



# 集团18咨询公司。

影响漫威 airbnb  
把主机变成超级hosts



## 第一阶段里程碑更新



G18咨询项目组：

乔纳森约克吴安东尼

Jisoo KIM Elsa ZHAN



G18 CONSULTING  
WE EAT YOUR PROBLEM FOR BREAKFAST

道情张特伦斯

# 1. Introduction

- 项目的定义
- Airbnb的“超级主机”节目的背景 • 香港市场研究

# 1.1 项目定义

有哪些简单而可行的因素可以帮助香港的主持人成为超级主持人？🏆

# 1.2背景-“超级主机”

## 什么是“超级主机”？

Airbnb表现最好的主持人。

主机必须拥有一个良好的帐户  
在过去12个月内符合以下标准的人员：

- 完成至少10次或3次旅行  
预订总共至少有100晚；
- 保持90%的+应答率；
- 保持低于1%的取消率，属于Airbnb的用户除外  
情况政策；
- 保持了4.8分的总体评级。

## 为什么是“超级主机”？

更客人的客人，  
额外的收入潜力，独家奖励，并从Airbnb  
获得优先支持

- 每周的浏览量增加了5%
- 入住率提高81%
- 日收入同比增加60%  
普通主机平均
- 从AirBnB提供的现金奖励  
指导

## Airbnb的特点是什么？

鉴于Airbnb的收入主要来自两个来源，它也被激励去做了

鼓励更多的主持人成为“超级主持人”：

- **由主持人收取的佣金：每次支付**

