

基于多源数据融合的城市景点资源分析： 以南京为例

课程名称	大数据系统原理与应用
开课院系	信息管理学院
学生学号	231820102 魏来
及姓名	231820299 周千雯
授课教师	石进
提交日期	2025.10.31

（本项目已上传至 <https://github.com/guangchacha/multi-source-data-fusion-nanjing-scenic>）

摘要—现代城市是复杂的动态系统，城市旅游景点作为典型公共空间，是城市运行动态的微观缩影。为探究城市动态系统中天气、景区类型与游客情感的关联机制，本研究以南京为案例，整合微博签到文本、景区 POI 及气象数据等多源信息开展研究。

研究首先基于 Python 与 DeepSeek API 构建文本情感分析方法，从微博评论文本中提取包含情感极性、强度及具体情绪类别的结构化情感特征；其次采用 K-means 聚类算法结合轮廓系数法，将南京景区划分为 6 类，并通过 LDA 主题模型挖掘情感的语义归因；最后通过多变量统计检验与描述性统计分析，验证天气、景区类型与情感的相关性。

结果表明：景区聚类呈现功能与情感协同特征，城市休闲游憩类景区情感表现最优，核心文化古迹类景区情感分化最显著；天气对情感的影响具有类型特异性，远郊自然风光类景区受能见度、降水影响最显著，核心文化古迹类景区则在雾天、雨天因“氛围增益”弱化天气负面影响；LDA 主题模型揭示情感的场景归因，自然类景区情感多与景观观赏相关，人文类景区情感多与文化共鸣绑定。

本研究为城市文旅资源优化配置、公共服务精准供给及治理效能提升提供科学依据，同时拓展了 UGC 数据在城市文旅研究中的应用场景。

关键词—城市动态系统；多源数据融合；情感分析；天气相关性；LDA 主题分析

目录

I. 引言	4
1.1 研究背景	4
1.2 研究动机与意义	4
1.2.1 研究动机	4
1.2.2 研究意义	5
II. 数据与研究方法	5
2.1 数据来源	5
2.1.1 微博签到数据	5
2.1.2 南京景点 POI 数据	6
2.1.3 南京气象数据	6
2.2 数据预处理	6
2.2.1 地理空间数据标准化与空间过滤	6
2.2.2 时序数据标准化与气象字段关联	6
2.2.3 预处理后数据集结构	7
2.3 情感分析 (DeepSeek API)	7
2.4 聚类分析	7
2.5 多维度描述性统计分析	8
2.5.1 天气细分与情感的关联分析	8
2.5.2 景区聚类与情感类型的匹配分析	8
2.5.3 情感正负倾向的群体差异分析	8
2.6 统计检验	9
2.7 LDA 主题分析	9
2.8 可视化方法	9
2.8.1 词云图可视化方法	9
2.8.2 Dash 交互式可视化方法	9
III. 实验过程与结果	10
3.1 情感分布结果	10
3.2 景点聚类结果	12
3.3 多维度关联结果	15
3.3.1 降水条件对不同景区类型情感的差异化影响	15
3.3.2 能见度等级与景区情感的联动规律	15
3.3.3 情感类型与景区类型、天气的三维关联	16
3.4 统计检验	19
3.4.1 方差分析与非参数检验：天气 - 情感关联的显著性探索	19
3.4.2 调节效应分析：能见度与降水的差异化影响	19
3.5 LDA 主题结果	21
3.6 可视化结果	22
IV. 讨论	22
4.1 情绪与天气的耦合规律	22
4.2 城市景区类型对情感分布的影响	22
4.3 研究局限	23
4.3.1 微博数据的时间局限性	23
4.3.2 LDA 主题的解释力有限	23
V. 研究总结与展望	23
5.1 研究核心结论	23
5.2 文旅实践启示	24
5.3 未来展望：多源数据融合的潜力与挑战	25
5.3.1 多源数据融合的潜力	25

5.3.2 多源数据融合的挑战	25
参考文献	27

I. 引言

1.1 研究背景

现代城市是一种高度复杂的开放动态系统，是人口流动、基础设施运行、环境变化与社会互动等多元要素的有机集合体。[1]随着城市化进程加速，至2020年人口普查，中国城市人口占比已突破63%且持续攀升[2]，我国在较短的时间内经历了西方发达国家在较长的历史时期内发生的城镇化历程，一方面推动社会经济高速发展，一方面显现许多城市问题。[3]传统依赖碎片化统计数据与人工观测的城市管理模式，已难以应对现代城市运行的多维挑战：不管突发性环境事件的及时预警，还是公共资源的动态调整，都迫切需要一套更精细、智能的决策范式支撑。

要突破这一困境，关键在于大数据技术的发展和运用。相较于传统数据，大数据展现出显著的差异化特征：其数据体量庞大，数据来源多元化，且具备随时间与空间动态变化的属性。[4]如今，测绘、遥感、传感器等技术手段的快速发展，已构建起覆盖城市空间的感知网络，催生了海量、异质的多源数据，为城市规划和资源调配提供基本依据。[5]具体来说，要对城市人口活动范围和交通流量状况进行监测，可以利用公交卡刷卡数据、社交网络数据等来进行分析[6]；利用兴趣点（Point of Interest, POI）数据则可以展示出城市产业空间界限[7]。通过对这些数据深刻挖掘，可以帮助我们揭示城市活动中隐藏的规律，为城市政策和规划提供智能决策辅助。

基于此，多源数据的融合成为提升治理效能的关键技术路径。通过多源数据的协同融合，能从多元维度视角勾勒出城市运行全貌，突破“数据孤岛”，提升整体治理体系效率，将碎片化管理转化成整体性治理[8]。

在此背景下，本研究以城市旅游景点这一典型公共空间为研究载体，探索多源数据融合支撑下的城市动态系统分析方法。景区作为城市文旅功能的核心载体，其运行状态（如游客流量、公众情感、服务负载等）是城市动态系统的“微观缩影”——通过融合景区时空数据（经纬度、日期）、天气数据与社会感知数据（游客情感分析），不仅可以揭示景区自身的运行规律，更能为理解城市公共空间的动态演化机制提供窗口，最终为城市文旅资源优化配置、公共服务精准供给与社会治理效能提升提供科学支撑。

1.2 研究动机与意义

1.2.1 研究动机

城市景点的价值不仅体现在其文化底蕴与景观特色上，更在于它能为居民和游客带来的真实体验感受。但长期以来，人们对景点的评价多依赖零散的口头反馈或抽样问卷等，这类方式往往覆盖范围有限或滞后于实际体验，难以全面捕捉大众对不同景点的真实态度。而微博等社交媒体的普及改变了这一现状——用户在游览后即时发布

的图文内容中，天然蕴含着对景点的情感表达，这些海量 UGC（用户生成内容）能帮助我们感知公众体验的。因此，借助情绪识别技术从微博签到数据中挖掘城市居民对不同景点的感受，成为填补大众真实体验感知缺口的可行路径[9]。

与此同时，在生活人们中常有这样的体会：晴天逛公园时心情更舒畅，雨天游览古迹则可能多了几分惆怅。这种气象与情绪的关联并非偶然，但对此多数认知仍停留在主观感受层面，缺乏数据支撑的系统分析。比如究竟是降水，还是能见度对情绪影响更显著？不同气象条件下，大众对自然类景点和人文类景点的情感反应是否存在差异？这些问题的答案尚未明确。随着气象数据的公开可及与情感分析技术的成熟，探索两者间的耦合机制有了坚实的技术基础，这也成为本研究想要深入挖掘的核心问题。

此外，当前城市文旅管理多依赖客流量、营收等指标，这些数据虽能反映运营状况，却缺乏对游客满意度等定量数据的深度挖掘[10]。因此，如何将抽象的体验感受转化为可量化的指标，为文旅管理提供超越传统指标的深度洞察，成为推动研究开展的现实动力。

1.2.2 研究意义

从理论层面看，本研究以南京多景点为样本，验证了社交媒体情绪数据在景点体验评估中的有效性，拓展了 UGC 数据在城市文旅研究中的应用场景。同时，通过实证分析气象情况与公众景点体验情绪的关联模式，为气候环境与旅游体验的关联研究提供了数据支撑的证据。在南京城市尺度上将社交媒体情绪数据与气象数据进行融合，构建景点体验的多源数据评估模型。

在实践层面，研究结论能为文旅管理部门与普通游客提供切实价值。对文旅管理部门而言，基于气象与情绪的耦合规律，能为制定差异化文旅服务策略、合理调配公共资源提供科学依据。对普通游客来说，研究结论可作为出行决策的参考，辅助其结合气象条件选择更适配的游览目的地。最终，研究成果能推动城市文旅从的精准服务，助力提升整体文旅体验质量。

