**《在线民宿房源五星评级影响因素分析》**

1. **背景介绍**

随着我国经济的发展和人民生活水平的提高，旅游从少数人消费得起的奢侈品变为大众喜爱且有能力负担的消费品。出行人数增多，旅游消费急剧增长，与旅游业息息相关的住宿业也迎来了新的行业搅局者：在线民宿。所谓的在线民宿就是以线上平台运营与线下闲置房源相结合为运营模式，为游客提供住宿的接待场所。在线民宿与传统酒店的主要差异在于：（1）房源分散，单点房源量较少；（2）产品个性化，经营主体多元化；（3）依赖互联网+模式。由于在线民宿具有性价比高、预订便捷、选择多样的优点，因此倍受消费者的青睐

中国的在线民宿行业萌芽于2011年。在互联网浪潮的席卷下，途家、蚂蚁短租等在线民宿预订平台成为首批开拓者，2016年Airbnb作为老牌劲旅强势崛起，随后行业经历了野蛮生长的蜜月期。如今，在资本、市场和技术的多方面因素助力下，中国的在线民宿行业欣欣向荣，根据数据显示，2018年，中国在线民宿房源数量从2017年的81.5万增长到2018年的107.2万，而在线民宿线上交易额也从2017年的78.1亿元增长到2018年的127.9亿元，预计2019年全年将突破200亿元大关。与此同时，截止2019年第一季度，途家平台的房源渗透率最高，达到了69%，环比增长2.4%，榛果民宿以43%的房源渗透率紧随其后，在线预订平台房源数量得到了整体提升。

虽然行业整体形势大好，但是依然存在一些问题与挑战。主要问题是现有房东群体专业性有待提升。对于在线民宿而言，平台方主要是为房源提供技术加持，而房东才是房源经营的主体。根据房东经营方式数据显示，截止2019年第一季度，在线民宿房东经营方式仍然以个人/家庭为主，专业公司经营比例仅占7%。而房东专兼职对比的数据也表明，现有房东群体专职比例虽然较2018年略有上升，但是整体占比仍然不到四成。个人/家庭、兼职为主的经营模式意味着房东群体缺乏相关的住宿管理经验，而且没有足够的时间进行打理和营销。这种专业性的缺失会导致两个直接后果：第一房源硬性条件不完善。部分房东仅仅提供住所，而对房间数量、容纳人数及厨房配套等硬件设施并未多加考虑。这种不以消费者为中心的经营方式无法有效挽留消费者。第二营销策略不合理。如果说房源硬性条件是硬实力，那么房东的营销策略则是软实力。随着消费者消费习惯向移动端迁移，预订平台上的用户评价、房东的优质等级以及打折活动等都会被纳入消费考量范围，倘若房东不注重平台的营销策略，就会在赛道上被拥有同等硬件质量但更善于平台营销的竞争者远远甩开。

由此可见，对于在线民宿行业而言，如何提高房东专业性就变成了一个重要的问题，这主要表现在如何改善房源硬性条件，以及如何优化平台营销策略。为此，平台需要了解房源硬性条件的优劣与营销策略的不同将会如何影响房源的评分指标。为了能够科学地回答该问题，收集了某在线民宿平台北京地区的475条房源信息。有针对性地提取了关于：（1）房源硬性条件（包括房源配置和所处地段）和（2）平台营销能力（包括房东展示和用户评价）相关的指标。具体而言，在房源硬性条件方面，采集了房源类型、容纳数量、厨房配套、所处城区、地铁配套、与天安门的距离，共6个指标。在平台营销能力方面，采集了性价比等级、房东等级、评论数、标题长度、价格、打折类型，共6个指标。通过这些指标，希望能够对在线民宿房源的硬性条件与平台营销能力有所描述，进而通过定序回归分析的方法建立房源评分与房源配置、地段、房东展示、用户评价之间的相关性，从而为改善房源硬性条件，改进平台营销能力提供建议与参考，详细分析如下。

