



在线民宿房源五星评级影响因素分析

狗熊会 | 精品案例 主讲人：

目录

Contents

第一章

背景介绍

第二章

数据说明

第三章

描述统计

第四章

建模分析

第五章

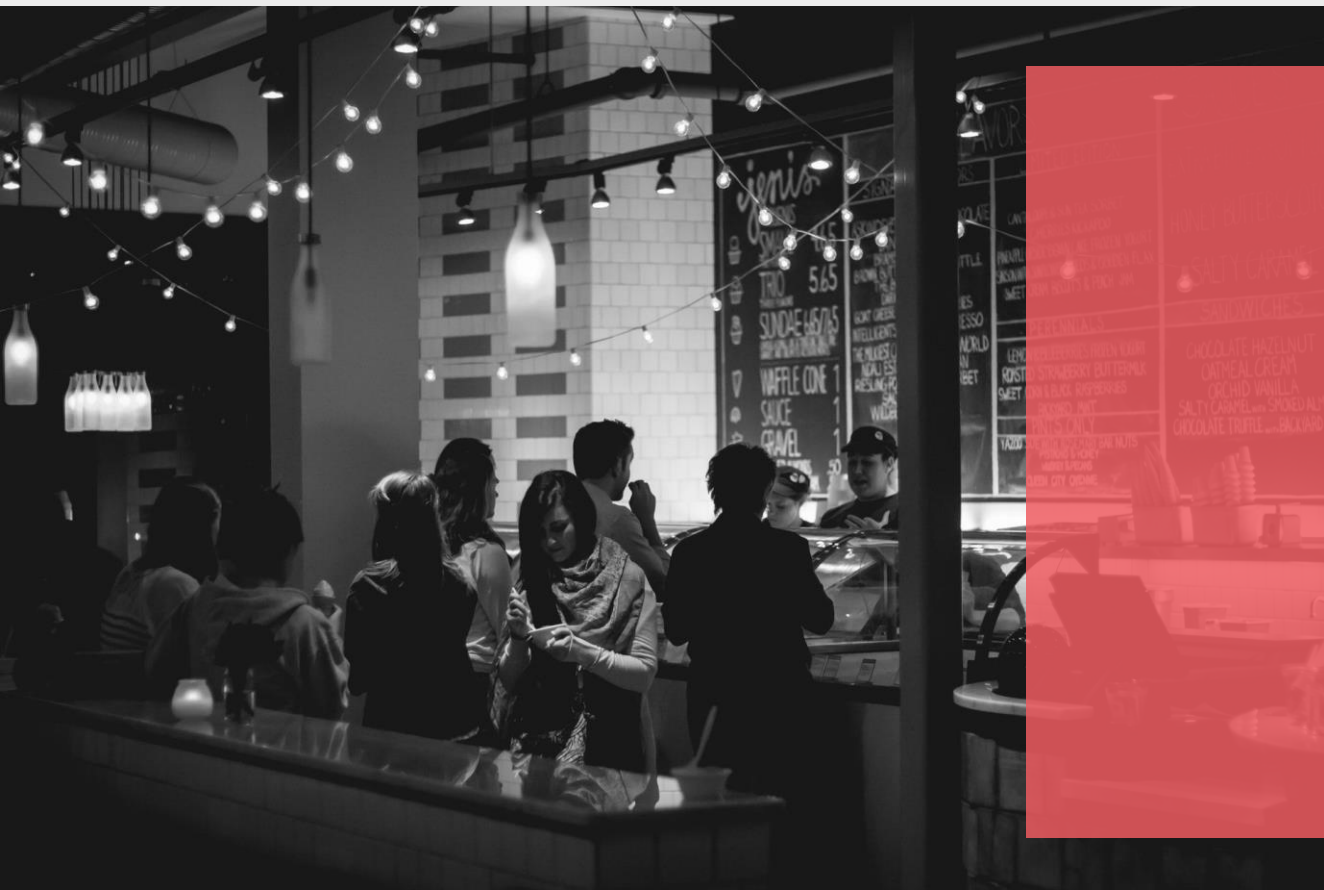
结论与建议



1

背景介绍

在线民宿市场



在线民宿与传统酒店的主要差异：



房源分散，单点房源量较少



产品个性化，经营主体多元化



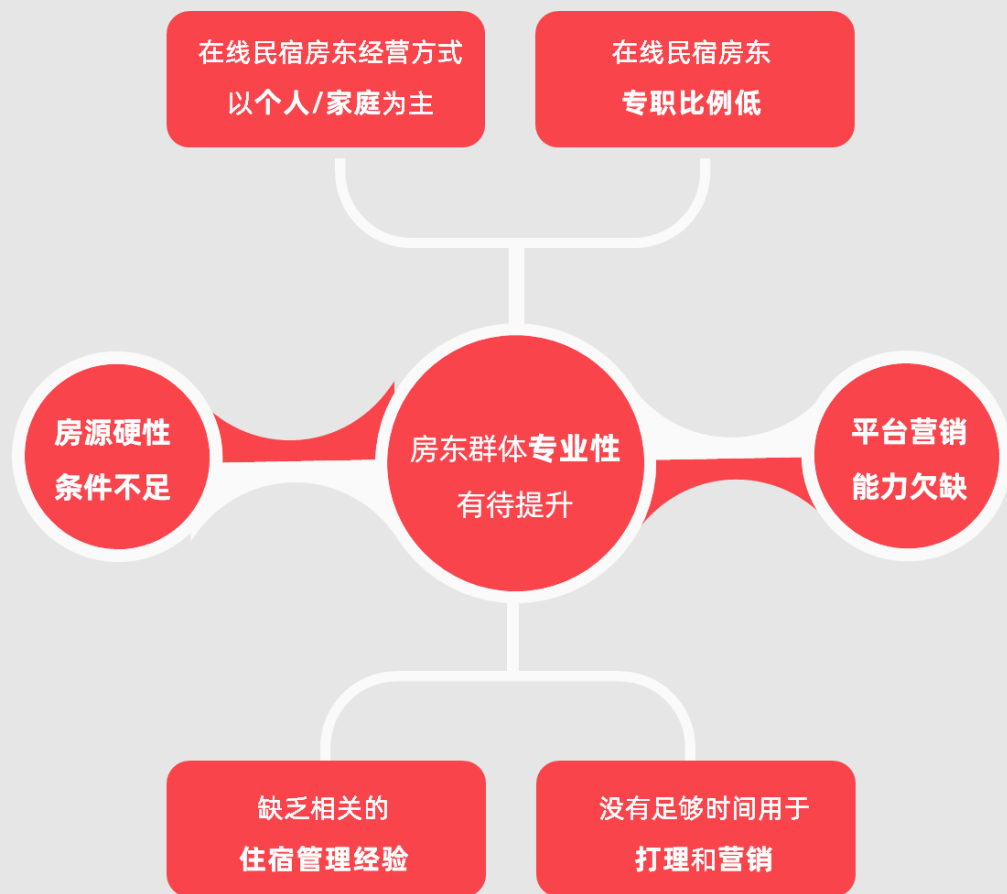
依赖互联网+模式

在线民宿市场



破土萌芽，野蛮生长		有序发展，规范洗牌		资本角力，欣欣向荣	
途家	国内首家中高端度假公寓预订平台	途家收购蚂蚁短租 整合携程和去哪儿网旗下的公寓民宿业务		深化与同城艺龙的合作关系，优势互补，共享在线房源	
蚂蚁短租	国内短租领域开拓者，背靠赶集网				
小猪短租	国内共享住宿代表企业，58同城加持	小猪短租估值超过10亿美元，正式进入独角兽行列		发力乡村民宿，对现有乡村客栈进行升级改造优化	
	Airbnb	Airbnb强势入局，专注在华业务		加速拓展中国非一线城市房源，进一步细分民宿预订品类	
	榛果民宿	美团点评正式宣布切入分享住宿领域—榛果民宿App		成立榛果管家体系，为民宿房东提供标准化服务参考	

问题与挑战



BOOM..!





