

# 基于机器学习的贵州文创产品营销策略优化研究

徐倩

贵州大学数学与统计学院, 贵州 贵阳

收稿日期: 2026年3月10日; 录用日期: 2026年3月23日; 发布日期: 2026年5月29日

## 摘要

贵州拥有丰富的民族文化、自然景观和非遗资源, 这些得天独厚的资源为文创产品的开发、创新与传播提供了深厚且独特的支撑。本文以贵州文创产品消费者为研究主体, 首先采用网络爬虫技术获取消费者评论数据, 再利用SPSS软件对调查问卷进行可行性分析并获取有效回收问卷数据; 其次采用随机抽样方式对消费者性别、年龄、学历、收入、职业和婚姻状况等基本情况进行分析, 并采用Logistic回归模型对消费者购买频率的相关影响变量进行量化; 最后运用K-means聚类方法将消费者划分为核心消费群体、发展消费群体、潜在消费群体和低消费群体这4类群体, 并给出对应营销策略。研究结果可为贵州文创产品相关企业制定电商营销策略提供部分理论支撑与实践参考, 具有现实意义。

## 关键词

文创产品, 消费者行为, K-Means聚类, Logistic回归模型

# Optimization Research on Marketing Strategies of Guizhou Cultural and Creative Products Based on Machine Learning

Qian Xu

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: March 10, 2026; accepted: March 23, 2026; published: May 29, 2026

## Abstract

Guizhou has abundant ethnic culture, natural landscapes and intangible cultural heritage resources.

文章引用: 徐倩. 基于机器学习的贵州文创产品营销策略优化研究[J]. 电子商务评论, 2026, 15(5): 866-875.  
DOI: 10.12677/ecl.2026.155587

These unique resources provide profound and distinctive support for the development, innovation and dissemination of cultural and creative products. This paper takes consumers of Guizhou cultural and creative products as the research subject. Firstly, it uses web crawling technology to obtain consumer comment data, and then uses SPSS software to conduct feasibility analysis on the questionnaire and obtain effectively recovered data. Secondly, it adopts random sampling to analyze consumers' basic information such as gender, age, education, income, occupation and marital status, and uses Logistic regression model to quantify the relevant influencing variables of consumers' purchase frequency. Finally, it uses K-means clustering method to divide consumers into four groups: core consumer group, developing consumer group, potential consumer group and low consumer group, and puts forward corresponding marketing strategies. The research results can provide some theoretical support and practical reference for enterprises related to Guizhou cultural and creative products to formulate e-commerce marketing strategies, which is of practical significance.

## Keywords

Cultural and Creative Products, Consumer Behavior, K-Means Clustering, Logistic Regression Model

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

在国家文化自信建设持续推进与文化产业高质量发展的时代背景下，文创产品已逐步成为衔接传统文化资源与现代消费市场的核心载体。作为我国西南地区民族文化资源积淀最为深厚的省份之一，贵州省依托其多民族聚居的独特格局、禀赋优越的自然生态条件以及源远流长的历史文化底蕴，构建起具有鲜明地域辨识度的文化符号体系。

然而，当下贵州文创产业发展状况与消费需求之间依然存在着较为突出的矛盾[1]：1) 文创产品同质化严重，原创设计能力相对薄弱。部分企业在开发过程中，仅停留在对苗绣图案、银饰纹样或蜡染符号的简单拼贴与移植，缺乏对文化内涵、历史语境与精神价值的深入挖掘，未能实现从“符号借用”到“文化再创造”的转变[2][3]。2) 科技赋能浅层化，IP 打造能力欠缺。当前部分文创产品在引入数字技术时，往往停留在形式层面的叠加，例如简单加入二维码讲解、基础 AR 展示或线上展示页面优化等[4]。3) 营销手段单一，运营韧性不足。部分文创企业长期依赖旅游场景或节庆消费形成收入来源，销售周期具有明显的季节性特征。其营销模式也仍停留于传统路径，如线下门店陈列、节庆促销或单一电商平台销售[5][4]。

