

“正大杯”第十四届全国研究生市场调查  
与分析大赛陕西省选拔赛

解码短视频时代的秘密—  
“短视频+电商”模式效果及发展趋势浅析

调查报告

团队成员：宋帆、郭永强、安琨谊、黄思烨、孙怀玲

## 摘要

在自媒体时代的快速发展的背景下,短视频已经成为了一种新的内容形式和传播渠道,对于传达信息、吸引观众、产生共鸣以及商业化价值等方面都具有重要意义。本文主要针对短视频内容对消费者产生的影响进行调查,从多角度出发,结合消费者的行为、态度、对短视频的满意度以及通过短视频购物的意愿,利用多种模型与统计方法进行分析,并提供合理、可行的对策和建议。

首先,本次调查采用分层随机抽样和滚雪球的调查方法收集问卷,在网络上按地区分层随机发放问卷,通过网络进行回收,有利于大范围的数据收集。在正式调查中,共计收回 1067 份问卷(即通过检验的问卷)。对其调查数据进行信效度检验,以确保调查数据的准确性和可靠性。

其次,对调查得到的数据进行处理和分析。对调查数据进行描述性分析,得到了被调查者的短视频基本使用信息;对调查对象的职业与短视频使用时间、短视频的整体满意度与短视频使用时间分别进行交叉列联表分析与卡方检验,结果表明职业、满意度与短视频的使用都有一定的关联性;此外,对月收入与月均网上消费水平进行了相关性分析,结果显示月收入与网上消费有一定的相关性,接着对学历与直播间商品类型进行卡方检验,分析得到学历与购买的线上商品类型具有明显相关性;进一步,利用随机森林模型研究短视频内容对消费者行为和态度的影响分析,并采用基于 ANOVA F 值的 SelectKBest 方法计算各变量的 F 值作为评分,与随机森林重要性进行比较,结果表明,消费者对短视频中活动的参与度和消费者观念想法的改变是导致消费者购买决策的主要原因,与随机森林模型得分结论一致。最后,构建多元线性回归模型,探究用户消极使用电商平台行为的原因,分析得到消极行为、负面内容、使用习惯、隐私顾虑以及平台质量之间存在密切关系。

通过以上分析,本文最后给出一些结论与建议。结论:第一,职业与整体满意度都是短视频使用频率的影响因素;第二,大多数人对短视频的整体满意度较高,短视频与电商平台的结合如直播带货的方式带动了人们的消费水平,为我们的生活提供了便利;第三,通过分析消费者选择购买决策的原因,我们发现消费者对短视频中活动的参与度和消费者观念想法的改变是主要影响因素;第四,使得用户消极使用电商平台的主要原因是隐私顾虑、负面内容以及平台质量。建议:第一,优化短视频的内容与创意;第二,加强平台与政府监管力度;第三,用户要加强自我约束,合理利用短视频平台;第四,消费者要理性消费,养成良好的消费行为和消费习惯。通过以上结论与建议,为我国短视频产业的规范提供有益参考,促进其健康、可持续地发展。

**关键词:** 短视频 消费行为与态度 随机森林模型 多元回归模型 方差分析

# 目录

图表清单 .....	V
一、 绪论 .....	1
(一) 选题背景 .....	1
(二) 文献综述 .....	2
(三) 研究过程 .....	3
二、 调查方案的设计 .....	4
(一) 调查对象的选择 .....	4
(二) 调查方法 .....	4
(三) 调查样本量的确定 .....	4
(四) 问卷设计 .....	5
1. 主要调查内容 .....	5
2. 数据类型的编码 .....	5
3. 调查质量控制 .....	5
(五) 调查的组织 .....	6
(六) 调查数据的处理 .....	6
1. 编码 .....	6
2. 分组 .....	7
3. 测量 .....	8
(七) 数据的质量诊断与评估 .....	8
三、 调查数据的分析 .....	9
(一) 信效度检验 .....	9
1. 信度检验 .....	9
2. 效度检验 .....	9
(二) 调查数据的基本描述性统计 .....	10

(三) 短视频的基本使用情况的特征分析 .....	10
1. 基于短视频的基本使用情况的分析 .....	10
2. 短视频的评价与消费者购买意愿 .....	13
3. 短视频与电商平台的结合 .....	15
4. 短视频未来发展优缺点分析 .....	18
四、短视频内容对消费者影响因素分析 .....	18
(一) 短视频内容对消费者行为和态度的影响分析 .....	18
1. 模型变量选择及相关性分析 .....	18
2. 随机森林模型 .....	19
3. 基于 ANOVA F 值的 SelectKBest 法 .....	22
(二) 消极行为的影响因素研究 .....	23
1. 模型理论 .....	23
2. 变量选择与解释 .....	24
3. 建模分析 .....	27
4. 模型的进一步分析 .....	28
五、 结论与建议 .....	29
(一) 结论 .....	29
1. 短视频对用户的生活与学习有一定的积极影响 .....	29
2. 短视频+电商的形式促进了线上消费的发展 .....	30
3. 短视频的消极影响 .....	31
(二) 建议 .....	32
1. 优化短视频的内容质量与创意 .....	32
2. 加强平台与政府监管力度 .....	32
3. 用户应加强自我约束, 理性对待短视频 .....	32
参考文献 .....	34

附注 1 .....	35
附注 2 .....	43

## 图表清单

图 1 本文流程图 .....	3
图 2 调查对象的性别分布 .....	10
图 3 调查对象的学历分布图 .....	10
图 4 不同职业/目标下男女在社交媒体平台上的用户比较 .....	11
图 5 短视频内容对用户的影响意愿评分 .....	15
图 6 线上消费的原因分析图 .....	17
图 7 词云图 .....	18
图 8 相关性热力图 .....	19
图 9 随机森林的步骤流程图 .....	21
图 10 随机森林决策树的变量图 .....	21
图 11 归一化重要性与变量的关系图 .....	22
图 12 (a)以 F 值为评分的雷达图；(b)以归一化后的 F 值为评分的雷达图 .....	23
图 13 自变量的具体细节 .....	25
图 14 降维后的数据散点图 .....	25
图 15 矩阵散点图 .....	26
图 16 残差直方图和 P-P 图 .....	28
表 1 各个成员的任务分配 .....	6
表 2 问题编码 .....	6
表 3 问卷信度分析 .....	9
表 4 KMO 和巴特利特检验结果 .....	9
表 5 职业与每天使用短视频时间的卡方检验 .....	11
表 6 职业与每天在短视频上花费的时间的交叉列联表 .....	12
表 7 职业与最常观看内容类型的卡方检验 .....	12
表 8 职业与最常观看内容类型的交叉列联表 .....	13
表 9 短视频的整体满意度与每天观看短视频的时间的卡方检验 .....	14
表 10 短视频的整体满意度与每天观看短视频的时间之间的交叉列联表 .....	14
表 11 月收入与月均网上消费水平进行相关性 .....	16
表 12 学历与直播间购买商品类型的卡方检验 .....	16
表 13 不同学历选择在直播间购买的商品类型的差异 .....	16
表 14 短视频内容对消费者行为和态度的影响因子 .....	18
表 15 模型的误差值 .....	22
表 16 变量名称与编号 .....	24

表 17 处理后的数据 .....	25
表 18 回归系数 .....	27
表 19 F 检验结果 .....	27
表 20 自变量的 $p$ 值 .....	28
表 21 VIF 值 .....	28

# 一、绪论

## （一）选题背景

短视频，即短片视频，是一种互联网内容传播方式，在移动终端普及和网络提速的背景下逐渐受到人们的青睐。根据第四届中国短视频大会发布的《中国短视频发展研究报告（2023）》显示，近7年来，网络视听产业规模增长了7.18倍，而短视频产业规模增长了153倍，占比增长了17.8倍。2018年至2022年，发布过自制短视频的用户比例从28.2%攀升至46.9%，我国有近4.75亿短视频用户参与了内容创作，短视频成为普通人记录生活和表达自我的重要形式。中国短视频用户规模已从2018年12月的6.48亿增长至2023年6月的10.26亿，用户使用率从78.2%增长至95.2%。目前，中国短视频产业规模已接近3000亿元。随着产业规模的持续扩大和生态的逐步成熟，短视频产业正成为舆论宣传的重要阵地和视听产业发展的主要引擎。

随着短视频用户规模的不断扩大和市场范围的增加，推动了新业态和新模式的不断涌现。短视频内容与电商的融合加速，各平台从内容引流到电商营销持续不断。此外，短视频内容的多样化和不断融入各行业领域，也推动了品牌营销的创新。对于品牌主来说，短视频不仅带来了庞大的受众群体和商业潜力，还提供了丰富多样的营销内容，使其能够更有效地连接目标受众。然而，高昂的成本问题，包括内容生产、传播和用户获取成本，制约了短视频营销的盈利能力。持续、规模化的内容生产是维持用户活跃度和创造更多粉丝付费可能性的关键。针对这一问题，各类媒体平台开始关注消费者的心理，聚焦内容生产，打造优质内容。优质内容的生产成为各平台价值的主要表现形式，也增强了用户的付费意愿。因此，通过分析用户的消费心理和了解其需求，探索优质内容变现模式，促使用户付费意愿持续提升显得尤为重要。

与传统营销相比，短视频营销具有明显的优势。基于大数据和人工智能，短视频平台不断收集和分析用户使用数据，根据用户偏好进行个性化内容推荐。然而，短视频广告的监测和效果评估仍存在挑战，缺乏完善的监测系统。尤其是在多平台投放和多视频达人制作的情况下，广告主难以全面把控投放时机、渠道选择和创意优化，对短视频营销的影响不利。因此，根据短视频数据分析消费者的消费心理，并以此为基础制定营销策略，是企业 and 产品品牌宣传的关键。

调查中，为了兼顾考虑多个目标变量的相似性以及同层单元在空间上的邻近性。我们采用分层随机抽样方法。并对抽取的样本进行信效度检验和描述性分析，保证调查数据的有效性。对数据进行分析时，我们通过统计和数学模型建立了完整的数据分析体系。



综上，本次调查项目针对短视频用户的心理进行调查研究，针对短视频的使用现状，深入探究了短视频内容对用户的行为和态度的影响以及短视频所带来的社会影响。并针对短视频所带来的影响，提供合理的市场对策与建议。

## （二）文献综述

近几年，许多学者对短视频现状做了大量研究。蔡丽蓉（2024）分析了抖音短视频平台的运营现状和运营环境，并从消费者、品牌、队伍和信息技术等层面，提出了适合抖音平台的新媒体品牌营销策略<sup>[1]</sup>。马佳音（2024）分析了在5G和疫情的大背景下，短视频行业稳步发展，抖音和快手竞争激烈，头部效应凸显；功能不断扩展，覆盖用户生活各方面；AI技术赋能内容生产，重心转向内容增长；短视频与直播紧密结合，本地服务功能增强，迎来发展下半场<sup>[2]</sup>。李欣颖（2023）探讨短视频平台用户受短视频信息质量影响在该平台上的购物行为，有助于短视频平台发展电商业务<sup>[3]</sup>。张莹（2024）分析互联网给农产品销售带来的有益影响，结合农产品在互联网渠道的营销现状，总结“互联网+农产品”营销的优势、问题和改进的策略<sup>[4]</sup>。孔驹等人（2024）详细分析了抖音电商的营销方法，探究其对大学生消费心理的影响<sup>[5]</sup>。董肖宇（2024）提出了一种新的时尚短视频内容策略，达到提高短视频知名度、增加时尚产品销量的目的<sup>[6]</sup>。蔡军和曹亚方（2024）从盈利模式的构成要素角度，对其利润对象、利润源、利润屏障和利润杠杆进行梳理；分析快手盈利模式的优点及问题，发现快手盈利模式存在内容低俗化、同质化的问题<sup>[7]</sup>。王温青等人（2023）依据山东省和河北省836份果品和药材种植户样本数据，采用内生转换回归模型验证了短视频运营对农户特色农产品种植收入的影响<sup>[8]</sup>。

