商务数据分析

ISBN 978-7-5647-7585-8
9 787564 775858
定价：49.80元

需要全本请访问 www.ertongbook.com

精品课程配套教材
21 世纪应用型人才“十三五”规划教材

商务数据分析

主　编　王珊珊　蔡映珍　欧阳红巍

副主编　朱建平　张亚伟　朱志辉
曾淑云　沙昉奕　黄　彦
李鹏举　刘小娇　杨江浩
蒋小龙　刘超君　林碧坤

电子科技大学出版社
University of Electronic Science and Technology of China Press
• 成都 •

前　言

国家对电子商务产业的发展十分重视这是有目共睹的。2015 年 3 月十二届全国人大三次会议上，李克强总理在政府工作报告中首次提出“互联网+”行动计划。电子商务以前所未有的速度迅猛发展着，已经成为世界各国增强经济竞争实力，赢得全球资源配置优势的有效手段。电子商务作为现代流通方式，已广泛渗透到生产、流通、消费等各个领域，对改变传统经营管理模式、有效降低交易成本、提高经济运行效率、促进区域经济发展起着越来越重要的作用，已被公认为是 21 世纪世界格局的新型经济模式和催化剂。

为了让电子商务在更多的企业得到应用，电子商务运营必须产生良好的运营效果。宏观层面看，数据能在量化层面提示电商生态的进化与成长；微观层面看，企业的兴衰总是率先在运用数据方式上显现出端倪。因此，未来国家与企业均会重视商务数据分析这一新兴行业。这就要求在大力推动高等院校培养商务数据分析与应用人才的同时，撰写出更多的高质量教材，从而满足对开展电子商务数据分析人才培养的需求。

为提高电子商务大类专业学生的数据分析能力，以就业为导向，注重学生岗位工作能力的培养。本书在内容上特意增加了电子商务企业数据化运营等相关内容，旨在对接电商企业需求，培养学生的专业核心技能。同时，为了便于学生自主学习，将内容设置成为若干任务。每个任务均设置了“任务情境”，介绍知识背景，激发学生学习兴趣；每个任务配备了“任务导入”“相关知识”“任务实战”“同步习题”，方便学生总揽内容，巩固学习成果。并通过“知识链接”，紧密对接电子商务、商务数据分析行业最新业态变化，拓宽学生视野，启发学生积极开展思考。本教材主要面向高等院校电子商务、商务数据分析与应用、营销以及其他相关专业学生，对从事电子商务和商务数据分析等实际工作与研究的人员也有较高的参考价值。

全书由王珊珊、蔡映珍、欧阳红巍担任主编；朱建平、张亚伟、朱志辉、曾淑云、沙昉奕、黄彦、李鹏举、刘小娇、杨江浩、蒋小龙、刘超君、林碧坤担任副主编。具体编写分工如下：王珊珊负责编写任务一、任务二；蔡映珍负责编写任务三、任务四；欧阳红巍负责编写任务五；任务六、任务七由朱建平、张亚伟、朱志辉、曾淑云、沙昉奕、黄彦共同编写；任务八、任务九由李鹏举、刘小娇、杨江浩、蒋小龙、刘超君、林碧坤共同编写。王珊珊负责组织设计编写大纲以及全书的统稿和修改，并完成最后的定稿。

本书在撰写和出版过程中，引用和参考了国内外大量著作、教材和资料，也得到了有关专家和学者的指导，在此一并表示感谢。同时，部分参考资料由于无法确定来源和作者，因此没有在参考文献中列出，为此表示深深的歉意。由于编者水平有限，书中难免有不足和错漏之处，敬请广大读者和同行批评指正。

作　者

2019 年 10 月

目　　录

任务一　认知电子商务数据分析——数据意味着什么 …… 1

一、任务导入 …… 1
二、相关知识 …… 3
三、任务实战 …… 30
四、拓展实训 …… 35
同步习题 …… 36

任务二　运营数据分析——为店铺做好诊断 …… 37

一、任务导入 …… 37
二、相关知识 …… 38
三、任务实战 …… 50
四、拓展实训 …… 55
同步习题 …… 56

任务三　流量数据分析——实现精准引流 …… 57

一、任务导入 …… 57
二、相关知识 …… 61
三、任务实战 …… 87
四、拓展实训 …… 91
同步习题 …… 92

任务四　转化数据分析——店铺诊断与优化 …… 93

一、任务导入 …… 93
二、相关知识 …… 94
三、任务实战 …… 111
四、拓展实训 …… 124
同步习题 …… 125

任务五　客单价数据分析——帮助提升销售额 …… 127

一、任务导入 …… 127
二、相关知识 …… 128
三、任务实战 …… 135

四、拓展实训……137
同步习题……138

任务六　客户数据分析——客户关系管理的基础……139

一、任务导入……139
二、相关知识……141
三、任务实战……167
四、拓展实训……178
同步习题……179

任务七　商品数据分析——撰写产品分析报告……181

一、任务导入……181
二、相关知识……183
三、任务实战……200
四、拓展实训……207
同步习题……207

任务八　市场行情数据分析——撰写市场调研报告……209

一、任务导入……209
二、相关知识……211
三、任务实战……222
四、拓展实训……234
同步习题……235

任务九　竞争对手数据分析——制定企业竞争战略……236

一、任务导入……236
二、相关知识……238
三、任务实战……252
四、拓展实训……259
同步习题……260

参考文献……262

需要全本请访问 www.ertongbook.com

任务一　认知电子商务数据分析
——数据意味着什么

知识目标

1. 理解电子商务数据分析的相关概念；
2. 熟悉电子商务数据分析的流程；
3. 掌握电子商务数据分析的方法；
4. 掌握电子商务数据分析工具；
5. 理解电子商务数据分析模型。

