



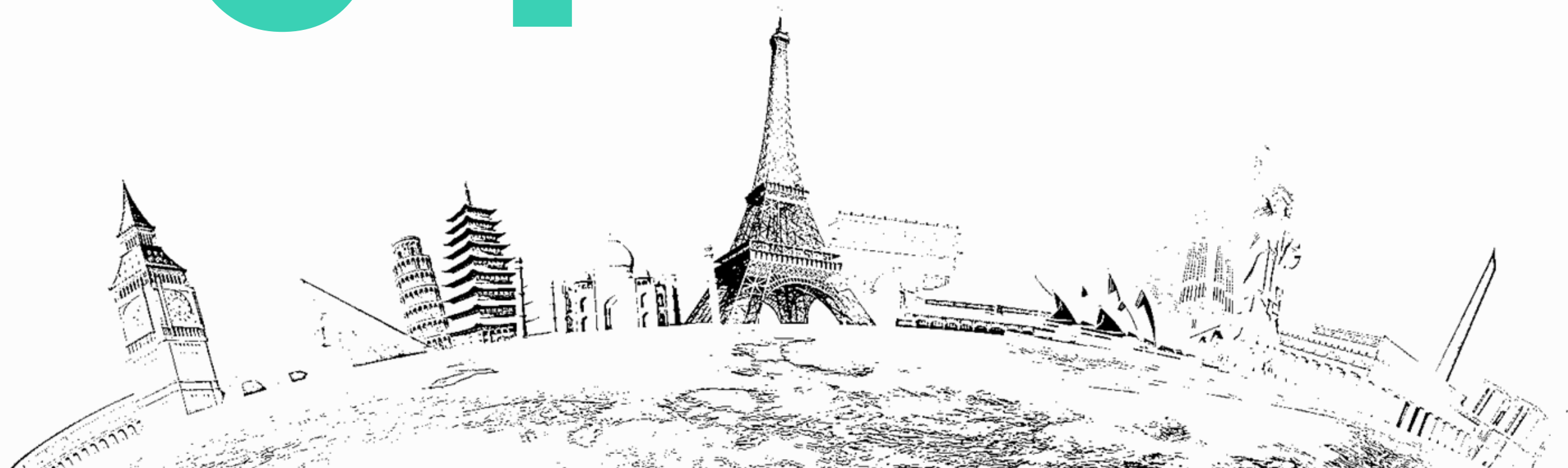
世界这么大 我想去看看

北美在线旅游产品选择攻略



01 章节 PART

背景介绍



故事是这样开始的.....



