



数据分析 从入门到进阶

陈红波 刘顺祥 等编著

- 系统介绍数据分析的策略、方法（数据分析之道）
- 实例讲解引导，快速掌握数据分析工具（数据分析之术）
- 免费提供本书案例涉及的全部数据集、程序代码



机械工业出版社
CHINA MACHINE PRESS

数据分析从入门到进阶

陈红波 刘顺祥 等编著



机械工业出版社

本书由资深数据分析师精心编写，通过大量案例介绍了数据分析工作中常用的数据分析方法与工具。本书包括 5 章内容，分别是数据分析入门、数据分析——从玩转 Excel 开始、海量数据管理——拿 MySQL 说事儿、数据可视化——Tableau 的使用、数据分析进阶——Python 数据分析。本书通俗易懂、通过大量贴近企业真实场景的案例，帮助读者在提高数据处理技能的同时加深对数据分析思维的理解。

本书适合有志于从事数据分析工作或已从事初级数据分析工作的人士自学，也可作为产品经理、运营人员、市场人员、对数据分析感兴趣的企业高管以及创业者的参考用书。

图书在版编目 (CIP) 数据

数据分析从入门到进阶 / 陈红波等编著. —北京: 机械工业出版社, 2019.6
ISBN 978-7-111-62882-8

I. ①数… II. ①陈… III. ①数据处理 IV. ①TP274

中国版本图书馆 CIP 数据核字 (2019) 第 105979 号

机械工业出版社 (北京市百万庄大街 22 号 邮政编码 100037)

策划编辑: 王 斌 责任编辑: 王 斌

责任校对: 张艳霞 责任印制: 张 博

三河市国英印务有限公司印刷

2019 年 7 月第 1 版第 1 次印刷

184mm×260mm·24.25 印张·4 插页·602 千字

0001—3000 册

标准书号: ISBN 978-7-111-62882-8

定价: 79.90 元

电话服务

客服电话: 010-88361066

010-88379833

010-68326294

封底无防伪标均为盗版

网络服务

机 工 官 网: www.cmpbook.com

机 工 官 博: weibo.com/cmp1952

金 书 网: www.golden-book.com

机工教育服务网: www.cmpedu.com

前言

随着大数据技术的快速发展，人们对数据的价值越来越重视，数据采集、存储、安全技术也变得日益重要，数据分析和数据挖掘技术得到了日益广泛的应用。利用数据分析技术从海量数据中提取的信息具有极高的价值，例如，支持企业高层进行业务决策、发现新的销售和市场机会、提升组织的社交媒体营销能力、提高用户忠诚度以及复购率、降低用户流失率、提前预测风险并进行防范等。

对于数据的重视以及数据分析技术的发展与应用，带动了企业对数据分析人才需求的快速增长。在未来一段时间内，数据分析人才缺口会很大。从事数据分析工作需要专门的技能，一名优秀的数据分析师既要熟练掌握数据分析之“道”——数据分析的策略、方法（也可以将其理解为做数据分析的思路），也要熟练掌握数据分析之“术”——数据分析工具的使用。此外，数据分析师还需要熟悉行业知识、公司业务及流程，了解企业产品和运营活动的设计思路，才能根据数据分析的结论驱动业务增长落地，使得数据分析工作的价值和自身的价值得以实现。

本书结合数据分析工作的实际情况，通过大量案例介绍了数据分析的方法和工具，内容涵盖了 Excel、SQL、Tableau 以及 Python 这几个常用的数据分析工具的使用，融会贯通地介绍了数据分析的道与术。通过本书，读者可以由浅入深、循序渐进地学习数据分析，为日常工作中数据的处理与分析打下坚实的基础。

本书内容

第 1 章数据分析入门。主要内容包括什么是数据分析、数据分析的职业发展及分类，以及数据分析之“道”（数据分析需要掌握的理论知识）与数据分析之“术”（各类软件工具的运用）。

第 2 章数据分析——从玩转 Excel 开始。主要内容包括 Excel 概述、高效处理数据的 Excel 函数家族、十分有用的 Excel 数据分析技巧、酷炫的 Excel 图表可视化、让你的 Excel 报表动起来（VBA）。

第 3 章海量数据管理——拿 MySQL 说事儿。主要内容包括 MySQL 数据库的安装、将数据写入到数据库中、重要的单表查询、复杂的多表查询、通过索引提高数据的查询速度、数据库的增删改操作。

