

米其力 大業力

米其力大業力

BEHIND EVERY
GOOD DECISION

米其力大業力
米其力大業力

米其力大業力
米其力大業力



目录

[封面](#)

[赞誉](#)

[献言](#)

[作者简介](#)

[前言](#)

[致谢](#)

[序言](#)

[第一部分 你好，分析](#)

[第1章 要么分析、要么灭亡](#)

[第2章 什么是分析](#)

[第3章 七大分析方法](#)

[第二部分 深入理解](#)

[第4章 BDIR：商业分析的五步框架](#)

[第5章 预测分析](#)

[第6章 数据和分析工具](#)

[第三部分 领导工具包](#)

[第7章 分析和领导者](#)

[第8章 分析竞争](#)

[第9章 分析师领导者的攻略手册](#)

[第10章 付诸行动](#)

[第11章 常见的陷阱](#)

[第四部分 实践中的分析：十个案例研究](#)

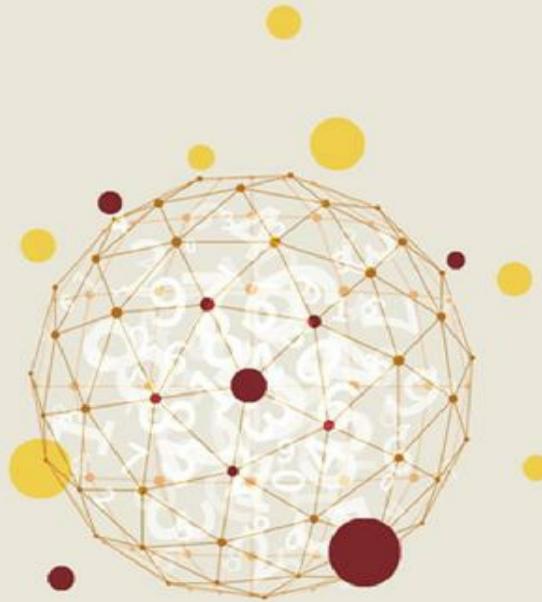
[附录 商业统计101](#)

>>【看书累了吗？休息一会！更多新书朋友圈 每日免费分享微信
xueba987。渺沧海一粟：为终身学习者赋能 2019年7月】

封面



华章经管



BEHIND EVERY GOOD DECISION

How Anyone Can Use Business
Analytics to Turn Data into
Profitable Insight

从回放回溯出发，教你如何收集数字、分析，并且验证假设，拟定执行方案。

数字决策力

从商业分析中挖掘利润

[美]

皮扬卡·贾殷 (Piyanka Jain)
普尼特·夏尔马 (Puneet Sharma) 著
李雅琦 译 朱中 审校

 机械工业出版社
China Machine Press

>> 【看书累了吗？休息一会！更多新书朋友圈 每日免费分享微信
xueba987。渺沧海一粟：为终身学习者赋能 2019年7月】

赞誉

坚实的分析基础是众多新兴趋势的发展依据，也是任何公司达到巅峰状态的根本。本书带领新手缓缓进入分析的领域，告诉经理人如何与他们的分析师取得最好的结果，它能建立一套架构，使任何企业能了解驱动其成功和失败的因素。本书是业务经理人的必读佳作。

——施与公司（GiveCorps）财务长布莱恩·惠伦（Brian Whalen）

本书引人深思，是切合需求的实务工作者手册。本书以智能架构和解决方案，填补了分析理论与实务间的缺口。分析专家和业务经理人往往讲不同的语言，加上数据数量与日俱增和商业模式的不断变化，使得情况更为复杂。本书提供了用良好的方法和洞见克服这些问题的思路。强烈推荐给想利用数据做决策的人。

——万豪国际（Marriott International）

人力资本分析资深总监席卡·纳雷·皮里·凡卡特斯瓦拉

（Shekar Nalle Pilli Venkateswara [NV] ）

我的祖父常说：“如果你第一次都没有时间做好，怎么还会有时间去做第二次呢？”遵循本书所说的程序，会使你直探正确的问题，找到分析数据的正确模式，引领你得出正确的观点和建议——一次就成功。本书有许多出色的例子和故事，简单易懂。讲述常见陷阱的一章更告诉我们，抄快捷方式或节省时间的小路，省略几个步骤会有什么样的后果。我强烈推荐置身于商业环境中的每个人都阅读本书。

——应用全球服务（Applied Global Services）常务董事

约翰·卡明斯（John C.Cummings）

本书的研究细致、周密，见解非常深刻并有很多经典的现实案例。本书对初学者和经验老到的专业人士都有非常大的帮助。

——全球营销、顾客和市场分析差距公司

(Global Marketing Customer and Marketing Analytics Gap Inc.) 主管

素密·库马尔 (Sumit Kumar)

本书提供了必要且耳目一新的观点，读者可以去阅读分析和洞见是如何产生的。皮扬卡和普尼特在教会大家从分析到形成观点的“流程”方面做得非常出色。本书说到做到，准确地融合了人员、流程、技术和策略，引导读者理解如何运用商业分析技术和方法来形成行动观点且没有把太多的焦点放在数据和技术上。本书讲述了很多成功和失败的故事，使读者能向那些掌握了分析技术的企业学习，并从那些单纯凭直觉决策的失败案例中汲取教训。

本书简单易懂，可以在任何企业组织中应用。它指出，分析不必只依赖于某一种技术式工具，也可以在直觉和正确的数据基础上展开。本书是所有决策制定者和数据分析师的必读书目，可以使他们的企业更上一层楼。

——不定式商业智能与大数据实务 (Business Intelligence and Big Data at Infinitive) 高级主管迈克·兰普勒 (Mike Lempner)

我未曾见过哪本分析方面的书能写得如此令人爱不释手。连专家都会混淆的观念、术语和方法，都被转换成非常容易理解的、有趣的内容。我不仅向领导者、分析师和专家顾问强力推荐本书，也向希望改善运营活动的所有领域人士推荐。

——Visa公司劳动力规划副总裁

(资深业务领导人) 梅安克·贾殷 (Mayank Jain)

本书是商业和技术专业人员想要建立数据驱动型职场文化的必读之物。

——GigaOM研究主任乔·梅特兰 (Jo Maitland)

本书在几个方面表现出色：①重新强调对企业日益增多的数据进行分析的必要性以在数字经济中维持竞争力；②以有针对性的方法，将不同的内容提供给不同的受众（分析师、经理人、领导者）；③打破数据分析的“象牙塔”印象，提供明确的指引，告诉我们何时和如何使用简单可行

的方法，产生速赢的成果；④强调将分析纳入决策所需的软技能及其重要性。

——圆点金融集团（Round Point Financial Group）副总裁

克里斯蒂安·贝格（Christiaan Berge）

本书为解决问题提供了十分重要的指南，汇总了任何分析专业人员从产生观点到形成策略所需的所有内容，适合各种管理岗位的读者阅读。作者以出色的例子，让我们综合预览简单的分析技术和复杂的预测分析，这是史无前例的。

——史威夫特资本（Swift Capital）副总裁米希尔·寇克（Mihir Korke）

本书是一本非常引人入胜且有趣味的书籍。由于数字时代每天产生数千兆字节的数据，所以从这些数据中获得洞见会带来巨大的竞争优势。作为一名商业领袖，我见过很多数据分析师玩弄数字很在行，但是在理解商务环境上却遭遇了滑铁卢。初出茅庐的分析师如果想要明白自己在驱动企业成功上的潜力，那么本书是必读书目。作为一种战略业务杠杆，“分析”这个领域仍然还处于发展阶段。本书也可以帮助商业领袖弄清楚如何有效地利用分析去发现经营问题。向皮扬卡和普尼特致敬。

——印孚瑟斯公司北美能源、通信和服务业务流程外包

（Notrh American-Engergy, Communications & Services at Infosys BPO）

销售总监安舒·莫汉（Anshul Mohan）

本书对想窥探神奇的商业分析世界的应届大学毕业生和想通过进一步利用分析的力量来提高各自企业效率的商业老手来说都是必读书目。在这个时代，好决策的产生越来越依赖于数据和分析，皮扬卡和普尼特展示了许多不同领域中的成功企业是如何引领经商的新航线的。

——普华永道咨询公司总经理凯莫尔·凯罗凯亚（Kemal Karakaya）

本书是一本简单但是应用性很强的数据分析指南——对即将到来的机遇和前方可能遭遇的冰山做出更好的解读至关重要。我极力推荐本书。

——意大利联合信贷银行集团风险管理公司

(Group Risk Management Unicredit Spa)

副高级总裁阿布舍克·达巴里 (Abhishek Darbaris)

>> 【看书累了吗？休息一会！更多新书朋友圈 每日免费分享微信
xueba987。渺沧海一粟：为终身学习者赋能 2019年7月】

献言

献给我的女儿Jia，我生命中的魔力；

献给我的丈夫Parth，我的精神支柱；

献给我的父母，我的力量源泉。

——皮扬卡·贾殷 (Piyanka Jain)

献给一直相信我的Shweta，

以及给我的生活带来快乐的孩子Nikhil和Samir。

——普尼特·夏尔马 (Puneet Sharma)

>> 【看书累了吗？休息一会！更多新书朋友圈 每日免费分享微信
xueba987。渺沧海一粟：为终身学习者赋能 2019年7月】

作者简介

皮扬卡·贾殷 (Piyanka Jain)

Aryng管理顾问公司的CEO，这家公司专攻商业分析。作为分析领域备受尊崇的业界领头羊，她曾在美国营销协会 (American Marketing Association)、预测分析世界 (Predictive Analytics World)、GigaOm 和谷歌用户分析会 (Google Analytics User Conference) 等商业及分析研讨会上担任专题演讲人。演讲主题为通过数据驱动型决策来取得竞争优势。担任分析领导人15年之久，她对商业产生了超过1.5亿美元的影响。她是技巧丰富的问题终结者，总是想方设法找出各种模型，提出各

种观点，来推动客户企业的变革并影响商业驱动力。她认为客户满意度、客户授予的权利以及正面积极的接触是最好的反馈，有了这些之后，金钱的效益自然会随之而来。

她拥有两个硕士学位，她的论文侧重于应用数学和统计学。她还热衷于徒步旅行，喜欢跑步和瑜伽。她与丈夫Parth和女儿Jia生活在加利福尼亚的森尼维耳市。

皮扬卡·贾殷的联系方式如下：

Twitter: @AnalyticsQueen

Blog: www.aryng.com/blog/

Email: pj@aryng.com

普尼特·夏尔马（Puneet Sharma）

行动公司（Move Inc.）的分析、增长黑客技术和使用者研究部副总裁。热衷于运用分析与消费者观点来驱动显著的商业效益。

他富有创意，作为资深领导人也屡获嘉奖，15年来一直在金融服务和互联网公司倡导高效能营销并提供产品策略营销方案。专长于通过消费者的信息和深层分析来推动决策制定，并提供可行性建议，以实现行之有效的变革。

普尼特在马里兰大学取得工商管理硕士（MBA）学位之后，先后在PayPal、Capital One和HSBC等《财富》（Fortune）前500强公司担任营销、产品和分析领导人的职务，负责以消费者为基础的零售业务，抓住了获利丰厚但在过去却无人问津的机会。他与他的妻子和两个儿子现在生活在加利福尼亚的森尼维耳市。

普尼特·夏尔马的联系方式如下：

Email: puneet08@gmail.com

>>【看书累了吗？休息一会！更多新书朋友圈 每日免费分享微信xueba987。渺沧海一粟：为终身学习者赋能 2019年7月】

前言

皮扬卡·贾殷

我痛恨统计学。1999年在得州农工大学（Texas A & M University）的环境工程论文课题中，我真正想做的是跟踪调查使危险废弃物场地变得不安全的事件。在实验室的研究过程中，当事实愈渐清晰时，坦白地讲，能解决这个问题的唯一方法就是采用非线性回归技术建立逆形态分布模型。

我一直都喜欢数字和数学，曾相对我所就读的年级提前4年用自己的方法推导出勾股定理。我一直质疑每一项杰出技术所带来的影响，包括我父亲的突破性发明——垃圾焚烧发电。数学和工程学成就了今天的我，但在1999年，我努力确保统计数据不仅可以分析一个问题，还可以解决这个问题，并有真正切实的商业影响。我与数据化敌为友，并找到了我的激情所在。

我在明尼苏达大学（University of Minnesota）计算机科学与工程专业完成了第二硕士学位。利用蚂蚁信息素概念，通过统计建模与仿真，我设计了一个自我更新的网络路由表。我曾教三年级和四年级的本科生学习微积分和统计学，我有能力利用统计和分析技术去解决各种不同类型的问题。

当Adobe公司希望通过这些技能来提高其客户体验时，我就知道从这一刻开始，分析将成为我人生的一部分。一如既往地，我努力确保研究的数据可以被转化为行动和积极的商业影响。在一个热爱数据的公司里，我和我的团队能够通过市场营销和广告，改善客户体验，并最终对公司的盈利带来影响。

后来我在不同的公司任职，在贝宝（PayPal），我遇到了普尼特·夏尔马。在那里，我们每人带领着一个分析团队。

普尼特·夏尔马

取得了计算机科学学位和金融领域的工商管理硕士学位之后，我开启了早期的职业生涯。当我开始了解数据和分析，以及我的数据分析师朋友一天的生活后，我就知道我不得不屈从于这种渴望——我称之为我

的“欲望”。美国第一资本投资国际集团（Capital One Financial Corp）招聘数据分析师时，我提交了申请。经过严格的能力测试，Capital One邀请我在董事会里担任数据分析经理，以好好发挥我的技术和业务技能。我在我热爱的事业里蓬勃发展——将分析转换为商业影响和有形的美元价值。公司相信数据能报告所有关键的战略和决策，并且从一开始就这样设计。我和其他500个人一起参加培训，学习统计、建模和分析。我知道进入数据分析领域是一个正确的决定。

然而在这之后，我很惊讶地得知，和我有工作来往的其他公司并不像我们这样对利用数据来带动商业影响感兴趣，有的公司甚至毫无兴趣。分析是一个“支持功能”，根据优先级不同的事业部，分析的产出可能会也可能不会被考虑。现有系统的设立并不利于我们获取有深远影响的各种商业信息。物力和财力被花在理论分析上，却并不一定能直接产生商业影响。一个项目如果有强大的数据驱动执行支持，我和我的团队就能够带来惊人的影响。但是实际并非如此，我曾见证数百万美元资金分文不动地躺在那里，仅作为企业结构的一个要素和分析的感知价值。经过一段时间强烈的自我怀疑，质疑我自己是否真的有能力推动持续的影响，我发现，需求是有的，而我还有不足，我从那些失败中吸取了教训。作为一个技术公司的分析负责人，我还有很多经验、教训需要学习。

我和皮扬卡·贾殷多年来经常一起共进午餐并时不时地讨论一番，我发现我们原来都有同样的经历。我们开始清楚地认识到这些模式，并想办法来填补这些缺口。这不只是关于分析，也是因为我们现在拥有可以强有力地提高业绩的环境。分析的复杂性是一个障碍，但是我们知道大约只有20%的时间需要那些资源密集的过程。我们通过共同的努力开始阐明分析的力量、企业对分析的态度、常见的失误，以及用简单的方法来驱动影响。

皮扬卡·贾殷

这几年我在会议上发表讲话的时候，也听到了来自其他机构同僚的亲身经历。他们招架不住这些数据、大数据、分析、预测分析，虽然许多人都跟上了潮流，但是他们在分析上的投资并没有得到多少投资回报（ROI）。整个场面模糊又复杂，有一个需求是我们可以填补的。

2011年，我创立了Aryng——一家管理咨询公司，专注于分析。普尼特·夏尔马担当顾问的角色，我则设计了一个叫作BADIRTM的商业影响分析方案。BADIRTM代表商业问题（business question）、分析计划

(analysis plan)、数据收集 (data collection)、得出结论 (insights) 和提出建议 (recommendations)，它是一个可将数据转化为更明智决策的简单五步框架。

当召开研讨会时，我们听到了越来越多这样的说法，怎样将如此简洁基础的方法用在工业和学术领域呢？本书囊括了所有这一切。我们很高兴带你一起走上这趟旅程。

>>【看书累了吗？休息一会！更多新书朋友圈 每日免费分享微信xueba987。渺沧海一粟：为终身学习者赋能 2019年7月】

致谢

本书是我们的信念和商业分析知识的结晶。虽然在过去的几个月里我们全身心地投入在本书上，但是如果没有朋友和家人的帮助与支持，我们根本无法完成本书，他们不仅慷慨地提供了各种专业知识，还大度地包容我们。对于这些善良又聪慧的人所提供的支持，我们心怀感激。

我们是分析专家，但并非专业作家。感谢Lakshmi Jayaraman坚持不懈地听我们诉说故事，加入设计和创新情境，并参与书写和编辑来支持本书。我们同样还感激你利用你的设计专长和插图知识提供的帮助，你的坚持使本书充满乐趣和可读性。

特别感谢我们的审稿人，你们的反馈确保了我们的阐述合情合理，并清楚地传达了本书的意图，所以谢谢你们：Bonny Elgamil、Claire Dean、Nirmal Baid、Desmond Chan、Marcia Trask、Ketan Babaria、Parth Sethia、Sangita Khater、Sunil Bafna、Anshul Mohan、Mahendra Kumar、Narasimman Jayaraman、Kamala Subramaniam、Cole Roberts和Betty Li。

感谢Ramkumar Ravichandran对分析数据工具的研究综述，感谢Kameshwari Viswanadha的输入设计和评审，感谢Vijay Aviur提供我们在本书中谈到的故事。

感谢Shailesh、Ashish和Satvik团队，为本书画了生动的插图。感谢Terri Griffith对本书写作过程的指导以及提供的许多简介。

最后，我们要感谢我们家庭的支持，尤其是我们各自的另一半Parth和Shweta。在我们潜心研究或是缺席大部分周末的时候，她们为我们分担了很多。

>> 【看书累了吗？休息一会！更多新书朋友圈 每日免费分享微信xueba987。渺沧海一粟：为终身学习者赋能 2019年7月】

序言

为什么写本书

没有什么能比火箭科学更复杂了。分析学并不是火箭科学。当然，分析学非常专业，涉及各种格式的大量信息，而我们花费数百万美元希望通过分析加上一些人性的洞察、简单的数学和复杂的统计去找某种模式。好吧，这听起来挺难，但就像你刚开始接触的任何课题一样，有基本原理、方法和简单诀窍，帮你掌握它并利用它来实现你自己的利益。尽管成分复杂，但分析（尤其是商业分析）其实就是一个简单的解决问题工具。本书的目的就是解释分析是什么，以及怎样去进行分析。

我们已经告诉了你分析其实可以很简单。那我们是否告诉过你，企业中的每个成员都足够了解如何在企业日常事务中带来积极变化？或者告诉过你这种简单的分析只花费复杂分析成本的一小部分，却可以帮你解决80%的实际业务问题吗？嗯，它可以。

并不是所有分析问题都是需要数据科学家建立并读取复杂模型的大型项目。事实上，这些问题中的80%可以由管理人员和决策者日常解决，他们只掌握简单工具和方法就可以了。他们知道什么时候应该以及如何最好地利用数据科学家和分析师去解决更复杂的商业问题。本书专为这些商务人士提供基本知识，并帮助数据科学家最大限度地发挥自己的能力。本书力争结合两方最好的资源来推动结果。

那我们期望分析会影响什么样的结果呢？如果操作得当，通过商业分析可以发现差异，创造满意的客户、更好的产品、改进的流程、高效的员工、提高收入和利润，从而让股东开心。结合我们在大型企业领导分析部门30多年的工作经验，我们知道，有些企业做法得当，而另外一些企业还在苦苦挣扎。这种模式对我们来说越来越明显。因为成功的企业对分析、丰富的资源、清楚的沟通和简单的方法有信心，所以它们已经能够

利用分析来推动经济增长，而那些还在挣扎的企业却有各种失误，但有一条简单的出路，那就是学习本书的内容。

每次出现一个热门的流行语时，各公司就急急忙忙地跟风。了解分析可能带来的潜在价值之后，各企业开始对分析还有大数据产生错误预期。它们认为分析太复杂了，需要能在数据实验室里解决所有问题的“科学家”。他们认为数据应该定期给出答案，然后大数据工具就可以修复这一切……不可能！我们希望在本书里消除这些或是其他的误解。

但是，你如何挣脱这种分析无力？有一个方法可以解决这种病症，BADIR框架可以帮你用五个简单步骤从数据中得出决策。这种方法非常普遍适用，可应用于商业或企业分析的很多方面。为了说明这一点，本书中贯穿了这方面的例子。

这些例子来自我们在硅谷（Silicon Valley）各大公司及最近在Aryng管理咨询公司分析团队的工作经历。

本书写给谁

写给所有人：第一部分和第三部分写给每一位有兴趣了解分析的读者。
第一部分是对分析的介绍：什么是分析、为什么要分析和常用分析方法。第四部分提供了真实的案例，对分析如何影响从政治到体育，从执法机关到技术领域这些典型的商业环境进行说明。

写给想学习和实践分析的读者：第二部分详细介绍了商业分析的BADIR模型，对预测分析进行了概述，并介绍了用于商业智能和分析的工具。这部分将为你提供必要的知识，将商业分析应用到你日常大部分的工作中，并在必要时针对复杂分析与专家进行充分的沟通。本书附录中还提供了必要的统计知识，方便你需要时参考。

写给领导者：商业领袖和分析大师会发现第三部分值得花时间阅读，因为通过很好地利用数据，你可以引领企业走向盈利。本书提供了专门的工具包和方法，帮助你创建数据赋能型企业，建立一个分析议程，扩大分析在各类组织中的影响，并轻松地避免掉入常见的陷阱。

本书所用的术语

我们在本书中使用了以下术语，为了帮助你熟悉它们，我们在这儿给出

了这些术语的定义。

A/B测试（A/B testing）或测试和学习（test and learn）：所谓测试，是对根据分析、直觉或客户输入值所得出观点的有效性进行测量的过程。A/B测试能对受控实验中所得到的结论效果进行评价，在一个对照实验中评估这些见解的效果。

分析（analytics）：利用数据和分析来驱动影响，运用结构化的方法来解决商业问题的科学。

BADIRTM：一个从数据到决策的浅显易懂的秘诀。它将数据科学和决策科学（即商业技术技能和软技能）相结合，将数据转换成能产生影响力的观点。这个分析过程适用于各个行业，本书第4章对此进行了介绍。

商业分析（business analytics）：利用历史数据和简单的统计方法，再加上一定的商业背景，去获取观点并对组织施加影响。商业分析包括相关分析和总量分析。

商业领袖（business leader）：在职能部门、业务部门或整个企业扮演领导角色的个人，包括首席执行官、总经理、产品负责人、营销负责人、客户关系负责人。在这些角色中，他们负责各自业务部门的盈亏（P&L），对公司的战略发挥至关重要的作用。

商业用户（business user）或商务专业人士（business professional）：在各行业中，执行职能战略并负责日常决策的个人，他们直接或间接地影响公司业绩。我们交叉使用这些术语，商业用户的作用可以体现在营销、产品、运营或其他角色上。

数据驱动（data driven）：有些公司是创新驱动公司，有些则是技术驱动公司。数据驱动公司利用数据引导更好的决策，为组织的种种目标提供支持。在本书中，我们交替使用数据启用、数据引导和数据驱动这些术语，读者可以采用其中任何一个，为各自企业的愿景和目标提供支持。

数据科学家（data scientist）：有能力使用先进的统计工具做高级分析的个人。在本书中，有时也用“分析专家”来进行描述。

关键绩效指标（KPI）：衡量某机构或项目成功与否的关键绩效指标。收入和利润往往是企业层面的关键绩效指标。

P&L：利润和亏损（盈亏）。

预测分析（predictive analytics）：对历史数据应用先进的统计技术，以期能预测未来事件。预测分析技术包括线性回归和决策树等，通常需要数据科学家和预测分析专家来完成。

3个关键问题TM（3 key questionsTM）：Aryng咨询公司为企业制定可行性分析框架的另一个秘诀。3个关键问题帮助领导者像剥洋葱一样对企业展开层层分析，明确KPI，找到这些KPI的驱动因素，并协调内部项目，来移除影响KPI的驱动因素。

>> 【看书累了吗？休息一会！更多新书朋友圈 每日免费分享微信
xueba987。渺沧海一粟：为终身学习者赋能 2019年7月】

第一部分 你好， 分析

第1章 要么分析、要么灭亡

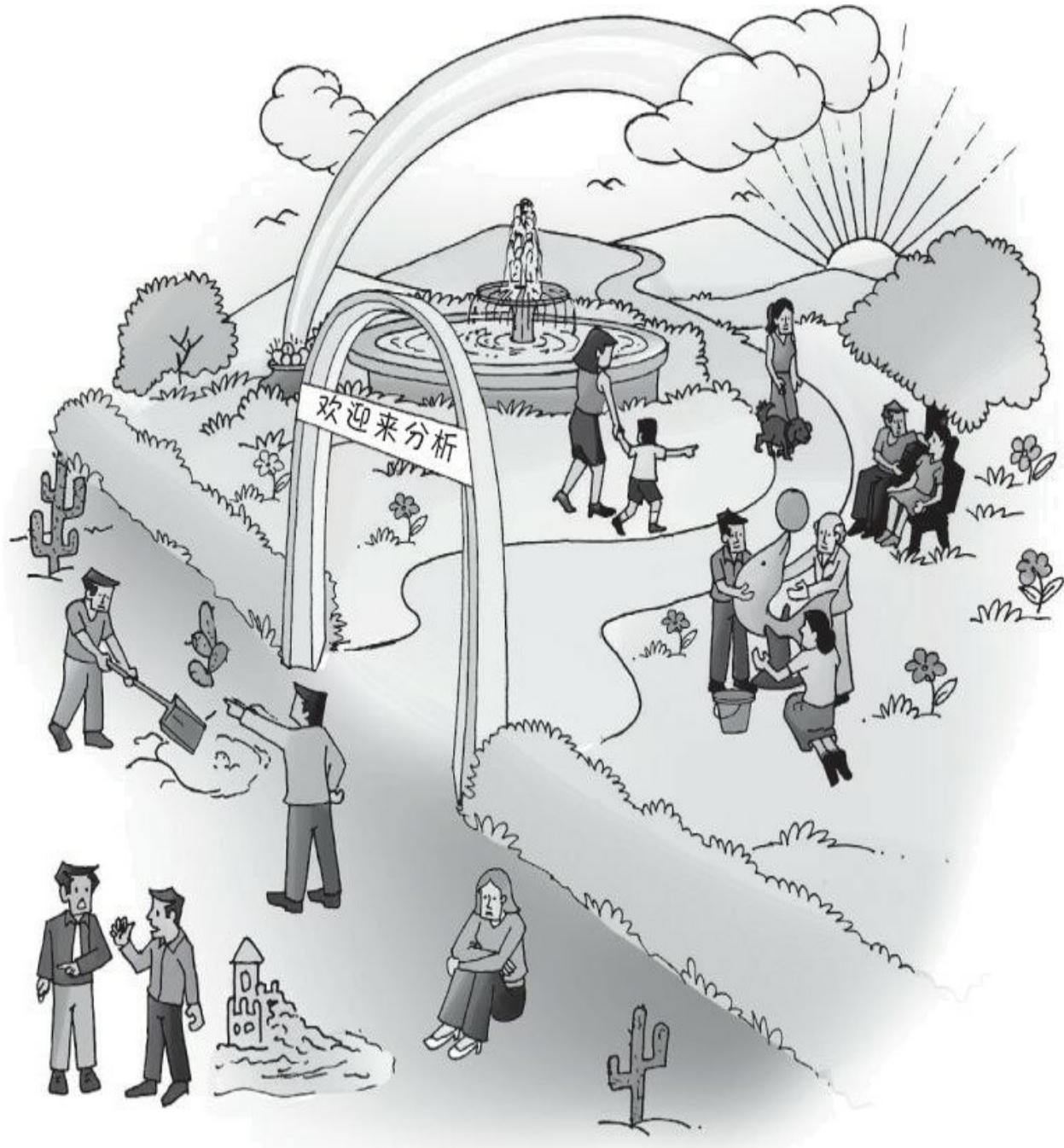
第2章 什么是分析

第3章 七大分析方法

第1章 要么分析、要么灭亡

本章将讨论：

- 为什么要分析？因为不能纯粹只靠运气。
- 什么是分析？



电路城公司（Circuit City）是美国占统治地位的电子零售商，自20世纪八九十年代发展壮大，在沿海岸线的战略位置缔造了它的商店帝国。2008年8月，电路城公司最大的竞争对手百思买公司（Best Buy）报告它实现了销售增长，季度利润达到2亿美元，但电路城公司同期却亏损了2.39亿美元。到2009年3月9日，电路城公司关闭了它最后的567家零售店。

电路城公司的魔力消失了。但是到底发生了什么？

20世纪90年代末，网络电子商务的增长率预计将达到2300%。百思买在2000年向世界推出了它的网站，但电路城公司却没有。

1999年，电路城公司公布大型家电的年销售额达16亿美元，仅次于西尔斯百货（Sears）的销量。2000年，电路城公司决定停止销售家用电器，以节省仓库的储存和运输成本。但紧接着，在20世纪第一个10年的中期，房地产市场异常繁荣，家用电器的销售额达到自18世纪中期家电产品面世以来的历史新高。电路城公司错过了这一浪潮。

到2004年，百思买已取得了“A”级的地理位置，而电路城公司为节省资金却选择了低成本的偏远位置。消费者当然愿意蜂拥到可以在10分钟车程内就到达的商店。

电路城公司为降低人力成本，在保持员工人数不变的情况下，将各地商店收入最高的销售人员解雇。许多高薪销售员都是公司最优秀、最忠诚的伙伴，为公司创造了最大的收益，陪公司走过了成长的风风雨雨。结果，电路城公司却让他们把自己的技能和经验带到了百思买和其他竞争对手那里。

很多战略失误都会导致大型企业的衰落。上面提到的因素其实是可以通过数据分析来予以强调的。如果能密切关注行业增长预测和竞争对手对新技术的采用情况，电路城公司就会及早扩大它的在线业务。

在削减成本的决策中，投资回报率（ROI）分析本可以阻止电路城公司解雇其高薪员工，因为他们带来了更高比例的收入。如果电路城公司分析过数据来平衡供需，从而得知库存管理的问题，那么上层决策者就不会采用这些不合理的成本削减方案来偿还因滞销库存而引起的债务。管理层在公司范围内做出了两个关键的决定来降低成本：①关闭家用电器业务；②将零售店位置转移到“B”级区域或郊区位置。数据可能表明了这两项是公司成本的主要来源，但这些决定却没有经过合理的测试。如果先减少家用电器库存并观察在“B”级地理位置一些网点的业绩表现，就能推翻这些决策，阻止公司关闭利润丰厚的家电业务和向郊区转移这种导致客户流失的错误决策。

分析不仅仅事关数据，还影响决策。如果电路城公司能有效利用数据分析，并根据分析结果做出正确决策，那么该公司可能就不会倒闭了。

什么是分析

所谓分析，是运用结构化的方法来解决商业问题，用数据和分析来产生影响力的科学。许多企业正努力面对战略性的经营问题，比如成本削减和业务创新。要解决这些问题，往往会影响企业预算和组织结构。数据分析在这类决策中可以成为一个强大的工具。正确的分析方法是以数据和观点为支撑的明智的解决方案。

直觉+数据=强大的洞察力→良好的决策

不仅是洞察，还要是可操作的洞察——可以提升思维过程，驱动决策并激励积极行动的洞察。像宝洁公司（Procter & Gamble）、亚马逊（Amazon）、领英（LinkedIn）和Capital One等企业领袖，就是通过采用数据引领、假设驱动和分析支撑的策略来支持想象和直觉，称霸各自的领域，我们将在本书中逐渐介绍这些方法。

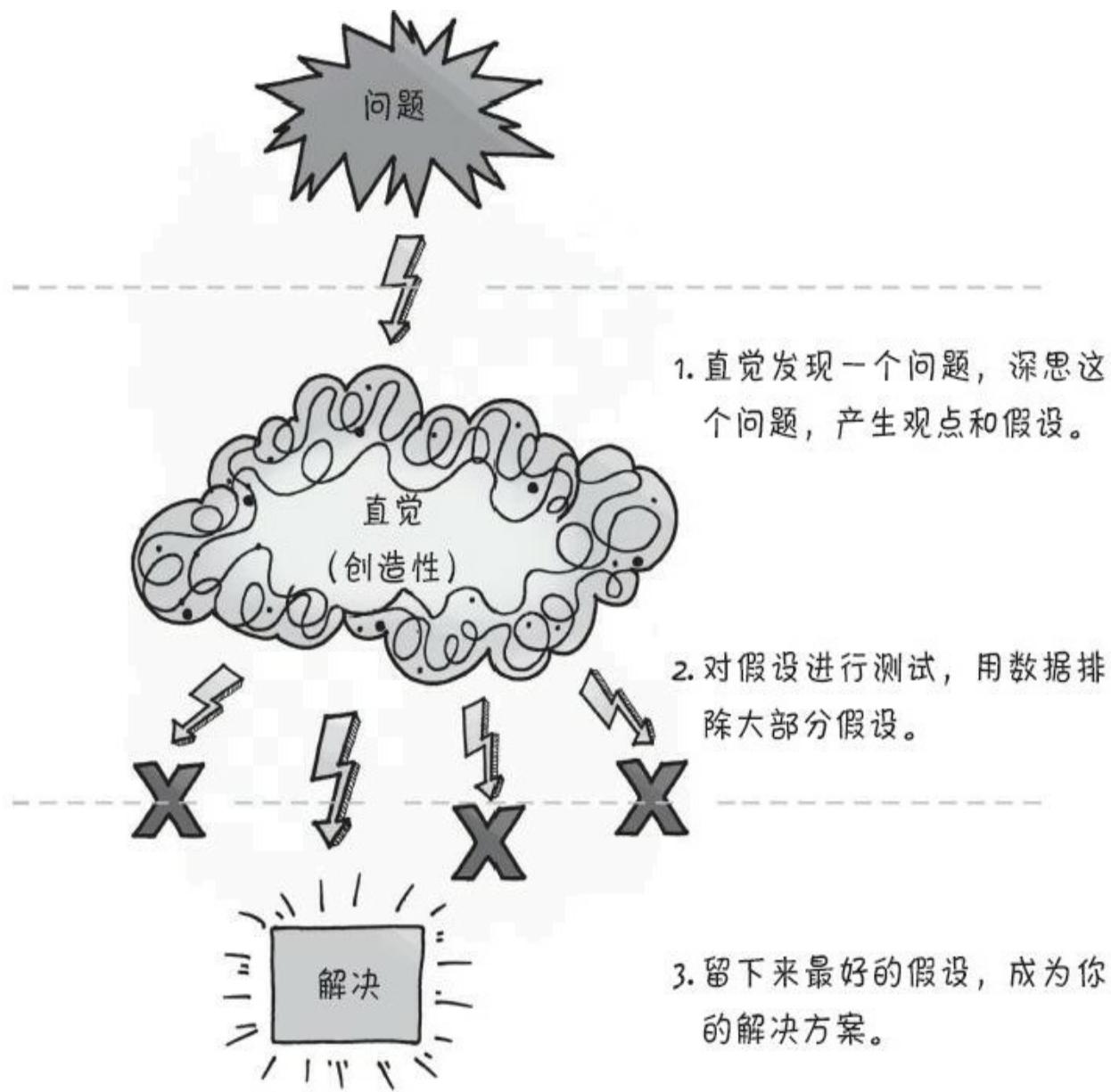
欢迎来到商业分析的世界。

2011年，宝洁公司的销售额超过826亿美元。宝洁公司位列《财富》杂志“世界最受尊敬的企业”排行榜第五名，比上一年的排名第六名又上升了一位。

宝洁公司在全球有127000名员工，在180个国家和地区销售300多个品牌的产品。宝洁公司从1837年成立以来，在长达177年的时间里，一直位于家居用品行业的领先地位，它是如何做到的呢？这个市值1400亿美元的消费品公司随着时间的推移不断更新自己，利用假设驱动的分析策略保持它的生命力和自身优势。

什么是假设驱动的分析？想象一下，假设你在太平洋里游来游去想寻找金子。你难道不希望将目标缩小到最有可能找到金子的那个地方吗？这样你就不必在广阔如从中国到美国的水域里大海捞针了。

与此类似，数据就是一片海洋，假设可以有助于缩小范围，将我们定位到最有可能找到答案的地方。假设是由人类的直觉产生的，它是依靠集体智慧和利益相关者的经验以及对业务和周边环境的理解而做出的。利用数据对假设的有效性进行验证，可以得出一致的解决方案。这种解决方案的优势在于两点：数据和假设。



假设驱动的分析策略如何起作用？

首先要回答一系列简单合理的问题：“现在正在发生什么”以及“为什么会发生”。这些问题已帮助像宝洁公司这样的企业找到可以用来解决难题的行动。没有任何观点支持的行动意味着所做的决定没有任何前瞻性，不知道哪一种方法会产生结果。

2010年，宝洁公司首席执行官麦睿博（Bob McDonald）制定战略，将公司业务无缝衔接，使流程数字化，利用分析实现业务优化和高效决策，但宝洁公司同时也要削减成本。到2012年，宝洁公司裁减了1600个非制

造业岗位，信息技术部门已经削减了超过9亿美元的总开支。虽然减少了员工人数和开支，但宝洁公司获得的总收入却有所增加。

它是怎么做到的呢？宝洁公司聘请了分析专家来开发技术，让公司在不可预知的环境和压力下能做出明智、及时的决策。正是这样的分析预测了宝洁公司在未来一年的市场份额和其他效能指标。这种决策能力的核心是一系列的分析模型，这些模型可以揭示商业中各种事件的峰值和低谷，确定事件发生的原因，并指出企业可以采取的措施，从而利用或减轻这些事件的影响。

宝洁公司的分析是由文化理念驱动的，也就是在钻研复杂的数据分析前先提出问题并做出假设。思考“什么”能监控出货量、销售量和市场占有率等关键业务指标。当发生异常时，追问“为什么”，从而了解正在发生的事件背后的详情，找出大到国家和领土，小至生产线和商店层面的影响因素，通过分析确定宝洁公司可以撬动的“行动杠杆”，比如价格、广告和产品组合，并预测此行动可能会导致的结果。通过有效地运用“什么”和“为什么”，宝洁公司的决策者能够直接着手于问题的解决方案，从根本上提高经营效率。

这种战略对像宝洁公司这样的大型企业无疑是有效的。现在，当面对着不断缩小的预算和无穷无尽的数据集时，如何才能像大型企业一样维持公司的正常运转和繁荣？如何才能最高效地实施你的商业战略并使最明智的商业决策成为可能？

你也可以通过提出一连串正确的“什么”和“如何”这样的问题，利用简单的结构化过程——假设驱动分析和对观点进行测试，来确定要实施的策略。



这不是多么高深复杂的问题，而是一个公开的秘密。像宝洁公司一样，先通过如表1-1所列的假设驱动、数据引导过程，提出这些简单的问题。一个假设驱动的分析策略有很多优势，包括更深入地洞察、更快地获取信息并采取行动。通过假设驱动方法得到的具有可操作性的观点可以精确到马上就可以实施的程度。

表1-1 假设驱动的分析策略：是什么和怎么做

<p>{ 目的 }</p> <p>“你想做什么?”</p> <p>或者</p> <p>“本不应该发生却发生了的是什么?”</p>	<p>这个问题看起来简单，但它的答案决定了你的初始目标。接下来，它会为你设置背景，将你的想法按既定目标去进行调整</p> <p>它为优化和有效转化提供了理论基础，因此推动了战略决策和行动的积极性。这个问题也可以帮助你深入到难题或是紧迫事务的根部，它们需要得到立即的关注和有效的解决方案。它也可以帮助你确定各种参与者，更重要的是，明确影响成功的障碍。这个问题的答案将开启成功的大门，让你探索和发现更多的东西</p>
<p>{ 假设 }</p> <p>“是什么在影响某事件?”</p> <p>或者</p> <p>“为什么某事件会发生?”</p>	<p>回答这些问题能让你明确你的潜在催化剂和关键影响因素，帮助确定关键的利益相关者和他们的目标。</p> <p>有一点要铭记于心：这个问题的答案是不明确的，即它仍然是一个假设，但这是你能得到的最好线索，在开始解决棘手问题前，它能帮你做出最好的选择</p>
<p>{ 分析 }</p> <p>“我们可以做 A 或 B 或 C 或 D 来解决这个问题。”</p> <p>或者</p> <p>“根据分析，做 A 和 B 可能会增加利润率。”</p>	<p>然后进行分析，用数据证明或证伪假设并得出具有可操作性的结论观点。这些结论观点告诉你需要解决什么样的问题，禁止什么样的举措，以及需要采取什么样的行动来解决手头的问题。有关如何分析的其他信息，请参见第 4 章</p>

(续)

{ 测试 }

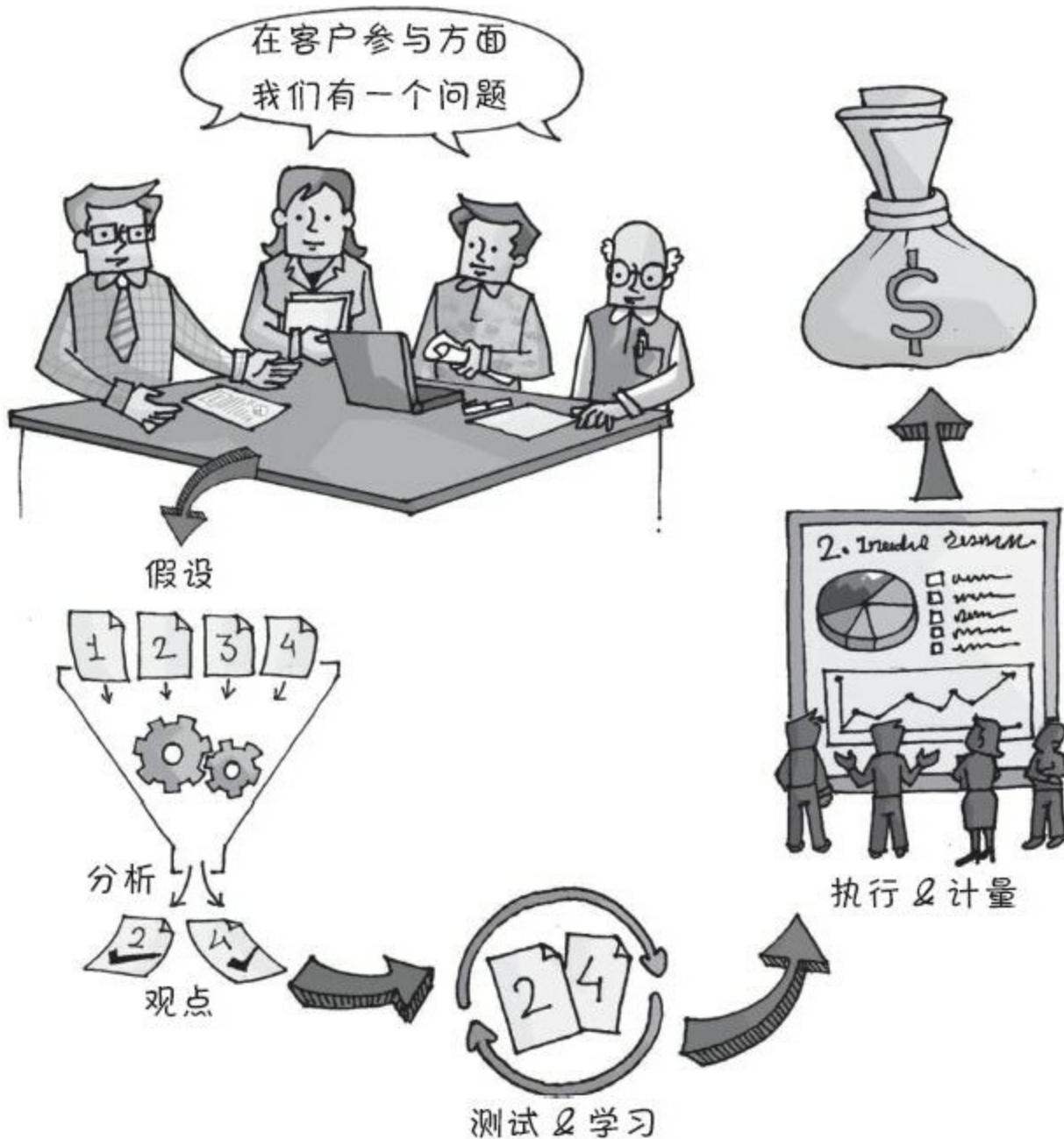
“根据分析，B 可以增加 X% 的利润率而 A 可以增加 Y% 的利润率。”

或者

“消费者调查告诉我们，消费者只是喜欢 B。”

在最终执行前测试这些结论或观点，对加强这些论点的可操作性是很重要的。消费者在决定一个产品的可行性和其未来的销售前景中起了关键作用。你可能通过“什么”和“为什么”已经发现了一个产品功能需求，这些结论或观点可以支持你的假设。精心地设计一个试验产品，展示潜在客户群的想法和购买力（就像对顾客分部展开的测试一样），看看你在这两个方面的计算是否正确

下图展示出一个完整的假设驱动过程。



一个新常态

这有一个大胆的声明。每家公司在几乎相同的程度下都暴露在同样的因素中。区分一个企业的成功与否，不是看它遇到了什么样的挑战，而是看它对挑战的回应。那些能熟练搭上经济快车并安然度过竞争海啸的公司，就可以继续生存和发展下去。此外，如果在应对市场挑战时做出糟糕的内部决策，则会削弱公司的竞争力，直到它最后崩溃。

自2008年以来的经济疲软和它缓慢的复苏使商业世界变得日益复杂。销售直线下降、预算缩减以及客户要求越来越高，这些影响已经引起了各企业不同的反应：有些乘风破浪做出了更明智的决策，而有些则被淹没在潮流之中，也就是说，很多因素可以导致一个企业的成功或失败。来自外部市场的巨大压力可能阻碍企业盈利；消费者信任度的起伏、投资波动和经济动荡也会成为企业的巨大发展障碍。竞争和迅速发展的新兴技术可以给周密的计划造成严重的破坏。这些情况迫使企业改变工具和方法以解决商业问题。

分析能够让我们在非常特殊的内部商业动态基础上做出明智的决策。最近一项基于60个案例数据的研究发现，每花费1美元在分析工作上，就能带来10.66美元的回报。



“但说实话，分析真的适合我吗？分析是否可以在减少预算的基础上帮我扩大业务？”当然没问题。

分析可以为每个人服务

从日常管理者到领导者到科学家再到想买车的某个人，分析可以为每个人服务。如果你希望以最高的效率和最佳的效果经营你的公司，让做出最明智的商业决策变为可能；如果你是某企业的一员，而这个企业越来越无法容忍决策失误；如果你正为企业寻找新的优势和差异性；如果你正在极复杂的全球环境中苦苦挣扎，竞争加剧，产品上市时间加快，客

户愈发挑剔；如果不管所有的外部挑战和不断减少的预算，你仍希望提升你的业务；如果你想买一辆车，那么你是时候和分析打个招呼了。

小结

- 分析不仅仅是关于数据，也涉及决策。
- 有一个方法可以解决这种问题：从询问正确的“什么”和“如何”开始。
- 分析可以为每个人服务。

第2章 什么是分析

本章将讨论：

- 什么是分析？
- 商业分析与预测分析。
- 商业分析与商业智能。
- 大数据与商业分析。
- 什么是增长黑客？



场景1

假如你是本（Ben），阳光太阳能电池制造公司（Bright Sun Solar Cell Manufacturing Company）的质量监控经理，你刚从测试小组收到一份报告，说最后一批太阳能电池板的产品功效惊人得低。在你的声誉岌岌可危的情况下，你需要加足马力展开行动，弄清到底哪儿出了错。你的头脑里闪过各种猜测。“为什么最后一批的故障率这么高？”“我们怎么解

决？”“我如何才能确保未来的批次没有同样的问题？”

时间紧迫。每多一个错误都会加剧目前的收入下滑，损失正蔓延到数以百万计。

场景2

假如你是安德烈亚斯（Andreas），一个寻宝猎人，同时是月亮公主号的船长。你的梦想即将实现。一位私人投资者刚刚给你提供了一大笔钱，让你拿去用于寻找在太平洋丢失的宝藏。更好的消息是，她为你提供了一笔特别的奖金，前提是你能发现1715年沉没的遇难船，这艘船因为镀银的甲板而被称为“板舰队”（Plate Fleet，普雷特舰队）。

但这笔钱只够你和你的船员在海上航行30天。

你是哥伦布还是福尔摩斯

我们首先要说，虽然寻找宝藏和解决制造缺陷是两个非常不同的情景，但处理方案的路径，也就是有效的解决方案，将会非常相似。

如果你愿意选择做哥伦布，你可以和你的船员扬帆起航并潜入水中，寻找任何有可能埋有宝藏的地方。这会是个令人愉快的航线，毫无疑问你会发现自己沉浸在美丽的大海中：闪闪发光的碧水，醒来时嬉戏跳跃的海豚，还有古老非凡的珊瑚礁。



但与此同时，时间紧迫，资金越来越少。你是否正在尽可能有效且迅速地行动呢？在30天内，你能否找到沉没的板舰队？可能不会，因为你的探险和找到板舰队的目标是相互独立的。如果你的任务是寻找虎鲸或一种食用海藻，你的行动也可能跟目前是完全一样的。你采取的是探索性的方法。

那么如果你选择做福尔摩斯，你会怎么做呢？你会找出发生海难时的潜在位置，研究板舰队可能沉没的具体位置在哪里。你可以查看历史上的贸易航线和沉船记录，然后找到深度信息去排除一些可能性。但你要找的并不是其他沉船，你想确定的是板舰队沉没的地方。你会察看1715年的飓风路径记录。通过这些线索和事实，开始锁定可以找到宝藏的潜在领域。一旦锁定了10来个左右的潜在区域并且优先排序出前三个位置，你就能派你的潜艇或深海潜水员下去侦察了。

只根据合理的信息进行集中搜索，就增加了你在更短的时间内找到板舰队的机会。

如果失败了，你仍然有时间重新计划，再次攻克这个难题。这就是侦探性的方法。



作为本，想要识别和解决导致太阳能电池板失效的问题，你可以采用的方法和寻宝没有什么不同。如果用哥伦布式的探索性方法，怀着找到失败原因的希望，你可以收集到数量惊人的数据点。但是，正如你可以想

象到的，由于有很多条流水线，每条线上又有众多流程和设备，想要误打误撞快速找出症结所在，机会将非常渺茫。

也许问题就是某个单向阀的故障，但是想象一下，在数以百计需要查看的东西中找到这个单向阀的概率会有多大呢？当然，也要看你有多幸运了，但是你愿意把毕生事业都只押在运气上吗？

本，来见识一下数据驱动的决策制定

采用侦探性的方法，通过仔细研究数据去寻找线索，就能找到症结所在。这就是商业分析101（基础窍门）。我们可以用提问来开始：哪里发生了故障？是不是所有的生产线都在生产有缺陷的产品，或者只是个独立事件？故障是什么时候开始的？产品的故障到底在哪里？故障对应的工艺和设备到底是什么？通过这些问题，你就能找到指引。

最需要注意的是：你并不需要知道所有引导型问题的答案。你可以根据已知信息建立一个可靠的假设，虽然可能并不完全，然后用这个假设去锁定最让人生疑的问题。



通过扮演神探福尔摩斯的角色，你很快就会发现是10号流水线的高温导致了产品故障。温度问题往往是由原材料质量或者硬件故障引起的，比如热交换器故障。

通过明确要寻找的首要问题，你很容易就缩小了问题的原因范围，在这个实例中，就是热交换器上一个故障的阀门。通过追根求源，节省了时间和资源，而你的主管也因此认可了你的能力。

分析是高深复杂的吗

分析这个词确实会让人联想起一个模糊的图像，就像统计学课程上一个又长又复杂的方程，但是，商业分析其实并没有那么复杂。事实上，我们在学校所学的统计技术，只有极少部分可以被应用到商业场景中。而在这些能够应用于商业分析的技术中，又只有一小部分能满足投资报酬率（ROI）、可解释性、可维护性、周转时量和可扩展性等商业约束。因此，在商业情景中，简单的分析技术会更适用。

2006年，网飞公司（Netflix）发起了一项数据挖掘竞赛，如果谁可以将网飞公司的电影评分预测准确度提高10%及以上，就可以得到100万美元的奖金。这便是典型的商业“遇见”统计的情形。大量数据挖掘者前来竞争。一年后，进展奖团队报告他们花费了超过2000小时达到8.43%的提升。两年后，一个综合方案（不同模型的组合）终于达到了10%的改进，成为最终赢家。获奖的解决方案由数百个模型混合而成，但网飞公司直到现在也没有使用它。为什么？因为这种解决方案落实起来很昂贵，而且它所能带来的收入增加并不能弥补模型评分和维护的成本。



事实证明，商业中最有效的分析技术就是少量的简单技巧，大多数专业人士都可以学习并掌握。你会惊讶地发现，实际上，决定交易成败的，

其实是联结商业和数学之间的一些大众技能。

商业中的分析

只有当能够产生影响力的时候，分析才是有用的。根据你的商业业务，这种影响可以是收入增长、流程效率或者是产品改良。

以产生影响为目的的分析=数据科学+决策科学

能产生影响力的分析有两个组成部分：

- 数据科学：技术环节，旨在从数据中获取观点。
- 决策科学：商业环节，旨在联合利益相关者，将通过数据科学环节获取的观点融入决策过程，并将它们转化为行动。

好的分析不仅仅有炫酷复杂的模型，也涉及使用软技能、了解业务并呈现在商业环境中有益的相关见解，从而驱动商业影响。除非分析可以驱动商业影响，否则它就不能被称为分析，而只是统计，只是数据科学。深陷于数据之中的分析师往往只专注于获得最酷的观点，却忽略了人为因素的影响。

