

免费提供

*** 精品教学资料包

服务热线: 400-615-1233
www.huatengzy.com

Python基础 及其在经管中的应用

Python基础及其在经管中的应用

主编 张海鸥 刘先花 刘庆红

高等职业教育立体化精品教材

Python基础 及其在经管中的应用

主编 张海鸥 刘先花 刘庆红
副主编 王哲 张庆丰 孙冲 王可阳

以“数据与人生”模块为引领，帮助学生塑造正确的**价值观和择业观**
以**实用**为主，培养学生的**Python实践运用**能力
以**项目**为驱动，培养学生**综合化**的职业能力

特约编辑: 张海红
责任编辑: 李特军
责任校对: 陈何真璐
封面设计: 刘文东

ISBN 978-7-5504-5877-2



定价: 45.00元



西南财经大学出版社
Southwestern University of Finance & Economics Press



西南财经大学出版社
Southwestern University of Finance & Economics Press

中国·成都

高等职业教育立体化精品教材

Python基础 及其在经管中的应用

主 编 张海鸥 刘先花 刘庆红
副主编 王 哲 张庆丰 孙 冲 王可阳



西南财经大学出版社
Southwestern University of Finance & Economics Press

中国·成都

图书在版编目(CIP)数据

Python 基础及其在经管中的应用/张海鸥,刘先花,刘庆红主编;王哲等
副主编. —成都:西南财经大学出版社,2023. 8
ISBN 978-7-5504-5877-2

I. ①P… II. ①张…②刘…③刘…④王… III. ①软件工具—程序
设计 IV. ①TP311. 561

中国国家版本馆 CIP 数据核字(2023)第 136145 号

Python 基础及其在经管中的应用

Python JICHU JIQI ZAI JINGGUAN ZHONG DE YINGYONG

主 编 张海鸥 刘先花 刘庆红
副主编 王 哲 张庆丰 孙 冲 王可阳

特约编辑:张海红
责任编辑:李特军
责任校对:陈何真璐
封面设计:刘文东
责任印制:朱曼丽

出版发行	西南财经大学出版社(四川省成都市光华村街 55 号)
网 址	http://cbs.swufe.edu.cn
电子邮件	bookcj@swufe.edu.cn
邮政编码	610074
电 话	028-87353785
印 刷	三河市骏杰印刷有限公司
成品尺寸	185mm×260mm
印 张	14.25
字 数	356 千字
版 次	2023 年 8 月第 1 版
印 次	2023 年 8 月第 1 次印刷
书 号	ISBN 978-7-5504-5877-2
定 价	45.00 元

版权所有,翻印必究。



前言

Preface

党的二十大报告指出,“推动构建新发展格局,推进高质量发展”,科技的变革不断推动着财务领域的创新,如 ERP、财务共享、财务中台以及财务机器人等应用层出不穷。然而作为财务人员,我们不能停留在已有的技术应用上,必须不断探索新的技术途径,适应新发展格局的需要。Python 作为编程语言中的热点,其在 WEB 开发、爬虫程序、科学计算、自动化运维、数据分析+数据可视化、云计算、桌面软件、游戏、人工智能等领域都非常活跃,Python 技能成为财务人员必备的基本技能。通过学习 Python,我们可以更加高效地处理海量数据,Python 的处理速度令人惊叹。同时,Python 的循环和判断也可以帮助我们解决复杂的计算逻辑。在新发展格局下,我们应该持续学习和应用 Python 等前沿技术,积极探索财务领域的创新,为推进高质量发展贡献我们的力量。

职业教育是培养技能型、应用型人才的重要途径,教材则是职业教育中不可或缺载体。本书在各个环节中,充分贯彻了基于工作过程的思想,旨在为高职学生提供面向基层、面向生产服务和管理第一线的实用性知识和技能,使他们真正成为一线企业所需要的人才。除了掌握专业知识,职业教育学生更应该具备对应工作岗位的典型工作技能和应用能力。因此,本书的职业教育理念贯穿教材始终,使学生掌握工作所需的实用技能,达到学以致用目的。

