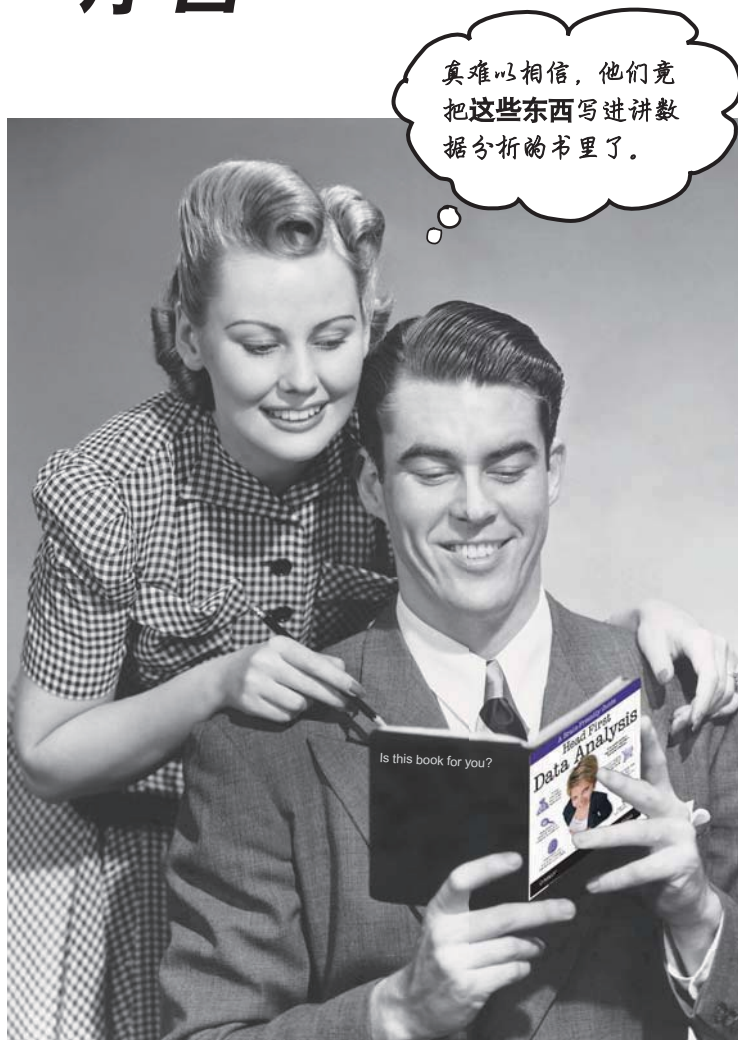


序言



本节回答一个热门问题：“作者为什么非要把这些东西写进一本讲数据分析的书里？”

谁适合阅读本书？

请先回答几个问题：

- 1 你觉得，数据中隐含了无穷的智慧，只要有合适的工具，就能利用这些智慧，对吗？
- 2 你想学习、理解和记忆如何创建靓丽的图形、试验假设条件、进行回归分析或整理混乱的数据，对吗？
- 3 你喜欢笑语喧哗的晚宴甚于枯燥、无聊的学术演讲，对吗？

如果以上问题全部回答“对！”——这本书适合你。

谁该和本书说拜拜？

请先回答几个问题：

- 1 你是一个经验老道的数据分析师，正在调查数据分析领域最前沿的课题，对吗？
- 2 你从未用过Microsoft Excel或OpenOffice calc，对吗？
- 3 你惧怕尝试新事物，宁可上山打虎也不愿标新立异，对吗？你认为要是用拟人的手法叙述控制组和目标函数，技术书籍就难免有失严肃，对吗？

只要有一个问题回答“对！”——你与本书无缘。



[营销部捎话——只要有信用卡
就可以买书哦。]

我们了解你在想什么

“这怎么能是一本严肃的数据分析图书呢？”

“这些图都是用来干嘛的？”

“我真能这样学数据分析吗？”

我们了解你的大脑在想什么

你的大脑渴望新事物。大脑总是不停地搜索、探查、等待不同寻常的事物，它天生如此，这正是你活力的来源。

那么，大脑怎么对待你所碰到的常规、普通、一般的事情呢？——它会竭尽全力阻止这些事情，以免干扰自己真正的工作——记录重要事项。大脑不会费力保存这些琐事；这些琐事从来不会成功地闯过“明显不重要事项”的关卡。

你的大脑如何知道哪件事重要？假想有一天你出门旅行，迎面扑来一只吊睛白额大虎，你的头脑和身体会有什么反应？

神经元发动……情绪激动……化学物质激增

于是，你的大脑知道——

这事绝对重要！记住！

但，想像你是呆在家里，或者是呆在图书馆里，也就是说，是在一个安全、温暖、没有老虎的地方。

你正在复习迎考，要不然就是在努力弄明白一些艰深的技术，你的老板认为花个把星期就能搞定，顶多十天。

唯一的问题是：你的大脑想好好帮你一把，它试图保证不让这种“明显不重要”的内容去破坏珍惜的资源，这些珍稀的资源最好用来保存真正“重大”的事情，像老虎啊，像火灾险情啊，像你绝不该在大学生网站Facebook的网页上贴上那些聚会照片啊。没有什么便当的办法可以告诉大脑“喂，大脑，我对你感激之至，可惜啊，不管这本书多无聊，也不管我的情感地动仪如何纹丝不动，我真的希望你把这些材料都记住。”



我们认为该系列图书的读者都是学习者。

既然要学习，怎样才能学会呢？首先，你得搞懂，然后，切勿遗忘；一字一句硬塞不是办法。根据最新的认知科学、神经生物学及教育心理学研究结果，学习远不仅仅是读书认字。*Head First* 知道怎么让你的脑筋动起来。

下面是部分深入浅出 (*Head First*) 教学原则：

将知识图形化。图形比单调的文字好记得多，可以提高学习效率（记忆学习和转移学习的学习效率最多能提高89%）；图形还能让知识更容易理解，相比将文字放在页脚和下一页，**将文字放在相关图形当中或图形周围**，学习者成功解决相关问题的可能性将成倍增长。



采用对话式的个性化风格。最近的研究表明，要是回避一本正经的语气，代之以对话般的风格，以第一人称平易近人地给学生上课，学生的课后测验成绩最多可提高40%。多讲几个故事，少来一点高谈阔论，语气宜随和。别太郑重其事。想想看，一局笑语喧哗的晚宴和一场演讲，哪一样更让你惦记？

引导读者深入思考：换句话说，除非读者主动调动自己的神经元，否则脑袋里不会发生什么大变化。只有激发读者的兴趣，引起读者的好奇，刺激读者的灵感，读者才能解决问题，得出结论，获得新知识。为此，讲授者要设计各种难题、练习，提出引人深思的提问，还要多让读者做一些让左右脑半球和多种感官都动起来的活动。



牢牢吸引读者的注意力。大家都有这样的体验——“我是真想学，但看完第一页就晕了”。大脑注意的是不同寻常的、有趣的、奇怪的、引人注意的、出人意料的事情。学习一种新颖艰深的技术不一定非得枯燥不可，如果它不是这样乏味，大脑会学得更快。

影响读者的情感。现已知道，人的记忆能力在很大程度上取决于要记忆的内容对情感的影响。我们关心什么，就会记住什么；我们对什么事有感觉，就会记住什么。这里讲的情感并非天灾人祸给人带来的撕心裂肺的伤痛情感，而是惊讶、好奇、感觉有趣、想追根究底之类的情感，以及在猜对一个字谜、在学会别人感觉难以学会的事情或是在意识到自己懂的东西居然比工程部那位开口闭口“我比你更有技术”的张三还多时，油然而生的“我是老大”的感觉。



元认知：对思考的思考

如果真想学东西，而且想学得更快更深入，就要关注自己如何集中注意力。要思考自己的思考方式；研究自己的研究方式。

大多数人在成长过程中都不曾学习元认知和学习理论方面的知识。人们期望我们学知识，但极少有人教我们如何学。

但想象得到，捧着本书的你，的确想学习数据分析知识，同时可能不想花费太多时间。要想利用在本书中读到的知识，就得记住读过的知识，为此必须理解这些知识。为了淋漓尽致地发挥本书或任何书本或学习经验的作用，请管好你的大脑，请管好大脑对待本书的态度。

诀窍在于让大脑把正在学习的新资料当做“正经大事”——对幸福至关重要的大事，像老虎一样重要的大事。若非如此，你就会陷入一场持久战：你竭力要记住新知识，大脑却竭力要把这些新知识踢出去。

既然如此，如何让大脑像对待吃人的老虎一样对待数据分析知识呢？

有两种办法，一种缓慢而乏味，一种迅速而有效。慢办法是简单记忆。你显然明白，只要不停地把同样的东西往大脑里灌，即使是最乏味的知识，也能学会、记牢。只要重复灌的次数足够多，大脑就会想：“这些东西给他的感觉并不重要，但他不停地看这些相同的东西，一遍，一遍，再一遍。因此我猜这些东西肯定很重要。”

快办法是做一切增进大脑活动的事，尤其是不同类型的大脑活动。上一页讲了很多这样的活动，事实证明，这些活动全都能促使大脑以有利于己的方式工作。例如，研究表明，将文字放在文字所描述的图片当中（相反的做法是将文字放在页面中的其他位置，如注释位置或正文位置），会促使大脑努力搞清楚文字和图片之间的关系，进而发动更多神经元。更多神经元发动 = 更有机会让大脑明白某件事值得注意，可能还值得记住。

对话式的写作风格对此很有帮助。人们在与人对话时注意力会更集中，原因是别人期待他们有所表现。令人惊讶的是，大脑不一定会在意“对话”是在人和书之间进行！反之，要是写作风格了无新意，乏味枯燥，大脑的感觉就和在挤满消极听众的屋子里听演讲没什么两样：没必要保持清醒。

不过，图形和对话式风格只是起步……



我们的做法

我们使用**丰富的图片**，这是因为，大脑追逐图像，而非文字。在大脑的活动中，一张图片胜过千言万语。当同时使用图片和文字进行说明时，我们将文字填写在图片当中，当文字出现在它所描述的事物当中时，大脑的工作更有效率；相反，若将说明性文字放在注释或其他正文当中，则无此效果。