爸妈来参加毕业典礼，想带他们到处溜达溜达

今年去哪儿浪起来啊？



主要是在西海岸吧，先玩儿西雅图附近的奥林匹克和瑞尼尔两个国家公园，再去黄石公园看看老忠实和大棱镜，最后要还能去趟旧金山和拉斯维加斯就完美啦~

听起来不错，都去哪儿呢？



打算跟团，不能让爸妈跟着我风餐露宿的跑来跑去

这么多地儿，跟团还是自驾啊？



还没呢，网上产品太多了，景点安排的线路不同，时间不同，还有住宿交通的条件也不一样

定团了？也让我参考参考呗？



看了快一个星期，不过都不确定看到的产品是不是同类中价格合适的，所以一直下不了决心

恩，选个性价比还不错的就行



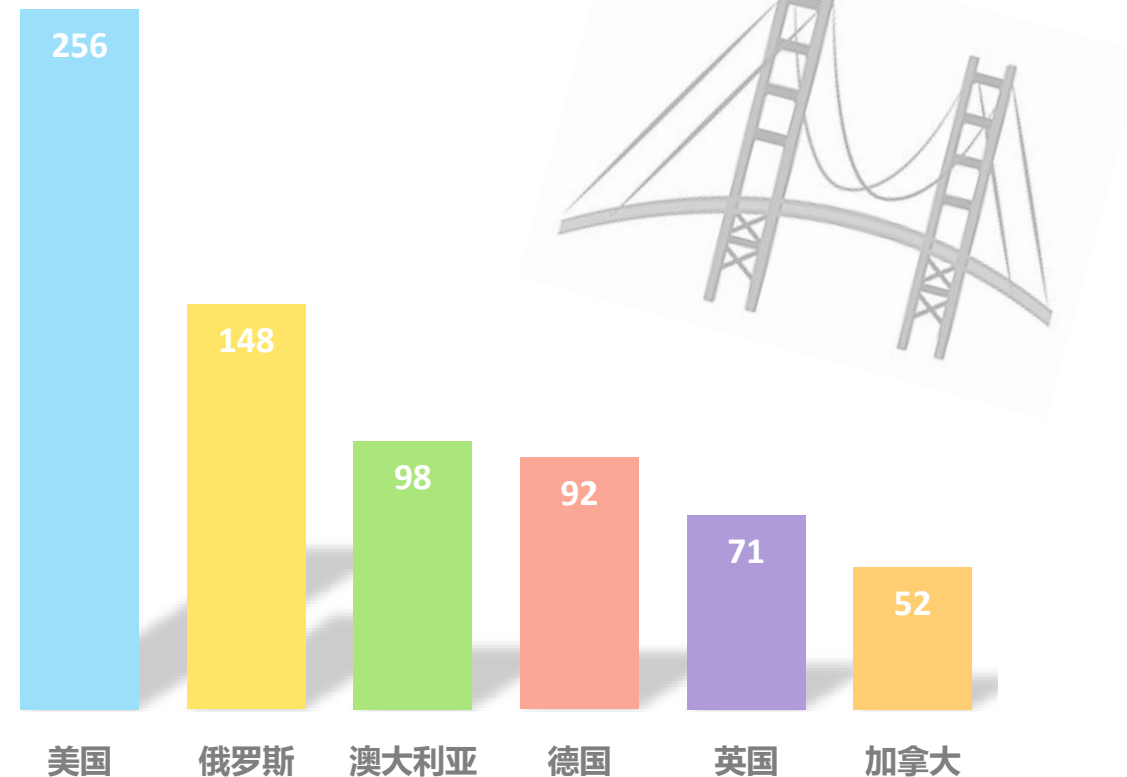
是啊，要是有人能帮我看看我挑的产品的性价比就好了

这个是得谨慎一点，不然随便定了个价格高、但其他条件又差不多产品的，就亏大发了





- 美国在欧美出境游中**人数**最多
- 美国路线广、景点多、**资源**丰富
- 留学人群增加带来了的潜在旅游**市场**



数据来源：中国产业信息网
词云图：根据熊大旅游网产品中的地名频率绘制

如何判断旅游产品是否属于高性价比产品？



行程安排
游玩天数、景点个数和出发地等因素

01

服务内容
跟团方式、供应商和酒店质量等因素

02

销售评价

是否已有销售记录以及顾客反馈

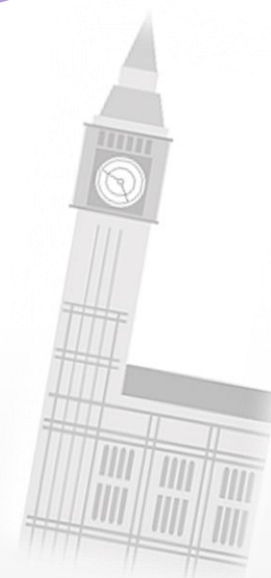
03

回归预测

预测不同情境下的产品期望价格。同等条件下比该价格低的产品认为性价比比较高

04


性价比



02

章节 PART

数据说明



数据说明



来源：熊大旅游网

产品：北美目的地跟团游

观测：2619条（去缺失）

时间：8月20号到8月26号一周的数据

变量类型		变量名称	详细说明	取值范围	备注
因变量		产品价格	单位：美元/人次	337~53,073	一周的平均报价
自变量	行程安排	游玩路线	定性变量 (5水平)	东海岸、西海岸、中南部、 美国+加拿大、夏威夷	
		出团日期	定性变量 (3水平)	仅工作日有团、仅周末有团、 工作日和周末都有团	
		行程天数	单位：天数	2~21	总出行时间
		景点总个数	单位：个	1~77	熊大网给出了每条产品的 景点总个数
		经典景点数	单位：个	1~35	熊大网给出了每条产品 经典景点的个数
	服务内容	参团方式	定性变量 (3水平)	跟团游、半自助、私家团	半自助游：部分时间自由活动 私家团：一单即为一团
		供应商	定性变量 (4水平)	熊大旗下、境外华人旅社、 国内旅社、其它	共有35家供应商， 按属性分为4大类
		产品等级	定性变量 (5水平)	2钻、3钻、4钻、 5钻、无	从住宿、交通弄、自费项目等 多个方面为产品定级
		交通信息	定性变量 (2水平)	1：行车时间长 0：行车时间短	如果行车时间过长，网站 会给出行车时间，
		用餐信息	定性变量 (2水平)	1：全部自理 0：部分自理	
	销售评价	评价信息	定型变量 (2水平)	有评价，没有评价	是否有评价信息
		销量信息	定性变量 (2水平)	有销量、没有销量	是否有销量数据