1. **数据来源与指标设计**

首先，考虑什么指标能够衡量房源的硬性条件和平台营销能力。由于消费者是房源的直接体验者，因此消费者的反馈是一个最为恰当的指标。关键问题是，如何量化消费者的反馈？在线民宿预订平台的房源评分。房源评分指标是平台通过收集历史消费者对于房源的评分，通过特定算法得出的综合性评价指标，取值范围在0分到5分之间。对该评分作直方图（见图2-1），从中可以发现24%的样本得分为5分，这说明通过提升房源质量来达到消费者的需求从而获得高分的目标是能够实现的。同时，也提示我们关心什么样的房源硬性条件和平台营销能力可以得到5分的满分。为此，定义房源评分等级Y=1（如果房源评分为5分），否则Y=0。这就是本案例的因变量。

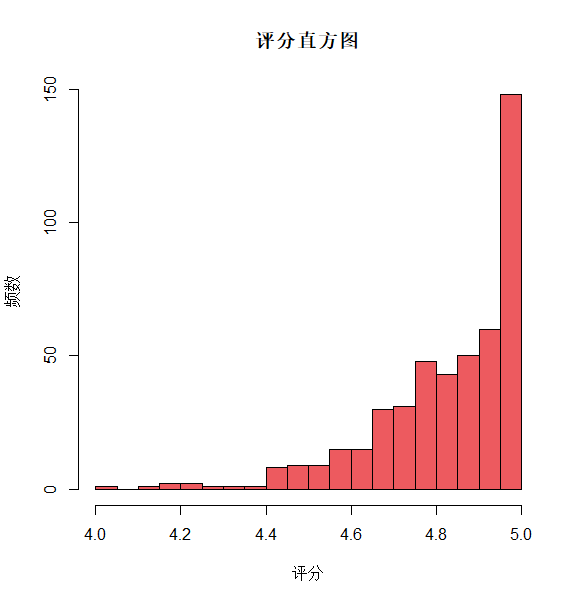


图2-1 某在线民宿平台北京地区房源评分直方图

确定了因变量之后，本案例再考虑解释变量。由于案例的研究问题是如何提升房源的硬性条件和平台营销能力，因此解释变量的挑选从这两方面着手。根据日常对房源的评价经验，刻画房源的硬性条件特征可以有很多指标或者标签。例如，可以考虑房源的配置，主要包括房源的类型（如独立和整套）、容纳数量和厨房配套；还可以考虑房源的所处地段，包括该房源的所处城区、地铁配套、与核心区域（天安门）的距离。针对平台营销能力，通过对该平台的要素观察，能够反映平台营销能力的指标主要有两个方面：房东展示和用户评价。房东展示主要包括标题长度、房源价格和打折类型（不打折、有条件打折、直接打折），而用户评价包括性价比等级、房东等级和评论数。综上所述，以上指标可以帮助更好地描述房源的硬性条件和平台营销能力，而这些指标都可以成为解释性变量；见表2-1：