技能目标

1. 具备运用思维导图绘制数据分析平台功能架构的能力；
2. 具备运用 AARRR 模型分析产品的能力；
3. 具备讲解数据分析案例的能力。

基本素养

1. 具有数据敏感性；
2. 善于用数据思考和分析问题；
3. 具备收集、整理和清洗数据的能力；
4. 具有较好的逻辑分析能力。

一、任务导入

时趣大数据助力李宁精准跨界明星营销

李宁与韩国少女时代郑秀妍（Jessica）的跨界合作，让“Li-Ning×Jessica”系列一经推出就备受追捧。在李宁首度宣布与 Jessica 合作当日，李宁官微创下自身官微互动记录，型自西卡也成为热门话题榜冠军。参与互动的 30%以上的网友都明确表示具有购买欲望，之后“Li-Ning×Jessica”跨界合作产品在李宁官方商城正式开始预售，瞬间就吸引了数以万计的客户蜂拥而至。

利用明星效应进行营销我们都不陌生，但所用明星对产品目标消费者影响力的大小，企业主常常难以把控。所有的成功都不是一次单纯的偶然，李宁与 Jessica 跨界合推新产品系列的成功是大数据下的产物，是基于社交媒体的数据收集、处理的一次精准营销。

那么时趣是怎样通过大数据分析帮助李宁选定合作明星 Jessica，又是如何利用数据进行后续精准营销的呢？

首先，基于大数据的人群洞察——找到品牌与消费者的最优连接者。

作为中国最大的运动品牌之一，李宁需要清晰地把控自己新品系列的调性和受众。李宁把90后、女性这些未来消费主力军，作为他们主要的产品战略方向。基于这个战略，李宁联手合作伙伴大数据专家——时趣，在社交媒体上进行了全方位的人群洞察，以帮助李宁在年轻人群、女性群体身上找到自己新品系列的基础调性：小清新、运动、时尚、阳光。如何把新品信息精准抵达目标消费者，还需要一个“连接”，而李宁通过时趣基于社交媒体的卓越的大数据分析能力，找到了这个具有与李宁新产品系列属相相似的“连接”——韩国顶级女团少女时代成员郑秀妍。

其次，匹配明星信息——确保信息精准抵达。

时趣数据中心通过采集、清洗、存储、计算并整合新浪微博海量微博内容数据及相关客户数据、关系数据等，积累了庞大数量的活跃客户数据、关系数据、微博数据、标签数据，并保持每日新增客户数据及微博数据。利用足够多的客户数据，才能分析出客户的喜好与购买习惯，及时并且全面地了解客户的需求与想法，做到“比客户更了解客户自己”。因此，在纵向上，时趣基于数据对当红艺人社交网络影响力进行了详细的分析与总体评价；在横向上，时趣从性别、年龄、地域、兴趣标签、语义情感等几个维度上把Jessica粉丝的集中倾向属性和李宁新产品系列的调性进行了综合匹配，确保其推送信息可以精准抵达目标消费者，并且这个抵达过程速度极快，范围很广，单位抵达成本小得惊人。Li-Ning与Jessica的合作效果如图1-1所示。