我们使用**反复论述法**，即以不同的方式、通过不同的媒介对同一主题进行反复描述，给读者营造**丰富的感受**，目的是让这些主题有更多机会印在大脑的多个区域。

我们以**出人意料的方式**叙述概念和使用图片，因为，大脑追逐新鲜事物；我们在图片和创意中或多或少加入了一些**情感性的内容**，因为，大脑关注情感的生物化学反应。让人有所**感触的东西**更可能让人记住，即使这点感触不过是一丝幽默、一丝惊讶或一丝兴趣。

我们使用**个性化的对话式写作风格**，因为，当大脑认为你是在进行对话而不是在消极地听报告时，就会调整到注意力更集中的状态。即使在**读书时**，大脑也是这个习惯。

我们安排了**80多个活动**，因为，相比读书，在**做事时**，大脑经过调整，能学会和记住更多东西。我们安排的练习有难度，但不会让人束手无策，这正是大多数人愿意做的练习。

我们使用**多种教学风格**，因为，有的人可能喜欢一步一步按顺序来，有的人可能喜欢先看懂大图，还有一些人可能只想看看例子。我们将以多种方式反复讲述相同的主题，不管读者的个人爱好如何，**他们都将因此受益匪浅**。

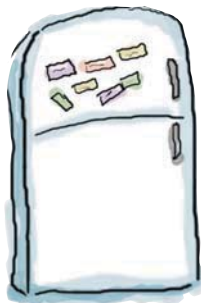
我们安排了让**左右脑半球**分别负责的内容，因为，大脑开动部位越多，就学得越多，记得越多，注意力更持久。由于一侧大脑工作往往意味着另一侧大脑得到休息，左右半脑的分工合作使得长时间学习的学习效率得到提高。

我们还安排了一些**场景和练习**，在场景中展现**不同的观点**，因为，当大脑被迫进行评估和判断时，会调整到深度学习状态。

我们在练习中安排了一些**难点**，即提出一些无法简单回答的问题。因为，你的大脑在不得不**处理某件事情**时，会调整到学习和记忆状态。开动脑筋吧，“光看别人做运动无法让自己**体态健美**”。别担心，我们尽力保证，你努力学习的都是该学的，你不会为了对付一个费解的例子或为了分析一段用词过于晦涩或行文过于简练的段落而**多用一个脑细胞**。

我们以**人物为例**，把人物安排在场景、实例、图片等内容中。至于原因嘛，因为你是人群中的一员啊，你的大脑对人比对事更关注。





把这张图剪下来，贴在冰箱上。

你的任务：征服大脑

我们的工作到此为止，剩下的就看你的了。从下面这些提示出发，顺从大脑的判断，看看哪些对你有用，哪些对你没用，尝试一下新事物吧。

- 1 **慢慢读。理解的内容越多，要记忆的内容越少。**
忌死读。停一停，想一想，碰到书中的提问时，别直接翻看答案；想象真的有人在问你这个问题。强迫自己的大脑想得越深，学会、记住的概率就越大。
- 2 **自己做练习，自己记笔记。**
我们安排了练习和笔记，但是，要是我们替你完成，就像让别人替你锻炼身体一样；只动眼不动手也不可取，**要动动笔**。大量证据证明，学习时的身体动作能提高学习效率。
- 3 **阅读“世上没有傻问题”部分**
世上没有傻问题。这些问题并非可看可不看，**这是核心内容的组成部分！请勿忽略。**
- 4 **请将下面这段话作为最后一段床头阅读文字，或起码作为最后一段高深的床头阅读文字。**
有一部分学习过程（尤其是短暂记忆转变为长期记忆的过程）发生在放下书本之后，大脑需要有自己的时间进行更多处理。如果在这段处理时间内学新东西，将会丢失一些刚学会的东西。
- 5 **开口大声讨论。**
说话会刺激大脑的其他部分。如果你正在努力理解一些知识，或者正在努力增加以后记住这些知识的概率，请大声说出这些知识。还有一种更好的做法，试着向别人大声解释这些知识。你会学得更快，可能还会发现一些阅读时不曾发现的名堂。
- 6 **大量喝水。**
充沛的体液会让大脑处于最佳工作状态，脱水（早在感到口渴前就会发生）则会让认知功能下降。
- 7 **聆听大脑的声音。**
留意你的大脑是否超负荷工作。若你发现自己开始心不在焉，或者刚刚读过的东西转眼忘记，就该休息。一旦过了某个学习点，哪怕拼命塞，也无法提高学习效率，反而有可能影响学习。
- 8 **找到感觉。**
大脑需要知道事情是否重要。让自己融入各种场景，为照片设想旁注，就连抱怨一个并不好笑的玩笑，也比什么感觉都没有强。
- 9 **勤加练习！**
学会数据分析的唯一办法就是勤加练习，这正是本书的要求。数据分析是一门技术，精于此道的唯一办法就是大量实践。本书将给你带来大量实践机会：每一章中都有一个等待你解决的问题，千万别跳过这些问题不看——大量学习都发生在解决问题的过程中。我们为每一个问题提供了答案，要是卡了壳（有些细微之处很容易给人带来麻烦），别不敢看！不过，请尽量先解决问题再看答案，务必让你的办法行之有效，然后才继续看书中的下一部分内容。

自述

本书是经验之谈，并非参考书籍，我们故意抽掉了会妨碍讲述书中相关知识的東西。本书对你已经见识过和学习过的知识作了一些假设，因此第一次通读本书的时候，需要从头读起。

本书并非软件工具指导书。

许多以“数据分析”为题的图书都是顺着Excel函数表把认为和数据分析有关的部分一路讲下去，然后针对每个函数给几个实例。但《深入浅出数据分析》讲的是如何成为数据分析师，尽管你在本书中会学到相当多的软件工具，但它们不过是手段而已，目的是学习如何进行出色的数据分析。

我们希望你懂得如何使用基本的电子表格公式。

用过电子表格的SUM求和公式吗？要是没用过，你可能先要突击一下才能开始学习本书。尽管许多章节根本不要求使用电子表格，但其他有此要求的章节却假定你会使用各种公式。要是熟悉SUM工具，那么你基础不错。

本书超越统计学。

本书充满统计知识，作为数据分析师，你应该尽量多掌握一些统计知识，读完《深入浅出数据分析》之后，最好再读一读《深入浅出统计学》（*Head First Statistics*）。不过，数据分析不仅涵盖统计学，还牵涉许多其他领域，本书中选用的非统计题材主要用于讲解来源于现实生活的具体、实用的数据分析经验。

活动并非可做可不做。

练习和活动不是点缀，而是本书的核心组成部分。这些练习和活动有的是为了帮助记忆，有的是为了帮助加深理解，还有的是为了帮助应用所学知识，切勿忽略。唯一的

例外是填字游戏，这不是非做不可，不过，填字游戏会让大脑得到在不一样的语境下思考所学到的单词和术语的大好机会。

反复论述是刻意而重要的安排。

深入浅出系列图书有一个明显特色：我们希望你真正掌握学到的知识，我们希望你在看完本书的同时就记住学到的知识。大多数参考书都不把记忆和回忆当做一个目标，但本书的目标是学会，所以，常常会看到同一概念多次出现。

本书意犹未尽。

我们乐于看到你在书籍合作网站上找到更多实用而有趣的资料，下列网站可为你提供这些资料：

<http://www.headfirstlabs.com/books/hfda/>

“动动脑”练习没有答案。

有一些“动动脑”练习没有标准答案；另有一些练习可以参考“动动脑”活动的学习经验部分判断自己的答案是否正确，以及在什么情况下会正确。部分“动动脑”练习给出了提示，为你指明正确方向。

技术顾问组

Eric Heilman



Tony Rose



Bill Mietelski



技术顾问：

Eric Heilman，美国乔治敦大学沃尔什外交学院优秀毕业生，国际经济学学位。在哥伦比亚特区读大学期间，曾在美国国务院和白宫国家经济委员会工作。他在芝加哥大学完成经济学毕业论文，目前在位于美国马里兰州贝塞斯达（Bethesda）的乔治敦大学预备学校任统计分析和数学教师。

Bill Mietelski，软件工程师，三度担任深入浅出（*Head First*）技术顾问。他急不可待地想给自己的高尔夫技术做个数据分析，好在球场上领风骚。

Anthony Rose，在数据分析领域从业近十年，目前任Support Analytics公司总裁、数据分析及图表顾问。Anthony拥有财务与管理专业工商管理硕士学位，他对数据分析的热爱由此开始。工作之余，他常常出现在马里兰州哥伦比亚市的高尔夫球场上，陶醉在好书中，品味着美味的葡萄酒，或者和年幼的女儿们及迷人的妻子一起消磨时光。

致谢

我的编辑：

Brian Sawyer，一位不可思议的编辑。和Brian一起工作就像和舞蹈家共舞，各种各样重要的工作纷至沓来，虽令人不十分理解，看上去却很不错，让人干得兴高采烈。我们的合作振奋人心，他的支持、反馈和创意是无价之宝。

O'Reilly团队：

Brett McLaughlin一开始就看到了这个项目的未来，引领项目走过艰难岁月，始终如一地支持项目。Brett孜孜不倦地强调你对深入浅出（*Head First*）书籍的体验，让人备受鼓舞。他运筹帷幄。

Karen Shaner提供后勤支持，在剑桥寒冷的清晨给我们带来很多快乐。**Brittany Smith**贡献了一些非常棒的图形元素，供我们反复使用。

给我启示的睿智者：

本书有大量出色的创意，许多创意在以“数据分析”为题的书籍中颇不常见，但这些创意很少是我个人的独创。我从Dietrich Doerner、Gerd Gigerenzer、Richards Heuer、Edward Tufte等超级智星的作品中汲取了大量经验。把他们的作品统统读一遍吧！“反查”（anti-resume）这个创意出自Nassim Taleb的《黑天鹅》（真希望他出第二部，带来更多创意）；**Richards Heuer**好心地给我回信讨论本书，还给我出了很多有用的主意。

