

《城市降雨对地铁短期OD客流的影响路径与机制研究 ——基于格兰杰因果检验和双重机器学习模型》

团队名称:

组 长:

组 员:

答辩日期:

1 团队成员介绍 与分工

我们 100% 的作业任务在 2023/07/01 - 2023/07/17 本次专业实习期间完成。

项目开发进度表

项目名称：城市降雨对地铁短期OD客流的影响路径与机制研究——基于格兰杰因果检验和双重机器学习模型

序号	任务名称	开始日期	结束日期	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	作业选题	2023/07/01	2023/07/05																	
2	文献阅读和算法学习	2023/07/02	2023/07/14																	
3	数据采集和预处理	2023/07/05	2023/07/09																	
4	时间序列预测算法	2023/07/08	2023/07/11																	
5	聚类算法	2023/07/09	2023/07/11																	
6	格兰杰因果检验	2023/07/12	2023/07/15																	
7	双重机器学习算法	2023/07/12	2023/07/17																	
8	可视化绘图	2023/07/13	2023/07/17																	
9	报告撰写和ppt制作	2023/07/15	2023/07/17																	

小组成员共同完成：

选题和文献阅读、数据采集及预处理、可视化绘图、报告撰写及PPT制作

：时间序列预测算法

：K-Means聚类算法

：因果推断算法

2 项目背景及意义

- 2.1 研究背景

地铁作为城市出行中必不可少的交通方式，是城市出行中重要的组成部分，而降雨天气对于交通系统的影响是极大的，天气类型对交通安全、交通需求和交通模式具有显著影响。另外天气因素对于出行行为也有较大影响，恶劣的天气的出现将减少出行活动。

地铁OD (Origin-Destination) 客流是乘客出行动态分布的一个最直观的展现，它包括了旅客的进站和出站的时间信息和站点信息，可以充分反应居民乘坐地铁的出行状态，是网络化运营条件下地铁行车组织、客运组织和乘客服务的基础支撑数据，蕴含着非常有价值的信息，对高效协调轨道交通运出能力和客运需求具有重要意义。

