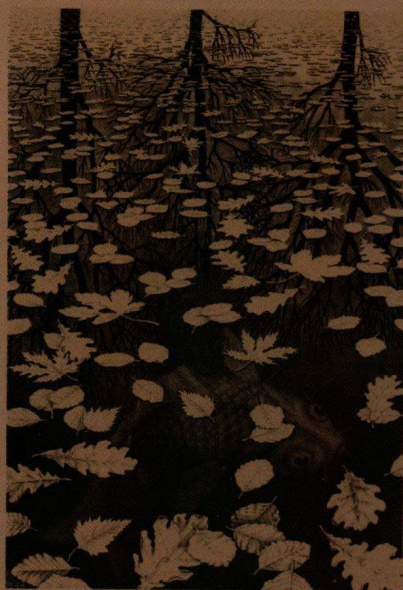


第一本关于Analytics的中文专著



Analytics:
On Big Data and
Evidence-based Decisions

证析

——大数据与基于证据的决策

郑毅◎著 田溯宁◎序



华夏出版社
HUAXIA PUBLISHING HOUSE



Analytics:
On Big Data and
Evidence-based Decisions

证 析
——大数据与基于证据的决策

郑毅◎著 田溯宁◎序



华夏出版社
HUAXIA PUBLISHING HOUSE

图书在版编目(CIP)数据

证析:大数据与基于证据的决策/郑毅著. —北京:华夏出版社,
2012.5(2012.8重印)

ISBN 978-7-5080-6949-4

I. ①证… II. ①郑… III. ①企业管理-数据-分析
IV. ①F270.7

中国版本图书馆CIP数据核字(2012)第086598号

证析——大数据与基于证据的决策

作 者 郑 毅

责任编辑 李欣利

出版发行 华夏出版社

经 销 新华书店

印 刷 三河市李旗庄少明印装厂

装 订 三河市李旗庄少明印装厂

版 次 2012年5月北京第1版

2012年8月北京第2次印刷

开 本 720×1030 1/16开

印 张 27.25

字 数 400千字

定 价 68.00元

华夏出版社 地址:北京市东直门外香河园北里4号 邮编:100028

网址:www.hxph.com.cn 电话:(010)64663331(转)

若发现本版图书有印装质量问题,请与我社营销中心联系调换。

序言 探索知识新边疆

中国宽带资本基金董事长 田溯宁博士

拿到郑毅这本《证析》，便爱不释手，一气呵成地读了下来。一方面因为他的观点正是我所想证实与了解的，同时也为他证据充分、文采飞扬的内容所感动。

前一段到英国参加万事达董事会，参观战争博物馆时，看到了时任海军大臣的丘吉尔在1912年1月的一张关于英国舰队未来的动力是用本土的煤还是新科技进口的石油为原料的便条，丘吉尔力排众议，有远见地确定石油为动力。这个决策使英国海上力量又称霸了几十年。当时我在想，如果今天的海洋是数据信息，那什么是征服这个数据海洋的“油动力”？

人类文明的历史也是知识发现、地域边疆拓展的历史。科技创新在地域疆土的扩张过程中扮演着重要的角色。从西班牙人的三桅船，到英国的蒸汽机动力，地理的扩张随之带来财富的积累，生活质量的提高。人们的世界观也因此发生巨大的变化。大航海使我们知道地球是圆的，天文望远镜、显微镜的发明及一系列理论的创新，使我们认识到了宇宙与微生物，前者使人类感到自己的渺小与“仅有一个地球”的珍贵，后者让我们发现的抗菌素等医药使我们的生命质量得以提高。

二十世纪，尤其是二战之后，野蛮殖民，依靠征疆拓土获得财富的方式早已发生变化。五十年代的计算机技术，使人类对数据、信息的存储、处理与理解发生了质的飞跃。从IBM主机的产生，到七十年代个人计算机的出现，社会、企业、个人的生活已经离不开“计算”技术，正如同工业的革命“动力”使人们的体力得以扩张，“计算机”使人们“智力”得以提高。但我们所有过去的“计算”构架，基本上是处理、存储“结构化”

