

普通高等教育



计算机系列规划教材

数据可视化分析

(Excel 2016+Tableau)

★ 吕峻闽 张诗雨 主编 ★
★ 李 化 龚轩涛 高玲玲 汤来锋 副主编 ★



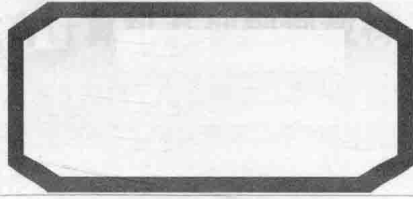
中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>



普通高等教



数据可视化分析

(Excel 2016+Tableau)

吕峻阁 张诗雨 主编

李化 龚轩涛 副主编
高玲玲 汤来锋

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书主要介绍数据可视化分析相关的知识，所使用的工具为 Excel 2016、Tableau。本书使用一个模拟企业的案例数据贯穿所有内容，主要介绍了数据分析的概念，以及实现数据可视化的方法。读者通过本书的学习，能够掌握数据分析理论，并能够制作数据可视化分析报告及商业智能仪表盘。

本书既可作为师生学习数据分析的教材，也可供初学数据分析的用户使用。无任何数据分析基础的用户都可无障碍阅读本书，学习数据可视化分析的相关知识。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目 (CIP) 数据

数据可视化分析：Excel 2016+Tableau / 吕峻闽，张诗雨主编. —北京：电子工业出版社，2017.9
普通高等教育计算机系列规划教材
ISBN 978-7-121-32190-0

I. ①数… II. ①吕… ②张… III. ①表处理软件—可视化软件—高等学校—教材 IV. ①TP391.13

中国版本图书馆 CIP 数据核字 (2017) 第 161259 号

策划编辑：徐建军 (xujj@phei.com.cn)

责任编辑：徐 萍

印 刷：三河市良远印务有限公司

装 订：三河市良远印务有限公司

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：787×1 092 1/16 印张：12.25 字数：294 千字

版 次：2017 年 9 月第 1 版

印 次：2017 年 9 月第 1 次印刷

印 数：4 000 册 定价：35.00 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，(010) 88258888。

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

本书咨询联系方式：(010) 88254570。

本书编委会成员

(按拼音排序)

陈昌平 陈 婷 陈 婷 陈小宁 高玲玲

龚轩涛 郭 进 何臻祥 黄纯国 靳紫辉

李长松 李 化 刘 强 罗 丹 罗文佳

吕峻闽 马 明 汤来锋 王 强 王书伟

魏雨东 夏钰红 肖 忠 徐鸿雁 杨大友

姚一永 银 梅 袁 勋 张诗雨

大数据时代，无论是管理者、经营者还是政策的制定者，都面临着管理好数据、发现数据中的规律以及从数据中获得价值的挑战。数据可视化分析是数据管理技能之一。在未来，90%的市场决策和经营决策都应该通过数据分析来研究确定的。读者通过本书的学习，能够掌握数据分析基本原理与一些有效的数据分析方法，并能灵活运用到实践中。

本书共分为7章，内容包括数据分析的基本概念、数据分析方法、数据的处理、数据展示、数据可视化报告及商业智能仪表盘的制作。对于各个知识点的讲解，都配有模拟企业案例，并使用 Excel 2016 来实现。在商业智能仪表盘章节扩展了工具，同时使用 Excel 2016 和 Tableau 来实现。

本书的特点为使用模拟企业案例贯穿数据分析理论和概念。模拟企业案例数据涉及人力资源、商品库存、商品销售三方面，并结合相关专业知识点进行数据分析及可视化，形成数据可视化分析报告及商业智能仪表盘。

读者通过本书的学习，可以具备以下能力：

懂分析——能够掌握数据分析基本原理与一些有效的数据分析方法，并能灵活运用到实践中。

懂工具——能够运用 Excel、Tableau 工具实现数据分析理论，制作数据可视化分析报告及商业智能仪表盘。

懂设计——能够运用图表有效表达数据分析的观点，使分析结果一目了然。

本书由吕峻闽、张诗雨任主编，由李化、龚轩涛、高玲玲、汤来锋任副主编并负责编写相应各章节，参加本书编写的还有陈昌平、陈婷、陈小宁、郭进、何臻祥、黄纯国、靳紫辉、李长松、刘强、罗丹、罗文佳、马明、王强、王书伟等。同时，西南财经大学天府学院信息技术教学中心和现代技术中心的各位老师为本书提供了许多帮助，编委会成员为本书的编写也提供了很大的帮助。在此，编者对以上人员致以最诚挚的谢意！