表2-1： 案例数据说明表



1. **描述性分析**

在对房源评分等级的影响因素进行模型探究之前，首先对各变量进行描述性分析，以初步判断评分等级的影响因素，为后续研究做铺垫。

1. 因变量：房源评分等级

在本案例中，共有五星房源112间，非五星房源363间，全五星好评的房源数量相对较少，可见打造一套“完美民宿”还是具有一定难度。

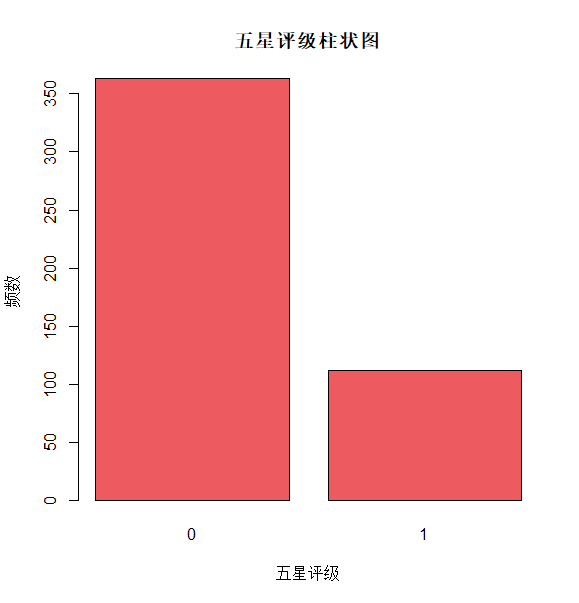


图3-1：五星评级房源的分布情况

1. 自变量：房源硬性条件

从图3-2可以看出，在北京地区可容纳2人的房源数量最多，其余容纳数量的房源数量较少。针对二人房源数量远远超过其他类型房源的现象，初步猜想，第一是市场需求因素，在线预订民宿的消费者大多为游客，两人结伴而行的形式较为常见，房源需求端影响供给端；第二是房价因素，北京作为首都，寸土寸金，容纳人数3人及以上的房源需要较大的占地面积，对于房东来说，成本相对较高。

此外，从不同容纳数量的房源五星占比来看，单人房源占比27%最高，二人房源以25%的比例紧随其后，其余容纳类型占比均在20%上下波动。随着房客数量的增加，五星房源的比例呈下降趋势，这可能是由于房客数量的增加带来了需求的增加，从而导致消费者满意度降低。比如，单人房源可能仅仅需要满足基本的起居需求，而四人房源还需要考虑娱乐端（电脑、电视、投影仪等）的需求。其次，房客数量较多，个人评判标准各异，众口难调，因此容纳人数较多的房源达成五星房源的门槛较高。

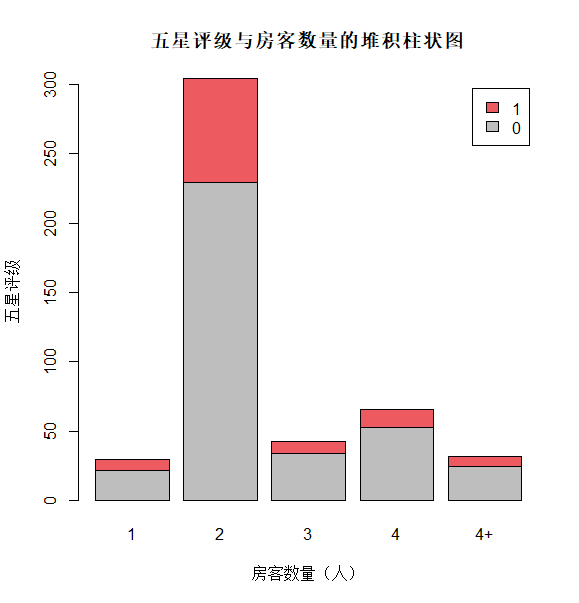


图3-2：五星评级与房源容纳数量的分布情况

从图3-3的房源类型分布情况可以看出，整套类型的房源数量最多，合住类型的房源数量极少，仅有13套。此类现象可能与市场竞争有关，合住类型的房源在市场上面临着青年旅馆这一强势对手，而且青年旅馆与在线民宿相比成本更低，消费者的需求也更低，因此在在线民宿行业中出租合住房源可能不是一个合适的选择。

从所处城区的房源分布情况来看，北京各城区的房源数量差异较大。总的来说可以分为三个梯队，第一梯队为朝阳区；第二梯队包括东城区和海淀区；第三梯队包括西城区和丰台区。初步猜想，这可能与各城区的旅游资源有关，朝阳、东城和海淀集中了北京市的大多数知名旅游景点，西城和丰台旅游资源相对匮乏，而且距离市中心较远。