朋友与同事：

感谢**Lou Barr**为本书提供知识产权、职业道德、逻辑学及美学支持；**Vezen Wu**给我讲解关系模型；**Aron Edidin**在我大学求学期间曾赞助我学习一门超棒的情报分析课；我的牌友**Paul**、**Brewster**、**Matt**、**Jon**和**Jason**给我上了关于均衡使用启发法和最优化决策法的昂贵一课。

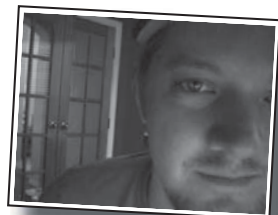
离开这些人我没法活：

技术顾问组工作出色，他们揪出成堆的错误，提出大量建议，给予我巨大支持。在本书撰写过程中，我对一位心思缜密的统计师——我的朋友**Blair Christian**依赖甚深，书中每一页都能看到他的影子。谢谢你为我做的一切，Blair。

我的家人**Michael Sr.**、**Elizabeth**、**Sara**、**Gary**和**Marie**给了我巨大的支持，尤其要感谢我的妻子**Julia**的坚定支持，她是我的一切。谢谢我的全家！



Brian Sawyer



Brett McLaughlin



Blair 与 Niko Christian



Julia Burch

1 数据分析引言

分解数据



数据无处不在。

如今，不管是不是自称数据分析师，人人都得处理堆积如山的数据。熟谙一切数据分析技术方法的分析者会比其他人技高一筹：他们知道如何处理所有的数据材料，如何将原始数据转变成推进现实工作的妙策，如何分解和构建复杂的问题和数据集，进而牢牢把握工作中的各种问题的要害。

Acme化妆品公司需要你出力

这是你走上数据分析师岗位的第一天，刚刚收到了首席执行官发来的销售数据，需要查阅一下。数据反映了Acme公司旗舰产品——貌洁超强保湿霜的销售情况。

过去六个月的销量有哪些变化？

这六个月的实际总销量与目标总销售量相比如何？

	9月	10月	11月	12月	1月	2月
总销量	\$5 280 000	\$5 501 000	\$5 469 000	\$5 480 000	\$5 533 000	\$5 554 000
目标销量	\$5 280 000	\$5 500 000	\$5 729 000	\$5 968 000	\$6 217 000	\$6 476 000
广告费	\$1 056 000	\$950 400	\$739 200	\$528 000	\$316 800	\$316 800
社会网络费	\$0	\$105 600	\$316 800	\$528 000	\$739 200	\$739 200
单价（每盎司）	\$2.00	\$2.00	\$2.00	\$1.90	\$1.90	\$1.90

看出Acme的开支规律了吗？

你觉得这些单价有什么变化？为什么下降了？

看看这些数据，不必抽丝剥茧——只要放慢速度就行。

看出什么了吗？表格让你对Acme的业务了解了多少？对Acme的貌洁超强保湿霜了解了多少？

优秀的分析师总想看到数据。

首席执行官希望数据分析师帮他提高销量

他希望你“帮他分析分析”。

这要求很含糊，不是吗？听起来挺简单，可你的工作会那么顺吗？不错，他希望提高销量；不错，他认为这些数据中有些东西能帮助实现这个目标。可到底是哪些东西呢？怎么帮呢？

这位就是首席执行官。

欢迎加入我们的团队。看看我们的数据，给我分析分析，说说我们该如何提高销量。等你的结论。

他这话是什么意思？

动动脑

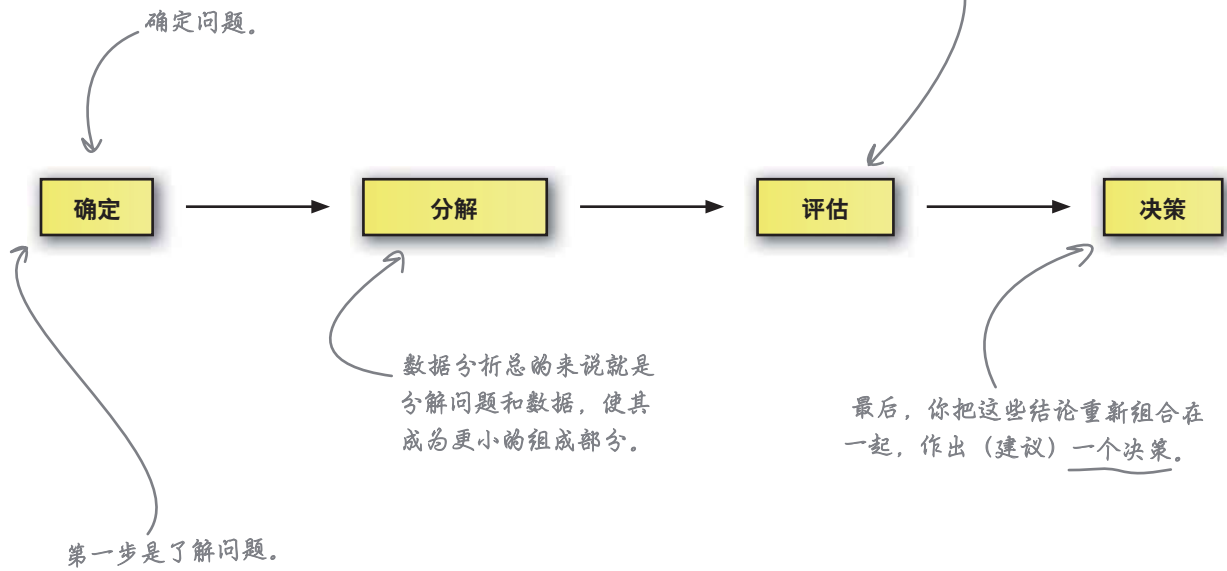
想想首席执行官主要想从你这里得到什么，同时思考这个问题：做数据分析到底意味着什么？

数据分析就是仔细推敲证据

数据分析这个词涵盖大量形形色色的工作和大量形形色色的技巧。就算有人明白告诉你她是数据分析师，你依然无法确定她的**专长**。

你可能会打赌说她懂 Excel，但仅此而已！

但是，所有优秀的分析师，无论专长及目标如何，都会在工作过程中按顺序执行下面这个**固定基本流程**，同时通过经验数据来仔细推敲各种问题。



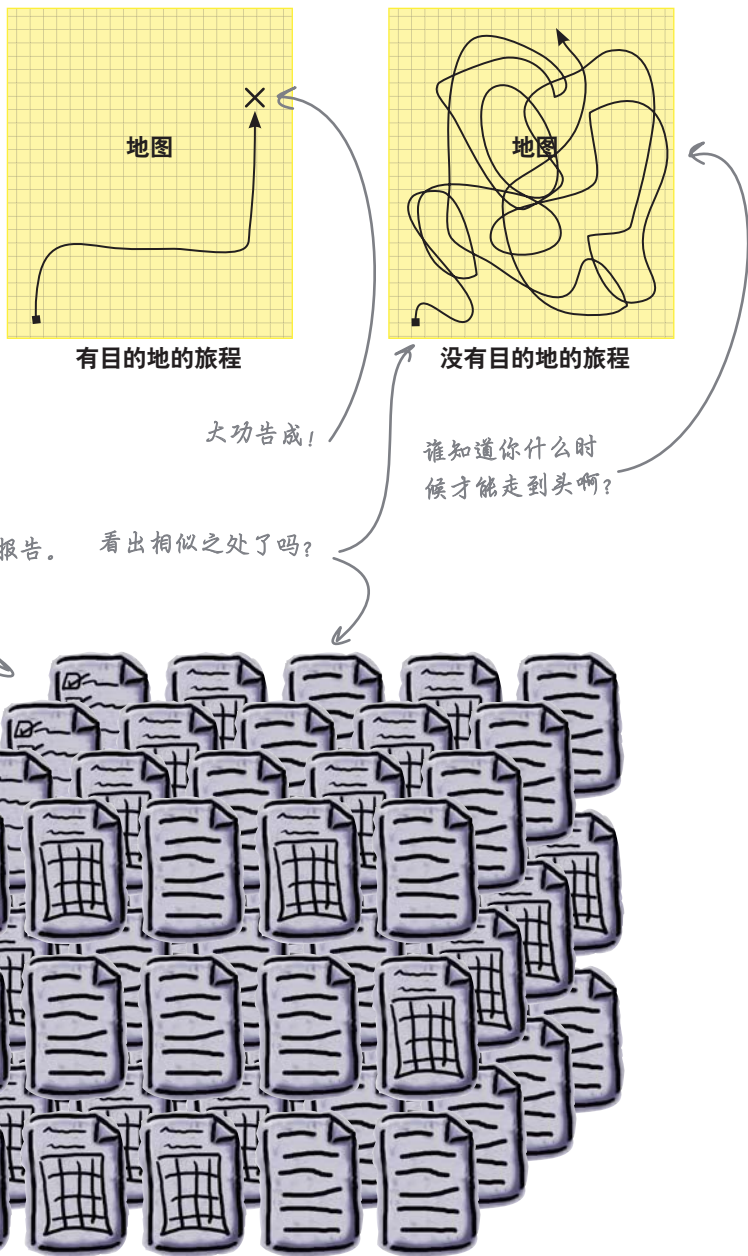
在本书的每一章中，你会一次又一次地按顺序执行这些步骤，很快，这些步骤就会完全成为你的第二本能。

所有的数据分析师最终都会被打造成能作出**更好决策**的人才，你要学的就是在浩如烟海的数据中洞察先机，作出更好决策。

确定问题

未明确确定自己的问题或目标就进行数据分析就如同未定下目的地就上路旅行一样。

当然，您可能会碰到一些有意思的现象，有时还可能盼着能兜来兜去地撞上点好东西，但是，谁会说你将有所发现？



见过长达百万页、图表不计其数的分析报告吗？

偶尔，分析师的确会需要几百张纸或一小时的幻灯片来阐述一个观点，但如此一来，分析师常常不够注重自己要解决的问题，他们抛给别人一些信息，借此推卸自己解决问题和建议决策的义务。

