试探共享经济的信誉系统：以 Airbnb“星级评分”的影响因素为切入点

1612555 顾　容　菱
1613641 张　靖　昱
1711342 李　　纪
1812972 艾 合 买 提
1811144 李　　岳

摘要

信誉系统是当今互联网共享平台的一个基本组成部分，而 Airbnb 作为最成功和最有影响力的共享平台之一，其信誉系统尤其是“星级评分”系统具有较高的调查价值。本文的基础数据为 2018 年 4 月-2019 年 9 月美国 4 个城市共 126887 个 Airbnb 房源的公开月度数据，以及 2019 年 12 月 6 日在 TripAdvisor 爬取的 1386 个酒店和 1211 个旅游租赁房屋的数据。本文首先通过比较 Airbnb 和 TripAdvisor 上所有房源、不同房间类型的房源和不同地区的房源“星级评分”的分布，得出 Airbnb 上的评分具有明显的偏态性，且明显高于另一平台上的评分。为消除因两平台房源差异性带来的影响，本文进行了房源的坐标和描述匹配；在此基础上，对匹配所得的 32 个相同房源进行评分比较，针对 Airbnb 平台上的星级评分仍普遍高于 TripAdvisor 上的现象，本文将其归因于平台效应。同时，本文对两平台上评分和排名的关系进行回归分析。与此同时，本文通过研究 Airbnb 平台上房源的进入和退出过程、用 Markov 链对评分进行动态模拟以及预测房源退出率等，发现“星级评分”的偏态分布可以用“幸存者偏差”和“信号传递理论”来解释。但通过研究评分的稳态分布以及其与评价数量的关系，我们发现评分的动态过程还受到大数定律以及均值回归定理影响。

关键字： Airbnb TripAdvisor 星级评分 平台效应 动态过程 Python

目录

1 引言

从 Airbnb 和 Uber 相继成立的 2008 年算起，“共享经济”从一个陌生的概念，到资本追逐的泡沫，再到今日的冷静与成熟已经走过了十年的历史。环顾四周，我们已经习惯了使用滴滴叫专车、用摩拜租自行车、在旅行时使用小猪短租来预定民宿；但押金问题、监管问题、平台竞争等问题依然困扰着世界各国的各种共享经济企业，至今在经济学领域仍然没有一个可靠的定量模型来证明共享经济作为一种经济模式能够为社会的运转提高多少效率。

那“共享经济”究竟是什么呢？准确地说它是指利用互联网等现代信息技术，以使用权分享为主要特征，整合海量、分散化资源，满足多样化需求的经济活动总和。互联网共享平台的设计主要涉及三个方面：买卖双方的匹配、价格机制以及信任体系。共享平台的核心功能是实现具有高度异质性特征的海量分散的个体买方的需求与个体卖方的供给之间的匹配，包括产品或服务信息、需求信息、价格信息等。我们认为信誉系统是当今互联网共享平台的一个基本组成部分，特别是对于具有不同用户、商品和服务的 P2P（peer-to-peer）平台，而 Airbnb 作为最成功和最有影响力的 P2P 平台之一，其信誉系统具有较高的调查价值。

Airbnb（中文名：爱彼迎）平台上的用户可分为房东和房客两类，双方可以互相进行评分或评论。我们今天探讨的主题“星级评分”就是该信誉系统中“四舍五入到最近的半星的平均房源评级”，它仅仅代表了一个是聚合（即，平均值和四舍五入）的值。

在有关 Airbnb 星级评分的研究中，部分学者指出 Airbnb 上大多数的” 星级评分” 等于或高于 4.5 星，其分布呈现高度偏态（Teubner et al. 2017; Zervas et al. 2015）[7] [1] 。而这引起了人们对 Airbnb 信誉系统（Reputation System）的功能和有效性的一些怀疑。虽然到目前为止，这种被观察到的偏态分布仍然是一个黑盒子，但最近的文献提出了几种可能的解释，包括消费者的自我选择、羊群行为、对负面经历的过少或没有报道等。后者可能是由于一些原因，如私人接触、互惠、害怕报复，或宣传等（Bridges and Vásquez 2016; Fradkin et al. 2017; Zervas et al. 2015）[4][5][1] 。虽然人们越来越多地讨论这些可能的解释，但很少有人从实证上对它们进行检验、而仅仅停留于理论的解释。

而本文的研究目的就是通过实证分析，探讨平台效应以及动态过程等是如何影响“星级评分”分布。我们的主要观点是，Airbnb 平台的“双边信誉系统”是导致其评分分布与 TripAdvisor（中文名：猫途鹰）这个平台上评分分布的差异的主要原因；Airbnb 评级分数分布受到“幸存者偏差”的影响，即评级较高的房主退出市场的可能性较低导致了这一偏态分布；但上述原因均不能全面解释“偏态分布”。

本文的其余部分组织如下。在第 2 部分中，我们概述了 Airbnb 的背景和其信誉系统；在第 3 部分中，我们介绍了相关文献综述以及理论概念；在第 4 部分中，我们介绍了本文的数据来源以及预处理；在第 5 和 6 部分中，我们基于 Python 进行了相关数据分析。

2 Airbnb 及其信誉系统介绍

Airbnb 成立于 2008 年 8 月，总部设在美国加州旧金山市，是全球共享经济的鼻祖。Airbnb 是一个旅行房屋租赁社区，用户可通过网络或手机应用程序发布、搜索度假房屋租赁信息并完成在线预定程序。目前它已经成为最重要的住宿共享平台。它在全球 65000 多个城市提供了 300 多万个房源（Airbnb，2018）。该平台目前的市场估值为 300 亿美元，与许多酒店行业的现有企业如希尔顿集团（200 亿美元）和万豪酒店（340 亿美元）的市值相当（Forbes，2017）。Airbnb 提供的房源种类繁多，从城市公寓、客房和度假屋，到更具异国情调的房源，如树屋、城堡、冰屋和船屋（Forbes，2016）。

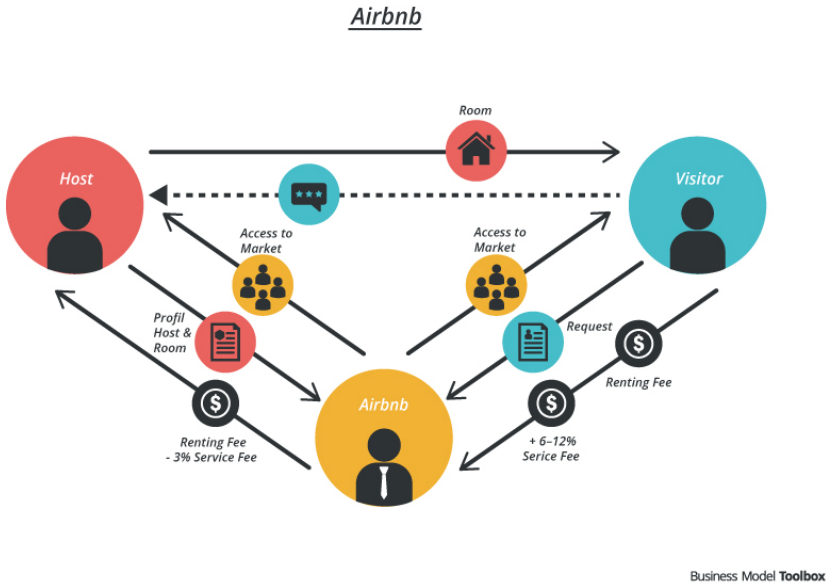

图 1: Airbnb 商业模式

该平台上的用户可分为房东（卖方）和房客（买方）两类。作为典型的互联网共享平台，Airbnb 采用非中心化模式。具体而言，买方在输入搜索请求后，Airbnb 平台会根据这一请求，展示符合买方需求的所有房源即卖方信息。买方可以观察到所有符合条件的卖方信息，甚至可以通过平台提供的搜索算法进一步压缩搜索范围和最终筛选出符合自身要求的卖方，并取得联系最终达成交易。

在这一匹配过程中，爱彼迎只是充当了信息提供者的角色，在信息充分展示后，买卖双方自行决定是否进行交易，而并非由平台统一控制匹配过程。在买卖双方搜寻和匹配的过程中，为提升匹配效率，平台往往会采取不同的策略，包括采用特殊的匹配算法、采用信誉机制（Reputation System）以及构建用户界面（User Interface）等。

而对 Airbnb 来说，它的核心挑战和优势之一就是在这一匹配过程中，采用信誉机制、尽可能地维护用户之间（即房东和房客之间）的信任（Gebbia 2016；Mohlmann 2015）。Airbnb

使用各种机制来建立信任，包括要求用户提供一定程度的个人资料、鼓励用户验证身份等。而以评分和评论为核心的双边信誉系统无疑是实现这一目标最重要的手段。所谓双边，即指房东和房客双方可以互相进行评分或评价，而不仅仅是房客单方面为房东评分。

- **评论：**
  完成交易后，双方有 14 天的时间为此次行程撰写评论。为了促进评论的公正性和真实性，只有有在双方都完成了评价后或 14 天的评价期结束后，评价才会显示在页面上。评论会显示在 Airbnb 网站的各个页面上。它们最显著地出现在用户配置文件页面和单个属性页面上，这些页面按时间倒序排列。

- **评分：**
  即“星级评分（Star Rating）”。在该系统中，完成交易后，房客会给他们的房东打 1 到 5 星的等级。在每个列表的页面上，都会显示所有交易的平均得分并四舍五入为半单元。“星级评分”可分为“总体星级评分”和“分类星级评分”两类，主要参考依据为：

  – 整体体验。整体而言，入住体验如何？
  – 干净卫生。房客觉得居住空间干净整洁吗？
  – 如实描述。房源页面上的内容是否符合实际情况？例如，房客应该能够在房源描述中找到最新的信息和照片。
  – 性价比。房客觉得房源物有所值吗？
  – 沟通交流。在住宿之前及住宿期间，沟通顺畅吗？房客通常会在意房东能否快速、可靠、经常地回复消息和问题。
  – 入住顺利。办理入住的过程是否顺利？
  – 位置便利。房客对所在街区感觉如何？这可能意味着对交通工具、购物中心、市中心等设施的距离远近和前往路线的准确描述，并说明需要注意的特殊事项，例如噪音和家庭安全。
  – 便利设施。房客觉得住宿期间提供的便利设施如何？房客通常会关心列出的所有便利设施是否可用、好用、状态良好。

  房东可以在进度页面的评价下方查看星级评分。但与其他主要旅游点评平台不同，**Airbnb 不会公开与单个点评相关的星级评级，只会公开文本评论以及每个房源的点评汇总统计数据**。即，该评分仅仅代表了一个**聚合的（即，平均值和四舍五入）的值，而无法检索哪个客人提交了哪个等级**。

在一定程度上，Airbnb 的“星级评分”系统是维护用户间信任的最显著、也是最具争议的方法之一。这样的评级对电子 P2P 商务信任的形成以及交易的完全实现起着核心作用（Ert et al. 2016; Zervas et al. 2015）。

图 2: Airbnb 双边信誉系统：评论和评分

3 文献综述及理论背景

论及 Airbnb，许多学者都提到 Airbnb 上的“星级评分”的分布呈现高度偏态，其中大部分评分等于或高于 4.5 分（满分 5.0 分）。最近的文献对这种显著的偏态分布提出了几种可能的解释。

