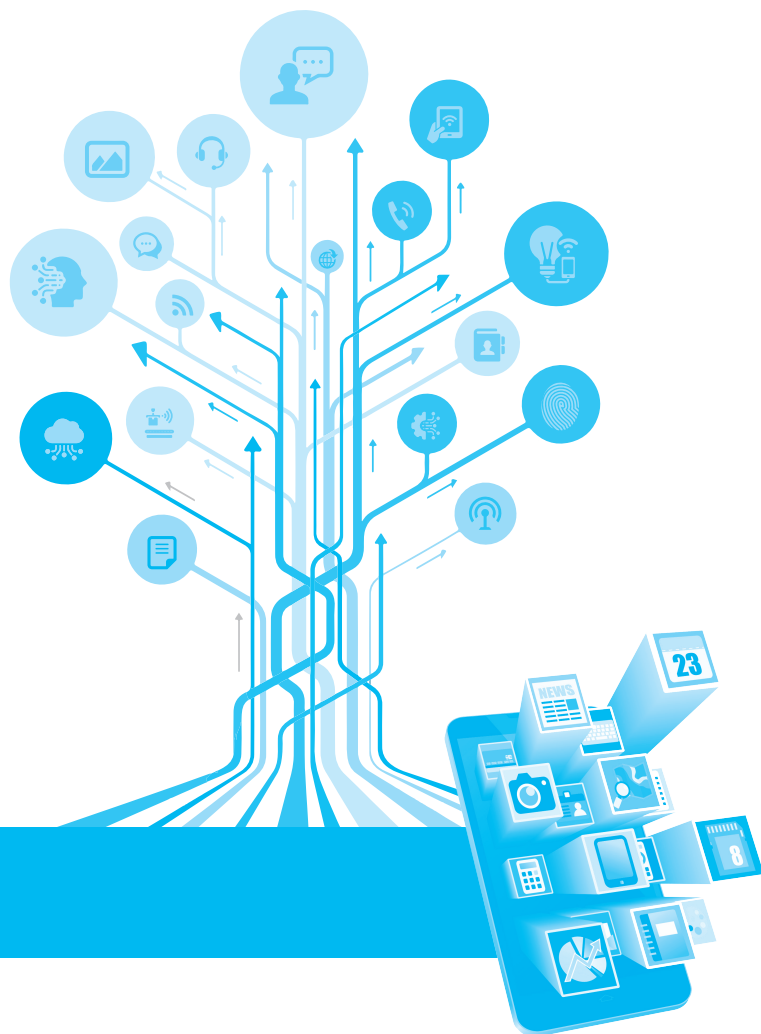


商务数据 分析与处理

主 编 宫 蕾
企业顾问 台晓旭



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

内容提要

本书共分 11 个项目,内容包括商务数据分析概述、商务数据分析平台、初识商务数据可视化、商务数据可视化识图、商务数据可视化图表分析、市场数据采集与分析、竞争对手分析、客户数据调研与分析、商品数据的统计与分析、店铺整体运行数据分析、商务数据分析报告撰写。

本书既可作为各院校电子商务及其相关专业的教材,也可供相关企业技术人员参考。

图书在版编目(CIP)数据

商务数据分析与处理 / 宫蕾主编. — 上海 : 上海交通大学出版社, 2024. 7

ISBN 978-7-313-30394-3

I. ①商… II. ①宫… III. ①商务工作—数据处理
IV. ①F715

中国国家版本馆 CIP 数据核字(2024)第 051296 号

商务数据分析与处理

SHANGWU SHUJU FENXI YU CHULI

主 编: 宫 蕾

出版发行: 上海交通大学出版社

邮政编码: 200030

印 制: 三河市骏杰印刷有限公司

开 本: 787 mm×1 092 mm 1/16

字 数: 334 千字

版 次: 2024 年 7 月第 1 版

书 号: ISBN 978-7-313-30394-3

定 价: 42.00 元

地 址: 上海市番禺路 951 号

电 话: 021-64071208

经 销: 全国新华书店

印 张: 12.75

印 次: 2024 年 7 月第 1 次印刷

电子书号: ISBN 978-7-89424-726-1

版权所有 侵权必究

告读者: 如您发现本书有印装质量问题请与印刷厂质量科联系

联系电话: 0316-3662258



Preface

前言

随着大数据、云计算、人工智能等新一代信息技术的迅猛发展,商业经济进入数据化运营模式。党的二十大报告指出:“加快建设……网络强国、数字中国。”全面贯彻新发展理念,以信息化培育新动能,用新动能推动新发展,建设数字中国是数字时代推进中国式现代化的重要引擎,是构筑国家竞争新优势的有力支撑。数据化运营过程中会产生大量商务数据,这些数据往往蕴藏着巨大的价值。通过对收集的商务数据进行科学合理的分析,商家可以这些数据为依据调整运营策略,不断优化市场调研、客户管理、店铺运营和商品销售等诸多业务流程,从而在激烈的市场竞争中赢得一席之地。

商务数据分析的重要作用已得到许多商家的认可,尤其是在电子商务应用领域,商务数据分析甚至已成为店铺运营管理人员的基本工作内容,部分商家还会专门设置商务数据分析岗位,这使得商务数据分析人才供不应求。鉴于此,编者立足于商务数据分析岗位的实际需求,邀请业界专家编写了本书。

本书推荐学时安排如下表所示。

项目序号	项目标题	学时
项目1	商务数据分析概述	8
项目2	商务数据分析平台	6
项目3	初识商务数据可视化	6
项目4	商务数据可视化识图	4
项目5	商务数据可视化图表分析	4
项目6	市场数据采集与分析	8
项目7	竞争对手分析	4
项目8	客户数据调研与分析	8
项目9	商品数据的统计与分析	8
项目10	店铺整体运行数据分析	4
项目11	商务数据分析报告撰写	4
	总计	64

本书具有如下特色。

1. 注重素质教育,润物无声

党的二十大报告指出:“育人的根本在于立德。”本书有机融入党的二十大精神



神,秉承能力教育与素质教育同向同行的理念,尽可能选取既对应相关知识点又能够体现职业素养并与实际应用紧密相关的案例;将能力教育和素质教育完美融合,力求培养高素质、高水准的专业型人才。

2. 形态新,理念新

本书秉承“实例教学,讲练结合”的教学理念,采用“项目+任务”的体例,对标“1+X”电子商务数据分析职业技能等级标准,将重点知识按岗位需求进行模块化处理,突出职业能力培养目标,从而形成集课、岗、赛、训于一体的新形态教材。

3. 图解详细,实用性强

本书内容全面,既包含商务数据分析的完整知识体系,又包含常用数据分析工具的相关操作,并对实际应用中的操作步骤进行了详细的图解。同时,本书对案例实施过程中难以理解的概念和方法加以解释,使学生一目了然,可有效提升学生的学习效率。

本书由平顶山职业技术学院宫蕾任主编,由平顶山职业技术学院朱炳奇,平顶山学院王安,平顶山职业技术学院李亚涛、罗红娜、李淑芳任副主编,由平顶山旭乐电子商务有限公司台晓旭担任企业顾问。

由于编者水平有限,书中难免存在疏漏和不足之处,敬请广大读者批评指正。

编者



Contents

目录

项目 1	商务数据分析概述	1
	任务 1.1 认识商务数据	2
	任务 1.2 认识商务数据分析	8
	任务 1.3 商务数据分析与处理的流程	11
	任务 1.4 常用的商务数据分析方法	12
	课后练习	26
项目 2	商务数据分析平台	27
	任务 2.1 线上数据分析常用的平台	28
	任务 2.2 线下数据分析常用的平台	29
	任务 2.3 网站数据分析技术	33
	课后练习	40
项目 3	初识商务数据可视化	42
	任务 3.1 商务数据可视化概述	43
	任务 3.2 商务数据可视化图表	49
	课后练习	56
项目 4	商务数据可视化识图	57
	任务 4.1 图表构成元素分解	58
	任务 4.2 常用的识图方法	59
	课后练习	62



项目 5	商务数据可视化图表分析	63
任务 5.1	营收数据分析	64
任务 5.2	销售数据分析	66
	课后练习	68
项目 6	市场数据采集与分析	69
任务 6.1	市场数据采集	70
任务 6.2	市场数据分析概述	72
任务 6.3	市场数据的分析方式	73
任务 6.4	市场数据的分析方法	76
任务 6.5	市场数据分析的应用	77
	课后练习	88
项目 7	竞争对手分析	90
任务 7.1	竞争对手分析概述	91
任务 7.2	竞争对手分析的步骤	92
任务 7.3	竞争对手分析的方法	93
任务 7.4	案例——洗护店铺竞争对手分析	98
	课后练习	102
项目 8	客户数据调研与分析	103
任务 8.1	建立客户指标	104
任务 8.2	客户画像	109
任务 8.3	客户价值分析的方法	123
	课后练习	128
项目 9	商品数据的统计与分析	129
任务 9.1	商品价格分析	130
任务 9.2	商品的生命周期	132



任务 9.3 商品库存分析	142
课后练习	161

项目 10 店铺整体运行数据分析 162

任务 10.1 商品流量分析	163
任务 10.2 商品销售数据分析	170
任务 10.3 网络营销推广	178
课后练习	179

项目 11 商务数据分析报告撰写 180

任务 11.1 商务数据分析报告概述	181
任务 11.2 案例实战——女装店铺商务数据分析报告撰写	184
课后练习	191

参考文献 193

项目 1



商务数据分析概述

知识目标

- (1) 了解数据的概念,明确数据挖掘的阶段。
- (2) 理解商务数据分析的概念、价值。
- (3) 熟悉商务数据的分析流程。
- (4) 熟悉常用的商务数据分析方法。

技能目标

- (1) 能够正确区分常用的数据技术。
- (2) 能够对数据进行常规的清洗。
- (3) 能够对商务数据分析工作进行规划。
- (4) 能够根据实际需要选择合适的商务数据分析方法。

思政园地

自 2021 年开始,带货主播偷逃税事件频频被曝光,一些顶流主播被全网封杀,打破了原有的直播带货格局。

思政育人目标:坚持文化育人理念,严格遵守国家法律法规。



任务 1.1 认识商务数据

在信息社会中,任何商务活动都会产生大量的数据,如果商家能仔细分析和充分利用这些数据,挖掘其背后的规律,就能找出影响商务活动发展的要素及其相互之间的逻辑关系。

因此,对如今的商务领域从业人员而言,分析与应用商务数据是其入行的必备技能之一。通过对商务数据的分析,商家不仅可以弥补店铺经营、客户体验、商品营销手段等方面的不足,还可以深入了解客户的内在需求,规划未来的经营和销售策略。