有时情况更糟糕：问题根本没有确定下来，而且分析师不想让别人意识到他只是在数据中兜圈子。

你如何确定问题？

客户将帮助你确定问题

客户是分析结果的服务对象。你的客户可能是你的上司、你所在公司的首席执行官，或甚至就是你本人。

客户将根据你的分析作决策，你需要尽量从他那里多了解一些信息，才能确定问题。

本文中这位首席执行官想提高销量，但这只是最初答案。你需要更多更确切地摸清他的心思，才能拟定一个能够解决问题的分析方案。



要点

你的客户可能：

- 相当了解或不甚了解自己的数据
- 相当了解或不甚了解自己的问题或目标
- 相当了解或不甚了解自己的业务
- 目标明确或优柔寡断
- 头脑清醒或稀里糊涂
- 富有直觉或善于分析

阅读本章时，请注意页尾这些图示，它们代表你所处的各个分析阶段。

你对客户了解越深，你的分析越有可能派上用场。



世上没有傻问题

问： 我总是在数据里兜来兜去。您是说我得先在脑子里有些特定的目标，才能哪怕只是过一眼我的数据？

答： 没必要先在脑子里形成问题才去浏览数据。但要记住，**仅仅过一眼**并不是数据分析。数据分析总的来说就是认清问题，以及继而解决问题。

问： 我听说过探索性数据分析，就是从数据中找出一些可能想进一步进行评估的点子。这种数据分析方法中并没有什么“问题确定”步骤！

答： 确实有这种分析方法。在探索性数据分析中，问题就是要找到一些值得进行测试的假设条件，这完全是个具体问题。

问： 很好。给我多讲讲对自己的问题不甚了解的客户吧。那种人也需要数据分析师吗？

答： 当然！

问： 听起来似乎那种人更需要专业帮助。

答： 的确如此，优秀的数据分析师帮助客户思考自己的问题；他们不会等着客户告诉他们该做什么。要是有人能够向客户指出他们毫无察觉的问题，客户会真心诚意地感谢此人。

问： 听起来挺傻。谁想多搞出些问题？

答： 聘用数据分析师的人认为，具备分析技能的人能够改善他们的业务。有些人把问题视为机会，而向客户指出如何发现机会的数据分析师则能让客户赢得竞争优势。

动动笔



总体问题是我们需要提高销量。为了更好地摸清这位首席执行官的真正意图，你想问这位首席执行官什么问题呢？写出5个问题。

1

.....

2

.....

3

.....

4

.....

5

.....

Acme公司首席执行官给了你一些反馈

你提出的问题
可能与此不同。

这份邮件对你的问题进行了答复。其中有很多知识点……

这是一些能让首席执行官确定你的分析目标的典型提问。

不停地问“是多少”，使你的各种目标和确信观点得到量化。

预见客户的想法，他一定会关心竞争对手的情况。

对某些数据感到好奇吗？问吧！

发件人：Acme化妆品公司首席执行官
收件人：Head First
主题： 回复：确定问题

您希望销量提高多少？

我需要让销量重新回到目标值，你可以在表格里看到这个目标值。我们所有的预算都是按照这个目标值确定的，如果达不到目标值，我们就会有麻烦。

您觉得我们怎样才能办到呢？

哦，想办法是你的事。不过策略是要让更多的人买更多的产品，我所说的“人”是十岁出头的少女消费者（11~15岁）。你们要通过对这样那样的产品进行市场营销来提高销量。你是数据人才，想想办法！

您觉得销量提高多少是可行的？目标销量合理吗？

这些少女消费者手头宽裕——做保姆的工钱，父母给的零花钱，等等。

我想，通过向她们推销貌洁超强保湿霜，销量可以扶摇直上。

我们的竞争对手销量如何？

我没有可靠的数字，但在我印象中他们打算超过我们。我得说他们的保湿霜总收入要比我们高50%~100%。

广告和社会网络营销预算是怎么回事？

我们正在尝试一些新手段，总预算是第一个月收入的20%。过去这笔预算全部用在广告上，但我们现在会分出一些用于社会网络上。要是广告费一直维持这个水平，我真不敢想象会有什么结果。

确定

分解

评估

决策

把问题和数据分解为更小的组块

数据分析的下一步就是把从客户那里了解到的问题和手头的数据放在一起，把这些问题分解为颗粒级的小问题，让它们在分析时发挥最大作用。



将大问题划分为小问题

你需要将问题划分为可管理、可解决的组块。你面对的问题常常含糊不清，例如：

“我们如何提高销量？”

“我们最好的客户希望我们给他们什么？”

“哪种促销方式最有可能产生效果？”

“我们的广告做得怎么样了？”

你无法直接回答大问题。但是，通过回答从大问题分解出来的小问题，你就可以找到大问题的答案。

回答小问题，解决大问题

将数据分解为更小的组块

数据的处理也是如此。人们无意告诉你你所需要的精确答案的量化值，你必须自己提炼重要的因子。

如果你拿到的是汇总情况，就像Acme给你的那些数据，你就会想知道哪些因子对你至关重要。

如果你拿到的是原始数据表，你就会想对这些因子进行汇总，让数据更有用。

	9月	10月	11月	12月	1月	2月
总销量	\$5 280 000	\$5 501 000	\$5 469 000	\$5 480 000	\$5 533 000	\$5 554 000
目标销量	\$5 280 000	\$5 500 000	\$5 729 000	\$5 968 000	\$6 217 000	\$6 476 000
广告费	\$1 056 000	\$960 400	\$739 200	\$528 000	\$316 800	\$316 800
社会网络费	\$0	\$105 600	\$316 800	\$528 000	\$739 200	\$739 200
单价（每盎司）	\$2.00	\$2.00	\$2.00	\$1.90	\$1.90	\$1.90

12月目标销量 \$5 968 000

11月单价 \$2.00

这些可能就是你要查看的组块。

稍后会详细解释这些时髦的行话！

让我们给分解工作来个特写……

现在再来看看了解到的情况

让我们从数据开始。你手头有一份Acme销售数据汇总，尝试分解最重要因子的最好起步办法是找出高效的比较因子。

找到感兴趣的比较对象，分解汇总数据。

10月份的总销量与目标销量相比如何？

1月份的总销量与2月份的总销量相比如何？

	9月	10月	11月	12月	1月	2月
总销量	\$5 280 000	\$5 501 000	\$5 469 000	\$5 480 000	\$5 533 000	\$5 554 000
目标销量	\$5 280 000	\$5 500 000	\$5 729 000	\$5 968 000	\$6 217 000	\$6 476 000
广告费	\$1 056 000	\$950 400	\$739 200	\$528 000	\$316 800	\$316 800
社会网络费	\$0	\$105 600	\$316 800	\$528 000	\$739 200	\$739 200
单价（每盎司）	\$2.00	\$2.00	\$2.00	\$1.90	\$1.90	\$1.90

广告费和社会网络费怎样随着时间变化而相对变化？

单价的降低与总销量的变化一致吗？

进行有效的比较是数据分析的核心，本书通篇都在讲述这个工作。

在这个案例中，您想通过比较各项汇总数据在脑子里形成一个概念，即Acme公司的貌洁超强保湿霜业务是如何开展的。



你已经确定了问题：想出提高销量的办法。但通过这个问题几乎无法知道别人对你的工作期望，于是你从首席执行官那里搞到了大量有用的言论。

这些言论给出了关于如何开展化妆品业务的重要基准假设。希望首席执行官关于这些假设的看法是正确的，因为它们将是分析的基础！首席执行官的论点里最重要的有哪些呢？

这个评论本身就是一类数据。它的哪几个部分最为重要呢？

这是一些相关问题

哪些东西最有用？

发件人：Acme化妆品公司首席执行官
 收件人：Head First
 主题： 回复：确定问题

您希望产量提高多少？

我需要让销量重新回到目标值，您可以在表格里看到这个目标值。我们所有的预算都是按照这个目标值确定的，如果达不到目标值，我们就会有麻烦。

您觉得我们怎样才能办到呢？

哦，想办法是你的事。不过策略是要让更多的人买更多的产品，我所说的“人”是十岁出头的少女消费者（11~15岁）。你们要通过对这样那样的产品进行市场营销来提高销量。你是数据人才，想个办法吧！

您觉得销量提高多少是可行的？目标销量合理吗？

这些豆蔻年华的少女消费者手头宽裕——做保姆的工钱，父母给的零花钱，等等。

我想，通过向她们推销貌洁超强保湿霜，销量可以扶摇直上。

我们的竞争对手销量如何？

我没有可靠的数字，但在我印象中他们打算超过我们。我得说他们的保湿霜总收入要比我们高50%~100%。

广告和社会网络营销预算是怎么回事？

我们正在尝试一些新手段，总预算是第一个月收入的20%。过去这笔预算全部用在广告上，但我们现在会分出一些用于社会网络上。要是广告费一直维持这个水平，我真不敢想会有什么结果。



根据你所得到的分析数据，总结一下客户确信无疑的观点以及你的想法。分析以上邮件和你的数据，将它们分解为能够描述你的现状的更小的组块。

客户确信无疑的观点

你对数据的想法

- ①
- ②
- ③
- ④

- ①
- ②
- ③
- ④



动动笔 解答

清点一下你和客户确信无疑的观点。你发现了什么？

客户确信无疑的观点

你写的答案可能略有不同。

- 1 貌洁超强保湿霜的消费者是处于豆蔻年华的少女（具体说是11~15岁），她们基本上是唯一消费群。
- 2 Acme试图重新分配广告费和社会网络费，但迄今为止，这个新办法是否成功尚未可知。
好……现在的人都这么做。
- 3 我们看出产品在少女消费者中的销售潜力是无限的。
- 4 Acme的竞争对手极为强大。
这值得记住。

你对数据的想法

- 1 2月份的销量与上年9月份相比略有上升，但成绩平平。
大问题
- 2 销量与目标相去甚远，而且从11月份开始南辕北辙。
- 3 降价看来无助于销量达标。
接下来他们该怎么办？
- 4 削减费用可能会影响Acme的销量达标能力。