2 项目背景及意义

- 2.2 研究假设

H1：城市降雨能够降低地铁OD客流的总客流量。

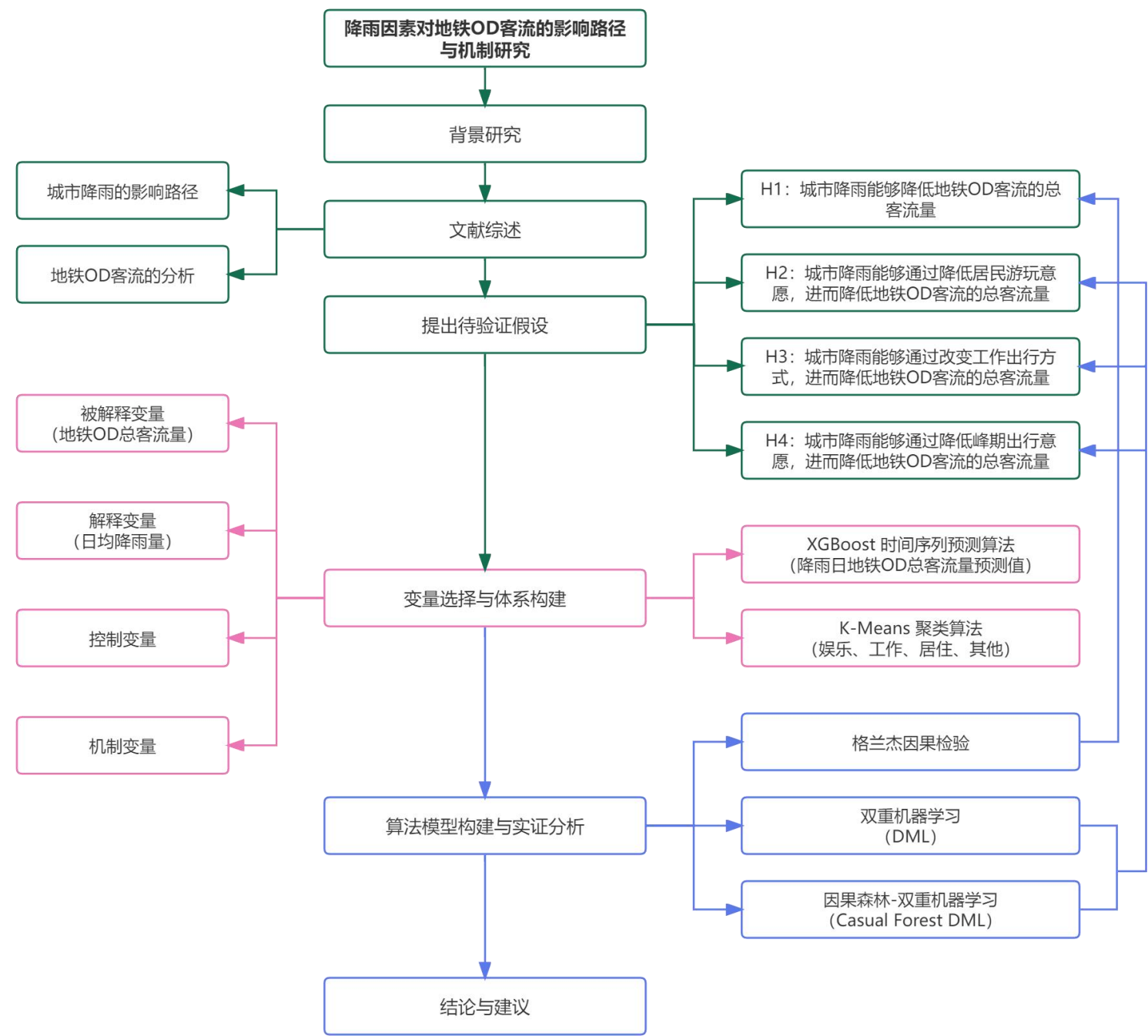
H2：城市降雨能够通过降低居民游玩意愿，进而降低地铁OD客流的总客流量。

H3：城市降雨能够通过改变工作出行方式，进而降低地铁OD客流的总客流量。

H4：城市降雨能够通过降低高峰期出行意愿，进而降低地铁OD客流的总客流量。

2 项目背景及意义

- 2.3 技术路线流程图



3 项目开发过程及 方法

- 3.1 数据来源

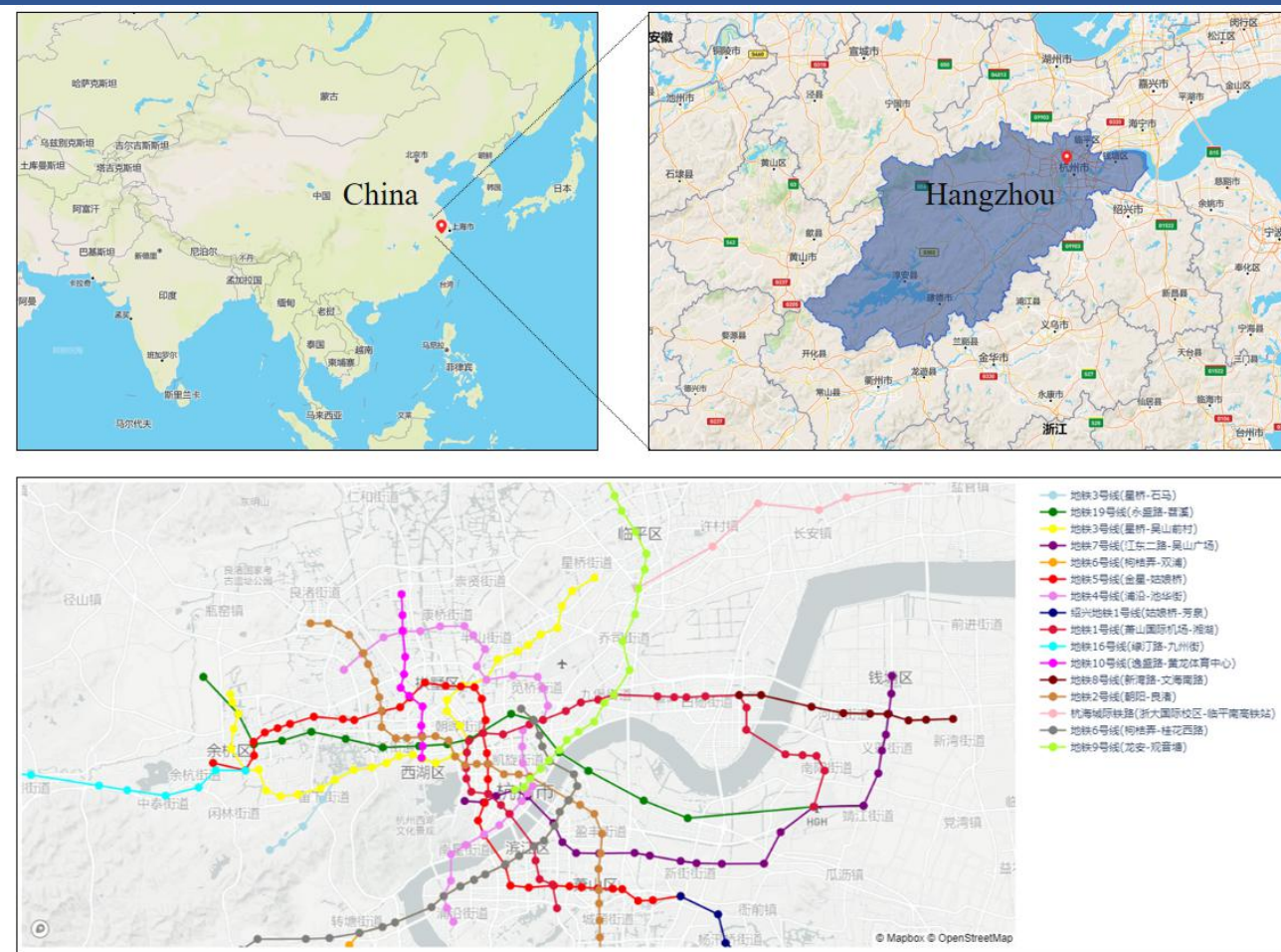
研究对象：杭州市的地铁OD客流和降雨数据

数据区间：2019年1月1日至1月25日

降雨数据：来自欧洲中期天气中心

地铁OD客流数据：来自阿里云天池大数据竞赛所开放的数据资源

其他控制变量：来自“2345天气王平台”查询的历史天气数据，包括气温、空气污染水平和风力级别。



3

项目开发过程及方法

- 3.2 变量设置与选择

- 1.被解释变量：地铁OD客流水平 (MOD)
- 2.解释变量：杭州城市降雨量
- 3.控制变量：最低气温 (Ltp)、平均气温 (Atp)、最高气温 (Htp)、空气污染指数 (Air)、风力等级 (Wnd)
- 4.机制变量：降低居民游玩意愿 (Rwp)、改变工作出行方式 (Cwm)、降低峰期出行意愿 (Ptw)