有人选择了一个主机的属性和  
在付款时，Airbnb收取支付金额的10%  
作为佣金

- **从旅客处收取的交易费用：在以下时间内**

旅客支付住宿费用，他们被收取3%的交易费用。这个

这笔钱增加了Airbnb的收入。



## 1. 2背景市场研究



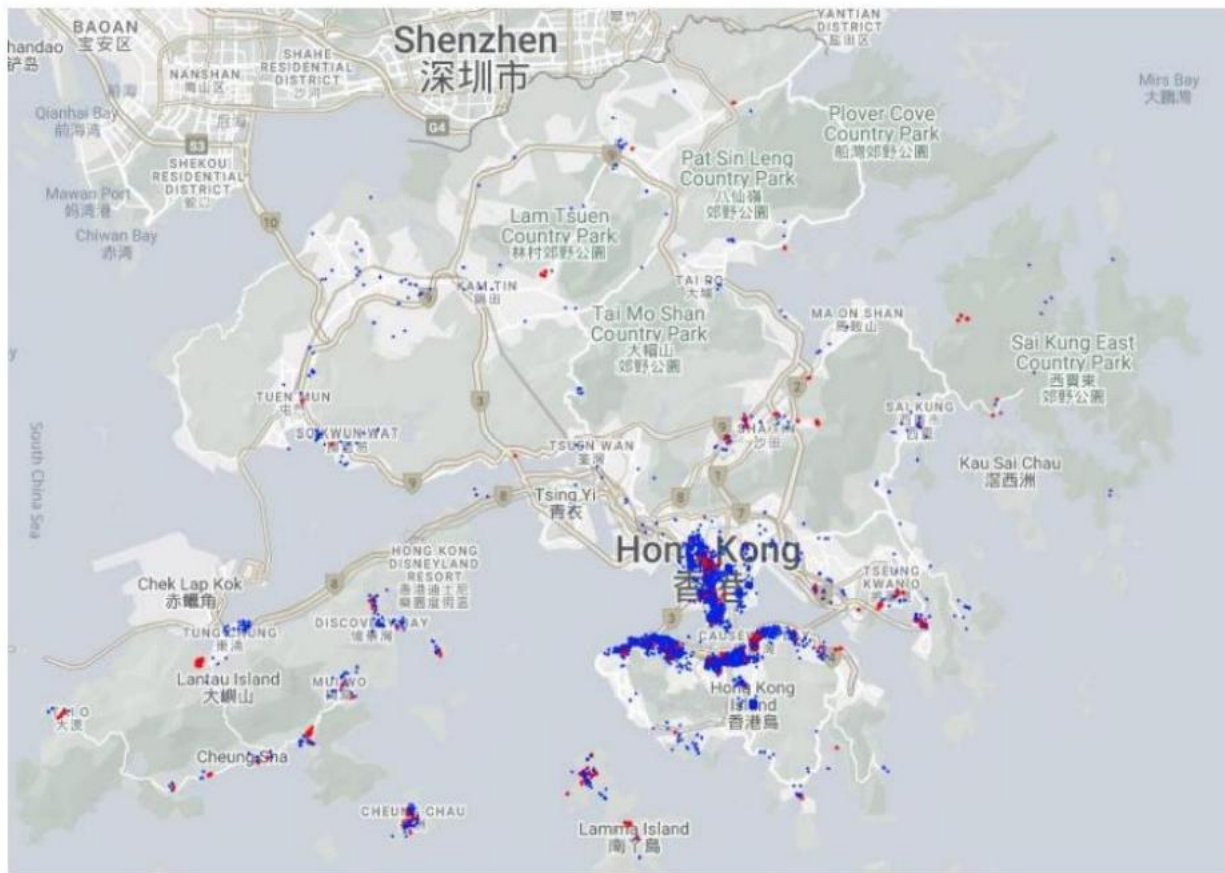
19.4%



10.5%



# 1.2背景资料-香港市场研究



包括2009年至2022年的主机

5056个列表，532个超级主机  
最后一次搜集的数据是在2022年9月16日



邻居的影响？

在沙田、西贡及香港特区上市  
香港岛通常是由  
“Superhosts”。

黄大新葵青上市  
而埔的税率最低  
of “Superhosts”.



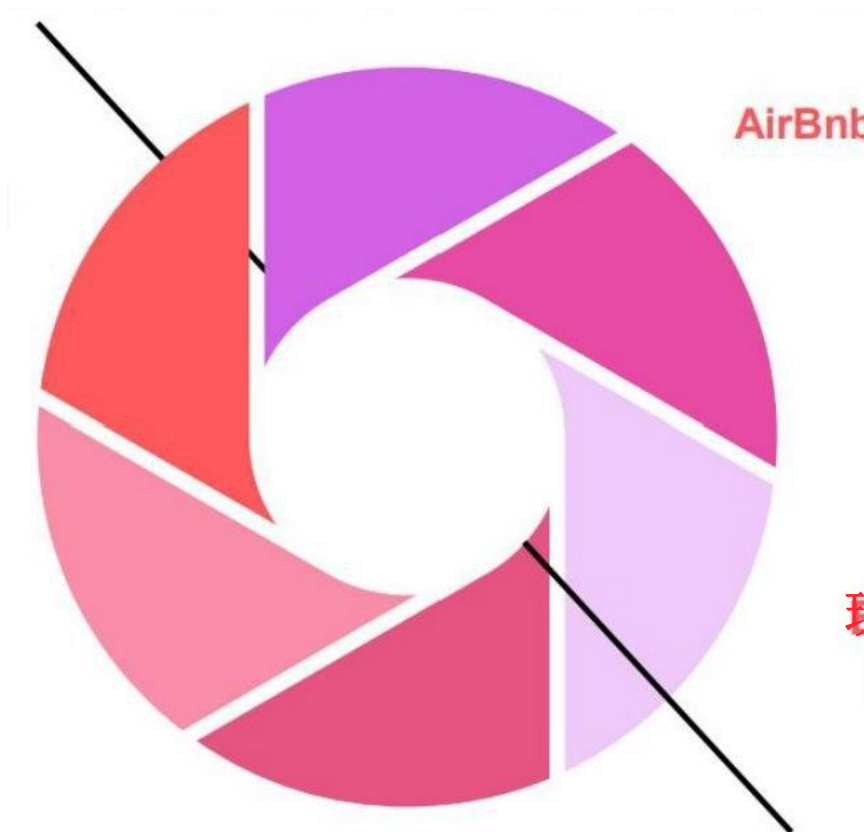
# 1.3利益相关者分析

提供实用的技巧，以帮助有抱负的主机和现有的主机成为“超级主机”。

利益相关者  
受影响

有抱负  
的主机

现有和  
潜在客人



利益相关者  
有影响的

现有  
主机

## 2. 分析和发现

• 数据清理和准备 • 模型和发现



## 2.1 数据准备

### 数据集：

- 在香港有5056份清单，75栏
- 变量概述：
  - 主机信息：位置、iD、验证状态
  - 清单信息：最大/最小的夜晚可用，价格，设施，房间类型，位置
  - 关于评审分数的信息：评审分数及其价值的分类

