

东华大学

硕士学位论文

大数据背景下 A 淘宝小微电商企业精准营销策略研究

专业名称：工商管理（MBA）

作者姓名：朱梦圆

指导教师：陈梅梅

学校代码：10255

学 号：2131934

中图法分类号：C93

大数据背景下A淘宝小微电商企业精准营销策略研究
**STRATEGIC RESEARCH OF ACCURATE
MARKETING OF A TAOBAO SMALL
E-COMMERCE UNDER THE BACKGROUND OF
BIG DATA**

学科专业：工商管理（MBA）

作者姓名：朱梦圆

指导教师：陈梅梅

答辩日期：2017 年 5 月 18 日

东华大学

工程管理硕士学位论文答辩委员会成员名单

姓名	职称	职务	工作单位	备注
徐波	教授	答辩委员会主席	上海对外经贸大学	
沈蕾	教授	答辩委员会委员	东华大学	
陈亚荣	副教授	答辩委员会委员	东华大学	
李莫愁	讲师	答辩委员会秘书	东华大学	

东华大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律结果由本人承担。

学位论文作者签名：

日期：2017 年 3 月 16 日

大数据背景下 A 淘宝小微电商企业精准营销策略研究

摘 要

在现如今的大数据时代，对大数据的掌握程度已经成为了新经济价值的来源，数据已经成为了能创造经济利益的一种新的商业资本。最近几年，随着电子商务的迅猛发展，电子商务数据的数据挖掘及其在精准营销上的运用显得越来越重要。通过对大数据资源的挖掘与开发利用，得到数据间有用的潜在信息，并将其运用于营销中去，进而降低广告成本，提高利润，已经是业界的热门话题。

本文介绍从小微企业运营的视角出发，利用对大数据进行数据挖掘处理后的结果，对企业进行精准营销提出建设性的意见。通过对销售数据的分析与计算，来探索顾客的购买行为模式，对电子商务中的客户购买行为进行预测，使得商家可以精准定位目标客户及其偏好，进而有针对性地进行商品的选品、上架推广与销售，最终达到提高销售利润的目标。本文着重从选品和上架渠道的选择这两个角度进行对策研究，主要运用分析与统计的方法对产品的特性进行分类总结及建模分析，先运用网络数据采集器对淘宝平台的大数据进行销售数据的采集，将采集到的销售数据根据各销售特点进行归纳分析整理，得出消费者的购物行为模式，初步制定选品策略；再利用离散选择模型的 MNL 模型（Multinomial Logit Model 模型）来构建因子选择模型并进行实证分析，从而初步预测顾客在产品选择中的服务承诺的属性偏好，进而与真实数据对比得出预测正确率，得出商品上架过程中推广方式的属性选择。根据以上两方面结论，

最终追踪出企业的目标客户群及应对方式。

本研究的创新点在于：在普遍关注用户信息的研究现状下，本文从中小企业运营的商户角度出发，从分析店铺运营中面临的实际问题着眼，研究对于小微电商而言有价值的运营活动。以淘宝店家为例，通过对淘宝平台的一整个月香水销售记录数据的分组、整理与对比来进行实证分析，通过对淘宝以及天猫的平台分析，地域分析，卖家分析，买家分析，进行精准定位，得出目标客户群的范围，再通过选择合适的工具来构建顾客选择模型，重点关注哪些已获知的淘宝卖家属性会对消费者的购买行为产生影响，以量化的方式对客户购买数据进行数据挖掘分析，得出相关的影响因素，对客户购买行为进行预测，最终达到企业在后期的购买服务与推广中有方向的进行产品的精准营销的目的。

另外，国内大多数相关研究都是通过对调查问卷数据定量建模来对顾客的购买行为得出定性的结论，而本文则基于电商的实际交易记录，做目标市场的销售分析和消费者需求偏好的建模研究，得出定量的结论，通过 MNL 模型得出了顾客购买商品的实际概率，并达到了较高的准确率，而非泛泛的定性分析，既可以对电商的精准营销提供依据，又因为模型更贴近实际，使得该研究有很高的实际操作的可能。

关键词：电子商务，精准营销，小微电商企业，大数据，数据挖掘，离散选择模型

**STRATEGIC RESEARCH OF ACCURATE MARKETING OF A TAOBAO
SMALL E-COMMERCE UNDER THE BACKGROUND OF BIG DATA**

ABSTRACT

In the current era of big data, the mastery degree of big data has become the source of new economic value. Data turns to be a new commercial capital which will create economic interest. As e-commerce is rapidly developed in recent years, the data exploration and the accurate application in marketing seem to be more and more important. It has become the hot topic to get the useful potential information and apply into marketing by exploiting and applying the big data resource, so as to lower down the ad. cost and improve profit.

This essay has provided constructive recommendation to the accurate marketing of enterprises by referring to the result of data exploration in the aspect of operation of small businesses. It has done some research on the mode of purchasing behavior and predicted the purchasing behavior in e-commerce by summarizing how to apply the data exploration technology and referring to the sales data, so as to guarantee that merchants are able to accurately aim at the target clients and their hobbies and then pointedly select merchandise, place these merchandise on the shelves, do promotion and selling, and finally realizing the target of improving selling profit. This article has emphasized the merchandise selection and placement on the shelves and arranged classification summarization and modeling analysis of product features by statistic analysis. It has collected the selling data and concluded the strategic of product selection on the basis of selling characteristic first. Then it has set up the factor selection mode by MNL mode and arranged empirical analysis to conclude the shopping behavior mode and then preliminarily predicted the product preference of clients. Moreover, it has concluded the accuracy by comparison with the real data and finally speculated the promotion selection of merchandise placement on the shelves and got the targeted client group and the copying style.

Here's the innovation of the research in this essay: it has started from the analysis of the practical problems confronted by merchants during the store operation and done some research on the activities which are valuable for small e-commerce from the aspect of middle and small-sized enterprises under the condition of paying attention to the subscriber information

ABSTRACT

generally. Let's take the stores in Taobao for example. They have arranged empirical analysis by grouping, summarizing and comparing the selling data of perfume for the whole month on Taobao platform, accurately aiming at getting the scope of target client group by Taobao& Tmall platform analysis, territory analysis and vendor& buyer analysis. After that, they've created the client selection mode by selecting suitable tool and paying more attention to those received vendor attributes which will make influence on purchasing behavior and exploited& analyzed the purchasing data in the way of quantification to conclude the relative influence factors and predict the purchasing behavior. Finally it'll be available for them to do accurate marketing of products in the targeted direction in the purchasing service and promotion in future.

Most of the domestic researches have got the qualitative conclusion to the purchasing behavior by quantitative modeling of the questionnaire data. This essay has speculated the qualitative conclusion by selling analysis of target market and the modeling research of buyer demand preference on the basis of actual deal record of e-commerce. What's more, it has concluded the actual probability of purchasing behavior by MNL model and got the higher accuracy. It's easy to find out that it's not rough qualitative analysis which will provide evidence for the accurate marketing and guarantee that it's possible to do the actual operation of the research due to the mode getting close to reality.

Supervised by Professor Chen Meimei

Keywords: E-commerce, Precision Marketing, Small Business, Big Data, Data mining, Multinomial Logit Model

目 录

摘 要.....	I
第一章 绪论.....	1
1.1 研究的背景	1
1.2 研究的目的与意义.....	4
1.2.1 研究的目的.....	4
1.2.2 研究的意义.....	5
1.3 研究思路及框架.....	7
1.4 研究方法与技术路线.....	8
1.4.1 研究方法.....	8
1.4.2 技术路线.....	9
第二章 相关理论及综述.....	10
2.1 精准营销的概念及特点.....	10
2.2 数据挖掘的理论基础.....	12
2.2.1 数据挖掘的概念与特点.....	12
2.2.2 电子商务中数据挖掘的过程.....	13
2.2.3 数据挖掘的应用.....	14
2.3 离散选择模型的理论基础.....	15
2.3.1 模型的概念及建立基础.....	15
2.3.2 模型的应用范围.....	16
2.3.3 顾客选择的 MNL 模型	17
2.3.4 MNL 模型参数估计	20
2.3.5 MNL 模型假设检验	21

2.4 大数据精准营销的研究现状综述.....	21
第三章 A 小微企业电商的运营现状	23
3.1 小微电商在运营中所面临的问题.....	23
3.1.1 电商运营概述.....	23
3.1.2 淘宝平台现状.....	24
3.1.3 淘宝平台上的小微电商所面临的问题.....	25
3.2 现有的解决途径及其局限性.....	26
3.3 A 小微电商企业运用基于大数据精准营销的必要性.....	28
3.4 数据的采集与预处理.....	29
第四章 基于销售数据分析的选品策略.....	32
4.1 基于目标市场的销售数据分析.....	32
4.1.1 渠道分析及结论.....	33
4.1.2 地区分析及结论.....	34
4.2 基于消费者的需求偏好分析.....	36
4.2.1 产品属性分析及结论.....	36
4.2.2 品牌分析及结论.....	40
4.3 选品策略的研究结论.....	41
第五章 基于顾客购买行为分析的上架策略.....	42
5.1 数据描述.....	43
5.2 数据实验.....	44
5.3 上架策略的研究结论.....	48
第六章 总结与展望.....	49
6.1 全文结论及创新之处.....	49
6.1.1 全文结论.....	49

6.1.2 创新之处.....	50
6.2 展望.....	50
参考文献	52
致 谢	56

图目录

图 2-1 CRISP-DM 参考模型	13
图 4-1 各平台销售数据对比	33
图 4-2 不同类型店铺销售数据对比	33
图 4-3 淘宝销售数据图	34
图 4-4 区域销售额	35
图 4-5 省份销售额 TOP 10	35
图 4-6 城市销售额 TOP 10	36
图 4-7 产品属性分析对比图	37
图 4-8 产品销售额走势图	37
图 4-9 产品规格及产地分析图	39
图 4-10 产品含量及分类分析图	39
图 4-11 热销品牌 TOP10	40

表目录

表 4-1 各平台数据表格 33

表 4-2 产品规格分析 38

表 5-1 样本变量的描述统计 43

表 5-2 数据概要 45

表 5-3 各自变量对模型整体影响的显著性检验 45

表 5-4 各自变量的估计值和显著性检验 46

表 5-5 模型预测的正确率 47

第一章 绪论

1.1 研究的背景

近年来，网络购物已然逐步成为人们日常消费的重要渠道。随着 2003 年淘宝网以及 2004 年京东商城成立，淘宝与京东逐渐成为我国两个最大的电商平台，随着电子商务网站的快速发展，电商品牌已逐渐固化，进入成本也在渐渐升高，但消费者对某一品牌的忠诚度却远没有到来。在现在的电商格局中，一心想赚快钱，往往盲目地追求新的热门的商业模式，这必然导致浮躁的电商风气。成功无模板，失败有先例，历次双十一的“成功”经验似乎一再向我们展示着价格战是当今互联网零售中挥之不去的主题，但深入到每一个实际运营的电商公司中去，我们很容易看到，最常见的运营模式，如通过一个或几个超低价的爆款产品拉流量，带动平台或店铺的访问量，最终达到带动业绩贡献的目的，这种方式带来的结果，通常是一天的“赔本赚吆喝”的虚假繁荣，第二天即被打回原形。保证品质的商家有苦说不出，玩各种脑筋的商家则让客户有苦说不出。不难看出，凭借融资，通过打价格战、玩烧钱来大肆宣传、倒流量和做公关数据的做法是很容易导致失败的，因为它在盲目消耗资源的同时，未能带来口碑的同步提升。

电商需要重新回到商业中来，商业的核心是市场细分，本质是洞悉人性。在这个资源极度丰富、信息过载的网络时代，对客户而言，最稀缺的并不是资源，而是时间，强暴式介入的推广手段有时候不仅不能达到宣传的效果，甚至招来客户反感，最终可能连产品的目标客户都丧失了。通过市场细分，首先洞察用户的需求，再根据公司的自身优势，制定适合自己产品的传播渠道，这种精准营销的方式已成为业界的共识。所以，电子商务新的利润增长点在于关注消费者需求。

站在消费者的立场上来看，网购的优势主要表现在：一、便宜：网络营销经营成本较低，无需大库存无需店面租金，减少了零售环节，也由于减少了中间环节进而消减了多重赋税的费用；二、方便：时间与空间的拓展，使得消费

者通过轻轻点击鼠标就可以在各大电商网站之间、全国甚至世界范围内挑选商品，可以对产品进行全方位的比价，下单不受时间控制，可以在任何时间购物也是网购的一大便利之处；三、主动选择性：消费者可以主动基于兴趣去搜索需要的产品进行网购，而非受传统广告所迫，降低了购物的机会成本；四、社会化：网络购物一大优势在于素不相识的购物者可以因为购买同一款产品通过评价进行沟通，基于在电商网站上消费者之间的线上互动也是如今买家选择网络购物的一个重要原因。

电子商务网站正在追求以体验为核心，满足消费者各种各样的需求，以此在竞争中脱颖而出。洞察现今社会的消费行为的变化，洞察人们情感与价值观的变化，结合公司的优势，挖掘其能带给用户的功能需求与情感需求，最终完成选品、重新定义产品的命名、页面的设计、素材的制作，直到末端的投放端口。完成这一系列工作后，才能说基本做到了精准营销。推广销售仅仅是最后一个步骤，前面的几个环节才是企业经营的根本之所在。每一步的效率都依赖于前一步是否执行到位。

在这些环节中，企业第一步要做的就是精准选品，通过定位目标客户群挑选恰当的产品，并选择合适的营销工具将商品上架推广。

大部分的淘宝卖家都是贸易商，而非拥有自己的工厂生产自己的产品然后搭建店铺进行销售的生产商。尤其对于中小卖家而言，选品占有非常重要的地位。小微卖家在建店之初，第一件事就是要确定主营的产品线，根据公司优势，整合手上所有的资源，集中精力打造一条主营产品线。

在完成初步确认店铺要销售的产品线的工作后，接下来就是要精准定位目标客户群，满足其深度的需求。涂子沛先生在《大数据》一书中说，以前小数据时代，可以研究个体但不能研究群体，因为只有关于个人的数据。而在大数据时代，人的行为都以数据形式被记录下来，这里面就有很多商机。换句话说，消费者将其行为习惯以数据的方式制成“密码”，我们对数据的研究，其实就是对这种密码的一个翻译过程。大数据时代，是客户的“行为密码”在各平台共享的时代，越快越深的将其破译，越多的商机就将被挖掘出来。数据挖掘就是在做密码破译的工作，有条件的公司都应该尽快尽早的投身到对要销售的产品数据进行数据挖掘的工作中去，以增进对消费者的了解，使得自身企业更精准的匹配到合适的目标客户群。

在企业运营中，要提高效率降低成本，被广为运用的一个方法就是通过优化运营，现代的科学运营优化手段的基石通常在于海量的销售数据。要做到“低成本、高效率”的运营，达到精准决策的目的，企业必需尽可能的推行以事实为依据的决策思想、大量使用数据分析来优化企业的各个运营环节，通过对数据的分析与整合，把业务流程和决策过程中不必要的耗费挤出去，将潜在的价值挖掘出来，从而达到降低成本、制服对手、在市场上立于不败之地。