II. 数据与研究方法

2.1 数据来源

本研究的核心数据包括微博签到数据、景点 POI 数据与气象数据三类，均通过公开渠道获取，具体信息如下：

2.1.1 微博签到数据

该数据来源于 GitHub 平台公开分享的地理数据资源库 (<https://giser2000.github.io/geodata.github.io/>) [11]，包含用户在微博平台发布的景点相关评论内容及对应的签到位置信息，核心字段涵盖用户评论文本、签到时间、地理位

置经纬度 (lon/lat) 等。数据通过网页端公开获取，记录了研究时段内南京地区用户游览景点后的实时反馈，为情感分析与空间关联研究提供了基础数据支撑。

2.1.2 南京景点 POI 数据

景点兴趣点 (POI) 数据同样取自上述 GitHub 公开地理数据资源库 (<https://giser2000.github.io/geodata.github.io/>)，属于南京市开放 POI 数据集。数据包含地点名称 (name)、地理坐标 (经度 lon、纬度 lat)、一级分类标签 (tag1)、二级分类标签 (tag2) 等结构化字段，其中地理坐标信息可直接用于空间定位与地图匹配，分类标签可支撑景点 POI 单独提取，为微博签到数据的空间关联与景点特征分析提供基础地理框架。

2.1.3 南京气象数据

气象数据来源于天气网历史气象数据平台 (<https://lishi.tianqi.com/nanjing/202311.html>) [12]，选取 2023 年 11 月 (与微博签到数据时段匹配) 的南京市逐日气象记录，核心指标包括当日天气类型 (晴、多云、雨等)、气温、风向等。该平台数据基于气象观测站点记录整理，为分析气象条件与游客情感的耦合关系提供了精准的时序环境数据。

2.2 数据预处理

为确保数据质量与分析适用性，本研究基于“空间关联 - 时序匹配”的逻辑，结合 QGIS 空间分析工具与数据清洗方法，完成数据预处理，具体流程如下：

2.2.1 地理空间数据标准化与空间过滤

(1) 坐标系统一

微博签到数据与南京景点 POI 数据原始坐标存在差异，为避免空间定位偏差，首先将两类数据的坐标统一转换为 WGS84 坐标系 (World Geodetic System 1984) —— 该坐标系是全球通用的地理信息标准[13]，也是微博签到位置服务默认采用的坐标体系，可确保经纬度数据的空间一致性。

(2) 基于 QGIS 的空间关联与过滤

借助 QGIS 完成景点与微博数据的关联[14]，首先将南京 POI 数据导入 QGIS，在 tag1 筛选“旅游景点标签”，并新建图层文件，单独导出南京旅游景点的 POI 数据；随后导入微博签到数据，利用 QGIS “点在多边形内”空间查询功能，筛选出落在各景点内的微博记录，剔除与景点无空间关联的冗余签到数据，确保最终用于情感分析的微博数据均来自游客真实游览场景，并对景区的分类进行人工筛查更改。

2.2.2 时序数据标准化与气象字段关联

以“日期”为关联键，将预处理后的南京气象数据 (含天气类型、气温、风向等字

段) 与微博签到数据进行表连接, 形成“微博评论 - 空间位置 - 当日气象”的三维关联数据集, 为分析气象与情感的耦合关系奠定数据基础。

2.2.3 预处理后数据集结构

预处理完成后, 最终形成的分析数据集包含 15 个字段, 核心字段及说明如下表所示:

数据类别	核心字段	字段说明
微博签到数据	message、ts_created、lon、lat	经筛选后的景区关联微博记录
景点 POI 数据	name、tag1、tag2	含景点分类标签 (如“文物古迹”)
气象数据	天气	匹配至当日微博数据的气象条件

2.3 情感分析 (DeepSeek API)

为实现对游客评论文本的情感信息提取, 本研究采用文本情感分析方法, 将非结构化的微博评论转化为包含情感极性、情感强度及具体情绪类别的结构化特征, 为后续相关性分析与聚类分析提供量化支撑。

该分析任务以单条游客评论文本为输入, 输出包含三类核心信息的结构化结果: 其一为情感极性, 用于判断文本情感倾向, 分为正向积极与负向消极两类; 其二为情感强度, 采用 0 至 10 的数值刻度对情感表达的强烈程度进行量化, 数值越高代表情感倾向越显著; 其三为具体情绪类别, 基于文本语义进一步细分情感类型, 如愉悦等, 实现对情感表达的精准刻画。

在技术实现层面, 本研究通过 Python 编程语言结合 DeepSeek API 完成情感分析流程。Python 的高效数据处理能力与 DeepSeek API 的自然语言理解能力形成协同, 先对原始评论文本进行预处理以去除冗余信息, 再调用 API 接口实现情感特征的自动化提取, 最终生成统一格式的结构化数据。

这类结构化情感特征不仅能够直观反映游客对不同景区的情感反馈, 还可与景区属性数据、气象数据进行多维度关联, 为后续分析天气与景区类型对情感的影响机制, 以及基于情感特征开展景区聚类提供可靠的量化基础。

2.4 聚类分析

为识别具有相似特征的景区群体, 本研究采用 K-means 聚类算法[15], 基于景区的多维度特征进行群体划分, 以 3 类变量作为聚类分析的核心特征, 分别是空间特征: 景区经纬度、类型特征: 景区二级分类标签和情感特征: 平均情感得分。

为确保聚类效果，对数据进行标准化处理，首先对分类变量采用独热编码（One-Hot Encoding）转换为数值型；之后对所有数值型变量（经纬度、平均情感得分）进行标准化，消除不同量纲对聚类距离的影响。最终形成的聚类输入矩阵包含经编码和标准化后的所有特征变量。

为确定最优簇数（K 值），采用轮廓系数（Silhouette Score）评估不同簇数下的聚类效果（取值范围 $[-1,1]$ ，越接近 1 表示聚类越合理）。通过遍历 $K=2$ 到 $K=10$ 的区间，计算每个 K 值对应的轮廓系数，选择系数最大的 K 值作为最优簇数，并绘制轮廓系数随 K 值变化的趋势图用于结果验证[17]。

基于最优簇数，使用 `sklearn.cluster.KMeans` 算法进行聚类[16]：设定随机种子确保结果可复现；并增加初始化次数（`n_init=10`）以降低初始聚类中心对结果的影响，提高聚类稳定性。

此外，由于聚类变量维度较高，采用主成分分析（PCA）将高维数据降维至二维空间（保留 2 个主成分），通过散点图直观展示不同聚类群体的分布特征，其中每个点代表一个景区，颜色区分不同聚类标签，并标注各聚类的景区数量[18]。

最终聚类结果包含各景区的名称、聚类标签、平均情感得分、主要类型、经纬度等信息，以 CSV 文件形式保存，同时输出轮廓系数图和 PCA 降维后的聚类散点图，用于后续分析与解读。

2.5 多维度描述性统计分析

以人工命名的聚类群体为基础框架，结合天气细分维度与情感特征，通过可视化工具（Matplotlib、Seaborn）构建多维度对比分析体系，核心包括以下 3 类分析方向。

2.5.1 天气细分与情感的关联分析

按“降水与否”（降水天 / 无降水天）、“能见度等级”（低能见度 / 高能见度）及原始天气类型（晴、雾~晴、多云~小雨等）对数据进行分组，采用热力图呈现不同聚类群体在各天气条件下的平均情感得分，直观识别天气对不同聚类群体的情感影响差异；借助条形图对比同一聚类群体在“降水 / 无降水”“低 / 高能见度”条件下的情感得分差异，通过箱线图展示各天气类型下情感得分的离散程度，揭示天气对情感波动的影响强度。

2.5.2 景区聚类与情感类型的匹配分析

以人工命名的聚类群体为分类依据，通过百分比热力图分析不同聚类群体中各情感类型（如愉悦、怀旧、烦躁等）的占比分布特征；构建景区聚类、天气和情感类型的三维交叉分析框架，采用热力图呈现不同聚类群体在典型天气下（晴、雾~晴、多云~小雨）TOP3 情感类型的平均强度，挖掘“天气 - 聚类群体 - 情感强度”的联动规律。

2.5.3 情感正负倾向的群体差异分析

针对各人工命名的聚类群体，通过堆叠条形图展示其内部“正面 / 负面 / 中性”情感的占比结构，对比不同群体的整体情感倾向，为群体特征命名提供实证支撑。

2.6 统计检验

为避免分析结果的偶然性，采用多组统计检验方法验证“天气 - 聚类群体 - 情感”关联的显著性，确保结论可靠性。

为验证天气对情感的影响因聚类群体而异的现象是否显著，拟进行方差齐性检验与方差分析，首先通过 Levene 检验（以中位数为中心，增强稳健性）验证不同聚类群体在各天气类型下情感得分的方差齐性；若满足方差齐性，采用双向方差分析

