



大数据用户画像设计与实践（下）

主讲人：姚劲





01

用户画像基础

02

数据指标体系

架构设计

03

标签数据开发

千亿级实时数仓架构设计与实践

04

用户画像产品化和用户画像应用

千亿级实时数仓架构设计与实践

The background is a solid green color with several large, overlapping, organic shapes in a slightly lighter shade of green. These shapes resemble stylized human figures or abstract forms, creating a layered, textured effect.

01

用户画像基础

消费信息

快餐	2000.78元
酒水	864.38元
饮料	280.00元
零食	50.66元

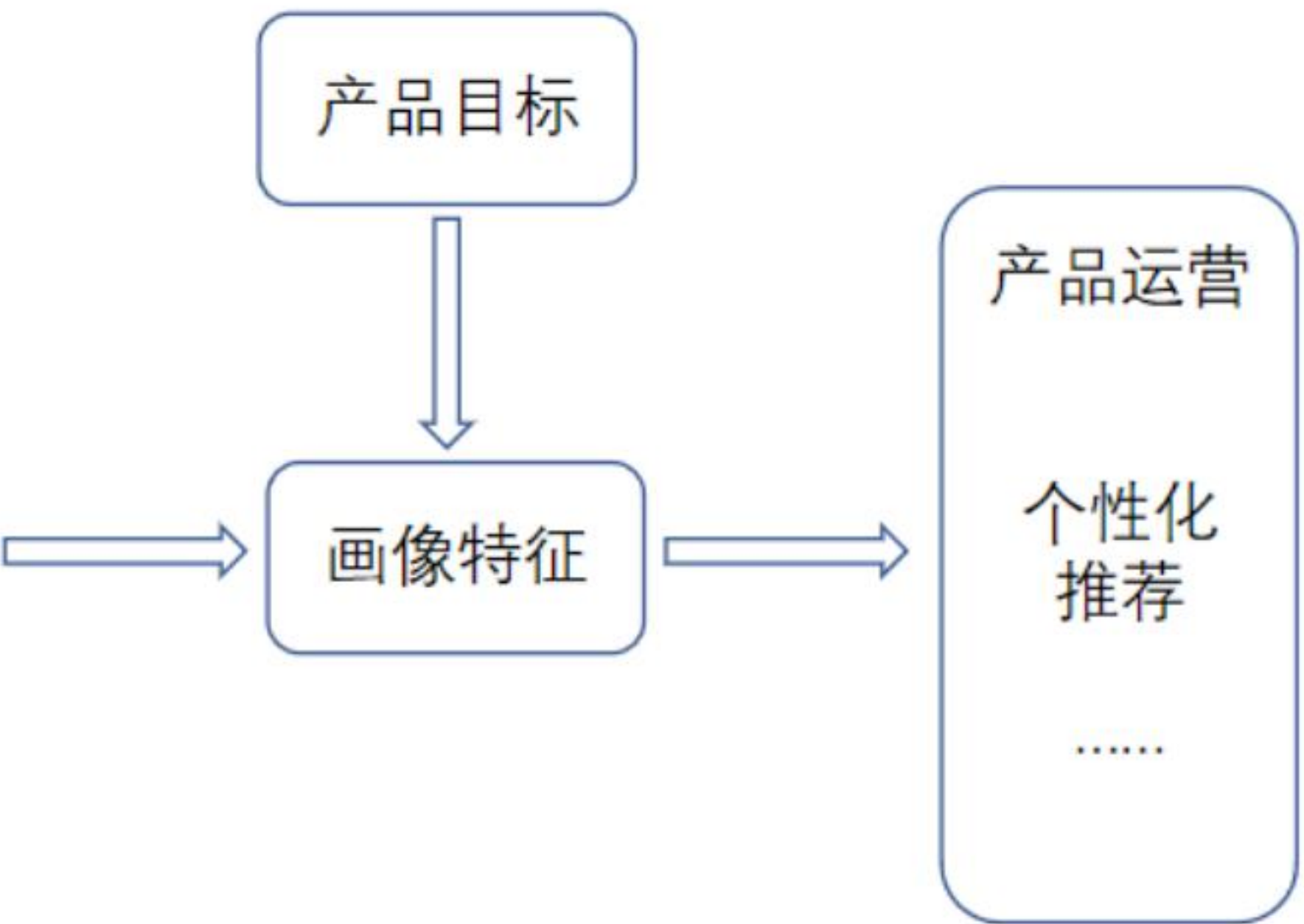
行为数据

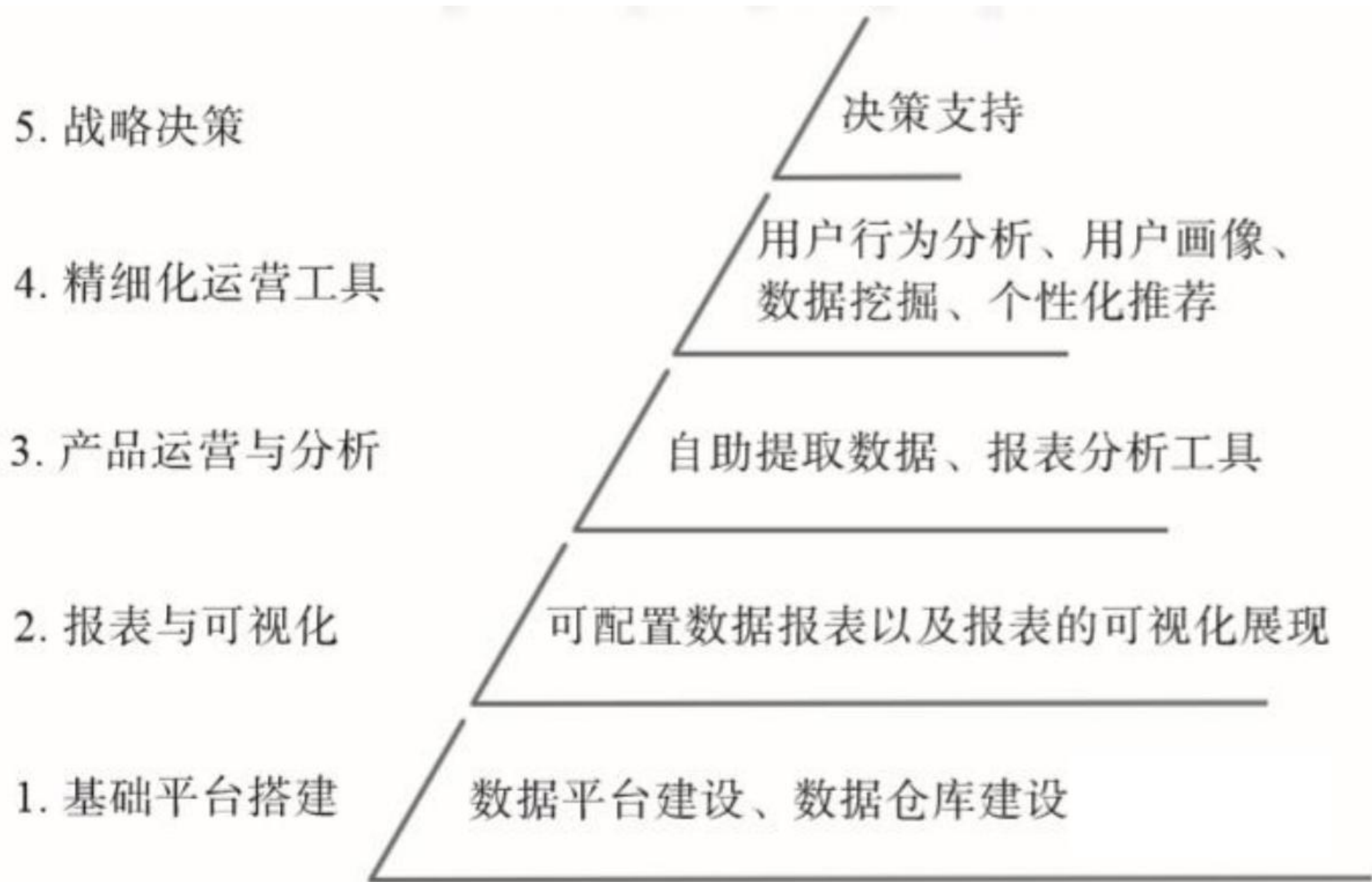
浏览时长	30s
点击次数	5次
购买	300.00元
...