的数据（某人、某时完成了某一事件）。这种结构型数据为提高企业的效率起到了革命性的作用。同时历史上，这些结构数据的大量积累、沉淀，如同远古的矿藏能源一样，存储了大量的“数据原矿”。九十年代之后，互联网的出现，数据的产生与积累又如火山爆发般地扩大，而文字数据开始变为图片、视频、互动、三维等更大量的形态各异的数据，这些数据的特征是“非结构”，它与事件、社会情绪、天气等诸多不可预测的因素有关。“结构数据”很像牛顿力学，而互联网“非结构数据”则像量子力学。从生物学角度来看，“结构数据”很像人的理性左脑，而“非结构数据”则如同控制我们情绪的右脑。而对世界的认识需要对两种数据的分析。

近十年来，计算领域正向从对数据的交易、记录到对数据的分析、比较、提炼、再分析以产生正确认识的过程发展。这个领域正是“证析”。正如此，“大数据”成为近两年计算、投资领域最热门的话题。迅速发展的“云计算”，则为“大数据”提供了规模、便宜的计算平台。

对“大数据”分析意义重大。如果个人的智慧以智商为衡量标准，对“大数据”的理解与运用是关于企业的“智商”，它的目的是提高企业的决策质量。它的技术基础是通过各种统计、算法，加上前所未有的“云计算”能力。我们对大规模的各种数据进行“冶炼”，从数据中发现关联，发现知识，成为我们对未来判断决策的依据。它的数据基础在于企业、社会的大量积累，它的计算方法则是这些年运用、实践的各种算法，尤其是互联网企业的算法与软件模型，此书都有详细的描述。

“大数据”在中国的研究与实践对现代化进程有着重大意义。

这个领域在理论与实践均刚刚起步。如果我们下大力气，从实践到理论，从理论到实践来抓起，有可能在此领域参与、推动甚至领导世界上此行业的发展。我相信数十年后，人类对数据－信息－知识过程的创新，会如同蒸汽机、计算机的诞生一样将是一个重大领域的创新。对建设创新性

国家的中国无疑是历史性的机遇。

其次，中国现代化中城市化的很多问题，如交通、教育、医疗、社会管理等用传统的方法走不通，或者成本太高，而基于数据的证析可能是解决此类问题的选择。

最后，这个领域的发展孕育着巨大的产业机会，随着人类对数据的依赖性不断增加，我们每个人生下来的数据（生理与心理数据）均会被记录下来，我们将有一个“数据银行”，因此，数据加工业，数据工厂等新产业形式将层出不穷。它将带来几十万、上百万新型软件工程师、数据分析师等新的工作与就业机会，中国应有可能在过去几十年成为世界工厂后，在未来几十年成为世界的“数据工厂”。

郑毅写这本书的经历是值得尊敬的，他曾服务于多家公司，在公司的商业环境不具备合适的研究条件的情况下，他肯于为自己喜欢的证析领域研究付出巨大代价创造条件，长期潜心学习、理解、著学立说。这样的精神在当今商品化、短视与功利化的氛围中尤显可贵。他大量广泛的阅读，认真严谨的索引，并且把技术放在人文、社会的背景下阐述，此书可起到范例作用。多少也印证了他的名字“明心正义”，“士不可以不弘毅”的中国优秀治学传统与理想主义精神。此书不足之处，是对“云计算”与“大数据”这对孪生姐妹的关系剖析未能深入，我期待他下一部书在这方面进一步的探索。

前言

什么是证析？

“证析”是英文单词“analytics”的中文翻译。Analytics 是一个看起来很熟悉的“新词”。它和 analysis（分析）很像，从字面上可以推测这个词和数据及分析相关。但我遍查包括《牛津高阶英汉双解词典》、《朗文高阶英汉双解词典》在内的几本流行的英文词典，里面只收录了 analysis、analyze、analytical 等词条，而没有发现 analytics。如果按照英语的构词法，-ics 这个词根的意思是“关于……”，中文译成“……学”，例如 economics 译成经济学，physics 译成物理学，mathematics 译成数学。以此类推，analytics 应译为分析学。