鉴于此，如何改变贵州文创产品在电商消费体系的现状，提高其综合竞争力具有十分重要的意义。本文围绕贵州文创产品消费数据展开研究，利用网络爬虫技术和问卷调查收集数据，通过 SPSS 软件、Logistic 回归模型、K-means 聚类等方法获取消费者偏好信息、识别消费者行为以及探讨消费者购买频率的影响因素。

## 2. 数据收集与分析

本文以贵州文创产品购买消费者作为研究对象，结合 python 语言进行数据挖掘，同时线上发放问卷得到相应反馈信息，进一步了解消费者的真实需求、消费偏好等重要数据资源。



评论数据中存在大量出现频率较高但含义不大的词语,为了得到有意义的分析结果,本文利用《哈工大停用词表》,在分词结果上过滤掉了类似于“我”、“在”、“的”、“都”等词和符号。

词云图可以直观地展示消费者对于贵州文创产品的评价,本文利用 Python 对预处理后的数据进行高频词提取,绘制了图 1 所示的词云图。在词云图中,词汇越大,说明该词出现的频率越高,消费者的关注度越高。可以观察到,消费者对贵州文创产品的关注高频词汇主要是“设计”、“质量”、“物流”等,说贵州文创产品的视觉设计、质量问题和物流快慢是消费者是否购买的主要原因。

## 2.2. 问卷调查分析

在完成问卷的设计后,需要对问卷的有效性进行预调查。设计的调查问卷主要有三个部分:基本信息、文创产品的市场容量、消费者购买认知。本文通过在线平台随机发放了 220 份问卷,回收了 204 份,有效问卷数为 174 份,有效问卷的回收率达到 85%。基于回收到的有效问卷数据,对问卷进行了信度和效度评估。

首先对问卷中的量表进行信度检验,采用 Cronbach 信度测量问卷项目的内在一致性系数,信度系数的取值范围为[0, 1]。Cronbach 信度测量公式如下:

$$\alpha = \frac{k}{k-1} \left( 1 - \frac{\sum S_i^2}{S^2} \right) \quad (1)$$

其中,  $k$  为量表的总题数,  $S_i^2$  为第  $i$  题得分的题内方差,  $S^2$  为全部题项得分的方差。信度系数越大,说明测量的可信度越大,本文利用 SPSS 软件计算 Cronbach's Alpha 系数作为评价标准,得到信度系数为 0.889,表明量表的可靠性高且信度好,问卷通过信度检验。

其次对问卷中的量表进行效度检验,效度分析的方法很多,这里采用常用的统计方法——因子分析法。通过 KMO 和 Bartlett 球状检验因子,进行结构效度分析,利用 SPSS 计算得到问卷的 KMO 取样适切性量数为 0.857, Bartlett 球形检验对应的 P 值小于 0.01,说明数据很适合做因子分析,问卷的结构设计良好。

在正式调查时,一共发放问卷 1158 份,其中有效问卷 1002 份。同样地,对回收问卷也进行了信度和效度检验。得到信度系数为 0.901, KMO 的值为 0.855, Bartlett 球形检验对应的 P 值小于 0.01。其中 KMO 检验和 Bartlett 球形检验具体结果见表 3。

Table 3. Table of KMO test and Bartlett's test of sphericity

表 3. KMO 检验和 Bartlett 球形检验表

KMO 值	-	0.855
	近似卡方	1272.884
Bartlett 球形检验	df	653
	P	0.000***

## 3. 消费者特征及行为分析

### 3.1. 消费者基本情况分析

本次调查以贵州文创产品消费群体为研究对象,采用随机抽样的方式开展调查,并借助问卷网站发布网络问卷进行数据收集。本节在对问卷数据进行整理与分析的基础上,总结消费者的基本特征与行为表现,并以此作为后续消费者群体划分的重要依据。

### 1) 消费者性别

从消费者性别来看,在本次调查中女性的占比为 53.38%,男性的占比为 46.62%,女性高于男性,其性别分布见图 2。这表明相比于男性消费者,女性消费者更喜欢网上购物,且精美传统的刺绣和银饰等非遗传文创产品更容易吸引女性朋友。

