

景区游客画像分析模型

一、模型概述

本模型旨在通过对 X 市景区游客多维度信息的分析，聚焦于游客来源地分布、年龄分布以及性别分布等方面，借助 GIS 技术直观呈现所选景区游客的特征画像，为景区运营管理、旅游市场营销以及旅游资源的精准配置等提供详实的数据支撑，助力提升景区服务质量，增强景区对不同游客群体的吸引力，推动 X 市旅游产业的科学发展。

二、数据来源与整理

景区票务系统数据：景区票务售卖与入园登记系统中记录了游客的购票信息，包含身份证号码（可据此判断游客来源地）、出生日期（用于推算年龄）以及部分可能体现性别的相关字段（如姓名等结合常见规律判断）等内容，这些数据是分析游客来源地、年龄和性别分布的重要基础，能直接反映购买门票进入景区游客的基本特征情况。

在线旅游平台预订数据：各大在线旅游平台（OTA）上游客预订 X 市景区门票、旅游产品等相关信息，同样涵盖游客出发地（对应来源地）、年龄区间（部分平台有收集）以及性别选项等内容，通过与景区票务系统数据相互补充，进一步扩大数据样本量，尤其是对于通过线上渠道预订游览的游客群体特征能更全面地掌握，使分析结果更具代表性。

景区游客问卷调查数据：在景区内随机对游客开展问卷调查，收集游客自愿提供的个人信息，包括来自哪里、年龄范围、性别等，同时还能获取游客对景区的评价、游览目的等额外信息，从游客主观反馈角度完善数据，对一些系统中可能缺失或不准确的信息进行补充核实，提高整体数据质量。

数据整理流程：首先，将从上述多渠道收集到的数据汇总到统一的数据仓库中。接着，运用数据清洗工具和人工核查相结合的方法，去除重复、错误以及逻辑不清晰的数据记录，例如纠正票务系统中因录入错误导致的身份证号码异常、剔除

在线旅游平台上明显虚假的预订信息等。然后，针对不同来源的数据，按照统一的格式提取游客来源地（精确到省级行政区）、年龄（划分合适的年龄区间，如 0 - 18 岁、19 - 30 岁等）、性别等关键字段信息，并进行分类整理，为后续的统计分析做好准备。

三、核心算法与分析逻辑

游客来源地分布统计：借助数据库管理系统（如 MySQL、Oracle 等）或数据分析软件（如 Python 的数据分析库 pandas 等），依据整理好的数据中代表游客来源地省级行政区的字段，运用分组和计数函数（如 SQL 语句中的 GROUP BY 和 COUNT 函数，或 pandas 中的 groupby 和 size 方法），将来自同一省级行政区的游客数据归为一组，统计每组内的游客记录数量，再通过计算各省级行政区游客数量占总游客数量的比例，得出各省级行政区游客在所选景区的占比情况，以此清晰呈现游客来源地在全国各省级行政区的分布特征。

游客年龄分布统计：按照事先划分好的年龄区间，同样运用上述分组和计数方法，依据代表游客年龄的字段，将游客数据归入相应的年龄区间组内，统计各年龄区间内的游客数量，并计算其占总游客数量的比例，得到景区游客年龄的分布情况，了解不同年龄段游客对景区的青睐程度差异。

游客性别分布统计：依据代表游客性别的字段，运用计数函数分别统计男性游客和女性游客的数量，再计算各自占总游客数量的比例，明确景区游客的性别分布特征，为景区针对性地开展服务和营销提供依据。

四、模型输出与可视化

GIS 地图展示游客来源地分布：利用地理信息系统（GIS）技术，以全国地图为底图，将各省级行政区按照所选景区游客的来源地占比情况进行可视化呈现。可以采用不同的颜色深浅或圆形大小等方式来表示占比的高低（例如，占比越高，颜色越深或圆形越大），在地图上直观展示出哪些省级行政区是所选景区的主要客源地，方便快速查看游客的地域来源特征以及地域差异。同时，为地图添加交

互功能,当鼠标指针悬停在某一省级行政区上时,在地图上方或旁边弹出信息框,清晰展示该省级行政区的名称以及对应的游客数量和占比数值。

柱状图展示年龄分布：生成以年龄区间（如 0 - 18 岁、19 - 30 岁等）为横坐标，各年龄区间游客数量占比为纵坐标的柱状图，不同颜色的柱子代表不同年龄区间，柱子的高低直观反映各年龄区间游客在景区游客总量中的占比情况，通过柱状图清晰对比不同年龄段游客的分布差异，便于整体把握景区游客的年龄结构特点。