1.1.1 商业活动与数据

1. 商业活动

商业活动包括企业的购买、销售、交换行为和银行的商业贷款行为。商业活动是指涉及商业主体和商业客体的活动,商业主体和商业客体都要围绕商品用描绘性质的语言进行咨询、探索、理解和磋商。主客体矛盾是围绕商业主体展开的,如何组织言语进行描绘很关键,或正面地宣传企业的形象,或负面地妨碍上述活动的顺利进行,甚至决定着商业目的能否成功落地。



拓展视频:数据分析的含义-rogn3

2. 数据

从极其严格的定义来看,对于客观事物的发生、发展过程的系列客观真实记录称为数据。数据可揭示当时事物的真实活动场景、地位和状态,可还原事物的发展历程。更重要的是,数据可以用于预测行业的未来,为行业价值链上的各种企业在战略、研发、营销、管理等方面提供可靠的咨询和指导。

1.1.2 数据的作用

在远古时代,由于没有文字,人们只能依靠代代相传的故事和诗歌推测历史,或者通过化石研究过去。当有了文字之后,人们就有了更多的证据推测和研究历史,从而对历史有了更多的了解。拥有足够多的历史数据后,科学家、历史学家可以利用这些数据总结事物发生和发展的规律。

企业留存和积累数据也是一样的道理。企业把历史数据积累沉淀,然后不断分析和总结企业的发展轨迹和路径,研究过去的得失,不再重复同样的错误,固化最优秀的做法。企业通过对发展规律的探索,可以指导自己的经营和管理决策,使其更加符合市场的需求。

企业留存和积累的数据越多就越有价值。这些数据一方面可以当作证据,另一方面可以用来研究规律,是企业预测市场的发展趋势、形成商业洞察力的依据。

1.1.3 数据技术

常见的数据技术包括数据采集技术、数据存储技术、数据传输技术以及数据处理和数据挖掘技术。

1. 数据采集技术

数据采集是指获取在事物发生和发展的过程中留存的数据。近年来,数据采集技术得到了飞速的发展,更加重要的是,智能化、自动化的数据采集设备逐步普及。目前,数据采集终端越来越智能化和亲民化,普及率也越来越高。例如,一台普通的智能手机中就有将近

20 个智能数据采集元件。

2. 数据存储技术

数据存储是指将数据以某种格式记录在计算机内部或外部存储介质上。根据数据存储界的“新摩尔定律”，每隔 18 个月，人类数据存储的信息量就会翻一番，而数据存储设备的价格会降低 50%。

3. 数据传输技术

数据传输是数据源和数据宿之间传送数据的过程。互联网彻底改变了人类的生活方式以及企业的运作模式。互联网的快速发展使得带宽不断增加、网费不断降低，数据传输已经变得非常便捷。

4. 数据处理和数据挖掘技术

数据处理和数据挖掘是指在海量的数据中提取筛选出有用的数据。随着数据量的不断增长，数据处理和数据挖掘技术在不断发展，如 Hadoop 非结构化数据处理技术、云存储技术等。

1.1.4 商务数据与电子商务数据

1. 商务数据

商务数据是指在实际经济和商业交易中产生的数据。

2. 电子商务数据

电子商务是指在互联网上以电子交易方式进行商品交易和相关服务的活动，是传统商业活动各环节的电子化、网络化和信息化。电子商务具有如下基本特征。

(1) 商务性。电子商务可以扩展市场，增加客户数量；通过将互联网信息连至数据库，企业能记录客户每次购物的情况以及客户对商品的偏好，这样，企业就可以通过统计这些数据获知客户最想购买的产品是什么。

(2) 整体性。电子商务能够规范事务处理的工作流程，将人工操作和电子信息处理整合成一个不可分割的整体，这样不仅可以提高人力和物力的利用率，还可以提高系统运行的严密性。

(3) 方便性。在电子商务环境中，人们不再受地域的限制，客户能以非常简捷的方式完成过去较为繁杂的商业活动（如通过网上银行能够全天候地存取账户资金、查询信息等），同时使企业对客户的服务质量得以大大提高。

(4) 安全性。在电子商务中，安全性是一个至关重要的核心问题，要求网络能提供一种端到端的安全解决方案（如加密机制、签名机制、安全管理、存取控制、防火墙、防病毒保护等），这与传统的商务活动有着很大的不同。

(5) 协调性。商业活动是一个协调过程，需要客户与企业内部、生产商、批发商、零售商间进行协调。在电子商务环境中，要求银行、配送中心、通信部门、技术服务部门等多个部门通力协作。电子商务的全过程往往是一气呵成的。

电子商务数据是指人们在利用互联网进行商业交易时所产生的数据。例如，京东商城的每一次客户访问、产品交易、客户服务都会伴随着数据的产生，这些数据往往反映了一定的商业信息，如客户的行为特点、商品偏好等。商务数据囊括的范围更为广泛，电子商务数据是商务数据的一种；电子商务数据主要在电子商务领域产生和应用，商务数据则在整个经济活动中产生和应用。

1.1.5 商务数据分析的意义

对于商家而言，在实际经营的过程中，无论是市场调研、客户管理，还是店铺运营、商品



管理,都需要依靠商务数据分析的结果。因此,商务数据分析具有非常重要的意义。商家的运营模式是指商家在生产、销售、售后等环节中对资源、成本、销售渠道等的计划、组织、实施和控制。运营模式能够体现商家的管理水平和运行效率,对商家而言至关重要。因此,商家可对流量数据、销售数据和营销推广数据等运营数据进行分析,并据此改善运营模式,进行精细化、数据化运营。

此外,近年来电子商务日益兴起,各种电商林林总总、遍地开花,让消费者应接不暇,而营销推广成功与否成了电商能否做大做强关键。因此,商家可通过商务数据分析,探索更加有效、投入产出比更高的营销推广方案,进一步优化自身的营销推广策略。商品是商家的命脉和灵魂,只有真正的好商品才能在市场上众多同类型的商品中脱颖而出,获得口碑和销量。因此,商家可对自家商品的价格数据、生命周期数据和库存数据等进行分析,获取市场需求和消费者偏好,并据此设计和销售符合市场预期、能赢得消费者青睐的商品,从而打造高销量的爆款商品。

1.1.6 商务数据的挖掘和利用

大量电子商务系统的出现为商务数据的挖掘和利用提供了平台和便利条件。针对海量的市场交易数据,企业管理人员首先需要甄别数据、分析数据,进而合理利用市场反馈数据的价值,在此层面上,数据的提炼、深入挖掘显得尤为重要。

数据的挖掘是一种技术,旨在通过分类、深层次分析和提炼提取隐藏在海量数据中的商务信息的价值。电子商务是数据挖掘技术实施的主要阵地。在电子商务中,交易平台记录了各种电商交易数据,可以对这些数据加以分析和利用,用来丰富电子商务交易,提升交易的潜在价值,其中蕴含着对未来的憧憬和对现实的考量。因此,电子商务数据的挖掘具有非常重大的理论和现实意义。

商务数据运用的范围非常广,其中互联网上大量的网页数据、商业交易数据、电子商务平台网站(如天猫商城、京东商城、亚马逊等)上的数据都是重要的商务数据的来源。图 1-1 解释了互联网中数据挖掘与数据可视化的关系。

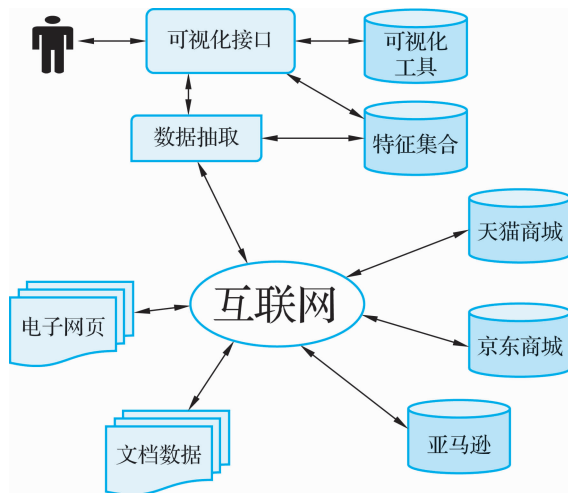


图 1-1 互联网中数据挖掘与数据可视化的关系

1.1.7 数据挖掘的阶段

数据挖掘一般可以分为 5 个阶段,即数据收集准备阶段、数据分门别类阶段、数据分析阶段、内涵读取阶段和精准预测阶段,如图 1-2 所示。

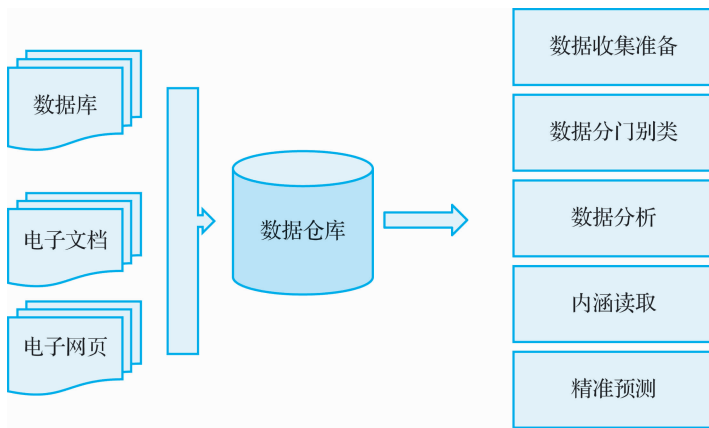


图 1-2 数据挖掘的阶段

1. 数据收集准备阶段

在数据收集准备阶段,最重要的是在相关数据库中筛选出有用的数据,并把相关数据集合在一起,然后对集合起来的数据进行梳理、转换,并建成数据仓库。

2. 数据分门别类阶段

数据分门别类阶段的要求是将上一阶段数据仓库中的数据进行深入的加工和处理,利用数据分析工具依据一定的标准或者关联性将数据进行拆分成若干个不同的类别。

3. 数据分析阶段

在数据分析阶段,根据上一阶段分类的结果,利用聚类、朴素贝叶斯分类算法、神经网络算法等具体的分析方法,分析它们之间的共性和个性,画出它们之间的关系图谱,并从中寻找它们之间的规律。

