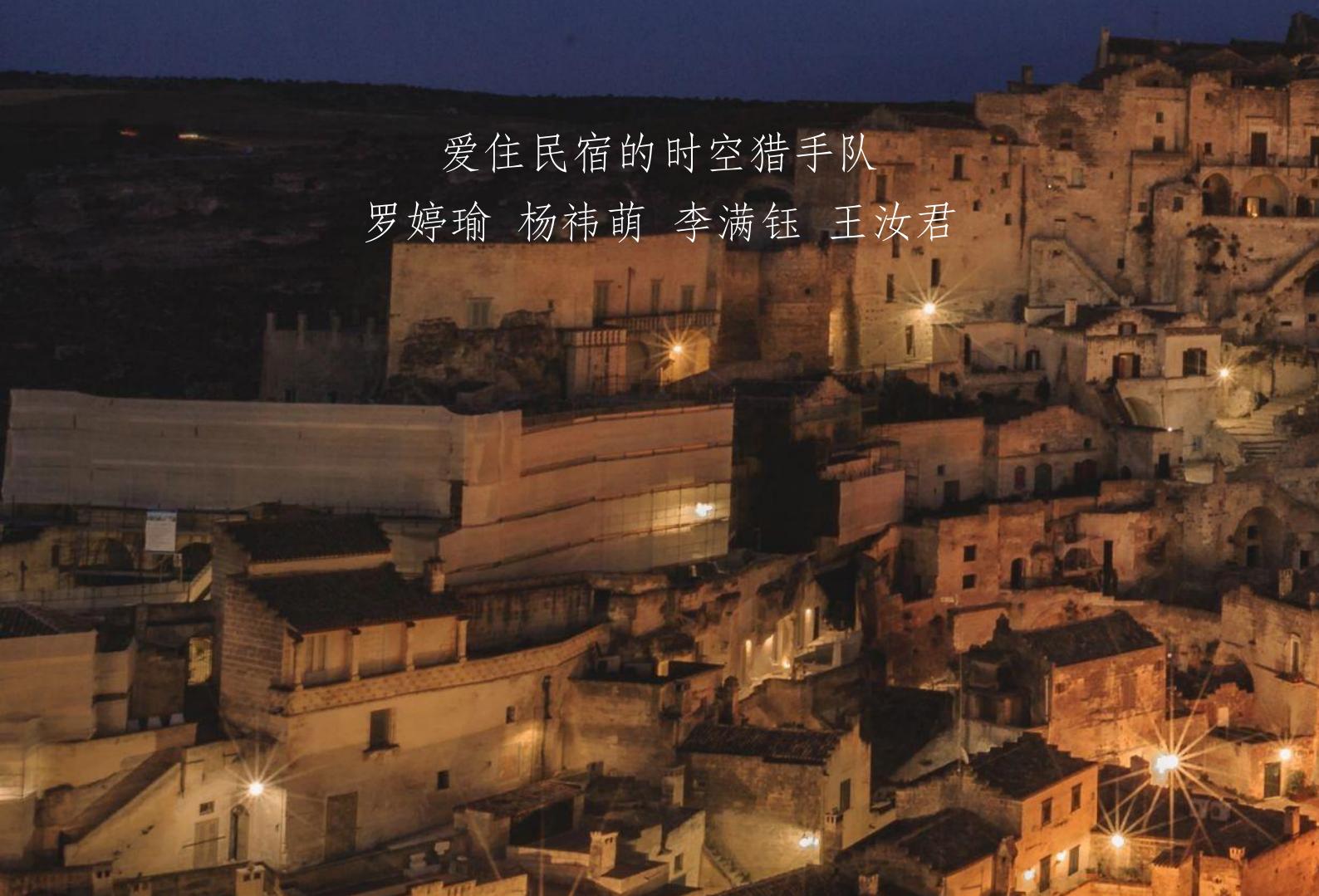


# 何曾住处 限西东

基于供需空间格局特征的武汉市民宿市场的调查研究

爱住民宿的时空猎手队

罗婷瑜 杨祎萌 李满钰 王汝君



## 摘要

随着现代旅游业的快速发展和大众旅游消费方式的转变，游客对旅游住宿的需求越来越趋向多样化和个性化，民宿产业迅速兴起，席卷了一股民宿热潮；政府部门也高度重视民宿新型业态的打造，民宿产业正在各地政府的积极推动下不断发展。然而，当前民宿市场发展尚处于满地开花的自由式发展阶段，民宿市场供给呈现何种空间格局？消费者对民宿市场存在何种需求？如何基于供需匹配关系优化民宿市场空间格局？这些问题的解答是民宿市场健康发展的前提。基于此，本团队面向民宿市场供需平衡发展目标，采取网络爬虫和问卷调查相结合的方法，深入开展武汉市民宿市场调查分析，识别武汉市民宿市场空间格局特征，挖掘民宿市场发展背后隐藏的价值信息，提出民宿市场发展的优化策略，为实现民宿市场空间布局优化和推动民宿市场健康发展提供决策参考。

本团队从美团、爱彼迎、携程等 APP 上共爬取评论数据 78149 条，在经过数据清洗后，使用中文分词、词共现分析、情感倾向分析和 LDA 主题模型等算法和模型，实现对文本评论数据隐藏信息的挖掘与分析。依据结合路网数据及爬虫得到的民宿点位置，采用网络分析的方式计算民宿到周围设施点的最短路径，分析其空间可步性及可达性，实现对现有民宿空间格局的综合评估，为未来民宿的选址提供科学参考。在小范围预调查的基础上，确定正式调查方案，有针对性地设计调查问题；基于民宿意愿的类型划分，运用描述性统计的方法分析消费者接受性、消费者满意度和潜在消费者特征，运用 logistic 二元回归模型和排序选择模型对不同消费群体消费行为特征及需求进行分析，运用层次分析法和空间探索性分析开展武汉市民宿发展适宜性评价，进而提出民宿市场发展的重点方向和优化策略。

本团队的调研分析表明：①民宿消费者特征明显，25~30 岁群体是民宿市场消费的主力军，年轻群体更注重环境氛围感、人情味、价格合理；②中、老年群体更关注民宿带来的本地风土人情，高消费、中层年龄阶段、流动性较大的人群作为潜在消费者值得发展；③卫生问题、安全性问题和规范性问题是民宿行业发展的主要限制性因素，是影响消费者不愿意规避民宿旅游的主要原因；④武汉市民宿可步性和可达性较高，但实用性相对不足；近四成的武汉市民宿发展适宜性达到 6 分及以上，武汉市民宿市场具有发展潜力。

针对民宿调查分析结果，我们提出以下几点民宿市场发展建议：①加强民宿优势点建设，重点在交通便利、环境氛围和价格实惠方面予以优化完善；②充分结合自身优势打造民宿品牌标签，进行客源市场细分，对不同群体进行不同营销方案；③建立统一的清洁标准，提高整体卫生水平；④将入住流程规范化、透明

化，提供完整的服务体系，提供购买凭证；⑤加强安全措施保障，最大程度上满足用户的安全需求；⑥针对潜在客户群体关注点，着重“品质”、“聚会”等话题，打造故事性视频进行引流。

**关键词：**民宿；问卷调查；网络爬虫；空间格局；空间适宜性；优化策略

## 目 录

一、问题综述.....	8
(一) 研究背景.....	8
1. 政策环境分析.....	8
2. 行业现状及调查目的.....	8
(二) 文献评述.....	9
1. 国外相关研究.....	9
2. 国内相关研究.....	9
3. 文献评述.....	10
(三) 问题归纳.....	10
二、研究思路与创新.....	13
(一) 研究思路.....	13
(二) 研究特色与创新.....	14
1. 基于大数据的文本挖掘.....	14
2. 地理视角下的民宿空间格局网络分析.....	14
3. 顾及消费者需求特征的民宿适宜性评价.....	14
4. 基于民宿与酒店对比分析的优化策略.....	14
三、基于武汉市民宿评论数据的挖掘与分析.....	16
(一) 数据来源背景.....	16
(二) 数据搜集.....	16
1. 有效数据爬取.....	16
2. 无效数据处理.....	20
(三) 数据的处理.....	21
1. 中文分词.....	21
2. 分词的原理.....	22
3. 词频统计.....	22
4. 词共现分析.....	24
5. 基于情感词典的情感倾向性分析.....	25
6. 基于 LDA 模型的主题分析.....	26
四、民宿空间格局特征分析.....	30
(一) 民宿基本格局.....	30
(二) 民宿可步性.....	32
1. 基础步行指数计算原理.....	32
2. 路网分析.....	32
3. 可视化展示.....	34
(三) 民宿可达性.....	35
1. 整体实验设计.....	35
2. 路网分析.....	35
3. 可视化展示.....	36
五、问卷调查方案和实施策划.....	38
(一) 调查方案设计.....	38
1. 调查目的.....	38
2. 调查内容.....	38

3. 调查对象和方法.....	38
4. 抽样设计.....	39
5. 主要数据分析方法.....	39
(二) 调查组织分工.....	41
1. 调查组织分工.....	41
2. 调查实施进度.....	41
六、质量控制.....	44
(一) 预调查数据检验及问卷修改.....	44
1. 数据检验.....	44
2. 问卷修改.....	45
(二) 正式调查数据处理与检验.....	46
1. 数据处理.....	46
2. 数据检验.....	46
七、总体认知描述分析.....	50
(一) 消费者基本特征.....	50
1. 性别分布.....	50
2. 年龄分布.....	50
3. 地区分布.....	51
4. 职业分布.....	52
5. 月消费分布.....	52
6. 消费者在外住宿的目的分布.....	53
7. 消费者在外住宿的频率分布.....	53
8. 消费者住宿信息认知渠道.....	54
9. 消费者民宿居住意愿.....	55
(二) 民宿情况调查与分析.....	55
1. 选择民宿的原因分析.....	55
2. 民宿价位分布.....	56
3. 民宿类型分布.....	57
4. 民宿风格分布.....	57
5. 民宿设施偏好.....	58
6. 民宿出行情况分析.....	58
7. 民宿出行情况分析.....	59
8. 现有民宿满意程度分析.....	60
(三) 酒店情况调查与分析.....	60
1. 偏好酒店原因分析.....	60
2. 酒店价位分布.....	61
3. 酒店出行情况分析.....	61
4. 酒店选择因素分析.....	62
八、调查问卷交叉分析.....	65
(一) 消费者在外住宿目的分析.....	65
(二) 在外住宿频率的分析.....	66
(三) 民宿风格.....	67
(四) 消费者的选择民宿的关注因素分析.....	68
(五) 消费者的选择民宿的原因分析.....	68

# 目 录

九、民宿消费者特征 Logistic 回归模型分析.....	71
(一) 模型的建立.....	71
(二) 结果分析.....	71
1. 性别.....	72
2. 年龄.....	72
3. 职业.....	72
4. 月消费.....	73
十、民宿消费者影响因素分析.....	75
(一) 排序选择模型的建立.....	75
(二) 模型估计与分析.....	75
(三) 模型结果分析.....	76
1. 性别.....	76
2. 年龄.....	76
3. 消费水平.....	76
4. 住宿目的.....	77
十一、武汉市民宿适宜性评价.....	79
(一) 模型构建.....	79
(二) 综合指数计算.....	81
(三) 模型结果分析.....	82
十二、结论和建议.....	85
(一) 调查结论.....	85
1. 武汉市民宿供给格局.....	85
2. 武汉市民宿消费群体特征.....	85
3. 民宿消费者体验及需求情况.....	85
(二) 优化策略.....	86
1. 针对现有客户群体的建议.....	86
2. 针对潜在客户群体的建议.....	87
十三、附录.....	89
附录一：参考文献.....	89
附录二：调查问卷.....	90
附录三：问卷编码.....	97
附录四：二元选择模型变量表.....	104
附录五：排序选择模型变量表.....	106



# 第一章 问题综述

## 一、问题综述

民宿主要指利用本地闲置资源，为游客提供体验当地文化与生活方式的小型住宿设施。民宿不仅提供了基本的旅游住宿、体验功能，还包含娱乐等多项服务。从类型来看，民宿可分为农村型与城市型，除了带有浓厚乡村地域色彩的农家乐，立足于城市空间发展的民宿群同样结合了当地特色文化和民宿特点形成独有的个性化特色和商业性质，引起了学术界和产业界的广泛关注，城市型民宿也是本文将重点调研的对象。

### (一) 研究背景

#### 1. 政策环境分析

近年来，民宿市场在国家政策推动下有序发展。

2018年9月国务院办公厅颁发《完善促进消费体制机制实施方案（2018-2020年）》鼓励发展租赁式公寓、民宿客栈等旅游短租服务；推进经济型酒店连锁经营，鼓励发展各类生态、文化主题酒店和特色化、中小型家庭旅馆。

2019年7月国家文化和旅游部发布最新版《旅游民宿基本要求与评价》，代表着民宿行业标准的基本形成。于2021年进行了一次修改。这些政策的制定为民宿市场的行业规范指明方向。

2019年12月武汉市政府下发《关于促进全市旅游民宿规范管理和健康发展的意见》明确旅游民宿开办基本条件和部门监管职责，优化简化审批流程，解决长期困扰旅游民宿经营者无法办证、管理无章可循等突出问题。

2021年7月，在武汉市文化和旅游局的支持下，武汉市旅游民宿协会举办武汉市旅游民宿管理提升项目，旨在促进武汉民宿行业更规范、更标准，培育出更多的民宿经营人才，打响武汉民宿这张名片。

在这样的政策环境中，民宿行业成为共享经济的新亮点，具有可观的发展前景。

#### 2. 行业现状及调查目的

在民宿经济环境方面，中国旅游与民宿发展协会发布的研究报告中指出，2015至2020年民宿市场交易额逐年上涨，2021至今民宿行业发展按下加速键。

在社会文化需求方面，民宿产业充分挖掘原生地域文化，满足大众内在的精

神需求。

民宿市场焕发活力，具有不容小觑的发展潜力以及研究价值。基于此，我们也将从多方面展开调查研究。旨在运用多种方法分析民宿消费者行为及需求，结合空间格局对武汉市民宿进行多维度分析，最后建立系统的评价体系，为民宿市场提供多角度发展策略。

## （二）文献评述

由于国内外民宿产业的起步时间不同，发展时间和相应的研究热点也存在一定差异，本文从国内、国外两个方面展开相关评述。

### 1. 国外相关研究

二十世纪六十年代，美国和日本部分地区也相继出现民宿业态，在一些地区得到了长足的发展。国外对民宿的研究始于二十世纪八十年代，理论和实证研究成果较为丰富，主要从民宿产业发展、相关主体行为、市场营销<sup>[1][2]</sup>、民宿旅游<sup>[3]</sup>等方面开展研究。

很多学者表示民宿产业的发展是一把双刃剑，发展民宿能够为边远地区缺乏其他工作的当地人提供持续性收入，且能提升民宿经营者对当地自然环境及文化遗产保护的责任感 (BHALLA P, 2016)。但民宿行业的发展也为社会带来一定弊端，经营民宿会增加当地经济对于旅游业的依赖性 (CHURYENA, 2015)，降低社会凝聚力和公平性。

在对民宿旅游者的研究方面，调整压力和放松是游客选择民宿的主要消费动机 (QU J.A, 2018)。千禧年轻一代是民宿消费群体的重要组成部分 (LIN Y-C, 2019)”，可以设立更多类型的活动来吸引年轻人。在民宿经营者方面，自治性、冒险精神、主动性、竞争积极性和创新性是经营者创业的重要驱动因素 (Lumpkin, 1996)，生活质量、生活态度、社区的作用和政府的作用也与民宿成功发展呈正向关系 (Yusnita, 2016)。

### 2. 国内相关研究

二十世纪八十年代，台湾本地居民利用自有的空闲房屋招揽顾客，民宿业态由此开始发展。中国大陆地区的民宿在 2010 年后才开始倾向于专业化、高端化方向发展。本世纪以来民宿的研究逐渐受到国内多个领域学者的重视，关于民宿的研究主要内容集中在服务质量与满意度评价、民宿设计、民宿开发等方面。

# 第一章 问题综述

在消费者对民宿的满意度研究方面，民宿档次低、雷同性大、管理人员综合素质不高、设施设备不齐全、政策体系不健全等是游客对民宿满意度偏低的原因（范少花，2016），特色不突出、体验情境设置不足等也是当前民宿应该注意改进的问题（董之昱，2017）。

在民宿设计与建议方面，传统的酒店强调“住”带给客人的服务性感受，设计的模式和建设的程序相对固定化，而民宿在设计建设过程中高度关注人性化和个性化（赵菁，2017）。在民宿的设计中要将生活符号艺术化、文化符号系统化、形象符号商品化、商品符号产业化，才能突出民宿周边自然和人文风情的独特性和专属性（王明泰，2015）。

## 3. 文献评述

在文献研究中，对民宿的发展历程、乡村民宿开发、民宿设计等相关问题的研究已经较为透彻，对本次的调研具有一定的参考价值，指明了调研方向。但仍有部分问题并没有较为全面且深入的研究，需要我们进一步做出调研与分析。

关于民宿消费者的研究中，针对消费者行为以及消费者对民宿的各方面态度及需求的研究存在一定局限性。部分文献使用的分析方法过于片面，不能全面地反应民宿消费者的真实需求。部分文献虽然从多个维度进行分析，但是选取的角度比较浅显，只能笼统地对消费者情况进行概述。

针对民宿评价体系的建立的问题，大部分学者仅从较为宏观的价值层面提出对标准化评价标准的倡议，并未提出具体的、切实可行的解决方法。近年来部分学者给出了相关的评价办法，但是研究方法较为局限，覆盖面较窄，缺乏深度和客观性。这样的情况从根本上不利于民宿市场的发展，更无法切实地对民宿发展建设给出系统性地建议。

本次调研则立足于现状，从消费者视角出发，结合多种数据分析方法和建立模型等方法分析消费者行为，同时结合空间格局分析对民宿市场进行多维度的剖析，针对武汉市民宿建立系统而全面的评价体系，从新的角度提供建设性意见，并结合多个角度为武汉市民宿市场的发展提供策略。

### （三）问题归纳

在文献研究中，对民宿的发展历程、民宿开发、民宿设计等相关问题的研究已经较为透彻，对本次的调研具有一定的参考价值。但仍有部分问题并没有较为全面且深入的研究，需要我们进一步做出调研与分析。

针对消费者行为及对民宿态度及需求的研究存在一定局限性，部分文献使用的分析方法单一，不能全面反应消费者真实需求，部分文献调查结果深度不够。在民宿评价方面，大部分学者仅从较为宏观的价值层面提出对标准化评价体系的倡议，并未提出具体的、切实可行的解决方法。近年来部分学者提出相关评价措施，但是研究方法覆盖面较窄，缺乏深度和客观性。

此外，目前的研究多集中在旅游学、经济学、管理学及社会学等角度，从地理、空间角度出发的研究较少，缺乏对民宿市场供需空间格局的统筹，容易带来民宿发展空间不合理、供需不平衡等问题。这样的情况从根本上不利于民宿市场的发展，更无法切实对民宿发展及建设提供系统建议。

基于以上问题，本次调研立足于现状，从消费者视角出发，结合多种数据分析方法和建立模型等方法分析消费者行为，同时结合空间格局分析对民宿市场进行多维度的剖析，针对武汉市民宿建立系统而全面的评价体系，从新的角度提供建设性意见，并结合多个角度为武汉市民宿市场的发展提供策略。



## 第二章 研究思路与创新

## 二、研究思路与创新

### (一) 研究思路

本调研采取网络爬虫、地理空间格局计算和问卷调查相结合的方法，以武汉市为例，对民宿市场进行调研并进行一系列评价，最终提出结论与建议。

首先，以选择民宿意愿为区分，针对不同消费群体发放问卷调研，探索其消费行为特征及需求。本小组针对消费者对民宿的接受性，潜在消费者特征，消费者满意度等进行了描述性分析。并在此基础上，运用 logistic 二元选择模型和排序选择模型对消费者特征进行定量分析。

其次，通过爬取美团、携程、爱彼迎等网站上全部可抓取评论数据，使用中文分词、构建语义网络与 LDA 主题模型等算法和模型，实现对文本评论数据的情感倾向性判断以及所隐藏信息的挖掘与分析。

再者，基于地理视角分析武汉市目前现有民宿的空间格局，依据路网数据以及爬虫下来的民宿点位置，采用网络分析的方式，模拟民宿到周围设施点的最优路线来计算其可步性及可达性指数。

接着，针对文本挖掘及空间格局计算得到的结果，再从消费群体对民宿的需求出发，运用层次分析法构建出武汉市民宿适宜性评价的 AHP 模型。

最后，本文综合数据挖掘、空间格局和问卷调研分析结果，研究提出未来民宿市场发展方向及优化建议。

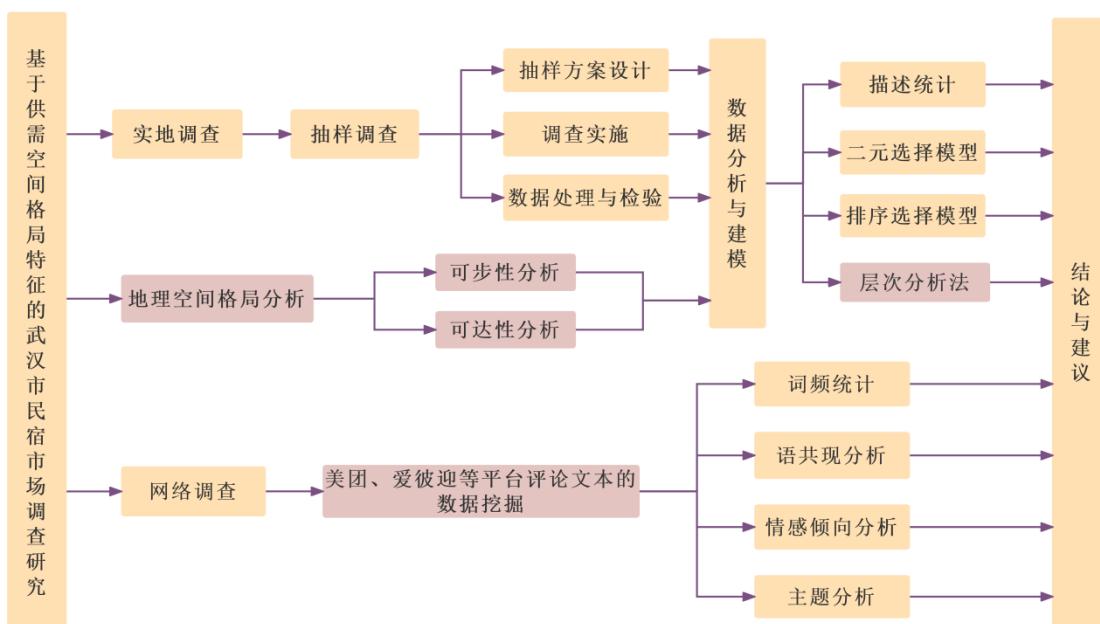


图 2 整体研究思路

### (二) 研究特色与创新

#### 1. 基于大数据的文本挖掘

通过爬取美团、携程、爱彼迎等网站上全部可抓取评论数据，使用中文分词、构建语义网络与 LDA 主题模型等算法和模型，实现对文本评论数据的情感倾向性判断以及所隐藏信息的挖掘与分析，以短时间内获得实地调查无法完成的数据信息，且可以提供实时的用户使用反馈信息。