考虑到众多读者可能没有编程基础,本书从零开始介绍 Python 语言,并且由浅入深、循序渐进。与目前市场上的 Python 教材不同,本书根据《大数据财务分析》一书中职业技能等级标准中工作任务与职业能力对应表以及高职财务会计类专业教学标准、课程标准中明确规定的“主要教学内容”梳理出职业能力要求对应的知识点和技能,围绕大数据技术在财务管理中的应用设计教学项目,划分任务,将知识、技能、素质融入案例,使教学内容更加具有针对性,便于学生考证以及在将来工作岗位上灵活运用。

本书的主要特点有:

(1) 教材功能:通过“数据与人生”模块让学生热爱职业岗位,树立正确的价值观、择业观,形成良好的职业道德和职业意识。用短小、精练的微课讲述抽象的理论以及任务实施,发挥教材对学生职业能力培养的载体作用。

(2) 教材内容:以职业能力为本位,以应用为目的,实用为主、够用为度,满足职业岗位的需要,与职业技能等级标准接轨,融入行业与企业新动态、新技术。

(3) 教材组织:将“工作情景”“任务分析”作为工作的主线贯穿于完成全部过程,淡化理论知识的系统性、完整性,追求工作过程的综合化和模块化。

(4) 教材资源:配备了演示视频二维码链接资源、教学课件、程序安装包、代码包、课外拓展、资源推荐以及习题答案等。同时,在学习通平台开设“Python 基础及其在经管中的应用”课程,形成可学、可练、可测的全方位立体化教学资源,为院校开展同类课程教学提供便利。

本书建议教学课时数为 48 个学时,实际教学时可依据教学时间及教学对象进行灵活调整。具体每个项目的建议教学课时如下:

教学内容	建议学时	
	理论学时	实践学时
项目一 关于大数据	2	0
项目二 Python 环境搭建	0	2
项目三 Python 语法基础	4	4
项目四 Python 程序控制结构与函数	4	6
项目五 NumPy 数组运算	2	4
项目六 数据处理	4	6
项目七 数据图表展示	2	4
项目八 Python 在经管中的案例应用	2	2
小计	20	28
总学时	48	

本书由吉林省经济管理干部学院张海鸥、刘先花、刘庆红担任主编,王哲、张庆丰、孙冲、王可阳担任副主编。编者在编写本书的过程中得到了本校会计学院梁雪老师以及长春职业技术学院张立辉老师的指导,在此一并表示感谢。

由于编者水平有限,书中难免存在不足之处,欢迎广大读者提出宝贵意见和建议。

编者

2022 年 4 月



目 录

Contents

项目一 关于大数据 1

- 任务一 认知大数据····· 2
- 任务二 培养大数据思维····· 3
- 任务三 大数据的处理流程····· 4
- 任务四 常见的大数据分析模型····· 6
- 任务五 大数据的运用····· 8

项目二 Python 环境搭建 11

- 任务一 Python 环境的安装与设置····· 12
- 任务二 使用 Jupyter Notebook 创建 Python 程序····· 24

项目三 Python 语法基础 37

- 任务一 财务人员的初始 Python 程序····· 38
- 任务二 财务字符串型数据存储····· 48
- 任务三 财务高级类型数据存储····· 55

项目四 Python 程序控制结构与函数 68

- 任务一 财务中的判断语句····· 69
- 任务二 消灭重复计算····· 78
- 任务三 可重复使用的财务算法····· 90
- 任务四 函数精进····· 98

项目五 NumPy 数组运算 109

- 任务一 使用 NumPy 表示及存储数据····· 110
- 任务二 NumPy 数据统计操作····· 117
- 任务三 NumPy 财务分析····· 121

项目六	数据处理	127
任务一	Pandas 中的财务表格	128
任务二	读写数据	134
任务三	清洗数据	140
任务四	统计数据	157
项目七	数据图表展示	173
任务一	认识数据可视化	173
任务二	Matplotlib 销售数据可视化	182
项目八	Python 在经管中的应用案例	205
任务一	线性回归销售额预测	206
任务二	k-means 聚类分析员工成绩	212
参考文献		221