03

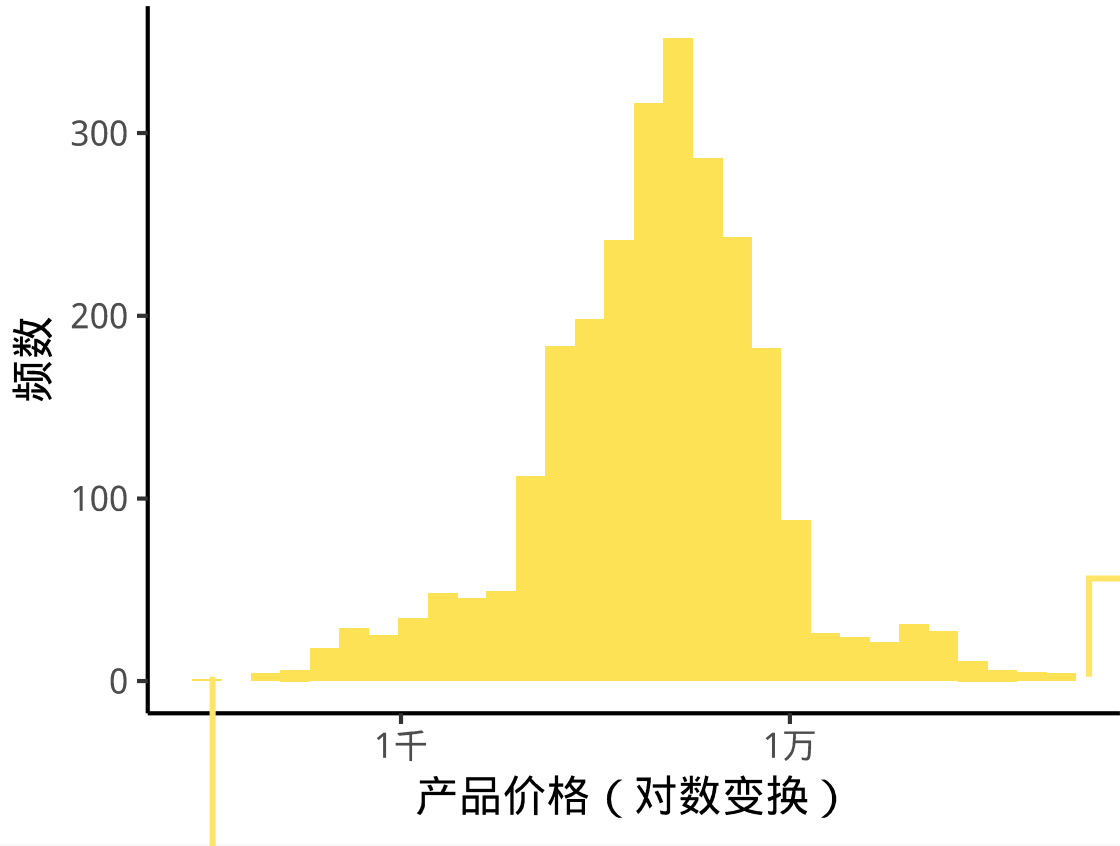
章节 PART

描述分析



产品价格直方图

均值	最小值	Q _{.25}	中位数	Q _{.75}	最大值
5730	337	3080	4720	6780	53073

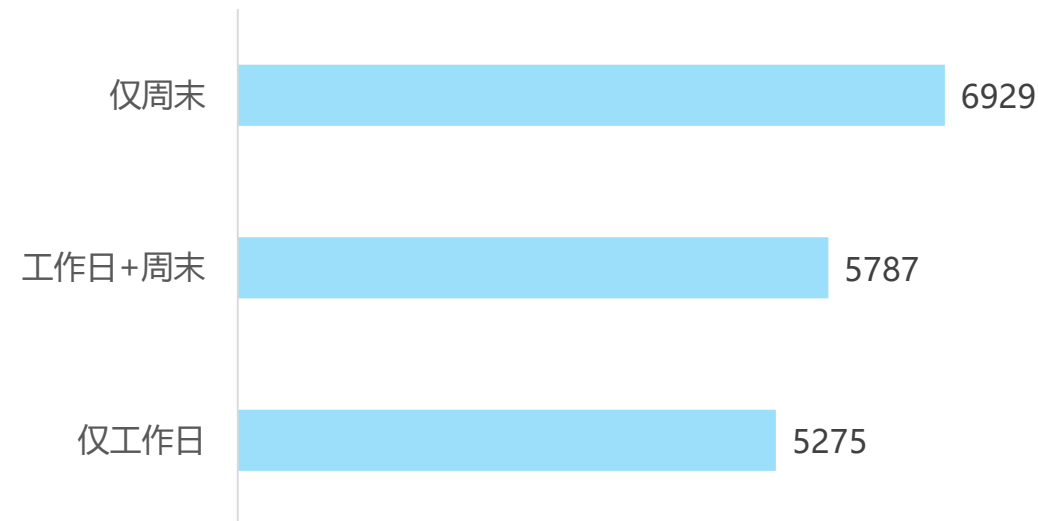
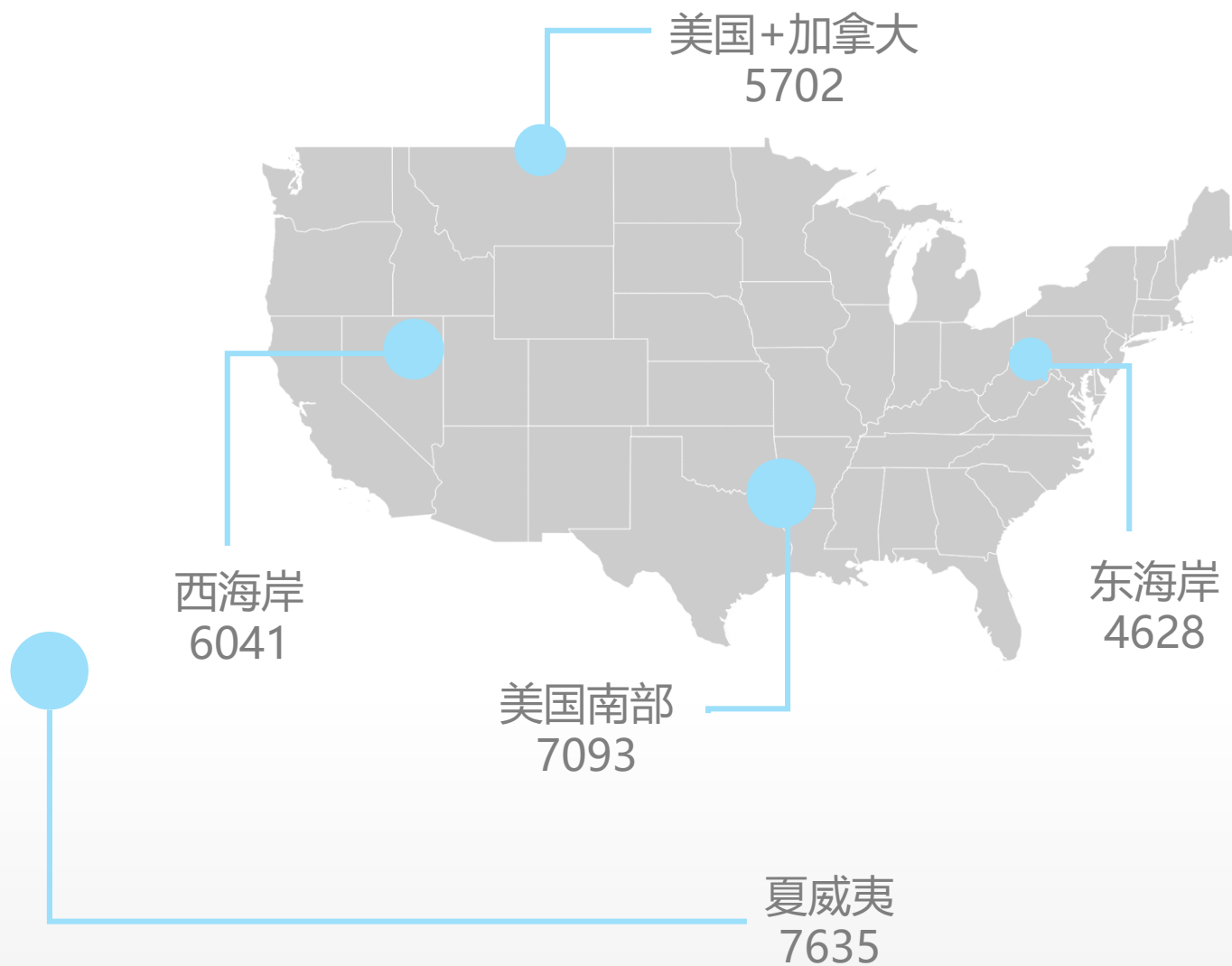


- 产品价格呈现明显的右偏分布
- 最大值超过5万抬高了整体均值
- 多数产品价格集中在3千到1万之间

产品价格

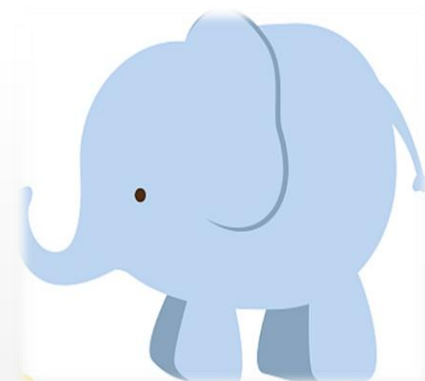


不同路线的产品价格均值



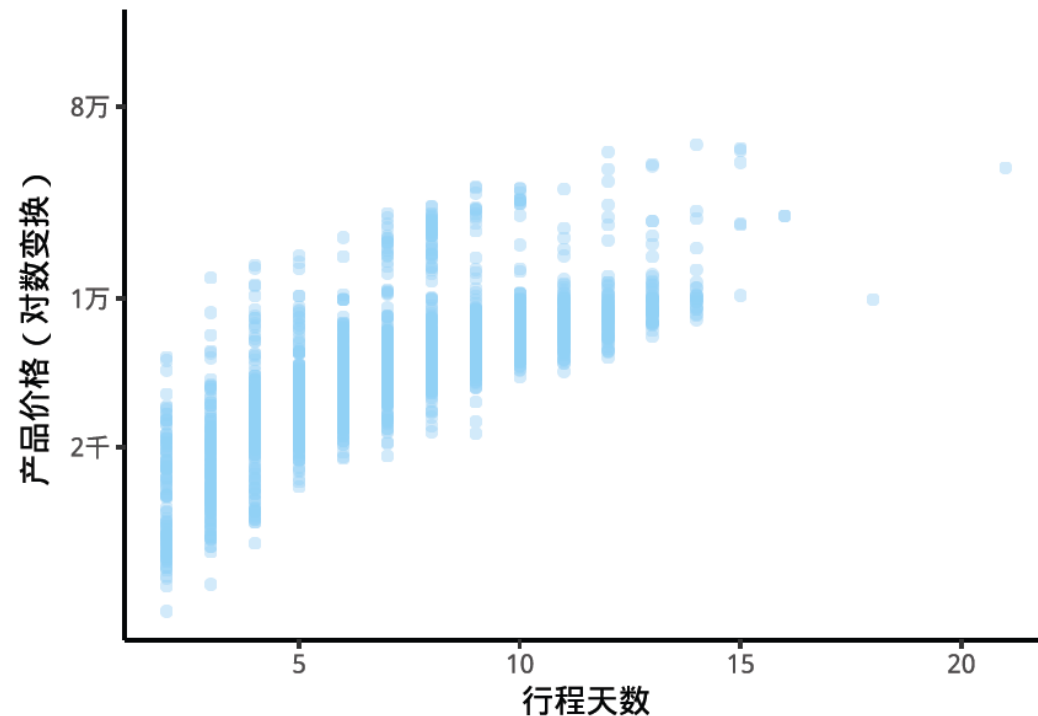
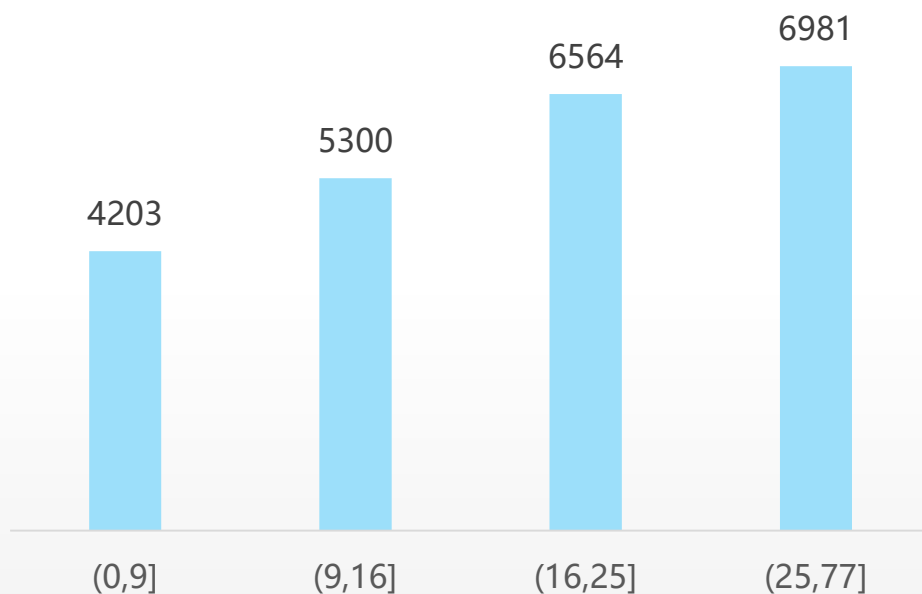
- 多数产品在工作日和周末都发团
- 仅周末出团的产品价格明显比有工作日出团的产品价格高