#### 2. 地理视角下的民宿空间格局网络分析

本文从空间位置的角度出发，探求各个民宿的位置对其整体情况的影响。由于在现实中民宿到各个设施点需要步行或乘车，而其位置的好坏不能从“距离”这样一个简单的指标得来，因此针对现有路网，对民宿到其他位置的路线进行 OD 矩阵计算，得到最优路线并对其可步性及可达性进行数值化的呈现。对后续的 AHP 建模也有一定的帮助。

#### 3. 顾及消费者需求特征的民宿适宜性评价

运用 Logistic 模型对消费者的性别、年龄、职业、月消费对民宿的接受程度进行定量分析；借助排序选择模型，对在外住宿的消费者的特点进行分析。借助 AHP 模型，将民宿适宜性评价指标分解为 3 个维度 6 个评价指标，最终构建出武汉市民宿适宜性评价的 AHP 模型以对武汉市现有民宿进行整体性分析。

#### 4. 基于民宿与酒店对比分析的优化策略

本文通过对比民宿和酒店，深入的了解消费者对民宿的顾虑以及民宿与酒店的差异，并对这些因素进行分析，改变目前的市场格局，打破某些影响因素方面的局限性，使得调查的结果和所提出策略建议更具有可行性和针对性。



### 第三章 基于武汉市民宿 评论数据的挖掘与分析

## 三、基于武汉市民宿评论数据的挖掘与分析

### (一) 数据来源背景

国内经济的持续增长，人民生活质量水平的提升，为共享民宿的产生和发展都奠定了良好的基础。武汉市作为中国经济地理中心，素有“九省通衢”之称，是中国内陆最大的水陆空交通枢纽和长江中游航运中心，其高铁网辐射大半个中国。在这个背景下，武汉市的旅游人数逐年上升。此外，武汉是国内大学生数量最多的城市，全城共 84 所大学，大学生的住宿需求也为民宿热潮添砖加瓦。

### (二) 数据搜集

在线评论作为口碑传播的一种重要形式，对消费者的决策行为有着重要影响。实证分析结果表明：评论质量、评论时效性、评论数量和评论者资深度对消费者决策行为产生正向影响，同时，这 4 个变量均可通过评论可信度来影响消费者决策。本研究通过爱彼迎、美团、小猪民宿等 APP，运用 Web Scraper 与 R 语言编写数据采集逻辑、运行程序，收集了 2022 年 4 月 30 日至 2022 年 5 月 1 日武汉市的民宿信息以及评论数据等，探寻消费者心声，从而探究消费者对于武汉市民宿的总体印象、个人体验等，为未来武汉市民宿行业的发展提供指导与规划。另一方面，能够为武汉类似地区乃至全国的民宿行业提供发展规划，维护民俗行业健康良好的持续发展。

#### 1. 有效数据爬取

本次数据采集共搜集评论 78149 条(其中美团 27268 条，携程民宿 18634 条，爱彼迎 13979 条，小猪民宿 9655 条，途家民宿 8613 条)，有效率达 96.98%，此次分析中，将非结构化的文本数据转化为可检索的结构化数据，用表格形式储存，结构示意图如下：

表 9 从美团民宿平台爬取数据结构示意表

民宿名称	价格	评分	位置	用户 ID	发布时间	评论内容
森系浪漫投影房.....	338 ¥	5.0	武汉武昌区中山路 99 号	时光***27	2022 年 01 月	离地铁站挺近的，出门就可以买菜，很方便，菜市场菜也挺齐

### 第三章 基于武汉市民宿评论数据的挖掘与分析

						全的，厨具齐全。空间很大，公寓特别安静，总之棒棒哒！
【格调家情 迷影院房】.....	360¥	4.8	武汉汉宫 银座公寓	j**3	2021年07 月	很舒服的一次入住体验，位置很好找，去昙华林粮道街很方便，步行就可以到达，适合情侣入住。

表 10 从携程民宿平台爬取数据结构示意表

民宿名称	价格	评分	位置	用户 ID	发布时间	评论内容
【铂金涵舍】.....	¥346	4.9 分	观湖铂金公寓（江岸区云林街）	M***6	2022-03-14	住的地方交通很方便，周边吃的也挺多，晚上夜景很漂亮，房间性价比很高
【融舍·拾心】.....	¥398	5.0 分	观湖铂金酒店公寓	D***8	2021-05-05	很喜欢，装修风格简单舒适，湖景很漂亮 很热情主动 挺方便的，附近有地铁站 设施完备，没啥问题 这家我们住的很满意，力荐！

表 11 从小猪民宿平台爬取数据结构示意表

民宿名称	价格	评分	位置	用户名称	发布时间	评论内容
------	----	----	----	------	------	------

### 第三章 基于武汉市民宿评论数据的挖掘与分析

绝美沙湖 汉街全景 两居复 式.....	¥338	4.9 分	武昌区友谊 大道武汉友 谊国际大酒 店	沫7n	2018 年 11 月	房子很精致， 看得出房东布 置很用心，很 温馨，确实性 价比很高噢
《偶 遇》 .....	¥559.3	4.9 分	武昌区中山 路 424 号	60Emap	2019 年 03 月	武昌火车站 E 出口就是了， 老板人真好， 房子不错

表 12 从爱彼迎平台爬取数据结构示意表

民宿名称	价格	评分	位置	用户名 称	发布时间	评论内容
oneday 设计师桉 白.....	¥510	5.0 分	中国，湖北 省，武汉市， 江岸区，江岸 区汉飞青年城	媛钰	2022 年 3 月	太喜欢这个大窗 户了，简直是梦 中情床啊！拍照 很出片，生活用 品很齐全，还有 投影仪，很不 错！！

### 第三章 基于武汉市民宿评论数据的挖掘与分析

【往日】 (特价短租) 法式 复古 loft.....	¥340	4.8 分	中国, 湖北省, 武汉市, 武昌区, 武昌区友谊国际家居广场	毛毛	2022 年 1 月	房东姐姐很好.信息也回复的很快.那个吹风机有两个坏啦 还有一个是好的 但是吹出来的风很小不大....不过反馈以后也及时得到回应.这一点还是很好的.再就是跑步机着实给我弄了好一会哈哈哈哈 不过最后 OK 了还是很开心的 综艺来说体验感还不错 下次有机会还来~
--------------------------------------	------	-------	--------------------------------	----	------------	--

表 13 从途家民宿平台爬取数据结构示意表

民宿名称	价格	评分	位置	用户 ID	入住时间	评论内容
【星光】.....	¥161	5.0 分	汉阳区, 武汉汉阳越秀·星月湾畔	许***	2022-03-15	可以可以很不错入住的体验十分地舒适床很舒服
天空之城 一居室.....	¥182	4.9 分	武汉江汉区湖北武汉柚米国际社区(长青公园店)	用***5	2022-01-06	性价比很高，入住周边首选！

分析发现携程、美团、爱彼迎三个平台评论有效率处于前三位，分别截取三大平台详细评论界面示意图如下：



仄\*\*\*4

超棒

情侣出游 | 2021-11-19

途家网提供点评

好喜欢这种装修，非常有感觉！房间打扫的还是挺干净的，晚上关灯看投影感觉可幸福哈哈哈哈。建议房东可以准备一个USB的插线板，万一有人像我一样只带了线忘记带插头就方便很多！居然忘记拍照了，可恶

### 第三章 基于武汉市民宿评论数据的挖掘与分析

图 28 携程用户民宿评论结构示意图



图 29 美团用户民宿评论结构示意图

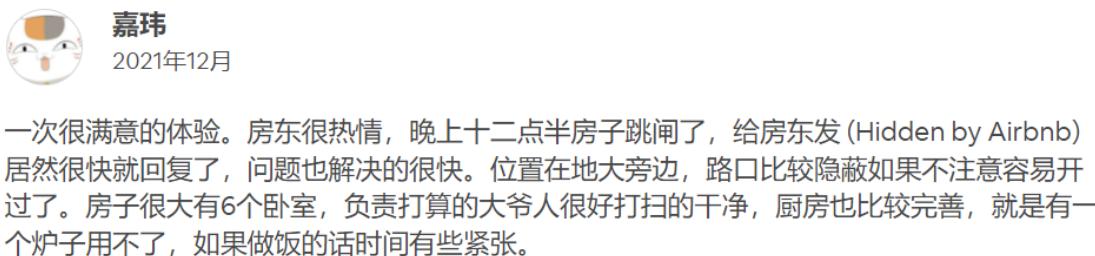


图 30 爱彼迎用户民宿评论结构示意图

## 2. 无效数据处理

初始得到的数据中，存在重复性评论、语义不明的评论、空评论等，这些评论无法传递有效信息，并且在后续文本分析中容易造成高稀疏性文本模型。所以，要进行数据的清洗，得到有效数据。首先，根据“百度停用词表”，“哈工大停用词表”，“四川大学机器学习实验室停用词表”等，使用 R 语言的 jiebaR 分词工具去除评论文本中停用词，并逐条解析评论的文本。最后，对 2436 条无效数据进行了识别与删除，具体删除类型如下：

(1) 系统自动评论，这些评论并非消费者的真实评论，无法显示消费者的真实意愿，不具有分析价值，应予以删除，例如：

喵喵喵 233	2022-01	系统自动评论:该房东在房客入住 1 天前单方面取消了订单。
浦***北	2021-09	系统自动评论:该房东在房客入住当天单方面取消了订单。
R***er	2021-12	系统自动评论:该房东在房客入住 35 天前单方面取消了订单。

表 14 无效数据示例 1

(2) 评论为空，主要分为三种，第一种为“无”，第二种为“null”，第三种为空格，这些数据对后续的数据分析没有价值，容易产生具有偏差的数据，应予以删除，例如：

表 15 无效数据示例 2

今天吃什么 5678	2019-05	无
想***甜	2021-11	Null
Ca***	2022-01	

(3) 评论内容为仅含数字、符号、字母等，这些被视为随意发表的评论，没有分析价值，因此删除，例如：

表 16 无效数据示例 3

夏日***	2018-07	111
机***！	2022-01	...
醉***山	2019-05	aaaaa
al***	2020-10	◆◆

(4) 评论为同一 ID 在同一天发出的多条相同评论，视作被雇佣的“水军”发言，评论内容不具有分析价值，需要被删除，例如：

表 17 无效数据示例 4

子***安	2021-08	好好好
子***安	2021-08	好好好
子***安	2021-08	好好好

### (三) 数据的处理

#### 1. 中文分词

中文分词指的是将一个汉字序列切分成一个一个单独的词。分词就是将连续的字序列按照一定的规范重新组合成词序列的过程。

### 第三章 基于武汉市民宿评论数据的挖掘与分析

本报告主要利用 R 软件中的 jieba 包来进行分词处理，属于概率语言模型分词。概率语言模型分词的任务是：在全切分所得的所有结果中求某个切分方案 S，使得  $P(S)$  最大。

#### 2. 分词的原理

首先，基于前缀词典实现高效的词图扫描，生成句子中汉字所有可能成词情况所构成的有向无环图（DAG）。

之后，动态规划查找最大概率路径，找出基于词频的最大切分组合。

最后，对于未登录词，采用了基于汉字成词能力的 HMM 模型，使用了 Viterbi 算法，得到一个概率最大的 BEMS 序列，按照 B 开头，E 结尾的方式，对待分词的句子重新组合，就得到了分词结果。

#### 3. 词频统计

为了较全面地反映消费者对民宿居住的总体感受，本节选取中高频关键词作为分析对象，不仅可以避免因词量太少而导致的结果模糊，又能保证核心关键词的完整。部分词频统计分析如下表格：

表 18 部分词频统计分析表

序号	词名	词频	词性	序号	词名	词频	词性
1	房间	34414	n	16	喜欢	6741	v
2	房东	33394	n	17	装修	6030	v/n
3	位置	22515	n	18	地铁站	6007	n
4	干净	21220	a	19	很多	5915	a
5	方便	19729	a	20	很棒	5686	a
6	不错	14742	a	21	交通	5675	n
7	房子	12117	n	22	热情	5480	a
8	入住	11063	v	23	环境	5285	n

<b>9</b>	舒服	9091	a	<b>24</b>	推荐	5090	v
<b>10</b>	体验	8988	n	<b>25</b>	齐全	4849	a
<b>11</b>	民宿	8036	n	<b>26</b>	性价比	4666	n
<b>12</b>	姐姐	7841	n	<b>27</b>	整洁	4540	a
<b>13</b>	照片	7807	n	<b>28</b>	好看	4368	a
<b>14</b>	楼下	7142	n	<b>29</b>	温馨	4368	a
<b>15</b>	武汉	6970	ns	<b>30</b>	设施	4333	n

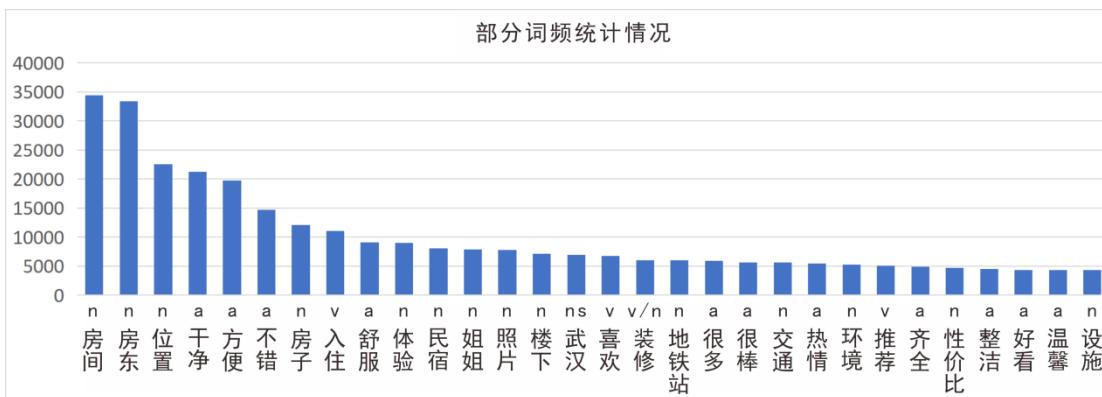


图 31 部分词频统计情况图

将分词根据词频绘制词云图，其是对文本内容的高度浓缩和精练处理，能更直观地反映出具体文本的内容。分析后得到的词云图如下所示。



图 32 部分词云图

#### 4. 词共现分析

为了理清关键词之间的关系，利用 gephi 软件，对数据中前 20 个高频关键词进行共现分析，生成  $20 \times 20$  的共现矩阵，生成词共现网络图。词共现网络图的各个关键词之间的连线越粗，说明相应关键词联系越紧密；关键词节点越大，说明它与其他关键词的连接越多。下图为关键词共现网络图：

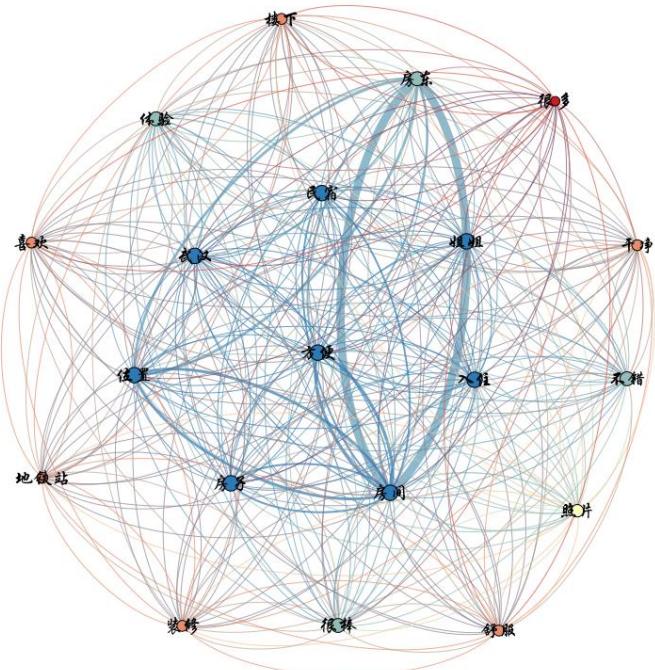


图 33 关键词共现网络图

#### 5. 基于情感词典的情感倾向性分析

情感分析就是对文本信息进行情感倾向挖掘，情感倾向分析目前主要使用的方法就是利用已有的情感词典，运用 python 对每个评论出现的词语与情感词典中的词汇进行匹配，并汇总情感词进行评分，最后得到评论语料的情感倾向。

在这里，主要采用了基于 BosonNLP 情感词典和采用的是中国知网发布的“情感分析用词语集”。

##### (1) 基于 BosonNLP 情感词典

首先对评论数据进行分句、分词，然后与 BosonNLP 词典进行逐个匹配，并记录匹配到的情感词分值。最后，计算每条评论的分值总和，如果分值大于 0，表示情感倾向为积极的；如果小于 0，则表示情感倾向为消极的，如果等于 0，表示为中立。



图 34 基于 BosonNLP 词典的情感分析流程图

##### (2) 基于知网情感词典

首先，对评论分句，分词，并去除停用词；然后，判断评论中的情感词数目，若含有积极词，则积极词数目加 1；若含有消极词，则消极词数目加 1。但是还需要判断该情感词前面是否存在程度副词，如果存在，则需要根据程度副词的种类赋予不同的权重，来乘以情感词数。如果句尾存在？！等符号，则情感词数目要增加一定值。最后，根据如下的分析规则得到每条评论的情感倾向。

$$P(d_i) = \begin{cases} 1 & (sumPos(d_i) > sumNeg(d_i)) \\ -1 & (sumPos(d_i) < sumNeg(d_i)) \\ 0 & (sumPos(d_i) = sumNeg(d_i)) \end{cases} \quad \text{公式 (8)}$$

上式中  $P(d_i)$  表示文本的情感极性值， $sumPos(d_i)$  表示文本中包含的正面情

感词个数,  $sumNeg(d_i)$ 表示文本中包含的负面情感词个数。

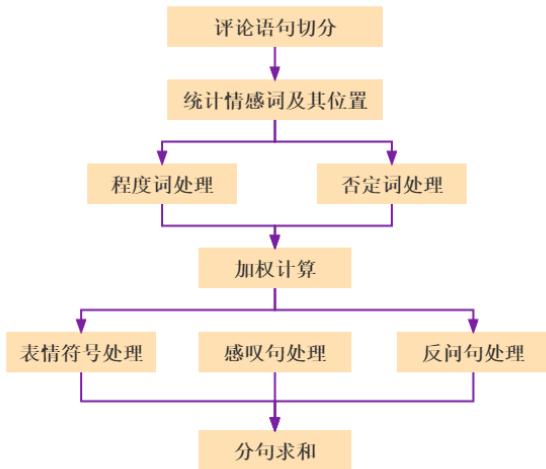


图35 基于知网词典的情感分析流程图

#### (3) 结果显示

基于两种词典的武汉市民宿评论数据的情感倾向分析结果如下图所示, 其中积极的情绪都占据了很大的比例, 分别为 91.7% 和 86.1%。

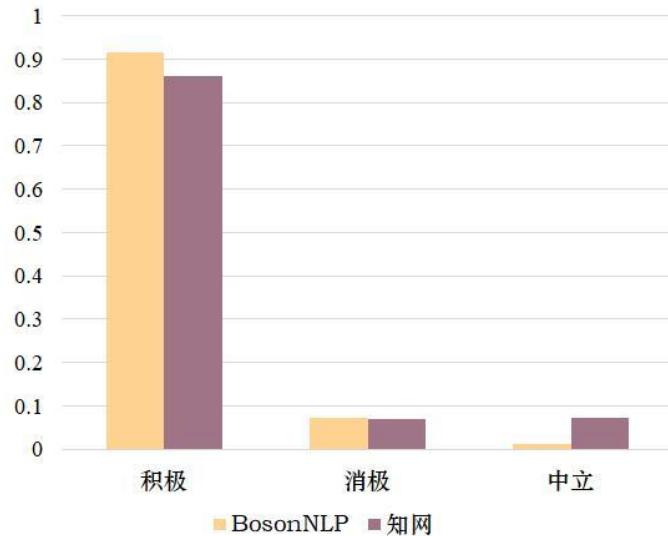


图36 情感倾向分析结果图

## 6. 基于 LDA 模型的主题分析

LDA 主题模型是评论文本每条评论的主题以概率分布的形式展现出来。其通常由文档 ( $d$ )、主题 ( $z$ ) 和词 ( $w$ ) 三层结构组成, 能够有效地对文本进行建模, 挖掘数据集中的潜在主题, 进而分析数据集的集中关注点及其相关特征词。

### 第三章 基于武汉市民宿评论数据的挖掘与分析

模型首先假设每条评论文本都是包含多种主题，且主题都按照一定的多项分布的混合比例随机组合而成：

$$Z | \theta = \text{Multinomial}(\theta) \quad \text{公式 (9)}$$

与此同时，每个主题都包含了词汇表中的多个词语，且都按照一定的多项分布的比例混合而成：

$$W | Z, \varphi = \text{Multinomial}(\varphi) \quad \text{公式 (10)}$$

在评论  $D_j$  条件下生成词  $w_i$  的概率表示为：

$$P(W_j | D_j) = \sum(W_i | Z = S) \times P(Z = S | D_j) \quad \text{公式 (11)}$$

式中， $P(W_i | Z = S)$  表示词  $W_i$  属于第  $S$  个主题的概率， $P(Z = S | D_j)$  表示第  $S$  个主题在评论  $D_j$  中的概率。

采用 Metropolis-Hastings 算法中的吉布斯抽样对参数  $q$ 、 $\varphi$  进行近似估计。可得出词  $W_i$  在主题  $Z_s$  中、主题  $Z_s$  在评论  $D_j$  中的参数估计，分别记为：

$$\varphi_{s,i} = (N_{s,i} + \beta_i) / (\sum N_{s,i} + \beta_i) \quad \text{公式 (12)}$$

$$\theta_{j,s} = (N_{j,s} + \alpha_s) / (\sum N_{j,s} + \alpha_s) \quad \text{公式 (13)}$$

对参数的近似估计使用 MCMC 算法中的一个特例 Gibbs 抽样。本研究将先验函数  $\alpha$  和  $\beta$  设置为经验值： $\alpha=50/K$ ,  $\beta=0.1$ ，主题个数  $K$  采用统计语言模型中常用的评价标准困惑度来选取，即  $K=50$ 。

在这里，具体给出其中聚类 4 个主要主题词云图，如下所示：



图37 主题1词云图



图38 主题2词云图



图39 主题3词云图



图40 主题4词云图

根据武汉市民宿评论的 4 个潜在主题的特征词提取，主题 1 中的高频特征词，即位置、交通、地铁站、江汉路等与所在地理位置相关的主题；主题 2 中的高频特征词，即老板、热情、及时、耐心等与房东及服务相关的主题；主题 3 中的高频特征词，即环境、投影、浴缸、卫生等与民宿环境及设施相关的主题；主题 4 中的高频特征词，即不错、棒、舒服、好看等与消费者的体验相关的主题。



