

**本科毕业设计[论文]**

**喜马拉雅FM免费增值策略对用户留存的影响的实证研究**

|  |  |
| --- | --- |
| 院 系 | 管理学院 |
| 专业班级 | 信息管理与信息系统1802班 |
| 姓 名 | 刘棫欣 |
| 学 号 | U201812791 |
| 指导教师 | 赵玲 |

2022年 6 月 5 日

**学位论文原创性声明**

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包括任何其他个人或集体已经发表或撰写的成果作品。本人完全意识到本声明的法律后果由本人承担。

作者签名： 年 月 日

**学位论文版权使用授权书**

本学位论文作者完全了解学校有关保障、使用学位论文的规定，同意学校保留并向有关学位论文管理部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权省级优秀学士论文评选机构将本学位论文的全部或部分内容编入有关数据进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于 1、保密 囗 ，在 年解密后适用本授权书。

2、不保密 囗 。

（请在以上相应方框内打“√”）

作者签名： 年 月 日

导师签名： 年 月 日

# **摘 要**

本研究从知识付费移动音频的角度出发，旨在研究免费增值策略对用户留存的影响并且探究出最优策略。研究爬取喜马拉雅FM的所有在线音频、声音、用户的信息作为数据源，通过数据筛选与变量提取进行模型选择。模型构建中，本研究将专辑与主播的特征作为控制变量，通过逐层回归，探究免费增值策略对专辑在收费节点用户留存的影响。在得到影响程度的基础上，加入收费节点的二次项进一步回归和进行分样本回归，尝试寻找最优免费增值策略。

结果显示：免费增值策略的收费节点和专辑中的章节平均价格对专辑的收费节点处的用户留存率有显著的负向影响，是否支持VIP服务对专辑收费节点处的用户留存率即付费转化率有显著的正向影响。针对收费节点对于付费转化率产生显著负向影响这一结果不符合研究假设的现象，本文进行了进一步地分析，发现由于主播的高专业度、高频更新以及粉丝的高粘性，可能会抵消收费节点靠后带来的价格变化对付费转化率的抑制作用；同时，依据损失厌恶，有声书的免费部分如果展现的是消费者最想看到的，而付费部分隐藏消费者厌恶的或者预期没有那么高的，用户会更愿意做出购买行为。

综合结果来看，本研究为特定主播发布特定专辑的免费增值策略的选择提供了指导建议，具体措施建议为根据需要选择是否支持VIP服务，若支持，建议收费节点尽量靠前，若不支持，建议收费节点为30%，并且制定合理但较市场价低的章节价格。

**关键词：**付费音频；免费增值模式；付费转化率；数据爬取；回归分析

# 

# Abstract

From the perspective of knowledge paid mobile audio, the study aims to study the impact of Free-Premium Strategy on user retention and explore the optimal strategy. The study crawls all online audio, sound, and user information of Himalaya FM as a data source, and selects models through data screening and variable extraction. In the construction of the model, the characteristics of albums and anchors are used as control variables, and the choosing model is layer-by-layer regression. Based on this, further regression and sub-sample regression are carried out by adding the quadratic term of the Charging Point, trying to find the optimal strategy.

The results show that the average price of the Charging Point and the sound in the album has a significant negative impact on the user retention rate at the Charging Point of the album, whereas whether the VIP service is supported has a significant positive impact. In view of the fact that the results of the Charging Point have a significant negative impact on the paid conversion rate, which is not in line with the research hypothesis, this paper further analyzes and finds that due to the high professionalism of the anchor, high frequency updates and the high stickiness of the fans, it may offset the inhibition effect of the price change brought about by the Charging Point on the paid conversion rate. At the same time, according to the loss aversion, if the free part of the Audiobook shows what the consumer wants to see the most, and the paid part hides the consumer's disgust or expectation is not so high, users will be more willing to make purchases.

Above all, this study provides suggestions for the setting of Free-Premium Strategy for specific anchors when releasing specific albums. The specific measures are recommended to choose whether to support VIP services according to needs, if supported, it is recommended that the Charging Point should be as far ahead as possible, if not supported, it is recommended that the Charging Point should be 30%, and the track price should be reasonable but lower than the market price.

**Key Words：**Paid Audio; Free-Premium Mode; Conversion Rate; Data Crawler; Regression Analysis

**目 录**

**[摘要](#_Toc27728)** [I](#_Toc27728)

**[Abstract](#_Toc5884)** [II](#_Toc5884)

**[1 绪论](#_Toc4125)** [1](#_Toc4125)

[1.1 研究背景和问题 1](#_Toc15422)

[1.2 研究目的和意义 2](#_Toc10367)

[1.3 研究思路与研究方法 3](#_Toc11327)

[1.4 研究创新与难点 4](#_Toc14371)

**[2 文献综述](#_Toc15615)** [6](#_Toc15615)

[2.1 知识付费平台的定义和类型 6](#_Toc11012)

[2.2 知识付费平台的用户行为研究 7](#_Toc18586)

[2.3 知识付费平台的免费增值策略研究 9](#_Toc24726)

[2.4 文献评述 11](#_Toc3929)

**[3 研究场景与数据描述](#_Toc1050)** [13](#_Toc1050)

[3.1 喜马拉雅FM平台的免费增值策略 13](#_Toc13098)

[3.2 数据获取与清洗 15](#_Toc3248)

[3.3 变量描述 19](#_Toc24560)

[3.4 变量的描述性统计 21](#_Toc19174)

**[4 研究设计](#_Toc3922)** [25](#_Toc3922)

[4.1 研究假设 25](#_Toc22484)

[4.2 模型构建 27](#_Toc3050)

[4.3 相关分析 28](#_Toc6663)

**[5 免费增值策略对用户留存影响的实证分析](#_Toc14083)** [29](#_Toc14083)

[5.1 逐层回归分析 29](#_Toc30404)

[5.2 模型检验 30](#_Toc27985)

[5.3 结果解读 33](#_Toc14958)

[5.4 关于结果不符合假设的进一步分析 33](#_Toc2619)

[5.5 免费增值最优策略的探索 35](#_Toc22804)

**[6 结论与启示](#_Toc18104)** [38](#_Toc18104)

[6.1 研究结论 38](#_Toc28784)

[6.2 研究启示 39](#_Toc22476)

[6.3 研究不足与展望 41](#_Toc32479)

**[致谢](#_Toc4642)** [42](#_Toc4642)

**[参考文献](#_Toc7961)** [43](#_Toc7961)

**[附录一 数据获取部分代码](#_Toc17247)** [46](#_Toc17247)

**[附录二 取声音title中编号的部分代码](#_Toc22200)** [49](#_Toc22200)

**[附录三 相关分析完整结果](#_Toc14111)** [55](#_Toc14111)

# 1 绪论

## 1.1 研究背景和问题

### 1.1.1 研究背景

在线音频需求巨大，2021年，iiMedia Research（艾媒咨询）数据资料显示，中国移动音频的用户规模从2019年开始一直保持着快速发展，并且预计到下一年，移动音频用户的规模将达到6.9亿。在线音频伴随科技发展产生，在国内起源以及发展至今至少拥有了10年的历史。它起源于2011年蜻蜓FM的创立，后来喜马拉雅和荔枝FM接着上线，直至今日，在线音频对人类生活产生着举足轻重的影响，并且拥有的广大的用户群体与极强的用户粘性。中国移动音频行业发展至今，拥有海量的用户基础，社交音频热度为行业带来更多关注，用户和市场规模不断增长。

但是由于没有立足用户需求，在线音频行业依旧存在大量问题，尤其是知识变现的问题。2021年Clubhouse为音频社交带来更多的关注，进一步发现在线音频行业即使拥有如此庞大的用户群体以及市场需求，却依旧存在许多可提升的问题，如版权贵、收费难等等。尤其是收费难的问题，知识付费的时代下，大部分人能够接受为知识买单，为内容创作者的的劳动成果买单，但是有声音频的收费模式并没有立足用户需求与用户体验。再加上版权问题，有时候平台甚至会出现，一张专辑的价格远远大于实体书的价格。这些不合理的收费标准与策略使得用户体验降低，用户流失率高，从而引起主播的流失问题，带来一系列的恶性循环。

为了解决收费难的问题，从而改善收益，在线音频平台尝试将免费增值模式作为营销策略，但是效果并不理想。免费增值模式的目的是营销，免费提供产品或服务始终是吸引客户的非常有效的方式，在互联网上非常有效。即使免费用户不付费，用户交互可以提供广泛的口碑。而且，提供免费的东西，促进网络效应即增加使用特定产品或服务的人，从而提高产品的价值是建立受众的最简单方法之一。

但是免费增值模式的实施也面临着许多困难与风险，由于低价，服务与产品需要大量的付费用户才能够获得收益，同时也需要大量的免费用户基础才能够在目前较低的付费转换率中发展成足够的付费用户。目前只有大约1%的人会购买溢价优惠（行业基准为5%）。但是从长远来看，免费增值模式更能吸引消费者，通过免费创建一个客户群，以说服该免费客户进行付费客户的转化，是未来更有潜力的商业模式，也是众多行业的关注点。

### 1.1.2 研究问题

综合以上背景，有声音频内容创作平台前期发展迅速，但是随着内容付费、知识付费的助推产生了新的问题，例如收费难，盈利成问题等；为了解决上述问题，平台尝试将免费增值模式作为营销策略，但是因为不规范的收费标准，用户诟病收费高，收费不合理等问题，所以效果不佳。

行业发展至此，虽然依旧拥有巨大的市场需求，但是行业内部发展前景却不容乐观。分析发现，平台因为垄断，没有根据用户的需求以及市场价格进行营销策略规范化，虽然前期拥有巨大的收益，但是随着时间的推移，用户留存不理想，用户满意度低，主播的收益不可观等问题暴露。

综上，本研究尝试去解决的问题是平台需要作出如何的改变才能够与主播、用户达成三赢，让越来越多的优质主播愿意加入，并且有动力创作优质作品，让更多的用户愿意为内容付费并且提高用户满意度。

## 1.2 研究目的和意义

### 1.2.1 研究目的

本研究目的是通过优化免费增值模式下的收费策略从而优化有声音频平台的发展，通过聚焦免费增值模式，寻找使得收费节点处用户留存最优的策略，从而让作品的价值符合用户的预期，让用户更加自愿并且乐意做出付费行为，提高付费转化率，从而进一步改善主播的收益与平台的运营，形成更良好的生态圈。

具体而言，本研究以免费增值模式的收费策略为切入点，从作品层面进行分析，对收费策略中的收费节点，收费价格，收费模式对付费转化率的影响进行实证研究，并且构建模型尝试寻找最优收费策略。在本研究中，笔者控制了专辑本身的特征以及主播的特征因素（目前已有学者对专辑与主播的特征进行了探究），着重对收费策略的影响进行研究，旨在为平台提供建议，为特定主播发布特定专辑时提供个性化的收费策略推荐，包括收费价格，收费模式，收费节点，从而使得专辑收费节点处的用户留存率即付费转化率更高，更有效地改善平台的收益。

### 1.2.2 研究意义

本研究在理论上，可以提高在线音频作品的用户留存，一定程度上提高付费转化率；从而进一步改善在线音频内容创作平台与内容创作者的收益，实现用户、主播、平台的三赢。在知识付费无限的时代红利下，在线音频的发展需要立足于用户的需求。但是在线音频开始采取收费模式这一行为对用户的影响，一直没有研究涉及，具体的影响程度还需要被量化。而且本研究对于不同收费模式下，在线音频的用户粘性以及用户行为进行研究，在一定程度上可以帮助在线音频平台更好地发展以及迭代。尤其是对在线音频本身的定价以及平台的VIP定价模式有重要的作用。

本研究在实际应用中，一定程度上可以促进知识付费、知识变现。知识付费是版权意识的觉醒，是新时代知识即服务文化的衍生物。而有声音频则是基于电子书且最受欢迎的知识付费内容，针对有声音频研究，可以促进知识付费行为的研究，改善用户满意度，从而促进知识付费文化的发展，促进知识变现。

## 1.3 研究思路与研究方法

### 1.3.1 研究思路

通过抓包爬虫获取喜马拉雅FM的所有数据，并且依据初步描述性统计分析对数据集进行筛选；确定数据集之后，对专辑中的章节名称进行排序；然后对这些免费增值模式的专辑，用收费节点前后的播放量变化来衡量用户留存即付费转化率，量化付费的影响，最后通过实证分析进行结果呈现，对结果与假设进行解释与进一步分析。

在此过程中，着重研究喜马拉雅FM中的付费专辑，对主播特征以及专辑特征进行控制，控制这些因素的影响，对专辑中的免费与付费章节的用户留存与定价、收费节点、收费模式建立模型。根据回归结果进行进一步分析解释，探索不同收费模式、不同价格、不同收费节点对付费转化率的影响；为在线音频单个专辑或者章节本身的定价以及平台的VIP定价模式提供参考。

通过建模量化，结合平台，我们可以得到各个因素对付费转化率的影响程度，然后对平台优化，使得用户拥有更好的体验，让大部分用户能够自愿为知识付费，也使得有声音频价值更好地符合用户预期。本研究可以帮助在线音频本身的定价以及平台的VIP定价模式提供参考，让有声音频行业更好地结合知识付费发展。

### 1.3.2 研究方法

文献调查法：本研究通过知网、Springer、IEEE Xplore、INFORMS、谷歌学术等网站进行文献搜索，对本研究所探究的领域目前的国内外研究现状以及研究防范进行了详细全面的了解，对以移动音频内容创作平台、免费增值模式或者知识付费为研究对象的研究，以及以用户留存、用户流失、付费转化率为研究因素的研究，都进行了广泛的查阅与梳理，涵盖不同的行业，不同的研究方法，研究角度，从而对研究假设的提出，研究问题的量化提供了指导意见以及理论基础。

实证分析法：从众多可能影响付费转化率的因素中，聚焦收费模式、收费节点、收费价格，控制其余因素进行逐层回归的实证分析，探究这些因素是否会对用户行为造成影响及其影响的程度，通过构建模型对每一个因素的影响程度进行量化，根据不同的结果结合喜马拉雅平台的特点进行解释，提出方案。

## 1.4 研究创新与难点

### 1.4.1 研究创新

选题立意点新颖：本研究的核心是移动音频创作平台的免费增值模式下收费标准对付费转化率的影响，目前尚未有研究涉及。本研究内容属于创新，将目前最有效的免费增值模式进行深入研究，可以更全面的挖掘该模式的特点；同时关注付费转化率是发展用户的长期选择，付费转化率衡量作品的价值，是提升经济效益最直观最长远最有效的指标；而且，针对作品与内容创作者本身的特点，收费标准是用户更关注的因素，研究的所有重点都明确并且重要，在现有研究未涉及的问题踏出了新的一步。从专辑层面分析，可以减少因为市场变化而导致的用户群体的变化对付费转化率的影响，达到更长期的持续化发展。

研究范围创新：传统的知识付费营销策略研究几乎不涵盖免费增值模式，只有少数对免费试用策略进行过研究，同时少数之于传统的免费增值模式的研究几乎都选择的是游戏行业或者软件行业的免费版本—付费版本模式，而本研究选择在线音频内容创作平台的免费—付费模式作为免费增值模式，同时去除免费试听模式的干扰，研究范围聚焦专辑层面的免费—付费，立足于现实生活的用户体验，能够更准确的反映用户对免费增值的态度。