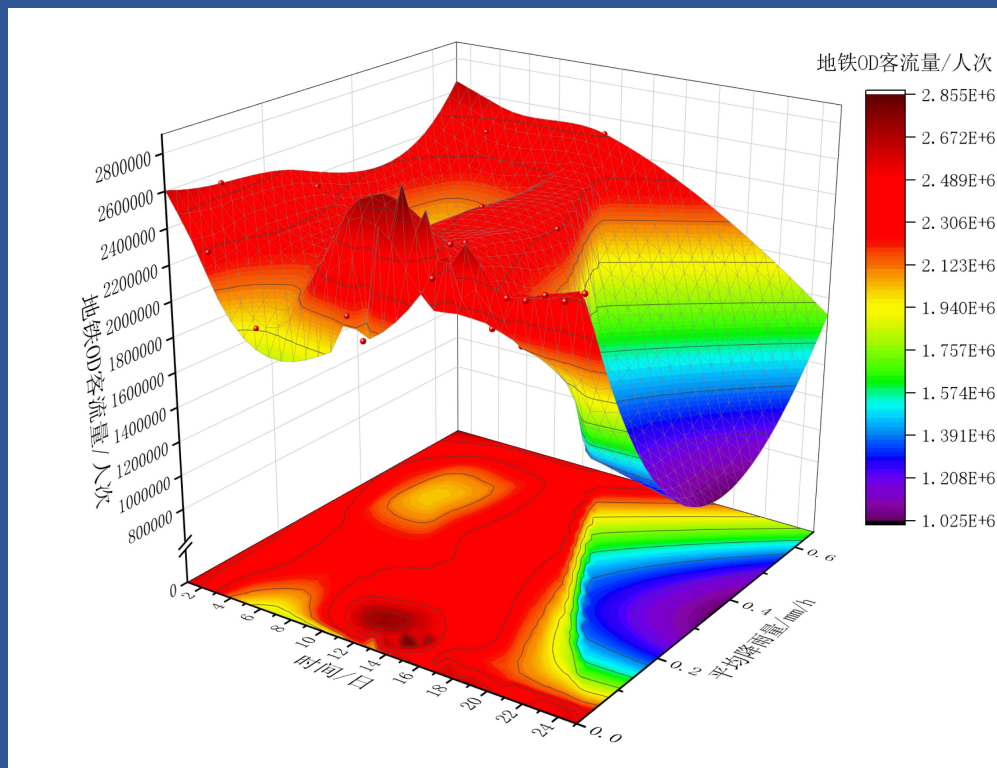
主要变量描述性统计

变量类型	变量	样本量	均值	标准差	最小值	最大值
被解释变量	MOD	25	2345490	183526	1948260	2664404
解释变量	Rain	25	0.156	0.200	0	0.662
控制变量	Ltp	25	2.880	1.878	-1	5
	Atp	25	5.520	1.311	1.5	8
	Htp	25	8.160	2.095	4	13
	Air	25	85.48	33.58	28	155
	Wnd	25	1.760	0.597	1	3

3 项目开发过程及 方法

- 3.3 变量可视化

根据前述统计得到的各日期下的降雨量和地铁OD客流水平，绘制3D曲面图：其中，平均降雨量较低的日期，地铁OD客流量普遍较高；而出现降雨的日期，地铁OD客流量随降雨量的变化呈现了不同程度的下降。



杭州市平均降雨量与地铁OD客流量变化曲面 (2019年1月1日-1月25日)

3 项目开发过程及 方法

- 3.4 时间序列预测算法

通过XGBoost进行时间序列数据预测，得到实际降雨日的地铁OD客流在非降雨的假设下的预测值。

1月29日站点27的进展人数和出站人数预测值（部分）

站点编号	起始时间	终止时间	进站人数（预计）	出站人数（预计）
27	2019/1/29 6:00	2019/1/29 6:10	51	1
27	2019/1/29 6:10	2019/1/29 6:20	67	1
27	2019/1/29 6:20	2019/1/29 6:30	71	6
27	2019/1/29 6:30	2019/1/29 6:40	86	5
27	2019/1/29 6:40	2019/1/29 6:50	134	11
27	2019/1/29 6:50	2019/1/29 7:00	169	13
27	2019/1/29 7:00	2019/1/29 7:10	200	17
27	2019/1/29 7:10	2019/1/29 7:20	290	20
27	2019/1/29 7:20	2019/1/29 7:30	335	20
27	2019/1/29 7:30	2019/1/29 7:40	373	39
27	2019/1/29 7:40	2019/1/29 7:50	390	122
27	2019/1/29 7:50	2019/1/29 8:00	415	54
27	2019/1/29 8:00	2019/1/29 8:10	404	73
27	2019/1/29 8:10	2019/1/29 8:20	299	84
27	2019/1/29 8:20	2019/1/29 8:30	302	110
27	2019/1/29 8:30	2019/1/29 8:40	286	48
27	2019/1/29 8:40	2019/1/29 8:50	198	38
27	2019/1/29 8:50	2019/1/29 9:00	151	40