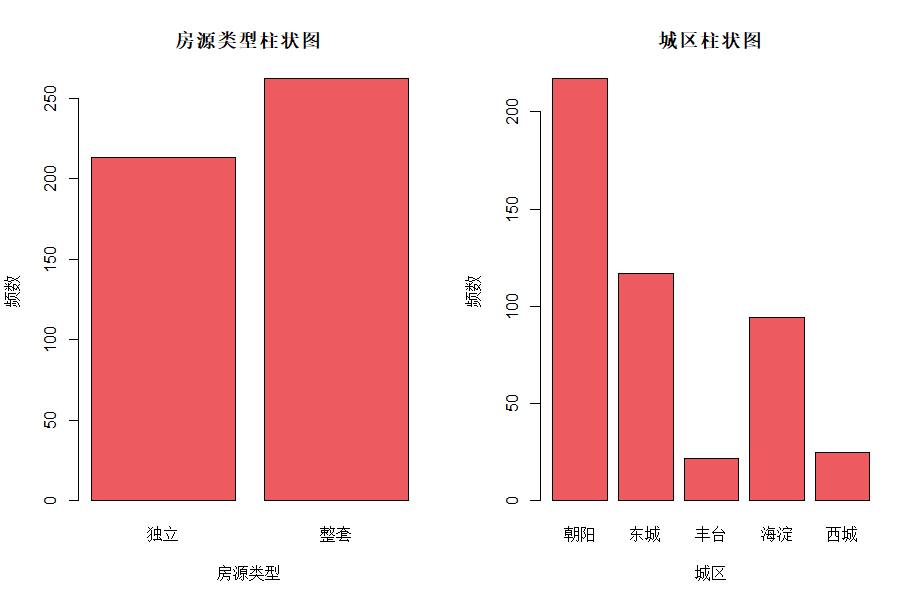


图3-3：房源类型与所处城区的分布情况

图3-4描述了在是否拥有地铁配套的状况下五星房源的概率分布，从中可以看到，在五星房源中，拥有地铁配套的占比远小于其在非五星房源中的占比，这说明地铁配套因素可能在某种程度上影响了五星评级。

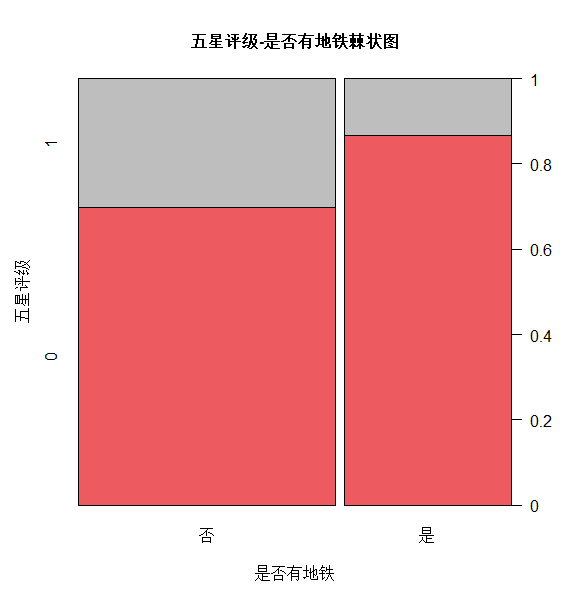


图3-4：五星评级-地铁指标棘状图

1. 自变量：平台营销能力

从图3-5中可以看到，在房源的标题长度上，整体呈左偏分布，五成房源的标题长度达到40字以上，说明该平台的房东更倾向于详细描述自己的房屋，以此让消费者得到更多信息；在房源的价格上，整体呈右偏分布，八成房源的价格在500元以下，比较符合大多数消费者的消费能力；在打折类型上，有条件打折的房源最多，相较于不打折和直接打折，这种“中庸”的打折之道在有效揽客和控制成本之间达到了平衡。