项目一

关于大数据

知识目标

1. 理解并掌握大数据的基本概念。
2. 掌握大数据的处理流程。
3. 了解大数据常见的分析模型。

能力目标

1. 能够运用大数据思维解决实际问题。
2. 能够将大数据灵活应用到生活中。

素养目标

1. 要以遵纪守法、顺应社会伦理道德为最根本要求,正确使用大数据,杜绝大数据的消极影响是大数据时代每个人的责任与义务。
2. 学生通过学习大数据思维,学会用不同的角度和思维思考问题,培养创新能力,推动数据创新,让大数据创造出更高的价值。

工作情景

一些新型的中介机构,如美国 Movenbank、德国 Kreditech 贷款评分公司,都在积极尝试利用社交网络数据构建个人信用分析平台,将社交网络资料转化成个人互联网信用。这些中介机构希望通过说服 Facebook、LinkedIn 等其他社交网络对金融机构开放用户相关资料和用户在各网站中的活动记录,运用大数据分析技术,分析用户在社交网络中好友的信用状况,以此为根据生成客户信用评分。

分析:社交网络平台已经积累了海量用户信息,包括用户在注册应用时填写的个人信息、日常发布的动态、好友关系等,即社交网络用户大数据。中介机构运用大数据思维和大数据分析处理技术,可以分析出用户在社交网络平台中的信用状况,以此产生客户信用评分。其中包括的专业知识有大数据的概念、大数据的思维、大数据的处理流程、大数据常见的分析模型和大数据的应用,帮助读者对大数据和财务大数据形成初步的认知。

任务一 认知大数据



微课
大数据的认知

【任务描述】

近年来,信息技术发展迅速,普及度也越来越高,推动着各行各业的发展,随之而来的是呈指数级增长的数据。对于这些海量数据,传统的计算技术和信息系统的处理能力已经捉襟见肘,因而促进了大数据的产生。大数据时代的到来不仅是信息技术的巨大变革,也和人们的生活息息相关。在全球范围内,大数据也成为许多国家的战略发展目标,受到极大的重视。那么究竟什么是大数据? 大数据有哪些特点?

【相关知识】

首先对于“什么是大数据”这个问题,目前领域中认可度比较高的是关于大数据 4V 的说法。大数据 4V 中的“V”指的是 volume(数量)、variety(多样性)、velocity(速度)、value(价值),这 4 个 V 也可以说是大数据的 4 个特点,分别代表数据量大、数据类型繁多、处理速度快和价值密度低。

1. 数据量大

数据量大指的是所涉及的数据量大到无法在人们所需的时间范围内,通过传统信息技术或主流软件达到存储、撷取、处理等功能。依照这个衡量标准,许多生活中的场景已经达到了大数据的层面。例如,马路上的交通摄像头、治安监控,工厂中的生产监测,手机应用软件微信、微博、短信,等等,都在时时刻刻生成大量的数据,这些也属于大数据的范畴。

2. 数据类型繁多

大数据的数据种类众多,包含图片、视频、音频等,且大数据的数据来源也十分广泛,金融、生物、医疗、交通、航空航天等领域的数据都呈“爆炸式”增长。除此之外,大数据有不同的格式,包括结构化、半结构化和非结构化,多种类型的数据也给数据处理带来了更大的难度。

3. 处理速度快

大数据时代数据产生的速度超乎想象,对于高速且大量的数据,要做到快速分析并生成实时分析结果,传统的数据挖掘技术显然是不能满足的。那么为什么要追求“快”呢? 因为“快”是可以和利润直接联系在一起的,尤其是在电商平台中。当消费者在某一商品页面停留一定时长后,平台会直接判定消费者对此类商品有需求,随之会更多地推送同类商品,增加消费者的购物体验,从而实现利润的增长。