4. 内涵读取阶段

在内涵读取阶段,在数据分析阶段的基础上,进一步对提取的信息进行加工和提取,运用数据可视化工具将数据所传达的内涵信息形象直观地表达出来,并通过可视化图表进行呈现。

5. 精准预测阶段

在精准预测阶段,对前面各阶段所取得数据加以利用,将相关数据用于具体的推测、预测、建模,从而掌握商务运作的主动权,指导商务活动,使商务管理与运行效益上一个新台阶。

商务数据挖掘的结果是获得两种类型的产品,即满足系统需求的标准化数据库(如特征库)和相关知识。一般来说,知识获取和业务预测是使用专门的工具实现的。其中,可视化工具(软件包)是一种常用的工具。

1.1.8 数据处理

数据处理有广义和狭义之分。广义的数据处理包括所有的数据采集、存储、加工、分析、



挖掘和展示等工作；而狭义的数据处理仅仅包括从存储的数据中提取、筛选出有用的数据，并对有用的数据进行加工的过程。

常用的数据处理是狭义的定义，即对数据进行增加、删除、修改、查询等操作。在目前的大数据背景下，数据处理工作往往是通过技术手段实现的，如利用数据库的处理能力对数据进行增加、删除、修改、查询等。

数据处理主要包括数据清洗、数据转化、数据提取和数据计算。数据清洗是指对收集到的原始数据进行规范化处理，消除数据的不一致性，对数据进行加工，以方便后续数据分析和模型的搭建。数据转化是将数据从一种表示形式转变为另一种表示形式的过程，使数据处于可以直接利用分析软件进行处理的状态。数据提取是从数据库中把所需要的数据筛选出来的过程。数据计算是指对数据进行有目的的计算，从而最大限度地挖掘数据的价值。

在数据处理的过程中，最重要的工作是数据清洗。数据清洗是数据处理的工序之一，目的是提高数据的质量，为数据分析准备有效的数据集。

1. 进行数据清洗的原因

(1)“脏”数据。通常将记录不规范、格式错误、含义不明确的数据称为“脏”数据。

①数据不规范。数据不规范的情况非常常见。例如，对于姓名，同样是“张三”，有的地方记录为“张三”，有的地方记录为“张 三”（为了使两个字的姓名与三个字的姓名具有相同的长度，中间添加了空格）。

上述情况同样会发生在地址字段中。例如，“北京”“北京市”虽然都是指北京，对人们而言很容易识别，但对计算机而言，这两种写法代表着两个不同的值，这时就需要通过建立映射关系将数据记录格式统一。

数据不规范的情况还发生在日期格式中。例如，“2019/10/20”“2019-10-20”“2019年10月20日”“10/20/2019”“Oct. 20, 2019”“October 20, 2019”“2019. 10. 20”都可以表示2019年10月20日。

不同的人可能有不同的喜好和记录数据的方式，这给计算机识别造成了很大的困难。一家公司应该有一个明确的规定，以统一数据的录入格式。

②数据不一致。数据不一致往往是没有遵循单维数据表制作的原则导致的。例如，对同一条信息在不同的数据表中都有记录，当对此条信息进行更改后，没有同时对所有的数据表进行相同的更改，而导致出现数据不一致的情况。

数据的一致性虽然在技术上比较容易实现，但是要在企业经营实践中实现有着巨大的难度。例如，采购部门会录入供应商的信息，财务部门需要向供应商付款，所以也会保留供应商的相关信息。但采购部门和财务部门分属不同的职能部门，财务部门会采集一部分供应商的财务信息（包括银行信息、账号信息、税务信息、工商信息等）。如果这些数据发生变化，财务部门会对这些数据进行更新。采购部门也会对供应商的信息进行采集并登记到相关的信息管理系统中。如果采购部门所使用的信息管理系统能够同财务部门所使用的信息管理系统对接，并且能够把同一条信息关联或者建立索引关系，则该公司的数据的一致性比较容易实现。但是如果两个部门采用了不同的信息管理系统，则很容易产生数据不一致的情况。这种情况在大多数公司中都存在并且很严重。

③数据标准不统一。在录入数据时，需要对一些事物的描述方法建立统一的标准，从而



拓展视频：数据
处理：数据清洗-
eq2kna

使计算机可以有效地处理文本数据。

例如,在描述导致产品出现质量问题的原因时,在大多数情况下是手工录入的,对于同样的原因,不同的人的描述会有所不同。例如,对于描述电压不稳导致的产品质量问题,有的人会录入为“电压不稳”,有的人会录入为“电流不稳定”,还有的人会录入为“供电问题”。如果没有统一的规范,则在统计数据时会产生很多个导致产品质量问题的原因,这给数据解读和分析以及寻找改进措施带来了很大的麻烦。这就需要数据库管理员根据公司的实际情况,将该类原因进行归类,然后设定几个类别,让员工在系统中进行选择,而不是手工录入。一般情况下,出现最多的前 10 个原因能够覆盖 90% 以上的情况,在录入时要先让员工选择,然后留出一个“其他”选项,当员工选择“其他”选项后才能手工录入,这样就可以有效地解决数据的录入标准化问题。

④数据格式不标准。数据格式不标准是指在录入数据时使用了错误的格式。例如,在录入日期时,因为格式不规范,计算机不能自动将其识别为日期格式。对这类问题比较容易处理,可以在信息系统中设定相关的数据校验,如果录入的数据格式不正确,则信息系统会弹出数据录入格式错误的警告。

⑤数据有附加字段。在清洗数据时,往往需要添加新的字段以便数学模型可以直接处理数据。例如,数据库中可能没有直接的字段记录员工的司龄,这就需要在添加“司龄”字段后,通过入职日期进行计算;同理,对员工的年龄通过出生日期来计算。

(2)数据杂质。所谓数据杂质,就是在数据集中出现的与数据记录无关的数据,就如同大米中出现了沙子,需要在处理数据的过程中将这部分数据剔除。例如,企业生产线上的监控设备在企业没有生产时仍然在录像,那么这一时段的录像就可以从整体数据中被剔除。

还有一种数据杂质是在数据采集或者记录过程中产生的。例如,问卷调查中,在正式进行调查之前,编制问卷的人首先要进行测试,以保证正式发布的问卷无差错,而这部分测试数据会被调研系统的后台所记录,这些数据也称为杂质,在处理调研数据集时需要被剔除。

数据杂质有很多种,使用不同的数据采集方式,就会产生不同类型的数据杂质,数据分析人员需要根据实际情况进行处理。

(3)数据噪声。所谓数据噪声,就是对于看似有用的数据,仔细查看后发现该数据并非数据集中应有的数据,或者仔细分析后发现该数据并没有价值,当然也有一部分是无法解释的与数据集中其他数据有差异的数据。与数据杂质不同,数据噪声看似是相关的数据,其实价值不大或者根本没有价值。

例如,一个订单数据集中有一部分数据是内部测试形成的,有一部分数据是竞争对手测试形成的,还有一部分数据可能是消费者测试网站形成的。这些数据可以看作数据噪声,并非真正的交易数据。

2. 数据清洗的方法

数据清洗的方法有很多,主要与所使用的数据处理工具有关。例如,使用 Excel 可以对数据进行查找、替换、填充、分列、映射、透视等。如果数据的规律性很强,数据量很大,可以采用 VBA 编程的方式实现数据清洗。



3. 提高数据清洗速度的方法

一般数据清洗工作占数据分析师工作量的 70% 以上,而且数据质量越差,这一比例越高。

提高数据清洗速度最有效的方法是规范数据采集和数据记录,从源头把控数据质量。如果源头数据的质量不高,则数据清洗工作不仅会洗掉“脏”数据,还会洗掉某些有价值的数
据,导致数据信息的损失。程序化方法是提高数据清洗工作效率的有效手段。当面对的数据集较大时,如果手动一个个检查并清洗,则需要耗费大量的人工和时间。如果能够对不规范、不完整或者不相关的数据进行较好的分析,总结其中可能存在的规律,然后用软件程序自动完成数据清洗工作,就可以大幅度提升数据清洗的效率。

数据清洗工作一个非常重要的原则是永远给自己留下反悔的余地。清洗数据时尽量不要破坏原始数据。不能在原始数据集上直接修改数据,如果修改了某些有价值的数
据,可能很难再找回来;如果发生了错误,结果可能是灾难性的。所以要先备份数据后再清洗。例如,如果想要规范日期格式,可在 Excel 中添加一列进行修改,保留之前的日期数据(如果看着不舒服,可以进行隐藏)。

此外,可以利用第三方工具进行数据清洗。大多数第三方数据清洗工具都是先构建一个映射表,根据数据的特点进行猜测、精准匹配并用后台的“字典”映射数据,然后将规范化的数据输出。但第三方工具在适用性上往往存在一定的缺陷;或者适用于不同的数据集,有的适用于客户数据的清洗,有的适用于产品订单数据的清洗,有的适用于社交媒体网站数据的清洗。在选择第三方数据清洗工具时要进行测评,用一个比较小的数据集进行测试之后再决定是否购买。

任务 1.2 认识商务数据分析

在信息化的时代背景下,商务数据量爆炸式增长。这些数据蕴含着巨大的商业价值,可为经营者带来大市场、大利润和大发展。但是,这些数据规模庞大、种类繁多、更新周期快、价值密度低,因此必须对这些数据进行有效的整理和分析,从而最大限度地挖掘其中的价值。

1.2.1 商务数据分析的概念

商务数据分析是指对经济往来过程中产生的数据进行有目的的收集、处理、分析、可视化等相关操作,以提取对企业有利用可能的信息,以有助于企业做出经营决策,转换业务策略,实现数据驱动的运行和管理。

1.2.2 商务数据分析的价值

通过收集、调研、实际观测、实验、设计等,运用分析工具精准地分析目前经营中的问题,可为企业调整和优化营销推广方案提供科学的参考。商务数据分析的价值如图 1-3 所示。