### 1.4.2 研究难点

1）数据处理难

本研究选择了喜马拉雅FM这一有声音频创作行业市场占有率高达75%以上的平台，从中爬取我们所需要的数据，并且进行数据筛选，从67万张专辑，对应上亿级个声音，10万级个主播的数据集中进行筛选与清洗，数据量庞大，且数据变量繁杂，工作量多。

2）变量界定难

（1）付费转化率的界定

不同学者研究不同行业以及不同问题时给予了不同的界定，但是对于有声音频维度，付费转化率的界定涉及到范围的选取，备选方案有付费节点前后的用户留存率，免费付费章节的用户播放量比值，付费节点的用户留存率相对于专辑本身的自然留存率的变化等。

（2）收费标准的界定

喜马拉雅FM提供5种收费模式，但是按照粒度与划分方式不同，这5种收费方式又可以分成三种划分方式。因此选取哪一种收费模式的界定作为被解释变量需要多角度多方面的衡量确定。

（3）收费节点的界定

收费节点可以采用相对位置或者绝对位置进行界定，相对位置可以采用收费所在集的位置百分比，以及收费的时长位置，绝对位置有收费的集数以及收费的个数。综上，在界定变量时需要考虑全面并且对所有可能的备选方案纳入分析，使得模型更准确并且稳健。

# 2 文献综述

本部分将从知识付费平台的定义与类型、知识付费平台的用户行为研究，知识付费平台的免费增值策略研究三个维度进行文献调查，进行国内外研究现状调查与呈现。

## 2.1 知识付费平台的定义和类型

知识付费平台比较广泛，有喜马拉雅、腾讯课堂、百学习、MOOC、网易云课堂、千聊等的流量型平台；有百度问答、知乎、知识星球、分答等的分享型平台；以及小鹅通、短书等的SAAS工具。其中移动音频创作平台是指以有声音频的形式进行知识传播的内容付费平台，是目前受众最广的知识付费类型平台，主要平台有喜玛拉雅FM、蜻蜓FM、荔枝FM等。

国内外针对知识付费平台的研究主要聚焦用户满意度的影响因素，宗利永、夏子惠[1]以订阅式知识付费平台用户评论为研究对象，其中选取的平台为以知乎live为代表的订阅式知识付费平台。通过建立用户评论挖掘系统框架，结合用户打分及评论文本情感分析结果。结果表明影响订阅式知识付平台用户体验的因素分别是：内容数量、内容类型、内容质量、舒解是影响积极用户体验评价的因素，即激励因素；系统可靠性、广告、价格是影响消极用户体验评价的因素，即保健因素；收听功能为激励—保健因素。

国内外针对知识付费平台中的移动音频创作平台，例如喜马拉雅，懒人听书的研究很少，已有的研究基本针对在线音频制作平台的营销策略、盈利模式与知识付费的发展进行案例讨论。其中敖雅倩[2]对以喜马拉雅FM为例的移动在线音频行业的盈利模式与盈利策略进行了分析，总结了其盈利特点并且结合市场现状给出了建议。李琳[3]对喜马拉雅FM平台的付费内容进行了详细的模式研究，使用多种研究方法，发现喜马拉雅遵循从付费内容生产—运营—消费的营销，也总结了用户需付费购买的内容在音频平台中遵循的整体传播路径，并且针对目前的一些问题提出优化措施。针对移动音频创作平台的知识付费，有少数研究进行了探索，其中张思嘉[4]对懒人听书的知识付费模式营销战略进行了分析，从市场需求宏观环境分析了该模式的发展，同时对用户进行了调研，从多方面分析最终提出了营销策略改良的思考。类似的，丁翱林[5]对喜马拉雅FM的营销策略进行了案例分析，针对喜马拉雅的知识付费特点，提出了目前知识付费平台的营销现状及存在的三大问题，通过分析喜马拉雅FM这一知识付费平台成功案例，从喜马拉雅FM的发展以及现状出发，对喜马拉雅的营销策略进行了详细的分析，总结了五个喜马拉雅在营销策略上根据市场变化，互联网发展进行的创新调整，对其他知识付费平台提供了一些创新思路。蒋岩汐[6]以知识付费产品为研究对象，包括知乎严选，有声书等知识付费产品，先简要分析了知识付费产业的现状，并由此介绍了知识付费产品的国内外研究现状，以商品二因素为基础介绍了知识付费产品的价值与使用价值，为后文的研究奠定相关的理论基础。通过构建知识付费产品价值评估模型并分析各阶段知识付费产品的形成流程和对得到平台进行内容分析的方式，对知识付费产品的价值形成机制进行了研究。

也有部分研究对于其他现象对知识付费行为造成的影响进行了研究，比如用户交互行为，用户学习行为等。其中王雨萍[7]将用户交互与知识付费结合，对虚拟社区用户的知识付费行为的因素进行了研究。该研究[7]表明知识经济与虚拟社区的发展让人们拥有更多的机会与选择，在虚拟社区的交互中，用户能够在进行知识付费前对产品或者服务有更加客观全面的了解，总结到知识付费有着巨大的上升潜力。罗钦芳[8]以知乎live为研究对象，对消费者知识付费产品的购买决策的影响因素进行了探究，发现产品价格对销量的影响是非线性且复杂的，并且日销量不仅受外部因素如日交易记录影响，还收内部因素如产品本身的特质影响。刘征驰等[9]从知识的隐匿性给消费者以及知识提供者带来的交易双方信息不对称，面临极高交易风险的问题出发，将网络中的社群交互因素纳入知识付费两期动态定价博弈模型，深入探讨社群学习机制对最优定价策略的影响，发现消费者在决策过程中对于社群交互的依赖很高，这样也可以减少知识价值隐匿性带来的购买风险[9]。刘征驰等[10]还通过改善非授权分享给知识付费平台以及知识创作者带来的定价分歧，给知识付费平台的运营提供了更据参考性的策略。

## 2.2 知识付费平台的用户行为研究

在上文中知识付费平台的理论研究之外，学者们针对知识付费平台用户行为进行了进一步的实证研究，面广但尚不深入，其中涉及的知识付费的平台有流媒体、知乎、得到、喜马拉雅等，涉及的用户行为有流失/留存，知识付费行为等。

之于平台的用户留存与流失的研究中，学者们将用户留存率定义为判断产品或服务的价值最重要的标准，用户是否愿意留下来，取决于产品的价值是否达到用户的预期。不同行业的用户留存率的提升，或是避免用户流失，以及提高付费转化率，发展高价值用户都有学者进行了较为深入的探讨。游戏行业中，部分研究[11-12]从游戏行业的用户层面对流失进行了预测。通过机器学习、特征工程、随机森林等技术对用户的流失几率进行预测，并评估预测性流失模型可带来的业务影响。类似的，Sangaralingam等[13]提出了一种基于可扩展数据混搭的机器学习算法，对高价值的客户群进行了识别，并使用独立的客户识别现有客户的潜在不利过渡。除了通过机器学习的方法进行用户层面的流失预测，部分研究通过其他方法，分析了不同行业的流失原因。柯金燕[14]通过深度访谈对移动游戏体验者的社交体验、满意度和留存意向影响因素进行了探究。聚焦知识付费平台中的用户流失/留存研究，吴爽[15]通过问卷调查，对知识付费平台高知女性用户流失行为进行影响因素的探究，问卷调查对象为高知女性，从用户层面以及用户需求角度分析了用户选择终止服务的原因。周勤琪[16]以短视频社交媒体的用户为研究对象，流失影响因素研究社交媒体参与度中的内容消费、内容生产、社交参与、社交领袖均与用户卸载流失有显著影响。不同行业的用户流失研究虽然方法不一，但研究思路都是从用户层面去分析用户流失原因以及预测流失。

之于用户在付费节点的流失与留存即知识付费行为，大部分研究关注知识付费行为的影响因素，有从用户本身层面进行分析用户需求，也有从内容特征进行分析，还有从内容发布者或者创作者的特点进行分析。聚焦知识付费行为的研究中，学者们进一步针对不同的角度与维度进行了许多影响元素的研究。Su等[17]选择以内容创作者层面，对社交媒体包括微信，微博，Facebook等不同内容发布者对用户内容付费的意愿影响进行了协方差分析，探究了社交媒体的用户付费意愿与内容创作者特征的关系。结果发现得到发布者知名与否的情况下，被调查者对付费PGC与UGC的购买意愿不同，与此同时，该付费意愿还受到内容发布者的态度和先前态度的显著影响。杨东红等[18]从知识付费移动音频的角度，研究用户对有声阅读平台中付费音频的付费行为。根据相关性分析发现传播者性别对用户知识付费行为没有影响，在控制平台变量后，付费移动音频的价格、时长和信息展现形式与付费行为成反比，试听、传播者简介长度、作品量、粉丝数和认证信息长度与知识付费行为成正比。邓胜利等[19]对不同类型以及不同程度的交互行为对于预测知识付费用户的贡献度及其变化趋势进行了研究。以知乎社区爬取了 400 万用户社会交互行为作为数据，分析结果发现，用户和其他用户的交互影响大于用户和平台交互的影响。王雪莲等[20]选取“喜马拉雅”平台的18个节目样本，利用爬虫技术抓取用户评价数据，并采用内容分析方法对数据进行分析，从用户维度进行需求分析，发掘并构建了在线知识付费节目的顾客体验价值维度体系。

但是，虽然学者们进行了大量的实证研究，目前针对知识付费意愿的研究结果存在不一致的问题，Li等[21]的研究结果表明经济因素是对于用户在线知识付费的行为影响不显著，但是Jing等[22]发现感知费用与知识付费行为存在高度相关关系（其中r=0.647），但是赵宇翔等[23]在研究被调查者从免费到付费知识问答平台的用户转移时却发现成本和知识付费行为存在高度的负相关关系（其中r=-0.584），罗钦芳[8]等认为产品价格对知识付费产品的销量影响是非线性且复杂的。于是卢恒等[24]对国内外的在线用户为知识付费的意愿相关的40篇实证研究文献进行了元分析，将这些现有相关研究的文献的实证结果进行了整理，得到的结论表明对知识付费意愿具有高度相关性的因素有用户需求、体验、感知专业性、可靠性、感知知识质量、感知有趣性、感知收益和付费态度等，也发现用户具有知识付费经历与否是不同研究结果异质性的部分来源。

## 2.3 知识付费平台的免费增值策略研究

随着免费增值商业模式的推广，知识付费平台开始尝试这一模式的应用。但是Free-Premium模式在知识付费中的应用依旧不够完善，行业还处于探索阶段。

经过文献调查，笔者发现知识付费平台的免费增值策略研究几乎没有。所以本研究将文献搜索范围扩大至免费增值模式的研究，发现目前的研究针对不同行业中免费增值模式都有探索，涉及面广但是都尚未深入。

本研究对免费增值的概念进行了全面的资料整理，以清晰地界定研究对象。免费增值模式即Freemium或者Free-Premium模式，该模式可以追溯到20世纪80年代，但是这个专有名词是在2006年创造的。免费增值的定义中，比较容易混淆的两个概念是免费试用与免费试听。

免费试用指的是在限定的时间段内免费为潜在用户提供部分或完整的服务或者产品，之于服务，通常采用时间限定，之于具体的有形的产品通常直接赠送产品以供试用。免费试用适用于产品差异化的市场，但是由于差异化带来的价格差异也会很大，对产品本身的品质要求较高；免费增值指的是时间或者空间上的先免费后收费，包括免费试听。严格来说，免费增值与免费试用的定义是有明显的区别与界线的，根据定义，免费试用是一种客户获取模型，可在有限的时间内免费向潜在客户提供部分或完整的产品。而免费增值是一种客户获取模式，可以免费向潜在客户提供对部分软件产品的访问，没有时间限制。

综上，免费增值似乎与免费试用相同，但没有时间限制。不过在具体研究中，学者们对于这两个概念并没有十分严格地进行区分，有交叉重叠的部分。例如研究[28]认为免费试用属于免费增值，选择的是软件版本的免费—高级策略 (Free-Trail) 与高级—免费策略对付费转化率的影响，其中的Free-Trail没有时间限制，按照定义该模式属于免费增值模式，但是该研究认为属于免费试听并且将免费试听归为免费增值，得到的结论是高级—免费策略相比较于免费试用策略对付费转化率有更多的促进影响，即用户一直自由使用功能有限的服务，反而会限制付费行为。夏万虎[25]对免费增值模式的B2B用户付费的影响因素进行实证研究中，也将免费试用归属于免费增值模式，以教育机构行业为研究对象，构建数学模型，研究了免费试用中增值功能的时长、是否试用增值的功能、试用强度与是否购买、付费金额的相关关系。但是其他研究只针对免费试听研究，并未提及免费增值以及两者关系。免费试听被学者定义为免费增值而非免费试用，因为免费试听是指提供部分内容的体验，尤其对于有声音频，限制的是内容的时长而非使用的时长。在本研究中，我们研究的是免费增值模式，不涉及免费试用；也不涉及免费试听，有声音频两种免费增值模式，都属于产品的部分免费增值，但本研究不对免费试听进行涉及，经过筛选去除了限时免费的专辑，排除了免费试用与免费试听的影响，研究结果更准确。

基于本节上述的文献研究，本研究确定了免费增值模式的选择，笔者进一步对目前免费增值研究的国内外现状进行了调查。曾润滋[26]对免费增值的相关文献进行了综述，对免费增值的定义、发展进行了总结归纳。该研究指出早期学者们对Free-Premium的研究基本上基于理论与案例，聚焦Free-Premium的形成，应用，分类。之后才开始关注免费增值模式下的付费转化率，以及结合价格等因素为不同行业的免费增值模式进行最优策略的制定。其中，郭敏[27]基于财务视角分析了免费增值商业模式的经济效果，以迅雷作为案例分析，为其他使用或预期使用免费增值商业模式的企业提供了借鉴。在以上对免费增值模式单一维度下的经济效益研究基础上，有学者对免费试用与其他经济策略的选择进行了研究，Li等[28]就选择免费增值和广告两种营销方式，探究了企业在不同的因素影响下应该如果选择或者组合这两种营销策略。

现有研究对于免费增值模式下的付费转化率或者用户留存率涉及很少。根据免费增值的分类不同，现有研究中Koch等[29]针对免费版本—付费版本的Free-Premium模式进行研究，衡量免费服务的局限性是否会影响免费和高级版本的评估。

本研究将免费增值与免费试用严格区分，虽然本研究不涉及免费试用，但是由于目前免费增值的研究不够全面，笔者进而对免费试用的研究也进行了整理，以供参考。陈建[30]对免费试用的经济溢出效应进行了分析，证实了免费试用经济溢出效应是客观存在的。目前针对免费试用策略的研究主要关注免费试用对服务本身的影响。Bram等[31]探究了免费试用对高科技服务的影响，研究结果表明消费者具有前瞻性，结果强调，免费试用是一把双刃剑，试用期间的时间和消费者的使用强度是这些促销活动有效性的关键，还讨论了对管理人员的影响。Hua等[32]将用户的学习行为纳入免费试用的效果中，发现通过免费试用，用户的学习效果使用户熟悉服务，并有助于提高用户的实用性。