第 4 章数据可视化——Tableau 的使用。主要内容包括数据可视化概述、Tableau 概述、数据可视化图表、仪表板的制作与发布。

第5章数据分析进阶——Python 数据分析。主要内容包括数据分析的利器——Python、Jupyter 的使用技巧、数据读取——从 pandas 开始、常见的数据处理技术、探索性数据分析、线性回归模型的应用。

本书特点

- 由浅入深，循序渐进：本书在简要概述了数据分析的基本概念之后，首先讲解了数据分析入门工具 Excel 的操作技巧，然后结合案例讲解了 VBA 的知识点，帮助读者快速掌握表格处理技术；结合 MySQL 数据库对 SQL 语言的讲解可以让读者轻松地处理海量数据；Tableau 是用来进行数据可视化分析的重要工具；Python 作为本书的进阶部分内容，可以帮助读者高效处理数据和通过建模进行数据分析。书中讲解的知识点环环相扣、逐层深入，比较符合初学者学习数据分析的认知规律。
- 案例丰富，轻松易学：本书在介绍各类数据分析工具时结合了大量的实际案例，能够让读者快速理解并掌握各个知识点，简单易学、轻松上手。
- 内容全面，讲解详细：本书定位在数据分析的入门与进阶，从数据分析理论到数据处理、从可视化分析到建模分析，知识点覆盖得很全面。全书最后附有彩插，将书中对应效果图直观呈现，方便读者参考。
- 配套资源丰富，免费提供：本书中的案例涉及的数据集、代码等资源都免费提供给读者学习使用，可通过扫描封底二维码 IT 有得聊，并输入本书书号中的五位数字获取。

适用对象

本书适合有志于从事数据分析工作或已从事初级数据分析工作的人士自学，也适合产品经理、运营人员、市场人员、对数据分析感兴趣的企业高管以及创业人员等参考。

本书作者

本书由陈红波、刘顺祥等编著，参与本书编写的人员还有孙宗鹏、朱烨、陶颖。此外，还要对帮助本书出版的所有朋友致以衷心的感谢！由于作者水平有限，书中难免出现错误和不足之处，敬请广大读者批评指正。

希望本书能够成为您数据分析入门的领航者。

陈红波 刘顺祥
2019年1月11日

目录

前言

第 1 章 数据分析入门	1
1.1 什么是数据分析	1
1.1.1 数据分析的含义	1
1.1.2 数据分析的操作步骤	2
1.1.3 数据分析的应用场景	5
1.1.4 数据分析的发展趋势	8
1.2 数据分析的职业发展及分类	8
1.3 数据分析之道	10
1.3.1 三类统计分析策略	10
1.3.2 数据分析的常用方法	17
1.4 数据分析之术	22
1.4.1 必备的 Excel 处理方法	22
1.4.2 高超的数据库查询技巧	23
1.4.3 纯熟的数据可视化技能	23
1.4.4 高大上的统计编程技术	26
第 2 章 数据分析——从玩转 Excel 开始	28
2.1 Excel 概述	28
2.1.1 强大的数据处理技能	28
2.1.2 实用的数据分析技巧	29
2.1.3 丰富的数据可视化图表	30
2.1.4 便捷的自动化数据处理	30
2.1.5 快速实现业务报表开发	32
2.2 高效处理数据的 Excel 函数家族	34
2.2.1 常用的统计分析函数	35
2.2.2 灵活的文本处理函数	49
2.2.3 便捷的数值运算函数	59
2.2.4 经典的逻辑判断函数	68
2.2.5 实用的日期计算函数	74