4. 价值密度低

虽然大数据有许多传统关系数据库不能比拟的优势,但实际上大数据的价值密度是很低的。从海量的数据中提取有价值的信息,这个过程就像大浪淘沙,数据量越大,价值密度就越低。例如,对于校园里的监控设备,只有案件发生后,调查人员调取案发时间段的监控录像才是有价值的。然而,为了记录一段很短但有价值的录像,校方需要投入大量成本用于购买监控设备、网络设备、存储设备等,来保存连续的监控数据。

任务二 培养大数据思维

【任务描述】

在大数据时代下,有人能够利用大数据得到“金山银山”,而有人却只能“望洋兴叹”,其中的关键在于大数据思维。也就是说,大数据的发展除了要有数据和技术,大数据思维也是必要因素之一。例如,用户在使用淘宝搜索或浏览某样商品后,推荐主页会出现许多同类型或相关产品。那么这种现象产生的原因是什么?推荐商品的依据是什么?为什么在推荐同类商品的基础上还推荐了相关产品?本任务主要介绍大数据的三大思维方式:全样思维、效率思维和相关思维。

【相关知识】

1. 全样思维

由于过去在数据采集、存储以及处理方面的能力有限,人们通常采用抽样分析的方法进行数据分析。通过在全集数据中抽取部分样本数据进行分析,从而推断出全集数据的总体特征。通常,样本数据的规模是远小于全集数据的,在可承受代价范围内实现数据分析,大大节省了时间和成本,抽样分析也是质量检测、社会调查普遍采用的一种研究方式。例如,工厂要检测一批100个零件的质量合格率,抽取10个为样本做质量检测,检测结果为1件不合格、9件合格,那么可以推断这批100个零件的合格率为90%。这样做不需要对100个零件逐一检查,就能够得到全部零件的合格率。

然而,抽样分析也有着相应的缺点,即具有不稳定性。例如,某次抽样恰好抽到了全部不合格的样本,或恰好抽到全部合格的样本,这样就给分析结果带来很大的不稳定性。而在大数据时代,分布式文件系统和分布式数据库技术为我们提供了近乎无上限的存储空间和分布式并行编程框架(MapReduce)处理海量数据的能力,从而可以实现全集分析,并迅速给出分析结果,从而使得科学分析的结果更加稳定。

2. 效率思维

在使用抽样分析方法时,针对所抽取的样本分析必须要做到十分精确,这样才能确保分析结果应用到全集数据时的准确性。因为,抽样分析中即便是微小的误差,应用到全集数据时,都可能就会变成一个很大的误差,分析结果将“差之毫厘,谬以千里”。因此,为了使分析结果更加准确,我们在抽样分析时必须追求精确,也就意味着同效率相比,精确性才是首要的。而在大数据时代,我们采用的分析方法是全样分析而非抽样分析,解决了误差被放大的问题,效率便成为重中之重。

举例说明效率思维在电商中的重要性。当用户访问电商平台时,平台的后端会记录用户每一次点击,再由大数据分析平台对点击流数据进行处理,而后为用户推荐可能感兴趣的物品。例如,用户在浏览运动鞋时,平台会立即推荐各种款式、各种品牌的运动鞋,也会推荐

运动装、运动包等,精确性在可接受范围内就可以,效率才是首要目标。

3. 相关思维

之前分析数据的目的更加侧重于因果关系,例如,国家实施垃圾分类的政策以及分析微博等应用软件的数据后得出,大家对垃圾桶的讨论度很高,从而建议超市和购物平台上架各种类型的垃圾桶,这反映了一种因果关系。然而在大数据时代,人们更加追求相关性,因果关系不再那么重要。例如,当用户在淘宝购买了薯片,淘宝会推荐可乐,这两件商品没有必然的联系,也不存在因果关系,而是在大数据的分析之下得知薯片和可乐存在相关性。