# 第四章 民宿空间 格局特征分析

### 四、民宿空间格局特征分析

#### (一) 民宿基本格局

为了了解现有民宿基本格局，对民宿分布与武汉市现有人口布局进行分析。由于民宿数据大多数集中在主城区，且考虑到现实因素，故此步骤利用武汉市主城区进行分析。

首先，人口数据来源为第七次人口普查数据。本小组使用 ArcGIS 软件运用空间连接的方式将人口数据连接到每一个街区的面数据中，并对街区人口数量以及人口密度进行计算及可视化展示。由低到高采用自然分级法分为五级，由浅到深进行展示。加上民宿点后的可视化成果如下。

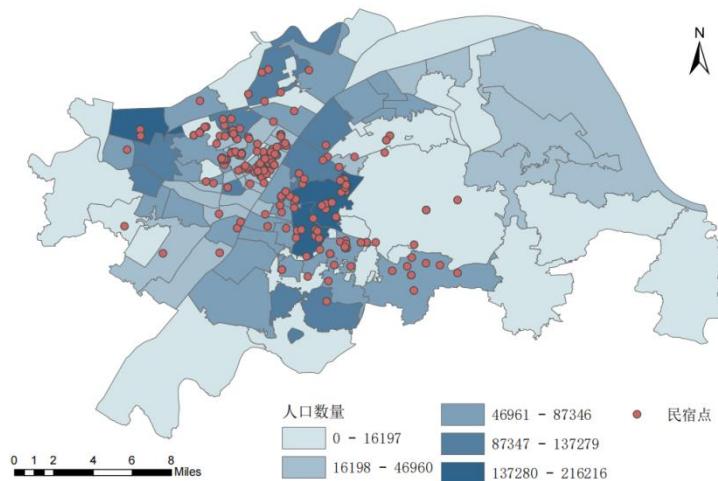


图 41 武汉市主城区人口-民宿分布图

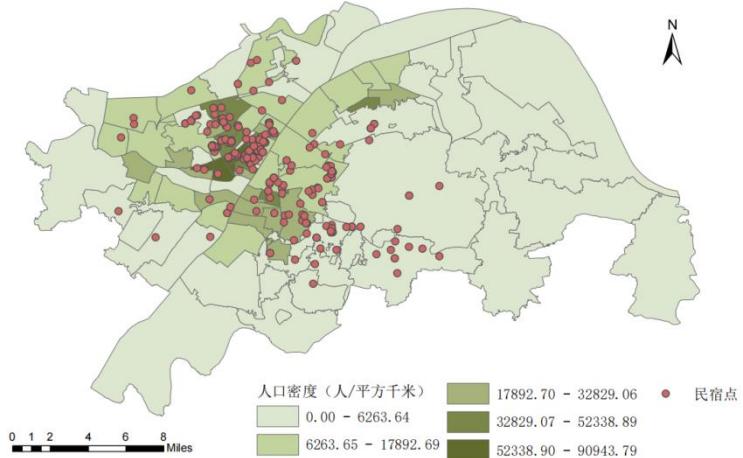


图 42 武汉市主城区人口密度-民宿分布图

随后对其进行数据分析。首先，运用空间连接计算出每个街道含有的民宿个

数，进而进行数据分析工作，旨在探求人口密度以及民宿个数之间的相关关系。

首先第一种方法就是对 75 个街区进行相关性检验，得到的结果为负相关，但其显著性不足，因此可能是极值和数据样本数不大的影响。该结果不适用于对其相关性的研究。

表19 人口密度-民宿个数相关性

		密度	Count_
密度	皮尔逊相关性	1	-0.141
	Sig. (双尾)		0.227
Count_	个案数	75	75
	皮尔逊相关性	-0.141	1
Count_	Sig. (双尾)	0.227	
	个案数	75	75

本实验采用的第二种方法就是查看在人口密度分为五类的街区中，有多少街区是有民宿点的，这样可以减少数据量过低以及特殊值引起的误差。对其进行数据分析，结果如下：

表20 街区所含民宿比例表

分级结果	有民宿街区个数	无民宿街区个数	总计街区数	有民宿所占比例
1	7	13	20	35.00%
2	14	8	22	63.64%
3	9	6	15	60.00%
4	7	2	9	77.78%
5	6	3	9	66.67%

由上表可知，在分级结果为“1”的街区中，其中有民宿的街区仅占 35.00%，与其他级别的街区有很大的不同，而其余四个级别的街区所含民宿的比例相差不大。由此可知，在人口密度方面，现有民宿的格局更不愿意选择人口过于稀疏的街区，这一点也和可视化展示的结果类似。

## (二) 民宿可步性

民宿作为一个暂时性的住宿场所，其交通通达性是非常重要的一个性质。然而，对于其评价系统却不能采用简单的两点间距离大小来界定。因此采用地理信息科学中的“可步性”进行进一步的分析。

### 1. 基础步行指数计算原理

步行指数是目前唯一的国际性量化测度步行性的方法，该方法于 2007 年由美国研究者提出，本实验将利用步行指数理论测度武汉市的民宿可步性。

设施的初始权重将随着其与出发点距离的增加而有规律地衰减。本小组采用三次曲线，根据 walkscore.com 网站的距离衰减规律标准，按标准步行速度 80m/min，得到以下距离衰减规律，如下表所示：

表21 距离衰减规律表

时间	范围	距离衰减规律	备注
-5min	400m	不发生距离衰减， $y=1$	
-20min	1600m	$y = -153.6558x^3 + 419.4604x^2 - 395.9706x + 201.1086$	快速衰减，当位于1600m时，衰减到12%；
-30min	2400m	$y = -92.8x^3 + 566.6x^2 - 1153.1x + 786.6$	衰减缓慢，当大于2400m时，衰减率大于1； 2400m以外对步行指数无影响

(注：式中  $y$  为衰减率，单位为%， $x$  为小区点到公共设施的距离，单位为 km)

### 2. 路网分析

首先通过对道路数据进行拓扑关系的构建与检查，保证网络数据构建正确。并设置网络属性，用于控制导航，用作网络阻抗的成本属性，或禁止双向穿越或单向穿越（如单行线）的约束属性。确保“可步性”的科学性，以求模拟最接近的步行路线。

OD 成本矩阵是计算多个起始点与多个目的点之间连通成本（时间、费用、路长）的表格。成本矩阵解算器可从网络中查找并测量从多个来源到多个目的地

成本最低的路径。将起点设置为民宿，终点设置为公交地铁站，计算其 OD 成本矩阵。将计算得到的公共设施点到小区的距离，带入相应的衰减函数，计算公共设施的距离衰减率，将衰减率与公共设施的原始权重相乘，最后得到每个设施点经衰减后的权重值，即可得到小区在 400~1600m 范围内可达的邮局设施的权重。同理可以计算其 400m 内和 1600~2400m 距离范围内的衰减后的权重。

小区到其他商圈及景点的距离衰减权重同理。



图 43 民宿-地铁公交站 OD 矩阵连接图

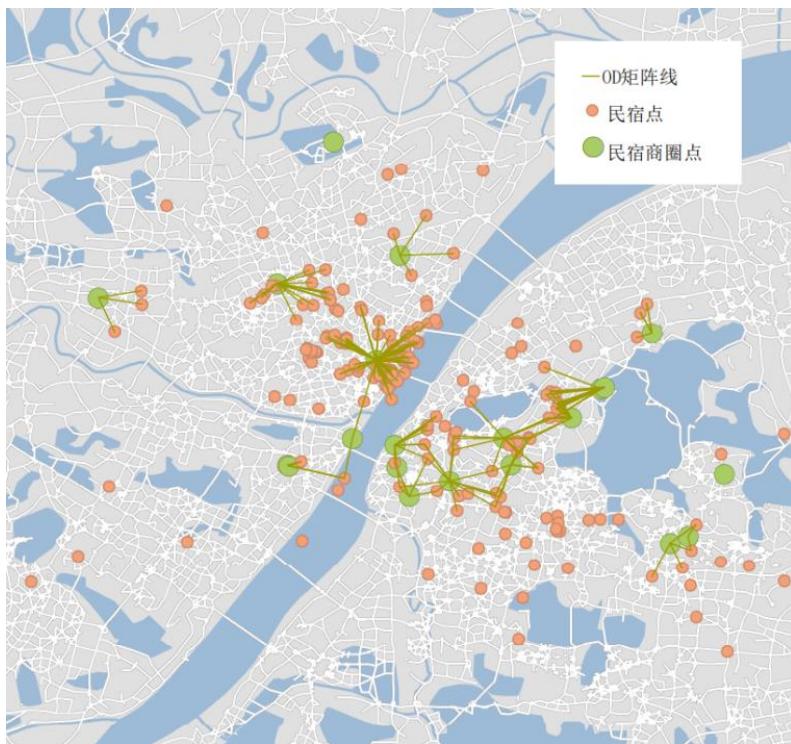


图 44 民宿-商圈景点 OD 矩阵连接图

### 3. 可视化展示

最后通过数据连接将计算得到的各设施经距离衰减后的权重结果添加到民宿数据的属性表里。最终得到的数据为每个民宿都有一列属性为“公交地铁权重”以及“商圈景点权重”，便于后续的计算及表示。

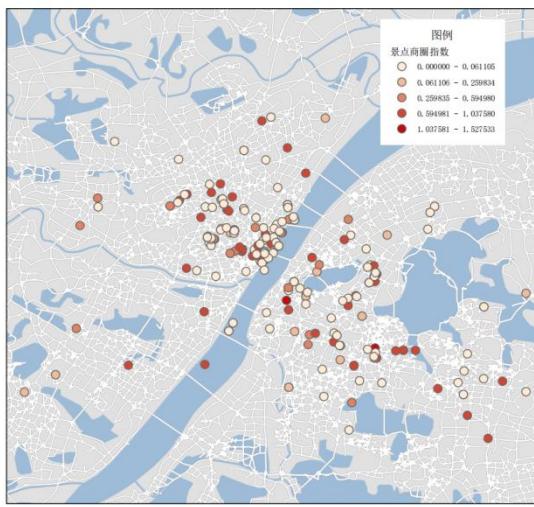


图 45 景点商圈分级图

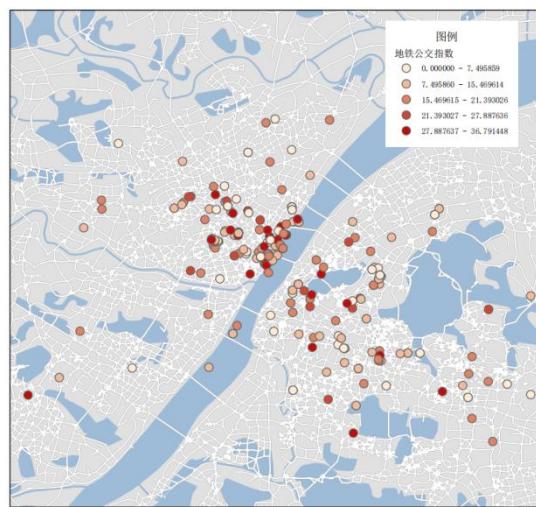


图 46 地铁公交分级图

再通过插值得到的权重分布图如下所示。

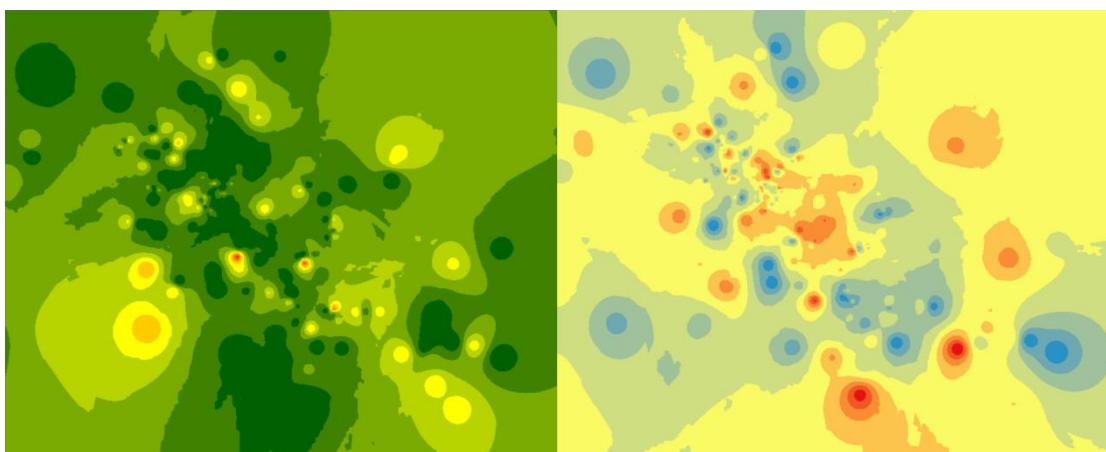


图 47 商圈景点插值图

图 48 公交地铁插值图

### (三) 民宿可达性

#### 1. 整体实验设计

和前面的商圈景点以及地铁公交不同的是，当需要应急去医院时往往不是步行前往，因此应计算车行距离，采用交通可达性进行医院指数的计算。

#### 2. 路网分析

基本操作与可达性一致。不同的地方就在于计算步行指数要采用步行距离衰减的公式，而交通可达性需要计算车行时间，公式为  $\text{Drivetime} = \text{shape\_length}/1000$ 。这里表示车行速度  $60\text{km/h}$  下的车行时间。将起点设置为民宿，终点设置为医院，建立 OD 成本矩阵。

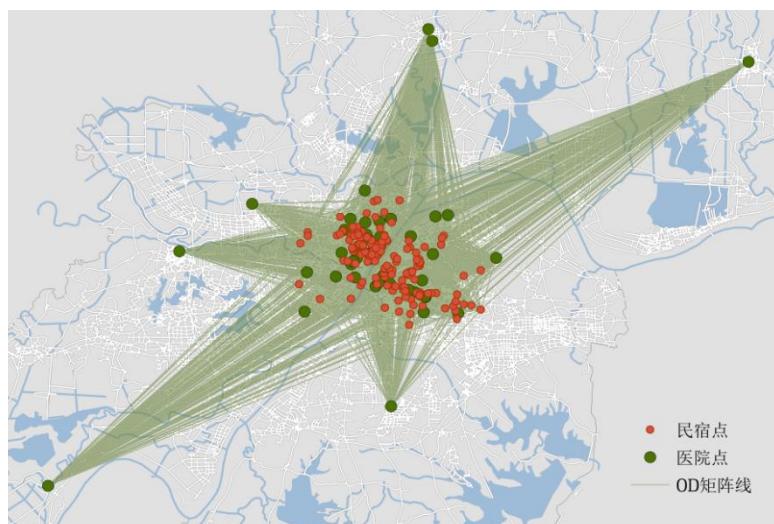


图 49 民宿-医院 OD 矩阵连接图

分类汇总对车行时间进行求和，最终生成可达性计算表。利用公式，可达性 $=[\text{Sum\_Total}] / ([\text{Cnt\_Origin}] - 1)$ ，计算出可达性指数。

### 3. 可视化展示

最后再将可达性表连接到起始点上，就得到了“医院权重”，将其进行自然分级法进行可视化表达。

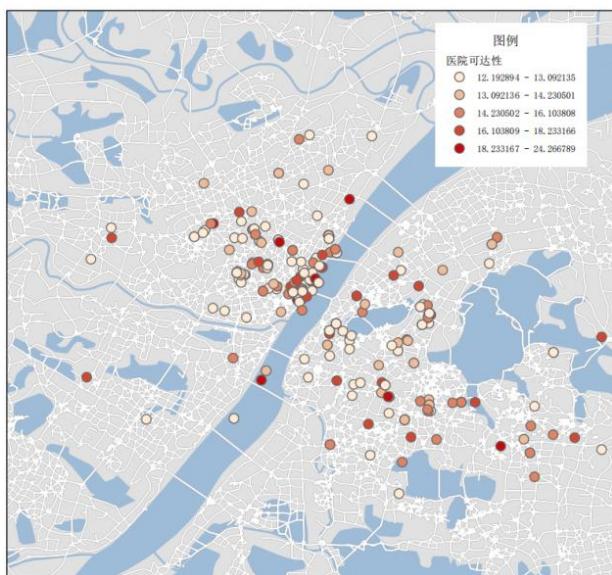


图 50 医院权重分级图

再通过插值得到的权重分布图如下所示。



图 51 医院权重插值图



## 第五章 问卷调查 方案和实施策划

### 五、问卷调查方案和实施策划

#### (一) 调查方案设计

##### 1. 调查目的

通过问卷调查，来了解大众对民宿行业的态度以及民宿行业内现存的问题，搜集调查影响消费者选择民宿的影响因素，深入研究民宿行业的消费者特征及其行为特点。搜集不愿意选择民宿居住的消费者不选择的原因以及其相应的基本信息，从而对潜在用户进行挖掘。搜集消费者对民宿及其配套设施的满意度，从而为民宿行业未来发展方向提供一定的参考，从而建立对民宿行业评分系统。搜集消费者对民宿行业未来发展的期望，根据调查的结果为民宿行业的创业者提供一定的实质性的建议，以便于业内民宿对于自身各方面因素进行改善，实现民宿经济的可持续发展。

##### 2. 调查内容

依据调查目的，本调查主要围绕全国消费者对于民宿行业的态度及其市场现状进行研究，进而为民宿行业的发展提供有效的建议。

调查内容的结构及主题主要体现在问卷结构的设计上，整体分为四部分：

- (1) 被调查者基本情况
- (2) 民宿行业的情况及消费者满意度
- (3) 影响消费行为的重要性因素
- (4) 消费者对于民宿的建议

问卷中分别采用了单选题、多选题、填空题和量表题四种题型。为了更好对所需调查项目进行考量和保证答题者答题准确性，将相应调查项目进行了高度概括，最终整理得出的调查问卷见附录。

##### 3. 调查对象和方法

我们希望通过调查了解现有民宿消费者的满意度，并发掘潜在用户，进而探究民宿行业的发展前景，因此我们需要调查的不仅是住过民宿的用户，还需要调查潜在用户群体对民宿的了解及看法。因此本调查以 34 个省份中住过民宿以及未住过民宿的互联网网民作为调查对象，分别有针对性地设计调查问卷，应用问

卷网以及问卷星制作网络问卷，通过在互联网平台上分享链接的方式进行问卷发放。共计 1165 份有效问卷数据。

#### 4. 抽样设计

##### (1) 样本框的确定

本项调查以互联网中住过和未住过民宿的消费者作为调查对象。根据可实施条件，运用随机抽样的方法，通过控制调查样本的年龄、性别结构，选取了不同年龄阶段的消费者作为调查对象。其中 25 岁以下，25-30 岁，31-40 岁，40 岁以上分别作为一个独立的年龄阶层加以区分。性别比例则控制在 1: 1 左右。

##### (2) 样本量及样本分配量的确定

本项调查的样本容量计划为 1200 人，鉴于各年龄阶层对住宿的需求程度不同，为降低抽样难度，对事先确定的各年龄阶层配额抽样。对于 30 岁以下的消费者，其日常出差、旅游等活动较多，且勇于尝试新事物，故计划的样本容量较大，约为 60%，即 720 人。30 岁以上的消费者由于家庭、经济实力等原因，更愿意保持原有状态，因此计划的样本容量占总体样本容量的数额相对较少，约为 40%，即 480 人。

#### 5. 主要数据分析方法

##### (1) 列联分析

列联分析是用来判断样本之间的变化趋势与相关程度的统计方法。进行列联分析时，使用卡方检验测量两道题之间的相关性，原假设为：两道不同题目之间是无关的，备择假设为：两道不同题目之间是相关的。卡方检验的统计量为  $\chi^2$ ，

$$\chi^2 = \sum_j \sum_i \frac{(A-T)^2}{T} \quad \text{公式 (1)}$$

其中 A 为实际频数，T 为理论频数，i、j 分别代表所在行、列的编号。 $\chi^2$  值越大，说明实际频数与理论频数的差别越明显。自由度为  $v = (R - 1)(C - 1)$ ，R 和 C 分别代表行数和列数，显著性水平为 0.05。当显著性水平小于 0.05 时，两道题目之间是相关的。

##### (2) 独立样本 t 检验

独立样本 t 检验是用来判断两个独立样本之间的均值差异情况的统计方法。进行 t 检验时，原假设为：两个独立样本的均值相等，备择假设为：两个独立样本的均值不相等。在大样本、方差未知的条件下，t 检验的统计量为

$$t = \frac{[(\bar{x}_1 - \bar{x}_2) - (u_1 - u_2)]}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad \text{公式 (2)}$$

这是一个双侧检验，在 0.05 的显著性水平下，查 t 检验表得临界值，确定接受域与拒绝域，看 t 统计量是否落在接受域内，如果落在接受域内，则接受原假设，否则拒绝原假设。也可以比较 p 值与给定显著性水平的大小，p 值越小，越能拒绝原假设。

### (3) Logistics 回归分析

Logistic 回归分析广泛地应用于因变量为二分类别变量的回归模型，此二分类别变量的编码不是 0 就是 1。Logistic 回归分析的假定是观察值样本在因变量上的概率呈 S 型分布。模型的显著性检验包括整体模型检验和个别参数检验两个部分。在参数估计方面，Logistic 回归分析是通过极大似然估计使因变量观察次数的概率极大化，进而得到自变量参数的最佳估计值，同时 Logistic 回归模型通过概似值进行整体检验。定量分析性别、年龄、职业、消费水平等因素对是否住过民宿的显著影响因素，进而确定住过民宿的消费者所具有的显著特征。

### (4) 描述性统计

在对搜集到的数据出入计算机进行初步处理后，我们进行了描述性统计分析。我们通过频数分析方法计算各描述统计量指标，绘制了频率分布条形图和饼图，并对调查对象的人群特征分布、满意度、改善意见等情况有了一个整体的印象。

### (5) 结构方程模型

结构方程模型是一种建立、估计和检验因果关系模型的方法。模型中既包含有可观测的显在变量，也可能包含无法直接观测的潜在变量。在已有的因果理论基础上，用与之相应的线性方程系统表示该因果理论的一种统计分析技术。目的在于探索事物间因果关系并将这种关系用因果模型、路径图等来表达。一般来说，结构方程模型由测量和潜在变量两部分组成：测量部分求出观察指标与潜在变量之间的关系；潜在变量部分求出潜在变量与潜在变量之间相关系数。因此，结构方程模型分为测量模型与潜在结构模型两部分。

### (6) AHP 层次分析法

层次分析法是指将一个复杂的多目标决策问题作为一个系统，将目标分解为多个目标或准则，进而分解为多指标（或准则、约束）的若干层次，通过定性指标模糊量化方法算出层次单排序（权数）和总排序，以作为目标（多指标）、多方案优化决策的系统方法。

层次分析法是将决策问题按总目标、各层子目标、评价准则直至具体的备投方案的顺序分解为不同的层次结构，然后用求解判断矩阵特征向量的办法，求得每一层次的各元素对上一层次某元素的优先权重，最后再加权和的方法递阶归并各备择方案对总目标的最终权重，此最终权重最大者即为最优方案。