## 2.4 文献评述

综上，知识付费平台包括移动有声音频的定义与发展，知识付费平台的用户行为，免费增值模式下的付费转化率三个维度都有研究涉及。

但是，目前针对移动有声音频平台的研究还主要局限于营销方式的理论研究，纵使少量研究对用户付费行为的预测进行了探索，但是都局限于用户层面，而未探讨作品以及收费策略本身。在进一步的文献搜索中，笔者发现本研究探讨的领域，即知识付费的免费增值模式下的付费转化，目前研究基本属于空白。

为了进一步为本研究提供理论基础，在文献调查中，笔者将范围扩大至免费增值模式或者付费转化的研究。在此部分中，笔者通过文献调查发现目前学者还没有对知识付费中的用户付费转化给予一些关注，但是其他行业已然在此领域进行了一些探索，如游戏行业，短视频，流媒体等行业中付费转化率以及用户留存率的提高，减少用户流失都有涉及，其中的免费增值模式都是软件的免费版本—付费版本，其中用户留存与付费转化的计算方式都结合具体的研究对象与平台进行定义与计算，为本研究的付费转化界定提供了参考。其中对于用户留存提升的研究基本上是以用户为维度研究，方法论基本上是从用户层面进行特征提取从而进行流失预测，不同于本文关注点。

综合来看，目前移动音频创作平台的免费增值模式下付费转化率影响因素尚未有研究涉及，但是随着在线音频的付费，用户对喜马拉雅FM专辑的高定价以及混乱设置收费节点早有诟病；同时，移动音频创作平台的用户庞大，作品的可替代选择高，对用户付费行为影响的因素越来越倾向于付费模式的选择，价格，以及收费节点的设置。所以本研究通过文献调查决定坚持从作品层面出发，研究专辑的用户留存是如何受免费增值模式收费策略的影响，尝试给在线音频平台带来更好更稳定的发展。

# **3 研究场景与数据描述**

## 3.1 喜马拉雅FM平台的免费增值策略

喜马拉雅FM创立于2012年。2012年，中国的互联网正处于繁荣发展时期，互联网的一切都是免费的，这一点不管是之于用户的认知，平台所持的理念还是当时电子支付技术的欠缺等方面来看，都是互联网使用者与提供者的共识。之后，渐渐开始发展形成了内容付费的模式，大部分的内容平台的盈利模式都是通过用免费的内容或者服务来引流，奠定用户基础，当活跃用户达到一定程度的规模以后，各大平台会采取流量广告的形式来获利，之后增加辅以部分的增值服务。这一尝试虽然让大量潜在内容制作者开始原因参与到内容制作中，并且透明化，但是最开始的内容制作者主要集中在各类出版社、文化有限公司或者大型的在线媒体，低价并不足以吸引他们进行优质内容的生产。所以喜马拉雅的付费在用户固有的免费思维以及没有优质的内容的双轮效应作用下，生态环境停滞不前甚至不断恶化。随着版权意识的增强以及国家对盗版的整治，喜马拉雅逐渐开始与出版商、优质主播、自媒体进行签约，形成版权认证。喜马拉雅经过上述的发展以后，至今形成了一套完整的付费体系。

除了大量普通用户的专辑属于免费（Free：即不需要任何费用即可享受专辑的内容）之外，付费专辑基本上属于PUGC。喜马拉雅的付费专辑拥有两种收费模式，两种收费模式根据是否支持VIP又可以细分为五种模式。

### 3.1.1 喜马拉雅FM平台的免费增值收费模式

1）单独购买专辑的单个或某些声音

PayByTrack模式。根据是否支持VIP服务又可分为以下两种收费模式（其中喜马拉雅平台本身提供的会员服务 (VIP) 指的是会员在会员期间可以享受专辑）：

（1）只能单独购买专辑的单个声音

ByTrackOnly模式。即可以通过购买专辑的某个或者某些声音永久的享受该声音，也叫单集购买，所有的声音组成一张专辑。

（2）单独购买声音或者开通VIP

ByTrackOnly/Vip模式。该模式指用户可以选择开通VIP并且不购买声音，在会员期间享受非VIP用户需要购买才能享受的声音，或者不开通VIP选择单独购买声音永久享受声音。

2）购买整张专辑

PayByAlbum模式。根据是否支持VIP服务又可以分为以下三种收费模式：

（1）只能开VIP才能享受专辑

Only Free For Vip模式。即只能通过开通VIP享受专辑，并且只能在VIP期间享受专辑，没有其他获取专辑的付费模式。

（2）单独购买专辑或者开通VIP

ByAlbumOnly/Vip模式。该模式指用户可以选择开通VIP并且不购买专辑，在会员期间享受非VIP用户需要购买才能享受的专辑，或者不开通VIP选择单独购买声音永久享受声音。

（3）只能单独购买专辑

ByAlbumOnly模式。即只能通过购买整张专辑永久的享受该专辑。

### 3.1.2 喜马拉雅的免费增值服务

在上述的付费的模式下，喜马拉雅采用了提供免费增值服务的方式作为营销策略。喜马拉雅FM提供两种免费增值服务，专辑中部分章节的免费以及付费，单个付费章节的小段免费试听。

（1）专辑中部分章节的免费

喜马拉雅的付费专辑中有近一半的专辑属于该模式，即付费专辑并非所有章节都收费，而是一般采取某些声音免费，其余声音付费的收费模式。针对声音情节连续的部分专辑，一般采取前几十个声音免费，后续所有声音付费的模式。针对声音之间的内容不相互影响的专辑，内容制作者一般会选择一些可能受众面广的声音作为免费章节，吸引用户，这样的免费增值模式既可以让用户自行选择优质内容并且试听体验，也可以提高内容制作者的用户基础。

（2）单个付费章节的小段免费试听

喜马拉雅中的付费专辑中的收费声音，也分为两种类型，一种是完全无法试听的付费声音，另一种则是可以免费试听一小段的付费声音，通常这个试听时长为90s或者180s。

喜马拉雅FM的以上两种免费增值服务是不互斥的，可以同时存在。

### 3.1.3 付费转化率的界定与量化

在金融经济学中，付费转化率的计算为付费用户数/活跃用户数，微观的角度看，付费转化率代表了单个用户对单个收费产品的购买几率，其中购买行为可以分为理性购买与冲动型购买两种。宏观的角度看，付费转化率代表了用户的付费意愿、消费能力和消费观念。付费转化率也可以衡量产品的付费用户转化能力，产品的付费节点、收费价格是否被接受。

流失定义为用户停止使用来自服务提供商的各种服务比如游戏，音乐等软件服务的使用，而留存则可以用总活跃用户数与流失数之差来计算。本研究用专辑付费节点前后的用户留存率来量化付费转化率，即付费节点前后用户的收听的比例。一般的研究从用户的维度去衡量用户留存率，或者付费转换率，通过建立特征工程对单个用户的留存或者流失几率进行预测，从而去反映用户维度的付费转化率。这类基于用户层面的流失或者留存进行付费转化预测通常通过机器学习进行特征工程预测，难点在于找到一个合适的ML技术，以及确定一个合适的类标签方法。通常基于机器学习得预测能够很好地利用现有用户的数据也能够比较准确并且完整的，但是这是黑盒的，用户的行为永远是无法量化，不明晰的。而对于有声音频创作平台，用户量十分庞大，并且不断增长变化，同时由于有声音频的选择多，种类丰富，以及最近兴起的知识网红经济效益，用户的付费转化率更多的取决于有声音频的特征以及主播的特征，从产品本身出发分析用户付费转化率的影响因素更能长久地改善收益；所以本研究的问题聚焦专辑层面，研究主播发布的专辑对用户付费的转换能力，以及其影响因素。通过专辑付费节点前后用户的播放量即用户留存来反映用户的购买行为，更能够反映该研究的问题。

## 3.2 数据获取与清洗

### 3.2.1数据源介绍

数据源：喜马拉雅FM平台中所有的数据信息，该数据为截面数据，取自于某一个时间点，具体为2021年9月22日；

方法：通过抓包以及爬虫获取喜马拉雅FM三个终端（电脑端网页版、手机端网页版、APP）数据存储到MongoDB中。各个终端的URL如表3-1所示。具体的爬取思路为用Python对所有专辑的分类的信息进行爬取，然后对分类里所有的专辑信息进行爬取，然后对专辑里的声音信息进行爬取，最后是主播信息的爬取。所有信息的爬取都基于对抓包的URL中的信息进行了查看以后，选择了可能需要的变量进行了存储，存储媒介为MongoDB数据库，具体的Python爬取代码见附录一。

表3-1 数据源——喜马拉雅FM三个终端的URL

|  |  |  |
| --- | --- | --- |
| 研究平台 | 数据源 | URL |
| 喜马拉雅FM | 电脑端网页版 | https://www.ximalaya.com |
| 手机端网页版 | https://m.ximalaya.com |
| APP | https://mobile.ximalaya.com |

### 3.2.2 数据的三个维度

总的来说，喜马拉雅FM的数据可以分为三个维度：用户 (User) ，由用户发布的专辑 (Album) ，专辑的所有章节 (Track) 。

（1）用户

用户可以分为主播与普通用户。主播即喜马拉雅平台的内容制作者或者产生者，主要在喜马拉雅平台发布专辑以供大家订阅；普通用户即一般的消费者，大多登陆平台，寻找自己感兴趣的专辑听，也有发布专辑的权限，但是一般普通用户不会也无法通过喜马拉雅盈利，我们得到的数据中包含用户提供给喜马拉雅所有的基本信息，包括性别，专业认证，粉丝数等属性。

（2）专辑

专辑是联系喜马拉雅整个生态的纽带，即有声音频。拥有许多不同的分类标准，按照专辑制作者的专业度等特点，可以分为PGC专业内容、UGC及PUGC内容；PGC专业内容即Professionally Generated Content，对应喜马拉雅的所有专业的主播发布的专辑；UGC即User Generated Content，指的是用户生成内容，即用户原创内容；PUGC，Professional Generated Content + User Generated Content，即“专业用户生产内容”或“专家生产内容”）指在移动音视频行业中，将UGC+PGC相结合的内容生产模式。由于知识付费的到来，有声书也开始由免费变为了付费内容，喜马拉雅对专辑设置了不同的收费模式。喜马拉雅的专辑涵盖面广，有泛知识领域的商业财经、历史文化类的专辑，小说类专辑；适合少儿与个人发展的教育内容，有适合更年龄阶段的内容；同时，内容上既有音频播客的形式，随着视频的兴起，也逐渐形成了音频直播+有声音频的形式。专辑是整个研究最核心的对象，获取到的信息有专辑的播放量，评论数，订阅数，声音数等一系列指标。

（3）声音

每一张专辑由声音（也称章节）组成，每一个声音以一段音频的形式存在，声音的时长有长有短，不同主播录制的声音拥有不同的声色与特点。不等数量的声音组成专辑，声音的内容都应该符合专辑的主题。（直播的音频也可以被存为专辑的声音，一般时长以小时计数，用于用户回听）。

### 3.2.3 数据集清洗

由于原始数据庞大并且复杂，我们需要进行数据清洗，过滤掉不需要的数据，得到符合条件的数据集作为分析对象。

1）以专辑为清洗对象

三个维度中选择以专辑为清洗对象的原因：首先，我们研究的核心是专辑层面的付费用户转化率；其次，声音数据量庞大不方便操作，用户维度无法准确的代表数据集。综合以上两点，本研究决定以专辑维度为数据清洗的对象。接着筛选出这些符合条件的专辑对应的用户与声音维度的数据集共同作为分析对象。

2）筛选规则

（1）筛选已经完结的专辑。去除未完结的专辑，由于我们需要分析整个专辑的所有章节随着剧集呈现的播放趋势，所以不应该将未完结的专辑纳入研究对象，故首要规则为已完结。

（2）筛选PUGC以及PGC专辑。过滤掉普通用户上传的专辑。由于喜马拉雅用户的特殊性，我们拥有的专辑可能是普通用户即消费者的私人专辑，基本上没有可分析的意义，所以我们只分析属于PUGC以及PGC的专辑。

（3）除去直播宣传的专辑。在数据集中发现有少量专辑中的章节时长长达2小时，经过查看发现是直播回听的音频，研究主要因变量是作品播放量，而直播形式的本质是变相营销宣传，所以本研究过滤掉有直播宣传的专辑，排除直播宣传的干扰。

（4）筛选章节中有编号且属于免费增值模式的专辑。由于喜马拉雅有声书的专辑占70%以上，而在这些有声书专辑中专辑的播放量本身会随着章节的变化，需要确保章节是按照剧情排序的，所以需要对章节中的title进行编号提取，并且排序，筛选出都有编号并且无重复错误编号的专辑，具体提取章节编号并且进行筛选的过程通过Python进行，详细代码在附录二。在编号符合条件的基础上，还需要对专辑的特征进行识别，选择出属于免费增值模式且只有一个收费节点的专辑作为研究对象。

3）清洗的方法

首先查看是否有直接衡量或者代表以上两个规则的一个分类属性或多个分类属性的组合指标，如果有直接对该指标进行筛选处理即可；如果没有，用多个连续型指标量化规则，对所有的数据集的这些连续型指标进行基本的描述性统计分析，包括均值，方差，四分位数，频率分布直方图等，从而过滤掉离群值。

### 3.2.4 数据集概览

原始数据集经过数据清洗以后的数据规模（如图3-1所示）。数据集中4610张专辑特征如图3-2所示，所有专辑的Playtimes与Price的折线图都符合该特点，均为前部分声音免费，从某一章节开始付费，且只有一个收费节点，无多个收费节点，同时这些专辑中的所有声音都没有免费试听这一免费增值服务。