我们现在处于大互联网时代，对消费者的众多个人数据常常可以通过各种渠道获得，消费者的决策历程也由 AIDMA 模型（引起注意→引起兴趣→唤起欲望→留下记忆→购买行动）发展成为 AISAS 模型（引起注意→引起兴趣→进行搜索→购买行动→与人分享），数据挖掘的方法与以前相比，有了质的飞跃。如刘德寰教授所说，现在数据挖掘的商业本质是：挖掘的过程要复杂再复杂，因为所有的市场研究方法都能用于数据挖掘，而数据挖掘呈现的结果要简单而直接。数据挖掘需要对商业的理解，并不是调用大数据就是数据挖掘。数据挖掘首先需要了解消费者，做到对市场的准确区分，保证有效的市场、产品和定位，公司内部各部门人员，如市场研究人员、数据分析人员、信息技术人员和营销人员间相互合作，才能够通过做好数据挖掘，最终达到开源节流的目的，实现公司利润的进一步提升。

近几年，随着云计算技术和网络通信技术等服务迅猛发展，对于大数据的运用让营销的可度量、可精准正在一步步实现中。电商通过对用户的已经公开的信息进行分析、挖掘、归类，利用大数据从而进行精准营销的行为已不再被归罪于对隐私的侵犯，用户已逐渐接受并享受由此带来的便利。通过对电子商务的销售数据进行数据挖掘来指导公司的营销活动，也显得越来越重要。阿里巴巴、京东等电商巨头都已经在各方面进行了大数据的搜集与运用。但由于大数据的获取难度较高，其分析的复杂性和多样性有较高的门槛，导致国内针对小微企业的电商用户进行大数据精准营销实施和效果的研究比较缺乏。

1.2 研究的目的与意义

1.2.1 研究的目的

大数据的概念自 2008 年被正式提出来以后，在社会各界都得到了广泛的关注。基于云计算技术得以迅速成长起来的大数据技术给各行各业都带来了深远的影响，尤其是以营利为目的的企业，大数据和数据挖掘技术的结合，给企业的营销带来了一个全新思路，企业可以运用大数据思维，通过对数据的采集、分析和有效处理，对用户进行科学的分类，精准的筛选出目标客户群后“投其所好”的进行个性化的营销活动和宣传推广，在恰当的时间，利用恰当的沟通方式，将恰当的产品和活动推荐给恰当的人。不仅如此，基于大数据的精准营销在对用户进行精准细分、降低了企业的营销推广的成本的同时，由于对客户偏好做了深层次的研究，推送的正好是客户需要的，从另一个角度来说，还给用户带来了良好的用户体验。

但是，大数据下的精准营销也有其缺陷所在。

首先，拥有大量的数据未必一定可以直接带来精准营销的结果。

其次，传统的广告模式固然成本高，但市场细分的成本也会随着细分化的深度的提高成几何比率增长，反过来提高了成本收益比。尤其对于小微企业而言，公司资源有限，用于营销的预算更是有限，所以营销部门希望的是通过适度的细分化，合理地配置资源，来达到收益的最大化。

但是不同的产品对应的市场具有不同的特点，自然相应的营销方式也不相同，如何制定判断标准，进而用好这些营销组合，对大多数公司来说是一件困难的事情。尤其是小微企业主，没有足够的资源支持他们进行大量试错，资源的浪费往往意味着死亡。

对此，国内外学者展开了深入研究，提出各种模型来探索价格和各种营销因素对顾客品牌选择的影响，努力在其中找到一个平衡点。而顾客对于品牌的选择是一个离散的选择过程^[33]，因此离散选择模型(Multinomial Logit Model)就得到了广泛地应用。

A 小微企业电商已有近 2 年的淘宝网店运营基础，在母婴用品等方面已有很好的经验。鉴于公司新开发了“香水”品类的购货渠道，公司希望开辟一片

新的市场,而 A 公司没有既没有历史销售数据可以借鉴,又没有理论可以参考。

本文正是从 A 公司运营的角度为切入点,运用数据采集器收集淘宝销售数据,在对大量低密度的销售数据进行初步的整理后,先通过对销售数据的归类与整理,在目标市场与消费者需求偏好两方面进行分析,得出选品策略,再通过对整体数据建立 MNL 模型,来初步预测顾客对四大品牌香水的购买行为,进而与真实数据对比得出正确率。最后再通过综合以上分析,得出选品与上架的策略建议,初步对消费群的购买方向作出预测,制定出可落地的小微电商精准营销实施方案。

本研究需要解决两个问题:一是当 A 公司有香水方面的资源时,如何通过平台销售数据的收集及整理,找到利基市场,并据此做出选品决策;二是当顾客已经决定购买某类香水的情况下,公司如何利用对平台销售数据的分析,对推广渠道做出选择。

1.2.2 研究的意义

传统营销强调产品创意、营销策略以及对受众的覆盖程度。而在如今的大数据背景下,营销成了一个技术活,与以往的盲目推送行为相比,精准营销更强调顾客的体验,具体做法是以海量数据为前提,通过数据挖掘的方式寻找特定受众,并运用技术手段对经过筛选后的客户群进行个性化的广告传播。因此,在海量低密度的数据背景下,怎样采集、分析和运用数据是很多企业面临的难题。

现阶段,在应对这个问题中,企业通常采用两种方法:一是自己解决,即自行建立相关的数据采集与分析系统;二是借助外力,即从各类信息供应商处获取数据和技术,借助专家来解决这一问题。

大数据能给企业带来巨大的价值和良好的用户体验这一点是毋庸置疑的,但是是不是大数据就一定能够带来精准营销呢?

首先,数据采集仍然相当困难。这一问题的成因有客观和主观的两方面原因,一方面部分小城镇和农村及偏远地区的网络普及度还不够,硬件上达不到数据的多平台共享的基本要求;另一方面,鉴于对隐私的保护和商业竞争力的追求,平台与平台之间、个人与个人之间的数据开放度也因此受限。

其次，由于数据量离散、庞大且低密度的特点，以及数据本身的个性化与实时性，众多企业在对大数据进行有效的分析处理的问题上显得应对乏力。2011 年，IBM 针对数据运营情况，对全球一百个国家和地区的三十个行业的三千名高管做了一次调查，得到的反馈是六成的受访者认为尚无法做到有效的利用所有数据。IBM 在次年对六十四个国家及地区在十九个行业从事营销工作的一千七百名首席营销官继续做了一次深层次调查后，调查结果竟显示七成的首席营销官表示对于大数据的挑战，他们的企业尚未做好充分的应对准备。

对消费者画像的思路受到技术条件的制约，让企业在营销渠道中无法精准描绘目标客户，精准营销也因此变得力不从心。

电子商务的运营者通过销售数据进行数据挖掘来指导公司的营销活动，在资金实力不足的中小企业显得越来越重要。公司用于营销的预算是有限的，所以营销部门希望通过合理地配置资源来达到收益的最大化。但是不同的产品对应的市场具有不同的特点，自然相应的营销方式也不相同，如何用好这些营销组合，对大多数公司来说是一件困难的事情。

在普遍关注用户信息的研究现状下，本文从中小企业经营的商户角度出发，以淘宝店家为例，重点关注哪些已获知的淘宝卖家属性会对消费者的购买行为产生影响，再通过对淘宝以及天猫的平台分析，地域分析，卖家分析，买家分析，进一步精准定位，初步得出目标客户群的范围，以便在后期的购买服务与推广中可以做到广告的精准化投放和商品的个性化推荐，据此进行产品的精准营销。

国内外学者有着众多模型来探索价格和各种营销因素对顾客品牌选择的影响，模型建构复杂、专业化要求高的设计，对于小微企业而言并不能很好的落地施行。顾客的品牌选择是一个离散的选择过程，因此本文选取了模型建构简单且在数据分析中得到广泛应用的离散选择模型。本文将通过基于离散选择模型的 MNL 模型（Multinomial Logit Model）来定量构建预测模型并得出顾客的购买概率，从而使得小微电子商务的商家可以更有针对性地进行选品与销售。

另外，国内大多数相关研究都是通过对调查问卷数据定量建模来对顾客的购买行为得出定性的结论，而本文是通过电商的实际交易记录来做定量的建模研究，得出定量的结论，使模型更贴近实际。通过 Multinomial Logit Model 得

出了顾客购买商品的实际概率，并达到了较高的准确率，而非泛泛的定性分析，既可以对电商的精准营销提供依据，又有很高的实际操作的可能。

1.3 研究思路及框架

新卖家或初入某新领域的老卖家，在根据自身优势选定了要卖的产品类型后，第一步要做的事就是选品，如何根据用户的行为预测分析进行选品，以及选择合适的工具进行引流，这是淘宝运营的重点和难点。

小微电商在没有获得历史成交数据以前，对于产品上架前的选品对策以及如何有针对性的选择引流工具这两个问题尚无系统性的解决方案，常常依靠的是运营者的经验和 CEO 的直觉。

A 小微企业电商已有近 3 年的淘宝网店运营基础，在母婴用品等方面已有很好的经验。鉴于公司新开发了“香水”品类的购货渠道，公司希望开辟一片新的市场，而 A 公司既没有历史销售数据可以借鉴，又没有理论可以参考，作为一个新卖家，最稳妥的做法，是对该品类在平台上的销售情况做一个充分的了解，选择平台卖的最好的产品作为进货时的首选产品。

本研究正是基于这一现实问题，提出了如何通过对网络可获得的公开信息进行采集、归类、挖掘分析，根据用户的行为预测分析结果进行选品并选择合适的工具进行引流，从选品策略和产品上架的营销渠道的选择这两方面，尝试运用精准营销的思维方式，寻找可落地的解决方案。

本研究从大数据精准营销的定义出发，以淘宝卖家的电子商务环境为研究情境，从运营的角度，依次分析店铺运营中涉及到的问题，分析具有可操作性的精准营销方案，进而以香水销售为例，通过基于离散选择模型的 **Multinomial Logit Model** 来定量构建预测模型并得出顾客的购买概率，最终得出对优化运营有价值的结论，为指导小微商家有针对性地进行商品的选品与销售提出了一个新的解决思路和方法。

作为一个专题研究类论文，本文的研究框架具体如下：

第一章是绪论，主要针对本文选题的背景，本研究的目的与意义，以及论文的思路和整体结构安排做了简略的介绍；

第二章是本研究的理论综述环节，精述了大数据、精准营销、数据挖掘及离散选择模型的相关理论基础以及大数据精准营销的研究现状综述。

第三章具体描述了本研究的选题原因及探索过程。论文从小微电商的运营现状出发，论述了其面临的问题、现有解决途径的局限性，以及本研究的必要性与可行性，并对本研究例证所涉及的数据处理做了简要的描述。

第四章和第五章详细阐述了通过对经过预处理的数据进行各种分析与处理，最终达到对小微电商在上架前的精准营销的能起到的指导意义。第四章通过对销售数据的分析，确定了选品策略，第五章通过构建模型进行测算，对营销方案作出选择。

第六章总结了整个专题对于优化小微电商运营的作用，分析了研究的意义及研究局限，并对未来的研究方向提出了展望。

1.4 研究方法与技术路线

1.4.1 研究方法

基于客户是理性人的假设，即其做出的选择符合效用最大化原则为基础进行分析。本文要研究的问题分两部分：一是当公司有香水方面的资源时，如何通过分析平台销售数据找到利基市场并据此做出选品决策；二是当顾客已经决定购买某类香水的情况下，公司如何利用平台销售数据对推广渠道做出选择。鉴于本研究的背景是没有店铺历史销售数据的小微电商的选品问题，我们先忽略消费者基本信息的因素。我将以淘宝网某品类的一个完整自然月的数据为基础，从以下两个方面，通过相关筛选分析，结合定性和定量分析的方法，对客户的行为进行预测：

在选品策略方面，主要通过销售渠道的数据选择商品上架平台，通过区域分析得出在国内哪些地区的销售额比较可观，继而选择产品推送区域。

从消费者偏好方面，对于不同的平台结构通过数据进行分析，并对何种品类的哪种品牌的产品卖的好以及平均价格的定位作出一定的分析。具体到价格区间，发现目标客户群的潜在购买力以及消费区间。通过对于产品属性的分析，得出产品哪些特性会特别得到消费者的青睐。

1.4.2 技术路线

本研究中，我将使用免费的网页数据采集器，随机抓取淘宝平台一整月的“香水”类目的销售记录作为样本，对网页数据采集器的要求是可以搜索到的成交数据中需要包含尽量多的信息。第二步是对数据进行分类筛选：哪些是网站用身份识别的信息，不具备数据勘探与挖掘的价值；哪些数据几乎重复，可剔除一半；哪些数据需同时考虑，得出有用的数据。最终，研究对象由哪些因子构成。再用定性研究的方法对目标市场和消费者偏好作出筛选，用定量研究的方法对其购买行为进行预测，主要是通过将70%的样本作为训练样本，通过筛选，得出各变量对模型整体影响的显著性检验，再通过得出的结果建立模型，最后用此模型对剩下的30%的测试样本进行预测，验证模型预测的正确率，最终综合以上两方面分析，作出选品与上架的策略建议。

第二章 相关理论及综述

2.1 精准营销的概念及特点

大数据精准营销是在大数据环境下的一种营销方式，它基于多平台的海量数据，依托数据挖掘技术，对顾客进行精准定位，在此基础上与客户建立一对一的沟通交流，为客户提供个性化的服务体验，其核心在于让产品的广告、信息和活动在恰当的时间，通过恰当的渠道，通过恰当的方式，精准传递给恰当的人。

自 F.W. 泰勒提出科学管理以来，人们第一次将管理从经验上升为科学，而大数据精准营销可以说是这个理论的一次重要的拓展和延伸。它以科学管理为方向，依托多维度的数据采集，以及数据分析技术的运用，达到洞察用户需求，进而预测客户行为的目的。运用对大数据挖掘分析的结果，可以实现对客户进行多角度的画像，并据此进行个性化营销。进一步来说，企业可以通过将有了初步画像是消费者进行归类，得出其细分市场进而有针对性的设计营销方案，使得广告和营销活动更加精准有效，进而为公司带来更高的投资回报。

精准营销的定义中有三个关键点：精确定位、可衡量和高回报^[30]。企业可以借助先进的数据库技术进行数据的收集工作，利用统计分析技术对数据进行整理分析工作，让企业的目标客户得到更加准确的细分和定位，进而达到明晰目标客户群的目的。

大数据精准营销主要特点是：

（1）大数据精准营销突破了传统营销只能依靠经验进行定性定位的局限，它利用数据的量化分析技术对市场进行精准定位，从而解决了企业在传统营销方式的进行过程中因为广告的无关推送而导致的巨额浪费的问题。

（2）大数据精准营销运用信息技术手段，通过数据往来，实时满足不同客户的个性化需求，从而建构起企业自身的客户数据资料，逐步培养企业忠诚稳定的顾客群，随着客户群的累积加上口碑效应带来的链式反应，形成粉丝经济，最终达到企业长期稳定高速发展的目的。