### 数据准备所采取的步骤：

1. 将具有对象的数据类型的数据转换为适当的数据类型；
2. 删除缺失值超过75%的列；
3. 用适当的条目填充缺失的数据；
4. 正在删除变量；以及
5. 创建分类/虚拟变量。



## 2. 2理解主机vs。超级主机

理解差异

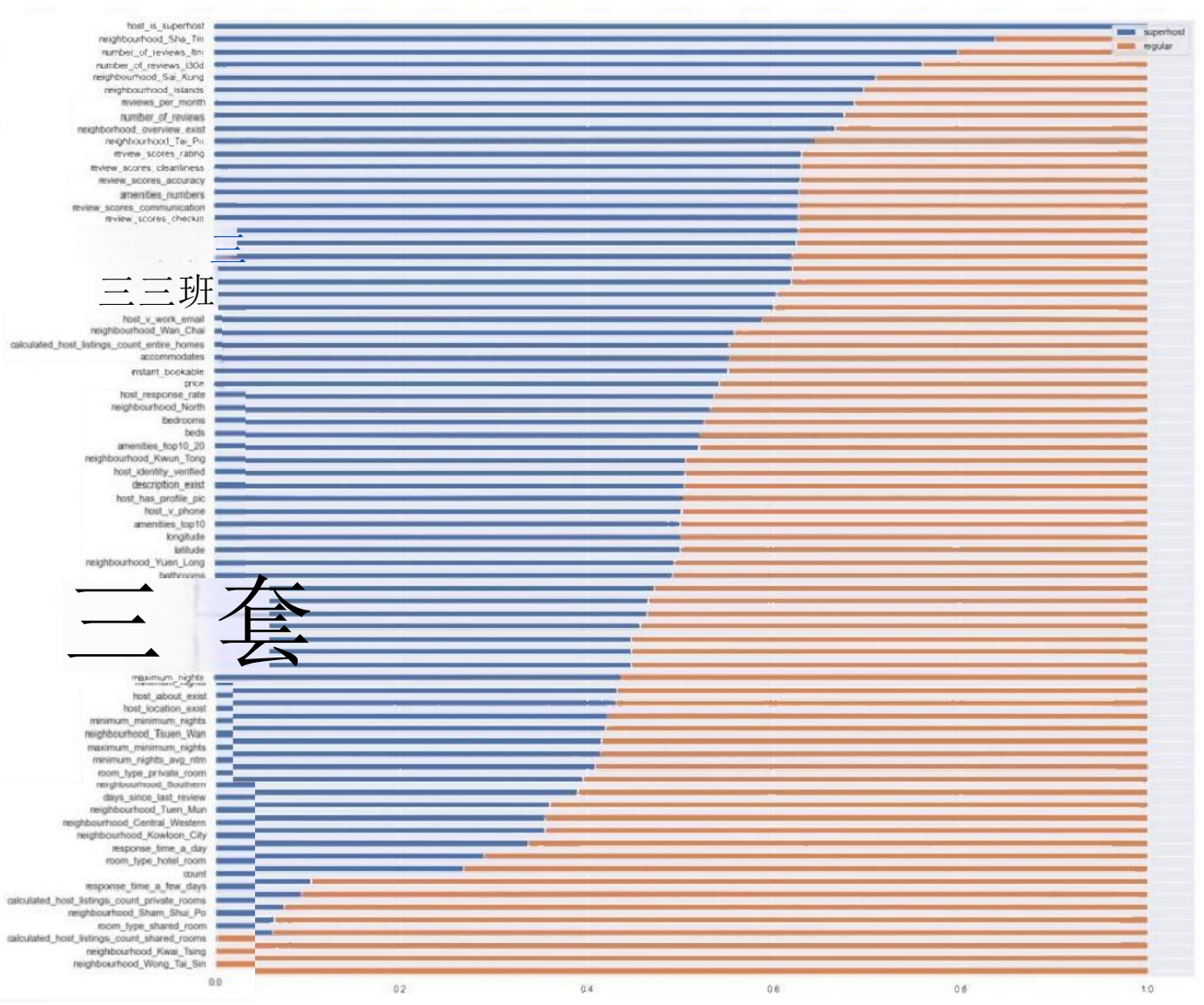
1. 邻里关系很重要！
2. 评论和超级主机是相关的
3. 评论的数量对成为超级节目主持人也很重要
4. ...
5. 有很多东西。。。。。。。。。