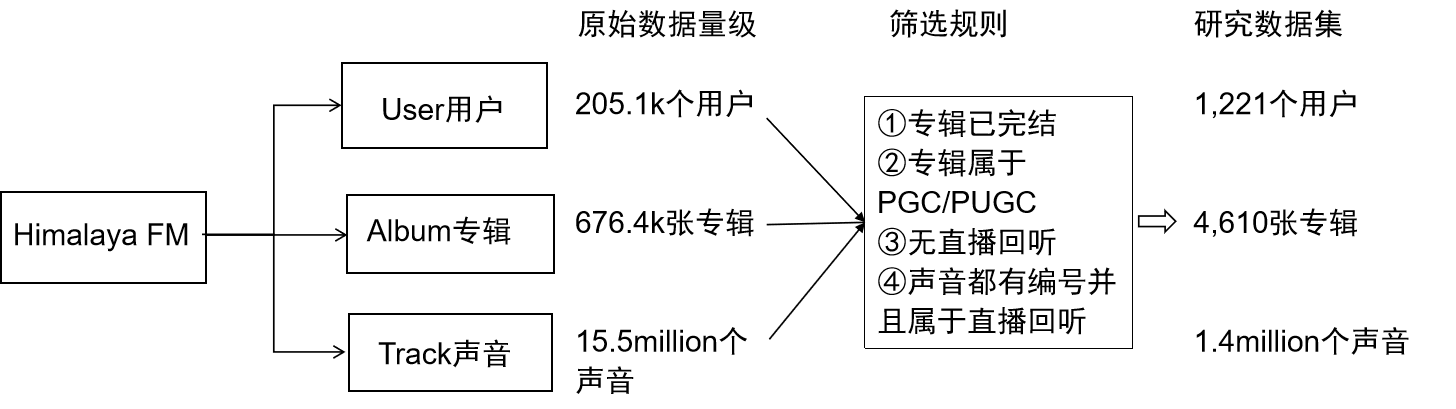


图3-1 数据集概览

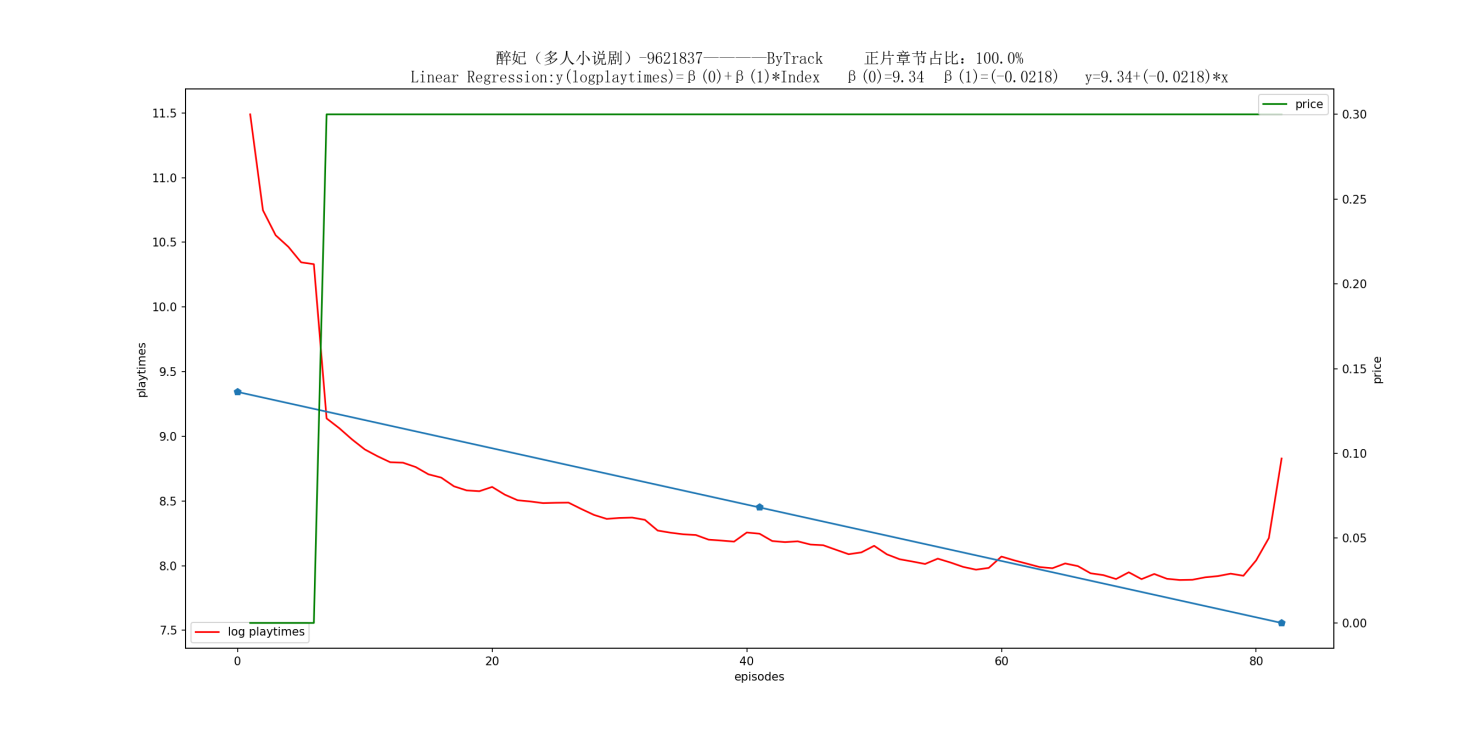


图3-2 专辑特征示例

## 3.3 变量描述

本研究从专辑、主播、声音三个层面对数据集中可能会对用户留存率以及专辑播放量有影响的变量进行了选择，其中选择并进一步处理得到了28个变量。这28个变量的描述（如表3-2所示）。

表3-2 变量描述

|  |  |  |  |
| --- | --- | --- | --- |
| 变量名 | 变量类型 | 变量含义 | 变量取值 |
| R | float | 收费节点的用户留存率；首个收费的章节播放量与最后一个免费章节播放量的比值，%。 | |
| AvgR | float | 平均留存率（自然留存率）%， (n-1) √ (last/first)，首尾声音的留存率开n-1次方，其中n为声音数； | |
| tP | float | 专辑单个声音的价格； | |
| PP | float | 收费节点，relative position，收费点在所有声音的位置，%； | |
| PP\_1 | float | 收费节点，relative position，收费点开始的时长占比，%，所有FreeTracks的SumDuration/总Duration； | |
| PP\_2 | int | 收费节点，absolute position，收费声音的编号； | |
| Vip | bool | 是否开通VIP就能享受专辑，0/1； | |

（续）表3-2 变量描述

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 变量名 | 变量类型 | 变量含义 | 变量取值 | | |
| Vip\_1 | int | 获取专辑的购买类型 | | 0：免费 (Free)，  1：可单集购买 (BuyTrack)，  2：整张专辑购买 (BuyAlbum)； | |
|  |  |  | |
| Vip\_2 | int | 付费类型 | | 1：只能购买声音 (TrackOnly)，  2：开通VIP或者购买声音 (VipOrTrack)，  3：只能通过开通VIP享受专辑 (VipOnly)，  4：开通VIP或者购买声音 (VipOrAlbum)，  5：只能购买整张专辑 (AlbumOnly)； | |
|  |  |  | |
|  |  |  | |
| avgD | float | 声音的平均时长（单位为Min）； | | | |
| avgPT | float | 平均播放量 (Playtimes/TrackCount)； | | |  |
| cId | int | 专辑所属的分类Id； | | | |
| cC | int | 专辑的评论总数目； | | | |
| T | float | 从创建到完结所经历的时间，单位为天； | | | |
| tC | int | 专辑里的声音数； | | | |
| uA | int | 主播发布的专辑数目； | | | |
| uAG | int | 主播的等级 | 其中-1表示该用户没有实名认证，严格来说不算是主播，1-16代表主播的等级； | | |
| uG | int | 主播性别 | -1表示未提供，  0表示保密，  1表示男性，  2表示女性； | | |
|  |  |  |
|  |  |  |
|  |  |  |
| uT | int | 主播创作的声音数； | | | |
| uF | int | 主播的粉丝数； | | | |
| uP | int | 主播是否是喜马拉雅认证的专业主播，取值为0/1； | | | |
| UpF | float | 该专辑更新一集需要的时间，单位为天/集； | | | |
| Age | float | 该专辑从创建到我们收集的时间，单位为天； | | | |

## 3.4 变量的描述性统计

总共4610张专辑，将分类变量与连续变量分别呈现，分类变量用占比表示，连续变量由均值、方差、四分位数、观测值呈现指标的分布；具体分为专辑层面，包括基本特征，以及付费特征，呈现该专辑本身的一些信息；主播基本特征，呈现这些专辑所属的主播的一些基本信息；最后是描述性统计信息的归纳总结。

### 3.4.1 专辑层面

由avgDurationInMin（avgD，专辑每集平均时长）、avgPlayTimes（avgPT，每集平均播放量）、trackCount（tC，章节数目）、subscribeCount（sC，专辑的订阅数）、commentsCount（cCount，评论数）、categoryId（cId，专辑所属分类id）、TimeOfMaking（T，专辑从创建到完结所需的时间）、UpdateFrequency（UpF，平均每更新一集所需时间，单位为天）、Age（Age，专辑从创建到本研究收集节点的时间）总共9个指标来描述专辑的基本特征，用Retention (R) 衡量付费转化率，即收费节点前后的用户留存率；用PayPoint (PP) 衡量收费节点，即收费集在所有声音中的相对位置，百分比（如表3-3所示）。收费模式选择了三种不同的粒度，分别是Vip：是否支持会员服务；Vip\_1：购买类型；Vip\_2：五种收费模式（如表3-4所示），专辑所属分类的分布（如表3-5所示），总共以上多个维度的指标来描述专辑的特征。

表3-3 album的基本特征

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 变量 | Mean | Std | Min | Q1 | Median | Q3 | Max | Count |
| avgD | 16.6 | 4.54 | 0.95 | 12.86 | 18.62 | 20.28 | 32.55 | 4610 |
| avgPT | 8101 | 51204.54 | 11.65 | 360.59 | 1266.19 | 4258.22 | 2147028 | 4610 |
| cC | 80.67 | 512.13 | 0 | 3 | 11 | 37 | 26125 | 4610 |
| sC | 11077.49 | 101417.92 | 11 | 652 | 1948.5 | 5988.75 | 5227481 | 4610 |
| tC | 306.69 | 253.74 | 51 | 134 | 244 | 397 | 3120 | 4610 |
| T | 273.44 | 217.28 | 0 | 103.62 | 222.74 | 385.5 | 1427.66 | 4610 |
| UpF | 1.07 | 0.93 | 0 | 0.47 | 0.91 | 1.33 | 13.12 | 4610 |
| Age | 798.27 | 345.27 | 103.7 | 517.56 | 811.75 | 1058.42 | 2266.32 | 4610 |
| PP | 0.18 | 0.1 | 0.01 | 0.11 | 0.18 | 0.21 | 1 | 4610 |
| R | 0.55 | 0.22 | 0.01 | 0.39 | 0.55 | 0.71 | 1.67 | 4610 |
| tP | 0.22 | 0.11 | 0.03 | 0.2 | 0.2 | 0.22 | 4.04 | 4610 |

表3-4 三种收费模式的分布

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 变量 | 取值与占比 | | | | | |
| Vip | 0：不支持VIP | | 1：VipFree |  |  |  |
|  | 271 | 4339 | |  |  |  |
|  | 5.88% | 94.12% | |  |  |  |
| Vip\_1 | 1：ByTrack | | | 2：ByAlbum | | |
|  | 3877 | | | 733 | | |
|  | 84.10% | | | 15.90% | | |
| Vip\_2 | 1:OnlyTrack | 2:VipFree/ByTrack | | 3:VipOnly | 4:VipFree/ByAlbum | 5:OnlyAlbum |
|  | 236 | 3641 | | 0 | 698 | 35 |
|  | 5.12% | 78.98% | | 0.00% | 15.14% | 0.76% |

表3-5 专辑所属的分类频率分布表

|  |  |  |  |
| --- | --- | --- | --- |
| categoryId | categoryName | f | frequency |
| 2 | 音乐 | 2 | 0.05% |
| 3 | 有声书 | 4154 | 99.86% |
| 4 | 娱乐 | 7 | 0.17% |
| 5 | 外语 | 3 | 0.07% |
| 6 | 儿童 | 35 | 0.84% |
| 7 | 健康养生 | 3 | 0.07% |
| 8 | 商业财经 | 30 | 0.72% |
| 9 | 历史 | 135 | 3.25% |
| 10 | 情感生活 | 6 | 0.14% |
| 12 | 相声评书 | 3 | 0.07% |
| 13 | 个人成长 | 86 | 2.07% |
| 15 | 广播剧 | 26 | 0.63% |
| 23 | 影视 | 5 | 0.12% |
| 24 | 二次元 | 44 | 1.06% |
| 39 | 人文 | 51 | 1.23% |
| 92 | 少儿素养 | 20 | 0.48% |

根据描述性统计结果得到：专辑的基本特征中，章节平均时长大部分在20min之内，并且均值为16.6；章节平均播放量在5000以内，存在少量播放量极少以及播放量极高的专辑，均值为8100；专辑评论数拥有较大的方差，大多数在37之内，但是极大值达到了2万评论数；专辑的订阅数也拥有较大的方差，平均数目在1.1万用户，75%以上的专辑拥有不超过6000的订阅数，存在部分专辑拥有较多的播放量，极大值大于500万；章节数目也表现为高峰度，并且大多数专辑的章节数在400以内，而少数专辑章节数在1000以上，章节数均值为273章；专辑从创建到完结所需的时间，大部分在1年以内，极大值大于3年。大部分专辑的平均每天更新的集数大概在1左右，有少量专辑平均每天更新集数大于10；专辑已经存在的年限中均值为2年左右，大部分专辑在3年以内，少量专辑高达5-6年。

被解释变量即付费转化率基本呈正态分布，大部分专辑付费转化率小于1；解释变量中大部分专辑收费节点在20%以内，并且均值也在18%左右；大部分专辑平均每个声音价格在0.22以内，并且均值也为0.22。

分类变量中，专辑的收费模式中94%以上的专辑都支持VIP服务，只有极少部分专辑必须得单独购买专辑或者章节才能享受专辑；在购买类型中，84%以上的专辑提供按照单个声音的购买模式，少数指提供整张专辑购买的方式。专辑的分类中，有声书分类占比高达99%，这些专辑基本上都是有声书音频。

### 3.4.2 主播层面

由userFollowers（uF，主播的粉丝数）、userAlbums（uA，主播已发布的专辑数）、userTracks（uT，主播已发布的声音数）、userAnchorGrade（uAG，主播的等级）、以及分类变量userGender（uG，主播的性别）来呈现主播的基本特征，具体为专业度，受欢迎度（如表3-6以及表3-7所示）。

根据描述性统计结果，可以得到：主播的基本特征中，主播粉丝数大部分在百万级别以内，但是均值达到了600万，存在部分主播粉丝数在千万级别；主播已发布的专辑数大部分在90左右且均值也在90左右，只有少部分主播发布了上百张专辑；主播的发布的声音数中，大部分取值在2.5万以内，均值也在2.5万左右，存在少量主播发布的声音数达到了几乎25万。并且主播中提供了性别的主播男女比接近2:1；主播中被喜马拉雅官方授予荣誉认证的主播占大多数，比例在85%。

表3-6 user的基本特征

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 变量 | | Mean | | Std | | Min | | Q1 | | Median | Q3 | | Max | Count | |
| uA | 89.67 | | 127.08 | | 1 | | 18 | | 42 | | 93 | 596 | | | 4610 |
| uF | 603165.14 | | 1057262.87 | | 57 | | 68425.25 | | 185665.5 | | 850824 | 19131367 | | | 4610 |
| uT | 25010.63 | | 42102.56 | | 56 | | 3579 | | 10093 | | 25087 | 244969 | | | 4610 |
| uAG | 12.42 | | 1.84 | | -1 | | 12 | | 12 | | 14 | 16 | | | 4610 |

表3-7 user的专业度以及性别占比

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 变量 | 取值和占比 | | | |
| uP | 0：未授予喜马拉雅荣誉认证 | | 1：喜马拉雅官方荣誉主播认证 | |
|  | 681 |  | 3929 |  |
|  | 14.77% |  | 85.23% |  |
| uG | -1：不明 | 0：保密 | 1：男 | 2：女 |
|  | 109 | 1521 | 1803 | 1177 |
|  | 2.36% | 32.99% | 39.11% | 25.53% |