2

数据说明

数据来源



数据获取

某在线民宿预订平台北京地区475条房源数据

数据处理

房源评价等级：

五星房源——房源评分为5分

非五星房源——房源评分低于5分

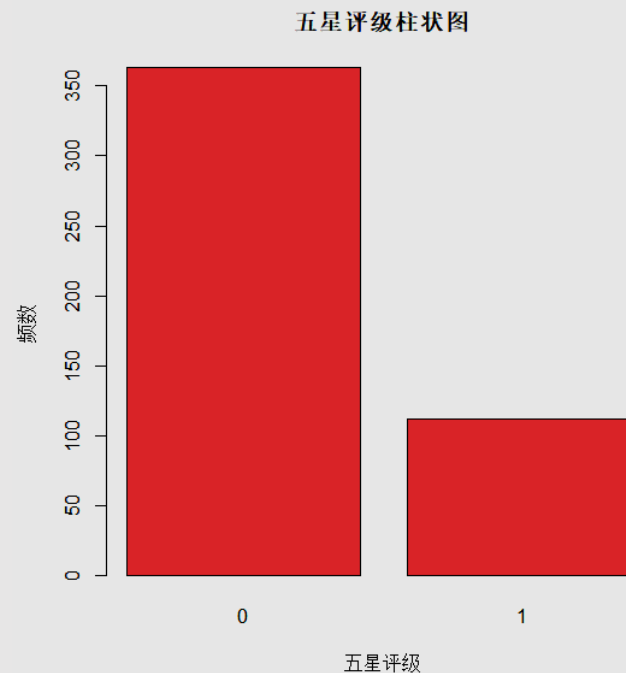
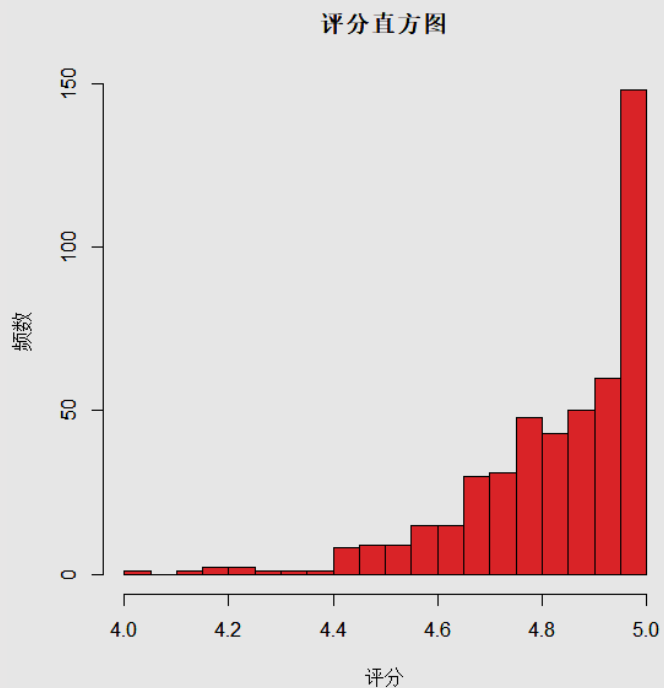
数据说明

变量类型			变量名	详细说明	取值范围	备注
因变量			房源评分等级	定性变量 (2水平)	0,1	0：非五星房源 1：五星房源 五星房源占比24%
自变量	房源硬性条件	房源配置	房源类型	定性变量 (2水平)	独立， 整套	整套类型占比55.16%
			容纳数量	定序变量	1,2,3,4,4+	1-3位占比79.37%
			厨房配套	定性变量 (2水平)	有， 无	有厨房占比85.47%
		房源地段	所处城区	定性变量 (5水平)	朝阳， 东城， 丰台， 海淀， 西城	朝阳区占比最高为45.68%
			地铁配套	定性变量 (2水平)	有， 无	有地铁站占比39.37%
			与天安门的距离	定量变量 单位：千米	0.5~28.0	右偏分布， 中位数为5.62
	平台营销能力	房东展示	标题长度	定量变量 单位：字	3~54	左偏分布， 中位数为40
			房源价格	定量变量 单位：元	70~1480	右偏分布， 中位数为299
			打折类型	定性变量 (3水平)	不打折， 有条件打折， 直接打折	有条件打折占比最高为44.42%
		用户评价	是否高性价比	定性变量 (2水平)	是， 否	高性价比占比23.16%
			是否超赞房东	定性变量 (2水平)	是， 否	超赞房东占比58.53%
			评论数	定量变量 单位：条	3~313	右偏分布， 中位数为40

