胸有成竹！

数据分析的SAS EG进阶

人大经济论坛 主编　常国珍 编著

未来数据分析相关的就业岗位会有1000万人才缺口
CDA数据分析师系列丛书携你与时俱进！

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

CDA数据分析师 系列丛书

胸有成竹！

数据分析的SAS EG进阶

人大经济论坛 主编　常国珍 编著

電子工業出版社
Publishing House of Electronics Industry

内 容 简 介

本书共 5 章，涉及使用 SAS EG 做数据分析的主要分析方法。其中，第 1 章为数据分析方法概述，第 2 章至第 4 章为横截面数据分析方法。第 5 章为时间序列分析方法。每章都根据所涉及的知识点的不同，选取了实用的案例，并为读者准备了相应的思考和练习题。

本书是一本面向商业数据分析初学者的教材，从具体的商业数据分析案例入手，使读者掌握数据分析的目的、理念、思路与分析步骤。本书力图淡化技术，对于方法的介绍也尽量避免涉及过多的数学内容，和高等数学相关的内容只在线形回归和主成分分析这两节中涉及到，而且都辅以图形作形象的展现。因此本书的读者只需要具有高中水平的数学基础即可。

图书在版编目（CIP）数据

胸有成竹！数据分析的 SAS EG 进阶 / 人大经济论坛主编；常国珍编著. —北京：电子工业出版社，2015.2
（CDA 数据分析师系列丛书）

ISBN 978-7-121-25243-3

Ⅰ. ①胸… Ⅱ. ①人… ②常… Ⅲ. ①数据处理—教材数据处理系统—教材 Ⅳ. ①TP274

中国版本图书馆 CIP 数据核字（2014）第 302725 号

策划编辑：张慧敏
责任编辑：徐津平
印　　刷：三河市鑫金马印装有限公司
装　　订：三河市鑫金马印装有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编　100036
开　　本：787×980　1/16　印张：11.5　字数：295 千字
版　　次：2015 年 2 月第 1 版
印　　次：2015 年 2 月第 1 次印刷
印　　数：4 000 册　　定价：49.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

序言：这是一个用数据说话的时代

在 CDA（注册数据分析师）Level I 级教材付诸印刷之际，关于数据分析这个职业及其价值的报道就有很多。比如，下面两条报道就充分体现了在大数据时代下，数据分析的价值。这在以前是从来没有过的。

LinkedIn 的最新投票结果显示，‘统计分析和数据挖掘’是 2014 年最大的求职法宝。LinkedIn 对全球超过 3.3 亿用户的工作经历和技能进行分析，公布 2014 年最受雇主喜欢、最炙手可热的 25 项技能，其中位列榜首的是统计分析和数据挖掘。

麦肯锡公司的一份研究预测称，到 2018 年，在“具有深入分析能力的人才”方面，美国可能面临着 14 万到 19 万人的缺口，而“可以利用大数据分析来做出有效决策的经理和分析师” 缺口则会达到 150 万人。数据科学家将成为 2015 年最热门的职业。

早在 2010 年 2 月，肯尼斯•库克尔在《经济学人》上发表了一份关于管理信息的特别报告——《数据，无所不在的数据》，文中写道：“世界上有着无法想象的巨量数字信息，并以极快的速度增长……从经济界到科学界，从政府部门到艺术领域，很多地方都已感受到了这种巨量信息的影响。”2011 年，麦肯锡发布了《大数据：下一个具有创新力、竞争力与生产力的前沿领域》，使人们在这篇文章里认识到了数据的力量。于是，一夜之间，面向数据分析市场的新产品、新技术、新服务、新业态正在不断涌现。从个人、企业到国家层面，都把数据作为一种重要的战略资产，逐渐认识到了数据的价值，不同程度地渗透到每个行业领域和部门，大大提升了企业的经营利润，推动了经济的发展。

这是一个用数据说话的时代，也是一个依靠数据竞争的时代。目前世界 500 强企业中，有 90%以上都建立了数据分析部门。IBM、微软、Google 等知名公司都积极投资数据业务，建立数据部门，培养数据分析团队。各国政府和越来越多的企业意识到数据和信息已经成为企业的智力资产和资源，数据的分析和处理能力正在成为日益倚重的技术手段。