### 3.4.3 小结

对描述性统计结果进行总结，可以发现：专辑的特征中大部分受欢迎程度都在中等偏下，表现为右偏，只有少数专辑拥有极高的订阅数、评论数；同时大部分专辑的时长都在20min左右，没有太大的差别；但是专辑的声音数也表现为右偏，只有少数专辑的声音数较多，以千为单位计数；绝大多数专辑的付费转化率小于1，但是有少数付费转化率大于1，需要对此现象进一步研究；收费节点的设置也绝大多数在20%以内，经过调查发现喜马拉雅的付费专辑收费节点目前是喜马拉雅FM平台规定的，所以20%应该是平台规定；章节的价格也在0.2左右，是平台的特点；平台大部分专辑支持VIP免费，以及以章节收费；主播特征中，大部分主播属于中小型受欢迎的主播，只有少数大咖拥有极高的粉丝数以及作品数。

# **4 研究设计**

## 4.1 研究假设

杨东红[18]等人已经从付费移动音频的维度，对用户为专辑的付费行为探究了，得到了专辑一些基本特征以及主播基本特征对用户付费行为的影响，本研究便不在重复研究。同时，为了排除交互行为对付费转化率的影响，我们将作品流行度，专辑已发布的年限一并作为控制变量。将以上变量作为控制变量的前提下，本研究探索的问题是在给定一个特定的专辑，在章节数，类别，平均时长确定的基础上，结合专辑发布者即主播的特征比如主播等级，主播专业度，主播已发布的作品数，收费节点，收费价格，收费模式会对付费转化率产生这样的影响，并寻找一个最优策略，对于该影响做出以下假设（假设概览如表4-1以及图4-1所示）。

表4-1 研究假设概览

|  |  |
| --- | --- |
| 解释变量 | 假设 |
| 收费节点 (PayPoint) | H1：收费节点对专辑收费节点处的用户留存率即付费转化率具有显著的正向影响 |
| 章节价格 (TrackPrice) | H2：章节价格对专辑收费节点处的用户留存率即付费转化率具有显著的负向影响 |
| 是否支持开通VIP免费 (Vip) | H3：支持VIP免费对专辑收费节点处的用户留存率即付费转化率具有显著的正向影响。 |

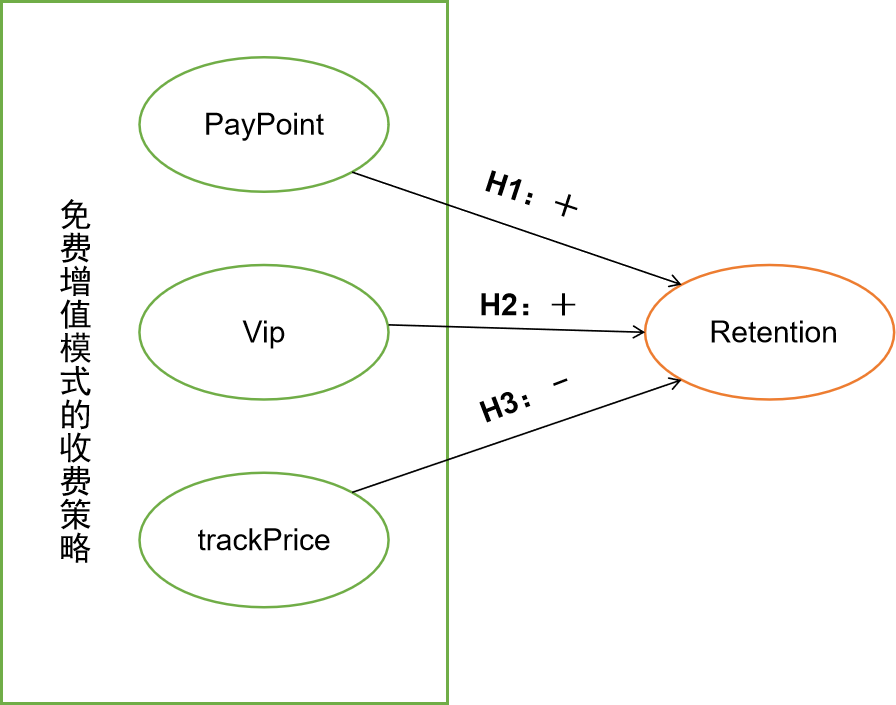


图4-1 研究假设

### 4.1.1 收费节点

收费节点，宏观层面，指的是收费的时间点，对象，范围等。本研究针对喜马拉雅平台专辑的特点，以专辑为维度，将收费节点定义为，一张专辑里声音刚开始部分免费到某一个声音开始付费，节点为开始付费的集数，或者相对位置，相对时长。用户是否付费收到收费节点的影响，尤其是连续情节的专辑，随着情节的走向，声音对用户的吸引力不一。并且由于收费节点越靠后就说明该专辑的收费比例越少，用户对价格的敏感性一般比较高，更倾向于愿意付费，所以针对收费节点做出以下假设：

H1：收费节点对专辑收费节点处的用户留存率即付费转化率具有显著的正向影响。

### 4.1.2 章节价格

付费转化率最直接的影响因素便是价格，价格反映了作品的价值，同时用户是否愿意付费表现了该价格是否在用户对该专辑价值的预期之内。虽然章节价格的范围取值很小，但是专辑所含的章节数巨大，所以价格上的区别会对用户付费行为产生明显影响。同时，用户一般更容易接受低价的产品，所以更愿意做出付费行为，所以针对章节价格做出以下假设：

H2：章节价格对专辑收费节点处的用户留存率即付费转化率具有显著的负向影响。

### 4.1.3 是否支持VIP免费的模式

是否支持VIP免费是指开通喜马拉雅的会员服务是否可以享受该专辑。喜马拉雅的收费模式分为两种，细分又可以分为五种，其中是否支持VIP免费的模式代表了该用户的付费行为是有限的还是既能选择VIP短暂享受专辑也能选择购买专辑永久享受。购买专辑或者章节在有声书的获得中是比较少有的选择方式，因为用户对专辑的享受一般是一次性的，购买专辑虽然可以永久享受专辑，但是性价比并不高，所以支持VIP免费服务更能促进活跃用户向付费用户转化。所以针对是否支持VIP免费做出以下假设：

H3：支持VIP免费对专辑收费节点处的用户留存率即付费转化率具有显著的正向影响。

## 4.2 模型构建

在上述28个变量中本研究选择的模型如下表。依据研究问题，即，在一个主播创建专辑时，根据给定的主播特征例如专业度，受欢迎度，粉丝基础，专辑特征例如专辑每章平均时长，专辑章节数，专辑的类别，更新频率去探索设置不同的收费节点、章节价格，收费模式会对收费节点前后的用户留存率即付费转化率有怎样的影响，由于我们的数据属于截面数据，并且是专辑完结以后收集的信息，考虑到专辑本身的原始积累的流行度，例如播放量，评论数也会对用户留存率造成影响，所以我们需要控制主播特征，专辑特征，以及专辑流行度对因变量用户留存率的影响，将这12个变量作为控制变量；解释变量选择收费节点，章节价格，收费模式，探究这三个变量对付费转换率的影响。从而为平台以及主播提供推荐的收费方式，使得用户接受的同时提高收益。模型的变量选择与赋值如表4-2所示，根据该选择所构建的模型为：

 (4-1)

表4-2 模型变量选择与赋值

|  |  |  |  |
| --- | --- | --- | --- |
|  | 变量 | 含义 | 赋值情况 |
| 解释变量 | R | 付费转化率； | =Pt (FirstPayTrack) /Pt (LastFreeTrack) |
| 被解释变量 | Vip | 是否支持开通VIP服务在VIP期间享受专辑 | 0：否  1：是 |
|  | PP | 收费节点，%； | 例：一张拥有一百集的专辑从第十集开始收费，则PP=10% |
|  | tP | 专辑每个章节的平均价格； | 整张专辑收费：tP=Price/trackCount |
| 控制变量 | avgPT | 平均播放量； | Playtimes/TrackCount |
|  | avgD | 声音的平均时长（单位为Min）； | |
|  | cId | 专辑所属的分类Id； | |
|  | cC | 专辑的评论总数目； | |
|  | Age | 从创建到完结所经历的时间，单位为天； | |
|  | tC | 专辑里的声音数； | |
|  | UpF | 平均每更新一集所需时间（天）； | |

（续）表4-2 模型变量选择与赋值

|  |  |  |  |
| --- | --- | --- | --- |
|  | 变量 | 含义 | 赋值情况 |
| 控制变量 | uA | 主播发布的专辑数目； | |
|  | uAG | 主播的等级； | -1表示该用户没有实名认证，严格来说不算是主播，1-16代表主播的等级； |
|  | uF | 主播的粉丝数； | |
|  | uG | 主播性别； | -1：不明，0：保密，1：男性，2：女性； |
|  | uP | 主播是否被授予喜马拉雅荣誉认证； | 0：否，1：是 |

## 4.3 相关分析

经过QQ图，概率密度分布图以及正态性检验得到被解释变量符合正态分布，故不再对变量进行处理，直接进行相关分析。对连续型变量用皮尔逊相关分析，分类变量与连续变量之间的相关性用斯皮尔曼分析。若变量的相关性符合要求，则模型变量的选择正确，进一步通过偏相关查看结果，若不符合则需要修正。相关分析的详细结果在附录三，表4-3为控制变量后的偏相关分析结果。

相关分析结果显示，解释变量之间不存在高度相关，虽然存在部分解释变量之间在5%水平上相关显著，但是相关系数都小于0.1，所以这些解释变量相关性很低即几乎不存在相关。解释变量 (R) 与被解释变量之间在5%水平上都存在一定程度的显著相关，解释变量之间以及解释变量与被解释变量之间的相关性符合规范，进一步将控制变量进行控制，得到偏相关分析结果。偏相关分析结果表明被解释变量在1%的水平上与解释变量均显著相关，解释变量之间除了收费节点与章节价格在5%水平上显著相关，但是相关系数<0.1，几乎不相关，其余解释变量之间均不相关。相关性结果符合要求，可以进行接下来的实证逐层分析。

表4-3 解释变量与被解释变量偏相关分析结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Vip | PP | tP | R |
| Vip | 1 |  |  |  |
| PP | -0.007 | 1 |  |  |
| tP | -0.021 | -0.094\*\*\* | 1 |  |
| R | 0.322\*\*\* | -0.092\*\*\* | -0.069\*\*\* | 1 |

注：\*\*\*、\*\*、\*分别表示在1%、5%和10%的显著性水平上显著。

# 5 免费增值策略对用户留存影响的实证分析

## 5.1 逐层回归分析

由于我们的解释变量有三个，并且属于不同的层面，解释变量对被解释变量的影响可能受解释变量之间的影响，所以我们采用逐层回归，起始为控制层，即只将控制变量纳入回归分析；然后再在此基础上，用前进法将各个自变量放入回归中，总共有四层回归，不同层的回归中控制变量以及上一层解释变量的影响都没有改变。所以三个解释变量的纳入是有必要的，下图提供了控制层（如表5-1）和最终完整的回归模型（如表5-2）结果，以供对比。

表5-1 逐层回归—控制层回归结果

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 变量 | 系数 | 标准误 | t | VIF | 调整R² | F | 常数项 |
| avgD | -0.009\*\*\* | 0.001 | -11.06 | 1.73 | 0.1962 | 94.73\*\*\* | 0.614\*\*\* |
| avgPT | 0 | 0 | -0.51 | 2.15 |
| cId | -0.002\*\*\* | 0 | -4.41 | 1.14 |
| cC | 0\*\* | 0 | 2.05 | 2.27 |
| Age | 0\*\*\* | 0 | -12.19 | 1.83 |
| tC | 0\* | 0 | 1.93 | 1.25 |
| UpF | 0.005 | 0.003 | 1.44 | 1.27 |
| uA | 0\*\*\* | 0 | 3.32 | 1.87 |
| uAG | 0.011\*\*\* | 0.002 | 5.4 | 1.65 |
| uF | 0\*\*\* | 0 | 4.16 | 2.01 |
| uG | 0.05\*\*\* | 0.004 | 13.15 | 1.14 |
| uP | -0.005 | 0.009 | -0.56 | 1.19 |

表5-2 逐层回归—最终层回归结果

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 变量 | 系数 | 标准误 | t | VIF | 调整R² | F | 常数项 |
| Vip | 0.271\*\*\* | 0.012 | 23.084 | 1.04 | 0.2897 | 126.31\*\*\* | 0.4295\*\*\* |
| PP | -0.207\*\*\* | 0.03 | -6.923 | 1.22 |
| tP | -0.13\*\*\* | 0.025 | -5.118 | 1.15 |

（续）表5-2 逐层回归—最终层回归结果

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 变量 | 系数 | 标准误 | t | VIF | 调整R² | F | 常数项 |
| avgD | -0.007\*\*\* | 0.001 | -9.004 | 1.76 | 0.2897 | 126.31\*\*\* | 0.4295\*\*\* |
| avgPT | 0 | 0 | 0.145 | 2.16 |
| cId | -0.001\*\*\* | 0 | -2.57 | 1.2 |
| cC | 0\* | 0 | 1.792 | 2.28 |
| Age | 0\*\*\* | 0 | -13.065 | 1.9 |
| tC | 0 | 0 | 0.067 | 1.33 |
| UpF | 0.011\*\*\* | 0.003 | 3.371 | 1.3 |
| uA | 0 | 0 | 1.427 | 2.01 |
| uAG | 0.009\*\*\* | 0.002 | 4.769 | 1.66 |
| uF | 0\*\*\* | 0 | 3.707 | 2.01 |
| uG | 0.041\*\*\* | 0.004 | 11.355 | 1.16 |
| uP | -0.008\*\*\* | 0.008 | -0.916 | 1.19 |

## 5.2 模型检验

### 5.2.1正太残差检验

回归分析的前提是回归模型残差符合正态分布，即残差随机。残差随机表示模型中的随机误差与被解释变量无关，解释变量的选择是准确并且足够的。用SPSS对模型进行残差正态性检验得到（如图5-1和图5-2所示），残差符合正态分布，模型符合残差正态性要求。



图5-1 模型的正太残差检验（直方图）



图5-2 模型的正太残差检验（QQ图）

### 5.2.2稳健性检验

为进一步验证模型，本研究对逐层回归中最终层即所有控制变量，解释变量都纳入回归的模型进行稳健性检验，方法为因变量替换法以及自变量替换法。

下表（如表5-3所示）中序(0)表示原始的回归模型，序(1)通过替换解释变量，进行稳健性检验，本研究已经探究了收费节点的用户留存率与收费节点，章节价格，是否支持会员服务的关系，在稳健性检验时，我们将因变量替换为付费转化率相对于专辑自然留存率的留存率变化。用R\_1=R-AvgR，进行回归，序(1)的结果表明收费节点、收费价格也显著负向影响相对于自然留存率的留存变化，是否支持VIP服务也显著促进除去自然留存率影响的留存率变化。所以在专辑自然留存率的基础上，该模型也能解释付费节点的自然留存率与解释变量之间的关系。

序(2-5)均通过替换自变量进行稳健性检验，其中序(2-3)通过替换其中的付费节点变量进行稳健性检验，PP\_1表示收费开始的时长占比，PP\_2表示收费开始的绝对位置，即所在的集数。结果表明，这两个收费节点的衡量也在0.01的水平上显著负向影响付费转化率，并且该替换对模型的其他解释变量的解释性没有影响。

