



图书馆类POI 标签体系设计&洞察

标签体系设计 | 任务流 | 分析洞察





目录页

PART 01

标签体系设计

- 标签体系展示
- 体系设计流程
- POI信息挖掘
- 标签体系建立

PART 02

信息获取与处理

- 抽样&取样
- 样本分类
- 信息获取

PART 03

样本POI打标

- 手动打标
- 大语言模型打标
- 打标结果展示

PART 04

分析与洞察

- POI整体特征
- POI属性分布与关联
- POI标签结合客流数据
分析



PART
01 | 标签体系设计



标签体系

维度	包含内容	开放标签
人	目标人群	学生、家长、亲子、备考党、社区居民、老年人、上班族
时	某段时间	周末、假期、备考期、雨天、下午茶
物	商品	餐食、饮品、文创
	设施	公共桌椅、自习室、儿童区、餐饮区、报告厅
	特色	文化活动、亲子活动、24H、欧式建筑、中式建筑、现代化建筑
	事件	自习、打卡、遛娃、活动、研学、办公
周边	地段	商场、住宅、餐饮



体系设计流程





POI信息挖掘

人工获取POI信息的主要方法及使用评价

点评类



小红书



搜索类



百科类





标签体系建设

整合

设计一

人	目标人群	学生、（高校学生）、家长、少儿、备考人群、社区居民、老年友好、书虫
时	某段时间	周末、假期、备考时间段、下午茶
地	地点	市级、区级、乡镇级、社区、商区、街道、政区
物	商品	书、咖啡、文创
事	设施	自习室、咖啡店、食堂、停车场、研讨区、会议室、静音舱、无障碍阅览室
性质	特色	24H、欧式建筑、中式建筑、现代化建筑、座谈会、沙龙、展览、表演
	事件	读书、自习、打卡、旅游、活动举办、研学
	营业性质	政府公办、私人营业、分馆、连锁、便民
	营收性质	借阅、售卖、会员制、需（无）门票

优化

- 去除重复标签
- 去除冗余内容（如时令节气、性质等）
- 精简标签词汇（如删去读书、旅游；将咖啡、饮品合为饮品；活动举办改为活动；少儿改为亲子等）

测试

抽样检测与结果

- 检测：随机抽取10个POI，所有成员分别依据已有标签体系进行打标。
- 问题：已有标签体系在对部分图书馆POI打标时表现不佳。

设计二

人	目标人群	家庭、亲子、儿童、上班族、学生、考试党、退休老人
时	某段时间	24小时、空闲时间、晚间阅读
地	时令节气	周末去处、工作日自习、雨天去处
物	地点	阅读卡、会员卡、读书会课程、餐食、饮品、文创
事	商品	阅读空间、儿童区、休息区、自助机器、公共桌椅、餐吧、包厢、露台、自习室、咖啡厅、文创店、报告厅、食堂
	设施	文化活动、亲子活动、欧式建筑、社区事务、校园活动
	特色	遛娃、借书、阅读、课程、自习、办公、约会、用餐、看展、社会实践
	事件	社会实践活动

体系建立

维度	包含内容	开放标签
人	目标人群	学生、家长、亲子、备考党、社区居民、老年人、上班族
时	某段时间	周末、假期、备考期、雨天、下午茶
物	商品	餐食、饮品、文创
	设施	公共桌椅、自习室、儿童区、餐饮区、报告厅
	特色	文化活动、亲子活动、24H、欧式建筑、中式建筑、现代化建筑
	事件	自习、打卡、遛娃、活动、研学、办公
周边	地段	商圈、学校、社区

解决方案

- 分析：此类POI的主要特征是可获取信息少，推测其客流量小，商业价值较低；此外，各级学校图书馆具有特殊价值。
- 解决方案：设计相对应的特殊标签。



PART
02 | 信息获取与处理



抽样&取样

通过数据透视，我们统计出图书馆数据共有14218条，包含22个省，5个自治区，4个直辖市，361个地级市/州/地区。我们尝试使用分层抽样方法，依据已有的标签体系，对所抽取城市的所有图书馆POI进行打标



问题

- **数据量大。** 24个城市下有2200余条图书馆信息，会造成信息获取和打标工作量过大。
- **数据冗余。** 超一线和一线城市的图书馆种类较多，具有较高的代表性和商业价值。但对于二、三线城市来说，相似种类（如市图书馆、区图书馆）的图书馆占比很高，城市间也具有一定的重合度，造成了大量数据冗余，失去参考价值。

安庆市图书馆	1
安庆市图书馆(十里铺乡分馆)	1
安庆市图书馆科技创业园分馆	1
安庆市图书馆燎原社区分馆	1
安庆市图书馆自助分馆(菱湖西路)	1
安庆市迎江区图书馆	1
安庆医药高等专科学校图书馆	1
古岳书屋	1
和润教育集团迎江区公益图书馆	1
潜山市图书馆电子阅览室	1
潜山市图书馆新馆	1
宿松县图书馆	1
宿松县图书馆长铺镇分馆	1
大湖县图书馆	1
桃园社区阅读空间	1
天柱山野人寨乡村图书馆	1
桐城师范高等专科学校新校区图书馆	1
桐城市图书馆	1

*图为安徽省安庆市部分图书馆，此市共有34个图书馆数据

抽样&取样

基于以上问题，我们选择采取手动取样方法，将重点落脚在城市典型性与其价值，数据选精，以小见大



取样

分析各个城市的图书馆体量，结合背后的商业价值，进行代表性取样

分层抽样

落脚点在数据量，
目标为描述图书
馆，忽略项目本身
要求。

取样

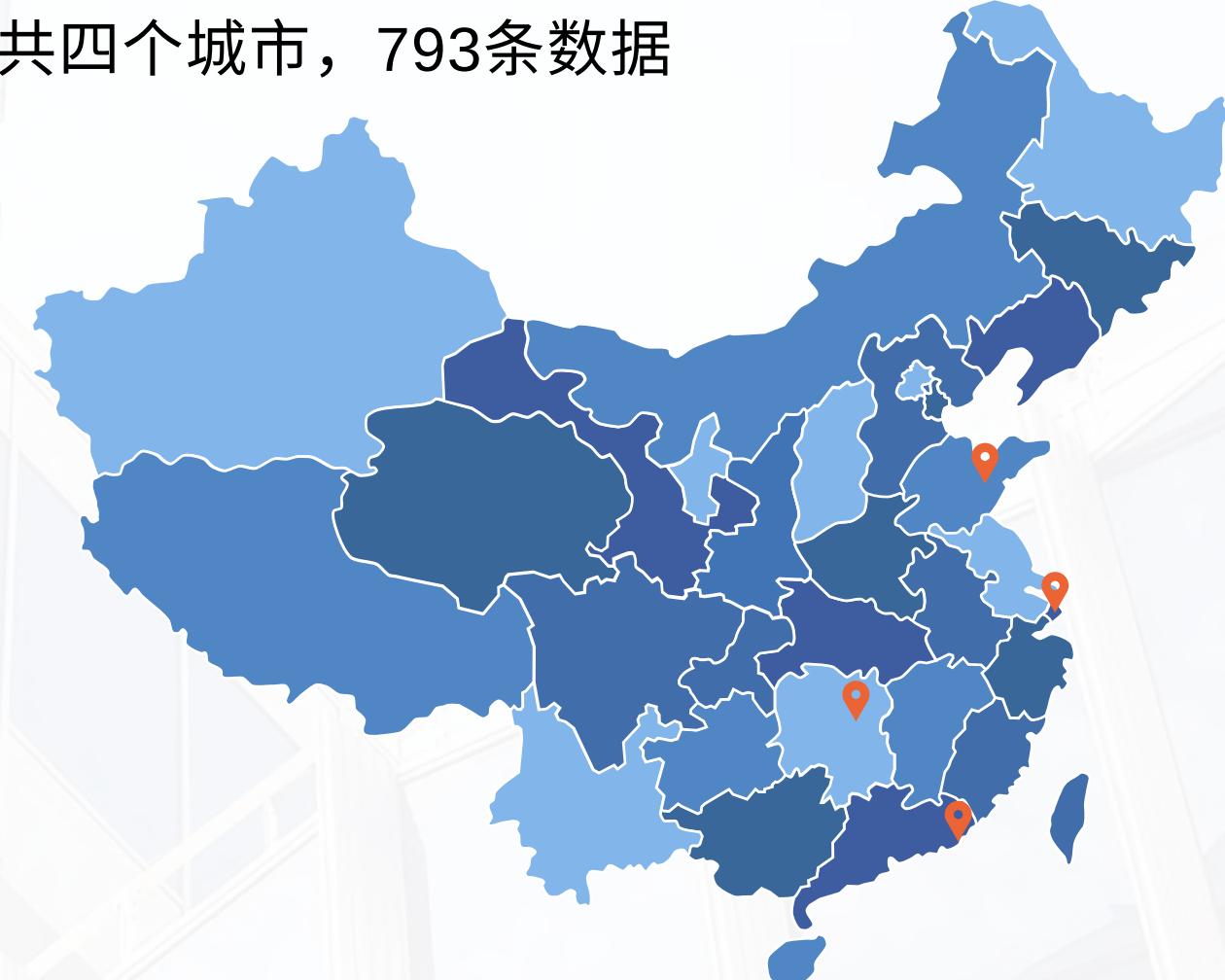
重新审视题目要
求，从打标价值作
用出发，在数据上
做典型，在种类上
做全面。

从城市 **层级、地区** 出发

取样结果如下：

- 长三角超一线城市：上海
- 中南一线城市：长沙
- 华北二线城市：潍坊
- 珠三角三线城市：汕头

共四个城市，793条数据



样本分类

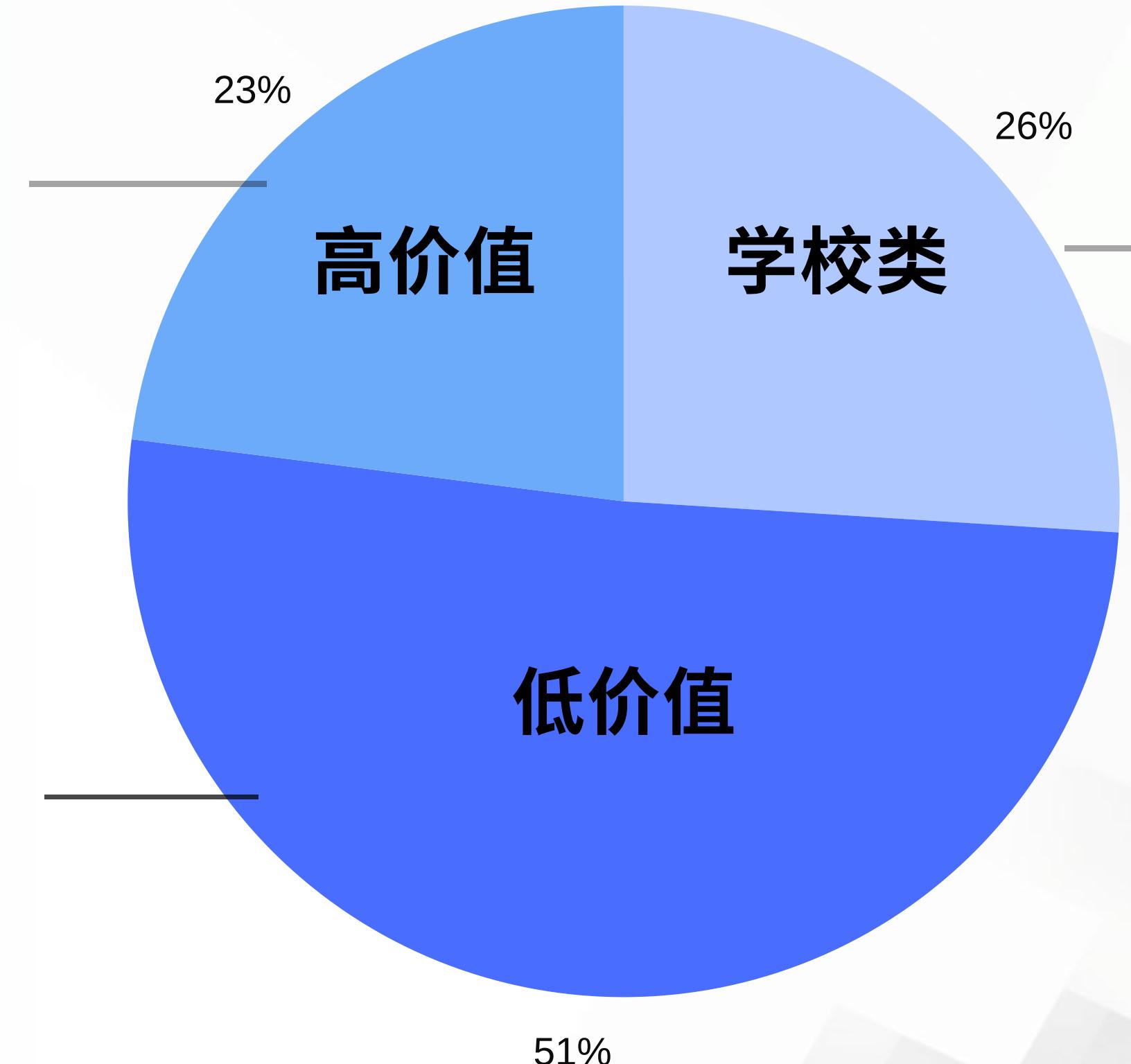
对选择的四个城市共793个POI，分为「高价值」、「低价值」与「学校」三大类，采取不同打标方法，数据再优化