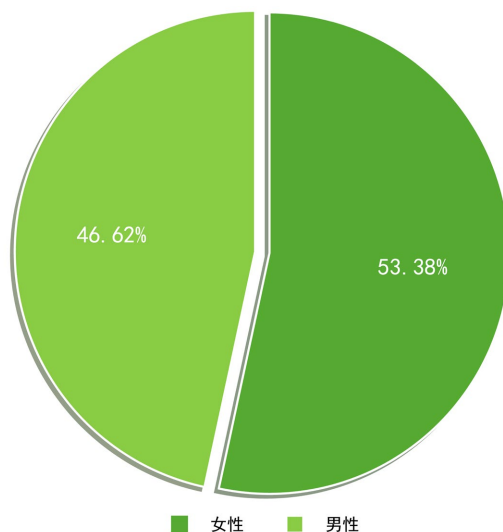


Figure 2. Consumer gender ratio distribution  
图 2. 消费者性别比例分布

### 2) 消费者年龄段

如图 3 所示,在本次调查中,26~35 岁的消费者占比最大,达到 30.05%,其次是 18~25 岁和 36~45 岁的消费者,占比分别为 21.42%、23.11%。18 岁以下、46~55 岁、56 岁以上的消费者占比分别为 11.36%、9.66%、4.39%。由于 18 岁以下人群不具备成熟的消费能力,所以本次研究主要针对 18 岁以上的消费群体。

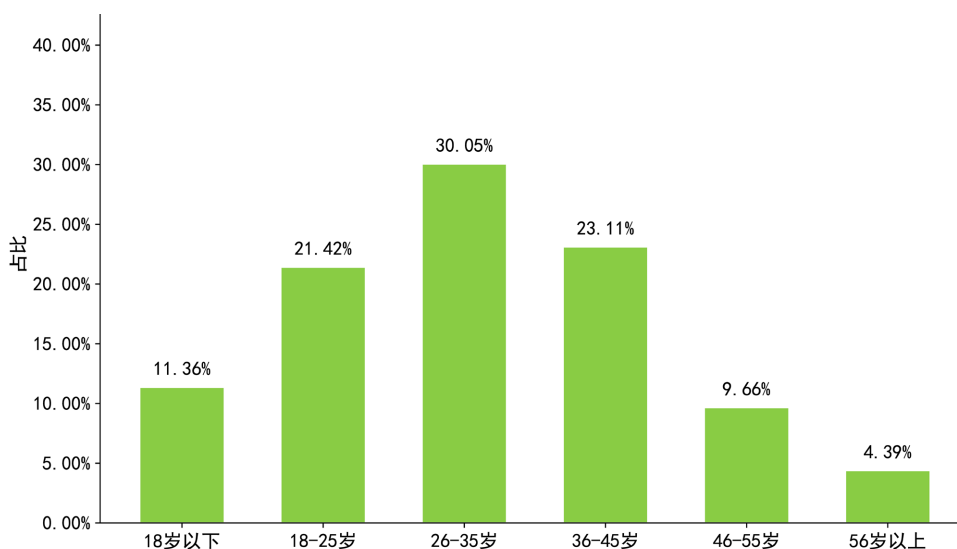
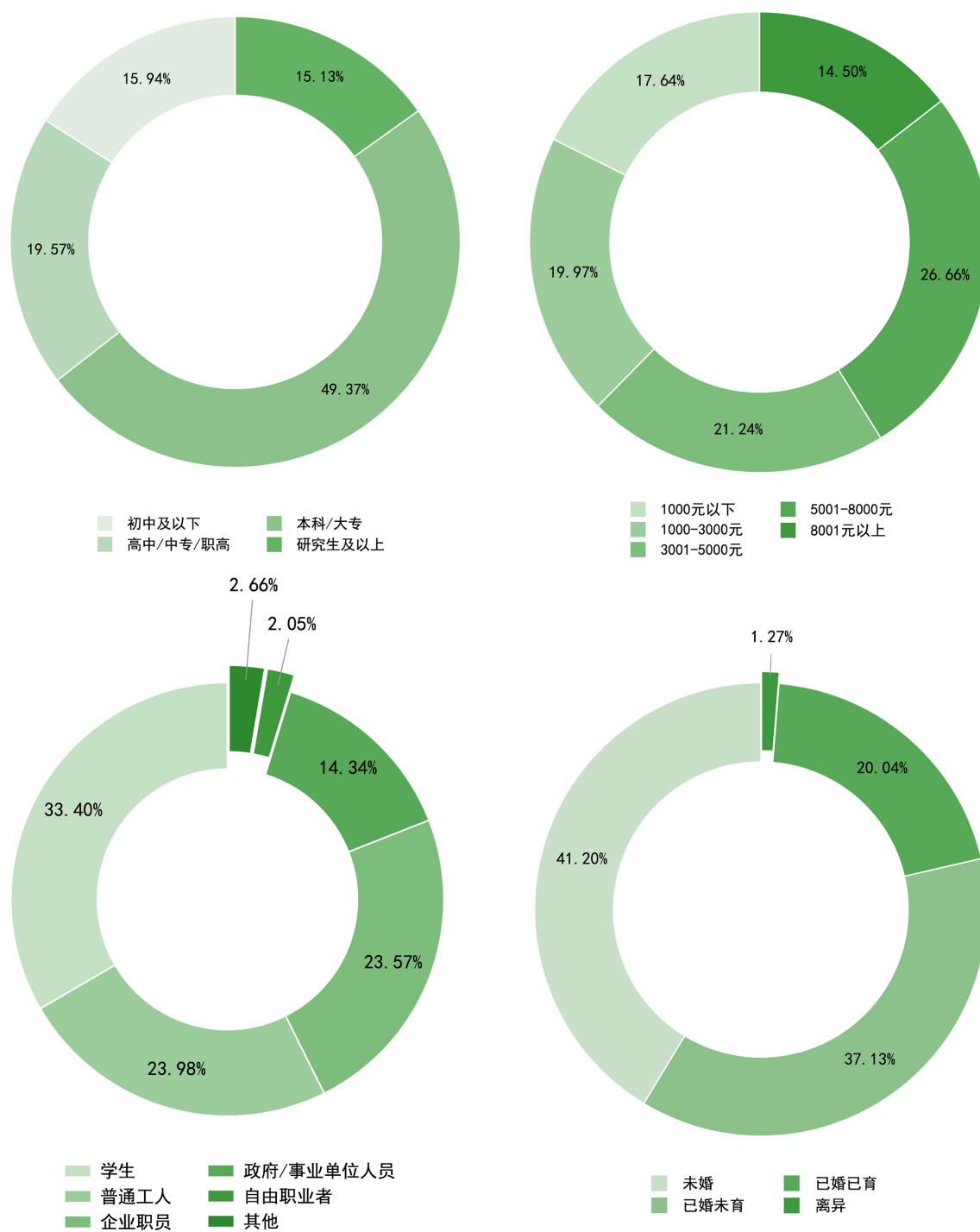