3

描述统计

房源评分等级

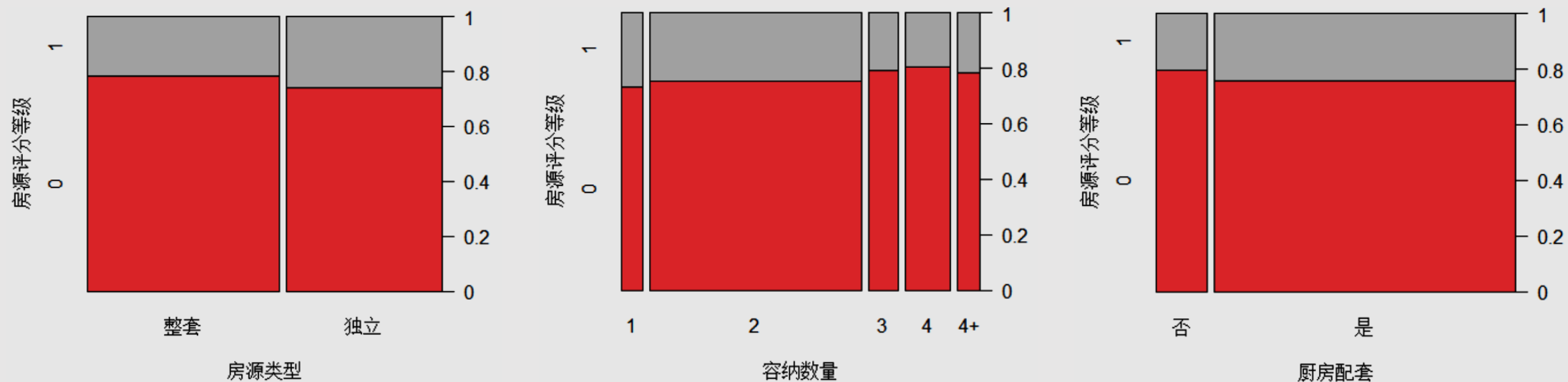


房源评分整体呈**左偏**分布，极少数房源评分**低于4.4分**



在所有房源中，五星房源共有**112套**，非五星房源共有**363套**，五星房源占整体的**23.58%**

房源硬性条件-房源配置



在房源类型上，**独立类型**的房源获五星评级比例**略高**，相比整套房源，独立房间总体上**私密性**更好，更受房客欢迎

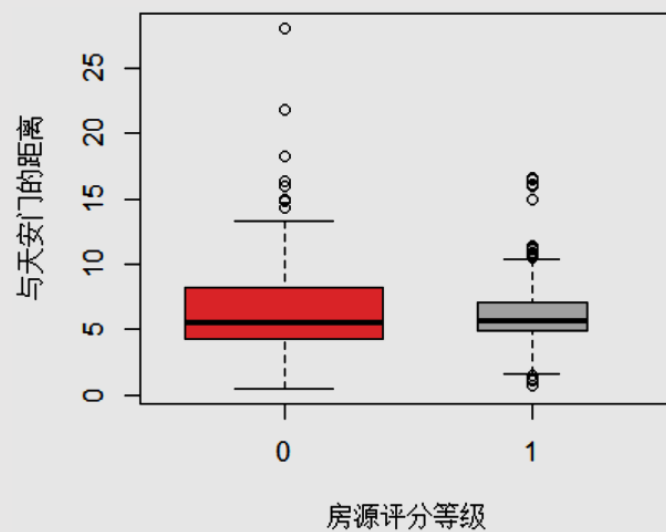
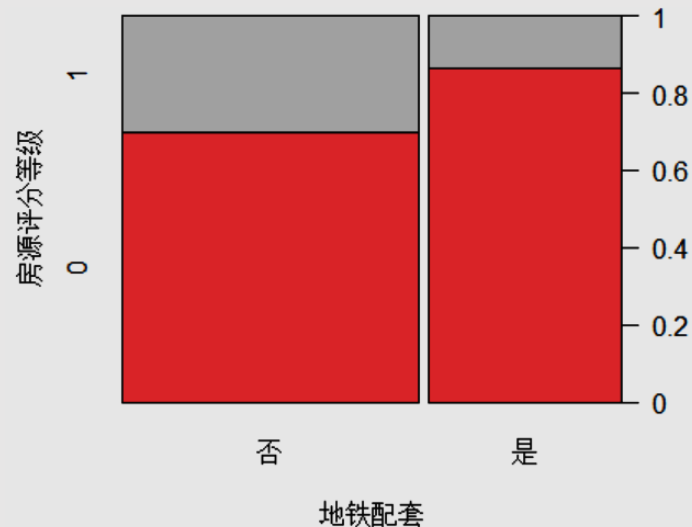
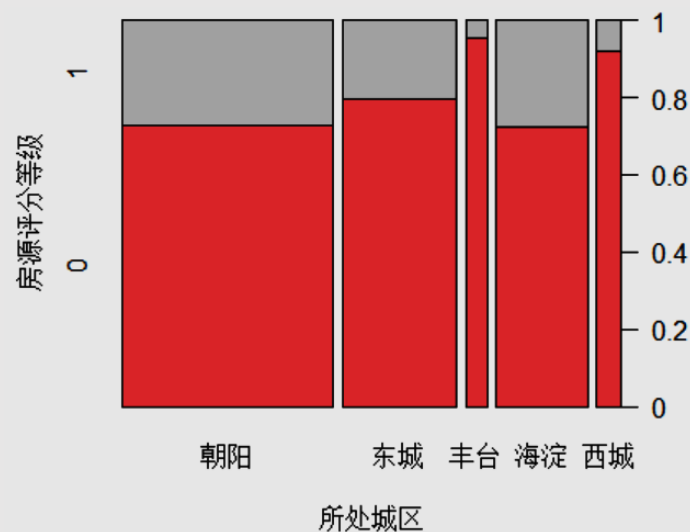