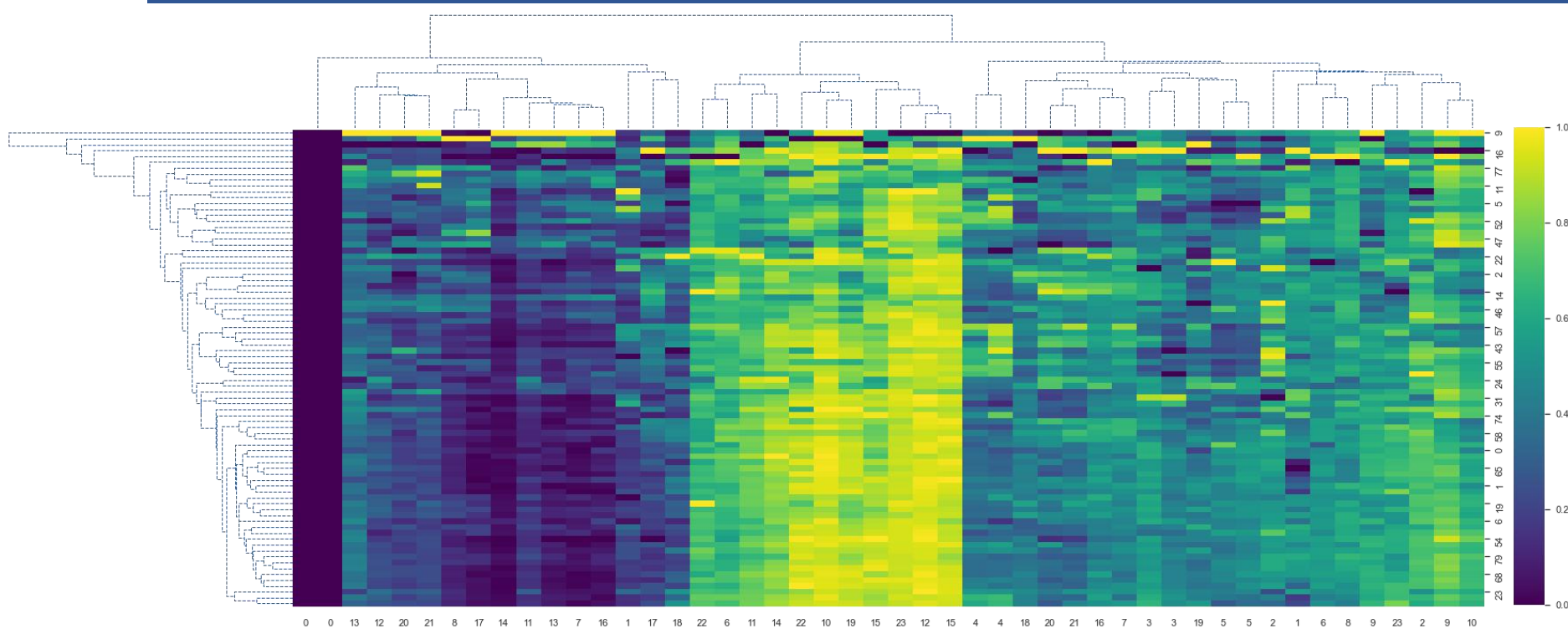
3 项目开发过程及 方法

- 3.5 K-Means聚类算法

地铁站点所处的位置及其周边环境，会影响居民的出行意愿和客流波动。针对常态下的进出站客流数据，利用无监督的K-Means算法识别站点类型。

采用K-Means算法，对具有相似进站和出站客流特征的站点进行聚类。绘制地铁站点聚类热图。

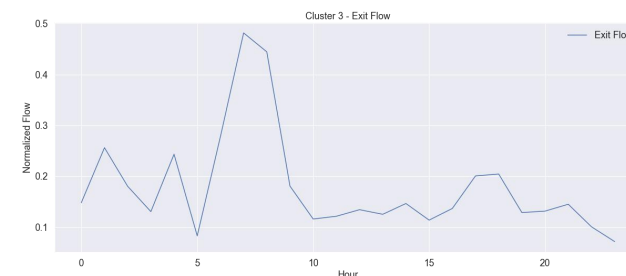
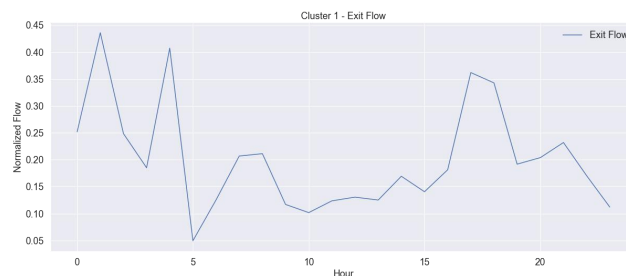
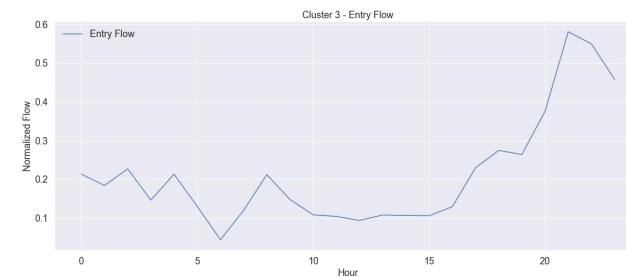
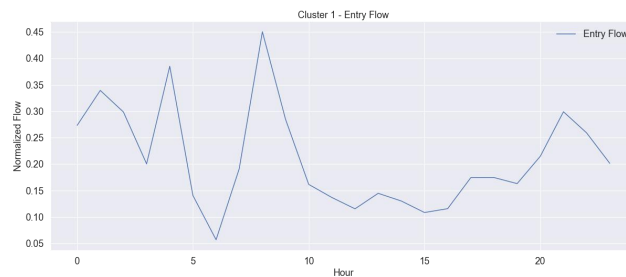
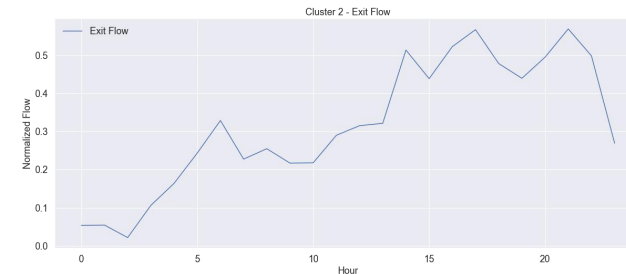
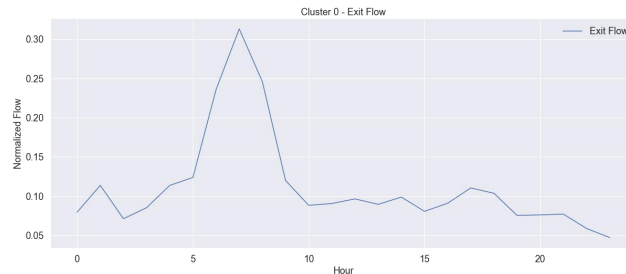
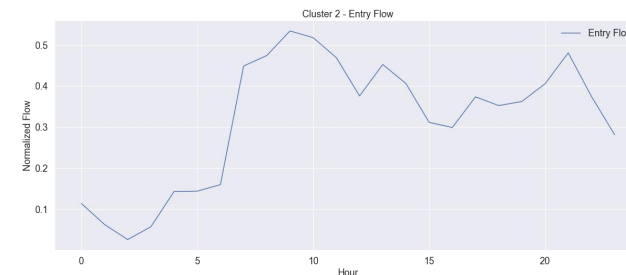
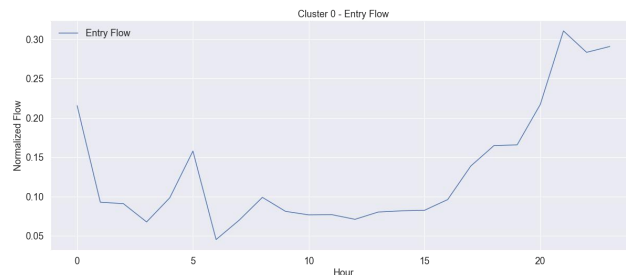
杭州市地铁站点K-Means聚类热图