### （二）调查组织分工

#### 1. 调查组织分工

在熟悉了调查项目性质、目的、要求等方面内容之后，我们首先系统地学习调查过程中应掌握的调查知识以及访问技巧，提升调查人员素质。受到疫情的影响以及安全问题的考量，我们的问卷采用了线上调查的方式，其显著缺点为回收效率低，有效性较差等问题。因此，我们邀请了各自的亲人朋友协助问卷的发放工作作为问卷的预调查。第二阶段改善了问题的设计并继续扩大样本容量。第三阶段则是调整现有问卷的年龄结构，进行有针对性的问卷收集，平衡年龄结构，使其趋于合理化。

#### 2. 调查实施进度

本次调查的周期为 2022 年 1 月 21 日至 2022 年 3 月 25 日，历时两个月。

第一阶段为调研前期，时间为 1 月 21 日至 2 月 3 日，主要任务为确定调查题目，拟定调查目的，编写调查大纲等工作。

第二阶段为调查执行期，时间为 2 月 4 日至 3 月 4 日，主要任务是编写调查问卷，查阅参考文献，进行数据资料的收集和整理，对数据进行统计分析和描述性分析。

第三阶段为报告撰写期，时间为 3 月 4 日至 3 月 25 日，主要任务是对统计分析结果和描述性分析结果进行图表及文字说明，并对结论进行汇总。

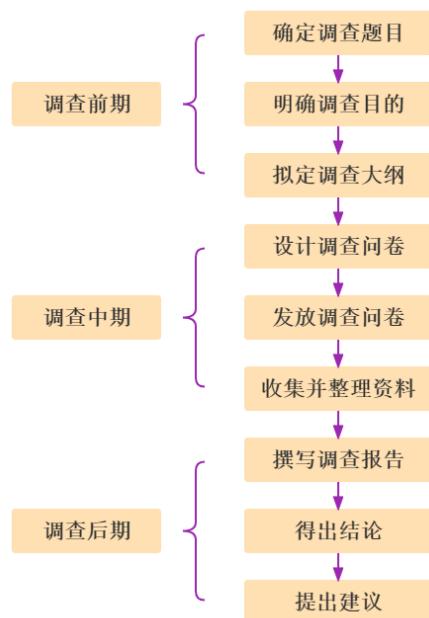


图 3 调查实施过程



## 第六章 质量控制

### 六、质量控制

#### (一) 预调查数据检验及问卷修改

在完成问卷设计之后，我们通过网络完成了问卷的试发放，共有效回收问卷 50 份。根据回收问卷对数据进行分析以及信度、效度检验，从而检查设计题目的鉴别度，测量问卷结果的可靠性以及问卷的有效性。同时针对问卷的完成情况来对相应的问题和问卷结构进行调整与改进。

##### 1. 数据检验

###### (1) 区分度分析

在整理回收的问卷后，首先对量表进行各个题项的区分度分析。主要目的在于检验编制的量表或检验个别题项的可靠程度。本次调查采用高低组平均数差异检验法进行检验。低区分度的题目不能有效的鉴别被调查者。

首先对住过民宿的被调查者填写的满意度量表进行项目分析，将每个被调查者所填量表的 11 项各得分相加求得总分，将每个样本的总得分从高到低排序，将总得分前 27% 的视为高分组，后 27% 的视为低分组。然后对高低两组的各个条目得分进行独立样本 t 检验以剔除差异不显著的项目。得到前 27% 的得分为 52 分以上，后 27% 的得分为 26 分以下。运用样本差异性 t 检验法对两组的 11 个题项分别进行检验，结果中的 t 值全部显著( $p$  值小于 0.05)，表明量表中的 11 个题项均具有鉴别度，即量表中所有题都能鉴别不同受试者的反应程度。综上所述，调查问卷通过了检验，可以用于正式调查。

###### (2) 信度检验

问卷设计质量的信度检验指的是对问卷测量结果准确性的分析，即对设计的问卷在多次重复使用下得到的数据结果的精确性和可靠性检验。通过信度分析，可以反映被测特征的真实程度。我们采用 Cronbach 信度测量问卷项目的内在一致性系数，信度系数的取值范围为[0,1]。Cronbach 信度测量公式如下：

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum s_i^2}{s_t^2}\right) \quad \text{公式 (3)}$$

其中， $k$  为量表的总题数， $s_i^2$  为第  $i$  题得分的题内方差， $s_t^2$  为全部题项总得分的方差。从公式中可以看出，Cronbach 系数评价的是量表中各调查项目得分的内部一致性。信度系数越大，说明测量的可信度越大。

一份设计较为良好的问卷信度系数应在 0.80 以上, 0.70 至 0.80 之间算是可以接受的范围, 若分量表的内部一致性系数在 0.60 以下或者总量表的信度系数在 0.80 以下, 应考虑重新修订量表或增删题项。我们对问卷中量表题的四个维度进行信度检验, Cronbach 系数结果如下:

表 1 Cronbach 系数表 1

层面	Cronbach 系数	项数	信度评价
价格	0.943	2	很好
地理位置	0.954	3	很好
交通通达程度	0.972	2	很好
设施及服务	0.928	4	很好

### (3) 效度检验

效度即有效性, 它是指所测量到的结果反映所想要考察内容的程度, 测量结果与要考察的内容越吻合, 则效度越高; 反之, 则效度越低。效度分为三种类型: 内容效度、准则效度和结构效度。其中结构效度是指测量结果体现出来的某种结构与测值之间的对应程度。

本项调查运用 SPSS 统计分析软件, 通过 KMO 和 Bartlett 球形检验因子分析适合情况。KMO 取值在 0 和 1 之间。当所有变量间的简单相关系数平方和远远大于偏相关系数平方和时, KMO 值接近 1, 意味着变量间的相关性越强, 原有变量越适合作因子分析, 并且因子分析的结果越好。Bartlett 球度检验值越大则表明变量之间的独立性越高, 各个条目存在共享公因子的可能性越低, 越适合做因子分析。计算后的问卷 KMO 系数为 0.897, 表明适合进行因子分析, P 值为 0.000, 因此问卷的结构设计较好。

上述检验结果表明, 预调查所采用的问卷能够达到本次调查目的。

## 2. 问卷修改

针对预调查获得的 50 份问卷, 在完成对数据的基本的检测后我们对问卷回答情况做了初步分析与研究, 并结合实际情况对问卷做出以下调整:

(1) 在年龄结构设置方面, 由于观察到各年龄阶层对旅游的需求程度不同, 我们对事先确定的各年龄阶段分层做出了调整。40 岁以下的消费者, 因其娱乐休闲意识较为强烈, 且具有的选择多样性较大, 各年龄分层有明显特征; 而 40

岁以上的消费者由于家庭，身体机能等原因，有较为趋同的选择倾向。故我们将年龄选项调整为：“25岁以下，25-30岁，30-40岁，40岁以上”。

(2) 在对“选择民宿各因素重要程度”的调查中，我们了解到原本设立的比重题不太利于问卷填写者们的作答，不仅作答难度加大，也不能起到直观评价填写者内心评估结果的作用。于是我们将其修改为基本的矩阵量表题型。

(3) 在对“选择民宿各因素重要程度”的问题中，我们发现民宿的治安问题、用户的私密性问题、及处理应急情况的基础设施方面也是很多人关心的因素，于是我们在此题中添加了“安全应急性”这一选项。且在满意度调查题中也增添了相应题项。

综上，此预调查结果对问卷修改有指导作用，针对调研过程中出现的各类问题，对问卷进行了修改与完善，使问卷更加科学化和合理化。对调研开展赋予了更深层次的意义。

## (二) 正式调查数据处理与检验

### 1. 数据处理

#### (1) 数据的接收与导出

由于使用电子问卷调研的形式，问卷录入的过程能有效避免被调查者漏答及回答内容形式不正确的问题。调查过程中，我们将被调查者配合度较低、填写速度较快、较随意的问卷、以及中途放弃填写的问卷标注为无效问卷。同时我们在问卷中插入了检测专注度的题目，在筛除被调查者没有通过专注度检验的样本后，回收的有效样本整体质量较高。没有缺失数据、问卷不完整的情况。

#### (2) 数据编码及缺省值处理

本问卷的导出采用数据编码，将每份问卷中每个问题对应的编码录入计算机中，使用 SPSS 及 Excel 实现这一过程，随后进行数据处理分析。问卷中的缺省值一般由两种原因造成：第一种是被访者应该填写而未填写；第二种是由于跳转而造成的数据缺失。本问卷通过问卷星网络平台及问卷网进行发放，每道客观题均为必答题，不存在第一种情况；问卷的跳转逻辑经调查员设计好后，经过多次预调查试验，已经确认准确无误。核查后可知本问卷不存在缺省值。

### 2. 数据检验

正式调查共回收有效问卷 1165 份，我们共采用以下几种分析方法对问卷数

据进行分析，包括区分度分析、信度分析、效度分析、游程检验等，对问卷数据的内在一致性，随机性进行检验。

### (1) 区分度分析

对每个被调查者所填满意度量表的 14 项各得分相加求得总分，得到前 27% 的得分为 51 分以上，后 27% 的得分为 29 分以下。运用样本差异性 t 检验法对两组的 14 个题项分别进行检验，结果中的 t 值全部显著，量表中的 14 个题项均具有鉴别度，即量表中所有题都能鉴别不同受试者的反应程度。

### (2) 信度分析

对正式问卷中量表题的五个维度进行信度检验，我们将量表题中 14 个题项按调查内容分成五组，分别考察消费者对民宿价格、地理位置、交通通达程度、设施及服务、安全及应急性方面的满意度。采用 Cronbach 系数作为检验标准，分析结果如下：

表 2 Cronbach 系数表 2

层面	Cronbach 系数	项数	信度评价
价格	0.931	2	很好
地理位置	0.942	3	很好
交通通达程度	0.966	2	很好
设施及服务	0.913	4	很好
安全及应急性	0.911	3	很好

根据系数表可知问卷分类合理，量表内在一致性高。

### (3) 效度分析

仿照预调查数据的检验方法，我们用各项目分数与总和分数的相关系数做量表题的内容效度分析，判断对民宿价格、地理位置、交通通达程度、设施及服务、安全及应急性方面的分类是否合理。得到的结果如下所示：

表 3 效度分析结果表

层面	价格	地理位置	交通通达程度	设施及服务	安全及应急性
相关系数	0.932	0.953	0.974	0.926	0.926

## 第六章 质量控制

置信度	0.000	0.000	0.000	0.000	0.000
-----	-------	-------	-------	-------	-------

分析可知，P 值均小于 0.05，各样本单个题目的评分与该题总分是具有强正相关，问卷中量表测量的内容具有适当性和相符性。

### (4) 游程检验

在调查过程中，我们需要保证抽样的随机性，因此我们对其中的分类变量进行游程检验。根据样本标志排列所形成的游程的多少进行判断的检验，用上下交错形成的游程个数来检验样本是否是随机的。游程检验的检验统计量为 R，即游程个数。 $X_1, X_2 \dots X_N$  是一列由 0 或 1 构成的序列，假设检验问题为： $H_0$ ：数据出现顺序随机， $H_1$ ：数据出现顺序不随机。在  $H_0$  成立的情况下， $X_i \sim b(i, p)$ ，p 是 1 出现的概率，游程个数 R 的分布与 p 有关。在本调查中，我们采取单样本变量值的随机性检验，可以构造出 Z 统计量：

$$Z = \frac{R - 2 \frac{n_0 n_1}{n_0 + n_1} - 1}{\sqrt{\frac{2n_0 n_1 (2n_0 n_1 - n_0 - n_1)}{(n_0 + n_1)^2 (n_0 + n_1 + 1)}}} \quad \text{公式 (4)}$$

$$Z \sim N(0, 1)$$

因此用正态分布表得到 P 值和检验结果，在给定水平  $\alpha$  后，可以用近似公式得到拒绝域的临界值  $c_1$  和  $c_2$ ，满足  $P(R < c_1) < \alpha$  及  $P(R > c_2) < \alpha$ 。我们以被调查者性别为变量 X 来检验问卷数据是否是随机的。检验结果显著性为 0.146，大于 0.05，不拒绝原假设，故可以认为问卷数据是随机的，得出的结论能够进行推广，具有说服力。



# 第七章 总体认知 描述分析

## 七、总体认知描述分析

### (一) 消费者基本特征

#### 1. 性别分布

在调查的消费者中，男性所占比例为 45.60%，女性占比 54.40%，男女比例基本持平，符合 1: 1 的比例。在调查者中，愿意选择民宿的男性占比 71.40%，愿意选择民宿的女性占比 74%。由此可以看出，女性对民宿的包容度稍大于男性。民宿在营销方面可以适当偏向女性消费者作为目标。

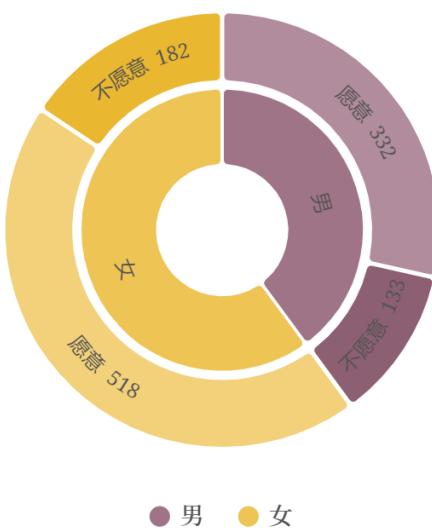


图 4 性别分布图

#### 2. 年龄分布

在愿意选择民宿的消费者中，25 岁以下占比 32.82%，占比最大，其次是 25-30 岁，为 31.88%，再次是 30-40 岁，为 29.76%，最后是 40 岁以上，仅占比 5.53%。可以看出年轻人是民宿市场的主力军。由于年轻人出差、旅游的频率比较高，其对于新事物的接受程度也比较大，因此民宿现有的市场应集中在较为年轻的群体之中。

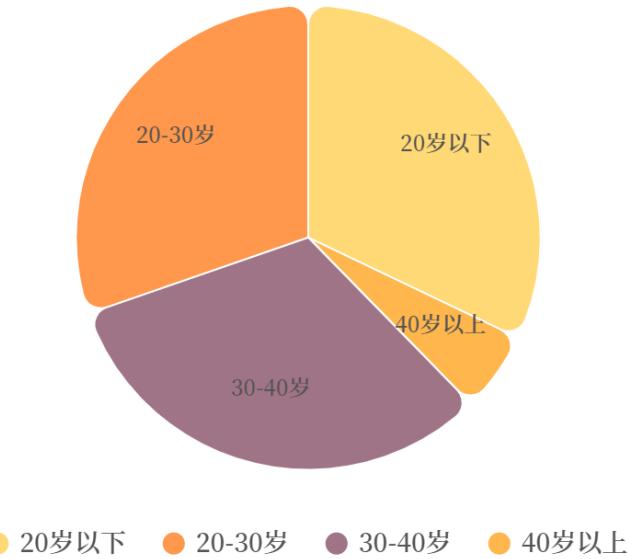


图 5 年龄分布图

### 3. 地区分布

在调查的消费者中，集中在东部地区人数较多，也基本符合我国目前的人口现状。

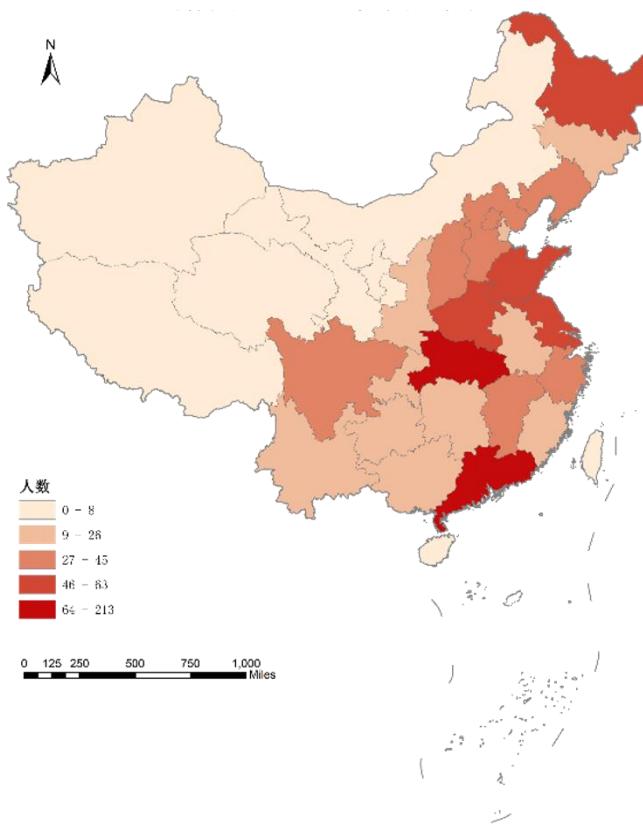


图 6 省份分布图

### 4. 职业分布

在调查的消费者中，全日制学生占 17.91%，文职/办公人员占 12.85%，销售人员占 11.13%，生产人员占 10.63%。行政/后勤人员和管理人员两个群体所占比例大致相当约 8%左右。在“是否愿意选择民宿”这个问题上可以明显看出，管理人员，文职人员，专业人士和教师对民宿的消费意愿非常强，这几个群体也具备消费能力，适合作为未来民宿的消费发展人群。

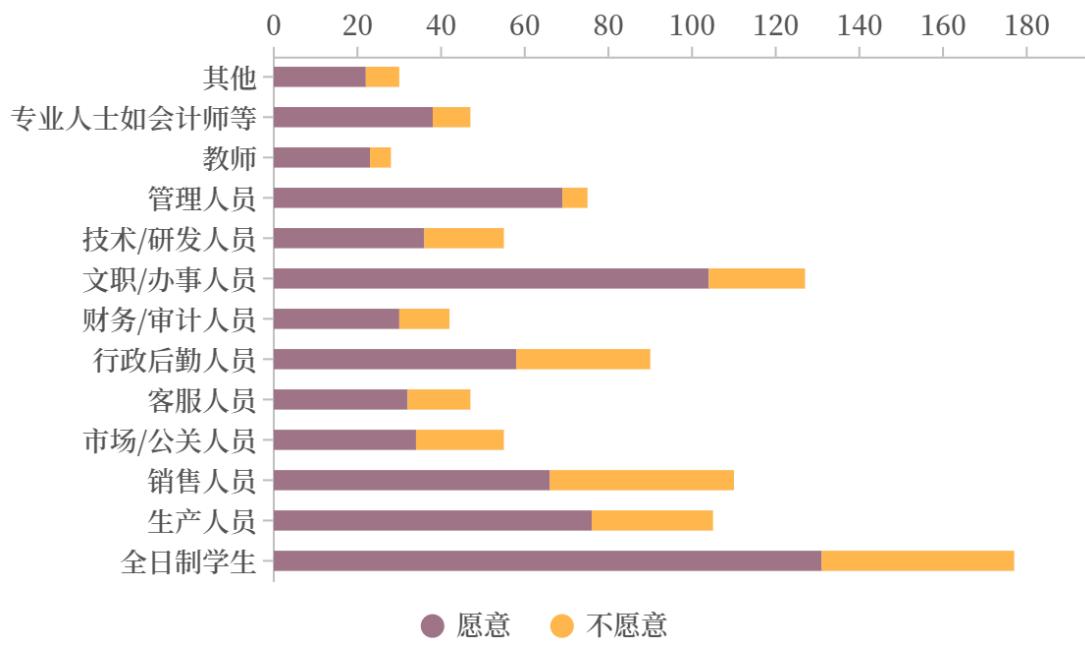


图 7 职业分布图

### 5. 月消费分布

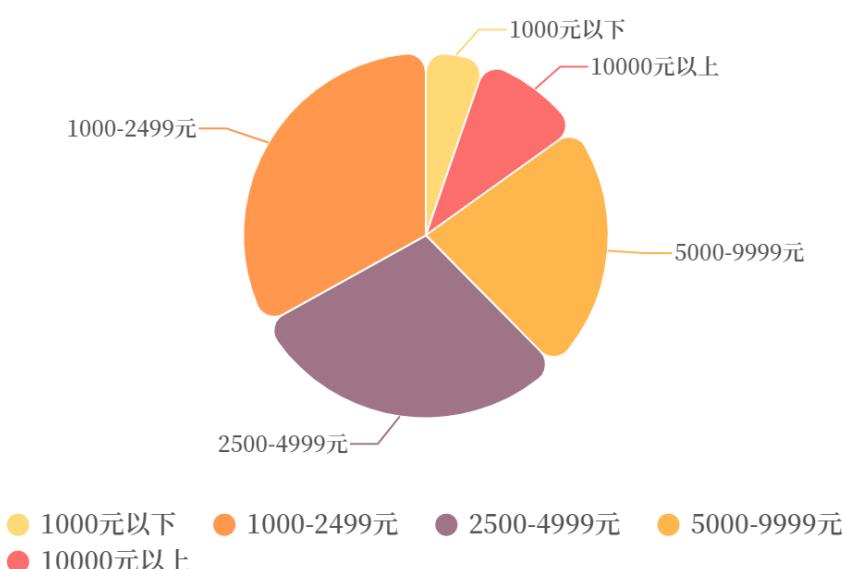


图 8 月消费分布图

由图可知，在本调查的所有消费群体中，月消费水平在 1000—2499 元的消费者占比最大，为 33.05%。其次，月消费水平为 2500—4999 元、5000—9999 元的消费者占比为 29.36%、22.49%。而月消费水平在 10000 以上和 1000 元以下的消费者占比为 9.78% 和 5.32%。由此可见，本次调查的消费群体主要为中消费群体，低消费群体和高消费群体的占比比较少。

## 6. 消费者在外住宿的目的分布

由图 7-6 可知，在调查的 1165 个消费者中，消费者在外住宿的目的为旅游的为 905 人，即 77.68% 的消费者都曾有过此目的，出差与聚会的人数分别为 499 人与 435 人，即 42.83% 和 37.34% 的消费者曾因为出差与聚会在外住宿。因此，可以得到，住客在外住宿的主要目的为旅游，而出差与聚会也是比较重要的目的。