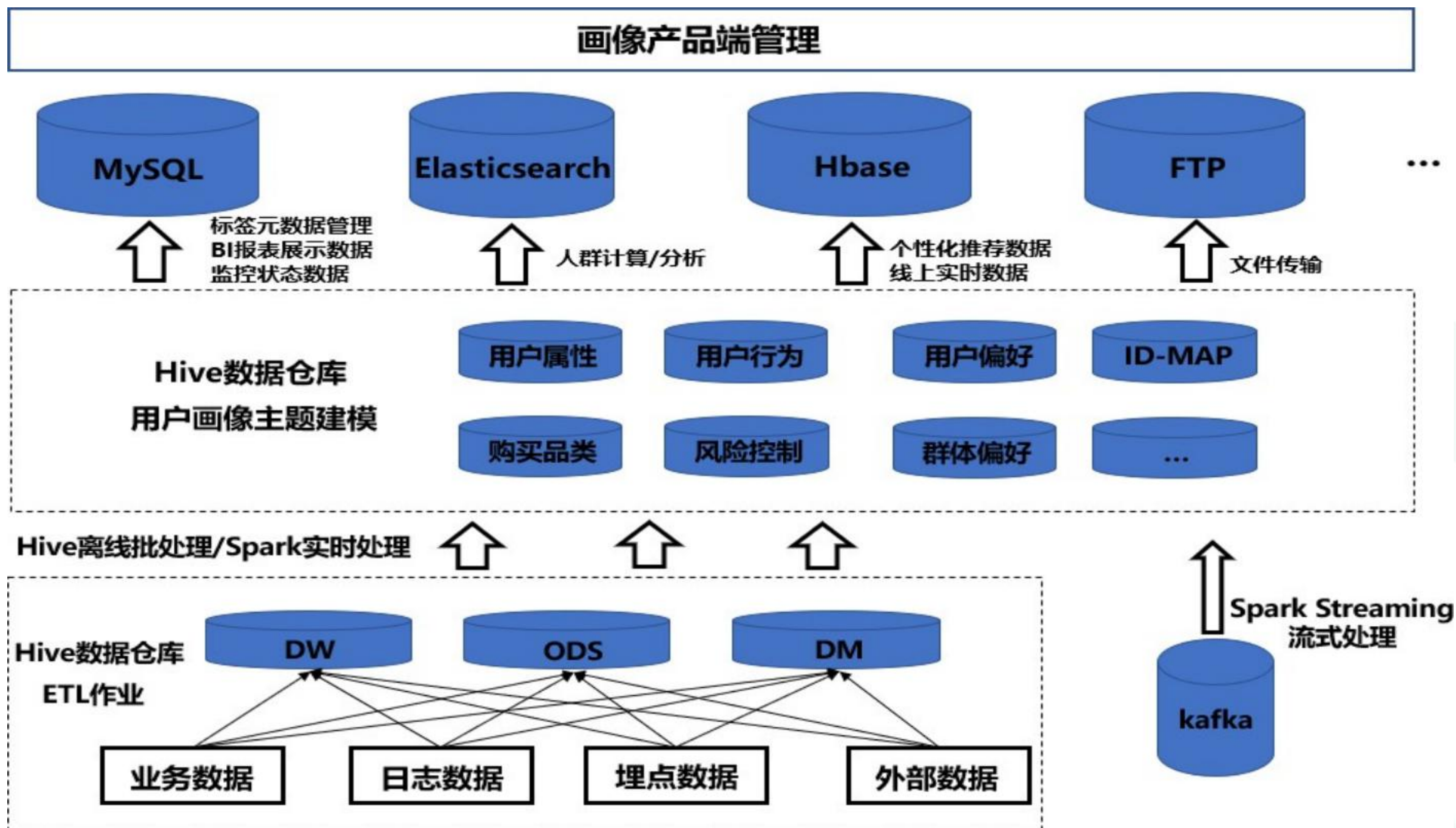


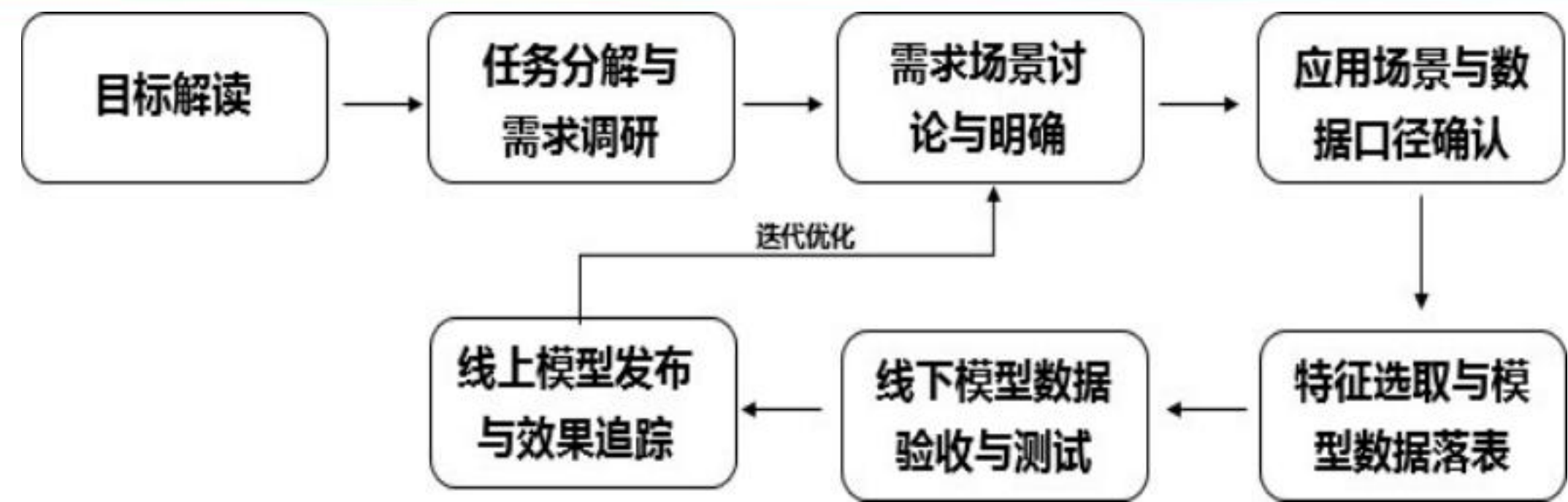
基本属性

用户id	348392706
手机号	185....9389
是否会员	是
会员时长值	200天
人脸认证	已认证
微信认证	已认证
营业执照法人	李四
imei	7673278
身份证证	是









第一阶段：目标解读

在建立用户画像前，首先需要明确用户画像服务于企业的对象，根据业务方需求，未来产品建设目标和用户画像分析之后预期效果；

第二阶段：任务分解与需求调研

经过第一阶段的需求调研和目标解读，我们已经明确了用户画像的服务对象与应用场景，接下来需要针对服务对象的需求侧重点，结合产品现有 业务体系和“数据字典”规约实体和标签之间的关联关系，明确分析纬度；

第三阶段：需求场景讨论与明确

在本阶段，数据运营人员需要根据前面与需求方的沟通结果，输出《产品用户画像规划文档》，在该文档中明确画像应用场景、最终开发出的标 签内容与应用方式， 并就该份文档与需求方反复沟通确认无误。

第四阶段：应用场景与数据口径确认

经过第三个阶段明确了需求场景与最终实现的标签纬度、标签类型后，数据运营人员需要结合业务与数据仓库中已有的相关表，明确与各业务场 景相关的数据口径。在该阶段中，数据运营方需要输出《产品用户画像实施文档》，该文档需要明确应用场景、标签开发的模型、涉及到的数据库 与表，应用实施流程；

第五阶段：特征选取与模型数据落表

本阶段中数据分析挖掘人员需要根据前面明确的需求场景进行业务建模，写好HQL逻辑，将相应的模型逻辑写入临时表中，抽取数据校验是否符 合业务场景需求。

第六阶段：线下模型数据验收与测试

数据仓库团队的人员将相关数据落表后，设置定时调度任务，进行定期增量更新数据。数据运营人员需要验收数仓加工的HQL逻辑是否符合需求， 根据业务需求抽取查看表中数据范围是否在合理范围内，如果发现问题及时反馈给数据仓库人员调整代码逻辑和行为权重的数值。

第七阶段：线上模型发布与效果追踪

经过第六阶段，数据通过验收之后，就可以将数据接口给到搜索、或技术团队部署上线了。上线后通过对用户点击转化行为的持续追踪，调整优 化模型及相关权重配置。

任务类型	任务名称	任务内容	所需时间	重点内容
标签开发	性别标签开发	数据调研、熟悉数据字典，开发标签（包括统计类、算法类、实时类的标签）	xx天	数据调研、和业务方确认数据口径，标签开发上线。初期上线满足应用需求
	会员标签开发		xx天	
	活跃度标签开发		xx天	
	RFM标签开发		xx天	
	
ETL调度开发	任务依赖关系梳理	梳理各任务之间的依赖关系	xx天	满足定时调度，监控预警、失败重试，各调度任务之间的复杂依赖关系
	监控脚本开发	开发标签监控、人群计算监控、服务层监控等相关脚本	xx天	
	调度脚本开发	根据梳理的各任务间调度依赖，开发调度流脚本	xx天	
	上线调度系统	调度流脚本上线调度试运行/正式运行	xx天	
打通服务层接口	push系统业务对接沟通	画像人群数据和push系统的打通方案，开发方式确定	xx天	打通数据仓库数据和各业务系统之间的通路，提供稳健的服务
	外呼系统业务对接沟通	和外呼团队了解外呼业务场景，确定打通方式	xx天	
	广告系统业务对接沟通	和广告团队了解目前广告场景，确定打通方式	xx天	
	客服系统业务对接沟通	和客服团队确认系统打通方式	xx天	
	
画像产品化	产品经理与业务人员、技术开发对接沟通	确定产品功能、画原型、明确开发排期	xx天	产品交互友好，能支持到业务方对用户进行分析、精细运营的需求
	JAVA Web端开发	开发测试、内测	xx天	
	产品上线	通知各业务方使用产品	xx天	
开发调优	标签脚本、调度脚本的重构优化	梳理现有标签开发、调度、校验告警、同步到服务层等相关脚本，明确可以优化的地方，迭代优化	xx天	减少ETL调度时间，降低调度时消耗资源
面向业务方推广应用	写画像使用文档	面向数据分析师、业务人员等群体撰写详细的画像使用文档，包括相关表及元数据、产品使用手册等	xx天	帮助业务人员将画像数据应用到业务中去，提高用户活跃、提高GMV
	提供业务支持	针对业务场景，为业务方提供画像解决方案	xx天	

02

数据指标体系

用户画像建模其实就是对用户进行打标签，从对用户打标签的方式来看，一般分为三种类型：1、基于统计类的 标签；2、基于规则类的标签、3、基于挖掘类的标签。这三种类型标签的区别：

- **统计类标签**：这类标签是最为基础也最为常见的标签类型，例如对于某个用户来说，他的性别、年龄、城市、星座、近7日活跃时长、近7日活跃天数、近7日活跃次数等字段可以从用户注册数据、用户访问、消费类数据中统计得出。该类标签构成了用户画像的基础；
- **规则类标签**：该类标签基于用户行为及确定的规则产生。例如对平台上“消费活跃”用户这一口径的定义为近30天交易次数 ≥ 2 。在实际开发画像的过程中，由于运营人员对业务更为熟悉、而数据人员对数据的结构、分布、特征更为熟悉，因此规则类标签的规则确定由运营人员和数据人员共同协商确定；
- **机器学习挖掘类标签**：该类标签通过数据挖掘产生，应用在对用户的某些属性或某些行为进行预测判断。例如根据一个用户的行为习惯判断该用户是男性还是女性，根据一个用户的消费习惯判断其对某商品的偏好程度。该类标签需要通过算法挖掘产生。

在项目工程实践中，一般统计类和规则类的标签即可以满足应用需求，开发中占有较大比例。机器学习挖掘类 标签多用于预测场景，如判断用户性别是男是女，判断用户购买商品偏好、判断用户流失意向等。一般地机器学习标签开发周期较长，耗费开发成本较大，因此其开发所占比例较小。

标签主题	标签类型	开发方式	是否互斥关系	用户维度
A: 用户属性 B: 用户行为 C: 用户消费 D: 风险控制	1: 分类型 2: 统计型	1: 统计型 2: 算法型	1: 互斥关系 2: 非互斥关系	C: cookieid U: userid