行程安排



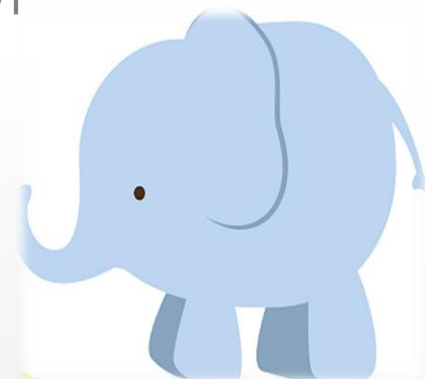
- 线路包含的所有景点个数越多，价格越高
- 经典景点个数的趋势和所有景点相似，故此处略图

所有景点个数按组分的平均价格

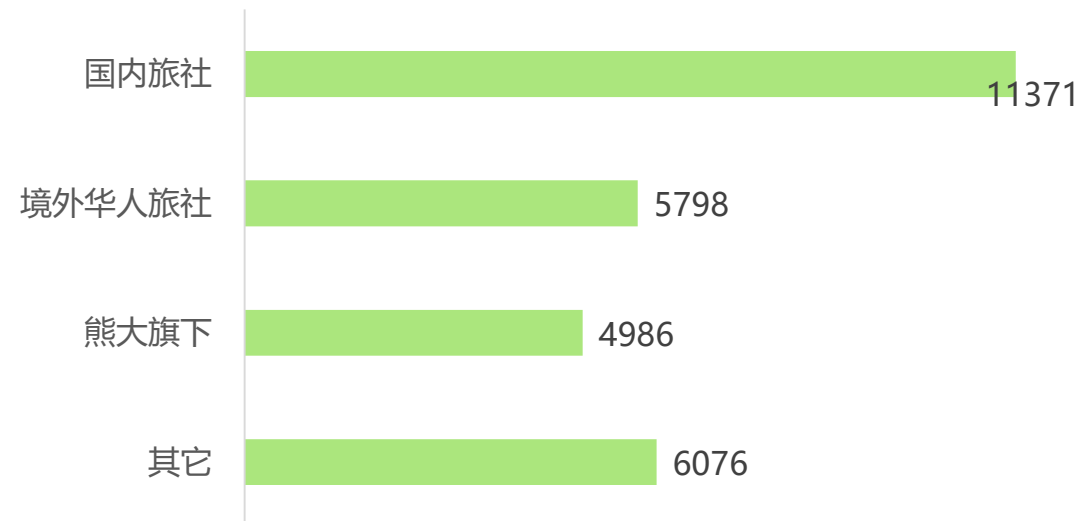
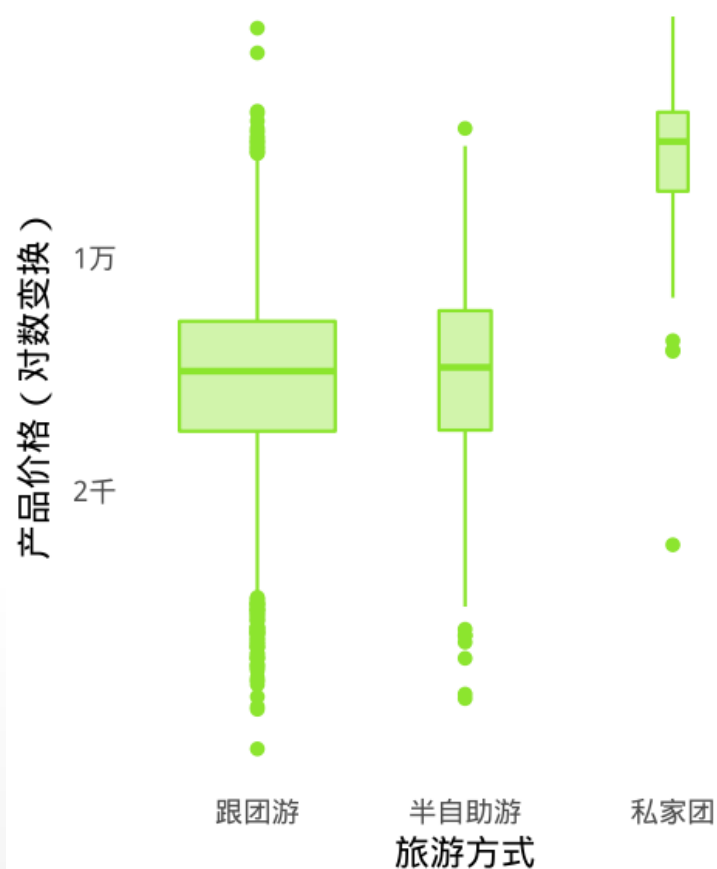