成功的商业分析专家并不是统计学家（尽管统计学家能成为伟大的分析专家），而是能够热情而恰当地与向他们求助的商业伙伴相处的人。确实，分析师需要选择正确的技术去分析并提供见解。但是一个成功的分析师还会利用他的影响力和软技能，与利益相关者或商业同行建立结盟。这就确保了当宝贵的论点被挖掘出来后，对口业务已经准备好采取行动，并将这些论点转换为商业影响。更多信息请见第4章和第10章。

成功的分析师使用问答的形式和相关人员研讨，了解他们应当要求什么样的数据（真正的商业问题）和可能的答案线索（假设）。他们可以利用这些答案去挖掘合适的数据，做适当的分析，并提出可行的建议。

这并不代表技术分析和建模技巧就不重要，但没有商业掌舵能力和人际交往技巧，分析就成了一艘迷航之船。



80/20分析法则

由于分析版图仍处于形成阶段，相关定义和术语仍在不断变化之中，甚至可能在本书出版后还会依然如此。在本书中，我们将使用以下术语：

- 商业分析：对过去的数据使用简单的分析方法。
- 高级分析：除商业分析外的所有其他分析，包括预测分析。

关于大数据和预测分析，存在许多的传说，它们都涉及复杂的技能，动辄需要上百万美元的投资，这让我们很难去相信简单分析的力量。然而事实是，只有20%~30%的决策真正需要使用预测分析这样的高级技术。通过商业分析或简单的分析技术，可以明智而谨慎地做好70%~80%的商业决策，这些技术，是任何一个专业人员都可以学习并在Excel电子表格中实现的。预测分析是商业分析的延伸，它通过从历史数据分析中获得的认知来预测未来的项目或表现，可以用来提高在商业分析中所获论点的准确性（见表2-1）。



表2-1 商业分析和预测分析的优缺点

商业分析	预测分析
[+] 快速且简单	[−] 需要较多的时间和资源
[+] 大多数商务专业人士可以学习，只需通用工具，如 Excel	[−] 需要专门的建模技能和高级统计工具
[−] 不详尽，所以可能会遗漏能驱动最大商业绩效的重要参数	[+] 可以评估数百个指标，从而确定最重要的参数，驱动最大的商业绩效

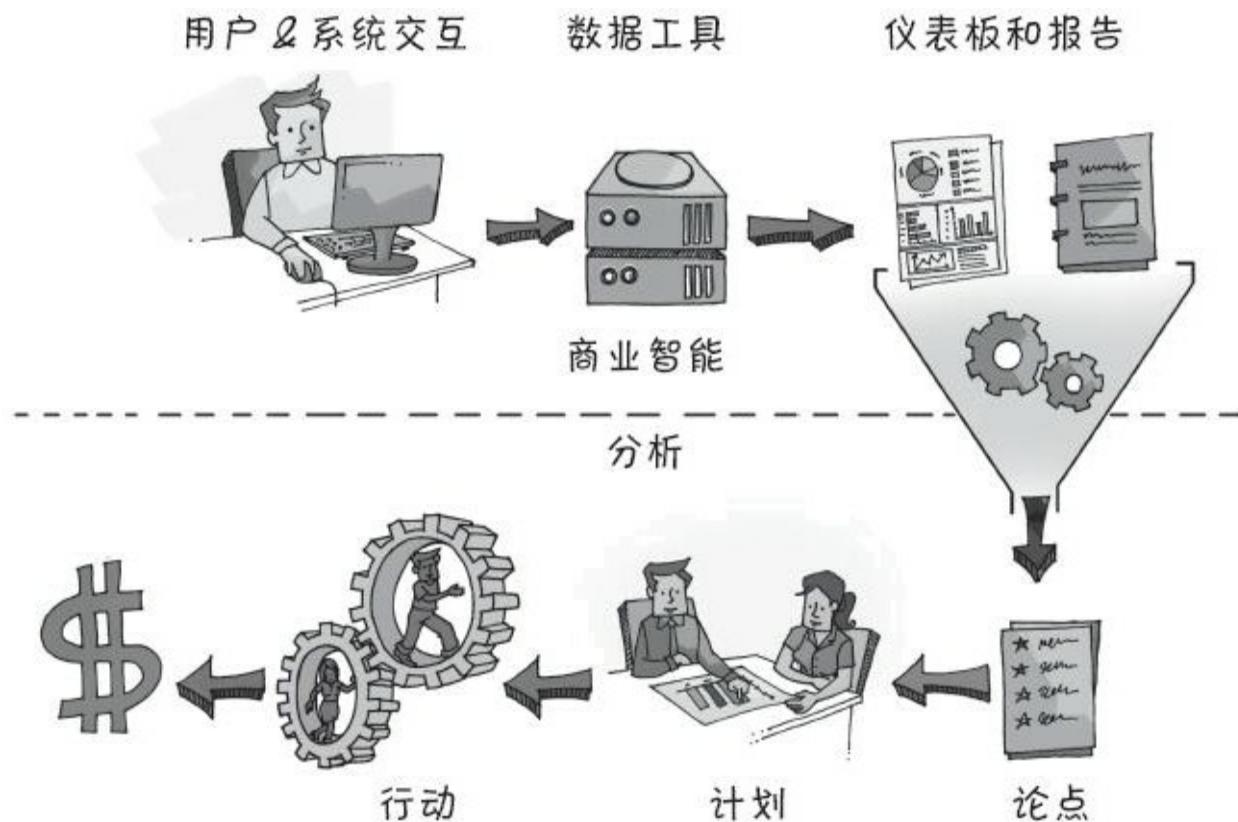
我们鼓励在任何数据驱动决策的过程中首先使用商业分析，因为它可以通过有限的投入获得高额投资回报率。随后对于一小部分预期收益可抵消高昂初始投资的问题，再使用高级分析，这样才能得到可接受的投资回报率。

分析，不是报告

报告和商业智能（BI）经常与分析相混淆。商业智能和分析实际上是两个不同的过程，涉及不同的工具，也出于不同的目的。当用户与某系统交互时（例如，当你从当地超市购买了1加仑牛奶时），数据将被生成、收集、清理，然后通过Teradata、Hadoop和Oracle这些数据仓库解决方案被存储下来。接下来，可以通过报表，但现在越来越多的是图形仪表板来访问数据。商业智能包括从收集数据到访问数据的所有操作。

此外，商业智能提供的数据将被分析所用。然后，分析将数据转化为论点、决策、行动，并且最终到收入或其他影响指标。SAS、R和Excel等分析工具可用来将数据转换为论点。你会注意到，其中一些（如Excel）并不是特别深奥的工具。

例如，某个玩具店的仪表板告诉经理，本周的收入比上周下降了10%，客流量下降了2%。通过分析发现，某个分部比如常客，比上周下降了2%。这表明该问题与常客的参与度有关，可引导商店经理关注相关领域的业务。



分析与测试

分析是通过研究业务事件和分析历史数据得出论点，而测试则是当你没有作为决策依据的历史数据时所进行的受控实验。以电子商务网站的红色结账按钮为例。基于焦点小组的输入，产品经理坚定地认为，蓝色结账按钮将成为用户首选，可以带来更多的结账和购买。但是，该网站从来没有用过蓝色结账按钮，所以没有历史数据与红色结账按钮的效果进行比较。这个业务问题无法通过分析解决，它需要测试。

因此，产品经理设置了一个实验来测试他的假设。他创建了一个测试网页，将部分网站流量暴露给红色结账按钮，另一部分暴露给蓝色结账按钮。然后，再观察客户的付款行为。现在，他有了实验数据可以用来分析，以确定蓝色按钮是否真的可以提高实际购买率。

此外，分析只能证明是否具有影响关系 ($A \leftarrow \rightarrow B$)，而测试则可以证明是否存在因果关系 ($A \rightarrow B$)。在我们的示例中，假设产品经理通过分析已经看到平均订单量和网页加载时间之间的关系，即加载时间越长，订货量越小，但是他还不知道的是订单大小在影响加载时间还是加载时间在影响订单大小，或两者都共同受其他因素影响。如果他真的想了解原因，就需要设置一个测试实验来改变加载时间，并查看它对订单大小的影响。

总之，分析是对过去数据展开分析，以获取论点并找出数据之间的关系。测试是通过受控实验来创建新的样本数据，以获得论点并证明因果关系。

增长黑客推动用户增长

什么是增长黑客（growth hacking）？它是一种组织内部程序，以推动某个增长指标实现快速可扩展性增长为唯一目标。Facebook通过增长黑客将消费者从4500万增长到10亿，而Twitter则利用增长黑客将用户群扩大了10倍。包括LinkedIn、Quora和PayPal等在内的其他公司也已采用并坚持采用增长黑客策略。简单地说，增长黑客的要义就是要突破营销预算的限制，去创造产品市场本身。

大多数企业同时关注多个指标：收入、用户黏性和许多其他类似的战略指标。与此相反，许多技术型的初创公司现在正在组建增长黑客突击队（SWAT team），只瞄准某个指标的增长，比如客户数量。本书第8章

介绍了3个关键问题框架，一旦增长指标被锁定并达成一致，通过这个框架，公司就可以找到增长的影响因素。

为了实现目标，增长黑客团队作为公司内的启动部门而发挥作用。他们是集少数专家，用小额预算快速做出决策的创新型团队，通常涵盖了产品开发、设计、分析和营销专长，努力确认出增长指标的影响因素，然后实施以推动指数级的增长。最为关键的原则是：这种增长是可扩展的，而不是一次性的。使可扩展性成为可能，要通过大量实验，快速学习，从所学中进行连续迭代，然后再频繁地实施以学习更多，所有这些都是通过创造性分析来实现的。

这样做的结果是通过汇聚的方法将所有领域的投入都聚合进来，全面加速增长目标的实现。增长黑客团队使用“拉动”战略来吸引消费者。也就是说，它通过产品体验来提高参与度，从而了解用户的动机和行为，并使参与客户立即感受到价值。这种方法不同于传统的模块化方法，营销部门不必等产品组构建了一个产品版本之后才联系接洽客户。在确认客户行为数据的模式和持续验证产品变化影响方面，分析是一个强大的工具。

Twitter对其现有用户群的行为分析表明，如果新用户在Twitter上使用简单的登录操作关注了至少10个人，他们就更有可能留下来。于是，Twitter的策略便是开发一个新功能，向新用户推荐可关注的前10名用户列表。并且，它随后还添加了更多功能，帮助新用户获得其他人的关注。这大大增加了用户回访的概率，并且越来越多的用户选择留下来。Facebook早期的增长黑客的操作是，允许用户在个人的网站和博客上嵌入自己的Facebook小插件，如徽章和个人资料，使Facebook超越了自身的局限。此举使Facebook在互联网上获得了令人难以置信的关注度，吸引了数百万用户前来注册。而对已有用户来说，又可以很方便地通过自己的门户网站积攒人气。

增长黑客不是公式化的程序，这正是它的强项。它是为企业、企业目标及其产品定制的解决方案，需要组织领导和超越他们本身技能的边界的团队成员的共同支持。

无须大肆宣传

大数据就像房间里的大象一样显而易见，是我们不容忽视的。事实上，它在分析情境中通常不是很相关，不过人们还是经常使用这个术语。原

因如下。

什么是大数据

大数据通常可用三个v开关的词来解释，即很多种类（variety）的大量（volume）数据在高速（velocity）流动。这三个v开头的词引起了存储和可视化的问题。数据分析传统上处理的是结构化数据。大数据的种类多样性是由结构化数据（可以存储在列和行中）的膨胀和非结构化数据（图片、文本、电影）的爆炸所导致的。传统的商业智能系统可以很好地处理结构化数据，但非结构化数据却是一个问题。数据量庞大这个问题已经存在很长时间了，但它只是另外两个v使得传统的商业智能系统面对大数据出现不稳定之外的另一个问题。大数据是一个商业智能问题，而不是分析问题。大数据很难被存储或渲染，因此需要特殊的工具和技术（所以有了大数据炒作）。但是一旦商业智能可以解决大量数据的结构化问题了，那么，对这些数据展开分析也就不成问题了。

如果操作得当，分析是从识别大数据的子集开始，通过假设智能地缩小范围，就像神探福尔摩斯寻找太平洋的宝藏一样。

因此，大数据不是分析的同义词，我们不会在本书中谈论大数据。我们将讨论如何利用你可以得到的数据做出更明智的决策。

结论

我们不能想象采用哥伦布的方法来寻找宝藏的景象，但我们已经看到许多企业完全遵循这种探险家思维去对待数据，在信息的汪洋大海中苦苦挣扎，浪费了大量的时间和资源，却对真正的商业价值一无所获。

但与之相对，高效管理者和分析师可以在他们搜索与问题相关的答案时，使用有指导性的福尔摩斯探索性数据驱动方法来指引他们努力的方向。毫无疑问，这些管理者和分析师发现了一座金山，并影响企业的价值。

在第3章中，我们将深入探讨刚才谈到的一些分析方法。

小结

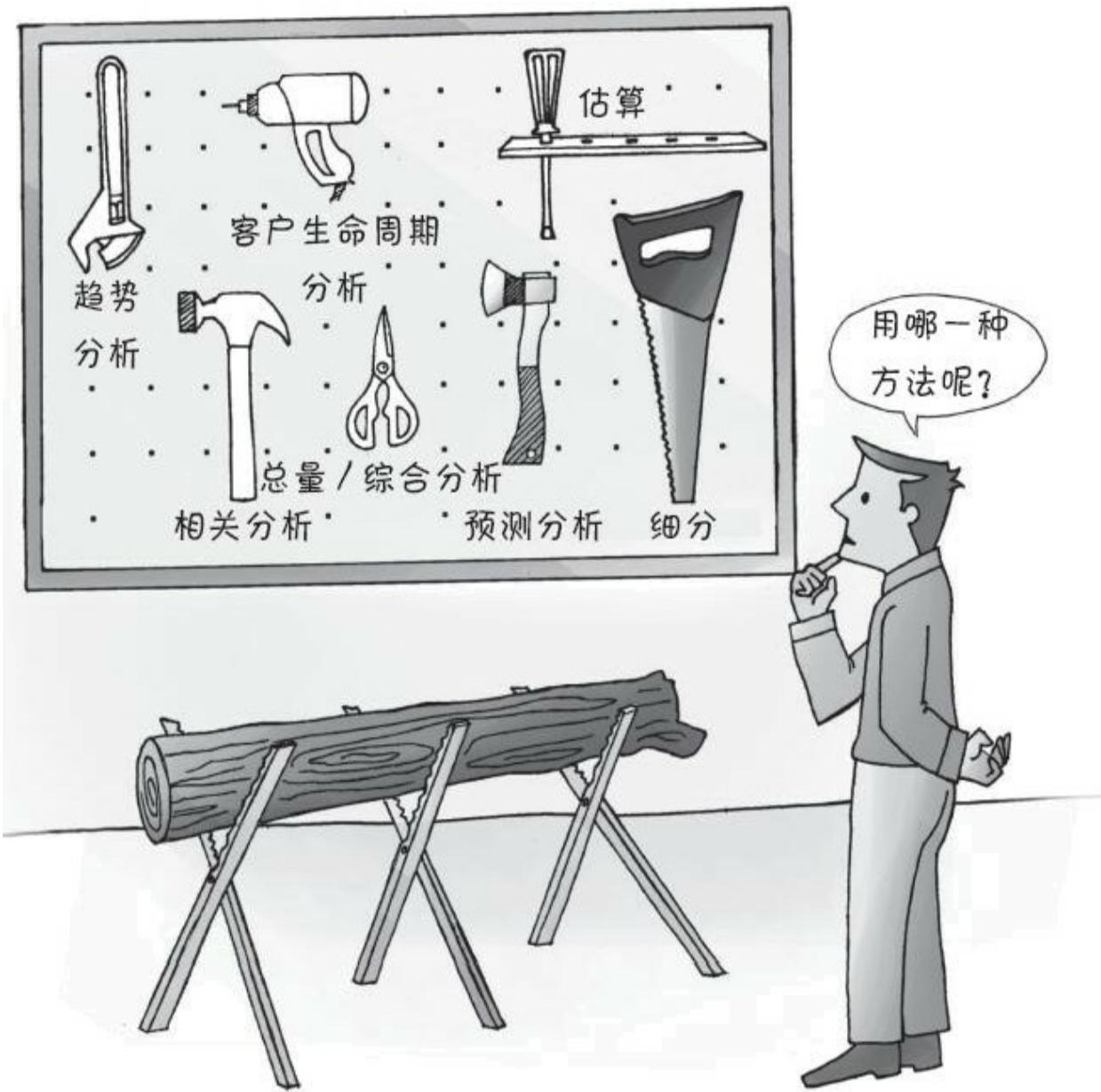
- 分析=数据科学+决策科学。

- 分析没那么高深复杂，任何商业专业人士都期待利用数据来做出更好的决策。
- 分析不是报告。
- 分析不受大数据挑战的约束（如果操作得当，分析永远都是处理数据子集）。

第3章 七大分析方法

本章将讨论：

- 七大常用的分析方法。
- 每种方法的适用性。
- 各种分析方法的典型应用案例：营销、产品和客户服务。



假设你是山葡萄酒厂的老板，小葡萄酒厂生意不错。除了葡萄酒的生产和销售，你还在美丽的葡萄园为人提供婚礼服务。现在，婚礼业务的年收入为100万美元，你正在希望增加业务量。你现在花费27000美元的营销预算购买了各种促销工具，主要是搜索引擎和在线的结婚门户网站，但你对于是否能从中获得很大收益并不自信。哪种营销工具能达到最好的效果并能促进销售呢？

这是一个简单的商业分析挑战：营销渠道优化。你需要做的全部工作就是分析你投资的每个促销工具的运营结果，比较成本与收益，看哪种方式可实现最低价格达到最好效果，以确定最有效的营销方案。

你需要首先提出一些试探性的问题：什么渠道会比较好？对于这些资源，我们应该投资多少？是否应该改变营销组合，以获得更好的回报？

你的目标是找出商业问题：好的线索来自哪里？我们如何更多地得到这些好的线索？下一步是提出假设：这些问题的答案是什么。接下来，再通过分析确定最佳解决方案。

每个企业都有像山葡萄酒厂这样的情况——可能是多种情况。这些都是应用简单数据分析的理想情况。它不必是耗时的和昂贵的，并且肯定不是高深复杂的。在第5章中，我们将会进一步了解稍微高深复杂点或者说魔幻一点的预测分析，然后你就可以自行判断了。

最常用的七种分析方法

当面临涉及数据的管理决策时，你可以在任何时候使用分析方法。要解决常见业务问题，涉及的分析方法并不多。具体采用哪种方法将由你所面临的特定问题类型来决定。下面我们将介绍常用的分析方法，以及它们的使用背景。最常用的七种方法如表3-1所示。

表3-1 七种最常用的分析方法

分析方法	描述	应用
综合 / 总量分析	用于描述总体或分部，或对两个分部进行比较	描述性分析、性能分析、深度营销分析、赢家输家分析
相关分析	寻找两种或两种以上事物之间的关系，以期能够用一种事物解释另一种事物或者驱动另一种事物	前后对比，控制测试、驱动因素、仪表盘
趋势分析	一段时间内进行的总量分析或相关分析，即在一段时间内的趋势	销售、收入趋势；趋势的改变或分部趋势，或在一段时间内的影响因素
估算	缺乏历史数据时，用结构性方法来做一个近乎准确的推测	内部数据有限的商业案例，依赖外部数据和假设的商业案例
预测分析 / 时间序列	根据当前和历史数据，对未来事件的预测	转换、消费者参与的影响因素、预测
细分	将客户或产品进行有意义的划分，确定更好的目标市场，通过用户定制实现更高价值	对客户或产品分组，以实现目标和定制
客户生命周期分析	着眼于购买过程的不同阶段，可以确定某类客户处于什么阶段，以及知道如何将他们转移到下一个购买阶段	客户进展阶段，包含考察、购买、使用及销售渠道

综合/总量分析

综合/总量分析是最简单和最常用的分析方法，也是大多数其他方法论的第一步。它常用来描述一个总体或一个分部，或对两个分部进行比较。所谓总体，是你想了解的事物或人的集合。它可能是客户或前景，

也可能是产品，如某月的产品或在你网站上的一个特定种类里出售的项目。

在山葡萄酒厂，你也想试图了解通常是哪些人在预定婚宴，以做好交流准备。你查看过去三年的预订记录，如在300个客户中，对年龄、性别和位置做综合分析。分析显示，他们之中85%为女性，平均年龄为33岁，其中有60%的人住在俄勒冈。

现在你可以使用这个结论来准备你的对话交流了，也可以将俄勒冈设定为你的营销目标。你在哪里可以找到30多岁的女人？是否有一些商店专门服务于这些人？他们偏爱什么样的出版物？他们在某些餐馆吃饭吗？你能在这些地方做广告吗？

除了描述总体以外，综合分析也可用于对两个分部进行比较。

你想解决的另一个商业问题是将填写婚礼预订在线表格的人转换为你的客户。你可以对两种表格的使用结果进行测试：短表格和标准表格，看哪种表格能争取到更多的客户。

总之，综合分析可以用来回答描述性或比较性的商业问题，如：

- 我的客户是谁？
- 不同地区的客户有什么不同？
- 年轻人与老年人相比，会更多地通过平板电脑访问我们的数码产品吗？
- 在上一次的营销活动中，哪些措施起作用，哪些不起作用？

相关分析

相关分析通过寻找两种或多种事物之间的关系，希望实现用某种事物解释或影响另一种的目的。

在山葡萄酒厂寻找好客源时，你可以将用户提交的询价单作为客源的代理变量。然后，就可以把最初的假设缩小为四个最合理的假设：

- 通过某些葡萄酒指南和谷歌付费搜索能找到更好的客源。

- 移动用户有更好的客源。
- 看到定价页面会破坏客源。
- 某些地点能产生更好的客源，比如俄勒冈当地。

你对每一个假设使用相关分析：影响是否成交的因素包括流量来源、移动用户与非移动用户、人们看没看到定价页面，以及客户的地理位置。通过这一分析，你能够迅速否决两个假设，留下两个有待进一步探索：不同的流量来源确实会影响最后是否成交，地理位置也是一个很大的因素。

表3-2是一个例子，告诉我们不同网站来源的访问成交比例和相关的相关分析。

表3-2 转换数据和相关分析

网站来源 / 中位数	询价成交率 (%)
谷歌 / 自然搜索	4
谷歌 / 每次点击成本	4
必应 / 每次点击成本	1
(直接) / (无)	2
myportlandwedding.com/	5
facebook.com/ 引荐	0
必应 / 自然搜索	4
雅虎 / 自然搜索	5
stinnocentwine.com/ 推荐	4
wineryweddingguide.com/	10
weddingwire.com/ 引荐	11
eolaamityhills.com/ 推荐	0
apps.facebook.com/ 推荐	0
vibranttable.com/ 引荐	8
google.com/ 引荐	7
其他：长尾	3
平均	3

根据商机资源的好坏来重新分配预算，可以帮助你增加收益。第一，对商机薄弱的网站来源，要减少投资——必应搜索、Facebook页面和某些葡萄酒指南的每次点击成本。第二，增加对三个特定访问工具的投资，它们都是能够带来高效商机的网站来源：谷歌付费竞价排名和某些婚礼指南网站。通过简单的相关性分析，你会发现同样还是使用27000美元的营销预算，却能实现了山葡萄酒厂12%（即120000美元）的收入增加。

额。当然，要注意每个渠道的饱和情况，以及询价成交率在一定极限后会出现下降。

在我们的经验中，如果借助现有的数据，一个类似这样的数据分析只需要两个小时，完全可以由一个懂得业务背景的非专业分析者在Excel中操作完成，实现小投入获得大收益，没有高深复杂的科学，只是坚实的商业分析，任何专业的人都可以完成。我们将在第4章中详细介绍如何做到这一点。

总结一下，相关分析用来回答这一类商业问题：

- 为什么上季度的收入低于预期？（通过找出与收入不足相关的部门。）
- 为什么曼哈顿的威瑞森商店的客流量下降？（通过寻找影响客流量的相关因素。）

趋势分析

趋势分析是一段时间内的综合分析或相关分析，即分析一段时期内的趋势。它通常用于观察销售业绩或收入增长在一段时期内的变化情况。我们的目标是识别趋势中的反常情况，以及查明在那段期间的受影响部门和影响因素。它被用于回答像这样的问题：

- 我们的客户群已经转移到年轻人身上了吗？（按时间的推移的年龄趋势。）
- 为什么增长趋势在下滑？（通过查看一段时间内的增长趋势，并按不同的部门划分，找到令增长放缓的相关因素。）
- 为什么iPhone 5销售放缓？（通过查看iPhone 5在过去几个月的销售情况，识别影响销售的内部和外部的相关因素。）

估算

估算是在缺乏历史数据时，进行近似估计的一种结构化方法。从技术角度看，估算并不属于一种分析方法，因为并未使用历史数据。但由于它采用基于假设结构化方法以及利用有限的外部和内部数据点，并在商业决策中被广泛使用，因此我们将它作为分析的一部分。估算常被用在进

入新市场的典型商业案例中，了解尚未发布产品的潜在适销对路市场空间，并迅速扩大决策的影响或变化。估算通常用来回答下面这样的问题：

- 在美国，每年销售多少网络路由器，其中成功卖给消费者的比例是多少？
- 在洛杉矶，每年销售多少婚纱，我们能占领多少市场？

预测分析

预测分析利用当前和历史数据去预测未来事件。FICO（财务模块）评分和天气预报都是预测分析常见的例子。预测分析利用的是未来某个时点的目标值与当前或过去的其他相关指标之间的基础相关性。通过对当前和历史相关因素的观察分析，去较为准确地预测某个指标的未来状态。

时间序列是预测分析的一个特殊应用，它不使用其他的相关因素，而是使用目标值自身的历历史数据进行估值。这种关联关系称为自相关或与自我关联。

预测分析可解决的问题看起来与相关分析的非常相似，举例如下：

- 客户流失的影响因素是什么？
- 什么因素促使客户参与？
- 为什么成交率会下降？是哪些因素造成的？
- 什么环境因素导致了人类自身免疫性问题？

是的，即使是在疾病诊断中，也可以应用预测分析。在通常情况下，预测分析相当复杂，已接近高深复杂科学的边缘。在第5章中，我们将更详细地介绍预测分析，同时还会向想要进一步了解预测分析的读者介绍更多的其他资源和培训知识。

细分

细分是一种分析方法，它将客户或产品进行划分为有意义的分部，通过

客户定制，驱动更高的价值，实现更好的目标。被划分到同一个分部的人或产品彼此相似，与其他分部的特征具有明显差异。

比如，将客户分为当前客户和潜在客户就是一种简单的细分。以亚马逊公司为例，你还可以进一步根据他们所购产品类别进行细分，如服装买家和图书买家，正如你所能想象的，细分的方法有很多种，这取决于为什么要进行细分。简单的细分方法有很多，如RFM（最近购买、频率、金额）分析、产品版本与人口统计。在营销中，RFM是一种最常见的细分技术。它的标准方法是按这三个变量各自划分三个层次（低、中、高），然后将客户分为27类。

这里还有一个利用产品版本进行细分的例子。如果你进行的是一种营销活动，出售刚刚发布的消费者软件9.0版本，那么消费者群可以按照潜在消费者、试用下载者和1.0版本到8.0版本用户进行划分，看看谁是最好的采纳者。你可能会惊讶地发现，版本跳级的用户（像那些使用6.0版本或7.0版本的用户比使用8.0版本的用户）更容易升级产品。

细分可以使用更先进的统计方法，即聚类，这种方法可以接受大量变量，抵消不同组别或类别之间的内生性影响，让不同类别之间的差异更突出。

当你的目标是回答像这样的问题时，你就需要细分技术了，如：

- 如何定制我们的产品或服务？（向谁提供哪些产品？）
- 如何定制购买激励措施？（向谁提供10美元折扣，向谁提供50美元折扣？）
- 我们的产品组合是什么样子的？（更有利可图的产品组合与其他组合比较会有何区别？）

客户生命周期分析

客户生命周期（CLC）分析是判断客户购买过程的不同阶段，以确定客户群目前正处于什么阶段，并且决策如何将他们转移到下一个阶段。

组群分析（cohort analysis，也称断代分析、人口特性分析）是一种特殊类型的客户生命周期分析，它将客户起始日或激活日进行分类。例如，

7月群组可能需要15天来适应你的产品，但12月群组可能需要30天。销售管道分析也是客户生命周期分析的另一个应用。客户生命周期分析回答这样的问题：我们的销售成交可能是什么？也就是说，在呈现出商机的人群中，有多少符合资格、有多少成为机会、有多少会试用产品、有多少会成为客户？客户生命周期分析还可以回答下面这样的问题：

- 客户是如何改善我们的产品的？
- 苹果产品的用户是如何体验苹果产品的？也许一些是先使用iTunes，然后是iCloud，在大概90天左右会购买，然后一年以后可能购买iPhone。

客户生命周期分析通常与细分一起共同发挥作用。你可以想象，通过iTunes获得的苹果用户会与通过企业Mac电脑获得的苹果用户具有不同的客户生命周期。

现在你已对常用分析方法有了大致的了解了，让我们用例3-1来试试你掌握得如何。

例3-1选择适当的方法

你会选择哪种或哪些方法来解决这些问题？答案在书末的注释中。

- (1) 为什么一项产品发布以后，成交率却降低了？
- (2) 纽约州有多少所小学？
- (3) 确定在过去几周内，“玩具及所有产品”的收入增长是否放缓？如果是，原因是什么？
- (4) 你能告诉我哪一种策略在上一次的市场营销活动中效果最好吗？
- (5) 我们伦敦办公室的员工比新加坡办公室的员工更年轻吗？
- (6) 我们的客户从听说我们到下载免费游戏，然后再付款购买高级功能的时间周期是怎样的？
- (7) 在我们的100万个客户里，我应该给其中哪20万人发送未来的营销方案，以获得最佳的投资回报率？

(8) 客户使用我们的打印机产品有哪些不同的用途？这对我们来说意味着什么？

分析方法的应用

让我们看看这些方法在营销、产品和客户服务等商业领域的一般应用情况。

如果你在营销部门

考虑行业的广度：金融服务、消费品、电子商务、汽车、技术、媒体等。一位首席营销官（CMO）对业务计划通常期望以下三个关键结果。



从本质上讲，首席营销官希望寻求更广泛的首选目标销售漏斗，实现每个阶段更高的转化率。这将能使收入最大化，实现最佳投资回报率。以山葡萄酒为代表。公司通过以下这些问题，可以让最优销售漏斗得到数据的支持。

- 向谁和到哪里营销？
- 在每个渠道上花费多少钱？
- 什么因素会影响响应和成交？
- 谁最能回应什么样的信息、优惠和产品？

- 什么导致客户流失？

虽然这可能看起来有点像预测分析，我们认为使用简单的业务分析技术，首席营销官可以实现更好的投资回报率，以做出有见地的明智决定。表3-3提供了一个示例：如何使用非常简单的分析方法即可实现目标（具体的分析方法在第二列）。

表3-3 市场营销的商业分析技术

1. 以更具成本效益的方式带来更多的未来客户：	
• 根据现有客户简介，通过识别新渠道增加适销对路的市场	综合分析 估算
• 根据过去的营销活动采用更好的目标信息和优惠以提高响应	测试 相关分析
• 优化渠道，提高投资回报率，降低客户获取成本	相关分析
2. 获取更多的上门客户：	
• 识别转换影响因素。是否要做某些履行选项、用户体验、审查选项、购物车选项、付款选项、优惠和促销来提高客户转化率	测试 相关分析
3. 为维持当前客户，需要：	
• 细分客户，促进他们的参与	简单细分：RFM

(续)

• 发起鼓励客户参与的营销活动，刺激不同客户群的购买	
• 了解影响不同客户群的参与因素（如某些优惠、折扣、捆绑销售、忠诚会员等）	相关分析
• 分析营销方案的效果：符合客户心意的和不符合的	测试 综合分析和相关分析
• 了解客户流失的影响因素：识别造成客户不选择你的业务的因素	相关分析

如果你在产品管理或业务部（BU）

首席产品官（CPO）或业务部负责人在领导产品小组或业务部门时会面临类似的挑战。同样，他可以使用许多类似的技术和策略（见表3-4）。

表3-4 首席产品官或业务部负责人的商业分析技术

1. 为各种客户群识别新的产品和功能：了解每一部分的消费者需求，并提供针对性的产品：	
• 划分客户群，了解需求差异（根据过去的产品使用、人口统计等）	简单细分
• 识别不同客户群的产品和特性差异	测试
• 优化新产品想法或特性	估算
2. 考虑优先包括哪些产品功能，这可以通过了解预期的业务影响来确定	估算
3. 通过激励消费者采取行动（如购买产品或注册）来优化客户体验，提高产品使用率	测试 相关分析
• 识别阻碍用户成功使用你的产品的因素	测试 相关分析

如果你在客户服务部门

同样，客户服务业务主管也可以利用分析来解决关键业务问题（见表3-5）。

表3-5 为客户服务的商业分析技术

1. 优化最珍贵的资源：通过客户细分优化 FCR（首次呼叫解决）和 NPS（净推荐值）等指标，以及成本	相关分析 简单细分
2. 圆满解决客户来电问题，这反过来意味着了解首要问题和解决驱动因素	综合分析 相关分析
3. 谁是最有价值的客户？每个客户群的特点是什么？每个客户群的保障援助等级是什么	综合分析 相关分析
4. 在呼叫前转移客户到自我服务，以减少客户服务开销，这需要了解客户呼叫和创建有效自我服务的主要原因	相关分析

总结

在大多数情况下，简单商业分析是非常有效率的，你不用等待分析团队来给你答案。在多数情况下，你需要的数据已经在手边，你只需要将它们找出来，正视它们就行了。

在接下来的章节中，我们将通过一步一步的指导，告诉你如何靠自己进行商业分析，为你的业务做出更明智的决策，采取更好的行动。

小结

- 80%的业务问题都可以使用四种最常见的分析方法来解决，而且都可以通过Excel来完成。
- 综合分析用于描述或比较。
- 相关分析常用于发现事物之间的关系，用一件事物去影响另一件事物。
- 趋势分析用于分析随时间推移的趋势。
- 估算利用有限的内部数据去产生精确的估计值。

>> 【看书累了吗？休息一会！更多新书朋友圈 每日免费分享微信 xueba987。渺沧海一粟：为终身学习者赋能 2019年7月】

第二部分 深入理解

第4章 BDIR：商业分析的五步框架

第5章 预测分析，并称高深学科

第6章 数据和分析工具

第4章 BDIR：商业分析的五步框架

本章将讨论：

- 秘方：从“数据到决策”的五个步骤。
- 如何使用能解决80%商业问题的四种方法来分析数据。



SmartShoes是一家时尚鞋类和配饰领域的新兴在线零售商，它的增长速度令人难以置信。然而，一个新的配件部门，在强劲的开始后不仅失去了发展目标，收入也在节节败退。

首席执行官汤姆停下手头工作和分析总监杰夫见了个面，汤姆说：“我想了解部门和全国近三年的配件销售收入数据。”一个星期后，杰夫将资料交给汤姆并汇报说：“整个美国的配件营业收入都在下降。”汤姆怒不可遏，问：“这就是你能找到的信息？这些我早就知道了。我现在没有时间，我们明天早上7点再谈。”

杰夫出师不利。他或许知道如何处理数据，却错过了关键的第一步。“深入真正的商业问题”这才是汤姆真正的要求。这是一个常见的商

业失误。虽然杰夫在利用数据和分析方面很用心，并且花费了大量的时间和金钱去构建复杂的分析模型，但他仍然没有找出症结并解决问题。

本章将向你介绍一个BADIR分析框架，它将通过一组五个精简的步骤，简化从数据到决策的过程，用简单但功能强大的分析解决80%的商业问题。这里的BADIR，分别代表商业问题（business question）、分析计划（analysis plan）、数据收集（data collection）、得出结论（insights）和提出建议（recommendations）（见图4-1）。



任何经理都可以依据这五个步骤有效地执行分析解决方案，在一般情况下，都无须使用复杂的方法。我们的经验表明，如果在企业中没有得出分析结果，通常是因为我们跳过了一个或多个步骤，或没有按正确的步骤执行。例如，商业问题可能尚未被解释清楚，或者利益相关者无法找到有意义的假设去进行探索和测试。学完这一章，你将能够通过遵循这组框架去解决大多数需要进行分析的商业问题。

如第2章所讨论的，有效分析是数据科学与决策科学的结合。从技术层面来看，要求具有与数据科学相关的技术技能，得出研究结论。从商业

层面来看，要求能够利用决策科学去影响受众和理解商业的软技能，从而产生影响。商业层面的要求经常被渴望基于数据得出结论的组织所忽略。BADIR框架通过将这两个强大的科学结合一起来达成效果。

以影响为目的的分析=数据科学+决策科学

步骤1：商业问题

在第一步，你可以获得：

- 减少重复劳动。
- 提供可行性建议。
- 被认可为决策过程中的合作伙伴。

你可能很快就注意到，虽然BADIR是数据分析的过程，但它并不是从数据开始的。这个框架是从理解数据需要回答的真正商业问题开始的。而且，不是任何问题都行。提出错误的问题，可能导致无用的解决方案。越接近分析需要解决的核心商业问题，就能越快得到正确的答案。正确的商业问题能做出更明智的决策并产生影响，让你直接走上通往既定目标的康庄大道。通过减少重复劳动，加速整个过程，并帮助提供可行性建议，从而实现商业价值。因此，你和你的团队将被视为有价值的商业合作伙伴。但是这一切都始于花时间去了解现在正在发生什么，并提出相关问题。

想要发现真正的商业问题，就要了解问题背后的意图和商业考量。

从既定目标到真正商业问题的框架

传统的六个问题框架为：何事、何人、何处、何时、何缘由以及何方法。



一开始就提出问题，目的是设身处地找出问题所在，称为神探福尔摩斯法，或侦探性方法。如果没有最终找出真正的问题，就像是在整个海洋（数据）中寻找黄金那样，称为哥伦布法，或探索性方法。在这里，你应该使用侦探性方法。提出相关问题，使你马上就能够了解当前因素、过去事件或推动分析请求的未来战略（见图4-2）。

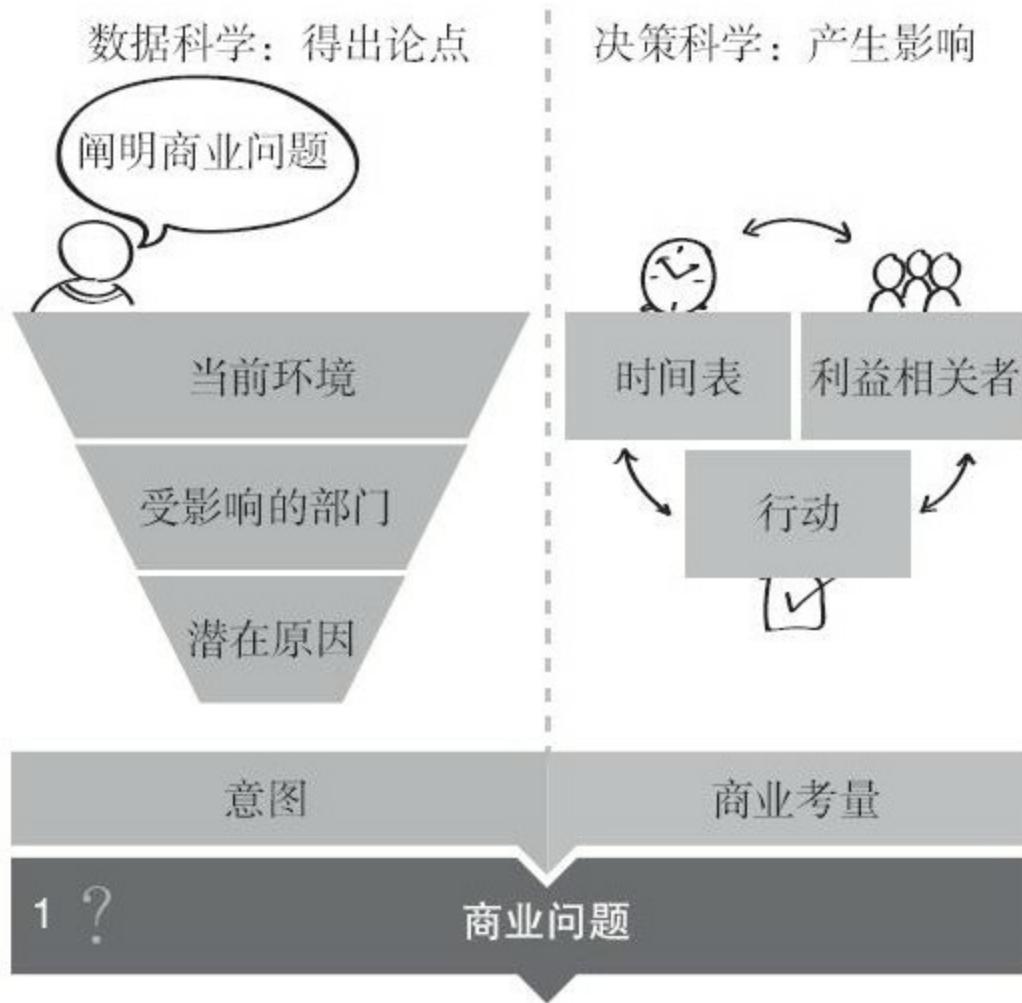


图4-2 步骤1：商业问题

步骤1：商业问题——数据科学层面

结合当前环境、受影响的部门以及分析需求背后的潜在原因，才能明确需求背后的意图（见图4-3）。



图4-3 提出问题以确定意图

步骤1：商业问题——决策科学层面

结合你对利益相关者的了解和他们准备在既定时间段内采取的行动，你就能够确理解问题背后的商业考量（见图4-4）。

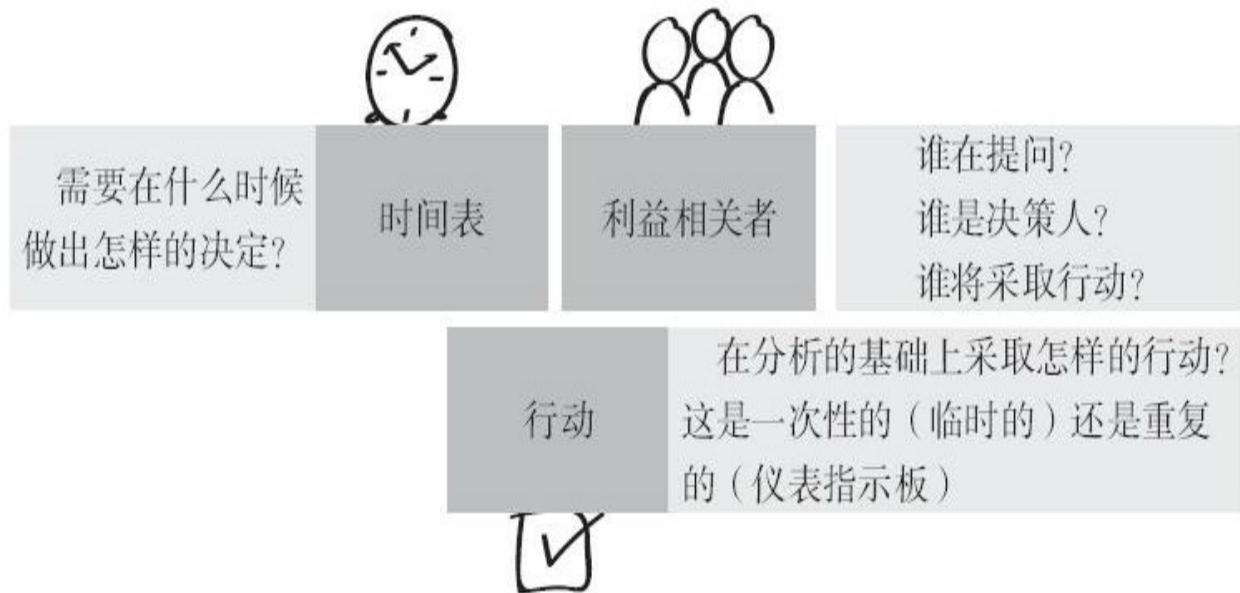


图4-4 提出问题，理解商业事项

“需要做什么决定”和“在什么时候做出决定”等一系列问题的答案将告诉你，你有多少时间来完成你的分析。你是否需要在一夜之间或3个月之

内或某段时限内完成些什么？通过了解“谁在要求我们做分析”“谁会受分析和后续建议的影响”和“谁会根据结果采取行动”等，确定利益相关者。最重要的是，通过分析，还可以了解你的利益相关者准备采取什么行动。

如果通过对这些问题的思考，你发现利益相关者无法采取行动，那么则无须进行此项分析。如果正确完成步骤1，能过滤掉十之八九的数据和结论的需求，因为这些信息或许证明“知道也挺好，但是没什么用”，因为它们不能回答驱动商业发展的关键问题。

将意图与商业考量相结合，有利于将技术层面与商业层面合并，从而改进真正的商业问题，更好地瞄准你的分析。这样做可以节省时间，增加相关性，并有效地利用资源，从而降低成本并减轻工作量。

简单地说，提越多越好的问题，那么你的分析结果就会越好。在步骤1中，要深入商业问题，有一些事项需要特别注意（见例4-1）。

例4-1在寻找商业问题时应避免的步骤

×请勿.....