Figure 3. Consumer age ratio distribution  
图 3. 消费者年龄比例分布



**Figure 4.** Top left: Educational background of consumers; Top right: Income level of consumers; Bottom left: Occupation distribution of consumers; Bottom right: Marital status of consumers

**图 4.** 左上: 消费者学历分布; 右上: 消费者收入情况; 左下: 消费者职业分布; 右下: 消费者婚姻状况

### 3) 其他基本信息

图 4 展示的是消费者学历、收入、职业和婚姻状况等分布情况, 旨在全面分析不同消费者对贵州文创产品的消费现状和特征。从图中可以看出, 对于消费者的学历分布, 学历为本科/大专的占比最大, 为

49.37%，其次是学历为高中/中专/职高的受访者，占比为 19.57，初中及以下、研究生及以上分别占比为 15.94%、15.13%。对于消费者的月收入情况，占比最大的月收入区间是 5001~8000 (26.66%)，其次是 3001~5000 元区间，占比为 21.24%，占比最少的是 8001 元以上，为 14.50%。对于消费者职业状态分布，已经工作的人群占比最大，已经超过样本总体的一半，其次是学生群体，占比为 33.40%，因为调查中以工作人群和学生占比最多，故本次研究对象多为工作人群和学生。对于消费者婚姻状况，可以发现未婚人群所占比例最大，为 41.20%，已婚未育、已婚已育、离异人群占比分别为 37.13%、20.04%、1.27%。

