**Task 2**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 联表查询 |  |  |  |  |  |  |
| 通常，在我们洗完了驱动表之后，我们需要将之与不同维度的表去做联表查询，从而统计黑白的标签在不同维度的特征下的分布情况（这里指的统计标签  ，更多的是看账户维度和事件维度的打标情况，比如同一个账户可能标记了好几个黑事件，也可能标记了好几个白事件）   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | 步骤 |  |  |  |  |  |  | | 1. 将上一步的驱动表分别和设备信息表、会员信息表关联，增量表时间窗口选取为20200101-20200114天，全量表选取20200114的日期即可 | | | | | | | | 注意：与设备信息表关联时用设备指纹 device\_sign进行关联，与会员信息表关联时则用 userid进行关联 | | | | | |  | |  |  |  |  |  |  |  | | 2. 统计特征的打标情况，并对设备信息表的字段做分箱，同时统计：黑事件数、白事件数、总事件数、黑账户数、白账户数、总账户数 | | | | | | | | | | | | | |

【分箱在sql里面我们可以通过case when来简单地实现分bin】

**用到的表：**

**由task1 得到的驱动表：drivertable，**

**ods\_apdid\_request\_table 设备信息表 （增量表，宽表）**

**acc\_base\_dd 会员信息表 （全量表，较大，PB级别）**

1. **设备信息表中的字段包括但不限于：**

**f\_apdid 设备指纹 和device\_sign 一样，只不过在不同的表里称呼不一样**

**f\_ae2 是否root Y或者N**

**f\_ie2 是否hook Y或者N**

**f\_ae10 设备具体型号**

**f\_ad109 设备是否有被模拟行为，篡改地理位置 1或0**

**【root和hook字段分别代表着：root意味着是否有越狱，如果root为Y，则检测到大概率设备已经越狱；与之相似，hook意味着有无刷机改机，如果hook为Y，则检测到大概率设备已经被刷机改机或者有篡改行为发生】**

1. **会员信息表中的字段包括但不限于**

**user\_create\_date 用户注册日期**

**【drivertable 里面理论上也包含target字段，就是所谓的打好了标签的别名字段 这里只是一个示例】**

**drop table if exists equipment\_drivertable;**

**create table equipment\_drivertable as  -- 关联设备信息表**

**select ods\_apdid\_request\_table.\*,drivertable.target**

**from ods\_apdid\_request\_table**

**left join drivertable on ods\_apdid\_request\_table.f\_apdid = drivertable.devise\_sign where ods\_apdid\_request\_table.dt between '20200101' and '20200114';**

**select**

**sum(CASE when target = 1 then '1' else '0' end) as count\_black,  -- 黑事件数**

**sum(CASE when target = 0 then '1' else '0' end) as count\_white, -- 白事件数**

**count(target) as count\_all, -- 总事件数**

**(select count(distinct(f\_apdid)) from equipment\_drivertable where target = 1) as black\_account, -- 黑账户数**

**(select count(distinct(f\_apdid)) from equipment\_drivertable where target = 0) as white\_account, -- 白帐户数**

**(select count(distinct(f\_apdid)) from equipment\_drivertable ) as all\_account – 总账户数**

**from equipment\_drivertable**

**drop table if exists member\_drivertable;**

**create table member\_drivertable as   -- 关联会员信息表**

**select acc\_base\_dd.\*,drivertable.target**

**from acc\_base\_dd**

**left join drivertable on acc\_base\_dd.user\_id = drivertable.userid where acc\_base\_dd.dt = '20200114';  -- 关联会员信息表**