.....在这个阶段尝试提出解决问题的方法。

通过以后彻底调查，你也许会发掘对你的业务有价值的信息，而上面的尝试会先入为主。

.....提出诱导性问题。

这只会导致答案偏颇或根本不会得到答案。反过来，这种错误的分析会歪曲你提供的见解和建议。

Pets & Pets是一家英国在线宠物资源公司。它有良好的口碑，以极具竞争力的价格为客户供应有尽有的宠物产品，还是宠物和宠物产品信息资源的可靠来源。多年来，它已经积累了在线访问者的大量信息。产品总监（CPO）布兰妮要求分析总监亚历克斯按照国家、地区、产品和功能的不同，统计过去10年的在线销售成交数据。

亚历克斯意识到此举涵盖了Pets & Pets的整个历史，将会涉及海量的数据，他先提出了一些问题，弄明白为什么布兰妮想要这些信息。亚历克

斯了解到，在他休假期间，公司针对狗绳在线销售流程进行了新功能测试，以提高访问成交率。但是，该功能实际上导致访问成交率下降。虽然只有1%的英国网站访问量受到了影响，但是每月的损失接近60万美元。

亚历克斯还了解到，新功能的漏洞比工程师承认的还要多，在设计上还有些问题。总裁（CEO）给布兰妮施加了巨大压力。布兰妮需要决定是否取消测试并放弃这个新功能，或者通过调整来立即提高访问成交率。放弃并不是理想的方案，因为布兰妮的团队已经在这个引发高度关注的项目上工作了两个多月。她倾向于后一种选择，但她想知道要做什么必要的调整，才能控制住访问成交率下降的现象。

布兰妮将在两天后面见总裁并做汇报。为了完成这项调查，亚历克斯将真正的商业问题缩小为“在英国推出狗绳在线结账功能后，访问成交率下降的原因是什么”和“产品开发、项目管理和质量保证部门应采取什么措施来减少损失”？

请注意，明确表达的商业问题与布兰妮刚开始提出的要求有很大区别（见图4-5）。在找出商业问题的同时，亚历克斯也发现了这一需求背后的影响因素——各种商业考量、利益相关者和时间要求。如果没有切中问题，他会发现自己跟SmartShoes公司的杰夫一样，会浪费时间和资源，反馈给布兰妮大量数据，而布兰妮肯定会奇怪怎么亚历克斯就是不能理解她的要求呢。

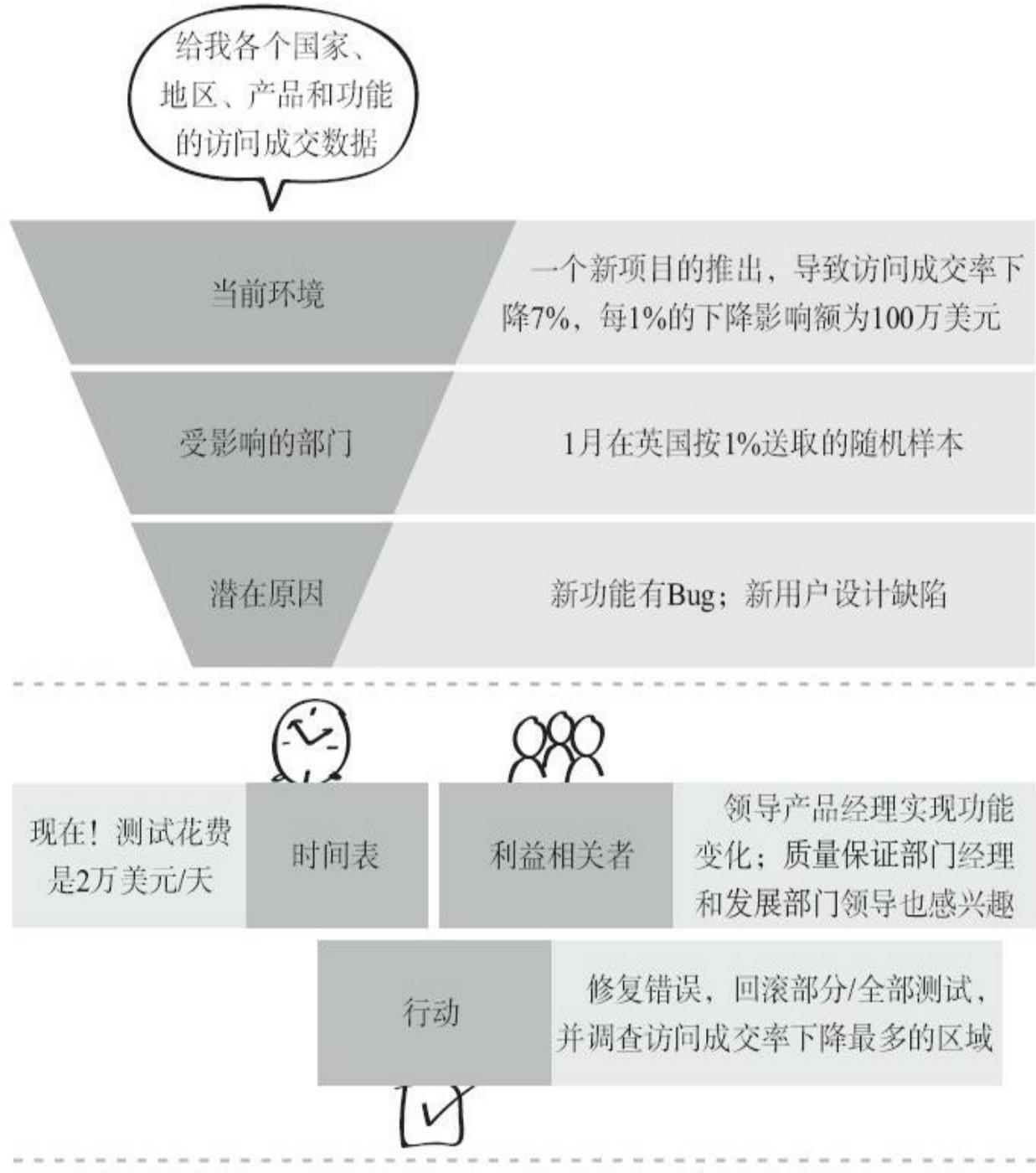


图4-5 Pets & Pets商店：阐明业务问题

步骤2：分析计划

在这一步，你可以获得：

- 与关键利益相关者进行协调。
- 管理范围。
- 根据达成一致的时间安排分配资源。
- 确定分析任务完成的范围和时间。
- 影响的视野。

回到SmartShoes公司的问题，汤姆已经冷静下来。杰夫现在明白，对于配件收入下降，尤其是在美国市场的惨败，汤姆要对董事会负责，这是他需要立即解决的问题。通过相关询问调查，杰夫也终于弄清了真正的商业问题是“为什么美国市场的配件收入会减少以及我们可以提出什么解决方案”。明天上午9点要开董事会议，汤姆需要在这之前拿出一个翔实可靠的行动计划。

时间紧迫，杰夫和他的团队又一头扎进工作中。他指挥团队去查看所有关于配件销售的数据，然后建立了一个预测模型，根据过去三年的数据找出美国配件业务收入的影响因素。他告诉他的团队：“快一点，否则我们就完蛋了。”

杰夫再次错过了一个关键的步骤：分析计划。

对，就是坐下来，理清事情的来龙去脉，写出一份计划。这些感觉确实像是零星的琐事，但却是确保分析工作成功的关键步骤。该计划至少有助于使你的分析与业务需求保持一致。它也迫使你按照时间计划表来思考，而不是把某个任务强加给你，然后要求马上完成。这让你有机会去找到并排列整合各种任务所需的资源。最后，让你理清问题，达成一致并获得认同；这可以证明避免争论是至关重要的（而且你可以放心，论据早晚会出现）。

步骤2：分析计划——数据科学层面

分析计划有五个组成部分：分析目标、假设、方法、数据规范和项目计划（见图4-6）。

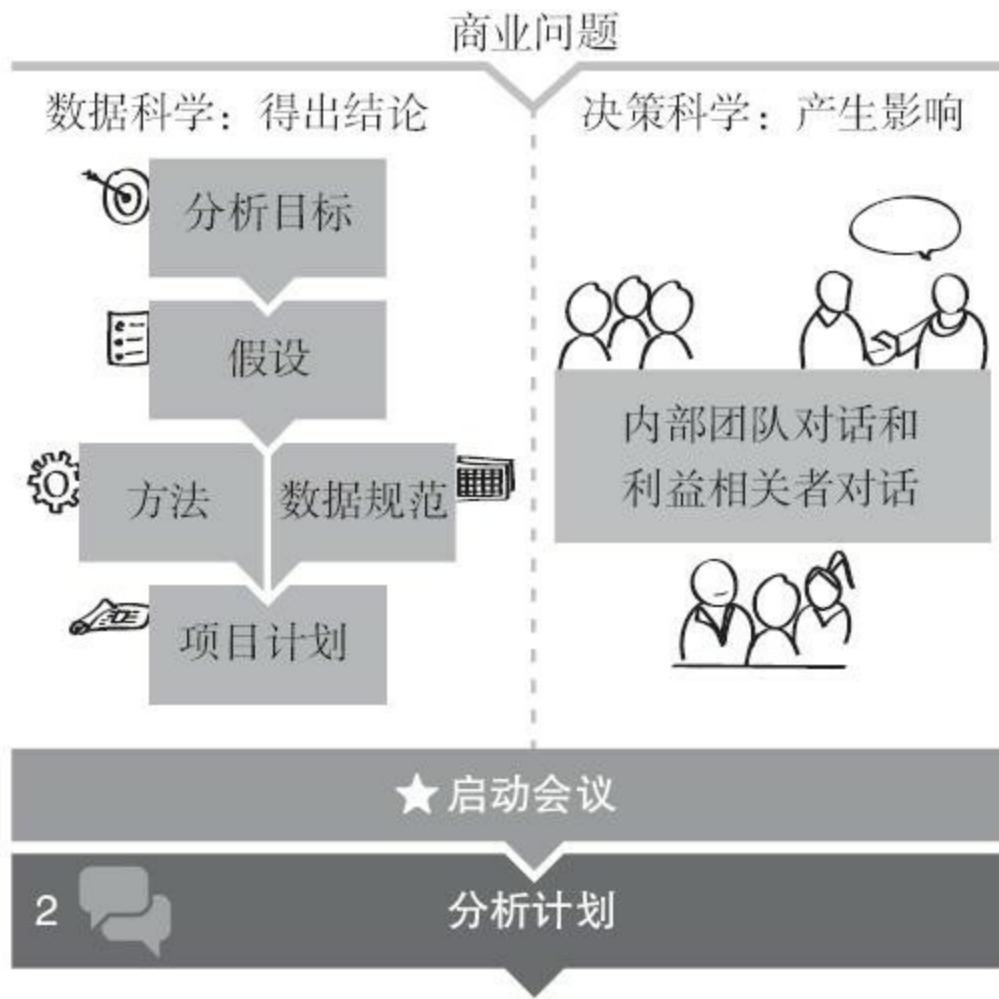


图4-6 步骤2：分析计划

步骤2：分析计划——数据科学层面→分析目标

首先创建一个SMART（明确的、可测量的、可实现的、相关的和有时限的）分析目标，回答步骤1明确的商业问题或其子问题（见图4-7）。它能告诉你根据既有数据可以直接回答的问题有哪些。

如果商业问题是“我们如何在三年内将1亿美元的业务增长到3亿美元”，那么这实在是一个很重要的问题。为了解决这个问题，除了分析之外，还需要融合商业的许多方面，包括战略、竞争格局分析、内部流程审核、技术审核、测试、市场研究、客户需求等。由于分析只能为一部分业务问题提供解决方案，因此需要将分析缩小到利益相关者在项目启动会议上赞同的某个分析目标。

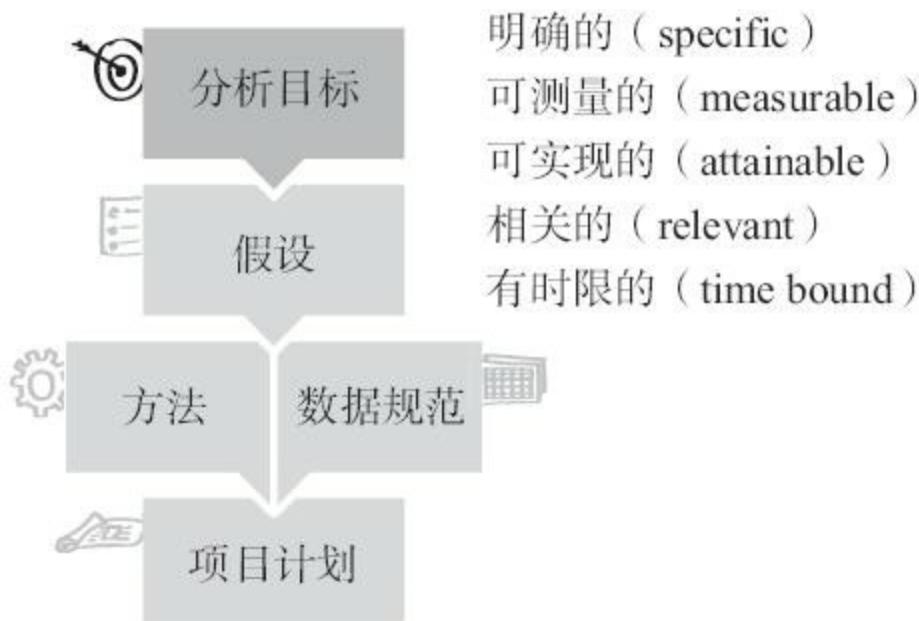


图4-7 分析计划的组成部分：分析目标

即使全部商业问题都可以用分析来解决，你也可以选择将项目分成多个子部分来提高效率。在这种情况下，单个部分的分析目标将是整个商业问题的一个子集。例如，如果业务问题是“我如何把我最新的次贷产品推向市场”，第一部分可能是“谁是新产品的最佳目标细分客户群”，第二部分可能是“应该使用哪些营销渠道——电子邮件、直接邮递广告、电话、网站，以及以什么频率的推广能避免取消订阅”。

Pets & Pets公司的亚历克斯和他的团队在准备分析计划时注意到：

- **商业问题：**在英国推出狗绳在线结账功能后，导致用户访问成交率下降的原因有哪些？质检员、产品设计师和产品经理可以采取哪些行动来解决？
- **分析目标：**确定访问成交率的影响因素，以及测试群体中的访问成交率低于对照组中的访问成交率的区段。

步骤2：分析计划——数据科学层面→假设

在确定应收集哪些数据之前，提出假设、定好标准来证明或证伪每个假设。每个假设都是有根据的猜测，猜想你通过数据分析试图解决问题的诱因是什么。假设描述了一个可能的答案，例如商业问题背后的影响因

素或原因。

假设通常采取以下形式：

A → B (A导致B，且B是商业问题的一部分)

人们经常认为假设来自数据，但其实并不是这样。假设大多来自所有关键利益相关者之间的广泛研讨，基于他们认为可能影响现状的因素。研讨的目标是找出多种可能的假设。在这个阶段，不应反驳任何假设，哪怕是那些看起来牵强或荒谬的假设。你也不必找出所有的假设，只需将你已经收集的所有假设进行优先排序，按照合理性或可测试性以及它们可能造成的影响进行排列（见图4-8）。

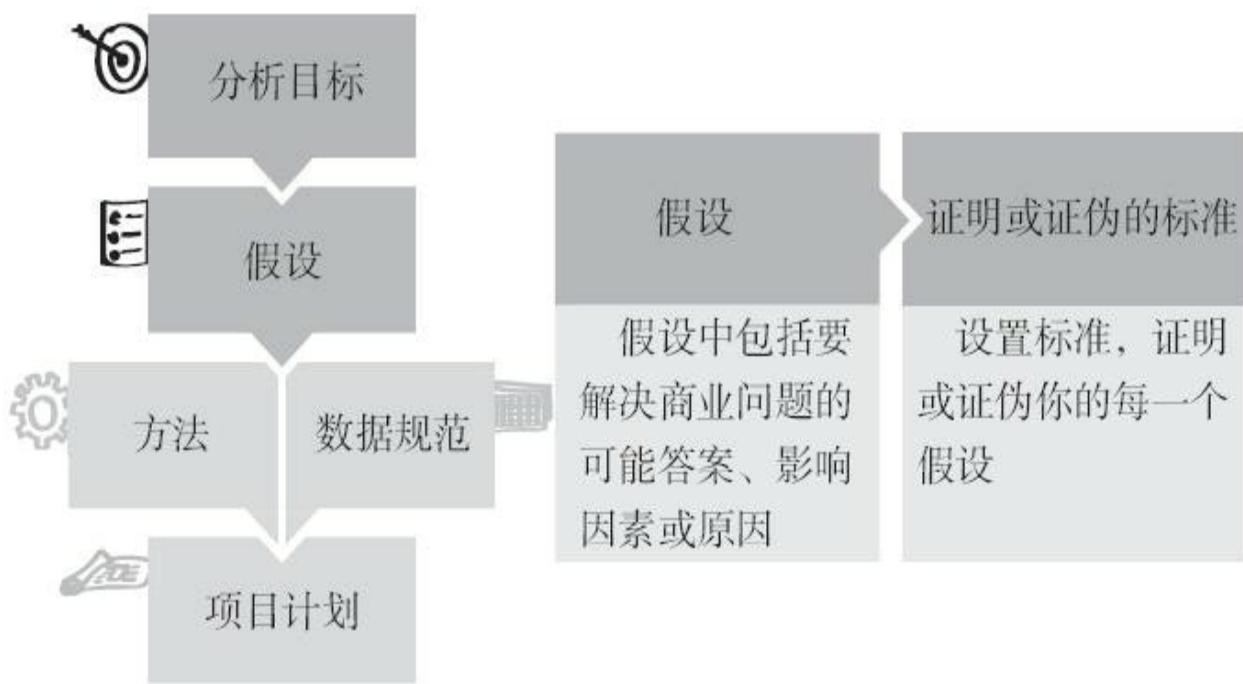


图4-8 分析计划的组成部分：假设

建议举行两次单独会议：

- (1) 第一次会议应该生成假设。
- (2) 第二次会议应该能将假设按优先顺序排好。

提出假设的最佳来源是工作在业务第一线的人，通常是产品经理、营销

经理，还有能提出问题的客服或销售人员，或是根据研究结论将真正采取行动的人。

排列优先级的会议可以有少量权威人士参加，通常以资深利益相关者居多。他们应该首先商定优先级标准，如潜在的影响力或获取数据的难易程度，然后每个人投票表决哪些是最佳选择。100分机制很适合评定优先级。在这种方法中，针对所有假设，每个利益相关者可以在100分内给出评分。然后，根据每个假设所得的总分对它们进行优先级排序。

Aryng团队完成了一个分析咨询项目，该项目是为了帮助一个希望减少客户流失的金融服务公司。在集思广益的研讨会中，有人提出风险政策的变化可能会剥夺潜在买家的资格，因此导致客户流失。如果没有这个假设，就不会有人更深入地研究那些晦涩的专业政策。风险政策变化通常不会存储在数据库中，因此这个假设只能通过商业知识来得出。项目研究最终发现，该政策原来就是导致潜在客户流失的主因。这就很容易做出调整，使收入显著增加。

Pets & Pets公司的亚历克斯和他的团队得出了几个假设，均来自与网络产品团队、开发和营销团队研讨时的收获。然后，他们优先选出其中三个假设：软件漏洞、未经测试的IE6浏览器代码的问题，以及新版Chrome浏览器的问题。这是排在前三位的，可能解释在英国推出新功能后访问成交率下降原因的假设。

步骤2：分析计划——数据科学层面→方法

商业问题的类型决定了你应选择的分析方法或途径（见图4-9）。就像第3章所讨论的，常用的分析方法一共有7种，其中最常被使用的是综合分析、相关分析、趋势分析和估算。

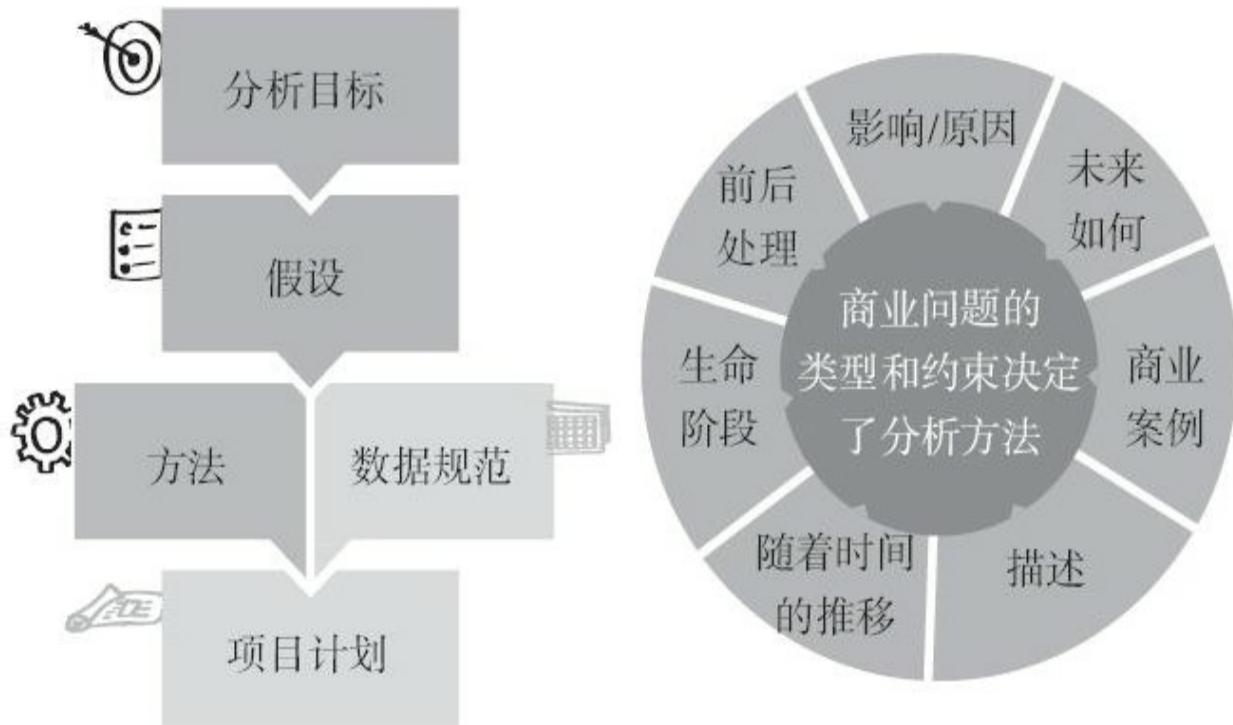


图4-9 分析计划的组成部分：方法

Pets & Pets公司的亚历克斯和他的团队确定相关分析方法适用于这个项目，因为他们需要找到与访问成交率相关的事件来解释下降的原因。

步骤2：分析计划——数据科学层面→数据规范

根据假设、证明或证伪的标准，以及所选择的方法，收集必要的数据进行分析（见图4-10）。数据采集时应该注意两点：

- 只收集有助于证明或证伪某个假设的相关数据。
- 在进行数据收集之前，应先写明数据规范（见图4-10）。

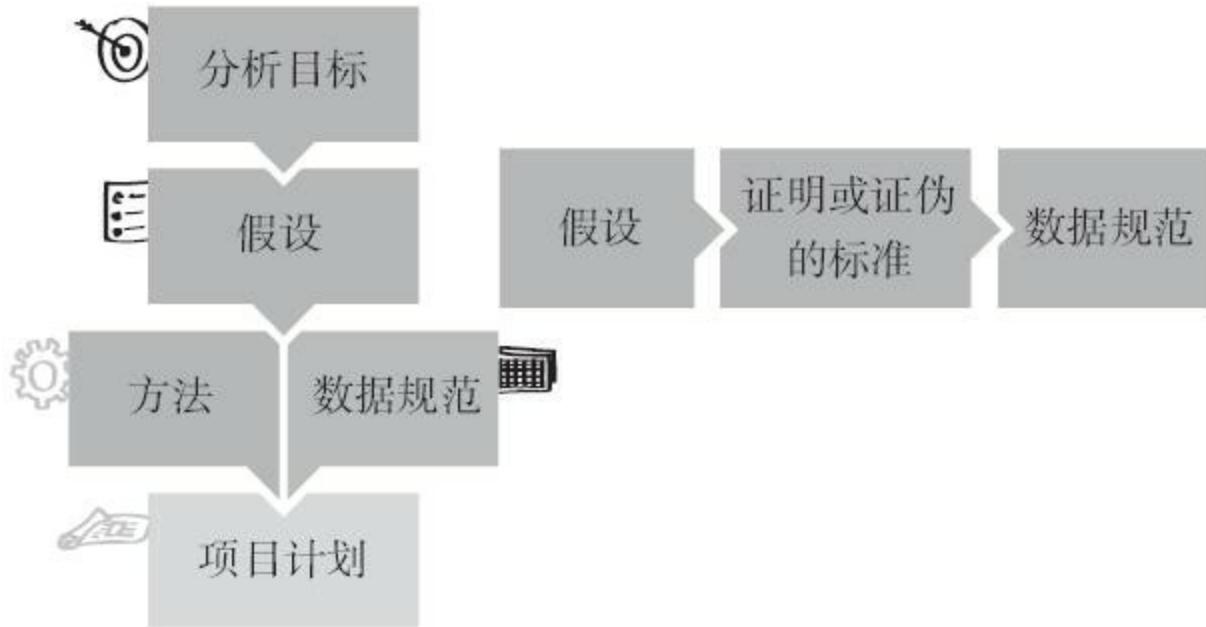


图4-10 分析计划的组成部分：数据规范

数据收集只有在主要利益相关者商定好完整的分析计划后才能开始。

首先明确每个假设所需的数据，然后将其输入到一个主要数据规范中。在编写数据规范时，确定与此分析相关的粒度级别——按时间、地理位置、事件或部门来表示的粒度。假如你正在搜集销售数据，你是按照每周、每月还是每年来提取数据？根据所需的聚合级别和所提取的原始数据的级别，你可能需要相应地将数据进行汇总。在这样做时，要确保将它绑定到唯一的ID。比如，如果你已确定以每月为粒度级别，那么 Month_ID 将成为你唯一的ID，你假设中的所有不同的数据元素都会绑定到这个ID上，例如漏洞、不兼容的浏览器等。

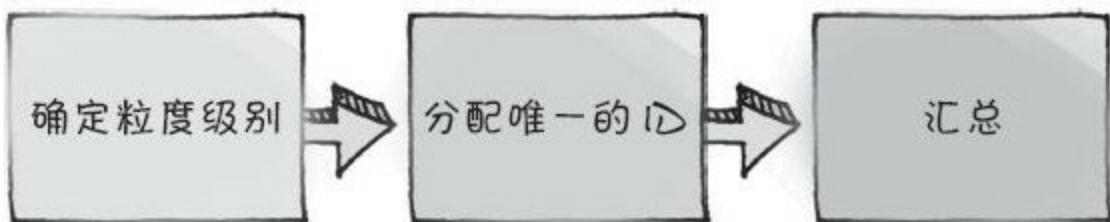


表4-1 Pets & Pets案例：假设与所需数据

假设	证明或证伪的标准	所需数据
新功能中的漏洞导致访问成交率下降	漏洞数量占交易尝试数量的百分比，在访问成交率下降后明显高于下降之前的	测试组与对照组的漏洞数量；每组的访问成交率
IE6 浏览器功能没有针对用户体验进行优化	一种浏览器的访问成交率相对历史水平呈下降趋势；足够大的访问量	测试组与对照组的 IE6 浏览器访问成交率；与上相同
Chrome 可能出现的问题	测试组中 Chrome 访问成交率低于对照组	测试组与对照组中各浏览器的访问成交率

在 Pets & Pets 的案例中，他们需要得到交易尝试这一级的数据，因为他们要了解为什么某些交易尝试留下来成为用户，而其他交易尝试却没有（见表 4-1）。表 4-2 显示了他们的数据规范。

表 4-2 Pets & Pets 公司：数据规范

变量名	定义
Txn_atmpt_id	交易尝试 ID
Bug_Flag	是 / 否，标志交易是否有错误
Browser_Type	用于交易的浏览器类型
Test_Flag	测试 / 控制标志
Conversion_Flag	1/0 标志，表示是否成交

注：今后发布的所有交易测试数据都按这种粒度。

基于关注漏洞的第一个假设，Pets & Pets 团队确认他们需要测试组和对照组在漏洞期间和非漏洞期间的访问成交数据，以证明或证伪那个假设。类似地，用同样的方法，他们确定了其他两个假设所需的确切数

据。

步骤2：分析计划——数据科学层面→项目计划

项目计划将所有组成部分联系在一起，只有在明确了解上述分析计划之后，才制订项目计划（见图4-11）。



图4-11 分析计划的组成部分：项目计划

项目计划中包括的关键要素有：

- 资源：**这个项目可能需要哪些人员、预算、硬件、软件、其他IT资源、其他公司资源和外部资源（如果有）？
- RASCI：**谁负责（Responsible）推动这个项目、谁是有责任的（accountable）、谁扮演支持（supportive）角色、需要向谁请教（consulted）和谁需要知情（informed）。这些角色是至关重要的，应该定义在前面，这样人们才不会凌驾在彼此之上或事后指责别人。有些公司使用此模型的变体，如DACI、RACI和ARCI，它们具有相同的元素。我们还发现了将支持部门纳入进来也有价值，因为这可能涉及整个部门，例如IT或项目管理。
- 时间表：**时间表包括与利益相关者有关的时间节点和适当的登记点，确保每个人都明确了解他应在何时履行职责以及何时需要做出关键决策。对于每个项目，不论项目大小，我们都建议设置至少三个时间节点

点：项目开始、有结论后初次报告和最终报告。

- **风险：**在此阶段识别所有潜在风险非常有价值。例如，项目越长，与资源缺失或意外事项相关的风险就越大。当计划A失败或者不能按时完成时，拥有计划B甚至计划C总是好的。
- **阶段：**将可交付成果分成更小的阶段，可以快速得出第一套具有操作性的结论，同时确保你处于正确的轨道。例如，前5个假设可以构成阶段1，并且一旦阶段1结果返回，接下来的5个假设可以是阶段2的一部分。然后，执行团队可以在阶段2还在工作时，就先执行阶段1的结论。
- **确定优先级：**最后，需要确保该项目相对所有参与者的优先级，以便满足此项目的时间安排。正在进行的项目可能必须重新确定优先级，以适应这个新项目。这需要与各利益相关者适当进行协商。

步骤2：分析计划——决策科学层面

你会注意到，分析计划的所有组成部分都涉及内部团队对话和利益相关者对话，以产生各自的结果——从分析目标的协调，到项目计划的签订。在构建分析计划时，请确保内部团队和利益相关者都在你的工作路线安排之中。与内部团队可以讨论协作的角色、职责、资源和以前的报告等。与利益相关者则可以讨论你的分析计划，特别强调在目标、方法、优先假设和时间进程安排方面去达成一致。分析计划的顶点是一个正式的启动会议——强烈推荐。在这次会议上，应确保让利益相关者签署分析计划，并取得那些将投身配合分析工作的人的承诺。分析计划不是解决方案，而是所有相关各方都能看到的清晰路径。

Pets & Pets公司的亚历克斯利用他在与布兰妮第一次会议中得到的信息，以及随后与其他利益相关者的头脑风暴会议信息制订计划。这个计划给了亚历克斯他需要的信息以有效地开展工作，比如，提出了可能的假设，确定了适当的方法（相关分析，在本案例中），且合并了项目计划。最后，他咨询了布兰妮的意见，然后稍微调整了计划。带着充满信心的计划，布兰妮第二天会见了公司CEO。在那次会议上，她提出了拟议的分析计划，而不是解决这个问题的办法。亚历克斯明确表示，他的团队需要一周的时间来解决核心问题，并提出了建议，布兰妮和CEO都同意了。随着计划被确定下来，他们正在一步步接近答案。

步骤3：数据收集

数据收集是BADIR框架中的步骤3（见图4-12）。通常，人们从这里开始数据分析工作，但在你查看数据之前，你需要先了解问题并达成一致同意的分析计划，以便可以策略化你的数据收集工作。



图4-12 BADIR分析框架：步骤3——数据收集

为了加快进程，SmartShoes公司的杰夫跳过了分析计划的这一步，然后，为了进一步节省时间，他的团队又跳过了数据清理和验证步骤。他们没有去想需要什么数据或去哪里拉取这些数据。结果，他的团队不得不去收集所有的数据，通宵加班。

步骤3：数据收集——数据科学层面

记住，GIGO（无用输入，无用输出）。分析结论的质量和分析建议的价值取决于你一开始使用数据的质量。数据收集，以及数据清理和验证，都很重要。

让我们谈谈数据收集的两个阶段和最常见错误的避免。

（1）数据拉取：根据数据规范收集数据。

(2) 数据清理和验证：清理数据使其可用，并验证数据以确保数据准确。

数据拉取

根据分析计划中的数据规范去收集数据，而分析计划又是由分析目标、假设和所选方法决定的。根据你所处的角色，该过程可能会有所不同。

- 如果你是商业专业人士，请让分析人员根据规范提取数据，或者从Tableau、Pentaho或MicroStrategy等商业智能工具中提取数据则更好（请参阅第6章）。虽然通过这些数据仓库获得的只是综合数据，但这样的数据对于大多数简单分析是足够的，因此它不是什么重大的限制。
- 如果你是分析人员，则可能直接使用SQL或类似语言从数据库中提取数据，然后进行整合。

提取数据，特别是直接从数据库中提取数据时，先提取一个小样本进行考察，以确保那确实是你所需要的。此外，了解每个指标的预期数据类型，并确保它与你在样本中看到的一致。

数据错误的常见来源。第一种数据错误产生于多个来源。你可能会从各种不同的应用程序（如CRM系统、Web分析工具、产品环境、数据库或机器数据）中提取数据，最终获得来自不同数据源的类似数据点。为此，应确保分析中使用数据的一致性，要知道每个数据的来源（见表4-3）。

表4-3 数据收集指南

提示	常见错误
<ul style="list-style-type: none"> • 数据应绑定到唯一的 ID • 在抽取大量数据之前，使用小数据样本对 SQL 进行测试 • 使用每个指标的预期数据类型匹配收集的数据 • 将所搜集数据与每个指标的预期数据类型相匹配 	<ul style="list-style-type: none"> • 包含已覆盖或错误分类的历史数据 • 数据部分附加在历史记录中的某个时间，但都代表了完整的历史数据 • 多种来源的类似数据

你可能会发现，并非每个数据源都以相同的方式在定义数据，或者，它们对数据的定义与你所想要的定义不同。例如，不同的来源可能会以不同的方式定义销售收入或成交率，这可能会对你的数据分析产生影响。

在Aryng公司与财富500强金融服务机构的咨询业务期间，我们发现，某个特定产品的成交情况因为这个原因而被过度报告了。这种特定的产品被销售给商家，承诺能比市场上的其他支付选项具有更高的成交率。然而，这种说法最后被证明不是正确的。该产品的成交情况类似于市场上的其他产品。误报背后的罪魁祸首是不同来源对该产品的启动和成交进行了不同的定义。当这种错误最终被披露时，商人有很强的抵制。

数据错误的第二个最常见原因是数据的历史覆盖，导致该列数据的内容不再与描述相匹配。这两个问题都需要数据清理、验证和对照，这些都是耗时的活动。然而，使用数据标记（例如勾稽关系测试）来发现数据问题将加速该过程。

数据清理和验证

在准备开始收集数据前需要验证数据。首先对小数据样本进行完整性检查，并将它与你期望的内容进行比较。注意空值（特定数据库中不存在的数据）和数据的基数（该列的不同值的数量）。这是你期望找到的吗？例如，经常在自我报告的数据（比如职位）中，客户可能选择“其他”作为选项，或者他们可能不会填充它。虽然“其他”这个答案不是空值，但是如果信息未被填充，则可以将其存储为空或空白。因此，检查基数将帮助你发现“其他”以及“空白”报告为“空”的问题（见例4-2）。

例4-2数据收集：验证数据的小提示

- 检查空值与基数。
- 确保数据类型符合你的期望。
- 通过目察50个奇数行进行样本健全性检查。
- 通过检查总金额之间的勾稽，检查收入和其他关键指标与其他报表中的数据是否相匹配。
- 进行快速单变量分析，并观察大部分空、零、无效或相同的值，以及分类为“其他”的值。

同样，检查数据类型是否符合你的期望。最重要的是，你需要将一些关键指标与你手中的现成数据进行勾稽核对，与其他明显不同的来源数据进行匹配，对数据进行稽核。例如，你可能已收集公司日本部门的收入数据，以了解组合情况。验证数据的最简单方法之一是将来自你的数据集的日本部门收入与来自日本的关键报告进行核对。如果它们不匹配，你会知道你的数据有问题。大多数时候，不同的来源或不同的定义是主要的罪魁祸首。

Pets & Pets公司的亚历克斯完成了他的计划，提交给他的经理。然后，团队基于分析计划中的数据规范，仅仅收集必要的有效数据，这只需要很少的时间来完成和验证。他们不必工作一整夜。相反，他们正在计划一个愉快的周末。

步骤4：得出结论

在这一步，你可以获得：

- 节省时间的结构化方法；
- 能得出相关研究结论的正确方法；
- 通向影响和可操作性。

请参见图4-13，了解怎样得出研究结论的过程概述。

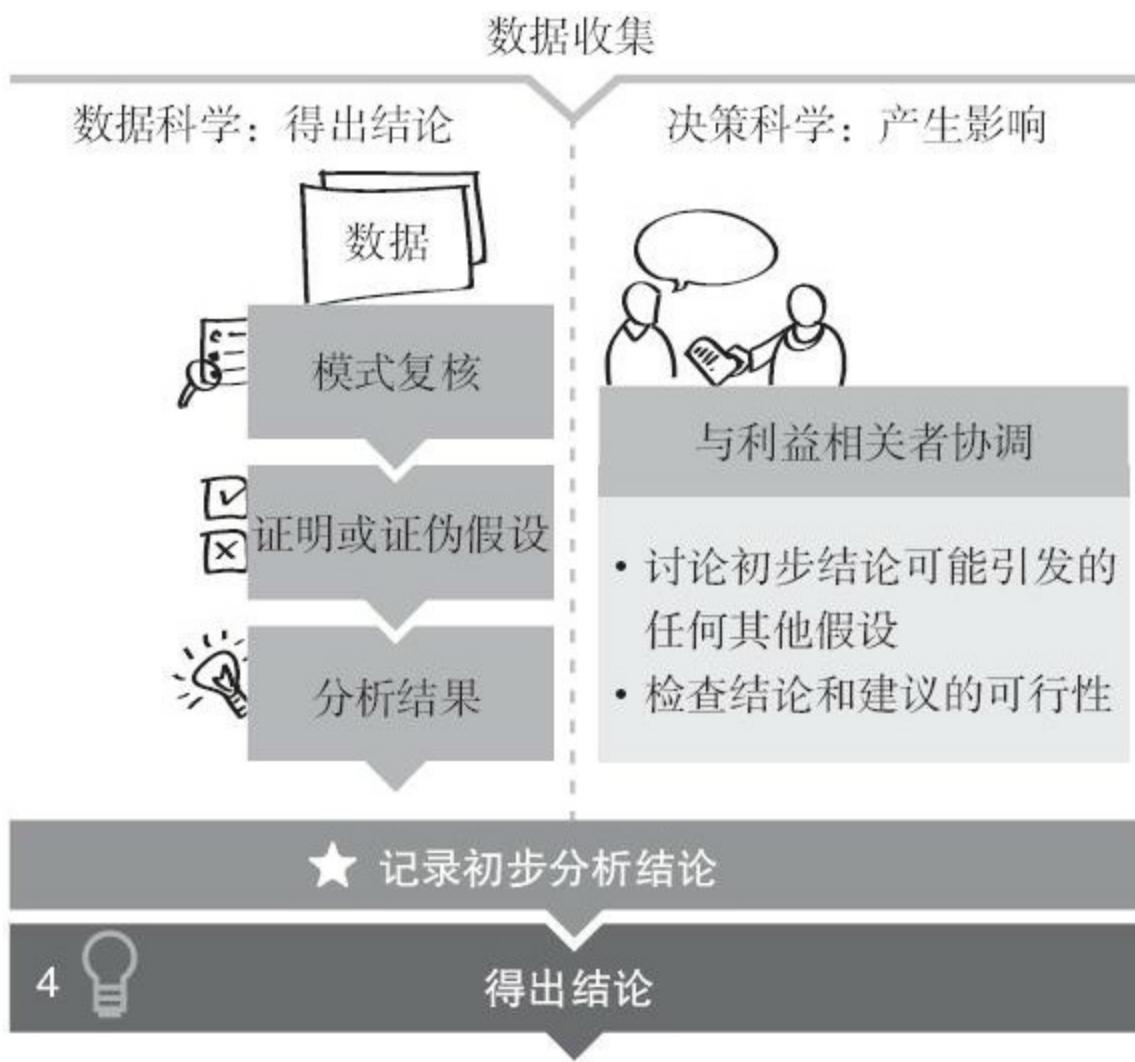


图4-13 BADIR分析框架：步骤4——得出结论

缺少结构化方法的指引，又跳过了关键步骤，SmartShoes公司的杰夫团队在正在尝试从他们建立的模型中找出研究结论。这是一个逻辑回归模型，他们希望通过这个模型找出影响收入的因素。在经历了一个不眠之夜后，他们找到了三个可能影响收入下滑的因素。一个是数据依据错误，因为他们跳过了数据验证步骤。另一个是如此明显，唯一可能的反应是“呃，这是新闻？”然而，还有一个从相关分析中找出的原因则比较有意思。

汤姆非常失望，他一直希望奇迹出现来缓解他的压力。现在，不仅这个模型难以理解或无法向董事会解释，而且他的团队没有能提出一个关于如何解决配件业务收入问题的建议。汤姆请杰夫明天回来提出建议。现在他真的害怕他会保不住自己的工作，并开始打算找猎头了。

步骤4：得出结论——数据科学层面

当处于最高级别时，一旦你获得了数据，就可以选择任意的分析方法（见图4-14）。

- 模式复核：可以帮助验证数据模式，判断是否真的存在商业问题，或者关键变量中是否存在异常模式。
- 证明或证伪假设：查看每个假设，一次只检查一个，并检查证明或证伪每个假设所需的相关数据。这将有助于去除一些假设，并确定你应该将精力集中在哪里。
- 报告结果：最后，用量化语言报告分析结果，以指导分析假设的优先级设定。

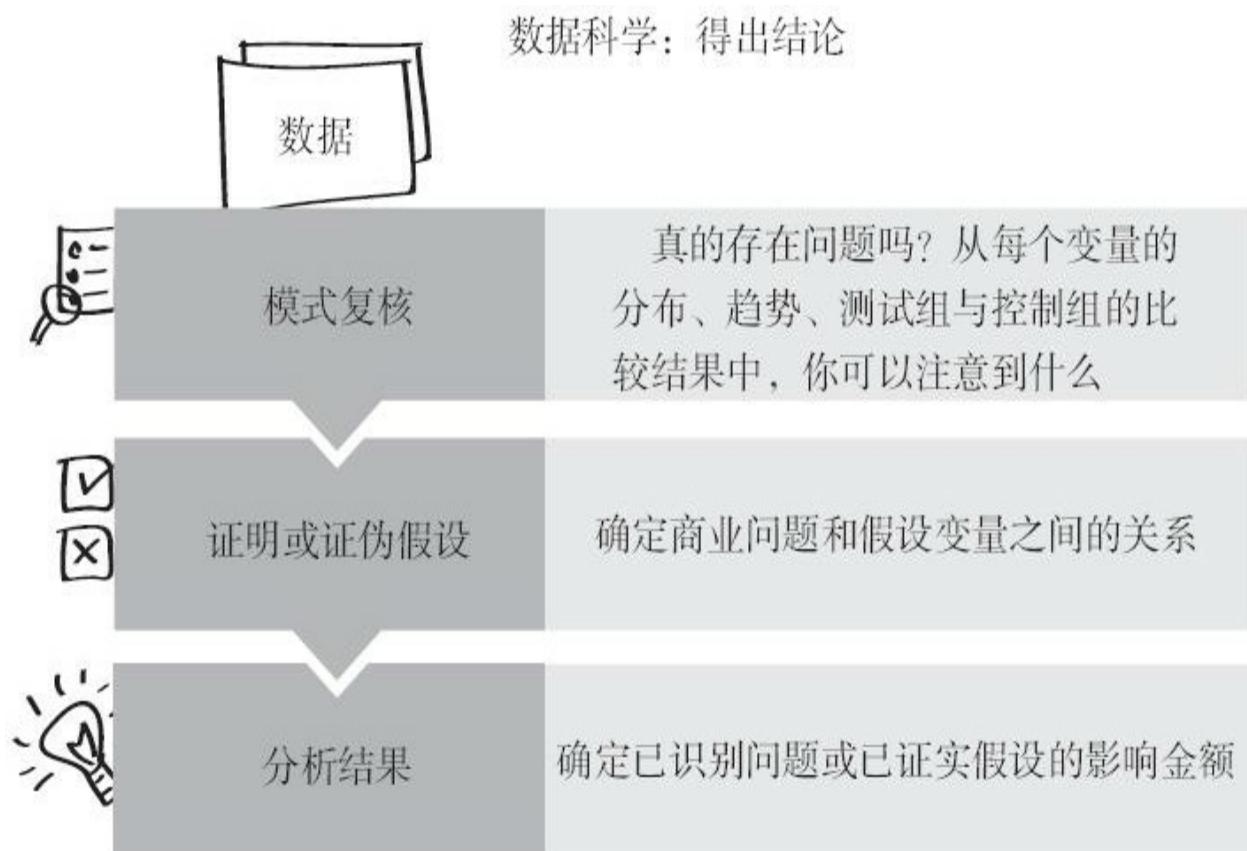


图4-14 从已搜集的数据中得出结论

在任何分析中，请注意总问题或机会的大小，以及你所确定的关键影响因素的解释的问题百分比。

在第3章中定义的七种分析方法中，综合分析、相关分析、趋势分析以及预估是其中最常用的四种。让我们依次来看看。

综合分析

综合分析是证明或证伪假设的一个捷径，在描述性分析（如特征描述或比较分析，例如竞选分析和输赢分析）中最常用到（见图4-15）。

●模式复核：使用此方法时，请查看相关的指标，并按不同的分部将数据分类。常见的例子是判断哪个营销活动是成功的，以及发生在哪个部门。通过查看最佳和最差网站跳出率、退订率、开放率和点击率（CTR），就可以解决此问题。



图4-15 综合分析

Zameify是一家游戏公司，它的营销团队希望从人群类别和参与度两个

方面确定iPhone玩家是否比台式机玩家更有价值。这将直接影响公司的营销策略和目标。在他们的分析计划中，营销和分析团队成员确定了他们想要查看的关键属性、分析期、iPhone应用中的访问次数与网站访问次数和用户消费金额。他们收集了台式机和iPhone用户在产品整个生命周期内的数据元素。按照最常见的访问渠道（iPhone与台式机）进行分类，然后比较每个独立用户（UU）的综合关键指标。结果如表4-4所示。

表4-4 Zameify公司：综合分析

	iPhone	台式机
年龄（平均）	25岁	32岁
收入（平均）	\$125 000	\$95 000
访问次数 / 独立用户	2.4	1.3
支出金额 / 独立用户	\$42	\$15

从上面的比较可以看出，iPhone用户比台式机用户更富裕和更有价值。因此，专门针对这些用户展开营销应该是有意义的。

●分析结果：一旦你查看了感兴趣的指标，并提出了关于关注点或战略变化的清晰见解，就可以量化此类变更对商业的潜在影响。

对于Zameify公司来说，这意味着如果公司专注于扩大iPhone用户的范围，使用户总数从100万增加到200万，那么公司的收入可能增加4200万美元。这是一个很好的理由，说明公司完全可以花10万美元的营销支出在iPhone用户上。

相关分析

在相关分析中，你需要考察与企业希望影响的目标相关的变量。此分析方法最常用于解决与理解商业或某事件的影响因素相关的商业问题（见图4-16）。

在Pets & Pets公司，存在的问题是功能发布后访问成交率不明原因的下降。亚历克斯和他的团队确定了分析目标，概述了三个假设，确定了证

明或证伪每一个假设所需的标准，并确定了所需的数据的范围。一旦团队完成了数据收集，就只需按照相关分析的三个步骤操作，证明或证伪这三个假设。



图4-16 相关分析

- 模式复核：相关分析从单变量分析开始，通过观察关键目标变量和其他与假设相关变量的分布，来判断是否存在问题（见图4-17）。它也是在实验中，可以利用相关分析来比较测试数据与对照数据，例如测试总体与对照总体的访问成交率。

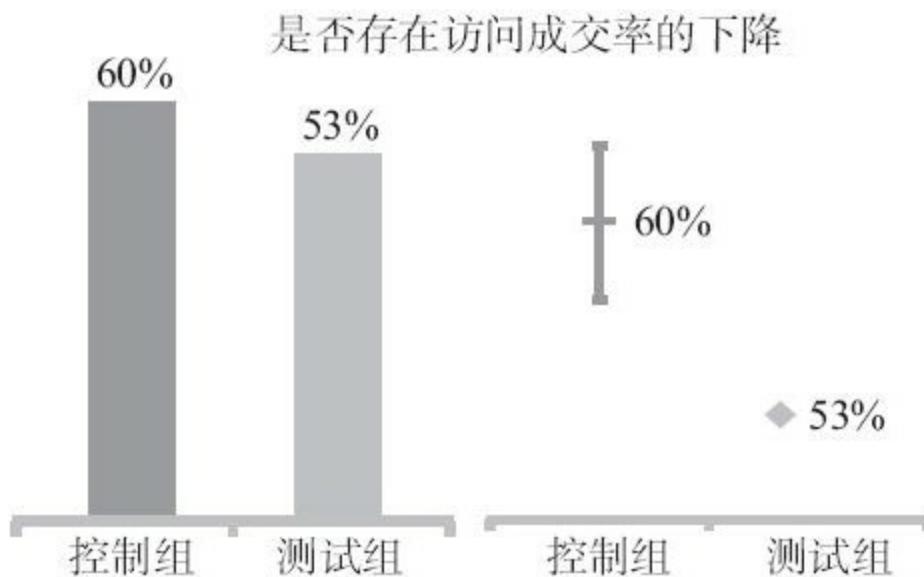


图4-17 单变量分析的示例

亚历克斯的团队比较了测试组和控制组的访问成交率，以确认是否确实存在访问成交率下降问题。团队成员发现，测试组中的访问成交率为53%，而对照组的为60%，组间差异具有统计学意义。

- 证明或证伪假设：下一步是双变量分析，一次检查两个变量之间的关系（见图4-18）。一个好的例子是分别查看访问成交率和浏览器类型、错误和关键产品之间的关系，确定哪些变量与访问成交率的下降相关。

假设1：导致访问成交率下降的错误

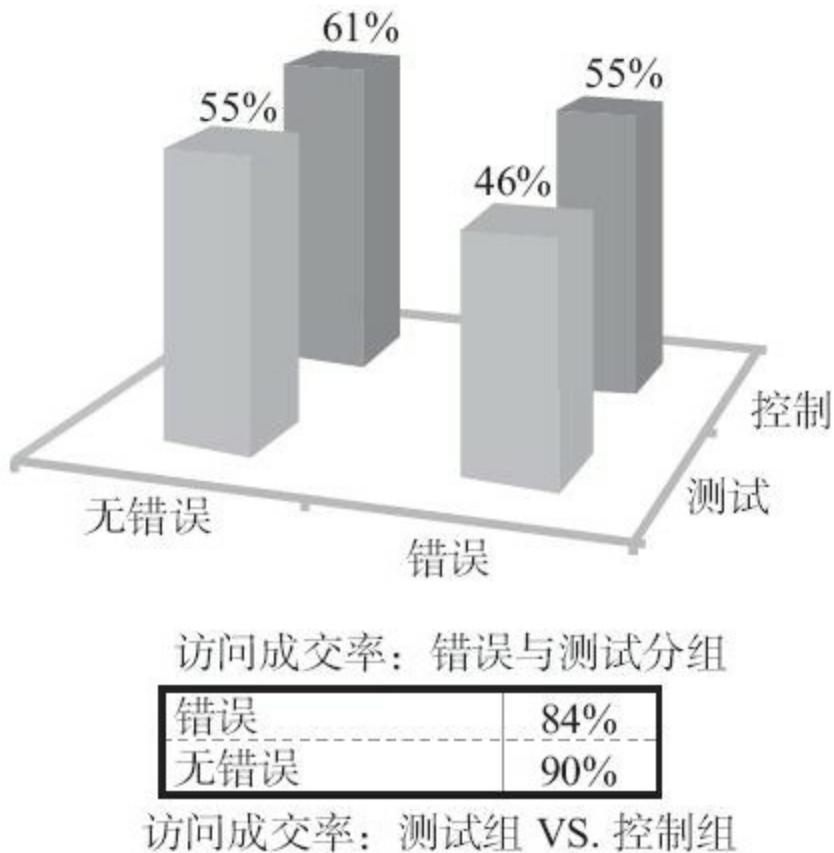


图4-18 Pets & Pets公司：双变量分析

多变量分析考察两个或多个假设变量与因变量之间的关系，并判断那些被证实的假设之间的交互影响。

亚历克斯的团队接受了第一个假设，即错误导致访问成交率下降。团队成员查看了测试组与对照组的访问成交率出现错误交易和无错误交易的比率。结果表明，出现错误会话后测试-控制总体的访问成交率都出现下降，这表明编程错误绝对是访问成交降低的因素之一。类似地，他们检验了所有其他假设，证明IE6浏览器也导致了访问成交率下降，而Chrome则没有。

亚历克斯的团队随后进行了三变量分析，以查看被证明影响访问成交率的变量组合，确定这两个变量的组合效果是否对访问成交率有影响。结果表明，当使用IE6浏览器同时又碰到错误会话时，访问成交率是最低的，结果如图4-19所示。

已证明假设：错误和IE6浏览器与访问成交率的关系

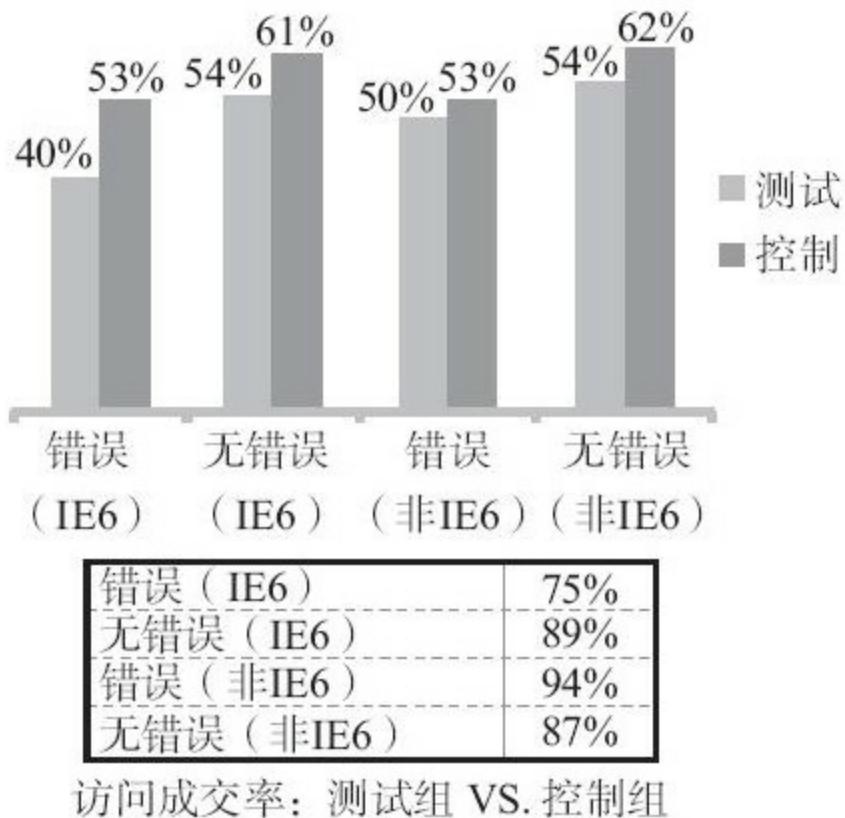


图4-19 Pets & Pets公司：三变量分析

●分析结果：最后，量化已证实假设的影响，即根据结论来提出建议（见图4-20）。将结论量化表示有助于建立案例，使其具有行动能力。要做到这一点，将所有的结论都按时序、相互依赖性或大小来排序。然后，量化第一结论的影响（即在一系列事件中首先发生的事件，其他所有事件都会受此事件的影响），接下来，在去除第一个事件的影响之后，再量化第二个事件，依此类推，直到所有的结论都被处理。

快速提示：将任何指标量化为用美元影响表示，能吸引利益相关者，因为这是每个小组都能理解的通用语言。例如，1%的下降导致200万美元的收入下降。

相互独立，完全穷尽 (MECE)

证明假设	测试人数	转换增量	转换影响	影响
错误 (IE6)	1 502	13%	0.9%	12.7%
无错误 (IE6)	4 275	7%	1.4%	19.4%
错误 (非 IE6)	4 061	3%	0.6%	7.9%
其他	11 557	8%	4.3%	60.0%
总人数	21 394	7%	7%	100%

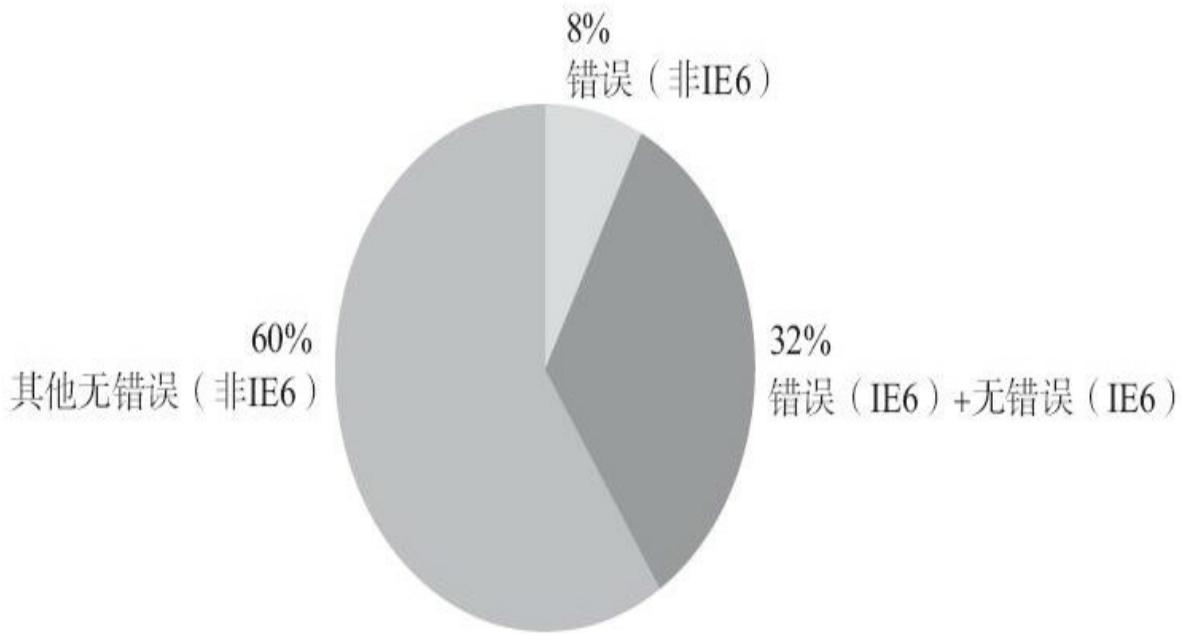


图4-20 Pets & Pets公司：量化影响

最后，我们来看看在瀑布图中直观表示的相同影响，以便于理解。如果能用表格将每个证明的假设与其整体影响的百分比联系起来，人们便可以一目了然哪些是主要影响因素，也会非常有帮助（见图4-21）。

结论：IE6 和错误是已知的下降原因

已证明的假设	测试数量	成交增量	成交影响	影响百分比
错误 (IE6)	1 502	13%	0.9%	12.7%
无错误 (IE6)	4 275	7%	1.4%	19.4%
错误 (非 IE6)	4 061	3%	0.6%	7.9%
其他	11 557	8%	4.3%	60.0%
总体数量	21 394	7%	7%	100%