首先，没有或少报负面经历可能会导致过高的“星级评分”（Fradkin et al. ,2017）[5]。少报本身可能有多种原因。“互惠互利”是其中的一种解释，即不管房主和房客的实际经历如何（例如，基于一种以牙还牙的逻辑，或出于对报复的恐惧），这种“互惠互利”都可能会导致双方为对方做出最积极的评价。

其次，与住在某人的公寓有关的高度个人接触和互动可能会减少过于消极与苛刻的评价。当我们与某人有过一定亲密的相处后，若再提交负面评价可能会带来社交上的尴尬，因此用户往往会收回他们的负面评价甚至投诉(Bridges & Vasquez , 2016; Ikkala & Lampinen, 2015）[4][2] 。

第三，评论的公开可用性可能会阻止房客提交负面评论，因为他们会因此间接承认自己对住宿选择的错误判断（Bridges and Vasquez 2016）[4] 。此外，从战略角度来看，所有过于苛刻的评分都可能损害房客被未来的东道主接纳的机会。但事实上，只有文本评论是公开显示的，并且直接指向作者；而星级评分仅仅代表了一个聚合的（即，平均值和四舍五入）的值，而且无法检索哪个客人提交了哪个等级。显然，Airbnb 已经在尝试使一些用户的个人星级评分现在会显示出来，但至少到目前为止，“星级评分”并不直接指向某一客户。

另有学者指出,Airbnb 是一种相当新颖的消费模式,大部分客户都是首次交易(Teubner, 2017）[6] 。与住酒店相比，Airbnb 上的客人可能根本不知道该期待什么，因此没有很高的期望或标准，可能会失望。这可能会使评分向积极的一端倾斜。Ert et al.[3] 曾比较 Airbnb 和 Booking.com 平台下欧洲城市的房源评分，结果发现，相同房源下爱彼迎的平均分数有

约高出 20%（Ert, E. et al., 2016）。

基于此，本文基于“星级评分”的动态过程考虑了另一个可能解释。我们认为，即使 Airbnb 这一平台上“星级评分”的“初始”分布呈现较弱的偏态分布，该平台上独特的“幸存过程（Survivorship Process）”最终也会导致大比例的高评级房源。也就是说，高评级房源的巨大份额可能部分归因于这些房子“生存能力”的增强。以下是这一动态过程的理论基础。

3.1 Markov 链

马尔可夫链（Markov 链）为**状态空间中经过从一个状态到另一个状态的转换的随机过程**，是由俄国数学家马尔可夫于 1906 年提出的。**运用马尔可夫链只需要最近或现在的动态资料便可预测将来**。基于 Markov 链马尔科夫过程是**研究离散事件动态系统状态空间**的重要方法，它的数学基础是随机过程理论；而基于 Markov 链的马尔科夫预测法是一种预测事件发生的概率的方法，它是**根据事件的目前状况预测其将来各个时刻（或时期）变动状况的一种预测方法**。此处我们选用 Markov 链来模拟并预测无规则的房源“星级评分”的动态过程。以下是两个相关概念：

- **无记忆性：** Markov 链假定下一状态的概率分布只能由当前状态决定，在时间序列中它前面的事件均与之无关。
- **状态转移：** 在马尔可夫链的每一步，系统根据概率分布，可以从一个状态变到另一个状态，也可以保持当前状态。**状态的改变叫做转移**，与不同的状态改变相关的概率叫做**转移概率**。
- **平稳分布：** 无论采用何种初始概率分布，马尔科夫链模型的最终状态收敛到同一个稳定的概率分布，即平稳分布。**平稳分布提供了使马氏链和初始状态无关的一个办法，并刻画了马氏链在长时间下的极限行为和平均行为。它只与初始状态和转移概率矩阵有关**。

3.2 幸存过程和幸存者偏差

幸存者偏差（Survivor Bias）是一种认知偏差。其逻辑谬误表现为只能看到经过某种筛选而产生的结果，而没有意识到筛选的过程，因此忽略了被筛选掉的关键信息。考虑到大多数市场并不代表静态结构，而是处于一个稳定的变化过程，因此，当我们考虑市场参与者和参与者类型时，该规律具有普遍性。

3.3 信号传递理论

该理论假设买卖双方存在利益冲突和信息不对称，而卖方（或者更一般地说，信号提供者）可以通过提供信号来减少信息不对称和相关的不确定性。例如这里可以将第三方的评级

作为信号。房东通过“星级评分”像买者提供一个信号：“星级评分”较高传递了“高声誉”的信号，“星级评分”较低则传递了“低声誉”的信号。这将有助于相关房源获得消费者的信任，从而获得经济上的成功。

3.4 大数定律和均值回归

除了上述理论，我们认为 Airbnb 房源的信誉动态也符合相当普通的统计现象：大数定律和均值回归。

- **大数定律：**当重复做同一实验的次数足够多时 (即)，这些结果的平均值收敛于期望值。应用到 Airbnb 房源的评级上，这表明，评论数量不断增加的房源获得最高评级的可能性会降低。
- **均值回归：**如果一个变量在第一次测量时是极端的，那么它在第二次测量时就会趋向于接近平均值。较高的信誉评级可能会引起客户对一个特别好的体验的期望，这可能会有效地增加失望的机会，反过来，促进较低的客户评级。

4 数据来源及预处理

4.1 Airbnb 数据来源

本文的基础数据来源于 InsideAirbnb.com (Cox 2017; Wired 2017) ，该网站为我们提供了共 44 个城市中可以被检索到的所有房源的详细数据，包括诸如房源价格、房间类型、评分和评价数量等变量。

为增加样本量以及时间跨度、从而使结论更具普遍意义，**本文选取了美国 New York、San Francisco、Austin、Los Angeles 等 4 个城市在 2018 年 4 月-2019 年 9 月共 126887 个房源的月度数据**，这 4 个城市基本可认定为 Airbnb 在美国的房源数量最多的几大城市。但由于 Austin 和 San Francisco 的在 2018 年 6 月的数据缺失，为了保持时间的连贯性，我们把 Austin 和 San Francisco 的 2018 年 5 月的数据各自复制了一份，用作这两个城市当年 6 月的数据。该原始数据共包括 108 个变量, 具体说明可见表 1 。

4.2 TripAdvisor 数据来源

考虑到 Airbnb 是作为一个直接与酒店竞争的住宿平台，它与酒店的评分的比较或许可以为我们的主题提供线索。因此，我们爬取了另一家主要旅游网站 TripAdvisor（中文名：猫途鹰）下的相关数据。

TripAdvisor 成立于 2000 年目前已成为全球最大的旅游点评媒体。该网站的网络预订功能不是很突出，其优势在于源自 TripAdvisor 全球超过 1000 万注册会员的海量评价信息

表 1: 2018 年 4 月-2019 年 9 月 4 座城市的 Airbnb 房源的原始数据说明
数据来源：http://insideairbnb.com/get-the-data.html

城市	起始时间	结束时间	文件数目（真实）	文件数目（扩充后）	单文件大小（随机采样值）
New York	2018/4	2019/9	18	18	180MB
San Francisco	2018/4	2019/9	17	18	30MB
Los Angeles	2018/4	2019/9	18	18	170MB
Austin	2018/4	2019/9	17	18	45MB

及超过 3 亿条旅游评论，让你看到即将入住房子的真实性。而且相较 Airbnb 较为单一的房屋类型，TripAdvisor 下的房源类型更具多样性。

具体爬取过程的步骤如下：

（1）导入 Requests（用于请求网页）、Time（用于获取当前时间以及等待多长时间后执行程序）、Re（正则表达式，获取网页中特定内容）、BeautifulSoup（通过 XPath 获取网页中指定内容）、Pandas（数据读取与储存）和 Numpy(数据处理) 等工具包；

（2）通过定义 get_hotel_url() 函数，传入酒店目录链接，从而获取该页面中所有的 hotels 链接；

（3）用 get_hotel_rental_inf() 函数，传入具体的 hotel 链接通过正则表达式及 BeautifulSoup 的 Selector 函数获取该酒店的具体信息；

（4）运用主函数，获取当前时间，循环爬取从该城市的第一页目录到最后一页，最后保存为 csv 文件。

作为结果，我们共爬取了该网站上在 2019 年 12 月 6 日的共 2340 家酒店（Hotel）（包括民宿（B&B））和 4286 家度假出租房（Vocational Rental）数据。表 2 和表 3 是对爬取的原始数据变量的说明。

4.3 Airbnb 数据集预处理

（1）整合 72 张分表

为整合 4 座城市历时 18 个月的分表，我们先在每张表上都创建两个新变量，分别为 date（比如 201909）与 city（比如 Austin）；同时除去重复变量：即 listing_id 相同的记录只取一条。此时，我们便以 id、date 和 city 作为某一时间段、某一地点某个房源的唯一标识。

（2）剔除无关变量及残缺变量

考虑到本文实证的需要，我们剔除了无关变量，并记录了剩余 43 个所需变量及其残缺值数量。同时去除 listing_id、room_type、property_type、price、latitude、longitude 等变量中有残缺的记录。具体如表 4 ：

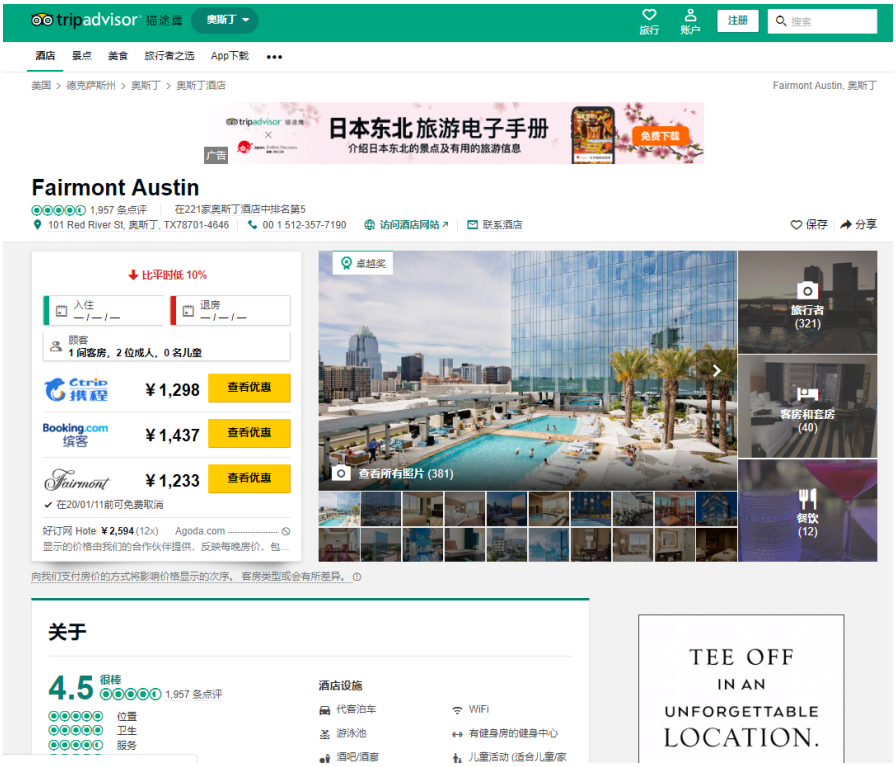

图 3: TripAdvisor 酒店属性页

表 2: 2019 年 12 月 6 日爬取的度假出租屋的原始数据变量说明
数据来源：https://www.tripadvisor.com/Hotels

变量名	含义	变量名	含义
title	酒店名称	owner_name	房东
rate	评分	owner_inf	房东信息
review	评论数	amenities	设施
id_locid	位置 id	overview	概述
latitude	纬度	description	描述
longitude	经度	rules	住房规则