标签主题：用于刻画属于那种类型的标签，如用户属性、用户行为、用户消费、风险控制等多种类型，可用A、B、C、D等 字母表示各标签主题；

标签类型：标签类型可划为分类型和统计型这两种类型，其中分类型用于刻画用户属于哪种类型，如是男是女、是否是会员、 是否已流失等标签，统计型标签用于刻画统计用户的某些行为次数，如历史购买金额、优惠券使用次数、近30日登陆次数等 标签，这类标签都需要对应一个用户相应行为的权重次数；

开发方式：开发方式可分为统计型开发和算法型开发两大开发方式。其中统计型开发可直接从数据仓库中各主题表建模加工 而成，算法型开发需要对数据做机器学习的算法处理得到相应的标签；

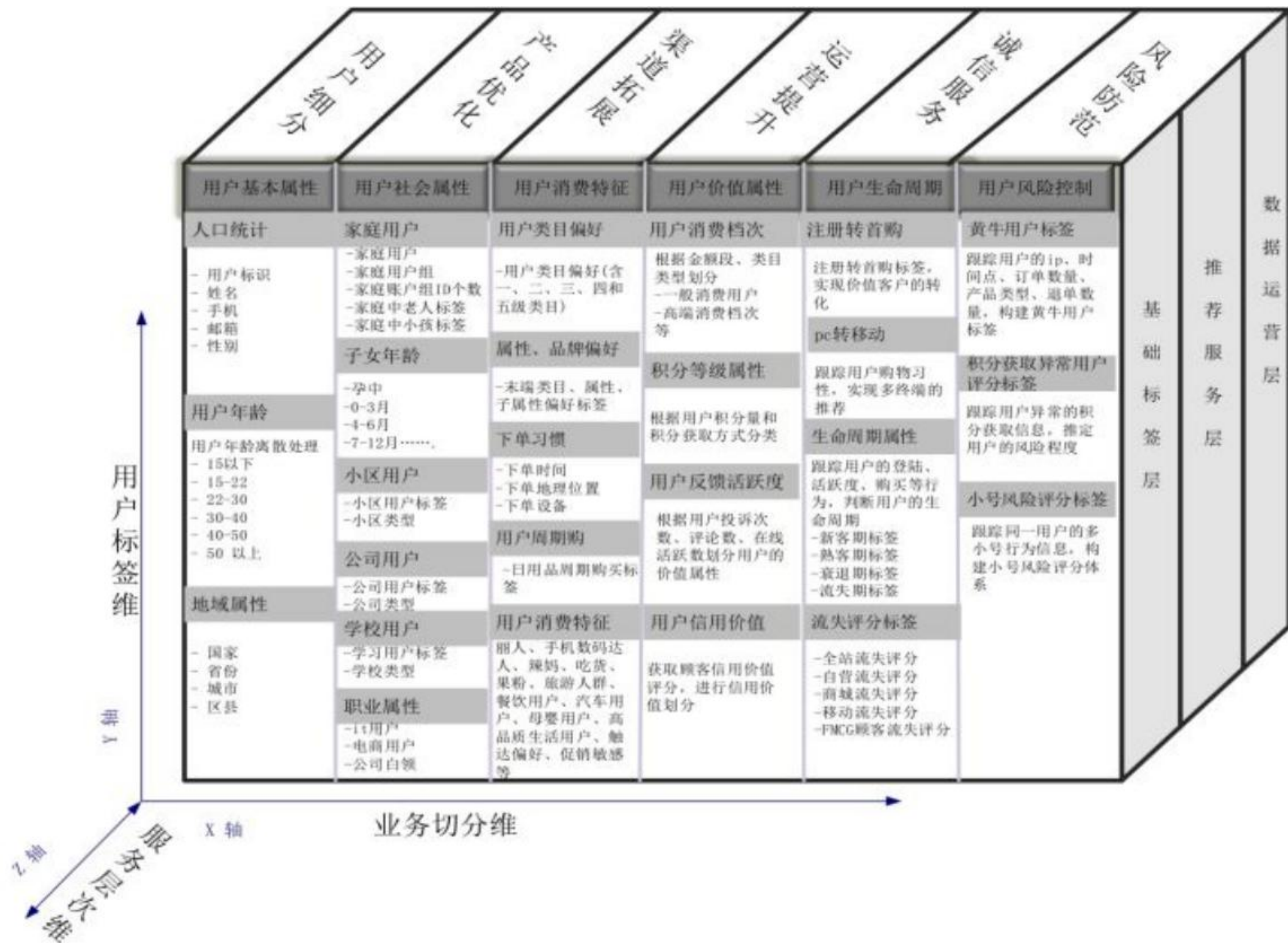
是否互斥标签：对应同一级类目下（如一级标签、二级标签），各标签之间的关系是否为互斥，可将标签划分为互斥关系和 非互斥关系。例如对于男、女标签就是互斥关系，同一个用户不是被打上男性标签就是女性标签，对于高活跃、中活跃、低 活跃标签也是互斥关系；

用户维度：用于刻画该标签是打在用户唯一标识（userid）上，还是打在用户使用的设备（cookieid）上。可用U、C等字 母分别标识userid和cookieid 维度。

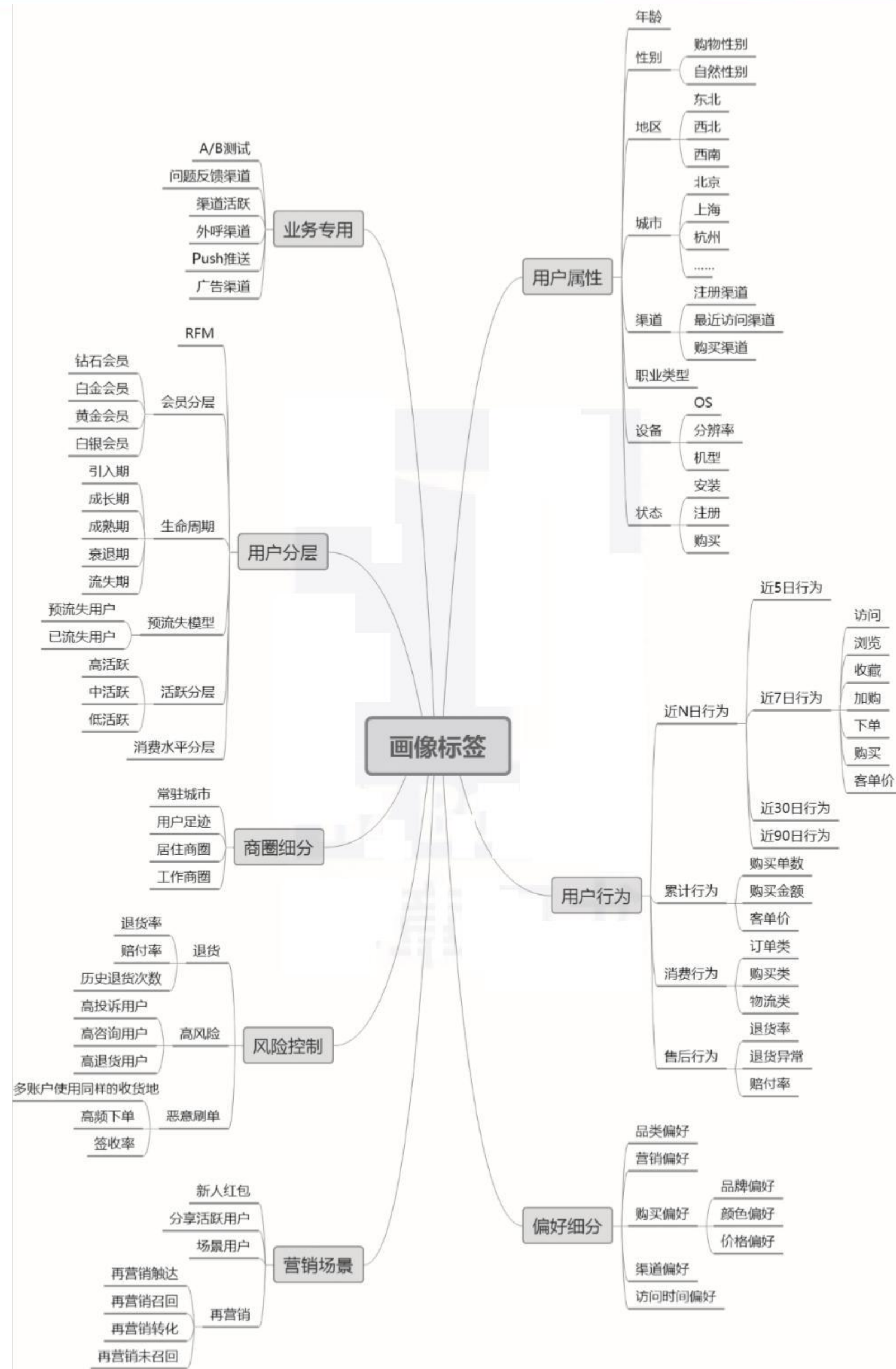
示例：

对于用户是男是女这个标签，标签主题是用户属性，标签类型属于分类型，开发方式为统计型，为互斥关系，用户 维度为userid。这样给男性用户打上标签 “A111U001_001”， 女性用户打上标签 “A111U001_002”， 其中 “A111U” 为上面介绍的命名方式，“001” 为一级标签的id，后面对于用户属性维度的其他一级标签可用 “002”、“003” 等方式追加命名，“_” 后面的 “001” 和 “002” 为该一级标签下的标签明细，如果是划分高、中、低活跃 用户的，对应一级标签下的明细可划分为 “001”、“002”、“003”。

标签id	标签名称	标签汉语	序号	标签主题	一级标签id	一级标签
A111H008_001	important value user	重要价值用户	1	用户属性	8	RFM
A111H008_002	important development use	重要发展用户	2	用户属性	8	RFM
A111H008_003	important maintain user	重要保持用户	3	用户属性	8	RFM
A111H008_004	important detainment user	重要挽留用户	4	用户属性	8	RFM
A111H008_005	general value user	一般价值用户	5	用户属性	8	RFM
A111H008_006	general development user	一般发展用户	6	用户属性	8	RFM
A111H008_007	general maintain user	一般保持用户	7	用户属性	8	RFM
A111H008_008	general detainment user	一般挽留用户	8	用户属性	8	RFM