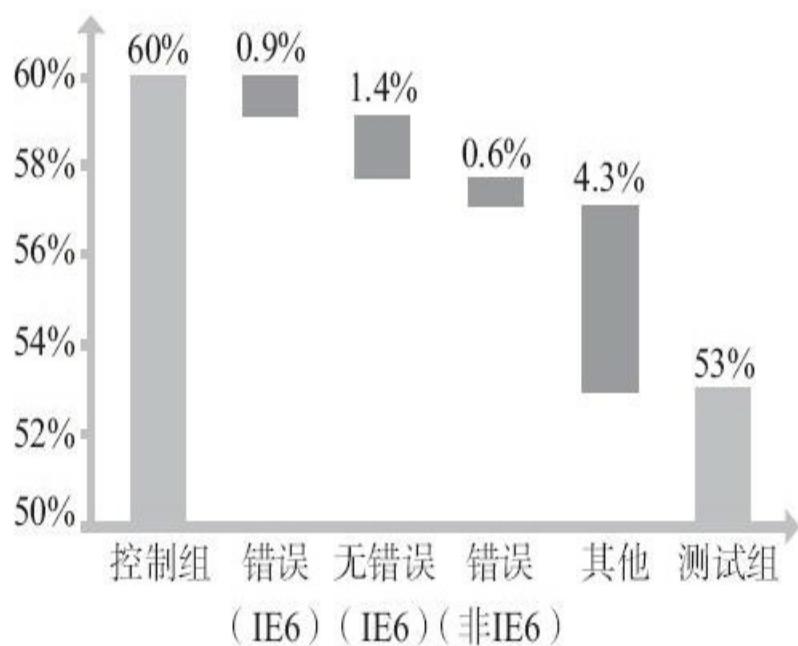


图4-21 Pets & Pets公司：瀑布图

亚历克斯的团队发现：

1.IE6上的错误造成了13%的下降。

2.IE6造成20%的下降。

3.其他错误导致8%的下降。

4.Chrome并不是问题所在。

5.60%的访问成交率下降仍然原因不明。

趋势分析

趋势分析是第三种常用的分析方法。它可用于分析销量和收入的趋势，以及趋势的改变。它还可以用于在一段时间内分析各个分部的情况和各种影响因素。趋势分析的步骤与相关分析的步骤类似（参见图4-22）。关键的区别是，需要查看一段时间内的数据。

分析销量和收入的趋势，并确定趋势的变化和分部或影响因素

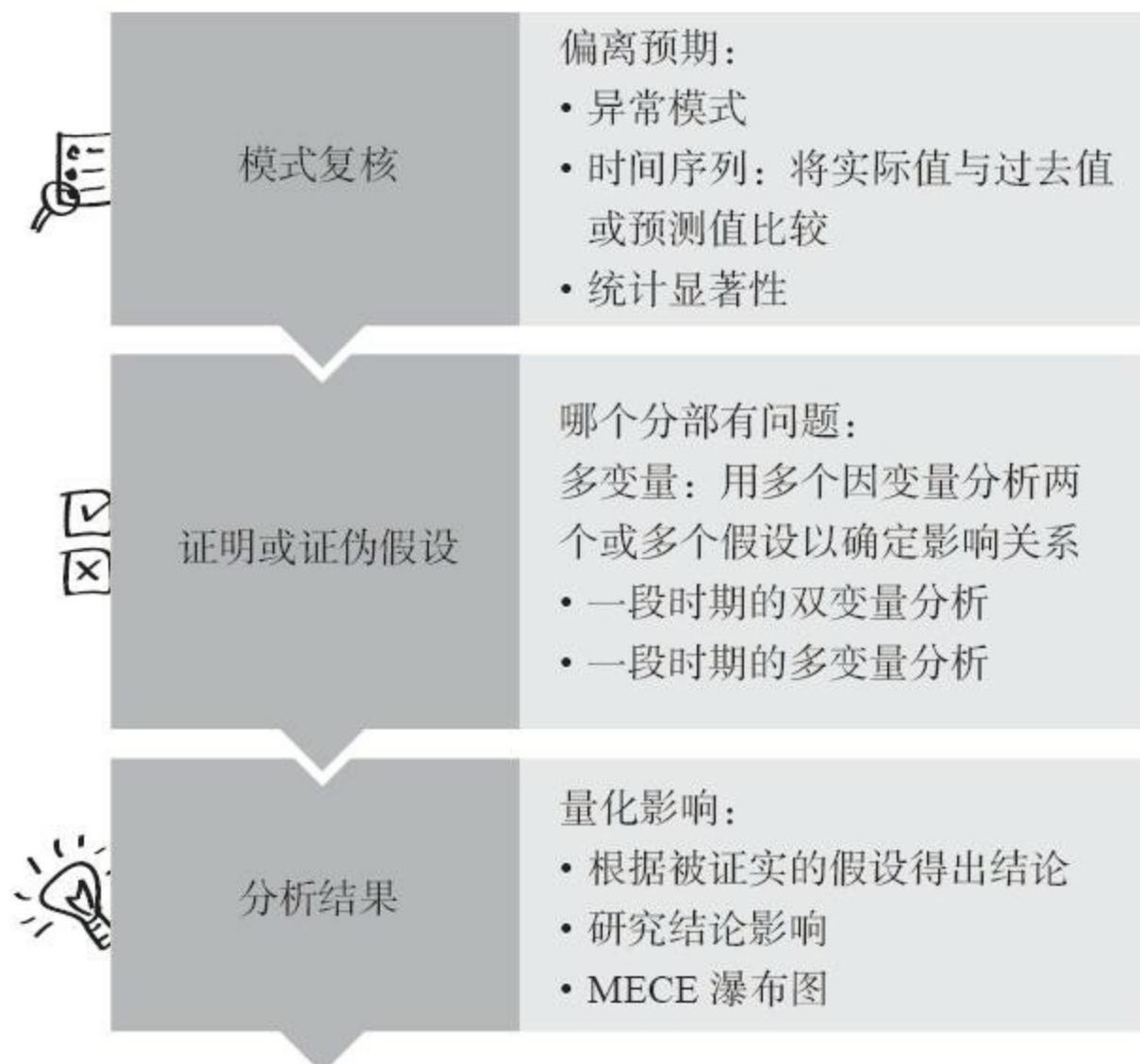


图4-22 趋势分析

- 模式复核：查看趋势的中断和一段时期内的模式规律（见图4-23）。
- 证明或证伪假设：通过双变量分析（图4-24）和三变量分析（见图4-25）证明或证伪关于这些趋势的关键假设。有时候，关键指标的趋势本身的趋势并不明显。在这些情况下，需要查看增长率和类似指标的趋势，以了解趋势线变化情况。
- 分析结果：量化每个结论的影响，如在相关分析中所做的那样（见图4-26）。

画出一段时间内的收入点状图，查看趋势的变化，并与预测值进行比较

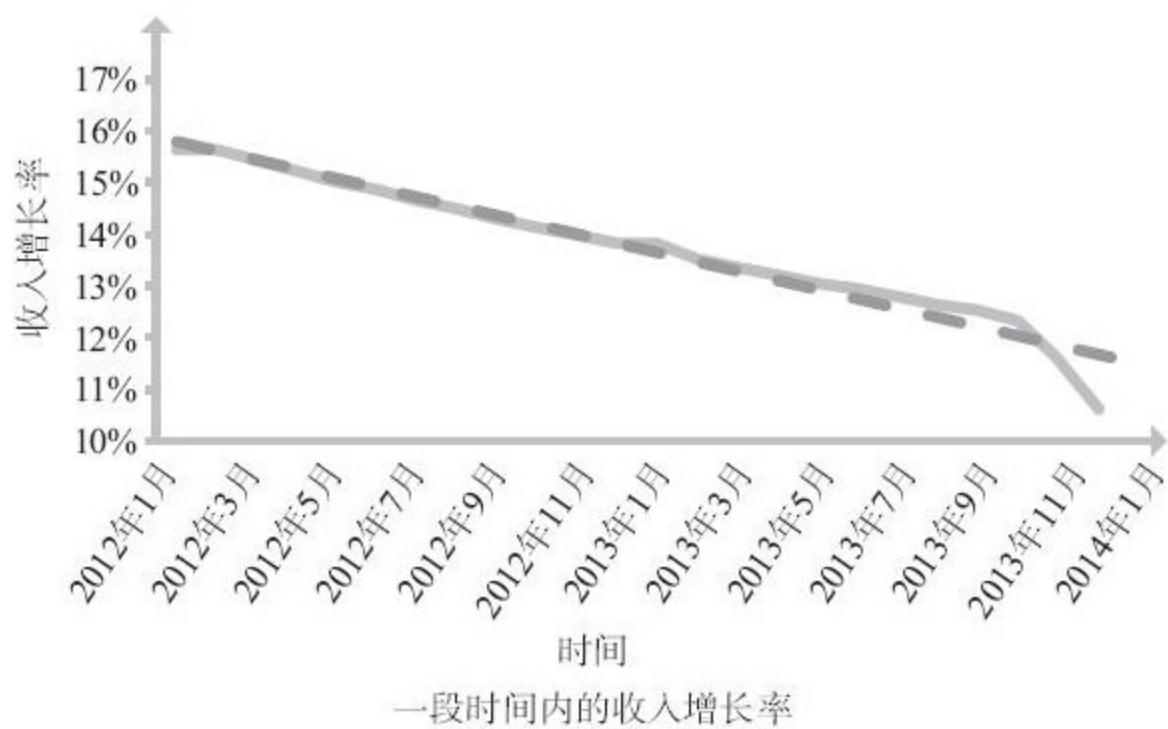
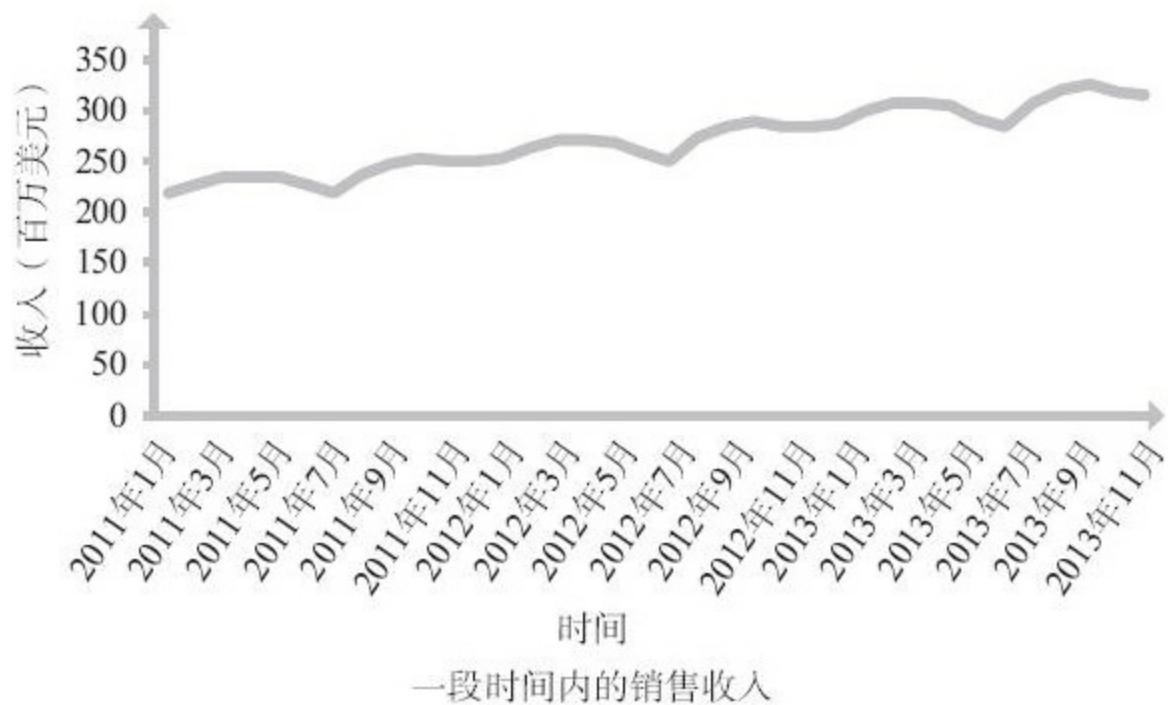


图4-23 Zameify公司：一段时间内的趋势

双变量：产品销售随时间的变化

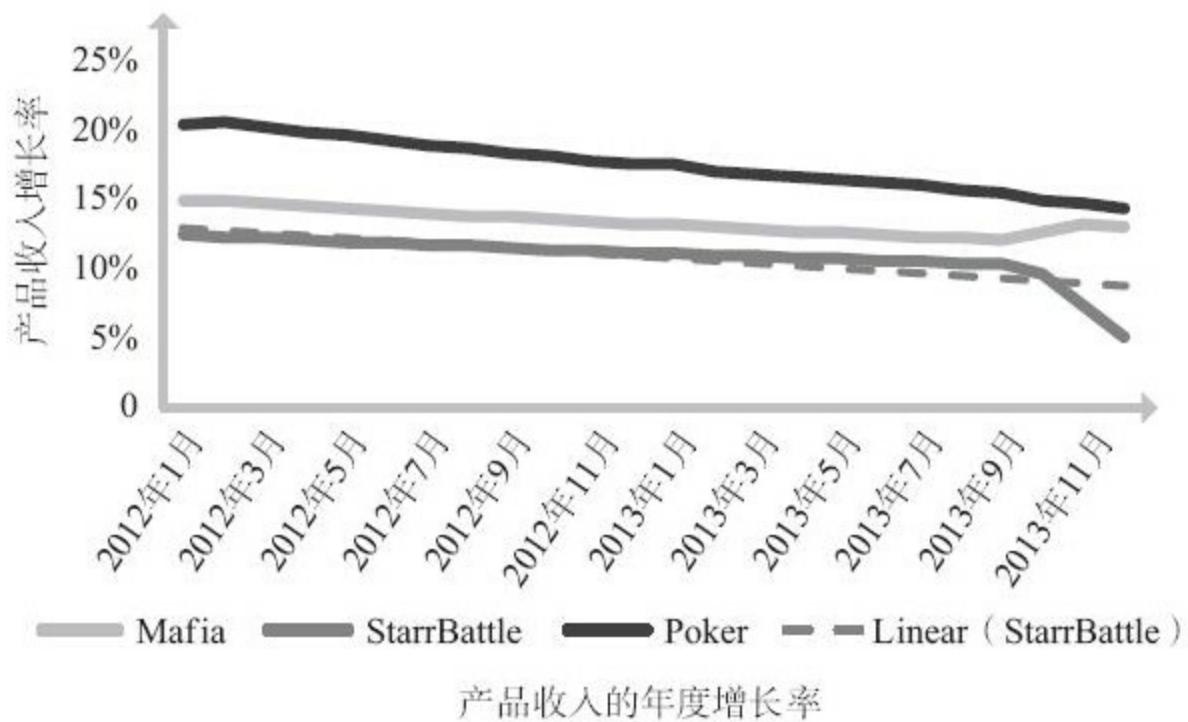
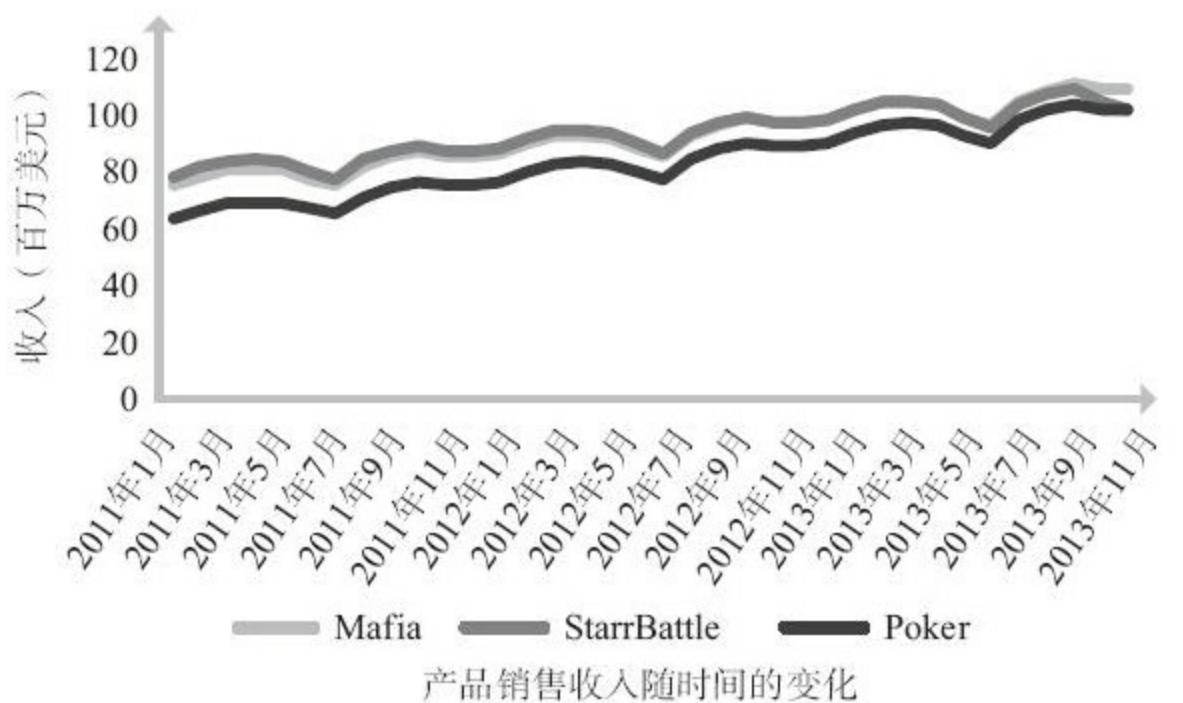


图4-24 Zameify公司：一段时期内的双变量分析

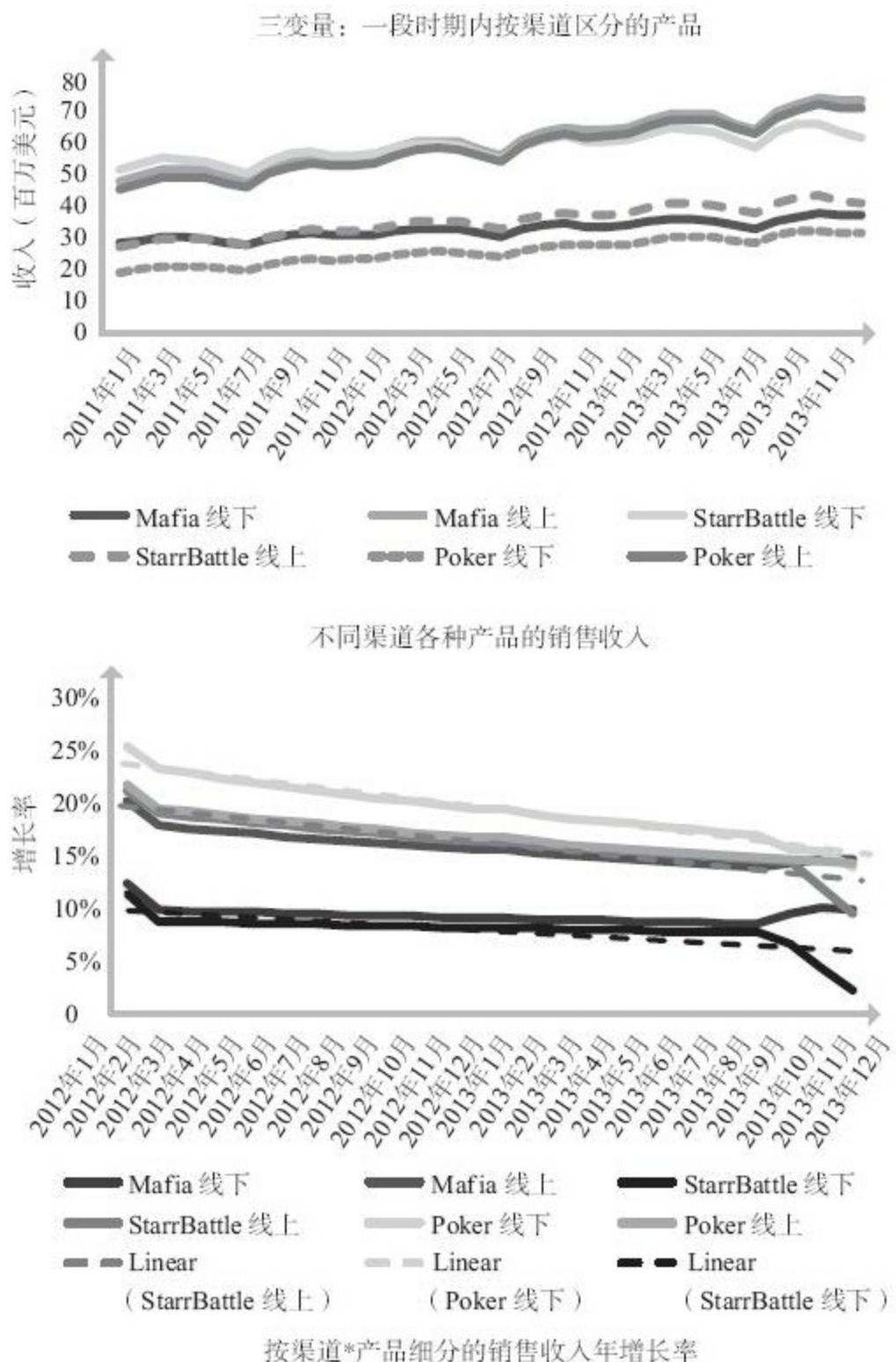


图4-25 Zameify公司：一段时期内的三变量分析

结论：销售放缓是由 StarrBattle 和 Poker 的线下部门引起的。

已证明的假设	月收入 (\$)	年度增长率差距	影响 %
StarrBattle 线下	61M	5%	52%
StarrBattle 线上	41M	5%	32%
Poker 线下	31M	3%	17%
总量	314M	1.9%	

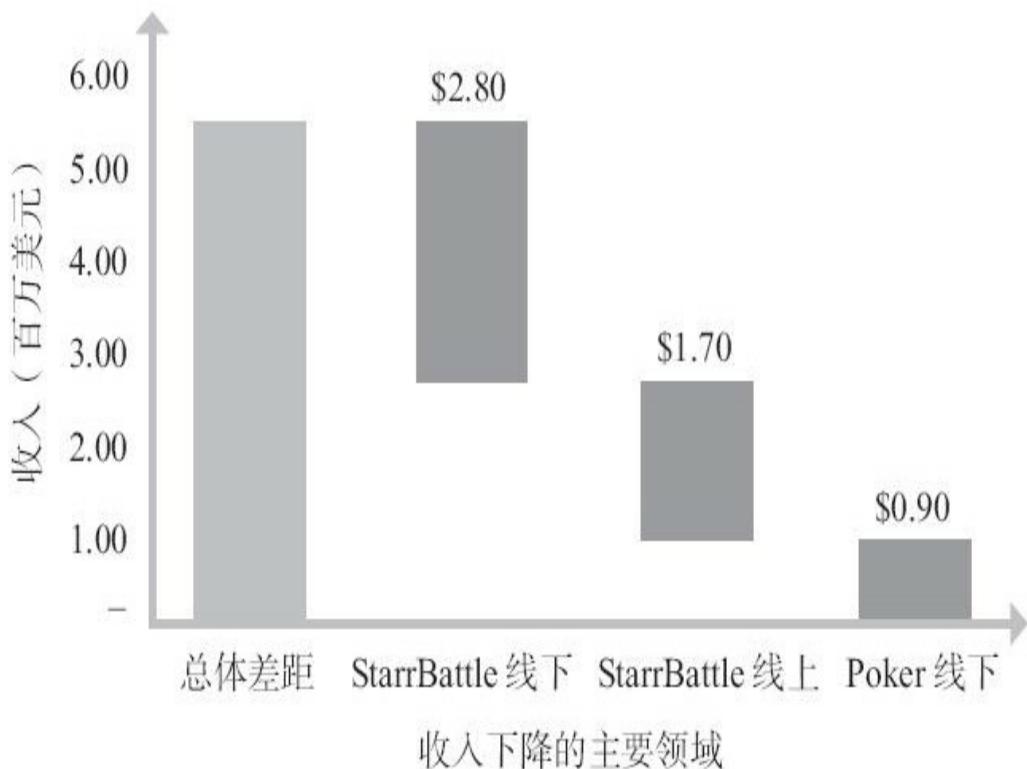


图4-26 Zameify公司：量化影响

我们之前介绍的游戏公司Zameify正在查找其收入增长放缓的原因。刚开始，该公司假定只是特定产品或渠道的收入放缓。调查人员发现，虽然总收入仍在增长，但收入增长率却明显下降。

因此，Zameify公司研究了各种产品随时间推移的收入分布情况，发现

StarrBattle这种产品的增长率在最近几个月都持续下降，而**Mafia**的销售则正在增长。同样，在查看各个销售渠道的收入分布情况时，**Zameify**公司发现线下销售的收入增长开始变得艰难。

利用三变量分析，将产品因素和渠道因素结合，**Zameify**公司发现，**StarrBattle**产品的增长率在线上和线下都有下降。此外，**Poker**产品的线下销售做得不太好。

接下来，**Zameify**公司开始量化每个结论的影响。该公司发现，有84%的收入增长下降是由**StarrBattle**导致的，影响金额高达450万美元。这显然是公司需要集中精力采取行动的地方。

估算

估算最常用于开发具有有限内部数据的企业案例。这是一个值得让你放入工具包的有用方法，因为它不仅可以使你的预算接近现实世界，而且还有助于在不容易获得数据的情况下，做好估算工作。

Edulane是一家为小学儿童提供在线夏令营解决方案的软件公司，目前正打算在加利福尼亚州湾区开展业务。市场营销经理马克认为，由于很多拥有高科技知识背景的父母在硅谷工作，因此硅谷将是一个线上夏令营的绝佳地点。在这之前，马克首先需要确定市场的规模。是100所学校还是1000所学校？他无法使用可用的开源数据来获得该湾区的准确小学数量。因此，他很快预估了该地区的小学数量（见图4-27）。

3

城市人口 = 50 000,
 郊区人口 = 10 000,
 每10 000人1所学校,
 20个城市+ 20个郊区,
 城市的财富因子是郊区的2倍

2

公式:
 学校数量 = 城市数×每个城市的人口数×单位
 人口对应的学校数×财富因子



图4-27 Eudulane公司：估算

任何估算都包含五部分内容：

(1) 分层：将问题划分为更小的分层，并识别与你所要估计的内容相关的但表现不同的分组。

马克认为，与城市部分相比，湾区郊区的小学密度会有所不同。因此，他将问题分解为郊区和城市两部分，这可能涉及使用不同的假设和计算。

(2) 相关性和影响因素：确定什么指标和因素可能对你正在估算的指标产生影响，用方程式来予以表达。

马克认为城市数、人口数和财富因子可能对小学数量有影响。因此他的方程是：

学校数量=城市数×每个城市的人口数×单位人口对应的学校数×财富因子

(3) 假设：对于方程中的各种因素，我们了解多少呢？对于在分层中确定的各个分组，可以采用不同的假设。

对于每个分组，马克就他所知对每个影响因素进行了赋值。他估计每10000人中有一所学校，估计郊区城镇的平均人口为10000，城市为50000。根据他对该地区的了解，他认为湾区有20个城市和20个郊区城镇。他还使用财富因子，因为城市地区更富有。他为城市和郊区镇的财富因子分别赋值2和1。

(4) 计算：这基本上只需通过数学来完成对每个分组学校数量的估计，同时设定高、中和低三种情况，以帮助设置边界。

使用这种方法，马克估计在中等情况下，大约有220所小学。

(5) 交叉稽核/正交实验法：这意味着用完全不同的影响因子来处理相同的估算问题。交叉稽核是一种很好的方法，能方便我们对先前的假设和计算进行现实的检查。为此，你将再次重复前四个动作，同时选择一个完全不同的方程来证实第一个方程的结果。如果它们不匹配，那么你可能需要重新评估你的假设或提出另一种交叉稽核的方法。

马克使用完全不同的影响因素构成的正交方法，重复了估算的前四步内容。他将目标市场总面积用平方英里表示，假设每平方英里将有一所小学，那么估算结果是共有350所小学。有了这个答案，马克证实了他的第一个估计，在湾区有220~350所小学。事实上，这是一个非常接近市场实际的估计。

马克的估计可以通过在雅虎目录中查找湾区小学来进一步证实，其中列出了大约180所学校。由于并不是所有的学校都会被列在目录中，所以马克的估计与实际情况相当接近。

步骤4：得出结论——决策科学层面

由于决策科学层面要求，我们找到结论，因此，现在该向关键相关利益者进行求证了。这也是一个利用这些结论去找出更多其他假设的绝佳机会。有时，可以用来证明或者证伪这些额外假设的数据已是现成的了，如果是这样，那么在最终的报告演示之前将这些数据与论点结合是非常必要的。如果这些数据不可用或者不容易被提取，则可将这些假设放在分析计划的下一个阶段，并在演示中具体命名为“下一步”呈现给利益相关者，使他们对分析状态清楚明了。

在Pets & Pets公司，亚历克斯和他的团队做事一直力求准确。他们发现了系统漏洞，而IE6是罪魁祸首。随后亚历克斯又与产品开发团队证实了这些研究结果。尽管工程师因被发现了错误感到尴尬，但却对亚历克斯的发现并不感到惊讶。他们在短时间内匆忙地启动了英国市场的试点。虽然他们想到过一些修改可能不利于IE6浏览器运行效果的措施，但是他们仍旧选择在最后期限前完成浏览器的优化和测试。接着为解决这个问题亚历克斯提出了自己的建议。工程团队的成员不仅予以支持，并且为他提供了估算时间和资源帮助完成这些修改。对此他们承诺一旦得到布兰妮的赞同，他们将会最优先考虑此事。随着他的研究结果得到验证，亚历克斯计划在两天内与布兰妮会面，并准备了可行性建议，其中包括一个解决问题的预估时间表。

步骤5：提出建议

杰夫向汤姆演示了80张幻灯片，通过强大的模型说明了技术明细（例如，错误分类、ROC曲线等），并且凸显了团队的建模能力。五分钟后，汤姆打断了幻灯片演示，问道：“利用这个模型，我们能做些什么不同的事情吗？”杰夫回答不出这个问题，所以汤姆失望地走了出去。杰夫不知道为什么汤姆不满意这样一个复杂的模型，这可是他的团队争分夺秒、加班熬夜，在很短的时间内建立的。

提出建议是BADIR框架最后的，也是最重要的一步。为什么？因为组织要求分析过程的首要目的是解决问题。如果你已经很好地完成了分析过程，那么你的建议对团队来说将会变得实用，而建议“提出”这一步将是你推动利益相关者采取行动的机会。

除了解决问题之外，你还可以用你的建议达到三个目的（当然，也为你的团队赢得荣誉）。

第一，你想要打动你的听众。这意味着你的建议需要简短、有见地，不

要纠结细节。当然你自己需要知道细节，但除非对方发问，否则不需要说明。第二，你希望被视为一个有价值的业务合作伙伴，如果你的提议被认为是合理和可信的，这点将会实现。第三，你想促使听众按照你的建议采取行动，并有所成效，如帮助他们成功地解决商业问题。

根据加里·克莱因（Gary Klein）在《如何做出正确决策》（Source of Power: How People Make Decisions）（麻省理工学院出版社，1999）一书中的观点，你的客户不会自己花时间去整合你所有的论点。因为他们已经委托你根据商业背景来做最好的解决方案，他们将会偏向第一个合理建议以寻求解决方案。因此你的推荐必须经过充分的研究，力求精准（见图4-28）。

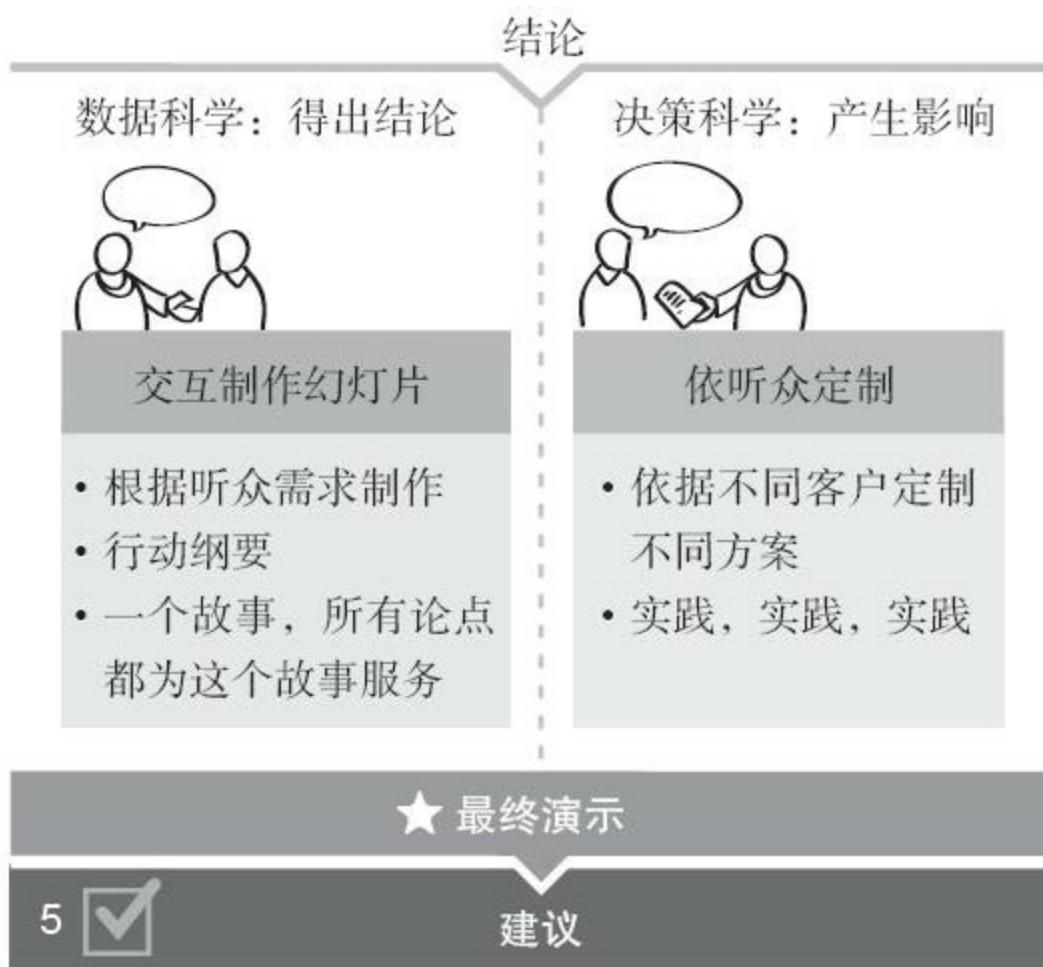


图4-28 BADIR分析框架：步骤5——提出建议

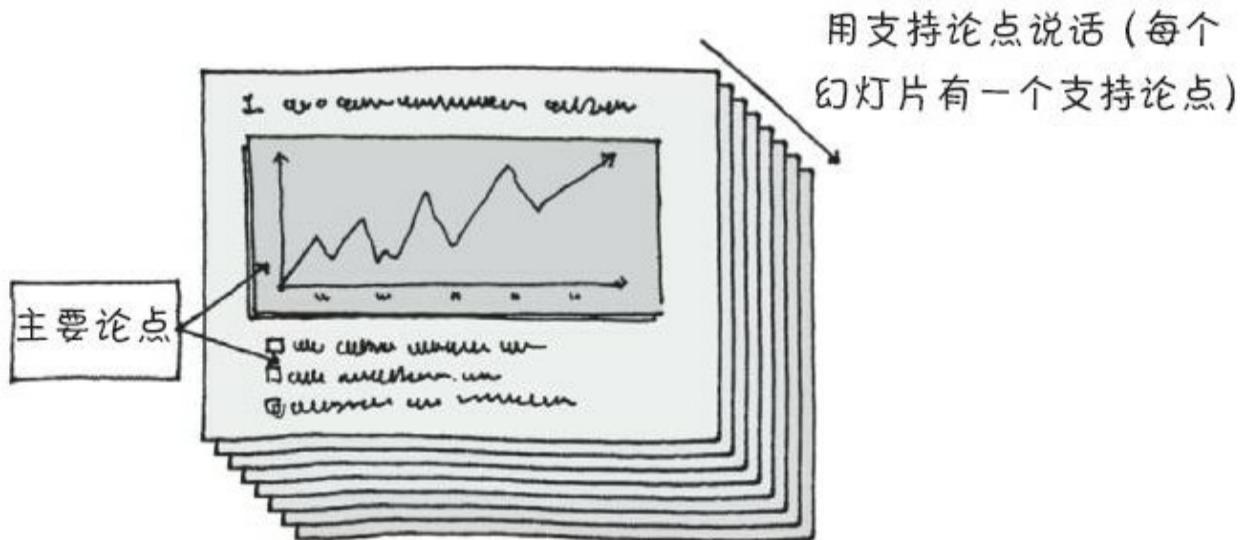
步骤5：提出建议——数据科学层面

简明一点，开头就提出行动纲要，且应该简短、清晰、有吸引力。不过，最重要的是为听众给出以下关键问题的答案：“这对我有什么作用？”尤其是如果听众没有参与最初讨论就来到这个阶段时，内容提要应当包括如下7个部分（针对不同的客户，有些部分是可选的）。

- (1) 目的：商业问题和/或分析目标是什么？
- (2) 背景（可选）：你想让大家知道怎样的分析背景？
- (3) 范围（可选）：分析中包括哪些和不包括哪些？
- (4) 方法（可选）：哪些定义（指标、分析对象）和方法应该解释一下？
- (5) 建议：你希望利益相关者对你的发现做出怎样的反应？建议应当实用且产生积极影响。
- (6) 有影响的关键论点：关键结论从哪里得出？支持你建议的每个观点的影响是什么？
- (7) 下一步：接下来预计该做什么？时间表是什么？

最好能有一张幻灯片来演示行动纲要。遵照行动纲要，接下来的幻灯片则报告论点和建议的其他必要细节。确保幻灯片的所有细节都联系实例，围绕关键信息，有论据进行支撑。最佳的决策会议一般最看重前几张幻灯片，并且几乎所有讨论都围绕第一张幻灯片展开。

统一调整幻灯片的格式，使其更容易让人接受，并且记得要图文并茂。如果你的报告很长，那么在各个主题之间最好使用分隔幻灯片进行切换。



幻灯片的制作是交互的。我们通常分三步：

第一步，将你所有的观点和表格堆叠在一起，然后逐渐削减，留下关键的结论和主要信息。

第二步，为内容提要添加关键建议和结论。可将具体细节放入附录中。

第三步，梳理幻灯片流程，添加幻灯片支持论点。最后删除多余的东西或者移到附录中。

步骤5：提出建议——决策科学层面

准备报告时须时刻记住你的报告对象是谁，了解他们想知道什么。为不同的客户制作不同的幻灯片。另外，需要注意幻灯片发布方式——人工发布或者通过电子邮件和企业内部网发布。如果你在一群真人面前演示报告，则需要给对方消化吸收的时间。你可以提出问题引发讨论。一个成功的演示报告应该会引发很多讨论，因此是需要提倡的。



祝贺你！现在，你已经了解了BADIR分析方法的全部内容（见图4-29）。



图4-29 BADIR：从数据到决策的五步框架

亚历克斯在一个四页的幻灯片上报告了具体的建议，以内容提要开始，随后引出具体行动。正式建议是具体的、可操作的并且得到利益相关者认可的。

(1) 通过研发产品修复系统漏洞，诊断IE6执行问题。这样，700万美元损失中的290万美元可得到修复。预计5个工作日完成开发，所需成本

为15000美元。

(2) 在研发产品解决上述问题时，进行假设驱动分析，解释其余访问成交率的下降原因。

布兰妮很高兴此项测试并没有造成任何损失。她恢复了信心，一旦这些问题被解决，新的流程最终会促成更高的成交转化率。亚历克斯不仅为该组织做出了重大的贡献，现在还被布兰妮和首席执行官视为一个重要的业务合作伙伴。

测试

如果你使用BADIR分析框架去获取论点，那么，在你完全推出产品或服务之前，需要进行测试以减轻风险。测试将对你基于分析所得的结果是否可真正发生进行验证。换言之，分析和结果之间是否真的有必然关系。例如，负责“高价值客户”的URetailnline产品经理发现，注册URetailnline会员（年度会员，两天免费送货）的客户就是高价值客户。这是一个很棒的论点，产品经理可以利用这一点去激励客户注册URetailnline会员。

然而，直到他使用一个小样本进行测试，并且看到业务量增加了，他才确定注册URetailnline会员样本（A）是向高价值客户（B）转变的真实原因（即 $A \rightarrow B$ ）。因为这也可能只是简单的反向原因：高价值客户可能会注册URetailnline会员（ $B \rightarrow A$ ）。如果后者是真的，那么刺激注册会员只可能增加会员注册数量，但是并不能增加高价值客户的数量。

要进行更深层的测试，请登录Aryng咨询公司网址aryng.com，查看BEDIR分析框架讲解中的完整A/B测试训练。

第5章将进一步讨论更先进的分析方法——预测分析概述。

小结

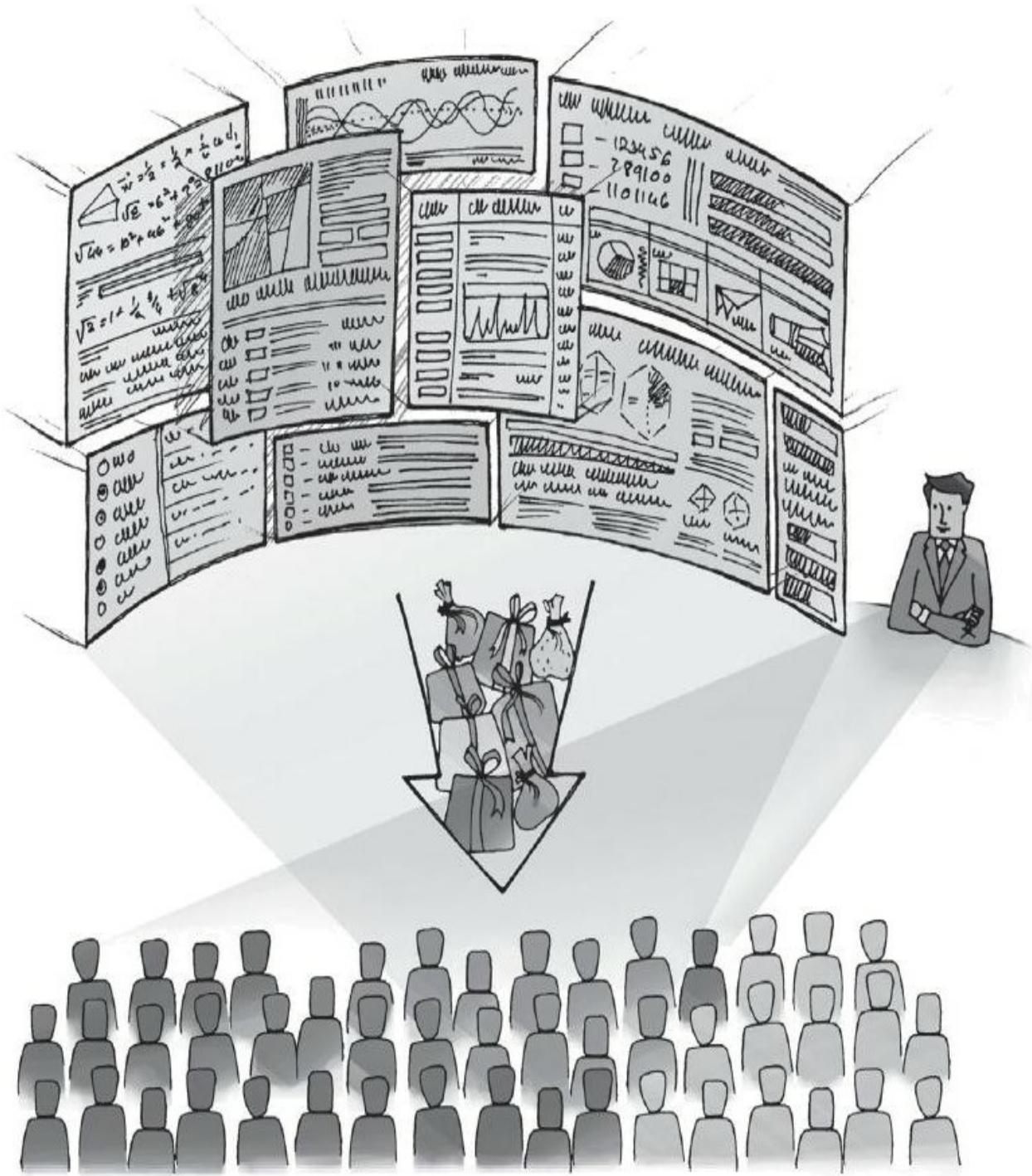
- BADIR是一个用于解决商业问题的高效数据-决策框架，适用于各个行业，简单易行。

- 同时遵循技术层面和商业层面的要求，利用数据和BADIR分析框架去得出结论并产生影响，就能一点一点地真正解决商业问题。

第5章 预测分析

本章将讨论：

- 预测分析简介。
- 怎样利用BADIR框架建立预测模型。



2月夜晚的风比平时的更大，布鲁斯·麦考利（Bruce McCauley）把他的银色本田雅阁开到加利福尼亚州西圣克鲁斯公寓的停车场上。他的心跳得异常快。他对这场求婚已经计划了好几个月了。他打开了一个红色的小盒子，戒指上的钻石在昏暗的路灯下闪闪发光，照亮了他的脸。他和凯特（Kate）已经在一起一年多了。他一早精心策划了第二天的求婚计

划，他想将这枚戒指藏在汽车的手套箱里。起风的傍晚逐渐变得寒冷多雾，布鲁斯睡觉前透过他卧室的窗户看着车，觉得它在外面浓雾里看起来格外梦幻，逐渐被吞没在这夜晚的城市里。

同时，在城市的另一边，中尉查德·福斯特（Chad Foster）——圣克鲁斯警察局第三分局的警官，开始值夜班。他打开电台开关，调到警用频道，按下话筒，通知调度“10-8在值，随时待命”。上夜班后，他把车停在一块空地上开始赶文案工作。然后，他的收音机响了。MASC（数学建模和模拟犯罪系统）刚刚发出了犯罪预测信号。一个有组织的犯罪活动即将发生在西圣克鲁斯的停车场上。MASC甚至准确定位到距离事发区250英尺远。“10-4，在路上”，他传回调度。

关掉车灯，他偷偷地审视了似乎毫无人类迹象的公寓停车场，然后看见两个影子从远处走向一辆车。当福斯特抓到他们时，他们已经闯进一辆银色的本田雅阁，准备逃脱。福斯特以盗用汽车罪将其逮捕。他认出其中一个嫌犯是以前曾因盗窃而被捕的重罪犯。罪犯的帮凶则因武装抢劫罪而刚出狱，正处于察看期。福斯特中尉拘留他们时，从一个小红盒子里发现了两样东西——一枚戒指和一个9毫米半的自动手枪。

福斯特微笑地将罪犯运回管辖区。利用优秀的数据和预测分析能力，MASC将警局从一个被动反应型组织变为主动出击型，从而实现了在损害发生前就阻止犯罪。麦考利和他的未婚妻第二天一大早就去感谢福斯特。他们感觉自己非常幸运，并且认为这都是警局的功劳，但他们从没意识到工作在幕后的预测分析发挥的作用。

事实证明，预测分析在这个部门发挥了更大作用，绝不仅仅只是对犯罪地点进行预测。随着服务人员减少了20%，服务电话增加了30%，该警局必须更有效地调配资源。部门领导认识到，他们有大量的高质量、充分验证的数据，所以他们开始关注如何充分使用这些数据等一些关键问题。通过分析，这些问题可以回答包括“我们如何才能减少响应时间”“每班需要多少人”“每班的具体时间如何安排（预测分析的重大胜利）”“如何有效分配我们的有限资源”和“在没执行任务的时候（福斯特经常用来赶文案工作的时间）我们如何分配职员的时间”。

分析工作的见效很快。该警察部门依靠预测分析进行了超过25次逮捕，在开始使用分析的19个月内降低了19%的夜盗率。虽然警局轮流值班的制度没有改变，但是分析法让他们知道了工作重点，提高了工作效率。



第5章介绍的都是分析的神奇一面——预测分析。它是复杂的，但它的影响可以是巨大的。本章分为两个部分，第一部分是预测分析的概述。它讨论了预测模型的适用范围和如何运用预测模型产生最大的商业影响力，对于所有读者都适用。第二部分探讨了如何使用BADIR框架建立预测模型，以及何时进一步深入预测分析。这一部分内容为希望使用预测分析来应对商业挑战的专业人员服务，方便他们与数据科学家之间达到合作。

我们不指望单靠这一章就把你变成一位熟练的预测分析模型专家，但我们希望它能让你在解决业务问题时具有丰富的预测分析知识，能成为一名高效的预测分析专业人员。

I：预测分析及其常见应用

什么是预测分析

过去能预知未来。预测分析也称高深学科，它采用统计技术分析当前和过去的事，做出对未来事件或行为的预测。它也可以用来提高商业分析结果的准确性。预测分析最常用的统计技术是回归法和分类方法。回归法为一个连续的变量（我们的预测目标）建立模型，通常使用线性回归，比如，建模说明一个客户的终身价值。分类法则是通过决策树或逻辑回归着眼于离散的目标。

以销售服装和配件的网上电子商务公司为例。公司拥有40%的购物车成交转化率。也就是说，100个把东西放在他们的购物车中的人，有40%会实际结账和支付。产品经理正试图确定最终成交的影响因素，以此增

加成交率。用不同颜色的购物车按钮进行测试，分析测试数据显示，蓝色按钮有高出2%的成交转化率，这是强大的信息。因为经理只需使用一个蓝色的购物车按钮就可以从相同的起始访问者中赚取到增量收入。此外，三步结账（控制）与两步结账（试验）之间的一个相关分析测试也让他发现三步结账具有较高的成交转化率。

大家可以想象，还有更多的变量会对成交转化率产生影响。要了解每个变量的相对影响，管理者需要借助一个组合方程，其中包括所有的影响变量，这样才能解释成交转化率的变化。因此，产品经理与数据科学家合作，建立了一个预测模型，使他能够理解成交转换的影响因素。在这个模型中，可以将这些假设属性放在一起，分析和理解一个变量（行为或事件）和成交转化率之间的预测相关性。然后，他只需简单地通过一个或多个预测变量就可影响最终的成交转化率（假设相关性也具有因果关系）。

他们发现，首要影响因素是谷歌结账的使用。它的结账过程中没有横幅广告，是主要的支付方式选择。此外，人们使用蓝色购物车按钮，三步结账，并且网页加载时间小于7秒。如果能做到上述几点，模型预测出成交转化率将增加7%。带着这些结论，产品经理现在可以进行测试，然后将这些变化运用到结账过程中，去获得成交转化率的增加。因此，预测分析通常会比单独使用商业分析更能产生成效，因为在这个过程中对大量的变量和它们之间的相互作用进行了评价（见表5-1）。

表5-1 案例：识别促使访问者转变为买家的关键参数

分析技术	参数预估	关键结果	利与弊
商业分析	红色和蓝色按钮	对测试数据进行相关性分析，表明使用蓝色结账按钮可获得高出 2% 的转化率	[+] 快速简单 [-] 不彻底 因此可能会错过最优参数，阻碍最大商业成果的实现
预测分析	多重参数建模	首要影响因素——使用谷歌结账，作为主要支付方式的选择。在结账过程中没有横幅广告，使用蓝色购物车按钮，三步结账，并且网页加载时间小于 7 秒	[+] 通过业务分析建立论点 [-] 时间和资源密集型 [+] 可以评估数以百计的指标，从而确定最重要的参数，以推动实现最大的业务成果

在日常生活中，我们其实经常遇到预测分析，只是没有进行确认罢了。FICO评级是美国最有名的和最广泛使用的信用评级模型。随着你消费并支付账单（或不支付账单），人口和行为数据就被收集，而FICO正是基于这些数据在进行预测分析。

FICO评分或者更准确地说，是FICO评分模型能使银行和信贷机构了解其客户的信用程度。这样，这些机构就可以确定哪些人的信用风险高，由此可以拒绝发放贷款给那些不太可能偿还贷款的人。如果你的组织要向客户发放贷款，那么它就可以使用一个类似FICO的模型。

预测分析的常见商业应用

除了信贷行业，传统上营销和广告会使用预测分析，目标是增加销售空间和提高营销活动的投资回报率。例如，基于信用模型（比如FICO和一个响应模型）的调查结果，你收到了信用卡优惠办理邀请。根据这些调查，适当的优惠会发给最有可能回应的具有顶级信誉的人。但问题是，公司是如何知道你是有信用的，为什么你更有可能做出回应？这是

由你之前的活动和行为，以及其他具有相似属性的个人的活动和行为来判断的。亚马逊检索工具可根据你的购买和浏览历史向你建议下一个你可能喜欢的产品。它是一个很好的促销预测模型的例子。当打开你的LinkedIn界面时，你就会看到预测模型为你选择的最相关的条幅广告。根据你的相关个人资料，你点击此广告的可能性非常大。

最近，我们也看到预测分析被广泛地应用在改善客户体验方面，例如保留和流失模型、客户服务操作、客户体验优化模型等。当客户放弃一个产品或服务时就产生流失，这是行业都担心的事情，但这是能被预测的。因为在客户放弃一件产品前会留下迹象，预测分析可以检测出这些迹象，接着主动识别出这些客户，并相应地重新激发他们的购买欲望。例如，你打电话给你的电话公司，可能被传送给一个远程在线的客户代表、一个国内客户运营中心或坐落于巴西的近海客户运营部。具体转接到哪里，是根据你的预测终生价值，以及特定的客户体验属性来决定的，比如你和一个非英语为母语的人交谈是否会感到舒服。

游戏、娱乐和移动行业在产品管理中广泛使用预测分析，用来优化它们的产品和推出的功能，提高产品的采用率和盈利能力，并增加成交率。当你看到Netflix上的电影推荐时，你就知道这是预测分析的应用结果。当你十几岁的女儿同样使用Netflix账户时，你可能不愿意总看到这些互相冲突的推荐。Netflix公司最近已经开始在各个共享账户里发布了多达5个分离文件来解决这个客户问题。