在容纳人数上，**单人房源**五星评级占比最高，**二人房源**紧随其后，其余容纳类型占比均在20%上下波动。这可能是由于容纳数量的增加带来了需求的增加，从而导致消费者满意度降低



在厨房配套上，**拥有厨房配套**的房源获得五星评级的比例**更高**，由于厨房配套属于日常生活中较为重要的设施，如果没有配备，会给房客的日常起居带来不便，从而影响房源的评分等级

房源硬性条件-房源地段



在城区位置上，**朝阳、海淀、东城区**房源获五星评级比例**更高**。初步猜想，这可能与各城区的旅游资源有关

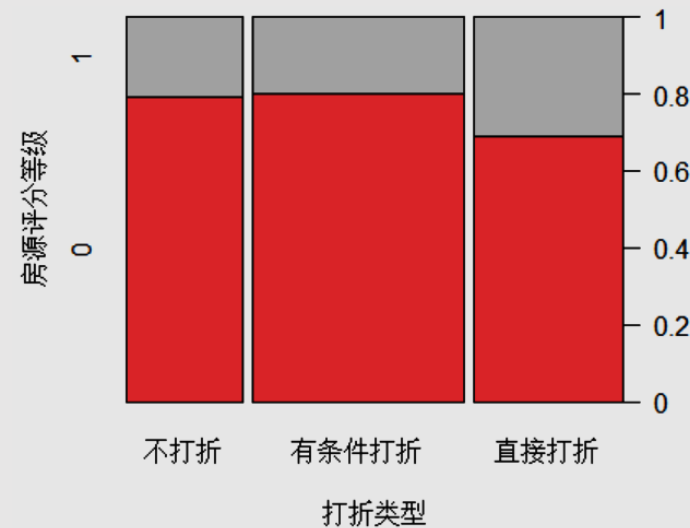
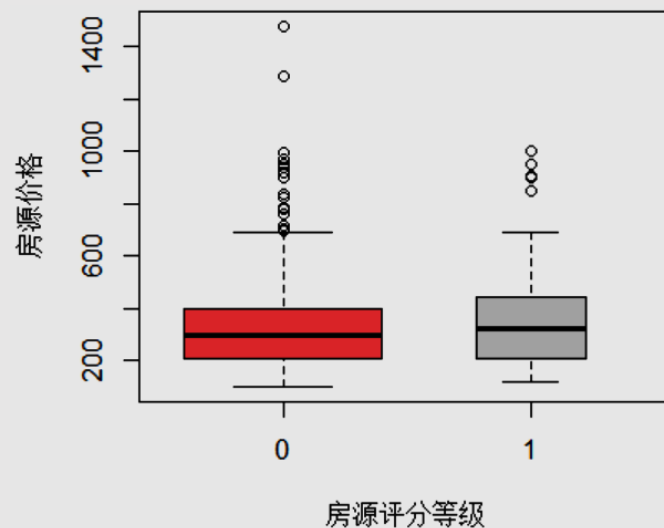
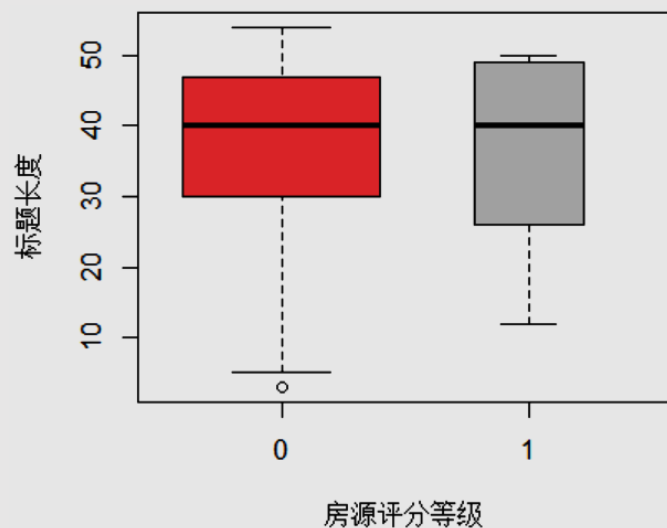