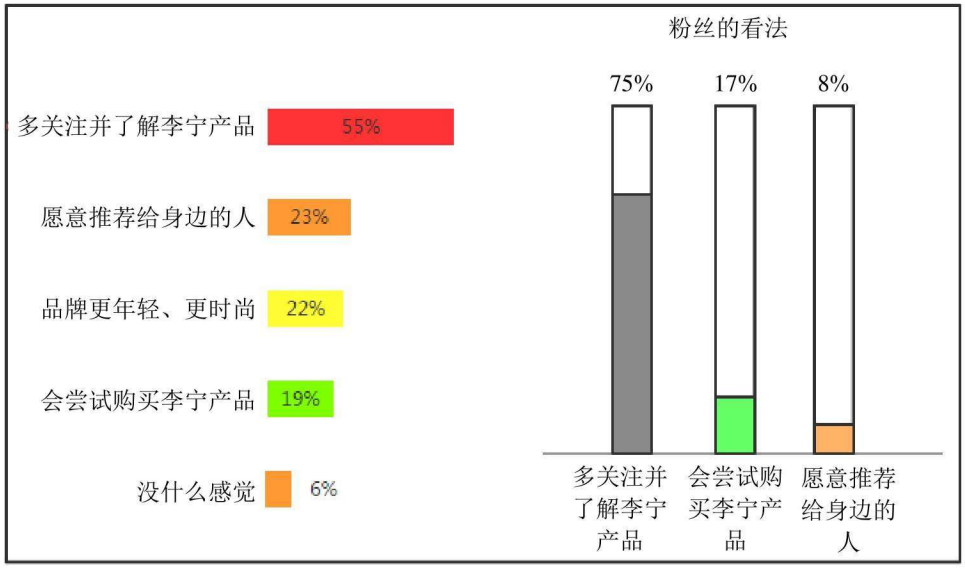


图1-1　Li-Ning与Jessica的合作效果

最后，建立预测响应机制——优化后续营销活动设计。

当然，要做到“精准营销”就必须准确地预计客户需求，因此，时趣不仅对大量历史数据进行了挖掘与分析，还建立了相应的预测响应机制，根据客户在社交媒体上的活动建立数据收集模型，通过模型完成数据的加工和分析，为品牌下一步的产品策划与营销提供更加有意义的数据参考。“Li-Ning×Jessica”相关话题中，“Jessica参与设计”“Jessica行程”等相关内容被多次提及，根据这一数据信息，李宁建立了李宁首尔工作室，随继推出“型自首尔”系列，

邀请 Jessica 亲自参与设计，并在其官方商城为“Jessica&Krystal 近距离接触”的活动造势，使新产品售卖热度继续升温。

总的来说，时趣已经具备通过大数据战略打破行业边界的能力，通过对庞大、复杂的客户数据进行挖掘、追踪、分析，对不同客户群体进行聚合，获得更为完整的客户或客户群体的模型，从而打造个性化、精准化、智能化的产品营销解决方案，最终以个性化营销和主动营销打破传统无差异的、被动的产品服务营销方式。

思考：

1．大数据分析在李宁的“Li-Ning-Jessica”系列产品推广过程中起到哪些作用？

2．如何评价李宁与明星这次合作的营销效果？请列出具体指标。

二、相关知识

电子商务数据分析指对电子商务经营过程产生的数据进行分析，在研究大量的数据的过程中寻找模式、相关性和其他有用的信息，从而帮助商家做出决策。

（一）电子商务数据分析的相关概念

1．数据分析

数据分析是指收集、处理数据并获取信息的过程。具体来说，数据分析是指在业务逻辑的基础上，运用简单有效的分析方法和合理的分析工具对获取的数据进行处理的一个过程。

（1）数据分析的目的。数据分析的目的是把隐藏在一大批看来杂乱无章的数据中的信息集中、萃取和提炼出来，以找出所研究对象的内在规律。在实际生活中，数据分析可帮助人们做出判断，以便采取适当行动。

（2）数据分析的价值。数据分析的价值包含三个方面，一是帮助领导做出决策；二是预防风险；三是把握市场动向，如图 1-2 所示。通过数据分析，可以帮助企业发现做得好的方向、需要改进的地方，以及指出企业出现的问题。

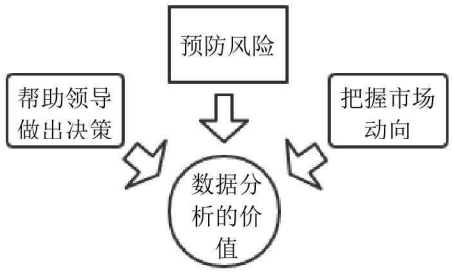


图 1-2　数据分析的价值

（3）数据分析的作用。数据分析在企业日常经营分析中，具有以下三方面的作用。

①现状分析，提供企业现阶段整体运营，以及企业各项业务的构成情况，包括各项业务的发展以及变动情况。

②原因分析，发现企业存在问题的原因，并依据原因制订相应的解决方案。

③预测分析，对企业未来的发展趋势做预测，便于企业制订运营计划。

（4）数据分析的应用。数据分析有极广泛的应用范围，在产品的整个生命周期内，从产品的市场调研到售后服务以及最终处置都需要适当运用数据分析。例如，企业会通过市场调查分析所得数据，以判定市场动向，从而制订合适的生产及销售计划。同样，在淘宝店铺运营过程中，数据分析也起着积极的作用。

（5）数据分析的分类。一般把数据分析分为三类：EDA（Exploratory Data Analysis，探索性数据分析）、CDA（Confirmatory Data Analysis，验证性数据分析）和定性数据分析。

EDA 是指对已有的数据在尽量少的先验假定下进行探索，侧重于在数据之中发现新的特征。EDA 讲究的是从客观数据出发，探索其内在的数据规律，让数据自己说话。

CDA 是指在进行分析之前一般都有预先设定的模型，侧重于已有假设的证实或证伪。定性数据分析是指对词语、照片、观察结果之类的非数值型数据进行分析。如图 1-3 所示。

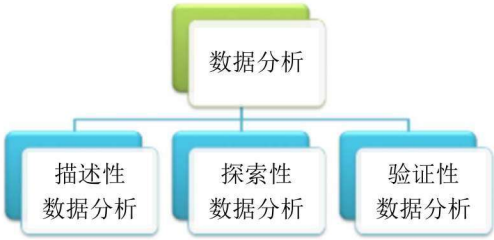


图 1-3　数据分析的分类

2．大数据

大数据本身是一个比较抽象的概念，单从字面来看，它表示数据规模的庞大。目前大数据尚未有一个公认的定义，已有的定义基本是从大数据的特征出发，对其特征的阐述和归纳。在这些定义中，比较有代表性的是 3V 定义，即认为大数据需满足 3 个特点：规模性（Volume）、多样性（Variety）和高速性（Velocity）。维基百科对大数据的定义则简单明了：大数据是指利用常用软件工具捕获、管理和处理数据所耗时间超过可容忍时间的数据集。