综合分析可得，贵州文创产品的主要消费群体为学生以及年轻上班族，这两类人群合计占比超过 50%，其次高收入区间和高学历区间的占比高于其他群体，因为学生群体和年轻上班族群体时间相对自由且消费水平更高。最后关于不同婚姻状况对文创产品消费行为的差异，未婚人群更可能对文创类产品产生兴趣。

### 3.2. 消费者行为特征分析

Logistic 回归模型能够用于分析各类因素对消费者购买行为的影响强度，从而为消费者行为的解释与预测提供有效依据，并为企业制定更具针对性的市场营销策略提供支持[7]。在电子商务快速发展的背景下，对消费者购买频率的相关影响因素进行系统分析具有重要的现实意义。

本文聚焦消费者购买贵州文创产品频率的影响因素，建立多元 Logistic 回归模型。其回归模型公式如下：

$$\text{logit}(P) = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m \quad (2)$$

其中， $P$  为事件发生的概率， $\beta_0$  为模型的截距项， $\beta_1 \dots \beta_m$  为自变量系数。

本文将月均购买频率设定为响应变量，并以此划分消费者购买水平：当月均购买次数为 2 次及以下时定义为低购买频率，3 次及以上则定义为高购买频率。在模型构建中，选取性别(A)、年龄(B)、职业(C)、月收入(D)、选择种类(E)、产品质量(F)、设计创意(G)、价格合理性(H)、购买便利性(R)以及品牌(J)共 10 个因素作为解释变量。首先利用极大似然估计方法建立初始逻辑回归模型并进行分析；对于统计上不显著的变量，结合 AIC 准则与逐步回归方法对自变量进行筛选，从而对模型进行进一步优化，最终得到的模型结果如表 4 所示。

**Table 4.** Regression results of the Logistic model

**表 4.** Logistic 模型回归结果

自变量	回归系数	标准差	显著性	Wald
D	0.979	0.783	0.023	5.692
E	1.524	1.010	0.041	4.837
F	0.925	0.158	0.027	10.136
G	1.242	0.906	0.034	6.242
H	-1.944	1.493	0.016	3.195
R	0.611	0.522	0.041	7.581
J	0.486	0.338	0.047	2.773

根据模型回归结果，最终可以建立模型为：

$$\ln\left(\frac{P}{1-P}\right) = 0.472 + 0.979D + 1.524E + 0.925F + 1.242G - 1.944H + 0.611R + 0.486J \quad (3)$$

从上述模型结果可知,在本次调查中,月收入、选择种类、产品质量、设计创意、价格合理性、购买便利性、品牌这几个变量的显著性均小于 0.05,在显著性水平为 0.05 的情况下均为显著,效果较好。结合实际电商经济特点,可以得出:1) 在如今互联网电商快速发展的时代,消费者的性别、年龄和职业对文创产品的购买频率影响不是特别显著。因为电商平台打破了传统的消费方式,给不同性别、年龄、职业的消费者提供了更公平和便利的购买方式和渠道。2) 贵州文创产品在电商平台下可提供的选择种类、价格是否合理、品牌知名度对消费者购买频率影响较大。丰富多样且精美的文创产品更能吸引消费者的目光,如果只是单一且同质化严重的文创产品,容易审美疲劳;合理的价格也是留住消费者的一个重要因素;打造出一些相对知名的品牌,给到消费者关于贵州文创的理念,也能促进消费者购买。

#### 4. 消费者群体细分及营销策略

本文针对消费者对贵州文创产品的消费金额和购买频率,运用聚类分析方法对消费群体进行细分,以达到更精准地洞察各类群体的需求特征,这对贵州文创产品在电商场景下实现精准营销、增强市场竞争力具有重要的现实意义。

##### 4.1. 基于 K-Means 聚类算法的消费者细分

随着电商行业的迅速发展,文创产品市场规模持续扩大。为系统探究不同消费群体对文创产品的认知与态度,本文基于问卷调查数据展开分析。采用 K-means 聚类算法[8],根据消费者基本特征对样本进行分类,在最大化组内同质性的同时增强组间异质性,以此精准识别并刻画不同消费群体,也在一定程度上简化了数据分析过程。