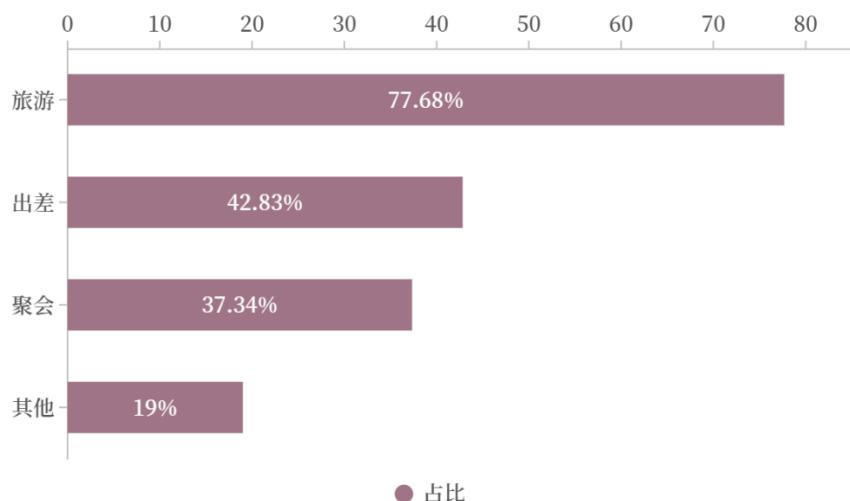


图 9 在外住宿的目的分布图

## 7. 消费者在外住宿的频率分布

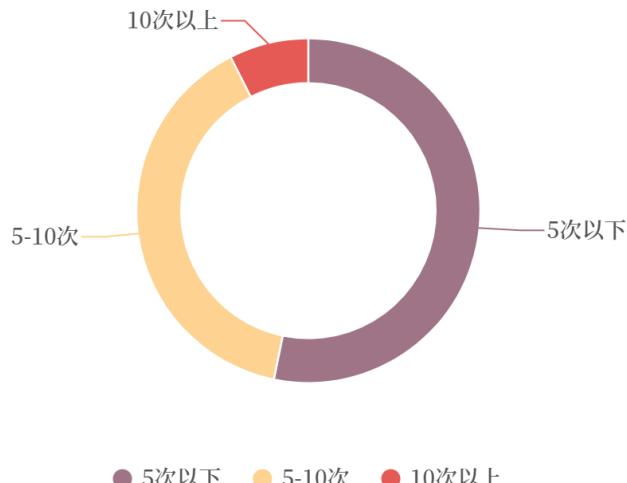


图 10 在外住宿的频率分布图

在调查的 1165 人中，有 53.22% 即 620 人每年在外住宿 5 次以下，有 39.31% 即 458 人每年在外住宿 5—10 次，有 7.47% 即 87 人每年在外住宿 10 次以上，由此可以得到，大部分人在外住宿的频率较低，每年在外住宿只有个位数。

### 8. 消费者住宿信息认知渠道

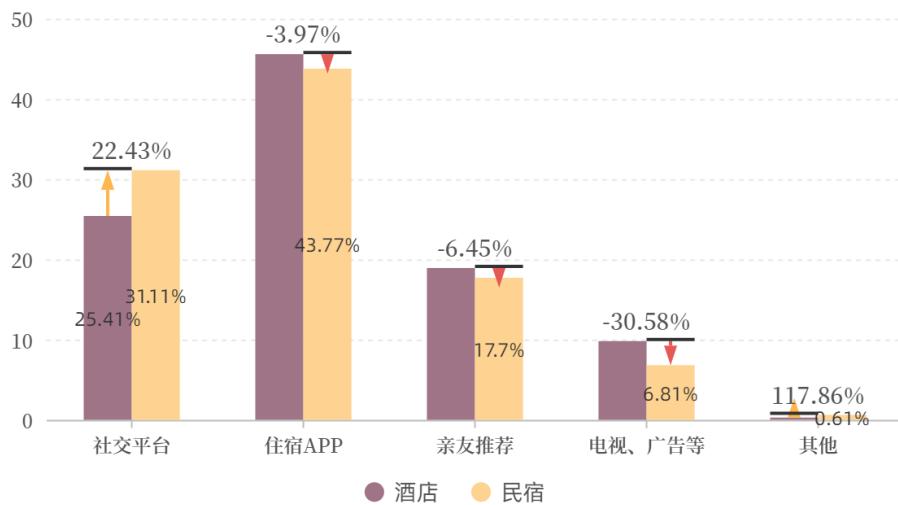


图 11 住宿信息认知渠道

认知渠道主要分为五大类，分为社交平台，含有住宿信息的 APP，亲友推荐，电视、广告等媒体和其他渠道。其中，对于酒店，住客的信息来源为以上五类的占比分别为 25.41%、45.58%、18.92%、9.81% 和 0.28%，由此可以得到：大部分人了解酒店信息的渠道为住宿 APP，其次分别为社交平台与亲友推荐。对于民宿，这五类信息来源的占比分别为 31.11%、43.77%、17.7%、6.81% 和 0.61%，住客的信息来源与酒店大体相似，即来源于住宿 APP 的比例最大，其次为社交平台。

与亲友推荐。

### 9. 消费者民宿居住意愿

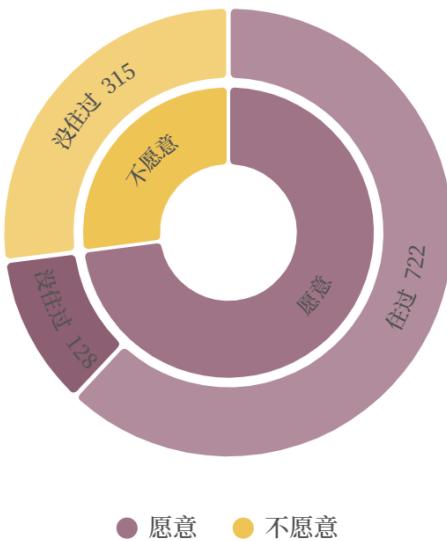


图 12 民宿居住意愿图

在 1165 个调查对象中,有 850 位被调查者愿意于民宿住宿,占比为 72.96%;其余的 315 位被调查者不愿意居住于民宿,占比为 27.04%。因此,可以得到:大部分被调查者都愿意接受于民宿住宿。

在愿意住宿民宿的 850 位被调查者中,有 722 位被调查者有过民宿住宿经历,其余的 128 位没有住过民宿,占所有被调查者的比例分别为 61.97% 和 10.99%,占愿意住宿于民宿的被调查者的比例为 84.94% 和 15.06%。由此,可以得到:大部分愿意住宿于民宿的消费者有过民宿住宿经验。

## (二) 民宿情况调查与分析

### 1. 选择民宿的原因分析

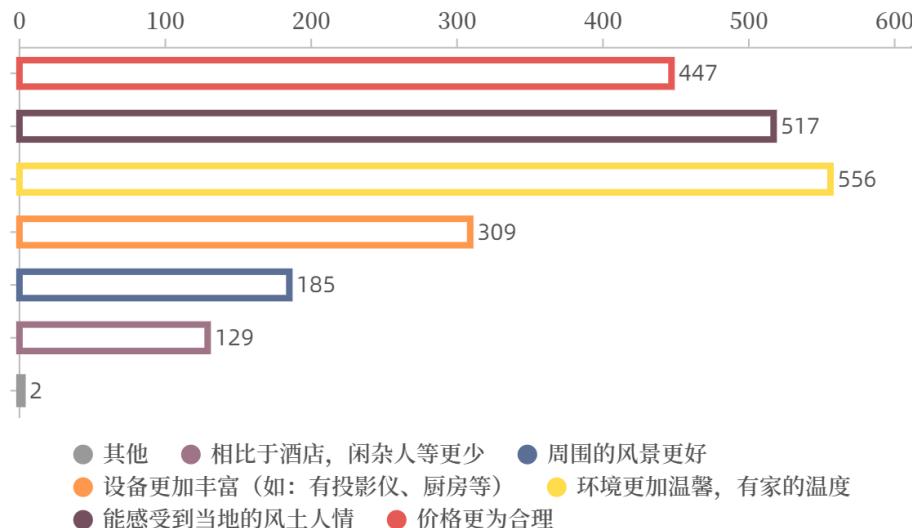


图 13 民宿住宿原因分析结果

住宿民宿的原因主要分为 6 个方面，分别是价格更为合理，能感受到当地的风土人情，环境更加温馨，有家的温度，设备更加丰富，周围的风景更好以及相比于酒店，闲杂人等更少。通过对调查数据进行分析，可以得到：民宿的环境更加温馨，有家的温度是消费者愿意选择民宿的主要意愿，其次为价格更为合适和能感受到当地的风土人情，占比分别为 47.73%、44.38% 和 38.37%。民宿住宿的几个原因的占比实际上比较平均，因此可以了解到住客的住宿原因比较多样。

### 2. 民宿价位分布

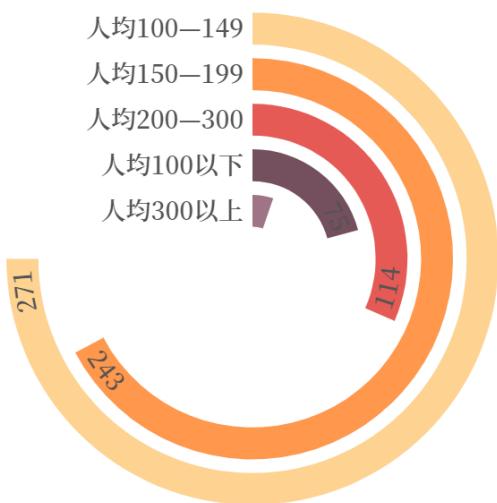


图 14 民宿价格分布

由图可得，人均 100—149 元的民宿最受消费者欢迎，占比为 37.53%，其次

人均 150—199 元的民宿也比较受欢迎，占比为 33.66%，之后的三个价位分布为人均 200—300 元、人均 100 元以下和人均 300 元以上，占比分别为 15.79%、10.39% 和 2.63%。由此可得：大众偏好于人均 100—199 元的民宿，人均 100 元以下和人均 300 元以上的民宿，由于价格过低和价格过高，大众选择意愿较弱。

### 3. 民宿类型分布

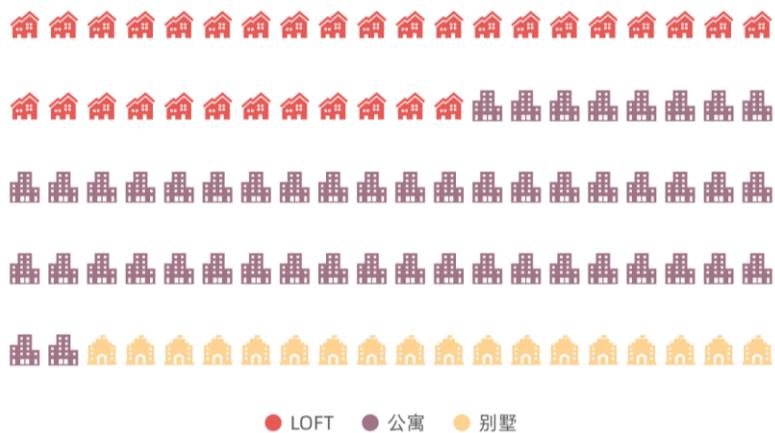


图 15 被调查者所选择过民宿类型的相对分布

民宿主要分为三大类型，为 LOFT、公寓和别墅，根据调查所得到的数据，有一半的消费者都曾经选择过类型为 LOFT 的民宿，75.94%的消费者选择过公寓类型的民宿，28.12%的消费者选择过别墅。由此，可以得出，三种类型的民宿都很受到大众欢迎，公寓最容易被选择，可以视为大众偏好公寓这一类型的民宿。

### 4. 民宿风格分布

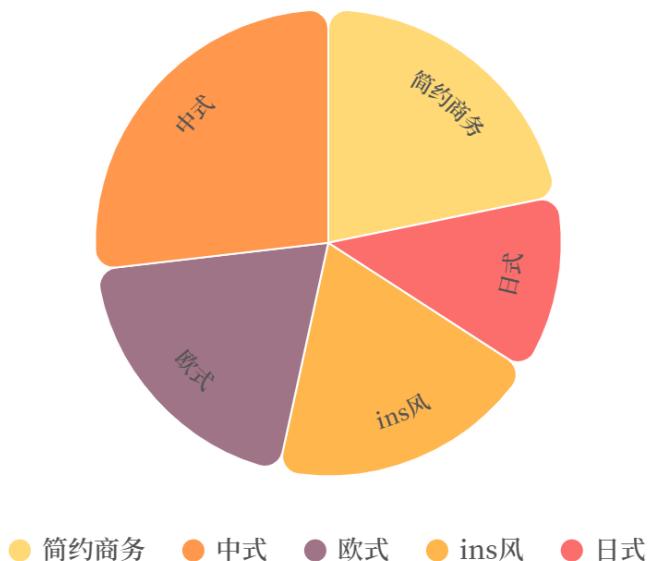


图 16 被调查者喜欢的民宿风格的相对分布

由图可知，占比最高的为中式风格，为 26.4%，其次为简约商务风，占比为 21.42%，欧式风格、ins 风、日式风格的占比分别为 19.41%、18.97% 和 12.16%。所以，大众的喜好比较平均，略微偏向于中式和简约商务风。

## 5. 民宿设施偏好

民宿所具有的设施主要可以分为 10 类，冰箱、投影仪、电视、浴缸、落地窗、厨具、停车位、洗衣机、烘干机、麻将机等。

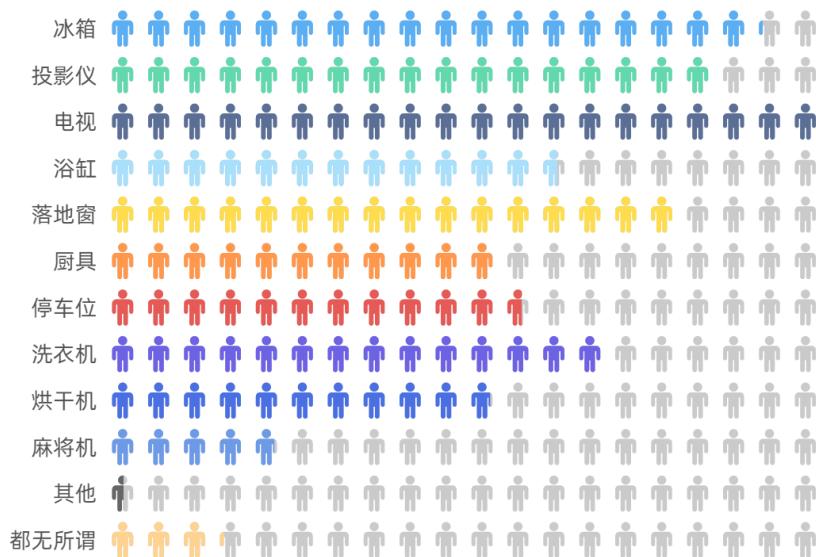


图 17 被调查者认为重要的民宿设施情况

可以看出，较高比例的消费者认为电视是一个很重要的设施，占比为 47.51%，冰箱、投影仪、落地窗紧随其后，占比 43.49%、40.30%、38.09%。另外，被调查者认为这些设施都无所谓的占比为 7.89%。因此，民宿在考虑规划时，应当优先增加电视、冰箱、投影仪等设备的投入。

## 6. 民宿出行情况分析

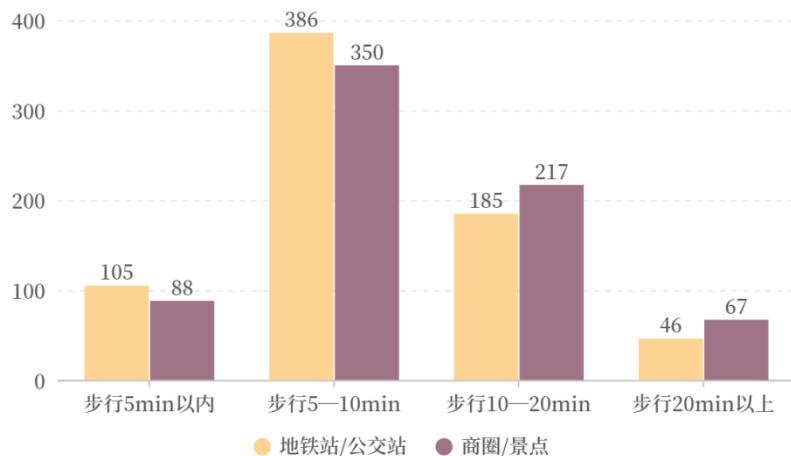


图 18 民宿出行情况统计

由图可得，对于地铁站和公交站，大部分的距离为步行 5—10 分钟可以到达的距离，其次为步行 10—20 分钟可以到达的距离，之后是步行 5 分钟以内可以到达的距离，最后也是最少的，步行 20 分钟以上才可以到达的距离。对于商圈和景点来说，分布与步行到达地铁站或公交站大致相同。因此，从民宿出发，到最近的地铁站/公交站、商圈/景点，距离大概为 400—800 米。

## 7. 民宿出行情况分析

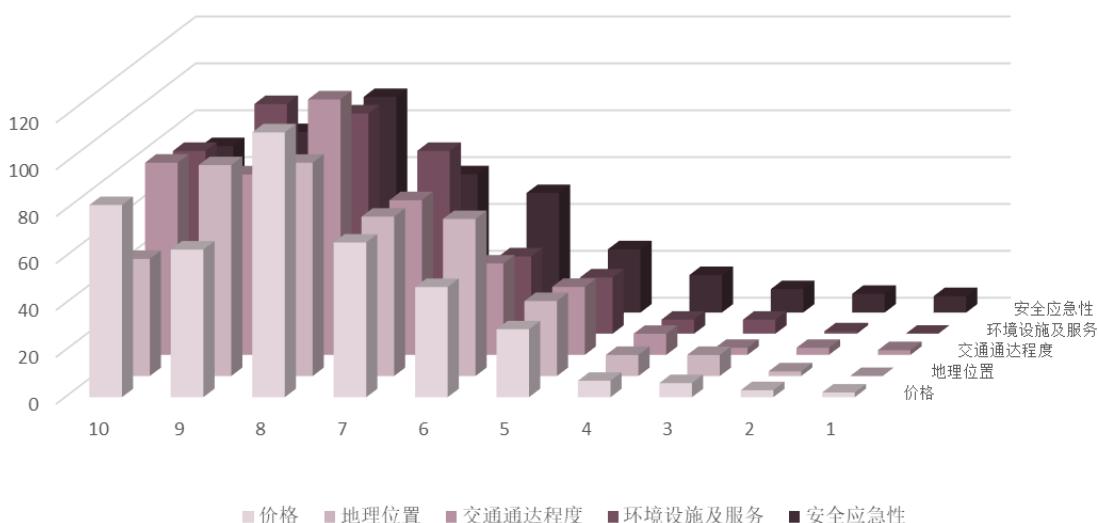


图 19 民宿选择影响因素结果

其中，1—10 为被调查者对重要程度的打分，10 为最重要，1 为最不重要。可以看出，价格、地理位置、交通通达程度、环境设施及服务、安全应急性这五部分总体上在选择民宿时还是十分重要的。总体上来说，人们对这五个方面都十分看重，其中最重要的为环境设施及服务，最不重要的为安全应急性。

## 8. 现有民宿满意程度分析

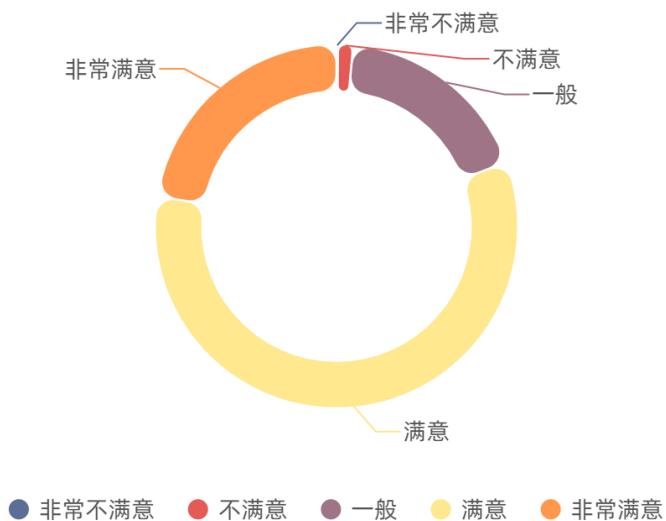


图 20 现有民宿满意程度

对现有民宿的满意程度进行分析，非常满意的占 22.36%，满意的占 58.37%，一般的占 17.8%，不满意的占 1.34%，非常不满意的占 0.13%。因此，从总体上来看，大众对于民宿的现有情况还是比较满意的，所以现有的民宿行业还是比较乐观的。

## （三）酒店情况调查与分析

### 1. 偏好酒店原因分析

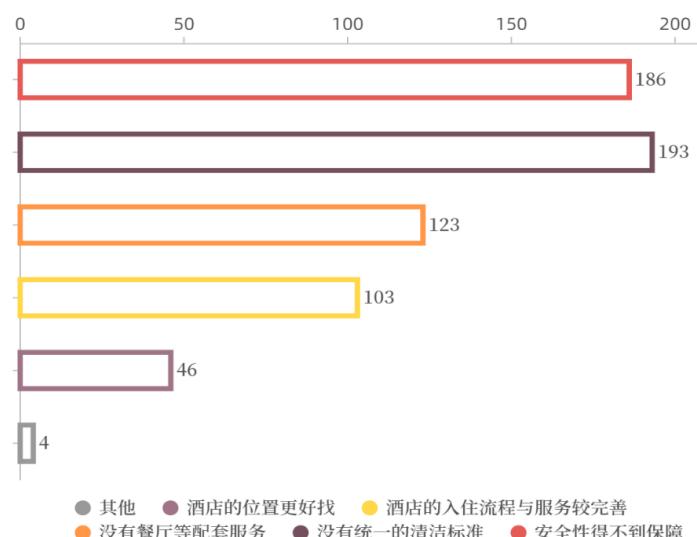


图 21 不选择民宿的原因调查结果

不愿意居住民宿人群总数的 61.27%认为，由于民宿没有统一的清洁标准，所以相比于酒店不愿意选择民宿；其次分别为没有专门的安保人员，安全性得不到保障，占比为 59.05%。其余三种原因为没有餐厅等配套服务、酒店的入住流程与服务较完善、酒店的位置更好找，占比分别为 39.05%、32.70% 和 14.60%。因此，民宿商家在进行民宿条件改善时，应该将重点放在提升自身的清洁标准，以及增加安保人员，保障住宿人员的安全性这两方面。

## 2. 酒店价位分布

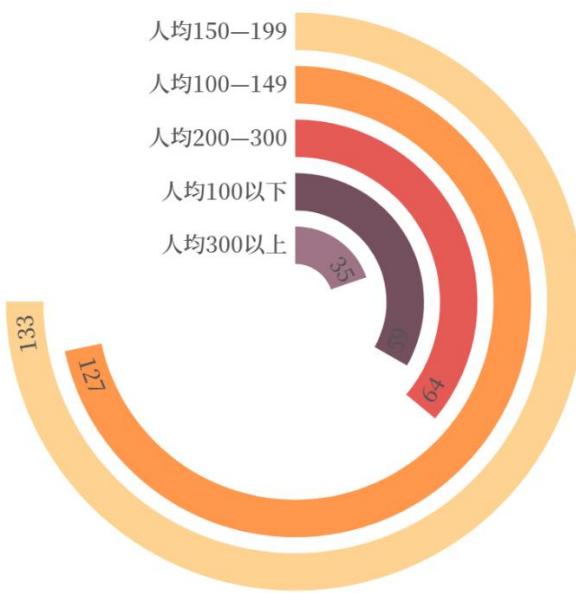


图 22 酒店价格分布

由图可得，人均 150—199 元的酒店最受消费者欢迎，占比为 31.82%，其次人均 100—149 元的酒店也比较受欢迎，占比为 30.38%，之后的三个价位分布为人均 200—300 元、人均 100 元以下和人均 300 元以上，占比分别为 15.31%、14.11% 和 8.37%。由此可得：大众偏好于人均 150—199 元的酒店，人均 100 元以下和人均 300 元以上的酒店，大众选择意愿较弱。与民宿相比，消费者选择过的民宿的价格比酒店低，民宿可能在价位上更加实惠。