作为一个数学和统计学的强国，数据分析、数据挖掘和大数据价值挖掘行业在我国仍属于朝阳行业，数据分析人才仍然比较稀缺。各行各业在平常工作中积累的各种各样的数据分析问题仍然没有得到及时有效地解决，有些问题，还是关乎本行业发展的至关重要的问题。数据积累越来越多，期待解决分析的数据问题也越来越多，人们逐渐习惯使用数据作为决策的重要参考依据。据艾瑞的研究报告，未来与数据分析相关的就业岗位会在 1000 万人左右，而目前来说国内合格的数据分析师不足 5 万人，建立一个科学有效的数据分析师培训体系迫在眉睫。

在这样一个用数据说话的时代，积累了丰富的数据分析培训经验的人大经济论坛承担起使命，几番调查研究，几番反复推演论证，在 2013 年，这个大数据的“元年”，CDA 注册数据分析师应运而生！

2003 年，人大经济论坛依托中国人民大学成立，在金融、管理、统计领域已积淀 11 个年头，在国内享有良好声誉。

2006 年，人大经济论坛数据分析培训中心设立，至今经历 8 个春秋，建立了大陆、台湾一线师资团队，培养人才已达 3 万余人。

2013 年，“中国数据挖掘与数据分析俱乐部 CDMC”在人大经济论坛旗下成立，2014 年改名为“CDA 数据分析师俱乐部”。来自政府、金融、电信、零售、电商、互联网、教育等行业人士加入会员，成功举办了数十场行业聚会。紧接着，积累了数据分析培训丰富经验的人大经济论坛在国内展开 CDA 数据分析师系统培训和认证考试，成功见证了 1000 余名数据分析师的成长。

2015 年，人大经济论坛将提供高水平、多层次的数据分析培训服务，以在行业积累 多年的影响力，吸引更好更多的优秀师资，瞄准行业内重要的数据分析问题和难点，攻坚突破，建立更加规范的行业培训体系，引领数据分析培训行业向规范化、有效化和前瞻化方向发展，为数据分析培训做出应有的贡献。

其实，数学（含统计）和英语一样重要，都是人们不可或缺的重要技能。既然英语全民这么重视，数学及其数据分析的技能更加需求于方方面面，更应被做大做强。让我们共同期待人大经济论坛办成另一个数据的“新东方”！

覃智勇

2015 年 1 月 1 日

前　言

感谢您选择“CDA 数据分析师” Level Ⅰ学习系列丛书”之《胸有成竹！数据分析的 SAS EG 进阶》。

该丛书按照数据分析师规范化学习体系而定，对于一名初学者，应该先掌握必要的概率、统计理论基础，包括描述性分析、推断性分析、参数估计、假设检验、方差分析、回归分析等内容，这在第一本书《从零进阶！数据分析的统计基础》中进行了专业详细的讲解。其次，数据分析需要按照标准流程进行，即数据的获取、储存、整理、清洗、归约等系列数据处理技术，这在《如虎添翼！数据处理的 SAS EG 实现》中利用 SAS EG 和编程技术进行了操作过程的详解。最后，经过处理的数据需要根据业务问题，利用相关方法进行建模分析，得出结果，结果检验，绘制图表并解读数据，这在《胸有成竹！数据分析的 SAS EG 进阶》中进行了详细的讲解和操作分析。

CDA 数据分析师丛书整体风格是“理论>技术>应用”的一个学习过程，最终目的在于商业业务应用、职场数据分析，为欲从事于数据分析领域的各界人士提供了一个规范化数据分析师的学习体系。

读者对象

本书是一本面向商业数据分析初学者的教材，从具体的商业数据分析案例入手，使读者掌握数据分析的目的、理念、思路与分析步骤。本书力图淡化技术，对于方法的介绍也尽量避免涉及过多的数学内容，和高等数学相关的内容只在线形回归和主成分分析这两节中涉及到，而且都辅以图形作形象的展现。因此本书的读者只需要具有高中水平的数学基础即可。但是本书强调每种方法的假设、适用条件和与商业数据分析主题的匹配。实践教学中，发现业务经验丰富和有较好商业模式理解的学员，在学习数据分析有更好的效果，这主要原因可能是因为这类学员有较强的思辨能力、分析能力、学习目的性和质量意识，而不是简单的模仿和套用数学公式。