知识链接

数据分析从小数据开始

大数据最引人注目的是“大”，但是有时从小一些的数据集开始也是可以的，只要你能够分辨出你所寻找的是什么。如果你的研究对象为开业小于一年的零售店，那么你不需要一个大的 IT 系统来帮助你获取、贮存以及分析所有信息。如果你想要获得更愿意购买你的商品的顾客的统计资料，有很多免费的资源（如 Google Adwords 以及其相关的分析工具）可以帮助你达到目的。一旦你明白数据分析如何使你受益以及使你明白去获得什么，你将会非常确定地想要更大规模以及更可信的数据和方法。

3．云计算

Google 作为大数据应用最为广泛的互联网公司之一，在 2006 年率先提出云计算的概念。云计算是一种大规模的分布式模型，通过网络将抽象的、可伸缩的、便于管理的数据能源、服

务、存储方式等传递给终端客户。根据维基百科的说法，狭义云计算是指 IT 基础设施的交付和使用模式，指通过网络以按照需求量的方式和易扩展的方式获得所需资源；广义云计算指服务的交付和使用模式，指通过网络以按照需求量和易扩展的方式获得所需服务。目前云计算可以认为包含三个层次的内容：服务（IAAS）、平台即服务（PAAS）和软件即服务（SAAS）。国内的阿里云与云谷公司的 XenSystem，以及在国外已经非常成熟的 Intel 和 IBM 都是云计算的忠实开发者和使用者。

云计算是大数据的基础平台与支撑技术。如果将各种大数据的应用比作一辆辆“汽车”，支撑起这些“汽车”运行的“高速公路”就是云计算。正是因为云计算技术对数据存储、管理与分析等方面的支撑，才使得大数据有用武之地。

4．数据挖掘

大数据分析的理论核心就是数据挖掘。数据挖掘的各种算法基于不同的数据类型和格式，能更加科学地呈现出数据本身的特点，从而帮助人们更快速地处理大数据。如果采用一个算法需要花好几年才能得出结论，那大数据的价值也就无从说起了。因此，算法不仅能够满足处理大数据的数据量要求，也能一定程度地满足处理大数据的速度要求。

数据挖掘的重点不在数据本身，而在于如何能够真正地解决数据运营中的实际商业问题。但是，要解决商业问题，就得让数据产生价值，就得做数据挖掘。

5．数据质量

更好的数据意味着更好的决策，数据分析的前提就是要保证数据质量。因此，在数据分析和数据挖掘之前，必须完成数据质量的处理工作，即对数据的集成和处理。

数据质量的处理工作主要包括两方面：数据的集成和数据的清洗，关注的对象主要有原始数据和元数据两方面。

（1）数据的集成。数据的集成主要解决信息孤岛的问题，包括两方面：数据仓库对源数据的集成和元数据系统对不同数据源中的元数据的集成。相应的，数据质量管理也关注两方面：对数据仓库中的真实数据的质量探查和剖析，以及对元数据系统中元数据的数据质量的检查。

元数据的管理目标是整合企业信息资产、支撑数据在使用过程中的透明可视，提升数据报告、数据分析、数据挖掘的可信度，所以元数据的数据质量检查着重在于对元数据信息的唯一性、一致性、准确性的检查。

（2）数据的清洗。数据质量处理主要是采用一些数据清洗规则处理缺失数据、去除重复数据、去除噪声数据、处理异常（但真实）的数据，从而保证数据的完整性、唯一性、一致性、精确性、合法性和及时性。

6．数据预测分析

数据预测分析是大数据技术的核心应用，如电子商务网站通过数据预测分析顾客是否会购买推荐的产品，信贷公司通过数据预测分析借款人是否会违约，执法部门用数据预测分析特定地点发生犯罪的可能性，交通部门利用数据预测分析交通流量等。预测是人类本能的一部分，只有通过大数据预测分析才能获取智能的、有价值的信息。

数据预测分析能帮助企业做出正确而果断的业务决策，让客户更开心，实施成功的数据预测分析有赖于以下几个要素。

（1）数据质量。数据是预测分析的血液。数据通常来自内部，如客户交易数据和生产数据，但通常还需要补充外部数据源，如行业市场数据、社交网络数据和其他统计数据。与流行的技术观点不同，这些外部数据未必一定是大数据。数据中的变量是否有助于有效预测才是关键所在。总之，数据越多，相关度和质量越高，找出原因和结果的可能性越大。

（2）数据分析师。数据分析师必须理解业务需求和业务目标，审视数据，并围绕业务目标建立预测分析规则，如如何增加电子商务的销售额、保持生产线的正常运转、防止库存短缺等。数据分析师需要拥有数学、统计学等多个领域的知识。

知识链接

数据分析师应具备的能力

对于数据分析师来说，熟悉业务逻辑和掌握数据分析的方法与工具同样重要。数据分析师不可能在没有理解业务逻辑的基础上，提出一个基于数据的解决方案。因为每个行业都有不同的趋势、行为和驱动因素。事实上，一个不懂业务逻辑的数据分析师，其分析结果不会产生任何使用价值；而一个只懂业务逻辑而不精通数据分析方法与工具的数据分析人员，充其量只能算数据分析爱好者。现实情况是懂数据分析的人很多，懂业务逻辑的人更多，但既懂业务又懂数据分析，称得上“数据分析师”的人却非常少。