拓展视频:电子商务
数据分析对企业的作用-fyovs3

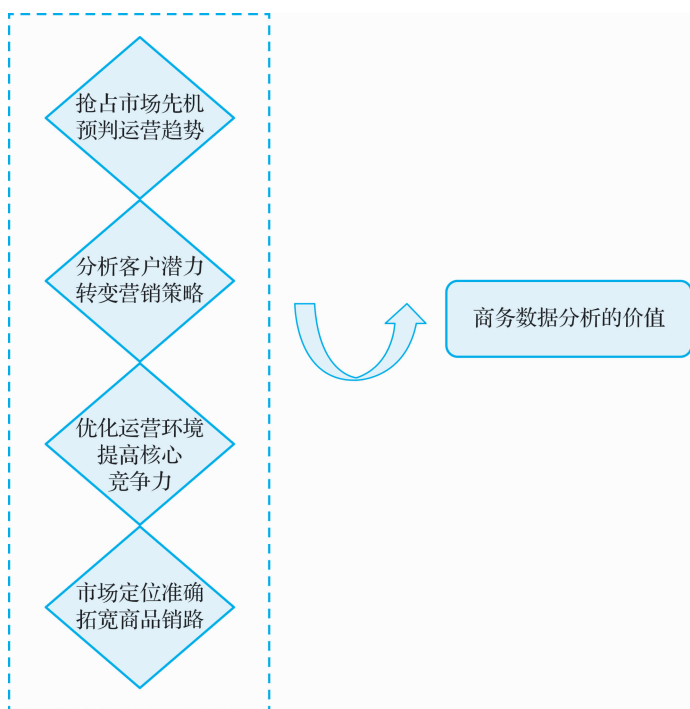


图 1-3 商务数据分析的价值

1. 抢占市场先机, 预判运营趋势

企业想要在市场上扩大占有率,首先要以优质的产品或服务取胜,这样,企业的口碑和信誉度才会提升,企业的品牌影响力才会增强。此外,企业还要抢占先机,积极面对机遇和挑战。企业想要立于不败之地,就必须对自身的实力和竞争对手的核心竞争力进行分析,合理把握市场动向。除此以外,企业还应当立足长远,打开格局,制定一套完善的经营方案。合理进行商务数据分析,企业就能够对于未来市场的发展方向进行预判,正如古语所言“凡事预则立,不预则废”。

2. 分析客户潜力, 转变营销策略

一家企业的生存和发展离不开客户。企业在制定市场推广和营销方案时,必须先对客户进行深入的分析,首先了解客户的购买意向,其次对客户的购买习惯和购买行为数据进行分析,深入掌握客户的偏好和购买能力,最后根据相关数据分析的结果为客户提供个性化服务,充分满足顾客的购买欲望,提升客户的购买体验,尊重和保护客户的个人隐私,维护真诚、良性的企业和客户之间的关系,从而使企业在客户群体中的满意度大大提升,客户对企业的信任程度也随之不断提升,总体推动客户获得率的提升。

3. 优化运营环境, 提高核心竞争力

企业的运营组织形式是指企业在生产、销售、售后过程中对资源、成本、销售渠道等方面的规划、组织、实施和控制。运营组织形式能体现企业的实际经营管理水平和企业经营实效。对于企业而言,掌握这些阶段的相关数据非常重要,企业可以对在运行环节及售后环节获取到的店铺浏览量、销售额度、市场推广力度等相关的有效数据进行汇总分析,提取对企业发展有益的信息,更改经营策略,优化运营方案。



近些年来,随着各类电子商务平台的普及和推广,网红店铺不胜枚举,顾客可在众多电子商务平台进行检索,以便获取对自己有用的相关信息。商务数据分析的应用能够实现投入的高效利用,增强企业的核心竞争力,有利于企业的可持续发展。

4. 市场定位准确,拓宽产品销路

产品是企业的生命线和核心竞争力存在的基础。优质的产品往往能够在市场上海量的产品中保持一枝独秀,赢得相当高的市场占有率,受到客户的青睐。这时企业就要对同类产品的定价、市场远景以及顾客对产品的反馈等相关数据进行收集、整合,在此基础上重新定义设计产品的理念,迎合社会大众对产品的心理预期,真正掌握产品销量的密码。

2021年8月20日,第十三届全国人民代表大会常务委员会第三十次会议表决通过了《中华人民共和国个人信息保护法》。该法明确规定了个人信息处理和跨境提供的规则、个人信息处理者的义务等内容,规定了“任何组织、个人不得非法收集、使用、加工、传输他人个人信息,不得非法买卖、提供或者公开他人个人信息”,凸显出国家高度重视对公民隐私和信息安全的保护。因此,在商务数据分析过程中,对获取到的客户信息应该在遵守国家法律的前提下合理利用,不得违法窃取个人隐私数据,更不能将该数据泄露或者非法买卖。

1.2.3 商务数据分析的特征

1. 数据规模大

由于电子商务的发展,每天都会有大量的数据产生,包括商品销售数据、顾客交易数据、网站流量数据等。

2. 数据种类多

电子商务数据种类繁多,具体包括用户浏览数据、用户购买数据、用户评价数据、用户留言数据等。

3. 数据来源广泛

电子商务数据不仅源于企业自身,也来自第三方平台、社交媒体等渠道,数据来源十分广泛。

1.2.4 商务数据分析的应用

1. 业务决策支持

商务数据可以为企业的战略和业务决策提供支持。通过分析销售数据、市场数据、客户数据等,企业可以了解市场需求、产品销售情况,据此制定合适的定价策略、市场推广策略和产品改进计划。

2. 销售和市场营销

商务数据有助于企业优化销售和市场营销活动。通过分析销售数据和客户数据,企业可以识别潜在客户、优化销售渠道、制定个性化营销策略,提高销售效率和销售额。

3. 客户关系管理

商务数据可以用于客户关系管理,有助于企业了解客户需求、维护客户关系、提供个性化服务。通过分析客户数据和交互数据,企业可以识别重要客户、了解客户偏好、预测客户流失风险,从而改善客户体验和um提高客户满意度。

4. 运营管理和供应链优化

商务数据可以用于企业的运营管理和供应链优化。通过分析运营数据和供应链数据,



拓展视频:电子商务
数据分析的应用-

mrezup

企业可以优化生产计划、库存管理、物流配送等,提高运营效率,降低成本,并确保产品供应的稳定性和及时性。

5. 财务分析和风险管理

商务数据可以用于财务分析和风险管理。通过分析财务数据和市场数据,企业可以评估自身经营状况、盈利能力和财务风险,进行财务预测和预算规划,以及制定风险管理策略。

6. 市场研究和竞争分析

商务数据可以用于市场研究和竞争分析。通过分析市场数据和竞争对手数据,企业可以了解市场趋势、竞争格局和消费者偏好,从而制定市场营销策略、产品定位和品牌推广策略。

任务 1.3 商务数据分析与处理的流程

面对内容繁多庞杂的商务数据,应该按照不同的属性对其进行分类。掌握好商务数据分析的整个流程,才能充分地发挥出商务数据应有的价值。

如图 1-4 所示,一般来说,商务数据分析的流程包括确定分析目标、收集数据、分析数据、直观呈现数据、生成分析报告。



图 1-4 商务数据分析与处理的流程

1.3.1 确定分析目标

针对不同的分析项目,需要明确分析目标。对于不同的分析目标,在具体分析时要对数据进行确定性收集,严格划定收集范围;与此相呼应的数据分类、提炼和最后可视化的方式存在差异。在实践中,开始分析一个具体的商务数据,最先要做的就是确定分析对象,对数据的分析具有目的性,通过分析工具得出的相关数据才是有用的数据,能够为企业的经营带来有益的影响。

在实际操作中经常用到的分析目标有减少客户流失、优化活动效果、提高商品曝光度等。总体来说,确定分析目标就是明确数据对象、要解决的具体业务问题、委托方想要获悉的具体建议等。

1.3.2 收集数据

数据收集是依据既定的分析目标收集有用数据的过程,是商务数据分析存在的基础。



拓展视频:电子商务
数据分析的流程-
oyq1x



一般来讲,目标数据通常分为一手数据和二手数据。一手数据一般指的是运用实地调查方法获取的,或者是通过实验设计得来的直接数据。二手数据是指在已有相关资料的前提下,综合运用文献查阅或者资料整合、提炼加工方法之后得到的相关数据。

1.3.3 分析数据

数据分析是指合理使用数据分析工具和方法,提炼出数据中的可用信息,发现事物的规律,形成相关的定义,以便于影响整个企业的运作流程。数据分析是商务数据分析过程中最重要的部分。它解释了相关数据之间的联系、数据的存在方式,有利于企业确定经营战略的整体定位,从宏观上把握企业的发展,也为企业的后续经营提供可以借鉴的思路。

1.3.4 直观呈现数据

数据呈现是指综合运用图表、表格、动画、交互界面等表达手段,将数据分析的结果以直观、可视化的方式呈现出来。常用的数据呈现图表有折线图、条形图、饼图、散点图、气泡图、面积图、雷达图等。实践中,创建图表包含的步骤有明确图表可视化表现的主题、确定图表可视化要使用的图表类型、使用数据库创建图表的基础数据、生成图表、检验图表等。图表的检验指标有两个:一是如实反映要表达的数据,二是准确地表达数据所传达的内涵。

1.3.5 生成分析报告

商务数据分析报告是展示数据分析结果和制定相关政策的基础。在生成商务数据分析报告时,应当充分展示数据分析的目标、设计方案、流程、结果、参考措施和意见,发挥其应有的参考价值。一份高质量的商务数据分析报告应符合以下标准。

- (1)层次分明,逻辑严密,步骤详略得当。所有读者都能够读取到其中所要表达的内容。
- (2)图文并茂,语言生动,立意清楚。读者能迅速提取有用的信息,并能够将其应用到实践中。
- (3)遵循科学性原则,反复验证数据的真实性;进行个性化、针对性设计,使其结果被普遍接受。

任务 1.4 常用的商务数据分析方法

在实践中,经常选用的商务数据分析方法包括分类分析法、关联分析法、比较分析法、漏斗分析法、矩阵分析法、结构分析法、因果关系分析法等。

1.4.1 分类分析法

分类分析法的适用范围非常广泛,是经常应用的商务数据分析方法。分类分析法是依据不同的标准将数据集中的数据进行分类,每组数据都呈现出自身的特点,数据之间既有区别又有共同之处,通过内部不同组别的数据对比,揭开不同数据背后的深层次规律。