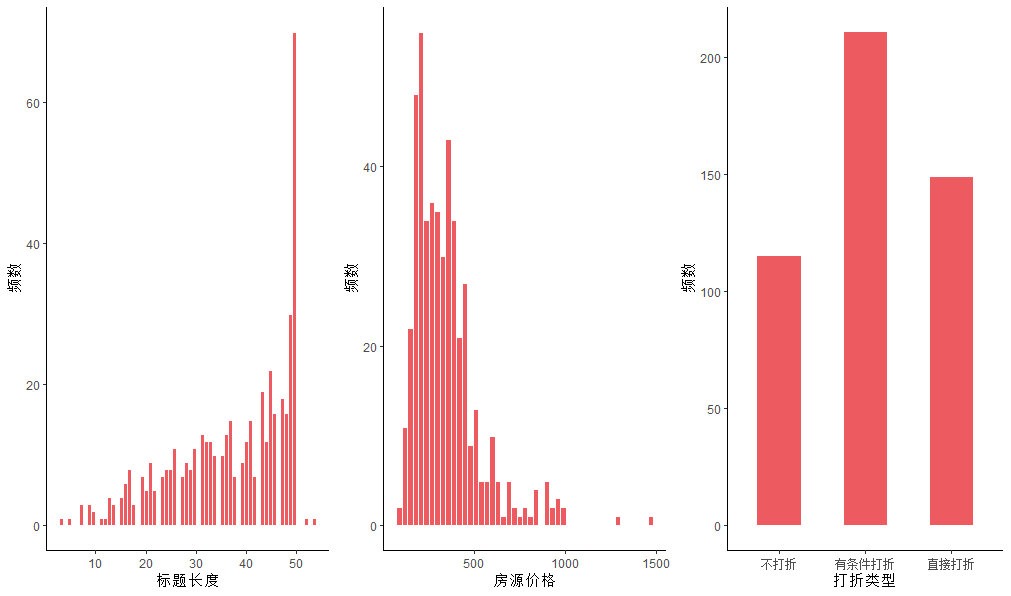


图3-5：房源平台营销能力（房东展示）变量分布图

从图3-6中可以看到，在房东等级上，房东等级为超级房东的房源有278套，占比58.53%，说明该平台半数以上的房东得到了消费者一定程度的认可；在性价比等级上，属于高性价比房源的共有110套，占比23.16%，说明消费者对于性价比等级的要求更加严苛；在评论数上，整体呈右偏分布，近五成的评论在40字以下，符合消费者日常的评论习惯。

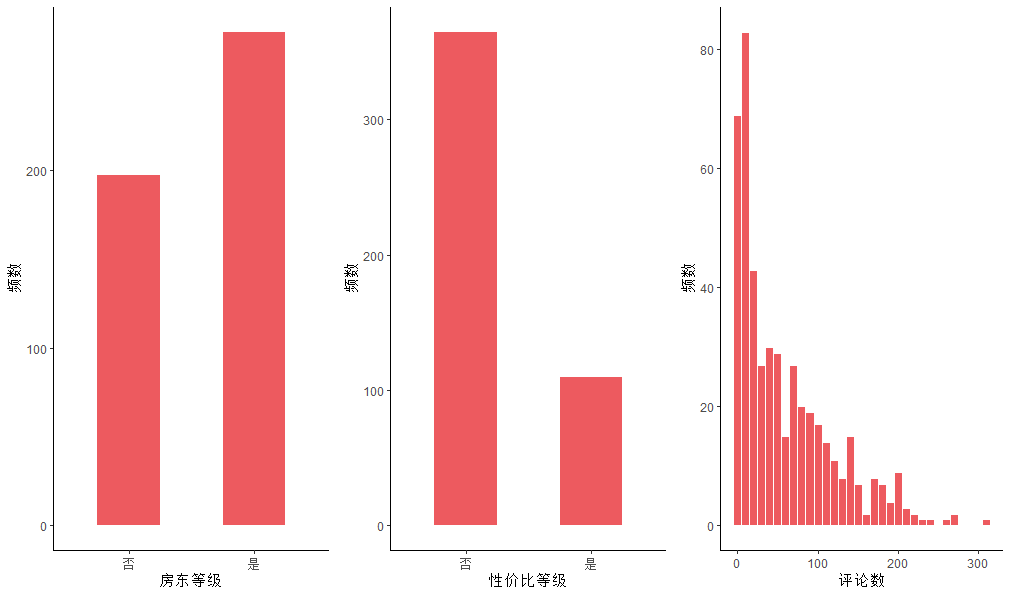


图3-6：房源平台营销能力（用户评价）变量分布图

在图3-7中可以看到，五星房源的平台评论数很少，一方面可能是因为五星房源大多上线时间较短，交易记录较少；另一方面是随着入住人数的增多，较难继续保持全五星好评。此外，可以看到评论数与五星评级存在一定的负相关。