任务三 大数据的处理流程

【任务描述】

大数据像是未经提炼的金矿石,人们需要对其进行一系列的加工处理,才能体现出它的价值。大数据的处理流程通常是数据采集与预处理、数据存储与管理、数据处理与分析三大步骤。人们通过获取传感器数据、互联网数据等数据源进行数据采集。首先,要对数据进行预处理,包括数据清洗、数据转换和数据脱敏,这项操作能够筛选出有价值的信息,使最终分析结果的可靠性更高。其次,对数据进行有效的存储与管理,才能够充分发挥数据的作用。最后,对数据进行处理和分析,最终得到对人们生产和生活有价值的结果。

【相关知识】

1. 数据采集

企业的财务大数据主要有两种,即企业内部大数据和企业外部大数据。

(1) 企业内部大数据主要来源为 ERP 系统或会计信息系统中的财务数据及业务数据。例如,用友 U8、金蝶 K3、SAP 等系统中的数据,存储在 Access、SQL Server、Oracle 等数据库中的数据。将这些数据从信息系统中导出后即可进行数据分析。

(2) 企业外部大数据主要包括政策法规文件、客户数据、供应商数据、国家统计数据以及行业数据等。

2. 数据清洗

数据清洗是发现并纠正数据文件中可识别错误的最后一道程序,包括一致性检查、处理无效值和缺失值等,相当于重新审查和校验数据的过程,目的在于删除重复信息、纠正存在的错误,并提供数据一致性。

(1) 一致性检查。一致性检查是根据每个变量的合理取值范围和相互关系,检查数据是否合乎要求,发现超出正常范围、逻辑上不合理或相互矛盾的数据。例如,用 1-7 级量表测量的变量出现了零值,体重出现了负数,这些都应视为超出正常值域范围。SPSS、SAS 和 Excel 等计算机软件都能够根据定义的取值范围,自动识别每个超出范围的变量值。逻辑上具有不一致性的答案可能以多种形式出现:例如,许多调查对象说自己开车上班,又报告没

有汽车；一些调查对象报告自己是某品牌的重度购买者和使用者，但同时又在熟悉程度量表上给了很低的分值。我们在发现不一致时，要列出问卷序号、记录序号、变量名称、错误类别等，以便于进一步核对和纠正。

(2) 处理无效值和缺失值。调查、编码和录入误差，数据中可能存在一些无效值和缺失值，需要给予适当的处理。常用的处理方法有估算、整例删除、变量删除和成对删除。

① 估算。最简单的办法就是用某个变量的样本均值、中位数或众数代替无效值和缺失值。这种办法简单，但没有充分考虑数据中已有的信息，误差可能较大。另一种办法就是根据调查对象对其他问题的答案，通过变量之间的相关分析或逻辑推论进行估计。例如，某一产品的拥有情况可能与家庭收入有关，可以根据调查对象的家庭收入推算拥有这一产品的可能性。

② 整例删除。整例删除指剔除含有缺失值的样本。很多问卷都可能存在缺失值，这种做法可能导致有效样本量大大减少，无法充分利用已经收集到的数据。因此，整例删除只适合关键变量缺失或者含有无效值和缺失值的样本比重很小的情况。

③ 变量删除。如果某一变量的无效值和缺失值很多，而且该变量对于所研究的问题不是特别重要，则可以考虑将该变量删除。这种做法减少了供分析用的变量数目，但没有改变样本量。

④ 成对删除。其指用一个特殊码(通常是 9、99、999 等)代表无效值和缺失值，同时保留数据集中的全部变量和样本。但是，在具体计算时只采用有完整答案的样本，因而不同的分析涉及的变量不同，其有效样本量也会有所不同。这是一种保守的处理方法，最大限度地保留了数据集中的可用信息。

采用不同的处理方法可能对分析结果产生影响，尤其是当缺失值的出现并非随机且变量之间明显相关时。因此，我们在调查中应当尽量避免出现无效值和缺失值，以保证数据的完整性。

3. 数据转换

数据转换是将数据从一种表示形式变为另一种表示形式的过程，通常有对数转换、平方根转换、平方根反正弦转换、平方转换、倒数变换等转换方法。

4. 数据脱敏

在某些特定情况下，对敏感数据还需要做进一步处理，即数据脱敏，来保护用户或企业的隐私等。数据脱敏是指对某些敏感信息按照脱敏规则进行数据的变形，实现敏感隐私数据的可靠保护。当涉及客户安全数据或者一些商业性敏感数据的情况时，在不违反系统规则的前提下，对真实数据进行改造并提供测试使用，如身份证号、手机号、银行卡号、客户号等个人信息都需要进行数据脱敏。数据脱敏主要有以下几种方法：