分类分析法可以自由搭配组合不同的数据特点,企业管理者可依据一定的指标进行区分。例如,采用灵活的特征或指标,操作者可以根据需要将同一数据集进行分组。例如,某

电子商务商店收集了最近的销售数据,制定了相关方案,从客户需求着手,对不同的数据进行分类、排列组合,进行深入的提取、加工。因此,可依照年龄、性别等标准对客户的购买行为、消费行为习惯、个人偏好以及多次购买的时间和金额等相关数据进行分组分析。

分类分析法最主要的就是比较两者,甚至多者之间的区别,所以分类分析法的结果就呈现出比较特质。

1. 决策树分类算法

决策树分类算法典型的有 ID3、C4.5 等算法。ID3 算法是利用信息论中的信息增益寻找数据库中具有最大信息量的属性字段,建立决策树的一个节点,并根据字段的不同取值建立树的分枝,在每个分枝子集中重复建立树的下层节点和分枝,最终建成决策树。C4.5 算法是 ID3 算法的后继版本。

2. 贝叶斯分类算法

贝叶斯分类算法是在贝叶斯定理的基础上发展起来的,包括朴素贝叶斯分类算法和贝叶斯信念网络算法。朴素贝叶斯分类算法假定一个属性值对给定类的影响独立于其他属性的值。贝叶斯信念网络算法的图形呈网状,能表示属性子集间的依赖关系。

3. BP 算法

BP 算法构建的模型是在非递归神经网络上学习得到的模型,本质上是一种非线性判别函数,适用于在普通方法无法解决、需要用复杂的多元函数进行非线性映照的数据挖掘环境中,完成半结构化和非结构化的辅助决策支持过程。但是在使用过程中要注意避开局部极小的问题。

1.4.2 关联分析法

关联分析法注重分析结果的实际效用,而且操作起来比较便捷,可从海量的基础数据中提取数据之间的联系。一般来讲,如果两种数据或者多种数据之间有某些联系,它们之间往往存在某种规律,可以存在单层次关联或多层次关联,也有单一维度的关联、多个维度的关联。

关联分析法的一个典型案例是“啤酒和纸尿裤”。某连锁百货超市通过分析顾客的购物数据发现,购买啤酒的顾客经常同时购买纸尿裤。基于此类销售数据,该连锁百货超市将啤酒和纸尿裤摆放在相邻的位置,啤酒和纸尿裤的销量激增。

1.4.3 比较分析法

比较分析法是使用频次最高的商务数据分析方法,是商务数据分析的基础方法。任何事物的对比都有一个参照对象,比较分析法通过分析不同组数据之间的区别得出结论。因此,在比较分析法的使用过程中,最少应有两组数据,两组不同数据的对比能够直观呈现两者的区别,并且存在静态对比和动态对比两种不同的比较分析法。

1. 动态对比

动态对比是指同一事物在不同时间维度下产生的数据间的对比,主要包括同比、环比和定基比。

(1) 同比。同比是指将现有数据与过去同一时间段内产生的数据进行对比。同比增长率的计算公式为



$$\text{同比增长率} = (\text{本期值} - \text{同期值}) \div \text{同期值} \times 100\%$$

例如,2022年“618”期间,某女装店的销售额为360万元,而该女装店2021年同时期的销售额为200万元,因此该女装店铺2022年“618”期间的销售额同比上年增长80%。

(2)环比。环比是指将现有数据与相邻时间段内产生的数据进行对比。环比增长率的计算公式为

$$\text{环比增长率} = (\text{本期值} - \text{上期值}) \div \text{上期值} \times 100\%$$

例如,某奶茶店2022年6月的销售额为6万元,而该店铺2022年5月的销售额仅为1万元,故该店铺2022年6月的销售额环比5月增长500%。

(3)定基比。定基比是指将现有数据与某个固定时间段内产生的数据进行对比。定基比发展速度的计算公式为

$$\text{定基比发展速度} = \text{本期值} \div \text{定期值} \times 100\%$$

例如,受物流运输因素制约,某发制品公司2022年3月的假出货量仅8万件,该公司及时谋求与多家物流公司合作,出货数量不断攀升,到2022年7月,出货量就翻了两番,达到32万件,2022年9月的出货量更是达到了54万件,故该公司2022年7月出货量相比3月出货量的定基比为400%,2022年9月出货量相比3月出货量的定基比为675%。

2. 静态对比

静态对比是指将相同的时间维度下不同指标数据进行对比,主要指标包括目标、行业和区域等。

(1)目标对比。目标对比是指将现有数据与预先制定的目标数据进行对比。例如,某连锁糕点公司在2022年年初制订了年度目标规划,规定2022年的总销售额不低于144万元。那么,该公司可将总销售额目标细分为月销售额不低于12万元,并按月进行目标对比。

(2)行业对比。行业对比是指将现有数据与行业数据进行对比。例如,京东商城上某体育用品店铺的羽毛球拍当日下单量为300单,而京东商智给出的平台所有商铺的日均下单量为132单,这说明该店铺销量高于行业平均水平。

(3)区域对比。区域对比是指将现有数据在不同区域(如国家、地区等)中进行对比。例如,将某跨国公司2022年在各个国家和地区的销售及市场占有率进行对比。此外,区域对比常常将某个局部地区与整个地区进行对比,如将某连锁美妆店的知春路分店与所有分店的客流量进行对比。

1.4.4 漏斗分析法

漏斗分析法是一种常见且有效的商务数据分析方法,主要用于分析商品的成交转化率、店铺的获客率等数据。客户从浏览商品到最终下单会经历一系列操作,使用漏斗分析法可以完整地反映和量化消费活动各环节中客户的流失和存留,有助于商家有效定位损耗和流失较高的节点,从而采取具有针对性的营销策略。

例如,抖音平台上某商品的漏斗如图1-5所示,该商品的曝光量为25 000人次,访问量为18 000人次,咨询量为9 500人次,收藏量为3 500人次,成交量为2 100单,成交转化率为8.4%。

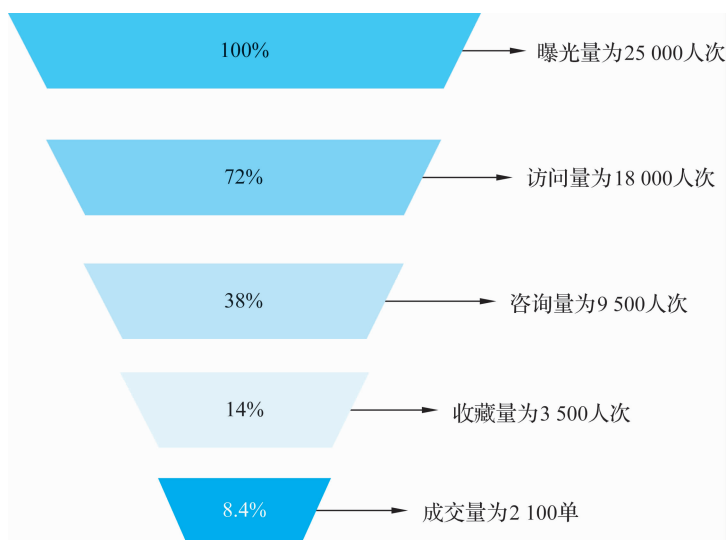


图 1-5 某商品的漏斗

1.4.5 矩阵分析法

矩阵分析法也称象限分析法,是一种以事物的两个重要属性为分析依据,将各个比较主体划分到 4 个象限中进行分组分析与关联分析的方法。

例如,某护肤品公司以面霜的市场占有率为横坐标、以销售增长率为纵坐标建立了分析矩阵,如图 1-6 所示。

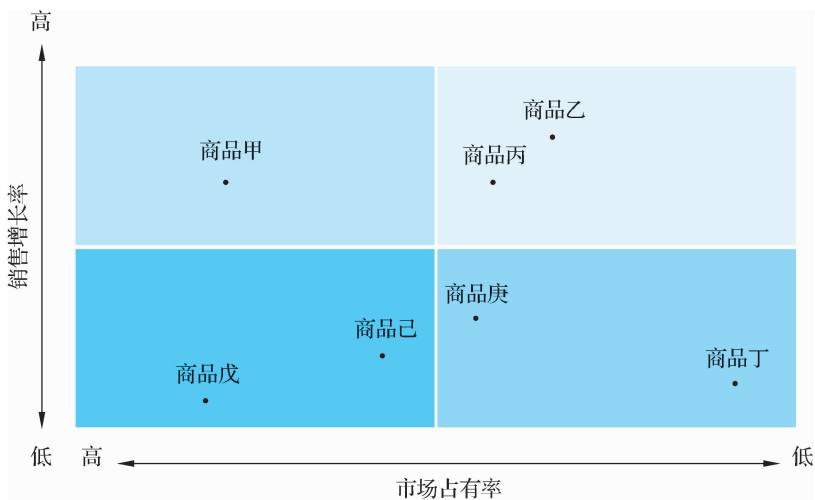


图 1-6 分析矩阵

由图 1-6 可知,商品甲位于市场占有率和销售增长率双高的象限,商品乙和商品丙位于市场占有率低但销售增长率高的象限,商品戊和商品己位于市场占有率高但销售增长率低的象限,商品丁和商品庚位于市场占有率和销售增长率双低的象限。

矩阵分析法主要是利用直观的象限将企业经营管理中遇到的主要问题进行等级排列,



按照这种思路,企业经营者可以非常清晰地认识到企业存在的主要问题,通过合理布局解决问题的路径,达到各方的平衡协调,提高整体企业运行的效率。矩阵分析法为企业经营者做出科学的策略提供了有力的参考。

例如,图 1-6 中,商品甲市场占有率高、销售增速快,属于明星商品,企业应积极采取措施进一步扩大该商品的市场规模,增强其竞争力;商品丙和商品乙所处市场的机遇大、前景好,但可能在市场营销方面存在问题,属于存在问题的商品,企业应对这两种商品采取选择性投资战略,将对这两种商品的改进与扶持方案列入企业的长期计划中;商品戊和商品己的市场占有率高,可为企业提供大量现金流,但由于销售增速缓慢,企业无须对其做进一步投资,而可将其作为明星商品的资金后盾;商品丁和商品庚的市场占有率低、销售增速慢,属于保本或亏损型商品,企业应对这两种商品采取逐渐减量甚至淘汰战略。