(3) 大数据的收集和数据分析技术除了可以对目标客户群的定位与行为分析以外,更可以根据对客户数据的持续跟踪与修正,让企业得以与客户建立长久的个性化沟通,降低了推广的盲目性,让企业的营销行为达到可度量、可调控等精准要求,从而最终让企业可以从网络推广、传统广告沟通的高成本束缚中摆脱出来,不仅可以降低营销成本,还能基于信息的准确送达而获得更好的营销效果,进而企业得以从激烈的竞争中脱颖而出,在市场中获得立足之地。

在实现大数据精准营销时,企业应该有结合不同平台上数据的资源与能力,将用户决策流程上的所有节点整合在一起,这样才能具体一对一的对客户作出全方位的描绘,形成真正个性化的用户视图。在资源与技术允许的前提下,进行数据的多平台化采集,这可以让企业对客户行为的刻画更加全面而精准。

对企业而言,大数据精准营销的价值主要体现在对用户的个性化营销上,而这一目的实现的基础,就是利用数据的分析和预处理,对企业的目标客户进行全面而准确的画像。大数据技术可以让企业对自己的实际用户更加了解。

在云计算和大数据时代,企业的营销理念已不再依靠媒体宣传进行单向的推送,而是逐渐向体验营销的方式进行转化。在传统的营销过程中,企业为了达到宣传自身产品的目的,在资源允许的前提下,更倾向于将广告投放于知名度高、浏览量大的媒体,以期通过高浏览量弥补转化率的不足;在产品确定上,由于企业的精力和资源有限,通常都会选择一款或几款主打产品进行营销活动。而如今,基于大数据技术和用户画像的日益成熟,企业在进行营销时,转向了以客户需求为导向,通过以受众为导向选择广告媒体,而非盲目追求该渠道的知名度和浏览量,因为大数据技术可让企业了解特定媒体的用户群特征,该用户群与本企业的目标受众的匹配度如何,这些用户的喜好如何,对哪些内容比较关注,对页面什么位置的关注度较高,这些信息都可以对广告的精准推送产生影响。同时,企业也可以通过销售数据的收集、分析与整理,挖掘用户对产品功能和价格的偏好,利用大数据技术的预测功能,对客户发生购买决策的可能的时间点和能接受的营销方式做出推断,切实做到以用户为导向,进行产品的选择、价格的定位以及活动的安排,最终做到有针对性的进行营销活动。大数据技术已经做到了,当不同用户关注同一媒体的相同界面时,推送关联文章不同,广告的内容不同,甚至产品和价格都可以不一样,真正实现了对用户的个性化营销。

企业在运用大数据进行精准营销时，还能通过关联推荐，吸引客户增加在店内的停留时间，提高用户在店铺的客单价，从另一方面降低引流成本，提高公司收益。在大数据营销过程中，可以通过对用户进行分群识别和分析后，根据同类客户的历史购买行为，对该用户群的购买行为、购买决策时间点以及购买其他产品的相关度做出预测，达到实现大数据营销的关联销售的目的。

2.2 数据挖掘的理论基础

2.2.1 数据挖掘的概念与特点

大数据的数据量大、类型繁多、价值密度低的特点，让数据挖掘在信息产业界引发起极大的关注。海量且多维度的数据只是第一步，如不将这些庞杂无序的数据通过分类整理做出结论，仅仅是庞大的数据，是没有任何意义的。因此，数据挖掘技术在各领域都得到了广泛的应用。尤其是竞争激烈的商家，更是迫切需要将数据转换为有用的信息，以达到降低成本、提高效率的目的。通过数据挖掘处理后的信息不仅可以广泛地运用于各领域，例如生产控制、商务管理、工业设计以及科学探索，甚至在现代战争中也有出色的表现。那么，什么是数据挖掘呢？

数据挖掘，简称 DM (Data Mining)，它是通过利用半自动化或者是自动化的方法来对数据中隐藏的具体有价值或意义的信息与规则进行挖掘，是从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们事先不知道的、但又是潜在有用的信息和知识的过程，通过挖掘数据仓库中存储的大量数据，从中发现有意义的新的关联模式和趋势。

数据挖掘是一项多学科综合应用的技术，其最吸引人的地方是它能建立预测模型而不是回顾型的模型。利用功能强大的数据挖掘技术，可以使企业把数据转化为有用的信息，为决策提供依据，从而使得企业在市场竞争中能够获得优势地位。在大数据时代，数据挖掘其实是沟通客户与企业之间的桥梁，客户以数据的方式对其购买行为作出表达，企业则根据对数据的梳理构建对客户行为模式的洞察，借此深层次了解到消费者的关注重点，以此为依据去改进产品，更好的满足客户的需求。

2.2.2 电子商务中数据挖掘的过程

数据挖掘是一门应用技术，其交叉产业标准过程(cross-industry standard process for data mining, CRISP-DM)是当今数据挖掘业界通用流行的标准之一。该参考模型通过商业理解、数据理解、数据准备、建立模型、模型评估和模型发布这几个步骤完成对数据的采集、处理及结果程序的过程，如图 2-1 所示

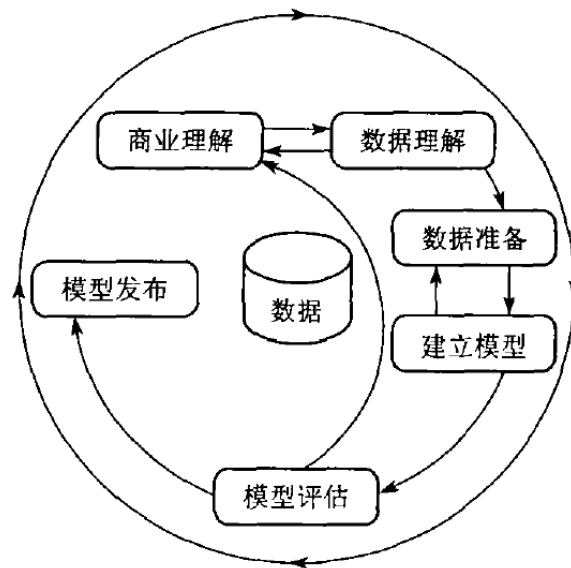


图 2-1 CRISP-DM 参考模型

在实际操作中，数据挖掘工作通常分成以下三个步骤来进行：

1、数据准备

数据准备是指从最初的数据采集阶段开始的一系列活动。目的是对采集到的数据进行质量的检验，包括剔除无用的数据以提高数据挖掘的效率与准确性，经过筛选后对其进行初步的分析处理，发现其潜在关联的可能性，并形成对完成目标有意义的处理假设。数据准备阶段可以分为数据集成、数据选择以及数据预处理这三个子步骤。其中，数据集成是在多个数据库中挑选出所需要的数据进行合并处理；数据选择是对提取的数据集合进行适合的选择、整理、转换进而归类；数据预处理则是对数据进行筛选、清理、集成、变换以及过滤的工作。这些步骤的实施保证了在接下来的挖掘操做过程中，既提高了数据挖掘的质量，又更好地克服了数据挖掘工具本身所具有的局限性，从而为数据挖掘的

实施提供了必要的前提。

2.数据挖掘

数据挖掘的操做是指根据挖掘目标,通过适当的挖掘方法的方式,以期获得有意义的数据。数据挖掘的方法有很多,目前常见的有九种,分别是:基于大量历史数据且具备学习功能的记忆基础推理法;通过选择正确的品项、经由对共同发生矩阵的探讨挖掘出联想规则、运用技术手段进行去噪这三个步骤实现的购物篮分析法;在处理归类于预测问题上有着极强的能力的决策树分析法;通过对生物系统细胞演化模式的学习发展而出的遗传算法;运用群集侦测基础在数据中找出以前未知的相似群体的聚类分析法;以关系为主体,对各节点间的关系进行多维分析的连接分析;通过对一串例子的学习归纳出一个足以区分的样式的神经网络法;应用于解决分类问题的判别分析法,非常适用于因变量为定性、自变量为定量的情况;当判别分析中的群体不符合正态分布假设时,用于预测事件发生几率的罗吉斯回归分析法。

3.结果表述和解释

在经过以上对数据的筛选、建模、对参数的优化,进而构建模型实施后,我们已基本得出了对目标进行决策支持的结论,结果的表述和解释步骤就是将结论组织为阅读模式呈现出来,其具体表现为:对分析处理后的信息进行整理,将其中最有价值的信息梳理出来,以决策支持工具的方式,将其提供给决策者。如果决策者对提交的信息不满意,则应对上述的过程进行重复,直到获得可供支持决策的信息。

2.2.3 数据挖掘的应用

传统的营销中,对客户的研究常常通过问答的方式进行,所获得的答案通常是在提问的情况下被引导出来的,这就意味着在一定程度上,问题的本身对答案就是一种限制。更不提这种方式获取的答案会在时间上滞后,也可能会因为空间上的不匹配,对结果产生一定的偏差。

激烈的竞争让成本收益成为每一个企业关注的重点,而这就决定了细分消费者群体是大势所趋。数据挖掘在近年来忽然如此火爆,这是其中一个重要的原因。在如今这个信息网络高速发展的时代,销售中的边际成本在迅速降低,

长尾理论越来越广泛的得到验证，特色淘宝小店的兴起，代表了满足小众需求的市场亦极具商业价值，通过对相似需求的消费者的整合，细分市场的规模收益被放大了，过去所谓“碎片化”的市场在今天看来忽然变得大有可为起来。通过对客户社交图谱的多维度精准描绘，营销者可以 360 度观察一个消费者的态度、性格、行为，甚至其背后的动机，进而对其未来的行为作出预判。

在电子商务与线下销售中，数据挖掘技术被运用的最为广泛的就是对客户购买行为的分析。数据挖掘在营销领域得到广泛运用，是建立在“消费者未来的消费倾向可以从过往的行为中找到线索”这一假设的基础上的。客户购买行为分析是指通过对海量的客户交易数据的收集，对数据间隐藏的潜在意涵进行挖掘整理，寻找出存在于其中的交易、流转等方面的规律，并据此勾画客户行为轨迹，帮助企业深度了解客户，开发客户价值，提升服务质量，并最终达到提高企业营利的目的。

数据挖掘中的关联分析法，可以通过寻找数据间的关联关系，帮助企业对顾客的购买历史数据进行分析，探寻顾客的购买行为模式，挖掘顾客购买不同商品之间的关系，实现交叉销售。恰当的商品推荐策略可以减少客户对商品盲目查找的时间增强客户的购买体验、增加客户在店内停留的时间、提高每次交易的商品数量和交易金额，也有利于企业提前进行存货安排，以及制定交叉销售的活动策略，使电子商务活动更为高效。

在实际运作中，这一类数据挖掘分析的方法可用于店铺关联设置中。如通过一个爆款产品将客户引入店铺，再关联其他相关产品，引起店内跳转，产生一次甚至多次的店铺内流量，提高访问深度和在店内停留时间，最终目的是提高客单价，在一定程度上也就降低了平均点击花费和单次引流的成本。由于此种方式运作在客户进店以后，不在本论文的讨论范围内，关于此项应用文中不再进行深入的展开。

2.3 离散选择模型的理论基础

2.3.1 模型的概念及建立基础

离散选择模型也叫基于选择的结合分析模型，它是一种非常有效且实用的

市场研究技术。该模型是在实验设计的基础上，通过模拟所要研究产品或服务的市场竞争环境，来测量消费者的购买行为，从而获知消费者是如何在不同的产品或服务属性和价格条件下进行选择的。在本研究中，因为当顾客在选择商品时，经常要面对多个可选择因素，这些因素都有可能作为首选因素直接影响到顾客最终的购买决策，这里就需要离散选择模型来帮助我们对客户的选择行为作出预测。

离散选择模型是一种非线性的多元统计分析的方法，它建立在实验设计的基础上，描述了决策者如何在给出的备选物中做出选择。以经济学中的理性人假设为基础，所有的经济学问题都可以转化为选择的问题，离散选择的理论便是基于这一假设的基础，逐渐发展而来。

离散选择模型是一类模型的总称。在这类模型中，经济学家可获知的唯一现象是某个主体选择了某个对象，如一个学生毕业后是选择继续深造还是就业，在同类产品比如可乐的购买选择中最终购买了可口可乐还是百事可乐等。数据分析人员通过对数据的处理来估计参数的分布，从而得出未来该对象被选择的概率。

离散选择模型主要是建立在随机效用理论的基础之上的，选择实验起源于心理学的研究。历史上最早的关于选择试验的论述出现于 1927 年，Louis L. Thurstone 在 *Psychological Review* 上发表文章，提出了成对选择试验。随机效用理论(Random Utility Theory, 简称 RUT)虽然是由 Thurstone 首先提出来的，但他无法将之扩展到两个以上选项的模型中去。丹尼尔·麦克法登(Daniel L. McFadden)教授在扩展问题上做出了进一步的研究工作。鉴于他在离散选择模型等微观计量经济学领域的重大贡献，麦克法登教授被授予了 2000 年诺贝尔经济学奖。他在 1974 年发表在 *Frontiers in Econometrics* 上的论文给出了关键的理论扩展，之后的若干篇论文陆续证明了该理论在实际的问题中可以得到应用。

2.3.2 模型的应用范围

离散选择模型最初是由生物统计学家在研究流行病、病毒以及发病率时发展起来的。这一技术最初被用于为实验结果建模，实验结果通常是以比值的方式呈现（比如对不同数量的杀虫剂施放于昆虫时是否影响其死亡）。由于经济

学家研究的许多变量是离散的或是以离散形式度量的。一个人要么在职，要么失业；美国总统最终不是民主党就是共和党当选。因此离散选择模型在经济学中常常被运用到对未来的影响预测上。当事件的发生概率永远落在 0 和 1 之间时，我们就将用到 Logit 模型了。

2.3.3 顾客选择的 MNL 模型

早期的经济理论和计量经济学的分析，局限于数值连续的经济变数，类别选择的问题虽然无所不在却没有一个严谨的分析架构。而在类别选择问题上，丹尼尔·麦克法登（Daniel L. McFadden）教授发展出的一套完整的理论和实证方法，正好弥补了这个空隙。麦克法登教授关于类别选择问题思想基础是效用最大化原则，具体来说就是：不论要选择的类别是什么，对做选择的经济个体来说，每一个类别都有或多或少的效用，一个类别的脱颖而出必然是因为该类别能产生出最高的效用。而顾客选择首选因素的依据就是使自己的效用最大。

本文通过消费者对产品作出选择的实际交易数据为基础，来测量消费者的购买行为，从而预测客户未来将如何在不同产品的多种属性条件下进行选择。根据麦克法登教授的研究，消费者的行为根据最大效用理论服从耿贝尔分布，将据此推导出其顾客 n 选择第 j 个产品的概率，具体步骤如下：

根据该假设计算出分布函数，根据有随机性干扰的情况下，选择 i 的条件概率公式，代入分布函数，可求出顾客 n 选择第 j 个产品的概率；然后将在第五章将数据代入模型，再用 SAS 软件对参数进行估计和假设检验，最终对消费者的选择得出预测。