2.2.6	高效的匹配查找函数	80
2.3	十分有用的 Excel 数据分析技巧	94
2.3.1	数据透视表实现统计分析	95
2.3.2	描述性统计分析	113
2.3.3	相关系数与协方差	115
2.3.4	线性回归模型预测	118
2.3.5	移动平均模型预测	121
2.4	酷炫的 Excel 图表可视化	123
2.4.1	Excel 基础图表	123
2.4.2	高级可视化图表	135
2.5	让你的 Excel 报表动起来	141
2.5.1	VBA 基础语法	142
2.5.2	录制宏，解放你的双手	152
2.5.3	VBA 应用：学生成绩信息统计	154
2.5.4	VBA 应用：报表发送前的一键优化	155
2.5.5	VBA 应用：数据库字典的超链接	156
2.5.6	VBA 应用：一键合并拆分工作簿	158
2.5.7	VBA 应用：从数据库获取并更新数据	163
第 3 章	海量数据管理——拿 MySQL 说事儿	174
3.1	MySQL 数据库的安装	174
3.2	将数据写入到数据库中	178
3.2.1	常用的数据类型	178
3.2.2	手工建表	181
3.2.3	数据插入	183
3.2.4	外部数据的批量导入	186
3.3	重要的单表查询	191
3.3.1	SQL 查询的 7 个核心关键词——以链家二手房数据为例	192
3.3.2	基于 CASE WHEN 的常用查询——以电商交易数据为例	201
3.3.3	几种常见的嵌套查询——以学员考试成绩为例	205
3.3.4	基于分组排序的辅助列功能——以销售员业绩数据为例	210
3.4	复杂的多表查询	214
3.4.1	纵向表合并——以超市交易数据为例	215
3.4.2	表连接操作——以校园一卡通记录数据为例	217
3.5	通过索引提高数据的查询速度	224
3.5.1	常见的索引类型	224
3.5.2	索引的查询和删除	229

3.5.3	关于索引的注意事项	230
3.6	数据库的增删改操作	231
3.6.1	数据库的增操作	231
3.6.2	数据库的删操作	232
3.6.3	数据库的改操作	235
第 4 章	数据可视化——Tableau 的使用	238
4.1	数据可视化概述	238
4.1.1	什么是数据可视化	238
4.1.2	为什么要实现数据可视化	240
4.2	Tableau 概述	241
4.2.1	软件安装	241
4.2.2	连接数据源	243
4.2.3	数据源界面	249
4.2.4	工作区界面	251
4.2.5	高级操作	259
4.2.6	保存结果	264
4.3	数据可视化图表	265
4.3.1	条形图	266
4.3.2	柱形图	267
4.3.3	折线图	268
4.3.4	面积图	271
4.3.5	符号图	272
4.3.6	树状图	272
4.3.7	气泡图	273
4.3.8	文字云	273
4.3.9	饼图	274
4.3.10	热图	274
4.3.11	盒须图	275
4.3.12	双轴图	276
4.3.13	动态图表	277
4.3.14	参数图表	279
4.3.15	漏斗图	282
4.4	仪表板的制作与发布	286
4.4.1	仪表板的制作	286
4.4.2	可视化成果发布	288
第 5 章	数据分析进阶——Python 数据分析	291

5.1	数据分析的利器——Python	291
5.1.1	Anaconda——Python 集成开发环境的安装	292
5.1.2	Python 编程工具的选择	295
5.2	Jupyter 的使用技巧	297
5.2.1	代码运行组合键	297
5.2.2	代码框操作组合键	297
5.2.3	注释组合键	297
5.2.4	帮助组合键	298
5.2.5	代码与笔记的切换组合键	298
5.3	数据读取——从 pandas 开始	299
5.3.1	文本文件的读取	299
5.3.2	电子表格的读取	301
5.3.3	数据库数据的读取	303
5.4	常见的数据处理技术	306
5.4.1	数据的概览与清洗	306
5.4.2	数据的引用	312
5.4.3	多表合并与连接	314
5.4.4	数据的汇总	319
5.5	探索性数据分析	324
5.5.1	异常数据的检测与处理	324
5.5.2	数据的描述	331
5.5.3	数据的推断	348
5.6	线性回归模型的应用	358
5.6.1	简单线性回归模型——刹车距离的研究	359
5.6.2	多元线性回归模型——产品市场销售额的研究	363
5.6.3	模型的显著性检验—— F 检验	365
5.6.4	回归系数的显著性检验—— t 检验	367
5.6.5	基于回归模型识别异常点	370
5.6.6	模型的预测	372

第 1 章

数据分析入门

大数据技术的发展使得数据采集、存储、安全技术变得越来越成熟，并且当前人们对于采集到的各种数据的价值越来越重视，从而带动了数据分析技术的发展。数据分析初学者首先得了解什么是数据分析，然后明确自身的职业发展方向，逐步掌握数据分析理论及数据分析工具的使用，夯实数据分析相关技能，再结合企业实际业务进行数据分析挖掘，从而实现业务流程优化，提高工作效率，并能辅助企业基于数据对市场变化进行快速判断，以便采取有效的行动。