现在让我们根据我们的经历来给大家介绍一些现实生活的例子——预测分析如何显著地推动商业成效。

增加市场空间

一个软件零售商将适销对路的领域扩张了4倍，从而获得了3500万美元的新收入。

背景：一个专为迎合创意专业人士的知名软件供应商为它的次优（第二受欢迎）产品（称为产品X）寻找新的潜在客户，现行做法是向用户推销产品X的早期版本和试用下载新版本。分析的任务是为这个产品找到一个新的收入来源。

行动：零售商有一个包含5000万个潜在客户的数据库，这些用户都下载过其他产品的免费版。我们认为这是开发潜在客户的绝好机会，因为这

一个产品与用户所需的产品有相似的属性，但它们并没有展现产品X的任何突出优势。使用逻辑回归法，根据产品X过去的用户属性，我们用采用概率积分（采用产品X）方式来为每一个潜在客户评分，然后向前400万个潜在客户提供了产品X的最新版本报价。

影响：该模型将供应商的潜在市场扩张了4倍，增加了3500万美元的直接收益。另外，这一潜在客户群成为未来发行所有新版产品X的增量收入来源。

降低产品摩擦

一家支付公司通过减少产品摩擦，获得了1800万美元的增量收入。

背景：一家在线支付公司的产品结账率很低。产品经理曾经根据市场调查了解消费者在结账前放弃产品的原因，初步市场研究报告发现了许多客户摩擦点，但该组织不知道每个摩擦点的相对大小。此外，领导层并不相信消费者自我报告信息。实现1%的成交率改善将预示着1000万美元的收入增加，这被认为是一个巨大的机会。

行动：我们使用逻辑回归法建立了一个成交模型，用它来分析是什么造成了成交率不高。我们还使用了决策树和商业逻辑来判断最有可能的原因并量化各个因素的可能影响。然后，我们与产品经理合作，根据模型找出的首要原因发布了一个测试版本。

影响：仅仅关注了一个成交影响因素，就使该组织获得了1800万美元的增量收入。

激活休眠客户

一家金融服务公司通过激活休眠客户，获得了2000万美元的收入。

背景：一家金融服务公司有一批账户已经休眠超过12个月的客户，无论是降低APR（年利率）还是承诺增加他们的信贷额度，对这些客户都没什么影响。

行动：我们使用逻辑回归法建立一个损耗模型，用来确定造成客户休眠的因素。为了量化留住这些客户的意义，我们还为每个关键的利润影响因素建立了模型，确定从每个保留客户中能获得预期利润。最后，我们

为每个消费群体开发了多个业务套餐，在上述模型的作用下仍然能保住盈利。

影响：确定并推出优惠，重新激活了这些客户，产生增量收入2000万美元。

产品推荐引擎

一家支付公司通过产品推荐引擎获得了超过2000万美元的增量利润，降低了70%的营销支出。

背景：一家支付公司有多种支付方式满足客户（即商户）需要。据观察，拥有多种产品的商家会更多地参与进来，并对产品有更广泛深入的使用，这将提高公司的盈利能力。所以，营销团队决定通过营销活动，向每个商家提供每一种支付方式，但这却造成了商家的混乱和高退订率。所以，营销主管想知道如何找到次优产品（NBP）推荐给各商家，从而优化商家对支付方式的接纳并提升公司的盈利能力。

行动：我们使用一个多级决策树来判断，针对不同的商家群看哪些支付方式具有较高的使用率。接下来我们结合数据和历史利润分布研究各个商家群每种支付方式的预期增量利润（EIP）。通过这个推荐引擎，EIP最高的产品将作为次优产品推荐给具有类似属性的商家群。接着在数据库中对该模型进行评分，用于推播式营销和集客式营销。

影响：通过使用NBP推荐引擎，营销团队现在知道了应向哪些人推荐哪些产品，使得营销支出的占比减少了70%。此外，使用NBP评分，可提供正确的产品，仅对外营销就增加了6倍成交转化率，带来了超过2000万美元的利润。

总结

支付公司、软件零售商和圣克鲁斯警察及其他许多人发现，预测分析能针对市场不同的群体，更好地规划产品走向，建立了更广泛的消费基础，并增加了公司的利润。

II：建立预测模型——理论和实践

本部分内容对于两类读者最为有用：从事商业分析工作的商业专家和成

熟组织的数据分析师。我们将介绍预测分析和模型构建的术语和理论。了解这些将有助于商业专业人员更轻松地以及更专业地加入与同行科学家的讨论中。接着，我们将介绍如何建立一个预测模型，并在本章中介绍一些额外的资源，帮助你将来进一步深入学习预测分析。

术语

让我们从变量、预测因子和时间窗口开始。我们将使用本章第一部分购物车的例子来说明术语合适的用法。

- 因变量：**指特定预测分析的对象。它们是由模型想要解决的商业问题所决定的。例如，成交转化率就是一个因变量。
- 自变量或独立变量：**指可能与因变量有关系，但相互间没有关系的未知数，由以解决商业问题为目的的假设决定。之前例子中的蓝色按钮和结账过程中出现的广告横幅便是独立变量。
- 预测因子：**已被发现与因变量有关系的未知数。预测因子和独立变量这两个术语有时可以互换使用。
- 时间窗口：**如果你想建一个模型，有两个时间窗口很重要：观察窗口和预测窗口（见图5-1）。观察窗口（T0-T1）是你的一段观察期以及收集独立变量进行分析的阶段。你需要确定观察窗口的起点T0，确保建模所用的样本数据来自观察期。以我们的购物车为例，如果你决定对所有的购物车活动观察一个星期，那么你的观察窗口即为一个星期。

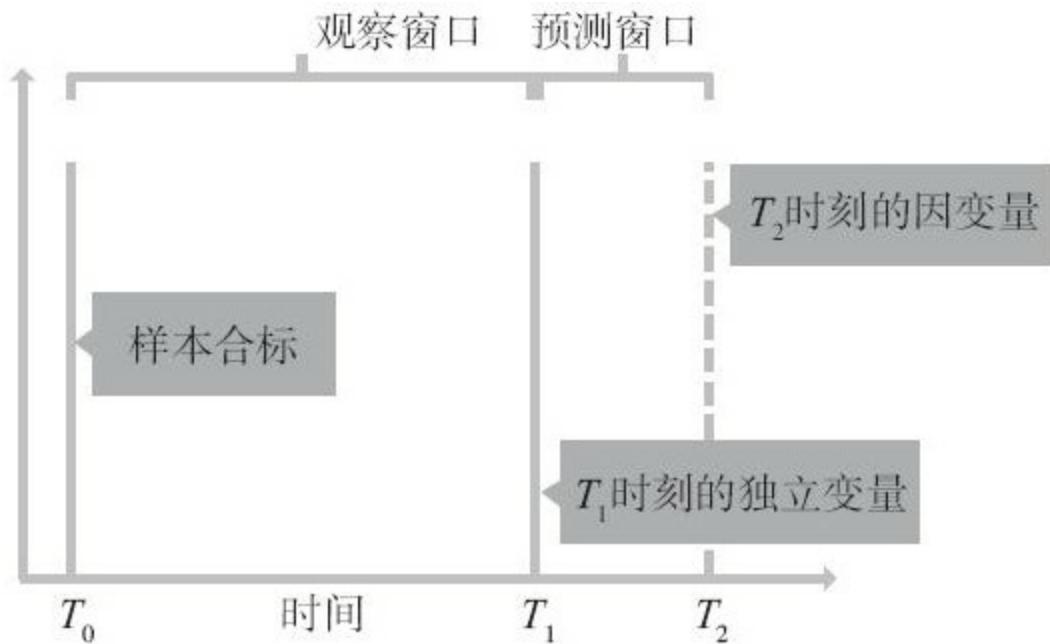


图5-1 观察和预测窗口

预测窗口（T1-T2）是用来预测因变量的时间段。你需要T2时刻的因变量值，即在预测窗口结束时的值。预测时间窗口可长可短，取决于你要预测的未来有多远。一般来说，预测窗口越长，预测的误差越大。然而，一个较短的预测窗口对商业的作用也较小，因为在预测的目前状态之前，业务采取行动的时间很短。在预测窗口，你可以设定30分钟的预测期，预测各段对话的成交转换率。在这段时间内，你将观察成交是否发生。

- 相关性：指两个或多个随机变量之间的线性关系的统计测量，以相关系数（r）来表示，取值在+1和-1之间。预测模型是利用T2时刻的因变量和T1时刻的独立变量之间的内在关系。如果相关性越强或是找出越多的潜在相关因素，模型的准确性就越高，这就是为什么一定要提出一系列好的假设。

记住，所有的相关性都从BADIR框架的步骤2所提出的假设开始。以购物车案例为例。分析是从假设蓝色按钮与更高的访问成交率有关开始的。当你发现越来越多的潜在相关性时，你的模型会有越来越强的预测能力。

- 错误分类：一个评价模型性能的方式。这是错误预测总数和预测总数

之间的比例。它是扩大容忍度以获得更多可能的结果（敏感性）和收紧容忍度以得到更具体的结果之间的权衡（准确性）。错误分类率越低，模型就越好。让我们回到圣克鲁斯警察局的例子。如果警察想抓住的不仅仅是两个偷车贼，他们就必须修改他们的模型（增加敏感性和降低准确性），但这样的话他们可能会抓住一些好人。如果他们增加准确性只抓坏人，那么他们可能会放过一些不符合这些特定属性的坏人。

这两种类型的错误分类，敏感性衡量的是事实为正，模型也预测为正的比例（真阳性，TP），而准确性衡量的是事实为负，模型也预测为负的比例（真阴性，TN）。与TP相对的是假阴性（FN），与TN相对的是假阳性（FP）。举例来说，通过分析，犯罪模型检测出的坏人的抓捕率为13个人中抓到了11个人（TP=11； FN=2； 敏感性=11/13），但在这个过程中也从16个好人中抓到3个坏人（TN=13， FP=3， 准确性=13/16）（见图5-2）。如果任务是必须抓到这两个坏人，模型可能提高模型参数去抓更多的坏人（提高敏感性），但最终也可能会抓到更多的好人（减少准确性）。根据商业目标，有时我们需要放弃更高的准确性，以获得更高的敏感性，反之亦然。这需要由建模者应与他的业务伙伴一起做决定。



图5-2 模型的效果

常用的预测技术

线性回归法、逻辑回归法、决策树、k-均值聚类分析、时间序列预测、生存分析和神经网络分析是商业上常用的预测技术（见表5-2）。其中，前三个是最常用的。

表5-2 预测技术和它们的应用

技术	描述	应用
线性回归法	标量因变量与一个或多个独立变量之间线性关系的模型方法	客户终身价值；收购成本
逻辑回归法	线性回归的特例：因变量是二维虚拟变量	流失或损耗模型、舞弊检测模型、响应模型
决策树	树图，常常用来确定最好的总体分类类型（根据独立变量）来优化因变量的预测	交叉销售产品预测；客户分群
k-均值聚类分析	细分技术，将总体（观察数）划分为不同类别，同一类别的变量均值接近	无监督的客户细分，即依据相似属性归类，不受客户流失或客户终生价值等任何目标变量影响
时间序列预测	基于同一事物已知的过去价值，预测未来事件的技术	预测一段时期的销售情况
生存分析	常用来预测事件结果出现的时间	达到信用限额；客户任期
神经网络分析	现有统计模型的概括；黑箱；难以理解，但比其他技术更强大	舞弊检测、响应模型和其他

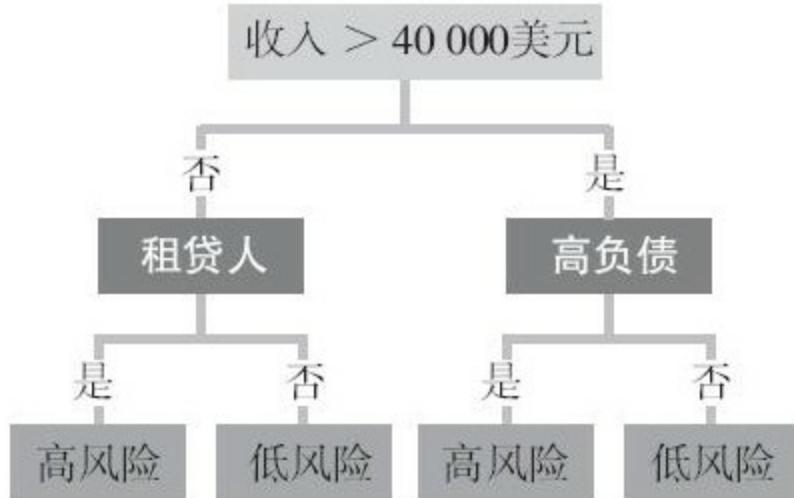


图5-3 决策树法

- 线性回归法：一种对连续因变量（y）与一个或多个连续的独立变量（x₁, x₂, ...）之间的关系进行量化建模的统计方法。最常见的应用是预测客户终生价值（CLV），例如预测一个客户一生能带来的总收入（y=CLV）。通过建立模型，你发现CLV的首要预测因子是客户至今为止的历史年收入x₁、年度所得x₂等。线性回归预测y是x₁, x₂, ..., x_n的一个函数。这条函数线能使每个数据点到它的距离平方和最小。用公式表达如下：

$$y=a+b_1x_1+b_2x_2+\dots+b_nx_n$$

- 逻辑回归法：这是回归法的一种特例，在逻辑回归中，因变量是不连续的（见图5-3）。逻辑回归的因变量是离散的或分类的，主要为0-1二分变量。当面对一系列独立的决策或离散的行为，如客户流失和舞弊预测时，逻辑回归法是最常用的。假设我们正在预测一位客户是否会流失。在这个等式中，p表示流失概率，用流失发生率 $\log(p/(1-p))$ 取代上面公式中的y即可。当你建立模型时，你会发现客户是否流失的首要预测因素是花费在帮助页面的时间x₁、收入水平x₂等。逻辑回归模型预测客户是否流失是x₁, x₂, ..., x_n的一个函数。用公式表达的话，它看起来像这样：

$$\ln\left(\frac{p}{1-p}\right) = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

●决策树：根据递归划分产生的数据，制成自上向下的分类结构。即通过反复的评估和每个节点划分，直到该模型能够提供所需的结果。决策树是一种贪心算法，有助于迅速确定最重要的预测因素。这是最直观的技术，使自己融入有规则的情境中。在我们的例子中，我们用决策树来帮助金融机构预测向客户扩大信用的风险。如果客户有超过40000美元的收入，没有高负债，或如果客户的收入虽然低于40000美元，但他们不是租货人，那么决策树会输出规则：他们的信用良好，可以信贷给他们。在这里，因变量为风险（好与坏）、收入、高负债，信贷者状况则为预测因素。

怎样建立预测模型

无论是预测分析还是商业分析，BADIR框架均与利用数据驱动决策有关（见图5-4）。当用作预测分析时，使用的具体活动和事件会在BADIR步骤2到步骤5中有所区分。



图5-4 使用五步BADIR框架进行预测建模

步骤1：商业问题

预测分析和商业分析在这一步是一样的，在确定使用什么方法之前，这一步是非常关键的。如果你的分析没有出错的话，你应从第一阶段描述的简单商业分析方法开始，到后续阶段中，再使用更先进的技术（如预测分析）来证明。

上面的示例中，我们通过优化次优产品的（NBP）建议，帮助支付公司降低了70%的营销支出，利润增值为2000万美元。产品团队已经使用简单的分析观察到，多项产品所有者比单一产品所有者更有利可图。他们在过去进行大规模销售，将每个产品推荐给每个商人，结果增加了退订率，使商家烦恼该采用哪种支付方式。在步骤1中，我们使用商业问题框架，发现真正的商业问题是：“我们应该给哪些人（商家）提供哪种次优产品，以及这样会获得多少利润？”产品开发主管是关键角色，但营销主管也是一个关键利益相关者，因为她的团队将利用该模型去向商家营销。如果有必要的话，也可以进行未来三个月的预测分析。

步骤2：分析计划

预测分析和商业分析相比，其分析目标和假设的推出也是相同的。在方法论阶段，可决定使用简单的分析技术还是更复杂的高级分析技术。预期收益、决策时间和项目可用资源都是决定因素。如果选择预测分析，则还需要额外的参数，包括时间窗口和样本选取战略，这些将被列入数据规范内。确定项目计划时，要与内部和外部都进行了交谈，分析计划才算是完整的。然后，在启动会议上，还需要利益相关者同意此计划，相关人员要同意加入。

仍以上述支付公司为例，我们将分析目标缩小为：“依据一个积极的商家经历，将其过去采用的产品作为指标对其进行优化，对个别商家做出次优产品评估，增加利润。”我们基于利益相关者的头脑风暴，产生超过200个假设，并选择将他们全包括进这一阶段，因为这对预测分析是有利的。相比于商业分析，预测分析的重要优势在于评估大量假设的能力，并使用先进的统计方法来快速地完成。但请记住，如果你有更多的假设，无论你采用哪种方法，数据收集都会花费更多的时间。当下一个最好产品是一个离散的结果时，我们决定使用决策树和逻辑回归作为预测分析方法。有时候，使用不同的技术可建立多个模型，从而选择错误率最低的最佳模型。我们为所需历史数据进行数据规范。为优化公司利润和体验的推荐引擎，我们设计了一个创新的启发式近似法，为下一个具体产品的采用概率添加利润指标。在启动会议上，客户非常看重我们的方法，因为这为他们提供了一个简单的解决方案来优化客户体验，客户通过下一个产品的采用实现了利润的增加。一旦我们的客户达成既定的目标，利益相关者将承诺给出他们的最优资源和时间来促成这个项目的成功，这是任何分析项目实施和成功的关键。

步骤3：数据收集

在这一步中，你会发现历史数据经常需要重新创建，需要大量的整理和验证。在这里，需要两种不同的数据集：一个用来建立模型（训练和在观察期内验证样本），另一个用来检验模型（在预测期外验证样本）。用于训练和在观察期内验证的数据来自同一时间段。训练样本用于建立模型的逻辑，以预测目标变量，而验证样本是用来评价模型的性能，看是否能够预测目标变量。然后，再用另一时间段内的样本——称为预测期验证样本，对模型进行验证。例如，首先使用2012年的客户和交易数据训练以及在观察期内验证预测未来次优产品模型，然后再使用2013年的数据作为预测期验证样本测试此模型的性能。所以，如果这一步需要80%的时间分布也不奇怪。在支付公司案例中，从分析计划到收集、整理和验证数据花费了我们将近6周的时间。

步骤4：得出结论

到了这一步，预测分析需经过大量的统计过程，包括变量的选取（选择最优独立变量），建立模型来确认自变量与因变量之间的关系，以及模型验证，直到最终确定方程和独立变量。

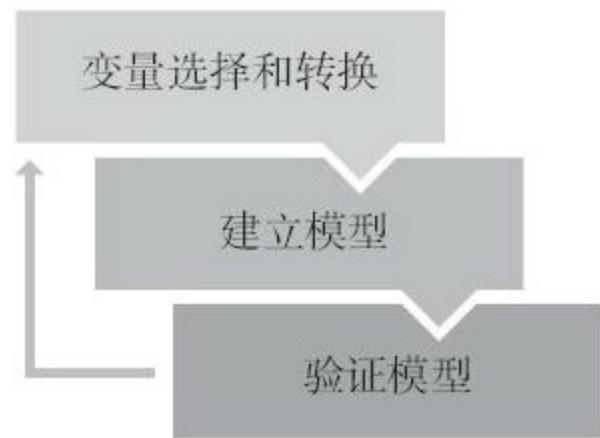


图5-5 变量选择

变量的选择是一个迭代的过程，其中需要不停地转换和处理独立变量，以增加对因变量的预测能力（见图5-5）。对于NBP模型，其中的一个独立变量是账户创建日期。处于原始状态的变量（如日期）是不能用于建立模型的，因为它不是一个连续变量，也是没有意义的离散值。因此，用账龄代替账户创建日期变量，即商家变成客户需要多久。然后，我们计算账龄与NBP之间的相关性。在此基础上，还需要对账龄进一步变换。我们发现，比起其他形式的数学变换，账龄的平方根与NBP有更

高的相关性。类似地，对所有200多个变量，都像这个变量一样被适当地进行了变换，并选取与NBP具有最高相关关系的变量形式。

接着，这个被选中的变量子集被代入模型中，其中一个方程（如在回归的情况下）或一个树形结构（如在决策树的情况下）中的因变量，将用独立变量的某种关系式来进行表示。

模型的建立也是迭代的，以上过程需要被不断重复，直到模型被确立下来。在这过程中，需不断检查错误分类和误差的影响，直到从商业视角来看，模型可被接受为止。在训练样本的基础上建立了模型之后，还需要在观察期内验证模型的稳定性。而利用观察期外的样本再次验证确认该模型，则可以确保训练样本没有受季节性因素或临时干扰的影响。

有时，为了提高预测整体精度，你可能需要建立多个模型并将它们合并（也称为整合）。正如第4章所讨论的，2006年Netflix奖就是授给多个模型的集合的。集成模型可以提高预测精度，但却很难予以解释和实施，所以这是建立模型时需要重点考虑的事情。使用预测分析很容易晕头转向，尤其是当涉及多个模型时。记住，无论你做什么，你都必须向那些冷静的或置身分析之外的人进行解释。他们只想找到答案——你的建议、你的回答必须要快，让他们能够理解，并易于实施。

一旦模型被验证确立，就该与利益相关者进行沟通，以确保能取得最终需要的预测因子等输入变量，保障模型的性能。然后，你可能需要适当地进行调整，比如，由于商业背景发生了变化，那么就需要去掉某些预测因素，因此历史数据的收集或模型参数也会发生改变。

例如，对于支付公司，我们开始有200多个变量（从约200个假设中开始）。我们将训练样本根据商户目前选择的支付方式，将基础产品类型分成许多子样本，然后使用变量选择将其缩小到60个变量，并用逻辑回归法确定前35个。接下来，我们为每个分组建立一个多级决策树（见图5-6）。一个多级决策树中自变量有超过两个的离散值，如产品A、产品B、产品C等。在最后一组决策树中，我们使用了30个随机的预测因素去制造规则，以方便利用数据库信息进行评分。例如某条规则为，如果商家使用的基础产品A，年销售额已经超过500000美元，在最后6个月要求客户服务超过3次，在过去6个月中的损失额超过20000美元，那么， $NBP=$ 产品Y；预计每年能带来增量利润（EIP）=300美元。

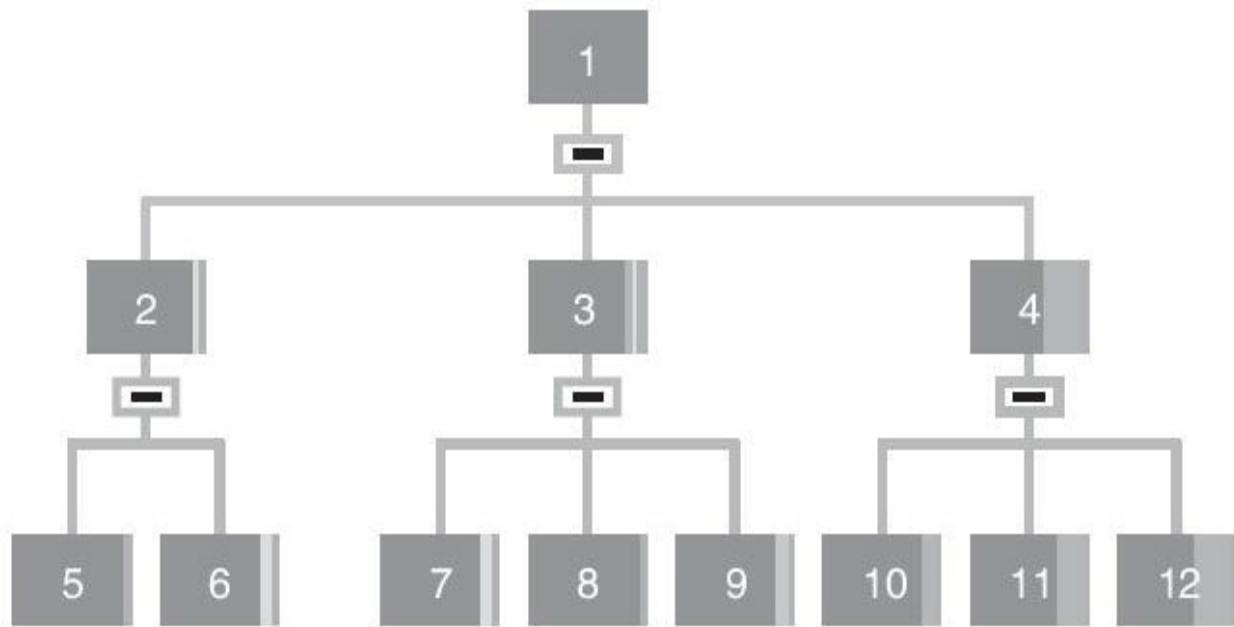


图5-6 每种基础产品的多级决策树

该模型产生了许多类似的规则，被用于评估数据库中当前的所有商户，因此每个商户都有一个NBP和EIP，然后这两个值可被营销和产品经理用于对外宣传活动，被客户服务代表用于店内促销。

步骤5：提出建议

BADIR框架的最后一步是向利益相关者展示模型，提出建议，促使其付出行动。这可能涉及模型的部署，例如在NBP评分案例中，需在数据库中评估模型逻辑；或在客户流失模型中，需靠学习来解决问题，对客户流失模型往往不采取数据库评分的方式，因为通过模型学习到的内容是用来控制流失影响因素的，例如，糟糕的客户服务体验。

预测分析项目往往在范围和影响上比商业分析项目更大，所以你可能需要针对不同的利益相关者群体定制不同版本的最终报告。在这个阶段，预测分析需要模型评分，以及针对不同听众进行多次报告演示，这是它和商业分析的主要区别。

在支付公司案例中，我们给利益相关者——产品和营销负责人提出了最终结论。它包括七个主要幻灯片，以行动纲要为开头，并涵盖模型所有细节的附录为补充。在获得利益相关者的赞许时，我们在数据库中部署了NBP模型。销售团队首先在营销活动中使用模型对产品进行了增值促

销。依靠模型，他们第一次认识到应该向谁专门去推销某项产品。他们还有一个对照样本，向一组随机选择的商家提供同样的报价。对比对照样本，NBP部分的成交转化率是它的6倍。此外，NBP部分的采用者获得了1.5倍的利润。该模型迅速取得成功，并在今年帮助营销团队获得了超过2000万美元的增量利润。NBP模型最终也被其他部门使用，包括绑定客户服务电话。

预测分析功能强大，但要小心过度使用

正如你从上面的例子中可以看到的，预测分析只是多种分析工具之一。它的作用强大，但必须被明智地使用（见图5-7）。

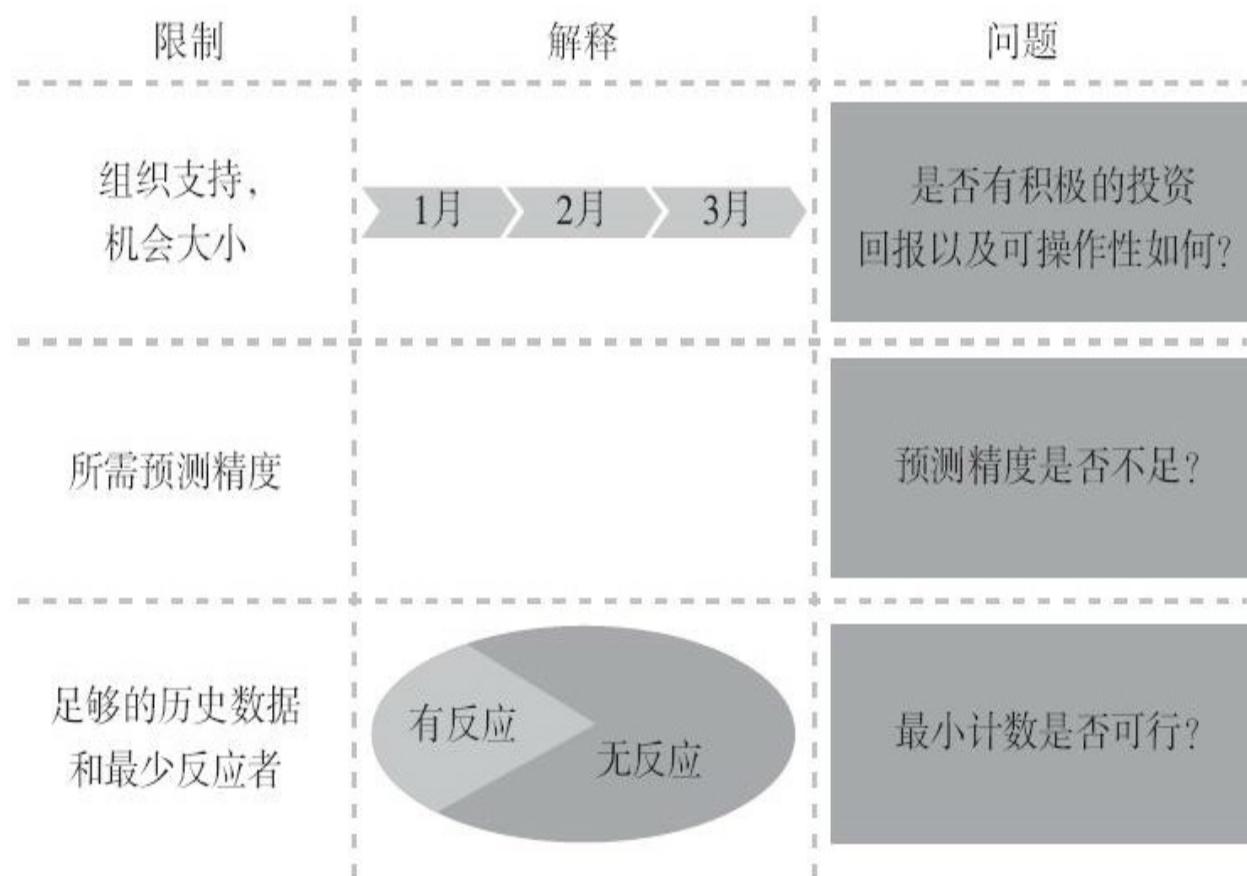


图5-7 什么时候使用预测分析

预测分析需要耗费的资源和时间是简单分析的10~20倍。它需要先进的工具和技巧、历史数据的支持、可操作化、现场验证和持续的维护。这就是为什么我们不能利用预测分析先发制人地解决每一个业务问题。如果做对了，可通过商业分析发现机会，为预测分析奠定基础。而且，它

需要有较大的成功概率和显著回报的承诺，并得到组织的支持，然后才能让预测分析有意义。

这里有一个商业分析战胜预测分析的例子。我们的一个金融B2B客户让我们调查它的风险政策和对客户体验的影响。风险小组要求客户一旦达到一定的信贷限额即须核实他们的身份，并提供额外的文件。客户发现这一要求后十分沮丧，因此该政策造成了每年1200万美元的客户流失。为拯救他们的客户，风险团队建立了生存模型（预测分析技术）来预测每个客户达到信用限额所需的时间，这样公司可以主动向客户提供额外的文件，把代理时间侧重于处理传入的文件。该模型的建立预计需要三四个月，并且我们的项目将把预测时间作为输入变量放到影响分析中。

根据我们在信用限额方面的工作经历，我们申请快速进行一些逻辑分析，以便加快工作进程。利用BADIR框架的相关分析（商业分析技术），我们在10天之内建立了一个简单的逻辑分析，正确地预测了在未来3个月将有80%的客户达到他们的信用限额。我们分享了调查结果和其中的原理，然后开始执行此项目。风险小组迅速采用并实施了我们的逻辑分析，将其用作预先检测机制。这导致生存模型项目报废，虽然它尤其能为个别客户更准确地预测达到限额的时间，但我们的整体水平预测结果对风险团队来说已足够了。它使用简单的逻辑模型为团队节约了4个月的时间，节约了有关的成本和资源。此外，还立刻能为客户节省1200万美元的收入流失，并减少与客户之间的冲突发生。

市场营销人员、产品经理或运营经理将BADIR数据运用到决策框架中，通过数据可以优化他们80%的日常工作流程（如上面的例子），并且不需要依赖于稀缺和昂贵的分析资源。对于剩下的20%，潜在的投资回报率证明了预测技术的作用，他们可以与同行分析人员一起合作。这展现了分析可以为组织获得良好的运营。相反，当组织和他们的领导人被围绕的大数据和预测性分析的炒作误导，最终一边倒地去投资先进数据分析工具和资源时，经常导致较差的投资回报率。

更多的预测分析资源

这里是一些我们为实践者推荐的阅读、活动和预测分析资源。

- 阅读与预测分析有关的入门书籍。我们推荐奥利维亚·帕尔·陆克文（Olivia Parr Rudd）的《数据挖掘实践：营销数据建模、风险和客户关

系管理》（Data Mining Cookbook: Modeling Data for Marketing, Risk and Customer Relationship Management）（Wiley, 2000）。

- 如果你是一个SAS（商业分析与商业智能软件）用户，请读洛拉·德尔维什（Lora Delwiche）和苏珊·斯劳特（Susan Slaughter）的第5版《SAS入门》（The Little SAS Book: A Primer）（SAS研究所, 2012）。
- 利用如Coursera这样的免费在线课程，学会R软件（一个免费的统计软件包）编程。
- 利用免费的统计工具，如Knime，学习建立可视化模型。
- 使用1~5天的时间参加专业培训课程学习分析方法，有效地与数据科学家合作推动你的业务影响。你不必亲手建造模型，除非你正试图成为一名数据科学家。预测建模是相当复杂的，有许多陷阱，有时甚至训练有素的统计学家也会出错。

最后，我们不建议任何人通过参加分析学交流会议和阅读博客来学习预测分析，这样会出现很多错误，因为许多热门关键词已被一些厂商滥用，为他们自己谋求获利。最好的学习方法是在一个经验丰富的建模者的监督指导下学习建立模型。

我们已经提到了一些数据和分析工具。在接下来的章节中，我们将更密切地观察它们，并告诉你如何确定你需要什么。

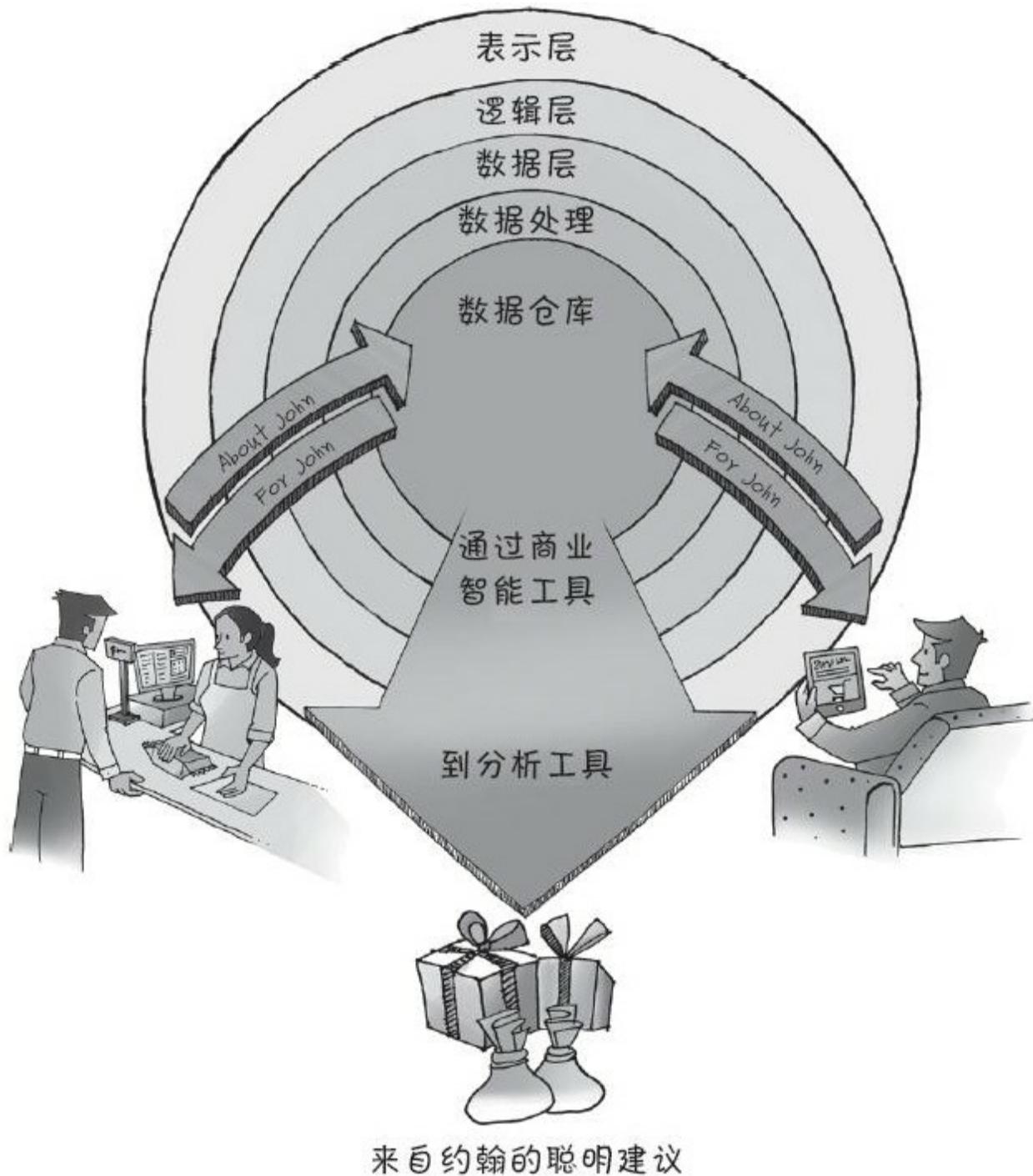
小结

- BADIR可以最大限度地扩大预测分析的潜力，否则预测分析也只是一次统计学练习而已。
- 预测分析是一个功能强大的工具，它可以生成显著的业务成效，但它需要耗费大量的资源和时间。我们要合理地使用它，且只有在投资回报率合理的情况下才使用这种方法。

第6章 数据和分析工具

本章将讨论：

- 四种类型的商业智能（BI）和各种分析工具。
- 如何选择你需要的最佳工具，以及每个类别中的常用工具是什么。



费尔南多（Fernando）刚刚加入StartMatrix游戏公司，担任产品开发主管。他向团队询问产品的运营仪表盘数据，想迅速了解公司的动态。团

队成员不情愿地为他提供了仪表板，并开始解释他们的业务动态、商业模式以及更多内容。他得知，这是一个免费增值的商业模式，赚取的大部分盈利都来自鲸鱼（高消费者），另一部分则来自广告宣传收入。随着费尔南多的进一步追查，他开始问一些尖锐的问题，并调用了报告的拖拽功能。这时，问题开始暴露出来。“从缴费基数看，我们真的只有2.68美元的平均支出吗？”“报告显示，大部分的基数是从一个表格获得的，而第一套产品在除了个人电脑之外的地方甚至都不可用。”团队承认这些数据确实不可靠：“要想知道平均花费，我们需要从财务系统里查看不同的报告，但这些数据自上个月以来就不包括新用户了。”接着，交谈继续进行。会议结束的时候，费尔南多和他的团队认为，他们根本不能相信他们现有的任何数据。

在这家公司，即使有小部分数据是准确的，但因为它们与大量不准确的数据混杂在一起，所以也是不可信的。你可以想象，由于团队成员不信任数据，就不能根据数据做决定。为了解他们的客户和产品，他们通过受控测试或实验，创造了新的数据集，并对此进行了分析，分析的依据不再是历史数据。但问题是，没有根据历史数据得出的可靠假设，大部分的测试都无法得出突破性的论点，不能推动产品参与。在这种情况下，分析成熟度评估发现，果然，数据很不成熟。这将成为他们通过数据去了解客户使用产品情况的最大障碍。他们需要一个新的信息架构、一个修改后的数据存储系统，以及商业智能工具。

在商业分析中，最重要的是要有正确的方法论。即使是一个简单的工具，如Excel，也可以很好地解决80%的业务问题。然而，作为一个专业的商务人士，你需要更好的、规整的数据以便使用，促进分析结果的得出。本章阐述了用什么样的工具存储、报告和分析数据，以提高一个组织的数据成熟度（见图6-1）。如果你知道你的数据是有缺陷的，或如果你在收集自己的数据上有困难，本章将会帮助你解决这个问题，并与专业人士进行一场有意义的交流。这将有希望让你找到获得纯正数据的路径。此外，在数据科学家的帮助下，你还可以增加挑战力度，实现你的业务目标。

数据和分析工具有四大类：

- (1) 数据存储和处理工具
- (2) 商业智能（BI）或报告工具

(3) 商业分析工具

(4) 高级分析工具

数据存储和处理工具

数据存储和处理有五个组成部分：表示层、逻辑层、数据层、数据处理和数据仓库。

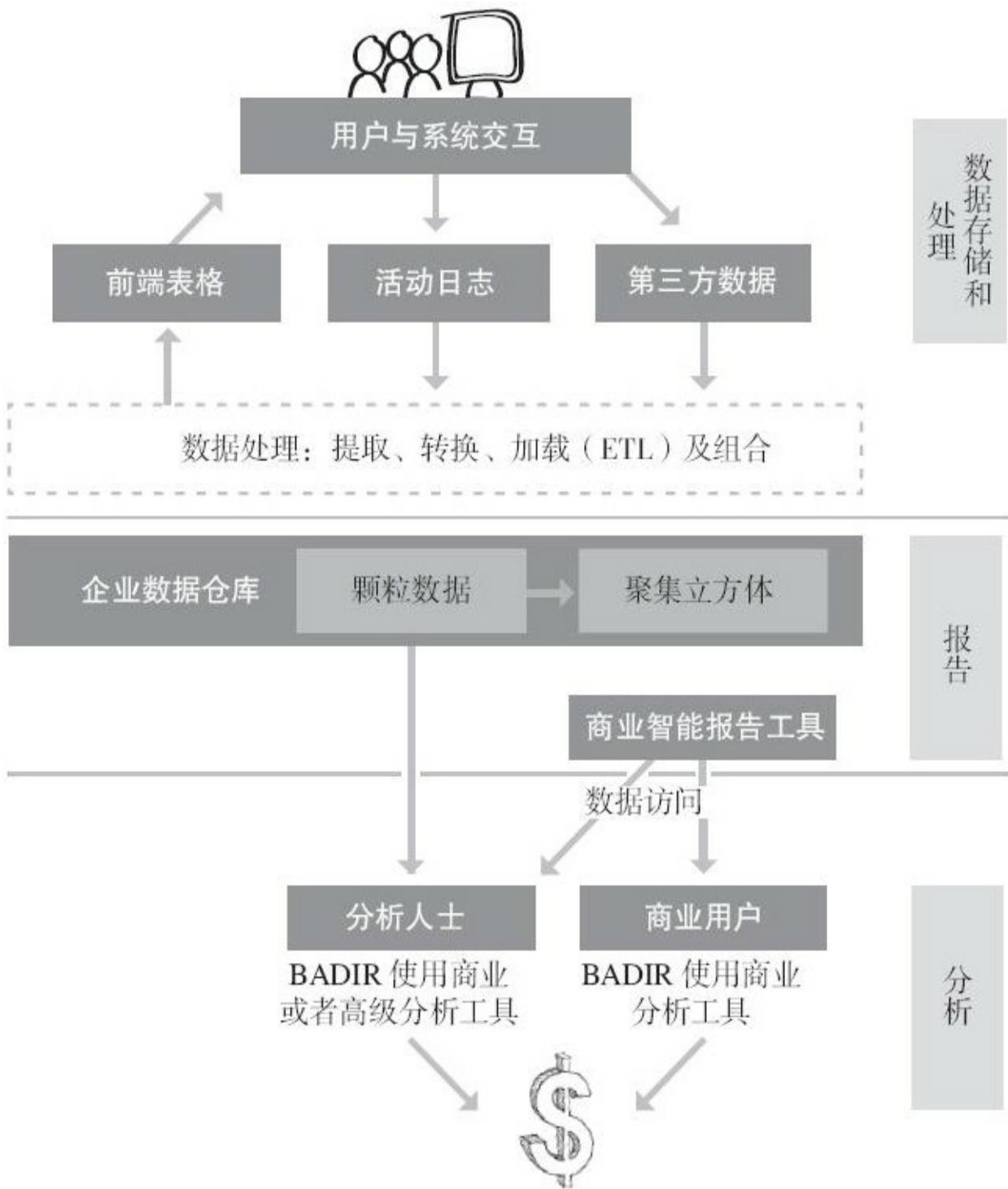


图6-1 一个组织中的典型数据流

表示层

如图6-1所示，当一位客户与企业业务交互时，就会产生数据，这就是

表示层。这些数据可以是各种类型的，如客户数据、业务交易、网络和移动活动数据等。下面用一个例子来予以说明。客户约翰走进银行，并按要求提供了所有的详细信息，希望能开一个支票账户。当银行员工将他的信息输入系统中时，数据便被存储在数据层中，最终通过一系列ETL（数据仓库技术）过程被写入数据仓库。事务处理系统接收数据并创建关于约翰的一行或多行数据。它分配给约翰一个识别号。当约翰走向出纳员并将钱存入他的账户时，数据层中的响应字段会被更新。之后，当约翰登录到他的在线账户检查余额时，将产生网络活动，他的记录再次被更新。如果约翰把钱转到某个人的账户中，交易信息会根据他的余额进行检查，然后转账将被执行，并且他的纪录会再次被更新。

逻辑层

现在让我们来看看后端发生了什么。每当约翰与系统交互时，前端就通过一个中间逻辑层与交易系统交互，服务于约翰。逻辑层存储所有需要执行命令的逻辑、数学、数学计算、分析决策结构和其他操作。这些操作的输出为约翰提供与产品相关的交互体验，包括访问账户信息或查看建议。它负责移动前端和后端之间的数据，并确保所有约翰的请求被正确处理。

数据层

这是数据操作的发生层。数据层存储数据，用于服务逻辑层。

数据处理

由前端所产生或被第三方系统接收的大部分数据是无结构、无组织的。它们需要被处理、整理、合并，并以符合逻辑的方式来支持数据库中的分析。原始数据到基础度量和业务实体的转换及相关性（数量大小、何时、何地、何人等）是业务规则和逻辑应用的关键步骤，以结构化的方式支持测量和报告。这些操作称为ETL（提取、转换和加载），ETL的作用是对数据库进行定期维护，及时更新约翰的所有活动。

数据仓库

在ETL完成后，结构化数据流到企业数据仓库（EDW）的各个表中。企业数据仓库将任意系统中的所有信息存储在一个地方，不影响交易数据，只需一个简单版本的真实信息，而不需要许多独立的文件和表格。

数据仓库也确保所有的历史数据是可用的。企业级数据仓库可能有关于特定信息类型的特定表格，比如包括人口统计信息的客户表、活动快照、风险和市场概要，或带有交易金额、交易数量和购买产品类型等信息的交易表格。

数据收集和存储系统应该有能力，并有效地处理大量的数据，可供分析师快速查看。许多组织使用关系型数据存储技术，如IBM DB2、微软SQL服务、Oracle和Teradata，来支持这些结构和传统的数据集。随着数据和数据存储的需求上升，新加入者已经引进了成本较低的解决方案技术来扩展规模，并且减少了延迟响应的时间。这种现象的表现之一是大数据。大数据作为一个包罗万象的容器，10多年前用于数据仓库和商业智能领域，装载着许多方法论和技术，负责处理大的、复杂的数据集。为便于讨论，我们将使用使大数据的百万兆字节的存储、处理和查询缩小到PB级字节的技术。传统的数据库（如Netezza和Teradata）已经通过分布式处理模式解决了问题。然而，虽然使用这些技术来缩放比例是可以实现的，但它需要大量的专有硬件和许可成本。新的计算平台（如Hadoop）开放源码，通过廉价的商品硬件进行缩放，并支持结构化（表格）和非结构化数据，是继大数据后的新的驱动力。有些公司（如雅虎和脸谱网）走在了前面，因为它们需要PB级别的数据集来强化自己的网站，所以推进了Hadoop的演变。这项技术作为一种替代传统的常规数据处理，现在已被除这些数字企业之外的公司所采用。

从分析的角度来看，数据存储和处理的最关键方面是它的整体信息架构。信息体系结构是关于信息如何流动、存储内容和存储地点的设计。在费尔南多和StarMatrix公司的案例中，信息重新架构是我们提倡的第一个行动，甚至应当被放在正确的数据存储和处理系统被设置到位之前。

商业智能（BI）或报告工具

商业智能（BI）或报告工具可以以静态和动态报告的形式快速、轻松地访问聚合数据。鉴于其可将数据进行集合且可视化地输出，使得商业用户和管理人员无须专业人员的帮助就能够自助获得服务数据。

通常，企业会错误地将这些工具连接到粒度非聚合数据，而令这些工具被错误地超负荷工作，使得商业智能工具的性能不稳定。这致使工具失去功能性，并且商业用户会发现，他们再次需要数据分析人员的帮助，并根据他们期望进行监视和管理的内容范围来获得相关数据。原始粒度

数据需要进行聚合，以便通过商业智能工具保持对数据的快速访问。聚合过程分为两个部分。

(1) 创建相关的数据子集：例如，按周和月而不是按天来选择分析过去3年的数据，而非全部历史数据，以此来减少数据量。

(2) 向上钻取至适当的粒度：粒度数据可以有成千上万个度量值。为了便于用户访问适当的信息，需要通过优先化和聚合将数据筛选至可管理的数量（100~300）。具体数量是用户试图回答的商业问题的函数，并且需要对其度量值进行优先化。首先是使用3个关键问题框架（第8章将会讨论）来确定一个部门或公司的前10~15个战略项目，接下来是使用BADIR来回答这些战略项目背后的商业问题，然后制定假设，获得对商业最重要的前100~300个度量值。

这个向上钻取的度量值子集通过存储，成为一个可以用商业智能工具进行连接、访问的聚合数据立方体，且该聚合数据的可视化效率远高于粒度数据。一旦后端商业智能数据立方体被正确设置，所有接受过这些工具培训的商业专业人员都可以构建自己的报告。

市场上有大量的报告工具，要做出明智的选择，可使用以下准则来选取一款能够满足大多数需求的工具。IBM、Oracle、Teradata、Microsoft和SAP都有与自己的特定应用程序相关的商业智能工具。市场上其他一些常见的工具有Tableau、Datameer、MicroStrategy和Qliktech。云和开源商业智能供应商（如Birst、GoodData和Jaspersoft）近年来已经成为企业商业智能技术的更便宜、更可定制的替代选择。在选择正确的报告工具时，以下几个方面是最关键的。

可视化

数据可视化已经从单纯的条形图和饼图发展成为一门科学，并且有许多专门的研究人员涉入其中。正确的可视化可以快速、直观地突出数据中的结论和异常。在分析人员进行数据发现或创建高级管理人员简报时，这是一个强大的工具。在评估可视化功能时，要评估该工具是否具有创造性的、丰富灵活的图形和可视化库，以及该工具是否具有任何推荐功能，可为正确的数据集呈现正确的图表。

通道

在当前这个流动型、地方型、社交驱动型的社会环境下，数据是否可以访问和共享变得非常关键。是否可以从移动工具、云、个人电脑（PC）、邮件、文本、通知、推特和其他社交媒体活动等全面的渠道和平台访问报告？

速度和易用性

商业智能工具是否成功在很大程度上是通过采用率来衡量的。报告是否能够快速生成并且易于访问？是否容易导航？是否可以轻松地以所需格式共享报告？它是否具有直观的图形用户界面（GUI），以便用户能够容易地操作报告，例如进行拖放操作？

与其他工具集成

虽然数据集成和数据组合通常属于数据处理的领域，但是快速访问不同数据集的能力意味着能更快地让你的数据创造价值，即使它是一次性的解决方案。这个工具能否无缝连接并整合来自关系数据库和大数据工具的数据？它是否能够将来自关系数据库多个工具（如销售事务）的数据与来自Salesforce.com的客户数据相结合？

数据和用户限制

如果你期望增长你的用户基础和数据量，并对你在3~5年内拥有的用户规模和数据规模进行一些估算和规划，那么商业智能解决方案的可扩展性应符合你的预期才行。你需要考虑，此工具对数据集和用户数量是否有任何限制？它的性能是否会随着数据量或复杂性的增加而下降？

成本

哪些成本与工具相关联，例如许可、实施、操作？

易于部署和维护

部署和维护成本可能会快速增加，并在几个月内超过前期软件成本。工具的实施是否便捷、快速、经济实惠？供应商和内部资源提供何种类型的支持？作为评估的一部分，商业智能工具如果具有较低的总体拥有成本，可能会促进用户选择该工具。

分析能力

大多数商业智能解决方案具有某种程度的分析功能，可帮助分析人员简化使用相同接口访问和分析数据的工作流程。此工具是否具有执行简单分析（如综合分析、相关分析和趋势分析）的功能？此工具是否具有简单的函数统计功能，如平均值和标准偏差？

信息设计能力

该工具是否具有数据建模功能，例如星形模式、维度、KPI等的设计？

商业分析工具

这些工具使用户能够应用简单的商业分析方法，直观地分析手头的数据。这些方法包括综合分析、相关分析和趋势分析，我们在第3章和第4章中已进行过讨论。任何有权访问数据的商业专业人员都应该能够使用一个良好的商业分析工具自行进行分析。他们可以找出隐藏的模式，解释波动，估量机会，并做出高水平的预测。