您已经成功地将问题分解为更小、更好管理的组块。

现在是时候对这些组块进行更具体的评估了……



评估组块

好戏上场了。你知道需要想办法，你知道哪些数据组块能让你做到这一点。现在，仔细、专注地看看这些组块，形成自己的判断。



正如分解时一样，评估分解组块的关键就是比较。

通过对这些因子进行相互比较，你看出了什么？

针对问题的观察结果

貌洁超强保湿霜的消费者是处于豆蔻年华的少女消费者（具体是11~15岁）。她们基本上是唯一的消费群体。

Acme正在尝试增加用于扩展社会网络的广告费，但迄今为止，新做法是否成功尚未可知。

我们看出产品在少女消费群体中的销售潜力是无限的。

Acme的竞争者极为危险。

想象一下！

抽取两个因素，依次阅读。

你看出什么了？

针对数据的观察结果

2月份的销量与上年9月份的销量相比略有上升，但尚属持平。

销量与目标相去甚远。

看来降价无助于销量达标。

削减费用可能会影响Acme的销量达标能力。

你几乎拥有所有合适的组块，唯独缺少重要的一块……

分析从你介入的那一刻开始

让自己介入分析的意思是作出自己的明确假设，并且以自己的信用为自己的结论打赌。

无论你是否正在构建复杂的模型还是在作简单的决策，数据分析就是你的一切：你的信念，你的判断，你的信用。

只要你在分析中明白地展现自己，成功就更有希望。

本人介入

给你带来的好处

你将知道要在数据中发现什么。
你将避免作出过头的结论。
你将对工作成败负责。

给客户带来的好处

客户将更尊重你的判断。
客户将理解到你的判断是有局限性的。

本人不介入

给你带来的坏处

你将无法追踪基准假设如何影响你的结论。
你将成为逃避责任的懦夫！

给客户带来的坏处

客户将不会信任你的分析，因为他不知道你的动机和动力。
客户可能会产生客观的错觉，或变得冷漠而理性。

在撰写最终报告的时候，一定要提到你自己，这样客户才知道你的结论出自何处。

哎呀！你不想卷入这种问题吧。



提出建议

作为数据分析师，你的工作就是让自己和客户仔细研究你对数据的评估，洞察先机，从而有能力作出更好的决策。



为了实现这个目的，你必须将自己的设想和判断以合适的格式整合起来，供客户拮取。

也就是说，你的作品要能简单则简单，但不可简单过头！你的工作是确保自己的意见传达到位，让人们根据你的意见作出正确的决策。

除非将分析形成报表供制定决策，否则，分析将毫无用处。

你提交给客户的报告要以得到客户理解、鼓励客户以数据为基础作出明智的决策为重点。

动动笔



看看你在前面几页搜集到的信息。

你建议Acme如何提高销量？为什么？

.....

.....

.....

.....

.....

.....

.....

.....

报告写好了

Acme化妆品公司 分析报告

背景

貌洁超强保湿霜的客户是少女消费者（具体是11~15岁）。她们基本上是唯一的客户群。Acme正在尝试增加用于扩展社交网络的广告费，但迄今为止，这个新做法是否成功尚未可知。我们看出产品在少女消费者中的销售潜力巨大。Acme的竞争对手极为危险。

数据解说

2月份的销量与上年9月份相比略有增长，但仍属持平。销量与目标相去甚远，削减广告费用可能会影响Acme的销售达标能力。降价看来无助于销量达标。

建议

销量相对目标下降可能与广告费相对从前的广告费下降有关。没有充分的证据让我们相信社交网络建设已如我们所愿取得成功。我将把广告费重新调整到9月的水平，看看少女消费者是否有反应。针对少女消费者做广告是让总销售额重新达到销售目标的手段。

这是我们一开始从首席执行官那里得到的材料。

这是你的分析大餐。

你的结论可能与此不同。

在报告中写下自己和客户的假设是个不错的办法。

用简单的图形解说自己的结论。

首席执行官会怎么想呢？



首席执行官欣赏你的工作

干得好，我完全被说服了。我立刻就下订单多投广告，我迫不及待地想知道结果！

你的报告简练、专业、直截了当

报告说清楚了首席执行官的需求，甚至比首席执行官本人说得更清楚。

你审视数据，通过首席执行官把事情弄得更明白，把首席执行官确信的观点和你自己对数据的理解相比较，然后提出决策建议。

干得好！



你的建议将给Acme的业务带来哪些影响？

Acme的销量会上升吗？

一则新闻

表面上看起来是一篇正面报道。

数据邦商务时报

貌洁保湿霜在少女消费者市场完全饱和

据我报化妆品行业独立分析师报告，少女消费者保湿霜市场已经完全被Acme公司保湿霜旗舰产品“貌洁”占据，据《数据邦商务时报》调查，95%的少女消费者称“非常频繁”使用貌洁保湿霜，通常每天两次以上。

当我报记者告诉Acme首席执行官这个调查结果时，他非常惊讶。“我们承诺以

平易近人的价格给少女消费者最奢华的保湿体验”，他说，“得知貌洁在少女消费者中如此走红我很高兴，希望以后由我们的数据分析部门告诉我这些消息，而不是报社。”

Acme在市场上的实际竞争对手——竞争化妆品公司回应了记者采访：“我们基本上已经撤出了少女消费者市场。我们雇来扰乱市场的少女消费者受到了朋友们的嘲笑，因为据说使用了廉价、低档的产品。貌洁品牌太强了，和他们

竞争是在浪费营销费用。菩萨保佑，给貌洁来个打击，比如，让他们的代言人在镜头里被逮到……

这对你的分析有意义吗？

表面上来这对Acme是个好消息，但是，如果市场已经饱和，再多投广告可能就不会有太大效果。

幸好我接到了这个电话，我取消了少女消费者市场广告。马上给我再搞一个有用的方案吧。

很难想象少女消费者市场广告会有效。要是绝大部分少女消费者每天都用貌洁保湿霜，而且用两次以上，销量还有机会提高吗？

你需要寻找别的机会提高销量，但首先需要搞清楚你的分析有何差池。



考考你

你在分析过程中得到了一些**错误的或不完整的信息**，使你对上述有关少女消费者的情况把握不准。是哪些信息不完整呢？

首席执行官确信的观点 让你误入歧途

这是首席执行官嘴里的貌洁销售情况：

首席执行官确信的貌洁销售情况

貌洁超强保湿霜的消费者是处于豆蔻年华的少女消费者（具体说是11~15岁）。她们基本上是唯一的消费群。

Acme试图重新分配广告费和社会网络费，但迄今为止，这个新办法是否成功尚未可知。

我们看出产品在少女消费者中的销售潜力是巨大的。

Acme的竞争对手极为危险。

这是一种心智模型……

看看这些确信观点与数据的吻合情况，二者一致还是矛盾？所描述的内容有差别吗？

	9月	10月	11月	12月	1月	2月
总销量	\$5 280 000	\$5 501 000	\$5 469 000	\$5 480 000	\$5 533 000	\$5 554 000
目标销量	\$5 280 000	\$5 500 000	\$5 729 000	\$5 968 000	\$6 217 000	\$6 476 000
广告费	\$1 056 000	\$950 400	\$739 200	\$528 000	\$316 800	\$316 800
社会网络费	\$0	\$105 600	\$316 800	\$528 000	\$739 200	\$739 200
单价（每盎司）	\$2.00	\$2.00	\$2.00	\$1.90	\$1.90	\$1.90

数据没有体现少女消费者市场的任何情况。他假定少女消费者是产品的唯一购买者，而且少女消费者有能力购买更多的貌洁保湿霜。

看了上述新闻后，你可能想重新评估首席执行官确信的这些观点。

我们又回到了起点！



你对外界的假设和你确信的观点就是你的心智模型

在这个案例中，心智模型带来了问题，如果新闻报导是真实的，那么首席执行官关于少女消费者市场的确信观点就是错误的，而这些确信观点正是你用来解释数据的模型。

现实世界非常复杂，因此我们用心智模型来理解现实。你的大脑就像一个工具箱，只要有新信息进来，大脑就会拿出一个工具利用这个新信息。

心智模型可能是一些与生俱来的先天禀赋，也可能是后天学会的理论，不管是哪种情况，都会大大影响你对数据的解释。



心智模型有时助益良多，有时带来麻烦。本书就是你的妥善利用心智模型速成班。

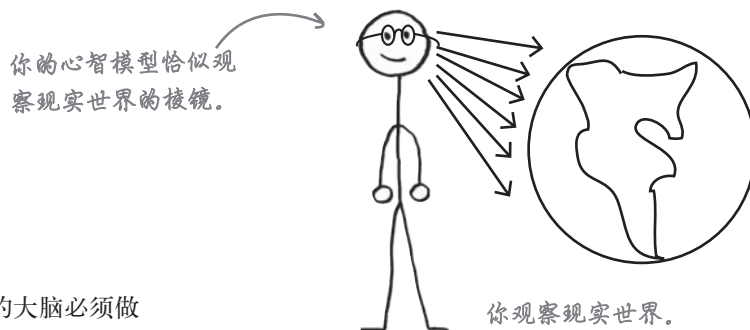
重中之重是明确心智模型，并且像对待数据一样严肃认真地对待心智模型。

务必尽量明确你的心智模型。



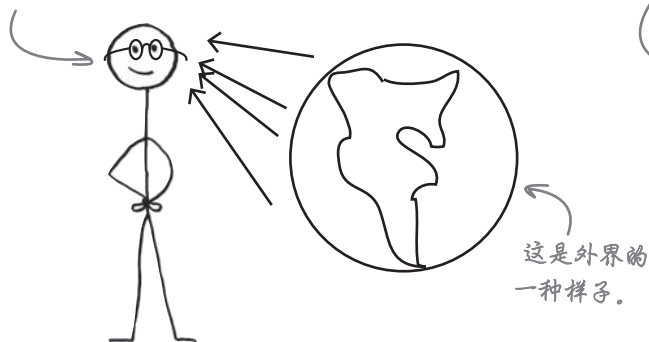
统计模型取决于心智模型

心智模型决定你的观察结果，是你观察现实的棱镜。



你无法看到一切，因此你的大脑必须做出选择，以便集中注意力，这就是所谓的心智模型大大决定观察结果。

一种心智模型会让你注意到外界的一些特点……



……另一种心智模型会让你注意到外界的另一一些特点。



如果你了解自己的心智模型，那么你会发现重点、开发最相关最有用统计模型的可能性就更大。

你的统计模型取决于你的心智模型，如果用了错误的心智模型，分析就会胎死腹中。

最好使用正确的心智模型！





动动笔

让我们再次审视这些数据，想一想，有没有其他的心智模型适合这些数据。

	9月	10月	11月	12月	1月	2月
总销量	\$5 280 000	\$5 501 000	\$5 469 000	\$5 480 000	\$5 533 000	\$5 554 000
目标销量	\$5 280 000	\$5 500 000	\$5 729 000	\$5 968 000	\$6 217 000	\$6 476 000
广告费	\$1 056 000	\$950 400	\$739 200	\$528 000	\$316 800	\$316 800
社会网络费	\$0	\$105 600	\$316 800	\$528 000	\$739 200	\$739 200
单价（每盎司）	\$2.00	\$2.00	\$2.00	\$1.90	\$1.90	\$1.90

发挥你的创造力！



- 1 列出一些假设情况：若貌洁保湿霜的确是少女消费者喜爱的润肤品，则假设成立。

.....