- 行程天数越长，产品价格越高
- 但当时间长到一定程度后，价格上升越慢

行程安排



- **旅游方式**以跟团游产品数量最多，价格相对低
- 私家团产品最少，价格远高于跟团和半自助



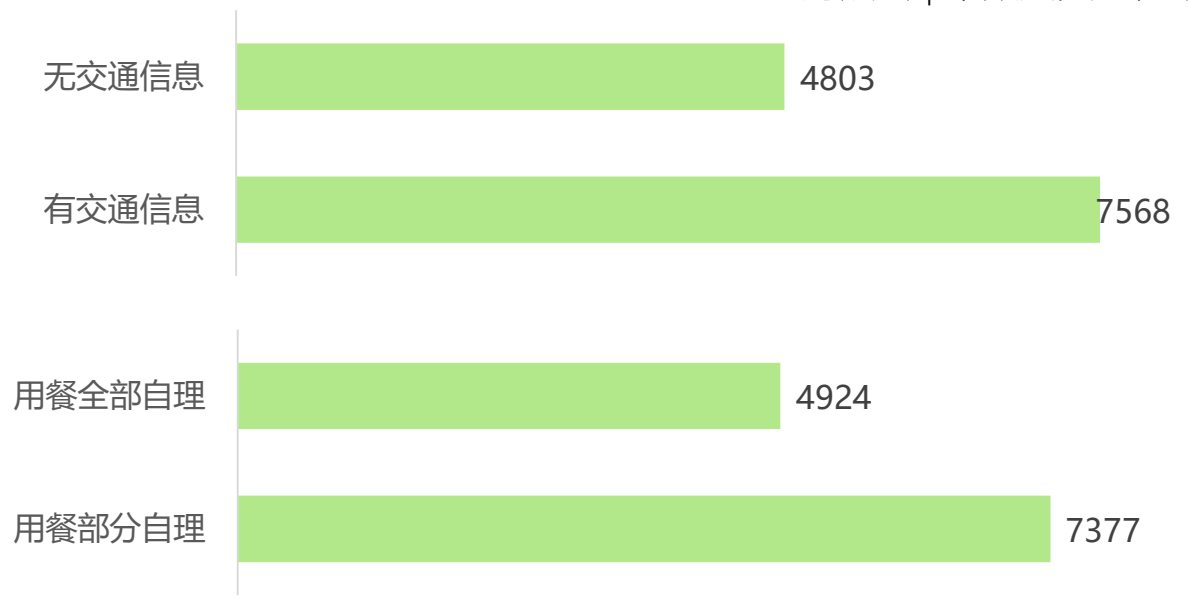
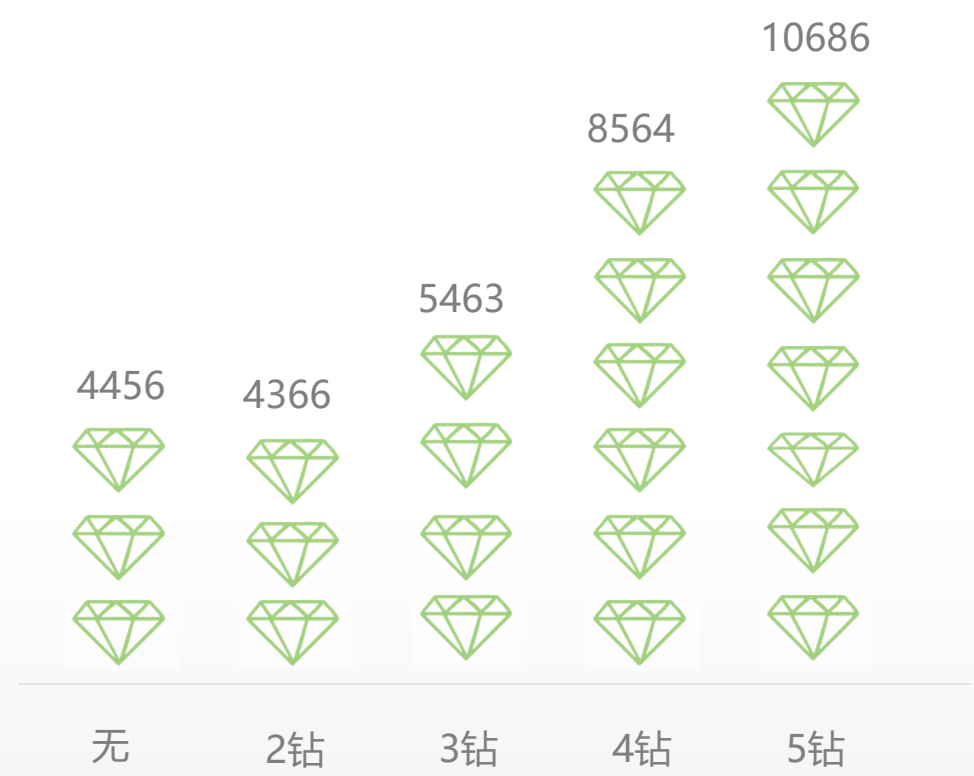
- 这里按照供应商的类别分为国内旅社、境外华人旅社、熊大旗下旅社和包括零售商、部分欧洲旅社等在类的其它旅社
- 不同**供应商**的报价差异较大。熊大旗下的公司报价相对低，零售商和国内旅社报价高