3 项目开发过程及 方法

- 3.6 聚类结果

聚类划分站点类型，分别为：工作型、娱乐型、峰期型、其他型。



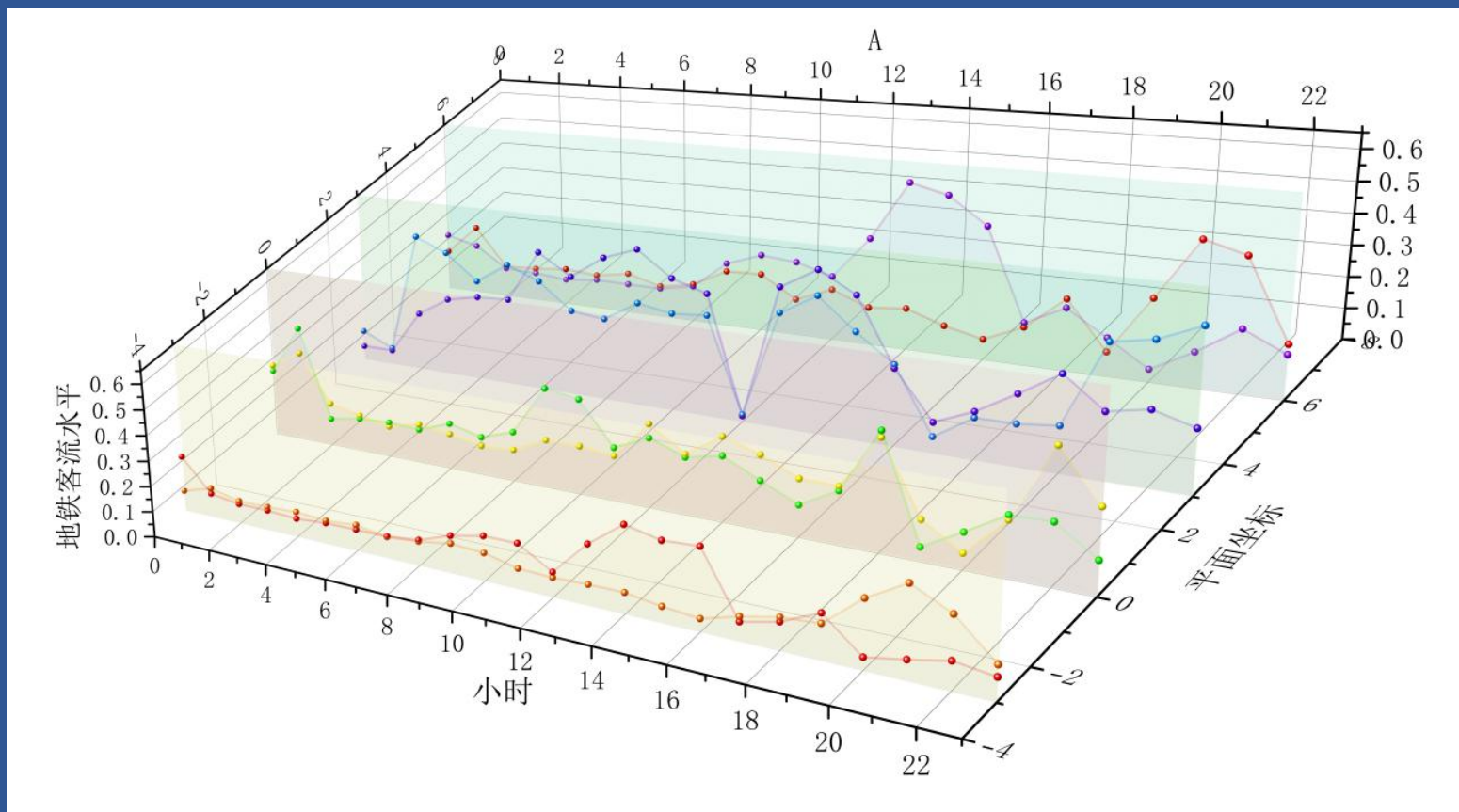
3 项目开发过程及 方法

- 3.7 聚类结果可视化

聚类划分站点类型，分别为：工作型、娱乐型、峰期型、其他型。

对具有相似进站和出站客流特征的站点进行聚类。绘制地铁站点聚类结果3D瀑布图。

杭州市地铁站点K-Means聚类3D瀑布图



3 项目开发过程及 方法

- 3.8 格兰杰因果检验

格兰杰因果检验是一种统计方法，其基本原理是：当一个变量（自变量）发生变化时，另一个变量（因变量）也会发生变化。它可以用来检验某一事件与另一事件是否存在因果关系，从而推断出某一行为或某一事件是否会导致另一事件的发生。它可以更好地控制和排除其他可能影响因变量的因素，从而更准确地推断出自变量是否会引起因变量的变化。