.....

.....

.....

.....

- 2 列出一些假设情况：若貌洁保湿霜处于在竞争中失去顾客的危境地，则假设成立。

.....

.....

.....

.....

.....



动动笔 解答

你刚才用新眼光观察了汇总数据。**不同**的心智模型该如何与之契合呢？

	9月	10月	11月	12月	1月	2月
总销量	\$5 280 000	\$5 501 000	\$5 469 000	\$5 480 000	\$5 533 000	\$5 554 000
目标销量	\$5 280 000	\$5 500 000	\$5 729 000	\$5 968 000	\$6 217 000	\$6 476 000
广告费	\$1 056 000	\$950 400	\$739 200	\$528 000	\$316 800	\$316 800
社会网络费	\$0	\$105 600	\$316 800	\$528 000	\$739 200	\$739 200
单价（每盎司）	\$2.00	\$2.00	\$2.00	\$1.90	\$1.90	\$1.90

① 列出一些假设情况：若貌洁保湿霜的确是少女消费者喜爱的润肤品，则假设成立。

少女消费者几乎把所有保湿霜预算都给了貌洁。

Acme需要开发新的貌洁保湿霜市场才能提高销量。

这里歌舞升平。

貌洁保湿霜没有匹敌的竞争对手，它是迄今为止最好的产品。

社会网络是目前售卖产品最经济有效的方式。

少女消费者愿意在保湿霜上花更多的钱。

② 列出一些假设情况：若貌洁保湿霜处于在竞争中失去顾客的危险境地，则假设成立。

少女消费者改用新的保湿霜产品，Acme需要夺回失地。

貌洁保湿霜被认为“不够酷”，是“给傻子用的”。

这里刀光剑影。

“干”皮肤在年轻人中日渐流行。

社会网络营销是个无底洞，我们需要重投广告的怀抱。

提高貌洁的价格将损失市场份额。

客户拥有完全错误的心智模型并不稀奇，忽略心智模型中有可能最重要的部分其实也是家常便饭……



心智模型应当包括你不了解的因素

一定要指出**不确定**因素，只要能明确不确定因素，你就会小心防范并想办法填补知识空白，继而提出更好的建议。

考虑不确定因素及盲点会让人感觉不爽，但回报显著。这种“反查”方法会揭示出**未知**信息，而不是已知信息，例如，你要雇用**一个舞蹈家**，他不会跳的舞可能比会跳的舞更让你感兴趣。

反思这些会
感觉痛苦。

人们往往在雇用职员
后才发现他们有些事
做不来，可为时已晚。

数据分析也是如此，了解自己的知识缺陷非常重要。

未雨绸缪方能防备不测风云。

Head First反查表

我所没有的经历：

被捕
吃小龙虾
骑自行车
铲雪

我不知道的事情：

圆周率前50位数
我今天用手机打了多少分钟电话
生命的意义

我不知道该怎么做的事情：

做法式面包
跳恰恰舞
弹吉他

我没读过的书：

《红楼梦》
《三刻拍案惊奇》

动动笔



为了搞清楚首席执行官**不知道**的事情，你会问哪些问题？

.....

.....

.....

.....

.....

.....

.....

首席执行官承认 自己有所不知

发件人：Acme化妆品公司首席执行官

收件人：Head First

主题： 回复：管理不确定因素

关于貌洁保湿霜的销售情况，你觉得自己在哪方面最缺乏了解？

哦，这是个有意思的问题，我总是觉得我们真正了解客户对产品的感受，但由于我们并没有直接把产品卖给消费者，所以，在将产品发给经销商后，我们确实不知道接下来的情况。所以，不错，我们确实不知道貌洁保湿霜出库后的情况。

让客户开始思考，挺不错。

关于广告对提高销量的贡献，你有多少信心？

哦，正如大家一向所说，一半有用，一半没用，而且永远也不知道哪一半是哪一半。但很明显，貌洁品牌是消费者购买最多的产品，因为貌洁与其他保湿霜产品并无太大差别，所以广告是打响品牌的关键。

广告效果并不十分确定。

除了少女消费者，还有谁可能会买这些产品？

这我可不知道，毫无线索，因为产品全靠牌子，我们只考虑了少女消费者，我们绝不插手其他消费者群体。

这是一个巨大的盲点！

有没有我该知道的其他难以排解的不确定因素？

当然有，还不少呢。你吓死我了，我再也不觉得自己对产品了如指掌了，你的数据分析让我觉得我对产品的把握所知甚少。

还有谁可能买貌洁保湿霜？

除了少女消费者，有其他买家吗？

确定

分解

评估

决策

世上没有傻问题

问： 首席执行官最后一句话挺搞笑：数据分析让人感觉自己所知甚少，这话不对吧？

答： 这要看你怎么对待。如今越来越多的问题能够通过数据分析技术解决，而在过去，人们要靠直觉来解决这些问题。

问： 所以和以前相比，心智模型越来越不可信了？

答： 许多由心智模型完成的工作都是为了帮助你填补信息空白。好的一面是，数据分析工具让你有能力以系统而自信的方式填补这些空白，因此，“指定大量不确定因素”这一做法的目的就是帮助你发现盲点，这要求拥有过硬的数据工作经验。

问： 但我非得用心智模型来填补“对外界的了解”这项知识的空白吗？

答： 确实如此……

问： 我这么说是因为，即使我目前对外界的运行规律了如指掌，但十分钟后外界就会变成另外一个样子。

答： 对极了。你无法无所不知，世界总是在不断变化，这就是严谨地指定问题并管理心智模型不确定因素之所以成为工作重点的原因。你只有那么些时间、那么些资源来解决分析问题，因此，回答上述问题将有助你有效率、有效果地完成工作。

问： 通过统计模型了解到的信息能为心智模型所用吗？

答： 当然能。今天的研究所发现的事实和现象往往成为明天的研究的假设情况。这样想：你不可避免地会从统计模型得出错误结论，人无完人嘛。当这些结论成为心智模型的一部分后，你希望它们突显出来，这样才能认清情况，以便在需要时回头改变这种结论。

问： 所以心智模型可以通过经验进行试验？

答： 对，而且应该进行试验。你无法试验每一件事，但可以试验模型中的每一件事。

问： 如何改变心智模型？

答： 你即将了解……

首席执行官下令搞来了更多数据，帮助你寻找少女消费者以外的市场。让我们看一看。

Acme给你发来了一长串原始数据

所获得的新数据若未经过任何处理，即称为原始数据，为了让他人提供的数据在你要进行的数据运算中发挥作用，几乎总是要调节数据。

千万要保存原始数据，避免进行任何数据处理。即使是最好的数据分析师也会失误，必须能够将自己的工作结果与原始数据进行比较。

资料太多了……你
能用不了这么多。

日期	经销商	销量 (件)	运区	费用
9/1/08	野奢女化妆品公司	5253	20817	\$76 643
9/3/08	野奢女化妆品公司	6148	20817	\$88 531
9/4/08	信悦公主公司	8931	20012	\$128 806
9/14/08	野奢女化妆品公司	2031	20817	\$29 246
9/14/08	信悦公主公司	8029	20012	\$115 618
9/15/08	泛美批发公司	3754	20012	\$54 058
9/20/08	野奢女化妆品公司	7039	20817	\$107 362
9/21/08	信悦公主公司	7478	20012	\$107 683
9/25/08	泛美批发公司	2646	20012	\$38 102
9/26/08	野奢女化妆品公司	6361	20817	\$91 588
10/4/08	信悦公主公司	9481	20012	\$136 626
10/7/08	泛美批发公司	8698	20012	\$123 811
10/8/08	野奢女化妆品公司	6333	20817	\$91 195
10/12/08	泛美批发公司	4813	20012	\$69 307
10/15/08	信悦公主公司	1550	20012	\$22 320
10/20/08	野奢女化妆品公司	3230	20817	\$46 512
10/25/08	野奢女化妆品公司	2064	20817	\$29 722
10/27/08	泛美批发公司	8298	20012	\$119 491
10/28/08	信悦公主公司	8300	20012	\$119 520
11/2/08	泛美批发公司	6791	20012	\$97 793
11/4/08	信悦公主公司	3775	20012	\$54 360
11/10/08	野奢女化妆品公司	8320	20817	\$119 808
11/10/08	野奢女化妆品公司	6160	20817	\$88 704
11/10/08	泛美批发公司	1894	20012	\$27 274
11/15/08	信悦公主公司	1697	20012	\$24 437
11/24/08	信悦公主公司	4825	20012	\$69 480
11/28/08	野奢女化妆品公司	6188	20817	\$89 107
11/28/08	泛美批发公司	4157	20012	\$59 861
12/3/08	野奢女化妆品公司	6841	20817	\$96 510
12/4/08	信悦公主公司	7483	20012	\$107 755
12/6/08	泛美批发公司	1462	20012	\$21 053
12/11/08	泛美批发公司	8680	20012	\$124 992
12/14/08	野奢女化妆品公司	3221	20817	\$46 382
12/14/08	信悦公主公司	6257	20012	\$90 101
12/24/08	泛美批发公司	4504	20012	\$64 858
12/25/08	信悦公主公司	6157	20012	\$88 661
12/28/08	野奢女化妆品公司	5843	20817	\$85 579
1/7/09	野奢女化妆品公司	4415	20817	\$63 576
1/10/09	信悦公主公司	2726	20012	\$39 254
1/10/09	泛美批发公司	4937	20012	\$71 093
1/15/09	野奢女化妆品公司	9602	20817	\$138 269
1/18/09	泛美批发公司	7025	20012	\$101 160
1/20/09	信悦公主公司	4726	20012	\$68 054
1/21/09	野奢女化妆品公司	7489	20817	\$107 842
1/25/09	信悦公主公司	6280	20012	\$133 532
1/28/09	泛美批发公司	9994	20012	\$143 914
2/8/09	野奢女化妆品公司	7418	20817	\$106 819
2/9/09	信悦公主公司	6201	20012	\$89 294
2/10/09	泛美批发公司	8100	20012	\$116 640
2/12/09	信悦公主公司	9437	20012	\$135 893
2/13/09	泛美批发公司	3354	20012	\$48 298
2/18/09	野奢女化妆品公司	9810	20817	\$141 264
2/23/09	泛美批发公司	5442	20012	\$78 365
2/25/09	信悦公主公司	7813	20012	\$112 507
2/26/09	野奢女化妆品公司	5258	20818	\$75 715