在地铁配套上，在五星房源中，**拥有地铁配套**的占比**远小于**其在非五星房源中的占比，地铁配套因素可能在某种程度上影响了五星评级



在与天安门的距离上，以中位数计，五星房源与非五星房源没有明显差异，但非五星房源在距离分布上更为**分散**

平台营销能力-房东展示



在标题长度上，以中位数计，五星房源与非五星房源没有明显差异，但非五星房源在标题长度上更为**分散**

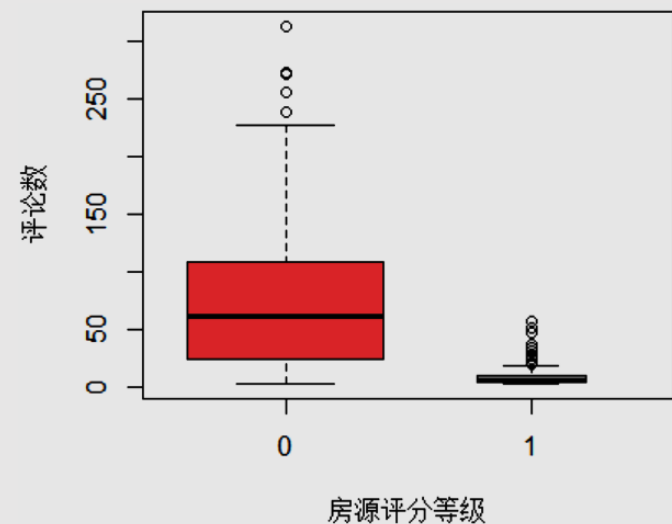
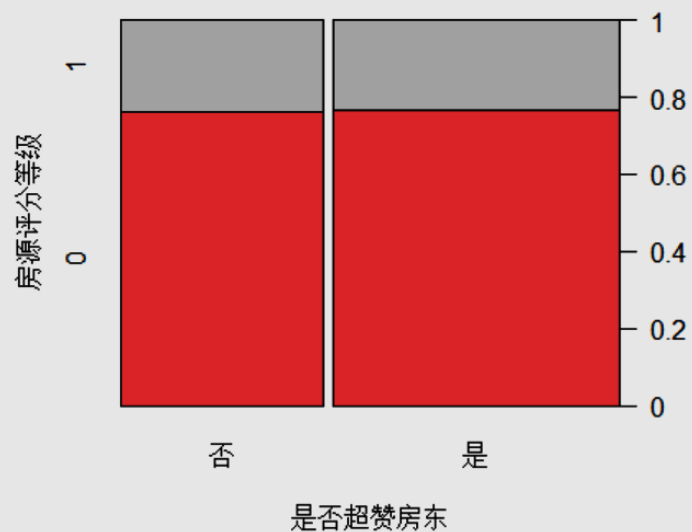
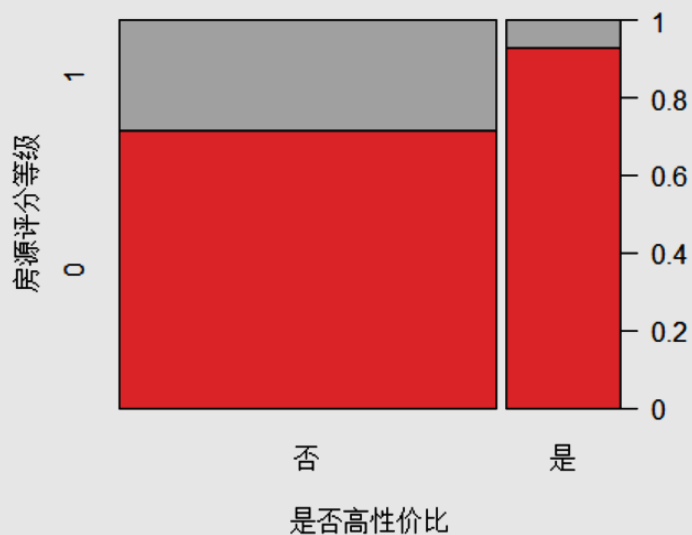