（Two-way ANOVA）检验“景区聚类”“天气类型”及两者交互效应对情感得分的显著影响；若不满足方差齐性或参数检验结果不显著，改用 Kruskal-Wallis 非参数检验，再次验证不同天气类型下各聚类群体情感得分的整体差异显著性。

除方差分析外，还将“降水与否”“能见度等级”转换为数值变量（如无降水天 = 1、降水天 = 0；高能见度 = 1、低能见度 = 0），通过 Pearson 相关系数计算不同聚类群体中降水与情感、能见度与情感的相关强度，验证不同景区聚类下对天气与情感之间关联的调节作用。

2.7 LDA 主题分析

为辅助解释情感聚类结果的内在成因，本研究采用 LDA（潜在狄利克雷分配）主题模型对微博签到数据进行分析。首先对文本进行预处理，包括清洗无效内容、分词并过滤无关词汇，再构建词典与词袋向量作为模型输入；通过优化后的主题模型训练，从文本中提取 4 类具有明确语义的主题，并将每条文本匹配至最可能的主题；最后结合情感聚类结果，分析不同聚类群体中各主题的分布特征，揭示主题与情感倾向的关联机制，为聚类结果提供文本语义层面的解释。

2.8 可视化方法

2.8.1 词云图可视化方法

词云图利用词频与情感强度作为权重参数，将出现频率高或情感强度大的词以更大字体显示，从而形成情绪与主题的视觉显著差异。

绘制过程中主要步骤包括文本清洗与分词，采用 jieba 对微博评论文本进行中文分词，去除停用词与无意义符号；词频统计与情感权重计算，结合 DeepSeek API 提取的情感标签与强度值，将高频且情感显著的词汇标注；可视化绘制，使用 wordcloud 与 matplotlib 生成多景区对比词云图，颜色方案依据情感极性区分。

2.8.2 Dash 交互式可视化方法

Dash 是基于 Plotly 与 Flask 的 Python 可视化框架，能够通过 Web 界面动态展示分析结果。本研究通过 Dash 构建了一个多源数据情感关联分析平台，主要模块包括

多维数据可视化，利用坐标经纬度散点图展示各景区游客气象、情绪与景区类别的分布；词云与评论展示，情感占比饼图等。

III. 实验过程与结果

3.1 情感分布结果

为系统呈现气象条件与游客情感类型的关联特征，本研究通过可视化分析与文本语义解读，从“分布规律”与“语义归因”两个维度展开探索。

首先聚焦“典型天气中各情感类型的占比”热力图，该图以天气类型为行、情感类型为列，通过颜色深浅直观呈现不同天气下各类情感的占比强度。从图中可见，愉悦情感在各天气类型中均占据最高比例，其中晴天时愉悦占比达 46.4%，多云小雨时为 44.7%，雾晴时为 43.5%，反映出南京景区整体游客情感以正向体验为主。同时，怀旧情感在雾晴天气下占比 15.7%，显著高于晴天（13.6%）与多云小雨（13.2%），这与雾天为历史景区营造的朦胧氛围密切相关，验证了天气对情感类型的场景化塑造作用。

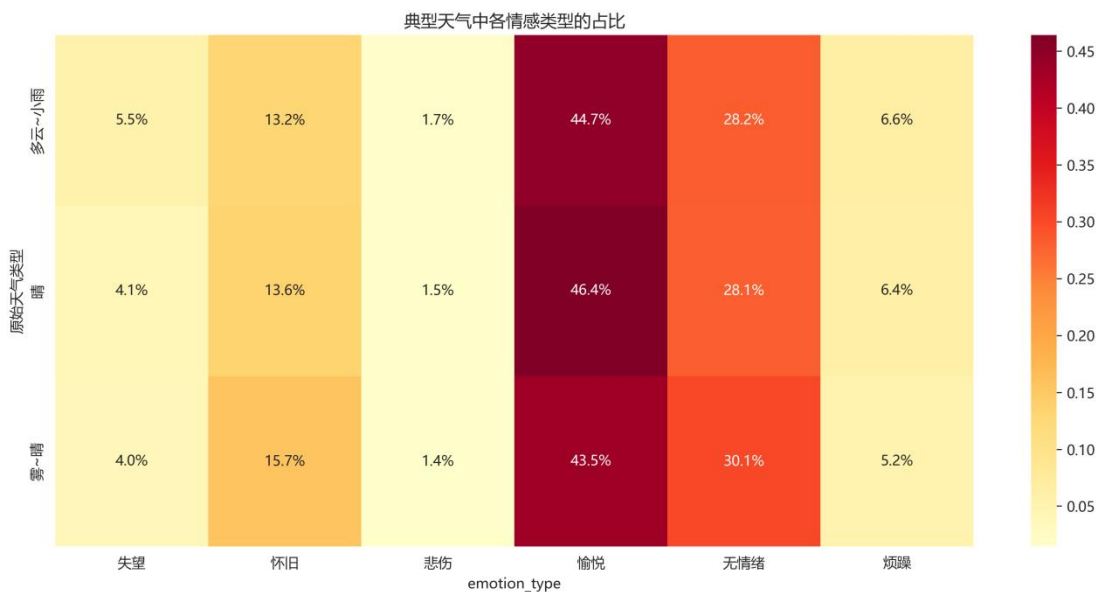


图 1 典型天气中各情感类型的占比

在此基础上，结合“愉悦”“怀旧”情感的关键词云，可从语义层面深化对情感分布的归因解释。

图 2 为“愉悦”情感的关键词词云图，内含诸如动物园、玄武湖、可爱等字眼，集

中于亲子、休闲类场景，与亲子休闲娱乐类、城市休闲游憩类景区的情感占比特征完全契合。



图2“愉悦”情感关键词云

图3为“怀旧”情感的关键词词云图，金陵、梧桐树、南京博物院等字眼则锚定历史文化场景，呼应了核心文化古迹类景区的高怀旧占比，雾天的朦胧感与秋天梧桐的历史感共同强化了这种情感归因。



图3“怀旧”情感关键词云

综上，气象条件通过景观观赏性、出行便利性、氛围营造等路径，对不同类型景区的游客情感产生差异化影响；而景区类型的功能定位（休闲、历史、科普等）则进一步强化了这种影响的场景特异性，最终形成气象、类型、情感的动态关联网。

3.2 景点聚类结果

基于经纬度、客观景区类型、平均情感得分三类核心变量，本研究通过 K-means 算法和轮廓系数法确定最优簇数，结果如下图，当簇数 $k=6$ 的时候，轮廓系数达到最高值，这表情此时的聚类效果最优，因此选择讲景区聚为 6 类来进行后续分析。

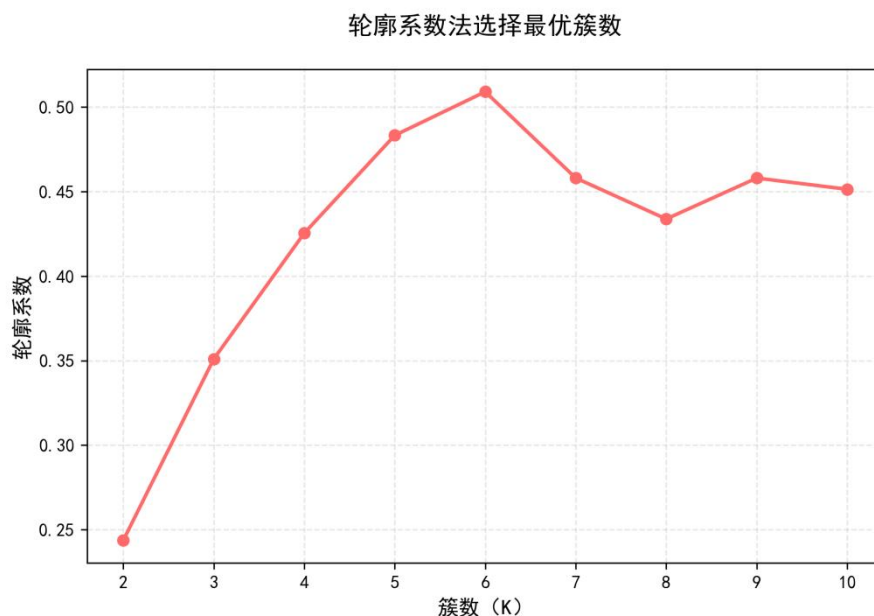


图 4 轮廓系数折线图

为直观展示景区聚类的分布特征，通过主成分分析法，将高维聚类特征降维至二维空间并绘制散点图。横轴为“主成分 1”（解释方差 21.0%），纵轴为“主成分 2”（解释方差 18.4%），两者共解释约 39.4% 的原始数据方差，虽然方差解释率较低，但因为数据本身具有的高异质性，且后续聚类结果能有效支撑研究目标，所以认为低解释率在目前场景下是可接受的[19]。从图中可见，各类别在二维空间中呈现明显的分布差异：聚类 4 作为样本量最大的群体，聚集明显，集中于主成分 1 负向区域；聚类 1、聚类 3、聚类 5 样本量较小，分布相对分散；聚类 2 和聚类 6 分别在主成分 1 的 0 值附近和正向区域形成聚集。