TLDR: 有很多因素

然而，与成为超级主机相关的是，并不是所有的主机都是可操作的或简单的。





三三班

三套

# 2. 2模型和结果

模型1: 逻辑回归 在...上 特征 数据

- Host\_is\_Superhost: 相关变量 (二进制)
- 采用AIC前向选择模型, 进一步缩小了变量的数量。
- 进行VIF分析, 检查多重共线性问题。我们决定在哪里放弃  
“host\_has\_profilepic”
- | 我们最终的逻辑回归模型包含9个变量。

变量VIF变量		维
0设施_编号5.237735	0设施_编号5.072229	
host_acceptance_rate3.853621	1host_acceptancerate 3.780598	2主人_v_电子邮件7.852591
2主人_y_电子邮件172.21985		
3邻域概述exist2.294009		
4host_identity_venfied 2.728111		
5响应时间_a_第1.16天1903		
6响应时间_a_数天1.102534	3 neighborhood_oveview_exist 2.290653	4host_identity_venfied 2.721392
7卧室2.695326	5响应时间_a_第1.157507天	6响应时间_a_几天1.091049
8 instant_bookable1 .8454647床位2.661999		
9主机_about_存在5.039361	8即时预订1.844996	9ghost_about_存在4.616332
10主机有个人资料图片20.95678		



广义线性模型回归结果

Dep。变量：  
模型  
型号系列：  
联系函数  
方法：  
日期：  
时间  
不迭代：协方差  
类型：

host\_is\_superhost  
分划学习  
二项式  
对数  
几率  
红外着  
陆系统  
2022年12月11日星期  
日  
18:30:38  
不健壮的

不观察结果：  
Df残留物：  
Df型号：  
比例：  
对数似然  
偏差：  
皮尔逊chi2：

4050  
4039  
10  
1.0000  
-932.70  
1865.4  
2.81e+03

	coef	std err	z	P> z	[0.025	0.975]
—						—
拦截	-2.7300	0.216	-12.648	0.000	-3.153	-2.307
设施数量	0.0698	0.006	10.891	0.000	0.05	0.082
主机接受率	2.7683	0.242	11.427	0.000	2.294	3.243
主人_v_电子邮件	-1.9317	0.196	-9.856	0.000	-2.316	-1.548
邻域概述_存在	1.1165	0.144	7.736	0.000	0.834	1.399
已验证主机身份	-0.8830	0.142	-6.217	0.000	-1.161	-0.605
响应时间_a_一天	-1.5392	0.300	-5.124	0.000	-2.128	-0.950
响应时间数天	-1.9418	0.597	-3.251	0.001	-3.112	-0.771
卧室	-0.1859	0.050	-3.706	0.000	-0.284	-0.088
即时预订	-0.5843	0.147	-3.976	0.000	-0.872	-0.296
host_about_存在	-0.2795	0.136	-2.050	0.040	-0.547	-0.012



# 广义线性模型回归结果

Dep. 变量: host\_is\_superhost  
模型分划学习法  
型号系列: 二项式  
联系函数对数几率  
方法: IRLS  
日期: 2022年12月11日星期日  
时间: 18: 30: 38  
不迭代: 7  
协方差类型: 非鲁棒性

不观察结果: 4050  
Df残留物: 4039  
Df型号: 10  
比例: 1.0000  
对数似然: -932.70  
偏差: 1865.4  
皮尔逊chi2: 2.81e+03