（3）预测分析软件。数据分析师必须借助预测分析软件来评估分析模型和规则，预测分析软件通过整合统计分析和机器学习算法发挥作用。

（4）运营软件。找到了合适的预测规则并将其植入应用，就能以某种方式产生代码，预测规则也能通过业务规则管理系统和复杂事件处理平台进行优化。

7．数据可视化

数据可视化是指将数据分析结果用简单且视觉效果好的方式展示出来，一般运用文字、表格、图标和信息图等方式展示。Word、Excel、PowerPoint、水晶易表等都可以作为数据可视化的展示工具。现代社会已经进入一个速读时代，好的可视化图表可以清楚地表达数据分析的结果，大大节约人们思考的时间。

数据分析的使用者包括大数据分析专家和普通客户，他们对于大数据分析最基本的要求就是数据可视化，因为数据可视化能够直观地呈现大数据的特点，让数据自己说话，让观者直接看到结果，这就是数据可视化的作用。

（二）电子商务数据分析的流程

最初的数据可能杂乱无章且无规律，要通过作图、制表和各种形式的拟合来计算某些特征量，探索规律性的可能形式。这时就需要研究用何种方式去寻找和揭示隐含在数据中的规律性。首先在探索性分析的基础上提出几种模型，再通过进一步的分析从中选择所需的模型，最后使用数理统计方法对所选定模型或估计的可靠程度和精确程度做出推断。数据分析流程如图 1-4 所示，具体步骤如下。

需要全本请访问 www.ertongbook.com



图 1-4　数据分析流程

1．识别数据分析需求

识别数据分析需求可以为收集数据、分析数据提供清晰的目标，是确保数据分析过程有效性的首要条件。在开始数据分析之前，就应该冷静思考在数据分析过程中想要获得什么。例如，是想要更精确地确定网店的客户群，还是想要扩大网店的客户群？或者是为了评估产品改版后的效果是否比之前有所提升？或者是找到产品迭代的方向？还是进行科学的排班以至于不必在闲时浪费人力、在忙时缺少人手？明确通过数据分析要获得什么以及数据分析的目标是什么是至关重要的。

明确了数据分析的目的，接下来需要确定应该收集的数据都有哪些。

2．收集数据

当通过数据分析来揭示变化趋势时，数据量越大越好。对于任何类型的统计分析，样本量越大，所得到的结果越精确。仅仅是追踪公司一周的销售数据的价值是很难看出未来发展趋势的，3 个月的会好一些，6 个月的更佳。即使无法确定寻找的是什么，也要确保收集的数据所包含的信息要尽可能详尽和精确。试着弄清楚获得所需最优数据的途径，然后开始收集。如果没有数据，就不能够进行分析。

收集数据即是如何将数据记录下来的环节。在这个环节中需要着重说明的是两个原则，即全量而非抽样，以及多维而非单维。

（1）全量而非抽样。数据分析的人员能够做到完全全量的对数据进行收集和分析。

（2）多维而非单维。将数据针对客户行为实现 5W1H 的全面细化，将交互过程的什么时间、什么地点、什么人、因为什么原因、做了什么事情全面记录下来，并将每一个板块进行细化，时间可以从起始时间、结束时间、中断时间、周期间隔时间等细分；地点可以从地市、小区、气候等地理特征、渠道等细分；人可以从多渠道注册账号、家庭成员、薪资、个人成长阶段等细分；原因可以从爱好、人生大事、需求层级等细分；事情可以从主题、步骤、质量、效率等细分。通过这些细分维度，增加分析的多样性，并从中挖掘规律。

有目的的收集数据是确保数据分析过程有效的基础，需要对收集数据的内容、渠道、方法进行策划，主要考虑：①将识别的数据分析需求转化为更具体的要求，如评价供方时，需要收集的数据可能包括其过程能力、测量系统不确定性等相关数据；②明确由谁在何时何处，通过何种渠道和方法收集数据；③记录表应便于使用；④采取有效措施，防止数据丢失和虚假数据对系统的干扰。

3．数据的处理与集成

数据的处理与集成主要是完成对于已经采集到的数据进行适当的处理、清洗去噪以及进一步的集成存储。对收集数据进行抽取，从中提取出关系和实体，经过关联和聚合之后采用统一

定义的结构来存储这些数据。在数据抽取时，需要对数据进行清洗和整理，保证数据质量及可信性。常用的数据清洗和整理方法有三种：去重、排序和分组。

（1）去重。去重是指删除重复的数据，以减少对后续数据分析步骤的干扰。去重工作采用 Excel 工具，具体步骤如下。

步骤 1：从网上获取原始空调型号一共有 12 个，如表 1-1 所示。

表 1-1　原始空调型号

序号	原始空调型号
1	Midea/美的 KFR-26GW/WCBD3@
2	AUX/奥克斯 KFR-35GW/NFI19+3
3	Midea/美的 KFR-35GW/WDBD3@
4	Midea/美的 KFR-26GW/WCBD3@
5	Midea/美的 KFR-23GW/DY-PC400 （D3）
6	Midea/美的 KFR-26GW/WCAB3@
7	Gree/格力 KFR-26GW/ （26592）NhAc-3
8	TCL KFRd-23GW/BF33-I
9	Midea/美的 KFR-35GW/WCBD3@
10	TCL 移动水冷气扇小空调
11	Midea/美的 KFR-26GW/WCBD3@
12	AUX/奥克斯 KFR-35GW/BpNFI19+3