对业务场景的角度出发
对标签体系进行归类



03

标签数据开发



标签数据开发是用户画像体系中最重要的一环，主要包括离线标签开发、实时标签开发、用户特征库开发、人群计算、打通数据服务层等开发内容。

统计类标签是指用户年龄、性别、购买金额、累计购买次数、近x日登次数等描述用户状态的标签。

▷ 案例：近30日购买行为标签

- ① 拆解二级标签。近30日购买行为拆解为：付款订单量 (ACTION_U_01_001)、总付款金额 (ACTION_U_01_002)、加入购物车次数 (ACTION_U_01_003)
- ② 将需要计算的标签从目标表中抽出来 (ACTION_U_01_001、ACTION_U_01_002、ACTION_U_01_003)

```
select 'ACTION_U_01_001' as labelid,          # 付款订单量标签id
      cast(user_id as string) as userid,
      count(distinct order_id) as labelweight # 付款订单量
from dw.order_info_fact    # 商品订单表
where pay_status = 1      # 订单状态是已支付
   and to_date(add_time) >= 'month_day_ago' # 付款日期大于等于30日前
   and to_date(add_time) <= 'yesterday_date' # 付款日期小于等于昨天
union all
select 'ACTION_U_01_002' as labelid,          #总付款金额标签id
      cast(user_id as string) as userid,
      sum(order_total_amount) as labelweight  #总付款金额
from dw.order_info_fact
where pay_status = 1          # 订单状态是已支付
   and to_date(add_time) >= 'month_day_ago'
   and to_date(add_time) <= 'yesterday_date'
union all
```

- ③ 增量获取用户最新状态，做全连接关联。即通过全连接（full outer join）方式，当有最新状态时获取最细状态，否则保留原来的状态标签；
- ④ 任务执行完后，将数据插入Hive数据表中；

userid	labelweight	data_date	labelid
44463729	3	20190101	ACTION_U_01_001
46763725	15	20190101	ACTION_U_01_001
54597977	20	20190101	ACTION_U_01_001
35601021	15	20190101	ACTION_U_01_001
47801349	7	20190101	ACTION_U_01_001
32101029	157	20190101	ACTION_U_01_002
65102349	228	20190101	ACTION_U_01_003
23551034	16	20190101	ACTION_U_01_003

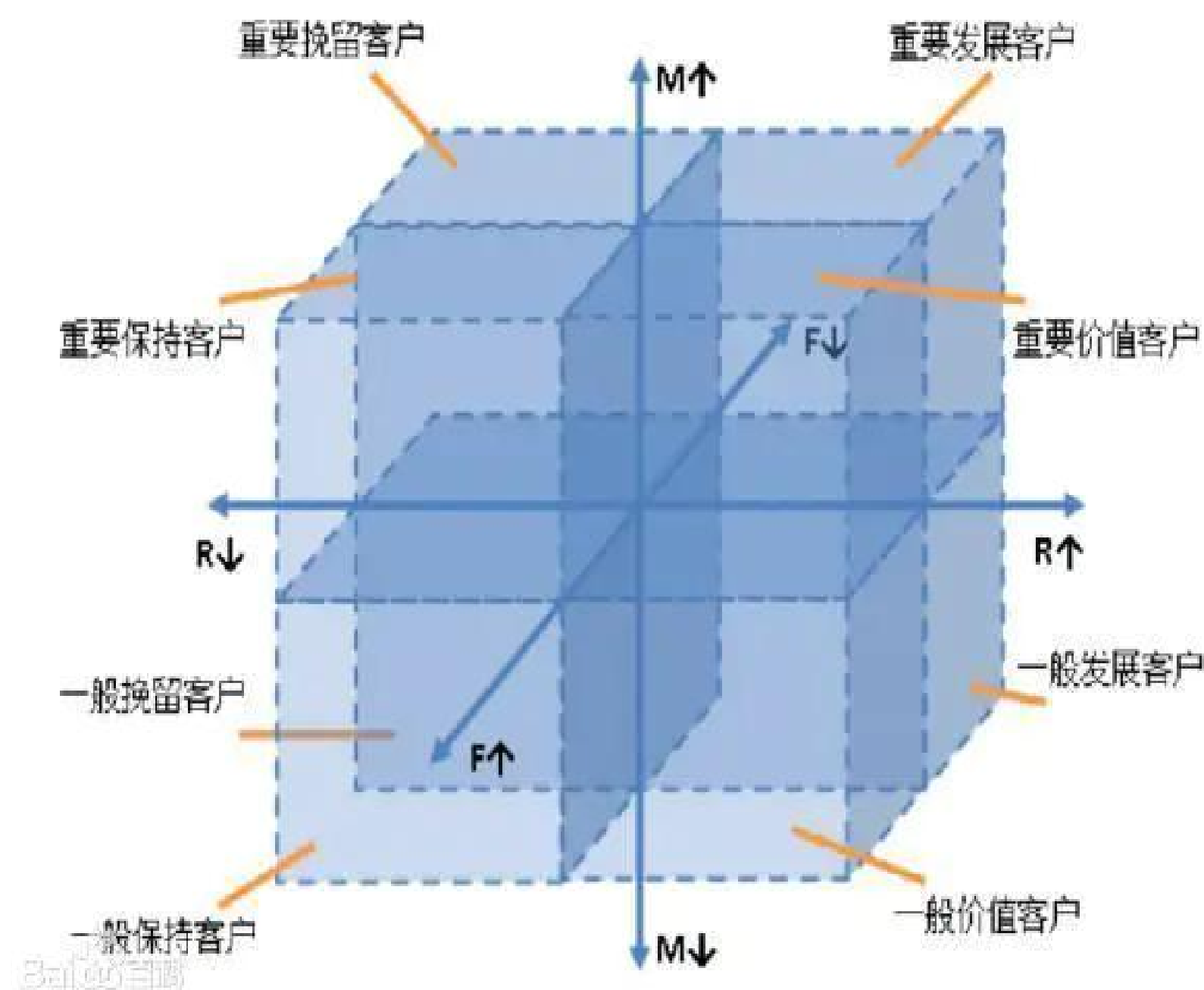
一般是根据业务场景，在业务层面上指定的规则标签。这类标签受主观判断因素的影响，在开发前需要做数据调研，结合运营业务规则开发。

除了写脚本开发标签外，还可自动打标签。比如用户触发的行为中，有超过80%的记录是3C商品，那么自动打上“数码达人”的标签。

▷ 案例：用户价值类标签

RFM模型是衡量用户价值的重要工具，包含3个指标，8类人群：

- (1) Recency：最近一次消费。指上一次购买的时候。
- (2) Frequency：消费频率。消费频率是顾客在限定的期间内所购买的次数。
- (3) Money：消费金额。消费金额是所有数据库报告的支柱，也可以验证“帕雷托法则” (Pareto's Law)——公司80%的收入来自20%的顾客。





在开发对应标签前需要进行数据调研。结合业务场景对3个维度的指标在时间定义或数值上进行确定和划分。

- ① 得到用户最近一次交易时间的分布。例如按照二八比例，将最近一次交易时间距今<90天的，定义为“近”，>90天的定义为“远”
- ② 得到用户近一年交易订单量分布。将历史订单<=3单的划分为“低频”，>3单的划分为“高频”
- ③ 得到将一年交易额分布。将交易金额<300的，定义为“低额”；>300的，定义为“高频”
- ④ 根据以上3个维度，进行交叉分析（ $R \leq 90 =$ “近”， $R > 90 =$ “远”； $F \leq 3 =$ “低频次”， $F > 3 =$ “高频次”； $M \leq 300 =$ “低额”， $M > 300 =$ “高额”），划分为以下8类人群

M (近1年交易金额)	F (近1年交易订单量)	R (最近一次交易时间)	定义	用户量占比	平均客单价	GMV总量	CMV贡献占比
高金额	高频率	近	重要价值用户				
高金额	高频率	远	重要保持用户				
高金额	低频率	近	重要发展用户				
高金额	低频率	远	重要挽留用户				
低金额	高频率	近	一般价值用户				
低金额	高频率	远	一般保持用户				
低金额	低频率	近	一般发展用户				
低金额	低频率	远	一般挽留用户				

- ⑤ 从用户消费订单表 (dw.user_consume_order_info) 里面读取用户最近一次消费距今天数、累计消费次数、累计消费金额这3个维度的数据，并注册视图user_rfm
- ⑥ 按照最近一次购买距今天90天、购买次数3次、消费500元来对用户3个维度进行高低划分。划分后的结果注册到视图user_rfm中将最终结果划分到8类人群中，再将结果插入用户标签表中

userid	labelweight	data_date	labelid
44463729		20190101	ATTRITUBE_U_06_001
46763725		20190101	ATTRITUBE_U_06_001
54597977		20190101	ATTRITUBE_U_06_002
35601021		20190101	ATTRITUBE_U_06_002
47801349		20190101	ATTRITUBE_U_06_002
32101029		20190101	ATTRITUBE_U_06_003
65102349		20190101	ATTRITUBE_U_06_003
23551034		20190101	ATTRITUBE_U_06_003

挖掘类标签即算法类标签，需要用算法挖掘用户相关特征。一般用户相关的挖掘标签可包括根据购买商品预测用户男女性别、预测用户点击下单、判断用户已流失或即将流失。挖掘类标签开发环节包括：