通过本章内容的学习，读者将会掌握如下几个方面的知识点：

- 数据分析的概念及操作步骤。
- 数据分析的应用场景及发展趋势。
- 数据分析的职业发展及分类。
- 数据分析常用的策略及方法。
- 数据分析常用工具及介绍。

1.1 什么是数据分析

1.1.1 数据分析的含义

数据分析是指使用适当的统计分析方法对收集的大量数据进行分析，将隐没在一大批看似杂乱无章的数据中的有价值的信息进行整合并提炼出来，找出所研究对象的内在规律。

一般情况下，初期收集的原始数据都是相对比较粗糙的，需要通过一定的技术手段进行加工，最后提炼出方便用户理解的知识。如图 1-1 所示，底层的粗

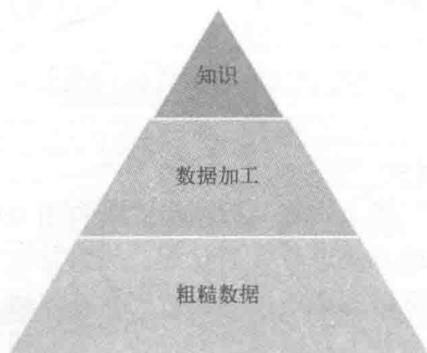


图 1-1 数据分析的含义

糙数据经过一系列的加工处理，然后将处理产生的相关信息与实际业务相结合，进行规律性总结，生成知识（解决方案或商业预测）。

实践证明，数据分析是非常有价值的，它与现实生活是密切相关的。例如，信用卡的审批额度、电商网站对消费者的产品推荐、游戏活动的奖品设置、超市的捆绑式促销、病人疾病的诊断预测等，数据分析可以渗透到这些业务环节中，帮助实现业务流程优化，提高工作效率，并能辅助用户进行快速判断，以便采取有效的行动。

1.1.2 数据分析的操作步骤

数据分析有一套比较规范的操作步骤，作为数据分析人员必须掌握好它，才能减少工作失误，提高工作效率。数据分析的操作步骤如图 1-2 所示。

1. 明确目的

数据分析的第一步就是要明确分析目的。和大家在生活中处理某件事情一样，先确定目标，然后再去动手实施。例如，某电商 APP 上线后，前期导入了大量新用户但是用户质量不是很好（包括登录、付费等表现），需要通过数据分析查明原因。作为数据分析师，需要明确此次分析的目的是找出大量注册用户登录时间不长、付费金额低的原因，可以首先从注册用户本身的属性着手展开分析。

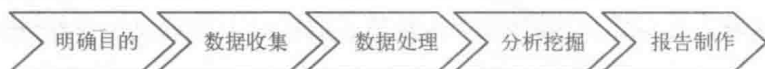


图 1-2 数据分析的操作步骤

2. 数据收集

数据是进行数据分析的前提，“巧妇难为无米之炊”说的就是这个道理。因此，数据的收集显得尤为重要。按收集方式的不同，数据收集可以分为线上收集和线下收集；按收集渠道的不同，又可以分为内部收集和外部收集。数据收集的两种不同分类如图 1-3 所示。

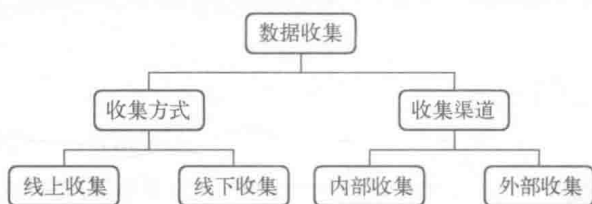


图 1-3 数据收集的分类

线上收集的数据指的是利用互联网技术自动采集的数据。例如，企业内部通过数据埋点的方式进行数据收集，然后将收集来的数据存储到数据库中。此外，利用爬虫技术获取网页数据或借助第三方工具获取网上数据等都属于线上收集方式。一般情况下，互联网科技企业、互联网电商企业、互联网游戏企业等都采用此种方式收集用户行为数据，因为其效率高且错误率较低。

线下收集的数据相对比较传统，对技术要求不高。例如，通过传统的市场调查问卷获取

数据即为线下收集。此外，通过手工录入获取数据、出版物收集的权威数据以及通过其他人提供的电子表格获取数据等都属于线下收集方式。这种收集数据的方式效率低且容易出现偏差。一般情况下，传统制造型企业、线下零售企业、市场调研咨询类企业等都采用此种方式收集数据。