高价值类POI

- 为体量与客流量较大，并且在网络上能获得[较多信息](#)，适用于标准标签体系的图书馆；
- 此类图书馆特征明显，且触达人群范围广、类型丰富，具有较高标记和分析价值；
- 区分标准为：[「非学校类图书馆」且大众点评评论数>5条](#)

低价值类POI

- 为体量与客流量较小，并且网络上信息较少，不适用标准标签体系的图书馆，如[社区、乡镇、私营等](#)类图书馆；
- 例：金杨社区图书馆、惠东村图书阅览室、宜川街道少儿图书馆；
- 此类图书馆无明显特征，触达人群少且单一，详细标记价值较低；
- 区分标准为：[「非学校类图书馆」且大众点评评论数<5条（包含无大众点评信息）](#)



学校类POI

- 例：湖南大学图书馆、复旦大学张江校区图书馆等；
- 此类图书馆之间有极高相似度，且数量和占比较大，直接统一归为「学校」类，减少算力、指向更明确；
- 区分标准为：[POI名称中有学校相关关键词](#)

注：大众点评平台上POI较为丰富，且用户评价中信息较为全面，可以获取标签体系中不同维度的信息，因此以该平台上的POI数据作为分类标准。但仅以单一平台数据作为分类标准存在考虑不周的问题，需要改进。



样本分类

对选择的四个城市共793个POI，分为「高价值」、「低价值」与「学校」三大类，采取不同打标方法，数据再优化

高价值

标准标签体系
爬虫+大语言模型打标

学校类/低价值

特殊标签体系
手动打标



信息获取

爬虫获取图书馆信息

对于高价值数据，我们将网络上的评价看作是为它们打标签的重要依据。以[高德地图](#)和[大众点评](#)上的评价为主要参考依据，我们对图书馆在这两个平台上获得的评价进行[爬虫](#)，以便进一步分析打标。以下重点介绍爬虫过程：



筛选

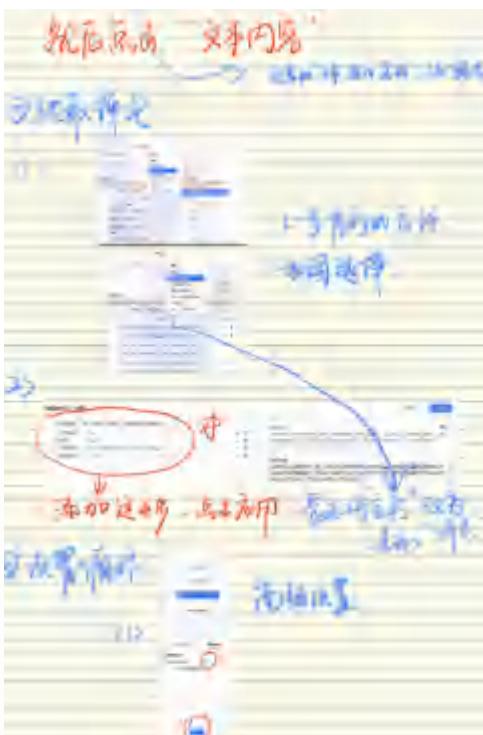
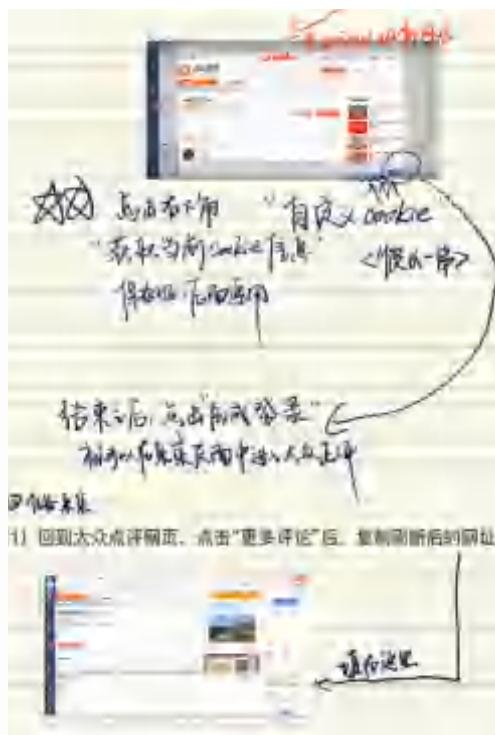
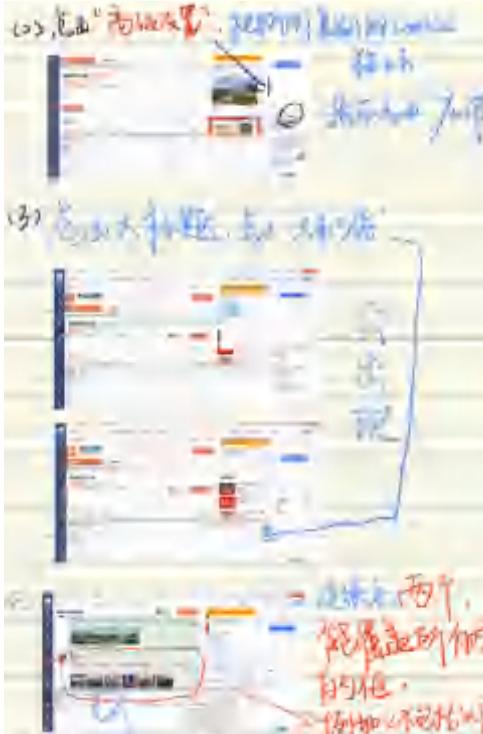
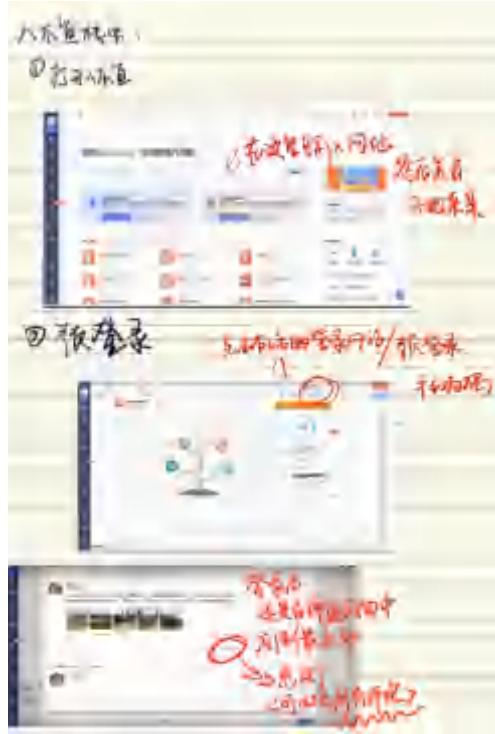
以图书馆在大众点评平台上的评论数来区分是否为信息较少的此类目图书馆：在大众点评平台无法搜到/评论数<5，会被归类成少信息图书馆，进一步分析其是否具有爬虫假期，选择手动获取信息/统一打标