## 3. 酒店出行情况分析

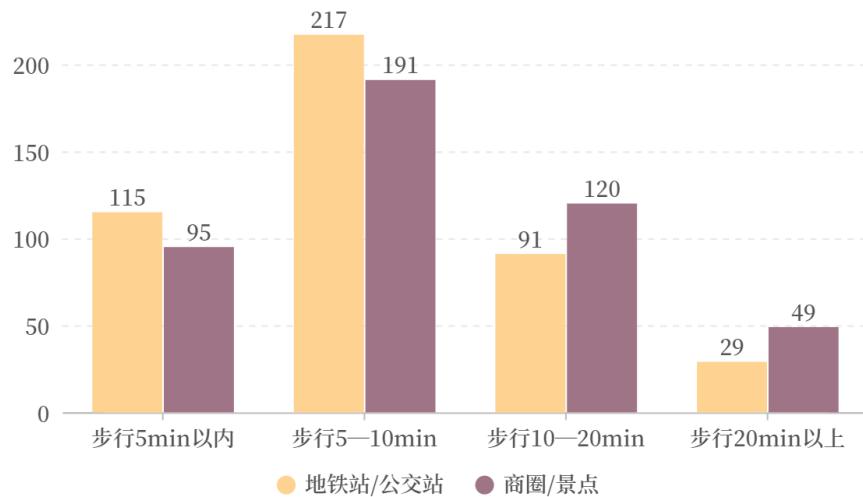


图 23 酒店出行情况统计

由图可得，对于地铁站和公交站来说，大部分的距离为步行 5—10 分钟可以到达的距离，其次为步行 10—20 分钟可以到达的距离，之后是步行 5 分钟以内可以到达的距离，最后也是最少的，步行 20 分钟以上才可以到达的距离。对于商圈和景点来说，分布与步行到达地铁站或公交站大致相同。因此，从酒店出发，到最近的地铁站/公交站、商圈/景点，距离大概为 400—800 米。

将酒店出行情况结果与民宿出行情况结果进行对比，可以得到：两个结果的分布大体相似，出行到附近的地铁站/公交站和商圈/景点的距离大部分为步行 5—10 分钟可以到达的距离，即 400—800m。

#### 4. 酒店选择因素分析

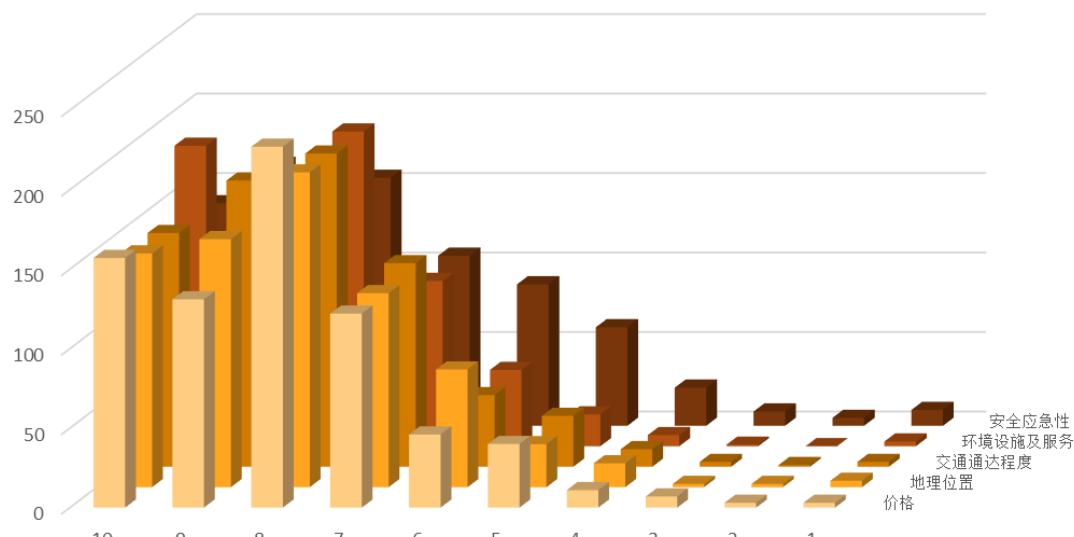


图 24 酒店选择影响因素结果

其中，1—10 为被调查者对重要程度的打分，10 为最重要，1 为最不重要。可以看出，价格、地理位置、交通通达程度、环境设施及服务、安全应急性这五部分总体上在选择酒店时还是十分重要的。总体上来说，人们对这五个方面都十分看重，其中最重要的为环境设施及服务，最不重要的为安全应急性。

将酒店选择影响结果与民宿选择影响结果进行对比，这两幅柱状图的结果在大体趋势上是相同的，另外最重要和最不重要的都为环境设施与服务、安全应急性。因此，商家在规划住所时，应该优先考虑环境设施与服务这一方面。



# 第八章 调查问卷

## 交叉分析

## 八、调查问卷交叉分析

为了更准确的分析不同题目间的相互关系及其影响程度,我们运用交叉分析的方法更深入挖掘并分析问卷内涵,进而结合民宿市场现状对未来发展提出实质性的建议。

### (一) 消费者在外住宿目的分析

将消费者年龄与在外住宿目的交叉分析,研究不同年龄段消费者的需求情况,可得到的以下结果:

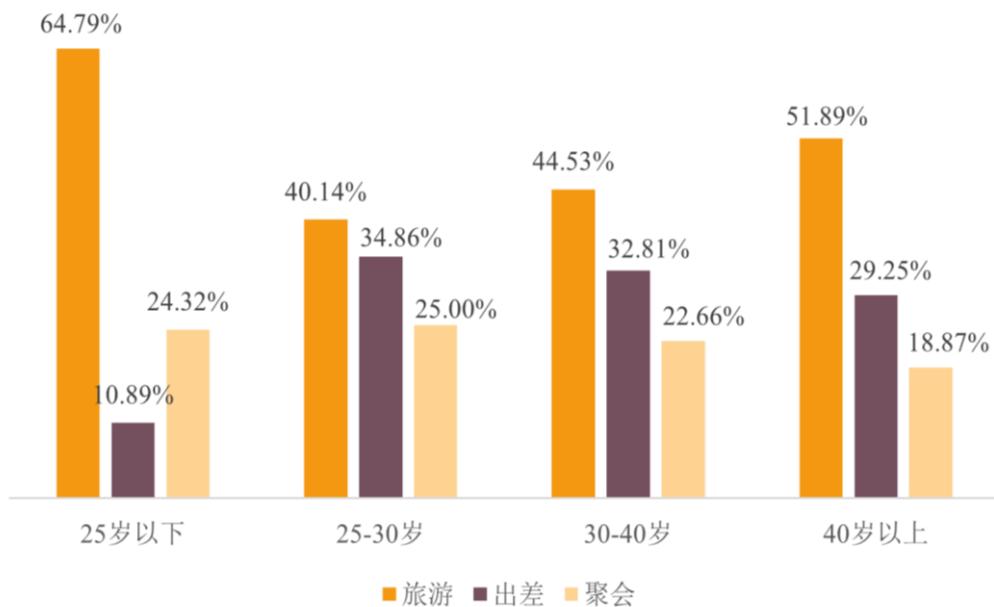


图 25 消费者年龄与在外住宿目的

数据表明 25 岁以下的消费者中,以旅游为目的的人数超过 64.79%,仅有 10.89%选择出差为目的。在 25-30 岁、30-40 岁顾客群体中,旅游、出差、聚会的选择比例较为均衡,但仍表现出以旅游为主要目的的趋势,分别达到 40.14% 和 44.53%,选择出差的消费者占比分别上升至 34.86% 和 32.81%。在 40 岁以上的消费者群体中,选择旅游为目的的消费者比例出现上升达到 51.89% 超过半数,而选择出差、聚会的比例略微降低,分别为 29.25% 和 18.87%。

25 岁以下的消费者中绝大部分会因旅游而选择外出住宿,25 岁以上的消费群体中因旅游而外出住宿的比例急剧下降,随着年龄增长,选择出差的人数比例出现明显上升,而在聚会方面的需求则变化不大,甚至有略微下降的趋势。这样的比例变化趋势也和不同年龄段的工作、家庭情况变化相呼应。

## (二) 在外住宿频率的分析

将消费者职业与在外住宿频率交叉分析，可得到的以下结果：

表 4 不同职业消费者每年在外住宿频率卡方检验

	值	自由度	渐进显著性（双侧）
皮尔逊卡方	125.037 <sup>a</sup>	24	0.000
似然比	127.718	24	0.000
线性关联	14.054	1	0.000
有效个案数	1165		

a. 7 个单元格（17.9%）的期望计数小于 5。最小期望计数为 2.09。

由表可知 Pearson 的卡方独立性检验对应  $p < 0.05$ ，拒绝接受原假设，即职业与外出住宿次数有显著相关关系。

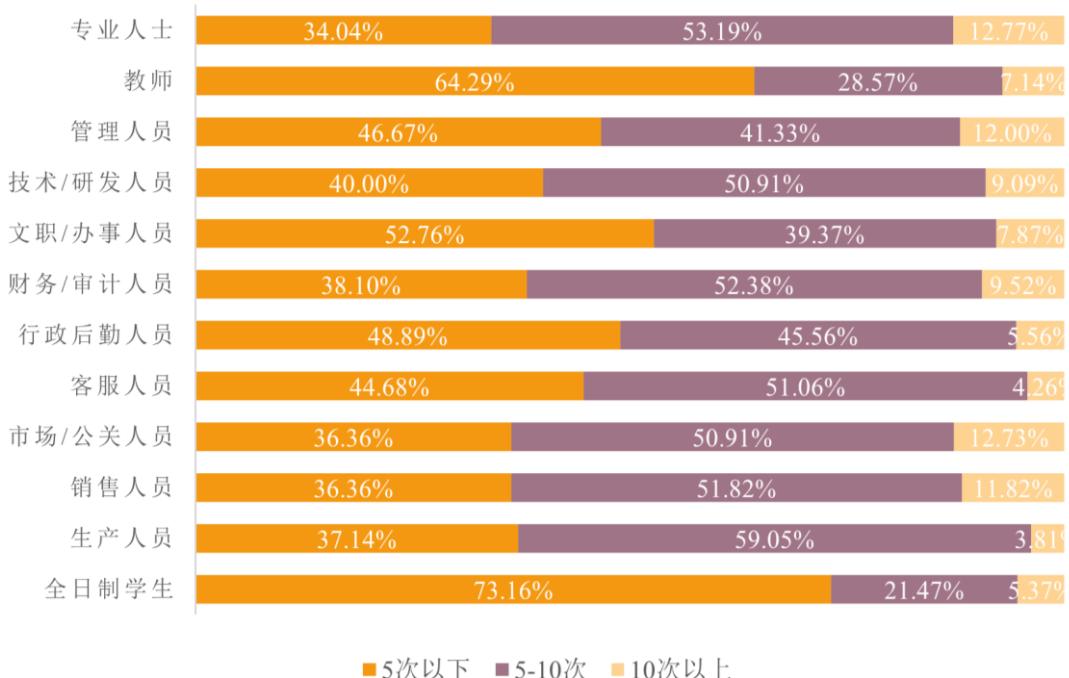


图 26 消费者年龄与每年在外住宿频次

由表可知，教师、文职/办事人员中，分别有 64.29%、52.76% 的比例仅外出住宿 5 次以下，销售人员、市场/公关人员、管理人员、专业人士（会计师、律师、医护人员、记者等）中外出住宿 10 次以上的比例在所有职业中较高。不同职业的频次差异和各自的职业特点息息相关，办公地点相对固定、活动强度较小的职业外出需求小，因此外出住宿频次也相对较低；流动性较强、外出需求较大的职业从业人员外出住宿频次也相应较高，是值得我们重点关注的发展对象。

### （三）民宿风格

将消费者年龄与所选的民宿风格交叉分析，可得到的以下结果：

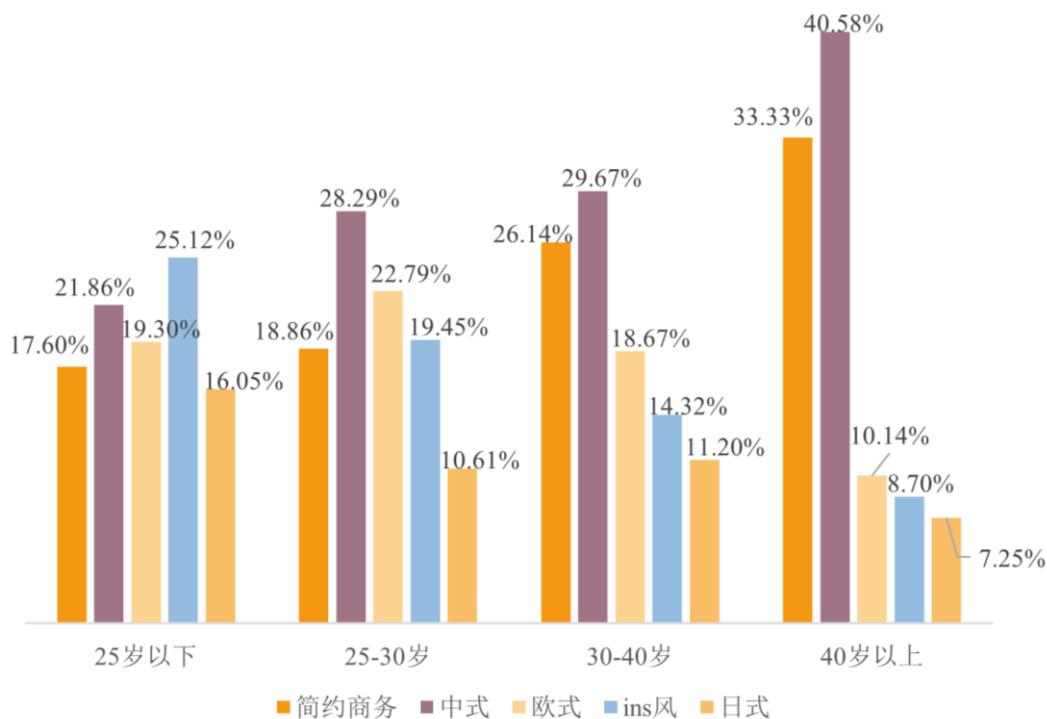


图 27 消费者年龄与民宿风格

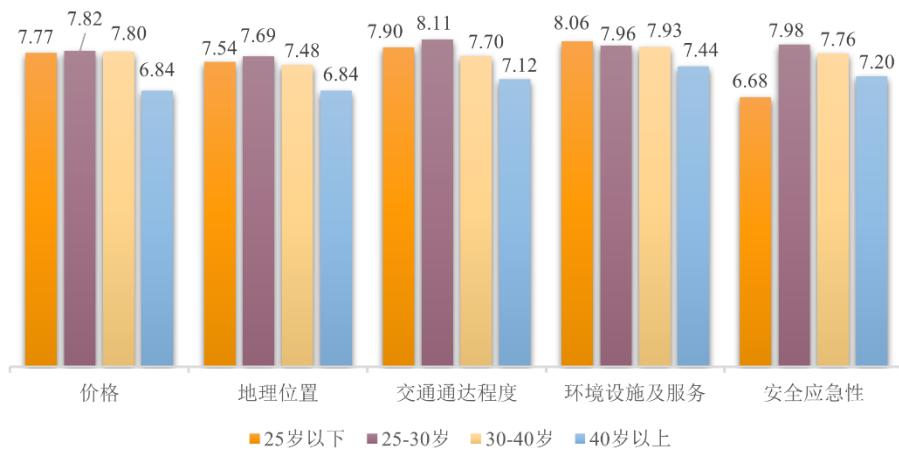
由表可知，25 岁以下的消费者群体对民宿风格的偏好比较多样，有超过 25% 的消费者选择 ins 风格的民宿。25-30 岁、30-40 岁的消费群体中，分别有 28.29%、29.67% 的消费者喜爱中式风格的民宿。在 40 岁以上的群体中，喜爱中式风格的民宿占到了 40.5%。此外，对于简约商务风格的民宿的偏好，随着年龄段的增长而增加，25 岁以下的群体中仅有 17.67% 的人选择该风格，40 岁以上则增长至 33.33%。

综合来看，年轻人较倾向于 ins 风等此类在网络上比较流行的风格，愿意选择的风格也趋于多样化。而随着年龄的增加，对“网红”风格追捧的热情逐渐减弱，消费者的选择更加趋同，偏向于选择较为简单务实的商务风格。此外，对于蕴含

我国传统文化的中式风格也受到各年龄阶段的喜爱，具有很大的发展潜力和推广内涵。

### (四) 消费者的选择民宿的关注因素分析

将消费者年龄与在选择民宿时考虑的因素的交叉分析，研究不同年龄段消费者的关注情况，可得到的以下结果：

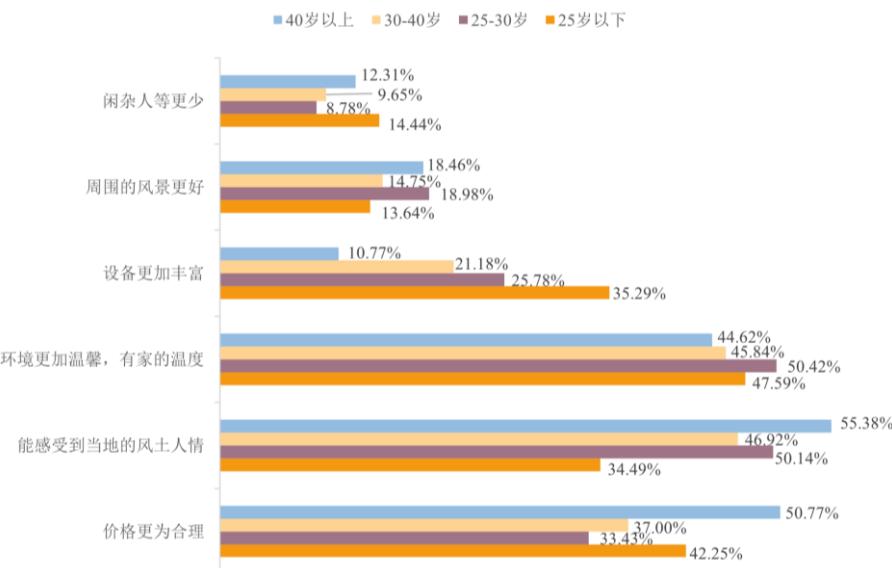


从整体得分情况来看，高分群体表现出年轻化的特征，从侧面说明年轻人是民宿消费市场的主力军。绝大部分消费者都认为环境设施及服务是影响民宿选择的重要因素。此外，年轻消费者更关注交通通达程度，25-30岁群体将其看作选择民宿的最重要因素。30-40岁年龄层较关注民宿的价格。不同年龄层对安全应急性的重要程度评分有着显著差异，25岁以下客户评分较低，这一数据随着年龄上升而增高。

综合可知，民宿环境设施及服务是各年龄段选择民宿时最关注的对象，年轻消费者对交通、价格关注度较高，随着年龄增长对安全应急性越来越重视。在经营者对民宿进行营销时，可充分结合各自关注特点对不同年龄层消费者进行民宿宣传。

### (五) 消费者的选择民宿的原因分析

将消费者年龄与愿意选择民宿的原因的交叉分析，研究民宿不同方面特点对不同年龄段消费者的吸引情况，可得到的以下结果：



民宿温馨的环境是绝大部分消费者原因选择民宿的原因，25岁以下群体认为民宿相比于酒店较为合理的价格也是民宿的显著的优势，丰富的设备也会让他们更愿意选择民宿。25-30岁、30-40岁群体则更关注民宿带来的本地风土人情，这也是40岁以上群体选择民宿的最主要的原因。

综合来看，25岁以下群体及40岁以上群体会因民宿实惠的价格而选择民宿，民宿风土人情的呈现、家的温馨感也是民宿相比酒店的两大优势。在民宿推广中可结合客户不同年龄段特点注重对民宿不同方面的建设与宣传。



# 第九章 民宿消费者特征 Logistic 回归模型分析

## 九、民宿消费者特征 Logistic 回归模型分析

### (一) 模型的建立

在回归分析中，预测变量与效标变量通常是连续变量，如果自变量是间断变量，在投入模型前要转化为虚拟变量。如果因变量是间断变量且为二分名义变量，则可使用 Logistic 回归分析法，Logistic 回归分析中自变量为等距或比率变量，因变量则是二分的类别变量。Logistic 回归分析的假定是观察值样本在因变量上的概率分布呈 S 形分布。在参数估计方面，Logistic 回归分析是通过极大似然估计使因变量观察次数的概率极大化，进而得到自变量参数的最佳估计值，同时 Logistic 回归模型通过概似值进行整体检验。

综合所调查的性别、年龄、职业、月消费等信息，以愿意选择民宿作为因变量，通过构建 Logistic 回归模型探索出于不同类别的消费群体针对民宿市场的接受程度，从而预测出对潜在的目标群体。模型中的自变量，由于本次设置的问卷调查中为分类选择，例如年消费水平分成了几种类别，如果直接编码为 1-5，令其作为自变量纳入分析，与实际情况不符。另外对于无序多分类变量，如职业，它们之间不存在数量上的高低，因此不可能为其给出一个单独的回归系数估计值，来表示职业每上升一个单位时因变量的变化趋势。因此，本调查采用统计上标准的做法，采用虚拟变量进行拟合，然后根据分析结果对模型进行简化。

模型中的因变量是二分名义变量，两个水平变量分别为“愿意”，水平数值编码为 1；“不愿意”水平数值编码为 0。模型中的自变量均为二分变量，转化为虚拟变量投入模型，这些因素会导致愿意选择民宿为水平数值 1，这些因素不会导致愿意选择民宿为水平数值 0。因而采用 Logistic 回归模型，以研究这些因素对选择民宿的影响程度。将全部自变量均纳入回归模型中，再通过个别自变量显著性的参数估计剔除不显著的变量，以得到最终的 Logistic 回归模型。

### (二) 结果分析

表 5 为影响是否愿意选择民宿因素的 Logistic 回归模型适配度检验结果。模型系数 Omnibus 检验的卡方值为 55.858,  $p=0.000 < 0.05$ , 通过显著性检验。Hosmer-Lemeshow 检验值 8.326, 伴随概率  $p=0.305 > 0.05$ , 未达显著水平，整体回归模型的适配度良好，表示自变量可以有效解释因变量。

表 5 整体模型适配度检验摘要表

检验方法	检验值	p
Omnibus 检验	55.858	0.000
Hosmer-Lemeshow 检验	8.326	0.305

选用向后逐步筛除方法，对初步建立的回归方程根据极大似然估计的统计量的概率值删除对因变量影响不显著的自变量，显著性水平给定为 0.05。经过多轮筛选，排除不显著的变量后，得到的回归结果如下：

表 6 模型参数显著性检验摘要表

自变量名称	系数	Wald 值	P 值	Exp(B)
age_2	0.402	6.478	0.011	1.495
profession_c	-0.555	6.484	0.011	0.574
profession_h	0.635	6.519	0.011	1.888
profession_i	1.726	15.573	0.000	5.620
profession_l	0.763	3.914	0.048	2.144
cost_2	0.680	16.461	0.000	1.973
cost_3	0.385	5.340	0.021	1.469
常量	0.444	12.436	0.000	1.559