				=====		
				P> z	[0.025	0.975]
					—	
	coef	std errZ				
-				0.00	-3.153	-2.307
拦截-2.7300		0.216	-12.648	0	0.057	0.082
设施数量0.0698		0.006	10.891	0.00	2.294	3.24
主机接受率	2.768	0.242	11.427	0	-2.316	3 -
主人_v_电子邮件	3 -	0.196	-9.856	0.00	0.834	1.548
邻居概述主机身份验证	存在 1.9317	0.144	7.736	0	-1.161	1.39
响应时间_a_一天	1.116	0.142	-6.217	0.00	-2.128	9 -
响应时间_a_数天	5 -	0.300	-5.124	0	-3.112	0.605
卧室	0.8830	0.597	-3.251	0.00	-0.284	-0.950
即时可簿记	-1.5392	0.050	-3.706	0	-0.872	-0.771
主机存在	-1.9418	0.147	-3.976	0.00	-0.547	-0.088
	-0.1859	0.136	-2.050			-0.296
	-0.5843					-0.012
	-0.2795					
				=====		



广义线性模型回归结果

Dep。变量: host\_is\_superhost不观测结果: 4050  
模型GLM Df残留物: 4039  
模型族: 二项式的Df型号: 10  
链接功能: logit比例尺: 1.0000  
方法: 红外着陆系统对数似然-932.70  
日期: 2022年12月11日星期日偏差: 1865.4  
时间: 18: 30: 38皮尔逊chi2: 2.81e+03  
不迭代: 7  
协方差类型: 不健壮的

	coef	std err	z	P> z	[0.025	0.975]
拦截	-2.7300	0.21	-12.648	0.000	-3.153	-2.307
设施数量	0.0698	6	10.891	0.000	0.057	0.082
主机接受率	2.7683	0.006	11.427	0.000	2.294	3.243
主人_v_电子邮件	-1.9317	0.242	-9.856	0.000	-2.316	-1.548
邻居概述存在主机身份验证	1.1165	0.196	7.736	0.000	0.834	1.399
	-0.8830	0.144	-6.217	0.000	-1.161	-0.605
		0.142				
响应时间_a_一天	-1.5392	0.300	-5.124	0.000	-2.128	-
响应时间_a_数天	-1.9418	0.597	-3.251	0.001	-3.112	0.950
						-
						0.771
卧室	-0.1859	0.050	-3.706	0.000	-0.284	-0.088
即时可簿记-0.58430.147-3.976	0.000	-0.872-0.296				
host_about_存在	-0.2795	0.136	-2.050	0.040	-0.547	-0.012





# 2. 2模型和结果

## 模型 2:逻辑的 回归 在...上 设施设施 数据

- 更进一步，进一步了解“便利设施”。
- 过滤掉在香港上市时最常见的产品。选取35个变量，占香港所列便利设施的80%以上。
- 1进行了另一个逻辑回归来看看与超级主机状态最相关的？
- 使用AIC前向选择模型和VIF来测试多重共线性问题。

可变因素维夫

0洗发水6. 2523719厨房4. 788073	
1铁2. 925847	10烘干机1. 460525
2热水水壶1. 430071	11盘子和银器3. 743478
3First_aid_kit1. 788546	12热水3. 334319
4电梯2. 708379	13个额外的枕头和毯子1. 721110
5咖啡机1. 603831	14火extinguisher2. 999285
6电视3. 687530	15烹饪基础知识2. 539328
7有线电视1. 756567	16电吹风机6. 879310
8Dodicated_workspace1. 419282	17冰箱3. 522900

- 18空调14. 585210
- 19语句6. 893271
- 20吊架4. 298193
- 21长期停留允许15. 029286
- 22Luggage\_dropoff\_allowed 1. 835552
- 23一氧化碳报警1. 640830
- 24Lock\_on\_bedroomdoor 1. 853370



广义线性模型回归结果

Dep. 变量: 主人\_是\_超级主机

不Observations: 5056

型号: GLM

Df残留物: 5031

模型族: 二项式的

Df型号: 24

联系函数对数几率

比例: 1.0000

方法: IRLS

对数似然 -1256.7

日期: 2022年12月12日, 星期一

偏差: 2513.4

时间: 19: 37: 25

皮尔逊chi2: 5.36e+03

不迭代:

协方差类型: 非鲁棒性

</

顶部 3:  
洗发水  
钢铁  
咖啡机



广义线性模型回归结果

Dep。变量：主人\_是\_超级主机  
型号：GLM  
模型族：二项式的  
联系函数对数几率  
方法： IRLS  
日期：2022年12月12日，星期一  
时间：19：37：25  
不迭代：  
协方差类型不稳健

不Observations：  
Df残留物：  
Df型号：  
比例：  
对数似然  
偏差：  
皮尔逊chi2：

5056  
5031  
24  
1.000  
0 -  
1256.7  
2513.

负  
影响：为什么？  
厨房

	coef	std err	Z	P> z	[0.025	0.975]
拦截	-4.5907	0.346	-13.260	0.000	-5.269	-3.912
沙姆波克	1.2221	0.179	6.831	0.000	0.871	1.573
铁	0.8694	0.131	6.656	0.000	0.613	1.125
热水水壶	0.4263	0.169	2.519	0.012	0.095	0.758
急救箱	0.5245	0.121	4.329	0.000	0.287	0.762
电梯	1.0319	0.115	8.942	0.000	1.258	-0.806
咖啡机	0.8132	0.155	5.233	0.000	0.509	1.118
电视机	0.7333	0.152	4.825	0.000	0.435	1.031
有线电视	0.7780	0.188	4.148	0.000	0.410	1.146
专用工作区	0.5728	0.129	4.447	0.000	0.320	0.825
厨房	-0.8450	0.149	5.684	0.000	-1.136	-0.554
耶尔博士	0.2279	0.123	1.855	0.064	-0.013	0.469
盘子和银器	1.0339	0.202	5.108	0.000	0.637	1.431
热水	-0.4396	0.149	-2.940	0.003	-0.733	-0.147
额外的枕头和毯子	-0.3403	0.152	-2.237	0.025	-0.639	-0.042
灭火器	0.4785	0.130	3.681	0.000	0.224	0.733
烹饪基础知识	0.3323	0.154	2.159	0.031	0.031	0.634
你的头发	0.8791	0.207	4.251	0.000	0.474	1.285
制冷机	-0.5541	0.168	-3.303	0.001	-0.883	-0.225
必需品	-0.2636	0.203	-1.299	0.194	-0.661	0.134
吊架	0.1213	0.159	0.761	0.447	-0.191	0.434
Long term stays allowed	0.2415	0.275	0.879	0.379	-0.297	0.780
Luggage dropoff allowed	-0.0908	0.141	-0.643	0.520	-0.368	0.186
一氧化碳报警器	0.3394	0.130	2.607	0.009	0.084	0.594
锁在卧室门上	-0.1268	0.12g	0.980	0.327	-0.386	0.12