本书以 SAS Enterprise Guide(以下简称 SAS EG)为演示软件，但是操作方法可以自由的转换到 SPSS Statistics 这类图形化统计软件，同时也是学习 SAS 编程的捷径。

工具介绍

SAS EG 是一个以项目为导向的 Windows 应用软件，它被用于实现对 SAS 系统大多数分析能力的快速访问。它通常会被统计专家、业务分析员以及 SAS 程序员使用。利用 SAS 多平台的强大能力，SAS EG 能够使用户访问本地或 SAS 服务器上的数据、管理数据、编写基本报表和汇总，做基本和复杂的数据分析，运用最高质量的 SAS 图形能力，最后将结果输出或发送到 SAS 服务器或其他基于服

务器或 Windows 的应用中。在 SAS EG 中进行的工作也可以容易地被其他的 EG 使用者分享。通过生成 SAS 代码，大多数在 SAS EG 中进行的工作也可以被 EG 外部的 SAS 使用者共享。

SAS EG 面向企业中数据轻度使用客户，它的同类产品是 SPSS。而与 R、Stata 和 Eviews 等科研教学类软件有明显不同。SAS EG 基本继承了 SAS Base 的所有功能，可以方便地调用其他模块的程序。可以说在商业数据分析领域，SAS EG 是 SAS Base 的升级换代产品。SAS EG 和 SPSS 类似，都是可以直接使用鼠标点击操作的，这降低了使用人员的入门难度，而且记录脚本可以便于使用者学习 SAS 语言。它的文档管理功能是目前统计软件中最强大的。其中的流程图使单次分析过程一目了然，这与 SPSS 等有明显差别。SPSS 较难记录分析过程，而 SAS EG 可以将分析过程记录下来，便于使用者反复使用和组织内部共享分析文档。在统计方法方面，SAS EG 的菜单中实现的统计方法少而精炼，满足 90% 以上的商业分析需求，而且其拓展性强大，可以调用 SAS 其他模块的过程，可以实现 SPSS 无法很好实现的时间序列和面板数据分析。在和其他软件衔接方面，SAS EG 以 SAS Base 为基础，而 SAS Base 在某些公司作为 ETL 工具，可见 SAS 具有强大的数据管理功能，可以和企业内部数据库做透明访问。

目前各大金融机构、国有企业和著名外企，尤其是咨询公司都在使用 SAS 产品。SAS Base 是面向数据处理程序员的，入门难度较大，只在专门的数据分析部门使用。而 SAS EG 的用户多为业务部门的工作人员，入门难度较低。在公司内部培训的过程中，发现公司数据分析人员和业务人员对学习 SAS EG 有较大兴趣，部门领导也倾向于让员工多学习 SAS EG 的课程。而且 SAS 公司也逐步将其部分产品免费化，其中 University-Edition 就是一个有益的尝试，其操作方式和 SAS EG 类似。相信在统计技能大众化的今天，SAS EG 有着巨大的发展潜力。

当前 R 和 Python 等开源软件方兴未艾，但是这类软件学习曲线缓慢，使很多初学者的热情在进入数据分析的核心领域之前就已经消逝殆尽。真正商业数据分析的目的是为了业务的分析需求，构造稳健的数据挖掘模型。数据挖掘产品的质量是通过对分析流程的严格掌控而得以保障的。SAS EG 产品正是针对分析流程设计的，这对于数据分析初学者大有裨益。而开源软件在这方面基本上没有支持，而要求其使用者具有丰富的实战经验。因此使用 SAS EG 这个产品作为演示工具，无论将来读者使用何种分析工具，都可以通过本书的学习获得分析流程的经验。

阅读指南

本书包括 5 章，涉及使用 SAS EG 做数据分析的主要分析方法。其中，第 1 章为数据分析方法概述，第 2 章至第 4 章为横截面数据分析方法。第 5 章为时间序列分析方法。每章都根据所涉及的知识点的不同，选取了实用的案例，并为读者准备了相应的思考和练习题。