表 3: 2019 年 12 月 6 日爬取的酒店和民宿的原始数据变量说明
数据来源：https://www.tripadvisor.com/Hotels

变量名	含义	变量名	含义
title	酒店名称	good	好
rate	评分	ordinary	一般
Location_rate	位置评分	bad	较差
Cleanliness_rate	卫生评分	verybad	很差
Service_rate	服务评分	features	房间特点
Value_rate	性价比评分	types	房间类型
description	描述	low_price	最低价
location	位置	high_price	最高价
rank	排名	latitude	纬度
review	评论数	longitude	经度
great	很棒	locationid	位置 id

（3）异常值处理

考虑到各类评分的数值范围，我们分别剔除了 review_scores_rating 的变量值不属于 [20,100] 的记录以及 review_scores_accuracy、review_scores_cleanliness、review_scores_checkin、review_scores_communication、review_scores_location、review_scores_value 的变量值不属于 [0,10] 的记录。最后发现有关评分的所有数值都无异常值。同时考虑到后文对房源匹配是需要准确的位置信息，所以去除 is_location_exact 中非“t”（t 表示正确）的记录。

（4）创建“星级评分”变量（star_rating）

参考 Teubner et al.（2018）对 review_scores_rating 到 star_rating 的转换方法，我

表 4: 爱彼迎总数据集中变量及残缺值数目说明

列名	缺失值数量	列名	缺失值数量
listing_id	0	accommodates	0
name	137	bathrooms	1780
summary	36738	bedrooms	841
space	273567	beds	546
description	4606	bed_type	0
neighborhood_overview	392118	amenities	0
notes	616249	square_feet	1250667
transit	385918	price	1055091
access	410368	weekly_price	0
interaction	429108	monthly_price	0
house_rules	382126	number_of_reviews	0
host_id	0	review_scores_rating	0
host_name	546	review_scores_accuracy	0
host_about	399573	review_scores_cleanliness	0
neighbourhood	53812	review_scores_checkin	0
market	3241	review_scores_communication	0
latitude	0	review_scores_location	0
longitude	0	review_scores_value	0
is_location_exact	0	reviews_per_month	37
property_type	0	date	0
room_type	0	city	0
star_rating	0		

们创建了新变量 star_rating。以下是具体转换规则：

（5）形成总表和分表

在上述的基础上，我们形成表“总 _ listings.csv”，共 1268644 个房源。之后，我们提取 2019 年 9 月的房源数据，成表“201909_ listings.csv”；同时，提取不同 city 的数据，依次成表“201909_ 城市 _ listings.csv”。下表是关于“201909_ listings.csv”中部分变量的描述性统计。

4.4 TripAdvisor 数据集预处理

对于 TripAdvisor 爬取数据的处理思路与上述类似：

表 5: review_scores_rating 转化为 star_rating 的规则说明

star_rating	review_scores_rating	star_rating	review_scores_rating
1	[20,24]	3.5	[65,74]
1.5	[25,34]	4	[75,84]
2	[35,44]	4.5	[85,94]
2.5	[45,54]	5	[95,100]
3	[55,64]		

表 6: “201909_ listings.csv” 数值变量的描述性统计

Variable	Mean	Std	Min	25%Q	Median	75%Q	Max
latitude	36.843	3.728	30.119	34.046	37.731	40.712	40.915
longitude	-97.451	21.386	-122.513	-118.356	-97.759	-73.96	-73.717
accommodates	3.45	2.425	1	2	2	4	32
bathrooms	1.307	0.735	0	1	1	1	15.5
bedrooms	1.345	0.976	0	1	1	2	22
beds	1.864	1.572	0	1	1	2	50
number_of_reviews	39.54	62.927	0	4	14	47	900
review_scores_rating	94.548	8.077	20	93	97	100	100
review_scores_accuracy	9.661	0.79	2	10	10	10	10
review_scores_cleanliness	9.4	0.988	2	9	10	10	10
review_scores_checkin	9.783	0.669	2	10	10	10	10
review_scores_communication	9.774	0.695	2	10	10	10	10
review_scores_location	9.655	0.706	2	9	10	10	10
review_scores_value	9.445	0.875	2	9	10	10	10
reviews_per_month	1.687	1.898	0.01	0.27	1	2.55	62.5
star_rating	4.767	0.429	1	4.5	5	5	5

（1）分别整合酒店和度假出租屋的数据

我们先在每张表上都创建两个新变量,分别为 date(代表爬取数据的时间,比如 20191202) 与 city（比如 Austin)；同时除去重复变量：即 title 相同的记录只取一条。此时，我们便以 title、date 和 city 作为某一时间段、某一地点某个房源的唯一标识。

（2）剔除残缺变量

本文记录了爬取的变量及其残缺值数量，如下所示。此后，我们去除了 rate、number_of_reviews、latitude、longitude、description 中有残缺的记录，同时将 “nan/none” 这个字符也视为残缺值。

（3）异常值处理以及创建“星级评分”变量（star_rating）

首先对变量 rate 的处理。在该列下，分别有：[0-50] 区间内取值的评分，[0-50] 区间内

表 7: TripAdvisor“度假租赁房”数据集中变量及残缺值数目说明

列名	缺失值	列名	缺失值
title	0	description	0
rate	2858	rules	0
number_of_reviews	2858	owner_name	0
id_locid	342	owner_inf	151
latitude	21	amenities	42
longitude	21	date	0
overview	0	city	0

表 8: TripAdvisor“酒店”数据集中变量及残缺值数目说明

列名	缺失值	列名	缺失值
title	0	great	301
rate	301	good	301
Location_rate	507	ordinary	301
Cleanliness_rate	507	bad	301
Service_rate	507	verybad	301
Value_rate	507	features	63
description	593	types	164
location	0	low_price	608
rank	303	high_price	608
number_of_reviews	269	latitude	1
locationid	0	longitude	1
date	0	city	0

取值的评分，部分字符串、如 star_5 的，star_45。对于第一种情况，我们将相关数值除以 10；对于字符串这种情况，我们只保留数值。最后将其放入新变量 star_rating 中，使得该变量下数据的取值范围在 [0,5]。 其次是对 number_of_reviews 的处理，由于该变量下有字符串的形式（比如，1 review），我们将其处理为只保留数字的变量。

（4）形成总表和分表

在上述的基础上，我们形成表“总 _Hote.csv”，共 1386 个房源；以及表“总 _Rental.csv”，共 1211 个房源。同时，对于 “总 _Hotel.csv” 这个表，将 title 变量中包含字符串 “Bed And Breakfast” 的提取出来，成表 “总 _TripBB.csv”，作为 TripAdvisor 上民宿数据；将剩余

纯酒店的部分，另成表“总 _TripH.csv”。此后，分别形成各城市的表“城市 _Hotel.csv”和“城市 _Rental.csv”。以下是关于“总 _ Hotel.csv”和“总 _Rental.csv”中部分变量的描述性统计。

表 9: “总 _ Rental.csv”数值变量的描述性统计

Variable	Mean	Std	Min	25%Q	Median	75%Q	Max
star_rating	4.564	0.794	1	4.5	5	5	5
number_of_reviews	12.225	21.878	1	1	4	12	213
id_locid	11598479	4282832	1306225	8695528	11564464	15370456	19489568
latitude	34.62	5.199	-56.108	30.272	34.061	40.723	42.587
longitude	-95.518	18.692	-122.508	-118.221	-97.747	-73.996	169.954

表 10: “总 _Hotel.csv”数值变量的描述性统计

Variable	Mean	Standard deviation	Minimum	25%Q	Median	75%Q	Max
star_rating	3.877	0.74	1	3.5	4	4.5	5
Location_rate	42.974	6.412	10	40	45	45	50
Cleanliness_rate	41.4	7.299	10	40	45	45	50
Service_rate	40.374	6.887	10	40	40	45	50
Value_rate	39.057	6.214	10	35	40	45	50
number_of_reviews	1156.535	1863.428	1	108.75	536.5	1428.75	24998

5 数据分析 1：TripAdvisor 与 Airbnb 房源的“星级评分”

本文首先对 Airbnb 上 2019 年 9 月和 TripAdvisor 上 2019 年 12 月的“星级评分”的分布进行实证研究。由于 Airbnb 的最新数据仅截止于 2019 年 9 月，而我们做该作业的时间已到 2019 年 12 月，所以存在 3 个月的时间差。虽然 3 个月对两个平台上的房源的评分不会构成非常大的影响，但仍在此处指出，读者在阅读以下结果时可自行判断。

5.1 TripAdvisor 与 Airbnb 房源“星级评分”的总体比较

这部分我们对 TripAdvisor 上不同类型房源与 Airbnb 的房源的“星级评价”分布先进行总体比较。我们选择用 Pyecharts 库中的 Pie 函数来做“radius 型玫瑰图”，使得数据可

以更清晰化地呈现，如图 4 。

图 4 的左上角显示了 Airbnb“星级评分”的分布情况，可以在发现这 4 个城市中，Airbnb 房源的评分极高：超过一半（66.67%）的房源拥有最高的 5 星评级，92% 的房源评分为 4.5 星及以上，评分均值 4.77，呈现严重偏态。但观察右上角猫途鹰网站上酒店的评分分布，显然没有那么偏态：只有 4.62% 的酒店拥有最高的五星评级，35.08% 的酒店评分为 4.5 星及以上，评分均值仅为 3.88。

产品异质性可能是造成上述房源评分巨大差异的一个潜在原因。我们在比较 TripAdvisor 上 B&B 和 Vocational Rental 两类产品的“星级评分”时，它们在视觉上和统计上的差异却并没有那么大。从均值来看，两类产品与 Airbnb 上房源的均值更加接近：B&B 的评分均值为 4.42，Vocational Rental 的评分均值为 4.56。但不可否认的是，这三者在尾部分布上仍存在差异：相较于 Airbnb 上 92% 的房源“星级评分”在 4.5 星及以上，仅有 76.92% 的 B&B 和 79.03% 的 Vocational Rental 的评分在这个区间。由此，我们可以得出的一个基本观察是，即便是在一个平台上，评分均值也会明显受到产品组合的影响。