前言

Preface

大数据开启了一次重大的时代转型，在这个数据呈爆炸式增长的转型社会，无论是管理者、经营者还是政策的制定者，都面临着管理好数据、发现数据中的规律以及从数据中获得价值的问题，也就是说数据分析技能已经成为未来必不可少的工作技能之一。在未来，90%的市场决策和经营决策都应该是通过数据分析来研究确定的。读者通过本书的学习，能够洞悉数据背后的意义，并能够用随手可及的工具进行数据可视化分析，轻松应对时代转型。

本书共分为7章，内容包括数据分析的基本概念、数据分析方法、数据的处理、数据展示、数据可视化报告及商业智能仪表盘的制作。对于各个知识点的讲解，都配有模拟企业案例，并使用 Excel 2016 来实现。在商业智能仪表盘章节扩展了工具，同时使用 Excel 2016 和 Tableau 来实现。

本书的特点为使用模拟企业案例贯穿数据分析理论和概念，模拟企业案例数据涉及人力资源、商品库存、商品销售三方面，并结合相关专业知 识进行数据分析及可视化，形成数据可视化分析报告及商业智能仪表盘。

读者通过本书的学习，可以具备以下能力：

1 懂分析——能够掌握数据分析基本原理与一些有效的数据分析方法，并能灵活运用 到实践中。

2 懂工具——能够运用 Excel、Tableau 工具实现数据分析理论，制作数据可视化分析 报告及商业智能仪表盘。

3 懂设计——能够运用图表有效表达数据分析的观点，使分析结果一目了然。

本书由吕峻闽、张诗雨任主编，由李化、龚轩涛、高玲玲、汤来锋任副主编并负责 编写相应各章节，参加本书编写的还有陈昌平、陈婷、陈小宁、郭进、何臻祥、黄纯国、 靳紫辉、李长松、刘强、罗丹、罗文佳、马明、王强、王书伟等。同时，西南财经大学 天府学院信息技术教学中心和现代技术中心的各位老师为本书提供了许多帮助，编委会 成员为本书的编写也提供了很大的帮助。在此，编者对以上人员致以最诚挚的谢意！

为了方便教师教学，本书配有电子教学课件，请有此需要的教师登录华信教育资源网（www.hxedu.com.cn）注册后免费下载，如有问题可在网站留言板留言或与电子工业出版社联系（E-mail: hxedu@phei.com.cn）。

虽然我们精心组织，努力工作，但错误之处在所难免；同时由于编者水平有限，书中也存在诸多不足之处，恳请广大读者朋友们给予批评和指正，以便在今后的修订中不断改进。

编者

前言

Preface

本书在编写过程中，得到了许多专家和学者的帮助，在此向他们致以衷心的感谢。同时，本书的出版也得到了许多领导和同事的支持，在此向他们致以衷心的感谢。本书的出版，是作者多年教学经验的总结，也是作者多年教学实践的结晶。本书的出版，旨在为广大的读者提供一本实用的教材，帮助读者更好地理解和掌握本书的内容。

本书共分八章，第一章为绪论，主要介绍本书的编写目的、适用范围、主要特点等。第二章至第七章分别介绍了本书的主要内容，第八章为附录，主要介绍本书的参考文献、索引等。本书在编写过程中，力求做到概念清晰、重点突出、循序渐进、由浅入深，力求做到既注重理论知识的传授，又注重实践能力的培养。

本书在编写过程中，参考了大量的文献资料，力求做到概念清晰、重点突出、循序渐进、由浅入深，力求做到既注重理论知识的传授，又注重实践能力的培养。本书在编写过程中，力求做到概念清晰、重点突出、循序渐进、由浅入深，力求做到既注重理论知识的传授，又注重实践能力的培养。

本书在编写过程中，力求做到概念清晰、重点突出、循序渐进、由浅入深，力求做到既注重理论知识的传授，又注重实践能力的培养。本书在编写过程中，力求做到概念清晰、重点突出、循序渐进、由浅入深，力求做到既注重理论知识的传授，又注重实践能力的培养。

