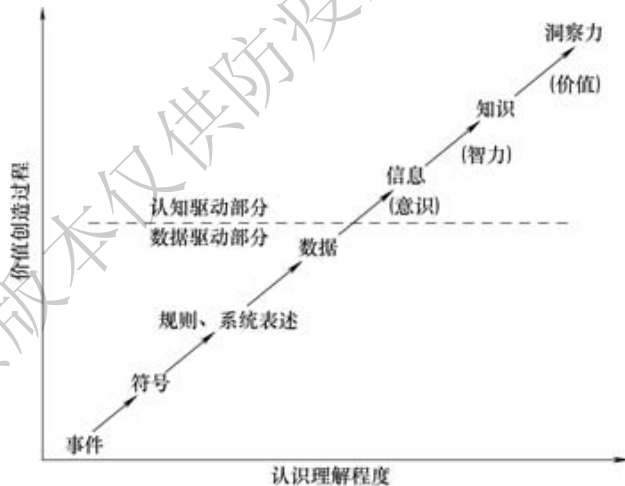


图2 从数据到信息、知识、洞察力的过程

从数据科学整个应用过程来看，不同阶段工作拥有各自的特定目标和探索工具，由不同的人来完成。例如数据获取与管理由首席数据官负责执行；数据存储与处理由数据构架师负责完成；而数据分析与建模由数据分

析师负责执行；当结合到某个特定领域时，特别是试图发现商业价值和启发并进一步应用时，则由商业数据分析师来执行完成。如图3所示。

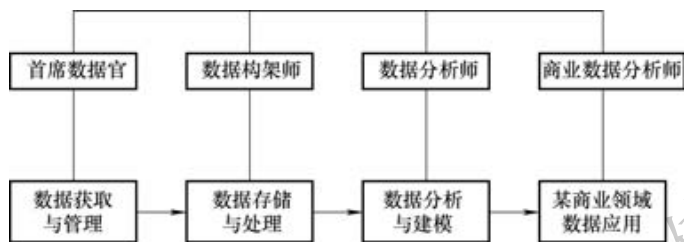


图3 数据应用中各阶段工作的目标任务和相应负责人

对于商业数据分析师和数据分析师来说，他们的工作尽管可能有交叉重叠之处，但是侧重点是不一样的。本质上，商业数据分析师探寻数据中隐藏的商业规律，在商业和IT之间作为中间沟通人，提高管理者的洞察力，进而改进业务流程，提高公司绩效。

3. 数据科学与商业数据分析专业

大数据时代，数据成为最重要的资源。2017年，中国网民规模达到7.72亿人，互联网普及率为55.8%，手机网民规模达到7.53亿人。2016年12月，工业和信息化部发布了《大数据产业发展规划（2016—2020年）》，这是为贯彻落实《中华人民共和国国民经济和社会发展第十三个五年规划纲要》和《促进大数据发展行动纲要》，加快实施国家大数据战略，推动大数据产业健康快速发展，在行动纲领上提出的指导规划。中国信息通信研究院发布的《中国大数据发展调查报告（2017年）》显示，2016年中国大数据市场规模达168亿元，增速达到45%，预计2017—2020年仍将保持30%以上的增长。

大数据的发展离不开物联网、云计算、人工智能等技术，离不开大数据相关人才。人才是产业发展的重要基石和支撑。在国外，伊利诺伊大学厄巴纳-香槟分校从2011年起举办“数据科学暑期研究班”；哥伦比亚大学从2013年起开设“应用数据科学”课程，并从2013年起开设相关培训项目，从2014年起设立硕士学位，2015年设立博士学位；纽约大学从2013年秋季起设立“数据科学”硕士学位。

加州大学伯克利分校在2016年春季学期的时候，学校首次开设的数据科学基础课程只吸引了300多名学生，而到2017年，已经有1000多名学生注册这门课程。学校一位数据科学教授表示：“在伯克利，从来没有哪门课程发展得如此之快。”

在中国，香港中文大学从2008年起设立“数据科学商业统计”硕士学位；复旦大学从2007年起开设数据科学讨论班，2010年开始招收数据科学博士研究生，并从2013年起开设“数据科学”课程；北京航空航天大学于2012年设立大数据工程硕士学位。

2017年4月，教育部公布高校新增专业——“数据科学与大数据技术”本科专业，有32所高等院校成为第二批成功申请“数据科学与大数据技术”本科新专业的高校。这样，加上此前第一批成功申请该专业的北京大学、对外经济贸易大学和中南大学，已有35所高校成功申请“数据科学与大数据技术”专业。

为了深刻认识大数据、人工智能等新技术对商业运营管理和商业模式所带来的冲击，及时做出应对策略，国内外一些知名大学的管理学院或商学院纷纷创建商业数据分析（Business Analytics，又称为商业分析）硕士项目或专业。比如，美国南加州大学的马歇尔商学院开设了商业数据分析的硕士项目，北京大学光华管理学院设立了商业分析硕士项目等，哈尔滨工业大学经济与管理学院创建了商务数据分析、商务智能辅修专业（学位）。

4. 关于 Business Analytics 的译法

如果针对 Analytics 以一种学术界常见的追根溯源方式来探究，那么通常可将 Analytics 理解成“n. 分析学，解析学，分析论”。

在大数据时代背景下，Analytics 是否被赋予了新的含义呢？当前，如果从学术角度来考察“Analytics”，一般是指 INFORMS 学术组织（www.informs.org）对其所赋予的新的含义：Analytics is defined as the scientific process of transforming data into insight for making better decisions。中文大意是：分析学被定义成将数据转变成用于更好地决策的洞察力的科学过程。可以发现，这个新定义中，Analytics 的对象是“数据”，对数据的加工过程则是一种科学的过程，最终目标是为人们的决策提供某种参考借鉴的依据，而且突出了“更好地决策”的高级目标。

国外有一本期刊《分析学》（Analytics），将 Analytics 定义为：利用数学、运筹学、统计影响商业决策。

而对于“数据分析”，我们特别欣赏和赞同麦克·沃森（Michael Watson）和德里克·内尔森（Derek Nelson）的观点。他们对“数据分析”所给出的定义为：“数据分析是利用数据获得洞察力，帮助人们更好地做决策的学科集合。它由下面三个分支组成：①描述性数据分析，提供数据描述，报告数据并对数据可视化处理；②预测性数据分析，利用数据预测趋势，识别数据关系；③规范性（指导性）数据分析，根据所拥有的已知数据，

依据未来所希冀的方向，为执行一系列的最佳决策提供指导方针。换句话说，描述性数据分析的目的是对已经发生什么和正在发生什么，提供一种认识理解；预测性数据分析的目的是告诉人们，将要可能发生什么；而规范性数据分析的目的是告诉人们如何应对。”（源自《管理数据分析——原理、方法、工具及实践》，中译本，机械工业出版社2017年出版）

鉴于以上介绍的 Analytics 和“数据分析”的学术含义，将 Analytics 翻译成“数据分析”就非常顺理成章。查询国内最近几年以来已经引进翻译的涉及 Analytics 方面的书籍，比如中译本《数据分析竞争法——企业赢之道》（2009年出版）、《精益数据分析》（2015年出版），以及前不久我们自己的译著《管理数据分析——原理、方法、工具及实践》，也是这样翻译的。本书是一本介绍数据分析方法在管理领域应用的初中级读本，通俗易懂、简明扼要，特别适合管理人员学习阅读，因此，我们将 Business Analytics 译成“商业数据分析”就十分自然了。

5. 本书的翻译和出版

本书的翻译和出版，是为了介绍如何将商业数据分析的理论及方法应用于各个行业，是一本学习和研究商业数据分析的基础教材。

本书第一译者曾于2013年10月参加哈尔滨工业大学经济与管理学院的“商务数据分析”新专业的课程建设小组，从那时起就一直关注和积累商业数据分析方面的知识。

本书的翻译和出版要特别感谢哈尔滨工业大学经济与管理学院的李一军教授、胡运权教授、于渤教授，以及学院院长叶强教授多年来给予译者的教诲和关怀。本书翻译工作具体分工如下：哈尔滨工业大学经济与管理学院王忠玉翻译第3、4、6、7、8章和附录B、附录C、附录D；黑龙江大学王天元翻译第5章和附录E、附录F、附录G；绥化学院王伟翻译第1、2章和附录A。最后，由王忠玉统稿。译者还要感谢曾经参与本书翻译和出版工作的其他老师和同学，他们是：哈尔滨工业大学经济与管理学院金融专业硕士研究生温雅欣，上海财经大学硕士研究生黄慧凡，哈尔滨工业大学经济与管理学院学生徐明泽、赵桓、吴春媛、张宇昂、王初旭、邢喆、王欢、王晓楠、高旭彤、张天爱、钟世超、彭娟，以及中央财经大学学生张丽波等。另外，在后期统稿时，哈尔滨工业大学经济与管理学院的陈政同学参与了部分工作。

虽然译者精心钻研和翻译，但仍可能存在纰漏和错误，希望广大读者指正。联系方式：wangzhy@hit.edu.cn。

王忠玉、王天元、王伟

作者介绍

马克·J. 施耐德詹斯 (Marc J. Schniederjans), 内布拉斯加大学林肯分校工商管理学院 C. Wheaton Battey 特聘教授, 在此之前曾经在其他三所大学任教。施耐德詹斯教授是决策科学研究所 (Decision Sciences Institute, DSI) 的研究员, 并在 2014—2015 年担任 DSI 主席。他曾经营过自己的货车租赁业务。目前, 他是供应管理学会 (ISM)、生产和运营管理学会 (POMS) 和 DSI 的成员。

施耐德詹斯教授主要讲授“运营管理和科学”, 并赢得了许多教学奖, 是金钥匙 (Golden Key) 荣誉会员和 Alpha Kappa Psi 商业兄弟会的荣誉会员。他发表了 100 多篇杂志文章, 写作出版了 20 本书。他最近的一本书名为《重塑供应链生命周期》, 广泛地研究了运营管理和决策科学。他还在学术会议上提交了 100 多篇研究论文。

施耐德詹斯教授不仅在五个期刊编辑委员会任职, 包括《计算机与运筹学》《国际信息与决策科学期刊》《国际服务业信息系统学报》《运营管理》《生产和运营管理学》, 而且担任《运营管理研究》杂志的区域编辑, 以及《国际战略决策科学》杂志和《国际社会系统科学与管理评论杂志: 国际杂志》(韩国) 的副主编。除此之外, 施耐德詹斯教授还担任多个商业和政府机构的顾问和培训师。

达拉·G. 施耐德詹斯 (Dara G. Schniederjans), 罗德岛大学工商管理学院供应链管理助理教授, 讲授“供应商关系管理和运营管理”等课程, 曾在《决策支持系统》《运筹学研究会》《商业过程管理》等期刊上发表过多篇文章。她曾与人合作编写过两本教材和一本大众读物, 也曾参加编写有关定量统计方法的读物。她担任过《国际社会系统科学杂志》社会科学商业伦理专题的客座顾问, 目前还担任 DSI 的网站协调员。