然而，analytics 不是一门象牙塔里的学问，从一开始 analytics 就有现实世界的意义，analytics 的目的是提升现实世界中决策的质量。简单地说，analytics 就是基于证据尤其是数字化证据进行决策的实践，它区别于那些基于经验、直觉、甚至胆识进行的决策。Analytics 更多应用于商业领域的决策，和 business 一道构成 business analytics。Thomas Davenport 等人对 analytics 定义“Analytics 指的是广泛应用数据、使用统计与量化分析方法、使用描述性与预测性模型以及基于事实的管理方法影响决策和行动的实践^①”。《Analytics》杂志给出的 analytics 定义是“用数学、运筹学、统计

① Thomas H. Davenport, Jeanne G. Harris, *Competing On Analytics: The New Science of Winning*, Harvard Business School Publishing Corporation, 2007, p7.

中译本：康蓉 吴越 译：《数据分析竞争法——企业赢之道》，商务印书馆，2009 年。

学影响商业决策^①”。在其产生与发展过程中，大学、科研院所绝不是其唯一的驱动力，其驱动力更多来自业界通过搜集和分析数据以从中获取最大化的价值的努力，业界利用数据指导从产品设计、物流规划到推广营销的一系列决策。

如果像经济学简称经济、物理学简称物理一样，analytics 应该译成分析。SAS、IBM、Google 等跨国企业所提供的产品与服务中也都将 analytics 译成“分析”。然而，这很难不产生混淆，analytics 经常应用于商业领域，将 analytics 译成“分析”会将 business analysis 和 business analytics 两个差别很大的领域都称为“业务分析”。Business analysis 是“分析业务领域的结构、过程和问题，并向解决方案提供者呈现业务需求^②”。在 IT 领域，business analysis 的目的主要是分析与获取 IT 系统的需求，其结果可能是获取诸如“所有使用者都能远程无线登录客户关系管理系统”这样的需求。虽然定量数据对于获取这样需求有所帮助，但它不是必要条件。这样的需求不一定是以数据为基础的，要得到这样的需求不一定会用到数学或统计学的分析手段。同时，“分析”已经变成一个意义很宽泛的常用词，例如时政分析、电影人物分析、精神分析中的分析都符合“把一事物、一种现象、一个概念分成较简单的组成部分，找出这些部分的本质属性和彼此之间的关系（跟‘综合’相对）^③”的定义，但都不是以数据为基础、不是以数学、统计学为主要手段进行的分析。

虽然用数据支持决策已经有很长的历史，但数据介入决策所能达到的深度、广度以及所具备的潜力却是前所未有的。技术的发展使得计算资源变得更加丰富、便利和便宜，从而使得数据搜集的能力、数据存储的能力以及数据处理的能力极大地丰富，为 analytics 的普及提供了技术的基础。

① Analytics Magazine 网站首页。

② Howard Podeswa, The Business Analyst's Handbook, Course Technology PTR, 2009, p297.

③ 现代汉语词典。

数学、统计学、计算机科学的发展使得人们能够运用更加复杂的数学模型与算法分析数据，丰富了 analytics 分析数据的手段。而哲学、经济学、管理学、社会学、心理学的发展使得理性在决策中占据越来越重要的地位，这些学科的发展为 analytics 提供了思想与理论的准备。商业与贸易的全球化使得大型生产企业生产、运营、物流等流程的复杂程度超出了人类直觉所能掌控的范围，通信手段的发达使得一对一的沟通与个性化生产成为可能这也要求企业能够利用自动化的机器决策在大量低价值交易中获利，商业需求为 analytics 的成长提供了土壤。2008 年起 IBM 等跨国公司开始在 analytics 的名义下重新组织或完善自己的产品线。以 Analytics 为名的杂志也于 2008 年开始发行，大学也开始授予 analytics 的理学硕士学位。