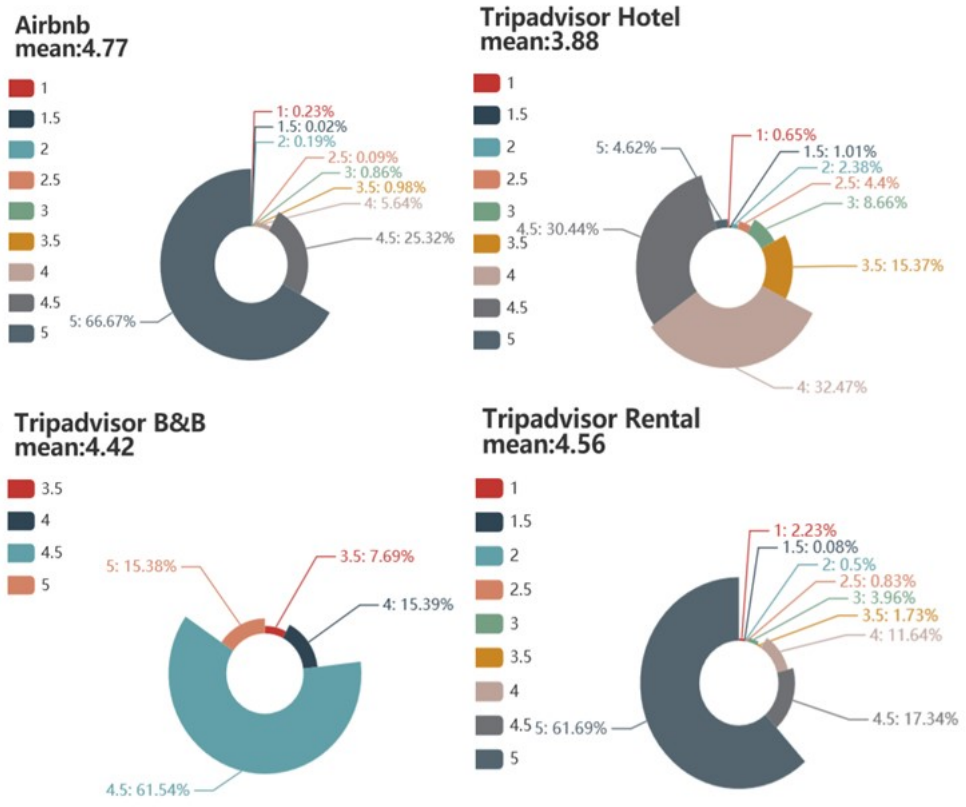

图 4: TripAdvisor 上不同类型房源与 Airbnb 的房源的“星级评价”百分比

5.2 TripAdvisor 与 Airbnb 房源“星级评分”的具体比较

基于上述 Airbnb 房源“星级评分”的严重偏态性这一观察，我们提出了另一种可能：在特定细分领域，这个评分分布的偏态性会不会减弱？为了更好地了解房产评级分布背后的潜在异质性，我们根据不同的房源属性对 2019 年 9 月 Airbnb 平台上的房源进行分类研究。

5.2.1 不同房源类型对评分的影响

首先，我们根据 Airbnb 上不同的房源类型，对其“星级评分”的分布进行研究。考虑到房源类型有很多，我们选出了最具代表性的（房源数多）的五种类别，利用 Seaborn 中的 Catplot 按房源类型分类作图，如图 5 所示。据此，我们发现不同的房源类型并不能多大程度上改变“星级评分”的偏态性：虽然个人占有公寓（Condominium）、客房（Guest suite）和住宅（House）的“星际评分”高于公寓（Apartment）和 B&B，但总体来看，这些类型的房源评分为 4.5 星及更高的比例至少为 90%。

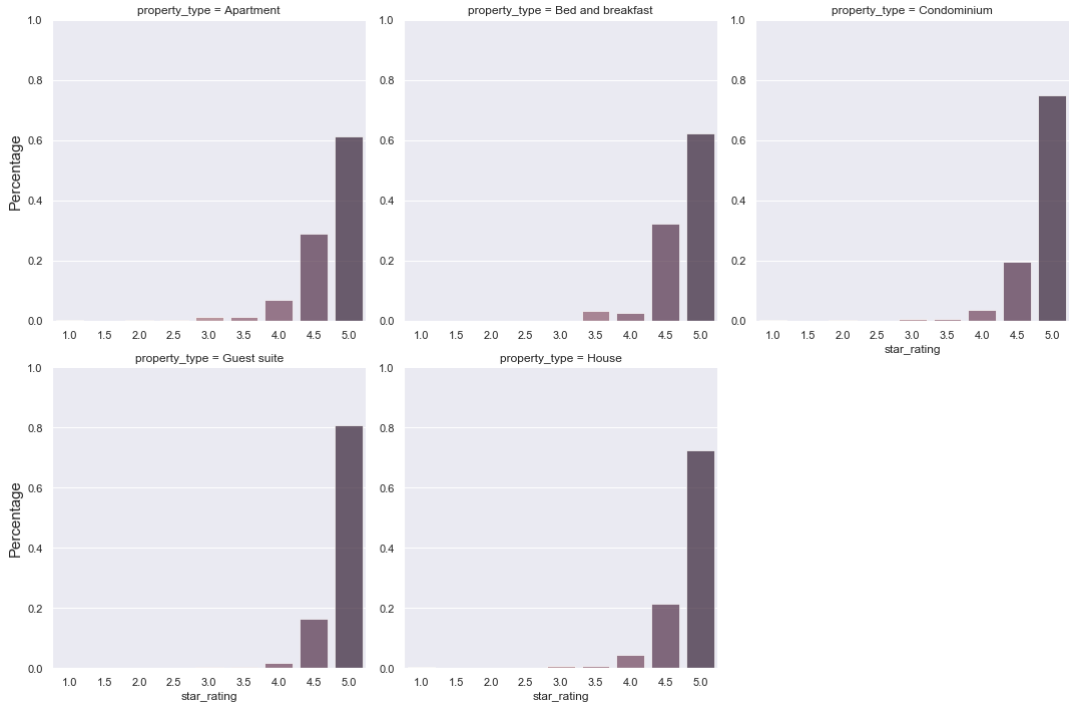

图 5: Airbnb 上不同房源类型的评价等级分布

5.2.2 不同地理位置对评分的影响

其次，我们根据美国这 4 个城市所在的地理市场，绘制房源“星级评分”的分布图。我们通过 Pyecharts 中的 Radar 函数精确地绘出四个城市 Airbnb 和 Tripadvisor 上“星级评

分”分布的雷达图，如下图 6 所示。

我们发现，对于 Airbnb 上的房源来说，4 座城市 4.5 星和 5 星评级相对频率发生了一定的变化：Austin 的评分总体较低，星级评分集中于 4.5 星，而 5 星几乎没有；其余三个城市五星级评分最多，San Francisco 的五星级评分分布最密集。但四个城市的评分仍呈现出严重的偏态分布，评分在 4.5 星或以上的比例仍然很高。而 TripAdvisor 上酒店（包括 B&B）的“星级评分”也在不同城市显示出了较大的差异。

在 5.1 部分，我们提到两个平台的“星级评分”均值差约为 1 颗星；在这部分，我们发现这种均值差在不同城市也有较大的差异。在我们绘制的 4 个城市中，San Francisco 的差异最大，达 1.13 颗星；纽约的差异最小，为 0.63 颗星。

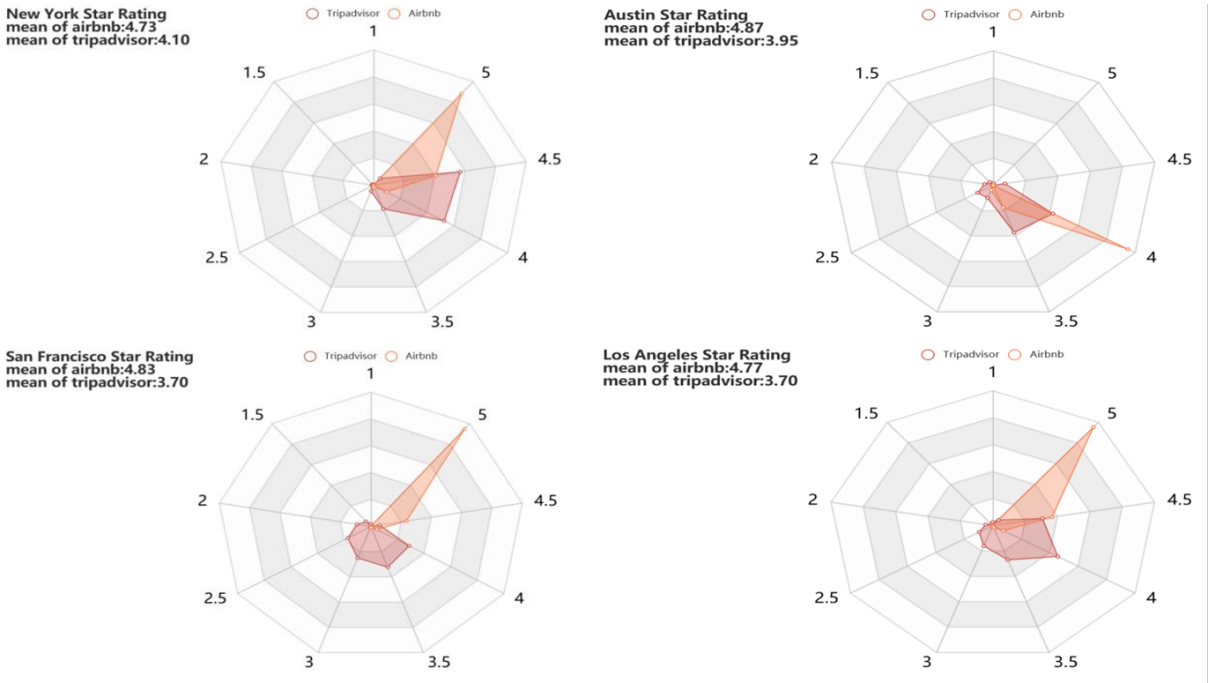

图 6: 四个城市 Airbnb 和 TripAdvisor 上“星级评分”分布的雷达图

5.3 TripAdvisor 和 Airbnb 上相同房源“星级评分”的比较

在上述比较中，使用两个平台上房源不尽相同，这种差异性可能会对评分产生影响。为了消除这方面的差异性、更好地了解这些跨平台房源评分差异的根源，我们通过匹配进而找出 Airbnb 和 TripAdvisor 这两个平台上相同的房源，在尽可能控制房源差异性的情况下，进一步进行研究。

5.3.1 Tripadvisor 和 Airbnb 上相同房源的匹配

将 Airbnb (201909_listings.csv) 与 TripAdvisor (总 _ Hotel.csv 和总 _Rental.csv) 这两个文件进行匹配。(注：两个文件中保留的变量不同)

(1) 根据经纬度匹配初步匹配地理位置相近的房源

首先采用经纬度 <=500m 的匹配模式。经验证，Airbnb 的经纬度精确度为小数点后 12 位左右，TripAdvisor 的经纬度精确度为小数点后 5-6 位。若小数点后前 5-6 位相同，Airbnb 和 TripAdvisor 的距离最多相差 1.4168m，可以忽略不计。同时考虑到运算时间问题，匹配时使用了每个城市的分表而没有使用总表。

使用 Geopy 工具包中的 distance 可以通过经纬度来计算距离，再筛选出两房源位置小于 500 米的做为一组匹配数据。

(2) 根据字符串匹配在地理位置相近的房源中精确匹配相同房源

在房源地理位置相近的基础上，通过对房源的描述关键词的相似度进行计算，从而确定是否为统一房源。

利用 Jaccard 算法来计算两相近房源关键词字符串的 Jaccard 系数。Jaccard 值的定义为 A 与 B 交集的大小与并集大小的比值：

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Jaccard 值越大说明相似度越高，当相似度超过一定程度基本可以确定是两者是同一房源。