去重的第一步是标识重复项，单击“数据/高亮重复项”按钮，结果显示 Midea/美的 KFR-26GW/WCBD@有两条重复项，如图 1-5 所示。

步骤 2：单击“数据/删除重复项”按钮，打开“删除重复项”对话框，在“删除重复项”对话框的列中选择“空调型号”多选项，单击“删除重复项”按钮，删除两条重复项，如图 1-6 所示。

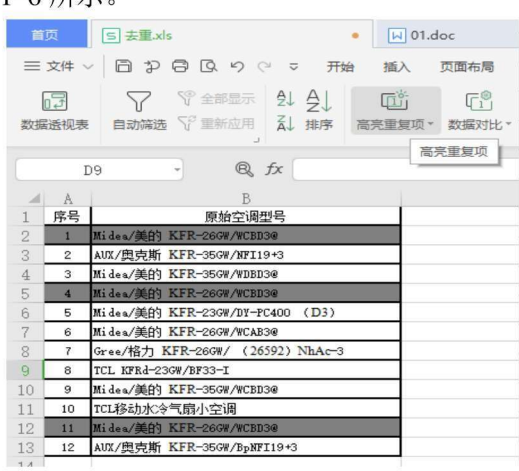


图 1-5　高亮重复项

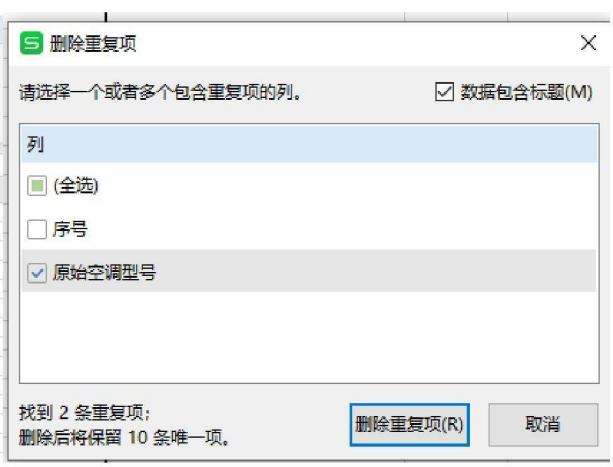
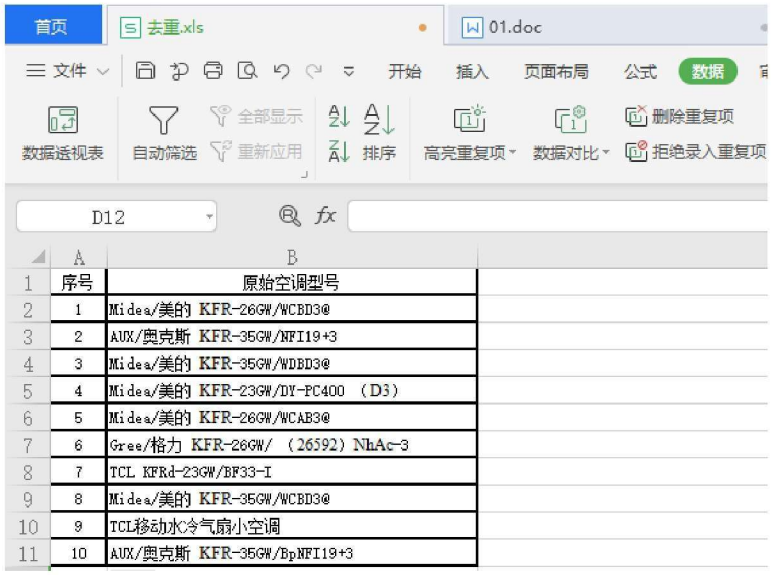


图 1-6　删除重复项

步骤 3：删除后保留 10 条唯一项，将 B2 中“原始空调型号修改序号，如图 1-7 所示。



	A	B
1	序号	原始空调型号
2	1	Midea/美的 KFR-26GW/WCBD3@
3	2	AUX/奥克斯 KFR-35GW/NFI19+3
4	3	Midea/美的 KFR-35GW/WDBD3@
5	4	Midea/美的 KFR-23GW/DY-PC400 (D3)
6	5	Midea/美的 KFR-26GW/WCAB3@
7	6	Gree/格力 KFR-26GW/ (26592) NhAc-3
8	7	TCL KFRd-23GW/BF33-I
9	8	Midea/美的 KFR-35GW/WCBD3@
10	9	TCL移动水冷气扇小空调
11	10	AUX/奥克斯 KFR-35GW/BpNFI19+3

图 1-7　去重后的空调型号

去重可以节省存储空间，通过删除重复数据，可以大大降低需要的存储介质数量，进而降低成本，提升磁盘的写入性能，节省网络带宽。

知识链接

大数据时代数据的特征

大数据是什么？其实很简单，大数据其实就是海量资料巨量资料，这些巨量资料来源于世界各地随时产生的数据。在大数据时代，任何微小的数据都可能产生不可思议的价值。大数据有 4 个特点，分别为：Volume（大量）、Variety（多样）、Velocity（高速）、Value（价值），一般我们称之为 4V。