本书在编写过程中，力求做到概念清晰、重点突出、循序渐进、由浅入深，力求做到既注重理论知识的传授，又注重实践能力的培养。本书在编写过程中，力求做到概念清晰、重点突出、循序渐进、由浅入深，力求做到既注重理论知识的传授，又注重实践能力的培养。

本书在编写过程中，力求做到概念清晰、重点突出、循序渐进、由浅入深，力求做到既注重理论知识的传授，又注重实践能力的培养。本书在编写过程中，力求做到概念清晰、重点突出、循序渐进、由浅入深，力求做到既注重理论知识的传授，又注重实践能力的培养。

目录

Contents

第 1 章 数据分析概述	(1)
1.1 何谓数据分析	(1)
1.1.1 数据分析定义概述	(2)
1.1.2 数据分析方法	(4)
1.2 数据分析步骤	(4)
1.2.1 明确目的	(4)
1.2.2 收集数据	(5)
1.2.3 处理数据	(5)
1.2.4 分析数据	(6)
1.2.5 撰写报告	(6)
1.3 数据分析前景	(6)
1.3.1 数据分析的作用	(6)
1.3.2 数据分析就业前景	(7)
习题	(11)
第 2 章 案例背景	(12)
2.1 模拟企业背景介绍	(12)
2.1.1 A 企业概况	(12)
2.1.2 A 企业数据来源说明	(12)
2.2 模拟企业基础数据	(13)
2.2.1 人力资源方面部分数据描述和说明	(13)
2.2.2 商品库存方面部分数据描述和说明	(14)
2.2.3 商品销售方面部分数据描述和说明	(15)
第 3 章 数据处理	(18)
3.1 数据基本概念	(19)
3.1.1 字段与记录	(19)

3.1.2	数据类型	(19)
3.1.3	数据表	(20)
3.2	数据来源	(26)
3.3	数据导入	(26)
3.3.1	文本文件数据导入	(27)
3.3.2	网络数据源导入	(30)
3.4	数据清洗	(33)
3.4.1	重复数据的处理	(33)
3.4.2	缺失数据的处理	(38)
3.5	数据加工	(41)
3.5.1	数据抽取	(41)
3.5.2	字段合并	(43)
3.5.3	字段匹配	(44)
3.5.4	数据计算	(45)
3.5.5	数据分组	(48)
3.5.6	数据转换	(49)
3.6	数据抽样	(50)
	习题	(53)
第4章	数据分析方法	(54)
4.1	常用数据分析术语	(54)
4.1.1	平均数	(54)
4.1.2	绝对数/相对数	(55)
4.1.3	百分比/百分点	(57)
4.1.4	频数/频率	(58)
4.1.5	比例/比率	(58)
4.1.6	倍数/番数	(59)
4.1.7	同比/环比	(60)
4.2	数据基本分析方法	(61)
4.2.1	对比分析法	(61)
4.2.2	分组分析法	(64)
4.2.3	平均分析法	(67)
4.2.4	交叉分析法	(70)
4.2.5	漏斗图法	(72)
4.2.6	矩阵关联分析法	(74)
4.3	数据分析工具——数据透视表	(77)
4.3.1	基本概念	(77)
4.3.2	创建方法	(77)

4.3.3 案例实践	(80)
4.3.4 实用数据透视表技巧	(86)
习题	(92)
第5章 数据展示	(93)
5.1 表格展示	(93)
5.1.1 数据列突出显示	(93)
5.1.2 图标集	(95)
5.1.3 迷你图	(97)
5.2 图表展示	(98)
5.2.1 双坐标	(98)
5.2.2 折线图	(102)
5.2.3 柱状图	(106)
5.2.4 饼图	(111)
5.2.5 旋风图	(113)
5.2.6 瀑布图	(116)
5.2.7 数据地图	(118)
5.2.8 折线图与柱状图的组合	(120)
5.2.9 动态图	(123)
5.3 图表专业化	(129)
5.3.1 基本要素	(130)
5.3.2 基本配色	(131)
5.3.3 商务图表设计	(134)
5.4 拓展运用	(141)
习题	(143)
第6章 数据可视化分析报告	(144)
6.1 数据分析方法论	(144)
6.1.1 5W2H分析法	(144)
6.1.2 SWOT分析法	(146)
6.1.3 4P营销理论	(147)
6.1.4 用户行为分析理论	(147)
6.2 数据可视化分析报告结构	(148)
6.3 案例：人力资源数据分析报告	(150)
6.3.1 分析背景与目的	(150)
6.3.2 分析思路	(150)
6.3.3 人力资源数据分析	(150)
6.3.4 分析总结	(157)
6.4 案例：A公司会员分析报告	(157)