克里斯多夫·M. 斯塔基 (Christopher M. Starkey), 康涅狄格大学的经济学学生, 在管理和生产及运营管理学会会议上提交过论文; 他讲授“微观经济学原理”课程, 也曾经讲授过“宏观经济学原理”课程。他目前的研究兴趣包括宏观经济和货币政策, 以及决策方法。

前 言

我们每天都要面对大爆炸式的信息，我们尽力汇总整理信息，并运用信息来帮助决策，但有时候还是会被大量的数据所淹没。这可能导致我们得出错误的结论，同时做出错误的决定。例如一家全球性公司，在从世界各地收集数百万笔交易和客户行为数据时，单独就数据量而言，也会使得查找有用的客户信息这一任务几乎不可能完成。对于这家公司，以及更小的企业来说，解决方案是应用商业数据分析（Business Analytics, BA）。商业数据分析可以帮助人们整理大数据文件（称为“大数据”），在预测未来时，寻找有用的行为模式，并据此配置资源以实现最优化决策。商业数据分析是一个逐步深入的过程，帮助企业收集有用信息，以系统化程序来管理大数据，进而解决问题，找准提高公司绩效的机会。

本书为商业数据分析的基础教学而编写，既可服务于大学教育，也可服务于实践者。除了为商业数据分析提供最新的文献和研究外，本书利用简单的术语及其在实践中十分有用的辅助方法，来解释商业数据分析流程。就整体而言，本书所阐述的统计和定量工具中需要的数学知识不超过高中代数的水平。为了服务于读者需求，我们给出大量实例和案例，以训练读者运用普通的商业数据分析工具和软件。从业者可以发现，对商业数据分析方法进行研究，有益于考察所探讨的专题。大学师生可以发现，每章学习目标和对问题的讨论有助于满足他们的需求。

本书的目标是阐明什么是商业数据分析，为什么商业数据分析十分重要，以及如何应用商业数据分析。为实现这个目标，本书将介绍概念内容、相关软件和一些分析工具。

概念内容

本书的前面8章阐述有关的概念内容（参看第1章的1.4节关于本书内容的组织结构）。本书的概念内容远远超出商业数据分析的范围。本书将要解释：对于提供问题的解决方案来说，为什么商业数据分析是十分重要的，如何用它来实现竞争优势，以及如何协调组织更好地利用它。本书解释了在组织中应用商业数据分析时需要用到的管理知识，还有商业数据分析人员应该拥有的技能。本书还阐述了数据管理问题，例如数据收集、业务外包、数据质量及变革管理，因为这些内容和商业数据分析有关。

在建立管理基础来解释商业数据分析是什么和为什么重要之后，剩下各章则关注如何运用商业数据分析。为使这三个过程具体化，我们将商业数据分析解释为具有描述性、预测性以及规范性的分析步骤。对于每一个步骤，本书提供了一系列策略，以及商业数据分析流程的最佳实践指南。

相关软件

商业数据分析的绝大部分工作都要使用软件。令人遗憾的是，没有一个软件可以涵盖商业数据分析的所有方面。虽然有很多相关软件，但许多机构各自偏爱其中的一个。为了提供灵活性，本书介绍了多个软件，以供选择。在本书中，我们利用 SPSS®、Excel®和 Lingo®软件进行建模并解决问题。尽管本书提供了使用这些软件系统时的一些输入和指令，但主要内容是其输出。至于对运行软件不感兴趣的人，软件输出的资料也为他们提供了有价值的信息。对于那些在大学其他课程中会教授使用方法的基本软件，本书不再详细培训，建议对商业数据分析感兴趣的人要先了解这些基本软件的使用。

分析工具

本书附录对分析工具的内容进行了概述。商业数据分析是综合了统计、管理信息系统（MIS）和定量方法的学科。虽然书中的概念内容概括了如何进行商业数据分析流程，但如何在实际中应用商业数据分析则需要掌握定量化工具。因为一些实践者和大学课程对商业数据分析的技术方面不太感兴趣，所以本书只好在附录中给出这些定量内容。这些附录提供了用于支持各种分析的商业数据分析工具的大量解释。本书所解释和阐明的统计工具包括统计计数（排列、组合、重复）、概率方法（加法准则、乘法准则、贝叶斯定理）、概率分布（二项分布、泊松分布、正态分布、指数分布）、置信区间、抽样方法、一元回归与多元回归、绘图方法与假设检验。尽管管理信息系统已超出本书的范围，但本书仍然利用前面所提到的软件运用来阐明管理信息系统的搜索、聚类 and 典型的数据挖掘方法的应用。此外，本书中所解释和阐明的定量方法工具包括：线性规划、对偶性和灵敏度分析、整数规划、0-1 规划、预测建模、非线性最优化、模拟分析、盈亏平衡分析和决策理论（确定型、风险、不确定型决策、期望机会损失分析，完全信息期望值，不完全信息的价值）。

我们要感谢许多人的帮助，这些人为创作本书提供了所需的支持。首先，我们非常感谢我们的编辑 Jeanne Glasser Levine 和 Pearson 的优秀工作人员的支持。他们使得创作本书变成一种快乐，并与我们一起改进最终产品。和出版商合作编写书籍的几十年使我们认识到，顶级的出版商在图书出版

中发挥了非常大的作用。我们感谢 Alan McHugh，本书的封面（英文版）是他所设计的。他的不断探索精神和创新思想，对我们完成书稿做出了重要贡献。我们还要感谢 Jill Schniederjans 给予的大量编辑帮助。她减少了冗长的话语，增强了内容的趣味性。最后，我们要感谢 Miles Starkey 的帮助，他的存在和魅力提升了我们的精神状态，使我们保持在正确的方向上并按期完成书稿。

尽管许多人为本书写作提供了帮助，但其准确性和完整性由我们负责。对于本书可能存在的所有错误，我们事先表示歉意。

马克·J. 施耐德詹斯 (Marc J. Schniederjans)

达拉·G. 施耐德詹斯 (Dara G. Schniederjans)

克里斯多夫·M. 斯塔基 (Christopher M. Starkey)

试读版本仅供防疫期间线上教学使用

目 录

推荐序
译者的话
作者介绍
前言

第1部分 什么是商业数据分析

第1章 什么是商业数据分析 / 2

本章目标 / 2

1.1 术语 / 2

1.2 商业数据分析流程 / 5

1.3 商业数据分析流程与组织决策过程的关系 / 8

1.4 本书内容的组织结构 / 9

总结 / 10

问题讨论 / 10

参考文献 / 11

第2部分 为什么商业数据分析是十分重要的

第2章 商业数据分析十分重要的原因 / 14

本章目标 / 14

2.1 简介 / 14

2.2 为什么商业数据分析十分重要：提供解决问题的方案 / 15

2.3 为什么商业数据分析十分重要：提供战略竞争优势 / 17

2.4 商业数据分析十分重要的其他原因 / 18

总结 / 21

问题讨论 / 21

参考文献 / 21

第3章 哪些资源对支持商业数据分析至关重要 / 23

本章目标 / 23

第3部分 如何应用商业数据分析

- 3.1 简介 / 23
- 3.2 商业数据分析人员 / 23
- 3.3 商业数据分析中的数据 / 27
- 3.4 商业数据分析技术 / 29
- 总结 / 33
- 问题讨论 / 33
- 参考文献 / 34

第4章 如何整合组织资源支持 商业数据分析 / 36

- 本章目标 / 36
- 4.1 整合商业数据分析的组织结构 / 36
- 4.2 管理问题 / 43
- 总结 / 48
- 问题讨论 / 49
- 参考文献 / 49

第5章 什么是描述性数据分析 / 50

- 本章目标 / 50
- 5.1 简介 / 50
- 5.2 数据可视化与数据探索 / 51
- 5.3 描述性统计学 / 53
- 5.4 抽样与估计 / 57
- 5.5 概率分布简介 / 61
- 5.6 市场营销/策划案例：BA 过程中的
描述性数据分析 / 63
- 总结 / 70
- 问题讨论 / 70
- 习题 / 71

第6章 什么是预测性数据分析 / 72

- 本章目标 / 72
- 6.1 简介 / 72
- 6.2 预测模型 / 73
- 6.3 数据挖掘 / 75
- 6.4 市场营销/策划案例续：BA 过程中的
预测性数据分析 / 79
- 总结 / 89
- 问题讨论 / 89

习题 / 89

参考文献 / 91

第7章 什么是规范性数据分析 / 92

本章目标 / 92

7.1 简介 / 92

7.2 规范性数据分析模型 / 93

7.3 非线性最优化 / 94

7.4 市场营销/策划案例续：BA 过程中的
规范性数据分析 / 100

总结 / 104

补充内容 / 104

问题讨论 / 104

习题 / 105

参考文献 / 106

第8章 商业数据分析案例研究 / 107

本章目标 / 107

8.1 简介 / 107

8.2 案例研究：问题背景和数据 / 108

8.3 描述性数据分析 / 108

8.4 预测性数据分析 / 114

8.5 规范性数据分析 / 120

总结 / 125

问题讨论 / 125

习题 / 126

第4部分 附录

附录 A 统计工具 / 128

附录 B 线性规划 / 153

附录 C 线性规划的对偶性与灵敏度
分析 / 185

附录 D 整数规划 / 203

附录 E 预测 / 210

附录 F 模拟 / 228

附录 G 决策理论 / 235

试读版本仅供防务学院线上教学使用

第 1 部分

什么是商业数据分析

第 1 章 什么是商业数据分析

第 1 章 什么是商业数据分析



本章目标

- 给出商业数据分析的定义
- 解释分析、商业智能同商业数据分析的关系
- 阐述商业数据分析流程的三个步骤
- 阐述数据的四种测量尺度分类
- 解释商业数据分析流程与组织决策过程的关系

1.1 术语

商业数据分析^①始于数据集（简单的数据集或者数据文件），或者数据库（包含人物、地点等信息的多个文件数据集）。随着数据库规模的不断扩大，数据需要合理地存储。计算机云（用于数据的远程存储、检索和计算的硬件和软件的统称）及数据仓库（用于报告和数据分析的多个数据库）可用于存储数据。随着数据库存储区越来越大，一个新的术语应运而生，即**大数据**（Big Data）。大数据描述了那些太大、太复杂，以至于软件系统几乎无法处理的数据集的集合（Isson 和 Harriott, 2013, 第 57 ~ 61 页）。Isson 和 Harriott（2013, 第 61 页）定义了**小数据**：大数据以外的数据。例如，那些帮助个体企业记录客户资料的较小的文件即可视为小数据。作为一种整理数据以寻找有用信息的手段，分析有了崭新的用武之地。