### （三）研究过程

为了更加清楚的展示本文的调查研究过程，设计了如图 1 所示的流程图。

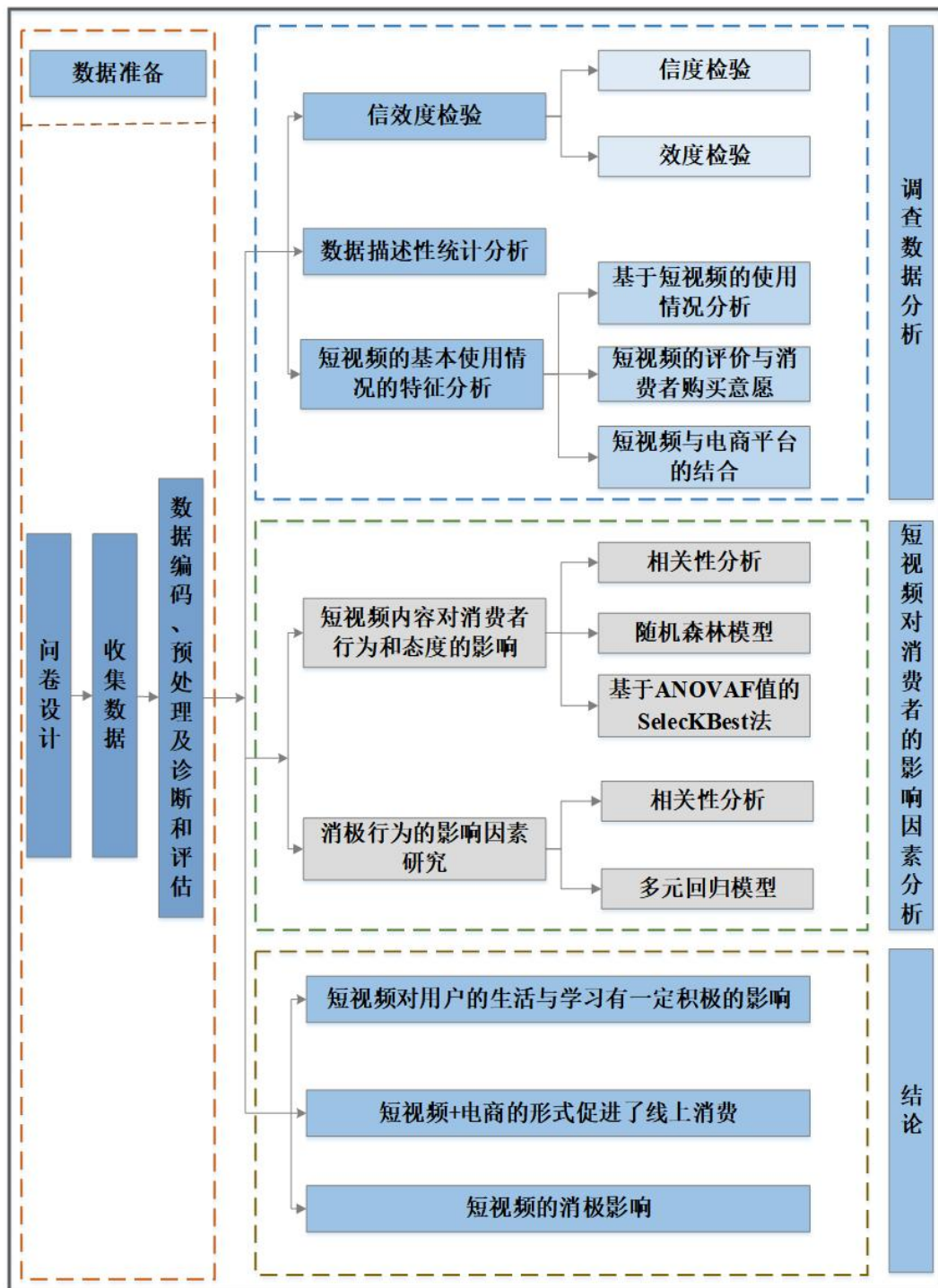


图 1 本文流程图

## 二、调查方案的设计

### （一）调查对象的选择

本次问卷主要面向不同性别、不同学历和不同职业的社会群体。考虑到不同人群对短视频的观看偏好的差异性,我们设置了调查对象的年龄和相关问题的内容,以确保调查对象的有效性和代表性。

### （二）调查方法

鉴于经济和人力的限制,我们采取了分层按比例抽样的调查方法。为了提高样本的代表性和数据的质量,我们通过网络发布问卷的方式进行地区分布,并通过网络进行发布和回收,以便更高效地收集数据。

### （三）调查样本量的确定

为了确保样本能够代表总体,选取适当的样本量是必要的。不同的抽样方法有不同的计算公式。一般情况下,样本大小主要取决于:

- (1) 研究对象的变化程度,即变异程度;
- (2) 要求和允许的误差大小,即精度要求;
- (3) 要求推断的置信度;
- (4) 总体的大小;
- (5) 抽样的方法。

为保证样本选择的客观性和随机性,我们采用了分层随机抽样的方法。这种方法可以确保每个层次的样本数量与其在总体中的比例相匹配,从而保证了抽样的代表性。分层随机抽样的计算公式涉及确定每个层中抽样的样本数量。假设总体中有  $N$  个单位(个体、样本等),总共分为  $K$  个层次,其中第  $i$  个层次有  $N_i$  个单位。设想在第  $i$  层中需要抽取个  $n_i$  个单位作为样本如果要在每个层次中的样本比例与总体中的单位比例相同,则可以按照公式(1)计算每个层次的样本量。

$$n_i = \left(\frac{N_i}{N}\right) \times n \quad (1)$$

其中,  $n_i$  是第  $i$  个层次的样本数量,  $N_i$  是第  $i$  个层次的单位数量,  $N$  是总体中单位数量,  $n$  是要抽取的样本总量。这样计算出来的  $n_i$  可以确保每个层次的样本数量与其在总体中的比例相匹配,从而保证了抽样的代表性。

鉴于受访者的收入水平可能反映其经济实力和消费能力,不同收入水平的人可能在短视频内容的需求和购买行为上存在差异。因此,我们针对被调查者的月收入进行了分层随机抽样,以确保问卷样本的代表性。最终,我们选取了 1067 份有效代表样本。

## （四）问卷设计

本次调查的问卷结构主要分为三个部分,共设有三十五个题目,采用结构化设计。内部题干之间的逻辑验证等措施确保了数据的一致性,同时防止了受调查者将其个人主观偏见带入题目中,从而更好地提高了数据的有效性和真实性。这三个部分分别涉及短视频用户的基本使用情况、用户对短视频的评价以及直播带货的影响。

### 1. 主要调查内容

为了准确采集受访者的个人信息,问卷设置了个人性别、学历、职业以及个人的月收入等问题。这有助于更好地了解不同性别、学历、职业的社会群体对短视频的使用情况、习惯、偏好以及对虚拟现实或增强现实技术的兴趣等方面的差异。同时,通过调查被调查者在直播带货中购买的商品类型、购买支出以及线下消费与直播带货花销的对比,可以对被调查者在直播带货中的购买行为、态度和看法等方面进行综合分析。为了全面了解被调查者对短视频内容的满意度、态度和评价,以及他们认为存在的问题,问卷还设置了评价等级、存在问题选择以及简述建议等问题。这有助于了解受访者对短视频内容的情绪波动、影响程度、购买意愿以及对内容真实性的态度等方面。

### 2. 数据类型的编码

为了满足分析的需要,我们对每一个问题数据进行编码,得到数值型数据,便于进行数据分析。对于需要进行评价和建议的问题,我们采用了字符型数据,这样可以清晰地展现被调查者的观点和建议,确保数据的准确性和全面性。

### 3. 调查质量控制

为了确保数据质量,我们需要对收集的数据进行质量控制。例如,我们可以检查调查对象的基本情况是否与正常逻辑相符,被调查者填写的内容是否存在前后矛盾,以及收集的数据是否有缺失等情况。通过对这些问题的筛选,我们可以得到最终有效的数据。

本次问卷设计的原则是确保数据收集的真实性、准确性、一致性和全面性。通过上述阐述,我们可以确保本次问卷设计是基本可行的。

（五）调查的组织

调研前期通过查阅大量资料和参考文献，进而确定影响短视频用户使用的相关因素，确定相关指标，并依据指标制定问卷。通过问卷调查法向社会群体发送问卷进行预调查，根据预调查结果，参考调查对象想法，并结合指导教师意见，小组成员召开会议商讨后确定正式问卷。

正式发放问卷从 2024 年 1 月 5 日开始，持续到 2024 年 1 月 20 日结束，共计耗时 16 天。调查结束后，我们查看了被调查者填写问卷所用的时间。若填写时间少于 1 分钟，我们将认为该被调查者随意填写问卷，将其问卷标记为无效。然后，按照上述的抽样方式抽取下一个样本。

问卷收集结束后，我们对数据进行处理，筛查异常值和缺失值，并对数据进行整合和分析，以得出结论。

表 1 各个成员的任务分配

成员	任务分配
安坤谊	设计问卷，数据分析，撰写报告
孙怀玲	收集问卷，数据处理，撰写报告
黄思烨	撰写报告，PPT 制作
宋帆、郭永强	数据分析与建模，撰写报告

（六）调查数据的处理

在数据收集完成后，我们需要通过合理的方式对其进行编码、分组、测量、数据质量核查、调整样本结构以及筛选。

1. 编码

为了方便后续分析，我们对收集到的对象进行编码。每个编码符号代表一条信息或一串数据，以实现数据的编码。通过自定义变量名称并在原始数值的基础上进行替换，我们对数据进行了重新编码，从而改变了原始数据的表示形式。编码的目的在于识别每一个记录，区分处理方法，进行分类和校核，以使数据尽可能排列整齐。例如，在我们获得的数据中，我们为每份问卷进行了编号，这样做可以克服不同群体数据参差不齐的问题，同时节省时间、提高速度，并且便于后续的数据分组。

表 2 问题编码

Q1	性别
Q2	学历

Q3	职业
Q4	月收入
Q5	您是否自己制作过短视频，您制作短视频时的想法是什么
Q6	如果您使用短视频软件，您一般使用哪些平台
Q7	您一般在什么时间段使用短视频软件
Q8	您一般在什么时间段使用短视频软件
Q9	您使用短视频软件的主要目的是什么
Q10	您在短视频平台上更喜欢哪种互动方式
Q11	您最喜欢观看哪些类型的短视频内容
Q12	您对将来在短视频中使用虚拟现实（VR）是否感兴趣
Q13	您关心以下哪些方面的个人隐私和数据安全问题
Q14	您每个月在直播带货中购买的商品类型主要为
Q15	您每个月在直播带货中的购买支出大致为
Q16	直播网购花销占您网购总花销的比例大约为
Q17	您接触网络带货直播后您的网购消费额较之前增多的程度
Q18	网络带货直播使您减少线下消费，增加线上消费的程度
Q19	您选择在直播间购买商品的主要原因是
Q20	在观看直播带货时，您最关注的是
Q21	您在直播中购买过一次商品后，回购的意愿
Q22	您觉得直播带货对消费者来说最主要的好处是什么
Q23	您认为直播带货的主要弊端
Q24	您拒绝在直播间购买商品的主要原因
Q25	您认为当下电商直播首要应该做出哪些改进
Q26	您对直播带货的看法
Q27	您对直播带货的未来看法
Q28	您因为短视频的内容而产生情绪波动
Q29	您会因为短视频的内容而改变自己的想法或观念的程度
Q30	您会因为短视频的内容而购买相关的商品或服务的意愿
Q31	您是否会因为短视频的内容而参与相关的活动或话题的意愿
Q32	您对短视频中内容的真假、是非所持的态度
Q33	您对短视频的整体满意度是多少
Q34	您认为短视频有哪些缺点需要改进
Q35	您对短视频未来的发展有什么建议或期待

## 2. 分组

为了满足统计研究的需要，我们将原始样本数据按照一定的标准划分成不同的组别，这样处理后的数据称为分组数据。通过拆分文件，我们可以明确分组的依据，并按照这些组别来组织输出。分组的依据可以是性别、职业、月收入等。例如，我们可以根据被调查者的性别将样本数据分为男性和女性两组；根据被调

查者的月收入情况将样本分为六个组别；或者根据被调查者的职业将样本分成九组等。经过分组处理后的数据具有更高的直观性、可读性和分析性。

### 3. 测量

对数据进行分组后，我们分别进行名义测量、次序测量、间距测量和比率测量。数据的质量要求一般包括测试需求的满足、覆盖被测范围和测试边界，以及数据的完整性、一致性和合理性等方面。

**Step1:** 使用卡方分析来测量不同分类之间的关联程度，并使用线性方程和回归模型对不同分组进行参数化估计。

**Step2:** 使用方差分析和均值检验等方法对有序分类的数值进行数据分析，以获得更高的顺序性。

**Step3:** 使用一定单位的实际测量值进行加减运算，具有很高的量化程度。

经过整理的数据应纳入配置管理，以便根据需求和变更进行数据的维护和更新，以确保满足数据测量的要求

**Step4:** 数据质量核查。数据质量核查是数据分析应用的基础，为了确保可靠性，我们对数据进行质量核查。数据质量直接影响数据分析结果的可信度。我们的目的是筛选出文件中模糊、不完整、不一致或模棱两可的答案。例如，调查问卷中所有答案都选择了第一个选项的问卷可能被视为废弃问卷，因为回答者可能并未认真回答或心不在焉，这样的问卷参考价值较低。以上情况都需要筛选掉，不考虑在后续的数据分析阶段中。

**Step5:** 调整样本结构。通过采用加权估计的方法对样本结构进行调整，以确保样本对总体的代表性。

### （七）数据的质量诊断与评估

确保数据的准确性对于短视频使用信息的可信度至关重要，因此数据检验过程至关重要。问卷调查的质量主要受到人的特征、情境因素和测量工具等方面的影响。此外，在数据收集、记录或输入过程中出现的人为错误可能导致数据异常。因此，我们必须从逻辑和数学两个方面对最终数据进行检查。

首先，我们进行逻辑检查。逻辑检查从定性角度检查数据是否符合逻辑，内容是否合理，以及各项目或数字之间是否存在相互矛盾的现象。举例来说，如果在问卷中被调查者选择没有制作过短视频，那么与制作短视频相关的问题就失去了参考价值，我们只能将其剔除。其次是数学检查，即从定量角度检查各项数据在计算方法和计算结果上是否存在错误。

在发现数据中的异常值和缺失值后，我们对其进行处理。对于缺失值，我们可以选择使用预测模型或 KNN 插补等方法进行填补。对于异常值的处理，如果数据的异常值是由处理过程中的错误或导致的，并且异常值的数量较少，我们可以直接删除异常值。此外，我们还可以使用数据转换或聚类转换数据取对数的方法来减少极端值的影响。对于部分数据，我们选择使用均值、中位数或众数替换异常值。根据调查数据的具体情况，对于较少的异常数据，我们可以选择人工处理数据值。如果异常值较多，在统计模型中我们应该将其分组处理，即将异常值与正常值分组处理，然后分别建立模型，最后将结果合并。