其中，Jaccard 算法的优点有：

- 时间复杂度足够低，能够在有限的时间内计算出结果。
- 原理简洁易懂，易于解释。
- 比较通用，提供了足够支持我们工作的可靠性。

(3) 测试相似度阈值

相似度的阈值需要通过测试来确定，且阈值的确定对数据结果的影响极大：设置大了，容忍度提高，造成误判；设置小了，造成漏判。

在确定阈值时，分别计算各个城市两组数据的中位数和平均值等统计数据，并进行比较，通过比较两组数据统计信息的一致性来判断两组数据集合中的数据是否一致。

经过仔细筛查过后，选取了 0.7 作为阈值，接着人工验证了 50 个左右的匹配数据，确认取相似度阈值为 0.7 符合要求且较为准确。

(4) 筛选过程及结果总结

表 11 呈现的是坐标匹配和描述匹配后剩余的房源数。最后匹配的结果为 32 个旅游租赁住房。

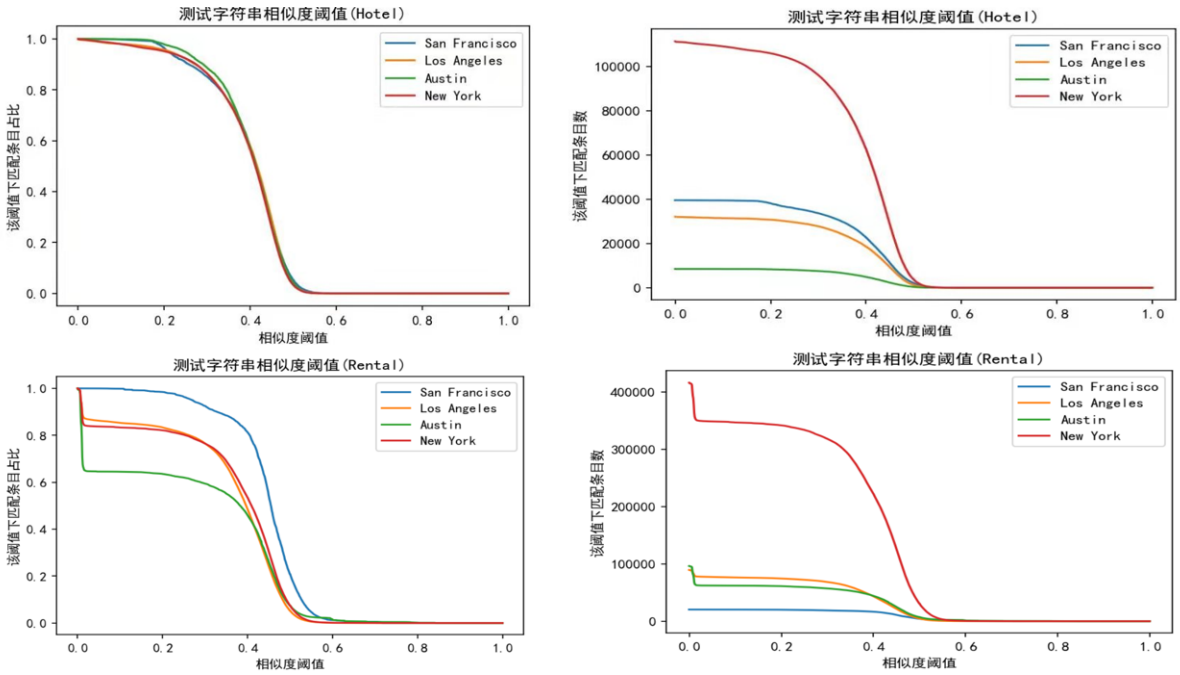

图 7: 相似度阈值测试

表 11: 每个筛选阶段剩余房源数

	坐标匹配 (Hotel)	描述匹配 (Hotel)	坐标匹配 (Rental)	描述匹配 (Rental)	去除没有评论的房源后 (总剩余房源数)
Austin	9880	0	96763	447	15
Los Angeles	38318	0	90370	7	5
New York	133122	0	421154	105	11
San Francisco	44077	0	20376	111	1

5.3.2 Airbnb 与 TripAdvisor 相同房源的星级评分比较

为比较两个平台共有的 32 家房源的评分分布，我们选用 Pyecharts 库中的 Pie 函数来做“area 型玫瑰图”，即每个部分的弧度相同，面积大小来表示占比，如图 8 所示。可以看到，在 Airbnb 上，4.5 星及以上的房源占比高达 95.75%，而在 TripAdvisor 上该比例仅为 65.62%；前者房源的均值为 4.75，而后者仅为 4.58。可以看到，相同房源在 Airbnb 的评分高于 TripAdvisor 上的评分。

由上可得，仅去除“房源的差异性”并不能解释两个平台间的“星级评分”偏差。这或许与“平台效应”有关。结合本文文献综述部分的“互惠原则”和部分学者的观点，我们猜

测由于 Airbnb 使用双边审核系统（即，房客可以对房东进行评分，同时房东可以对房客进行评分），而在 TripAdvisor 上，只有房客才能对房东进行评分，导致了两个平台间的评分差异。

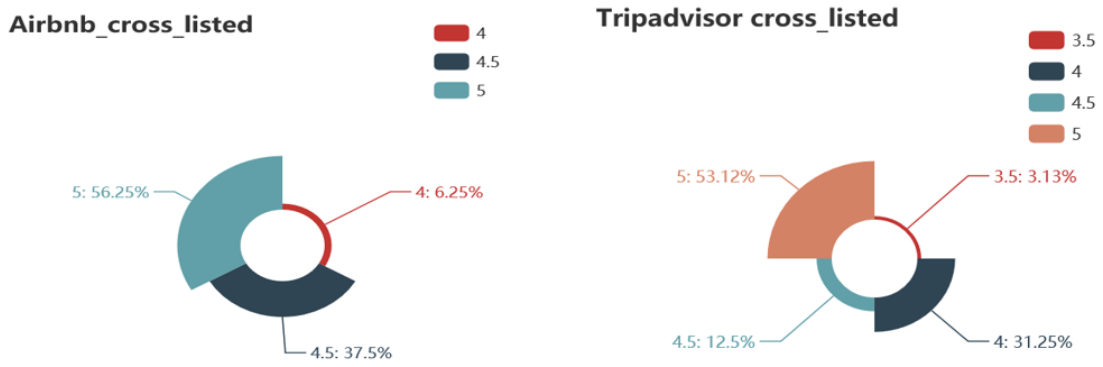

图 8: Airbnb 与 TripAdvisor 相同房源的星级评分比较

5.3.3 Airbnb 对 TripAdvisor “星级评分” 预测

上述的讨论主要集中在了解 TripAdvisor 和 Airbnb 评分分布的差异上，但与此同时，两个平台房源的相对排名关系也是值得研究的。例如，考虑 Airbnb 和 TripAdvisor 上列出的两个房源。假设在 Airbnb 上，房源 A 的评级高于房源 B，这种情况也适用于 TripAdvisor 吗？更广泛地说，就是在一个平台上的评分能够预测对方的评分吗？

为了回答这个问题，我们使用匹配出来的 32 个房源的数据，用 Airbnb 的星级评分来预测 TripAdvisor 的星级评分。虽然 TripAdvisor 的评分普遍低于 Airbnb 上的评分，但前者的评分仍然可以在原则上完美预测 Airbnb 的评分 (反之亦然)。例如，从直觉上来讲，TripAdvisor 和 Airbnb 的用户可能有相似的品味，但他们对 “5 星评分标准” 的解读不同，TripAdvisor 的用户给出了更严格的评分。因此，这些绝对数值分布的差异并不预先决定这部分的分析结果。

我们们以匹配出来的 32 个房源作为分析对象、运用 Statsmodels 中的 OLS 线性回归模型进行多元线性回归，回归结果如表 12 所示。我们首先以 Airbnb 上的星级评分为因变量，TripAdvisor 上的星级评分为因变量进行回归。我们发现，两个平台上相同的房源的评分之间存在着较为显著的正相关，TripAdvisor 的每一颗星的增长大致相当于 Airbnb 的四分之一颗星的增长；但同时，我们可以看到调整后的 R^2 较低 (0.281)，表示一个平台上的评分只能解释另一个平台上评分的微小变化。

但上述分析的一个问题是，我们正在比较不同地理位置和价格区间的房源。然而，大多

数买者会将他们的房源搜索限制在目标预算内的特定地点。因此，虽然评分并不能进行全面的预测，但在界定严格的细分市场中，评分可能具有更强的解释力。例如，TripAdvisor 的用户可能更喜欢价格较高的住宿，而 Airbnb 的用户则更偏爱价格低的房源。然而，在比较每个价格区间内的房地产时，它们的相对偏好是相同的。受此启发，我们在表 12 的第（2）和（3）列中依次加入 city 虚拟变量和 price。我们发现，加入价格变量后，调整后的 R^2 下降，说明解释力度下降；而加入 city 虚拟变量后，的 R^2 上升至 0.401，说明评级的联合分布受到了地理位置的较大影响。总体而言，这些结果表明，虽然 Airbnb 上评级较高的房产在 TripAdvisor 上的评级平均也较高，但在两个平台上，甚至在界定严格的细分市场中，评级的联合分布受到地理位置的影响，但也存在大量无法解释的变化。

表 12: Airbnb 对 TripAdvisor“星级评分”的预测

	(1)	(2)	(3)
TripAdvisor Rating	0.035*	0.036*	0.025
	(0.010)	(0.010)	(0.010)
City Dummies	No	No	Yes
Price Dummies	No	Yes	No
N	32	32	32
R^2	0.302	0.308	0.401
$Adj.R^2$	0.281	0.260	0.312

注：(1) 因变量：32 个相同房源在 Airbnb 上的星级评分。

(2) 显著性水平。$*p < 0.1, **p < 0.05, ***p < 0.01$

5.3.4 Airbnb 与 TripAdvisor “排名” 关系

这一部分，我们将注意力转移到两个平台相同房源的排名而不是它们的“星级评分”。这种非参数的比较可以作为一种稳健性检验，因为消费者会使用不同的“排名算法”，即评分往往是相对的而不是绝对的。他们会对更喜欢的房源打出相对高的评分、对不喜欢的房源打出相对低的评分，但不一定会为这种高分和低分规定一个具体的、绝对的数值范围。基于上述考虑，我们对“星级评分”做出如下调整：

（1）根据“星级评分”，对 32 个房源在每个平台上分别进行排序：“星级评分”越高，房源的排名越靠前；

（2）当出现两个房源“星级评分”相同时，我们使用评论的数量再进行排序：评论数量

越多，房源的排名越靠前。这种判断是与直觉相吻合的，即对于消费者来说，拥有 100 条评论的五星级酒店可能比只有一条评论的五星级酒店风险更小。在此基础上创建新的 rank 变量；

（3）定义 cal_tau() 函数用于计算 Kindall 相关系数，并画出两个平台上 rank 变量在所有城市以及 3 个城市的相关性分析图。如图 9 所示。由于 San_Francisco 仅剩下一个房源，没有出现在图中。