服务内容



描述分析

- 产品等级越高，产品平均价格越高
- 3钻产品数量最多，5钻产品数量少

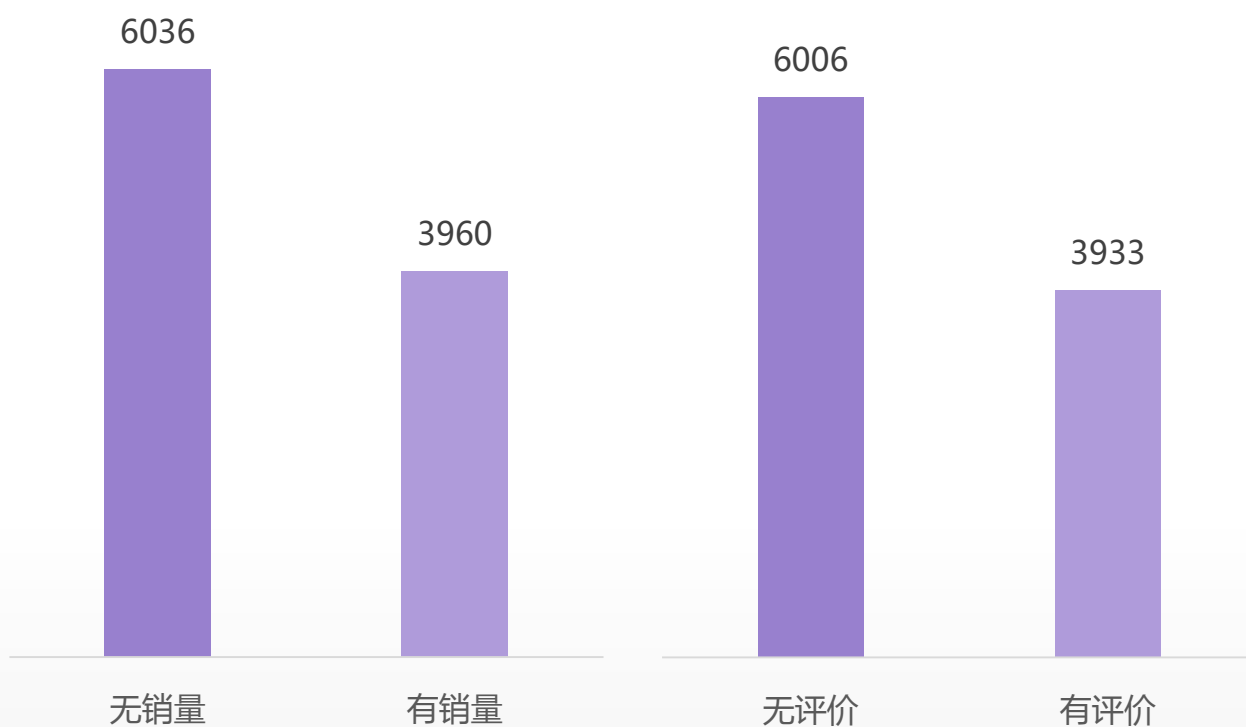


- 提供交通信息的产品费用明显较高
- 如果用餐全部自理，产品费用明显较低

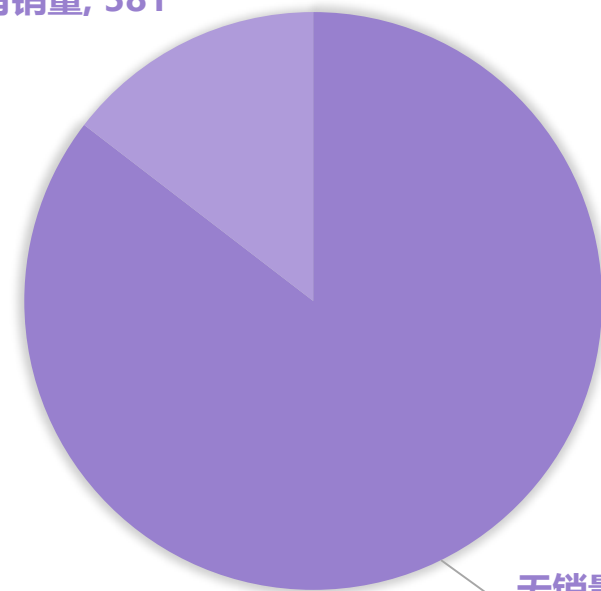
服务内容



- 有销售数据的产品价格明显低于无销售量的产品
- 有评价信息的产品价格明显低于无评价的产品



有销量, 381



无销量, 2234

- 无销量产品 and 无评价产品 占有所有北美旅游产品的80%以上。
- 有无销售和有无评价之间强相关

销售评价



04

章节 PART

回归建模



探究连续型变量和多个解释变量的关系，常用线性回归模型

在本案例中，

- 因变量：是每条旅游产品的周平均价格；由于因变量右偏明显，这里采取对因变量取对数的策略
- 自变量：包括行程安排、服务内容、销售评价三个方面
- 研究目的：12个自变量是如何影响旅游产品的价格，并根据建立的模型对不同情境下的产品进行预测

模型选择

最终模型：对数线性模型



由于变量个数较多，这里采用逐步回归的方法、以AIC准则为依据选择变量

我们对最终筛选出的变量按其所属的类别进行解读

变量名称		系数估计(10^{-1})	p值	基准组
行程安排	游玩线路	南部	1.84	<0.001***
		西海岸	1.57	<0.001***
		夏威夷	2.95	<0.001***
		美国+加拿大	0.25	0.41
	出团日期	仅工作日	-1.55	<0.001***
		工作日和周末	-1.77	<0.001***
	行程天数	1.83	<0.001***	
	经典景点数量	-0.10	<0.001***	

变量解读

当其他因素不变的情况下

- 南部、西海岸、夏威夷分别比东海岸价格平均高出18.4%、15.7%、29.5%
- 仅工作日出团和工作日周末都出团分别比仅周末出团的产品价格平均低了15.5%和17.7%
- 行程每增加1天，产品价格平均高出18.3%
- 经典景点每增加一个，产品价格平均低1%



变量解读

当其他因素不变的情况下

变量名称			系数估计(10 ⁻¹)	p值	基准组
服务内容	参团方式	跟团游	-11.65	<0.001***	私家游
		半自助游	-11.33	<0.001***	
	供应商	熊大旗下	-0.60	0.111	其它
		国内旅社	2.88	<0.001***	
		境外华人旅社	-0.40	0.259	
	产品等级	2钻	-0.69	0.059 .	无等级
		3钻	-0.91	<0.001***	
		4钻	0.21	0.395	
		5钻	2.34	<0.0001***	
	交通信息	提供行车时长	-0.34	0.072 .	没有行车时长
销售评价	销售信息	用餐部分自理	2.12	<0.001***	用餐全部自理
		有销量	-1.10	<0.001***	无销量

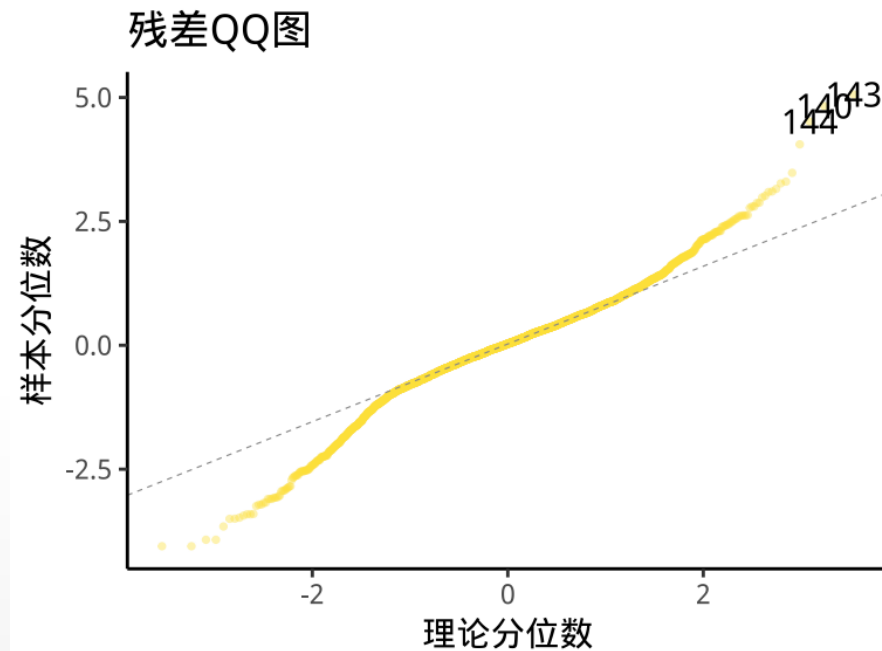
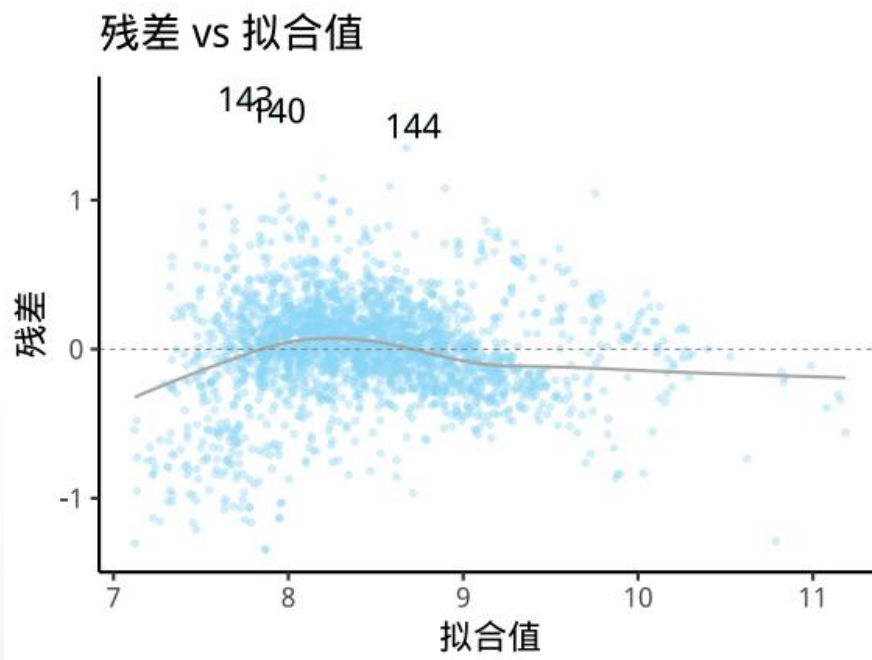
- 跟团游和半自助游比私家游产品价格平均低116.5%、113.%
- 国内旅社显著比其它旅社高28.8%；熊大网和境外华人旅社平均低6%和4%， 但不显著
- 2钻和3钻产品比无等级产品价格平均低6.9%、9.1%；5钻产品则平均高23.4%
- 提供行车时长的产品比不提供产品价格平均低3.4%
- 用餐部分自理比全部自理的产品价格平均高出21.2%
- 已经有销售信息的比无销售的产品价格平均低11%