数据太太太多了！
我该怎么办？该从
哪儿开始？



放松

数据多往往是好现象

在密集的数据中兜圈子很容易让人“迷路”，要是你迷失了目标，忘记了假设，只要集中注意力完成该完成的数据处理就能扭转局势，优秀的数据分析的根本在于密切关注需要了解的数据。

确定

分解

评估

决策



练习

好好看看这些数据，想一想**首席执行官的心智模型**。这些数据符合所有顾客都是少女消费者的想法吗？还是看得出有其他的消费者？

日期	经销商	运量 (件)	运区	费用
9/1/08	野蛮女友化妆品公司	5253	20817	\$75 643
9/3/08	野蛮女友化妆品公司	6148	20817	\$88 531
9/4/08	忸怩公主公司	8931	20012	\$128 606
9/14/08	野蛮女友化妆品公司	2031	20817	\$29 246
9/14/08	忸怩公主公司	8029	20012	\$115 618
9/15/08	泛美批发公司	3754	20012	\$54 058
9/20/08	野蛮女友化妆品公司	7039	20817	\$101 362
9/21/08	忸怩公主公司	7478	20012	\$107 683
9/25/08	泛美批发公司	2646	20012	\$38 102
9/26/08	野蛮女友化妆品公司	6361	20817	\$91 598
10/4/08	忸怩公主公司	9481	20012	\$136 526
10/7/08	泛美批发公司	8598	20012	\$123 811
10/9/08	野蛮女友化妆品公司	6333	20817	\$91 195
10/12/08	泛美批发公司	4813	20012	\$69 307
10/15/08	忸怩公主公司	1550	20012	\$22 320
10/20/08	野蛮女友化妆品公司	3230	20817	\$46 512
10/25/08	野蛮女友化妆品公司	2064	20817	\$29 722
10/27/08	泛美批发公司	8298	20012	\$119 491
10/28/08	忸怩公主公司	8300	20012	\$119 520
11/3/08	泛美批发公司	6791	20012	\$97 790
11/4/08	忸怩公主公司	3775	20012	\$54 360
11/10/08	野蛮女友化妆品公司	8320	20817	\$119 808
11/10/08	野蛮女友化妆品公司	6160	20817	\$88 704
11/10/08	泛美批发公司	1894	20012	\$27 274
11/15/08	忸怩公主公司	1697	20012	\$24 437
11/24/08	忸怩公主公司	4825	20012	\$69 480
11/28/08	野蛮女友化妆品公司	6188	20817	\$89 107
11/28/08	泛美批发公司	4157	20012	\$59 861
12/3/08	野蛮女友化妆品公司	6841	20817	\$98 510
12/4/08	忸怩公主公司	7483	20012	\$107 755
12/6/08	泛美批发公司	1462	20012	\$21 053
12/11/08	泛美批发公司	8680	20012	\$124 992
12/14/08	野蛮女友化妆品公司	3221	20817	\$46 382
12/14/08	忸怩公主公司	6257	20012	\$90 101
12/24/08	泛美批发公司	4504	20012	\$64 858
12/25/08	忸怩公主公司	6157	20012	\$88 661
12/28/08	野蛮女友化妆品公司	5943	20817	\$85 579
1/7/09	野蛮女友化妆品公司	4415	20817	\$63 576
1/10/09	忸怩公主公司	2726	20012	\$39 254
1/10/09	泛美批发公司	4937	20012	\$71 093
1/15/09	野蛮女友化妆品公司	9602	20817	\$138 269
1/18/09	泛美批发公司	7025	20012	\$101 160
1/20/09	忸怩公主公司	4726	20012	\$68 054

把答案写在这儿。



练习 解答

从数据中看出什么了？首席执行官“只有豆蔻年华的少女消费者才买貌洁保湿霜”的想法对吗？还是看得出有其他的消费者？

这些公司像是会向少女消费者销售化妆品的。

可以肯定地看到，Acme将产品卖给各家公司，这些公司再将产品卖给年轻的消费者。野蛮女友化妆品公司和恹恹公主公司看来肯定名副其实，可名单上还有一家经销商——泛美批发公司，单从名字无法看出其客户群，但可能值得调查一下。

这些人是谁？

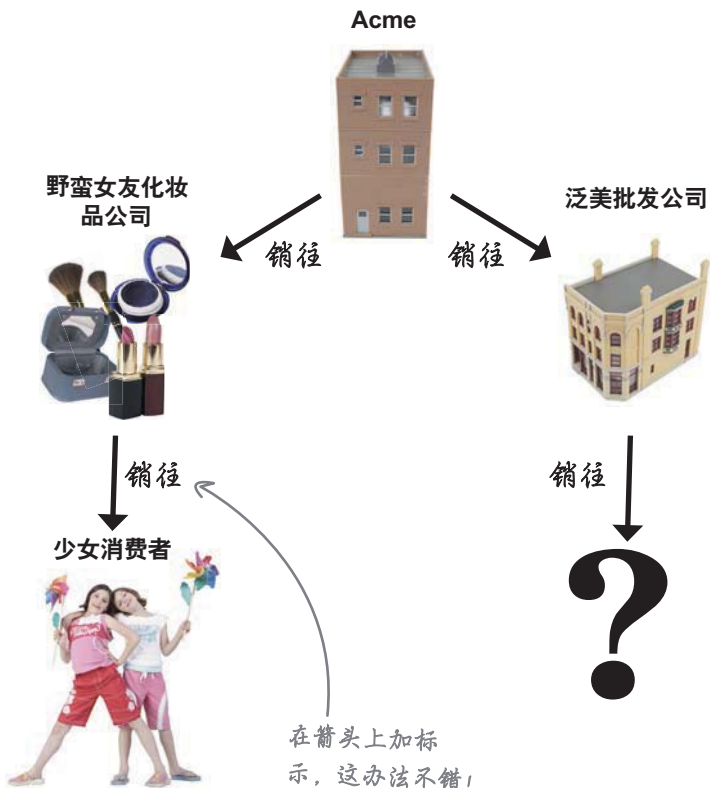
日期	经销商	运量(件)	运区	费用
9/1/08	野蛮女友化妆品公司	5253	20817	\$75 643
9/3/08	野蛮女友化妆品公司	6148	20817	\$88 531
9/4/08	恹恹公主公司	8931	20012	\$128 606
9/14/08	野蛮女友化妆品公司	2031	20817	\$29 246
9/14/08	恹恹公主公司	8029	20012	\$115 618
9/15/08	泛美批发公司	3754	20012	\$54 058
9/20/08	野蛮女友化妆品公司	7039	20817	\$101 362
9/21/08	恹恹公主公司	7478	20012	\$107 683
9/25/08	泛美批发公司	2646	20012	\$38 102
9/26/08	野蛮女友化妆品公司	6361	20817	\$91 598
10/4/08	恹恹公主公司	9481	20012	\$136 526
10/7/08	泛美批发公司	8598	20012	\$123 811
10/9/08	野蛮女友化妆品公司	6333	20817	\$91 195
10/12/08	泛美批发公司	4813	20012	\$69 307
10/15/08	恹恹公主公司	1550	20012	\$22 320
10/20/08	野蛮女友化妆品公司	3230	20817	\$46 512
10/25/08	野蛮女友化妆品公司	2064	20817	\$29 722
10/27/08	泛美批发公司	8298	20012	\$119 491
10/28/08	恹恹公主公司	8300	20012	\$119 520
11/3/08	泛美批发公司	6791	20012	\$97 790
11/4/08	恹恹公主公司	3775	20012	\$54 360
11/10/08	野蛮女友化妆品公司	8320	20817	\$119 808
11/10/08	野蛮女友化妆品公司	6160	20817	\$88 704
11/10/08	泛美批发公司	1894	20012	\$27 274
11/15/08	恹恹公主公司	1697	20012	\$24 437
11/24/08	恹恹公主公司	4825	20012	\$69 480
11/28/08	野蛮女友化妆品公司	6188	20817	\$89 107
11/28/08	泛美批发公司	4157	20012	\$59 861
12/3/08	野蛮女友化妆品公司	6841	20817	\$98 510
12/4/08	恹恹公主公司	7483	20012	\$107 755
12/6/08	泛美批发公司	1462	20012	\$21 053
12/11/08	泛美批发公司	8680	20012	\$124 992
12/14/08	野蛮女友化妆品公司	3221	20817	\$46 382
12/14/08	恹恹公主公司	6257	20012	\$90 101
12/24/08	泛美批发公司	4504	20012	\$64 858
12/25/08	恹恹公主公司	6157	20012	\$88 661
12/28/08	野蛮女友化妆品公司	5943	20817	\$85 579
1/7/09	野蛮女友化妆品公司	4415	20817	\$63 576
1/10/09	恹恹公主公司	2726	20012	\$39 254
1/10/09	泛美批发公司	4937	20012	\$71 093
1/15/09	野蛮女友化妆品公司	9602	20817	\$138 269
1/18/09	泛美批发公司	7025	20012	\$101 160
1/20/09	恹恹公主公司	4726	20012	\$68 054



深入挖掘数据

要看的数据很多，但任务很明确：
找出除少女消费者以外购买产品的群体。

你发现了一家名叫泛美批发公司的
公司，它是谁？谁买它的产品？



练习

泛美批发公司应Acme的要求发来了这份貌洁客户明细表。这些信息能帮助你弄清楚谁在购买产品吗？

写下从这些数据中看出的
购买貌洁保湿霜的消费者

泛美批发公司貌洁保湿霜半年销售明细（至2009年2月）

经销商	件数	%
威猛胡须保养公司	9 785	23%
Haohan.com	20 100	46%
四犊子剃须品公司	8 093	19%
男用化妆品公司	5 311	12%
总计	43 289	100%

.....