Analytics 是本书的主题，是学科和时代发展到现阶段的新产物，正如英语中使用一个偏僻的单词 analytics 以和 analysis 区分一样，本书为了行文方便也自创^①一个汉语词汇“证析”特别指代最近兴起的 analytics 实践。证，是证据的证，这个证据更多强调的是定量的证据，也就是数据；析，

① 就我目前所掌握的信息，analytics 在汉语中没有统一的译名。它经常被直接翻译成“分析”，例如 IBM 公司新成立了一条服务产品线叫做 Business Analytics and Optimization，中文名字是“业务分析及优化”；SAS 公司的 analytics 解决方案的中文网站上的名字是“分析挖掘技术”；Google 的产品 Google Analytics 的中文网站将该产品称为“Google Analytics（分析）”。本人暂时还未看到有关 analytics 的中文原创图书出版，也许本书将是一个首创。而引进的有关 analytics 的英文图书中 analytics 也没有一个合适的翻译，以 Competing on Analytics 一书为例，中国大陆出版的简体中文版《数据分析竞争法：企业赢之道》由吴越翻译，商务印书馆 2009 年出版，书中将 analytics 根据行文方便译成“数据分析法”、“数据分析”和“分析”。该书在中国台北出版的繁体中文版书名为《魔鬼都在数据里》由胡玮珊译，书中将 analytics 根据不同语境译成“分析”、“分析竞争力”以及“检视分析法”。该书的日文版书名为《分析力を武器とする企業強さを支える新しい戦略の科学》，将 analytics 翻译成“分析力”。如前所述，“分析”这个译法无法和其他语境下 analysis 的翻译相区分，而其他翻译方法要么显得太学术化、要么显得太累赘。因为 analytics 一词是本书的核心，所以这里花一定篇幅讨论这个词应该怎么翻译。

本文行文过程中，在不引起混淆的地方仍然沿用“分析”一词，例如“数据分析”，而用“证析”一词特指以对数字的分析为主要特点的方法、项目以及思想。并且，将从事证析工作的人也称作“分析师”而不是“证析师”。

仍然是分析的析，“析万物之理”，分析数据以产生新的洞察，据此影响决策，从而提升绩效。证析就是指代对量化证据进行分析以影响决策的实践。当人们想到使用数据指导商业决策时，往往过于强调证析中“析”的一面，强调使用数理统计模型、数据挖掘工具等数学手段分析数据，这是一个相对被动的过程。在证析中“证”的一面同样重要，也就是需要主动地搜集数据、搜集证据以指导决策。并且，“分析”一词中的“分”字强调分解的手段，强调还原论的方法论。而在证析的具体实践中，采用还原论还是整体论的方法论并不重要，重要的是寻找到能够指导决策的、证明什么样的做法是有效的证据。

如就其技术与实践的渊源而论，业务证析（Business Analytics）和商业智能（Business Intelligence）之间存在着密切的联系，很多从事业务证析工作的团队目前仍隶属于公司的商业智能团队。根据《数据库系统百科全书》的商业智能词条给出的定义“确切地说，它（商业智能）指的是一组工具和技术，这组工具和技术使得企业能够将其业务数据转换成决策流程中所需的及时、准确的信息，它将信息以合适的形式传递给合适的人。……从体系架构的观点看，商业智能系统的核心是以一致的、整合的形式储存企业历史数据的数据仓库^①”。然而，与商业智能的强调技术手段和工具不同，业务证析强调的不是技术手段而是强调用技术手段解决业务问题。传统的商业智能包括诸如从业务系统中抽取数据、将数据以合理的方式组织与存储在数据仓库、及制作报表等工作。虽然这些工作为业务证