以下是模型构建的详细过程。

本文给出的假设是我们讨论 N 个顾客有 J 种产品可以选择，他也可以选择 not 购买，即：顾客 $n(n=1, 2, \dots, N)$ 有 $J+1$ 种商品可供选择，分别为 $j=0, 1, \dots, J$ 。对于面临 $J+1$ 种选择的第 i 个顾客，假设选择 j 的效用是：

$$U_{nj} = X'_n \beta_j + \varepsilon_{nj}, \quad n = 1, 2, \dots, N; j = 0, 1, \dots, J \quad (2.1)$$

其中， X'_n 是一个向量，表示顾客的内在特征， β_j 也是一个向量，表示第 j 种商品被选择时，对应顾客内在特征参数， ε_{nj} 是独立的随机项，且服从 $\varepsilon_{nj} \sim N(0, 1)$ 。

用 P_{nj} 表示顾客 n 选择商品 j 的概率，那么根据效用最大化原则，可以得出顾客 n 选择商品 j 的概率为：

$$P_{nj} = P(U_{nj} > U_{nk}, k \neq j) \quad (2.2)$$

根据消费者的行为服从极值 I 型分布的假设，Multinomial Logit Model 是通过假设每一个 ε_{nj} 服从极值 I 型分布来获得的，未观测到的效用函数的密度函数是：

$$f(\varepsilon_{nj}) = e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}} \quad (2.3)$$

积分后得到累积分布函数为：

$$F(\varepsilon_{nj}) = \exp(-e^{-\varepsilon_{nj}}) \quad (2.4)$$

下面跟随 McFadden(1974)来推导 Multinomial Logit Model 的选择概率。

由前面随机效用部分可知，顾客 n 选择选项 i 的概率为：

$$P_{ni} = P(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \forall j \neq i) = P(\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj} \forall j \neq i) \quad (2.5)$$

如果 ε_{ni} 给定，这个表达式是每一个 ε_{nj} 在 $\varepsilon_{ni} + V_{ni} - V_{nj}$ 上的累积分布函数，通过(2.4)得可表示为 $\exp(-\exp(\varepsilon_{ni} + V_{ni} - V_{nj}))$ 。既然 ε_{nj} 是独立的，这个累积分布对于所有的 $j \neq i$ 是单个的累积分布相乘的结果，即：在有随机干扰的情况下，顾客 n 选择 i 的条件概率如下：

$$P_{ni} | \varepsilon_{ni} = \prod_{j \neq i} e^{-e^{-(\varepsilon_{ni} + V_{ni} - V_{nj})}} \quad (2.6)$$

将以上代入分布求 P_{ni} ：

$$\begin{aligned}
P_{ni} &= \int (P_{ni}|\epsilon_{ni})P(\epsilon_{ni}) d\epsilon_{ni} \\
&= \int \left[\prod_{j \neq i} \exp(-\exp(\epsilon_{ni} + V_{ni} - V_{nj})) \right] e^{-\epsilon_{ni}} \exp(-\exp(-\epsilon_{ni})) d\epsilon_{ni} \\
&= \int_{\epsilon_{ni}=-\infty}^{+\infty} \left[\prod_j \exp(-\exp(\epsilon_{ni} + V_{ni} - V_{nj})) \right] e^{-\epsilon_{ni}} d\epsilon_{ni} \\
&= \int_{\epsilon_{ni}=-\infty}^{+\infty} \exp(-\sum_j \exp(-(\epsilon_{ni} + V_{ni} - V_{nj}))) e^{-\epsilon_{ni}} d\epsilon_{ni} \\
&= \int_{\epsilon_{ni}=-\infty}^{+\infty} \exp(-e^{-\epsilon_{ni}} \sum_j \exp(-(V_{ni} - V_{nj}))) e^{-\epsilon_{ni}} d\epsilon_{ni}
\end{aligned}$$

令 $t=e^{-\epsilon_{ni}}$ 换元得:

$$\begin{aligned}
P_{ni} &= \int_{t=\infty}^0 \exp(t \sum_j \exp(-(V_{ni} - V_{nj}))) dt = \frac{\exp(t \sum_j \exp(-(V_{ni} - V_{nj})))}{\sum_j \exp(-(V_{ni} - V_{nj}))} \Big|_0^{\infty} \\
&= \frac{1}{\sum_j e^{-(V_{ni}-V_{nj})}} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} \quad (2.7)
\end{aligned}$$

又由(2.1)得:

$$P_{ni} = \frac{e^{x'_{ni}\beta_i}}{\sum_{j=0}^J e^{x'_{nj}\beta_j}} \quad (2.8)$$

如果对任意向量 q , 定义 $\beta_j^* = \beta_j + q$, 然后利用 β_j^* 而非 β_j 对上式定义的概率进行重新计算, 结果与使用 β_j 得到的概率相同, 因为含有 q 的项都消掉了。也就是说, 增加任何客户属性变量 q 对结果都没有影响, 因此可以得出结论, 客户的属性, 如性别、职业等, 对结果不产生影响。为此可以标准化 $\beta_0 = 0$ 。(因为概率和为 1, 所以确定 $J+1$ 个概率只需要 J 个参数向量。) 因此, 概率为:

$$P_{ni} = \frac{e^{x'_{ni}\beta_i}}{1 + \sum_{j=1}^J e^{x'_{nj}\beta_j}} \quad (2.9)$$

通过上式, 可以计算 J 个对数几率比:

$$\ln\left(\frac{P_{ni}}{P_{nj}}\right) = \ln \frac{\frac{e^{x'_n \beta_i}}{1 + \sum_{k=1}^J e^{x'_n \beta_k}}}{\frac{e^{x'_n \beta_j}}{1 + \sum_{k=1}^J e^{x'_n \beta_k}}} = x'_n(\beta_i - \beta_j), \quad j \neq i \quad (2.10)$$

若模型有 $J+1$ 个类别，则可表述如下：

$$\ln\left(\frac{P_{n0}}{P_{nJ}}\right) = x'_n(\beta_0 - \beta_J) \quad (2.11)$$

$$\ln\left(\frac{P_{n1}}{P_{nJ}}\right) = x'_n(\beta_1 - \beta_J) \quad (2.12)$$

.....

$$\ln\left(\frac{P_{n(J-1)}}{P_{nJ}}\right) = x'_n(\beta_{J-1} - \beta_J) \quad (2.13)$$

其中 J 为参考类别，也就是说，以 j 为标准，与其他选择做对比。

2.3.4 MNL 模型参数估计

为了估计模型参数 β_i ，定义虚拟变量

$$d_{ij} = \begin{cases} 1, & \text{如果第 } i \text{ 个顾客选择商品 } j \\ 0, & \text{否则} \end{cases} \quad (2.14)$$

那么对应的似然函数可表示为：

$$L = \prod_{i,j} P_{ij} d_{ij} \quad (2.15)$$

为了便于计算，将(2.15)写成对数似然函数形式：

$$\ln L = \ln \prod_{i,j} P_{ij} d_{ij} = \sum_{i=1}^N \sum_{j=0}^J d_{ij} \ln P_{ij} \quad (2.16)$$

参数 β_i 可以通过对(2.16)做极大似然估计得到。在本文中，我是通过 SAS 软件来实现上面的模型估计过程的。

2.3.5 MNL 模型假设检验

本文对 MNL 模型进行两类检验：T 检验、似然比检验。

一、T 检验

与一般的回归模型类似，T 统计量主要用于检验单个变量对离散选择模型影响的显著性检验。由正态分布表可知，当 T 统计量的绝对值大于 1.96 时，可以有 95% 的把握认为对应的变量是影响选择概率的主要因素。我将在第五章中运用 T 检验，得出对客户选择产生影响的变量，以进一步作出分析。

二、似然比检验

似然比(Likelihood Ratio, LR)是反映真实性的一种指标,属于同时反映灵敏度和特异度的复合指标。似然比检验的实质是在比较有约束条件下的似然函数最大值与无约束条件下似然函数最大值。这里设有约束条件下的似然函数的最大值为 L_1 ，无约束条件下似然函数的最大值为 L_2 ，则似然比卡方统计量为：

$$-2(\log L_1 - \log L_2)$$

该统计量服从自由度为 $K_1 - K_2$ ($K_1 < K_2$) 的卡方分布，这里 K_1 和 K_2 分别表示以上两模型的参数的个数。在第五章中，我将对模型的拟合优度进行似然比检验，以检验该模型的假设是否可以被接受。

2.4 大数据精准营销的研究现状综述

大数据一词自本世纪初出现以来，短短十几年间，得到了各界的广泛认可与追捧。研究机构 Gartner 对大数据的定义为：经过新模式的有效处理后具有更强的决策力、洞察发现力与流程优化能力的海量、高增长率与多样化的信息资产。这表明大数据不只是海量数据的堆积，其战略意义不在于掌握庞大的数据资料，而在于对这些数据资产进行专业化的处理，进而得出所需要的信息，最终达到对市场进行反映、对客户需求的满足以及实现自身盈利的目的。

大数据时代的到来有其非常复杂的背景和条件。综合前人的研究观点，可以得出以下两个结论：一、云计算技术的低成本高容量的存储能力和快速计算能力，是大数据技术得以产生并迅猛发展的基础。经过多年的发展和积累，云计算技术已经基本实现了从终端到服务的聚合，多平台上纷繁复杂的各类数据

通过云端整合成庞大的数据集，这些大数据再被各平台获取，经过过滤提炼挖掘分析等各种方式进行处理后，最终才能得到对各自有用的信息。也就是说，离开了云计算技术的存储和计算能力作为技术支持，大数据也不过仅仅是大量的数据而已。二、如果说云技术的成熟时大数据技术得以诞生和发展的推手的话，那么大数据技术大规模的被使用，则从另一个方面拉动了大数据技术的发展。国际文献资料中心曾对数据量作出如下统计，2010年起，世界范围内的数据量达到 ZB 数量级，且将以每两年翻一番的速度增长，预计到了 2020 年，全球数据量将达到 35ZB。

大数据具有数据体量大、类型多、价值密度低以及实时处理的特点，这就从数据来源、处理方式甚至企业的数据思维等方面对市场营销的管理变革提出了挑战。与传统营销强调创意、策略及覆盖度相比，大数据背景下主导的营销理念，是一种以技术为驱动、以海量数据挖掘为前提，对特定受众的个性化广告传播。

精准营销正是在这一背景下被提出来的。较公认的说法是精准营销的概念是由菲利普·科特勒在 2005 年首次提出的，他同时提出了精准营销的三个关键点：精确定位、可衡量和高投资回报。我国著名精准营销学者徐海亮提出：精准营销就是在精准定位的基础上，依托现代信息技术手段，建立个性化的顾客沟通体系，实现企业可度量的低成本扩张。刘征宇认为精准营销是通过定量和定性相结合的方法对目标市场的不同消费者进行细致分析，根据他们不同的消费心理和行为特征，企业采用有针对性的现代技术、方法和指向明确的策略，实现对目标市场不同消费者群体强有效性、高投资回报的营销沟通。

目前为止，学术界对于精准营销尚未形成一套完整的理论，这里根据前人的研究成果，可以将大数据精准营销理解为：依托云计算条件下发达的互联网分享机制，运用现代信息技术对既有的大数据进行分析与预测，帮助企业向特定客户应用特定策略投放特定产品的营销方式。

第三章 A 小微企业电商的运营现状

为了使研究更有实际意义,要了解作为一个新卖家或初入某新的区域的老卖家在没有获得历史成交数据前,如何通过对公开数据的抓取,如八爪鱼这种网页数据采集器,收集到交易数据,进行对数据分析从而做出决策,我们就必须站在运营者的角度,从一个店铺在经营过程中会遇到的问题出发,让数据挖掘技术得到实际运用,帮助店铺运营者处理实际问题。通过寻找在运营过程中遇到的问题,选择合适的方法进行处理,搜寻待处理因子,再通过建立模型,筛选出起到重要影响作用的因子,从而帮助运营者更好的达成预设的运营目标。

下面我将以 A 小微电商为例,一步步深入到电商的运营中去,分析其实际存在的问题,并据此筛选得到最有价值的数据,使得淘宝运营商可依据分析出的结果,让有限的资金得到最高效的运用。

3.1 小微电商在运营中所面临的问题

3.1.1 电商运营概述

运营是对运营过程的计划、组织、实施和控制,是与产品生产和服务创造密切相关的各项管理工作的总称。

近年来,电商的发展非常迅速。随着 2003 年淘宝网以及 2004 年京东商城成立,电商逐渐走入人们的日常生活中,成为人们日常消费的重要通道,大家的接受度越来越高,在生活中所占的比重也越来越大。2014 年,阿里巴巴、京东、聚美优品的成功上市,网购行业的网络零售市场格局趋向于稳定的态势,巨头模式日益显现,淘宝、天猫、京东的品牌渗透率稳居前三,遥遥领先于同类对手。第四名的唯品会虽然以特卖的形式后来居上,其品牌渗透率也仅在 18.8%。电商品牌逐渐固化,进入成本渐渐升高,电商的运营逐渐由平台的运营转向在已搭建好的平台上做优化与运营。由于本研究的重点为解决小微商户遇到的问题,而京东、唯品会等平台的进入门槛相对较高。所以,这里重点讲一下淘宝

平台的运营。

小微电商的运营，不同团队在不同阶段的理解都不尽相同。

总的来说，一个运营体系通常包括五个要素：产品、客户、流量（引流）、转化（优化）和用户黏性（营销）。

通常说的流量，是指页面的访问量，是用来描述访问一个店铺的用户数量以及用户所浏览的产品数量等的一个指标，常用的统计指标包括独立用户数量（一般指 IP）、总用户数量（含重复访问者）、产品浏览数量、用户在店铺的平均停留时间等流量问题。引流就是通过运营手段提高店铺流量，降低店铺流量的获取成本。

转化率是指在一个统计周期内，完成转化行为的次数占推广信息总点击次数的比率。提高转化率，就是要通过营销策划、店铺设计等方式来提高店铺潜在客户与实际发生购物行为的客户的比值。

用户黏性，是衡量用户忠诚度计划的重要指标，电商的特殊性，决定了店铺的用户黏性的培养，无法通过市场销售来完成，因此通常做法是通过淘宝提供的数据，跟踪客户在页面的停留时间，在店铺的搜索路径以及在店铺的平均停留时间等，了解客户需求，再通过 CRM 系统和售后跟踪服务，来增加店铺的忠实客户。

电商运营的中心问题就是通过解决流量、转化率和用户黏度这三个问题把产品和客户这两个端点结合起来，达到公司盈利的目的。

3.1.2 淘宝平台现状

电商平台间竞争的加剧，使得淘宝在近些年也做出了很多的改变。

首先，为了增强竞争力，淘宝对于内容上的注重加强了很多。微淘最初在淘宝网并不受到重视，在这次的变革中，微淘被重新放在了手机淘宝的重要位置；另外，淘宝开设了淘宝头条栏目，并将其放在了非常显眼的位置，还增加了产品推荐的栏目。

其次，2016 年阿里巴巴成为了新浪微博的第二大股东，两家联姻的直接结果是淘宝与微博的账号可以互相关联，消费者在淘宝的浏览历史出现在了微博的广告推荐里面，将社交与购物紧密的结合了起来。虽然由于此举引发的隐私