.....

.....

.....

.....

.....



练习 解答

泛美批发公司的销售明细告诉你是谁在购买貌洁保湿霜了吗？

泛美批发公司貌洁保湿霜半年销售明细（至2009年2月）

经销商	件数	%
威猛胡须保养公司	9 785	23%
Haohan.com	20 100	46%
四犊子剃须品公司	8 093	19%
男用化妆品公司	5 311	12%
总计	43 289	100%

看来是男人在买貌洁保湿霜！Acme原来的销售表没有显示出是男人在买保湿霜，但泛美批发公司却将貌洁保湿霜转卖给了剃须品经销商！

泛美批发公司确认了你的印象

是，老头们喜欢这东西，虽说用小姑娘的东西让他们有点难为情，但用来做剃须后保养棒极了。

这恐怕是大买卖。

看来，有一个群体在买貌洁保湿霜，而Acme竟还没有意识到。

一切顺利的话，就靠这个潜在群体提高Acme的销量了。



我被迷住了，这条妙计可能让我们的业务来个天翻地覆的变化，能让我过一遍你得出这个结论的过程吗？我们该怎么利用这个新信息呢？

你已经进入分析冲刺阶段。

现在该写报告了。记住，让客户详细地浏览你的思考过程——你是如何得出这个看法的？

根据这个看法，你建议客户如何改进业务？这条信息如何能帮助他提高销量？



动动笔



- 心智模型有哪些改变？
- 有何证据证明你的结论？
- 有难以排解的不确定因素吗？

.....

.....

.....

.....

.....

.....



动动笔 解答

你如何扼要复述你的工作？你对首席执行官提出了哪些建议以期提高销量？

一开始，我试图想办法提高少女消费者市场的销量，因为我们相信这些消费者是貌洁保湿霜唯一的客户群。当我们发现少女消费者市场已经饱和后，我深入挖掘数据，寻找提高销量的源泉。在这个过程中，我改变了心智模型，结果表明热衷于使用貌洁的人比我们意识到的要多——尤其是上了年纪的男人。由于这个消费群并不宣扬自己对产品的热衷，我建议大幅度增加对这个群体的广告宣传，用更易被男性接受的特色销售同样的产品，这将提高销量。

世上没有傻问题

问： 如果为了解决问题而需要获得更详细的信息，我该做到什么程度呢？是不是要亲自去采访客户？

答： 对新数据的挖掘深度最终取决于你自己的最佳判断，在这个例子中，你不断摸索，终于找到了新的市场领域，这个发现足以让你制定有说服力的销售策略。我们将在后续章节中进一步讨论何时该停止搜集数据。

问： 看来，起初的错误心智模型是第一次分析失败的罪魁祸首。

答： 是啊，最初的错误假设注定了分析会得出错误的答案，因此，从一开始就务必要基于正确的假设建立模型显得如此重要，并且，要做好准备，一旦所得到的数据有违你的假设，就要立即回头重新详加思考。

问： 分析会有大结局吗？我所追求的是定论。

答： 数据分析肯定会得出重大问题的答案，但绝不会料事如神，即使你今天无所不知，明天又会有新情况。向年长男子促销的建议可能在今天是有用的，但Acme永远需要分析师为他们出点子抓销售。

问： 听起来挺没劲。

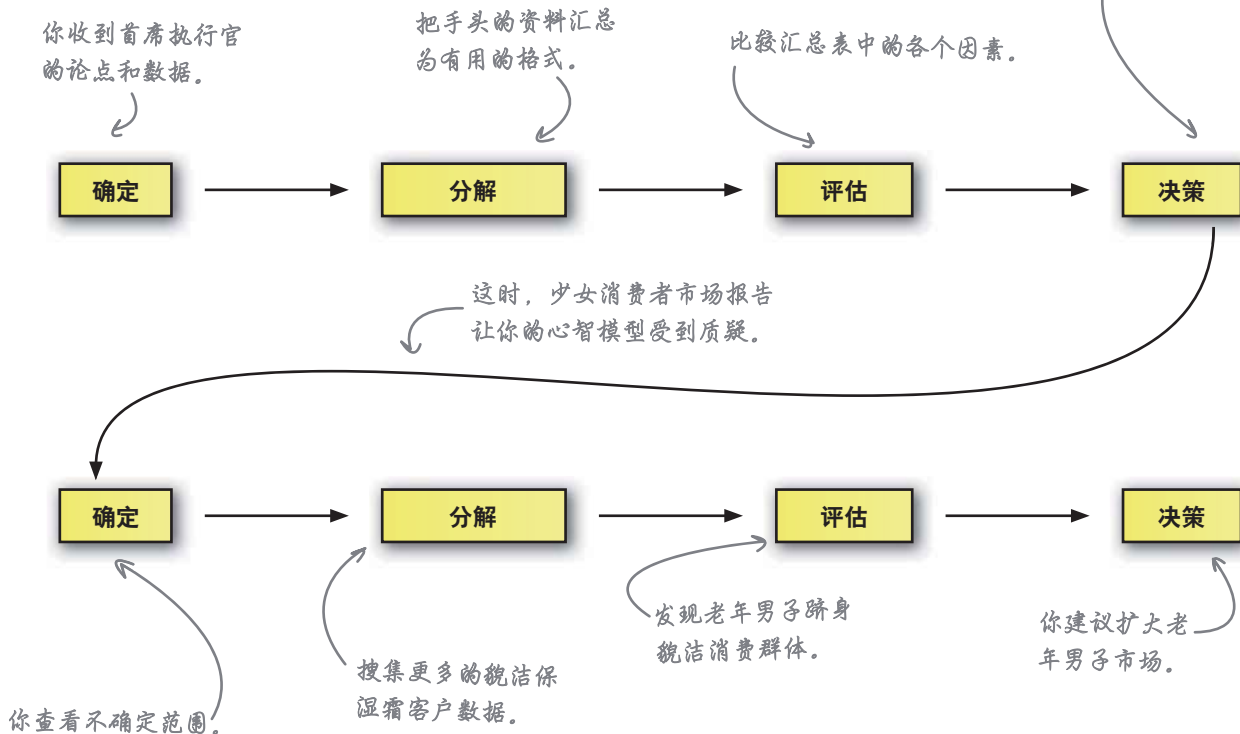
答： 恰好相反！分析师好比侦探，总有一些秘密等着他们去发现，这正是数据分析的乐趣所在！回顾问题、提炼模型、基于新模型观察外界，这些都是分析师工作的基本组成部分，并非特例，而是规律。



回顾你的工作

下面最后看一眼你所经历的所有步骤，目的是得出如何帮助Acme提高貌洁保湿霜销量的结论。

你提出提高对少女消费者市场的广告力度，这可能有助于销量回升。



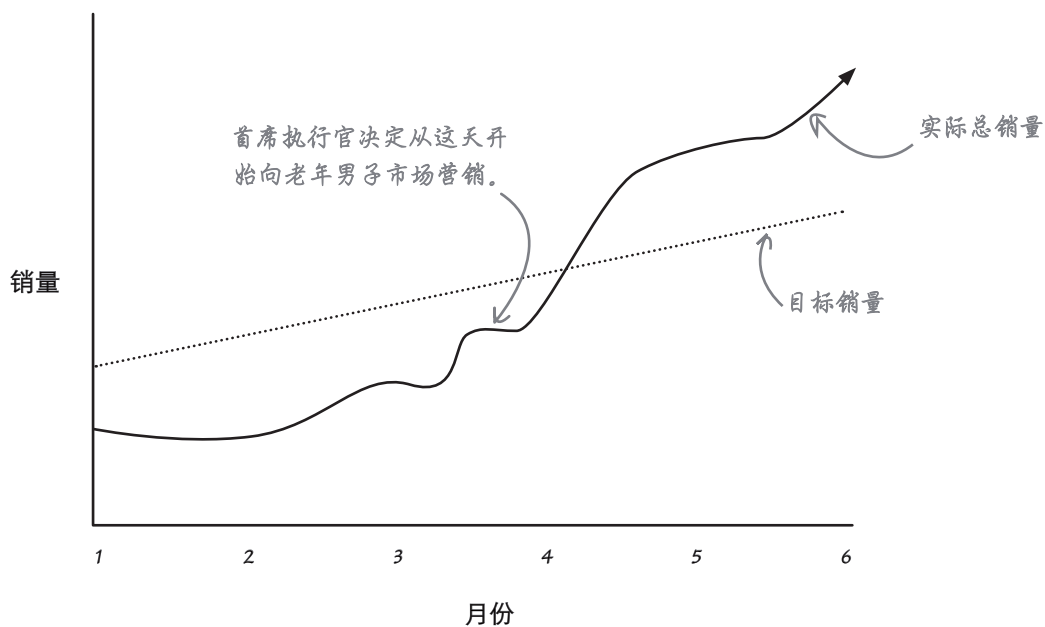
啊，我卖出去了。让我们追随老头们前进吧！



你的分析让客户作出了英明的决策

看了你的报告后，首席执行官迅即调动营销团队创建“须洁”品牌——无非就是“貌洁”保湿霜换个新名字罢了。

Acme旋风般地把须洁保湿霜推向老年男子市场，下面是结果：



销量一飞冲天！两个月的销量超过了你在文章开头看到的所有目标销量。

你的分析出成果了！

