

山东大学计算机科学与技术学院

大数据分析实践课程实验报告

学号: 202300130153	姓名: 吴宇轩	班级: 23 数据
实验题目: 手机数据采集与分析实践		
实验学时: 2	实验日期: 2025. 12. 14	

作品描述:

1. 数据集来源:

阿里天池数据集: bike-rental dataset <https://tianchi.aliyun.com/dataset/93509>

数据集描述: 共享单车的出租过程与环境和季节性环境高度相关。例如, 天气条件, 降水, 星期几, 季节, 一天中的小时等等可能会影响出租行为。核心数据集与两年历史记录相对应, 该历史记录对应于美国华盛顿特区 Capital Bikeshare 系统的 2011 年和 2012 年。

2. 思路:

通过对 day.csv 和 hour.csv 的联合分析, 我们发现了以下三个核心行为模式:

1. “双峰”通勤效应 vs “钟形”休闲效应 (Hourly Pattern)

这是数据中最显著的模式。我们将数据按“工作日(Working Day)”和“非工作日(Non-working Day)”拆分, 观察每天 24 小时的流量变化。

工作日 (Working Days): 呈现明显的早晚高峰双峰结构。

早高峰出现在 08:00。

晚高峰出现在 17:00 – 18:00。

结论: 这表明共享单车在工作日主要作为通勤工具 (解决“最后一公里”问题)。

非工作日 (Weekends/Holidays): 呈现平滑的钟形曲线 (正态分布)。

流量从上午 10 点开始逐渐上升, 在 12:00 – 15:00 达到峰值。

结论: 周末的骑行更多是为了休闲娱乐、健身或旅游, 用户倾向于在气温较舒适的下午出行。

2. 用户分层的行为差异 (User Segmentation)

我们将用户拆分为 Registered (注册会员) 和 Casual (临时用户) 进行对比, 发现他们的行为逻辑截然不同。

注册用户 (Registered):

对工作日高度敏感, 构成了早晚高峰的主力军。

对天气有一定的忍耐度 (因为是刚需通勤)。

临时用户 (Casual):

主要活跃在周末和每天的中午/下午时段。、

对环境非常敏感, 呈现“看天骑车”的特性。

商业洞察: 临时用户是旅游和周末经济的主要贡献者, 而注册用户提供了稳定的工作日收入流。

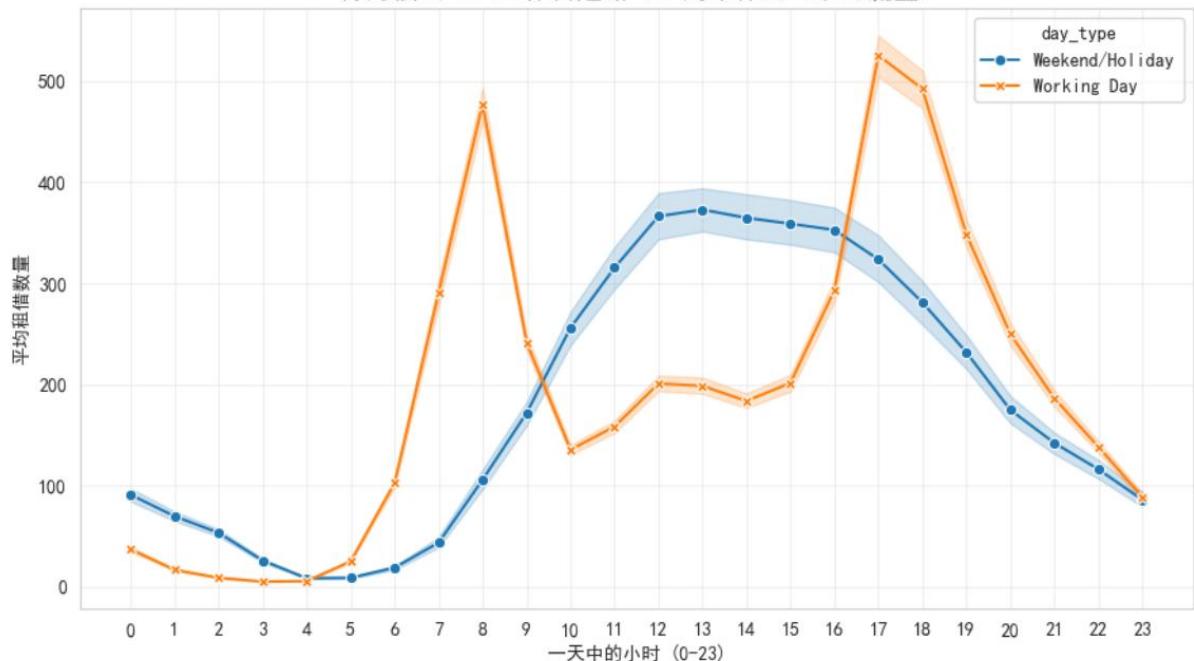
3. 季节与温度的非线性关系 (Environmental Impact)