### 三、调查数据的分析

#### （一）信效度检验

##### 1. 信度检验

为确保此次问卷调查收集到的问卷的质量，对问卷进行信度检验，检验问卷所得结果的一致性和稳定性程度，采用克隆巴赫信度系数法。克隆巴赫信度系数法的公式如下所示。

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

其中， $k$  为量表所包括的总题数， $\sum_{i=1}^n S_i^2$  为量表题项的方差总和， $S^2$  为量表题项加总后方差。

根据发放问卷的量表，进行信度检验，计算克隆巴赫 Alpha 信度系数，计算结果如表 3 所示。

表 3 问卷信度分析		
克隆巴赫信度系数	基于标准化的克隆巴赫信度系数	项数
0.888	0.889	9

克隆巴赫 Alpha 信度系数的值和标准化克隆巴赫 Alpha 信度系数的值分别为 0.888 和 0.889，均大于 0.8，说明问卷的信度水平较高，数据真实可靠，能够满足实证研究的要求。

##### 2. 效度检验

表 4 KMO 和巴特利特检验结果		
KMO 取样适切性量数		0.896
巴特利特球形检验	近似卡方	4766.948
	自由度	36
	显著性	0.000



效度分析能够有效判断问卷设计题目的合理性,分析题目能否充分反映出变量的研究信息与维度。本文针对调查问卷采取 KMO 和 Bartlett 的检验,如表 4 所示,得到 KMO 取样适切性量数为 0.896,大于 0.6,同时 Bartlett 球形检验  $P$  值显著小于 0.05,表明问卷的结构设计比较合理。

上述检验结果表明,所采用的问卷能够达到本次调查目的,可以进行数据分析与建模。

### （二）调查数据的基本描述性统计

本次调查我们采集调查对象性别分布中,女性和男性占比大致相同。由于我们是随机发放的问卷,因此,调查对象的性别分布能够真实反映对短视频直播到货的看法。调查对象性别结构如图 2 所示。由图 2 可以看出,在本次调查的 1200 个样本中,男性被调查者约占 54%,女性被调查者约占 46%。

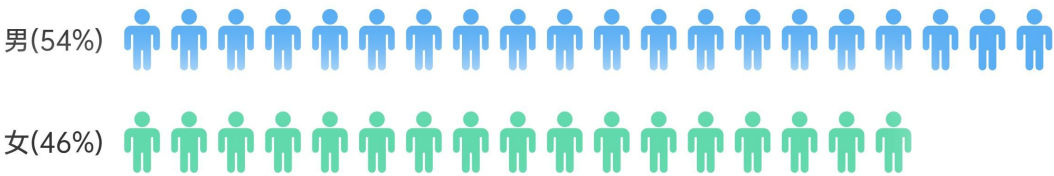


图 2 调查对象的性别分布

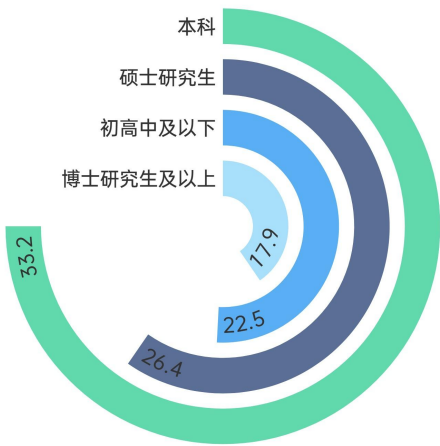


图 3 调查对象的学历分布图

从调查对象的学历来看,调查对象的学历分布如图 3 所示,被调查者的学历大多数是本科生,其次是硕士研究生,少部分是初高中及以下和博士研究生及以上,说明本次调查的调查结果真实可信。

### （三）短视频的基本使用情况的特征分析

#### 1. 基于短视频的基本使用情况的分析

## 短视频用户画像

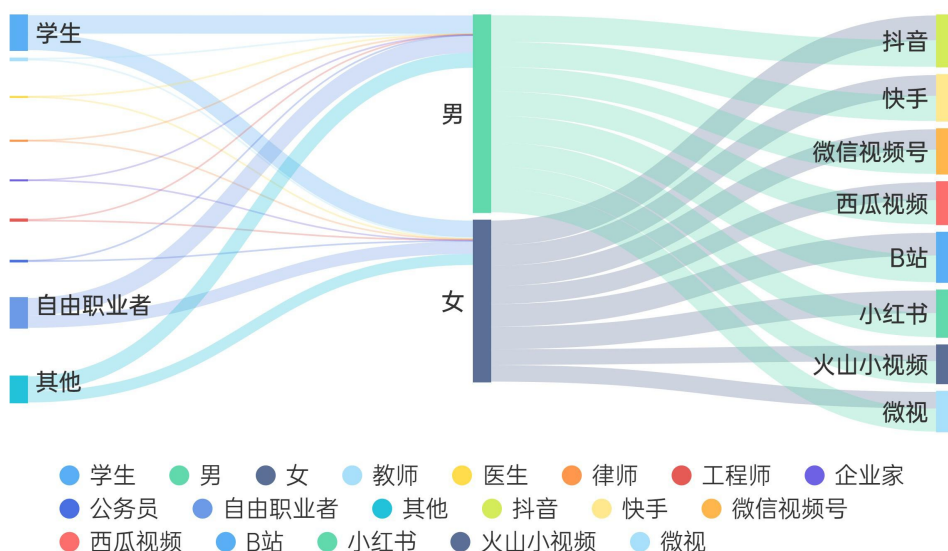


图 4 不同职业/目标下男女在社交媒体平台上的用户比较

为了比较不同职业中男性和女性在不同社交媒体平台上的数量，绘制如图 4 所示的桑基图。总体观察显示男性在大多数平台上的用户量稍多于女性，但也有些平台女性用户数量较多。

### (1) 职业与短视频使用时间的交叉分析

我们将调查对象的职业与调查对象每天在短视频上使用的时间进行交叉列联表分析以及卡方检验，研究职业与短视频使用时间的相关程度。首先进行假设： $H_0$ ：职业与短视频使用时间无相关性； $H_1$ ：职业与短视频使用时间有相关性。其次选取卡方统计量即公式(3)。

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (3)$$

其中  $O_i$  为实际频数， $E_i$  为理论频数， $n$  为总频数，卡方检验以及交叉列联表分析结果如表5、表6所示。

表 5 职业与每天使用短视频时间的卡方检验

	值	自由度	渐进显著性（双侧）
皮尔逊卡方	105.669	40	0.000
似然比	102.161	40	0.000
线性关联	0.922	1	0.337
有效个案数	1067		

由卡方检验结果可知，皮尔逊卡方检验所得的  $P$  值为  $0 < 0.05$ ，故拒绝原假设，认为不同职业与每天使用短视频时间这两者之间有显著性差异，即职业与观

看短视频时间明显相关。

表 6 职业与每天在短视频上花费的时间的交叉列联表

每天在短视频上花费的时间/h (频数)	职业									合计
	学生	教师	医生	律师	工程师	企业家	公务员	自由职业者	其他	学生
<0.25h	75	1	0	0	0	0	1	78	73	228
0.25h-0.5h	48	6	0	1	7	0	1	37	45	145
0.5h-1h	59	7	5	2	3	2	3	34	29	144
1h-2h	58	7	0	1	4	0	3	57	35	165
2h-3h	52	0	0	0	3	0	2	43	45	145
>3h	93	1	0	1	0	2	1	81	61	240
合计	385	22	5	5	17	4	11	330	288	1067

从表 6 可以得出,学生每天在短视频上花费的时间最多,其次是自由职业者,企业家每天在短视频上花费的时间最少。

学生每天花费在短视频上的时间大多在一小时以上,花费时间较长,学生群体的时间相对宽裕,而且更年轻化,因此每天有较多的时间使用短视频进行娱乐、购物等。教师、工程师、自由职业者每天花费在短视频上的时间大多在 1 小时以下,这类群体会更关注自己的职业情况,时间相对紧缺,因此每天花费在短视频上的时间也较少。公务员,每天花费在短视频上的时间大多分布在 30-60 分钟和 1-2 小时,由于公务员职业的特殊性,这类群体时间较紧缺,但需要每天关注时政新闻,因此每天花费了一定的时间在短视频上。被调查者的职业为医生、律师和企业家的人数较少,且每天花费在短视频上的时间也很少。由此可见,不同职业的人群每天花费在短视频上的时间不相同,时间越充分、越年轻化的群体每天花费在短视频上的时间较多。

## (2) 职业与观看短视频内容类型的交叉分析

我们将进一步研究职业与观看短视频内容类型的关联程度,首先作出假设: $H_0$ :职业与观看短视频内容类型之间无显著性差异; $H_1$ :职业与观看短视频内容类型之间有显著性差异。然后进行卡方检验以及交叉列联表分析,所得结果如表 7、表 8 所示。

表 7 职业与最常观看内容类型的卡方检验

值	自由度	渐进显著性 (双侧)
---	-----	------------

皮尔逊卡方	22.023	8	0.005
似然比	28.489	8	0.000
线性关联	1.665	1	0.197
有效个案数	1067		

通过卡方检验表 7 可知, 皮尔逊卡方检验所得的  $P$  值为  $0.005 < 0.05$ , 故拒绝原假设, 认为不同职业与最喜欢观看的短视频内容类型这两者之间有显著性差异, 即职业与观看的短视频内容类型具有明显相关性。

表 8 职业与最常观看内容类型的交叉列联表

最常观看的短视频内容类型 (频数)	职业									合计
	学生	教师	医生	律师	工程师	企业家	公务员	自由职业者	其他	
娱乐	81	2	1	0	4	0	1	86	93	268
美食	42	4	0	1	2	0	1	58	47	155
美妆	67	3	0	2	1	2	2	42	39	158
新闻	45	7	2	1	4	1	3	33	36	132
游戏	52	0	0	0	3	0	2	67	42	166
教育	96	6	2	1	3	1	2	44	31	186
合计	385	22	5	5	17	4	11	330	288	1067

从表 8 可以看出, 被调查者中最常观看的短视频内容为娱乐类型的, 其次为教育、游戏、美食类型, 此外自由职业者是观看娱乐视频最多的, 而教师与学生常常会选择观看教育类型的视频, 公务员一般会观看新闻以及教育类型, 因此可以看到职业对观看短视频内容的选择是有一定影响的, 即不同职业观看的内容是有差异的。

人们选择短视频的原因多种多样, 包括多样的视频内容、短时间的消费形式、创新的表达方式、社交互动的娱乐性、放松与感官刺激的追求, 以及方便易用的平台等。如今短视频已经给人们提供了娱乐、放松、学习和创造的机会, 成为日常生活中不可或缺的一部分。

## 2. 短视频的评价与消费者购买意愿

### (1) 短视频的整体满意度

本小节我们通过交叉列联表与卡方检验来分析人们对短视频的整体满意度与每天观看短视频的时间之间的差异性, 通过结果探究人们对短视频的整体评价, 在问卷中, 我们将满意度等级分为 0-10, 其中 0 为非常不满意, 10 为非常满意。假设  $H_0$ : 人们对短视频的整体满意度与每天观看短视频的时间之间没有相关性;  $H_1$ : 人们对短视频的整体满意度与每天观看短视频的时间之间有相关性。

短视频的整体满意度与职业的卡方检验以及交叉列联表结果如表 9、表 10 所示。

表 9 短视频的整体满意度与每天观看短视频的时间的卡方检验

	值	自由度	渐进显著性（双侧）
皮尔逊卡方	48.119	50	0.000
似然比	52.130	50	0.000
线性关联	0.252	1	0.616
有效个案数	1067		

通过卡方检验表可知，皮尔逊卡方检验所得的  $P$  值为  $0.000 < 0.05$ ，故拒绝原假设，认为视频的整体满意度与每天观看短视频的时间这两者之间有显著性差异，即视频的整体满意度与每天观看短视频的时间之间具有明显相关性。

表 10 短视频的整体满意度与每天观看短视频的时间之间的交叉列联表

对短视频的 整体满意度	每天观看短视频的时间/h						
等级	<0.25h	0.25h-0.5h	0.5h-1h	1h-2h	2h-3h	>3h	总计
0	2	1	1	1	0	0	5
1	1	0	0	0	1	3	5
2	9	5	5	5	6	4	34
3	9	7	4	1	4	7	32
4	13	4	5	3	4	12	41
5	32	15	21	19	17	28	132
6	46	28	22	29	20	54	199
7	40	32	39	44	40	52	247
8	50	34	37	38	40	55	254
9	15	15	7	14	11	17	79
10	11	4	3	11	2	8	39
合计	228	145	144	165	145	240	1067

从表 10 可知，有很少一部分人对短视频整体满意程度为 0-2，其观看时间大致在 10-30 分钟之间，但大多数人对短视频的满意度中等偏高，多集中在 6-8 分之间，给短视频打分在 6-8 区间的人数比例已经超过了一半，还有一部分人的满意程度为非常满意，这部分人观看短视频的时间也比较长，一般在 2-3 小时或 3 小时以上，因此绝大部分人对短视频的态度还是乐观积极的，整体满意度较高。

从每天观看短视频的时间上来看，观看时间在 30-120 分钟的人数最多，然而也有一部分人观看时间较长在 3 小时以上，而往往这些人对短视频的打分更高。因此从整体上看，人们对短视频的满意度与观看时间是有一定联系的，对短视频打分较低即对其比较排斥的人打分一般都低于 5，所以他们的观看时间都集中在 30 分钟之内，而对短视频整体满意度中等偏高的人来说，观看时长都集中在 2 小时以上。