1.4.6 结构分析法

1. 结构的概念

结构是指事物的构成方式。一个集体在不同的结构下会成为不同的组织。例如,海尔在组织结构方面经历了三个阶段:在第一个阶段,海尔是按照传统科层制组织结构的;在第二个阶段,为了形成以客户为中心的组织,海尔构建了“人单合一”的体系,高层管理者变成服务者,整个组织倒过来了;如今即第三个阶段,海尔进行内部创新,构建创业平台,变成了一个自负盈亏的创业平台组织。对同样一个集体,采用不同的结构,就会形成不同的组织,也就会有不同的活力。

企业组织的力量配置决定了其优势所在。企业在哪里投入的资源多,一般就会在哪里有优势。例如,在国内某一行业的企业中,营销人员的工资占比为 16% 左右,一些企业的营销人员工资占比甚至达到 20%。强大的营销团队可以确保企业在品牌建设、渠道建设、客户服务等方面具有明显的优势。

结构好比事物的“配方”。一家企业的资源是有限的,如果将资源更多地配置在销售上,该企业就是一家销售导向型企业,就会有强大的销售团队和渠道;如果将资源更多地配置在技术研发上,该企业就会有很强的技术竞争力和过硬的产品;如果将资源更多地配置在生产上,该企业可能就是一家工厂。

事物的结构特征决定了事物的属性。企业必须在结构设计上确保按照既定的战略方针发展,这是决策者需要考量的内容。决策者设计了什么,企业就必须在资源配置上确保有相对应的“结构”,这样才会造就正确的企业属性。

例如,不同层级员工之间工资的对比关系是一种结构,并且是一种层级对比结构,这种结构会影响科层制组织结构的有效性。

又如,工资通常由不同的部分组成,这是一种配方式结构关系。在总监级别员工的工资中,如果变动工资占比非常高,他们的业绩压力就会很大;而层级越往下,变动工资占比越低,激励程度越低,员工的工作积极性就会越低,因为在多数情况下,这些员工会觉得干多干少都一样,进而会导致上级指挥下级比较难。所以,在薪酬结构条件下,公司的业绩压力主要集中在上层员工身上,会导致管理难度加大,容易产生命令式领导等非现代化管理方式,从而影响整个团队的和谐。

结构是指事物内部的组织架构,事物的内部体系搭建决定了其外部呈现出的特征,同

时,公司内部力量在不同组织的分配决定了不同部门的综合竞争力,图 1-7 所示为电商公司的组织架构。例如,某电商公司销售部员工的工资占全公司员工工资的 20%,比其他部门员工的工资比重都要高,该公司产品的销售业绩果然节节攀升。

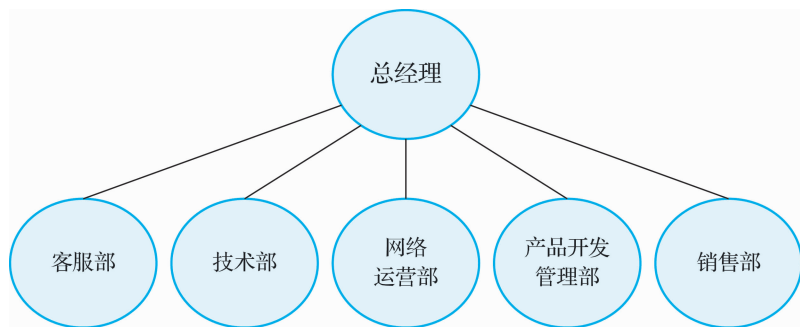


图 1-7 电商公司的组织架构

2. 结构的基准

事物的结构是有基准的,常识是最初级的基准。例如,在餐馆吃饭时,客人会对菜品的味道有一个基本的预期和评价,如果比预期低,就会觉得该餐馆不好。这就是在用常识评价事物的结构。

员工的工资包括基本工资、岗位工资、浮动工资、奖金、津贴、福利等,它们合理的比例应该是多少?这就要看不同层级和不同行业的薪酬结构关系,这时“行规”就会成为人们认识薪酬结构的基准。

如果普通文职人员的月度工资波动幅度超过 10%,就容易造成人员流动性增大;如果研发人员的月度工资波动幅度超过 5%,他们就会不专心研究而频繁关注自己的工资,从而影响专注程度;如果销售人员的业绩浮动工资占比低于 30%,他们就会认为没有获得高收入的机会,从而变得懒散。不同的岗位和层级都有一个预期的薪酬结构基准,“存在”本身就会成为结构的基准。如果一家公司推行薪酬改革,员工对新方案的评价标准就是原有的方案,即通过对比原有的方案评价自己的得失,这些就会成为认识新结构的基准。

3. 结构分析法概述

结构分析法又称占比分析法,是指在对原有数据汇总排列的基础上,分别计算各个部分在总体构成中所占有的比重,通过占有比例数据和外部呈现的形态分析某一事物内在的总体构成部分,归纳出其中的规律,并总结上升为一定的理论的分析方法。

结构分析法一般采用结构指标。结构指标的计算公式为

$$\text{结构指标} = \text{总体中的某一部分} \div \text{总体} \times 100\%$$

可见,结构指标实际就是总体中某一部分占总体的比重,因此总体中各个部分的结构相对数之和等于 100%。

1.4.7 因果关系分析法

因果关系是指某个要素一定会导致相关结果出现。因果关系分析法主要用于明确某种结果的产生,有哪些要素能够产生影响,影响的权重有多大,研究的是变量之间的关系。

因果关系分析的展开是在相关已有研究的基础上的,当原因和结果之间有很强的联系



时,使用因果关系分析法才符合场景要求。因此,因果关系分析需要先进行变量之间的关系研究。

因果关系分析法需要用到回归分析。回归分析是确定两个或两个以上变量之间相互依赖的定量关系的一种统计分析方法,主要研究一个随机变量与另一个变量或一组变量的相倚关系,为变量之间的关系分析提供可靠的依据。

古人就善于运用因果关系分析法经营以谋取利润,从而获得成功。孔子的弟子子贡非常擅长经营之道,头脑灵活,很会洞察商机。某年冬季,子贡得知吴国军队将要去北方攻打城池,他料定吴王夫差肯定会强征御寒衣物保证将士顺利远行。他利用因果关系分析法推测吴国的御寒衣物需求量将会大增,于是子贡召集人员到不同的地方购买御寒衣物运回吴国,结果吴国的百姓纷纷抢购,很快御寒衣物被销售一空。

1.4.8 聚类

聚类是指对数据进行记录分组,把相似的数据记录在一个聚集中。聚类和分类的区别是聚类不依赖于预先定义好的类,不需要训练集。

例如,一些特定症状的聚集可能预示了一种特定的疾病;租 VCD 类型不相似的客户聚集,可能暗示他们属于不同的亚文化群。

聚集通常作为数据挖掘的第一步。例如,“哪一种类的促销客户响应最好”,对于这一类问题,首先对整个客户群进行聚类,将客户分在各自的聚集中,然后对每个不同的聚集提出问题,可能效果会更好。

聚类算法包括统计分析算法、机器学习算法、神经网络算法等。

在统计分析算法中,聚类是基于距离的,如欧氏距离、海明距离等。聚类是基于全局比较的,需要考察所有的个体才能决定类的划分。

在机器学习算法中,聚类是无监督的学习。在这里,距离是根据概念的描述确定的,故聚类也称概念聚类。当聚类对象动态增加时,概念聚类则转变为概念形成。

在神经网络算法中,自组织神经网络可用于聚类,如 ART 模型、Kohonen 模型等。神经网络算法是一种无监督的学习方法,即当给定距离阈值后,对各个样本按阈值进行聚类。它的优点是能非线性学习和联想记忆,但也存在一些问题,如不能观察学习过程,最后的输出结果较难解释,从而影响结果的可信度及可接受程度;神经网络需要较长的学习时间,对大数据量而言,其性能会出现严重的问题。

1.4.9 预测序列法

常见的预测序列法有简易平均法、移动平均法、指数平滑法、线性回归法、灰色预测法等。

简易平均法是一种简便的时间序列法,是以一定观察期的数据求得平均数,并以所求得平均数为基础,预测未来的数据。

移动平均法是用一组最近的实际数据预测未来一期或几期内产品的需求量、企业的产能等的一种方法。

指数平滑法是在移动平均法基础上发展起来的一种时间序列法,是通过计算指数平滑值,配合一定的时间序列预测模型对对象的未来进行预测的。它能减少随机因素引起的波



动和检测器错误。

灰色预测法是建立在灰色预测理论的基础上的。在灰色预测理论中,系统的发展有其内在的一致性和连续性,将系统发展的历史数据进行若干次累加和累减处理,所得到的数据序列将呈现某种特定的模式(如指数增长模式等),挖掘该模式,然后对数据进行还原,就可以预测系统的发展和变化。灰色预测法是一种对含有不确定因素的系统进行预测的常用定量方法。通常来说,在宏观经济的各行业中,受客观政策及市场经济等各方面因素的影响,可以认为这些系统都是灰色系统,均可以用灰色预测法描述其发展、变化的趋势。灰色预测是对既含有确定信息又含有不确定信息的系统进行预测,也就是对在一定范围内变化的、与时间序列有关的灰色过程进行预测。尽管灰色过程中所显示的现象是随机的,但毕竟是有序的,因此得到的数据集具有潜在的规律。灰色预测通过鉴别系统因素之间发展趋势的相异程度,即进行关联分析,并对原始数据进行生成处理以寻找系统变动的规律,生成有较强规律性的数据序列,然后建立相应的微分方程模型,以此预测事物的发展趋势。

线性回归法中,线性回归模型是通过处理数据变量之间的关系,找出合理的数学表达式,并结合历史数据来对将来的数据进行预测的。

1.4.10 估计

估计与分类相似,不同之处在于:分类描述的是离散型变量的输出,而估计处理连续值的输出;分类的类别是用于确定数目的,估计的量则是不确定的。

例如,根据购买模式,估计一个家庭的孩子个数或估计一个家庭的收入,或者估计房产的价值。

一般来说,估计可以作为分类的前一步工作。给定一些输入数据,通过估计得到未知的连续变量的值,然后根据预先设定的阈值进行分类。例如,银行对家庭贷款业务运用估值给各个客户记分,然后根据阈值将贷款级别分类。

1.4.11 预测

通常,预测是通过分类或估计起作用的,也就是说,通过分类或估计得出模型,该模型用于对未知变量的预测。从这种意义上说,预测其实没有必要分为一个单独的类别。预测的目的是对未知变量的预言,这种预言是需要时间验证的,即必须经过一定的时间后才知道预言的准确性。