(1) 数据替换。用设置的固定虚构值(如“*”)替换真值。例如，将手机号码替换为 000 * * * * 0000。

(2) 无效化。通过对字段数据值进行截断、加密、隐藏等方式让敏感数据脱敏，使其不再具有利用价值。例如，将地址替换为“* * * * *”。

(3) 随机化。采用随机数据代替真值，保持替换值的随机性以模拟样本的真实性。例如，用随机生成的“猪猪饲养员”代替真实姓名“孙小桐”。

(4) 偏移和取整。这种方式通过随机移位改变数字数据，偏移和取整在保持数据的安

全性的同时,也保证了范围的大致真实性,比之前几种方案更接近真实数据,在大数据分析场景中意义比较大。例如,把日期“2021-12-21 15:46:36”变为“2021-12-21 15:00:00”。

(5) 掩码屏蔽。针对账户类数据的部分信息进行脱敏。例如,将身份证号码替换后,显示为“000000 * * * * * * * 0000”。

(6) 灵活编码。在需要特殊脱敏规则时,可执行灵活编码。例如,用固定字母和固定位数的数字代替合同编号真值。

以孙小桐的工资条为例,展示数据脱敏前后对比,如图 1-1 所示。

姓名	电话	身份证号	家庭住址	日期	工资
孙小桐	0000000000 00	0000000000 0000000000	朝阳区前进 大街	2021-12-21 15:46:36	10000
数据		↓		脱敏	
姓名	电话	身份证号	家庭住址	日期	工资
猪猪饲养员	000****000 0	000000***** ***0000	*****	2021-12-21 15:00:00	*

图 1-1 数据脱敏实例

任务四 常见的大数据分析模型

【任务描述】

数据分析要从业务的角度分析其目标,并对现有的数据进行探查,发现其中的规律,大胆假设并进行验证,依据各模型算法的特点选择合适的模型进行测试验证,分析并对比各模型的结果,最终选择合适的模型进行应用。

【相关知识】

1. 4P 模型

4P 模型又称 4P 营销理论(the marketing theory of 4Ps)。它产生于 20 世纪 60 年代的美国,随着营销组合理论的提出而出现,如图 1-2 所示。1953 年,尼尔·博登(Neil Borden)在美国市场营销学会的就职演说中首先提出了“市场营销组合”(marketing mix)这一术语,其意指市场需求或多或少地在某种程度上受到营销变量或营销要素的影响。1960 年,美国密歇根州立大学的杰罗姆·麦卡锡(Jerome McCarthy)教授在其《基础营销》一书中将这些要素概括为 4 类,即产品(product)、价格(price)、渠道(place)、促销(promotion)。1967 年,菲利普·科特勒(Philip Kotler)在其所著的《营销管理:分析、规划与控制》一书中进一步确认了以 4P 为核心的营销组合方法。