## (2) 短视频内容的影响分析

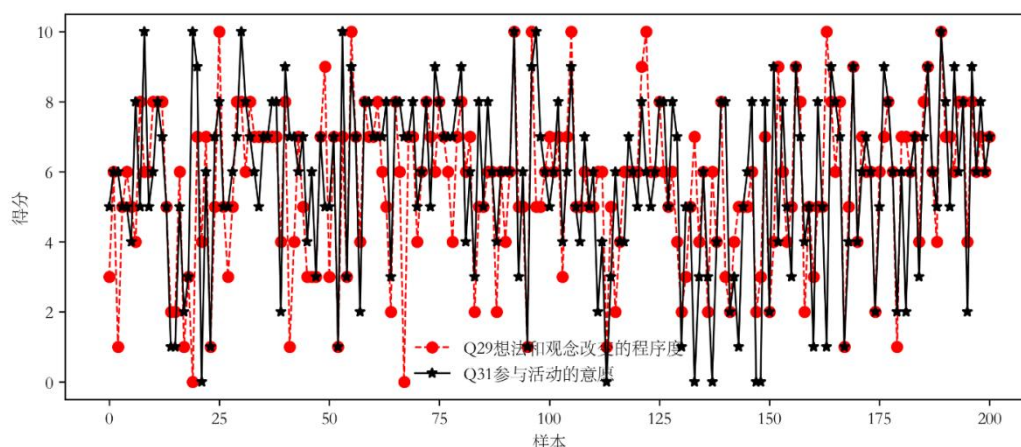


图 5 短视频内容对用户的影响意愿评分

通过被调查者根据短视频的内容而改变自己的想法或观念的程度以及短视频的内容而购买相关的商品或服务的意愿这两项所给的得分绘制如图 5 所示。

由图 5 可知，被调查者所给的得分主要集中在 5-8 之间，仅有个别被调查者打分为 0 或 10，说明短视频对我们的观念以及参与度是有一定影响的，但影响不会很大。因此可以认为短视频的发展是比较成熟、适当合理的，它并没有完全占据人们的生活，仍给我们的生活留下了自由发展的空间，我们既能通过短视频获取知识了解新闻而改变理念，同时也让更多的人参与进来，更愿意接受新鲜、创意、有趣的内容，并通过社交网络分享和传播，产生共鸣，更有利于信息的分享。

## (3) 消费者购买意愿分析

由于绝大部分人对短视频都比较认可，因此我们有必要进一步分析短视频能获得认可的原因。除了短视频提供的娱乐新闻等内容，它的崛起还有更深层次的原因，在飞速发展的数字化时代，短视频和直播内容更生动、直观，传播更有效率，能够有效地挖掘、释放和激发老百姓潜在的新消费需求。

在未来，短视频可以更好地进入到垂类领域和传统产业的生产层面，参与传统产业升级再造，为产业带来新的发展机遇期，通过更加直观的实景视频展示、视频直播让消费者身临其境，满足消费者多样化和个性化的需求，从而让越来越多的人愿意使用短视频提供的平台进行消费。

## 3. 短视频与电商平台的结合

### (1) 月收入与月均网络直播购买支出的相关性分析

短视频提供了一种全新的消费者购买决策方式。通过短视频平台，消费者可以看到商品以及其他消费者的使用体验和评价，这大大提高了消费者购买决策的

准确性，此外，一些短视频平台还提供了直接购买商品的功能，让消费者更加便捷地完成购买。因此，短视频平台在一定程度上会刺激绝大部分人的网上消费，这里我们将月收入与月均网上消费水平进行相关性分析，从而得到月收入的高低与网上消费金额大小之间的关系，相关性结果如表 11 所示。

表 11 月收入与月均网上消费水平进行相关性

		月收入	网上月均消费
月收入	皮尔逊相关性	1	0.443
	显著性	——	0.0
网上月均消费	皮尔逊相关性	0.443	1
	显著性	0.0	——

由上表可知，显著性概率  $P$  值小于 0.05，因此我们可以认为月收入与网上消费有一定的相关性，且相关系数大于 0，即月收入的高低与每月通过直播平台进行的消费水平成正比，月收入越高，在短视频等各个直播电商平台的消费也越高，月收入若较低，则在短视频等各个直播电商平台的消费也会降低。

由此也可以看出，视频内容在不断丰富化多样化，而且如今短视频平台已经与电商结合合并合作，推出直播带货等形式增加人们的消费欲望。尽管短视频的功能在不断强大，带动消费水平，但根据以上分析可知，人们在直播间的消费还是理性的，不会盲目消费，能根据自身的经济水平来决定是否在直播间购买所需商品。

## (2) 学历与直播间购买商品类型的分析

下面我们分析不同学历的被调查者选择在直播间购买的商品类型的差异，商品类型主要包括美妆、食品、服饰、健身用品、学习用品、家用电器、家居装饰品七类。对其进行交叉列联表分析以及卡方检验，假设  $H_0$ : 学历与直播间购买的商品类型无显著差异， $H_1$ : 学历与直播间购买的商品类型有显著差异，得到的卡方检验结果与交叉列联表如表 12、13 所示。

表 12 学历与直播间购买商品类型的卡方检验

	值	自由度	渐进显著性（双侧）
皮尔逊卡方	33.224	24	0.009
似然比	33.944	24	0.0860
线性关联	0.012	1	0.911
有效个案数	1067		

通过卡方检验表可知，皮尔逊卡方检验所得的  $P$  值为  $0.009 < 0.05$ ，故拒绝原假设，认为学历与在直播间选择购买的商品类型这两者之间有显著性差异，即学历与购买的线上商品类型具有明显相关性。

表 13 不同学历选择在直播间购买的商品类型的差异

每月在直播	学历	总计
-------	----	----



带货中主要购买的商品类型（频数）	初高中及以下	本科	硕士研究生	博士研究生及以上	
美妆类	38	40	45	24	147
食品类	34	65	38	28	165
服饰类	26	61	37	25	149
健身用品类	28	39	23	21	111
学习用品类	16	35	39	27	111
家用电器类	28	29	32	21	116
家居装饰类	25	30	27	9	91
其他	26	24	22	17	89
从不购买	19	31	19	19	88
总计	240	354	282	191	1067

由表 13 可以看到，被调查者在直播间中购买的商品类型最多的是食品类，其次为服饰类与美妆类，购买较少的是健身用品与家用电器类。由于直播间商品的种类更多，选择性更大，因此本科生一般会选择在直播间购买食品、服饰与美妆类商品，硕士研究生会购买学习用品与美妆类商品，而博士研究生购买的商品类型比较宽泛。

### （3）线上消费原因分析

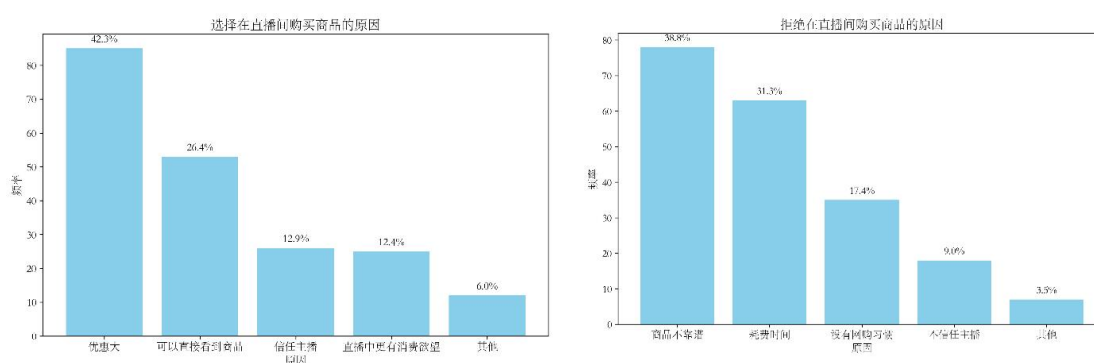


图 6 线上消费的原因分析图

从图 6 可看到人们选择在直播间购买商品以及拒绝在直播间购买商品的原因，将近半数 42.3% 的人在直播间购买商品的主要因素是直播间的商品优惠力度更大，其次是因为能够更形象直观地看到商品，还有一少部分人出于信任主播、观看直播的过程中更有消费欲望，所占比例为 25.3%。在选择拒绝购买直播间商品的调查者中，有 38.8% 的被调查者认为线上的商品无法保证其质量，31.3% 的被调查者由于时间问题而不选择购买直播间商品，还有少部分的人没有网购习惯，仅有 9.0% 的人是因为不信任主播。综上，影响消费者购买直播间商品的最重要原因是线上商品的质量与优惠力度，而这也是消费者最关心的问题，因此电商平台要将商品质量与商品价格进行衡量，既要提供优惠力度，更要保证商品的品质，而消费者也要根据平台的安全性与自身的实际情况合理利用直播消费。



#### 4. 短视频未来发展优缺点分析



图 7 词云图

该词云图展现了用户对短视频平台的期待和担忧。他们希望平台能够提供高质量、真实可信的内容，避免低俗和虚假信息的传播。用户迫切希望看到更多正能量和有意义的教育内容，并呼吁平台加强审核和管理，确保内容的安全性。同时，用户也希望平台能够更加个性化地推荐内容，满足用户的个性化需求。总的来说，用户期待短视频平台能够提供更优质的内容和服务，从而提升用户的体验和满意度。

#### 四、短视频内容对消费者影响因素分析

### （一）短视频内容对消费者行为和态度的影响分析

### 1. 模型变量选择及相关性分析

进一步为了研究短视频内容对消费者行为和态度的影响,我们主要选择如下十个变量进行研究,如表 11 所示,以 Q30(因短视频的内容而购买相关的商品或服务的意愿)为因变量,其他九个变量为自变量,建立相应的数学模型。

表 14 短视频内容对消费者行为和态度的影响因子

影响因子	问题编号	变量定义
消费观与习惯	Q17	自变量
	Q18	
	Q21	
	情绪	
	Q28	
	想法、观念	
参与度意愿	Q29	
	Q31	

满意度	Q33	
购买意愿	Q30	因变量

首先为了更清楚的观察各变量之间的相关性，绘制相应的相关性热力图，如图 7 所示。

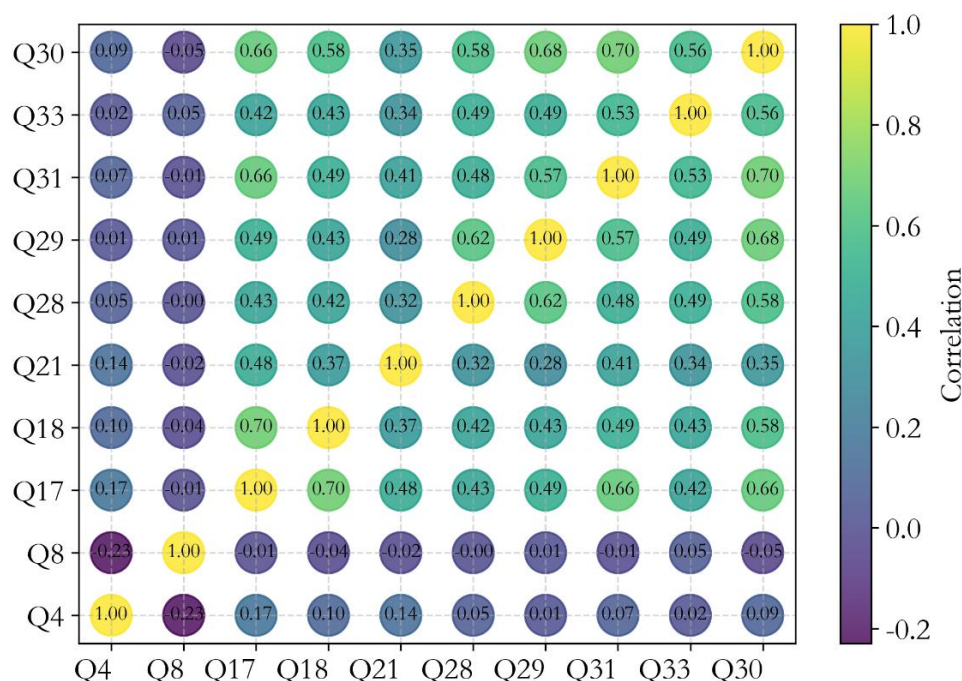


图 8 相关性热力图

通过图 8 可以看出 Q30 与其他变量均呈现正相关性，说明短视频对消费者的消费习惯、观念和想法的影响越大，则消费者对商品的购买意愿越强烈，其中消费者的参与度意愿和购买意愿呈现显著相关，可见消费者在浏览短视频过程中，参与的活动或互动越多，那么对该商品或服务购买意愿越强烈。

## 2. 随机森林模型

### (1) 模型的建立

随机森林预测模型是一种机器学习算法，其基于决策树，能够用于实现多种任务，包含回归和分类等。该算法是采用集成方法来实现的，也就是利用大量的决策树来进行模型的训练和预测。每个决策树都可以生成自己的预测结果，而随机森林模型则将这些预测结果结合起来，从而得到更加准确的预测结果。该算法的集成学习思想也是其优势之一，能够通过结合多个分类器来解决复杂问题，提高模型的性能表现<sup>[9]</sup>。

随机森林预测模型的核心思想是利用多个决策树的集成来提高预测准确性和稳定性，同时降低单个决策树的方差和偏差。为了实现这一目标，随机森林通

过引入两种随机性：自主抽样和特征子集抽样，使得每个决策树都具有一定的独立性和多样性，从而减少了过度拟合和共线性的风险。通过对多个决策树的预测结果进行投票，随机森林可以有效地降低单个决策树的噪声和误差，并提高整体模型的泛化能力和鲁棒性，不易产生过拟合现象。