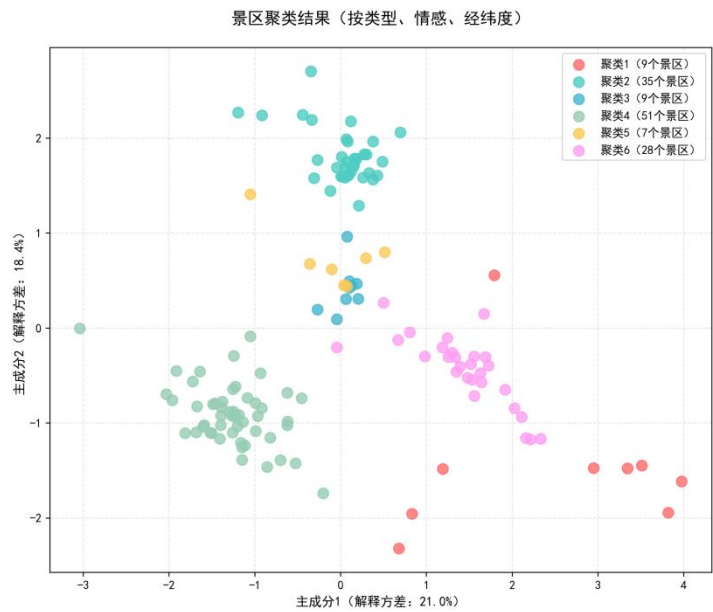


图 5 景区聚类结果

基于以上内容，最终的最优聚类数量为 6 类，共覆盖 119 个南京景区，各聚类的核心特征通过表 1 可直观区分，整体呈现“公园类景区情感得分最高、文物古迹类景区数量最集中、休闲娱乐类景区规模最小”的分布特征。

聚类标签	平均情感得分	主要景区类型	景区数量
聚类 1	2.34	风景区、公园	9
聚类 2	2.12	文物古迹	35
聚类 3	2.60	博物馆	9
聚类 4	3.12	公园	51
聚类 5	1.81	休闲娱乐类	7
聚类 6	2.37	风景区	28

根据各聚类在核心特征和空间分布上的差异化表现，对 6 个聚类群体进行人工标签化命名，确保命名能较为精准的反映每类景区的特征。

聚类一以风景区 / 公园为代表，包含濂渚洲公园、溧水秦淮源公园、固城湖国家城市湿地公园等景区，空间分布上集中在溧水、高淳一带，远离南京市区，以休闲、生态类景区为核心，记录数较少，偏冷门景区，以家庭自驾为主，为聚类一命名为近

郊生态休闲类景区。

聚类二则以文物古迹为主，包含鸡鸣寺、愚园、明孝陵、夫子庙及侵华日军南京大屠杀遇难同胞纪念馆等著名景区，涵盖南京从六朝到近代的核心历史遗存，是南京市文化旅游的必打卡地和名片，空间分布上多位于市中心区和文化核心区域，为聚类二命名为核心文化古迹类景区。

聚类三以博物馆为唯一类型，以中国南京云锦博物馆、六朝博物馆、南京博物院等博物馆为代表，是游客获取历史、艺术、科技知识的核心载体，将其标为文博知识体验类景区。

聚类四以公园为核心，遍布南京市各区域，以玄武湖公园、青奥体育公园、羊山公园、白鹭洲公园等为代表，全域均衡分布，融入南京市民的日和休闲娱乐生活中，兼具生态功能和便民属性，命名为城市休闲游憩类景区。

聚类五以休闲娱乐类为主，集中于主城区或近郊，如红山动物园或绿博园、百家湖摩天轮等广受年轻游客和家庭游客欢迎的景区，具有趣味性和体验性，将其标为休闲娱乐体验类景区。

聚类六以风景区为核心，主要分布于南京市非核心区位置，以山地风景区为主，如汤山风景区、方山风景区、栖霞山等，兼有少量历史遗址，如南京北站(旧址)，为聚类六命名为远郊自然观光类景区。

基于聚类群体的核心特征与空间分布完成人工命名后，各聚类所呈现的差异化特征可从城市文旅资源管理、服务优化与体验提升等层面提供实践支撑，具体应用价值体现在以下三方面：

其一，为文旅资源分类管理提供依据。不同聚类群体的类型属性与功能定位差异显著，如核心文化古迹类景区承载南京历史文化展示核心功能，城市休闲游憩类景区承担市民日常休闲与生态服务角色，远郊自然观光类景区则侧重户外自然体验供给。这种分类可辅助文旅管理部门明确各类资源的核心价值与保护开发方向，例如针对核心文化古迹类景区强化历史遗存保护与文化内涵挖掘，针对城市休闲游憩类景区优化便民设施与日常运维，实现分类施策，精准管理。

其二，助力差异化文旅服务策略制定。各聚类的空间分布与客群特征为服务优化提供参考：近郊生态休闲类景区与远郊自然观光类景区远离市区，可针对性完善周边交通接驳与停车配套，适配家庭自驾为主的客群需求；休闲娱乐体验类景区聚焦年轻与家庭游客，可增加互动体验项目与亲子服务设施；文博知识体验类景区则可结合游客知识获取需求，优化展览讲解服务与教育活动设计，提升不同群体的游览体验适配度。

其三，支撑文旅资源整合与线路规划。聚类结果可作为线路设计的核心依据：针对文化体验需求，整合核心文化古迹类与文博知识体验类景区，打造“历史文化深度游”线路；针对日常休闲与短途度假需求，组合城市休闲游憩类景区形成“城区休闲圈”，或串联近郊生态休闲类与远郊自然观光类景区设计“自然生态游”产品，推动南

京文旅资源从“单点开发”向“集群化整合”转型，进一步提升文旅供给的多样性与针对性。

3.3 多维度关联结果

接下来基于可视化图表与统计分析，系统呈现气象条件、景区类型与游客情感的关联特征，为理解南京城市文旅体验的动态影响机制提供实证支撑。

3.3.1 降水条件对不同景区类型情感的差异化影响

从降水/无降水天气下不同景区类型的情感得分条形图可见，降水条件对景区情感的影响呈现显著的类型异质性。在这里需要注意的是，受数据集时期限制，降水指的是“小雨”类型的天气，并非大雨。

远郊自然观光类景区在降水天的情感得分（约 5.8）远高于无降水天（约 3.4），这类景区以溧水、高淳的自然景观为主，降水天的湿润环境和低密度游客的情况提升了游客的休闲体验；核心文化古迹类景区则相反，无降水天的情感得分（约 3.0）高于降水天（约 2.1），因该类景区（如夫子庙、明孝陵）多为露天游览，且多在市中心，游客众多，降水会增加游客的出行不便感；城市休闲游憩类、亲子休闲娱乐类等城区型景区，降水与无降水天的情感得分差异较小，体现出配套设施完善对天气负面影响的缓冲作用。

