

用户画像项目

达摩盘 基础特征 任务工单

八维文化与产业研究院

2025 年 5 月

1 基本信息

工单对应的项目名称	用户画像
工单编号	大数据-用户画像-11-达摩盘基础特征
创建时间	2025. 5. 1
创建人	郭洵

2 任务描述

任务描述	<p>作为〈电商平台广告运营人员〉</p> <p>我想要〈构建一个精准的用户画像标签体系，优化电商广告投放策略，提高广告效果和转化率〉</p> <p>以便于〈让更多的商家在平台上投放广告〉</p> <p>做任务前必看“达摩盘简介.docx”</p> <p>基础特征包括年龄、性别、年代、身高、体重、星座 6 类标签</p>
------	---



● 用户年龄标签

- 标签名称：用户年龄标签
- 标签定义：根据用户在淘宝平台上的浏览、搜索、收藏、加购、购买等全网行为数据，综合判断其所属的年龄段区间。年龄段分为6类：18-24岁、

25-29 岁、30-34 岁、35-39 岁、40-49 岁、50 岁以上。

➤ 数据来源：淘宝平台用户行为日志数据（包括但不限于点击、浏览、搜索、收藏、加购、下单、支付、商品评价等）、商品类目属性数据

➤ 具体处理逻辑：

若用户资料中存在有效出生年份，直接按当前年份-出生年份计算年龄并归类。

若无有效出生年份，进入多行为加权判定流程，判断条件包括类目偏好、品牌偏好、价格敏感度、时间行为、搜索词分析、社交互动行为、设备信息。

类目偏好得分计算：针对每个商品类目，预设其对应于各年龄段的偏好系数（由历史数据分析得出，不完整），对用户近期行为涉及的所有类目，按比例加权汇总，得到各年龄段的总分。

类目名称	18-24	25-29	30-34	35-39	40-49	50以上
潮流服饰	0.9	0.8	0.6	0.4	0.2	0.1
家居用品	0.2	0.4	0.6	0.8	0.9	0.7
健康食品	0.1	0.2	0.4	0.6	0.8	0.9

品牌偏好：

品牌名	18-24	25-29	30-34	35-39	40-49	50以上
ZARA	0.9	0.7	0.5	0.3	0.2	0.1
海澜之家	0.1	0.3	0.5	0.7	0.8	0.9

价格敏感度：

价格区间	18-24岁	25-29岁	30-34岁	35-39岁	40-49岁	50岁以上
低价商品	0.8	0.6	0.4	0.3	0.2	0.1
中价商品	0.2	0.4	0.6	0.7	0.8	0.7
高价商品	0.1	0.2	0.3	0.4	0.5	0.6

时间行为:

时间段	18-24岁	25-29岁	30-34岁	35-39岁	40-49岁	50岁以上
凌晨	0.2	0.1	0.1	0.1	0.1	0.1
早晨	0.1	0.1	0.1	0.1	0.2	0.3
上午	0.2	0.2	0.2	0.2	0.3	0.4
中午	0.4	0.4	0.4	0.4	0.4	0.3
下午	0.4	0.5	0.5	0.5	0.5	0.4
晚上	0.8	0.7	0.6	0.5	0.4	0.3
夜间	0.9	0.7	0.5	0.3	0.2	0.1

凌晨: 00:00 – 05:59

早晨: 06:00 – 08:59

上午: 09:00 – 11:59

中午: 12:00 – 13:59

下午: 14:00 – 17:59

晚上: 18:00 – 21:59

夜间: 22:00 – 23:59

搜索词分析：

搜索词类别	18-24岁	25-29岁	30-34岁	35-39岁	40-49岁	50岁以上
时尚与潮流	0.9	0.7	0.5	0.3	0.2	0.1
性价比	0.2	0.4	0.6	0.7	0.8	0.8
健康与养生	0.1	0.2	0.4	0.6	0.8	0.9
家庭与育儿	0.1	0.2	0.4	0.6	0.8	0.7
科技与数码	0.8	0.6	0.4	0.3	0.2	0.1
学习与发展	0.4	0.5	0.6	0.7	0.8	0.7

时尚与潮流：“潮流”、“新款”、“时尚”等。

性价比：“优惠”、“折扣”、“便宜”等。

健康与养生：“健康食品”、“保健品”、“养生”等。

家庭与育儿：“婴儿用品”、“儿童玩具”、“亲子活动”等。

科技与数码：“智能手机”、“电脑配件”、“智能设备”等。

学习与发展：“在线课程”、“书籍”、“技能提升”等。

社交互动行为：

社交互动行为	18-24岁	25-29岁	30-34岁	35-39岁	40-49岁	50岁以上
参与直播	0.9	0.7	0.5	0.4	0.3	0.2
发表评论	0.8	0.6	0.5	0.4	0.3	0.2
分享商品	0.7	0.6	0.5	0.4	0.3	0.2
加入社区讨论	0.8	0.6	0.5	0.4	0.3	0.2
关注达人/店铺	0.7	0.6	0.5	0.4	0.3	0.2