① Ling Liu, M. Tamer Özsu (Eds.), Encyclopedia of Database Systems, Springer, 2009, p288. 另外，Business Intelligence 仿照 Artificial Intelligence 译成“人工智能”而约定俗成的译为商业智能或者商务智能。Business Intelligence 一词首次出现于 1958 年 H. P. Luhn 题为“A Business Intelligence System”（H. P. Luhn, A Business Intelligence System, IBM Journal, October, 1958, p314 - 319）的文章中。文中 Luhn 给 Intelligence 更为宽泛的定义，“理解现有事实之间的关系从而指导行动以达成特定目标的能力”。可以看出，在 Luhn 的定义中并没有一味强调技术手段的重要性，与刚刚诞生两年的 Artificial Intelligence 之间没有必然的联系。在此定义下的 intelligence 中“智能”的含义并不那么突出，其含义更接近 intelligence 的另外一种翻译“情报”。

析提供了重要数据支撑，但这些工作绝不是业务证析的全部，业务证析需要利用商业智能系统中所搜集的数据进行数学建模与分析以支撑业务决策。并且，业务证析的数据来源也不限于数据仓库。例如为测试店面中不同的货品摆放方式对营业额的影响，一家连锁超市开展了一次实验，并用实验的数据设计出最佳的货品摆放方式。这是一个业务证析的过程，但这个业务证析过程中数据的产生、搜集以及分析都不必经过数据仓库。

证析和数据挖掘之间也存在区别。首先，数据挖掘是“从大量数据中抽取或挖掘知识的过程”^①。数据挖掘是业务证析的重要技术手段，但并不是唯一的技术手段。例如上面例子提到的这次实验，用到的技术手段就是统计学中的实验设计与分析，它处理的数据点只有几十个而不是数据挖掘的定义中所提到的大量数据。除数据挖掘之外，模拟仿真、数学优化等工具也是加工数据、产生洞察的重要技术手段。其次，数据挖掘是一种二手数据分析（secondary data analysis）的手段，它对已有的数据进行分析，通常从事数据挖掘的人不会直接参与到数据搜集的过程中^②。而这个例子中，业务证析的工作也包括设计数据采集方案并采集数据等工作。有时业务证析在为决策提供支撑时遇到的最大困难是如何设计有效的度量方法，例如，如何度量广告对销售的刺激效果、如何度量服务改善对客户满意度的提升等，一旦设计了合适的数据采集方式，业务证析的工作也完成了一半。所以，从这个角度讲，业务证析的工作比数据挖掘的工作更加主动，它会主动参与数据搜集方案的设计以及分析结果应用等工作。最后，数据挖掘可能会探索复杂的分析技术，但业务证析不是以技术复杂性为目的，业务证析的目的是提升决策质量。无论技术是否复杂，只要能分析数据、提升决策质量就是对业务证析有用的技术手段。数据挖掘是一项应用技

① Jiawei Han, Micheline Kamber, *Data Mining: Concepts and Techniques*, second edition, Elsevier Inc., 2006, p5.

② David Hand, Heikki Mannila, Padhraic Smyth, *Principles of Data Mining*, MIT, 2001, p132.

术，随着业务证析应用于新的业务领域、新数据的出现以及新的业务需求的出现，也需要数据挖掘不断发展技术、拓展其应用领域。在前几年，数据挖掘曾热极一时，但经过一段时间的实践发现，虽然数据挖掘能解决很多问题，但并不如厂商所鼓吹的那样包办一切，而有时为了分析数据也不一定需要数据挖掘这个技术手段。在这种背景下，厂商与企业开始关注以解决问题为出发点的业务证析^①。