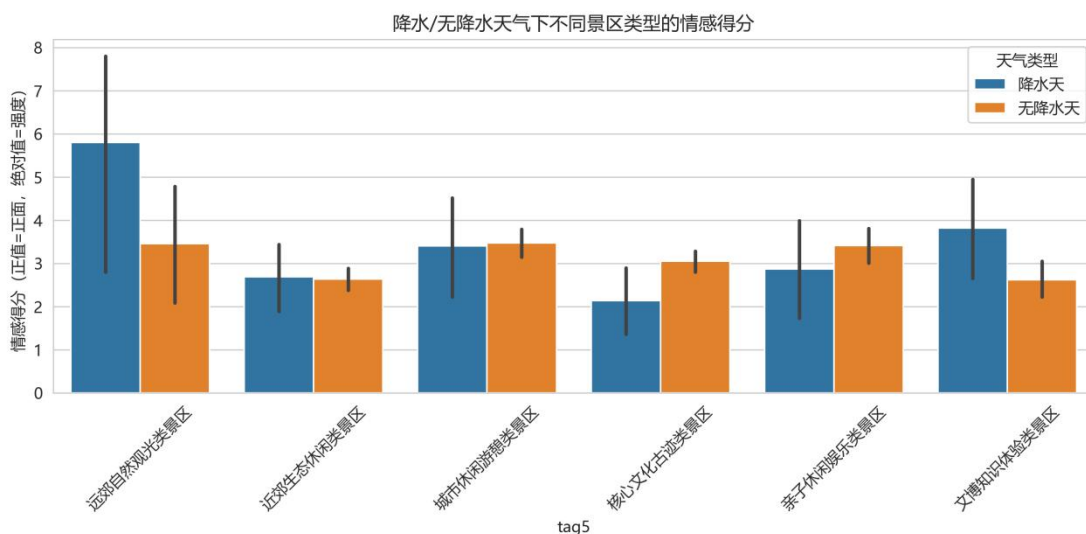


图 6 降水/无降水天气下不同景区类型的情感得分

3.3.2 能见度等级与景区情感的联动规律

低/高能见度下不同景区类型的情感得分分析揭示了能见度对情感的场景依赖性。需要注意的是，这里的低能见度为“雾”的天气类型，高能见度为“晴”的天气类型。

远郊自然观光类景区在高能见度下情感得分（约 4.0）显著高于低能见度（约

0.5)，这类景区以观赏性为核心吸引力，雾天会削弱山水、植被的视觉体验；核心文化古迹类景区在低能见度下的情感得分（约 3.0）反而略高于高能见度（约 2.8），雾天为明孝陵、老门东等历史景区营造的朦胧氛围感，强化了游客的历史怀旧情感；文博知识体验类（如南京博物院）、亲子休闲娱乐类（如红山动物园）、城市休闲游憩类的情感得分受能见度影响较小，但大多都是高能见度情况下，情感得分较高，因这类景区多在室内或核心体验对视觉条件的依赖度较低。

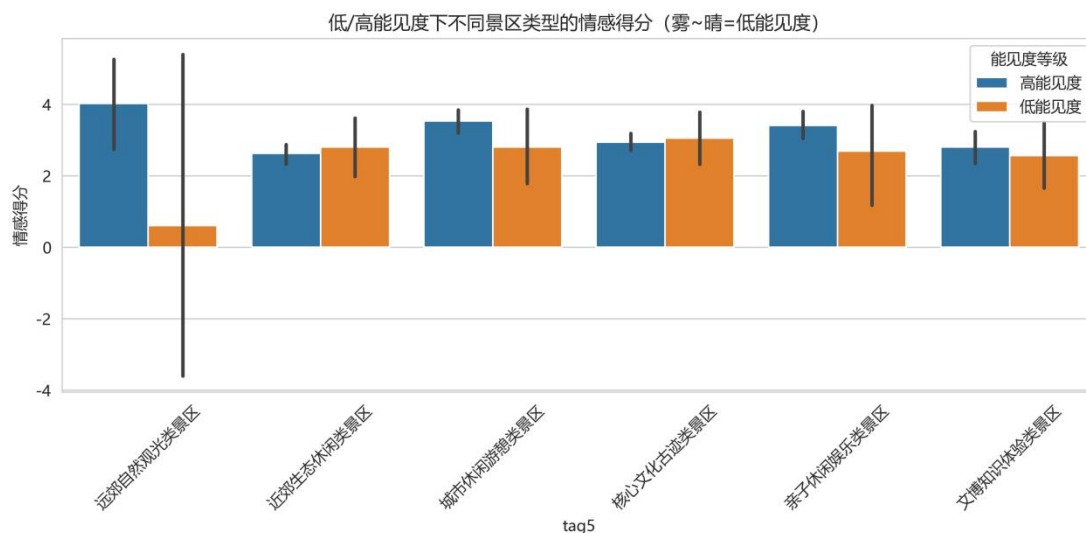


图 7 低/高能见度下不同景区类型的情感得分

3.3.3 情感类型与景区类型、天气的三维关联

为明确不同景区类型在情感类型分布上的差异，本研究绘制“不同景区类型中各情感类型的占比”热力图。该图以“景区类型”为行、“情感类型”为列，通过颜色深浅直观呈现各类情感在不同景区类型中的占比强度。

从图中可见，情感类型的分布呈现显著的景区类型特异性：亲子休闲娱乐类景区的“愉悦”情感占比最高（53.9%），这类景区以红山森林动物园、绿博园等为代表，其“趣味性、互动性”的场景设计天然适配家庭游客的情感需求；核心文化古迹类景区的“怀旧”情感占比达 17.3%，远高于其他类型，南京的明孝陵、夫子庙等历史遗存承载了六朝、明清的文化记忆，易触发游客对历史的追忆；文博知识体验类景区（如南京博物院）的“无情绪”占比最高（32.7%），反映出游客在知识获取场景中更偏向理性认知，情感表达相对克制；各类景区的“悲伤”“烦躁”等负面情感占比普遍较低（≤2.1%），说明南京景区整体的游客情感体验以正向、中性为主。

这种分布特征为后续分析“气象如何调节不同类型景区的情感表现”提供了基础——不同情感类型的占比差异，本质上源于景区功能定位的内在属性差异。

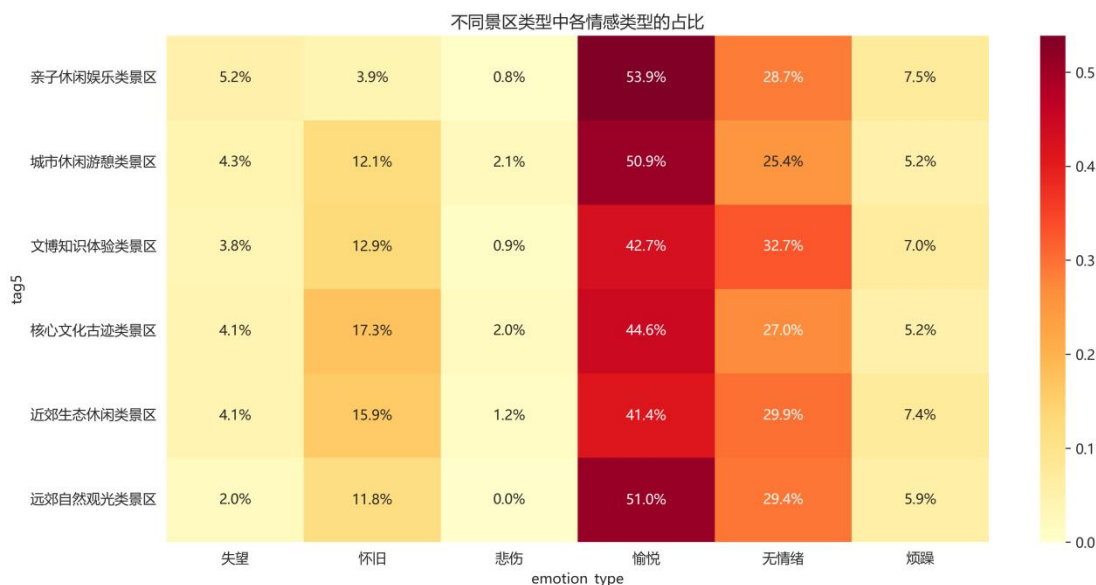


图 8 不同景区类型中各情感类型的占比

接下来，为量化分析不同天气类型对各类景区情感表现的影响，本研究绘制“全天气类型 × 景区类型的平均情感得分”热力图。该图以“景区类型”为行、“原始天气类型”为列，通过颜色深浅与数值标注直观呈现不同天气 - 类型组合下的平均情感得分。

从图中可提炼出天气 - 类型的情感适配特征：远郊自然观光类景区在“多云 ~ 小雨”天气下情感得分最高（5.8），这类景区以溧水、高淳的自然景观为主，再次证实湿润天气下的清新环境和低密度游客的情况提升了休闲体验；但在“雾 ~ 晴”天气下得分仅 0.6，可能因雾天削弱了自然景观的视觉观赏性；文博知识体验类景区（如南京博物院）在“多云 ~ 小雨”天气下得分最高（3.8），阴雨天气既减少了游客流量，又使知识获取场景更具沉浸感；晴天得分降至 2.6，猜测是因晴天游客增加干扰了观赏体验；核心文化古迹类景区的情感得分呈现“雾 ~ 晴 > 晴 > 多云 ~ 小雨”的趋势（3.1 > 3.0 > 2.1），雾天为明孝陵、老门东等历史景区营造的历史感，强化了游客的怀旧与文化共鸣；而雨天的出行不便则降低了情感表现；城市休闲游憩类和亲子休闲娱乐类都在晴天情感得分最高，雾天情感得分较低，说明这两种类型的景区依赖于光线充足、温暖的天气，雾天的出行体验会降低客群的出行意愿和舒适度。