=====

==== 年

=====

## 2.2模型和结果

模型3：在特征数据上的随机森林分类

-对上述最终模型1所确定的变量进行随机森林回归，  
以从另一个角度验证每个变量的重要性。

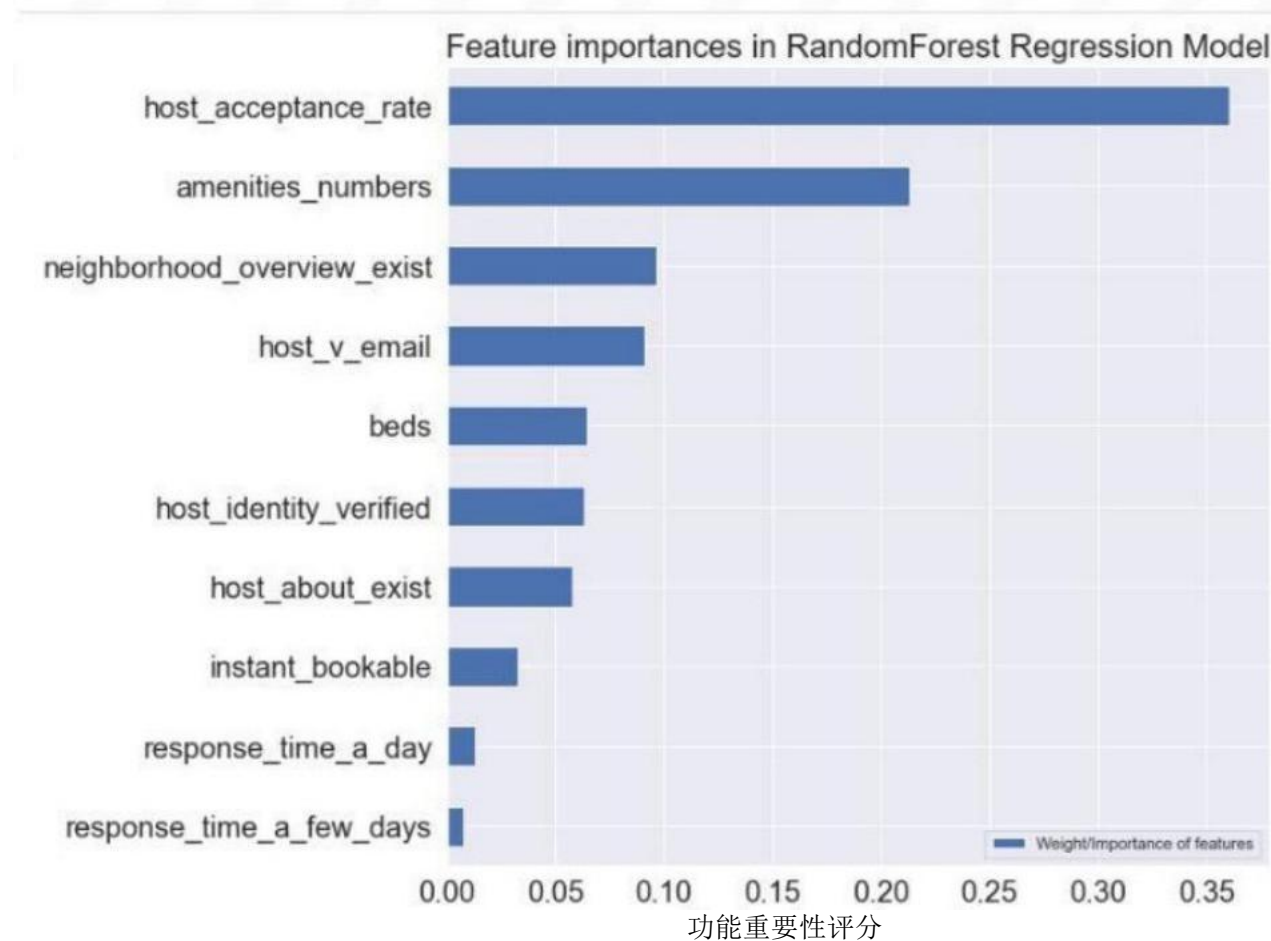
-这两种模型都产生了相似的结果

可用的便利设施和主机的数量”