(1) 产品:注重开发的功能,要求产品有独特的卖点,把产品的功能诉求放在第一位。

(2) 价格:根据不同的市场定位制定不同的价格策略,产品的定价依据是企业的品牌战略,注重品牌的含金量。



微课
大数据常见的
分析模型

(3) 渠道:企业并不直接面对消费者,而是注重经销商的培育和销售网络的建立,企业与消费者是通过分销商进行联系的。

(4) 促销:企业通过不同的促销方式来刺激消费,如以“满减”“买赠”等方式吸引消费者,增加销量。

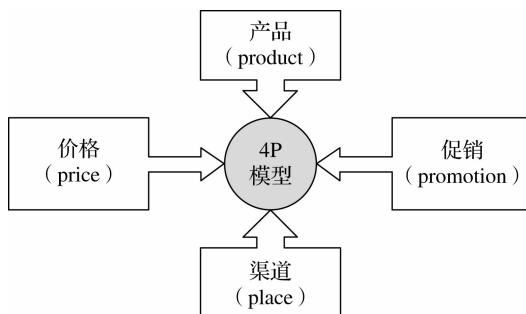


图 1-2 4P 模型

2. RFM 模型

除 4P 模型外,RFM 模型也是常见的大数据分析模型。

RFM 模型通过一个客户的最近一次消费 (recency)、消费频率 (frequency) 及消费金额 (monetary) 三项指标来描述该客户的价值状况,如图 1-3 所示。该模型是衡量客户价值和客户创造利益能力的重要工具和手段。在众多的客户关系管理 (customer relationship management, CRM) 的分析模式中,RFM 模型被广泛提到。

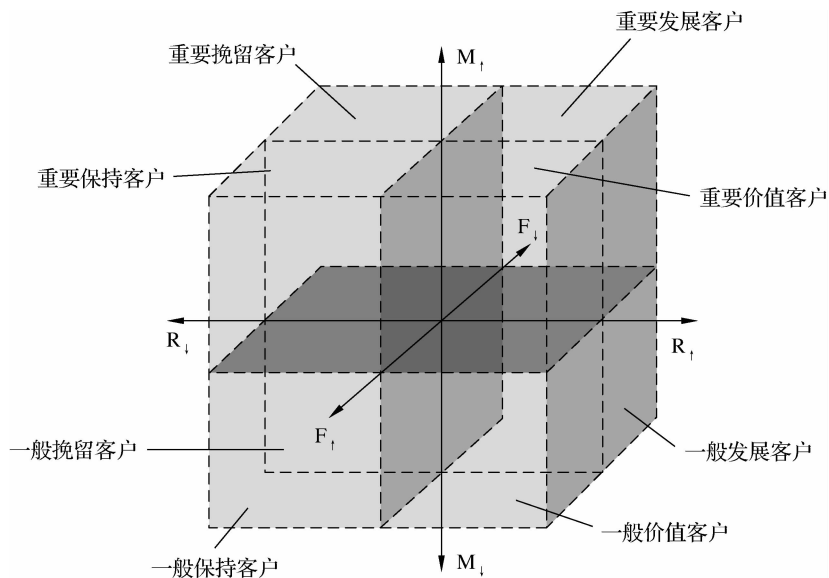


图 1-3 RFM 模型

(1) 最近一次消费。理论上,上一次消费时间越近的顾客是价值越高的顾客,对提供即时的商品或是服务也越可能会有反应。营销人员若想业绩有所增长,只能靠获取竞争对手的市场占有率,而如果要密切地注意消费者的购买行为,那么最近的一次消费就是营销人员第一个要利用的工具。

(2) 消费频率。消费频率是顾客在限定期间内所购买的次数。可以说最常购买的顾客也是满意度最高的顾客。越常购买的消费者,品牌忠诚度也就越高。

(3) 消费金额。消费金额是所有数据库报告的支柱,一个客户的消费金额值越大表示该客户的消费能力越强。

RFM 模型较为动态地显示了一个客户的全部轮廓,这对个性化的沟通和服务提供了依据,也能够较为精确地判断该客户的长期价值。企业可以通过改善三项指标的状况,从而为更多的营销决策提供支持。

任务五 大数据的运用

【任务描述】

大数据相当于一种取之不尽,用之不竭的资源,有同黄金和石油一般的价值,又可无限再生。同样一组数据,用不同的分析方法、从不同的思维角度分析,能得到截然不同的结果,或应用到不同的领域。因此,我们能够直观看到的价值仅仅是沧海一粟,随着不断地深度挖掘和分析探索,大数据可以创造出巨大的价值。