季节性：骑行量在 秋季 (Fall) 达到最高，其次是夏季和春季，冬季最低。有趣的是，虽然夏天温度高，但华盛顿特区的秋季（通常指 9-11 月）气候最宜人，因此骑行量反而超过夏季。

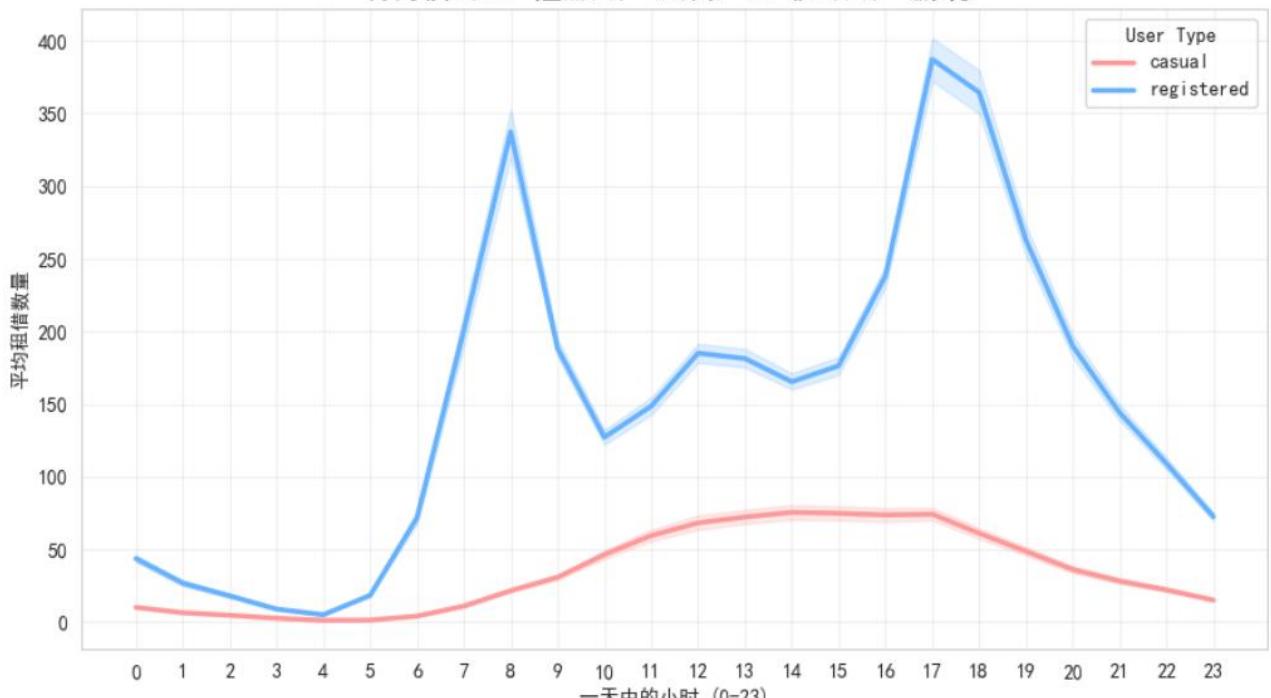
温度拐点：骑行量与温度呈正相关，但存在一个“舒适区”。当体感温度过高（酷热）或过低（严寒）时，骑行量都会下降。最佳骑行体感温度大约在 $25^{\circ}\text{C} - 30^{\circ}\text{C}$ 之间（对应归一化数据的 0.6-0.7 左右）。