在商业文献中，存在通常彼此关联的三个术语：分析、商业数据分析和商业智能。**分析**可以被定义为涉及使用统计技术（集中趋势测量、图表等）、信息系统软件（数据挖掘技术、排序程序）和运筹学方法（如线性规划）的一种用于探索数据、数据可视化、发现数据传递模式或趋势的过程。

① Business Analytics, 简称 BA, 又译为商业数据分析学, 实际上 analytics 的本意为分析学, 鉴于国内大多数译者将 Business Analytics 译成“商业数据分析”, 因此本书也采用这个通行的译法。——译者注

简言之，分析将数据转变为有用的信息。“分析”是一个较为陈旧的术语，它可应用于所有学科，而不仅仅适用于商业领域。一个典型例子是：对气象数据进行收集，并将其转化成为统计资料，进而用于预报天气。

分析有许多不同的类型，这里有必要给予分类，以便认识、理解其用途。我们将其分为三种类型：描述性数据分析、预测性数据分析和规范性数据分析，依据的是国际运筹学与管理科学学会（INFORMS）关于分析分类的建议（参见表 1.1）。这三种类型的分析中，每一种都可以视为是独立的。例如，有些公司仅仅使用描述性数据分析，为其所面临的决策提供信息；而其他一些公司，则可能使用多种类型的分析方法来搜集有价值的信息，帮助规划和决策。

表 1.1 分析的类型

分析的类型	定义
描述性数据分析	运用简单的统计技术来描述数据集或者数据库中包含了什么，比如制作描述顾客年龄的条形图，用于百货公司按年龄定位目标客户群
预测性数据分析	运用高等统计技术、信息软件或运筹学方法来识别预测变量，并建立预测模型来识别那种用描述性数据分析无法观测的趋势和关系。比如将多元回归用于揭示年龄、体重和锻炼对减肥食物销售的影响（或显示没有关系）。知晓某种关系的存在，有助于理解为什么一组自变量的集合会影响诸如公司绩效这个因变量
规范性数据分析 (也称指导性数据分析)	运用决策科学、管理科学、运筹学（应用数学技术）来对资源进行最优配置。比如百货公司针对目标客户的广告预算有限，运用线性规划模型来最优地分配不同广告媒体的预算

对于三种类型的分析来说，它们的目标和方法各不相同，如表 1.2 所示。这些不同之处将分析和商业数据分析截然区分开来。分析聚焦于从数据来源产生有价值的信息，而商业数据分析则更进一步，利用分析来改进公司绩效。分析的过程可以包含分析三种类型中的任何一种，而商业数据分析则主要综合三种类型的分析产生新的、独特的、有价值的信息，以此帮助商业组织决策。另外，这三种类型的分析需要依顺序使用（首先是描述性数据分析，其次是预测性数据分析，最后是规范性数据分析）。因此，商业数据分析可以被定义为如下过程：开始于商业数据收集，然后依次应用描述性数据分析、预测性数据分析和规范性数据分析，目的是支撑和论证商业决策过程及组织绩效。Stubbs（2011，第 11 页）认为，商业数据分析优于简单分析，这有赖于以下三个方面：与商业的关联清晰；所获得的结果是可执行的；绩效和价值可测量。

表 1.2 分析目标与方法

分析类型	目标	方法举例
描述性数据分析	识别大数据集或数据库中可能存在的趋势。目的是对数据大致情况和识别趋势，或对未来商业行为潜力的准则有初步了解	描述性统计学，包括集中趋势指标（均值、中位数、众数）、离散趋势指标（标准差）、表格、图、排序法、频数分布、概率分布和抽样法
预测性数据分析	建立预测模型，识别与预测未来趋势	统计方法，如多元回归与方差分析 信息系统方法，如数据挖掘和排序 运筹学方法，如预测模型
规范性数据分析	最优配置资源，以便充分利用预测出来的趋势或未来的机会	运筹学方法，如线性规划与决策论

商业智能（BI）被定义为基于商业目的将数据转化成有意义和有用信息的一系列过程和技术。有些人认为，商业智能是一个包含分析、商业数据分析及信息系统在内的宽泛学科（Bartlett, 2013, 第4页），也有人认为，商业智能主要聚焦于收集、存储和探索大型的数据库组织结构，并据此获得对决策和规划有用的信息（Negash, 2004）。一般来讲，商业智能的一个主要部分是将数据存储于计算机云存储器或数据仓库。尽管这些数据可以用于分析，但数据仓库本身不属于分析或者商业数据分析。在实际应用中，商业智能聚焦于查询和生成报告，但在报告中有商业数据分析的内容。商业智能寻求解答的问题如：“现在正在发生什么”“在哪里发生”及“依据以往经验需要什么商业措施”。商业数据分析则从另一个角度，回答像是“为什么发生”“有哪些潜在的新趋势”“将要发生什么”及“对未来的最佳策略是什么”这些问题。

总的来看，商业数据分析包含了传统分析的过程，同时对分析出的结果又有额外的要求，即必须对公司绩效有着可测量的影响。商业数据分析包含类似商业智能的报告，但和商业智能仅仅报告和存储这些结果相比，商业数据分析则侧重于通过分析来对结果进行解释。本书会经常提到分析、商业数据分析以及商业智能。表 1.3 中所列的特点有助于认识它们之间的差异。

表 1.3 分析、商业数据分析和商业智能的特点

特点	分析	商业数据分析	商业智能
公司绩效规划作用	正在发生什么，将要发生什么	当下正在发生什么，将要发生什么，最佳处理策略是什么	当下正在发生什么，为处理它我们已经做了哪些事
使用描述性分析作为分析的主要部分	是	是	是

(续)

特 点	分 析	商业数据分析	商业智能
使用预测性分析作为分析的主要部分	是	是	否 (仅过去)
使用规范性分析作为分析的主要部分	是	是	否 (仅过去)
混合使用以上三者	否	是	否
聚焦于商业	可能	是	是
聚焦于存储和维护数据	否	否	是
要求聚焦于提升公司价值和绩效	否	是	否

1.2 商业数据分析流程

完整的商业数据分析流程包含三个主要步骤 (参见图 1.1)。商业数据分析流程的结果必定与商业活动有关,并且在一定程度上有助于提升公司绩效。

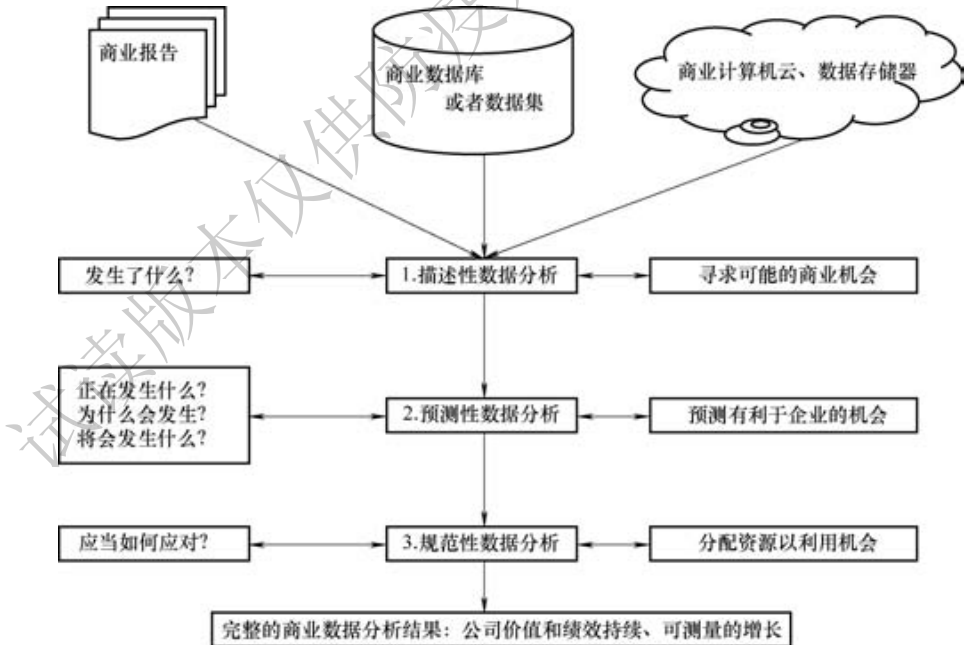


图 1.1 商业数据分析流程

图 1.1 所表示的商业数据分析流程的逻辑基于这个问题：企业所拥有的数据蕴含着哪些有价值或者有助于解决问题的信息？如图 1.1 所示，组成商业数据分析的每个步骤都有额外的问题需要解答。回答这些问题，需要通过组成商业数据分析流程的三个步骤，来挖掘数据中有价值的信息。商业数据分析流程就像是挖矿，找到新的、独特的、有价值的信息从而走向成功，直观地说，如同在矿井中找到金子。SAS 是分析领域中的一家重要的公司（www.sas.com），他们提出了问询钻取（Query Drilldown）的概念，是指挖掘问题和找出答案，从而提炼出有用的信息。许多公司通常采用商业数据分析来解决特定的问题，而其他一些公司则利用商业数据分析探索和发现新知识，以引导企业的规划和决策，从而提高公司绩效。

一些数据源可能难以管理，过于复杂，通常令人困惑。整理数据与试图了解它的信息价值，这需要应用商业数据分析流程的第一步：描述性数据分析。一开始只是简单地将数据分组，可采用表 1.4 列出的四种分类法。同时，将数据合并入电子表格，如 Excel，准备交叉制表和列联表也是将数据限定在一个易于管理的数据结构中的手段。利用集中趋势和离散程度的简单测量也会有利于捕获到提升公司绩效的潜在机会。其他的描述性分析归纳方法，诸如表、曲线图和图形，都能够帮助决策者实现数据的可视化，更好地理解潜在的机会。

经过步骤 1 的描述性数据分析，人们可以识别一些代表商业机会或未来可能（但尚未被定义）的趋势的模式或因素。但要解释数据中发生了什么（即过去发生了什么），还需要做更多的努力（需要更多的挖掘），比如做聚焦于商机目标的详细统计报告。企业基于大量数据对预测性变量进行统计学搜寻，发现的行为模式如果将来发生，企业就会加以利用。例如，企业可能在综合销售信息里发现，在经济停滞时期，如果采取某些广告策略，某一特定收入水平的客户群就会购买某些商品。这些销售额、顾客群和广告变量可能是表 1.4 中数据的任意数量化形式，但它们必须满足前文提到的商业数据分析的三个条件：与商业的关联清晰；所获得的结果是可执行的；绩效和价值可测量。

表 1.4 数据的测量尺度分类

数据测量尺度分类	解 释
分类数据	数据可根据一个或多个特征进行分组。分类数据通常包含基数或百分比 例：产品市场可基于销售价格划分为“高端”或“低端”市场 这个术语常用于包含分类数据的数据集以及交叉表、列联表所概括的观察值

(续)