内部收集的数据指的是获取的数据都来源于企业内部数据库、日常财务数据、销售业务数据、客户投诉数据、运营活动数据等。此类数据的获取相对较为方便，数据分析人员可以根据实际业务需求对内部收集的数据进行处理分析。

外部收集的数据指的是数据不是企业内部产生的，而是通过其他手段从外部获取的。例如，利用爬虫技术获取的网页数据，从公开出版物收集的权威数据，市场调研获取的数据以及第三方平台提供的数据等。外部数据的收集不像内部收集那么容易，且大部分都是碎片化、零散的数据。因此，数据分析人员需要对这些数据进行清洗和整合，然后再去进行分析。

总之，不管以何种方式收集过来的数据，都是企业宝贵的财富。数据分析人员需要多和这些数据打交道，多去研究数据背后隐藏的规律，为业务决策提供支持。

3. 数据处理

一般通过不同途径收集过来的原始数据都是相对比较粗糙且无序的。此时，需要利用数据处理软件进行一系列的加工处理，降低原始数据的复杂程度，最终汇总成用户可以解读的业务指标。数据处理包括前期的脏数据清洗、缺失值填充、数据分组转换、数据排序筛选等，后期的业务指标计算、报表模板填充等。常用的数据处理工具包括 Excel 之类的电子表格软件、各类数据库软件、Python、R、SAS、SPSS 等，这些工具都包含数据处理模块，方便用户对数据进行快速清洗，然后进行分析。

4. 数据分析与数据挖掘

基于处理好的数据，数据分析人员才可以对其进行分析和挖掘，结合实际业务得出相关结论，提供给管理层进行决策。因此，数据分析人员需要掌握数据分析和数据挖掘的常用方法，才能为后期的数据报告的制作打下坚实的基础。

数据分析的侧重点在于对业务的熟练掌握，一个优秀的数据分析人员往往对公司业务了如指掌。例如，产品日常活动的前期设计、中期上线跟踪、后期效果评估以及最终的建议与反馈等，数据分析人员都要非常熟悉。当然，除了熟练掌握业务之外，数据分析人员对数据分析常用的分析策略和分析方法也必须掌握。一般情况下，数据分析策略分为描述性统计分析、探索性统计分析、推断性统计分析，如图 1-4 所示。

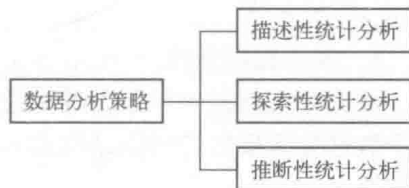


图 1-4 数据分析常用的分析策略

常用数据分析方法包括对比分析法、分组分析法、预测分析法、漏斗分析法、A/B 测试分析法，如图 1-5 所示。通过这些数据分析方法，可以挖掘出数据隐藏的价值，从而降低企业成本，提高营业利润等。

数据挖掘的侧重点在于对模型和算法的理解，一个优秀的数据分析人员必须拥有扎实的数学基础和熟练的编码能力。数据的复杂性、多样性、动态性等特点会使得数据挖掘变得很

困难。因此，在数据挖掘过程中，应该要清楚每一步需要做什么，达到什么样的效果，有问题及时调整方案策略，从而确保整个数据挖掘项目的最终成功。

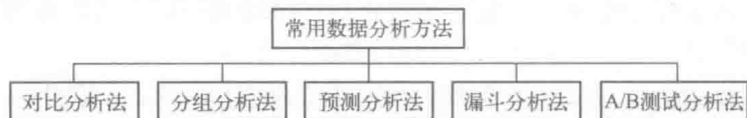


图 1-5 数据分析常用的分析方法

数据挖掘的规范化步骤可以采用 SIG 组织在 2000 年推出的 CRISP-DM 模型，如图 1-6 所示。该模型将数据挖掘项目的生存周期定义为六个阶段。六个阶段分为商业理解 (Business Understanding)、数据理解 (Data Understanding)、数据准备 (Data Preparation)、建立模型 (Modeling)、模型评估 (Evaluation)、结果部署 (Deployment)。数据挖掘的流程并非要完全参照这个顺序执行，数据分析人员可以根据实际业务场景进行调整，通过不断地测试和验证，才能做好一个完整的数据挖掘项目。此外，数据挖掘具有循环特性，并不是一次部署完就结束挖掘过程，需要通过不断的迭代优化，获得最优结果。