对于简单分析，我们选择的工具是Microsoft Excel。它是Microsoft Office包中的电子表格应用程序。它是迄今为止最广泛使用的商业分析工具，并且近年来已经实现了一些强大的改进和插件，使得能够用其进行更复杂的分析。Excel是商业用户处理非大型数据集或执行非高级分析（简单分析、报告、模拟、场景或建模）的理想选择。

Excel使用传统的表格结构，即行和列的数据。它可以轻松地导出到各种工具。它可以连接到数据库表（如Microsoft访问表），并具有可以将其连接到关系数据库和非关系数据库的插件。它通常用于综合分析、相关分析和趋势分析和估算，进行简单预测建模和时间序列预测。最新版本的Excel还增加了高级统计和数学功能。在Excel中可视化地表示分析结果是非常容易的，其具有静态和动态图形选项。它支持多平台操作，可以在PC、云和移动设备上进行访问和操作。



如果你想要通过不具备分析功能的报告工具访问数据，我们建议以CSV文件格式将数据集导出到Excel中，然后利用Excel作为分析工具。

Excel是一个易于掌握的工具，其流行性源于它具有非常直观和易于学习的图形用户界面。只要通过一两周的简单学习，你就可以使用Excel进行简单的商业分析和报告。但这个工具有一个限制，就是它只能处理最多100万行数据。然而，使用Power Pivot这个最新的扩展，就可以将可处理的数据扩展到1000万行。

除了Excel，上面提到的许多商业智能工具（包括Tableau、Datameer和MicroStrategy）都有良好的商业分析能力。我们下面讨论的许多高级分析工具还有对应的初级版本，便于用户快速获得可视化数据并快速进行商业分析。

高级分析工具

高级分析工具是分析工具的提升。在过去，许多工具是在学校和大学开发与使用的，然后被统计学家和分析师用于企业界。一些工具已经拥有成熟的市场并成为主流。

一些工具在小众产业、学术界、政府、研究机构和交易场所中被应用，其他的一些工具则被像IBM、SAP和Teradata这样的大型运营商收购，成为这些运营商垂直整合过程中的一部分，并随后发展成为价值数十亿美元的实体。

要选择正确的高级分析工具，关键应考虑的因素包括可视化能力、可处理的数据类型、成本、与其他工具的集成、数据和用户限制以及运营效率。此外，还应考虑商业智能工具是否具有以下两个特点。

- 易学性：该工具是否有图形用户界面？需要多少编码来执行分析？培训和培训资源的成本和可用性如何？

- 分析类型：除了简单的分析，它是否具有高级分析功能，是否可以进行预测分析、时间序列预测、细分（聚类）、生命周期分析和场景建模？它是否可以进行蒙特卡罗模拟？

市场上有许多高级分析工具，包括KNIME、R语言、SAS Enterprise Miner和Angoss Knowledge Studio。其中，KNIME和R语言均是自由的分析工具。KNIME是一种图形用户界面（GUI）工具，而R语言不是。

KNIME

对于那些精通先进统计概念的专业分析人员来说，KNIME是一个免费且易于使用的图形软件。虽然它不像其他软件那么稳定，但仍然非常实用，对于一些没有高级工具预算的小团队来说非常实用。

R语言

R语言也是许多数据挖掘者所选择的一个工具，但它有一定的使用难度，因为它不是图形化的工具。

SAS

SAS Enterprise Miner是许多财富100强公司所使用的世界级软件。它是一种图形工具，但定价较高，因此只有那些大中型企业才会选择该工具。SAS在JMP软件中有一个更便宜、轻便的选项，它也是可视化的，但是存在一些数据量的限制。

Angoss

Angoss Knowledge Studio是一款基于图形用户界面的工具，它能使决策变得轻而易举。因为它是可视的，所以分析结果易于理解。该工具相比较而言便宜一些，因此对于企业而言也是一个不错的选择。

总结

并不是每一个高级分析工具都是完美的。不同工具的成本和学习难度可能存在很大的差异。因此，需要根据你的特定需求和情况选择适合的工具。这些工具几乎完全由数据分析师专家和科学家使用，因为有效地使用这些工具需要一定的知识深度。商业用户最好与他们的分析专家进行沟通，共同解决那些必须通过高级分析进行处理的20%的问题。

小结

- 如果作为商业专业人士，你无法通过可视商业智能工具轻松访问数据，那么请与你的分析主管一起快速启用该功能。
- 如果你所掌握的数据不准确，请立即查看信息架构和数据工具。
- 一旦你可以访问数据，Microsoft Excel会是你最好的朋友，可以让你快速获得有助于解决80%商业问题的数据结论。

>>【看书累了吗？休息一会！更多新书朋友圈 每日免费分享微信xueba987。渺沧海一粟：为终身学习者赋能 2019年7月】

第三部分 领导工具包

第7章 分析和领导者

第8章 分析竞争

第9章 分析师领导者的攻略手册

第10章 付诸行动

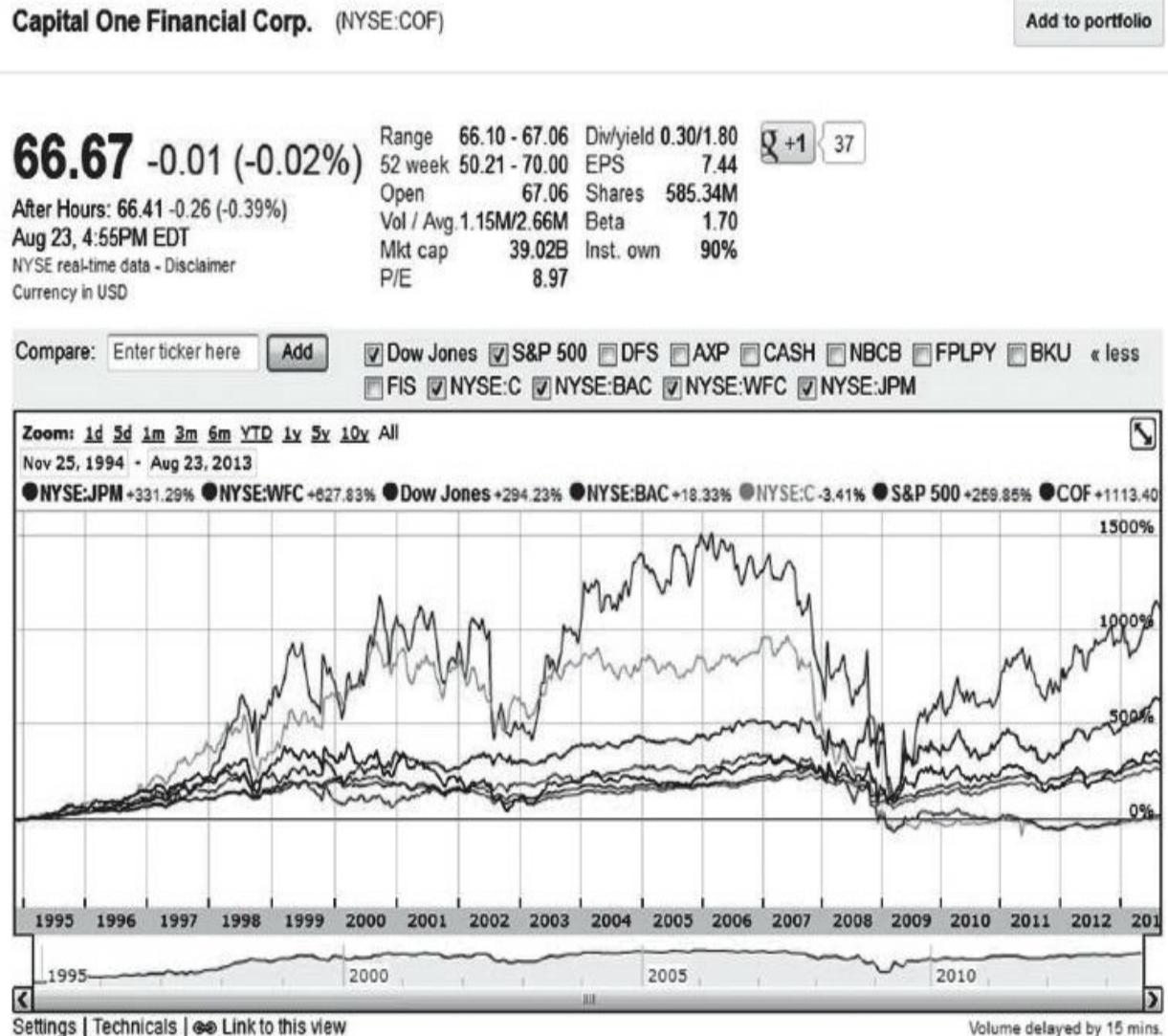
第11章 常见的陷阱

第7章 分析和领导者

本章将讨论：

- 领导者在建立数据赋能型企业和团队的过程中可发挥的作用。

- 领导者可用来评估企业的分析成熟度的四个关键点。



据称，Capital One的现任首席执行官理查德·费尔班克（Richard Fairbank）曾说过，他希望10个决策中有7个决策能按照计划实际执行。20世纪80年代，所有的金融服务公司都在发行通用型信用卡。它们向所有客户都收取19.4%的APR和20美元的年费。当时，费尔班克和Capital One的联合创始人奈杰尔·莫里斯（Nigel Morris）想要利用客户数据来创建客户群，并向不同的客户群体提供不同的信用卡优惠。因此，他们通过将信用卡首次提供给一个庞大的新客户群，征服了次贷市场，从而彻底改变了贷款行业。

这种数据赋能型创新对公司产生了深远影响。1996~2004年，Capital

One的年均收益率和每股收益增长率都在20%以上。2008年，当经济崩溃时，Capital One依然保持了增长，但它的复苏则是另一个故事。虽然其他多元化经营银行（如花旗银行、美国银行、摩根大通和富国银行）的股票表现挣扎着回到它们以前的荣耀，但Capital One的表现要优于所有银行，甚至强于道琼斯和标准普尔500指数的大盘指数。

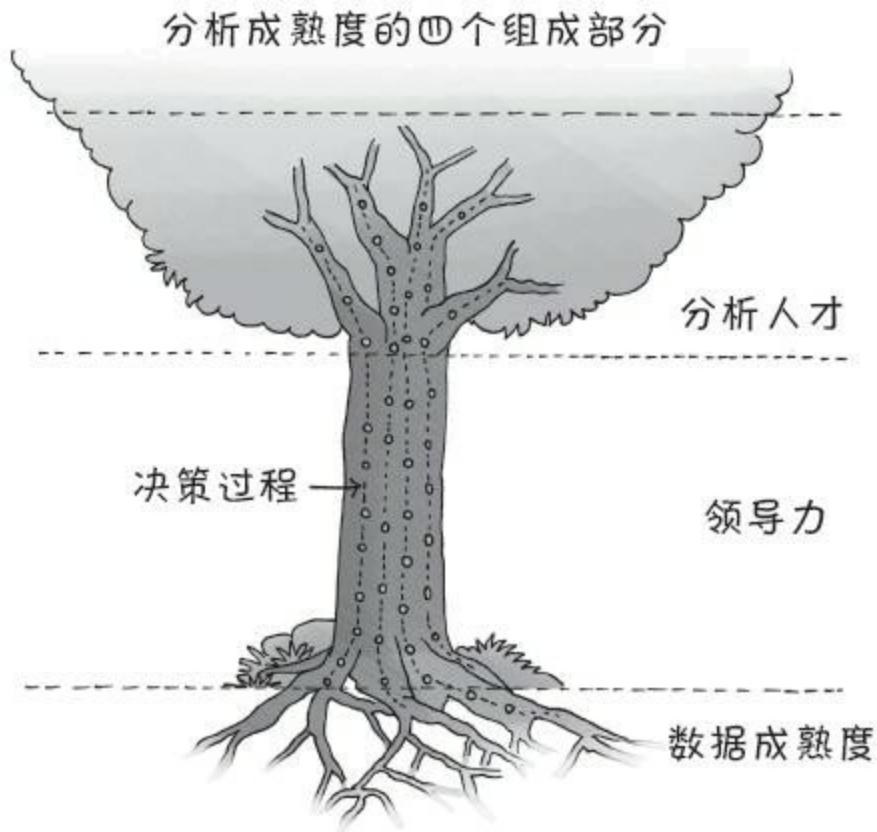
费尔班克是一个数据驱动型领导者。要成为一名数据驱动型领导，需要：

- 致力于了解他的客户和所有影响客户的动力，以推动增长和创新。
- 发自内心地认为数据将有助于推动卓越的商业决策和成果。

只有这样，作为公司的促进者和推动者，数据型领导者才能成为有证据为基础的决策者。请使用本节中的工具包，通过分析来助力你的领导力吧。首先，你需要评估和构建你所在企业的分析成熟度。

定义你所在企业的分析成熟度

要成功地借助分析的力量，你的企业需要增强这四个方面：领导力、分析人才、决策和数据成熟度。如果你的企业已经积累了大量数据，但是营销或产品团队无法利用它们来实现大家都能看到的增长，那应说明四个方面之一是有问题的。看看你的企业在这几个方面存在哪些问题，并努力纠正所有缺点。



第一个方面：领导力

从一开始，费尔班克和他的团队就以数据和分析作为支柱，建立了 Capital One，并且他们建立了公司文化和人才队伍。

麦肯锡（McKinsey）在2011年关于大数据的报告中指出：“分析大数据集……将成为竞争的关键基础，支撑生产力增长、创新和消费者盈余的新浪潮……每个领域的领导者，而不仅仅是数据导向型管理者，都必须应对大数据的影响。”数据分析不仅仅是一时流行的风尚，麦肯锡研究人员预计“在可预见的未来，数据会呈现指数式增长”。

为了使你的公司能够对所有这些数据进行分析，第一步应该对你的企业进行详细评估，包括与关键利益相关者进行访谈，以及进行一项全公司范围的调查，以确定数据分析实践经验的程度。这将确立该企业是准备成为数据驱动型还是数据赋能型，以及确定其弱点。作为领导者，你需要利用许多管理技能和管理模式来运营你的公司。在创建分析成熟度方面，领导力的最大差距是愿景、承诺和问责。

在你的企业里，是否有一个每个人都知悉并理解的愿景或使命？这种愿

景或使命是引入分析之前所需的第一个也是最重要的行动。

此外，数据装备型领导必须在进行关键决策时，坚持以数据作为决策依据。他们不仅要依靠数据来证明或反驳他们自己关于商业机会的假设，而且还应从数据中学习以发现新的机会并识别差距。如果你的企业还处于分析成熟阶段的早期阶段，那么可在构建企业愿景时，将企业的开发分析能力纳入其中，用来指导全公司的投资决策。

当然，你只能管理你可以计量的事物。为了让领导团队利用数据驱动决策的力量去推动增长，你的团队需对他们做出的决策负责。无论你是使用零基预算、平衡记分卡还是其他方式来跟踪领导团队的决策和绩效，都需要在企业内的所有级别中实现可量化的问责制。

可能还有其他方法可用来优化决策，比如运气、胆量、宗教、魔法，但这些都不能提供一致的、可重复的结果。基于可计量数据和文档分析，可以帮助你做出更明智的决策，从而实现一致的、可重复的成功，你可以在这些成功的基础上继续学习并利用其成长。费尔班克所说的10次里面有7次就是这样做的！

第二个方面：分析人才

Capital One有一种独特的方式，使它能够在初级和中级管理层获得分析人才，让具有强大分析能力的人员参与大多数涉及高影响力决策的工作职能。该过程包括书面定量筛选试验、两轮案例面试和两轮行为面试。定量测试非常困难，即使一些GMAT分数很高的候选人也未必能够通过该测试。对于案例面试，每个面试官都接受过一些正式的培训，因此结果将是可比较的。在汇报会上，所有面试官都分享他们对参与面试者的评分，并详细讨论各个评分之间是否存在很大差异。最后，他们将员工的工作绩效，即他们的年度绩效评分与面试过程中的评分相关联。他们发现，得分最高的候选人事实上确实比其他人在工作上表现更好。

领导阶层设定目标，但不能直接提供结果。这需要分析人才，且这些分析人才不一定是分析师。最好的第一步是引入一个高级分析师领导，让这个人有影响力，并能在企业的职能和战略研讨中发表意见。

麦肯锡大数据还预测，到2018年，企业需要收集待处理的数据量将会增多，因此会产生140000~190000个数据分析人员和高达150万个分析商业经理的人才缺口。事实上，人才短缺问题已显而易见。这既是挑战也

是机遇。随着更多的商业专业人士具备数据-决策的技能，且分析能力又成为企业对其人员的核心需求，这一缺口是可以消除的。许多专业分析机构正在通过培训和招聘来填补这一缺口。



2011年，埃森哲研究发现，美国工人在提高技能的压力下，迅速增加了技术技能，但在提升解决问题、分析或管理技能方面进展缓慢。因此，雇主应该投资于增强商业专业人士的数据决策技能，这将消除三个技能缺口中的另外两个，即解决问题和分析技能。

商业方面的分析人才

无论你是每天从事分析工作的商业专业人员，还是企业主或领导者，你都需要使用数据来查找可为你和你的企业带来更好结果的各类信息。无论是为大型商铺进行季节性库存规划还是为零售店配备员工，了解数据趋势（例如过往销售量或每日客流量）都有助于优化库存和资源。

参与决策的商业专业人员需要通过某种工具（如Tableau、Business Object或Micro策略）轻松访问数据。此外，他们还需要发展三种重要的

能力：

- 实践商业分析和测试：商业专业人员需要对一个结合数据科学和决策科学的数据-决策框架进行理解并亲身实践。对于数据科学，他们应了解基本的商业分析方法，能够在Excel中执行操作，并且对A/B测试有基本的了解。这将有助于解决他们80%的商业问题。
- 与分析师有效合作：他们需要与数据科学家和分析师有效地合作，以取得他们（对于80%的可解决问题）对实际工作的支持，并与他们合作解决需要高级分析的复杂问题。
- 引导性高级分析：他们需要对预测分析（包括回归、决策树和细分分析）进行全面的了解，以便在需要时能有效地与分析师交流。

数据方面的分析人才

分析专业人士还需要数据和决策科学技能来分析数据并开发人际交往和商业技能，以弥合数据与商业之间的差距。大多数企业发现，招聘受过商业分析培训的人很不容易，因为正式的分析教育直到最近才出现。但是，你可以对那些表现出分析能力的内部人士进行培养。通过投资分析培训和聘用高级分析专业人员来领导或扩大团队，可以将企业转变得更具分析成熟度。

数据科学家和分析师团队需要结合以下四种技能：

- 实践商业分析和测试：分析师还需要具备对于包含基本商业分析方法的数据-决策框架的实践经验。他们特别需要强大的决策科学技能，通过影响力和沟通来建立同盟，从而显著影响商业。如果企业涉及测试，他们需要有坚实的测试和学习基础，也称为A/B测试。
- SQL技能：数据分析师必须熟练地从多个来源提取和整理数据。编写SQL查询以及掌握Teradata和Oracle等工具的经验非常重要。对大数据工具（如Hadoop）拥有一些了解也将起到帮助作用。
- 实践高级分析：分析师需要具有先进分析技术的实践经验，例如预测分析，包括回归和决策树、时间序列，以及根据商业需要，对所获得的文本进行分析的实践经验。

- 统计工具：要执行高级统计分析，分析师需要具有一个或多个统计工具（例如SAS、R语言、SPSS或Knime）的使用经验。

分析小组成员应具备上述各类技能的组合。例如，数据分析师在第二项技能方面很强大，但是对第一项技能仅有所了解。商业分析专家在第一项技能方面很强大，而对第二项和第三项技能可有较强的理解力。预测分析专家在第三项和第四项技能方面很强大，而对第一项技能仅有良好的掌握。很少有专业人员在所有四个领域都很强大，所以你需要经常将团队的人员与这些技能相匹配。

第三个方面：决策

Capital One的决策过程非常卓越，这对像它这类数据驱动型企业是非常重要的。公司有一个严格的计划过程来评估每个新的想法。产品或市场营销经理提出的想法必须建立在分析过去的客户数据，估计提议项目的绩效，并且映射该项目可能遇到的每种境况和风险，以确保公司免受经济损失的基础上。一旦一个想法获得批准，就会在完全推出之前进行试用或测试。除了前瞻性的规划过程，Capital One还有一个系统性审查过程。在项目启动并且开始产生结果之后，领导者必须监视整个过程以评估该项目是否按照预期运行。如果出现任何异常，他们将需要在运行过程中改变或停止项目。所有这些决定不仅需要相应的高级管理层的审查，还要经过在不同群体的正式审查过程。

在诸如战略、营销、产品、运营、设计和创新的商业功能中有许多机会，使用数据分析可以做出更好的决策。例如，公司的营销预算分配是一个良好的决策机会，因为在这个过程中有很多方法来划分预算，并存在许多目标群体和子群体。你是想要增加客户的终生价值还是希望获得新客户？大多数公司看到很多可能性，但没有足够的预算来实现这些可能性，而这正是数据驱动型决策的价值所在。

例如，CEO不能在企业的所有职能和层面上做出个人决策。公司内部的有效决策结构将授权商业负责人以及广大成员，通过了解其所处的商业环境并清楚知悉公司愿景来做出决策。

解决方案是建立一个让企业中每个人都理解的透明决策过程。在战略层面上，这一过程应清楚地表明：

- 获得资助的项目类型。

- 项目选择的评估标准。这可以包括财务和非财务指标，如客户满意度、创新、可持续性和社区影响。

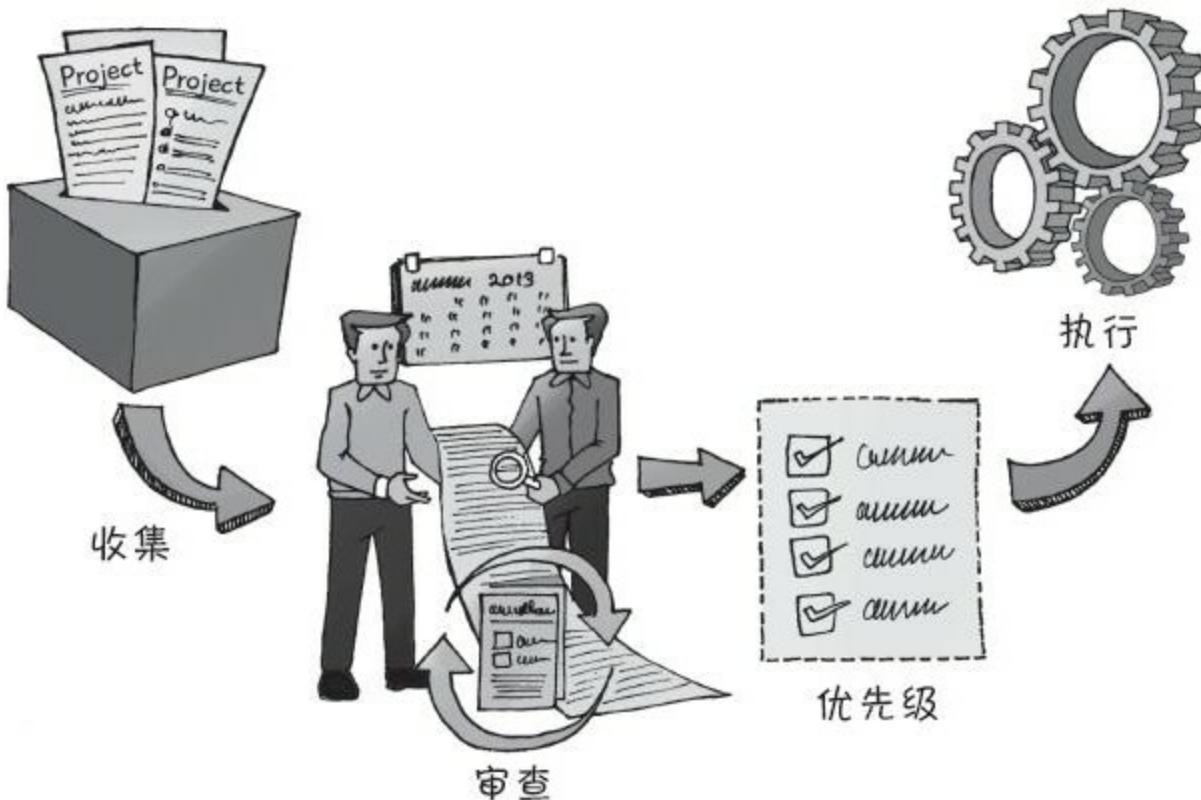
决策过程需要：

(1) 收集所有提议方案，以及关于其预期投资和回报的信息。在这个过程中，可以采用战略规划文件，在这类战略规划文件中，关键参与者提出他们的想法，然后再对这些规划文件进行审查。像PayPal这样的公司每年都会进行这样的决策活动，小公司则可以每月或每季度进行。所有在要求日期内提交的项目都会被考虑。

(2) 一个由利益相关者组成的团队将定期审查新项目，例如进行每月或每季度的正式规划和审查。在审查会议上，利益相关者会审阅最近一组决策所产生的结果，对这些结果进行评估并从中学习，以利于对新的项目做出决策。通常，审查和规划发生在同一个周期中。虽然我们建议每月或每季度进行一轮审查和规划，但有些创业公司实则每天都这样做。无论审查和规划的周期多长，你都要确保在进行下一个计划周期前有足够的时间吸收从上一组决策中所学到的经验。

(3) 一个项目优先于另一个项目的标准。3个关键问题框架（将在第8章中讨论）是确定关键绩效指标（KPI）及其驱动因素的有用工具。在确定企业应该投资的高层项目及决策标准时，这些问题的答案迟早会有用处。

(4) 清楚地描述项目将如何执行，包括过程、标准和指标。企业中的每个决策者都应该了解新项目的实施计划、决策标准、每个项目预计推进的指标，以及该指标与企业的首要优先事项和关键绩效指标的关系。



第四个方面：数据成熟度

Capital One的另一个突出之处在于它投资建立了一个坚实的数据成熟度基础。仅有少数企业能保证它们的大多数数据在到达数据仓库时仍是清楚的，而Capital One便是这样的企业之一。事实上，因为数据非常清楚和可靠，当趋势发生突变时，Capital One不需要再次去验证数据。相反，趋势的突变意味着肯定发生了某件事。21世纪初，Capital One升级了系统，将数百个独立系统及企业数据仓库简化并集成到一个统一的数据仓库中。公司领导人从所有工作职能中挑选了500名员工，作为这个为期18个月的项目的组成人员，确保让表现最好的人员和最有经验的人员去参与这一关键战略计划。

我们的许多例子证明，对数据分析的投资是值得的。然而，为了获得回报，数据团队必须能够收集和存储适当的数据，并确保分析师和商业用户能够无缝访问这些数据。然后，企业可以生成那种带来制胜决策和成功策略的数据指引。这是分析回报的来源，也就是我们所指的投资回报率。

分析应以良好数据作为坚实的基础。良好数据是指及时收集并安全地存

储在一个既保护数据的完整性，又确保已就绪且易于访问的准确的数据。通过适当的数据收集和存储过程，投资于基础设施的企业已经开始将利用数据作为资产。不良数据是指不完整、错误、不一致和矛盾的数据，这些数据会严重影响企业通过数据了解客户和产品的能力，也会严重影响决策者需要的数据可信度。

数据成熟度包括四个主要方面：基础设施、访问、可用性和仪器化过程。

数据成熟度首先依赖于灵活的数据基础设施，这些基础设施可以适当地进行扩展，并支持企业所需的访问类型。此外，当执行分析并且需要先前不可用的数据或度量时，敏捷的基础设施可以快速地实现对新数据字段和源的捕获。这个过程称为仪器化过程。新的数据源或字段然后可以被送到现有的基础牢固的数据基础设施中。



分析师和商业专业人士是具有独特且不同的数据访问需求的用户。因此，他们依赖于基础设施的可用性。商业专业人士应通过商业智能工具（如Tableau或Business Object）轻松访问受限的综合数据。数据分析人员需要直接访问企业数据仓库（EDW），其中存储了详细的基础用户和事件级别数据。此访问可以通过SQL查询或类似的协议来实现。数据基础设施需要支持这两种数据访问。

有效的数据基础设施应具有如下品质：

- 其设计和结构开放、灵活、安全。
- 可扩展处理更多的数据和用户。

- 可扩展处理更多和不同类型的数据。
- 可处理复杂查询和大量数据。
- 可以与大量且不断变化的系统、技术和工具进行交互。
- 授权用户（无论是数据分析师还是商业用户）都可轻松安全访问。

令人惊讶的是，开发有效的数据基础设施最困难的方面并不是硬件或技术的选择，而是你的信息系统的工作设计和架构。如果你的企业中没有信息架构人才，请聘请外部顾问来设计能让你的内部团队可以理解的信息流。但是请注意，不需要过度投资收集所有可能的数据。如果你正确解决了这3个关键问题，那么你和你的领导团队将会确定你们在商业理解上的差距，突出所需的数据类型，并且请记住，随着你的商业发展，所需的数据也将增加。因此，仪器化过程是数据成熟度必不可少的一部分。

Capital One是建立分析驱动型企业的典型代表。好消息是，你的企业也可以通过某种方法成为这样一个分析驱动型的企业，并且可以在你的时间和预算限制内实现。最重要的一点是你要希望、计划并努力使其成为这样一个企业。第8章讨论如何引导你的企业在市场上进行分析竞争。

小结

- 为了最好地利用分析，首先要评估你所在企业的分析成熟度。
- 用于评估分析成熟度的四个方面是领导力、分析人才、决策和数据成熟度。

第8章 分析竞争

本章将讨论：

- 为企业制定分析议程的3个关键问题。
- 数据赋能型公司的最佳企业结构。

一名哲学教授站在学生面前。上课伊始，他默默地拿起一个非常大的空

的广口瓶，然后将一些大石块装入瓶中。他问学生瓶子是否满了，学生说满了。然后他又拿起一盒小石子倒入瓶子里，轻轻地摇了摇瓶子。随后，这些小石子进入大石块之间的空隙。他再次问学生瓶子是否满了，学生又说满了。接下来，教授又拿起一盒沙子倒入瓶子里，沙子填满了石子间的缝隙。学生一致认为现在这个瓶子已经满了，但教授继续用啤酒填充瓶子。



虽然已众所周知，但这个故事值得反复讲述。教授将生活比喻为一个广口瓶，大石块代表你生活中的大事物，例如家庭和健康。如果你失去了其他小事物，你的生活中还有家庭和健康这些重要的大事物。所以，你要先照顾大石块。我们将这个比喻衍生到一家企业。大石块是驱动企业发展的3~5个任务，例如维持客户满意度的特定百分比、资产回报、员工留任，以及增进利益相关者的价值。大石块不应该经常改变，因为它们代表了公司的目标。小石子是实现大石块的重要举措，可能每年都略有变化。沙子是企业实现和维护大石块的一切别的事物。重要的是，要确保企业内的举措和行动不会损害这些大石块。那么啤酒代表什么？它代表了成功后的庆祝。

一个精心制定的分析议程有助于通过管理和重新校准企业驱动因素，将企业的目标与关键绩效指标（KPI）结合起来。让我们谈谈如何利用“3个关键问题”框架创建一个最佳分析议程，然后再来看看你的企业结构可以如何来支持这个议程。

I：“3个关键问题”框架

在进行分析成熟度评估后，你确定你的企业已准备好成为数据赋能型企业。现在的重要问题是：“你应该从哪里开始，你应该如何处理这些数据？”显然，你需要一个分析议程。本章将告诉你为实现分析目标，你预期将需完成的事项，我们将其视为整个企业的高级分析计划。最重要的是，如果要让这项计划有效的话，那么就一定应当包含在企业跨职能计划当中。以下是这3个关键问题（见图8-1）。

首先，需要对你的数据提出以下3个关键问题，然后再根据问题答案去运行你的业务：

- (1) 我现在做得怎么样？
- (2) 我的业务驱动因素是什么？
- (3) 我的客户是谁？他们的需求是什么？



图8-1 3个关键问题法

接下来，让你的整个企业都充分了解并构建理解的三大支柱，即

- (1) 用一个测量框架来回答：我现在做得怎么样？

(2) 用投资组合分析来回答：业务驱动因素是什么？

(3) 用客户分析来回答：我的客户是谁？他们的需求是什么？

接下来让我们依次看看这三大支柱因素。

第一个支柱因素：测量框架

我现在做得怎么样？

通常的回答是财务方面的，例如收入、交易利润率、投资回报率等。但是，我们还可以用其他方式来回答这个问题。例如，你可以使用平衡计分卡，它能将财务信息与非财务指标（如客户满意度和股东价值）相结合。无论在哪种情况下，你都希望定义主要指标或KPI，然后确定二级和三级影响因素。这些可以是企业内部单位的个别指标，但应与总体商业目标和优先级相一致。

测量框架建立全面和可重复的步骤，用于评估公司是否达到了你自己设定的目标。一旦创建了测量框架，你就可以选择最有可能推动公司KPI的正确项目（分析或其他）。

关键是要将指标分解到可以受项目影响的水平。如图8-2所示，交易利润率可以作为KPI。所有其他指标，例如收入、流失的客户数量等，都是影响因素。在这里，我们从数据中得出的结论是：减少流失的客户数量可以增加收入和利润。一旦确定，你可以调整内部项目，比如进行客户流失分析，以处理这些影响因素。

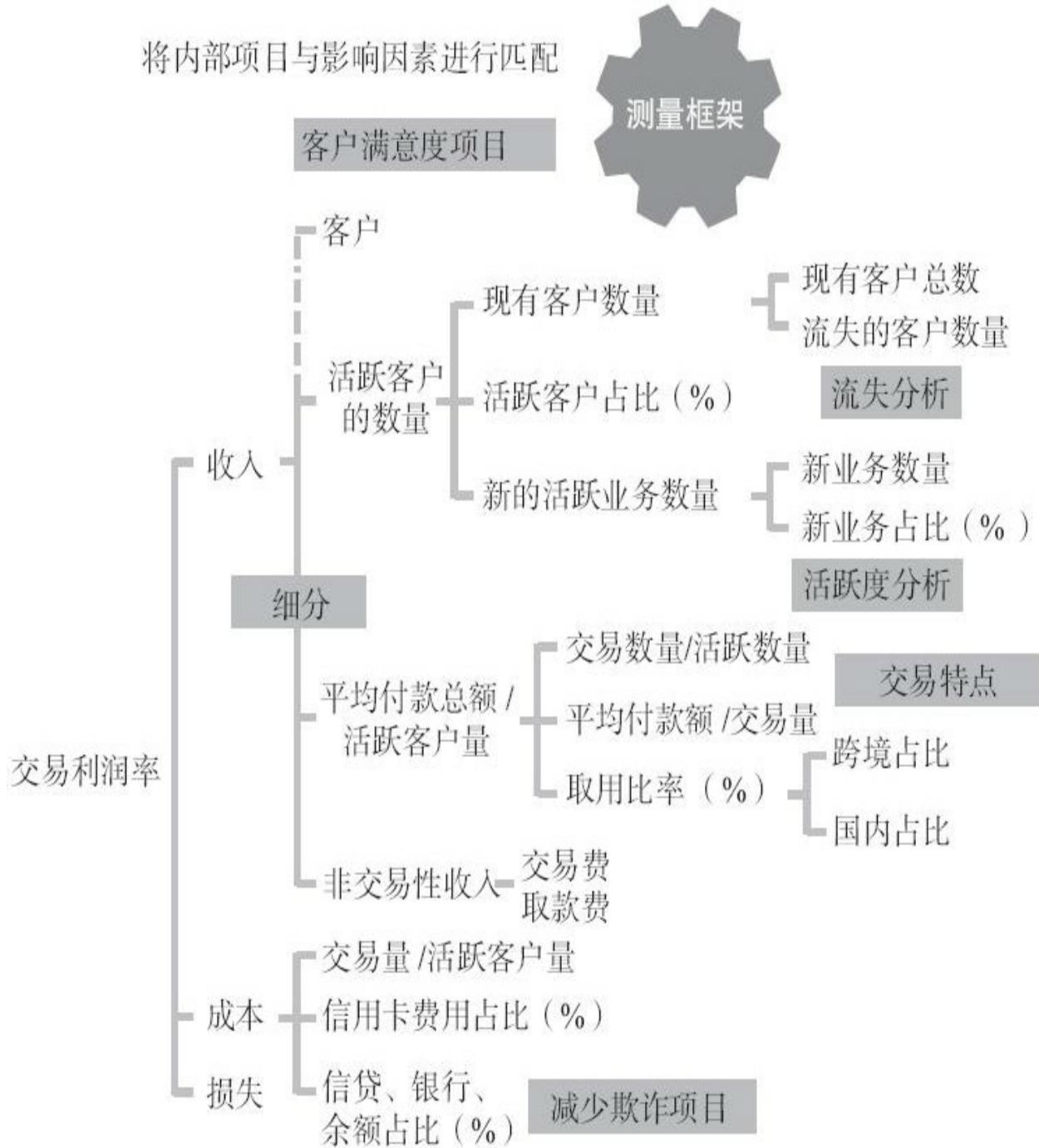


图8-2 测量框架

测量框架还能为影响因素铺设假设，并利用分析来确认框架内的正确影响因素，消除那些对KPI无甚助益的影响因素。但是，我们发现，其中一些杠杆因素是难以施加影响的。在图8-2中，企业不能轻易地改变诸如“平均交易价值”这样的杠杆因素。这是一个杠杆不可调节的例子。你

还必须通过利益相关者会议来识别可调节和不可调节的杠杆，以便你可以将项目与可调节的杠杆进行匹配。在图8-2中，你的商业关注重点应当是可调节的杠杆，例如流失的客户数量，而不是不可调节的杠杆，例如平均付款值（APV），即所购商品或服务的价值。

第二个支柱因素：组合分析

业务驱动因素是什么？

这个问题能让你理解你的商业动力。你的各个不同产品的绩效如何？根据它们之间的相对绩效，你能发现哪些新的机会？

现在你已经从测量框架中确认了正确的影响因素，你可以观察这些影响因素在不同的产品或服务分部的表现，开始了解你的投资组合的动力。一般来说，首先要查看整体利润报表，了解你的产品、服务、商业单位或其他业务部门的相对业绩，然后通过创建一系列 2×2 或 $2\times 2\times 2$ 矩阵来进行比较，从而以不同的计量指标反映公司的组合情况。

在图8-3中，已用图示标出了增长、利润和相对规模情况。当你充分了解你的产品与市场地位之间的关系，并将这些信息与多种商业指标（如增长、盈利能力、规模、收入、渗透、电子钱包份额、竞争优势和客户价值）相结合时，就可以确定你最大的增长潜力在哪里以及最大的危险障碍潜伏于何处。

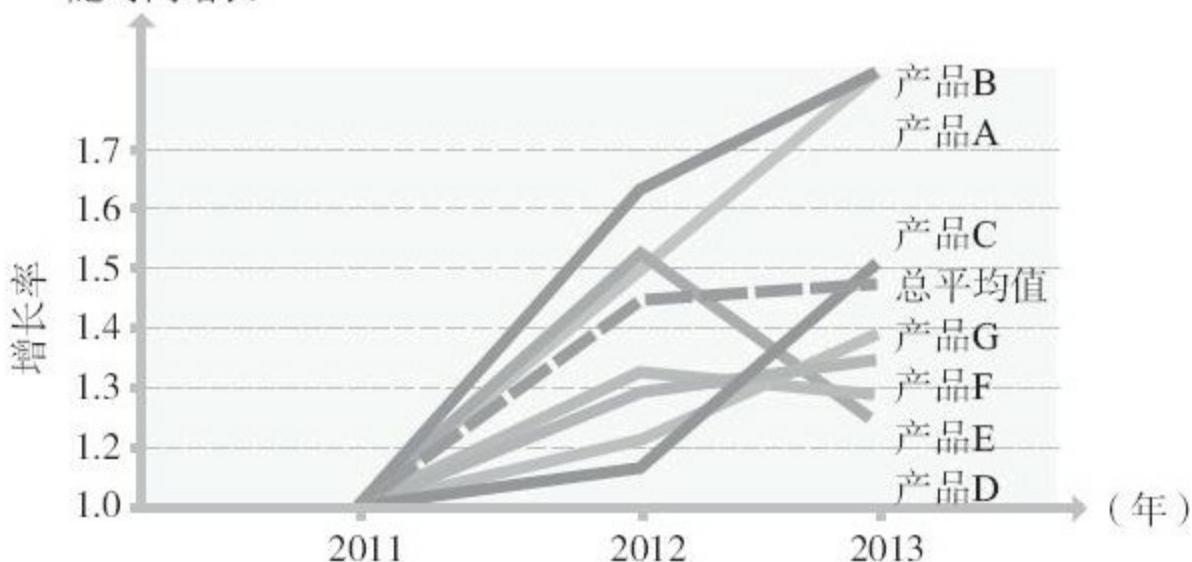
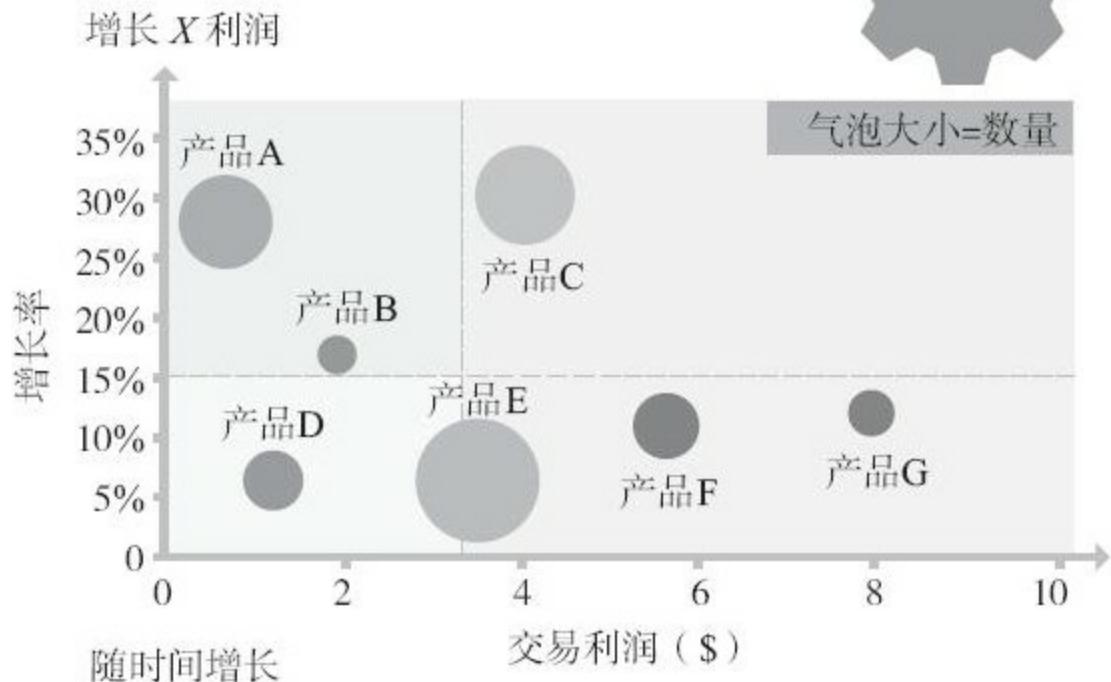


图8-3 组合分析：确定重点

一旦你对产品组合有了充分的了解，就可以利用数据来分析你的风险和机会，并推动那些对公司有所助益的项目。例如，通过查看图8-3中对增长和利润有推动力的产品，你可能会发现某种产品（产品C）具有高

增长和高利润。因此在确定你的黄金产品或明星产品后，你可以构建一个交叉销售预测模型来识别那些有可能购买黄金产品的重要客户。这是一个如何通过分析使项目对业务产生推动影响的例子。这种分析矩阵还可以提供进一步的论点，比如想一想为什么产品G是一个具有高利润但低增长的产品。只要理解了为什么它的增长会低于平均水平，就能帮助你最大限度地改进产品。

图8-4显示了如何根据组合分析的结论去认识项目。



产品	成交量				交易盈亏 (bp)			收益和成本影响因素 (%)			
	Txns (MM)	总 Txn 占比	复合年增长率	金额	收入 (MM)	利润 (bp)	成本 (bp)	损失 (bp)	境外	CC	银行
A	136	37%	22%	\$13 754	\$479	122	-19	-28	24%	48%	30%
B	107	29%	11%	\$12 741	\$356	231	-78	-43	16%	66%	30%
XY	97	27%	15%	\$10 756	\$290	152	-127	-40	46%	96%	0
Z	23	6%	34%	\$8 756	\$210	180	-111	-25	26%	56%	14%
总计	363	100%	29%	\$34 007	\$1 035	168	-72	-32	34%	79%	15%

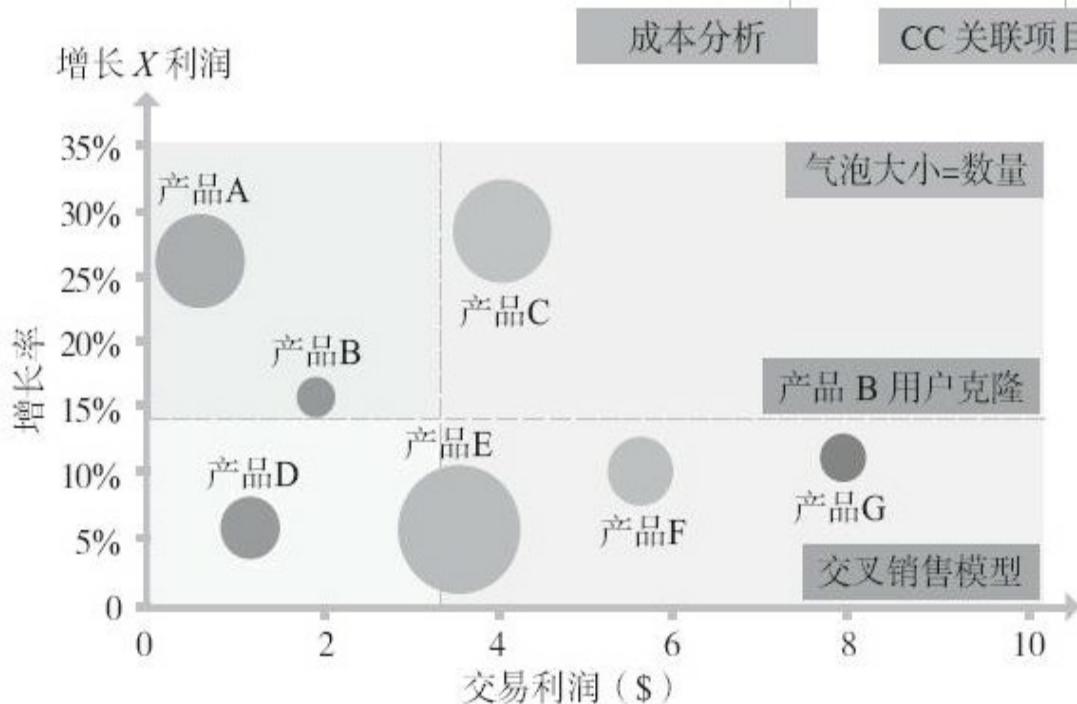


图8-4 组合分析：将内部项目与产品组合进行匹配

第三个支柱因素：客户分析

我的客户是谁？他们的需求是什么？

客户的行为是怎样的？他们如何使用我的产品？他们需要什么？这些是你在进行客户分析时的关键问题。客户的特点跨几个类别，包括：

- **人口统计特征：**你的客户分布特征是什么？有多少是女性？平均年龄是多少？已婚已育的占多大比例？具体的人口特征统计指标将取决于你的特定产品和服务。例如，在Curves有一家主要面向女性的减肥中心，那么对这家店来说，性别就不是一个有用的人口统计指标，但年龄和购买习惯可能会是有用的统计指标。
- **需求：**确定客户的明确和隐含需求。他们是否按产品预期功能对你的产品进行了充分利用？他们对产品是否有特定的功能需求？你的客户是否需要安装或操作支持？如果需要的话，需要如何支持？了解客户需求的其他方法是市场调研。调研时可采用一些定量工具，例如NPS（净推荐值）调查、网站上的产品调查、客户反馈表、客户服务电话，以及一些定性工具，例如客户服务输入、焦点组和迭代闭群反馈等。
- **行为：**了解客户的行为。如果你是一家水疗中心的老板，了解客户的行为有助于你了解一次性客户与经常性客户之间的区别，以及他们不同的行为方式。除了水疗服务之外，经常性的客户是否还喜欢其他产品？

通过查看相关客户数据，你将获得关于客户的新观点。很快你会拥有大量不同属性的客户。不同客户属性所构成的大矩阵意味着有很多客户信息可供提炼，但由于矩阵很大，这些客户属性往往混合在一起。你可以使用简单的细分，例如RFM（近期、频次、货币金额）或高级细分，例如聚类技术，将你的客户群分成5~7个最有意义的细分群体。在图8-5中，我们将客户群分为7个细分群体。

一旦将客户按有意义的属性进行细分之后，可将客户群体绘制成有意义的2×2矩阵，用以指导公司的客户吸引策略。图8-6中显示了如何按不同的客户群体去确定项目从而吸引客户。例如，你可以针对公司的获利情况和成长阶段，并根据客户类型（国际客户、奖励爱好者或是技术专家）制定定价策略。

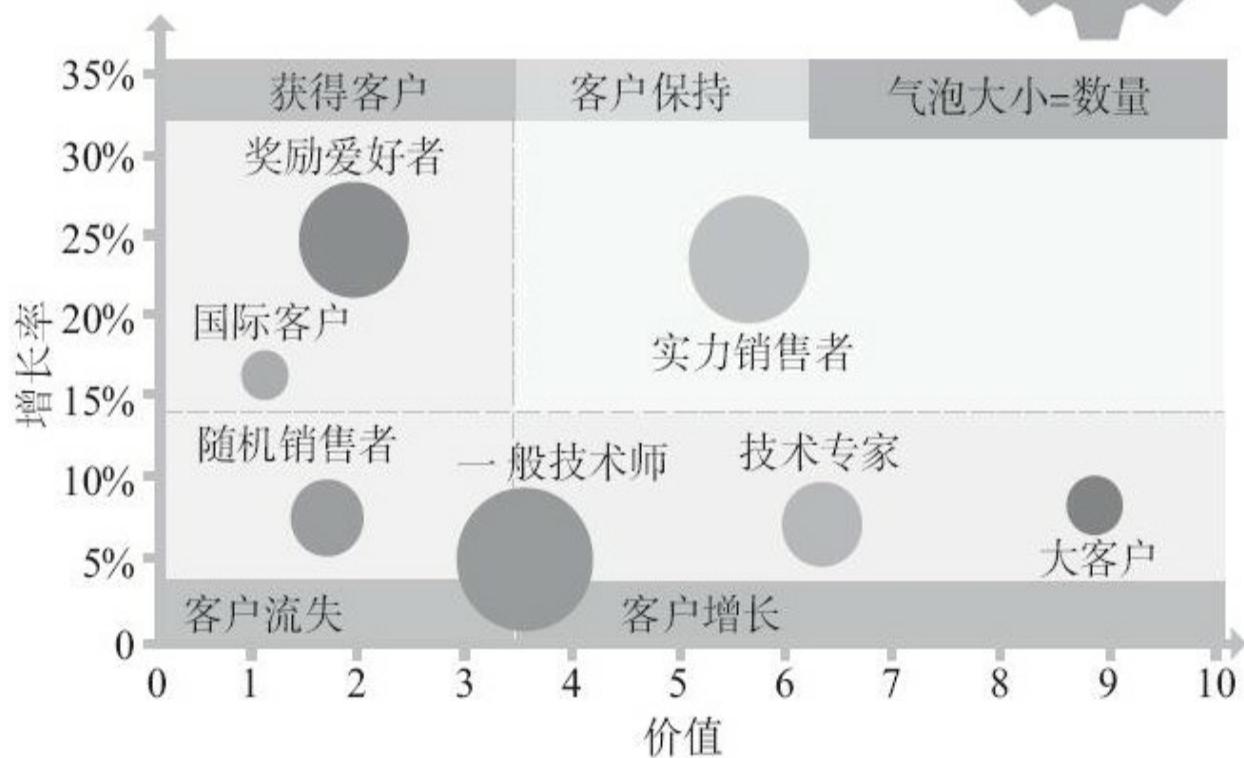


图8-5 客户分析：确定客户需求

客户细分		客户不变	获得客户	客户增长	客户流失		
		实力销售型	国际客户	奖励爱好者	技术专家	大客户	随机销售者
战略	客户细分						
保留战略	X			X	X		
吸引	X	X	X	X	X		
定价		X	X	X			
交叉销售	X	X	X	X	X		
经营效率		X	X			X	X
信用	X		X	X	X		
.....							