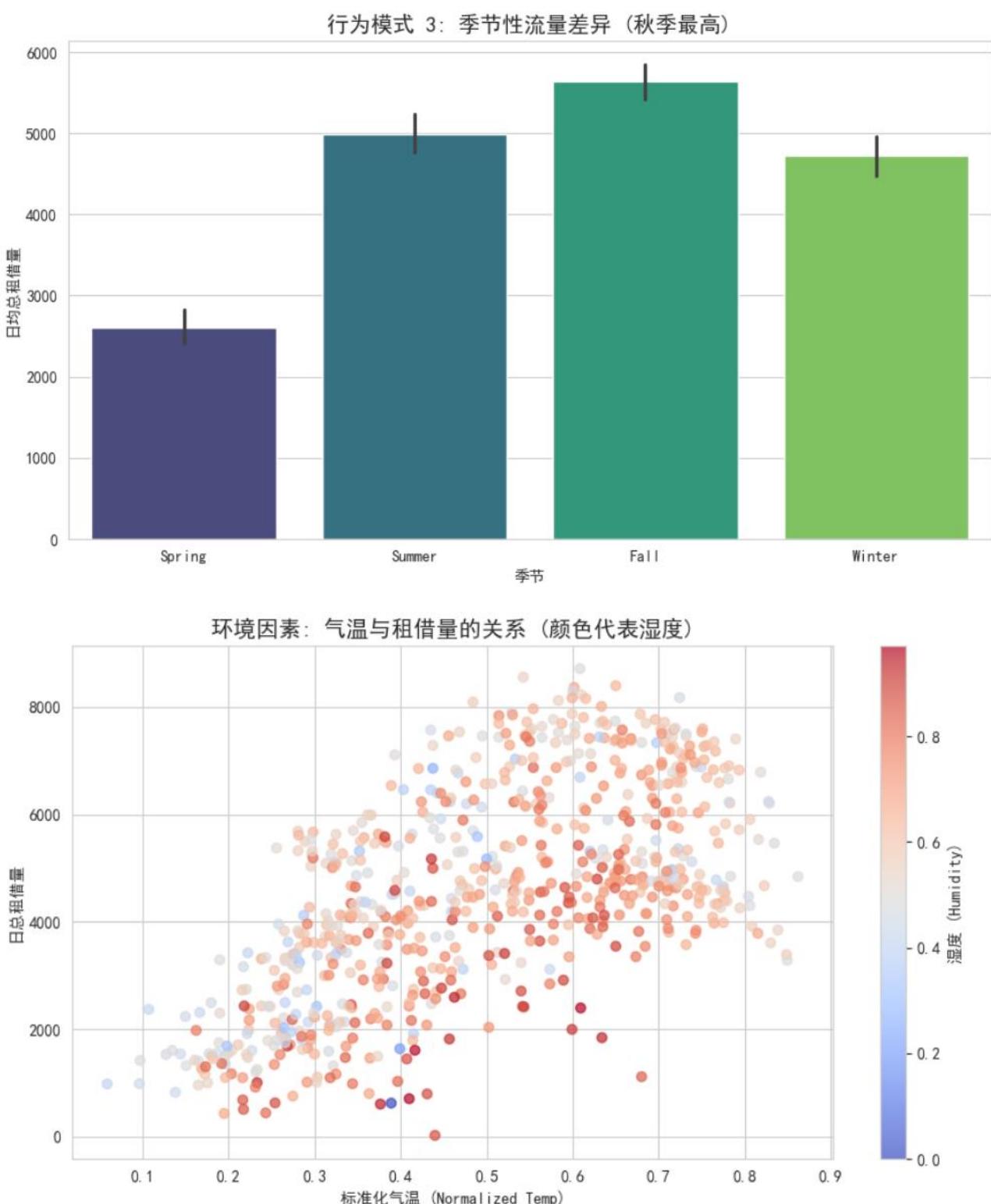
结果图片：

行为模式 1：工作日通勤 vs 周末休闲（小时流量）



行为模式 2：注册用户(刚需) vs 临时用户(游玩)





图片解读:

第一张图 (Line Plot): 你会清晰地看到红线 (工作日) 有两个尖锐的峰 (8 点和 17 点) , 而蓝线 (周末) 是一个平滑的弧线。这直观地证明了该系统是混合型系统 (兼具通勤和旅游功能)。

第二张图 (Line Plot): 对比 Casual 和 Registered。你会发现 Casual 用户 (红色) 的曲线非常平缓, 主要集中在白天; 而 Registered 用户 (蓝色) 几乎完全复刻了通勤的高峰曲

线。这意味着如果你想推广会员卡，应该针对通勤族；如果你想卖单次票，应该针对游客。

第三张图(Bar Plot)：展示了秋季(Fall)是骑行量最高的季节。这可能与华盛顿特区的气候有关（夏天可能过于闷热潮湿，冬天太冷，秋天最适合户外活动）。

第四张图(Scatter Plot)：展示了气温越高（向右），骑行人数通常越多，但颜色（湿度）也揭示了，如果在高温高湿（右上角红色点）的情况下，人数可能会不如预期多。