序(4-5)通过替换其中的收费模式变量( Vip )进行稳健性检验，Vip\_1表示喜马拉雅的两种购买类型，购买整张专辑或者单个章节；Vip\_2表示五种具体收费模式，是将按照Vip\_1按照Vip的不同取值进一步分成的五种模式。结果表明，这两个收费模式的衡量也在0.01的水平上显著正向影响付费转化率，该替换对模型的其他解释变量的解释性没有影响。所有的替换的结果都与序(0)即基本模型的结论一致，证实了本研究模型的结果是稳健的。

表5-3 稳健性检验结果

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | (0)R | (1)R\_1 | (2)R | (3)R | (4)R | (5)R |
| tP | -0.132\*\*\* (0.025) | -0.118\*\*\* (0.025) | -0.132\*\*\* (0.025) | -0.132\*\*\* (0.025) | -0.171\*\*\* (0.028) | -0.208\*\*\* (0.028) |
| Vip | 0.271\*\*\* (0.012) | 0.266\*\*\* (0.012) | 0.271\*\*\* (0.012) | 0.271\*\*\* (0.012) |  |  |
| Vip\_1 |  |  |  |  | 0.037\*\*\* (0.01) |  |
| Vip\_2 |  |  |  |  |  | 0.038\*\*\* (0.004) |
| PP | -0.207\*\*\* (0.03) | -0.204\*\*\* (0.03) |  |  | -0.195\*\*\* (0.032) | -0.174\*\*\* (0.032) |
| PP\_1 |  |  | -0.209\*\*\* (0.029) |  |  |  |
| PP\_2 |  |  |  | -0.001\*\*\* (0) |  |  |
| avgD | -0.007\*\*\* (0.001) | -0.007\*\*\* (0.001) | -0.007\*\*\* (0.001) | -0.007\*\*\* (0.001) | -0.007\*\*\* (0.001) | -0.006\*\*\* (0.001) |
| avgPT | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) |
| cId | -0.001\*\* (0) | -0.001\*\* (0) | -0.001\*\*\* (0) | -0.001\*\* (0) | -0.002\*\*\* (0) | -0.002\*\*\* (0) |
| cC | 0\* (0) | 0\*\* (0) | 0\* (0) | 0 (0) | 0\*\* (0) | 0\*\* (0) |
| Age | 0\*\*\* (0) | 0\*\*\* (0) | 0\*\*\* (0) | 0\*\*\* (0) | 0\*\*\* (0) | 0\*\*\* (0) |
| tC | 0 (0) | 0\*\* (0) | 0 (0) | 0\*\*\* (0) | 0 (0) | 0\*\* (0) |
| UpF | 0.011\*\*\* (0.003) | 0.013\*\*\* (0.003) | 0.011\*\*\* (0.003) | 0.008\*\* (0.003) | 0.007\*\* (0.003) | 0.006\*\* (0.003) |
| uA | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 0\* (0) | 0\* (0) |
| uAG | 0.009\*\*\* (0.002) | 0.009\*\*\* (0.002) | 0.009\*\*\* (0.002) | 0.009\*\*\* (0.002) | 0.011\*\*\* (0.002) | 0.012\*\*\* (0.002) |
| uF | 0\*\*\* (0) | 0\*\*\* (0) | 0\*\*\* (0) | 0\*\*\* (0) | 0\*\*\* (0) | 0\*\*\* (0) |
| uG | 0.04\*\*\* (0.004) | 0.04\*\*\* (0.004) | 0.04\*\*\* (0.004) | 0.04\*\*\* (0.004) | 0.047\*\*\* (0.004) | 0.048\*\*\* (0.004) |

（续）表5-3 稳健性检验结果

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | (0)R | (1)R\_1 | (2)R | (3)R | (4)R | (5)R |
| uP | -0.008\*\*\* (0.008) | -0.009  (0.008) | -0.008 (0.008) | -0.005 (0.008) | -0.005 (0.009) | -0.004 (0.009) |
| Cons | 0.429\*\*\* (0.028) | -0.535\*\*\* (0.027) | 0.429\*\*\* (0.028) | 0.4\*\*\* (0.027) | 0.619\*\*\* (0.032) | 0.554\*\*\* (0.031) |
| 样本量 | 4610 | 4610 | 4610 | 4610 | 4610 | 4610 |
| 调整R² | 0.2897 | 0.2874 | 0.2901 | 0.2921 | 0.2097 | 0.2195 |

注：括号中代表标准误。

## 5.3 结果解读

在控制专辑的特征（专辑平均时长、专辑平均播放量、专辑所属分类，专辑的评论数、专辑章节数、专辑的发布时长、章节平均更新频率）、以及主播的特征（主播是否经过喜马拉雅官方荣誉认证、主播的等级、主播的粉丝数、主播发布的专辑数、主播的性别）之后，回归结果表明付费节点对付费转化率具有显著的负向影响，回归系数为-0.206，在1%的水平上显著，该结果与假设H1相悖；章节价格对付费转化率具有显著的负向影响，回归系数为-0.13，在1%的水平上显著，该结果与假设H2相符，假设H2得到验证成立；收费模式是否支持VIP服务对付费转化率具有显著的正向影响，回归系数为0.271，在1%的水平上显著，该结果与假设H3相符，假设H3得到验证成立。

## 5.4 关于结果不符合假设的进一步分析

关于假设H1与模型结果不一致的问题，本研究通过进一步的散点图分析以及方差分析对具体的原因进行探究。首先我们对付费转化率与收费节点进行简单线性回归并且画出二维散点图。观察发现存在大量专辑虽然收费节点靠前，但是付费转化率依旧很高，同时也存在部分专辑虽然收费节点很靠后，但是付费转化率却很低。我们用PP=0.5分隔开这些专辑，着重观察收费节点<0.5，但是付费转化率在回归线上方以及收费节点>0.5，但是付费转化率在回归线下方的专辑。即下图中区①和区②的专辑特征（如图5-3所示）。

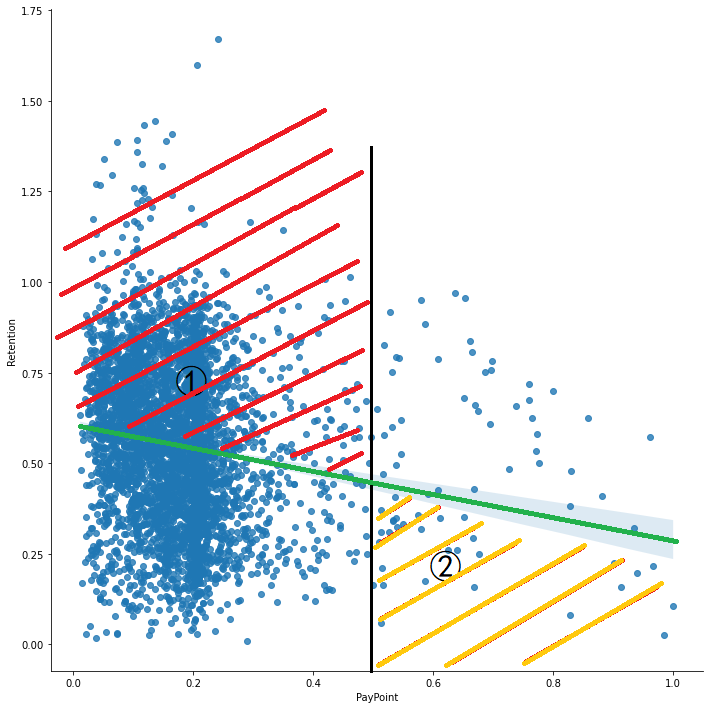


图5-3 收费节点与付费转化率的散点图（标注区域）

本研究对区①的专辑进行了标记，标记为1，其余部分专辑标记为0，区①这部分专辑虽然拥有靠前的收费节点，但是收费对其付费转化率的负向影响很小或者无负向影响。为了进一步研究这部分专辑的特征，探究是否是主播本身的粉丝基础强大以及一些外部因素使得这部分专辑的付费转化率几乎不受节点以及收费的影响，我们对这部分专辑与其他专辑的主播特征进行了t检验以及逻辑回归和方差齐性检验分析，具体结果如表5-4：

表5-4 方差齐性检验

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | G1(0) | Mean1 | G1(1) | Mean2 | MeanDiff | p-Value | 方差齐次检验 |
| uA | 2336 | 80.13 | 2274 | 99.47 | -19.34 | 0.000\*\*\* | sd(0)< sd(1) |
| uF | 2336 | 4.30E+05 | 2274 | 7.80E+05 | -350000 | 0.000\*\*\* | sd(0)< sd(1) |
| uP | 2336 | 0.83 | 2274 | 0.875 | -0.045 | 0.000\*\*\* | sd(0)< sd(1) |
| UpF | 2336 | 1.17 | 2274 | 0.97 | 0.2 | 0.000\*\*\* | sd(0)> sd(1) |

由结果分析可以得到：两部分专辑的样本量差不多，在主播发布的专辑数 (uA) 变量上，区①的专辑所属的主播已发布的专辑数均值比其余专辑的主播多，且在0.01的显著性水平上支持该结论。所以区①专辑的主播发布的专辑数更多；在主播的粉丝数 (uF) 变量上，区①的专辑所属的主播拥有的粉丝数均值比其余专辑的主播多，且在0.01的显著性水平上支持该结论。所以区①专辑的主播所拥有的粉丝数更多；在主播的专业度 (uP) 变量上，区①的专辑所属的主播专业度均值更偏向于1即该主播拥有喜马拉雅荣誉认证称号，且在0.01的显著性水平上支持该结论。所以区①专辑的主播专业度更高；在更新间隔（uF，更新一集所需的时间）变量上，区①的专辑更新时长均值比其余专辑少，且在0.01的显著性水平上支持该结论。所以区①专辑的更新间隔更短；

根据理论分析，区②出现该现象的原因可以解释为付费节点靠后的专辑本身可能是没有足够的粉丝基础，再加上粘性不高的用户在前期长期免费听专辑内容，到某一章节开始收费时，更容易产生不满情绪，带来更高的损失厌恶，然后更可能拒绝为专辑付费。

## 5.5 免费增值最优策略的探索

在上述回归分析中，结果表明在控制了专辑与主播特征后，免费增值的收费策略中，总体而言，收费节点对用户留存有显著负向作用，并且是否支持VIP服务的影响程度最高，章节价格的影响程度最低。为了更深度探讨收费策略对用户留存的影响，进行了进一步的二次回归分析，以及分样本回归（结果如表5-5所示）。序(1)在回归模型的基础上加入了收费节点的二次项进行回归，序(2-3)为序(1)基础上将样本按照是否支持VIP服务分为两个子样本进行回归。

表5-5 二次项回归以及分样本回归结果

|  |  |  |  |
| --- | --- | --- | --- |
| 变量 | (1) | (2) | (3) |
| C(Vip)[T.1] | 0.270\*\*\* (-0.012) |  |  |
| PP | **-0.297\*\*\*** (-0.072) | **-0.352\*\*\*** (-0.076) | **0.710\*\*\*** (-0.173) |
| np.power(PP, 2) | **0.162** (-0.1) | **0.216\*\*** (-0.106) | **-0.868\*\*\*** (-0.218) |
| tP | -0.131\*\*\* (-0.025) | -0.151\*\*\* (-0.028) | 0.038 (-0.042) |
| const | 0.450\*\*\* (-0.028) | 0.732\*\*\* (-0.028) | 0.061 (-0.053) |
| Age | -0.000\*\*\* (0) | -0.000\*\*\* (0) | 0 (0) |
| UpF | 0 (0) | 0 (0) | 0 (0) |

（续）表5-5 二次项回归以及分样本回归结果

|  |  |  |  |
| --- | --- | --- | --- |
| 变量 | (1) | (2) | (3) |
| avgD | -0.007\*\*\* (-0.001) | -0.007\*\*\* (-0.001) | 0.002 (-0.002) |
| avgPT | 0 (0) | 0 (0) | 0 (0) |
| cId | -0.001\*\*\* (0) | -0.001\*\*\* (0) | 0 (-0.001) |
| cCount | 0.000\*\*\* (0) | 0.000\*\* (0) | 0.000\*\* (0) |
| tC | 0 (0) | 0 (0) | 0 (0) |
| uA | 0 (0) | 0 (0) | 0 (0) |
| uAG | 0.009\*\*\* (-0.002) | 0.010\*\*\* (-0.002) | 0.001 (-0.003) |
| uF | 0.000\*\*\* (0) | 0.000\*\*\* (0) | 0 (0) |
| uG | 0.040\*\*\* (-0.004) | 0.041\*\*\* (-0.004) | 0.005 (-0.01) |
| uP | -0.008 (-0.008) | -0.009 (-0.009) | -0.003 (-0.018) |
| Observations | 4,610 | 4,339 | 271 |
| Adjusted R2 | 0.288 | 0.198 | 0.065 |
| Std. Error | 0.184 (df=4593) | 0.187 (df=4323) | 0.105 (df=255) |
| F Statistic | 117.624\*\*\*  (df=16; 4593) | 72.621\*\*\*  (df=15; 4323) | 2.242\*\*\*  (df=15; 255) |

针对收费节点 (PP) 与用户留存率进行进一步分析得到：不区分是否支持VIP服务的全样本回归分析中全样本中，图像可视化结果显示（如下图5-4所示），收费节点越靠前用户留存率越高，不存在最优收费节点，但是当收费节点在91.67%左右时，用户留存最低；在支持VIP服务的子样本中，回归分析结果显示（如下图5-5所示），收费节点越靠前用户留存率越高，不存在最优收费节点，但是当收费节点在61.13%左右时，用户留存最低；在不支持VIP服务的子样本中，回归分析结果显示（如下图5-6所示），使得用户留存最大的收费最优节点为30.68%。

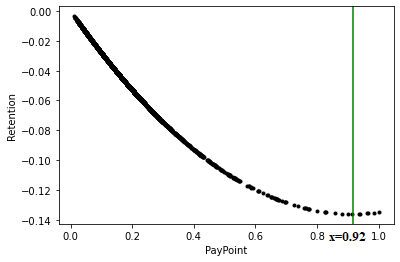


图5-4 收费节点与用户留存的关系图（全样本）

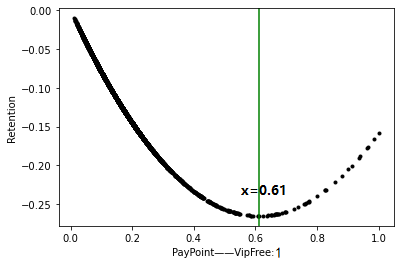


图5-5 收费节点与用户留存的关系图（子样本：Vip=1，支持VIP免费）

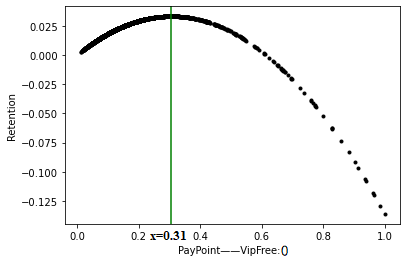


图5-6 收费节点与用户留存的关系图（子样本：Vip=0，不支持VIP免费）

# **6 结论与启示**

## 6.1 研究结论

本研究通过爬虫抓包，将喜马拉雅所有有声音频以及主播的信息获取，通过一系列筛选去除了经过了直播宣传的专辑，声音间情节不连续的专辑，未完结的专辑，排除了直播宣传以及非专业专辑的影响，得到4610张符合研究条件的专辑对象。这4610张专辑都是喜马拉雅免费增值模式的专辑，即起始部分章节免费，从某一节点开始付费，仅一个收费节点，并且经过了筛选以后去除了包含另一免费增值模式即免费试听的专辑，排除了这一模式的影响。