客户欺诈分析

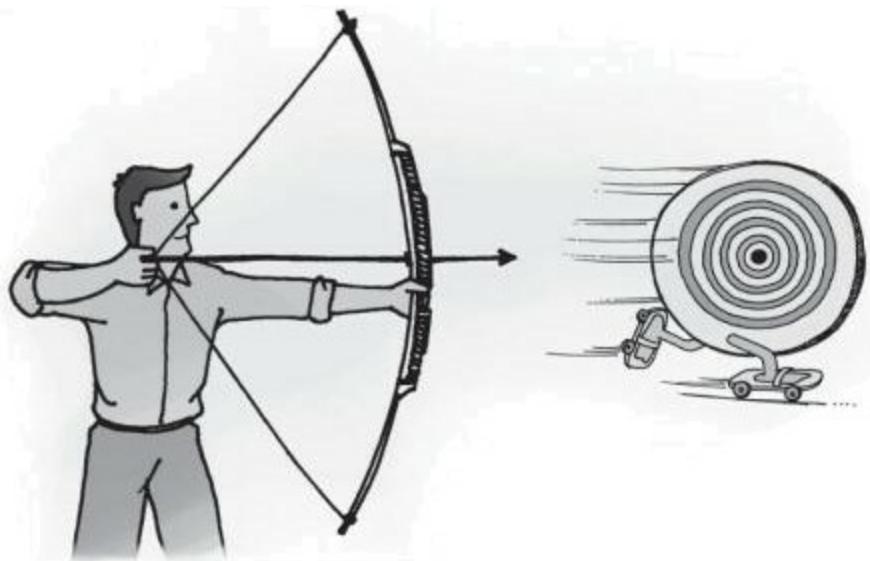
图8-6 吸引客户

简而言之，基于三大支柱因素的新观点将有助于你重新确定你的商业影响因素。这些都是可以帮助最大化投资回报率（ROI）和推动商业发展的杠杆工具。这种分析在这两个方面都发挥了重要作用。了解你所希望的产品或服务的增长方式，将有助于确定你与客户的互动方式。了解与客户的互动方式，将决定你未来可提供的服务类型。你的测量框架最终将这些元素联系在一起，并让你了解如何制定公司战略以及如何安排优先事项。

虽然企业中的高管已经了解了他们最重要的KPI，但是随着这些KPI在企业内部从上至下的分解，人们对它的理解反而被稀释了。这适得其反，因为KPI是对影响因素的凝练，最终揭示了是哪些因素在影响KPI。为了有效地利用数据以做出更好的决策，所有员工（无论是哪个商业部门的），都需要接受培训，这有利于他们在日常工作流程中提出正确的数据问题，在他们的职责范围内做出正确的决策。这最终会令员工拥有推动KPI和ROI所需的突破性观察力。

请注意，回答“3个关键问题”将使你能够深入了解不断变化的商业环

境，从而产生新的观点，但这不可能一蹴而就。在年收入为1亿~2亿美元的公司中，可能需要4~6周，而在年收入为10亿美元的公司中则需要3~4个月的时间。你所获得的理解不是静态的，它可以随着商业增长或变化、竞争带来的新趋势，或是行业前景和技术的变化而发生改变。例如，人们对智能手机的热衷改变了社交媒体行为。这“3个关键问题”从本质上来说，可帮助你随时知悉企业在商业过程中，哪些杠杆因素发生了变化，并可以以此继续推进那些有所助益的杠杆因素。



II：企业结构

分析在企业中起到了极具价值的作用。它支持旨在加强企业KPI的关键决策，并影响企业的资产负债表。最终，分析可以对公司的成功发挥关键作用，特别是当企业能够充分地借助分析工具的力量时。

团队结构

根据企业的性质、规模、市场、竞争力、文化，以及许多其他因素的不同，每个企业都有其独特的组织结构。然而，在理想的世界中，我们建议应嵌入分析结构。

在嵌入式系统中，分析团队是商业功能的一部分，它每天都提供专门的服务，同时直接向分析主管报告。团队成员对他们的项目负责人负责。设置实体可能是实现这种整合效果的最佳方式。

许多企业现在将其分析团队称为卓越分析中心。这个实际称谓当然可以

随着企业的文化和风格的变化而有所不同。这里的要点是，在每个利润中心，需要有一个明确的分析团队领导。如果是规模较小的企业（价值5万~50万美元的公司），则分析主管可直接向CEO报告；如果是规模较大的企业（年收入超过5亿美元的公司），则分析主管可直接向总经理报告（见图8-7）。

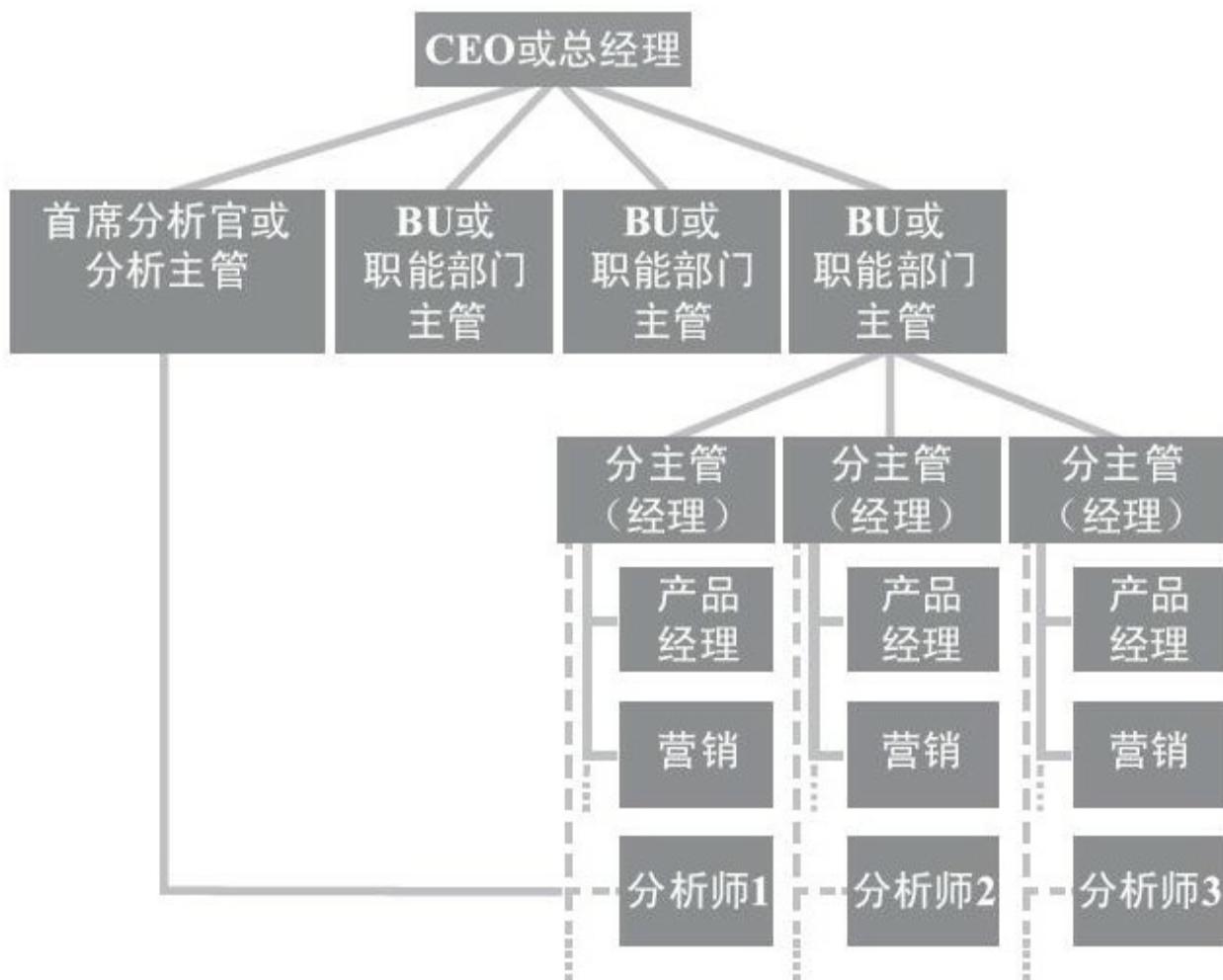


图8-7 团队结构模型

分析团队负责人的核心职能是：

- (1) 影响和管理分析议程。
- (2) 建立一种数据驱动的文化——指导并培训分析师，并对商业专业人士进行培训。

(3) 制定数据基础设施建设策略，以便在整个企业中有效地使用分析。

分析主管不一定是最好的分析师或杰出的数据科学家。相反，分析主管的角色是管理，另外还兼具部分预测和传播职能。分析主管应该是一个有能力的分析师，但更重要的是，他必须能够传达分析的力量，加强企业的KPI并提高公司的利润。

团队组成

实际的团队组成应由分析议程、工作量和性质来决定，但从广义上讲，我们推荐如下（见图8-8）：

- 在一家公司中，如果商业专业人员不具备数据-决策的技能，则一个分析团队可以由3名商业分析师、2名数据分析师和1名预测分析专家组成，这样可支持一个由20名商业专业人士组成的团队，包括营销、产品、运营、设计、策略和其他功能。
- 在一家公司中，如果商业专业人员已经拥有数据-决策的技能，那么分析团队可以由数据分析师和预测分析专家组成，可以支持多达20名商业专业人员。

在第9章中，我们将深入讨论商业层面，你可以通过第9章了解个人需要做哪些事来通过研究结论扩大影响力。

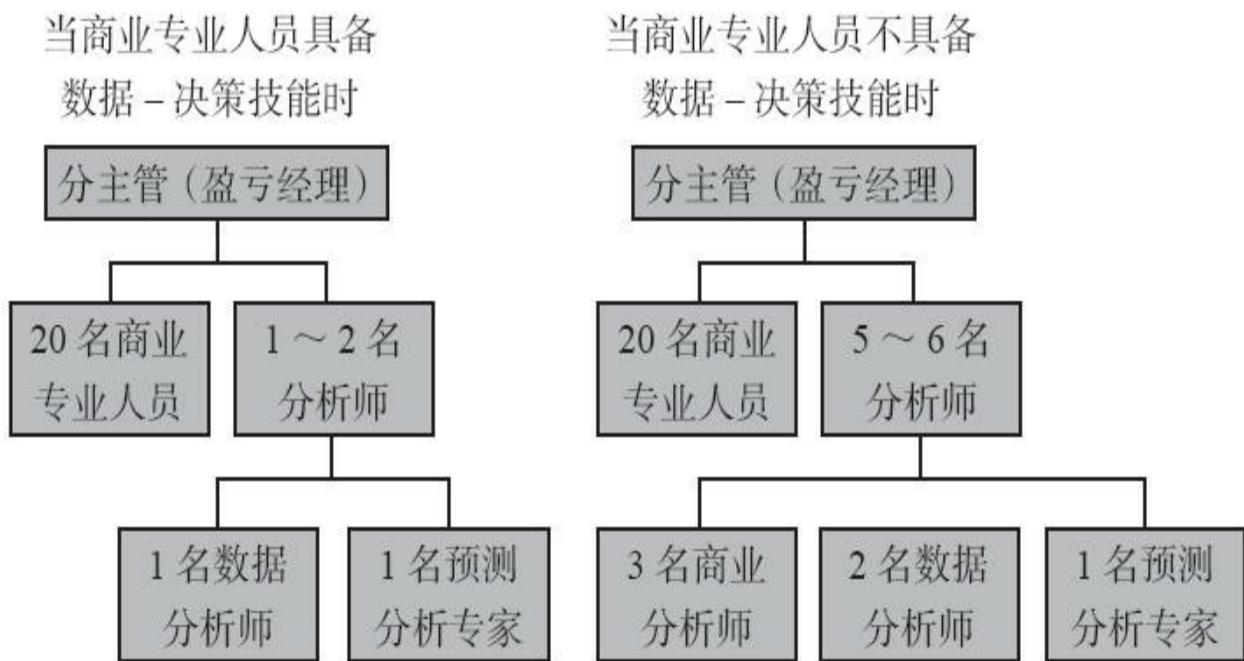


图8-8 团队组成

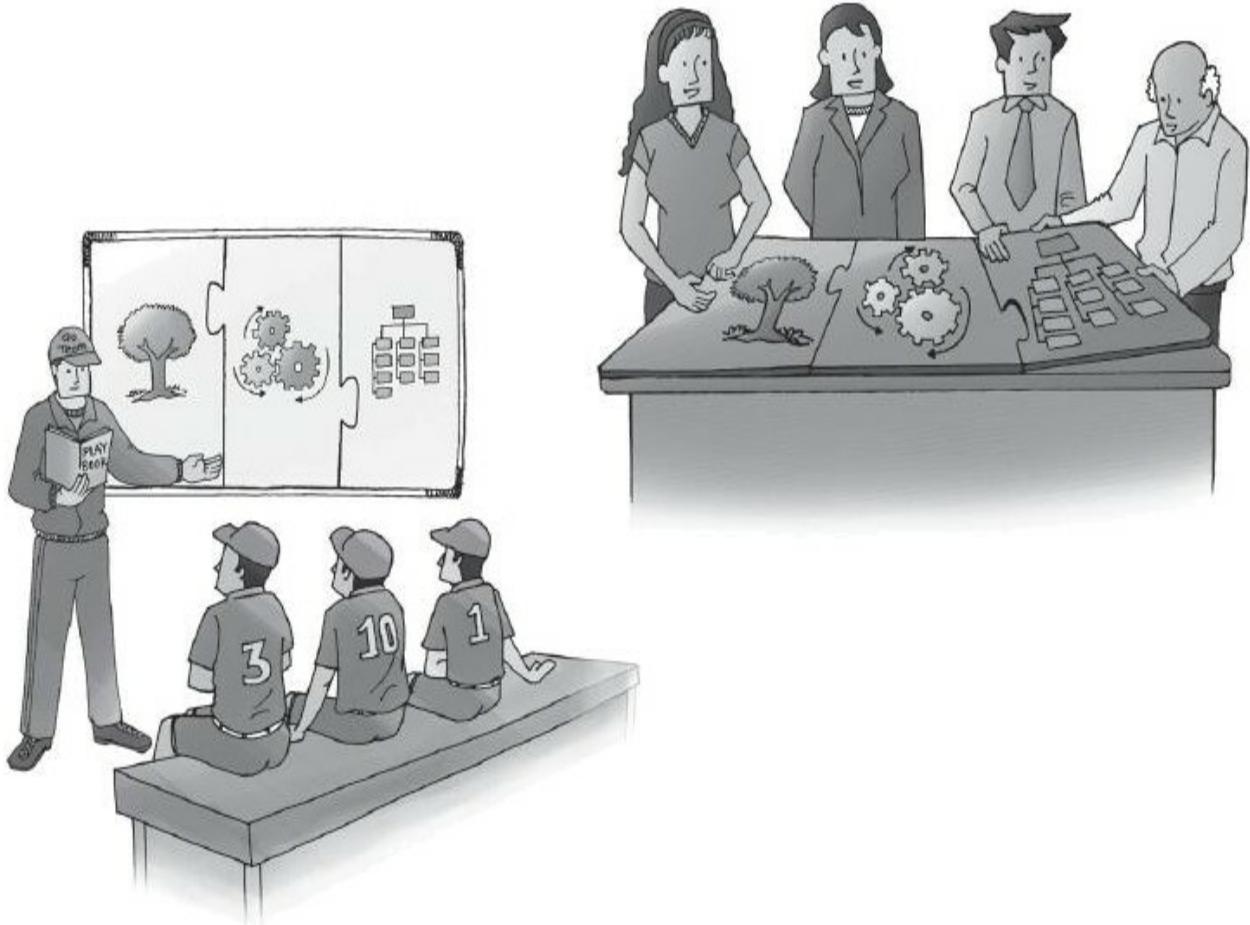
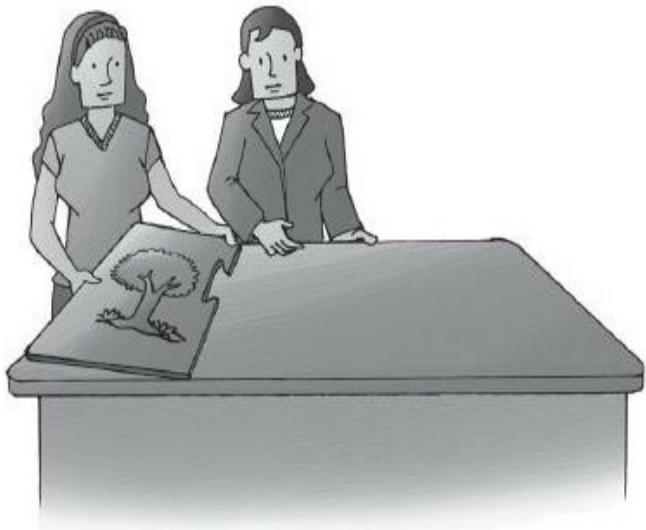
小结

- 你可以按照“3个关键问题”的结构化流程制定分析议程。
- 嵌入分析结构可帮助企业获得最佳投资回收率（ROI）。

第9章 分析师领导者的攻略手册

本章将讨论：

- 新承担分析责任的领导者在上任后前90天的攻略手册。



现在你已完成了之前的步骤，评估了企业的分析成熟度，确定了你的分析议程，为分析目标调整了适当的企业组织结构，并确定了人员配置、角色和技能。那么，你该如何将这些协调起来呢？这个攻略手册，对于

新上任的分析部门责任人特别有用。我们建议以下面这两个简单的步骤设置分析型的企业（见图9-1）。

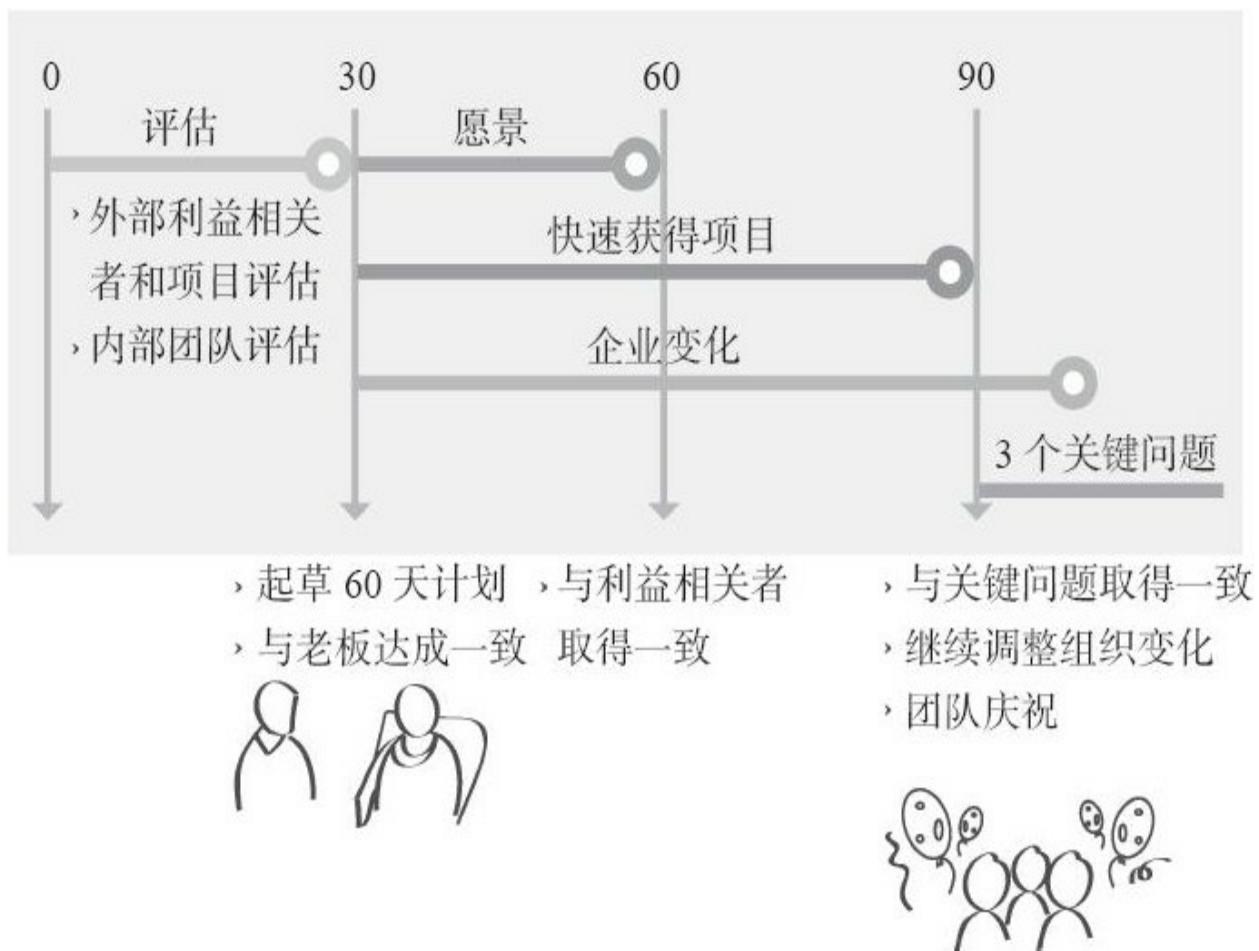


图9-1 为期90天的计划

步骤1：前30天——评估

在提出任何计划之前，请你的老板（CEO或总经理）给你30天的时间准备一份对企业的评估。

高级外部评估

(1) 审查过去和当前的工作，包括过去12个月中完成的分析项目类型及其商业影响。这将让你了解企业内部的分析能力和分析需求。它还将告诉你，你的商业合作伙伴是否提出了正确的问题。

(2) 与你的团队合作领导者和商业合作伙伴交流，了解他们的心态。询问他们的关键商业问题、分析目标，以及他们希望如何与分析团队合作。这将让你了解他们是否将你的分析团队作为支持团队或作为真正的商业伙伴。它将照亮你的职场舞台和职场必须遵循的路径，使你与企业利益相关者保持同盟，以获得企业的最终成功。简而言之，你需要对你的企业文化进行真实评估，重视企业的分析准备，采用数据-决策策略，加强企业分析成熟度以及领导和员工的共同责任意识。

(3) 最后，当你与利益相关者谈话时，你会发现一些速效方案，在这些方案中你的团队可以立即产生有意义的商业影响，而且仅需要投入少量的时间、资源和资金。建议你开始实践“3个关键问题”的方法，以确定最具影响力的数据议程。

内部团队评估

(1) 从评估分析团队的技能开始。当然，当你查看团队的分析和数据技能时，也不要忘了评估他们的软技能。

(2) 确保查看他们过去的项目、工作性质，以及他们如何使用仪表板、简单分析或综合预测分析。这将帮助你评估团队的生产力、分析能力、效率和驱动影响的能力。

(3) 从过去的重点领域、团队动力、技能和所选供应商等方面来评估你的团队。

(4) 审核和评估你的数据基础设施，以确保它能满足你预期的工作需求。

(5) 了解当前用于确定分析项目优先级的流程（如果存在）。如果没有，那么你可以计划设置一个。

步骤2：接下来的60天——计划和执行

这个60天计划是基于你在步骤1中做出的评估。你应与你的经理讨论这一点，以确定优先级和资源分配。

(1) 这个讨论应该包括你的团队使命和目标。

(2) 根据预计所需的工作量和人力，按数量、技能和预算来安排关键

项目。这些关键项目应以你在前30天与企业领导人进行访谈的结果为基础。

(3) 最后，创建一个管理计划。在该计划中，你需要估计你将花费在企业内各个分析策略的每个重点领域的时间，以及你将解决每个领域中的哪些问题。重点领域可能涵盖当前项目的分析议程以及企业变化，例如团队结构、招聘和资源规划、培训和指导，以及传播数据驱动型文化。

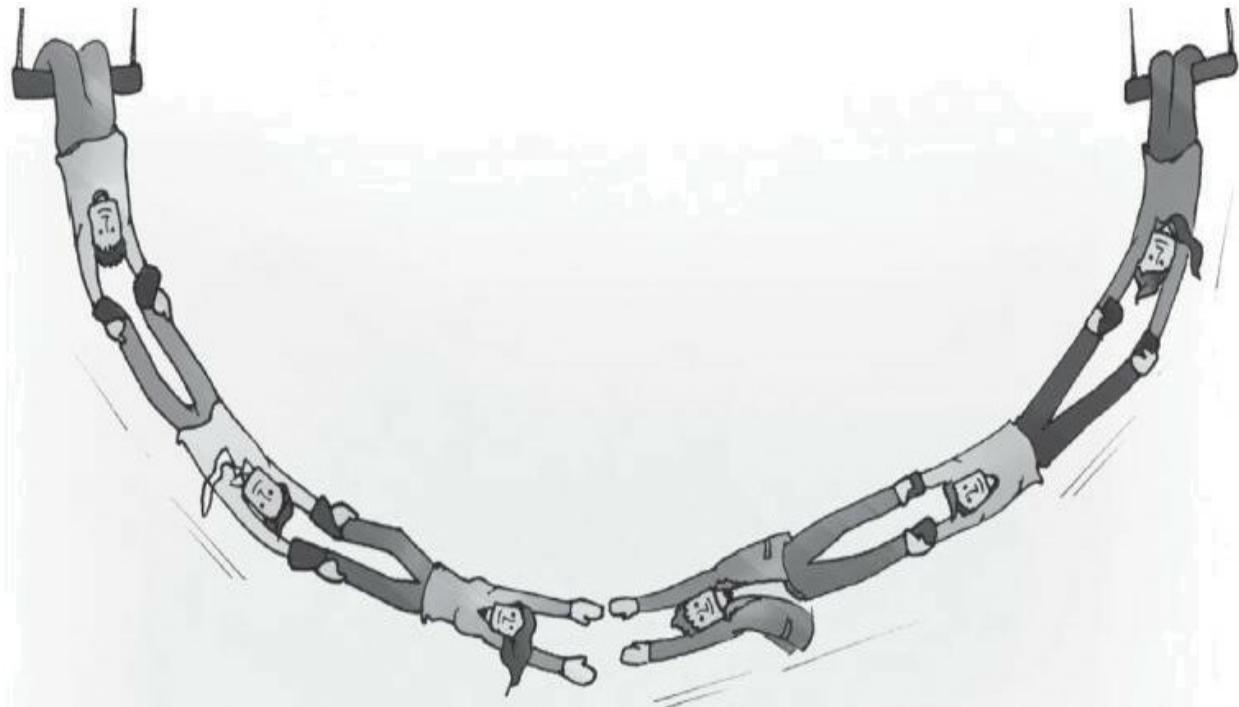
一旦执行了该60天计划，你势必会取得一些胜利，为你增加在企业中的可信度。现在就可以遵循更正式的“3个关键问题”框架，传达你的分析议程。当你执行该分析议程时，可以继续进行实现更成熟的数据驱动文化所必需的企业变革（人员、流程和工具）。记住要定期（每周、每月或每季度）与利益相关者协调。

祝你和你的团队取得成功。

第10章 付诸行动

本章将讨论：

- 如何真正推动企业中的影响。
- 有效利用影响、优先级和协调来扩大企业在商业或分析方面的影响。



比尔·谢里（Bill Sherry）在2005年被任命为圣荷西国际机场的航空主任。对于机场来说，这是一个艰难的时期，企业急需45亿美元用于机场扩建，这给所有主要利益相关者带来了很大的财政压力，包括航空公司承运人和市议会都有影响。比尔恰在这时上任了，他将这个10年项目的预算削减了30亿美元。他是如何做到的呢？当然是通过协调利益相关者与企业的共同目标相一致，并为这样的一致性去营造动机。2005年9月，他将一个由48名利益相关者组成的团体锁在会议室里，这些人包括

市议会议员、航空公司高管、财务顾问和工程师。这个会议的目的是取得这些人的一致同意，即用15亿美元预算来建造机场。这群人最终同意了该项计划，是因为他们被“强迫”留在这个会议室。解决问题的关键就是得到利益相关者的一致同意，所以这个计划也因此从一个城市计划最终变为“我们的计划”。比尔主持了这个关于建立独立机场和创建团队之间公平竞争的结构化头脑风暴会议。大家被分成多个团队，每个团队代表了不同的利益相关者团体。每个团队被要求在理论上“建立自己的航空港”。随后第二天，比尔又战略性地安排了一场酒会。因此，利益相关者共同制订了一个完全可行的计划，以原始成本的1/3，即15亿美元建造了一个机场。所有的事情在短短三天内就得到了解决！

想象一下，在商业层面，如果能与所有利益相关者保持利益一致，那么企业的力量和影响将会多么强大。这不仅仅是只针对机场这一案例，而是适用于所有公司、政府和有社会影响的组织。在领导项目的过程中，这一过程看起来可能会是苛刻和充满压力的，但它最终取得了良好的结果，推动了企业的增长。通常，作为产品经理、营销专业人员、运营经理或分析师，你的工作都是在有限的资源和紧迫的期限要求内，被期许做出正确的决策并无缝地执行。你的计划有望加强目标KPI并增加可衡量的商业价值。当然，管理层还希望立刻得到结果。

对于分析师来说，执行分析并提出结论和建议就足够了。对于商业专业人士来说，设计一个透彻全面的可快速得出结论的路线图，用来加强你的计划并执行，应该就够了，但事实并非如此。

事实是，它不仅仅需要你将这个决策付诸行动，从而实现商业价值，还要求主要利益相关者（营销经理、产品经理、运营经理和商业前线的其他人）互相达成一致。它需要让支持你的分析师与数据和数据结论保持协调，如果预备采取该措施，则还需要你的老板，甚至可能是他人的老板，来批准预算。只有所有这些利益相关者都知道并愿意与企业决策目标相一致，这些决策才能最终转化为该企业的行动。在理想情况下，你必然希望你的项目在整个企业中产生真正的影响。

为了在跨职能环境中执行并使决策得以实现，除了需要具备可行的分析结论外，你还需要掌握一些管理技术，具体来说，需要培养的技能是影响、优先级和协调。

影响

在跨职能组织中，如果能认识到以下基本原则，你的影响力便会开始形成了。

- 协同决策和利用他人专长要比个人决策更有效。
- 从一开始就要求利益相关者参与项目将加速项目的执行，不管他们是你合作伙伴还是预算审批人员。

无论项目是想为你的客户建立一个更好的产品或经验，提高营销活动推动客户参与度，还是在操作过程中节省成本，分析都可以发挥重要的作用，甚至可能是至关重要的作用，以最快的方式实现你的目标。尽管实现目标往往不取决于一个人的专业知识，而在于你的团队，但你属于跨职能项目的一部分，或者作为一个贡献者或领导者，因此，要让每个人都参与进来，大家一起跨职能地、积极高效地合作是非常有优势的。

无论是利用网站的变化来提高客户参与度，在医院急救登记期间改善病人的体验，还是减少制造单位的产品成本，你都需要带动你的跨职能团队来合力驱动，实现大家追求的共同愿景和目标。

从一开始就与利益相关者站在统一队列，提高你的影响力。任何倡议或项目都有四个阶段（见图10-1）。

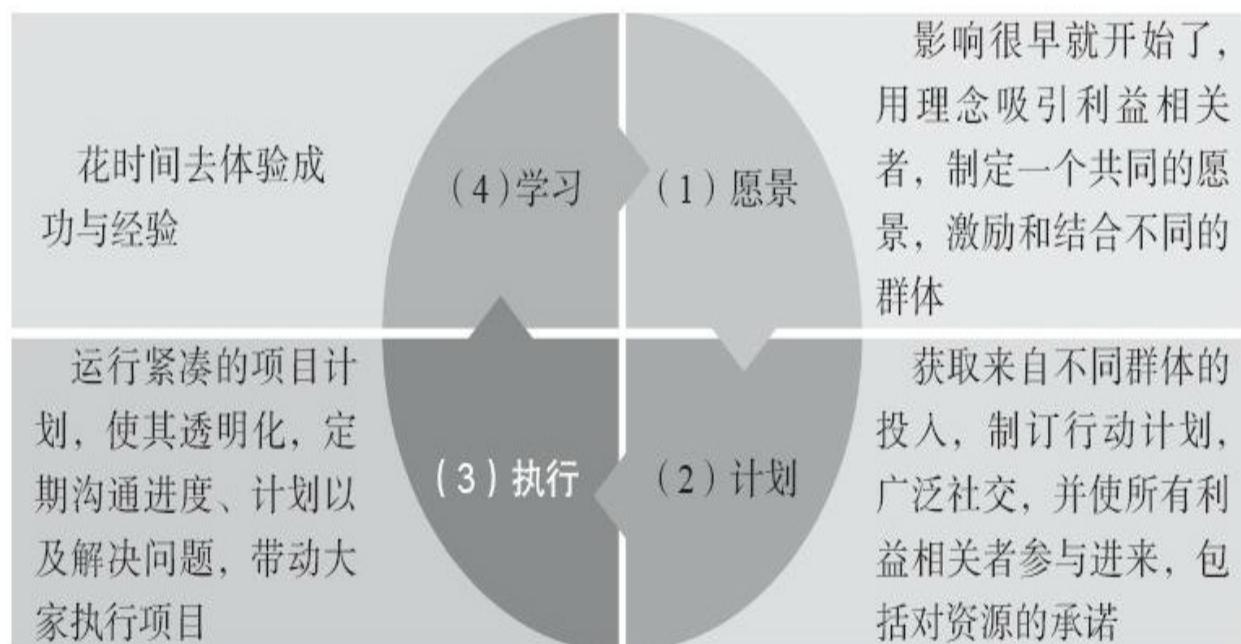


图10-1 如何发挥你的作用

- (1) 愿景：定义“那个”你想推动团队的走向。
- (2) 计划：制订详细的计划范围、利益相关者、资源和时间表。
- (3) 执行：按计划执行，如有需要，在执行过程中进行调整。
- (4) 学习：从这项计划中学习并将结果反馈到下一个项目。

愿景

在大家为愿景达成共识时，影响就已经开始了。关键是制定一个共同的愿景，用来激励和统一不同群体的目标。这里有一些快速的技巧用来实现这一目标：

- 如果群体不统一，一对一地建立共识。
- 记住，在构思阶段没有想法是不好的。
- 有一个发起者，在参与之前可以支持行动。

计划

制订行动计划的过程可以让所有涉及的组织都参与进来。一旦计划敲定，就要广泛地联系所有的利益相关者参与其中。你不仅需要按你的方法和时间表让他们参与其中，还需要他们对资源进行承诺。一个快速的办法是从利益相关者中获得资源的正式书面承诺，使他们认可计划以及角色的安排。

执行

严密执行计划是成功的关键。你需要管理项目时间表，包括一个定期沟通进程的计划，以及解决任何可能出现的问题的计划。让大家负责行动项目和时间分配是至关重要的。

有效地沟通是一个快速的办法，包括定期向所有利益相关者发送更新邮件，所有利益相关者都应该列在会议可选人选名单中。更新报告应该简短并且及时，帮助大家追踪事态发展，特别是当他们不参与日常活动时。

学习

最后，当计划成功时，你还要从中进行学习。你和团队应该对此进行庆祝！为确保你能慷慨地表达信任，要使所有的利益相关者都参加庆祝活动。此时是分享你的成功和你所学到知识的最佳时间。这将有助于组织中的其他团队学习关于客户以及产品的信息，可为他们的各自项目投入服务。一个快速的方式是给予所有参与者信任，并花时间庆祝。

优先级

难道你不希望每天有40个小时来完成我们一天中很多待办的事项吗？而且在工作中，我们所需要做的事比一辈子能完成的还要多，所以要让你的执行团队人数加倍和优先级需求变得越来越重要。优先级是调整有限的资源与最具影响的主动权，带动组织的愿景和帮助目标的实现。优化是一个非常重要的过程，能有效地管理任何类型的组织、业务单元或功能。

优先级以其最简单的形式包括从每个活动和在活动中排名最佳的投资回报率或最高回报中了解投资（I）和回报（R）。问题是活动不一定会有回报，投资附属于活动。它们通常是来自不同群体的伟大想法。有许多活动没有直接的财务目标，而是出于其他原因，如提高利益相关者的价值和提高品牌感知。

优先级的区分是很困难的，我们推荐以下四个标准化的步骤。

（1）组建一个核心团队，由不同群体代表加入。例如，如果你是一个产品营销集团的负责人，那么为区分优先级，创建一个小的领导团队，这涉及你的分析团队代表，以及从产品开发、分析，及包括在你的项目交付前涉及的任何其他团体。如果没有这个涉及不同群体的利益相关者的民主决策过程，就会有产生矛盾的风险和可能的利益冲突，这可能就会适得其反。

（2）考虑优先级设置指标和标准。第一步是让所有的利益相关者对优先级指标的标准达成一致，并确保它们与组织的目标和优先权一致。分级和预测方法可以帮助预测项目对以下因素的影响，如预期收入、投资回报率，或减少流失等因素。除了优先级标准，你还可以建立过滤要求的标准。常见的过滤指标包括战略一致性（即这是一个很大的项目，但它并不符合我们的整体定位）、执行性（即我们能做到吗？）、数据的

可用性（即我们有良好的和易于访问的数据吗？）、最低投资回报率（即投资的最低预期回报率是多少？）、最大的投资（即我们不会做任何需要超过20个产品开发日的项目）。无论你做什么，当你要在项目中应用它们时，请商定优先级和过滤标准。这将证实你的做法是公平和透明的。

（3）建立一个接受新想法的程序，然后定期、频繁地优化它们。优先排序是一个过程，应该由集团在定期安排会议中提出。同样重要的是要创造一个透明的过程，这样可以帮助培养新的想法和创新。这些项目的想法越充实，要优化它们就越容易。



(4) 基于新的投入调整优先事项。由于这些会议的召开，无论是优化过程还是优化标准都将会发生变化。你需要做好准备，灵活应对。

协调

当与许多企业内部部门（BU）在一个大型组织中一起工作时，进行正式和结构化的协调将有助于各组织和企业的跨团队策划活动协调一致，减少冗余，发挥最大的影响。协调促进工作总是从一个共同的学习议程和定期会议或工作午餐时的分享路线图与知识开始。这将确保所有的组织都在相同的进程阶段。它还可以帮助你避免过度操劳，因为这既浪费时间又使人有挫败感。最后，准备一个每周或每月的沟通会，通常用电子邮件更新每个人正处的状况就可以了。我们还发现它有助于大家了解组织中的不同群体，了解为其他项目工作的人。



无论你服务于组织的哪个部门，上面的步骤肯定都会对你的工作产生重要的影响。第11章将介绍商业分析的常见陷阱。

小结

- 有效的影响能加速各种措施的执行。

- 区分优先级是一个必要的过程，其中，民主性实施可以在促进创新的同时，保持组织始终将重心放在关键点上。

- 跨组织协调可以避免冗余，创造和谐的组织，让各团队建立在彼此协作的成功上。

第11章 常见的陷阱

本章将讨论：

- 作为商业领袖、分析领导者、商业用户或分析师，你可能犯的关键错误是什么？如何避免犯这些错误？



才刚工作的第二天，年轻的杰伊（Jay）就被要求去领导一个成交转换模型项目。他的雇主是一家年轻的、新的全球技术发展公司，它只有一个规模很小的全球分析组，且仍在朝着以分析为基础的决策性文化而努力。这个项目异常复杂，因为它涉及成百上千个变量，而且他才工作两天，这对一个新人来说太复杂了。他几乎还不知道同事的办公室在哪里。

迅速并成功地完成该项目的压力是很大的，杰伊错过了我们所教的最重要的课程：要与所有关键利益相关者碰面并达成一致。他与一位高层领导人的愿景一致，但他并没有专心去让其他利益相关者也支持他。他甚至跳过与产品经理人的接触，而这些产品经理人才是真正负责实施杰伊的团队所想出的任何方案的人。

杰伊的团队对他们极其复杂的模型苦苦思索了整整两个月，最终制作了一个很好的成交转换模型，指出五个影响成交转换的主要变量。然后，他将调查结果提交给更广泛的人群，包括他最初商议的高级负责人、高级产品负责人和负责特定产品的产品经理。评论之前，他们只是看了看执行总结：“嗯，从我们其他研究中，就已经知道这些推动成交转换的变量了。”杰伊意识到了他的错误。从那一刻起，会议气氛开始变糟，杰伊慌忙地捍卫自己的团队，同时强调了他们想出的一些新的见解。然而一个状态在那个时刻已被建立：产品团队与分析法。它花了6个月的时间与产品团队建立信任关系。尽管如此，他也是通过不断重复地提供新的见解做到这一点的。

每个人都会犯错误，你也会，关键是要从这些错误中进行学习。

这是为了避免像杰伊一样所犯的错误，如BADIR框架中不可或缺的一部分是要问正确的问题，并与利益相关者联系。

这不是一个简单的可以为节省时间而跳过的步骤。我们知道它是多么重要，多么容易被忽视。

让我们讨论参与分析的四类用户和他们经常发生的常见错误。只有意识到这些错误信息才能帮助你避免它们。这四类用户分别是商业领导、业务经理、分析领导者和分析师。本章为每个角色提供了常见陷阱的备忘录，因为识别它们是非常重要的，只有这样才能避免它们。

商业领袖：陷阱

(1) 不知道如何为你的组织衡量成功。作为一个商业领袖，你需要清楚你想要传递的关键结果，以及如何衡量它们是否成功。如果不清楚这一点，你将无法真正将数据和分析应用到业务优势中。更重要的是，没有成功的指标显示，你将不能够使其他业务经理和企业高管信服你的数据启用决策价值。你应该开始使用KPI衡量主要成果，确保KPI与组织的目标一致。



(2) 不知道成功的影响因素和这些影响因素的调整计划。这要比了解KPI更深一步。只有知道KPI影响因素，你才可以将项目与它们结合起来完成预期的业务成果。利用分析来了解是什么在影响着KPI，然后建立一个过程，通过使用“3个关键问题”来回顾一下哪些项目受到了影响。这会帮助你发现新的项目机会和调整现有项目的优先级。对于大部分已使用“3个关键问题”框架的客户，我们发现，50%的项目需要调整优先级，并且一旦这个过程受“可调整影响因素”的影响，那么资金会很容易被抽回。

(3) 缺乏数据支持的决策文化。很少组织从开始就有数据支持文化。因此必须首先倡导，然后在工作中演示，最后进行培养和培育。所以，首先需要领导希望数据驱动，这是建立数据文化的关键。有一种方法可以应对这种情况，使领导者加快这一进程：

- 将决策权交给决策者，无论结果是好还是坏使他们对结果负责，否则，人们永远不会对决策进行回顾和吸取教训。

- 建立清晰、透明的决策过程，反对独裁决策。
- 雇用一个强有力的分析领导者，并赋予这个人广泛的权限来设置分析议程，而不是只担任任务执行者。
- 建立一个机制来倾听和回应你的客户，并设计解决方案和总结经验。没有这个机制，你的业务将难以实现客户驱动。
- 适当地投资于数据和分析基础设施，而不是开发一个自发修补的信息架构——它很少能够方便快捷地获取相关数据。

(4) 用直觉来做决策。独裁的决策制定过程比较典型地，往往由人们根据商业问题或解决方案的直觉来做出决定。你会限制分析师对领导团队已经决定的事情进行挑战或者质疑吗？如果数据不支持你的决定或如果它完全抵触被赞成的决定怎么办？只有当领导者保持开放和灵活的心态，提供最强的信念作为前提时，企业才可以利用数据进行竞争。如果数据与这些假设冲突，只有当领导愿意为之改变时，企业才可以在数据利用上处于竞争优势。

(5) 直接把数据扔给数据专家，就要他们拿出结论。直接把公司的数据库扔给10位数据科学家，要求他们建立复杂的模型，是无助于产生推动业务的结论的，除非企业领导者首先了解需要使用数据来解决的问题。通过提出与业务驱动程序相一致的试探性问题，并使用明智的假设支持数据科学家，分析才可从数据中得到新的具可操作性的结论。

(6) 将分析视为成本中心而非利润中心。通常只有利润中心能在管理层获得席位。如果你是一位总经理或首席执行官，你可能对你团队的所有功能都具有领导力，如市场营销和产品开发。如果你是此功能的领导人，你将有所有部门的负责人组成你的决策团队。如果分析在一个组织内只被认为是一个昂贵的支持功能，分析领导人在利润贡献方面找不到一席之地，这是决策的下下之选。这往往是因为分析的成果影响不能直接归因于分析团队。比如，一个200万美元的增长或节约成本的成果完全归因于功能性部门的实施，从而分析就只视为一个支持功能。

在这里，我们想强调，如果分析在决策层面没有一席之地时，将是一个非常大的遗憾。首先，它削弱了分析对价值的贡献和对利润的影响。分析在你的组织里只是一种如果需要才使用的功能。其次，它将分析从商业中排除了——无法深刻地理解商业，并且根据数据分析结果来调整投

入。分析的输出和分析师所意识到的业务背景是一样强大的。再次，最重要的是，分析师如果对决策制定不做贡献，那么他只能是为商业服务了一半。在执行表中，分析负责人可以客观地从之前业务和当前业务的对比趋势中发表见解，以决定需要支持哪些早期的假设，这些都不受利润下降压力的影响。这是分析在一个组织内能达到的更高战略水平，支持突破性的创新和传达盈利目标。

真正要用好你对分析的投资，首先就需要在管理上给分析一个位置和话语权，让分析师可在这样的环境下工作，才能从他们见解的深度中找到全新的想法。

(7) 没有将分析与企业盈亏挂钩。最有效的分析团队是融进了业务组的，要承担一定的盈亏责任。这能使团队成员从内部了解业务，成为业务组真正的合作伙伴。如果分析组自成一体，则难以理解业务优先级，或不了解业务，也不能使业务人员融入分析团队。自成一体的分析组也可能在不了解潜在业务影响的情况下，陷入应付多组要求和使用一个随意的机制去优化要求的陷阱。但是，如果将独立的分析组设置为一个矩阵组织，使分析者能与业务环境更加接近，则也可以驱动影响。

(8) 分析团队内部冲突。无论分析资源以什么方式在公司内组合，针对不同的垂直业务都会有不同的分析团队。如果它能使团队解决更多的问题，并协助更多的本地化决策，这固然是好的。然而，当一个业务组因对一个团队缺乏信任或者因为小组成员努力寻求一个能够加强他们预设答案的结果而分配相同的分析问题给多个团队时，就会发生团队之间的冲突。



在这两种情况下，不仅破坏了分析资源，而且还造成了不信任，导致资源浪费，破坏员工工作动力，并可能会使最好的人才离开组织。因此，最好的做法是根据企业或者团队的职能，只限制给一个分析团队分配一个分析问题。如果某个团队缺乏信心是一个问题，也需要独立解决。

(9) 不能充分理解分析的力量。作为一个商业领袖，你需要知道分析的潜力，尤其是你可以做些什么。否则，你的领导团队可能不重视和支持用分析来推动组织的影响。我们已经看到，一些领导曾对分析有过消极的经验，例如，在试图解决一个潜在的低影响问题时，分析结果的影响很弱。同样，一些领导很少看到分析的效益，因此觉得不需要聘请高级分析负责人。这可能导致恶性循环，分析当前资源并没有能力提供高影响的分析结果，于是管理者不会投资于有效性分析，导致企业永远没有机会得到高影响结果。



对于那些很早就开始进行数据驱动的大部分客户，我们会在一个行政研讨会上去展示分析的力量：它是什么，它是如何做到的，以及其他组织如何有效地使用它。这可以激励企业领导采取行动去建立数据驱动的文化。

(10) 将分析工作外包。一个数据驱动型组织是在内部配备系统，以了解客户以及他们如何使用企业的产品。然后，再利用这些知识去做未来的产品，取悦客户。在我们作为内部分析领导的经历中，我们发现，外部顾问经常假设一些误导性的观点或是根本不能为企业创造价值。这是因为他们没有商业背景，或缺少时间将他们自己沉浸在公司细微差别的内部数据中。即使我们自己作为外部顾问，我们仍认为，聘用外部顾问发表日常的分析结论是有限的，不是长期的解决方法。一个良好的内部分析团队有全面的数据知识，在特定的业务背景下，能更好地准备提出和回答与业务相关的问题。对外部顾问的过分依赖也会向你的内部资源发出信号，表明你不相信自己的内部能力，当然毫无疑问也不利于提高士气。

然而，外部顾问可以通过加快公司向数据驱动转型的脚步来获得利益。他们可以提供一个有价值的由外到内的视觉和使用他们的专业知识，帮助企业建立方法和流程，成为真正的数据驱动型企业。外部顾问也可以用来支持你的内部团队。

对于这些论点的总结，详见表11-1。

表11-1 成为数据驱动公司商业领袖的成功关键

为你的组织定义怎样才算成功
揭开你的商业影响因素，并根据这些影响因素调整你的内部项目
建立一个数据支持的决策文化
让你的团队学会根据数据做出决策
提出好的、有影响力的数据问题，而不是将分析资源扔给数据
在决策中给分析一席之地；雇用正确的分析领导
将分析资源融入需要优化的业务中
不要向不同的分析团队问相同的问题
发展内部分析能力——人、流程和工具

业务经理：陷阱

(1) 只为数据驱动决策开空头支票。随着大数据正在流行，业务经理如果没有使用数据来推动决策，似乎会落后于时代。不过，如果你真的只想使用分析作为一种工具来影响议程，并忽略可能造成议程争议的数据观点，那真的没有任何益处。越来越多的企业领导者坚持数据驱动的决策。对业务经理来说，真心懂得用数据辅佐决策的魔力将是一个稳健的职业决策。这将阻止你根据直觉做出决定，避免缺乏或没有有效的数据或方法来支持建议的行动。

(2) 依靠别人得出论点。随着分析更好地整合进业务中，满足企业与日俱增的数据需求，可促进企业管理者更有效地使用简单的分析得到数据结论。好处是，你无须等待分析团队去优化考虑要求，就可以得到良好的和早期的观点。在利用Excel进行简单分析的培训中，你可以从随手可得的数据中证实假设和收集观点。反过来，这将有助于业务部门更有效地加入分析团队，以验证从这些简单分析中得来的假设和结论，或提出更复杂的问题或寻找解决方案。

(3) 试图衡量一切。你一定见过，甚至促成过——比如，300个指标的

仪表板。在我们的一个项目里，一次会议上，产品所有者提出了接近500个每天被执行和报告的指标。如果你需要深入分析的话，做好每一件事都是必要的，但是没有人能够或者应该每天读一个有500个指标的报告。你不会想付钱让你的人去这么做。

企业管理者应该推动团队思索涉及两三个关键成功指标和20~30个这个关键指标的影响因素。这不仅能将业务集中于真正能推动关键事物发展的尖端，也使得它有可能在数周而不是数月内就建立一个决策支持报告。最大的好处是，它迫使团队和领导人更深入地思考经营业务的前30大可见指标。



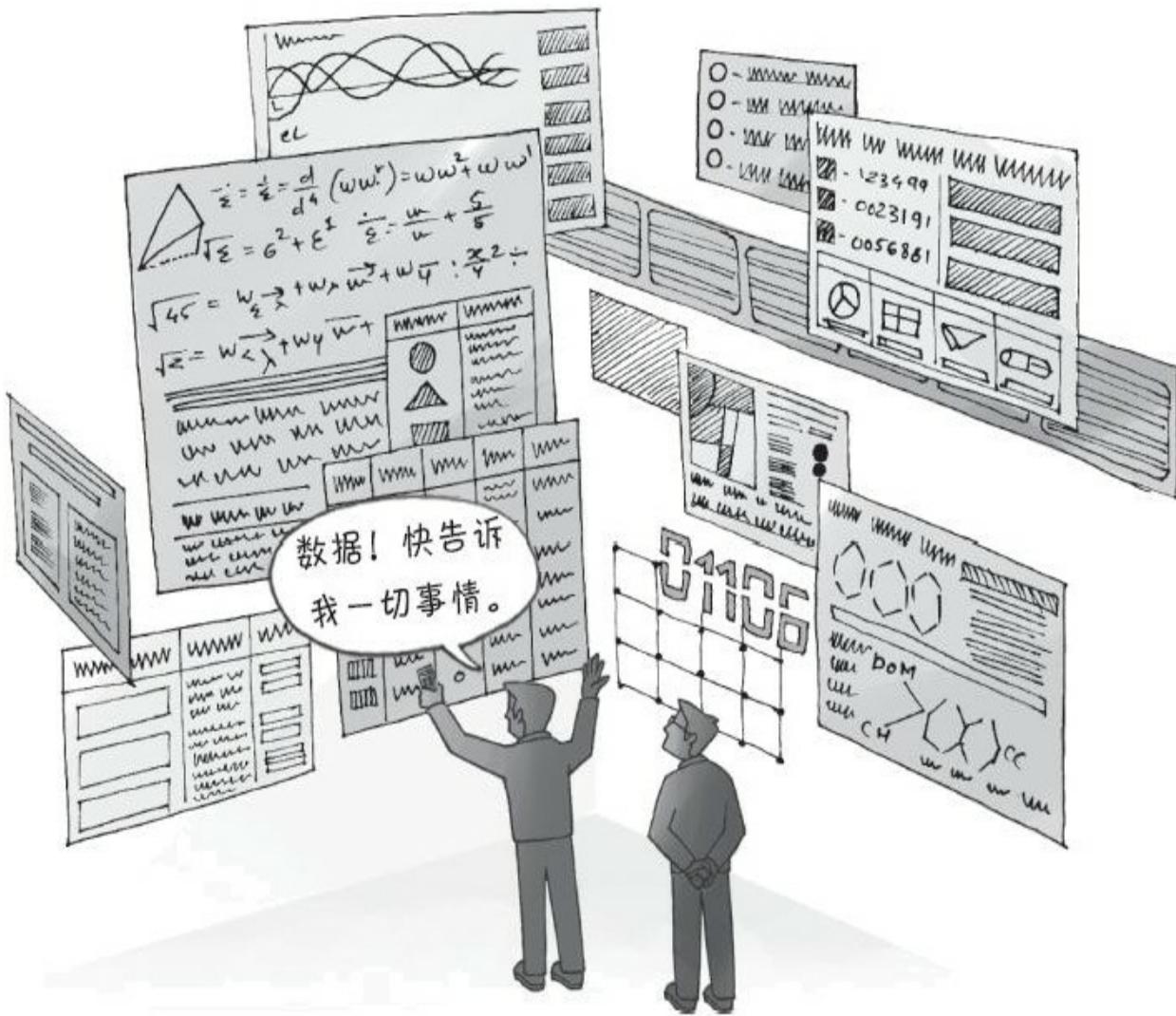
(4) 不知道提什么问题。套用一个流行的表达：“要有好结论，先得提出好的问题。”你应该在开始分析前就根据你的业务知识提出好的问题。这一条无论是对分析师还是业务经理都同样适用。闭门造车的管理人员甚至不知道要提什么问题或怎样的分析能实现什么目的。分析不是一个水晶球，而是一个强大的商业性工具供你使用和塑造。你可通过阅读、参加会议、与有经验的专家交流和从学习分析中获益。我们的目标是让你熟悉分析，促使你提出能获得良好结论的问题。这一步可参考第8章的“3个关键问题”框架。另一种看待“要提什么问题”的方法是思考哪些问题能产生观点，帮助你投入到工作中，获得商业影响。不同的问题