【相关知识】

大数据与我们的生活息息相关,被应用在零售、金融、生物信息、医疗、影视、餐饮、电信、能源、互联网、物流、安全等许多领域,以下重点对在零售与金融领域的应用加以介绍。

1. 大数据在零售领域的应用

谈到大数据在零售行业的应用,不得不提到一个经典案例“啤酒与尿布”。沃尔玛公司拥有世界上最大的数据仓库系统之一,为了能够准确了解顾客在其门店的购买习惯,沃尔玛对其顾客的购物行为进行了购物篮关联规则分析,从而知道顾客经常一起购买的商品有哪些。沃尔玛庞大的数据仓库里集合了其所有门店的详细原始交易数据。在这些原始交易数据的基础上,沃尔玛利用数据挖掘工具对这些数据进行分析 and 挖掘,得到的结果是,跟尿不湿一起购买最多的商品是啤酒。

既然尿不湿与啤酒一起被购买的机会很多,那么沃尔玛就在他们所有的门店里将尿不湿与啤酒并排摆放在一起,结果尿不湿与啤酒的销售量双双增长。按常规思维,尿不湿与啤酒毫不相干,若不是借助数据挖掘技术对大量交易数据进行挖掘分析,沃尔玛是不可能发现数据内这一有价值的信息的。

2. 大数据在金融领域的应用

保险行业主要通过保险代理人与保险客户进行联系,对客户的基本信息和需求掌握很少,因此极端依赖外部保险代理人和渠道(银行)。在竞争不激烈的情况下,这种联系客户的

方式是可行的。但是随着互联网保险的兴起,用户会被分流到互联网渠道,特别是年轻人,他们会更喜欢通过互联网渠道来满足自己的需求。未来线上客户将成为保险公司客户的重要来源。

保险行业的产品是一个长周期性产品,保险客户再次购买保险产品的转化率很高,所以经营好老客户是保险公司的一项重要任务。保险公司内部的交易系统不多,交易方式比较简单,数据主要集中在产品系统和交易系统中。保险公司的主要数据有人口属性信息、信用信息、产品销售信息和客户家人信息等,但是缺少客户兴趣爱好、消费特征、社交等信息。

保险行业的数据业务场景是围绕保险产品和保险客户进行的,典型的数据应用有利用用户行为数据来制定车险价格、利用客户外部行为数据来了解客户需求、向目标用户推荐产品等。

例如,依据个人属性和外部养车 App 的活跃情况,为保险公司找到车险客户;依据个人属性和移动设备位置信息,为保险企业找到商旅人群,推销意外险和保障险等;依据家人数据和人生阶段信息,为用户推荐理财保险、寿险、保障保险、养老险和教育险等;依据自身数据和外部数据,为高端人士提供财产险和寿险等;利用外部数据,提升保险产品的精算水平,提高利润水平和投资收益。

保险公司也需要同外部渠道进行合作,以开发出适合不同业务场景的保险产品,如航班延误险、旅游天气险、手机被盗险等新的险种。保险公司不仅仅是靠这些新险种盈利,还可以找到潜在客户,为客户提供其他保险产品。另外,保险公司应该借助于移动互联网联系客户,利用数据分析来了解客户,降低对外部渠道的依赖,降低保险营销费用,提高直销渠道投入和直销销售比。



数据与人生

大数据最根本的价值在于为人类提供了认识复杂系统的新思维和新手段。大数据作为金融高科技手段,使金融业在业务流程、业务拓展和客户服务等方面得到全面的智慧提升,实现金融产品、风控、获客、服务的智慧化。我国近几年流行的移动支付迅速发展,各方面发展均处于国际领先水平,同步发展的还有智慧理财、智慧银行、智慧证券、智能投顾、智慧保险等。大数据对人类的贡献浸润在生活的方方面面,也可以说人类生活早已离不开大数据。同时,数据本身是没有属性的,如何使用或用在何处取决于人本身。当然,创新的同时不能忽略法律法规和人伦道德。我们要学习大数据,深入大数据,做有能力、有价值、有道德的新时代人才。

 职业知识测试

1. 请阐述大数据的 4V 特性。
2. 请阐述大数据的三大思维方式。
3. 请阐述大数据预处理有哪些步骤及分别是什么。
4. 请列举两个常见的大数据分析模型并简述其原理或作用。