本文基于问卷数据,选取文创产品购买频率与月均消费两个核心变量,均采用 4 分制有序赋值:购买频率 2 次及以下、3~6 次、7~9 次、10 次及以上依次赋值为 1~4 分;月均消费 50 元及以下、51~100 元、101~150 元、151 元及以上分别赋值为 1~4 分,以此量化消费者购买行为特征与消费水平。随后对调查者的总消费金额和购买频率使用 Python 语言进行聚类分析,可将消费者群体分为四类:核心消费群体、发展消费群体、潜在消费群体和低消费群体,最终结果见表 5。

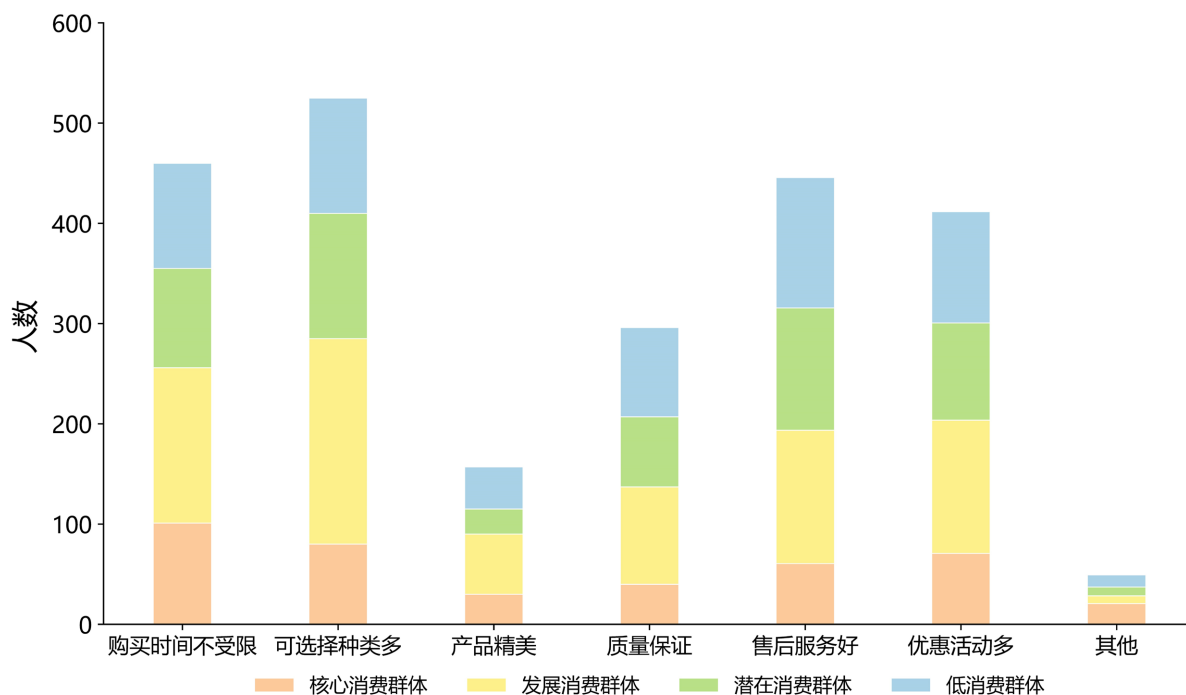
Table 5. Consumer segmentation characteristics

表 5. 消费者分类特征

名称	核心消费群体	发展消费群体	潜在消费群体	低消费群体
总消费金额	高(4 分)	高(3)	中等(2 分)	低(1 分)
购买频率	高(4 分)	中频(3 分)	中频(2 分)	低(1 分)
人数	158	296	237	271

根据聚类结果,并结合消费者的基本情况特征分析可知:核心消费群体以 18~25 岁青年为主,普遍拥有本科及以上学历,月收入与消费支出水平较高,喜爱购物、娱乐,乐于追逐潮流与新鲜事物。发展消费群体年龄集中在 26~35 岁,具有高学历、高收入、职业稳定等特征,多已婚已育,作为家庭消费主力,文创产品消费潜力突出。潜在消费群体以 36~45 岁人群为主,月收入处于中高水平,大多已婚已育,可支配时间充裕,具备较强的潜在消费能力。低消费群体多为自由职业者与临时工作人员,年龄偏大、学历与收入水平相对较低,消费观念偏传统保守,对潮流产品接受度不高,对贵州文创产品的消费意愿与潜力相对有限。