这种差异化规律揭示了气象条件对景区情感的影响并非单一维度，而是与景区的“类型属性、空间区位、功能定位”深度耦合，最终形成多元的天气 - 情感 - 类型关联网络。

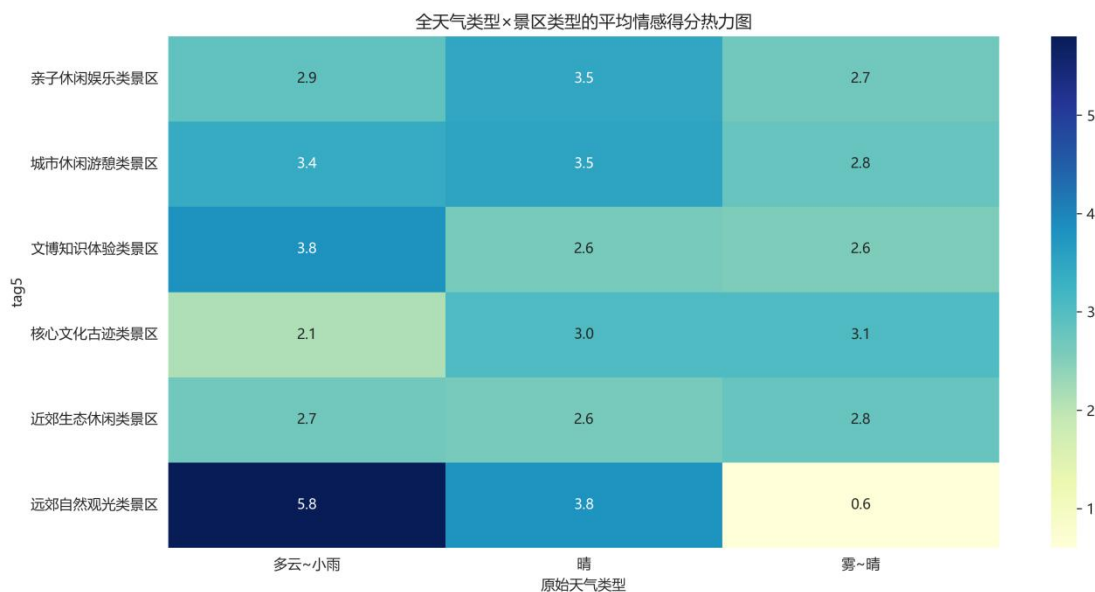


图9 全天气类型 × 景区类型的平均情感得分

为了进一步挖掘气象、景区类型和情感的深层联动机制，基于前两者的分析基础，构建了“景区 × 天气 × TOP3 情感类型的平均强度”热力图，从情感强度维度细化分析不同天气和情感类型组合下的不同景区的情感表现差异。该图以“景区类型”为行、“天气类型~情感类型”为列，通过颜色深浅与数值标注呈现各类组合的平均情感强度。

结合前文对情感类型占比与天气~景区类型得分的分析，从该图可进一步发现情感强度的场景共同点和特异性规律，总体上来说，各类型景区的愉悦情绪类型强度都较高，其次是怀旧，最后是无情绪，值得注意的是，城市休闲游憩类景区在多云~小雨天气下，怀旧情绪明显高于其他类型景区，这可能是因为云雨天气为南京城区公园增添几分老城的历史感，引发游客和市民的怀旧情绪；此外，远郊自然风光类景区在雾~晴天气下愉悦情感强度高达 9.0，是所有组合中强度最高的，这类景区的自然景观（如山水、植被）在雾天呈现出诗意的视觉效果，触发游客强烈的审美愉悦。

这种在相同点之外的强度差异进一步验证了气象可以影响场景氛围塑造，对不同类型景区的情感深度产生差异化影响的结论，最终形成情感强度的场景分异格局[20]。

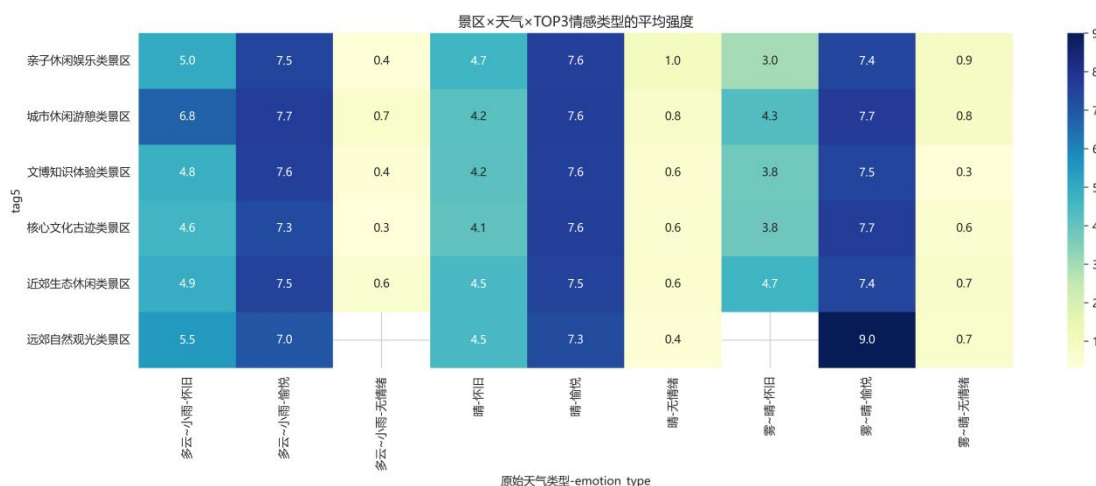


图 10 景区 × 天气 × TOP3 情感类型的平均强度

3.4 统计检验

为从统计学角度验证“天气 - 景区类型 - 情感”的关联显著性，本研究通过方差分析、非参数检验及相关性分析展开多维度验证。

3.4.1 方差分析与非参数检验：天气 - 情感关联的显著性探索

首先通过方差齐性检验确认全天气类型分组满足齐性 ($p=0.1548>0.05$)，在此基础上开展双向方差分析。

结果显示，景区类型对情感得分的主效应显著 ($F=4.3626, p=0.0006<0.05$)，说明不同类型景区的情感表现存在本质差异，这与前文聚类分析、可视化结果的结论完全契合；天气类型的主效应不显著 ($F=0.7936, p=0.4523>0.05$)，且景区类型与天气的交互效应也不显著 ($F=1.4915, p=0.1355>0.05$)。

为进一步验证，补充 Kruskal-Wallis 非参数检验，结果仍显示不同天气类型的情感得分无显著差异 (统计量 = 2.4052, $p=0.3004>0.05$)。

这类天气主效应不显著的结果，并非否定天气对情感的影响，而是由研究场景的特殊性导致：南京作为四季分明、天气类型丰富的城市，本研究所取的时段样本较为单一，时间和天气类型覆盖范围小，无法覆盖较为极端的天气，削弱了天气的差异化作用；景区部分为室内场馆，大幅降低了天气对游客体验的直接冲击，使得天气的影响更多体现为边际差异而非显著分异；情感得分的计算融合了“正负向情感的综合评价”，天气可能对“正向情感强度”或“负向情感占比”存在局部影响，但在综合得分中被抵消。

3.4.2 调节效应分析：能见度与降水的差异化影响

通过分析“不同景区类型中能见度、降水与情感的相关性”，可揭示天气对情感的类型特异性影响。

3.4.2.1 能见度的调节效应

首先，能见度与情感的相关性呈现明显的类型分异远。远郊自然观光类景区的正相关性最高（0.227），这类景区以景观观赏性为核心吸引力（如溧水天生桥、高淳慢城），高能见度下的视觉通透感整体提升游客对自然景观的审美愉悦，其次这类景区多为家庭驾车自驾，低能见度降低了出行便利性，导致情感得分降低；核心文化古迹类、近郊生态休闲类景区呈现弱负相关（-0.007、-0.011），反映出雾天为明孝陵、老门东等历史景区营造的朦胧历史感，反而能强化游客的怀旧与文化共鸣，这种“氛围增益”抵消了能见度降低的负面影响。

景区类型	能见度-情感相关系数
远郊自然观光类景区	0.227
城市休闲游憩类景区	0.044
亲子休闲娱乐类景区	0.038
文博知识体验类景区	0.018
核心文化古迹类景区	-0.007
近郊生态休闲类景区	-0.011

3.4.2.2 降水的调节效应

降水与情感的相关性同样体现类型特异性。远郊自然观光类景区呈现显著负相关（-0.155），这类景区多位于远郊（如固城湖湿地、无想山），降水会增加游客的出行成本（如自驾路况、景区内步行便利性），同时削弱自然景观的体验感，导致情感反馈明显降低；文博知识体验类景区也存在负相关（-0.083），雨天的“低流量”虽能提升场馆的沉浸感，但降水导致的“出行不便”“体感阴冷”仍会对整体体验产生负面影响；核心文化古迹类景区呈现弱正相关（0.057），夫子庙、中华门等景区的“历史街巷”在雨天形成的“烟雨金陵”场景，反而能强化游客对历史氛围的感知。