6.4.1	分析背景与目的	(157)
6.4.2	分析思路	(158)
6.4.3	会员客户分析	(158)
6.4.4	分析总结	(163)
6.5	案例: 库存管理数据分析报告	(163)
6.5.1	分析背景与目的	(163)
6.5.2	分析思路	(163)
6.5.3	库存分析	(164)
6.5.4	分析总结	(166)
	习题	(166)
第7章	商业智能仪表板	(167)
7.1	商业智能仪表板简介	(167)
7.2	商业智能仪表板设计要点	(168)
7.3	利用 Excel 制作商业智能仪表板	(169)
7.4	利用 Tableau 制作商业智能仪表板	(171)
7.4.1	Tableau 简介	(171)
7.4.2	利用 Tableau 制作销售分析仪表板	(174)
7.4.3	利用 Tableau 制作会员分析仪表板	(179)
	习题	(184)
参考文献		(185)
4.1.1	平均数	(18)
4.1.2	绝对数/相对数	(15)
4.1.3	百分比/百分点	(17)
4.1.4	频数/频数	(18)
4.1.5	比例/比率	(18)
4.1.6	指数/指数	(19)
4.1.7	同比/环比	(19)
4.2	数据基本分析方法	(19)
4.2.1	对比分析法	(19)
4.2.2	分组分析法	(19)
4.2.3	平均分析法	(19)
4.2.4	交叉分析法	(19)
4.2.5	漏斗图法	(19)
4.2.6	矩阵图分析法	(19)
4.3	数据可视化工具——数据可视化	(19)
4.3.1	基本要素	(19)
4.3.2	创建方法	(19)

第1章

数据分析概述

1.1 何谓数据分析

数据是指以时间为轴，记录人物、地点、事件和方法等生活各个维度的数字字符。数据会随着时间不断累积，也会随着科技、生活观念等变化而呈现出不同的特性。消费者去商场用现金支付的方式购买了一件心仪的衣服，商店的日销售报告中记录了此次交易的金额、数量、款式和型号。当消费者采用的是刷卡的支付方式时，银行的日流水单以及商店的POS机刷卡记录就产生了一笔实时交易数据。如果该消费者还是这家商店的会员，那么该商店就拥有了该消费者部分基本信息以及多次购买产品的交易记录。随着互联网、自动化科技的发展，消费者更多地参与了线上交易，那么线上交易平台会产生消费者常用地址、联系方式、偏好产品、产品型号、消费额度和消费频率等全面而及时的消费数据。

在传统的商业和社会环境下，人们对于数据的利用是非常有限的。企业和商家们利用自身的营销数据汇编成财务报告、信息披露报告，用来为管理层或者潜在投资者提供企业经营状况的参考资料。政府各个机构拥有的人口、宏观指标、地区发展、部门业务发展等各方面的数据为定期政府报告等特定事项提供数据服务。数据成为我们衡量过去发展状况和业绩水平的一种度量衡。传统意义上对于数据利用存在的缺失是不可忽视的。

首先，传统意义上对于数据的利用形成了无数个数据孤岛。宏观数据、调研数据、社会化数据和企业数据之间存在数据孤岛，而政府内部和企业内部同样存在数据孤岛。各个政府部门、甚至每个政府部门内部都有自身因专门的业务内容而产生专项数据，但是这些专项数据仅仅在服从专项需求时才被局部调用。企业内部也是这样，除了必要的信息披露之外，企业各部门之间的数据也是缺乏协同和共享机制的。数据孤岛的存在比

我们想象得还要多而广泛，也极大地降低了对数据的利用。

其次，传统数据存在缺失和错误的比率较大。在商务系统和互联网尚未发展的情况下，企业和社会运营的数据很多是通过人工的方式进行记录的，因此很容易存在数据缺失和数据失实的情况。更重要的是，很多数据的产生需要大量的人力、物力去完成，在不可估量商业价值的情况下，数据的累积往往具有很强的延时性。