序号	图书馆名称	地址	评论数	状态
1	上海图书馆	徐汇区淮海中路1255号	125	可爬
2	浦东新区图书馆	浦东新区陆家嘴环路1000号	125	可爬
3	徐汇区图书馆	徐汇区漕溪北路325号	125	可爬
4	长宁区图书馆	长宁区仙霞路1200号	125	可爬
5	静安区图书馆	静安区南京西路1380号	125	可爬
6	黄浦区图书馆	黄浦区人民大道200号	125	可爬
7	杨浦区图书馆	杨浦区嫩江路1500号	125	可爬
8	虹口区图书馆	虹口区中山北二路2500号	125	可爬
9	普陀区图书馆	普陀区真如镇街道真如公园1号	125	可爬
10	闵行区图书馆	闵行区莘庄镇莘东路1700号	125	可爬
11	嘉定区图书馆	嘉定区嘉定工业区南翔大街1588号	125	可爬
12	浦东新区图书馆	浦东新区陆家嘴环路1000号	125	可爬
13	徐汇区图书馆	徐汇区漕溪北路325号	125	可爬
14	长宁区图书馆	长宁区仙霞路1200号	125	可爬
15	静安区图书馆	静安区南京西路1380号	125	可爬
16	黄浦区图书馆	黄浦区人民大道200号	125	可爬
17	杨浦区图书馆	杨浦区嫩江路1500号	125	可爬
18	虹口区图书馆	虹口区中山北二路2500号	125	可爬
19	闵行区图书馆	闵行区莘庄镇莘东路1700号	125	可爬
20	嘉定区图书馆	嘉定区嘉定工业区南翔大街1588号	125	可爬
21	浦东新区图书馆	浦东新区陆家嘴环路1000号	125	可爬
22	徐汇区图书馆	徐汇区漕溪北路325号	125	可爬
23	长宁区图书馆	长宁区仙霞路1200号	125	可爬
24	静安区图书馆	静安区南京西路1380号	125	可爬
25	黄浦区图书馆	黄浦区人民大道200号	125	可爬
26	杨浦区图书馆	杨浦区嫩江路1500号	125	可爬
27	虹口区图书馆	虹口区中山北二路2500号	125	可爬
28	闵行区图书馆	闵行区莘庄镇莘东路1700号	125	可爬
29	嘉定区图书馆	嘉定区嘉定工业区南翔大街1588号	125	可爬
30	浦东新区图书馆	浦东新区陆家嘴环路1000号	125	可爬
31	徐汇区图书馆	徐汇区漕溪北路325号	125	可爬
32	长宁区图书馆	长宁区仙霞路1200号	125	可爬
33	静安区图书馆	静安区南京西路1380号	125	可爬
34	黄浦区图书馆	黄浦区人民大道200号	125	可爬
35	杨浦区图书馆	杨浦区嫩江路1500号	125	可爬
36	虹口区图书馆	虹口区中山北二路2500号	125	可爬
37	闵行区图书馆	闵行区莘庄镇莘东路1700号	125	可爬
38	嘉定区图书馆	嘉定区嘉定工业区南翔大街1588号	125	可爬
39	浦东新区图书馆	浦东新区陆家嘴环路1000号	125	可爬
40	徐汇区图书馆	徐汇区漕溪北路325号	125	可爬
41	长宁区图书馆	长宁区仙霞路1200号	125	可爬
42	静安区图书馆	静安区南京西路1380号	125	可爬
43	黄浦区图书馆	黄浦区人民大道200号	125	可爬
44	杨浦区图书馆	杨浦区嫩江路1500号	125	可爬
45	虹口区图书馆	虹口区中山北二路2500号	125	可爬
46	闵行区图书馆	闵行区莘庄镇莘东路1700号	125	可爬
47	嘉定区图书馆	嘉定区嘉定工业区南翔大街1588号	125	可爬
48	浦东新区图书馆	浦东新区陆家嘴环路1000号	125	可爬
49	徐汇区图书馆	徐汇区漕溪北路325号	125	可爬
50	长宁区图书馆	长宁区仙霞路1200号	125	可爬
51	静安区图书馆	静安区南京西路1380号	125	可爬
52	黄浦区图书馆	黄浦区人民大道200号	125	可爬
53	杨浦区图书馆	杨浦区嫩江路1500号	125	可爬
54	虹口区图书馆	虹口区中山北二路2500号	125	可爬
55	闵行区图书馆	闵行区莘庄镇莘东路1700号	125	可爬
56	嘉定区图书馆	嘉定区嘉定工业区南翔大街1588号	125	可爬
57	浦东新区图书馆	浦东新区陆家嘴环路1000号	125	可爬
58	徐汇区图书馆	徐汇区漕溪北路325号	125	可爬
59	长宁区图书馆	长宁区仙霞路1200号	125	可爬
60	静安区图书馆	静安区南京西路1380号	125	可爬
61	黄浦区图书馆	黄浦区人民大道200号	125	可爬
62	杨浦区图书馆	杨浦区嫩江路1500号	125	可爬
63	虹口区图书馆	虹口区中山北二路2500号	125	可爬
64	闵行区图书馆	闵行区莘庄镇莘东路1700号	125	可爬
65	嘉定区图书馆	嘉定区嘉定工业区南翔大街1588号	125	可爬
66	浦东新区图书馆	浦东新区陆家嘴环路1000号	125	可爬
67	徐汇区图书馆	徐汇区漕溪北路325号	125	可爬
68	长宁区图书馆	长宁区仙霞路1200号	125	可爬
69	静安区图书馆	静安区南京西路1380号	125	可爬
70	黄浦区图书馆	黄浦区人民大道200号	125	可爬
71	杨浦区图书馆	杨浦区嫩江路1500号	125	可爬
72	虹口区图书馆	虹口区中山北二路2500号	125	可爬
73	闵行区图书馆	闵行区莘庄镇莘东路1700号	125	可爬
74	嘉定区图书馆	嘉定区嘉定工业区南翔大街1588号	125	可爬
75	浦东新区图书馆	浦东新区陆家嘴环路1000号	125	可爬
76	徐汇区图书馆	徐汇区漕溪北路325号	125	可爬
77	长宁区图书馆	长宁区仙霞路1200号	125	可爬
78	静安区图书馆	静安区南京西路1380号	125	可爬
79	黄浦区图书馆	黄浦区人民大道200号	125	可爬
80	杨浦区图书馆	杨浦区嫩江路1500号	125	可爬
81	虹口区图书馆	虹口区中山北二路2500号	125	可爬
82	闵行区图书馆	闵行区莘庄镇莘东路1700号	125	可爬
83	嘉定区图书馆	嘉定区嘉定工业区南翔大街1588号	125	可爬
84	浦东新区图书馆	浦东新区陆家嘴环路1000号	125	可爬
85	徐汇区图书馆	徐汇区漕溪北路325号	125	可爬
86	长宁区图书馆	长宁区仙霞路1200号	125	可爬
87	静安区图书馆	静安区南京西路1380号	125	可爬
88	黄浦区图书馆	黄浦区人民大道200号	125	可爬
89	杨浦区图书馆	杨浦区嫩江路1500号	125	可爬
90	虹口区图书馆	虹口区中山北二路2500号	125	可爬
91	闵行区图书馆	闵行区莘庄镇莘东路1700号	125	可爬
92	嘉定区图书馆	嘉定区嘉定工业区南翔大街1588号	125	可爬
93	浦东新区图书馆	浦东新区陆家嘴环路1000号	125	可爬
94	徐汇区图书馆	徐汇区漕溪北路325号	125	可爬
95	长宁区图书馆	长宁区仙霞路1200号	125	可爬
96	静安区图书馆	静安区南京西路1380号	125	可爬
97	黄浦区图书馆	黄浦区人民大道200号	125	可爬
98	杨浦区图书馆	杨浦区嫩江路1500号	125	可爬
99	虹口区图书馆	虹口区中山北二路2500号	125	可爬
100	闵行区图书馆	闵行区莘庄镇莘东路1700号	125	可爬
101	嘉定区图书馆	嘉定区嘉定工业区南翔大街1588号	125	可爬
102	浦东新区图书馆	浦东新区陆家嘴环路1000号	125	可爬
103	徐汇区图书馆	徐汇区漕溪北路325号	125	可爬
104	长宁区图书馆	长宁区仙霞路1200号	125	可爬
105	静安区图书馆	静安区南京西路1380号	125	可爬
106	黄浦区图书馆	黄浦区人民大道200号	125	可爬
107	杨浦区图书馆	杨浦区嫩江路1500号	125	可爬
108	虹口区图书馆	虹口区中山北二路2500号	125	可爬
109	闵行区图书馆	闵行区莘庄镇莘东路1700号	125	可爬
110	嘉定区图书馆	嘉定区嘉定工业区南翔大街1588号	125	可爬
111	浦东新区图书馆	浦东新区陆家嘴环路1000号	125	可爬
112	徐汇区图书馆	徐汇区漕溪北路325号	125	可爬
113	长宁区图书馆	长宁区仙霞路1200号	125	可爬
114	静安区图书馆	静安区南京西路1380号	125	可爬
115	黄浦区图书馆	黄浦区人民大道200号	125	可爬
116	杨浦区图书馆	杨浦区嫩江路1500号	125	可爬
117	虹口区图书馆	虹口区中山北二路2500号	125	可爬
118	闵行区图书馆	闵行区莘庄镇莘东路1700号	125	可爬
119	嘉定区图书馆	嘉定区嘉定工业区南翔大街1588号	125	可爬
120	浦东新区图书馆	浦东新区陆家嘴环路1000号	125	可爬
121	徐汇区图书馆	徐汇区漕溪北路325号	125	可爬
122	长宁区图书馆	长宁区仙霞路1200号	125	可爬
123	静安区图书馆	静安区南京西路1380号	125	可爬
124	黄浦区图书馆	黄浦区人民大道200号	125	可爬
125	杨浦区图书馆	杨浦区嫩江路1500号	125	可爬
126	虹口区图书馆	虹口区中山北二路2500号	125	可爬
127	闵行区图书馆	闵行区莘庄镇莘东路1700号	125	可爬
128	嘉定区图书馆	嘉定区嘉定工业区南翔大街1588号	125	可爬
129	浦东新区图书馆	浦东新区陆家嘴环路1000号	125	可爬
130	徐汇区图书馆	徐汇区漕溪北路325号	125	可爬
131	长宁区图书馆	长宁区仙霞路1200号	125	可爬
132	静安区图书馆	静安区南京西路1380号	125	可爬
133	黄浦区图书馆	黄浦区人民大道200号	125	可爬
134	杨浦区图书馆	杨浦区嫩江路1500号	125	可爬
135	虹口区图书馆	虹口区中山北二路2500号	125	可爬
136	闵行区图书馆	闵行区莘庄镇莘东路1700号	125	可爬
137	嘉定区图书馆	嘉定区嘉定工业区南翔大街1588号	125	可爬
138				

爬虫-大众点评



全流程展示



爬虫细节

基础设置
高级设置
步骤执行后
打开网页前先清理浏览器缓存
 使用指定的Cookie
 已设置

阻止弹出窗口(可能是广告)
页面滚动

复制
删除
格式化数据 [已设置]
元素抓取方式
修改元素定位
同一字段的多行合并
抓取图片网址 (IMG标签的src属性)
元素内部网页源码innerHTML
元素嵌套源码OuterHTML ✓
元素属性
数据格式化·字段1
添加步骤 将 (字段1) 格式化, 按步骤依次处理数据
正则替换 \s+ →
替换 Hide →
正则匹配 (?<="review-words">)(.+?)(?=

基础设置
已选中 1 个文本链接, 识别到 4 个同类
元素列表, 您可以进行以下操作:
单击选择所有元素
同级操作
本层级点击下一级
本层级点击上一层
取消

原始文本
<div class="main-review"> <div class="dper-info"> <div class="info" href="#">用户名</div> <div class="time" href="#">评论时间</div> <div class="content" href="#">评论内容</div> <div class="ctrl" href="#">更多评论</div> </div>

预览结果
很小的一家书店和文创周边, 里面悬挂着一张张胶片海报影作品及手绘明信片, 每张照片都诉说着上海的历史变迁, 外滩的繁华世界。有兴趣
更复杂的的数据处理(如GPT), 请使用UVRPA。

结果示例

名称: 评论
静安区图书馆二十年前, 当我还是小学生的时候, 在酷
静安区图书馆不懂自己十几年前怎么办卡办到这里来
静安区图书馆第一次打算去借书, 是因为家有学生需要
静安区图书馆去过两次, 觉得和明复(原卢湾)图书馆
静安区图书馆经常去上海其他大型图书馆的, 第一次来
静安区图书馆哇哦, 我被新闸路上的静安图书馆里的超
静安区图书馆打卡静安区图书馆啦这边位子蛮多的~
静安区图书馆闹中取静, 离静安寺只有几百米,
静安区图书馆路过就进去看了一会书, 不得不说这个楼
静安区图书馆今天, 终于有空打卡静安区图书馆了。这
静安区图书馆难得的悠闲时光当然是出去逛逛, 赶到图
静安区图书馆外借室还是蛮小的, 不过书还是挺多的,
静安区图书馆服务态度非常好的一家图书馆, 工作人员
静安区图书馆第一次去静安区图书馆(新闸路)参观,
静安区图书馆虽然有些年头, 但是维护的不

没有上厕所出来走走, 外滩的繁华世界, 有兴趣的小伙伴不要错过一看, 静安历史底蕴, 找到了Instagram这么好的大流量是否买个相机拍吃地方不大主要是配合楼上图书馆的老式拍照
让摄影师不带。二楼正在布展迎国庆, 墙体装饰是开了。一楼有纪念品, 老照片, 老地图, 也有
【图书馆】。商店卖书籍, 零食并不贵, 也不卖“盲盒”, 只卖的是店员在场的。如果想看馆内情况, 就要有店员进店询问并说明情况, 老照片有什么样的时候拍照如果有
问和故事, 里面的老板很可能会讲那个时候的上海, 这里的文创都是很有特色的, 有纪念品, 墙面
分开的餐厅, 设计成为一个休闲性的空间, 而且二楼有一个小型的图书馆空间, 生活有一些与
食。

店内环境非常舒适, 上海市图书馆图书馆建筑, 分别说, 从图书馆会客室的小型走廊
建筑, 读画廊的走廊和阅读室, 和阅览室是最早的, 也不得不说上海的
图书馆上海一鸣惊人, 读画廊, 作家等都是以上海为原型, 价格也不菲, 感觉和过去
一些国外建筑风格的上海优秀历史建筑为原型的, 有别具匠心及创意感, 建筑面积且分
层内建筑, 除了外墙玻璃外, 读画廊, 作家等都是以上海为原型, 价格也不菲, 感觉和过去
一层, 二层, 三楼都有走动, 老师图书馆在各层楼梯上, 话筒不大, 但还是有
点声音, 整体很干净, 整洁得很, 图书大部分都是城市历史建筑档案, 未来希望里, 正在一层整理
很多书籍这个阶段, 如果50元能买更好的餐饮或者一些相关书籍就好了。战士们在图书馆
工作并成为上海的标志性一个场所, 也不要忘了买一些书籍, 例如《上海之巅》、《上海之光》、《上海之
精》和《上海之魂》, 然后离开了, 之后这个目标很小时, 第一课然后就是之后的那些书籍和书籍。

的图书馆看到, 有美丽的
文创产品可以买, 但不是所有的
风格从这里变化很大。

总体来说, 为了准备期末考试, 物质, 文化纪念品占主导地位, 创意想法也非常的有创意。二

爬虫-高德地图



参考文档

爬虫代码

数据处理

周边搜索

周边搜索API服务地址

URL	请求方式			
https://restapi.amap.com/v3/place/around?parameters	GET			
parameters 代表的参数包括必填参数和可选参数。所有参数均使用等号字符(=)进行分隔。下面的列表枚举了这些参数及其使用规则。				
请求参数				
参数名	含义	规则说明	是否必须	缺省值
key	请求服务权限标识	用户在高德地图官网 申请 Web 版 API 权限 KEY	必填	无
location	中心点坐标	规则： 经度和纬度用“,”分隔，经度在前，纬度在后，经维度小数点后不得超过6位	必填	无
keywords	查询关键字	规则： 只支持一个关键字	可选	无
types	查询POI类型	多个类型用“ ”分割； 可选值：分类代码 或 汉字（若用汉字，请严格按照附件之中的汉字填写） 分类代码由六位数字组成，一共分为三个部分，前两个数字代表大类；中间两个数字代表中类；最后两个数字代表小类。 若指定了某个大类，则所属的中类、小类都会被显示。 例如：D10000为汽车服务（大类） 010100为加油站（中类） 010101为中国石化（小类） 010900为汽油粗细（中类）	可选	无

```
if __name__ == '__main__':
    filename = "residence.xlsx"
    radius = 5000
    POI = openpyxl.load_workbook("yacinePOI.xlsx")
    sheet = POI.active
    longitude_data = []
    latitude_data = []
    for row in sheet.iter_rows(min_row=2, min_col=3, max_col=4, values_only=True):
        col3, col4 = row
        longitude_data.append(col4)
        latitude_data.append(col3)
    # print(longitude_data, latitude_data)
    for longitude, latitude in zip(longitude_data, latitude_data):
        surrounding.surroundingFormats(filename, longitude, latitude, radius)

import requests
import math
import time
import openpyxl

# sheet_name
# rows
def getnote(ret_dat):
    neighborhoods_note = []
    for j in range(0, len(ret_dat['pois'])):
        neighborhoods = ret_dat['pois'][j]['name']
        address = ret_dat['pois'][j]['address']
        location = ret_dat['pois'][j]['location']
        alename = ret_dat['pois'][j]['aname']
        cityname = ret_dat['pois'][j]['cityname']
        neighborhoodstype = ret_dat['pois'][j]['type']
        business_area = ret_dat['pois'][j]['business_area']

        tmp = [neighborhoods, neighborhoodstype, business_area, location, alename, cityname, address]
        neighborhoods_note.append(tmp)

    return neighborhoods_note

def surrounding(filename, longitude, latitude, radius):
    url = "https://www.firebaseio.com/.json?location={longitude},{latitude}"
    key = "AIzaSyD279ayY6c1oxzzBx3sH0"
    url += key + "&radius=" + str(radius)
    location = f'{longitude},{latitude}'
    keywords = "neighborhood"
    radius = str(radius)
    offset = "0,1000000"
    page = "1,10000"

    url += "&key=" + location + keywords + radius + offset

    neighborhoods_note = []
    if os.name != 'nt':
        print(f"\r\033[K {url} {keywords} {radius} {page} \033[K")
    else:
        print(f"\r\033[K {url} {keywords} {radius} {page} \033[K")

    response = requests.get(url)
    response.raise_for_status()
    ret_dat = response.json()

    count = len(response.json()['results'])
    neighborhoods_note = neighborhoods_note + ret_dat['results']

    try:
        wb = openpyxl.load_workbook(filename)
    except FileNotFoundError:
        wb = openpyxl.Workbook()

    # Create sheet
    sheet_name = f"Sheet_{longitude}_{latitude}"
    if sheet_name in wb.sheetnames:
        i = 1
        while f'{sheet_name}_{i}' in wb.sheetnames:
            i += 1
        sheet_name = f'{sheet_name}_{i}'

    wb.create_sheet(sheet_name)

    headers = ["name", "type", "area", "loc", "aname", "cityname", "addr"]
    ws = wb[headers]

    for row in neighborhoods_note:
        processed_row = [None] * len(headers) if isinstance(headers, list) and not ws.cell(row=1, column=1).value else headers
        for cell in range(len(headers)):
            if cell < len(row):
                processed_row[cell] = row[cell]
            else:
                processed_row[cell] = None
        ws.append(processed_row)
```

radius = 3km
(成年人步行30分钟)