详细的章节内容如下。

第 1 章　数据分析方法概述

数据分析的目的是使工作更有效率、资源分配更合理、对事物的发展脉络更为清晰或是提高对未来预测的准确性。阅读本章可以使读者在具体接触数据分析之前，了解整个数据分析的脉络，明确将要学习的内容。

第 2 章　描述数据特征

数据统计指标描述是数据分析的重点，对数据的直觉也是通过对数据的探索建立起来的。数据可

视化则是将统计指标转换成图形和图表。通过本章的学习，读者可以掌握完成一份市场分析报告的基本技能。

第 3 章　描述性数据分析方法

该部分是上一章的自然延伸，是大数据背景之下描述类数据分析方法的主要手段。分别针对变量过多和观测样本过多这两个问题，进行变量和观测这两个维度的信息压缩。通过本章的学习，可以完成客户画像、因素分析等较高质量的分析报告。

第 4 章　预测性数据分析方法

传统意义上的数据分析建模特指预测性数据分析。在完整本章的学习之后，对于横截面数据分析方法就算结束了。通过本章的学习，可以构造精细的精准营销、流失预警和信用评级等分类模型。

第 5 章　时间序列

本章主要介绍两种单变量时间序列分析方法。分别是趋势分解法和基于动态差分方程的 ARIMA 法。对于非统计学背景的读者，只要学会分析软件提供的图表就可以掌握该分析方法，满足一般的商业指标预测需要。

为方便读者学习，本书提供了书中实例的源文件下载，请读者进入人大经济论坛（http://bbs.pinggu.org/），注册后搜索“CDA 教材源文件”关键词下载相应的源文件。

本书特点

本书作为市场上第一本以 SAS EG 为统计工具的面向商业数据分析的书籍，和其他统计软件图书有很大的不同，文体结构新颖，案例贴近实际，讲解深入透彻。主要表现在以下几方面：

场景式设置

本书从实际电信、银行等商业案例中进行精心归纳、提炼出各类数据分析的运用场景，方便读者搜寻与实际工作相似的问题。

开创式结构

本书案例中的“解决方案”环节是对问题的思路解说，结合“操作方法”环节中的步骤让人更容易理解。“原理分析”环节则主要解释所使用代码的工作原理或者详细解释思路。“知识扩展”环节包括与案例相关的知识点的补充，可拓展读者的视野，同时也有利于理解案例本身的解决思路。

启发式描述

本书注重培养读者解决问题的思路，以最朴实的思维方式结合启发式的描述，帮助读者发现规律、总结规律和运用规律，从而启发读者快速找出问题的解决方法。

学习方法

俗话说打把势全凭架势，像不像，三分样。只有对分析的流程熟悉了，才能实现从模仿到灵活运用的提升。在产品质量管理方面，对流程的掌控是成功的关键，在数据分析当中，流程同样是重中之重。数据分析是一个先后衔接的过程，一个步骤的失误会带来完全错误的结果。一个分析的流程大致包括抽样、数据清洗、数据转换、建模和模型评估这几个步骤。如果抽样中的取数逻辑不正确，就有可能使因果关系倒置，得到完全相反的结论。数据转换方法如果选择不正确，模型就难以得到预期的结果。而且，数据分析是一个反复试错的过程，每一步都要求有详细的记录和操作说明，否则分析人

员很可能迷失方向。

学习数据分析最好的方法就是动手做一遍，本书语言通俗但高度凝炼，很少有公式，这会让读者产生麻痹大意的思想，如果不动手做一遍，很难体会到书中表述的思想。本书按照相关商业数据分析主题提供了相应的演练用数据，也同时给出了相关方面的参考资料，供学员学习。

售后服务

本书读者可以在人大经济论坛的“数据挖掘与商业智能（http://bbs.pinggu.org/forum-133-1.html）”版块就书中的问题进行提问，也欢迎大家就自己遇到的业务问题和大家讨论。同时，也可以向作者发邮件，作者邮箱为 guozhen.c@gmail.com。