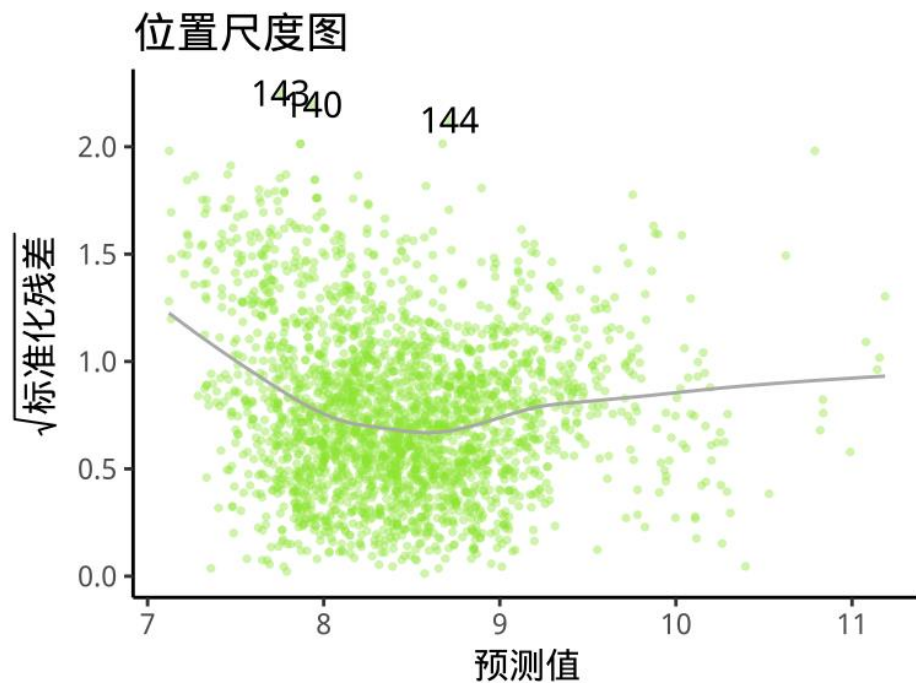
- 模型的拟合优度为75.9%，模型拟合效果良好
- 模型整体的F检验高度显著（ p 值 <0.001 ），说明产品价格确实与所考虑的3类因素有关