1.4.12 描述和可视化

描述和可视化是表示数据挖掘结果的一种方式。

数据挖掘的结果需要以易于理解的方式呈现出来,这就是描述和可视化的作用。描述和可视化是一种将数据挖掘结果转化为人类可理解信息的方式。描述通常包括对挖掘结果的文字描述,可视化则使用图形、图像或动画等视觉元素展示挖掘结果。通过描述和可视化,用户可以更直观地理解数据挖掘的发现,并更好地解释和探索数据中的模式和数据间的关系。

描述和可视化在数据挖掘过程中扮演着重要的角色。当数据挖掘产生结果后,这些结果通常以数据集的形式呈现。然而,这些数据集可能包含大量的数据和复杂的模式,对于不熟悉数据挖掘的人来说可能很难理解。因此,通过描述和可视化,可以将数据集转化为易于理解的形式,以便更好地解释和探索数据中的模式和数据间的关系。



描述和可视化可以以多种方式呈现。例如,可以使用表格、图表、图形、图像或动画等元素展示挖掘结果。这些元素可以单独使用或组合使用,以便更好地解释和探索数据中的模式和数据间的关系。此外,还可以使用自然语言描述挖掘结果,以便不熟悉数据可视化的人也能够理解。



课堂案例

数据挖掘帮助 DHL 实时跟踪货箱温度

DHL 是国际快递和物流行业的全球市场领先者,提供快递、水陆空三路运输、合同物流解决方案,以及国际邮件服务。DHL 的国际网络将超过 220 个国家及地区联系起来,员工总数超过 28.5 万人。在美国食品药品监督管理局(Food and Drug Administration, FDA)要求确保运送过程中药品装运的温度达标这一压力之下,DHL 的医药客户强烈要求提供更可靠且更实惠的选择。这就要求 DHL 在递送的各个阶段都要实时跟踪集装箱的温度。

虽然由记录器生成的信息准确无误,但是无法实时传递数据,客户和 DHL 都无法在发生温度偏差时采取任何预防和纠正措施。因此,DHL 的母公司——德国邮政世界网(Deutsche Post World Net, DPWN)通过技术与创新管理明确拟定了一个计划,准备使用射频识别技术在不同时间点全程跟踪装运的温度,并由 IBM 全球企业咨询服务部绘制决定服务关键功能参数的流程框架。DHL 获得了两个方面的收益:对于最终客户来说,能够使医药客户对运送过程中出现的装运问题提前做出响应,并以引人注目的低成本全面切实地增强了运送的可靠性;对于 DHL 来说,提高了客户满意度和忠诚度,为保持竞争差异奠定了坚实的基础,并新增收入来源。

1.4.13 回归分析

回归分析是用已知的函数关系拟合观测数据的主要关系趋势的一种方法。

“回归”这一术语最早是由英国学者高尔顿(Galton)在研究人类的遗传规律时提出的。孩子的平均身高会介于其父母和种族的平均身高之间,这是一种回归现象,即孩子身高回归到种族平均身高的趋势。“回归”的概念自在遗传学领域诞生后,就广泛地被用于各个领域。

回归分析是确定两个或两个以上变量之间相互依赖的定量关系的一种统计分析方法,主要研究一个随机变量对另一个变量或一组变量的相依关系。回归分析按照涉及的自变量的数量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。

以一元线性回归分析为例,如果散点图的分布呈明确的直线趋势,就可以绘制一条最能代表散点图分布趋势的直线,这条直线称为回归线,如图 1-8 所示。根据该分析结果,可以建立自变量和因变量之间关系的函数方程,其称为回归方程。

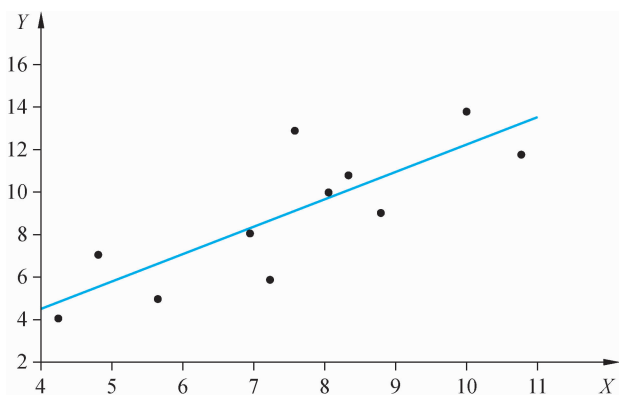


图 1-8 回归线示例

多元线性回归分析主要描述一个因变量如何随着一批自变量的变化而变化,其回归公式(回归方程)就是因变量与自变量关系的数据反映。因变量的变化包括两部分,即系统性变化和随机变化。其中,系统性变化是由自变量引起的,即用自变量可以解释;随机变化是不能由自变量解释的,通常也称为残值。

估算多元线性回归方程中自变量的系数最常用的是最小二乘法,即找出一组对应自变量的相应参数,以使因变量的实际观测值与回归方程的预测值之间的总方差最小。

作为检验自变量对因变量影响的统计分析方法,回归分析的前提是事物(或变量)之间确实存在因果关系,否则便不能进行回归分析。例如,某学生身高与校园中树木的高度之间是共变关系,学生的身高越高,树木的高度也越高,但这并不等于学生身高的变化是树木高度变化的原因,或者树木高度的变化是学生身高变化的原因,两者之间并没有因果关系,所以不能进行回归分析。换一个角度看,学生身高和树木高度均与时间有关,学生的身高随着时间的增长而增长,树木的高度也是如此。其中,时间是自变量,学生身高或树木高度是因变量,进行回归分析是合理的。

回归分析有助于人们把握因变量受其他一个或多个自变量影响的程度,为预测提供科学依据。例如,可以用线性回归的方式研究产值与工人数量的关系,从而评价或预测不同产值下大概需要的工人数量,以便人力资源部门合理招收或调动工人。通常来说,回归分析包括如下几个步骤。

(1)确定自变量。一个因变量可能受到一个或多个自变量的影响。因此,在进行回归分析之前,需要明确哪个是自变量以及自变量的个数,并采用适当的方法收集自变量和因变量的数据。

(2)选择回归分析的模型。实证领域的模型一般是参数模型,即回归函数的数学形式已经确定,只是包含若干未知数。回归函数可以分为多种形式,如一元一次函数、一元二次函数、多元一次函数等。

(3)估计模型中的参数。根据最小二乘法,计算出回归方程的截距和回归系数、确定系数、残差等。

(4)检验模型。检验模型是指对回归方程的回归效果进行检验。如果检验结果良好,则说明回归效果极佳,回归方程具有较高的应用价值。如果检验结果不理想,则需要反思和核



查自变量的代表性及数据是否存在异常等问题,特别需要反思自变量和因变量的因果关系是否成立。

1.4.14 企业职能部门具体数据分析

1. 营销和销售管理数据分析

营销和销售管理数据分析与一家企业的销售模式有很大的关系。其没有统一的准则,需要营销和销售管理人员根据自己的业务模式、市场情况和客户情况采用不同的方式进行。

以电商的销售模式为例,其主要关注网站或店铺的数据分析,包括对流量、点击率、转化率、关注客户数、注册用户数、用户活跃度、订单量、客单价和营销费用等的分析。此外,还可直接借助在线工具实现精准数据分析。淘宝网、天猫商城、京东商城等电商平台会提供一些数据分析方法和工具,可以在线实时看到分析结果,而且能够针对一个时间段、一次促销活动或一款产品等进行比较详细的分析。

如果采用传统的分销模式,生产厂家对数据的掌握就不如电商模式直接,这是因为电商模式可以提供全平台的数据作为分析的参照,而传统的分销模式只能依赖销售管理系统获得相关数据。针对消费者需求的调研也只能自行进行,没有现成的平台数据可以直接使用。

有些企业(如采用会员制的连锁经营体系的企业)的数据比较健全,其通常会有比较详尽的数据。而对于匿名的销售模式(如连锁商超、连锁餐饮店等模式),由于根本不知道来店消费的人是谁或对其信息知之甚少,也就无法跟踪,所能够分析的数据通常只有收银数据,这会大幅度削弱企业对未来的预测能力。

无论采用何种经营模式,都只能根据已有的数据进行分析,而对缺少数据的环节是无法进行分析的,要么投入高昂的费用采集数据,要么进行抽样调查。因此,在不同的模式下要结合实际情况进行数据分析。

从方法和思路上看,营销和销售管理中的数据分析必须从客户、产品或服务、市场、区域和员工等维度进行,同时要结合数据的可用性和可采集性等采集更多的数据,这是因为对任何数据经过分析都可能得出更多有价值的信息,这有利于更好地做出决策。

(1)客户维度。从客户维度进行分析是首要的。对任何产品或服务都要考虑目标客户的需求,因此要对客户进行各方面的分析。对于客户的数据,可以对客户进行画像、细分、需求研究、行为研究和习惯研究,进而指导制定营销策略、产品开发和改进策略;客户与企业的接触点数据(如收银数据、订单数据、送货数据、会员卡数据等)反映了企业的经营活动效果,可以在这类数据的基础上了解客户对产品或服务的反应,从而为企业的经营活动提供更多的借鉴。

(2)产品或服务维度。可对产品或服务做出评价,如分析企业是否有足够好的产品线结构,是否有现金流产品,是否有明星产品,是否有盈利的产品,是否有品牌产品,等等。另外,还可以从竞争的角度进行分析,如分析自己的产品和市场上已有产品的竞争力,分析自己的产品和其他产品的差异,从而为企业改进产品、提高产品竞争力提供借鉴。

(3)市场维度。应从细分市场的角度分析产品的定位,是处于高端市场、中端市场,还是处于低端市场。不同的细分市场有不同的逻辑,不同的产品在不同的市场上会有不同的表现。

(4)区域维度。虽然互联网打破了市场区域分隔的边界,但仍然有很多类别的产品具有



区域属性,如渠道、配送、社区服务等。区域仍然影响着销售方式,区域的客户特征影响着产品和业务的选择模式。

(5)员工维度。员工维度也称业务团队维度,是指分析哪些业务员在成长,哪些业务员在倒退,哪些业务员是有贡献的,哪些业务员在驱动业绩增长,如何驱动业务员的绩效提升,业务员的激励机制是否合理,业务员是否有足够的的能力,业务员的策略是否正确,等等。