致谢

本书由人大经济论坛策划，常国珍负责编写和完成统稿。

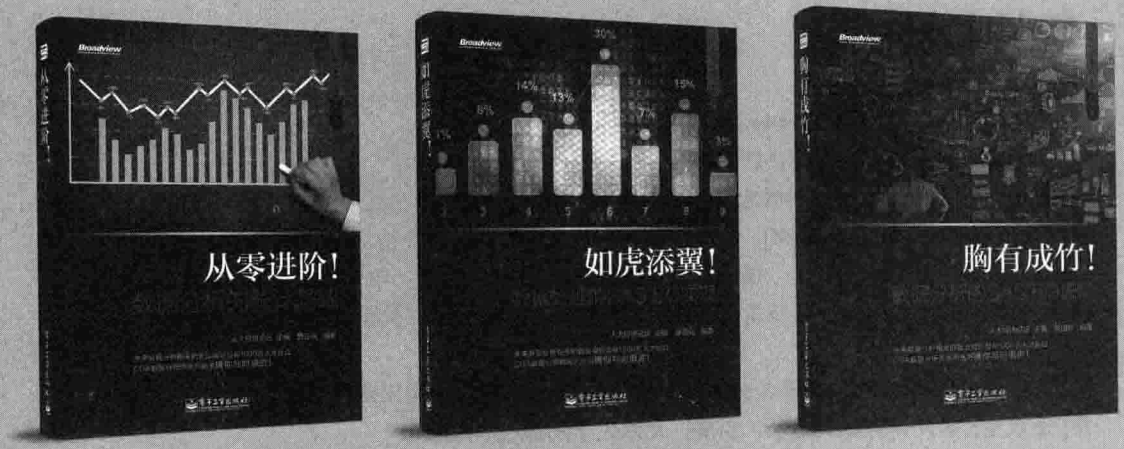
丛书从策划到出版，倾注了电子工业出版社计算机图书分社张慧敏、石倩、官杨、张童等多位编辑的心血，特在此表示衷心的感谢！

为保证丛书的质量，使其更贴近读者，我们组织了人大经济论坛的多位版主和高级会员参与了本书的预读工作，他们是杨同梅、田佳、孙华枫、原瑜芬、叶阵雨、郑赟、李剑宇、江翊雪、陈鹏、刘莎莎、丁亚军。感谢各位预读员的辛勤、耐心与细致，使得本丛书能以更加完善的面目与各位读者见面，特别感谢覃智勇圆满地组织了本次预读工作和审校工作。

尽管作者们对书中的案例精益求精，但疏漏仍然在所难免，如果您发现书中的错误或某个案例有更好的解决方案，敬请登录社区网站向作者反馈，我们将尽快在社区中给出回复，且在本书再次印刷时修正。

再次感谢您的支持！

未来数据分析相关的就业岗位会有1000万人才缺口
CDA数据分析师系列丛书携你与时俱进！



丛书介绍 CDA数据分析师

该丛书按照数据分析师规范化学习体系而定，对于一名初学者，应该先掌握必要的概率、统计理论基础，包括描述性分析，推断性分析，参数估计，假设检验，方差分析，回归分析等内容，这在第一本书《从零进阶！数据分析的统计基础》中进行了专业详细的讲解。其次，数据分析需要按照标准流程进行，即数据的获取、储存、整理、清洗、归约等系列数据处理技术，这在《如虎添翼！数据处理的SAS EG实现》中利用SAS EG和编程技术进行了操作过程的详解。最后，经过处理的数据需要根据业务问题，利用相关方法进行建模分析，得出结果，结果检验，绘制图表并解读数据，这在《胸有成竹！数据分析的SAS EG进阶》中进行了详细的讲解和操作分析。

CDA数据分析师丛书整体风格是“理论>技术>应用”的一个学习过程，最终目的在于商业业务应用、职场数据分析，为欲从事于数据分析领域的各界人士提供了一个规范化数据分析师的学习体系。

作者介绍

人大经济论坛（bbs.pinggu.org）：于2003年成立，致力于推动经管学科的进步，传播优秀教育资源，目前已经发展成为国内最大的经济、管理、金融、统计类的在线教育和咨询网站，也是国内最活跃和最具影响力的经管类网络社区。