①用户行为特征工程开发、②算法调优、③上线工程化调度等环节，开发周期较长。

► **案例：对大量未打标签的文章、帖子等文本数据自动分类，自动打标签。**

1、特征选取及开发

① 标注：对一批文章进行人工准确分类，作为训练样本；

② 训练：计算机从标注好的文档集中挖掘出能够有效分类的规则，生成分类器。模型中用到的算法和数据处理技术包括文本分词、TF-IDF算法、朴素贝叶斯分类算法；

③ 分类：将生成的分类器应用在待分类的文档集中，从而获得文档的分类结果。

2、文本分词处理

将连续的字序按照一定规范重新组合成次序列的过程，中文分词是讲一个个汉字序列切分重一个个独立的单词。

3、数据结构处理

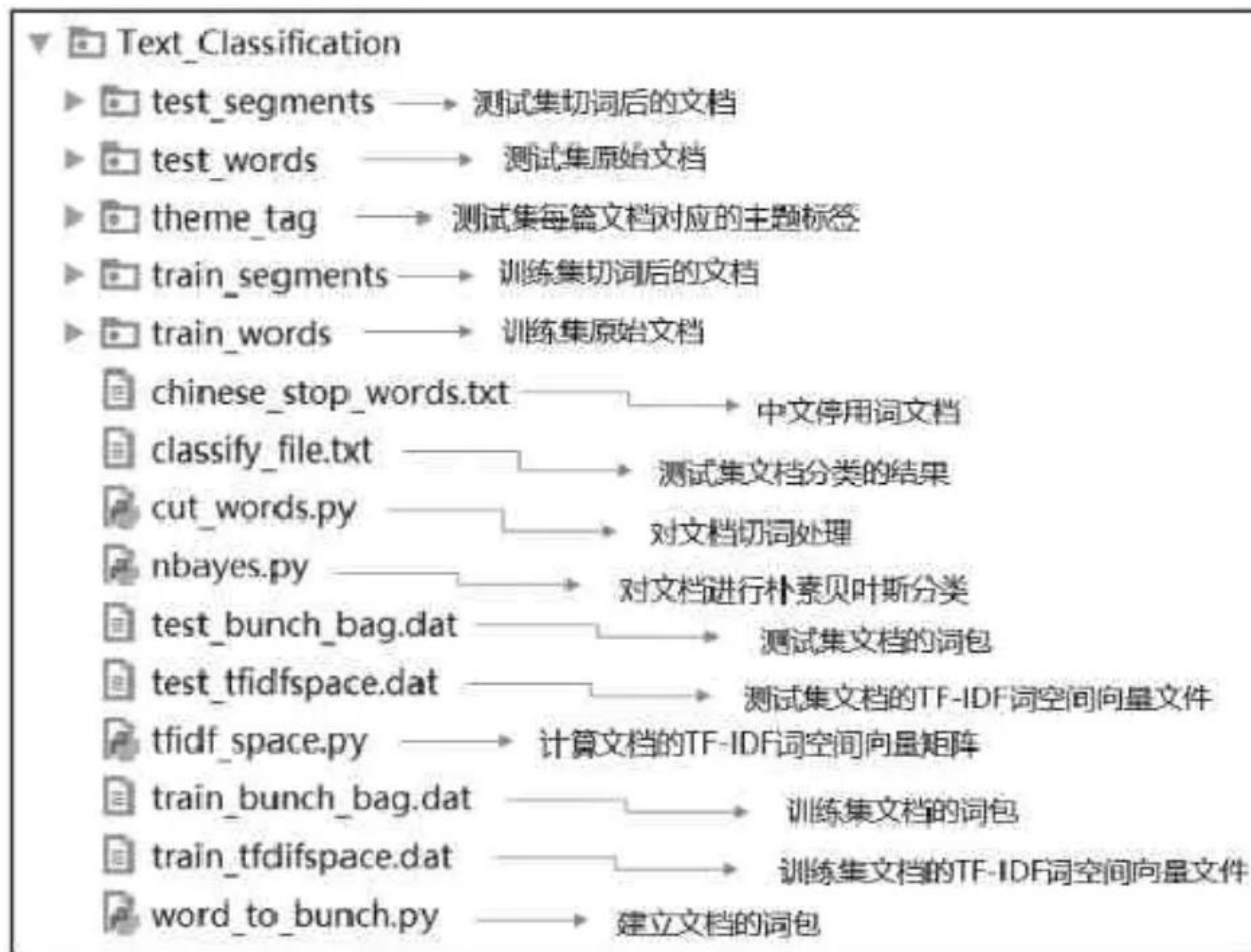
为了便于后续生成词向量空间模型，这些分词后的文本信息需要转换成文本向量信息并对象化。

4、文本TF-IDF权重

该步骤中将上一步存储的结构化数据构建成一个TF-IDF词向量空间，空间中的词均来自该训练集，各个词的权重矩阵也都一并保存下来。

5、朴素贝叶斯分类

至此，文本分类打标签流程中各模块的数据处理方式就完了，文件结构图如下：



离线标签的开发，即批次ETL任务，一般为T+1日的数据。流式计算标签主要是实时数据，比如实时订单分析，或者给首次登录App的新人用户弹窗推送、发放红包，实时分析用户场景并进行推送。

特征库是对用户每一次行为（浏览、首次、搜索、购买等）及该行为对应的标签（或商品类）进行详细的记录，以便挖掘用户喜好。是为个性化推荐、精准营销、商业分析提供中间层数据。一般来说，我认为这一块是最重要的。

- 用户标签：静态记录当前状态
- 用户特征库：多维度汇总。例如：一个用户经常浏览、购买奶粉等婴儿用品，则她可能是个妈妈；用户经常收藏、点赞搞笑视频，可用于挖掘用户偏好；用户经常搜索美妆美容类商品，可能用户是女性。

① 特征库规划

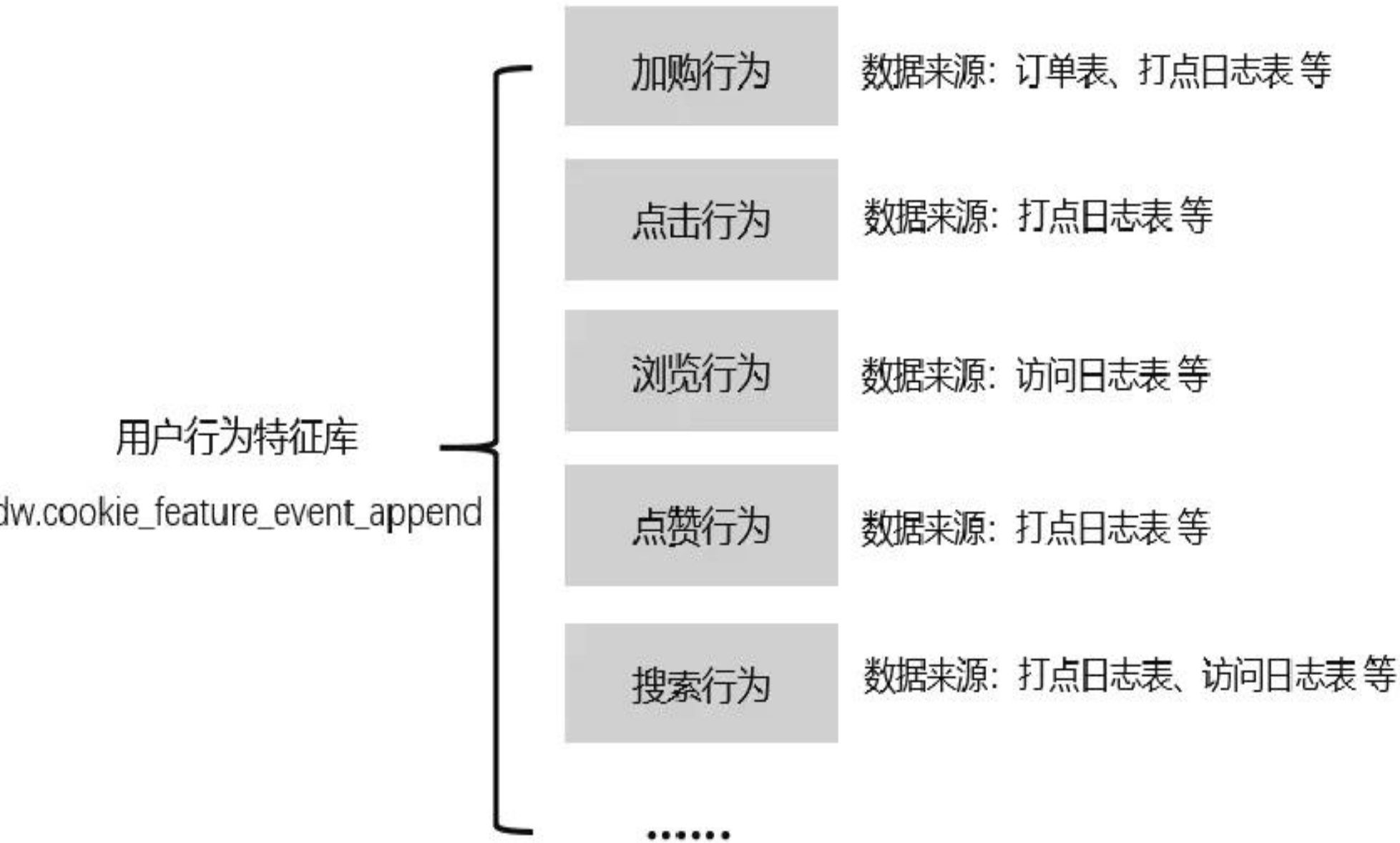
用户与商品相关的行为日志数据保护了用户对商品行为的明细。根据应用需要，创建表 dw.cookie_feature_event_append 来构建用户特征。

字段名	中文名	类 型	备 注
cookie_id	cookie_id	String	非空
tag_id	标签 id	String	非空
tag_name	标签名称	String	非空
tag_type_id	标签类型	String	按业务线来划分标签
act_num	事件统计值	Bigint	行为次数
data_date	日期分区	String	非空
act_name	事件名称	String	如点赞 / 打赏 / 加购 / 点击 / 收藏 / 浏览等行为

- 该表中的act_name是用户行为事件名称，大致可以分为3类：
- a、埋点日志数据：用户访问页面时点击的按钮，搜索的关键词通过打点日志上报记录；
 - b、访问日志数据：用户访问的页面、访问时间从访问日志数据挖掘；
 - c、订单数据：用户下单信息及商品信息。

② 数据开发

数据开发过程中，主要从订单表、访问日志表、打点日志表中对用户当日的行为（加购、点击、浏览、点赞等）抽取数据，然后清洗加载到用户特征库对应表（cookie_feature_event_append）



③ 其他特征库规划

除了用户特征库，还有围绕本公司的产品进行特征库的规划与开发

用户行为对应不同的权重，比如用户购买的权重肯定比加到购物车重，加入购物车的权重大于收藏，收藏大于浏览。

1、TF-IDF算法

TF-IDF是一种统计方法，用以评估一字词对于一个文件集或一个语料库中的其中一份文件的重要程度。字词的重要性随着它在文件中出现的次数成正比增加，但同时会随着它在语料库中出现的频率成反比下降。TF-IDF加权的各种形式常被搜索引擎应用，作为文件与用户查询之间相关程度的度量或评级。

2、时间衰减系数

在用户画像的应用中，用户的某些行为会随时间衰减，而某些行为不会随时间衰减。时间衰减是指随着时间的推移，用户的历史行为和当前行为的相关性不断减弱。

3、标签权重配置

用户标签的权重最终还是需要进一步结合标签所处的业务场景、距离当前时间、用户行为产生该标签的行为次数等因素，最终得到用户标签权重的综合打分公式：用户标签权重=行为类型权重×时间衰减×用户行为次数×TF-IDF计算标签权重