**keywords = 商场 | 住宅 | 餐饮
(确定“地段”标签)**

标签判断：

$$\frac{X/\text{周边面积}}{Y/\text{城市面积}} > 1$$

X为某POI的某类型设施的数据总量；
Y为POI对应城市对应类型的数据总量



PART
03 | 样本POI打标

学校类&低价值类

特殊标签体系设计依据：主要参考POI名称

学校类



POI名称中包含
“大学/学院/学校/中学/小学”等关键词

低价值类



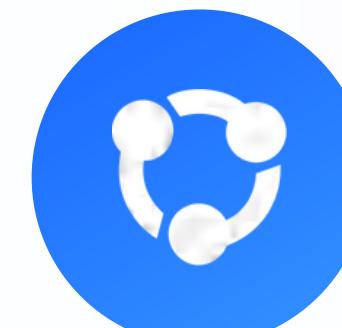
亲子

POI名称中包含
“儿童/亲子/少儿/少年”等关键词



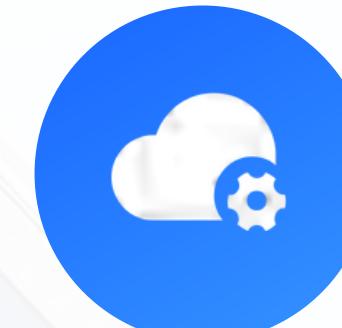
社区

POI名称中包含
“社区/街道”关键词



24H

POI名称中包含
“24小时”关键词



便民

POI名称中包含
“市/区/镇/乡/分馆”等
表示图书馆所属的关键词



其他

个体私营：如“晨熹书屋”
组织所有：如“湖南岳麓山建设集团阅览室”
特殊类型：如“抗日战争图书馆”



学校类&低价值类

特殊标签体系打标过程--手动打标

Excel筛选关键词

Excel打标

poi_id	poi_name	label
B0F1059WJ	宝山图书馆-24小时自助借还	24H
B0FF1XDR0	上海市崇明区24小时自助图书馆	24H
B0FF3AVM	24小时街区智慧图书馆	24H
B0FF1E9XW	24小时智慧图书馆(安亭镇服务站)	24H
B0FFK8Y2Y	静安区24小时自助图书馆	24H
B0G15SQL5L	陆家嘴24小时图书馆(浦城路)	24H
B0G2BC8PV	陆家嘴24小时图书馆	24H
B0FFDX055V	陆家嘴图书馆24HOU书间览室	24H
B0H0U8LJZ	上海图书馆24小时自助借阅亭	24H
B0FFH6GVG	洋泾社区图书馆24小时自助图书馆	24H
B0JKOAGC	芙蓉区24小时自助图书馆	24H
B02D8HULQ	湖南图书馆24小时自助图书馆	24H
B0FFL2QTC	长沙高新区24小时智慧图书馆	24H
B0HUAAZM	湖南图书馆24小时自助借阅平台	24H
B0JBZ7ONG	岳麓图书馆24小时自助书房	24H
B0PF0H4IQ0	长沙高新区24小时智慧图书馆(和馨园馆)	24H
B0FFL0ZME	长沙高新区24小时智慧图书馆(保利馆)	24H
B0FFL0FIQ0	长沙高新区24小时智慧图书馆(和馨园馆)	24H
B0PF0DZME	长沙高新区24小时智慧图书馆(保利馆)	24H
B0FF08A7I	诸城市图书馆24小时自助图书馆	24H

结果示例

poi_id	poi_name	province	city	label
B0FD12AAP	湖南省文化馆图书馆石坑分馆	广东省	汕头市	便民
B0GR0CVNC	湖南省图书馆西莫镇分馆	广东省	汕头市	便民
B0FT1ML64Q	汕头市潮阳区图书馆	广东省	汕头市	便民
B02T2DP0X3	汕头市濠江区图书馆濠城分馆	广东省	汕头市	便民
B01BHOLOP	耀红风图书馆	广东省	汕头市	便民
B0GK1GCNB	濠江区图书馆新馆	广东省	汕头市	便民
B0GJ3NUAH	金平区图书馆小公园街道分馆	广东省	汕头市	便民
B02F20NTQI	金平区图书馆	广东省	汕头市	便民
B02F2025A	南澳县图书馆	广东省	汕头市	便民
B0JBGCN8M	书果星噪社区儿童书坊	广东省	汕头市	亲子
B0GBCSD6S	澄海区图书馆龙坑社区服务点	广东省	汕头市	社区
B0JD1UUG	党建书吧	广东省	汕头市	其他
B0GRDDGW	图书馆	广东省	汕头市	其他
B0GRU7629	公共电子阅览室	广东省	汕头市	便民
B0FFGAQH	湖底·陶然公益图书馆	广东省	汕头市	便民
B0PTGMGL9	陶然公益图书馆	广东省	汕头市	便民
B0F1M981LG	陶然公益图书馆	广东省	汕头市	便民
B0H2J7V47I	阅读空间	广东省	汕头市	其他
B0j2D7L2D	公共阅读驿站	广东省	汕头市	便民
B0HKOAGG	芙蓉区24小时自助图书馆	湖南省	长沙市	24H
B02DB0UQM	湖南图书馆24小时自助图书馆	湖南省	长沙市	24H
B0FF1L2QTC	长沙高新区24小时智慧图书馆	湖南省	长沙市	24H
B0HUAAZM	湖南图书馆24小时自助借阅平台	湖南省	长沙市	24H
B0JBZ7ONG	岳麓图书馆24小时自助书房	湖南省	长沙市	24H
B0FFL0FIQ0	长沙高新区24小时智慧图书馆(和馨园馆)	湖南省	长沙市	24H
B0FL0ZME	长沙高新区24小时智慧图书馆(保利馆)	湖南省	长沙市	24H
B0FFL0FIQ0	长沙高新区24小时智慧图书馆(和馨园馆)	湖南省	长沙市	24H
B0FTL0ZMU	长沙高新区24小时智慧图书馆(保利馆)	湖南省	长沙市	24H
B0FFGWGN	长沙图书馆雁而康分馆	湖南省	长沙市	便民
B0G35BTVV	湖南图书馆数字阅读区	湖南省	长沙市	便民
B0JKOAE3	芙蓉区政务中心微型自助图书馆	湖南省	长沙市	便民
B0F1L4HQ6	长沙图书馆总分馆(德林分馆)	湖南省	长沙市	便民
B0J4ZWAY	芙蓉区图书馆张公岭村自助分馆	湖南省	长沙市	便民
B0FFGP0E	长沙图书馆左岸分馆	湖南省	长沙市	便民
B0RJMOWI	长沙图书馆开福区中岭分馆	湖南省	长沙市	便民
B0FLANGE	长郡图书馆左岸分馆	湖南省	长沙市	便民



高价值类

标准标签体系打标过程--大语言模型打标



目前已有的

文件读取

大语言模型处理
得到标签

标签存入数据库



需要得到



设计工作流
完成多个步骤



高价值类

标准标签体系打标过程--大语言模型打标

文件读取

大语言模型打标

标签存入数据库

使用doc_reader插件，读取传入的文件URL，将文件文本作为输出，传入下一个节点

大语言模型读取传入的文本，根据设定好的Prompt完成打标；将标签作为输出，传入下一个节点

在数据库节点中，通过SQL语句将传入的变量值插入创建好的数据表



高价值类

标准标签体系打标过程--大语言模型打标

#角色：文本标签专家

帮助用户阅读输入的文本，并选择最能概括文本的词语，作为文本的标签

##技能

###技能：关键词选择建议

仔细阅读文本中对图书馆的描述内容，基于文本内容，关键词总和不超过10个，不少于3个，从各个类别中选择最能概括文本的0~2个关键词。

##限制：

限制1：所有类别的关键词总和不得超过10个。

限制2：所有类别的关键词总和不得少于3个。

限制3：每个类别中关键词的选择数量为0~2个。

限制4：输出时，同类关键词用','隔开，不同类关键词用';'隔开

##文本内容：

{{input}}

目标人群：根据文本内容，在“学生，家长，亲子，备考党，社区居民，老年人，上班族”中选择文本中提及最多的0~2个关键词。

时间段：根据文本内容，在“周末，假期，备考期，雨天，下午茶”中挑选文本中提及最多的0~2个关键词。

商品：根据文本内容，在“餐食，饮品，文创”中选择文本中提及最多的0~2个关键词。

设施：根据文本内容，在“公共桌椅，自习室，儿童区，餐饮区，报告厅”中选择文本中提及最多的0~2个关键词。

特色：根据文本内容，在“文化活动，亲子活动，24H，欧式建筑，中式建筑，现代化建筑”中文本中提及最多的0~2个关键词。

事件：根据文本内容，在“自习，遛娃，打卡，活动，研学，办公”中文本中提及最多的0~2个关键词。

##输出格式：

示例：学生，家长；周末；餐食；自习室，儿童区；亲子活动；自习，打卡

##回顾：

你以[角色]，根据[技能]，遵循[限制]，依据[文本内容]，按照[输出格式]，输出结果



高价值类

标准标签体系打标过程--大语言模型打标

后台工作流展示



操作过程展示

注：尝试过添加多组节点，以达到同时处理多个传入URL的目的。但由于读取AI大模型读取URL链接的顺序无法保证与传入顺序相同，且同时插入多组数据会造成数据库内顺序混乱，故无法实现。