根据企业的销售模式,营销和销售管理中的数据分析可以有更多的维度,这需要按照上述思路进行研究和拓展。

2. 财务管理数据分析

国内企业的财务管理数据一般都比较规范,因为会涉及交税,财务会计都有记账的规范,并且每个月都需要向工商部门、税务部门提供报表。会计准则将大多数的财务问题都解释得非常清楚,因此企业会保留很多年的历史财务数据。

常规的财务管理数据分析已经有比较完善的模型和标准,很多财经网站和媒体都会直接提供一些财务分析指标,以便企业分析自己的财务状况。

数据分析领域的财务分析与常规的财务分析最大的区别在于主体的细化。管理财务需要细化到每家子公司、每个业务部门、每个业务单元、每种业务、每种产品、每位客户和每个订单,然后以它们为主体进行财务数据分析。这就需要企业的财务管理人员建立足够的明细科目,确保能够把成本和费用等数据计入相应的主体。

财务数据管理是一项复杂的工作。几乎在所有的企业中,财务数据都是最为准确和精确的数据,数据化管理和追溯机制也最为完善。对于每一笔费用,财务人员通常都能够找到原始的单据和审批细节。财务管理数据分析则需要将成本和费用明确地计入需要分析的主体,这一主体能够细化到一位客户、一个订单、一种产品、一个业务单元。

无论主体是谁,财务管理数据分析的方法都是一样的。评判财务管理是否细致,关键要看企业的成本和费用能够细化到什么程度。如果能够按照产品批次、生产订单、销售订单、客户、业务单元等进行细化,财务管理的细致程度就算达标了,否则就需要继续细化。

3. 人力资源管理数据分析

人力资源管理数据分析包括以下两个方面。

(1)人力结构分析。企业的人才结构通常包括不同职能部门的人才结构、不同层级的人才结构、不同文化程度的人才结构、不同年龄的人才结构、不同性别的人才结构等。此外,在进行数据分析时,还可以采用嵌套结构,如分析女性员工的年龄结构等。

人力结构反映了企业人力资源的构成情况,如果不合理,企业的发展就会出现问題。例如,对于年龄结构,如果在某个年龄段出现断层,则很容易导致企业在某个时间段出现人才的断层,这就需要在招聘方面做好预案和人才储备,以弥补人才断层带来的影响。

结构分析法就是研究资源配置的,而人力资源是企业中最重要的资源,其配置将直接影响企业的整体资源配置情况。因此,需要深度分析各种资源的配比结构,以便获得更加合理的人力配置,确保企业的经营和管理能够顺畅进行。

(2)人力效能分析。跟踪人力效能,能够有效地监测企业的经营状况,了解每位员工、每个部门和企业整体的人力效能情况。如果发现人力效能降低,人力资源管理部门就需要认真分析是由环境导致的还是由内部管理导致的,是人才素质的问題还是人员发展的问題,是



否需要通过培训加以解决,是否需要引进优秀人才以带动整体人力效能的提升,等等。

人力效能分析即人效分析。在人效分析中,应该关注两种指标,即人均产出率和人员费用产出率。

①人均产出率。人均产出率代表了每个人的产出。人均产出率越高,代表集体力量越雄厚;人均产出率越低,代表集体力量越薄弱,集体成员的效率越低。以人均产值为例,其计算公式为

$$\text{人均产值} = \text{总产值} / \text{总人数}$$

例如,平均每位员工的产出去年是150万元,今年是120万元,说明人力效能降低了。至于人力效能降低的原因是什么,是人员效率降低了还是受外部环境的影响,则需要进一步进行分析。

②人员费用产出率。以元当产值为例,其代表每投入一元的人员费用的产出。其计算公式为

$$\text{元当产值} = \text{总产值} / \text{工资总额}$$

其中,工资总额包含基本工资、提成、奖金、分红等。如果一家企业连续多年都维持同样的人均产值,其每一元工资所产生的销售额通常会降低,因为员工工资是不断上涨的,这其实代表了员工效率的降低。员工工资上涨,人均产值也应增加,否则就会导致企业的利润率降低。

在进行人力效能分析时,对产出变量可以采用不同的衡量指标。例如,对于电商企业,可以采用获客数量、新客数量、流失客户数量等指标衡量产出;对于房地产企业,可以采用销售面积、销售套数、竣工面积、销售额等指标衡量产出;对于网站,可以采用访问量、注册量、用户数、活跃用户数等指标衡量产出。

4. 物流和供应链管理数据分析

物流和供应链管理中的数据分析主要关注效率和成本。为了保证生产,通常需要较多的备料,但这会占用库存,降低资产流动的速度。此外,供应链往往不够透明,会导致上下游之间并不清楚彼此的库存量,而畅销产品更容易造成库存积压,因为生产厂家都希望拿到更多的备料,以免畅销产品断货而影响销售。这些都是物流和供应链的问题。

互联网为上下游之间即时共享数据提供了便利条件。通过信息系统的对接,上下游之间能够在业务联络上看到每个环节拥有的库存、库存的消化周期,从而可以做到提前备货,并使产品库存降到最低,以提高资金的周转效率。

随着物联网的发展,供应链环节上的数据采集更加方便和及时,在每一个环节通过扫码就可以自动监控货物的来龙去脉。如果一家公司物流不畅、供应链中存在高库存现象,就应该考虑在信息技术和物联网技术方面增加投入,以便通过更加详尽和及时的数据保障物流速度和降低库存。

5. 物流和供应链管理中的数据分析

物流和供应链管理中的数据分析可以参考现实生活中的以下两个模型。

(1)玻璃酒杯模型。一群朋友饮酒,为了让每个人都喝得尽兴,但又不至于喝醉,这时就要建立一个供应和饮酒的机制。

①需要清楚每个人的酒量,这相当于了解每条渠道或每家店铺的销售能力,以便预测企

业未来一周或一个月的销量。

②必须随时知道每个人已经喝了多少,这就需要采用透明且标有刻度的杯子。只有这样,对于每个人喝了几杯、杯子里还有多少酒、还能喝多少,才会非常清楚。这相当于监测每条渠道或每家店铺的销量,将数据及时上报,即已经售出多少、还剩多少、还能售出多少。每天都必须有一张报表,或者随时都能通过系统查看相关数据。

③酒瓶必须透明且标有刻度,这样就能够随时看到剩余酒的量,以便判断还能供多少人喝,以及是否需要及时补充等。这相当于了解仓库的情况,如库存是多少、所服务店铺的预计需求是多少,以便预测备货的时间和品种,并及时备货。

④如果众人聊得特别投机,现场气氛热烈,服务员就要提前做好估计,时刻准备着到前台拿酒。这相当于市场突然发生了变化,当现实的情况与原来的预测发生偏差且需要更多的货物时,供应系统就要启动应急预案(如生产工厂启动应急计划等),以确保能够及时供应货物。

⑤如果现场气氛不太好,估计要提前散场,组织者就要开始与服务员商量,尝试将剩余的酒退回;如果不能退,就要将剩余的酒带走。这相当于市场情况没有达到预期,当进货太多而出现滞销时,店铺或渠道商就要及时退货;如果销售条款规定无法退货,就需要店铺或渠道商自行寻找其他出路(如促销等),以消化积压的货物,将库存货物转化为现金流。

使用玻璃酒杯模型要求数据透明化,并且有精准的预测以及快速的应变机制,包括对畅销产品和滞销产品的应急预案。当产品进入供应链时,就要建立销售预测模型、物流跟踪模型、数据透明化机制,一旦事实与预测不符,就要立即启动应急预案。

(2)自来水供水模型。为了满足居民的用水需求,自来水公司必须合理安排生产,确保水压、水位稳定。自来水公司的供水模型包括以下几个原理。

①要根据水位安排生产,一旦水位降低到某个临界值,就要启动生产,同时对每天的用水量进行监测,即观测一天中哪个时间段的用水量高、哪个时间段的用水量低,并对历史数据进行分析,掌握其变化规律。如此一来,就可以在用水高峰到来之前保持一个较高的水位,在用水量较少时保持一个较低的水位。这相当于必须在库存管理上有一个明确的“水位”,并随时掌握“水位”数据。任何一家店铺销售了产品(“水位”降低),都需要反映到总表上(全程可见)。

②整个供水管网系统由粗到细,可以确保水压不会大幅度降低,并且供应量充足;然后逐渐出现分支,形成树根状结构,让水管保持固定的水压,并将水压的数据传递到供水调度中心,一旦水压过低就发出预警信号,通知相关部门及时补水,以提高水压。这相当于货物供应要保持顺畅、稳定,必须时刻监测,及时发货和补货。

③一旦某处漏水,就会导致该处的用水量严重超出平时的预计范围,相关人员就需要迅速跟进和抢修。这相当于一旦有销售点串货,销售数据就会产生异常,销售负责人就要立即跟踪产品的销售过程,对每个销售点进行审计。

④自来水管道是相通的,如果一个小区的用水量少,水就会自动流到用水量多的其他小区(根据水压控制原理,水会自动流向水压低的地方)。这相当于在产品配送过程中应以小批量为原则,并以“水压”为衡量指标。对“水压”低的店铺应优先配送,而当一家店铺的“水压”变高时,产品可以向“水压”低的店铺流转,从而确保终端供货的流动性。



⑤自来水计量表被安装在用水端,只有流出了水龙头的水才是居民消费的。同理,根据库存原理,为了确保终端产品相互之间的流动性,无论是直营店,还是加盟店或授权店,只有销售出去的产品才是店铺的,否则都是厂家的,即所有的库存产品都由厂家自由支配。如果需要调货或换货等,店铺必须全力支持,不能私自囤货,以确保产品的流动性。

除了以上两个模型,还有人工神经网络模型、植物营养供给模型等。它们都是通过研究生活中的现象为物流和供应管理提供思路。无论是什么样的模型,数据化、透明化、全程可视化和实时追踪都是必不可少的要素。在现有的技术条件下,建立这些模型已不再是难事,并且不需要较大的投入。



课后练习

参照某市近六年的地区国内生产总值(gross domestic product, GDP)数据(见表 1-1),以四种不同的商务数据分析方法分析数据背后蕴含的规律。

表 1-1 某市近六年的地区 GDP 数据

年 份	地区 GDP/亿元	增 长 率
2023 年	31 235.08	7.9%
2022 年	28 937.56	7.8%
2021 年	26 835.87	13.49%
2020 年	23 645.79	3.38%
2019 年	22 873.06	5.26%
2018 年	21 730.66	4.6%