综上，统计检验的“不显著”结论是研究场景复杂性的体现，而非天气对情感无影响的证据。通过调节效应分析可见，天气对情感的影响是隐性且类型特异性的——它通过“景观观赏性、出行便利性、氛围营造感”等路径，对不同类型景区的情感产生差异化调节，这种“隐性关联”需结合可视化分析、文本语义解读才能完整揭示。

景区类型	降水-情感相关系数
核心文化古迹类景区	0.057
亲子休闲娱乐类景区	0.032
城市休闲游憩类景区	0.004
近郊生态休闲类景区	-0.003
文博知识体验类景区	-0.083
远郊自然观光类景区	-0.155

3.5 LDA 主题结果

为从语义层面解析不同景区类型、天气条件下的情感归因，本研究采用潜在狄利克雷分配（LDA）主题模型，对游客评论进行无监督主题提取，最终识别出 4 个具有明确场景特征的情感主题，以下结合关键词与场景特征进行归纳命名与分析。

主题	关键词									
0	晒太阳	鸽子	动物园	老门	分享	感受	太美	秋色	灵谷	日记
1	拍照	记忆	震撼	特种兵	晚风	还有	快乐	微笑	月亮	公园
2	枫叶	出来	有点	最好	民国	计划	世界	闪现	总统府	红叶
3	遇见	天气	开学	知道	出门	自由	烟雨	不错	体验	时光

基于主题关键词的语义聚合与场景关联，对 4 个主题的命名及核心特征如下：

主题 0：自然休闲历史赏秋主题。关键词包含“晒太阳、鸽子、动物园、老门、秋色、灵谷”等，聚合了红山动物园的亲子休闲场景、老门东的历史街区场景、灵谷寺的秋日赏景场景，体现自然互动与历史赏秋的复合情感体验。

主题 1：城市休闲与夜游欢乐主题。关键词涵盖“拍照、晚风、快乐、微笑、月亮、公园”等，锚定玄武湖、绿博园等城市公园的夜游场景，以及市民日常休闲的“晚风散步、公园拍照”场景，突出城市休闲和夜间欢乐的情感氛围。

主题 2：历史人文与季节景观主题。关键词包含“枫叶、民国、总统府、红叶”等，聚焦总统府的民国历史场景、栖霞山的秋日红叶场景，体现历史探秘以及季节景观的文化体验特征。

主题 3：天气感知与时光体验主题。关键词涉及“天气、烟雨、体验、时光”等，关联雨天的烟雨金陵场景、日常出行的自由时光场景，反映天气感知和时光沉浸的情感归因。

3.6 可视化结果

对各景区评论语料独立生成词频统计表。以情感强度为加权系数构建词云权重矩阵，利用 WordCloud 函数绘制多张景区词云图。例如，夫子庙景区的高频词包括“热闹”、“小吃”、“人多”、“夜景”，反映出典型的城市夜游特征。

Dash 平台搭建中，利用 dash 与 dash_bootstrap_components 设计前端布局，包括导航栏、景区选择下拉框、时间滑条、图表区域等；后端使用 pandas 与 plotly.express 进行数据处理与可视化渲染；通过回调函数实现“景区选择—情感趋势—词云展示”动态联动。运行后，用户可在浏览器中实时查看不同景区的情感走势、关键词云与评论样本，实现多维度探索分析。

IV. 讨论

4.1 情绪与天气的耦合规律

本研究通过统计检验与可视化分析，揭示了情绪与天气耦合的非普遍性与类型特异性规律。从整体统计结果看，天气类型对情绪得分的主效应不显著，核心原因在于数据样本量较小和城市景区具备完善的室内设施和遮蔽空间，削弱了降水、低能见度等天气对游客体验的直接冲击，使得天气的影响更多体现为“边际调节”而非“显著干预”。

但调节效应分析进一步表明，情绪与天气的耦合存在明确的类型特异性：远郊自然观光类景区中，能见度与情绪呈正相关（0.227）、降水与情绪呈负相关（-0.155），反映出这类景区的情绪体验高度依赖“景观观赏性”与“出行便利性”，高能见度提升自然景观的审美价值，降水则增加出行成本；而核心文化古迹类景区中，降水与情绪呈弱正相关（0.057）、能见度与情绪呈弱负相关（-0.007），雾天与雨天营造的“朦胧氛围”反而强化了历史场景的文化共鸣，形成“天气 - 氛围 - 情绪”的正向耦合。这种差异本质是景区功能定位的分化：自然类景区依赖天气对“景观体验”的调节，人文类景区依赖天气对“文化氛围”的塑造，最终形成情绪与天气的多元耦合路径。

4.2 城市景区类型对情感分布的影响

聚类分析与方差检验结果（景区类型主效应 $p=0.0006$ ）共同证实，城市景区类型

是决定情感分布的核心变量，且这种影响通过功能定位与体验需求的匹配实现。具体表现为三类特征：

一是高情感稳定性群体，以城市休闲游憩类景区为代表，这类景区（如玄武湖公园、青奥体育公园）兼具日常休闲与生态功能，覆盖“高频次、短时长”的市民体验与“低压力、轻互动”的游客体验，功能的复合性使得情感反馈始终维持在正向区间，且受天气、流量等外部因素影响较小；二是高情感分化群体，以核心文化古迹类景区为典型，明孝陵、夫子庙等景区因历史文化价值触发正向情感，而侵华日军南京大屠杀遇难同胞纪念馆因历史主题严肃性导致情感得分偏低，这种分化源于景区文化属性的内在差异；三是低情感波动群体，包括文博知识体验类（平均情感得分 2.60）与亲子休闲娱乐类景区（平均情感得分 1.81），前者因“知识获取”场景的理性属性，情感表达相对克制，后者因“亲子需求匹配度高”，情感反馈集中于“愉悦”维度，波动范围狭窄。

综上，景区类型通过不同的功能定位和场景设计预设预期体验，引导情感表达，最终塑造出差异化的情感分布格局，这为城市文旅资源的情感化运营提供了类型适配的实践依据。

4.3 研究局限

4.3.1 微博数据的时间局限性

本研究采用的微博数据仅覆盖 6 天样本区间，存在两方面局限：一是时间跨度较短，无法捕捉季节交替和特殊节假日（如春节、国庆）对游客情感的长期影响；二是天气样本覆盖不完整，6 天内以温和天气为主，缺乏暴雨、暴雪、高温等极端天气的样本，导致天气和情感关联的强效应场景未被识别，可能低估了极端天气对远郊自然风光类、核心文化古迹类景区的情感冲击，后续研究需扩大时间跨度至 1-3 个月，以覆盖更丰富的天气与季节场景。

4.3.2 LDA 主题的解释力有限

LDA 主题模型提取的 4 个主题虽能反映核心场景特征，但解释力仍存在边界：一方面，主题关键词的语义聚合存在模糊性，例如主题 0 中“老门”既可能指向“老门东”历史街区，也可能指向“中华门”城楼，未完全实现场景的精准对应；另一方面，主题与情感的关联仅停留在“关键词语义推断”层面，未通过量化分析（如主题情感得分、主题占比与情感强度的相关性）验证主题对情感的解释力，无法明确自然休闲与历史赏秋主题是否显著提升正向情感，后续需结合情感得分数据，构建主题与情感的量化关联模型，增强主题结果的解释深度。

V. 研究总结与展望

5.1 研究核心结论

总的来说，本研究以南京 119 个景区为研究对象，整合微博文本数据、气象数据与景区属性数据，通过聚类分析、统计检验、LDA 主题建模等方法，系统揭示了天气、景区类型与游客情感的关联机制，核心结论如下。

其一，景区聚类呈现功能和情感协同特征。通过 K-means 聚类结合轮廓系数法确定 6 类景区，其中城市休闲游憩类（51 个，平均情感得分 3.12）情感表现最优，核心文化古迹类（35 个，平均情感得分 2.12）情感分化最显著，远郊自然风光类（28 个）与近郊生态休闲类（9 个）情感受天气调节更明显，证实功能定位决定情感基调，空间区位影响情感稳定性的分布规律。

其二，天气对情感的影响具有类型特异性。整体统计检验显示天气主效应不显著，但调节效应分析证实天气的隐性影响。远郊自然风光类景区受能见度（正相关 0.227）、降水（负相关 -0.155）影响最显著；核心文化古迹类景区在雾天、雨天因“氛围增益”弱化天气负面影响；城市休闲游憩类、文博知识体验类景区因设施完善，对天气的缓冲能力更强。