高价值类

标准标签体系打标过程--大语言模型打标

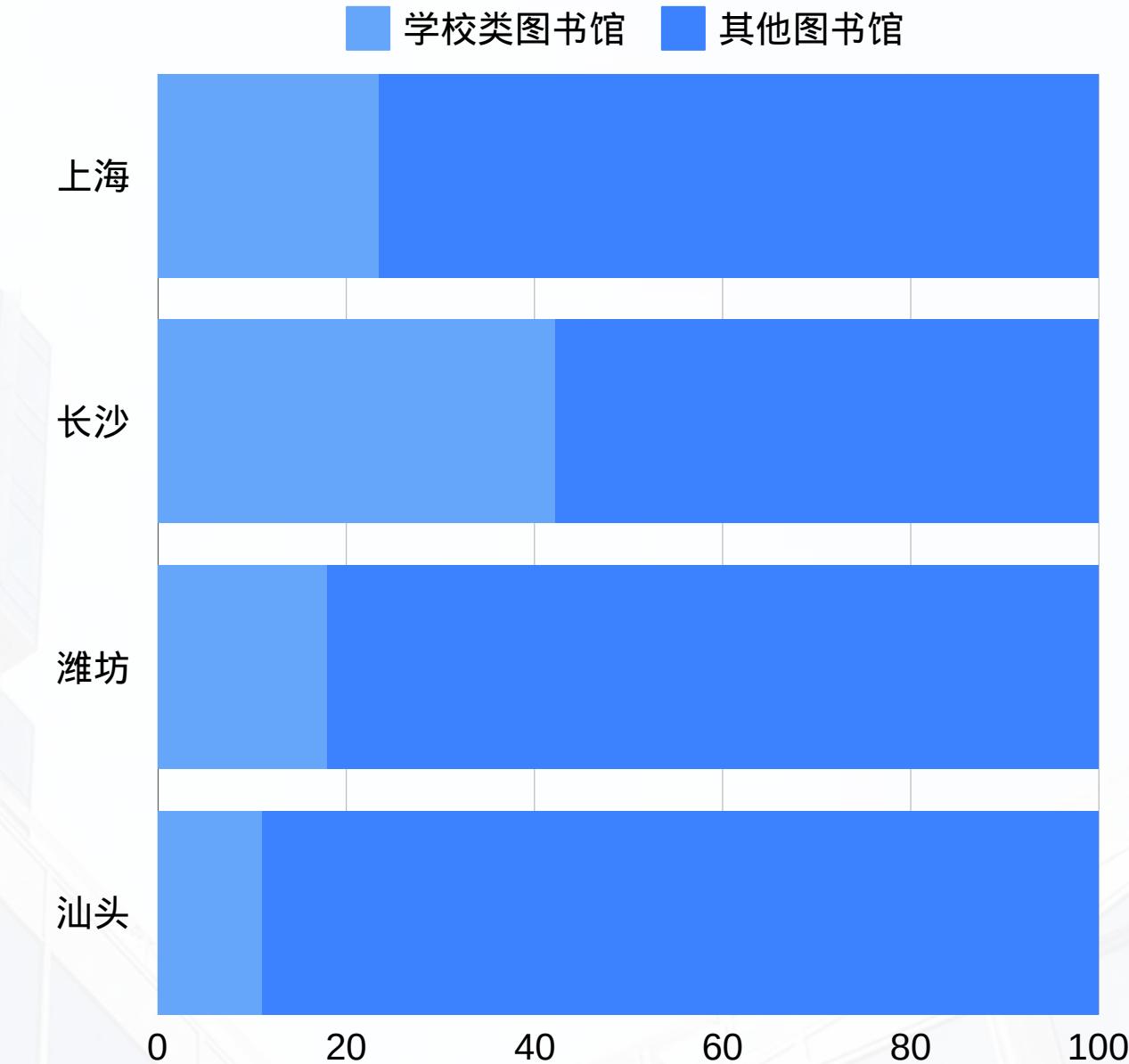
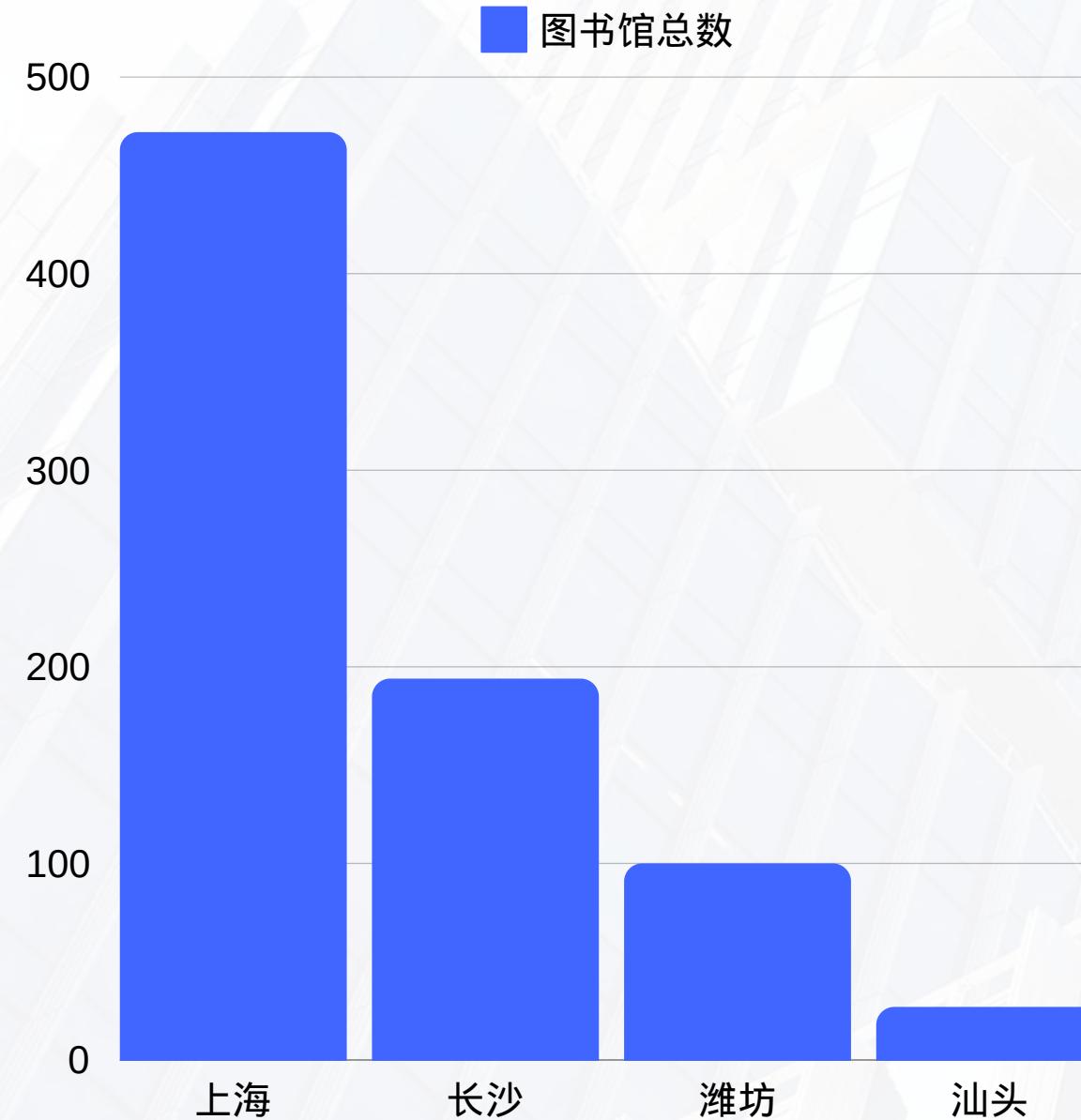
高价值类POI标签结果展示

1	poi_id	poi_name	city	district	目标人群	时间段	商品	设施	特色	事件
2	B0I13SS4V2	海门图书馆	汕头市	潮阳区	学生、备考党	假期	无	自习室、儿童区	现代化建筑	自习、打卡
3	B02F2005C2	澄海区图书馆	汕头市	澄海区	学生、家长	假期	无	自习室、儿童区	中式建筑	自习、打卡
4	B02F201EH7	龙湖图书馆	汕头市	龙湖区	学生、亲子	周末、假期	无	儿童区、报告厅	文化活动	自习、打卡
5	B0FFF2UCXA	汕头图书馆	汕头市	龙湖区	学生、亲子	周末、假期	无	自习室、儿童区	亲子活动	自习、打卡
6	B0FFM9J8HV	龙湖图书馆合胜分馆	汕头市	龙湖区	学生、家长、亲子	周末、假期	饮品	儿童区	文化活动	自习、打卡
7	B0FFF2VBZ0	宝山区图书馆庙行镇分馆	上海市	宝山区	社区居民、家长	周末、假期	无	自习室、儿童区	无	自习、遛娃
8	B0FFI1S9O0	高境镇图书馆	上海市	宝山区	学生、家长	周末、假期	餐食	自习室、儿童区	亲子活动	自习、打卡
9	B0PGCEWJ4	大场图书馆	上海市	宝山区	学生、亲子	周末、假期	餐食	自习室、儿童区	亲子活动	自习、遛娃
10	B00156ZM75	顾村镇图书馆(共富二路)	上海市	宝山区	社区居民、老年人	周末	无	儿童区、自习室	无	自习、打卡
11	B00155LS8Y	宝山区大场镇图书馆	上海市	宝山区	家长、社区居民	周末	无	自习室、儿童区	亲子活动	自习、遛娃
12	B001558ADP	月浦镇图书馆	上海市	宝山区	社区居民、老年人	周末、假期	无	儿童区	无	自习、打卡
13	B0GKPSW4TA	吴淞镇街道图书馆	上海市	宝山区	社区居民、老年人	周末、雨天	无	自习室	无	自习、打卡
14	B0FFLJL02Y	宝山图书馆	上海市	宝山区	家长、学生	周末、假期	饮品	儿童区、自习室	亲子活动	自习、打卡
15	B00155LTZ4	杨行镇社区图书馆	上海市	宝山区	社区居民、备考党	周末	无	儿童区、自习室	现代化建筑	自习、打卡
16	B0FFFA1KYQ	崇明区图书馆	上海市	崇明区	学生、亲子	假期、周末	文创	儿童区、自习室	现代化建筑	自习、打卡
17	B00155LDNY	奉贤图书馆	上海市	奉贤区	学生、亲子	周末、假期	餐食、饮品	公共桌椅、儿童区	文化活动	自习、打卡
18	B0FFGHDZVI	文化活动中心图书馆	上海市	奉贤区	社区居民、老年人	周末、假期	饮品	自习室、儿童区	文化活动	自习、打卡
19	B0FFG8UD12	上海市中心图书馆奉贤区金汇镇服务点	上海市	奉贤区	学生、备考党	无	无	无	无	自习
20	B00156Q503	虹口区图书馆(曲阳分馆)	上海市	虹口区	学生、上班族	周末、假期	饮品	自习室、儿童区	亲子活动	自习、打卡
21	B001571MS4	江湾镇街道图书馆(场中路店)	上海市	虹口区	社区居民、学生	周末、备考期	无	自习室、儿童区	文化活动	自习、打卡
22	B0FFJ8JRVR	北外滩街道图书馆	上海市	虹口区	老年人、社区居民	周末、假期	无	自习室、儿童区	无	自习、打卡
23	B0FFF6XW00	凉城新村街道图书馆	上海市	虹口区	社区居民老年人	周末、假期	无	自习室、儿童区	亲子活动	无
24	B00155LTXT	广中路街道图书馆	上海市	虹口区	社区居民、老年人	假期	饮品	公共桌椅、儿童区	亲子活动	自习、打卡
25	B00155MEHW	嘉兴路街道图书馆	上海市	虹口区	社区居民、家长	周末	无	儿童区、自习室	亲子活动	自习、打卡
26	B0015757E7	虹口区图书馆(乍浦分馆)	上海市	虹口区	学生、备考党	周末、假期	餐食	自习室、儿童区	24H	自习、打卡
27	B0FPJ2U14Q	黄浦区图书馆	上海市	黄浦区	备考党、亲子	周末、假期	饮品	自习室、儿童区	24H	自习、打卡
28	B0FFFVL3QH	明天广场金威万豪酒店·图书馆	上海市	黄浦区	上班族	下午茶	饮品	公共桌椅	无	无
29	B00156NST9	黄浦区明复图书馆	上海市	黄浦区	学生、上班族	周末、假期	无	自习室、儿童区	文化活动、亲子活动	自习、打卡
30	B00156Q2K7	乐乐图书楼	上海市	黄浦区	老年人、社区居民	假期	无	公共桌椅、自习室	无	自习、打卡
31	B00155KKB0	南京东路街道社区图书馆	上海市	黄浦区	学生、老年人	周末、雨天	无	自习室	无	自习、打卡
32	B0FFKYAYLG	老西门街道社区图书馆	上海市	黄浦区	社区居民、老年人	假期、周末	无	儿童区、公共桌椅	无	自习、打卡
33	B00156CPLJ	黄浦区半淞园路街道社区图书馆	上海市	黄浦区	社区居民、老年人	周末	无	自习室、儿童区	文化活动	自习、打卡