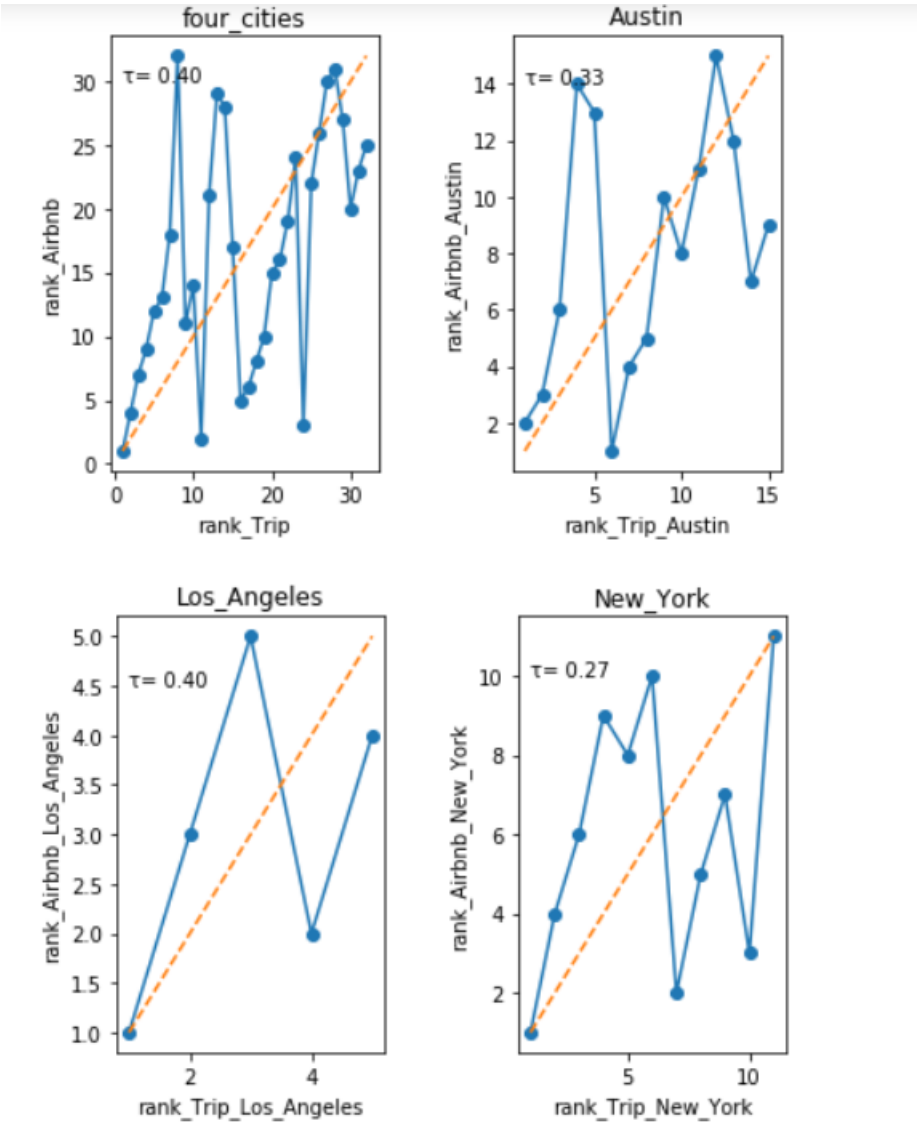

图 9: 32 个相同房源在 Airbnb 和 TripAdvisor 上的排名关系

上述使用的 Kendall 相关系数主要用于测量评分数据一致性水平。通常情况下：该取

表 13: 2018 年 4 月-2019 年 9 月 Airbnb 房源的进入退出情况

listing_id	201901	201902	201903	201904	201905	201906	201907	201908	201909
109	80	80	80	80	80	80	80	80	80
344	93	93	93	93	93	93	93	93	93
958	97	97	97	97	97	97	97	97	97
2515	93	0	0	0	0	0	0	0	0
2732	96	96	96	96	96	96	96	96	96
3021	81	81	81	81	81	81	0	0	0
3330	97	97	97	97	0	0	0	0	0

值 < 0.2 说明一致性程度较差；0.2−0.4 之间说明一致性程度一般；0.4−0.6 之间说明一致性程度中等；0.6−0.8 之间说明一致性程度较强；0.8−1.0 之间说明一致性程度很强。在图 8 中，我们观察到该相关系数的取值主要在 0.2−0.4 之间，说明两个平台的相关性一般，即在这两个平台上，房源的相对排名具有一定的差异。

6 数据分析 2：Airbnb 上房源的相关动态过程

显然，上节两个平台的评分比较仍然不能充分说明 Airbnb 评分偏态的问题，因此我们从一个新的角度——房源的相关动态过程来解释影响 Airbnb 上“星级评分”的“偏态分布”的可能原因。

6.1 Airbnb 上房源进入退出的基本概况

为研究 Airbnb 上各月房源的动态情况,我们通过分别提取总表中各月的 review_scores_rating 和 date 两个变量，再通过 Pandas 中 Merge 函数，形成 2018 年 4 月-2019 年 9 月所有房源的面板数据。数据结构如表所示。

在此基础上，我们将考虑每个月 Airbnb 平台上房源的数量以及进入、退出的情况，如图 **??** 所示。总体而言，这四座城市在该时间段内的平台上的房源数量稳定在 65000 以上，并呈现上升趋势。而每月市场房源进入和推出数基本保持一致，约占 7%。

6.2 Airbnb“星级评分”的 Markov 动态模拟

关于各个房源“星级评分”的动态，我们采用 Markov 链进行模拟。我们将一个房源的“星级评分”作为一种状态, 考虑房源不同的“星级评分”状态之间的转换。我们将所有 3.5 星或以下的评分都被聚合为一个联合状态，所以共有 4 类状态，分别是：≤ 3.5，4，4.5,5。

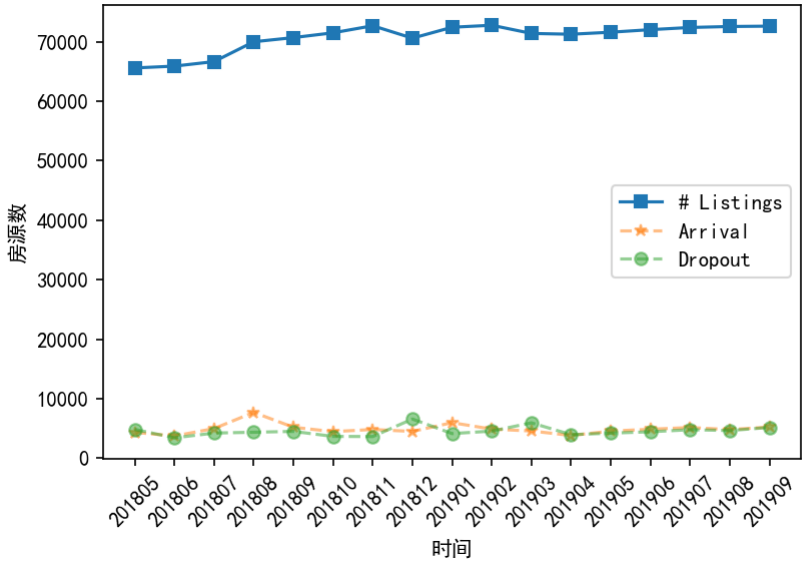

图 10: Airbnb 上 2018 年 5 月-2019 年 9 月房源的动态过程

我们利用 Pandas、Operator 和 Numpy 三个工具包，计算 2018 年 4 月-2019 年 8 月 4 个城市所有房源的状态转移概率。比如，这些概率可以表示，一个拥有 4.5 星的房源在接下来的一个月内将其评级提高到 5.0 星或降低到 4.0 星的可能性有多大。此外，这还会得到某星级的房源退出的概率，以及新进入的房源到达某一个星级的概率。在图 11 中，我们总结了“星级评分”间的状态转移概率概况。可以看到，对于某评分星级，它在下一个月仍然保持原有“星级评分”的概率是最高的。一个新出现在 Airbnb 上的房源有评级为 5.0 星的概率为 5.8%，4.5 星为 1.5%，4.0 星为 0.9%，3.5 星或以下为 0.5%(数值以蓝色显示)。在退出率方面，较低的评级分数伴随着较高的退出率相关 (0.5% < 7.1% < 11.5% <15.2%)。其中，与评级为 5.0 星的房源相比，评级为 3.5 星或更低的房源的退出率要高出 14.4%。

将该结果与“幸存者偏差”和“信号传递”理论结合，我们可以大致猜测：房东通过“星级评分”像买者提供一个信号：“星级评分”较高传递了“高声誉”的信号，“星级评分”较低则传递了“低声誉”的信号。前者能吸引更多的需求，从而更易获得经济上的成功；而那些低声誉的房源，更容易退出。这种“幸存过程”导致大比例的高评级房源，从而导致了偏态分布。

6.3 Airbnb 房源退出率预测

“幸存过程”强调了房源的退出。这部分我们来进一步分析“退出率”，分析该数值与星级评分、房价、城市和房间类型之间的关系。这里，我们用 1 代表该房源在该月“退出市场”，用 0 表示该房源该月“未退出市场”。

我们首先观察各变量间的相关关系。

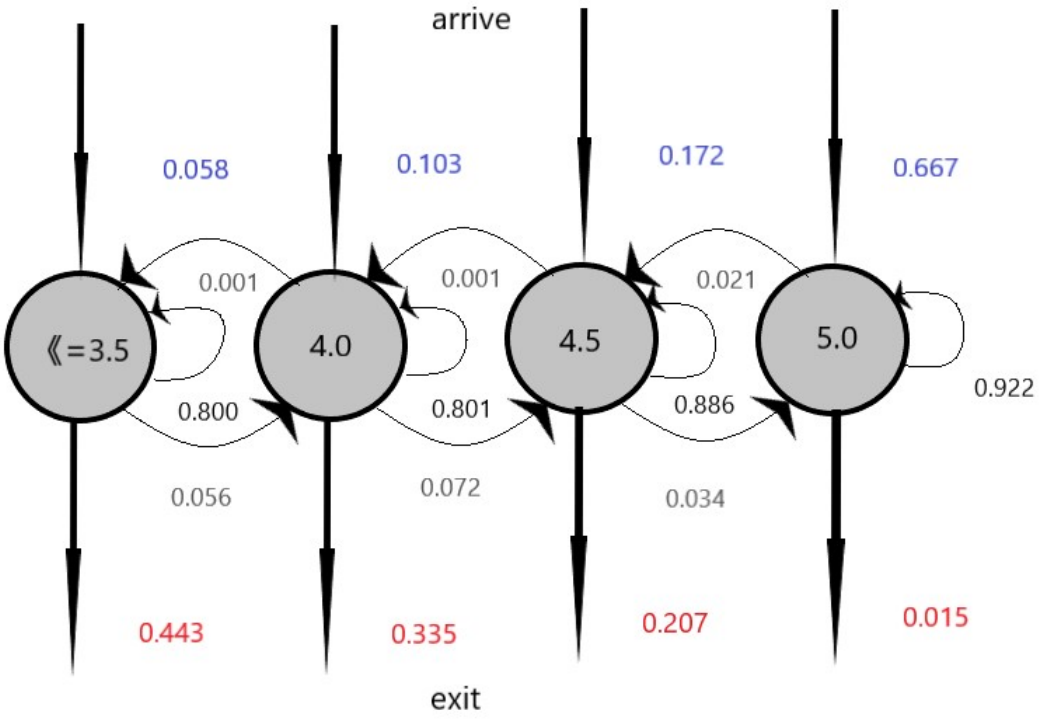

图 11: Airbnb“评分星级”状态转移网络

- **Airbnb 房源当月是否退出市场与评价和房价之间的相关性:**

  我们用 Pearson 系数刻画 Airbnb 房源当月是否退出市场与评价和房价之间的相关性，用 Seaborn 的 headmap 图可视化相关性系数，如图 12 所示。可知房价和前一个月的评价等级之间与房源在下一个是否退出市场相关性很小，基本没有相关性。