特征重要性反映了某个特征加入预测模型时，该特征对预测精度的贡献程度。可以通过计算每个特征在树中的平均信息增益或平均基尼指数得到特征重要性。这里我们使用平均信息增益来刻画特征重要性，用以下公式来度量。

$$H(S) = - \sum_{k=1}^K \frac{|D_k|}{|S|} \log_2 \frac{|D_k|}{|S|} \quad (4)$$

$$H(S|A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} H(S_i) = - \sum_{i=1}^n \frac{|S_i|}{|S|} \sum_{k=1}^K \frac{|D_{ik}|}{|S_i|} \log_2 \frac{|D_{ik}|}{|S_i|} \quad (5)$$

$$G(S, A) = H(S) - H(S|A) \quad (6)$$

其中，式（4）中的  $H(S)$  是数据集  $S$  的熵， $|D_k|/|S|$  表示第  $k$  类样本的总数占数据集  $S$  样本总数的比例；式（5）是特征  $S$  对于数据集  $S$  的经验条件熵  $H(S|A)$ ， $S_i$  表示以属性集  $A$  为划分，分成  $n$  个分支，第  $i$  个分支的节点集合；式（6）是特征集  $A$  对数据集  $S$  的信息增益  $G(S, A)$ ，它反映了以特征集  $A$  划分前后数据集  $S$  的熵的差值，因此，越大越好。

随机森林预测模型的构建过程如图 9 所示，模型步骤如下：

**Step1:** 从数据集中随机选择  $n$  个样本，每个样本可以重复选择(即有放回采样)，作为训练集。

**Step2:** 从所有特征中随机选择  $k$  个特征 ( $k < m$ ， $m$  为总特征数)，作为候选分裂特征。

**Step3:** 使用候选分裂特征构建一棵决策树模型，不进行剪枝，并且每个节点内的样本数保存一致。

**Step4:** 重复步骤 1~3 直到生成指定数量的决策树，构成一个随机森林。

**Step5:** 对于回归问题，使用随机森林中所有决策树的平均值作为最终预测结果；对于分类问题，使用随机森林中所有决策树的投票结果作为最终预测结果。

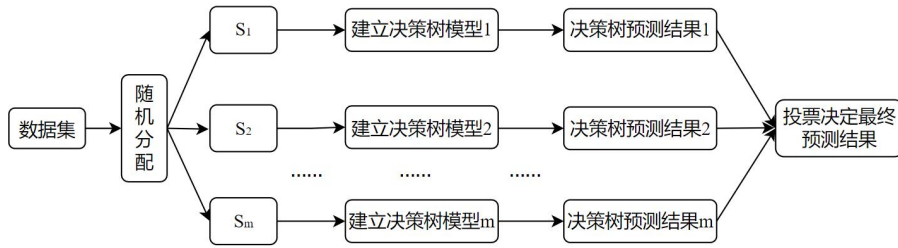


图 9 随机森林的步骤流程图

根据上述理论，建立随机森林模型，需要确定两个重要的参数，即随机森林决策树的数目(ntree)和决策树在每个节点处用于选择最优变量的变量数目(mtry)。

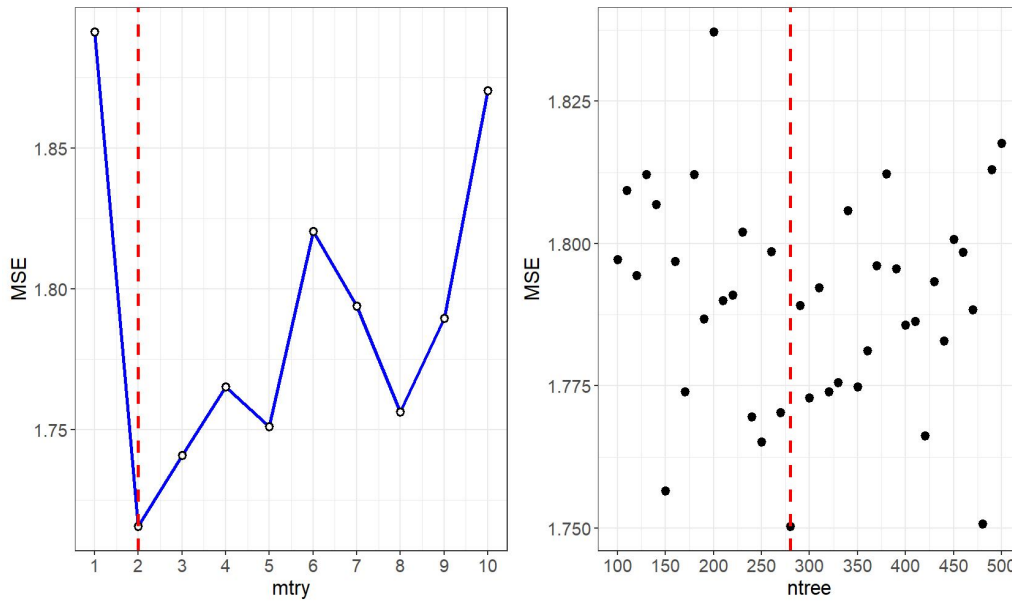


图 10 随机森林决策树的变量图

通过十折交叉验证法，计算参数 ntree 对应的平均均方误差，如图 10 所示，虚线所在的点为 MSE 最小值点，即当 ntree=280 时，平均均方误差最小。对于参数 mtry，分别拟合出不同 mtry 值下模型的 MSE 和可解释方差，结果如图所示，虚线所在的点为 MSE 最小值点，即当 mtry=2 时，平均均方误差最小，可解释方差最大，此时模型最优。因此，确定最优参数值为 ntree=280，mtry=2。

## (2) 模型的评价与结论

利用确定的最优参数建立随机森林模型，并且计算出各个变量的重要性，即 %IncMSE，它表示当去除某一特征时，模型的 MSE 增加的百分比，反映了该特征在模型中的重要程度，即当 %IncMSE 越大，则该特征对模型的贡献越大，接着对其进行归一化处理，即  $x_{scaled} = (x - x_{min}) / (x_{max} - x_{min})$ ，以便更清晰地展示每个变量的相对重要性，结果如图 11 所示，其中 Q31、Q29 和 Q17 的重要度均超过了 0.5，为了提供一个参考标准，我们认为得分超过 0.5 的变量具有显著的影响力，这一阈值是基于模型中所有变量重要性得分的分布确定的，由此可见消

费者对短视频中活动的参与度和消费者观念想法的改变是导致消费者选择购买决策的主要原因。

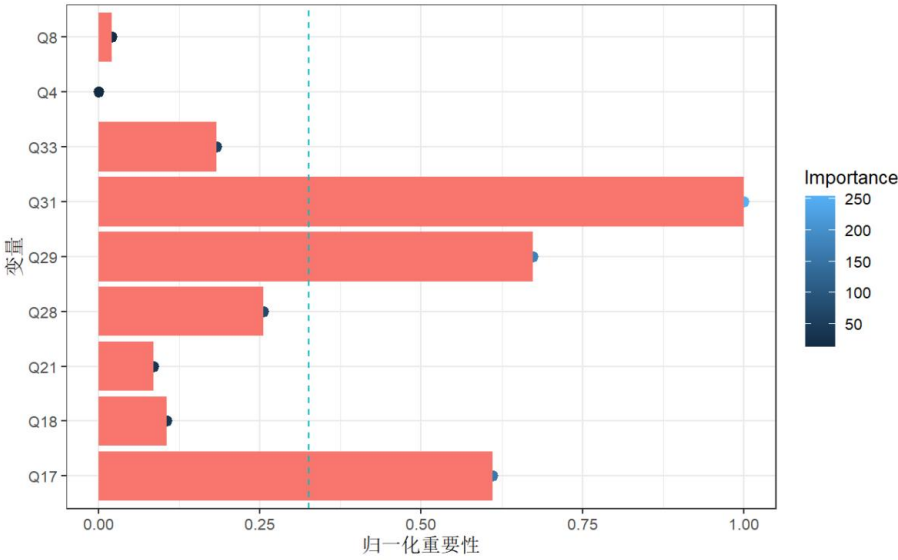


图 11 归一化重要性与变量的关系图

建立的随机森林模型，不仅揭示了影响消费者购买意愿的关键因素，而且还能预测消费者未来的购买意愿。我们计算了上述模型的三种误差指标：MAE、MSE 和 RMSE，以评估模型的预测准确性。如表 15 所示。

表 15 模型的误差值

	MAE	MSE	RMSE
误差值	0.4616479	0.4030551	0.6348662

由表 15 中的结果可知，该模型的误差均小于 1，表明模型具有较好的拟合度和较强的解释力。

### 3. 基于 ANOVA F 值的 SelectKBest 法

为了与随机森林重要性进行比较，接下来采用基于 ANOVA F 值的 SelectKBest 方法计算各变量的 F 值作为评分。

基于 ANOVA F 值的 SelectKBest 方法是一种特征选择方法，用于从给定的特征集合中选择最佳的  $K$  个特征。它基于方差分析(ANOVA)的原理，通过计算每个特征与目标变量之间的 F 值来评估特征的重要性<sup>[10]</sup>。这种方法的优点是可以通 F-值来量化特征与目标变量之间的关系强度，从而可以得到影响目标变量的主要因素。

**Step1:** 计算 F 值。ANOVA F 值是用来估计特征变量和目标变量之间的线性关系的统计方法，其表达式为

$$F = \frac{SSA / (k - 1)}{SSE / (N - k)} \quad (7)$$

其中 $k$ 是组数， $N$ 是总样本数， $SSA$  为组间平方和， $SSE$  为组内平方和，分别为  
 $SSA = \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2$ ， $SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$ ，式中 $n_i$ 是第 $i$ 组的样本大小， $\bar{x}_i$ 是第 $i$ 组的样本均值， $\bar{x}$ 是所有样本的总均值， $x_{ij}$ 是第 $i$ 组中第 $j$ 个观测值， $\bar{x}_i$ 是第 $i$ 组的样本均值。

Step2: 对所有特征的  $F$  值进行排序。

Step3: 将  $F$  值进行归一化，作为各个特征的评分，并且绘制评分雷达图。

通过上述的步骤，计算得到的以  $F$  值为评分的雷达图如图 12(a)所示，以归一化后的  $F$  值为评分的雷达图如图 12(b)所示。

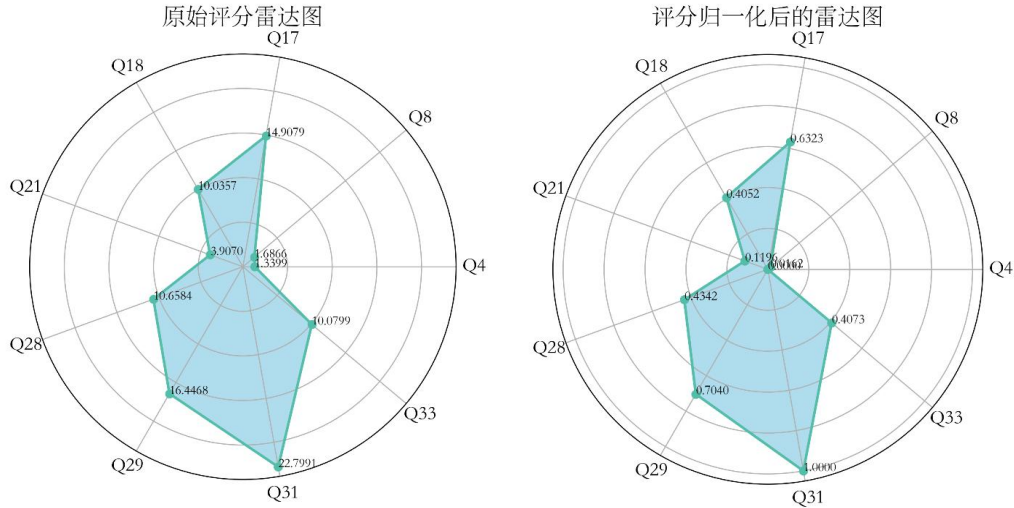


图 12 (a)以  $F$  值为评分的雷达图；(b)以归一化后的  $F$  值为评分的雷达图

雷达图直观的展示了各个变量评分情况，其中  $Q31$  评分最高，其次是  $Q29$ 。由此，可见消费者观念、想法与消费者对短视频内容或活动的参与是影响消费者购买意愿的主要原因，与之前随机森林模型得分结论一致。

## (二) 消极行为的影响因素研究

### 1. 模型理论

多元线性回归模型是一种用于分析自变量与因变量之间关系的统计模型。在实际应用中，它通常用于探索多个自变量对一个因变量的影响，并且可以识别出各自变量对因变量的相对重要性<sup>[11]</sup>。

模型的矩阵形式表示如下：



$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1m} \\ 1 & \vdots & \cdots & \vdots \\ 1 & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_m \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_0 \\ \vdots \\ \varepsilon_m \end{bmatrix} \quad (8)$$

可简写为

$$\begin{cases} Y = X\beta + \varepsilon \\ \varepsilon \sim N(0, \sigma^2 E_n) \end{cases} \quad (9)$$

其中  $E_n$  是  $n$  阶单位矩阵。 $x_1, \dots, x_m$  是非随机的变量， $y$  是随机的因变量， $\beta_0, \beta_1, \dots, \beta_m, \sigma^2$  都是与  $x_1, \dots, x_m$  无关的未知参数，其中  $\beta_0, \beta_1, \dots, \beta_m$  是回归系数， $\varepsilon$  是随机误差项。