进行城市降雨与地铁OD客流水平的双向因果检验。

其中，原假设H1a：“城市降雨”不是“地铁OD客流水平”的格兰杰原因；原假设H1b：“地铁OD客流水平”不是“城市降雨”的格兰杰原因。

城市降雨与地铁OD客流水平的双向格兰杰因果检验结果

F检验			卡方检验		因果图
原假设	F(1,21)	Prob > F	chi2(1)	Prob > chi2	
H1a	7.65	0.0116*	8.75	0.0031**	
H1b	0.22	0.6464	0.25	0.6187	

3 项目开发过程及 方法

- 3.9 双重机器学习算法

由于传统的双重差分法需要严格的假设，实际数据很难满足，为此，引入双重机器学习模型，进一步探究城市降雨对于地铁OD客流水平的影响机制。我们将机制变量分为三个角度，即降低居民游玩意愿、改变工作出行方式和降低峰期出行意愿，分别使用双重机器学习检验影响的显著性，从而验证H2-H4假设。

四类地铁站点客流水平的因果显著性检验结果

类别	变量	原假设	Coefficient	z	P> z	[95% conf . interval]	
0	Cwm	H3	0.364	5.47	0.000***	0.234	0.495
1	Ptw	H4	0.281	5.23	0.000***	0.176	0.386
2	Rwp	H2	0.035	1.28	0.200	-0.019	0.089
3	-	-	0.328	9.23	0.000***	0.259	0.398

结果显示，在四类地铁站点的三类机制变量中，城市降雨通过降低居民游玩意愿所引起影响不明显，而主要的影响路径是改变工作出行方式和改变峰期出行意愿。这说明，H2假设没有得到有效验证，而H3、H4假设验证成立。