这里需再次强调，判断是否是采用证析的方法解决业务问题不是以所采用的数学分析手段是否足够高级、足够复杂为依据。例如，为监控员工的满意度，微软公司 2006 年申请了一项专利技术，利用这项专利技术计算机可以监控员工的血压、心跳、甚至面部表情。当复杂的数据监控与分析结果表明员工的挫折感很强时，系统会发出警报，也就是部门的主管该采取一些措施提升士气了。当然，这属于业务证析的范畴，因为它能利用数据帮助企业及时发现员工中存在的问题，并能及时做出干预，不过它的实现极为复杂，并耗资不菲^②。为了对员工的满意度进行分析，还有另外的做法。有一家公司给每一位员工发了一袋玻璃球，这些球有红色、黄色、蓝色。员工每天下班时根据自己今天的情绪是高兴、一般或沮丧，向本部门的瓶子里投入一颗红球、黄球或蓝球。第二天早上，公司总经理通过统计每个部门瓶子里不同颜色球的个数来分析员工的情绪如何，如果统计显示某部门的瓶子中蓝球突然增多，也就是昨天员工情绪不太好，则提醒总经理要采取什么措施了。它的分析很简单，简单到用一张纸和一支笔

① Google 趋势 (trends.google.com) 是 google 提供的一项服务，可以获取不同关键词的查询量。从 Google 趋势中可以看出，人们对于 data mining 的关注一直在平稳下降，而 2006 年起人们对于 analytics 一词的关注持续上升，2011 年人们对于 analytics 一词的关注度是对 data mining 一词关注度的 15 倍以上。

② Stephen Baker, Numerati, Marine Books, 2008, p19.

就能完成^①；成本也很低廉，几百块钱就能实现。但是，它也属于业务证析的范畴，因为它也是搜集量化证据（每个瓶子里不同颜色的球的个数），对量化数据进行分析并影响决策的过程。虽然本书对于如此简单的业务证析讨论得不多，但这里还是要强调复杂并不是业务证析的特征和目的。

同样，并非所有利用数据指导决策的实践都是业务证析。想象下面的情况，你和你的合伙人午餐期间正在为是否投资于一个项目争得面红耳赤。最终，你们决定用一个古老的方式解决争执。你从餐桌的牙签筒里随手抽出一把牙签，并约定如果这把牙签是奇数支，你们就投资于这个项目，否则就此罢休。虽然这也是用数据（牙签的数量）影响决策（是否投资项目），但读者能看出这种方法和占卜求签并无二致，肯定也没有人反对不将其列入业务证析的范畴。有些决策虽然看似是以数据为基础做出的，但它们并不能被称为业务证析，因为所谓的数据和分析只是为主观臆断披上一层科学的外衣，不过这层皇帝的新装并不像上面这个例子中数牙签做决策那样容易让人识破。如果希望给出关于证析的一个清晰界限很可能会进入形而上学的讨论中，所以本书只会通过一些例子或议论说明证析是什么不是什么，而不会给出一个结论性的封闭定义。

Harrah's 一个案例

下面将以一个公司的实践为例解释什么是业务证析。Gary Loveman，1989 年在美国顶尖学府麻省理工学院（MIT）获得经济学博士学位。和现在相比，Gary Loveman 读书时所能获得的数据十分有限，他认为当时摆弄的数学只是象牙塔里学问们出于个人兴趣的消遣，而对真实世界的决策没有帮助，这让他一度感到不满甚至沮丧^②。毕业后，他也和很多博士一样

① 例子转引自 Mark Graham Brown 著，谢军容译，《超越平衡计分卡：利用分析型指标提高商业智慧》，中国财政经济出版社，2010。

② Michael Schrage, "Q&A: The Experimenter", Technology Review Website, posted February 18, 2011, <http://www.technologyreview.com/business/32351>.