模型中的参数  $\beta_0, \beta_1, \dots, \beta_m$  用最小二乘法来进行估计，即：

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (10)$$

残差向量为  $e = Y - \hat{Y} = Y - X\hat{\beta}$ ，则随机误差  $\sigma^2$  的最小二乘估计为：

$$\sigma^2 = \frac{e^T e}{n - m - 1} \quad (11)$$

进行相应处理代入回归模型即可得到  $y$  的估计值：

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_m x_m \quad (12)$$

即数据的拟合值为  $\hat{Y} = \hat{X}\hat{\beta}$ 。

## 2. 变量选择与解释

### (1) 自变量与因变量

我们将探讨多个自变量对消极行为的影响。这些自变量包括负面内容、使用习惯、隐私顾虑、平台质量等。消极行为作为因变量，表示受访者对短视频的整体满意程度 (1-10 分)。自变量的具体细节如表 16 所示。

表 16 变量名称与编号

变量类型	变量名称(题号)	编号
自变量	负面内容(Q34)	$X_1$
	使用习惯(Q8,Q9,Q11)	$X_2$
	隐私顾虑(Q13)	$X_3$
	平台质量(Q6)	$X_4$
因变量	消极行为(Q33)	$Y$

为了更加清晰的说明变量情况，如图 13 所示，直观的进行了展示。



图 13 自变量的具体细节

因变量指消极行为，即：受访者对短视频的整体满意程度是多少(1-10 分)。

## (2) 数据处理

每个变量的得分都是通过加权求和得到的，权重是根据每个选项出现的频率以及归一化得到的，其中负面内容，隐私顾虑，消极行为，使用习惯等变量的值越大，消极行为的分值越高，相反，平台质量的值越大，消极行为的值越低。由于数据个数是 1067，相对较多，存在冗余样例的可能性越大，所以，首先利用 PCA 技术，将数据集降到 2 维，可视化结果如下：

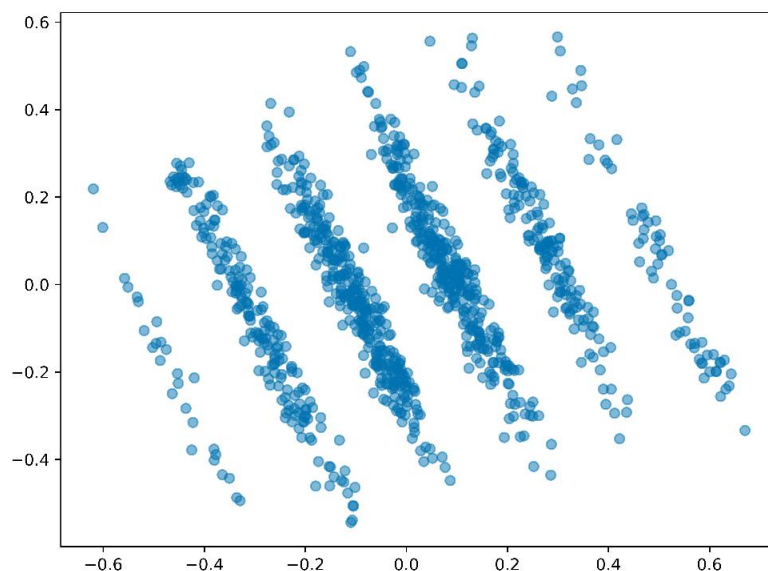


图 14 降维后的数据散点图

从图 14 中看出，数据可以聚为 6 类，通过样例选取技术，最终选择了具有代表性的 12 个样例，结果如表 17 所示。

表 17 处理后的数据

ID	负面内容	使用习惯	隐私顾虑	平台质量	消极行为
----	------	------	------	------	------



1	0.535	0.65341	0.2	0.003356	0.9
2	0.768	0.5623	0.2	0.311838	0.5
3	0.4	0.59818	0.1	0.252472	0.2
4	0.5	0.91047	0	0.369175	0.7
5	0.961	0.0827	0.6	0.294732	0.8
6	0.752	0.64538	0.2	0.351067	1
7	0.775	0.15301	0.5	0.458986	0.4
8	0.751	0.42516	0.2	0.461393	0.3
9	0.598	0.41712	0.4	0.69507	0.1
10	0.628	0.61319	0.2	0.80931	0.4
11	0.226	0.78521	0.2	0.744955	0.4
12	0.636	0.71467	0.1	0.965243	0.3

在建立多元回归模型之前，需要对数据变量进行相关性分析，利用 python 软件，计算得到了各变量之间的相关系数以及绘制矩阵散点图来观察变量间的相关性。一方面是相关系数矩阵，结果如下：

$X_1$	1	-0.711	-0.588	-0.178	0.290
$X_2$	-0.711	1	-0.912	0.159	0.081
$X_3$	0.588	-0.912	1	-0.075	0.037
$X_4$	-0.178	0.159	-0.075	1	-0.579
$Y$	0.290	0.081	0.037	-0.579	1

另一方面是矩阵散点图：

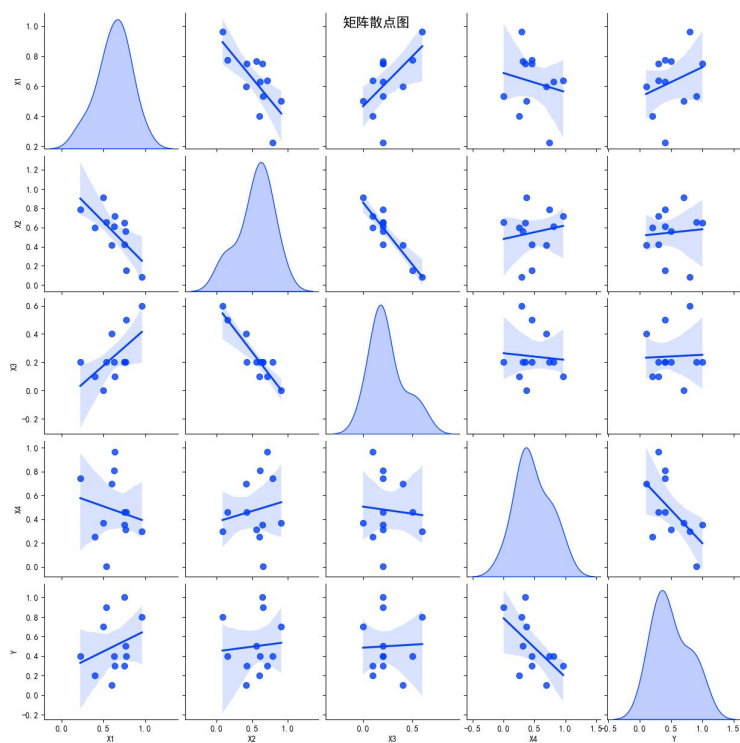


图 15 矩阵散点图

由相关系数矩阵和矩阵散点图可以看出负面内容，使用习惯，隐私顾虑等变

量与消极行为呈正相关，当这些变量的值变大时，用户越有可能产生消极行为，从而客户流失率增高，短视频公司应该多推荐一些正能量，积极的视频，以便留住用户。同时，当平台质量越高时，用户使用该 APP 的时间更长，粘性系数越好。

### 3. 建模分析

#### (1) 结果解释

通过计算，得到的结果如表 18 所示。

表 18 回归系数

常量	$X_1$	$X_2$	$X_3$	$X_4$
-1.198	1.151	2.642	2.603	-0.745

则多元回归方程为：

$$Y = -1.958 + 1.151X_1 + 2.642X_2 + 2.603X_3 - 0.745X_4 \quad (13)$$

根据回归结果，我们可以得出以下结论：负面内容是回归模型中的截距项，表示当所有自变量的值为 0 时，消极行为的预测值为 1.151，即使其他自变量的影响被控制在 0 的情况下，仍然存在一定程度的消极行为。同时，使用习惯和隐私顾虑与消极行为呈正相关关系，即增加使用习惯和隐私顾虑可能会导致更多的消极行为。相反，平台质量与消极行为呈负相关关系，即提高平台质量可能会减少消极行为的发生。

#### (2) 模型的检验

对所建的回归模型进行检验，利用  $F$  检验，进行假设：

$$H_0: \beta_1 = \dots = \beta_m = 0 \quad \text{VS} \quad H_1: \beta_1, \dots, \beta_m \text{不全为} 0$$

当假设成立时，检验统计量：

$$F = \frac{SSR/m}{SSE/(n-m-1)} \sim F(m, n-m-1) \quad (14)$$

式中回归平方和：  $SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$ ；残差平方和：  $SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ 。

对给定的显著性水平  $\alpha$ ，取 0.05，检验拒绝域  $F > F_{\alpha}(m, n-m-1)$ 。通过计算得到  $F(4, 12) = 12.021$ ， $p\text{-value} = 0.003$  小于显著性水平  $\alpha$ ，拒绝原假设  $H_0$ ，接受  $H_1$ 。即可认为该回归模型通过了  $F$  检验。

表 19 F 检验结果

R 方	调整后 R 方	标准估算的错误	统计量 F	p-value
0.934	0.800	0.1295	12.021	0.003

表 20 自变量的 $p$ 值				
$X_1$	$X_2$	$X_3$	$X_4$	常量
0.006	0.001	0.004	0.002	0.006

由表格得出，调整后的  $R$  方为 0.800，即总体自变量对因变量的解释程度达到 80%，拟合效果较好，即建立的模型较为稳定。统计量  $F$  的值为 12.021， $p$ -value 是 0.003，小于 0.05，说明回归模型是高度显著的，自变量和因变量的线性关系是显著的。

#### 4. 模型的进一步分析

##### (1) 共线性诊断

共线性诊断用于检测自变量之间是否存在高度线性相关性。共线性可能导致参数估计不稳定和解释性下降的问题，常用方法包括相关系数、方差膨胀因子（VIF）和特征值分解等，本文采样 VIF 值来判断自变量间的共线性，结果如表 21 所示。

表 21 VIF 值			
$X_1$	$X_2$	$X_3$	$X_4$
2.135	9.510	7.245	1.101

每个自变量的 VIF 值均小于 10，认为自变量间不存在共线性。

##### (2) D-W 值

D-W 值是 Durbin-Watson 统计量的简称，用于检验回归模型中是否存在自相关，其取值范围在 0 到 4 之间，接近 2 表示不存在自相关，偏离 2 可能暗示存在自相关，需要进一步检验和处理，对确保模型准确性和稳定性具有重要意义。本文的结果为 2.579，则认为残差间不存在一阶自相关。

##### (3) 直方图和 P-P 图

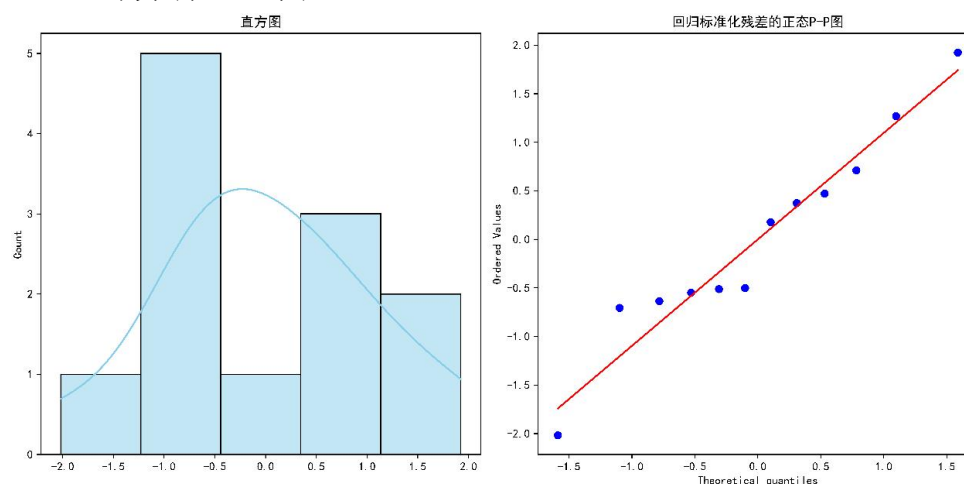


图 16 残差直方图和 P-P 图

从图 16 中可以看出，残差一定程度上符合正态分布，符合最小二乘回归的假设，可以用多元回归模型对变量进行解释，同时，回归方程和系数均显著，所以该模型较优，且结果符合人们直观的看法。

## 五、结论与建议

### （一）结论

#### 1. 短视频对用户的生活与学习有一定的积极影响

##### （1）短视频可以提升信息传递的效率

近年来，短视频平台在一众数字媒体中“异军突起”，并日益渗透至用户的日常学习和生活中。

相对于传统的信息传递方式来说，短视频在传递信息方面有更为明显的优势，一方面是短视频有更强的画面感，用户获取信息的方式由原来的图文方式转化成了现在的视频获取，这能让我们更好的获取到信息，使得用户获取信息的方式更加多元化，同时也使我们不出家门就能领略祖国的大好风光；另一方面，短视频的价值密度相对比较高，以前用户交流的方式相对单一，如用微信、QQ 等聊天工具，而如今用户不仅可以用短视频软件进行交流，而且可以对短视频进行点赞，转发，用户的交流方式更加便捷且多元化，所以短视频能够在一定程度上提升信息的传递效率，这个过程也会带来价值增量。

##### （2）短视频能够丰富人们的生活

物质文明的极度丰富使得人们生活呈现出快节奏，娱乐时间多碎片化的特点，传统的精神文化形式和内容很难满足人们现时代日益增长的精神文化需求。而短视频适合在移动状态和短时休闲状态观看的特点恰好满足了这种快节奏、碎片化的生活方式下的精神文化需求，丰富了人们的精神生活。