在房源价格上，五星房源中**高性价比**的占比很小，说明性价比并无法完全让房客满意



在打折类型上，**直接打折**的五星评级比例**最高**，其余两项差异不大，说明干脆利落的打折活动更能给房客带来愉悦感从而影响房源评级。

平台营销能力-用户评价



在性价比等级上，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



在房东等级上，五星房源和非五星房源没有显著差异



在评论数目上，**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



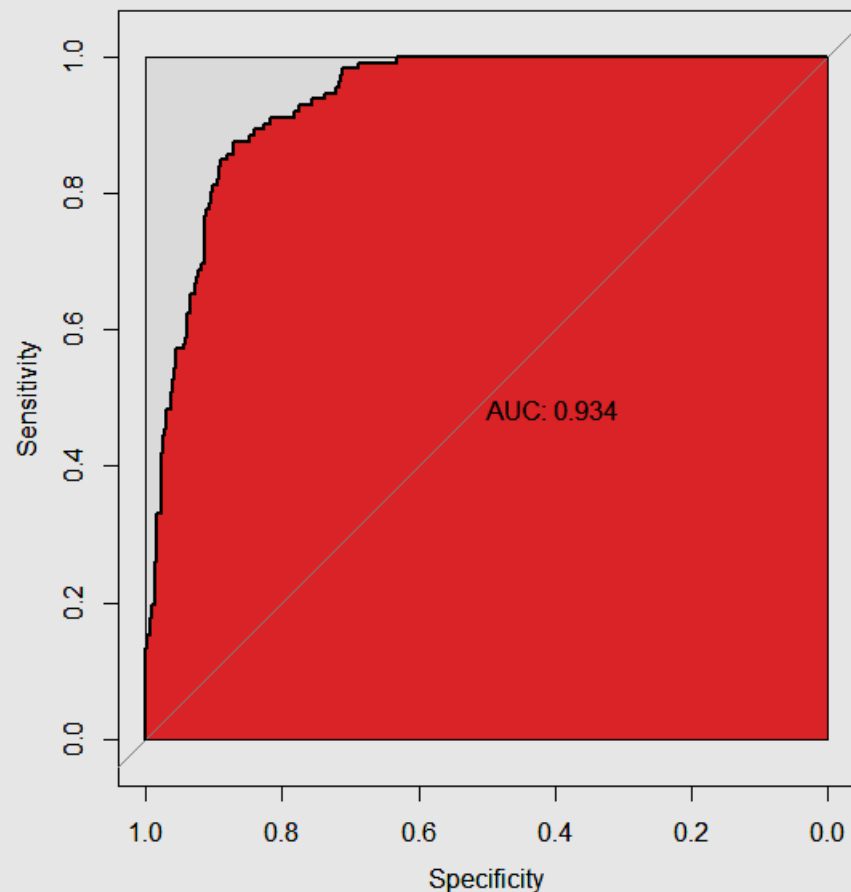
4 建模分析

逻辑回归模型结果

变量	回归系数	p值	备注
截距项	0.669	0.300	
房源类型-独立	0.991	0.011	基准组：整套
地铁配套-有	-0.888	0.010	基准组：无地铁配套
与天安门的距离	-0.129	0.007	
是否高性价比-是	-1.213	0.023	基准组：是否高性价比-否
是否超赞房东-是	1.138	0.001	基准组：是否超赞房东-否
评论数	-0.111	<0.001	
房源价格	0.003	0.016	
AIC : 273.95			