争议层出不穷，但信息共享带动的精准营销的流量还是让双方都达到了共赢的目的。

第三，自阿里巴巴提出“新零售”的概念以来，已用与百联合作的实际行动表明了淘宝的方向与态度。纯电商的未来，至少现阶段的小商户还看不出来。另外，卖家用户的增多，买家用户增速变化，也打破了原有淘宝市场的平衡，淘宝也为此做出了不少的规则调整，如增加了直通车的广告位；将新品扶持的时间由原本的两周调整为两天；设置了层级，对不同层级的卖家用户设置了不同的流量限制^[38]。

综上所述，为了应对其他电商平台的竞争，淘宝做出了多项改革，针对买家的改变主要在提升用户体验方面，一方面提升内容的占比以吸引住用户，另一方面利用数据优势推荐更符合用户的宝贝，降低用户的搜索时间，提升用户的购物体验。针对卖家的改变主要体现在控制方面，基于卖家用户的不断增多，淘宝更趋向于让好的店铺更有竞争力，让不好的店铺逐步退出。他们通过对流量的合理分配来达到这个目的，也就是说淘宝对于免费的流量将不再在各卖家间均匀分配，而是成交额高的店铺将得到更多的流量，成交额低的店铺只有极少的流量，以此来动态的调整卖家的数量，有运营能力的卖家会更愿意留在淘宝，运营能力差的卖家成交额也会随着这一改革变得更低，进而不得不逐步退出淘宝。因此，运营能力就成了一家淘宝店铺能否盈利甚至存活的关键所在。

3.1.3 淘宝平台上的小微电商所面临的问题

在大数据时代，数据本身也是一种资本，它能够为企业创造出新的经济利益，因此，企业对大数据的掌握程度可以直接转化为新经济价值的来源。随着电子商务的普及，对电商数据的数据挖掘技术的研究并将其运用于精准营销显得越来越重要。大中型企业和电商商家可以通过资源共享甚至试错的方式获取数据信息，可是小微企业主对数据的运用尚停留在较浅显的层面。

电子商务的商家开始店铺的运营，在根据店铺优势选定了要卖的产品类型后，第一步要做的事就是根据用户的行为预测分析进行选品，以及选择合适的工具进行引流，这是淘宝运营的重点和难点。如果前期没有投入足够的时间精力和资源手段对选品进行研究，后期产品上架后才发现选出的产品卖不动，就

需要付出更多的时间和成本，为之前犯下的错误买单，先用传统粗放的铺货模式再用试错的方式去筛选 SKU，这种运营模式对小微电商而言是很沉重的负担。

中小卖家的选品战略，一般来说都是以结果为导向的。目前国内电子商务企业对用户的行为预测分析，常用的参数集中在消费者的最近购买时间，产品的购买频率，消费者平均订单金额，单次最高交易额，客单价等等，但是这些数据对于淘宝上一般的商家而言，除了曾在本店消费过的客户，商家可以通过向淘宝购买服务的方式获取以外，未在本店消费过的客户的数据无从获得，对于新建店的小微商家或计划开辟一片新的领域的老商家而言，精准营销就比较空泛。

店铺引流同样面临这个问题。淘宝主推的直通车、钻展及淘客，是业内商家常用的引流工具。这种付费引流带来的流量能引起多少转化率最终能给企业带来多少利润是另一个课题，在选择使用这些引流工具前，如何通过既有数据，在面对竞争激烈的市场时，如何通过对免费共享数据资源的挖掘与开发利用，得到数据间有用的潜在信息，并将其运用到营销中去，进而降低广告成本，提高利润，目前似乎尚无有效的理论支持。

3.2 现有的解决途径及其局限性

目前，电商企业在选品上通常采用的策略还是通过便宜爆款刷单烧广告走量。这一流程的起点是刷流量，终点却往往不是因流量而引入的利润。原因如下：

首先，为了获得流量店铺需要怎么做。对流量的分类通常有两种：按照流量的来源分类，一般有两个渠道：站内流量和站外流量；按照流量的获得成本分类，可分免费流量和付费流量。一般而言，免费流量通常来源于：一、站外搜索引擎，如百度知道维护、百度百科维护、SEM 等；二、SNS 社区，如购物分享、推广软文等；三、站内流量，如友情链接、店铺收藏等。近年来，由于互联网技术的发展，信息的增长速度呈现指数级递增的态势。信息过滤技术的发展滞后于信息过载的速度，加上淘宝平台本身基于布局设置做出的规则调整，导致免费流量的效果对小微电商带来的效果已日渐减少，某些企业从投入产出比的角度来看，其免费流量带来的效益甚至不能覆盖投入于引入免费流量

的推广人员的人力成本，因此大部分电商将更多的精力投入于淘宝提供的付费推广工具上。

其次，推广工具有哪些。由于淘宝改革，自然流量少了很多，获取流量的主要途径为直通车、钻展、站内站外的活动、淘宝客以及 SEO 优化，其他可选的属性还有：假一赔三、正品保障、7 天无理由退货、保证如实描述、支持货到付款、退换处理、全球购卖家、旺铺等。这些引流途径的选择都将直接涉及到店铺的投入产出比，对店铺利润带来直接影响。但是，这些引流途径具体能对客户最终选择起到多少效果有待考察，因此引流的途径作为影响因子，是我们关注的重点。

再次，运用这些推广工具需要付出些什么。以直通车为例，淘宝直通车是阿里巴巴公司推出的为专职淘宝卖家量身定制的，按点击付费的效果营销工具。经过雅虎中国和淘宝网的资源整合，这一全新的搜索竞价模式，可以根据客户在淘宝的搜索和流量，对有购买意愿的客户与购买直通车精准词的商家进行匹配，为卖家实现宝贝的精准推广。他的竞价结果不只可以在雅虎搜索引擎上显示，还可以以全新的图片加文字的形式在淘宝网上充分展示。每件商品可以设置 200 个关键字，卖家针对每个竞价词进行自由报价，并且可以看到自己的定价在雅虎和淘宝网上的排名，最终依照产品在直通车上实际被点击次数乘以自己的报价来付费。每个关键词最低出价为 0.05 元 最高出价 99 元，每次可做最低 0.01 元的加价。

直通车的这种通过精准的搜索匹配的推广方式，在给宝贝带来曝光量的同时，也给店铺带来了精准的潜在买家，其优势非常明显。通过淘宝直通车进行推广的宝贝，通常想在淘宝够买该类宝贝的人都能看到，大大提高了宝贝的曝光率。由于淘宝对于关键词的关联设置，能看到广告的只有实际在搜索的或者通过点击被淘宝认定为有购买意愿的人，因此带来的点击都是有购买意向的点击，带来的客户都是购买意向的买家，针对性很强。淘宝直通车的另一个显著的优点在于能给店铺带来人气。当买家通过对直通车推荐宝贝链接的点击进入店铺后，通常会到店铺里去看一看，虽然做推广的是单个宝贝，但是逛店的行为通常能带动店铺内的跳转，产生一次甚至多次的店铺内流量，最终可能带来多个产品的成交，这种以点带面的关联效应可以降低店铺的整体推广成本，并对提高整店的关联营销效果有积极的营销。总的来说，精准营销，以及通过宝

页间的关联设置带动整体连锁反应，是直通车推广的最大优势。

直通车最大的缺点就是投入较高。首先，淘宝对购买直通车的商家有最低要求，初创公司或个人小卖家很难在开始时就直接利用此工具进行目标客户的精准营销。其次，直通车是用竞价的方式在做的。稍热门的关键词在直通车上的叫价都非常高，综合实力不强的商家只能短暂的采用这种方法宣传。最后，直通车的收费是根据点击收费而非转化收费，一天花费上万元的公司不在少数，其转化率却各有不同，直接导致各店铺的收益率天差地别。尤其是初创企业，一旦收入达不到预期，只能更换营销方式，店铺人气的增长也成了昙花一现的虚假繁荣。

最后，如何获知商家选择的推广工具给实际运营带来怎样的影响。以往的运营中，这一点只能通过积累店铺销售数据获知。近期每日近万家淘宝店的倒闭，其试错成本的高昂往往占了很大的因素。

3.3 A 小微电商企业运用基于大数据精准营销的必要性

随着互联网技术的发展，人们越来越多的行为通过网络来实现，购物、学习、娱乐、社交等等，都因为有了云端的介入，让人与人之间的各类信息的交互越来越频繁，而这些数据也被互联网忠实的记录了下来。如果可以通过网络的共享而获得，进而经过数据挖掘分析，形成对客户的全方位洞察，再通过个性化的分析，对用户未来的行为方式和决策思维作出预判，完成企业对用户的精准的广告投放和营销推送，小微电商企业就能最大限度的避免浪费，提高客户转化率，达到精准营销的效果。

选品上架是电商运营的第一步。众所周知，产品是营销 4P 要素之首，精品化的选品策略要求商品做到以下几点：质量好、利润高、需求大、轻售后。好的产品，常常伴随着高客单价、高复购率、低更新频率和轻售后，间接带来高毛利率和高转化率。这是电商企业能够长期稳定发展的基础。

淘宝平台上的小微电商对于产品上架前的选品对策以及如何有针对性的选择引流工具这两个问题尚无系统性的解决方案，常常依靠的是运营者的经验和 CEO 的直觉。如何利用基于网络可获得的含有消费者及商家信息的消费数据，对客户未来的选择模式进行预测，从而加以利用，并提前进行准备或调整，

以达到精准营销的目的就成了当务之急。

3.4 数据的采集与预处理

要了解香水品类的销售情况，作为一个新卖家，最稳妥的做法是对该品类在平台上的销售情况做一个充分的了解，选择平台卖的最好的产品作为进货时的首选产品，要了解的细节包括：产品的品牌，产品成交的价格（据此作出定价策略，并对公司盈利进行分析及预测），客户对平台的选择偏好（为公司在平台上的选择提供参考），对产品属性的偏好（作为进货时对产品属性的选择依据），客户在选购中关注的重点（后期推广中以此为依据，对页面的设计作出方向性的指引），客户的分布规律（鉴于淘宝的区域推广是付费功能，该数据将作为在淘宝推广工作中选择推广区域的重要依据）。以上这些细节均可通过淘宝页面获得。本研究，我使用了八爪鱼软件，随机抓取了淘宝平台 2016 年 6 月一整月的“香水”类目的销售记录作为样本。

八爪鱼是一款网页数据采集器，由于其易操作型和强悍的功能性，在数据采集软件中应用相对比较广泛。它通过配置程序按照人工操作流程记录程序规则，软件完成对规则的配置后，依此进行数据采集，代替人工步骤，完成对所需要的页面信息进行抓取的工作。

我在采集销售数据时按照如下工作流程，对数据采集工作作出了规则设置：

- 1、在淘宝主搜索页输入“香水”，将该淘宝搜索列表页的地址复制到八爪鱼里面，创建一个打开网页步骤；
- 2、点击页面底部的“下一页”按钮，创建一个循环翻页的操作；
- 3、点击列表页中的第一款商品及第二款商品的名称，系统将智能创建列表；
- 4、创建完成后点击循环，进入每个商品的详情页中，此时在详情页中再对所需要的数据进行采集。这是比较关键的一步，所需要的数据均位于该页面中，我们需要对需求进行详细了解后，对该页面所需采集的数据作出整理，最终，我选择了店铺名称、产品名称、价格、交易成功数等等数据作为采集对象；
- 5、选择完成后，对各字段名称进行修改，对有干扰的冗余信息进行初步的降噪处理。
- 6、对流程稍作整理后，即可完成对基本规则的建立，就可以开始对数据的采

集工作了；

7、在完成了数据采集后，最后选择导出为 excel 格式，即初步完成了数据采集的工作；

8、以上数据由于淘宝的设置，很多内容混杂在一起，在进行分析前，需要对数据进行整理，如通过网页地址分析，筛选出参加直通车的产品；通过对服务承诺内容的分类整理，将 7 天无理由退货、品牌授权、包邮、公益宝贝、运费险等等内容分类整理出来，将不合法值、空值等等问题剔除出去，最终清除掉脏的数据，保证了数据的质量，为下一步进行数据分析做好准备。

我最终整理出来的成交数据中包含的信息有：产品类型，平台，店铺名，店铺信誉，产品评分，店铺所在地，标题，宝贝 ID，品牌，产品规格，是否包邮，标题，标价，成交价，销量，交易成功数，评论数，收藏数，服务承诺（含假一赔三、正品保障、7 天无理由退货、保证如实描述、支持货到付款、退换处理、全球购卖家、淘客、旺铺、直通车共十种属性），支付方式等。经过筛选后，得到的有效的完整数据信息有 927174 组。

在完成数据采集后，第二步，需要对数据进行分类筛选：哪些是网站用于做身份识别的信息，不具备数据勘探与挖掘的价值；哪些数据几乎重复，可剔除一半；哪些数据需同时考虑，得出有用的数据。最终，研究对象由哪些因子构成。

A 小微电商企业有香水的渠道商资源，因此，我将产品类型选择为“香水”；筛去网站用于做身份识别的信息，如宝贝 ID；通过综合比较，筛去有交叉且有干扰作用的数据，如销量和交易成功数中的“销量”；不考虑对决策不产生影响的因子，如店铺名，店铺信誉，产品评分等。最终留下对选品有影响的因子：产品规格、成交价等与客户相关的因子；交易成功数、服务承诺等与店铺相关的因子；平台、区域销售额等与上架相关的因子。

本文要研究的问题分两部分：一是当公司有香水方面的资源时，如何通过分析平台销售数据找到利基市场并据此做出选品决策；二是当顾客已经决定购买某类香水的情况下，公司如何利用平台销售数据对推广渠道做出选择。鉴于本研究的背景是没有店铺历史销售数据的小微电商的选品问题，我们先忽略消费者基本信息的因素。

在选品策略方面，主要通过销售渠道的数据选择商品上架平台，通过区域

分析得出在国内哪些地区的销售额比较可观，继而选择产品推送区域。

从消费者偏好方面，对于不同的平台结构通过数据进行分析，并对何种品类的哪种品牌的产品卖的好以及平均价格的定位作出一定的分析。具体到价格区间，发现目标客户群的潜在购买力以及消费区间；通过对于产品属性的分析，得出产品哪些特性会特别得到消费者的青睐。最终为 A 公司关于香水的精准营销决策提供依据。

第四章 基于销售数据分析的选品策略

A 小微企业电商已有近 2 年的淘宝网店运营基础，在母婴用品等方面已有很好的经验。鉴于公司新开发了“香水”品类的购货渠道，公司希望开辟一片新的市场，而 A 公司没有既没有历史销售数据可以借鉴，又没有理论可以参考。

本研究正是基于这一现实问题，提出了如何通过对网络可获得的公开信息进行采集、归类、挖掘分析，根据用户的行为预测分析结果进行选品并选择合适的工具进行引流，从选品策略和产品上架的营销渠道的选择这两方面，运用精准营销的思维方式为其提出解决方案。

通过利用网络数据采集器，我将对收集整理好的 927174 组共 2688.8 万个数据进行分析，由于这些数据具有数据量大、类型繁多、价值密度低的特点，因此在分析前我对数据先做了一定的预处理，利用此预处理的结果，从渠道及地区两方面，对目标市场做出定位，从产品属性及品牌方面，对消费者需求偏好做出定位，从而为 A 公司香水晶类的精准营销提供决策依据。