（1）大量。大数据的特征首先就体现为“大”，从先 MP3 时代，一个小小的 MB 级别的 MP3 就可以满足很多人的需求，然而随着时间的推移，存储单位从过去的 GB 到 TB，乃至现在的 PB、EB 级别。随着信息技术的高速发展，数据开始爆发性增长。社交网络（微博、推特、脸书）、移动网络、各种智能工具，服务工具等，都成为数据的来源。淘宝网近 4 亿的会员每天产生的商品交易数据约 20TB；脸书约 10 亿的用户每天产生的日志数据超过 300TB。迫切需要智能的算法、强大的数据处理平台和新的数据处理技术，来统计、分析、预测和实时处理如此大规模的数据。

（2）多样。广泛的数据来源，决定了大数据形式的多样性。任何形式的数据都可以产生作用，目前应用最广泛的就是推荐系统，如淘宝、网易云音乐、今日头条等，这些平台都会通过对用户的日志数据进行分析，从而进一步推荐用户喜欢的东西。日志数据是结构化明显的数据，还有一些数据结构化不明显，例如图片、音频、视频等，这些数据因果关系弱，就需要人工对其进行标注。

（3）高速。大数据的产生非常迅速，主要通过互联网传输。生活中每个人都离不开互联网，也就是说每个人每天都在向大数据提供大量的资料。并且这些数据是需要及时处理的，因为花费大量资本去存储作用较小的历史数据是非常不划算的，对于一个平台而言，也许保存的数据只有过去几天或者一个月之内，再远的数据就要及时清理，不然代价太大。基于这种情况，大数据对处理速度有非常严格的要求，服务器中大量的资源都用于处理和计算数据，很多平台都需要做到实时分析。数据无时无刻不在产生，谁的速度更快，谁就有优势。

（4）价值。这也是大数据的核心特征。现实世界所产生的数据中，有价值的数据所占比例很小。相比于传统的小数据，大数据最大的价值在于通过从大量不相关的各种类型的数据中，挖掘出对未来趋势与模式预测分析有价值的数据，并通过机器学习方法、人工智能方法或数据挖掘方法深度分析，发现新规律和新知识，并运用于农业、金融、医疗等各个领域，从而最终达到改善社会治理、提高生产效率、推进科学研究的效果。

（2）排序。整理数据时，排序也是重要的方法之一，因为数据经过排序后，会方便商家从中识别哪个数据最大，哪个数据最小，进而发现数据反映出的问题。

（3）分组。商家日常都会收集数据，日积月累，数据量就会变得很大。面对这些毫无规律的数据，商家会感觉不知如何进行数据分析。但如果能对这些数据进行分组整理，分析起来就容易找到头绪。

4．数据分析

数据分析是整个大数据处理流程里最核心的部分，因为在数据分析的过程中，会发现数据的价值所在。数据分析是指将收集到的数据通过加工、整理和分析后，将其转化为信息的过程。

常用的数据分析方法有排列图、因果图、分层法、调查表、散布图、直方图、控制图、关联图、系统图、矩阵图等。

在数据分析的基础上，还要进一步将分析方法应用在业务需求中。基于业务主题的分析可以涉及多个领域，从客户参与网店推广活动的转化率分析，到客户的留存时长分析，再到内部的各环节衔接的及时率和准确度分析等，每一方面都有独特的指标、维度以及分析方法的要求。

5．数据解释

对于广大的数据信息客户来讲，最关心的并非是数据的分析处理过程，而是对大数据分析结果的解释与展示。因此，在一个完善的数据分析流程中，数据结果的解释步骤至关重要。如果数据分析的结果正确但是没有采用适当的解释方法，则所得到的结果很可能让客户难以理解，极端情况下甚至会误导客户。

数据解释的方法很多，比较传统的就是以文本形式输出结果或者直接在计算机上显示结果。这种方法在面对小数据量时是一种很好的选择，但是大数据时代的数据分析结果往往是海量的，同时结果之间的关联关系极其复杂，采用传统的解释方法基本不可行。可以考虑从下面两个方面提升数据解释能力。

（1）引入可视化技术。可视化作为解释大量数据最有效的手段之一率先被科学与工程计算领域采用。通过对分析结果的可视化，可以用更形象的方式向客户展示结果，同时图形化的方式比文字更易理解和接受。

常见的可视化技术有标签云（Tagcloud）、历史流（Historyflow）、空间信息流

（Spatialinformation）等。数据可视化工具中，报表类工具有 JReport、Excel、水晶报表、FineReport 等；BI 分析工具有 BO、BIEE 等；国内的数据可视化工具有 BDP 商业数据平台、大数据魔镜、数据观、FineBI 商业智能软件等。用户可以根据具体的应用需要选择合适的可视化技术和工具。

（2）客户参与分析过程。另一方面，可以让客户能够在一定程度上了解和参与具体的分析过程，从而提升数据解释效果。客户参与分析过程有两种方式：既可以采用人机交互技术，利用交互式的数据分析过程来引导客户逐步地进行分析，使客户在得到结果的同时更好地理解结果的由来；也可以采用数据起源技术，通过该技术帮助追溯整个数据分析的过程，以帮助客户理解数据分析结果。