其三，LDA 主题揭示情感的场景归因：4 个核心主题验证了景区类型、天气与情感的语义关联，例如自然类景区的情感多与景观观赏相关，人文类景区的情感多与文化共鸣绑定。

5.2 文旅实践启示

依据不同类型景区的天气敏感性差异，需针对性制定运营策略，实现天气影响的优化或转化。

例如，远郊自然风光类景区（如溧水天生桥、高淳慢城）应建立气象客流服务联动机制。高能见度天气（如晴天、多云）时，主推自然景观深度游产品，包括徒步、摄影打卡等，充分发挥景观观赏性优势；降水或低能见度天气时，重点优化交通接驳，增加景区内接驳车频次，同时提供场景化服务如雨中赏景向导，减少天气对游客体验的负面影响。

核心文化古迹类景区（如夫子庙、明孝陵）需挖掘天气与文化的关联场景，将天气转化为文化体验的增益要素。雾天可推出历史氛围沉浸游，例如老门东烟雨街巷游览，借朦胧氛围强化历史共鸣；雨天则设计室内文化体验活动，包括非遗手工体验、历史讲座等，规避户外游览不便的问题。

城市休闲游憩类景区（如玄武湖、青奥体育公园）应强化全天气适配设施建设。增设室内休闲空间，如景观驿站、文化展馆，同时根据天气灵活调整活动形式：晴天开展户外市集，雨天组织亲子手工课，通过设施与活动的双重适配，维持游客情感体验的稳定性。

此外，还可以结合景区类型与天气的情感适配特征，从需求导向出发设计多元文旅线路，提升线路对天气变化的适应性与体验协同性。

文化深度游线路可整合核心文化古迹类与文博知识体验类景区（如总统府、南京

博物院、老门东），该线路不受天气限制：雨天侧重室内展馆体验，引导游客深入了解文化展品；晴天则延伸至户外历史街区，实现室内外文化体验的衔接，满足游客深度文化探索需求。

自然休闲游线路串联远郊自然风光类与近郊生态休闲类景区（如无想山、固城湖湿地），并依据天气预报动态调整：高能见度天气主推自然景观一日游，让游客充分感受自然风貌；降水天气则优化为近郊休闲结合生态科普的轻体验线路，降低游客出行风险，同时保障基础休闲需求。

季节限定游线路需结合 LDA 主题的季节特征，强化季节类型情感的协同体验。秋季推出红叶与民国主题线路（如栖霞山、总统府），契合游客对秋日景观与历史文化的双重兴趣；冬季推出暖阳与休闲主题线路（如玄武湖晒太阳、红山动物园亲子游），匹配冬季游客的休闲保暖与亲子互动需求。

最后，基于 LDA 主题的语义特征，还可以针对不同类型景区设计场景化营销内容，强化天气与场景的情感联想，提升营销精准度与吸引力。

城市休闲游憩类景区营销需突出晚风、月亮、公园的夜游场景。在傍晚时段通过社交媒体推送玄武湖夜游指南，结合市民与短途游客的夜间休闲需求，吸引目标群体参与。

远郊自然风光类景区营销应强调秋色、红叶、能见度的景观优势。晴天时发布溧水天生桥秋日摄影攻略，借助优质景观视觉素材，激发游客的审美需求与出行意愿。

核心文化古迹类景区营销需挖掘烟雨、民国、历史的文化氛围。雨天推出夫子庙秦淮河烟雨游船主题活动，通过短视频传播活动场景，强化天气与文化的情感关联，提升景区文化体验的独特性感知。

5.3 未来展望：多源数据融合潜力与挑战

5.3.1 多源数据融合潜力

多源数据融合为拓展天气、景区类型和情感的关联的研究深度提供了多重可能：一是补充景区运营数据（如游客流量、开放时间、设施使用频率），可分析流量密度如何调节天气对情感的影响，例如验证雨天低流量是否提升文博知识体验类景区的沉浸感；二是纳入气象监测的微观数据（如温度、湿度、风力、能见度的小时级变化），可细化天气要素的边际效应；三是结合游客轨迹数据（如景区内停留时间、游览路线），可揭示空间移动路径如何中介天气与情感的关联，例如分析雨天游客缩短户外停留时间是否导致核心文化古迹类景区的情感得分降低。

5.3.2 多源数据融合的挑战

实践层面，多源数据融合面临三方面挑战：一是数据异构性问题，微博文本数据、气象数据、景区运营数据的格式差异较大，需通过数据标准化（如文本分词、气象指标归一化）实现融合，增加了数据预处理的复杂度；二是数据隐私与获取难度，

游客轨迹、景区运营数据多属于商业或管理机密，公开获取渠道有限，可能导致样本代表性不足；三是数据冗余与噪声，多源数据中可能存在重复信息（如微博与景区打卡数据的游客重叠）、无效信息（如无意义的文本评论、异常的气象监测值），需通过数据清洗与去重技术保障数据质量，这对研究的技术门槛提出了更高要求。

尽管存在挑战，多源数据融合仍能突破单一数据的局限，构建从宏观到微观的多层次分析框架，为城市文旅体验的动态优化提供更精准的实证支撑。

参考文献

- [1] 吴晓军.复杂性理论及其在城市系统研究中的应用[D].西北工业大学,2005.
- [2] 国家统计局, 人口普查城镇人口数 (2020 年), 2020, 国家数据平台. (Online). Available: <https://data.stats.gov.cn/easyquery.htm?cn=C01&zb=A0305&sj=2025>, Accessed: 2025-10-29.
- [3] 甄茂成,党安荣,许剑.大数据在城市规划中的应用研究综述[J].地理信息世界,2019,26(01):6-12+24.
- [4] Laney D. 3D data management: Controlling data volume, velocity and variety[J]. META group research note, 2001, 6(70): 1.
- [5] 叶宇,魏宗财,王海军.大数据时代的城市规划响应[J].规划师,2014,30(08):5-11.
- [6] 刘晓,柳林,邹健.基于公交刷卡数据的城市通勤时空分析[J].计算机应用与软件,2021,38(05):69-74+105.
- [7] 张景奇,史文宝,修春亮.POI 数据在中国城市研究中的应用[J].地理科学, 2021,41(01):140-148.DOI:10.13249/j.cnki.sgs.2021.01.015.
- [8] 刘尚钦, 张福浩, 仇阿根, 等. 基于城市信息单元的多源时空数据融合框架[J]. 集成技术, 2023, 12(3): 34-47.
- [9] 刘思叶,田原,冯雨宁,等.游客微博主题情感分析方法比较研究[J].北京大学学报(自然科学版),2018,54(04):687-692.DOI:10.13209/j.0479-8023.2018.011.
- [10] 陈盈盈, 黄乐鹏, 吕平. 沉浸式文旅发展现状影响因素研究——以杭州市为例[J]. 统计学与应用, 2025, 14(3): 43-51. <https://doi.org/10.12677/sa.2025.143056>
- [11] giser2000. 地理数据资源库 [EB/OL]. (2025-10-29). <https://giser2000.github.io/geodata.github.io/>.
- [12] 天气网. 南京 2023 年 11 月历史天气 [EB/OL]. (2025-10-29). <https://lishi.tianqi.com/nanjing/202311.html>.
- [13] Decker B L. World geodetic system 1984[J]. 1986.
- [14] QGIS Development Team, “QGIS Documentation – User Guide, Training Manual and API Reference,” QGIS Project, <https://docs.qgis.org/>, accessed Sep. 29, 2025.
- [15] McQueen J B. Some methods of classification and analysis of multivariate observations[C]//Proc. of 5th Berkeley Symposium on Math. Stat. and Prob. 1967: 281-297.
- [16] sklearn Developers, “sklearn.cluster.KMeans — scikit-learn 1.3.0 documentation,” <https://scikit-learn.org/1.3/modules/generated/sklearn.cluster.KMeans.html>, accessed Sep. 29, 2025.
- [17] P. J. Rousseeuw, “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,” Journal of Computational and Applied Mathematics, vol. 20, pp. 53-65, 1987.

- [18]I. T. Jolliffe, **Principal Component Analysis**, 2nd ed. New York: Springer, 2002.
- [19] K. Leufkens, Data Visualization using Principal Component Analysis within the Digital Humanities. Master's thesis, Dept. Media and Culture Stud., Utrecht Univ., Utrecht, The Netherlands, Feb. 2020.
- [20]Jeuring J H G, Peters K B M. The influence of the weather on tourist experiences: Analysing travel blog narratives[J]. *Journal of Vacation Marketing*, 2013, 19(3): 209-219.