进一步,本文通过统计不同类型消费群体购买贵州文创产品的原因,得到如图 5 所示的结果。



**Figure 5.** Analysis diagram of purchase reasons by different consumer groups  
**图 5.** 不同消费群体购买原因分析图

由图 5 可知, 购买时间灵活、产品种类多是四类消费群体购买文创产品的主要原因: 核心消费群体看重优良服务及售后服务, 可通过优质售后巩固其忠诚度; 发展消费群体对优惠活动兴趣显著; 潜在消费群体与低消费群体均关注产品质量, 需在保质基础上制定合理价格(低消费群体对价格尤为敏感)。

#### 4.2. 不同类型消费者体验优化策略

通过深入分析不同消费群体的特征与偏好, 针对贵州文创产品提出差异化营销策略:

1) 针对核心消费群体, 以优质服务强化忠诚度。完善售后体系对提升复购与粘性至关重要, 建立高效响应机制, 及时处理咨询与售后问题, 保障消费体验; 定期收集消费者反馈, 持续优化产品与服务质量, 同时简化购买流程、提升平台使用体验。

2) 针对发展消费群体, 以优惠活动刺激消费。推出文创组合套餐与灵活定价, 结合节日促销、会员折扣等方式提升吸引力; 建立积分与分享激励机制, 通过好友邀请、积分兑换文创周边等形式, 提高购买频次。

3) 针对潜在消费群体, 以体验式服务降低入门门槛。推行无风险售后政策, 配合首单优惠、新人专享价等方式, 降低初次消费成本, 吸引观望人群尝试购买。

4) 针对低消费群体, 以轻量化方案激活需求。推出小规格、低单价的精美文创产品, 降低决策成本; 基于浏览与历史订单进行精准推送, 发放专属优惠券与复购福利, 唤醒沉睡消费者并逐步培养消费习惯。

### 5. 总结

本文调查了消费者购买贵州文创产品的现状, 通过网络爬虫技术获取电商平台用户的评论, 再根据问卷调查获取研究数据, 随后利用 SPSS 软件、Logistic 回归模型、K-means 聚类等方法展开研究。根据结果分析可以发现贵州文创产品的主要消费群体集中在学生和年轻上班族, 他们时间相对自由且消费水

平更高；将消费金额、购买频率作为依据把消费者划分为四类群体，并给出了差异化营销策略。同时为贵州文创产品相关企业制定电商营销策略提供理论支撑与实践参考，也丰富了文创电商领域消费者行为研究的范式。

## 参考文献

- [1] 贵州省文化和旅游厅. 贵州特色文化创意产品繁花渐放[EB/OL]. 贵州省文化和旅游厅官网. [https://whhly.guizhou.gov.cn/xwzx/tt/202506/t20250613\\_88135317.html](https://whhly.guizhou.gov.cn/xwzx/tt/202506/t20250613_88135317.html), 2025-06-13.
- [2] 倪鹏飞, 王一冉. 符号消费理论视阈下的贵州民族文创产品设计研究[J]. 工业设计, 2026(2): 122-125.
- [3] 陈舒, 刘黎. 乡村振兴视域下贵州非遗文创发展路径研究[J]. 天工, 2022(24): 57-59.
- [4] 辛路娟. 互联网+视域下非遗文创产品设计策略研究[J]. 天工, 2022(21): 24-26.
- [5] 唐慧娟. 基于消费心理学的品牌塑造与营销——以图书馆文创产品为例[J]. 现代阅读, 2025, (12): 13-15.
- [6] 黄强. 基于文本挖掘的汽车用户需求分析[J]. 汽车与新动力, 2024, 7(1): 80-84.
- [7] 黄依伊, 肖志坚, 钱桢华. 电商平台用户购买行为的 Logistic 回归模型分析——以淘宝平台商贸数据为例[J]. 商业经济, 2025(8): 83-86.
- [8] 郑可莉. 基于用户画像与 K-means 算法的电子商务精准营销策略研究[J]. 贵阳学院学报(自然科学版), 2025, 20(2): 84-89.