4.1 基于目标市场的销售数据分析

关于目标市场的选择，我从销售渠道与区域两方面开展数据分析工作，具体如下：

4.1.1 渠道分析及结论

表 4-1 各平台数据表格

平台	店铺数量	销量	销售额	平均交易价格	人气	市场占有率
淘宝	1760	1022101	46043621.67	45	571262	55.07
天猫	163	674597	37563710.65	55.7	355912	44.93

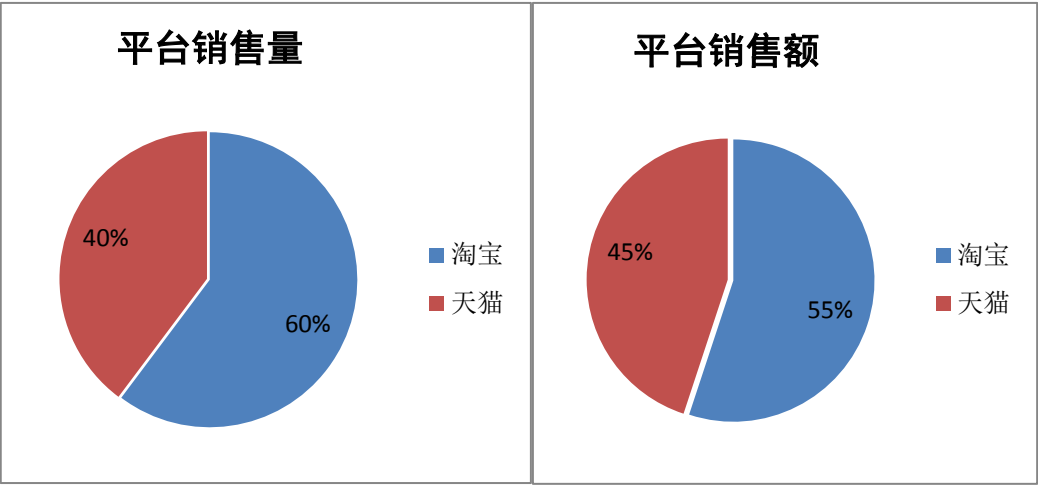


图 4-1 各平台销售数据对比

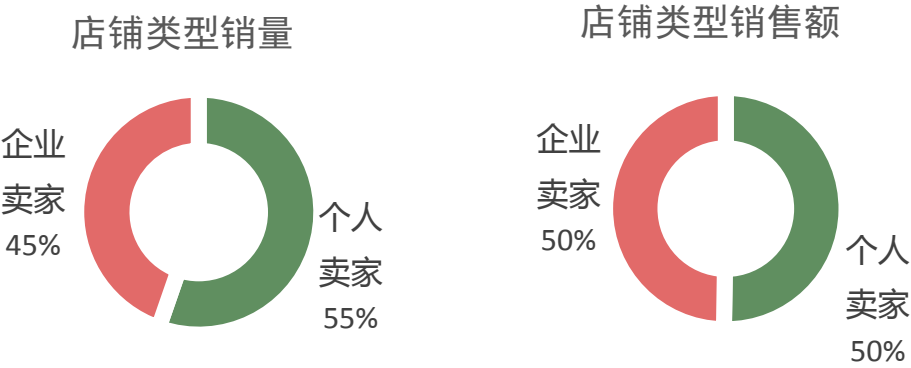


图 4-2 不同类型店铺销售数据对比

本文对比了淘宝平台以及天猫平台的销售额以及销售量的数据，可以看出无论是从平台销售量来讲，还是从平台销售额方面，淘宝都领先于天猫，一方面说明淘宝平台的使用人数相对多一些，另一方面从平台销售额来分析，淘宝 55% 的市场占有率，天猫 45% 的市场占有率，相对于平台销售量的差距有所减少，从侧面说明淘宝的价格以及打折力度应该强于天猫平台。由第二个图形中可以看出，淘宝全网铺类型主要包括个人和企业。发展趋势倾向企业店铺，2016 企业店铺入驻达到高峰期。香水行业中企业店铺市场占有率达到 49.8%。对于淘宝平台，个人卖家比较更多，而天猫平台则是个人买家于企业卖家平分的情况，当然也不排除由于淘宝平台的商户类型比较多比较杂，导致价格会不稳定，而天猫平台则多数是官网的旗舰店，所以价格略贵的可能。

4.1.2 地区分析及结论

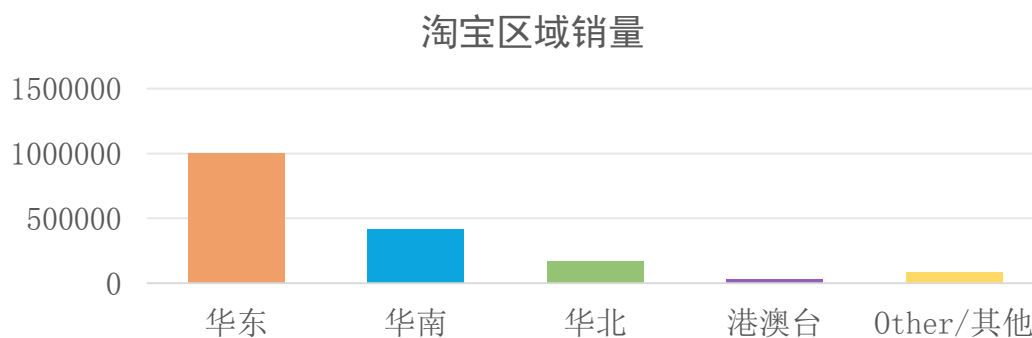


图 4-3 淘宝销售数据图

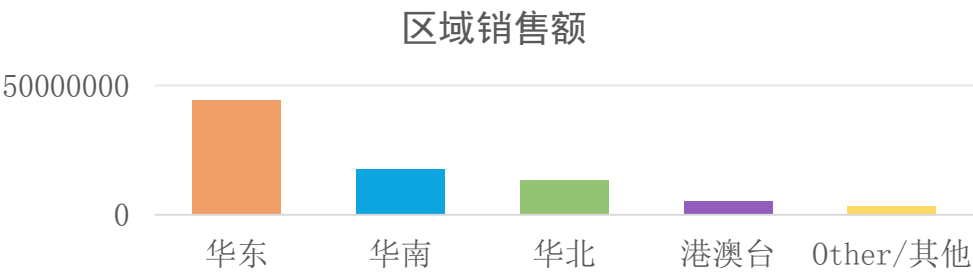


图 4-4 区域销售额

从上面两张淘宝区域销售量以及区域销售额图中可以看出，华东，华南，华北地区是重要的购买力地区，一方面是由于华东地区包括上海、江浙地区等，华南包括广州深圳等地区，华北包括北京天津等，相对于其他地区的消费水平比较高，大中型城市比较多，同时职场女性或者是适龄女性对于化妆品的需求比较旺盛，相对于其他地区属于发达地区，销量比较可观。

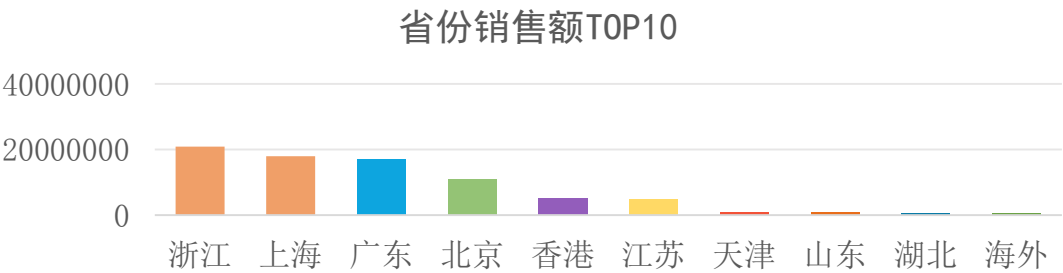


图 4-5 省份销售额 TOP 10

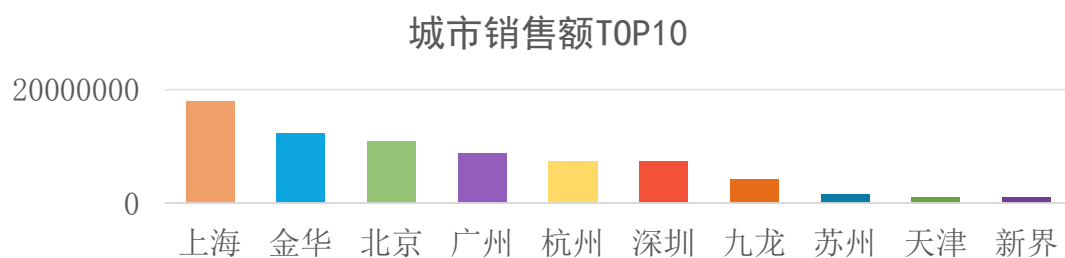


图 4-6 城市销售额 TOP 10

从省市以及城市销售额来看，上海位居第一，其次是金华、北京，具体到省市到统计中可以看到，华东地区有上海，金华，杭州，苏州四个城市上榜，华北地区有北京和天津，华南地区包括广州、深证，九龙，新界等地区，基本与省份销售额以及地区销售额一致。

4.2 基于消费者的需求偏好分析

关于消费者的需求偏好，我从产品属性和品牌选择两方面开展数据分析工作：

4.2.1 产品属性分析及结论

最受卖家及消费者关注的就是关于产品的属性以及价格，下面从以下两个方面来通过数据分析产品以及价格对于消费者购买的影响。

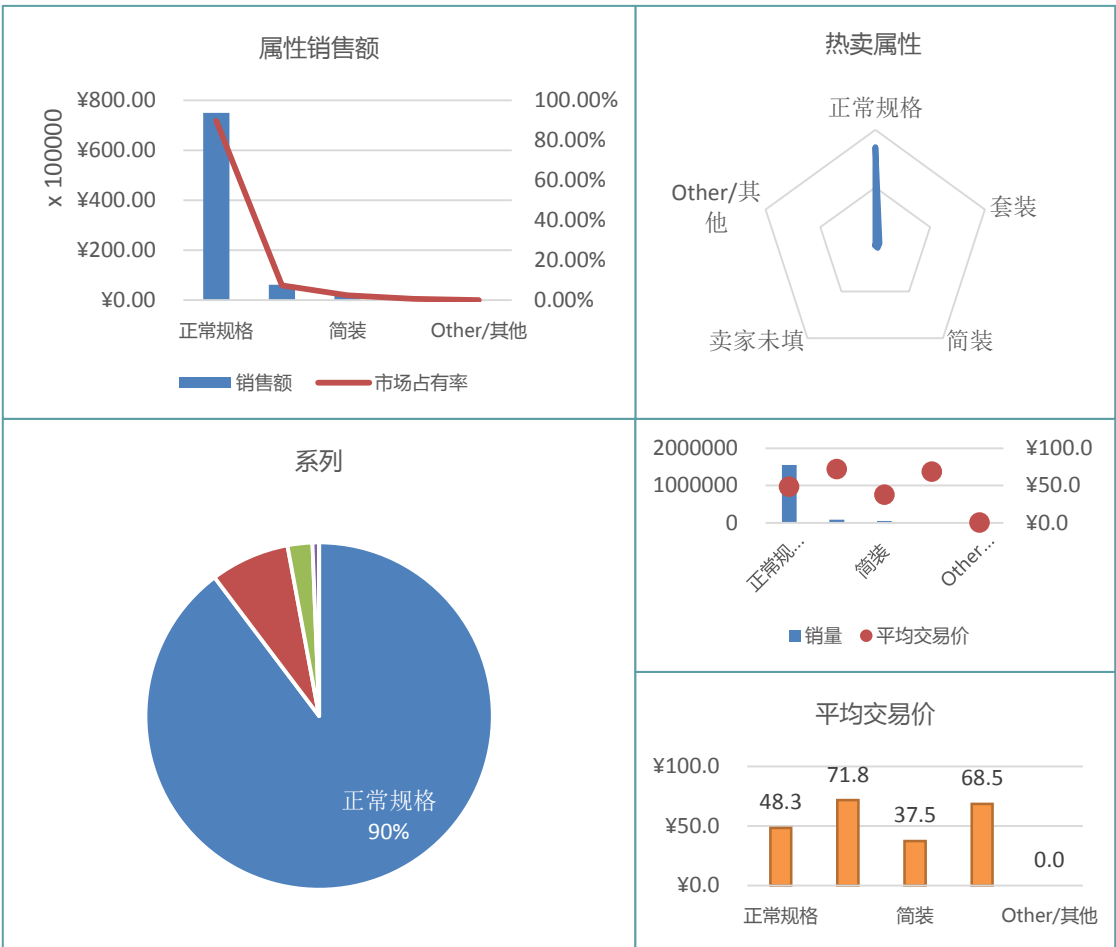
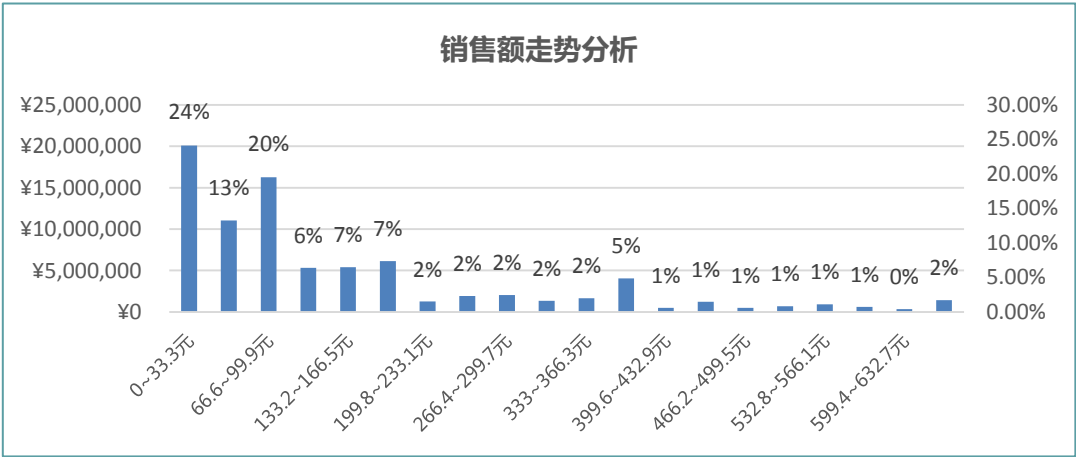


图 4-7 产品属性分析对比图



从图中可以看出，产品价格方面，从淘宝全网来看，香水行业出货量最多的价格区间在 11 元到 22 元之间，其中出货量占比 27.9%，销售额高的价格区间在 22 元到 33 元之间，其中销售额占比 14.3%。另外在 333-432 元的区间内也有显著的增加，说明大多数的消费者一部分还是选用比较低端一些的化妆品为主，并不追求品牌的产品；另一部分则倾向于购买品牌香水，集中购买在几款知名品牌；而中间价位的产品则较少有人选购。