在我们的调查报告中，人们最常用的短视频软件是抖音（快手），这是由于抖音（快手）采用中心化的分发逻辑，而这种机制在于容易产生爆款。人们在短视频中既可以公开评论和分享，也可以私下交流，两种交流方式的结合不仅考虑到了用户的全面性，而且增加了自我表达以及与同龄人互动的渠道和机会。我们调查发现，短视频的内容包含多种类型，包括搞笑、美食、美妆、旅游、体育、情感等，而人们对各种内容的喜爱度具有异质性，且都可以看到自己所喜爱的内容，使得人们的生活更加丰富。

短视频的出现使人们通过在线社群实现娱乐消遣、信息获取、人际交往、情绪调节等方面的需求。

### （3）短视频能提升用户的学习体验

由于短视频制作门槛低、受众范围广、信息时效强等特性，用户几乎能够不受时空限制地学习、分享甚至生产知识。这在一定程度上改变了学习资源的生产与传播方式，提升学习者的参与感和主体性，使得人们获取信息和教育资源的机会更加均等。

在本次调查中，我们发现有一部分用户自己制作过短视频，且喜欢在短视频软件上进行互动。而在线互动与视频创作也为用户学习能力的培养提供了新的机遇。基于短视频的互动、分享和创造能够激发用户综合运用知识并以丰富的表达形式呈现的潜能，同时在这种非正式学习过程中增强同伴支持与自我效能感。

## 2. 短视频+电商的形式促进了线上消费的发展

### （1）短视频的发展促进电商销售

由于短视频的用户众多，因此能够提升产品曝光度和销售转化率，而短视频形式生动直观，因此能够有效展示产品特点和使用场景，吸引用户注意，减少消费者时间成本。

在本次调查中发现，通过短视频与商家的合作，商家可以让利部分宣传费用给消费者，使得消费者的购买意愿增加；并且，基于短视频和直播构建出的社交网络实现裂变式传播，在短视频中加入购买链接或二维码，每个用户既是购买者，也是传播者。基于用户个体的去中心化传播网络，为长尾商品增加发展空间。在本次调查中还发现，有一定比例的消费者的购买原因是由于足够信任带货主播，这是因为主播通过个人魅力搭建了人与货的沟通场景，增加人对货的信任感，提高购买转化率。

### （2）形成娱乐—发现—购物—娱乐的业务闭环

满足用户价值和商业价值的产品才能可持续的输出生命力，在娱乐——发现——购物——娱乐的闭环中，用户获得了娱乐买了东西，平台和主播赚了钱；边玩边消费的模式，将传统电商的搜索式购物转变为直播电商的发现式购物。

在传统电商中，用户往往是有购物需求时才来使用产品，属于主动消费；用户的日活相对较低，所以，淘宝有了省钱消消乐、淘金币、芭芭农场等应用，本质上是先提高日活时长再促进转化率。

直播电商中，用户一开始时抱着看短视频、直播的娱乐需求来使用产品，而在娱乐过程中，受平台引导、主播的感染而“发现”购物需求，更侧重被动消费。目前，抖音、快手的内容分发策略都牢牢抓住了用户的喜爱偏好，让用户“抖音

5 分钟，现实 2 小时”，这对于电商领域来说具有极高的日活时长。购物也增加了平台的内容频道，扩大了用户的娱乐形式。整个闭环中，环环相扣，短视频和直播提高电商的购买转化率，电商购物丰富娱乐形式。

### 3. 短视频的消极影响

#### （1）观看内容良莠不齐，影响价值观

作为短视频的观看者，短视频的内容成为影响用户的最直接因素。短视频的内容存在有为吸引眼球博关注的荒诞闹剧、有为引流变现卖假货的虚假营销、还有用心险恶居心叵测的炒作造谣。如此种种，对于用户来说，视频内容不仅会误导欺骗他们，而且会潜移默化地影响用户的价值观。一旦错误的价值观形成，就会使得行为失范，心理失衡。

#### （2）分享视频形成圈层，影响交往能力

在本次调查中发现，一方面，用户每天使用短视频的时间大多集中在 1-2 小时，有少部分用户每天使用短视频的时间高达 3 小时以上，极易成瘾，而这是非常不利于人们的生活的。

另一方面，作为短视频的分享者，通过在网上点赞、评论、分享自己所喜爱的短视频，可以找到有共同话题的好友。这种网络世界中的交流不受现实生活中的物质、样貌、身份的束缚，会更易使人卸下防备，信任对方，加之有相同的兴趣爱好，自然相“见”恨晚，相聊甚欢，如今的“低头族”就很好地解释了这个现象：网络社交正逐渐取代了现实生活中的人际交往。

由于平台依赖算法为每位用户提供个性化推荐，那么会很容易形成一个个网络圈层。大家都习惯于在网络上向自己圈层中的人表达想法，会逐渐丧失现实生活中的人际交往能力，使得我们在面对面交流中缺乏表达欲望和互动需求。

#### （3）监管力度不足，侵犯用户权利

目前，短视频平台用户增长仍呈迅猛之势，但短视频平台的监管难以跟上步伐。例如，抖音、快手头部短视频平台分别在 2018 年、2021 年被“约谈”，主要原因为涉及软色情、侵权等问题。一方面，短视频平台存在着大量“擦边球”、“软色情”内容，不仅有损社会风气，更不利于青少年的健康成长；另一方面，近两年在短视频平台中流行的“街拍视频”，视频发布者有意或无意的行为同样侵犯路人的合法权益。

在短视频中，用户可以点击其他人的虚拟形象，浏览个人信息，如地区、年龄和个人签名，进入自己的主页，查看他们的公开作品。任何访问者都可以轻易

地获得所有信息，而不留下任何浏览标记和跟踪，在一定程度上，侵犯了用户的隐私权，甚至可能会引起诈骗事件。

## （二）建议

### 1. 优化短视频的内容质量与创意

在用户观看短视频中，发现短视频的质量良莠不齐且同质化严重，因此建议优化短视频的内容质量与创意。

在抖音账号以往的内容中，如若存在比较受欢迎的视频，可以尝试从粉丝用户的评论中发掘出他们为何偏爱本条视频的原因，分析视频中的哪一点比较容易引起他们的兴趣。这样我们在以后的同类视频内容创作中，就尽可能围绕目标用户的关注点来展开，从而获得目标用户的好感。

在制作短视频是应注意内容质量和创意，确保内容有吸引力和独特性，不断创造和尝试新颖的表现形式；并且针对目标受众的特点和兴趣制作内容，提高用户的观看体验和参与度；与用户互动，收集用户反馈和建议，不断优化内容和服务，增强用户粘性；关注短视频平台的算法变化，了解用户喜好和趋势，灵活调整内容策略。

### 2. 加强平台与政府监管力度

促进短视频的健康传播最直接的措施就是各大运营平台要时刻牢记政策法规，时刻牢记社会效益。短期上来看，对部分不宜传播的内容可以设置壁垒，甚至采用封号等措施，从长期来看还可以建立必要的内容审核机制，通过采取创作者实名登记制度和用户投诉举报制度来增强约束力，做到违法违规内容可自查、可溯源，发挥平台应有的自我净化功能。短视频平台可以使用大数据来提高人工智能技术来自动识别不正确的视频信息，并将这些不好的视频从平台上逐出。平台还可以对不规则的内容进行采样检查。

政府也应加强第三方监督，尽快完善相关法律法规且严格执行，为各方明确责任和义务，划清行为的边界线。

### 3. 用户应加强自我约束，理性对待短视频

网民作为网络文化的参与主体，应树立正确的价值观，自觉抵制恶俗内容。面对潜在的虚假、低俗信息时应自觉地分析、判断，做到不被不实信息所欺骗，不被低俗信息所诱惑。坚持自身的道德观念，只有这样，才可以不被短视频所操控，深迷其中，迷失自我。

并且在短视频平台中，应注意保护个人隐私。我们生活在大数据时代下，个

人信息难免会受到智能设备的监控和采集，需要提升自身的隐私信息保护意识，学会辨别信息，对相关信息被储存和挪用有一个清晰地认知，有节制地让自己的数据被记录。



## 参考文献

- [1] 蔡丽蓉. 新媒体背景下抖音短视频平台品牌营销策略[J]. 北方经贸, 2024(02): 90-94.
- [2] 马佳音. 2023 短视频: 创新求变, 稳步前行[J]. 国际品牌观察, 2024(02): 63-69.
- [3] 李欣颖. 基于信息质量理论的短视频平台用户购买意向研究[J]. 情报科学, 2023, 41(11): 103-110.
- [4] 张莹. 互联网与农产品直播营销策略分析[J]. 现代营销(下旬刊), 2024(01): 154-156.
- [5] 孔驹, 杨青, 祁钰. 短视频电商对大学生消费心理的影响—以抖音电商为例[J]. 科学咨询(科技·管理), 2024(01): 106-109.
- [6] 董肖宇. 基于 KOL 营销的时尚短视频内容策略[J]. 采写编, 2024(02): 181-183.
- [7] 蔡军, 曹亚方. 快手短视频平台的盈利模式研究[J]. 时代经贸, 2024, 21(02): 158-160.
- [8] 王文青, 乔立娟, 王建忠. 短视频运营对农户特色农产品种植收入的影响[J]. 西北农林科技大学学报(社会科学版), 2024, 24(02): 151-160.
- [9] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, 26(3): 7.
- [10] 岳梦阳. 方差分析方法及其应用研究[J]. 广东经济, 2016, 000(014): 289-290.
- [11] 李海娟. 数理统计在经济学中的应用[J]. 老字号品牌营销, 2022(08): 185-187.

## 附注 1

### 短视频的魅力与挑战—— 以带货、娱乐和新闻等多视角的调查研究

亲爱的朋友：

您好！我们是西安某大学在校学生。短视频是一种新兴的媒体形式，它具有形式多样、内容丰富、传播快速、互动强等特点，可以满足不同用户的多种需求。为了完成一项研究试题，需要了解您对您对短视频的一些情况和看法，现诚挚的邀请您参与此次调查活动，感谢您能抽出几分钟时间来参与本次问卷调查，请您根据您的实际情况提供您的看法和意见。能够获得您的看法和想法，我们感到十分荣幸，您所提供的所有信息我们只作为项目研究，并保证为您的信息保密。谢谢！

请阅读本项声明，然后回答问题：

基本打分情况包括：0-10，0分为无，分数越高，程度依次增加。

#### 一、基本信息情况

##### 1. 您的性别为

☐ 男

☐ 女

##### 2. 您的学历为

☐ 初高中及以下

☐ 本科

☐ 研究生

☐ 博士及以上

##### 3. 您的职业是

☐ 学生

☐ 教师

☐ 医生

☐ 律师

- ☐ 工程师
- ☐ 企业家
- ☐ 公务员
- ☐ 自由职业者
- ☐ 其他

4. 您的月收入是

- ☐ 3000 元以下
- ☐ 3000-5000 元
- ☐ 5000-8000 元
- ☐ 8000-12000 元
- ☐ 12000-20000 元
- ☐ 20000 元以上