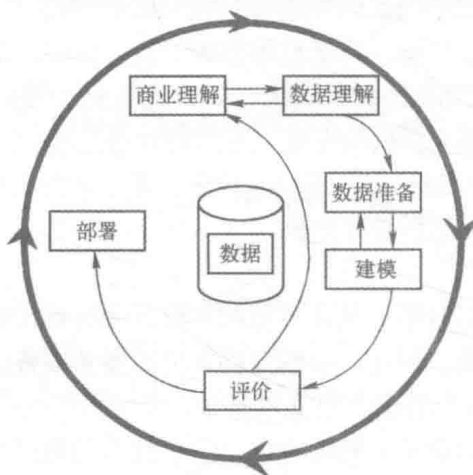


图 1-6 CRISP-DM 模型流程图 (来源网络)

5. 制作数据报告

数据分析的最终结果需要汇总成一份数据报告，最常见的是 PPT 格式的报告。因此，关于数据分析报告的制作显得尤为重要。数据分析报告的制作要求目的明确、结构清晰、有理有据。

报告开始部分一般为目录和前言，简单扼要地列出本次汇报需要陈述的章节；中间部分为正文，主要是对目录的各章节节点展开叙述；结尾部分进行报告总结并提出相关建议和解决措施。数据分析报告的结构如图 1-7 所示。

开始部分的目录是数据分析报告的整体纲要，要求简洁扼要、结构清晰、逻辑有序，让阅读者能快速了解整个汇报的内容。目录切记要归纳总结，不要分太多章节，大致包含分析

目的、分析要点、结论与建议。前言是对分析报告的目的、背景、思路、方法、结论等内容的基本概括，然后引出分析报告的正文内容。

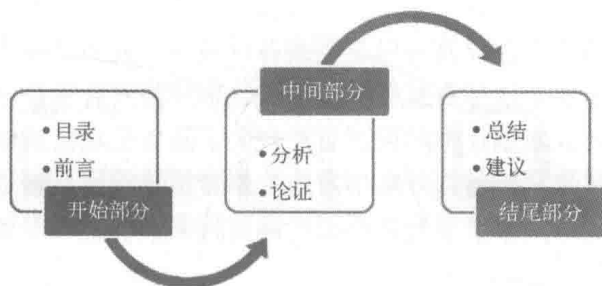


图 1-7 数据分析报告结构图

正文部分的分析和论证是数据分析报告的核心部分，按目录的章节排序分别进行阐述，详解分析思路并进行论证。分析和论证要求条理清晰、层层剖析、有理有据。

结尾部分的结论和建议是依据前面的分析结果得出的相关结论。结论要求准确、干练、有价值，切不可冗余拖沓。在准确的结论基础之上，提出自己的见解和建议，为管理者进行决策时提供参考依据。

最后，数据分析报告的风格要前后一致，内容也可以加入一些动态展示效果，让阅读者赏心悦目，心情舒畅。当然，数据分析报告的核心还是分析、结论与建议，过分重视分析报告的美观程度而忽视分析报告的本质是不可取的。数据分析人员应抱着科学严谨的态度，将对业务的理解与分析挖掘技术相结合，得出可靠且令人信服的分析报告，提供给管理层进行业务决策。

1.1.3 数据分析的应用场景

随着大数据分析技术的发展，日常生活、政府及企业对数据分析的应用需求越来越多，利用大数据分析技术可以为人们的日常生活带来便利，促进社会变革和科技发展。日常生活方面，人们的衣食住行都离不开数据的支撑，大数据分析技术的发展正在逐渐改变人们的生活方式。政府方面，国家出台了一系列政策，明确提出实施国家大数据战略，建设数据强国的目标。政府在交通、天气、农牧业、医疗卫生、教育等众多领域出台了一系列的大数据应用政策，支持其发展。企业方面，公司都在积极架构大数据分析部门，支持其他业务部门的发展，旨在降低企业运营成本、提高企业运作效率、创造更多利润。下面介绍数据分析众多的应用场景，可以看到数据分析与人们的日常生活结合得有多么紧密。

1. 日常生活应用场景

(1) 电商购物

电商平台的崛起让用户不需要出门就能购买到自己需求的商品，提高了用户购买日常生活物品的便利性。移动互联网技术的发展让用户可以随时随地购物，但这一切的发展都离不开大数据技术的支撑。随着数据量的日益增长，包括大数据存储、大数据处理、大数据分析在内的各类大数据技术也在不断发展。利用大数据分析技术，电商企业可以对用户的偏好进