饼图展示性别分布：制作饼图，将整个圆分为两部分，分别代表男性游客和女性游客，扇形面积大小对应各自占总游客数量的比例，同时在饼图上标注男性、女性以及对应的占比数值，一目了然地呈现景区游客的性别构成情况。

五、模拟数据示例与可视化展示

假设对 X 市某景区进行分析，有以下模拟的游客画像相关数据：

省级行政区	游客数量（人）	占比（%）
广东省	500	25
江苏省	300	15
浙江省	250	12.5
山东省	200	10
其他省级行政区（汇总）	750	37.5

游客年龄分布（模拟划分年龄区间及对应数据）：

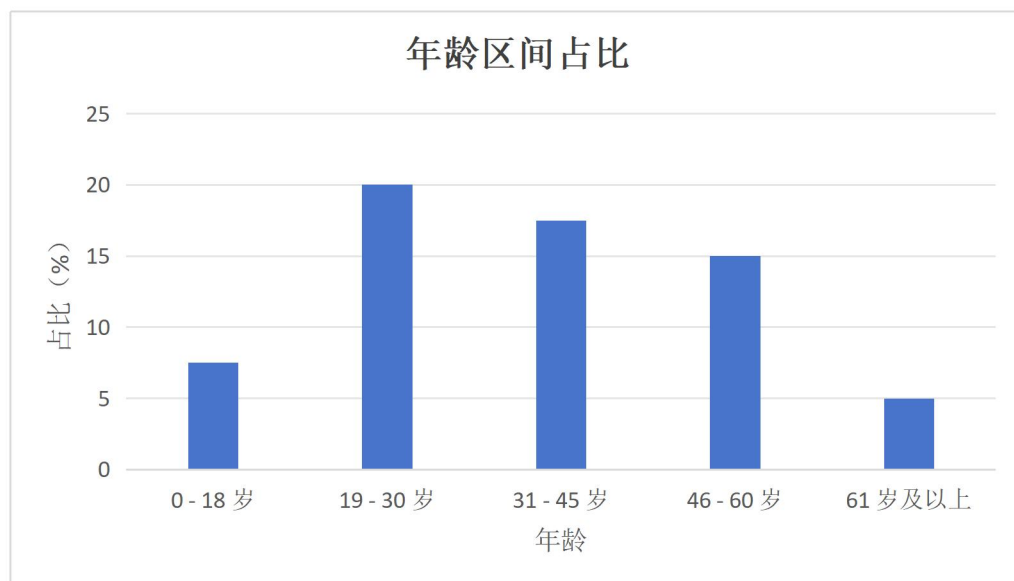
年龄区间	游客数量（人）	占比（%）
------	---------	-------

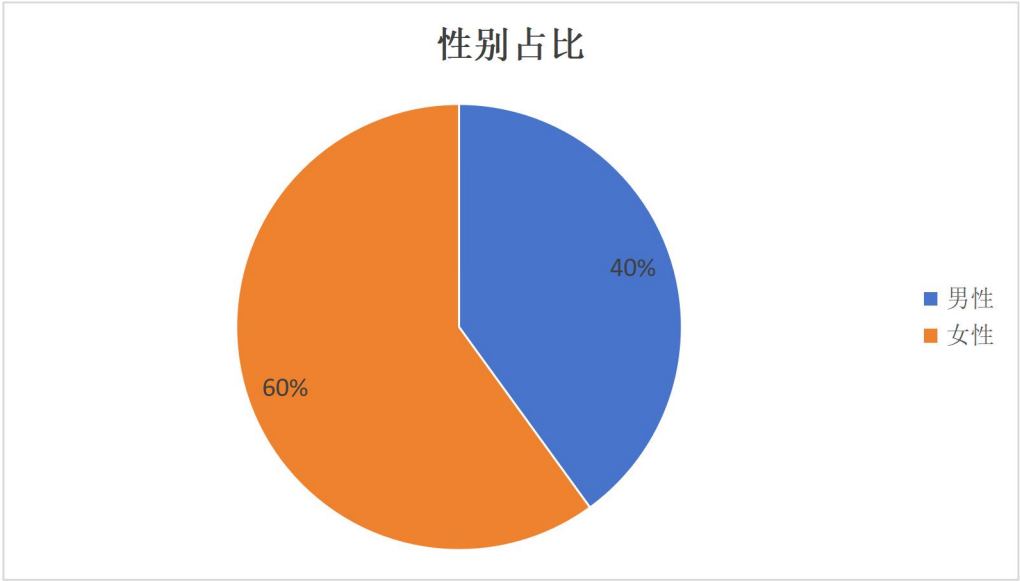
年龄区间	游客数量（人）	占比（%）
0 - 18 岁	150	7.5
19 - 30 岁	400	20
31 - 45 岁	350	17.5
46 - 60 岁	300	15
61 岁及以上	100	5