人大经济论坛从2006年起在国内最早开展数据分析培训，累计培训学员数万人，在大数据的趋势背景下，作为Certified Data Analyst Institute（注册数据分析师协会，简称CDA协会）的中国唯一授权中心，根据CDA协会的数据分析师Level Ⅰ（业务数据分析师）、Level Ⅱ（建模分析师）、Level Ⅲ（数据分析专家）的等级标准，致力于培养正规化、科学化、专业化的数据分析师队伍，为企事业单位输送更多优秀数据分析人才（Certified Data Analyst Institute，亦称“注册数据分析师协会”，成立于美国特拉华州，主要宗旨为汇聚国际先进的数据分析技术，建设国际性规范化数据分析师职业标准，推进数据分析师的行业发展及认证工作，目前标准行业认证为CDA数据分析师）。

反侵权盗版声明

目　录

第 1 章

数据分析方法概述

从事脑力工作的目的是什么？是使工作更有效率、资源分配更合理、对事物的发展脉络更为清晰或是提高对未来预测的准确性。数据分析可以在这些方面为我们提供帮助。数据分析方法起源于各个学科的实际工作，通过学者与专业人士的凝练与升华，形成了方法论体系。数据挖掘专家韩家炜说“数据—信息—知识是一个自然而然的过程。”。这一分析流程的掌握对理解数据分析至关重要。本章就是使读者在具体接触数据分析之前，了解整个数据分析的脉络，明确将要学习的内容。

本章分三个部分，首先，介绍商业场景之下数据分析的概念，明确数据分析的意义在于完成商业目标，而其立足点是商业理解。其次，概括介绍数据挖掘中两类方法，由于 SAS EG 这个产品是一个统计分析工具，该部分介绍的一些内容，比如网络分析、神经网络超出了本书的范围。对其进行介绍，主要是希望让读者了解主要的数据挖掘方法，为后续深入学习数据挖掘方法做铺垫。最后，介绍两个数据挖掘的方法论，其中 CRISP-DM 是指导数据分析整个项目的流程，它是从实际项目中总结出来的，有的公司将该方法论固化为数据挖掘工作流运用于数据挖掘项目的管理；SEMMA 方法是数据分析中具体操作层面的流程，主要的数据挖掘软件，比如 SAS EM 和 SPSS Modeler 的菜单布局完全依照该方法，甚至 SAS EG 这样的统计分析软件在菜单布置中也参考该方法。

1.1 数据分析概述

1.1.1 数据分析过程

数据分析的目的是为业务发展答疑解惑。他描述了“过去发生了什么”、“现在正在发生什么”和“未来可能发生什么”。根据分析的级别，分为常规报表、即席查询、多维分析（又称为钻取或者 OLAP）、警报、统计分析、预报（或者时间序列预测）、预测型建模（预测性（predictive）模型）和优化，如图 1-1 所示。

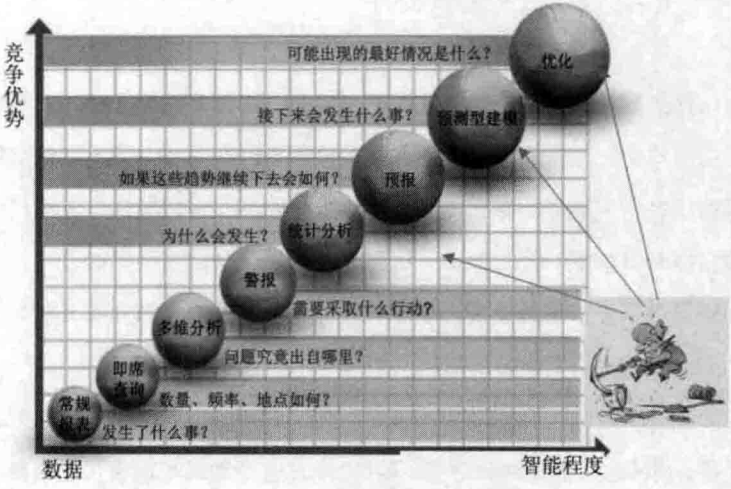


图 1-1

（图 1-1 摘自：SAS 公司《SAS 数据挖掘技术概览》）

（1）常规报表：常规报表广为人知，它们通常按照一定的周期产生，对过去一段时间、一定范围内所发生的事实进行记录。它们对了解业务现状非常有用，但是却无法据此进行长期决策。标准报表主要用于回答“发生了什么”和“什么时候发生”这样的问题。典型的标准报表包括月度或季度的财务报告。