可以让你找到想知道的信息，提供有见解的问题答案。那些“很高兴知道”的问题会满足他们的好奇心或看起来很有趣，但答案不会导致业务上的任何变化。当问题导致了有影响力的结论时，你的组织必须有意愿去改变。如果不愿意采取行动或如果变化的潜在影响低，你只会浪费你稀缺的分析资源。

(5) 不设定优先顺序。一旦你有了20个问题的清单，你的团队就可以一次一个地解决问题，这样看起来也有逻辑，但这需要大量的时间，很多公司都不能接受，尤其是现在的业务发展如此迅速。分析20个问题需要时间和资源原本可以用在其他地方。你可以用一些时间利用一个业务过滤器去优化这20个问题，然后缩小到前3~5个，获取能够提供商业价值的最好问题。关于剩余的问题，可以预留另外3~5个最具影响潜力的问题到第二阶段使用，前提是资源仍然可用。

(6) 相信你肯定没有足够清洁的数据来得出结论。我总是从我们的客户中听到这一点。虽然在罕见的情况下，你可能真的没有清洁的数据，但是对大多数公司来说这并不是事实。与其因为你缺乏数据而胆怯害怕，不如提出你的关键问题，然后让你的团队在真正需要的数据上下功夫。你和你的团队将会对你们利用已拥有数据所取得的进步惊讶不已。一个证据充分的、可信的答案比没有答案更好。例如一位客户要求我们拿出一个客户终生价值的估计值。客户并没有太多的数据，但有一些可用。我们建立了系统，捕捉6个月的新的和清洁的数据来估计重复行为，并获得到终生价值，然后我们推算出未来一年和两年的值。这足以让公司走上正确的方向。

(7) 期待数据自己给出答案。这就是为什么要以正确的问题和假设开始分析，而不是从数据开始。数据不会说话，但它可以响应。仅仅要求分析师带着“所有相关的数据”回来，是不会对推进商业战略或目标有太多益处的。



这不仅非常耗时，而且对你的分析团队来说也是很难的。这表明，你并不知道你的业务影响因素。总之，数据只是数据，它只能通过用来证明或证伪假设而发挥作用。

(8) 对最新流行语相信太多。对许多预测分析者来说，大数据和机器学习在数据驱动管理领域是热门的东西。它们是真正的流行语，了解它们是有必要的，但这些概念对你试图解决的大部分问题都是昂贵的矫枉过正，不可能指望这些热门概念去为你解决即时的业务需求。最后，有效的数据驱动决策仍然依赖于提出正确的业务问题，并使用业务分析，利用假设解决它们。在这之后，如果你还有用预测分析帮助解决商业问题的需要，那么受益于你最初的快速分析，它已有了一个明确的范围。

(9) 将分析视为无比高深复杂。也许是一些先进分析技术的预测功能

使企业管理者痴迷。预测分析是一门科学和一门结构化的学科，并不是魔术。在第4章中，我们讨论了常用的分析方法，使你熟悉了它们的可能性和局限性。然后，你可以毫无畏惧地讨论分析，并要求你的团队提供所有适合你得问题类型的分析，而不必以为这是多么神秘高深的事情。

对于这些陷阱的总结，详见表11-2。

表11-2 作为业务经理使用分析见解成功的关键

根据数据做出决定；如果这样做的话，你会始终如一地推动结果
学习分析；没有人在邻居家里存钱，你为什么要这样
找出你在运行业务时需要的 20 大指标，并掌握它们
学习使用“3 个关键问题”框架来对你的数据提问
不断优化你的分析和你的项目
在担心你没有的资源之前，先利用你已有的数据
不要指望数据自己就给出结论，相反，要提正确的问题去找出结论
不要为最新的流行语分心，留意好你的 20 大指标即可

分析领导者：陷阱

(1) 没有行政支持。对于一个从来不重视、不使用分析的组织来说，要实现数据驱动是一个管理改变的过程，而不是轻松的一夜之间的转变。组织里有行政权力的人需要站出来捍卫分析驱动决策。如果做不到这一点，分析领导者在组织中就不可能有重大影响。作为一个分析领导者，你需要有行政力量的支持，才可能去影响组织的发展方向。你还需要得到高级领导的支持，拥有质疑现状的权限，并在决策过程、组织性结构和基础设施中利用决策分析去为组织带来改变。某天，分析职能将在公司委员会中拥有一席之地。但在此之前，你需要行政领导的支持。

(2) 与商业目标不吻合。我们的一位CEO朋友对分析非常失望。这是因为，我们发现他在大约18个月前聘请了一位首席数据科学家，并对获得大数据怀了很大的希望。CEO随后聘请了12名统计学家组成的团队，

用先进而复杂的算法开始解决复杂问题。正如预期的那样，这些问题花了很长时间来解决，但不幸的是并没有带来业务的变化。CEO花了一年时间和数百万美元后，对他的分析投资感到很失望。迫于董事会的压力，他解雇了自己和他的团队。事实结果显示，首席数据科学家完全是出于技术分析，并没有根据企业的商业实际进行分析。分析研究项目的工作单纯仅是智力回报。但在学术之外，分析工作应与重要的商业目标一致。如果不这样做，你将不会有一致的分析支持。而且，没有支持，你会发现项目会在半截就被停止（例如，突然来了一个令人失望的季度报告，然后就有人想削减成本）。在其他情况下，你可能能够完成分析项目，但得不到能够帮助你实施建议的资源。而且，直到建议被实施，组织从分析中都没有得到价值。当与企业的商业目标相一致时，分析可以成为企业成功不可或缺的关键。

（3）对于分析议程做出反应而不是转换。如果分析领导者不要求在管理层中有一席之位，很可能他们就不会意识到企业正在考虑的战略思想或举措。不知道这点，你将不能够影响组织的方向。你和你的团队将只会反应、接受高级管理层的指示。但反过来，你其实可以利用你的位置，主动提出措施。利用自己的特殊位置，通过数据和数据趋势去了解你的顾客。结合公司的发展方向，以及使用一个框架如“3个关键问题”去认知差距。你可让组织关注战略的不同级别，将内部项目与不同级别的战略目标结合，通过快速给出具有可操作性的结论来驱动巨大的影响。

（4）思想狭隘。分析领导者的作用可以很大，然而，他们又很容易因为支持中小型影响项目而分心——结果只取得了一系列小规模影响的胜利。优秀的分析领导者必须考虑重大事项并参与到企业的战略讨论中。你不能等待被邀请，而必须以你的方式进入讨论，并借助你的大思维力量（这就是为什么你需要一个背后的行政领导支持）。使用结构化的方法，如采用“3个关键的问题”框架，你可以将组织侧重于战略重点，调整内部项目的优先次序，并促进及时的行动建议，推动大规模影响。

（5）在业务和分析之间没有跨职能地达成一致。这常常会导致多个分析团队为不同的群体解决同一问题。这种与业务群体的不协调往往会导致某些职能领导质疑分析团队的价值。你需要将这种冲突可能降到最低，并与各职能部门真正地合作，高效地工作。

（6）没有建立正确的团队。你不能独自完成这一切，只有当整个团队都好时，你才可能好。这就是为什么你必须雇用拥有分析和沟通能力，

并且有技术、懂领导的人才。你需要具备商业分析、高级分析和数据技能的复合型人才。你也喜欢愿意解决难题、负责任的人。同时拥有这些才能的人几乎很少，所以要找拥有基本技能并且可以训练其他人的团队成员。记住，用错一个人都意味着工作的质量会受到影响，而你必须要学会收拾残局。

(7) 不能持续培养和激励团队。有时分析工作可能是令人沮丧的，有时候，工作会充满压力和紧迫的期限，而有时处理利益相关者关系甚至会造成冲突。没有什么能比加班工作，损耗脑细胞试图去回答一些问题，但却毫无结果会更快地使分析师失去动力了。此外，团队中的人需要感受到他们的专业技能在成长，并且变得更加受市场欢迎。为此，你需要确保：

- 你的团队正以最好的期望致力于解决正确的问题，以产生好的影响。
- 当工作赢得表扬时，你要慷慨大方。
- 当事物发展触礁时，你要感同身受。
- 大家都受打击失望时，你需要准备激励。
- 关注团队成员学习、参加会议、获得专业认可的机会。

是的，你要正视明星队员可能被挖走的风险，但那些正在做出令人满意工作的人，他们驱动了重大的影响，因此，要让他们享受认同和赞赏，为他们争取去学习和发展的机会，争取有竞争力的工资，永远表现出将团队利益放在心中的领导者形象。

(8) 没有对数据基础设施的输入。如果你不参与仪器和基础设施计划的讨论，你、你的团队和你的业务工作需要将被数据的质量和访问速度所阻碍。因此，亲自参与规划、架构和设计数据基础设施，如服务器和BI工具是非常重要的。你不必成为IT人员，但你确实希望要在流程中能输入数据，以便做出正确的权衡，使你的团队和业务团队能够轻松地访问数据。请记住，如果它不容易使用，就没有人愿意用这些设施了。

(9) 建立帝国还是传递企业价值。如果你首先投资于招聘人员和建立一个庞大的团队，迟早会有一个高管质疑你是否能创造价值。我们的经验表明，在要求增加预算，使你能够增加人手之前，最好是先用较小的

团队为你的业务带来优秀的业绩。有了优秀的业务结果才能吸引更多的项目，这将使你有机会轻松地建立你的团队。你的拥护者将会主动要求一个更大的分析团队。

对于这些陷阱的总结，详见表11-3。

表11-3 作为分析领导者成功的关键

有行政主管的主持，可能是你的经理
根据商业目标调整你的内部资源和时间
使用“3个关键问题”来推动分析议程，最大潜能地推动分析结果
思维开阔一些——你能用什么指标来利用你现在拥有的资源
通过分析和业务功能，建立跨功能调整
根据分析议程建立正确的团队驱动力
培养和激励团队，激励团队成员他们可以超越自己
提供输入数据基础设施，使数据访问对你的团队以及你的业务同行来说更容易
专注于提供商业价值，而不是建立大团队

分析师：陷阱

(1) 不与利益相关者保持一致。你必须从项目开始时就与利益相关者保持一致。否则，你可能错过利用他们的观点的机会，而且会发现自己重蹈覆辙。因此，在项目开始时就检查所有的参与人员，并在项目执行期间也进行检查。随着你不断前进，请务必向大家更新汇报你的研究结果。没有人喜欢惊奇，所以给他们定期的工作进展汇报。当然，一定要在你做最后陈述之前来做这些事，否则，得到最大惊奇的可能是你。超越紧张的人际关系，疏远关键的利益相关者会导致你最好的结论和建议无法实施，缺乏你希望实现的业务影响的支持。这听起来不太好，但解决起来很简单。

(2) 未能确定关键利益相关者。主要的利益相关者通常是产品经理、营销总裁、销售副总裁，但可能也有其他的人，因为他们也可能发挥重

要作用。这可能是业务副总裁、合规主管，或者甚至是首席财务官。这里的重点是不要错过任何相关者，正如杰伊在本章开头的案例研究中所做的。BADIR框架建议避免遗漏任何利益相关者，所以，想想谁会采取行动，谁需要给予批准，谁会批准预算等。

(3) 不知道真正的商业问题。经理给你了一个分析项目简介，尽管怀有最好的意图，但可能并没有包含整个故事。如果不花费时间去发现和理解真实的和潜在的商业问题和背景，你可能难以理解经理真正想让你的团队去解决的问题。这就是为什么你必须探索。思考如何使用分析去实现大的影响。具有讽刺意味的是，即使你的结论不能产生任何影响或驱动决策，你的经理仍然可能感到满意，但你不应该这样。

(4) 不为你的项目定义成功的标准。在缺乏成功标准的定义时，你如何判断是否达到了你的目标？所以你需要为特定的目标，以及一个具体的目标先界定正确的标准。成功是否意味着首次客户成交率的增加或现有客户收益的增加？两者都会增加收入，但如果企业希望增加新客户数量，但目前却只从现有的客户中增加收入，那么这就不能被认为是成功了。避免关于成功的争论方式是定义项目成功的标准、准绳以及你想在分析计划中实现的具体和特定目标或数量。你需要问几个关键问题：

- 什么是具体的成功指标（例如，收入、客户增长、减少流失）？
- 成功的标准是什么（例如，5%的收入增加或超过一个特定的基准）？
- 是否有办法去测量正确的指标（例如，是否正在收集能够决定成功所需要的数据）？

(5) 从数据开始。从数据开始做分析看起来很自然，但这却是一个大错误。还记得我们在第2章关于探险家和侦探的讨论吗？即使是在小公司，面临海量数据时也很容易迷失方向。所以应当从一个清楚的业务问题开始，并利用假设进行导航。正如第4章中所讨论的，假设会让你专注于你应该寻找的方向和你应该寻找什么样的数据。如果你从数据而不是假设开始，就可能会错过一些关键变量或与预算或时间表冲突，或两者兼而有之。

(6) 没有分析计划。计划是管理预期的关键部分。计划应列出了分析项目的范围、可用的资源和时间安排。它代表了一个机会，你要让所有利益相关者与计划结合。没有分析计划，你就难以达成任何经理的期

望。计划将清楚地表明是否可以容纳这样的请求，是否有成本和资源的影响。计划可以让你知道你需要多少资源。当然，如果期望高的经理要求减少利用额外的资源（人、技术、预算或时间），你可以跟他们谈论扩张计划可以达成新的期望。

(7) 使用不正确的方法。你应熟悉你可以应用的方法，了解它们的优点和缺点，它们在哪里可以发挥最大的作用，以及何时和如何有效地部署它们。选择错误的方法会破坏你的最佳成果。解决方法很简单：从一开始就得打下坚实的基础。你需要了解你的方法论。

(8) 不为特定顾客定制你的分析报告。你需要你的特定顾客定制分析报告。例如，企业管理人员需要比操作人员有不同程度的细节了解。如果你与高级商业领袖组成的听众深入去探讨细节，你可能会在5分钟之内就失去他们。高级商业领袖通常更喜欢在深入细节之前获得更大的结论、影响力和建议。当然，凡事也总有例外。所以，了解你的顾客，具体定制你的报告可以让你的演讲更加成功。这意味着虽然工作量多了一点，但它比其他方式更好，其他方式可能会在中间就切断你的路，但都无法为你的工作创造奇迹。

对于这些陷阱的总结，详见表11-4。

表11-4 作为分析者成功的关键

识别所有利益相关者，并与他们一致
花时间去了解真正的商业问题，然后开始挖掘研究论点
花时间为你的观众定制演示文稿
如果你想要强有力的可操作的结论，不要从数据开始工作

(续)

确保有一个书面的分析计划，即使它只是一个半页的文件
花时间找出你将使用的方法，并在分析计划中写出来
为你的项目定义成功的标准

认识到这些常见的错误往往就可以避免它们。在其他情况下，该解决方案涉及简单、基本的最佳管理实践和对分析的熟知。祝你好运。

小结

- 如果你是一位商业领袖，你需要设定成功的标准，找出这些成功指标的影响因素，并让你的整个团队朝着这个方向努力。
- 如果你是一位业务经理，你需要提出正确的数据问题和学习做简单的分析。
- 如果你是一位分析领导者，你需要制定一个分析议程，并与企业的成功指标协调一致，使整个组织的数据方便访问。
- 如果你是一位分析师，你需要与关键利益相关者协调一致，使用BADIR分析框架，用数据去解决业务问题。

>>【看书累了吗？休息一会！更多新书朋友圈 每日免费分享微信xueba987。渺沧海一粟：为终身学习者赋能 2019年7月】

第四部分 实践中的分析：十个案例研究

本部分主要介绍一些商业应用案例以及高级分析在不同行业的应用，案例涉及学校领域、体育领域和医疗保健领域等。我们希望通过案例激励并激发读者的创造性思维，结合使用分析性框架和智能工具来解决商业上的挑战。

案例研究1：奥巴马赢得连任

背景：对于贝拉克·奥巴马（Barack Obama）总统来说，连任并不是十拿九稳的。面对巨大的政治挑战，竞争将变得越来越激烈。当时，奥巴马面临的两个最大挑战是筹集足够的竞选资金，以及说服美国民众，他才是总统最合适的人选。

行动：奥巴马竞选团队（OFA）最大限度地利用分析、调查、测试、可视化和报告最佳实践，制定了一个数据驱动的战略框架。让50多位数据

分析师通过投票数据来预测数百万美国人的个人行为。

他的团队有三个简单的目标：①让以前投奥巴马的民众坚持再一次投他；②从新增人口中挖掘新的选民；③通过有针对性的消息散布争取摇摆不定的选民。

微目标模型：奥巴马团队使用的数据是从大量的调查、与选民互动、志愿者服务、捐助者和网站用户中得来的，结合一个1.8亿人的选民档案，团队建立了几个微目标模型来评估两个因素：选民会投票吗？选民会支持奥巴马吗？通过预估每个州和每个媒体市场对奥巴马和罗姆尼（Romney）的支持率，竞选团队计算奥巴马赢得每个州支持的概率。这些模型每周都会进行调整，然后再将这些数据通知竞选活动组织，要大家投票给奥巴马。

沟通分析：OFA有一个工具专门用来评估当地报刊上的言论，确定奥巴马讲话的哪一部分是最常被引用的、各地理区域的人们反应是怎样的。演讲稿撰写人会跟踪被报道的信息，看它们是否与其他竞选信息被放在一起，传达给选民。

民众资金筹集的策略也以分析结果为支撑，为促进选民参与带来了积极的效果。

测试：A/B测试被广泛地应用。通过电子邮件的内容来确定哪些邮件有最好的筹款回应率。OFA一鼓作气检测了主题行和邮件副本的18种变化，其中高居榜首的“我的花费将要超支”，筹集了超过260万美元。该网站在捐赠页面上进行了240个A/B测试，成功地优化了捐款页面，增加了49%的转化率。

响应分析和有针对性的沟通：竞选小组检测捐款人此前所捐最高金额的不同比例，而不是要求一个固定的捐款金额，并发现这些信息能更成功地筹集资金。基于收件人过去电子邮件活动的集资策略，能更好地实现与目标进行沟通。例如，根据目标过去是否在网上捐赠过，或是否在过去主动捐款来锁定不同的目标。

影响：我们都知道竞选的结果。但许多人可能不知道的是，模型预测的结果误差非常小，误差范围为 $\pm 2.5\%$ 。OFA最终推测奥巴马总统在摇摆州的选票结果会是51%比48%，这几乎就是最终真正的结果。分析可以锁定地方市议会竞选活动，进而完成一场全国性的竞选活动。

资料参考

www.technologyreview.com/featuredstory/509026/how-obamas-team-used-big-data-to-rally-voters/

engagedc.com/download/Inside%20the%20Cave.pdf

techpresident.com/news/23214/how-analytics-made-obamas-campaign-communications-more-efficient

<http://www.businessweek.com/articles/2012-11-29/the-science-behind-those-obama-campaign-e-mails>

案例研究2：美国女子场地自行车赛赢得奥运比赛

背景：在女子奥运会场地自行车比赛中，美国已20年没有赢得过奖牌，直到在2012年奥运会上夺得了银牌。美国队不像其他队一样有充足的资助资金，比如英国队就有一个运动员配备10名工作人员的资源配置支撑。比赛开始前三个月，美国队进入奥运会训练中心，它需要缩小5秒的差距才有可能赢得一枚奖牌，很多队员和赞助人认为这是天方夜谭。

行动：教练斯凯·克里斯托弗森（Sky Christopher）是一名非常优秀的运动员，他参加过1996年和2000年奥运会，同时也是生物和基因组健康公司的创始人。他的团队通过传感器和摄像机每天24小时收集选手的自我量化数据，包括基因测试、环境信息、睡眠模式、血糖水平、血压水平、心率，并将这些数据相互进行关联。除了报告看得见的关键指标，团队也会报告骑手的其他指标，如日常活动和营养结构、早晨常规的户外训练以及他们每次的表现，最后还对各种变化对队员表现的影响做了趋势分析。

根据分析结果，团队制定了比赛策略、健康和体能恢复路线，并且对队员的日常活动方式和习惯做了调整。总之，队员采纳了分析所产生的新建议来改善身体素质和提高日常表现。

影响：美国队最初以5秒的弱势进入比赛，但在半决赛中以0.08秒的差距击败了原来具有压倒性优势的澳大利亚队，赢得银牌。

资料参考

<http://datameer.com/learn/videos/us-womens-olympic-cycling-team-big-data-story.html>

fora.tv/2013/06/27/The_US_Womens_Cycling_Teams_Big_Data_Story

[//enhancedkinetics.com/interviews/sky-christopherson-biometric-hacking/](http://enhancedkinetics.com/interviews/sky-christopherson-biometric-hacking/)

www.wired.com/2012/06/insidetracker/

www.dailymail.co.uk/news/article-2177666/London-2012-Olympic-athletes-swap-personal-fitness-data-latest-gadgets-bid-2012-edge.html

案例研究3：孟菲斯市打击犯罪

背景：2005年，孟菲斯被认为是美国犯罪最猖獗的城市之一。那时，孟菲斯大学教授理查德·尼科夫斯基（Richard Janikowski）曾与孟菲斯警察部门合作设计了一个预测分析工具，并应用其创建了一个打击犯罪的模型，被称为“蓝色粉碎行动”。

行动：核心策略是利用犯罪数据统计来识别犯罪率高发区和犯罪高发时间。孟菲斯警署已搜集并绘制了每宗罪案资料的数据地图。蓝色粉碎行动使用这些数据来建模和创建“重点区域”地图，并推测了影响因素，如一天的某个时段和一周的某一天。当模型显示了一种可能的活动模式时，就可以根据实时情况派出警察增援，积极主动地阻止犯罪或减少犯罪的发生率。警局也可以从不同的执法单位调配部署资源，如打击有组织犯罪部门、特别行动队、交通队和酒驾（DUI）执法部门等。

执法人员如果有手持设备可以在现场提交报告，就能大大加强该部门的数据收集，也为正在行动的其他执法人员创建了一个及时可用的数据库。

影响：从2005年到2012年，在该计划执行的第一个7年里，暴力犯罪下降了23%，入室盗窃案在孟菲斯的下降率超过全国平均水平的5倍。IBM的案例研究估计，按照如果没有分析技术的帮助，孟菲斯想要减少以上比例的犯罪率需要投入额外警力所需的花费来计算，孟菲斯取得了863%的投资回报。

资料参考

<http://www.commercialappeal.com/news/2010/sep/19/blue-crush-gives-ibm-a-boost/>

<http://www.memphispolice.org/BLUE%20Crush.htm>

<http://www.commercialappeal.com/news/2013/jan/27/blue-crush-controversy/>

<http://wreg.com/2013/05/01/the-brain-behind-operation-blue-crush-retires/>

案例研究4：控制疾病暴发

背景：曼尼托巴（Mantoba）是加拿大最大的猪肉出口省，美国是它最大的客户。由于一种失控的疾病在猪身上暴发，曼尼托巴的猪肉出口业出现了经济上的滑铁卢。按照惯例，曼尼托巴农业、食品和农村计划部门通过口头传达的方式从（包括农民、实验室工作人员和兽医等）各种来源收集关于疾病的潜在暴发数据，来制定预防和控制动物疾病暴发的策略。

行动：2010年，该部门设立了一个电子追踪系统，使用位置识别来追踪牲畜的移动。跟踪系统推出后不久，就在三个临近农场的一群猪中发现了一种传染性胃肠炎（TGE）。

该数据库收集了家畜的信息。疾病跟踪可视化有助于快速确定疾病暴发以及暴发的规模，并能查看感染家畜的移动。通过分析，他们能够计算出感染区以外的潜在暴发区。预测模型有助于评估暴发疾病的严重性。假定的情景规划有助于了解疾病传播的潜在模式，以及可以制定相应策略来解决这些问题。相关性分析有助于识别出进入和离开疾病感染农场的动物。

影响：通过快速识别和分析风险，疾病控制时间缩短80%。没有多余的农场受到猪传染性胃肠炎疫情的影响，而且出口限制风险明显降低。通过分析，防止了数百万美元的相关损失。

资料参考

http://www.ibm.com/smarterplanet/us/en/leadership/mafri/assets/pdf/MAFRI_

<http://www.ibm.com/smarterplanet/us/en/leadership/mafri/>

案例研究5：揭露J.K.罗琳化名出新书

背景：罗勃·盖布瑞斯（Robert Galbraith）写了一部名为《杜鹃的呼唤》（The Cuckoo's Calling）的犯罪小说，这是他的处女作。英国《星期日泰晤士报》在推特上收到一条匿名提示消息，声称盖布瑞斯实际上是J.K.罗琳，也就是《哈利·波特》系列的作者。

行动：在与出版商确定此事之前，《星期日泰晤士报》的美术编辑理查德·布鲁克斯（Richard Brooks）和他的团队做了一些调查，发现两位作者的出版商和代理商的确一模一样。他们咨询了两位计算机科学家，来确认《杜鹃的呼唤》和罗琳的书是否能显示出语言上的相似之处。

法律语言学家通过统计分析挖掘文本、博客、推特和亚马逊书评来寻找作者的蛛丝马迹。作家会选择一些字词来传达特定的信息，这些字词带着他们本人都意识不到的个人独特信息。他们给了计算机科学家5本书来检验上述假说，其中4本书都来自大家熟知的作者，另外一本则是罗琳的《临时空缺》（The Casual Vacancy）。

在第一项测试中，每本书都通过计算机程序JGAAP运行分析，一共用了4种不同的分析将《杜鹃的呼唤》与其他书对比，每一种分析都专注于写作的不同方面。一种是比较词对，或比较相邻的词组。例如，它会显示作者对事物的分类，作者描述为“昂贵”的物品有汽车、衣服、食物等。第二项测试寻找“n个字符”或相邻的字符序列。例如，搜索“跳”（jump）的序列，将出现跳跃（jump）、跳跃的第三人称形式（jumps）、跳跃的过去式（jumped）和跳跃的进行时（jumping）。第三项测试集中在最常见的单词，像a、and、of和the等词。通过观察每本书中100个最常见的词，比较常用词的使用频率差异会发现，一个作家使用某词的频率为5%，而另一个作家则为7%。第四项测试只查看单词长度来分析这本书有多少比例是由3个字母或4个字母的单词等组成的。第四项测试的模式显示出了非常典型的罗琳用词风格。

另一位科学家对8本书进行了第二次分析，也就是对4个作家每人又分别测试了一本书。使用语言学软件程序对这8本书和《杜鹃的呼唤》进行了分析。所有的书都比较以下6种特点：词长、句长、段长、字母频率、标点符号频率和词的用法。通过测试，证明《杜鹃的呼唤》与罗琳的一本书最相似。

影响：布鲁克斯相当自信地认为推特的匿名提示所言不虚，他立即联系到了罗琳并得到及时确认，她确实是《杜鹃的呼唤》的作者。

资料参考

phenomena.nationalgeographic.com/2013/07/19/how-forensic-linguistics-outed-j-k-rowling-not-to-mention-james-madison-barack-obama-and-the-rest-of-us/

案例研究6：本书书名的确定

背景：我们是分析专家。但是，作为第一次出书的作家，我们感觉就像从消防水管里喝水一样，需要了解的有关书籍写作和出版过程的内容太多。摆在我们面前的首要挑战就是确定本书书名，这个书名要能吸引目标读者的眼球，进而吸引读者来买书。

行动：蒂姆·格拉尔（Tim Grahl）的《第一个1000份副本：书籍营销指南》（Your First 1000 Copies: The Step-by-Step Guide to Marketing Your Book, Out: think, 2013）引起了我们的注意，这本书有一个完整的章节讲述如何确定一本书的书名。格拉尔的书名与我们的需求相关，所以我们立刻就决定要买他的书。他列出的数据驱动过程与我们产生了共鸣，为了找出最好的书名，我们也学着使用他的搜索过程。

(1) 我们在脸谱网、推特、商务化人际关系网上向目标读者描述书籍的概要，并鼓励他们集思广益发表对书名的创意。将网友给出的书名列表与我们内部团队头脑风暴产生的列表相结合。

(2) 我们从上面的主列表中挑选，有45个书名进入最终的候选名单。

(3) 我们接着号召目标读者投票，从而挑出候选名单中的前8名。

(4) 然后我们用一个叫作PickFu的服务设计了一系列的A/B测试。利用PickFu的特征对读者进行调查，以确定他们为什么做这种选择。根据这些信息，我们调整了最终候选名单中的主书名和副书名组合。PickFu让我们挑出了前四名的组合，然后就止步不前了。利用目标读者的人数统计，前四名的投票结果差别不大，因此书名A和B之间没有明显的差别。我们需要用另一款工具来裁决最终获胜者。

(5) 我们把测试和调查结果移到谷歌关键词广告 (Google Adwords) 中，向所有输入目标搜索关键词（例如，分析、商业分析等）的用户展示了前四名主书名和副书名的组合，根据点击通过率决定了最终标题。

影响：本书英文书名Behind Every Good Decision与其他几个书名相比大比分获胜，其他的标题有《提高决策精度》（Adding Precision to Your Decision）、《分析最小化》（Minimizing Analytics）、《结果最大化》（Maximizing Results）等。《数字决策力》（Behind Every Good Decision）的点击率是第二名的4倍。通过让目标观众参与决策制定，也为本书造势。因此我们要在此感谢格拉尔和分析方法！

案例研究7：谷歌的创新秘诀

背景：谷歌上市（IPO）10年内就成为世界第二大有价值的公司。谷歌的成功可以归功于它能吸引、留住并培养杰出的人才。谷歌因为基于“人员分析”方法的独特人才管理实践，才有了今天的成就。谷歌公司认为，只有拥有最具创新力的创新型人才，企业才有可能将创新最大化。

行动：谷歌人力资源部门（HR）在处理员工的相关问题时，要根据数据分析过程做出决策。人员分析团队在每个人力资源小组中都派一名成员融入其中，该成员直接向人力资源副总裁汇报工作。他们通过分析产生结论并提供建议。那些起源于分析方法的特殊人力资源政策如下。

(1) 优秀的管理者与卓越的管理者：“氧气项目”分析了大量的数据，发现卓越的管理者是留住员工、提高员工绩效的关键。分析还发现与优秀的管理者相比，卓越的管理者有8个特质，这其中包括对手下定期一对一辅导，还有经常会根据不同的员工给出针对性的反馈，这些特质比专业技术知识更加有价值。如今，员工每年都会根据这8个特质给管理者进行两次评级。

(2) 快乐与健康：谷歌的创新实验室（PiLab）通过实验找到保持高效工作环境的方式。健康是生产力的关键因素，创新实验室发现，谷歌只要简单地缩小食堂餐具的尺寸就可以帮助员工减少热量的摄入，保持身体健康。

(3) 招聘：谷歌有一个科学的招聘方法，甚至开发了一种算法来预测哪些候选人的成功概率最高。分析还表明，对于潜在的员工进行4次面

试后，其他额外的面试都没有什么价值，从而大大缩短了招聘时间。

(4) 好与最好：人力资源分析证明，出色的技术专家的表现和一般的专家间会有300倍的差异。这个证明有助于得到所需资源来雇用、留住和培养杰出的技术天才。

(5) 工作场所设计推动合作：谷歌研究数据显示，创新的增加来自三个因素的结合：学习、协作和乐趣，所以谷歌的工作场所旨在最大限度地提高这些因素。实际上，谷歌甚至追踪员工在食堂排队花费的时间来最大限度地提高协作力。

(6) 增加发现和学习：谷歌知道大部分人都是边学边做。因此，与其侧重传统的课堂学习，不如强调在实践中学习，增加项目轮换，从失败中学习，并邀请像阿尔·戈尔（Al Gore）和嘎嘎小姐（Lady Gaga）这样的名人来做演讲。

影响：谷歌的员工都有惊人的生产力，几乎无人能比，平均每个员工每年创造100万美元的收入。

资料参考

<http://tlnt.com/2013/02/26/how-google-is-using-people-analytics-to-completely-reinvent-hr/>

案例研究8：扭转汉弥尔顿县高中的辍学率

背景：田纳西州的汉弥尔顿县注意到自己县的学生分数总是低于国家基准和标准化测试成绩，而且辍学率也居高不下。行政管理人员意识到如果能弄清背后的原因，他们就可以采取行动来解决这些问题。

行动：克里克·凯利（Kirk Kelley）博士是问责制和测试部门（Accountability and Testing）主管，她发现辍学率高和学习落后的问题需要针对个别学生来解决。首要问题是识别出高危学生，不要等他们的问题严重到要辍学的地步。但是，那些警示信号在实际中往往非常复杂多样，使老师很难跟踪和观察这些信号。

凯利帮助开发了一种学生表现建模工具，从汉弥尔顿县的78所学校中采集学生的个人信息数据，并创建了预测文件来把某学生标记为“不稳

定”或“偏离轨道”。该模型确定了哪些因素是学生失败的最强警报器，并帮助需要积极指导的不稳定学生。学校的负责人用创造性的手段接触这些学生，甚至通过脸谱网和学生建立联系。

该县还分析了教师奖励计划。研究人员调查了8年级考试成绩和高中考试成绩之间的历史相关关系，预测每个学生成绩的基线，然后根据教师突破基线的能力来给予相应的奖励。

影响：汉弥尔顿全县毕业率提高了8个点，增至80%。过去几年数学和阅读标准化考试成绩也提高了10%以上。

资料参考

www.ibm.com/smarterplanet/us/en/leadership/hamiltoncounty/assets/pdf/IBM

案例研究9：防制洗钱

背景：据英国广播公司监听部门（BBC Monitoring Service）透露，每年在哥伦比亚洗钱的数额高达约27亿美元。所以哥伦比亚对银行提出了更严格的反洗钱报告要求，这意味着，哥伦比亚最大的私人银行——拥有超过600万客户的哥伦比亚银行需要新的方法来分析数据。因为哥伦比亚银行正处于收购对手银行Conavi的关键时刻，所以完成这一要求难度极高。哥伦比亚银行已经有了严格的规定和程序来标记那些欺诈性交易的，但不是简单修改一下这些规定和程序就可以检测到不断变化的犯罪行为。想要达到新要求，银行需要投入大量的人工劳力去审查那些被标记为欺诈性交易的账户。

行动：银行建立了预测模型，检测可能与洗钱或恐怖主义融资相关的可疑交易。该模型每天分析200万次交易，主要通过分析关键变量，如季节性变化、年龄、职业、身份，识别出偏离预期模式的商业客户和个人客户。该模型还可检测出货币交易中交易的发起人和受益人之间微妙的可疑关系，而这很难进行人工跟踪。

影响：该模型将可疑交易的报告率提高了40%。模型的主要功效在于降低了错误分类，使交易审查量从每年120000件下降到只有6000件。银行提高了80%的生产力，并且可以进行资源重组，把原来花费在交易审查上的资源分配到其他创收活动上，比如新增业务。

资料参考

public.dhe.ibm.com/common/ssi/ecm/en/ycl03092usen/YTC03092USEN.pdf

案例研究10：红牛赞助F1逆袭，后来居上

背景：2004年，红牛能量饮料公司收购福特的一级方程式（F1）赛车队，原来叫捷豹赛车队（Jaguar Racing），现在被称为无限红牛赛车队（Infinity Red Bull）。五年来，该队没有赢得过一次F1比赛。车队最好的排名也就是2004年排在11支车队中位列第七位。

行动：为了分析车队的表现来创造提升的机会，红牛从每一场比赛收集并查看了100千兆字节的数据。大部分是遥测仪在比赛过程中记录的赛车数据。100多个传感器在收集温度、重力、旋转等这些信息，工程师团队在一个偏远的车库里对这些数据实时进行分析。这些数据同时也被发送到红牛公司在英国的运营中心。运营中心有24名分析师，用4排屏幕分析数据流，与美国国家航空航天（NASA）的任务控制中心很像。

比赛期间，赛车各个方面的专家综合分析计算机的输出数据。他们配备耳机，与赛车手、工程师和维修人员联系，帮助团队大幅度提高对地面情况的反应时间。

影响：2012年，红牛车队的明星车手塞巴斯蒂安·维特尔（Sebastian Vettel）在20场比赛中赢了5场，自始至终遥遥领先。在本赛季的最后一场比赛中，赛道上的一起事故可能会导致红牛队与冠军无缘。然而，车队并没有让赛车手下车去评估损坏程度，而是让他继续待在赛道上，工程师通过传感器传来的数据进行评估。这大大节省了时间，使得维特尔能够顺利完成比赛，最终排名第六，但仍然保持在领先地位，为他自己和他的团队赢得了2012年一级方程式赛车世界锦标赛。

无限红牛赛车队现在占一级方程式赛车的主导地位，就像21世纪早期迈克尔·舒马赫（Michael Schumacher）的法拉利队一样。红牛队自2010年后每年都获得双料冠军——个人车手冠军和团队冠军。

资料参考

qz.com/131429/how-upstart-red-bull-racing-came-to-dominate-formula-1-with-superfast-internet-and-tons-of-data/

>> 【看书累了吗？休息一会！更多新书朋友圈 每日免费分享微信
xueba987。渺沧海一粟：为终身学习者赋能 2019年7月】

附录 商业统计101

这是关于基本商业统计学的简短入门，让我们先来看一下定义。

所谓总体（population），是你正在研究的人或事物的集合（即“数据集”）。它可能是你公司所有的员工，也可能是2月生产的所有零部件，或是一个特定的客户群体。

在商业领域，我们的目标是了解和描述某一总体的特征，例如客户的年龄。两种常见的描述性参数有平均值和方差（或标准偏差）。平均值是测量集中趋势的度量标准，标准差是对离散程度的测量。

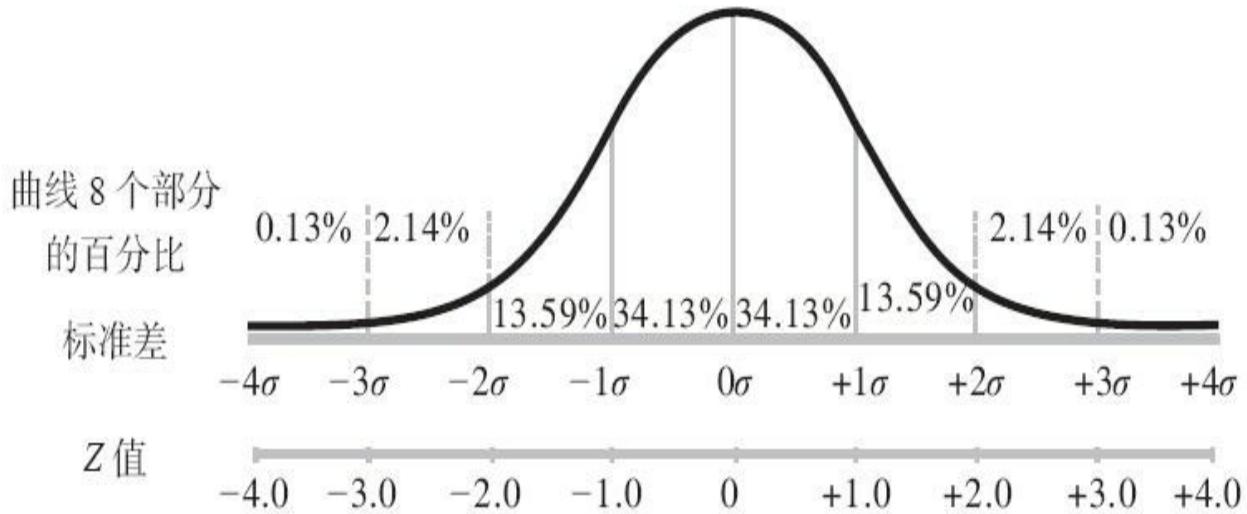
平均值

有三种测量平均值的方法。

(1) 均值（mean）是一个给定数集的总和再除以集合的数据量规模。它是一个对正态分布最常用的求平均值的方法。在我们的日常生活中有很多事项都遵从正态分布，分布曲线看起来像一个钟形曲线（见图A-1）。例如，一节课早上8点开始，这个曲线将大致代表学生到达的时间；平均值大概是早上8点左右。

如果 μ 代表平均值， n 代表数据集中的数据量大小， x_i 代表每个数据点，那么可使用下面公式计算得到：

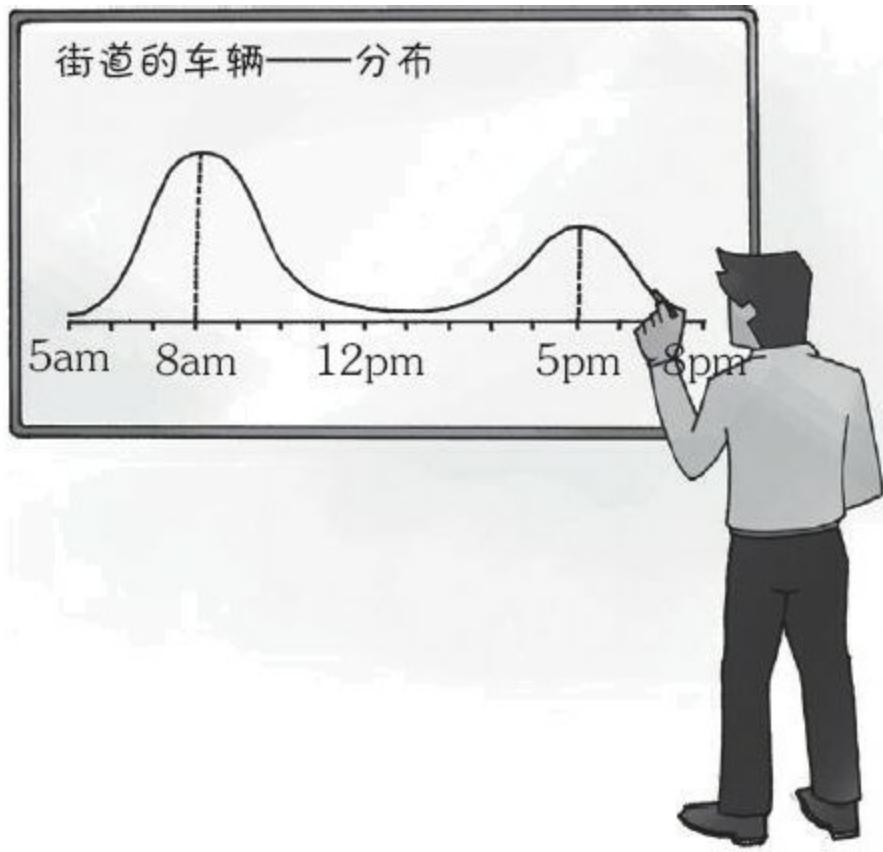
$$\mu = \frac{\sum_{i=1}^n x_i}{n}$$



图A-1 标准的钟形曲线

(2) 中位数 (median) 是数据集中间位置的数（当样本有异常时作为平均值使用）。

(3) 众数 (mode) 是最频繁出现在一组数集中的数，这种行为被称为一种模态。在非正态性的分布中，众数能比均值或中位数更好地代表平均值。例如，如果观察道路上的汽车数量，它通常是非正态的双峰式分布，其中的数字峰值大约是早上8点和下午5点。对城市规划者来说，了解街道上的平均车流量没有了解道路上最常见的车流量峰值重要。



标准差

标准差 (σ) 是方差的平方根——测量一组数字偏离平均值 (μ) 的距离。它用来描述数据集的偏差，使用下列公式计算：

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}}$$

如果某种情景可以近似为正态分布，那么知道均值和标准差，我们就可以粗略地估计有68%的值会在平均值的一个标准差范围内，95%的值会落入平均值的两个标准差范围内。这是任何正态分布的一个标准特性，当我们不知道人口属性，不能测量中间68%的值时，可以使用一个已知的样本标准差来推断人口分布。

在综合分析中，我们使用平均值和标准差描述某个总体或比较两个分组。让我们回顾一下第3章的山葡萄酒案例。

在300名顾客的样本中，有85%是女性，平均年龄为33岁，其中60%人住在俄勒冈。如果年龄的标准差是2，那么我们知道68%的客户（在这一群体中）会处在31岁和35岁之间。

标准差越小，人口特征统计和解决方案就更加有的放矢。可以帮助我们了解到偏差是2年，而不是10年。如果是10年的话，我们会考虑将该组进行再次细分，因为该组中可能存在许多子群。

Z值

这是一个术语，用来描述一个数据值距平均值有多少个标准差。它用来评估某个特定的点是典型的还是不典型的。Z值的计算公式是：

$$Z = \frac{X - \mu}{\sigma}$$

其中X就是你需要计算Z值的那个数。

玛丽，20岁，想在山葡萄酒厂预定她的婚礼场地。因此她的Z值为 $6.5 [= (20 - 33) / 2]$ 。由于她距离平均值有两个以上的标准差，所以她不是你的典型客户。假如说玛丽愿意多花钱，要求你提供现在没有的一些额外设施。因为你已经知道她不是一个典型的客户，所以你没有必要去购买那些额外设施，除非你的许多典型客户都有这些需求。

这就是使用商业统计学来做商业决策的方法。

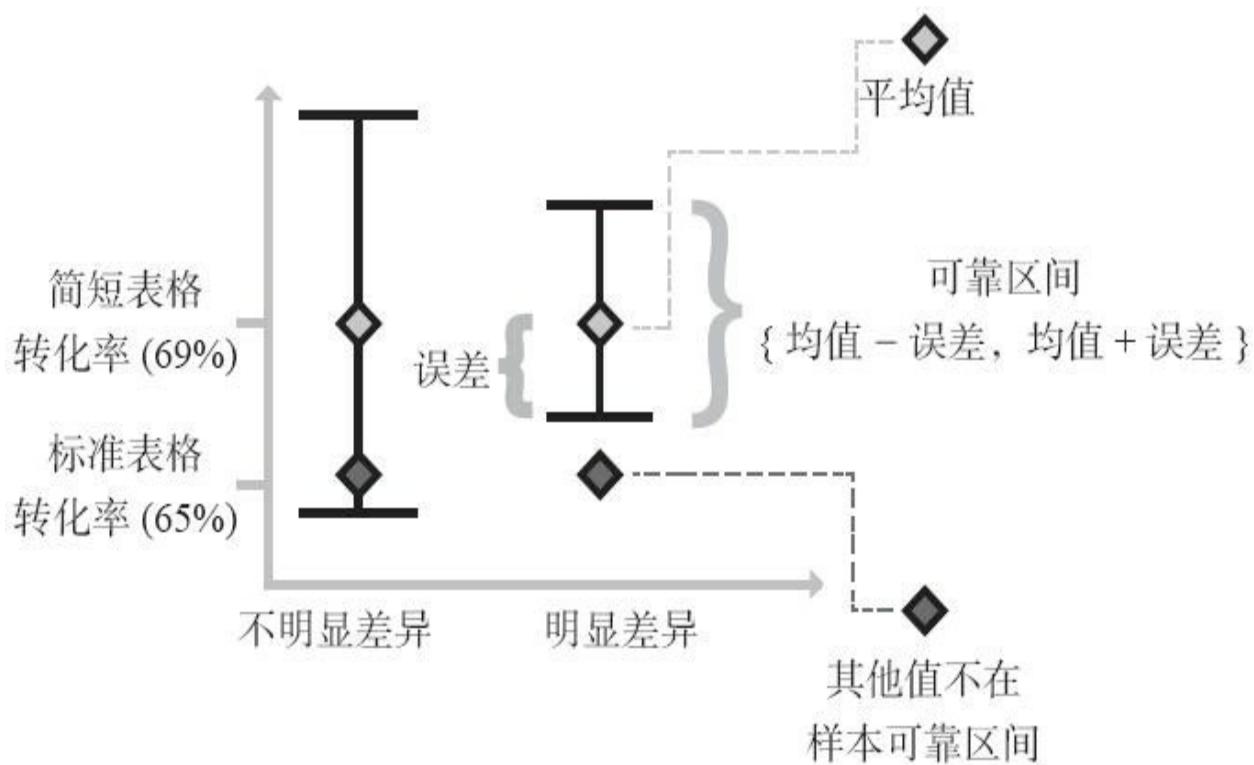
误差

在山葡萄酒厂，假设你想要对比使用新的简短表格与目前的标准表格来提交婚礼信息请求的人数。假设你测试了两种表格形式的成交转化率。假如简短表格的转化率是69%，传统标准表格是65%。

如果把表格发给下一周的访客，你觉得得到的数据结果会完全一样吗？当然不可能。这是因为当你从总体中取样时，会带有差异，这就是统计

误差。误差是样本大小、样本百分比或标准差和Z临界值的函数，是一个常数。如果我们要达到95%的置信度，我们就令Z临界值取1.96（请看图A-1的正态分布图）。

因此，如果想证明第一周69%的转化率和65%的转化率之间的差异是真正有意义的，那么百分比差异需要大于误差。假设对于山葡萄酒厂的表格来说，通过公式发现误差率是2%。由于4%的差值大于误差率2%，就可以有95%的信心证明简短表格比传统标准表格能带来更高的成交转化率。误差在统计学上意义非凡。如果误差率是6%，那么4%的差异就小于误差率。说明两种表格的成交转化率差异没有什么显著的统计学意义，即两种表格的成交转化率是相似的（见图A-2）。



图A-2 统计学差异

事实上，媒体在政治报道中就玩弄过统计学中的误差概念。根据2012年美国总统选举的预选投票报告，有49%的选民说会支持奥巴马，47%支持罗姆尼。2%的优势在信息反馈抽样误差的范围内，这就意味着在这场选举中，实际上双方在统计学上是打了平手，两人都没有获胜。但是，我们都知道，这些民意调查会故意煽情说已有一位总统明确地被选出来了。

误差可使用下面给出的公式进行计算。

两个样本均值之间的误差（例如要比较两个样本之间的平均年龄）：

$$z \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

其中， n_1 和 n_2 是样本大小； S_1 和 S_2 是样本标准差。

两个样本比例之间的误差率（例如，通过两个样本的电子邮件群发阅读率的对比，选择哪种主题行的邮件更易被阅读）由以下决定：

$$z \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

其中， n_1 和 n_2 是样本大小； p_1 和 p_2 是两个样本比例（响应比率）； z 是 z -临界值，是由置信水平（CL）所决定的（对于大样本，如果置信水平为95%，那么 $z=1.96$ ）。

相关性

相差系数测量两个或多个随机变量之间的线性统计关系，用 r 表示（皮尔森相关系数）， r 的值在+1和-1之间。 r 值越接近+1或-1，则相关性越高。如果 $r=0$ ，则表示没有相关性。

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

对连续变量的相关系数可以用上面的公式计算，也可以通过绘制X与Y的分布图来直观显示。如果X随着Y增加而增加或随着Y增加而减小，那么X与Y具有相关性。r值可以显示相关性的强度。

让我们用示例表来解释不连续变量事件，以山葡萄酒厂为例（见表A-1）。对于不连续变量（如访问量来源）和汇总形式的变量（通过谷歌点击进入的7224访问量中有4%的成交转化率），通过交叉表能发现相关性（访问量来源与转化率）。如果不同的来源显示不同的转化率（谷歌/点击付费转化率为4%，而必应/点击付费转化率为1%），且该差异在统计学上是有意义的，那么就可以说这两个变量（来源和转化率）之间有一定的相关性。这一相关性直接表明哪些来源能产生不错的成交转化率，如网站winerywedding guide.com或weddingwire.com。营销开支可以据此相应地重新分配。

表A-1 山葡萄酒厂转化率

来源 / 媒介	访问量	转化率 (%)
google/organic	14 440	4
google/cpc	7 224	4
Bing/cpc	4 360	1
(direct) / (none)	3 651	2
myportlandwedding.com/	2 546	5
facebook.com/referral	1 505	0
bing/organic	962	4
yahoo/organic	628	5
stinnocentwine.com/referral	488	4
wineryweddingguide.com/	437	10
weddingwire.com/referral	381	11
eolaamityhills.com/referral	370	0
apps.facebook.com/referral	256	0
vibranttable.com/referral	146	8
google.com/referral	138	7
Others-long tail	2 641	3
Average	40 174	3

Table of Contents

[封面](#)

[赞誉](#)

[献言](#)

[作者简介](#)

[前言](#)

[致谢](#)

[序言](#)

[第一部分 你好，分析](#)

[第1章 要么分析、要么灭亡](#)

[第2章 什么是分析](#)

[第3章 七大分析方法](#)

[第二部分 深入理解](#)

[第4章 BDIR：商业分析的五步框架](#)

[第5章 预测分析](#)

[第6章 数据和分析工具](#)

[第三部分 领导工具包](#)

[第7章 分析和领导者](#)

[第8章 分析竞争](#)

[第9章 分析师领导者的攻略手册](#)

[第10章 付诸行动](#)

[第11章 常见的陷阱](#)

[第四部分 实践中的分析：十个案例研究](#)

[附录 商业统计101](#)