接受率以及存在

社区概况对成为“超级主机”的影响最大

。



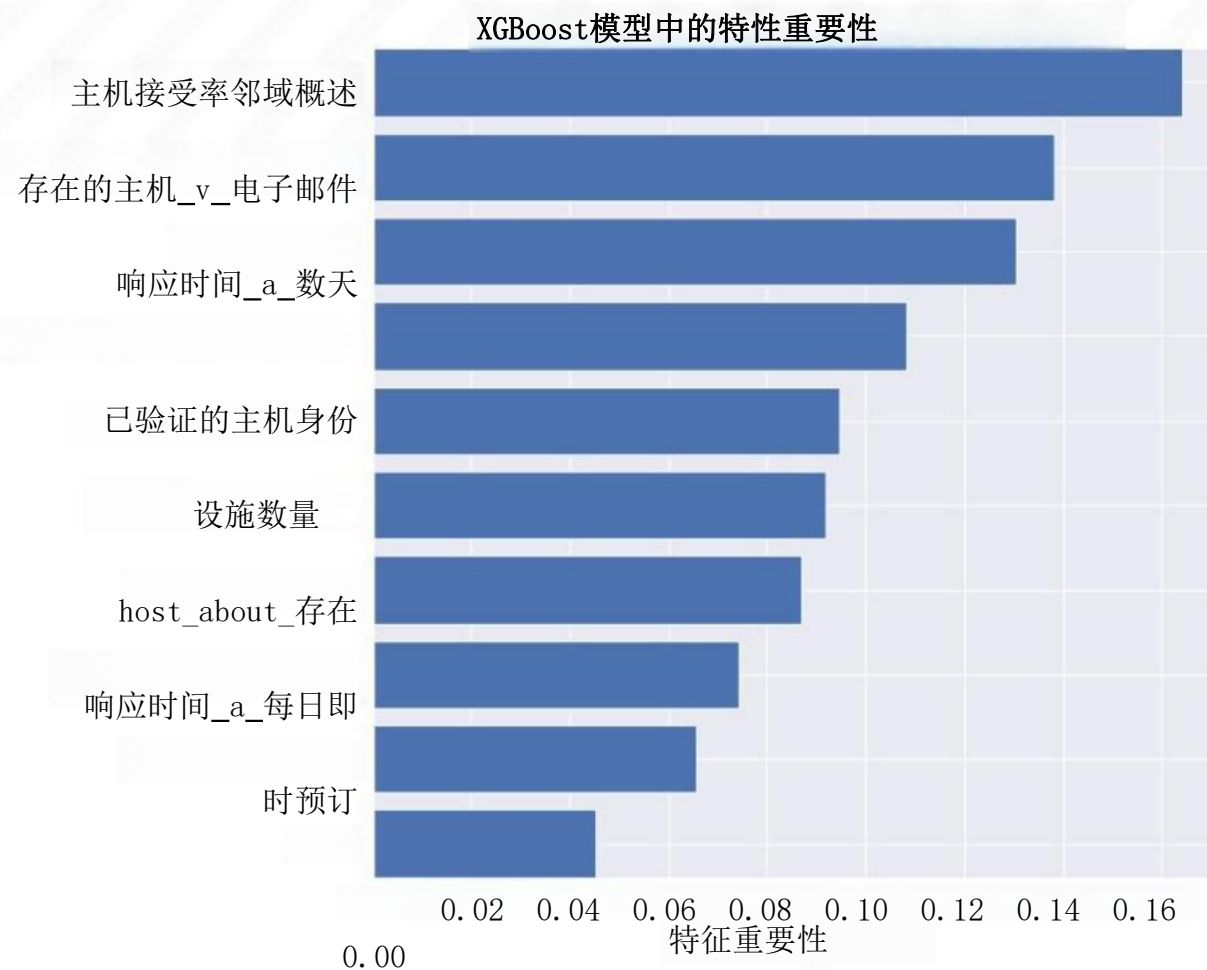




## 2. 2模型和结果

### 型号4: XGBoost

- XGBoost模型来验证每个模型的重要性  
在最终的模型1中确定的变量
- 有验证电子邮件和慢响应时间的主机有最高的重要性得分。
- 我们必须谨慎，这个模型不能告诉我们  
一个特征是积极的还是消极的影响  
成为“超级主持人”的前景。



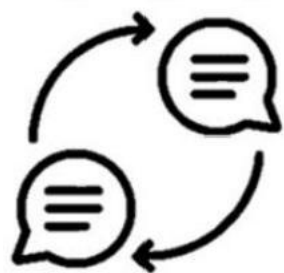


# 3.

建议 • 对主机的建议

# 3. 建议

Airbnb应该鼓励香港的房东。



响应客户机内的一个  
一天



尽可能多地接受预订

删除未来日期的列表  
如果他们不能做到，请提前付款

接受预订



包括概述  
附近的清单



提供更多的设施。不仅仅是  
是主食，而且是为你的  
客人提供额外的装饰

顶级设施：  
铁，电视，装备厨房，  
配备浴室，  
工作空间、安全设备



# 谢谢！ 有什么问题吗？

你可以在

✉ [ProjectMarvel@G18Consulting.co](mailto:ProjectMarvel@G18Consulting.co)