表 4-2 产品规格分析

序号	系列	销量	销售额	平均交易价	人气	市场占有率
1	正常规格	1551991	¥75,007,996.26	¥48.3	849101	89.71%
2	套装	85644	¥6,152,876.36	¥71.8	39807	7.36%
3	简装	51564	¥1,932,707.10	¥37.5	32905	2.31%
4	卖家未填	7499	¥513,752.60	¥68.5	5361	0.61%
5	Other/其他	0	¥0.00	¥0.0	0	0.00%

上表可以看出，正常规格的销量占比最高，占据了市场占有率的近九成，而套装选择的较少，说明消费者的需求只是根据喜好选择某种类或某种香型的香水，而非购买某品牌全套的产品，也有部分消费者的消费倾向会更喜欢购买相应的简装产品，此种消费者可能是自用为主，讲究经济实惠，不追求包装，按需购买。

下图的产地分析，我们可以看出，进口香水与国产香水几乎平分秋色，而鉴于低价位产品的销量占比远高于高端进口产品，说明国内香水以低价位走量的营销策略为主，而进口香水则胜在单价较高，因此销量仅略逊于国产香水而已。

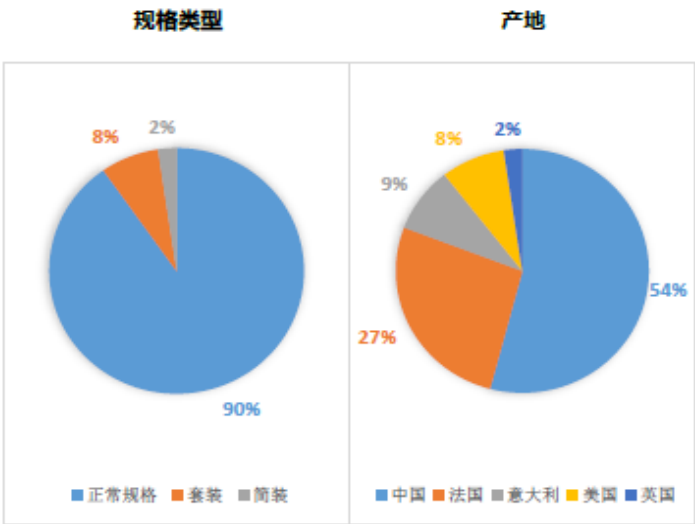


图 4-9 产品规格及产地分析图

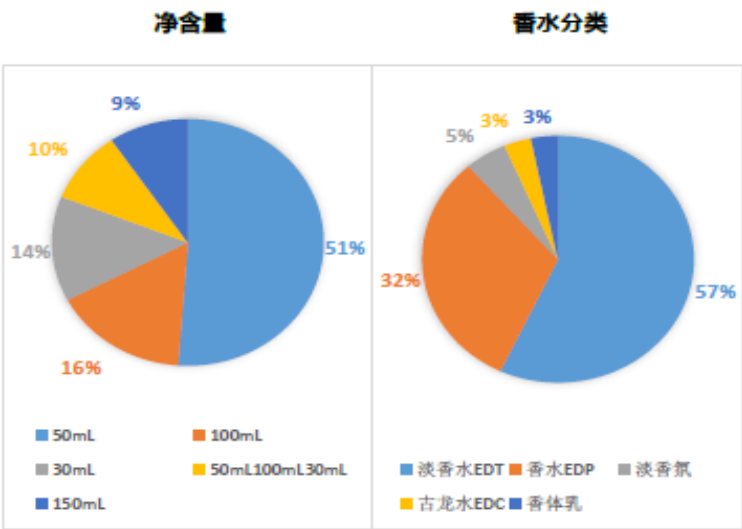


图 4-10 产品含量及分类分析图

通过分析香水的规格类型、产地、净含量、香水分类，在淘宝全网销售额数据对比发现，香水行业的产品包含：正常规格、中国、50mL、淡香水 EDT 的属性会更受消费者的青睐，能满足市场的需求。

鉴于香水的定价净含量越高单位价格越低，因此，由于消费者更愿意选择性价比比较高的品种，所以在同等价位的情况下，消费者倾向于选择净含量高的品种。

4.2.2 品牌分析及结论

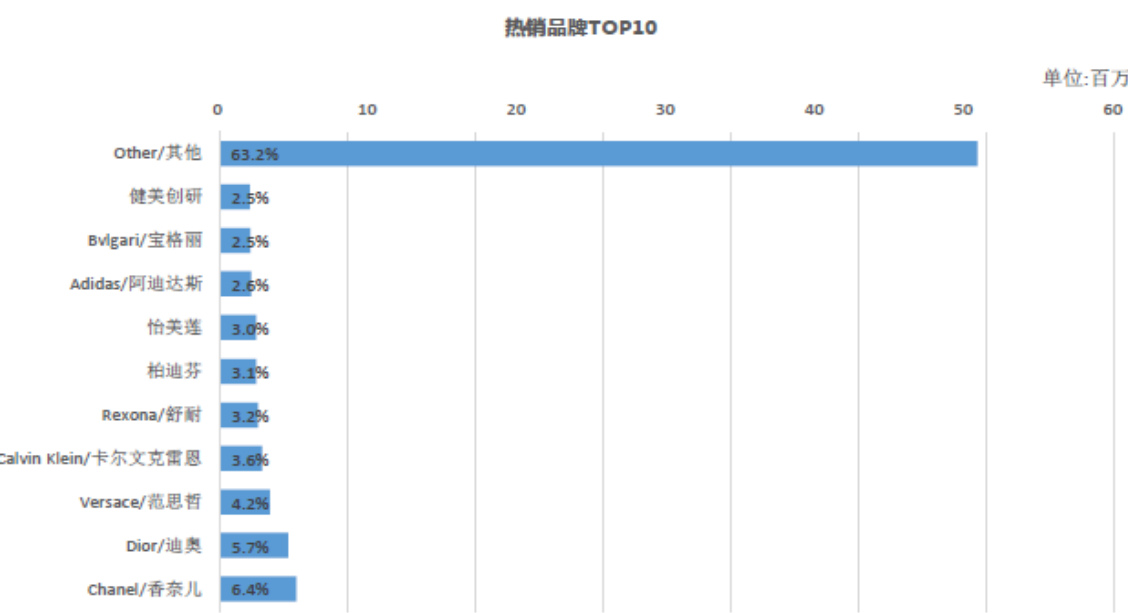


图 4-11 热销品牌 TOP10

为了精准定位目标客户群，我们需要分析淘宝平台上的热销品牌。通过以上图表，我们可以看到，消费者大多数还是选择其他的不知名品牌，这与消费者购买价格有一定的关系，在知名的国际品牌当中，香奈儿、迪奥、范思哲、卡尔文·克雷恩是消费者购买最多的四个品牌。而这部分人群基本定位在高端的人群，从消费价格可以看出也是在均值为 330-430 元的区间内。说明国外一线品牌还是受到消费者的欢迎和认可，有能力够买的消费人群还是会为其买单。

4.3 选品策略的研究结论

以上，我们通过对销售数据的整理分析，可以得出以下结论：

- 一、在目标市场的渠道选择上，应选择在平台销售量和市场占有率均偏高的淘宝平台销售；
- 二、在选择推广区域时，可选择浙江、上海、广东、北京、香港、江苏等省市为主的地域进行有针对性的推广；
- 三、产品属性应选择 400 元左右的国产简装 50ml 正常规格的 EDT 淡香水产品；
- 四、如销售知名品牌，以香奈儿、迪奥、范思哲和卡尔文克雷恩为首选。

综上所述，通过对在淘宝上爬出的九十多组数据的整理分析，以及 A 公司精品化的选品策略，我们已经可以初步将 A 公司的精准营销方案细化为：四款品牌香水，其品牌为香奈儿、迪奥、范思哲和卡尔文克雷恩，其属性为：位于 330-430 区间、正常规格、50ml、国产简装淡香水，销售平台为淘宝，推广区域选择浙江、上海、广东、北京、香港、江苏等省市为主的地域。

接下来，我们将以符合以上属性的四款品牌香水为目标产品，对其进行上架策略的分析。

第五章 基于顾客购买行为分析的上架策略

当顾客在选择商品时，经常要面对多个可选择的因素，这些因素都有可能作为首选因素直接影响到顾客最终的购买决策，这里就需要离散选择模型来帮助我们对客户的选择行为作出预测。

通过对顾客购买行为的分析，我们知道，客户在选择购买产品时影响最大的因素除了产品还有服务。因此，我们将卖家的十项服务承诺（假一赔三、正品保障、7 天无理由退货、保证如实描述、支持货到付款、退换处理、全球购卖家、淘客、旺铺、直通车），作为影响顾客在不同品牌之间进行选择的产品属性影响因子，设为虚拟变量引入模型，通过运用 MNL 模型来预测顾客在不同品牌之间选择的概率，进而为 A 公司在产品上架过程中的推广渠道的选择（如：提供何种服务，是假一赔三还是 7 天无理由退货，是否提供退换处理，是否加入全球购，是否购买直通车服务等）提供决策依据。

要对第三章结论中的目标产品的数据进行分析，首先需要在已初步整理出的 927174 组数据中将符合要求的数据分离出来。经过初步筛选，得到符合第三章结论的 63345 组目标产品数据，再在其中挑选出“服务承诺”属性，作为建模的数据基础。接下来，需要对这部分数据进行分析处理，我们将 70%（前 50000 组）样本作为训练样本进行建模分析，得出模型后，再用模型对剩下 30%（13345 组）的测试样本进行预测，最终检验预测结果。

接下来我将把变量代入第二章的 MNL 模型，用 SAS 软件对其进行检验，以下为实证分析，以及根据此分析结果对产品的上架做出推广渠道的选择。

5.1 数据描述

表 5-1 样本变量的描述统计

		变量名	值	频数	百分比
香水品牌	香奈儿	Brand1	1	42940	67.79
	迪奥		2	7605	12.01
	范思哲		3	8205	12.95
	卡尔文·克雷恩		4	4595	7.25
产品属性	7天无理由退货	VAR18	0	13851	21.87
			1	49494	78.13
	假一赔三	VAR19	0	63265	99.87
			1	80	0.13
	全球购卖家	VAR20	0	59383	93.75
			1	3962	6.25
	退换处理	VAR21	0	53936	85.15
			1	9409	14.85
	淘客	VAR22	0	35797	56.51
			1	27548	43.49
	旺铺	VAR23	0	8788	13.87
			1	54557	86.13
	保证如实描述	VAR24	0	10291	16.25
			1	53054	83.75
	正品保障	VAR25	0	55070	86.94
			1	8275	13.06
	支持货到付款	VAR26	0	57724	91.13
			1	5621	8.87
	直通车	VAR27	0	12756	20.14
			1	50589	79.86

我们的目的是为了研究十项服务承诺这一产品属性对客户选择产生的影

响，因此为了对影响顾客的在不同品牌之间进行选择的因素进行分析，我将数据采集中的“服务承诺”所包含的十种属性作为十个虚拟变量引入模型。

这里定义一个 4 分类因变量，当它取 1, 2, 3, 4 时分别表示顾客选择香奈儿，迪奥，范思哲，卡尔文•可雷恩。

VAR18 描述淘宝卖家是否定制了 7 天无理由退货这项产品属性，VAR18 等于 1 表示淘宝卖家定制了这项服务，0 表示淘宝卖家没有定制此项服务，VAR19-VAR27 依次类推，描述卖家是否定制了相应的服务承诺。于是就可以运用模型来预测顾客在不同品牌之间选择的概率。

于是可以得到下面的 MNL 模型：

$$\ln\left(\frac{P_{nj}}{P_{nj}}\right) = \alpha + \beta_{1j}\text{VAR18} + \beta_{2j}\text{VAR19} + \beta_{3j}\text{VAR20} + \beta_{4j}\text{VAR21} + \beta_{5j}\text{VAR22} \\ + \beta_{6j}\text{VAR23} + \beta_{7j}\text{VAR24} + \beta_{8j}\text{VAR25} + \beta_{9j}\text{VAR26} + \beta_{10j}\text{VAR27}$$

其中 P_{nj} 为顾客 n 选择第 j 种品牌香水的概率， j 的取值为 (4, 3, 2, 1)。

因为在 SAS 中用来进行 Multinomial Logit Model 估计的程序默认将反应变量的最后一个类别当作参照类。

5.2 数据实验

我将 70%（前 50000 组）样本作为训练样本，通过反复筛选变量得出以下结果。表 5-2、表 5-3、表 5-4 分别是数据概要、各自变量对模型整体影响的显著性检验以及各自变量的估计值和显著性检验。

1. 模型的检验

表 5-2 数据概要

Data Summary			
Response	brand1	Response Levels	4
Weight Variable	None	Populations	5
Data Set	TAOBA02_1	Total Frequency	50000
Frequency Missing	0	Observations	50000

表 5-3 各自变量对模型整体影响的显著性检验

Maximum Like lihood Analysis of Variance			
Source	DF	Chi-Square	Pr > ChiSq
Intercept	3	2373.38	<.0001
VAR18	3	345.41	<.0001
VAR19	3	18.71	0.0003
VAR25	3	364.87	<.0001
LikelihoodRatio	3	6.93	0.0743

上表给出了各自变量对模型的显著性检验,可以看到变量 VAR18、VAR19 和 VAR25 均通过了显著性检验。最后一行是对模型的拟合优度似然比检验,似然比统计量为 6.93,统计结果为不显著($p=0.0743$),这说明这一拟合模型是可以被接受的。

下表最后一列给出了每一个自变量对每一个 MNL 模型的显著性检验结果,其中除了 VAR18 的第二个系数估计值和 VAR19 的第三个系数估计值不显著以外,其余自变量的系数估计值均通过了显著性检验。

表 5-4 各自变量的估计值和显著性检验

Analysis of Maximum Likelihood Estimates					
Parameter	Function Number	Estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	-1.814	0.0511	1259.06	<.0001
	2	-1.5286	0.0453	1136.7	<.0001
	3	-1.0986	0.0383	823.74	<.0001
VAR18	1	-0.6591	0.0557	140.08	<.0001
	2	-0.0479	0.0477	1.01	0.3152
	3	-0.6547	0.0415	248.87	<.0001
VAR19	1	1.0529	0.3257	10.45	0.0012
	2	-2.2058	1.0106	4.76	0.0291
	3	-0.6481	0.5217	1.54	0.2142
VAR25	1	-0.1434	0.0654	4.8	0.0285
	2	-0.5337	0.0624	73.19	<.0001
	3	-1.0723	0.0591	329.76	<.0001

2.模型系数解释以及结论

在顾客选择香奈儿(Brand1=1)为参照系的前提下,对于7天无理由退货(VAR18)这一自变量来讲, $e^{-0.6591} = 0.52$ 表示淘宝卖家选择7天无理由退货这一产品属性时,顾客选择卡尔文·可雷恩的概率是未选择7天无理由退货这一产品属性时的概率的0.52倍;同样对于假一赔三(VAR19)来说, $e^{1.0529} = 2.87$ 表示淘宝卖家选择假一赔三这一产品属性时,顾客选择卡尔文·可雷恩的概率是没选择这一产品属性时的概率的2.87倍;最后,对于正品保障(VAR25)来说, $e^{-0.1434} = 0.87$ 表示淘宝卖家选择正品保障这一产品属性时,顾客选择卡尔文·可雷恩的概率是没选择正品保障这一产品属性的0.87倍。