游客性别分布：

性别	游客数量（人）	占比（%）
男性	800	40
女性	1200	60

对应的可视化展示如下：





以下是使用 Python 的 `pyecharts` 库生成满足需求的可视化地图的方法：

收起

python

```
from pyecharts import options as optsfrom pyecharts.charts import Map
```

```
# 假设数据已经整理成如下格式的字典
```

```
data = {  
    "广东省": [500, 25],  
    "江苏省": [300, 15],  
    "浙江省": [250, 12.5],  
    "山东省": [200, 10],  
    "其他省级行政区（汇总）": [750, 37.5]}
```

```
def generate_visualization_map():  
    # 创建地图对象  
    map_chart = (  
        Map()  
        .add(  
            "",  
            [(k, v[1]) for k, v in data.items()], # 省份和占比数据  
            maptype="china",  
            is_map_symbol_show=False,  
        )  
        .set_global_opts(  
            title_opts=opts.TitleOpts(title="景区游客来源地占比地图"),
```

```

        visualmap_opts=opts.VisualMapOpts(

            min_=min([v[1] for v in data.values()]), # 占比最小值

            max_=max([v[1] for v in data.values()]), # 占比最大值

            is_calculable=True,

            range_color=["#313695", "#4575b4", "#74add1", "#abd9e9", "#e0f3f8", "#ffffbf",
"#fee090", "#fdae61", "#f46d43", "#d73027"], # 颜色渐变范围

        ),

        tooltip_opts=opts.TooltipOpts(

            trigger="item",

            formatter=lambda x: f"{x['name']}<br>游客数量: {data[x['name']][0]} 人<br>占比: {data[x['name']][1]}%"

        )

    )

)

# 渲染地图为 HTML 文件

map_chart.render("tourist_source_map.html")

generate_visualization_map()

```

在上述代码中：

- 首先定义了数据字典 **data**，其中键为省级行政区名称，值为包含游客数量和占比的列表。
- 在 Map 图表的创建过程中，通过 **add** 方法添加数据，指定地图类型为 **china** 并隐藏地图符号。
- **set_global_opts** 方法用于设置全局选项，包括标题、视觉映射和提示框。视觉映射根据数据中的占比范围设置颜色渐变，提示框的 **formatter** 函数自定义了鼠标悬停时显示的信息格式，包括省份名称、游客数量和占比数值。
- 最后使用 **render** 方法将地图渲染为 **tourist_source_map.html** 文件，在浏览器中打开该文件即可查看可视化地图效果。

请注意，在实际应用中，你需要根据实际的数据获取和处理方式来替换示例中的数据部分，确保数据的准确性和完整性。同时，如果需要更精细的地图定制，如添加地理边界、调整地图投影等，可以进一步查阅 **pyecharts** 库的官方文档。

（此处应插入一个简单的全国地图示例，各省级行政区用不同颜色或大小体现所

选景区游客来源地占比情况，鼠标悬停有信息框展示省份名称、游客数量及占比数值示例图；插入一个柱状图示例，横坐标为各年龄区间，纵坐标为占比数值，柱子高低体现各年龄区间占比差异，鼠标悬停有交互效果示例图；插入一个饼图示例，两扇形分别代表男性和女性，面积体现占比大小，标注性别及占比数值，鼠标悬停有交互效果示例图）

六、应用场景与价值

景区精准营销：景区运营者可依据模型输出的游客画像信息，制定精准的市场营销策略。针对主要客源地省级行政区，加大在当地的宣传推广力度，如开展旅游推介会、投放广告等；根据不同年龄、性别的游客偏好，设计差异化的旅游产品和活动，例如，对于年轻游客居多的情况，推出更具刺激性、时尚感的体验项目，提高景区对目标游客群体的吸引力，提升营销效果，增加景区游客量。

服务优化与资源配置：了解游客的年龄、性别等特征后，景区可以更合理地配置服务资源。比如，若老年游客占比较大，增加无障碍设施、设置更多休息区域，并配备相应的医疗急救服务；针对女性游客较多的情况，在景区内优化卫生间布局、增加母婴设施等，通过优化服务提升游客满意度，增强景区的竞争力。

旅游产品开发与创新：旅游企业结合景区游客画像分析结果，开发更贴合不同游客群体需求的旅游产品。例如，针对来自文化底蕴深厚地区的游客，设计深度文化体验游线路；对于亲子游游客较多的情况（通过年龄分布判断），打造亲子互动主题的旅游项目，促进 X 市旅游产品的多样化发展，满足不同游客的个性化需求，推动旅游产业的可持续发展。