利用AIC准则进行逐步回归

逻辑回归模型ROC曲线&混淆矩阵



		Prediction	
		1	0
Label	1	81	31
	0	31	332

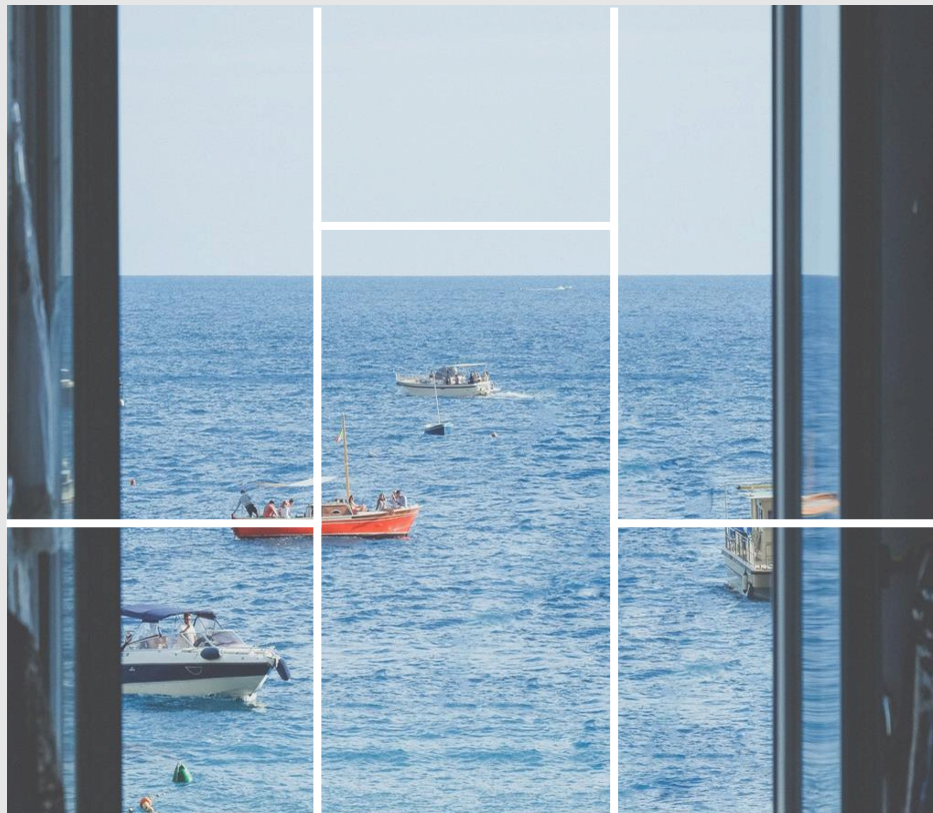
$$\text{准确率 (ACC)} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{81+332}{81+332+31+31} = 86.95\%$$

$$\text{精确率 (PPV)} = \frac{TP}{TP+FP} = \frac{81}{81+31} = 72.32\%$$

$$\text{灵敏度 (TPR)} = \frac{TP}{TP+FN} = \frac{81}{81+31} = 72.32\%$$

$$\text{特异度 (TNR)} = \frac{TN}{TN+FP} = \frac{332}{332+31} = 91.46\%$$

模型解读&评价



房源配置

相较于整套类型，**独立类型**的房源更容易成为五星房源



房源地段

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



用户评价

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



房东展示

房源价格越高，房源评分等级越高



模型评价

AUC值为**0.934**，ROC曲线贴合左上角，说明模型分类效果很好



5

总结讨论

结论与建议

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（房源类型、地铁配套、与天安门的距离）和平台营销能力因素（房源价格、是否高性价比、是否超赞房东、评论数）
建议1	合理规划房屋类型 房东可以把整套房子拆分成独立房间出租，从而提高房客满意度
建议2	避免地铁配套误区 房东在考虑房源交通便利性的同时，应更注重与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
建议3	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
建议4	摆脱性价比的束缚 现在的主要目标客群年轻群体经济能力较好，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仍需以品质优先

狗熊会 | 扫描二维码

关注狗熊会

获取更多案例资源



谢谢观赏