参与直播：包括观看直播、与主播互动（点赞、评论）、购买直播推荐商品等。

发表评论：对商品或服务进行评价，分享使用体验。

分享商品：将感兴趣的商品分享给朋友或社交网络。

加入社区讨论：参与淘宝内的社区讨论，如圈子、话题讨论等。

关注达人/店铺：关注感兴趣的达人账号或店铺，获取最新动态。

设备信息：

设备信息类别	18-24岁	25-29岁	30-34岁	35-39岁	40-49岁	50岁以上
iOS设备	0.7	0.6	0.5	0.4	0.3	0.2
Android设备	0.8	0.7	0.6	0.5	0.4	0.3
桌面端 (PC/Mac)	0.3	0.4	0.5	0.6	0.7	0.8
最新系统版本	0.8	0.7	0.6	0.5	0.4	0.3
高端设备	0.4	0.5	0.6	0.7	0.8	0.7

iOS 设备：使用苹果公司的 iPhone 或 iPad 等设备。

Android 设备：使用基于 Android 系统的各种品牌手机或平板电脑。

桌面端（PC/Mac）：通过个人电脑或 Mac 访问淘宝。

最新系统版本：用户是否倾向于更新到操作系统的最新版本。

高端设备：指价格较高、配置较好的设备。

将上述所有维度的得分加权汇总，得到每个年龄段的总分，取最高分作为最终年龄段判断结果。

维度	权重
类目偏好	30%
品牌偏好	20%
价格敏感度	15%
时间行为	10%
搜索词分析	10%
社交互动行为	10%
设备信息	5%

计算举例：

某个用户的年龄段标签，各个维度的偏好得分如下：

维度	18-24岁得分	25-29岁得分	30-34岁得分	35-39岁得分	40-49岁得分	50岁以上得分
类目偏好	0.7	0.6	0.5	0.4	0.3	0.2
品牌偏好	0.8	0.6	0.4	0.3	0.2	0.1
价格敏感度	0.9	0.7	0.5	0.4	0.3	0.2
时间行为	0.6	0.5	0.4	0.3	0.2	0.1
搜索词分析	0.7	0.6	0.5	0.4	0.3	0.2
社交互动行为	0.8	0.6	0.5	0.4	0.3	0.2
设备信息	0.7	0.6	0.5	0.4	0.3	0.2

对于每个年龄段，我们将对应的维度得分乘以其权重，然后求和得到最终得分。以 18-24 岁为例：

总分 18-24 = $0.21 + 0.16 + 0.135 + 0.06 + 0.07 + 0.08 + 0.035 = 0.75$

同理，可以计算其他年龄段的总分：

总分 25-29 = 0.69

总分 30-34=0.585

总分 35-39=0.47

总分 40-49=0.365

总分 50 以上=0.26

该用户年龄标签为 18-24 岁。

➤ 输出标签属性值：18-24 岁、25-29 岁、30-34 岁、35-39、40-49 岁、50 岁以上。

➤ 更新频率：每日更新一次，基于用户最近 30 天的行为数据进行重新计算。对于无新行为的用户，保留上一次标签结果。

● 用户性别标签

➤ 标签名称：用户性别标签（全网行为综合判断）

➤ 标签定义：

根据用户在淘宝平台上的全网行为（浏览、收藏、加购、购买等），结合其购买商品的一级品类分布及行为频次，综合判断用户性别倾向。标签分为三类：

女性用户：以女性相关品类购买行为为主。

男性用户：以男性相关品类购买行为为主。

家庭用户：购买行为覆盖男女装、儿童用品或家庭日用品，且无明显性别倾向。

➤ 数据来源：

1. 用户行为数据：

浏览、收藏、加购、购买的商品记录（一级品类）。

行为发生的时间戳（用于计算近 30 天/90 天活跃度）。

2. 商品一级品类列表：

女装/女士精品、女鞋、美容护肤/美体/精油、女士内衣/男士内衣/家居服、饰品/流行首饰/时尚饰品新、孕妇装/孕产妇用品/营养、童鞋/婴儿鞋/亲子鞋、童装/婴儿装/亲子装、尿片/洗护/喂哺/推车床等（女性相关品类）。