在得到研究数据集以后，本研究对可能对付费转化率有影响的指标进行了筛选，并且结合喜马拉雅FM平台的特征对指标进行了量化与转换。本研究通过控制专辑与主播的特征（专辑每集平均时长、每集平均播放次数、章节数目、评论数、所属类别、平均每集更新间隔、从创建到本研究收集的时间、主播等级、主播发布的专辑数、主播粉丝数、主播是否是喜马拉雅荣誉认证主播、主播性别总共12个指标），进行逐层回归，将控制层作为第一层，按照前进法将解释变量一一纳入回归，最终得到一个完整的多元线性回归模型。模型结果解释了收费节点、章节价格、是否支持VIP服务三个解释变量在专辑与主播特征得到控制的基础上是如何影响被解释变量即付费转化率的。本研究还对模型进行了残差正太检验，多重共线性检验，以及稳健性检验，检验结果均表明模型结果有效并且稳健可靠。回归模型结果表明：支持会员服务会对专辑的付费转化率具有显著的正向影响，与假设相符；章节价格会对专辑的付费转化率产生显著的负向影响，与假设相符；收费节点的位置对付费转化率有显著的负向影响，与假设相悖；收费节点的位置对付费转化率有显著的负向影响，即收费节点越靠后用户的付费意愿越低，与预期不符。

于是为了探究这个现象的原因，本研究对付费节点靠前但是付费转化率依旧很高的专辑的特征进行了探究，通过T检验以及方差齐性检验与其余专辑进行对比，得到结果：相比收费位置对用户转化率正向影响的专辑，收费节点对播放量负向影响很小甚至没有的这部分专辑，在主播专业度，粉丝数，更新频率，发布的专辑数四个指标上，表现为更高的水平。

在得到初步回归的结果后，本研究针对其中收费策略中的是否存在最优收费节点进行了进一步分析，通过二次项回归，以及按照是否支持VIP服务进行分样本回归，得到了是否支持VIP服务下的最优收费节点。

**6.2 研究启示**

通过研究结果可以看出在控制了专辑与主播的特征基础上，支持会员享受专辑的服务会促进专辑的付费转换率提升，所以建议主播在创建发布专辑时选择支持VIP享受专辑的服务，虽然相比购买专辑或者章节，主播从单个用户中可以获得的收益减少，但是由于喜马拉雅的用户群体庞大，而且专辑数量多，专辑的声音数也通常较多，所以支持VIP服务不管是之于平台还是主播都是长期效益递增的。在稳健性检验中，发现另外两种付费模式也符合模型，即购买类型以及收费模式。结果表明：购买类型的两种类型与收费的五种模式虽然都显著影响但是回归系数都小于0.1，所以主播在选择专辑的收费模式时，优先考虑是否支持VIP服务，并且建议选择支持；而购买类型选择整张专辑购买或者单个章节购买对付费转化率的影响不大。

同时，章节的价格会对付费转换率有负向影响，单章章节的价格高，该价格可能与用户预期的章节价值不符合，所以用户可能更不愿意做出付费决策。并且，在所有的专辑层面，用户可能会对比其他专辑的单章价格，在喜马拉雅FM拥有众多可选的替代选择的基础上，如果发现有更低价的替代选择，用户会选择转而去听同等类别甚至同等题材但是价格更低的专辑。所以主播在确定章节价格的时候，需要依据平台的章节总体收费决定该发布的专辑应该使用什么样的价格。总体而言，喜马拉雅的付费专辑中75%以上的专辑平均每个声音的价格在0.2左右，所以主播可以结合自身的粉丝基础，专业度，以及发布的专辑特征比如专辑的章节数，选择0.2±0.05元的价格作为专辑的单集价格，若章节数过高，建议适当降低单集价格。

最后，收费节点，包括相对位置（收费的位置，占总集数的比例，以及收费开始的时长占比），绝对位置（收费开始的集数）对付费转化率是显著的负向影响，与研究的假设相反。这说明存在大量的专辑，其付费转化率受免费增值模式下的收费节点的选择影响很小，可能不受影响，甚至靠后收费节点的位置会降低付费转化率。因此，为了查看这部分专辑的特征，深度归结该现象的原因，本研究进行了进一步的分析。结果显示这部分专辑的主播拥有更多的粉丝基础，更高的专业度，更频繁的更新速度，以及更多的经验和作品，这类主播发布的专辑的收费节点设置对用户付费行为没有强抑制作用，所以该类主播的专辑付费转化率受收费节点限制的程度较小。

结合经济学理论解释：首先，专业度高，粉丝基础好的的主播粉丝黏性更高，而且通常是强粘性，即该用户对专辑的选择以及付费主要因为主播，并且愿意与主播共患难。在喜马拉雅FM中，随着“知识网红”经济的发展，主播像是一个偶像，用户对专辑的选择更倾向于爱屋及乌，而且同等类型的专辑替代选择很多，所以有声专辑的收益越来越取决于主播的专业度以及能力；同时收费节点靠前可能使得用户在还没有充分了解专辑的情况下对专辑的期待更高，所以做出购买行为，按照损失厌恶理论，有声书的免费部分如果展现的消费者最想看到的，而在付费部分隐藏消费者厌恶的或者预期没有那么高的，用户会更愿意做出购买行为。而收费节点靠后会让专辑的内容更多的暴露给用户，更多的内容透明通常会使得用户对专辑的预期价值降低。付费节点靠后的专辑本身可能是没有足够的粉丝基础，再加上粘性不高的用户在前期长期免费听专辑内容，到某一章节开始收费时，更容易产生不满情绪，带来更高的损失厌恶，更可能拒绝为专辑付费。这一结论和Godinho de Matos等[33]得到连续大量观看视频反而会导致购买增值服务的概率减少的研究结论类似，过度暴露产生的损失厌恶对购买行为会造成负向影响。

基于以上结果，在进一步的探索最优收费策略中，本研究引入收费节点二次项，以及按照是否支持VIP服务进行了分样本回归。结果表明：在全样本回归分析以及支持VIP服务的子样本回归分析结果中，收费节点越靠前用户留存越高，均不存在最优收费节点，但是都存在使得用户留存最低的收费节点，分别为全样本：91.67%；支持VIP服务：61.13%；在不支持VIP服务的子样本回归分析中，存在使得用户留存最高的最优收费节点即30.68%。

综合来看，在主播的专业度、粉丝基础给定，以及专辑类别、基本特征确定的情况下，为了获得最优的用户留存，从而实现更高的付费转化率，是否支持VIP服务各位主播可以依据需要选择，如果对于引流要求高，粉丝基础薄弱，建议选择支持VIP服务，在此基础上建议收费节点尽量靠前；粉丝基础雄厚的主播可以选择不支持VIP服务，建议收费节点定为30%，两种方案下都需要制定合理较低的章节价格（0.2±0.05元/集）。

**6.3 研究不足与展望**

本文研究的对象是在线音频制作平台，但是在平台的选择上仅聚焦了喜马拉雅FM这一个市场占有率70%以上的平台。所以本研究在平台的选择上具有局限性。后续的学者可以将市场上在线音频制作平台都纳入研究对象，可能会得到更准确的研究结果。

其次，本研究仅研究了在线音频付费平台的一种免费增值模式，在研究过程中通过数据集筛选排除了免费试听模式对本研究的免费增值模式的影响，但是免费试听与本研究的免费增值模式都属于平台的仅有的两种免费增值模式，后续学者可以进一步探究免费试听对付费转化率的影响以及两种模式共同作用对付费转化率的影响。

基于以上两点局限，未来会全面获取研究对象的数据，并且探究更复杂的免费增值组合模式对付费转换率的影响，同时未来将进一步通过模型对收益进行建模分析，为特定用户发布特定专辑提供使得收益最优的收费节点设置一以及定价，收费标准建议，为各大有声音频创作平台个性化发展主播提供高效的建议。

# 致谢

How time flies!在华科四年的本科生活就快要结束了。

在毕业设计的完成过程中，我想要由衷地感谢我的毕设导师赵玲教授以及香港中文大学（深圳）的张强教授，耐心地指导我的毕业设计，从选题，数据分析，模型设计，方法论，论文撰写等方面给予了及时且宝贵的意见，感谢两位老师认真负责地教导我学习以及论文，并且友善的对我未来的规划以及学习给予真诚的指导。

四年的本科生活，我很感谢国家，社会，学校对我各方面的帮助以及教育；感谢我的家人，一直支持我，让我可以勇敢无畏地追逐自己的梦想；感谢所有教授过我的老师，让我能够在专业学习中，不断挑战自己，不断进步；感谢所有一直陪伴选择我的朋友，陪我一起探索以及度过了美好充实的本科生活，并且携手一起走向更美好的未来；感谢从未想过放弃，坚持真理，自信努力的自己。感谢所有的一切，让我能够成长为一个更加全面发展的个体，让我不断对自己与这个世界拥有更深刻的理解，让我知道个人的有限，知识的无限以及世界的多元；让我体会到自我了解、探索、认知以及自我实现、挑战的快乐；让我更有同理心，去接纳冲突与矛盾并从中学习。

论文写到这，我的毕设基本完成，我的本科生活也即将结束，但是我的人生之旅才刚刚起步。我曾想，在这样信息流的时代，我们接收知识的速度往往赶不上忘却它们的速度，那么，我们学习的意义又是什么？在这四年里，我想我应该找到了答案，我们所经历的一切都是一堂课，所有的记忆都是知识，它们融入到我们的人格中，成为我们的一部分，体现在我们的一言一行，一思一想中。也许付出并没有理想的结果，也许多年后回顾重大决定时会认为有更好的方案，但是所接受的教育教会了我对于所有的一切，都应该用感恩的态度去思考。最后，在此对未来的自己寄予无限的希冀，愿永远美好，永远善良，永远保持谦逊求知，更好地借用计算机语言去解释这个世界，去解决问题。

# 参考文献

1. 宗利永,夏子惠. 订阅式知识付费平台用户满意度影响因素研究——基于喜马拉雅FM用户评论文本挖掘的分析[J].图书情报研究,2020, 13(04): 44-50.
2. 敖雅倩. “喜马拉雅FM”的盈利模式与盈利策略探析[J].新媒体研究,2018, 4(13): 42-43. DOI:10.16604/j.cnki.issn2096-0360.2018.13.016.
3. 李琳. 移动网络音频平台的付费内容研究[D].贵州民族大学,2021. DOI:10.27807/d.cnki.cgzmz.2021.000354.
4. 张思嘉. 有声读物APP“懒人听书”运营策略研究[D].河南:河南大学,2020.
5. 丁翱林. 探析知识付费平台营销策略及启示——以喜马拉雅FM为例[J].新闻研究导刊,2021, 12(02): 116-117.
6. 蒋岩汐. 知识付费产品价值形成机制研究[D].江西财经大学,2021. DOI:10.27175/d.cnki.gjxcu.2021.001021.
7. 王雨萍. 虚拟社区用户知识付费行为影响因素研究[D].吉林大学,2021. DOI:10.27162/d.cnki.gjlin.2021.000920.
8. 罗钦芳. 知识付费产品销量的跨层次研究[D].厦门大学,2018.
9. 刘征驰,马滔,叶宇阳. 隐匿价值、社群学习与知识付费动态定价[J/OL].中国管理科学:2021: 1-12 DOI:10.16381/j.cnki.issn1003-207x.2019.1447.
10. 刘征驰,黄雅文,马滔. 非授权分享、定价分歧与知识付费平台运营策略[J].中国管理科学,2021, 29(03): 168-175. DOI:10.16381/j.cnki.issn1003-207x.2018.0950.
11. K. Rothmeier, N. Pflanzl, J. A. Hüllmann and M. Preuss, "Prediction of Player Churn and Disengagement Based on User Activity Data of a Freemium Online Strategy Game," in IEEE Transactions on Games, 2021, 13(01): 78-88, doi: 10.1109/TG.2020.2992282.
12. J. Runge, P. Gao, F. Garcin and B. Faltings, "Churn prediction for high-value players in casual social games"[C]. 2014 IEEE Conference on Computational Intelligence and Games, 2014: 1-8, doi: 10.1109/CIG.2014.6932875.
13. K. Sangaralingam, N. Verma, A. Ravi, S. W. Bae and A. Datta, "High Value Customer Acquisition & Retention Modelling – A Scalable Data Mashup Approach"[C]. 2019 IEEE International Conference on Big Data (Big Data), 2019: 1907-1916, doi: 10.1109/BigData47090.2019.9006106.
14. 柯金燕. 移动游戏中的社交体验对用户留存意向影响的实证研究[D].中国社会科学院研究生院,2020. DOI:10.27642/d.cnki.gskyy.2020.000009.
15. 吴爽. 知识付费平台高知女性用户流失行为影响因素研究[D].南昌大学,2020. DOI:10.27232/d.cnki.gnchu.2020.001648.
16. 周勤琪. 短视频社交媒体的用户流失影响因素研究[D].哈尔滨工业大学,2019. DOI:10.27061/d.cnki.ghgdu.2019.004543.
17. W. Su, Y. Li, H. Zhang, et al, How the attributes of content distributors influence the intentions of users to pay for content shared on social media. Electronic Commerce Research 2021. <https://doi.org/10.1007/s10660-021-09482-z>.
18. 杨东红,贺红梅,徐畅. 移动音频有声阅读平台用户知识付费行为研究[J].情报学,2020, 38(07): 105-111+172. DOI:10.13833/j.issn.1007-7634.2020.07.015.
19. 邓胜利,蒋雨婷. 用户交互特征对知识付费行为预测的贡献度研究[J].图书情报工作,2020, 64(08): 93-102. DOI:10.13266/j.issn.0252-3116.2020.08.011.
20. 王雪莲,高凯凯,李嫄. 在线知识付费节目的顾客体验价值研究——基于内容分析方法对“喜马拉雅”平台数据的发掘分析[J].价格理论与实践,2021(06): 141-144. DOI:10.19851/j.cnki.cn11-1010/f.2021.06.119.
21. Z Y. LI, Y. CHENG, From Free To Fee: Exploring The Antecedents Of Consumer Intention To Switch To Paid Online Content[J]. Journal of Electronic Commerce Research, 2014, 15(4): 281-299.
22. J. Jing and C. Lu, “A Study on Factors Affecting Chinese Users’ Willing to Pay for Online Paid Knowledge Contents: Focusing on Theory of Planned Behavior,”[J]. Journal of Digital Convergence, 2020, 18(2): 151–162.
23. 赵宇翔,刘周颖,朱庆华. 从免费到付费：认知锁定对在线问答平台中提问者转移行为的影响研究[J].情报学报,2020, 39(05): 534-546.
24. 卢恒,张向先,肖彬,刘苡声. 在线用户知识付费意愿的影响因素及其调节变量：元分析研究[J].图书情报工作,2021, 65(13): 44-54. DOI:10.13266/j.issn.0252-3116.2021.13.005.
25. 夏万虎. 免费增值模式的B2B用户付费的影响因素实证研究[D].南京大学,2020. DOI:10.27235/d.cnki.gnjiu.2020.001429.
26. 曾润滋. 免费增值模式文献综述[J].现代商业,2021(13): 26-28. DOI:10.14097/j.cnki.5392/2021.13.008.
27. 郭敏. 迅雷免费增值商业模式价值创造研究[D].河南财经政法大学,2021. DOI:10.27113/d.cnki.ghncc.2021.000041.
28. Z. Li, G. Nan and M. Li, "Advertising or Freemium: The Impacts of Social Effects and Service Quality on Competing Platforms,"[J] IEEE Transactions on Engineering Management, 2020, 67(1): 220-233, doi: 10.1109/TEM.2018.2871420.
29. O.F. Koch, A. Benlian, The effect of free sampling strategies on freemium conversion rates.[J] Electron Markets 27, 2017: 67–76. <https://doi.org/10.1007/s12525-016-0236-z>.
30. 陈建. 免费试用经济溢出效应分析[J].经济研究导刊,2014(17):207-209.
31. F. Bram, G. Els, Try It, You’ll Like It—Or Will You? The Perils of Early Free-Trial Promotions for High-Tech Service Adoption.[J] Marketing Science, 2016, 35(5):810-826. https://doi.org/10.1287/mksc.2015.0973.
32. Z. Hua, Y. Fan, X. Xu and L. Bao, Optimal Length of Free Trial for Online Service Considering User's Learning Effect.[J] IEEE Transactions on Engineering Management, 2019, 66(4): 583-597, doi: 10.1109/TEM.2018.2877783.
33. M. Godinho de Matos, P. Ferreira, The effect of binge-watching on the subscription of video on demand: Results from randomized experiments.[J] Information Systems Research, 2020, 31(4): 1337-1360.