- **Airbnb 房源当月是否退出市场与城市的相关性:**

  考虑到城市变量为分类变量，这里我们构建卡方统计量来计算变量之间的相关性。卡方值越大，说明关联越强。卡方值越小，说明越不相关。结果如图 13 所示，自由度 dof=8 的情况下卡方值为 85.0216，有相关性。P 表示两变量相互独立的概率，p=0 表示一定相关。

- **Airbnb 房源当月是否退出市场与房间类型的相关性:**

  这里我们仍然构建卡方统计量来计算这两个分类变量之间的相关性。结果如图 14 所示，自由度 dof=6 的情况下卡方值为 102.0006，数值很大，有相关性。P 表示两变量相互独立的概率，p=0 表示一定相关。

在此基础上，我们用 Statsmodels 中的 OLS 线性回归模型进行多元线性回归。其中，

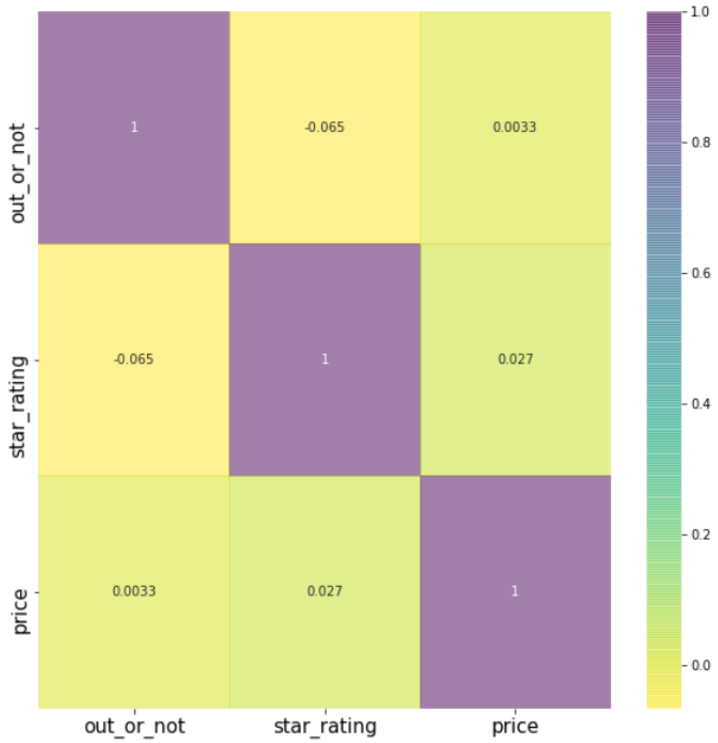

图 12: Airbnb 房源当月是否退出市场与评价和房价之间的相关性

因变量为 out_or_not（是否退出）；自变量为 price，城市虚拟变量（Los-Angeles，New-York，San-Francisco，Austin）和房源类型虚拟变量（Private_room，Shared_room，Entire_home,apt_room）等九个变量。回归结果如图 14 所示。

$P > |t|$ 可以用来评价拟合的结果和相关性大小，值越小表示相关性越高，同时拟合结果也越准确。可以看到，较高的评级分值对房源退出率有显著的负面影响 (b = -0.0386，p<0.000)，一定程度上验证了“幸存过程”的存在。此外，退出率与价格关系较小，但与房源的房间类型和所在城市有关。

6.4 Airbnb “星级评分”的平稳分布

应用马尔可夫链模型，我们还可以算出平稳分布。

我们以 2019 年 9 月 Airbnb 各星级评分作为初始分布：≤ 3.5:0.014，4.0:0.032，4.5:0.144，5.0:0.384，结合一步转移概率矩阵，最后计算可得平稳分布：≤3.5:0.012，4.0:0.031，4.5:0.147，5.0:0.384。具体迭代过程如图 16 所示。可以看到，相较于初始分布，平稳分布后的星级评分并未有过大比例的提升。因此，我们猜测这不仅仅是幸存过程的问题。

```
from scipy import stats
# chi2_contingency:卡方检验,chisq:卡方统计量值,expected_freq:期望频数
print(' chisq = %6.4f\n p-value = %6.4f\n dof = %i\n expected_freq = %s'
      %stats.chi2_contingency(cross_table))
```

```
chisq = 85.0216
p-value = 0.0000
dof = 8
expected_freq = [[ 6744.7227544  26558.1585186  29048.18904801  5033.92967899
 67385.        ]
[  517.2772456   2036.8414814   2227.81095199   386.07032101
  5168.        ]
[ 7262.         28595.         31276.          5420.
 72553.        ]]
```

图 13: Airbnb 房源当月是否退出市场与评价和城市之间的相关性

```
#计算卡方值和自由度
from scipy import stats
# chi2_contingency:卡方检验,chisq:卡方统计量值,expected_freq:期望频数
print(' chisq = %6.4f\n p-value = %6.4f\n dof = %i\n expected_freq = %s'
      %stats.chi2_contingency(cross_table_type))
```

```
chisq = 102.0006
p-value = 0.0000
dof = 6
expected_freq = [[41371.10029909 24363.47663088  1650.42307003 67385.        ]
[ 3172.89970091  1868.52336912   126.57692997  5168.        ]
[44544.         26232.          1777.         72553.        ]]
```

图 14: Airbnb 房源当月是否退出市场与评价和房间类型之间的相关性

6.5 Airbnb 评价数量与评分的动态关系

基于以上分析，我们又调查了“星级评分”与“评价数量”之间的关系。与星级评分类似，“客户评价”也是 Airbnb 信誉机制的重要组成部分。

首先，我们利用 Seaborn 工具包中的 distplot 函数作出了 2019 年 9 月 Airbnb 平台上评价数的概率分布图，如图 17 显示。可以看出，最常见的评论数量实际上是“0”，每个房源获得的评价数基本都集中在 100 以内。据统计，该样本的总体均值为 39.5，标准差为 63.0，中位数为 14.0。

接下来，我们将评价数量与“星级评分”结合来看，即不同的评价数量下，星级评分的分布。图 18 是 Seaborn 中的 Violinplot 函数绘制的“琴形图”，面积越大的地方代表数据越多。通过对评论数进行分组，我们发现从左到右，随着评价数量的增加，“星级评分”也越来越高。

Parameter:

	0	1
const	const	0.172037
x1	price	0.000003
x2	LA	0.041888
x3	AUS	0.065989
x4	SF	0.021676
x5	NY	0.042484
x6	Shared room	0.093464
x7	Private room	0.041711
x8	Entire home/apt	0.036862
x9	star_rating	-0.038648

OLS Regression Results

Dep. Variable:	out_or_not	R-squared:	0.007
Model:	OLS	Adj. R-squared:	0.007
Method:	Least Squares	F-statistic:	69.94
Date:	Mon, 23 Dec 2019	Prob (F-statistic):	3.17e-101
Time:	23:52:33	Log-Likelihood:	-4187.2
No. Observations:	72553	AIC:	8390.
Df Residuals:	72545	BIC:	8464.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.1720	0.007	24.911	0.000	0.159	0.186
x1	2.958e-06	2.22e-06	1.334	0.182	-1.39e-06	7.3e-06
x2	0.0419	0.002	18.411	0.000	0.037	0.046
x3	0.0660	0.003	20.921	0.000	0.060	0.072
x4	0.0217	0.003	6.428	0.000	0.015	0.028
x5	0.0425	0.002	19.253	0.000	0.038	0.047
x6	0.0935	0.005	18.950	0.000	0.084	0.103
x7	0.0417	0.003	13.417	0.000	0.036	0.048
x8	0.0369	0.003	11.939	0.000	0.031	0.043
x9	-0.0386	0.002	-17.145	0.000	-0.043	-0.034

Omnibus:	47612.701	Durbin-Watson:	1.902
Prob(Omnibus):	0.000	Jarque-Bera (JB):	377591.245
Skew:	3.303	Prob(JB):	0.00
Kurtosis:	12.015	Cond. No.	8.59e+18

图 15: Airbnb 房源“退出率”的 OLS 线性回归结果

进一步来看，我们观察具体评论数下“星级评分”的情况。我们利用 Matplotlib 中的 scatter 函数作出散点图描述，如图 19 所示。这里值得注意的一点是，大部分房源的评论数其实都集中在 50 以内，因此得到低评论数的房源数量与获得高评论数的房源数量之间差距极大，如果让房源数与点的面积成正比来进行作图的话，园点的面积差距大，重叠度很高，无法完整呈现规律，所以需要缩小房源数量的差异。在图中，点的大小实际是与该处房源数的平方根成正比，这样就使得点的大小差异较为合适。

综上所述，当评论数量增加时，“星级评分”的分布会越来越窄，评论数量不断增加的房源获得最高评级的可能性会降低。

这个现象或许可以用统计中的大数定律和均值回归来解释。大数定律表明，随着评论数量的不断增加，房源获得最高评级的可能性会降低；而均值回归定理可以说明：较高的信誉评级可能会引起客户对一个特别好的体验的期望，这可能会有效地增加失望的机会，反过

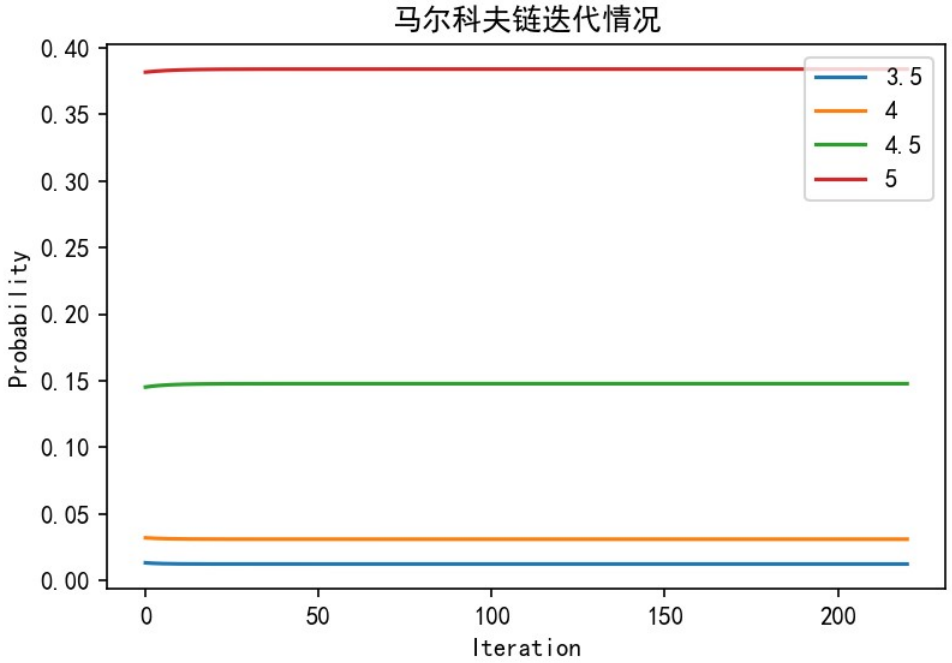

图 16: Airbnb 星级评分的迭代过程及其“平稳分布”