男装、流行男鞋、男鞋、运动服/休闲服装、运动鞋、网游装备/游戏币/帐号/代练、电玩/配件/游戏/攻略等（男性相关品类）。

家居用品、厨房电器、生活电器、收纳整理、居家布艺、家庭保健、儿童用品等（家庭相关品类）。

➤ 具体处理逻辑：

1. 行为权重分配：

购买行为：权重 50%（单次购买行为对性别判断的贡献最大）。

加购/收藏行为：权重 30%（用户有明确购买意向但未完成支付）。

浏览行为：权重 20%（仅作为辅助参考，需结合其他行为）。

2. 品类分类规则：

女性相关品类：

女装/女士精品、女鞋、美容护肤/美体/精油、女士内衣/男士内衣/家居服、饰品/流行首饰/时尚饰品新、孕妇装/孕产妇用品/营养、童鞋/婴儿鞋/亲子鞋、童装/婴儿装/亲子装、尿片/洗护/喂哺/推车床等。

男性相关品类：

男装、流行男鞋、男鞋、运动服/休闲服装、运动鞋、网游装备/游戏币/帐号/代练、电玩/配件/游戏/攻略等。

家庭相关品类：

居家日用、厨房电器、生活电器、收纳整理、居家布艺、家庭保健、儿童用品等。

3. 评分计算规则：

对每个用户，计算其在女性相关品类和男性相关品类的行为得分：

女性得分 = Σ (购买行为 \times 50% + 加购/收藏行为 \times 30% + 浏览行为 \times 20%) \times 对应品类权重。

男性得分 = Σ (购买行为 \times 50% + 加购/收藏行为 \times 30% + 浏览行为 \times 20%) \times 对应品类权重。

家庭用户判断条件：

若用户在家庭相关品类的购买行为占比 $\geq 30\%$ ，且女性得分与男性得分的差值 $\leq 10\%$ ，则归类为家庭用户。

4. 阈值划分：

女性用户：女性得分 $>$ 男性得分 + 10%。

男性用户：男性得分 $>$ 女性得分 + 10%。

家庭用户：满足上述家庭用户判断条件。

➤ 输出标签属性值：女性用户、男性用户、家庭用户

➤ 更新频率：

每日更新：基于用户最近 30 天的行为数据动态计算，确保标签反映最新行为

趋势。对于无新行为的用户，保留上一次标签结果。

全量更新：每月对历史用户进行一次全量重新计算，修正长期行为偏差。

➤ 补充说明：

排除干扰行为：若用户存在刷单行为（如短时间内高频浏览/加购非目标品类），需通过异常行为检测模块过滤后，再进行标签计算。

● 用户体重

➤ 标签名称：用户体重（千克）

标签定义：基于用户在淘宝系平台（淘宝、天猫、聚划算等）主动填写或设备同步的体重信息，经规则处理后生成的标准化体重标签。

➤ 数据来源：

1. 订单信息：用户购买体重秤、健康类商品时填写的体重数据（如商品评价、售后反馈）。

2. 会员资料：用户主动在淘宝/天猫个人资料中填写的体重字段。

3. 健康设备同步：接入淘宝生态的智能设备（如体脂秤）自动上传的体重数据。

4. 活动表单：用户参与大促活动（如聚划算健康品类活动）时提交的体重信息。

➤ 具体处理逻辑：

1. 数据抽取：

从上述来源提取原始体重字段（含数值及单位，如“60kg”“120斤”“70.5”）。

2. 数据清洗与转换:

单位统一:

若含“斤”单位, 数值除以 2 (如“120 斤” → 60kg);

若含“克”“g”, 数值除以 1000 (如“65000g” → 65kg);

无单位默认“千克”。

异常值过滤:

剔除 <30kg (低于儿童最低合理值) 或 >200kg (成人极端值) 的记录;

排除非数值字符 (如“50 公斤左右”标记为无效)。

3. 数据合并规则:

优先级权重 (由高到低):

1. 健康设备同步数据 (实时性强、精度高, 权重 1.0);

2. 订单信息 (与消费场景强相关, 权重 0.8);

3. 会员资料 (用户主动维护, 权重 0.6);

4. 活动表单 (可能存在随意填写, 权重 0.4)。

冲突处理:

取权重最高的来源数据;

同权重时取最新时间戳数据。

4. 数值标准化, 结果四舍五入保留 1 位小数 (如“65.4kg” → 65.5kg)。

➤ 输出标签属性值:

有效值: 数值 (如“68.5”);

无效值: 标记为“NULL” (无合规数据或数据冲突无法处理)。

➤ 更新频率:

T+1 日更新：每日凌晨拉取前一日增量数据，按规则覆盖历史值；

设备数据实时更新：若用户当日有设备同步记录，触发标签即时更新。

● 用户身高

➤ 标签名称：用户身高（厘米）

➤ 标签定义：

记录用户在淘宝、天猫、聚划算等阿里系电商平台主动填写的实名制身高数据，以及在购物场景（如服饰鞋帽购买、运动装备购买、孕婴用品购买等）中通过商品尺码选择、客服沟通记录、退换货备注等渠道间接确认的身高信息，单位精确到厘米。

➤ 数据来源：

1. 用户资料库（主来源）：用户账号后台「个人信息身体数据」模块填写的实名认证信息
2. 订单备注系统：服饰/鞋类订单中「商品尺码」选项选择的对应身高参数（如 180 对应 XL 码）
3. 客服工单系统：用户与服饰类目客服沟通中明确声明的身高记录
4. 退货原因库：用户选择「尺寸不符」退货时填写的实际身高数据
5. 智能硬件数据：接入淘宝运动的智能体脂秤设备同步数据（需用户授权）

具体处理逻辑：

1. 数据清洗：

过滤<100cm 或>250cm 的异常值

排除「666」「888」等特殊数字

过滤非整数数值

2. 多源数据优先级（权重）：

智能硬件数据（权重 1.0）

实名认证信息（权重 0.9）

近 30 天购物场景数据（权重 0.85）

历史购物场景数据（权重 0.7）

客服沟通记录（权重 0.6）

3. 冲突处理规则：

取加权平均值： $\Sigma(\text{各来源值} \times \text{权重}) / \text{总权重}$

当标准差>5cm 时，取最近 3 次有效记录众数

新数据覆盖旧数据原则：近 30 天数据覆盖率提升 20%

4. 补全策略：

服饰类目用户：用「购买尺码身高对照表」推算（男装：L=175cm，XL=180cm，每码±5cm；女装：L=170cm，XL=175cm）

母婴用户：用「儿童年龄身高百分位表」推算（需结合购买商品年龄段）

➤ 输出标签属性值：

精确值，如 175，单位厘米

➤ 更新频率：

实时更新：智能硬件数据、新订单数据

T+1 更新：用户资料修改、客服记录

周级更新：尺码推算规则库

月级更新：年龄身高对照表

- 星座

- 标签名称：星座标签

- 标签定义：根据用户填写的生日信息，解析对应的十二星座，用于分析用户的性格特征、消费偏好等画像属性。

- 数据来源：

用户淘宝/天猫个人资料中填写的生日字段，如未填写为未知。

- 具体处理逻辑：

提取月份（`month`）和日期（`day`），按严格分界点匹配星座：

摩羯座（12月22日~1月19日）

水瓶座（1月20日~2月18日）

双鱼座（2月19日~3月20日）

白羊座（3月21日~4月19日）

金牛座（4月20日~5月20日）

双子座（5月21日~6月21日）

巨蟹座（6月22日~7月22日）

狮子座（7月23日~8月22日）

处女座（8月23日~9月22日）

	<p>天秤座（9月23日～10月23日）</p> <p>天蝎座（10月24日～11月22日）</p> <p>射手座（11月23日～12月21日）</p> <p>➤ 输出标签属性值：“白羊座”，“金牛座”，“双子座”，“巨蟹座”，“狮子座”，“处女座”，“天秤座”，“天蝎座”，“射手座”，“摩羯座”，“水瓶座”，“双鱼座”，“未知”</p> <p>➤ 更新频率：</p> <p>实时更新：用户主动修改生日后，触发实时计算更新星座标签。</p> <p>离线批量更新：每日凌晨全量刷新，覆盖因数据源变更或规则调整导致的标签变动。</p>
工时预估	4人日
产出物	<p>1、达摩盘基础特征设计文档</p> <p>2、达摩盘基础特征代码</p> <p>3、达摩盘基础特征测试文档</p> <p>4、达摩盘基础特征上线记录</p>
验收标准	<p>1、1、达摩盘基础特征设计文档需包括 ADS 层表设计、原始数据分析、关键标签实现方案，性能优化方案，需写出明确的实现思路。</p> <p>2、2、达摩盘基础特征代码需全部标签的功能，需包括不分层、分层两种实现方式，代码需有明确的注释。</p>

	<p>3、3、达摩盘基础特征测试文档需含测试记录及测试 sql，同时分层、不分层两种实现方式的数据需一致。</p> <p>4、达摩盘基础特征上线记录需包括原始数据导出（csv 格式 10 万条以下），处理后的数据导出（csv 格式 10 万条以下），运行日志，运行过程截图。</p>
备注	代码注释需包括工单编号：大数据-用户画像-11-达摩盘基础特征