04

用户画像产品化与应用

输入用户userid或cookieid，可查看该用户各维度信息

标签视图

标签查询

标签编辑管理

用户分群

用户分析

请输入用户id

查询

姓名: 用户甲

浙江省 杭州市

Userid: 10000619

cookieid: 000003e4-d757-4490-8321-761cc41be1d1

用户属性

性别: 男

年龄: 26

注册时间: 2017-03-02 18:00:00

历史付费金额: 1500

RFM: 重要发展用户

购物性别: 男性

会员等级: 金卡会员

操作系统: iPhone 6s

历史购买次数: 6

用户活跃度: 高活跃

购买品类: 多品类购买

是否反馈问题: 否

用户行为

近30日购买次数: 2次

近30日购买金额: 200

高频活跃时间段: 晚上

是否短信黑名单: 否

是否邮件黑名单: 否

近30日活跃天数: 12天

最近下单距今天数: 18天

push周活跃度: 3天

订单优惠券使用率: 33.3%

首单距今天数: 300天

输入用户id后，可以查看该 用户的属性信息、行为信息、 风控属性等信息。从多方位 了解一个具体的用户特征

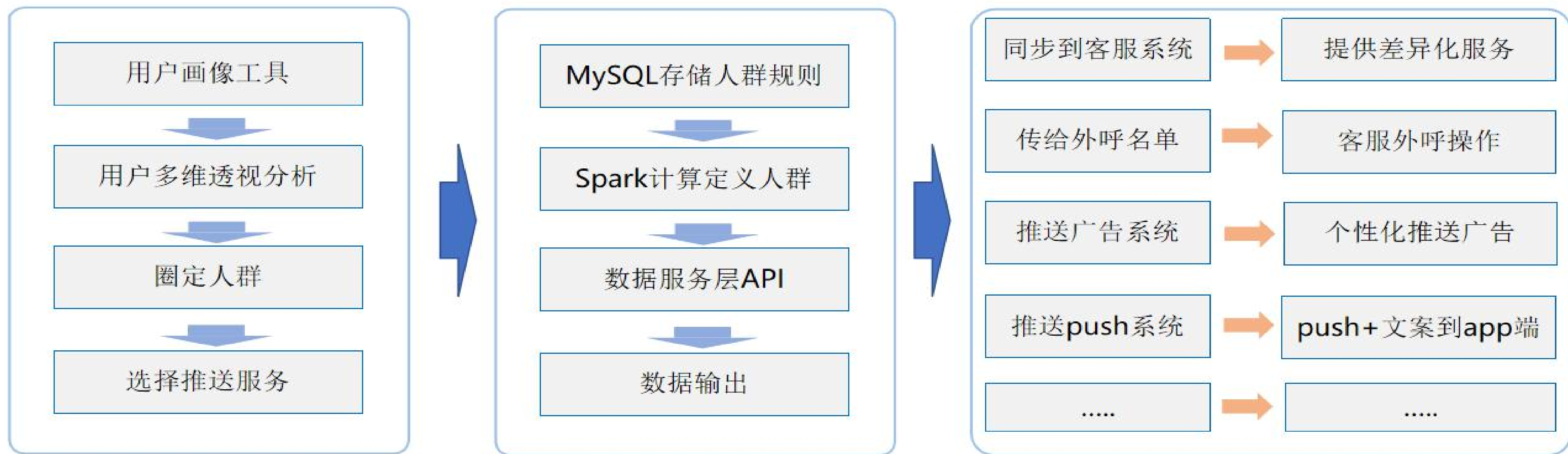
标签编辑管理中，开发人员可在web端编辑画像相关的元数据。编辑后，元数据插入到MySQL数据库中

	添加标签						
	标签id	标签名称	所属主题	一级大类	标签类型	开发方式	操作
标签视图	A111H001_001	男	用户属性	购物性别	分类型	统计	编辑 删除
标签查询	A111H001_002	女	用户属性	购物性别	分类型	统计	编辑 删除
标签编辑管理	A112H001_001	近30日浏览天数	用户行为	近30日活跃	统计型	统计	编辑 删除
用户分群	A112H001_002	近30日加购次数	用户行为	近30日活跃	统计型	统计	编辑 删除
用户分析	A112H001_003	近30日付款金额	用户行为	近30日活跃	统计型	统计	编辑 删除
	A112H001_004	近30日下单次数	用户行为	近30日活跃	统计型	统计	编辑 删除
	A113H001_001	首单距今天数	用户行为	用户活跃	统计型	统计	编辑 删除
	A114H001_001	是否短信黑名单	用户行为	黑名单	分类型	统计	编辑 删除
	A114H001_002	历史购买次数	用户属性	用户价值	统计型	统计	编辑 删除
<div><div><</div><div>1</div><div>2</div><div>3</div><div>4</div><div>5</div><div>6</div><div>...</div><div>29</div><div>></div></div>							

用户点击编辑按钮 可对已经添加的元数据进行重新编辑

自定义人群提供根据现有用户标签设置圈定用户群体的功能，业务人员可利用“**多维透视分析**”功能进行人群的对比分析，通过预计算对该运营规则圈定的人群数量做测算。保存后将生成圈定人群的规则，而后依据该规则生成圈定人群推送到服务端

下面介绍提供产品化服务的调度流程



Web端的用户画像产品功能

数据计算层逻辑

输出到服务端

点击添加分组，可重新添加人群的 分组。支持多标签组合进行个性化 定义人群

点击“编辑”或“删除”，可对已保存的人群分组进行重新编辑，或者删除该人群分组

点击推送到不同的业务系统时，会将对应人群的标签数据推送到相应的数据库中。如：外呼系统可通过FTP传输待呼用户的数据文件、客服系统对应MySQL数据库、push系统对应hbase数据库等

添加分组

标签视图

标签查询

标签编辑管理

用户分群

用户分析

人群名称	人群定义	创建时间	创建人	触达用户量	操作
高价值付费用户	XXXXXXXXX	2018-01-02 19:00	甲	17500	<div>编辑删除</div> <div>外呼系统客服系统广告系统push系统</div>
七天退款	XXXXXXXXX	2018-01-02 19:00	甲	30000	<div>编辑删除</div> <div>外呼系统客服系统广告系统push系统</div>
加购易放弃	XXXXXXXXX	2018-01-02 19:00	甲	2000	<div>编辑删除</div> <div>外呼系统客服系统广告系统push系统</div>
七天高拒签收	XXXXXXXXX	2018-01-02 19:00	乙	150	<div>编辑删除</div> <div>外呼系统客服系统广告系统push系统</div>
问题用户群	XXXXXXXXX	2018-01-02 19:00	乙	1500	<div>编辑删除</div> <div>外呼系统客服系统广告系统push系统</div>
高价值高活跃	XXXXXXXXX	2018-01-02 19:00	乙	100000	<div>编辑删除</div> <div>外呼系统客服系统广告系统push系统</div>
流失用户群	XXXXXXXXX	2018-01-02 19:00	甲	800000	<div>编辑删除</div> <div>外呼系统客服系统广告系统push系统</div>
高访问低下单	XXXXXXXXX	2018-01-02 19:00	乙	10000	<div>编辑删除</div> <div>外呼系统客服系统广告系统push系统</div>

<123456...29>

保存后对应的用户群组的名称，点击叉号可以删除该人群

点击新建人群，回到上页内容，组合标签创建对应的用户群组

选择分析该批人群的维度。
可供筛选的包括用户属性、
用户行为等不同主题下的
二级标签维度*或自己创建的*
人群；
筛选维度后，可从这些维
度透视查看该批用户群的
特征

点击可保存
该用户群

用户群透视分析

重要价值用户群x

+新建人群

重要价值用户群(总人数:325000)

用户属性

购买客单价

下单次数

购物性别

活跃度

RFM

历史购买状态

短信黑名单

邮件黑名单

√

√

对比维度:

活跃度x

查询

保存用户群

√

保存

标签视图

标签查询

标签编辑管理

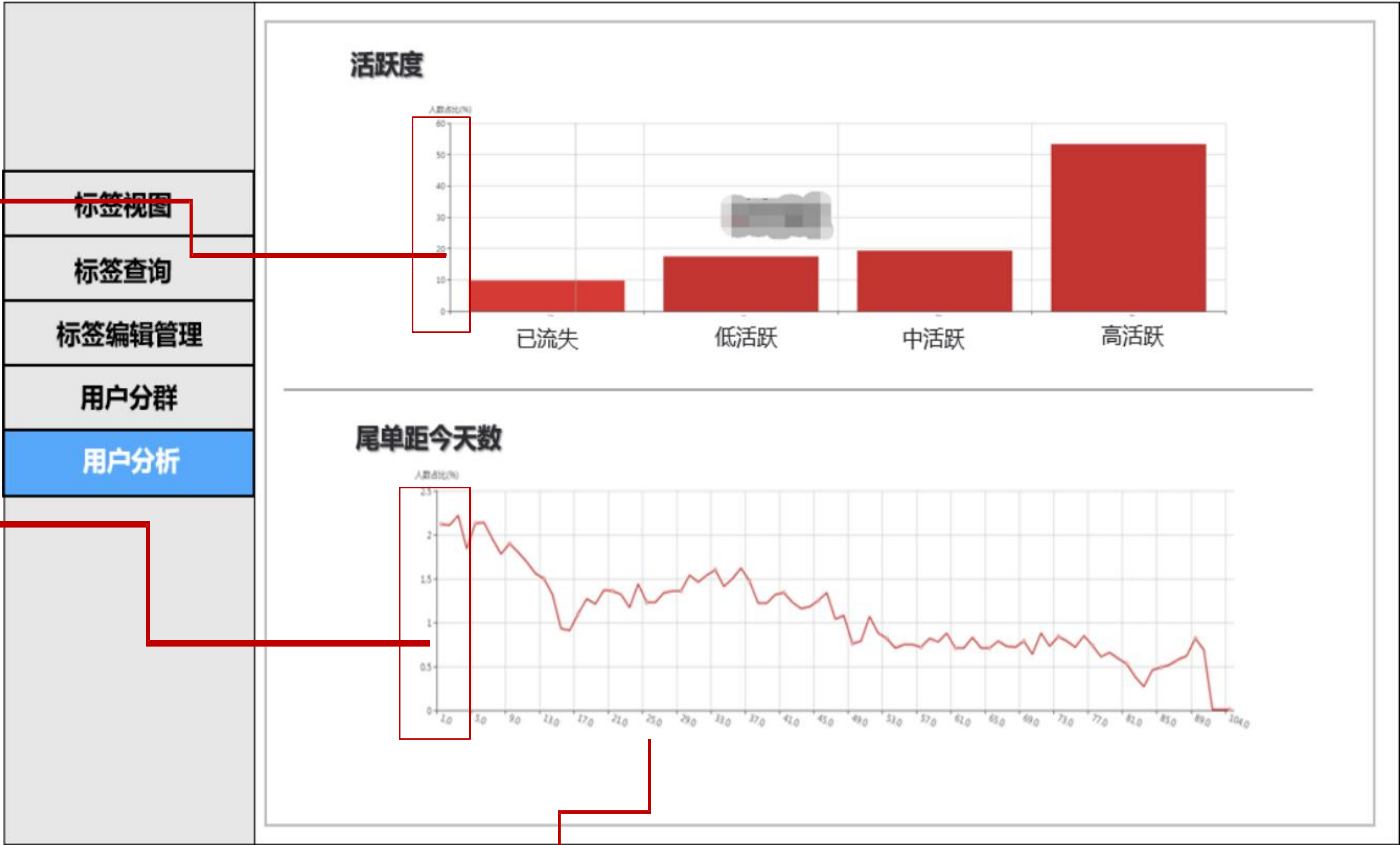
用户分群

用户分析

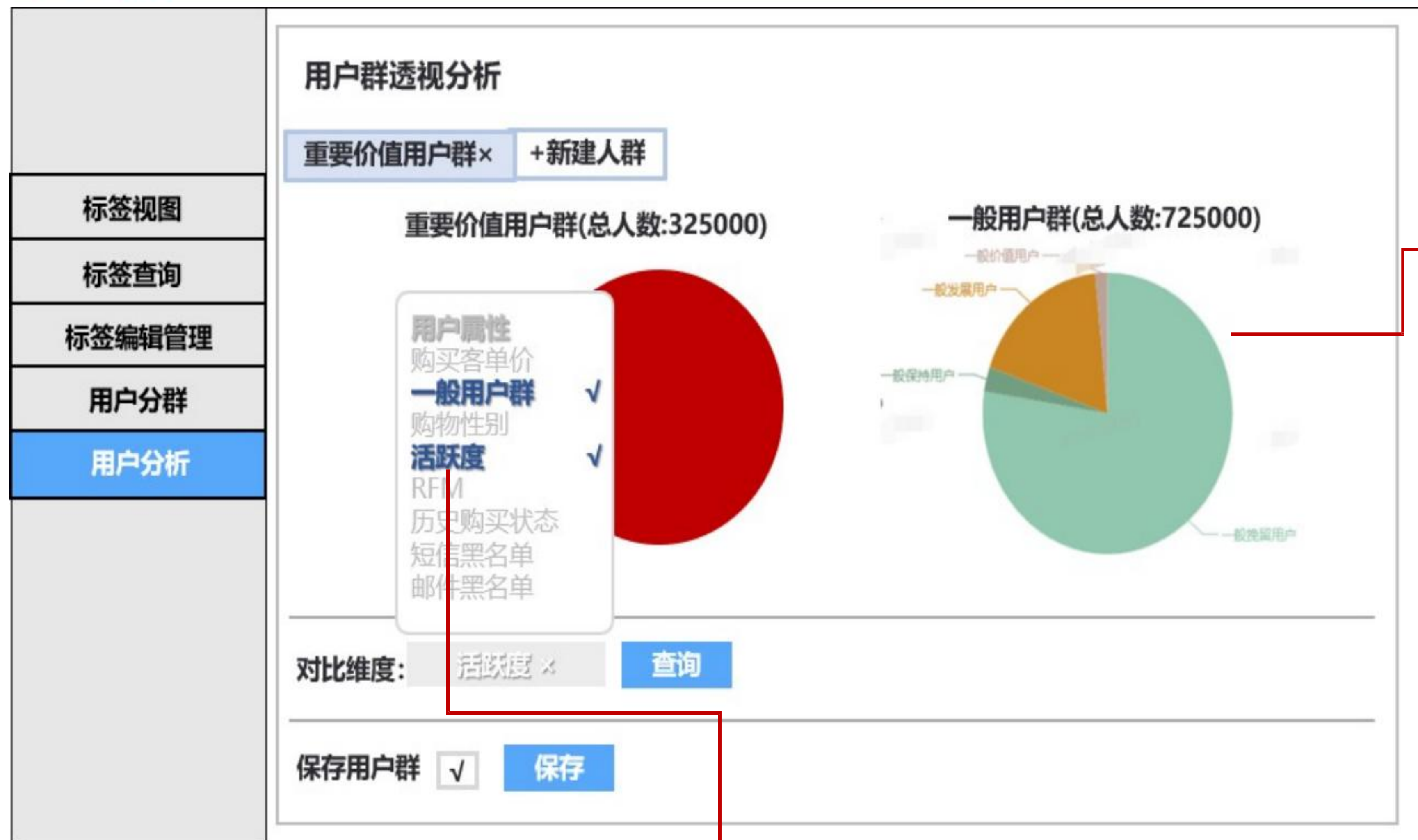
承接上页ppt，对筛选出来的用户群及分析该批用户群的维度，以可视化的方式展现该批用户群特征

不同活跃度情况对应人数占比

不同尾单距今天数
对应人数的占比



尾单距今天数



用户进入电商类网站或APP的一个典型流程包括，进入首页后搜索关键词、点击商品板块或点击推荐商品进入详情页，在详情页浏览点击加购后退出该页面搜索其它商品继续浏览，最后进入订单页进行支付，或浏览途中退出APP。这系列行为就是用户的行为轨迹，对于用户这样的访问行为，我们称为session。



Session中记录了用户在什么时间点、通过什么样的行为、浏览了什么页面/商品。一般session的切割为固定时长，如定义APP端session的切割时长为5分钟时，用户每次访问行为如果距离上一次访问行为在5分钟之内，则记为同一次访问，如果距离上次访问大于5分钟则记为两次不同的访问。通过session_id可用来标识用户的访问，同一次连续访问的session_id相同，否则不同。

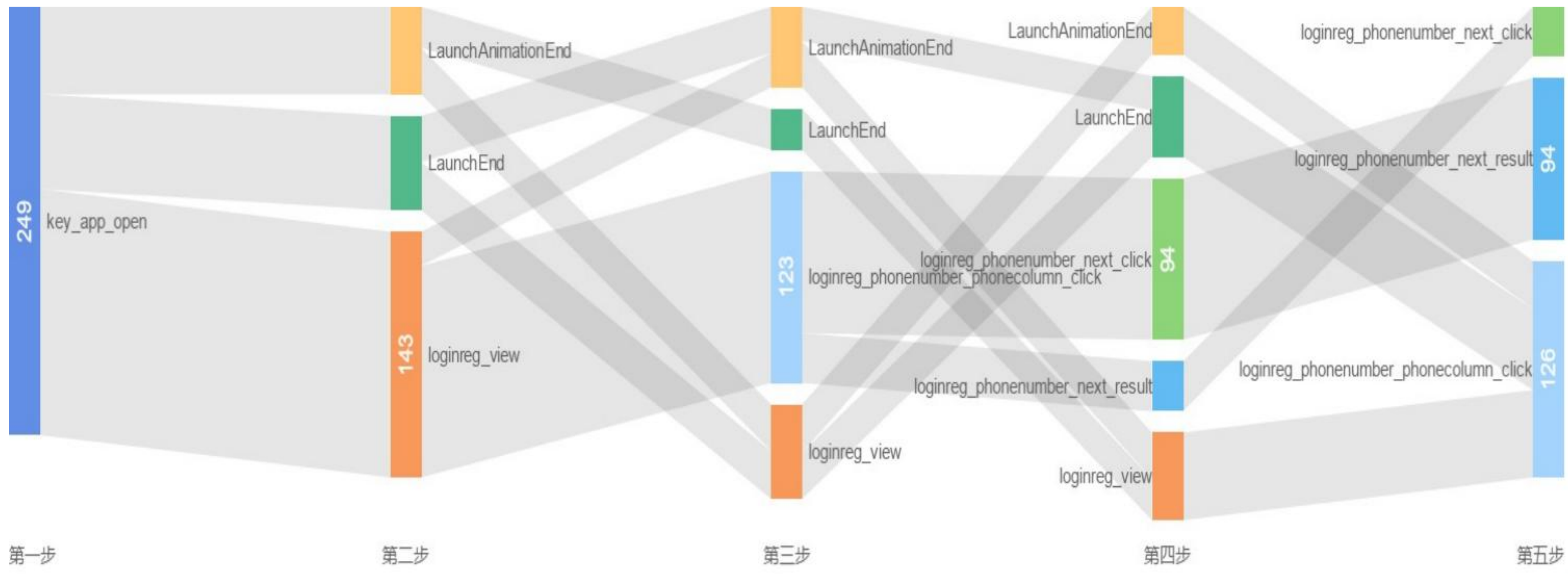
基于session对用户进行分析具有非常重要的作用，可以从用户的访问次数、访问路径、访问商品品类等多个维度分析用户特征。进一步地，分析用户首次访问的session对于挖掘影响用户购买行为具有重要的意义。