数据测量尺度分类	解 释
有序数据	将数据排列或排序，用以显示相关偏好 例1：足球队排名不是基于所得分数而是胜场 例2：基于产品质量的企业排名
等距数据（区间数据）	按照等比例尺度排列数据，数据中每个值和其他值的距离都是相同的。它是有序数据 例1：A 温度表 例2：标度为李克特量表式的测量仪器（即1, 2, 3, 4, 5, 6, 7），1和2、2和3被视为是等距的，以此类推 注意，在有序数据中，企业排名中第一名和第二名的差异可以很大。而在等距数据中，它们必须是成比例的关系
比率数据（比值数据）	将数据表示成连续尺度上的比率 例：具备绿色制造项目的企业比率是不具备该项目企业比率的2倍

为了确定在步骤1描述性数据分析中所发现的趋势和行为是否真的存在或有效，以及能否被应用于预测未来，人们在步骤2中会有更进一步的分析，即商业数据分析流程中的预测性数据分析。商业数据分析流程的这一过程中有多种方法可供采用。一种常用的方法是多元回归（关于多元回归和方差分析的讨论，参见附录A和附录E）。无论在描述性数据分析中发现的预测性变量间是否存在统计学关系，这个方法都是理想的建模方法。其发现的关系可能是因变量可预测地同公司价值或某种绩效相联系。例如，某企业可能想从几种促销手段（可测量的自变量，在模型中由电视广告、无线广播、个人推销或杂志的费用表示）中寻找一种产生销售额（因变量，公司绩效的一种测量）的最有效的方法。我们必须确保，多元回归模型的应用是有效和可靠的，这也是我们使用方差分析和其他验证性分析来辅助建模的原因。使用高级统计程序探索数据库对最佳预测变量进行验证和确认，是商业数据分析流程在这一步骤的重要部分。这会告诉正在发生什么，以及它为什么会发生在模型的变量间。

一元或多元回归模型通常可以预测未来的趋势线。当回归不符合实际时，我们可使用其他预测模型（指数平滑、平滑平均）做预测性数据分析，来生成所需预测（参见附录E）。识别出的未来趋势是步骤2的主要成果，而预测性数据分析正是用来发现它们的。这有助于告诉我们将要发生什么。

如果企业通过商业数据分析流程的步骤2预测了趋势，并了解到未来的情景，那么就能把握住未来潜在的机会。在步骤3规范性数据分析中，运筹学方法可被用于最优化分配一个企业有限的资源，从而利用所预测出的未

来趋势中的机会。人力、技术和资金上的限制使得任何一家公司都无法在同一时间利用所有的机会。规范性数据分析能够让企业合理分配有限的资源，以尽可能完全地实现目标。例如，线性规划（一种约束优化方法）已被用于供应链设计，来使利润最大化（Paksoy 等，2013）。（线性规划和其他优化方法，参见附录 B、附录 C、附录 D）。商业数据分析流程的步骤 3 解决了“未来怎样做分配和决策才是最优的”的问题。

总之，商业数据分析流程中这三个主要部分——描述性数据分析、预测性数据分析及规范性数据分析，能帮助企业发现机会、预测趋势——从而预测未来的机会，还能帮助企业选择行动路线，目的是最优化分配企业资源，从而使公司价值和绩效最大化。第 5~7 章将详细介绍商业数据分析流程以及各种方法。

1.3 商业数据分析流程与组织决策过程的关系

商业数据分析能够解决问题，并识别出改善公司绩效的机会。在这个过程中，组织也可以决定指导行动的策略，确立竞争优势。解决问题和识别所遵循的战略机会是一个组织典型的决策任务。后者，机会的识别，可以视为需要解决的战略选择问题。可以发现，在 1.2 节所描述的商业数据分析流程和经典的组织决策过程有着密切的相似之处，这一点并不令人意外。如图 1.2 所示，商业数据分析流程与经典的组织决策过程有着内在的关联。

组织决策过程（ODMP）由 Elbing（1970）提出。如图 1.2 所示，它关注制定决策，为的就是解决问题。但同时也能用于从数据中寻找机会，以及决定采取什么样的行动以充分利用它们。组织决策过程五个步骤是从察觉不均衡或意识到有需要决策的问题开始的。类似地，商业数据分析流程的第一步，就是辨别出数据库中可能存在的有助于解决问题和发现机会的信息，从而来改善公司绩效。而在组织决策过程第二步，也就是探索问题，需要确定问题的规模、影响，同时考察是否有其他因素，为的是诊断问题到底是什么。同样地，商业数据分析流程中描述性数据分析也寻求那些可能有助于解决问题和发现机会的因素。组织决策过程中描述问题步骤和商业数据分析流程中预测性数据分析在发现策略、路径或者趋势上都有着类似结构，这些发现都清晰地定义了组织要去解决的问题。最后，在组织决策过程的最后一步策略选择和实施时，同商业数据分析流程最后的规范性数据分析包含的任务类型如出一辙（做出最优的资源分配，并加以执行）。

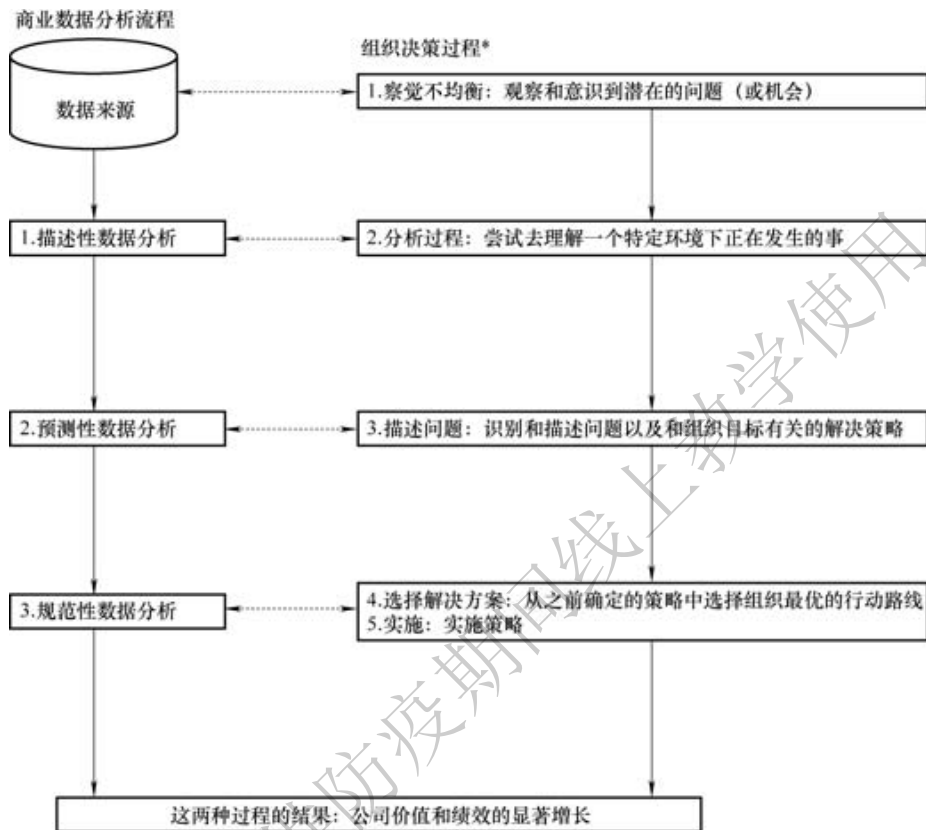


图1.2 商业数据分析流程与组织决策过程的比较

(来源: 改编自 Elbing (1970), 第12~13页表1。)

组织决策过程中一直蕴含的决策功能和商业数据分析流程是极其相似的, 二者采用相同的逻辑, 都可以提升组织的决策技巧和能力。

1.4 本书内容的组织结构

本书旨在回答关于商业数据分析的三个问题:

- 什么是商业数据分析?
- 为什么商业数据分析是十分重要的?
- 如何应用商业数据分析?

为了回答这三个问题, 本书相应地分为三部分。第1部分“什么是商业数据分析”, 包含第1章, 回答“是什么”的问题。第2部分中“为什

么”的问题由第2章和第3章给予解答。

如何应用商业数据分析非常重要，因此本书余下几章和附录都将围绕这个问题展开。第4章“如何整合组织资源支持商业数据分析”，解释了组织需要如何支持商业数据分析。第5章“什么是描述性数据分析”、第6章“什么是预测性数据分析”、第7章“什么是规范性数据分析”，分别说明商业数据分析流程的三个步骤，进一步阐明如何应用商业数据分析。第8章“商业数据分析案例研究”，提供了商业数据分析的具体例子。第8章之后是面向分析的一系列附录内容，以供补充分析和讨论。

第3部分涵盖了利用计算机软件进行定量分析的相关内容。为体现应用软件的多样性，书中使用了SPSS、Excel以及LINGO。既可使用SPSS和LINGO来复现书中的分析案例，也可单独利用Excel及其加载项。鉴于软件的不断变化和教育背景的差异，本书不再提供更多的关于软件方面的解释。

某些章除基本内容之外，还有帮助强化学习的内容。每一章开始都给出本章学习目标，章尾则有总结、问题讨论，根据需要还会列出参考文献。此外，第5~8章介绍了一些问题的案例和解决方案，当然也有额外的习题。

几个附录给出了关于方法论更详细的一些说明。将它们安排在附录中是为了使内容更流畅，同时为有经验的读者提供一种弹性方式，使他们可以仅选择想使用的技术性内容来阅读。

总结

这一章介绍一些重要术语，将商业数据分析定义为如下过程：此流程有助于确保有用的信息用于制定商业决策和抓住商机。我们也介绍了数据的测量尺度分类，来帮助理解商业数据分析中可能采用的各种测量；同时解释了商业数据分析流程和组织决策过程的关系——这两者是相互补充的。本章最后介绍了全书的内容组织结构，说明了这些结构如何安排以利于学习。

认识知道商业数据分析是什么，这一点特别重要，知道它为什么重要也是同样重要的。第2章就将回答后面这个问题。

问题讨论

1. 分析和商业数据分析的区别是什么？
2. 商业数据分析和商业智能的区别是什么？
3. 为什么商业数据分析流程有着特定的顺序？

4. 商业数据分析流程和组织决策过程如何相似?
5. 为什么等距数据必须是成比例关系的?

参考文献

Bartlett, R. (2013) A Practitioner's Guide to Business Analytics. McGraw-Hill, New York, NY.

Elbing, A. O. (1970) Behavioral Decisions in Organizations. Scott Foresman and Company, Glenview, IL.

Isson, J. P. , Harriott, J. S. (2013) Win with Advanced Business Analytics. John Wiley & Sons, Hoboken, NJ.

Negash, S. (2004) "Business Intelligence." Communications of the Association of Information Systems. Vol. 13, pp. 177-195.

Paksoy, T. , Ozkeylan, E. , Weber, G. W. (2013) "Profit-Oriented Supply Chain Network Optimization." Central European Journal of Operations Research. Vol. 21, No. 2, pp. 455-478.

Stubbs, E. (2011) The Value of Business Analytics. John Wiley & Sons, Hoboken, NJ.