再次，数据的价值被低估，缺乏专业的数据分析人员对数据的商业和社会价值进行分析。我们会发现，传统意义上的数据管理是基于某些特定的目的和需求，例如，定期的信息披露，盈余管理和预测等。但是这些目的和需求都不是为了能够创造价值而设立的，更多的是一种业务和管理层面的辅助。在缺乏商业利益动机的前提下，也就没有专业数据分析师存在的必要。

然而，大数据时代拥有的数据量是足够大的。在互联网的世界里，每分钟 Facebook 平均有 600 次的访问量，并有新增用户 28 万；Amazon 每分钟销售高达 8.3 万美元；全球 IP 网一分钟能够传输 639TB 的数据；你需要花费 5 年的时间才能看完互联网上一秒钟传输的视频。同时，大数据时代的数据开始逐步走向多元化的趋势。数据来源包括移动数据、店面交易、网络行为、定位信息、电商、用户调查、社会网络以及企业 CRM 等。

当今社会是一个大数据的社会，信息高度发达，数据信息更是呈爆炸式增长，每天全世界都在产生着巨大的数据，大到一个跨国公司，小到一个社区的小卖部，都不可避免地和各种数据打交道。面对众多的数据，无论是管理者、经营者还是政策的制定者，都面临着管理好数据、发现数据的规律以及从数据中获得价值的问题。在这一章中，我们将首先探讨什么是数据分析。

1.1.1 数据分析定义概述

数据分析是数据的摄取，进行数据分析离不开数据的支持。数据实际上是一种观测值，是信息的外在表现形式，也是实验、策略、观察和调查的结果，以数量形式来表现。原始的数据往往具有数量巨大且杂乱无章的特点，很多时候给人的感觉就是让人眼花缭乱，不知所云。这样的数据是没有任何意义的，需要对其进行分析。

那么数据分析就是将大量且杂乱无章的数据进行整理、归纳和提炼，从中寻找出数据的内在规律，从而获得需要的信息。数据分析的过程，实际上就是对数据进行汇总和理解吸收的过程，也是为了提取有用的信息和形成结论而对数据加以研究和概括总结的过程。通过对数据进行分析，以求最大化地开发数据，发挥数据的作用。

数据分析是一种有组织有目的处理数据并使数据成为信息的过程，其根本目的是集中、萃取和提炼。在实际工作中，其最终是为了帮助经营者和决策者做出判断，以便采取正确有效的行动。例如，企业的高层希望通过市场分析和研究，把握当前产品的市场动向，从而制订合理的产品研发和销售计划。因此，在经济生活中，经济决策实际上就是一种“数据决策”，“用数据说话”是众多企业经营者和决策者的共识。

大数据时代，数据的形态和体量都发生了很大的变化，缺乏数据分析的数据本身是

不具备商业价值的。数据分析的确能够为大数据时代带来质的飞跃。常规报表、查询、多维分析、警报——这数据分析的前四个等级都只能展示已经发生的历史状况，但是数据分析不仅仅如此。统计分析能够帮助我们找到触发事件发生的相关因素、确认最为有效的潜在交易方案。预报可以告诉我们未来股市预期变动或者是企业未来盈利水平预期。预测建模可以帮助金融机构预测新的金融产品的潜在客户。运筹优化能够帮助企业在限定的条件下把握最优的业务机会。

数据分析的核心思路就是要与实际业务、商业目的和运营目标相结合，进而为社会、经济和个体创造价值。数据分析与业务流程相结合可以体现为五个基本步骤，包括认知、运营、交互、销售和维护。商业运营要与数据分析的关键指标紧密联系，用数据提高产品市场营销效率和推广效率。大数据的维护和累积能够为商业运营描绘完整的企业画像、客户画像。大数据画像包括了解企业或者客户的基本信息、需求倾向、用户行为等。通过追踪核心的数据指标，进一步完善企业或者客户画像，进而将其转化成为产品元素和营销战略。通过数据分析，我们可以知道通过什么渠道以最小的成本将竞争对手的客户转化为自身的客户，进而创造营业收入。通过大数据与运营维护的结合可以很大程度上提高客户满意度，降低客户的流失率。