行分析，然后进行商品推荐，从而提高用户的购买效率；电商企业对用户反馈的评论进行收集并分析，可以用来对产品进行优化，从而提高用户对产品的体验。

(2) 外卖订购

外卖平台的发展让消费者在家就可以享受到各种美食，随着平台用户规模的增加，大数据技术的支撑显得尤为重要，包括商家数据的接入、客户消费订单数据、定位信息数据以及实时外卖路线规划等都与大数据技术的应用息息相关。随着市场需求的升级，餐饮外卖行业的发展环境迎来进一步优化，同时，外卖用户大数据分析渗透程度将更深。例如，提升平台和商家的经营效率，增进用户对于平台和商家的满意程度以及扩展配送商品品类等，都要基于大数据分析。

(3) 物流配送

物流的配送效率直接体现在用户从下单到收到商品之间的间隔时间上，高效的物流配送也是建立在大数据分析基础之上。通过大数据分析可以对物流资源配置进行优化，合理规划物流路线，从而降低物流成本，提升物流配送效率。物流网点的选址、交通网络规划、辐射区域规划，都可以通过大数据分析进行辅助决策。此外，对车队的能耗数据、路线跟踪、调配信息等数据进行整合并分析，进行数字化管理，可以有效控制车队的运营成本。

(4) 交通出行

大数据分析技术在交通出行方面的应用也很广泛。例如，利用大数据分析技术可以实时监控车辆通行密度，合理规划行驶路线；实现即时的信号灯调度，提高已有线路运行能力。此外，近几年来发展迅猛的打车平台和共享单车也是利用大数据分析技术快速匹配司乘信息，从而提高用户乘车便利性，降低能源损耗，提高出行效率。

(5) 游戏产业

游戏厂商可以基于用户数据根据用户的偏好行为进行分析，可以主动推荐符合其偏好的游戏产品，减少用户搜索感兴趣游戏的时间。此外，对用户在游戏平台内产生的大量行为数据进行分析挖掘，可以迅速定位产品存在的问题并进行优化改进，提高用户忠诚度，降低用户流失率。市场推广渠道的数据分析可以帮助渠道进行优化，从而降低获取客户的成本并实现优质客户的新增导入。

2. 基于行业的应用场景

(1) 天气预报

基于历史海量数据的预测分析结合气象知识，天气预报的准确性和实效性将会大大提高，预报的及时性将会大大提升。此外，对于重大自然灾害，例如台风、龙卷风等，大数据分析技术可以更加精确地判断其运动轨迹和危害的等级，有利于帮助人们提高应对自然灾害的能力，减少损失。天气预报准确度的提升和预测周期的延长将会有利于农业生产的安排。

(2) 农牧业

借助于大数据技术收集农牧产品的产地、产量、品种、流向、销售等各种信息，在大量数据分析基础上得到农牧产品的指导信息、流通信息等。通过不同的应用场景，可以使得农牧业从业者获取农牧产品的市场行情、相关技术等信息，从而做好预判。此外，企业基于大

数据分析可以获得农牧产品的流通数据、市场消费需求、市场布局情况等专业的分析报告。政府可以通过大数据的整合分析，为农牧业生产提供合理建议，引导市场供需平衡，避免产能过剩，造成不必要的资源和社会财富浪费。

(3) 医疗卫生

根据医院病人的就诊信息，通过大数据分析得出涉及食品安全的信息，及时进行监督检查，降低已有不安全食品的危害；基于用户在互联网的搜索信息，掌握流行疾病在某些区域和季节的爆发趋势，及时进行干预，降低其危害；基于覆盖区域的居民健康档案和电子病历数据库，快速检测传染病，进行全面的疫情监测，并通过集成疾病监测和响应程序，快速进行响应。

(4) 教育行业

大数据分析技术可以被政府教育部门运用到教学改革实践中。通过对学生成绩、行为表现、心里活动等数据的分析，可以让教育工作者理解学生在个性化层面是怎样学习的，从而制定相关策略来提高学生的成绩。此外，基于大数据分析可以将学习兴趣相同的学生进行分组，从而提高共同学习效率，还可以为每位学生创建适合自己的学习环境和个性化的学习方案和学习路径。