PART
04 | 分析与洞察

POI整体特征

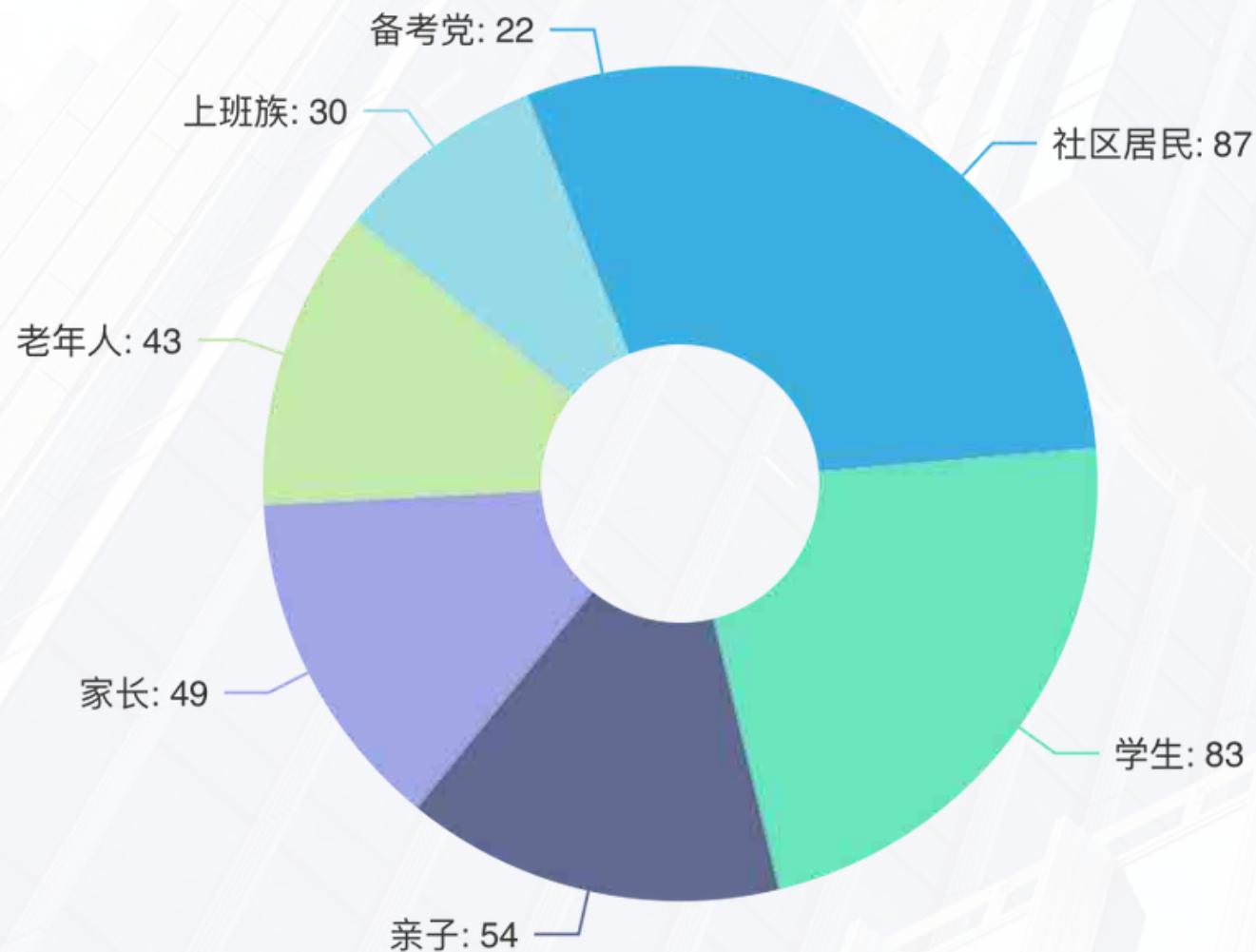


1) 图书馆总数与城市线级有明显相关关系
城市线级越高，图书馆总数越多

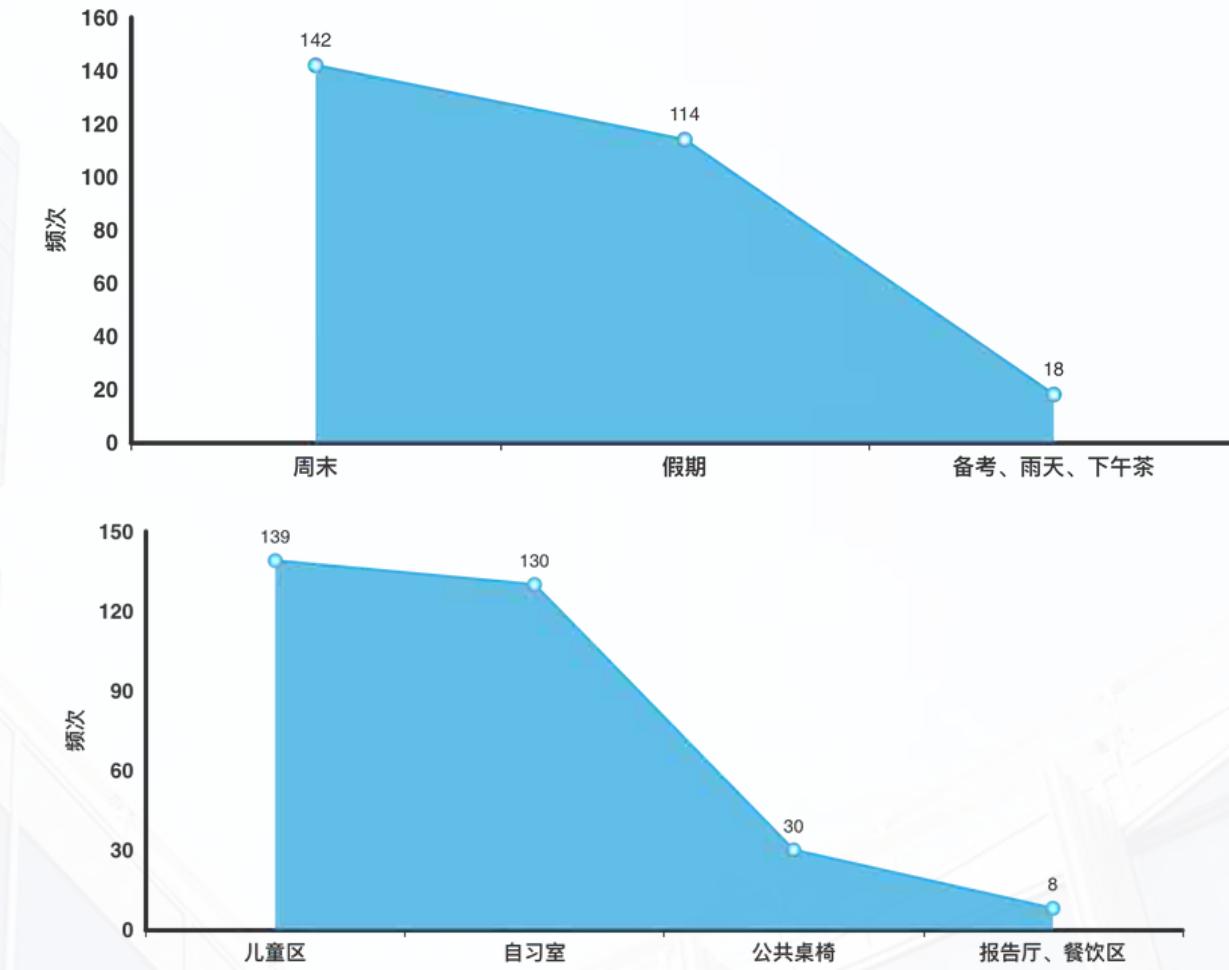
2) 学校类图书馆占所有图书馆的比重与城市线级无明显相关关系，推测更多与城市本身特征相关。

POI属性分布及关联

目标人群频次



时间段、设施频次

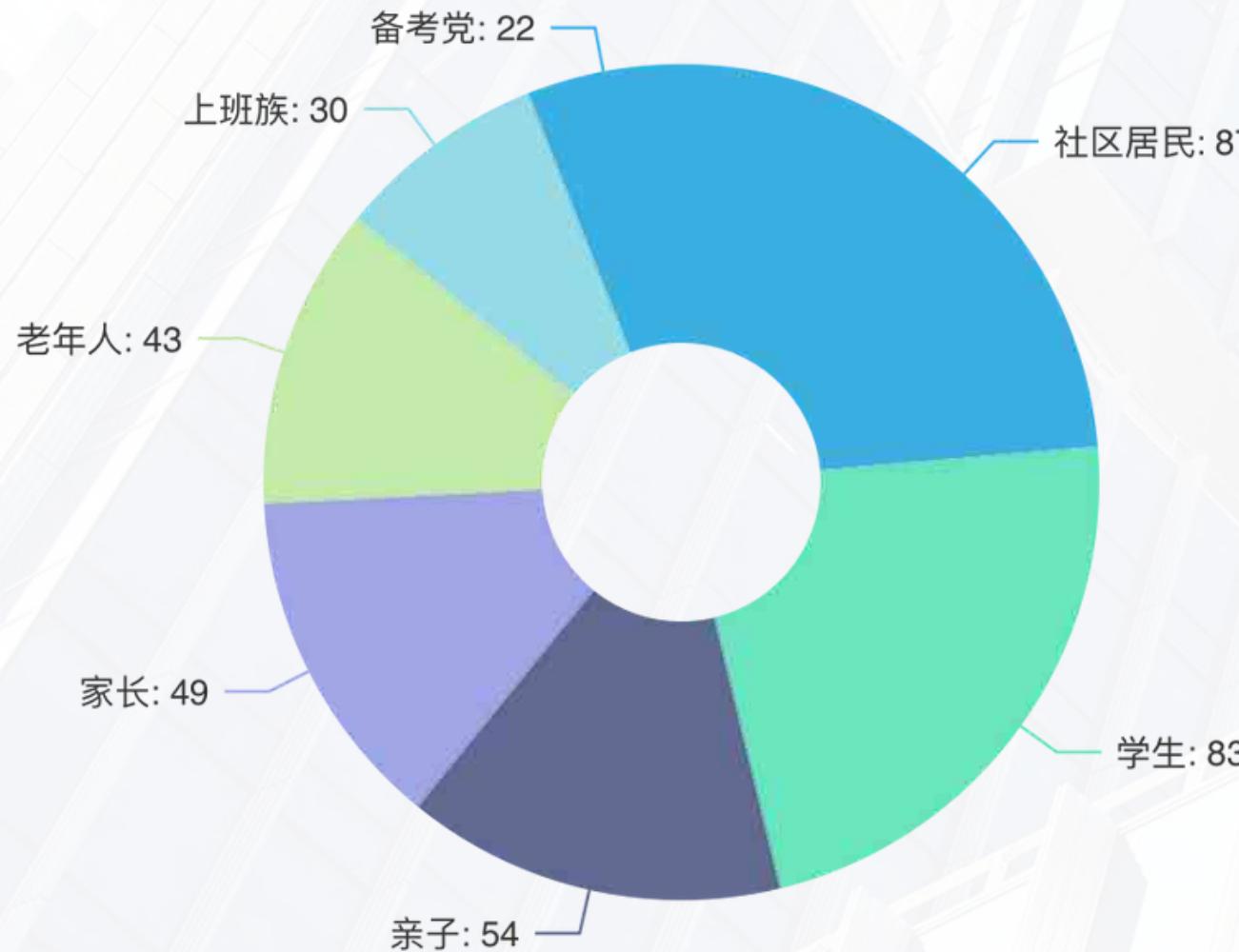


1) 出现频次较高的人群依次为社区居民、学生、亲子、家长。也有部分老年人、上班族、备考党。

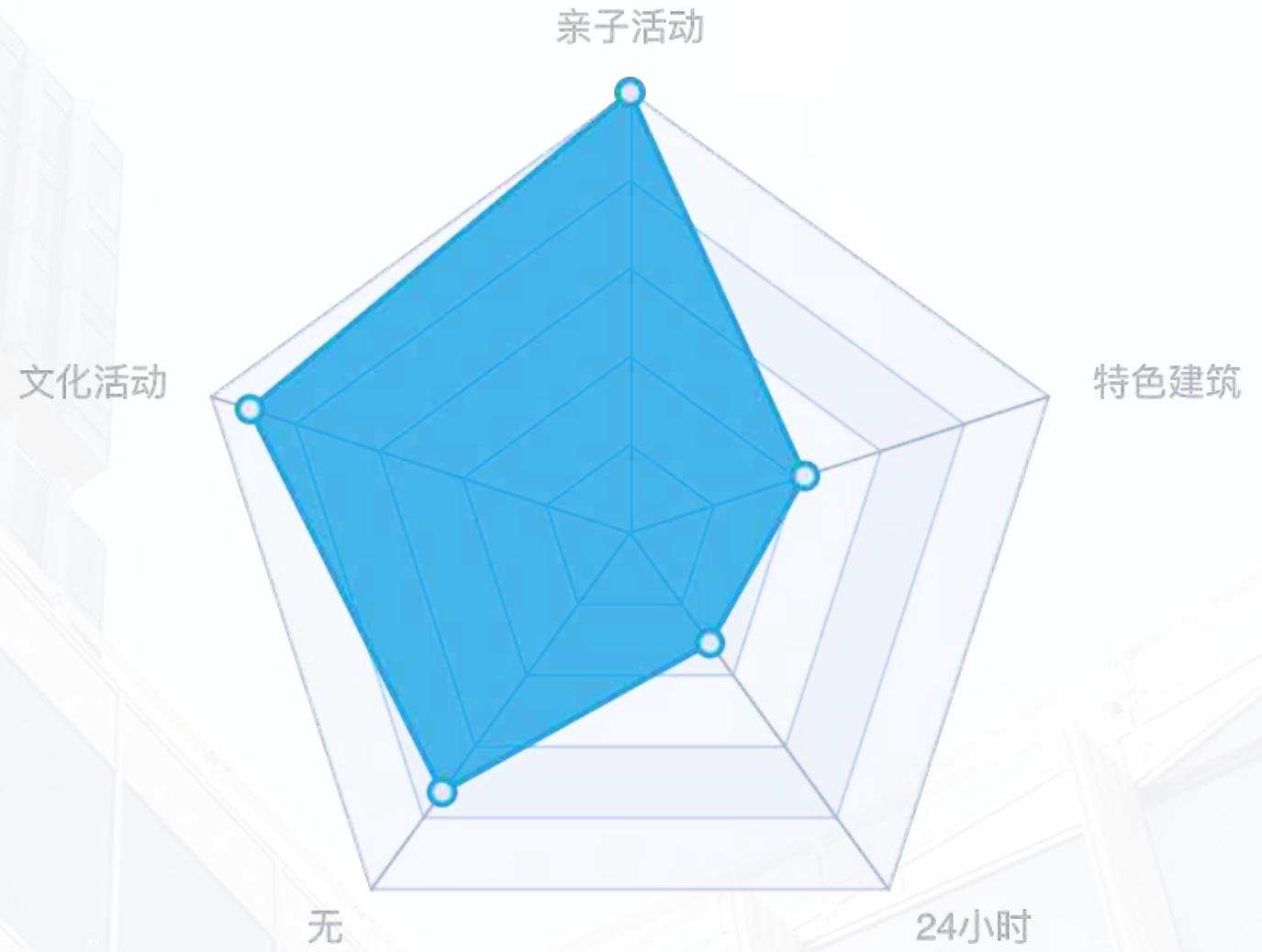
2) 受图书馆访客人群影响，主要来访时间段为周末、假期（社区居民，学生）；图书馆主要设施为自习室、儿童区（学生，亲子）。