试读版本仅供防疫期间线上教学使用

试读版本仅供防疫期间线上教学使用

第 2 部分

为什么商业数据分析是十分重要的

第 2 章 商业数据分析十分重要的原因

第 3 章 哪些资源对支持商业数据分析至关重要

第 2 章 商业数据分析十分重要的原因



本章目标

- 阐述在解决商业问题方面，商业数据分析为什么是重要的
- 解释商业数据分析为什么在识别新的业务计划中是重要的
- 描述商业数据分析有助于解决哪些类型的问题
- 解释商业数据分析如何帮助一个组织取得竞争优势
- 解释各种不同类型的竞争优势及其与商业数据分析的关系
- 解释对企业来说商业数据分析的重要性

2.1 简介

通信和信息系统正在以难以置信的速度和包罗万象的方式收集着人们生活中方方面面的数据。此外，企业要为意见调查、日常经营而收集各种各样的数据。云信息系统可以提供大量容易获得的数据，数据仓库系统则有能力将大数据存储在大型数据库中，当前需要从数据中获取信息，从而判断是否进行数据投资。正如 Demirkan 和 Delen (2013) 所证明的，将大数据放入云计算当中，可以及时、便捷地进行商业数据分析。企业认识到，要保有竞争力，就一定需要这种信息，而商业数据分析正是获取这些信息的一种途径。

大数据文件和小数据文件存在的问题，是它们可以很容易掩盖所需的信息。有时，文件里数据的细小变动都会导致含义的变化。20 世纪 60 年代的电视剧《密谍》[⊖]里经常出现“I want information”，当我们平时用于书面

⊖ *The Prisoner*，是于 1967 年 10 月 1 日至 1968 年 2 月 4 日期间，在英国 ITV 电视台播出的电视系列剧。剧中身为英国谍报员的主角，某天决定辞职。当他回家收拾行李时被迷晕绑架到一处不明之地，岛上有许多和他一样的“囚犯”，每个人的真实身份被隐匿，并且以编号来称呼。被赋予“6 号”这个编号的主角，每集都会遇到村中的新领导者“2 号”前来质问他“辞职的理由”与“所知的情报”，但主角却坚拒回答。“6 号”屡次尝试逃出“村子”，却总是遭到“2 号”阻拦，并以各种手段隐藏幕后的秘密。——译者注

或口语当中，这句话是指有人想要获得情报。但在《密谍》中，它却意味着“in”和“formation”，即“I want in formation”，用于让囚犯听话并和他人做一样的事情。单词中的一个空格即可改变本来的含义。细小差异就可以造成含义的变化，这对于想要在大数据中寻求相关商业信息的企业来讲是一个挑战。商业数据分析便应运而生。

2.2 为什么商业数据分析十分重要：提供解决问题的方案

虽然这么说可能有些自夸，但商业数据分析是未来解决重要商业问题的最佳手段。在商业数据分析三个步骤中的每一步（由第1章可以知道），对一系列问题的解答都能够并且应该被视为分析的逻辑结果。商业数据分析为什么成为有价值的决策工具？为什么学习和使用它是十分重要的？对这些问题的回答就给出了理解这些问题的信息和知识基础。

如表2.1所示，典型商业数据分析可以回答的一些问题，不仅和商业数据分析中的每一步息息相关，而且和时间有所关联。我们举个例子来更好地理解商业数据分析所给出信息的重要价值，说明它为什么对企业来讲是至关重要的。

表 2.1 商业数据分析所要回答的问题

商业数据分析的三个步骤	时 间 段		
	过 去	现 在	未 来
描述性数据分析	过去发生了什么	基于过去，现在发生了什么	基于过去，未来将要发生什么
预测性数据分析	过去它是怎样发生的 为什么它发生在过去	数据中可能存在什么趋势 现在应采取什么行动	如果考虑当前趋势和预测，未来可能发生结果的范围和可能性是什么
规范性数据分析	我们怎样最好地利用从趋势和预测中得到的信息	我们怎样优化利用资源来使未来的公司绩效最大化	在未来，怎样才能不断地应用商业数据分析来优化公司绩效

（来源：Isson 和 Harriott（2013）第 169 页。）

例如，一个信用合作社为房主贷款提供了一系列的套餐服务，这些促销宣传会定期地出现在各种媒介（纸媒、电视、收音机）上，目的是为了吸引新客户选择这些套餐。在营销过程中，信用合作社并不知道所产生的交易是由于促销活动，还是他们正常商业活动周期的结果。为了搞明白这

一点，信用合作社进行了商业数据分析。商业数据分析中所得到的信息已由表 2.2 给出（基于表 2.1 的问题）。首先，按照过去→现在→将来的顺序阅读描述性数据分析一行，再按照相同的方式阅读下面两个分析步骤。这个例子展示了可以从商业分析问题的求解中获得的信息类型。

表 2.2 对信用合作社例子所进行的商业数据分析

商业数据分析 的三个步骤	时 间 段		
	过 去	现 在	将 来
描述性数据分析	图形结果显示，过去广告推销活动导致了新贷款的适度增加	基于对贷款活动的排序，新房主在新贷款申请时刚好经历了贷款的适度增加	基于迄今为止贷款的直方图，没有明显的模式，只是随着时间推移，新贷款销售均匀为常数。没有观察到商业周期改变了贷款模式的影响
预测性数据分析	统计学上表明过去新贷款数量的增加和营销活动存在关联，但为什么产生新的贷款却取决于营销的方式	利用多元回归，此模型预测资金分配给电视广告与平面广告，将比投资于广播更为有效	利用回归模型中的方差统计，可以给出因促销的投资变化而产生新贷款的置信区间
规范性数据分析	将市场预算资金更多地投入到电视和纸媒的广告上，可以更有效地接触到目标受众	已知资金受限资源，用线性规划模型来配置营销预算，使得新贷款的促销结果最大化	在一段时间内持续跟踪由促销活动带来的新贷款项目，与分析的预期结果相对比

上述信用合作社案例中对问题的求解，就是一个典型的商业组织解决问题和寻求机会的过程。回答这些问题要靠一系列的商业数据分析流程，而不仅仅是单独依靠统计学、计算机搜索程序或者运筹学就能解决的。这些结果的价值体现在，它能为信用合作社管理层提供一个重要而精确的行动指南。通过持续不断地将商业数据分析作为决策支持系统，公司不仅可以发现为什么需要商业数据分析，而且可以知道，商业数据分析为何可以成为获得竞争优势的手段。基龙等人（Kiron, etc. 2012）曾经报道，在 2012 年的一项商业调查中，企业凭借商业数据分析使得该组织的决策制定有更好的数据支撑，并获得了竞争优势。

2.3 为什么商业数据分析十分重要：提供战略竞争优势

如果公司能制定实现目标的成功规划，那么这个公司就是市场上的赢家。如果公司没有有效地规划未来，那么这个公司就是市场上的失败者。经营企业重要的一点就是制定规划。只有制定出正确的规划，才会有想要的结果。

企业组织制定的规划一般分成三种类型，如图 2.1 所示。大多数企业组织规划的结构中都存在着双向的活动，但规划过程通常遵循从战略层面到战术层面再到操作层面的顺序。在图 2.1 中，向上流表示从较低层次向上传递的信息，而下降流表示从更高层次的管理向下传递到较低层次的执行命令。Teece (2007) 以及 Rha (2013) 在关于商业数据分析流程和战略规划中三个步骤的研究中，都向我们展示了相同的结论。

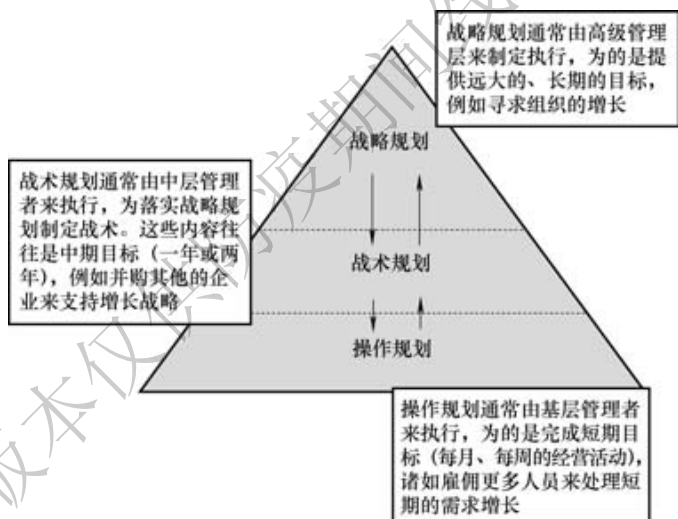


图 2.1 组织规划的类型

（来源：改编自 Schniederjans 和 Le Grand (2013) 第 9 页图 1.2。）

企业为了成为市场上的赢家，要有效制定规划并向下传递正确的指令，这需要优良的信息作为决策基础。有些信息是非常有价值的，因为它为企业取得了竞争优势（企业在更高层次上具有执行力，处于当前同行业或市场的领先地位）。商业数据分析可以为企业提供全部三种规划所需的有用信息，以使企业取得竞争优势。商业数据分析帮助企业获得竞争优势的例子

已由表 2.3 给出。

表 2.3 商业数据分析可以帮助实现竞争优势的途径

竞争优势类型	描述	商业数据分析可以帮助实现竞争优势的途径
价格领导	从市场角度来看, 提供产品或服务, 在行业中为客户提供最低的成本, 同时使公司获得可接受的利润	确定主要竞争对手; 监测、报告和预测具有准确竞争力的价格, 使公司能够保持很低的成本, 同时能够保持和测量利润率
可持续发展	为了确保既不伤害环境, 也不伤害公司在企业的资源使用情况下的盈利底线	查明需要的资源分配, 以避免破坏环境, 提高如何重新分配资源并帮助它们分配优化的方式, 实现最佳平衡点
运营效率	相比于竞争对手, 要改进内部业务和活动, 减少客户的成本。降低的成本可以为客户提供基于效率的更低价格的优势	确定需要更正或修改的业务领域, 并提出替代方案, 以提高效率。这对选择何种替代方案以确保商业效益最大化也是有益的
服务有效性	可以提供给客户比其他公司更简单和易于接受的业务。这突显了公司的服务特色, 同时降低了客户的时间成本, 从而提高了客户价值	获取客户意见, 确定哪些方面需要完善; 解释为什么需要修复, 提供修复建议, 提高服务运营的有效性, 对改进进行量化和报告
创新	通过观察现有市场, 供应突破性的产品, 引入全新或更好的产品或服务, 来扰乱竞争对手的业务意向	获取并确认客户对于新产品的建议或者对于现有产品的意见, 监测客户动态、提供改良的新产品, 监测新产品的市场表现并报告结果
产品差异化	提供其他竞争者不能提供的多种产品、服务或功能	找到竞争对手没有提供的产品, 提出可以提供的新服务, 预测新产品在利润上的潜力