3

项目开发过程及方法

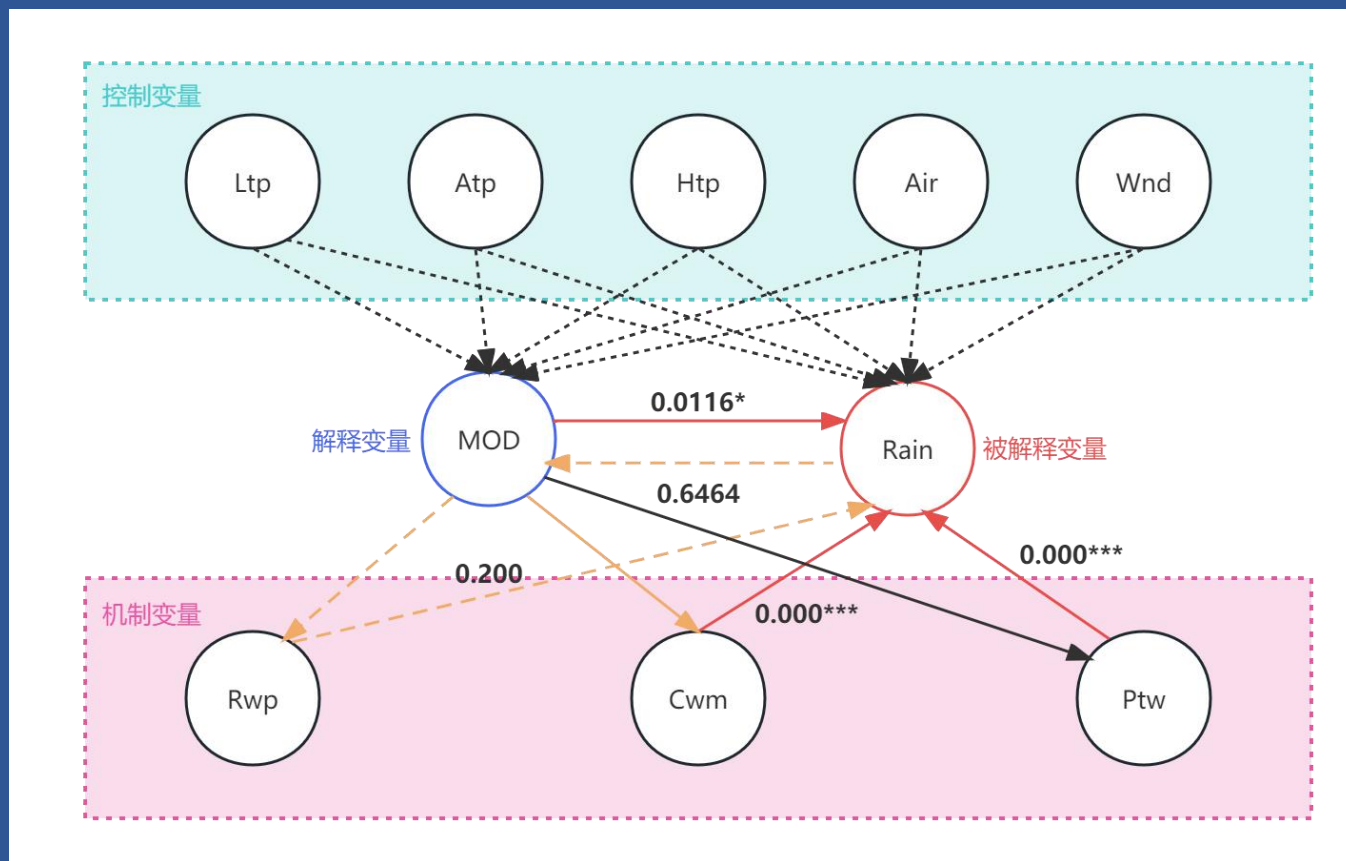
- 3.10 因果图结论

我们可以得到如下推论：工作出行方式和高峰期出行意愿，更容易受到降雨因素带来的抑制作用，在降雨的影响下，乘坐私人交通工具等出行方式可能更受青睐，乘客更倾向于在阴雨天气选择错峰出行。

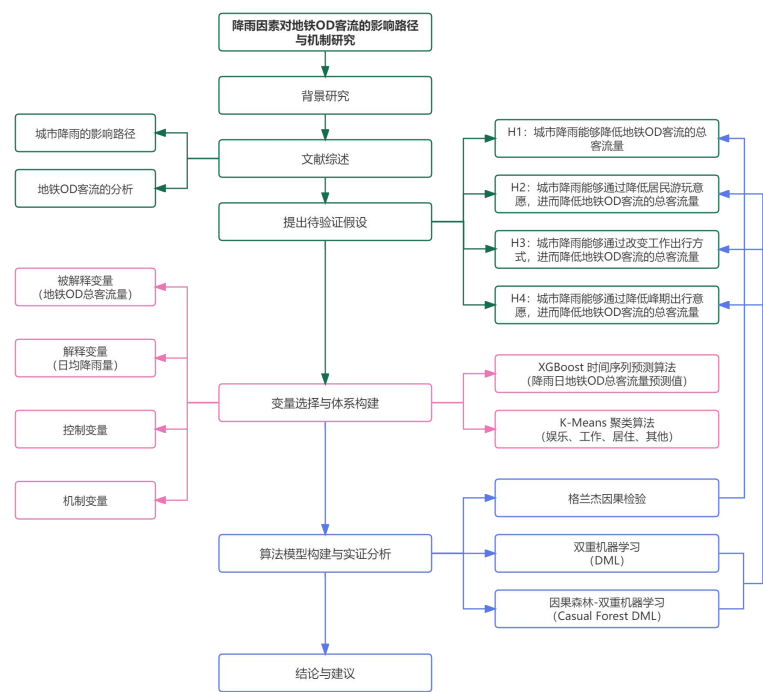
我们认为，大型商超和娱乐型场所往往呈现聚集特征，其体量较为庞大，客流量水平较为稳健，因此，仅考虑降雨因素，难以对娱乐型客流水平产生显著影响。

最终，得到的因果图如图所示：

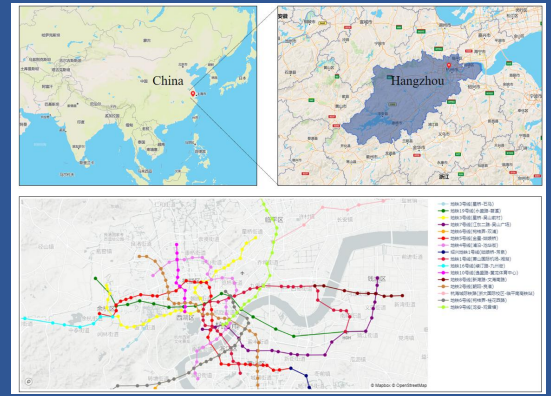
城市降雨与地铁OD客流水平间因果关系图



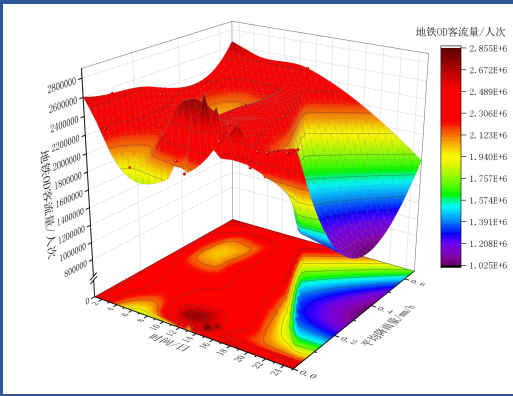
4 项目成果展示



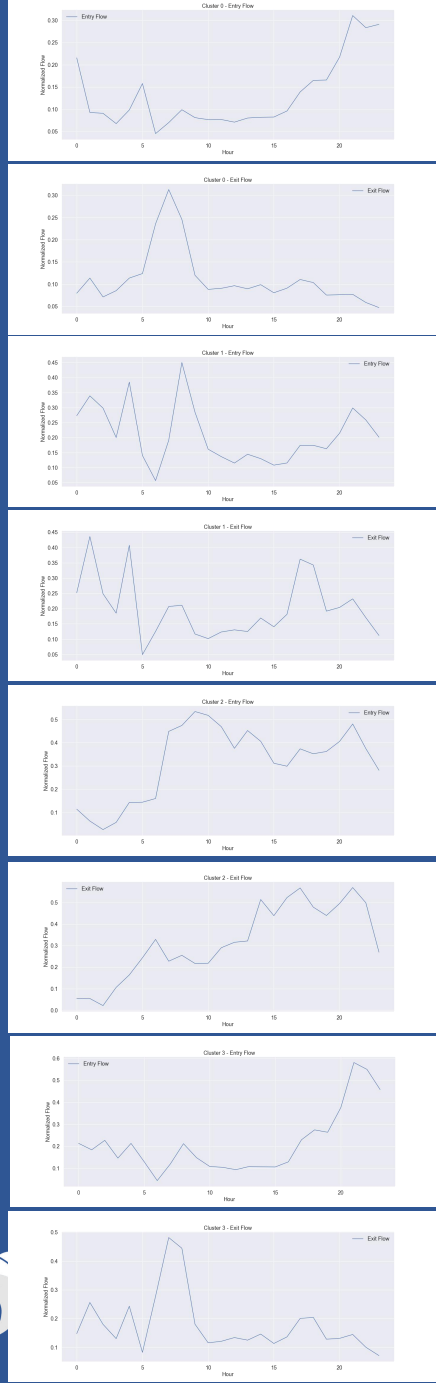
a.