在大学谋得一个教职，在哈佛大学商学院任教授。1994 年他在《哈佛商业评论》上发表的一篇有关客户服务的文章^①引起商业界的广泛关注。很多公司向他伸出了橄榄枝，他最终于 1997 年接受了主营赌场业务的 Harrah's Entertainment 的邀请，担任该公司首席运营官。当时他只准备待两年，为此请了两年学术假期。但实际情况是他运营 Harrah's 之后就一发而不可收，再没有返回哈佛大学而是在 Harrah's 一直待下来，并于 2003 年接任该公司的首席执行官，任职至今。

教授运营的赌场，注定成为博彩业中的异数。当时拉斯维加斯的其他赌场都投巨资于一些动人心魄的炫目表演招徕顾客。而顾客对于 Harrah's 的优质服务印象更深，这归功于 Loveman 的使用数学运营赌场的新理念。为增加客户忠诚度，Harrah's 使用数据库营销技术以及基于分析的科学决策来拉开和竞争对手之间的距离。他们深入分析客户的数据，开展各种实验，并用分析与实验的结果指导营销与服务，为顾客提供符合他们需求的服务。

和航空公司流行的做法一样，Harrah's 也推行了一套名为“完全回馈”（最初称作 Total Gold，后改称 Total Reward）的会员卡制度。会员卡分三个等级：金卡、白金卡、钻石卡。顾客在 Harrah's 时时刻刻都要用到会员卡：停车、玩老虎机、用餐、住宿。会员卡搜集了大量客户信息，使得 Harrah's 不仅知道顾客的年龄、性别、家庭住址，还知道顾客喜欢玩二十一点还是老虎机，玩了多长时间，多久来一次赌场，在赌场内如何用餐，是否过夜等信息^②。到 2010 年，Harrah's 已积累超过 4 000 万会员的

① James L. Heskett, Thomas O. Jones, Gary W. Loveman, W. Earl Sasser, Jr., and Leonard A. Schlesinger, "Putting the Service - Profit Chain to Work", Harvard Business Review, March - April, 1994, p164 - 174.

② Gary Loveman, "Diamond in the Data Mine", Harvard Business Review, May, 2003, p109 - 113.

信息，这是博彩业最大的客户数据库^①。

传统市场营销中的市场战略制定与数据库营销是两个割裂的流程，往往首先制定了宏观的市场策略，然后利用数据库营销等手段配合完成市场策略的实施。然而，Harrah's与之不同，数据分析已成为Harrah's整个运营流程中的一部分，他们会用数据指导市场策略的制定。通过对会员卡数据及调研数据的研究发现，客户用于赌场娱乐的消费中只有36%花在Harrah's，其他钱都花在别的赌场。这对Harrah's意味着机会，Harrah's将市场策略建立在提高客户的忠诚度上面，让客户尽可能多地在Harrah's消费。

为提高客户忠诚度，Harrah's首先要弄清楚究竟是哪些人在Harrah's消费。数据表明26%的客户贡献了82%的收入。然而，出人意料的是，这部分顾客并不是大多数赌场一直在争取的那些金领豪客，而是包括教师、医生、银行职员、技师在内的中产阶级的中老年顾客，他们足够的时间和收入来玩一把。和那些豪客不同，他们的赌博风格并不是一掷千金，而更喜欢在那些一晚上只有几十到几百美元输赢的老虎机上消磨时间。数据告诉Harrah's，这些客户很少在赌场过夜，他们都是下班的路上或者周末过来玩玩。并且，数据还告诉Harrah's 60美元的筹码更能取悦这部分客户，而他们对于30美元的筹码外加一晚免费住宿以及免费的两顿大餐不是那么感兴趣，明显前者的成本会更低一些。

为了赢得顾客的忠诚，Harrah's通过对顾客的消费模式的分析，利用复杂的量化模型计算出顾客的长期价值（customer worth）——关注一个顾客理论上长期会在Harrah's总共消费多少，而不只关注某一晚上顾客的消费。Harrah's的市场策略是尽量让顾客有更好的体验。系统能根据顾客的

^① Karl Taro Greenfeld, "How to Survive in Vegas", Bloomberg Businessweek, August9 - August15, 2010, p70 - 75.