字段名称	中文名称	备注	示例
cookieid	用户id		d4a6-41e5-a670-85d217d055a6
eventkey1	第一次行为事件	浏览/收藏/加购等行为	add_to_bag
eventkey2	第二次行为事件	浏览/收藏/加购等行为	add_to_bag
eventkey3	第三次行为事件	浏览/收藏/加购等行为	add_to_bag
eventkey4	第四次行为事件	浏览/收藏/加购等行为	add_to_bag
eventkey5	第五次行为事件	浏览/收藏/加购等行为	add_to_bag
.....
level3_name	3级品类名称	用户行为浏览/收藏/加购等行为次数最多对应商品所属三级品类	运动鞋
goods_event_num	商品数量	用户行为对应商品数量	11
goods_list	商品清单	用户行为对应商品明细，记录商品id及对应行为次数	{001:3;002:5;003:11;...}
avg_origin_price	商品平均原价	浏览/收藏/加购等行为对应商品	66
avg_promote_price	商品平均促销价	浏览/收藏/加购等行为对应商品	59
price_range	价格区间	浏览/收藏/加购行为对应商品	[50,100)
access_time	浏览时长	访问从开始到结束时间	600
click_num	点击次数		32
goods_number	商品数量		16
is_paid	是否下单	记录用户是否下订单	1

cookieid	eventkey	level3_name	ods_event_n	goods_list	avg_origin_price	avg_promote_price	price_range	access_time	click_num	goods_number	is_paid	data_date
6945-4AAE-9911-FC1525850DA1	goodsdetail_view	洗发水	3	["3766198":"1"]	310	294	>200	180	34	4	0	20181225
4C5E-402F-8D8F-D955ADFF7FB3	addtobag_click	牙膏	3	["2226516":"1"]	9.99	0	[0.20]	788	91	1	0	20181225
8B2D-4715-BCF7-86A04A807107	addtobag_click	沐浴露	6	["1001174":"1","530748":"1"]	32.99	10	(20.40]	553	84	4	0	20181225
E0D4-464A-BA5B-99779B0161CF	goodsdetail_view	香皂	4	["4389212":"1"]	15.99	7.99	[0.20]	1620	216	13	0	20181225
E0D4-464A-BA5B-99779B0161CF	goodsdetail_view	礼盒	4	["4389212":"1"]	15.99	7.99	[0.20]	1620	216	13	0	20181225
C96B-4FBE-A0A0-D9D441F02B9A	goodsdetail_view	洗手液	8	["4365470":"1","4946434":"1"]	9.99	0	[0.20]	359	84	12	0	20181225
C96B-4FBE-A0A0-D9D441F02B9A	goodsdetail_view	洗手液	8	["4365470":"1","4946434":"1"]	9.99	0	[0.20]	359	84	12	0	20181225
DACC-4413-B22F-2249E32051C1	goodsdetail_view	卫生纸	3	["3582844":"1"]	24.99	0	(20.40]	945	109	8	0	20181225
FCCC-4068-97C8-FA0B60C73E4C	goodsdetail_view	厨房用纸	12	["1980724":"2","2946772":"1","4217	10.99	0.5	[0.20]	1161	167	35	1	20181225
9800-431B-8812-31D435AE5E0E	goodsdetail_view	洗洁精	11	["3685230":"1","1751462":"1","4774	139.72	23.03	(100.150]	825	169	47	0	20181225
D44D-4754-9BBE-213329C489E5	goodsdetail_view	洁厕剂	6	["172567":"1","596632":"1"]	26.49	13.49	(20.40]	897	135	25	0	20181225
8744-4D90-B28C-ADA1560A6078	addtobag_click	洁厕剂	3	["3661866":"1"]	9.99	4.99	[0.20]	799	152	2	0	20181225
8744-4D90-B28C-ADA1560A6078	addtobag_click	洗衣液	3	["3661866":"1"]	9.99	4.99	[0.20]	799	152	2	0	20181225
C28A-4B1B-82E0-4F747E819ED8	goodsdetail_view	彩妆工具	3	["2518138":"1"]	71	0	(60.80]	264	61	1	0	20181225
84E8-4F27-93D4-085863EAB225	goodsdetail_view	漱口水	4	["3272696":"1"]	23.99	0	(20.40]	248	65	4	0	20181225
B240-44DC-BBAC-37151BDBF7B6	goodsdetail_view	美发工具	3	["1656288":"3"]	769	0	>200	1512	197	11	0	20181225
2795-4ED7-B15B-3BE334687CE2	goodsdetail_view	香皂	6	["4336200":"1","4072692":"1"]	71.5	34.5	(60.80]	489	114	14	1	20181225
067F-4F0A-B609-BACB981A2BE2	goodsdetail_view	礼盒	3	["3624142":"1"]	34.99	28.99	(20.40]	1220	98	18	0	20181225
7627-40FB-949D-CAEC4A3B4023	goodsdetail_view	洗手液	7	["4622712":"2","197939":"1"]	21.7	15.7	(20.40]	269	29	3	0	20181225
7627-40FB-949D-CAEC4A3B4023	goodsdetail_view	洗手液	7	["4622712":"2","197939":"1"]	21.7	15.7	(20.40]	269	29	3	0	20181225
E020-4F91-996A-A07D70C1FEF1	goodsdetail_view	面膜	17	["4562986":"3","3479876":"6","4942	21.4	9.64	(20.40]	484	146	21	0	20181225
E020-4F91-996A-A07D70C1FEF1	goodsdetail_view	精华	17	["4562986":"3","3479876":"6","4942	21.4	9.64	(20.40]	484	146	21	0	20181225
5029-4AA9-8E5E-076D4C75602C	goodsdetail_view	防晒	3	["4190982":"2"]	52	38.99	(40.60]	307	73	6	0	20181225
6624-452B-B7E7-1B97F47462BF	goodsdetail_view	爽肤水	3	["2790564":"1"]	30.99	6.99	(20.40]	413	40	4	0	20181225
6624-452B-B7E7-1B97F47462BF	goodsdetail_view	美甲	3	["2790564":"1"]	30.99	6.99	(20.40]	413	40	4	1	20181225
32DC-4B99-8687-71B143E40671	goodsdetail_view	男士香水	4	["4656856":"1"]	37.99	22.99	(20.40]	419	66	6	0	20181225
7114-4AC8-82A6-383198C069F6	goodsdetail_view	垃圾袋	3	["3291054":"1"]	6.99	2.99	[0.20]	255	89	6	0	20181225
6847-47A2-8A01-93AB0153C1BF	goodsdetail_view	卫生巾	4	["4496272":"1"]	304	288.99	>200	158	60	1	0	20181225
6847-47A2-8A01-93AB0153C1BF	goodsdetail_view	女生香水	4	["4496272":"1"]	304	288.99	>200	158	60	1	0	20181225
6FBE-4646-ADD8-6EFF5D2CF919	goodsdetail_view	洗手液	3	["977190":"4"]	33.99	27.99	(20.40]	135	59	4	1	20181225
FD17-420A-962B-3A2832517FC9	addtobag_click	洗手液	3	["2225960":"2"]	35.99	0	(20.40]	807	257	2	0	20181225
AD1E-482F-A55A-423D347991CE	goodsdetail_view	面膜	19	["1327832":"2","3634942":"2","1577	32.36	18.71	(20.40]	635	194	25	0	20181225
F085-4302-8EF8-5DEC8C8DBDF2	goodsdetail_view	精华	3	["2557358":"1"]	80	37.99	(60.80]	504	72	5	0	20181225
DB98-4CD4-A3E9-517CC055387F	goodsdetail_view	防晒	4	["4830686":"1"]	19.43	14.9	[0.20]	144	35	3	0	20181225
9442-44BF-89C2-68361176C2DD	addtobag_click	垃圾袋	4	["4480258":"1"]	22.99	9.99	(20.40]	497	88	2	0	20181225
4C79-4D80-9688-A06926B98BE1	goodsdetail_view	卫生巾	27	["231855":"1","2535354":"1","23337	33.58	19.77	(20.40]	1592	249	34	0	20181225
2B19-490C-991A-4390120D3C1D	addtobag_click	女生香水	4	["4764064":"1"]	13.04	11.18	[0.20]	326	83	1	0	20181225
31D6-408E-ABDE-5ABDD812DDFE	goodsdetail_view	洗手液	4	["4705066":"1"]	29.99	0	(20.40]	178	91	2	0	20181225
D7F6-4FA5-AE54-57AF407D808A	goodsdetail_view	洗手液	32	["3259126":"1","4864050":"1","4861	11.18	2.56	[0.20]	1399	178	40	0	20181225
D7F6-4FA5-AE54-57AF407D808A	goodsdetail_view	卫生纸	32	["3259126":"1","4864050":"1","4861	11.18	2.56	[0.20]	1399	178	40	1	20181225
83DB-4549-ACE6-7DA822DD685F	goodsdetail_view	厨房用纸	14	["4382552":"1","2809962":"1","8556	41.44	12.63	(40.60]	380	89	5	0	20181225
0A5F-415A-A40D-88F1D145C1CD	goodsdetail_view	洗洁精	4	["5610018":"1"]	59	37.99	(40.60]	1399	79	21	0	20181225

根据用户的访问路径分析，对于产品设计的改进有很大帮助，分析用户从登录，搜索，浏览详情页到购买的行为路径，根据用户在各环节的转化率发现用户行为偏好和影响订单转换的主要因素。

					eventLst num			
				key_app_open,loginreg_view,null,null,null	136			
key_app_open,	loginreg_view,	loginreg_phonenumber_phonecolumn_click,	loginreg_phonenumber_next_click,	loginreg_phonenumber_next_result	94			
key_app_open,	loginreg_view,	loginreg_phonenumber_phonecolumn_click,	loginreg_phonenumber_next_result,	loginreg_phonenumber_next_click	29			
		key_app_open,LaunchEnd,	loginreg_view,LaunchAnimationEnd,	loginreg_phonenumber_phonecolumn_click	28			
		key_app_open,LaunchEnd,	LaunchAnimationEnd,	loginreg_view,loginreg_phonenumber_phonecolumn_click	27			
		key_app_open,LaunchAnimationEnd,	loginreg_view,LaunchEnd,	loginreg_phonenumber_phonecolumn_click	27			
			key_app_open,loginreg_view,	loginreg_phonenumber_phonecolumn_click,null,null	26			
				key_app_open,null,null,null,null	25			
		key_app_open,LaunchAnimationEnd,	LaunchEnd,	loginreg_view,loginreg_phonenumber_phonecolumn_click	24			
		key_app_open,loginreg_view,	LaunchAnimationEnd,	LaunchEnd,	loginreg_phonenumber_phonecolumn_click	20		
		key_app_open,loginreg_view,	LaunchEnd,	LaunchAnimationEnd,	loginreg_phonenumber_phonecolumn_click	18		
			key_app_open,LaunchEnd,	loginreg_view,LaunchAnimationEnd,	guidepage_skip	17		
		key_app_open,	key_app_open,	loginreg_view,loginreg_view,	loginreg_phonenumber_phonecolumn_click	17		
key_app_open,	loginreg_view,	loginreg_phonenumber_phonecolumn_click,	loginreg_phonenumber_next_click,	loginreg_verification_sendsms...	15			
				key_app_open,LaunchAnimationEnd,	null,null,null	15		
			key_app_open,	loginreg_view,loginreg_phonenumber_next_click,	null,null	14		
			key_app_open,	LaunchAnimationEnd,	LaunchEnd,	loginreg_view,guidepage_skip	13	
			key_app_open,	LaunchAnimationEnd,	loginreg_view,LaunchEnd,	guidepage_skip	13	
			key_app_open,	LaunchEnd,	LaunchAnimationEnd,	loginreg_view,guidepage_skip	12	
			key_app_open,	loginreg_view,	LaunchEnd,	LaunchAnimationEnd,	guidepage_skip	11



用户访问session流量桑基图



谢谢大家