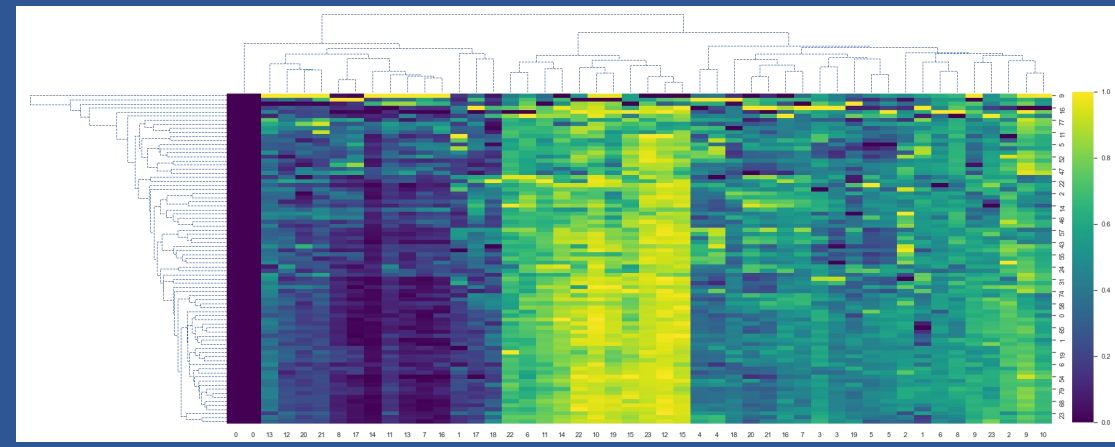
b.



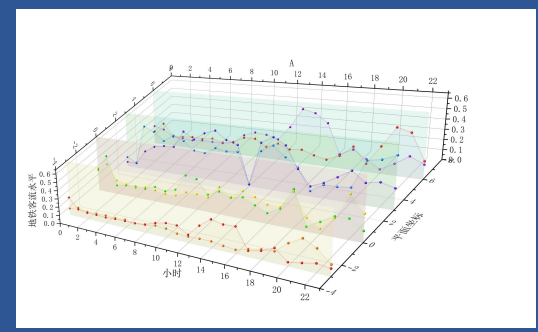
c.



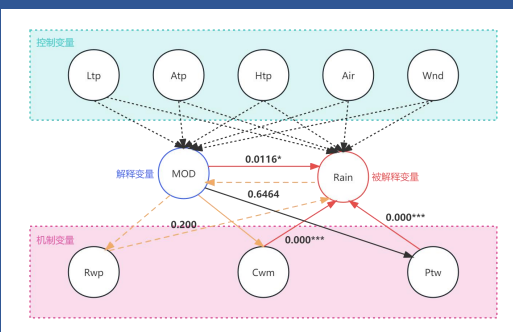
d.



e.



f.



5

项目总结与感悟

- 5.1 创新点

我们结合轨道交通线网矩阵，对地铁OD级别的客流进行了预测。

现有的研究大多针对于进出站客流量、日客流量、小时客流量及断面客流量的预测，其中OD预测的研究内容相对较少。而OD客流体现着进出站客流的流向和分布，是乘客出行需求的直观体现，是网络化运营条件下地铁行车组织、客运组织和乘客服务的基础支撑数据，蕴含着非常有价值的信息。

我们使用近年提出的双重机器学习算法研究降雨因素与地铁客流的影响机制，研究方法新颖。

在因果推断中，传统的PSM-DID等差分模型需要很强的假设。这些模型假设除了处理因素外，其他所有可能影响结果的因素都是恒定的，或者它们随时间的变化是恒定的，这在现实中很难实现。有的模型需要平行趋势假设，即处理组和控制组的趋势在没有处理之前是相同的。在许多情况下，这些假设可能会被违反，从而导致偏误的估计结果。这种方法的应用，突破了传统的因果推断方法对数据要求较为严格的局限性，为研究城市降雨量与地铁OD客流水平之间的因果推断提供了创新方案。

5

项目总结与感悟

- 5.2 其他努力与不足

我们还尝试向上海市公共数据开放中心提出了上海地铁客流数据使用申请。

遗憾的是，截止2023/07/17，还没有收到审核回应。

上海市公共数据开放平台

欢迎您, ChestnutSilver2020

帮助中心退出登录

基本信息

数据使用

使用申请

我的消息

收藏订阅

应用管理

互动管理

我的沙箱

普惠数据

数据使用 / 使用申请

数据使用

申请编号	数据名称	资源类型	数据提供单位	申请时间	审核时间	状态	操作
TN004444320230712090216454543	全市地铁进出客流量	数据接口	上海市交通委员会	2023-07-12 09:02:16		审核中	查看

相信地铁客流数据蕴含的价值远不止于此。还可以进一步对从接驳行程、POI密度分析、地铁线路异质性分析等角度出发，也许还会有不一样的收获。

THANKS!