# 附录一：数据获取部分代码

def random\_user\_agent():#根据已有代理随机分配

with open('./config/user\_agent.txt') as f:#文件

agents\_temp = f.readlines()

user\_agents = [line.rstrip() for line in agents\_temp]

return random.choice(user\_agents)

def get\_response(url, html=False, \*\*kwargs):#请求网站爬取数据

if 'headers' not in kwargs:

kwargs['headers'] = {

'User-Agent': random\_user\_agent(),

"Cookie": "",

}

elif 'User-Agent' not in kwargs['headers']:

kwargs['headers']['User-Agent'] = random\_user\_agent()

while True:

try:

with requests.get(url, \*\*kwargs) as r:

if r.status\_code == 429:#响应码正常显示网页

kwargs['headers']['User-Agent'] = random\_user\_agent()

continue

if r.status\_code == 404:#网页丢失

return None

r.raise\_for\_status()

# parse response in html or json

if html:

dom = etree.HTML(r.text)

else:

dom = json.loads(r.text)#

return dom

#response indicates whether return user-agent and cookies

ua = kwargs['headers']['User-Agent']

cookies = r.cookies

if response:

return dom, ua, cookies

else:

return dom

except requests.exceptions.ConnectTimeout:

continue

except requests.exceptions.Timeout:

continue

except requests.exceptions.HTTPError as e:

print(e)

time.sleep(5)

continue

except Exception as e:

time.sleep(5)

continue

#以获取专辑信息为例

def get\_album\_info(self,albumId):#获取专辑信息

retry = 3

while retry>0:

try:

res = self.s.get(f'https://mobile.ximalaya.com/mobile-album/album/page/ts-1630507657566?ac=WIFI&albumId={albumId}&device=android&isAsc=true&isQueryInvitationBrand=true&isVideoAsc=true&pageId=1&pageSize=20&pre\_page=0&source=5&supportWebp=true')

data = res.json()['data']

data.pop('tracks',None)

data['\_id'] = data['album']['albumId']

data['collection\_time'] = datetime.now()

return data

except BaseException as e:

retry -=1

print(e)

if 'album' not in data.keys():

print(data)

time.sleep(3)

#将数据写入MongoDB数据库

def write\_album(self,data): self.db.db.album\_info.update\_one(filter={'\_id':data['\_id']},update={'$set':data},upsert=True)

# 附录二：取声音title中编号的部分代码

import pymongo

import pandas as pd

import numpy as np

import os,re

import json

from pycnnum import cn2num, num2cn

class ConnDB:

def \_\_init\_\_(self):

self.client = pymongo.MongoClient(host='localhost', port=27017)

self.db = self.client.xmly

self.data=self.db.restAtrackcount

self.track=self.db.restTrackTitle

self.restsituation=self.db.allsituation

print('Successfully Connected!')

def close(self):

self.client.close()

class UseDB:

def \_\_init\_\_(self):#连接MongoDB数据库

self.cdb = ConnDB()

###获取专辑id

def getalbumid(self):

l=list()

l1=list()

with self.cdb.data.find({},{'trackcount':1}) as cursor:

for item in cursor:

l.append(item["\_id"])

l1.append(item['trackcount'])

print(len(l))

self.OnlyFirstNum(l, l1)#取第一个数字

self.OnlySecondNum(l, l1)#取第二个数字

self.OnlyLastNum(l, l1)#取最后一个数字

self.Di(l, l1)#第几章第几节

self.DiWithNum(l, l1)#第几章第几节+1,2,3...

self.ZW(l, l1)#+中文简体、繁体

self.Multi\_spot(l, l1)#1.1/1.1.1

self.Multi\_xiegang(l, l1)#1-1/1-1-1

###此处以其中一种方法为例

def OnlyFirstNum(self,l,l1):

jishu=0

lT=list()

for aId,tc in zip(l,l1):

notdigitfromo=0

feizheng=0

fromzero=False

fromone=False

fromgtone=False

alltitle=list()

notnumtitle=0

digititle=list()

notnumtlist=list()

newlist=list()

lengthOfDuplicate=0

OK=0

with self.cdb.track.find({'albumId':aId},{'title':1}) as cursor:

for item in cursor:

x=re.findall(r"\d+",item['title'])

track=list()

if x!=[]:

###只取第一个数字

digititle.append(int(x[0]))

###只取第二个数字

digititle.append(int(x[1]))

###只取最后一个数字

digititle.append(int(x[int(len(x))-1]))

###1.1/1.1.1

x=re.findall(r"\d+\.\d+\.\d",item['title'])

###1-1

x=re.findall(r"\d+\-\d+\-\d",item['title'])

if x==[]:

x=re.findall(r"\d+\-\d{1,4}",item['title'])

###第几章第几节

x=re.findall(r"第+\d+[集章回季卷册篇片节课讲周天期部]",item['title'])

if x!=[]:

x=re.findall(r"第+[一二三四五六七八九十百千]+[集章回季卷册篇片节课讲周天期部]",item['title'])

###第几章第几节+1,2,3...

NUM=re.findall(r"\d$",item['title'])

x=re.findall(r"第.+第+[一二三四五六七八九十百千\d]+[集章回季卷册篇片节课讲周天期部]+\d",item['title'])

###中文，简体/繁体编号

NUM=re.findall(r"\d$",item['title'])

x=re.findall(r"[(（]+[一二三四五六七八九十壹贰叁肆伍陆柒]+[)）]",item['title'])

track.append({'\_id':item['\_id'],'title':int(x[0]),'albumId':aId,'originaltitle':item['title']})

# print(track)

elif x==[]: track.append({'\_id':item['\_id'],'title':item['title'],'albumId':aId})

notnumtlist.append(item['title'])

x1=re.findall("片+[花头]",item['title'])#标记非正片

y1=re.findall("预告",item['title'])

z1=re.findall("[引楔]+子",item['title'])

c1=re.findall("彩蛋",item['title'])

f1=re.findall('番外',item['title'])

tb1=re.findall('特别篇',item['title'])

fkc=re.findall('发刊词',item['title'])

if x1!=[] or y1!=[] or z1!=[] or c1!=[] or f1!=[] or tb1!=[]or fkc!=[]:

feizheng+=1

self.cdb.t.insert\_many(track)

notnumtitle=len(notnumtlist)

##Duplicate 查看是否含有重复编号

result={}

s=0

dupEpisode=list()###重复章节

for i in np.unique(digititle):

result[i]=digititle.count(i)

for k,v in result.items():

if v!=1:

lengthOfDuplicate+=v

s=str(str(k)+'—'+str(v))

dupEpisode.append(s)

lenthOfD=len(dupEpisode) #重复数目

###判断编号是否连续

digititle.sort(key=int)

sumof=0

c=0

if lenthOfD!=0:

digititle=list(set(digititle))

digititle.sort(key=int)

if digititle!=[]:

i=int(digititle[0])

if i==0:

fromzero=True##标记从0开始的

elif i==1:

fromone=True##标记从1开始的

elif i>1:

fromgtone=True##标记从大于1开始的

c=0

for l in digititle:

c=i-int(l)

i+=1

if c<0:

c=-c

sumof+=c

percentOfNum=(len(digititle)+lengthOfDuplicate-lenthOfD)/tc

percentofnotnum=notnumtitle/tc

if (percentofnotnum+percentOfNum)!=1:

jishu+=1

lT.append(aId)

###判断是否专辑所有的章节都有编号，并且连续无重复

if sumof==0 and lenthOfD==0 and notnumtitle==0:

OK=1#标记为OK：1

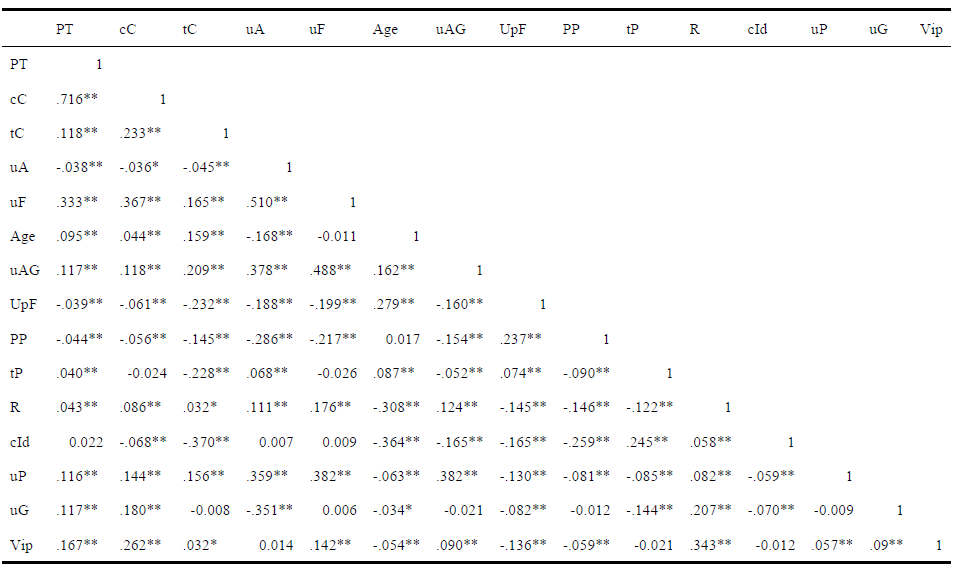
newlist.append({'\_id':aId,'fromzero':fromzero,'fromone':fromone,'fromgtone':fromgtone,'sumOfchazhi':sumof,'lengthOfD':lenthOfD,'dupEpisode':dupEpisode,'alltitle':alltitle,'digittitle':digititle,'notnumtitle':notnumtitle,'notnumtlist':notnumtlist,'oriCount':notdigitfromo,'feizhengtcount':feizheng,'lengthOfDuplicate':lengthOfDuplicate,'percentOfNum':percentOfNum,'OK':OK})

self.cdb.restsituation.insert\_many(newlist) ##存入MongoDB中

print(aId)

print(lT)

# 附录三：相关分析完整结果





**本科生毕业设计（论文）任务书**

|  |  |
| --- | --- |
| 题 目 | **喜马拉雅FM免费增值策略对用户留存的影** |
|  | **响的实证研究** |

（任务起止日期：2021年11月2日～2022年6月5日）

|  |  |
| --- | --- |
| 院 系 | **管理学院** |
| 专业班级 | **信息管理与信息系统1802班** |
| 姓 名 | **刘棫欣** |
| 学 号 | **U201812791** |
| 指导教师 | **赵玲** |

教研室（系、所）负责人 2021年10月28日审查

院（系）负责人 2021年11月2日批准

|  |
| --- |
| 课题内容：  ①数据源：收集喜马拉雅FM的所有数据，该数据为截面数据，取自于某一个时间点；  ②数据集：进行数据清理，得到用于分析的初步数据集；  ③分析：  建立猜想：针对专辑，专辑里的章节，发布专辑的主播三个层面可能对用户行为造成影响的变量，建立猜想。  对这些变量进行分类转换成一个更高维的指标，比如用户行为通过流行度这一指标来衡量，专辑的流行度通过专辑和章节层面的一些由用户产生的数据变量量化。  然后对这些指标与流行度进行回归分析，建模测试模型，再通过functional data analysis对这个模型进行验证分析。最后从模型中找到付费模式下流行度与免费专辑流行度随着章节的走向趋势最接近的各类指标值，尤其是价格指标，从而找到收费对用户行为影响最小的价格，使得平台最大收益化。 |
| 课题任务要求：  研究对象：对在线音频收听流行度的影响因素探究。  从以下三个层面出发，喜马拉雅的专辑，专辑对应的章节以及发布专辑的主播，来探究在线音频中对用户行为影响的可能因素；同时探索免费与付费等不同收费模式对专辑流行度的影响程度，进而通过建模找到使得流行度与收费的关系，找到使得该影响最小化的临界收费价格，辅助定价。 |
| 主要参考文献（由指导教师选定）：   1. Su, W., Li, Y., Zhang, H. et al. How the attributes of content distributors influence the intentions of users to pay for content shared on social media. Electron Commer Res (2021). https://doi.org/10.1007/s10660-021-09482-z 2. 李琳. 移动网络音频平台的付费内容研究[D].贵州民族大学,2021.DOI:10.27807/d.cnki.   cgzmz.2021.000354.   1. 喜马拉雅 FM 知识付费模式营销战略分析 中国传媒大学广告学院毕业论文 2. 喜马拉雅 FM 盈利模式与盈利策略探析 中图分类号 G2 文献标识码 A 文章编 号 2096-0360 (2018）13-0042-02   [5] 宗利永,夏子惠.订阅式知识付费平台用户满意度影响因素研究——基于喜马拉雅FM用户评论文本挖掘的分析[J].图书情报研究,2020,13(04):44-50.  [6] Bram Foubert, Els Gijsbrechts (2016) Try It, You’ll Like It—Or Will You? The Perils of Early Free-Trial Promotions for High-Tech Service Adoption. Marketing Science  Published online in Articles in Advance 31 May 2016. <http://dx.doi.org/10.1287/>mksc.  2015.0973 |
| 同组设计者  无 |
| 指导教师签名：  年 月 日 |