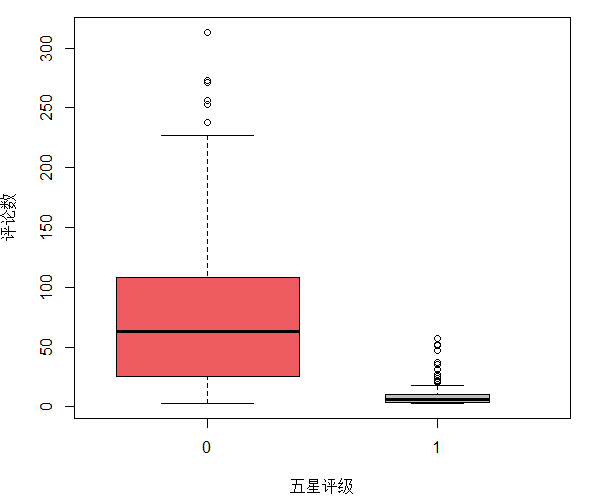


图3-7：五星评级评论数量对比箱线图

从图3-8中可以看到，在性价比指标上，五星房源中高性价比的占比很小，说明性价比并无法完全让房客满意。在打折类型指标上，五星房源中不打折类型的占比最少，直接打折的占比最多，说明折扣对于房客的满意度影响较大，这些初步的结论都比较符合大众的预期。

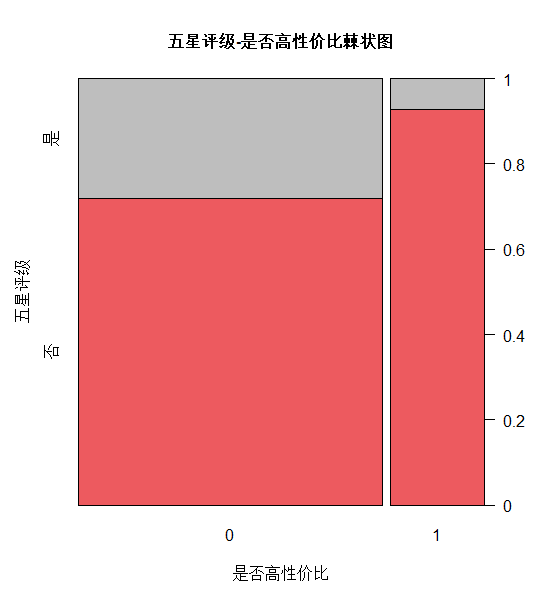
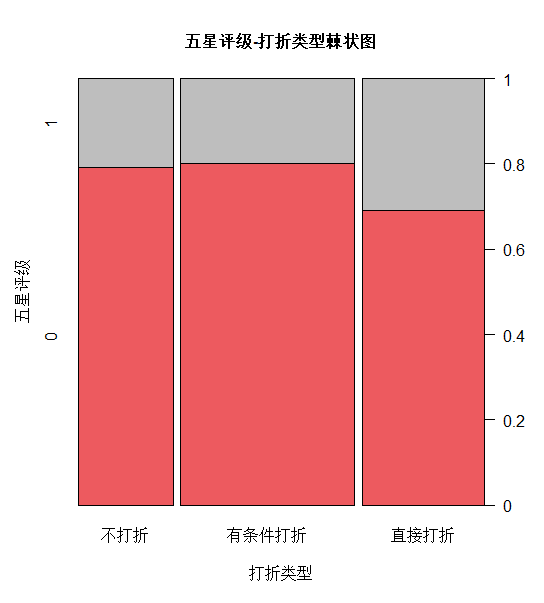


图3-8：五星评级-性价比等级、打折类型的棘状图

综上，通过对本案例数据的描述性分析，可以推测：对房源评价等级产生影响的因素包括：房源硬性条件（房源类型、容纳数量）和平台营销能力（评论数量、打折类型）；从影响作用来看，硬性条件因素更为明显一些。