数据分析完成后一般会要求撰写数据分析报告，它是对整个数据分析过程的总结，是给企业决策者的一个参考报告，可以为决策者提供科学、严谨的决策依据。一份优秀的数据分析报告，需要有一个明确的主题、一个清晰的目录、图文并茂的数据阐述以及条理清晰的内容呈现。最后还需要加上结论和建议，并提供解决问题的方案和想法，以便决策者在决策时作为参考。

（三）电子商务数据分析的方法

从根本目的上来说，数据分析的任务在于通过抽象数据形成对业务有意义的结论。因为单纯的数据是毫无意义的，直接看数据是没有办法发现其中的规律的，只有通过使用分析方法将数据抽象处理后，人们才能看出隐藏在数据背后的规律。

1．数据分析方法分类

选取恰当的数据分析方法是整个数据处理过程中的关键步骤，一般从分析方法复杂度上来讲，可以将数学分析方法分为三个层级，即常规分析方法、统计学分析方法和自建模型。

（1）常规分析方法。常规分析方法不对数据做抽象的处理，主要是直接呈现原始数据，多用于针对固定的指标且具有周期性的分析主题。常规分析方法直接通过原始数据来呈现业务意义，主要的分析方法有两种——趋势分析和占比分析，其对应的分析方法分别为同环比分析及帕累托分析。同环比分析，其核心目的在于呈现本期与往期之间的差异，如销售量增长趋势；而帕累托分析则是呈现单一维度中的各个要素占比的排名，如“各个城市本期的销售量增长趋势的排名”，以及“前百分之八十的增长量都由哪几个城市贡献”这样的结论。

（2）统计学分析方法。统计学分析方法能够基于以往数据的规律来推导未来的趋势，根据使用的原理多分为以下几大类：有目标结论的有指导学习算法、没有目标结论的无指导学习算法以及回归分析法。

（3）自建模型。自建模型在分析方法中是最为高阶也是最具有挖掘价值的，多用于金融领域，甚至业界专门为自建模型的人群起了一个名字叫作“宽客”，这群人就是靠数学模型来分析金融市场的。统计学分析方法所使用的算法是具有局限性的，虽然统计学分析方法能够运用于各种场景中，但是它存在不精准的问题，在有指导和没有指导的学习算法中，得出的结论很多都不精准，而在金融这种锱铢必较的领域中，这种算法显然不能达到需求的精准度，因此数学家在这个领域专门自建模型，通过输入数据，得出投资建议。在统计学分析方法中，回归分析法是最接近数学模型的，但回归分析法用到的公式的复杂程度有限，而数学模型是完全自

由的，能够对指标进行任意的组合，从而确保最终结论的有效性。

2．常用的数据分析方法

（1）描述性统计分析。所谓描述性统计分析，就是在表示数量的中心位置的同时，还能表示数量的变异程度（即离散程度）。描述性统计分析一般包括两种方法：频数分布分析和列联表分析。

描述性统计量包括均值、方差、标准差、最大值、最小值、极差、中位数、分位数、众数、变异系数、中心矩、原点矩、偏度、峰度、协方差和相关系数。

知识链接

描述性统计指标

平均数——N 个数相加除以 N；

中位数——每一组中数据按大小排列，排在中间位置的数据；

众数——一组数据中出现次数最多的数；

峰度系数——一种对称分布曲线峰顶尖峭程度指标，峰度系数＜0，则呈现平阔峰分布，峰度系数＞0，则呈现尖峭峰分布；

偏度系数——数据对称性指标，偏度系数＜0 为负偏态分布，偏度系数＞0 为正偏态分布，偏度系数＞1 为高度偏态分布，0.5＜偏度系数＜1 为中等偏态分布；

方差——各个数据分别与其平均数之差的平方的和的平均数；

标准差——总体各单位标准值与其平均数离差平方的算术平均数的平方根。

（2）回归分析法。回归分析法是研究一个随机变量（Y）对另一个（X）或一组（X_1，X_2，…，X_k）变量的相依关系的统计分析方法。回归分析（Regression Analysis）法是确定两种或两种以上变数间相互依赖的定量关系的一种统计分析方法，其运用十分广泛。回归分析法按照涉及的自变量多少，可分为一元回归分析和多元回归分析；按照自变量和因变量之间的关系类型，可分为线性回归分析和非线性回归分析。

回归分析法简单说就是几个自变量加减乘除后就能得出因变量来。例如，想知道活动覆盖率、产品价格、客户薪资水平、客户活跃度等指标与购买量存在何种关系，就可以运用回归分析法，把这些指标及购买量的数据输入系统，运算后即可分别得出这些指标与购买量存在何种关系的结论，以及通过进一步的运算得出相应的购买量。

回归分析工具是一种非常有用的预测工具，既可以对一元线性或多元线性问题进行预测分析，也可以对某些可以转化为线性问题的非线性问题预测其未来的发展趋势。一般线性回归分析主要有以下 5 个步骤：

①根据预测对象，确定自变量和因变量；

②制作散点图，确定回归模型类型；

③估计参数，建立回归模型；

④检验回归模型；

⑤利用回归模型进行预测。

利用回归分析法进行预测时，常用的是一元线性回归分析，又称简单线性回归。

需要全本请访问 www.ertongbook.com