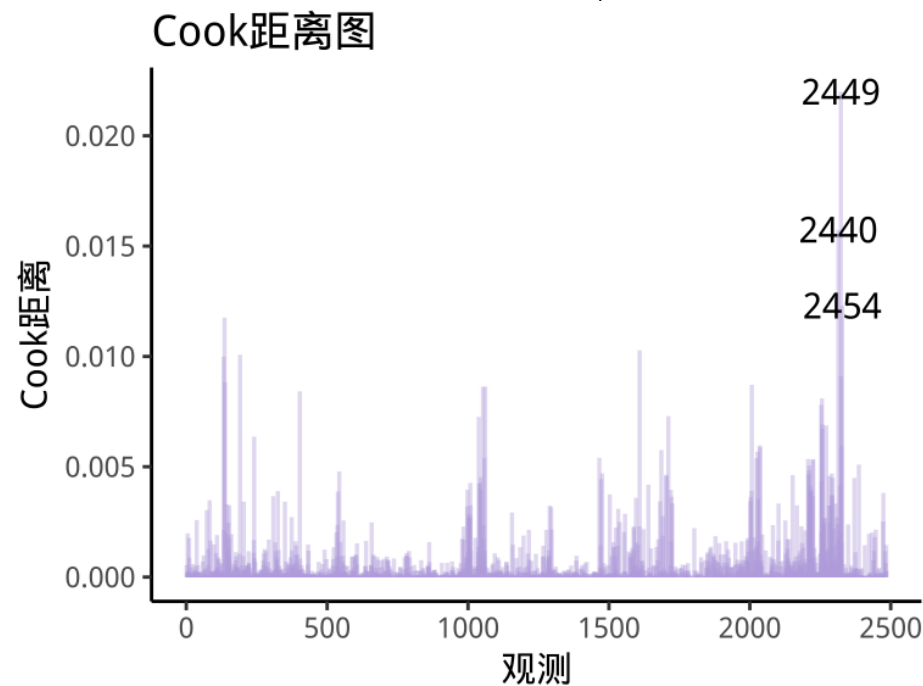
模型检验

- 残差和拟合值没有明显的系统关系，说明模型中的自变量能够解释大部分的价格波动。
- 残差的分布尾部概率较小，但没有明显偏离正态性





- 残差基本上沿着预测值均匀分布，没有明显的异方差性



- 有3个样本像强影响点，但是并没有出现在其他检验图中，且剔除后对模型影响不大，可保留

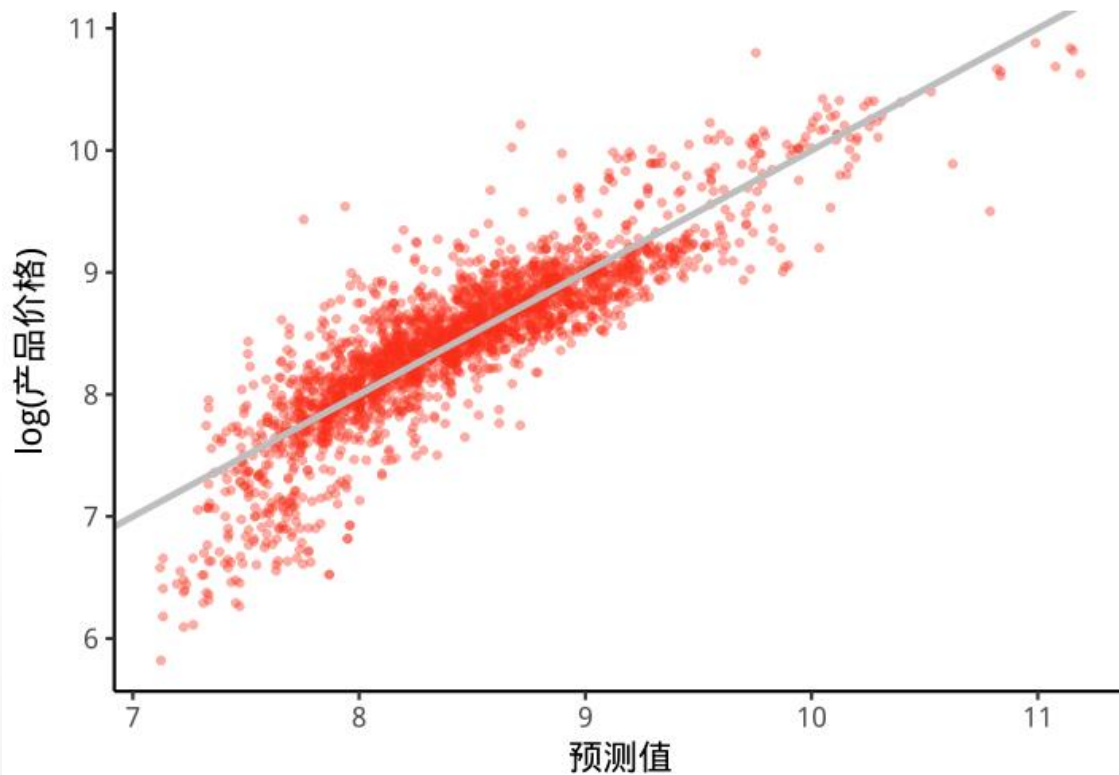
模型检验

- 模型中几乎不存在多重共线性，所有变量的方差膨胀因子VIF值在1和4之间。

采用5折交叉验证的方式检验模型的预测效果

- 利用该模型得到的预测值和真实值随机分布在 $Y=X$ 两侧，说明模型的预测能力较好
- 预测的均方误差为0.11，相对 $\log(\text{产品价格})$ 来说较小，说明模型预测效果良好

5折交叉验证检验模型的预测效果



注：这里采用的
是一次随机分配
样本的结果



预测效果

05

章节 PART

应用场景



该怎么样使用这个模型来找到既适合自己、性价比又高的产品呢？

情境模拟

- 和同学约好了来美国东部耍一圈，没啥要求，能省则省

东海岸 / 随时出发 / 玩10~13天 / 5个以上经典景点 / 跟团游 / 旅社随意 / 等级随意 / 车程得短 / 用餐可自理 / 可暂无销售



穷屌国内大学生

- 打算和女盆友一起去夏威夷度假，钱都不是问题

夏威夷 / 周日出发 / 玩5~7天 / 1个以上经典景点 / 私家团 / 旅社随意 / 4钻或5钻 / 车程无所谓 / 用餐全部自理 / 可暂无销售



刚脱单的高级白领

- 带爸妈去西部多看看，该花得花，但不铺张浪费

西海岸 / 随时出发 / 玩10~15天 / 10个以上经典景点 / 跟团游 / 熊大网旗下 / 3钻或4钻 / 车程无所谓 / 用餐可自理 / 有销售量



快秃顶的留美程序员

根据每位游客不同需求给出期望产品价格和选择产品的建议

产品选择

1: 由于部分变量不固定，所以模型将给出一个预测的价格区间。



■ 要省软妹币

模型估计最低价¹: ¥ 5225 - ¥ 17476

熊大产品: 20个 (¥ 5186~¥ 10310)

最低价产品: 美国华盛顿+尼亚加拉瀑布+大烟山国家公园
10日9晚跟团游(3钻)

性价比: 价格低于模型估计最低期望价格, 是**高性价比**产品



■ 得让女朋友玩好

模型估计最高价: ¥ 9720 - ¥ 24343

熊大产品: 3个 (¥ 7610~¥ 9899)

最高价产品: 美国夏威夷欧胡岛6日5晚私家团·专属司兼导+
私人定制+2人成行+舒适酒店

性价比: 价格低于模型估计最高期望价格, 是**高性价比**产品



■ 带爸妈多看看美国

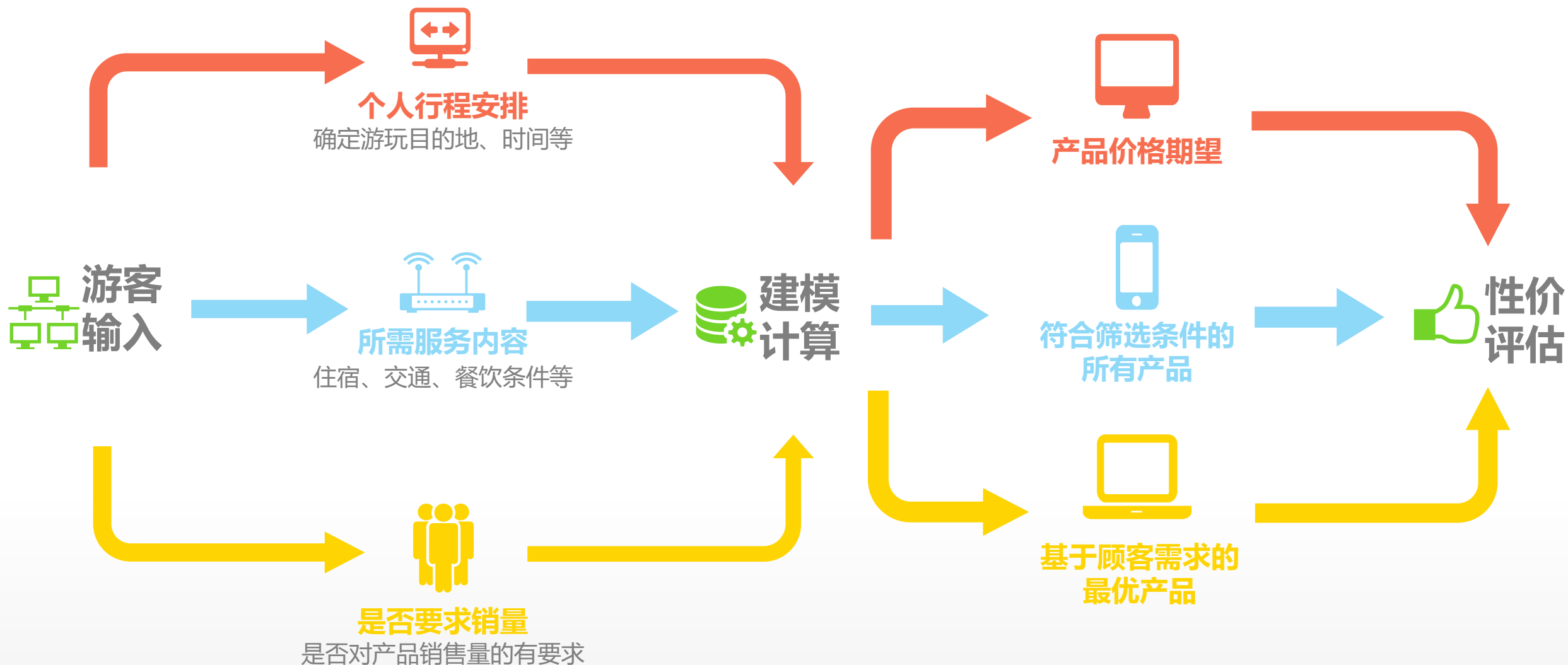
模型估计价格区间: ¥ 6701~¥ 22026

熊大产品: 8个 (¥ 5186~¥ 11200)

景点最多产品: 美国黄石国家公园+盐湖城+旧金山+拉斯维加斯11日
10晚跟团游(3钻)·全网特惠价5186元

性价比: 价格低于模型估计最低期望价格, 是**高性价比**产品

线上性价评估产品设计





谢谢观赏！



扫描二维码，关注狗熊会
获取更多案例资源