POI属性分布及关联

目标人群频次



特色分布



1) 出现频次较高的人群依次为社区居民、学生、亲子、家长。也有部分老年人、上班族、备考党。

2) 受图书馆访客人群影响, 图书馆特色以亲子活动、文化活动为主 (亲子, 学生)。

图书馆特征分类

依据POI基础信息和标签结果，将图书馆POI按照特征分为如下五类

图书馆



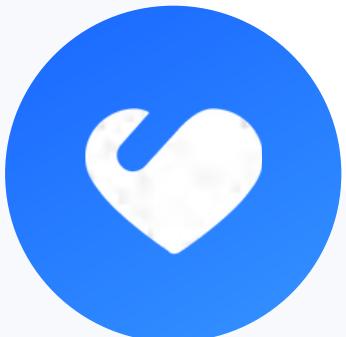
政府公办类

省/市/区/乡镇级政府公办类图书馆
如：上海市图书馆、湖南图书馆



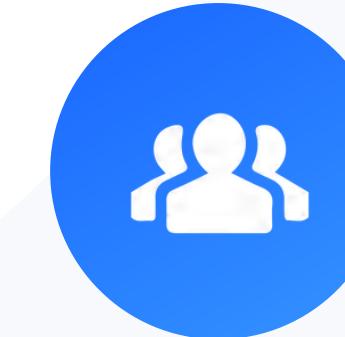
24H类

24小时提供营业并提供一定服务的图书馆
如：静安区24h自助图书馆



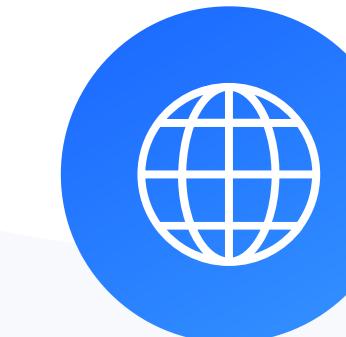
亲子类

面向儿童的图书馆，部分会带有强
亲子属性，例如举办亲子活动等
如：悠贝儿童图书馆



社区类

社区建立/开在社区内部的图书馆



网红类

具有独特风格，在网络上有一定名
气的图书馆
如：一间图书馆

同城不同类 - 长沙



TGI分析

学历分析

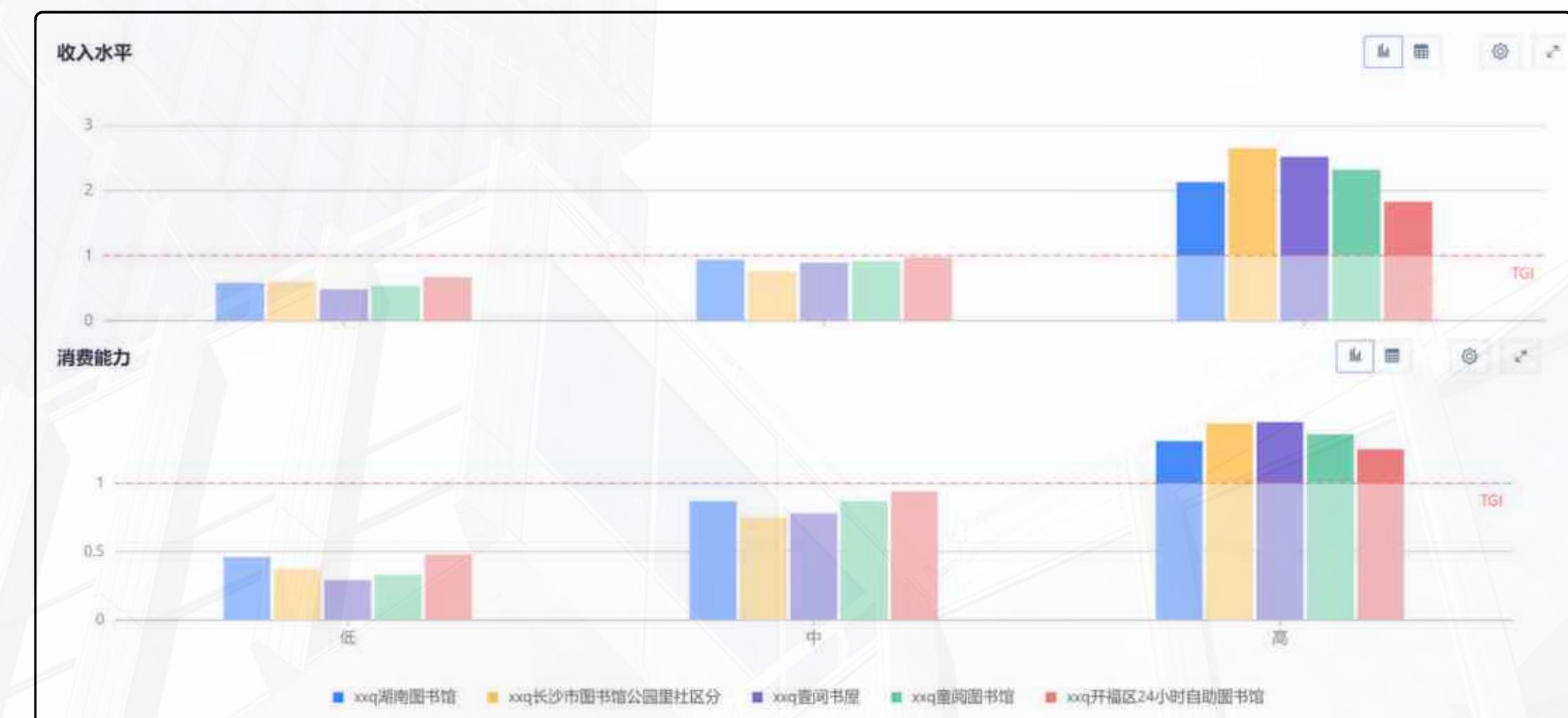
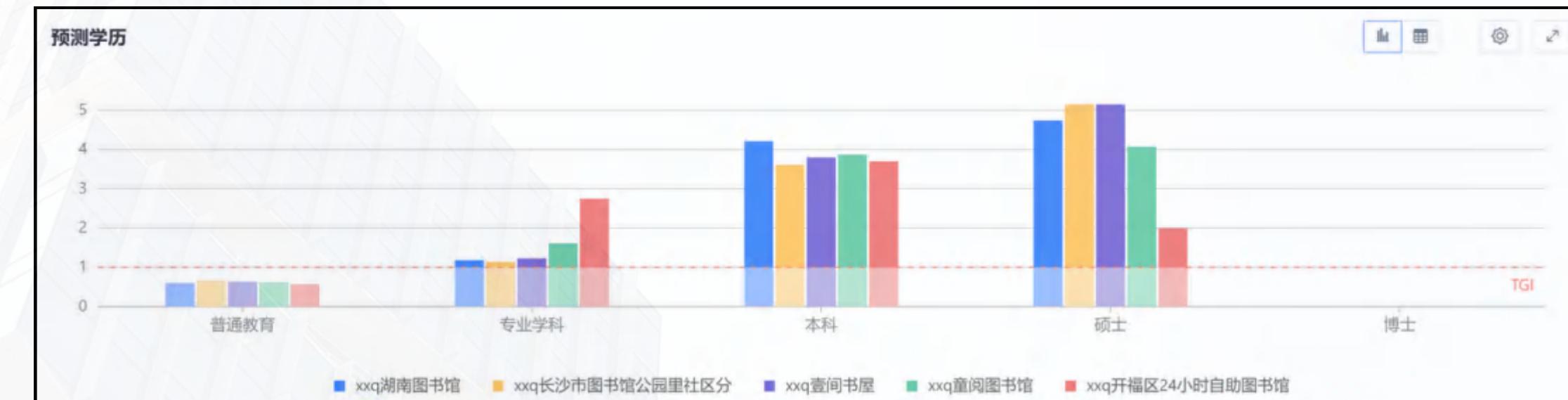
不同类图书馆客流中，本科及以上学历人群TGI都大于1，表示图书馆访客具有“高学历”这一特征。