目前数据分析实践的运用主要体现在物联网、定位服务、客户制成以及反欺诈领域。首先是物联网领域。以 UPS（美国快递公司）为例，UPS 每天通过 5 万台快递车派送约 1630 万个包裹。UPS 在每台快递车上都安装了传感器，并且通过传感器传输数据分析，制定每天每台车少跑一英里的运营战略，该战略为 UPS 每年实现了约 3000 万美元的盈利。其次是定位服务。以美洲银行为例，美洲银行为其客户提供汉堡王的优惠券。该优惠券以美洲银行客户刷卡记录数据为基础，判断汉堡王潜在竞争对手的客户，并对这些客户进行了定向、定位的优惠券推送。该项营销战略既维护了美洲银行客户，也为汉堡王实现了创收。再次是客户支撑。通过文本挖掘、自然语言处理、情感分析等手段，对客户评论、客户投诉、海外舆情、媒体报道数据进行分类处理，进而充分掌握客户潜在的需求，达到及时有效维护客户的商业目的。最后是反欺诈领域。最典型的例子就是保险公司骗保。我们可以通过神经网络分析等多元的数据分析方法及时识别和判断已有的欺诈模式和潜在的欺诈人群，进而有效地进行客户管理，确保企业运营和效益。

在传统的数据分析模式下，通常是先提出假设检验，然后带着问题去进行数据分析。在大数据时代下，更重要的是关注小数据完善和收集的同时，构建完善的数据交互平台。在先有数据的基础上，在数据中找寻新的思路和创新机遇，进而实现价值的飞跃。在数据爆炸和新媒体时代的背景下，文字、图片、视频、网络数据等新兴的数据模式使得我们需要掌握和运用全新的数据处理方式。同时，还需要对数据进行生命周期的管理，对非结构数据进行筛选和标签化。数据分析看重的是数据的多元性和数据的质量，需要构建起大数据谱系，同时结合数据的特性采用不同的数据分析方法、分析工具和分析模型。因此，数据分析需要较为综合的思维和能力。

1.1.2 数据分析方法

对数据进行分析的方法有很多,归纳起来主要包括统计分析方法、运筹学分析方法、财务分析方法和图表分析方法。下面分别简单介绍以上所述分析方法。

(1) 统计分析方法:是指对收集到的数据进行整理归类并解释的分析过程,主要包括描述性统计或推断性统计。其中,描述性统计以描述和归纳数据的特征以及变量之间的关系为目的,主要涉及数据的集中趋势、离散程度和相关程度,其代表性指标是平均数、标准差和相关系数等。推断性统计是用样本数据来推出总体特征的一种分析方法,包括总体参数估计和假设检验,代表性方法是Z检验、T检验和卡方检验等。

(2) 运筹学分析方法:是在管理领域中运用的数学方法,该方法能够对需要管理的对象(如人、财务和物等)进行组织从而发挥最大效益。运筹学分析常使用数学规划分析,如线性规划、非线性规划、整数规划和动态规划等,也可以运用运筹学中的理论(如图论、决策论和库存论等)来进行分析预测。运筹学分析方法常用在企业的管理中,如服务、库存、资源分配、生产和产品可靠性分析等诸多领域。

(3) 财务分析方法:是以财务数据及相关数据为依据和起点来系统分析和评估企业过去和现在的经营成果、财务状况以及变动情况,从而了解过去、分析现在和预测未来,达到辅助企业的经营和决策的目的。财务分析方法包括比较分析法、趋势分析法和比率分析法等。

(4) 图表分析方法:是一种直观形象的分析方法,其将数据以图表的形式展示出来,使数据形象、直观和清晰,让决策者更容易发现数据中的问题,提高数据处理和效率。图表分析主要针对不同的数据分析类型,采用不同的图表类型将数据单独或组合展示出来,常见的图表如柱形图、条形图、折线图和饼图等。

