

第九届“泰迪杯”数据挖掘挑战赛——

C 题：游客目的地印象分析

一、问题背景

提升景区及酒店等旅游目的地美誉度是各地文旅主管部门和旅游相关企业非常重视和关注的工作，涉及到如何稳定客源、取得竞争优势、吸引游客到访消费等重要事项。游客满意度与目的地美誉度紧密相关，游客满意度越高，目的地美誉度就越大。因此掌握目的地游客满意度的影响因素，切实提高游客满意度、最终提升目的地美誉度，不仅能够保证客源稳定，而且对于旅游企业科学监管、资源优化配置以及市场持续开拓具有长远而积极的作用。

二、解决问题

1、景区及酒店印象分析

依据附件 1 中景区及酒店网评文本，按表 1 格式计算出目的地 TOP20 热门词，并保存为文件“印象词云表.xls”

表 1 印象词云表

评论热词	热度

2、景区及酒店的综合评价

根据附件 1 景区及酒店网评文本及附件 2 景区及酒店得分建立合理的数学模型及相应算法，按满分为 5 分对景区及酒店的服务、位置、设施、卫生、性价比五个方面进行评分，并按照均方误差（Mean Squared Error, MSE）进行模型评价。

3、网评文本的有效性分析

出于各种原因，网络评论常常出现内容不相关、简单复制修改和无有效内容等现象，妨碍了游客从网络评论中获得有价值的信息，也为各网络平台的运营工作带来了挑战。请从文本分析的角度，建立合理的模型，对附件 1 景区及酒店网络评论的有效性进行分析。

4、景区及酒店的特色分析

旅游业繁荣发展给游客带来了选择困难的问题，评分接近的景区或酒店很难根据评分进行取舍。请建立合理的模型和算法，从景区及酒店的网评文本中挖掘他们各自的特色和亮点，以吸引游客提升竞争优势。请选择综合评价高、中、低三个层次的各 3 家景点和 3 家酒店，结合模型的结果，分析他们各自的特色。

三、附件数据说明

附件 1 数据来源于互联网公开渠道，具体表结构字段如下：

表 2 景区评论

景区名称	评论日期	评论详情
------	------	------

A1	2019-04-13	去之前最好提前在网订好门票 rmb39 元（景区 rmb45 元），有自动取票机非常方便快捷，如果有带孩子去建议购买套票（1.2m 以下免费），套票包括大门票及 7 项娱乐设施，有旋转木马、碰碰车、海盗船、丛林飞鼠、自控飞机等，玩起来非常的刺激，惊险。
A2	2020-07-12	同程门票优惠是最大的！赞！大家导航导景区售票口。要不会被带到高速上钻山洞。
A2	2020-08-16	还行，挺方便的适合情侣或者一家人过去爬爬山，我算毕竟幸运了，拍到了正面照片！
.....

表 3 酒店评论

酒店名称	评论时间	评论详情	入住类型
H01	2019-12-16	经常来东莞出差，这家酒店接待很热情，服务态度好得不得了，赞一个，往前走不远就有小吃一条街，吃饭什么的都还挺方便，床是真的舒服，说好工作躺床上休息一下一睁眼就天亮了。房间没有异味，很干净整洁的布置，也有矿泉水等免费赠饮，不错，性价比很高，退房的前台小姐姐很不错哦	标准客房
H01	2019-02-23	位置比较偏，但就不会很吵啦酒店很人性化啦，还配了免费的充电器，以备我的不时之需(这点真的很棒！)卫生也比较干净没有什么异味	豪华城景房
H01	2020-09-03	停车方便，前台小姐姐服务很好，声音也很好听喔	豪华江景房
.....

附件 2 提供了景区和酒店的考查维度及专家打分，具体表结构字段如下：

表 4 景区评分

景区名称	总得分	服务得分	位置得分	设施得分	卫生得分	性价比得分

表 5 酒店评分

酒店名称	总得分	服务得分	位置得分	设施得分	卫生得分	性价比得分