## 二、短视频使用情况

5. 您是否自己制作过短视频，您制作短视频时的想法是什么？（多选）

- ☐ 记录日常，与朋友互动
- ☐ 宣泄自己的情绪
- ☐ 经营账号，获取利益
- ☐ 炫耀自己，获取关注
- ☐ 其他

6. 如果您使用短视频软件，您一般使用哪些平台？（多选）

- ☐ 抖音
- ☐ 快手
- ☐ 微信视频号
- ☐ 西瓜视频
- ☐ 哔哩哔哩（B 站）

- ☐ 小红书
- ☐ 火山小视频
- ☐ 微视
- ☐ 其他

7. 您一般在什么时间段使用短视频软件？（多选题）

- ☐ 早上起床前
- ☐ 上午
- ☐ 中午
- ☐ 下午
- ☐ 晚上
- ☐ 睡觉前

8. 您每天大约花多少时间在短视频软件上？（单选题）

- ☐ 15 分钟以下
- ☐ 15-30 分钟
- ☐ 30-60 分钟
- ☐ 1-2 小时
- ☐ 2-3 小时
- ☐ 3 小时以上

9. 您使用短视频软件的主要目的是什么？（多选题）

- ☐ 娱乐消遣
- ☐ 获取信息
- ☐ 学习知识
- ☐ 购物带货
- ☐ 展示自我
- ☐ 社交交友

☐ 其他

10. 您在短视频平台上更喜欢哪种互动方式？（单选）

☐ 私下交流

☐ 公开评论和分享

☐ 两者都喜欢

☐ 均不喜欢

11. 您最喜欢观看哪些类型的短视频内容？（多选题）

☐ 搞笑

☐ 美食

☐ 美妆

☐ 旅游

☐ 音乐

☐ 舞蹈

☐ 体育

☐ 游戏

☐ 宠物

☐ 情感

☐ 新闻

☐ 科技

☐ 教育

☐ 其他

12. 您对将来在短视频中使用虚拟现实（VR）或增强现实（AR）技术感兴趣程度？

☐ 0-10 分打分情况（0 分最低，10 分最高）

13. 您在使用短视频平台时，您关心以下哪些方面的个人隐私和数据安全问题？（多选题）

- ☐ 个人信息泄露
- ☐ 数据被滥用
- ☐ 第三方访问权限
- ☐ 地理位置隐私
- ☐ 消息和对话隐私
- ☐ 不关心

### 三、直播网购相关情况

14. 您每个月在直播带货中购买的商品类型主要为？（单选）

- ☐ 美妆类
- ☐ 食品类
- ☐ 服饰类
- ☐ 健身用品类
- ☐ 学习用品类
- ☐ 家用电器类
- ☐ 家居装饰品类
- ☐ 其他
- ☐ 从不购买

15. 您每个月在直播带货中的购买支出大致为？（单选）

- ☐ 0-100 元
- ☐ 101-300 元
- ☐ 301-500 元
- ☐ 501-1000 元
- ☐ 1000 元以上

16. 直播网购花销占您网购总花销的比例大约为（单选）

- ☐ 10%以下

☐ 10%-30%

☐ 31%-50%

☐ 51%以上

17. 您接触网络带货直播后您的网购消费额较之前增多的程度？

☐ 0-10 分打分情况（0 分从未改变，10 分增多程度最高）

18. 网络带货直播使您减少线下消费，增加线上消费的程度？

☐ 0-10 分打分情况（0 分从未改变，10 分增多程度最高）

19. 您选择在直播间购买商品的主要原因是？（单选）

☐ 优惠力度大

☐ 可以直观看到商品

☐ 相信主播的推荐

☐ 在直播中热烈的氛围下更有消费欲望

☐ 其他

20. 在观看直播带货时，您最关注的是？（单选）

☐ 带货主播的可信度

☐ 货品品牌的优劣

☐ 价格是否足够优惠

☐ 其他

21. 您在直播中购买过一次商品后，回购的意愿？

☐ 0-10 分打分情况（0 分最低，10 分最高）

22. 您觉得直播带货对消费者来说最主要的好处是什么？（单选）

☐ 通过与主播商家合作，商家可以让利出部分宣传费用给消费者

☐ 有主播背书，售后可追责

☐ 直播带货让一些冷门产品进入消费者视野，带来了更多选择

☐ 在直播间可通过实时交流获取商品信息

☐ 减少消费者时间和金钱成本

☐ 其他

23. 您认为直播带货的主要弊端？（单选）

☐ 虚假宣传，数据造假

☐ 产品质量堪忧

☐ 消费者权益难

☐ 容易泄露个人隐私

☐ 传播不良价值观，进行低俗直播

☐ 其他

24. 您拒绝在直播间购买商品的主要原因？（单选）

☐ 认为直播中的商品不靠谱

☐ 观看直播太耗费时间

☐ 本身没有在网上购物的习惯

☐ 不信任主播与电商的合作

☐ 其他

25. 您认为当下电商直播首要应该做出哪些改进？（单选）

☐ 完善相关法规（透明主播相关收入等）

☐ 要求平台加强管理（提高入驻门槛，建立黑名单等）

☐ 加大监管与处罚力度

☐ 其他

26. 您对直播带货的看法？（单选）

☐ 是一种正常的商业现象，使购物更便捷

☐ 并不认为这是一个优质行业

☐ 不太了解，保持中立态度

27. 您对直播带货的未来看法？（单选）



- ☐ 高速发展
- ☐ 发展速度放缓
- ☐ 发展规模跟现在差不多
- ☐ 开始萎靡，陷入低谷

#### 四、短视频的影响

28. 您因为短视频的内容而产生情绪波动？

- ☐ 0-10 分打分情况（0 分最低，10 分最高）

29. 您会因为短视频的内容而改变自己的想法或观念的程度？

- ☐ 0-10 分打分情况（1 分程度最低，10 分程度最高）

30. 您会因为短视频的内容而购买相关的商品或服务的意愿？

- ☐ 0-10 分打分情况（0 分意愿最低，10 分意愿最高）

31. 您是否会因为短视频的内容而参与相关的活动或话题的意愿？

- ☐ 0-10 分打分情况（0 分意愿最低，10 分意愿最高）

32. 您对短视频中内容的真假、是非所持的态度？（单选）

- ☐ 自我甄别
- ☐ 不管，纯粹消遣
- ☐ 从不怀疑
- ☐ 从不相信

33. 您对短视频的整体满意度是多少？

- ☐ 0-10 分打分情况（0 分满意度最低，10 分满意度最高）

34. 您认为短视频有哪些缺点需要改进？（多选题）

- ☐ 质量低
- ☐ 内容雷同
- ☐ 传播误导
- ☐ 侵权多

☐ 欺诈多

☐ 其他

35. 您对短视频未来的发展有什么建议或期待（或者短视频的您眼中的优缺点）？（开放题）

## 附注 2

程序

```
1.import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
from sklearn.preprocessing import MinMaxScaler
warnings.filterwarnings("ignore")#忽略警告
plt.rcParams['font.sans-serif']='SimHei'#设置字体为 SimHei
plt.rcParams['axes.unicode_minus']=False#解决负号 "-" 显示异常
data = pd.read_excel(r"（影响）打分数据.xlsx")
data
2.import matplotlib.pyplot as plt
# 原因和频率数据
reasons = ['优惠大', '可以直接看到商品', '信任主播', '直播中更有消费欲望', '其他']
frequencies = [85, 53, 26, 25, 12]
# 创建柱状图
font = {
    'family': 'STSong',
    'serif': 'Times New Roman',
    'weight': '660',
    'size': 12
}
plt.rc('font', **font)
plt.figure(figsize=(10, 6), dpi=300)
plt.rcParams['font.sans-serif'] = 'SimHei'
plt.rcParams['axes.unicode_minus'] = False
```

```

plt.rcParams["font.weight"] = "bold"
# 计算频率占比
total = sum(frequencies)
percentages = [freq / total * 100 for freq in frequencies]
bars = plt.bar(reasons, frequencies, color='skyblue')
# 添加频率占比标签
for bar, percentage in zip(bars, percentages):
    height = bar.get_height()
    plt.text(bar.get_x() + bar.get_width() / 2, height + 1, f'{percentage:.1f}%',
ha='center', va='bottom')
# 添加标题和标签
plt.title('选择在直播间购买商品的原因')
plt.xlabel('原因')
plt.ylabel('频率')
# 显示柱状图
plt.show()
3.# 创建柱状图
font = {
    'family': 'STSong',
    'serif': 'Times New Roman',
    'weight': '660',
    'size': 12
}
plt.rc('font', **font)
plt.figure(figsize=(10, 6), dpi=300)
plt.rcParams['font.sans-serif'] = 'SimHei'
plt.rcParams['axes.unicode_minus'] = False
plt.rcParams["font.weight"] = "bold"
import matplotlib.pyplot as plt
# 原因和频率数据
reasons = ['商品不靠谱', '耗费时间', '没有网购习惯', '不信任主播', '其他']
frequencies = [78, 63, 35, 18, 7]
# 计算频率占比
total = sum(frequencies)
percentages = [freq / total * 100 for freq in frequencies]
bars = plt.bar(reasons, frequencies, color='skyblue')
# 添加频率占比标签
for bar, percentage in zip(bars, percentages):

```

```

        height = bar.get_height()
        plt.text(bar.get_x() + bar.get_width() / 2, height + 1, f'{percentage:.1f}%',
ha='center', va='bottom')
# 添加标题和标签
plt.title('拒绝在直播间购买商品的原因')
plt.xlabel('原因')
plt.ylabel('频率')
# 显示柱状图
plt.show()
4.font = {
    'family': 'STSong',
    'serif': 'Times New Roman',
    'weight': '660',
    'size': 15
}
plt.rc('font', **font)
plt.figure(figsize=(7, 5), dpi=300)
plt.rcParams['font.sans-serif'] = 'SimHei'
plt.rcParams['axes.unicode_minus'] = False
plt.rcParams["font.weight"] = "bold"
# 计算相关性矩阵
cor_matrix = data.corr()
# 将相关性矩阵转换为数据框
cor_df = cor_matrix.unstack().reset_index()
cor_df.columns = ['Var1', 'Var2', 'Correlation']
plt.scatter(data=cor_df, x='Var1', y='Var2', c='Correlation', cmap='viridis', s=400,
alpha=0.8)
# 添加相关性系数文本标签
for _, row in cor_df.iterrows():
    plt.text(row['Var1'], row['Var2'], f'{row["Correlation"]:.2f}', ha='center',
va='center', fontsize=10, color="black")
plt.colorbar(label='Correlation')
plt.xticks(rotation=0, ha='right')
plt.gca().spines['top'].set_visible(True)
plt.gca().spines['right'].set_visible(True)
plt.grid(True, linestyle='--', alpha=0.5)
plt.tight_layout()
plt.show()
5.import pandas as pd

```

```

import numpy as np
from sklearn.feature_selection import SelectKBest, f_classif
# 假设我们要为分类问题确定变量的重要性
# 我们需要一个目标变量，这里没有提供。我们将创建一个虚拟目标变量来
演示
# 在实际情况中，目标变量应该是分类问题的实际标签
target = data.iloc[:, -1]
df = data.iloc[:, :-1]
# 使用 SelectKBest 选择得分最高的 k 个特征
selector = SelectKBest(score_func=f_classif, k='all')
selector.fit(df, target)
# 获取每个特征的得分
scores = selector.scores_
# 打印每个变量的得分
for column, score in zip(df.columns, scores):
    print(f"The score for {column} = {round(score,4)}")
score = pd.DataFrame()
score["ID"] = df.columns
scores = pd.DataFrame(scores)
score["得分"] = scores
score
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
# 假设你有一个 DataFrame，包含 ID 和得分
# 这里假设 DataFrame 叫做 df，包含两列：ID 和 得分
# 假设你已经从某个地方得到了 scores 列表
# 提取 ID 和 Score 列
ids = score['ID']
scores = score['得分']
# 设置雷达图的角度和分数
angles = np.linspace(0, 2 * np.pi, len(ids), endpoint=False).tolist()
scores = scores.values.tolist()
scores += scores[:1]
angles += angles[:1]
# 创建雷达图
fig, ax = plt.subplots(figsize=(6,6), subplot_kw=dict(polar=True), dpi=300)
ax.fill(angles, scores, color='#7ac7e2', alpha=0.6)

```

```

ax.plot(angles, scores,'o-', color='#54beaa',linewidth=2)
# 在每个数据点上添加具体的得分数值（保留四位小数）
for angle, score, label in zip(angles[:-1], scores[:-1], ids):
    ax.text(angle, score, '{:.4f}'.format(score), color='black', fontsize=10,
ha='left')
ax.set_yticklabels([])
ax.set_xticks(angles[:-1])
ax.set_xticklabels(ids)
ax.set_title("原始评分雷达图")
plt.show()
The score for Q4 = 1.3399
The score for Q8 = 1.6866
The score for Q17 = 14.9079
The score for Q18 = 10.0357
The score for Q21 = 3.907
The score for Q28 = 10.6584
The score for Q29 = 16.4468
The score for Q31 = 22.7991
The score for Q33 = 10.0799
6.target = data.iloc[:, -1]
df = data.iloc[:, -1]
# 使用 SelectKBest 选择得分最高的 k 个特征
selector = SelectKBest(score_func=f_classif, k='all')
selector.fit(df, target)
# 获取每个特征的得分
scores = selector.scores_
# 打印每个变量的得分
for column, score in zip(df.columns, scores):
    print(f"The score for {column} = {round(score,4)}")
score = pd.DataFrame()
score["ID"] = df.columns
scores = pd.DataFrame(scores)
score["得分"] = scores
score
# 创建一个 MinMaxScaler 对象
scaler = MinMaxScaler()
# 将得分进行归一化处理
scores = score['得分']
scores_normalized = scaler.fit_transform(scores.values.reshape(-1, 1))

```

```

score["得分"] = scores_normalized
score
The score for Q4 = 1.3399
The score for Q8 = 1.6866
The score for Q17 = 14.9079
The score for Q18 = 10.0357
The score for Q21 = 3.907
The score for Q28 = 10.6584
The score for Q29 = 16.4468
The score for Q31 = 22.7991
The score for Q33 = 10.0799
7.ids = score['ID']
scores = score['得分']
# 设置雷达图的角度和分数
angles = np.linspace(0, 2 * np.pi, len(ids), endpoint=False).tolist()
scores = scores.values.tolist()
scores += scores[:1]
angles += angles[:1]
# 创建雷达图
fig, ax = plt.subplots(figsize=(6,6), subplot_kw=dict(polar=True),dpi=300)
ax.fill(angles, scores, color='#7ac7e2', alpha=0.6)
ax.plot(angles, scores,'o-', color='#54beaa',linewidth=2)
# 在每个数据点上添加具体的得分数值（保留四位小数）
for angle, score, label in zip(angles[:-1], scores[:-1], ids):
    ax.text(angle, score, '{:.4f}'.format(score), color='black', fontsize=10,
ha='left')
ax.set_yticklabels([])
ax.set_xticks(angles[:-1])
ax.set_xticklabels(ids)
ax.set_title("评分归一化后的雷达图")
plt.show()
8.import jieba
from matplotlib import pyplot as plt
from wordcloud import WordCloud
from PIL import Image
import numpy as np
path = r'C:\Users'
font = r'C:\Windows\Fonts\STXINGKA.TTF'
def tcg(texts):

```

```

        cut = jieba.cut(texts) # 分词
        string = ''.join(cut)
        return string
text = (open(path + r'\绘图.txt', 'r', encoding='utf-8')).read()
string = tcg(text)

img = Image.open(path + r'\地图.jpg')
img_array = np.array(img)
stopword = ['视频','无','的','更加','更','多','没有','越来越','加大','不要','和','
暂无','是','有','以','应该','也','一点','可以','再','吧','为']
wc = WordCloud(
    background_color='white',
    width=1000,
    height=800,
    mask=img_array, # 设置背景图片
    font_path=font,
    stopwords=stopword,
)
wc.generate_from_text(string)
plt.figure(dpi=1000) # 设置 dpi 参数
plt.imshow(wc, interpolation='bilinear')
plt.axis('off')
plt.show()
wc.to_file(path + r'\词云图.jpg', dpi=1000)

```