## 1. 性别

在二十几个因素的筛选中，性别对于是否选择民宿没有明显的影响关系，因此在最后的几个因素中没有被留下。也证明了性别差异对民宿市场没有直接的影响。

## 2. 年龄

自变量 Age\_2 为 25-30 岁所对应的 Exp(B)胜算比值为 1.495，表示样本认同这一因素会导致愿意选择民宿整体的胜算是这一因素不会导致愿意选择民宿整体的胜算的 1.495 倍。这表明 25-30 岁阶段的消费者是 90 后，这一人群具有更加宽裕的时间，并大多有稳定的收入，也拥有更多的精力与兴趣，所以显著性很高，更愿意选择民宿作为他们的暂时住所。

## 3. 职业

模型估计结果显示，消费者有四种职业类别（销售人员；文职/办事人员；技

术/研发人员；专业人士如会计师、律师、医护人员、记者等）通过了模型系数的显著性检验，它们的系数分别为：-0.555、0.635、1.726、0.763，这说明不同的职业影响着消费者是否愿意选择民宿。从 EXP(B)值来看，销售人员群体愿意选择民宿整体的胜算是这一因素不会导致愿意选择民宿整体的胜算的 0.574 倍。这可能是销售人员对于一些民宿的销售套路比较了解，因此更愿意选择营销比较少的酒店作为他们的住所。而文职/办事人员，技术/研发人员，专业人士如会计师、律师、医护人员、记者等职业对于民宿的包容度更高，更愿意选择民宿。

## 4. 月消费

模型估计结果显示，月消费级别为2、3级（1000-2499元、2500-4999元）通过了模型系数的显著性检验，它们的系数分别为：0.680、0.385，所对应的Exp(B)胜算比值分别为1.973和1.469，表示样本认同这一因素会导致愿意选择民宿整体的胜算是这一因素不会导致愿意选择民宿整体的胜算的1.973和1.469倍。这两个月消费级别所包含的人数过半，属于中低水平消费人群。这表明改消费水平人群更愿意选择民宿作为住所，这也是民宿行业目前的主力人群。而高消费水平人群则是需要进一步挖掘的潜在消费群体。



# 第十章 民宿消费 者影响因素分析

## 十、民宿消费者影响因素分析

在前面，我们对在外住宿的消费者的特点进行分析，得到了消费者特点与是否愿意住民宿之间的关系模型，得到了愿意住民宿群体的合理定位。下面我们将进一步深入分析，考虑消费者深度的几个层次，通过建立排序选择模型，研究不同因素对选择民宿入住的影响。

### (一) 排序选择模型的建立

在这里，我们选择住过民宿的人群去居住民宿的频率作为是否选择民宿入住的量化指标。根据最大效用最大化的原则，选择具有极限值的Logistic分布，离散选择模型采用 Logit模型：

$$\text{logit}(p(y < g)) = \ln\left(\frac{F(\mu_g - X'\beta)}{1 - F(\mu_g - X'\beta)}\right) = \mu_g - X'\beta \quad \text{公式 (5)}$$

从消费者的年龄、消费水平、住宿目的等，探究各变量对在外住宿时选择民宿的影响，建立的模型为

$$Y_I = \begin{cases} 1 & y^* \leq \sigma_1 \\ 2 & \sigma_1 \leq y^* \leq \sigma_2 \\ 3 & y^* \geq \sigma_2 \end{cases} \quad \text{公式 (6)}$$

带估计的潜回归模型：

$$Y^* = X\beta + \mu \quad \text{公式 (7)}$$

其中 $\beta$ ,  $\sigma_1$ ,  $\sigma_2$ 为待估计参数，具体变量和符号见附录五。

### (二) 模型估计与分析

表 7 模型参数显著性检验摘要表

自变量名	系数估计值	标准误	Wald 值	显著性
cost_1	2.433	0.589	17.029	0.000
cost_2	1.760	0.312	31.706	0.000
cost_3	0.886	0.282	9.854	0.002
cost_4	0.545	0.287	3.601	0.058

<b>toor</b>	0.534	0.208	6.593	0.010
<b>business</b>	-0.748	0.179	17.399	0.000
<b>party</b>	-0.589	0.156	14.190	0.000

平行回归假设检验是为了检验在不同等级的有序多分类结果中，解释变量的效应是否保持一致，不会随等级的不同而变化，该检验用于判断模型是否具有系统性差别。利用问卷收集的数据，我们对平行回归假设进行 LR 似然比检验，检验结果如表所示：

表 8 平行回归假设检验结果

模型	-2 对数似然值	$\chi^2$	自由度	显著性
零假设	418.886			
广义	398.645	20.242	13	0.057

P值大于显著性水平  $\alpha = 0.05$ ，不能拒绝原假设。

### (三) 模型结果分析

#### 1. 性别

在十几个因素的筛选中，性别对于在外住宿时是否选择民宿没有明显的影响关系，故可以在影响因素表中去掉，这也说明了性别差异对民宿市场没有直接的影响。

#### 2. 年龄

在十几个因素的筛选中，年龄对于在外住宿时是否选择民宿的影响不大，20-30年龄段对选择民宿具有略微明显的优势。

#### 3. 消费水平

自变量 cost\_1、cost\_2、cost\_3 和 cost\_4 的系数估计值分别为 2.433、1.760、0.886 和 0.545 意味着这四个组别的人群选择民宿的可能性是 cost\_5（月消费水平大于 10000 元）的组的 11.39、5.81、2.42 和 1.72 倍，这表示中低水平消费人群相较于高消费群体更愿意选择民宿作为在外住宿的场所。

### 4. 住宿目的

在外住宿的目的包括了旅游、出差、聚会和其他，设置了3个虚拟变量，以其他类别作为基准，根据模型结果，可以旅游的系数估计值为0.534，是其他的1.71倍，说明以旅游为目的的人群会比其他目的更倾向于选择居住民宿，而聚会和出差没有显著影响。



# 第十一章 武汉市 民宿适应性评价

## 十一、武汉市民宿适宜性评价

### (一) 模型构建

从消费群体对民宿的需求出发,运用层次分析法(AHP)将民宿适宜性评价指标分解为目标层、准则层、指标层,从实惠性、可步性和安全应急性3个维度选取6个评价指标,最终构建出武汉市民宿适宜性评价的AHP模型(图52)。

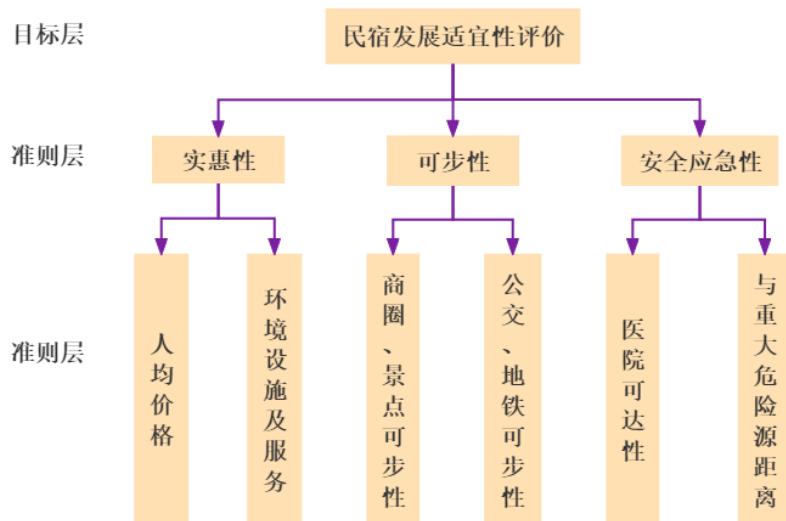


图 52 民宿综合性评价的层次分析模型

建立层次分析结构模型后,确立了上下层之间的关系,要解决的问题则转化为确定各层次各因素之间的权重。在这里,通过对专家进行咨询,并结合获得的1156份问卷调查的结果,得到各要素之间的相对重要性,构造各准则层、指标层的判断矩阵(表23至表26),对所得的判断矩阵进行一致性检验。

表 22 判断矩阵元素标度方法

标度	含义
1	表示2个因素相比,具有同样重要性
3	表示2个因素相比,一个因素比另一个因素稍微重要
5	表示2个因素相比,一个因素比另一个因素明显重要
7	表示2个因素相比,一个因素比另一个因素强烈重要
9	表示2个因素相比,一个因素比另一个因素极端重要
2, 4, 6, 8	上述两相邻判断的中值
倒数	因素 <i>i</i> 与 <i>j</i> 比较的判断为 <i>a<sub>ij</sub></i> ,则因素 <i>j</i> 与 <i>i</i> 比较的判断为 <i>a<sub>ji</sub> = 1/a<sub>ij</sub></i>

表 23 判断矩阵  $A-B$ 

$A$	$B_1$	$B_2$	$B_3$	$W$
$B_1$	1	1	3	0.4434
$B_2$	1	1	2	0.3874
$B_3$	1/3	1/2	1	0.1692

$\lambda_{\max} = 3.0183$ ,  $CI = 0.0092$ , 查表得  $RI_3 = 0.58$ ,  $CR = CI/RI_3 = 0.0157 < 0.10$ , 判断矩阵满足一致性检验。

表 24 判断矩阵  $B_I-C$ 

$B_1$	$C_1$	$C_2$	$W$
$C_1$	1	2	0.5000
$C_2$	1/2	1	0.5000

$CR = 0 < 0.10$ , 判断矩阵满足一致性检验。

表 25 判断矩阵  $B_2-C$ 

$B_2$	$C_3$	$C_4$	$W$
$C_3$	1	1/2	0.5000
$C_4$	1	1	0.5000

$CR = 0 < 0.10$ , 判断矩阵满足一致性检验。

表 26 判断矩阵  $B_3-C$ 

$B_3$	$C_5$	$C_6$	$W$
$C_5$	1	2	0.6667
$C_6$	1/2	1	0.3333

$\lambda_{\max} = 2.000$ ,  $CI = 0$ ,  $CR = 0 < 0.10$ , 判断矩阵满足一致性检验。

对指标层进行层次总排序计算, 得到所有元素相对于目标层的重要性排序权重, 结果如表 27 所示。

表 27 适宜性评价指标及权重值

目标层	准则层	准则层权重	指标层	指标层权重	对目标权重
民宿综	实用性 $B_1$	0.4434	人均价格 $C_1$	0.5000	0.2217

合性评价 A			环境设施及服务 $C_2$	0.5000	0.2217
	可步性 $B_2$	0.3874	商圈、景点可步性 $C_3$	0.5000	0.1937
	安全应急性 $B_3$	0.1692	公交、地铁可步性 $C_4$	0.5000	0.1937
			医院可达性 $C_5$	0.6667	0.1128
			与重大危险源距离 $C_6$	0.3333	0.0564

从结果可以看出，人均价格，环境设施及服务，商圈、景点可步性以及公交、地铁可步性对民宿适宜性影响较大，也是消费者会优先考虑的因素。

## (二) 综合指数计算

(1) 利用 ArcGIS 建立武市民宿地理信息数据库。

(2) 对各指标进行赋分。利用自然断点法划分为 10 个级别，并根据级别分别赋予 1-10 分，尽量减少主观因素的影响。

(3) 计算综合指数。利用公式：

$$P = \sum_{i=1}^{13} W_i F_i \quad (i = 1, 2, \dots, 13) \quad \text{公式 (14)}$$

计算出武汉市每个民宿综合指数，其中  $W_i$  为表 xx 中的对目标权重值， $F_i$  为上述步骤中，对各指标赋予的分值。

(4) 民宿适宜性评价。将计算所得的综合指数划分为 5 个等级(表 28)，武市民宿适宜性评价结果如图 53，图 54 所示。

表 28 武市民宿适宜性等级划分

综合指数	等级	等级说明
8.01-10.00	A	整体情况好，在实惠性、可步性和安全应急性方面均较合理，需要进行保持和完善
6.01-8.00	B	整体情况较好，需要根据评分有针对性的改进不足之处
4.01-6.00	C	整体情况中等，基本满足消费者需求，需进一步改进
2.01-4.00	D	整体情况较差，存在较大的改进空间
0.00-2.00	E	整体情况差

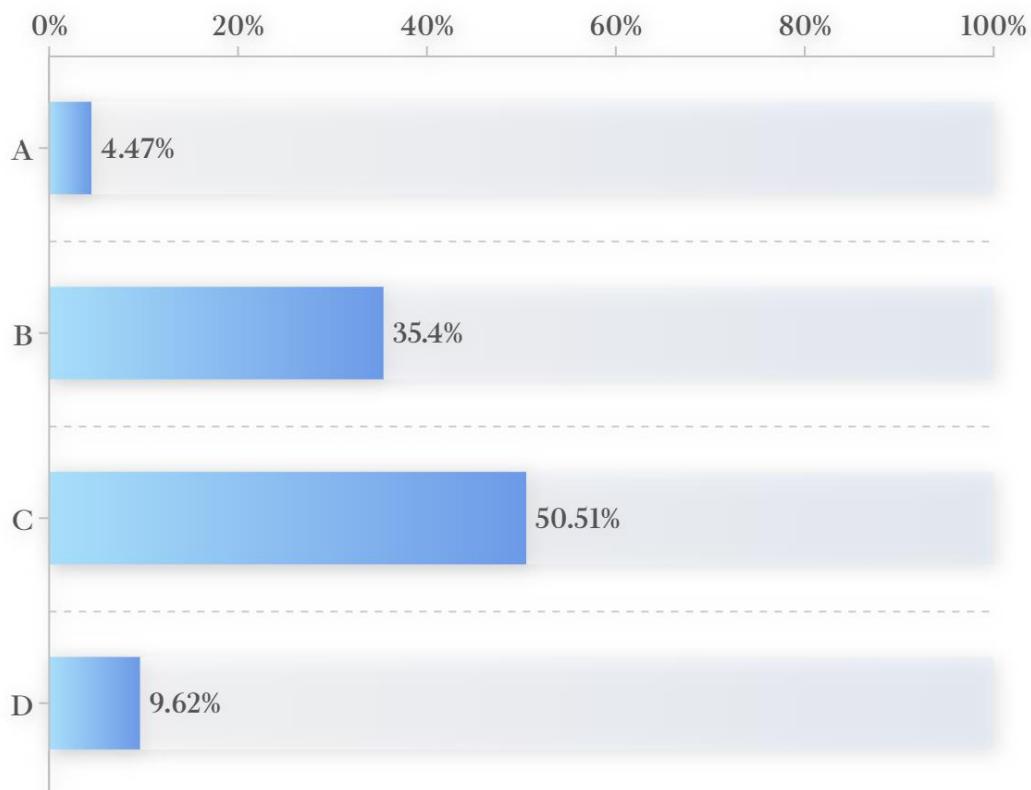


图 53 武市民宿适宜性评价等级分布

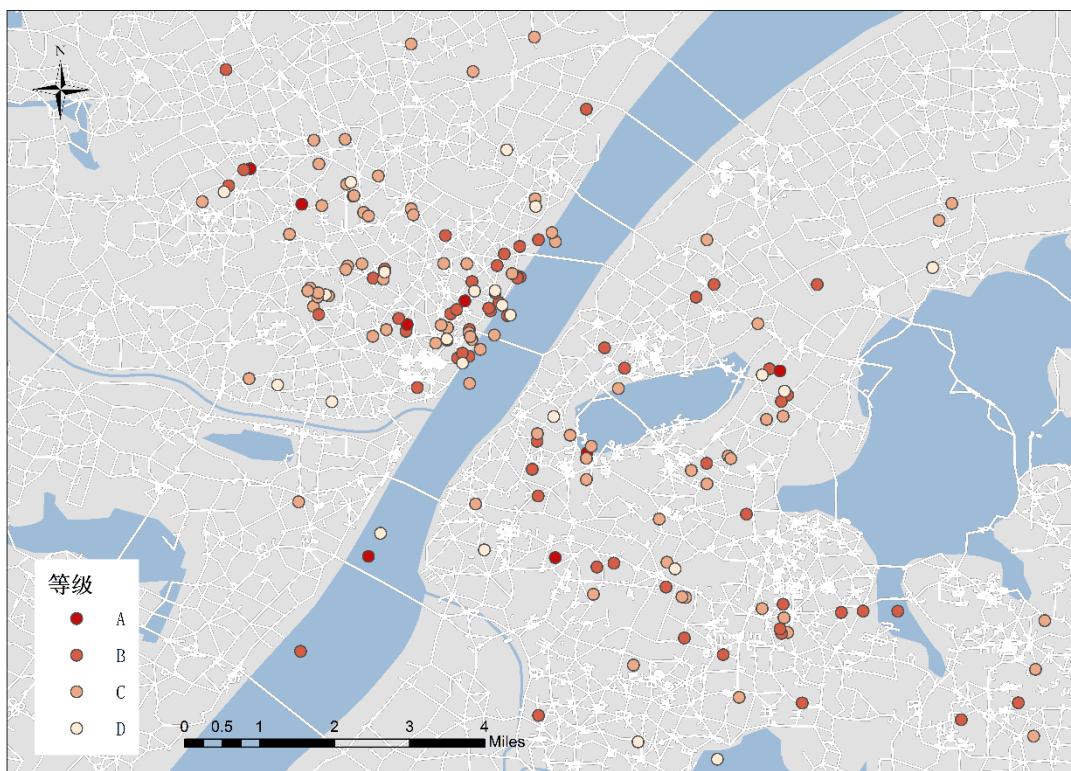


图 54 武市民宿适宜性评价结果图

### (三) 模型结果分析

## 第十一章 武市民宿适应性评价

本模型是从消费者的需求出发，平衡专家意见和问卷调查结果进行构建的。总体来看，接近四成的民宿等级达到了 B 等级及以上，没有 E 等级的民宿，说明武汉市民宿的适宜性水平整体情况较好，但 C 等级的民宿数量达到了 50%，说明还存在一定的提升空间。



图 55 武市民宿各准则层等级分布图

从实用性来看，武汉市民宿的平均得分为 4.55，有 45.7% 的民宿得分在 5 分以上。结合图 55 可以看出，评分主要集中在 2-6 分的区间内，由此可以看出，在实用性方面仍存在较大的提升空间，环境设施及服务水平进一步提高，提高民宿的性价比。

从可步性来看，武汉市民宿的平均得分为 7.21，整体情况较好。结合图 55 可以看出，评分主要集中在 6-10 分的区间内，说明武汉市民宿的地理位置优越，交通通达情况较好，基本满足消费者的出行需求。

从安全应急性来看，武汉市民宿的平均得分为 5.09，有接近一半的民宿评分达到 5 分及以上。结合图 55 可以看出，评分主要集中在 2-8 分的区间内，说明武汉市民宿在安全应急性方面存在一定的不足，但基本上满足了远离重大危险源和与医院之间可达性较好的要求。



## 第十二章 結論与建议

### 十二、结论和建议

#### (一) 调查结论

##### 1. 武汉市民宿供给格局

###### (1) 近四成的武汉市民宿发展适宜性达到 6 分及以上

AHP 模型评估结果表明近四成的民宿发展适宜性达到 6 分以上，即实惠性、可步性和安全应急性整体较均衡或在某一方面特别突出，对于在某一方面达到 8 分及以上的民宿，需要在其他方面适当改进。根据 AHP 评价结果，还可以发现，4-6 分的民宿达到了 50%，说明武汉市现有民宿综合实惠性、可步性和安全应急性三个方面来看，基本满足了需求，但在某些方面需要着力改善，是接下来可以重点提高的对象。

###### (2) 武汉市民宿交通通达情况好，实惠性不高。

在可步性方面，平均得分为 7.21，近一半的民宿评分主要集中在 6-10 分的区间内，说明武汉市现有民宿地理位置优越，距商圈、景点近，交通通达情况好，距公交、地铁站近，且道路连通性好，基本满足消费者出行需求。实惠性不高，主要体现在价格与环境设施及服务水平的不匹配，要有针对性的进行改善。

##### 2. 武汉市民宿消费群体特征

###### (1) 消费者特征明显，25-30 岁群体是民宿市场消费的主力军。

对民宿有较高选择倾向的特征有 25-30 岁，月消费额在 1000-4999 元之间，全日制学生或文职/办事人员，技术/研发人员、专业人士（如会计师、律师等），这一群体是消费的主力军。

###### (2) 高消费、中层年龄阶段、流动性较大的人群作为潜在消费者值得发展。

在未住过民宿的人群中有近 8 成表现出尝试民宿的意愿，是非常重要的潜在客户。在有较高外宿需求的群体中，没有统一的清洁标准、入住流程与服务相对不够完善是不想选择民宿的主要原因，分别占 29.5% 和 28.4%。可针对不同群体需求特征设置不同的活动，并完成有针对性的优化，进一步拓宽民宿客户群体。

##### 3. 民宿消费者体验及需求情况

###### (1) 年轻群体更注重环境氛围感、人情味、价格合理。

民宿温馨的环境是绝大部分消费者原因选择民宿的原因，达到了 26%。25 岁以下群体认为民宿相比于酒店较为合理的价格也是民宿的显著的优势，丰富的设备也会让他们更愿意选择民宿。

### （2）中、老年群体更关注民宿带来的本地风土人情。

30-40 岁群体则更关注民宿带来的本地风土人情，这也是 40 岁以上群体选择民宿的最主要的原因。民宿经营者应该继续发展这些优点，可以在很大程度上稳固民宿的客户群体，结合宣传力度吸引更多的客户。

### （3）消费者不愿意住民宿主要是卫生问题、安全性问题、规范性问题。

我们对客户相比于酒店不愿意选择民宿的原因进行分析，结果表明主要的原因有以下三点：①没有专门的安保人员，安全性得不到保障②没有统一的清洁标准③没有餐厅等配套服务。大部分消费者群体主要由于安全性、卫生性、规范性的原因不愿意选择民宿。如果民宿经营者应在一定程度上提高这几方面的水平，可以在很大程度上增加民宿的客户群体，进一步拓展民宿市场。

## （二）优化策略

根据本次调研的民宿市场情况及消费者的需求情况，我们分别针对现有客户群体和潜在客户群体对武汉市民宿产业给出以下建议。

### 1、针对现有客户群体的建议

#### （1）加强民宿优势点建设，在交通便利、环境氛围、价格实惠方面进一步提高。

处在地铁、公交站附近的民宿更应发挥自身地理位置的优势，在入户指示方面更加清晰地指示；而处于不利位置，较为偏远的民宿房东可以设置专车接送等方法来弥补交通方面的缺陷。环境氛围方面民宿经营者应更加注重民宿细节装饰方面的布置以及相关设施的完善。在价格方面可以面向学生等群体进行阶段性的优惠活动，吸引更多的消费者。

#### （2）充分利用网络资源，对不同群体进行不同营销方案

针对 25-30 岁的群体，采用自媒体博主视频、写图文推广的方式着重展现温馨度及真实的体验感，附上附近游玩的简短小攻略，并实现订单直连，简化入住流程。

#### （3）充分结合自身优势特点打造品牌标签，锁住客源。

民宿在整体的氛围感、与本土文化结合方面表现出了较明显的特点，在日后

的宣传和推广中，更应该充分发扬这一特点并转换为较为独特的优势。例如在装饰风格、配套设施上更注意细节方面的挖掘，注重与本土文化的结合，充分做到“扬长”，把“氛围感”、“人情味”打造成民宿独有的产业特色，使用“免费住”等营销方案锁住客源。