背景资料及历史的消费模式计算出顾客的痛苦点（pain point）——如果他输钱超过痛苦点，今晚的赌博会成为一个痛苦的回忆，离开 Harrah's 后一去不返。例如，Shelly，34 岁，白人女性，来自中产阶级街区，喜欢玩老虎机。系统预测她一晚赌博的痛苦点是输 900 美元。如果系统发现 Shelly 马上就要输到 900 美元，这时 Harrah's 会派出一位幸运大使（luck ambassador）将她从老虎机前劝走。幸运大使将告诉 Shelly “看来你今天老虎机上的运气不太好，是时候去餐厅大餐一顿了。送你一张 20 美元代金券，一小时内有效。” Shelly 虽然输了钱，但回忆起来，这一晚的经历还不是太糟，至少还有免费晚餐可以享用，她以后有机会还会再次拜访 Harrah's。所以，Harrah's 对数据的利用已经不是简单的汇总数据，也不只是对数据进行动作迟缓的事后分析，已能实时收集数据、实时分析数据，并对顾客的行为实时做出反应。

Harrah's 对数据的实时分析与应用在赌场运营的方方面面都有所体现。老虎机是赌场的一个热门项目，有的顾客为了能玩上老虎机甘愿排队。在排队期间，顾客很可能等得不耐烦而离开。更为重要的是，顾客在排队期间是不消费的，不带来任何利润。当某个老虎机排队人太多时，Harrah's 会开放某些其他区域的老虎机，这些老虎机会有一些优惠政策，并把信息通知到排队的顾客。

随着证析能力的发展，Harrah's 将对数据的分析拓展到包括人力资源管理在内的企业内部运营方方面面。例如，Harrah's 的经理会统计员工在面对顾客时微笑的次数，因为经过分析发现微笑次数和顾客满意度极为相关。Harrah's 还能通过建立数学模型计算每张赌桌或其他客户接触点所需员工数量，从而将适量的具有合适技能的员工分配到相应岗位上。Harrah's 经分析还发现由快乐健康的员工为顾客提供服务能够带来更高的客户满意度，据此他们更加关注员工的健康福利计划。Harrah's 还能量化

评估健康和福利计划对员工的敬业程度和财务绩效的影响。他们敦促更多的员工定期去企业自办诊所接受预防性体检与医疗服务，这项措施降低了企业在紧急医疗救护所花费的成本，在一年时间内为企业节约了数百万美元^①。

Harrah's 在客户服务方面所取得的进步与 Gary Loveman 推行的一切从事实出发的企业文化分不开。如果一位主管在他面前说“我认为……”，这位主管要倒霉了。Harrah's 已经搜集到足够的客户信息，所有员工所做出的决策与建议都应从数据出发，应该是“我知道……”，而不是简单地从个人猜测出发的“我认为……”。^② 并且，Gary Loveman 要求公司在每推出一项新的吸引客户的方案前，先通过小规模实验以验证方案的合理性，之后才能在整个公司范围内推广。如果哪位主管未经实验验证就上马新方案，则很可能职位不保，因为在 Harrah's 员工可能被开除的理由有三：“偷窃，调戏妇女，或没有实行有控制组的实验”。

为了用数字支撑决策，Harrah's 每年在信息系统上的投资超过 1 亿美元，其核心是坐落于拉斯维加斯公司总部的数据库服务器，在这里公司的 IT 部门和市场部门并肩工作。这一切投入都物有所值，在 Gary Loveman 的带领下，Harrah's 从拥有 15 家赌场的地方性企业成长为拥有美国本土 39 家赌场、13 家海外赌场的全球性企业。并且 Harrah's 的收入连续 5 年保持两位数的增长，到 2008 年收入达 89 亿美元，成长为全球最大的博彩公司^③。

证析所包括的内容

从 Harrah's 的例子可以看到，证析的目的是使用数学手段、利用客观

① Thomas H. Davenport, Jeanne Harris, Jeremy Shapiro, “Competing on Talent Analytics”, Harvard Business Review, October 2010, p53-58.

② David O. Becker, “Gambling on Customers”, the McKinsey Quarterly, 2003 Number2, p46-59.

③ Harrahs 网站, <http://investor.harrahs.com>.