2.4 商业数据分析十分重要的其他原因

商业数据分析在提供信息用于决策领域方面有着数不尽的应用。

2.4.1 商业数据分析为什么十分重要的应用原因

后面几章将给出商业数据分析在决策上的若干潜在应用。表 2.4 描述了几个简单的例子。

表 2.4 商业数据分析在增强决策中的应用

决策中的应用	描 述
提高客户盈利能力	商业数据分析可以提供竞争对手产品的详细信息（当前的定价和价格趋势）。这些信息可用于确定价格，以使企业通过权衡更低价格的销量与利润，或提高价格来增加利润，以保持利润处于最大化水平，怎么做取决于竞争对手的定价
降低风险	利用所有商业数据分析步骤中提供的信息类型，企业不必猜测，而是计算销售、预算、人力和技术等方面的数据来确定未来指南，将概率估计作为降低误判风险的指导
商品促销策略优化	在商业数据分析流程的规范性步骤所用的定量工具，可用于确定库存商品的最佳布局设计，降低库存成本水平，甚至安排销售人员，进而实现促销效果的最大化
人力资源决策	根据 Fitx-enz（2013，第 223 ~ 245 页）的观点，商业数据分析中的预测性步骤的人力资源指标可以被计算出来（劳动生产率、每个全职员工（FTE）的收入、每个全职员工的成本等）。反过来，这回答了怎样做才能提高人力资源员工素质、什么类型的培训将是最有效的、激励性薪酬如何刺激绩效的问题
公司绩效跟踪	在商业数据分析的描述性步骤过程中，除了收集常规数据，还可以不断收集、监控和测量特定的企业绩效参数，然后对那些测量不断更新并加以分析，以便提供最新绩效成就指数用于比较绩效。此外，对预期绩效的分析，可用于指导制定规划以及达成目标

2.4.2 商业数据分析带来的新数据资源的重要性

随着新计算机的产生和电信技术的进步，数据也出现了一些新的类型。因此，这就需要新的分析类型应用于商业数据分析。数字分析（Digital Analytics）是描述利用数字信号源来传送数据时的术语。基于数据分析的这些新来源的例子包括文本分析与非结构化数据分析。文本分析（Text Analytics）被定义为一组语言、统计以及基于计算机的技术，这类技术可对来自文本源中的信息内容加以建模和安排（Basole 等，2013）。寻找文字图案的内容就是针对数据库的搜索过程，目的是提供有用信息。文本分析也被称为文本数据挖掘，就是利用数据挖掘软件观察数据库，为的是寻找和验证各种可以做出预测的信息。

利用文本分析搜索和量化文本数据，这开启了基于技术驱动型数据收集技术来搜集客户和市场信息的巨大机遇。技术驱动型数据的一个例子是社交媒体数据。社交媒体数据被定义为个人或团体通常在技术平台所执行

的，涉及共享、创建、讨论和修改等的口头或电子内容。全球性社交平台有推特（Twitter）和脸谱（Facebook）等。应用于社交媒体数据中的方法或技术，可以分析包括照片或图片、视频、网络论坛、网络日志、论坛、社交博客、维基、社交网络和播客等信息，但不限于这些内容。这些数据来源都是社交媒体分析的基础，在此基础上，可以学习新的社交媒体行为和信息的类型。它们为商业数据分析提出了巨大挑战，这是因为在以有用方式来量化信息方面出现了数据过多的困难。它们还提供了可能找到创造竞争优势信息的机会。社交媒体分析如何帮助找到汽车缺陷的例子，已经由亚伯拉罕斯等人（Abrahams 等，2012）的一项研究所给出。通过利用被汽车热爱者用于社交媒体（在线论坛）的文本挖掘，可以对一系列的质量缺陷进行识别、分类以及优先处理，以便为汽车制造商所纠正。

另一个类似于数字信号源分析的是移动分析。移动分析（Mobile Analytics）被定义为从移动设备，如智能手机、iPad 等终端设备所获得的任何数据。所有这些移动技术被用于从人的互动交往中获得数据（Evans，2010）。这些设备是移动的，随用户从一个地方到另一个地方，这为分析者提供了区分可用信息不同类型。例如，移动技术允许分析者不仅可以跟踪那些潜在客户谈论某种产品的使用（如社交媒体分析），而且可以跟踪客户在哪里对产品进行决策行为，还可以帮助解释为什么会做出那些决策。例如，移动通信技术可能揭示购买发胶的所在地，这个地方其实是邻近投放发胶广告的广告牌区域，这就有助于揭示广告牌促销的有效性。

数据如果被置于数据库当中，同时以符合逻辑的方式归档、访问、被引用甚至使用时，则称为结构化数据。当数据或信息（要么是数字要么是非数字）不能放入数据库或没有预定义的结构时，则这些数据称为非结构化数据。非结构化数据的例子包括图像、文本及其他数据，它们有些会因某些原因不能放置在基于内容的逻辑搜索数据库中。这类数据可以被数字化地存储在文件中，但是不能用任何逻辑模型或分类处理方式来进行检索。大多数包含在电子邮件和 Web 上的数据都是非结构化的。考查非结构化数据的另一种方式是看它会留下来什么，而且不能被放置在结构化数据库中。随着时间的推移，人们投入在开发可用于非结构化数据的复杂算法和其他基于计算机的技术上的努力会更多，以减少属于非结构化数据的数量。每天都会产生大量图形数据或其他非结构化数据，商业数据分析所面临的挑战是，要更加努力地理解非结构化数据，并将其结构化，而同时人们也在努力地获取非结构化数据的信息价值。商业数据分析的部分价值和重要性就是接受这个挑战。

总结

本章试图解释对企业组织来说，商业数据分析为什么是一个重要课题，并讨论商业数据分析是如何解答重要问题的，还有它是如何帮助企业赢得竞争优势的。此外，本章阐述了商业数据分析在组织规划中的作用。最后，本章介绍其他的数据分析类型，解释它们的优点和面临的挑战。

我们在下一章将进一步解释商业数据分析所需资源。正如任何管理任务一样，要想成功利用商业数据分析，就需要在人力和技术资源方面加以投资。第3章探讨怎样进行资源配置，以使商业数据分析效用最大化，并解释为什么需要投资。

问题讨论

1. 为什么商业数据分析流程的每一步都有过去、现在和未来的维度？
2. 什么是竞争优势？它与商业数据分析是怎样的关系？
3. 为什么在商业数据分析中具有决策的能力很重要？
4. 商业数据分析如何有助于实现可持续发展？
5. 什么是数字分析？

参考文献

Abrahams, A., Jiao, J., Wang, G., Fan, W. (2012). "Vehicle Defect Discovery from Social Media." *Decision Support Systems*. Vol. 54, No. 1, pp. 87-97.

Basole, R., Seuss, C., Rouse, W. (2013). "IT Innovation Adoption by Enterprises: Knowledge Discovery through Text Analytics." *Decision Support Systems*. Vol. 54, No. 2, pp. 1044-1054.

Demirkan, H., Delen, D. (2013). "Leveraging the Capabilities of Service-Oriented Decision Support Systems: Putting Analytics and Big Data in Cloud." *Decision Support Systems*. Vol. 55, No. 1, pp. 412-421.

Evans, B. (2010). "The Rise of Analytics and Fall of the Tactical CIO." *Informationweek*. December 6, No. 1286, p. 14.

Fitzenz, J. (2013). "Predictive Analytics Applied to Human Resources." In Isson, J. P., Harriott, J. S. (2013) *Win with Advanced Business Analytics*. John Wiley & Sons, Hoboken, NJ.

Isson, J. P., Harriott, J. S. (2013). *Win with Advanced Business Analytics*. John Wiley & Sons, Hoboken, NJ.

Kiron, D. , Kirk - Prentice, P. , Boucher-Ferguson, R. (2012). “Innovating with Analytics.” MIT Sloan Management Review. Vol. 54, No. 1, pp. 47-52.

Rha, J. S. (2013). “Ambidextrous Supply Chain Management as a Dynamic Capability: Building a Resilient Supply Chain” (Doctoral Dissertation).

Teece, D. J. (2007). “Explicating Dynamic Capabilities: The Nature and Micro-foundations of (Sustainable) Enterprise Performance.” Strategic Management Journal. Vol. 28, No. 13, pp. 1319-1350.

第 3 章 哪些资源对支持商业 数据分析至关重要



本章目标

- 解释确立商业数据分析项目时需要人员、数据和技术的原因
- 解释商业数据分析人员应掌握的技能及其原因
- 描述商业数据分析工作的特性
- 描述数据库百科全书的内容
- 解释数据来源的分类
- 描述内部和外部数据来源
- 描述信息技术基础设施
- 描述数据库管理系统及其如何支持商业数据分析

3.1 简介

为了完全理解为什么需要商业数据分析，就必须认识商业数据分析人员所承担角色的性质。此外，还有必要认识商业数据分析项目所需要的资源有哪些，目的是更好地领会商业数据分析所提供的信息的价值。各个企业对商业数据分析资源的需求会有所不同，以便满足特定的决策支持要求。有些公司可能选择有限的投资，而另一些公司则可能有 BA 团队或 BA 专家部门。无论资源投资水平如何，商业数据分析项目至少需要在商业数据分析人员、数据和技术资源方面加以投资。

3.2 商业数据分析人员

确定商业数据分析员工所需人员的一种方法是，考察提供 BA 服务的组织在商业数据分析领域的认证需要什么。INFORMS (www.informs.org/Certification-Continuing-Ed/Analytics-Certification) 是一家重要的学术和专业组织，于 2013 年宣布启动一项认证分析专业 (CAP) 项目。另一个更为成

熟的组织 Cognizure (www.cognizure.com/index.aspx), 提供各种服务产品, 包括商业数据分析服务。此组织提供了通用认证商业数据分析专业 (BAP) 考试, 测评商业数据分析工作人员的现有技能, 并确定需要改进的领域 (www.cognizure.com/cert/bap.aspx)。这是验证 BA 领域技术能力、专业知识以及专业标准的工具。此认证包括三个考试, 涵盖表 3.1 所列的内容领域。

表 3.1 Cognizure 认证机构认证考试内容领域

测评	专 题	涵 盖 内 容	示 例
I	统计方法	1. 数据可视化和探索数据 2. 描述性统计 3. 概率分布 4. 抽样和估计 5. 统计推理 6. 回归分析 7. 预测建模和分析	1. 图形和图表 2. 平均值、中位数、众数 3. 正态分布 4. 置信区间 5. 假设检验 6. 多元回归 7. 对原始数据的拟合曲线函数与模型
II	运筹学方法	1. 线性优化 2. 整数优化 3. 非线性优化 4. 模拟方法 5. 决策分析 6. 预测	1. 线性规划 2. 整数规划 3. 非线性规划 4. 蒙特卡罗方法 5. 预期价值分析 6. 指数平滑
III	案例研究	现实世界的实践知识	上述解决现实世界问题方法的应用