1. **模型分析**

为了更深入地分析各因素对房源评价等级的影响，本案例将建立房源评价等级关于硬性条件因素和平台营销能力因素的逻辑回归模型，使用定量化的方式更为精细地刻画两方面因素的影响大小，并且使用该模型来预测房源评价等级。具体而言，对数据建立逻辑回归模型，为得到一个尽量简单同时又具有良好预测能力的模型，采用AIC准则进行逐步回归，结果如表4-1所示。经测算，本模型的AUC高达93.4%。这说明本模型的预测精度非常高。详细分析结果如下：

表4-1： 逻辑回归结果



从表4-1中可以看出，模型整体是高度显著的，AIC值为273.95。在控制其他因素不变时，可以得到以下结论：（1）在房源配置方面，相较于整套类型，独立类型的房源更容易成为五星房源；这说明在线民宿的客户主要是年轻群体，这类群体注重私密性但同时又乐意交友，而合住类型恰好能够满足私密与交友的双重需求。（2）在房源地段方面，所处城区的差异对民宿的五星评级来说并不显著，而拥有地铁配套、与天安门距离越远的房源更不易成为五星房源；在北京便利的公共交通网下，各大城区地理上的距离劣势被抹去，与此同时，由于共享单车的出现弥补了从住处到地铁站的交通空白，民宿的地铁配套设施变得不再重要，相反还可能因为过于嘈杂而被抛弃。鉴于天安门在中国人民心中拥有特殊的历史地位，与天安门的距离近的房源拥有了情怀加分。（3）在房东展示方面，房源价格越高，房源评分等级越高；这说明了消费者在某种程度上还是偏向“一分价钱一分货”的观念，对于价格较低的房源容易先入为主地预设房源质量较差从而影响最后打分。（4）在用户评价方面，超赞房东更容易经营出五星房源，而房源的高性价比指标和评论数在某种程度上拖累了房源评级；这说明了在消费者的打分标准中，房东的服务素养是较为重要的，情感上的愉悦能够在一定程度上弥补现实房源的缺点。此外，高性价比指标和评论数的结论说明当下消费者更加理性和客观，不能单纯将性价比与房源质量挂钩，发表评论不仅仅是出于赞赏，也可能是批评与建议。

1. **结论与建议**

本案例对于某在线民宿预订平台北京地区的房源数据进行统计分析，得出结论：影响北京地区在线民宿房源评价等级的主要因素有硬性条件因素和平台营销能力因素。在硬性条件方面，主要包括房源类型、地铁配套、与天安门的距离。与天安门距离较近，远离地铁口，出租独立房间的房源更加容易获得五星评级；在平台营销能力方面，主要包括房源价格、是否高性价比、是否超赞房东、评论数。房源价格较高，房东属性超赞且拥有较少评论数的非高性价比房源更容易获得五星评级。

根据模型结果提出以下建议，第一，合理规划房源类型。房东可以考虑将整套房子拆分成独立房间进行出租，从而提高房客的满意度；第二，避免地铁配套误区。房东在考虑交通便利性的同时，应更注重于地铁站保持适当距离防止噪音干扰，同时尽可能选在天安门附近的核心区域，方便房客获得较好的旅行体验；第三，重视用户评价。房东切勿片面追求评论数量，评论数较多可能一定程度上反映了房客给出差评的激动情绪。因此房东在经营过程中需要注意服务态度和形象，及时解决房客在评论中提及的问题，从而提高满意度；第四，摆脱性价比束缚。现在的年轻房客经济能力较好，在选择房源时会更加注重品质高低，而非单纯追求性价比，所以房东对于民宿的设计中仍需以品质优先，适当的高价优质房源会吸引更多房客。

由于房源评价等级的影响因素有很多，因此在未来的研究中可以考虑在模型中加入更多因素，比如评论情感分析（正面、负面）、房东资料完整度等。另外，若要将模型推广到其他城市，还要进一步考虑城市特有因素（如：在旅游城市是否为海景房等）。