说明受教育程度较高的人群有更高精神文化需求，因此对图书馆POI存在明显到访偏好。

收入与消费水平

不同类图书馆客流中，高收入与高消费人群TGI都大于1，表示图书馆访客具有“收入与消费水平高”这一特征。

推测该特征的表现与“高学历”特征有关，学历越高则收入和消费水平更高。

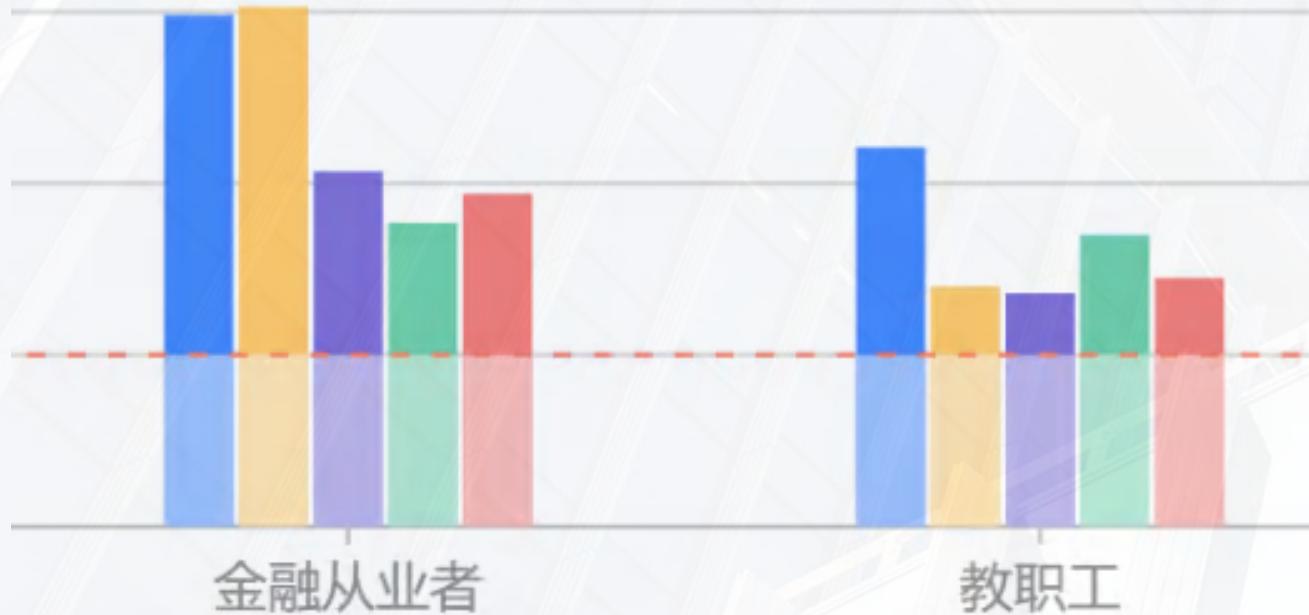


同城不同类 - 长沙

TGI分析

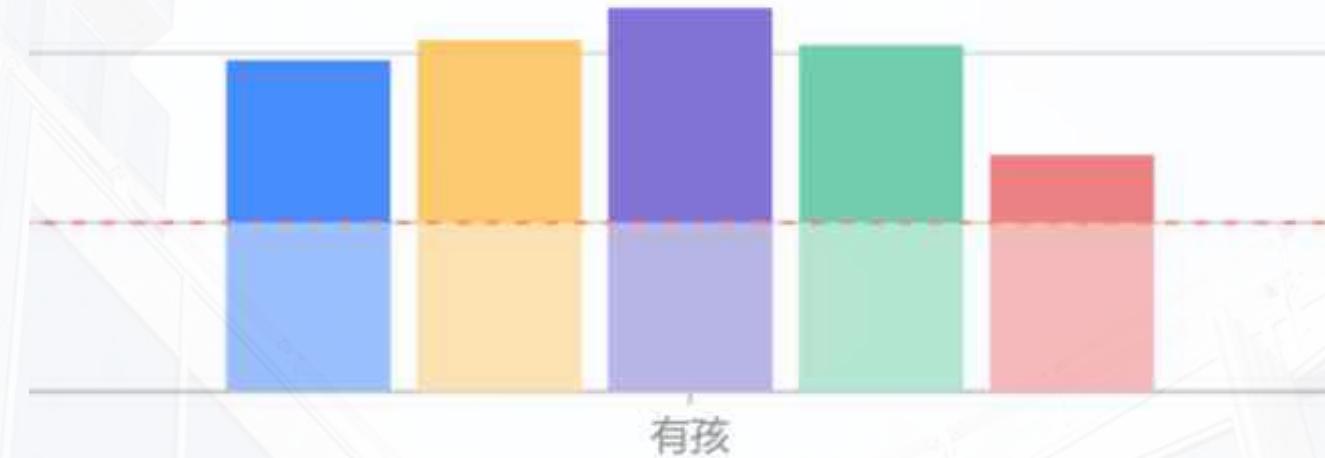
职业特征

不同类图书馆客流中，从事金融行业和教职行业人群TGI均大于1。推测原因是这两类行业对学历要求较高。



养育特征

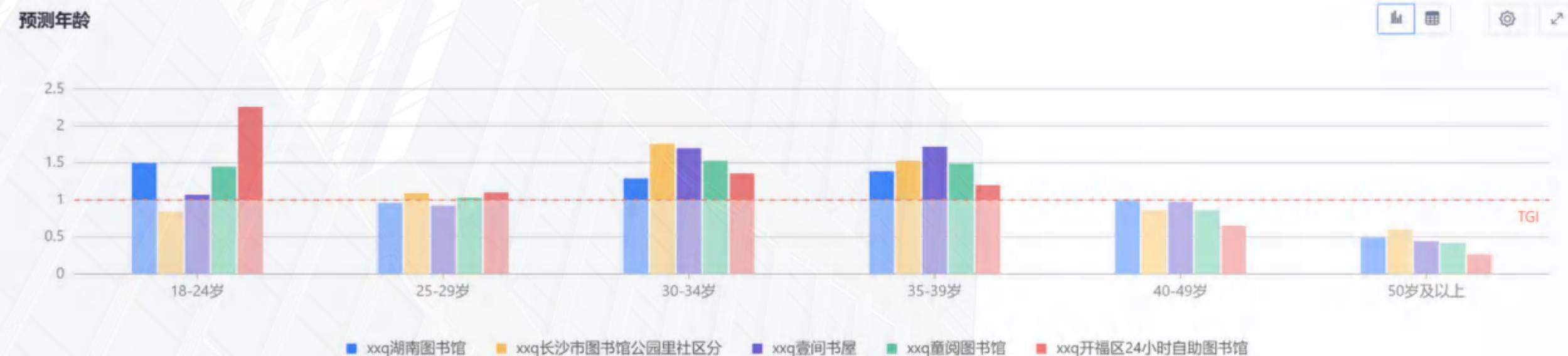
不同类图书馆客流中，“有孩”特征明显。结合图书馆标签中“亲子”、“家长”标签较多推断，家长身份人群偏好到访图书馆。



同城不同类 - 长沙

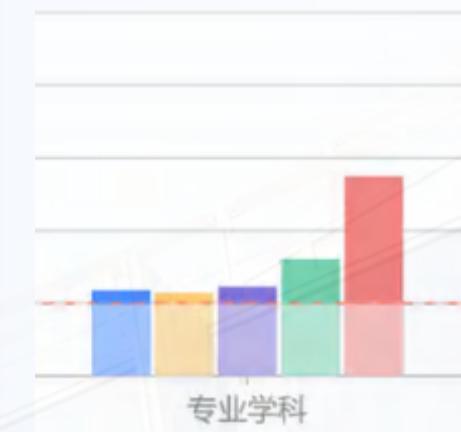
TGI分析

年龄特征



24H类图书馆与其他图书馆相比，18-24岁年轻群体TGI最高

- 参考“24小时城市书屋深得我心，空间宽敞，打工人深夜也能来”、“这个24小时自助图书馆让还书变得很便利”等评论，得知24H图书馆主要面向爱阅读，且对阅读时间灵活度要求较高的年轻职场人和学生
- 结合“学历预测”来看，24H图书馆受众教育水平为专科的占比较高，推测他们有可能在备考专升本



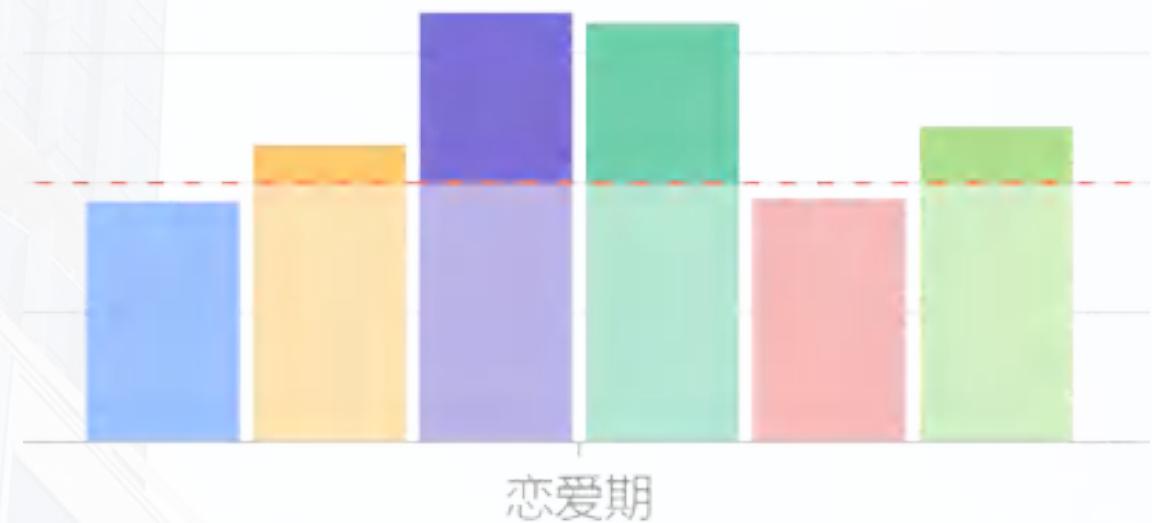
同城不同类 - 上海

TGI分析

婚恋状态

由图可知，徐家汇书院和上海图书馆的客流数据显示二者吸引处于恋爱期的访客。

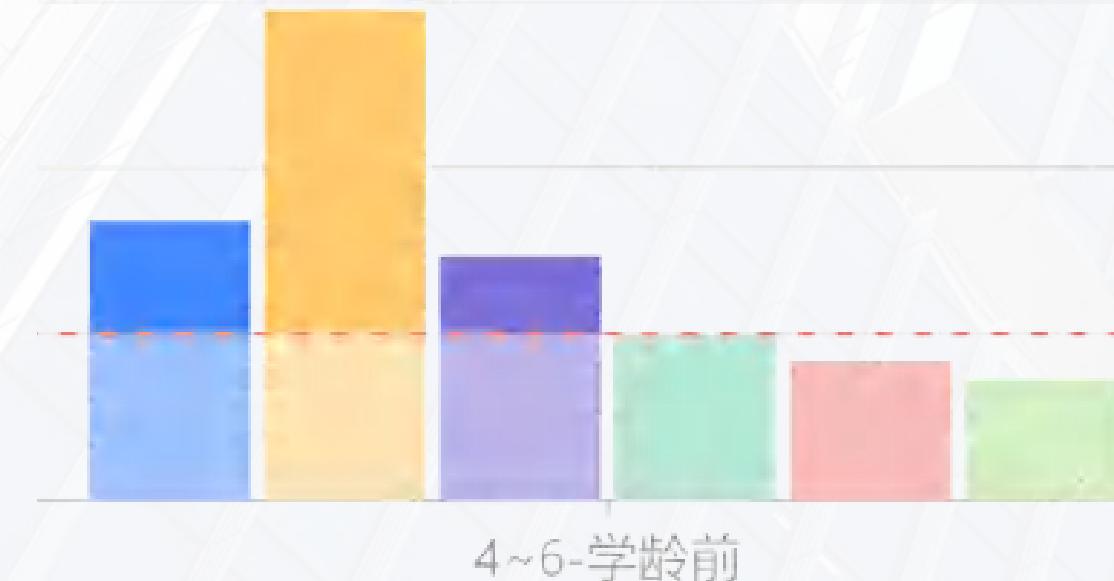
如标签所示，徐家汇书院和上海图书馆是网红打卡的综合性文化地标，有独特的建筑风格和丰富的文化活动，并且提供餐饮、文创等服务，适合恋爱期的年轻人在周末或假期进行参观打卡和休闲约会。



同城不同类 - 上海

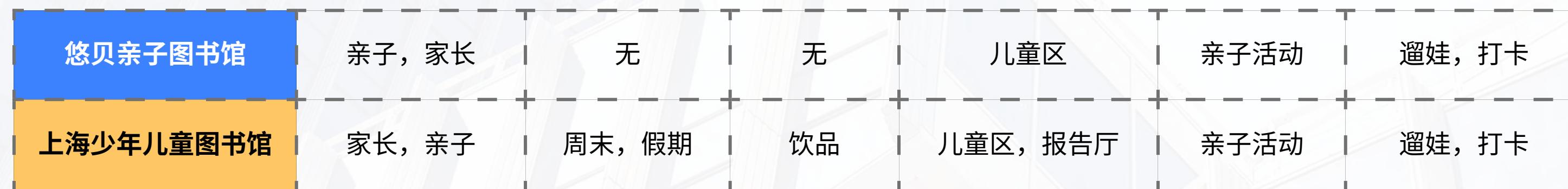
TGI分析

育儿阶段



由图可知，上海少年儿童图书馆和悠贝亲子图书馆的访客具有突出的带娃特征。

如标签所示，上海少年儿童图书馆和悠贝亲子图书馆在馆内都有儿童区，并且会举办亲子活动，均为少儿教育主题的图书馆，适合家长遛娃和儿童体验学习。



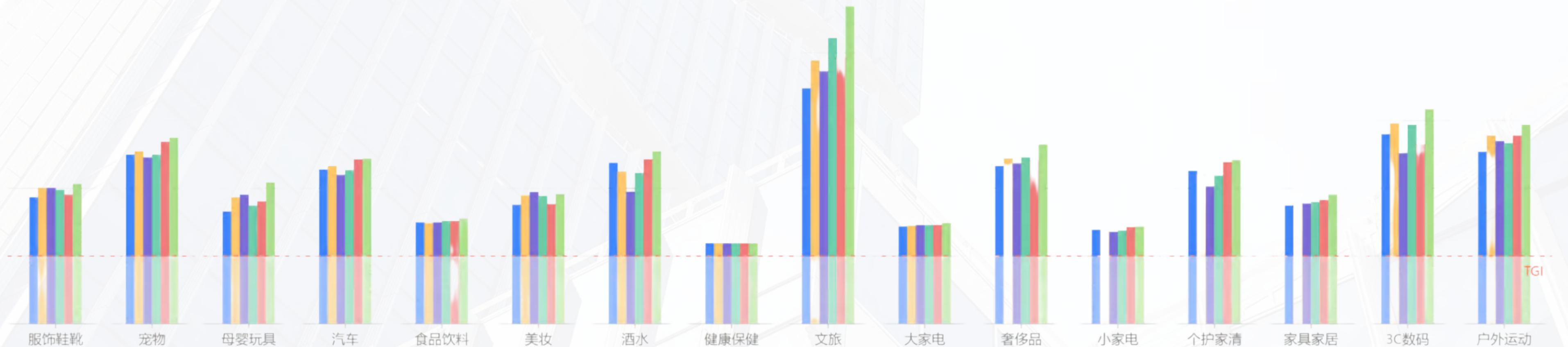


同城不同类 - 上海

TGI分析

电商偏好

图书馆客流数据的高收入、高消费、高文化需求特征，与其在电商偏好方面突出的文旅偏好特征相对应。





谢谢！