1.2 数据分析步骤

数据分析通常可以分为明确目的、收集数据、处理数据、分析数据和撰写报告这几个步骤。

1.2.1 明确目的

在进行数据分析时,首先需要明确分析的目的。在接收到数据分析的任务时,首先需要搞清楚为什么要进行这次分析、这次数据分析需要解决的是什么问题、应该从哪个方面切入进行分析以及什么样的分析方法最有效等问题。在确定总体目的后,可以对目标进行细化,将分析的目标细化为若干分析要点,厘清具体的分析思路并搭建分析框架,搞清楚数据分析需要从哪几个角度来进行,采用怎样的分析方法最有效。

只有这样才能为接下来的工作提供有效的指引，保证分析完整性、合理性和准确性，使数据分析能够高效进行，分析结果保证有效和准确。

总之，在开展数据分析之前，要考虑清楚，为什么要开展数据分析？通过这次数据分析要解决什么问题？只有明确数据分析的目标，数据分析才不会偏离方向，否则得出的数据分析结果不仅没有指导意义，甚至可能将决策者引入歧途，导致严重的后果。

1.2.2 收集数据

收集数据是在明确数据分析的目的后，获取需要数据的过程，其为数据的分析提供直接的素材和依据。收集数据的来源包括两种方式，第一种方式就是直接来源，也称为第一手数据，数据来源于直接的调查或现实的结果。第二种方式称为间接数据，也可称为第二手数据，数据来源于他人的调查或实验，是结果加工整理后的数据。通常，数据来源主要有以下几种方式：

(1) 公司或机构自己的业务数据库，存放着大量相关业务数据。这个业务数据库就是一个庞大的数据资源，需要有效地利用起来。

(2) 公开出版物，比如《中国统计年鉴》、《中国社会统计年鉴》、《世界经济年鉴》、《世界发展报告》等统计年鉴或报告。

(3) 互联网，网络上发布的数据越来越多，特别是搜索引擎可以帮助我们快速找到所需要的数据。例如，国家及地方统计局网站、行业协会网站、政府机构网站、传播媒体网站和大型综合门户网站等，上面都可能有我们需要的数据。

(4) 进行市场调查。在数据分析时，如果要了解用户的想法和需求，通过以上三种方式获得数据会比较困难，这时就可以采用市场调查的方法收集用户的想法和需求数据。市场调查是指运用科学的方法，有目的、有系统地收集、记录、整理有关市场营销的信息和资料，分析市场情况，了解市场现状及其发展趋势，为市场预测和营销决策提供客观、正确的数据资料。市场调查可以弥补其他数据收集方式的不足，但进行市场调查所需的费用较高，而且会存在一定的误差，故仅作参考之用。

所以，在实际工作中，获取数据的方式有很多，根据不同的需要有不同的获取途径，如对本公司的经营状况的分析，可以从公司自由的业务数据库获取。对于一些专业数据，可以从公开的出版物获取，如年鉴或分析报告等。随着互联网的发展，获取数据的途径更为广阔，通过搜索引擎，可以快速找到需要的数据，如到国家或地方统计局的网站、行业组织的官方网站或行业信息网站等。

1.2.3 处理数据

在获得数据后，需要对数据进行处理。数据处理是指对收集到的数据进行加工整理，形成适合数据分析的样式，它是数据分析前必不可少的一步工作。数据处理的基本目的是从大量的、杂乱的且难以理解的数据中抽取并推导出对解决问题有价值

意义的数

数据处理常常需要对数据进行清理、转换、提取、汇总和计算。通常情况下,收集到的数据都需要进行一定的处理才能用于后面的数据分析工作,即使是再“干净”的原始数据也需要先进行一定的处理才能使用。

1.2.4 分析数据

数据分析需要从数据中发现有关信息,一般需要通过软件来完成。在进行数据分析时,数据分析人员根据分析的目的和内容确定有效的数据分析方法,并将这种方法付诸实施。当前数据分析一般都是通过软件来完成的,简单实用的有大家熟悉的 Excel,专业高端的软件有 SPSS 和 SAS 等。

1.2.5 撰写报告

在完成数据分析后,需要将分析结果展示出来并形成分析报告。数据分析报告一般包括封面、目录、分析内容和总结这几个部分。数据分析报告是对数据分析过程的总结和归纳,分析报告需要描述出数据分析的过程和分析的结果,并且要给出分析的结论。数据分析报告应该结构清晰且主次分明。分析报告应该具有一定的逻辑性,一般可以按照发现问题、总结问题原因和解决问题这一流程来描述。在分析报告中,每一个问题都必须要有明确的结论,一个分析对应一个结论,切忌贪多,结论应该基于严谨的数据分析,不能主观臆测。同时,分析报告应该通俗易懂,使用图表和简洁的语言来描述,不要使用过多的专业名词,要让看报告的人能够看懂。