(5) 金融行业

银行基于客户资料的大数据分析，对申请贷款的客户进行信用评分，从而确定是否给客户发放贷款以及发放贷款的额度。此外，银行可以对客户数据进行细分研究，通过聚类分析发现不同类型客户的特征，挖掘不同客户的特点，从而为客户提供优质的服务。

利用大数据挖掘技术对投资的理财产品进行组合策略分析，从而降低投资风险，提高资金使用效率。此外，对已有的投资产品的组合模型进行优化分析，为投资者提供更为精准的数据分析。

保险业可以通过大数据技术对客户数据进行挖掘，研究欺诈客户的行为特征，进行实时监控与预警，降低企业风险。

(6) 零售行业

客户群体的细分以及精细化运营同样适用于零售行业，根据客户的消费喜好和趋势，进行商品的精准营销，降低营销成本。

利用大数据分析技术缩短产品生产时间，根据顾客反馈意见，快速进行决策并迅速修正产品缺陷，给用户更好的体验，从而提高产品的服务质量。

建立用户预测趋势的模型，对消费者购买方式和地点进行预测，从而能够调整库存量，提高产品周转效率，满足消费者的需求。

(7) 制造业

对制造业企业的销售业绩、利润率、成本等数据的分析，有助于了解企业销售状况，从而制定相应的销售策略，扩大生产利润。

对采购及库存数据的分析，有助于全面掌握企业采购及库存状态，为优化采购流程、降低库存积压提供决策依据。

针对产品故障数据进行预警分析，了解产品的故障状态，对于发生概率较高的故障问

题、排名靠前的故障产品型号，可以改进生产工艺流程，降低产品故障率。

1.1.4 数据分析的发展趋势

1. 技术发展趋势

大数据技术的发展使得数据采集、存储、安全等技术变得越来越成熟。人们对于采集并存储的数据的价值越来越重视，从而带动了数据分析和数据挖掘技术的发展。利用大数据分析技术从海量数据中提取的信息具有极高的价值，例如，支持企业高层进行业务决策、识别新的销售和市场机会、提升组织的社交媒体营销能力、提高用户忠诚度以及复购率、降低用户流失率、提前预测风险并进行防范等。未来大数据分析技术的发展方向大致可以分为以下两点：一是对海量的结构化和半结构化数据进行深度分析，挖掘数据背后隐藏的知识；二是对非结构化数据进行深度挖掘，将文本、图形、声音、影视、超媒体等类型中蕴藏的丰富信息转化为有用的知识。

2. 产业发展趋势

大数据技术的发展带动了包括数据软件和硬件相结合的高科技服务行业，提供专业大数据解决方案的咨询服务业，从事数据采集、处理、加工及分析为一体的数据服务产业的产生和发展。此外，2016年以来国家政策持续推动大数据产业发展，“十三五规划”中明确提出实施大数据战略，把大数据作为基础性战略资源，全面实施促进大数据发展行动，加快推动数据资源共享开放和开发应用，助力产业转型升级和社会治理创新。国家众多部门相继出台政策推动大数据产业的发展，随着政策的逐步落地，大数据产业的发展速度也将越来越快。

3. 人才发展趋势

大数据技术的发展带动了企业对于大数据分析人才需求的快速增长，由于当前国内大数据人才培养的滞后，导致大数据分析人才的缺口很大，因此未来一段时间内大数据分析人才依然炙手可热。目前国内主流招聘网站上发布的数据分析相关岗位的数量呈现持续快速增长，可以看出企业对于大数据分析人才的需求量很大。从数据分析师的职位分布区域来看，主要分布在北京、上海、广州、深圳、杭州等经济发达的大城市。从数据分析师的职位分布行业来看，主要集中在互联网、金融、游戏、生物医药、房地产、制造业等行业。互联网、金融行业的数据分析师职位数占比很高，因为这些行业在日常运营工作中积累了大量业务数据，而且数据依然在快速增长。

1.2 数据分析的职业发展及分类

数据分析的职业发展分为技术路线和管理路线。技术路线分为数据分析助理、数据分析专员、初级数据分析师、中级数据分析师、高级数据分析师、资深数据分析师、数据科学家等；管理路线分为数据分析主管、数据分析经理、数据分析总监、首席数据官等。

从工作内容上划分，数据分析的职位主要分为两大类：业务数据分析和数据挖掘算法研发，如图 1-8 所示。下面分别对这两类职位的基本要求和职能进行详细介绍。