（2）即席查询：即席查询往往通过对一系列数据（组合）的要求来“回答”一些常见的业务问题。即席报表主要用于解决类似“多少”、“频次如何”和“在哪里”这样的问题。记录每种产品每天销量的定制报表就属于即席报表。

（3）多维分析（又称为钻取或者 OLAP 技术）：OLAP 技术可以帮助了解更多细节信息，它可以帮助客户自己操纵数据，找出诸如“多少”、“什么”和“哪里”之类问题的答案。OLAP 技术主要解决的是“问题出在哪里”和“我如何找到问题的答案”这样的问题。例如，对不同类型的客户通话行为进行排序，找出他们的通话特征就需要运用到 OLAP 技术。

（4）警报：当问题发生时你可以通过告警及时获知，并且可以在将来发生类似情况时引起注意。告警可以通过电子邮件、网络频道、记分卡或者仪表盘的形式给出。警报的过程需要确认的是引起注意的触发点，以及一旦报警需要采取什么行动。比如，销售总监在销售情况与销售目标差距大时会收到告警信息。

（5）统计分析：我们可以运行一些更加复杂的分析。例如，方差分析和回归分析等。我们可以基于数据提出一些假设，然后再利用数据构建统计分析模型来“回答”这些假设是否成立。统计分析解决的问题主要是“行为/事件为什么发生”和“我失去了怎样的机会”。例如，银行希望了解什么样的人，更可能对他们的房子进行转按揭操作，那么他们就会用到统计分析的方法。

（6）预报（或者时间序列预测）：它能够帮助建立恰当的库存，从而使得既不会脱销，也不会积压库存。时间序列预测主要解决的问题是“未来的趋势会怎样”和“如果这样的趋势继续会怎样”。例如，零售商可以根据销售历史，预测未来特定店铺的特定产品的销售量，而这样的预测过程就是时间序列预测。

（7）预测型建模（预测性模型）：如果你有 1000 万个客户需要做一次直邮，谁最有可能响应？怎样对现有客户进行有效分群？哪些客户最可能流失？预测性模型可以回答这类问题。预测性模型主要关心的是将来可能发生的情况，以及不同的预测情况对业务的影响。例如，商户可以预测客户可能会对哪种产品更有兴趣，以及哪些客户会对特定产品更有兴趣。

（8）优化：优化往往带来创新，它使企业可以在有限资源下实现利润最大化。优化强调的是更好地利用各种资源的途径。例如，在特定资源条件下，如何安排并使利润最大化，就是优化需要解决的问题。

前 4 类分析提供了关于以往和当前情况的描述，让业务人员对历史情况有一个深入的认识。但是这往往是不够的，这就像在驾驶的时候只看两边和后视镜，而挡住前面的玻璃，对前面发生的情况一无所知。第 5 类到第 7 类分析提供了向前看的途径，可以预测未来发展的情况，及早发现问题，做到提前准备。而最后一类分析是在掌握了未来发展状况之后，对业务进行优化，制定最优的决策方案。

从上面介绍中可以看到，数据分析是和业务紧密联系在一起的，其目的就是满足商业决策的需求。这种决策是以事实和数据分析的结果为基础，结合经验和行业的洞察作出决策。在解读和判断数据模型时，需要融入对业务的理解、融入基于经验的灵感，很多时候是无法用单纯的公式或规则来替代人的智慧和艺术灵感的。因此，数据分析是技术与艺术的结合。如果可以量化分析某些问题，那么就去分析，但别忘记加入你的经验、知识和理性的推断。

1.1.2　数据分析的商业驱动

可以认为数据分析涉及到公司运营的方方面面，这包括对企业部门经营情况的评估、内部员工的管理、生产流程的监管、产品结构优化与新产品开发、财务成本优化、市场结构的分析和客户关系的管理。其中，关于客户与市场的数据分析是“重头戏”。下面以客户全生命周期管理为例介绍数据分析运用场景和挖掘主题，如图 1-2 所示。