最后,好的分析报告一定要有建议和解决方案。作为决策者,需要的不仅仅是找出问题,更重要的是建议或解决方案,以便决策者在做决策时参考。所以,数据分析师不仅需要掌握数据分析方法,还要了解和熟悉业务,这样才能根据发现的业务问题,提出具有可行性的建议或解决方案。

1.3 数据分析前景

1.3.1 数据分析的作用

数据分析在管理上有着十分重要的作用,它产生的价值来源于详尽而真实的数据,是一个企业的管理走向正规化、决策走向合理化的重要环节。数据分析在实际工作中能够及时纠正经营和生产中的错误,使企业的管理者能够了解企业现阶段的经营状况,知道企业业务的发展和变动的情况,及时对企业的运营有一个深入了解。通过数据分析,

可以对企业的计划进度进行分析,实时了解经营情况。同时,在了解企业当前状况的同时,提供了科学管理的依据。数据分析也可以有效帮助决策者对未来的发展趋势进行预测,为制定经营方向、运营目标以及决策提供有效的参考和依据,最大限度地规避风险。

1.3.2 数据分析就业前景

就算你不是数据分析师,数据分析技能也是未来必不可少的工作技能之一。在数据分析行业发展成熟的国家,90%的市场决策和经营决策都是通过数据分析研究确定的。

数据分析师,是指在互联网、金融、电信、医疗、旅游、零售等多个行业专门从事数据的采集、清洗、处理、分析,能够利用统计数据、定量分析和信息建模等技术制作业务报告、进行行业研究、评估和预测,从而为企业或所在部门提供商业决策的新型数据分析人才。

2015年2月,美国白宫正式命名DJ Patil担任首席数据科学家和制定数据策略的副首席技术官。DJ Patil曾在LinkedIn、eBay、PayPal、Skype和风险投资公司Greylock Partners等诸多硅谷知名公司工作过,积累了丰富的经验,在上任之后将会扮演负责政府大数据应用开发专家的角色,尤其是针对奥巴马的医疗改革方案。美国政府正在用实际行动告诉全世界,其已经意识到要充分利用他们的数据。

IDC(互联网数据中心)预测,目前每年数据的生产量超过8ZB,2020年将达到40ZB,如图1-1所示,属于大数据的时代已经到来。

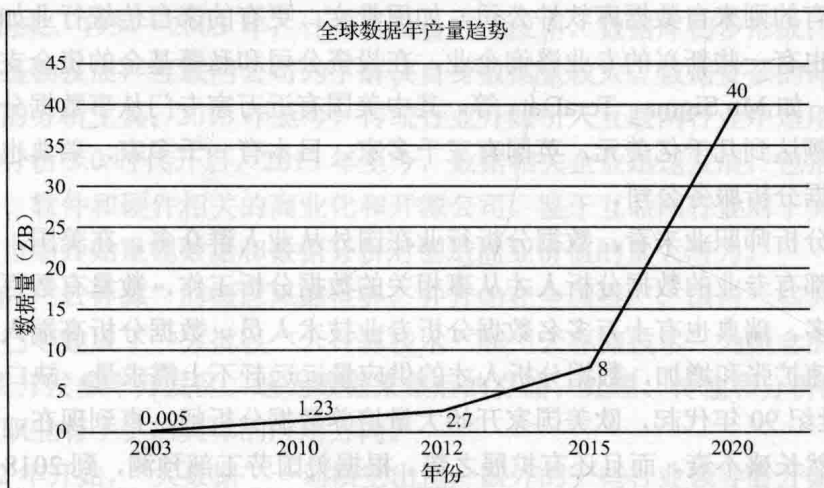


图 1-1 全球数据年产量趋势

数据生产量“拐点”已至,将开始爆发式增长。我们正处在一个数据量爆发增长的时代,当今的信息产业呈现出前所未有的繁荣,新的互联网技术不断涌现,从传统互联网的PC终端,到移动互联网的智能手机,再到物联网传感器,技术革新使数据生产能力呈指数级提升。