(来源: 改编自 Cognizure Organization 网站, www.cognizure.com/cert/bap.aspx。)

表 3.1 所列出的大部分内容领域, 将在后面几章和附录中加以讨论与说明。人们很容易清楚地看出, Cognizure 认证项目所需的三个考试是在前面几章所讨论的商业数据分析流程的三个步骤 (描述性数据分析、预测性数据分析和规范性数据分析) 的内容, 认证项目图 3.1 中的专题适用于商业数据分析流程中的三个主要步骤。基本统计工具适用于描述性数据分析步骤, 更高级的统计工具适用于预测性数据分析步骤, 运筹学工具适用于规范性数据分析步骤。一些工具既可应用于描述性数据分析步骤, 还可应用于预测性数据分析步骤。类似地, 仿真工具既可应用于在预测性数据分析步骤, 又可应用于规范性数据分析步骤中解答问题, 这取决于使用它们的方式。将所有工具相结合, 则是案例研究的现实应用。运用案例研究旨在提供实践经验, 其中所有工具用于解答重要问题或寻求机会。

其他组织也提供专门化的认证项目。除了分析工具之外, 这些认证包括其他领域的知识和技能。例如, IBM 提供了各种专业的 BA 认证 (www-03.ibm.com/certify/certs/ba_index.shtml)。这些包括多达数十种统计、信息

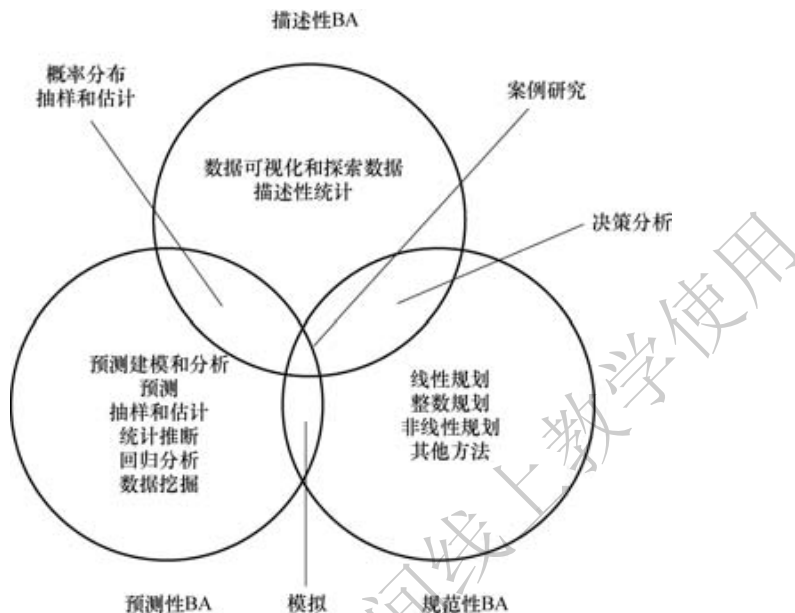


图 3.1 认证内容领域及其与商业数据分析中三个步骤的关系

系统和与 BA 相关的分析方法的认证，还包括与商业数据分析人员（管理员、设计师、研发人员、解决方案专家和技术专家）相关的专业技能集合，如表 3.2 所示。

表 3.2 BA 人员类型

BA 人员类型	描述
管理员	在 IBM BA 和 BI 软件平台中，管理员管理服务器（负载平衡、安装和配置）。他们管理计算机门户报告，管理调度程序和负责故障排除，还负责用户授权和安全认证
设计师	设计师是负责利用关系数据模型撰写报告的团队成员，还负责专业报告的质量提升、定制和管理
研发人员	作为团队成员，研发人员所需的技能与 BA 过程密切相关，涉及分析应用程序、数据仓储、建模、运筹学和统计方法的使用，以及向客户展现实时数据流
解决方案专家	作为团队成员，解决方案专家运用适当的方法和开发途径，负责 BA 程序的分析、规划、设计、部署和操作。这需要众多不同 BA 软件应用的了解，还需要了解统计学、信息系统和运筹学
技术专家	技术专家是负责安装和配置 BA 和 BI 应用的团队成员

（来源：改编自 IBM 网站 www-03.ibm.com/certify/certs/ba_index.shtml。）

参与者在商业数据分析中职位和角色的多样性，导致了 BA 职位需要不同的技能集合或能力。在一般意义上，BA 职位需要业务、分析和信息系统技能方面的能力。如表 3.3 所示，业务技能涉及人员和流程的基本管理。商业数据分析人员必须与组织内部的商业数据分析员工（BA 团队成员）以及公司内部的其他功能区域（BA 客户和用户）进行沟通，使其有用。因为他们服务于公司内的各种功能领域，商业数据分析人员需要拥有客户服务技能，以便他们能够与公司的人员进行互动，并了解问题的性质。商业数据分析人员还需要向公司内部的用户销售他们的服务。此外，有些人必须领导商业数据分析团队或部门，这需要相当强大的人际管理技能和领导能力。

表 3.3 选择商业数据分析人员技能或能力要求的类型

技能或能力的类型	可能的角色描述
业务方面	<ul style="list-style-type: none"> • 领导力 • 对相关人员的管理能力和沟通能力 • 管理 BA 项目（确定优先次序、安排进度等） • 管理 BA 过程（采用合适的规则、程序） • 确定项目的要求 • 培训 BA 人员的开发能力
分析方面	<ul style="list-style-type: none"> • 了解如何使用统计方法 • 了解如何使用运筹学方法 • 知道如何对定量数据进行数据挖掘，并对非结构化数据进行文本分析
信息系统方面	<ul style="list-style-type: none"> • 维护和使用计算机门户网站 • 识别和提取数据 • 保持数据质量

商业数据分析的基础是认识理解表 3.1 所列出的分析方法，还有未列出的其他分析方法。除了工具集，还需要知道它们如何被整合到商业数据分析流程中以利用数据（结构化的或非结构化的）获得由分析需求所指导的客户的信息。

总之，对于从事商业数据分析职业的人员来说，应该知道如何与人进行互动，并利用必要的分析工具将数据用于有用的信息，这些信息可以在信息系统中被处理、存储和共享，以便指导公司达到更高水平的公司绩效。

3.3 商业数据分析中的数据

人们需要利用结构化和非结构化数据（第2章曾经介绍过）来生成分析。将数据组织成一个可以理解的框架作为开始，统计学家通常将数据分成有意义的一些数组。

3.3.1 数据分类

有许多方法可以对 BA 数据进行分类。其中常用的一种是依据内部或外部来源加以分类（Bartlett, 2013, 第 238 ~ 239 页）。内部数据的典型例子见表 3.4。当公司试图解决内部生产或服务运营问题时，就可能需要内部来源的数据。典型的外部数据来源（参看表 3.5）数量众多，为 BA 提供了巨大的多样性和独特的挑战。数据可以定量（如销售额）方式来测量，或以定性方式采用偏好调查（例如，某一种产品和另一种产品相比，消费者更加喜爱某一产品），或通过网络上消费者对竞争性产品利弊的讨论量（喋喋不休）来定性测量。

表 3.4 BA 可利用的典型内部数据来源

内部数据类型	描述
记账和提示系统	记账系统和提示系统会打印出账单，并监控基于价值的客户细分市场的客户付款信息
业务	行业编码、会计信息、员工信息等可在日常业务过程中收集
客户	当客户为支付产品、服务或者签收时，可以获得客户的名字、地址、带来的盈利、特殊合同、市场细分情况等
客户关系管理系统	客户关系管理（CRM）系统是收集和提供客户的历史，以及如投诉、不再与一个公司发生联系等行为信息
人力资源	有关员工、薪水、能力等信息可通过对就业历史记录的日常工作来收集
企业资源计划系统	企业资源计划（ERP）系统用于内部沟通，目的是为管理问题和所关注的事情，以及生产与销售产品所需要的操作活动提供直接反馈信息
产品	信息是通过从采购经由售后再到监控盈利能力、持久性和质量而收集来的
生产	用于优化生产、库存控制以及将产品投递给客户的供应链的信息均是在生产过程中收集起来的
问卷	客户行为的信息是通过问卷获得的，以便测量客户服务和产品质量等
网页日志	信息是利用公司网站通过 Cookie 和其他手段来收集的，以便了解客户的导航行为和产品的利益点

表 3.5 商业数据分析可利用的典型外部数据来源

外部数据类型	测量工具
满意度	<ul style="list-style-type: none"> • 收入、利润 • 市场份额、销售额 • 产品/服务调查数据 • 忠诚度 • 品牌意识 • 每位客户的平均支出
客户人口统计	<ul style="list-style-type: none"> • 地理位置（距市场的距离） • 收入水平 • 市场规模
竞争	<ul style="list-style-type: none"> • 市场份额 • 竞争对手盈利能力 • 广告促销努力 • 偏好调查 • 产品上的 Web 信息
经济	<ul style="list-style-type: none"> • 人口统计 • 收入分配统计

绝大部分外部数据来源可在文献中找到。例如，美国人口普查和国际货币基金组织（IMF）是用于宏观经济水平上建模的十分有用的数据来源。同样，被调查群体和调查数据也可能来自于 Nielsen（www.nielsen.com/us/en.html）、Claritas 的消费心态或人口统计数据（www.claritas.com）、Equifax 的财务数据（www.equifax.com）、Dun & Bradstreet（www.dnb.com）等。

3.3.2 数据问题

无论数据来源如何，它必须被放置在使得商业数据分析人员可以使用的结构中。我们将在下一节讨论数据仓库，但这里我们重点介绍对于任何数据库或数据文件的可用性至关重要的几个数据问题。这些问题是数据质量和数据隐私。数据质量可以定义为服务于其收集目的的数据所要求的质量。它意味着各种不同的应用会要求不同的质量，但有一些共同点，通常包括准确地表示现实、测量所要测度的内容、永恒性以及完整性。数据质量较高，将有助于确保竞争力，帮助客户服务，并提高盈利能力。当数据质量较差时，就会提供矛盾的信息，导致误导性决策。例如，文件有缺失数据，这可能会阻碍某些表单的统计建模，并且信息的不正确编码可以完

全使数据库无用。确保数据质量需要数据管理者付出努力，以此清除错误信息的数据，并修复或替换缺失数据。我们将在后面几章讨论这些质量数据测量的一些内容。

数据隐私是指保护共享数据，以方便只允许访问它的用户。这是一个安全问题，需要在知道所需内容与共享太多而导致更多风险之间加以权衡。不受限制的访问存在许多风险。例如，竞争对手可以通过访问地址来窃取公司的客户。产品质量故障的数据泄露可能损害品牌形象，客户可能变得不信任分享信息的公司。为了避免这些问题，公司需要遵守关于客户隐私的现行立法，并制订专门用于数据隐私的计划。