来，促进较低的客户评级。同样，如果预订低评级的房源可能会带来意想不到的积极体验。

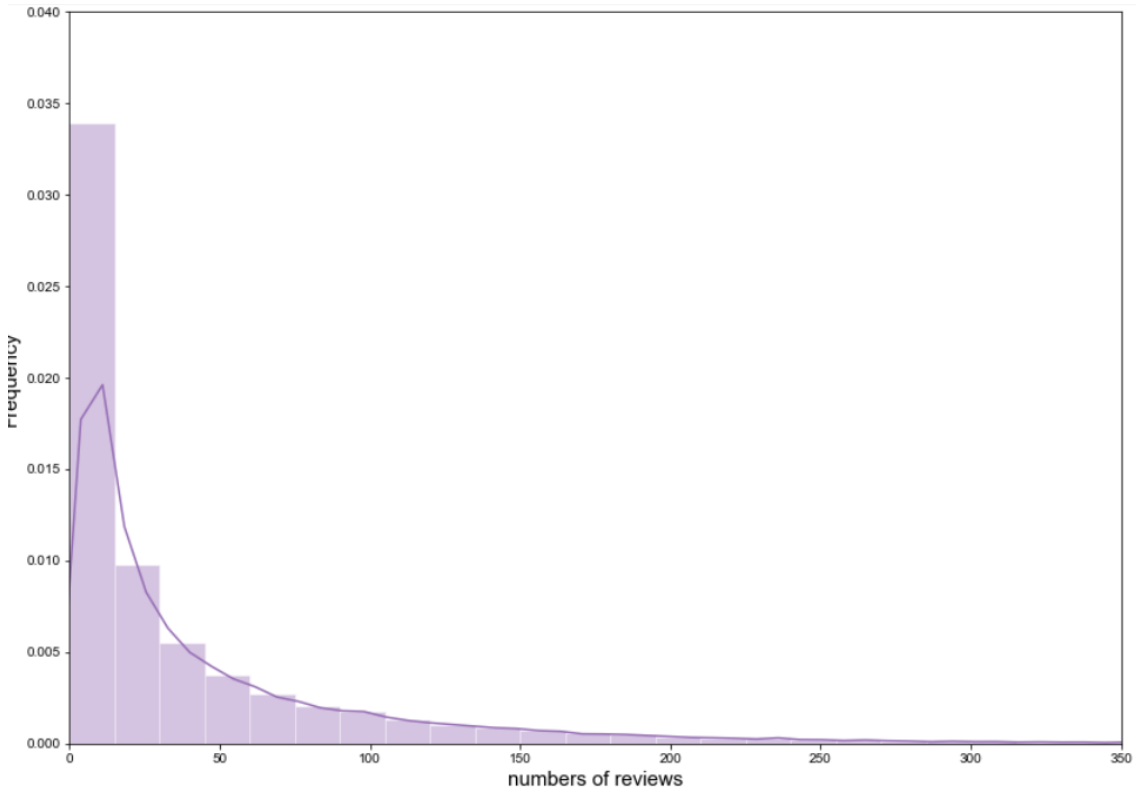

图 17: 2019 年 9 月 Airbnb 平台上房源的评价数分布图

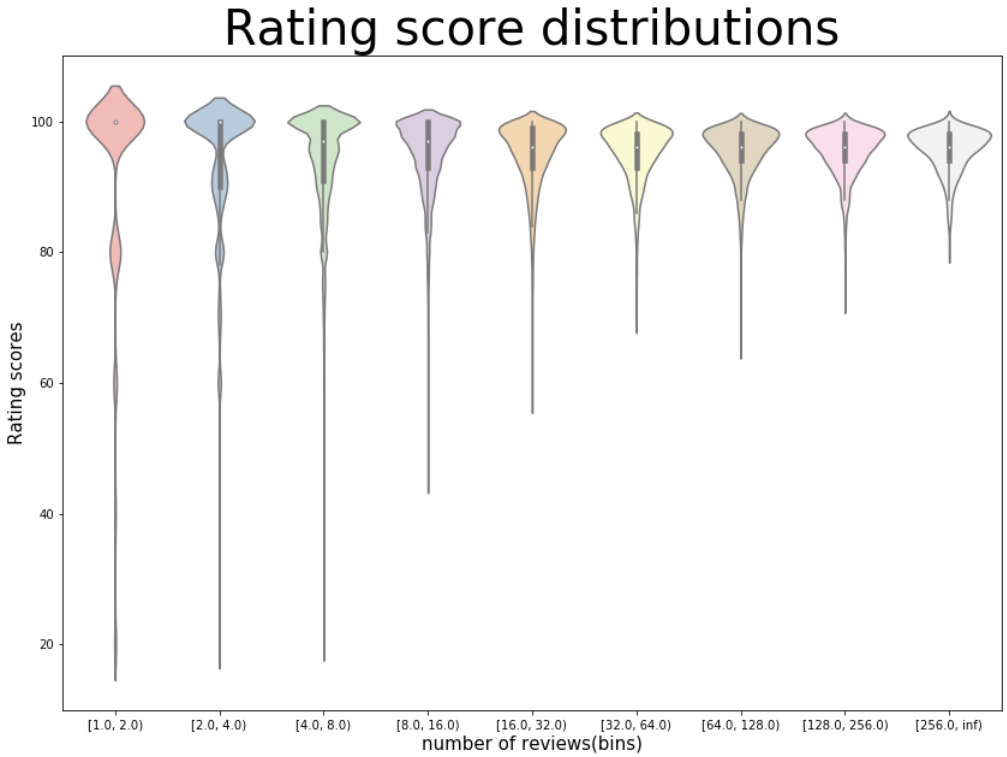

图 18: 不同评价数组中获得的评价等级的分布图（琴形图）

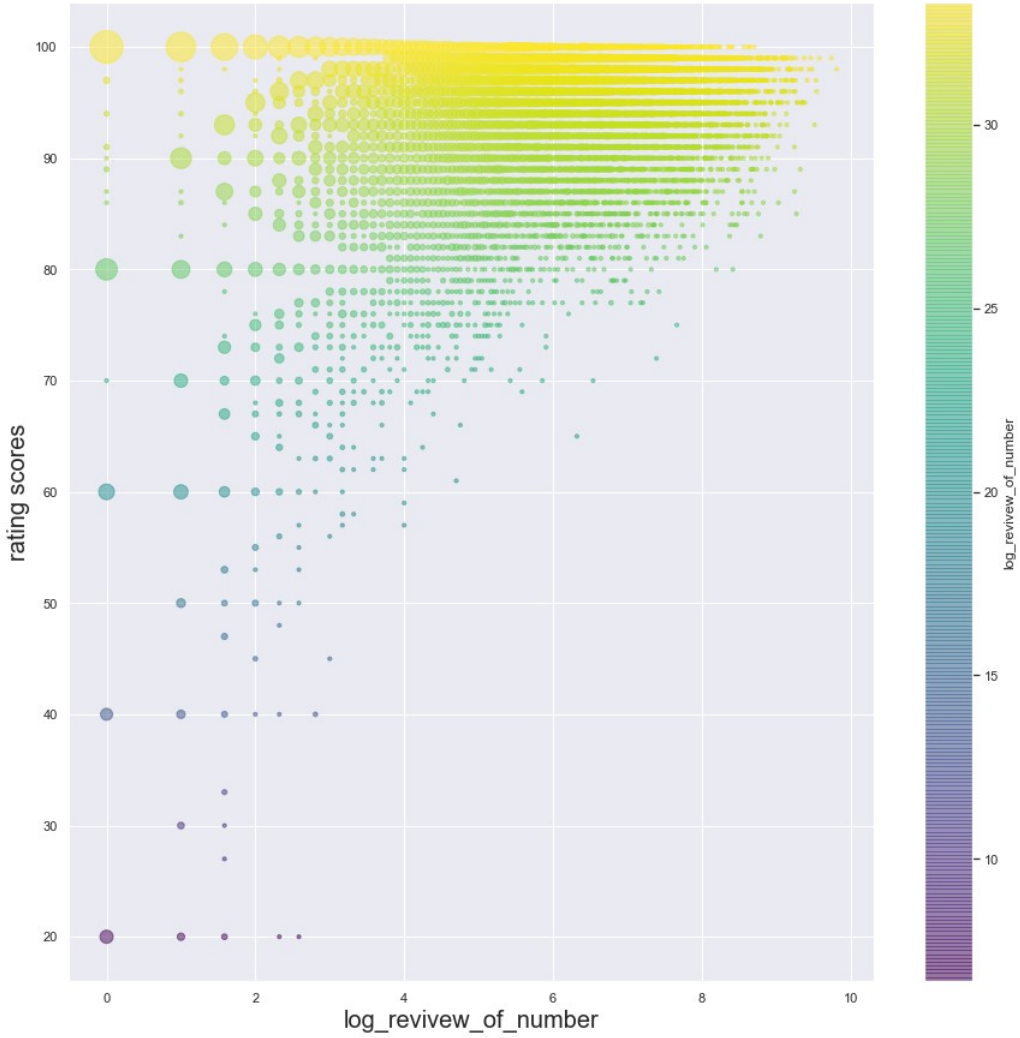

图 19: 不同评价数组中获得的评价等级的分布图（散点图）

参考文献

[1] Zervas G., Proserpio D., and Byers J. “A first look at online reputation on Airbnb, where every stay is above average”. In: *Working Paper.* (2015).

[2] Ikkala T. and Lampinen A. “Monetizing network hospitality: Hospitality and sociability in the context of Airbnb”. In: *CSCW’ 15 Proceedings* (2015), pp. 1033–1044.

[3] Ert E., Fleischer A., and Magen N. “Trust and reputation in the sharing economy: The role of personal photos in Airbnb”. In: *Tourism Management* 55.1 (2016), pp. 62–73.

[4] Bridges J. and Vásquez C. “If nearly all Airbnb reviews are positive, does that make them meaningless?” In: *Current Issues in Tourism* (2016), pp. 1–19.

[5] Fradkin A., Grewal E., and Holtz D. “The determinants of online review informativeness: Evidence from field experiments on Airbnb”. In: *Working Paper* (2017).

[6] Teubner T. “The web of host-guest connections on Airbnb—A social network perspective”. In: *Working Paper* (2017).

[7] Teubner T., Hawlitschek F., and Dann D. “Price determinants on Airbnb: How reputation pays off in the sharing economy”. In: *Journal of Self-Governance and Management Economics* 5.4 (2017), pp. 53–80.