### 2、针对潜在客户群体的建议

#### （1）建立统一的清洁标准，提高整体卫生水平。

清洁卫生情况不够理想成为许多客户不愿意选择民宿的主要原因，卫生情况能够加强也是绝大部分消费者的期待。民宿行业应该加快对卫生的整顿，加速清洁标准化进程，提高民宿基本设施用品的清洁程度，从整体到局部实现清洁问题的改进。此外，在线上预定方面，需要在订购页面中标明清洁时间与消毒频率等，将卫生水平公开透明化。

#### （2）将入住流程规范化、透明化，提供完整的服务体系，提供购买凭证。

民宿的入住流程相比于酒店缺乏一定的标准化，在民宿产业逐步壮大的当下，更应以平台为单位完成民宿从预定到入住到退房中每一个环节的规范化。且将这一流程以清晰的方式展现给消费者，从而实现服务流程的透明化，增强用户对民宿入住体系的信任度及支持度。

#### （3）加强安全措施保障，最大程度上满足用户的安全需求。

许多民宿由于所在小区原因无法对安保人员的保障进行落实，但可以从其他方面提高民宿的安全指数。例如提高门锁的安全性，实行入住方式的闭环管理，防止民宿入住信息的泄露。同时经营者也应注意民宿的内部安全性，如用电安全、以及住户隐私安全的落实。

#### （4）针对潜在客户群体关注点，着重“品质”、“聚会”等话题，打造故事性视频进行引流。

针对中层年龄、中消费群体等潜在客户群体，着重“品质”、“聚会”等话题，用打造故事的方式录制视频，发布在短视频平台等新媒体中，从“用途”角度吸引顾客。



## 第十三章 附录

### 十三、附录

#### 附录一：参考文献

- [1] 莫赞,赵琦智,罗敏瑶.在线评论对消费者购买决策的影响研究[J].南宁师范大学学报(哲学社会科学版),2021,42(03):36-44.DOI:10.16601/j.cnki.issn2096-7349.2021.03.004.
- [2] LANIER P, CAPLES D, COOK H. How big is small? A study of bed& breakfasts, country inns, and small hotels[J]. Cornell Hotel and Restaurant Administration Quarterly, 2000, 41(5):90-95.
- [3] Xia B, Dong S, Ba D, et al. Research on the spatial differentiation and driving factors of tourism enterprises' efficiency: Chinese scenic spots, travel agencies, and hotels. Sustainability, 2018, 10(4): 901.
- [4] 谢萍,詹坤婷,常洁.成都民宿空间分布特征及影响因素[J].国土资源科技管理,2021,38(04):108-124.
- [5] 刘大均.成都市民宿空间分布特征及影响因素研究[J].西华师范大学学报(自然科学版),2018,39(01):89-93.
- [6] 龙飞,刘家明,朱鹤,李涛.长三角地区民宿的空间分布及影响因素[J].地理研究,2019,38(04):950-960.
- [7] 王文鑫. 重庆市民宿空间分布特征及影响因素研究[D].西南大学,2021.
- [8] 明跃强. 全域旅游视角下的阳朔民宿空间分布特征及影响因素研究[D].桂林理工大学,2021.
- [9] 刘贤菊. 重庆民宿消费者市场研究[D].重庆师范大学,2017.
- [10] 吴欢欢,李天宇,符琳蓉,陆林.黄山市民宿空间集聚特征及影响因素研究[J].资源开发与市场,2021,37(08):914-920.
- [11] 沈士琨,史春云.苏南、浙北地区民宿空间分布特征及其影响因素[J].热带地理,2022,42(01).

## 附录二：调查问卷

随着近期国内旅行比例的大幅提高，补偿式出游需求强劲，民宿恢复迅速，再次迎来红利期，国内民宿市场发展潜力巨大。同时，消费者的消费观点更加理智成熟，对民宿品质和品牌的需求与意识进一步提升。本问卷旨在对目前民宿市场现状进行一些调查和探究，进而为民宿行业的发展提供有效建议。希望您能在百忙之中抽空真实地填写本份问卷(填写时间预计为5分钟)，感谢您的配合！

本调查问卷信息将会严格保密，请您放心填写。联系方式(QQ)：1290833241

1. 您的性别

男 女

2. 您的年龄

25岁以下 25—30岁 30岁—40岁 40岁以上

3. 您的职业

全日制学生

生产人员

销售人员

市场/公关人员

客服人员

行政后勤人员

人力资源

财务/审计人员

文职/办事人员

技术/研发人员管理人员

教师

顾问/咨询

专业人士（如会计师，律师，建筑师，医护人员，记者等）

其他

4. 您每个月的月消费大约为

1000元以下

1000—2499元

## 第十三章 附录

2500—3999 元

4000—5499 元

5500—7000 元

7000—9999 元

10000—14999 元

15000—19999 元

30000 元以上

5. 您在外住宿（民宿、酒店等）的目的

旅游   出差   聚会   其他

6. 您每年在外住宿（民宿、酒店等）的频率

5 次以下   5-10 次   10 次以上

7. 您是否愿意选择民宿？

是——【转第 8 题】   否——【转第 9 题】

8. 您愿意选择民宿的原因（最多选择三项）

价格更为合理

能感受到当地的风土人情

环境更加温馨，有家的温度

设备更加丰富（如：有投影仪、厨房等）

周围的风景更好

相比于酒店，闲杂人等更少

其他

9. 相比于酒店，您不愿意选择民宿的原因（最多选择三项）

没有专门的安保人员，安全性得不到保障

没有统一的清洁标准

没有餐厅等配套服务

酒店的入住流程与服务较完善

酒店的位置更好找

其他

10. 你是否住过民宿

是——【转 A 部分】

否——【转 B 部分】

### 【A 部分】

11. 你住过的民宿普遍价位是多少

人均 100 以下

人均 100—149

人均 150—199

人均 200—300

人均 300 以上

12. 您每年住民宿的频率

5 次以下    5-10 次    10 次以上

13. 您获取民宿信息的渠道有哪些？

小红书、微博、抖音等社交平台

美团、爱彼迎、携程、去哪儿旅行等

亲友推荐

电视、广告等

其他

14. 您选择过什么类型的民宿

LOFT

公寓

别墅

15. 您所选择过的民宿中，到最近地铁站/公交站的距离大约为

步行 5min 以内

步行 5—10min

步行 10—20min

步行 20min 以上

16. 您所选择的民宿中，到最近商圈/景点的距离大约为？

步行 5min 以内

## 第十三章 附录

步行 5—10min

步行 10—20min

步行 20min 以上

17.除了基本设施外，以下哪些民宿设施对你来说很重要

冰箱

投影仪

电视

浴缸

落地窗

厨具

停车位

洗衣机

烘干机

麻将机

其他

都无所谓

18.您喜欢怎样的民宿风格

简约商务

中式

欧式

ins 风

日式

其他

19.在选择住所时可能考虑的因素时，您认为的重要程度是

价格

1    2    3    4    5    6    7    8    9    10

地理位置（距离市中心或景点）

1    2    3    4    5    6    7    8    9    10

交通通达程度（距离公交站、地铁站）

○1 ○2 ○3 ○4 ○5 ○6 ○7 ○8 ○9 ○10

环境设施及服务

○1 ○2 ○3 ○4 ○5 ○6 ○7 ○8 ○9 ○10

安全应急性（到医院的距离）

○1 ○2 ○3 ○4 ○5 ○6 ○7 ○8 ○9 ○10

20.您对现有民宿的整体满意程度为

很不满意

不满意

一般

满意

很满意

21.请您对以下民宿各方面满意程度进行打分

人均价格

优惠与折扣

到商圈的距离

到景点的距离

位置好找

出行方式便利性

出行方式多样性

卫生

基础设备齐全程度

其他设备丰富度

房东服务态度

住房私密性

周边治安问题

周边应急设施（如医院、药店）

22. 您对现有民宿的建议

\_\_\_\_\_ (填空题)

## 第十三章 附录

### 【B 部分】

23.您获取酒店信息的渠道有那些

- 小红书、微博、抖音等社交平台
- 美团、爱彼迎、携程、去哪儿旅行等
- 亲友推荐
- 电视、广告等
- 其他

24.您住过酒店的普遍价位是多少

- 人均 100 以下
- 人均 100—149
- 人均 150—199
- 人均 200—300
- 人均 300 以上

25.您所选择过的酒店中，到最近地铁站/公交站的距离大约为

- 步行 5min 以内
- 步行 5—10min
- 步行 10—20min
- 步行 20min 以上

26.您所选择过的酒店中，到最近地铁站/公交站的距离大约为

- 步行 5min 以内
- 步行 5—10min
- 步行 10—20min
- 步行 20min 以上

27.在选择住所时可能考虑的因素时，您认为的重要程度是

价格

1     2     3     4     5     6     7     8     9     10

地理位置（距离市中心或景点）

1     2     3     4     5     6     7     8     9     10

交通通达程度（距离公交站、地铁站）

1 2 3 4 5 6 7 8 9 10

□环境设施及服务

1 2 3 4 5 6 7 8 9 10

□安全应急性（到医院的距离）

1 2 3 4 5 6 7 8 9 10

28.您对民宿有什么期待

\_\_\_\_\_ (填空题)

我们的问卷到此结束，感谢您的参与！

## 第十三章 附录

### 附录三：问卷编码

变量序号	变量名	变量类型	取值范围	选项	对应题号	对应问题
1	Q1	数值型	1-2	1-男 2-女	0	性别
2	Q2	数值型	1-4	1- 25 岁以下 2- 25-30 岁 3- 30-40 岁	0	年龄
3	Q3	数值型	1-13	1-全日制学生 2-生产人员 3-销售人员 4-市场/公关人员 5-客服人员 6-行政后勤人员 7-财务/审计人员 8-文职/办事人员 9-技术/研发人员 10-管理人员 11-教师 12-专业人士(如会计师、 律师、医护人员、记者 等) 13-其他	0	职业
4	Q4	数值型	1-5	1- 1000 以下 2- 1000-2499 3- 2500-4999 4- 5000-9999 5- 10000 以上	0	月消费水平
5	Q5	数值型	1-4	1-旅游 2-出差 3-聚会 4-其他	0	在外住宿的 目的
6	Q6	数值型	1-3	1- 5 次以下 2- 5-10 次 3- 10 次以上	0	每年在外住 宿的频率

7	Q7	数值型	1-2	1-愿意 2-不愿意	0	是否愿意选择民宿
8	Q8	数值型	1-7	1-价格更为合理 2-能感受到当地的风土人情 3-环境更加温馨,有家的温度 4-设备更加丰富(如:有投影仪、厨房等) 5-周围的风景更好 6-相比于酒店,闲杂人等更少 7-其他	7	愿意选择民宿的原因
9	Q9	数值型	1-6	1-没有专门的安保人员,安全性得不到保障 2-没有统一的清洁标准 3-没有餐厅等配套服务 4-酒店的入住流程与服务较完善 5-酒店的位置更好找 6-其他	7	不愿意选择民宿的原因
10	Q10	数值型	1-2	1-是 2-否	8、9	是否住过民宿
11	Q11	数值型	1-5	1-人均 100 以下 2-人均 100—149 3-人均 150—199 4-人均 200—300 5-人均 300 以上	10	住过民宿的普遍价位
12	Q12	数值型	1-5	1-5 次以下 2-5-10 次 3-10 次以上	11	每年住民宿的频率
13	Q13	数值型	1-5	1-小红书、微博、抖音等社交平台 2-美团、爱彼迎、携程、去哪儿旅行等 3-亲友推荐 4-电视、广告等 5-其他	12	获取民宿信息的渠道

## 第十三章 附录

14	Q14	数值型	1-3	1-LOFT 2-公寓 3-别墅	13	民宿类型
15	Q15	数值型	1-4	1-步行 5min 以内 2-步行 5—10min 3-步行 10—20min 4-步行 20min 以上	14	民宿到最近地铁站/公交站的距离
16	Q16	数值型	1-4	1-步行 5min 以内 2-步行 5—10min 3-步行 10—20min 4-步行 20min 以上	15	民宿到最近商圈/景点的距离
17	Q17	数值型	1-12	1-冰箱 2-投影仪 3-电视 4-浴缸 5-落地窗 6-厨具 7-停车位 8-洗衣机 9-烘干机 10-麻将机 11-其他 12-都无所谓	16	认为重要的民宿设施
18	Q18	数值型	1-6	1-简约商务 2-中式 3-欧式 4-ins 风 5-日式 6-其他	17	喜欢的民宿风格
19	Q19.1	数值型	1-10	1→10: 非常不重要→非常重要	18	选择民宿时，认为民宿价格的重要程度
19	Q19.2	数值型	1-10	1→10: 非常不重要→非常重要	18	选择民宿时，认为民宿地理位置的重要程度
19	Q19.3	数值型	1-10	1→10: 非常不重要→非常重要	18	选择民宿时，认为民宿交通通达度的重要程度

<b>19</b>	Q19.4	数值型	1-10	1→10: 非常不重要→非常重要	18	选择民宿时，认为民宿环境设施及服务的重要程度
<b>19</b>	Q19.5	数值型	1-10	1→10: 非常不重要→非常重要	18	选择民宿时，认为民宿安全应急性的重要程度
<b>20</b>	Q20	数值型	1-5	1-很不满意 2-不满意 3-一般 4-满意 5-很满意	19	对现有民宿的满意程度
<b>21</b>	Q21.1	数值型	1-5	1-很不满意 2-不满意 3-一般 4-满意 5-很满意	20	对民宿人均价格的满意程度
<b>21</b>	Q21.2	数值型	1-5	1-很不满意 2-不满意 3-一般 4-满意 5-很满意	20	对民宿优惠与折扣的满意程度
<b>21</b>	Q21.3	数值型	1-5	1-很不满意 2-不满意 3-一般 4-满意 5-很满意	20	对民宿到商圈的距离的满意程度
<b>21</b>	Q21.4	数值型	1-5	1-很不满意 2-不满意 3-一般 4-满意 5-很满意	20	对民宿到景点的距离的满意程度
<b>21</b>	Q21.5	数值型	1-5	1-很不满意 2-不满意 3-一般 4-满意 5-很满意	20	对民宿位置好找的满意程度

## 第十三章 附录

21	Q21.6	数值型	1-5	1-很不满意 2-不满意 3-一般 4-满意 5-很满意	20	对民宿出行方式便利性的满意程度
21	Q21.7	数值型	1-5	1-很不满意 2-不满意 3-一般 4-满意 5-很满意	20	对民宿出行方式多样性的满意程度
21	Q21.8	数值型	1-5	1-很不满意 2-不满意 3-一般 4-满意 5-很满意	20	对民宿卫生的满意程度
21	Q21.9	数值型	1-5	1-很不满意 2-不满意 3-一般 4-满意 5-很满意	20	对民宿基础设备齐全程度的满意程度
21	Q21.10	数值型	1-5	1-很不满意 2-不满意 3-一般	20	对民宿其他设备丰富度的满意程度
21	Q21.11	数值型	1-5	1-很不满意 2-不满意 3-一般 4-满意 5-很满意	20	对民宿房东服务态度的满意程度
21	Q21.12	数值型	1-5	1-很不满意 2-不满意 3-一般 4-满意 5-很满意	20	对民宿住房私密性的满意程度
21	Q21.13	数值型	1-5	1-很不满意 2-不满意 3-一般 4-满意 5-很满意	20	对民宿周边治安问题的满意程度

21	Q21.14	数值型	1-5	1-很不满意 2-不满意 3-一般 4-满意 5-很满意	20	对民宿周边应急设施（如医院、药店）的满意程度
22	Q22	文本型			21	对现有民宿的建议
23	Q13	数值型	1-5	1-小红书、微博、抖音等社交平台 2-美团、爱彼迎、携程、去哪儿旅行等 3-亲友推荐 4-电视、广告等 5-其他	22	获取民宿信息的渠道
24	Q24	数值型	1-5	1-人均 100 以下 2-人均 100—149 3-人均 150—199 4-人均 200—300 5-人均 300 以上	23	住过民宿的普遍价位
25	Q24	数值型	1-4	1-步行 5min 以内 2-步行 5—10min 3-步行 10—20min 4-步行 20min 以上	25	酒店到最近地铁站/公交站的距离
26	Q24	数值型	1-4	1-步行 5min 以内 2-步行 5—10min 3-步行 10—20min 4-步行 20min 以上	25	酒店到最近商圈/景点的距离
27	Q27.1	数值型	1-10	1→10: 非常不重要→非常重要	26	选择酒店时，认为酒店价格的重要程度
27	Q27.2	数值型	1-10	1→10: 非常不重要→非常重要	26	选择酒店时，认为酒店地理位置的重要程度

## 第十三章 附录

27	Q27.3	数值型	1-10	1→10: 非常不重要→非常重要	26	选择酒店时，认为酒店交通通达程度的重要程度
27	Q27.4	数值型	1-10	1→10: 非常不重要→非常重要	26	选择酒店时，认为酒店环境设施及服务的重要程度
27	Q27.5	数值型	1-10	1→10: 非常不重要→非常重要	26	选择酒店时，认为酒店安全应急性的重要程度
28	Q28	文本型			27	对民宿的期待

## 附录四：二元选择模型变量表

变量名	变量符号	变量定义
性别	$sex$	$male = \begin{cases} 1 & \text{男} \\ 0 & \text{其他} \end{cases}$ 以女性作为基础类型
年龄	$age\_1$ $age\_2$ $age\_3$ $age\_4$	$age\_1 = \begin{cases} 1 & 25 \text{ 岁以下} \\ 0 & \text{其他} \end{cases}$ $age\_2 = \begin{cases} 1 & 25 - 30 \text{ 岁} \\ 0 & \text{其他} \end{cases}$ $age\_3 = \begin{cases} 1 & 30 - 40 \text{ 岁} \\ 0 & \text{其他} \end{cases}$ $age\_4 = \begin{cases} 1 & 40 \text{ 岁以上} \\ 0 & \text{其他} \end{cases}$ 以 40 岁以上作为基础类型
月消费水平	$cost\_1$ $cost\_2$ $cost\_3$ $cost\_4$ $cost\_5$	$cost\_1 = \begin{cases} 1 & 1000 \text{ 元以下} \\ 0 & \text{其他} \end{cases}$ $cost\_2 = \begin{cases} 1 & 1000 - 2499 \text{ 元} \\ 0 & \text{其他} \end{cases}$ $cost\_3 = \begin{cases} 1 & 2500 - 4999 \text{ 元} \\ 0 & \text{其他} \end{cases}$ $cost\_4 = \begin{cases} 1 & 5000 - 9999 \text{ 元} \\ 0 & \text{其他} \end{cases}$ $cost\_5 = \begin{cases} 1 & 10000 \text{ 元以上} \\ 0 & \text{其他} \end{cases}$ 以月消费 10000 元以上作为基础类型
职业	$profession\_a$ $profession\_b$ $profession\_c$ $profession\_d$ $profession\_e$ $profession\_f$	$profession\_a = \begin{cases} 1 & \text{全日制学生} \\ 0 & \text{其他} \end{cases}$ $profession\_b = \begin{cases} 1 & \text{生产人员} \\ 0 & \text{其他} \end{cases}$

## 第十三章 附录

	$profession_g$ $profession_h$ $profession_i$ $profession_j$ $profession_k$ $profession_l$ $profession_m$	$profession_c = \begin{cases} 1 & \text{销售人员} \\ 0 & \text{其他} \end{cases}$ $profession_d = \begin{cases} 1 & \text{市场/公关人员} \\ 0 & \text{其他} \end{cases}$ $profession_e = \begin{cases} 1 & \text{客服人员} \\ 0 & \text{其他} \end{cases}$ $profession_f = \begin{cases} 1 & \text{行政后勤人员} \\ 0 & \text{其他} \end{cases}$ $profession_g = \begin{cases} 1 & \text{财务/审计人员} \\ 0 & \text{其他} \end{cases}$ $profession_h = \begin{cases} 1 & \text{文职/办事人员} \\ 0 & \text{其他} \end{cases}$ $profession_i = \begin{cases} 1 & \text{技术/研发人员} \\ 0 & \text{其他} \end{cases}$ $profession_j = \begin{cases} 1 & \text{管理人员} \\ 0 & \text{其他} \end{cases}$ $profession_k = \begin{cases} 1 & \text{教师} \\ 0 & \text{其他} \end{cases}$ $profession_l = \begin{cases} 1 & \text{专业人士 (如} \\ & \text{会计师、律师、} \\ & \text{医护人员、记者等)} \\ 0 & \text{其他} \end{cases}$ $profession_m = \begin{cases} 1 & \text{聚会} \\ 0 & \text{其他} \end{cases}$
--	--	--

以其他作为基础类型

## 附录五：排序选择模型变量表

变量名	变量符号	变量定义
住民宿的次数	$Y$	$Y = \begin{cases} 1 & 5 \text{ 次以下} \\ 2 & 5 - 10 \text{ 次} \\ 3 & 10 \text{ 次以上} \end{cases}$
性别	$sex$	$male = \begin{cases} 1 & \text{男} \\ 0 & \text{其他} \end{cases}$ 以女性作为基础类型
年龄	$age\_1$ $age\_2$ $age\_3$ $age\_4$	$age\_1 = \begin{cases} 1 & 25 \text{ 岁以下} \\ 0 & \text{其他} \end{cases}$ $age\_2 = \begin{cases} 1 & 25 - 30 \text{ 岁} \\ 0 & \text{其他} \end{cases}$ $age\_3 = \begin{cases} 1 & 30 - 40 \text{ 岁} \\ 0 & \text{其他} \end{cases}$ $age\_4 = \begin{cases} 1 & 40 \text{ 岁以上} \\ 0 & \text{其他} \end{cases}$ 以 40 岁以上作为基础类型
月消费水平	$cost\_1$ $cost\_2$ $cost\_3$ $cost\_4$ $cost\_5$	$cost\_1 = \begin{cases} 1 & 1000 \text{ 元以下} \\ 0 & \text{其他} \end{cases}$ $cost\_2 = \begin{cases} 1 & 1000 - 2499 \text{ 元} \\ 0 & \text{其他} \end{cases}$ $cost\_3 = \begin{cases} 1 & 2500 - 4999 \text{ 元} \\ 0 & \text{其他} \end{cases}$ $cost\_4 = \begin{cases} 1 & 5000 - 9999 \text{ 元} \\ 0 & \text{其他} \end{cases}$ $cost\_5 = \begin{cases} 1 & 10000 \text{ 元以上} \\ 0 & \text{其他} \end{cases}$ 以月消费 10000 元以上作为基础类型

## 第十三章 附录

在外住宿原因	<i>toor</i> <i>business</i> <i>party</i>	$toor = \begin{cases} 1 & \text{旅游} \\ 0 & \text{其他} \end{cases}$ $business = \begin{cases} 1 & \text{出差} \\ 0 & \text{其他} \end{cases}$ $party = \begin{cases} 1 & \text{聚会} \\ 0 & \text{其他} \end{cases}$ 以其他作为基础类型
--------	--	---