通过表 5-4 的参数,这里得出 Multinomial Logit Model 为:

$$\begin{cases} \ln\left(\frac{p_4}{p_1}\right) = -1.8140 - 0.6591 * \text{VAR18} + 1.0529 * \text{VAR19} - 0.1434 * \text{VAR25} \\ \ln\left(\frac{p_3}{p_1}\right) = -1.5286 - 0.0479 * \text{VAR18} - 2.2058 * \text{VAR19} - 0.5337 * \text{VAR25} \\ \ln\left(\frac{p_2}{p_1}\right) = -1.0986 - 0.6547 * \text{VAR18} - 0.6481 * \text{VAR19} - 1.0723 * \text{VAR25} \end{cases}$$

与 $p_1 + p_2 + p_3 + p_4 = 1$ 联立得预测概率为:

$$\begin{cases} p_1 = \frac{1}{1 + A} \\ p_2 = \frac{e^{-1.0986-0.6547*\text{VAR18}-0.6481*\text{VAR19}-1.0723*\text{VAR25}}}{1 + A} \\ p_3 = \frac{e^{-1.5286-0.0479*\text{VAR18}-2.2058*\text{VAR19}-0.5337*\text{VAR25}}}{1 + A} \\ p_4 = \frac{e^{-1.8140-0.6591*\text{VAR18}+1.0529*\text{VAR19}-0.1434*\text{VAR25}}}{1 + A} \\ A = e^{-1.8140-0.6591*\text{VAR18}+1.0529*\text{VAR19}-0.1434*\text{VAR25}} \\ + e^{-1.5286-0.0479*\text{VAR18}-2.2058*\text{VAR19}-0.5337*\text{VAR25}} \\ + e^{-1.0986-0.6547*\text{VAR18}-0.6481*\text{VAR19}-1.0723*\text{VAR25}} \end{cases}$$

用以上模型对剩下 30% (13345 个) 的测试样本进行预测的结果如表所示:

表 5-5 模型预测的正确率

RESULT	频数	百分比
预测正确	8983	67.32

表格 5-5 说明,我们将剩下的 30%的数据代入模型对客户行为进行预测,正确率达到了 67.32%,说明该模型是基本可以被接受的。

5.3 上架策略的研究结论

通过以上分析，我们在上架时营销工具的选择上，可以得出以下结论：

一、在实际运营中费用偏高的淘客、旺铺、直通车等属性均未能通过显著性检验，说明这些服务的选择与否对顾客的选择不产生直接影响。

二、客户的关注点在假一赔三、正品保障和 7 天无理由退货三项属性，从一个侧面可以看出，顾客最在乎的，是买到的是否是正品，尤其是“假一赔三”属性最受重视。

三、A 公司在香水类目的产品在上架时的营销工具的选择上，可以先不做直通车、淘客等费用较高的淘宝工具的推广，将运营的重点放在正品保障方面，在页面设计的配合和产品推广的定位上，也应让客户全方面了解产品是“正品”为展示的重点。

四、由于数据来源限制，没有考虑客户的相关信息，以上模型及分析只能在较简单的情况下进行预测，对于加入平台属性、客户的地域信息、考虑自身作为卖家的属性等维度后的情况下的处理还有待加强。在 A 公司后期的实际运营中，关于提高转化率和增强客户黏性的问题中，均将有大量数据可用于构建模型，帮助分析决策。未来可通过对店铺的基础统计类数据的统计分析建模，利用数据挖掘对成交数据进行分析，从而对客户购买趋势进行预测，最终形成一套精准营销的理论体系。

第六章 总结与展望

6.1 全文结论及创新之处

6.1.1 全文结论

综合以上两章的分析结果，可以将结论归纳为以下两方面：

一、选品策略：

当店铺决定销售品牌香水时，应选择在平台销售了和市场占有率均偏高的淘宝平台销售；推广时，可将区域市场选择在以浙江、上海、广东、北京、香港、江苏等省市为主的地域；产品属性应选择 400 元左右的国产简装 50ml 正常规格的 EDT 淡香水产品；如销售知名品牌，以香奈儿、迪奥、范思哲和卡尔文克雷恩为首选。

二、上架管理：

根据建模分析的结果，在实际运营中费用偏高的淘客、旺铺、直通车等属性均未能通过显著性检验，说明这些服务的选择与否对顾客的选择不产生直接影响。顾客最在乎的是假一赔三、正品保障和 7 天无理由退货三项属性，从一个侧面可以看出，顾客最在乎的是买到的是否是正品，尤其是“假一赔三”属性最受重视。因此在上架时的营销工具选择上，可先不做直通车、淘客等推广，将运营的重点放在正品保障方面，推广时也应让客户全方面了解产品是“正品”为展示的重点。

6.1.2 创新之处

在电子商务中，数据挖掘有着广泛的应用，一方面可以通过数据挖掘研究客户的购买行为，另一方面也可以从历史的销售数据中挖掘交易规律。如何将其运用到小微电商的实际运营中去，是本文的重点。在本文的研究中，首先介绍了选题的背景及研究思路，接着对小微电商的运营现状做出了论述，并就其面临的问题以及现有解决途径的局限性做出了阐述，再通过对采集的实际交易数据进行处理与分析，从选品策略与上架管理两方面，对小微电商的运营提出了优化建议。

本文主要有三个方面的创新：

（1）国内大多数相关研究都是通过对调查问卷数据定量建模来对顾客的购买行为得出定性的结论，而本文是通过电商的实际交易记录来做定量的建模研究，得出定量的结论，使模型更贴近实际。

（2）与亚马逊之类的统一销售数据不同，本文是对淘宝这种电商联盟的交易数据进行建模分析，复杂度更高，结论更有意义。

（3）本文避开了模型建构复杂、专业化要求高的设计，选取了模型建构简单且在数据分析中得到广泛应用的 MNL 模型（Multinomial Logit Model）来定量构建预测模型并得出顾客的购买概率，并达到了较高的准确率，而非泛泛的定性分析，从而使得小微电子商务的商家可以更有针对性地进行选品与销售，既可以对电商的精准营销提供依据，又有很高的实际操作的可能。

6.2 展望

数据挖掘是一个新兴的交叉学科领域，具有非常广阔的应用前景。而电

子商务领域已过了快速发展的红利期，目前正处于平稳发展阶段，有数据显示，如今在淘宝平台上的运营商，仅有 5% 是不亏损的，其中有盈利的仅仅占到 2%，小微电商的盈利模式问题亟待解决。如何通过对数据挖掘技术的运用，强化对客户的服务、促进市场最优化、加速资金周转、实现企业的创新发展，是我们这一代电商人无可避免的责任。基于数据挖掘技术在电子商务平台上的运用有待更深入的研究，本文也还有许多问题需要解决，这需要在实际的工作应用中不断的进行积累和完善。在以下几个方面，还需要做进一步的研究和开发。

本研究在建立模型之初由于数据来源限制，没有考虑客户的相关信息，使得系统只能在较简单的情况下进行预测，对于加入平台属性、客户的地域信息、考虑自身作为卖家的属性等维度后的情况下的处理有待加强。可以在进一步完善此类信息后再次套用本模型进行筛选和测试，增强模型的适用性。

另外，由于本论文研究初始出发点是考虑新成立的电商或老电商的新部门，在没有历史销售数据作为参考的前提下，做出的建模分析，因此仅在引流阶段构建了模型。实际运营中，关于提高转化率和增强客户黏性的问题中，均有大量数据可用于构建模型帮助分析决策。后继研究可通过店铺的基础统计类数据的获得，对既有数据进行统计分析建模，最终形成利用数据挖掘对成交数据进行分析从而对客户购买趋势的预测，形成一套精准营销的理论体系。相信随着成本效益原则的日渐深入人心，数据挖掘技术也将不断地被推动着深入发展和广泛应用，进而创造出更高的经济价值和社会利益。

参考文献

- [1] V.Kumar, R. Venkatesan, and W.J.Reinartz . Knowing What to Sell, When, and to Whom. Harvard Business Review[J], March 2006
- [2] Carl F. Mela, Sunil Gupta and Donald R. Lehmann The long-term impact of promotion and advertising on consumer brand choice [J], Management Science, 2004. 50
- [3] V. Kumar, Rajkumar Venkatesan& Werner Reinartz. Performance Implications of Adopting a Customer-Focused Sales Campaign[J], Journal of Marketing Vol. 72 (September 2008)
- [4] Kenneth E. Train. Discrete Choice Methods with Simulation[M], Cambridge University Press, 2003
- [5] Karen E. Dynan. Habit Formation in Consumer Preferences: Evidence from Panel Data [J], The American Economic Review, 2000
- [6] Bert De Reyck Z D. MABS: Spreadsheet-based decision support for precision marketing[J]. European Journal of Operational Research, 2006, 171(3)
- [7] 王晓艳, 宋亦平. 中国年轻白领群体香水消费偏好的联合分析 [EB/OL]. 中国市场学会年会暨全国会员代表大会, 2006
- [8] 金贞花. 中国香水市场现状及消费者购买行为分析[N]. 延边大学学报(社会科学版), 2009
- [9] 王文璐. 基于用户大数据的海尔集团线上精准营销研究[M]. 东北农

业大学, 2015

[10]顾海斌. 基于大规模电商数据的用户消费行为分析方法研究[D]. 吉林大学, 2016

[10]芦文娟, 韩德昌. 网络营销模式下消费者行为研究综述[EB/OL]. 经济学动态, 2010

[11]彭兰, 基于电子商务环境的顾客消费偏好研究[N]. 上海师范大学学报(哲学社会科学版), 2011

[12]戴丽娜. 改革开放以来中国消费者研究探析[J]. 社会科学, 2012

[13]邬贺铨. 大数据时代的机遇与挑战[J]. 求是杂志, 2013

[14]朱建平等. 大数据时代下数据分析理念的辨析[J]. 统计研究, 2014(2)

[15]王梦雪. 数据挖掘综述[J]. 软件导刊, 2013(10)

[16]喻金平, 邱波. 数据挖掘技术在网络销售中的应用[J]. 软件导刊, 2007(23)

[17]樊明太. 2000 年度诺贝尔经济学奖得住郝克曼和麦克法登在经济学上的贡献[J]. 税务研究, 2001

[18]王济川, 郭志刚. Logistic 回归模型——方法与应用[M], 北京: 高等教育出版社, 2001

[19]刘枚莲等. 基于网络消费者偏好预测的推荐算法研究[J], 图书情报工作, 唐洁, 交通信息对出行者路径选择行为影响研究[D], 上海交通大学硕士学位论文 2010. 01

[20]刘鹰. 浅析我国电商企业的大数据应用现状[J], 商业时代, 2014(25)

[21]刘军, 吕俊峰. 大数据时代及数据挖掘的应用[J], 国家电网报,

2012 (010)

[22]王忠良. 再造营销: 大数据时代的广告营销变革[J], 销售与管理, 2012(12)

[23]王功辉. 国内移动商务研究定量分析[J]. 情报探索, 2010(10)

[24]施风芹, 张涛. 电子商务与网络营销[M]. 中国水利水电出版社, 2011

[25]张申. 浅析电子商务中数据挖掘的应用技术[J]. 天津商务职业学院学报, 2015(02)

[26]王珊, 王会举, 覃雄派, 周烜. 架构大数据:挑战、现状与展望[J]. 计算机学报, 2011(10)

[27]刘淑鑫. 面向大数据的网络用户行为分析研究[D]. 东华大学, 2015

[28]韩晶晶. 用大数据分析预测人类行为[J]. 科技创业, 2012(10)

[29]童楠楠, 朝乐门. 大数据时代下数据管理理念的变革: 从结果派到过程派[J]. 情报理论与实践, 2017(2)

[30]龚敏, 刘广丹. 基于大数据的精准营销应用研究综述[J], 2016(07)

[31]万红玲. 大数据时代下的精准营销[J], 2014(1)

[32]范苗苗. 电商数据勘探与挖掘研究[D], 2015

[33]郭捷, 李永壮. 基于离散选择模型的顾客选择偏好分析[J]. 技术经济与管理研究, 2012(8)

[34]张宏. 基于 Multinomial Logit 模型的产品线中消费者选择购买行为研究[M], 安徽农业科学, 2006

[35]王文彬, 李辉. 企业社会责任对消费者真实购买行为的影响研究--基于 Multinomial Logit 模型的实证分析[M]. 广西师范大学学报, 2012

[36]田莎. 关于用户选择移动通信运营商影响因素的研究——基于

Multinomial Logit 模型的实证分析[D]. 西南财经大学, 2012

[37]刘鹰. 浅析我国电商的大数据应用现状[J]. 商业时代, 2014(25)

[38]王步芳, 刘凤. 阿里开辟精准营销时代[J]. 企业管理, 2012(05)

致 谢

时光匆匆，短暂的 MBA 学习生涯就要结束了，过程中的美好让我在毕业之际充满了不舍。在重回校园的这段时光里，我不仅学到了丰富的管理类的知识，这些知识让我在经历过工作的实践后，对管理学的知识有了更深层次的认识。教授们的学术作风更是深深的感染了我，他们不仅有很高的学术水平，更难得的是他们务实的作风和复合型知识的应用手法，让我大开眼界，极大的拓宽了我的视野，开阔了我思考问题的角度，对我未来的工作和生活产生了及其深远的影响。

在这里我还结识了一群和我一起学习生活的伙伴们。对事业的追求和对知识的渴求将我们紧紧的联系在一起，在短暂的学习生涯中，我们结下了深厚的友谊，在这里我深深的感激学校给我的这个深造的机会。

我要郑重感谢我的导师陈梅梅教授。本文是在导师陈教授的精心指导下才能完成的。从选题，到开题的构思，论文结构的梳理，整体成文的写作，陈教授给了我悉心的指导。导师知识结构丰富，学术水平高深，学风严谨、视野宽阔、要求严格，其为人、品格、学识都让我心生敬仰，没有她的指导，我是无法顺利完成论文的。在撰写论文的过程中，我的导师给予了我无尽的关怀和帮助，让我不仅巩固了学到的知识，更通过对研究问题的思考，加深了我对工作的认识，对我将来的职业生涯产生了深远的影响。她是我的良师益友，也将成为我终身学习的楷模。这里我再次郑重的感谢我的指导老师，真的非常非

常感谢您对我的关怀与帮助！

我也要感谢我的 MBA 同学们，这两年间我们在学习和工作上都有了深入的交流，我们共同学习共同成长，这些经历是我们在学校完成 MBA 学习之外的重要补充。

感谢我的家人，有了他们的理解和支持，我才能顺利的完成 MBA 的学习。他们对我的理解、期望和鼓励是推动我不断前进的勇气和动力。

感谢所有关心帮助过我的人！在我完成了 MBA 课程的学习后，我将会努力将其运用到工作中去，回报学校，回报社会，回报所有关心、帮助我的人，谢谢！