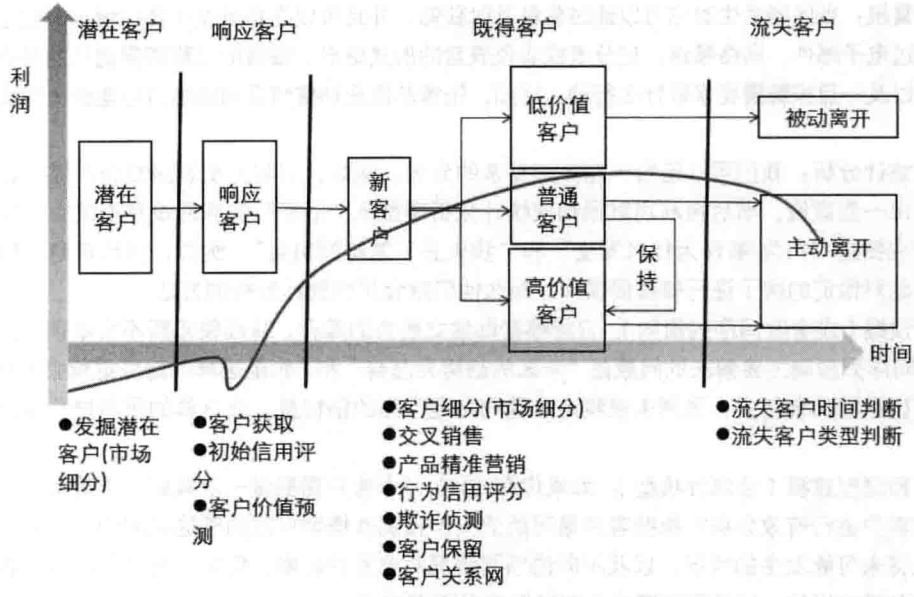


图 1-2

（1）发掘潜在客户（市场细分）：关于这个主题的分析，更多的是基于地区、性别和年龄段等粗粒度的指标，结合产品设计定位和目标客户群体进行匹配。比如，高档母婴产品的潜在客户应该是新建高档小区中的住户。这类分析是运用最早的，在广告投放、新店寻址等场景下大量使用。

（2）客户获取：当客户初次了解我们的产品和服务后，有可能会犹豫不决，拖延很久才可能真正成为我们的客户，而大部分客户在这期间会由于兴趣逐渐减退而最终流失。比如，信用卡新客户在填好个人信息，并收到信用卡后却迟迟没有开卡。这时就可以运用数据挖掘技术，对营销人员得到的客户基本信息进行一个初步筛选，找出购买倾向性较高的客户进行深度跟踪营销。这么做既减少了人工成本，又降低了打扰客户的次数，从而减少了投诉。同时在与潜在客户的交流中，也会为其制定更个性化的产品或服务组合。

（3）初始信用评分：当客户最终购买我们的产品时，在涉及赊销情况的时候，就会用到初始信用评分技术。这是根据客户的性别、年龄以及居住场所等基本信息对客户的信用进行预判。这类情况不只在银行信贷中会遇到，在很多企业中都会遇到。企业的应收账款就是一种自然的商业信用，建立好优秀的初始信用评分体系，可以使企业在不增大财务风险的情况下快速开拓市场。比如，IBM 全球融资部（IGF）是一个为赊购买入 IBM 产品的小公司提供金融服务的部门，其在上世纪 80 年代开发的客户信用评分模型对开拓全球市场功不可没。现在这个技术也成为了提高客户满意度的一种方式。比如，中国移动的先付费客户的欠费额度和京东的“打白条”服务。

（4）客户价值预测：为了更好地为客户提供服务的同时增加企业利润，需要根据客户的基本信息进行其价值预测。其中价值既包括以消费水平为代表的直接价值，也包括客户口碑宣传的间接价值。

（5）客户细分（市场细分）：根据客户的基本信息，从人口学、工业统计信息、社会状态、产品使