收集和检索数据以及计算分析需要使用计算机和信息技术。商业数据分析人员所做的大部分工作涉及管理信息系统，以从各种来源中收集、处理、存储和检索数据。

3.4 商业数据分析技术

为支持员工进行日常业务操作，公司需要信息技术（IT）基础设施，如表 3.6 所示。这些类型的技术是商业数据分析运作的基本需求。

表 3.6 信息技术（IT）基础设施

类 型	说 明
计算机硬件	这是指信息系统中用于输入、处理和输出活动的物理设备。硬件包括各种尺寸的计算机，各种输入、输出和存储设备；包括连接计算机的电子通信设备，如移动手持设备
计算机软件	这是指信息系统中对硬件组件起控制和协调作用的预编程指令集，包括像 ERP 这类的系统软件，也包括移动设备上的小型应用程序（计算机软件应用程序）
网络和通信技术	物理设备和软件连接着多种硬件设备，将数据从一个物理位置传输到另一个物理位置。它包括用于连接到网上以共享语音、数据、图像、声音和视频的计算机和通信设备，还包括因特网、内部网（仅限公司内部员工访问的基于因特网技术的内部网络）和外部网（将私人内部网扩展到企业外的授权用户）
数据管理技术	软件控制支配着物理存储介质上的数据组织。它包括数据库管理系统、数据仓库、数据集市和联机分析处理（OLAP），以及数据、文本和 Web 挖掘技术

就商业数据分析而言，表 3.6 所列出的数据管理技术是非常重要的。数据库管理系统（DBMS）是一种数据管理技术软件，它允许公司集中并有效地管理数据，并通过应用程序提供对存储数据的访问。DBMS 通常用作应用程序和结构化数据的物理数据文件之间的接口。数据库管理系统承担着在哪里存储以及如何存储数据才能使其更有效的任务。此外，有些 DBMS 可以处理非结构化数据，例如，面向对象的 DBMS 能存储和检索非结构化的数据，如图形、图像、照片和语音数据。想要处理目前大多数公司收集的大数据，这些技术是必不可少的。

数据库管理系统提供用于组织、管理和访问数据库的工具。数据定义语言、数据字典、数据库百科全书和数据操纵语言是四个更为重要的功能。数据库管理系统具有数据定义功能，以此规定数据库中内容的结构，在创建数据库表和创建用于识别内容的特性时使用。随着数据库内容成规模地增大，这些表和特性是检索中的重要因素。这些特性记录在数据字典（一种自动或手动创建的文件，存储着数据的大小、描述、格式及其他用于刻画数据的属性）中。数据库百科全书是列出公司当前数据项目，以及可以用于构建或购买的数据文件的表格，它的典型内容如表 3.7 所示。对于商业数据分析来说，数据库管理系统所包含的数据操纵语言工具是非常重要的。这些工具用于在数据库中检索特定的信息，例如结构化查询语言（SQL），用户通过查询和响应可在数据库中找到特定的数据。

表 3.7 数据库百科全书的内容

数据库内容项	描 述
目的	数据库存在的意义，包括利用数据时使用的其他报告和分析
时间	数据何时收集或何时运用的时间窗口
资源	内部（审计、会计等）和外部（客户等）的资源
示意图	展示表格和其他数据文件关联的示意图
成本	收集数据的成本，包括采购价格
数据可用性	数据何时是可用的时间窗口
收集技术	收集方法，包括观察、数据挖掘、普查和焦点小组
收集工具	网页、客户自发形成、电子调查等

对于决策者来说，数据仓库是用于存储具有潜在意义的当前和历史数据的数据库，可供任何需要访问的人使用。数据仓库中的数据是禁止更改的。数据仓库还提供一系列的查询工具、分析工具和图形报告工具。有些公司使用内部门户网站，为整个公司提供数据仓库的信息。

数据集市是数据仓库内的重点关注子集或较小的分组。公司通常建立企业范围的数据仓库，这里的中央数据仓库服务于整个组织，较小的、分散的数据仓库（称为数据集市）聚焦于组织数据中的有限部分，并被放置于独立的数据库中，便于特定用户访问。例如，公司可能会开发关于产品质量的小型数据库，专门关注于优质客户和产品问题。与企业范围的数据仓库相比，专注于最受关注项目的数据集市可以更快地被创建起来，而且成本更低。

一旦捕获数据并将其存入数据库管理系统中，人们就可以使用商业数据分析工具对其进行分析，包括 OLAP，还有各种针对数据、文本以及 Web 的挖掘技术。OLAP 是一种允许用户在多个维度上观察数据的软件。例如，可以根据年龄、性别、地理位置等条目来查看雇员，用 OLAP 找出 35 岁以上、男性以及在国家西部地区的雇员人数。即便数据存储在非常大的数据库中，OLAP 也可以让用户在线快速获得特定问题的查询答案。

数据挖掘是一种发现驱动（Discovery-driven）的软件应用，通过在大数据或大型数据库中寻找隐藏的模式和关系，从中推断规则来预测未来的行为，从而提升对商务数据的洞察力。观察到的模式和规则被用来指导做决策。运用它们还可以预测这些决策的影响。它是一种理想的预测性数据分析工具，在第 1 章所提到的 BA 过程时曾使用过。采用数据挖掘技术可获得的信息类型如表 3.8 所示。

表 3.8 采用数据挖掘技术可获得的信息类型

信息类型	描述	示例
关联	发生的事情和某个事件的关联	报纸上投放的广告促进了销售
分类	通过检查经过分类的已有项目并用推断出的规则来指导分类，从而识别出描述项目所在群体的模式	识别出哪些客户倾向于需要更多的客户服务，哪些客户需要的更少
聚类	对未经分组的数据进行分类，从而有助于发现数据中的不同组别	对一个单一的、庞大的客户群体按照不同特点进行分组识别，比如找出在飞机上更喜欢喝茶的旅客

(续)

信息类型	描述	示例
预测	预测值可以识别客户行为模式	通过一个典型客户，估计未来销售额的走向
序列	事件的时间关联	识别出购买新房子的人和将在 90 天内购买新车的人之间的关联

文本挖掘（第 2 章曾经提及）是一个软件应用程序，用于从非结构化数据集中提取关键元素，发现文本材料中的模式和关系，还能概括信息。鉴于存储的大多数信息都是非结构化数据（电子邮件、图片、备忘录、文字稿、调查结果、业务收据等），挖掘与寻找有用信息的需求在未来将会要求更多地使用文本挖掘工具。

Web 挖掘旨在发现 Web 用户的行为模式、趋势，提升对客户行为的洞察力。例如，营销人员使用网络搜索引擎提供的商业智能服务，可以跟踪各种词汇和短语的流行程度，了解消费者感兴趣的内容和正在购买的东西。

除了前面讨论的通用软件应用之外，还有一些 BA 分析师日常在商业数据分析流程三个分析步骤中所使用的软件应用程序（参见第 1 章）。这包括 Microsoft Excel[®] 电子表格应用程序、SAS 应用程序和 SPSS 应用程序。Microsoft Excel (www.microsoft.com/) 电子表格系统带有专门用于 BA 分析的加载项。这些加载项拓宽了 Excel 在 BA 领域的应用。Analysis ToolPak（分析工具库）是一个 Excel 加载项，其中包含用于描述性和预测性 BA 过程步骤的各种统计工具（如图形和多元回归）。另一个 Excel 加载项 Solver（规划求解）包含 BA 过程中规范性分析步骤中所使用的运筹学最优化工具（如线性规划）。

SAS[®] Analytics Pro (www.sas.com/) 软件提供了一个桌面统计工具集，允许用户以可视化方式对信息进行访问、操作、分析和呈现。它允许用户从任意来源获取信息，并将其转换为有用和有意义的信息，这种可视化方式使得决策者能够快速认识并掌握数据中的关键信息。它专为需要以一种易于理解的方式来探索、检查和呈现数据的分析师、研究者、统计人员、工程师以及科学家而设计，并且可以多种格式分发发现的结果。它是一个主要用于 BA 过程中描述性分析和预测性分析步骤的统计工具包。

IBM 的 SPSS 软件 (www-01.ibm.com/software/analytics/spss/) 为用户提供了广泛的统计和决策工具。这些工具用于数据收集、统计操作、结构

化和非结构化数据的趋势建模以及优化分析。通过获取统计工具包，SPSS 软件可以涵盖 BA 过程中的所有三个步骤。

商业数据分析流程也可以使用其他软件来进行。本书将使用的是 Lindo Systems (www.lindo.com) 的 LINGO® 软件。LINGO 是一个综合工具，旨在使构建和求解优化模型更加迅捷、轻松和高效。LINGO 提供了一个完整集成包，包括用于表达模型优化的直观语言、用于创建和编辑问题的全功能环境以及一组内置求解器。此求解器用于求解线性规划、非线性规划、二次规划、随机线性规划和整数线性规划方面的最优化建模。

总之，任何进行商业数据分析的组织所需要的技术，都会涉及一种通用信息系统结构，包括数据库管理系统、BA 分析师所需要的能对组织有所贡献的更为具体的软件。具有更多 BA 需求的组织，基本上会需要更多的技术来支持商业数据分析工作，但对所有想要运用 BA 寻求竞争优势的企业来说，需要对技术进行大量投资，这是因为商业数据分析是一项依赖技术的工作。

总结

商业数据分析的重要意义和它的成本是成正比的。在这一章中，我们探讨了成本问题，也通过商业数据分析带来的诸多好处证实了 BA 的必要性。本章讨论企业需要哪些资源来支持 BA 项目。由此，我们确认三个主要资源领域：人员、数据和技术。为了找到具有相应技能的商业数据分析人员，我们阐述了 BA 认证考试的内容，还讨论专业人员的类型，并讨论了 BA 数据的内部来源和外部来源可作为数据分类的手段。最后，商业数据分析技术涵盖了支持个体分析师的软件包的通用信息系统。

在这一章中，我们重点介绍要想实现可行的商业数据分析操作所需的资源投入。在第 4 章，我们开始第 3 部分“如何应用商业数据分析”。具体来说，在下一章中，我们将关注本章所提到的资源如何配置在组织中，以便实现目标。

问题讨论

1. 如何运用 BA 认证考试内容，解释 BA 分析师的技能集合？对于他们来说，哪些技能是必需的？
2. 对于想要从事商业数据分析的人员来说，为什么领导力是一项重要的技能？
3. 在商业数据分析中，为什么对数据依来源进行分类是重要的？
4. 什么是数据质量？为什么它在商业数据分析中是极其重要的？
5. 数据仓库和数据集市之间有什么差异吗？

参考文献

Bartlett, R. (2013). *A Practitioner's Guide to Business Analytics*. McGraw-Hill, New York, NY.

Laursen, G. H. N., Thorlund, J. (2010). *Business Analytics for Managers*. John Wiley & Sons, Hoboken, NJ.

Stubbs, E. (2013). *Delivering Business Analytics*. John Wiley & Sons, Hoboken, NJ.

Stubbs, E. (2011). *The Value of Business Analytics*. John Wiley & Sons, Hoboken, NJ.